



# (12) 发明专利申请

(10) 申请公布号 CN 113348245 A

(43) 申请公布日 2021.09.03

(21) 申请号 201980064619.X

(22) 申请日 2019.07.31

(30) 优先权数据

62/712,809 2018.07.31 US

62/751,421 2018.10.26 US

62/775,865 2018.12.05 US

62/822,639 2019.03.22 US

62/873,031 2019.07.11 US

(85) PCT国际申请进入国家阶段日

2021.03.30

(86) PCT国际申请的申请数据

PCT/US2019/044480 2019.07.31

(87) PCT国际申请的公布数据

W02020/028555 EN 2020.02.06

(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 I·史雷梅克 S·坎南

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)

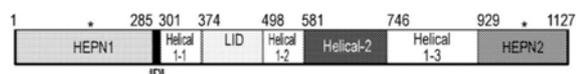
权利要求书30页 说明书434页 附图239页

(54) 发明名称

新型CRISPR酶和系统

(57) 摘要

本公开内容提供了用于靶向核酸的系统、方法和组合物。特别地,本发明提供了突变的Cas13蛋白及其在修饰靶序列中的用途,以及突变的Cas13核酸序列和编码突变的Cas13蛋白的载体,以及载体系统或CRISPR-Cas13系统。



1. 一种工程化腺苷脱氨酶,其包含一个或多个突变,其中所述工程化腺苷脱氨酶具有胞苷脱氨酶活性,其中所述腺苷脱氨酶蛋白或其催化结构域基于hADAR2-D的氨基酸序列位置包含突变E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T中的一个或多个,以及在同源ADAR蛋白中的相应突变。

2. 一种工程化腺苷脱氨酶,其包含一个或多个突变,其中所述工程化腺苷脱氨酶具有胞苷脱氨酶活性。

3. 根据权利要求2所述的工程化腺苷脱氨酶,其中所述工程化腺苷脱氨酶具有腺苷脱氨酶活性。

4. 根据权利要求2所述的工程化腺苷脱氨酶,其中所述工程化腺苷脱氨酶是融合蛋白的一部分。

5. 根据权利要求2所述的工程化腺苷脱氨酶,其中所述融合蛋白包含功能性结构域。

6. 根据权利要求2所述的工程化腺苷脱氨酶,其中所述功能性结构域能够引导所述工程化腺苷脱氨酶结合至靶核酸。

7. 根据权利要求2所述的工程化腺苷脱氨酶,其中所述功能性结构域是权利要求50至55中任一项所述的CRISPR-Cas蛋白。

8. 根据权利要求2所述的工程化腺苷脱氨酶,其中所述CRISPR-Cas蛋白是死形式的CRISPR-Cas蛋白或CRISPR-Cas切口酶蛋白。

9. 根据权利要求2所述的工程化腺苷脱氨酶,其中基于hADAR2-D的氨基酸序列位置,所述一个或多个突变包含:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T,以及在同源ADAR蛋白中的相应突变。

10. 根据权利要求2所述的工程化腺苷脱氨酶,其中基于hADAR2-D的氨基酸序列位置,所述一个或多个突变包含:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E和S661T,以及在同源ADAR蛋白中的相应突变。

11. 一种多核苷酸,其编码上述任一项权利要求所述的工程化腺苷脱氨酶或其催化结构域。

12. 一种载体,其包含权利要求11所述的多核苷酸。

13. 一种药物组合物,其包含配制成用于通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪或可植入装置递送的权利要求1-10中任一项所述的工程化腺苷脱氨酶或其催化结构域。

14. 一种工程化细胞,其表达权利要求1-10中任一项所述的工程化腺苷脱氨酶或其催化结构域。

15. 根据权利要求14所述的工程化细胞,其中所述细胞瞬时表达所述工程化腺苷脱氨酶或其催化结构域。

16. 根据权利要求15所述的工程化细胞,其中所述细胞非瞬时表达所述工程化腺苷脱氨酶或其催化结构域。

17. 一种用于在靶核酸中修饰核苷酸的工程化、非天然存在的系统,其包含:

a) 死CRISPR-Cas或CRISPR-Cas切口酶蛋白,或者编码所述死Cas或Cas切口酶蛋白的核苷酸序列;

b) 指导分子,其包含与靶序列杂交并设计为与所述死CRISPR-Cas或CRISPR-Cas切口酶蛋白形成复合物的指导序列;和

c) 核苷酸脱氨酶蛋白或其催化结构域,或者编码所述核苷酸脱氨酶蛋白或其催化结构域的核苷酸序列,其中所述核苷酸脱氨酶蛋白或其催化结构域共价或非共价连接至所述死CRISPR-Cas或CRISPR-Cas切口酶蛋白,或者所述指导分子适于在递送后与其连接。

18. 根据权利要求17所述的系统,其中所述腺苷脱氨酶蛋白或其催化结构域基于hADAR2-D的氨基酸序列位置包含突变E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T中的一个或多个,以及以及在同源ADAR蛋白中的相应突变。

19. 根据权利要求17所述的系统,其中所述腺苷脱氨酶蛋白或其催化结构域基于hADAR2-D的氨基酸序列位置包含突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E和S661T,以及在同源ADAR蛋白中的相应突变。

20. 根据权利要求17所述的系统,其中所述CRISPR-Cas蛋白是Cas9、Cas12、Cas13、Cas14、CasX或CasY。

21. 根据权利要求17所述的系统,其中所述CRISPR-Cas蛋白是Cas13b。

22. 根据权利要求17所述的系统,其中所述CRISPR-Cas蛋白是Cas13b-t1、Cas13b-t2或Cas13b-t3。

23. 根据权利要求17所述的系统,其中所述CRISPR-Cas是根据权利要求50至367中任一项所述的工程化CRISPR-Cas蛋白。

24. 一种用于在靶核酸中修饰核苷酸的方法,其包括:

将权利要求1-10中任一项所述的工程化腺苷脱氨酶或权利要求17-23中任一项所述的系统递送至所述靶核酸,其中所述脱氨酶使在所述靶核酸上的一个或多个靶基因座处的核苷酸脱氨基。

25. 根据权利要求24所述的方法,其中所述核苷酸脱氨酶蛋白或其催化结构域已被修饰以增加对DNA-RNA异源双链体的活性。

26. 根据权利要求24所述的方法,其中所述核苷酸脱氨酶蛋白或其催化结构域已被修饰以降低脱靶效应。

27. 根据权利要求24所述的方法,其中所述靶核酸在细胞内。

28. 根据权利要求24所述的方法,其中所述细胞是真核细胞。

29. 根据权利要求24所述的方法,其中所述细胞是非人动物细胞。

30. 根据权利要求24所述的方法,其中所述细胞是人细胞。

31. 根据权利要求24所述的方法,其中所述细胞是植物细胞。

32. 根据权利要求24所述的方法,其中所述靶核酸在动物内。

33. 根据权利要求24所述的方法,其中所述靶核酸在植物内。

34. 根据权利要求24所述的方法,其中所述靶核酸在体外包含在DNA分子中。

35. 根据权利要求24所述的方法,其中所述工程化腺苷脱氨酶或者所述系统的一种或

多种组分作为核糖核蛋白复合物递送到所述细胞中。

36. 根据权利要求24所述的方法,其中所述工程化腺苷脱氨酶或者所述系统的一种或多种组分通过一种或多种颗粒、一种或多种囊泡或者一种或多种病毒载体递送。

37. 根据权利要求24所述的方法,其中所述一种或多种颗粒包含脂质、糖、金属或蛋白。

38. 根据权利要求24所述的方法,其中所述一种或多种颗粒包括脂质纳米颗粒。

39. 根据权利要求24所述的方法,其中所述一种或多种囊泡包括外泌体或脂质体。

40. 根据权利要求24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载体,一种或多种慢病毒载体或者一种或多种腺相关病毒载体。

41. 根据权利要求24所述的方法,其中所述方法通过在感兴趣的基因组基因座处操纵一个或多个靶序列来修饰细胞、细胞系或生物体。

42. 根据权利要求24所述的方法,其中在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用治疗由G→A或C→T点突变或病原性SNP引起的疾病。

43. 根据权利要求24所述的方法,其中所述疾病选自癌症、血友病、β地中海贫血,马凡氏综合征和维斯科特-奥尔德里奇综合征。

44. 根据权利要求24所述的方法,其中在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用治疗由T→C或A→G点突变或病原性SNP引起的疾病。

45. 根据权利要求24所述的方法,其中在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用使在所述靶基因座处的靶基因失活。

46. 根据权利要求24所述的方法,其中通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪、可植入装置或权利要求302所述的载体系统递送所述工程化腺苷脱氨酶或者所述系统的一种或多种组分。

47. 根据权利要求24所述的方法,其中所述核苷酸的修饰改变在所述靶基因座处编码的基因产物或所述基因产物的表达。

48. 根据权利要求1-10中任一项所述的工程化腺苷脱氨酶或权利要求17-23中任一项所述的系统,其中所述腺苷蛋白或其催化结构域基于hADAR2-D的氨基酸序列位置包含在S375上的突变,以及在同源ADAR蛋白中的相应突变。

49. 根据权利要求48所述的工程化腺苷脱氨酶或系统,其中所述在S375上的突变是S375N。

50. 一种工程化CRISPR-Cas蛋白,其包含一个或多个HEPN结构域,以及还包含一个或多个修饰的氨基酸,其中所述氨基酸:

- a. 与指导RNA相互作用,所述指导RNA与所述工程化CRISPR-Cas蛋白形成复合物;
- b. 在所述工程化CRISPR-Cas蛋白的HEPN活性位点、结构域间接头结构域、盖结构域、螺旋结构域1、螺旋结构域2或桥螺旋结构域中;或者
- c. 其组合。

51. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中所述HEPN结构域包含RxxxxH基序。

52. 根据权利要求51所述的工程化CRISPR-Cas蛋白,其中所述RxxxxH基序包含R{N/H/K}X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H序列。

53. 根据权利要求52所述的工程化CRISPR-Cas蛋白,其中:

X<sub>1</sub>是R、S、D、E、Q、N、G或Y，

X<sub>2</sub>独立地是I、S、T、V或L，和

X<sub>3</sub>独立地是L、F、N、Y、V、I、S、D、E或A。

54. 根据权利要求50所述的工程化CRISPR-Cas蛋白，其中所述CRISPR-Cas蛋白是VI型CRISPR-Cas蛋白。

55. 根据权利要求54所述的工程化CRISPR-Cas蛋白，其中所述VI型CRISPR-Cas蛋白是Cas13。

56. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其中所述VI型CRISPR-Cas蛋白是Cas13a、Cas13b、Cas13c或Cas13d。

57. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变：T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、A656、V795、A796、W842、K871、E873、R874、R1068、N1069或H1073。

58. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变：H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、W842、K871、E873、R874、R1068、N1069或H1073。

59. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变：T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294或E400。

60. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或H1073。

61. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、H407、S658、N653、K655、N652、H567、

N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或H1073。

62. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

63. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

64. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

65. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在桥螺旋结构域中包含对应于PbCas13b的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

66. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N480、N482、N652或N653。

67. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N480或N482。

68. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N480或N482。

69. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:N652或N653。

70. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:N652或N653。

71. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

72. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

73. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、S757、N756或K741。

74. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于PbCas13b的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、

R791、G566、K590、R638、S757、N756或K741。

75. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或N756。

76. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或N756。

77. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、V795、A796、R791、G566、S757或N756。

78. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、V795、A796、R791、G566、S757或N756。

79. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846或R874。

80. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在桥螺旋结构域中包含对应于PbCas13b的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846或R874。

81. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500或G566。

82. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-2中包含对应于PbCas13b的螺旋结构域1-2中的下述氨基酸的氨基酸的一个或多个突变:H567、H500或G566。

83. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。

84. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。

85. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R762、V795、A796、R791、S757或N756。

86. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:R762、V795、A796、R791、S757或N756。

87. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、K590、R638或K741。

88. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、K590、R638或K741。

89. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨

基酸的氨基酸的一个或多个突变:T405、H407、N486、K484、N480、H452、N455或K457。

90. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:T405、H407、N486、K484、N480、H452、N455或K457。

91. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或K741。

92. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于PbCas13b的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或K741。

93. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。

94. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。

95. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、R791、G566、S757或N756。

96. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、R791、G566、S757或N756。

97. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。

98. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。

99. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R762、R791、S757或N756。

100. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:R762、R791、S757或N756。

101. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、K590、R638或K741。

102. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、K590、R638或K741。

103. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨

基酸的氨基酸的一个或多个突变:H407、N486、K484、N480、H452、N455或K457。

104. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:H407、N486、K484、N480、H452、N455或K457。

105. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R56、N157、H161、R1068、N1069或H1073。

106. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于PbCas13b的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R56、N157、H161、R1068、N1069或H1073。

107. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R56、N157或H161。

108. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于PbCas13b的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R56、N157或H161。

109. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R1068、N1069或H1073。

110. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于PbCas13b的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:R1068、N1069或H1073。

111. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。

112. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。

113. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。

114. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。

115. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

116. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

117. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨

基酸的氨基酸的一个或多个突变：S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

118. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

119. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：N486、K484、N480、H452、N455或K457。

120. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变：N486、K484、N480、H452、N455或K457。

121. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、N486、K484、N480、H452、N455或K457。

122. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、N486、K484、N480、H452、N455或K457。

123. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

124. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

125. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变：R53、Y164、K943或R1041。

126. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变：R53或Y164。

127. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变：K943或R1041。

128. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其在HEPN结构域中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变：R53、Y164、K943或R1041。

129. 根据权利要求55所述的工程化CRISPR-Cas蛋白，其在HEPN结构域1中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨

氨酸的一个或多个突变:R53或Y164。

130. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:K943或R1041。

131. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

132. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、R56、N157或H161。

133. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

134. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

135. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、R56、N157或H161。

136. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

137. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943或R1041。

138. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183或K193。

139. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K943或R1041。

140. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943或R1041。

141. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183或K193。

142. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨

氨酸的一个或多个突变:K943 or R1041。

143. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

144. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、R56、N157或H161。

145. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

146. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

147. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、R56、N157或H161。

148. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

149. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K183或K193。

150. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:K183或K193。

151. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943或R1041。

152. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943或R1041。

153. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、K943或R1041; 优选地R53A、R53K、R53D或R53E; K943A、K943R、K943D或K943E; 或者R1041A、R1041K、R1041D或R1041E。

154. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、K943或R1041; 优选地R53A、R53K、R53D或R53E; K943A、K943R、K943D

或K943E;或者R1041A、R1041K、R1041D或R1041E。

155. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸Y164的氨基酸的突变,优选地Y164A、Y164F或Y164W。

156. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1的氨基酸Y164的氨基酸的突变,优选地Y164A、Y164F或Y164W。

157. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398或E399。

158. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398或E399。

159. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸H407的氨基酸的突变,优选地H407Y、H407W或H407F。

160. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R402、K393、R482、N480、D396、E397、D398或E399。

161. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:R402、K393、R482、N480、D396、E397、D398或E399。

162. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K457、D434或K431。

163. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K457、D434或K431。

164. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。

165. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。

166. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

167. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

168. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762或R791。

169. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762或R791。

170. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

171. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在桥螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

172. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500或K570。

173. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-2中的下述氨基酸的氨基酸的一个或多个突变:H500或K570。

174. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

175. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

176. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762或R791。

177. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应

于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762或R791。

178. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870或R877。

179. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870或R877。

180. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K826、K828、K829、R824、R830、Q831、K835、K836或R838。

181. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:K826、K828、K829、R824、R830、Q831、K835、K836或R838。

182. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653或N652。

183. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653或N652。

184. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:Q646或N647。

185. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:Q646或N647。

186. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N653或N652。

187. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:N653或N652。

188. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741或K744。

189. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨

氨酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741或K744。

190. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R600、K607、K612、R614、K617或R618。

191. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:R600、K607、K612、R614、K617或R618。

192. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R285、R287、K292、E296、N297或K294。

193. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在IDL结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的IDL结构域中的下述氨基酸的氨基酸的一个或多个突变:R285、R287、K292、E296、N297或K294。

194. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R285、K292、E296或N297。

195. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在IDL结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的IDL结构域中的下述氨基酸的氨基酸的一个或多个突变:R285、K292、E296或N297。

196. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:T405、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R285、R287、K292、E296、N297、Q646、N647或K294。

197. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R402、K393、N653、N652、R482、N480、D396、E397、D398或E399。

198. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、K655、R762或R1041;优选地R53A或R53D;K655A;R762A;或者R1041E或R1041D。

199. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N297、E296、K292或R285;优选地N297A、E296A、K292A或R285A。

200. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在IDL结构域(的中央通道)中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的IDL结构域(的中央通道)中的下述氨基酸的氨基酸的一个或多个突变:N297、E296、K292或R285;优选地N297A、E296A、K292A或R285A。

201. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌

(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:Q831、K836、R838、N652、N653、R830、K655或R762;优选地Q831A、K836A、R838A、N652A、N653A、R830A、K655A或R762A。

202. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N652、N653、R830、K655 or R762;优选地N652A、N653A、R830A、K655A或R762A。

203. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K655或R762;优选地K655A或R762A。

204. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:Q831、K836、R838、N652、N653、R830、K655或R762;优选地Q831A、K836A、R838A、N652A、N653A、R830A、K655A或R762A。

205. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:N652、N653、R830、K655或R762;优选地N652A、N653A、R830A、K655A或R762A。

206. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:K655或R762;优选地K655A或R762A。

207. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R614、K607、K193、K183或R600;优选地R614A、K607A、K193A、K183A或R600A。

208. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2的反式亚基环中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2的反式亚基环中的下述氨基酸的氨基酸的一个或多个突变:Q646或N647;优选地Q646A或N647A。

209. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53或R1041;优选地R53A或R53D或R1041E或R1041D。

210. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53或R1041;优选地R53A或R53D或R1041E或R1041D。

211. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K457、D397、E398、D399、E400、T405、H407或D434;优选地D397A、E398A、D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。

212. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K457、D397、E398、D399、E400、T405、H407或D434;优选地D397A、E398A、

D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。

213. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其中所述氨基酸对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸:氨基酸46-57、73-79、152-164、1036-1046和1064-1074。

214. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R156、N157、H161、R1068、N1069和H1073。

215. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R285、R287、K292、K294、E296和N297。

216. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K826、K828、K829、R824、R830、Q831、K835、K836和R838。

217. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870和R877。

218. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸T405的氨基酸的突变。

219. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸H407的氨基酸的突变。

220. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K457的氨基酸的突变。

221. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸H500的氨基酸的突变。

222. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K570的氨基酸的突变。

223. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K590的氨基酸的突变。

224. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N634的氨基酸的突变。

225. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R638的氨基酸的突变。

226. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N652的氨基酸的突变。

227. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N653的氨基酸的突变。

228. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K655的氨基酸的突变。

229. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸S658的氨基酸的突变。

230. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K741的氨基酸的突变。

231. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K744的氨基酸的突变。

232. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N756的氨基酸的突变。

233. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸S757的氨基酸的突变。

234. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R762的氨基酸的突变。

235. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R791的氨基酸的突变。

236. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K846的氨基酸的突变。

237. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K857的氨基酸的突变。

238. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K870的氨基酸的突变。

239. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R877的氨基酸的突变。

240. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K183的氨基酸的突变。

241. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K193的氨基酸的突变。

242. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R600的氨基酸的突变。

243. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K607的氨基酸的突变。

244. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K612的氨基酸的突变。

245. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R614的氨基酸的突变。

246. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K617的氨基酸的突变。

247. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K826的氨基酸的突变。

248. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌

(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K828的氨基酸的突变。

249. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K829的氨基酸的突变。

250. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R824的氨基酸的突变。

251. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R830的氨基酸的突变。

252. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸Q831的氨基酸的突变。

253. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K835的氨基酸的突变。

254. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K836的氨基酸的突变。

255. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R838的氨基酸的突变。

256. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R618的氨基酸的突变。

257. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸D434的氨基酸的突变。

258. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K431的氨基酸的突变。

259. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R53的氨基酸的突变。

260. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K943的氨基酸的突变。

261. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R1041的氨基酸的突变。

262. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸Y164的氨基酸的突变。

263. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R285的氨基酸的突变。

264. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R287的氨基酸的突变。

265. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K292的氨基酸的突变。

266. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸E296的氨基酸的突变。

267. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸N297的氨基酸的突变。

268. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸Q646的氨基酸的突变。

269. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N647的氨基酸的突变。

270. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R402的氨基酸的突变。

271. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K393的氨基酸的突变。

272. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N653的氨基酸的突变。

273. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N652的氨基酸的突变。

274. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R482的氨基酸的突变。

275. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N480的氨基酸的突变。

276. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸D396的氨基酸的突变。

277. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸E397的氨基酸的突变。

278. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸D398的氨基酸的突变。

279. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸E399的氨基酸的突变。

280. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K294的氨基酸的突变。

281. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸E400的氨基酸的突变。

282. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸R56的氨基酸的突变。

283. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N157的氨基酸的突变。

284. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸H161的氨基酸的突变。

285. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸H452的氨基酸的突变。

286. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N455的氨基酸的突变。

287. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌

(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K484的氨基酸的突变。

288. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸N486的氨基酸的突变。

289. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸G566的氨基酸的突变。

290. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸H567的氨基酸的突变。

291. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸A656的氨基酸的突变。

292. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸V795的氨基酸的突变。

293. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸A796的氨基酸的突变。

294. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸W842的氨基酸的突变。

295. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸K871的氨基酸的突变。

296. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸E873的氨基酸的突变。

297. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R874的氨基酸的突变。

298. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸R1068的氨基酸的突变。

299. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸N1069的氨基酸的突变。

300. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(Prevotella buccae) Cas13b (PbCas13b) 的氨基酸H1073的氨基酸的突变。

301. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌(Leptotrichia shahii) Cas13a (LshCas13a) 的下述氨基酸的氨基酸的一个或多个突变:R597、N598、H602、R1278、N1279或H1283。

302. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌(Leptotrichia shahii) Cas13a (LshCas13a) 的下述氨基酸的氨基酸的一个或多个突变:R597、N598、H602、R1278、N1279或H1283。

303. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于沙氏纤毛菌(Leptotrichia shahii) Cas13a (LshCas13a) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R597、N598、H602、R1278、N1279或H1283。

304. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌(Leptotrichia shahii) Cas13a (LshCas13a) 的下述氨基酸的氨基酸的一个或多个突变:R597、N598或H602。

305. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R597、N598或H602。

306. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的下述氨基酸的氨基酸的一个或多个突变:R1278、N1279或H1283。

307. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:R1278、N1279或H1283。

308. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的下述氨基酸的氨基酸的一个或多个突变:R146、H151、R1116或H1121。

309. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的下述氨基酸的氨基酸的一个或多个突变:R146、H151、R1116或H1121。

310. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R146、H151、R1116或H1121。

311. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的下述氨基酸的氨基酸的一个或多个突变:R146或H151。

312. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R146或H151。

313. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的下述氨基酸的氨基酸的一个或多个突变:R1116或H1121。

314. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:R1116或H1121。

315. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种(*Prevotella sp.*) P5-125 Cas13b (PspCas13b)的下述氨基酸的氨基酸的一个或多个突变:H133或H1058。

316. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种(*Prevotella sp.*) P5-125 Cas13b (PspCas13b)的下述氨基酸的氨基酸的一个或多个突变:H133或H1058。

317. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于普雷沃菌属种(*Prevotella sp.*) P5-125 Cas13b (PspCas13b)的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:H133或H1058。

318. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种(*Prevotella* sp.) P5-125 Cas13b (PspCas13b)的氨基酸H133的氨基酸突变。

319. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于普雷沃菌属种(*Prevotella* sp.) P5-125 Cas13b (PspCas13b)的HEPN结构域1中的氨基酸H133的氨基酸突变。

320. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种(*Prevotella* sp.) P5-125 Cas13b (PspCas13b)的氨基酸H1058的氨基酸突变。

321. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于普雷沃菌属种(*Prevotella* sp.) P5-125 Cas13b (PspCas13b)的HEPN结构域2中的氨基酸H1058的氨基酸突变。

322. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为A、P或V,优选地A。

323. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为疏水性氨基酸。

324. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为芳香族氨基酸。

325. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为带电荷的氨基酸。

326. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为带正电的氨基酸。

327. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为带负电的氨基酸。

328. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为极性氨基酸。

329. 根据权利要求57至321中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为脂肪族氨基酸。

330. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是或源于以下的种: 别样杆菌属(*Alistipes*)、厌氧唾液杆菌(*Anaerosalibacter*)、拟杆菌属(*Bacteroides*)、拟杆菌门(*Bacteroidetes*)、伯杰菌属(*Bergeyella*)、布劳特氏菌属(*Blautia*)、丁酸弧菌属(*Butyrivibrio*)、二氧化碳噬纤维菌属(*Capnocytophaga*)、肉食杆菌属(*Carnobacterium*)、绿曲挠丝状菌属(*Chloroflexus*)、金黄杆菌属(*Chryseobacterium*)、梭菌属(*Clostridium*)、Demequina、优杆菌科(*Eubacteriaceae*)、优杆菌属(*Eubacterium*)、黄杆菌属(*Flavobacterium*)、梭杆菌属(*Fusobacterium*)、Herbinix、Insolitispirillum、毛螺菌科(*Lachnospiraceae*)、纤毛菌属(*Leptotrichia*)、李斯特氏菌属(*Listeria*)、香味菌属(*Myroides*)、杆状杆菌(*Paludibacter*)、细菌杆菌(*Phaeodactylibacter*)、紫单胞菌科(*Porphyromonadaceae*)、卟啉单胞菌属(*Porphyromonas*)、普雷沃菌属(*Prevotella*)、假丁酸弧菌属(*Pseudobutyrvibrio*)、冷弯菌属(*Psychroflexus*)、赖兴氏杆菌(*Reichenbachiella*)、红杆菌属(*Rhodobacter*)、里氏杆菌属(*Riemerella*)、Sinomicrobium、海旋菌属(*Thalassospira*)、瘤胃球菌属

(*Ruminococcus*); 优选地, 沙氏纤毛菌 (*Leptotrichia shahii*)、斯氏李斯特氏菌 (*Listeria seeligeri*)、毛螺科菌细菌 (*Lachnospiraceae bacterium*) (如Lb MA2020、Lb NK4A179、Lb NK4A144)、*Clostridium aminophilum* (如Ca DSM 10710)、鸡肉杆菌 (*Carnobacterium gallinarum*) (如Cg DSM 4847)、丙酸杆菌 (*Paludibacter propionicigenes*) (如Pp WB4)、魏斯特斯蒂芬氏李斯特氏菌 (*Listeria weihenstephanensis*) (如Lw FSL R9-0317)、李斯特氏菌科细菌 (*Listeriaceae bacterium*) (如Lb FSL M6-0635)、瓦氏细单胞菌 (*Leptotrichia wadei*) (如Lw F0279)、荚膜红杆菌 (*Rhodobacter capsulatus*) (如Rc SB 1003、Rc R121、Rc DE442)、口腔纤毛菌 (*Leptotrichia buccalis*) (如Lb C-1013-b)、*Herbinix hemicellulosilytica*、优杆菌科细菌 (*Eubacteriaceae bacterium*) (如Eb CHKCI004)、布劳特氏菌属种 (*Blautia sp.*) Marseille-P2398、纤毛菌属种 (*Leptotrichia sp.*) 口腔分类单元879str.F0557、聚集绿屈扰菌 (*Chloroflexus aggregans*)、*Demequina aurantiaca*、海旋菌种 (*Thalassospira sp.*) TSL5-1、假丁酸弧菌属种 (*Pseudobutyrvibrio sp.*) OR37、丁酸弧菌属种 (*Butyrvibrio sp.*) YAB3001、纤毛菌属种 (*Leptotrichia sp.*) Marseille-P3007、*Bacteroides ihuae*、紫单孢菌科细菌 (*Porphyromonadaceae bacterium*) (如Pb KH3CP3RA)、*Listeria riparia*、*Insolitispirillum peregrinum*、别样杆菌属种 (*Alistipes sp.*) ZOR0009、酿脓拟杆菌 (*Bacteroides pyogenes*) (如Bp F0041)、拟杆菌属细菌 (*Bacteroidetes bacterium*) (如Bb GWA2\_31\_9)、动物溃疡伯格菌 (*Bergeyella zoohelcum*) (如Bz ATCC 43767)、犬咬二氧化碳噬纤维菌 (*Capnocytophaga canimorsus*)、希诺地米碳酸噬胞菌 (*Capnocytophaga cynodegmi*)、*Chryseobacterium carnipullorum*、济州岛金黄杆菌 (*Chryseobacterium jejuense*)、嗜鳃黄杆菌 (*Chryseobacterium ureilyticum*)、嗜分支黄杆菌 (*Flavobacterium branchiophilum*)、柱状黄杆菌 (*Flavobacterium columnare*)、黄杆菌属种 (*Flavobacterium sp.*) 316、*Myroides odoratimimus* (如Mo CCUG 10230、Mo CCUG 12901、Mo CCUG 3837)、*Paludibacter propionicigenes*、*Phaeodactylibacter xiamenensis*、牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) (如Pg F0185、Pg F0568、Pg JCVI SC001、Pg W4087)、马铃薯红单胞菌 (*Porphyromonas gulae*)、卟啉单胞菌属种 (*Porphyromonas sp.*) COT-0520H4946、树蛙普雷沃菌 (*Prevotella aurantiaca*)、颊普雷沃菌 (*Prevotella buccae*) (如Pb ATCC 33574)、*Prevotella falsenii*、中间普雷沃菌 (*Prevotella intermedia*) (如Pi 17、Pi ZT)、*Prevotella pallens* (如Pp ATCC 700821)、*Prevotella pleuritidis*、解糖普雷沃菌 (*Prevotella saccharolytica*) (如Ps F0055)、普雷沃菌属种 (*Prevotella sp.*) MA2016、普雷沃菌属种 (*Prevotella sp.*) MSX73、普雷沃菌属种 (*Prevotella sp.*) P4-76、普雷沃菌属种 (*Prevotella sp.*) P5-119、普雷沃菌属种 (*Prevotella sp.*) P5-125、普雷沃菌属种 (*Prevotella sp.*) P5-60、扭曲冷弯曲菌 (*Psychroflexus torquis*)、*Reichenbachiella agariperforans*、鸭疫里默氏杆菌 (*Riemerella anatipestifer*)、*Sinomicrobium oceani*、坏死梭杆菌 (*Fusobacterium necrophorum*) (如Fn subsp.funduliforme ATCC 51357、Fn DJ-2、Fn BFTR-1、Fn subsp.Funduliforme)、坏疽梭杆菌 (*Fusobacterium perfoetens*) (如Fp ATCC 29250)、溃疡梭杆菌 (*Fusobacterium ulcerans*) (如Fu ATCC 49185)、厌氧唾液杆菌属种 (*Anaerosalibacter sp.*) ND1、产亚硝酸真杆菌 (*Eubacterium siraeum*)、黄化瘤胃球菌

(*Ruminococcus flavefaciens*) (如Rfx XPD3002)或白色瘤胃球菌(*Ruminococcus albus*)。

331. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是Cas13a蛋白。

332. 根据权利要求331所述的工程化CRISPR-Cas蛋白,其中所述Cas13a蛋白是或源于以下的种:拟杆菌属(*Bacteroides*)、布劳特氏菌属(*Blautia*)、丁酸弧菌属(*Butyrivibrio*)、肉食杆菌属(*Carnobacterium*)、绿曲挠丝状菌属(*Chloroflexus*)、梭菌属(*Clostridium*)、*Demequina*、优杆菌属(*Eubacterium*)、*Herbinix*、*Insolitinspirillum*、毛螺菌科(*Lachnospiraceae*)、纤毛菌属(*Leptotrichia*)、李斯特氏菌属(*Listeria*)、杆状杆菌(*Paludibacter*)、紫单胞菌科(*Porphyromonadaceae*)、假丁酸弧菌属(*Pseudobutyrvibrio*)、红杆菌属(*Rhodobacter*)或海旋菌属(*Thalassospira*);优选地,沙氏纤毛菌(*Leptotrichia shahii*)、斯氏李斯特氏菌(*Listeria seeligeri*)、毛螺科菌细菌(*Lachnospiraceae bacterium*) (如Lb MA2020、Lb NK4A179、Lb NK4A144)、*Clostridium aminophilum* (如Ca DSM 10710)、鸡肉杆菌(*Carnobacterium gallinarum*) (如Cg DSM 4847)、丙酸杆菌(*Paludibacter propionicigenes*) (如Pp WB4)、魏斯特斯蒂芬氏李斯特氏菌(*Listeria weihenstephanensis*) (如Lw FSL R9-0317)、李斯特氏菌科细菌(*Listeriaceae bacterium*) (如Lb FSL M6-0635)、瓦氏细单胞菌(*Leptotrichia wadei*) (如Lw F0279)、荚膜红杆菌(*Rhodobacter capsulatus*) (如Rc SB 1003、Rc R121、Rc DE442)、口腔纤毛菌(*Leptotrichia buccalis*) (如Lb C-1013-b)、*Herbinix hemicellulosilytica*、优杆菌科细菌(*Eubacteriaceae bacterium*) (如Eb CHKCI004)、布劳特氏菌属种(*Blautia sp.*)Marseille-P2398、纤毛菌属种(*Leptotrichia sp.*)口腔分类单元879str.F0557、聚集绿屈扰菌(*Chloroflexus aggregans*)、*Demequina aurantiaca*、海旋菌种(*Thalassospira sp.*)TSL5-1、假丁酸弧菌属种(*Pseudobutyrvibrio sp.*)OR37、丁酸弧菌属种(*Butyrivibrio sp.*)YAB3001、纤毛菌属种(*Leptotrichia sp.*)Marseille-P3007、*Bacteroides ihuae*、紫单胞菌科细菌(*Porphyromonadaceae bacterium*) (如Pb KH3CP3RA)、*Listeria riparia*或*Insolitinspirillum peregrinum*。

333. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是Cas13b蛋白。

334. 根据权利要求333所述的工程化CRISPR-Cas蛋白,其中所述Cas13b蛋白是或源于以下的种:别样杆菌属(*Alistipes*)、拟杆菌属(*Bacteroides*)、拟杆菌门(*Bacteroidetes*)、伯杰菌属(*Bergeyella*)、二氧化碳噬纤维菌属(*Capnocytophaga*)、金黄杆菌属(*Chryseobacterium*)、黄杆菌属(*Flavobacterium*)、香味菌属(*Myroides*)、杆状杆菌(*Paludibacter*)、细菌杆菌(*Phaeodactylibacter*)、卟啉单胞菌属(*Porphyromonas*)、普雷沃菌属(*Prevotella*)、冷弯菌属(*Psychroflexus*)、赖兴氏杆菌(*Reichenbachiella*)、里氏杆菌属(*Riemerella*)或*Sinomicrobium*;优选地,别样杆菌属种(*Alistipes sp.*)ZOR0009、酿脓拟杆菌(*Bacteroides pyogenes*) (如Bp F0041)、拟杆菌属细菌(*Bacteroidetes bacterium*) (如Bb GWA2\_31\_9)、动物溃疡伯格菌(*Bergeyella zoohelcum*) (如Bz ATCC 43767)、犬咬二氧化碳噬纤维菌(*Capnocytophaga canimorsus*)、希诺地米碳酸噬胞菌(*Capnocytophaga cynodegmi*)、*Chryseobacterium carnipullorum*、济州岛金黄杆菌(*Chryseobacterium jejuense*)、嗜鳃黄杆菌(*Chryseobacterium ureilyticum*)、嗜分支黄

杆菌 (*Flavobacterium branchiophilum*)、柱状黄杆菌 (*Flavobacterium columnare*)、黄杆菌属种 (*Flavobacterium sp.*) 316、*Myroides odoratimimus* (如 Mo CCUG 10230、Mo CCUG 12901、Mo CCUG 3837)、丙酸杆菌 (*Paludibacter propionigenes*)、*Phaeodactylibacter xiamenensis*、牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) (如 Pg F0185、Pg F0568、Pg JCVI SC001、Pg W4087)、马铃薯红单胞菌 (*Porphyromonas gulae*)、卟啉单胞菌属种 (*Porphyromonas sp.*) C0T-0520H4946、树蛙普雷沃菌 (*Prevotella aurantiaca*)、颊普雷沃菌 (*Prevotella buccae*) (如 Pb ATCC 33574)、*Prevotella falsenii*、中间普雷沃菌 (*Prevotella intermedia*) (如 Pi 17、Pi ZT)、*Prevotella pallens* (如 Pp ATCC 700821)、*Prevotella pleuritidis*、解糖普雷沃菌 (*Prevotella saccharolytica*) (如 Ps F0055)、普雷沃菌属种 (*Prevotella sp.*) MA2016、普雷沃菌属种 (*Prevotella sp.*) MSX73、普雷沃菌属种 (*Prevotella sp.*) P4-76、普雷沃菌属种 (*Prevotella sp.*) P5-119、普雷沃菌属种 (*Prevotella sp.*) P5-125、普雷沃菌属种 (*Prevotella sp.*) P5-60、扭曲冷弯曲菌 (*Psychroflexus torquis*)、*Reichenbachiella agariperforans*、鸭疫里默氏杆菌 (*Riemerella anatipestifer*) 或 *Sinomicrobium oceani*。

335. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其中所述Cas13c蛋白是Cas13c蛋白。

336. 根据权利要求335所述的工程化CRISPR-Cas蛋白,其中所述Cas13c蛋白是或源于以下的种:梭杆菌属 (*Fusobacterium*) 或厌氧唾液杆菌 (*Anaerosalibacter*); 优选地,坏死梭杆菌 (*Fusobacterium necrophorum*) (如 Fn subsp. *funduliforme* ATCC 51357、Fn DJ-2、Fn BFTR-1、Fn subsp. *Funduliforme*)、坏疽梭杆菌 (*Fusobacterium perfoetens*) (如 Fp ATCC 29250)、溃疡梭杆菌 (*Fusobacterium ulcerans*) (如 Fu ATCC 49185) 或厌氧唾液杆菌属种 (*Anaerosalibacter sp.*) ND1。

337. 根据权利要求55所述的工程化CRISPR-Cas蛋白,其中所述Cas13c蛋白是Cas13d蛋白。

338. 根据权利要求337所述的工程化CRISPR-Cas蛋白,其中所述Cas13d蛋白是或源于以下的种:优杆菌属 (*Eubacterium*) 或瘤胃球菌属 (*Ruminococcus*), 优选地,产亚硝酸真杆菌 (*Eubacterium siraeum*)、黄化瘤胃球菌 (*Ruminococcus flavefaciens*) (如 Rfx XPD3002) 或白色瘤胃球菌 (*Ruminococcus albus*)。

339. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的催化活性增加。

340. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的催化活性降低。

341. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的gRNA结合增加。

342. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的gRNA结合降低。

343. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的特异性增加。

344. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas

蛋白相比,所述工程化CRISPR-Cas蛋白的特异性降低。

345. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的稳定性增加。

346. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的稳定性降低。

347. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其还包含使催化活性失活的一个或多个突变。

348. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的脱靶结合增加。

349. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的脱靶结合降低。

350. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的靶标结合增加。

351. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的靶标结合降低。

352. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白具有更高蛋白酶活性或多核苷酸结合能力。

353. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,PFS识别被改变。

354. 根据权利要求1所述的工程化CRISPR-Cas蛋白,其还包含功能性异源结构域。

355. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其还包含NLS。

356. 根据权利要求50所述的工程化CRISPR-Cas蛋白,其还包含NES。

357. 一种工程化CRISPR-Cas蛋白,其包含一个或多个HEPN结构域,且长度小于1000个氨基酸。

358. 根据权利要求357所述的工程化CRISPR-Cas蛋白,其中所述蛋白的尺寸小于950个、小于900个、小于850个、小于800个、或小于750个氨基酸。

359. 根据权利要求357所述的工程化CRISPR-Cas蛋白,其中所述HEPN结构域包含RxxxxH基序。

360. 根据权利要求359所述的工程化CRISPR-Cas蛋白,其中所述RxxxxH基序包含R[N/H/K]X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H序列。

361. 根据权利要求360所述的工程化CRISPR-Cas蛋白,其中:

X<sub>1</sub>是R、S、D、E、Q、N、G或Y,

X<sub>2</sub>独立地是I、S、T、V或L,和

X<sub>3</sub>独立地是L、F、N、Y、V、I、S、D、E或A。

362. 根据权利要求357所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白是VI型CRISPR Cas蛋白。

363. 根据权利要求362所述的工程化CRISPR-Cas蛋白,其中所述VI型CRISPR Cas蛋白是Cas13a、Cas13b、Cas13c或Cas13d。

364. 根据权利要求357所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白与功

能性结构域相关联。

365. 根据权利要求357所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白包含与权利要求57-329中任一项所述的突变等价的一个或多个突变。

366. 根据权利要求365所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白包含在螺旋结构域中的一个或多个突变。

367. 根据权利要求357所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白是死形式或具有切口酶活性。

368. 一种多核苷酸,其编码权利要求1至367中任一项所述的工程化CRISPR-Cas蛋白。

369. 根据权利要求319所述的多核苷酸,其是密码子优化的。

370. 一种CRISPR-Cas系统,其包含权利要求1至367中任一项所述的工程化CRISPR-Cas蛋白或者权利要求318或319所述的多核苷酸,以及能够与所述工程化CRISPR-Cas蛋白形成复合物并能够与靶核酸序列杂交并且引导所述复合物与所述靶核酸序列的序列特异性结合的核苷酸组分。

371. 一种载体系统,其包含一个或多个载体,所述一个或多个载体包含编码权利要求370所述的工程化CRISPR-Cas蛋白的组分的一个或多个多核苷酸分子。

372. 一种修饰靶核酸的方法,其包括:在包含所述靶核酸的细胞或生物体中引入根据权利要求1至367中任一项所述的工程化CRISPR-Cas蛋白,根据权利要求368或369所述的多核苷酸,根据权利要求370所述的CRISPR-Cas系统,或者根据权利要求371所述的载体或载体系统,使得所述工程化CRISPR-Cas蛋白修饰在所述细胞或生物体中的所述靶核酸。

373. 根据权利要求372所述的方法,其中所述工程化CRISPR-Cas系统是通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪、可植入装置或权利要求371所述的载体系统递送引入。

374. 根据权利要求372所述的方法,其中所述工程化CRISPR-cas蛋白与一个或多个功能性结构域相关联。

375. 根据权利要求372所述的方法,其中所述靶核酸包含基因组基因座,并且所述工程化CRISPR-Cas蛋白改变在所述基因组基因座处编码的基因产物或所述基因产物的表达。

376. 根据权利要求372所述的方法,其中所述靶核酸是DNA或RNA,且其中所述靶核酸中的一个或多个核苷酸被碱基编辑。

377. 根据权利要求372所述的方法,其中所述靶核酸是DNA或RNA,且其中所述靶核酸被切割。

378. 根据权利要求377所述的方法,其中所述工程化CRISPR-Cas蛋白还切割非靶核酸。

379. 根据权利要求377所述的方法,其还包括使活性可视化,以及任选地,使用可检测标记。

380. 根据权利要求377所述的方法,其还包括检测所述CRISPR-Cas系统的一个或多个组分与所述靶核酸的结合。

381. 根据权利要求377所述的方法,其中所述细胞或生物体是原核细胞或生物体。

382. 根据权利要求377所述的方法,其中所述细胞或生物体是动物细胞或生物体。

383. 根据权利要求377所述的方法,其中所述细胞或生物体是植物细胞或生物体。

384. 一种用于检测样品中的靶核酸的方法,其包括:

将样品与下述接触：

权利要求50至367中任一项所述的工程化CRISPR-Cas蛋白；

至少一种指导多核苷酸，其包含能够结合至所述靶核酸并被设计为与所述工程化CRISPR-Cas形成复合物的指导序列；和

包含非靶序列的基于RNA的掩蔽构建体；

其中所述工程化CRISPR-Cas蛋白表现出附带RNase活性并切割所述检测构建体的所述非靶序列；和

检测来自所述非靶序列的切割的信号，从而检测在所述样品中的所述靶核酸。

385. 根据权利要求384所述的方法，其还包括将所述样品与用于扩增所述靶核酸的试剂接触。

386. 根据权利要求385所述的方法，其中用于扩增的所述试剂包括等温扩增反应试剂。

387. 根据权利要求386所述的方法，其中所述等温扩增试剂包括基于核酸序列的扩增、重组酶聚合酶扩增、环介导的等温扩增、链置换扩增、解旋酶依赖性扩增或切口酶扩增试剂。

388. 根据权利要求384所述的方法，其中所述靶核酸是DNA分子，并且所述方法还包括将所述靶DNA分子与包含RNA聚合酶位点的引物和RNA聚合酶接触。

389. 根据权利要求384所述的方法，其中所述掩蔽构建体：

抑制可检测阳性信号的产生直至所述掩蔽构建体被切割或失活，或者

掩蔽可检测阳性信号或产生可检测阴性信号直至所述掩蔽构建体被切割或失活。

390. 根据权利要求384所述的方法，其中所述掩蔽构建体包含：

a. 沉默RNA，其抑制由报告构建体编码的基因产物的产生，其中当表达时所述基因产物产生所述可检测阳性信号；

b. 核酶，其产生所述可检测阴性信号，并且其中当所述核酶失活时产生所述阳性可检测信号；

c. 核酶，其将底物转化为第一颜色，并且其中当所述核酶失活时所述底物转化为第二颜色；

d. 适体和/或包含多核苷酸栓系的抑制剂；

e. 多核苷酸，其附着有可检测配体和掩蔽组分；

f. 通过桥分子保持在聚集体中的纳米颗粒，其中至少一部分的所述桥分子包含多核苷酸，并且其中当所述纳米颗粒被释放在溶液中时所述溶液经历色移；

g. 通过连接分子连接至一个或多个淬灭剂分子的量子点或荧光团，其中至少一部分的所述连接分子包含多核苷酸；

h. 与嵌入剂复合的多核苷酸，其中在切割所述多核苷酸后所述嵌入剂改变吸光度；或者

1. 通过多核苷酸栓系的两个荧光团，其当从所述多核苷酸释放时经历荧光移位。

391. 根据权利要求390所述的方法，其中所述适体

a. 包含螯合酶的多核苷酸栓系的抑制剂，其中在通过作用于底物而从所述适体或多核苷酸栓系的抑制剂释放后所述酶产生可检测信号；或者

b. 是抑制性适体，其抑制酶并阻止酶催化底物产生可检测信号，或其中所述多核苷酸

栓系的抑制剂抑制酶并阻止酶催化底物产生可检测信号;或者

c. 整合一对试剂,所述试剂当从所述适体释放时组合以产生可检测信号。

392. 根据权利要求390所述的方法,其中所述纳米颗粒是胶体金属。

393. 根据权利要求384所述的方法,其中所述至少一个指导多核苷酸包含错配。

394. 根据权利要求384所述的方法,其中所述错配是在所述一个或多个指导序列上的单核苷酸改变的上游或下游。

395. 一种细胞或生物体,其包含根据权利要求1至367中任一项所述的工程化CRISPR-Cas蛋白,根据权利要求368或369所述的多核酸,根据权利要求370所述的CRISPR-Cas系统或者根据权利要求371所述的载体或载体系统。

## 新型CRISPR酶和系统

[0001] 相关申请的交叉引用

[0002] 本申请要求2018年07月31日提交的美国临时申请号62/712,809, 2018年10月26日提交的美国临时申请号62/751,421, 2018年12月05日提交的美国临时申请号62/775,865, 2019年03月22日提交的美国临时申请号62/822,639和2019年07月11日提交的美国临时申请号62/873,031的权益。以上确认的申请的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)授予的拨款号HG009761、MH110049和HL141201在政府支持下完成的。政府享有本发明的某些权利。

[0005] 电子序列表引用

[0006] 电子序列表(“BROD-2660WP\_ST25.txt”;大小为1,997,857字节,创建日期为2019年07月25日)的内容以引用方式整体并入本文。

### 技术领域

[0007] 本发明总体上涉及用于控制涉及序列靶向(如基因转录物的扰动或核酸编辑)的基因表达的系统、方法和组合物,其可以使用涉及规律间隔成簇短回文重复序列(CRISPR)及其组分的载体系统。

### 背景技术

[0008] 细菌和古细菌适应性免疫的CRISPR-CRISPR相关(Cas)系统是一些这样的系统,其显示出蛋白质组成和基因组基因座结构的极端多样性。CRISPR-Cas系统基因座具有超过50种的基因家族并且不存在严格的通用基因,这表明了基因座结构的快速进化和极端多样性。到目前为止,采用了多分支方法,针对93种Cas蛋白存在约395个谱分析的全面cas基因鉴定。分类包括特征基因谱分析加上基因座结构的特征。提出了一种新的CRISPR-Cas系统分类,其中这些系统宽泛地分成两类,具有多亚基效应子复合物的第1类和通过Cas9蛋白来示例的具有单亚基效应子模块的第2类。与第2类CRISPR-Cas系统相关联的新型效应蛋白可以被开发为强有力的基因组工程化工具,并且推定的新型效应蛋白的预测及其工程化和优化是重要的。新型Cas13b直系同源物及其用途是期望的。

[0009] 在证明CRISPR-Cas9可以改换用途用于基因组编辑之后,对利用CRISPR系统的兴趣导致发现了数种具有新性质的新型Cas酶和CRISPR系统(1-3)。在这些新发现中,值得注意的是第2类VI型CRISPR-Cas13系统,其使用可编程CRISPR-RNA(crRNA)指导序列来使用单一酶以靶向RNA(1-6)。与靶单链RNA结合的Cas13激活了一般RNase活性,其切割靶点并且非特异性地降解周围的RNA(4)。已将VI型系统用于RNA敲低、转录物标记、RNA编辑和超灵敏病毒检测(3、4、7-12)。根据Cas13蛋白的特性将CRISPR-Cas13系统进一步分成四种亚型(Cas13a-d)(2)。所有Cas13蛋白家族成员均含有两个高级原核生物和真核生物核苷酸结合(HEPN)结构域。本申请中的任何文献的引用或确定并不是承认该文献可作为本发明的现有

技术。

[0010] 迫切需要替代性的和稳健的系统和技术用于靶向核酸或多核苷酸（例如，DNA或RNA或者任何杂交物或其衍生物），具有广泛应用，特别是开发具有改变的功能性的效应蛋白，例如包括但不限于增加或降低的特异性、增加或降低的活性、改变的特异性和/或活性、替代的PAM 识别等。本发明满足了这种需求并提供了相关的优点。将本申请的新型 RNA靶向系统添加到基因组、转录组和表观基因组靶向技术的库中可以通过直接检测、分析和操纵来改变对特定靶位点的研究和干扰或编辑。为了有效地利用本申请的RNA靶向系统来进行RNA靶向而没有有害影响，至关重要的是要了解这些RNA靶向工具的工程化和优化的方面。

## 发明内容

[0011] 在一个方面中，本公开内容提供了一种工程化的CRISPR-Cas蛋白，其包含一个或多个HEPN结构域，以及还包含一个或多个修饰的氨基酸，其中所述氨基酸：与指导RNA相互作用，所述指导RNA与所述工程化 CRISPR-Cas蛋白形成复合物；在所述工程化CRISPR-Cas蛋白的HEPN 活性位点、结构域间接头结构域、盖结构域、螺旋结构域1、螺旋结构域2或桥螺旋结构域中；或者其组合。

[0012] 在一些实施方式中，所述HEPN结构域包含RxxxxH基序。在一些实施方式中，所述RxxxxH基序包含R {N/H/K} X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H (SEQ ID NO:78) 序列。在一些实施方式中，在R {N/H/K} X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H序列中，X<sub>1</sub>是R、S、D、E、Q、N、G或Y，X<sub>2</sub>独立地是I、S、T、V或L，以及X<sub>3</sub>独立地是L、F、N、Y、V、I、S、D、E或A。

[0013] 在一些实施方式中，CRISPR-Cas蛋白是VI型CRISPR Cas蛋白在一些实施方式中，VI型CRISPR Cas蛋白是Cas13。在一些实施方式中，VI型CRISPR Cas蛋白是Cas13a、Cas13b、Cas13c或Cas13d。

[0014] 在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、A656、V795、A796、W842、K871、E873、R874、R1068、N1069或H1073。

[0015] 在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、W842、K871、E873、R874、R1068、N1069、H1073。

[0016] 在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b)

的下述氨基酸的氨基酸的一个或多个突变是：T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294或E400。

[0017] 在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：K393、R402、N482、T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或H1073。

[0018] 在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：K393、R402、N482、H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或H1073。

[0019] 在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：W842、K846、K870、E873或R877。在一些实施方式中，在螺旋结构域1中对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变是：W842、K846、K870、E873或R877。在一些实施方式中，在螺旋结构域1-3中对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是：W842、K846、K870、E873或R877。在一些实施方式中，在桥螺旋结构域中对应于PbCas13b的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是：W842、K846、K870、E873或R877。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：K393、R402、N480、N482、N652或N653。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：K393、R402、N480或N482。在一些实施方式中，在盖结构域中对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是：K393、R402、N480或N482。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：N652或N653。在一些实施方式中，在螺旋结构域2中对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是：N652或N653。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、S757、N756或K741。在一些实施方式中，在螺旋结构域中对应于PbCas13b的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是：S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、S757、N756或K741。在一些实施方式中，对应于

PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或N756。

[0020] 在一些实施方式中，在螺旋结构域1中对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变是：H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或N756。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：H567、H500、R762、V795、A796、R791、G566、S757或N756。在一些实施方式中，在螺旋结构域1中对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变是：H567、H500、R762、V795、A796、R791、G566、S757或N756。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：K871、K857、K870、W842、E873、R877、K846或R874。在一些实施方式中，在桥螺旋结构域中对应于PbCas13b的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是：K871、K857、K870、W842、E873、R877、K846或R874。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：H567、H500或G566。

[0021] 在一些实施方式中，在螺旋结构域1-2中对应于PbCas13b的螺旋结构域1-2中的下述氨基酸的氨基酸的一个或多个突变是：H567、H500或G566。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。在一些实施方式中，在螺旋结构域1-3中对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是：K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：R762、V795、A796、R791、S757或N756。在一些实施方式中，在螺旋结构域1-3中对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是：R762、V795、A796、R791、S757或N756。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：S658、N653、A656、K655、N652、K590、R638或K741。在一些实施方式中，在螺旋结构域2中对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是：S658、N653、A656、K655、N652、K590、R638或K741。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：T405、H407、N486、K484、N480、H452、N455或K457。在一些实施方式中，在盖结构域中对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是：T405、H407、N486、K484、N480、H452、N455或K457。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或K741。在一些实施方式中，在螺旋结构域中对应于PbCas13b的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是：S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或K741。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是：H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。在一些实施方式中，在螺旋结构域1中对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变是：H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。在一些实施方式中，对应于PbCas13b的下述氨基酸的氨基酸的一个或多

个突变是:H567、H500、R762、R791、G566、S757或N756。在一些实施方式中,在螺旋结构域1中对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变是:H567、H500、R762、R791、G566、S757或N756。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。在一些实施方式中,在螺旋结构域1-3中对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。

[0022] 在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:R762、R791、S757或N756。在一些实施方式中,在螺旋结构域1-3中对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是:R762、R791、S757或N756。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:S658、N653、K655、N652、K590、R638或K741。在一些实施方式中,在螺旋结构域2中对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是:S658、N653、K655、N652、K590、R638或K741。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:H407、N486、K484、N480、H452、N455或K457。

[0023] 在一些实施方式中,在盖结构域中对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:H407、N486、K484、N480、H452、N455或K457。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:R56、N157、H161、R1068、N1069或H1073。在一些实施方式中,在HEPN结构域中对应于PbCas13b的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R56、N157、H161、R1068、N1069或H1073。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:R56、N157或H161。在一些实施方式中,在HEPN结构域1中对应于PbCas13b的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变是:R56、N157或H161。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:R1068、N1069或H1073。在一些实施方式中,在HEPN结构域2中对应于PbCas13b的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变是:R1068、N1069或H1073。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。在一些实施方式中,在盖结构域中对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。在一些实施方式中,在盖结构域中对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。在一

些实施方式中,对应于 PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:N486、K484、N480、H452、N455或K457。

[0024] 在一些实施方式中,在盖结构域中对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:N486、K484、N480、H452、N455或K457。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:K393、R402、N482、N486、K484、N480、H452、N455或K457。在一些实施方式中,在盖结构域中对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:K393、R402、N482、N486、K484、N480、H452、N455或K457。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。在一些实施方式中,对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变是:S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

[0025] 在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K943或R1041。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53 或Y164。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K943 或R1041。在一些实施方式中,在HEPN结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K943或R1041。在一些实施方式中,在HEPN结构域1中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变是:R53或Y164。在一些实施方式中,在HEPN结构域2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变是:K943或R1041。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K943、R1041、R56、N157、H161、R1068、N1069或H1073。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、R56、N157 或H161。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K943、R1041、R1068、N1069或H1073。在一些实施方式中,在HEPN结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN 结构域中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K943、R1041、R56、N157、

H161、R1068、N1069或H1073。在一些实施方式中,在HEPN结构域1中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、R56、N157或H161。在一些实施方式中,在HEPN 结构域2中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变是:K943、R1041、R1068、N1069或H1073。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183、K193、K943或R1041。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183或K193。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K943或 R1041。

[0026] 在一些实施方式中,在HEPN结构域中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183、K193、K943或R1041。在一些实施方式中,在HEPN结构域1中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183或K193。在一些实施方式中,在HEPN 结构域2中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变是:K943 或R1041。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069 或H1073。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183、K193、R56、N157或H161。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K943、R1041、R1068、N1069或H1073。在一些实施方式中,在HEPN结构域中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069或H1073。在一些实施方式中,在HEPN 结构域1中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K183、K193、R56、N157或H161。

[0027] 在一些实施方式中,在HEPN结构域2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变是:K943、R1041、R1068、N1069或 H1073。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K183或K193。在一些实施方式中,在HEPN结构域1中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变是:K183或K193。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K943或R1041。在一些实施方式中,在HEPN结构域中对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R53、Y164、K943或R1041。在一些实施方式中,对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个

突变是：R53、K943或R1041；优选地，R53A、R53K、R53D 或R53E；K943A、K943R、K943D或K943E；或者，R1041A、R1041K、R1041D或R1041E。在一些实施方式中，在HEPN结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是：R53、K943或R1041；优选地，R53A、R53K、R53D或R53E；K943A、K943R、K943D或K943E；或者，R1041A、R1041K、R1041D或R1041E。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸Y164 的氨基酸的突变是，优选地Y164A、Y164F或Y164W。

[0028] 在一些实施方式中，在HEPN结构域1中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1的氨基酸Y164 的氨基酸的突变是，优选地Y164A、Y164F或Y164W。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398或E399。在一些实施方式中，在盖结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是：T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398或E399。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸H407的氨基酸突变是，优选地 H407Y、H407W或H407F。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：R402、K393、R482、N480、D396、E397、D398或E399。在一些实施方式中，在盖结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是：R402、K393、R482、N480、D396、E397、D398或E399。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：K457、D434 或K431。在一些实施方式中，在盖结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是：K457、D434或K431。

[0029] 在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。在一些实施方式中，在螺旋结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是：H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些实施方式中，在螺旋结构域1中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变是：H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b

(PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：H500、K570、N756、S757、R762或R791。在一些实施方式中，在螺旋结构域1中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变是：H500、K570、N756、S757、R762或R791。

[0030] 在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836 或R838。在一些实施方式中，在桥螺旋结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是：K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：H500或K570。在一些实施方式中，在螺旋结构域1-2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-2中的下述氨基酸的氨基酸的一个或多个突变是：H500或K570。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些实施方式中，在螺旋结构域1-3中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是：N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：N756、S757、R762或R791。在一些实施方式中，在螺旋结构域1-3中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是：N756、S757、R762、R791、K846、K857、K870或R877。在一些实施方式中，在螺旋结构域1-3中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3 中的下述氨基酸的氨基酸的一个或多个突变是：N756、S757、R762、R791、K846、K857、K870或R877。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：K826、K828、K829、R824、R830、Q831、K835、K836 或R838。在一些实施方式中，在螺旋结构域1-3中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变是：K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653 或N652。在一些实施方式中，在螺旋结构域2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是：K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653或N652。

[0031] 在一些实施方式中，对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是：Q646或N647。在一些实施方式中，在螺旋结构

域2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是:Q646或N647。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:N653或N652。在一些实施方式中,在螺旋结构域2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是:N653或 N652。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K590、N634、R638、N652、N653、K655、S658、K741或K744。在一些实施方式中,在螺旋结构域2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是:K590、N634、R638、N652、N653、K655、S658、K741或K744。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R600、K607、K612、R614、K617或R618。在一些实施方式中,在螺旋结构域2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域 2中的下述氨基酸的氨基酸的一个或多个突变是:R600、K607、K612、R614、K617或R618。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R285、R287、K292、E296、N297或K294。在一些实施方式中,在盖结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:R285、R287、K292、E296、N297或K294。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R285、K292、E296或N297。在一些实施方式中,在盖结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:R285、K292、E296或N297。

[0032] 在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:T405、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R285、R287、K292、E296、N297、Q646、N647或K294。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R402、K393、N653、N652、R482、N480、D396、E397、D398或E399。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53、K655、R762或R1041;优选地,R53A或R53D;K655A;R762A;或者,R1041E或R1041D。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:N297、E296、K292或R285;优选地,N297A、E296A、K292A或R285A。在一些实施方式中,在IDL结构域(的中央通道)中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的IDL结构域(的中央通道)中的下述氨基酸的氨基酸的一个或多个突变是:N297、E296、K292或R285;优选地,N297A、E296A、K292A或R285A。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:Q831、K836、R838、N652、N653、R830、K655或R762;优选地,Q831A、K836A、R838A、N652A、N653A、R830A、K655A或R762A。在一些实施方式

中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:N652、N653、R830、K655 or R762;优选地,N652A、N653A、R830A、K655A或R762A。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K655或R762;优选地,K655A 或R762A。在一些实施方式中,在螺旋结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是:Q831、K836、R838、N652、N653、R830、K655或R762;优选地,Q831A、K836A、R838A、N652A、N653A、R830A、K655A或R762A。

[0033] 在一些实施方式中,在螺旋结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变是:N652、N653、R830、K655 or R762;优选地,N652A、N653A、R830A、K655A或R762A。在一些实施方式中,在螺旋结构域 2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是:K655或R762;优选地,K655A或R762A。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R614、K607、K193、K183或R600;优选地,R614A、K607A、K193A、K183A或R600A。在一些实施方式中,在螺旋结构域 2中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变是:Q646或N647;优选地,Q646A或N647A。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R53或R1041;优选地,R53A或R53D或R1041E或 R1041D。在一些实施方式中,在HEPN结构域中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R53或R1041;优选地,R53A或R53D 或R1041E或R1041D。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K457、D397、E398、D399、E400、T405、H407或D434;优选地,D397A、E398A、D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。在一些实施方式中,在盖结构域中对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变是:K457、D397、E398、D399、E400、T405、H407或D434;优选地,D397A、E398A、D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。在一些实施方式中,所述氨基酸对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸:氨基酸46-57、73-79、152-164、1036-1046和1064-1074。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R156、N157、H161、R1068、N1069和H1073。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R285、R287、K292、K294、E296和N297。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:K826、K828、K829、R824、R830、Q831、K835、K836和R838。在一些实施方式中,对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870和R877。

[0034] 在一些实施方式中,氨基酸突变对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b





式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸N157。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸H161。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸H452。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸N455。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸K484。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸N486。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸G566。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸H567。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸A656。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸V795。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸A796。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸W842。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸K871。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸E873。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸R874。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸R1068。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸N1069。在一些实施方式中,氨基酸突变对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的氨基酸 H1073。

[0036] 在一些实施方式中,对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的下述氨基酸的氨基酸的一个或多个突变是:R597、N598、H602、R1278、N1279或H1283。在一些实施方式中,对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的下述氨基酸的氨基酸的一个或多个突变是:R597、N598、H602、R1278、N1279或H1283。在一些实施方式中,在HEPN结构域中对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R597、N598、H602、R1278、N1279或H1283。在一些实施方式中,对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的下述氨基酸的氨基酸的一个或多个突变是:R597、N598或H602。在一些实施方式中,在HEPN结构域1中对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变是:R597、N598或H602。在一些实施方式中,对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的下述氨基酸的氨基酸的一个或多个突变是:R1278、N1279或H1283。在一些实施方式中,在HEPN结构域2中对应于沙氏纤毛菌(*Leptotrichia shahii*) Cas13a (LshCas13a)的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变是:R1278、N1279或H1283。在一些实施方式中,对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的下述氨基酸的氨基酸的一个或多个突变是:R146、H151、R1116或H1121。在一些实施方式中,对应于喉管卟啉单胞菌(*Porphyromonas gulae*) Cas13b (PguCas13b)的下

述氨基酸的氨基酸的一个或多个突变是:R146、H151、R1116或H1121。在一些实施方式中,在HEPN结构域中对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:R146、H151、R1116或 H1121。在一些实施方式中,对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的下述氨基酸的氨基酸的一个或多个突变是: R146或H151。在一些实施方式中,在HEPN结构域1中对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的HEPN结构域 1 中的下述氨基酸的氨基酸的一个或多个突变是:R146或H151。在一些实施方式中,对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:R1116或 H1121。在一些实施方式中,在HEPN结构域2中对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变是:R1116或H1121。在一些实施方式中,对应于普雷沃菌属种 (*Prevotella* sp.) P5-125 Cas13b (PspCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:H133或H1058。在一些实施方式中,对应于普雷沃菌属种 (*Prevotella* sp.) P5-125 Cas13b (PspCas13b) 的下述氨基酸的氨基酸的一个或多个突变是:H133 或H1058。在一些实施方式中,在HEPN结构域中对应于普雷沃菌属种 (*Prevotella* sp.) P5-125 Cas13b (PspCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变是:H133或H1058。在一些实施方式中,氨基酸突变对应于普雷沃菌属种 (*Prevotella* sp.) P5-125 Cas13b (PspCas13b) 的氨基酸H133。在一些实施方式中,在HEPN结构域1 中的氨基酸突变对应于普雷沃菌属种 (*Prevotella* sp.) P5-125 Cas13b (PspCas13b) 的HEPN结构域1中的氨基酸H133。在一些实施方式中,氨基酸突变对应于普雷沃菌属种 (*Prevotella* sp.) P5-125 Cas13b (PspCas13b) 的氨基酸H1058。在一些实施方式中,在HEPN结构域2 中的氨基酸突变对应于普雷沃菌属种 (*Prevotella* sp.) P5-125 Cas13b (PspCas13b) 的HEPN结构域2中的氨基酸H1058。

[0037] 在一些实施方式中,所述氨基酸被突变为A、P或V,优选地,A。在一些实施方式中,所述氨基酸突变为疏水性氨基酸。在一些实施方式中,所述氨基酸被突变为芳香族氨基酸。在一些实施方式中,所述氨基酸被突变为带电荷的氨基酸。在一些实施方式中,所述氨基酸被突变为带正电的氨基酸。在一些实施方式中,所述氨基酸被突变为带负电的氨基酸。在一些实施方式中,所述氨基酸被突变为极性氨基酸。在一些实施方式中,所述氨基酸被突变为脂肪族氨基酸。在一些实施方式中,所述工程化CRISPR-Cas蛋白还包含功能性异源结构域。

[0038] 在一些实施方式中,所述Cas13蛋白是或源于以下的种:别样杆菌属 (*Alistipes*)、厌氧唾液杆菌 (*Anaerosalibacter*)、拟杆菌属 (*Bacteroides*)、拟杆菌门 (*Bacteroidetes*)、伯杰菌属 (*Bergeyella*)、布劳特氏菌属 (*Blautia*)、丁酸弧菌属 (*Butyrivibrio*)、二氧化碳噬纤维菌属 (*Capnocytophaga*)、肉食杆菌属 (*Carnobacterium*)、绿曲挠丝状菌属 (*Chloroflexus*)、金黄杆菌属 (*Chryseobacterium*)、梭菌属 (*Clostridium*)、Demequina、优杆菌科 (*Eubacteriaceae*)、优杆菌属 (*Eubacterium*)、黄杆菌属 (*Flavobacterium*)、梭杆菌属 (*Fusobacterium*)、Herbinix、Insolitispirillum、毛螺菌科 (*Lachnospiraceae*)、纤毛菌属 (*Leptotrichia*)、李斯特氏菌属 (*Listeria*)、香味菌属 (*Myroides*)、杆状杆菌 (*Paludibacter*)、细菌杆菌 (*Phaeodactylibacter*)、紫单胞菌科 (*Porphyromonadaceae*)、卟啉单胞菌属 (*Porphyromonas*)、普雷沃菌属 (*Prevotella*)、假丁酸弧菌属

(*Pseudobutyrvibrio*)、冷弯菌属(*Psychroflexus*)、赖兴氏杆菌(*Reichenbachiella*)、红杆菌属(*Rhodobacter*)、里氏杆菌属(*Riemerella*)、*Sinomicrobium*、海旋菌属(*Thalassospira*)、瘤胃球菌属(*Ruminococcus*)；优选地，沙氏纤毛菌(*Leptotrichia shahii*)、斯氏李斯特氏菌(*Listeria seeligeri*)、毛螺科菌细菌(*Lachnospiraceae bacterium*) (如Lb MA2020、Lb NK4A179、Lb NK4A144)、*Clostridium aminophilum* (如Ca DSM 10710)、鸡肉杆菌(*Carnobacterium gallinarum*) (如Cg DSM 4847)、丙酸杆菌(*Paludibacter propionicigenes*) (如Pp WB4)、魏斯特斯蒂芬氏李斯特氏菌(*Listeria weihenstephanensis*) (如Lw FSL R9-0317)、李斯特氏菌科细菌(*Listeriaceae bacterium*) (如Lb FSL M6-0635)、瓦氏细单胞菌(*Leptotrichia wadei*) (如Lw F0279)、荚膜红杆菌(*Rhodobacter capsulatus*) (如Rc SB 1003、Rc R121、Rc DE442)、口腔纤毛菌(*Leptotrichia buccalis*) (如Lb C-1013-b)、*Herbinix hemicellulosilytica*、优杆菌科细菌(*Eubacteriaceae bacterium*) (如Eb CHKCI004)、布劳特氏菌属种(*Blautia sp.*) Marseille-P2398、纤毛菌属种(*Leptotrichia sp.*) 口腔分类单元879str.F0557、聚集绿屈扰菌(*Chloroflexus aggregans*)、*Demequina aurantiaca*、海旋菌种(*Thalassospira sp.*) TSL5-1、假丁酸弧菌属种(*Pseudobutyrvibrio sp.*) OR37、丁酸弧菌属种(*Butyrvibrio sp.*) YAB3001、纤毛菌属种(*Leptotrichia sp.*) Marseille-P3007、*Bacteroides ihuae*、紫单孢菌科细菌(*Porphyromonadaceae bacterium*) (如Pb KH3CP3RA)、*Listeria riparia*、*Insolitispirillum peregrinum*、别样杆菌属种(*Alistipes sp.*) ZOR0009、酿脓拟杆菌(*Bacteroides pyogenes*) (如Bp F0041)、拟杆菌属细菌(*Bacteroidetes bacterium*) (如Bb GWA2\_31\_9)、动物溃疡伯格菌(*Bergeyella zoohelcum*) (如Bz ATCC 43767)、犬咬二氧化碳噬纤维菌(*Capnocytophaga canimorsus*)、希诺地米碳酸噬胞菌(*Capnocytophaga cynodegmi*)、*Chryseobacterium carnipullorum*、济州岛金黄杆菌(*Chryseobacterium jejuense*)、嗜鳃黄杆菌(*Chryseobacterium ureilyticum*)、嗜分支黄杆菌(*Flavobacterium branchiophilum*)、柱状黄杆菌(*Flavobacterium columnare*)、黄杆菌属种(*Flavobacterium sp.*) 316、*Myroides odoratimimus* (如Mo CCUG 10230、Mo CCUG 12901、Mo CCUG 3837)、*Paludibacter propionicigenes*、*Phaeodactylibacter xiamenensis*、牙龈卟啉单胞菌(*Porphyromonas gingivalis*) (如Pg F0185、Pg F0568、Pg JCVI SC001、Pg W4087)、马铃薯红单胞菌(*Porphyromonas gulae*)、卟啉单胞菌属种(*Porphyromonas sp.*) COT-0520H4946、树蛙普雷沃菌(*Prevotella aurantiaca*)、颊普雷沃菌(*Prevotella buccae*) (如Pb ATCC 33574)、*Prevotella falsenii*、中间普雷沃菌(*Prevotella intermedia*) (如Pi 17、Pi ZT)、*Prevotella pallens* (如Pp ATCC 700821)、*Prevotella pleuritidis*、解糖普雷沃菌(*Prevotella saccharolytica*) (如Ps F0055)、普雷沃菌属种(*Prevotella sp.*) MA2016、普雷沃菌属种(*Prevotella sp.*) MSX73、普雷沃菌属种(*Prevotella sp.*) P4-76、普雷沃菌属种(*Prevotella sp.*) P5-119、普雷沃菌属种(*Prevotella sp.*) P5-125、普雷沃菌属种(*Prevotella sp.*) P5-60、扭曲冷弯菌(*Psychroflexus torquis*)、*Reichenbachiella agariperforans*、鸭疫里默氏杆菌(*Riemerella anatipestifer*)、*Sinomicrobium oceani*、坏死梭杆菌(*Fusobacterium necrophorum*) (如Fn subsp.funduliforme ATCC 51357、Fn DJ-2、Fn BFTR-1、Fn subsp.Funduliforme)、坏疽梭杆菌(*Fusobacterium perfoetens*) (如

Fp ATCC 29250)、溃疡梭杆菌(*Fusobacterium ulcerans*) (如Fu ATCC 49185)、厌氧唾液杆菌属种(*Anaerosalibacter* sp.)ND1、产亚硝酸真杆菌(*Eubacterium siraeum*)、黄化瘤胃球菌(*Ruminococcus flavefaciens*) (如Rfx XPD3002)或白色瘤胃球菌(*Ruminococcus albus*)。

[0039] 在一些实施方式中,所述Cas13蛋白是Cas13a蛋白。

[0040] 在一些实施方式中,所述Cas13a蛋白是或源于以下的种:拟杆菌属(*Bacteroides*)、布劳特氏菌属(*Blautia*)、丁酸弧菌属(*Butyrivibrio*)、肉食杆菌属(*Carnobacterium*)、绿曲挠丝状菌属(*Chloroflexus*)、梭菌属(*Clostridium*)、Demequina、优杆菌属(*Eubacterium*)、Herbinix、*Insolitispirillum*、毛螺菌科(*Lachnospiraceae*)、纤毛菌属(*Leptotrichia*)、李斯特氏菌属(*Listeria*)、杆状杆菌(*Paludibacter*)、紫单胞菌科(*Porphyromonadaceae*)、假丁酸弧菌属(*Pseudobutyrvibrio*)、红杆菌属(*Rhodobacter*)或海旋菌属(*Thalassospira*);优选地,沙氏纤毛菌(*Leptotrichia shahii*)、斯氏李斯特氏菌(*Listeria seeligeri*)、毛螺科菌细菌(*Lachnospiraceae bacterium*) (如Lb MA2020、Lb NK4A179、Lb NK4A144)、*Clostridium aminophilum* (如Ca DSM 10710)、鸡肉杆菌(*Carnobacterium gallinarum*) (如Cg DSM 4847)、丙酸杆菌(*Paludibacter propionicigenes*) (如Pp WB4)、魏斯特斯蒂芬氏李斯特氏菌(*Listeria weihenstephanensis*) (如Lw FSL R9-0317)、李斯特氏菌科细菌(*Listeriaceae bacterium*) (如Lb FSL M6-0635)、瓦氏细单胞菌(*Leptotrichia wadei*) (如Lw F0279)、荚膜红杆菌(*Rhodobacter capsulatus*) (如Rc SB 1003、Rc R121、Rc DE442)、口腔纤毛菌(*Leptotrichia buccalis*) (如Lb C-1013-b)、Herbinix hemicellulosilytica、优杆菌科细菌(*Eubacteriaceae bacterium*) (如Eb CHKCI004)、布劳特氏菌属种(*Blautia* sp.) Marseille-P2398、纤毛菌属种(*Leptotrichia* sp.) 口腔分类单元879str.F0557、聚集绿屈扰菌(*Chloroflexus aggregans*)、Demequina aurantiaca、海旋菌种(*Thalassospira* sp.) TSL5-1、假丁酸弧菌属种(*Pseudobutyrvibrio* sp.) OR37、丁酸弧菌属种(*Butyrivibrio* sp.) YAB3001、纤毛菌属种(*Leptotrichia* sp.) Marseille-P3007、*Bacteroides ihuae*、紫单胞菌科细菌(*Porphyromonadaceae bacterium*) (如Pb KH3CP3RA)、*Listeria riparia*或*Insolitispirillum peregrinum*。

[0041] 在一些实施方式中,所述Cas13蛋白是Cas13b蛋白。

[0042] 在一些实施方式中,所述Cas13b蛋白是或源于以下的种:别样杆菌属(*Alistipes*)、拟杆菌属(*Bacteroides*)、拟杆菌门(*Bacteroidetes*)、伯杰菌属(*Bergeyella*)、二氧化碳噬纤维菌属(*Capnocytophaga*)、金黄杆菌属(*Chryseobacterium*)、黄杆菌属(*Flavobacterium*)、香味菌属(*Myroides*)、杆状杆菌(*Paludibacter*)、细菌杆菌(*Phaeodactylibacter*)、卟啉单胞菌属(*Porphyromonas*)、普雷沃菌属(*Prevotella*)、冷弯菌属(*Psychroflexus*)、赖兴氏杆菌(*Reichenbachiella*)、里氏杆菌属(*Riemerella*)或*Sinomicrobium*;优选地,别样杆菌属种(*Alistipes* sp.) ZOR0009、酿脓拟杆菌(*Bacteroides pyogenes*) (如Bp F0041)、拟杆菌属细菌(*Bacteroidetes bacterium*) (如Bb GWA2\_31\_9)、动物溃疡伯格菌(*Bergeyella zoohelcum*) (如Bz ATCC 43767)、犬咬二氧化碳噬纤维菌(*Capnocytophaga canimorsus*)、希诺地米碳酸噬胞菌(*Capnocytophaga cynodegmi*)、*Chryseobacterium carnipullorum*、济州岛金黄杆菌(*Chryseobacterium*

jejuense)、嗜鳃黄杆菌 (*Chryseobacterium ureilyticum*)、嗜分支黄杆菌 (*Flavobacterium branchiophilum*)、柱状黄杆菌 (*Flavobacterium columnare*)、黄杆菌属种 (*Flavobacterium* sp.) 316、*Myroides odoratimimus* (如 Mo CCUG 10230、Mo CCUG 12901、Mo CCUG 3837)、丙酸杆菌 (*Paludibacter propionicigenes*)、*Phaeodactylibacter xiamenensis*、牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) (如 Pg F0185、Pg F0568、Pg JCVI SC001、Pg W4087)、马铃薯红单胞菌 (*Porphyromonas gulae*)、卟啉单胞菌属种 (*Porphyromonas* sp.) COT-0520H4946、树蛙普雷沃菌 (*Prevotella aurantiaca*)、颊普雷沃菌 (*Prevotella buccae*) (如 Pb ATCC 33574)、*Prevotella falsenii*、中间普雷沃菌 (*Prevotella intermedia*) (如 Pi 17、Pi ZT)、*Prevotella pallens* (如 Pp ATCC 700821)、*Prevotella pleuritidis*、解糖普雷沃菌 (*Prevotella saccharolytica*) (如 Ps F0055)、普雷沃菌属种 (*Prevotella* sp.) MA2016、普雷沃菌属种 (*Prevotella* sp.) MSX73、普雷沃菌属种 (*Prevotella* sp.) P4-76、普雷沃菌属种 (*Prevotella* sp.) P5-119、普雷沃菌属种 (*Prevotella* sp.) P5-125、普雷沃菌属种 (*Prevotella* sp.) P5-60、扭曲冷弯曲菌 (*Psychroflexus torquis*)、*Reichenbachiella agariperforans*、鸭疫里默氏杆菌 (*Riemerella anatipestifer*) 或 *Sinomicrobium oceani*。

[0043] 在一些实施方式中,所述Cas13蛋白是Cas13c蛋白。

[0044] 在一些实施方式中,所述Cas13c蛋白是或源于以下的种:梭杆菌属 (*Fusobacterium*) 或厌氧唾液杆菌 (*Anaerosalibacter*); 优选地,坏死梭杆菌 (*Fusobacterium necrophorum*) (如 Fn subsp. funduliforme ATCC 51357、Fn DJ-2、Fn BFTR-1、Fn subsp. Funduliforme)、坏疽梭杆菌 (*Fusobacterium perfoetens*) (如 Fp ATCC 29250)、溃疡梭杆菌 (*Fusobacterium ulcerans*) (如 Fu ATCC 49185) 或厌氧唾液杆菌属种 (*Anaerosalibacter* sp.) ND1。

[0045] 在一些实施方式中,所述Cas13蛋白是Cas13d蛋白。

[0046] 在一些实施方式中,所述Cas13d蛋白是或源于以下的种:优杆菌属 (*Eubacterium*) 或瘤胃球菌属 (*Ruminococcus*), 优选地,产亚硝酸真杆菌 (*Eubacterium siraeum*)、黄化瘤胃球菌 (*Ruminococcus flavefaciens*) (如 Rfx XPD3002) 或白色瘤胃球菌 (*Ruminococcus albus*)。

[0047] 在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的催化活性增加。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的催化活性降低。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的gRNA结合增加。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的gRNA结合降低。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的特异性增加。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的特异性降低。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的稳定性增加。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的稳定性降低。在一些实施方式中,所述工程化CRISPR-Cas蛋白还包含使催化活性失活的一个或多个突变。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的脱靶结合增加。在一些实施方式中,与对应的野生型CRISPR-Cas

蛋白相比,所述CRISPR-Cas蛋白的脱靶结合降低。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的靶标结合增加。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的靶标结合降低。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白具有更高蛋白酶活性或多核苷酸结合能力。在一些实施方式中,与对应的野生型CRISPR-Cas蛋白相比,PFS识别被改变。在一些实施方式中,所述工程化CRISPR-Cas蛋白还包含功能性异源结构域。在一些实施方式中,所述工程化CRISPR-Cas蛋白还包含NLS。

[0048] 在另一个方面中,本公开内容提供了一个或多个HEPN结构域,且长度小于1000个氨基酸。在一些实施方式中,所述蛋白的尺寸小于950个、小于900个、小于850个、小于800个、或小于750个氨基酸。在一些实施方式中,所述HEPN结构域包含RxxxxH基序。在一些实施方式中,所述RxxxxH基序包含R[N/H/K]X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H序列。在一些实施方式中,X<sub>1</sub>是R、S、D、E、Q、N、G或Y,X<sub>2</sub>独立地是I、S、T、V或L,和X<sub>3</sub>独立地是L、F、N、Y、V、I、S、D、E或A。在一些实施方式中,所述CRISPR-Cas蛋白是VI型CRISPR Cas蛋白。在一些实施方式中,所述VI型CRISPR Cas蛋白是Cas13a、Cas13b、Cas13c或Cas13d。在一些实施方式中,所述CRISPR-Cas蛋白与功能性结构域相关联。在一些实施方式中,所述CRISPR-Cas蛋白包含与本文所述的突变等价的一个或多个突变。在一些实施方式中,所述CRISPR-Cas蛋白包含在螺旋结构域中的一个或多个突变。在一些实施方式中,所述CRISPR-Cas蛋白是死形式或具有切口酶活性。

[0049] 在另一个方面中,本公开内容提供了一种多核苷酸,其编码本文的工程化CRISPR-Cas蛋白。在一些实施方式中,所述多核酸是密码子优化的。

[0050] 在另一个方面中,本公开内容提供了一种CRISPR-Cas系统,其包含本文的工程化CRISPR-Cas蛋白,以及能够与所述工程化CRISPR-Cas蛋白形成复合物并能够与靶核酸序列杂交并引导所述复合物与所述靶核酸序列的序列特异性结合的核苷酸组分。

[0051] 在另一个方面中,本公开内容提供了一种载体系统,其包含一个或多个载体,所述一个或多个载体包含编码所述工程化CRISPR-Cas蛋白的组分的一个或多个多核苷酸分子。

[0052] 在另一个方面中,本公开内容提供了一种修饰靶核酸的方法,其包括:在包含所述靶核酸的细胞或生物体中引入本文所述工程化CRISPR-Cas蛋白,所述多核苷酸,所述CRISPR-Cas系统,或者所述载体或载体系统,使得所述工程化CRISPR-Cas蛋白修饰在所述细胞或生物体中的所述靶核酸。

[0053] 在一些实施方式中,所述工程化CRISPR-Cas系统是通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪、可植入装置或本文的载体系统递送引入。在一些实施方式中,所述工程化CRISPR-cas蛋白与一个或多个功能性结构域相关联。在一些实施方式中,所述靶核酸包含基因组基因座,并且所述工程化CRISPR-Cas蛋白改变在所述基因组基因座处编码的基因产物或所述基因产物的表达。在一些实施方式中,所述靶核酸是DNA或RNA,且其中所述靶核酸中的一个或多个核苷酸被碱基编辑。在一些实施方式中,所述靶核酸是DNA或RNA,且其中所述靶核酸被切割。在一些实施方式中,所述工程化CRISPR-Cas蛋白还切割非靶核酸。在一些实施方式中,所述方法还包括使活性可视化,以及任选地,使用可检测标记。在一些实施方式中,所述方法还包括检测所述CRISPR-Cas系统的一个或多个组分与所述靶核酸的结合。在一些实施方式中,所述细胞或生物体是原核细胞或生物体。在一些实施方式中,所述细胞或生物体是动物细胞或生物体。在一些实施方式中,所述细胞或生物体是

植物细胞或生物体。

[0054] 在另一个方面中,本公开内容提供了用于检测样品中的靶核酸的方法,其包括:将样品与下述接触:本文的工程化CRISPR-Cas蛋白;至少一种指导多核苷酸,其包含能够结合至所述靶核酸并被设计为与所述工程化CRISPR-Cas形成复合物的指导序列;和包含非靶序列的基于RNA的掩蔽构建体;其中所述工程化CRISPR-Cas蛋白表现出附带RNase活性并切割所述检测构建体的所述非靶序列;和检测来自所述非靶序列的切割的信号,从而检测在所述样品中的所述靶核酸。

[0055] 在一些实施方式中,所述方法还包括将所述样品与用于扩增所述靶核酸的试剂接触。在一些实施方式中,用于扩增的所述试剂包括等温扩增反应试剂。在一些实施方式中,所述等温扩增试剂包括基于核酸序列的扩增、重组酶聚合酶扩增、环介导的等温扩增、链置换扩增、解旋酶依赖性扩增或切口酶扩增试剂。在一些实施方式中,所述靶核酸是DNA分子,并且所述方法还包括将所述靶DNA分子与包含RNA聚合酶位点的引物和RNA聚合酶接触。在一些实施方式中,所述掩蔽构建体:抑制可检测阳性信号的产生直至所述掩蔽构建体被切割或失活,或者掩蔽可检测阳性信号或产生可检测阴性信号直至所述掩蔽构建体被切割或失活。

[0056] 在一些实施方式中,所述掩蔽构建体包含:a.沉默RNA,其抑制由报告构建体编码的基因产物的产生,其中当表达时所述基因产物产生所述可检测阳性信号;b.核酶,其产生所述可检测阴性信号,并且其中当所述核酶失活时产生所述阳性可检测信号;c.核酶,其将底物转化为第一颜色,并且其中当所述核酶失活时所述底物转化为第二颜色;d.适体和/或包含多核苷酸栓系的抑制剂;e.多核苷酸,其附着有可检测配体和掩蔽组分;f.通过桥分子保持在聚集体中的纳米颗粒,其中至少一部分的所述桥分子包含多核苷酸,并且其中当所述纳米颗粒被支付在溶液中时所述溶液经历色移;g.通过连接分子连接至一个或多个淬灭剂分子的量子点或荧光团,其中至少一部分的所述连接分子包含多核苷酸;h.与嵌入剂复合的多核苷酸,其中在切割所述多核苷酸后所述嵌入剂改变吸光度;或者1.通过多核苷酸栓系的两个荧光团,其当从所述多核苷酸释放时经历荧光移位。

[0057] 在一些实施方式中,所述适体a.包含螯合酶的多核苷酸栓系的抑制剂,其中在通过作用于底物而从所述适体或多核苷酸栓系的抑制剂释放后所述酶产生可检测信号;或者b.是抑制性适体,其抑制酶并阻止酶催化底物产生可检测信号,或其中所述多核苷酸栓系的抑制剂抑制酶并阻止酶催化底物产生可检测信号;或者c.螯合剂对试剂,所述试剂当从所述适体释放时组合以产生可检测信号。在一些实施方式中,所述纳米颗粒是胶体金属。在一些实施方式中,所述至少一个指导多核苷酸包含错配。在一些实施方式中,所述错配是在所述一个或多个指导序列上的单核苷酸改变的上游或下游。

[0058] 在另一个方面中,本公开内容提供了一种细胞或生物体,其包含本文的工程化CRISPR-Cas蛋白,本文的多核酸,本文的CRISPR-Cas系统或者本文的载体或载体系统。

[0059] 在另一个方面中,本公开内容提供了一种工程化腺苷脱氨酶,其包含一个或多个突变,其中所述工程化腺苷脱氨酶具有胞苷脱氨酶活性。

[0060] 在一些实施方式中,所述工程化腺苷脱氨酶具有胞苷脱氨酶活性。在一些实施方式中,所述工程化腺苷脱氨酶是融合蛋白的一部分。在一些实施方式中,所述融合蛋白包含功能性结构域。在一些实施方式中,所述功能性结构域能够引导所述工程化腺苷脱氨酶结

合至靶核酸。在一些实施方式中,所述功能性结构域是本文的CRISPR-Cas蛋白。在一些实施方式中,所述CRISPR-Cas蛋白是死形式的CRISPR-Cas蛋白或 CRISPR-Cas切口酶蛋白。在一些实施方式中,基于hADAR2-D的氨基酸序列位置,所述一个或多个突变包含:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T,以及在同源ADAR蛋白中的相应突变。在一些实施方式中,基于hADAR2-D的氨基酸序列位置,所述一个或多个突变包含:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E 和S661T,以及在同源ADAR蛋白中的相应突变。

[0061] 在另一个方面中,本公开内容提供了一种多核苷酸,其编码所述的工程化腺苷脱氨酶或其催化结构域。在另一个方面中,本公开内容提供了包括多核苷酸。

[0062] 在另一个方面中,本公开内容提供了一种药物组合物,其包含配制成用于通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪或可植入装置递送的所述工程化腺苷脱氨酶或其催化结构域。

[0063] 在另一个方面中,本公开内容提供了一种工程化细胞,其表达所述工程化腺苷脱氨酶或其催化结构域。在一些实施方式中,所述细胞瞬时表达所述工程化腺苷脱氨酶或其催化结构域。在一些实施方式中,所述细胞非瞬时表达所述工程化腺苷脱氨酶或其催化结构域。

[0064] 另一个方面,本公开内容提供了一种用于在靶核酸中修饰核苷酸的工程化、非天然存在的系统,其包含:a)死CRISPR-Cas或CRISPR-Cas 切口酶蛋白,或者编码所述死Cas或Cas切口酶蛋白的核苷酸序列;b) 指导分子,其包含与靶序列杂交并设计为与所述死CRISPR-Cas或 CRISPR-Cas切口酶蛋白形成复合物的指导序列;和c)核苷酸脱氨酶蛋白或其催化结构域,或者编码所述核苷酸脱氨酶蛋白或其催化结构域的核苷酸序列,其中所述核苷酸脱氨酶蛋白或其催化结构域共价或非共价连接至所述死CRISPR-Cas或CRISPR-Cas切口酶蛋白,或者所述指导分子适于在递送后与其连接。

[0065] 在一些实施方式中,所述腺苷脱氨酶蛋白或其催化结构域基于 hADAR2-D的氨基酸序列位置包含突变E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T中的一个或多个,以及以及在同源ADAR 蛋白中的相应突变。在一些实施方式中,所述腺苷脱氨酶蛋白或其催化结构域基于hADAR2-D的氨基酸序列位置包含突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E和S661T,以及在同源ADAR蛋白中的相应突变。

[0066] 在一些实施方式中,所述CRISPR-Cas蛋白是Cas9、Cas12、Cas13、Cas 14、CasX或CasY。在一些实施方式中,所述CRISPR-Cas蛋白是 Cas13b。在一些实施方式中,所述CRISPR-Cas蛋白是Cas13b-t1、Cas13b-t2或Cas13b-t3。在一些实施方式中,所述CRISPR-Cas是工程化 CRISPR-Cas蛋白。

[0067] 在另一个方面中,本公开内容提供了一种用于在靶核酸中修饰核苷酸的方法,其包括:将所述工程化腺苷脱氨酶或所述系统递送至所述靶核酸,其中所述脱氨酶使在所述靶核酸上的一个或多个靶基因座处的核苷酸脱氨基。

[0068] 在一些实施方式中,所述核苷酸脱氨酶蛋白或其催化结构域已被修饰以增加对

DNA-RNA异源双链体的活性。在一些实施方式中,所述核苷酸脱氨酶蛋白或其催化结构域已被修饰以降低脱靶效应。在一些实施方式中,所述靶核酸在细胞内。在一些实施方式中,所述细胞是真核细胞。在一些实施方式中,所述细胞是非人动物细胞。在一些实施方式中,所述细胞是人细胞。在一些实施方式中,所述细胞是植物细胞。在一些实施方式中,所述靶核酸在动物内。在一些实施方式中,所述靶核酸在植物内。在一些实施方式中,所述靶核酸在体外包含在DNA分子中。在一些实施方式中,所述工程化腺苷脱氨酶或者所述系统的一种或多种组分作为核糖核蛋白复合物递送到所述细胞中。在一些实施方式中,所述工程化腺苷脱氨酶或者所述系统的一种或多种组分通过一种或多种颗粒、一种或多种囊泡或者一种或多种病毒载体递送。在一些实施方式中,所述一种或多种颗粒包含脂质、糖、金属或蛋白。在一些实施方式中,所述一种或多种颗粒包括脂质纳米颗粒。在一些实施方式中,所述一种或多种囊泡包括外泌体或脂质体。在一些实施方式中,所述一种或多种病毒载体包括一种或多种腺病毒载体,一种或多种慢病毒载体或者一种或多种腺相关病毒载体。在一些实施方式中,所述方法通过在感兴趣的基因组基因座处操纵一个或多个靶序列来修饰细胞、细胞系或生物体。在一些实施方式中,在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用治疗由G→A或C→T点突变或病原性SNP引起的疾病。在一些实施方式中,所述疾病选自癌症、血友病、β地中海贫血,马凡氏综合征和维斯科特-奥尔德里奇综合征。在一些实施方式中,在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用治疗由T→C或A→G点突变或病原性SNP引起的疾病。在一些实施方式中,在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用使在所述靶基因座处的靶基因失活。在一些实施方式中,通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪、可植入装置或所述载体系统递送所述工程化腺苷脱氨酶或者所述系统的一种或多种组分。在一些实施方式中,所述核苷酸的修饰改变在所述靶基因座处编码的基因产物或所述基因产物的表达。

[0069] 结合以下对所说明的示例性实施方式的详细说明,示例性实施方式的这些和其他方面、目的、特征和优点对于本领域普通技术人员而言将变得显而易见。

## 附图说明

[0070] 通过参考对说明性实施方式进行阐述的以下详细说明,将获得对本发明的特征和优点的更好理解,在这些实施方式中利用了本发明的原理,并且在这些附图中:

[0071] 图1A-1D。PbuCas13b-crRNA二元复合物的晶体结构。(图1A) PbuCas13b的线性结构域组织。活性位点的位置用星号表示。(图1B) 与PbuCas13b复合的crRNA发夹。(图1C) PbuCas13b的总体结构。两个视图彼此旋转180度。结构域的颜色与线性结构域图一致。crRNA涂成红色。(图1D) PbuCas13b的空间填充模型,每个视图彼此旋转180度。

[0072] 图2A-2E。PbuCas13b crRNA识别。(图2A) PbuCas13b crRNA的图 (SEQ ID NO:1)。正向重复残基涂成红色,间隔子残基的颜色为浅蓝色。(图2B) 接近K393的crRNA的3'末端位置和PbuCas13b内的配位残基。(图2C) 在PbuCas13b复合物内crRNA的结构。颜色与嵌板一致(图2A)。(图2D) 碱基识别交换。上图,核酸酶活性;下图,热稳定性。散列填充表示野生型碱基识别。(图2E) PbuCas13b内配位和处理crRNA的盖结构域残基的诱变。上图,SHERLOCK反应中的RNase活性;下图,crRNA处理。切割条带和预期大小用红色标记物表示,带有大小的梯形显示在左侧。

[0073] 图3.PbuCas13b和crRNA (SEQ ID NO:2) 之间的分子间接触的示意图。

[0074] 图4A-4C.PbuCas13b与LshCas13a结构和活性位点的比较。(图 4A) PbuCas13b和LshCas13a (pdb 5wtk) 的结构域组织的线性比较。crRNA 显示在右侧。(图4B) PbuCas13b的两个视图旋转90度。插图按与(图 4C) 相同的方向放大了活性位点的残基。(图4C) LshCas13a的颜色与(图4A) 一致。标记同源残基。

[0075] 图5A-5H.PbuCas13b的位点定向诱变;在哺乳动物细胞中的RNA 干扰。(图5A) 所有all PbuCas13b位点定向突变对哺乳动物比中RNA 干扰的影响。最强的干扰敲低用浅蓝色表示。(图5B) 带有标记的强突变的PbuCas13b,并用红色表示。(图5C-5H) 按区域分隔的突变。

[0076] 图6A-6D。(图6A) PbuCas13b的表面静电。(图6B) 从图A旋转180度的PbuCas13b的表面静电。(图6C) 除去盖结构域的PbuCas13b 的表面静电,显示了内部带正电的通道。(图6D) 推定crRNA处理活性位点的表面静电。

[0077] 图7.pgCas13b C末端截短的REPAIR测定。

[0078] 图8A-8G。(图8A) PbuCas13b正向重复结构。(图8B) 理想的A 型RNA。(图8C) 正向重复碱基对和二级结构 (SEQ ID NO:3) 的图。(图8D) 多重一。(图8E) 多重二。(图8F) 多重三。(图8G) PbuCas13b 正向重复序列 (SEQ ID NO:4-9) 的比对。星号表示保守的核苷酸。

[0079] 图9.带有突变的crRNA的PbuCas13的切割活性,以及crRNA突变体的热稳定性的扩展数据。

[0080] 图10A-10D。(图10A) 用于处理测定的crRNA底物的示意图 (SEQ ID NO:10-11)。(图10B) 显示未处理的互补DR的凝胶。(图10C) PbuCas13b突变体的crRNA处理。(图10D) 测量一般RNase活性的 SHERLOCK测定。

[0081] 图11A-11C.使用底物RNA和镁离子的PbuCas13b的熔解曲线。(图 11A) RNA底物对PbuCas13b热稳定性的影响。(图11B) PbuCas13b RNA 切割和热稳定性的影响。(图11C) 镁对PbuCas13b热稳定性的影响。

[0082] 图12.PbuCas13b与RNA底物的有限蛋白水解作用。PbuCas13b的有限蛋白水解作用。T=胰蛋白酶,C=糜蛋白酶,P=胃蛋白酶

[0083] 图13A-13C.Cas13b桥螺旋。(图13A) 带有桥螺旋的Cas13b用红色突出显示。RNA用粉色表示。(图13B) 带有桥螺旋的Cas12 (Cpf1) 用青色突出显示。RNA用浅蓝色表示,DNA用深蓝色表示。(图13C) 来自PbuCas13b和LbCas12的桥螺旋的人工序列比对 (SEQ ID NO:12-13)。

[0084] 图14.所有Cas13b家族成员的Cas13b邻接树。插图,具有PbuCas13b 的Cas13b子集(粗体)。

[0085] 图15.Cas13b亚组的基于结构的比对 (SEQ ID NO:14-22)。

[0086] 图16.所有Cas13bs的基于结构的比对 (SEQ ID NO:23-37)。

[0087] 图17A-17D.图中显示了所有凝胶的未裁剪图像。(图17A) crRNA 处理凝胶1。(图17B) crRNA处理凝胶2。(图17C) crRNA处理凝胶 3。(图17D) 有限的蛋白水解凝胶。

[0088] 图18.PbuCas13b晶体结构的分组拓扑图。

[0089] 图19显示了一个pymol文件,该文件显示了Cas13b活性位点中配位核苷酸的位置。

[0090] 图20显示了一个示例性RNA环延伸。

- [0091] 图21显示了示例性融合点,核苷酸脱氨酶通过其连接至Cas13b。
- [0092] 图22显示了RESCUE v9突变的筛选。
- [0093] 图23显示了RESCUEv9对T翻转指导的作用的验证。
- [0094] 图24显示了RESCUEv9对C翻转指导的作用的验证。
- [0095] 图25显示了在内源性靶向方面RESCUEv9的性能。
- [0096] 图26显示了RESCUE v10突变的筛选。
- [0097] 图27显示了用于C翻转的30bp指导的检测结果。
- [0098] 图28显示了来自使用RESCUE v1至v8在Cas13b6和Cas13b12之间比较的Gluc/Cluc结果。
- [0099] 图29显示了来自使用RESCUE v1至v8在Cas13b6和Cas13b12之间比较的组分编辑结果。
- [0100] 图30显示了来自使用RESCUEv8在Cas13b6和Cas13b12之间比较的对内源性靶向(T翻转)的影响的结果。
- [0101] 图31显示了RESCUE对碱基转换的作用。
- [0102] 图32显示了CCN的3' 基序靶向的检测结果。
- [0103] 图33A显示了具有与ADAR融合的dCas13b的构建体的示意图。图 33B显示了构建体的检测结果。
- [0104] 图34显示了N末端标签和接头的测序。
- [0105] 图35显示了脱靶的定量。
- [0106] 图36显示了脱靶编辑的测试。
- [0107] 图37显示了具有(GGS) 2/Q507R的内源性基因靶点的检测结果。
- [0108] 图38和图39显示了在(GGS) 2/Q507R上的突变的eGFP筛选。
- [0109] 图40A显示了具有Cas13b截短的构建体。图40B显示了构建体的检测结果。
- [0110] 图41显示了用于筛选的多道开关靶点指导(SEQ ID NO:38-39)。
- [0111] 图42A-42E显示了在RESCUEv10上的验证检测。图42A显示了 RESCUEv10的验证(第50、52轮)。图42B显示了RESCUEv10的验证(第53、54轮)。图42C显示了RESCUEv10的验证(第58轮)。图42D显示了RESCUEv10的验证(第59轮)。图42E显示了RESCUEv10 的验证(第61轮)。
- [0112] 图43显示了RESCUEv10的NGS分析。
- [0113] 图44显示了已鉴定的提高特异性的突变。
- [0114] 图45显示了RESCUE对内源性靶向(C翻转和T翻转)结果的影响。
- [0115] 图46显示了使用RESCUE v6和v9靶向 $\beta$ -连环蛋白。
- [0116] 图47显示了新的 $\beta$ -连环蛋白分泌的Gluc/Cluc报告基因。
- [0117] 图48显示了RESCUEv10靶向 $\beta$ -连环蛋白的结果。
- [0118] 图49显示了RESCUEv10靶向ApoE4。
- [0119] 图50显示了能够使用RESCUE产生的在PCSK9中的示例性突变。
- [0120] 图51显示了来自Cas13b-t1对哺乳动物细胞中Gluc敲低的结果。
- [0121] 图52显示了来自Cas13b-t2对哺乳动物细胞中Gluc敲低的结果。
- [0122] 图53显示了来自Cas13b-t3对哺乳动物细胞中Gluc敲低的结果。
- [0123] 图54A-54C显示了Cas13b-t1、Cas13b-t2和Cas13b-t3的基因座。

- [0124] 图55A-55C显示了Cas13b-t1、Cas13b-t2和Cas13b-t3 (SEQ ID NO: 40-45) 的基因座上的更多细节。
- [0125] 图56显示了Cas13b-t1、Cas13b-t2和Cas13b-t3与其他Cas13b直系同源物 (SEQ ID NO: 46-64) 的比对。
- [0126] 图57显示了筛选的RESCUE突变的总结。
- [0127] 图58是说明选择更好的 $\beta$ -连环蛋白突变体的实验结果的图。
- [0128] 图59显示了说明RESCUE第12轮结果的图。
- [0129] 图60是说明 $\beta$ -连环蛋白迁移测定的示意图。
- [0130] 图61是显示由 $\beta$ -连环蛋白诱导的细胞迁移测定结果的图。
- [0131] 图62显示了说明特异性突变消除A-I脱靶的图。
- [0132] 图63显示了说明靶向Stat1/3磷酸化位点降低信号传导的图。
- [0133] 图64显示了说明靶向Stat1/3磷酸化位点降低信号传导的图 (STAT1 未处理 (左图) 和STAT1 IFN  $\gamma$  处理 (右图))。
- [0134] 图65显示了说明靶向Stat1/3磷酸化位点降低信号传导的图, 其中图65A显示了STAT3 IL6活化的结果和图65B显示了STAT3未处理的结果。
- [0135] 图66显示了说明RESCUE第12轮结果的图。
- [0136] 图67显示了说明可能的RESCUE第13轮结果的图。
- [0137] 图68是显示由 $\beta$ -连环蛋白诱导的细胞迁移测定结果的图。
- [0138] 图69显示了说明对Gluc敲低的死的和活的微小直系同源物比较结果的图。
- [0139] 图70显示了说明Cas13b-t1的检测功能的图。
- [0140] 图71显示了说明Cas13b-t3的检测功能的图。
- [0141] 图72显示了说明指导、非靶向比较的图。
- [0142] 图73A-73G: 针对胞苷脱氨的ADAR2脱氨酶结构域的定向进化。(图73A) 定向进化方法的示意图, 包括合理诱变、酵母筛选, 以及活性哺乳动物细胞验证。(图73B) RESCUE版本0-16在胞苷上的活性, 其侧翼是Gluc转录物上的5' U和3' G。左图: 报告了RESCUEv0-v16 的荧光素酶报告基因活性。右图: 报告了RESCUEv0-v16的编辑水平百分比。(图73C) 热图描述了在胞苷上RESCUEv0-v16的编辑水平百分比, 其侧翼是在Gluc转录物上的不同碱基。(图73D) 在转染的RESCUE 质粒水平不同的情况下, 在胞苷上RESCUEv0-v16的编辑水平百分比, 其侧翼是Gluc转录物上的5' U和3' G。(图73E) RESCUEv16和RESCUEv8对Gluc转录物上所有可能的16个胞苷侧翼碱基基序的编辑活性。使用具有来自靶胞苷的T翻转或C翻转的指导设计。(图73F) 将通过RESCUEv16进行的胞苷脱氨与使用指导RNA以及ADAR2dd、全长ADAR2或无蛋白的编辑进行比较。(图73G) 还显示了在催化脱氨位点处突变体的放大晶体结构图, 所述催化脱氨位点具有带有翻转碱基的RNA。
- [0143] 图74A-74G: 通过RESCUE对内源性和疾病相关靶点的C到U编辑。(图74A) RESCUEv16在覆盖多个基序的一组内源性基因上的编辑效率。(图74B) 描述在一组三个内源性基因上的RESCUE版本v0-v16的编辑效率的热图。(图74C) RESCUEv16在一组相关T>C疾病突变的合成版本上的编辑效率。(图74D) 使用前crRNA指导阵列对C到U和A 到I进行多重编辑的示意图。(图74E) 在 $\beta$ -连环蛋白转录物上同时进行C到U和A到I的编辑。(图74F) 通过引入不利的碱基翻转 (SEQ ID NO: 65-66), 合理防止相邻腺苷位点的脱靶活性的示意图。

(图74G) 使用合理引入不利的碱基翻转的高斯荧光素酶(左图)和KRAS(右图)的中靶C和脱靶A位点的编辑百分比。

[0144] 图75A-75F:RESCUEv16的转录组范围特异性。(图75A) RESCUE v16和B6-REPAIRv1、B12-REPAIRv1和B12-REPAIRv2的中靶C到U 编辑以及C到U和A到I的转录组范围脱靶的总结。(图75B) RESCUEv16 A到I和C到U脱靶的Manhattan曲线。中靶C到U编辑以橙色突出显示。(图75C) ADAR2dd残基和双链RNA底物之间相互作用的示意图,在用于提高特异性的诱变筛选中使用的残基以红色突出显示(SEQ ID NO:67-68)。(图75D) 针对RESCUEv16和95RESCUEv16 突变体显示了具有靶向指导(y轴)的C到U活性和具有非靶向指导(x轴)A到I活性的荧光素酶值。以蓝色突出显示的突变体具有有效靶向的C到U活性,但失去了其残留的A到I活性,表明A到I的特异性得到了改善。(图75E) RESCUE v16和最高特异性突变体的中靶C到U 编辑以及C到U和A到I的转录组范围脱靶的总结。(图75F) RESCUEv16(+S375A) A到I和C到U脱靶的Manhattan曲线(SEQ ID NO:65-66)。中靶C到U编辑以橙色突出显示。

[0145] 图76A-76H:针对细胞生长和信号传导的由C到U RNA编辑导向的表型结局。(图76A) 靶向STAT3磷酸化残基以改变相关信号通路的 RNA示意图(SEQ ID NO:69-74)。(图76B) RESCUEv16在STAT3(左图)和STAT1(右图)中相关磷酸化残基的编辑百分比。(图76C) 如通过STAT驱动荧光素酶表达所测量的,通过RNA编辑的STAT3(左图)和STAT1(右图)信号传导的抑制。(图76D) 靶向CTNNB1磷酸化残基以促进稳定化的RNA示意图(SEQ ID NO:75-77)。(图76E) 通过RESCUE编辑磷酸化残基激活 $\beta$ -连环蛋白,导致细胞生长增加的示意图。(图76F) RESCUEv16在CTNNB1中相关磷酸化残基的编辑百分比。(图76G) 如通过CTNNB1驱动(TCF/LEF)荧光素酶表达所测量的,通过RNA编辑的CTNNB1信号传导的活化。(图76H) 通过RNA 编辑的CTNNB1信号传导活化导致的细胞生长的定量。

[0146] 图77A-77B:筛选失活的G1uc突变,以产生胞苷脱氨荧光素酶报告基因。(图77A) 此前显示出对荧光素酶活性的某些影响[引用G1uc论文]的一组各种G1uc突变体的荧光素酶活性。值表示平均值 $\pm$ S.E.M( $n=3$ )。(图77B) 一组亮氨酸到脯氨酸G1uc突变体的荧光素酶活性。关注了亮氨酸到脯氨酸突变体报告基因,因为其产生了用于胞苷脱氨的CCN基序位点(中央C被脱氨)。这允许测定所有四个CCN基序对 RESCUE脱氨活性的影响。值表示平均值 $\pm$ S.E.M( $n=3$ )。

[0147] 图78:在G1uc中CCG、ACG、GCG、CCA和CCU位点上 RESCUEv0-v16的胞苷脱氨活性。值表示平均值 $\pm$ S.E.M( $n=3$ )。

[0148] 图79A-79B:各种量的RESCUEv0-16的胞苷脱氨活性。(图79A) 如通过在G1uc转录物的UCG位点上荧光素酶活性的恢复所测量的, RESCUEv0-v16活性的剂量应答。值表示三次重复的平均值。(图79B) 如通过在CTNNB1转录物的T41I位点上荧光素酶活性的恢复所测量的, RESCUEv0-v16活性的剂量应答。值表示三次重复的平均值。

[0149] 图80:RESCUEv6-v9以不同的指导和RESCUE质粒量对G1uc转录物中的UCG位点进行编辑的百分比。值表示平均值 $\pm$ S.E.M( $n=3$ )。

[0150] 图81:使用具有G或A错配指导,用RESCUEv16和v8,对具有16个可能的5'和3'碱基组合的G1uc位点的编辑百分比。值表示平均值 $\pm$ S.E.M( $n=3$ )。

[0151] 图82:使用不同U错配位置的指导RNA在G1uc转录物中的UCG位点上的RESCUEv1和RESCUEv2-v8的编辑百分比。将RESCUE版本与RanCas13b和PspCas13b两者进行比较。值表示

平均值 $\pm$ S.E.M(n= 3)。20/22表示针对RanCas13b的错配距离是20和针对PspCas13b的错配距离是22。

[0152] 图83:使用具有不同U错配位置的30bp和50bp指导在Gluc转录物中的UCG位点上的RESCUEv16的编辑百分比。值表示平均值 $\pm$  S.E.M(n=3)。

[0153] 图84A-84D:用于定向进化的各种酵母报告基因的编辑速率。(图 84A)通过具有靶向和非靶向指导的RESCUEv3、v7和v16的GFP突变 Y66H的荧光校正百分比。通过对10,000个细胞进行流式细胞术来测量荧光。(图84B)通过具有靶向和非靶向指导的RESCUEv3、v7和v16 的GFP突变Y66H的编辑校正百分比。值表示平均值 $\pm$ -S.E.M(n=3)。(图84C)通过具有靶向和非靶向指导的RESCUEv7和v16的HIS3突变P196L的编辑校正百分比。值表示平均值 $\pm$ -S.E.M(n=3)。(图84D) 通过具有靶向和非靶向指导的RESCUEv7和v16的HIS3突变S129P的编辑校正百分比。值表示平均值 $\pm$ -S.E.M(n=3)。

[0154] 图85A-85B:使用重组蛋白的含RESCUEv2突变的ADAR2脱氨结构域的生物化学脱氨活性。(图85A)包含RESCUEv2突变的ADAR2 脱氨酶结构域蛋白的腺苷脱氨活性,其中22bp双链RNA底物包含与胞苷错配的中心腺苷。将反应孵育不同时间点,并且具有和不具有脱氨酶结构域。(图85B)包含RESCUEv2突变的ADAR2脱氨酶结构域蛋白的胞苷脱氨活性,其中22bp双链RNA底物包含与尿苷错配的中心胞苷。将反应孵育不同时间点,并且具有和不具有脱氨酶结构域。将反应孵育不同时间点,并且具有和不具有脱氨酶结构域。

[0155] 图86A-86E:RESCUEv16、全长ADAR2(具有RESCUEv16突变)、ADAR2脱氨酶结构域(具有RESCUEv16突变)和不含任何蛋白的胞苷脱氨酶活性比较。(图86A)在具有不同5' 碱基的Gluc转录物中某一位点的编辑百分比,所述不同5' 碱基具有靶向指导以及RESCUEv16、全长 ADAR2(具有RESCUEv16突变)、ADAR2脱氨酶结构域(具有 RESCUEv16突变)和无蛋白。值表示平均值 $\pm$ -S.E.M(n=3)。(图 86B)在具有不同5' 碱基的Gluc转录物中某一位点的编辑百分比,所述不同5' 碱基具有非靶向指导以及RESCUEv16、全长ADAR2(具有 RESCUEv16突变)、ADAR2脱氨酶结构域(具有RESCUEv16突变) 和无蛋白。值表示平均值 $\pm$ -S.E.M(n=3)。(图86C)在具有RESCUEv16 和指导RNA的Gluc转录物中UCG位点的编辑,所述指导RNA含有不同错配位置。值表示平均值 $\pm$ -S.E.M(n=3)。(图86D)在具有全长 ADAR2(具有RESCUEv16突变)和指导RNA的Gluc转录物中UCG位点的编辑,所述指导RNA含有不同错配位置。值表示平均值 $\pm$ -S.E.M(n=3)。(图86E)在具有ADAR2脱氨酶结构域(具有RESCUEv16 突变)和指导RNA的Gluc转录物中UCG位点的编辑,所述指导RNA 含有不同错配位置。值表示平均值 $\pm$ -S.E.M(n=3)。

[0156] 图87A-87C:错配位置平铺,以发现在内源性靶位点上针对 RESCUEv16的最佳编辑指导设计。(图87A)具有不同碱基基序的内源性靶位点的编辑百分比,所述碱基基序具有RESCUEv16和在位置7、9、11和13具有错配的指导 and U型碱基翻转。值表示平均值 $\pm$ -S.E.M(n= 3)。(图87B)具有不同碱基基序的内源性靶位点的编辑百分比,所述碱基基序具有RESCUEv16和在位置7、9、11和13具有错配的指导 and C 型碱基翻转。值表示平均值 $\pm$ -S.E.M(n=3)。(图87C)具有不同碱基基序的内源性靶位点的编辑百分比,所述碱基基序具有RESCUEv16 和在位置3、5、7、9和11具有错配的指导 and C型和U型碱基翻转。值表示平均值 $\pm$ -S.E.M(n=3)。

[0157] 图88:如通过在KRAS位点的编辑百分比所测量的,各种量的RESCUEv0-16的胞苷脱

氨活性。值表示三次重复的平均值。

[0158] 图89:使用RESCUEv16和具有不同错配位置的指导在合成报告基因上各种与疾病相关的图标编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0159] 图90:使用RESCUEv16和具有不同C型和U型错配位置的指导在两个ApoE4胞苷(rs429358和rs7412)处的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0160] 图91A-91C:在指导双链体窗的RESCUE版本的特异性。(图91A) 高斯荧光素酶突变体C82R的编辑位点的示意图,其中靶向的C以红色突出显示,以及对附近的腺苷碱基编号并用灰色突出显示。(图91B) 在高斯荧光素酶突变体C82R中在附近腺苷碱基处的编辑百分比,其是由RESCUEv0、RESCUEv8和RESCUEv16靶向的。(图91C) 通过改变 RESCUEv0-v16的量,在腺苷20处,腺苷向鸟苷的编辑百分比。值表示三次重复的平均值。

[0161] 图92A-92D:RESCUEv0-v16和RESCUEv16S的腺苷脱氨酶活性。(图92A) 使用靶向指导RNA,利用RESCUEv0-v16和RESCUEv16S,通过G1uc转录物的腺苷脱氨的荧光素酶校正。值表示平均值 $\pm$ S.E.M(n=3)。(图92B) 使用非靶向指导RNA,利用RESCUEv0-v16和RESCUEv16S,通过G1uc转录物的腺苷脱氨的荧光素酶校正。值表示平均值 $\pm$ S.E.M(n=3)。(图92C) 使用靶向指导RNA,利用RESCUEv0-v16 和RESCUEv16S,G1uc转录物的腺苷向肌苷的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。(图92D) 使用非靶向指导RNA,利用RESCUEv0-v16 和RESCUEv16S,G1uc转录物的腺苷向肌苷的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0162] 图93A-93C:使用不同量的RESCUEv0-16和RESCUEv16S,在 $\beta$ -连环蛋白靶位点上的胞苷脱氨活性和脱靶活性。(图93A) CTNNB1 T41I 的编辑位点的示意图,其中靶向的C以红色突出显示,以及附近脱靶腺苷碱基用灰色突出显示。(图93B) 使用不同量的RESCUEv0-16和 RESCUEv16S,胞苷向尿苷(T41A)的编辑百分比。值表示三次重复的平均值。(图93C) 使用不同量的RESCUEv0-16和RESCUEv16S,在脱靶腺苷处的腺苷向鸟苷的编辑百分比。值表示三次重复的平均值。

[0163] 图94A-94E:在内源性靶点上RESCUEv16和RESCUEv16S的中靶和脱靶编辑。(图94A) 使用RESCUEv16和RESCUEv16S,具有不同碱基基序的内源性靶位点的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。(图94B) 使用RESCUEv16和RESCUEv16S靶向,在NRAS I21I中,在邻近腺苷碱基处的编辑百分比。(图94C) 使用RESCUEv16和 RESCUEv16S靶向,在NF2 T21M中,在邻近腺苷碱基处的编辑百分比。(图94D) 使用RESCUEv16和RESCUEv16S靶向,在RAF1 P30S中,在邻近腺苷碱基处的编辑百分比。(图94E) 使用RESCUEv16和 RESCUEv16S靶向,在CTNNB1 P44S中,在邻近腺苷碱基处的编辑百分比。

[0164] 图95A-95B:RESCUE启用的氨基酸变化的总结。(图97A) 通过 RESCUE使用胞苷脱氨可能的氨基酸转化。(图97B) 显示通过RESCUE 的所有潜在氨基酸改变的可能性的密码子表。

[0165] 图96:RESCUE v16S能够有效编辑内源性基因。

[0166] 图97:RESCUE v16S保持一些A至I活性。

[0167] 图98:将RESCUE v16用于靶向STAT,以减少INF  $\gamma$  /IL6诱导。

[0168] 图99A-99B:RESCUE靶向诱导细胞生长。

[0169] 图100:显示示例性转录追踪方法的示意图。

[0170] 图101显示了根据本文一些实施方式的可编程的胞苷向尿苷转化的示例性系统和方法。

[0171] 图102显示了根据本文一些实施方式的使用碱基编辑器校正突变和/或靶向翻译后信号传导或催化的示例性方法。

[0172] 图103A-103E:在报告基因和内源性转录物中用于胞苷脱氨的ADAR2脱氨酶结构域的进化。图103A。高斯荧光素酶报告基因转录物 (SEQ ID NO:712-714) 的催化残基突变体 (C82R) 的RNA靶向的示意图。图103B。描述在胞苷上RESCUEr0-r16的编辑水平百分比的热图,胞苷的侧翼为在Gluc转录物上的各种碱基。有利性更高的编辑基序显示在上部,而有利性更低的基序 (5' C) 显示在下部。图103C。在具有U型翻转或C翻转指导的Gluc转录物上在所有可能的16个胞苷侧翼碱基基序上RESCUE的编辑活性。图103D。RESCUE、ADAR2dd (不具有Cas13)、全长ADAR2 (不具有Cas13) 或无蛋白之间的活性比较。图103E。在覆盖多个基序的一组内源性基因上RESCUE的编辑有效性。图125中显示了整个指导嵌板中所示的针对每个位点的最佳指导。

[0173] 图104A-104F:在细胞生长和信号传导上RESCUE的表型结局。图104A。b-连环蛋白结构域和RESCUE靶向引导 (SEQ ID NO:715-717) 的示意图。图104B。通过RESCUE编辑的b-连环蛋白活化和细胞生长的示意图。图104C。在CTNNB1转录物中在相关位置RESCUE的编辑百分比。图104D。如通过b-连环蛋白驱动 (TCF/LEF) 荧光素酶表达所测量的,通过RNA编辑的Wnt/b-连环蛋白的活化。图104E。在HEK293FT 细胞中RESCUE CTNNB1靶向和非靶向指导的代表性显微图像。图104F。由于在HEK293FT细胞中通过RNA编辑的CTNNB1信号传导活化的胞内生长的定量。

[0174] 图105A-105D:通过指导体工程化进行的RESCUE和REPAIR多路复用和特异性增强。图105A。使用前-crRNA指导阵列进行的多路复用C到U和A到I编辑的示意图。图105B。在CTNNB1转录物上同时进行的C到U和A到I编辑。图105C。具有鸟苷碱基翻转以防止在相邻腺苷位点处的脱靶活性的合理工程化示意图 (SEQ ID NO:718-719)。图105D。使用合理引入不利的碱基翻转,针对高斯荧光素酶 (左图) 和KRAS (右图),在中靶C和脱靶A位点处的编辑百分比。

[0175] 图106A-106G:RESCUE的全转录组特异性。图106A。与REPAIR 相比,针对RESCUE的中靶C到U编辑以及C到U和A到I全转录组脱靶的总结。图106B。RESCUE A到I (左图) 和C到U (右图) 脱靶的Manhattan曲线。中靶C到U编辑用橙色突出显示。图106C。ADAR2dd 残基以及双链RNA底物之间的相互作用的示意图,用红色突出显示针对改善特异性的诱变筛选中使用的残基 (SEQ ID NO:720-721)。图106D。针对RESCUE和95RESCUE突变体显示的,针对具有靶向指导 (y轴) 的C到U活性以及具有非靶向指导 (x轴) 的A到I活性的荧光素酶值。以蓝色突出显示的突变体具有更高的特异性,并保持C到U的活性。RESCUE以红色突出显示。产生REPAIRv2的T375G突变以橙色显示。图106E。RESCUE、REPAIR和最高特异性突变体的中靶C到U编辑,以及C到U和A到I全转录组脱靶的总结。图106F。RESCUE-S (+S375A) A到I (左图) 和C到U (右图) 脱靶的Manhattan曲线。中靶C到U 编辑以橙色突出显示。图106G。针对RESCUE (上图) 和RESCUE-S (下图) 在中靶Gluc编辑位点 (蓝色三角形) 周围的代表性RNA测序读数。A到I编辑以红色突出显示;C到U (T) 编辑以蓝色突出显示;测序错误以黄色突出显示 (SEQ ID NO:722-767)。

[0176] 图107A-107B:靶向RNA胞苷到尿苷编辑使得能够实现新的碱基转换。图107A。通过RESCUE使用胞苷脱氨基化的可能氨基酸转换,具有相应的翻译后修饰和生物学活性。图107B。定向进化方法的示意图,包括合理诱变、酵母筛选,以及活性的哺乳动物细胞验证。合理诱变始于靶向已知与RNA底物接触的残基,如在来源于ADAR2dd (23) 的晶体结构的上部示意图所示的。使用饱和诱变靶向的残基以红色突出显示。对于定向进化,使用HIS3生长报告基因以使得能够在酵母中使用C到U 编辑和HIS3基因恢复来进行ADAR2dd突变体的阳性选择。针对C到U 编辑活性在哺乳动物细胞中评价每轮酵母进化的排在前面的突变体,然后将排在前面的突变体用于下一轮的酵母进化。

[0177] 图108。比较使用靶向和非靶向指导的在酵母中的 RanCas13b-REPAIR和PspCas13b-REPAIR腺苷脱氨基化活性。在EGPF 中Y66H突变的A到I校正恢复GFP荧光,并通过流式细胞术进行测量。由于使用来自鸭疫里默氏杆菌 (*Riemerella anatipestifer*) 的催化失活 Cas13b直系同源物 (dRanCas13b) 的REPAIR比使用来自普雷沃菌属种 (*Prevotella* sp.) P5-125的催化失活Cas13b直系同源物 (dPspCas13b) 的REPAIR更有效,我们开始使用dRanCas13b-ADAR2dd融合蛋白开发RESCUE。

[0178] 图109A-109B:用于产生胞嘧啶脱氨基化荧光素酶报告基因的失活 Gluc突变的筛选。图109A。此前显示对荧光素酶活性具有一定影响的一组各种各样的Gluc突变体的荧光素酶活性(33)。值表示平均值 $\pm$ S.E.M (n=3)。图109B。一组亮氨酸到脯氨酸Gluc突变体的荧光素酶活性。关注亮氨酸到脯氨酸突变体报告基因是因为其产生了针对胞苷脱氨基化的CCN基序位点(中央C是脱氨基化的)。这允许测定所有四种CCN 基序对RESCUE脱氨基化活性的影响。值表示平均值 $\pm$ S.E.M(n=3); WT,野生型Gluc序列。

[0179] 图110。在Gluc中UCG、CCG、ACG、GCG、CCA和CCU位点上 RESCUEr0-r16的胞苷脱氨基化活性。值表示平均值 $\pm$ S.E.M(n=3)。

[0180] 图111A-111C:不同量的RESCUEr0-r16的胞苷脱氨基化活性。图 111A。RESCUEr0-r16活性的剂量应答,通过在Gluc转录物中UCG位点上的荧光素酶活性的恢复来测量。值表示三次重复的平均值。图111B。 RESCUEr0-r16活性的剂量应答,通过在Gluc转录物中UCG位点上的C 到U编辑来测量。值表示三次重复的平均值。图111C。RESCUEr0-r16 活性的剂量应答,通过在CTNNB1转录物中T41I位点上的荧光素酶活性的恢复测量。值表示三次重复的平均值。

[0181] 图112。以不同指导和RESCUE质粒的量,通过RESCUEr6-r9在Gluc 转录物中UCG位点进行的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0182] 图113A-113E:针对定向进化的各种酵母报告基因的编辑速率。图 113A。使用靶向和非靶向指导,通过RESCUEr3、r7和r16进行的GFP 突变Y66H的荧光校正百分比。在10,000个细胞上通过进行流式细胞术测量荧光。T,靶向指导;NT,非靶向指导。图113B。使用靶向和非靶向指导,通过RESCUEr3、r7和r16进行的GFP突变Y66H的编辑校正百分比。T,靶向指导;NT,非靶向指导。图113C。使用靶向和非靶向指导,通过RESCUEr7和r16进行的HIS3突变P196L的编辑校正百分比。T,靶向指导;NT,非靶向指导。图113D。使用靶向和非靶向指导,通过RESCUEr7和r16进行的HIS3突变S129P的编辑校正百分比。T,靶向指导;NT,非靶向指导。图113E。在RESCUE诱导后的不同小时,使用各种错配距离的靶向指导和非靶向指导,通过RESCUEr3、r7和r16 进行的HIS3突变S22P的编辑校正百分比。NT,非靶向指导。

[0183] 图114A-114C:使用具有U、C、G或A错配的指导,使用RESCUEr16 和r8,使用所有16种可能的5'和3'碱基组合的Gluc位点的编辑百分比。图114A.使用具有U或C错配的指导,使用RESCUEr8,使用所有16种可能的5ÅL和3ÅL碱基组合的Gluc位点的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。图114B.使用具有G或A错配的指导,使用RESCUEr8,使用所有16种可能的5ÅL和3ÅL碱基组合的Gluc位点的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。图114C.使用具有G或A错配的指导,使用RESCUEr16,使用所有16种可能的5ÅL和3ÅL碱基组合的Gluc位点的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0184] 图115.使用具有不同U错配位置的30bp和50bp指导,在Gluc转录物的UCG位点上RESCUE的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0185] 图116.使用具有不同U错配位置的指导RNA,在Gluc转录物中的UCG位点上RESCUEr1和RESCUEr3-r8的编辑百分比。将候选轮与RanCas13b和PspCas13b两者进行比较。值表示平均值 $\pm$ S.E.M(n=3)。20/22表示针对RanCas13b的错配距离为20和针对PspCas13b的错配距离为22。由于REPAIR使用ADAR2dd与dPspCas13b的融合体(7),我们将我们的RESCUE候选轮与PspCas13b和RanCas13b的融合体进行了比较,并发现其具有同等活性。

[0186] 图117A-117B:在ADAR2脱氨酶结构域的晶体结构上的RESCUE突变的视图。图117A.RESCUE突变体与紫色建模的翻出的胞苷一起显示在ADAR2晶体结构(蓝色)中。图117B.在催化脱氨基化位点处的突变体的放大晶体结构视图,其中带有翻出碱基的RNA也显示为紫色。

[0187] 图118A-118D:RESCUEr0-r16和RESCUEr16-S的腺苷脱氨酶活性。采用REPAIR,腺苷脱氨酶的效率取决于相对于靶腺苷的位置的指导设计选择和碱基翻转选择(7),因为ADAR2dd倾向于在错配气泡中脱氨基。在指导:靶标dsRNA双链体内的靶碱基的位置是特别重要的,因为Cas13指导可以被置于任何位置而没有任何序列限制,并且ADAR2dd的最佳活性窗很小(7)。对于RESCUE,我们检测了靶胞嘧啶上所有可能的指导碱基翻转,并发现胞苷脱氨酶的最佳碱基翻转是C或U,UCG基序的最佳编辑是使用30nt指导RNA,靶向碱基翻转位置26个碱基对是来自靶标的5ÅL末端。图118A.使用靶向指导RNA,利用RESCUEr0-r16和RESCUEr16-S,通过Gluc转录物的腺苷脱氨酶进行的荧光素酶校正。值表示平均值 $\pm$ S.E.M(n=3)。图118B.使用非靶向指导RNA,利用RESCUEr0-v16和RESCUEr16-S,通过Gluc转录物的腺苷脱氨酶进行的荧光素酶校正。值表示平均值 $\pm$ S.E.M(n=3)。图118C.使用靶向指导RNA,通过RESCUEr0-r16和RESCUEr16-S进行的Gluc转录物的腺苷到肌苷的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。图118D.使用非靶向指导RNA,通过RESCUEr0-r16和RESCUEr16-S进行的Gluc转录物的腺苷到肌苷的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0188] 图119A-119D:在REPAIR上添加的个体RESCUE突变(RESCUEr0)或从RESCUEr16除去的个体突变的评价。图119A.如通过荧光素酶活性恢复所测量的,在靶向荧光素酶转录物上的位点的REPAIR(RESCUEr0)上添加的个体RESCUE突变的C到U脱氨酶活性的评价。值表示平均值 $\pm$ S.E.M(n=3);WT,RESCUEr0序列。图119B.如通过编辑百分比所测量的,在靶向荧光素酶转录物上的位点的REPAIR(RESCUEr0)上添加的个体RESCUE突变的C到U脱氨酶活

性的评价。值表示平均值 $\pm$ S.E.M ( $n=3$ ) ;WT, RESCUEr0序列。图119C。如通过荧光素酶活性恢复所测量的, 靶向荧光素酶转录物的位点的、除去个体突变的RESCUEr16构建体的C到U脱氨酶活性的评价。值表示平均值 $\pm$ S.E.M ( $n=3$ ) ;WT, RESCUEr16序列。图119D。如通过编辑百分比所测量的, 靶向荧光素酶转录物上的位点的、除去个体突变的 RESCUEr16构建体的C到U脱氨酶活性的评价。值表示平均值 $\pm$ S.E.M ( $n=3$ ) ;WT, RESCUEr16序列。

[0189] 图120A-120D: 使用重组蛋白, 含有RESCUEr0、r2、r8、13和r16 突变的ADAR2脱氨酶结构域的生化脱氨基化活性。图120A。含有各种候选突变的ADAR2脱氨酶结构域蛋白的腺苷脱氨基化活性, 22bp双链 RNA底物含有与胞苷错配的中心腺嘌呤。在具有和不具有脱氨酶结构域的情况下, 将反应孵育不同时间点。值表示平均值 $\pm$ S.E.M ( $n=3$ , 一些误差线被符号遮挡)。图120B。含有各种候选突变的ADAR2脱氨酶结构域蛋白的胞苷脱氨基化活性, 22bp双链RNA底物含有与尿苷错配的中心胞苷。在具有和不具有脱氨酶结构域的情况下, 将反应孵育不同时间点。值表示平均值 $\pm$ S.E.M ( $n=3$ , 一些误差线被符号遮挡)。图 120C。在RNA和DNA底物上的RESCUE r0和r16胞苷脱氨酶活性, 包括与互补DNA退火的RNA (RNA:DNA) 中的胞苷, 与互补RNA退火的DNA (DNA:RNA) 中的脱氧胞苷, 在双链DNA (dsDNA) 中的脱氧胞苷和在ssDNA中的脱氧胞苷。所有双链模板均含有与胸苷错配的胞苷。值表示平均值 $\pm$ S.E.M ( $n=3$ )。图120D。在RNA和DNA底物上的 RESCUE r0和r16腺苷脱氨酶活性, 包括与互补DNA退火的RNA (RNA:DNA) 中的腺苷, 与互补RNA退火的DNA (DNA:RNA) 中的脱氧腺苷, 在双链DNA (dsDNA) 中的脱氧腺苷和在ssDNA中的脱氧腺苷。所有双链模板均含有与腺苷错配的胞苷。值表示平均值 $\pm$ S.E.M ( $n=3$ )。

[0190] 图121A-121D: RESCUEr16、全长ADAR2 (具有RESCUEr16突变)、ADAR2脱氨酶结构域 (具有RESCUEr16突变) 和无任何蛋白的胞苷脱氨酶活性的比较。图121A。使用靶向指导以及 RESCUEr16、全长ADAR2 (具有RESCUEr16突变)、ADAR2脱氨酶结构域 (具有RESCUEr16 突变) 和无任何蛋白, 通过Cluc活性恢复测量的腺苷脱氨酶活性。值表示平均值 $\pm$ S.E.M ( $n=3$ )。图121B。使用靶向指导以及RESCUEr16、全长ADAR2 (具有RESCUEr16突变)、ADAR2脱氨酶结构域 (具有 RESCUEr16突变) 和无任何蛋白, 通过Gluc活性恢复测量的胞苷脱氨酶活性。值表示平均值 $\pm$ S.E.M ( $n=3$ )。图121C。使用靶向指导以及 RESCUEr16、全长ADAR2 (具有 RESCUEr16突变)、ADAR2脱氨酶结构域 (具有RESCUEr16突变) 和无蛋白, 使用不同5ÅL 碱基, 在Gluc 转录物中的位点的编辑百分比。值表示平均值 $\pm$ S.E.M ( $n=3$ )。图121D。使用非靶向指导以及RESCUEr16、全长ADAR2 (具有RESCUEr16突变)、ADAR2脱氨酶结构域 (具有 RESCUEr16突变) 和无蛋白, 使用不同5ÅL 碱基, 在Gluc转录物中的位点的编辑百分比。值表示平均值 $\pm$ S.E.M ( $n=3$ )。

[0191] 图122A-122C: RESCUEr16、全长ADAR2 (具有RESCUEr16突变)、ADAR2脱氨酶结构域 (具有RESCUEr16突变) 和无任何蛋白的胞苷脱氨酶活性的比较。图122A。使用RESCUEr16和包含不同错配位点的指导RNA, 在Gluc转录物中的UCG位点的编辑。值表示平均值 $\pm$ S.E.M ( $n=3$ )。图122B。使用全长ADAR2 (具有RESCUEr16突变) 和包含不同错配位点的指导RNA, 在Gluc转录物中的UCG位点的编辑。值表示平均值 $\pm$ S.E.M ( $n=3$ )。图122C。使用ADAR2脱氨酶结构域 (具有RESCUEr16突变) 和包含不同错配位点的指导RNA, 在Gluc转录物中的UCG位点的编辑。值表示平均值 $\pm$ S.E.M ( $n=3$ )。

[0192] 图123A-123C: 使用没有长度为30或50nt的直接重复和各种错配的指导, 在Gluc转

录物上的RESCUEr16的胞苷脱氨基化活性。图123A。使用没有直接重复和各种错配的30nt指导,在Gluc转录物上的 RESCUEr16的胞苷脱氨基化活性。值表示平均值 $\pm$ S.E.M(n=3)。图123B。使用没有直接重复和各种错配的50nt指导,在Gluc转录物上的 RESCUEr16的胞苷脱氨基化活性。值表示平均值 $\pm$ S.E.M(n=3)。图123C。使用具有直接重复和各种错配的30nt指导,在Gluc转录物上的 RESCUEr16的胞苷脱氨基化活性。值表示平均值 $\pm$ S.E.M(n=3)。

[0193] 图124A-124F:具有掺入其中的RESCUE突变的替代RNA编辑技术的胞苷脱氨基化活性。图124A。使用具有不同碱基翻转和各种错配的30 nt指导,在Gluc转录物上使用RESCUE突变的MS2募集的ADAR脱氨酶结构域(24)的胞苷脱氨基化活性。通过荧光素酶活性的恢复测量活性。值表示平均值 $\pm$ S.E.M(n=3);NT,非靶向指导。图124B。使用具有不同碱基翻转和各种错配的30nt指导,在Gluc转录物上使用 RESCUE突变的MS2募集的ADAR脱氨酶结构域(24)的Gluc编辑百分比。值表示平均值 $\pm$ S.E.M(n=3);NT,非靶向指导。图124C。使用具有不同碱基翻转和各种错配的30nt指导,在Gluc转录物上使用含有RESCUE突变的脱氨酶结构域的关联ADAR指导RNA技术(24)的胞苷脱氨基化活性。通过荧光素酶活性的恢复测量活性。值表示平均值 $\pm$ S.E.M(n=3);NT,非靶向指导。图124D。使用具有不同碱基翻转和各种错配的30nt指导,在Gluc转录物上使用含有RESCUE突变的脱氨酶结构域的关联ADAR指导RNA技术(24)的Gluc编辑百分比。值表示平均值 $\pm$ S.E.M(n=3);NT,非靶向指导。图124E。使用具有不同碱基翻转和各种错配的30nt指导,在Gluc转录物上使用RESCUE 突变的指导RNA募集的ADAR脱氨酶结构域(11)的胞苷脱氨基化活性。通过荧光素酶活性的恢复测量活性。值表示平均值 $\pm$ S.E.M(n=3); NT,非靶向指导。图124F。使用具有不同碱基翻转和各种错配的30nt 指导,在Gluc转录物上使用RESCUE突变的指导RNA募集的ADAR脱氨酶结构域(11)的Gluc编辑百分比。值表示平均值 $\pm$ S.E.M(n=3); NT,非靶向指导。

[0194] 图125A-125C:错配位置平铺,以发现在内源性靶位点上针对 RESCUE的最佳编辑指导设计。图125A。使用RESCUE和在位置7、9、11和13处具有错配和U型碱基翻转的指导,具有不同碱基基序的内源性靶位点的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。图125B。使用RESCUE和在位置7、9、11和13处具有错配和C型碱基翻转的指导,具有不同碱基基序的内源性靶位点的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。图125C。使用RESCUE和在位置3、5、7、9和11处具有错配和C型和U型碱基翻转的指导,具有不同碱基基序的内源性靶位点的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0195] 图126A-126B:如通过在各种内源性位点和不同量处的编辑百分比所测量的,RESCUEr0-r16的胞苷脱氨基化活性。图126A。热图描述了在一组三个内源性基因上RESCUEr0-r16的编辑效率。值表示三次重复的平均值。图126B。如通过在内源性KRAS位点处的编辑百分比所测量的,不同量的RESCUEr0-r16的胞苷脱氨基化活性。值表示三次重复的平均值。

[0196] 图127A-127B:在合成报告基因上各种疾病相关突变的编辑百分比。图127A。在一组相关T>C疾病突变的合成版本上RESCUE的编辑效率,采用每个靶位点的最佳可能错配指导。编辑率在1%和42%之间变化,并且按照编辑效率显示条件。在表31中列出了针对合成位点的所有编辑率。值表示平均值 $\pm$ S.E.M(n=3)。图127B。使用RESCUE和具有各种错配位置的指导的疾病相关突变的编辑。值表示平均值 $\pm$ S.E.M(n=3)。

[0197] 图128。使用具有各种C和U错配位置的指导的RESCUE在ApoE4 胞嘧啶处的编辑百

分比。ApoE4变体(rs429358和rs7412)显著增加阿尔茨海默氏症的风险,并且被通过RESCUE进行编辑,在两个位点上的比率高达5%和12%。在表31中列出了针对合成位点的所有编辑率。值表示平均值 $\pm$ S.E.M(n=3)。

[0198] 图129A-129F:通过RESCUE进行的RNA编辑和STAT1/STAT3的信号调节。STAT3和STAT1是转录因子,其在通过JAK/STAT途径进行的信号转导中起重要作用,通常通过细胞因子和生长因子的磷酸化来激活。为了证明通过RNA编辑进行的信号传导调节,我们在48小时过程中通过用RESCUE编辑STAT3上的Y705和S727以及STAT1上的Y701和S727磷酸化位点来改变STAT通路的激活。图129A。靶向STAT3的磷酸化残基以改变相关信号传导通路的STAT3结构域和RESCUE指导的示意图(SEQ ID NO:768-770)。图129B。通过RESCUE在STAT3中在相关磷酸化残基处进行的编辑百分比。在HEK293FT细胞中,我们观察到S727 STAT3位点的6%编辑,以及Y701和S727 STAT1位点分别的11%和7%编辑。图129C。如通过使用具有不同碱基翻转的指导进行的STAT3驱动荧光素酶表达所测量的,通过RNA编辑进行的STAT3信号传导的抑制。这些编辑导致STAT3和STAT1活性的13%抑制。图129D。使用具有不同碱基翻转的指导,通过RESCUE在STAT1中S727F磷酸化残基处进行的编辑百分比。图129E。使用具有不同碱基翻转的指导,通过RESCUE在STAT1中Y701C磷酸化残基处进行的编辑百分比。图129F。如通过STAT驱动荧光素酶表达所测量的,使用RESCUE通过RNA编辑进行的STAT1信号传导的抑制。

[0199] 图130A-130B:在HUVEC细胞中b-连环蛋白磷酸化和细胞生长的调节。图130A。通过在HUVEC细胞中的RNA编辑进行的CTNNB1信号传导的活化造成的细胞生长的定量。RESCUE刺激HUVEC生长达到与过表达b-连环蛋白磷酸化无效突变体的细胞中观察到的水平相当的水平。NT,非靶向指导。图130B。在HUVEC细胞中RESCUE CTNNB1 靶向和非靶向指导的代表性显微图像。

[0200] 图131。在带有不同dRanCas13b的C末端截短的靶点周围,使用不同5'和3'侧翼碱基在转录物上的RESCUE C到U和A到I活性。

[0201] 图132A-132C:在指导双链体窗中的候选轮的特异性。图132A。高斯荧光素酶突变体C82R的编辑位点的示意图,其中被靶向的C用红色突出显示,以及将附近的腺嘌呤碱基编号并用灰色突出显示(SEQ ID NO:771)。图132B。通过RESCUEr0、RESCUEr8和RESCUEr16进行靶向,在高斯荧光素酶突变体C82R中的附近腺嘌呤碱基的编辑百分比。图132C。通过不同量的RESCUEr0-r16,在腺嘌呤20处,腺嘌呤至鸟苷的编辑百分比。值表示三次重复的平均值。

[0202] 图133A-133D:在通过RESCUE r0、r8和r16进行的单路和多路靶向中在靶点胞苷附近的脱靶。图133A。KRAS转录物的编辑位点的示意图,其中被靶向的C用红色突出显示,以及将附近的腺嘌呤碱基编号并用灰色突出显示(SEQ ID NO:772)。图133B。通过RESCUEr0、RESCUEr8和RESCUEr16进行靶向,在KRAS转录物中的附近腺嘌呤碱基的编辑百分比。图133C。CTNNB1转录物的多路复用编辑位点的示意图,其中两个被靶向的C位点用红色突出显示,以及将附近的腺嘌呤碱基编号并用灰色突出显示(SEQ ID NO:773)。FIG 133D。通过RESCUEr0、RESCUEr8和RESCUEr16进行多路复用靶向,在CTNNB1转录物中的附近腺嘌呤碱基的编辑百分比。

[0203] 图134A-134F:RESCUE和RESCUE-S全转录组脱靶的表征。图134A。具有靶向指导的RESCUE对荧光素酶转录物上的位点的全转录组脱靶编辑的预测作用。图134B。具有靶向指

导的RESCUE对荧光素酶转录物上的位点的全转录组脱靶编辑的预测致癌作用。图134C。具有靶向指导的RESCUE对荧光素酶转录物上的位点以每个转录物的脱靶编辑数量来可视化的全转录组脱靶。图134D。具有靶向指导的RESCUE-S对荧光素酶转录物上的位点的全转录组脱靶编辑的预测作用。图134E。具有靶向指导的RESCUE-S对荧光素酶转录物上的位点的全转录组脱靶编辑的预测致癌作用。图134F。具有靶向指导的RESCUE-S对荧光素酶转录物上的位点以每个转录物的脱靶编辑数量来可视化的全转录组脱靶。

[0204] 图135A-135C:全转录组脱靶的5'和3'侧翼碱基的表征。图135A。具有靶向指导的RESCUE针对荧光素酶转录物上的位点,使用所有16种可能的5'AL和3'AL侧翼碱基的每一种的脱靶数量。图135B。具有靶向指导的RESCUE-S针对荧光素酶转录物上的位点,使用所有16种可能的5'AL和3'AL侧翼碱基的每一种的脱靶数量。图135C。在具有靶向荧光素酶转录物的RESCUE构建体的条件下,显著差异表达的转录物的数量。

[0205] 图136A-136B:使用重组蛋白,含有RESCUEr0、RESCUEr16和 RESCUEr16-S突变的ADAR脱氨酶结构域的生化脱氨化活性。图136A。含有各种候选突变的ADAR2脱氨酶结构域蛋白的腺苷脱氨化活性,22bp双链RNA底物含有与胞嘧啶错配的中心腺嘌呤。在具有和不具有脱氨酶结构域的情况下,将反应孵育不同时间点。值表示平均值 $\pm$ S.E.M(n=3,一些误差线被符号遮挡)。图136B。含有各种候选突变的ADAR2脱氨酶结构域蛋白的胞苷脱氨化活性,22bp双链RNA底物含有与尿苷错配的中心胞嘧啶。在具有和不具有脱氨酶结构域的情况下,将反应孵育不同时间点。值表示平均值 $\pm$ S.E.M(n=3,一些误差线被符号遮挡)。

[0206] 图137A-137D:RESCUE和RESCUE-S的腺苷脱氨酶活性。图137A。使用靶向指导RNA,利用RESCUE和RESCUE-S,通过Gluc转录物的腺苷脱氨化进行的荧光素酶校正。值表示平均值 $\pm$ S.E.M(n=3)。图137B。使用非靶向指导RNA,利用RESCUE和RESCUE-S,通过Gluc转录物的腺苷脱氨化进行的荧光素酶校正。值表示平均值 $\pm$ S.E.M(n=3)。图137C。使用靶向指导RNA,利用RESCUE和RESCUE-S,Gluc转录物的腺苷到肌苷的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。图137D。使用非靶向指导RNA,利用RESCUE和RESCUE-S,Gluc转录物的腺苷到肌苷的编辑百分比。值表示平均值 $\pm$ S.E.M(n=3)。

[0207] 图138A-138C:使用各种量的RESCUEr0-r16和RESCUEr16-S在b-连环蛋白靶位点上的胞苷脱氨化活性和脱靶活性。图138A。CTNNB1 T41I的编辑位点的示意图,其中被靶向的C以红色突出显示,以及附近脱靶腺嘌呤碱基以灰色突出显示(SEQ ID NO:774)。图138B。通过各种量的RESCUEr0-r16和RESCUEr16-S进行的胞嘧啶到尿苷(T41A)的编辑百分比。值表示三次重复的平均值。图138C。通过各种量的 RESCUEr0-r16和RESCUEr16-S在脱靶腺嘌呤处进行的腺嘌呤到鸟苷的编辑百分比。值表示三次重复的平均值。

[0208] 图139A-139C:通过RESCUE和RESCUE-S进行的STAT1和STAT3的编辑。图139A。通过C到U和A到I编辑在STAT3处编辑的位点的示意图(SEQ ID NO:775-778)。图139B。通过RESCUE和RESCUE-S在STAT1和STAT3中在酪氨酸残基处进行的A到I编辑百分比。值表示平均值 $\pm$ S.E.M(n=3);NT,非靶向指导。图139C。通过RESCUE和RESCUE-S在STAT1和STAT3中在丝氨酸残基处进行的C到U编辑百分比。值表示平均值 $\pm$ S.E.M(n=3);NT,非靶向指导。

[0209] 图140A-140E:在内源性靶点上的RESCUE和RESCUE-S的中靶和脱靶编辑。图140A。使用RESCUE和RESCUE-S,具有不同碱基基序的内源性靶位点的编辑百分比。值表示平均值

+/-S.E.M (n=3)。图140B。通过RESCUE和RESCUE-S进行靶向,在NRAS I21I中的邻近腺嘌呤碱基的编辑百分比。图140C。通过RESCUE和RESCUE-S进行靶向,在NF2 T21M中的邻近腺嘌呤碱基的编辑百分比。图140D。通过RESCUE和RESCUE-S进行靶向,在RAF1 P30S中的邻近腺嘌呤碱基的编辑百分比。图140E。通过RESCUE和RESCUE-S进行靶向,在CTNNB1 P44S中的邻近腺嘌呤碱基的编辑百分比。

[0210] 图141:RESCUE使得能够进行的氨基酸变化的总结。密码子表显示了通过RESCUE可能发生的所有潜在氨基酸变化。

[0211] 本文中的附图仅用于说明目的,而不一定按比例绘制。

## 具体实施方式

### [0212] 定义

[0213] 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。在分子生物学中常用的术语和技术的定义可以在以下文献中找到: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 等编辑); the series Methods in Enzymology(Academic Press,Inc.):PCR 2: A Practical Approach(1995) (M.J.MacPherson,B.D.Hames和G.R.Taylor 编辑):Antibodies,A Laboratory Manual(1988) (Harlow和Lane编辑): Antibodies A Laboratory Manual第2版2013 (E.A.Greenfield编辑); Animal Cell Culture(1987) (R.I.Freshney编辑); Benjamin Lewin,Genes IX, Jones and Bartlet出版,2008 (ISBN 0763752223);Kendrew等(编辑), The Encyclopedia of Molecular Biology,Blackwell Science Ltd.出版,1994 (ISBN 0632021829);Robert A.Meyers编辑,Molecular Biology and Biotechnology:a Comprehensive Desk Reference,VCH Publishers出版,Inc., 1995 (ISBN 9780471185710);Singleton等,Dictionary of Microbiology and Molecular Biology第2版,J.Wiley&Sons (New York,N.Y.1994),三月,Advanced Organic Chemistry Reactions,Mechanisms and Structure第4版, John Wiley&Sons (New York,N.Y.1992);和Marten H.Hofker和Jan van Deursen,Transgenic Mouse Methods and Protocols,第2版(2011)。

[0214] 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。

[0215] 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,并且该描述包括事件或情形发生的情况和不发生的情况。

[0216] 通过端点表述的数值范围包括在对应范围内的所有数值和分数,以及所表述的端点。

[0217] 如本文所用的术语“约”或“近似”当涉及诸如参数、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,诸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。应当理解,修饰语“约”或“近似”所涉及的值本身也是具体地且优选地公开的。

[0218] 如本文所用,“生物样品”可以含有全细胞和/或活细胞和/或细胞碎片。生物样品可以含有(或来源于)“体液”。本发明涵盖以下实施方式,其中体液选自羊水、房水、玻璃体液、胆汁、血清、母乳、脑脊液、耳屎(耳垢)、乳糜、食糜、内淋巴液、外淋巴液、渗出液、粪便、女性射液、胃酸、胃液、淋巴液、粘液(包括鼻引流和痰)、心包液、腹膜液、胸膜液、脓液、炎性分泌物、唾液、皮脂(皮油)、精液、痰涎、滑液、汗液、眼泪、尿液、阴道分泌物、呕吐物以及这些物质中一者或多者的混合物。生物样品包括细胞培养物、体液、来自体液的细胞培养物。体液可以例如通过穿刺或其他收集或采样程序从哺乳生物体获得。

[0219] 术语“受试者”、“个体”和“患者”在本文中可互换使用,是指脊椎动物,优选地是哺乳动物,更优选地是人类。哺乳动物包括但不限于鼠、猴、人类、农场动物、运动型动物和宠物。还包括在体内获得或在体外培养的生物实体的组织、细胞和他们的子代。

[0220] 每当在本文中提及Cas13时,将理解的是,除非另外明确指出,是指根据本发明的突变的或工程化Cas13。每当在本文中提及Cas13时,将理解的是,除非另外明确指出,是指根据本发明的突变的或工程化 Cas13a、Cas13b、Cas13c或Cas13d。每当在本文中提及Cas13时,将理解的是,除非另外明确指出,是指根据本发明的突变的或工程化Cas13b。

[0221] 在下文中描述各种实施方式。应当指出的是,具体实施方式不旨在作为详尽的描述或作为对本文所论述的更广泛方面的限制。结合特定实施方式描述的一个方面不必限于该实施方式,而是可以与任何其他一个或多个实施方式一起实践。在本说明书通篇提及“一个实施方式”、“实施方式”、“示例性实施方式”意指结合实施方式描述的特定特征、结构或特性包括于本发明的至少一个实施方式中。因此,在本说明书通篇各处出现短语“在一个实施方式中”、“在实施方式中”或“示例性实施方式”不一定全部指代同一实施方式,但也有可能如此。此外,如本领域技术人员从本公开将明显了解的,在一个或多个实施方式中,特定特征、结构或特性可以按任何适合的方式组合。此外,虽然本文所述的一些实施方式包括其他实施方式中所包括的一些特征而非其他特征,但不同实施方式的特征的组合有意处于本发明的范围内。举例来说,在所附权利要求中,任何所要求保护的实施方式可以按任何组合使用。

[0222] 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方式并入,如同每个单独公布、公布的专利文献或专利申请被确切地且单独地指明为以引用方式整体并入。

[0223] 综述

[0224] 在一个方面中,本文公开的实施方式涉及一种工程化CRISPR-Cas 蛋白,其包含一种或多种修饰的氨基酸。在某些实施方式中,与相应野生型CRISPR-Cas蛋白相比,工程化CRISPR-Cas蛋白增加或降低以下的一种或多种:PFS识别/特异性、gRNA结合、蛋白酶活性、多核苷酸结合能力、稳定性、特异性、靶标结合、脱靶结合和/或催化活性。在某些实施方式中,CRISPR-Cas蛋白包含一个或多个HEPN结构域,以及包含一种或多种修饰的氨基酸。修饰的氨基酸可以与指导RNA相互作用,所述指导RNA与CRISPR-Cas蛋白形成复合物,和/或在CRISPR-Cas蛋白的HEPN活性位点、结构域间接头结构域、盖结构域、螺旋结构域或桥螺旋结构域,或其组合中。在一些实例中,工程化CRISPR-Cas蛋白包含一个或多个HEPN结构域,以及还包含一种或多种修饰的氨基酸,其中所述氨基酸与指导RNA相互作用,所述指导RNA与CRISPR-Cas蛋白形成复合物;在工程化CRISPR-Cas蛋白的HEPN活性位点、结构域间接头结

构域、盖结构域、螺旋结构域1、螺旋结构域2或桥螺旋结构域,或其组合中。

[0225] 在另一个方面中,本文公开的实施方式提供了新鉴定的CRISPR-Cas 直系同源物子集,其尺寸小于此前发现的CRISPR-Cas直系同源物,包括对其的进一步修饰和用途。在特定实施方式中,CRISPR-Cas直系同源物小于约1000个氨基酸,并且可以任选地作为融合蛋白的一部分提供。

[0226] 本文还提供了工程化核苷酸脱氨酶。在某些实施方式中,工程化核苷酸脱氨酶是腺苷脱氨酶,可以将其工程化以包含胞苷脱氨酶活性。在实施方式中,可以将工程化核苷酸融合至Cas蛋白,包括本文公开的 CRISPR-Cas蛋白。

[0227] 在另一个方面中,本文公开的实施方式包括此类修饰的CRISPR-Cas 蛋白的系统和用途,包括但不限于诊断、碱基编辑疗法和检测方法。包含CRISPR Cas蛋白的融合蛋白(包括本文公开的那些)和核苷酸脱氨酶也可以用于碱基编辑。还提供了所公开的蛋白和系统的递送,包括递送至各种细胞以及通过各种颗粒、囊泡和载体。

[0228] 一般CRISPR-CAS系统

[0229] 在通常情况下,CRISPR-Cas或CRISPR系统统称为与CRISPR相关(“Cas”)基因表达或指导其活性的转录物和其他元件,包括编码Cas基因的序列、tracr (反式激活CRISPR) 序列(例如,tracrRNA或活性部分 tracrRNA)、tracr配对序列(在内源性CRISPR系统背景下包含“正向重复”和tracrRNA处理的部分正向重复)、指导序列(在内源性CRISPR 系统背景下也称为“间隔子”),或本文中使用的术语“RNA”(例如,指导Cas(如Cas9)的RNA,例如,CRISPR RNA和反式激活(tracr) RNA 或单一指导RNA(sgRNA)(嵌合RNA))或者来自CRISPR基因座的其他序列和转录物。在通常情况下,CRISPR系统的特征是促进靶序列位点上CRISPR复合物形成的元件(在内源性CRISPR系统背景下也称为原间隔子)。当CRISPR蛋白是2类VI型效应物时,不需要tracrRNA,在本发明的工程化系统中,正向重复可以包含天然存在的序列或非天然存在的序列。本发明的正向重复不限于天然存在的长度和序列。正向重复的长度可以是36nt,但是更长或更短的正向重复可以变化。例如,正向重复可以是30nt或更长,如30-100nt或更长。例如,正向重复的长度可以是30nt、40nt、50nt、60nt、70nt、70nt、80nt、90nt、100nt 或更长。在一些实施方式中,本发明的正向重复可以包括插入在天然存在的正向重复序列的5' 和3' 末端之间的合成核苷酸序列。在某些实施方式中,插入的序列可以是自我互补的,例如,20%、30%、40%、50%、60%、70%、80%、90%或100%自我互补。此外,本发明的正向重复可以包括核苷酸插入,如适体或结合至衔接子蛋白(用于与功能性结构域结合)的序列。在某些实施方式中,包含此类插入的正向重复的一端大约是短DR的前一半,而一端大约是短DR的后一半。

[0230] CRISPR-Cas蛋白(在本文中与“Cas蛋白”、“Cas效应物”互换使用) 可以包括Cas9、Cas 12(例如,Cas12a、Cas12b、Cas12c、Cas12d等)、Cas13(例如,Cas13a、Cas13b(如Cas13b-t1、Cas13b-t2、Cas13b-t3)、Cas13c、Cas13d等)、Cas14、CasX和CasY。在一些实施方式中, CRISPR-Cas蛋白可以是VI型CRISPR-Cas蛋白。例如,VI型CRISPR-Cas 蛋白可以是Cas13蛋白。Cas13蛋白可以是Cas13a、Cas13b、Cas13c或 Cas13d。在一些实例中,CRISPR-Cas蛋白是Cas13a。在一些实例中, CRISPR-Cas蛋白是Cas13b。在一些实例中,CRISPR-Cas蛋白是Cas13c。在一些实例中,CRISPR-Cas蛋白是Cas13d。

[0231] 在一些实施方式中,工程化CRISPR-Cas蛋白包含一个或多个HEPN 结构域并且长

度小于1000个氨基酸。例如,蛋白的尺寸可以小于950、小于900、小于850、小于800或小于750个氨基酸。

[0232] 在某些示例性实施方式中,RISPR-Cas蛋白包含至少一个HEPN结构域,包括但不限于本文所述的HEPN结构域,本领域公知的HEPN结构域以及通过与共有序列基序相比而被识别为HEPN结构域的结构域。本文中提供若干此类结构域。在一个非限制性实例中,共有序列可以来源于本文所提供的C2c2或Cas13b直系同源物的序列。在某些示例性实施方式中,效应蛋白包含单个HEPN结构域。在某些其他示例性实施方式中,效应蛋白包含两个HEPN结构域。

[0233] 在一个实例实施方式中,效应蛋白包含一个或多个包含RxxxxH基序序列的HEPN结构域。RxxxxH基序序列可以是但不限于来自本文所述的HEPN结构域或本领域中已知的HEPN结构域。RxxxxH基序序列还包括通过组合两个或更多个HEPN结构域的部分而建立的基序序列。如所指出的,共有序列可以来源于以下文献中公开的直系同源物的序列:题为“新颖的VI型CRISPR直系同源物和系统(Novel CRISPR Enzymes and Systems)”的美国临时专利申请62/432,240,2017年3月15日提交的题为“新颖的VI型CRISPR直系同源物和系统(Novel Type VI CRISPR Orthologs and Systems)”的美国临时专利申请62/471,710和律所案号为47627-05-2133并于2017年4月12日提交的题为“新颖的VI型CRISPR直系同源物和系统(Novel Type VI CRISPR Orthologs and Systems)”的美国临时专利申请。

[0234] 在本发明的一个实施方式中,HEPN结构域包含至少一个RxxxxH基序,其包含R{N/H/K}X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H的序列。在本发明的一个实施方式中,HEPN结构域包含RxxxxH基序,其包含R{N/H}X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H的序列。在本发明的一个实施方式中,HEPN结构域包含R{N/K}X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H的序列。在某些实施方式中,X<sub>1</sub>是R、S、D、E、Q、N、G、Y或H。在某些实施方式中,X<sub>2</sub>是I、S、T、V或L。在某些实施方式中,X<sub>3</sub>是L、F、N、Y、V、I、S、D、E或A。

[0235] 在形成CRISPR复合物的情形中,“靶序列”是指指导序列被设计成与其具有互补性的序列,其中靶序列与指导序列之间的杂交促进CRISPR复合物的形成。靶序列可以包含任何多核苷酸,诸如DNA或RNA多核苷酸。在一些实施方式中,靶序列位于细胞的细胞核或细胞质中。在一些实施方式中,可以通过搜索满足以下任何或所有条件的重复基序,在计算机上鉴定正向重复序列:1. 在II型CRISPR基因座侧翼的2Kb基因组序列窗口中;2. 跨度为20至50bp;和3. 间隔20至50bp。在一些实施方式中,可以使用这些标准中的2个,例如1和2、2和3或1和3。在一些实施方式中,可以使用所有3个标准。

[0236] 在本发明的实施方式中,术语指导序列和指导RNA,例如,能够将CRISPR-Cas效应蛋白指导至靶基因座的RNA,可以互换使用,如本文引用的文献,如W0 2014/093622(PCT/US2013/074667)。在通常情况下,指导序列(或间隔子序列)是与靶多核苷酸序列具有足够互补性的任何多核苷酸序列,可与目标序列杂交并引导与CRISPR复合物的序列特异性结合。在一些实施方式中,当使用合适的比对算法进行最佳比对时,指导序列与其对应的靶序列之间的互补程度约为或大于约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或更高。最佳比对可以借助于用于比对序列的任何合适算法来确定,其非限制性实例包括史密斯-沃特曼算法(Smith-Waterman算法)、尼德曼-翁施算法(Needleman-Wunsch算法)、基于巴罗斯-维勒变换(Burrows-Wheeler Transform)的算法(例如,巴罗斯-维勒比对仪(Burrows Wheeler Aligner)、ClustalW、Clustal X、BLAT、Novoalign(Novocraft

Technologies;在www.novocraft.com 上可得)、ELAND(Illumina, San Diego, CA)、SOAP(在soap.genomics.org.cn上可得)和Maq(在maq.sourceforge.net上可得)。在一些实施方式中,指导序列(或间隔子序列)的长度是约或多于约5、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、75个或更多个核苷酸。在一些实施方式中,指导序列的长度短于约75、50、45、40、35、30、25、20、15、12个或更少个核苷酸。优选地,指导序列是10-40个核苷酸长,如20-30个或20-40个核苷酸长或更长,如30个核苷酸长或约30个核苷酸长。在某些实施方式中,针对CRISPR-Cas效应物,指导序列是10-30个核苷酸长,如20-30个或20-40个核苷酸长或更长,如30个核苷酸长或约30个核苷酸长。在某些实施方式中,指导序列是10-30个核苷酸长,如20-30个核苷酸长,如30个核苷酸长。指导序列指导CRISPR复合物与靶序列的序列特异性结合的能力可以通过任何合适的测定法来评估。例如,可将足以形成CRISPR复合物的CRISPR系统组分(包括待测的指导序列)提供给具有相应靶序列的宿主细胞,例如通过转染编码CRISPR序列组分的载体,然后通过评估靶序列内的优先切割,例如通过本文所述的Surveyor测定。类似地,可通过提供靶序列,CRISPR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),并比较在测试和对照指导序列反应之间在靶序列处裂解的结合或速率来在试管中评估靶多核苷酸序列的切割。其他测定是可能的,并且对本领域技术人员而言将是显而易见的。

[0237] 在经典CRISPR-Cas系统中,指导序列与其对应的靶序列之间的互补程度可以是约或大于约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或100%;指导或RNA或crRNA可以是长度约或大于约5、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、75个或更多个核苷酸;或者指导或RNA或crRNA可以是长度小于约75、50、45、40、35、30、25、20、15、12个或更少个核苷酸;以及有利地tracrRNA的长度是30或50个核苷酸。然而,本发明的一个方面是减少脱靶相互作用,例如,减少指导与具有低互补性的靶序列相互作用。实际上,在实例中,显示出本发明涉及突变,其导致CRISPR-Cas系统能够区分具有大于80%至约95%的互补性,例如,83%-84%或88-89%或94-95%互补性的靶序列和脱靶序列(例如,区分具有18个核苷酸的靶标与具有1、2或3个错配的18个核苷酸的脱靶)。因此,在本发明的背景下,指导序列与其对应靶序列之间的互补程度大于94.5%或95%或95.5%或96%或96.5%或97%或97.5%或98%或98.5%或99%或99.5%或99.9%或100%。脱靶是在序列和指导之间的互补性小于100%或99.9%或99.5%或99%或99%或98.5%或98%或97.5%或97%或96.5%或96%或95.5%或95%或94.5%或94%或93%或92%或91%或90%或89%或88%或87%或86%或85%或84%或83%或82%或81%或80%,有利地,脱靶是在序列和指导之间的互补性是100%或99.9%或99.5%或99%或99%或98.5%或98%或97.5%或97%或96.5%或96%或95.5%或95%或94.5%。

[0238] 在某些实施方式中,可以通过引入错配来利用切割效率的调节,例如,1个或更多个错配,如间隔子序列和靶序列之间1或2个错配,包括沿着间隔子/靶点不匹配的位置。例如,双链错配越集中(即,不是3'或5'),则切割效率受到的影响越大。因此,通过选择沿着间隔子的错配位置,可以调节切割效率。举例来说,如果期望靶点的切割小于100%(例如,在细胞群中),则可以在间隔子序列中引入1个或更多个,如优选地在间隔子与靶序列之间的2个错配。沿着错配位置的间隔子越中心,切割百分数越低。

[0239] 如本文所述的根据本发明的方法包括在如本文所述的真核细胞中（在体外，即在分离的真核细胞中）诱导一种或多种核苷酸修饰，包括将如本文所述的载体递送至细胞。所述突变可以包括通过指导RNA或 sgRNA在细胞的每个靶序列上引入、缺失或取代一个或多个核苷酸。突变可以包括通过指导RNA在所述细胞的每个靶序列上引入、缺失或取代 1-75 个核苷酸。突变可以包括通过指导RNA在所述细胞的每个靶序列上引入、缺失或取代1、5、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。突变可以包括通过指导RNA在所述细胞的每个靶序列上引入、缺失或取代5、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。突变可以包括通过指导RNA在所述细胞的每个靶序列上引入、缺失或取代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。突变可以包括通过指导RNA在所述细胞的每个靶序列上引入、缺失或取代20、21、22、23、24、25、26、27、28、29、30、35、40、45、50或 75个核苷酸。突变可以包括通过指导RNA在所述细胞的每个靶序列上引入、缺失或取代40、45、50、75、100、200、300、400或500个核苷酸。

[0240] 对于最小化毒性和脱靶效应，控制Cas mRNA或蛋白和指导RNA递送的浓度将是重要的。可以通过在细胞或非人真核生物动物模型中测试不同的浓度并使用深度测序来分析潜在脱靶基因组基因座的修饰程度来确定Cas mRNA或蛋白和指导RNA的最佳浓度。

[0241] 通常地，在内源性CRISPR系统的背景下，CRISPR复合物（包含与靶序列杂交并与一个或多个Cas蛋白复合的指导序列）的形成导致在靶序列中或附近（例如，距离1、2、3、4、5、6、7、8、9、10、20、50 个或更多个碱基对内）的切割，但是可能取决于二级结构，特别是在RNA靶点的情况下。在一些情况下，在内源性CRISPR系统背景下，CRISPR 复合物（包含与靶序列杂交并与一个或多个Cas蛋白复合的指导序列）的形成导致在靶序列中或附近（例如，距离1、2、3、4、5、6、7、8、 9、10、20、50个或更多个碱基对内）一条或两条链（如适用）的切割。

[0242] 在根据本发明特别优选的实施方式中，指导RNA（能够将Cas引导至靶基因座）可以包含（1）能够与真核细胞中的靶基因座（多核苷酸靶基因座，如RNA靶基因座）杂交的指导序列；（2）正向重复（DR）序列，该序列位于单个RNA中，即sgRNA（以5' 至3' 方向排列）或crRNA。

[0243] 在本发明的实践中，关于CRISPR/Cas系统、其组分、及此类组分的递送（包括方法、材料、递送媒介物、载体、颗粒、以及其制造和使用（包括关于量和配制品））的一般信息，参考以下：美国专利号8,999,641、8,993,233、8,945,839、8,932,814、8,906,616、8,895,308、8,889,418、8,889,356、8,871,445、8,865,406、8,795,965、8,771,945和8,697,359；美国专利公开号US 2014-0310830（美国申请系列号14/105,031）、US 2014-0287938 A1（美国申请系列号14/213,991）、US 2014-0273234 A1（美国申请系列号14/293,674）、US2014-0273232 A1（美国申请系列号 14/290,575）、US 2014-0273231（美国申请系列号14/259,420）、US 2014-0256046 A1（美国申请系列号14/226,274）、US 2014-0248702 A1（美国申请系列号14/258,458）、US 2014-0242700 A1（美国申请系列号 14/222,930）、US 2014-0242699 A1（美国申请系列号14/183,512）、US 2014-0242664 A1（美国申请系列号14/104,990）、US 2014-0234972 A1（美国申请系列号14/183,471）、US 2014-0227787 A1（美国申请系列号 14/256,912）、US 2014-0189896 A1（美国申请系列号14/105,035）、US 2014-0186958（美国申请系列号14/105,017）、US 2014-0186919 A1（美国申请系列号14/104,

977)、US 2014-0186843 A1(美国申请系列号 14/104,900)、US 2014-0179770 A1(美国申请系列号14/104,837)和US 2014-0179006 A1(美国申请系列号14/183,486)、US 2014-0170753(美国申请系列号14/183,429);欧洲专利EP 2 784 162 B1和EP 2 771 468 B1;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);和PCT专利公开号PCT专利公开号WO 2014/093661 (PCT/US2013/074743)、WO 2014/093694 (PCT/US2013/074790)、WO 2014/093595 (PCT/US2013/074611)、WO 2014/093718 (PCT/US2013/074825)、WO 2014/093709 (PCT/US2013/074812)、WO 2014/093622 (PCT/US2013/074667)、WO 2014/093635 (PCT/US2013/074691)、WO 2014/093655 (PCT/US2013/074736)、WO 2014/093712 (PCT/US2013/074819)、WO 2014/093701 (PCT/US2013/074800)、WO 2014/018423 (PCT/US2013/051418)、WO 2014/204723 (PCT/US2014/041790)、WO 2014/204724 (PCT/US2014/041800)、WO 2014/204725 (PCT/US2014/041803)、WO 2014/204726 (PCT/US2014/041804)、WO 2014/204727 (PCT/US2014/041806)、WO 2014/204728 (PCT/US2014/041808)、WO 2014/204729 (PCT/US2014/041809)。还参考了美国临时专利申请号61/758,468; 61/802,174;61/806,375;61/814,263;61/819,803和61/828,130,其分别在2013年01月30日;2013年03月15日;2013年03月28日;2013年04月20日;2013年05月06日和2013年05月28日提交。还参考了在2013年06月17日提交的美国临时专利申请号61/836,123。另外参考了均在2013年06月17日提交的美国临时专利申请号61/835,931、61/835,936、61/836,127、61/836,101、61/836,080和61/835,973。还参考了在2013年08月05日提交的美国临时申请号61/862,468和61/862,355;在2013年08月28日提交的61/871,301;在2013年09月25日提交的 61/960,777和在2013年10月28日提交的61/961,980。进一步参考了:PCT专利申请号:PCT/US2014/041803、PCT/US2014/041800、PCT/US2014/041809、PCT/US2014/041804和PCT/US2014/041806,其均在2014年06月10日提交;在2014年06月11日提交的 PCT/US2014/041808;和在2014年10月28日提交的PCT/US2014/62558,以及美国临时专利申请系列号:61/915,150、61/915,301、61/915,267和 61/915,260,其均在2013年12月12日提交;61/757,972和61/768,959,其在2013年01月29日和2013年02月25日提交;61/835,936、61/836,127、61/836,101、61/836,080、61/835,973和61/835,931,其在2013年06月17日提交;62/010,888和62/010,879,均在2014年06月11日提交;62/010,329和62/010,441,分别在2014年06月10日提交;61/939,228 和61/939,242,均在2014年02月12日提交;61/980,012,其在2014年04月15日提交;62/038,358,其在2014年08月17日提交;62/054,490、62/055,484、62/055,460和62/055,487,其分别在2014年09月25日提交;以及62/069,243,其在2014年10月27日提交。还参考了美国临时专利申请号62/055,484、62/055,460、62/055,487,其在2014年09月25日提交;美国临时专利申请号61/980,012,其在2014年04月15日提交;和美国临时专利申请号61/939,242,其在2014年02月12日提交。尤其参见指定美国的PCT申请,申请号PCT/US14/41806,2014年06月10日提交。参考了于2014年01月22日提交的美国临时专利申请号 61/930,214。参考了美国临时专利申请号61/915,251;61/915,260和 61/915,267,其均在2013年12月12日提交。参考了于2014年04月15日提交的美国临时专利申请号USSN 61/980,012。尤其参见指定美国的 PCT申请,申请号PCT/US14/41806,2014年06月10日提交。参考了 2014年01月22日提交的美国临时专利申请号61/930,214。参考了美国临时专利申请号61/915,251;61/915,260和61/915,267,其均在

2013年 12月12日提交。

[0244] 还提到了2014年12月12日提交的美国申请号62/091,455,受保护的指导RNA (PROTECTED GUIDE RNAS (PGRNAS));2014年12月 24日提交的美国申请号62/096,708,受保护的指导RNA (PROTECTED GUIDE RNAS (PGRNAS));2014年12月24日提交的美国申请号 62/091,462,CRISPR转录因子的死指导 (DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS);2014年12月23日提交的美国申请号 62/096,324,CRISPR转录因子的死指导 (DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS);2014年12月12日提交的美国申请号 62/091,456,用于CRISPR-CAS系统的受护航和官能化的指导 (ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS);2014年12月12日提交的美国申请号62/091,461,用于关于造血干细胞 (HSC) 的基因组编辑的CRISPR-CAS系统和组合物的递送、用途以及治疗应用 (DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs));2014年12月19日提交的美国申请号62/094,903,通过基因组范围的插入捕获测序对双链断裂和基因组重排的无偏鉴定 (UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING);2014年12月24日提交的美国申请号62/096,761,用于序列操纵的系统、方法及优化的酶以及指导序列支架的工程化 (ENGINEERING OF SYSTEMS,METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION);2014 年12月30日提交的美国申请号62/098,059,RNA靶向系统 (RNA-TARGETING SYSTEM);2014年12月24日提交的美国申请号 62/096,656,具有稳定化结构域或与稳定化结构域相关联的CRISPR (CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS);2014年12月24日提交的美国申请号62/096,697,具有 AAV或与AAV相关联的CRISPR (CRISPR HAVING OR ASSOCIATED WITH AAV);2014年12月30日提交的美国申请号62/098,158,工程化CRISPR复合物插入靶向系统 (ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS);2015年04月22日提交的美国申请号,用于胞外核外报告的细胞靶向 (CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING);2014年12月24日提交的美国申请号62/054,490,使用粒子递送组分靶向病症和疾病的 CRISPR-CAS系统和组合物的递送、用途以及治疗应用 (DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS);2014年09月25日提交的美国申请号62/055,484,用于用优化的功能性CRISPR-CAS系统进行序列操纵的系统、方法和组合物 (SYSTEMS,METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS);2014年12月04日提交的美国申请号62/087,537,用于用优化的功能性CRISPR-CAS系统进行序列操纵的系统、方法和组合物 (SYSTEMS,METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS);2014年09月24日提交的美国申请号62/054,651,用于对体内多种癌症突变的竞争建模的 CRISPR-CAS系统和组合物的递送、用途以及治疗应用 (DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO);2014年10月23日提交的美国申请号62/067,886,用于对体内多种癌

症突变的竞争建模的CRISPR-CAS系统和组合物的递送、用途以及治疗应用 (DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO); 2014年12月24日提交的美国申请号62/054,675,在神经元细胞/组织中CRISPR-CAS系统和组合物的递送、用途以及治疗应用 (DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES); 2014年09月24日提交的美国申请号62/054,528,在免疫疾病或病症中CRISPR-CAS系统和组合物的递送、用途以及治疗应用 (DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS); 2014年09月25日提交的美国申请号62/055,454,用于使用细胞穿透肽 (CPP) 靶向病症和疾病的 CRISPR-CAS系统和组合物的递送、用途以及治疗应用 (DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP)); 2014年09月25日提交的美国申请号62/055,460,多功能性CRISPR复合物和/或优化的酶连接的功能性CRISPR复合物 (MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES); 2014年12月04日提交的美国申请号62/087,475,用优化的功能性CRISPR-CAS系统进行功能性筛选 (FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS); 2014年09月25日提交的美国申请号 62/055,487,用优化的功能性CRISPR-CAS系统进行功能性筛选 (FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS); 2014年12月04日提交的美国申请号 62/087,546,多功能性CRISPR复合物和/或优化的酶连接的功能性 CRISPR复合物 (MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES); 以及2014年12月30日提交的美国申请号62/098,285,对肿瘤生长和转移的CRISPR介导的体内建模以及遗传筛选 (CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS)。

[0245] 此外,关于CRISPR-Cas系统的一般信息,请注意以下内容(也通过引用并入本文):

[0246] ➤ 使用CRISPR/Cas系统进行多重基因组工程 (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);

[0247] ➤ 使用CRISPR-Cas系统细菌基因组的RNA指导的编辑 (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);

[0248] ➤ 通过CRISPR/Cas介导的基因组工程一步产生在多基因中携带突变的小鼠 (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);

[0249] ➤ 哺乳动物内源性转录和表观遗传状态的光学控制 (Optical control of mammalian endogenous transcription and epigenetic states)。Koneremann 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);

[0250] ➤通过RNA指导的CRISPR Cas9进行双切口以增强基因组编辑的特异性(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);

[0251] ➤RNA指导的Cas9核酸酶的DNA靶向特异性(DNA targeting specificity of RNA-guided Cas9 nucleases)。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);

[0252] ➤使用CRISPR-Cas9系统的基因组工程化(Genome engineering using the CRISPR-Cas9 system)。Ran,FA.,Hsu,PD.,Wright,J.,Agarwala,V., Scott,DA.,Zhang, F.Nature Protocols Nov;8(11):2281-308(2013-B);

[0253] ➤在人细胞中的基因组规模的CRISPR-Cas9敲除筛选(Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells)。Shalem,O.,Sanjana, NE.,Hartenian,E., Shi,X.,Scott,DA.,Mikkelsen,T.,Heckl,D.,Ebert, BL.,Root,DE.,Doench,JG.,Zhang, F.Science Dec 12.(2013)。[网络公开发表];

[0254] ➤在与指导RNA和靶DNA的复合物中cas9的晶体结构(Crystal structure of cas9 in 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);

[0255] ➤在哺乳动物细胞中CRISPR核酸内切酶Cas9的全基因组结合(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);

[0256] ➤用于基因组编辑和癌症建模的CRISPR-Cas9敲入小鼠(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):440-455DOI: 10.1016/j.cell.2014.09.014 (2014);

[0257] ➤用于基因组工程化CRISPR-Cas9的开发和应用(Development and Applications of CRISPR-Cas9 for Genome Engineering),Hsu PD,Lander ES,Zhang F.,Cell.Jun 5;157(6):1262-78(2014);

[0258] ➤使用CRISPR/Cas9系统在人细胞中的遗传筛选(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);

[0259] ➤ 用于CRISPR-Cas9介导的基因失活的高活性sgRNA的合理设计 (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., (于2014年09月03日网络发表) Nat Biotechnol. Dec; 32(12): 1262-7 (2014);

[0260] ➤ 使用CRISPR-Cas9对哺乳动物大脑中的基因功能进行体内询问 (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., (于2014年10月19日网络发表) Nat Biotechnol. Jan; 33(1): 102-6 (2015);

[0261] ➤ 工程化CRISPR-Cas9复合物实现基因组规模的转录激活 (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);

[0262] ➤ 用于诱导基因组编辑和转录调节的分离的Cas9结构 (A split-Cas9 architecture for inducible genome editing and transcription modulation), Zetsche B, Volz SE, Zhang F., (于2015年02月02日网络发表) Nat Biotechnol. Feb; 33(2): 139-42 (2015);

[0263] ➤ 在肿瘤生长和转移小鼠模型中基因组范围的CRISPR筛选 (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, 2015年03月12日 (在小鼠中的多重筛选), 和

[0264] ➤ 使用金黄色葡萄球菌Cas9的体内基因组编辑 (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., (于2015年04月01日网络发表), Nature. Apr 9; 520(7546): 186-91 (2015);

[0265] ➤ Shalem等, “使用CRISPR-Cas9的高通量功能性基因组学 (High-throughput functional genomics using CRISPR-Cas9),” Nature Reviews Genetics 16, 299-311 (2015年5月)。

[0266] ➤ Xu等, “改进的CRISPR sgRNA设计的序列决定子 (Sequence determinants of improved CRISPR sgRNA design),” Genome Research 25, 1147-1157 (2015年08月)。

[0267] ➤ Parnas等, “在原代免疫细胞中的全基因组CRISPR筛选以剖析调控网络 (A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks),” Cell 162, 675-686 (2015年07月30日)。

[0268] ➤ Ramanan等, “CRISPR/Cas9切割病毒DNA有效抑制乙型肝炎病毒 (CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus),” Scientific

Reports 5:10833.doi:10.1038/srep10833 (2015年06 月02日)

[0269] ➤ Nishimasu等,“金黄色葡萄球菌Cas9的晶体结构(Crystal Structure of *Staphylococcus aureus* Cas9),”*Cell* 162,1113-1126 (2015年08月27 日)

[0270] ➤ Zetsche等, (2015), “Cpf1是2类CRISPR-Cas系统的单一RNA指导的核酸内切酶(Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system),”*Cell* 163,759-771 (Oct.22,2015) doi: 10.1016/j.cell.2015.09.038.Epub Sep.25,2015

[0271] ➤ Shmakov等, (2015), “各种2类CRISPR-Cas系统的发现和功能表征(Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems),”*Molecular Cell* 60,385-397 (Nov.5,2015) doi: 10.1016/j.molcel.2015.10.008.Epub Oct 22,2015

[0272] ➤ Dahlman等,“使用具有催化活性Cas9核酸酶的直系同源基因控制 (Orthogonal gene control with a catalytically active Cas9 nuclease),” *Nature Biotechnology* 33,1159-1161 (2015年11月)

[0273] ➤ Gao等,“具有改变的PAM特异性的工程化Cpf1酶(Engineered Cpf1 Enzymes with Altered PAM Specificities),”*bioRxiv* 091611;doi: dx.doi.org/10.1101/091611Epub Dec.4,2016

[0274] ➤ Smargon等, (2017), “Cas13b是由附件蛋白Csx27和Csx28差异调节的VI-B型CRISPR关联RNA指导的RNase(Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28),”*Molecular Cell* 65,618-630 (Feb. 16,2017) doi:10.1016/j.molcel.2016.12.023.Epub Jan 5,2017

[0275] 在本发明的实践中,可以考虑将每一个都通过引用并入本文,并在下面简要讨论:

[0276] ➤ Cong等基于嗜热链球菌(*Streptococcus thermophilus*) Cas9以及产脓链球菌(*Streptococcus pyogenes*) Cas9设计了一种在真核细胞中使用的工程化II型CRISPR-Cas系统,并证明Cas9核酸酶可以由短RNA 指导,以诱导人和小鼠细胞中DNA的精确切割。他们的研究进一步表明,Cas9转化为切口酶可用于以最小的诱变活性促进真核细胞中同源性定向的修复。此外,他们的研究表明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源基因组位点中的几个,从而证明了RNA指导的核酸酶技术的易编程性和广泛适用性。利用RNA来编程细胞中的序列特异性DNA切割的能力定义了一类新的基因组工程工具。这些研究进一步表明,其他CRISPR 基因座可能也可以移植到哺乳动物细胞中,并且还可以介导哺乳动物基因组切割。重要的是,可以设想,可以进一步改进CRISPR-Cas系统的几个方面以提高其效率和多功能性。

[0277] ➤ Jiang等使用了簇状、规则间隔、短回文重复序列(CRISPR)相关的 Cas9内切核酸酶与双RNA配合,在肺炎链球菌(*Streptococcus pneumoniae*)和大肠杆菌(*Escherichia coli*)的基因组中引入了精确的突变。该方法依赖于靶向基因组位点的双RNA:Cas9定向切割来杀死未突变的细胞,从而避免了对选择标记或反选择系统的需求。这项研究报告,通过改变短CRISPR RNA (crRNA)的序列,对双RNA: Cas9特异性进行重新编程,以使编辑模板上的单核苷酸和多核苷酸发生变化。研究表明,同时使用两种crRNA可以进行多重诱变。此外,

当该方法与重组结合使用时,在肺炎链球菌中,使用所述方法回收的细胞中几乎有100%包含所需的突变,而在大肠杆菌中,回收的细胞中有65%包含该突变。

[0278] ➤Wang等(2013年)使用CRISPR/Cas系统一步生成具有多个基因突变的小鼠,这些基因传统上是通过胚胎干细胞中的顺序重组和/或具有单个突变的小鼠耗时的交叉杂交而在多个步骤中生成的。CRISPR/Cas 系统将大大加速功能冗余基因和上位基因相互作用的体内研究。

[0279] ➤Koneremann等(2013)满足了本领域对通用和可靠技术的需求,这些技术能够对基于CRISPR Cas9酶的DNA结合结构域以及转录激活因子样效应物进行光学和化学调节。

[0280] ➤Ran等(2013-A)描述了一种将Cas9切口酶突变体与成对的指导RNA 结合以引入靶向双链断裂的方法。这解决了来自微生物CRISPR-Cas 系统的Cas9核酸酶通过指导序列靶向特定基因组基因座的问题,该指导序列可以耐受与DNA靶标的某些错配,从而促进不期望的脱靶诱变。由于基因组中的各个缺口均以高真度进行修复,因此双链断裂需要通过适当偏移的指导RNA同时进行缺口,并扩展了特异性识别的靶标碱基的数目。作者证明,使用配对切口可以在细胞系中将脱靶活性降低50到1,500倍,并促进小鼠受精卵中的基因敲除,而不会牺牲对靶标的切割效率。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程序。

[0281] ➤Hsu等(2013)表征了SpCas9在人类细胞中的靶向特异性,以告知靶位点的选择并避免脱靶效应。该研究评价了293T和293FT细胞中>100 个预测的基因组脱靶基因座的>700个指导RNA变体和SpCas9诱导的 indel突变水平。作者认为,SpCas9能够以序列依赖的方式耐受对错配的数量、位置和分布敏感的不同位置的指导RNA和靶DNA之间的错配。作者进一步表明,SpCas9介导的切割不受DNA甲基化的影响,并且可以对SpCas9和sgRNA的剂量进行滴定,以最大程度地减少脱靶修饰。此外,为促进哺乳动物基因组工程应用,作者报告提供了一种基于网络的软件工具,以指导靶序列的选择和验证以及脱靶分析。➤Ran等(2013-B)描述了一组通过哺乳动物细胞中的非同源末端连接 (NHEJ) 或同源性定向修复(HDR)进行Cas9介导的基因组编辑的工具,以及用于下游功能研究的修饰细胞系的生成。为了最大程度地减少脱靶切割,作者进一步描述了使用Cas9切口酶突变体与成对的指导RNA进行双重切口的策略。作者提供的方案是通过实验得出指导,用于选择靶位点,评估切割效率和分析脱靶活性。研究表明,从靶标设计开始,基因修饰可在短短的1-2周内完成,而修饰的克隆细胞系则可在2-3周内获得。

[0282] ➤Shalem等描述了一种在基因组范围规模内问询基因功能的新方法。他们的研究表明,以18,080个基因为目标的基因组规模的CRISPR-Cas9 基因敲除(GeCKO)文库的递送具有64,751个独特的指导序列,可在人细胞中进行阴性和阳性选择筛选。首先,作者展示了使用GeCKO 文库鉴定癌症和多能干细胞中细胞活力所必需的基因。接下来,在黑色素瘤模型中,作者筛选了基因丢失与维罗非尼抗药性有关的基因,维罗非尼是一种抑制突变蛋白激酶BRAF的药物。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,以及新型命中的NF2、CUL3、TADA2B和TADA1。作者观察到靶向相同基因的独立指导RNA之间的高度一致性和较高的命中率,从而证明了用 Cas9进行基因组规模筛选的希望。

[0283] ➤Nishimasu等报道了酿脓链球菌Cas9与sgRNA及其靶DNA的复合物的晶体结构,

分辨率为 $2.5\text{\AA}$ 。该结构揭示了由靶标识别和核酸酶叶组成的双叶结构,在界面处带正电荷的沟槽中容纳了sgRNA:DNA 异源双链体。识别叶对于结合sgRNA和DNA是必不可少的,而核酸酶叶则包含HNH和RuvC核酸酶结构域,它们的位置正确,分别用于切割靶DNA的互补链和非互补链。核酸酶叶还包含一个羧基末端结构域,负责与原型间隔子邻近基序(PAM)相互作用。这种高分辨率的结构和伴随的功能分析揭示了Cas9靶向RNA指导的DNA的分子机制,从而为合理设计新的通用基因组编辑技术铺平了道路。

[0284] ➤ Wu等绘制了在小鼠胚胎干细胞(mESC)中酿脓链球菌的催化无活性 Cas9 (dCas9)的全基因组结合位点,该酿脓链球菌载有单个指导RNA (sgRNA)。作者表明,所测试的四个sgRNA中的每一个都将dCas9 靶向到数十至数千个基因组位点,通常以sgRNA中的5个核苷酸种子区和NGG原型间隔子邻近基序(PAM)为特征。TChromatin的不可及性降低了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,在用催化活性Cas9转染的mESC 中,对295个dCas9结合位点进行了靶向测序,仅发现了一个高于背景水平突变的位点。作者提出了Cas9结合和裂解的两种状态模型,其中种子匹配触发了结合,但裂解需要与靶DNA进行广泛配对。

[0285] ➤ Platt等建立了一种Cre依赖性Cas9敲入小鼠。作者展示了使用腺相关病毒(AAV)、慢病毒或颗粒介导的神经元,免疫细胞和内皮细胞中指导RNA的体内和离体基因组编辑。

[0286] ➤ Hsu等(2014)是一篇综述文章,一般性地讨论了从酸奶到基因组编辑的CRISPR-Cas9的历史,包括细胞的遗传筛选。

[0287] ➤ Wang等(2014)涉及使用基因组规模的慢病毒单一指导RNA (sgRNA) 文库,用于阳性和阴性选择的合并、功能丧失的遗传筛选方法。

[0288] ➤ Doench等创建了一个sgRNA池,将六个内源性小鼠和三个内源性人类基因组中所有可能的靶位点进行平铺,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无效等位基因的能力。作者表明,优化PAM可以提高活性,并且还提供了用于设计sgRNA的在线工具。

[0289] ➤ Swiech等证明了AAV介导的SpCas9基因组编辑可以实现大脑中基因功能的反向遗传研究。

[0290] ➤ Konermann等(2015)讨论了将多个效应结构域(例如,转录激活因子、功能性和表观基因调节因子)在指导上的适当位置(例如,带有或不带有接头的茎或四环)连接的能力。

[0291] ➤ Zetsche等证明了Cas9酶可以被分为两部分,因此可以控制Cas9用于激活的组装。

[0292] ➤ Chen等通过证明小鼠的全基因组体内CRISPR-Cas9筛选揭示了调节肺转移的基因,从而进行了多重筛选。

[0293] ➤ Ran等(2015)涉及SaCas9及其编辑基因组的能力,并证明不能从生化分析中推断。Shalem等(2015)描述了催化无活性的Cas9 (dCas9) 融合物用于合成抑制(CRISPRi)或激活(CRISPRa)表达的方式,展示了使用Cas9进行基因组规模筛选(包括阵列和合并筛选),使基因组位点失活的敲除方法和调节转录活性的策略方面的进展。

[0294] 末端编辑

[0295] ➤Shalem等(2015)描述了催化无活性的Cas9(dCas9)融合物用于合成抑制(CRISPRi)或激活(CRISPRa)表达的方式,显示了Cas9在用于基因组规模筛选(包括阵列筛选和合并筛选),使基因组位点失活的敲除方法以及调节转录活性的策略方面的进展。

[0296] ➤Xu等(2015)评估了基于CRISPR的筛选中有助于单一指导RNA(sgRNA)效率的DNA序列特征。作者探索了CRISPR/Cas9敲除的效率和切割位点的核苷酸偏好。作者还发现,CRISPRi/a的序列偏好与CRISPR/Cas9敲除的序列偏好有很大不同。

[0297] ➤Parnas等(2015)将全基因组的CRISPR-Cas9文库引入树突状细胞(DC)中,以鉴定控制细菌脂多糖(LPS)诱导肿瘤坏死因子(Tnf)的基因。确定了Tlr4信号的已知调节剂和以前未知的候选物,并将其分为三个功能模块,这些功能模块对LPS的经典响应具有明显的影响。

[0298] ➤Ramanan等(2015)证明了感染细胞中病毒游离DNA(cccDNA)的裂解。HBV基因组以称为共价闭环状DNA(cccDNA)的3.2kb双链游离型DNA形式存在于被感染的肝细胞核中,这是HBV生命周期中的关键组成部分,其复制不受当前疗法的抑制。作者表明,特异性靶向HBV高保守区的sgRNA可以强有力地抑制病毒复制和耗尽cccDNA。

[0299] ➤Nishimasu等(2015)报道了SaCas9的晶体结构与单个指导RNA(sgRNA)及其双链DNA靶标复合,其中包含5'-TTGAAT-3' PAM和5'-TTGGGT-3' PAM。SaCas9与SpCas9的结构比较突出了结构保守性和差异性,解释了它们独特的PAM特异性和直系同源sgRNA识别。

[0300] 而且,“用于高度特异性基因组编辑的二聚体CRISPR RNA指导FokI核酸酶(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,其识别延伸的序列并且可以在人细胞中高效编辑内源基因。此外,提及PCT申请号PCT/US14/70057,律所案号47627.99.2060和BI-2013/107,发明名称为“使用颗粒递送组合物,用于治疗病症和疾病的CRISPR-CAS系统和组合物的递送、用途和治疗应用(DELIVERY,USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS)(要求以下美国临时专利申请的一个或多个或全部的优先权:2014年09月24日提交的62/054,490;2014年06月10日提交的62/010,441;和均在2013年12月12日提交的61/915,118、61/915,215和61/915,148) (“颗粒递送PCT”),其通过引用并入本文,其涉及制备含sgRNA和Cas9蛋白的颗粒的方法,该方法包括将包含sgRNA和Cas9蛋白的混合物(和任选地HDR模板)与包含或基本上由表面活性剂、磷脂、可生物降解的聚合物、脂蛋白和乙醇;以及来自此过程的颗粒混合。例如,其中Cas9蛋白和sgRNA以适宜的(例如,3:1至1:3或2:1至1:2或1:1)摩尔比,在适宜温度下(例如,15-30C,例如,20-25C,例如,室温),以适宜时间(例如,15-45,如30分钟),优选地在无菌、无核酸酶的缓冲液(例如,1X PBS)中混合在一起。单独地,将颗粒组分,诸如或包含:表面活性剂,例如,阳离子脂质,例如,1,2-二油酰基-3-三甲基铵丙烷(DOTAP);磷脂,例如,二豆蔻酰基磷脂酰胆碱(DMPC);可生物降解的聚合物(例如乙二醇聚

合物或PEG)和脂蛋白(例如低密度脂蛋白)(例如,胆固醇)溶于醇,最好是C<sub>1-6</sub>烷基醇(如甲醇、乙醇、异丙醇),例如100%乙醇。将两种溶液混合在一起以形成含有Cas9-sgRNA复合物的颗粒。因此,在将整个复合物配制成颗粒之前,可以将sgRNA与Cas9蛋白预先复合。可以用不同摩尔比的不同已知成分来制备制剂,这些成分可以促进核酸向细胞内的传递(例如,1,2-二油酰基-3-三甲基铵丙烷(DOTAP)、1,2-二十四烷酰基-sn-甘油-3-磷酸胆碱(DMPC)、聚乙二醇(PEG)和胆固醇)。例如,DOTAP:DMPC:PEG:胆固醇摩尔比可以是DOTAP 100、DMPC 0、PEG 0、胆固醇0;或者DOTAP 90、DMPC 0、PEG 10、胆固醇0;或者DOTAP 90、DMPC 0、PEG 5、胆固醇5;或者DOTAP 100、DMPC 0、PEG 0、胆固醇0。因此,该申请包括将sgRNA、Cas9蛋白和形成颗粒的组分混合在一起;以及这种混合产生的颗粒。本发明的各方面可以涉及颗粒;例如,使用类似于颗粒输送PCT的方法的颗粒,例如,通过将包含本发明中crRNA和/或CRISPR-Cas的混合物与形成颗粒的组分混合,例如,如颗粒输送PCT,通过这种混合形成一个或多个颗粒(或者,当然,如本发明中涉及crRNA和/或CRISPR-Cas的其他颗粒)。

#### [0301] 指导序列

[0302] 在本发明的实施方式中,术语指导序列和指导RNA,即,能够将Cas 导向靶基因组基因座的RNA,如前面引用的文献诸如W02014/093622 (PCT/US2013/074667)中所述互换使用。在通常情况下,指导序列是与靶多核苷酸序列具有足够互补性以与靶序列杂交并且引导CRISPR复合物序列特异性地结合至靶序列的任何多核苷酸序列。在一些实施方式中,当使用合适的比对算法最佳比对时,指导序列与其相应靶序列之间的互补程度为约或大于约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或更大。最佳比对可以借助于用于比对序列的任何合适算法来确定,其非限制性实例包括史密斯-沃特曼算法(Smith-Waterman算法)、尼德曼-翁施算法(Needleman-Wunsch算法)、基于巴罗斯-维勒变换(Burrows-Wheeler Transform)的算法(例如,巴罗斯-维勒比对仪(Burrows Wheeler Aligner)、ClustalW、Clustal X、BLAT、Novoalign (Novocraft Technologies;在www.novocraft.com上可得)、ELAND (Illumina, San Diego, CA)、SOAP(在soap.genomics.org.cn上可得)和Maq(在maq.sourceforge.net上可得)。在一些实施方式中,指导序列的长度是约或多于约5、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、75个或更多个核苷酸。在一些实施方式中,指导序列的长度短于约75、50、45、40、35、30、25、20、15、12个或更少个核苷酸。优选地,指导序列是10-30个核苷酸长,如30个核苷酸长。指导序列指导CRISPR复合物与靶序列的序列特异性结合的能力可以通过任何合适的测定法来评估。例如,可将足以形成CRISPR复合物的CRISPR系统组分(包括待测的指导序列)提供给具有相应靶序列的宿主细胞,例如通过转染编码CRISPR序列组分的载体,然后通过评估靶序列内的优先切割,例如通过本文所述的Surveyor测定。类似地,可通过提供靶序列,CRISPR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),并比较在测试和对照指导序列反应之间在靶序列处裂解的结合或速率来在试管中评估靶多核苷酸序列的切割。其他测定是可能的,并且对本领域技术人员而言将是显而易见的。可以对指导序列进行选择以靶向任何靶序列。在一些实施方式中,靶序列是在细胞基因组中的序列。示例性靶序列包括靶基因组中独特的那些。

[0303] 在通常情况下,以及在本说明书通篇,术语“载体”是指能够转运它所连接的另一个核酸的核酸分子。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

个游离端、不包含游离端(例如,环状)的核酸分子;包含DNA、RNA或二者的核酸分子;以及本领域已知的其他种类多核苷酸。一种类型的载体是“质粒”,该质粒是指一种环状双链DNA环,可以诸如通过标准分子克隆技术向该环中插入另外的DNA区段。另一种类型的载体是病毒载体,其中病毒来源的DNA或RNA序列存在于包装到病毒(例如,逆转录病毒、复制缺陷型逆转录病毒、腺病毒、复制缺陷型腺病毒以及腺相关病毒)中的载体中。病毒载体还包括由转染到宿主细胞中的病毒携带的多核苷酸。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附加型哺乳动物载体)。其他载体(例如,非附加型哺乳动物载体)在引入到宿主细胞后被整合到宿主细胞的基因组中,并且因此随着宿主基因组一起复制。此外,某些载体能够引导它们可操作地连接的基因的表达。此类载体在此被称为“表达载体”。用于真核细胞并在真核细胞中表达的载体在本文中可称为“真核表达载体”。在重组DNA技术中实用的常见表达载体常常是质粒形式。

[0304] 重组表达载体可以包含处于适用于核酸在宿主细胞中表达的形式之本发明的核酸,这意味着重组表达载体包含一个或多个调节元件,这些调节元件可以基于用于表达的宿主细胞来选择,可操作地连接至有待表达的核酸序列。在重组表达载体内,“可操作地连接”旨在意指感兴趣的核苷酸序列以允许核苷酸序列表达(例如,在体外转录/翻译系统中或当该载体被引入到宿主细胞时在宿主细胞中)的方式连接至一个或多个调节元件。

[0305] 术语“调节元件”旨在包括启动子、增强子、内部核糖体进入位点(IRES)以及其他表达控制元件(例如,转录终止信号,诸如多聚腺苷酸化信号和聚U序列)。此类调控元件描述于例如Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990)中。调节元件包括引导核苷酸序列在许多类型的宿主细胞中连续表达的那些元件和引导核苷酸序列仅在某些宿主细胞中表达的那些元件(例如,组织特异性调节序列)。组织特异性启动子可以引导主要在希望的感兴趣的组织诸如肌肉、神经元、骨骼、皮肤、血液、特定器官(例如,肝脏、胰脏)、或特定细胞类型(例如,淋巴细胞)中的表达。调节元件还可以时间依赖性方式诸如细胞周期依赖性 or 发育阶段依赖性方式引导表达,这可以是或也可以不是组织特异性或细胞类型特异性的。在一些实施方式中,体包含一个或多个pol III启动子(例如,1、2、3、4、5、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、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、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,逆转录病毒劳斯氏肉瘤病毒(RSV) LTR启动子(任选地具有RSV增强子)、巨细胞病毒(CMV)启动子(任选地具有CMV增强子)[参见,例如,Boshart等,Cell, 41:521-530(1985)]、SV40启动子、二氢叶酸还原酶启动子、 $\beta$ -肌动蛋白启动子、磷酸甘油激酶(PGK)启动子、以及EF1 $\alpha$ 启动子。术语“调节元件”还涵盖增强子元件,诸如WPRE;CMV增强子;HTLV-I的LTR中的R-U5'区段(Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988);SV40增强子;以及兔 $\beta$ -球蛋白的外显子2与3之间的内含子序列(Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981)。本领域技术人员将了解的是,表达载体的设计可以取决于诸如有待转化的宿主细胞的选择、所希望的表达水平等因素。载体可以引入到宿主细胞中从而产生由在此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)。

[0306] 有利的载体包括慢病毒和腺伴随病毒并且此类载体类型还可以针对靶向的特定细胞类型来选择。

[0307] 如在此所用,术语VI型CRISPR-Cas座位效应蛋白的“crRNA”或“指导RNA”或“单一指导RNA”或“sgRNA”或“一种或多种核酸组分”包括与靶核酸序列具有足够互补性以与靶核酸序列杂交并引导核酸靶向复合物与靶核酸序列的序列特异性结合的任何多核苷酸序列。

[0308] 在某些实施方式中,如本文所提供的CRISPR系统可以利用包含指导序列的crRNA或类似的多核苷酸,其中所述多核苷酸是RNA、DNA 或者RNA和DNA的混合物,和/或所述多核苷酸包含一种或多种核苷酸类似物。所述序列可以包含任何结构,包括但不限于天然crRNA的结构,如凸起、发夹或茎环结构。在某些实施方式中,包含指导序列的多核苷酸与第二多核苷酸序列形成双链体,所述第二多核苷酸序列可以是RNA 或DNA序列。

[0309] 在某些实施方式中,本发明的指导包括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。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/或碱基部分被修饰在本发明的一个实施方式中,指导核酸包含核糖核苷酸和非核糖核苷酸。在一个这样的实施方式中,指导序列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式中,指导序列包含一种或多种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联、硼酸磷酸酯键联的核苷酸、包含在核糖环的2' 和4' 碳原子之间的亚甲基桥的锁定核酸(LNA)或桥接核酸(BNA)。修饰的核苷酸的其他实例包括2' -0- 甲基类似物、2' -脱氧类似物、2-硫代尿苷类似物、N6-甲基腺苷类似物或 2' -氟类似物。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( $\Psi$ )、N1-甲基假尿苷(me1 $\Psi$ )、5-甲氧基尿苷(5moU)、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' -0-甲基(M)、2' -0-甲基-3' -硫代磷酸酯(MS)、硫代磷酸酯(PS)、S-约束乙基(cEt),或2' -0-甲基-3' -硫代PACE(MSP)。此类化学修饰的指导序列与未修饰的指导序列相比可以包含增加的稳定性和增加的活性,不过中靶对脱靶特异性不可预测。(参见,Hendel,2015, Nat Biotechnol.33(9):985-9,doi:10.1038/nbt.3290,2015年06月29日网络发表;Allerson等,J.Med.Chem.2005,48:901-904;Bramsen等,Front. Genet.,2012,3:154;Deng等,PNAS,2015,112:11870-11875;Sharma等, MedChemComm., 2014,5:1454-1471;Li等,Nature Biomedical Engineering, 2017,1,0066DOI:10.1038/s41551-017-0066)。

[0310] 在一些实施方式中,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性部分修饰。(参见Kelly等,2016,J.Biotech.233:74-83)。在某些实施方式中,指导序列在结合至靶DNA的区域中包含核糖核苷酸,并在结合至Cas9、Cpf1 或C2c1的区域中包含一个或多个脱氧核糖核苷酸和/或核苷酸类似物。在本发明的一个实施方式中,将脱氧核糖核苷酸和/或核苷酸类似物并入工程化指导结构(诸如但不限于5' 端和/或3' 端、茎环区和种子区)中。在某些实施方式中,修饰不在茎环区的5' 柄(5' -handle)中。指导序列的茎环区的5' 柄中的化学修饰可能会废除其功能(参见Li等, Nature Biomedical Engineering,2017,1:0066)。在某些实施方式中,指导序列的至少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个核苷酸经化学修饰。在一些实施方式中,指导序列的3' 或5' 端的3-

5个核苷酸经化学修饰。在一些实施方式中,在种子区中仅引入较小的修饰,诸如2'-F修饰。在一些实施方式中,在指导序列的3'端引入2'-F修饰。在某些实施方式中,指导序列的5'和/3'端的3至5个核苷酸用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PACE(MSP)化学修饰。这样的修饰可以提高基因组编辑效率(参见Hendel等,Nat.Biotechnol.(2015)33(9):985-989)。在某些实施方式中,指导序列的所有磷酸二酯键被硫代磷酸酯(PS)取代以增强基因破坏的水平。在某些实施方式中,指导序列的5'和/3'端的多于5个核苷酸用2'-O-Me、2'-F或S-约束乙基(cEt)化学修饰。这种化学修饰的指导序列可以介导增强的基因破坏水平(参见Ragdarm等,0215,PNAS,E7110-E7111)。在本发明的一个实施方式中,指导序列被修饰成在其3'和/或5'端包含化学部分。这样的部分包括但不限于胺、叠氮化物、炔、硫代基、二苯并环辛炔(DBCO)或若丹明。在某些实施方式中,化学部分通过接头诸如烷基链缀合至指导序列。在某些实施方式中,修饰的指导序列的化学部分可用于将指导序列附接至另一分子,诸如DNA、RNA、蛋白质或纳米粒子。这种化学修饰的指导序列可用于识别或富集一般由CRISPR系统编辑的细胞(参见Lee等,eLife,2017,6:e25312,DOI:10.7554)。

[0311] 在一些实施方式中,对指导序列的修饰是化学修饰、插入、缺失或拆分。在一些实施方式中,化学修饰的实例包括但不限于并入2'-O-甲基(M)类似物、2'-脱氧类似物、2-硫代尿苷类似物、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( $\Psi$ )、N1-甲基假尿苷(me1 $\Psi$ )、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PACE(MSP)。在一些实施方式中,指导序列包含一种或多种硫代磷酸酯修饰。在某些实施方式中,指导序列的至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或25个核苷酸经化学修饰。在某些实施方式中,种子区中的一个或多个核苷酸经化学修饰。在某些实施方式中,在3'端的一个或多个核苷酸经化学修饰。在某些实施方式中,5'柄中的核苷酸均未经化学修饰。在一些实施方式中,种子区中的化学修饰是次要修饰,诸如并入2'-氟类似物。在具体实施方式中,种子区的一个核苷酸被2'-氟类似物替代。在一些实施方式中,3'端中的5或10个核苷酸经化学修饰。在Cpf1 CrRNA的3'端处的此类化学修饰提高基因切割效率(参见Li等,Nature Biomedical Engineering,2017,1:0066)。在具体实施方式中,3'端中的5个核苷酸被2'-氟类似物替代。在具体实施方式中,3'端中的10个核苷酸被2'-氟类似物替代。在具体实施方式中,3'端中的5个核苷酸被2'-O-甲基(M)类似物替代。

[0312] 在一些实施方式中,指导序列的5'柄的环经修饰。在一些实施方式中,指导序列的5'柄的环被修饰为具有缺失、插入、拆分或化学修饰。在某些实施方式中,环包含3、4或5个核苷酸。在某些实施方式中,环包含序列UCUU、UUUU、UAUU或UGUU。

[0313] 在一个方面中,引导物包括通过非磷酸二酯键化学连接或共轭的部分。在一个方面中,在一个非限制性实例中,该引导物包括通过非核苷酸环化学连接或缀合的正向重复序列部分和靶向序列部分。在一些实施方式中,这些部分通过非磷酸二酯共价接头连接。共价接头的实例包括但不限于选自以下的化学修饰:氨基甲酸酯、醚、酯、酰胺、亚胺、脒、氨基三嗪、脞、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、富勒烯、亚砷、脲、硫脲、酰肼、胍、三唑、光不稳定键、C-C键形成基团,如Diels-Alder环加成对或闭环易位对和Michael反应对。

[0314] 在一些实施方式中,首先使用标准亚磷酰胺合成方法合成指导的各部分(Herdewijn,P.编著,Methods in Molecular Biology Col 288, Oligonucleotide Synthesis:Methods and Applications,Humana Press,New Jersey(2012))。在一些实施方式中,使用本领域已知的标准方案,可以将非靶向引导部分官能化以包含合适的用于连接的官能团(Hermanson, G.T.,Bioconjugate Techniques,Academic Press(2013))。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、氯代羰基、咪唑基羰基、氢叠氮化物、氨基脒、硫代氨基脒、硫醇、马来酰亚胺、卤代烷基、亚磺酰基、烯丙基、炔丙基、二烯、炔和叠氮化物。一旦指导的非靶向部分被官能化,就可以在两个寡核苷酸之间形成共价化学键或键。示例性的化学键包括但不限于基于氨基甲酸酯、醚、酯、酰胺、亚胺、脒、氨基三嗪、脎、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、富勒烯、亚砷、脲、硫脲、酰肼、肟、三唑、光不稳定键、C-C键形成基团,如Diels-Alder环加成对或闭环易位对和Michael反应对的那些。

[0315] 在一些实施方式中,指导的一个或多个部分可以是化学合成的。在一些实施方式中,化学合成使用具有2'-乙酰氧基乙基原酸酯(2'-ACE)的自动化固相寡核苷酸合成机(Scarange等,J.Am.Chem.Soc.(1998) 120: 11820-11821;Scarange,Methods Enzymol.(2000) 317:3-18)或2'-硫代氨基甲酸酯(2'-TC)化学(Dellinger等,J.Am.Chem.Soc.(2011) 133: 11540-11546;Hendel等,Nat.Biotechnol.(2015) 33:985-989)。

[0316] 在一些实施方式中,通过修饰糖、核苷酸间磷酸二酯键、嘌呤和嘧啶残基,可以使用各种生物缀合反应、环、桥和非核苷酸键将引导部分共价连接。Sletten等,Angew.Chem.Int.Ed.(2009) 48:6974-6998; Manoharan,M.Curr.Opin.Chem.Biol.(2004) 8:570-9;Behlke等, Oligonucleotides(2008) 18:305-19;Watts等,Drug.Discov.Today(2008) 13:842-55;Shukla等,ChemMedChem(2010) 5:328-49。

[0317] 在一些实施方式中,可以使用点击化学将指导部分共价连接。在一些实施方式中,可以使用三唑接头指导部分共价连接。在一些实施方式中,可以使用涉及炔烃和叠氮化物的Huisgen 1,3-偶极环加成反应将导向部分共价连接,以产生高度稳定的三唑连接基(He等,ChemBioChem(2015) 17:1809-1812;WO 2016/186745)。在一些实施方式中,通过连接5'-己炔部分和3'-叠氮化物部分将指导部分共价连接。在一些实施方式中,5'-己炔指导部分和3'-叠氮化物指导部分中的一个或两个都可以用2'-乙酰氧基乙基原酸酯(2'-ACE)基团保护,随后可以使用Dharmacon方案将其除去(Scarange等,J.Am.Chem.Soc.(1998) 120: 11820-11821; Scarange,Methods Enzymol.(2000) 317:3-18)。

[0318] 在一些实施方式中,指导部分可以通过接头(例如,非核苷酸环)共价连接,该接头包含诸如间隔子、附着、生物缀合物、生色团、报道基团、染料标记的RNA和非天然存在的核苷酸类似物的部分。更具体地,用于本发明目的的合适的间隔子包括但不限于聚醚(例如,聚乙二醇、多元醇、聚丙二醇或乙烯和丙二醇的混合物),聚胺基团(例如,斯潘宁、亚精胺及其聚合物衍生物)、聚酯(例如,聚(丙烯酸乙酯))、聚磷酸二酯、亚烷基及其组合。合适的附接物包括可以被添加到接头以向接头增加附加性质的任何部分,例如但不限于荧光标记。合适的生物缀合物包括但不限于肽、糖苷、脂质、胆固醇、磷脂、二酰基甘油和二烷基甘油、脂肪酸、烃、酶底物、类固醇、生物素、洋地黄毒苷、碳水化合物、多糖。合适的生色团、报道基团和染料标记的RNA包括但不限于荧光染料,例如荧光素和若丹明、化学发光、电化学

发光和生物发光标记化合物。将两个RNA组分缀合的示例接头的设计也在WO 2004/015075中进行了描述。

[0319] 接头(例如,非核苷酸环)可以是任何长度。在一些实施方式中,接头的长度约等于0-16个核苷酸。在一些实施方式中,接头的长度约等于0-8个核苷酸。在一些实施方式中,接头的长度约等于0-4个核苷酸。在一些实施方式中,接头的长度约等于2个核苷酸。示例性接头设计也在W02011/008730中进行了描述。

[0320] 在一些实施方式中,当使用合适的比对算法进行最佳比对时,互补度约为或大于约50%、60%、75%、80%、85%、90%、95%、97.5%、99%或更高。可以使用任何合适的用于比对序列的算法来确定最佳比对,其非限制性实例包括史密斯-沃特曼算法(Smith-Waterman算法)、尼德曼-翁施算法(Needleman-Wunsch算法)、基于巴罗斯-维勒变换(Burrows-Wheeler Transform)的算法(例如,巴罗斯-维勒比对仪(Burrows Wheeler Aligner)、ClustalW、Clustal X、BLAT、Novoalign (Novocraft Technologies;在www.novocraft.com上可得)、ELAND (Illumina, San Diego, CA)、SOAP(在soap.genomics.org.cn上可得)和Maq(在maq.sourceforge.net上可得)。可以通过任何合适的测定来评定指导序列(在RNA靶向指导RNA或crRNA内)引导核酸靶向复合物序列特异性地结合至靶核酸序列的能力。例如,足以形成核酸靶向复合物的RNA靶向CRISPR-Cas系统的组分,包括有待测试的指导序列,可以提供给具有相应靶核酸序列的宿主细胞,诸如通过用编码核酸靶向复合物的组分的载体转染,继而诸如通过如本文所述的Surveyor测定评定靶核酸序列内的优先靶向(例如,切割)。类似地,可以在试管中通过提供靶核酸序列、核酸靶向复合物的组分(包括有待测试的指导序列)和不同于测试指导序列的对照指导序列,以及在测试指导序列与对照指导序列反应之间比较靶序列处的结合或切割速率来评估靶核酸序列的切割。其他测定可能存在,并且将为本领域技术人员所想到。可以选择指导序列并且因此选择核酸靶向指导以靶向任何靶核酸序列。靶序列可以是DNA。靶序列可以是任何RNA序列。在一些实施方式中,靶序列可以是选自由以下组成的组的RNA分子内的序列:信使RNA(mRNA)、前体mRNA、核糖体RNA(rRNA)、转移RNA(tRNA)、微小RNA(miRNA)、小干扰RNA(siRNA)、小细胞核RNA(snRNA)、小细胞核RNA(snoRNA)、双链RNA(dsRNA)、非编码RNA(ncRNA)、长非编码RNA(lncRNA)和小细胞质RNA(scRNA)。在一些优选的实施方式中,靶序列可以是选自由mRNA、前体mRNA和rRNA组成的组的RNA分子内的序列。在一些优选的实施方式中,靶序列可以是选自由ncRNA和lncRNA组成的组的RNA分子内的序列。在一些更优选的实施方式中,靶序列可以是mRNA分子或前体mRNA分子内的序列。

[0321] 在一些实施方式中,选择RNA靶向指导RNA或crRNA以降低RNA靶向指导RNA或crRNA内的二级结构程度。在一些实施方式中,当最佳折叠时,RNA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。最佳折叠可以通过任何合适的多核苷酸折叠算法来确定。一些程序是基于计算最小吉布斯自由能(Gibbs free energy)。一种此类算法的实例是如Zuker和Stiegler(Nucleic Acids Res.9(1981),133-148)所描述的mFold。另一个示例性折叠算法是维也纳大学(University of Vienna)的理论化学研究所开发的使用质心结构预测算法的在线网络服务器RNAfold(参见例如,A.R.Gruber等,2008,Cell 106(1):23-24;和PA Carr and GM Church,2009,Nature Biotechnology 27(12):1151-62)。

[0322] 在一些实施方式中,设计或选择靶向核酸的指导序列以调节指导分子之间,例如不同指导分子的茎环区域之间的分子间相互作用。应当理解,碱基对形成茎环的指导中的核苷酸也能够与第二个指导进行碱基配对以形成分子间双链体,并且这种分子间双链体将不具有与CRISPR复合物形成相容的二级结构。因此,对于选择或设计DR序列以调节茎环形成和CRISPR复合物形成是有用的。在一些实施方式中,大约或低于约75%、50%、40%、30%、25%、20%、15%、10%、5%、1%或更低的核酸靶向指导是在分子间双链体中。应当理解,茎-环变异通常将在 DR-CRISPR效应子相互作用所施加的限度内。调节茎环形成或改变茎环与分子间双链体之间平衡的一种方法是改变DR茎环的茎中的核苷酸对。例如,在一个实施方式中,G-C对被替代为A-U或U-A对。在另一个实施方式中,A-U对被替代为G-C或C-G对。在另一个实施方式中,天然存在的核苷酸被替代为核苷酸类似物。调节茎环形成或改变茎环与分子间双链体之间的平衡的另一种方法是修饰DR的茎环的环。不希望受到理论的束缚,该环可以看作是插入序列,其侧翼是两个彼此互补的序列。当该插入序列不是自身互补的时,其作用将是破坏分子间双链体的形成。当多重复用指导时,适用相同的原理:虽然靶向序列可能不同,但修改不同指导的DR中的茎环区域可能是有利的。此外,当对指导进行多重复用时,可以通过平衡每个指导的活动来调节不同指导的相对活动。在某些实施方式中,确定分子间茎环与分子间双链体之间的平衡。该确定可以通过物理或生化手段进行,并且可以在存在或不存在CRISPR效应子的情况下进行。

[0323] 在某些实施方式中,指导RNA或crRNA可以包含,基本上由正向重复(DR)序列和指导序列或间隔子序列组成或由其组成。在某些实施方式中,指导RNA或crRNA可以包含,基本上由或由与指导序列或间隔子序列融合或连接的正向重复序列组成。在某些实施方式中,正向重复序列可以位于指导序列或间隔区序列上游(即,5')。在其他实施方式中,正向重复序列可以位于指导序列或间隔区序列下游(即,3')。在其他实施方式中,可以存在多重DR(如双重DR)。

[0324] 在某些实施方式中,crRNA包含茎环,优选单个茎环。在某些实施方式中,正向重复序列形成茎环,优选单个茎环。

[0325] 在某些实施方式中,指导RNA的间隔子长度为15至35nt。在某些实施方式中,指导RNA的间隔区长度为至少15个核苷酸。在某些实施方式中,间隔区长度为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;27-30nt,例如27、28、29或30nt;30-35nt,例如30、31、32、33、34或35nt;或35nt或更长。

[0326] “tracrRNA”序列或类似术语包括与crRNA序列具有足够互补性以进行杂交的任何多核苷酸序列。在通常情况下,互补度是指沿着两个序列中较短的一个序列的sca序列和tracr序列的最佳比对。最佳比对可以通过任何合适的比对算法来确定,并且可以进一步考虑二级结构,例如sca序列或tracr序列内的自互补性。在一些实施方式中,最优比对时,tracr序列和sca序列沿着两者较短者的长度的互补程度约为或大于约25%、30%、40%、50%、60%、70%、80%、90%、95%、97.5%、99%或更高。在某些实施方式中,tracrRNA可能是不需要的。实际上,来自动物溃疡伯格菌(*Bergeyella zoohelcum*)及其直系同源物的CRISPR-Cas效应蛋白不需要tracrRNA来确保RNA靶标的切割。

[0327] 在进一步细节中,如下针对DNA靶进行测定。在此测定中使用两种大肠杆菌菌株。

一种携带编码来自细菌菌株的内源性效应蛋白座位的质粒。另一种菌株携带空质粒(例如 pACYC184, 对照菌株)。将所有可能的7或8bp PAM序列呈递在抗生素抗性质粒(具有氨苄青霉素抗性基因的pUC19)上。将PAM定位成靠近原型间隔区1的序列(内源性效应蛋白座位中的第一间隔区的DNA靶)。克隆了两个PAM文库。一个具有原型间隔区的8个随机bp 5' (例如总的65536个不同PAM序列=复杂度)。另一个文库具有原型间隔区的7个随机bp 3' (例如总复杂度是16384个不同的PAM)。将两个文库克隆成具有平均500个质粒/可能的PAM。用5' PAM和3' PAM文库在单独的转化中转化测试菌株和对照菌株并且将转化的细胞分别接种在氨苄青霉素板上。使用质粒的识别和随后的切割/干扰使得细胞对氨苄青霉素易感并且阻止了生长。转化后大约12h, 收获由测试菌株和对照菌株形成的所有群落并且分离出质粒 DNA。使用质粒DNA作为用于PCR扩增和随后的深度测序的模板。未转化的文库中的所有PAM的表现度显示转化细胞中的PAM的预期表现度。对照菌株中发现的所有PAM的表现度显示真实的表现度。测试菌株中的所有PAM的表现度显示哪个PAM未被酶识别并且与对照菌株的比较允许提取出缺失的PAM的序列。在特定实施方式中, 切割(如PNA 切割)不是PAM依赖性的。实际上, 对于动物溃疡伯格菌 (*Bergeyella zoohelcum*) Cas13b效应蛋白及其直系同源物, RNA靶标裂解似乎是PAM 非依赖性的, 因此本发明的表1Cas13b可以以PAM非依赖性的方式起作用。

[0328] 为了使毒性和脱靶作用最小化, 控制递送的靶向RNA的指导RNA 的浓度将很重要。可以通过在细胞或非人真核生物动物模型中测试不同的浓度并使用深度测序来分析潜在脱靶基因组基因座的修饰程度来确定 Cas mRNA或蛋白和指导RNA的最佳浓度。给出基于靶修饰的最高水平, 同时最小化的脱靶修饰的水平的浓度应被选择用于体内递送。RNA靶向系统有利地衍生自CRISPR-Cas系统。在一些实施方式中, RNA靶向系统的一个或多个元件来自特定生物, 该特定生物包括如本文所述的表1-4 Cas13效应蛋白系统的内源RNA靶向系统。

[0329] 死指导序列

[0330] 在一个方面中, 本发明提供了以允许形成CRISPR Cas复合物并成功与靶标结合的方式进行修饰的指导序列, 而同时不允许或不允许成功的核酸酶活性(即, 无核酸酶活性/没有indel活性)。为了说明的目的, 这种修饰的指导序列被称为“死指导”或“死指导序列”。就核酸酶活性而言, 可以认为这些死指导或死指导序列是催化失活的或构象失活的。实际上, 死指导序列就促进催化活性或区分靶上和靶外结合活性的能力而言可能不足以参与生产性碱基配对。简而言之, 该测定涉及合成CRISPR 靶RNA和包含与靶RNA错配的指导RNA, 将其与RNA靶向酶结合, 并基于裂解产物产生的条带的存在情况基于凝胶分析裂解, 并基于相对条带强度定量裂解。

[0331] 因此, 在相关方面, 本发明提供了非天然存在的或工程化组合物RNA 靶向CRISPR-Cas系统, 其包含本文所述的功能性RNA靶向酶, 以及指导RNA (gRNA) 或crRNA, 其中所述gRNA或crRNA包含死亡的RNA。指导序列, 其中gRNA能够与靶序列杂交, 从而将靶向CRISPR-Cas系统的RNA定向到细胞中的目标基因组位点, 而该系统的非突变RNA靶向酶没有可检测的RNA裂解活性。应当理解, 如本文其他地方所述, 根据本发明的任何gRNA或crRNA可以用作包含死指导序列的死亡 gRNA/crRNA。

[0332] 可以通过任何合适的测定来评估死指导序列指导CRISPR复合物与 RNA靶序列的序列特异性结合的能力。例如, 可将足以形成CRISPR-Cas 复合体的CRISPR-Cas系统组分

(包括待测试的死指导序列) 提供给具有相应靶序列的宿主细胞, 如通过用编码系统组成部分的载体转染, 然后评估靶序列内的优先切割。

[0333] 如本文进一步解释的, 几个结构参数允许适当的框架到达这样的死指导。死指导序列通常可以短于相应的指导序列, 从而导致活性RNA裂解。在特定实施方式中, 死指导比对应的相同指导短5%、10%、20%、30%、40%、50%或更短。

[0334] 如下述解释和本领域公知的, gRNA或crRNA的一个方面——RNA 靶向特异性是正向重复序列, 该序列应适当地连接至此类指导。特别是, 这意味着正向重复序列的设计取决于RNA靶向酶的来源。可用于验证的死指导序列的结构数据可用于设计CRISPR-Cas特异性等同物。例如, 两个或多个CRISPR-Cas效应蛋白的直向同源核酸酶结构域HEPN之间的结构相似性可用于转移设计等效的死指导。因此, 本文的死指导可以在长度和序列上进行适当修饰以反映这种CRISPR-Cas特异性等同物, 从而允许形成CRISPR-Cas复合物并成功结合靶RNA, 而同时不允许成功的核酸酶活性。

[0335] 死指导允许人们使用gRNA或crRNA作为基因靶向的手段, 而无核酸酶活性的结果, 同时提供了激活或抑制的直接手段。可以修饰包含死指导的指导RNA或crRNA, 使其以允许激活或抑制基因活性的方式进一步包括元件, 尤其是本文其他地方所述的允许基因效应物(例如, 基因活性的激活剂或阻遏物) 功能性定位的蛋白质衔接子(例如, 适体)。一个例子是掺入适体, 如本文和现有技术中所述。通过工程化包含死指导的gRNA或crRNA来掺入与蛋白质相互作用的适体(Konermann等, “Genome-scale transcription activation by an engineered CRISPR-Cas9 complex,” doi:10.1038/nature14136, 其通过引用并入本文), 可以组装多个不同的效应物结构域。其可以在自然过程之后进行建模。

[0336] 一般的CAS13

[0337] 本发明提供了特别的Cas13效应物、核酸、载体和使用方法。Cas13 的特征和功能也可以是本文所述的其他CRISPR-Cas蛋白的特征和功能。

[0338] 如本文所用, 术语Cas13b-s1辅助蛋白、Cas13b-s1蛋白质、Cas13b-s1、Csx27和Csx27蛋白可互换使用, 术语Cas13b-s2辅助蛋白、Cas13b-s2 蛋白、Cas13b-S2, Csx28和Csx28蛋白可以互换使用。

[0339] 在特定实施方式中, 野生型Cas13效应蛋白具有RNA结合和切割功能。

[0340] 在特定实施方式中, (野生型或突变的) Cas13效应蛋白可以具有 RNA和/或DNA切割功能, 优选RNA切割功能。在这些实施方式中, 可以基于本文提供的效应蛋白提供本发明的方法, 所述效应蛋白理解如本文所讨论的在真核细胞中(在体外, 即在分离的真核细胞中) 诱导一种或多种突变, 包括将本文所讨论的载体递送至细胞。所述突变可以包括通过指导RNA或sgRNA或crRNA在细胞的每个靶序列上引入、缺失或取代一个或多个核苷酸。所述突变可以包括通过指导RNA或sgRNA 或crRNA在细胞的每个靶序列上引入, 缺失或取代一个或多个核苷酸。所述突变可包括通过指导RNA或sgRNA或crRNA在所述细胞的每个靶序列上引入, 缺失或取代1-75个核苷酸。突变可以包括通过指导RNA 或sgRNA或crRNA在所述细胞的每个靶序列上引入, 缺失或取代1、5、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。突变可以包括通过指导RNA或sgRNA或crRNA在所述细胞的每个靶序列上引入, 缺失或取代5、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。突变可以包括通过指

导RNA或sgRNA或crRNA在所述细胞的每个靶序列上引入,缺失或取代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。突变可以包括通过指导RNA或sgRNA或crRNA在所述细胞的每个靶序列上引入,缺失或取代20、21、22、23、24、25、26、27、28、29、30、35、40、45、50或75个核苷酸。突变可以包括通过指导RNA或sgRNA或crRNA在所述细胞的每个靶序列上引入,缺失或取代40、45、50、75、100、200、300、400或500个核苷酸。

[0341] 为了使毒性和脱靶作用最小化,控制递送的Cas13 mRNA和指导RNA的浓度将很重要。可以通过在细胞或非人真核生物动物模型中测试不同的浓度并使用深度测序来分析潜在脱靶基因组基因座的修饰程度来确定Cas13 mRNA或蛋白和指导RNA的最佳浓度。使毒性和脱靶效应最小化的指导序列和策略可以在WO 2014/093622 (PCT/US2013/074667) 中;或,通过此处的突变。

[0342] 编码Cas13的核酸分子有利地是密码子优化的。密码子优化序列的一个例子,在这种情况下是为在真核生物例如人中表达而优化的序列(即,为在人中表达而优化),或为本文所讨论的另一种真核生物,动物或哺乳动物而优化的序列;参见例如,WO 2014/093622 (PCT/US2013/074667) 中的SaCas9人密码子优化序列。尽管这是优选的,但应理解,其他实例也是可能的,并且已知除人类以外的宿主物种的密码子优化或针对特定器官的密码子优化。在一些实施方式中,对编码Cas的酶编码序列进行密码子优化以在特定细胞如真核细胞中表达。真核细胞可以是诸如哺乳动物的特定生物的细胞或衍生自其的细胞,所述哺乳动物包括但不限于本文所述的人或非人真核生物或动物或哺乳动物,例如小鼠,大鼠,兔,狗,家畜或非人类的哺乳动物或灵长类动物。在一些实施方式中,可能不包括用于修饰人类种系遗传特性的过程和/或用于修饰可能导致人类遭受痛苦而对人或动物没有任何实质性医学益处的动物的过程,以及由此过程产生的动物。在通常情况下,密码子优化是指通过替换至少一个密码子(例如约或大于约1、2、3、4、5、10、15、20、25、50或50个以上的密码子)和在该宿主细胞的基因中更常用或最常用的密码子,同时又保留了天然氨基酸序列。各种物种对特定氨基酸的某些密码子表现出特定的偏好。密码子偏好(生物体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翻译的密码子的性质和特定密码子。转移RNA(tRNA)分子的可用性。所选tRNA在细胞中的优势通常反映了肽合成中最常使用的密码子。所选tRNA在细胞中的优势通常反映了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最佳基因表达。密码子使用情况表是易于获得的,例如,可在[www.kazusa.or.jp/codon/](http://www.kazusa.or.jp/codon/)中获得“密码子使用数据库”,这些表可以通过多种方式进行修改。参见,Nakamura, Y.等,“Codon usage tabulated from the international DNA sequence databases:status for the year 2000”Nucl. Acids Res. 28:292 (2000)。还可以使用用于优化特定序列以在特定宿主细胞中表达的密码子的计算机算法,例如Gene Forge (Aptagen; Jacobus, PA)。在一些实施方式中,编码Cas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或更多或所有密码子)对应于特定氨基酸最常用的密码子。

[0343] 在一些实施方式中,Cas13可以直接在靶序列的位置或附近(例如,在靶序列和/或内部)切割一条或两条核酸链,如在靶序列的互补序列之内或在与靶序列相关的序列之内,例如,从目标序列的第一个或最后一个核苷酸开始在大约1、2、3、4、5、6、7、8、9、10、15、20、

25、50、100、200、500或更多个碱基对之内。在一些实施方式中,Cas13蛋白可指导靶序列内和/或靶序列的互补序列内或与靶序列相关的序列中的一条或两条链的不止一个切割(例如一,二,三,四,五或更多个切割)。靶序列和/或距第一个或最后一个碱基对约1、2、3、4、5、6、7、8、9、10、15、20、25、50、100、200、500或更多个碱基对内靶序列的核苷酸。在一些实施方式中,切割可以是钝的,即,产生钝末端。在一些实施方式中,切割可以是错开的,即产生粘性末端。。在一些实施方式中,载体编码可以相对于相应的野生型酶突变的靶向核酸的Cas13 蛋白,使得突变的靶向核酸的Cas13蛋白缺乏切割含有靶序列的靶多核苷酸的一条或两条链的能力,例如,在HEPN结构域中的改变或突变以产生基本上没有所有RNA切割活性的突变Cas13,例如,该突变酶的RNA 切割活性大约不超过25%、10%、5%、1%、0.1%、0.01%或以下非突变形式的酶的核酸裂解活性;一个例子可以是当突变形式的核酸切割活性与非突变形式相比为零或可忽略时。所谓衍生,申请人是指衍生的酶在很大程度上与野生型酶具有高度的序列同源性,但是它已经以本领域已知或本文描述的某种方式进行了突变(修饰)。

[0344] 通常,在内源性RNA靶向系统的情况下,RNA靶向复合物(包含与靶序列杂交并与一种或多种RNA靶向效应蛋白复合的指导RNA或 crRNA)的形成导致RNA链断裂(一个或多个)在靶序列中或其附近(例如,在1、2、3、4、5、6、7、8、9、10、20、50个或更多个碱基对之内)。如本文所用,术语“与感兴趣的靶基因座相关的序列”是指靠近靶序列附近的序列(例如在1、2、3、4、5、6、7、8、9、10、20、50或更多个来自靶序列的碱基对,其中靶序列包含在感兴趣的靶基因座内)。

[0345] 密码子优化序列的一个例子,在这种情况下是为在真核生物例如人中表达而优化的序列(即,为在人中表达而优化),或为本文所讨论的另一种真核生物,动物或哺乳动物而优化的序列;参见例如,WO 2014/093622 (PCT/US2013/074667) 中的SaCas9人密码子优化序列(根据本领域的知识和本公开内容,密码子优化编码核酸分子,尤其是关于效应子蛋白(例如,Cas13)的密码子优化在本领域技术人员的能力范围内)。尽管这是优选的,但应理解,其他实例也是可能的,并且已知除人类以外的宿主物种的密码子优化或针对特定器官的密码子优化。在一些实施方式中,对编码靶向RNA的Cas13蛋白的酶编码序列进行密码子优化以在特定细胞如真核细胞中表达。真核细胞可以是诸如哺乳动物的特定生物的细胞或衍生自其的细胞,所述哺乳动物包括但不限于本文所述的人或非人真核生物或动物或哺乳动物,例如小鼠,大鼠,兔,狗,家畜或非人类哺乳动物或灵长类动物。在一些实施方式中,可能不包括用于修饰人类种系遗传特性的过程和/或用于修饰可能导致人类遭受痛苦而对人或动物没有任何实质性医学益处的动物的过程,以及由此过程产生的动物。在通常情况下,密码子优化是指通过置换至少一个密码子(例如,约或大于约1、2、3、4、5、10、15、20、25、50或更多个密码子)和在该宿主细胞的基因中更常用或最常用的密码子,同时又保持了天然氨基酸序列。各种物种对特定氨基酸的某些密码子表现出特定的偏好。密码子偏好(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翻译的密码子的性质和特定密码子的转移RNA(tRNA)分子可用性。所选tRNA在细胞中的优势通常反映了肽合成中最常使用的密码子。所选 tRNA在细胞中的优势通常反映了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最佳基因表达。密码子使用情况表是易于获得的,例如,可在[www.kazusa.or.jp/](http://www.kazusa.or.jp/)

codon/中获得“密码子使用数据库”，这些表可以通过多种方式进行修改。参见，Nakamura, Y. 等, “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” Nucl. Acids Res. 28:292 (2000)。还可以使用用于优化特定序列以在特定宿主细胞中表达的密码子的计算机算法, 例如 Gene Forge (Aptagen; Jacobus, PA)。在一些实施方式中, 编码DNA/RNA 靶向Cas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或更多或所有密码子)对应于特定氨基酸最常用的密码子。

[0346] (i) Cas13或编码它的核酸分子或(ii) crRNA可以分开递送;且有利地, (i) 和(ii)之一, 例如组装的复合物, 是通过颗粒或纳米颗粒复合物递送的。可以在靶向RNA的指导RNA或crRNA之前递送靶向 RNA的效应蛋白mRNA, 以留出时间表达靶向核酸的效应蛋白。靶向RNA的效应蛋白(Cas13)mRNA可以在施用RNA靶向的指导RNA或 crRNA之前1-12小时(最好在2-6小时左右)施用。或者, 可以一起施用靶向RNA的效应蛋白mRNA和靶向RNA的指导RNA或crRNA。有利地, 可以在初次施用靶向RNA的效应子(Cas13)蛋白mRNA+指导 RNA后1-12小时(优选约2-6小时)施用指导RNA或crRNA的第二加强剂量。靶向RNA的效应蛋白mRNA和/或指导RNA或crRNA的其他给药可能对获得最有效的基因组修饰水平有用。

[0347] 在一个方面中, 本发明提供了使用RNA靶向系统的一个或多个元件的方法。本发明的靶向RNA的复合物提供了修饰单链或双链, 线性或超螺旋靶RNA的有效手段。本发明的靶向RNA的复合物具有多种用途, 包括修饰(例如, 缺失, 插入, 转运, 失活, 活化)多种细胞类型的靶 RNA。因此, 本发明的靶向RNA的复合物在例如基因治疗, 药物筛选, 疾病诊断和预后方面具有广泛的应用。示例性的RNA靶向复合物包括与感兴趣的靶基因座内的靶序列杂交的指导RNA或crRNA复合的RNA 靶向效应蛋白。

[0348] 在一个实施方式中, 本发明提供了一种切割靶RNA的方法。该方法可以包括使用与靶RNA结合并实现所述靶RNA的切割的RNA靶向复合物来修饰靶RNA。在一个实施方式中, 本发明的靶向RNA的复合物被引入细胞中时, 可在RNA序列中产生断裂(例如, 单链断裂或双链断裂)。例如, 该方法可用于裂解细胞中的疾病RNA。例如, 可以将包含要被整合的序列侧接上游序列和下游序列的外源RNA模板引入细胞中。上游和下游序列与RNA中整合位点的任一侧共享序列相似性。需要时, 供体 RNA可以是mRNA。外源RNA模板包含要整合的序列(例如, 突变的 RNA)。整合序列可以是细胞内源或外源的序列。待整合的序列的实例包括编码蛋白质的RNA或非编码RNA(例如, 微小RNA)。因此, 用于整合的序列可以可操作地连接至合适的一个或多个控制序列。备选地, 要整合的序列可以提供调节功能。选择外源RNA模板中的上游和下游序列以促进目的RNA序列和供体RNA之间的重组。上游序列是与用于整合的靶位点上游的RNA序列具有序列相似性的RNA序列。类似地, 下游序列是与整合靶位点下游的RNA序列具有序列相似性的RNA序列。外源RNA模板中的上游和下游序列可以与目标RNA序列具有75%, 80%, 85%, 90%, 95%, 或100%的序列同一性。优选地, 外源RNA模板中的上游和下游序列与靶RNA序列具有约95%, 96%, 97%, 98%, 99%, 或100%的序列同一性。在一些方法中, 外源RNA模板中的上游和下游序列与靶RNA序列具有约99%或100%的序列同一性。上游或下游序列可包含约20bp至约2500bp, 例如, 约50、100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500, 1600、1700、1800、1900、2000、2100、2200、2300、2400或2500bp。在一些方法中, 示例性的上游或下游序列具有约200bp至约2000bp, 约600bp至约1000bp, 或更特别地约700bp至约1000bp。在一些方法中, 外源RNA模板可以进一

步包含标记。这样的标记可以使针对目标集成的筛选变得容易。合适的标记的例子包括限制性位点, 荧光蛋白或选择标记。可以使用重组技术构建本发明的外源RNA模板(参见, 例如, Sambrook等, 2001和Ausubel等, 1996)。在通过整合外源RNA模板来修饰靶RNA的方法中, 通过核酸靶向复合物将断裂(例如, 双链或单链RNA中的双链或单链断裂)引入RNA序列, 该断裂被修复。通过与外源RNA模板的同源重组, 从而将模板整合到RNA靶标中。双链断裂的存在促进了模板的整合。在其他实施方式中, 本发明提供了一种修饰RNA在真核细胞中表达的方法。该方法包括通过使用与DNA或RNA(例如, mRNA或pre-mRNA)结合的核酸靶向复合物来增加或减少靶多核苷酸的表达。在一些方法中, 可以使靶RNA失活以影响细胞中表达的修饰。例如, 在RNA靶向复合物与细胞中的靶序列结合后, 靶RNA被灭活, 使得该序列不被翻译, 不产生编码蛋白或该序列不作为野生型起作用。顺序呢。例如, 可以使蛋白质或微RNA编码序列失活, 从而不产生蛋白质或微RNA或前微RNA转录物。RNA靶向复合物的靶RNA可以是真核细胞内源性或外源性的任何RNA。例如, 靶RNA可以是存在于真核细胞核中的RNA。靶RNA可以是编码基因产物(例如蛋白质)的序列(例如mRNA或mRNA前体)或非编码序列(例如ncRNA, lncRNA, tRNA或rRNA)。靶RNA的例子包括与信号生化途径相关的序列, 例如与信号生化途径相关的RNA。靶RNA的实例包括疾病相关RNA。“与疾病相关的”RNA是指与受感染的组织或组织相比, 在受疾病影响的组织得到的细胞中, 以异常水平或异常形式产生翻译产物的任何RNA。

[0349] 在一些实施方式中, 该方法可以包括使靶向RNA的复合物与靶RNA结合以实现所述靶RNA的切割, 从而修饰靶RNA, 其中所述靶向RNA的复合物包含靶向核酸的效应子(Cas13与指导RNA或crRNA复合的蛋白质, 所述指导RNA或crRNA与所述靶RNA内的靶序列杂交。在一个方面中, 本发明提供了一种在真核细胞中修饰RNA表达的方法。在一些实施方式中, 该方法包括使靶向RNA的复合物结合至RNA, 使得所述结合导致所述RNA的表达增加或减少。其中靶向RNA的复合物包含与指导RNA复合的核酸靶向效应子(Cas13)蛋白。修饰靶RNA的方法可以在真核细胞中, 其可以在体内, 离体或体外。在一些实施方式中, 该方法包括从人或非人动物中取样细胞或细胞群, 并修饰该细胞。培养可以在离体的任何阶段进行。所述一种或多种细胞甚至可以重新引入非人类动物或植物中。对于重新导入的细胞, 特别优选的是细胞是干细胞。

[0350] 两种不同适体的使用(每个均与不同的靶向RNA的指导RNA相关)允许使用激活剂-适配器蛋白融合体和阻抑剂-蛋白融合体, 并使用不同的靶向RNA的指导RNA或crRNA激活蛋白的表达。RNA, 同时抑制另一个。它们以及它们不同的指导RNA或crRNA可以以多重方式一起或基本上一起施用。可以同时全部使用大量此类修饰的靶向RNA的指导RNA或crRNA, 例如10或20或30等, 而效应蛋白(Cas13)仅一种(或最少数量)由于递送相对较少数量的效应蛋白分子可与大量修饰的引导物一起使用, 因此需要递送这些分子。衔接子蛋白可以与一种或多种激活剂或一种或多种抑制子缔合(优选连接或融合)。例如, 衔接蛋白可以与第一活化剂和第二活化剂相关。第一和第二活化剂可以相同, 但是优选它们是不同的活化剂。可以使用三个或更多或甚至四个或更多的活化剂(或阻遏物), 但包装尺寸可能会限制其数量高于5个不同的功能域。优选在与衔接蛋白直接融合的情况下使用接头, 其中两个或更多个功能域与衔接蛋白缔合。合适的链接器可能包括GlySer链接器。

[0351] 还可以设想, 靶向RNA的效应蛋白指导RNA复合物整体上可以与两个或更多个功能

域结合。例如,可能有两个或多个与靶向RNA的效应子蛋白相关的功能域,或者可能有两个或多个与指导RNA或crRNA相关的功能域(通过一个或多个衔接子蛋白),或者可能有一个或多个与靶向RNA的效应子蛋白相关的更多功能域和与指导RNA或crRNA相关的一个或多个功能域(通过一种或多种衔接蛋白)。

[0352] 衔接子蛋白与活化剂或阻遏物之间的融合体可包括接头。例如,可以使用GlySer接头GGGS。它们可以以3((GGGS)3 (SEQ ID NO: 79))或6、9甚至12或更多的重复序列使用,以根据需要提供合适的长度。可以在指导RNA与功能域(激活物或阻遏物)之间,或在靶向核酸的效应蛋白与功能域(激活物或阻遏物)之间使用接头。链接器使用户可以设计适当数量的“机械灵活性”。

[0353] CRISPR效应子(Cas13)蛋白或其mRNA(或更普遍地为其核酸分子)和指导RNA或crRNA也可以分开给药,例如在给药前1-12小时(最好在2-6小时左右)指导RNA或crRNA或一起引导。初次给药后1-12小时(最好在2-6小时左右)可以再给予第二次指导RNA或crRNA加强剂量。

[0354] Cas13效应蛋白有时在本文中称为CRISPR酶。应当理解,效应蛋白是基于酶或衍生自酶,因此在一些实施方式中,术语“效应蛋白”当然包括“酶”。然而,还应理解,效应蛋白可以按要求以某种方式具有DNA或RNA结合活性,但不一定具有切割或切刻的活性,包括死Cas效应蛋白的功能。

[0355] 细胞靶标包括造血干/祖细胞(CD34+);人T细胞;和眼睛(视网膜细胞)-例如,感光细胞前体细胞。

[0356] 本发明的方法可以进一步包括模板的递送。模板的递送可以通过同时或与任何或所有CRISPR效应蛋白(Cas13)或向导或crRNA的递送同时进行,或者可以通过相同或不同的递送机制进行。

[0357] 在某些实施方式中,本文描述的方法可以包括提供一种Cas13转基因细胞,其中提供或引入编码一种或多种指导RNA的一种或多种核酸,并在该细胞中与包含一个或多个启动子的调控元件可操作地连接。或更多感兴趣的基因。如本文所用,术语“Cas13转基因细胞”是指已在基因组上整合了Cas13基因的细胞,例如真核细胞。根据本发明,细胞的性质,类型或来源没有特别限制。同样,Cas13转基因导入细胞的方式可以改变,并且可以是本领域已知的任何方法。在某些实施方式中,通过将Cas13转基因导入分离的细胞中获得Cas13转基因细胞。在某些其他实施方式中,通过从Cas13转基因生物中分离细胞来获得Cas13转基因细胞。举例来说,但不限于,本文所指的Cas13转基因细胞可以衍生自Cas13转基因真核生物,例如Cas13敲入真核生物。参考WO 2014/093622 (PCT/US13/74667),其通过引用并入本文。可以修改转让给Sangamo BioSciences, Inc的,靶向Rosa基因座的美国专利公开号20120017290和20110265198的方法,以利用本发明的CRISPR Cas系统。也可以修改转让给Cellestis的,旨在靶向Rosa基因座的美国专利公开号20130236946的方法,以利用本发明的CRISPR Cas系统。通过进一步的示例,参考Platt等人。等(Cell;159(2):440-455(2014)),描述了Cas9敲入小鼠,其通过引用并入本文。Cas13转基因可进一步包含Lox-Stop-polyA-Lox (LSL)盒,从而使得可通过Cre重组酶诱导Cas13表达。或者,可以通过将Cas13转基因引入分离的细胞中来获得Cas13转基因细胞。转基因的递送系统是本领域众所周知的。举例来说,Cas13转基因可通过载体(例如,AAV,腺病毒,慢病毒)和/或颗粒和/

或颗粒递送而在例如真核细胞中递送,如本文其他地方所述。

[0358] 本领域技术人员将理解,如本文所指的细胞,例如Cas13转基因细胞,除了具有整合的Cas13基因或当与具有RNA功能的RNA复合时由 Cas13的序列特异性作用引起的突变之外,还可包含其他基因组改变。将 Cas13引导至靶基因座的过程,例如一种或多种致癌突变,例如但不限于在Platt等人(2014),Chen等人(2014)或Kumar等人(2009)中描述的。

[0359] 在一些实施方式中,Cas13序列与一个或多个核定位序列(NLS)融合,例如约或大于约1、2、3、4、5、6、7、8、9、10或在一些实施方式中,Cas13在氨基末端处或附近,在大约或附近包含约或大于约1、2、3、4、5、6、7、8、9、10或更多个NLS。羧基末端或附近的NLS数量超过约1、2、3、4、5、6、7、8、9、10或更多,或这些的组合(例如零或至少一个或多个NLS)氨基末端,在羧基末端为零或一个或多个NLS)。当存在一个以上的NLS时,可以彼此独立地进行选择,以使单个NLS可以一个以上的副本存在和/或与一个或多个其他NLS一起存在一个或多个副本中。在本发明的一个优选实施方式中,Cas13包含至多6个NLS。在一些实施方式中,当NLS的最接近的氨基酸在约1、2、3内时,NLS被认为在N端或C端附近。N或C端沿着多肽链的4、5、10、15、20、25、30、40、50或更多个氨基酸。NLS的非限制性实例包括源自以下的NLS序列:SV40病毒大T抗原的NLS,其具有氨基酸序列PKKKRKV(SEQ ID NO:80);和来自核纤溶酶的NLS(例如具有序列KRPAATKKAGQAKKKK的核纤溶蛋白二分NLS)(SEQ ID NO:81);具有氨基酸序列PAAKRVKLD(SEQ ID NO:82)或RQRRNELKRSP(SEQ ID NO:83)的c-myc NLS;具有序列NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY(SEQ ID NO:84)的hRNPA1 M9 NLS;IBB结构域的Importin- $\alpha$ 序列RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV(SEQ ID NO:85);肌瘤T蛋白的序列VSRKRPRP(SEQ ID NO:86)和PPKKARED(SEQ ID NO:87);人p53的序列POPKKKPL(SEQ ID NO:88);小鼠c-ab1 IV的序列SALIKKKKKMAP(SEQ ID NO:89);流感病毒NS1的DRLRR序列(SEQ ID NO:90)和PKQKKRK序列(SEQ ID NO:91)。肝炎病毒 $\delta$ 抗原的序列RKLKKKIKKL(SEQ ID NO:92);小鼠Mx1蛋白的序列REKKKFLKRR(SEQ ID NO:93);人聚(ADP-核糖)聚合酶的序列KRKGDEVGDGVDEVAKKKSKK(SEQ ID NO:94);类固醇激素受体(人)糖皮质激素的序列RKCLQAGMNLEARKTKK(SEQ ID NO:95)。在通常情况下,一个或多个NLS具有足够的强度来驱动Cas在真核细胞核中的积聚。在通常情况下,核定位活动的强度可能来自Cas中NLS的数量,所使用的特定NLS或这些因素的组合。核中积累的检测可以通过任何合适的技术来进行。例如,可将可检测的标记融合到Cas上,从而可以将细胞内的位置可视化,例如结合用于检测细胞核位置的手段(例如,对细胞核特异的染色剂,例如DAPI)。也可以从细胞中分离细胞核,然后通过用于检测蛋白质的任何合适的方法,例如免疫组织化学,蛋白质印迹或酶活性测定来分析其内容。也可以间接确定细胞核中的蓄积,例如通过测定CRISPR复合物形成的作用(例如测定靶序列上的DNA裂解或突变,或测定受CRISPR复合物形成和/或影响的基因表达活性的改变)与未暴露于Cas或复合物或未暴露于缺少一种或多种NLS的对照的对照相比)。

[0360] 指导RNA,例如sgRNA或crRNA编码序列和/或Cas13编码序列,可以在功能上或操作上与调节元件连接,因此调节元件驱动表达。启动子可以是组成型启动子和/或条件启动子和/或诱导型启动子和/或组织特异性启动子。启动子可以选自RNA聚合酶,pol I,pol II,pol III,T7,U6,H1,逆转录病毒劳斯肉瘤病毒(RSV)LTR启动子,巨细胞病毒(CMV)启动子,SV40启动子,二氢叶酸还原酶启动子, $\beta$ -肌动蛋白启动子,磷酸甘油激酶(PGK)启动子和EF1

$\alpha$ 启动子。有利的启动子是启动子是 U6。

[0361] 在一些实施方式中,CRISPR效应子(Cas 13n)蛋白可能形成诱导系统的组成部分。该系统的可诱导性质将允许使用一种能量来时空控制基因编辑或基因表达。能量的形式可以包括但不限于电磁辐射,声能,化学能和热能。诱导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)或光诱导系统(植物色素,LOV结构域或隐色素)。一个实施方式中,CRISPR效应蛋白可以是光诱导转录效应子(LITE)的一部分,以序列特异性方式指导转录活性的变化。光的成分可以包括CRISPR效应蛋白,光响应性细胞色素异二聚体(例如来自拟南芥)和转录激活/抑制域。在US 61/736465和US 61/721,283以及WO 2014018423 A2中提供了可诱导的 DNA结合蛋白的进一步实例及其使用方法,其通过引用整体并入本文。

[0362] 每当在本文中提及Cas13时,将理解的是,除非另外明确指出,否则本文所描述的根据本发明的突变的Cas13是指。除非另外明确指出,否则本文中每当提及Cas13时,均意指本文所述的根据本发明的突变的 Cas13a,Cas13b,Cas13c或Cas13d。除非另外明确指出,否则本文中每当提及Cas13时,优选是指本文所述的根据本发明的突变的Cas13b。

[0363] 在一个方面中,本发明提供了如本文所述的突变的Cas13,例如优选但不限于如本文其他地方所述的Cas13b,其具有一种或多种导致减少脱靶效应的突变,即用于实现修饰的改进的CRISPR酶。靶向位点,但降低或消除针对脱靶的活性,例如与指导RNA结合时,以及改进的CRISPR 酶,例如与指导RNA结合时,可提高CRISPR酶的活性。应当理解,本文下文所述的突变酶可用于本文其他地方所述的根据本发明的任何方法。如本文其他地方所述的任何方法,产品,组合物和用途均与突变的 CRISPR酶同样适用,如下文进一步详述。

[0364] Slaymaker等,最近,他描述了一种产生具有增强的特异性的Cas9 直向同源物的方法(Slaymaker等,2015“Rationally engineered Cas9 nucleases with improved specificity”)。该策略可用于增强Cas13蛋白的特异性。用于诱变的主要残基优选是HEPN结构域内的所有正电荷残基。另外的残基是在不同直系同源物之间保守的带正电荷的残基。

[0365] 一方面,本发明还提供了用于调节Cas13结合活性和/或结合特异性的方法和突变。在某些实施方式中,使用缺乏核酸酶活性的Cas13蛋白。在某些实施方式中,采用修饰的指导RNA来促进Cas13核酸酶的结合但不促进核酸酶活性。在这样的实施方式中,可以增加或减少靶上结合。同样,在这样的实施方式中,可以增加或减少脱靶结合。此外,关于靶上结合与脱靶结合的特异性可以增加或降低。

[0366] 可以以各种组合使用的方法和突变,以增加或减少靶上与脱靶活性的活性和/或特异性,或增加或减少靶上与脱靶结合的结合和/或特异性。可用于补偿或增强为促进其他作用而进行的突变或修饰。为促进其他作用而进行的此类突变或修饰包括对Cas13的突变或修饰和/或对指导RNA 的突变或修饰。本发明的方法和突变用于调节Cas13核酸酶活性和/或与化学修饰的指导RNA结合。

[0367] 一方面,本发明提供了用于调节如本文所定义的本发明的Cas13蛋白的结合和/或结合特异性的方法和突变,其包括功能域,例如核酸酶,转录激活因子,转录阻遏物等。例如,可以通过引入突变,例如本文其他地方所述的Cas13突变,使Cas13蛋白变为无核酸酶,或具有改变或降低的核酸酶活性。核酸酶缺陷的Cas13蛋白可用于RNA引导的靶序列依赖的

功能域的传递。本发明提供了用于调节Cas13蛋白的结合的方法和突变。在一个实施方式中,功能结构域包含VP64,提供RNA指导的转录因子。在另一个实施方式中,功能域包含Fok I,提供RNA引导的核酸酶活性。提到了美国专利公开号2014/0356959,美国专利公开号2014/0342456,美国专利公开号2015/0031132和Mali,P.等2013,Science 339(6121):823-6,doi:10.1126/science.1232033,在线发表于2013年1月3日,并且通过本文的教导,本发明理解了以下方法和材料这些文件结合本文的教导而应用。在某些实施方式中,靶上结合增加。在某些实施方式中,脱靶结合减少。在某些实施方式中,靶上结合减少。在某些实施方式中,靶上结合增加。因此,本发明还提供增加或降低功能化Cas13结合蛋白的靶上结合vs脱靶结合的特异性。

[0368] Cas13作为RNA指导的结合蛋白的用途不限于无核酸酶的Ca13。当与某些指导RNA一起使用时,包含核酸酶活性的Cas13酶也可以充当RNA引导的结合蛋白。例如,短指导RNA和包含与靶错配的核苷酸的指导RNA可以促进RNA指导的Cas13结合至靶序列,而很少或没有靶裂解。(参见例如Dahlman,2015,Nat Biotechnol.33(11):1159-1161,doi:10.1038/nbt.3390,2015年10月5日在线发布)。一方面,本发明提供了调节具有核酸酶活性的Cas13蛋白结合的方法和突变。在某些实施方式中,靶上结合增加。在某些实施方式中,脱靶结合减少。在某些实施方式中,脱靶结合减少。在某些实施方式中,脱靶结合增加。在某些实施方式中,靶上结合相对于脱靶结合的特异性增加或降低。在某些实施方式中,指导RNA-Cas13酶的核酸酶活性也被调节。

[0369] RNA-RNA双链体的形成对于整个靶区域的裂解活性和特异性很重要,不仅是最接近PAM的种子区域序列。因此,截短的指导RNA显示降低的切割活性和特异性。一方面,本发明提供了使用改变的指导RNA来增加切割的活性和特异性的方法和突变。

[0370] 在某些实施方式中,本发明的CRISPR-Cas蛋白(例如Cas13)的催化活性被改变或修饰。应当理解,如果催化活性不同于相应的野生型CRISPR-Cas蛋白(例如,未突变的CRISPR-Cas蛋白)的催化活性,则突变的Cas13具有改变的或修饰的催化活性。催化活性可以通过本领域已知的方法来确定。举例来说,但不限于,可以通过测定插入缺失百分比(例如在给定时间后或以给定剂量后)在体外或体内测定催化活性。在某些实施方式中,催化活性提高至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或至少100%。在某些实施方式中,催化活性降低。在某些实施方式中,催化活性至少降低5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或(基本上)100%。本文的一个或多个突变可以使催化活性失活,所述催化活性可以基本上是所有可催化的活性,低于可检测的水平,或者没有可测量的催化活性。

[0371] 经改造的CRISPR-Cas蛋白的一个或多个特征可以与相应的野生型CRISPR-Cas蛋白不同。此类特征的示例包括催化活性,gRNA结合,CRISPR-Cas蛋白的特异性(例如,编辑定义的靶标的特异性),CRISPR-Cas蛋白的稳定性,脱靶结合,靶标结合,蛋白酶活性,切口酶活性,PFS识别。在一些实例中,一种工程CRISPR-Cas蛋白可能包含一个或多个相应野生型CRISPR-Cas蛋白的突变。在一些实施方式中,该工程CRISPR-Cas蛋白的催化活性比相应的野生型CRISPR-Cas蛋白高。野生型CRISPR-Cas蛋白。在一些实施方式中,与相应的野生型CRISPR-Cas蛋白相比,工程CRISPR-Cas蛋白的催化活性降低。在一些实施方式中,工程

CRISPR-Cas的gRNA结合。与相应的野生型CRISPR-Cas蛋白相比,该蛋白增加了。在一些实施方式中,与相应的野生型CRISPR-Cas蛋白相比,工程化CRISPR-Cas蛋白的gRNA结合降低了。与相应的野生型 CRISPR-Cas蛋白相比,CRISPR-Cas蛋白的特异性增加。在一些实施方式中,与相应的野生型CRISPR-Cas蛋白相比,CRISPR-Cas蛋白的特异性降低。方式中,与在 一些实施方式中,工程CRISPR-Cas蛋白还包含一种或两种以上的相应的野生型CRISPR-Cas蛋白。与一些相应的野生型 CRISPR-Cas蛋白相比,CRISPR-Cas蛋白的稳定性降低。在一些实施方式中,与靶点结合的脱靶结合增加。与一些野生型CRISPR-Cas蛋白相比, CRISPR-Cas的脱靶结合增加。与相应的野生型CRISPR-Cas蛋白相比, CRISPR-Cas蛋白降低了。在一些实施方式中,与相应的野生型 CRISPR-Cas蛋白相比,CRISPR-Cas蛋白的靶结合增加了。因此,与相应的野生型CRISPR-Cas蛋白相比,CRISPR-Cas蛋白的靶结合度降低了。在一些实施方式中,与相应的野生型相比,改造后的CRISPR-Cas蛋白具有更高的蛋白酶活性或多核苷酸结合能力。CRISPR-Cas蛋白。在一些实施方式中,PFS识别是可变的与相应的野生型CRISPR-Cas蛋白相比呈红色。

[0372] 在某些实施方式中,本发明的Cas13蛋白的gRNA (crRNA) 结合被改变或修饰。应当理解,如果gRNA结合不同于相应的野生型Cas13的 gRNA结合(即未突变的Cas13),则突变的Cas13具有改变的或修饰的 gRNA结合。gRNA结合可以通过本领域已知的方法来确定。举例来说,但不限于,可以通过计算结合强度或亲和力(例如基于平衡常数,  $K_a$ ,  $K_d$ 等)来确定gRNA结合。在某些实施方式中,gRNA结合增加。实施方式中,gRNA结合增加至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或至少100%。在某些实施方式中,gRNA结合降低。在某些实施方式中,gRNA结合降低至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或(基本上)100%。

[0373] 在某些实施方式中,本发明的Cas13蛋白的特异性被改变或修饰。应当理解,如果特异性不同于相应的野生型Cas13(即未突变的Cas13) 的特异性,则突变的Cas13具有改变的或修饰的特异性。特异性可以通过本领域已知的方法来确定。举例来说,但不限于,可以通过比较靶上活性和靶外活性来确定特异性。在某些实施方式中,特异性提高。在某些实施方式中,特异性提高至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或至少100%。在某些实施方式中,特异性降低至少5%,优选至少10%,更优选至少20%。至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或(基本上)100%。

[0374] 在某些实施方式中,本发明的Cas13蛋白的稳定性被改变或修饰。应当理解,如果稳定性不同于相应的野生型Cas13(即未突变的Cas13) 的稳定性,则突变的Cas13具有改变或修饰的稳定性。稳定性可以通过本领域已知的方法来确定。举例来说,但不限于,可以通过确定Cas13 蛋白的半衰期来确定稳定性。在某些实施方式中,稳定性提高。在某些实施方式中,稳定性提高至少5%。优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或至少100%。在某些实施方式中,稳定性降低至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或(基本上)100%。

[0375] 在某些实施方式中,本发明的Cas13蛋白的靶结合被改变或修饰。应当理解,如果

靶结合不同于相应的野生型Cas13的靶结合(即未突变的Cas13),则突变的Cas13具有改变的或修饰的靶结合。靶结合可以通过本领域已知的方法来确定。举例来说,但不限于,可以通过计算结合强度或亲和力(例如基于平衡常数,  $K_a$ ,  $K_d$ 等)来确定靶结合。在某些实施方式中,靶结合增加。方式中,靶结合增加至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%或至少100%。在某些实施方式中,目标结合减少。在某些实施方式中,目标结合减少至少5%,最好在至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或(基本上)100%。

[0376] 在某些实施方式中,本发明的Cas13蛋白的脱靶结合被改变或修饰。应当理解,如果脱靶结合不同于相应的野生型Cas13的脱靶结合(即未突变的Cas13),则突变的Cas13具有改变或修饰的脱靶结合。脱靶结合可以通过本领域已知的方法确定。举例来说,但不限于,可以通过计算结合强度或亲和力(例如基于平衡常数,  $K_a$ ,  $K_d$ 等)来确定脱靶结合。在某些实施方式中,脱靶结合增加。在某些实施方式中,脱靶结合增加至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%或至少100%。在某些实施方式中,脱靶结合减少。在某些实施方式中,脱靶结合减少至少5%,优选至少10%,更优选至少20%,例如至少30%,至少40%,至少50%,至少60%,至少70%,至少80%,至少90%,或(基本上)100%。

[0377] 在某些实施方式中,本发明的Cas13蛋白的PFS(或PAM)识别或特异性被改变或修饰。应当理解,如果PFS识别或特异性不同于相应的野生型Cas13(即未突变的Cas13)的PFS识别或特异性,则突变的Cas13具有改变或修饰的PFS识别或特异性。PFS的识别或特异性可以通过本领域已知的方法来确定。举例来说,但不限于,可以通过PFS(PAM)筛查确定PFS的识别或特异性。在某些实施方式中,至少一种不同的PFS被Cas13识别。在某种实施方式中,除突变型Cas13识别出一个PFS外,野生型Cas13也不能识别出一个PFS。在某些实施方式中,至少一种PFS被突变的Cas13识别,而相应的野生型Cas13无法识别,并且不再识别野生型PFS。被突变的Cas13识别的PFS比野生型Cas13识别的PFS更长,例如长1、2或3个核苷酸。在某些实施方式中,被突变的Cas13识别的PFS比被识别的PFS短。由野生型Cas13组成,例如短1、2或3个核苷酸。

[0378] 本发明提供了一种非天然存在或工程化组合物,其包含

[0379] 突变的Cas13效应蛋白,以及

[0380] crRNA,

[0381] 其中crRNA包含:a)能够与靶RNA序列杂交的引导序列,和b)正向重复序列,

[0382] 由此形成包含与引导序列复合的Cas13效应蛋白的CRISPR复合物,该引导蛋白与靶RNA序列杂交。该复合物可以在体外或离体形成,并引入细胞中或与RNA接触。或可以在体内形成。

[0383] 在一些实施方式中,例如对于Cas13b,本发明的非天然存在或工程化组合物可包含增强VI-B型CRISPR-Cas效应蛋白活性的辅助蛋白。

[0384] 在某些此类实施方式中,增强Cas13b效应子蛋白活性的辅助蛋白是csx28蛋白。在这样的实施方式中,VI-B型CRISPR-Cas效应蛋白和VI-B型CRISPR-Cas辅助蛋白可以来自相同来源或不同来源。

[0385] 在一些实施方式中,本发明的非天然存在或工程化组合物包含抑制 Cas13b效应子蛋白活性的辅助蛋白。

[0386] 在某些此类实施方式中,抑制Cas13b效应子蛋白活性的辅助蛋白是 csx27蛋白。在这样的实施方式中,VI-B型CRISPR-Cas效应蛋白和VI-B型CRISPR-Cas辅助蛋白可以来自相同来源或不同来源。在本发明的某些实施方式中,VI-B型CRISPR-Cas效应蛋白来自表1。

[0387] 在一些实施方式中,本发明的非天然存在或工程化组合物包含两个或更多个 crRNA。

[0388] 在一些实施方式中,本发明的非天然存在或工程化组合物包含与原核细胞中的靶RNA序列杂交的指导序列。

[0389] 在一些实施方式中,本发明的非天然存在或工程化组合物包含与真核细胞中的靶RNA序列杂交的指导序列。

[0390] 在一种实施方式中,Cas13效应蛋白包含一个或多个核定位信号(NLS)。

[0391] 在本实施方式中,本发明的Cas13效应蛋白是源自表1-4的蛋白质,或在其中,或由其构成,或基本上由其组成,或由其组成,或涉及或与之有关。并且包含本文其他地方所述的本发明的一种或多种突变。

[0392] 在本发明的非天然存在或工程化组合物的一些实施方式中,Cas13效应子蛋白与一个或多个功能域结合。所述结合可以通过效应子蛋白与功能域的直接连接,或通过与 crRNA的结合。在非限制性实例中,crRNA 包含可与感兴趣的功能域结合的添加或插入的序列,所述功能域包括例如与核酸结合衔接子蛋白结合的适体或核苷酸。功能域可以是功能异源域。

[0393] 在某些非限制性实施方式中,本发明的非天然存在或工程化组合物包含切割靶RNA序列的功能性结构域。

[0394] 在某些非限制性实施方式中,本发明的非天然存在或工程化组合物包含修饰靶RNA序列的转录或翻译的功能域。

[0395] 在本发明的组合物的一些实施方式中,Cas13效应子蛋白与一个或多个功能域结合;例如,Cas13效应蛋白与一个或多个功能域结合。并且效应蛋白在HEPN结构域内包含一个或多个突变,由此复合物可以传递表观遗传修饰子或转录或翻译激活或抑制信号。该复合物可以在体外或离体形成,并引入细胞中或与RNA接触。或可以在体内形成。

[0396] 在本发明的非天然存在或工程化组合物的一些实施方式中,Cas13b 效应蛋白和辅助蛋白来自同一生物。

[0397] 在本发明的非天然存在或工程化组合物的一些实施方式中,Cas13b 效应蛋白和辅助蛋白来自不同生物。

[0398] 本发明还提供了VI型CRISPR-Cas载体系统,其包含一个或多个载体,所述载体包括:

[0399] 可操作地连接至编码Cas13效应蛋白的核苷酸序列的第一调节元件,和

[0400] 第二调控元件可操作地连接至编码crRNA的核苷酸序列。

[0401] 在某些实施方式中,本发明的载体系统还包含与VI-B型CRISPR-Cas 辅助蛋白的核苷酸序列可操作地连接的调节元件。

[0402] 适当时,对编码VI型CRISPR-Cas效应蛋白的核苷酸序列(和/或任选地编码VI-B CRISPR-Cas辅助蛋白的核苷酸序列)进行密码子优化以在真核细胞中表达。

[0403] 在本发明的载体系统的一些实施方式中,对编码Cas13效应子蛋白(和任选地)辅助蛋白的核苷酸序列进行密码子优化以在真核细胞中表达。

[0404] 在某个实施方式中,本发明的载体系统包含在单个载体中。

[0405] 在本发明的载体系统的一些实施方式中,一种或多种载体包括病毒载体。

[0406] 在本发明的载体系统的一些实施方式中,一种或多种载体包含一种或多种逆转录病毒,慢病毒,腺病毒,腺相关或单纯疱疹病毒载体。

[0407] 本发明提供了一种递送系统,其被配置为递送Cas13效应子蛋白和非天然存在或工程化组合物的一种或多种核酸组分,包括

[0408] 如本文所述的根据本发明的突变的Cas13效应子蛋白,和

[0409] crRNA,

[0410] 其中crRNA包含a)与细胞中的靶RNA序列杂交的引导序列,和b)正向重复序列,

[0411] 其中Cas13效应蛋白与crRNA形成复合物,

[0412] 其中所述指导序列特异性结合至靶RNA序列,

[0413] 由此形成包含与引导序列复合的Cas13效应蛋白的CRISPR复合物,该引导蛋白与靶RNA序列杂交。该复合物可以在体外或离体形成,并引入细胞中或与RNA接触。或可以在体内形成。

[0414] 在本发明的递送系统的一些实施方式中,该系统包含一种或多种载体或一种或多种多核苷酸分子,该一种或多种载体或多核苷酸分子包含编码Cas13效应蛋白的一种或多种多核苷酸分子和一种或多种核酸。非天然成分或工程成分的成分。

[0415] 在某实施方式中,本发明的递送系统包括递送载体,其包含脂质体,颗粒,外来体,微泡,基因枪或一种或多种病毒载体。。

[0416] 在某个实施方式中,本发明的非天然存在或工程化组合物用于治疗的治疗方法或研究程序中。

[0417] 在某个实施方式中,本发明的非天然存在的或工程化载体系统用于治疗的治疗方法或研究程序中。

[0418] 在某个实施方式中,本发明的非天然存在或工程化递送系统用于治疗的治疗方法或研究程序中。

[0419] 本发明提供了一种修饰感兴趣的靶基因表达的方法,该方法包括使靶RNA与一种或多种非天然存在或工程化组合物接触,所述组合物包含:

[0420] 如本文所述的根据本发明的突变的Cas13效应子蛋白,和

[0421] crRNA,

[0422] 其中crRNA包含a)与细胞中的靶RNA序列杂交的引导序列,和b)正向重复序列,

[0423] 其中Cas13效应蛋白与crRNA形成复合物,

[0424] 其中所述指导序列特异性结合至细胞中的靶RNA序列,

[0425] 从而形成一种CRISPR复合物,该复合物包含与指导序列复合的Cas13效应蛋白,该指导序列与靶RNA序列杂交,

[0426] 从而改变感兴趣的靶基因座的表达。该复合物可以在体外或离体形成,并引入细

胞中或与RNA接触。或可以在体内形成。

[0427] 在某个实施方式中,修饰感兴趣的靶基因表达的方法还包括使靶RNA与增强Cas13b效应子蛋白活性的辅助蛋白接触。

[0428] 在修饰感兴趣的靶基因表达的方法的一些实施方式中,增强Cas13b 效应子蛋白活性的辅助蛋白是csx28蛋白。

[0429] 在某个实施方式中,修饰感兴趣的靶基因表达的方法进一步包括使靶RNA与抑制Cas13b效应子蛋白活性的辅助蛋白接触。

[0430] 在修饰感兴趣的靶基因表达的方法的一些实施方式中,抑制Cas13b 效应子蛋白活性的辅助蛋白是csx27蛋白。

[0431] 在某个实施方式中,修饰感兴趣的靶基因表达的方法包括切割靶 RNA。

[0432] 在某个实施方式中,修饰感兴趣的靶基因表达的方法包括增加或减少靶RNA的表达。

[0433] 在修饰感兴趣的靶基因表达的方法的一些实施方式中,靶基因在原核细胞中。

[0434] 在修饰感兴趣的靶基因表达的方法的一些实施方式中,靶基因在真核细胞中。

[0435] 本发明提供了包含修饰的目的靶的细胞,其中所述目的靶已经根据本文公开的任何方法修饰。

[0436] 在本发明的一种实施方式中,该细胞是原核细胞。

[0437] 在本发明的一种实施方式中,该细胞是真核细胞。

[0438] 在某个实施方式中,细胞中感兴趣目标的修饰会导致:

[0439] 包含至少一种基因产物的表达改变的细胞;

[0440] 包含至少一种基因产物的表达改变的细胞,其中所述至少一种基因产物的表达增加;或者

[0441] 包含至少一种基因产物的表达改变的细胞,其中所述至少一种基因产物的表达降低。

[0442] 在某个实施方式中,该细胞是哺乳动物细胞或人类细胞。

[0443] 本发明提供了本文公开的细胞或包含本文公开的细胞或通过本文公开的任何方法修饰的细胞或其后代的细胞系。

[0444] 本发明提供了一种多细胞生物,其包含一个或多个本文公开的细胞或一个或多个根据本文公开的任何方法修饰的细胞。

[0445] 本发明提供了一种植物或动物模型,其包含一个或多个本文公开的细胞或一个或多个根据本文公开的任何方法修饰的细胞。

[0446] 本发明提供了来自本文公开的细胞或细胞系或生物或植物或动物模型的基因产物。

[0447] 在某种实施方式中,表达的基因产物的量大于或小于未改变表达的细胞的基因产物的量。

[0448] 在某些实施方式中,Cas13蛋白起源于以下种类:别样杆菌属 (Alistipes), 厌氧杆菌,拟杆菌,拟杆菌,贝氏杆菌,Blautia,Butyrivibrio, Capnocytophaga, 食肉杆菌,绿弯曲菌,Chryseobacterium,Clostridium, Demequina,Feubacteriaceae, Eubacteriaceae,Herbinix,Insolitispirillum, Lachnospiraceae,Leptotrichia,李斯

特菌, Myroides, Paludibacter, Phaeodylibacter, Porphyromonadaceae, Porphyromonas, Prevotella, Pseudobutyrvibrio, Psychroflexus, Reichenbachiella, Rhodobacter, Riemerella, 中微芽孢杆菌。如本文所用, 当Cas13蛋白起源于一个物种时, 它可能是该物种中的野生型Cas13蛋白, 或者是该物种中的野生型 Cas13蛋白的同源物。作为物种中的野生型Cas13蛋白的同源物的Cas13 蛋白可以包含野生型Cas13蛋白的一种或多种变异(例如, 突变, 截短等)。

[0449] 在一些实施方式中, Cas13蛋白起源于沙丙氏梭菌, 李斯特菌, 链霉菌属细菌(例如 Lb MA2020, Lb NK4A179, Lb NK4A144), 氨基梭菌(例如Ca DSM 10710), 鸡冠状杆菌(例如 Cg DSM 4847), 丙酸杆状杆菌(如Pp WB4), 魏斯特氏李斯特菌(如Lw FSL R9-0317), 李斯特菌属细菌(如Lb FSL M6-0635), Wepteptrichia wadei(如Lw F0279), 荚膜红细菌(如Rc SB 1003, Rc R121, Rc DE442), buccalis Leptotrichia buccalis(例如Lb C-1013-b), Herminix hemicellulosilytica, Eubacteriaceae细菌(例如Eb CHKCI004), Blautia, 马赛-P2398, Leptotrichia sp. 口腔分类单元879str.F0557, 绿屈挠菌, Demequina aurantiaca, Thalassospira sp TSL5-1, Pseudobutyrvibrio sp. OR37, Butyrvibrio sp. YAB3001, Leptotrichia sp. 马赛P3007, 拟杆菌, 猪肺菌科细菌(如Pb KH3CP3RA), 利斯特氏菌, 百日草不通杆菌, Alistipes sp. ZOR0009, 化脓性杆菌(如Bp F0041), 拟杆菌属细菌(如Bb GWA2\_31\_9), 兽疫伯尔氏菌(如Bz ATCC 43767), 食管噬菌性, 食环梭状芽孢杆菌, 嗜酸梭状芽孢杆菌, 黄杆菌属316, Myroides odoratimimus(如Mo CCUG 10230, Mo CCUG 12901, Mo CCUG 3837), 丙酸杆菌(Pluudibacter propionicigenes), 厦门细菌(Phaeodactylibacter xiamenensis), 牙龈卟啉单胞菌(Porphyromonas gingivalis)(如Pg F0185, Pg F0568, Pg JCVI SC001, Pg W4087, Porphyromonas.COT-0520H4946, 橘色丙酸杆菌, 颊普雷沃菌(如Pb ATCC 33574), 假小肠丙酸杆菌, 中间媒介小球藻(如Pi 17, Pi ZT), 淡水小球藻(如Pp ATCC 700821), 普氏小球藻, 解毒小球藻(saccharchara)(如Ps F0055), Prevotella sp. MA2016, Prevotella sp. MSX73, Prevotella sp. P4-76, Prevotella sp. P5-119, Prevotella sp. P5-125, Prevotella sp. P5-60, Psychroflexus torquis, Reichenbachiella agariperforans, Riemerella 厌食菌, 海洋微生物, 坏死梭状芽孢杆菌(例如Fn亚种ATCC 51357, Fn DJ-2, Fn BFTR-1, Fn亚种 Funduliforme), 产气梭状芽孢杆菌(例如Fp ATCC 29250), 溃疡杆菌(例如Fu ATCC 49185), 厌氧消化细菌ND 1, 西非杆菌(Eubacterium siraeum), 黄褐球菌(Ruminococcus flavefaciens)(例如Rfx XPD3002) 或黄褐球菌(Ruminococcus albus)。

[0450] 在某些实施方式中, Cas13是Cas13a, 起源于拟杆菌属, Blautia, Butyrvibrio, 食肉杆菌, 绿弯曲菌, 梭状芽孢杆菌, Demequina, 真细菌, Herpinix, 不溶螺旋体, 漆膜螺旋菌, Leptotrichia, 李斯特菌, 古菌, 衣藻, 红细菌或海藻。

[0451] 在一些实施方式中, Cas13是Cas13a, 起源于沙丙氏梭菌, 李斯特菌, 链霉菌属细菌(例如Lb MA2020, Lb NK4A179, Lb NK4A144), 氨基梭菌(如Ca DSM 10710), 鸡食杆菌(如食管杆菌), Cg DSM 4847, 丙酸杆菌(如Pp WB4), 魏氏李斯特菌(如Lw FSL R9-0317), 李斯特氏菌属细菌(如Lb FSL M6-0635), 韦氏杆菌属(Leptotrichia wadei)(如Lw F0279), 荚膜红细菌例如Rc SB 1003, Rc R121, Rc DE442), 丁酸杆菌(Leptotrichia buccalis)(例如Lb C-1013-b), Herminix hemicellululilytica, 真细菌科细菌(例如Eb CHKCI004),

Blautia, 马赛-P2398, Leptotrichia sp. 口腔分类单元879str.F0557, 绿屈挠菌, Demequina aurantiaca, Thalassospira sp.TSL5-1, Pseudobutyrvibrio sp.OR37, Butyrvibrio sp.YAB3001, Leptotrichia sp. 马赛P3007, 艾美拟杆菌, 卟啉单胞菌属细菌 (例如Pb KH3CP3RA), 河岸李斯特菌或百日草不溶菌。

[0452] 在某些实施方式中, Cas13是Cas13b, 起源于以下种类: 阿利培斯, 拟杆菌, 拟杆菌, 贝氏杆菌, Capnocytophaga, 金黄杆菌, 黄杆菌, Myroides, 古菌, 杆状杆菌, 卟啉单胞菌, 普雷维氏菌, Psychrofexus, Riechenbach 或Reichenbach微生物。

[0453] 在某些实施方式中, Cas13是Cas13b, 起源于Alistipes sp.ZOR0009, 化脓性杆菌 (如Bp F0041), 拟杆菌属细菌 (如Bb GWA2\_31\_9), 兽疫伯尔氏菌 (如Bz ATCC 43767), 食管噬菌性, 食环梭状芽胞杆菌, 嗜酸梭状芽胞杆菌, 黄杆菌属316, 香豌豆花 (如Mo CCUG 10230, Mo CCUG 12901, Mo CCUG 3837), 丙酸杆菌 (Pluudibacter propionicigenes), Phaeodactylibacter xiamenensis, 牙龈卟啉单胞菌 (Porphyromonas gingivalis) (如Pg F0185, Pg F0568, Pg JCVI SC001, Pg W4087, Porphyromonas, COT-0520H4946, 橘色丙酸杆菌, 颊普雷沃菌 (如Pb ATCC 33574), 假小肠丙酸杆菌, 中间媒介小球藻 (如Pi 17, Pi ZT), 淡水小球藻 (如Pp ATCC 700821), 普氏小球藻, 解毒小球藻 (saccharchara) (如Ps F0055), Prevotella sp.MA2016, Prevotella sp.MSX73, Prevotella sp.P4-76, Prevotella sp.P5-119, Prevotella sp.P5-125, Prevotella sp.P5-60, Psychrofexus torquis, Reichenbachiella agariperforans, Riemerella。在一些实例中, 当Cas13是死拟南芥Riemerella Anatipestifer Cas13。在一些实例中, Cas13是Prevotella sp.P5-125。实例中, Cas13 是死去的Prevotella sp.P5-125。

[0454] 在某些实施方式中, Cas13是Cas13c, 起源于Fusobacter或 Anaerosalibacter属。

[0455] 在某些实施方式中, Cas13是Cas13c, 起源于坏死镰刀菌 (例如Fn 亚种 Fatuliforme ATCC 51357, Fn DJ-2, Fn BFTR-1, Fn亚种 Funduliforme), Fusobacter perfoetens (例如Fp ATCC 29250), 溃疡杆菌 (例如Fu ATCC 49185) 或厌氧弯曲杆菌 (Anaerosalibacter sp.ND1)。

[0456] 在某些实施方式中, Cas13是Cas13d, 起源于真细菌或鲁米诺球菌属。

[0457] 在某些实施方式中, Cas13是Cas13d, 起源于大肠杆菌, 黄褐球菌 (如Rfx XPD3002) 或白蜡球菌。

[0458] 在某些实施方式中, 本发明提供了分离的Cas13效应蛋白, 其包含或基本上由表1-4列出或由其组成, 或由表1-4列出或由其组成, 并且包含如本文其他地方所述的一个或多个突变。表1-4Cas13效应蛋白如本文结合表1-4更详细地讨论。本发明提供了编码Cas13效应蛋白的分离的核酸。在本发明的一些实施方式中, 分离的核酸包含DNA序列, 并且进一步包含编码crRNA的序列。本发明提供了分离的真核细胞, 其包含编码Cas13效应蛋白的核酸。因此, 在本文中, “Cas13效应蛋白”或“效应蛋白”或“Cas”或“Cas蛋白”或“RNA靶向效应蛋白”或“RNA靶向蛋白”或类似表达应理解为包括Cas13a, Cas13b, Cas13c或Cas13d; 诸如“靶向RNA 的CRISPR系统”的表达应理解为包括Cas13a, Cas13b, Cas13c或Cas13d CRISPR系统, 并且在某些实施方式中可以被理解为表1-4的Cas13效应蛋白CRISPR系统。结合指导RNA或sgRNA的参考文献应结合本文讨论的Cas13系统crRNA一起阅读, 例如, 在本发明中其他系统中的sgRNA 可被视为或类似于crRNA。

[0459] 本发明提供了一种鉴定本发明的Cas13效应子蛋白的合适的指导序列的要求的方法(例如,表1-4),所述方法包括:

[0460] 选择生物体内的一组必需基因

[0461] 设计能够与这些基因的编码区以及这些基因的5'和3'UTR区域杂交的靶向指导序列文库

[0462] 产生不与所述生物的基因组内任何区域杂交的随机指导序列作为对照指导

[0463] 制备包含RNA靶向蛋白和第一抗性基因的质粒,以及包含所述靶向向导和所述对照向导以及第二抗性基因的向导质粒文库,

[0464] 将所述质粒共引入宿主细胞

[0465] 将所述宿主细胞引入用于所述第一抗性基因和第二抗性基因的选择性培养基上

[0466] 对生长中的宿主细胞的必需基因进行测序

[0467] 通过将细胞的损耗与对照向导进行比较,确定用靶向向导转化的细胞的损耗的重要性;和

[0468] 基于耗尽的指导序列确定合适的指导序列的要求。

[0469] 在这种方法的一方面,确定RNA靶向蛋白的合适的指导序列的PFS 序列是通过比较耗尽的细胞中的指导所靶向的序列。在这种方法的一方面,该方法还包括在不同的重复实验中比较不同条件下的指导序列丰度。在这种方法的一方面,选择对照指导序列是因为确定它们在重复实验中显示出指导序列消耗中的有限偏差。在这种方法的一个方面,消耗的显著性被确定为:(a)比最消耗的对照指导还多的消耗;(b)大于平均消耗加上对照指导的标准差的两倍的消耗。在这种方法的一方面,宿主细胞是细菌宿主细胞。在这种方法的一方面,共引入质粒的步骤是通过电穿孔,并且宿主细胞是电感受态宿主细胞。

[0470] 本发明提供了修饰与感兴趣的靶基因座相关或在其上的序列的方法,该方法包括将包含Cas13效应蛋白和一种或多种核酸组分的非天然存在或工程化组合物递送至所述基因座,其中该效应蛋白当与所述一种或多种核酸成分形成复合物时,所述复合物与感兴趣的基因座结合后,效应子蛋白诱导与目标感兴趣的基因座相关或在其上的序列的修饰。在一个优选的实施方式中,修饰是引入链断裂。在一个优选的实施方式中,与感兴趣的靶基因座相关或在其上的序列包含RNA或由RNA组成。

[0471] 本发明提供了修饰与感兴趣的靶基因座相关或在其上的序列的方法,该方法包括将包含Cas13效应子蛋白,任选的小辅助蛋白和一个或多个核酸的非天然存在或工程化组合物递送至所述基因座。核酸组分,其中效应子蛋白与一种或多种核酸组分形成复合物,并且当所述复合物与感兴趣的基因座结合后,效应子蛋白诱导与感兴趣的靶基因座相关或在其上的序列的修饰。在一个优选的实施方式中,修饰是引入链断裂。在一个优选的实施方式中,与感兴趣的靶基因座相关或在其上的序列包含 RNA或由RNA组成。

[0472] 本发明提供了一种修饰与感兴趣的靶位点相关或在其上的序列的方法,该方法包括向所述与靶位点相关或在该位点的序列递送包含Cas13 基因座效应子蛋白和一个或多个核酸的非天然存在或工程化组合物。核酸组分,其中Cas13效应蛋白与一种或多种核酸组分形成复合物,并且当所述复合物结合至目的基因座时,效应蛋白诱导与感兴趣的靶基因座相关或在其上的序列的修饰。在一个优选的实施方式中,修饰是引入链断裂。在一个优选的实施方式中,Cas13效应蛋白与一种核酸组分形成复合物。有利地,工程或非天然存在的核

酸组分。与感兴趣的靶基因座相关或在其上的序列的修饰的诱导可以是Cas13效应子蛋白-核酸引导的。在一个优选的实施方式中,一种核酸组分是CRISPR RNA (crRNA)。在一个优选的实施方式中,一种核酸组分是成熟的crRNA或指导RNA,其中该成熟的crRNA或指导RNA包含间隔子序列(或指导序列)和正向重复(DR)序列或其衍生物。在一个优选的实施方式中,间隔区序列或其衍生物包含种子序列,其中种子序列对于在靶基因座处的序列的识别和/或杂交是关键。在本发明的优选实施方式中,crRNA是可以与短DR序列相关的短crRNA。在本发明的另一个实施方式中,crRNA是可以与长DR序列(或双重DR)相关的长crRNA。本发明的方面涉及具有一种或多种非天然存在或工程改造或修饰或优化的核酸组分的Cas13效应蛋白复合物。在一个优选的实施方式中,核酸组分包含RNA。在一个优选的实施方式中,复合物的核酸组分可包含与正向重复序列连接的指导序列,其中该正向重复序列包含一个或多个茎环或优化的二级结构。在本发明的优选实施例中,正向重复可以是短DR或长DR(双DR)。在一个优选的实施方式中,可以修饰正向重复序列以包含一种或多种蛋白结合RNA适体。在一个优选的实施方式中,可以包括一个或多个适体,例如优化的二级结构的一部分。这样的适体可能能够结合噬菌体外壳蛋白。噬菌体外壳蛋白可以选自QB, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205,  $\phi$  Cb5,  $\phi$  Cb8r,  $\phi$  Cb12r,  $\phi$  Cb23r, 7s和 PRR1。在一个优选的实施方式中,噬菌体外壳蛋白是MS2。本发明还提供了复合物的核酸成分,其长度为30个或更多个,40个或更多个或50个或更多个核苷酸。

[0473] 本发明提供了与感兴趣的靶位点相关或在其上的基因组编辑或修饰序列的方法,其中该方法包括将Cas13复合物引入任何所需的细胞类型,原核或真核细胞中,由此Cas13效应蛋白复合物有效地起作用以干扰RNA。在真核或原核细胞中。在优选的实施方式中,细胞是真核细胞,并且RNA从哺乳动物基因组转录或存在于哺乳动物细胞中。在人细胞中RNA编辑或基因组编辑的优选方法中,Cas13效应蛋白可以包括但不限于本文公开的Cas13效应蛋白的特定种类。

[0474] 本发明还提供了修饰感兴趣的靶基因座的方法,该方法包括将包含Cas13效应蛋白和一种或多种核酸成分的非天然存在或工程化组合物递送至所述基因座,其中Cas13效应蛋白形成复合物。在具有一种或多种核酸成分的情况下,并且在所述复合物与感兴趣的基因座结合后,效应子蛋白诱导目标感兴趣的基因座的修饰。在一个优选的实施方式中,修饰是引入链断裂。

[0475] 在这样的方法中,感兴趣的靶基因座可以包含在RNA分子内。在这样的方法中,感兴趣的靶基因座可以体外包含在RNA分子中。

[0476] 在这样的方法中,感兴趣的靶基因座可以包含在细胞内的RNA分子中。该细胞可以是原核细胞或真核细胞。该细胞可以是哺乳动物细胞。通过本发明引入细胞的修饰可以是这样的,使得改变细胞和细胞的后代以改善生物产物如抗体,淀粉,醇或其他所需细胞输出的产生。通过本发明引入细胞的修饰可以使得细胞和细胞的后代包括改变产生的生物产物的改变。

[0477] 哺乳动物细胞可以是非人类哺乳动物,例如灵长类动物,牛,绵羊,猪,犬,啮齿动物,鼠科,例如猴,牛,绵羊,猪,狗,兔,大鼠或小鼠细胞。该细胞可以是非哺乳动物的真核细胞,例如禽鸟(例如,鸡),脊椎动物(例如,鲑鱼)或贝类(例如,牡蛎,白鲑,龙虾,虾)细胞。该

细胞也可以是植物细胞。植物细胞可以是单子叶植物或双子叶植物,也可以是农作物或谷物植物,例如木薯,玉米,高粱,大豆,小麦,燕麦或水稻。植物细胞还可以是藻类,树木或生产植物,水果或蔬菜(例如,柑橘树等树木,例如橙,葡萄柚或柠檬树;桃或油桃树;苹果或梨树;坚果树等)的细胞。杏仁树或核桃树或开心果树;茄属植物;芸苔属植物;Lectica属植物;菠菜属植物;辣椒属植物;棉花,烟草,芦笋,胡萝卜,白菜,西兰花,花椰菜,番茄,茄子,胡椒,生菜,菠菜,草莓,蓝莓,覆盆子,黑莓,葡萄,咖啡,可可等)。

[0478] 本发明提供了一种修饰感兴趣的靶基因座的方法,该方法包括将包含Cas13效应蛋白和一种或多种核酸组分的非天然存在或工程化组合物递送至所述基因座,其中所述效应蛋白与所述靶蛋白形成复合物。一种或多种核酸组分,并且在所述复合物与感兴趣的基因座结合后,效应子蛋白诱导目标感兴趣的基因座的修饰。在一个优选的实施方式中,修饰是引入链断裂。

[0479] 在这样的方法中,感兴趣的靶基因座可以包含在RNA分子内。在一个优选的实施方式中,感兴趣的靶基因座包含RNA或由RNA组成。

[0480] 本发明还提供了修饰感兴趣的靶基因座的方法,该方法包括将包含 Cas13效应蛋白和一种或多种核酸成分的非天然存在或工程化组合物递送至所述基因座,其中Cas13效应蛋白形成复合物。在具有一种或多种核酸成分的情况下,并且在所述复合物与感兴趣的基因座结合后,效应子蛋白诱导目标感兴趣的基因座的修饰。在一个优选的实施方式中,修饰是引入链断裂。

[0481] 优选地,在这样的方法中,感兴趣的靶基因座可以体外包含在RNA 分子中。同样优选地,在这种方法中,感兴趣的靶基因座可包含在细胞内的RNA分子中。该细胞可以是原核细胞或真核细胞。该细胞可以是哺乳动物细胞。该细胞可以是啮齿动物细胞。该细胞可以是小鼠细胞。

[0482] 在任何描述的方法中,感兴趣的靶基因座可以是目的基因组或表观基因座。在所描述的任何方法中,复合体可以与多个引导件一起被递送以用于多重使用。在任何描述的方法中,可以使用一种以上的蛋白质。

[0483] 在本发明的其他方面,核酸组分可以包含CRISPR RNA (crRNA) 序列。由于效应蛋白是Cas13效应蛋白,因此核酸组分可以包含CRISPR RNA (crRNA) 序列,并且通常可以不包含任何反式激活crRNA (tracr RNA) 序列。

[0484] 在任何描述的方法中,效应蛋白和核酸组分可以通过编码该蛋白和/ 或核酸组分的一个或多个多核苷酸分子来提供,并且其中一个或多个多核苷酸分子被可操作地配置为表达该蛋白。和/或核酸成分。一种或多种多核苷酸分子可包含一种或多种可操作地配置为表达蛋白质和/或核酸组分的调控元件。一种或多种多核苷酸分子可包含在一种或多种载体内。在任何描述的方法中,感兴趣的靶基因座可以是基因组,表观基因组或转录组基因座。在所描述的任何方法中,复合体可以与多个引导件一起被递送以用于多重使用。在任何描述的方法中,可以使用一种以上的蛋白质。

[0485] 在任何描述的方法中,链断裂可以是单链断裂或双链断裂。在优选的实施方式中,双链断裂可指RNA的两个部分的断裂,例如当单链RNA 分子自身折叠时形成的RNA的两个部分或由包含自身的RNA分子形成的推定的双螺旋。互补序列允许RNA的部分折叠并与其自身配对。

[0486] 调节元件可包含诱导型启动子。多核苷酸和/或载体系统可以包括诱导系统。

[0487] 在任何描述的方法中,一个或多个多核苷酸分子可包含在递送系统中,或所述一个或多个载体可包含在递送系统中。

[0488] 在任何描述的方法中,可以通过脂质体,包括纳米颗粒,外来体,微囊泡,基因枪或一种或多种病毒载体的颗粒递送非天然存在或工程化组合物。

[0489] 本发明还提供了一种非天然存在的或工程化组合物,其是具有如本文所讨论的或在本文所述的任何方法中定义的特征的组合物。

[0490] 因此,在某些实施方式中,本发明提供了一种非天然存在或工程化组合物,例如特别是能够或被配置为修饰感兴趣的靶基因座的组合物,所述组合物包含Cas13效应蛋白和一种或多种核酸。组分,其中效应蛋白与一种或多种核酸组分形成复合物,并且当所述复合物与感兴趣的基因座结合后,效应蛋白诱导目标靶基因的修饰。效应蛋白可以是Cas13a, Cas13b, Cas13c或Cas13d效应蛋白,优选为Cas13b效应蛋白。

[0491] 在另一方面,本发明还提供了非天然存在或工程化组合物,例如特别是能够或被配置为修饰感兴趣的靶基因座的组合物,所述组合物包含:(a) 指导RNA分子(或指导RNA分子,例如第一指导RNA分子和第二指导RNA分子)或编码该指导RNA分子的核酸(或一个或多个编码指导RNA分子的组合的核酸);(b) 一种Cas13效应蛋白。在某些实施方式中,效应蛋白可以是Cas13b效应蛋白。

[0492] 在另一方面,本发明还提供了一种非天然存在的或工程化组合物,其包含:(I.) 一个或多个CRISPR-Cas系统多核苷酸序列,其包含(a) 能够与多核苷酸基因座中的靶序列杂交的指导序列,(b) tracr mate(即正向重复)序列,和(II.) 编码Cas13效应蛋白的第二个多核苷酸序列,其中在转录时,引导序列指导CRISPR复合物与靶序列的序列特异性结合,和其中,CRISPR复合物包含与引导序列复合的Cas13效应蛋白,该引导序列与靶序列杂交。在某些实施方式中,效应蛋白可以是Cas13b效应蛋白。

[0493] 在某些实施方式中,可能不需要tracrRNA。因此,本发明在某些实施方式中还提供了一种非天然存在的或工程化组合物,其包含:(I.) 一个或多个CRISPR-Cas系统多核苷酸序列,其包含(a) 能够与多核苷酸基因座中的靶序列杂交的指导序列。(b) 正向重复序列,和(II.) 编码 Cas13效应蛋白的第二个多核苷酸序列,其中在转录时,引导序列指导 CRISPR复合物与靶序列的序列特异性结合,并且其中CRISPR该复合物包含与(1) 与靶序列杂交的引导序列,和(2) 正向重复序列复合的Cas13 效应蛋白。优选地,效应蛋白可以是Cas13b效应蛋白。在没有限制的情况下,申请人假设在这种情况下,正向重复序列可包含足以将crRNA加载到效应蛋白上的二级结构。借助于实例而非限制,这种二级结构可在正向重复序列内包含茎环(基本上由其组成)或由其组成(例如,一个或多个茎环)。

[0494] 本发明还提供了包含一种或多种载体的载体系统,所述一种或多种载体包含一种或多种多核苷酸分子,所述多核苷酸分子编码非天然存在或工程化组合物的组分,所述非天然存在的或工程化组合物具有本文所述的任何特征所定义的组合物。方法。

[0495] 本发明还提供了包含一种或多种载体或一种或多种多核苷酸分子的递送系统,所述一种或多种载体或多核苷酸分子包含编码非天然存在或工程化组合物的组分的一种或多种多核苷酸分子,所述非天然存在的或工程化组合物是具有本文讨论的或本文描述的任何方法中定义的特征。

[0496] 本发明还提供了用于治疗方法的非天然存在或工程化组合物,或一种或多种编码所述组合物成分的多核苷酸,或包含编码所述组合物成分的一种或多种多核苷酸的载体或递送系统。治疗的治疗方法可以包括基因或基因组编辑或基因治疗。

[0497] 本发明还提供了方法和组合物,其中效应蛋白的一个或多个氨基酸残基可以被修饰,例如,表1-4蛋白或由其组成或基本上由表1-4蛋白组成的工程或非天然存在的Cas13效应蛋白。在一个实施方式中,该修饰可以包括效应蛋白的一个或多个氨基酸残基的突变。一种或多种突变可以在效应蛋白的一种或多种催化活性结构域中。与缺乏所述一种或多种突变的效应蛋白相比,效应蛋白可以具有降低或消除的核酸酶活性。效应蛋白可能不指导在感兴趣的靶位点切割一条RNA链。在一个优选的实施方式中,一个或多个突变可包含两个突变。在一个优选的实施方式中,在Cas13效应蛋白,例如工程化或非天然存在的Cas13效应蛋白中,一个或多个氨基酸残基被修饰。在本发明的实施方式中,效应子蛋白包含一个或多个HEPN结构域。在一个优选的实施方式中,效应蛋白包含两个HEPN结构域。在另一个优选的实施方式中,效应蛋白在该蛋白的C末端包含一个HEPN结构域,在该蛋白质N末端包含另一个HEPN结构域。在效应蛋白的催化活性结构域中,其包括HEPN结构域或与HEPN结构域同源的催化活性结构域。在某些实施方式中,效应蛋白包括以下一种或多种突变:R116A,H121A,R1177A,H1182A(其中氨基酸位置对应于源自兽疫伯克氏菌ATCC 43767的第29组蛋白质的氨基酸位置)。本领域技术人员将理解,可以将不同的Cas13蛋白中的相应氨基酸位置突变为相同的效果。在某些实施方式中,一个或多个突变完全或部分消除了该蛋白的催化活性(例如,裂解率改变,特异性改变)在某些实施方式中,该效应蛋白是一种“死亡”效应蛋白,例如死的Cas13效应蛋白(即dCas13b)。在HEPN结构域1中有更多的突变。在某些实施方式中,效应蛋白在HEPN结构域2中有一个或多个突变。在某些实施方式中,效应蛋白在HEPN结构域1和HEPN结构域2中具有一个或多个突变。效应蛋白可以包含一个或多个异源功能域。一个或多个异源功能域可包含一个或多个核定位信号(NLS)域。一个或多个异源功能域可包含至少两个或多个NLS域。一个或多个NLS结构域可位于效应蛋白(例如,Cas13b效应蛋白)的末端处或附近或附近,并且如果有两个或多个NLS,则两者中的每一个可位于或邻近或附近或附近。在效应蛋白(例如,Cas13效应蛋白)的末端附近。一个或多个异源功能域可包含一个或多个转录激活域。在一个优选的实施方式中,转录激活结构域可以包含VP64。一个或多个异源功能域可包含一个或多个转录抑制域。在优选的实施方式中,转录抑制结构域包含KRAB结构域或SID结构域(例如SID4X)。一个或多个异源功能结构域可包含一个或多个核酸酶结构域。在一个优选的实施方式中,核酸酶结构域包含FokI。

[0498] 本发明还提供了一个或多个异源功能域,具有以下一种或多种活性:甲基化酶活性,脱甲基酶活性,转录激活活性,转录抑制活性,转录释放因子活性,组蛋白修饰活性,核酸酶活性,单-链RNA切割活性,双链RNA切割活性,单链DNA切割活性,双链DNA切割活性和核酸结合活性。至少一个或多个异源功能结构域可以在效应蛋白的氨基末端处或附近,和/或其中至少一个或多个异源功能域在效应蛋白的羧基末端处或附近。所述一个或多个异源功能域可以与效应蛋白融合。所述一个或多个异源功能结构域可被束缚至效应蛋白。一个或多个异源功能域可以通过接头部分连接至效应蛋白。

[0499] 在某些实施方式中,如本文所述的Cas13效应子蛋白可以与包含短的CRISPR重复序列的基因座相关,所述短的CRISPR重复序列长在30至40bp之间,更通常在34至38bp之

间,甚至更通常在36至37bp之间,例如30、31、32、33、34、35、36、37、38、39或40bp长。在某些实施方式中,CRISPR重复序列是在80至350bp长之间的长或双重复,例如在80至200bp之间,甚至更典型地在86至88bp之间,例如80、81、82、83、84、85、86、87、88、89或90bp长

[0500] 在某些实施方式中,原型间隔子邻近基序(PAM)或PAM样基序指导本文所述的效应子蛋白(例如Cas13效应子蛋白)复合物与感兴趣的靶位点的结合。PAM可以是5' PAM(即位于原间隔子5'末端的上游)。在其他实施例中,PAM可以是3' PAM(即,位于原间隔子的5'端的下游)。在其他实施例中,需要5' PAM和3' PAM。在本发明的某些实施方式中,指导效应子蛋白(例如Cas13效应子蛋白)的结合可能不需要PAM或PAM样基序。在某些实施方式中,5' PAM是D(例如,A,G或U)。在某些实施方式中,Cas13b效应子的5' PAM是D。在本发明的某些实施方式中,在重复序列处的切割可以产生crRNA(例如,短或长的crRNA),其包含由短核苷酸(例如5、6、7、8、9或10nt或更长,如果有的话)侧接的完整间隔子序列。是双重复序列)在5'末端的重复序列(可称为crRNA“标签”),其余重复序列在3'末端。本文描述的蛋白质可能要求crRNA标签与靶标5'侧翼序列之间缺乏同源性。该要求可能与Samai等人在下文中进一步描述的要求类似,“Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity” Cell 161,1164-1174, 2015年05月21日,据认为该要求可区分从CRISPR阵列本身入侵的核酸的天然靶标,以及重复序列的存在会导致与crRNA标签完全同源并阻止自身免疫。

[0501] 在某些实施方式中,Cas13效应蛋白经过工程改造,可以包含一种或多种减少或消除核酸酶活性,从而减少或消除RNA干扰活性的突变。也可以在邻近残基处,例如在参与核酸酶活性的那些氨基酸附近的氨基酸处进行突变。在一些实施方式中,一个或多个推定的催化核酸酶结构域被失活,效应蛋白复合物缺乏切割活性,并起RNA结合复合物的作用。在一个优选的实施方式中,所得的RNA结合复合物可以与一个或多个本文所述的功能性结构域连接。

[0502] 在某些实施方式中,一个或多个功能域是可控的,即可诱导的。

[0503] 在本发明的某些实施方式中,指导RNA或成熟crRNA包含,基本上由正向重复序列和指导序列或间隔子序列组成或由其组成。基本上由指导序列或间隔序列连接的正向重复序列组成或由其组成。在本发明的优选实施方式中,成熟crRNA包含茎环或优化的茎环结构或优化的二级结构。在优选的实施方式中,成熟crRNA在正向重复序列中包含茎环或优化的茎环结构,其中茎环或优化的茎环结构对于裂解活性是重要的。在某些实施方式中,正向重复序列优选包含单个茎环。在某些实施方式中,通过引入影响茎环RNA双链体结构的突变来修饰效应蛋白复合物的切割活性。在优选的实施方式中,可以引入维持茎环的RNA双链体的突变,从而保持效应蛋白复合物的切割活性。在其他优选的实施方式中,可以引入破坏茎环的RNA双链体结构的突变,从而完全消除效应蛋白复合物的切割活性。

[0504] 本文提供的CRISPR系统可利用包含指导序列的crRNA或类似多核苷酸,其中多核苷酸是RNA,DNA或RNA和DNA的混合物,和/或其中多核苷酸包含一种或多种核苷酸类似物。该序列可以包含任何结构,包括但不限于天然crRNA的结构,例如凸起,发夹或茎环结构。在某些实施方式中,包含引导序列的多核苷酸与可以是RNA或DNA序列的第二多核苷酸序列形成双链体。

[0505] 本公开还提供了包含工程化CRISPR-Cas蛋白的细胞,组织,生物体, CRISPR-Cas

系统,编码CRISPR-Cas系统的一种或多种组分的多核苷酸和/或包含所述多核苷酸的载体。本发明还提供了编码效应子蛋白的核苷酸序列,该核苷酸蛋白经密码子优化以在本文所述的任何方法或组合物中在真核生物或真核细胞中表达。在本发明的一个实施方式中,密码子优化的效应子蛋白是本文讨论的任何Cas13效应子蛋白,并且对于在真核细胞或生物中,例如但不限于本文其他地方提到的这种细胞或生物,对密码子进行了优化。酵母细胞或哺乳动物细胞或生物,包括小鼠细胞,大鼠细胞和人细胞或非人真核生物,例如植物。

[0506] 在本发明的某些实施方式中,至少一个核定位信号(NLS)附着于编码Cas13效应蛋白的核酸序列。在优选的实施方式中,至少一个或多个C端或N端NLS连接(因此编码Cas13效应子蛋白的核酸分子可以包括NLS编码,以便表达的产物具有NLS连接或连接)。在一个优选的实施方式中,为了在真核细胞,优选人细胞中的最佳表达和核靶向,连接了C端NLS。本发明还包括用于递送多种核酸组分的方法,其中每种核酸组分对不同的感兴趣的靶基因座具有特异性,从而修饰了多个感兴趣的靶基因座。复合物的核酸组分可以包含一种或多种蛋白结合RNA适体。一种或多种适体可能能够结合噬菌体外壳蛋白。

[0507] 在另一方面,本发明提供了包含修饰的感兴趣的靶基因座的真核细胞,其中所述感兴趣的靶基因座已根据本文所述的任何方法进行了修饰。另一方面提供了所述细胞的细胞系。另一方面提供了包含一个或多个所述细胞的多细胞生物。

[0508] 在某些实施方式中,感兴趣的靶基因座的修饰可导致:真核细胞包含改变的至少一种基因产物的表达;该真核细胞包含改变的至少一种基因产物的表达,其中该至少一种基因产物的表达增加;该真核细胞包含改变的至少一种基因产物的表达,其中该至少一种基因产物的表达降低;或包含已编辑基因组的真核细胞。

[0509] 在某些实施方式中,真核细胞可以是哺乳动物细胞或人类细胞。

[0510] 在进一步的实施方式中,本说明书中描述的非天然存在或工程化组合物,载体系统或递送系统可用于:位点特异性基因敲除;定点基因组编辑;RNA序列特异性干扰;或多重基因组工程。

[0511] 还提供了来自本文所述的细胞,细胞系或生物体的基因产物。在某些实施方式中,表达的基因产物的量可以大于或小于来自细胞,细胞系或生物体的基因产物的量。没有改变表达或未编辑基因组。在某些实施方式中,与未改变表达或未编辑基因组的细胞的基因产物相比,该基因产物可能会发生改变。

[0512] 在另一个方面中,本发明提供了一种鉴定新颖的核酸修饰效应子的方法,其包括:从编码推定的核酸修饰酶基因座的一组核酸序列中鉴定推定的核酸修饰基因座,所述核酸序列位于距所述抗体的限定距离内。基因座的保守基因组元件,其包含至少一种超过定义的大小限制的蛋白质,或两者兼有;将鉴定出的推定的核酸修饰基因座分组为包含同源蛋白质的子集;通过基于一个或多个以下子项从一个或多个子集中选择核酸修饰基因座,从而鉴定出候选核酸修饰基因座的最终集合;包含具有低域同源性的推定效应蛋白的基因座的子集与相对于其他子集中的基因座的已知蛋白质域相匹配,包含相对于其他亚组的保守基因组与保守基因组距离最小的推定蛋白质的子集,包含具有较大效应子蛋白的基因座的子集相对于其他子集中的大型效应蛋白而言,与推定的相邻辅助蛋白具有相同的方向,该子集包含相对于其他基因座而言具有较低现有核酸修饰分类的推定效应子蛋白,该子集包含的基因座与已知核酸修饰基因座相比具有更低的邻近性子集,以及每个子集中的候选基

因座总数。

[0513] 在一个实施方式中,该核酸序列集是从基因组或宏基因组数据库获得的,例如包含原核基因组或宏基因组序列的基因组或宏基因组数据库。

[0514] 在一个实施方式中,距保守基因组元件的定义距离在1kb和25kb 之间。

[0515] 在一个实施方式中,保守的基因组元件包括重复元件,例如CRISPR 阵列。在一个具体的实施方式中,距保守基因组元件的限定距离在 CRISPR阵列的10kb之内。

[0516] 在一个实施方式中,推定的核酸修饰(效应子)基因座中包含的蛋白质的定义的大小限制大于200个氨基酸,或更具体地,定义的大小限制大于700个氨基酸。在一个实施方式中,推定的核酸修饰位点在900 至1800个氨基酸之间。

[0517] 在一个实施方式中,使用重复的或模式发现分析对一组核酸(如 PILER-CR)进行鉴定,从而确定了保守的基因组元件。

[0518] 在一个实施方式中,本文描述的方法的分组步骤至少部分地基于结构域同源性搜索或HHpred蛋白结构域同源性搜索的结果。

[0519] 在一个实施方式中,定义的阈值为BLAST最近邻居截止值0到 $1e^{-7}$ 。

[0520] 在一个实施方式中,本文所述的方法进一步包括过滤步骤,该过滤步骤仅包括具有900至1800个氨基酸之间的推定蛋白质的基因座。

[0521] 在一个实施方式中,本文所述的方法还包括对候选核酸修饰效应子的核酸修饰功能进行实验验证,包括产生一组编码核酸修饰效应子的核酸构建体并进行一个或多个生化验证测定,例如通过在细菌菌落中使用 PAM验证,体外裂解测定,Surveyor方法,在哺乳动物细胞中进行实验, PFS验证或其组合。

[0522] 在一个实施方式中,本文所述的方法进一步包括从所鉴定的核酸修饰基因座制备包含一种或多种蛋白质的非天然存在或工程化组合物。

[0523] 在一个实施方式中,所鉴定的基因座包括2类CRISPR效应子,或者所鉴定的基因座缺少Cas1或Cas2,或者所鉴定的基因座包括单个效应子。

[0524] 在一个实施方式中,单个大效应蛋白的长度大于900或大于1100个氨基酸,或包含至少一个HEPN结构域。

[0525] 在一个实施方式中,至少一个HEPN结构域在效应蛋白的N端或C 端附近,或位于效应蛋白的内部位置。

[0526] 在一个实施方式中,单个大效应蛋白在N和C端包含一个HEPN结构域,在该蛋白内部包含两个HEPN结构域。

[0527] 在一个实施方式中,所鉴定的基因座进一步包含在CRISPR阵列的 2kb至10kb内的一个或两个小的推定的辅助蛋白。

[0528] 在一个实施方式中,一种小的辅助蛋白质少于700个氨基酸。在一个实施方式中,小辅助蛋白的长度为50至300个氨基酸。

[0529] 在一个实施方式中,小辅助蛋白包含多个预测的跨膜结构域,或包含四个预测的跨膜结构域,或包含至少一个HEPN结构域。

[0530] 在一个实施方式中,小辅助蛋白包含至少一个HEPN结构域和至少一个跨膜结构域。

[0531] 在一个实施方式中,该位点不包含来自CRISPR阵列的最大25kb的额外蛋白质。

[0532] 在一个实施方式中,CRISPR阵列包含正向重复序列,其长度约为 36个核苷酸。在一个具体的实施方式中,正向重复在5'端包含GTTG/ GUUG,其与3'端的CAAC反向互补。

[0533] 在一个实施方式中,CRISPR阵列包含间隔序列,该间隔序列包含约 30个核苷酸的长度。

[0534] 在一个实施方式中,所鉴定的基因座缺少小辅助蛋白。

[0535] 本发明提供了鉴定新颖的CRISPR效应子的方法,其包括:a)在编码CRISPR阵列的基因组或宏基因组数据库中鉴定序列;b)在CRISPR阵列的10kb内,在所述选择的序列中鉴定一个或多个开放阅读框(ORF);c)基于大小在900-1800个氨基酸之间的推定的CRISPR效应蛋白的存在来选择基因座,d)选择编码50-300个氨基酸的推定的辅助蛋白的基因座;e)鉴定编码推定的CRISPR效应子和CRISPR辅助蛋白的基因座,并任选地基于结构分析对它们进行分类。

[0536] 在一个实施方式中,CRISPR效应子是VI型CRISPR效应子。在一个实施方式中,步骤(a)包括:i)比较基因组和/或宏基因组数据库中的序列与至少一个预先确定的编码CRISPR阵列的种子序列,并选择包含所述种子序列的序列;ii)基于CRISPR算法鉴定CRISPR阵列。

[0537] 在一个实施方式中,步骤(d)包括鉴定核酸酶结构域。在一个实施方式中,步骤(d)包括鉴定RuvC,HPN和/或HEPN结构域。

[0538] 在一个实施方式中,CRISPR阵列10kb之内不存在编码Cas1或Cas2的ORF。

[0539] 在一个实施方式中,步骤(b)中的ORF编码一个50-300个氨基酸的辅助蛋白。

[0540] 在一个实施方式中,将步骤(d)中获得的推定的新型CRISPR效应子用作种子序列,以进一步比较基因组和/或宏基因组学序列,并随后选择权利要求1的步骤a)至d)中所述的目的基因座。一种实施方式中,通过以下方法获得预先鉴定的种子序列:(a)在基因组或宏基因组数据库中鉴定CRISPR基序,(b)在所述鉴定的CRISPR基序中提取多个特征,(c)对CRISPR基因座进行分类使用无监督学习,(d)基于所述分类鉴定保守的基因座元件,和(e)从中选择适合作为种子序列的推定的CRISPR效应子。

[0541] 在一个实施方式中,特征包括蛋白质元件,重复结构,重复序列,间隔子序列和间隔子作图。在一个实施方式中,基因组和宏基因组数据库是细菌和/或古细菌基因组。基因组和宏基因组序列可从Ensembl和/或NCBI基因组数据库中获得。在一个实施方式中,步骤(d)中的结构分析基于二级结构预测和/或序列比对。以及手动管理获得的簇。在另一个方面,一种包含一个或多个氨基酸突变的突变的Cas13蛋白,其中所述氨基酸与与所述突变的Cas13蛋白形成复合物的指导RNA相互作用;或在HEPN活性位点中,是一个有盖域,该域是用两个 $\beta$ 发夹(参见图1,图18)覆盖crRNA 3'末端的域,一个螺旋域,选自螺旋1或工程Cas13蛋白的螺旋2结构域,域间接头(IDL)结构域或桥螺旋结构域。在某些实施方式中,螺旋结构域1是螺旋结构域1-1、1-2或1-3。在实施方式中,螺旋结构域2是螺旋结构域2-1或2-2。在一个方面中,与天然存在的对应Cas13蛋白相比,工程化Cas13蛋白具有更高的蛋白酶活性或多核苷酸结合能力。

[0542] 在一些实施方式中,Cas13蛋白是Cas13a、Cas13b、Cas13c或Cas13d。在一些实施方式中,Cas13蛋白是Cas13b。在一些实施方式中,该等氨基酸与指导RNA相互作用,所述指导RNA与突变的Cas13蛋白形成复合物。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b(PbCas13b)的以下氨基酸:T405,H407,K457,H500,K570,K590,N634,R638,N652,

N653, K655, S658, K741, K744, N756, S757, R762, R791, K846, K857, K870和R877。在一些实施方式中,该等氨基酸在HEPN活性位点中。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:氨基酸46-57、73-79、152-164、1036-1046和1064-1074。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:R156, N157, H161, R1068, N1069和H1073。在一些实施方式中,该等氨基酸位于突变的Cas 13蛋白的结构域间接头结构域中。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:R285, R287, K292, K294, E296和N297。在一些实施方式中,该等氨基酸位于突变的Cas 13蛋白的桥螺旋结构域中。在一些实施方式中,该等氨基酸对应于颊普雷沃菌 Cas13b (PbCas13b) 的以下氨基酸:K826, K828, K829, R824, R830, Q831, K835, K836和R838。

[0543] 在另一个方面中,本公开内容提供了一种改变Cas13蛋白活性的方法,该方法包括:基于至少一部分Cas 13蛋白质的三维结构,识别Cas13 蛋白中的一个或多个候选氨基酸,其中一个或多个候选氨基酸与指导 RNA相互作用,该指导RNA与Cas13蛋白形成复合物,或者位于Cas13 蛋白的HEPN活性位点,域间接头结构域或桥螺旋结构域中;突变所述一个或多个候选氨基酸,从而产生突变的Cas13蛋白,其中所述突变的 Cas13蛋白的活性不同于所述Cas13蛋白。

[0544] 在一些实施方式中,Cas13蛋白是Cas13a、Cas13b、Cas13c或Cas13d。在一些实施方式中,Cas13蛋白是Cas13b。在一些实施方式中,该等氨基酸与指导RNA相互作用,所述指导RNA与突变的Cas13蛋白形成复合物。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:T405, H407, K457, H500, K570, K590, N634, R638, N652, N653, K655, S658, K741, K744, N756, S757, R762, R791, K846, K857, K870和R877。在一些实施方式中,该等氨基酸在HEPN活性位点中。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:氨基酸46-57、73-79、152-164、1036-1046和1064-1074。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:R156, N157, H161, R1068, N1069和H1073。在一些实施方式中,该等氨基酸位于突变的Cas 13蛋白的结构域间接头结构域中。在一些实施方式中,该等氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:R285, R287, K292, K294, E296和N297。在一些实施方式中,该等氨基酸位于突变的Cas 13蛋白的桥螺旋结构域中。在一些实施方式中,该等氨基酸对应于颊普雷沃菌 Cas13b (PbCas13b) 的以下氨基酸:K826, K828, K829, R824, R830, Q831, K835, K836和R838。

[0545] 在一些实施方式中,Cas13b是Cas13b。在一些实施方式中,Cas13b 是一个Cas13直系同源物,其大小比迄今发现的Cas13系统小。在一些实施方式中,Cas13b是Cas13b-t1, Cas13b-t1a, Cas13b-t2或Cas13b-t3。在一些实施方式中,Cas13b是Cas13b-t1。在一些实施方式中,Cas13b 是Cas13b-t2。在一些实施方式中,Cas13b是Cas13b-t3。CAS13直系同源物

[0546] 术语“同源物”(在本文中也称为“直系同源物”)和“同源物”(在本文中也称为“同源物”)在本领域中是众所周知的。通过进一步的指导,本文所用的蛋白质的“同源物”是与该蛋白质的同源物具有相同或相似功能的相同物种的蛋白质。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。如本文所用,蛋白质的“同源物”是不同种类的蛋白质,其执行与其作为同源物的蛋白质相同或相似的功能。在特定实施方式中,本文所指的Cas13

蛋白的同源物或直系同源物具有至少60%，优选至少70%的序列同源性或同一性。优选至少80%，更优选至少85%，甚至更优选至少90%，例如至少95%具有下表1-4中列出的Cas13效应蛋白。在一个优选的实施方式中，Cas13b效应蛋白可以是表1-4中鉴定的生物或该生物所属的属或来自该表。

[0547] 已经发现许多Cas13直向同源物以共同的基序为特征。因此，在特定实施方式中，Cas13b效应蛋白是包含与由DKHXFGAFLNLRHN (SEQ ID NO:96)，GLLFFVSLFLDK (SEQ ID NO:97)，SKIXGFK (SEQ ID NO:98)，DMLNELXRCP (SEQ ID NO:99)，RXZDRFPYFALRYXD (SEQ ID NO:100) 和LRFQVBLGXY (SEQ ID NO:101) 组成的序列中的一个或多个具有至少70%序列同一性的序列的蛋白质。在进一步的特定实施方式中，Cas13b效应蛋白包含具有与这序列中的至少2、3、4、5或全部6个至少70%序列同一性的序列。在另外的特定实施方式中，与这些序列的序列同一性为至少75%，80%，85%，90%，95%或100%。在另外的特定实施方式中，Cas13b效应蛋白是包含与GLLFFVSLFL (SEQ ID NO:102) 和RHQXRFYPY (SEQ ID NO:103) 具有100%序列同一性的序列的蛋白。在另外的特定实施方式中，Cas13b效应子是Cas13b 效应子蛋白，其包含与RHQDRFPY (SEQ ID NO:104) 具有100%序列同一性的序列。

[0548] 在特定实施方式中，Cas13b效应蛋白是与来自颊普雷沃菌，牙龈卟啉单胞菌，解糖丙酸杆菌，抗黎氏杆菌的Cas13b蛋白具有至少65%，优选至少70%，75%，80%，85%，90%，95%或更高的序列同一性的Cas13b 效应蛋白。在进一步的特定实施方式中，Cas13b效应子选自酿脓拟杆菌 (*Bacteroides pyogenes*)，普氏菌种 (*Prevotella* sp.) MA2016，鸭疫里默氏杆菌 (*Riemerella anatipestifer*)，喉管卟啉单胞菌 (*Porphyromonas gulae*)，牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) 和 *Porphyromonas* sp. COT-0520H4946 的Cas13b蛋白。

[0549] 应当理解，可以在本发明范围内的表1Cas13b酶的直系同源物可以包括包含多个直系同源物的表1Cas13b酶的片段的嵌合酶。这样的直系同源物的例子在本文其他地方描述。嵌合酶可包含表1Cas13b酶的片段和来自另一种CRISPR酶的片段，例如生物体的表1Cas13b酶的直系同源物，该生物体包括但不限于 *Bergeyella*，*Prevotella*，*Porphyromonas*，*Bacteroides*，*Alistipes*，*Riemerella*，*Myroides*，*Flavobacterium*，*Capnocytophaga*，*Chryseobacterium*，*Phaeodactylibacter*，*Paludibacter* 或 *Psychroflexus*。嵌合酶可以包含第一片段和第二片段，以及所述片段，其中所述第一片段和第二片段中的一个片段是或来自表1的Cas13b酶的，而另一个片段是或来自不同物种的CRISPR酶直系同源物的。在某些情况下，Cas13b是Cas13b-t。例如，Cas13b可以是Cas13b-t1 (例如，Cas13b-t1a)，Cas13b-t2或Cas13b-t3 (例如，参见图54A-54C)。

[0550] 在实施方式中，本文所指的靶向Cas13 RNA的Cas13效应蛋白也包括效应蛋白的功能变体或其同源物或直向同源物。如本文所用，蛋白质的“功能变体”是指这种蛋白质的变体，其至少部分地保留了该蛋白质的活性。功能性变体可以包括突变体 (可以是插入，缺失或替代突变体)，包括多晶型物等，包括如本文结合表1所讨论的。功能性变体中还包括的是这样的蛋白质与另外的、通常不相关的核酸，蛋白质，多肽或肽的融合产物。功能变体可以是天然存在的，也可以是人造的。在一个实施方式中，可以对编码Cas13 RNA靶向效应蛋白或其直系同源物或同源物的核酸分子进行密码子优化以在真核细胞中表达。真核生物可以如本文所述。核酸分子可以被工程化或非天然存在。

[0551] 在一个实施方式中，靶向Cas13 RNA的效应蛋白或其直系同源物或同源物可包含

一个或多个突变。所述突变可以是人为引入的突变,并且可以包括但不限于催化域中的一个或多个突变,例如,将一个或多个突变引入一个或多个HEPN结构域。

[0552] 在一个实施方式中,Cas13蛋白或其直系同源物或同源物可以用作与功能域融合或可操作地连接的通用核酸结合蛋白。示例性功能结构域可包括但不限于翻译引发剂,翻译活化剂,翻译阻遏物,核酸酶,特别是核糖核酸酶,剪接体,珠,光诱导/可控结构域或化学诱导/可控结构域。

[0553] 在一个有利的实施方式中,本发明包括表1-5提到的Cas13效应蛋白。在某些示例性实施方式中,Cas13效应蛋白来自表1-5中鉴定的生物。在某些示例性实施方式中,Cas13效应蛋白来自选自以下的生物体: *Bergeyella zoohelcum*, *Prevotella intermedia*, *Prevotella buccae*, *Porphyromonas gingivalis*, *Bacteroides pyogenes*, *Alistipes* sp.ZOR0009, *Prevotella* sp.MA2016, *Riemerella anatipestifer*, *Prevotella aurantiaca*, *Prevotella saccharolytica*, *Myroides odoratimimus* CCUG 10230, *Capnocytophaga canimorsus*, *Porphyromonas gulae*, *Prevotella* sp.P5-125, *Flavobacterium branchiophilum*, *Myroides odoratimimus*, *Flavobacterium columnare*, 或 *Porphyromonas* sp.COT-0520H4946。在另一个实施方式中,一种或多种指导RNA被设计成结合一种或多种诊断疾病状态的靶 RNA序列。

[0554] 在某些示例性实施方式中,CRISPR效应蛋白是选自表1的Cas13b 蛋白。

[0555] 表1

|   |   |   |
|---|---|---|
| [0556] 动物溃疡伯格菌<br>( <i>Bergeyella zoohelcum</i> )<br>(SEQ ID No. 105) | I | MENKTSLGNNIYNPFKPQDKSYFAGYFNAAMENTDSVFRELGKRLKGG<br>EYTSENFFDAIFKENISLVEYERYVKLLSDYFPMARLLDKKEVPIKERKEN<br>FKKNFKGIKAVRDLRNFYTHKEHGEVEITDEIFGVLDMLKSTVLTVKK<br>KKVKTDKTKEILKKSIEKQLDILCQKKLEYLRDTARKIEEKRRNQRRERGE<br>KELVAPFKYSDKRDDLIAAIYNDAFDVYIDKKKDSLKESSKAKYNTKSDP<br>QQEEGDLKIPISKNGVVFLLSLFLTKQEIHAFKSKIAGFKATVIDEATVSEA<br>TVSHGKNSICFMATHEIFSHLAYKCLKRKRVRTAEINYGEAENAEQLSVYA |
|---|---|---|

[0557]

|   |          |   |
|---|----------|---|
|   |          | <p>KETLMMQMLDELKVPDVVYQNLSEVDVQKTFIEDWNEYLKENNGDVGT<br/> MEEEEQVIHPVIRKRYEDKFNYFAIRFLDEFAQFPTLRFQVHLGNLHDSRP<br/> KENLISDRRIKEKITVFGRLSELEHKKALFIKNTETNEDREHYWEIFPNPNY<br/> DFPKENISVNDKDFPIAGSILDREKQPVAGKIGIKVLLNQYVSEVDKAV<br/> KAHQLKQRKASKPSIQNIIIEIVPINESNPKEAIVFGGQPTAYLSMNDIHSIL<br/> YEFFDKWEKKKEKLEKKGEKELRKEIGKELEKKIVGKIQAQIQQIIDKDTN<br/> AKILKPYQDGNSTAIKKEKLIKDLKQEQNILQKLDKDEQTVREKEYNDFIA<br/> YQDKNREINKVRDRNHKQYLKDNLKRKYPEAPARKEVLYYREKGGKVA<br/> VWLANDIKRFMPTDFKNEWKGEQHSLLQKSLAYYEQCKEELKNLLPEK<br/> VFQHLPPFKLGGYFQKQYLYQFYTCYLDKRLEYISGLVQQAENFKSENKV<br/> FKKVENECFKFLKKQNYTHKELDARVQSILGYPIFLERGFMDKPTIIGKVK<br/> TFKGNEALFADWFRYYKEYQNFQTFYDTENYPLVELEKKQADRKRKTKI<br/> YQQKKNVFTLLMAKHIFKSVFKQDSIDQFSLEDLYQSREERLGNQERAR<br/> QTGERNTNYIWNKTVDLKLCDGKITVENVKLKNVGDFFIKYEYDQQRVQAF<br/> LKYEENIEWQAFLIKESKEEENYPYVVEREIEQYEVRRREELLKEVHLLIEE<br/> YILEKVKDKKILKKGDNQNFKYIILNGLLKQLKNEDEVESYKVFNLNTEPE<br/> DVNINQLKQEAATDLEQKAFVLTIRNKFVHNQLPKKEFWDYCQEKYGKI<br/> EKEKTYAEYFAEVFKKEKEALIK</p>   |
| <p>中间普雷沃<br/> 菌(Prevotella<br/> intermedia)<br/> (SEQ ID No.<br/> 106)</p>                      | <p>2</p> | <p>MEDDKKTTDSIRYELKDKHFWAAFLNLARHNVYITVNHINKILEEGEINR<br/> DGYETTLKNTWNEIKDINKDRLSKLIKHFPFLEAATYRLNPTDTTKQKE<br/> EKQAEAQSLESRLKSFVFIYKLRDLRNHSHYKHSKSLERPKFEEGLLEK<br/> MYNIFNASIRLVKEDYQYNKDINPDEDKHLDRTEEFNYFYTKDNENGI<br/> TESGLLFFVSLFLEKKDAIWMQKLRGFKDNRENKKKMTNEVFCRSRML<br/> LPKLRQSTQTQDWILLDMLNELIRCPKSLYERLREEDREKFRVPIEIADE<br/> DYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLRQIDLGTYHFSIY<br/> KKQIGDYKESHHLTHKLYGFERIQEFTKQNRPDWRKFKVTFNSFETSKE<br/> PYIPETTPHYHLENQKIGIRFRNDNDKIWPSLKTNSEKNEKSKYKLDKSFQ<br/> AEAFLSVHELLPMMFYLLKKTENTDNDNEIETKKKENKNDKQEKHKIE<br/> EIIENKITEIYALYDTFANGEIKSIDELECYCKGKDIEIGHLPKQMIAILKDE<br/> HKVMATEAERKQEEMLVDVQKSLESLDNQINEEIEENVERKNSSLSKSGKIA<br/> SWLVNDMMRFQPVQKDNQKPLNNSKANSTEYQLLQRTLAFFGSEHER<br/> LAPYFKQTKLISSNPHFPLKDTWEKCNNILSFYRSYLEAKNFLESKLP<br/> EDWEKNQYFLKLEPKTKPKTLVQGWKNGFNLRGIFTEPIRKFWMKHR<br/> ENITVAELKRVGLVAKVIPLFSEYKDSVQPFYNYHFNVGNINKPDEKN<br/> FLNCEERRELLRKKKDEFKMTDKEKEENPSYLEFKSWNKFRELRLVR<br/> NQDIVTWLLCMELFNKKIKELNVEKIYLNINTNTTKKEKNTEEKNGEE<br/> KNIKEKNLNRIMPRLPIKVYGRENFKNKKKIRNRTFTVYIEEKGT<br/> KLLKQGNFKALERDRRLGGLFSFVKTPSKAESKNTISKLRVEYELGEYQ<br/> KARIEIKDMLALEKTLIDKYNSLDTDNFNKMLTDWLELKGEPDKASFQN<br/> DVDLLIAVRNAFHNQYPMRNRIAFANINPFLSSANTSEEKGLGIANQLK<br/> DKTHKTIEKIIIEKPIETKE</p> |
| <p>颊普雷沃菌<br/> (Prevotella<br/> buccae)<br/> (SEQ ID No.<br/> 107)<br/> WP_004343<br/> 973.1</p> | <p>3</p> | <p>MQKQDKLFVDRKKNIAIFAPKYITIMENKEKPEPIYYELTDKHFVAAFLN<br/> LARHNVYTTINHINRRLEIAELKDDGYMMGIGKSWNEQAKKLDKKVRLR<br/> DLIMKHFPFLEAAAYEMTNSKSPNNKEQREKEQSEALSNLKNVLFIFL<br/> EKLQVLRNYSHYKYSESPKPIFETSLKNMYKVFANVRLVCRDYM<br/> HENIDMQRDFTHLNRKKQVGRTKNIIDSPNFHYHFADKEGNMTIAGLLFF<br/> VSLFLDKKDAIWMQKLLKGFKDGRLREQMTNEVFCRSRISLPKLLKEN<br/> VQTKDWMQLDMLNELVRCPSLYERLREKDRSFKVPFDIFSDDYNAEE<br/> EPFKNTLVRHQDRFPYFVLRVYFDLNEIFEQLRFQIDLGTYHFSIYNKRIGDE<br/> DEVRLHLLHLYGFARIQDFAPQNPQPEWRKLVKDLDFHETSQEPYISKTA<br/> PHYHLENEKIGIKFCSAHNNLFPQLQTDKTCNGRSKFNLTGTFQTAEAFLSV<br/> HELLPMMFYLLKTDYSRKESADKVEGIIRKEISNIYAIYDAFANNEINSI<br/> ADLTRLQNTNILQGHLPKQMSILKGRQKDMGKEAERKIGEMIDDTQRR<br/> LDLLCKQTNQKIRIGKRNAGLLKSGKIADWLVNDMMRFQPVQKQNNIP<br/> INNSKANSTEYRMLQRALALFGSENFRLKAYFNQMNVLVGNNDNPHPLAE<br/> TQWEHQTNILSFYRNYLEARKKYLKGLKPQNWKQYQHFLILKVQKTNR</p>  |

[0558]

|  |   |   |
|--|---|---|
|  |   | <p>NTLVGTGWKNSFNLRGIFTQPIREWFEKHNNKRIYDQILSFDRVGFVAK<br/>                 AIPLYFAEEYKDNVQPFYDYPFNIGNRLKPKKRQFLDKKERVELWQKNK<br/>                 ELFKNYPSEKKTDLAYLDFLSWKKFERELRLIKNQDIVTWLMFKELFNM<br/>                 ATVEGLKIGEIHLRDIDTNTANEESNNILNRIMPMLKLPVKTYETDNKGNIL<br/>                 KERPLATFYIEETETKVLKQGNFKALVKDRRLNGLFSFAETTDLNLEEHP<br/>                 SKLSVDLELIKQYTTTRISIFEMTLGLEKKLIDKYSTLPTDSFRNMLERWLQC<br/>                 KANRPELKNYVNSLIAVRNAFVSHNQYPMYDATALFAEVKKFTLFPVSDTK<br/>                 KIELNIAPQLEIVGKAIKEIEKSENKN</p>   |
| <p>牙龈卟啉单胞菌<br/>                 (Porphyromonas gingivalis)<br/>                 (SEQ ID No. 108)</p> | 4 | <p>MNTVPASENKQSRTEVEDDPQYFGLYLNLAARENLIEVESHVRIKFGKKKL<br/>                 NEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSK<br/>                 TGVDPDSAQRLIRELYSLDFLRNDFSHNRLDGTTFEHLEVSPDISSFITGT<br/>                 YSLACGRAQSRFAVFFKPDDFVLAKNRKEQLISVADGCECLTVSGFAFI<br/>                 CLFLDREQASGMLSRRIRGFKRTDENWARAVHETFLCDLCIRHPHDLRESSN<br/>                 TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLSESLD<br/>                 EESRLLWDGSSDWAALTKRIRHQDRFPYMLRFEIEMDLLKIRFRVDL<br/>                 GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br/>                 YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV<br/>                 HDLRKLLMELLCEGSFSRMQSDFLRKANRILDETAEGKLFQFSALFPEMR<br/>                 HRFIPPQNPKSKDRREKAETTLEKYKQEIKGRKDKLNSQLLSAFDMDQRQ<br/>                 LPSRLLEDEWMNIRPASHSVKLRITYVKQLNEDCRLRLRKRFRKDGDKARA<br/>                 IPLVGEMATFLSQDIVRMIIESETKKLITSAAYNEMQRSQAQYAGEENRRQ<br/>                 FRAIVAEARLLDPSSGHPFLSATMETAHRYTEGFYKCYLEKKREWLAKIF<br/>                 YRPEQDENTKRRISVFFVDPDGEARKLLPLLIRRRMKEQNLDQDWIRNKQA<br/>                 HPIDLPSHLFDSKVMELLKVKDGGKKWNEAFKDWSTKYPDGMQPFYGL<br/>                 LRRELNIHGKSVSYIPSDGKGFADCYTHLMEKTVRDKKRELRTAGKPVPP<br/>                 DLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPG<br/>                 LKNIDSILDEENQFSLAVHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWS<br/>                 KYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLEGEYDRCRIKIFDWAFALE<br/>                 GAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILI<br/>                 RNKAAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWK<br/>                 KYEMIIRKILPILDPENRFFGKLLNMSQPINDL</p> |
| <p>酿脓拟杆菌<br/>                 (Bacteroides pyogenes)<br/>                 (SEQ ID No. 109)</p>       | 5 | <p>MESIKNSQKSTGKTLQKDPYFGLYLNLMALLNVRKVENHIRKWLGDVAL<br/>                 LPEKSGFHSLTTDNLSSAKWTRFYKSRKFLPFLFEMFSDSKSYENRRE<br/>                 TAECLDTIDRQKISSLLKEVYKGLQDIRNAFPHYHIDDQSVKHTALISSEM<br/>                 HRFIENAYSFALQKTRARFTGVFVETDFLQAEKGDNKKFFAIGGNEGKIK<br/>                 LKDNALIFLICFLDREEAFFKFLSRATGFKSTKEKGF LAVRETFCALCCRQP<br/>                 HERLLSVNPREALLMDMLNELNRCPDILFEMLDEKDQKSFLLPGLGEEQA<br/>                 HILENSLDELCEAIDDPFEMIASLSKRVRKYKRFYMLRLRYIEEKNLLPFI<br/>                 RFRIDLGCLELASYPKKMGEENNYERSVTDHMAAFGRITDFHNEAVLQ<br/>                 QITKGITDEVRFSLYAPRYAIYNNKIGFVRTSGSDKISFPTLKKKGEGHC<br/>                 VAYTLQNTKSFYFISYDLRKLILLSFLDKDKAKNIVSGLLEQCEKHWKDL<br/>                 SENLFDAIRTELQKEFPVPLIRYTLPRSKGGKLVSSKLADKQEKYSEFER<br/>                 RKEKLTEILSEKDFDLSQIPRRMIDEWLNVLPTSREKLLKGYVETLKLDCR<br/>                 ERLRVFEKREKGEHPLPRIGEMATDLAKDIIRVIDQGVKQRTSAYSE<br/>                 IQRCLAQYAGDNRRLHDSIIRELRLKDTKNGHPFLGKVLPRPGLGHEKIL<br/>                 YQRYFEEKKEWLEATFYPAASPKRVPRFVNPTGKQKELPLIIRNLMKER<br/>                 PEWRDVKQRKNSHPIDLPQLFENEICRLLKDKIGKEPSGKLLWNEMFKL<br/>                 YWDKEFPNGMQRFYRCKRRVEVFDKVVVEYSEEGGNYKYYEALIDE<br/>                 VVRQKISSSKEKSKLQVEDLTVSRRVFKRAINEKEYQLRLLCEDDRLLF<br/>                 MAVRDLYDWKEAQLDLKIDNMLGEPVSVSQVIQLEGGQPDVAVIKAE<br/>                 KLKDVSKLMRYCYDGRVKGMLMPYFANHEATQEQVEMELRHVEDHRRR<br/>                 VFNWVFALEKSVLKNEKLRFRFYEESSQGGCEHRRCIDALRKASLVSEEEYE<br/>                 FLVHIRNKS AHNQFPDLEIGKLPNVTSGFCECIWSKYKAIICRIIPFIDPER<br/>                 RFFGKLLLEQK</p>                               |
| <p>别样杆菌<br/>                 (Alistipes)</p>   | 6 | <p>MSNEIGAFREHQFAYAPGNEKQEEATFATYFNLALS NVEGMMFGEVESN<br/>                 PDKIEKSLDTPAILRQIASFIWLSKEDHPDKAYSTEEVKVIVTDLVRRLC</p>   |

[0559]

|   |    |   |
|---|----|---|
| sp.)<br>ZOR0009<br>(SEQ ID No.<br>110)                    |    | FYRNYFSHCFYLDTQYFYSDDELVDTTAIGEKLPYNFHHFITNRLFRYSLPEI<br>TLFRWNEGERKYEILRDGLIFFCCLFLKRGQAERFLNELRFFKRTDEEGRI<br>KRTIFTKYCTRESHKHIGIEEQDFLIFQDIIGDLNRVPKVCDCGVVDSLKENE<br>RYIKNRETSNESDENKARYRLIREKDKFPYYLMRYIVDFGVLPICITFKQN<br>DYSTKEGRGQFHYQDAVAQEERCYNFVVRNGNVYYSYMPQAQNVVR<br>ISELQGTISVEELRNMVYASINGKDVNKSVEQYLYHLHLLYEKILTISGQTI<br>KEGRVDVEDYRPLLDKLLRPASNGEELRRELKLLPKRVCDLLSNRFDC<br>SEGVSAREKRLKAILLRHEQLLSQNPALHIDKIKSVIDYLYLFFSDDEKFR<br>QQPTEKAHRGLKDEEFQMYHYLVGDYDSHPLALWKELEASGRKLPEMR<br>KLTSATSLHGLYMLCLKGTVEWCRKQLMSIGKGTAKVEAIADRVLGLKY<br>DKLKEYTPEQLEREVKLVVMHGYAAAATPKPKAQAAIPSKLTELRFYSF<br>LGKREMSFAAFIRQDKKAQKLWLRNFYTVENIKTLQKRQAAADAACKK<br>LYNLVGEVERVHTNDKVLVLAQRYRERLLNVGSKCAVTLDNPERQOK<br>LADVYEVQNAWLSIRFDDLDFTLTHVNLSNLRKAYNLIPRKHILAFKEYL<br>DNRVKQKLCEECRNVRKEDLCTCCSPRYSNLTSWLKENHSESSIEREAA<br>TMMLLDVERKLLSFLDERRKAIIEYGKFPFSALVKECRLADAGLCGIRN<br>DVLHDNVISYADAIGKLSAYFPKEASEAVEYIRRTKEVREQRREELMANS<br>SQ  |
| 普雷沃菌<br>(Prevotella<br>sp.) MA2016<br>(SEQ ID No.<br>111) | 7a | MSKECKKQRQEKKRRLQKANFSISLTGKHVFGAYFNMARTNFVKTINYIL<br>PIAGVRGNYSENQINKMLHALFLIQAGRNEELTTEQKQWEKKLRLNPEQ<br>QTKFQKLLFKHFPVLGPMMAADVADHKAYLNKKKSTVQTEDETFAMLKG<br>VSLADCLDIICLMADTLTECRNFYTHKDPYNKPSQLADQYLHQEMIACK<br>LDKVVVASRRILKDREGLSVNEVEFLTGDHHLHQEVLKDEFGNAKVKDG<br>KVMKTFVEYDDFYFKISGKRLVNGYTVTTKDDKPVNVNTMLPALSDFGL<br>LYFCVFLFSKPYAKLFIDEVRLFEYSPFDDKENMIMSEMLSIYRIRTPRLHK<br>IDSHDSKATLAMDFGELRRCPEMELYNLLDKNAGQPFFHDEVKHPNSHTP<br>DVSKRLRYDDRFPPTALRYIDETELFKRIRFQLQLGFSFRYKFDKENCIDG<br>RVRVRRIQKEINGYGRMQEVADKRMDKWGDLIQKREERSVKLEHEELYI<br>NLDQFLEDTADSTPYVTDRRPAYNIHANRIGLYWEDSQNPQKYKVFEN<br>GMYIPELVVTEDEKKAIPKMPAPRCALSVYDLPAMLFYEYLREQQDNEFP<br>SAEQVIEYEDDYRKFKAVAEGKLPFKRPKEFRDFLKKEYPKLRMADIP<br>KKLQFLCASHGLCYNNKPEVYERLDRLTLQHLEERELHIQNRLEHYQKD<br>RDMIGNKDNQYGKKSFSVDRHGALARYLAQSMMEVQPTKLDKKEKGH<br>DKLTGLNYNVLTAYLATYGHQPVEEGFTPTLQHLINAHLLIGGSNPHF<br>FINKVLALGNRNIEELYLHYLEEELKHIRSRIQSSNPDKALSALPFIHH<br>DRMRYHERTSEEMMALAARYTTIQLPDGLFTPYILEILQKHYTENSIDLQN<br>ALSQDVPVKNLPTCNAAYLITLFYQTVLKDNAQPFYLSDKTYTRNKDGE<br>KAESFSFKRAYELFVNLNNKKTFFPFEMIPFLTSDEIQERLSAKLLDGD<br>GNPVPEVGEKGPATDSQGNITWKRRIYSEVDDYAEKLTDRDMKISFKG<br>EWEKLPRWKQDKIKRRDETRRQMRDELLQRMPRYIRDIDKNERTLRRY<br>KTQDMVLFLLAEKMFTNIISEQSSEFNWKQMRLSKVCNEAFLRQTLTFRV<br>PVTVGETTIYVEQENMSLKNYGEFYRFLTDRLMSLLNIVETLKPENEG<br>DLVIRHTDLMSELAAVDQYRSTIFMLIQSIENLIITNNAVLDDPADGFWV<br>REDLPKRNNFASLLELINQLNNVELTDDERKLLVAIRNAFSHNSYNIDFSLI<br>KDVKHLPEVAKGILQHLQSMGLGVEITK |
| 普雷沃菌<br>MA2016<br>(SEQ ID No.<br>112)                     | 7b | MSKECKKQRQEKKRRLQKANFSISLTGKHVFGAYFNMARTNFVKTINYIL<br>PIAGVRGNYSENQINKMLHALFLIQAGRNEELTTEQKQWEKKLRLNPEQ<br>QTKFQKLLFKHFPVLGPMMAADVADHKAYLNKKKSTVQTEDETFAMLKG<br>VSLADCLDIICLMADTLTECRNFYTHKDPYNKPSQLADQYLHQEMIACK<br>LDKVVVASRRILKDREGLSVNEVEFLTGDHHLHQEVLKDEFGNAKVKDG<br>KVMKTFVEYDDFYFKISGKRLVNGYTVTTKDDKPVNVNTMLPALSDFGL<br>LYFCVFLFSKPYAKLFIDEVRLFEYSPFDDKENMIMSEMLSIYRIRTPRLHK<br>IDSHDSKATLAMDFGELRRCPEMELYNLLDKNAGQPFFHDEVKHPNSHTP<br>DVSKRLRYDDRFPPTALRYIDETELFKRIRFQLQLGFSFRYKFDKENCIDG<br>RVRVRRIQKEINGYGRMQEVADKRMDKWGDLIQKREERSVKLEHEELYI<br>NLDQFLEDTADSTPYVTDRRPAYNIHANRIGLYWEDSQNPQKYKVFEN   |

[0560]

|   |   |   |
|---|---|---|
|   |   | GMYIPELVVTEDEKKAIPKMPAPRCALSVYDLPAMLFYEYLREQQDNEFP<br>SAEQVIEYEDDYRKFFKAVAEGKLPFKRPKEFRDFLKKEYPKLRMADIP<br>KKLQFLCASHGLCYNNKPETVYERLDRLTLQHLEERLHIQNRLEHYQKD<br>RDMIGNKDNQYGKKSFSQDVRHGALARYLAQSMMEWQPTKLDKEKGGH<br>DKLTGLNYNVLTAYLATYGHPQVPEEGFTPTRTLEQVLINAHLIGGSNPHP<br>FINKVLALGNRNIEELYLHYLEEELKHRSRIQSLSSNPSDKALSALPFIHH<br>DRMR YHERTSEEMMALAARYTTIQLPDGLFTPYILEILQKHYTENSDLQN<br>ALSQDVPVKLNPTCNAAYLITLFYQTVLKDNAQPFYLSDKTYTRNKDGE<br>KAESFSFKRAYELFVSNNNKKTDFPFEMIPFLTSDEIQRLSAKLLDGD<br>GNPVPEVGEKGPATDSQGNITWKRRIYSEVDDYAEKLTDRDMKISFKG<br>EWEKLPWKQDKIKRRDETRRQMRDELLQRMPIYIRDIKDNERTLRRY<br>KTQDMVFLFLAEKMFNTNIISEQSSEFNWKQMRLSKVCNEAFLRQTLTFRV<br>PVTVGETTIYVEQENMSLKNYGEFYRFLTDDRMSLLNNIVETLKPENEG<br>DLVIRHTDLMSELAAAYDQYRSTIFMLIQSIENLIITNNAVLDDPADAGFWV<br>REDLPKRNNFASLLELNLNNVELTDDERKLLVAIRNAFVSHNSYNIDFSLI<br>KDVKHLPEVAKGILQHLQSMGLGVEITK  |
| 鸭疫里默氏<br>杆菌<br>( <i>Riemerella<br/>anatipestifer</i> )<br>(SEQ ID No.<br>113) | 8 | MEKPLLPNVYTLKHKFFWGAFLNIARHNAFITICHINEQLGLKTPSNDDKI<br>VDVVCETWNNILNNDHDLKKSQTELELKHFPFLTAMCYHPPKKEGKK<br>KGHQKEQQKEKESEAQSQAALNPSKLIIEALEILVNQLHSLRNYYSHYKH<br>KKPDAEKDIFKHLKAFDASLRMVKEDYKAHFTVNLTRDFAHLNRKGGK<br>NKQDNPDFNRYRFEKDGFFTESGLLFFTNLFLDKRDAYWMLKKVSGFKA<br>SHKQREKMTTEVFCRSRILLPKLRLESRYDHNQMLLDMLSELSRCPKLLY<br>EKLSEENKHKHFQVEADGFLDEIEEEEQNPFKDTLIRHQDRFPYFALRYDLN<br>ESFKSIRFQVDLGTYHYCIYDKKIGDEQEKRHRLTRTLLSFGRLQDFTEINRP<br>QEWKALTKDLDYKETSQPFISKTPHYHITDNKIGFRLGTSKELYPSLEI<br>KDGANRIAKYPYNSGFVAHAFISVHELLPLMFYQHLTGKSEDLKETVRH<br>IQRIYKDFEEERINTIEDLEKANQGRPLGAFPKQMLGLLQNKQPDSEKA<br>KIKIEKLIETKLLSHRLNKLKSSPKLGKRREKLIKTVLADWLKDFMR<br>FQPVAAYDAQNQPIKSSKANSTEFWFIRRALALYGGENRLEGYFKQTNLI<br>GNTNPHPFLNKFVWKACRNLVDFYQQYLEQREKFLAIAKNQPWEPEYQY<br>CLLLKIPKENRKNLVKQWEQGGISLPRGLFTEAIRETLSEDLMLSKPIRKEI<br>KKHGRVGFISRAITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKREEHYEY<br>WQQNKQSPTESQRLELHTSDRWKDYLLYKRWQHEKCLRRLRYNQDV<br>MLWMLTLELTKNHFKELNLNYHQLKLENLAVNVQEAADAKLNPLNQTL<br>MVLVPKVYPATAFGEVQYHKTPRTVYIREEHTKALKMGNFKALVKDRR<br>LNGLFSFIKEENDTQKHPISQLRRLRRELEIYQSLRVDAFKETLSLEEKLLNK<br>HTSLSSLENEFRALLEEWKKEYAASSMVTDEHIAFIASVRNAFCHNQYPF<br>YKEALHAPIPLFTAQPTTEEKDGLGIAEALLKVLREYCEIVKSQI |
| 桔红色普雷<br>沃菌<br>( <i>Prevotella<br/>aurantiaca</i> )<br>(SEQ ID No.<br>114)    | 9 | MEDDKTTGSISYELKDKHFWAAFLNLARHNVYITINHINKLEIREIDND<br>EKVLDIKTLWQKGNKDLNQAARLRELMTHFPFLETAIYTKNKEDKKEV<br>KQEKQAEASLESKDCFLFLDKLQEARNYSHYKYSEFSKEPEFEGL<br>LEKMYNIFGNINQLVINDYQHNDINPDEDFKHLDRKGQFKYSFADNEG<br>NITESGLLFFVSLFLEKDAIWMQQKLNGLFKDNLENKKKMTHVFCRSRI<br>LMPKLRLESTQTDWILLDMLNELIRCPKSLYERLQGGDDREKFKVFPDPA<br>DEDYNAEQEPFKNTLIRHQDRFPYFVLRVYFDYNEIFKNLRFQIDLGTYHFS<br>IYKKLIGGQKEDRHLTHKLYGFERIQEFAKQNRPEDEWKAIKVDLDYETS<br>NKRYISETTPHYHLENQKIGIRFRNGNKEIWPSTKNDENNEKSKYKLDK<br>QYQAEAFLSVHELLPMMFYLLKKEKPNNDENASIVEGFIKREIRNIFK<br>LYDAFANGEINNIDDLEKYCADKGIPKRHLPKQMVAILYDEHKDMVKEA<br>KRKQKEMVKDTKKLLATLEKQTQKEKEDDGRNVKLLKSGEIARWLVND<br>MMRFQPVQKDNEGKPLNNSKANSTEQMLQRSLALYNNEEKPTRYFRQ<br>VNLIESNNPHPFLKWTKEECNNILTFYYSYLTKKIEFLNKLKPEDWKKN<br>QYFLKKEPKTNRETQVQGWKNGFNLPRGIFTEPIREWFKRHQNSKEYE<br>KVEALDRVGLVTKVIPLFFKKEYFKDKEENFKEDTQKEINDCVQPFYFNF<br>YNVGNIIHKPEKDFLHREERIELWKKKDKFKGYKEKIKSKKLEKDKKE<br>EFRSYLEFQSWNKFERELRLVRNQDIVTWLCKELIDKLIKIDELNIEELKK  |

[0561]

|   |    |  |
|---|----|--|
|   |    | LRLNNIDTDTAKKEKKNILNRVMPMELPVTVVEIDDSHKIVKDKPLHTIYI<br>KEAETKLLKQGNFKALVKDRRLNGLFSFVKTNSEAESKRNPISKLRVEYE<br>LGEYQEARIEI IQDMLALEEKLINKYKDLPTNKFSEMLNSWLEGKDEADK<br>ARFQNDVDFLIAVRNAFVSHNQYPMHNKIEFANIKPFSLYTANNSEEKGLG<br>IANQLKDKTKETTDKIKKIEKPIETKE   |
| 解糖普雷沃<br>菌(Prevotella<br>saccharolytic<br>a) (SEQ ID<br>No. 115)  | 10 | MEDKPFWAFFNLRHNVYLVTVNHINKLLDLEKLYDEGKHKEIFEREDIF<br>NISDDVMNDANSNGKKRKLDIKKIWDLDLTLRKYQLRELILKHFPIQ<br>PAIIGAQTKERTTIDKDKRSTSTSNDSLKQTGEGDINDLLSLSNVKSMFFRL<br>LQILEQLRNYYSHVKHSKATMPNFDEDLNWMRYIFIDSVNKVKEDYS<br>SNSVIDPNTSFSHLIYKDEQGGIKPCRYPFSTKDGSIKAFGLLFFVSLFLEKQ<br>DSIWMQKIPGFKKASENYMKMTNEVFCRNHILLPKIRLETVYDKDWML<br>LDMLNEVVRCPKSLYKRLTPAAQNKFKVPEKSSDNANRQEDDNPFSRIL<br>VRHQNRFPYFVLRFFDLNEVFTTLRFQINLGCYHFAICKQIGDKKEVHH<br>LIRTYGFSRLQNFQNTREPEEWNTLVKTTEPSSGNDGKTVQGVPLPYIS<br>YTIPHYQIENEKIGIKIFDGDTA VDTDIWPSVSTEKQLNKPKDYTLTPGFK<br>ADVFLSVHELLPMMFYQLLLCEGMLKTDAGNAVEKVLIDTRNAIFNLY<br>DAFVQEKINTITDLENYLQDKPILIGHLPKQ MIDLLKGHRDMLKAVEQK<br>KAMLIKDTERRKLLDKQLKQETDVA AKNTGTLKNGQIADWL VNDM<br>MRFQPVKRDKEGNPINCSCANSTEYQMLQRAFAFYATDSCRLSRYFTQL<br>HLIHSNLSHLFSRFEYDKQPNLIAFYAAYLKAKLEFLNELQPQNWASDN<br>YFLLLAPKNDRQKLAEGWKNGFNLPRGLFTEKIKTFWNEHKTIVDISDC<br>DIFKNRVGQVARLIPVFFDKKFKDHSQPFYRYDFNVGNVSKPTEANYLSK<br>GKREELFKSYQNKFKNIPAEKTKYREYKNFSLWKKFERELRIKLNQDI<br>LIWLMCKNLFDEKIKPKKDILEPRIAVSYIKLDSLQNTSTAGSLNALAKV<br>VPMTLAIHIDSPKPKGKAGNNEKENKEFTVYIKEEGTKLLKWNFKTLLA<br>DRRIKGLFSYIEHDDIDLKQHPLTKRRVDLELDLYQTCRIDIFQQTGLEA<br>QLLDKYSIDLNTDNFYQMLIGWRKKEGIPRNIKEDTDFLKDVRNAFVSHNQ<br>YPDSKKIAFRIRKFNPKELILEEEEEGLGIATQMYKEVEKVVNRKRIELFD                   |
| HMPREF97<br>12_03108<br>[拟香味类香<br>味菌<br>(Myroides<br>odoratimimu<br>s) CCUG<br>10230] (SEQ<br>ID No. 116) | 11 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEEVNKRNT<br>FGKLAKRDNGLKNYIIHVFKDELSISDFEKRVAIFASYFPILETVDKKS<br>ERNRTIDLTLRQRIRQFREMLISLVTA VDQLRNFYTHYHSDIVIENKVL<br>FLNSSFVSTALHVKDKYLKTDKTKEFLEKTI AAEILDILIEAYKKKQIEKKN<br>TRFKANKREDILNAIYNEAFWSFINDKDKDKDKETVVAKGADAYFEKNH<br>HKSNDPDFALNISEKGIYLLSFFLTNKEMDSLKANLTGFKGKVDRESGN<br>SIKYMATQRIYSFHTYRGLKQKIRTSEGVKETLLMQMIDELSKVPNVVY<br>QHLSTTQQNSFIEDWNEYKYDYEDDVETDDLSRVIHVPVIRKRYEDRFNYF<br>AIRFLDEFFDPTLRFQVHLGDYVHRRRTKQLGKVESDRIIEKVTVFARL<br>KDINSKASYFHSLEEQDKEELDNKWTLPNPSYDFPKEHTLQHQGEQK<br>NAGKIGIYVKLRDTQYKEKAAL EARKSLNPKERSATKASKYDIITQIEA<br>NDNVKSEKPLVFTGQPIAYLSMNDIHSMLFSLT DN AELKKTPEEVEAKLI<br>DQIGKQINEILSKD TD TKILKYYKDN DLKETD TDKITRDLARDK EIEKLIL<br>EQKQRADDYNYTSSTKFNIDKSRKRKHL LFN AEGKIGVWLANDIKRFM<br>FKESKSKWKGYQHTELQKLFAYFDTSKSDLELILSNMVMVKDYPIELIDL<br>VKKSRTLVDLNLKYLEARLEYIENVITRVKNSIGTPQFKTVRKECFTLK<br>SNYTVVSLDKQVERILSMPLFIERGFMDDKPTMLEGKSYQHKEK FADW<br>FVHYKENSNYQNFYDTEVYEITTEDKREKAKVTKKIKQQKNDVFTLM<br>MVNYMLEEVLKLSNDRLSLNELYQTKEERIVNKQVAKDTQERNKNIY<br>WNKVVDLQLCDGLVHIDNVKLDIGNFRKYENDSRVKEFLTYQSDIVWS<br>AYLSNEVDSNKLYVIERQLDNYESIRSKELLKEVQEIECSVYNQVANKES<br>LKQSGNENFKQYVLQGLPIGMDVREMLILSTDVFKKEEIIQLGQAGEV<br>EQDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISDNEYAAYEYMEIFRSI<br>KEKYAN |
| 中间普雷沃<br>菌 (SEQ ID<br>No. 117)  | 12 | MEDDKKTTDSIRYELKDKHFWAAFLNLRHNVYITVNHINKILEEDEINR<br>DGYENTLENSWNEIKDINKDRLSKLIKHFPFLEATTYRQNPTDTTKQKE<br>EKQAEAQSLESLSKSFVFIYKLRDLRNHYSHYKHSKSLERP KFEEDLQN<br>KMYNIFDVS IQFVKEDYKHNTDINPKKDFKHLDRKRKGFHYSFADNEG  |

[0562]

|  |           |   |
|--|-----------|---|
|  |           | <p>NITESGLFFVSLFLEKKDAI WVQK KLEGFKCSNKSYQKMTNEVFCRSRM<br/> LLPKLRLESTQTQDWILLDMLNELIRCPKSLYERLQGVNRKKFYVSFDPA<br/> DEDYDAEQEPFKNTLV RHQDRFPYFALRYFDYNEVFANLRFQIDLGT<br/> YHFSIYK KLI GGQKEDRHLTHKLYGFERIQEFDKQNRPD EWKAI V<br/> KDSDTFKKKEEKEEKPYISETTPHYHLENKKIGIAFKNHNIWPSTQ<br/> TEL TNNKRKKY NLGTSIKAEAFLSVHELLPMMFY YLLLKTENTKND<br/> NKVGGK KETKKQGHKIEAIIESKIKDIYALYDAFANGEINSEDELKEY<br/> LKGKDIKIVHLPKQMIAI LKNEHKDMAEKAEAKQEKMKLATENR<br/> LKTLDKQLKGKI QNGKRYNSAP KSGEIASWLVNDMMRFQPVQK<br/> DENGESLNN SKANSTEYQLLQRTLAF FGSEHERLAPYFKQTKLIE<br/> SSNPHFLNDTEWEKCSNLSFYRSYLKARKNFL ESLKPEDWEKNQY<br/> FLMLKEPKTNRETLVQGWKNGFNLRGFFTEPIRKWFMEHWKSIK<br/> VDDLKRVGLVAKV TPLFFSEKYKDSVQPFYNYPFNVGDVN KP<br/> KEEDFLHREERIELW DKKKDKFKGYKAKKKFKEMTDKEKKEHRS<br/> YLEFQSWNKFERELRLVRNQDIVTWLLCTELIDKLIKIDELNIKEL<br/> KKLRLKDIN TDTAKKEKNILNRVMPMELPVTVYKVNKGGYI IKN<br/> KPLHTIYIKEAETK LLKQGNFKALVKDRRLNGLFSFVKTPSEAESE<br/> SNPISKLRVEYELGKYQN ARLDIIEDMLALEKKLIDKYN SLDT<br/> DNFHNMLTGWLELKG EAKKARFQNDVKLLTAVRNAFSHNQYPMY<br/> DENLFGNIERFSLSSNIIESKGLDIAAKLKEEVSKAAKKIQNEED<br/> NKKEKET</p>  |
| <p>犬咬二氧化碳嗜纤维菌<br/> (Capnocytophaga canimorsus)<br/> (SEQ ID No. 118)</p> | <p>13</p> | <p>MKNIQRLGKGNFSPFKKEDKFFYFGGFLNLANNIEDFFKEIITRFGIVITD<br/> ENKKPKETFGEKILNEIFKKDISIVDYEKWVNIFADYFPFTKYL SLYLEEM<br/> QFKNR VICEFRDVMKELLKTVEALRN FYTHYDHEPIKIEDRVFYFLDKVLL<br/> DVS LTVKNKY LKTDKTK EFLNQHIGEELEKELCKQRKDYL VGKGRKIDKE<br/> SEIINGIYNNAFKDFICKREKQDDKENHNSVEKILCNKEPQNKQKSSATV<br/> WELCSKSSSKYTEKSFNRENDKHCLEVPISQK GIVFLSFFLNKGEIYALT<br/> SNIKGFKAKITKEEPV TYDKNSIRY MATHRMFSFLAYKGLKRKIRTSEINY<br/> NEDGQASSTYEKETLMLQMLDELNKVPDVVYQNLSE DVQKTFIEDWNE<br/> YLKENNGDVGTMEEEQVIHPVIRKRYEDKFNYFAIRFLDEFAQFPTLRFQ<br/> VHLGNYLCDKRTKQICDTTTEREVKKKITVFGRLSELENKKAIFLNEREI<br/> KGWEVFPNPSYDFPKENISVNYKDFPIVGSILDREKQPVSNKIGIRVKIAD<br/> EQREIDKAIKEKLRNPNRKNRANQDEKQKERLVNEIVSTNSNEQGEPVVF<br/> IGQPTAYLSMNDIHSVL YEFLINKISGEALET KIVEKIETQIKQIIGKDATTKI<br/> LKPYTNANSNSINREKLLRDLQEQQILKTLLEEQQREKDKKDKKSKRKH<br/> HELYPSEKGVAVWLANDIKRFMPKAFKEQWRGYHSLDKQLKYLAYEQ<br/> SKEELKNLLPKEVFKHFPFKLKG YFQQYLNQFYTDYLRRLSYVNELL<br/> NIQNFKNDKDALKATEKECFKFFRKQNYIINPINIQISILVYPIFLKRGFLD<br/> EKPTMIDREKFKENKDTLADWFMHYKNYKEDNYQKFYAYPLEKVEEK<br/> EKFKRNKQINKQKKNVYTLMMVEYIIQKIFGDKFVEENPLVLKGIFQSK<br/> AERQQNNTHAATTQERNLNGILNQPDKDIQGITVKGVKLDIGNFRKY<br/> EIDQRVNTFLDYEPKEMAYLPNDWKEKEKQGLPNNVIDRQISKYE<br/> TVRSKILLKDVQELEKIISDEIKEEHRHDLKQGGKYNFKYIILNGLLRQLK<br/> NENVENYKVFKLNTNPEKVNITQLKQEA TDLEQKAFVLT YIRNKFAHNQ<br/> LPKKEFDWYCYEKY GKIEKETYAEYFAEVFKREKEALIK</p> |
| <p>古叶单胞菌<br/> (Porphyromonas guilae)<br/> (SEQ ID No. 119)</p>           | <p>14</p> | <p>MTEQSERPYNGTYTLEDKHFWA AFLNLARHNA YITLTHIDRQLAYSKA<br/> DITNDQD VLSFKALWKNFNDLERKSRLRSLILKHFSFLGAA YGKKLFE<br/> SKSSGNKSSKNKELTKKEKEELQANALS LNLKLSILFDLQKLKDFRNY<br/> SHYRHSGSSELPLFDGNMLQRLYNVFDVSVQRVKIDHEHNDEVDPHYHF<br/> NHLVRKGGKDRYGHNDNPSFKHFFVDGEGMVTEAGLLFFVSLFLEKRD<br/> AIWMQKKIRGFKGGTETYQ QMTNEVFCRSRISLPKLESLRMDDWMLL<br/> DMLNELVRCPKPLYDRLREDDRACFRVPVDILPDEDDTDGGGGEDPFKNT<br/> LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGT YHFAIYKKMIGE QPEDR<br/> HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDKPYISQTS PHYH<br/> IEKGKIGLRFMPEGQHLWPSPEVGTTRTGRSKYAQDKRLTAEAFLSVHEL<br/> MPMMFY YFLLREKYSEEVS AERVQGRKRVIEDVYAVYDA FARDEINTR<br/> DELDA CLADK GIRRGHLP RQMIAILSQEHKDMEEKIRKKLQEMMADTDH<br/> RLDMLDRQTRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDASG</p>  |

[0563]

|  |           |  |
|--|-----------|--|
|  |           | <p>KPLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPF<br/> HETRWESHNTILSFYRSYLARKAFLEIRIGRSDRVENRPFLLLKEPKTDRQ<br/> TLVAGWKGEFHLPRGIFTEAVRDCLIEMGHDEVASYKEVGFMAKAVPLY<br/> FERACEDRVQPFYDSPFNVGNLSKPKKGRFLSKEERAEEWERGKERFRD<br/> LEAWSYSAARRIEDAFAGIEYASPGNKKKIEQLLRDLSLWEAFESKLVKVR<br/> ADRINLAKLKEILEAQEHPYHDFKSWQKFERELRLVKNQDIITWMMCR<br/> DLMEENKVEGLDTGTLYLKDIPNVQEQGSLNVLNRVKPMRLPVVVYR<br/> ADSRGHVHKEEAPLATVYIEERDTKLLKQGNFKSVKDRRLNGLFSFVDT<br/> GGLAMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLTRYPHLPDE<br/> SFREMLESWSDPLLAKWPELHGKVRLLIAVRNAFSHNQYPMYDEAVFSSI<br/> RKYDPSSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA</p>  |
| <p>普雷沃菌<br/> P5-125 (SEQ<br/> ID No. 120)</p>  | <p>15</p> | <p>MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQNENNEN<br/> LWFHPVMSHLYNAKNGYDKQPEKTMFIERLQSYFPLKIMAENQREYS<br/> NGKYKQNRVEVNSNDIFEVLKRAFVGLKMYRDLTNHYKTYEELNDGC<br/> EFLTSTEQPLSGMINNYTVALRNMNERYGYKTEDLAFIQDKRFKFKVD<br/> AYGKKKSQVNTGFFLSLQDYNGDTQKKLHLSGVGIALLICFLDKQYINIF<br/> LSRLPIFSSYNAQSEERRIIRSFINSIKLPKDRIHSEKSNKSVAMDMLNEV<br/> KRCPDELFTTSAEKQSRFRIISDDHNEVLMKRSSDRFVPLLLQYIDYGKL<br/> FDHIRFHVNMGKLRYLKADKTCIDGQTRVRVIEQPLNGFGRLEEAETM<br/> RKQENGTFGNSGIRIRDFENMKRDDANPANYPIVDYTYHYILENNKVE<br/> MFINDKEDSAPLLPIEDDRYVVKTIPTSCRMSTLEIPAMAFHMFVFGSKKT<br/> EKLIVDVHNRYKRLFQAMQKEEVTAENIASFGIAESDLPQKILDLSGNAH<br/> GKDVDAFIRLTVDDMLTDERRIKRKFDDRKSIRSADNMKGKRGFKQIST<br/> GKLADFLAKDIVLFQPSVNDGENKITGLNYRIMQSAIAVYDSGDDYEAK<br/> QQFKLMFEKARLIGKGTTEPHPLYKVFARSIPANAVEFYERYLIERKFYL<br/> TGLSNEIKKGNRVDVPFIRRDQNKWKTTPAMKTLGRIYSEDLPVELPRQMF<br/> DNEIKSHLKSQPMEGIDFNANVTYLIAEYMKRVLDDDFQTFYQWNRN<br/> YRYMDMLKGEYDRKGSQHCFTSVEEREGLWKERASRTERYRKQASNK<br/> IRSNRQMRNASSEEIETILDKRLSNSRNEYQKSEKVIIRYRVQDALLFLA<br/> KKTTLTELADFDGERFKLKEIMPDAEKGILSEIMPMSFTFEKGGKYYTITSE<br/> GMKLLKNYGDFVFLASDKRIGNLLELVGSDIVSKEDIMEEFNKYDQCRPEI<br/> SSIVFNLEKWAFTDYPELSARVDREEKVDKFSILKILLNNKNINKEQSDILR<br/> KIRNAFDHNNYPDKGVVEIKALPEIAMSIIKAFGEYAIMK</p>  |
| <p>嗜枝黄杆菌<br/> (Flavobacteri<br/> um<br/> branchiophil<br/> um) (SEQ ID<br/> No. 121)</p> | <p>16</p> | <p>MENLNKILDKENEICISKIFNTKGIAPITEKALDNIKSKQKNDLNKEARLH<br/> YFSIGHSFKQIDTKKVFDYVLIIEELKDEKPLKFITLQKDFTFEFSIKLQKLI<br/> NSIRNINNHYVHNFNDINLNKIDSNVFHLKESFELAIIEKYYKVNKKYPL<br/> DNEIVLFLKELFIKIDENTALLNYFTNLSKDEAIEYILTFTITENKIWNINNEH<br/> NILNIEKGKYLTFEAMFLITIFLYKNEANHLLPKLYDFKNNKSKQELFTFF<br/> SKKFTSQDIDAEEGHLIKFRDMIQYLNHYPTAWNNDLKLESENKNKIMTT<br/> KLIDSIIIEFELNSNYPFATDIQFKKEAKAFLFASNKKRNQTSFSNKSYNIEE<br/> RHNPHIKQYRDEIASALTPISFNVKEDKFKIFVKKHVLEEYFPNSIGYEKFL<br/> EYNDFTEKEKEDFGLKLYSNPKTNKLIERIDNHKLKSHGRNQDRFMDFS<br/> MRFLAENNYFGKDAFFKCYKFDYDTEQDEFLQSNENDDVVFHKGKVT<br/> TYIKYEEHLKNYSYWDPCFVEENNSMSVKISIGSEEKIKIQRNLMYIFLE<br/> NALYNENVENQGYKLVNYYRELKDVESIASLDLIKSNPDFKSKYKKI<br/> LPKRLHNYAPAKQDKAPENAFETLLKADFREEQYKLLKKAHEKKNK<br/> EDFVKRNKGKQFKLHFIRKACQMMYFKEKYNTLKEGNAAFEKDPVIEK<br/> RKNKEHEFGHHKNLNITREEFNDCYCKWMFAFNGNDSYKYLRLDFSEK<br/> HFFDNQYKLNLFESSVLEAFYAKTKELFKKWIETNKPTNNENRYTLENY<br/> KNLILQKQVFVYHFSKYLIDKNLLSENVIQYKLENVEYLISDFYFQ<br/> SKLSIDQYKTCGKLFNKLKSNKLEDCLLYEIAINYIDKKNVHKIDIQKILT<br/> SKIILTINDANTPYKISVPFNKLERYTEMIAIKNQNNLKFILDLPLYSKN<br/> KIKKGDSDAGYEIIIKNLEIEDINTINNKIINDSVKFEVLMLEKEYFILKD<br/> KCILSKNYIDNSEIPSLKQFSKVIKENENEIINYRNAACHFHLPLLETFDNL<br/> LLNVEQKFIKEELQNVSTINDLSKQPQYLLIFIKFKHNNFYLNLFNKNESK<br/> TIKNDKEVKKNRVLQKFINQVILKKK</p> |

[0564]

|  |           |   |
|--|-----------|---|
| <p>拟香味类香味菌 (SEQ ID No. 122)</p>                          | <p>17</p> | <p>MKDILTTDTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEEVNRKNT<br/>                 FGKLAKRDNGLNKNYIIHVFKDELSISDFEKRVAIFASYFPILETVDKKSIK<br/>                 ERNRTIDLTLRQRQFREMLISLVTAVDQLRNFYTHYHSDIVIENKVLV<br/>                 FLNSSFVSTALHVKDKYKTKDKTKEFLKETIAAELDILIEAYKKKQIEKKN<br/>                 TRFKANKREDILNAIYNEAFWSFINDKDKDKDKETVVAKGADAYFEKNH<br/>                 HKSNDPDFALNISEKGIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGN<br/>                 SIKYMATQRIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVPNVVY<br/>                 QHLSTTQQNSFIEDWNEYKYDYEDDVETDDLSRVTHPVIRKRYEDRFNY<br/>                 FAIRFLDEFFDFPTLRFQVHLGDYVHRRRTKQLGKVESDRIIKEKVTVFAR<br/>                 LKDINSKASYFHSLEEQDKEELDNKWTLFPNPSYDFPKEHTLQHQGEQK<br/>                 NAGKIGIYVKLRDQYKEKAALAEARKSLNPKERSATKASKYDIITQIEA<br/>                 NDNVSEKPLVFTGQPIAYLSMNDIHSMLFSLLDNAELKKTPEEVEAKLI<br/>                 DQIGKQINEILSKDQTDTKLKKYKDNLDKETDQTDKITRDRLARDKKEIEKLL<br/>                 EQKQRADDYNYTSSTKFNIDKSRKRKHLNFNAEKGGKIGVWLANDIKRFM<br/>                 FKESKSKWKGYQHIELQKLFAYFDTSKSDLELILSNMVMVKDYPIELIDL<br/>                 KKSRTLVDLFLNKYLEARLEYIENVITRVKNSIGTPQFKTVRKECFTFLKKS<br/>                 NYTVVSLDKQVERILSMPLFIERGFMDKPTMLEGKSYKQHKKEFADWF<br/>                 VHYKENSNYQNFYDTEVYEITTEDKREKAKVTKKIKQQQKNDVFTLMM<br/>                 VNYMLEEVKLSSNDRLSLNELYQTKEERIVNKQVAKDTQERNKNYIWN<br/>                 KVVDLQLCDGLVHIDNVKLDIGNFRKYENDSRVKEFLTYQSDIVWSAY<br/>                 LSNEVDSNKLYVIERQLDNYESIRSKELLKEVQEIECSVYNQVANKESLK<br/>                 QSGNENFKQYVLQGLLPIGMDVREMLILSTDVFKKKEEIIQLGQAGEVEQ<br/>                 DLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISDNEYAAYMEIFRSIKE<br/>                 KYAN</p>  |
| <p>柱状黄杆菌 (Flavobacterium columnare) (SEQ ID No. 123)</p> | <p>18</p> | <p>MSSKNESYNKQKTFNHYKQEDKYFFGGFLNADDNLRQVGKEFKTRINF<br/>                 NHNNNELASVFKDYFNKEKSVAKREHALNLLSNYFPVLERIQKHTNHNH<br/>                 EQTREIFELLDTIKKLRDYTHHYHKPITINPKIYDFLDDTLDDVLITIKKK<br/>                 KVKNDSRELLKEKLRPELTQLKNQKREELIKKGGKLLLENLENAVFNHC<br/>                 LIPLEENKTDDKQNKTVSLRKYRKSKEPNEETSITLQSGLVFLMSFFLHR<br/>                 KEFQVFTSGLERFKAKVNTIKKEEISLNKNIVMITHWSYSYNYNFKGLK<br/>                 HRIKTDQGVSTLEQNNTTHSLTNTNTKEALLTQIVDYLSKVPNEIYETLSE<br/>                 KQKKEFEEDINEYMRENPENEDSTFSSIVSHKVIKRYENKFNYPAMRFL<br/>                 DEYAEPLTRFMVNFVDYIKDRQKILESIQFDSERIIKHEIHLFEKLSLVE<br/>                 YKKNVYLKETSNDLSRFPLFPNPSYVMANNPIFYIDSRSNLDLVEYNQK<br/>                 KKAQSQNKRRNLTFEKYNKEQSKDAIAMLQKEIGVKDLQQRSTIGLLSC<br/>                 NELPSMLYEVIVKDIKGALENKIAQKIREQYQSIRDFTLDSQPKNIPTTL<br/>                 IKTINTDSSVTFENQPIDIPRLKNALQKELTLTQEKLLNVKEHEIEVDNYNR<br/>                 NKNTYKFKNQPKNVDDKQLQRKYVYRNEIRQEANWLASDLIHFMKN<br/>                 KSLWKGYMHNELQSFLAFFEDKKNDCIALLETVFNLKEDCILTGLKNLF<br/>                 LKHGNFIDFYKEYLKLKEDFLSTESTFLENGFIGLPPKILKELSKRLKYIFI<br/>                 VFQKRQFIIKELEEKNNLYADAINLSRGIFDEKPTMIPFKKPNDEFASF<br/>                 VASYQYNNYQSFYELTPDIVERDKKKYKNLRAINVKIQDYLYKLMVD<br/>                 TLYQDLFNQPLDKSLSDFYVSKAEREKIKADAKAYQKLNDSLWVNVKVIH<br/>                 LSLQNNRITANPKLDIGKYKRALQDEKIATLLTYDARTWTYALQKPEKE<br/>                 NENDYKELHYTALNMELQEYEVRSKELLKQVQELEKKILDKFDYDFSN<br/>                 ASHPEDLEIEDKKGKRHPNFKLYITKALLKNESEIINLENIDIEILLKYDY<br/>                 NTEELKEKIKNMEDEDEKAKIINTKENYNKITNVLIKKALVLIIRNKMAHN<br/>                 QYPPKFIYDLANRFVPKKEEYFATYFNRVFETITKELWENKEKDKTQV</p> |
| <p>牙龈卟啉单胞菌 (SEQ ID No. 124)</p>                          | <p>19</p> | <p>MTEQNEKPYNGTYTLEDKHFWA AFLNLARHNAYITLAHIDRQLAYS<br/>                 ADITNDEDILFFKGQWKNLDNLERKARLRLSLKHFSLFEGAAYGKKLF<br/>                 ESQSSGNKSSKKKELSKKEEELQANALSLDNLKSILFDLQKLDKDFRNY<br/>                 YSHYRHPESSELPLFDGNMLQRLYNVFDVSVQRVCRDHEHNDKVDPHR<br/>                 HFNHLVRKGGKDKYGNNDNPFKHHFVDREGTVTEAGLLFFVSLFLEKR<br/>                 DAIWMQKKIRGFKGGTEAYQOMTNEVFCRSRISLPLKLESLRDTDDWML<br/>                 LDMLNELVRCPSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNT<br/>                 LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPEDRH</p>  |

[0565]

|  |           |   |
|--|-----------|---|
|  |           | <p>LTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDKPYITQTPHYHI<br/> EKGKIGLRFVPEGQHLWPSPEVGATRTGRSKYAQDKRLTAEAFLSVHEL<br/> MPMMFYFLLREKYSEEVSAAEKVQGRIKRVIEDVYAVYDAFARDEINTR<br/> DELDACLADKGIIRRGHLPRQMIAILSQEHKDMEEKVRKKLQEMIADTDH<br/> RLDMLDRQTDRKIRIGRKNAGLPKSGVVADWLVRDMMRFQPVAKDTSG<br/> KPLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHFPL<br/> HETRWESHNILSFYRSYLEARKAFLQSIGRSDRVENHRFLLLKEPKTDRQ<br/> TLVAGWKGEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEVGFMAKAVPLY<br/> FERASKDRVQPFYDYPFNVGNSLKPCKGRFLSKEKRAEEWESGKERFRL<br/> AKLKEILEAKEHPYHDFKSWQKFERELRLVKNQDIITWMMCRDLMEEEN<br/> KVEGLDTGTLYLKDIRTDVQEQQSLNVLNRVKPMRLPVVVYRADSRGH<br/> VHKEQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAM<br/> EQYPISKLRVEYELAKYQTARVCAFEQTELEESLLTRYPHLPDKNFRKM<br/> LESWSDPLLDKWPDLHGNVRLLIARNAFVSHNQYPMYDETLFSSIRKYD<br/> PSSPDAIEERMGLNIAHRLSEEVKQAKEMVERIIQA</p>   |
| <p>卟啉菌<br/> (Porphyromo<br/> nas<br/> sp.)COT-052<br/> OH4946<br/> (SEQ ID No.<br/> 125)</p> | <p>20</p> | <p>MTEQSERPYNGTYTLEDKHFWA AFLNLARHNAYITLTHIDRQLAYSKA<br/> DITNDQDVL SFKALWKNFNDLERKSRLRSLILKHFSFLEGAAYGKKLFE<br/> SKSSGNKSSKNKELTKKEKEELQANALSLDNLSILFDLQKLDKFRNYY<br/> SHYRHSESELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNDKVDPHRH<br/> FNHLVRKGGKDRYGHNDNPSFKHHFVDSEGMVTEAGLLFFVSLFLEKRD<br/> AIWMQKKIRGFKGGTETYQMTNEVFCRSRISLPKLLKLESLRTDDWMLL<br/> DMLNELVRCPKPLYDRLREDDRACFRVPVDILPDEDDTDGGGEDPDKNT<br/> LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKMKMIGEQPEDR<br/> HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDKPYISQTPHYH<br/> IEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSKYAQDKRLTAEAFLSVHEL<br/> MPMMFYFLLREKYSEEVSAAEKVQGRIKRVIEDVYAIYDAFARDEINTLK<br/> ELDACLADKGIIRRGHLPKQMIGILSQRKDMEEKVRKKLQEMIADTDHR<br/> LDMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGK<br/> PLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHFPLH<br/> ETRWESHNILSFYRSYLARKAFLERIGRSDRVENC PLLLLKEPKTDRQT<br/> LVAGWKGEFHLPRGIFTEAVRDCLIEMGYDEVGSYREVGFMAKAVPLYF<br/> ERACEDRVQPFYDSPFNVGNSLKPCKGRFLSKEDRAEEWERGKERFRDL<br/> EAWSHSAARRIKDAFAGIEYASPGNKKKIEQLLRDLSLWEAFSLVRA<br/> DKINLAKLKEILEAQEHPYHDFKSWQKFERELRLVKNQDIITWMMCRD<br/> LMEENKVEGLDTGTLYLKDIRPNVQEQQSLNVLNRVKPMRLPVVVYRA<br/> DSRGHVHKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTG<br/> GLAMEQYPISKLRVEYELAKYQTARVCFELTLRLEESLLSRYPHLPDES<br/> REMLSWSDPLLAKWPELHGKVRLLIARNAFVSHNQYPMYDEAVFSSIR<br/> KYDPSSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA</p> |
| <p>中间普雷沃<br/> 菌 (SEQ ID<br/> No. 126)</p>  | <p>21</p> | <p>MEDDKKTKESTNMLDNKHFWA AFLNLARHNAYITVNHINKVLELKNKK<br/> DQDIINDNDQDILA IAKTHWEKVNGLNKTERLRELMTKHFPLETAIYTKN<br/> KEDKEEVKQEKQAKAQSFDLSLKHCLFLFLEKLQEARNYSHYKYSESTK<br/> EPMLEKELLKMYNIFDDNIQLVIKDYQHNKDINPDEDFKHLDRTEEEFN<br/> YYFTTNKKGNITASGLLFFVSLFLEKKDAIWMQKLRGFKDNRESKMKM<br/> THEVFCRSRMLLPKLRLESTQTDWILLDMLNELIRCPKSLYERLQGEYR<br/> KKFNVPFDSADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLR<br/> FQIDLGTYHFSIYKKLIGGQKEDRHLTHKLYGFERIQEFAKQNRTEDEWKAI<br/> VKDFDITYETSEEPYISETAPHYHLENQKIGIRFRNDNDEIWP SLKTNGENN<br/> EKRYKLDKQYQAEAFLSVHELLPMMFYLLLLKKEPNNDKKNASIVEG<br/> FIKREIRDIYKLYDAFANGEINNIDDLEKYCEDKGIPKRHLPKQMVAILYD<br/> EHKDMAEEAKRKQKEMVKDTKLLATLEKQTQGEIEDGGRNIRLLKSGE<br/> IARWLVNDMMRFQPVQKDNENPLNNSKANSTEYQMLQRSLALYNKEE<br/> KPTRYFRQVNLINSSNPHFLKWKWEECNILSFYRSYLTKKIEFLNKLK<br/> PEDWEKNQYFLKLEPKTNRETLVQGWKNGFNLPRGIFTEPIREWFKRH<br/> QNDSEYEKVE TLDVRGLVTKVIPLFFKKEDSKDKEEYLLKDKAQKEINNC<br/> VQPFYGFYNVGNIIHKPDEKDFLPSEERKKLWGDKKYKFKGYKAKVKS</p>   |

[0566]

|   |              |   |
|---|--------------|---|
|   |              | KKLTDKEKEEYRSYLEFQSWNKFERELRLVRNQDIVTWLLCTELIDKLV<br>EGLNVEELKKLRLKIDIDDTAKQEKNILNRVMPMQLPVTVYEIDDSHNI<br>VKDRPLHTVYIEETKTLLKQGNFKALVKDRRLNGLFSFVDTSSSETELKS<br>NPISKSLVEYELGEYQNARIETIKDMLLEETLIEKYKTLPTDNFSDMLNG<br>WLEGKDEADKARFQNDVKLLVAVRNAFASHNQYPMRNRIAFANINPFSLS<br>SADTSEEKLDIANQLKDKTHKIIKRIIEIEKPIETKE   |
| PIN17_0200<br>[中间普雷沃<br>菌 17] (SEQ<br>ID No. 127) | AFJ0752<br>3 | MKMEDDKKTKESTNMLDNKHFWA AFLNLARHN VYITVNHINKVLELKN<br>KKDQDIIDNDQDILA I KTHWEK VNGDLNKTERLRELMTKHFPFLETAIYT<br>KNKEDKEEVKQEKQAKAQSFDSLKHCLFLFLEKLQEARNYYSHYKYSES<br>TKPEMLEKELLKMYNIFDDNIQLVIKDYQHNDINPDEDFKHLDRTEEE<br>FNYYFTTNKKNITASGLLFFVSLFLEKKDAIWMQQLRGFKDNRESKKEK<br>MTHEVFCRSRMLLPKLRLESTQTQDWILLDMLNELIRCPKSLYERLQGEY<br>RKKFNVPFDSADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNL<br>RFQIDLGTYHFSIYKLLIGGQKEDRHLTHKLYGFERIQEFAKQNRDTEWK<br>AIVKDFDTYETSEEPYISETAPHYHLENQKIGIRFRNDNDEIWPSLKTNGE<br>NNEKRKYKLDKQYQAEAFLSVHELLPMMFYLLKKEEPNNDKKNASI<br>VEGFIKREIRDIYKLYDAFANGEINNIDDLEKYCEDKGIPKRHLPKQMVAI<br>LYDEHKDMAEEAKRKQKEMVKDTKKLLATLEKQTQGEIEDGGRNIRLL<br>KSGEJARWLVNDMMRFQPVQKDNENPLNNSKANSTEQMLQRSLALY<br>NKEEKPTRYFRQVNLINSSNPHFLKWKWEECNILSFYRSYLTKKIEFL<br>NKLKPEDWEKNQYFLKLEKPTNRET LVQGWKNGFNLPRGIFTEPIREW<br>FKRHQNDSEEYK VETLDRVGLVTKVIPLFFKEDSKDKEEYKFKDAQK<br>EINNCVQPFYGFYVNGNIHKPDEKDFLPSEERKKLWGDKKYFKGYKA<br>KVKS KLT DKEKEEYRSYLEFQSWNKFERELRLVRNQDIVTWLLCTELID<br>KLKVEGLNVEELKKLRLKIDIDDTAKQEKNILNRVMPMQLPVTVYEID<br>DSHNIVKDRPLHTVYIEETKTLLKQGNFKALVKDRRLNGLFSFVDTSSSE<br>TELKSNPISKSLVEYELGEYQNARIETIKDMLLEETLIEKYKTLPTDNFSD<br>MLNGWLEGKDEADKARFQNDVKLLVAVRNAFASHNQYPMRNRIAFANIN<br>PFSLSADTSEEKLDIANQLKDKTHKIIKRIIEIEKPIETKE |
| 中间普雷沃<br>菌 (SEQ ID<br>No. 128)                    | BAU186<br>23 | MEDDKKTTDSISYELKDKHFWA AFLNLARHN VYITVNHINKVLELKNKK<br>DQDIIDNDQDILA I KTHWEK VNGDLNKTERLRELMTKHFPFLETAIYSKN<br>KEDKEEVKQEKQAKAQSFDSLKHCLFLFLEKLQETRNYYSHYKYSESTK<br>EPMLEKELLKMYNIFDDNIQLVIKDYQHNDINPDEDFKHLDRTEEDFN<br>YYFTRNKKGNITESGLLFFVSLFLEKKDAIWMQQLRGFKDNRESKKEK<br>THEVFCRSRMLLPKLRLESTQTQDWILLDMLNELIRCPKSLYERLQGEDR<br>EKFKVPFDPADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLR<br>FQIDLGTFFHSIYKLLIGGQKEDRHLTHKLYGFERIQEFAKQNRDTEWKAI<br>VKDLDTYETSNERYISETTPHYHLENQKIGIRFRNDNDEIWPSLKTNGEN<br>NEKSKYKLDKQYQAEAFLSVHELLPMMFYLLKKEEPNNDKKNASIVE<br>GFIKREIRDMYKLYDAFANGEINNIDDLEKYCEDKGIPKRHLPKQMVAIL<br>YDEHKDMVKEAKRKQKEMVKDTEKLLAALEKQTQEKTEDGGRNIRLLK<br>SGEJARWLVNDMMRFQPVQKDNENPLNNSKANSTEQMLQRSLALYN<br>KEEKPTRYFRQVNLINSSNPHFLKWKWEECNILSFYRSYLTKKIEFLN<br>KLKPEDWEKNQYFLKLEKPTNRET LVQGWKNGFNLPRGIFTEPIREWFK<br>RHQND SKEYE KVEALDRVGLVTKVIPLFFKEDSKDKEEYKFKDAQK<br>NNCVQPFYGFYVNGNIHKPDEKDFLHREERIELWDKKDKFKGYKAKV<br>KSKKLT DKEKEEYRSYLEFQSWNKFERELRLVRNQDIVTWLLCTELIDKL<br>KVEGLNVEELKKLRLKIDIDDTAKQEKNILNRVMPMQLPVTVYEIDDS<br>HNIVKDRPLHTVYIEETKTLLKQGNFKALVKDRRLNGLFSFVDTSSSEAE<br>LKSNPISKSLVEYELGEYQNARIETIKDMLLEETLIEKYKNLPTDNFSDM<br>LNGWLEGKDEADKARFQNDVKLLVAVRNAFASHNQYPMRNRIAFANINP<br>FSLSSADTSEEKLDIANQLKDKTHKIIKRIIEIEKPIETKE |
| HMPREF64<br>85_0083 [颊<br>普雷沃菌<br>ATCC            | EFU319<br>81 | MQKQDKLFVDRKKNAIFAFPKYITIMENKEKPEPIYYELTDKHFWA AFLN<br>LARHN VYTTINHINRRLEIAELKDDGYMMGIKGSWNEQAKKLDKVRRL<br>DLIMKHFPFLEAAAYEMTNSKSPNNKEQREKEQSEALSNLKNVLFIFL<br>EKLQVLRNYYSHYKYSEESPKPIFETSLKNMYKVF DANVRLVKRDYMH   |

[0567]

|  |           |   |
|--|-----------|---|
| 33574] (SEQ ID No. 129)  |           | HENIDMQRDFTHLNRKKQVGRTKNIIDSPNFHYHFADKEGNMTIAGLLFF VSLFLDKKDAIWMQKCLKGFKDGRNLREQMTNEVFCSRISLPLKLEEN VQTKDWMQLDMLNELVRCPSLYERLREKDRESFKVPFDIFSDDYNAEE EPFKNTLVRHQDRFPYFVLRVYFDLNEIFEQLRFQIDLGTYHFSIYNKRIGDE DEVRHLTHHLYGFARIQDFAPQNQPPEWRKLVKDLDFHETSQEPYISKTA PHYHLENEKIGIKFCSAHNNLFPQLTDKTCNGRSKFNLTGTQTAEAFLSV HELLPMMFYLLLT KDYSRKESADKVEGIIRKEISNIYAIYDAFANNEINSI ADLTRRLQNTNILQGHLPKQMISILKGRQKDMGKEAERKIGEMIDDTQRR LDLLCKQTNQKIRIGKRNAGLLKSGKIADWLVDMMRFQPVQKQDQNNIP INNSKANSTEYRMLQRALALFGSENFRKAYFNQMNVLVGNNDNPHFLAE TQWEHQTNILSFYRNYLEARKKYLKGLKQPONWKQYQHFLILK VQKTNR NTLVTGWKNSFNLPGRGIFTQPIREWFEKHNSKRIYDQILSFDRVGFVAK AIPLYFAEEYKDNVQPFYDYPFNIGNRLKPKKRQFLDKKERVELWQKNK ELFKNYPSEKKKTDLAYLDFLSWKKFERELRLIKNQDIVTWLMFKELFNM ATVEGLKIGEIHLRDIDTNTANEESNNILNRIMPMLKPVKTYETDNKGNIL KERPLATFYIEETETKVLKQGNFKALVKDRRLNGLFSFAETTDLNLEEHPI SKLSVDLELIK YQTTTRISIFEMTLGLEKKLIDKYSTLPTDSFRNMLERWLQC KANRPELKNYVNSLIAVRNAFVSHNQYPMYDATLFAEVKKFTLFPVSDTK KIELNIAPQLEIVGKAIKEIEKSENKN   |
| HMPREF91 44_1146 [苍白普雷沃菌 (Prevotella pallens) ATC C 700821] (SEQ ID No. 130) | EGQ184 44 | MKEEEKGKTPVVSTYNKDDKHFWAAFLNLARHNVYITVNHINKILGEGE INRDGYENTLEKSWNEIKDINKDRLSKLIHKHFPFLEVTTYQRNSADTTK QKEEKQAEASLESLLKKSFFVFIYKLRDLRNHYSHYKHSKSLERPKFEED LQEKMYNIFDASIQLVKEDYKHNTDIKTEEDFKHLDRKGQFKYSFADNE GNITESGLLFFVSLFLEKKDAIWVQKLEGFKCSNESYQKMTNEVFCSR MLLPKLRLQSTQTQDWILLDMLNELIRCPKSLYERLREEDRKKFRVPIEIA DEDYDAEQEPFKNALVRHQDRFPYFALRYFDYNEIFTNLRQIDLGTYHF SIYKKQIGDYKESHHLTHKLYGFERIQEFTKQNRPDEWRKFKVTFNSFETS KEPYIPETTPHYHLENQKIGIRFRNDNDKIWPSLKTNSEKNEKSKYKLDKS FQAEAFLSVHELLPMMFYLLLTEN TDNDNEIETKKKENKNDKQEKHK IEIIEENKITEIYALYDAFANGKINSIDKLEEYCKGKDIEIGHLPKQMIAILKS EHKDMATEAKRKQEEMLADVQKSLESLDNQINEEIEIENVERKNSSLKSGEI ASWLVDMMRFQPVQKDNENGNPLNNSKANSTEYQMLQRSLALYNKEE KPTRYFRQVNLIESSNPHPLNTEWEKCNNILSFYRSYLEAKNFLESK PEDWEKNQYFLMLKEPKTNCETLVQGWKNGFNLPGRGIFTEPIRKWFMEH RKNITVAELKRVGLVAKVIPLFFSEEYKDSVQPFYNYLFVGNINKPDEK NFLNCEERRELLRKKKDEFKMTDKEKEENPSYLEFQSWNKFERELRLV RNQDIVTWLLCMELFNKKIKELNVEKIYLNKINTNTTKKEKNTEEKNGE EKIIKEKNILNRIMPMLPIK VYGRENFSKNKKKIRRNTFFT VYIEEKG T KLLKQGNFKALERDRRLGGLFSFVKTHSKAESKNTISKSRVEYELGEYQ KARIEIKDMLALEETLIDKYNSLDTDNFHNMLTGWLKLDKDEPKASFQN DVDLLIAVRNAFVSHNQYPMRNRIFAANINPFLSSANTSEEKGLGIANQLK DKTHKTIEKIIIEKPIETKE |
| HMPREF97 14_02132 [拟香味类香味菌 CCUG 12901] (SEQ ID No. 131)                      | EHO087 61 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEEVNRNT FGKLAKRDNGNLKNYIIHVFKDELSDFEKRVAFASYFPILETVDKKSIK ERNR TIDLTL SQRIRQFREMLISLVTAVDQLRNFYTHYHSEIVIENKVL D FLNSSLVSTALHVKDKYLKTDKTEFLKETIAAELDILIEAYKKKQIEKKN TRFKANKREDILNAIYNEAFWSFINDKDKDKETVVAKGADAYFEKNHHK SNDDPDFALNISEKGIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIK YMATQRIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKV PNVVYQH LSTTQQNSFIEDWNEYKYDYEDDVETDDL SRVIHPVIRKRYEDRFNYFAI RFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVESDRIIEKVTVFARLK DINSAKANYFHSLEEQDKEELDNKWTLFPNPSYDFPKEHTLQHQQEQKN AGKIGIYVKL RDTQYKEAAL EEARKSLNPKERSATKASKYDIITQIIEAN DNVKSEKPLVFTGQPIAYLSMNDIHSMLFSLT DN AELKKTPEEVEAKLID QIGKQINEILSKD TDTKILKKYKDN DLKETD TDKITRDLARDKEIEKLILE QKQRADDYNYTSSTKFNIDKSRKRKHLLFNAEKGKIGVWLANDIKRFMT   |

[0568]

|   |              |  |
|---|--------------|--|
|   |              | EEFKSKWKG YQHTELQKLFAYYDTSKSDLDLILSDMVMVKDYPIELIALV<br>KKSRTLVDLFLNKYLEARLGYMENVITRVKNSIGTPQFKTVRKECFTLK<br>SNYTVVSLDKQVERILSMPLFIERGFMDKPTMLEGKSYQQHKEKFADW<br>FVHYKENSNYQNFYDTEVYEITTEDKREKAKVTKKIKQQKNDVFTLM<br>MVNYMLEEVKLKSSNDRLSLNELYQTKEERIVNKQVAKDTQERNKNI<br>WNVVDLQ LCEGLVRIDKVKLKDIGNFRKYENDSRVKEFLTYQSDIVWS<br>AYLSNEVDSNKLYVIERQLDNYESIRSKELLKEVQEIECSVYNQVANKES<br>LKQSGNENFKQYVLQGLVPIGMDVREMLILSTDVVKFKEEIIQLGQAGEVE<br>QDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISDNEYA EYMEIFRSIK<br>EKYTS   |
| HMPREF97<br>11_00870<br>[拟香味类香<br>味菌 CCUG<br>3837] (SEQ<br>ID No. 132)  | EKB060<br>14 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEEVNRKT<br>FGKLAKRDNGNLKNYIIHVFKDELSISDFEKRAIFASYFPILETVDKKSIK<br>ERNRTIDLTLRSQRIRQFREMLISLVTAVDQLRNFYTHYHHSEIVIENKVL<br>FLNSSLVSTALHVKDYLKTDKTKEFLETIAAELDILIEAYKKKQIEKKN<br>TRFKANKREDILNAIYNEAFWSFINDKDKDKETVVAKGADAYFEKNHHK<br>SNPDFALNISEKGIYVLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIK<br>YMATQRIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKV PNVVYQH<br>LSTTQQNSFIEDWNEYKYDYEDDVETDDLRSVIHPVIRKRYEDRFNYFAI<br>RFLDEFFDFPTLRFQVHLGDYVHDRRTKQLGKVESDRIIKEKVTVFARLK<br>DINSAKASYFHSLEEQDKEELDNKWTLPNPSYDFPKEHTLQHQQEQKN<br>AGKIGIYVKLRDTQYKEKAAL EEARKSLNPKERSATKASKYDIITQIIEAN<br>DNVKSEKPLVFTGQPIAYLSMNDIHSMLFSLTDNAELKKTPEEVEAKLID<br>QIGKQINEILSKD TDTKILKKYKDNLDKETD TDKITRDLARDKEIEKLILE<br>QKQRADDYNYTSSTKFNIDKSRKRKHLFNAEKGGKIGVWLANDIKRFMF<br>KESKSKWKG YQHTELQKLFAYDTSKSDLELILSDMVMVKDYPIELIDL<br>RKSRTLVDLFLNKYLEARLGYIENVITRVKNSIGTPQFKTVRKECF AFLKES<br>NYTVASLDKQIERILSMPLFIERGFMDKPTMLEGKSYQQHKEKFADWV<br>HYKENSNYQNFYDTEVYEITTEDKREKAKVTKKIKQQKNDVFTLMMV<br>NYMLEEVKLKPSNDRSLNELYQTKEERIVNKQVAKDTQERNKNIWNV<br>VVDLQ LCEGLVRIDKVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSGYLS<br>NEVDSNKLYVIERQLDNYESIRSKELLKEVQEIECIVYNQVANKESLKQSG<br>NENFKQYVLQGLPRGTDVREMLILSTDVVKFKEEIMQLGQVREVEQDL<br>YSLIYIRNKFAHNQLPIKEFFDFCENNYRPSIDNEYA EYMEIFRSIKEKY<br>AS |
| HMPREF96<br>99_02005<br>[动物溃疡伯<br>格菌 ATCC<br>43767] (SEQ<br>ID No. 133) | EKB541<br>93 | MENKTSLGNNIYNPFKPDKSYFAGYFNAAMENTDSVFRELGKRLKKG<br>EYTSENFFDAIFKENISLVEYERYVKLLSDYFPMARLLDKKEVPIKERKEN<br>FKKNFKGIIKAVRDLRN FYTHKEHGEVEITDEIFGVLD EMLKSTVLT VKK<br>KKVKTDKTKEILKKSIEKQLDILCQKKLEYLRDTARKIEBKRRNQ RERGE<br>KELVAPFKYSDKRDDLIAAIYNDAFDVYIDKKKDSLKESKAKYNTKSDP<br>QQEEDLKIPIKNGVVFLLSLFLTKQEIHAFKSKIAGFKATVIDEATVSEA<br>TVSHGKNSICFMATHEIFSHLAYKLLKRKVRTAEINYGEAENAEQLSVYA<br>KETLMMQMLDEL SKVPDVVYQNLSEDVQKTFIEDWNEYLKENNGDVGT<br>MEEEQVIHPVIRKRYEDKFNYFAIRFLDEFAQFPTLRFQVHLGNYLHDSRP<br>KENLISDRRIKEKITVFGRLSELEHKALFIKNTETNEDREHYWEIFPNPNY<br>DFPKENISVNDKDFPIAGSILDREKQPVAGKIGIKVLLNQYVSEVDKAV<br>KAHQLKQRKASKPSIQNIIEEIVPINESNPKEAIVFGGQPTAYLSMNDIHSIL<br>YEFFDKWEKKKEKLEKKGEKELRKEIGKELEKIVGKIQAQIQIIDKDTN<br>AKILKPYQDGNST AIDKEKLIKDLKQEQNILQKLDKDEQTVREKEYNDFIA<br>YQDKNREINKVRDRNHKQYLKDNLKRKYPEAPARKEVLYYREKGVVA<br>VWLANDIKRFMPTDFKNEWKGEQHSLQKSLAYYEQCKEELKNLLPEK<br>VFQHLPFKLG YFQQKYL YQFYTCYLDKRLEYISGLVQQAENFKSENKV<br>FKKVENECFKFLKKNYTHKELDARVQSILGYPIFLERGFMDKPTIIGK<br>TFKGNEALFADWFRYYKEYQNFQTFYDTENYPLVELEKKQADRKRKTKI<br>YQQKNDVFTLLMAKHIFKSVFKQDSIDQFSLEDLYQSREERLGNQERAR<br>QTGERNTNYIWNKTVDLKLCDGKITVENVKLKNV GDFIKY EYDQRVQAF<br>LKYEENIEWQAFLIKESKEEENYPYVVEREIEQYEVRRREELLKEVHLIEE  |

[0569]

|   |              |   |
|---|--------------|---|
|   |              | YILEKVKDKKEILKKGDNQNFKYYILNGLLKQLKKNEDVESYKVFNLNTEPE<br>DVNINQLKQEATDLEQKAFVLTYYIRNKFVHNQLPKKEFWDYCQEKYGKI<br>EKEKTYAEYFAEVFKKEKEALIK   |
| HMPREF91<br>51_01387<br>[解糖普雷沃<br>菌 F0055]<br>(SEQ ID No.<br>134) | EKY000<br>89 | MMEKENVQGSIIYYEPTDKCFWAAFYNLARHNAYLTIAHINSFVNSKKG<br>INDDDKVLDIIDDWSKFDNDLLMGARLNKLIKHPFLKAPLYQLAKRKT<br>RKQQGKEQQDYEEKGDEDEPEVIEAIAANAFKMANVRKTLHAFLKQLED<br>LRNHFSHYNYNSPAKKMEVKFDDGFCNKLYYVFDAALQMVKDDNRMN<br>PEINMQTDFEHLVRLGRNRKIPNTFKYNFTNSDGTINNGLLFFVSLFLEK<br>RDAIWMQKKIKGFKGGTENYMRMTNEVFCRNRMVIPKLRLETDYDNHQ<br>LMFDMNLNLRVRCPLSLYKRLKQEDQDKFRVPIEFLDEDEADNPYQENA<br>NSDENPTEETDPLKNTLVRHQHRFPYFVLRVFDLNEVFKQLRFQINLGCY<br>HFSIYDKTIGERTEKRHLTRTLFGFDRLQNFVSKLQPEHWKMNVMKHLDE<br>ESSDKPYLSDAMPHYQIENEKIGIHFLLKTDTEKKEVWPSLEVEVSSNRN<br>KYKSEKNLTADAFLSTHELLPMMFYQLLSSEEKTRAAAGDKVQGVLS<br>YRKKIFDIYDDFANGTINSMQKLDERLAKDNLLRGNMPQQMLAILEHQE<br>PDMEQKAKEKLDRLITETKKRIGKLEDQFKQKVRIGKRRADLPKVGSIAD<br>WLVNDMMRFQPAKRANADNTGVPDSKANSTEYRLLQEALAFYSAYKDR<br>LEPYFRQVNLIGGTNPHFPLHRVDWKKCNHLLSFYHDYLEAKEQYLSHL<br>SPADWQKHQHFLLLKVRKDIQNEKDWKSLVAGWKNGFNLPRGLFTE<br>SIKTWFSTDADKVQITDTKLFENRVGLIAKLIPLYDYDKVYNDKPPFYQY<br>PFNINDRYKPEDTRKRFTAASSKLWNEKKMLYKNAQPDSSDKIEYPQYL<br>DFLSWKKLERELRMLRNQDMMVWLMCKDLFAQCTVEGVFADLKLSSQ<br>LEVVDNVQDNLNLVNNVSSMILPLSVYPSDAQGNVLRNSKPLHTVYVQE<br>NNTKLLKQGNFKSLLKDRRLNGLFSFIAAEGEDLQQHPLTKNRLEYELSIY<br>QTMRISVFEQTLQLEKAILTRNKTLCGNNFNLLNSWSEHRTDKKTLQPD<br>IDFLIAVRNAFVSHNQYPMSTNTVMQGIEKFNIQTPKLEEKDGLGIASQLAK<br>KTKDAASRLQNIINGGTN            |
| A343_1752<br>[牙龈卟啉单<br>胞菌 JCVI<br>SC001]<br>(SEQ ID No.<br>135)   | EOA105<br>35 | MTEQNEKPYNGTYTLEDKHFWAFFNLARHNAYITLTHIDRQLAYSKA<br>DITNDEDILFFKGQWKNLDNDLERKARLRLSLILKHFSFLEGAAYGKKLFES<br>QSSGNKSSKKKELTKKEEELQANALSLDNLKSILFDLQKLDKFRNYYS<br>HYRHPESSELPLFDGNMLQRLYNVFDVSVQRVVRDHEHNDKVDPHRFH<br>NHLVRKGGKDRCGNNDNPFKHHFVDREKVTEAGLLFFVSLFLEKRD<br>IWMQKKIRGFKGGTETYQQMTNEVFCRSRISLPKLESLRTDDWMLLD<br>MLNELVRCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTLV<br>RHQDRFPYFALRYFDLKKVFTSLRFHIDLGYHFAYKKNIGEQPEDRHLLT<br>RNLYGFGRIQDFAEEHRPEEWKRLVRDLDFYFETGDKPYITQTPHYHIEK<br>GKIGLRFVPEGQLLWPSPEVGATRTGRSKYAQDKRFTAFAFLSVHELMP<br>MMFYFLLREKYSEEASAERVQGRIVIEDVYAVYDAFARGEIDTLDR<br>LDAACLADKGIIRGHLPRQMAILSSEHKDMEEKVRKQLQEMIADTDHRL<br>DMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPAKDTSGKP<br>LNNSKANSTEYRMLQRALALFGGEEKERLTPYFRQMNLTGGNNPHPLHE<br>TRWESHTNLSFYRSYLKARKAFLQSIGRSDRVENHRFLLKPKTDRQTL<br>VAGWKGEFHLPRGIFTEAVRDCLIEMLGLDEVGSYKEVGFMAKAVPLYFE<br>RACKDRVQPFYDYPFNVGNSLKPCKGRFLSKEKRAEEWESGKERFRDLE<br>AWSHSAARRIEDAFAGIENASRENKKKIEQLLQDLSLWETFESKLVKAD<br>KINIAKLLKEILEAKEHPYLDKFSWQKFERELRLVKNQDIITWMMCRDLM<br>EENKVEGLDTGTLYLKDIRTDVHEQGSNLVLRVKNRMLPVPVYRADS<br>RGHVHKEQAPLATVYIEERDTKLLKQGNFKSFVDRRLNGLFSFVDTGA<br>LAMEQYPISKLRVEYELAKYQTARVCAFEQTLELESLLTRYPHLPDKNF<br>RKMLSWSDPLLDKWPDLHGNVRLIAVRNAFVSHNQYPMYDETFLSSIR<br>KYDPSSPDAIEERMGLNIAHRLSEEVKQAKEMVERIIQA |
| HMPREF19<br>81_03090<br>[酿脓拟杆菌<br>F0041] (SEQ<br>ID No. 136)      | ERI8170<br>0 | MESIKNSQKSTGKTLQKDPYFGLYLNMALLNVRKVENHIRKWLGVAL<br>LPEKSGFHSLTDTNLSAKWTRFYKSRKFLPFLEMFDSDKKSyenRRE<br>TTECLDTIDRQKISSLLKEVYGLQDIRNAFVSHYHIDDQSVKHTALISSEM<br>HRFIENAYSFALQKTRARFTGVFVETDFLQAEKGDNKKFFAIGGNEGIK<br>LKDNALIFLICLFLDREEAFKFLSRATGFKSTKEKGFLAVRETFCALCCRQP  |

[0570]

|  |              |  |
|--|--------------|--|
|  |              | HERLLSVNPREALLMDMLNELNRCPDILFEMLDEKDKQSFLPLLGEEEQA<br>HILENSLNDELCEAIDDPFEMIASLSKRVRVKYKRFYLMRLYIEEKNLLPFI<br>RFRIDLGCLELASYPKKMGEENNYERSVTDHAMAFAFGRLTDFHNEDAVLQ<br>QITKGITDEVRFSLYAPRYAIYNNKIGFVRTGGSDKISFPTLKKKGGEGHC<br>VAYTLQNTKSFGEFISYDLRKILLSSFLDKDKAKNIVSGLLEQCEKHWKDL<br>SENLFAIRTELQKEFPVPLIRYTLPRSKGGKLVSSKLADKQEKYESEFER<br>RKEKLTEILSEKDFDLSQIPRRMIDEWLNVLPTSREKLLKGYVETLKLDCR<br>ERLRVFEKREKGEHPVPPRIGEMATDLAKDIIRMVIDQGKQKRITSAYYS<br>EIQRCLAQYAGDDNRRHLDLSIIRELRLKDTKNGHPFLGKVLRLPGLGHEK<br>LYQRYFEEKKEWLEATFYPAASPKRVPRFVNPTGKQKELPLIIRNLKME<br>RPEWRDVKQRKNSHPIDLPSQLFENEICRLLKDKIGKEPSGKLLKWNEMFK<br>LYWDKEFPNGMQRFYRCKRRVEVFDKVVEYVEEYSEEGGNYKKYVEALID<br>EVVRQKISSSKEKSKLQVEDLTLVRRVFKRAINEKEYQLRLLCEDRLLF<br>MAVRDLYDWKEAQLDLKIDNMLGEPVSVSQVIQLEGGQPD AVIKAEC<br>KLKDVSKLMRYCYDGRVKGMLMPYFANHEATQEQVEMELRHIEDHRRR<br>VFNWVFALEKSVLKNEKLRRFYEESSQGGCEHRRCIDALRKASLVSEEEYE<br>FLVHIRNKSAHNQFPDLEIGKLPNVTSGFCECIWSKYKAIICRIIPFIDPER<br>RFFGKLLKQK  |
| HMPREF15<br>53_02065<br>[牙龈卟啉单<br>胞菌 F0568]<br>(SEQ ID No.<br>137) | ERJ656<br>37 | MNTVPASENKGSRTVEDDPQYFGLYLNLARENLIEVESHVRIKFGKKKL<br>NEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSK<br>TGVDPDSAQRLIRELYSLDFLRNDFSHNRDLGTTFEHLEVSPDISSFITGT<br>YSLACGRAQSRFADFFKPDDFVLAKNRKEQLISVADGKECLTVSGLAFFI<br>CLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDL CIRHPHDLRESSN<br>TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLSNSLN<br>EESRLLWDGSSDWAEALTKRIRHQDRFPYLMRLRFIEEMDLLKIRFRVDL<br>GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br>YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPRSMGFISV<br>HDLRKLMLMELLCEGFSRMSQDFLRKANRILDETAEGKLFQSFALPEMR<br>HRFIPPQNPKSKDRREKAETTLEKYKQEIKGRKDKLNSQLLSAFDMDQRQ<br>LPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLKFRKDGDKARA<br>IPLVGEMATFLSQDIVRMIIESETKKLITSAYYNEMQRSLAQYAGEENRHQ<br>FRAIV AELRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAKT<br>YRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRRMEQNDLQDWIRNKQA<br>HPIDLPSHLFDSKIMELLKVKDGKKWNEAFKDWWSKYPDGMQPFYQ<br>LRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTQVQDKKRELRTAGKPVPP<br>DLAADIKRSFHRAVNEREFMLRLVQEDDRMLMAINMMTDREEDILPG<br>LKNIDSILDEENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWS<br>KYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALE<br>GAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILI<br>RNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWK<br>KYEMIIRKILPILDPENRFFGKLLNMSQPINDL |
| HMPREF19<br>88_01768<br>[牙龈卟啉单<br>胞菌 F0185]<br>(SEQ ID No.<br>138) | ERJ819<br>87 | MNTVPASENKGSRTVEDDPQYFGLYLNLARENLIEVESHVRIKFGKKKL<br>NEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSK<br>TGVDPDSAQRLIRELYSLDFLRNDFSHNRDLGTTFEHLEVSPDISSFITGT<br>YSLACGRAQSRFADFFKPDDFVLAKNRKEQLISVADGKECLTVSGLAFFI<br>CLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDL CIRHPHDLRESSN<br>TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLSNSLN<br>EESRLLWDGSSDWAEALTKRIRHQDRFPYLMRLRFIEEMDLLKIRFRVDL<br>GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br>YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV<br>HDLRKLMLMELLCEGFSRMSQSGFLRKANRILDETAEGKLFQSFALPEMR<br>HRFIPPQNPKSKDRREKAETTLEKYKQEIKGRKDKLNSQLLSAFDMNQRQ<br>LPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLKFRKDGDKARA<br>IPLVGEMATFLSQDIVRMIIESETKKLITSAYYNEMQRSLAQYAGEENRRQ<br>FRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAKT<br>YRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRRMEQNDLQDWIRNKQA   |

[0571]

|  |                |  |
|--|----------------|--|
|  |                | HPIDLPSHLFDSKIMELLKVKDGGKKKWNEAFKDWSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPGLKNIDSILDEENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILIRNKAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYEMIIRKILPILDHENRFFGKLLNNMSQPINDL   |
| HMPREF19<br>90_01800<br>[牙龈卟啉单胞菌<br>W4087]<br>(SEQ ID No. 139) | ERJ873<br>35   | MNTVPASENKGQSRTVEDDPQYFGLYLNLARENLIEVESHVRIKFGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSKTGVDPDSAQRLIRELYSLDFLRNDFSHNRLDGTTFEHLEVSPDISSFITGTYSLACGRAQSRFADFFKPDDFVLAKNRKEQLISVADCEKCLTVSGLAFFICLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDLRHPHDLRESSNTKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLSESLNEESRLLWDGSSDWAALTKRIRHQDRFPYMLRFFIEEMDLLKIRFRVDLGEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEASYPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPRSMGFISVHDLRKLMLMELLCESFSRMSQDFLRKANRILDETAEGKLFQFSALFPEMRHRFIPPQNPKSKDRREKAETTLEKYKQEIKGKDKLNSQLLSAFDMDQRRQLPSRLLDEWMNIRPASHSVKLRITYVKQLNEDCRLRLQKFRKDGDKGKARAIPLVGEMATFLSQDIVRMIIESETKKLITSAYYNEMQRSQAQYAGEENRHQFRAIVAE LRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKREWLAKTFYRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNLDQWIRNKQAHPIDLPSHLFDSKIMELLKVKDGGKKKWNEAFKDWSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVRDKKRELRTAGKPVPPDLAAIYKRSFHRAVNEREFMLRLVQEDDRLMLMAINKIMTDREEDILPGLKNIDSILDKENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILIRNKAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL |
| M573_1170<br>42 [中间普雷沃菌 ZT]<br>(SEQ ID No. 140)                | KJJ8675<br>6   | MKMEDDKKTTESTNMLDNKHFWAFLNLARHNVYITVNHINKVLELKNKKDQDIIDNDQDILAIKTHWEKVNGLNKTRELRLMTHFPFLETAIYTKNKEDKEEVKQEKQAEAQSLKDCFLFLEKLQEARNYSHYKYSES TKPEMLEEGLLEKMYNIFDDNIQLVIKDYQHNDINPDEDFKHLDRKGQFKYSFADNEGNITESGLLFFVSLFLEKKDAIWMQKQLTGFKDNRESKKKMTHEVFCRRRMLLPKLRLESTQTQDWILLDMLNELIRCPKSLYERLQGEYRK KFNVPFSDADEDYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFTNLRQIDLGTYHFSIYKLLIGGQKEDRHLTHKLYGFERIQEFAKQNRDPDEWKALVKDLDTYETSNERYISETTPHYHLENQKIGIRFRNGNKEIWP SLKTNGENNEKSKYKLDKPYQAEAFLSVHELLPMMFYLLKKEEPNNDKKNASIVEGFIKREIRDMYKLYDAFANGEINNIGDLEKYCEDKGIPKRHLPKQMVAILYDEPKDMVKEAKRKQKEMVKDTKKLLATLEKQTQEEIEDGGRNIRLLKSGEIARWL VNDMMRFQPVQKDNENGNPLNNSKANSTEYQMLQRSALALYNKEKPTRYFRQVNLINSSNPHFLK WTKWEECNILSFYRNYLTKKIEFLNKLKPEDWEKNQYFLKKEPKTNRETLVQGWKNGFNLPGRIFTEPIREWFKRHQNSKEYEKVEALKRVLVTKVIPLFFKEEYFKEDAQKEINNCVQPFYSFPYNVGNIIHKPDEKDFLPSEERKKLWGDKKDKFKGYKAKVSKKLTDEKKEEYRSYLEFQSWNKFERELRLVRNQDIVTWLLCTELIDKMKVEGLNVEELQKLRLKDIDTDTAKQEKNILNRIMPMLPVTVYEIDDSHNIVKDRPLHTVYIEETKTKLLKQGNFKALVKDRRLNGLFSFVDTSSKAELKDKPISKSVVEYELGEYQNARIETIKDMLLEKTLIKKYEKLPTDNFSDMLNGWLEGKDESDKARFQNDVKLLVAVRNAFSHNQYPMRNRIFAANINPFSLSADISEEKKLDIANQLKDKTHKIIKIEIEKPIETKE                    |
| A2033_1020<br>5 [拟杆菌]  | OFX180<br>20.1 | MENQTQKGGKIYIYTKNEDKHYFGSFLNLANNNIEQIIIEFRIRLSLKDEKNIKEIINNYFTDKKSYTDWGERGINILKEYLPVIDYLDLAITDKEFEKIDLK  |

[0572]

|   |                    |   |
|---|--------------------|---|
| <p>(Bacteroidetes bacterium)G WA2_31_9] (SEQ ID No. 141)</p>                            |                    | <p>QKETAKRKYFRTNFSLLDITIIDLRNFYTHYFHKPISINPDVAKFLDKNLLN<br/>VCLDIKKQKMKTDKTKQALKDGLDKELKLIELKKAELKEKKIKTWNITE<br/>NVEGAVYNDAFNHMVYKNNAGVTILKDYHKSILPDDKIDSELKLNFSISG<br/>LVFLLSMFLSKKEIEQFQSNLEGFKGKVVIGENGEYEISKFNNSLKYMATH<br/>WIFSYLTFKGLKQVRKNTFDKETLLMQMIDELNKVPHEVYQTLSEKQON<br/>EFLEDINEYVQDNEENKKSMENTSIVVHPVIRKRYDDKFNFAIRFLDEFA<br/>NFPTLKFFVTAGNFVHDKREKQIQGSMILTSRMIKEKINVFGLTEIAKY<br/>KSDYFSNENTLETSEWELFPNPSYLLIQNNIPVHIDLHNTEEAKQCQIAID<br/>RIKCTTNPAAKRNTRKSKEEIKIYQKNKNIKYGDPTALLSSNELPALIYEL<br/>LVNKKSGKELENIIVEKIVNQYKTIAGFEKGQNLNSLITKLLKSEPNE<br/>KINAEKIILAINRELEITENKLNIIKNNRAEFRTGAKRKHIFYSKELGQEAT<br/>WIAYDLKRFMPEASRKEWKGFFHSELQKFLAFYDRNKNDKALLNMFV<br/>NFDNDQLIGNDLNSAFREFHFDKFEKYLKIRDEILEGFKSFISNFKDEPKL<br/>LKKGKIDYRVFDKRYIYIKSTNAQKEQLLSKPICLPRGIFDNKPTYIEGVK<br/>VESNSALFADWYQYTYSDKHEFQSFYDMPRDYKEQFEKFEFNKSIQN<br/>KKNLNKSDKFIYFRYKQDLKIKQIKSQDLFIKLMVDELNVVFKNNIELNL<br/>KKLYQTSDERFKNQLIADVQKNREKGDTSNKMENFIWNMTIPLSLCN<br/>GQIEEPKVKLKDIGKFRKLETDDKVIQLEEDKSKVWKKLEIEDELENMP<br/>NSYERIRREKLLKGIQEFHFLLEKEKFDGINHPKHFEQDLNPNFKTYVIN<br/>GVLRKNSKLNYTEIDKLLDLEHISIKDIETSAKEIHLAYFLIHRNKVFGHNQ<br/>LPKLEAFELMKKYKKNNEETYAEYFHKVSSQIVNEFKNSLEKHS</p>   |
| <p>SAMN0542 1542_0666 [济州岛金黄杆菌 (Chryseobacterium jejuense)] (SEQ ID No. 142)</p>        | <p>SDI2728 9.1</p> | <p>MEKTQTGLGIYDHTKLQDKYFFGGFFNLAQNNIDNVIKAFIIFPERKD<br/>KDINIAQFLDICFKDNDADSDFKKKNKFLRIHFPVIGFLTSNDKAGFKKK<br/>FALLKTISELRNFYTHYHKSIEFPSELFELDDIFVKTTSEIKLKKKDDK<br/>TQQLLNKNLSEEYDIRYQQQIERLKLKAQGKRVS LTDETAIRNGVFNA<br/>FNHLIYRDGENVKPSRLYQSSYSEPDPAENGISLSQNSILFLLSMFLERKET<br/>EDLKSrvKGFKAIIKQGEEQISGLKFMAHWVFSYLCFKGIKQLSTEFH<br/>EETLLIQIIDELSKVPDEVYSAFDSKTKEKFLLEDINEYMKEGNADLSLEDS<br/>KVIHPVIRKRYENKFNFAIRFLDEYLSSTSLKFQVHVGNVYVHRRVKHI<br/>NGTGFQTERIVKDRIKVFGRLSNISNLKADYIKEQLELPNDSNGWEIFPNP<br/>SYIFIDNNVPIHVLADEATKKGIELFKDKRRKEQPEELQKRKGKISKYNI<br/>MIYKEAKGDKLRIDEPLALLSLNEIPALLYQILEKGATPKDIELIKNKLT<br/>RFEKIKNYDPETPAPASQISKRLRNNTAKGQEAALNAEKLIIEREIEN<br/>TKLSSIEEKRLKAKKEQRRNTPQRSIFSNSDLGRIAAWLADDIKRMPAEQ<br/>RKNWKGYQHSQQLQSLAYFEKRPQEAFLLLKEGWDTSDGSSYWNWV<br/>MNSFLENNHFEKFKNYLMKRVKYFSELAGNIKQHTHTKFLRKFQKQ<br/>MPADLFPKRHYILKDLETEKNKVLKPLVFSRGLFDNNPTFIKGVKVTEN<br/>PELFAEWYSYGYKTEHVQHFYGWERDYNELLDSELQKGNFAKNSIYY<br/>NRESQLDLIKLQDLKIKIKIQDLFLKRIAEKLFENVFNYPTTSLDEFYL<br/>TQEERAEKERIALAQLSREEGDNSPNIKDDFIWSKTIAFRSKQIYEPAILK<br/>DIGKFNRFVLDDEESKASKLLSYDKNKIWNKEQLERELSIGENSYEVIRRE<br/>KLFKEIQNLELQILSNWSWDGINHPREFEMEDQKNTRHPNFKMYLVNGIL<br/>RKNINLYKEDEDFWLESKENDFKTLPSEVLETKSEMVLFLVILIRNQF<br/>AHNQLPEIQFYNFIRKNYPEIQNNTVAELYLNLIKLA VQKLDNS</p> |
| <p>SAMN0544 4360_11366 [肉鸡白痢金黄杆菌 (Chryseobacterium carnipullorum)] (SEQ ID No. 143)</p> | <p>SHM528 12.1</p> | <p>MNTRVTGMGVSYDHTKKEDKHFFGGFLNLAQDNITAVIKAFCIKFDKNP<br/>MSSVQFAESCFTDKDSDTDFQNKVRYVRTHLPVIGYLNYYGGDRNTFRQK<br/>LSTLLKAVDSLRFYTHYHSPALSTELFELDTVFASVAVEVKQHMK<br/>KDDKTRQLLSKSLAEELDIRYKQQLERLKLKEQGKNIDLRDEAGIRNGV<br/>LNAAFNHLIYKEGEIAKPTLSYSSFYGADSAENGITISQSGLLFLLSMFLG<br/>KKEIEDLKSIRIRGFKAKIVRDGEENISGLKFMAHWVFSYLSFKGMKQRLS<br/>TDFHEETLLIQIIDELSKVPDEVYHDFDTATREKVFEDINEYIREGNEFSL<br/>GDSTIHPVIRKRYENKFNFAVRFLEFIKFPPLRFQVHGLGNFVHRRRIK<br/>IHGTGFQTERVVKDRIVFGKLEISSLKTEYIEKELDLSDTGWEIFPNPS<br/>YVFIDNNIPIYISTNKTFKNGSSEFIKLRKKEKPEEMKMRGEDKKEKRDIAS<br/>MIGNAGSLNSKTPLAMLSLNEMPALLYEILVKKTTPEEIELIIEKELDSHFE<br/>NIKNYDPEKPLPASQISKRLRNNTTDKGGKVINPEKLIHLINKEIDATEAKF</p>  |

[0573]

|  |                  |   |
|--|------------------|---|
|  |                  | ALLAKNRKELKEKFRGKPLRQTIFSNMELGREATWLADDIKRFMPDILRK<br>NWKGYQHNQLQQSLAFFNSRPKEAFTILQDGDWDFADGSSFWNGWIINSF<br>VKNRSFEYFYEAIFYEGRKEYFSSLAENIKQHTSNHRNLRRFIDQQMPKGL<br>FENRHYLLENLETEKNKILSKPLVFPRLGFDTKPTFIKGIKQVDEQPELFAEW<br>YQYGYSTEHVFNFYGWERDYNDLLESELEKDNDFSKNSIHYSRTSQLE<br>LIKQDLKIKIKIQDLFLKLIAGHIFENIFKYPASFSLDELTYLQTEERLNK<br>EQEALIQSQRKEGDHSDNIKDNFIGSKTVTYESKQISEPNVKKLDIGKFNK<br>FLDDKVKTLTLLSYNEDKVWNKNDLLELSIGENSYEYVIRREKLFKKIQNF<br>ELQTLTDWPWNGTDHPEEFGTTDNKGVNHPNFKMYVNVNGILRKHTDWF<br>KEGEDNWLENLNETHFKNLSFQELETKSKSIQTAFLIIMIRNQFAHNQLPA<br>VQFFEFIQKKYPEIQGSTTSELYLNFNINLA VVELLELLEK   |
| SAMN0542<br>1786_10111<br>19 [解脲金<br>黄杆菌<br>( <i>Chryseobact<br/>erium<br/>ureilyticum</i> )]<br>(SEQ ID No.<br>144) | SIS7048<br>1.1   | METQILGNISYDHTKTEDKHFFGGFLNTAQNNIDLLIKAYISKFESSPRKL<br>NSVQFPDVCFKKNDSDADFQHKLQFIRKHLVPIQYLKYGGNREVLKEKF<br>RLLLQAVDSLRFNYTHFYHKPIQLPNELLTLLDTIFGEIGNEVRQNKMKD<br>DKTRHLLKKNLSEELDFRYQEQLERLRKLKSEGKQVLDLDRDTEAIRNGVL<br>NAAFNHLIFKDAEDFKPTVSYSSYYYDSDTAENGISISQSGLLFLLSMFLG<br>RREMEDLKSRRVGRFKARIKHEEQHVSGLKFMATHWVFSEFCFKGIKTRL<br>NADYHEETLLIQLIDELSKVPDELYRSFDVATRERFIEDINEYIRDGKEDKS<br>LIESKIVHPVIRKRYESKFNYFAIRFLDEFVNFPTLRFQVHAGNYVHDRIK<br>SIEGTGFKTERLVKDRIKVFGLKSTISSLKAAYLAKAVNITDDTGWELLPH<br>PSYVFIDNNIPIHLTVDPSPFKNGVKEYQEKRLKQPEEMKNRQGGDKMH<br>KPAISSKIGKSKDINPESPVALLSMNEIPALLYEILVKKASPEEVEAKIRQKL<br>TAVFERIRDYDPKVPPLPASQVSKRLRNNTDLSYNKEKLVELANKEVEQT<br>ERKLALITKNRRECREKVKGKFKRQKVFKNALGTEATWLANDIKRFMP<br>EEQKKNWKGYQHSQSLAFFESRPGEARSLQAGWDFSDGSSFWNG<br>WVMNSFARDNTFDGFYESYLNMRMXYFLRLADNIAQQSSTNKLISNFIKQ<br>QMPKGLFDRRLYMLLEDLATEKNKILSKPLIFPRGIFDDKPTFKKGVQVSEE<br>PEAFADWYSYGYDVKHKFQEFYAWDRDYEELLREELEKDTAFTKNSIH<br>YSRESQIELLAKKQDLKVKVRIQDLYLKLMAEFLFENVFGHELALPLDQ<br>FYLTQEERLQEQEAIVQSQRPKGDDSPNIVKENFIWSKTIPFKSGRVFEP<br>NVKLDIGKFRNLLTDEKVDILLSYNNTTEIGKQVIENELIIGAGSYEFIRRE<br>QLFKEIQQMKRLSLRSVRGMGVPRLNLK   |
| 颊普雷沃菌<br>(SEQ ID No.<br>145)   | WP_004<br>343581 | MQKQDKLFDVRKKNAIFAPKPYITIMENQEKPEPIYYELTDKHFWAFLN<br>LARHNVYTTINHNRRLEIAELKDDGYMMDIKGSWNEQAKKLDKRVRLR<br>DLIMKHPPFLEAAA YEITNSKSPNNKEQREKEQSEALSNNLKNVLFIFLE<br>KLQVLRNYYSHYKYSEESPKPIFETSLLKNMYKVF DANVRLVKRDYMH<br>ENIDMQRDFTHLNRKKQVGRTKNIIDSPNFHYHFADKEGNMTIAGLLFFV<br>SLFLDKKDAIWMQKLLKGFKDGRNLREQMTNEVFCRSRISLPKLENV<br>QTKDWMQDMLNELVRCPSLYERLREKDRSEFKVPFDIFSDDYDAEEE<br>PFKNTLVRHQDRFPYFVLRVYFDLNEIFEQLRFQIDLGTYHFSIYNKRIGDE<br>DEVRLHLLHLYGFARIQDFAQQNQPEVWRKLVKDLDFEASQEPYIPKT<br>APHYHLENEKIGIKFCSTHNNLFPKTEKTCNGRSKFNLTQFTAEAFSL<br>VHELLPMMFYLLTKDYSRKESADKVEGIIRKEISNIYAIYDAFANGEIN<br>SIADLTCRLQKTNILQGHLPKQMSILEGRQKMEKEAERKIGEMIDDTQR<br>RLDLLCKQTNQKIRIGKRNAGLLKSGKIADWLVNDMMRFQPVQKDQNNI<br>PINNSKANSTEYRMLQRALALFGSENFRLKAYFNQMNLVGNDNPHPFLA<br>ETQWEHQTNILSFYRNYLEARKKYKGLKQPQNWQYQHFLILKVQKTNR<br>NTLVTGWKNSFNLPRGIFTQPIREWFEKHNSKRIYDQILSFDRVGFVAK<br>AIPLYFAEYKDNVQPFYDYPFNIGNKLPQKQGLDKKERVWQKKNK<br>ELFKNYPSEKKTDLAYLDFLSWKKFERELRLIKNQDIVTWLMFKELFNM<br>ATVEGLKIGEIHLRIDTNTANEESNNILNRIMPMLKPVKTYETDNKGNIL<br>KERPLATFYIEETETKVLKQGNFKVLAKDRRLNGLLSFAETTDIDLEKNPI<br>TKLSVDHELIKYQTTTRISIFEMTLGLEKLINKYPTLPTDSFRNMLERWLQ<br>CKANRPELKNYVNSLIAVRNAFVSHNQYPMYDATALFAEVKFKFTLFPVSDT<br>KKIELNIAPQLLEIVGKAIKEIEKSENKN |
| 牙龈卟啉单  | WP_005           | MNTVPASENKQSRVTEDDPQYFGLYLNLAENLIEVESHVRIKFGKKKL  |

[0574]

|                          |               |   |
|--------------------------|---------------|---|
| 胞菌 (SEQ ID No. 146)      | 873511        | NEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHD<br>TGVDPDASAQRILRELYSLDFLRNDFSHNRLDGTTFEHLEVSPDISSFITGT<br>YSLACGRAQSRFADFFKPDFFVLAKNRKEQLISVADGKECLTVSGLAFFI<br>CLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDLRHPHDLRLESSN<br>TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMMNLSNSLN<br>EESRLLWDGSSDWAELTKRIRHQDRFPYMLRFFIEEMDLLKIRFRVDL<br>GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br>YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV<br>HNLRLKLLMELLCEGSFSRMSQSDFLRKANRILDETAEGKLQFSALFPEMR<br>HRFIPPQNPKSKDRREKAETTLEKYKQEIKGRKDKLNSQLLSAFDMNQRR<br>LPSRLLDEWMNIRPASHSVKLRITYVKQLNEDCRLRLRFRKDGDKGKARA<br>IPLVGEMATFLSQDIVRMIISEETKLLITSAYYNEMQRSLAQYAGEENRRQ<br>FRAIVAEHLHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAKT<br>YRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQDWIRNKQA<br>HPIDLPSHLFDSKIMELLKVKDGGKKWNEAFKDWSTKYPDGMQPFY<br>LRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVQDKKRELRTAGKPVPP<br>DLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPG<br>LKNIDSILDEENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWS<br>KYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLEGEYDRCRIKIFDWAFALE<br>GAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILI<br>RNKAAHNQFPCAEMPLIYRDVSAKVGSIIEGSSAKDLPEGSSLVDSLWK<br>KYEMIIRKILPILDENRFFGKLLNMSQPINDL   |
| 牙龈卟啉单胞菌 (SEQ ID No. 147) | WP_005 874195 | MTEQNEKPYNGTYTLEDKHFWAFFNLARHNAYITLAHIDRQLAYS<br>ADITNDEDILFFKGQWKNLDNDLERKARLRSLIKHFSFLEGAAYGKKLF<br>ESQSSGNKSSKKKELTKKEKEELQANALSLDNLKSILFDLQKLDKDFRNY<br>YSHYRHPESSELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNDKVDPHR<br>HFNHLVRKGGKDKYGNNDNPFKHHFVDREKVTAEGLLFFVSLFLEKR<br>DAIWMQKKIRGFKGGTEAYQOMTNEVFCRSRISLPLKLESRTDDWML<br>LDMLNELVRCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPEDRH<br>LTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDKPYITQTPHYHI<br>EKKGIGLRFVPEGQLLWPSPEVGATRTGRSKYAQDKRFTAFAFLSVHEL<br>MPMMFYFLLREKYSEEASAQVQGRKRVIEDVYAVYDAFARDEINTR<br>DELDACLADKGIIRRHLPQMIALSQEHKDMEEKVRKQLQEMIAIDTH<br>RLDMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSG<br>KPLNNSKANSTEYRMLQRALALFGGKERLTPYFRQMNLTGGNNPHFPL<br>HETRWEHTNLSFYRSYLKARKAFLQSIGRSDREENHRFLLKPKTDRQ<br>TLVAGWKSEFHLPRGIFTEAVRDCLEMGYDEVGSYKEVGFMAKAVPLY<br>FERACKDRVQPFYDYPFNVGNLSLPPKGRFLSKEKRAEWEWGKERFRD<br>LEAWSHSARRIEDAFVGIYASWENKKKIEQLLQDLSLWETFESKLVK<br>ADKINIAKKEILEAKEHPYHDFKSWQKFERELRLVKNQDIITWMMCRD<br>LMEENKVEGLDTGTYLKDITDVEQEGSLNVLNHVKPMLRPLVVVYRA<br>DSRGHVHKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTG<br>ALAMEQYPIKLRVEYELAKYQATARVCAFEQTLLEESLTRYPHLPDES<br>FREMLESWSDPLLDKWPDLQREVRLIAVRNAFHNQYPMYDETIFSSIR<br>KYDPSSSLDAIEERMGLNIAHRLSEEVKLAKEMVERIIQA |
| 苍白普雷沃菌 (SEQ ID No. 148)  | WP_006 044833 | MKEEEKGKTPVVSTYNKDDKHFWAFFNLARHNAYITVNHINKILGE<br>INRDGYENTLEKSWNEIKDINKDRLSKLIKHFPFLEVTTYQRNSADTTK<br>QKEEKQAEASLESLSKSFVFIYKLRDLRNHYSHYKHSKSLERPKEED<br>LQEKMYNIFDASIQLVKEDYKHNNDIKTEEDFKHLDRKGQFKYSFADNE<br>GNITESGLLFFVSLFLEKKDAIWVQKLEGFKCSNESYQKMTNEVFCRSR<br>MLLPKLRQSTQTQDWILLMLNELIRCPKSLYERLREEDRKKFRVPIEIA<br>DEDYDAEQEPFKNALVRHQDRFPYFALRYFDYNEIFTNLRQIDLGTYHF<br>SIYKKQIGDYKESHHLTHKLYGFERIQEFTKQNRPDEWRKFKVTFNSFETS<br>KEYIPIETPHYHLENQKIGIRFRNDNDKIWPSLKTNSEKNEKSKYKLDKS<br>FQAEAFLSVHELLPMMFYLLKLTENTDNDNEIETKKKENKNDKQEKHK   |

[0575]

|                                 |                  |   |
|---------------------------------|------------------|---|
|                                 |                  | IEEIIENKITEIYALYDAFANGKINSIDKLEEYCKGKDIEIGHLPKQMIAILKS<br>EHKDMATEAKRKQEEMLADVQKSLESLDNQINEEIEIENVERKNSSLSGSEI<br>ASWL VNDMMRFQPVQKDNENPLNNSKANSTEYQMLQRSLALYNKEE<br>KPTRYFRQVNLIESSNHPFLNNTWEWEKCNNILSFYRSYLEAKKNFLES<br>PEDWEKNQYFLMLKEPKTNCETLVQGWKNGFNLPGRIFTEPIRKWFMEH<br>RKNITVAELKRVGLVAKVIPLFFSEEYKDSVQPFYNYLFVGNINKPDEK<br>NFLNCEERRELLRKKKDEFKKMTDKEKEENPSYLEFQSWNKFERELRLV<br>RNQDIVTWLLCMELFNKKKIKELNVEKIYLNKINTNTTKKEKNTEEKNGE<br>EKIIKEKNILNRIMPMLPIKVYGRENFSSKNKKKIRRNTFFTVYIEEKG<br>KLLKQGNFKALERDRRLGGLFSFVKTHSKAESKSNSTISKSRVEYELGEYQ<br>KARIEIHKDMLALEETLIDKYNSLDTDNFHNMLTGWLKLDKDEPKASFQ<br>DVDLLIAVRNAFSHNQYPMRNRIAFANINPFLSSANTSEEKGLGIANQLK<br>DKTHKTIEKIIIEKPIETKE  |
| 拟香味类香<br>味菌 (SEQ<br>ID No. 149) | WP_006<br>261414 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEEVNRNT<br>FGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAFASYFPILETVDKKS<br>ERNRTIDLTLRQIRQFREMLISLVAVDQLRNFYTHYHHSEIVIENKVL<br>FLNSSLVSTALHVKDYLKTDKTKEFKETAIAEELDILIEAYKKKQIEKKN<br>TRFKANKREDILNAIYNEAFWSFINDKDKDKETVVAKGADAYFEKNHHK<br>SNPDFALNISEKIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIK<br>YMATQRIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVNPVYQH<br>LSTTQQNSFIEDWNEYKYDYEDDVETDDLSRVIHVPVIRKRYEDRFNYFAI<br>RFLDEFFDFPTLRFQVHLGDYVHRRRTKQLGKVESDRIIKEKVTVFARLK<br>DINSAKANYFHSLEEQDKEELDNKWTLFPNPSYDFPKEHTLQHQGEQKN<br>AGKIGIYVCLRDTQYKEKAALAEARKSLNPKERSATKASKYDIITQIIAN<br>DNVKSEKPLVFTGQPIAYLSMNDIHSMLFSLTDAELKKTPEEVEAKLID<br>QIGKQINEILSKDITDKILKDYKDNLDKETDITDKITRDLARDKEEIEKLILE<br>QKQRADDYNYTSSTKFNIDKSRKRKHLFNAEKGKIGVWLANDIKRFMT<br>EEFKSKWKGYQHTLQKLFAYYDTSKSDLDLILSDMVMVKDYPIELIALV<br>KKSRTLVDLFLNKYLEARLGYMENVITRVKNSIGTPQFKTVRKECFTLK<br>SNYTVVSLDKQVERILSMPLFIERGFMDDKPTMLEGKSYQQHKEKFADW<br>FVHYKENSNYQNFYDTEVYEITTEDKREKAKVTKKIKQQQKNDVFTLM<br>MVNYMLEEVKLSSNDRLSLNELYQTKERIVNKQVAKDTQERNKNI<br>WNVVLDLQCEGLVRIDKVKLKDIGNFRKYENDSRVKEFLTYQSDIVWS<br>AYLSNEVDSNKLYVIERQLDNYESIRSKELLKEVIECSVYVNVANKES<br>LKQSGNENFKQYVLQGLVPIGMDVREMLLSTDVFKFIEEIIQLGQAGEVE<br>QDLYSLIYIRNKFAHNQLPIKEFFDFCENNYRSISDNEYAAYMEIFRSIK<br>EKYTS |
| 拟香味类香<br>味菌 (SEQ<br>ID No. 150) | WP_006<br>265509 | MKDILTTDTTEKQNRFYSHKIADKYFFGGYFNLASNNIYEVFEEVNRNT<br>FGKLAKRDNGNLKNYIIHVFKDELSISDFEKRVAFASYFPILETVDKKS<br>ERNRTIDLTLRQIRQFREMLISLVAVDQLRNFYTHYHHSEIVIENKVL<br>FLNSSLVSTALHVKDYLKTDKTKEFKETAIAEELDILIEAYKKKQIEKKN<br>TRFKANKREDILNAIYNEAFWSFINDKDKDKETVVAKGADAYFEKNHHK<br>SNPDFALNISEKIVYLLSFFLTNKEMDSLKANLTGFKGKVDRESGNSIK<br>YMATQRIYSFHTYRGLKQKIRTSEEGVKETLLMQMIDELSKVNPVYQH<br>LSTTQQNSFIEDWNEYKYDYEDDVETDDLSRVIHVPVIRKRYEDRFNYFAI<br>RFLDEFFDFPTLRFQVHLGDYVHRRRTKQLGKVESDRIIKEKVTVFARLK<br>DINSAKASYFHSLEEQDKEELDNKWTLFPNPSYDFPKEHTLQHQGEQKN<br>AGKIGIYVCLRDTQYKEKAALAEARKSLNPKERSATKASKYDIITQIIAN<br>DNVKSEKPLVFTGQPIAYLSMNDIHSMLFSLTDAELKKTPEEVEAKLID<br>QIGKQINEILSKDITDKILKDYKDNLDKETDITDKITRDLARDKEEIEKLILE<br>QKQRADDYNYTSSTKFNIDKSRKRKHLFNAEKGKIGVWLANDIKRFMT<br>KESKSKWKGYQHTLQKLFAYYDTSKSDLELILSDMVMVKDYPIELIDL<br>RKSRTLVDLFLNKYLEARLGYIENVITRVKNSIGTPQFKTVRKECFALKES<br>NYTVASLQKIERILSMPLFIERGFMDSKPTMLEGKSYQQHKEDEFADWFV<br>HYKENSNYQNFYDTEVYEITTEDKREKAKVTKKIKQQQKNDVFTLMV<br>NYMLEEVKLPSNDRSLNELYQTKERIVNKQVAKDTQERNKNIWNV   |

[0576]

|   |                  |   |
|---|------------------|---|
|   |                  | VVDLQLCEGLVRIDKVKLKDIGNFRKYENDSRVKEFLTYQSDIVWSGYLS<br>NEVDSNKLYVIERQLDNYESIRSKELLKEVQEIECIVYNQVANKESLKQSG<br>NENFKQYVLQGLLPRGTDVREMLLSTDVKFKKEEIMQLGQVREVEQDL<br>YSLIYIRNKFAHNQLPIKEFFDFCENNYRPISDNEYAAYEYMEIFRSIKEKY<br>AS   |
| 普雷沃菌<br>MSX73<br>(SEQ ID No.<br>151)                            | WP_007<br>412163 | MQKQDKLFVDRKKAIFAFPKYITIMENQEKPEPIYYELTDKHFWA AFLN<br>LARHNVTYTTINHINRRLEIAELKDDGYMMGIKGSWNEQAKKLDKVRRLR<br>DLIMKHFPFLEAAA YEITNSKSPNNKEQREKEQSEALSLNKNVLFIFLE<br>KLQVLRNYYSHYKYSEESPKPIFETSLLKNMYKVF DANVRLVKRDYMH<br>ENIDMQRDFTHLNRRKQVGRTKNIIDSPNFHYHFADKEGNMTIAGLLFFV<br>SLFLDKKDAIWMQKLLGFKDGRNLREQMTNEVFCRSRISLPKLLKLENV<br>QTKDWMQLDMLNELVRCPKSLYERLREKDRSFKVPFDIFSDDYDAEEE<br>PFKNTLVRHQDRFPYFVLRVYFDLNEIFEQLRFQIDLGTYHFSIYNKRIGDE<br>DEVRLHLLHLYGFARIQDFAPQNPPEWRKLVKDLDFHETSQEPYISKTA<br>PHYHLENEKIGIKFCSTHNNLFPKREKTCNGRSKFNLTQFTA EAFLSV<br>HELLPMMFYLLTKDYSRKESADKVEGIIRKEISNIYAIYDAFANNEINSI<br>ADLTCRLQKTNILQGHLPKQMISILEGRQKMEKEAERKIGEMIDDTQRR<br>LDLLCKQTNQKIRIGKRNAGLLKSGKIADWLVS DMMRFQPVQKDTNNA<br>INNSKANSTEYRMLQHALALFGSESSRLKAYFRQMNLVGNANPHPLAE<br>TQWEHQTNILSFYRNYLEARKKYLGKLPQNWKQYQHFLILKVKQTNR<br>NTLVTGWKNSFNLPRGIFTQPIREWF EKHNNSKRIYDQILSFDRVGFVAK<br>AIPLYFAEEYKDNVQPFYDYPFNIGNKLPQKQGGFDKKEVELWQKNK<br>ELFKNYPSEKNKTDLAYLDFLSWKKFERELRLIKNQDITWLMFKELFKT<br>TTVEGLKIGEIHLRDI DTNTANEESSNNILNRIMPMLKLPVKTYETDNKGNIL<br>KERPLATFYIEETETKVLKQGNFKVLAKDRRLNGLLSFAETTDIDLEKNPI<br>TKLSVDYELIKYQTTTRISIFEMTLGLEKKLIDKYSTLPTDSFRNMLERWLQ<br>CKANRPELKNYVNSLIAVRNAF SHNQYPMYDATLFAEVKKTLPSPVDT<br>KKIELNIAQQLLEIVGKAIKEIEKSENKN |
| 牙龈卟啉单<br>胞菌(SEQ<br>ID No. 152)                                  | WP_012<br>458414 | MTEQNERPYNGTYTLEDKHFWA AFFNLARHNA YITLAHIDRQLAYS KA<br>DITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSFLEGAAYGKKLFES<br>QSSGNKSSKKKELTKKEEELQANALSLDNLKSILFDLQKLDKDFRNYYS<br>HYRHPESELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNDKVDPHRH<br>NHLVRKGGKDRYGNNDNPFKHHFVDREKVT EAGLLFFVSLFLEKRD<br>IWMQKKIRGFGGTETYQMTNEVFCRSRISLPKLLKLESRTDDWMLLD<br>MLNELVRCPKSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTLV<br>RHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGE QPEDRH<br>RNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDKPYITQTTPHYHIEK<br>GKIGLRFVPEGQHLWPSPEVGATRTGRSKYAQDKRLTAEAFLSVHELMP<br>MMFYFLLREKYSDEASAERVQGRKRVIEDVYAVYDAFARGEINTRDE<br>LDA CLADK GIRRGHLPRQMIGILSQEHKDMEEKVRKKLQEMIVDTDHRL<br>DMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTS GKP<br>LNNSKANSTEYRMLQRALALFGGKERLTPYFRQMNLTGGNNPHPLHE<br>TRWESHTNLSFYRSYLKARKAFLQSIGRSDRVENHRFLLLEKPKTDRQTL<br>VAGWKGEFHLPRGIFTEAVRDCLIE MGLDEVGSYKEVGFMAKAVPLYFE<br>RACKDRVQPFYDYPFNVGNSLKP KGRFLSKEKRAEEWESGKERFRLAK<br>LKKEILEAKEHPYLD FKS WQKFERELRLVKNQDIITWMICRDLMEENKVE<br>GLDTGTLYLKDIRDVQE QGNLNLNRVKPMRLPVVVYRADS RGHVHK<br>EQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAMEQY<br>PISKLRVEYELAKYQTARVCAFEQTLELEESLLTRYPHLPDKNFRKMLES<br>WSDPLLDKWPDLHGNVRLLIAVRNAF SHNQYPMYDEAVFSSIRKYDPSS<br>PDAIEERMGLNIAHRLSEEVKQAKEMAERIIQA       |
| 丙酸杆菌<br>(Paludibacter<br>propionice<br>nes) (SEQ<br>ID No. 153) | WP_013<br>446107 | MKTSANNIYFNGINSFKKIFDSKGAIAPIAEKSCRNFDIKAQNDVNKEQRI<br>HYFAVGHTFKQLDTENLFEYVL DENLRAKRPTRFISLQFQDFKEFIENIKRLI<br>SDIRNINSHYIHRFDPLKIDA VPTNIIDFLKESFELAVIQIYLKEKGINYLQFS<br>ENPHADQKLVAF LHDKFLPLDEKKT SMLQNETPQLKEYKEYRKYFKTLS<br>KQAAIDQLLFAEKETDYIWNLFDSHPVLTISAGKYLSFYSCFLFLSMFLYK  |

[0577]

|                                 |                  |   |
|---------------------------------|------------------|---|
|                                 |                  | SEANQLISKIKGFKKNTTEEEKSKREIFTFFSKRFNSMIDIDSEENQLVKFRD<br>LILYLNHYPVAWNKDLELDSSNPAMTDKLSKSKIELEINRSFPLYEGNERF<br>ATFAKYQIWGKKHLGKSIEKEYINASFTDEEITAYTYETDTCPELKDAHK<br>KLADLKAAGLFGKRKEKNESDIKKTETSIRELQHEPNPIKDKLIQRIEKN<br>LLTVSYGRNQDRFMDFSARFLAEINYFGQDASFMYHFYATDEQNSELE<br>KYELPKDKKKYDSLKFHQGKLVHFISYKEHLKRYESWDDAFVIENNAIQL<br>KLSFDGVENTVTIQRALLIYLLEDALRNIQNNTAENAGKQLLQEYYSHNK<br>ADLSAFKQILTQQDSIEPQQKTEFKKLLPRLLNNSPAINHLQTPHSSLPL<br>ILEKALLAEKRYCSLVVKAKAEGNYDDFIKRNKGGKQFKLQFIRKAWNLM<br>YFRNSYLQNVQAAGHHKSFHIERDEFNDFSRYMFAFEELSQYKYLLNEM<br>FEKKGFFENNEFKILFQSGTSLENLYEKTQKQFEIWLASNTAKTNKPDNY<br>HLNNYEQQFSNQLFFINLSHFINYLKSTGKLQTDANGQIIEALNNVQYLI<br>PEYYYTDKPERSEKSGNKLYNKLKATKLEDALLYEMAMCYLKADKQI<br>ADKAKHPITKLLTSDVEFNITNKEGIQLYHLLVFPFKKIDAFIQLKMHKEQQ<br>DKKHPTSFLANIVNYLELVKNDKDIRKTYEAFSTNPVKRTLTYDDLAKID<br>GHLISIKFTNVTLELERYFIFKESLIVKKGNNIDFKYIKGLRNYNNEKK<br>KNEGIRNKAFHFGIPDSKSYDQLIRDAEVMFIANEVTPHATKYTDLNKKQ<br>LHTVCDKLMETVHNDYFSKEGDKKKKREAAGQKYFENIISAK   |
| 牙龈叶淋单<br>胞菌 (SEQ<br>ID No. 154) | WP_013<br>816155 | MTEQNEKPYNGTYTLEDKHFWAFFNLARHNAITLAHIDRQLAYSK<br>ADITNDEDILFFKGQWKNLNDLERKARLRSILKHFSELEGAAYGKKLF<br>ESQSSGNKSSKNKELTKKEKEELQANALSLDNLKSILFDLQKLDPRNY<br>YSHYRHPESSELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNDKVDPHR<br>HFNHLVVRKGGKDRYGNNDNPFKHHFVDREGTVTEAGLLFFVSLFLEKR<br>DAIWMQKKIRGFKGGTETYQQMTNEVFCRSRISLPLKLESLRTDDWML<br>LDMLNELVRCPSLYDRLREEDRARFRVPVDILSDEEDTDGAEEDPFKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFQIDLGTYHFAIYKKNIGEQPEDRH<br>LTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETGDKPYITQTPHYHI<br>EKKGIGLRFVPEGQHLWPSPEVGATRTGRSKYAQDKRFTAFAFLSAHEL<br>MPMMFYFLLREKYSEEASAERVQGRKRVIEDVYAVYDAFARDEINTR<br>DELDAACLADKGIIRRHLPQRMIGILSQEHKDMEEKIRKKLQEMMADTDH<br>RLDMLDRQTDKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSK<br>KPLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPL<br>HETRWESHNTILSFYRSYLKARKAFLQSIGRSDRVENHRFLLKPEKPTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMLDEVGSYKEVGFMAKAVPL<br>YFERACKDWVQPFYNYPFVNGNSLKPCKGRFLSKEKRAEEWESGKERFR<br>LAKLKEILEAKEHPYLDKFSWQKFERELRLVKNQDIITWMICGLMEEN<br>KVEGLDTGTLYLKDIRTDVQEQQSLNVLNRVKPMRPLVVVYRADSRRGH<br>VHKEQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAM<br>EQYPISKLRVEYELAKYQTARVCAFEQTLELEESLLTRCPHLPDKNFRKM<br>LESWSDPLLDKWPDLHRKVRLLIAVRNAFVSHNQYPMYDEAVFSSIRKYD<br>PSFPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA |
| 柱状黄杆菌<br>(SEQ ID No.<br>155)    | WP_014<br>165541 | MSSKNESYNKQKTFNHYKQEDKYFFGGFLNNAADDNLRQVGKEFKTRINF<br>NHNNELASVFKDYFNKEKSVAKREHALNLLSNYFPVLERIQKHTNHNK<br>EQTREIFELLLDTIKKLRDYYTHHYHKPITINPKIYDFLDDTLDDVLITIKK<br>KVKNDSRELLKEKLRPELTQLKNQKREELIKKGGKLEENLENAVFNHC<br>LRPFLEENKTDKQNKTVSLRKYRKSKEPNEETSITLTQSGLVFLMSFFLHR<br>KEFQVFTSGLEGFKAKVNTIKEEEISLNKNNIVYMITHWSYSYNNFKGLK<br>HRIKTDQGVSTLEQNNTTHTSLTNTNTKEALLTQIVDYLSKVPNEIYETLSE<br>KQKKEFEEDINEYMRENPENEDSTFSSIVSHKVIKRYENKFNYPAMRFL<br>DEYAEPLTRFMVNFVDYIKDRQKILESIQFDSERIIKKEIHLFEKLSLVTE<br>YKKNVYLKETSNIIDLSRFLFPNPSYVMANNIPFYIDSRNNLDEYLNQK<br>KKAQSQNKRNLTFEKYNKEQSKDAIAMLQKEIGVKDLQQRSTIGLLSC<br>NELPSMLYEVIVKDIKGAELNKIAQKIREQYQSIRDFTLSDPQKDNIPPTL<br>IKTINTDSSVTFENQPIDIPRLKNAIQKELTLTQEKLLNVKEHEIEVDNYNR<br>NKNTYKFKNQPKNKVDDKLLQRKYVFYRNEIRQEANWLASDLIHFMKN<br>KSLWKGVMHNEQLQSFLAFFEDKKNDCIALLETVFNLKEDCILTGLKLNLF  |

[0578]

|  |                  |  |
|--|------------------|--|
|  |                  | LKHGNFIDFYKEYLKLKEDFLNTESTFLENGLIGLPPKILKKELSKRFKYIFI<br>VFQKRQFIKELEEKKNLYADAINLSRGIFDEKPTMIPFKKPNPDEFASWF<br>VASYQYNNYQSFYELTPDIVERDKKKKYKNLRANKVKIQDYYLKLMVD<br>TLYQDLFNQPLDKSLDFYVSKAEREKIKADAKAYQKRNDSSLWNKVIH<br>LSLQNNRITANPKLKDIGKYKRALQDEKIATLLTYDDRTWTYALQKPEKE<br>NENDYKELHYTALNMELOEYEVRSKELLKQVQVELEKQILEEYTDFLST<br>QIHPADFEREGNPNFKKYLAHSILENEDDLKDLPEKVEAMRELDITITNPII<br>KKAIVLIIIRNKMAHNQYPPKFIYDLANRFVPKKEEYFATYFNRFVETITK<br>ELWENKEKKDKTQV  |
| 扭曲冷弯曲<br>菌<br>(Psychroflex<br>us torquis)<br>(SEQ ID No.<br>156) | WP_015<br>024765 | MESIIGLGLSFNPKYTADKHYFGSFLNLVENNLNAVFAEFKERISYKAKDE<br>NISSLIEKHFIDNMSIVDYEEKSILNGYLPIDFLDDELENNLNTRVKNFKK<br>NFILAEAIEKLRDYTHFYHDPITFEDNKEPILLELLDEVLLKTLTDVKKKY<br>LKTDKTKEILKDSLREEMDLLVIRKTDDELREKKTNPKIQHTDSSQIKNSIF<br>NDAFQGLLYEDKGNKKTQVSHRAKTRLNPKDIHQEERDFEPLSTSGL<br>VFLMSLFLSKKEIEDFKSNIKGFKGKVVKDENDHNSLKYMATHRVYSILAF<br>KGLKYRIKTDTFSKETLMMQMIDELSKVPDCVYQNLSETKQKDFIEDWN<br>EYFKDNEENTENLENSRVVHPVIRKRYEDKFNFAIRFLDEFANFKTLKF<br>QVFMGYIHDQRTKTIGTTNITTERTVKEKINVFGKLSKMDNLKHHFSSQ<br>LSDDENTDWEFFPNPSYNFLTQADNSPANNIPIYLELKNQIQIIEKDAIKA<br>EVNQTQNRNPNKPSKRDLNKLKTYEDFHQGDPTAILSLNEIPALLHLFL<br>VKPNNKTGQQIENIRIKIEKQFKAINHPSKNNKGIPKSLFADTNVRVNAIK<br>LKKDLEAELDMLNKKHIAFKENQKASSNYDKLLKEHQFTPKNKRPELKR<br>YVFKSEKGEATWLANDIKRFMPKDFKTKWKGQCQHSSELQRKLAFYDR<br>HTKQDIKELLSGCEFDHSLLDINAYFQKDNFEDFFSKYLENRIETLEGVLK<br>KLHDFKNEPTPLKGVFKNCFKFLKRQNYVTESPEIHKRILAKPTFLPRGVF<br>DERPTMKKGKNPLKDKNEFAEWFVEYLENKDYQKFYNAEEYRMRDAD<br>FKKNAVIKKQKLDKDFYTLQMVNYLLKEVFGKDEMNLQSELQTRQERL<br>KLQGIAKKQMNKETGDSSENTNRNQTYYIWNKDVVPSFFNGKVITDKVKLK<br>NIGKYKRYERDERVKTFIGYEVEDEKWMMYLPHNWKDRYSVKPINVIDL<br>QIQEYEEIRSHHELLKEIQNLEQYIYDHTTDKNILLQDGNPNFKMYVLNGLL<br>IGIKQVNPIDFIVLKQNTNFDKIDFTGIASCSELEKKTIIILAIRNKFAHNQLP<br>NKMIYDLANEFLKIEKNETYANYYLKVLKMKMISDLA |
| 鸭疫里默氏<br>杆菌 (SEQ<br>ID No. 157)                                  | WP_015<br>345620 | MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYKAHFTVNLTRDFAHLNRK<br>GKNKQDNPDFNRYRFEKDGFFTESGLLFFTNLFLDKRDAYWMLKKVSGF<br>KASHKQREKMTTEVFCSRILLPKLRLESRYDHNQMLLDMLSELRCPKL<br>LYEKLSEENKHFQVEADGFLDEIEEEQNPFKDTLIRHQDRFPYFALRYLD<br>LNEFSKSIRFQVDLGTYHYCIYDKKIGDEQEKRHLTRTLLSFGRLQDFTEI<br>NRPQEWKALTKDLDYKETSNOFFISKTPHYHITDNKIGFRLGTSKELYP<br>LEIKDGANRIAKYPYNSGFVAHAFISVHELLPLMFYQHLTGSSEDLK<br>VRHIQRIYKDFEERINTIEDLEKANQGRLPLGAFPKQMLGLLQNKQPDLS<br>EKAKIKIEKLIATKLLSHRLNTKLSKSPKLGKRREKLIKTVGLADWLVKD<br>FMRFPVAYDAQNQPIKSSKANSTEFWFIRRALALYGGEKNRLEGYFKQ<br>TNLIGNTNPHFPLNKFNWKACRNLVDFYQQYLEQREKFLAIAKHQPWEP<br>YQYCLLLKVPKENRKNLVKGEQGGISLPRGLFTEAIRETLTKDLTSLKPI<br>RKEIKKHGRVGFISRITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKKEE<br>HYEYWQQNKQSPTESQRLELHTSDRWKDYLLYKRWQHLEKLLRLYRN<br>QDIMLWMLTLELTKNHFKELNLNYHQLKLENLAVNVQEADAKLNPLNQ<br>TLPMLVPVKVYPTTAFGEVQYHETPIRTVYIREEQTKALKMGNFKALVK<br>DRRLNGLFSFIKEENDTQKHPISQLRLRRELEIYQSLRVDAFKETLSLEEK<br>LNKHASLSSENEFRTLLEEWKKKYAASSMVTDKHIAFIASVRNAFCHNQ<br>YPFYKETHAPILLFTVAQPTTEEKDGLGIAEALLKVLREYCEIVKSQI  |
| 胸膜炎普雷<br>沃菌<br>(Prevotella<br>pleuritidis)<br>(SEQ ID No.)       | WP_021<br>584635 | MENDKRLEESACYTLNDKHFWA AFLNLARHNVTYITVNHINKTLELKNKK<br>NQEIINDNDQDILAIKTHWAKVNGDLNKTDRLELMIKHFPFLEAAIYSNN<br>KEDKEEVKEEKQAKAQSFKSLKDCFLFLEKLQEARNYYSHYKYSESSKE<br>PEFEEGLLEKMYNTFDASIRLVKEDYQYNKDIDPEKDFKHLERKEDFNYL<br>FTDKDNKGKITKNGLLFFVSLFLEKKDAIWMQQKFRGFKDNRGNKEKMT  |

[0579]

|                          |                  |   |
|--------------------------|------------------|---|
| 158)                     |                  | HEVFCRSRMLLPKIRLESTQTQDWILLDMLNELIRCPKSLYERLQGAYRE<br>KFKVPFDSIDEDYDAEQEPFRNTLVRHQDRFPYFALRYFDYNEIFKNLRF<br>QIDLGTYHFSIYKKLIGGKKEDRHLTHKLYGFERIQEFTKQNRPKWQAI<br>KDLDTYETSNERYISETTPHYHLENQKIGIRFRNDNNDIWPSLKTNGEKN<br>KSKYNLDKPYQAEAFLSVHELLPMMFYLLKMENTDNDKEDNEVGTK<br>KKGKNNKQEKHKIEEIIENKIKDIYALYDAFTNGEINSIDELAEQREGKDI<br>EIGHLPKQLIVILKNKSKDMAEKANRQKQEMIKDTKKRLATLQKQVKEI<br>EDGGRNIRLLKSGEIRWLVNDMMRFQPVQKDNQKPLNNSKANSTEY<br>QMLQRSLALYNKEEKPTRYFRQVNLKSSNPHFLEDTKWEECYNILSFY<br>RNYLKAKIKFLNKLKPEDWKKNYFLMLKEPKTNRKTLVQGWKNGFNL<br>PRGIFTEPIKEWFKRHQNDSEEYKKVEALDRVGLVAKVIPLFFKEEYFKED<br>AQKEINNCVQPFYSPYVGNVGNHHPKPEEKNFLHCEERRKLDKDKKDFKQ<br>YKAKEKSKMTDKEKEEHRSYLEFQSWNKFERELRVRNQDILTWLLCT<br>KLIDKLIKIDELNIEELQKRLKIDIDTDTAKKEKNILNRVMPMRLPVTVYE<br>IDKSFNIVKDKPLHTVYIEETGTKLLKQGNFKALVKDRRLNGLFSFVKTS<br>EAESKSKPISKLRVEYELGAYQKARIDIKMDLALEKTLIDNDENLPTNKF<br>SDMLKSWLKGKGEANKARLQNDVGLLVAVRNAFVSHNQYPMYNSEVFK<br>GMKLLSLSSDIPEKEGLGIAKQLKDKIKETIERIEIEKEIRN   |
| 牙龈卟啉单胞菌 (SEQ ID No. 159) | WP_021<br>663197 | MNTVPASENKGQSRVTEDDPQYFGLYLNLARENLIEVESHVRIKFGKKL<br>NEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSK<br>TGVDPDSAQRLIRELYSLDFLRNDFSHNRDGTTFEHLVSPDISSFITGT<br>YSLACGRAQSRFADFFKPDDFVLAKNRKEQLISVADGKECLTVSGLAFFI<br>CLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDLRIRPHDRLESSN<br>TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNLSNSLN<br>EESRLLWDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKIRFRVDL<br>GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br>YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPRSMGFISV<br>HDLRKLMLMELLCEGFSRMSQDFLRKANRILDETAEGKLQFSALFPEMR<br>HRFIPPQNPKSKDRREKAETTLEKYKQEIQRKDKLNSQLLSAFDMDQRQ<br>LPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLKFRKDGDKARA<br>IPLVGEMATFLSQDIVRMIIEETKKLITSAYYNEQMSLAQYAGEENRHQ<br>FRAIVAEHLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAKT<br>YRPEQDENTKRRISVFFVDPGEARKLLPLLIRRRMKEQNDLQDWIRNKQA<br>HPIDLPSHLFDSKIMELLKVKDGKKWNEAFKDWWSSTKYPDGMQPFYQ<br>LRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTQVQDKKRELRTAGKPVPP<br>DLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPG<br>LKNIDSILDEENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWS<br>KYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFAL<br>GAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILI<br>RNKAAHNQFPCAAEMPLIYRDVSAKVGSIIEGSSAKDLPEGSSLVDSLWK<br>KYEMIIRKILPILDPENRFFGKLLNMSQPINDL |
| 牙龈卟啉单胞菌 (SEQ ID No. 160) | WP_021<br>665475 | MNTVPASENKGQSRVTEDDPQYFGLYLNLARENLIEVESHVRIKFGKKL<br>NEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSK<br>TGVDPDSAQRLIRELYSLDFLRNDFSHNRDGTTFEHLVSPDISSFITGT<br>YSLACGRAQSRFADFFKPDDFVLAKNRKEQLISVADGKECLTVSGLAFFI<br>CLFLDREQASGMLSRIRGFKRTNENWARAVHETFCDLRIRPHDRLESSN<br>TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNLSNSLN<br>EESRLLWDGSSDWAEALTKRIRHQDRFPYLMLRFIEEMDLLKIRFRVDL<br>GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br>YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV<br>HDLRKLMLMELLCEGFSRMSQSGFLRKANRILDETAEGKLQFSALFPEMR<br>HRFIPPQNPKSKDRREKAETTLEKYKQEIQRKDKLNSQLLSAFDMNQRQ<br>LPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLKFRKDGDKARA<br>IPLVGEMATFLSQDIVRMIIEETKKLITSAYYNEQMSLAQYAGEENRRQ<br>FRAIVAEHLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAKT<br>YRPEQDENTKRRISVFFVDPGEARKLLPLLIRRRMKEQNDLQDWIRNKQA  |

[0580]

|                          |               |   |
|--------------------------|---------------|---|
|                          |               | HPIDLPSHLFDSKIMELLKVKDGGKKKWNEAFKDWSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPGLKNIDSILDKENQFSLAVHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILIRNKAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYEMIIRKILPILDHENRFFGKLLNNMSQPINDL   |
| 牙龈卟啉单胞菌 (SEQ ID No. 161) | WP_021 677657 | MNTVPASENKGSRTVEDDPQYFGLYLNLARENLIEVESHVRIKFGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSTGVDPDSAQRLIRELYSLDFLRNDFSHNRLDGTTFEHLVSPDISSFITGYSLACGRAQSRFADFFKPDFFVLAKNRKEQLISVADGKECLTVSGLAFICLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDLRHPHDLRESSN TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLSENLSN EESRLLWDGSSDWAELTKRIRHQDRFPYMLRFFIEEMDLKIRFRVDLGEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV HDLRKLLLMELLCEGSFSRMQSGFLRKANRILDETAEGKLQFSALFPEMR HRFIPPQNPKSKDRREKAETTLEKYKQEIKGKDKLNSQLLSAFDMNQRLPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLKFRKDGDKARA IPLVGEMATFLSQDIVRMIIEETKKLITSAYYEMQSRSLAQYAGEENRRQ FRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAFTF YRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQDWIRNKQA HPIDLPSHLFDSKIMELLKVKDGGKKKWNEAFKDWSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVQDKKRELRTAGKPVPPDLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPGLKNIDSILDEENQFSLAVHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILIRNKAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYEMIIRKILPILDHENRFFGKLLNNMSQPINDL |
| 牙龈卟啉单胞菌 (SEQ ID No. 162) | WP_021 680012 | MNTVPASENKGSRTVEDDPQYFGLYLNLARENLIEVESHVRIKFGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSTGVDPDSAQRLIRELYSLDFLRNDFSHNRLDGTTFEHLVSPDISSFITGYSLACGRAQSRFADFFKPDFFVLAKNRKEQLISVADGKECLTVSGLAFICLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDLRHPHDLRESSN TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLSENLSN EESRLLWDGSSDWAELTKRIRHQDRFPYMLRFFIEEMDLKIRFRVDLGEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV HDLRKLLLMELLCEGSFSRMQSDFLRKANRILDETAEGKLQFSALFPEMR HRFIPPQNPKSKDRREKAETTLEKYKQEIKGKDKLNSQLLSAFDMDQRLPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLKFRKDGDKARA IPLVGEMATFLSQDIVRMIIEETKKLITSAYYEMQSRSLAQYAGEENRHQ FRAIV AELRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAFTF YRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQDWIRNKQA HPIDLPSHLFDSKIMELLKVKDGGKKKWNEAFKDWSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVRDKKRELRTAGKPVPPDLAAAIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPGLKNIDSILDKENQFSLAVHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILIRNKAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL |
| 牙龈卟啉单胞菌 (SEQ ID No. 163) | WP_023 846767 | MNTVPASENKGSRTVEDDPQYFGLYLNLARENLIEVESHVRIKFGKKKLNEESLKQSLLCDHLLSVDRWTKVYGHSSRRYLPFLHYFDPDSQIEKDHDSTGVDPDSAQRLIRELYSLDFLRNDFSHNRLDGTTFEHLVSPDISSFITGYSLACGRAQSRFADFFKPDFFVLAKNRKEQLISVADGKECLTVSGLAFICLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDLRHPHDLRESSN TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLSENLSN EESRLLWDGSSDWAELTKRIRHQDRFPYMLRFFIEEMDLKIRFRVDLGEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV HDLRKLLLMELLCEGSFSRMQSDFLRKANRILDETAEGKLQFSALFPEMR HRFIPPQNPKSKDRREKAETTLEKYKQEIKGKDKLNSQLLSAFDMDQRLPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLKFRKDGDKARA IPLVGEMATFLSQDIVRMIIEETKKLITSAYYEMQSRSLAQYAGEENRHQ FRAIV AELRLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAFTF YRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNDLQDWIRNKQA HPIDLPSHLFDSKIMELLKVKDGGKKKWNEAFKDWSTKYPDGMQPFYGLRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVRDKKRELRTAGKPVPPDLAAAIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPGLKNIDSILDKENQFSLAVHAKVLEKEGEGGDNSLSLVPATIEIKSKRKDWSKYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFALEGAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILIRNKAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWKKYEMIIRKILPILDPENRFFGKLLNNMSQPINDL |

[0581]

|   |               |  |
|---|---------------|--|
| ID No. 163)                                   |               | <p>TGVDPSAQRILRELYSLDFLRNDFSHNRLDGTTFEHLEVPDISSFITGT<br/> YSLACGRAQSRFADFVKPDDFVLAKNRKEQLISVADGKECLTVSGLAFFI<br/> CLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDLICIRPHDRLESSN<br/> TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMMNLSNSLN<br/> EESRLLWDGSSDWAELTKRIRHQDRFPYMLRFEIEMDLLKIRFRVDL<br/> GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br/> YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPRSMGFISV<br/> HDLRKLMLMELLCEGSFSRMQSDFLRKANRILDETAEGKLFQSFALPEMR<br/> HRFIPPQNPKSKDRREKAETTLEKYKQEIKGRKDKLNSQLLSAFDMNQRR<br/> LPSRLLDEWMNIRPASHSVKLRITYVKQLNEDCRLRLRKRFRKDGDKARA<br/> IPLVGEMATFLSQDIVRMIIESETKKLITSAYYEMQSRSLAQYAGEENRRQ<br/> FRAIVAEHLHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAKT<br/> YRPEQDENTKRRISVFFVDPDGEARKLLPLLIRRRMKEQNDLQDWIRNKQA<br/> HPIDLPSHLFDSKIMELLKVKDGKKKWNFAFKDWWSTKYPDGMQPFY<br/> LRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVDKRELRTAGKPVPP<br/> DLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPG<br/> LKNIDSILDEENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWS<br/> KYIRYRYDRRVPGLMSHFPEHKATLDEVKTLGLEYDRCRIKIFDWAFALE<br/> GAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILI<br/> RNKAAHNQFPCAAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWK<br/> KYEMIIRKILPILDPENRFFGKLLNMSQPINDL</p>  |
| 福氏普雷沃菌 (Prevotella falsenii) (SEQ ID No. 164) | WP_036 884929 | <p>MKNDNNSTKSTDYTLGDKHFWAAFLNLARHNVYITVNHINKVLELKNK<br/> KDQEIIDNDQDILAIKTLWGKVDTDINKKDRLELIMKHFPFLEAATYQQ<br/> SSTNNTKQKEEQAKAQSFESLKDCLFLFLEKLREARNYYSHYKHSKLE<br/> EPKLEEKLENMYNIFDTNVQLVIKDYEHNKDINPEEDFKHLGRAEGEFN<br/> YYFTRNKKGNITESGLLFFVSLFLEKKDAIWAQTKIKGFKDNRENKQKMT<br/> HEVFCRSRMLLPKLRLESTQTQDWILLDMLNELIRCPKSLYKRLQGEKRE<br/> KFRVFPDPADEYDAEQEPFKNTLVRHQDRFPYFALRYFDYNEIFNLRF<br/> QIDLGTYHFSIYKQIGDKEDRHLTHKLYGFERIQEFAKENRPDEWKAL<br/> VKDLDTFEESNEPYISETTPHYHLENQKIGIRNKNKKKKTIWPSLETKT<br/> VNERSKYNLGSFKAEAFLSVHELLPMMFYLLNKEEPNNGKINASKV<br/> EGIEKKIRDIYKLYGAFANEEINNEELKEYCEGKDIAIRHLPKQMIAILK<br/> NEYKDMAKKAEDKQKKMIKDTKKRLAALDKQVKGEVEDGGRNIKPLKS<br/> GRIASWLVNDMMRFQPVQRDRDGYPLNNSKANSTEYQLLQRTALFLGS<br/> ERERLAPYFRQMNLIQKDNPHPLKDTKWKEHNNILSFYRSYLEAKNFL<br/> GSLKPEDWKKNYFLKLEPKTNRETLVQGWKNGFNLPGRIFTEPIREW<br/> IRHQNESEYKVKDFDRIGLVAKVIPLFFKEDYQKEIEDYVQPFYGYPFN<br/> VGNIHNSQEGTFLNKKEREELWKGKTKFKDYKTEKNKEKTNKDKFK<br/> KKTDEEKEEFRSYLDQSWKKFERELRLVRNQDIVTWLLCMEIDKID<br/> ELNIEELQKLRLKDIDTDTAKKEKNILNRIMPMPVTVYETDDSNIIK<br/> DKPLHTIYKEAETKLLKQGNFKALVKDRRLNGLFSFVETSSEAEKSKPIS<br/> KSLVEYELGEYQRRARVEIKDMLRLEETLIGNDEKLPTNFRQMLDKWLE<br/> HKKETDDTDLKNDVKLLTEVRNAFSHNQYPMRDIAFANIKPFSLSANT<br/> SNEEGLGIACKLKDKTETIDRIIEIEEQATKR</p> |
| 胸膜炎普雷沃菌 (SEQ ID No. 165)                      | WP_036 931485 | <p>MENDKRLEESTCYTLNDKHFVAFLNLARHNVYITINHINKLLEIRQIDN<br/> DEKVLDIKALWQKVDKINQKARLRELMIKHFPFLEAATYSNNKEDKEEV<br/> KEEKQAKAQSFSLKDCFLFLEKLQEARNYSHYKSSESSKEPEFEEGLL<br/> EKMYNTFGVSIRLVKEDYQYNKDIDPEKDFKHLERKEDFNLFDTKDNK<br/> GKITKNGLLFFVSLFLEKKDAIWMQQLRGFKDNRGNKEKMTHEVFCRS<br/> RMLLPKIRLESTQTQDWILLDMLNELIRCPKSLYERLQGAYREKFKVPFDS<br/> IDEDYDAEQEPFRNTLVRHQDRFPYFALRYFDYNEIFKNLRFQIDLGTYHF<br/> SIYKKLIGDNKEDRHLTHKLYGFERIQEFAKQKRPNEWQALVKDLDIYET<br/> SNEQYISETTPHYHLENQKIGIRFKNKKDKIWPSETNGKENEKSKYNLD<br/> KSFQAEAFLSIHELLPMMFYDLLLKEEPNNDKNASIVEGFIKKEIKRMY<br/> AIYDAFANEEINSKEGLEEYCKNKGQERHLPKQMIAILTNKSKNMAEKA<br/> KRKQKEMIKDTKKRLATLQVQVKEIEDGGRNIRLLKSGEJARWLVNDM</p>   |

[0582]

|                                 |                  |  |
|---------------------------------|------------------|--|
|                                 |                  | MRFQSVQKDEKGPLNNSKANSTEYQMLQRSLALYNKEQKPTPYFIQVN<br>LIKSSNPHPFLEETKWEECNILSFYRSYLEAKKNFLESLEKPEDWKKNQYF<br>LMLKEPKTNRKTLVQGWKNGFNLPARGIFTEPIKEWFKRHQNDSEEEKV<br>EALDRVGLVAKVIPLFFKEEYFKEDAQKEINNCVQPFYSFPYVNGNIHKP<br>EENFLHCEERRKLWDKKDKFKGYKAKEKSKKMTDKEKEEHSYLEF<br>QSWNKFERELRLVRNQDIVTWLLCTELIDKLIKIDELNIEELQKRLKIDIT<br>DTAKKEKNNILNRIMPMQLPVTVEIDKSFNIVKDKPLHTIYIEETGTKLL<br>KQGNFKALVKDRRLNGLFSFVKTSSEAESKSKPISKLRVEYELGAYQKAR<br>IDIHKDMLALEKTLIDNDENLPTNKFSDMLKSWLKGKGEANKARLQNDV<br>DLLVAIRNAFASHNQYPMYNSVFKGMKLLSLSSDIPEKEGLGIKQLKDKI<br>KETIERIIEIEKEIRN  |
| 牙龈卟啉单<br>胞菌 (SEQ<br>ID No. 166) | WP_039<br>417390 | MTEQNERPYNGTYTLEDKHFWAFFNLARHNAITLAHIDRQLAYSKA<br>DITNDEDILFFKGQWKNLNDNLERKARLRSILKHFSLFEGAAYGKKLFE<br>QSSGNKSSKKELTKKEKEELQANALSLDNLKSILFDLQKLDKFRNYY<br>HYRHPESSELPLFDGNMLQRLYNVFDVSVQVRKRDHEHNDKVDPHRHF<br>NHLVRKGGKDRYGNNDNPFKHHFVDREGTVTEAGLLFFVSLFLEKRD<br>IWMQKKIRGFKGGTEAYQMTNEVFCRSRISLPKLESLRTRDDWMLLD<br>MLNELVRCPKSLYDRLREEDRARFRVPIDILSDEDDTDGTEEDPFKNTLV<br>RHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPEDRHLT<br>RNLYGFGRIQDFAEEHRPEEWKRLVRDLDFETGDKPYITQTPHYHIEK<br>GKIGLRFVPEGQHLWPSPEVGATRTGRSKYAQDKRLTAEAFSLVHELMP<br>MMFYFLLREKYSEEVSAAEKVQGRIKRVIEDVYAVYDAFARGEIDTLDR<br>LDAACLADKGIIRGHLPRQMAILSQECHKDMEEKVRKQLQEMIADTDHRL<br>DMLDRQTDKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSKGP<br>LNNSKANSTEYRMLQRALALFGGKERLTPYFRQMNLTGGNNPHPFLE<br>TRWESHTNLSFYRSYLKARKAFLQSIGRSDREENHRFLLKPKTDQRQTL<br>VAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEVGFMAKAVPLYFE<br>RACKDRVQPFYDYPFVGNLSLKPCKGRFLSKEKRAEWEWGKERFLAK<br>LKKEILEAKEHPYLDKFSWQKFERELRLVKNQDIITWMMCRDLMEENKV<br>EGLDTGTLYLKDIRTDVHEQGSNLVLRVVKPMRPLVVVYRADSRRGHVH<br>KEQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAMEQ<br>YPISKLRVEYELAKYQATARVCAFEQTLLEESLLTRYPHLPDKNFRKMLE<br>SWSDPLLDKWPDLHRKVRLLIAVRNAFASHNQYPMYDEAVFSSIRKYDPS<br>SPDAIEERMGLNIAHRLSEEKQAKEMAERIIQV |
| 古卟啉单胞<br>菌 (SEQ ID<br>No. 167)  | WP_039<br>418912 | MTEQSERPYNGTYTLEDKHFWAFFNLARHNAITLTHIDRQLAYSKA<br>DITNDQDVLFSKALWKNLNDNLERKSRLRSILKHFSLFEGAAYGKKLFE<br>SKSSGNKSSKNELTKKEKEELQANALSLDNLKSILFDLQKLDKFRNYY<br>SHYRHSGSSELPLFDGNMLQRLYNVFDVSVQVRKRDHEHNDKVDPHRH<br>FNHLVRKGGKDRYGHNDNPSFKHHFVDSEGMVTEAGLLFFVSLFLEKRD<br>AIWMQKKIRGFKGGTETYQMTNEVFCRSRISLPKLESLRMDWMLL<br>DMLNELVRCPKPLYDRLREDDRACFRVPVDILPEDDTDGGGEDPFKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPEDR<br>HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDFETGDKPYISQTSPPHYH<br>IEKGIGLRFMPEGQHLWPSPEVGATRTGRSKYAQDKRLTAEAFSLVHEL<br>MPMMFYFLLREKYSEEVSAAEKVQGRIKRVIEDVYAVYDAFARDEINTLK<br>ELDAACLADKGIIRGHLPKQMAILSQECHKDMEEKVRKQLQEMIADTDHRL<br>LDMLDRQTDKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDASGK<br>PLNNSKANSTEYRMLQRALALFGGKERLTPYFRQMNLTGGNNPHPFLE<br>DTRWESHTNLSFYRSYLKARKAFLERIGRSDRMENRPFLLKPKTDQRQ<br>TLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYREVGFMAKAVPLY<br>FERACEDRVQPFYDYPFVGNLSLKPCKGRFLSKEKRAEWEWGKERFRD<br>LEAWSHSAARRIEDAFAGIEYASPGNKKKIEQLLRDLSLWEAFESKLV<br>ADKINLAKLKEILEAQEHPYHDFKSWQKFERELRLVKNQDIITWMMCR<br>DLMEENKVEGLDTGTLYLKDIRTNVQEQGSNLVLRVVKPMRPLVVVYR<br>ADSRGHVHKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDT<br>GGLAMEQYPISKLRVEYELAKYQATARVCAFEQTLLEESLLTRYPHLPDK                                      |

[0583]

|                         |               |  |
|-------------------------|---------------|--|
|                         |               | NFRKMLESWSDPLLAKWPELHGKVRLLIAVRNAFASHNQYPMYDEAVFSSIRKYDPPSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA  |
| 古叶淋单胞菌 (SEQ ID No. 168) | WP_039 419792 | MTEQSERPYNGTYTLEDKHFWA AFLNLARHNAYITLTHIDRQLAYSKA<br>DITNDQDVL SFKALWKNLNDLERKSRLRSLILKHFSFLEGAAYGKKLFE<br>SKSSGNKSSKNKELTKKEKEELQANALSLDNLKSILFDLQKLDKFRNYY<br>SHYRHSGSSELPLFDGNMLQRLYNVFDVSVQVRKRDHEHNDKVDPHRH<br>FNHLVRKGGKDRYGHNDNPSFKHHFVDGEGMVTEAGLLFFVSLFLEKRD<br>AIWMQKKIRGFKGGTETYQ QMTNEVFCSRISLPKLESLRTDDWMLL<br>DMLNELVRCPKPLYDRLREKDRARFRVPVDILPEDDTDGGGEDPDKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKVIGEQPEDRH<br>LTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDFETGDKPYISQTTPHYHI<br>EKKGIGLRFVPEGQHLWPSPEVGTTRTGRSKYAQDKRLTAEAFLSVHEL<br>MPMMFYFLLREKYSEEVS AEKVQGRIKRVIEDVYAIYDAFARDEINTLD<br>ELDACLADKGIRRGHLPKQMIGILSQEHKNMEEKVRKQLQEMIADTDHR<br>LDMLDRQTDKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGK<br>PLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLD<br>ETR WESHNTILSFYRSYLARKAFLEIRGRSDRVENRPFLLKPKTDRQT<br>LVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEVGFMAKAVPLYF<br>ERACKDRVQPFYDSPFNVGNSLKP KGRFLSKEKRAEWEWGKERFRLA<br>KLLKEILEAQEHYPYHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENK<br>VEGLDTGTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVVYRADS RGHV<br>HKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGGLAME<br>QYPISKLRVEYELAKYQTARVCFELTLRLEESLSRYPHLPDES FREMLE<br>SWSDPLLAKWPELHGKVRLLIAVRNAFASHNQYPMYDEAVFSSIRKYDPS<br>SPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA     |
| 古叶淋单胞菌 (SEQ ID No. 169) | WP_039 426176 | MTEQSERPYNGTYTLEDKHFWA AFLNLARHNAYITLTHIDRQLAYSKA<br>DITNDQDVL SFKALWKNFNDLERKSRLRSLILKHFSFLEGAAYGKKLFE<br>SKSSGNKSSKNKELTKKEKEELQANALSLDNLKSILFDLQKLDKFRNYY<br>SHYRHSGSSELPLFDGNMLQRLYNVFDVSVQVRKRDHEHNDKVDPHYH<br>FNHLVRKGGKDRYGHNDNPSFKHHFV DSEGMVTEAGLLFFVSLFLEKRD<br>AIWMQKKIRGFKGGTGPYEQMTNEVFCSRISLPKLESLRTDDWMLL<br>DMLNELVRCPKPLYDRLREKDRACFRVPVDILPEDDTDGGGEDPDKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIGE QPEDR<br>HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDFETGDKPYISQTTPHYH<br>IEKKGIGLRFMPEGQHLWPSPEVGTTRTGRSKYAQDKRLTAEAFLSVHEL<br>MPMMFYFLLREKYSEEVS AEKVQGRIKRVIKDVYAIYDAFARDEINTLK<br>ELDAC SADKGIRRGHLPKQMIGILSQEHKNMEEKVRKQLQEMIADTDHR<br>LDMLDRQTDKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGK<br>PLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPFLD<br>ETR WESHNTILSFYRSYLARKAFLEIRGRSDRVENRPFLLKPKNDRQT<br>LVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEVGFMAKAVPLYF<br>ERACKDRVQPFYDSPFNVGNSLKP KGRFLSKEKRAEWEWGKERFRLA<br>KLLKEILEAKEHPYHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENK<br>VEGLDTGTLYLKDIRTDVHEQGSLNVLNRVKPMRLPVVVYRADS RGHV<br>HKEEQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGGLAME<br>QYPISKLRVEYELAKYQTARVCAFEQ TLELEESLLTRYPHLPDENFREML<br>ESWSDPLLGKWPDLHGKVRLLIAVRNAFASHNQYPMYDEAVFSSIRKYDP<br>SSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA |
| 古叶淋单胞菌 (SEQ ID No. 170) | WP_039 431778 | MTEQSERPYNGTYTLEDKHFWA AFLNLARHNAYITLTHIDRQLAYSKA<br>DITNDQDVL SFKALWKNFNDLERKSRLRSLILKHFSFLEGAAYGKKLFE<br>SKSSGNKSSKNKELTKKEKEELQANALSLDNLKSILFDLQKLDKFRNYY<br>SHYRHSESELPLFDGNMLQRLYNVFDVSVQVRKRDHEHNDKVDPHRH<br>FNHLVRKGGKDRYGHNDNPSFKHHFVDGEGMVTEAGLLFFVSLFLEKRD<br>AIWMQKKIRGFKGGTETYQ QMTNEVFCSRISLPKLESLRTDDWMLL<br>DMLNELVRCPKPLYDRLREDDRACFRVPVDILPEDDTDGGGEDPDKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIGE QPEDR   |

[0584]

|                                |                  |  |
|--------------------------------|------------------|--|
|                                |                  | HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDL DYFETGDKPYISQTS PHYH<br>IEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSKYAQDKRLTAEAFLSVHEL<br>MPMMFYFLLREKYSEEVS AEKVQGRIKRVIEDVYAIYDAFARDEINTLK<br>ELDA CLADK GIRRGHLPKQMIAILSQEHKDMEEKIRKKLQEMIADTDHRL<br>DMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTS GKP<br>LNNSKANSTEYRMLQRALALFGGEEKRLTPYFRQMNL TGGNPHPLHE<br>TRWESH TNILSFYRSYL RARKAFLE RIGRSDR MENRPFLL LKEPKTDRQTL<br>VAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYREVGFMAKAVPLYFE<br>RACEDRVQPFYDSPFN VGNLSL KPKGRFLSKEERAE EW ERGKERFRDLE<br>AWSHS AARRIEDAFAGIEYASPGNKKKIEQLLRDLSLWEAFESKLVKVRAD<br>KINLAKLKEILEAQEHYPHDFKSWQKFERELRLVKNQDIITWMMCRDL<br>MEENKVEGLDTGTL YLKD IRPNVQE QGSLNVLNRVKPMRLPVVYRAD<br>SRGHVHKEEA PLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGG<br>LAMEQY PISKLRVEYELAKYQTARVCFELTLRLEESLLTRYPHLPDESFR<br>KMLESWSDPLLAKWPELHGKVRLLI A VRNAF SHNQY PMYDEAVFSSIRK<br>YDPSSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQV   |
| 古叶琳单胞<br>菌 (SEQ ID<br>No. 171) | WP_039<br>437199 | MTEQSERPYNGTYTLEDKHFWA AFLN LARHNA YITLTHIDRQLAYS KA<br>DITNDEDILFFKQWKNLNDNLERKSRLRSLILKHFSFLEGAAYGKKFFES<br>KSSGNKSSKNKELTKKEKEELQANALSLDNLKSILFDLQKLKDFRNYYS<br>HYRHSGSSELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNDEVDPHYHF<br>NHLVRKGKKDRYGHNDNPSFKHHFVDGEGMVTEAGLLFFVSLFLEKRD<br>AIWMQKKIRGFKGGTEPYEQMTNEVFCSRISLPKLLKLESRTDDWMLL<br>DMLNELVRCPKPLYDRLREKDRACFRVPVDILPEDDTDGGGEDPFKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIGE QPEDR<br>HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDL DYFETGDKPYISQTT PHYH<br>IEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSKYAQDKRLTAEAFLSVHEL<br>MPMMFYFLLREKYSEEVS AEKVQGRIKRVIEDVYAIYDAFARDEINTLK<br>ELDA CLADK GIRRGHLPKQMIGILSQERKDMEEKVRKKLQEMIADTDHR<br>LDMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTS GKP<br>PLNNSKANSTEYRMLQRALALFGGEEKERLTPYFRQMNL TGGNPHPLFH<br>ETRWESH TNILSFYRSYL RARKAFLE RIGRSDR VENC PFLLLKEPKTDRQT<br>LVAGWKGEFHLPRGIFTEAVRDCLIEMGYDEVGSYREVGFMAKAVPLYF<br>ERACEDRVQPFYDSPFN VGNLSL KPKGRFLSKEERAE EW ERGKERFRLA<br>KLKKEILEAQEHYPHDFKSWQKFERELRLVKNQDIITWMMCRDLMEENK<br>VEGLDTGTL YLKD IRPNVQE QGSLNVLNRVKPMRLPVVYRADSRGHV<br>HKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGALAME<br>QYPISKLRVEYELAKYQTARVCAFEQTLELEESLLTRYPHLPDESFRMLE<br>SWSDPLLTKWPELHGKVRLLI A VRNAF SHNQY PMYDEAVFSSIWKYDPS<br>SPDAIEERMGLNIAHRLSEEVKQAKETIERIIQA |
| 古叶琳单胞<br>菌 (SEQ ID<br>No. 172) | WP_039<br>442171 | MTEQSERPYNGTYTLEDKHFWA AFLN LARHNA YITLTHIDRQLAYS KA<br>DITNDQDVL SFKALWKNLNDNLERKSRLRSLILKHFSFLEGAAYGKKLFE<br>SKSSGNKSSKNKELTKKEKEELQANALSLDNLKSILFDLQKLKDFRNYYS<br>SHYRHSGSSELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNKVDPHYH<br>FNHLVRKGKKDRYGHNDNPSFKHHFVDSEGMVTEAGLLFFVSLFLEKRD<br>AIWMQKKIRGFKGGTEPYEQMTNEVFCSRISLPKLLKLESRTDDWMLL<br>DMLNELVRCPKPLYDRLREKDRACFRVPVDILPEDDTDGGGEDPFKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIGE QPEDR<br>HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDL DYLETGDKPYISQTT PHYH<br>IEKGKIGLRFVPEGQHLWPSPEVGTTRTGRSKCAQDKRLTAEAFLSVHEL<br>MPMMFYFLLREKYSEEVS AEKVQGRIKRVIEDVYAIYDAFARDEINTLK<br>ELDTCLADK GIRRGHLPKQMITILSQERKDMKEKIRKKLQEMIADTDHRL<br>DMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDASGKP<br>LNNSKANSTEYRMLQRALALFGGEEKERLTPYFRQMNL TGGNPHPLHE<br>TRWESH TNILSFYRSYL RARKAFLE RIGRSDR VENC PFLLLKEPKTDRQTL<br>VAGWKDEFHLPRGIFTEAVRDCLIEMGYDEVGSYREVGFMAKAVPLYFE<br>RACEDRVQPFYDSPFN VGNLSL KPKGRFLSKEDRAE EW ERGGMERFRDLE  |

[0585]

|  |                  |  |
|--|------------------|--|
|  |                  | AWSHSAARRIKDAFAGIEYASPGNKKKIEQLLRDLSLWEAFESKLVKVRAD<br>KINLAKLKEILEAQEHYPYHDFKSWQKFERELRLVKNQDIITWMMCRDL<br>MEENKVEGLDTGTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVYRAD<br>SRGHVHKEAPLATVYIEERNTKLLKQGNFKSFVKDRRLNGLFSFVDTGGL<br>AMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLSRYPHLPDESFRE<br>MLESWSDPLAKWPELHGKVRLLIAVRNAFASHNQYPMYDEAVFSSIRKY<br>DPSSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQA   |
| 古叶淋单胞<br>菌 (SEQ ID<br>No. 173)   | WP_039<br>445055 | MNTVPATENKGSRTVEDDPQYFGLYLNLARENLIEVESHVRIKFGKKK<br>LNEESLKQSLLCDHLLSIDRWTKVYGHSSRRYLPFLHCFDPDPSGIEKDHS<br>KTGVDPDSAQRLIRELYSLDFLRNDFSHNRLDGTTFEHLKVSPISSFITG<br>AYTFACERAQSRFADFFKPDFFLLAKNRKEQLISVADGKECLTVSGFAFFI<br>CLFLDREQASGMLSRIRGFKRTDENWARAVHETFCDCIRHPHDLRESN<br>TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNLSNSLN<br>EESRLLWDGSSDWAEALTKRIRHQDRFPYMLRFEIEMDLLKIRFRVDL<br>GEIELDSYSKKVGRNGEYDRITDHALAFGKLSDFQNEEEVSRMISGEAS<br>YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV<br>HDLRKLMLMELLCEGSFSRMQSDFLRKANRILDETAEGKLFQSFALPEMR<br>HRFIPPQNPKSKDRREKAETTLEKYKQEIKGKDKLNSQLLSAFDMNQRR<br>LPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLRKRKFDGDKARA<br>IPLVGEMATFLSQDIVRMIISEETKLLITSAYYNEMQRSQAQYAGEENRRQ<br>FRAIVAEHLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAFTF<br>YRPEQDENTKRRISVFFVPDGEARKLLPLLIRRRMKEQNLDQWIRNKQA<br>HPIDLPSHLFDSKIMELLKVKDKGKKWNEAFKDWSTKYPDGMQPFYG<br>LRRELNIHGKSVSYIPSDGKKFADCYTHLMEKTVRDKKRELRTAGKPVPP<br>DLAAIYKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPG<br>LKNIDSILDEENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWS<br>KYIRYRYDRRVPGLMSHFPEHKATLDEVKTLLGEYDRCRIKIFDWAFAL<br>GAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILI<br>RNKAAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWK<br>KYEMIIRKILPILDHENRFFGKLLNMSQPINDL   |
| 犬咬二氧化<br>碳嗜纤维菌<br>(Capnocytop<br>haga<br>cynodegmi)<br>(SEQ ID No.<br>174) | WP_041<br>989581 | MENKTSLGNNIYNPFKPDQKSYFAGYLNAAAMENIDSVFRELGKRLKGGK<br>EYTSENFFDAIFKENISLVEYERYVKLLSDYFPMARLLDKKEVPIKERKEN<br>FKKNFRGIIKAVRDLRNFYTHKEHGEVEITDEIFGVLDMLKSTVLTVKK<br>KKIKTDKTKELKKSIEKQLDILCQKKLEYLKDRTARKIEEKRRNQREGEK<br>KLVPRFEYSDRRDDLIAAIYNDAFDVYIDKKKDSLKESKTKYNTESYPQ<br>QEEGDLKIPISKNGVVFLLSLFLSKQEVHAFKSKIAGFKATVIDEATVSHR<br>KNSICFMATHEIFSHLAYKLLKRVRTAEINYSEAENAEQLSIYAKETLM<br>MQMLDELSKVPDVVYQNLSEDVQKTFIEDWNEYLKENNGDVGTMEEE<br>QVIHPVIRKRYEDKFNYFAIRFLDEFAQFPFLRFQVHLGNYLHDSRPKHEL<br>ISDRRIKEKITVFGRLSELEHKKALFIKNTETNEDRKHYWEVFPNPNYDFP<br>KENISVNDKDFPIAGSILDREKQPTAGKIGIKVNLNQQYISEVDKAVKAH<br>QLKQRNNKPSIQNIIIEIVPINGSNPKEIIVFGGQPTAYLSMNDIHSILYEFF<br>DKWEKKKEKLEKKGEKELRKEIGKELEEKIVGKIQTQIQIIDKINAKILK<br>PYQDDDSTAIKELIKDLKQEQKILQKLKNEQTAREKEYQECIAYQEE<br>RKIKRSDKSRQKYLRLNQLKRYPEVPTREILYYQEKGVAVWLANDIK<br>RFMPPTDFKNEWKGEQHSLLQKSLAYYEQCKEELKNLLPQQKVFKHLPE<br>LGGHFQQKYLYQFYTRYLDRLEHISGLVQQAENFKNENKVFVKVENE<br>FKFLKKQNYTHKGLDAQSVLGYPIFLERGFMDKPTIIGKGTFKGNES<br>LFTDWFYRYKEYQNFQTFYDTENYPLVELEKKQADRKRETKIYQQKN<br>DVFTLLMAKHIFKSVFKQDSIDRFSLDLYQSREERLENQEKAKQTGERN<br>TNYIWNKTVDLNLCDGKVTVENVKLKNVGNFIKYEYDQRVQTFKLYEE<br>NIKWQAFLIKESKEEENYPIVEREIEQYEVKVRREELLKEVHLIEEYILEKV<br>KDKEILKKGDNQNFKYIILNGLLQKLNEDVESYKVFNLNTKPEDVNIN<br>QLKQEATDLEQKAFVLYIRNKFAHNQLPKKEFWDYCQEKYKIEKEKT<br>YAEYFAEVFKREKALMK |
| 普雷沃菌   | WP_042           | MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQNENNEN   |

[0586]

|                             |               |   |
|-----------------------------|---------------|---|
| P5-119 (SEQ ID No. 175)     | 518169        | LWFHPVMSHLYNAKNGYDKQPEKTMFIIERLQSYFPFLKIMAENQREYS<br>NGKYKQNRVEVNSNDIFEVLKRAFGVLKMYRDLTNHYKTYEELKIDGCE<br>FLTSTEQPLSGMISKYYTVALRNKERYGYKTEDLAFIQDNKIKTKDAY<br>GKRKSQVNTGFFLSLQDYNGDTQKKLHLSGVGIALLICLFLDKQYINIFLS<br>RLPIFSSYNAQSEERRIIIRSFGINSIKLPKDRIHSEKSNKSVAMDMLNEVKR<br>CPDELFTTSAEKQSRFRIISDDHNEVLMKRSTDRFVPLLLQYIDYGKLF<br>HIRFHVNMGKLR YLLKADKTCIDGQTRVRVIEQPLNGFGRLEEAEETMRK<br>QENGTFGNSGIRIRDFENVKRDDANPANYPYIVDTYTHYLENNKVEMFI<br>SDKGSSAPLLPLIEDDRYVVKTIPSCRMSTLEIPAMAFHMFVFGSKKTEKLI<br>VDVHNRYKRLFQAMQKEEVTAENIASFGIAESDLPQKILDLSIGNAHGKD<br>VDAFIRLTVDDMLTDTERRIKRFKDDRKSIRSADNKMGRGFKQISTGKL<br>ADFLAKDIVLFPQSVNDGENKITGLNYRIMQSAIAVYDSGDDYEAKQQF<br>KLMFEKARLIGKGTTEPHPLYKVFARSIPANAVDFYERYLIERKFYLTGL<br>CNEIKRGNRVDVPPFIRRDQNKWKTPAMKTLGRIYSEDLPVELPRQMFND<br>EIKSHLKS LQPMEGIDFNANVTYLIAEYMKRVLNDDFQTFYQWKRNYH<br>YMDMLKGEYDRKGS LQHCFTSVEEREGLWKERASRTERYRKLASNKIRS<br>NRQMRNASSEIEITILDKRLSNCRNEYQKSEKVIIRYRVQDALLFLAKK<br>TLTELADFDGERFKLKEIMPDAEKGILSEIMPMSFTFEKGGKKYTITSEGM<br>KLKNYGDFVFLASDKRIGNLLELVGSDIVSKEDIMEEFNKYDQCRPEISSI<br>VFNLEKWA FDTYPELSARVDREEKVD FKSILKILLNNKNINKEQSDILRKI<br>RNAFDHNNYPDKGIVEIKALPEIAMSIIKAFGEYAIMK  |
| 普雷沃菌 P4-76 (SEQ ID No. 176) | WP_044 072147 | MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQNENNEN<br>LWFHPVMSHLYNAKNGYDKQPEKTMFIIERLQSYFPFLKIMAENQREYS<br>NGKYKQNRVEVNSNDIFEVLKRAFGVLKMYRDLTNHYKTYEELKIDGCE<br>EFLTSTEQPLSGMINNYYTVALRNMNERYGKTEDLAFIQDKRFKFKVD<br>AYGKKKSQVNTGFFLSLQDYNGDTQKKLHLSGVGIALLICLFLDKQYINIF<br>LSRLPIFSSYNAQSEERRIIIRSFGINSIKQPKDRIHSEKSNKSVAMDMLNEI<br>KRCPNELFETLSAEKQSRFRIISNDHNEVLMKRSSDRFVPLLLQYIDYGKL<br>FDHIRFHVNMGKLR YLLKADKTCIDGQTRVRVIEQPLNGFGRLEEVETM<br>RKQENGTFGNSGIRIRDFENMKRDDANPANYPYIVDTYTHYLENNKVE<br>MFISDEETPAPLLPVIEDDRYVVKTIPSCRMSTLEIPAMAFHMFVFGSKKT<br>EKLIVDVHNRYKRLFQAMQKEEVTAENIASFGIAESDLPQKIIDLSIGNAH<br>GKD VDAFIRLTVDDMLADTERRIKRFKDDRKSIRSADNKMGRGFKQIST<br>GKLADFLAKDIVLFPQSVNDGENKITGLNYRIMQSAIAVYNSGDDYEAK<br>QQFKLMFEKARLIGKGTTEPHPLYKVFVRSIPANAVDFYERYLIERKFYL<br>IGLSNEIKKGNRVDVPPFIRRDQNKWKTPAMKTLGRIYDEDLPVELPRQMF<br>DNEIKSHLKS LQPMEGIDFNANVTYLIAEYMKRVLNDDFQTFYQWKRNY<br>YRYMDMLRGEYDRKGS LQSCFTSVEEREGLWKERASRTERYRKLASNKI<br>RSNRQMRNASSEIEITILDKRLSNRNEYQKSEKVIIRYRVQDALLFLA<br>KKTTELADFDGERFKLKEIMPDAEKGILSEIMPMSFTFEKGGKKYTITSE<br>GMKLLKNYGDFVFLASDKRIGNLLELVGSDTVSKEDIMEEFKKYDQCRPEI<br>SSIVNLEKWA FDTYPELSARVDREEKVD FKSILKILLNNKNINKEQSDILR<br>KIRNAFDHNNYPDKGVVEIRALPEIAMSIIKAFGEYAIMK |
| 普雷沃菌 P5-60 (SEQ ID No. 177) | WP_044 074780 | MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKVADIEGEQNENNEN<br>LWFHPVMSHLYNAKNGYDKQPEKTMFIIERLQSYFPFLKIMAENQREYS<br>NGKYKQNRVEVNSNDIFEVLKRAFGVLKMYRDLTNHYKTYEELKIDGCE<br>FLTSTEQPFSGMISKYYTVALRNKERYGYKAEDLAFIQDNRYKFTKDAY<br>GKRKSQVNTGSFSLQDYNGDTTKKLHLSGVGIALLICLFLDKQYINIFLS<br>RLPIFSSYNAQSEERRIIIRSFGINSIKQPKDRIHSEKSNKSVAMDMLNEVK<br>RCPDELFTTSAEKQSRFRIISDDHNEVLMKRSSDRFVPLLLQYIDYGKLF<br>DHIRFHVNMGKLR YLLKADKTCIDGQTRVRVIEQPLNGFGRLEEVETMR<br>KQENGTFGNSGIRIRDFENMKRDDANPANYPYIVETYTHYLENNKVEMFI<br>ISDEENPTPLPVIEDDRYVVKTIPSCRMSTLEIPAMAFHMFVFGSEKTEKL<br>IIDVHDRYKRLFQAMQKEEVTAENIASFGIAESDLPQKIMDLISIGNAHGKD<br>VDAFIRLTVDDMLTDTERRIKRFKDDRKSIRSADNKMGRGFKQISTGKL<br>ADFLAKDIVLFPQSVNDGENKITGLNYRIMQSAIAVYDSGDDYEAKQQF  |

[0587]

|  |                  |   |
|--|------------------|---|
|  |                  | KLMFEKARLIGKGTTEPHFPLYKVFVRSIPANA VDFYERYLIERKFYLIGLS<br>NEIKKGNRVDVPPFIRRDQNKWKTPAMKTLGRIYSEDLPVELPRQMFNDNEI<br>KSHLKSLPQMEGIDFNNANVTYLIAEYMKRVLNDDFQTFYQWKRNYRY<br>MDMLRGEYDRKGSQHCFTSIEEREGLWKERASRTERYRKLASNKIRSN<br>RQMRNASSEIEITILDKRLSNCRNEYQKSEKIIRRYRVQDALLFLAKKTL<br>TELADFDGERFKLKEIMPDAEKGILSEIMPMSFTFEKGGKIYTITSGGMKL<br>KNYGDFVFLASDKRIGNLLELVGSNTVSKEDIMEEFKKYDQCRPEISSIVF<br>NLEKWAFDTPPELPAVRDRKEKVDVFWSDVLSNNKDINNEQSYILRKIR<br>NAFDHNNYPDKGIVEIKALPEIAMSIIKAFGEYAIMK   |
| 厦门橙黄褐<br>指藻杆菌<br>( <i>Phaeodactyl<br/>ibacter<br/>xiamenensis</i> )<br>(SEQ ID No.<br>178) | WP_044<br>218239 | MTNTPKRRTLHRHPSYFGAFLNIARHNAFMIMEHLSTKYDMEDKNTLDE<br>AQLPNAKLFGLCKKRYGKPDVTEGVSRDLRRYFPFLNYPLFLHLEKQQN<br>AEQAATYDINPEDIEFTLKGFFRLLNQMRNNYSHYISNTDYGKFDKLPVQ<br>DIYEAAIFRLLDRGKHTKRFVDVFESKHTRHLESNNSEYRPSLANSPDHE<br>NTVAFVTCLFLERKYAFPFLSRLDCFRSTNDAAEGDPLIRKASHECYTMF<br>CCRLPQPKLESSDILLDMVNELGRCPSALYNLLSEEDQARFHIKREEITGF<br>EEDPDEELEQEIVLKRHSRFPYFALRYFDDTEAFQTLRFVYLGRWRTK<br>PVYKKRIYGQERDRVLTQSIRTFTRLSRLLPIYENVKHDAVRQNEEDGKL<br>VNPDVTSQFHKSQWQIESDDRAFLSDRIEHFSPHYNFGDQVIGLKFINDR<br>YAAIQNVFPKLPGEEKKDKDAKLVNETADAIISTHEIRSLFLYHYLSKKPIS<br>AGDERRFIQVDTETFIKQYIDTIKLFEDIKSGELQPIADPPNYQKNEPLPY<br>VRGDKEKTQERAQYRERQKEIKERRKELNLLQNRGLSIQYIPSRLE<br>YLLGYKKVPYEKLALQKLRAQRKEVKKRIKDIEKMRTPRVGEQATWLA<br>EDIVFLTPPKMHTPERKTTKHPQKLNNDQFRIMQSSLAYFSVNKKAIKKFF<br>QKETGIGLSNRETSHPFYRIDVGRGILDFYTGYLKYKMDWLDLDAIKK<br>VDNRKHGKKEAKKYEKYLPSSIQHKTPELDYTRLPVYLPRGLFKKAIK<br>ALAAHADVFQVEEEDNVIFCLDQLLDGDTQDFYNWQRYRSALTEKET<br>DNQLVLAHPYAEQILGTIKTLEGKQKNNKLGKAKQKIKDELIDLKRAK<br>RLLDREQYLRAVQAEDRALWLMIQERQKQKAEHEEIAFDQLDLKNITKIL<br>TESIDARLRIPDTKVDITDKLPLRRYGDLRRVAKDRRLVNLASYHYHAGL<br>SEIPYDLVKKLEEYDRRRVAFFEHVYQFEKEVYDRYAAELRNENPKGE<br>STYFSHWEYVAVAVKHSADTHFNELFKEKVMQLRNKFHHNEFPYFDWL<br>LPEVEKASAALYADRVDVAEGYYQKMRKLMRQ                                    |
| 黄杆菌<br>( <i>Flavobacteri<br/>um sp.</i> )316<br>(SEQ ID No.<br>179)                        | WP_045<br>968377 | MDNNITVEKTELGLGITYNHDKVEDKHYFGGFNLAQNNIDLVAQEFKK<br>RLLIQGKDSINIFANYFSDQCSITNLERGIKILAEYFPVVSIDLDEKNKSKS<br>IREHLILLLETINLRNYTHYHKKIIDGSLFPLDITILLKVVLEIKKKKL<br>KEDKTKQLLKKGLEKEMTILFNLMAEQKEKKIKGWNIDENIKGAVLNR<br>AFSHLLYNDELSDYRKSXYNTEDETLKDTL TESGILFLLSFFLNKKEQEQ<br>KANIKGYKGIASIPDEEITLKNNSLRNMATHWYSHLTYKGLKHRIKTD<br>HEKETLLVNMVDYLSKVPHEIYQNLSEQNKSLFLEDINEYMRDNEENHD<br>SSEASRVIHVPVIRKRYENKFA YFAIRFLDEFAEFPPLRFMVNVGNYIHDNR<br>KKDIGGTSITNRTIKQQINVFGNL TEIHKKNNDYFEKEENKEKTLEWELF<br>PNPSYHFQKENIPIDLEKSKE TNDLAKEYAKEKKKIFGSSRKKQONTAK<br>KNRETIINLVFDKYKTSRKT VTFEQPTALLSFNELNSFLY AFLVENKTKG<br>ELEKIIIEKIANQYQILKNCSSSTVDK TNDNIPKSIKKIVNTTDSYFEGKKI<br>DIEKLEKDITIEIEK TNEKLETIKENEESAQNYKRNERNTQKRKLYRKYVF<br>FTNEIGIEATWITNDILRFLDNKENWKGYQHSELQKFISQYDNYKKEALG<br>LLESEWNLESDAFFGQNLKRMFQSNSTFETFYKKYLDNRKNTLETYLSAI<br>ENLKTMTDVRPKVLKKKWT ELFRFFDKKIYLLSTIETKINELITKPINLSRG<br>IFEEKPTFINGKNPNKENNQHLFANWFIYAKKQ TILQDFYNLPLEQPKAIT<br>NLKKHKYKLEERSINNLKIEDIYIKQMVDFLYQKLFESFISLQDLYTSKE<br>KREIEKGKAKNEQTPDESFIWKKQVEINTHNGRIIAKT KIKDIGKFNLLT<br>DNKIAHLISYDDRIWDFSLNNDGDITK KLYSINTELESYETIRREKLLKQIQ<br>QFEQFLLEQETEYSAERKHPEKFEKDCNPNFKKYIIEGVLNKIIPNHEIEIE<br>ILSKEDVFKINFSDILLNNDNIKKGYLLIMIRNKFAHNQLIDKNLNFNLSLQ<br>LYSKNENENFSEYLNKVCQNIIEQEFKELK |
| 古叶琳单胞  | WP_046           | MTEQSERPYNGTYTTLEDKHFWA AFLNLARHNAYITLTHDRQLAYSKA   |

[0588]

|                                |                                    |   |
|--------------------------------|------------------------------------|---|
| 菌 (SEQ ID No. 180)             | 201018                             | DITNDQDVLSFKALWKNFDNDLERKSRLRSLILKHFSFLEGAAYGKKLFE SKSSGNKSSKNKELTKKEKEELQANALSLDNLKSILDFDLQKLDKFRNYY SHYRHSESELPLFDGNMLQRLYNVFDVSVQRVVRDHEHNDKVDPRRH FNHLVRKGGKDRYGHNDNPSFKHHFVDESEGMVTEAGLLFFVSLFLEKRD AIWMQKKIRGFKGGTETYQMTNEVFCRSRISLPKLLKLESLRTDDWMLL DMLNELVRCPKPLYDRLREKDRARFRVPVDILPDEDTDGDDGDPFKNT LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKMIGEQPEDR HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDFETGDKPYISQTTPHYH IEKGKIGLRFMPEGQHLWPSPEVGTTRTGRSKYAQDKRLTAEAFSLVHEL MPMFFYFLLREKYSEEVSAAEKVQGRIKRVIEDVYAIYDAFARDEINTLK ELDACLADKGIIRRHLPKQMIALSQEHKDMEEKIRKQLQEMIAADTDHRL DMLDRQTDKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGBK LNNKANSTEYRMLQRALALFGGEKKRLTPYFRQMNLTGGNNPHPLHE TRWESHNTLSFYRSYLARKAFLEIRIGRSDRMENRPFLLKKEPKTDRQTL VAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYREVGFMAKAVPLYFE RACEDRVQPFYDSPFNVGNSLKPCKGRFLSKEERAEWERGKERFRDLE AWSSAARRIEDAFAGIEYASPGNKKIEQLLRDLSLWEAFESKLVKVRAD KINLAKLKEILEAQEHPYHDFKSWQKFERELRLVKNQDITWMMCRDL MEENKVEGLDTGTLYLKDIRPNVQEQGSLNVLNRVKPMRLPVVYRAD SRGHVHKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTGG LAMEQYPISKLRVEYELAKYQTARVCVFELTLRLEESLLTRYPHLPDESFR KMLESWSDPLLAKWPELHGKVRLLIARNAFSHNQYPMYDEAVFSSIRK YDPSSPDAIEERMGLNIAHRLSEEVKQAKETVERIIQV          |
| WP_047431 796 (SEQ ID No. 181) | 金黄杆菌 (Chryseo bacterium sp.)YR4 77 | METQTIGHGIAYDHSKIQDKHFFGGFLNLAENNIKAVLKAFSEKFNVGNV DVKQFADVSLKDNLPDNDQKRVSLKMYFPVVDINIPNNRAKFRSDLT TLFKSVDQLRNFYTHYYHKPLDFDASLFILLDDIFARTAKEVRDQKMKDD KTRQLLSKSLSEELQKGYELQLERLKLNLGKKVNIHDQLGKNGVLNN AFNHLIYKDGESFKTKLTYSSALTSFESAENGIEISQSGLLFLLSMFLKRKEI EDLKNRNKGFKAKVVIDEDGKVNGLKFMATHWVFSYLCFKGLKSKLST EFHEETLLIQIIDELSKVPDELYCAFDKETRDKFIEDINEYVKEGHQDFSLE DAKVIHPVIRKRYENKFNYFAIRFLDEFVKFPSLRFQVHVGNVYVHRRIK NIDGTTFETERVVKDRIKVFGRLEISSYKAQYLSSVSDKHDETGWEIFPN PSYVFINNPIHISVDTSFKKEIADFKLLRAQVPDELKIRGAEKKRKFEIT QMIGSKSVLNQEEPIALLSLNEIPALLYEILINGKEPAEIERIHKDLNERQD VIKNYNPENWLPASQISRRLRSNKGERIINTDKLLQLVTKELLVTEQKLIKII SDNREALKQKKEGKYIRKFIFTNSELGREAIWLADDIKRFMPADVRKEWK GYQHSQQLQSLAFYNSRPKEALAILESSWNLKDEKIIWNEWILKSFTQNK FFDAFYNEYLKGRKKYFAFLSEHIVQYTSNAKNLQKFIKQOMPDLFEKR HYIIEDLQTEKNKILSKPFIFPRGIFDKKPTFIKGVKVEDSPESFANWYQYG YQKDHQFQKFYDWKRDYSDVFLEHLGKPFINNGDRRTLGMEELEKRIIK QDLKIKKIKIQDLFLRLIAENLFQKVFYSAKLPLSDFYLTQEERMEKENM AALQNVREEGDKSPNIKDNFIWSKMIPYKKGQIENAVKLDIGKLVLS LDDKVQTLTSLYDDAKPWSKIALENEFSIGENSYEVIRREKLFKEIQFSEI LFRSGWDGINHPAQLEDNRNPKFKMYIVNGILRKSAGLYSQGEDIWFY NADFNLDADVLETSELVQLAFLVTAIRNKFAHNQLPAKEFYFYIRAKY GFADEPSVALVYLNFTKYAINEFKKVM |
| 鸭疫里默氏杆菌 (SEQ ID No. 182)       | WP_049 354263                      | MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYKAHFTVNLTRDFAHLNRK GKNKQDNPDFNRYRFEKDGFFTESGLLFFTNLFLDKRDAYWMLKKVSGF KASHKQREKMTTEVFCRSRILLPKLRLESRYDHNQMLDMLSELSRCPKL LYEKLSEENKHHFQVEADGFLDEIEEEQNPFKDTLIRHQDRFPYFALRYLD LNESFKSIRFQVDLGTYHYCIYDKKIGDEQEKRLTRTLLSFGRLQDFTEI NRPQEWKALTKDLDYKETSNOFPISKTTPHYHITDNKIGFRLGTSKELYPS LEIKDGANRIAKYPYNSGFVAHAFISVHELLPLMFYQHLTGKSEDLKETS VRHIQRIYKDFEERINTIEDLEKANQGRPLGAFPKQMLGLLQNKQPDLS EKAKIKIEKLIATKLLSHRLNLTKLKSSPKLGRREKLIKTVLADWLVKD FMRFPQVAYDAQNQPIKSSKANSTEFWFIRRALALYGGKEKRLLEGYFKQ  |

[0589]

|                          |                  |   |
|--------------------------|------------------|---|
|                          |                  | TNLIGNTNPHFPLNKFNWKACRNLDVDFYQQYLEQREKFLEAIKNQPWEP<br>YQYCLLLKIPKENRKNLVKGWEEQGGISLPRGLFTEAIRETLSEDLMLSKPI<br>RKEIKKHGRVGFISRITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKREE<br>HYEYWQQNKPPQSPTESQRLELHTSDRWKDYLLYKRWQHLEKLRLYRN<br>QDVMLWMLTLELTKNHFKELNLYHQLKLENLAVNVQEAADAKLNPLN<br>QTLPMVLPVKVYPATAFGEVQYHKTPIRTVYIREEHTKALKMGNFKALV<br>KDRRLNGLFSFIKEENDTQKHPISQLRLRRELEIYQSLRVDFAFKETLSLEEK<br>LLNKHTSLSSLENEFRALLEEWKKEYAASSMVTDEHIAFIASVRNAFCHN<br>QYPFYKEALHAPIPLFTAQPTTEEKDGLGIAEALLKVLREYCEIVKSQI   |
| 牙龈卟啉单胞菌 (SEQ ID No. 183) | WP_052<br>912312 | MTEQNEKPYNGTYTLEDKHFWAFFNLARHNAITLAHIDRQLAYS<br>ADITNDEDILFFKGQWKNLDNDLERKARLRLSLKHFSLFLEGAAYGKKLF<br>ESQSSGNKSSKKKELTKKEKEELQANALSLDNLKSILFDFLQKLDKFRNY<br>YSHYRHPESSELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNDKVDPHR<br>HFNHLVRKGGKDKYGNNNDNPFKHHFVDREKVTAEAGLLFFVSLFLEKR<br>DAIWMQKKIRGFKGGTEAYQMTNEVFCRSRISLPLKLESRLTDDWML<br>LDMLNELVRCPKLLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNT<br>LVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPEDRH<br>LTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDFYFETGDKPYITQTPHYHI<br>EKGKIGLRFVPEGQLLWPSPEVGATRTGRSKYAQDKRFTAFAFLSVHEL<br>MPMMFYFLLREKYSEEASAQVQGRKRVIEDVYAVYDAFARDEINTR<br>DELDACLADKGIIRGHLPRQMIALSQEHKDMEEKVRKQLQEMADTDH<br>RLDMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSG<br>KPLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPL<br>HETRWESHNTILSFYRSYLKARKAFLQSIGRSDREENHRFLLKPKTDRQ<br>TLVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEVGFMAKAVPLY<br>FERACKDRVQPFYDYPFNVGNSLKPCKGRFLSKEKRAEWEWESGKERFRD<br>LEAWSHSAARRIEDAFVGIYASWENKKKIEQLLQDLSLWETFESKLKVK<br>ADKINIAKLLKEILEAKEHPYHDFKSWQKFERELRLVKNQDIITWMMCRD<br>LMEENKVEGLDTGTLYLKDIRTDVQEQGSLNVLNHNKPMRLPVVYRA<br>DSRGHVHKEEAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVDTG<br>ALAMEQYPISKLRVEYELAKYQTARVCAFEQTLLEESLLTRYPHLPDES<br>FREMLESWSDPLLDKWPDLQREVRLIAVRNAFASHNQYPMYDETFISSIR<br>KYDPSSSLDAIEERMGLNIAHRLSEEVLAKEMVERIIQA |
| 牙龈卟啉单胞菌 (SEQ ID No. 184) | WP_058<br>019250 | MTEQNEKPYNGTYTLKDKHFWAFFNLARHNAITLTHIDRQLAYS<br>ADITNDEDILFFKGQWKNLDNDLERKARLRLSLKHFSLFLEGAAYGKKLF<br>ESQSSGNKSSKKKELTKKEKEELQANALSLDNLKSILFDFLQKLDKFRNY<br>YSHYRHPESSELPMFDGNMLQRLYNVFDVSVQRVKRDHEHNDKVDPHR<br>HFNHLVRKGGKDKRCGNNDNPFKHHFVDREGKVTAEAGLLFFVSLFLEKR<br>DAIWMQKKIRGFKGGTETYQMTNEVFCRSRISLPLKLESRLTDDWML<br>LDMLNELVRCPKSLYDRLREEDRACFRVPVDILSDEDDTDGAEEDPFKN<br>TLVRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPEDR<br>HLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDFCFETGDKPYITQTPHYH<br>IEKGKIGLRFVPEGQHLWPSPEVGATRTGRSKYAQDKRFTAFAFLSVHEL<br>MPMMFYFLLREKYSEEVSARVQGRKRVIEDVYAVYDAFARDEINTR<br>DELDACLADKGIIRGHLPRQMIALSQKHKDMEEKVRKQLQEMADTDH<br>RLDMLDRQTDRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSG<br>KPLNNSKANSTEYRMLQRALALFGGEKERLTPYFRQMNLTGGNNPHPL<br>HETRWESHNTILSFYRSYLKARKAFLQSIGRSDRVENHRFLLKPKTDR<br>QTLVAGWKGEFHLPRGIFTEAVRDCLIEMLGLDEVGSYKEVGFMAKAVPL<br>YFERACKDRVQPFYDYPFNVGNSLKPCKGRFLSKEKRAEWEWESGKERFR<br>DLEAWSHSAARRIEDAFAGIENASRENKKKIEQLLQDLSLWETFESKLKVK<br>KADKINIAKLLKEILEAKEHPYLDKSWQKFERELRLVKNQDIITWMMCR<br>DLMEENKVEGLDTGTLYLKDIRTDVQEQGSLNVLNHNKPMRLPVVYR<br>ADSRGHVHKEQAPLATVYIEERDTKLLKQGNFKSFVKDRRLNGLFSFVD<br>TGALAMEQYPISKLRVEYELAKYQTARVCAFEQTLLEESLLTRYPHLPD<br>ENFRKMLESWSDPLLDKWPDLHRKVRLIAVRNAFASHNQYPMYDEAVF   |

[0590]

|                          |               |   |
|--------------------------|---------------|---|
|                          |               | SSIRKYDPSSPDAIEERMGLNIAHRLSEEVKQAKEMAERIIQA   |
| 柱状黄杆菌 (SEQ ID No. 185)   | WP_060 381855 | MSSKNESYNKQKTFNHYKQEDKYFFGGFLNADDNLRQVGKEFKTRINF<br>NHNNELASVFKDYFNKEKSVAKREHALNLLSNYPVLERIQKHTNHNF<br>EQTREIFELLDDTIKKLRDYTHHYHKPITINPKVYDFLDDTLLDVLITIKK<br>KKVKNDTSRELLKEKFRPELTQLKNQKREELIKGKLLLEENLENAVFNH<br>CLRPFLEENKTDKQNKTVSLRKYRKSKEPNEETSITLQSGLVFLISFLLHR<br>KEFQVFTSGLEGFKAKVNTIKEEEISLNKNNIVYMITHWSYSYNNFKGLK<br>HRIKTDQGVSTLEQNNTTHSLTNTNTKEALLTQIVDYLSKVPNEIYETLSE<br>KQKQEFEEEDINEYMRENPENEDSTFSSIVSHKVIKRYENKFNYPAMRFL<br>DEYAEPLTRFMVNFVDYIKDRQKILESIQFDSEIIEKKEIHLFEKLGVLV<br>EYKKNVYLKETSNDLSRFPLFPSPSYVMANNIPFYIDSRNNLDEYLNQ<br>KKKAQSQNRKRNLTFEKNKEQSKDAIAMLQKEIGVKDLQQRSTIGLLS<br>CNELPSMLYEIVKDIKGAELNKIAQKIREQYQSIRDFTLDSQKDNIPPT<br>LTKTISTDTSVTFENQPIDIPRLKNAQKELTLTQEKLLNVKQHEIEVDNY<br>NRNKNTYKFKNQPKDKVDDNKLQRKYVFYRNEIGQEANWLASDLIHF<br>KNKSLWGYMHNELQSFLAFFEDKKNDCIALLETVFNLKEDCILTDLK<br>NLFLKHGNFIDFYKEYLKLKEDFLNTESTFLENGFIGLPPKILKELSKRLN<br>YIFIVFQKRQFIKELEEKNNLYADAINLSRGIFDEKPTMIPFKPNPDEFA<br>SWFVASYQYNNYQSFYELTPDKIENDKKKKYKNLRAINKVQIDYYLKL<br>MVDTLYQDLFNQPLDKSLSDFYVSKTDREKIKADAKAYQKRNDLWLN<br>KVIHLSLQNNRITANPKLDIGKYKRALQDEKIATLLTYDDRTWTYALQK<br>PEKENENDYKELHYTALNMELQEYKVRSKLLKQVQELEKQILDKFDYD<br>FSNNATHPEDLEIEDKKGKRHPNFKLYITKALLKNESEIINLENIDIEILKY<br>YDYNTEKLKEKIKNMEDEKAKIVNTKENYNKNTVLIKKALVLIIRNK<br>MAHNQYPPKFIYDLATRFVPKKEEYFACYFNRFVETITTELWENKKA<br>KEIV |
| 牙龈卟啉单胞菌 (SEQ ID No. 186) | WP_061 156470 | MTEQNERPYNGTYTLEDKHFWAFFNLARHNAYITLTHIDRQLAYSKA<br>DITNDEDILFFKGQWKNLDNDLERKARLRSLILKHFSLGAAAYGKKLFE<br>NKSSGNKSSKKKELTKKEKEELQANALSLDNLKSLFDLQKLKDFRNY<br>SHYRHPESSELPLFDGNMLQRLYNVFDVSVQRVKRDHEHNDKVDPHRH<br>FNHLVRKGGKDRCGNNDNPFKHHFVDREGKVTEAGLLFFVSLFLEKRD<br>AIWMQKKIRGFKGGTEAYQMTNEVFCRSRISLPKLLKLESLRTDDWMLL<br>DMLNELVRCPSLYDRLREEDRARFRVPVDILSDEDDTDGTEEDPFKNTL<br>VRHQDRFPYFALRYFDLKKVFTSLRFHIDLGTYHFAIYKKNIGEQPEDRHL<br>TRNLYGFGRIQDFAEEHRPEEWKRLVRDLDFETGDKPYITQTPHYHIE<br>KKGIGLRFVPEGQHLWPSPEVGATRTRGRSKYAQDKRLTAEAFSLVHELM<br>PMMFYYFLLREKYSEEVSAAEKVQGRIKRVIEDVYAVYDAFARGEIDTLD<br>RLDACLADKGIIRGHLPRQMIAILSQEHKDMEEKVRKLLQEMIADTDHR<br>LDMLDRQTRKIRIGRKNAGLPKSGVIADWLVRDMMRFQPVAKDTSGK<br>PLNNSKANSTEYRMLQRALALFGGKERLTPYFRQMNLTGNNPHPLFH<br>ETRWEHTNILSFYRSYLKARKAFLQSIGRSDREENHRFLLKPKTDRQT<br>LVAGWKSEFHLPRGIFTEAVRDCLIEMGYDEVGSYKEVGFMAKAVPLYF<br>ERACKDRVQPFYDYPFNVGNSLKPCKGRFLSKEKRAEWEWESGKERFRLA<br>KLKKEILEAKEHPYLDKFSWQKFERELRLVKNQDIITWMMCRDLMEENK<br>VEGLDTGTLYLKDIREVQEQGSLNVLNRVKPMRLPVVYRADSRGHV<br>HKEQAPLATVYIEERDTKLLKQGNFKSVKDRRLNGLFSFVDTGGLAME<br>QYPISKLRVEYELAKYQTARVCAFEQTLELEESLLTRCPHLPDKNFRKML<br>ESWSDPLLDKWPDLQREVWLLIAVRNAFVSHNQYPMYDEAVFSSIRKYDP<br>SSPDAIEERMGLNIAHRLSEEVKQAKEMAERIIQA  |
| 牙龈卟啉单胞菌 (SEQ ID No. 187) | WP_061 156637 | MNTVPASENKGQSRVTEDDPQYFGLYLNLARENLIEVESHVRIKFGKKL<br>NEESLKQSLLCDHLLSVDRWTKVYGHSSRYLPFLHYFDPDSQIEKDHSK<br>TGVDPDSAQRLIRELYSLDFLRNDFSHNRLDGTTFEHLVSPDISSFITGT<br>YSLACGRAQSRFADFPPDDFVLAKNRKEQLISVADGKECLTVSGLAFFI<br>CLFLDREQASGMLSRIKRGKRTDENWARAVHETFCDLICIRHPDRLESSN<br>TKEALLDMLNELNRCPRILYDMLPEEERAQFLPALDENSMNNLENSLN<br>EESRLLWDGSSDWAEALTKRIRHQDRFPYMLRFLIEEMDLLKGRFRVDL  |

[0591]

|  |                          |  |
|--|--------------------------|--|
|  |                          | <p>GEIELDSYSKKVGRNGEYDRITTDHALAFGKLSDFQNEEEVSRMISGEAS<br/> YPVRFSLFAPRYAIYDNKIGYCHTSDPVYPKSKTGEKRALSNPQSMGFISV<br/> HDLRKLLLMELLCEGSFSRMQSGFLRKANRILDETAEGKLFQSFALPEMR<br/> HRFIPPQNPKSKDRREKAETTLEKYKQEIKGRKDKLNSQLLSAFDMNQRR<br/> LPSRLLDEWMNIRPASHSVKLRTYVKQLNEDCRLRLRFRKDGDKARA<br/> IPLVGEMATFLSQDIVRMIISEETKKLITSAYYNEMQSRSLAQYAGEENRRQ<br/> FRAIV AELHLLDPSSGHPFLSATMETAHRYTEDFYKCYLEKKREWLAKTF<br/> YRPEQDENTKRISVFFVDPDGEARKLLPLLIRRRMKEQNDLQDWIRNKQA<br/> HPIDLPSHLFDSKIMELLKVKDGGKKWNEAFKDWSTKYPDGMQPFYF<br/> LRRELNHIGKSVSYIPSDGKKFADCYTHLMEKTVQDKKRELRTAGKPVPP<br/> DLAADIKRSFHRAVNEREFMLRLVQEDDRLMLMAINMMTDREEDILPG<br/> LKNIDSILDKENQFSLAVHAKVLEKEGEGGDNLSLVPATIEIKSKRKDWS<br/> KYIRYRYDRRVPGMLSHFPEHKATLDEVKTLGLEYDRCRIKIFDWAFAL<br/> GAIMSDRDLKPYLHESSSREGKSGEHSTLVKMLVEKKGCLTPDESQYLILI<br/> RNKAAHNQFPCAEMPLIYRDVSAKVGSIEGSSAKDLPEGSSLVDSLWK<br/> KYEMIIRKILPILDPENRFFGKLLNMSQPINDL</p>   |
| <p>鸭疫里默氏<br/>杆菌 (SEQ<br/>ID No. 188)</p> | <p>WP_061<br/>710138</p> | <p>MFFSFHNAQRVIFKHLYKAFDASLRMVKEDYKAHFTVNLTRDFAHLNRK<br/> GKNKQDNPDFNRYRFEKDGFFTESGLLFFTNLFLDKRDAYWMLKKVSGF<br/> KASHKQSEKMTTEVFCRSRILLPKLRLESRYDHNQMLLDMLSELRCPKL<br/> LYEKLSEKDKKCFQVEADGFLDEIEEEQNPFKDTLIRHQDRFPYFALRYL<br/> DLNESFKSIRFQVDLGTYYHYCIYDKKIGYEQEKRHTRTLLNFGRLQDFTE<br/> INRPQEWKALTKDLDYNETSNQPFISKTTPHYHITDNKIGFRLRTSKELYP<br/> SLEVKDGANRIAKYPYNSDFVAHAFISISVHELLPLMPFYQHLTGKSEDLK<br/> ETVRHIQRIYKDFEERINTIEDLEKANQGRLPLGAFPKQMLGLLQNKQPD<br/> LSEKAKIKIEKLI AETKLLSHRLNTKLKSSPKLGKRREKLIKTVLADWL<br/> KDFMRFPVYVYDAQNQPIKSSKANSTESRLIRRALALYGGEKNRLEGYFK<br/> QTNLIGNTNPHFPLNKFNWKACRNLVDFYQYQLEQREKFLEAIKHQPWE<br/> PYQYCLLLKVPKENRKNLVKGWEGGSLPRGLFTEAIRETLTKDLTLK<br/> PIRKEIKKHGRVGFISRAITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKKE<br/> EHYEWQQNKPSPTESQRLELHTSDRWKDYLLYKRWQHLEKKLRLRYR<br/> NQDIMLWMLTLELTKNHFKELNLYHQLKENLAVNVQEADAKLNPLN<br/> QTLPMVLPVKVYPTAFGEVQYHETPIRTVYIREEQTKALKMGNFKALV<br/> KDRHLNGLFSFIKEENDTQKHPISQLRLRRELEIYQSLRVDAFKETLSLEEK<br/> LLNKHASLSLENEFRTLLEEWKCKYAASSMVTDKHIAFIASVRNAFCHN<br/> QYPFYKETLHAPILLFTVAQPTTEEKDGLGIAEALLRVLREYCEIVKSQI</p>   |
| <p>柱状黄杆菌<br/>(SEQ ID No.<br/>189)</p>    | <p>WP_063<br/>744070</p> | <p>MSSKNESYNKQKTFNHYKQEDKYFFGGFLNADDNLRQVGKEFKTRINF<br/> NHNNELASVFKDYFNKEKSVAKREHALNLLSNYPVLERIQKHTNHNH<br/> EQTREIFELLDTIKKLRDYTHHYHKPITINPKIYDFLDDTLDDVLITIKK<br/> KVKNDSRELLKEKLRPELTQLKNQKREELIKKGGKLLLENLENVFNHC<br/> LRPFLEENKTDDKQNKTVSLRKYRKSPPNEETSITLTQSGLVFLMSFFLHR<br/> KEFQVFTSGLEGFKAKVNTIKEEKISLNKNNIVYMITHWSYSYNNFKGLK<br/> HRIKTDQGVSTLEQNNTTHSLTNTNTKEALLTQIVDYLSKVPNEIYETLSE<br/> KQKQEFEEEDINEYMRENPENEDSTFSSIVSHKVIKRYENKFNYPAMRFL<br/> DEY AELPTLRFMVNFGDYIKDRQKKILESIFDSEIHKKEIHLFEKLGVLV<br/> EYKKNVYLKETSNIIDLSRFPPLFPSYVMANNIPFYIDSRSNLDEYLNQ<br/> KKAQSQNRKRNLTFEKYNKEQSKDAIAMLQKEIGVKDLQQRSTIGLLS<br/> CNELPSMLYEVIVKDIKGAELNKIAQKIREQYQSIRDFTLNSPQKDNIPPT<br/> LIKTISTDTSVTFENQPIDIPRLKNAIQKELALTOEKLLNVKQHEIEVNNYN<br/> RNKNTYKFKNQPKDKVDDNKLQRKYVFYRNEIGQEANWLASDLIHFMK<br/> NKSLWKGVMHNEQSF LAFFEDKKNDCIALLETVFNLKEDCILTDLKLN<br/> FLKHGNFIDFYKEYLKLKEDFLNTESTFLENGFIGLPPKILKKELSKRLNYIF<br/> IVFQKRQFIIKELEEKNNLYADAINLSRGIFDEKPTMIPFKKPNPDEFASW<br/> FVASYQYNNYQSFYELTPDKIENDKCKKYKNLRAINVKIQDYLLKLMV<br/> DTLYQDLFNQPLDKSLSDFYVSKTDREKIKADAKAYQKRND SFLWNKVI<br/> HLSLQNNRITANPKLKDIGKYKRALQDEKIATLLTYDDRTWTYALQKPE<br/> KENENDYKELHYTALNMELQEYEVRSKLLKQVQEQLEKQILDKIFYDFS</p> |

[0592]

|   |                        |  |
|---|------------------------|--|
|   |                        | NNATHPEDLEIEDKKGRHPNFKLYITKALLKNESEIINLENIDIEILIKYYD<br>YNTEKLKEKIKNMDEDEKAKIVNTKENYNKITNVLIKALVLIIRNKMA<br>HNQYPPKFIYDLATRFVPKKEEYFACYFNRFVETITTELWENKKKAKEI<br>V   |
| 鸭疫里默氏<br>杆菌 (SEQ<br>ID No. 190)   | WP_064<br>970887       | MEKPLPPNVYTLKHKFFWGAFLNIARHNAFITICHINEQLGLTTPNDDKI<br>ADVVCGTWNNILNNDHDLKKSQTELILKHFPLAAMCYHPPKKEGKK<br>KGSQKEQQKEKENEAAQSQAALNPSELIKVLKTLVKQLRTRLNYYSHHS<br>HKKPDAEKDIFKHLYKAFDASLRMVKEDYKAHFTVNLTDFAHLNRKG<br>KNKQDNPDFDRYRFEKDGFFTESGLLFFTNLFLDKRDAYWMLKKVSGFK<br>ASHKQSEKMTTEVFCRSRILLPKLRLESRYDHNQMLLDMLSELSRYPKLL<br>YEKLSEEDKKRFQVEADGFLDEIEEQNPFKDTLIRHQDRFPYFALRYLDL<br>NESFKSIRFQVDLGTYHYCIYDKKIGDEQEKRHLRTRLLSFGRLQDFTEIN<br>RPQEWKALTKDLDYKETSQPFISKTTPHYHITDNKIGFRLGTSKELYPSL<br>EVKDGANRIAQYPYNSDFVAHAFISVHELLPLMFYQHLTGKSEDLLKETV<br>RHIQRIYKDFEERINTIEDLEKANQGRPLGAFPKQMLGLLQNKQPDSE<br>KAKIKIEKLIATKLLSHRLNTKLSKSPKLGKRREKLIKTVLADWLKDF<br>MRFQPVAYDAQNQPISSEKANSTEFQLIQRALALYGGKRNREGYFKQT<br>NLIGNTNPHPLNKFVWKAACRNLVDFYQQYLEQREKFLAIAKNQPWEPY<br>QYCLLLKIPKENRKNLVKQWEQGGISLPRGLFTEAIRETSLKDLTSLKPIR<br>KEIKKHGRVGFISRAITLYFREKYQDDHQSFYDLPYKLEAKASPLPKKEH<br>YEYWQNKQPSTELQRLLEHTSDRWKDYLLYKRWQHLEKLRLYRNQ<br>DVMLWMLTLELTKNHFKELNLYHQLKLENLAVNVQEAADAKLNPLNQT<br>LPMVLPVKVYPATAFGEVQYQETPIRTVYIREEQTKALKMGNFALVKD<br>RRLNGLFSFIKEENDTQKHPISQLRRLRRELEIYQSLRVDFAFKETLNLEEKLL<br>KKHTSLSSVENKFRILLEEWKKEYAASSMVTDEHIAFIASVRNAFCHNQY<br>PFYEEALHAPIPLFTVAQQTTEEKDGLGIAEALLRVLREYCEIVKSQI  |
| 海洋解胶质<br>中华微菌<br>( <i>Sinomicrobi<br/>um<br/>oceanii</i> )(SEQ<br>ID No. 191)                 | WP_072<br>319476.<br>1 | MESTTTLGLHLKYQHDLFEDKHYFGGGVNLAVQNIESIFQAFERYGIQN<br>PLRKNQVPAINNIFHDNISISNYKEYLKFLKQYLPVVGFLKESNEINIFEFRE<br>DFEILINAIYKLRHFYTHYYHSPIKLEDRFYTCNLNELFVAVAIQVKKHKMK<br>SDKTRQLLNKNLHQLLQQLIEQKREKLDKKAEGEKVSLDTKSIENAVLN<br>DAFVHLLDKDENIRLNYSRLSEDIITKNGITLSISGLLFLSLFLQRKEAED<br>LRSRIEGFKGKGNELRFMATHWVFSYLNVKRIKHRLNTDFQKETLLIQIA<br>DELSKVPDEVYKTLDHENRSKFLEDINEYIREGNEDASLNESTVVHGVR<br>KRYENKFHYLVLRYLDEFVDFPFLRFQVHLGNYIHDRRDKVIDGTNFITN<br>RVIKEPIKVFGLSHVSKLSDYMESSLREHKNGWDFPNPSYNFVGHNI<br>PIFINLRSASSKGKELYRDLMIKSEKSKSREEGIPMERRDGKPTKIEISN<br>QIDRNIKDNFNDIYPGEPLAMLSLNEPALLFELLRRPSITPDIEDRMVE<br>KLYERFQIIRDYKPGDGLSTSKISKLRKADNSTRLDGKLLRAIQTETRN<br>AREKLHTLEENKALQKNRKRRTVYTTREQGREASWLAQDLKRFMPIASR<br>KEWRGYHHSQQLAFYDQNPQPLELLEQFWDLKEDTYVWNSWIHK<br>SLSQHNGFVPMYEGYLGRLGYKLESIIIGFLEEHLVLRKRYTQQHL<br>NVIFRERLYFIKTETKQKLELLARPLVFRGIFDDKPTFVQDKKVVDPHEL<br>FADWYVYSYKDDHSFQEFYHYKRDYNEIFETELSWDIDFKDNKRQLNPS<br>EQMDLFRMKWDLKIKKIKIQDIFLKIVAEDIYKIFGHKIPLSLSDFYISRQE<br>RLTLDEQAVAQSMRLPGDTSNQIKESNLWQTTVPYEKEQIREPKIKLKD<br>IGKFKYFLQQKVLNLLKYDPQHVWTKAELEEELYIGKHSYEVVREML<br>LQKCHQLEKHILEQFRFDGNSNHPRELEQGNHPNFKMYIVNGILTKRGELEI<br>EAENWWLELGNKSNLSDKVEVELLTMKTIPEQKAFLLILIRNKFAHNQLP<br>ADNYFHYASNLMLNKKSDTYSLFWFTVADTIVQEFMSL |
| 琼脂变色赖<br>兴巴赫氏菌<br>( <i>Reichenbac<br/>hiella<br/>agariperforan<br/>s</i> )(SEQ ID<br>No. 192) | WP_073<br>124441.<br>1 | MKTNPLIASSGEKPNYKKNFNTESDKSFKKIFQNKGSIAPIAEKACKNFEIKS<br>KSPVNRDGRHLHYFSVGHAFKNIDSKNVFRYELDESQMDMKPTQFLALQK<br>EFFDFQGALNGLLKHIRNVNSHYVHTFEKLEIQSINQKLITFLIEAFELAVIH<br>SYLNEEELSIEAYKDDPQSGQKLVQFLCDKFYPNKEHEVEERKTILAKN<br>KRQALEHLLFIEVTSIDWKLFEKHKVFTISNGKYLFSHACLFLSLFLYKS<br>EANQLISKIKGFKRNDNQYRSKRQIFTFSKFTSQDVNSEEQHLVKFRD<br>VIQYLNHYPSAWNKHLELKSQYPMQMDKLMRYIVEAEIYRSFPDQTDNH   |

[0593]

|  |  |   |
|--|--|---|
|  |  | <p>RFLFAIREFFGQSCLDTWTGNTPINFSNQEKGFSYEINTSAEIKDIETKL<br/>                 KALVLKGPLNFKEKKEQNRLEKDLRREKKEQPTNRVKEKLLTRIQHNMML<br/>                 YVSYGRNQDRFMDFAARFLAETDYFGKDAKFKMYQFYTSDEQRDLKE<br/>                 QKKELPKKEFEKLYHQSKLVDFYTYAEQQARYPDWDTPFVVENNAIQI<br/>                 KVTLFNGAKKIVSVQRNMLYLLLEDALYSEKRENAGKGLISGYFVHHQK<br/>                 ELKDQLDILEKETEISREQKREFKLLPKRLLHRYSPAQINDTTEWNPMEV<br/>                 ILEEAKAQEQRYQLLLEKAILHQTEEDFLKRNGKQKFLRFVKA WHLM<br/>                 YLKELYMNKVAEHGHHKSFHITKEEFNDFCRWMAFDEVPKYKEYLCD<br/>                 YFSQKGFNNAEFKDLIESSTSLNDLYEKTQRFEFGWSKDLTKQSDENKY<br/>                 LLANYESMLKDDMLYVNISHFISYLESKGKINRNAHGHIAYKALNNVPHL<br/>                 IEEYYYKDRLAPEEYKSHGKLYNKLKTVKLEDALLYEMAMHYLSLEPAL<br/>                 VPKVKTVKDILSSNIAFDIKDAAGHHLYLHLLIPFHKIDSFVALINHQSQQ<br/>                 EKDPDKTSFLAKIQPYLEKVKNSKDLKAVYHYKDTPHLTRYEDLNMIH<br/>                 SHIVSQSVQFTKVALKLEEYFIAKKSITLQIARQISYSEIADLSNYFTDEV<br/>                 NTAHFHDVPETAYSMILQIESEFLDREIKPQPKLSSELSTQQVSVCTAFL<br/>                 ETLHNNLFDKDDKERRLSKARERYFEQIN</p> |
|--|--|---|

[0594] 在某些示例实施方式中,所述CRISPR效应子蛋白是选自表2的Cas13a 蛋白。

[0595] 表2

[0596]

|        |   |  |   |
|--------|---|--|---|
| c2c2-5 | 1 | <p>毛螺科菌<br/>                 (Lachnospiraceae<br/>                 bacterium)<br/>                 MA2020<br/>                 (SEQ ID<br/>                 No. 193)</p> | <p>MQISKVNHKHVAVGQKDRERITGFIYNDPVGDEKSLEDVVA<br/>                 KRANDTKVLFNVFNTKDLYDSQESDKSEKDKKIEISKGAKFVA<br/>                 KSFNSAITLKKQNKIYSTLTSQQVIKELKDKFGGARIYDDIE<br/>                 EALTETLKKSFRENVRNSIKVLIENAAGIRSSLSKDEEELIQE<br/>                 YFVKQLVEEYTKTKLQKNVVKSIKNQNMVIQPDSDSQVLSLS<br/>                 ESREKQSSAVSSDTLVNCKEKDVLKAFLTDYAVLDEDERN<br/>                 SLLWKLRLNLVNLVYFGSESIRDYSYTKESVWKEHDEQKAN<br/>                 KTLFIDEICHITKIGKNGKEQKVLDEENRSRCKQININYRS<br/>                 ALNYAKNNTSGIFENEDSNHFWIHLIENEVERLYNGIENGEEF<br/>                 KFETGYISEKVWKA VINHLSIKYIALGKAVVNYAMKELSSPG<br/>                 DIEPGKIDDSYINGITSFDYEIHKAEESLQRDISMNVVFATNYL<br/>                 ACATVDTDKDFLLFSKEDIRSCTKKGDNLCKNIMQFWGGYS<br/>                 TWKNFCEEYLKDDKDALELLYSLKSMLYSMRNSSFHSTEN<br/>                 VDNGSWDTELIGKLFEEDCNRAARIEKEKFYNNNLHMFYSSS<br/>                 LLEKVLRLYSSHHERASQVPSFNRVFRKNFPSSLSEQRITP<br/>                 KFTDSKDEQIWQSAVYYLCKEIIYNDFLQSKEAYKLFREGV<br/>                 KNLDKNDINNQAADSFQAVVYYGKAIGNATLSQVCQAI<br/>                 MTEYNRQNNNDGLKKSAYAEKQNSNKYKHYPLFLKQVLQS<br/>                 AFWEYLDENKEIYGFISAQIHKSNVEIKAEDFIANYSSQQYKK<br/>                 LVDKVKKTPELQKWYTLGRLINPRQANQFLGSIRNYVQFVK<br/>                 DIQRRAKENGNPIRNYEVLESDSIIKILEMCTKLNGTTSNDIH<br/>                 DYFRDEDEYAEYISQFVNFVDVHSGAALNAFCNSESEGGKN<br/>                 GIYYDGINPIVNRNWVLCCLYKSPDLISKIISRVNENMIHDFH<br/>                 KQEDLIREYQIKGICSNKKEQQDLRTFQVLKNRVELRDIVEYS<br/>                 EIINELYGQLIKWCYLRERDLMYFQLGFHYLCLNNASSKEAD<br/>                 YIKINVDDRNISGAILYQIAAMYINGLPVYYKKDDMYVALKS<br/>                 GKKASDELNSNEQTSKKINYLKYGNLGDKKDQLYLAGL<br/>                 ELFENVAEHENIIIFRNEIDHFHYFYDRDRSMLDLYSEVDFRFF<br/>                 TYDMKLRKNVVNMLYNILLDHNIVSSVVFETGEKKVGRGDS<br/>                 EVIKPSAKIRLRANNGVSSDVFTYKVGSKDELKIATLPAKNEE</p> |
|--------|---|--|---|

[0597]

|        |   |   |   |
|--------|---|---|---|
|        |   |   | FLLNVARLIYYPDMEAVSENMVREGVVKVEKSNDKKGKISR<br>GSNTRSSNQSKYNNKSKNRMNYSMGSIFEKMDLKF  |
| c2c2-6 | 2 | 毛螺科菌<br>NK4A179<br>(SEQ ID<br>No. 194)              | MKISKVREENRGAKLTVNAKTAVVSENRSQEGILYNDPSRY<br>GKSRKNDIEDRDYIESRLKSSGKLYRIFNEDKNKRETDLQW<br>FLSEIVKINRRNGLVLSMMLSVDDRAFKAFAEKYAELSYTN<br>RRNKVSGSPAFETCGVDAATAERLKGIISETNFNRIKNNIDN<br>KVSEDIIDRIIAKYLKSLCRERVKRGKLLMNAFDLPYSDP<br>DIDVQRDFIDYVLEDFYHVRAKSQVSRSIKNNMMPVQPEGD<br>GKFAITVSKGGTESGNKRSAEKEAFKKFLSDYASLDERVRDD<br>MLRRMRRLVLYFYGSDDSKLSDVNEKFDVWEDHAARRVD<br>NREFIKLPLENKLANKTKDAERIRKNTVKELYNQNGICY<br>RQAVKAVEEDNNGRYFDDKMLNMFFIHRIEYGVKVIYANLK<br>QVTEFKARTGYLSEKIWKDLINYISIKYIAMGKAVVNYAMDE<br>LNASDKKEIELGKISEEYLSGISSFDYELIKAEEMLQRETAVY<br>VAFARHLSSQTVELDSSENSDFLLKPKGTMDKNDKNKLAS<br>NNILNFKDKETLRDITLQYFGGHSWTFDFDKYLAGGKDD<br>VDFTDLKDVYISMRNDSFHYATENHNGKWNKELISAMFE<br>HETERMTVVMKDKFYSNNLPMFYKNDLKLIDLYKDNV<br>ERASQVPSFNKVFVRKNFPALVRDKDNLGIELDLKADADKG<br>ENELKFYNALYMFKEIYYNAFLNDKNVRERFITKATKQVAD<br>NYDRNKERNLKDRIKSAGSDEKLLREQLQNYIAENDFGQRI<br>KNIVQVNPDYTLAQICQLIMTEYNQQNNGCMQKKSAAKDI<br>NKDSYQHYKMLLLVNLKAFLEFIKENYAFVLKPYKHDLCD<br>KADFVPDFAKYVKPYAGLISRVAGSSELQKWYIVSRFLSPAQ<br>ANHMLGFLHSYKQYVWDIYRRASETGTEINHSAEDKIAGVD<br>ITDVDAVIDLSVKLCGTISSEISDYFKDDEVYAEYISSYLDFFEY<br>DGGNYKDSLNRFCNSDAVNDQKVALYYDGEHPKLNRNILS<br>KLYGERRFLEKITDRVSRSDIVEYYKLLKETSQYQTKGIFDSE<br>DEQKNIKKFQEMKNIVEFRDLMDYSEIADELQGLINWIYLR<br>ERDLNMFQLGYHYACLNNSNKQATYVTLDYQGKKNRKN<br>GAILYQICAMYINGLPLYVVDKDSSEWTVSDGKESTGAKIGE<br>FYRYAKSFENTSDCYASGLEIFENISEHDNITELRNYIEHFRYY<br>SSFDRSFLGIYSEVDFRFFTYDLKYRKNVPTILYNILLQHFVNV<br>RFEFVSGKKMIGIDKKDRKIAKEKECARITIREKNGVYSEQFT<br>YKLKNGTVYVDARDKRYLQSIIRLLFYPEKVNMDemievke<br>KKKPSDNNTGKGYSKRDRQQDRKEYDKYKEKKKKEGNFLS<br>GMGGNINWDEINAQLKN |
| c2c2-7 | 3 | [梭菌属]<br>嗜氨基菌<br>DSM<br>10710<br>SEQ ID<br>No. 195) | MKFSKVDHTRSAVGIQKATDSVHGMLYTDPKKQEVNDLKD<br>RFDQLNVKAKRLYNVFNQSKAEEDDDEKRFVGVVKKLNRE<br>LKDLLFHREVSRYNSIGNAKYNYGKSNPEEIVSNLGMVES<br>LKGERDPQKVISKLLLYLRKGLKPGTDGLRMILEASCGLRK<br>LSGDEKELKVFLQTLDEDFEKKTFKKNLRSIENQNMAVQPS<br>NEGDPHIGITQGRFNSQKNEEKSAIERMMSMYADLNEDHRED<br>VLRKLRRLNVLYFNVDTEKTEEPTLPGEVDTNPVFEVWHDH<br>EKGKENDRQFATFAKILTEDRETRKKEKLAVKEALNDLKS<br>RDHNIMAYRCSIKVTEQDKDGLFFEDQRINRFWIHIESAVER<br>ILASINPEKLYKLRIGYLGEKVVKDLLNLSIKYIAVGVKAVFH  |

[0598]

|        |   |  |   |
|--------|---|--|---|
|        |   |  | <p>FAMEDLGKTGQDIELGKLSNSVSGGLTSFDYEQIRADETLQR<br/>                 QLSVEVAFAANNLFRAVVGQTGKKIEQSKSEENEDFLLWK<br/>                 AEKIAESIKKEGEGNTLKSILQFFGGASSWDLNHFCAAYGNE<br/>                 SSALGYETKFADDLRKAIYSLRNETFHFTTLNKGSFDWNAKL<br/>                 IGDMFSHEAATGIAVERTRFYSNNLPMFYRESDLKRIMDHLY<br/>                 NTYHPRASQVPSFNSVFRKNFRLFLSNTLNTNTSFDTEVYQ<br/>                 KWESGVYYLKFKEIYYNSFLPSGDAHHLFFEGLRRIRKEADNL<br/>                 PIVGKEAKKRNAVQDFGRRCDLKNLSLSAICQMIMTEYNEQ<br/>                 NNGNRKVKSTREDKRKPDIFQHYKMLLRLTQEAFIYIRRE<br/>                 EFKFIFDLPKTLYVMKPVVEEFLPNWKS GMFDSLVERVKQSPD<br/>                 LQRWYVLCKFLNGRLLNQLSGVIRS YIQFAGDIQRRAKANHN<br/>                 RLYMDNTQRVEYYSNVLEVVD FCIKGTSRFSNVFSDYFRDE<br/>                 DAYADYLDNYLQFKDEKIAEVSSFAALKTFCNEEEVKAGIY<br/>                 MDGENPVMQRNIVMAKLF GPDEV LKNVVPKVTREEIEEYYQ<br/>                 LEKQIAPYRQNGYCKSEEDQKKLLRFQRIKNRVEFQTITFESE<br/>                 IINELLGQLISWSFLRERDLLYFQLGFHYLCLHNDTEKPAEYK<br/>                 EISREDGTVIRNAILHQVAAMYVGG LPVYTLADKKLAAFEKG<br/>                 EADCKLSISKDTAGAGKKIKDFFRYSKYVLIKDRMLTDQNQK<br/>                 YTIYLAGLELFENTDEHDNITDVRKYVDHFKYATSDENAM<br/>                 SILDLYSEIHDRFFTYDMKYQKNVANMLENILLRHFVLRPEF<br/>                 FTGSKKVGEKKITCKARAQIEAENGMRSEDFTYKLSDGKK<br/>                 NISTCMIAARDQKYLNTVARLLYYPHEAKKSIVDTREKKNN<br/>                 KKTNRGDGTFNKQKGTARKEKDNGPREFNDTGFSNTPFAGF<br/>                 DPFRRNS</p>   |
| c2c2-8 | 5 | <p>鸡肉杆菌<br/>                 (Carnobact<br/>                 erium<br/>                 gallinarum)<br/>                 DSM 4847<br/>                 (SEQ ID<br/>                 No. 196)</p> | <p>MRITKVKIKLDNKLYQVTM QKEEKYGT LKLNESRKSTAEIL<br/>                 RLKKASFNKS FHSKTINSQENKNATIKKNGDYISQIFEKLVG<br/>                 VDTNKNIRKPKMSLTDLKDLPKKDLALFIKRKFKNDDIVEIK<br/>                 NLDLISLFYNALQKVPGEHFTDES WADFCQEMMPYREYKKN<br/>                 FIERKIILLANSIEQNKGF SINPETFSKRKRVLHQWAIEVQERG<br/>                 DFSILDEKLSKLAEIYNFKKMCKR VQDELNDLEKSMKKGKN<br/>                 PEKEKEAYKKQKNFKIKTIWKDY PYKTHIGLIEKIKENEELNQ<br/>                 FNIEIGKYFEHYFPIKKERCTE DEPYL NSETIATTVNYQLKN<br/>                 ALISYLMQIGKYKQFGLENQVLDSK KLQEIGIYEGFQTKFMD<br/>                 ACVFATSSLKNIIEPMRSGDILGKREFKEAIATSS FVNYHHFFP<br/>                 YFPFELKGMKDRESELIPFGEQTEAKQM QNIWALRGSVQQIR<br/>                 NEIFHSFDKNQKFNLPLQDKSNFEFDASENSTGKSQS YIETDY<br/>                 KFLFEAEKNQLEQFFIERIKSSGALEY YPLKSLEKLFACKEMK<br/>                 FSLGSQVVAFAPSYK KLVKKGHSYQTATEGTANYLGLSYYN<br/>                 RYELKEESFQAQYLLKLIYQYVFLPNFSQGN SPAFRET VKAI<br/>                 LRINKDEARKKMKNKFLRKYAFEQVREMEFKETPDQYM<br/>                 SYLQSEMREEKVRKA EKNDKGFEKNITMNF EKLLMQIFVKG<br/>                 FDVFLTTFAGKELLSSEEKVIKETEISLSK KINEREKTLKASIQ<br/>                 VEHQLVATNSAISYWLFC KLLDSRHLNELRNEMIKFKQSRIK<br/>                 FNHTQHAELIQNLLPIVELTILSN DYDEKNDSQNV DVSAYFED<br/>                 KSLYETAPYVQTDDRTRV SFRPILKLEKYHTKSLIEALLKDNP<br/>                 QFRVAATDIQEW MHKREEIGELVEKRKNLHTEWAEGQQT LG<br/>                 AEKREEYRDYCKKIDRFNWKANKVTLTYLSQLHYLITDLLG<br/>                 RMVGFSALFERDLVYFSRSFSELGGETYHISDYKNLSGVLRL</p> |

[0599]

|         |   |  |   |
|---------|---|--|---|
|         |   |  | <p>NAEVKPIKIKNIKVIDNEENPYKGNPEVKPFLDRLHAYLENV<br/>                 IGIKAVHKGIRNQT AHL SVLQLELSMIESMNNLRDLMAYDRK<br/>                 LKNAVTKSMIKILDKHGMILKLIKIDENHKNFEIESLIPKEIHLK<br/>                 DKAIKTNQVSEEYCQLVLALLTTNPGNQLN</p>  |
| c2c2-9  | 6 | <p>鸡肉杆菌<br/>                 DSM 4847<br/>                 (SEQ ID<br/>                 No. 197)</p> | <p>MRMTKVKINGSPVSMNRSKLNGLVWNGTTNTVNILTKKE<br/>                 QSFAASFLNKTLVKADQVKGYK VLAENIFIIFEQLEKSNSEKP<br/>                 SVYLNIRRLKEAGLKRFFKSKYHEEIKYTSEKNQSVPTKLN<br/>                 LIPLFFNAVDRIQEDKFDEKNWSYFCKEMSPYLDYKKSYLNR<br/>                 KKEILANSIQQNRGFSMPTAEEPNNLSKRKQLFQQWAMKFQE<br/>                 SPLIQQNNFAVEQFNKEFANKINELAAVYNVDELCTAITEKL<br/>                 MNFDKDKSNKTRNFEIKKLWKQHPHNKDKALIKLFNQEGNE<br/>                 ALNQFNIELGKYFEHYFPKTGKKESAESYLNLPQTIKTVGYQ<br/>                 LRNAFVQYLLQVGKLVHYNKGVLDSTLQEIGMYEGFQTKF<br/>                 MDACVFASSSLRNIIQATTNEDILTREKFKKELEKNVELKHDL<br/>                 FFKTEIVEERDENPAKKIAMTPNELDLWAIRGAVQVRNQIF<br/>                 HQQINKRHEPNQLKVGSENGDLGNVSYQKTIYQKLFDAEIK<br/>                 DIEIYFAEKIKSSGALEQYSMKDLEKLFSENKELTSLGGQVVA<br/>                 FAPSYKKLYKQGYFYQNEKTIELEQFTDYDFSNDVFKANY<br/>                 LIKLIYHYVFLPQFSQANNKLFKDTVHYVIQONKELNTTEKD<br/>                 KKNNKKIRKYAFEQVKLMKNESPEKYMQYLQREMQEERTI<br/>                 KEAKKTNEEKPNYNFEKLLIQIFIKGFDTFLRNFDLNLPAEE<br/>                 LVGTVKEKAEGLRKRKERIAKILNVDEQIKTGDEEIAFWIFAK<br/>                 LLDARHLSLRNEMIKFKQSSVKKGLIKNGDLIEQMOPILELC<br/>                 ILSNDSSEMEKESFDKIEVFLEKVELAKNEPYMQEDKLTVPKF<br/>                 RFMKQLEKYQTRNFIENLVIEPFEKVSEKIVLNWHEEKEKIA<br/>                 DLVDRKTKLHEEWASKAREIEEYNEKIKKNKSKKLDKPAEF<br/>                 AKFAEYKIICEAIENFNRLDHKVRLTYLKNLHYLMIDLGRM<br/>                 VGFSVLFERDFVYMGRSYSALKKQSIYLN DYDTFANIRDWE<br/>                 VNENKHLFGTSSSDLTFQETAEFKLNKKPMENQLKALLGVT<br/>                 NHSFEIRNNIAHLHVLNRNDGKGEVSLSCMNDLRKLMYSYD<br/>                 RKLKNAVTKAIKILDKHGMILKLTNNDHTKPFIEISLKPCKII<br/>                 HLEKSNHSFPMQVQSQEYCDLVKKMLVFTN</p> |
| c2c2-10 | 7 | <p>丙酸杆菌<br/>                 WB4 (SEQ<br/>                 ID No.<br/>                 198)</p>      | <p>MRVSKVKVKDGGKDKMVLVHRKTTGAQLVYSGQPVSNET<br/>                 SNILPEKKRQSFDLSTLNKTIKFDTAKKQKLVNDQYKIVEKIF<br/>                 KYPKQELPKQIKAEIILPFLNHKFQEPVKYWKNGKEESFNLT<br/>                 LLIVEAVQAQDKRKLQPYDWTWYIQTSDLLKKSIIENRI<br/>                 DLTENLSKRKKALLAWETEFTASGSIDLTHYHKVYMTDVLK<br/>                 KMLQDVKPLTDDK GKINTNAYHRGLKKALQNHQPAIFGTRE<br/>                 VPNEANRADNQLSIYHLEVVKYLEHYFPIKTSKRNTADDIA<br/>                 HYLKAQTLKTTIEKQLVNAI RANIIQGGKTNHHELKADTTSN<br/>                 DLIRIKTNEAFVLNLGTCAFAANNIRNMVDNEQTN DILGKG<br/>                 DFIKSLLKDNTNSQLYSFFFGEGLSTNKA EKETQLWGIRGAV<br/>                 QQIRNNVNHYKKDALKTVFNISNFENPTITDPKQQTNYADTI<br/>                 YKARFINELEKIPFAAQLKTGGAVSYTYIENLKSLLTTFQF<br/>                 SLCRSTIPFAPGFKKVFNGGINYQNAKQDES FYELMLEQYLR<br/>                 KENFAEESYNARYFMLKLIYNNLFLPGFTDRKAFADSVGVF<br/>                 QMQNKKAQEVNPRKKEAYAFEAVRPMTAADSIADY MAY</p>  |

[0600]

|             |        |   |   |
|-------------|--------|---|---|
|             |        |   | VQSELMQEQNKKEEKVAEETRINFEFVLQVFIKGFDSFLRA<br>KEFDVQMPQPQLTATASNQQKADKLNQLEASITADCKLTP<br>QYAKADDATHIAFYVFCKLLDAAHLSNLRNELIKFRESVNEF<br>KFHHLLEHIEICLLSADVVPDYRDLYSSEADCLARLRPFIEQG<br>ADITNWSDLFVQSDKHSPVIHANIELSVKYGTTKLEQIHKD<br>TQFKTTEANFTAWNTAQKSIEQLIKQREDHHEQVVKAKNAD<br>DKEKQERKREKSNFAQKFIEKHGDDYLDICDYINTYNWLDN<br>KMHFVHLNRLHGLTIELLGRMAGFVALFDRDFQFFDEQQIAD<br>EFKLGHFVNLHSIDKKLNEVPKTKIKEYDIRNKIIQINGNKINE<br>SVRANLIQFISSKRNYNNAFLHVSNDIEKEKQMYDIRNHIAH<br>FNYLTKDAADFLIDLINELRELLHYDRKLKNAVSKAFIDLFD<br>KHGMILKLLNADHKLKVESLEPKKIYHLGSSAKDKPEYQY<br>CTNQVMMAYCNMCRSLLEMKK  |
| c2c2-1<br>1 | 9      | 韦氏李斯特菌<br>( <i>Listeria<br/>weihenstephanensis</i> )FSL<br>R9-0317<br>(SEQ ID<br>No. 199)           | MLALLHQEVPSQKLHNLKSLNTESLTKLFPKPKFQNMISYPPS<br>KGAEHVQFCLTDIAVPAIRDLEIKPDWGIFFEKLPYTDWA<br>ESYIHYKQTTIQKSIEQNKIQSPDSPRKLVLQKYVTAFLNGEPL<br>GLDLVAKKYKLADLAESFKVVDLNEKDSANYKIKACLQQH<br>QRNILDDELKEDPELNQYGIEVKKYIQRYPPIKRAPNRSKHARA<br>DFLKKELIESTVEQQFKNAVYHYVLEQGKMEAYELTDPKTK<br>DLQDIRSGEAFSFKFINACAFASNNLKMILNPECEKDILGKGD<br>FKKNLPNSTTQSDVVKMIPFFSDEIQNVNFDEAIWAIRGSIQ<br>QIRNEVYHCKKHSWKSILKIKGFEPNMMKYTDSMQKLM<br>DKDIAKIPDFIEELKSSGIIRFYSHDKLQSIWEMKQGFSLTT<br>NAPFVPSFKRVYAKGHDYQTSKNRYDYDLGLTTDFILEYGEE<br>DFRARYFLTKLVYYQFMPWFTADNNAFRDAANFVLRNLK<br>NRQQDAKAFINIREVEEGEMPRDYMGYVQGGIAIHEDSTEDT<br>PNHFEEKFISQVFIKGFDSHMRSADLKFKNPRNQGLEQSEIEE<br>MSFDIKVEPSFLKNKDDYIAFWTFCKMLDARHSELRNEMIK<br>YDGHLTGEQEIHGLALLGVDSRENDWKQFFSSEREYKIMKG<br>YVGEELYQREPYRQSDGKTPILFRGVEQARKYGTETVIQRLF<br>DASPEFKVSKCNITEWERQKETIETIERRKELHNEWKPNK<br>KPQNNAFFKEYKECCDAIDAYNWHKNKTTLVYVNELHLLI<br>EILGRYVGYVAIADRDFQCMANQYFKHSGITERVEYWGDNR<br>LKSIIKLDLTKKEGLFVSEKNARNHIAHLNYLSLKSECTLLY<br>LSERLREIFKYDRKLKNAVSKSLIDILDRHGMSVVFANLKEN<br>KHRLVIKSLEPKLRHLGEEKIDNGYIETNQVSEEYCGIVKRL<br>LEI |
| c2c2-1<br>2 | 1<br>0 | 李斯特氏菌<br>( <i>Listeriaceae<br/>bacterium</i> )<br>FSL<br>M6-0635 =<br>纽约李斯特菌<br>( <i>Listeria</i> ) | MKITKMRVDGRTIVMERTSKEGQLGYEGIDGNKTTEIIFDKK<br>KESFYKSILNKTVRKPDEKEKNRRKQAINKAINKEITELMLA<br>VLHQEVPSQKLHNLKSLNTESLTKLFPKPKFQNMISYPPSKGA<br>EHVQFCLTDIAVPAIRDLEIKPDWGIFFEKLPYTDWAESYI<br>HYKQTTIQKSIEQNKIQSPDSPRKLVLQKYVTAFLNGEPLGLD<br>LVAKKYKLADLAESFKLVLDLNEKDSANYKIKACLQQHQRNI<br>LDELKEDPELNQYGIEVKKYIQRYPPIKRAPNRSKHARADFL<br>KKELIESTVEQQFKNAVYHYVLEQGKMEAYELTDPKTKDLQ<br>DIRSGEAFSFKFINACAFASNNLKMILNPECEKDILGKGNFKK<br>NLPNSTTRSDVVKMIPFFSDELQNVNFDEAIWAIRGSIQQR  |

[0601]

|             |        |  |  |
|-------------|--------|--|--|
|             |        | newyorken<br>sis)FSL<br>M6-0635<br>(SEQ ID<br>No. 200)                   | NEVYHCKKHSWKSILKIKGFEPNNMKYADSDMQKLMKD<br>DIAKIPFIEEKLKSSGVVRFYRHDELQSIWEMKQGFSLTTN<br>APFVPSFKRVYAKGHDYQTSKNRYNLDLTTFDILEYGEEDF<br>RARYFLTKLVYYQQFMPWFTADNNAFRDAANFVLRNKNR<br>QQDAKAFINIREVEEGEMPRDYMGYVQGGQIAIHEDSIEDTPN<br>HFEKFISQVFIKGFDRHMRSANLKFKNPRNQGLEQSEIEEMS<br>FDIKVEPSFLKNKDDYIAFWIFCKMLDARHLSLRNEMIKYD<br>GHLTGEQEIIIGLALLGVDSRENDWKQFFSSEREYEKIMKGYV<br>VEELYQREPYRQSDGKTPILFRGVEQARKYGTETVIQRLFDA<br>NPEFKVSKCNLAEWERQKETIETIKRRKELHNEWAKNPKKP<br>QNNAFFKEYKECCDAIDAYNWHKNKTTLAYVNELHLLIEIL<br>GRYVGYVAIADRFQCMANQYFKHSGITERVEYWGDNRLK<br>SIKKLDTFLKKEGLFVSEKNARNHIAHLNYSLKSECTLLYLS<br>ERLREIFKYDRKLKNAVSKSLIDILDRHGMSVVFANLKENKH<br>RLVIKSLEPKLRHLGGKKIDGGYIETNQVSEEYCGIVKRLLE<br>M   |
| c2c2-1<br>3 | 1<br>2 | 韦德纤毛<br>菌<br>(Leptotrichi<br>a<br>wadei)F027<br>9<br>(SEQ ID<br>No. 201) | MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLSELLSIRLDI<br>YIKNPDNASEEENRIRRENLKKFFSNKVLHLKDSVLYLKNRK<br>EKNAVQDKNYSEEDISEYDLKNKNSFVLLKILLNEDVNSEE<br>LEIFRKDVEAKLNKINSLKYSFEENKANYQKINENNVEKVG<br>KSKRNIIYDYRESAKRNDYINNVQEAFLDKLYKKEDIEKLFLL<br>IENSKKHEKYKIREYYHKIIGRKNNDKENFAKIIYEEIQVNNIK<br>ELIEKIPDMSELKKSQVYKYLDKEELNDKNIYAFCHFVEI<br>EMSQLLNKYVYKRLSNISNDKIKRIFEYQNLKLIENKLLNKL<br>DTYVRNCGKYNYYLQVGEIATSDFIARNRQNEAFLRNIGVS<br>SVAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYVSGE<br>VDKIYNENKQNEVKENLKMFSYDFNMDNKNEIEDFFANID<br>EAISSIRHGIVHFNLELEGKDIFAFKNIAPSEISKMFQNEINEK<br>KLKLIKFKQLNSANVFNYYEKDVIKYLKNTKFNFNKNIPFV<br>PSFTKLYNKIEDLRNTLKFVSWPKDKEEKDAQIYLLKNIYY<br>GEFLNKVFNKSVFFKITNEVIKINKQRNQTGHYKYQKFEN<br>IEKTVPEYLAIIQSREMINNQDKEEKNTYIDFIQQIFLKGFDY<br>LNKNNLKYIESNNNDNDNDFSKIKIKKDNKEKYDKILKNYE<br>KHNRNKEIPHEINEFVREIKLKGILKYTENLNMFYLLKLLNH<br>KELTNLKGSLKYQSANKEETFSDLELELNLNLDNNRVTE<br>FELEANEIGKFLDFNENKIKDRKELKKFDTNKIYFDGENIHKR<br>AFYNIKKYGMLNLLLEKIADKAKYKISLKLKEYSNKNEIEK<br>NYTMQQNLHRKYARPKKDEKFNDEDYKEYEKAIGNIQKYT<br>HLKNKVEFNELNLLQGLLLKILHRLVGYSIWERDLRFRLKG<br>EFPENHYIEEIFNFDNSKNVKYKSGQIVEKYINFYKELYKDNV<br>EKRSIYSDKKVKKQEKDLYIRNYIAHFNYIPHAESLLEV<br>LENLRKLLSYDRKLKNAIMKSIVDILKEYGFVATFKIGADKKI<br>EIQTLESEKIVHLKLNKLLKMLTDRNSEELCELVKVMFEYKA<br>LE |
| c2c2-1<br>4 | 1<br>5 | 荚膜红细<br>菌<br>(Rhodobact  | MQIGKVQGR TISEFGDPAGGLKRRKISTDGKNRKELPAHLSSD<br>PKALIGQWISGIDKIYRKPDSRKS DGKAIHSPTPSKMQFDARD<br>DLGEAFWKL VSEAGLAQSDYDQFKRRLHPYGD KFPADS   |

[0602]

|             |        |   |   |
|-------------|--------|---|---|
|             |        | er capsulatus) SB 1003 (SEQ ID No. 202) | GAKLKFEADPPEPQAFHGRWYGAMSKRGNDAKELAAALYE<br>HLHVDEKRIDGQPKRNPKTDKFAPGLVVARALGISSVLP<br>MARLARNWGEEIQTIFYVVDVAASVKEVAKAAVSAAQAFD<br>PPRQVSGRSLSPKVGAFALAEHLERVTGSKRCSFDPAAGPSVL<br>ALHDEVKKTYKRLCARGKNAARAFPADKTELLALMRH<br>NRVRNQVMRMGRVSEYRGQQAGDLAQSHYWTSAGQTEIK<br>ESEIFVRLWVGAFALAGRSMKAWIDPMGKIVNTEKNDRDLT<br>AAVNIRQVISNKEMVAEAMARRGIYFGETPELDRGAEGNE<br>GFV FALLRYLRGCRNQTFLGARAGFLKEIRKELEKTRWGK<br>AKEAEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFVA<br>HYASKEHFSTLYSEIVKAVKDAPEVSSGLPRLKLLLKRADGV<br>RGYVHGLRDTRKHAFATKLPPPPAPREDDPATKARYIALLR<br>LYDGPFRAYASGITGTALAGPAARAKEAATALAQSVNVTKA<br>YSDVMEGRSRLRPPNDGETLREYLSALTGETATEFRVQIGY<br>ESDSENARKQAEFIENYRRDMLAFMFEDYIRAKGFDWILKIE<br>PGATAMTRAPVLP<br>EPIDTRGQYEHWQAALYLMHFVPASDV<br>SNLLHQLRKWEALQGKYELVQDGDATDQADARREALDLVK<br>RFRDVLVFLKTGEARFEGRAAPFDLKPFRALFANPATFDRL<br>FMATPTTARPAEDDPEGDGASEPEL RVARTLRGLRQIARYNH<br>MAVLSDLFAKHKVRDEEVARLAEIEDETQEKSQIVAAQELRT<br>DLHDKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGRVYLG<br>DHLRLHRLMMDVIGRLIDYAGAYERDTGTFLINASKQLGAG<br>ADWAVTIAGAANTDARTQTRKDLAHFNVLDRADGTPDLTA<br>LVNRAREMMAYDRKRKNAVPRSILDMARLGLTLKWQMK<br>DHLLQDATITQAAIKHLDKVRLTVGGPAAVTEARFSQDYLQ<br>MVAAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQSQP<br>DQKPPNKAPSAGSRLPPPVGEVYEGVVVKVIDTGSGLFLAVEGV<br>AGNIGLHISRLRRIREDAIIVGRRYRFRVEIYVPPKSNTSKLNA<br>ADLVRID |
| c2c2-1<br>5 | 1<br>6 | 荚膜红细<br>菌 R121<br>(SEQ ID<br>No. 203)   | MQIGKVQGR TISEFGDPAGGLK RKISTDGKNR KELPAHLSSD<br>PKALIGQWISGIDKIYRKPDSRKS DGKAIHSPTPSKM QFDARD<br>DLGEAFWKL VSEAGLAQSDYDQFKRRLHPYGD KFKPADS<br>GAKLKFEADPPEPQAFHGRWYGAMSKRGNDAKELAAALYE<br>HLHVDEKRIDGQPKRNPKTDKFAPGLVVARALGISSVLP<br>MARLARNWGEEIQTIFYVVDVAASVKEVAKAAVSAAQAFD<br>PPRQVSGRSLSPKVGAFALAEHLERVTGSKRCSFDPAAGPSVL<br>ALHDEVKKTYKRLCARGKNAARAFPADKTELLALMRH<br>NRVRNQVMRMGRVSEYRGQQAGDLAQSHYWTSAGQTEIK<br>ESEIFVRLWVGAFALAGRSMKAWIDPMGKIVNTEKNDRDLT<br>AAVNIRQVISNKEMVAEAMARRGIYFGETPELDRGAEGNE<br>GFV FALLRYLRGCRNQTFLGARAGFLKEIRKELEKTRWGK<br>AKEAEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFVA<br>HYASKEHFSTLYSEIVKAVKDAPEVSSGLPRLKLLLKRADGV<br>RGYVHGLRDTRKHAFATKLPPPPAPREDDPATKARYIALLR<br>LYDGPFRAYASGITGTALAGPAARAKEAATALAQSVNVTKA<br>YSDVMEGRSSRLRPPNDGETLREYLSALTGETATEFRVQIGY<br>ESDSENARKQAEFIENYRRDMLAFMFEDYIRAKGFDWILKIE<br>PGATAMTRAPVLP<br>EPIDTRGQYEHWQAALYLMHFVPASDV   |

[0603]

|             |        |  |   |
|-------------|--------|--|---|
|             |        |  | SNLLHQLRKWEALQGKYELVQDGDATDQADARREALDLVK<br>RFRDVLVFLKTGEARFEGRAAPFDLKPFRALFANPATFDRL<br>FMATPTTARPAEDDPEGDGASEPELRVARTLRGLRQIARYNH<br>MAVLSDLFAKHKVRDEEVARLAEIEDETQEKSQIVAAQELRT<br>DLHDKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGRVYLG<br>DHLRLHRLMMDVIGRLIDYAGAYERDTGTFLINASKQLGAG<br>ADWAVTIAGAANTDARTQTRKDLAHFNVLDRADGTPDLTA<br>LVNRAREMMAYDRKRKNAVPRSILDMLARLGLTLKWQMK<br>DHLLQDATITQAAIKHLDKVRLTVGGPAAVTEARFSQDYLQ<br>MVAAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQSQPDQK<br>PPNKAPSAGSRLPPPQVGEVYEGVVVKVIDTGSGLFLAVEGV<br>AGNIGLHISRLRRIREDAIIVGRRYRFRVEIYVPPKSNTSKLNA<br>ADLVRID   |
| c2c2-1<br>6 | 1<br>7 | 荚膜红细<br>菌 DE442<br>(SEQ ID<br>No. 204) | MQIGKVQGRITSEFGDPAGGLKRRKISTDGKNRKELPAHLSSD<br>PKALIGQWISGIDKIYRKPDSRSDGKAIHSPTPSKMQFDARD<br>DLGEAFWKLVSEAGLAQSDYDQFKRRLHPYGDQKFPADS<br>GAKLKFEADPPEPQAFHGRWYGAMSKRGNDAKELAAALYE<br>HLHVDEKRIDGQPKRNPKTDFAPGLVVARALGISSVLRPG<br>MARLARNWGEIIIQTYFVVDVAASVKEVAKAAVSAAQAFD<br>PPRQVSGRSLSPKVGFALEHLERVTGSKRCSFDPAAGPSVL<br>ALHDEVKKTYKRLCARGKNAARAFPADKTELLALMRHTHE<br>NRVRNQMVVMGRVSEYRGQQAGDLAQSHYWTSAGQTEIK<br>ESEIFVRLWVGAFALAGRSMKAWIDPMGKIVNTEKNDRDLT<br>AAVNIRQVISNKEMVAEAMARRGIYFGETPELDRDLGAEGNE<br>GFV FALLRYLRGCRNQTFLHLAGARAGFLKEIRKELEKTRWVK<br>AKEAEHVVLTDKTVAAIRAIIDNDAKALGARLLADLSGAFVA<br>HYASKEHFSTLYSEIVKAVKDAPEVSSGLPRLKLLLKRADGV<br>RGYVHGLRDRKHAFATKLP PPPAPRELD DPATKARYIALLR<br>LYDGPFRAYASGITGTALAGPAARAKEAATALAQSVNVTKA<br>YSDVMEGRSSRLRPPNDGETLREYLSALTGETATEFRVQIGY<br>ESDSENARKQAEFIENYRRDMLAFMFEDYIRAKGFDWILKIE<br>PGATAMTRAPVLP EPIDTRGQYEHWQAALYVMHFVPASDV<br>SNLLHQLRKWEALQGKYELVQDGDATDQADARREALDLVK<br>RFRDVLVFLKTGEARFEGRAAPFDLKPFRALFANPATFDRL<br>FMATPTTARPAEDDPEGDGASEPELRVARTLRGLRQIARYNH<br>MAVLSDLFAKHKVRDEEVARLAEIEDETQEKSQIVAAQELRT<br>DLHDKVMKCHPKTISPEERQSYAAAIKTIEEHRFLVGRVYLG<br>DHLRLHRLMMDVIGRLIDYAGAYERDTGTFLINASKQLGAG<br>ADWAVTIAGAANTDARTQTRKDLAHFNVLDRADGTPDLTA<br>LVNRAREMMAYDRKRKNAVPRSILDMLARLGLTLKWQMK<br>DHLLQDATITQAAIKHLDKVRLTVGGPAAVTEARFSQDYLQ<br>MVAAVFNGSVQNPKPRRRDDGDAWHKPPKPATAQSQPDQK<br>PPNKAPSAGSRLPPPQVGEVYEGVVVKVIDTGSGLFLAVEGV<br>AGNIGLHISRLRRIREDAIIVGRRYRFRVEIYVPPKSNTSKLNA<br>ADLVRID |
| c2c2-2      |        | (SEQ ID<br>No. 205)                    | MGNLFGHKRWYEVDRDKKDFKIKRKVKVKRNYDGNKYILNI<br>NENNNKEKIDNNKFIRKYINYKKNNDNILKEFTRKFHAGNILFK   |

[0604]

|               |   |  |
|---------------|---|--|
|               |   | <p>LKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLGITKK<br/>                 KIIDEAIRQGITKDDKIEIKRQENEEEIEIDIRDEYTNKTLNDC<br/>                 SIILRIIENDELETKKSIYEIFKNINMSLYKIIKIIENETEKVFEN<br/>                 RYYEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFY<br/>                 LNVGGDKKSKNKKMLVEKILNINVDLTVEDIADDFVIKELEF<br/>                 WNITKRIEKVKKVNEFLEKRRNRRTYIKSYVLLDKHEKFKIE<br/>                 RENKKDKIVKFFVENIKNSIIEKIEKILAEFKIDELIKKLEKEL<br/>                 KKGNCDEIFGIFKHYKVNFDKSKKFSKKSDEEKELYKIIYRY<br/>                 LKGRIEKILVNEQKVRLKMEKIEIEKILNESILSEKILKRVKQ<br/>                 YTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELIT<br/>                 FFASTNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKILN<br/>                 SKIKIIRDLDFIDNKNITNNFIRKFTKIGTNERNRILHAISKER<br/>                 DLQGTQDDYINKVINIIQNLKISDEEVSKALNLDVVFKDKKNI<br/>                 TKINDIKISEENNDIKYLPFSKVLPEILNLYRNNPKNEPFDTI<br/>                 ETEKIVLNALIYVNKELYKLLILEDDLEENESKNIFLQELKKT<br/>                 GNIDEIDENIIENYKNAQISASKGNNKAIKKYQKKVIECYIG<br/>                 YLRKNYEELDFDSDFKMNIQEIKKQIKDINDNKTYERITVKTS<br/>                 DKTIVINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEY<br/>                 QNIIDILDEIMQLNLTNRNECITENWNLNLEEFIQMKKEIEKDFD<br/>                 DFKIQTKKEIFNNYEDIKNNILTEFKDDINGCDVLEKKLEKIV<br/>                 IFDDETKFEIDKKSNIQDEQRKLSNINKKDLKKKVDQYIKDK<br/>                 DQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIY<br/>                 YPKERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKF<br/>                 LFNIDGKNIRKNKISEIDAILKNLNDKLNGLYSKEYKEYIKKL<br/>                 KENDDDFAKNIQNKYKSFEDYNRVSEYKKIRDLVEFNLYL<br/>                 NKIESYLIDINWKLAIQMARFERDMHYIVNGLRELGIKLSGY<br/>                 NTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKFEKICYGF<br/>                 GIDLSENSEINKPENESIRNYISHFYIVRNPADYSIAEQIDRVS<br/>                 NLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGN<br/>                 NDILERLMKPKKVSVLELESYNSDYIKNLIHELLTKIENTNDTL</p> |
| <p>c2c2-3</p> | <p>韦德纤毛<br/>                 菌 (L<br/>                 wadei)<br/>                 (Lw2)<br/>                 (SEQ ID<br/>                 No. 206)</p> | <p>MKVTKVDGISHKKYIEEGKLVKSTSEENRTSERLSELLSIRLDI<br/>                 YIKNPDNASEEENRIRRENLKKFFSNKVLHLKDSVLYLKNRK<br/>                 EKNAVQDKNYSEEDISEYDLKNKNSFVLLKILLNEDVNSEE<br/>                 LEIFRKDVEAKLNKINSLKYSFEENKANYQKINENNVEKVG<br/>                 KSKRNIIYDYYRESAKRNDYINNVQEAFLKLYKKEDIEKLFFL<br/>                 IENSKKHEKYKIREYYHKIIGRKNNDKENFAKIIYEEIQNVNNIK<br/>                 ELIEKIPDMSELKKSQVYKYLLDKEELNDKNIKYAFCHFVEI<br/>                 EMSQLLKNYVYKRLSNISNDKIKRIFEYQNLKLIENKLLNKL<br/>                 DTYVRNCGKYNYYLQVGEIATSDFIARNRQNEAFLRNIGVS<br/>                 SVAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYVSGE<br/>                 VDKIYNENKQNEVKENLKMFSYDFNMDNKNEIEDFFANID<br/>                 EAISSIRHGIVHFNLELEGKDIFAFKNIAPSEISKMFQNEINEK<br/>                 KLKLIKFKQLNSANVFNYYEKDVIKYLKNTKFNFNKNIPFV<br/>                 PSFTKLYNKIEDLRNTLKKFFWSVPKDKEEKDAQIYLLKNIYY<br/>                 GEFLNKFVKNSKVFFKITNEVIKINKQRNQKTGHYKYQKFEN<br/>                 IEKTPVEYLAIQSREMINNQDKEEKNTYIDFIQQIFLKGFDY<br/>                 LNKNNLKYIESNNNDNDNDFSKIKIKKDNKEKYDKILKNYE<br/>                 KHNRNKEIPHEINEFVREIKLGKILKYTENLNMFYLLKLLNH</p>   |

[0605]

|        |   |  |
|--------|---|--|
|        |   | <p>KELTNLKGSLKYQSANKEETFSDELELINLLNLDNNRVTED<br/>                 FELEANEIGKFLDFNENKIKDRKELKKFDTNKIYFDGENIHKR<br/>                 AFYNIKKYGMLNLEKIADKAKYKISLKEKEYSNKKNEIEK<br/>                 NYTMQQNLHRKYARPKKDEKFNDEYKEYEKAIGNIQKYT<br/>                 HLKNKVEFNELNLLQGLLLKILHRLVGYTSIWERDLRFRLKG<br/>                 EFPENHYIEEIFNFDNSKNVKYKSGQIVEKYINFYKELYKDNV<br/>                 EKRSIYSDKKVKKLKQEKKDLYIRNYIAHFNYIPHAESISLEV<br/>                 LENLRKLLSYDRKLNKAIMKSIVDILKEYGFVATFKIGADKKI<br/>                 EIQTLESEKIVHLKNLKKKKLMTDRNSEELCELVKVMFEYKA<br/>                 LEKRPAATKAGQAKKKKGSYPYDVPDYAYPYDVPDYAYP<br/>                 YDVPDYA*</p>   |
| c2c2-4 | <p>斯氏李斯特菌<br/>                 (Listeria<br/>                 seeligeri)(S<br/>                 EQ ID No.<br/>                 207)</p> | <p>MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRITKVEVD<br/>                 RKKVLISRDKNGGKLVYENEMQDNTEQIMHHKSSFYKSVV<br/>                 NKTICRPEQKQMKLVHGLLQENSQEKIKVSDVTCLNINFL<br/>                 NHRFKKSLYFPENSPDKSEEYRIEINLSQLLEDLKKQGGTFI<br/>                 CWESFSKDMELYNWAENYISSKTKLIKKSIRNNRIQSTESRS<br/>                 GQLMDRYMKDILNKNKPFDIQSVSEKYQLEKLTSAKATFK<br/>                 EAKKNDKEINYKLNKSTLQNHHERQIIEELKENSELNQFNIEIRK<br/>                 HLETYFPIKKTNRKVGDIRNLEIGEIQKIVNHRLKNKIVQRILQ<br/>                 EGKLASYEIESTVNSNSLQKIKIEEAFALKFINACLFASNNLRN<br/>                 MVYVPVCKKDILMIGEFKNSFKEIKHKKFIRQWSQFFSQEITVD<br/>                 DIELASWGLRGAIAPIRNEIHLKHSWKKFFNNPTFKVKKSK<br/>                 IINGKTKDVTSEFLYKETLFDYFYSELDSVPELIINKMESSKIL<br/>                 DYYSSDQLNQVFTIPNFELSLTSAVFPFAPSFRVYLKGFQYQ<br/>                 NQDEAQPDPYLNKLNINYEKAFNSEAFQAQYSLFKMVYYQVF<br/>                 LPQFTTNDLKFSSVDFILTLNKERKGYAKAFQDIRKMNKDE<br/>                 KPSEYMSYIQSQLMLYQKKQEEKEKINHFEFKINQVFIKGFNS<br/>                 FIEKNRLTYICHPTKNTVPENDNIEIPFHTDMDDSNIAFWLMC<br/>                 KLLDAKQSELRNEMIKFSCSLQSTEEISTFTKAREVIGLALLN<br/>                 GEKGCNDWKELFDDKEAWKKNMSLYVSEELLQSLPYTQED<br/>                 GQTPVINRSIDLKVKYGTETILEKLFSSDDYKVSADIAKLH<br/>                 EYDVTEKIAQQESLHKQWIEKPGLARDSAWTKKYQNVINDIS<br/>                 NYQWAKTKVELTQVRHLHQLTIDLLSRLAGYMSIADRDFQF<br/>                 SSNYILERENSEYRVTSWILLSSENKNKNKYNDYELYNLKNAS<br/>                 IKVSSKNDPQLKVDLQQLRLTLEYLELFDNRLKEKRNNISHF<br/>                 NYLNGQLGNSILELFDADARDVLSYDRKLNKAVSKSLKEILSS<br/>                 HGMEVTFKPLYQTNHHLKIDKLQPKKIHLGKSTVSSNQVS<br/>                 NEYCQLVRTLLTMK</p> |

[0606]

|       |  |  |
|-------|--|--|
| C2-17 | <p>口腔纤毛菌<br/>                 (Leptotrichi<br/>                 a buccalis)<br/>                 C-1013-b<br/>                 (SEQ ID<br/>                 No. 208)</p> | <p>MKVTKVGGISHKKYTSEGRLVKSESEENRTDERLSALLNMR<br/>                 LDYIKNPSSSTETKENQKRIGKLLKFFSNKMVYLKDNTLSLK<br/>                 NGKKENIDREYSETDILESVDKKNFAVLKKIYLNENVNSE<br/>                 ELEVFRNDIKKLNKINSLKYSFEKNKANYQKINENNIEKVE<br/>                 GKSKRNIIYDYRESAKRDAYVSNVKEAFDKLYKEEDIAKL<br/>                 LEIENLTKLEKYKIREFYHEIIGRKNKDNKFNFAKIIYEEIQNVN<br/>                 MKELIEKVPDMSELKKSQVFYKYLDKEELNDKNIKYAFCH<br/>                 FVEIEMSQLLNKYVYKRLSNISNDKIKRIFEYQNLKLIENKL</p> |
|-------|--|--|

[0607]

|              |  |  |
|--------------|--|--|
|              |  | <p>LNKLDTYVRNCGKYNYYLQDGEIATSDFIARNRQNEAFLRNI<br/>             IGVSSVAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYV<br/>             SGEVDKIYNENKKNEVKENLKM FYSYDFNMDNKNEIEDFFA<br/>             NIDEAISSIRHGIVHFNLELEGGKDIKAFKNIAPSEISKMFQNEI<br/>             NEKKLKLKIFRQLNSANVFRYLEKYKILNYLKRTRFEFVNKNI<br/>             PFVPSFTKLYSRIDDLKNSLGIYWKTPK TNDNKTKEIIDAQI<br/>             YLLKNIYYGEFLNYFMSNNGNFFEISKEIHELKNDKRNKLTG<br/>             FYKLQKFEDIQEKIPKEYLANIQSLYMINAGNQDEEEKDTYID<br/>             FIQKIFLKGFM TYLANNGRLSLIYIGSDEETNTSLAEKKQEFD<br/>             KFLKKYEQNNNIKIPYEINEFLREIKLGNILKYTERLNM FYLIL<br/>             KLLNHKELTNLKG SLEKYQS ANKEEAFSDQLELINLLNLDNN<br/>             RVTEDFELEADEIGKFLDFNGNKVKDNKELKKFD TNKIYFDG<br/>             ENIIKHAFYNIKKY GMLNLLLEKIADKAGYKISIEELKKYSNK<br/>             KNEIEKNHKMQENLHRKYARPRKDEKFTDEDYESYKQAIEN<br/>             IEEYTHLKNKVEFNELNLLQGLLRILHRLVGYTSIWERDLRF<br/>             RLKGEFPENQYIEEIFNFENKKNV KYKGGQIVEKYIKFYKELH<br/>             QNDEVKINKYSSANIKVLKQEKDLYIRNYIAHFNYIPHAEIS<br/>             LLEVLENLRKLLSYDRKLLKNAVMSVVDILKEYGFVATFKIG<br/>             ADKKIGIQTLESEKIVHLK NLK KKKLMTDRNSEELCKLVKIM<br/>             FEYKMEEKKSEN</p>   |
| <p>C2-18</p> | <p>解半纤维<br/>             素哈尔滨<br/>             菌<br/>             (Herbinix<br/>             hemicellulo<br/>             silytica)(SE<br/>             Q ID No.<br/>             209)</p> | <p>MKLTRRRISGNSVDQKITA AFYRDMSQGLLYDSEDNDCTD<br/>             KVIESMDFERSWRGRILKNGEDDKNPFYMFVKGLVGSNDKI<br/>             VCEPIDVSDPDNLDILINKNLTGFGRNLKAPDSNDTLENLIR<br/>             KIAGAPIEEVLP ELK KIKEMI QKDIVNRKEQLLSIKNNRIPFS<br/>             LEGSKLVPSTKMKWLFKLIDVPNKTFNEKM LEKYWEIYDY<br/>             DKLKANITNRLDKTDKKARSISRAVSEELREYHKNLRTNYNR<br/>             FVSGDRPAAGLDNGGSAKYNPDKEEFLFLKEVEQYFKKYF<br/>             PVKSKHSNKS KDKSLVDKYKNYCSYKVVKKEVNRSIINQLV<br/>             AGLIQQGKLLYYFYNDTWQEDFLNSYGLSYIQVEEAFKKS<br/>             VMTSLSWGINRLTSFFIDDSNTVKFDDITTKKAKEAIESNYFN<br/>             KLRTCSRMQDHFKEKLAFFYPVYVKDKKDRPDDDIENLIVL<br/>             VKNAIESVSYLRNRTFHFKESSLELLKELDDKNSGQNKIDYS<br/>             VAAEFIKRDIENLYDVFREQIRSLGIAEYKADMISDCFKTCG<br/>             LEFALYSPKNSLMPAFKNVYKRGANLNKAYIRDKGPKETGD<br/>             QGQNSYKALEEYRELTWYIEVKNNDSYNA YKNLLQLIYYH<br/>             AFLPEVRENEALITDFINRTKEWNRKETEERLNTKNNKHKHN<br/>             FDENDDITVNTYRYESIPDYQGESLDDYLKVLQRKQMARAK<br/>             EVNEKEEGNNNYIQFIRDVVVWAFGAYLENKLNKYNKNEIQP<br/>             PLSKENIGLNDTLKELFPEEKVKSPFNIKCRFSISTFIDNKGKST<br/>             DNTSAEAVKTDGKEDEKDKKNIKRKDLLCFYFLRLLDENEI<br/>             CKLQHQFIKYRCSLKERRFPGNRTKLEKETELLAELEELMEL<br/>             VRFTMPSIPEISAKAESGYDTMIKKYFKDFIEKKVFKNPKTSN<br/>             LYYHSDSKTPVTRKYMALLMRSAPLHLYKDIFKGYYLITKKE<br/>             CLEYIKLSNIIKDYQNSLNELHEQLERIKLSEKQNGKDSL YL<br/>             DKKDFYKVKEYVENLEQVARYKHLQH KINFESLYRIFRIHVD<br/>             IAARMVGYTQDWERDMHFLFKALVYNGVLEERRFEAIFNNN<br/>             DDNNDGRIVKKIQNNLNNKNRELVSMLCWNKLNKNEFGAI<br/>             IWKRNP IAHLNHFTQTEQNSKSSLESLSLIRILLAYDRKRQN</p> |

[0608]

|       |   |  |
|-------|---|--|
|       |   | AVTKTINDLLLNDYHIRIKWEGRVDEGQIYFNIKEKEDIENEPI<br>IHLKHLHKKDCYIYKNSYMFDKQKEWICNGIKKEEVYDKSILK<br>CIGNLFKFDYEDKNKSSANPKHT   |
| C2-19 | [真杆菌属]<br>直 肠 菌<br>(SEQ ID<br>No. 210)  | MLRRDKEVKKLYNVFNQIQVGTKPKKWNNDEKLSPEENER<br>RAQQKNIKMKNYKWREACSKYVESSQRIINDVIFYSYRKAK<br>NKLRYMRKNEDILKKMQEAEKLSKFSGGKLEDFVAYTLRKS<br>LVVSKYDTQEFDSLAAVVVLEFCIGKNNISDHEREIVCKLLEL<br>IRKDFSKLDPNVKGSQGANIVRSVRNQNMIQVQPGDRFLFPQ<br>VYAKENETVTNKNVEKEGLNEFLNYANLDDEKRAESLRKL<br>RRILDVYFSAPNHYEKMDITLSDNIEKEKFNWVEKHECGKK<br>ETGLFVDIPDVLMEAEAENIKLDAVVEKREKVLNDRVRKQ<br>NIICYRYTRAVVEKYNSNEPLFFENNAINQYWIHHIENAVERI<br>LKNCKAGKLFKLRKGYLAEKVWKDAINLISIKYIALGKAVYN<br>FALDDIWKDKKKNKELGIVDERIRNGITSFDYEMIKAHENLQR<br>ELAVDIAFSVNNLARAVCDMSNLGNKESDFLLWKRNDIADK<br>LKNKDDMASVSAVLQFFGGKSSWDINIFKDAYKGGKKYNY<br>EVRFIDDLRKAICYCARNENFHFKTALVNDEKWNTLFGKIFE<br>RETEFCLNVEKDRFYSNNLYMFYQVSELRNMLDHLYSRSVS<br>RAAQVPSYNSVIVRTAFPEYITNVLGYQKPSYDADTLGKWY<br>SACYLLKEIYYNSFLQSDRALQLFEKSVKTLSDWDDKKQQR<br>AVDNFKDHFSDIKSACTSLAQVCQIYMTEYNQQNNQIKKVR<br>SSNDSIFDQPVYQHYKVLKKAIAANAFADYLKNNKDLFGFIG<br>KPFKANEIREIDKEQFLPDWTSRKYEALCIEVSGSQELQKWYI<br>VGKFLNARSLNLMVGSMSYIYVTDIKRRAASIGNELHVS<br>HDVEKVEKWVQVIEVCSLLASRTSNQFEDYFNDKDDYARYL<br>KSYVDFSNVDMPEYSALVDFSNEEQSDLYVDPKNPKVNRN<br>IVHSKLFAADHILRDIVEPVSKDNIEEFYSQKAEIAYCKIKGKE<br>ITAEQKAVLKYQKLNKRVELRDIVEYGEIINELGQLINWSF<br>MRERDLYFQLGFHYDCLRNDSKKPEGYKNIKVDENSIKDAI<br>LYQIIGMYVNGVTVYAPEKDGDKLKEQCVKGGVGVKVSFAF<br>HRYSKYLGLNEKTLYNAGLEIFEVVAEHEDIINLRNGIDHFKY<br>YLGDYRSMLSIYSEVDFRFFTYDIKYQKNVNLNLLQNILLRH<br>VIVEPIESGFKTIGEQTTPGAKLSIRSISDFTQYKVKGGTLIT<br>DAKDERYLETIRKILYYAENEEDNLKKSVVVTNADKYEKNK<br>ESDDQNKQKEKKNKDNKGKKNEETKSDAEKNNNERLSYNP<br>FANLNFKLSN |
| C2-20 | 真 杆 菌<br>(Eubacteria<br>ceae<br>bacterium)<br>CHKCI004<br>(SEQ ID<br>No. 211) | MKISKESHKRTAVAVMEDRVGGVVYVPGGSGIDLSNNLKKR<br>SMDTKSLYNVFNQIQAGTAPSEYEWKDYLSEAENKKREAQK<br>MIQKANYELRRECEDYAKKANLAVSRIIFSKPKKIFSDDDIIS<br>HMKKQRLSKFKGRMEDFVLIALRKSLLVSTYNQEVFDSRKA<br>ATVFLKNIGKKNISADDERQIKQLMALIREDYDKWNPDKDSS<br>DKKESGTVKIRSIEHQNMVIQPEKNKLSLSKISNVGKKTCTK<br>QKEKAGLDAFLKEYAQIDENSRMEYLKLRLLDITYFAAPS<br>SYIKGAAVSLPENINFSSSELNVWERHEAAKVNINFEIPESL<br>LNAEQNNKINKVEQHSLEQLRTDIRRRNITCYHFANALAA<br>DERYHTLFFENMAMNQFWIHHMENAVERILKKNVGTFLFKL<br>RIGYLSEKVVWKMNLNLSIKYIALGKAVYHFAALDDIWKADI  |

|               |   |  |
|---------------|---|--|
|               |   | <p>WKDASDKNSGKINDLTLKGISSFDYEMVKAQEDLQREMAV<br/>                 GVAFSTNNLARVTCKMDDLSDAESDFLLWNKEAIRRHVKYT<br/>                 EKGEILSAILQFFGGRSLWDESLFEKAYSSDSNYELKFLDDLKR<br/>                 AIYAARNETFHFKTAIDGGSWNTRFLGSLFEKEAGLCLNVE<br/>                 KNKFYSNNLVLFYKQEDLRVFLDKLYGKECSRAAQIPSYNTI<br/>                 LPRKSFSDFMKQLLGLKEPVYGSAILDQWYSACYLKFKEVY<br/>                 YNLFLQDSSAKALFEKAVKALKGADKKQEKAVERFRKRYW<br/>                 EISKNASLAEICQSYITEYNQQNNKERKQRSANDGMFNEPIY<br/>                 QHYKMLLKEALKMAFASYIKNDKELKFVYKPTKLFVVSQD<br/>                 NFLPNWNSEKYNTLISEVKNSPDLQKWYIVGKFMNARMLNL<br/>                 LLGSMRSLQYVSDIQKRAAGLGENQLHLSAENVGQVKKWI<br/>                 QVLEVCLLSVRISDKFTDYFKDEEEYASYLKEYVDFEDSAM<br/>                 PSDYSALLAFSNEGKIDLYVDASNPKVNRNIIQAKLYAPDMV<br/>                 LKKVVKKISQDECKEFNEKKEQIMQFKNKGDEVSWEQKQI<br/>                 LEYQKLKNRVELRDLSEYGELINELLGQLINWSYLRERDLY<br/>                 FQLGFHYSCLMNESKPKDAYKTIRRGTVSIENAVLYQIIAMYI<br/>                 NGFPVYAPEKGELKPCCKTGSAGQKIRAFQWASMVVEKKK<br/>                 YELYNAGLELFEVVEHDNIIDLNRNKIDHFKYYQGNDSILAL<br/>                 YGEIFDRFFTYDMKYRNNVLNHLQNILLRHNVIKPIISKDKK<br/>                 EVGRGKMKDRAAFLLEEVSDDRFTYKVKEGERKIDAKNRLY<br/>                 LETVRDILYFPNRAVNDKGEDVIICSKKAQDLNEKKADRDN<br/>                 HDKSKDTNQQKKEGKNQEEKSENKEPYSRMTWKPFAGIKLE</p>  |
| <p>[0609]</p> | <p>C2-21<br/>                 布劳特氏<br/>                 菌 (Blautia<br/>                 sp.) 马赛<br/>                 (Marseille)-<br/>                 P2398<br/>                 (SEQ ID<br/>                 No. 212)</p> | <p>MKISKVDHVKSGIDQKLSQRGMLYKQPQKKYEGKQLEEHV<br/>                 RNLSRKAKALYQVFPVSGNSKMEKELQIINSFIKNILLRLDSG<br/>                 KTSEEIVGYINTYSVASQISGDHIQELVDQHLKESLRKYTCVG<br/>                 DKRIYVPDIIVALLKSKFNSETLQYDENSELKILIDFIREDYLKE<br/>                 KQIKQIVHSIENNSTPLRIAEINGQKRLIPANVDNPKKSYIFEFL<br/>                 KEYAQSDPKGQESLLQHMRYLILLYLGPDKITDDYCEEIEA<br/>                 WNFSGSIVMDNEQLFSEEASMLIQDRIYVNNQIEEGRQSKDTA<br/>                 KVKKNKSKYRMLGDKIEHSINESVVKHYQEACKAVEEKDIP<br/>                 WIKYISDHVMSVYSSKNRVDLTKLSLPLAKNTWNTWISFIA<br/>                 MKYVDMGKGVYHFAMSDVDKVGKQDNLIIGQIDPKFSDGIS<br/>                 SFDYERIKAEDDLHRMSGYIAFAVNNFARAICSDFRKKNR<br/>                 KEDVLTVGLDEIPLYDNVVRKLLQYFGGASNWDDSIIDIIDDK<br/>                 DLVACIKENLYVARNVNFHFAGSEKQVKKQDDILEEIVRKET<br/>                 RDIGKHRYKVFYSNNVAVFYCDEDIIKLMNHLQYREKPYQA<br/>                 QIPSYNKVISKTYLPDLIFMLLKGNRTKISDPSIMNMFRTGTF<br/>                 YFLLKEIYYNDFLQASNLKEMFCEGLKNNVKNKKSEKPYQN<br/>                 FMRRFEELNMGMDFGEICQQIMTDYEQQNKQKKKTATAV<br/>                 MSEKDKKIRTLNDNTQKYKHFRLLYIGLREAFIYKDEKN<br/>                 KEWYEFLEPVRKREQPEEKEFVNKWKLNQYSDCSELILKDSL<br/>                 AAAYVVAHFINQAQLNHLIGDIKNYIQFISDIDRRAKSTGNP<br/>                 VSESTEIQIERYRKILRVLEFAKFFCGQITNVLTDYQDENDFS<br/>                 THVGHYVKFEKKNMEPAHALQAFSNSLYACGKEKKKAGFY<br/>                 YDGMNPVNRNITLASMYGKKNLLENAMNPVTEQDIRKYYS<br/>                 LMAELDSVLKNGAVCKSEDEQKNLRHFQNLKNRIELVDVLT<br/>                 LSELVNDLVAQLIGWVYIRERDMMYLQLGLHYIKLYFTDSV<br/>                 AEDSYLRTLBLEEGSIADGAVLYQIASLYSFNLPMPYVKPNKS</p> |

[0610]

|       |   |  |
|-------|---|--|
|       |   | SVYCKKHVNSVATKFDIFEKEYCNGDETVIENGLRRLFENINL<br>HKDMVKFRDYLAHFKYFAKLDESILELYSKAYDFFFSYNIKL<br>KKSVSYVLTNVLLSYFINAKLSFSTYKSSGNKTVQHRTTKISV<br>VAQTDYFTYKLRISIVKNKNGVESIENDDRRCEVVNIAARDKE<br>FVDEVCNVINYNSDK   |
| C2-22 | 纤毛菌口<br>分类单元<br>879 菌株<br>F0557(Lept<br>otrichia sp.<br>oral taxon<br>879 str.<br>F0557)<br>(SEQ ID<br>No. 213) | MGNLFGHKRWYEVDRDKDFKIKRKVKVKRNYDGNKYILNI<br>NENNNKEKIDNNKFIGEFVNYKKNNNVLKEFKRKFHAGNILF<br>KLKGKEEIRIENNDDFLETEEVVLVYIEVYGKSEKLEITKK<br>KIIDEAIRQGITKDDKKIEIKRQENEEEIEIDIRDEYTNKTLNDC<br>SIILRIIENDELETKKSIYEIFKNINMSLYKIIKIIENETEKVFEN<br>RYEEHLREKLLKDNKIDVILTNFMEIREKIKSNLEIMGFVKF<br>YLNVS GDKKKSENKKMFVEKILNTNVDLTVEDIVDFIVKELK<br>FWNITKRIEKVKKFNNEFLENRRNRITYIKSYVLLDKHEFKIE<br>RENKKDKIVKFFVENIKNNSIKEKIEKILAEFKINELIKKLEKEL<br>KKGNCDEIFGIFKHYKVNFDSSKFSNKSDEEKELYKIIYRY<br>LKGRIEKILVNEQKVRLKMEKIEIEKILNESILSEKILKRVKQ<br>YTLHEIMYLGKLRHNDIVKMTVNTDDFSRLHAKEELDLELIT<br>FFASTNMELNKIFNGKEKVTDFGFLNGQKITLKEKVPFVKL<br>NILKLNFINNENNIDEKLSHFYSFQKEGYLLRNKILHNSYGNL<br>QETKNLKG EYENVEKLIKELKSDEEISKSLSLDVIFEGKVDII<br>NKINSLKIGEYKDKKYLPSFSKIVLEITRKFREINKDKLFDIESE<br>KIILNAVKYVVKILYEKITSNEENEFLKTLDPDKLVKKSNNKKE<br>NKNLLSIEEYKNAQVSSSKGDKKAIKKYQNKVTNAYLEYL<br>ENTFTEIIDFSKFNLYDEIKTKIEERKDNKSKIIIDSISTNINITN<br>DIEYIISIFALLNSNTYINKIRNRFATSVWLEKQNGTKEYDYE<br>NIISILDEVLLINLLRENNITDILDKNHAIDAKIVENDETYIKNY<br>IFESNEEKLLKRLFCLELVDKEDIRKIFEDENFKFSFIKKNEIG<br>NFKINFGILSNLECNSEVEAKKIIGKNSKKLESFIQNIIDEYKSN<br>IRTLFSSEFLEKYKEEIDNLVEDTESENKNKFEKIYYPKEHKN<br>ELYIYKKNLFLNIGNPNFDKIYGLISKDIKNVDTKILFDDDIKK<br>NKISEIDAILKNLNDKLNYSNDYKAKYVVKLKENDDFFAK<br>NIQENYSSFGFEKDYNKVSEYKKIRDLVEFNLYLNKIESYLI<br>DINWKLAIQMARFERDMHYIVNGLRELGIKLSGYNTGISRA<br>YPKRNGSDGFYTTTAYYKFFDEESYKFEKICYGFGIDLENS<br>EINKPENESIRNYISHFYIVRNPFADYSIAEQIDRVSNLLSYSTR<br>YNNSTYASVFEVFKKDVNLDYDELKKKFRLLIGNNDILERLM<br>KPKKVS VLELESYNSDYIKNLIHELLTKIENTNDTL |
| C2-23 | 毛螺科菌<br>NK4A144<br>(SEQ ID<br>No. 214)  | MKISKVDHTRMAVAKGNQHRREISGILYKDPTKTGSIDFDE<br>RFKKNLCSAKILYHVFNGIAEGSNKYKNIVDKVNNNLDRVLF<br>TGKSYDRKSIIDIDTVLRNVEKINAFDRISTEEREQIIDLLEIQ<br>LRKGLRKGKAGLREVLLIGAGVIVRTDKKQEIADFLEILDED<br>NKTNQAKNIKLSIENQGLVSPVSRGEERIFDVSGAQK GKSS<br>KKAQEKEALS AFLDYADLDKNVRFEYLRKIRRLINLYFYVK<br>NDDVMSLTEIPAENVLEKDFDIWRDHEQRKEENGDFVGCPI<br>LLADR DVKKSNSKQVKIAERQLRESIREKNIKRYRFSIKTIEK<br>DDGTFFANKQISVFWIHRIENAVERILGSINDKKLYRLRLGY<br>LGEKVWKDILNFLSIKYIAVGKAVNFAMDDLQEKDRDIEPG  |

[0611]

|       |   |  |   |
|-------|---|--|---|
|       |   |  | <p>KISENAVNGLTSFDYEQIKADEMLQREVA VNVFAANNLAR<br/>                 VTVDIPQNGEKEDILLWNKSDIKKYKKNSSKKGILKSILQFFGG<br/>                 ASTWNMCMFEIAYHDQP GDYEENYLYDIIQIYSLRNKSFHF<br/>                 KTYDHGDKNWNRELIGK MIEHDAERVISVEREK FHSNNLPM<br/>                 FYKDADLKKILDLLYSDYAGRASQVPAFNTVLVRKNFPEFLR<br/>                 KDMGYKVHFNPEVENQWHS AVYYLYKEIYYNLFLRDKEV<br/>                 KNLFYTS LKNIRSEVSDKKQKLASDDFASRCEEIEDRSLPEIC<br/>                 QIIMTEYNAQNFGNRKVKSQRVIEKNKDIFRHYKMLLIKT LA<br/>                 GAFSLYLKQERFAFIGKATPIPYETTDVKNFLPEWKSGMYAS<br/>                 FVEEIKNNLDLQEWYIVGRFLNGRMLNQLAGSLRSYIQAED<br/>                 IERRAAENRNKLFSPDEKIEACKKAVRVL DLCIKISTRISAEF<br/>                 TDYFDSEDDYADYLEKYLKYQDDAIKELSGSSYAALDHFCN<br/>                 KDDLKFDIYVNAGQKPILQRNIVMAKLF GPDNILSEVMEKVT<br/>                 ESAIREYYDY LKKVSGYRVRGKCSTEKEQEDLLKFQRLKNA<br/>                 VEFRDVTEYAEVINELLGQLISWSYL RERDILLYFQLGFHYMC<br/>                 LKNKSFKPAEYVDIRRNGTIIHNAILYQIVSMYINGLDFYSC<br/>                 DKEGKTLKPIETGKG VSKIGQFIKYSQYLYNDPSYKLEIYNA<br/>                 GLEVFENIDEHDNITDLRKYVDHFKYYAYGNKMSLLDLYSEF<br/>                 FDRFFTYDMKYQKNVVNLENILLRHFVIFYPKFGSGKKDV<br/>                 GIRDCKKERAQIEISEQSLTSEDFMFKLDDKAGEEAKKFPARD<br/>                 ERYLQTIAKLLYYPNEIEDMNRFMKKGETINKKVQFNRRKKI<br/>                 TRKQKNSSNEVLSSTMGYL FKNIKL</p> |
| C2-24 | <p>聚集绿屈<br/>                 扰菌<br/>                 (Chloroflex<br/>                 us<br/>                 aggregans)(<br/>                 SEQ ID<br/>                 No. 215)</p> | <p>MTDQVRREEVAAGELADTPLAAAQTPAADA AVAATPAPAE<br/>                 AVAPTPEQAVDQPATTGESEAPVTTA QAAAHEAEPAEATGA<br/>                 SFTPVSEQQPQKPRRLKDLQPGMELEGK VTSIALYGFVDVG<br/>                 VGRDGLVHISEMSDRRIDTPSELVQIGD TVKVWVKSVDLDA<br/>                 RRISLTMLNPSRGEKPRRSRQSQAQPQRRQEV DREKLASL<br/>                 KVGEIVEGVITGFAPFGAFADIGV GKDGLIHISELSEGRVEKPE<br/>                 DAVKVGERYQFKVLEIDGEGTRISLSL RRAQRTQRMQQLEPG<br/>                 QIEGTVSGIATFGAFVDIGVGRDGLVHISALAPHRVAKVEDV<br/>                 VKVGDKVKVVLGVDPQSKRISL TMRLEEEQPATTAGDEAA<br/>                 EPAEEVTPTRRGNLERFAAAAQTARERSERGERSERGERRER<br/>                 RERRPAQSSPDYIVGEDDDESFEGNATIEDLLTKFGGSSRR<br/>                 DRDRRRRHEDDDDEEMERPSNRRQREAIRRTLQQIGYDE</p> |   |
| C2-25 | <p>桔红色脱<br/>                 醌菌<br/>                 (Demequin<br/>                 a<br/>                 aurantiaca)(<br/>                 SEQ ID<br/>                 No. 216)</p>   | <p>MDLTWHALLILFIVALLAGFLDTLAGGGGLLTVPALLLTGIPP<br/>                 LQALGTNKLQSSFGTGMATYQVIRKKRVHWRDVRWPMVW<br/>                 AFLGSAAGAVAVQFIDTDALLIIPVVLALVAAYFLFVPKSHL<br/>                 PPPEPRMSDPAYEATLVPIIGAYDGA FGPGTGSLYALSGVALR<br/>                 AKTLVQSTAIKTLNFATNFAALLVF AFAGHMLWTVGAVMI<br/>                 AGQLIGAYAGSHMLFRVNPLVLRVLIVVMSLGMLIRVLLD</p>   |   |
| C2-26 | <p>海旋菌<br/>                 (Thalassosp<br/>                 ira sp.)<br/>                 TSL5-1<br/>                 (SEQ ID</p>  | <p>MRIIKPYGRSHVEGVATQEP RRKLRLNSSPDISRDIPGFAQSH<br/>                 DALIAQWISAIDKIATKPKPKKPTQAQINLRTTLGDAAWQH<br/>                 VMAENLLPAATDPAIREKLHLIWQSKIAPWG TARPQAEKDG<br/>                 KPTPKGGWYERFCGVLSPEAITQNVARQIAKDIYDHLHVA<br/>                 A KRKGREPAKQGESSNPKGKFKPDRKRGLIEERAESIAKNALR</p>   |   |

[0612]

|              |  |   |
|--------------|--|---|
|              | <p>No. 217)</p>  | <p>PGSHAPCPWGPDDQATYEQAGDVAGQIYAAARDCLEEKKR<br/>                 RSGNRNTSSVQYLPRDLAAKILYAQYGRVFGPDTTIKAALDE<br/>                 QPSLFALHKAIKDCYHRLINDARKRDILRILPRNMAALFRLVR<br/>                 AQYDNRDINALIRLGKVIHYHASEQGKSEHHGIRDYWPSQQ<br/>                 DIQNSRFWGSQADIKRHEAFSRIWRHIALASRTLHDWAD<br/>                 PHSQKFSGENDDILLAKDAIEDDVFKAGHYERKCDVLFGAQ<br/>                 ASLFCGAEDFEKAILKQAITGTGNLRNATFHFKGKVRFEKEL<br/>                 QELTKDVPVEVQSAIAALWQKDAEGRTRQIAETLQAVLAGH<br/>                 FLTEEQNRHIFAALTAAMAQPGDVPLPRLRRVLARHDSICQR<br/>                 GRILPLSPCPDRAKLEESPALTCQYTVLKMLYDGPFWLAQ<br/>                 QNSTILNHYIDSTIARTDKAARDMNGRKLQAQEKDLITSRAA<br/>                 DLPRLSVDEKMGDFLARLTAATATEMRVQRGYQSDGENAQ<br/>                 KQAAFIGQFECDVIGRAFADFLNQSGFDFVLKADTPQDA<br/>                 AQCDVTALIAPDDISVSPQAWQQVLYFILHLPVDDASHLL<br/>                 HQIRKWQVLEGKEKPAQIAHDVQSVLMLYLDMDAKFTGG<br/>                 AALHGIEKFAEFFAHAADFRAVFPQSLQDQDRSIPRRGLREI<br/>                 VRFGLPLLQHMSGTVQITHDNVVAWQAARTAGATGMSPI<br/>                 ARRQKQREELHALAVERTARFRNADLQNYMHALVDVIKHR<br/>                 QLSAQVTLSDQVRLHRLMMGVLGRLVDYAGLWERDLYFVV<br/>                 LALLYHHGATPDDVFKGQKKNLADGQVVAALKPKNRKAA<br/>                 APVGVFDDLHDYGIYQDDRQSIRNGLSHFNMLRGGKAPDLS<br/>                 HWVNQTRSLVAHDRKLNVAKSVIEMLAAREGFDDWGIQ<br/>                 TDRGQHILSHGKIRTRQAQHFQKSRLHIVKKSAPKDNNTVK<br/>                 IRENLHGDAMVERVVQLFAAQVQKRYDITVEKRLDHLFLKP<br/>                 QDQKGNIGIHTHNGWSKTEKKRRPSRENKGNHEN</p> |
| <p>C2-27</p> | <p>SAMN044<br/>                 87830_139<br/>                 20 [假丁酸<br/>                 弧 菌<br/>                 (Pseudobut<br/>                 yrivibrio<br/>                 sp.)OR37]<br/>                 (SEQ ID<br/>                 No. 218)</p> | <p>MKFSKESHRTAVGVTESNGIIGLLYKDPLEKEKIEDVVNQ<br/>                 RANSTKRLFNLFGTEATSKDISRASKDLAKVVNKAIGNLKGN<br/>                 KKFNKKEQITKGLNTKIIVEELKNVLKDEKKLIVNKDIIDEAC<br/>                 SRLKTSFRTAKTKQAVKMILTAVLIENTNLSKEDEAFVHEY<br/>                 FVKKLVNEYNKTSVKKQIPVALSNQNMVIQPNVNGTLEISE<br/>                 TKKSKETKTTEKDAFRAFLRDYATLDENRRHKMRLCLRNLV<br/>                 NLYFYGETSVSKDDDFDEWRDHEDKKQNDLQVKKIVSIKTD<br/>                 RKGNVKEVLDVDAITDAIRTNNIACYRRALAYANENPDVFFS<br/>                 DTMLNKFVIHVENEVERIYGHINNNTGDYKYQLGYLSEKV<br/>                 WKGIIYLSIKYIAEGKAVVNYAMNALAKDNNNSNAFGKLE<br/>                 KRVNGITSFEYERIKAEETLQRECAVNIAFAANHLANATVDL<br/>                 NEKDSDFLLKHEKDKDTLGAVARPNILRNILQFFGGKSRWN<br/>                 DFDFSGIDEIQLLDDLKMIYSLRNSSFHFKTENIDNDSWNTK<br/>                 LIGDMFAYDFNMAGNVQKDKMYSNNVPMFYSTSDIEKMLD<br/>                 RLYAEVHERASQVPSFNSVFRKNFPDYLNKDLKITSAFGVD<br/>                 DALKWQSAVYVCKEYIYNDLQNPETFTMLKDYVQCLPID<br/>                 IDKSMDQKLKSERNAHKNFKEAFATYCKECSLSAICQMIM<br/>                 TEYNNQNKGNRKVISARTKDGDKLIYKHYKMILFEALKNVF<br/>                 TIYLEKNINTYGFLKPKLINNVAIEEFLPNYNGRQYETLVN<br/>                 RITEETELQKWYIVGRLLNPKQVNQLIGNFRSYVQYVNDVAR<br/>                 RAKQTGNNLSNDNIAWDVKNIIQIFDVCTKLNQVTSNILEDY<br/>                 FDDGDDYARYLKNFVDYTNKNNDHSATLLGDFCAKEIDGIKI<br/>                 GIYHDGTNPVNRNIIQCKLYGATGIISDLTKDGSILSVDYEIIK</p>   |

[0613]

|       |  |  |
|-------|--|--|
|       |  | <p>KYMQMQKEIKVYQQKGICKTKEEQQNLKKYQELKNIVELRN<br/>IIDYSEILDELQGQLINWGYLRERDLMYFQLGFHYLCLHNESK<br/>KPVGYNNAGDISGAVLYQIVAMYTNGLSLIDANGKSKKNAK<br/>ASAGAKVGSFCSYSKEIRGVDKDTKEDDDPIYLAGVELFENI<br/>NEHQQCINLRNYIEHFHYAKHDRSMLDLYSEVFDRFFTYD<br/>MKYTKNVPNMMYNILLQHLVVPAFEFGSSEKRLDDNDEQTK<br/>PRAMFTLREKNGLSSEQFTYRLGDGNSTVKLSARGDDYLRA<br/>VASLLYYPDRAPEGLIRDAAEADKFAKINHSNPKSDNRNRRG<br/>NFKNPKVQWYNNKTKRK</p>  |
| C2-28 | <p>SAMN029<br/>10398_000<br/>08 [丁酸弧<br/>菌<br/>(Butyrivibri<br/>o<br/>sp.)YAB30<br/>01] (SEQ<br/>ID No.<br/>219)</p> | <p>MKISKVDHRKTAVKITDNKGAEGFIYQDPTRDSSTMEQHSNR<br/>ARSSKVLNFNIFGDTKKSDDLKNTESLIIVNKAIKSLKGDKR<br/>NNKYEEITESLKTERRVNLALIQAGNEFTCSENNIEDALNKYLK<br/>KSFRVGNTKSALKKLLMAAYCGYKLSIEEKEEQNYFVDKLV<br/>KEYNKDITVLKYTAKSLKHQNMVVQPDTDNHVFLPSRIAGAT<br/>QNKMSEKEALTEFLKAYAVLDEEKRHNLRIILRKLVLNLYFYE<br/>SPDFIYPENNEWKEHDDRKNKTETVSPVKVNEEKNGKTFV<br/>KIDVPATKDLIRLKNIECYRRSVAETAGNPITYFTDHNISKFWI<br/>HHIENEVEKIFALLKSNWKDYQFSVGYISEKVVKEIINYSIK<br/>YIAIGKAVYNYALEDIKKNDGTLNFGVIDPSFYDGINSEFEYK<br/>IKAEETFQREVAVYVSFAVNLSSATVKLSEAQSDMLVLNK<br/>NDIEKIAYGNTKRNLQFFGGQSKWKEFDLDRYINPVNYTDID<br/>FLFDIKMVYSLRNESFHFTTTDTESDWNKNLISAMFEYECR<br/>RISTVQKNKFFSNLPLFYGENSLERVLHKLDDYVDRMSQ<br/>VPSFGNVFVRKKFPDYMKEIGIKHNLSSDNLKLQALYFLY<br/>KEIYNAFISSEKAMKIFVDLVNKLDTNARDKGRITHEAMA<br/>HKNFKDAISHYMTDCSLADICQKIMTEYNQONTGHRKKQT<br/>TYSSEKNPEIFRHYKMILFMLLQKAMTEYISSEEIFDFIMKPN<br/>PKTDIKEEEFLPQYKSCAYDNLIKLIADNVELQKWYITARLLS<br/>PREVNQLIGSFRSYKQFVSDIERRAKETNNSLSKSGMTVDVE<br/>NITKVLDLCTKLNFRSNELTDFYFDSKDDYAVYVSKFLDFGF<br/>KIDEKFPALLGEFCNKEENGKIGIYHNGTEPILNSNIKSKL<br/>YGITDVVSRAVKPVSEKLIREYLQQEVKIKPYLENGVCKNKE<br/>EQAALRKYQELKNRIEFRDIVEYSEIINELMGQLINFSYLRERD<br/>LMYFQLGFHYLCLNNYGAKPEGYYISVNDKRTIKGAILYQIV<br/>AMTYTYGLPIYHYVDGTISDRRKNKKTVDLTLNSETVGA<br/>YFIYYSDELFNDSLILYNAGLELFENINEHENIVNLRKYIDHFK<br/>YYVSQDRSLDIYSEVFDRYFTYDRKYKKNVMNLFNIMLK<br/>HFIITDFEFSTGEKTIGEKNATAKKECAKVRIRKGGSSDKFTY<br/>KFKDAKPIELSAKNTFLDGVARILYYPENVVLTDLVRNSEV<br/>EDEKRIEKYDRNHNSPTRKDKTYKQDVKKNYNKKTSKAFD<br/>SSKLDTKSVGNLSDNPVLKQFLSESKKR</p> |
| C2-29 | <p>布劳特氏<br/>菌 马 赛<br/>-P2398<br/>(SEQ ID<br/>No. 220)</p>  | <p>MKISKVDHVKSGIDQKLSSQRGMLYKQPQKKYEGKQLEEHV<br/>RNLSRKAKALYQVFPVSGNSKMEKELQIINSFIKNILLRLDSG<br/>KTSEEIVGYINTYSVASQISGDHIQELVDQHLKESLRKYTCVG<br/>DKRIYVPDIIALLKSKFNSETLQYDNSELKILIDFIREDYLKE<br/>KQIKQIVHSIENNSTPLRIAENGQKRLIPANVDNPKKSYIFEFL<br/>KEYAQSDPKGQESLLQHMRYLILLYLYGPKITDDYCEEIEA</p>  |

[0614]

|              |  |   |
|--------------|--|---|
|              |  | <p>WNFGSIVMDNEQLFSEEASMLIQDRIYVNQQIEEGRQSKDTA<br/>                 KVKKNKSKYRMLGDKIEHSINESVVKHYQEACKAVEEKDIP<br/>                 WIKYISDHVMSVYSSKNRVDLTKLSLPYLAKNTWNTWISFIA<br/>                 MKYVDMGKGVYHFAMSDVDKVGKQDNLIIGQIDPKFSDGIS<br/>                 SFDYERIKAEDDLHRMSGYIAFAVNNFARAICSDFRKKNR<br/>                 KEDVLTVGLDEIPLYDNVVKRLLQYFGGASNWDDSIIDIIDDK<br/>                 DLVACIKENLYVARNVNFHFAGSEKVQKKQDDILEEIVRKET<br/>                 RDIGKHVRKVFYSNNVAVFYCDEDIKLMNHLQYQREKPYQA<br/>                 QIPSYNKVISKTYLPDLIFMLLKGNRTKISDPSIMNMFRTGTF<br/>                 YFLLKEIYYNDFLQASNLKEMFCEGLKNNVKNKKSEKPYQN<br/>                 FMRRFEELNMGMDFGEICQQIMTDYEQQNKQKKKTATAV<br/>                 MSEKDKKIRTLNDNTQKYKHFRLLYIGLREAFIYKDEKN<br/>                 KEWYEFREPVKREQPEEKEFVNKWKLNQYSDCSELILKDSL<br/>                 AAAYVVAHFINQAQLNHLIGDIKNYIQFISDIDRRAKSTGNP<br/>                 VSESTEIQIERYRKILRVLEFAKFFCGQITNVLTDDYQDENDFS<br/>                 THVGHYVKFEKKNMEPAHALQAFSNSLYACGKEKKKAGFY<br/>                 YDGMNPVNRNITLASYGNKKLLENAMNPVTEQDIRKYYS<br/>                 LMAELDSVLKNGAVCKSEDEQKNLRHFQNLKNRIELVDVLT<br/>                 LSELVNDLVAQLIGWVYIRERDMMYLQLGLHYIKLYFTDSV<br/>                 AEDSYLRTLDEEGSIADGAVLYQIASLYSFNLPMPYVKPNKS<br/>                 SVYCKKHVNSVATKFDIFEKEYCNGDETVIENGLRFLFENINL<br/>                 HKDMVKFRDYLAHFKYFAKLDESILELYSKAYDFFFSYNIKL<br/>                 KKSVSYVLTNVLLSYFINAKLSFSTYKSSGNKTVQHRRTKISV<br/>                 VAQTDYFTYKLSIVKNKNGVESIENDDRRCEVVNIAARDKE<br/>                 FVDEVCNVINYNSDK</p> |
| <p>C2-30</p> | <p>纤毛菌马<br/>                 赛 -P3007<br/>                 (SEQ ID<br/>                 No. 221)</p> | <p>MKITKIDGISHKKYIKEGKLVKSTSEENKTDERLSELLTIRLDT<br/>                 YIKNPDNASEEENRIRRENLEFFSNKVLYLKDGIYKDRRE<br/>                 KNQLQNKNYSEEDISEYDLKNKNNFLVLKILLNEDINSEELE<br/>                 IFRNDFEKKLDKINSLSKYSLEENKANYQKINENNIKKVEGKSK<br/>                 RNIFYNYKDSAKRNDYINNIQEAFDKLYKKEDIENLFFLIEN<br/>                 SKKHEKYKIRECYHKIIGRKNDKENFATIIYEEIQNVNMMKEL<br/>                 IEKVPNVSELKKSQVFYKYLLNKEKLNDENIKYVFCHFVEIE<br/>                 MSKLLKNYVYKPSNISNDKVKRIFEYQSLKLIENKLLNKL<br/>                 DTYYVRNCGKYSFYLDQGEIATSDFIVGNRQNEAFLRNIIIGVSS<br/>                 TAYFSLRNILETENENDITGRMRGKTVKNNKGEEKYISGEIDK<br/>                 LYDNNKQNEVKKNLKMFYSYDFNMNSKKEIEDFFSNIDEAIS<br/>                 SIRHGIVHFNLELEGKDIFTFKNIVPSQISKKMFHDEINEKLLK<br/>                 LKIFKQLNSANVFRYLEKYKILNYLNRTRFEFVNKNIPFVPSF<br/>                 TKLYSRIDDLKNSLGIYWKTPKTNDDNKTKETDAQIYLLKNI<br/>                 YYGEFLNYFMSNNGNFFEITKEIHELKNDKRNKLTGFYKIQ<br/>                 KFENLQEKTPKEYLANIQSLYMINAGNQDEEEKDTYIDFIQKI<br/>                 FLKGFMTYLANNGRLSLIYIGSDEETNTSLAEKKQEFDKFLK<br/>                 KYEQNNNIEIPYEINEFVREIKLGKILKYTERLNMFYLLKLLN<br/>                 HKELTNLKSLEKYQSANKEEAFSDQLELINLLNLDNRRVTE<br/>                 DFELEADEIGKFLDFNGNKVKDNKELKFDNKNKIYFDGENIHK<br/>                 HRAFYNIKKYGMLNLEKISDEAKYKISIEELKNYSKKKNEIE<br/>                 ENHTTQENLHRKYARPRKDEKFTDEDYKKEYKAIRNIQQYT<br/>                 HLKNKVEFNELNLLQSLLLRILHRLVGYTSIWERDLRFRLKGE</p>  |

[0615]

|       |   |  |
|-------|---|--|
|       |   | <p>FPENQYIEEIFNFDNSKNVKYKNGQIVEKYINFYKELYKDDTE<br/>KISIYSDKKVKELKKEKDLIRNYIAHFNYIPNAEISLLEMLE<br/>NLRKLLSYDRKLNKAIMKSIVDILKEYGFVVFVKIEKDKKIRIE<br/>SLKSEEVVHLKLLKLDNDKKKEPIKTYRNSKELCKLVKVM<br/>FEYKMKEKksen</p>   |
| C2-31 | <p>爱华拟杆菌<br/>(Bacteroides<br/>s ihuae)<br/>(SEQ ID<br/>No. 222)</p>   | <p>MRITKVKVKESSDQKDKMVLHHRKVGEGTLVLDENLADLTA<br/>PIIDKYKDKSFEKSLKQTLVSEKEMNIPKCDKCTAKERCLSC<br/>KQREKRLKEVRGAIKTIAGVIAGRDIIIPRLNIFNEDEICWLIK<br/>KLRNEFTFKDVNKQVVKLNLPKVLVEYSKKNPDLFLAYQQ<br/>WIAAYLKNKKGHKKSILNNRVVIDYSDESKLSKRKQALELW<br/>GEEYETNQRIALESYHTSYNIGELVTLLPNPEEYVSDKGEIRP<br/>AFHYKLNVLQMHQSTVFGTNEILCINPINFENRANIQLSAYN<br/>LEVVKYFEHYFPIKKNLNLQAIYYLKVETLKERLSLQL<br/>ENALRMNLLQKGIKKEHFDKNTCSNTLSQIKRDEFFVLNLV<br/>EMCAFAANNIRNIVDKEQVNEILSKKDLCSLSKNTIDKELCT<br/>KFYGADEFSQIPVAIWAMRGSVQQRNEIVHYKAEIDKIFALK<br/>TFEYDDMEKDYSDTPFKQYLELSIEKIDSFIEQLSSNDVLNY<br/>YCTEDVNKLLNKCKLSLRRTSIPFAPGFKTIYELGCHLQDSSN<br/>TYRIGHYLMIGGRVANSTVTKASKAYPAYRFMLKLIYNHLF<br/>LNKFLDNHNKRFFMKAVAFVLKDNRENARNKFQYAFKEIR<br/>MMNNDESIASYMSYIHSLSVQEQEKKGDKNDKVRYNTEKFI<br/>EKVVFVKGFFDLSWLGVEFILSPNQEERDKTVTREEYENLMI<br/>KDRVEHSINSNQESHIAFFTFCKLLDANHLSDLRNEWIKFRSS<br/>GDKEGFSYNFAIDIIELCLLVDRVEQRRDGYKEQTELKEYLS<br/>FFIKGNESENTVWKGIFYFQQDNYTPVLYSPIELIRKYGTLELL<br/>KLIIVDEDKITQGEFEEWQTLKVVVEDKVTRRNELHQEWED<br/>MKNKSSFSQEKCSIYQKLCRDIDRYNWLDNKLHLVHLRKLH<br/>NLVIQILSRMARFIALWDRDFVLLDASRANDDYKLLSFFNFR<br/>DFINAKKTKTDELLAEFGSKIEKKNAPFIKAEDVPLMVECE<br/>AKRSFYQKVFFRNQLVADRNFIAHNYISKTAKCSLFEMII<br/>KLRTLMYYDRKLRNAVVKSIANVFDQNGMVLQLSLDDSHL<br/>KVDKVISKRIVHLKNNNIMTDQVPEEYKICRRLLEMKK</p> |
| C2-32 | <p>SAMN052<br/>16357_104<br/>5 [紫单孢<br/>菌科细菌<br/>(Porphyro<br/>monadacea<br/>e<br/>bacterium)<br/>KH3CP3R<br/>A] (SEQ<br/>ID No.<br/>223)</p> | <p>MEFRDSIFKSLQKEIEKAPLCFAEKLISGGVFSYYPSEKLF<br/>VGNHPFSLFRKTMFSPGFKRVMKSGGNYQANRDGRFYDL<br/>DIGVYLPKDGFGDEEWNARYFLMKLIYNQLFLPYFADAENH<br/>LFRECVDFVKRVNRDYNCKNNNSEEQAFIDIRSMREDESIAD<br/>YLAFIQSNIIIENKKEKTNKEGQINFNKFLLQVVFVKGFDLFLK<br/>DRTELNQLPELQGDGTRGDDLESKLGAVVAVDLKLDA<br/>TGIDADLNENISFYTFCKLLDSNHLRRLNEIIKYQSANSDFSH<br/>NEDFDYDRIISIELCMLSADHVSTNDNESIFPNNDKDFSGIRP<br/>YLSTDAKVETFDLYVHSDAKTPITNATMVLNWKYGTDKLF<br/>ERLMISDQDFLVTEKDYFVWVKELKDIIEKIKLREELHSLWV<br/>NTPKGKKGAKKNGRETTGEFSEENKKEYLEVCREIDRYVN<br/>LDNKLHFVHLKRMHSLIELLGRFVGFYLFERDYQYYHLEI<br/>RSRRNKDAGVVDKLEYNKIKDQNKYDKDDFFACTFLYEKA<br/>NKVRNFIAHFNYLTMWNSPQEEHNSNLGAKNSSGRQNLK<br/>CSLTELINELREVMSYDRKLNVAVTKAVIDLFDKHGMVIFR</p>  |

|       |  |  |   |
|-------|--|--|---|
|       |  |  | IVNNNNNDNKNKHHLELDDIVPKKIMHLRGIKLRQDGKPIPI<br>QTDSVDPLYCRMWKKLLDLKPTPF  |
| C2-33 | 河岸李斯特菌<br>( <i>Listeria riparia</i> )(SEQ ID No. 224)              |  | MHDAWAENPKKPQSDAFLKEYKACCEAIDTYNWHKNKATL<br>VYVNELHLLDILGRLVGYVAIADRDFQCMANQYLKSSGH<br>TERVDSWINTIRKRPDYIEKLDIFMKNKAGLFVSEKNGRNYIA<br>HLNYLSPKHKYSLLYLFEKLRMLKYDRKLNNAVTKSLIDLL<br>DKHGMCVVFANLKNKHLVIAASLKPKKIETFKWKKIK  |
| C2-34 | 陌生非适应螺菌<br>( <i>Insolitispirillum peregrinum</i> )(SEQ ID No. 225) |  | MRIIRPYGSSTVASPSPQDAQPLRSLQRQNGTFDVAEFSRRHP<br>ELVLAQWVAMLDKIIRKPAPGKNSTALPRPTAEQRRLRQQV<br>GAALWAEMQRHTPVPELKA VWDSKVHPYSKDNAPATAKT<br>PSHRGRWYDRFGDPETSAAVAEGVRRHLLDSAQPFRANGG<br>QPKGKGVIEHRALTIQNGTLLHHHQSEKAGPLPEDWSTYRA<br>DELVSTIGKDARWIKVAASLYQHYGRIFGPTTPISEAQTRPEF<br>VLHTAVKAYYRRLFKERKLPALERLERLLPRTGEALRHAVTV<br>QHGNRSLADAVRIGKILHYGWLQNGEPDPWPDDAALYSSRY<br>WGSQDQTDIKHSEAVSRVWRRALTAQRTLTSLWYPAGTD<br>AGDILLIGQKPDSDRNRLPLLYGDSTRHWTRSPGDVWFLK<br>QTLENLRNSSFHFKTLSAFTSHLDGTCESEPAEQQAALWQ<br>DDRQQDHQQVFLSLRALDATTYLP TGPLHRVNAVQSTDAT<br>LPLPRFRRVVTRAANTRLKGFVPEPVNRRRTMEDDPLLRCRY<br>GVLKLLYERGFRAWLETRPSIASCLDQSLKRSTKAAQTINGK<br>NSPQGV EILSRATKLLQAEGGGHGHIDLFDRLYAATAREM<br>RVQVGYHHDAAEARQQAEFIEDLKCEVVARAFCAYLKTLGI<br>QGDTRRQPEPLPTWPDLPDLPSSSTIGTAQAALYSVLHMPV<br>EDVGSLLHQLRRWLVALQARGGEDGTAITATIPLLELYLNRH<br>DAKFSGGGAGTGLRWDDWQVFFDCQATFDRVFPPGPALDS<br>HRLPLRGLREVLRFGRVNDLAALIGQDKITAAEVDRWHTAE<br>QTIAAQQQRREALHEQLSRKKGTDAEVDEYRALVTAIADHR<br>HLTAHVTLSNVVRHLRMLTTVLGRLVDYGGGLWERDLTFVTL<br>YEAHRLGGLRNLLSESRVNFVLDGQTPAALSCKNNAEENGM<br>ISKVLGDKARRQIRNDFAHFNMLQQGKKTINLTDEINNARKL<br>MAHDRKLNKNAITRSVTLLQQDGLDIVWTMDASHRLTDAKI<br>DSRNAIHLHKTHNRA NIREPLHGKSYCRWVAALFGATSTPSA<br>TKKSDKIR |

[0616]

[0617] 在某些示例实施方式中,靶向RNA的效应子蛋白是如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中公开的Cas13c效应子蛋白。下表4中提供了Cas13c 的示例野生型直系同源序列。在某些示例实施方式中,所述CRISPR效应子蛋白是来自表3或表4的Cas13c蛋白。

[0618] 表3

|        |  |  |  |
|--------|--|--|--|
| [0619] | 坏死梭杆菌<br>( <i>Fusobacterium necrophorum</i> )f |  | MEKFRRQNRNSIIKIIISNYDTKGIKELKVRYRKQAQLDTFIIKTEI<br>VNNDIFIKSIIEKAREKYRYSFLFDGEEKYHFKNKSSVEIVKKDIF<br>SQTPDNMIRNYKITLKISEKNPRVVEAEIEDLMNSTILKDGRRSA |
|--------|--|--|--|

[0620]

|  |  |
|--|--|
| <p>unduliforme 亚种 ATCC 51357 叠连群 (contig)00003 (SEQ ID No. 226)</p>      | <p>RREKSMTERKLIIEEKVAKNYSLLANCPMEEVDSIKIYKIKRFLTY RSNMLLYFASINSFLCEGIKGDNETEEIWHLKDNDVRKEKVVRE NFKNKLIQSTENYNSSLKNQIEEKEKLLRKEFKKGAFYRTIHKLL QQERIKELSEKSLTEDCEKIIKLYSKLRHSLMHYDYQYFENLFEN KKNDDLMDLNDLDFKSLPLIRKMKLNKNVNYLEDGDTLFLVQ KTKKAKTLYQIYDALCEQKNGFNKFINDDFFVSDGEENTVFKQIIN EKFAQSEMEFLEKRRISESEKKNEKLLKLDKSMKAHFRNINSEDTK EAYFWDIHSSRNYKTKYNERKNLVNEYTELLGSSKEKLLREEI TKINRQLLKLKQEMEEITKKNLSFRLEYKMKIAFGFLFCFDGNI SKFKDEFDASNQEKIIQYHKNGEKYLTSFLKEEEKEKFNLEKMQ KIIQKTEEDWLLPETKNNLKFYLLTYLLLPYELKGDFLGFVKK HYYDIKNVDFIDENQNNIQVSQTVKQEDYFYHKIRLFEKNTKK YEIVKYSIVPNEKLLQYFEDLGIDIKYLTVEQKSEVSEEKNNKVS LKNNGMFNKTILLFVFKYYQIAFKLFNDIELYSLFFLREKSGKPL EIFRKELESKMKDGYLNFQQLLYVVYEVLVKNKDLKILSKKID YRKDKSFSPEIAYLRNFLSHLNYSKFLDNFMKINTNKS DENKEV LIPSIKIQKMIQFIEKCNLQNDQDFDFNFVDFYMRKEKMFFIQLK QIFPDINSTEKQKMNEKEEILRNRYHLTDKKNEQIKDEHEAQS QL YEKILSLQKIYSSDKNNFYGRLKEEKLFLFEKQGGKLSMEEIKD KIAGDISDLLGILKKEITRDIKDKLTEKFRYCEEKLLNLSFYNHQD KKKEESIRVFLIRDKNSDNFKFESILDDGNSKIFISKNGKEITIQQC DKVLETLIEKNTLKISSNGKIISLIPHYSYSIDVKY</p>  |
| <p>坏死梭杆菌 DJ-2 叠连群 0065, 全基因组鸟枪法测序 (shotgun sequence)(SEQ ID No. 227)</p> | <p>MEKFRQRNRSSIIKIIISNYDTKGIKELKVRVYRKAQLDTEIITEI VNNNDIFIKSIIEKAREKYRYSFLFDGEEKYHFKNKSSVEIVKDDIF SQTPDNMIRNYKITLKIIEKNPRVVEAEIEDLMNSTILKDGRRSA RREKSMTERKLIIEEKVAENYSSLLANCPMEEVDSIKIYKIKRFLTY RSNMLLYFASINSFLCEGIKGDNETEEIWHLKDNDVRKEKVVRE NFKNKLIQSTENYNSSLKNQIEEKEKLLRKEFKKGAFYRTIHKLL QQERIKELSEKSLTEDCEKIIKLYSELRHPLMHYDYQYFENLFEN KENSELTKNLNLDIFKSLPLVRKMKLNKNVNYLEDNDTLFLVQ KTKKAKTLYQIYDALCEQKNGFNKFINDDFFVSDGEENTVFKQIIN EKFAQSEIEFLEKRRISESEKKNEKLLKLDKSMKAHFRNINSEDTKE AAYFWDIHSSRNYKTKYNERKNLVNEYTELLGSSKEKLLREEIT KINRQLLKLKQEMEEITKKNLSFRLEYKMKMAFGFLFCFDGNI SRFKDEFDASNQEKIIQYHKNGEKYLTYFLKEEEKEKFNLKKLQ ETIQKTGEENWLLPQNKNLKFYLLTYLLLPYELKGDFLGFVKK KHYYDIKNVDFMDENQSSKIIESKEDDFYHKIRLFEKNTKKYEIV KYSIVPDKKLLQYFKDLGIDTKYLILDQKSEVSGEKNKVS LKN NGMFNKTILLFVFKYYQIAFKLFNDIELYSLFFLREKSGKPFVFL KELKDKMIGKQLNFGQLLYVVYEVLVKNKDLSEILSERIDYRKD MCFSAEIAADLRNFLSHLNYSKFLDNFMKINTNKS DENKEV LIPSI KIQKMIKFIEECNLQSQIDDFDFNFVDFYMRKEKMFFIQLKQIFPD INSTEKQKMNEKEEILRNRYHLTDKKNEQIKDEHEAQS QLYEKI LSLQKIYSSDKNNFYGRLKEEKLFLFEKQEKKLSMEEIKDKIAG DISDLLGILKKEITRDIKDKLTEKFRYCEEKLLNLSFYNHQDKKK EESIRVFLIRDKNSDNFKFESILDDGNSKIFISKNGKEITIQQCDKV LETLIEKNTLKISSNGKIISLIPHYSYSIDVKY</p> |
| <p>坏死梭杆菌 BFTR-1 叠连</p>   | <p>MKVRYRKAQLDTEIITEI VNNNDIFIKSIIEKAREKYRYSFLFDG EEKYHFKNKSSVEIVKNDIFSQTPDNMIRNYKITLKIIEKNPRVV EAEIEDLMNSTILKDGRRSARREKSMTERKLIIEEKVAENYSLLA</p>  |

[0621]

|  |  |
|--|--|
| <p>群 0068 (SEQ ID No. 228)</p>   | <p>NCPIEEVDSIKIYKIKRFLTYRSNMLLYFASINSFLCEGIKGDNE<br/> TEEIWHLKDNDVRKEKVKENFNKLIQSTENYNSSLKNQIEEKE<br/> KLSSKEFKKGAFYRTIIKKLQQERIKELSEKSLTEDCEKIIKLYSEL<br/> RHPLMHYDYQYFENLFENKENSELTKNLNLDIFKSLPLVRKMKL<br/> NNKVNYLEDNDTLFVLQKTKKAKTLYQIYDALCEQKNGFNKFI<br/> NDFVSDGEENTVFKQIINEKFQSEMEFLEKRRISESEKKNEKLLK<br/> KLDSMKAHFRNINSEDTKEAYFWDIHSSRNYKTKYNERKNLVN<br/> EYTKLLGSSKEKKLLREEITKINRQLLKLKQEMEEITKKNLSFRL<br/> EYKMKIAFGFLFCEFDGNISFKDEFDASNQEKIIQYHKNGEKYL<br/> TSFLKEEEKEKFNLEKMQKIIQKTEEEDWLLPETKNNLFKFYLLT<br/> YLLLPHYELKGDFLGFVKKHYDIDKNVDFMDENQNNIQVSQTVE<br/> KQEDYFYHKIRLFEKNTKKYEIVKYSIVPNEKLLKQYFEDLGIDIK<br/> YLTGVSVEGKWLGENLGIDIKYLTVEQKSEVSEEKKNKVSLLK<br/> NGMFNKITLLFVFKYYQIAFKLFNDIELYSLFFLREKSEKPFVFL<br/> EELKDKMIGKQLNFGQLLYVVEVLVKNKDLKILSKKIDYRK<br/> DKSFSPEIAYLRNFLSHLNYSKFLDNFMKINTNKSNDENKEVLPSI<br/> KIQKMIQFIEKCNLQNDQDFDFNFVNDFYMRKEKMFQILKQIFP<br/> DINSTEKQKSEKKEILRKRYHLINKKNEQIKDEHEAQSQLYEKI<br/> LSLQKIFSCDKNNFYRRLKEEKLLFLEKQGGKISMKEIKDKIAS<br/> DISDLLGILKKEITRDIKDKLTEKFRYCEEKLLNISFYNHQDKKKE<br/> EGIRVFLIRDKNSDNFKFESILDDGSNKIFISKNGKEITIQCCDKVL<br/> ETLMIEKNTLKISSNGKIISLIPHYSYSIDVKY</p> |
| <p>坏死梭杆菌<br/> <i>funduliforme</i><br/> 亚种 1_1_36S<br/> cont1.14 (SEQ ID No. 229)</p>   | <p>MTEKKSIIFFKNKSSVEIVKDDIFSQTPDNMIRNYKITLKISEKNPR<br/> VVEAEIEDLMNSTILKDGRRSARREKSMTERKLIIEKVAENYSL<br/> LANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINSFLCEGIKGD<br/> DNETEEIWHLKDNDVRKEKVKENFNKLIQSTENYNSSLKNQIE<br/> EKEKLLRKESKKGAFYRTIIKKLQQERIKELSEKSLTEDCEKIIK<br/> YSELRHPLMHYDYQYFENLFENKENSELTKNLNLDIFKSLPLVR<br/> KMKLNNKVNYLEDNDTLFVLQKTKKAKTLYQIYDALCEQKNG<br/> FNKFINDFVSDGEENTVFKQIINEKFQSEMEFLEKRRISESEKKNE<br/> KLKKKFDSMKAHFRNINSEDTKEAYFWDIHSSRNYKTKYNERK<br/> NLVNEYTELLGSSKEKKLLREEITQINRKLKQEMEEITKKNLS<br/> LFRLEYKMKIAFGFLFCEFDGNISFKDEFDASNQEKIIQYHKNG<br/> EKYLYTYFLKEEEKEKFNLEKMQKIIQKTEEEDWLLPETKNNLFK<br/> FYLLTYLLLPHYELKGDFLGFVKKHYDIDKNVDFMDENQNNIQVS<br/> QTVEKQEDYFYHKIRLFEKNTKKYEIVKYSIVPNEKLLKQYFEDL<br/> GIDIKYLTGVSVEGKWLGENLGIDIKYLTVEQKSEVSEEKIKKFL</p>  |
| <p>极臭梭杆菌<br/> (<i>Fusobacterium</i><br/> <i>perfoetens</i>)AT<br/> CC 29250<br/> T364DRAFT_s<br/> caffold00009.9<br/> _C (SEQ ID<br/> No. 230)</p> | <p>MGKPNRSSIIKIIISNYDNKGIVKVRYNKQAQLDTFLIKSELKD<br/> GKFILYSIVDKAREKYRYSFEIDKTNINKNEILIIKKDIYSNKEDK<br/> VIRKYILSFEVSEKNDRTIVTKIKDCLETQKKEKFERENTRRLISE<br/> TERKLLSEETQKTYSKIACCPEDIDSVKIYKIKRYLAYRSNMLL<br/> FFSLINDIFVKGVDNNGEEVGEIWRIDSKEIDEKKTYYDLLVENF<br/> KKRMSQEFINYKQSIENKIEKNTNKIKEIEQKLLKKEKYKKEINRL<br/> KKQLIELNRENDLLEKDKIELSDEEIREIDIEKILKIYSDLRHKL<br/> MHYNYQYFENLFENKISKEKNEDVNLTELLDLNLFYRPLVRQLK<br/> LENKTNYLEKEDKITVLGVSDSAIKYYSYNNFLCEQKNGFNFI<br/> NSFFSNDGEENKSFKEKINLSLEKEIEMEKETNEKIKEINKNELQ<br/> LMKEQKELGTAYVLDIHSNDYKISHNERNKNVKLQNDIMNGN<br/> RDKNALDKINKLVELKIKMDKITKRNSILRLKYKLQVAYGFLM</p>   |

[0622]

|   |   |
|---|---|
|   | <p>E E Y K G N I K K F K D E F D I S K E K I K S Y K S K G E K Y L E V K S E K K Y I T K I L<br/> N S I E D I H N I T W L K N Q E E N N L F K F Y V L T Y I L L P F E F R G D F L G F V K K<br/> H Y Y D I K N V E F L D E N N D R L T P E Q L E K M K N D S F F N K I R L F E K N S K K<br/> Y D I L K E S I L T S E R I G K Y F S L L N T G A K Y F E Y G G E E N R G I F N K N I I P I F<br/> K Y Y Q I V L K L Y N D V E L A M L L T L S E S D E K D I N K I K E L V T L K E K V S P<br/> K K I D Y E K K Y K F S V L L D C F N R I I N L G K K D F L A S E E V K E V A K T F T N L<br/> A Y L R N K I C H L N Y S K F I D D L L T I D T N K S T T D S E G K L L I N D R I R K L I K<br/> F I R E N N Q K M N I S I D Y N Y I N D Y Y M K K E K F I F G Q R K Q A K T I I D S G K K<br/> A N K R N K A E E L L K M Y R V K K E N I N L I Y E L S K K L N E L T K S E L F L L D K<br/> K L L K D I D F T D V K I K N K S F F E L K N D V K E V A N I K Q A L Q K H S S E L I G I<br/> Y K K E V I M A I K R S I V S K L I Y D E E K V L S I I I Y D K T N K K Y E D F L L E I R R E<br/> R D I N K F Q F L I D E K K E K L G Y E K I I E T K E K K K V V V K I Q N N S E L V S E P<br/> R I I K N K D K K K A K T P E E I S K L G I L D L T N H Y C F N L K I T L</p>  |
| <p>溃疡梭杆菌<br/>(<i>Fusobacterium<br/>ulcerans</i>)ATC<br/>C 49185<br/>cont2.38 (SEQ<br/>ID No. 231)</p>   | <p>M E N K G N N K K I D F D E N Y N I L V A Q I K E Y F T K E I E N Y N N R I D N I I D K K<br/> E L L K Y S E K K E E S E K N K K L E E L N K L K S Q K L I L T D E E I K A D V I K I I K<br/> I F S D L R H S L M H Y E Y K Y F E N L F E N K K N E E L A E L L N L N L F K N L T L L<br/> R Q M K I E N K T N Y L E G R E E F N I I G K N I K A K E V L G H Y N L L A E Q K N G F<br/> N N F I N S F F V Q D G T E N L E F K K L I D E H F V N A K K R L E R N I K K S K K L E K<br/> E L E K M E Q H Y Q R L N C A Y V W D I H T S T T Y K K L Y N K R K S L I E E Y N K Q<br/> I N E I K D K E V I T A I N V E L L R I K K E M E E I T K S N S L F R L K Y K M Q I A Y A F<br/> L E I E F G G N I A K F K D E F D C S K M E E V Q K Y L K K G V K Y L K Y Y K D K E A<br/> Q K N Y E F P F E E I F E N K D T H N E E W L E N T S E N N L F K F Y I L T Y L L L P M E<br/> F K G D F L G V V K K H Y Y D I K N V D F T D E S E K E L S Q V Q L D K M I G D S F F<br/> H K I R L F E K N T K R Y E I I K Y S I L T S D E I K R Y F R L L E L D V P Y F E Y E K G T<br/> D E I G I F N K N I I L T I F K Y Y Q I I F R L Y N D L E I H G L F N I S S D L D K I L R D L K S<br/> Y G N K N I N F R E F L Y V I K Q N N N S S T E E E Y R K I W E N L E A K Y L R L H L L<br/> T P E K E E I K T K T K E E L K L N E I S N L R N G I C H L N Y K E I I E I L K T E I S E K<br/> N K E A T L N E K I R K V I N F I K E N E L D K V E L G F N F I N D F F M K K E Q F M F G<br/> Q I K Q V K E G N S D S I T T E R E R K E K N N K L K E T Y E L N C D N L S E F Y E T<br/> S N N L R E R A N S S S L L E D S A F L K K I G L Y K V K N N K V N S K V K D E E K R I<br/> E N I K R K L L K D S S I M G M Y K A E V V K K L K E K L I L I F K H D E E K R I Y V T<br/> V Y D T S K A V P E N I S K E I L V K R N N S K E E Y F F E D N N K K Y V T E Y Y T L E I<br/> T E T N E L K V I P A K K L E G K E F K T E K N K E N K L M L N N H Y C F N V K I I Y</p> |
| <p>厌氧唾液杆菌<br/>(<i>Anaerosalibac<br/>ter</i> sp.)ND1 基<br/>因组组装玛氏<br/>厌氧杆菌<br/>(<i>Anaerosalibac<br/>ter</i><br/><i>massiliensis</i>)N<br/>D1 (SEQ ID<br/>No. 232)</p> | <p>M K S G R R E K A K S N K S S I V R V I I S N F D D K Q V K E I K V L Y T K Q G G I D V I<br/> K F K S T E K D E K G R M K F N F D C A Y N R L E E E E F N S F G G K G K Q S F F V T<br/> T N E D L T E L H V T K R H K T T G E I I K D Y T I Q G K Y T P I K Q D R T K V T V S I T<br/> D N K D H F D S N D L G D K I R L S R S L T Q Y T N R I L L D A D V M K N Y R E I V C S<br/> D S E K V D E T I N I D S Q E I Y K I N R F L S Y R S N M I I Y Y Q M I N N F L L H Y D G E<br/> E D K G G N D S I N L I N E I W K Y E N K K N D E K E K I I E R S Y S I E K S I N Q Y I L<br/> N H N T E V E S G D K E K K I D I S E E R I K E D L K K T F I L F S R L R H Y M V H Y N Y<br/> K F Y E N L Y S G K N F I I Y N K D K S K S R R F S E L L D L N I F K E L S K I K L V K N R<br/> A V S N Y L D K K T T I H V L N K N I N A I K L L D I Y R D I C E T K N G F N N F I N N M<br/> M T I S G E E D K E Y K E M V T K H F N E N M N K L S I Y L E N F K K H S D F K T N N<br/> K K K E T Y N L L K Q E L D E Q K K L R L W F N A P Y V Y D I H S S K K Y K E L Y V E<br/> R K K Y V D I H S K L I E A G I N N D N K K K L N E I N V K L C E L N T E M K E M T K L<br/> N S K Y R L Q Y K L Q L A F G F I L E E F N L D I D K F V S A F D K D N N L T I S K F M E<br/> K R E T Y L S K S L D R R D N R F K K L I K D Y K F R D T E D I F C S D R E N N L V K L<br/> Y I L M Y I L L P V E I R G D F L G F V K K N Y Y D L K H V D F I D K R N N D N K D T F<br/> F H D L R L F E K N V K R L E V T S Y S L S D G F L G K K S R E K F G K E L E K F I Y K<br/> N V S I A L P T N I D I K E F N K S L V L P M M K N Y Q I I F K L L N D I E I S A L F L I A K</p>  |

[0623]

|  |   |
|--|---|
|  | KEGNEGSITFKKVIDKVRKEDMNGNINFSQVMKMALNEKVNCQ<br>IRNSIAHINMKQLYIEPLNIYINNNQNKKTISEQMEEIIDICITKGLT<br>GKELNKNINDYYMKKEKLVFNLKLRRNNLVSIDAQQKNMKE<br>KSILNKYDLNYKDENLNKEIKLVNDLNNKQKLLKETTEGESNY<br>KNALSKDILLNGIIRKNINFKIKEMILGIIQQNEYRYVNINIIDKI<br>RKEDHNIDLKINNKYIEISCYENKSNESTDERINFKIKYMDLKVK<br>NELLVPSCYEDIYIKKKIDLEIRYIENCKVYIDYIYKKNINLEF<br>DGKTLFVKFNKDVKKNNQKVNLESNYIQNIKFIVS |
|--|---|

[0624] 表4

[0625]

| 名称                   | 序列  |
|----------------------|---|
| EHO1<br>9081         | MTEKKSIIFFKNKSSVEIVKKDIFSQTPDNMIRNYKITLKISEKNPRVVEAEI<br>EDLMNSTILKDGRRSARREKSMTERKLIIEKVAENYSLLANCPMEEVDSI<br>KIYKIKRFLTYRSNMLLYFASINSFLCEGIKGDNETEEIWHLKDNDVRKE<br>KVKENFKNLIQSTENYNSSLKNQIEEKEKLLRRESKKGAFYRTIHKLQO<br>ERIKELSEKSLTEDCEKIIKLYSELRHPMHYDYQYFENLFENKENSELTK<br>NLNLDIFKSLPLVRKMKLNKVNYLEDNDFVLFVQTKKAKTLYQIYDA<br>LCEQKNGFNKFINDFVSDGEENTVFKQIINEKFQSEMEFLEKRISESEKK<br>NEKLLKKFDSMKAHFHNINSEDTKEAYFWDIHSSSNYKTKYNERKNLV<br>NEYTELLGSSKEKLLREEITQINRKLKQEMEEITKKNLSFRLEYKM<br>KIAFGFLFCFDGNISKFKDEFDASNQEKIIQYHKNGEKYLTYFLKEEKE<br>KFNLEKMQKIIQKTEEDWLLPETKNLKFYLLTYLLPYELKGDFLGF<br>VKKHYYDIKNVDFMDENQNNIQVSQTVEKQEDYFYHKIRLFKNTKKY<br>EIVKYSIVPNEKQYFEDLGIDIKYLTGSVESGEKWLGENLGIDIKYLTV<br>EQKSEVSEEKIKKFL   |
| WP_0<br>94899<br>336 | MEKDKKGEKIDISQEMIEEDLRKILILFSRLRHSVMHYDYEFYQALYSGK<br>DFVISDKNNLENRMISQLLDLNFKELSKVLIKDKAISNYLDKNTTIHVL<br>GQDIKAIRLLDIYRDICGSKNGFNKFINTMITISGEEDREYKEKVEHFNKK<br>MENLSTYLEKLEKQDNAKRNNKRVYNLLKQKLEQQKLEWFGGPVYVY<br>DIHSSKRYKELYIERKLVDRHSLFEEGLDEKNKKELTKINDELSKLNSE<br>MKEMTKLNSKYRLQYKLQAFGFILEEFDLNDTFINNFDDKDKDLIISNFM<br>KKRDIYLNRLDRGDNRLKNIKEYKFRDTEIFCNDRDNNLVKLYILMY<br>ILLPVEIRGDFLGFVKKNYDMKHVDFIDKKDKEDKDTFFHDLRLFENI<br>RKLEITDYSLSSGFLSKEHKVDIEKKINDFINRNGAMKLPEDITIEEFNKSLI<br>LPIMKNYQINFKLLNDIEISALFKIAKDRSITFKQAIDEIKNEDIKKNSSKND<br>KNNHKDKNINFTQLMKRALHEKIPYKAGMYQIRNNISHIDMEQLYIDPLN<br>SYMNSNKNITISEQIEKIIDVCVTGGVTGKELNNNIINDYYMKKEKLVFN<br>LKLKQNDIVSIESQEKNKREEFVFKYGLDYKDGEINIEVIQKVNSLQE<br>ELRNKETSKEKLNKETLFRDISLINGTIRKNINFKIKEMVLDIVRMDEIR<br>HINIHIYYKGENYTRSNIKFKYAIDGENKKYYLKQHEINDINLELKDKFV<br>TLICNMDKHPNKNKQTINLESNYIQNVKFIIP |
| WP_0<br>40490<br>876 | MENKGNNKKIDFDENYNILVAQIKEYFTKEIENYNNRIDNIIDKKELLYS<br>EKKEESEKNKKLEELNKLKSQKILKILTDEEIKADVIKIIFSDLRHSLMHY<br>EYKYFENLFENKNEELAE LLNLFKNLTLLRQMKIENKTNYLEGREF<br>NIIGKNIAKEVLGHYNLLAEQKNGFNFFINSFFVQDGTENLEFKKLIDEH<br>FVNAKRLERNIKKSKKLEKELEKMEQHYQRLNCA YVWDIHTSTTYKK<br>LYNKRKSLIEEYNKQINEIKDKEVITAINVELLRIKKEMEEITKSNLSFRK   |

[0626]

|                      |  |
|----------------------|--|
|                      | <p>YKMQIAYAFLEIEFGGNIKFKDEFDCSKMEEVQKYLKKGVKYLKYYK<br/> DKEAQKNYEFPFEEIFENKDTHTNEEWLENTSENNLKFYILTYLLPMEF<br/> KGDFLGVVKKHYDDIKNVDFDSEKELSQVQLDKMIGDSFFHKIRLFE<br/> KNTKRYEIIKYSILTSDEIKRYFRLELDVPYFEYKGTDEIGIFNKNILTIF<br/> KYYQIIFRLYNDLEIHGLFNISDLDKILRDLKSYGNKNINFREFLYVIKQN<br/> NNSSTEEYRKIWENLEAKYLRLLHLLTPEKEEIKTKTKEELEKLNESNLR<br/> NGICHLNYKEIEEILKTEISEKNKEATLNEKIRKVINFIKENELDKVELGFN<br/> FINDFFMKKEQFMFGQIKQVKEGNSDSITTERERKEKNNKLLKETYELNC<br/> DNLSEFYETSNNLRERANSSSLEDSEFLKIGLYKVKNNKVNKSKVDEE<br/> KRIENIKRLLKDDSSDIMGMYKAEVVKKLKEKLILIFKHDEEKRIYVTVY<br/> DTSKAVPENISKEILVKRNNKSEYFFEDNNKKYVTEYITLEITETNELK<br/> VIPAKKLEGKEFKTEKNKENLMLNNHYCFNVKIIY</p>   |
| WP_0<br>47396<br>607 | <p>MEEIKHKKNKSSIIIRVIVSNYDMTGIKEIKVLYQKQGGVDTFNLKTIINLES<br/> GNLEIISCKPKEREKYRYFNCKTEINTISITKKDKVLKKEIRKYSLELYFK<br/> NEKKDTVVAKVTDLLKAPDKIEGERNHRLKLSSTERKLLSKTLCKNYS<br/> EISKTPIEEISIKIYKIKRFLNYRSNFLIYFALINDFLCAGVKEDDINEVWLI<br/> QDKEHTAFLENRIEKITDYIFDKLSKDIENTKKNQFEKRIKYYKTSLEELKT<br/> ETLEKNKTFYIDSIKTKITNLENKITELSLYNSKESLKEDLIKISIFTNLRHS<br/> LMHYDYKSFENLFENIENEELKNLLDLNLFKSIRMSDEFKTKNRTNYLDG<br/> TESFTIVKKHQNLKLYTYNNLDCDKKNGFNFTINSFFVTDGIENTDFKN<br/> LILHFEKEMEEYKKSIEYYKIKISNEKNKSKKEKLKEKIDLLQSELINMRE<br/> HKNLLKQIYFFDIHNSIKYKELYSERKNLIEQYNLQINGVKDVTAINHINT<br/> KLLSLKNKMDKITKQNSLYRLKYKLIAYSFLMIEFDGDVSKFKNFDPPT<br/> NLEKRVEYLDKKEEYLYNTAPKNKFNFAKLEEELQKIQSTSEMGADYLN<br/> VSPENNLKFYILTYIMLPVEFKGDFLGFKVKNHYNYKNVDFMDESLLDE<br/> NEVDSNKLNEKIENLKDSSFFNKIRLFEKNIKKYIEIVKYSVSTQENMKEYF<br/> KQLNLDIPYLDYKSTDEIGIFNKNMILPIFKYYQNFKLCNDIEIHALLALA<br/> NKKQQNLEYAIYCCSKKNSLNYNELLKTFNRKTYQNLSFIRNKIAHLNY<br/> KELFSDLFNELDLNTRVCLIEFSQNNKFDQIDLGMNFINDYYMKKTRF<br/> IFNQRRRLDLNVPSKEKIIDGKRKQQNDSNNELLKKYGLSRTNIKDIFNKA<br/> WY</p>                                     |
| WP_0<br>35935<br>671 | <p>MKVRYRKQAQLDFTFIKTEIVNNDIFIKSIIKAREKYRYSFLFDGEEKYHF<br/> KNKSSVEIVKNDIFSQTPDNMIRNYKITLKISEKNPRVVEAEIEDLMNSTIL<br/> KDGRRSARREKSMTERKLIIEKVAENYSLLANCPIEEVDSIKIYKIKRFLT<br/> YRSNMLLYFASINSFLCEGKIGKDNTEEIIWHLKDNDVRKEKVKENFKN<br/> KLIQSTENYNSSLKNQIEEKEKLSSEKFKKGAFYRTIHKLQQERIKELSEK<br/> SLTEDCEKIIKLYSELRHPLMHYDYQYFENLFENKENSELTKNLNLDIFKS<br/> LPLVRKMKLNKVNYLEDNDFLVLQKTKKAKTLYQIYDALCEQKNGF<br/> NKFINDFVSDGEENTVFKQIINEKFQSEMEFLEKRRISESEKKNKLLKLL<br/> DSMKAHFRNINSEDTKEAYFWDIHSRNYKTKYNERKNLVNEYTKLLGS<br/> SKEKLLREEITKINRQLLKLKQEMEEITKKNSLFRLEYKMKIAFGFLFCE<br/> FDGNISKFKDEFDASNQEKIIQYHKNGEKYLTSLFKEEKEKFNLEKMQK<br/> IIQKTEEEDWLLPETKNNLKFYLLTYLLLPELKGDFLGFKVKKHYDDIK<br/> NVDFMDENQNNIQVSQTVEKQEDYFYHKIRLFEKNTKKYEIVKYSIVPNE<br/> KLKQYFEDLGIDIKYLTGSVESGEKWLGENLGIDIKYLTVEQKSEVSEEK<br/> NKKVSLKNNGMFNKTILLFVKYQYQIAFKLFNDIELYSLFFLREKSEKPF<br/> VFLEELKDKMIGKQLNFGQLLYVVYEVLVKNKDLKILSKKIDYRKDKS<br/> FSPEIAYLRNFLSHLNYSKFLDNFMKINTNKSNDENKEVLIPSIIKQKMIQFIE<br/> KCNLQNQIDFDNFVNDFYMRKEKMFQIKLQIFPDINSTEKQKSEKEEI<br/> LRKRYHLINKKNEQIKDEHEAQSQLYEKILSLQKIFSCDKNNFYRRLKEE</p> |

[0627]

|                      |   |
|----------------------|---|
|                      | KLLFLEKQGKKKISMKEIKDKIASDISDLLGILKKEITRDIKDKLTEKFRYC<br>EEKLLNISFYNHQDKKKEEGIRVFLIRDKNSDNFKFESILDDGNSKIFISKNG<br>GKEITIQCCKVLETLMIEKNTLKISSNGKIISLIPHYSYSIDVKY  |
| WP_0<br>35906<br>563 | MEKFRRQNRSSIIKIIISNYDTKGIKELKVRYRKQAQLDTFIIKTEIVNNDIFI<br>KSIIEKAREKYRYSFLFDGEEKYHFKNKSSVEIVKKDIFSQTPDNMIRNYK<br>ITLKISEKNPRVVEAEIEDLMNSTILKDGRRSARREKSMTERKLIIEKVAE<br>NYSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINSFLCEGIK GKDN<br>ETEIIWHLKDNDVRKEKVKENFNKLIQSTENYNSSLKNQIEEKEKLLRK<br>ESKKGAFYRTIHKLQQERIKELSEKSLTEDCEKIIKLYSELRHPMLHYDY<br>QYFENLFENKENSELTKNLNLDIFKSLPLVRKMMLNNKVNYLEDNDTLF<br>VLQKTKKAKTLYQIYDALCEQKNGFNKFINDDFFVSDGEENTVFKQIINEK<br>FQSEIEFLEKRISESEKKNELKKKLDKSMKAHFRNINSEDTKEAYFWDIHS<br>SRNYKTKYNERKNLVNEYTELLGSSKEKLLREEITKINRQLLKLKQEME<br>EITKKNLFRLEYKMKMAFGFLFCFDGNISRFKDEFDASNQEKIQQYHK<br>NGEKYLTFLKEEEEKEKFNKLLQETIQKTGEENWLLPQNKNLKFYKYL<br>LTYLLLPELKGDFLGFVKKHYDIKNVDFMDENQSSKIIIESKEDDFYHK<br>IRLFKNTKKYEIVKYSIVPDKKLLKQYFKDLGIDTKYLILDQKSEVSGEKN<br>KKVSLKNNGMFNKTILLFVFKYYQIAFKLFNDIELYSLFFLREKSGKPFV<br>FLKELKDKMIGKQLNFGQLLYVVVEVLVKNKDLSEILSERIDYRKDMCF<br>SAEIALDRNFLSHNYSKFLDNFMKINTNKSDENKEVLPSIKIQKMIKFIEE<br>CNLQSQIDFDNFVNDFYMRKEKMFIIQLKQIFPDINSTEKQKMNEKEEIL<br>RNRHYLTDKKNEQIKDEHEAQSQLYEKILSLQKIYSSDKNNFYGRLKEEK<br>LLFLEKQEKKKLSMEEIKDKIAGDISDLLGILKKEITRDIKDKLTEKFRYCE<br>EKLLNLSFYNHQDKKKEESIRVFLIRDKNSDNFKFESILDDGNSKIFISKNG<br>KEITIQCCKVLETLMIEKNTLKISSNGKIISLIPHYSYSIDVKY |
| WP_0<br>42678<br>931 | MKSGRREKAKSNKSSIVRVIIISNFDDKQVKEIKVLYTKQGGIDVIKFKSTE<br>KDEKGRMKFNFDCAYNRLEEEEFNSFGGKQSFVTTNEDLTELHVTK<br>RHKTTGEIHKDYTIQGGYTPIKQDRTKVTVSITDNKDHFDSDNLGDKIRLS<br>RSLTQYTNRILLDADVMKNYREIVCSDSEKVDETINIDSQEIYKINRFLSY<br>RSNMIIYYQMINNLLHYDGEEDKGGNDSINLINEIWKYENKKNDEKEKII<br>ERSYKSIEKSINQYILNHNTEVESGDKEKKIDISEERIKEDLKKTFILFSRLR<br>HYMVHYNKYFYENLYSGKNFIIYNKDKSKSRRFSELLDLNIFKELSKIKL<br>VKNRAVSNYLDKKTTHVLNKNINAIKLLDIYRDICETKNGFNFINMM<br>TISGEEDKEYKEMVTKHFNENMNKLSIYLENFKKHSDFKTNNKKKETYN<br>LLQELDEQKKLRLWFNAPYVYDIHSSKKYKELYVERKKYVDIHSKLI<br>AGINNDNKKLNEINVKLCELNTEMKEMTKLNSKYRLQYKLQLAFGFIL<br>EEFNLDIDKFVSAFDKDNLTISKFMKRETYLSKSLDRRDNRFKKLIKD<br>YKFRDTEIDFCSDRENVLKLYILMYILLPVEIRGDFLGFVKKNYDLKH<br>VDFIDKRNNNDNKDTFFHDLRLFEKNVCRLEVTSSLSGDFLGKKSREKF<br>GKELEKFIYKNVSIALPTNIDIKEFNKSLVLPMMKNYQIIFKLLNDIEISALF<br>LIAKKEGNEGSITFKKVIDKVRKEDMNGNINFSQVMKALNEKVNQCIR<br>NSIAHINMKQLYIEPLNIYINNNQNKKTISEQMEEIIDICITKGLTGKELNKN<br>IINDYYMKKEKLVFNKLRKRNNLVSIDAQQKNMKEKSILNKYDLNYKD<br>ENLNIEIILKVNDLNNKQKLLKETTEGESNYKNALSKDILLNGIIRKNIN<br>FKIKEMILGIIQQNEYRYVNINIYDKIRKEDHNIDLKINNKYIEISCYENKSN<br>ESTDERINFKIKYMDLKVKNELLPSCYEDIYIKKKIDLEIRYIENCKVVYI<br>DIYYKKYNINLEFDGKTLFVKFNKDVKKNNQKVNLESNYIQNIKFIVS         |
| WP_0<br>62627<br>846 | MEKFRRQNRNSIIKIIISNYDTKGIKELKVRYRKQAQLDTFIIKTEIVNNDIFI<br>IKSIIEKAREKYRYSFLFDGEEKYHFKNKSSVEIVKKDIFSQTPDNMIRNY<br>KITLKISEKNPRVVEAEIEDLMNSTILKDGRRSARREKSVTERKLIIEKVA  |

[0628]

|                               |   |
|-------------------------------|---|
|                               | <p>ENYSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINSFLCEGIKKE<br/> NETEEIWLKDNVDRKEKVKENFKNKLIQSTENYNSSLKNQIEEKEKLLR<br/> KESKKGAFYRTIIKKLQQERIKELSEKSLTEDCEKIIKLYSKLRHSLMHYD<br/> YQYFENLFENKETPELKDKLDLHLFKSLPLIRKMKLNKVNYLEDGDTLF<br/> VLQKTKKAKTLYQIYDALCEQKNGFNKFINDFFVSDGEENTVFKQIINEK<br/> FQSEMEFLGKRISSEKKNPKLKKKFDMSKAHFHNINSEDTKEAYFWDIH<br/> SSSNYKTKYNERKNLVNEYTELLGSSKEKLLREEITQINRKLKQEM<br/> EETKKNLSFRLEYKMKMAFGFLCFEFDGNISRFKDEFDASNQEKIIQYHK<br/> NGEKYLTYFLKEEKEKFNLKKLQETIQKTGKENWLLPQNKNLKFYFL<br/> LTYLLLPELKGDFLGFVKKHYYDIKNVDFMDENQSSKIIESKEDDFYHK<br/> IRLFEKNTKKYEIVKYSIVPDEKQYFKDLGIDTKYLILEQKSEVSGEKN<br/> KKVSLKNNGMFNKTILLFVFKYYQIAFKLFNDIELYSLFFLREKSGKPFV<br/> FLKELKDKMIGKQLNFGQLLYVIYEVLVKNKDLSEILSERIDYRKDMCFS<br/> AEIADLRNFLSHLNYSKFLDNFMKINTNKSDENKEVLPSIKIQKMIKFIEE<br/> CNLQSQIDFDNFVNDFYMRKEKMFQIKLQIFPDINSTEKQKMNEKEEIL<br/> RNRYHLTDKKNEQIKDEHEAQSQLYEKILSLQKIYSSDKNNFYGRLKEEK<br/> LLFLGKQGKLLSMEEIKDKIAGDISDLLGILKKEITRDIKDKLTEKFRYC<br/> EEKLLNLSFYNHQDKKKEESIRVFLIRDKNSDNFKFESILDDGSNKIFISK<br/> GKEITIQCCDKVLETLMIEKNTLKISSNGKIISLVPHYSYSIDVKY</p>   |
| <p>WP_0<br/>05959<br/>231</p> | <p>MEKFRQRNRNSIIKIIISNYDTKGIKELKVRyrKQAQLDTFIIKTEIVNNDIF<br/> IKSIEKAREKYRYSFLFDGEEKYHFNKSSVEIVKDFDIFSQTPDNMIRNY<br/> KITLKISEKNPRVVEAIEDLMNSTILKDGRRSARREKSMTERKLIIEKVA<br/> KNYSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINSFLCEGIKGD<br/> NETEEIWLKDNVDRKEKVRNFKNKLIQSTENYNSSLKNQIEEKEKLLR<br/> KEFKKGAFYRTIIKKLQQERIKELSEKSLTEDCEKIIKLYSKLRHSLMHYD<br/> YQYFENLFENKKNDDLKDLNLDLFLKSLPLIRKMKLNKVNYLEDGDT<br/> LFVLQKTKKAKTLYQIYDALCEQKNGFNKFINDFFVSDGEENTVFKQIIN<br/> EKFQSEMEFLEKRISSEKKNKLLKLDMSKAHFRNINSEDTKEAYFW<br/> DIHSSRNYKTKYNERKNLVNEYTELLGSSKEKLLREEITKINRQLLKLK<br/> QEMEEITKKNLSFRLEYKMKIAFGFLCFEFDGNISKFKDEFDASNQEKIIQ<br/> YHKNGEKYLSFLKEEKEKFNLEKMQKIIQKTEEDWLLPETKNNLKFY<br/> YLLTYLLLPELKGDFLGFVKKHYYDIKNVDFIDENQNNIQVSQTVKQKQ<br/> DYFYHKIRLFEKNTKKYEIVKYSIVPNEKQYFEDLGIDIKYLTVEQKSE<br/> VSEKKNKKVSLKNNGMFNKTILLFVFKYYQIAFKLFNDIELYSLFFLREKS<br/> GKPLEIFRKELESKMKDGYLNFQQLLYVVYEVLVKNKDLKILSKKIDY<br/> RKDKSFSPEIAYLRNFLSHLNYSKFLDNFMKINTNKSDENKEVLPSIKIQK<br/> MIQFIEKCNLQSQIDFDNFVNDFYMRKEKMFQIKLQIFPDINSTEKQKM<br/> NEKEEILRNRYHLTDKKNEQIKDEHEAQSQLYEKILSLQKIYSSDKNNFY<br/> GRLKEEKLLFLEKQGKLLSMEEIKDKIAGDISDLLGILKKEITRDIKDKL<br/> TEKFRYCEEKLLNLSFYNHQDKKKEESIRVFLIRDKNSDNFKFESILDDGS<br/> NKIFISKNGKEITIQCCDKVLETLMIEKNTLKISSNGKIISLVPHYSYSIDVKY</p> |
| <p>WP_0<br/>27128<br/>616</p> | <p>MGKPNRSSIIKIIISNYDNKGIKEVKVRYNKQAQLDTFLIKSELKDGKFIY<br/> SIVDKAREKYRYSFEIDKTNINKNEILIKKDIYSNKEDKVIRKYILSFEVSE<br/> KNDRTIVTKIKDCLETQKKEKFERENTRRLISETERKLLSEETQKTYSKIA<br/> CCSPEDIDSVKIYKIKRYLAYRSNMLLFFSLINDIFVKGVVKDNNGEEVGEI<br/> WRIIDSKIDEKKTIDYLLVENFKKRMSQEFINYKQSIENKIEKNTNKIKEIE<br/> QKLLKKEKYKKEINRLKQKLIENRENDLLEKDKIELSDEEIREKILKIY<br/> SDLRHKLMHYNYQYFENLFENKKSKEKNEDVNLTELLDLNLFYRPLV<br/> RQLKLENKTNYLEKEDKITVLGVSDSAIKYYSYNNFLCEQKNGFNFIN<br/> FFSNDGEENKSFKEKINLSLEKEIEIMEKETNEKIKEINKNELQLMKEQKE</p>  |

[0629]

|                               |  |
|-------------------------------|--|
|                               | <p>LGTA YVLDI HSLNDYKISHNER NKNV KLQNDIMNGNRDKNALDKINKKL<br/> VELKIKMDKITKRNSILRLKYKLQVAYGFLMEEYKGNIKKFKDEFDISKE<br/> KIKSYKSKGEKYLEVKSEKKYITKILNSIEDIHNITWLKNQEENNLFKFYV<br/> LTYILLPFEFGRDGLGFVKKHYDDIKNVEFLDENNDRLTPEQLEKMKNDS<br/> FFNKIRLFEKNSKKYDILKESILT SERIGKYFSLNTGAKYFEYGGREENRGI<br/> FNKNIIPIFKYYQIVLKL YNDVELAMLLTLSESEKDKINKIKELVTLKEKV<br/> SPKKIDYEKKYKFSVLLDCFNRIINLGKKDFLASEEVKEVAKTFTNLAYL<br/> RNKICHLNYSKFIDDLTIDTNKSTTDSEGKLLINDRIRKLIK FIRENNQKM<br/> NISIDYNYINDYMKKEKFIGQRKQAKTIIDSGKKANKRNKAEELLKMY<br/> RVKKNENLIYELSKKLNELTKSELFLDKKLLKDIDFTDVKIKNKSFFEL<br/> KNDVKEVANIKQALQKHSSSELIGIYKKEVIMAIKRSIVSKLIYDEEKVLSIII<br/> YDKTNKKYEDFLEIRREDINKFQFLIDEKKEKLGYEKIIETKEKKKVVV<br/> KIQNNSELVSEPRIIKNKDKKKAKTPEEISKL GILDLTNHYCFNLKITL</p>   |
| <p>WP_0<br/>62624<br/>740</p> | <p>MEKFRQRNRNSIIKIIISNYDTKGIKELKVRYRKQAQLDTFIIKTEIVNNDIF<br/> IKSIEKAREKYRYSFLFDGEEKYHFKNKSSVEIVKKDIFSQTPDNMIRNY<br/> KITLKISEKNPRVVEAEIEDLMNSTILKDGRRSARREKSMTERKLIIEEKVA<br/> KNYSLLANCPMEEVDSIKIYKIKRFLTYRSNMLLYFASINSFLCEGIKGD<br/> NETEEIWLKDNDVRKEK VRENFKNLIQSTENYNSSLKNQIEEKEKLLR<br/> KEFKKGFYRTIIKKLQKERIKELSEKSLTEDCEKIIKLYSKLRHSLMHYD<br/> YQYFENLFENKKNDDL MKDLNLDLFSKSLPLIRKMKLNKNVNYLEDGDT<br/> LFVLQKTKKAKTLYQIYDALCEQKNGFNKFINDFFVSDGEENTVFKQIIN<br/> EKFQSEMEFLEKRISESEKNEKLLKLLDSMKAHFRNINSEDTKEAYFW<br/> DIHSSRNYKTKYNERKNLVNEYTELLGSSKEKLLREEITKINRQLLKLK<br/> QEMEEITKKNSLFRLEYKMKIAFGFLFCFDGNISKFKDEFDASNQEKIIQ<br/> YHKNGEKYLTSFLKEEKEKFNLEKMQKIIQKTEEDWLLPETKNNLFKF<br/> YLLTYLLL PYELKGDGLGFVKKHYDDIKNVDFIDENQNNIQVSQTVEKQE<br/> DYFYHKIRLFEKNTKKYEIVKYSIVPNEKLLKQYFEDLGIDIKYLTGSVESG<br/> EKWLGENLGIDIKYLTVEQKSEVSEEKNNKVSLLKNNGMFNKTILLFVFK<br/> YYQIAFKLFNDIELYSLFFLREKSGKPLEIFRKELESKMKDGYLNFQQLLY<br/> VVEVLVKNKDLKILSKKIDYRKDKSFSPEIAYLRNFLSHLNYSKFLDN<br/> FMKINTNKSDENKEVLPSIKIQKMIQFIEKCNLQNDIDFDNFVNDFYMR<br/> KEKMFFIQLKQIFPDINSTEKQKMNEKEEILRNRYHLTDKKNEQIKDEHE<br/> AQSQLYEKILSLQKIYSSDKNNFYGRLKEEKLLFLEKQGKKLSMEEIKD<br/> KIAGDISDLLGILKKEITRDIKDKLTEKFRYCEEKLLNLSFYNHQDKKKEE<br/> SIRVFLIRDKNSDNFKFESILDDGSNKIFISKNGKEITIQCCKVLETLIEKN<br/> TLKISSNGKIISLIPHYSYSIDVKY</p> |
| <p>WP_0<br/>96402<br/>050</p> | <p>MENKNKPNRGSIVRIISNYDMKGIKELKVRYRKQAQLDTFILQTTLDKS<br/> NNSILINDFRVKAREKYRYSFTYDGKEKFSVPSNSIIVTKIDNAAPEKSKEI<br/> RKYKITLGIDEKCKTGSMITAAIEDLLEDDRVREGIRNPRRKASKTERKLI<br/> TESICHNYAQITQCPVEEIDAVKIYKVKRFLSYRSNMLLFFALINDFLCKN<br/> LKNEKGEKINEIWEMENKGNKIDFDENYILVAQIKEYFTKEIENYNN<br/> RIDNIIDKKELLYSEEKEESEKNNKLEELNKLESQKLLKILTDEEIKADVIK<br/> IIFSDLRHSLMHYKEYYFENLFENKKNEEELAE LLNLNLFKNLTLRQM<br/> KIENKTNYLEGDEKFNILGKDVRAKNALGHYDLLVEQKNGFNFNINSFF<br/> VQDG TENLEFKKFIDENFIKAQKELEEDIKNCKESVKKLEKLLKENPKKS<br/> EDLEKKLEKKQKLLKELKKELEKMKQHYKRLNCAYVWDIHSSTVYKLL<br/> YNERKNLIEKYNKQLNGLQDKNAITGINAQLLRIKKEMEEITKSNSLFRL<br/> KYKMQIAYAFLEMEYEGNIAKFKNEFDCSKTEKIQEWLEKSEEYLNYSM<br/> EKEEDGKNYKFHFKEISEIKDTHNEEWLENTSENNLKFYILTYLLLPMEF<br/> KGDGLGVVKKHYDDIKNVDFDTESEKELSQE QIDKMIGDSFFHKIRLFEK</p>   |

[0630]

|  |
|--|
| <p>NTKRYEIIKYSILTSDEIKKYFELLEKVPYLEYKGYDEIGIFNKNILPIFKY<br/> YQIIFRLYNDLEIHGLFNVSFDINKILSDLKSYGNENINFREFLYVIKQNNN<br/> SSTEEYQKIWEKLESKYLKEPLLTPPEKKEINKKTEKELKKLDGISFLRNK<br/> ISHLEYEKIIEGVLKTAVNGENKKTSETNADKVFLNEKIKKIIFIKENELD<br/> KIELGFNFINDFFMKKEQFMFGQIKQVKEGNSDSITTERKRKEENKRLKI<br/> TYGLNYNNLSKIYEFNTLREIVNSPLFLKDSTLLKKVDLSKVMLKEKPIC<br/> SLQYENNTKLEDDIKRILLKSSDIMGIYKAEVVKLKEKLVLFKYDEEK<br/> KIYVTVYDTSKAVPENISKEILVKRNSKEEYFFEDNKKKYTTQYYTLEIT<br/> KENELKVIPAKKLEGKEFKTEKKEENKMLMLNNHYCFNVKIIY</p> |
|--|

[0631] 在某些示例实施方式中,所述CRISPR效应子蛋白是选自表5的Cas13d蛋白。

[0632] 表5

[0633]

|  |   |
|--|---|
| <p>RfxCas13d<br/>(SEQ ID<br/>NO:233)</p> | <p>MIEKKKSFAGMGVSKSTLVSGSKVYMTTFAEGSDARLEKIVEGDSIRS<br/> VNEGEAFSAEMADKNAGYKIGNAKFSHPKGYAVVANNPLYTGPVQQ<br/> DMLGLKETLEKRYFGESADGNDNICIQVIHNILDIEKILAEYITNAAYAV<br/> NNISGLDKDIIGFGKFSTVYTYDEFKDPEHHRAAFNNNDKLINAIAQY<br/> DEFDNFLDNPRLG YFGQAFFSKEGRNYIINYGNECYDILALLSGLRHW<br/> VVHNNEEESRISRTWLYNLDKNLDNEYISTLNYLYDRITNELTNSFSKN<br/> SAANVNYIAETLGINPAEFAEQYFRFSIMKEQKNLGFNITKLREVM LDR<br/> KDMSEIRKNHKVFD SIRT KVYTMMD FVIYRYIEEDAKVAAANKSLP<br/> DNEKSLSEKDIFVINLRGSFNDDQKDALY YDEANRIWRKLENIMHNIK<br/> EFRGNKTREYKKKDAPRLPRILPAGR DVSAF SKL MYALTMFLDGKEIN<br/> DLLTTLINKFDNIQSFLKVMPLIGVNAKFVEEYAFFKDSAKIADELRLIK<br/> SFARMGEP IADARRAMYIDAIRILGTNLSYDELKALADTFSLDENGNK<br/> LKKGKHGMRNFIINNVISNKR FHYLIRYGDPAHLHEIAKNEAVVKFVL<br/> GRIADIQKKQNGKNQIDRYETCIGKDKGKSVSEKVDALTKIITGM<br/> NYDQFDKKRSVIEDTGRENAEREKFKKIISLYLTVIYHILKNIVNINARY<br/> VIGFHCVERDAQLYKEKGYDINLKKLEEKGFSSVTKLCAGIDETAPDK<br/> RKDV EKEMAERAKESIDSLESANPKLYANYIKYSDEKKAEEFTRQINR<br/> EKAKTALNAYLRNTKWNVIREDLLRIDNKTCTLFRNKAVHLEVARY<br/> VHAYINDIAEVNSYFQLYHYIMQRIIMNERYEKSSGKVSEYFDAVNDE<br/> KKYNDRLLKLLCVPFGYCIPRFKNLSIEALFDRNEAAKFDKEKKKVS<br/> NS</p> |
| <p>AdmCas13d<br/>(SEQ ID<br/>NO:234)</p> | <p>MNNKRKTKAKAAGLKS VFFDQKQAVLTTFAKGNNSQIEKKVVNSEV<br/> KDLRQPPAFDLELKEKTFYISGKNNINTSRENPLASASLPLSKRQRIRAE<br/> RIKRAREENRPYHNVKRVGEDDLRAKADLEKHYFGKEYSDNLKIQIIY<br/> NILDINKIISPYINDIVYSMNNLARNDEYIDGKIDVIGLSSTTDYSSFMS<br/> PNKDLEKEKKFSFHRENYKKFVEASKPYMRYYGKVFIRDVKKSKLST<br/> GKGEKIEVMYRSDEEIFTIFQILSYVRQSIMHNDIGNKSSILAIEKYPARF<br/> VGFLSDLLKTKTNDVNRMFIDNNSQTNFWVLF SIFGLQDHTSGADKIC<br/> RNFYDFVIKADSKNLGFLSKKIRELMLDLPNANMLRDHQFDTVRSKFY<br/> TLLDFIYQHYLEEKSRIDNMVEKLRMTLKEEEKEVLYAAEAKIVWNA<br/> IGAKVINKLVPMMNGDALKEIKRKNRDRKLPQSVIATVQVNSDANVF<br/> SGLIYFLTLFLDGKEINEMVSNLITKFENIDSL LHVDREIYKSDEKDL DL<br/> EIEKLALFFKGVVRPNAKTDTGAGEISKSFSIFQSAERIIIEELKFIKNVTR<br/> MDNEIFPSEGVFLDAANVLGVRGDDDFDFSNEFVGDDLHSDANKKIINK</p>  |

[0634]

|                                  |   |
|----------------------------------|---|
|                                  | INGTKEDRNLRNFIINNVVKSRRFQYIARHMNTHYVKQLANNETLNRFLNKMGDAKIINRYYESISGNTPNIEVRSQIDYLVKRLRSFSFEDLNDVKQKVRPGTNEISIEKEKKKALVGLCLTIQYLKYKNLVNINARYYTAFYCLERDSKLGKFGVDVWRDFESYALTNHFIKEGYLPVRKAEILRANLKHLDCEDEGFKYYRNQVTHLNAIRVAKYINEIKSVHSYFALYHYIMQRHLYDSLQAKAKDSSGFVIDALKKSFEHKKIYSKDLLHVLHSPFGYNTARYKNLSIEALFDKNESRPEVNPLSTND   |
| UrCas13d<br>(SEQ ID<br>NO:235)   | MAKKNMKPRELREAQKKARQLKAAEINNNAAPAIAAMPAAEVIAPAAEKKSSVKAAGMKSILVSENKMYITSFGKGN SAVLEYEVDNNDYNQTQLSSKDNSNIQLGGVNEVNITFSSKHGFESGVEINTSNPTHRSGESSPVRGDMGLKSELEKRFPGKTFDDNIHQIYLIYNILDIEKILAVYVTNIVYALNNMLGVKGSSESHDDFIGYLSTNNIYDVFIDPDNSSLDDKKANVRKSLSKFNALLKTKRLGYFGLEPKTKDNRVSQAYKKRVYHMLAIVGQIRQCVFHDKSGAKRFDLYSFINNIDPEYRDTLDYLVEERLKSINKDFIEDNKVNISLLIDMMKGYEADDIIRLYYDFIVLKSQKNLGFSSIKKREKMLDEYGFREFDKQYDSVRSKMYKLMDFLLFCNYRNDIAAGESLVRKLRFSMTDDEKEGIYADEAAKLWGKFRNDFENIADHMNGDVIKELGKADMDFDEKILDSEKKNASDLLYFSKMIYMLTYFLDGKEINDLLTTLISKFDNIK EFLKIMKSSAVDVECELTAGYKLFNDSQRITNELFIVKNIASMRKPAASAKLTMFRDALTILGIDDKITDDRISGILKKEKKGKGIHGLRNFITNNVIESRFVYLIKYANAQKIREVAKNEKVVMFVLGGIPDTQIERYYKSCVEFPDMNSSLGVKRSELARMIKNISFDDFKNVKQQAAGRENVAKERAKAVI GLYLTVMYLLVKNLVNVNARYVIAIHCLERDFGLYKEIPELASKNLKNDYRILSQTLCCLCDKSPNLFKKNERLRKCVEVDINNADSSMTRKYRNCIAHLTVVRELKEYIGDICTVDSYFSIYHYVMQRCITKRENDTKQEEKIKYEDDLLKNHGYTKDFVKALNSPFGYNIPRFKNLSIEQLFDRNEYLTEK |
| PIE0Cas13d<br>(SEQ ID<br>NO:236) | MEREVKKPPKSLAKAAGLKSTFVISPQEKELAMTAFGRGNDALLQKRIVDGVVRDVAGEKQQFQVQRQDESRFRLQNSRLADRTVTADDPLHR AETPRRQPLGAGMDQLRRKAILEQKYFGRTFDDNIHQIYLIYNILDIHKM LAVPANHIVHTLNLGGYGETDFVGMPLAGLPYDKLRVVKKKNNGDTVDIKADIAAYAKRPQLAYLGAIFYDVTGKSKRDAARGRVKREQDVYAILSLMSLLRQFCAHDSVRIWGQNTTAALYHLQALPQDMKDLLDDGWRRALGGVNDHFLDTNKVNLLTLFEYYGAETKQARVALTQDFYRFV VLKEQKNMGFSLRRLREELLKLPDAAYLTGQEYDSVRQKLYMLLDFL LCRLYAQERADRCEELVSALRCALSDEEKDTVYQAEAAALWQALGDTLRRKLLPLLKGGKLLQDKDKKKSDELGLSRDVLGVLFRPAQQGSRA NADYFCRLMHLSTWFMDGKEINTLLTTLISKLENIDSLRSVLESMLGLAYSFVPAYAMFDHSRYIAGQLRVVNNIARMRKAIGAKREMYRAAVVLLGVDSPEAAAAITDDLQIDPETGKVRPRSDSARDTGLRNFIANNVVE SRRFTYLLRYMTPEQARVLAQNEKLIAFVLSTVPDTQLERYCRTCGRE DITGRPAQIRYLTAQIMGVRYESFTDVEQRGRGDNPKKERYKALIGLY LTVLYLAVKNMVNVCNARYVIAFYCRDRDTALYQKEVCWYDLEEDKKSGKQRQVEDYTALTRYFVSQGYLNRHACGYLRSNMNGISNSLLTAYR NAVDHLNAIPPLGSLCRDIGRVDSYFALYHYAVQQYLNGRYYRKTPEQELFAAMAQHRTWCSDLVKALNTPFGYNLARYKNLSIDGLFDREGD HVVREDGEKPAE  |
| RffCas13d<br>(SEQ ID<br>NO:237)  | MKKKMSLREKREAEKQAKKAAYSASKNTDSKPAEKKAETPKPAEII SDNSRNKTA VKAAGLKSTIISGDKLYMTSFGKGNAAVIEQKIDINDYSF SAMKDTPSLEVDKAESKEISFSSHHPFVKNDKLLTYNPLYGGKDNPEK  |

[0635]

|                                |  |
|--------------------------------|--|
|                                | PVGRDMLGLKDKLEERYFGCTFNDNLHIQIINYILDIEKILAVHSANITT<br>ALDHMVDEDEDEKYLNSDYIGYMTINTYDVFMDPSKNSSLSPKDRKN<br>IDNSRAKFEKLLSTKRLGYFGFDYDANGKDKKKNEEIKRLYHLTAFAG<br>GQLRQWSFHSAGNYPRTWLYKLDLSDKEYLDTLDHYFDKRFNDIND<br>DFVTKNATNLYILKEVFPKANFKDIADLYYDFIVIKSHKNMGFSIKLR<br>EKMLECDGADRIKEQDMDSVRSKLYKLIDFCIFKYHEFPELSEKNVDI<br>LRAAVSDTKKDNLYSDEAARLWSIFKEKFLGFCDKIVVWVTGEHEKD<br>ITSVIDKDAYRNRSNVSYFSKLMYAMCFFLDGKEINDLLTTLINKFDNI<br>ANQIKTAKELGINTAFVKNYDFFNHSEKYVDELNIVKNIARMKKPSSN<br>AKKAMYHDALTILGIPEDMDEKALDEELDLILEKKTDPVTGKPLKGN<br>PLRNFIANNVIENSRIYLIKFCNPENVRKIVNNTKVTEFVLKRIPDAQIE<br>RYYKSCTDSEMNPPEKKITELAGKLDKDMNFGNFRNVRQSAKENMEK<br>ERFKA VIGLYLTVVYRVVKNLVDVNSRYIMAFHSLERDSQLYNVSD<br>NDYLALDTLVKEGDNSRSRYLAGNKRLRDCVKQDIDNAKKWVSD<br>KYNSITKYRNNVAHLTAVRNCAEFIGDITKIDSYFALYHYLIQRQLAKG<br>LDHERSGFDRNYPQYAPLFWHTYVKDVKALNAPFGYNIPRFKNLSI<br>DALFDRNEIKKNDGEKKSDD  |
| RaCas13d<br>(SEQ ID<br>NO:238) | MAKKSkgmslREKRELEKQKRIQKAAVNSVNDTPEKTEEANVSVN<br>VRTSAENKHSKSAKALGLKSGLVIGDELYLTSFGRGNEAKLEKKIS<br>GDTVEKLGIGAFEVAERDESTLTLESGRIDKKTARPKDPRHITVDTQK<br>FKEDMLGIRSVLEKKIFGKTFFDDNIHVQLAYNILDVEKIMAQYVSDIVY<br>MLHNTDKTERNDNLMGYMSIRNTYKTFCDTSNLPDDTKQKVENQKR<br>EFDKIKSGRLGYFGEAFMVNSGNSTKLRPEKEIYHIFALMASLRQSYF<br>HGYYKDTDYQGTWAYTLEDKLGKPSHEFRETIDKIFDEGFSKISKDF<br>GKMNKVNLQILEQMIGELYGSIERQNLTCDYDFIQLKHKHYLGF<br>RLRETMLETPAECYKAECYNSERQKLYKLIDFLIYDLYNRKPARIEE<br>IVDKLRESVNDEEKESIYSVEAKYVYESLSKVLDKSLKNSVSGETIKDL<br>QKRYDDEANRIWDISQHSISGNVNCFKLIYIMTLMLDGKEINDLLTT<br>LVNKFDNIAFIDVMDDELGLEHSFTDNYKMFADSKAICLDLQFINSFAR<br>MSKIDDEKSKRQLFRDALVILDIGNKDETWINNYLDSDFKLDKEGNK<br>LKGARHDFRNFIANNVIKSSRFKYLVKYSSADGMIKLKTNEKLIGFVL<br>DKLPETQIDRYYESCGLDNAVVDKKVRIEKLKSLGIRDMKFDDFSGVKT<br>SNKAGDNDKQDKAKYQAIISLYLMVLYQIVKNMIYVNSRYVIAFHCL<br>ERDFGMYGKDFGKYYQGCRKLTDFHIEEKYMKEGKLGKCNKKVGRYL<br>KNNISCTDGLINTYRNQVDHFVVRKIGNYAAIKSIGSWFELYHYVI<br>QRIVFDEYRFALNNTESNYKNSIHKHTYCKDMVKALNTPFGYDLPRY<br>KNLSIGDLFDRNNYLNKTKESIDANSSIDSQ |
| EsCas13d<br>(SEQ ID<br>NO:239) | MGKKIHARDLREQRKTDRTEKFADQNKKREAERAVPKKDAAVSVKS<br>VSSVSSKDNVTKSMAKAAGVKSVAVGNTVYMTSFGRGNDVLEQ<br>KIVDTSHEPLNIDDPAYQLNVVTMNGYSVTGHRGETVSAVTDNPLRR<br>FNGRKKDEPEQSVPTDMLCLKPTLEKKFFGKEFDDNIHQIYINIL<br>IEKILAVYSTNAIYALNNMSADENIENSDFFMKRTTDETFDDFEKKESTN<br>SREKADFDAFEKFIGNYRLAYFADAFYVNKKNPKGKAKNVLREDKEL<br>YSVLTLLIGKLRHWCVHSEEGRAEFWLYKLDELKDDFKNVLDVYVNR<br>VEEINNRFIENKVNILQILGSVYKNTDIAELVRSYEFLLITKYYKNMGFS<br>IKKLRESMLEGKGYADKEYDSVRNKLYQMTDFILYTGYNEDSDRAD<br>DLVNTLRSSLKEDDKTTVYCKEADYLWKKYRESIREVADALDGDNIK<br>KLSKSNIEIQEDKLRKCFISYADSVSEFTKLIYLLTRFLSGKEINDLVTTLI<br>NKFDNIRSFLEIMDELGLDRFTAEYSFFEGSTKYLAELVELNSFVKSCS<br>FDINAKRTMYRDALDILGIESDKTEEDIEKMIDNILQIDANGDKKLLKN   |

[0636]

|   |
|---|
| NGLRNFIASNVIDSNRFKYLVRYPGNPKKIRETAKCKPAVRFVLNEIPDA<br>QIERYYEACCPKNTALCSANKRREKLADMI AEIKFENFSDAGNYQKAN<br>VTSRTSEAEIKRKNQAIIRLYLTVMYIMLKNLVNVNARYVIAFHCV<br>DTKLYAESGLEVGNIEKNKTNLTMAMVGMVKLENGIHKTEFDKSF<br>AENA<br>ANRYLRNARWYKLLDNLKKSERAVVNEFRNTVCHLNAINININIK<br>E<br>KEVENYFALYHYLIQKHLENRFADKKVERDTGDFISKLEEHKTYCK<br>DF<br>VKAYCTPFGYNLVRYNLNTIDGLFDKNYPGKDDSDSEQK |
|---|

[0637] 变体和变异

[0638] 本公开提供了Cas蛋白的变体和突变形式。在一些实例中,本公开包括 Cas 13的变体和突变形式,例如Cas13b。与对应的野生型相比,Cas蛋白的变体或突变形式可以是催化失活的,例如不具有核酸酶活性或核酸酶活性降低。在某些实例中,Cas蛋白的变体或突变形式具有切口酶活性。

[0639] CAS13的突变

[0640] 在一些情况下,本公开提供了包含一种或多种氨基酸修饰的突变Cas13 蛋白,其中所述氨基酸:(a) 与和所述突变Cas 13蛋白形成复合物的引导RNA 相互作用;(b) 在所述突变Cas 13蛋白的HEPN活性位点、结构域间接头结构域或桥接螺旋结构域中;或其组合。

[0641] 术语“对应的氨基酸”或“对应于……的残基”指Cas13同系物或直系同源物中的特定氨基酸或其类似物,其与参照Cas蛋白中的氨基酸相同或在功能上等同。因此,如本文所用,提及指定的Cas 13蛋白的“对应于氨基酸位置 [X]的氨基酸位置”代表提及其他公认的Cas 13和结构同系物和家族中的等同位置的集合。本文所述的突变适用于所有Cas13蛋白,所述Cas13蛋白是所述Cas蛋白的直系同源物或同源物(例如PbCas13b)。例如,所述突变适用于 Cas13a、Cas13b、Cas13c、Cas13d、Cas13b-t1、Cas13b-t2或Cas13b-t3。

[0642] 在一方面,本发明涉及一种突变Cas13蛋白,其包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、A656、V795、A796、W842、K871、E873、R874、R1068、N1069或H1073。

[0643] 本文所用的PbCas13b优选地具有NCBI参照序列WP\_004343973.1的序列。应当理解,WP\_004343973.1是指野生型(即未突变的)PbCas13b。如本文中所使用的LshCas13a(沙氏纤毛菌(Leptotrichia shahii) Cas13a) 优选地具有 NCBI参照序列WP\_018451595.1的序列。应当理解,WP\_018451595.1是指野生型(即未突变的) LshCas13b。如本文所用的Pgu Cas13b(古卞淋单胞菌Cas13b) 优选地具有NCBI参照序列WP\_039434803.1的序列。应当理解,WP\_039434803.1是指野生型(即未突变的) Pgu Cas13b。本文所用的Psp Cas13b(普雷沃菌P5-125 Cas13b) 优选地具有NCBI参照序列 WP\_044065294.1的序列。应当理解,WP\_044065294.1是指野生型(即未突变的) Psp Cas13b。

[0644] 根据如本文所述的发明,在本发明的实施方式中,VI型系统包含突变 Cas13效应子蛋白(和可选地在Cas13b效应子蛋白的上游或下游编码的小辅助蛋白)。在某些实施方式

中,所述小辅助蛋白增强所述Cas13b效应子靶向 RNA的能力。

[0645] 从Cas13的结构中获得的见解使进一步的合理工程化成为可能,以改善 RNA靶向特异性、碱基编辑和核酸检测等的功能。基于阐明的Cas13效应子的晶体结构和本文所述的它的cRNA,可以假定合理工程化和诱变的功能含义,其非限制性突变例示于下表6中(参见PbCas13b;WP\_004343973.1)。

[0646] 表6

[0647]

| 残基        | 描述               | 预期的结果    |
|-----------|------------------|----------|
| T405      | 协调引导的第一个残基(U)    | 改变活性     |
| H407      | 与 U0 进行碱基堆积      | 可能涉及 PFS |
| H407Y/W/F | 与 U0 进行碱基堆积      | 改变 PFS   |
| K457      | A31 的直接读出        |          |
| H500      | 与 G11 的 bb 形成氢键  | 改变活性     |
| K570      | G25 的直接读出        | 改变活性     |
| K590      | U27 的 bb         | 改变活性     |
| N634      | A29 的 bb         | 改变活性     |
| R638      | A28 的 bb         | 改变活性     |
| N652      | U2 和 C36 的直接读出   | 改变活性     |
| N653      | C36 的直接读出        | 改变活性     |
| K655      | 与 na 3 的 bb 形成氢键 | 改变活性     |
| S658      | 协调引导的第一个残基       | 改变活性     |
| K741      | U27 的直接读出        | 改变活性     |
| K744      | 与 na 6 的 bb 形成氢键 | 改变活性     |
| N756      | C33 和 C5 的直接读出   | 改变活性     |
| S757      | A32 的直接读出        | 改变活性     |
| R762      | 与 G10 的 bb 形成氢键  | 改变活性     |
| R791      | A22 的 bb         | 改变活性     |
| K846      | 与 U18 的 bb 形成氢键  | 改变活性     |
| K857      | 与 C15 的 bb 形成氢键  | 改变活性     |
| K870      | 与 U19 的碱基形成氢键    | 改变活性     |
| R877      | U18 的直接读出        | 改变活性     |
|           | 通道               |          |
| K183      | 外通道边缘            | 改变活性     |
| K193      | 外通道边缘            | 改变活性     |
| R600      | 外通道边缘            | 改变活性     |

[0648]

|              |             |              |
|--------------|-------------|--------------|
| K607         | 外通道边缘       | 改变活性         |
| K612         | 外通道边缘       | 改变活性         |
| R614         | 外通道边缘       | 改变活性         |
| K617         | 外通道边缘       | 改变活性         |
| K826         | 桥接螺旋结构域     | 改变活性         |
| K828         | 桥接螺旋结构域     | 改变活性         |
| K829         | 桥接螺旋结构域     | 改变活性         |
| R824         | 桥接螺旋结构域     | 改变活性         |
| R830         | 桥接螺旋结构域     | 改变活性         |
| Q831         | 桥接螺旋结构域     | 改变活性         |
| K835         | 桥接螺旋结构域     | 改变活性         |
| K836         | 桥接螺旋结构域     | 改变活性         |
| R838         | 桥接螺旋结构域     | 改变活性         |
| R618         | 保守的外通道精氨酸   | 改变活性         |
| D434         | 保守环         | 改变活性         |
| K431         | 保守环         | 改变活性         |
|              | 活性位点口袋      |              |
| 46-57        | HEP1        |              |
| 73-79        | HEP1        |              |
| 152-164      | HEP1        |              |
| 1036-1046    | HEP2        |              |
| 1064-1074    | HEP2        |              |
| R53A/K/D/E   | HEP1        | 碱基特异性改变      |
| K943A/R/D/E  | HEP2        | 碱基特异性改变      |
| R1041A/K/D/E | HEP2        | 碱基特异性改变      |
| Y164A/F/W    |             | 影响活性位点处的碱基堆积 |
|              | 结构域间接头      |              |
| 285-299      |             |              |
| R285         | 中央通道活性口袋    | 改变活性         |
| R287         | 中央通道活性口袋    | 改变活性         |
| K292         | 中央通道活性口袋    | 改变活性         |
| E296         | 中央通道活性口袋    | 改变活性         |
| N297         | 中央通道活性口袋    | 改变活性         |
|              | 其他          |              |
|              | 反式活性位点环     | 改变活性         |
| Q646         | 反式活性位点环     | 改变活性         |
| N647         | 反式活性位点环     | 改变活性         |
|              | HEPN 界面     |              |
|              | crRNA 加工    |              |
| R402         | 去除 crRNA 加工 | 改变 crRNA 加工  |
| K393         | 去除 crRNA 加工 | 改变 crRNA 加工  |
| N653         | 去除 crRNA 加工 | 改变 crRNA 加工  |
| N652         | 去除 crRNA 加工 | 改变 crRNA 加工  |

|        |      |             |             |
|--------|------|-------------|-------------|
| [0649] | R482 | 去除 crRNA 加工 | 改变 crRNA 加工 |
|        | N480 | 去除 crRNA 加工 | 改变 crRNA 加工 |
|        |      | LID 结构域     |             |
|        | D396 | 功能未知的发夹     | 改变 crRNA 加工 |
|        | E397 | 功能未知的发夹     | 改变 crRNA 加工 |
|        | D398 | 功能未知的发夹     | 改变 crRNA 加工 |
|        | E399 | 功能未知的发夹     | 改变 crRNA 加工 |
|        | K294 | IDL         | 改变活性        |

[0650] 结构(亚)域

[0651] 在另一方面,本公开提供了包含一个或多个氨基酸突变的突变Cas13蛋白,其中所述氨基酸:与和所述工程化Cas 13蛋白形成复合物的引导RNA 相互作用;或在所述突变Cas 13蛋白的HEPN活性位点、lid结构域、螺旋结构域(选自螺旋1或螺旋2结构域)、结构域间接头(IDL)结构域(IDL)或桥接螺旋结构域或其组合中。

[0652] 基于所述Cas蛋白的晶体结构,可以识别不同的结构域。除序列比对外,所述晶体结构和结构域构建的信息还允许不同的直系同源物(例如Cas13b直系同源物)和同系物(其他Cas13蛋白,例如Cas13a、Cas13c或Cas13d)的对应的氨基酸被识别。举例来说而不受其限制,在如本文报道的和crRNA的复合物中的PbCas13b的晶体结构识别了以下结构域(也见图1A):HEPN1和HEPN2(催化结构域,其分别跨越氨基酸1至285和930至1127);IDL(结构域间接头,其跨越氨基酸286至301);螺旋结构域1和2,其中螺旋结构域在螺旋结构域1-1、1-2和1-3(其分别跨越氨基酸302至374、499至581和747至929)和螺旋结构域2(其跨越氨基酸582至746)中分裂;LID(其跨越氨基酸375至498)。螺旋结构域1(特别是螺旋结构域1-3)包含桥接螺旋作为可识别的亚结构域。因此,根据如本文所述的发明的特定突变,除了在Cas13多肽中具有指定的氨基酸位置之外,还可以与Cas13蛋白的特定结构域连接。因此,Cas13直系同源物或同系物中的对应的氨基酸可以在Cas13多肽中具有特定的氨基酸位置,并且属于对应的结构域(也见例如图4,其作为Cas13a和Cas13b的HEPN1和HEPN2中的对应的氨基酸的实例)。可以通过结构(亚)域中的位置,通过对应于特定Cas13蛋白(例如PbCas13b)的氨基酸的位置,通过与引导RNA的相互作用或其组合来识别突变。

[0653] 突变的类型可以是保守突变或非保守突变。在某些优选的实施方式中,将突变氨基酸突变成丙氨酸(A)。在某些优选的实施方式中,如果要突变氨基酸是芳族氨基酸,则将其突变成丙氨酸或另一种芳族氨基酸(例如H、Y、W或F)。在某些优选的实施方式中,如果要突变氨基酸是带电荷的氨基酸,则将其突变成丙氨酸或另一种带电荷的氨基酸(例如H、K、R、D或E)。在某些优选的实施方式中,如果要突变氨基酸是带电荷的氨基酸,则将其突变成丙氨酸或具有相同电荷的另一种带电荷的氨基酸。在某些优选的实施方式中,如果要突变氨基酸是带电荷的氨基酸,则将其突变成丙氨酸或具有相反电荷的另一种带电荷的氨基酸。

[0654] 本发明还提供了方法和组合物,其中所述效应子蛋白(例如工程化的或非天然存在的效应子蛋白或Cas13)的一个或多个氨基酸残基可以被修饰。在一个实施方式中,所述修饰可包括所述效应子蛋白的一个或多个氨基酸残基的突变。所述一个或多个突变可以在所述效应子蛋白的一个或多个催化活性结构域中,或在与crRNA相互作用的结构域中(例如引导序列或正向重复序列)。所述效应子蛋白相比于缺乏所述一个或多个突变的效应子蛋

白可具有降低或消除的核酸酶活性或可替代地提高的核酸酶活性。所述效应子蛋白可能不引导在靶RNA靶基因座处的RNA链的切割。在一个优选的实施方式中,一个或多个突变可包含两个突变。在一个优选的实施方式中,在Cas13b效应子蛋白(例如工程化的或非天然存在的效应子蛋白或Cas13b)中修饰一个或多个氨基酸残基。在某些情况下,所述CRISPR-Cas蛋白在所述螺旋结构域中包含一个或多个突变。

[0655] 本文所述的Cas13蛋白可包含一个或多个突变。在一些情况下,所述 Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 Cas13b (PbCas13b) 的以下氨基酸:T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、A656、V795、A796、W842、K871、E873、R874、R1068、N1069 或H1073。

[0656] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、W842、K871、E873、R874、R1068、N1069或H1073。

[0657] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294或E400。

[0658] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的以下氨基酸:K393、R402、N482、T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或 H1073。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:K393、R402、N482、H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或H1073。

[0659] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:W842、K846、K870、E873或R877。在某些情况下,所述Cas13蛋白在螺旋结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b螺旋结构域1中的以下氨基酸:W842、K846、K870、E873或R877。在某些情况下,所述Cas13蛋白在螺旋结构域1-3

中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b螺旋结构域1-3 中的以下氨基酸:W842、K846、K870、E873或R877。在某些情况下,所述Cas13蛋白在螺旋桥接域中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b螺旋桥接域中的以下氨基酸:W842、K846、K870、E873或R877。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:K393、R402、N480、N482、N652或N653。在某些情况下,Cas13蛋白包含一个或多个与PbCas13b的以下氨基酸相对应的氨基酸突变:K393、R402、N480或N482。在一些情况下,所述Cas13 蛋白在LID结构域中包含一个或多个氨基酸突变,所述氨基酸对应于 PbCas13b的LID结构域中的以下氨基酸:K393、R402、N480或N482。在一些情况下,Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于 PbCas13b的以下氨基酸:N652或N653。在一些情况下,所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b 的螺旋结构域2中的以下氨基酸:N652或N653。

[0660] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b 的以下氨基酸:H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、S757、N756或K741。在一些情况下,所述Cas13蛋白在螺旋结构域中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域中的以下氨基酸:S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、S757、N756或K741。

[0661] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或 N756。在一些情况下,所述Cas13蛋白在螺旋结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域1中的以下氨基酸:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或N756。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:H567、H500、R762、V795、A796、R791、G566、S757或N756。在一些情况下,所述Cas13蛋白在螺旋结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域1中的以下氨基酸:H567、H500、R762、V795、A796、R791、G566、S757或N756。

[0662] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:K871、K857、K870、W842、E873、R877、K846或R874。在一些情况下,所述Cas13蛋白在所述螺旋桥接域中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的所述螺旋桥接域中的以下氨基酸:K871、K857、K870、W842、E873、R877、K846或R874。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于

PbCas13b的以下氨基酸:H567、H500或G566。在一些情况下,所述Cas13 蛋白在螺旋结构域1-2中包含一个或多个氨基酸突变,所述氨基酸对应于 PbCas13b的螺旋结构域1-2中的以下氨基酸:H567、H500或G566。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于 PbCas13b的以下氨基酸:K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。在一些情况下,所述Cas13 蛋白在螺旋结构域1-3中包含一个或多个氨基酸突变,所述氨基酸对应于 PbCas13b的螺旋结构域1-3中的以下氨基酸:K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于 PbCas13b的以下氨基酸:R762、V795、A796、R791、S757或N756。在一些情况下,所述Cas13蛋白在螺旋结构域1-3中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域1-3中的以下氨基酸:R762、V795、A796、R791、S757或N756。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:S658、N653、A656、K655、N652、K590、R638或K741。在一些情况下,所述Cas13 蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于 PbCas13b的螺旋结构域2中的以下氨基酸:S658、N653、A656、K655、N652、K590、R638或K741。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:T405、H407、N486、K484、N480、H452、N455或K457。在一些情况下,所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b 的所述LID结构域中的以下氨基酸:T405、H407、N486、K484、N480、H452、N455或K457。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或K741。

[0663] 在一些情况下,所述Cas13蛋白在螺旋结构域中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域中的以下氨基酸:S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或 K741。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。在一些情况下,所述Cas13蛋白在螺旋结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域1中的以下氨基酸:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:H567、H500、R762、R791、G566、S757或N756。在一些情况下,所述Cas13蛋白在螺旋结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域1中的以下氨基酸:H567、H500、R762、R791、G566、S757或N756。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下氨基酸:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。在一些情况下,所述Cas13蛋白在螺旋结构域1-3中包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的螺旋结构域1-3中的以下氨基酸:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于PbCas13b的以下

氨基酸：R762、R791、S757或N756。在一些情况下，所述Cas13蛋白在螺旋结构域 1-3中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的螺旋结构域1-3中的以下氨基酸：R762、R791、S757或N756。

[0664] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：S658、N653、K655、N652、K590、R638 或K741。在一些情况下，所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的螺旋结构域2中的以下氨基酸：S658、N653、K655、N652、K590、R638或K741。

[0665] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：H407、N486、K484、N480、H452、N455 或K457。在一些情况下，所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述LID结构域中的以下氨基酸：H407、N486、K484、N480、H452、N455或K457。

[0666] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：R56、N157、H161、R1068、N1069或 H1073。在一些情况下，所述Cas13蛋白在所述HEPN结构域中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述HEPN结构域中的以下氨基酸：R56、N157、H161、R1068、N1069或H1073。

[0667] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：R56、N157或H161。在一些情况下，所述Cas13蛋白在所述HEPN结构域1中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述HEPN结构域1中的以下氨基酸：R56、N157 或H161。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：R1068、N1069或H1073。在一些情况下，所述Cas13蛋白在所述HEPN结构域2中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述HEPN结构域2中的以下氨基酸：R1068、N1069或H1073。

[0668] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。在一些情况下，所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述LID结构域中的以下氨基酸：K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。在一些情况下，所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述LID结构域中的以下氨基酸：K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。

[0669] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或 N482。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

[0670] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于

PbCas13b的以下氨基酸：S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

[0671] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：N486、K484、N480、H452、N455或K457。在一些情况下，所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述LID结构域中的以下氨基酸：N486、K484、N480、H452、N455或K457。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：K393、R402、N482、N486、K484、N480、H452、N455或K457。在一些情况下，所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的所述LID结构域中的以下氨基酸：K393、R402、N482、N486、K484、N480、H452、N455或K457。

[0672] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于PbCas13b的以下氨基酸：S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

[0673] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸：R53、Y164、K943 或R1041。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸：R53或Y164。

[0674] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸：K943或R1041。在一些情况下，所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的HEPN结构域中的以下氨基酸：R53、Y164、K943或R1041。在一些情况下，所述Cas13蛋白在HEPN结构域1中包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的HEPN结构域1中的以下氨基酸：R53或Y164。在一些情况下，所述Cas13蛋白在HEPN结构域2中包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的HEPN结构域2 中的以下氨基酸：K943或R1041。

[0675] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸：R53、Y164、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

[0676] 在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸：R53、Y164、R56、N157或H161。在一些情况下，所述Cas13蛋白包含一个或多个氨基酸突变，所述氨基酸对应于颊普雷沃菌PbCas13b

(PbCas13b)的以下氨基酸:K943、R1041、R1068、N1069或H1073。在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域中的以下氨基酸:R53、Y164、K943、R1041、R56、N157、H161、R1068、N1069或H1073。在一些情况下,所述Cas13蛋白在HEPN结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域1中的以下氨基酸:R53、Y164、R56、N157或H161。在一些情况下,所述Cas13蛋白在HEPN结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域2中的以下氨基酸:K943、R1041、R1068、N1069或H1073。

[0677] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的以下氨基酸:R53、Y164、K183、K193、K943或R1041。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的以下氨基酸:R53、Y164、K183或K193。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的以下氨基酸:K943或R1041。在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域中的以下氨基酸:R53、Y164、K183、K193、K943或R1041。

[0678] 在一些情况下,所述Cas13蛋白在HEPN结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域1中的以下氨基酸:R53、Y164、K183或K193。在一些情况下,所述Cas13蛋白在HEPN结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域2中的以下氨基酸:K943或R1041。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的以下氨基酸:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069或H1073。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的以下氨基酸:R53、Y164、K183、K193、R56、N157或H161。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的以下氨基酸:K943、R1041、R1068、N1069或H1073。在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域中的以下氨基酸:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069或H1073。在一些情况下,所述Cas13蛋白在HEPN结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域1中的以下氨基酸:R53、Y164、K183、K193、R56、N157或H161。

[0679] 在一些情况下,所述Cas13蛋白在HEPN结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域2中的以下氨基酸:K943、R1041、R1068、N1069或H1073。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的以下氨基酸:K183或K193。在一些情况下,所述Cas13蛋白在HEPN结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b(PbCas13b)的HEPN结构域1中的以下氨基酸:K183或K193。

[0680] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊

普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:R53、Y164、K943 或R1041。在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的HEPN 结构域中的以下氨基酸:R53、Y164、K943或R1041。在一些情况下,所述 Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的以下氨基酸:R53、K943或R1041;优选地,R53A、R53K、R53D或R53E;K943A、K943R、K943D或K943E;或R1041A、R1041K、R1041D或R1041E。

[0681] 在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的HEPN结构域中的以下氨基酸:R53、K943或R1041;优选地,R53A、R53K、R53D 或R53E;K943A、K943R、K943D或K943E;或R1041A、R1041K、R1041D或R1041E。

[0682] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的氨基酸Y164,优选地,Y164A、Y164F或Y164W。在一些情况下,所述Cas13蛋白在HEPN结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的氨基酸Y164,优选地,Y164A、Y164F或Y164W。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398或E399。

[0683] 在一些情况下,所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的所述LID 结构域中的以下氨基酸:T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398或E399。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的氨基酸H407,优选地,H407Y、H407W或H407F。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:R402、K393、R482、N480、D396、E397、D398或E399。在一些情况下,所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的所述LID结构域中的以下氨基酸:R402、K393、R482、N480、D396、E397、D398或E399。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:K457、D434或K431。在一些情况下,所述Cas13蛋白在所述LID结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的所述LID结构域中的以下氨基酸:K457、D434或 K431。

[0684] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。在一些情况下,所述Cas13蛋白在螺旋结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域中的以下氨基酸:H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸

突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的以下氨基酸:H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些情况下,所述Cas13蛋白在螺旋结构域 1中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的螺旋结构域1中的以下氨基酸:H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836,或R838。

[0685] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:H500、K570、N756、S757、R762或R791。在一些情况下,所述Cas13蛋白在螺旋结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的螺旋结构域1中的以下氨基酸:H500、K570、N756、S757、R762或R791。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸: K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些情况下,所述Cas13蛋白在所述螺旋桥接结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的所述螺旋桥接结构域中的以下氨基酸:K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或 R838。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:H500或K570。在一些情况下,所述Cas13蛋白在螺旋结构域1-2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域1-2 中的以下氨基酸:H500或K570。

[0686] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些情况下,所述Cas13蛋白在螺旋结构域1-3中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的螺旋结构域1-3中的以下氨基酸:N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸: N756、S757、R762或R791。在一些情况下,所述Cas13蛋白在螺旋结构域 1-3中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b) 的螺旋结构域1-3中的以下氨基酸:N756、S757、R762 或R791。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸: N756、S757、R762、R791、K846、K857、K870或R877。在一些情况下,所述Cas13蛋白在螺旋结构域1-3中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域1-3中的以下氨基酸:N756、S757、R762、R791、K846、K857、K870或R877。

[0687] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:K826、K828、K829、R824、R830、Q831、K835、K836或R838。在一些情况下,所述Cas13蛋白在螺旋结构域1-3中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域1-3中的以下氨基酸:

K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[0688] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653或N652。在一些情况下,所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域2中的以下氨基酸:K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653或N652。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:Q646或N647。在一些情况下,所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域2中的以下氨基酸:Q646或N647。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:N653或N652。在一些情况下,所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域2中的以下氨基酸:N653或N652。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:K590、N634、R638、N652、N653、K655、S658、K741或K744。在一些情况下,所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域2中的以下氨基酸:K590、N634、R638、N652、N653、K655、S658、K741或K744。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:R600、K607、K612、R614、K617或R618。在一些情况下,所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的螺旋结构域2中的以下氨基酸:R600、K607、K612、R614、K617或R618。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:R285、R287、K292、E296、N297或K294。在一些情况下,所述Cas13蛋白在所述IDL结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的所述IDL结构域中的以下氨基酸:R285、R287、K292、E296、N297或K294。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:R285、K292、E296或N297。在一些情况下,所述Cas13蛋白在所述IDL结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的所述IDL结构域中的以下氨基酸:R285、K292、E296或N297。

[0689] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:T405、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R285、R287、K292、E296、N297、Q646、N647或K294。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b) 的以下氨基酸:R402、K393、N653、N652、R482、N480、D396、E397、D398或E399。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b

(PbCas13b)的以下氨基酸:R53、K655、R762或R1041;优选地,R53A或R53D;K655A;R762A;或R1041E 或R1041D。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的以下氨基酸:N297、E296、K292或R285;优选地,N297A、E296A、K292A或R285A。在一些情况下,所述Cas13蛋白在所述IDL结构域(例如其中中央通道)中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的所述 IDL结构域(例如其中中央通道)中的以下氨基酸:N297、E296、K292或R285;优选地,N297A、E296A、K292A或R285A。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b)的以下氨基酸:Q831、K836、R838、N652、N653、R830、K655或R762;优选地,Q831A、K836A、R838A、N652A、N653A、R830A、K655A或R762A。

[0690] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的以下氨基酸:N652、N653、R830、K655或R762;优选地,N652A、N653A、R830A、K655A或R762A。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的以下氨基酸:K655或R762;优选地,K655A或R762A。在一些情况下,所述Cas13蛋白在螺旋结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的螺旋结构域中的以下氨基酸:Q831、K836、R838、N652、N653、R830、K655 或R762;优选地,Q831A、K836A、R838A、N652A、N653A、R830A、K655A 或R762A。在一些情况下,所述Cas13蛋白在螺旋结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的螺旋结构域中的以下氨基酸:N652、N653、R830、K655或R762;优选地,N652A、N653A、R830A、K655A或R762A。

[0691] 在一些情况下,所述Cas13蛋白在螺旋结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的螺旋结构域2 中的以下氨基酸:K655或R762;优选地,K655A或R762A。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b)的以下氨基酸:R614、K607、K193、K183或R600;优选地,R614A、K607A、K193A、K183A或R600A。在一些情况下,所述 Cas13蛋白在螺旋结构域2的反式亚基环中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的螺旋结构域2的反式亚基环中的以下氨基酸:Q646或N647;优选地,Q646A或N647A。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌PbCas13b (PbCas13b)的以下氨基酸:R53或R1041;优选地,R53A或 R53D,或R1041E或R1041D。在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b)的HEPN结构域中的以下氨基酸:R53或R1041;优选地,R53A或R53D,或R1041E或R1041D。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b)的以下氨基酸:K457、D397、E398、D399、E400、T405、H407或D434;优选地,D397A、E398A、D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。在一些情况下,所述Cas13蛋白在所述 LID结构域中包含一个或多个氨基酸突变,所述氨基酸对应于颊普雷沃菌 PbCas13b (PbCas13b)的所述LID结构域中的以下氨基酸:K457、D397、E398、D399、E400、T405、H407或D434;优选地,D397A、E398A、D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。

[0692] 在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌





酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸D398。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸E399。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸K294。在一些情况下,所述Cas13 蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸E400。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸R56。在一些情况下,所述Cas13 蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸N157。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸H161。在一些情况下,所述 Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸H452。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸N455。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌 Cas13b (PbCas13b) 的氨基酸K484。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸N486。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸G566。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸H567。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸A656。在一些情况下,所述Cas13 蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸V795。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸A796。在一些情况下,所述 Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸W842。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸K871。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌 Cas13b (PbCas13b) 的氨基酸E873。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸R874。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸R1068。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸 N1069。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b (PbCas13b) 的氨基酸H1073。

[0695] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于沙氏纤毛菌Cas13b (PbCas13b) 的以下氨基酸:R597、N598、H602、R1278、N1279或H1283。本公开还包括一种包含一个或多个氨基酸突变的突变Cas13蛋白,所述氨基酸对应于沙氏纤毛菌Cas13a (LshCas13a) 的以下氨基酸:R597、N598、H602、R1278、N1279或H1283。在一些情况下,所述 Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于沙氏纤毛菌Cas13a (LshCas13a) 的HEPN结构域中的以下氨基酸:R597、N598、H602、R1278、N1279或H1283。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于沙氏纤毛菌Cas13b (PbCas13b) 的以下氨基酸:R597、N598或H602。在一些情况下,所述Cas13蛋白在 HEPN结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于沙氏纤毛菌Cas13a (LshCas13a) 的HEPN结构域1中的以下氨基酸:R597、N598或H602。在一些情况下,所述Cas13

蛋白包含一个或多个氨基酸突变,所述氨基酸对应于沙氏纤毛菌Cas13b(PbCas13b)的以下氨基酸:R1278、N1279或H1283。在一些情况下,所述Cas13蛋白在HEPN结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于沙氏纤毛菌Cas13a(LshCas13a)的HEPN结构域2中的以下氨基酸:R1278、N1279或H1283。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于古吡啉单胞菌Cas13b(PguCas13b)的以下氨基酸:R146、H151、R1116或H1121。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于古吡啉单胞菌Cas13b(PguCas13b)的以下氨基酸:R146、H151、R1116或H1121。在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于古吡啉单胞菌Cas13b(PguCas13b)的HEPN结构域中的以下氨基酸:R146、H151、R1116,或H1121。

[0696] 在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于古吡啉单胞菌Cas13b(PguCas13b)的以下氨基酸:R146或H151。在一些情况下,所述Cas13蛋白在HEPN结构域1中包含一个或多个氨基酸突变,所述氨基酸对应于古吡啉单胞菌Cas13b(PguCas13b)的HEPN结构域1中的以下氨基酸:R146或H151。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于古吡啉单胞菌Cas13b(PguCas13b)的以下氨基酸:R1116或H1121。在一些情况下,所述Cas13蛋白在HEPN结构域2中包含一个或多个氨基酸突变,所述氨基酸对应于古吡啉单胞菌Cas13b(PguCas13b)的HEPN结构域2中的以下氨基酸:R1116或H1121。在一些情况下,所述Cas13蛋白包含一个或多个氨基酸突变,所述氨基酸对应于普雷沃菌P5-125 Cas13b(PspCas13b)的以下氨基酸:H133或H1058。本公开还提供了一种包含一个或多个氨基酸突变的突变Cas13蛋白,所述氨基酸对应于普雷沃菌P5-125 Cas13b(PspCas13b)的以下氨基酸:H133或H1058。在一些情况下,所述Cas13蛋白在HEPN结构域中包含一个或多个氨基酸突变,所述氨基酸对应于普雷沃菌P5-125 Cas13b(PspCas13b)的HEPN结构域中的以下氨基酸:H133或H1058。

[0697] 在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于普雷沃菌P5-125 Cas13b(PspCas13b)的氨基酸H133。在一些情况下,所述Cas13蛋白在HEPN结构域1中包含氨基酸突变,所述氨基酸对应于普雷沃菌P5-125 Cas13b(PspCas13b)的HEPN结构域1中的氨基酸H133。在一些情况下,所述Cas13蛋白包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b(PbCas13b)的氨基酸H1058。在一些情况下,所述Cas13蛋白在HEPN结构域2中包含氨基酸突变,所述氨基酸对应于颊普雷沃菌Cas13b(PbCas13b)的HEPN结构域2中的氨基酸H1058。

[0698] 本文所述的CRISPR-Cas蛋白可包含一个或多个突变氨基酸。在一些实施方式中,所述氨基酸突变为A、P或V,优选地为A。在一些实施方式中,所述氨基酸突变为疏水性氨基酸。在一些实施方式中,所述氨基酸突变为芳族氨基酸。在一些实施方式中,所述氨基酸突变为带电荷的氨基酸。在一些实施方式中,所述氨基酸突变为带正电荷的氨基酸。在一些实施方式中,所述氨基酸突变为带负电荷的氨基酸。在一些实施方式中,所述氨基酸突变为极性氨基酸。在一些实施方式中,所述氨基酸突变为脂族氨基酸。

[0699] 本公开还提供了改变CRISPR-Cas蛋白的活性的方法。在一些实施例中,此类方法包括基于Cas13蛋白的至少一部分的三维结构识别所述Cas13蛋白中的一个或多个候选氨基酸,其中所述一个或多个候选氨基酸与和所述Cas13蛋白形成复合物的引导RNA相互作用,或在所述Cas13蛋白的HEPN活性位点、结构域间接头结构域或桥接螺旋结构域中;和使

所述一个或多个候选氨基酸突变,从而产生突变Cas13蛋白,其中所述突变Cas13蛋白的活性不同于所述Cas13蛋白。

[0700] 不稳定的CAS13和融合蛋白

[0701] 在某些实施方式中,如本文所述,根据本发明所述的效应子蛋白与去稳定结构域(DD)相关联或融合。在一些实施方式中,所述DD是ER 50。

[0702] 在一些实施方式中,所述DD的对应的稳定配体是4HT。这样,在一些实施方式中,至少一个DD之一是ER 50,并且其稳定配体是4HT或CMP8。在一些实施方式中,所述DD是DHFR50。在一些实施方式中,所述DD的对应的稳定配体是TMP。这样,在一些实施方式中,至少一个DD之一是 DHFR50,并且其稳定配体是TMP。在一些实施方式中,所述DD是ER 50。在一些实施方式中,所述DD的对应的稳定配体是CMP8。因此,CMP8可以是所述ER50系统中的4HT的替代稳定配体。尽管可能可以/应该竞争性方式使用CMP8和4HT,但某些细胞类型可能更易受这两种配体中的一种或另一种影响,并且技术人员根据本公开内容和本领域中的知识可以使用 CMP8和/或4HT。

[0703] 在一些实施方式中,可以将一个或两个DD融合至所述Cas13的N末端,其中将一个或两个DD融合至Cas13的C末端。在一些实施方式中,所述至少两个DD与所述Cas13关联,并且所述DD是相同的DD,即所述DD是同系的。因此,所述DD中的两者(或两个或多个)可以是ER50 DD。这在一些实施方式中是优选的。或者,所述DD中的两者(或两个或多个)可以是DHFR50 DD。在一些实施方式中,这也是优选的。在一些实施方式中,所述至少两个DD与所述Cas13关联,并且所述DD是不同的DD,即所述DD 是异源的。因此,所述DD之一可以是ER50,而所述DD中的一个或多个或任何其他DD可以是DHFR50。具有两个或更多个异源的DD可以是有利的,因为它将提供更高水平的降解控制。在N或C末端处的多于一个的DD 的衔接融合可以增强降解;并且这种衔接融合可以是例如ER50-ER50-Cas13 或DHFR-DHFR-Cas13。设想在不存在任何稳定配体的情况下会发生高水平的降解,在不存在一个稳定配体和存在其他(或另一个)稳定配体的情况下会发生中等水平的降解,而在所述稳定配体中的两者(或两个或多个)存在的情况下会发生低水平的降解。也可以通过具有N末端ER50 DD和C末端DHFR50 DD来给予控制。

[0704] 在一些实施方式中,所述Cas13与所述DD的融合体包含所述DD和所述Cas13之间的接头。在一些实施方式中,所述接头是GlySer接头。在一些实施方式中,所述DD-Cas13还包含至少一个核输出信号(NES)。在一些实施方式中,所述DD-Cas13包含两个或更多个NES。在一些实施方式中,所述 DD-Cas13包含至少一个核定位信号(NLS)。这可以是NES的补充。在一些实施方式中,所述Cas13包含定位(核输入或输出)信号或基本上由其组成或由其组成,所述定位(核输入或输出)信号作为所述Cas13和所述DD之间的所述接头或其部分)。HA或Flag标记作为接头也在本发明的范围内。申请人使用NLS和/或NES作为接头,并且还使用短至GS长至(GGGGS)3的甘氨酸丝氨酸接头。

[0705] 去稳定结构域具有赋予多种蛋白不稳定性的普遍用途;参见例如 Miyazaki, J Am Chem Soc. Mar 7, 2012; 134 (9) : 3942-3945, 其通过引用并入本文。CMP8或4-羟基他莫昔芬可以是去稳定结构域。更普遍而言,发现哺乳动物DHFR的温度敏感突变体(DHFRts),遵循N末端规则的去稳定残基,在允许的温度下稳定,但在37°C下不稳定。在表达DHFRts的细胞中加入甲氨蝶呤(哺乳动物DHFR的高亲和力配体)可部分抑制所述蛋白的降解。这是一个重要

的证明,即小分子配体可以稳定蛋白,否则该蛋白将靶向细胞降解。雷帕霉素衍生物用于稳定mTOR的FRB结构域(FRB\*)的不稳定突变体并恢复融合激酶GSK-3 $\beta$ .6和7的功能。该系统证明配体依赖性稳定性代表了一种有吸引力的调节特定蛋白在复杂生物环境中的功能的策略。当雷帕霉素诱导的FK506结合蛋白和FKBP12的二聚化引起泛素互补时,控制蛋白活性的系统可能涉及DD发挥作用。可以将人FKBP12或ecDHFR蛋白的突变体工程化为在缺乏高亲和力配体Shield-1或甲氧苄啶(TMP)的情况下在代谢上不稳定。这些突变体是可用于实施本发明的一些可能的去稳定结构域(DD),并且DD与Cas13融合时的不稳定性赋予Cas13通过蛋白酶体对整个融合蛋白的降解。Shield-1和TMP以剂量依赖性方式结合并稳定DD。雌激素受体配体结合结构域(ERLBD,ERS1的残基305-549)也可以被工程化为去稳定结构域。由于雌激素受体信号转导通路涉及多种疾病,例如乳腺癌,因此已经对该通路进行了广泛研究,并且已经开发出许多雌激素受体激动剂和拮抗剂。因此,已知ERLBD和药物的相容对。有些配体与ERLBD的突变型而非野生型结合。通过使用编码三个突变(L384M、M421G、G521R)的12的这些突变结构域之一,使用不干扰内源雌激素敏感性网络的配体来调节ERLBD衍生DD的稳定性是可能的。可以引入一个额外的突变(Y537S)以进一步破坏ERLBD的稳定性并将其配置为潜在的DD候选物。该四突变体是有利的DD发展。所述突变ERLBD可以与Cas13融合,并且其稳定性可以使用配体来调节或扰动,其中所述Cas13具有DD。另一DD可以是基于Shield1配体稳定的突变FKBP蛋白的12kDa(107-氨基酸)标记;参见例如Nature Methods 5,(2008)。例如,DD可以是修饰的FK506结合蛋白12(FKBP12),所述修饰的FK506结合蛋白12与合成的生物学惰性小分子Shield-1结合并由其可逆地稳定;参见例如Banaszynski LA,Chen LC,Maynard-Smith LA,Ooi AG,Wandless TJ.A rapid,reversible,and tunable method to regulate protein function in living cells using synthetic small molecules.Cell.2006;126:995-1004;Banaszynski LA,Sellmyer MA,Contag CH,Wandless TJ,Thorne SH.Chemical control of protein stability and function in living mice.Nat Med. 2008;14:1123-1127;Maynard-Smith LA,Chen LC,Banaszynski LA,Ooi AG,Wandless TJ.A directed approach for engineering conditional protein stability using biologically silent small molecules.The Journal of biological chemistry. 2007;282:24866-24872;和Rodriguez,Chem Biol.Mar 23,2012;19(3):391-398—所有这些内容均通过引用并入本文,并且可以在本发明的实践中的选定的DD与Cas13关联中使用。可以看出,本领域的知识包括许多DD,并且所述DD可以与Cas13关联(例如有利地与接头融合),由此可以在存在配体的情况下使DD稳定,并且在其不存在时,DD可变得不稳定,从而使Cas13完全不稳定,或者可以在不存在配体的情况下使DD稳定,并且在存在配体时,DD可变得不稳定;所述DD允许Cas13被调控或控制(换言之,打开或关闭),从而允许CRISPR-Cas13复合体或系统被调控或控制(换言之,打开或关闭),从而提供了用于例如在体内或体外环境中调节或控制系统的手段。例如,当感兴趣的蛋白表达为与DD标记的融合体时,它在细胞中去稳定化并迅速降解,例如通过蛋白酶体。因此,稳定配体的缺失导致D关联的Cas被降解。当新的DD与感兴趣的蛋白融合时,其不稳定性赋予给了感兴趣的蛋白,导致整个融合蛋白的迅速降解。Cas的峰值活性有时有利于减少脱靶效应。因此,短时间的高活性爆发是优选的。本发明能够提供这样的峰。在某种意义上,所述系统是可诱导的。在一些其他意义上,所述系统在不存在稳定配体的情况下被抑制,并在不存在稳定配体

的情况下被去抑制化。

[0706] 死的CAS蛋白

[0707] 在某些实施方式中,本文所述的效应子蛋白是催化失活的或死的Cas蛋白。在一些情况下,根据如本文所述的发明的效应子蛋白(CRISPR酶;Cas13;效应子蛋白)是催化失活或死的Cas13效应子蛋白(dCas13)。在一些情况下,死的Cas蛋白(例如死的Cas13蛋白)具有切口酶活性。在一些实施方式中,所述dCas13效应子包含在所述核酸酶结构域中的突变。在一些实施方式中,所述dCas13效应子蛋白已被截短。在一些情况下,所述死的Cas蛋白可以与本文所述的脱氨酶(例如腺苷脱氨酶)融合。

[0708] 为了减小所述Cas13效应子和所述一个或多个功能结构域的融合蛋白的大小,可以截短所述Cas13效应子的C末端而仍保持其RNA结合功能。例如,可以在所述Cas13效应子的C末端处截短至少20个氨基酸,至少40个氨基酸,至少50个氨基酸,至少60个氨基酸,至少80个氨基酸,至少100个氨基酸,至少120个氨基酸,至少140个氨基酸,至少150个氨基酸,至少160个氨基酸,至少180个氨基酸,至少200个氨基酸,至少220个氨基酸,至少240个氨基酸,至少250个氨基酸,至少260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。Cas13截短的具体实例包括C末端 $\Delta$ 984-1090,C末端 $\Delta$ 1026-1090和C末端 $\Delta$ 1053-1090,C末端 $\Delta$ 934-1090,C末端 $\Delta$ 884-1090,C末端 $\Delta$ 834-1090,C末端 $\Delta$ 784-1090和C末端 $\Delta$ 734-1090,其中氨基酸位置对应于普雷沃菌P5-125 Cas13b蛋白的氨基酸位置。技术人员将明白可为其他Cas13b直系同源物或其他Cas13类型或亚类型(例如Cas13a、Cas13c或Cas13d)设计相似的截短。在一些情况下,截短的Cas13b由普雷沃菌P5-125 Cas13b的nt 1-984或者Cas13b直系同源物或同系物的对应的nt编码。Cas13截短的实例还包括C末端 $\Delta$ 795-1095,其中氨基酸位置对应于鸭疫里默氏杆菌Cas13b蛋白的氨基酸位置。Cas13截短的实例进一步包括C末端 $\Delta$ 875-1175、C末端 $\Delta$ 895-1175、C末端 $\Delta$ 915-1175、C末端 $\Delta$ 935-1175、C末端 $\Delta$ 955-1175、C末端 $\Delta$ 975-1175、C末端 $\Delta$ 995-1175、C末端 $\Delta$ 1015-1175、C末端 $\Delta$ 1035-1175、C末端 $\Delta$ 1055-1175、C末端 $\Delta$ 1075-1175、C末端 $\Delta$ 1095-1175、C末端 $\Delta$ 1115-1175、C末端 $\Delta$ 1135-1175、C末端 $\Delta$ 1155-1175,其中氨基酸位置对应于古吡啉单胞菌Cas13b蛋白的氨基酸位置。

[0709] 在一些实施方式中,可以截短所述Cas13效应子蛋白的N末端。例如,可以在所述Cas13效应子的N末端处截短至少20个氨基酸,至少40个氨基酸,至少50个氨基酸,至少60个氨基酸,至少80个氨基酸,至少100个氨基酸,至少120个氨基酸,至少140个氨基酸,至少150个氨基酸,至少160个氨基酸,至少180个氨基酸,至少200个氨基酸,至少220个氨基酸,至少240个氨基酸,至少250个氨基酸,至少260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。Cas13截短的实例包括N末端 $\Delta$ 1-125、N末端 $\Delta$ 1-88或N末端 $\Delta$ 1-72,其中所述截短的氨基酸位置对应于普雷沃菌P5-125 Cas13b蛋白的氨基酸位置。

[0710] 在一些实施方式中,可以截短所述Cas13效应子蛋白的N或C末端两者。例如,可以在所述Cas13效应子的C末端处截短至少20个氨基酸,并且可以在所述Cas13效应子的N末端







在所述Cas13效应子的C末端处截短至少20个氨基酸,至少40个氨基酸,至少60个氨基酸,至少80个氨基酸,至少100个氨基酸,至少120个氨基酸,至少140个氨基酸,至少160个氨基酸,至少180个氨基酸,至少200个氨基酸,至少220个氨基酸,至少240个氨基酸,至少260个氨基酸,至少300个氨基酸或至少350个氨基酸。例如,可以在所述Cas13效应子的N末端处截短至少260个氨基酸,并且可以在所述Cas13效应子的C末端处截短至少20个氨基酸,至少40个氨基酸,至少60个氨基酸,至少80个氨基酸,至少100个氨基酸,至少120个氨基酸,至少140个氨基酸,至少160个氨基酸,至少180个氨基酸,至少200个氨基酸,至少220个氨基酸,至少240个氨基酸,至少260个氨基酸,至少300个氨基酸或至少350个氨基酸。例如,可以在所述Cas13效应子的N末端处截短至少280个氨基酸,并且可以在所述Cas13效应子的C末端处截短至少20个氨基酸,至少40个氨基酸,至少60个氨基酸,至少80个氨基酸,至少100个氨基酸,至少120个氨基酸,至少140个氨基酸,至少160个氨基酸,至少180个氨基酸,至少200个氨基酸,至少220个氨基酸,至少240个氨基酸,至少260个氨基酸,至少300个氨基酸或至少350个氨基酸。例如,可以在所述Cas13效应子的N末端处截短至少300个氨基酸,并且可以在所述Cas13效应子的C末端处截短至少20个氨基酸,至少40个氨基酸,至少60个氨基酸,至少80个氨基酸,至少100个氨基酸,至少120个氨基酸,至少140个氨基酸,至少160个氨基酸,至少180个氨基酸,至少200个氨基酸,至少220个氨基酸,至少240个氨基酸,至少260个氨基酸,至少300个氨基酸或至少350个氨基酸。例如,可以在所述Cas13效应子的N末端处截短至少350个氨基酸,并且可以在所述Cas13效应子的C末端处截短至少20个氨基酸,至少40个氨基酸,至少60个氨基酸,至少80个氨基酸,至少100个氨基酸,至少120个氨基酸,至少140个氨基酸,至少160个氨基酸,至少180个氨基酸,至少200个氨基酸,至少220个氨基酸,至少240个氨基酸,至少260个氨基酸,至少300个氨基酸或至少350个氨基酸。

#### [0711] 分裂蛋白

[0712] 注意在该背景下并且对于本文所述的各种应用而言更普遍地,可以设想使用RNA靶向的效应子蛋白的分裂形式。确实,这不仅可以增加特异性,而且对于递送也可以是有利的。从Cas13酶的两个部分基本上包含功能性Cas13的意义上说,Cas13是分裂的。理想情况下,所述分裂应始终使催化结构域不受影响。所述Cas13可以充当核酸酶,或者它可以是死的Cas13,所述死的Cas13由于其催化结构域中的典型突变而本质上是具有非常少或没有催化活性的RNA结合蛋白。

[0713] 分裂的Cas13的每一半可以与二聚化配体融合。举例来说而不受限制地,采用雷帕霉素敏感的二聚化结构域允许产生化学上可诱导的分裂的Cas13以用于暂时控制Cas13的活性。因此,可以通过将Cas13分裂为两个片段来使其变为化学上可诱导的,并且雷帕霉素敏感的二聚化结构域可以用于所述Cas13的受控重组。所述分裂的Cas13的所述两个部分可以被认为是所述分裂的Cas13的N'末端部分和C'末端部分。所述融合通常在所述Cas13的分裂点。换言之,所述分裂的Cas13的所述N'末端部分的所述C'末端与所述二聚体的一半融合,而所述C'末端部分的N'末端与所述二聚体的另一半融合。

[0714] 从创建新分裂的意义上说,不必分裂所述Cas13。分裂点通常是由计算机设计的并被克隆到构建体中。所述分裂的Cas13的所述两个部分(所述N'末端和所述C'末端部分)一起构成完整的Cas13,所述完整的Cas13包含优选地至少70%或更多的所述野生型氨基酸(或编码它们的核苷酸),优选地至少80%或更多、优选地至少90%或更多、优选地至少

95%或更多和最优选地至少99%或更多的所述野生型氨基酸(或编码它们的核苷酸)。一些修整是可能的,并且设想了突变体。非功能性结构域可被完全去除。重要的是两个部分可以放于一起,以及恢复或重构所需的Cas13功能。所述二聚体可以是同二聚体或异二聚体。

[0715] 在某些实施方式中,如本文所述的Cas13效应子可用于例如突变特异性或等位基因特异性靶向,例如用于突变特异性或等位基因特异性敲减。

[0716] 此外,可将所述RNA靶向的效应子蛋白与另一功能性RNAe结构域(例如非特异性RNAe或Argonaute 2)融合,所述RNA靶向的效应子蛋白与所述另一功能性RNAe结构域协同作用以增加所述RNAe活性或确保所述消息的进一步降解。

[0717] 调节Cas13效应子蛋白

[0718] 本发明提供了调节CRISPR蛋白功能的辅助蛋白。在某些实施方式中,所述辅助蛋白调节CRISPR蛋白的催化活性。在本发明的实施方式中,辅助蛋白调节靶向的或序列特异性的核酸酶活性。在本发明的实施方式中,辅助蛋白调节附带核酸酶活性。在本发明的实施方式中,辅助蛋白调节与靶核酸的结合。

[0719] 根据本发明,待调节的核酸酶活性可以针对包含RNA或由RNA组成的核酸,包括但不限于mRNA、miRNA、siRNA和包含可切割的RNA链接以及核酸类似物的核酸。在本发明的实施方式中,所述待调节的核酸酶活性可以针对包含DNA或由DNA组成的核酸,包括但不限于包含可切割的DNA 链接和核酸类似物的核酸。

[0720] 在本发明的实施方式中,辅助蛋白增强CRISPR蛋白的活性。在某些这样的实施方式中,所述辅助蛋白包含HEPN结构域并增强RNA切割。在某些实施方式中,所述辅助蛋白抑制CRISPR蛋白的活性。在某些这样的实施方式中,所述辅助蛋白包含失活的HEPN结构域或完全缺失HEPN结构域。

[0721] 根据本发明,VI型CRISPR系统的天然存在的辅助蛋白包含在CRISPR 基因座处或附近编码的小蛋白,所述小蛋白的功能是修饰CRISPR蛋白的活性。通常,CRISPR基因座可被识别为包含推定的CRISPR阵列和/或编码推定的CRISPR效应子蛋白。在实施方式中,效应子蛋白可以是800至2000 个氨基酸,或900至1800个氨基酸,或950至1300个氨基酸。在实施方式中,辅助蛋白可以在推定的CRISPR效应子蛋白或阵列的25kb内或20kb 内或15kb内或10kb内,或在推定的CRISPR效应子蛋白或阵列的2kb至 10kb的范围内编码。

[0722] 在本发明的实施方式中,辅助蛋白是50至300个氨基酸,或100至300 个氨基酸或150至250个氨基酸或约200个氨基酸。辅助蛋白的非限制性实例包括本文所识别的csx27和csx28蛋白。

[0723] 本发明的CRISPR辅助蛋白的识别和使用不依赖于CRISPR效应子蛋白分类。可以发现本发明所述的辅助蛋白与多种CRISPR效应子蛋白关联或被设计为与多种CRISPR效应子蛋白共同起作用。本文所识别并使用的辅助蛋白的实例通常代表CRISPR效应子蛋白。应当理解,CRISPR效应子蛋白分类可涉及同源性,特征位置(例如REC结构域、NUC结构域、HEPN序列的位置),核酸靶标(例如DNA或RNA),tracr RNA的缺失或存在,正向重复的引导/间隔序列5' 或3' 的位置,或其他条件。在本发明的实施方式中,辅助蛋白的识别和使用超越此类分类。

[0724] 在靶向RNA的VI型CRISPR-Cas系统中,所述Cas蛋白通常包含两个保守的HEPN结构域,所述保守的HEPN结构域参与RNA切割。在某些实施方式中,所述Cas蛋白加工crRNA以产

生成成熟的crRNA。所述crRNA的引导序列识别具有互补序列的靶RNA,并且所述Cas蛋白降解靶链。更具体地,在某些实施方式中,在靶结合之后,所述Cas蛋白进行结构重排,所述结构重排集合两个HEPN结构域以形成活性HEPN催化位点,然后切割所述靶RNA。在所述Cas蛋白表面附近的所述催化位点的位置允许非特异性的附带ssRNA切割。

[0725] 在某些实施方式中,辅助蛋白在增加或减少靶标和/或附带RNA切割中至关重要。不受理论束缚地,激活CRISPR活性的辅助蛋白(例如csx28蛋白或直系同源物或包含HEPN结构域的变体)可被设想为能够与Cas蛋白相互作用并且能够将其HEPN结构域与Cas蛋白的HEPN结构域结合以形成活性HEPN催化位点,而抑制性辅助蛋白(例如缺失HEPN结构域的csx27)可被设想为能够与Cas蛋白相互作用并减少或阻断会集合两个HEPN结构域的Cas蛋白的构象。

[0726] 根据本发明,在某些实施方式中,增强VI型Cas蛋白或其复合物的活性包括使所述VI型Cas蛋白或其复合物与来自激活所述Cas蛋白的同一生物的辅助蛋白接触。在其他实施方式中,增强其复合物的VI型Cas蛋白的活性包括使所述VI型Cas蛋白或其复合物与来自同一亚纲内的不同生物的激活剂辅助蛋白(例如VI-b型)接触。在其他实施方式中,增强VI型Cas蛋白或其复合物的活性包括使所述VI型Cas蛋白或其复合物与不在所述亚类内的辅助蛋白(例如除具有VI-b型辅助蛋白的VI-b型以外的VI型Cas蛋白或与之相反的)接触。

[0727] 根据本发明,在某些实施方式中,阻遏VI型Cas蛋白或其复合物的活性包括使所述VI型Cas蛋白或其复合物与来自同一生物的阻遏Cas蛋白的辅助蛋白接触。在其他实施方式中,阻遏VI型Cas蛋白或其复合物的活性包括使所述VI型Cas蛋白或其复合物与来自同一亚纲内的不同生物的阻遏辅助蛋白(例如VI-b型)接触。在其他实施方式中,阻遏VI型Cas蛋白或其复合物的活性包括使所述VI型Cas蛋白或其复合物与不在所述亚纲内的阻遏辅助蛋白接触(例如除具有VI-b型阻遏辅助蛋白的VI-b型以外的VI型Cas蛋白或与之相反的)。

[0728] 在所述VI型Cas蛋白和所述VI型辅助蛋白来自同一生物的某些实施方式中,所述两种蛋白将在CRISPR工程化系统中共同发挥作用。在某些实施方式中,期望例如通过修饰所述蛋白中的任一种或两者或它们的表达来改变所述工程化CRISPR系统的功能。在所述VI型Cas蛋白和所述VI型辅助蛋白来自不同生物(所述生物可以在同一纲或不同纲内)的实施方式中,所述蛋白可以在工程化CRISPR系统中一起发挥作用,但是常常期望或要求修饰所述蛋白中的任一种或两者以使其一起发挥作用。

[0729] 因此,在本发明的某些实施方式中,可以修饰Cas蛋白和辅助蛋白中的任一种或两者以调节所述Cas蛋白和所述辅助蛋白之间的蛋白-蛋白相互作用的方面。在某些实施方式中,可以修饰Cas蛋白和辅助蛋白中的任一种或两者以调节蛋白-核酸相互作用的方面。调节蛋白-蛋白相互作用和蛋白-核酸相互作用的方法包括但不限于拟合分子表面、极性相互作用、氢键和调节范德华相互作用。在某些实施方式中,调节蛋白-蛋白相互作用或蛋白-核酸结合包括增加或减少结合相互作用。在某些实施方式中,调节蛋白-蛋白相互作用或蛋白-核酸结合包括偏向或不偏向所述蛋白或所述核酸的构象的修饰。

[0730] “拟合(fitting)”是指确定(包括通过自动或半自动方式)Cas13蛋白的一个或多个原子(和可选地,Cas13辅助蛋白的至少一个原子)之间的,或Cas13蛋白的一个或多个原子与核酸的一个或多个原子之间(或可选地,Cas13辅助蛋白的一个或多个原子与核酸之间)的相互作用,并计算这种相互作用稳定的程度。相互作用包括由电荷、空间考虑因素等

带来的吸引和排斥。

[0731] VI型CRISPR蛋白或其复合物(和/或在Cas13b的背景下的VI型CRISPR 辅助蛋白或其复合物)的三维结构在本发明的背景下提供了另外的用于识别 Cas13的直系同源物中的其他突变的工具。所述晶体结构也可以作为设计新的特定Cas13(和可选地,Cas13辅助蛋白)的基础。进一步描述了各种用于拟合的基于计算机的方法。可以通过使用对接程序(docking program)的计算机建模的使用来检查Cas13(和可选的辅助蛋白)与核酸的结合相互作用。对接程序(例如GRAM、DOCK或AUTODOCK)是已知的(参见Walters等,Drug Discovery Today,vol.3,no.4(1998),160-178,和Dunbrack等,Folding和设计2(1997),27-42)。该程序可以包括计算机拟合以确定所述结合配体的形状和化学结构有多好。可以对VI型系统的活性位点或结合位点进行计算机辅助的手动检查。例如GRID(定具有各种官能团的分子之间的可能的相互作用位点的程序)的程序(P.Goodford,J.Med.Chem,1985,28,849-57)也可以用于分析活性位点或结合位点以预测结合化合物的部分结构。可以使用计算机程序来估计两个结合配体(例如VI型CRISPR系统的组分,或核酸分子和VI型CRISPR系统的组分)的吸引力、排斥力或空间位阻。

[0732] 可以基于氨基酸特性(例如残基的极性、电荷、溶解度、疏水性、亲水性和/或两性性质)的差异或相似性来进行氨基酸置换,因此将氨基酸一起分组于功能组中是有用的。可以仅基于氨基酸的侧链而将它们分组在一起。在比较直系同源物时,由于结构或催化原因,可能存在保守的残基。这些组可以以维恩图的形式描述(Livingstone C.D.and Barton G.J.(1993)“Protein sequence alignments:a strategy for the hierarchical analysis of residue conservation”Comput.Appl.Biosci.9:745-756)(Taylor W.R.(1986)“The classification of amino acid conservation”J.Theor.Biol.119;205-218)。可以例如根据下表来进行保守置换,该表描述了公认的氨基酸的维恩图分组(参见下表7)。

[0733] 表7

| 组   |   | 亚组    |           |
|-----|---|-------|-----------|
| 疏水性 | F W Y H K M I L V A G<br>C<br>(SEQ ID NO:240) | 芳香族的  | F W Y H   |
|     |   | 脂肪族的  | I L V     |
| 极性  | W Y H K R E D C S T N<br>Q<br>(SEQ ID NO:241) | 带电荷的  | H K R E D |
|     |   | 带正电荷的 | H K R     |
|     |   | 带负电荷的 | E D       |
| 小的  | V C A G S P T N D<br>(SEQ ID NO:242)          | 微小    | A G S     |

[0735] 在工程化Cas13系统中,修饰可包括所述Cas13蛋白的一个或多个氨基酸残基的修饰(和/或在Cas13b的情况下可包括所述Cas13辅助蛋白的一个或多个氨基酸残基的修饰)。

[0736] 在工程化Cas13系统中,修饰可包括位于包含在未修饰的Cas13蛋白(和/或Cas13辅助蛋白)中为带正电荷的残基的区域中的一个或多个氨基酸残基的修饰。

[0737] 在工程化Cas13系统中,修饰可包括在所述未修饰的Cas13蛋白(和/或 Cas13辅助蛋白)中为带正电荷的一个或多个氨基酸残基的修饰。

[0738] 在工程化Cas13系统中,修饰可包括在所述未修饰的Cas13蛋白(和/或 Cas13辅助蛋白)中为不带正电荷的一个或多个氨基酸残基。

[0739] 所述修饰可包括修饰未修饰的Cas13蛋白(和/或Cas13辅助蛋白)中为不带电荷的一个或多个氨基酸残基的修饰。

[0740] 所述修饰可包括在所述未修饰的Cas13蛋白(和/或Cas13辅助蛋白)中为带负电荷的一个或多个氨基酸残基的修饰。

[0741] 所述修饰可以包括在未修饰的Cas13蛋白(和/或Cas13辅助蛋白)中为疏水性一个或多个氨基酸残基的修饰。

[0742] 所述修饰可包括在所述未修饰的Cas13蛋白(和/或Cas13辅助蛋白)中为极性的一个或多个氨基酸残基的修饰。

[0743] 所述修饰可包括带电荷的氨基酸对疏水性氨基酸或极性氨基酸的置换,所述带电荷的氨基酸可以是带负电荷的或带正电荷的氨基酸。所述修饰可包括带正电荷的或极性或疏水性氨基酸对带负电荷的氨基酸的置换。所述修饰可包括带负电荷的或极性或疏水性氨基酸对带正电荷的氨基酸的置换。

[0744] 本发明的实施方式包括序列(多核苷酸或多肽两者),所述序列可包含可发生的同源置换(本文使用了置换和取代两者以表示替代的残基或核苷酸对存在的氨基酸残基或核苷酸的交换),即,在氨基酸情况下的类似物间的置换,例如碱性对碱性、酸性对酸性、极性对极性等。也可发生非同源置换,即从一类残基到另一类残基,或者替代地涉及包含非天然的氨基酸,例如鸟氨酸(以下称为Z)、二氨基丁酸鸟氨酸(以下称为B)、正亮氨酸鸟氨酸(以下称为O)、吡啶基丙氨酸、噻吩基丙氨酸、萘基丙氨酸和苯基甘氨酸。变体氨基酸序列可包含可插入所述序列的任两个氨基酸残基之间的合适的间隔基,除了例如甘氨酸或 $\beta$ -丙氨酸残基的氨基酸间隔基外,所述合适的间隔基还包括烷基(例如甲基、乙基或丙基)。本领域技术人员可以很好地理解另一种形式的变异,其涉及类肽形式的一个或多个氨基酸残基的存在。为避免疑问,“所述类肽形式”用于表示变体氨基酸残基,其中 $\alpha$ -碳取代基在残基的氮原子上,而非所述 $\alpha$ -碳上。制备类肽形式的肽的方法是本领域已知的,例如Simon RJ等,PNAS (1992) 89(20),9367-9371和Horwell DC,Trends Biotechnol. (1995) 13(4),132-134。

[0745] 同源性建模:可以通过Zhang等,2012(Nature;490(7421):556-60)和Chen等,2015(PLoS Comput Biol;11(5):e1004248)的方法(计算蛋白-蛋白相互作用(PPI)的方法(预测结构域-基序界面介导的相互作用的计算化蛋白-蛋白相互作用(PPI)方法)来识别其他Cas13直系同源物中的对应的残基。PrePPI(预测PPI)(基于结构的PPI预测方法)使用贝叶斯统计框架(Bayesian statistical framework)来结合结构证据与非结构证据。所述方法涉及获取一对查询蛋白和使用结构比对以识别对应于其实验确定的结构或同源性模型的结构代表。通过考虑整体和局部几何关系,进一步使用结构对比以识别近的和远的结构邻近物两者。每当所述结构代表的两个邻近物形成蛋白数据库(蛋白Data Bank)中报告的复合物时,这定义了用于对两个查询蛋白之间的相互作用进行建模的模板。通过将代

表性结构叠加于他们在模板中的对应的结构邻近物上来创建复杂物的模型。这种方法在 Dey等,2013(Prot Sci;22:359-66) 中。

[0746] 附带活性

[0747] 最近将附带活性用于名为SHERLOCK的高灵敏度和特异性核酸检测平台,所述核酸检测平台可用于许多临床诊断(Gootenberg,J.S.et al.Nucleic acid detection with CRISPR-Cas13a/C2c2.Science 356,438-442(2017))。

[0748] 根据本发明,工程化CRISPR-Cas系统针对RNA核酸内切酶活性进行了优化,并且可以在哺乳动物细胞中表达和靶向以有效敲低细胞中的报告分子或转录物。

[0749] 具有等温扩增的工程化CRISPR-Cas的附带作用提供了基于CRISPR的诊断,所述诊断提供了具有高灵敏度和单碱基错配特异性的快速的DNA或 RNA检测。基于CRISPR-Cas的分子检测平台用于检测特定病毒株,区分病原细菌,对人DNA进行基因分型和鉴定无细胞的肿瘤DNA突变。此外,可以冻干反应试剂以实现冷链独立性和长期保存,并可以很容易地在理论上重构所述反应试剂以用于现场应用。

[0750] 在便携式平台上以高灵敏度和单碱基特异性快速地检测核酸的能力可有助于疾病诊断和监测、流行病学和一般实验室任务。尽管存在用于检测核酸的方法,但是它们在灵敏度、特异性、简单性、成本和速度之中有权衡。

[0751] 微生物的成簇、规则间隔的短回文重复(CRISPR)和CRISPR相关的(CRISPR-Cas)适应性免疫系统包含可编程的核酸内切酶,所述可编程的核酸内切酶可用于基于CRISPR的诊断(CRISPR-Dx)。可以使用CRISPR RNA(crRNA)对CRISPR-Cas进行重编程以提供用于特异性DNA感测的平台。激活的CRISPR-Cas在识别其DNA靶标之后参与附近的非靶标核酸(即RNA和/或ssDNA)的“附带”切割。这种crRNA编程的附带切割活性通过触发程序性细胞死亡或通过标记RNA或ssDNA的非特异性降解来使CRISPR-Cas能够检测体内特异性DNA的存在。在此描述了一种基于核酸扩增和对商购报道分子RNA的CRISPR-Cas介导的附带切割的具有高灵敏度的体外核酸检测平台,所述体外核酸平台允许对靶标的实时检测。

[0752] 非特异性ssDNA和RNA定向蛋白的保守性将不可避免地导致进一步的和潜在的改进的CRISPR蛋白,所述改进的CRISPR蛋白展现附带切割,并可用于检测,并为在扩增的和高度敏感的诊断系统(特别是SHERLOCK)中对核酸靶标进行多重检测提供了更大的广度。

[0753] 基于RNA的掩蔽

[0754] 在某些示例实施方式中,基于RNA的掩蔽构建体抑制可检测的阳性信号的产生,或者所述基于RNA的掩蔽构建体通过掩蔽所述可检测的阳性信号或替代地产生可检测的阴性信号来抑制可检测的阳性信号的产生,或者所述基于RNA的掩蔽构建体包含沉默RNA,所述沉默RNA抑制由报告构建体编码的基因产物的产生,其中所述基因产物在表达时产生可检测的阳性信号。

[0755] 在另一示例实施方式中,所述基于RNA的掩蔽构建体是产生阴性可检测信号的核酶,并且其中当核酶失活时产生阳性可检测信号。在一个示例实施方式中,所述核酶将底物转化为第一颜色,并且其中所述底物在所述核酶失活时转化为第二颜色。在另一示例实施方式中,所述基于RNA的掩蔽试剂是螯合酶的适配体,其中所述酶通过作用于底物而在从所述适配体释放时产生可检测的信号,或者所述适配体螯合一对在从所述适配体释放时结合以产生可检测的信号的试剂。

[0756] 在另一示例实施方式中,所述基于RNA的掩蔽构建体包含RNA寡核苷酸,可检测的配体寡核苷酸和掩蔽组分连接至所述RNA寡核苷酸。在某些示例实施方式中,所述可检测的配体是荧光团,并且所述掩蔽组分是淬灭分子。

[0757] 在另一方面,本发明提供了一种用于检测样品中的靶核酸(例如RNA)的方法,其包括:将样品或样品组分配到一个或多个单独的离散体积中,所述单独的离散体积包括包含CRISPR系统(所述CRISPR系统包含效应子蛋白)、一个或多个引导RNA、基于RNA的掩蔽构建体;在足以允许所述一个或多个引导RNA与一个或多个靶分子结合的条件下孵育所述样品或样品组;通过所述一个或多个引导RNA与所述一个或多个靶分子的结合来激活所述CRISPR效应子蛋白,其中激活所述CRISPR效应子蛋白导致所述基于RNA的掩蔽构建体的修饰,从而产生可检测的阳性信号;和检测所述可检测的阳性信号,其中检测到所述可检测的阳性信号表明所述样品中存在一个或多个靶分子。

[0758] 在一些实施方式中,所述用于检测样品中的靶核酸的方法包括:使样品与以下物质接触:工程化的CRISPR-Cas蛋白;至少一个引导多核苷酸,所述引导多核苷酸包含能够与靶核酸结合并设计为与所述工程化CRISPR-Cas形成复合物的引导序列;和包含非靶序列的基于RNA的掩蔽构建体;其中所述工程化CRISPR-Cas蛋白表现出附带的RNAe活性并切割所述检测构建体的所述非靶序列;和检测来自所述非靶序列的切割的信号,从而检测所述样品中的所述靶核酸。在一些实施方式中,所述方法进一步包括使所述样品与用于扩增所述靶核酸的试剂接触。在一些实施方式中,用于扩增的所述试剂包括等温扩增反应试剂。在一些实施方式中,所述等温扩增试剂包括基于核酸序列的扩增试剂、重组酶聚合酶扩增试剂、环介导的等温扩增试剂、链置换扩增试剂、解旋酶依赖性扩增试剂或切口酶扩增试剂。

[0759] 在一些实施方式中,所述靶核酸是DNA分子,并且所述方法进一步包括使所述靶DNA分子与包含RNA聚合酶位点和RNA聚合酶的引物接触。

[0760] 在一些实施方式中,所述掩蔽构建体:抑制可检测的阳性信号的产生,直到所述掩蔽构建体被切割或失活;或者掩蔽可检测的阳性信号或产生可检测的阴性信号,直到所述掩蔽构建体被切割或失活。

[0761] 在一些实施方式中,所述掩蔽构建体包含:a. 沉默RNA,其抑制由报告构建体编码的基因产物的产生,其中所述基因产物在表达时产生可检测的阳性信号;b. 产生阴性可检测信号的核酶,并且其中所述可检测的阳性信号在所述核酶失活时产生正;或者c. 核酶,其将底物转化为第一颜色,并且其中所述底物在所述核酶失活时为第二颜色;d. 适配体和/或包含多核苷酸拴系的抑制剂;e. 与可检测的配体和掩蔽组分连接的多核苷酸;f. 通过桥接分子保持聚集的纳米粒子,其中所述桥接分子的至少一部分包含多核苷酸,并且其中,当将纳米粒子分散于溶液中时,所述溶液经历色移;g. 通过连接分子与一个或多个猝灭分子连接的量子点或荧光团,其中所述连接分子的至少一部分包含多核苷酸;h. 与嵌合剂复合的多核苷酸,其中所述嵌合剂在切割所述多核苷酸之后改变吸光度;或i. 由多核苷酸拴系的两个荧光团,所述荧光团在从所述多核苷酸释放时发生荧光变化。在一些实施方式中,所述适配体a. 包含螯合酶的多核苷酸拴系的抑制剂,其中所述酶通过作用于底物从所述适配体或多核苷酸拴系的抑制剂释放时产生可检测的信号;或b. 是抑制酶并防止所述酶催化从底物产生可检测的信号的抑制性适配体,或其中所述多核苷酸拴系的抑制剂抑制酶并防止所述酶催化从底物产生可检测的信号;或c. 螯合一对试剂,所述试剂在从所述适配体释放

后结合以产生可检测的信号。在一些实施方式中,所述纳米颗粒是胶态金属。在一些实施方式中,所述至少一个引导多核苷酸包含错配。在一些实施方式中,所述错配在所述一个或多个引导序列上的单个核苷酸变异的上或下游。

[0762] 在另一方面,本发明提供了一种用于检测样品中的肽的方法,其包括:将样品或样品组分配到一个或多个单独的离散体积中,所述单独的离散体积包括包含肽检测适配体、包含效应子蛋白的CRISPR系统、一个或多个引导 RNA、基于RNA的掩蔽构建体;其中所述肽检测适配体包含掩蔽的RNA聚合酶位点并被配置为与一个或多个靶分子结合;在足以允许所述肽检测适体与所述一个或多个靶分子结合的条件下孵育所述样品或样品组,其中所述适配体与对应的靶分子的结合暴露所述RNA聚合酶结合位点,导致触发RNA的RNA合成;通过所述一个或多个引导RNA与所述触发RNA的结合来激活所述CRISPR效应子蛋白,其中激活所述CRISPR效应子蛋白导致所述基于RNA的掩蔽构建体的修饰,从而产生可检测的阳性信号;和检测所述可检测的阳性信号,其中检测到所述可检测的阳性信号表明所述样品中存在一个或多个靶分子。

[0763] 在某些示例实施方式中,所述一个或多个引导RNA设计为与诊断疾病状态的一个或多个靶分子结合。在某些其他示例实施方式中,所述疾病状态是感染、器官疾病、血液疾病、免疫系统疾病、癌症、脑和神经系统疾病、内分泌疾病、妊娠或分娩相关疾病、遗传疾病或环境获得性疾病、癌症或真菌感染、细菌感染、寄生虫感染或病毒感染。

[0764] 在某些示例实施方式中,所述基于RNA的掩蔽构建体抑制可检测的阳性信号的产生,或者所述基于RNA的掩蔽构建体通过掩蔽所述可检测的阳性信号或替代地产生可检测的阴性信号来抑制可检测的阳性信号的产生,或者所述基于RNA的掩蔽构建体包含抑制由报道构建体编码的基因产物的生成的沉默RNA,其中所述基因产物在表达时产生所述可检测的阳性信号,或者所述基于RNA的掩蔽构建体是产生所述阴性可检测信号的核酶并且其中所述阳性可检测信号在所述核酶失活时产生。在其他示例实施方式中,所述核酶将底物转化为第一状态并且其中所述底物在所述核酶失活时转化为第二状态,或者所述基于RNA的掩蔽试剂是适配体,或者所述适配体契合酶,其中所述酶在从所述适配体释放后通过作用于底物来产生可检测的信号,或者所述适配体契合一对在从所述适配体释放时结合以产生可检测的信号的试剂。在再进一步的实施方式中,所述基于RNA的掩蔽构建体包含RNA寡核苷酸,所述RNA寡核苷酸具有在所述RNA寡核苷酸的第一末端上的可检测的配体和所述RNA寡核苷酸的第二末端上的掩蔽组分,或者所述可检测的配体是荧光团并且所述掩蔽组分是淬灭分子。

[0765] 碱基编辑

[0766] 本公开还提供了一种碱基编辑系统。通常,这样的系统可包含与Cas蛋白融合的脱氨酶(例如腺苷脱氨酶或胞苷脱氨酶)。所述Cas蛋白可以是死的 Cas蛋白或Cas切口酶蛋白。在某些实施例中,所述系统包含与死的 CRISPR-Cas或CRISPR-Cas切口酶融合的腺苷脱氨酶的突变形式。所述腺苷脱氨酶的所述突变形式可具有腺苷脱氨酶和胞苷脱氨酶活性两者。

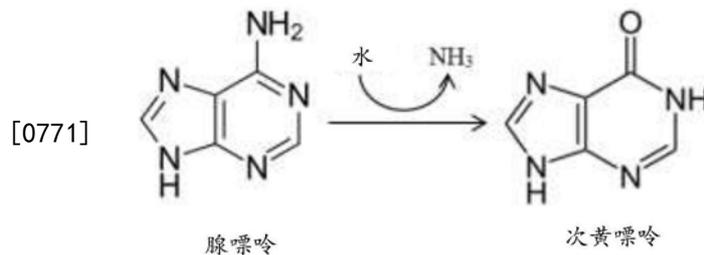
[0767] 在某些示例实施方式中,为了碱基编辑目的,dCas13b可以与腺苷脱氨酶或胞苷脱氨酶融合。在一些情况下,所述dCas13b是dCas13b-t1、dCas13b-t2 或dCas13b-t3。

[0768] 在一个方面,本公开提供了一种工程化腺苷脱氨酶。所述工程化腺苷脱氨酶在本

文中可包含一个或多个突变。在一些实施方式中,所述工程化腺苷脱氨酶具有胞苷脱氨酶活性。在某些实施例中,所述工程化腺苷脱氨酶具有胞苷脱氨酶活性和腺苷脱氨酶两者。图101显示了根据本文中的一些实施方式所述的可编程的胞苷至尿苷转化的示例系统和方法。在一些情况下,本文中由碱基编辑器进行的所述修饰可用于靶向翻译后信号传导或催化。图102 显示了示例方法。

[0769] 腺苷脱氨酶

[0770] 本文所用的术语“腺苷脱氨酶”或“腺苷脱氨酶蛋白”是指如本文所述的蛋白、多肽或或蛋白或多肽的一个或多个功能性结构域,其能够催化将腺嘌呤(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。在一些实施方式中,所述包含腺嘌呤的分子是腺苷(A),并且所述包含次黄嘌呤的分子是肌苷(I)。所述包含腺嘌呤的分子可以是脱氧核糖核酸(DNA)或核糖核酸(RNA)。



[0772] 根据本公开,可与本公开联合使用的腺苷脱氨酶包括但不限于被称为腺苷脱氨酶的作用于RNA (ADAR) 上的酶家族成员、被称为腺苷脱氨酶的作用于tRNA (ADAT) 上的酶家族成员和其他包含腺苷脱氨酶结构域 (ADAD) 的家族成员。根据本公开,所述腺苷脱氨酶能够靶向RNA/DNA和RNA双链核酸分子中的腺嘌呤。确实,Zheng等 (Nucleic Acids Res.2017,45 (6) :3369-3377) 证实ADAR可在RNA/DNA和RNA/RNA双链核酸分子上实现腺苷至肌苷的编辑反应。如下文所详述的,在特定实施方式中,已经修饰了所述腺苷脱氨酶以增加其编辑RNA/DNA异源双链体中的DNA或RNA双链体的能力,如下文详述。

[0773] 在一些实施方式中,所述腺苷脱氨酶源自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、鱼类,苍蝇和蠕虫。在一些实施方式中,所述腺苷脱氨酶是人、鱿鱼或果蝇腺苷脱氨酶。

[0774] 在一些实施方式中,所述腺苷脱氨酶是人ADAR,所述人ADAR包括 hADAR1、hADAR2、hADAR3。在一些实施方式中,所述腺苷脱氨酶是秀丽隐杆线虫ADAR蛋白,所述秀丽隐杆线虫ADAR蛋白包括ADR-1和 ADR-2。在一些实施方式中,所述腺苷脱氨酶是果蝇ADAR蛋白,所述果蝇 ADAR蛋白包括dAdar。在一些实施方式中,所述腺苷脱氨酶是鱿鱼长翼鱿鱼ADAR蛋白,所述长翼鱿鱼ADAR蛋白包括sqADAR2a和sqADAR2b。在一些实施方式中,所述腺苷脱氨酶是人ADAT蛋白。在一些实施方式中,所述腺苷脱氨酶是果蝇ADAT蛋白。在一些实施方式中,所述腺苷脱氨酶是人ADAD蛋白,所述人ADAD蛋白包括TENR (hADAD1) 和 TENRL (hADAD2)。

[0775] 在一些实施方式中,所述腺苷脱氨酶是Tad蛋白,例如大肠杆菌TadA。见Kim等, Biochemistry 45:6407-6416 (2006); Wolf等,EMBO J. 21:3841-3851 (2002)。在一些实施方式中,所述腺苷脱氨酶是小鼠ADA。见 Grunebaum等, Curr. Opin. Allergy Clin. Immunol.13:630-638 (2013)。在一些实施方式中,所述腺苷脱氨酶是人ADAT2。见 Fukui等, J. 核酸 2010:260512 (2010)。在一些实施方式中,所述脱氨酶(例如腺苷或胞苷脱

氨酶) 是Cox等, *Science*. 2017, November 24; 358 (6366) :1019-1027; Komore等, *Nature*. 2016 May 19; 533 (7603) :420-4和Gaudelli等, *Nature*. 2017 Nov 23; 551 (7681) : 464-471中描述的那些中的一种或多种。

[0776] 在一些实施方式中,所述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。在一些实施方式中,所述双链核酸底物是RNA-DNA杂合体双链核酸分子。在一些实施方式中,所述腺苷脱氨酶蛋白识别在所述双链底物上的结合窗口。在一些实施方式中,所述结合窗口包含至少一个靶腺苷残基。在一些实施方式中,所述结合窗口范围为约 3bp至约100bp。在一些实施方式中,所述结合窗口范围为约5bp至约50bp。在一些实施方式中,所述结合窗口范围为约10bp至约30bp。在一些实施方式中,所述结合窗口为约1bp、2bp、3bp、5bp、7bp、10bp、15bp、20bp、25bp、30bp、40bp、45bp、50bp、55bp、60bp、65bp、70bp、75bp、80bp、85bp、90bp、95bp或100bp。

[0777] 在一些实施方式中,所述腺苷脱氨酶蛋白包含一个或多个脱氨酶结构域。不受特定理论的约束,预期所述脱氨酶结构域作用为识别包含于双链核酸底物中的一个或多个靶腺苷(A)残基并将其转化为肌苷(I)残基。在一些实施方式中,所述脱氨酶结构域包含活性中心。在一些实施方式中,所述活性中心包含锌离子。在一些实施方式中,在所述A至I编辑过程中,在所述靶腺苷残基处的碱基配对被中断,并且所述靶腺苷残基被“翻转”出所述双螺旋以变得可被所述腺苷脱氨酶接触。在一些实施方式中,在所述活性中心或附近的氨基酸残基与一个或多个核苷酸5'至靶腺苷残基相互作用。在一些实施方式中,在所述活性中心或附近的氨基酸残基与一个或多个核苷酸3'至靶腺苷残基相互作用。在一些实施方式中,在所述活性中心或附近的氨基酸残基进一步与相反链上的所述靶腺苷残基互补的核苷酸相互作用。在一些实施方式中,所述氨基酸残基与所述核苷酸的2'羟基形成氢键。

[0778] 在一些实施方式中,所述腺苷脱氨酶包含人ADAR2全蛋白(hADAR2)或其脱氨酶结构域(hADAR2-D)。在一些实施方式中,所述腺苷脱氨酶是与hADAR2或hADAR2-D同源的ADAR家族成员。

[0779] 尤其是,在一些实施方式中,所述同源ADAR蛋白是人ADAR1(hADAR1)或其所述脱氨酶结构域(hADAR1-D)。在一些实施方式中,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。

[0780] 在一些实施方式中,所述腺苷脱氨酶包含hADAR2-D的野生型氨基酸序列。在一些实施方式中,所述腺苷脱氨酶在所述hADAR2-D序列中包含一个或多个突变,从而hADAR2-D的编辑效率和/或底物编辑偏好根据具体需求而改变。所述工程化腺苷脱氨酶可与Cas蛋白(例如Cas9,Cas 12(例如Cas12a,Cas12b,Cas12c,Cas12d等),Cas13(例如Cas13a,Cas13b(例如Cas13b-t1、Cas13b-t2、Cas13b-t3),Cas13c,Cas13d等),Cas14,CasX,CasY,或所述Cas蛋白的工程化形式(例如受攻击的(invective)、死的形式、切口酶形式))融合。在一些实施例中,本文提供的包括与死的Cas13b蛋白或Cas13切口酶融合的工程化腺苷脱氨酶。

[0781] hADAR1和hADAR2蛋白的某些突变已描述于Kuttan等,*Proc Natl Acad Sci U S A*. (2012) 109 (48) :E3295-304;Want等*ACS Chem Biol*. (2015) 10 (11) :2512-9;和Zheng等,*核酸s Res*. (2017) 45 (6) :3369-337中,其每个通过引用整体并入本文。

[0782] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的甘氨酸336或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置336处的所述甘

氨酸残基被天冬氨酸残基取代 (G336D)。

[0783] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的甘氨酸487或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置487处的所述甘氨酸残基被具有相对小的侧链的非极性氨基酸残基取代。例如,在一些实施方式中,在位置487处的所述甘氨酸残基被丙氨酸残基取代 (G487A)。在一些实施方式中,在位置487处的所述甘氨酸残基被缬氨酸残基取代 (G487V)。在一些实施方式中,在位置487处的所述甘氨酸残基被具有相对大的侧链的氨基酸残基取代。在一些实施方式中,在位置487处的所述甘氨酸残基被精氨酸残基取代 (G487R)。在一些实施方式中,在位置487处的所述甘氨酸残基被赖氨酸残基取代 (G487K)。在一些实施方式中,在位置487处的所述甘氨酸残基被色氨酸残基取代 (G487W)。在一些实施方式中,在位置487处的所述甘氨酸残基被酪氨酸残基取代 (G487Y)。

[0784] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的谷氨酸488或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置488处的所述谷氨酸残基被谷氨酰胺残基取代 (E488Q)。在一些实施方式中,在位置488处的所述谷氨酸残基被组氨酸残基取代 (E488H)。在一些实施方式中,在位置488处的所述谷氨酸残基被精氨酸残基取代 (E488R)。在一些实施方式中,在位置488处的所述谷氨酸残基被赖氨酸残基取代 (E488K)。在一些实施方式中,在位置488处的所述谷氨酸残基被天冬氨酸残基取代 (E488N)。在一些实施方式中,在位置488处的所述谷氨酸残基被丙氨酸残基取代 (E488A)。在一些实施方式中,在位置488处的所述谷氨酸残基被甲硫氨酸残基取代 (E488M)。在一些实施方式中,在位置488处的所述谷氨酸残基被丝氨酸残基取代 (E488S)。在一些实施方式中,在位置488处的所述谷氨酸残基被苯基丙氨酸残基取代 (E488F)。在一些实施方式中,在位置488处的所述谷氨酸残基被赖氨酸残基取代 (E488L)。在一些实施方式中,在位置488处的所述谷氨酸残基被色氨酸残基取代 (E488W)。

[0785] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的苏氨酸490或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置490处的所述苏氨酸残基被半胱氨酸残基取代 (T490C)。在一些实施方式中,在位置490处的所述苏氨酸残基被丝氨酸残基取代 (T490S)。在一些实施方式中,在位置490处的所述苏氨酸残基被丙氨酸残基取代 (T490A)。在一些实施方式中,在位置490处的所述苏氨酸残基被苯基丙氨酸残基取代 (T490F)。在一些实施方式中,在位置490处的所述苏氨酸残基被酪氨酸残基取代 (T490Y)。在一些实施方式中,在位置490处的所述苏氨酸残基被丝氨酸残基取代 (T490R)。在一些实施方式中,在位置490处的所述苏氨酸残基被丙氨酸残基取代 (T490K)。在一些实施方式中,在位置490处的所述苏氨酸残基被苯基丙氨酸残基取代 (T490P)。在一些实施方式中,在位置490处的所述苏氨酸残基被酪氨酸残基取代 (T490E)。

[0786] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的缬氨酸493或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置493处的所述缬氨酸残基被丙氨酸残基取代 (V493A)。在一些实施方式中,在位置493处的所述缬氨酸残基被丝氨酸残基取代 (V493A)。在一些实施方式中,在位置493处的所述缬氨酸残基被苏氨酸残基取代 (V493T)。在一些实施方式中,在位置493处的所述缬氨酸残基被精氨酸残基取代 (V493R)。在一些实施方式中,在位置493处的所述缬氨酸残基被天冬氨酸残基取代

(V493D)。在一些实施方式中,在位置493处的所述缬氨酸残基被脯氨酸残基取代(V493P)。在一些实施方式中,在位置493处的所述缬氨酸残基被甘氨酸残基取代(V493G)。

[0787] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的丙氨酸589或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置589处的所述丙氨酸残基被缬氨酸残基取代(A589V)。

[0788] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的天冬氨酸597或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置597处的所述天冬氨酸残基被赖氨酸残基取代(N597K)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置597处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置597处的所述天冬氨酸残基被精氨酸残基取代(N597R)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置597处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置597处的所述天冬氨酸残基被丙氨酸残基取代(N597A)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置597处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置597处的所述天冬氨酸残基被谷氨酸残基取代(N597E)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置597处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置597处的所述天冬氨酸残基被组氨酸残基取代(N597H)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置597处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置597处的所述天冬氨酸残基被甘氨酸残基取代(N597G)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置597处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置597处的所述天冬氨酸残基被酪氨酸残基取代(N597Y)。在一些实施方式中,在位置597处的所述天冬氨酸残基被苯基丙氨酸残基取代(N597F)。在一些实施方式中,所述腺苷脱氨酶包含突变N597I。在一些实施方式中,所述腺苷脱氨酶包含突变N597L。在一些实施方式中,所述腺苷脱氨酶包含突变N597V。在一些实施方式中,所述腺苷脱氨酶包含突变N597M。在一些实施方式中,所述腺苷脱氨酶包含突变N597C。在一些实施方式中,所述腺苷脱氨酶包含突变N597P。在一些实施方式中,所述腺苷脱氨酶包含突变N597T。在一些实施方式中,所述腺苷脱氨酶包含突变N597S。在一些实施方式中,所述腺苷脱氨酶包含突变N597W。在一些实施方式中,所述腺苷脱氨酶包含突变N597Q。在一些实施方式中,所述腺苷脱氨酶包含突变N597D。在某些示例实施方式中,在E488Q背景的上下文中进一步进行了上文所述的在N597处的突变。

[0789] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的丝氨酸599或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置599处的所述丝氨酸残基被苏氨酸残基取代(S599T)。

[0790] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的天冬氨酸613或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置613处的所述天冬氨酸残基被赖氨酸残基取代(N613K)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置613处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置613处的所述天冬氨酸残基被精氨酸残基取代(N613R)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置613处的突变,其在所述野生型序列中具有天

冬氨酸残基。在一些实施方式中,在位置613处的所述天冬氨酸残基被丙氨酸残基取代(N613A)。在一些实施方式中,所述腺苷脱氨酶包含在所述氨基酸序列的位置613处的突变,其在所述野生型序列中具有天冬氨酸残基。在一些实施方式中,在位置613处的所述天冬氨酸残基被谷氨酸残基取代(N613E)。在一些实施方式中,所述腺苷脱氨酶包含突变N613I。在一些实施方式中,所述腺苷脱氨酶包含突变N613L。在一些实施方式中,所述腺苷脱氨酶包含突变N613V。在一些实施方式中,所述腺苷脱氨酶包含突变N613F。在一些实施方式中,所述腺苷脱氨酶包含突变N613M。在一些实施方式中,所述腺苷脱氨酶包含突变N613C。在一些实施方式中,所述腺苷脱氨酶包含突变N613G。在一些实施方式中,所述腺苷脱氨酶包含突变N613P。在一些实施方式中,所述腺苷脱氨酶包含突变N613T。在一些实施方式中,所述腺苷脱氨酶包含突变N613S。在一些实施方式中,所述腺苷脱氨酶包含突变N613Y。在一些实施方式中,所述腺苷脱氨酶包含突变N613W。在一些实施方式中,所述腺苷脱氨酶包含突变N613Q。在一些实施方式中,所述腺苷脱氨酶包含突变N613H。在一些实施方式中,所述腺苷脱氨酶包含突变N613D。在一些实施方式中,结合E488Q突变而进一步进行了上文所述的在N613处的突变。

[0791] 在一些实施方式中,为了改善编辑效率,所述腺苷脱氨酶可包含以下一个或多个突变: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(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。

[0792] 在一些实施方式中,为了降低编辑效率,所述腺苷脱氨酶可包含以下一个或多个突变:E488F、E488L、E488W、T490A、T490F、T490Y、T490R、T490K、T490P、T490E、N597F(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在特定实施方式中,可能对使用具有降低的效能的腺苷脱氨酶来减少脱靶效应感兴趣。

[0793] 在一些实施方式中,为了减少脱靶效应,所述腺苷脱氨酶包含在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(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶包含在E488或选自R348、V351、T375、K376、E396、C451、R455、N473、R474、K475、R477、R481、S486、T490、S495、R510的一个或多个附加位置处的突变。在一些实施方式中,所述腺苷脱氨酶包含在T375处和任选地在一个或多个附加位置处的突变。在一些实施方式中,所述腺苷脱氨酶包含在N473处和任选地在一个或多个附加位置处的突变。在一些实施方式中,所述腺苷脱氨酶包含在V351处和任选地在一个或多个附加位置处的突变。在一些实施方式中,所述腺苷脱氨酶包含在E488和T375处和任选地在一个或多个附加位置处的突变。在一些实施方式中,所述腺苷脱氨酶包含在E488和N473处和任选地在一个或多个附加位置处的突变。在一些实施方式中,所述腺苷脱氨酶包含突变E488和V351和任选地在一个或多个附加位置处的突变。在一些实施方式中,所述腺苷脱氨酶包含在E488以及T375、N473和V351中的一个或多个处的突变。

[0794] 在一些实施方式中,为了减少脱靶效应,所述腺苷脱氨酶包含一个或多个突变,所述一个或多个突变选自R348E、V351L、T375G、T375S、R455G、R455S、R455E、N473D、R474E、K475Q、R477E、R481E、S486T、E488Q、T490A、T490S、S495T和R510E(基于hADAR2-D的氨基酸

序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶包含突变E488Q和选自R348E、V351L、T375G、T375S、R455G、R455S、R455E、N473D、R474E、K475Q、R477E、R481E、S486T、T490A、T490S、S495T和R510E的一个或多个附加突变。在一些实施方式中,所述腺苷脱氨酶包含突变T375G或T375S和任选地一个或多个附加突变。在一些实施方式中,所述腺苷脱氨酶包含突变N473D和任选地一个或多个附加突变。在一些实施方式中,所述腺苷脱氨酶包含突变V351L和任选地一个或多个附加突变。在一些实施方式中,所述腺苷脱氨酶包含突变E488Q,和T375G或 T375G,和任选地一个或多个附加突变。在一些实施方式中,所述腺苷脱氨酶包含突变E488Q和N473D和任选地一个或多个附加突变。在一些实施方式中,所述腺苷脱氨酶包含突变E488Q和V351L和任选地一个或多个附加突变。在一些实施方式中,所述腺苷脱氨酶包含突变E488Q以及T375G/S、N473D和V351L中的一个或多个。

[0795] 在某些实施例中,所述腺苷脱氨酶蛋白或其催化结构域已被修饰为包含在所述hADAR2-D氨基酸序列的E488(优选地E488Q)或同源ADAR蛋白中的对应的位置处的突变,和/或其中所述腺苷脱氨酶蛋白或其催化结构域已被修饰为包含在所述hADAR2-D氨基酸序列的T375(优选地T375G)或同源ADAR蛋白中的对应的位置处的突变。在某些实施例中,所述腺苷脱氨酶蛋白或其催化结构域已被修饰为包含在所述hADAR1d氨基酸序列的E1008(优选地E1008Q)或同源ADAR蛋白中的对应的位置处的突变。

[0796] 与双链核酸分子RNA结合的所述人ADAR2脱氨酶结构域的晶体结构显示了结合所述修饰位点的5'侧上的RNA的蛋白环。该5'结合环是造成ADAR家族成员之间底物特异性差异的一个原因。见Wang等,核酸Res., 44(20):9872-9880(2016),其内容通过引用整体并入本文。此外,在所述酶活性位点的附近识别了ADAR2特异性RNA结合环。见Mathews等,Nat.Struct. Mol.Biol.,23(5):426-33(2016),其内容通过引用整体并入本文。在一些实施方式中,所述腺苷脱氨酶包含在所述RNA结合环中的一个或多个突变以改善编辑特异性和/或效率。

[0797] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的丙氨酸454或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置454处的所述丙氨酸残基被丝氨酸残基取代(A454S)。在一些实施方式中,在位置454处的所述丙氨酸残基被半胱氨酸残基取代(A454C)。在一些实施方式中,在位置454处的所述丙氨酸残基被天冬氨酸残基取代(A454D)。

[0798] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的精氨酸455或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置455处的所述精氨酸残基被丙氨酸残基取代(R455A)。在一些实施方式中,在位置455处的所述精氨酸残基被缬氨酸残基取代(R455V)。在一些实施方式中,在位置455处的所述精氨酸残基被组氨酸残基取代(R455H)。在一些实施方式中,在位置455处的所述精氨酸残基被甘氨酸残基取代(R455G)。在一些实施方式中,在位置455处的所述精氨酸残基被丝氨酸残基取代(R455S)。在一些实施方式中,在位置455处的所述精氨酸残基被谷氨酸残基取代(R455E)。在一些实施方式中,所述腺苷脱氨酶包含突变R455C。在一些实施方式中,所述腺苷脱氨酶包含突变R455I。在一些实施方式中,所述腺苷脱氨酶包含突变R455K。在一些实施方式中,所述腺苷脱氨酶包含突变R455L。在一些实施方式中,所述腺苷脱氨酶包含突变R455M。在一些实施

方式中,所述腺苷脱氨酶包含突变R455N。在一些实施方式中,所述腺苷脱氨酶包含突变R455Q。在一些实施方式中,所述腺苷脱氨酶包含突变R455F。在一些实施方式中,所述腺苷脱氨酶包含突变R455W。在一些实施方式中,所述腺苷脱氨酶包含突变R455P。在一些实施方式中,所述腺苷脱氨酶包含突变R455Y。在一些实施方式中,所述腺苷脱氨酶包含突变R455E。在一些实施方式中,所述腺苷脱氨酶包含突变R455D。在一些实施方式中,结合E488Q突变而进一步进行了上文所述的在R455处的突变。

[0799] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的异亮氨酸456或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置456处的所述异亮氨酸残基被缬氨酸残基取代(I456V)。在一些实施方式中,在位置456处的所述异亮氨酸残基被亮氨酸残基取代(I456L)。在一些实施方式中,在位置456处的所述异亮氨酸残基被天冬氨酸残基取代(I456D)。

[0800] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的苯基丙氨酸457或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置457处的所述苯基丙氨酸被酪氨酸残基取代(F457Y)。在一些实施方式中,在位置457处的所述苯基丙氨酸被精氨酸残基取代(F457R)。在一些实施方式中,在位置457处的所述苯基丙氨酸被谷氨酸残基取代(F457E)。

[0801] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的丝氨酸458或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置458处的所述丝氨酸残基被缬氨酸残基取代(S458V)。在一些实施方式中,在位置458处的所述丝氨酸残基被苯基丙氨酸残基取代(S458F)。在一些实施方式中,在位置458处的所述丝氨酸残基被脯氨酸残基取代(S458P)。在一些实施方式中,所述腺苷脱氨酶包含突变458I。在一些实施方式中,所述腺苷脱氨酶包含突变458L。在一些实施方式中,所述腺苷脱氨酶包含突变458M。在一些实施方式中,所述腺苷脱氨酶包含突变458C。在一些实施方式中,所述腺苷脱氨酶包含突变458A。在一些实施方式中,所述腺苷脱氨酶包含突变458G。在一些实施方式中,所述腺苷脱氨酶包含突变458T。在一些实施方式中,所述腺苷脱氨酶包含突变458Y。在一些实施方式中,所述腺苷脱氨酶包含突变458W。在一些实施方式中,所述腺苷脱氨酶包含突变458Q。在一些实施方式中,所述腺苷脱氨酶包含突变458N。在一些实施方式中,所述腺苷脱氨酶包含突变458H。在一些实施方式中,所述腺苷脱氨酶包含突变458E。在一些实施方式中,所述腺苷脱氨酶包含突变458D。在一些实施方式中,所述腺苷脱氨酶包含突变458K。在一些实施方式中,所述腺苷脱氨酶包含突变458R。在一些实施方式中,结合E488Q突变而进一步进行了上文所述的在S458处的突变。

[0802] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的脯氨酸459或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置459处的所述脯氨酸残基被半胱氨酸残基取代(P459C)。在一些实施方式中,在位置459处的所述脯氨酸残基被组氨酸残基取代(P459H)。在一些实施方式中,在位置459处的所述脯氨酸残基被色氨酸残基取代(P459W)。

[0803] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的组氨酸460或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置460处的所述组氨酸残基被精氨酸残基取代(H460R)。在一些实施方式中,在位置460处的所述组氨酸残基

被异亮氨酸残基取代 (H460I)。在一些实施方式中,在位置460处的所述组氨酸残基被脯氨酸残基取代 (H460P)。在一些实施方式中,所述腺苷脱氨酶包含突变H460L。在一些实施方式中,所述腺苷脱氨酶包含突变H460V。在一些实施方式中,所述腺苷脱氨酶包含突变H460F。在一些实施方式中,所述腺苷脱氨酶包含突变 H460M。在一些实施方式中,所述腺苷脱氨酶包含突变H460C。在一些实施方式中,所述腺苷脱氨酶包含突变H460A。在一些实施方式中,所述腺苷脱氨酶包含突变H460G。在一些实施方式中,所述腺苷脱氨酶包含突变H460T。在一些实施方式中,所述腺苷脱氨酶包含突变H460S。在一些实施方式中,所述腺苷脱氨酶包含突变H460Y。在一些实施方式中,所述腺苷脱氨酶包含突变H460W。在一些实施方式中,所述腺苷脱氨酶包含突变H460Q。在一些实施方式中,所述腺苷脱氨酶包含突变H460N。在一些实施方式中,所述腺苷脱氨酶包含突变H460E。在一些实施方式中,所述腺苷脱氨酶包含突变 H460D。在一些实施方式中,所述腺苷脱氨酶包含突变H460K。在一些实施方式中,结合E488Q突变而进一步进行了上文所述的在H460处的突变。

[0804] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的脯氨酸462或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置462处的所述脯氨酸残基被丝氨酸残基取代 (P462S)。在一些实施方式中,在位置462处的所述脯氨酸残基被色氨酸残基取代 (P462W)。在一些实施方式中,在位置462处的所述脯氨酸残基被谷氨酸残基取代 (P462E)。

[0805] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的天冬氨酸469或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置469的所述天冬氨酸残基被谷氨酰胺残基取代 (D469Q)。在一些实施方式中,在位置469的所述天冬氨酸残基被丝氨酸残基取代 (D469S)。在一些实施方式中,在位置469的所述天冬氨酸残基被酪氨酸残基取代 (D469Y)。

[0806] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的精氨酸470或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置470处的所述精氨酸残基被丙氨酸残基取代 (R470A)。在一些实施方式中,在位置470处的所述精氨酸残基被异亮氨酸残基取代 (R470I)。在一些实施方式中,在位置470处的所述精氨酸残基被天冬氨酸残基取代 (R470D)。

[0807] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的组氨酸471或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置471处的所述组氨酸残基被赖氨酸残基取代 (H471K)。在一些实施方式中,在位置471处的所述组氨酸残基被苏氨酸残基取代 (H471T)。在一些实施方式中,在位置471处的所述组氨酸残基被缬氨酸残基取代 (H471V)。

[0808] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的脯氨酸472或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置472处的所述脯氨酸残基被赖氨酸残基取代 (P472K)。在一些实施方式中,在位置472处的所述脯氨酸残基被苏氨酸残基取代 (P472T)。在一些实施方式中,在位置472处的所述脯氨酸残基被天冬氨酸残基取代 (P472D)。

[0809] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的天冬氨酸473或同源ADAR蛋白中的对应的位置处的突。在一些实施方式中,在位置473处的所述天冬

氨酸残基被精氨酸残基取代 (N473R)。在一些实施方式中,在位置473处的所述天冬氨酸残基被色氨酸残基取代 (N473W)。在一些实施方式中,在位置473处的所述天冬氨酸残基被脯氨酸残基取代 (N473P)。在一些实施方式中,在位置473处的所述天冬氨酸残基被天冬氨酸残基取代 (N473D)。

[0810] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置 474处的所述精氨酸残基被赖氨酸残基取代 (R474K)。在一些实施方式中,在位置474处的所述精氨酸残基被甘氨酸残基取代 (R474G)。在一些实施方式中,在位置474处的所述精氨酸残基被天冬氨酸残基取代 (R474D)。在一些实施方式中,在位置474处的所述精氨酸残基被谷氨酸残基取代 (R474E)。

[0811] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的赖氨酸475或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置475处的所述赖氨酸残基被谷氨酰胺残基取代 (K475Q)。在一些实施方式中,在位置475处的所述赖氨酸残基被天冬氨酸残基取代 (K475N)。在一些实施方式中,在位置475处的所述赖氨酸残基被天冬氨酸残基取代 (K475D)。

[0812] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的丙氨酸476或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置476处的所述丙氨酸残基被丝氨酸残基取代 (A476S)。在一些实施方式中,在位置476处的所述丙氨酸残基被精氨酸残基取代 (A476R)。在一些实施方式中,在位置476处的所述丙氨酸残基被谷氨酸残基取代 (A476E)。

[0813] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的精氨酸477或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置477处的所述精氨酸残基被赖氨酸残基取代 (R477K)。在一些实施方式中,在位置477处的所述精氨酸残基被苏氨酸残基取代 (R477T)。在一些实施方式中,在位置477处的所述精氨酸残基被苯基丙氨酸残基取代 (R477F)。在一些实施方式中,在位置474处的所述精氨酸残基被谷氨酸残基取代 (R477E)。

[0814] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的甘氨酸478或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置478处的所述甘氨酸残基被丙氨酸残基取代 (G478A)。在一些实施方式中,在位置478处的所述甘氨酸残基被精氨酸残基取代 (G478R)。在一些实施方式中,在位置478处的所述甘氨酸残基被酪氨酸残基取代 (G478Y)。在一些实施方式中,所述腺苷脱氨酶包含突变G478I。在一些实施方式中,所述腺苷脱氨酶包含突变G478L。在一些实施方式中,所述腺苷脱氨酶包含突变G478V。在一些实施方式中,所述腺苷脱氨酶包含突变 G478F。在一些实施方式中,所述腺苷脱氨酶包含突变G478M。在一些实施方式中,所述腺苷脱氨酶包含突变G478C。在一些实施方式中,所述腺苷脱氨酶包含突变G478P。在一些实施方式中,所述腺苷脱氨酶包含突变G478T。在一些实施方式中,所述腺苷脱氨酶包含突变G478S。在一些实施方式中,所述腺苷脱氨酶包含突变G478W。在一些实施方式中,所述腺苷脱氨酶包含突变G478Q。在一些实施方式中,所述腺苷脱氨酶包含突变G478N。在一些实施方式中,所述腺苷脱氨酶包含突变G478H。在一些实施方式中,所述腺苷脱氨酶包含突变G478E。在一些实施方式中,所述腺苷脱氨酶包含突变 G478D。在一些实施方式中,所述腺苷脱氨酶包含突变G478K。在一些实施方式中,结合E488Q

突变而进一步进行了上文所述的在G478处的突变。

[0815] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的谷氨酰胺479或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置479处的所述谷氨酰胺残基被天冬氨酸残基取代(Q479N)。在一些实施方式中,在位置479处的所述谷氨酰胺残基被丝氨酸残基取代(Q479S)。在一些实施方式中,在位置479处的所述谷氨酰胺残基被脯氨酸残基取代(Q479P)。

[0816] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的精氨酸348或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置348处的所述精氨酸残基被丙氨酸残基取代(R348A)。在一些实施方式中,在位置348处的所述精氨酸残基被谷氨酸残基取代(R348E)。

[0817] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的缬氨酸351或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置351处的所述缬氨酸残基被亮氨酸残基取代(V351L)。在一些实施方式中,所述腺苷脱氨酶包含突变V351Y。在一些实施方式中,所述腺苷脱氨酶包含突变V351M。在一些实施方式中,所述腺苷脱氨酶包含突变V351T。在一些实施方式中,所述腺苷脱氨酶包含突变V351G。在一些实施方式中,所述腺苷脱氨酶包含突变V351A。在一些实施方式中,所述腺苷脱氨酶包含突变V351F。在一些实施方式中,所述腺苷脱氨酶包含突变V351E。在一些实施方式中,所述腺苷脱氨酶包含突变V351I。在一些实施方式中,所述腺苷脱氨酶包含突变V351C。在一些实施方式中,所述腺苷脱氨酶包含突变V351H。在一些实施方式中,所述腺苷脱氨酶包含突变V351P。在一些实施方式中,所述腺苷脱氨酶包含突变V351S。在一些实施方式中,所述腺苷脱氨酶包含突变V351K。在一些实施方式中,所述腺苷脱氨酶包含突变V351N。在一些实施方式中,所述腺苷脱氨酶包含突变V351W。在一些实施方式中,所述腺苷脱氨酶包含突变V351Q。在一些实施方式中,所述腺苷脱氨酶包含突变V351D。在一些实施方式中,所述腺苷脱氨酶包含突变V351R。在一些实施方式中,结合E488Q突变而进一步进行了上文所述的在V351处的突变。

[0818] 在一些实施方式中,所述腺苷脱氨酶包含所述hADAR2-D氨基酸序列的苏氨酸375或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置375处的所述苏氨酸残基被甘氨酸残基取代(T375G)。在一些实施方式中,在位置375处的所述苏氨酸残基被丝氨酸残基取代(T375S)。在一些实施方式中,所述腺苷脱氨酶包含突变T375H。在一些实施方式中,所述腺苷脱氨酶包含突变T375Q。在一些实施方式中,所述腺苷脱氨酶包含突变T375C。在一些实施方式中,所述腺苷脱氨酶包含突变T375N。在一些实施方式中,所述腺苷脱氨酶包含突变T375M。在一些实施方式中,所述腺苷脱氨酶包含突变T375A。在一些实施方式中,所述腺苷脱氨酶包含突变T375W。在一些实施方式中,所述腺苷脱氨酶包含突变T375V。在一些实施方式中,所述腺苷脱氨酶包含突变T375R。在一些实施方式中,所述腺苷脱氨酶包含突变T375E。在一些实施方式中,所述腺苷脱氨酶包含突变T375K。在一些实施方式中,所述腺苷脱氨酶包含突变T375F。在一些实施方式中,所述腺苷脱氨酶包含突变T375I。在一些实施方式中,所述腺苷脱氨酶包含突变T375D。在一些实施方式中,所述腺苷脱氨酶包含突变T375P。在一些实施方式中,所述腺苷脱氨酶包含突变T375L。在一些实施方式中,所述腺苷脱氨酶包含突变T375Y。在一些实施方式中,结合E488Q突变而进一步进行了上文所述的在T375Y处的突变。

[0819] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的精氨酸481或同源ADAR蛋白中的对应的位置处的突。在一些实施方式中,在位置481处的所述精氨酸残基被谷氨酸残基取代(R481E)。

[0820] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的丝氨酸486或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置486处的所述丝氨酸残基被苏氨酸残基取代(S486T)。

[0821] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的苏氨酸490或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置490处的所述苏氨酸残基被丙氨酸残基取代(T490A)。在一些实施方式中,在位置490处的所述苏氨酸残基被丝氨酸残基取代(T490S)。

[0822] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的丝氨酸495或同源ADAR蛋白中的对应的位置处的突。在一些实施方式中,在位置495处的所述丝氨酸残基被苏氨酸残基取代(S495T)。

[0823] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的精氨酸510或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置510处的所述精氨酸残基被谷氨酰胺残基取代(R510Q)。在一些实施方式中,在位置510处的所述精氨酸残基被丙氨酸残基取代(R510A)。在一些实施方式中,在位置510处的所述精氨酸残基被谷氨酸残基取代(R510E)。

[0824] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的谷氨酸593或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置593处的所述甘氨酸残基被丙氨酸残基取代(G593A)。在一些实施方式中,在位置593处的所述甘氨酸残基被谷氨酸残基取代(G593E)。

[0825] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的赖氨酸594或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置594处的所述赖氨酸残基被丙氨酸残基取代(K594A)。

[0826] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的以下位置中的任一个或多个或同源ADAR蛋白中的对应的位置处的突变: A454、R455、I456、F457、S458、P459、H460、P462、D469、R470、H471、P472、N473、R474、K475、A476、R477、G478、Q479、R348、R510、G593、K594。

[0827] 在一些实施方式中,所述腺苷脱氨酶包含所述hADAR2-D氨基酸序列的以下位置中的任一个或多个或同源ADAR蛋白中的对应的位置处的突变: 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。

[0828] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的以下位置中的任一个或多个或同源ADAR蛋白中的对应的位置处的突变,且任选地结合在E488处的突变:T375、V351、G478、S458、H460。在一些实施方式中,所述腺苷脱氨酶包含选自以下的一个

或多个突变,且任选地结合E488Q:T375G、T375C、T375H、T375Q、V351M、V351T、V351Y、G478R、S458F、H460I。

[0829] 在一些实施方式中,所述腺苷脱氨酶包含选自以下的一个或多个突变,且任选地结合E488Q:T375H、T375Q、V351M、V351Y、H460P。

[0830] 在一些实施方式中,所述腺苷脱氨酶包含突变T375S和S458F,且任选地结合E488Q。

[0831] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的以下位置中的两个或更多个或同源ADAR蛋白中的对应的位置处的突变,且任选地结合在E488处的突变:T375、N473、R474、G478、S458、P459、V351、R455、R455、T490、R348、Q479。在一些实施方式中,所述腺苷脱氨酶包含选自以下的两个或更多个突变,且任选地结合E488Q:T375G、T375S、N473D、R474E、G478R、S458F、P459W、V351L、R455G、R455S、T490A、R348E、Q479P。

[0832] 在一些实施方式中,所述腺苷脱氨酶包含突变T375G和V351L。在一些实施方式中,所述腺苷脱氨酶包含突变T375G和R455G。在一些实施方式中,所述腺苷脱氨酶包含突变T375G和R455S。在一些实施方式中,所述腺苷脱氨酶包含突变T375G和T490A。在一些实施方式中,所述腺苷脱氨酶包含突变T375G和R348E。在一些实施方式中,所述腺苷脱氨酶包含突变T375S和V351L。在一些实施方式中,所述腺苷脱氨酶包含突变T375S和R455G。在一些实施方式中,所述腺苷脱氨酶包含突变T375S和R455S。在一些实施方式中,所述腺苷脱氨酶包含突变T375S和T490A。在一些实施方式中,所述腺苷脱氨酶包含突变T375S和R348E。在一些实施方式中,所述腺苷脱氨酶包含突变N473D和V351L。在一些实施方式中,所述腺苷脱氨酶包含突变N473D和R455G。在一些实施方式中,所述腺苷脱氨酶包含突变N473D和R455S。在一些实施方式中,所述腺苷脱氨酶包含突变N473D和T490A。在一些实施方式中,所述腺苷脱氨酶包含突变N473D和R348E。在一些实施方式中,所述腺苷脱氨酶包含突变R474E和V351L。在一些实施方式中,所述腺苷脱氨酶包含突变R474E和R455G。在一些实施方式中,所述腺苷脱氨酶包含突变R474E和R455S。在一些实施方式中,所述腺苷脱氨酶包含突变R474E和T490A。在一些实施方式中,所述腺苷脱氨酶包含突变R474E和R348E。在一些实施方式中,所述腺苷脱氨酶包含突变S458F和T375G。在一些实施方式中,所述腺苷脱氨酶包含突变S458F和T375S。在一些实施方式中,所述腺苷脱氨酶包含突变S458F和N473D。在一些实施方式中,所述腺苷脱氨酶包含突变S458F和R474E。在一些实施方式中,所述腺苷脱氨酶包含突变S458F和G478R。在一些实施方式中,所述腺苷脱氨酶包含突变G478R和T375G。在一些实施方式中,所述腺苷脱氨酶包含突变G478R和T375S。在一些实施方式中,所述腺苷脱氨酶包含突变G478R和N473D。在一些实施方式中,所述腺苷脱氨酶包含突变G478R和R474E。在一些实施方式中,所述腺苷脱氨酶包含突变P459W和T375G。在一些实施方式中,所述腺苷脱氨酶包含突变P459W和T375S。在一些实施方式中,所述腺苷脱氨酶包含突变P459W和N473D。在一些实施方式中,所述腺苷脱氨酶包含突变P459W和R474E。在一些实施方式中,所述腺苷脱氨酶包含突变P459W和G478R。在一些实施方式中,所述腺苷脱氨酶包含突变P459W和S458F。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P和T375G。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P和T375S。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P和N473D。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P和R474E。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P和G478R。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P和G478R。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P

和S458F。在一些实施方式中,所述腺苷脱氨酶包含突变Q479P和P459W。可以结合E488Q突变而进一步进行本段所述的所有突变。

[0833] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的以下位置中的任一个或多个或同源ADAR蛋白中的对应的位置处的突变,任选地结合在E488处的突变:K475、Q479、P459、G478、S458。在一些实施方式中,所述腺苷脱氨酶包含选自以下的一个或多个突变,且任选地结合 E488Q:K475N、Q479N、P459W、G478R、S458P、S458F。

[0834] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的以下位置中的任一个或多个或同源ADAR蛋白中的对应的位置处的突变,任选地结合在E488处的突变:T375、V351、R455、H460、A476。在一些实施方式中,所述腺苷脱氨酶包含选自以下的一个或多个突,且任选地结合E488Q:T375G、T375C、T375H、T375Q、V351M、V351T、V351Y、R455H、H460P、H460I、A476E。

[0835] 在某些实施方式中,编辑的改善和脱靶修饰的减少是通过gRNA的化学修饰来实现的。如Vogel等(2014), *Angew Chem Int Ed*, 53:6267-6271, doi:10.1002/anie.201402634 (通过引用将其整体并入本文) 中所示例的化学修饰的gRNA减少了脱靶活性并改善了中靶(on-target)效率。一般而言,2'-O-甲基和硫代磷酸酯修饰的引导RNA改善细胞中的编辑效率。

[0836] 众所周知,ADAR表现出对编辑的A的任一侧上的的相邻核苷酸的偏好。(www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html, Matthews等(2017), *Nature Structural Mol Biol*, 23(5):426-433, 其通过引用整体并入本文)。因此,在某些实施方式中,出于基序偏好而选择性地优化所述gRNA、靶标和/或ADAR。已在体外证实故意的错配可以编辑非优选的基序 (academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272; Schneider et al (2014), *Nucleic Acid Res*, 42(10):e87); Fukuda et al. (2017), *Scientific Reports*, 7, doi:10.1038/srep41478, 通过引用整体并入本文)。因此,在某些实施方式中,为了改善非优选的5'或3'相邻碱基上的RNA编辑效率,引入相邻碱基中的故意的错配。

[0837] 在一些实施方式中,所述腺苷脱氨酶可以是tRNA-特异性腺苷脱氨酶或其变体。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌TadA的氨基酸序列位置的W23L、W23R、R26G、H36L、N37S、P48S、P48T、P48A、I49V、R51L、N72D、L84F、S97C、A106V、D108N、H123Y、G125A、A142N、S146C、D147Y、R152H、R152P、E155V、I156F、K157N、K161T。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌TadA的氨基酸序列位置的D108N。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌TadA的氨基酸序列位置的A106V、D108N。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌TadA的氨基酸序列位置的A106V、D108N、D147Y、E155V。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌TadA的氨基酸序列位置的A106V、D108N。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌TadA的氨基酸序列位

置的A106V、D108N、D147Y、E155V、L84F、H123Y、I156F。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、A142N。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、H36L、R51L、S146C、K157N。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、H36L、R51L、S146C、K157N、P48S。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、H36L、R51L、S146C、K157N、P48S、A142N。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、H36L、R51L、S146C、K157N、P48S、W23R、P48A。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、H36L、R51L、S146C、K157N、P48S、W23R、P48A、A142N。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的:A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、H36L、R51L、S146C、K157N、P48S、W23R、P48A、R152P。在一些实施方式中,所述腺苷脱氨酶可包含以下突变中的一个或多个和对应于以下突变的同源脱氨酶蛋白中的突变:基于大肠杆菌 TadA的氨基酸序列位置的:A106V、D108N、D147Y、E155V、L84F、H123Y、I156F、H36L、R51L、S146C、K157N、P48S、W23R、P48A、R152P、A142N。

[0838] 结果显示相对于其他碱基,优选地编辑所述ADAR脱氨酶结构域的靶窗口中的A相对的C。此外,与靶向的碱基的若干碱基内的U碱基配对的A显示了由CRISPR-Cas-ADAR融合体进行的低水平的编辑,其表明所述酶可以灵活地编辑多个A。这两个观察表明可以通过错配所有将用C编辑的A来指定CRISPR-Cas-ADAR融合体的活性窗口中的多个A进行编辑。因此,在某些实施方式中,所述活性窗口中的多个A:C错配被设计为生成多个A:I编辑。在某些实施方式中,为了抑制所述活性窗口中的潜在脱靶编辑,将非靶向的A与A或G配对。

[0839] 本文可交换地使用术语“编辑特异性”和“编辑偏好”来指在双链底物中的特定腺苷位点处的A至I编辑的程度。在一些实施方式中,所述底物编辑偏好由所述靶腺苷残基的5'最近邻近物和/或3'最近邻近物决定。在一些实施方式中,所述腺苷脱氨酶对所述底物的5'最近邻近物具有以下偏好:U>A>C>G(“>”表示更大偏好)。在一些实施方式中,所述腺苷脱氨酶对所述底物的3'最近邻近物具有以下偏好:G>C~A>U(“>”表示更大偏好;“~”表示相似偏好)。在一些实施方式中,所述腺苷脱氨酶对所述底物的3'最近邻近物具有以下偏好:G>C>U~A(“>”表示更大偏好;“~”表示相似偏好)。在一些实施方式中,所述腺苷脱氨酶对所述底物的3'最近邻近物具有以下偏好:G>C>A>U(“>”表示更大偏好)。在一些实施方式中,所述腺苷脱氨酶对所述底物的3'最近邻近物具有以下偏好:C~G~A>U(“>”表示更大偏好;“~”表示相似偏好)。在一些实施方式中,所述腺苷脱氨酶对包含所述靶腺苷残基的三

联体序列具有以下偏好:TAG>AAG>CAC>AAT>GAA>GAC (“>”表示更大偏好),中心A为所述靶腺苷残基。

[0840] 在一些实施方式中,腺苷脱氨酶的底物编辑偏好受所述腺苷脱氨酶蛋白中核酸结合结构域的存在或缺失的影响。在一些实施方式中,为了修改底物编辑偏好,将所述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。在一些实施方式中,所述dsRBD或dsRBM可以源自ADAR蛋白,例如hADAR1或hADAR2。在一些实施方式中,使用了包含至少一个dsRBD的全长ADAR蛋白和脱氨酶结构域。在一些实施方式中,所述一个或多个dsRBM或dsRBD在所述脱氨酶结构域的N末端。在其他实施方式中,所述一个或多个dsRBM或dsRBD在所述脱氨酶结构域的C末端。

[0841] 在一些实施方式中,腺苷脱氨酶的底物编辑偏好受所述酶的活性位点附近或活性位点中的氨基酸残基的影响。在一些实施方式中为了修改底物编辑偏好,所述腺苷脱氨酶可包含以下一个或多个突变:G336D、G487R、G487K、G487W、G487Y、E488Q、E488N、T490A、V493A、V493T、V493S、N597K、N597R、A589V、S599T、N613K、N613R(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。

[0842] 特别是,在一些实施方式中,为了减少编辑特异性,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V493A、N597K、N613K(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,为了增加编辑特异性,所述腺苷脱氨酶可包含突变T490A。

[0843] 在一些实施方式中,为了增加对紧接5'G(例如包含三联体序列GAC的底物)的靶腺苷(A)(中心A为所述靶腺苷残基)的编辑偏好,所述腺苷脱氨酶可包含以下一个或多个突变:G336D、E488Q、E488N、V493T、V493S、V493A、A589V、N597K、N597R、S599T、N613K、N613R(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。

[0844] 特别是,在一些实施方式中,所述腺苷脱氨酶包含突变E488Q或同源ADAR蛋白中的对应的突变以用于编辑包含以下三联体序列的底物:GAC、GAA、GAU、GAG、CAU、AAU、UAC,中心A为所述靶腺苷残基。

[0845] 在一些实施方式中,所述腺苷脱氨酶包含hADAR1-D的野生型氨基酸序列。在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR1-D序列中的一个或多个突变,以根据具体需求而改变hADAR1-D的编辑效率和/或底物编辑偏好。

[0846] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR1-D氨基酸序列的甘氨酸1007或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置1007处的所述甘氨酸残基被具有相对小的侧链的非极性氨基酸残基取代。例如,在一些实施方式中,在位置1007处的所述甘氨酸残基被丙氨酸残基取代(G1007A)。在一些实施方式中,在位置1007处的所述甘氨酸残基被缬氨酸残基取代(G1007V)。在一些实施方式中,在位置1007处的所述甘氨酸残基被具有相对大的侧链的氨基酸残基取代。在一些实施方式中,在位置1007处的所述甘氨酸残基被精氨酸残基取代(G1007R)。在一些实施方式中,在位置1007处的所述甘氨酸残基被赖氨酸残基取代(G1007K)。在一些实施方式中,在位置1007处的所述甘氨酸残基被酪氨酸残基取代(G1007Y)。此外,在其他实施方式中,在位置1007处的所述甘氨酸残基被亮氨酸残基取代(G1007L)。在其他实施方式中,在位置1007处的所述甘氨酸残基被苏氨酸残

基取代(G1007T)。在其他实施方式中,在位置 1007处的所述甘氨酸残基被丝氨酸残基取代(G1007S)。

[0847] 在一些实施方式中,所述腺苷脱氨酶包含在所述hADAR1-D氨基酸序列的谷氨酸1008或同源ADAR蛋白中的对应的位置处的突变。在一些实施方式中,在位置1008处的所述谷氨酸残基被具有相对大的侧链的极性氨基酸残基取代。在一些实施方式中,在位置1008处的所述谷氨酸残基被谷氨酰胺残基取代(E1008Q)。在一些实施方式中,在位置1008处的所述谷氨酸残基被组氨酸残基取代(E1008H)。在一些实施方式中,在位置1008处的所述谷氨酸残基被精氨酸残基取代(E1008R)。在一些实施方式中,在位置1008 处的所述谷氨酸残基被赖氨酸残基取代(E1008K)。在一些实施方式中,在位置1008处的所述谷氨酸残基被非极性或小的极性氨基酸残基取代。在一些实施方式中,在位置1008处的所述谷氨酸残基被苯基丙氨酸残基取代(E1008F)。在一些实施方式中,在位置1008处的所述谷氨酸残基被色氨酸残基取代(E1008W)。在一些实施方式中,在位置1008处的所述谷氨酸残基被甘氨酸残基取代(E1008G)。在一些实施方式中,在位置1008处的所述谷氨酸残基被异亮氨酸残基取代(E1008I)。在一些实施方式中,在位置1008处的所述谷氨酸残基被缬氨酸残基取代(E1008V)。在一些实施方式中,在位置 1008处的所述谷氨酸残基被脯氨酸残基取代(E1008P)。在一些实施方式中,在位置1008处的所述谷氨酸残基被丝氨酸残基取代(E1008S)。在其他实施方式中,在位置1008处的所述谷氨酸残基被天冬氨酸残基取代(E1008N)。在其他实施方式中,在位置1008处的所述谷氨酸残基被丙氨酸残基取代(E1008A)。在其他实施方式中,在位置1008处的所述谷氨酸残基被甲硫氨酸残基取代(E1008M)。在一些实施方式中,在位置1008处的所述谷氨酸残基被亮氨酸残基取代(E1008L)。

[0848] 在一些实施方式中,为了改善编辑效率,所述腺苷脱氨酶可包含以下一个或多个突变:E1007S、E1007A、E1007V、E1008Q、E1008R、E1008H、E1008M、E1008N、E1008K(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。

[0849] 在一些实施方式中,为了降低编辑效率,所述腺苷脱氨酶可包含以下一个或多个突变:E1007R、E1007K、E1007Y、E1007L、E1007T、E1008G、E1008I、E1008P、E1008V、E1008F、E1008W、E1008S、E1008N、E1008K(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。

[0850] 在一些实施方式中,腺苷脱氨酶的底物编辑偏好、效率和/或选择性受所述酶的活性位点附近或活性位点中的氨基酸残基的影响。在一些实施方式中,所述腺苷脱氨酶包含在hADAR1-D序列中的谷氨酸1008位置或同源 ADAR蛋白中的对应的位置处的突变。在一些实施方式中,所述突变是 E1008R或同源ADAR蛋白中的对应的突变。在一些实施方式中,所述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。

[0851] 在一些实施方式中,所述腺苷脱氨酶蛋白进一步包含一个或多个双链 RNA(dsRNA)结合基序(dsRBM)或结构域(dsRBD)或与其连接以识别双链核酸底物并与所述双链核酸底物结合。在一些实施方式中,所述腺苷脱氨酶和所述双链底物之间的相互作用由一个或多个附加蛋白因子(包括 CRISPR/CAS蛋白因子)介导。在一些实施方式中,所述腺苷脱氨酶和所述双链底物之间的相互作用由一个或多个核酸组分(包括引导RNA)进一步介导。

[0852] 在某些示例实施方式中,可用定向进化来设计能够催化除腺嘌呤至次黄嘌呤的脱

氨之外的附加反应的修饰ADAR蛋白。

[0853] 具有C至U的脱氨活性的修饰腺苷脱氨酶

[0854] 在某些示例实施方式中,可用定向进化来设计能够催化除腺嘌呤至次黄嘌呤的脱氨之外的附加反应的修饰ADAR蛋白。例如,所述修饰ADAR蛋白可能能够催化胞苷至尿嘧啶的脱氨。不受特定理论束缚的,改善C至U活性的突变可改变所述结合口袋的形状以使其更适合较小的胞苷碱基。在一些情况下,经修饰的ADAR包含在催化核心中的残基和/或与RNA靶标接触的残基上的突变。在催化核心中的残基上的突变的实例包括V351G和 K350I(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在与RNA靶标接触的残基上的突变的实例包括S486A和 S495N(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。

[0855] 在某些实施方式中,工程化所述腺苷脱氨酶以将所述活性转化为胞苷脱氨酶。这样的工程化腺苷脱氨酶也可保留其腺苷脱氨酶活性,即这样的突变腺苷脱氨酶可具有腺苷脱氨酶和胞苷脱氨酶活性两者。因此,在一些实施方式中,所述腺苷脱氨酶包含在选自以下的位置中的一个或多个突变: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。在特定实施方式中,所述腺苷脱氨酶包含一个或多个突变in a位置selected from V351,L444,V355,V525和 I520。在一些实施方式中,所述腺苷脱氨酶可包含在选自以下的位置中的一个或多个突变:E488、V351、S486、T375、S370、P462、N597(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。

[0856] 在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变: E488Q(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I(基于hADAR2-D的氨基酸序

列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施方式中,所述腺苷脱氨酶可包含以下一个或多个突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一些实施例中,本文中提供的包括突变腺苷脱氨酶,例如包含以下一个或多个突变的腺苷脱氨酶:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T(基于hADAR2-D的氨基酸序列位置,和对应于上述的同源ADAR蛋白中的突变),其与死的CRISPR-Cas蛋白或CRISPR-Cas切口酶融合。在一个特定实施例中,本文提供的包括突变腺苷脱氨酶,例如包含以下突变的腺苷脱氨酶:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E,和S661T(基于hADAR2-D的氨基酸序列位置,和对应于上述的同源ADAR蛋白中的突变),其与死的CRISPR-Cas蛋白或CRISPR-Cas切口酶融合。

[0857] 在一些实施方式中,具有C至U的脱氨活性的修饰腺苷脱氨酶包含在所述hADAR2-D氨基酸序列的以下位置中的任一个或多个或同源ADAR蛋白中的对应的位置处的突变:V351、T375、R455和E488。在一些实施方式中,所述腺苷脱氨酶包含突变E488Q。在一些实施方式中,所述腺苷脱氨酶包含一个或多个突变,所述一个或多个突变选自V351I、V351L、V351F、V351M、V351C、V351A、V351G、V351P、V351T、V351S、V351Y、V351W、V351Q、V351N、V351H、V351E、V351D、V351K、V351R、T375I、T375L、T375V、T375F、T375M、T375C、T375A、T375G、T375P、T375S、T375Y、T375W、T375Q、T375N、T375H、T375E、T375D、T375K、T375R、R455I、R455L、R455V、R455F、R455M、R455C、R455A、R455G、R455P、R455T、R455S、R455Y、R455W、R455Q、R455N、R455H、R455E、R455D、R455K。在一些实施方式中,所述腺苷脱氨酶包含突变E488Q并进一步包含选自以下的一个或多个突变:V351I、V351L、V351F、V351M、V351C、V351A、V351G、V351P、V351T、V351S、V351Y、V351W、V351Q、V351N、V351H、V351E、V351D、

V351K、V351R、T375I、T375L、T375V、T375F、T375M、T375C、T375A、T375G、T375P、T375S、T375Y、T375W、T375Q、T375N、T375H、T375E、T375D、T375K、T375R、R455I、R455L、R455V、R455F、R455M、R455C、R455A、R455G、R455P、R455T、R455S、R455Y、R455W、R455Q、R455N、R455H、R455E、R455D、R455K。

[0858] 在一些情况下,所述修饰ADAR可进一步包含降低脱靶活性的一个或多个突变。在修饰ADAR具有C至U的脱氨活性的情况下,这样的突变可降低A至I的脱靶活性和增加C至U的靶向脱氨活性。一般而言,这样的突变可以在与RNA靶标相互作用的残基上。这样的突变的实例包括S375N、S375C、S375A和N473I(基于hADAR2-D的氨基酸序列位置),和对应于上述的同源ADAR蛋白中的突变。在一个实施例中,所述ADAR具有S375N突变。在一个实施例中,本文提供的包括突变腺苷脱氨酶,例如包含以下突变的腺苷脱氨酶:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T,和S375N(基于hADAR2-D的氨基酸序列位置,和对应于上述的同源ADAR蛋白中的突变),其与死的CRISPR-Cas蛋白或CRISPR-Cas切口酶融合。

[0859] 与上述具有C至U的脱氨活性的修饰ADAR蛋白相关联的,本文所述的发明还涉及用于使感兴趣的靶RNA序列中的C脱氨的方法,其包括将本文公开的AD官能化组合物递送至靶RNA或DNA。

[0860] 在某些示例实施方式中,所述用于使靶RNA序列中的C脱氨的方法包括向所述靶RNA递送以下物质:(a)催化失活的(死的)Cas;(b)引导分子,所述引导分子包含与正向重复序列连接的引导序列;和(c)具有C至U的脱氨活性或其催化结构域的修饰ADAR蛋白;其中所述修饰ADAR蛋白或其催化结构域与所述死的Cas蛋白或所述引导分子共价或非共价连接或被调适为在递送后与其连接;其中所述引导分子与所述死的Cas蛋白形成复合物并指引所述复合物结合所述感兴趣的靶RNA序列;其中所述引导序列能够与包含所述C的靶序列杂合以形成RNA双链核酸分子;其中,任选地,所述引导序列包含在对应于所述C的位置处的未配对的A或U,导致错配在所述RNA双链核酸分子中形成;并且其中所述修饰ADAR蛋白或其催化结构域使所述RNA双链核酸分子中的所述C脱氨。

[0861] 与上述具有C至U的脱氨活性的修饰ADAR蛋白相关联的,本文所述的发明进一步涉及适于使感兴趣的基因座中的C脱氨的工程化、非自然存在的系统,其包含:(a)引导分子,所述引导分子包含与正向重复序列连接的引导序列,或编码所述引导分子的核苷酸序列;(b)催化失活的CRISPR-Cas蛋白,或编码所述催化失活的CRISPR-Cas蛋白的所述核苷酸序列;(c)修饰ADAR蛋白,其具有C至U的脱氨活性或其催化结构域,或编码所述修饰ADAR蛋白或其催化结构域的核苷酸序列;其中所述修饰ADAR蛋白或其催化结构域与所述CRISPR-Cas蛋白或所述引导分子共价或非共价连接或被调适为在递送后与其连接;其中所述引导序列能够与包含C的靶RNA序列杂合以形成RNA双链核酸分子;其中,任选地,所述引导序列包含在对应于所述C的位置处的未配对的A或U,导致错配在所述RNA双链核酸分子中形成;其中,任选地,所述系统是载体系统,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)与编码包含所述引导序列地所述引导分子的核苷酸序列可操作地连接的第一调节元素,(b)与编码所述催化失活的CRISPR-Cas蛋白的核苷酸序列可操作地连接的第二调节元素;和(c)编码具有C至U的脱氨活性或其催化结构域的修饰蛋白的核苷酸序列,其在所述第一调节元素或所述第二调节元素的控制下,或与第三调节元素可操作地连接;其中,如果

编码修饰ADAR蛋白或其催化结构域的所述核苷酸序列与第三调节元素可操作地连接,所述修饰ADAR蛋白或其催化结构域被调适为在表达后与所述引导分子或所述CRISPR-Cas蛋白连接;其中组分(a)、(b)和(c)位于所述系统的相同或不同载体上,任选地,其中所述第一调节元素、第二调节元素和/或第三调节元素是可诱导的启动子。

[0862] 在本发明的实施方式中,所述腺苷脱氨酶的所述底物是在所述引导分子与其DNA靶标结合之后形成的RNA/DNA异源双链核酸分子,其然后与所述CRISPR-Cas酶形成所述CRISPR-Cas复合物。所述RNA/DNA或DNA/RNA异源双链核酸分子在本文中也称为“RNA/DNA杂合体”、“DNA/RNA杂合体”或“双链底物”。

[0863] 根据本发明,所述腺苷脱氨酶的所述底物是在所述引导分子与其DNA靶标结合之后形成的RNA/DNRNA双链核酸分子,其然后与所述CRISPR-Cas酶形成所述CRISPR-Cas复合物。所述腺苷脱氨酶的所述底物也可以是在所述引导分子与其DNA靶标结合之后形成的RNA/RNA双链核酸分子,其然后与所述CRISPR-Cas酶形成所述CRISPR-Cas复合物。所述RNA/DNA或DNA/RNRNA双链核酸分子在本文中也称为“RNA/DNA杂合体”、“DNA/RNA杂合体”或“双链底物”。所述引导分子和CRISPR-Cas酶的具体特征描述于下文。

[0864] 本文所用的术语“编辑选择性”指由腺苷脱氨酶编辑的双链底物上的所有位点的部分。不受理论束缚的,预期腺苷脱氨酶的编辑选择性受所述双链底物的长度和二级结构(例如错配的碱基、突起和/或内环的存在)的影响。

[0865] 在一些实施方式中,当所述底物与长于50bp的双链核酸分子完美地碱基配对时,所述腺苷脱氨酶可使所述双链核酸分子内的多个腺苷残基(例如所有腺苷残基中的50%)脱氨。在一些实施方式中,当所述底物短于50bp时,腺苷脱氨酶的编辑选择性受所述靶腺苷位点处的错配的存在的影响。特别是,在一些实施方式中,以高效率使在相反链上具有错配的胞苷(C)残基的腺苷(A)残基脱氨。在一些实施方式中,跳过在相反链上具有鸟苷(G)残基的腺苷(A)残基而编辑所述腺苷(A)残基。

[0866] 在特定实施方式中,所述腺苷脱氨酶蛋白或其催化结构域被递送至所述细胞或在所述细胞内作为单独的蛋白表达,但被修饰为能够与所述Cas蛋白或所述引导分子连接。在特定实施方式中,这是通过正交RNA-结合蛋白或存在于菌体衣壳蛋白的多样性中的接头蛋白/适配体组合物的使用来确保的。这样的衣壳蛋白的实例包括但不限于以下:MS2、Q $\beta$ 、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、 $\phi$  Cb5、 $\phi$  Cb8r、 $\phi$  Cb12r、 $\phi$  Cb23r、7s和PRR1。适配体可以是已通过体外选择或SELEX(通过指数富集进行的系统进化)的重复回合而被工程化为与特定靶标结合的自然存在的或合成的寡核苷酸。

[0867] 在特定实施方式中,给所述引导分子提供了一个或多个可募集接头蛋白的不同的RNA环或不同的序列。可通过插入可募集接头蛋白的不同的RNA环或不同的序列来延伸引导分子,而不使其与所述Cas蛋白碰撞,所述接头蛋白可与所述不同的RNA环或不同的序列结合。修饰引导的实例和它们在为所述Cas复合物募集效应子结构域中的用途提供于Konermann(Nature 2015, 517(7536):583-588)中。在特定实施方式中,所述适配体是选择性结合哺乳动物细胞中的二聚化MS2噬菌体衣壳蛋白并被引入(例如茎环和/或四环中的)引导分子中的最小的发夹适配体。在这些实施方式中,所述腺苷脱氨酶蛋白与MS2融合。然后将所述腺苷脱氨酶蛋白与所述Cas蛋白和对应的引导RNA一起共递送。

[0868] 在一些实施方式中,本文所述的Cas-ADAR碱基编辑系统包含(a) Cas蛋白,所述Cas蛋白是催化失活的或是切口酶;(b) 包含引导序列的引导分子;和(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Cas蛋白或所述引导分子共价或非共价连接或被调适为在递送后与其连接;其中所述引导序列与所述靶序列基本上互补但包含对应于被指定为脱氨的A的未配对的C,导致通过所述引导序列和所述靶序列形成的DNA-RNA或RNA-RNA双链核酸分子中的A-C错配。至于在真核细胞中的应用,所述Cas蛋白和/或所述腺苷脱氨酶优选地为NLS-标记的。

[0869] 在一些实施方式中,将所述组分(a)、(b)和(c)作为核糖核蛋白复合物递送至所述细胞。可通过一个或多个脂质纳米颗粒来递送所述核糖核蛋白复合物。

[0870] 在一些实施方式中,将所述组分(a)、(b)和(c)作为一个或多个RNA分子(例如编码所述Cas蛋白、所述腺苷脱氨酶蛋白和任选地所述接头蛋白的一个或多个引导RNA和一个或多个mRNA分子)递送至所述细胞。可通过一个或多个脂质纳米颗粒来递送所述RNA分子。

[0871] 在一些实施方式中,将所述组分(a)、(b)和(c)作为一个或多个DNA分子递送至所述细胞。在一些实施方式中,所述一个或多个DNA分子包含于一个或多个载体(例如病毒载体(例如AAV))中。在一些实施方式中,所述一个或多个DNA分子包含一个或多个调节元素,所述一个或多个调节元素被可操作性地配置为表达所述Cas蛋白、所述引导分子和所述腺苷脱氨酶蛋白或其催化结构域,任选地,其中所述一个或多个调节元素包含可诱导的启动子。

[0872] 在一些实施方式中,所述引导分子能够与包含将要在所述靶基因座处的第一DNA链或RNA链中脱氨的腺嘌呤的靶序列杂合,以形成包含与所述腺嘌呤相对的未配对的胞嘧啶的DNA-RNA或RNA-RNA双链核酸分子。在双链核酸分子形成之后,所述引导分子与所述Cas蛋白形成复合物并引导所述复合物结合在所述感兴趣的靶基因座处的所述第一DNA链或所述RNA链。所述Cas-ADAR碱基编辑系统的引导方面的细节提供于下文中。

[0873] 在一些实施方式中,使用具有经典长度(例如针对AacCas约20nt)的Cas引导RNA以与所述靶DNA或RNA形成DNA-RNA或RNA-RNA双链核酸分子。在一些实施方式中,使用长于所述标准长度(例如针对AacCas >20nt)的Cas引导分子以与所述靶DNA或RNA(包括在所述Cas-引导RNA-靶DNA复合物之外)形成DNA-RNA或RNA-RNA双链核酸分子。在某些示例实施方式中,所述引导序列长度为约29-53nt,其能够与所述靶序列形成DNA-RNA或RNA-RNA双链核酸分子。在某些其他示例实施方式中,所述引导序列长度为约40-50nt,其能够与所述靶序列形成DNA-RNA或RNA-RNA双链核酸分子。在某些示例实施方式中,所述未配对的C和所述引导序列的5'末端之间的距离为20-30个核苷酸。在某些示例实施方式中,所述未配对的C和所述引导序列的3'末端之间的距离为20-30个核苷酸。

[0874] 至少在第一个设计中,所述Cas-ADAR系统包含(a)与Cas蛋白融合或连接的腺苷脱氨酶,其中所述Cas蛋白是催化失活的或是切口酶,和(b)包含引导序列的引导分子,所述引导序列被设计为引入在所述引导序列和所述靶序列之间形成的DNA-RNA或RNA-RNA双链核酸分子中的A-C错配。在一些实施方式中,所述Cas蛋白和/或所述腺苷脱氨酶在N或C末端中的任一者或两者之上是NLS-标记的。

[0875] 至少在第二个设计中,所述Cas-ADAR系统包含(a) Cas蛋白,所述Cas蛋白是催化失活的或是切口酶,(b) 包含引导序列和能够与接头蛋白(例如MS2衣壳蛋白或PP7衣壳蛋

白)结合的适配体序列(例如MS2 RNA基序或PP7 RNA基序)的引导分子,所述引导序列被设计为引入在所述引导序列和所述靶序列之间形成的DNA-RNA或RNA-RNA双链核酸分子中的A-C错配,和(c)与接头蛋白融合或连接的腺苷脱氨酶,其中所述适配体和所述接头蛋白的结合将所述腺苷脱氨酶募集至在所述引导序列和所述靶序列之间形成的所述DNA-RNA或RNA-RNA双链核酸分子以用于在所述A-C错配的A处进行靶向脱氨。在一些实施方式中,所述接头蛋白和/或所述腺苷脱氨酶在N 或C末端中的任一者或两者上是NLS-标记的。所述Cas蛋白也可以是NLS-标记的。

[0876] 不同适配体和对应的接头蛋白的使用也允许了正交基因编辑的实现。在一个实施例(其中将腺苷脱氨酶与胞苷脱氨酶联合使用以用于正交基因编辑/脱氨)中,用不同的RNA环修饰靶向不同基因座的sgRNA以分别募集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),从而分别导致所述感兴趣的靶基因座处的A或C的正交脱氨。PP7是噬菌体假单胞菌的RNA-结合衣壳蛋白。如MS2,它结合特定的RNA序列和二级结构。所述PP7 RNA-识别基序不同于的MS2的RNA-识别基序。因此,可以多重复用PP7和MS2以同时介导不同基因座处的不同作用。例如,靶向基因座A的sgRNA可用MS2环修饰,从而募集MS2-腺苷脱氨酶,而靶向基因座B的另一sgRNA可用PP7环修饰,从而募集PP7-胞苷脱氨酶。在相同细胞中,正交的基因座特异性修饰从而得以实现。可以延伸该原理以包含其他正交RNA-结合蛋白。

[0877] 至少在第三个设计中,所述Cas-ADAR CRISPR系统包含(a)插入Cas蛋白的内环或非结构化区域中的腺苷脱氨酶,其中所述Cas蛋白是催化失活的或是切口酶,和(b)包含引导序列的引导分子,所述引导序列被设计为引入在所述引导序列和所述靶序列之间形成的DNA-RNA或RNA-RNA双链核酸分子中的A-C错配。

[0878] 可借由晶体结构来识别适合于腺苷脱氨酶插入的Cas蛋白分裂位点。例如,对于AacCas突变体而言,其用于例如序列比对的对应的位置是什么应是显而易见的。至于其他Cas蛋白,如果直系同源物和预期的Cas蛋白之间存在较高程度的同源性,则可以使用所述直系同源物的晶体结构。

[0879] 所述分裂位置可位于区域或环内。优选地,所述分裂位置发生于所述氨基酸序列的中断不导致结构特征(例如 $\alpha$ -螺旋或 $\beta$ -折叠)的部分或全部破坏处。非结构化区域(未在所述晶体结构中出现的区域,因为这些区域的结构化程度不足以“冻存”于晶体中)常常是优选的选择。在本发明的实践中预见了在暴露于Cas的表面上的所有非结构化区域中的分裂。在所述非结构化区域或外环内的位置可不需恰巧为上文所提供的数字,而可在上文提供的位置处的任一侧基于所述环的大小而变化1、2、3、4、5、6、7、8、9或甚至10个氨基酸,只要所述分裂位置仍旧落于外环的非结构化区域内。

[0880] 本文所述的Cas-ADAR系统可用于靶向DNA序列内的特定的腺嘌呤以用于脱氨。例如,所述引导分子可与所述Cas蛋白形成复合物并引导所述复合物结合所述感兴趣的靶基因座处的靶序列。因为所述引导序列被设计为具有未配对的C,在所述引导序列和所述靶序列之间形成的异源双链核酸分子包含A-C错配,所述A-C错配引导所述腺苷脱氨酶接触与所述未配对的C 相对的A并使其脱氨,从而将其转化为肌苷(I)。由于肌苷(I)与C碱基配对并在细胞过程中像G一样运作,本文所述的A的靶向脱氨可用于纠正不良的G-A和C-T突变以及获得理想的A-G和T-C突变。在一些实施方式中,所述引导可包含一个或多个错配以增加特

异性。例如,所述引导可在脱靶腺苷对面包含一个或多个不良的鸟嘌呤错配。

[0881] 碱基切除修复抑制剂

[0882] 在一些实施方式中,所述AD-官能化CRISPR系统进一步包含碱基切除修复(BER)抑制剂。不期望受任何特定理论限制,对I:T配对的存在的细胞DNA-修复应答可能是导致细胞中核碱基编辑效率降低的原因。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的去除,所述去除可启动碱基切除修复并以所述I:T对至A:T对的回复为结果。

[0883] 在一些实施方式中,所述BER抑制剂是烷基腺嘌呤DNA糖基化酶的抑制剂。在一些实施方式中,所述BER抑制剂是人烷基腺嘌呤DNA糖基化酶的抑制剂。在一些实施方式中,所述BER抑制剂是多肽抑制剂。在一些实施方式中,所述BER抑制剂是结合次黄嘌呤的蛋白。在一些实施方式中,所述BER抑制剂是结合DNA中的次黄嘌呤的蛋白。在一些实施方式中,所述BER抑制剂是催化失活的烷基腺嘌呤DNA糖基化酶蛋白或其结合结构域。在一些实施方式中,所述BER抑制剂是不从所述DNA切除次黄嘌呤的催化失活的烷基腺嘌呤DNA糖基化酶蛋白或其结合结构域。其他能够抑制(例如位阻)烷基腺嘌呤DNA糖基化酶碱基-切除修复酶的蛋白在本公开的范围之内。此外,任何阻碍或抑制碱基-切除修复的蛋白也在本公开的范围之内。

[0884] 不期望受任何特定理论限制,碱基切除修复可被结合编辑链,阻碍编辑碱基,抑制烷基腺嘌呤DNA糖基化酶,抑制碱基切除修复,保护编辑碱基和/或促进未编辑链的修复的分子抑制。相信本文所述的BER抑制剂的使用可增加能够催化A至I的变化的腺苷脱氨酶的编辑效率。

[0885] 因此,在上文所述的AD-官能化CRISPR系统的所述第一个设计中,所述CRISPR-Cas蛋白或所述腺苷脱氨酶可与BER抑制剂(例如烷基腺嘌呤DNA糖基化酶的抑制剂)融合或连接。在一些实施方式中,所述BER抑制剂可包含于以下结构之一(nCas=Cas切口酶;dCas=死的Cas)中:

[0886] [AD]-[任选的接头]-[nCas/dCas]-[任选的接头]-[BER抑制剂];

[0887] [AD]-[任选的接头]-[BER抑制剂]-[任选的接头]-[nCas/dCas];

[0888] [BER抑制剂]-[任选的接头]-[AD]-[任选的接头]-[nCas/dCas];

[0889] [BER抑制剂]-[任选的接头]-[nCas/dCas]-[任选的接头]-[AD];

[0890] [nCas/dCas]-[任选的接头]-[AD]-[任选的接头]-[BER抑制剂];

[0891] [nCas/dCas]-[任选的接头]-[BER抑制剂]-[任选的接头]-[AD]。

[0892] 相似地,在上文所述的AD-官能化CRISPR系统的所述第二个设计中,所述CRISPR-Cas蛋白、所述腺苷脱氨酶或所述接头蛋白可与BER抑制剂(例如烷基腺嘌呤DNA糖基化酶的抑制剂)融合或连接。在一些实施方式中,所述BER抑制剂可包含于以下结构之一(nCas=Cas切口酶;dCas=死的Cas):

[0893] [nCas/dCas]-[任选的接头]-[BER抑制剂];

[0894] [BER抑制剂]-[任选的接头]-[nCas/dCas];

[0895] [AD]-[任选的接头]-[接头]-[任选的接头]-[BER抑制剂];

[0896] [AD]-[任选的接头]-[BER抑制剂]-[任选的接头]-[接头];

[0897] [BER抑制剂]-[任选的接头]-[AD]-[任选的接头]-[接头];

[0898] [BER抑制剂]-[任选的接头]-[接头]-[任选的接头]-[AD];

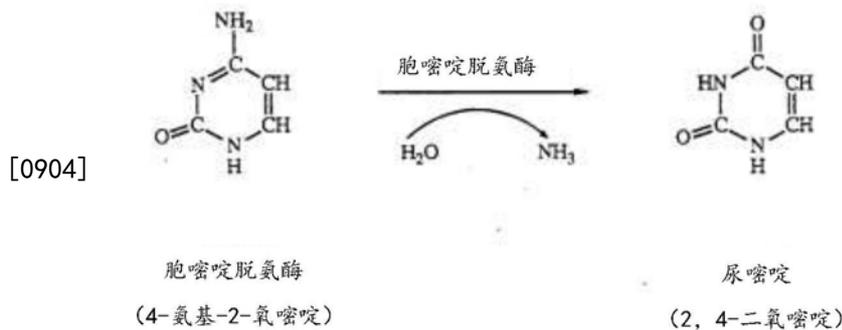
[0899] [接头]-[任选的接头]-[AD]-[任选的接头]-[BER抑制剂];

[0900] [接头]-[任选的接头]-[BER抑制剂]-[任选的接头]-[AD].

[0901] 在上文所述的AD-官能化CRISPR系统的所述第三个设计中,所述BER抑制剂可被插入至CRISPR-Cas蛋白的内环或非结构化区域内。

[0902] 胞苷脱氨酶

[0903] 在一些实施方式中,所述脱氨酶是胞苷脱氨酶。本文所用的术语“胞苷脱氨酶”或“胞苷脱氨酶蛋白”或“胞苷脱氨酶活性”指蛋白,多肽,或者能够催化如上文所示的将胞嘧啶(或分子的胞嘧啶部分)转化为尿嘧啶(或分子的尿嘧啶部分)的水解脱氨反应的蛋白或多肽的一个或多个功能性结构域在一些实施方式中,所述包含胞嘧啶的分子是胞苷(C),并且所述包含尿嘧啶的分子是尿苷(U)。所述包含胞嘧啶的分子可以是脱氧核糖核酸(DNA)或核糖核酸(RNA)。在某些实施例中,胞苷脱氨酶可以是作用于RNA上的胞苷脱氨酶(CDAR)。



[0905] 根据本公开,可结合本公开使用的胞苷脱氨酶包括但不限于被称为载脂蛋白B mRNA-编辑复合物(APOBEC)家族脱氨酶、诱导激活的脱氨酶(AID)或胞苷脱氨酶1(CDA1)的酶家族成员。在特定实施方式中,所述脱氨酶在APOBEC1脱氨酶、APOBEC2脱氨酶、APOBEC3A脱氨酶、APOBEC3B脱氨酶、APOBEC3C脱氨酶和APOBEC3D脱氨酶、APOBEC3E脱氨酶、APOBEC3F脱氨酶、APOBEC3G脱氨酶、APOBEC3H脱氨酶或APOBEC4脱氨酶中。

[0906] 在本发明的方法和系统中,所述胞苷脱氨酶或具有胞苷脱氨酶活性的工程化腺苷脱氨酶能够靶向DNA单链中的胞嘧啶。在一些示例性实施方式中,所述胞苷脱氨酶活性可在存在于所述结合组分(例如结合的CRISPR-Cas)之外的单链上编辑。在其他示例实施方式中,所述胞苷脱氨酶可在局部泡(例如通过在所述靶编辑位点而非所述引导序列处的错配形成的局部泡)处编辑。在一些示例性实施方式中,所述胞苷脱氨酶可包含帮助集中活性(如在Kim等, Nature Biotechnology (2017) 35 (4):371-377 (doi:10.1038/nbt.3803)中公开的那些活性)区域的突变。

[0907] 在一些实施方式中,所述胞苷脱氨酶源自一个或多个后生动物物种,所述一个或多个后生动物物种包括但不限于哺乳动物、鸟类、青蛙、鱿鱼、鱼类、苍蝇和蠕虫。在一些实施方式中,所述胞苷脱氨酶是人、灵长类动物、牛、狗、大鼠或小鼠胞苷脱氨酶。

[0908] 在一些实施方式中,所述胞苷脱氨酶是人APOBEC,所述人APOBEC包括但不限于hAPOBEC1或hAPOBEC3。在一些实施方式中,所述胞苷脱氨酶是人AID。

[0909] 在一些实施方式中,所述胞苷脱氨酶蛋白识别RNA双链核酸分子的单链泡中的一个或多个靶胞嘧啶残基并将其转化为尿嘧啶残基。在一些实施方式中,所述胞苷脱氨酶蛋白识别RNA双链核酸分子的所述单链泡上的结合窗口。在一些实施方式中,所述结合窗口包

含至少一个靶胞嘧啶残基。在一些实施方式中,所述结合窗口范围为约3bp至约100bp。在一些实施方式中,所述结合窗口范围为约5bp至约50bp。在一些实施方式中,所述结合窗口范围为约10bp至约30bp。在一些实施方式中,所述结合窗口为约1bp、2 bp、3bp、5bp、7bp、10bp、15bp、20bp、25bp、30bp、40bp、45bp、50bp、55bp、60bp、65bp、70bp、75bp、80bp、85bp、90bp、95bp或100bp。

[0910] 在一些实施方式中,所述胞苷脱氨酶蛋白包含一个或多个脱氨酶结构域。不打算受理论束缚,预期所述脱氨酶结构域发挥作用以识别包含于RNA 双链核酸分子的单链泡中的一个或多个靶胞嘧啶(C)残基并将其转化为尿嘧啶(U)残基。在一些实施方式中,所述脱氨酶结构域包含活性中心。在一些实施方式中,所述活性中心包含锌离子。在一些实施方式中,在所述活性中心或附近的氨基酸残基与靶胞嘧啶残基5'末端处的一个或多个核苷酸相互作用。在一些实施方式中,在所述活性中心或附近的氨基酸残基与靶胞嘧啶残基3'末端处的一个或多个核苷酸3'相互作用。

[0911] 在一些实施方式中,所述胞苷脱氨酶包含人APOBEC1全蛋白(hAPOBEC1)或其脱氨酶结构域(hAPOBEC1-D)或其C末端截短形式(hAPOBEC-T)。在一些实施方式中,所述胞苷脱氨酶是与hAPOBEC1、hAPOBEC-D或hAPOBEC-T同源的APOBEC家族成员。在一些实施方式中,所述胞苷脱氨酶包含人AID1全蛋白(hAID)或其脱氨酶结构域(hAID-D)或其C末端截短形式(hAID-T)。在一些实施方式中,所述胞苷脱氨酶是与hAID、hAID-D或hAID-T同源的AID家族成员。在一些实施方式中,所述hAID-T是C末端处截短20个氨基酸的hAID。

[0912] 在一些实施方式中,所述胞苷脱氨酶包含胞嘧啶脱氨酶的野生型氨基酸序列。在一些实施方式中,所述胞苷脱氨酶包含在所述胞嘧啶脱氨酶序列中的一个或多个突变,从而根据具体需求而改变所述胞嘧啶脱氨酶的编辑效率和/或底物编辑偏好。

[0913] APOBEC1和APOBEC3蛋白的某些突变已描述于Kim等,Nature Biotechnology (2017) 35 (4) :371-377 (doi:10.1038/nbt.3803);和Harris等,Mol. Cell (2002) 10:1247-1253,其每一个通过引用整体并入本文。

[0914] 在一些实施方式中,所述胞苷脱氨酶是APOBEC1脱氨酶或APOBEC3G脱氨酶,所述APOBEC1脱氨酶包含在对应于大鼠APOBEC1中的W90、R118、H121、H122、R126或R132的氨基酸位置处的一个或多个突变,所述APOBEC3G脱氨酶包含在对应于人APOBEC3G中的W285、R313、D316、D317X、R320或R326的氨基酸位置处的一个或多个突变。

[0915] 在一些实施方式中,所述胞苷脱氨酶包含在所述大鼠APOBEC1氨基酸序列的色氨酸90或同源APOBEC蛋白中的对应的位置(例如APOBEC3G的色氨酸285)处的突变。在一些实施方式中,在位置90处的所述色氨酸残基被酪氨酸或苯基丙氨酸残基(W90Y或W90F)取代。

[0916] 在一些实施方式中,所述胞苷脱氨酶包含在所述大鼠APOBEC1氨基酸序列的精氨酸118或同源APOBEC蛋白中的对应的位置处的突变。在一些实施方式中,在位置118处的所述精氨酸残基被丙氨酸残基取代(R118A)。

[0917] 在一些实施方式中,所述胞苷脱氨酶包含在所述大鼠APOBEC1氨基酸序列的组氨酸121或同源APOBEC蛋白中的对应的位置处的突变。在一些实施方式中,在位置121处的所述组氨酸残基被精氨酸残基取代(H121R)。

[0918] 在一些实施方式中,所述胞苷脱氨酶包含在所述大鼠APOBEC1氨基酸序列的组氨酸122或同源APOBEC蛋白中的对应的位置处的突变。在一些实施方式中,在位置122处的所

述组氨酸残基被精氨酸残基取代(H122R)。

[0919] 在一些实施方式中,所述胞苷脱氨酶包含在所述大鼠APOBEC1氨基酸序列的精氨酸126或同源APOBEC蛋白中的对应的位置(例如APOBEC3G 的精氨酸320)处的突变。在一些实施方式中,在位置126处的所述精氨酸残基被丙氨酸残基取代(R126A)或被谷氨酸取代(R126E)。

[0920] 在一些实施方式中,所述胞苷脱氨酶包含在所述APOBEC1氨基酸序列的精氨酸132或同源APOBEC蛋白中的对应的位置处的突变。在一些实施方式中,在位置132处的所述精氨酸残基被谷氨酸残基取代(R132E)。

[0921] 在一些实施方式中,为了缩小所述编辑窗口的宽度,所述胞苷脱氨酶可包含以下突变中的一个或多个:W90Y、W90F、R126E和R132E,(基于大鼠APOBEC1的氨基酸序列位置),和对应于上述的同源APOBEC蛋白中的突变。

[0922] 在一些实施方式中,为了降低编辑效率,所述胞苷脱氨酶可包含以下突变中的一个或多个:W90A、R118A、R132E,(基于大鼠APOBEC1的氨基酸序列位置),和对应于上述的同源APOBEC蛋白中的突变。在特定实施方式中,对使用具有降低的效能的胞苷脱氨酶以减小脱靶效应可能感兴趣。

[0923] 在一些实施方式中,所述胞苷脱氨酶使野生型大鼠 APOBEC1 (rAPOBEC1) 或其催化结构域。在一些实施方式中,所述胞苷脱氨酶包含在所述rAPOBEC1序列中的一个或多个突变,从而根据具体需求而改变rAPOBEC1的编辑效率和/或底物编辑偏好。

[0924] rAPOBEC1: MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELKTKCLLYEINWGGRHSIW RHTSQN  
TNKHVEVNFIEKFTTTERYFCPNTRCSITWFLSWSPCGECSRAITEF LSRYPHVTLFIYIARLYHHADPRNRQGLR  
DLISSGVTIQIMTEQESGYCWR NFVNYSNEAHWPYPHLLWVRLYVLELYCIILGLPPCLNILRRKQPQLTF  
FTIALQSCHYQRLPPHILWATGLK (SEQ ID NO:243)

[0925] 在一些实施方式中,所述胞苷脱氨酶是野生型人APOBEC1 (hAPOBEC1) 或其催化结构域。在一些实施方式中,所述胞苷脱氨酶包含在所述 hAPOBEC1序列中的一个或多个突变,从而根据具体需求而改变hAPOBEC1 的编辑效率和/或底物编辑偏好。

[0926] APOBEC1: MTSEKGPSTGDPTLRRRIEPWEFDVFDYDPRELKKEACLLYEIKWGMSRKI WRSSGKNT  
TNHVEVNFIEKFTSERDFHPSMCSITWFLSWSPCWECQAIR EFLSRHPGVTLVIYVARLFWHMDQQNRQGLRD  
LVNSGVTIQIMRASEYYH CWRNFVNYPGDEAHWPYPLWMLYALELHCIILSLPPCLKISRWWQN HLTFFR  
LHLQNCHYQTIPPHILLATGLIHPSVAWR (SEQ ID NO:244)

[0927] 在一些实施方式中,所述胞苷脱氨酶是野生型人 APOBEC3G (hAPOBEC3G) 或其催化结构域。在一些实施方式中,所述胞苷脱氨酶包含在所述hAPOBEC3G序列中的一个或多个突,从而根据具体需求而改变hAPOBEC3G的编辑效率和/或底物编辑偏好。

[0928] hAPOBEC3G: MELKYHPEMRFHWFHWSKWRKLHRDQEYEVTWYISWSPCKCTRDMATF LAEDPKVT  
LTIFVARLYYFWDPDYQEA LRSLCQKRDGPRATMKIMNYDEF QHCWSKFVYSQRELFEPPWNNLPKYILLHIMLG  
EILRHSMDPPTFTFNFN EPWVRGRHETYLCEVERMHNDTWLLNQRGFLCNQAPHKHGFLEGR HAELCFLD  
VIPFWKLDLDQDYRVTCFTSWSPCFSCAQEMAKFISKKNHVS LCIFTARIYDDQGRCEGLRTLAEAGAKISIMT  
YSEFKHCWDTFVDHQGCP FQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO:245)

[0929] 在一些实施方式中,所述胞苷脱氨酶是野生型海七鳃鳗(Petromyzon marinus) CDA1 (pmCDA1) 或其催化结构域。在一些实施方式中,所述胞苷脱氨酶包含在所述pmCDA1序

列中的一个或多个突变,从而根据具体需求而改变pmCDA1的编辑效率和/或底物编辑偏好。

[0930] pmCDA1: MTD AEYVRIHEKLDIYTFKKQFFNNKKS VSHRCYVLFELKRRGERRACFW GYAVNKPQS  
GTERGIHAEIFSIRKVEEYL RDNPGQFTINWYSSWSPCADCA EKILEWYNQELRGNGHTLKIWACKLYYEKNARN  
QIGLWNL RDNGVGLN VMSEHYQCCRKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMI QVKIL HHTKSPAV  
(SEQ ID NO:246)

[0931] 在一些实施方式中,所述胞苷脱氨酶是野生型人AID (hAID) 或其催化结构域。在一些实施方式中,所述胞苷脱氨酶包含在所述pmCDA1序列中的一个或多个突变,从而根据具体需求而改变pmCDA1的编辑效率和/或底物编辑偏好。

[0932] hAID: MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYL RNKNGCHVELLF  
LRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFL RGNPYLSLRIFTARLYFCEDRKAPEGLRRLHRAGVQI  
AIMTFKDYFYCW NTFVENHERTFKAWEGLHENSVR LSRQLRRILLPLYEVDDL RDAFRTLGLL D (SEQ ID NO:  
247)

[0933] 在一些实施方式中,所述胞苷脱氨酶是hAID的截短形式 (hAID-DC) 或其催化结构域。在一些实施方式中,所述胞苷脱氨酶包含在所述hAID-DC 中的一个或多个突变,从而根据具体需求而改变hAID-DC的编辑效率和/或底物编辑偏好 of hAID-DC。

[0934] hAID-DC: MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYL RNKNGCHVE  
LLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFL RGNPNLSLRIFTARLYFCEDRKAPEGLRRLHRAG  
VQIAIMTFKDYFYCW NTFVENHERTFKAWEGLHENSVR LSRQLRRILL (SEQ ID NO:248)

[0935] 所述胞苷脱氨酶的附加的实施方式公开于题为“核碱基编辑器及其用途 (Nucleobase Editor and Uses Thereof)”的W0 W02017/070632,其通过引用整体并入本文。

[0936] 在一些实施方式中,所述胞苷脱氨酶具有包围易受脱氨编辑的核苷酸的高效的脱氨窗口。因此,在一些实施方式中,所述“编辑窗口宽度”指在所述胞苷脱氨酶的编辑效率超过该靶位点的半最大值的指定靶位点处的核苷酸位置的数量。在一些实施方式中,所述胞苷脱氨酶的所述编辑窗口宽度的范围为约1至约6个核苷酸。在一些实施方式中,所述胞苷脱氨酶的所述编辑窗口宽度为1、2、3、4、5或6个核苷酸。

[0937] 不打算受理论束缚,预期在一些实施方式中,所述接头序列的长度影响所述编辑窗口宽度。在一些实施方式中,所述编辑窗口宽度随所述接头延伸 (例如从约3至约21个氨基酸) 而增加 (例如从约3至约6个核苷酸)。在非限制实施例中,16个残基的接头提供了约5个氨基酸的高效的脱氨窗口。在一些实施方式中,所述引导RNA的长度影响所述编辑窗口宽度。在一些实施方式中,缩短所述引导RNA导致所述胞苷脱氨酶的高效脱氨窗口缩小。

[0938] 在一些实施方式中,所述胞苷脱氨酶的突变影响所述编辑窗口宽度。在一些实施方式中,所述CD-官能化CRISPR系统的所述胞苷脱氨酶组分包含降低所述胞苷脱氨酶的催化效率的一个或多个突变,从而防止所述脱氨酶每个DNA结合事件脱氨多个胞苷。在一些实施方式中,在APOBEC1的残基90 (W90) 或同源序列中的对应的色氨酸残基处的色氨酸发生突变。在一些实施方式中,所述催化失活的CRISPR-Cas与包含W90Y或W90F突变的 APOBEC1突变体融合或连接。在一些实施方式中,在APOBEC3G的残基 285 (W285) 或同源序列中的对应的色氨酸残基处的色氨酸发生突变。在一些实施方式中,所述催化失活的CRISPR-Cas与包含W285Y或W285F突变的 APOBEC3G突变体融合或连接。

[0939] 在一些实施方式中,CD-官能化CRISPR系统的胞苷脱氨酶组分包含降低胞苷至所述脱氨酶活性位点的非最佳呈递的耐受性的一个或多个突变。在一些实施方式中,所述胞苷脱氨酶包含改变所述脱氨酶活性位点的底物结合活性的一个或多个突变。在一些实施方式中,所述胞苷脱氨酶包含改变将要被所述脱氨酶活性位点识别并与其结合的DNA的构象的一个或多个突变。在一些实施方式中,所述胞苷脱氨酶包含改变所述底物对所述脱氨酶活性位点的可及性的一个或多个突变。在一些实施方式中,在APOBEC1的残基126 或同源序列中的对应的精氨酸残基处的精氨酸发生突变。在一些实施方式中,所述催化失活的CRISPR-Cas与包含R126A或R126E突变的APOBEC1 融合或连接。在一些实施方式中,在APOBEC3G的残基320 (R320) 或同源序列中的对应的精氨酸残基处的色氨酸发生突变。在一些实施方式中,所述催化失活的CRISPR-Cas与包含R320A或R320E突变的APOBEC3G突变体融合或连接。在一些实施方式中,在APOBEC1的残基132 (R132) 或同源序列中的对应的精氨酸残基处的精氨酸发生突变。在一些实施方式中,所述催化失活的CRISPR-Cas与包含R132E突变的APOBEC1突变体融合或连接。

[0940] 在一些实施方式中,所述CD-官能化CRISPR系统的所述APOBEC1结构域包含选自W90Y、W90F、R126A、R126E和R132E的一个、两个或三个突变。在一些实施方式中,所述APOBEC1结构域包含W90Y和R126E的双重突变。在一些实施方式中,所述APOBEC1结构域包含W90Y和R132E 的双重突变。在一些实施方式中,所述APOBEC1结构域包含R126E和 R132E的双重突变。在一些实施方式中,所述APOBEC1结构域包含W90Y、 R126E和R132E三个突变。

[0941] 在一些实施方式中,在如本文公开的胞苷脱氨酶中的一个或多个突变将所述编辑窗口宽度减小至约2个核苷酸。在一些实施方式中,在如本文公开的胞苷脱氨酶中的一个或多个突变将所述编辑窗口宽度减小至约1个核苷酸。在一些实施方式中,在如本文公开的胞苷脱氨酶中的一个或多个突变减小所述编辑窗口宽度而仅轻微或适度地影响所述酶的编辑效率。在一些实施方式中,在如本文公开的胞苷脱氨酶中的一个或多个突变减小所述编辑窗口宽度而不降低所述酶的编辑效率。在一些实施方式中,在如本文公开的胞苷脱氨酶中的一个或多个突变可区分相邻胞苷核苷酸,否则所述相邻相邻胞苷核苷酸将被所述胞苷脱氨酶以类似的效率进行编辑。

[0942] 在一些实施方式中,所述胞苷脱氨酶蛋白进一步包含一个或多个双链 RNA (dsRNA) 结合基序 (dsRBM) 或结构域 (dsRBD) 或者与所述一个或多个双链RNA (dsRNA) 结合基序 (dsRBM) 或结构域 (dsRBD) 连接以用于识别双链核酸底物并与所述双链核酸底物结合。在一些实施方式中,所述胞苷脱氨酶与所述底物之间的相互作用由一个或多个附加蛋白因子(包括CRISPR/CAS蛋白因子)介导。在一些实施方式中,所述胞苷脱氨酶与所述底物之间的相互作用由一个或多个核酸组分(包括引导RNA)进一步介导。

[0943] 根据本发明,所述胞苷脱氨酶的所述底物是包含感兴趣的胞嘧啶的RNA 双链核酸分子的DNA单链泡,其在所述引导分子与其DNA靶标的结合之后可与所述胞苷脱氨酶接触,其然后与所述CRISPR-Cas酶,形成所述 CRISPR-Cas复合物,从而所述胞嘧啶脱氨酶与所述CRISPR-Cas复合物的一个或多个组分(即所述CRISPR-Cas酶和/或所述引导分子)融合或能够与所述一个或多个组分结合。所述引导分和CRISPR-Cas酶的具体特征详细描述于下文。

[0944] 所述胞苷脱氨酶或其催化结构域可为人、大鼠或七鳃鳗胞苷脱氨酶蛋白或其催化结构域。

[0945] 所述胞苷脱氨酶蛋白或其催化结构域可为载脂蛋白B mRNA-编辑复合物(APOBEC)家族脱氨酶。所述胞苷脱氨酶蛋白或其催化结构域可为诱导激活的脱氨酶(AID)。所述胞苷脱氨酶蛋白或其催化结构域可为胞苷脱氨酶 1(CDA1)。

[0946] 所述胞苷脱氨酶蛋白或其催化结构域可为APOBEC1脱氨酶。所述 APOBEC1脱氨酶可包含对应于大鼠APOBEC1中的W90A、W90Y、R118A、H121R、H122R、I26A、R126E或R132E的一个或多个突变,或APOBEC3G 脱氨酶包含对应于人APOBEC3G中的W285A、W285Y、R313A、D316R、D317R、R320A、R320E或R326E的一个或多个突变。

[0947] 所述系统可进一步包含尿嘧啶糖基化酶抑制剂(UGI)。在一些实施方式中,所述胞苷脱氨酶蛋白或其催化结构域与尿嘧啶糖基化酶抑制剂(UGI)一起被递送。所述GI可与所述胞苷脱氨酶蛋白或其催化结构域和/或催化失活的CRISPR-Cas蛋白连接(例如共价连接)。

[0948] 基因产物的翻译后修饰的调节

[0949] 在一些情况下,碱基编辑可用于调节基因产物的翻译后修饰。在一些情况下,可通过碱基编辑不可被修饰的氨基残基来使是翻译后修饰位点的氨基酸残基突变。这样的翻译后修饰的实例包括二硫键形成、糖基化、脂化、乙酰化、磷酸化、甲基化、泛素化、SUMO化或其任何组合。

[0950] 在一些实施方式中,本文所述的碱基编辑可调节Stat3/IRF-5通路,例如以减轻炎症。例如,IRF-5的Stat3、Thr10、Ser158、Ser309、Ser317、Ser451 和/或Ser462的Tyr705上的磷酸化可与白介素信号传导有关。可使用本文所述的碱基编辑器以使这些繁殖位点中的一个或多个突变以调节免疫、自身免疫和/或炎症。

[0951] 在一些实施方式中,本文所述的碱基编辑器可调节胰岛素受体底物(IRS) 通路。例如,Ser265、Ser302、Ser325、Ser336、Ser358、Ser407和/或 Ser408上的磷酸化可参与调节(例如抑制)ISR通路。替代地或附加地,可使小鼠中的丝氨酸307(或人中的丝氨酸)突变,从而调节所述磷酸化。例如,丝氨酸307磷酸化可导致IRS-1的降解并减少MAPK信号传导。可在胰岛素不敏感的条件下(例如胰岛素过度刺激和/或TNF $\alpha$ 治疗)诱导丝氨酸307磷酸化。在一些实施例中,可生成S307F突变以用于稳定IRS-1和所述通路中的其他组分之间的相互作用。可使用本文所述的碱基编辑以使这些繁殖位点中的一个或多个突变以用于for调节IRS通路。

[0952] 基因产物的稳定性的调节

[0953] 在一些实施方式中,碱基编辑可用于调节基因产物的稳定性。例如,可通过本文所述的碱基编辑器来使调节大鼠中蛋白降解的一个或多个氨基酸残基突变。在一些情况下,这样的氨基酸残基可在降解决定子中。降解决定子可指参与调节所述蛋白的降解速率的蛋白部分。降解决定子可包含短氨基酸序列、结构基序和暴露的氨基酸(例如赖氨酸或精氨酸)。一些蛋白可包含多个降解决定子。所述降解决定子为泛素依赖性(例如基于所述蛋白的泛素化调节蛋白降解)或非泛素依赖性。

[0954] 在一些情况下,可使用所述碱基编辑来使单个肽中的一个或多个氨基酸残基突变以用于蛋白降解。在一些实施例中,所述单个肽可为PEST序列,所述PEST序列是富含脯氨酸(P)、谷氨酸(E)、丝氨酸(S)和苏氨酸(T)的肽序列。例如,可增加包含PEST序列的NANOG的稳定性,例如以促进胚胎干细胞多能性。

[0955] 在一些实施例中,所述碱基编辑器可用于使SMN2突变(例如生成S270A 突变)以增加所述SMN2蛋白的稳定性,所述SMN2蛋白参与脊髓性肌萎缩。可通过碱基编辑器生成的SMN2中的其他突变包括描述于Cho S.等,Genes Dev.2010Mar 1;24(5):438-442中的那些。在某些实施例中,所述碱基编辑器可用于生成如Fortmann KT等,J Mol Biol.2015Aug 28;427(17):2748-2756中描述的I $\kappa$ B $\alpha$ 上的突变。可通过计算工具(例如 [slim.ucd.ie/apc/index.php](http://slim.ucd.ie/apc/index.php)上提供的线上工具)来识别降解决定子中的靶位点。其他靶标包括Cdc25A磷酸酶。

[0956] 可通过碱基编辑器靶向的基因的实例

[0957] 在一些实施例中,所述碱基编辑器可用于修饰PCSK9。所述碱基编辑器可引入终止密码子和/或降低PCSK9活性的疾病相关的突变。所述碱基编辑可引入PCSK9中的以下突变中的一个或多个:R46L、R46A、A53V、A53A、E57K、Y142X、L253F、R237W、H391N、N425S、A443T、I474V、I474A、Q554E、Q619P、E670G、E670A、C679X、H417Q、R469W、E482G、F515L 和/或H553R。

[0958] 在一些实施例中,所述碱基编辑器可用于修饰ApoE。所述碱基编辑器可靶向合成模型中的ApoE和/或源自患者的神经元(例如源自iPSC的那些)。可通过测序来测试所述靶向。

[0959] 在一些实施例中,所述碱基编辑器可用于修饰Stat1/3。所述碱基编辑器可靶向Y705和/或S727以用于减少Stat1/3激活。可通过基于荧光素酶的启动子来测试所述碱基编辑。通过碱基编辑靶向Stat1/3可阻碍单核细胞进行巨噬细胞分化和响应巨噬细胞的ox-LDL刺激的炎症。

[0960] 在一些实施例中,所述碱基编辑器可用于修饰TFEB(EB的转录因子)。所述碱基编辑器可靶向调节所述TFEB的易位的一个或多个氨基酸残基。在一些情况下,所述碱基编辑器可靶向调节自体吞噬的一个或多个氨基酸残基。

[0961] 在一些实施例中,所述碱基编辑器可用于修饰鸟氨酸氨基甲酰转移酶(OTC)。这样的修饰可用于修正鸟氨酸氨基甲酰转移酶缺乏症。例如,碱基编辑可通过将核苷酸134C转化为U来修正Leu45Pro突变。示例性方法示于图102中。

[0962] 在一些实施例中,所述碱基编辑器可用于修饰Lipin1。所述碱基编辑器可靶向可被mTOR磷酸化的一个或多个丝氨酸。Lipin1的碱基编辑可调节脂质堆积。所述碱基编辑器可靶向3T3L1前脂肪细胞模型中的Lipin1。可通过测量脂质堆积的减少(例如通过油红)来测试所述碱基编辑的影响。

[0963] 碱基编辑引导分子设计的考量

[0964] 在一些实施方式中,所述引导序列是RNA序列,所述RNA序列的长度为10至50nt之间,但尤其为约20-30nt,有利地约20nt、23-25nt或24nt。在碱基编辑的实施方式中,选择所述引导序列以确保其与包含将要脱氨的腺苷的靶序列杂合。这更详细地描述于下文。选择可涵盖增加脱氨地效能和特异性地进一步步骤。

[0965] 在一些实施方式中,所述引导序列长度为约20nt至约30nt且其与所述靶DNA链杂合以形成几乎完美配对的双链核酸分子,除了具有所述靶腺苷位点处的dA-C错配之外。特别是,在一些实施方式中,所述dA-C错配位于所述靶序列的中心的近处(因此在所述引导序列与所述靶序列的杂交之后的所述双链核酸分子的中心),从而限制所述腺苷脱氨酶至狭

窄的编辑窗口（例如约4bp宽）。在一些实施方式中，所述靶序列可包含多于一个将要脱氨的靶腺苷。在进一步的实施方式中，所述靶序列可进一步包含3'至所述靶腺苷位点的一个或多个dA-C错配。在一些实施方式中，为了避免所述靶序列中的非计划的腺嘌呤位点处的脱靶编辑，所述引导序列可被设计为包含在对应于非计划的腺嘌呤的位置处的未配对的鸟嘌呤以引入dA-G错配，所述dA-G错配催化地不利于某些腺苷脱氨酶（例如ADAR1和ADAR2）。见Wong等, RNA 7:846-858 (2001), 其通过引用整体并入本文。

[0966] 在一些实施方式中，使用具有标准长度（例如约20nt）的CRISPR-Cas引导序列以形成具有所述靶DNA的异源双链核酸分子。在一些实施方式中，使用长于所述标准长度（例如>20nt）的CRISPR-Cas引导分子以形成具有所述靶DNA（包括在所述CRISPR-Cas-引导RNA-靶DNA复合物之外的）异源双链核酸分子。使感兴趣的核苷酸的指定延伸范围内的一个或多个腺嘌呤脱氨可为有利的。在替代的实施方式中，维持所述标准引导序列长度的限定可为有利的。在一些实施方式中，所述引导序列被设计为引入CRISPR-Cas引导物的所述标准长度之外的dA-C错配，其可通过CRISPR-Cas减少位阻并增加所述腺苷脱氨酶和所述dA-C错配之间的接触的频率。

[0967] 在一些碱基编辑实施方式中，从所述PAM在DNA靶标上的位置处开始计算所述错配的核碱基（例如胞苷）的位置。在一些实施方式中，所述错配的核碱基位于距所述PAM 2-21nt, 或距所述PAM 13-21nt, 或距所述PAM 14-21nt, 或距所述PAM 14-20nt, 或距所述PAM 15-20nt, 或距所述PAM 16-20nt, 或距所述PAM 14-19nt, 或距所述PAM 15-19nt, 或距所述PAM 16-19nt, 或距所述PAM 17-19nt, 或距所述PAM约20nt, 或距所述PAM 约19nt, 或距所述PAM约18nt, 或距所述PAM约17nt, 或距所述PAM 约16nt, 或距所述PAM约15nt, 或距所述PAM约14nt处。在优选实施方式中，所述错配的核碱基位于距所述PAM 17-19nt或18nt处。

[0968] 错配距离是所述CRISPR-Cas间隔基的3'末端和所述错配的核碱基（例如胞苷）之间的碱基数，其中所述错配的碱基作为所述错配距离计算的一部分被包含。在一些实施方式中，所述错配距离为1-10nt或1-9nt或1-8nt或2-8nt 或2-7nt或2-6nt或3-8nt或3-7nt或3-6nt或3-5nt或约2nt或约3nt或约4nt或约5nt或约6nt或约7nt或约8nt。在优选实施方式中，所述错配距离为3-5nt或4nt。

[0969] 在一些实施方式中，本文所述的CRISPR-Cas-ADAR系统的所述编辑窗口为距所述PAM 12-21nt, 或距所述PAM 13-21nt, 或距所述PAM 14-21nt, 或距所述PAM 14-20nt, 或距所述PAM 15-20nt, 或距所述PAM 16-20nt, 或距所述PAM 14-19nt, 或距所述PAM 15-19nt, 或距所述PAM 16-19nt, 或距所述PAM 17-19nt, 或距所述PAM约20nt, 或距所述PAM约19nt, 或距所述PAM约18nt, 或距所述PAM约17nt, 或距所述PAM约16nt, 或距所述PAM约15nt, 或距所述PAM约14nt。在一些实施方式中，本文所述的CRISPR-Cas-ADAR系统的所述编辑窗口为距所述CRISPR-Cas间隔基的3'末端1-10nt, 或距所述CRISPR-Cas间隔基的3'末端1-9nt, 或距所述CRISPR-Cas间隔基的3'末端1-8nt, 或距所述CRISPR-Cas间隔基的3'末端2-8nt, 或距所述CRISPR-Cas间隔基的3'末端2-7nt, 或距所述CRISPR-Cas间隔基的3'末端2-6nt, 或距所述CRISPR-Cas间隔基的3'末端3-8nt, 或距所述CRISPR-Cas间隔基的3'末端3-7nt, 或距所述CRISPR-Cas间隔基的3'末端3-6nt, 或距所述CRISPR-Cas间隔基的3'末端3-5nt, 或距所述CRISPR-Cas间隔基的3'末端约2nt, 或距所述CRISPR-Cas间隔基的3'末端约3nt, 或距所述CRISPR-Cas间隔基的3'末端约4nt, 或距所述CRISPR-Cas间隔基的3'末端约5nt, 或

距所述CRISPR-Cas间隔基的3'末端约6nt,或距所述CRISPR-Cas间隔基的3'末端约7nt,或距所述CRISPR-Cas间隔基的3'末端约8nt处。

#### [0970] 接头

[0971] 本文所述的脱氨酶可通过接头与Cas蛋白融合。进一步设想RNA腺苷甲基化酶(N(6)-甲基腺苷)可与靶向本发明的效应子蛋白的RNA融合,并可靶向感兴趣的转录物。该甲基化酶引起可逆的甲基化,具有调节作用并可通过调控多个细胞相关的细胞通路来影响基因表达和细胞命运决定(Fu等,Nat Rev Genet.2014;15(5):293-306)。

[0972] ADAR或其他RNA修饰酶可通过接头与CRISPR-Cas或死的CRISPR-Cas蛋白(例如与CRISPR-Cas或死的CRISPR-Cas的C末端或N末端)连接(例如融合)。

[0973] 本文中融合蛋白联合使用的术语“接头”指与所述蛋白结合以形成融合蛋白的分子。一般而言,这样的分子除了与所述蛋白结合或保留所述蛋白之间的一些最小距离或其他空间距离之外没有特定生物活性。但是,在某些实施方式中,可选择所述接头以影响所述接头和/或所述融合蛋白的一些特性,例如所述接头的折叠、净电荷或疏水性。

[0974] 用于本发明的所述方法中的合适接头是本领域技术人员众所周知的且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,如本文所用的接头也可为共价键(碳-碳键或碳-杂原子键)。在特定实施方式中,所述接头用于将所述CRISPR-Cas蛋白和所述核苷酸脱氨酶分离足够确保每个蛋白保留其需要的功能特性的距离。优选的肽接头序列采用灵活的延伸构象且没有显示出发展有序二级结构的倾向。在某些实施方式中,所述接头可为化学部分,所述化学部分可为单体、二聚体、多聚体或聚合体。优选地,所述接头包含氨基酸。在灵活的接头中的典型的氨基酸包括Gly、Asn和Ser。因此,在特定实施方式中,所述接头包含Gly、Asn和Ser氨基酸中一个或多个的组合。其他几乎中性的氨基酸(例如Thr和Ala)也可用于所述接头序列中。示例性接头公开于Maratea等(1985),Gene 40:39-46;Murphy等(1986)Proc.Nat'l. Acad.Sci.USA 83:8258-62;U.S.Pat.No.4,935,233;和U.S.Pat.No.4,751,180中。例如,可使用GlySer接头GGG、GGGS或GSG。可以三次重复来使用GGG、GSG、GGGS或GGGG接头(例如(GGG)<sub>3</sub>(SEQ ID No. 249),(GGGS)<sub>3</sub>(SEQ ID NO:79)或5、6、7、9或甚至12(SEQ ID NO:250-254)或更多以提供合适的长度。在一些情况下,所述接头可为(GGGGS)<sub>3-15</sub>,例如,在一些情况下,所述接头可为(GGGGS)<sub>3-11</sub>,例如GGGGS(SEQ ID NO:255),(GGGGS)<sub>2</sub>(SEQ ID NO:256),(GGGGS)<sub>3</sub>(SEQ ID NO:79),(GGGGS)<sub>4</sub>(SEQ ID NO:257),(GGGGS)<sub>5</sub>,(GGGGS)<sub>6</sub>(SEQ ID NO:251),(GGGGS)<sub>7</sub>(SEQ ID NO:252),(GGGGS)<sub>8</sub>(SEQ ID NO:258),(GGGGS)<sub>9</sub>(SEQ ID NO:253),(GGGGS)<sub>10</sub>(SEQ ID NO:259)或(GGGGS)<sub>11</sub>(SEQ ID NO:260)。

[0975] 在特定实施方式中,接头(例如(GGGGS)<sub>3</sub>)优选地用于本文中。(GGGGS)<sub>6</sub>、(GGGGS)<sub>9</sub>或(GGGGS)<sub>12</sub>可优选地用作替代。其他优选的替代为(GGGGS)<sub>1</sub>(SEQ ID No 255),(GGGGS)<sub>2</sub>(SEQ ID No. 256),(GGGGS)<sub>4</sub>,(GGGGS)<sub>5</sub>,(GGGGS)<sub>7</sub>,(GGGGS)<sub>8</sub>,(GGGGS)<sub>10</sub>或(GGGGS)<sub>11</sub>。在又一进一步的实施方式中,LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No:261)用作接头。在又一附加实施方式中,所述接头是XTEN接头。在特定实施方式中,所述CRISPR-Cas蛋白是CRISPR-Cas蛋白且通过LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No:261)接头的方法与所述脱氨酶蛋白或其催化结构域连接。在进一步的特定实施方式中,所述CRISPR-Cas蛋白通过LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR(SEQ ID No:261)接头的方法以C末端与脱

氨酶蛋白的N末端或其催化结构域连接。此外，N-和C-末端的NLS也可以接头（例如PKKKRKVEASSPKKRKVEAS (SEQ ID No.262)）运作。

[0976] 接头的实例示于下表8中。

[0977] 表8

|        |           |   |
|--------|-----------|---|
|        | GGG       | GGTGGTAGT(SEQ ID NO:263)  |
|        | GGGx3(9)  | GGTGGTAGTGGAGGGAGCGGCGGTTCA(SEQ ID NO:264)  |
| [0978] | GGGx7(21) | ggtggaggaggctctggtggaggcggtagcggaggcggagggtcgGGTGGTAGT<br>GGAGGGAGCGGCGGTTCA(SEQ ID NO:265)   |
|        | XTEN      | TCGGGATCTGAGACGCCTGGGACCTCGGAATCGGCTACG<br>CCCGAAAGT(SEQ ID NO:266)   |
|        | Z-EGFR_S  | Gtggataacaaattaacaaagaaatgtggggcggcgtgggaagaaattcgtaacctgccga   |
| [0979] | hort      | acctgaacggctggcagatgaccgcggtttattgcgagcctggtggatgatccgagccagag<br>cgcgaaacctgctggcggaagcgaaaaactgaacgatgagcaggcgccgaaaaccggcg<br>gtggttctggt(SEQ ID NO:267) |
|        | GSAT      | Ggtggttctgccggtggctccggttctggctccagcgggtggcagctctggtgctccggca<br>cgggtactgcgggtggcactggcagcgggtccggtactggctctggc(SEQ ID<br>NO:268)                          |

[0980] 核苷酸脱氨酶或其他RNA修饰酶可通过一个或多个氨基酸与 CRISPR-Cas或死的CRISPR-Cas连接。在一些情况下，所述核苷酸脱氨酶可通过氨基酸411-429、114-124、197-241和607-624中的一个或多个与所述 CRISPR-Cas或死的CRISPR-Cas连接。所述氨基酸位置可对应于本文公开的 CRISPR-Cas直系同源物。在某些实施例中，所述核苷酸脱氨酶可通过对应于颊普雷沃菌CRISPR-Cas的氨基411-429、114-124、197-241和607-624的一个或多个氨基酸与所述死的CRISPR-Cas连接。

[0981] 一般使用方法

[0982] 在另一方面，本公开公开了使用本文所述的组合物和系统的方法。一般而言，所述方法包括通过引入细胞或生物来修饰靶核酸，所述生物包含所述靶核酸，所述工程化CRISPR-Cas蛋白，编码工程化CRISPR-Cas蛋白的多核苷酸，所述CRISPR-Cas系统，或包含所述多核苷酸的所述载体或载体系统，从而使所述工程化CRISPR-Cas蛋白修饰所述细胞或生物中的所述靶核酸。可通过由脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组装、基因枪、可植入的装置或本文所述的载体系统进行的递送来引入所述工程化 CRISPR-Cas蛋白或系统。所述细胞或生物可为真核细胞或生物。所述细胞或生物是动物细胞或生物。所述细胞或生物是植物细胞或生物。核酸纳米组装的实例包括DNA折纸和RNA折纸，例如US8554489、US20160103951、W02017189914和W02017189870（其通过引用整体并入本文）中描述的那些。基因枪可包括基因枪递送系统，所述基因枪递送系统是用于将外源DNA（转基因）递送至细胞的装置。效应分子可为涂敷有DNA（通常为质粒DNA）的重金属的基本粒子。CRISPR-Cas系统中的递送组分的实例描述于Svitashev等，Nat Commun.2016;7:13274中。

[0983] 在一些实施方式中，所述靶核酸包含基因座，且所述工程化CRISPR-Cas 蛋白修饰在所述基因座处编码的基因产物或所述基因产物的表达。所述靶核酸是DNA或RNA且其中所述靶核酸中的一个或多个核苷酸可被碱基编辑。所述靶核酸可为DNA或RNA且其中所述靶核酸被切割。所述工程化 CRISPR-Cas蛋白可进一步切割非靶标核酸。

[0984] 在一些实施方式中,所述方法可进一步包括可视化活性和任选地,使用可检测的标记。所述方法还可包括检测所述CRISPR-Cas系统的一个或多个组分与所述靶核酸的结合。

[0985] 在其他方面,所述使用方法包括检测样品中的靶核酸。在一些实施方式中,所述方法包括使样品与以下物质接触:本文的工程化CRISPR-Cas蛋白;至少一个引导多核苷酸,所述引导多核苷酸包含引导序列,所述引导序列能够与所述靶核酸结合并被设计为与所述工程化CRISPR-Cas形成复合物;和包含非靶标序列的基于RNA的掩蔽构建体;其中所述工程化CRISPR-Cas蛋白显示附带的核糖核酸酶活性且切割所述检测构建体的所述非靶标序列;和检测来自所述非靶标序列的切割的信号,从而检测所述样品中的所述靶核酸。所述方法可进一步包括使所述样品与用于扩增所述靶核酸的试剂接触。用于扩增的所述试剂可包含恒温扩增反应试剂。所述恒温扩增试剂可包含基于核酸序列的扩增、重组酶聚合酶扩增、环介导的恒温扩增、链置换扩增、解旋酶依赖性扩增或切口酶扩增试剂。所述靶核酸为DNA分子且所述方法可进一步包括使所述靶DNA分子与包含RNA聚合酶位点和RNA聚合酶的引物接触。

[0986] 所述掩蔽构建体:抑制可检测的阳性信号的生成直至所述掩蔽构建体被切割或失活为止,或掩盖可检测的阳性信号或产生可检测的阴性信号直至所述掩蔽构建体被切割或失活为止。所述掩蔽构建体可包含:抑制由报道构建体编码的基因产物的生成的沉默RNA,其中所述基因产物在表达时产生所述可检测的阳性信号;产生所述阴性可检测信号的核酶,且其中所述阳性可检测信号在所述核酶失活时生成;或将底物转化为第一颜色的核酶,且其中所述底物在所述核酶失活时转化为第二颜色;适配体和/或包含多核苷酸栓系的抑制剂;与可检测的配体和掩蔽组分连接的多核苷酸;通过桥接分子保持于聚集体中的纳米颗粒,其中所述桥接分子中的至少一部分包含多核苷酸,且其中所述溶液在所述纳米颗粒分散于溶液中时经历色移;通过连接分子与一个或多个猝灭分子连接的量子点或荧光团,其中所述连接分子中的至少一部分包含多核苷酸;具有嵌合剂的复合物中的多核苷酸,其中所述嵌合剂在所述多核苷酸的切割之后改变了吸光度;或通过多核苷酸栓系的两个荧光团,所述荧光团在从所述多核苷酸释放时发生荧光变化。

[0987] 所述适配体可包含契合酶的多核苷酸栓系的抑制剂,其中所述酶在通过作用于底物上而从所述适配体或多核苷酸-栓系的抑制剂释放时产生可检测的信号;或可为抑制酶且防止所述酶催化来自底物的可检测的信号的生成的抑制性适配体,或其中所述多核苷酸栓系的抑制剂抑制酶且防止所述酶催化来自底物的可检测的信号的生成;或契合试剂对,所述试剂对在从所述适配体释放时结合以生成可检测的信号。

[0988] 所述纳米颗粒可为胶态金属。所述胶态金属材料可包含水溶性金属粒子或分散于液体、水溶胶或金属固体中的金属化合物。所述胶态金属可选自元素周期表的组IA、IB、IIB和IIIB中的金属以及过渡金属,特别是组VIII的那些。优选的金属包括金、银、铝、钪、钇、铁、镍和钙。其他合适的金属还包括以下的各种氧化态:锂、钠、镁、钾、铷、铯、钒、铬、钴、铜、镓、锗、铟、铊、锡、钨、铼、铂和钯。所述金属优选地以粒子形式提供且源自合适地金属化合物,例如 $Al^{3+}$ 、 $Ru^{3+}$ 、 $Zn^{2+}$ 、 $Fe^{3+}$ 、 $Ni^{2+}$ 和 $Ca^{2+}$ 离子。

[0989] 当RNA桥被激活的CRISPR效应子切割时,观察到前述的色移。在一些示例性实施方式中,所述颗粒是胶态金属。在某些其他示例实施方式中,所述胶态金属是胶态金。在某些

示例实施方式中,所述胶态纳米颗粒是15nm 金纳米颗粒(AuNPs)。由于胶态金纳米颗粒地独特表面特性,当完全分散在溶液中时,在520nm处观察到最大吸光度,并且肉眼呈现红色。AuNP聚集后,它们会显示出最大吸收率的红移,并显示为较深的颜色,最终从溶液中沉淀出来,成为深紫色的聚集体。

[0990] 在一些实施方式中,至少一个引导多核苷酸包含错配。所述错配可在所述一个或多个引导序列上的单个核苷酸变异的上或下游。在某些实施方式中,可通过错配的引入来实现切割效率的调整,例如1个或多个错配,例如间隔基序列和靶序列之间的1个或2个错配,包括所述间隔基/靶标的错配位置。例如双错配的位置越中心(即不是not 3' 或5'),则切割效率受到的影响就越大。因此,通过选择所述间隔基上的错配位置,可以调整切割效率。举例来说,如果需要小于100%的靶标切割(例如细胞群体中),1个或更多个(例如优选地2个)间隔基和靶序列之间的错配可引入所述间隔基序列。沿着所述错配位置的所述间隔基越中心,则切割百分比越低。在某些示例实施方式中,可利用所述切割效率来设计可区分两个或更多个相差单个核苷酸(例如单个核苷酸多态性(SNP)、变异或(点)突变)的靶标的单个引导物。所述CRISPR效应子可降低对SNP(或其他单个核苷酸变异)的敏感性并继续以某种效率水平切割SNP靶标。因此,对于两个靶标或靶标组而言,引导RNA可被设计为具有与所述靶标的一个互补的核苷酸序列,即靶上SNP。所述引导RNA 进一步被设计为具有合成的错配。如本文所用的,“合成的错配”指引入至所述自然存在的SNP的上游或下游的非自然存在的错配,例如至多5个核苷酸上游或下游,例如4、3、2或1个核苷酸上游或下游,优选地至多3个核苷酸上游或下游,更优选地至多2个核苷酸上游或下游,最优选地1个核苷酸上游或下游(即与所述SNP相邻)。当所述CRISPR效应子与所述靶上SNP 结合时,只有单个错配会与合成的错配形成复合物且所述CRISPR效应子将继续被激活且可检测的信号将继续生成。当所述引导RNA与脱靶SNP杂合时,两个错配将形成,所述错配来自所述SNP和所述合成的错配,且没有可检测的信号生成。因此,本文公开的系统可被设计为区分群体内的SNP。例如,所述系统可用于区分因单个SNP而不同的致病菌株或检测某些疾病特异性SNP,例如但不限于疾病相关的SNP,例如但不限于癌症相关的SNP。

[0991] 在某些实施方式中,所述引导RNA被设计为以使所述SNP位于所述间隔基序列(从5' 末端开始)的位置1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29 或30上。在某些实施方式中,所述引导RNA被设计为以使所述SNP位于所述间隔基序列(从5' 末端开始)的所述间隔基序列(从5' 末端开始)的位置1、2、3、4、5、6、7、8或9上。在某些实施方式中,所述引导RNA被设计为以使所述SNP位于所述间隔基序列(从5' 末端开始)的位置2、3、4、5、6 或7上。在某些实施方式中,所述引导RNA被设计为以使所述SNP位于所述间隔基序列(从5' 末端开始)的位置3、4、5或6上。在某些实施方式中,所述引导RNA被设计为以使所述SNP位于所述间隔基序列(从5' 末端开始) 的位置3上。

[0992] 在某些实施方式中,所述引导RNA被设计为以使所述错配(例如所述合成的错配,即除SNP之外的附加突变)位于所述间隔基序列(从5' 末端开始) 的位置1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30上。在某些实施方式中,所述引导RNA被设计为以使所述错配位于所述间隔基序列(从5' 末端开始)的位置1、2、3、4、5、6、7、8或9上。在某些实施方式中,所述引导RNA被设计为以使所述错配位于所述间隔基序列(从5' 末端开始)的位置4、5、6或7上。在某些实施方式中,所述引导RNA被设计

为以使所述错配位于所述间隔基序列(从5'末端开始)的位置5上。

[0993] 在某些实施方式中,所述引导RNA被设计为以使所述错配位于所述SNP的上游2个核苷酸处(即一个间隔核苷酸)。在某些实施方式中,所述引导RNA被设计为以使所述错配位于所述SNP下游的2个核苷酸处(即一个间隔核苷酸)。在某些实施方式中,所述引导RNA被设计为以使所述错配位于所述间隔基序列(从5'末端开始)的位置5上且所述SNP位于所述间隔基序列(从5'末端开始)的位置3上。

[0994] 转录物追踪

[0995] 在另一方面,本公开提供了用于转录物追踪的组合物和方法。在一些实施方式中,转录物追踪允许研究者可视化细胞、组织、器官或动物中的转录物,提供关于RNA动力学和功能的重要的空间-时间信息。示例性方法示于图102中。

[0996] 在一些实施方式中,所述组合物可以是本文的具有一个或多个标记的CRISPR-Cas蛋白或包含如此标记的CRISPR-Cas蛋白的CRISPR-Cas系统。所述CRISPR-Cas蛋白或系统可与一个或多个转录物结合,从而可通过使用所述CRISPR-Cas蛋白上的所述标记来检测(例如可视化)所述转录物。

[0997] 在一些实施方式中,本公开包括用于表达具有一个或多个多核苷酸或多核苷酸标记的CRISPR-Cas蛋白的系统。所述系统可包含编码所述CRISPR-Cas蛋白和/或所述标记的多核苷酸。所述系统可进一步包含载体系统,所述载体系统包含这样的多核苷酸。例如,CRISPR-Cas蛋白可与荧光蛋白或其片段融合。荧光蛋白的实例包括GFP蛋白,EGFP,Azami-Green, Kaede,ZsGreen1和CopGFP;CFP蛋白,例如Cerulean,mCFP,AmCyan1,MiCy,和CyPet;BFP蛋白,例如EBFP;YFP蛋白,例如EYFP,YPet,Venus,ZsYellow和mCitrine;OFP蛋白,例如cOFP,mKO和mOrange;红色荧光蛋白,或RFP;来自任何其他物种的红色或远红荧光蛋白,例如异形珊瑚礁珊瑚(Heteractis reef coral)和海葵(Actinia)或Entacmaea海葵及其变异体。RFP包括例如Discosoma变异体,例如mRFP1,mCherry,tdTomato,mStrawberry,mTangerine,DsRed2和DsRed-T1,花水母亚目(Anthomedusa)J-Red和海葵(Anemonia)AsRed2。远红荧光蛋白包括例如Actinia AQ143,Entacmaea eqFP611,Discosoma变异体,例如mPlum和mRaspberry,和Heteractis HcRed1和t-HcRed。

[0998] 在一些情况下,所述系统可以诱导表达标记的CRISPR-Cas蛋白。例如,所述系统在可调节的启动子的控制下,在所述CRISPR-Cas蛋白和/或标记的编码下编码所述CRISPR-Cas蛋白和/或标记。这样的系统可以允许对所述标记的表达进行空间和/或时间控制,从而实现转录追踪的空间和/或时间控制。

[0999] 在某些情况下,所述CRISPR-Cas可以用可检测的标签标记。标记可以在细胞中进行。或者,可以首先进行标记,然后将标记的CRISPR-Cas蛋白递送到细胞、组织、器官或器官中。

[1000] 所述可检测的标记可以被检测到(例如通过成像,超声或MRI可视化)。这种可检测标签的实例包括可检测的寡核苷酸标签,其可以是但不限于包含独特核苷的寡核苷酸。在某些情况下,所述可检测的标记包含标记物质,可通过光谱法检测到,光化学,生物化学,免疫化学,电气,光学或化学手段。此类标签包括用于标记的链霉亲和素偶联物染色的生物素,磁珠(例如Dynabeads®),荧光染料(例如荧光素,得克萨斯红,若丹明,绿色荧光蛋白和类似物),放射性标记(例如<sup>3</sup>H,<sup>125</sup>I,<sup>35</sup>S,<sup>14</sup>C,或<sup>32</sup>P),酶(例如辣根过氧化物酶,碱性磷酸酶

和ELISA中常用的其他酶),和量热标记,例如胶态金或有色玻璃或塑料(例如聚苯乙烯,聚丙烯,乳胶等)珠。D可检测标签可以通过许多方法检测。例如,可以使用胶卷或闪烁计数器来检测放射性标记,可以使用光电探测器来检测发射的光来检测荧光标记。通常通过向所述酶提供底物并进行检测来检测酶标记,检测所述酶对底物的作用所产生的反应产物,并且通过简单地可视化有色标记来检测量热标记。可以使用的标记物质的实例包括本领域技术人员已知的标记物质,例如荧光染料,酶,辅酶,化学发光物质和放射性物质。特定实例包括放射性同位素(例如<sup>32</sup>P、<sup>14</sup>C、<sup>125</sup>I、<sup>3</sup>H 和<sup>131</sup>I),荧光素,若丹明,丹酰氯,伞形酮,荧光素酶,过氧化物酶,碱性磷酸酶,β-半乳糖苷酶,β-葡萄糖苷酶,辣根过氧化物酶,葡糖淀粉酶,溶菌酶,糖类氧化酶,微过氧化物酶,生物素和钆。在使用生物素作为标记物质的情况下,优选地,在添加生物素标记的抗体后,进一步添加与酶(例如过氧化物酶)结合的链霉亲和素。有利地,所述标记是荧光标记的。荧光标记的实例包括但不限于Atto dyes,4-乙酰氨基-4'-异硫氰基-2,2'二硫磺;吖啶和衍生物:吖啶,吖啶异硫氰酸酯;5-(2'-氨基)氨基萘-1-磺酸(EDANS);4-氨基-N-[3-乙基磺酰基]苯基]萘二甲酰亚胺-3,5-二磺酸盐;N-(4-苯胺基-1-萘基)马来酰亚胺;邻氨基苯甲酰胺;BODIPY;亮黄;香豆素和衍生物;香豆素,7-氨基-4-甲基香豆素(AMC,香豆素120),7-氨基-4-三氟甲基香豆素(Coumaran 151);花青染料;花青素;4',6-二氨基二氨基-2-苯基吖啶(DAPI);5'5"-二溴邻苯三酚-磺酞(溴邻苯三酚红);7-二乙氨基-3-(4'-异硫氰酸根合苯基)-4-甲基香豆素;二乙酸三亚乙基三胺;4,4'-二异硫氰酸根合二苯乙烯-2,2'-二磺酸;4,4'-二异硫氰酸根合苯乙烯-2,2'-二磺酸;5-[二甲基氨基]萘-1-磺酰氯(DNS,丹酰氯);4-二甲基氨基苯基偶氮苯基-4'-异硫氰酸酯(DABITC);曙红和衍生物;曙红,曙红异硫氰酸酯,赤藓醇和衍生物;赤藓醇B,赤藓红,异硫氰酸酯;乙啶;荧光素和衍生物;5-羧基荧光素(FAM),5-(4,6-二氯三嗪-2-基)氨基荧光素(DTAF),2',7'-二甲氧基-4'5'-二氯-6-羧基荧光素,荧光素,荧光素异硫氰酸酯,QFITC,(XRITC);荧光胺;IR144;IR1446;孔雀石绿异硫氰酸酯;4-甲基伞形酮原邻甲酚酞;硝基酪氨酸;副苯胺;苯酚红;B-藻红蛋白;邻苯二甲醛;苾和衍生物:苾,苾丁酸,琥珀酰亚胺基1-苾;丁酸量子点;反应性红4(Cibacron.TM亮红色3B-A),罗丹明和衍生物:6-羧基-X-罗丹明(ROX),6-羧基罗丹明(R6G),赖氨酸罗丹明B磺酰氯罗丹明(Rhod),罗丹明B,罗丹明123,罗丹明X异硫氰酸磺酸,磺胺若丹明B,磺胺若丹明101,磺酰若丹明101的磺酰氯衍生物(得克萨斯红);N,N,N',N'四甲基-6-羧基若丹明(TAMRA);四甲基若丹明;四甲基罗丹明异硫氰酸酯(TRITC);核黄素;玫瑰酸;钽螯合物衍生物;Cy3;Cy5;Cy5.5;Cy7;IRD 700;IRD 800;La Jolta蓝;酞菁;和菁。荧光标记可为荧光蛋白,例如蓝色荧光蛋白、蓝绿色荧光蛋白、绿色荧光蛋白、红色荧光蛋白、黄色荧光蛋白或任何光可转换蛋白。通过摄动分析,猝灭或供体和受体分子之间的电子传输,标记进一步可包括所述杂合复合物中的分子之间的能量转移。所述荧光标记可以是per或特瑞伦。或者,所述荧光标记可以是荧光条形码。有利地,所述标记可为光敏感的,其中所述标记是光激活的和/或光切割的一个或多个接头以释放所述。在另一种实施方式中,所述荧光标记可诱导自由基形成。是量子点。

[1001] 在一些实施方式中,本公开提供了用于递送所述标记的CRISPR-Cas蛋白或标记的CRISPR-Cas系统的系统。所述递送系统可包含任何递送载体,例如本文描述的那些,例如RNP、脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组装、基因枪、可植入的装置或本文所述的载体系统。

### [1002] 核酸靶向

[1003] 在某些实施方式中,本发明所述的CRISPR-Cas效应子蛋白是来自表1-4 或所述表中所示的这样的蛋白,或在所述蛋白中,或包含所述蛋白,或基本由所述蛋白组成,或由所述蛋白组成,或涉及所述蛋白,其中一个或多个氨基酸如本文他处所描述的那样发生突变。因此,在一些实施方式中,所述效应子蛋白可为RNA-结合蛋白,例如死的-Cas型效应子蛋白,其可任选地如本文所述的那样例如通过转录激活或抑制结构域,NLS或其他功能性结构域而官能化。在一些实施方式中,所述效应子蛋白可为切割RNA单链的RNA-结合蛋白。如果结合的RNA是ssRNA,则ssRNA被完全切割。在一些实施方式中,所述效应子蛋白可为RNA-结合蛋白,其切割一条RNA双链,例如,如果它包含两个RNAe结构域。如果结合的RNA是dsRNA,则dsRNA被完全切割。在一些实施方式中,所述效应子蛋白可为具有结合酶活性的RNA结合蛋白,即它与dsRNA结合,但仅切割RNA链中的一个。

[1004] CRISPR系统中的核糖核酸酶功能是已知的,例如,已经报道了针对某些III型CRISPR-Cas系统的mRNA靶向(Hale et al.,2014,Genes Dev,vol.28, 2432-2443;Hale et al.,2009,Cell,vol.139,945-956;Peng et al.,2015,Nucleic acids research,vol.43, 406-417) 和提供了显著优势。从而提供了通过存在的效应子蛋白靶向RNA的CRISPR-Cas系统、组合物或方法。

[1005] 所述靶RNA,即所述感兴趣的RNA,是本发明旨在靶向的RNA,本发明导致在所述靶RNA上的所述感兴趣的靶位点处的所述效应子蛋白的募集和结合。所述靶RNA可为任何合适的RNA形式。这在一些实施方式中可包含mRNA。在其他实施方式中,所述靶RNA可包含tRNA或rRNA。

### [1006] 自激活系统

[1007] 一旦编辑了细胞中所有RNA的拷贝,就不再需要在该细胞中继续进行 CRISPR-Cas效应子蛋白表达或活性。依靠所述CRISPR-Cas或crRNA的RNA 作为所述引导靶序列的自灭活系统可以通过阻止CRISPR-Cas或复合物形成的表达而关闭所述系统。

### [1008] 靶RNA的实例

[1009] 本文所述的组合物和系统可用于编辑各种类型的靶RNA。靶RNA的实例描述于下文。

### [1010] 干扰RNA (RNAi) 和微小RNA (miRNA)

[1011] 在其他实施方式中,所述靶RNA可包含干扰RNA,即参与RNA干扰通路的RNA,例如shRNA、siRNA等。在其他实施方式中,所述靶RNA可包含微小RNA (miRNA)。对干扰RNA或miRNA的控制可通过减少体内或体外干扰RNA或miRNA的寿命来帮助减少那些方法中所见的脱靶效应 (OTE)。

[1012] 如果选择性地表达所述效应子蛋白和合适的引导物(例如在适当的启动子(例如组织或细胞周期特异性的启动子和/或增强子)的控制下在空间或时间上)表达,则可能是用于保护细胞或系统(体内或体外)免受RNAi侵害。这对于不需要RNAi的邻近组织或细胞可能有用,或者用于比较未表达所述效应子蛋白和合适的引导区和的细胞或组织的目的(即在不受RNAi控制的地方或在没有RNAi的地方)所述效应子蛋白可用于控制或结合至包含或由RNA组成的分子,例如核酶,核糖体或核糖开关。在本发明的实施方式中,RNA引导可募集所述效应子蛋白与这些分子结合,使得所述效应子蛋白能够与它们结合。

### [1013] 核糖体RNA (rRNA)

[1014] 例如,阿扎利德抗生素(例如阿奇霉素)是众所周知的。他们靶向并破坏了50S核糖体亚基。可以将该效应子蛋白与合适的引导RNA一起靶向50S核糖体亚基,并招募并与50S核糖体亚基结合。因此,提供了与针对核糖体(尤其是50s核糖体亚基)靶标的合适的引导物协同作用的该效应子蛋白。该效应子蛋白与针对核糖体(尤其是50年代核糖体亚基)靶标的所述适当的引导物的配合使用可能包括抗生素的使用。特别是,抗生素的使用类似于阿扎利德抗生素(例如阿奇霉素)的作用。在一些实施方式中,原核生物的核糖体亚基,例如原核生物中的70S亚基,上述50S亚基,30S亚基以及16S和5S亚基都可以被靶向。在其他实施方式中,真核生物核糖体亚基,例如真核生物中的80S亚基,60S亚基,40S亚基以及28S,18S。可能以5.8S和5S亚单位为靶RNA。

[1015] 如本文所述,所述效应子蛋白可为RNA-结合蛋白。在一些实施方式中,所述效应子蛋白可为RNA切割单链的RNA结合蛋白。在任一种情况下,但是特别是在RNA的RNA结合蛋白切割单链的情况下,核糖体功能都可以被调节,特别是被减少或破坏。这可能适用于任何核糖体RNA和任何核糖体亚基,rRNA的序列是众所周知的。

[1016] 因此,设想通过使用本效应子蛋白与合适的引导至核糖体靶标的结合来控制核糖体活性。这可以通过切割或结合到所述核糖体来实现。特别地,设想减少核糖体活性。这可用于体内或体外测定核糖体功能,也可作为体内或体外控制基于核糖体活性的疗法的手段。此外,设想在体内或体外系统中控制(即减少)蛋白质合成,这种控制包括抗生素,研究和诊断用途。

### [1017] 核糖开关

[1018] 核糖开关(也称为辅酶)是与小分子结合的Messenger RNA分子的调控片段。这通常导致由mRNA编码的所述蛋白的产量发生变化。因此,因此设想通过使用本效应子蛋白与合适的引导结合到核糖开关靶标来控制核糖开关活性。这可以通过切割或绑定到核糖开关来实现。特别地,设想了核糖开关活性的降低。这可用于体内或体外测定核糖开关功能,也可用作体内或体外控制基于核糖开关活性的疗法的手段。此外,设想在体内或体外系统中控制(即减少)蛋白质合成。对于rRNA的这种控制可能包括抗生素,研究和诊断用途。

### [1019] 核酶

[1020] 核酶是具有催化性质的RNA分子,类似于酶(其当然是蛋白)。由于核酶是天然存在的和经过工程改造的,包含或由RNA组成,它们也可能被本RNA结合效应子蛋白靶向。在一些实施方式中,所述效应子蛋白可为RNA-结合蛋白切割所述核酶,从而使其失效。因此,设想通过使用本效应子蛋白与合适的引导结合到核酶靶标上来控制核酶活性。这可能是通过切割或结合到核酶来实现的。特别地,设想减少核酶活性。这可用于体内或体外测定核酶功能,也可用作体内或体外控制基于核酶活性的疗法的手段。

### [1021] RNA靶向应用

#### [1022] 基因表达,包括RNA加工

[1023] 所述效应子蛋白还可与合适的指南一起用于靶向基因表达,包括通过RNA加工的控制。对RNA加工的控制可能包括RNA加工反应,例如通过靶向RNAPol进行RNA剪接(包括其他剪接);病毒复制(尤其是卫星病毒,噬菌体和逆转录病毒,例如HBV,HBC和HIV和本文所列出的其他)包括植物中的病毒体;和tRNA生物合成。所述效应子蛋白和适当的引导也可用于

控制RNA激活(核糖核酸酶)。核糖核酸酶可促进基因表达,因此可通过破坏或减少核糖核酸来实现对基因表达的控制。酶和因此对基因表达的促进较少。

[1024] RNAi筛选

[1025] 鉴定其基因敲除与表型变化有关的基因产物,可以通过RNAi筛选来询问生物学通路并鉴定其组成部分。也可以通过使用所述效应子蛋白和适当的引导来去除或降低筛选中RNAi的活性,从而对这些筛选进行控制,从而恢复(先前干扰的)基因产物的活性或减少干扰/压制)。

[1026] 也可以治疗卫星RNA(satRNA)和卫星病毒。

[1027] 本文中关于核糖核酸酶活性的控制通常是指减少,负面破坏或敲下或敲出。

[1028] 体内RNA应用

[1029] 基因表达的抑制

[1030] 本文提供的靶标特异性RNA允许对靶标RNA进行非常特异性的切割。RNA水平的干扰可以无创地在空间和时间上进行调节,因为基因组没有被修饰。

[1031] 已经证明可以通过mRNA靶向治疗许多疾病。尽管这些研究大多数都与 siRNA的给药有关,但很显然,本文提供的RNA印迹效应子蛋白可以类似的方式应用。

[1032] mRNA靶标(以及相应的疾病治疗方法)包括VEGF,VEGF-R1和 RTP801(在AMD和/或DME治疗中),胱天蛋白酶2(在Naion治疗中), ADRB2(在眼压治疗中),TRPVI(在干眼症治疗中),Syk激酶(在哮喘治疗中),Apo B(在高胆固醇血症治疗中),PLK1,KSP和VEGF(在实体瘤治疗中),Ber-Abl(在CML治疗中)(Burnett和Rossi Chem Biol.2012,19(1):60-71)。相似地,RNA靶向已被证明可有效治疗RNA病毒介导的疾病,例如HIV(靶向HIV Tet和Rev),RSV(靶向RSV核衣壳)和HCV(靶向 miR-122)(Burnett和Rossi Chem Biol.2012,19(1):60-71)。

[1033] 可以设计特异性靶向转录的mRNA中包含突变或等位基因特异性序列的序列的引导RNA。这样的特异性敲低特别适用于与突变或等位基因特异性基因产物有关的疾病有关的治疗应用。例如,大多数家族性低脂蛋白血症(FHBL)是由所述APOB基因中的突变引起的。该基因编码载脂蛋白B蛋白的两个版本:一个短版本(ApoB-48)和一个较长版本(ApoB-100)。导致FHBL的几个ApoB基因突变导致两个版本的ApoB异常短。用本发明的RNA靶向效应子蛋白特异性靶向和敲除突变的ApoB mRNA转录物可能对FHBL的治疗有益。另一个实例,亨廷顿舞蹈病(HD)是由亨廷顿蛋白编码基因中的CAG三联体重复序列扩增引起的,从而导致蛋白质异常。用本发明的RNA靶向效应子蛋白特异性靶向和敲低编码Huntingtin蛋白的突变或等位基因特异性 mRNA转录物可能对HD治疗有益。

[1034] 通过调节RNA功能来调节基因表达

[1035] 除了通过切割mRNA对基因表达产生直接影响外,RNA靶向还可以用于影响细胞内RNA加工的特定方面,这可能允许对基因表达进行更细微的调节。一般而言,调节可以例如通过干扰蛋白与RNA的结合来介导,例如阻断蛋白的结合或募集RNA结合蛋白。确实,可以在不同水平确保调节,例如mRNA的剪接,转运,定位,翻译和翻转。类似地,在治疗方面,可以设想通过使用RNA特异性靶向分子来解决每个水平的(致病性)功能障碍。在这些实施方式中,在许多情况下,优选的是RNA剪切蛋白是“死的”CRISPR-Cas,其丧失了切割RNA靶标的能力但是保持了其与之结合的能力,例如本文所述的CRISPR-Cas的突变形式。

### [1036] 可变剪接

[1037] 由于所述剪接,许多所述人基因表达了多种mRNA。已经显示出不同的疾病与异常剪接有关,导致表达基因的功能丧失或功能获得。尽管其中一些疾病是由引起剪接缺陷的突变引起的,但许多并非如此。一种治疗选择是直接靶向剪接机制。本文描述的RNA效应子蛋白可以例如用于剪切或促进切片,包括或排除外显子和影响特定同工型的表达和/或刺激替代的蛋白产物的表达。这样的应用将在下面更详细地描述。

[1038] 靶RNA结合的RNA靶向效应子蛋白可以在空间上限制剪接因子进入 RNA序列。靶向剪接位点的RNA靶向效应子蛋白可以在该位点剪接,还可以将剪接重定向至相邻位点。例如RNA靶向效应子蛋白与5'剪接位点结合的结合可以阻碍剪接一些U1组分的募集,从而有利于该外显子的跳过。或者,靶向剪接增强子或沉默子的RNA靶向效应子蛋白可以防止靶位点处的交易性调节剪接因子结合并有效阻碍或促进剪接。外显子排斥可以进一步通过如本文所述的RNA靶向效应子蛋白将ILF2/3募集到外显子附近的前体 mRNA上来实现。作为另一个实例,可以附着碱性甘氨酸的结构域用于招募 hnRNP A1和外显子排阻(Del Gatto-Konczak et al. Mol Cell Biol. 1999 Jan; 19(1): 251-60)。

[1039] 在某些实施方式中,通过gRNA的适当选择,具体剪接变异体可有针对性的,而其他的剪接变异体将无法定位。

[1040] RNA靶向效应子蛋白可以与能够稳定剪接调节性茎环以进一步剪接的效应子结合。RNA收缩效应子蛋白可以与特定剪接因子的共有结合位点序列连接,以将所述蛋白募集到靶DNA上。

[1041] 与异常剪接相关的疾病的实例包括但不限于副肿瘤性视神经支配性肌阵挛共济失调(或POMA),这是由于失去了在突触中起作用的蛋白的调节剪接所引起的Nov蛋白的丧失,以及囊性纤维化,由囊性纤维化跨膜电导调节器的剪接不良引起的,导致产生无功能的氯离子通道。在其他疾病中,异常的RNA剪接会导致功能获得。例如在强直性肌营养不良中,这是由于 mRNA的3'UTR中CUG三联体重复扩增(从50到>1500次重复)引起的,从而导致剪接缺陷。

[1042] RNA靶向效应子蛋白可通过将剪接因子(例如U1)募集到所需外显子周围的5'剪切位点以促进切除内含子来用于包含外显子。可通过与富含精氨酸/丝氨酸的结构域的融合来介导这样的募集,所述富含精氨酸/丝氨酸的结构域作为剪接激活子发挥作用(Gravelly BR and Maniatis T, Mol Cell. 1998 (5): 765-71)。

[1043] 前导RNA效应因子蛋白可用于在所需基因座处限制剪接机制,从而防止外显子识别和另一种蛋白产物的表达。可以治疗的疾病的一个实例是 Duchenne肌营养不良症(DMD),它是由肌营养不良蛋白编码基因的突变引起的。几乎所有DMD突变都会导致移码,从而导致肌营养不良蛋白翻译受损。RNA标记效应子蛋白可以与剪接点或外显子剪接增强子(ESE)配对,从而阻止外显子识别,从而导致部分功能蛋白的翻译。这将使致命的Duchenne表型转化为不太严重的Becker表型。

### [1044] RNA修饰

[1045] RNA编辑是一种自然过程,通过轻微修饰RNA,可以增加指定序列的基因产物的多样性。通常,修饰涉及腺苷(A)到肌苷(I)的转化,从而产生不同于基因组编码的RNA序列。通常通过ADAR酶确保RNA修饰,从而使 pre-RNA靶标通过包含要编辑的腺苷酸的外显子与内

含子非编码元件之间的碱基配对,形成不完美的双链核酸分子RNA。A-I编辑的一个经典示例是谷氨酸受体GluR-B mRNA,该变化导致通道的电导特性发生了改变(Higuchi M, 等Cell.1993;75:1361-70)。

[1046] 在人中,ADAR1基因中的杂合功能无效突变导致皮肤疾病,人色素性皮肤病(Miyamura Y,等Am J Hum Genet.2003;73:693-9)。可以设想,本发明的RNA靶向效应子蛋白可以用于纠正功能异常的RNA修饰。

[1047] 聚腺苷酸化

[1048] mRNA的聚腺苷酸化对于核转运,mRNA的翻译效率和稳定性很重要,所有这些以及聚腺苷酸化的过程都取决于特定的RBP。大多数真核mRNA 在转录后会接受约200核苷酸的3' poly(A)尾巴。聚腺苷酸化涉及不同的RNA 结合蛋白复合物,它们刺激了聚(A)聚合酶的活性(Minvielle-Sebastia L等Curr Opin Cell Biol.1999;11:352-7)。可以设想,本文提供的靶向RNA的效应子蛋白可以用于干扰或促进RNA结合蛋白与RNA之间的相互作用。

[1049] 与参与多腺苷酸化的缺陷蛋白有关的疾病实例是眼咽肌营养不良(OPMD)(Brais B等,Nat Genet.1998;18:164-7)。

[1050] RNA输出

[1051] 在mRNA前的加工后,mRNA从细胞核输出到细胞质。这是通过涉及载体复合物生成的细胞机制来确保的,该复合物然后通过核孔转运并释放细胞质中的mRNA,并随后循环利用载体。

[1052] 已发现在RNA出口中起作用的蛋白(如TAP)的过度表达会增加在非洲爪蟾中无效出口的转录物的出口(Katahira J,等EMBO J.1999;18:2593-609)。

[1053] mRNA定位

[1054] mRNA定位可确保空间调节蛋白的产生。在特定的实施方式中,可以设想本文所述的所述效应子蛋白可用于将定位元件靶向于所述靶标的RNA。可以设计所述效应子蛋白以结合靶转录物并将它们穿梭到所述细胞中由其肽信号标签决定的位置。更特别地,例如,与核定位信号(NLS)融合的RNA 靶向效应子蛋白可用于改变RNA定位。

[1055] 定位信号的其他实例包括邮编结合蛋白(ZBP1),其确保 $\beta$ -肌动蛋白在几种不对称细胞类型中定位于细胞质,KDEL保留序列(定位于内质网),核输出信号(定位于细胞质),线粒体靶向信号(定位于线粒体),过氧化物酶体靶向信号(定位于过氧化物酶)和m6A标记/YTHDF2(定位于p抗体)。其他设想的方法是将RNA标记效应子蛋白与已知定位的蛋白融合(例如膜,突触)。

[1056] 或者,根据本发明的所述效应子蛋白可以例如用于定位敲低。通过将所述效应子蛋白融合至合适的定位信号,将效应子靶向特定的细胞腔室。仅存在于该区室中的靶RNA将被有效地靶向,而否则存在于不同细胞区室中的相同靶标将不被靶向,从而可以建立依赖于定位的敲低。

[1057] 翻译

[1058] 本文所述的RNA靶向效应子蛋白可用于增强或抑制翻译。可以设想,上调翻译是控制蜂窝电路的非常可靠的方法。此外,对于功能性研究,蛋白质翻译筛选可能优于转录上调筛选,后者的缺点是转录上调不会转化为增加的蛋白产量。

[1059] 设想本文所述的RNA靶向效应子蛋白可用于引入翻译起始因子,例如在感兴趣的

信使RNA的5'非翻译重复序列(5'UTR)附近产生EIF4G以驱动翻译。(如在De Gregorio等, EMBO J.1999;18(17):4865-74中描述的关于不可重编程的RNA结合蛋白)。作为另一实例,细胞质聚(A)聚合酶GLD2可通过RNA靶向效应子蛋白募集至靶mRNA。这将允许靶mRNA的定向聚腺苷酸化,从而刺激翻译。

[1060] 类似地,本文设想的RNA靶向效应子蛋白可用于抑制mRNA的翻译阻遏物,例如ZBP1(Huttelmaier S,等Nature.2005;438:512-5)。通过与靶RNA的翻译起始位点结合,可以直接影响翻译。

[1061] 此外,将RNA靶向效应子蛋白与稳定mRNA的蛋白融合(例如通过阻止其通过例如核糖核酸酶抑制剂降解),可能可以增加的蛋白从所述感兴趣的转录物生成。

[1062] 设想本文所述的RNA靶向效应子蛋白可通过结合在RNA转录物的5'UTR区域中并用于阻止所述核糖体形成和开始翻译而用于抑制翻译。

[1063] 进一步地,RNA干扰效应子蛋白可用于将CCR4-NOT死的烯基酶复合物的组分Caf1募集到靶RNAmRNA,从而导致死的烯基化或靶RNA转录物和蛋白翻译的抑制。

[1064] 例如,本发明的RNA靶向效应子蛋白可用于增加或减少治疗相关蛋白的翻译。可以使用RNA表达效应子蛋白下调或上调翻译的治疗应用实例是肌萎缩性侧索硬化症(ALS)和心血管疾病。据报道,ALS运动皮层和脊髓中神经胶质谷氨酸转运蛋白EAAT2的水平降低,以及ALS脑组织中多个异常的EAAT2 mRNA转录物也有报道。EAAT2蛋白和功能的丧失被认为是ALS兴奋性毒性的主要原因。EAAT2蛋白水平和功能的恢复可提供治疗益处。因此,如上所述,RNA阻断效应子蛋白可以有效地通过上调翻译抑制因子或稳定mRNA而上调EAAT2蛋白的表达。载脂蛋白A1是高密度脂蛋白(HDL)的主要蛋白组分)和ApoA1和HDL通常被认为具有抗动脉粥样硬化作用。可以设想,如上所述,RNA阻断效应子蛋白可通过阻断翻译阻遏物或稳定mRNA而有益地用于上调ApoA1的表达。

[1065] mRNA周转

[1066] 翻译与mRNA更新和调节的mRNA稳定性紧密相关。已经描述了特定蛋白与转录物和tristetraprolin(TTP)的稳定性有关(例如神经元中的ELAV/Hu蛋白,Keene JD,1999, Proc Natl Acad Sci U S A.96:5-7)。这些蛋白通过保护信息在细胞质中免受降解来稳定靶mRNA(Peng SS等,1988,EMBO J. 17:3461-70)。

[1067] 可以设想,本发明的靶向RNA的效应子蛋白可用于干扰或促进作用以稳定mRNA转录物的蛋白活性,从而影响mRNA更新。例如,使用RNA靶向效应子蛋白将人TTP募集至所述靶RNA将允许富含腺苷酸-尿苷酸的元素(富含AU的元素)介导的翻译抑制和靶降解。在许多mRNA的3'UTR中发现了富含AU的元素,这些元素编码原癌基因,核转录因子,细胞因子和促进RNA的稳定性。作为另一实例,该RNA靶向效应子蛋白可以与HuR融合,HuR是另一种mRNA稳定蛋白(Hinman MN和Lou H,Cell Mol Life Sci 2008;65:3168-81),并将其募集至靶RNA转录物以延长其寿命或稳定短命的mRNA。

[1068] 进一步设想本文所述的RNA靶向效应子蛋白可用于促进降解靶RNA转录物。例如,可以将m6A甲基转移酶募集至靶标转录物以将该转录物定位于导致靶标降解的P-体。

[1069] 作为又一个实例,如本文所述的RNA靶向效应子蛋白可以与非特异性核酸内切酶结构域Pil1TN末端(PIN)融合,以将其募集至靶转录物并允许其降解。

[1070] 患有副肿瘤性神经系统疾病(PND)-相关性脑脊髓炎和神经病的患者是在中枢神

神经系统以外的肿瘤中产生针对Hu蛋白的自身抗体的患者(Szabo A 等1991,细胞.;67:325-33),其然后穿过血脑屏障。可以设想,本发明的靶向 RNA的效应子蛋白可以用于干扰自身抗体与mRNA转录物的结合。

[1071] P由营养不良性肌强直蛋白激酶(DMPK)基因3'UTR中的(CUG)<sub>n</sub>扩展引起的1型营养不良型患者(DM1)的特征是这种转录物在细胞核中的积累。设想本发明的RNA靶向效应子蛋白与靶向(CUG)<sub>n</sub>重复序列的核酸内切酶融合可以抑制这种异常转录物的积累。

[1072] 与多功能蛋白的相互作用

[1073] 一些RNA结合蛋白与多个RNA上的多个位点结合,从而在多种过程中发挥作用。例如,已发现hnRNP A1蛋白结合外显子剪接沉默子序列,拮抗剪接因子,与端粒末端缔合(从而刺激端粒活性),并结合miRNA以促进 Drosha介导的加工,从而影响成熟。设想本发明的RNA-结合效应子蛋白可以在一个或多个位置处干扰RNA-结合蛋白的结合。

[1074] RNA折叠

[1075] RNA采用确定的结构以执行其生物学活性。交替三级结构中构象的转变对大多数RNA介导的过程至关重要。但是,RNA折叠可能与几个问题有关。例如,RNA可能会折叠成不正确的替代构象,并被其坚持,和/或正确的三级结构可能在热力学上没有比替代的结构更受青睐。发明所述的死的RNA 转录蛋白)可用于直接折叠(m)RNA和/或捕获其正确的三级结构。

[1076] RNA-靶向效应子蛋白在调控细胞状态中的用途

[1077] 在某种实现方式中,具有crRNA的复合物中的CRISPR-Cas与靶RNA 结合后被激活,随后切割附近的任何ssRNA靶(即“附带”或“旁观者”效应)。同源靶标,可以切割其他(非互补)RNA分子。这种混杂的RNA切割可能会引起细胞毒性,或其他影响细胞生理或细胞状态。

[1078] 因此,在本文实施方式中,非天然存在的或工程化的组合物,载体系统,或递送系统用于或用于诱导细胞休眠。如本文所述的-天然存在的或工程化组合物,载体系统,或递送系统用于或用于诱导细胞周期停滞。在某些实施方式中,非天然存在的或工程化组合物,载体系统,或本文所述的递送系统用于或用于减少细胞生长和/或细胞增殖,在某些实施方式中,非天然存在的或工程化组合物,如本文所述在某些实施方式中,本文所述的非天然或工程化组合物,载体系统,或递送系统用于或用于诱导细胞无能。在某些实施方式中,本文所述的非天然存在的或工程化组合物,载体系统,或递送系统用于或用于诱导细胞坏死。如本文所述的非天然存在的或工程化组合物,载体系统,或递送系统用于或用于在某些实施方式中,本文所述的非天然存在的或工程化的组合物,载体系统或递送系统用于或用于诱发程序性细胞死亡。

[1079] 在某些实施方式中,本发明涉及一种诱导细胞休眠的方法,包括引入或诱导本文所述的非天然存在的或工程化的组合物,载体系统或递送系统。本发明涉及一种诱导细胞周期停滞的方法,包括引入或诱导本文所述的非天然存在的或工程化的组合物,载体系统或递送系统。本发明所述的细胞生长和 /或细胞增殖的方法,包括引入或诱导非天然存在的或工程化组合物,载体系统或递送系统。在某些实施方式中,本发明涉及一种诱导细胞凋亡的方法,该方法包括引入或诱导非天然存在的或工程化的组合物。在某些实施方式中,本发明涉及一种诱导细胞坏死的方法,包括引入或诱导非天然存在。在某些实施方式中,本发明涉及一种诱导细胞死亡的方法,该方法包括引入或诱导非天然存在的或工程化组合物,

载体系统。在某些实施方式中,本发明涉及一种诱导程序性细胞死亡的方法,包括引入或诱导本文所述的非天然存在的或工程化的组合物,载体系统或递送系统。

[1080] 本文所述的方法和用途可以是治疗性或预防性和可复制的特定细胞,细胞(亚群)或细胞/组织类型。特别地,本文描述的方法和用途可以是治疗性或预防性和可切除的特定细胞,亚群,或表达一种或多种靶序列,例如一种或多种特定靶RNA(例如ss)的细胞/组织类型。不受限制地,靶细胞可以是表达特定转录物,例如指定类的神经元,引起例如自身免疫的免疫细胞或被特定(例如病毒)病原体感染的细胞等的癌症细胞。

[1081] 因此,在某些实施方式中,本发明涉及一种以存在不良细胞(宿主细胞)为特征的病理状态的方法,包括引入或诱导非天然存在的或工程化的组分,载体系统,或在某些实施方式中,本发明涉及使用本文所述的非天然或工程化组合物,载体系统或递送系统来治疗特征在于存在不良细胞的病理状况在某些实施方式中,本发明涉及本文所述的非天然存在的或工程化组合物,载体系统或递送系统,用于治疗特征在于存在不良细胞(宿主)的病理状况应当理解,优选地所述CRISPR-Cas系统靶向针对不良细胞的特异性靶。在某些实施方式中,本发明涉及使用非天然存在的或工程化的组合物,在某些实施方式中,本发明涉及本文所述的非天然存在的或工程化组合物,该载体系统,或递送系统用于治疗,预防或减轻癌症。用于治疗,预防或减轻癌症。在某些实施方式中,本发明涉及一种治疗,预防或减轻癌症的方法,包括引入或诱导本文所述的非天然存在的或工程化的组合物,载体系统或递送系统。应当理解,优选地所述CRISPR-Cas系统靶向癌症细胞特异的靶。在某些实施方式中,本发明与使用非天然存在的或工程化的组合物,载体系统有关。在某些实施方式中,本发明涉及本文所述的非天然存在的或工程化组合物,载体系统或递送系统。在某些实施方式中,本发明涉及一种用于治疗,预防或减轻病原体对细胞的感染的方法,包括引入或诱导非天然方法。如本文所述发生的或工程化的组分,载体系统,或交付系统。应当理解,优选地所述CRISPR-Cas系统的靶标是针对由病原体感染的细胞的特异性靶标(例如病原体衍生的靶标)。在某些实施方式中,本发明涉及非天然存在的或工程化组合物,载体系统,如本文所述用于治疗,预防或减轻自身免疫性疾病。在某些实施方式中,本发明涉及一种用于治疗,预防或减轻自身免疫性疾病的方法,包括引入或诱导非免疫性疾病。如本文所述的天然存在的或工程化组合物,载体系统,或递送系统。应当理解,优选地所述CRISPR-Cas系统靶向针对负责自身免疫疾病的细胞的特异性靶标(例如特异性免疫细胞)。

[1082] RNA-靶向效应子蛋白在RNA检测中的用途

[1083] 进一步设想了RNA印迹效应子蛋白可以用于Northern印迹测定中。Northern印迹涉及使用电泳按大小分离RNA样品。RNA标记效应子蛋白可用于特异性结合和检测所述靶RNA序列。

[1084] RNA靶向效应子蛋白可以与荧光蛋白(例如GFP)融合,并用于追踪活细胞中的RNA定位。更特别是,RNA靶向效应子蛋白可以不再被切割RNA而失活。在特定实施方式中,设想可以使用分裂的RNA靶向效应子蛋白,其信号依赖于结合为了确保更精确的可视化,对这两个亚蛋白进行了分析。或者,可以使用分裂的荧光蛋白,当多个RNA靶向效应子蛋白复合物结合至靶转录物时,该分裂的荧光蛋白被重建。进一步设想,转录物靶向mRNA上的多个结合位点,因此荧光信号可以放大真实信号,并可以用于病灶鉴定。作为另一种选择,可以将荧光蛋白重构组分裂内含蛋白。

[1085] RNA靶向效应子蛋白适用于确定RNA或特异性剪接变异体的定位, mRNA转录物的水平, 转录物的上调或下调和疾病特异性诊断。用于通过荧光显微镜或流式细胞术(例如荧光激活细胞分选(FACS))对(活)细胞中的RNA进行可视化, 从而可以高通量筛选细胞并在细胞分选后回收活细胞。此外, 可以在压力下同时评估不同转录物的表达水平, 例如使用分子抑制剂或低氧条件对细胞的癌症生长抑制作用。另一个应用将是使用两个光子显微镜在神经刺激过程中跟踪转录物到突触连接的定位。

[1086] 如本文所述, 本发明所述的根据本发明的所述组分或复合物可以用于原位杂合的多重稳健误差荧光中, (MER鱼类; Chen等Science; 2015; 348 (6233)), 例如用(荧光)标记的CRISPR-Cas效应子。

[1087] 体外APEX标记

[1088] 细胞过程取决于蛋白质, RNA和DNA之间的分子相互作用网络。准确检测蛋白-DNA和蛋白-RNA相互作用是了解此类过程的关键。体外接近标记技术采用了亲和标签和例如可光激活的探针相结合, 以在体外标记靶RNA 蛋白质或RNA附近的多肽和RNA。紫外线照射后, 可光激活基团与紧邻标记分子的蛋白质和其他分子发生反应, 从而对其进行标记。随后可以回收和鉴定标记的相互作用分子。用于将探针靶向所选的RNA序列。

[1089] 这些应用也可以在动物模型中用于疾病相关应用或难于培养的细胞类型的体内成像。

[1090] RNA-靶向效应子蛋白在RNA折纸/体外组装线中的用途-组合式 RNA折纸是指使用RNA作为集成模板创建二维或三维结构的纳米级折叠结构。折叠的结构被编码在RNA中, 因此合成的RNA序列决定了所得RNA 的形状(Geary等, 2014. Science, 345 (6198), pp. 799-804)。RNA折纸可以用作将其他组分例如蛋白排列到复合物中的支架。本发明的RNA标记效应子蛋白可以例如使用合适的引导RNA用于将感兴趣的蛋白靶向RNA折纸。

[1091] 这些应用也可以在动物模型中用于疾病相关应用或难于培养的细胞类型的体内成像。

[1092] RNA靶向效应子蛋白在RNA分离、纯化、富集或消耗中的用途

[1093] 进一步设想当与RNA复合时, RNA包裹效应子蛋白可以用于分离和/或纯化RNA。RNA印迹效应子蛋白可以例如与亲和标签融合, 该亲和标签可以用于分离和/或纯化。或纯化靶向效应子蛋白复合物的RNA-RNA。这样的应用对于分析细胞中的基因表达谱是有用的。

[1094] 在特定实施方式中, 可以设想RNA靶向效应子蛋白可用于靶向特定的非编码RNA(ncRNA), 从而阻断其活性, 从而提供有用的功能探针。如本文所述的蛋白可用于特异性富集特定RNA(包括但不限于增加稳定性等), 或可替代性地特异性耗尽特定RNA(例如但不限于特定剪接变异体、同种型等)。

[1095] LINC RNA功能和其他核RNA的询问

[1096] 当前的RNA敲低策略例如siRNA具有缺点, 因为所述蛋白质机制是胞质的, 所以它们主要限于靶向胞质的转录物。本发明的靶向效应子蛋白的 RNA的优点是对细胞功能不是必不可少的外源性系统, 它可以用于细胞中的任何区室。通过将NLS信号融合到RNA附加效应子蛋白上, 可以将其引导至细胞核, 从而使核RNA成为靶标。例如, 设想了探测lincRNA的功能。长的基因间非编码RNA(lincRNA)是一个尚未得到充分研究的领域。迄今为止, 大多数lincRNA具有未知功能, 可以使用本发明的RNA靶向效应子蛋白进行研究。

[1097] 蛋白RNA结合蛋白的鉴定

[1098] 鉴定与特定RNA结合的蛋白可能有助于理解许多RNA的作用。例如,许多lincRNA与转录和表观遗传调控因子结合,以控制转录。了解与指定的 lincRNA结合的蛋白可以帮助阐明组分的调控通路。可以设计本发明的RNA 靶向效应子蛋白以募集生物素连接酶至特定的转录物,以使用生物素标记局部结合的蛋白。然后将所述蛋白下拉并通过质谱分析以鉴定它们。

[1099] 复合物在RNA和底物上的组装

[1100] 发明的RNA靶向效应子蛋白可以进一步用于在RNA上组装复合物。这可以通过用多个相关蛋白(例如特定合成通路的组分)功能化RNA标记效应子蛋白来实现。或者,可以用这种不同的相关蛋白功能化多个靶向靶子蛋白的RNA,并靶向相同或相邻的靶标RNA。在RNA上组装复合物的有用应用是例如促进底物在蛋白质之间穿梭。

[1101] 合成生物学

[1102] 生物系统的开发具有广泛的用途,包括在临床应用中。设想可以使用本发明的可编程RNA靶向效应子蛋白融合以分裂毒性结构域的蛋白,以靶向细胞死亡,例如使用癌症连锁的RNA作为靶转录物。另外,在具有合成蛋白的生物系统中,涉及带有蛋白质-蛋白质相互作用的通路可以通过带有合适效应子(例如激酶或其他酶)的融合复合物来实现。

[1103] 蛋白剪接:内含肽

[1104] 蛋白剪切是翻译后的过程,在该过程中,称为内含肽的中间多肽催化其自身从被称为内蛋白的所说的裂开其的多肽中切下,并随后进行该内蛋白的连接。如本文所述在靶转录物上的两个或更多个RNA靶向效应子蛋白的组装可用于引导分裂内含肽的释放(Topilina和Mills Mob DNA.2014Feb 4;5(1):5),从而可以直接计算mRNA转录物的存在和随后释放的蛋白质产物,例如代谢酶或转录因子(用于下游激活转录通路)。该应用在合成生物学中可能具有重要意义(请参见上文)或大规模生物生产(仅在某些条件下生产产品)。

[1105] 可诱导的、适量的和自我激活的系统

[1106] 在一种实施方式中,将包含本发明的RNA靶向效应子蛋白和效应子组分的融合复合物设计为可诱导的,例如光诱导的或化学诱导的。这种诱导性允许在期望的时刻激活效应子组分。

[1107] 通过设计融合复合物实现了光诱导性,其中将CRY2PHR/CIBN配对用于融合。该系统对于光诱导活细胞中蛋白质相互作用特别有用(Konermann S 等,Nature.2013;500:472-476)。

[1108] 通过设计融合物,例如FKBP/FRB(FK506结合蛋白/FKBP雷帕霉素结合)配对来提供化学诱导性。使用该系统雷帕霉素对于结合蛋白是必需的(Zetsche等,Nat Biotechnol.2015;33(2):139-42本系统关于Cas9的用途)。

[1109] 进一步地,当以DNA形式引入所述细胞时,本发明的RNA表达效应子蛋白可以通过可诱导的启动子来调节,例如四环素或强力霉素控制的转录激活(Tet-On和Tet-Off表达系统),激素诱导基因表达系统例如蜕皮激素诱导的基因表达系统和阿拉伯糖诱导的基因表达系统。当以RNA的形式传递时,可以通过核糖开关调节RNA嵌入效应子蛋白的表达,它可以感应像四环素这样的小分子(如在Goldfless等,Nucleic Acids Res.2012;40(9):e64中描

述地)。

[1110] 在一个实施方式中,可以调节本发明的RNA表达效应子蛋白的所述递送以改变细胞中蛋白或crRNA的量,从而改变所需作用或任何不期望的脱靶效应的大小。

[1111] 在一个实施方式中,可以将本文所述的RNA靶向效应子蛋白设计为自灭活的。当作为RNA传递到细胞时,mRNA或作为复制RNA治疗剂(Wroblecka等,Nat Biotechnol.2015Aug;33(8):839-841),它们可以通过破坏自身的RNA来使表达及其后续作用失活,从而减少滞留和潜在的不良影响。

[1112] 对于本文所述的RNA靶向效应子蛋白的进一步体内应用,参考Mackay JP等(Nat Struct Mol Biol.2011Mar;18(3):256-61),Nelles等(Bioessays.2015 Jul;37(7):732-9)和Abil Z和Zhao H(Mol Biosyst.2015Oct;11(10):2658-65),其通过引用并入本文。尤其是,通过使用催化失活的CRISPR-Cas:增强翻译(例如CRISPR-Cas-翻译促进因子融合(例如eIF4融合));抑制翻译(例如gRNA靶向核糖体结合位点);外显子跳跃(例如gRNA靶向剪接供体和/或受体位点);外显子包涵(例如靶向特定外显子剪接供体的gRNA和/或受体位点至将其包括在内或将CRISPR-Cas融合至或招募剪接一些组分(例如U1 snRNA);访问RNA定位(例如CRISPR-Cas标记融合(例如EGFP融合));改变RNA定位(例如CRISPR-Cas定位信号融合(例如NLS或NES融合物));RNA降解(在这种情况下,如果依靠CRISPR-Cas的活性,则不使用催化失活的CRISPR-Cas,或者为了增加数量,可以使用分裂的CRISPR-Cas)使用);抑制非编码RNA功能(例如miRNA),例如通过gRNA与功能位点的降解或结合(可能通过CRISPR-Cas-信号序列融合重新定位在特定的位点滴定)。

[1113] 如本文之前和实施例中所述,CRISPR-Cas功能对于crRNA的5'或3'延伸和crRNA环的延伸具有鲁棒性。因此,设想可以将MS2环和其他募集结构域添加至crRNA,而不会影响复合物的形成和与靶转录物的结合。对crRNA的此类修饰以募集各种效应子结构域可适用于上述RNA靶向的效应子蛋白的用途。

[1114] CRISPR-Cas能够介导对RNA噬菌体的抗性。因此,设想CRISPR-Cas可用于针对仅RNA的病原体免疫例如动物,人和植物,包括但不限于埃博拉病毒和寨卡病毒。

[1115] 在某些实施方式中,CRISPR-Cas可以加工(切割)自己的阵列。这适用于野生型CRISPR-Cas蛋白和包含一个或多个本文所述的突变的氨基酸残基的突变的CRISPR-Cas蛋白。因此,可以设想为不同的靶转录物和/或应用设计的多个crRNA可以作为一个启动子驱动的作为pre-crRNA或as转录物的形式来递送。这种递送方法的优点在于,其在病毒系统中实质上更紧凑,更容易合成和易于递送。应当理解,对于本文的CRISPR-Cas的直系同源物,确切的氨基酸位置可以变化,如本领域已知的和如本文其他地方所述,可以通过蛋白比对来适当地确定。本发明的方面还包括本文所述的组合物和系统在基因组工程中的方法和用途,例如用于改变或操纵一个或多个基因或所述一个或多个基因种子在原核或真核细胞中的体外、体内或离体的表达。

[1116] 一方面,本发明提供了用于调节例如减少靶RNA在细胞中的表达的方法和组合物。在主题方法中,提供了本发明的CRISPR-Cas系统,其干扰RNA的转录,稳定性和/或翻译。

[1117] 在某些实施方式中,有效量的CRISPR-Cas系统用于切割RNA或以其他方式抑制RNA表达。在这方面,该系统具有与siRNA和shRNA相似的用途,因此也可以替代此类方法。该方法包括但不限于使用CRISPR-Cas系统替代例如干扰核糖核酸(例如siRNA或shRNA)或其转

录模板例如编码shRNA的 DNA。CRISPR-Cas系统例如通过被施用于包括靶细胞的哺乳动物而被引入靶细胞。

[1118] 有利地,本发明的CRISPR-Cas系统是特异性的。例如,尽管干扰核糖核酸(例如siRNA或shRNA)多核苷酸系统受到设计和稳定性问题以及脱靶结合的困扰,但是可以高特异性地设计本发明的CRISPR-Cas系统。

[1119] 在本发明的一个方面,本申请的也称为RNA或RNA靶向CRISPR系统的新型RNA靶向系统,其基于本文确定的CRISPR-Cas蛋白,且不需要生成定制蛋白即可靶向特异性RNA。RNA分子可对单个序列进行序列化,而单个酶则可识别特定的RNA靶标,换句话说,可使用所述RNA分子将该酶募集至特定的RNA靶标。

[1120] 在一些实施方式中,核酸靶向系统的一个或多个元件衍生自包含内源性 CRISPR RNA靶向系统的特定生物。在某些实施方式中,在Eubacterium和 Ruminococcus中发现了CRISPR RNA靶向系统。在某些实施方式中,效应子蛋白包含靶向的和附带的ssRNA切割活性。在某些实施方式中,效应子蛋白包含双重HEPN结构域。在某些实施方式中,效应子蛋白缺乏Cas13a 的Helical-1结构域的对应物。在某些实施方式中,效应子蛋白小于先前表征的2类CRISPR效应子,中值大小为928aa。该中值大小比Cas13c小190 aa (17%),比Cas13b小200aa (18%),比Cas13a小300aa (26%)。在某些实施方式中,效应子蛋白不需要侧翼序列(例如PFS、PAM)。

[1121] 在某些实施方式中,效应子蛋白基因座结构包括含有WYL结构域的辅助蛋白(因此在这些结构域的最初鉴定的组中保守的三个氨基酸之后表示;参见例如WYL结构域IPR026881)。在某些实施方式中,WYL结构域辅助蛋白包含至少一个螺旋-转-螺旋(HTH)或带-螺旋-螺旋(RHH) DNA结合结构域。在某些实施方式中,含有WYL结构域的辅助蛋白增加了靶向RNA的效应子蛋白的靶向ssRNA和附带的ssRNA切割活性。在某些实施方式中,含有WYL结构域的辅助蛋白包含N-末端RHH结构域,以及主要疏水的保守残基的模式,包括对应于原始WYL基序的不变的酪氨酸-亮氨酸双峰。在某些实施方式中,含有WYL结构域的辅助蛋白是WYL1。WYL1是主要与Ruminococcus 相关的单个WYL域蛋白。

[1122] 在其他示例实施方式中,靶向VI型RNA的Cas酶是Cas 13d。在某些实施方式中,Cas13d是西非杆菌(Eubacterium siraeum) DSM 15702 (EsCas13d) 或鲁米诺球菌(Ruminococcus sp.) N15.MGS-57 (RspCas13d) (参见例如Yan等, Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein, Molecular Cell (2018), doi.org/10.1016/j.molcel.2018.02.028)。RspCas13d和EsCas13d没有侧翼序列要求(例如PFS、PAM)。

[1123] CRISPR-CAS蛋白在优化功能RNA靶向系统中的应用

[1124] 在一方面,本发明提供了用于将功能组分特定递送至RNA环境的系统。可以使用本发明所述的包含RNA片段效应子蛋白的CRISPR系统来确保这一点,所述CRISPR系统允许将不同组分特异性靶向RNA。更特别地,此类组分包括RNA翻译,降解等的激活剂或阻遏物,例如激活剂或阻遏物。该系统的应用在本文其他地方描述。

[1125] 根据一个方面,本发明提供了一种非天然存在的或工程改造的组合物,其包含引导RNA,所述引导RNA包含能够与细胞中目的基因组位点中的靶序列杂交的引导序列,其中通过插入一个RNA来修饰所述引导RNA。或更多结合衔接蛋白的独特RNA序列。在特定实施方

式中,RNA序列可以结合两个或更多个衔接蛋白(例如适体),并且其中每个衔接蛋白与一个或多个功能域结合。已显示本文所述的CRISPR-Cas酶的引导RNA适合于引导序列的修饰。在特定实施方式中,通过在正向重复序列的3'内或正向重复序列的3'中插入正向重复序列的不同RNA序列5'来修饰引导RNA。当存在一个以上的功能域时,这些功能域可以相同或不同,例如,相同或两个不同的激活剂或阻遏物中的两个。一方面,本发明提供了本文讨论的组合物,其中所述一个或多个功能结构域连接至RNA靶向酶,从而在与靶RNA结合后,功能结构域处于空间取向,从而允许功能结构域在其中起作用。其归因功能;一方面,本发明提供了本文讨论的组合物,其中所述组合物包含具有至少三个功能域的CRISPR-Cas复合物,其中至少一个与RNA靶向酶缔合,并且其中至少两个与gRNA缔合。

[1126] 因此,一方面,本发明提供了非天然存在的或工程改造的CRISPR-Cas 复合物组合物,其包含本文讨论的引导RNA和作为RNA靶向酶的 CRISPR-Cas,其中任选地,RNA靶向酶包含至少一种突变。从而使RNA靶向酶具有不具有至少一种突变的酶的核酸酶活性的不超过5%,并且任选地具有至少一种或多种核定位序列的一种或多种。在特定的实施方式中,引导 RNA被额外地或替代地被修饰,以仍然确保RNA靶向酶的结合,但是防止被RNA靶向酶切割(如本文其他地方所详述)。

[1127] 在特定实施方式中,RNA靶向酶是CRISPR-Cas蛋白,与不具有至少一种突变的CRISPR-Cas酶相比,其核酸酶活性降低了至少97%或100%。在一个方面,本发明提供了本文讨论的组合物,其中CRISPR-Cas酶包含两个或多个如本文另外讨论的突变。

[1128] 在特定的实施方式中,如上所述提供了RNA靶向系统,其包含两个或更多个功能域。在特定实施方式中,两个或更多个功能域是异源功能域。在特定实施方式中,系统包含衔接蛋白,其是包含功能性结构域的融合蛋白,该融合蛋白任选地包含衔接蛋白和功能性结构域之间的接头。在特定实施例中,接头包括GlySer接头。另外地或可替代地,一个或多个功能域通过接头,任选地为GlySer接头,与RNA效应子蛋白连接。在特定的实施方式中,一个或多个功能结构域通过一个或两个HEPN结构域连接至RNA靶向酶。

[1129] 在一方面,本发明提供了本文讨论的组合物,其中与衔接蛋白或RNA 靶向酶相关的一个或多个功能域是能够激活或抑制RNA翻译的域。一方面,本发明提供了本文讨论的组合物,其中与衔接蛋白相关的一个或多个功能域中的至少一个具有一种或多种活性,包括甲基化酶活性,脱甲基酶活性,转录激活活性,转录抑制活性,转录释放。因子活性,组蛋白修饰活性,DNA 整合活性RNA切割活性,DNA切割活性或核酸结合活性,或分子开关活性或化学诱导性或光诱导性。

[1130] 在一方面,本发明提供了本文讨论的组合物,其包含适配体序列。在特定实施方式中,所述适配体序列是对相同接头蛋白具有特异性的两个或多个适配体序列。在一方面,本发明提供了本文所讨论的组合物,其中所述适配体序列是对不同接头蛋白具有特异性的两个或多个适配体序列。在一方面,本发明提供了本文所讨论的组合物,其中所述接头蛋白包含MS2、PP7、Q $\beta$ 、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、 $\phi$  Cb5、 $\phi$  Cb8r、 $\phi$  Cb12r、 $\phi$  Cb23r、7s、PRR1。因此,在特定实施方式中,所述适配体选自特异性结合上文所列的接头蛋白中任一个的结合蛋白。在一方面,本发明提供了本文所讨论的组合物,其中所述细胞是真核细胞。在一方面,本发明提供了本文所讨论的组合物,其中所述真核细胞是哺乳动物细胞、植物细胞或酵母细胞,其中所述哺乳

动物细胞任选地是小鼠细胞。在一方面,本发明提供了本文所讨论的组合物,其中所述哺乳动物细胞是人细胞。

[1131] 在一方面,本发明提供了本文上文所讨论的组合物,其中有多于一个的引导RNA或gRNA或crRNA,且这些靶向不同序列,其中当使用所述组合物时,有多重复用。在一方面,本发明提供了组合物,其中有多于一个的通过插入与一个或多个接头蛋白结合的不同RNA序列而修饰的引导RNA或gRNA或crRNA。

[1132] 在一方面,本发明提供了本文所讨论的组合物其中与一个或多个功能结构域相关联的一个或多个接头蛋白存在且与插入所述RNA中的所述不同RNA序列结合。

[1133] 在一方面,本发明提供了本文所讨论的组合物其中所述引导RNA被修饰为具有至少一个非编码的功能环;例如,其中所述至少一个非编码的功能环是抑制性的;例如,其中至少一个非编码的功能环包含Alu。

[1134] 在一方面,本发明提供了用于修饰基因表达的方法,其包括向宿主的施用或如本文所讨论的组合物中的一个或多个在宿主体内的表达。

[1135] 在一方面,本发明提供了本文讨论的方法,其包括所述组合物或编码其的核酸分子的递送,其中所述核酸分子与调节序列可操作地连接且在体内表达。在一方面,本发明提供了本文讨论的方法,其中通过慢病毒、腺病毒或AAV进行所述体内表达。

[1136] 在一方面,本发明提供了本文讨论的细胞的哺乳动物细胞系细胞,其中所述细胞系任选地是人细胞系或小鼠细胞系。在一方面,本发明提供了转基因哺乳动物模型,任选地大鼠,其中所述模型已用本文所讨论的组合物转化或是所述转化体的后代。

[1137] 在一方面,本发明提供了编码引导RNA或RNA靶向CRISPR-Cas复合物或如本文讨论的组合物的核酸分子。在一方面,本发明提供了载体,其包含:编码引导RNA(gRNA)或crRNA的核酸分子,其包含能够与细胞中的RNA靶序列杂合的引导序列,其中通过插入与两个或更多个接头蛋白结合的不同RNA序列来修饰所述gRNA或crRNA的正向重复,且其中每个接头蛋白与一个或多个功能结构域相关联;或者,其中所述gRNA被修饰为具有至少一个非编码的功能环。在一方面,本发明提供了包含编码以下物质的核酸分子的载体:非自然存在的或工程化CRISPR-Cas复合物组合物,所述CRISPR-Cas复合物组合物包含本文讨论的gRNA或crRNA,和RNA靶向酶,其中任选地,RNA靶向酶包含至少一个突变,从而RNA靶向酶具有少于5%的RNA靶向酶的核酸酶活性,所述RNA靶向酶不具有所述至少一个突变和任选地一包含至少一个或多个核定位序列。在一方面,载体可进一步包含可在真核细胞中运作的调节元素,所述调节元素与编码所述引导RNA(gRNA)或crRNA的核酸分子以及编码RNA靶向酶和/或所述任选的核定位序列的核酸分子可操作地连接。

[1138] 在一个方面,本发明提供了试剂盒,所述试剂盒包含本文所述的组分中的一个或多个。在一些实施方式中,所述试剂盒包含如本文所述的载体系统和关于使用所述试剂盒的说明书。

[1139] 在一个方面,本发明提供了一种筛选功能获得(GOF)或功能丧失(LOF)或筛选非编码RNA或潜在调控区(例如增强子,阻遏子)的方法,所述方法包含本文所述的细胞系。本文讨论的模型细胞或含有或表达RNA靶向酶的本文所述细胞或将本文讨论的组合物引入细胞系或模型细胞,其中gRNA或crRNA包含激活剂或阻遏物,并监测GOF或关于引入的gRNA或crRNA包括激活子的那些细胞的LOF,或者关于引入的gRNA或crRNA包括阻抑子的那些细胞

的LOF。

[1140] 在一个方面,本发明提供了非天然存在的或工程化的组合物的文库,每个组合物包含靶向CRISPR引导RNA (gRNA) 或crRNA的RNA,所述RNA 包含能够与细胞中目的靶RNA序列杂交的引导序列,RNA靶向酶,其中所述RNA靶向酶包含至少一种突变,使得所述RNA靶向酶具有不具有所述至少一种突变的RNA靶向酶的核酸酶活性的不超过5%,其中所述gRNA或crRNA被修饰通过插入与一种或多种衔接蛋白结合的不同RNA序列,并且其中衔接蛋白与一个或多个功能域结合,其中所述组合物包含一种或多种或两种或更多种衔接蛋白,其中每个蛋白质与一个或多个功能域结合,并且其中gRNA或crRNA包括全基因组文库,其包含多个靶向RNA的导向 RNA (gRNA) 或crRNA。一方面,本发明提供了本文所讨论的文库,其中与不具有至少一种突变的RNA靶向酶相比,靶向RNA的RNA靶向酶的核酸酶活性降低了至少97%或100%。一方面,本发明提供了本文所讨论的文库,其中衔接蛋白是包含功能性结构域的融合蛋白。在一个方面,本发明提供了如本文所讨论的文库,其中通过插入与一个或两个或多个衔接蛋白结合的不同RNA序列,不修饰gRNA或crRNA。在一方面,本发明提供了如本文所讨论的文库,其中一个或两个或更多个功能域与RNA靶向酶相关。一方面,本发明提供了如本文所讨论的文库,其中细胞的细胞群是真核细胞群。一方面,本发明提供了本文所讨论的文库,其中所述真核细胞是哺乳动物细胞,植物细胞或酵母细胞。在一方面,本发明提供了如本文所讨论的文库,其中哺乳动物细胞是人细胞。在一个方面,本发明提供了如本文所讨论的文库,其中细胞群是胚胎干 (ES) 细胞群。

[1141] 在一方面,本发明提供了如本文所讨论的文库,其中靶向是约100个或更多个RNA序列。在一方面,本发明提供了如本文所讨论的文库,其中靶向是约1000个或更多个RNA序列。在一个方面,本发明提供了如本文所讨论的文库,其中靶向是约20,000或更多个序列。在一方面,本发明提供了如本文所讨论的文库,其中靶向是整个转录组的。在一个方面,本发明提供了如本文所讨论的文库,其中所述靶向是针对相关或期望通路的一组靶序列。一方面,本发明提供了如本文所述的文库,其中所述通路是免疫通路。在一方面,本发明提供了如本文所讨论的文库,其中所述通路是细胞分裂通路。

[1142] 一方面,本发明提供了一种产生包含具有修饰的表达的基因的模型真核细胞的方法。在一些实施方式中,疾病基因是与患病或发展疾病的风险增加相关的任何基因。在一些实施方式中,该方法包括 (a) 将编码本文上文所述系统的组分的一种或多种载体引入真核细胞,和 (b) 使CRISPR复合物结合至靶多核苷酸,从而修饰a的表达。基因,从而产生包含修饰的基因表达的模型真核细胞。

[1143] 本文提供的结构信息允许询问引导RNA或crRNA与靶RNA和RNA靶向酶的相互作用,从而允许工程化或改变引导RNA结构以优化整个靶向 RNA的CRISPR-Cas系统的功能。例如,可以通过插入可以结合RNA的衔接蛋白来扩展引导RNA或crRNA,而不会与RNA靶向蛋白发生冲突。这些衔接蛋白可以进一步募集包含一个或多个功能域的效应子蛋白或融合蛋白。

[1144] 本发明的一个方面是上述元素包含在单一组合物中或包含在单独的组合物中。这些组合物可有利地应用于宿主以在基因组水平上引发功能作用。

[1145] 技术人员将理解,对引导RNA或crRNA的修饰允许结合衔接子+功能域,但不能正确定位衔接子+功能域(例如由于CRISPR-Cas复合物的三维结构内的空间位阻)是无意的修改。一个或多个修饰的引导RNA或crRNA可以通过在引导序列的正向重复或3'内引入正向重

复的5'的不同RNA序列来修饰。

[1146] 修饰的引导RNA或crRNA,灭活的RNA靶向酶(具有或不具有功能域) 和具有一个或多个功能域的结合蛋白可各自单独地包含在组合物中并单独或共同地施用于宿主。或者,这些组分可以以单一组合物的形式提供给宿主。可以通过本领域技术人员已知或本文描述的用于递送至宿主的病毒载体(例如慢病毒载体,腺病毒载体,AAV载体)进行对宿主的施用。如本文所述,使用不同的选择标记(例如用于慢病毒gRNA或crRNA选择)和gRNA或 crRNA的浓度(例如取决于是否使用多个gRNA或crRNA)可能有利于引起改善的效果。

[1147] 使用提供的组合物,本领域技术人员可以有利地和特异性地靶向具有相同或不同功能域的一个或多个基因座,以引发一个或多个基因组事件。所述组合物可以以多种方法用于在细胞中的文库中筛选和体内功能建模(例如 lincRNA的基因激活和功能鉴定;功能获得建模;功能丧失建模;用途)。本发明的组合物以建立用于优化和筛选目的的细胞系和转基因动物。

[1148] 本发明包括本发明的组合物用于建立和利用条件或诱导型CRISPR-Cas RNA靶向事件的用途。(参见例如Platt等,细胞(2014), dx.doi.org/10.1016/j.cell.2014.09.014,或本文引用的PCT专利公布,例如 WO 2014/093622 (PCT/US2013/074667),其不被认为是先于本发明或申请)。

[1149] 功能效应子的递送

[1150] RISPR-Cas13敲低允许通过使用人工转录因子来暂时减少基因表达,例如,通过突变Cas13蛋白质裂解结构域中的残基导致产生催化失活的Cas13 蛋白质。具有催化活性的Cas13与引导RNA或crRNA形成复合物,并位于由引导RNA或crRNA的靶向域指定的RNA序列上,但是,它不能切割靶标。失活的Cas13蛋白与效应子结构域,例如转录抑制结构域的融合使得能够将效应子募集到由引导RNA指定的任何位点。

[1151] 优化的功能RNA靶向系统

[1152] 因此,一方面,本发明提供了用于将功能性组分特异性递送至RNA环境的系统。可以使用包含本发明的RNA靶向效应子蛋白的CRISPR系统确保这一点,所述CRISPR靶向蛋白允许将不同组分特异性靶向RNA。更具体地,此类组分包括激活剂或阻遏物,例如RNA翻译,降解等的激活剂或阻遏物。

[1153] 根据一个方面,本发明提供了一种非天然存在的或工程改造的组合物,其包含引导RNA或crRNA,所述引导RNA或crRNA包含能够与细胞中的靶RNA靶序列杂交的引导序列,其中通过插入一个RNA来修饰所述引导 RNA或crRNA。或更多结合衔接蛋白的独特RNA序列。在特定实施方式中, RNA序列可以结合两个或更多个衔接蛋白(例如适体),并且其中每个衔接蛋白与一个或多个功能域结合。当存在一个以上的功能域时,这些功能域可以相同或不同,例如,相同或两个不同的激活剂或阻遏物中的两个。一方面,本发明提供了本文讨论的组合物,其中所述一个或多个功能结构域连接至 RNA靶向酶,从而在与靶RNA结合后,功能结构域处于空间取向,从而允许功能结构域在其中起作用。其归因功能;一方面,本发明提供了本文讨论的组合物,其中所述组合物包含具有至少三个功能域的CRISPR-Cas13复合物,其中至少一个与RNA靶向酶相关联,并且至少两个与gRNA或crRNA 相关联。gRNA或crRNA。

[1154] RNA靶向技术在植物和酵母中的应用

[1155] 定义:

[1156] 一般而言,术语“植物”涉及植物界的任何各种光合作用,真核,单细胞或多细胞生物,其特征在于通过细胞分裂而生长,含有叶绿体并且具有由纤维素组成的细胞壁。术语植物包括单子叶植物和双子叶植物。具体而言,所述植物旨在包括但不限于被子植物和裸子植物,例如阿拉伯树胶、苜蓿、苋菜、苹果、杏、朝鲜蓊、白蜡树、芦笋、鳄梨、香蕉、大麦、豆、甜菜、桦木、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、谷物、芹菜、栗子、樱桃、大白菜、柑橘、柑桔、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、花生、樱桃木、铁杉、山核桃、羽衣甘蓝、奇异果、大头菜、落叶松、生菜、韭菜、柠檬、石灰、蝗虫、松树、女仆、玉米、芒果、枫、瓜、小米、蘑菇、芥末、坚果、橡木、燕麦、油棕、秋葵、洋葱、橙子、观赏植物或花或树、木瓜、棕榈、欧芹、欧洲防风草、豌豆、桃子、花生、梨、泥煤、胡椒、柿子、木豆、松树、菠萝、车前草、李子、石榴、马铃薯、南瓜、拉迪基奥、萝卜、油菜籽、覆盆子、米饭、黑麦、高粱、红花、蜡黄、大豆、菠菜、云杉、南瓜、草莓、甜菜、甘蔗、向日葵、甘薯、甜玉米、橘子、茶、烟草、番茄、树木、小黑麦、草皮草、萝卜、藤蔓、核桃、豆瓣菜、西瓜、小麦、山药、紫杉和西葫芦。术语植物还包括藻类,藻类主要是光合自养生物,主要由于缺乏根,叶和其他代表高等植物的器官而统一。

[1157] 使用本文所述的使用RNA靶向系统调节基因表达的方法可用于在基本上任何植物上赋予期望的性状。使用本公开内容的核酸构建体和上述各种转化方法,可以针对本文所述的所需生理和农艺学特性对各种各样的植物和植物细胞系统进行工程改造。在优选的实施方式中,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和双子叶植物,例如包括谷类作物(例如小麦、玉米、水稻、小米、大麦)的作物,水果作物(例如番茄,苹果,梨,草莓,橙子,牧草作物(例如苜蓿),块根蔬菜作物(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)和用于实验目的的植物(例如拟南芥)。因此,所述方法和CRISPR-Cas系统可以在广泛的植物中使用,例如用于属于以下目的双子叶植物:木兰目、八角目、樟目、胡椒目、马兜铃目、睡莲目、毛茛目、罂粟目、瓶子草科、昆栏树目、金缕梅目、Eucomiales、塞子木目、杨梅目、山毛榉目、木麻黄目、石竹目、肉穗果目、蓼目、蓝雪目、五桠果目、山茶目、锦葵目、荨麻目、玉蕊目、紫堇目、杨柳目、白花菜目、杜鹃花目、岩梅目、柿树目、报春花目、蔷薇目、豆目、川苔草目、小二仙草目、桃金娘目、山茱萸目、山龙眼目、檀香目、大花草目、卫矛目、大戟目、鼠李目、无患子目、胡桃目、牻牛儿苗目、远志目、伞形目、龙胆目、花荵目、唇形目、车前目、玄参目、桔梗目、茜草目、川续断目和菊目。所述方法和CRISPR-Cas系统可以与单子叶植物一起使用,例如属于目泽泻目、水鳖目、茨藻目、霉草目、鸭跖草目、谷精草目、帚灯草目、禾本目、灯心草目、莎草目、香蒲目、凤梨目、姜目、槟榔目、环花目、露兜树目、天南星目、合目和兰目的植物;或可与属于裸子植物的植物一起使用,例如属于目松柏目、银杏目、苏铁目、南洋杉目、柏目和买麻藤目的那些。

[1158] 本文所述的靶向RNA的CRISPR系统和使用方法可在以下双子叶植物,单子叶植物或裸子植物属的非限制性列表中广泛使用的植物物种中:颠茄属、油丹属、漆树科、花生属、琼楠属、芸苔属、红花、球果、巴豆、黄瓜、柑桔、柑桔、辣椒、长春花、科科斯、咖啡、南瓜、胡萝

卜、杜鹃花属、花菱草属、榕属、草莓、海罌粟属、甘氨酸、棉酚、向日葵、橡胶树、轻海豚草、番茄属、羽扇豆、玛尼霍特、马约拉那、马卢斯、紫花苜蓿、烟草、烟叶、银胶菊属、罌粟、鳄梨属、菜豆、黄连木、豌豆属、梨属、李属、萝卜属、蓖麻属、千里光属、风龙属、千金藤属、白芥属、茄属、可可属、车轴草属、胡卢巴属、野豌豆、长春花、*Vilis*和豇豆属；和属葱属、须芒草属、*Aragrostis*、芦笋、燕麦属、狗牙根属、油棕属、羊茅属、羊茅黑麦草属、*Heterocallis*、大麦属、浮萍属、黑麦草、穆萨、稻属、黍属、狼尾草属、梯牧草属、早熟禾属、黑麦属、高粱、小麦属、玉蜀黍属、冷杉属、杉木属、麻黄属、云杉属、松属和黄杉属。

[1159] RNA靶向CRISPR系统和使用方法也可以在广泛的“藻类”或“藻类细胞”中使用；包括例如选自几种真核门的藻类，包括红藻门（红藻）、绿藻门（绿藻）、褐藻门（褐藻）、硅藻门（硅藻）、真眼点藻门和鞭毛藻门以及原核门蓝藻门（蓝绿藻）。术语“藻类”包括例如选自以下的藻类：双眉藻属、鱼腥藻属、纤维藻属、葡萄藻属、角毛藻属、衣藻属、小球藻属、绿球藻属、小环藻属、细柱藻属、杜氏藻属、球石藻属、裸藻属、红球藻属、等鞭金藻属、单鞭金藻属、单壳缝藻属、微球藻属、微拟球藻属、舟形藻属、肾鞭藻属、肾片藻属、菱形藻属、节球藻属、念珠藻属、*Oochromonas*、卵囊藻属、*Oscillartoria*、巴夫藻属、三角褐指藻属、扁藻属、颗石藻属、紫菜、假鱼腥藻属、塔胞藻属、裂丝藻属、聚球藻属、集胞藻属、四片藻属、海链藻属和束毛藻属。

[1160] 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。植物组织也包括植物细胞。如本文所用，术语“植物细胞”是指完整植物的完整单位或在体外组织培养中，培养基或琼脂上，悬浮于生长培养基或缓冲液或作为较高组织单位的一部分，例如植物组织，植物器官或整株植物。

[1161] “原生质体”是指通过使用例如机械或酶促方法已完全或部分去除了保护性细胞壁的植物细胞，从而形成了完整的生物体生化主管单元，可以重整其细胞壁，使其增殖并再生生长。在适当的生长条件下变成完整的植物。

[1162] 术语“转化(transformation)”广义上是指通过农杆菌或多种化学或物理方法之一通过引入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、器官或后代。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、叶片、幼苗、茎、愈伤组织、匍匐枝、微块茎和芽。植物组织还指这样的植物、种子、后代、繁殖体的任何克隆（论是有性或无性繁殖的）以及任何这些植物的后代，例如插条或种子。

[1163] 如本文所用，术语“转化的”是指已向其中引入外源DNA分子例如构建体的细胞、组织、器官或生物。可以将引入的DNA分子整合到受体细胞、组织、器官或生物体的基因组DNA中，从而使引入的DNA分子被传递至随后的后代。在这些实施方式中，“转化的”或“转基因的”细胞或植物还可以包括该细胞或植物的后代，以及从育种程序产生的后代，该育种程序采用这种转化的植物作为亲本在杂交中并表现出由其产生的改变的表型。引入的DNA分子的存在。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传递至后代。

[1164] 术语“后代”，例如转基因植物的后代，是由该植物或该转基因植物的出生，生于或衍生自该植物或该转基因植物的后代。引入的DNA分子也可以被瞬时引入受体细胞中，使得引入的DNA分子不会被随后的后代遗传，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。

[1165] 如本文所用,术语“植物启动子”是能够启动植物细胞中的转录的启动子,而不管其起源是否是植物细胞。示例性的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如农杆菌或根瘤菌获得的那些。

[1166] 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。真菌界内的毛舌藻包括子囊菌、担子菌、芽枝霉门、壶菌门、球囊菌门、微孢子虫和新丽鞭毛菌门。真菌细胞可包括酵母、霉菌和丝状真菌。在一些实施方式中,真菌细胞是酵母细胞。

[1167] 如本文所用,术语“酵母细胞”是指子囊菌门和子孢子菌门内的任何真菌细胞。酵母细胞可包括出芽的酵母细胞、裂变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的一部分。在一些实施方式中,酵母细胞是啤酒酵母、马克斯克鲁维酵母或东方伊萨氏菌细胞。其他酵母细胞可包括但不限于假丝酵母属(例如白色念珠菌)、耶氏酵母(例如解脂耶氏酵母)、毕赤酵母(例如巴斯德毕赤酵母)、克鲁维酵母属(例如乳酸克鲁维酵母和马克斯克鲁维酵母)、神经孢菌(例如克氏神经孢菌)、镰刀菌(例如尖孢镰刀菌(*Fusarium oxysporum*))和伊萨氏菌(例如东方伊萨酵母(又名库德里阿兹威毕赤酵母(*Pichia kudriavzevii*)) and *Candida acidothermophilum*)。在一些实施方式中,真菌细胞是丝状真菌细胞。如本文所用,术语“丝状真菌细胞”是指在丝状体中生长的任何类型的真菌细胞,即菌丝或菌丝体。丝状真菌细胞的实例可包括但不限于曲霉(例如黑曲霉)、木霉(例如里氏木霉)、根霉(例如米根霉(*Rhizopus oryzae*))和被孢霉(例如深黄被孢霉)。

[1168] 在一些实施方式中,真菌细胞是工业菌株。如本文所用,“工业菌株”是指在工业过程中使用或分离的任何真菌细胞菌株,例如,以商业或工业规模生产产品。工业菌株可以指通常在工业过程中使用的真菌种类,或者可以指也可以用于非工业目的(例如实验室研究)的真菌种类的分离物。工业过程的实例可以包括发酵(例如在食品或饮料产品的生产中),蒸馏,生物燃料的生产,化合物的生产和多肽的生产。工业菌株的实例可以包括但不限于JAY270 和ATCC4124。

[1169] 在一些实施方式中,真菌细胞是多倍体细胞。如本文所用,“多倍体”细胞可以指其基因组以一个以上拷贝存在的任何细胞。多倍体细胞可以是指天然存在于多倍体状态的细胞类型,也可以是指已被诱导以多倍体状态存在的细胞(例如通过特定的调控,改变,失活,激活或减数分裂,胞质分裂或DNA 复制的修饰)。多倍体细胞可以是指其整个基因组是多倍体的细胞,或者可以是指特定的特定基因组位点中是多倍体的细胞。不希望受理论的束缚,据认为,引导RNA的丰度在多倍体细胞的基因组工程中比在单倍体细胞中更经常是限速组分,因此使用本文所述的CRISPR-Cas CRISPR系统的方法可能会利用某些真菌细胞类型的优势。

[1170] 在一些实施方式中,真菌细胞是二倍体细胞。如本文所用,“二倍体”细胞可以指其基因组以两个拷贝存在的任何细胞。二倍体细胞可以指天然存在于二倍体状态的细胞类型,也可以指已经被诱导以二倍体状态存在的细胞(例如通过特定的调控,改变,失活,激活或减数分裂,胞质分裂或DNA复制的修饰)。例如,酿酒酵母菌株S228C可以维持在单倍体或二倍体状态。二倍体细胞可以是指其整个基因组是二倍体的细胞,或者它可以指的是在特定的特定基因组位点中是二倍体的细胞。在一些实施方式中,真菌细胞是单倍体细胞。如本文所用,“单倍体”细胞可以指其基因组以一个拷贝存在的任何细胞。单倍体细胞可以指天

然存在于单倍体状态的细胞类型,也可以指已经被诱导以单倍体状态存在的细胞(例如通过特定的调控,改变,失活,激活或减数分裂,胞质分裂或DNA复制的修饰)。例如,酿酒酵母菌株S228C可以维持在单倍体或二倍体状态。单倍体细胞可以是指其整个基因组是单倍体的细胞,或者它可以指的是在特定的特定基因组位点中是单倍体的细胞。

[1171] 如本文所用,“酵母表达载体”是指包含一个或多个编码RNA和/或多肽的序列并且可以进一步包含控制核酸表达的任何所需元件的核酸,以及能够在酵母细胞内部复制和维持表达载体的任何元素。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,多种载体和技术说明于Yeast Protocols,第二版本,Xiao,W.编辑(Humana Press,New York,2007),和 Buckholz,R.G.和Gleeson,M.A.(1991)Biotechnology (NY) 9(11):1067-72之中。酵母载体可包含但不限于着丝粒(CEN)序列,自主复制序列(ARS),可操作地连接至靶RNA序列或基因的启动子(如RNA聚合酶III启动子),终止子(如RNA聚合酶III终止子),复制起点和标记基因(例如营养缺陷型抗生素或其他选择性标记)。用于酵母的表达载体的实例可包括质粒、酵母人工染色体、2 $\mu$ 质粒、酵母整合质粒、酵母复制质粒、穿梭载体和游离质粒。

[1172] 植物和植物细胞基因组中靶向RNA的CRISPR系统组分的稳定整合

[1173] 在特定的实施方式中,设想引入了编码靶向CRISPR系统的RNA的组分的多核苷酸以用于稳定整合到植物细胞的基因组中。在这些实施方式中,可以根据何时,何处以及在什么条件下表达引导RNA和/或RNA靶向基因来调整转化载体或表达系统的设计。

[1174] 在特定的实施方式中,设想将靶向CRISPR系统的RNA的组分稳定地引入植物细胞的基因组DNA中。另外地或可替代地,设想引入靶向CRISPR系统的RNA的组分以稳定整合到植物细胞器的DNA中,所述植物细胞器例如但不限于质体,线粒体或叶绿体。

[1175] 用于稳定整合到植物细胞的基因组中的表达系统可以包含一种或多种以下元件:可用于在植物细胞中表达引导RNA和/或RNA靶向酶的启动子元件;一个5'非翻译区域,以增强表达;内含子元件,以进一步增强某些细胞如单子叶植物细胞中的表达;多克隆位点,为插入一个或多个引导RNA和/或靶向RNA的基因序列和其他所需元件提供方便的限制性位点;3'非翻译区,以有效终止表达的转录物。

[1176] 表达系统的元件可以在一个或多个表达构建体上,这些构建体可以是环状的,例如质粒或转化载体,或者是非环状的,例如线性双链DNA。

[1177] 在一个特定的实施方式中,靶向CRISPR的RNA表达系统至少包含:

[1178] 编码与植物中的靶序列杂交的引导RNA(gRNA)的核苷酸序列,其中该引导RNA包括引导序列和正向重复序列,和

[1179] 编码RNA靶向蛋白的核苷酸序列,

[1180] 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在植物细胞中可操作的相同或不同的调节元件的控制下。

[1181] 可以通过多种常规技术将包含靶向CRISPR系统的RNA的组分的DNA构建体引入植物,植物部分或植物细胞的基因组中。该方法通常包括以下步骤:选择合适的宿主细胞或宿主组织,将构建体引入宿主细胞或宿主组织,并从中再生植物细胞或植物。

[1182] 在特定实施方式中,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,气溶胶束注射的技术将DNA构建体引入植物细胞,或可使用生物弹射法(例如DNA轰击)将DNA构建体直接引入植物组织(另见Fu等, Transgenic Res.2000Feb;9(1):11-9)。粒子轰

击的基础是被感兴趣基因包被的粒子向细胞的加速,从而导致原生质被粒子穿透,并通常稳定地整合到基因组中(参见例如Klein等,Nature (1987),Klein等,Bio/Technology (1992),Casas 等,Proc.Natl.Acad.Sci.USA (1993))。

[1183] 在特定的实施方式中,可以通过农杆菌介导的转化将包含靶向CRISPR 系统的RNA的组分的DNA构建体引入植物中。可以将DNA构建体与合适的T-DNA侧翼区域组合,并引入常规的根癌土壤杆菌宿主载体中。通过感染植物或通过用含有一种或多种Ti(诱导肿瘤)质粒的农杆菌属细菌孵育植物原生质体,可以将外源DNA掺入植物的基因组中。(参见例如Fraley等,(1985),Rogers等,(1987)和美国专利号5,563,055)。

[1184] 植物启动子

[1185] 为了确保在植物细胞中的适当表达,本文所述的CRISPR-Cas CRISPR 系统的组分通常置于植物启动子,即在植物细胞中可操作的启动子的控制下。设想使用不同类型的启动子。

[1186] 组成型植物启动子是能够在植物的全部或几乎所有发育阶段表达其在所有或几乎所有植物组织中控制的开放阅读框(ORF)的启动子(称为“组成型表达”)。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。本发明设想了修饰RNA序列的方法,因此也设想了调节植物生物分子的表达。因此,在本发明的特定实施方式中,将RNA靶向CRISPR系统的一种或多种元件置于可调节的启动子的控制下是有利的。“调控启动子”是指不是组成型地而是以时间和/或空间调控的方式引导基因表达的启动子,包括组织特异性的,组织优选的和诱导型的启动子。在特定的实施方式中,靶向 CRISPR组分的一种或多种RNA在组成型启动子例如花椰菜花叶病毒的控制下表达。可以利用35S启动子优先选择的启动子来靶向特定植物组织内某些细胞类型(例如叶或根中的维管细胞或种子的特定细胞)中的增强表达。用于RNA靶向CRISPR系统的特定启动子的实例可在Kawamata等,(1997) Plant Cell Physiol 38:792-803;Yamamoto等,(1997) Plant J 12:255-65;Hire等,(1992) Plant Mol Biol 20:207-18,Kuster等,(1995) Plant Mol Biol 29:759-72,和Capana等,(1994) Plant Mol Biol 25:681-91中找到。

[1187] 可诱导的并且允许时空控制基因编辑或基因表达的启动子的实例可以使用能量的形式。能量的形式可以包括但不限于声能,电磁辐射,化学能和 /或热能。诱导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)或光诱导系统(光致色素、LOV结构域或隐色)。作为光诱导转录效应子(LITE)的一种,它以序列特异性方式引导转录活性的变化。光诱导系统的组分可以包括靶向CRISPR-Cas的RNA,光响应性细胞色素异二聚体(例如来自拟南芥)和转录激活/抑制域。可诱导的 DNA结合蛋白的进一步实例及其使用方法在美国61/736465和美国 61/721,283中提供,其通过引用整体并入本文。

[1188] 在特定的实施方式中,瞬时或可诱导的表达可以通过使用例如化学调节的启动子来实现,即,由此外源化学物质的应用诱导基因表达。基因表达的调节也可以通过化学可抑制的启动子获得,其中化学试剂的应用抑制基因表达。化学诱导型启动子包括但不限于由苯磺酰胺除草剂安全剂激活的玉米In2-2启动子(De Veylder等,(1997) Plant Cell Physiol 38:568-77),玉米 GST启动子(GST-1127,W093/01294)(其被用作发芽前除草剂的疏水性亲电化合物激活),和烟草PR-1a启动子(Ono等,(2004) Biosci Biotechnol Biochem

68:803-7) 被水杨酸激活。也可以在本文中使用由抗生素调节的启动子,例如四环素诱导型和四环素抑制型启动子(Gatz等,(1991) Mol Gen Genet 227:229-37;U.S.Patent Nos.5,814,618and 5,789,156)。

[1189] 特定植物器官的转运和/或表达

[1190] 表达系统可包含用于易位至和/或在特定植物细胞器中表达的元件。

[1191] 叶绿体靶向

[1192] 在特定的实施方式中,设想将靶向CRISPR系统的RNA用于特异性修饰叶绿体基因的表达和/或翻译或确保在叶绿体中的表达。为此目的,使用叶绿体转化方法或将靶向CRISPR组分的RNA分隔到叶绿体上。例如,在质体基因组中例如通过花粉的基因流引入遗传修饰可以减少生物安全性问题。

[1193] 叶绿体转化的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W02010061186中所述的涉及将转化盒从核基因组转移至质体的方法。

[1194] 或者,设想将一种或多种靶向CRISPR组分的RNA靶向植物叶绿体。这是通过将表达叶绿体转运肽(CTP)或质体转运肽的序列掺入表达构建体中来实现的,该序列可操作地连接至编码RNA靶向蛋白的序列的5'区域。CTP在易位到叶绿体中的处理步骤中被去除。所表达蛋白质的叶绿体靶向是本领域技术人员众所周知的(参见例如Protein Transport into Chloroplasts, 2010,Annual Review of Plant Biology,Vol.61:157-180)。在这样的实施方式中,还期望将一种或多种引导RNA靶向植物叶绿体。可用于通过叶绿体定位序列将引导RNA转移到叶绿体中的方法和构建体描述于例如美国 20040142476中,其通过引用并入本文。可以将这种构建体的变体并入本发明的表达系统中,以有效地使RNA靶向-引导RNA移位。

[1195] 在藻类细胞中引入包含CRISPR-RNA靶向系统的多核苷酸。

[1196] 转基因藻类(或其他植物,如油菜)在生产植物油或生物燃料(如醇(尤其是甲醇和乙醇)或其他产品)中可能特别有用。这些可以被设计成表达或过量表达用于石油或生物燃料工业的高含量的油或醇。

[1197] 美国8945839描述了一种使用Cas9工程化微藻类(莱茵衣藻细胞)物种的方法。使用类似的工具,本文所述的RNA靶向CRISPR系统的方法可以应用于衣藻属物种和其他藻类。在特定实施方式中,将RNA靶向蛋白和一种或多种引导RNA引入使用在组成性启动子例如Hsp70A-Rbc S2或 $\beta$ 2-微管蛋白的控制下表达RNA靶向蛋白的载体表达的藻类中。可选地,使用含有T7启动子的载体来递送引导RNA。或者,可以将靶向mRNA的RNA和体外转录的引导RNA递送至藻类细胞。电穿孔方案对于技术人员是可用的,例如来自GeneArt Chlamydomonas Engineering试剂盒的标准推荐方案。

[1198] 在酵母细胞中引入编码RNA靶向组分的多核苷酸

[1199] 在特定的实施方式中,本发明涉及RNA靶向CRISPR系统在酵母细胞中RNA编辑中的用途。转化酵母细胞的方法可用于引入编码靶向CRISPR系统组分的RNA的多核苷酸,这是本领域技术人员众所周知的,并且Kawai等,2010,Bioeng Bugs.2010Nov-Dec;1(6):395-403)对其进行了评论。非限制性实例包括通过乙酸锂处理(其可进一步包括载体DNA和PEG处理)轰击或通过电穿孔转化酵母细胞。

[1200] 植物和植物细胞中的RNA靶向CRISPR系统组分的瞬时表达

[1201] 在特定的实施方式中,设想了引导RNA和/或RNA靶向基因在植物细胞中瞬时表达。在这些实施方式中,仅当引导RNA和RNA靶向蛋白都存在于细胞中时,靶向RNA的CRISPR系统才能确保修饰RNA靶分子,从而可以进一步控制基因表达。由于RNA靶向酶的表达是瞬时的,因此从此类植物细胞再生的植物通常不包含外源DNA。在特定的实施方式中,RNA靶向酶由植物细胞稳定表达,并且引导序列是瞬时表达的。

[1202] 特别优选的实施方式中,可以使用植物病毒载体将靶向CRISPR系统组分的RNA引入植物细胞中(Scholthof等,1996,Annu Rev Phytopathol. 1996;34:299-323)。在另外的特定实施方式中,所述病毒载体是来自DNA病毒的载体。例如,双生病毒(例如卷心菜叶卷曲病毒、豆黄矮化病毒、小麦矮化病毒、番茄叶卷曲病毒、玉米条纹病毒、烟草叶卷曲病毒或番茄金黄色花叶病毒)或纳米病毒(例如蚕豆坏死黄色)病毒)。在其他特定的实施方式中,所述病毒载体是来自RNA病毒的载体。例如,烟草病毒(例如烟草拨浪鼓病毒、烟草花叶病毒),potex病毒(例如马铃薯病毒X)或大肠病毒(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,在避免生产转基因植物的背景下,这是令人感兴趣的。

[1203] 在特定实施方式中,用于RNA靶向CRISPR构建体的瞬时表达的载体是例如pEAQ载体,该载体针对农杆菌介导的瞬时表达进行了定制(Sainsbury F.等,Plant Biotechnol J. 2009Sep;7(7):682-93)。使用修饰的卷心菜叶卷曲病毒(CaLCuV)载体证明了基因组位置的精确靶向,可以在表达Cas13的稳定转基因植物中表达gRNA(参见Scientific Reports 5,Article number:14926(2015),doi:10.1038/srep14926)。

[1204] 在特定的实施方式中,可以将编码引导RNA或crRNA和/或RNA靶向基因的双链DNA片段瞬时引入植物细胞中。在这样的实施方式中,以足够的量提供引入的双链DNA片段以修饰细胞中的RNA分子,但是在经过预期的时间段之后或在一个或多个细胞分裂之后不持续存在。植物中直接DNA转移的方法是技术人员已知的(参见例如Davey等,Plant Mol Biol.1989 Sep;13(3):273-85)。

[1205] 在其他实施方式中,将编码RNA靶向蛋白的RNA多核苷酸引入植物细胞,然后由宿主细胞翻译和加工,从而产生足够数量的蛋白来修饰RNA分子细胞(在至少一个引导RNA的存在下),但其在经过一段预期的时间后或在一个或多个细胞分裂后不会持续存在。将mRNA引入植物原生质体以进行瞬时表达的方法是本领域技术人员已知的(参见例如Gallie,Plant Cell Reports(1993),13;119-122)。还设想了上述不同方法的组合。

[1206] 将靶向RNA的CRISPR组分递送至植物细胞

[1207] 在特定实施方式中,令人感兴趣的是将靶向CRISPR系统的RNA的一种或多种组分直接递送至植物细胞。这尤其对于非转基因植物的产生是有意义的。在特定的实施方式中,一种或多种RNA靶向组分是在植物或植物细胞外部制备并递送至细胞的。例如,在特定的实施方式中,在导入植物细胞之前在体外制备RNA靶向蛋白。靶向RNA的蛋白可以通过本领域技术人员已知的多种方法来制备,并且包括重组生产。表达后,将RNA靶向蛋白分离出来,必要时对其进行折叠,纯化并任选进行处理以去除任何纯化标签,例如His标签。一旦获得粗制的,部分纯化的或更完全纯化的RNA靶向蛋白,就可以将蛋白引入植物细胞。

[1208] 在特定的实施方式中,将RNA靶向蛋白与靶向目的RNA的引导RNA混合以形成预组装的核糖核蛋白。

[1209] 可以通过电穿孔,通过用RNA靶向相关的基因产物包被的颗粒轰击,通过化学转染

或通过某种其他跨细胞膜运输的方式,通过电穿孔将单个组分或预组装的核糖核蛋白引入植物细胞。例如,已经证明用预组装的CRISPR 核糖核蛋白转染植物原生质体以确保植物基因组的靶向修饰(如Woo等, Nature Biotechnology, 2015; DOI:10.1038/nbt.3389所述)。可以对这些方法进行修饰,以实现植物中RNA分子的靶向修饰。

[1210] 在特定的实施方式中,使用纳米颗粒将靶向CRISPR系统组分的RNA 引入植物细胞。可以将作为蛋白质或核酸或其组合的组分上载或包装在纳米颗粒中并施用于植物上(例如WO 2008042156和US 20130185823所述)。特别地,本发明的实施方式包括与编码RNA靶向蛋白的DNA分子一起上传或填充的纳米颗粒,如WO2015089419中所述的编码引导RNA和/或分离的引导RNA的DNA分子。

[1211] 将靶向RNA的CRISPR系统的一种或多种组分引入植物细胞的其他方法是使用细胞穿透肽(CPP)。因此,特别地,本发明的实施方式包括包含与 RNA靶向蛋白连接的细胞穿透肽的组合物。在本发明的特定实施方式中,将靶向RNA的蛋白质和/或引导RNA与一个或多个CPP偶联以有效地将其转运至植物原生质体内(Ramakrishna (2014, Genome Res. 2014年6月; 24(6):1020-7 for Cas9 in human cells)。在其他实施方式中, RNA靶向基因和/或引导RNA由与一个或多个CPP偶联的一个或多个环状或非环状DNA分子编码。CPPs通常被描述为少于35个氨基酸的短肽,其来源于蛋白质或嵌合序列,能够在细胞内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)。可以穿透生物膜并因此触发各种生物分子穿过细胞膜进入细胞质并改善其细胞内路径,从而促进生物分子与靶标的相互作用。CPP的实例包括:Tat, HIV 1型病毒复制所需的核转录激活蛋白,渗透素,卡波西成纤维细胞生长因子(FGF)信号肽序列,整联蛋白β3信号肽序列;聚精氨酸肽Args序列,鸟嘌呤富分子转运蛋白,甜箭肽等。

[1212] 为植物、藻类或真菌应用设想的靶RNA

[1213] 靶RNA,即感兴趣的RNA,是本发明靶向的RNA,导致在靶RNA上感兴趣的靶位点募集并与RNA靶向蛋白结合。靶RNA可以是任何合适形式的RNA。在一些实施方式中,这可以包括mRNA。在其他实施方式中,靶 RNA可以包括转移RNA (tRNA)或核糖体RNA (rRNA)。在其他实施方式中,靶RNA可以包括干扰RNA (RNAi),微RNA (miRNA),微开关,微酶,卫星 RNA和RNA病毒。靶RNA可以位于植物细胞的细胞质中,或位于细胞核中或植物细胞器中,例如线粒体,叶绿体或质体中。

[1214] 在特定的实施方式中,靶向RNA的CRISPR系统用于切割RNA或以其他方式抑制RNA表达。

[1215] 利用RNA靶向CRISPR系统通过RNA调控来调控植物基因表达、

[1216] 还可通过控制RNA加工,将RNA靶向蛋白与合适的引导RNA一起用于靶向基因表达。RNA加工的控制可以包括RNA加工反应,例如RNA剪接,包括替代剪接;RNA剪接。病毒复制(尤其是植物病毒,包括植物体内的类病毒体和tRNA生物合成。RNA靶向蛋白与合适的引导RNA结合也可用于控制RNA激活 (RNAa)。RNAa促进基因表达,因此可以通过破坏或减少RNAa从而减少基因表达的促进来实现对基因表达的控制。

[1217] 本发明的靶向RNA的效应子蛋白可以进一步用于植物中的抗病毒活性,特别是针对RNA病毒的抗病毒活性。可以使用对所选病毒RNA序列有选择性的合适的引导RNA,将效应

子蛋白靶向病毒RNA。特别地,效应子蛋白可以是切割RNA的活性核酸酶,例如单链RNA。因此,提供了本发明的RNA靶向效应子蛋白作为抗病毒剂的用途。可以以这种方式抵抗的病毒的示例包括但不限于烟草花叶病毒(TMV)、番茄斑萎病毒(TSWV)、黄瓜花叶病毒(CMV)、马铃薯病毒Y(PVY)、花椰菜花叶病毒(CaMV)(RT病毒)、李子痘病毒(PPV)、雀麦花叶病毒(BMV)和马铃薯X病毒(PVX)。

[1218] 本文进一步描述了调节植物,藻类或真菌中RNA表达的实例,作为靶向基因修饰的替代方案。

[1219] 特别令人感兴趣的是通过mRNA的调控切割对基因表达的调控。这可以通过将靶向RNA的元件置于如本文所述的受调控的启动子的控制下来实现。

[1220] 使用RNA靶向CRISPR系统恢复tRNA分子的功能

[1221] Pring等描述了植物线粒体和叶绿体中的RNA编辑,这些编辑可以改变mRNA序列以编码不同于DNA的蛋白质。(Plant Mol.Biol.(1993) 21(6): 1163-1170.doi:10.1007/BF00023611)。在本发明的特定实施方式中,可以将特异性靶向线粒体和叶绿体mRNA的靶向CRISPR系统的RNA的元件引入植物或植物细胞中,以在这种植物细胞器中表达不同的蛋白质,从而模拟体内发生的过程。

[1222] 使用RNA靶向CRISPR系统作为替代RNA干扰抑制RNA表达的方法

[1223] RNA靶向CRISPR系统具有类似于RNA抑制或RNA干扰的用途,因此也可以替代此类方法。在特定的实施方式中,本发明的方法包括使用靶向CRISPR的RNA作为例如RNA的替代物。干扰核糖核酸(例如siRNA或shRNA或dsRNA)。本文进一步描述了抑制植物,藻类或真菌中RNA表达的实例,作为靶向基因修饰的替代方案。

[1224] 使用RNA靶向CRISPR系统控制RNA干扰

[1225] 通过减少体内或体外干扰RNA或miRNA的寿命,对于干扰RNA或miRNA的控制可能有助于减少那些方法所产生的脱靶效应(OTE)。在特定实施方式中,靶RNA可以包括干扰RNA,即参与RNA干扰途径的RNA,例如shRNA、siRNA等。在其他实施方式中,靶RNA可以包括微小RNA(miRNA)或双链RNA(dsRNA)。

[1226] 在其他特定的实施方式中,如果RNA靶向蛋白和合适的引导RNA被选择性地表达(例如在受调控的启动子例如组织或细胞周期特异性启动子和/或增强子的控制下在空间或时间上)。这可用于“保护”细胞或系统(体内或体外)免受这些细胞中RNAi的侵害。这在不需要RNAi的邻近组织或细胞中或用于比较表达或不表达效应子蛋白和合适的引导的细胞或组织的目的可能是有用的(例如分别在控制RNAi的地方或不控制RNAi的地方)。RNA靶向蛋白可用于控制或结合包含RNA或由RNA组成的分子,例如核酶,核糖体或核糖开关。在本发明的实施方式中,引导RNA可以将RNA靶向蛋白募集到这些分子,使得RNA靶向蛋白能够与它们结合。

[1227] 根据本公开内容,本发明的靶向RNA的CRISPR系统无需过度实验即可用于植物体内RNAi技术领域,包括昆虫害虫管理、植物病害管理和除草剂抗性管理以及在植物测定中和用于其他应用(参见例如Kim等,in Pesticide Biochemistry and Physiology (Impact Factor:2.01) .01/2015;120.DOI: 10.1016/j.pestbp.2015.01.002;Sharma等,in Academic Journals(2015), Vol.12(18) pp2303-2312);Green J.M,in Pest Management Science, Vol 70(9), pp 1351-1357),因为本申请为所述系统的有根据的工程化提供了基

础。

[1228] 利用RNA靶向CRISPR系统修饰核糖开关并控制植物、藻类和真菌的代谢调节。

[1229] 核糖开关(也称为核酶)是信使RNA的调控片段,与小分子结合,进而调控基因表达。这种机制使细胞能够感知这些小分子的细胞内浓度。特定的核糖开关通常通过改变该基因的转录,翻译或剪接来调节其邻近基因。因此,在本发明的特定实施方式中,设想通过使用RNA靶向蛋白与合适的引导RNA结合以靶向核糖开关来控制核糖开关活性。这可以通过切割或结合至核糖开关来实现。在特定的实施方式中,设想了核糖开关活性的降低。近来,表征了结合硫胺焦磷酸盐(TPP)的核糖开关,并发现其调节植物和藻类中的硫胺生物合成。此外,似乎该元素是植物中初级代谢的必需调节剂(Bocobza 和Aharoni, Plant J. 2014年8月;79(4):693-703. doi:10.1111/tpj.12540. Epub 2014年6月17日)。TPP核糖开关还存在于某些真菌中,例如在*Neurospora crassa*中,它控制选择性剪接以有条件地产生上游开放阅读框(uORF),从而影响下游基因的表达(Cheah MT等, (2007) Nature 447(7143):497-500. doi:10.1038/nature05769)。本文所述的靶向CRISPR系统的RNA可用于操纵植物、藻类或真菌中的内源核糖开关活性,并因此改变其控制的下游基因的表达。在特定的实施方式中,靶向RNA的CRISPR系统可以用于体内或体外测定核糖开关功能以及研究其与代谢网络的相关性。在特定实施方式中,靶向CRISPR系统的RNA可以潜在地用于改造核糖开关,作为植物和平台中用于基因控制的代谢物传感器。

[1230] RNA靶向CRISPR系统在植物、藻类或真菌的RNAi中的应用

[1231] 鉴定其敲除与表型改变有关的基因产物,可以通过RNAi筛查生物学途径并鉴定组成部分。在本发明的特定实施方式中,还可通过使用本文所述的指南29或指南30蛋白和合适的引导RNA在这些筛选之上或期间施加控制,以去除或降低筛选中RNAi的活性,从而恢复活性。(先前受干扰的)基因产物的表达(通过消除或减少干扰/抑制)。

[1232] RNA靶向蛋白在体内和体外RNA分子可视化中的应用

[1233] 在特定的实施方式中,本发明提供了核酸结合系统。RNA与互补探针的原位杂交是一项强大的技术。通常,荧光DNA寡核苷酸用于通过杂交检测核酸。通过某些修饰,例如锁定核酸(LNA),已经获得了提高效率,但是仍然需要有效且通用的替代方案。因此,RNA靶向系统的标记元件可以用作高效、可调适的原位杂交系统的替代品。

[1234] RNA靶向CRISPR系统在植物和酵母中的进一步应用

[1235] RNA靶向CRISPR系统在生物燃料生产中的用途

[1236] 如本文所用,术语“生物燃料”是由植物和植物来源的资源制成的替代燃料。可以从有机物质中提取可再生的生物燃料,这些有机物质的能量是通过固碳过程获得的,或者是通过利用或转化生物质制得的。该生物质可以直接用于生物燃料,也可以通过热转化、化学转化和生化转化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。生物乙醇主要是通过纤维素(淀粉)的糖发酵过程生产的,淀粉主要来自玉米和甘蔗。另一方面,生物柴油主要来自油料作物,例如油菜籽、棕榈和大豆。生物燃料主要用于运输。

[1237] 增强植物特性以生产生物燃料

[1238] 在特定的实施方式中,使用本文所述的使用RNA靶向CRISPR系统的方法来改变细胞壁的特性,以便于关键水解剂的进入,从而更有效地释放糖以进行发酵。在特定的实施方

式中,纤维素和/或木质素的生物合成被修饰。纤维素是细胞壁的主要组分。纤维素和木质素的生物合成是共同调节的。通过减少植物中木质素的比例,可以增加纤维素的比。在特定的实施方式中,本文所述的方法用于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。更具体而言,本文所述的方法用于下调至少第一个木质素生物合成基因,该基因选自如WO 2008064289 A2中所公开的4-香豆酸酯3-羟化酶(C3H)、苯丙氨酸氨裂合酶(PAL)、肉桂酸4-羟化酶(C4H)、羟肉桂酰基转移酶(HCT)、咖啡酸O-甲基转移酶(COMT)、咖啡酰基辅酶A 3-O-甲基转移酶(CCoAOMT)、阿魏酸5-羟化酶(F5H)、肉桂醇脱氢酶(CAD)、肉桂酰基辅酶A还原酶(CCR)、4-香豆酸酯-CoA连接酶(4CL)、单木酚-木脂蛋白特异性糖基转移酶和醛脱氢酶(ALDH)。

[1239] 在特定的实施方式中,本文所述的方法用于产生在发酵过程中产生较低水平的乙酸的植物块(也参见WO 2010096488)。

[1240] 修饰酵母以生产生物燃料

[1241] 在特定的实施方式中,本文提供的RNA靶向酶用于通过重组微生物生产生物乙醇。例如,RNA靶向酶可用于工程改造微生物,例如酵母,以由可发酵糖产生生物燃料或生物聚合物,并任选地能够降解源自农业废物的植物来源的木质纤维素作为可发酵糖的来源。更具体地,本发明提供了方法,其中靶向RNA的CRISPR复合物用于修饰生物燃料生产所需的内源基因的表达和/或修饰为什么可能干扰生物燃料合成的内源基因。更特别地,该方法包括刺激一种或多种核苷酸序列的编码在酶例如丙酮酸向乙醇或另一目的产物的转化中的酶在微生物如酵母中的表达。在特定的实施方式中,该方法确保刺激一种或多种酶的表达,这允许微生物降解纤维素,例如纤维素酶。在另外的实施方式中,靶向CRISPR复合物的RNA用于抑制与生物燃料生产途径竞争的内源性代谢途径。

[1242] 修饰类和植物以生产植物油或生物燃料

[1243] 例如,转基因藻类或其他植物(例如油菜)在生产植物油或生物燃料(例如醇(尤其是甲醇和乙醇))中可能特别有用。这些可以被设计成表达或过量表达用于石油或生物燃料工业的高含量的油或醇。

[1244] US 8945839描述了一种使用Cas9工程化微藻类(莱茵衣藻细胞)物种的方法。使用类似的工具,本文所述的RNA靶向CRISPR系统的方法可以应用于衣藻属物种和其他藻类。在特定的实施方式中,将RNA靶向效应子蛋白和引导RNA引入使用表达的藻类的载体中,所述载体在组成型启动子如Hsp70A-Rbc S2或 $\beta$ 2-微管蛋白的控制下表达RNA靶向效应子蛋白。引导RNA将使用含有T7启动子的载体进行递送。或者,可以将体外转录的引导RNA递送至藻类细胞。电穿孔方案遵循GeneArt Chlamydomonas Engineering 试剂盒中推荐的标准推荐方案。

[1245] RNA靶向酶在植物中的特定应用

[1246] 在特定的实施方式中,本发明由于能够切割病毒RNA而可以用作植物系统中病毒去除的疗法。先前在人体系统中的研究表明,利用CRISPR成功靶向单链RNA病毒丙型肝炎(A.Price等,Proc.Natl.Acad.Sci,2015)。这些方法也可以适用于在植物中使用RNA靶向CRISPR系统。

[1247] 改良的植物

[1248] 本发明还提供了可通过本文提供的方法获得和获得的植物和酵母细胞。通过本文

所述方法获得的改良植物可通过基因的修饰表达而用于食品或饲料生产,例如确保对植物害虫、除草剂、干旱,低温或高温、过量水分等的耐受性。

[1249] 通过本文描述的方法获得的改良植物,特别是农作物和藻类,通过表达例如比野生型中通常可见的更高的蛋白质,碳水化合物,营养或维生素水平,可用于食品或饲料生产。在这方面,优选改良的植物,尤其是豆类和块茎。

[1250] 改良的藻类或其他植物(例如油菜)在生产植物油或生物燃料(例如醇(尤其是甲醇和乙醇))中特别有用。这些可以被设计成表达或过量表达用于石油或生物燃料工业的高含量的油或醇。

[1251] 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。本文所设想的植物部分可以是存活的、不能存活的、可再生的和/或不可再生的。

[1252] 本文还涵盖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、合子或体细胞的后代或杂种也包括在本发明的范围内。这样的植物可以包含插入或代替靶序列的异源或外源DNA序列。或者,此类植物可以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,此类植物将仅通过存在特定修饰而与其祖先植物不同。

[1253] 在本发明的一个实施方式中,CRISPR-Cas系统用于工程化病原体抗性植物,例如通过对细菌,真菌或病毒引起的疾病产生抗性。在某些实施方式中,病原体抗性可以通过对农作物进行改造以产生将被害虫摄入的CRISPR-Cas系统来实现,从而导致死亡。在本发明的一个实施方式中,CRISPR-Cas系统用于工程化非生物胁迫耐受性。在另一个实施方式中,使用CRISPR-Cas系统来工程化干旱胁迫耐受性或盐胁迫耐受性,或冷或热胁迫耐受性。Younis et al.2014,Int.J.Biol.Sci.10;1150评述了植物育种方法的潜在目标,所有这些都可通过使用本文所述的CRISPR-Cas系统进行校正或改良。一些非限制性目标作物包括拟南芥(*Arabidops Zea mays thaliana*)、水稻(*Oryza sativa* L)、李属(*Prunus domestica* L.)、陆地棉(*Gossypium hirsutum*)、仿古烟草(*Nicotianarusta*)、玉米(*Zea mays*)、苜蓿(*Medicago sativa*)、本氏烟草(*Nicotiana benthamiana*)和拟南芥(*Arabidopsis thaliana*)。

[1254] 在本发明的一个实施方式中,CRISPR-Cas系统用于处理农作物害虫。例如,可在植物害虫中操作的CRISPR-Cas系统可从植物宿主表达或直接转移至靶标,例如使用病毒载体。

[1255] 在一个实施方式中,本发明提供了一种从杂合非人类起始生物有效产生纯合生物的方法。在一个实施方式中,本发明用于植物育种。在另一个实施方式中,本发明用于动物育种。在这样的实施方式中,通过干扰或抑制参与双链断裂、染色体配对和/或链交换的至少一种靶基因来防止或抑制重组来生成纯合生物(例如植物或动物)。

[1256] CRISPR-CAS效应子蛋白复合物可以在植物中使用。

[1257] 在一些实施方式中,本发明包括修饰细胞或生物的方法。该细胞可以是原核细胞或真核细胞。该细胞可以是哺乳动物细胞。哺乳动物细胞可以是非人类的灵长类、牛、猪、啮齿动物或小鼠细胞。该细胞可以是非哺乳动物的真核细胞,例如家禽、鱼或虾。该细胞也可以是植物细胞。植物细胞可以是农作物,例如木薯、玉米、高粱、小麦或水稻。植物细胞也可

以是藻类,树木或蔬菜。通过本发明引入细胞的修饰可以是这样的,使得改变细胞和细胞的后代以改善生物产物(如抗体、淀粉、醇或其他所需细胞)输出的产生。通过本发明引入细胞的修饰可以使得细胞和细胞的后代包括改变产生的生物产物的改变。该系统可以包括一个或多个不同的载体。在本发明的一方面,效应子蛋白经密码子优化以表达所需的细胞类型,优选为真核细胞,优选为哺乳动物细胞或人细胞。可以将CRISPR-Cas系统(例如单个或多重)与作物基因组学的最新进展结合使用。此类CRISPR系统可用于执行有效且具有成本效益的植物基因或基因组或转录组查询或编辑或操作-例如,用于快速研究和/或选择和/或询问和/或比较和/或操作和/或植物基因或基因组的转化;例如,创建、鉴定、开发、优化或赋予植物特征或特性或转化植物基因组。因此,可以提高植物的产量,具有新的性状或特征组合的新植物或具有增强的性状的新植物。这样的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。因此,除非另有说明,否则本文中对动物细胞的引用在细节上作必要修改后也可适用于植物细胞。并且,本文中具有降低脱靶作用的酶和采用这种酶的系统可以用于植物应用中,包括本文中提及的那些。如所提供的,通过效应子蛋白和合适的引导(crRNA)修饰的工程植物及其后代。这些可能包括抗病或抗旱的农作物,例如小麦、大麦、水稻、大豆或玉米;经过修饰以去除或降低自花授粉的能力的植物(但可以替代地,可以替代地杂交);以及过敏性食品,例如花生和坚果,其中的免疫原性蛋白已通过效应子蛋白和合适的导向物靶向而被禁用,破坏或破坏。使用经典CRISPR-Cas系统的任何方面均可适用于与Cas蛋白无关的CRISPR系统,例如Cas13效应子蛋白系统。

#### [1258] 条件模型

[1259] 本发明的方法可用于产生植物,动物或细胞,所述植物,动物或细胞可用于建模和/或研究目的遗传或表观遗传条件,例如通过目的突变模型或疾病模型。如本文所用,“疾病”是指受试者中的疾病,病症或适应症。例如,本发明的方法可用于产生动物或细胞,该动物或细胞包括与疾病相关的一个或多个核酸序列的修饰,或其中表达一个或多个核酸序列的植物,动物或细胞。与疾病相关的改变。这样的核酸序列可以编码或翻译疾病相关的蛋白质序列,或者可以是疾病相关的控制序列。因此,应理解在本发明的实施方式中,植物、受试者、患者、生物或细胞可以是非人类受试者、患者、生物或细胞。因此,本发明提供了通过本方法产生的植物,动物或细胞或其后代。后代可以是所生产的植物或动物的克隆,或者可以是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的有性繁殖而产生的。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。在培养细胞的情况下,如果满足适当的培养条件,并且优选地如果该细胞适合于该目的,则可以建立细胞系(例如干细胞)。还设想了本发明产生的细菌细胞系。因此,还设想了细胞系。在某些方法中,可以使用常用的方法,使用疾病模型来研究突变或更普遍的变化(例如基因或基因产物的表达减少)对动物或细胞以及疾病的发展和/或进展的影响。在疾病的研究中。或者,这种疾病模型可用于研究药物活性化合物对疾病的影响。在某些方法中,疾病模型可用于评估潜在基因治疗策略的功效。即,可以修饰与疾病相关的RNA,从而显示或抑制或减少疾病的发展和/或进展,然后测试化合物对进展或抑制或减少的作用。

[1260] 在利用CRISPR-Cas效应子蛋白及其复合物和编码其的核酸分子以及使用其的方法用于实施本发明时,参考了以下内容:Genome-Scale CRISPR-Cas9 Knockout Screening

in Human Cells, Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., Zhang, F. Science Dec 12. (2013). [线上公布早于纸印]; 以一下最终编辑形式出版: Science. 2014 Jan 3; 343 (6166): 84-87. Shalem等涉及一种在基因组范围内询问基因功能的新方法。他们的研究表明, 以18,080个基因为目标的基因组规模的CRISPR-Cas9基因敲除 (GeCKO) 文库的递送具有64,751个独特的引导序列, 可在人细胞中进行阴性和阳性选择筛选。首先, 作者展示了使用GeCKO文库鉴定癌症和多能干细胞中细胞活力所必需的基因。接下来, 在黑色素瘤模型中, 作者筛选了基因丢失与维拉非尼抗药性有关的基因, 维拉非尼是一种抑制突变蛋白激酶BRAF的药物。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, 以及新型的hitsNF2、CUL3、TADA2B和TADA1。作者观察到靶向相同基因的独立引导RNA之间的高度一致性和较高的命中率, 从而证明了用Cas9进行基因组规模筛选的希望。还参考美国专利公开号US20140357530; 美国专利公开号US20140357530; 和PCT专利公开W02014093701, 其通过引用并入本文。

[1261] 在此使用的术语“与...相关联”是指功能域与CRISPR-Cas效应子蛋白或衔接蛋白的结合。关于一个分子如何与另一个分子“相关联”, 例如在衔接蛋白和功能结构域之间或在CRISPR-Cas效应子蛋白和功能结构域之间, 它被使用。在这种蛋白质-蛋白质相互作用的情况下, 可以以抗体识别表位的方式从识别的角度看待这种联系。或者, 一种蛋白质可以通过两者的融合与另一种蛋白质缔合, 例如一个亚基与另一亚基融合。融合通常通过将氨基酸序列与另一个氨基酸序列相加而发生, 例如通过将编码每种蛋白质或亚基的核苷酸序列剪接在一起。或者, 这可以基本上被视为两个分子之间的结合或直接键合, 例如融合蛋白。在任何情况下, 融合蛋白可在两个感兴趣的亚基之间 (即在酶和功能域之间或在衔接蛋白和功能域之间) 包括接头。因此, 在一些实施方式中, CRISPR-Cas效应子蛋白或衔接蛋白通过与功能域结合而与之结合。在其他实施方式中, CRISPR-Cas效应子蛋白或衔接蛋白与功能结构域相关联, 因为二者任选地经由中接头融合在一起。

#### [1262] 治疗应用

[1263] 由于本申请为系统的知情工程提供了基础, 因此根据本公开内容, 可以将本发明的系统应用于本发明的先前的RNA切割技术领域, 而无需进行过多的实验, 而无需进行过多的实验。本发明提供了由RNA, 有毒RNA和/或突变的RNA (例如剪接缺陷或截短) 的过表达引起的疾病的治疗性治疗。毒性RNA的表达可能与核内含物的形成以及脑, 心脏或骨骼肌中迟发性退行性改变有关。在研究得最好的实例中, 肌强直性营养不良, 看来毒性RNA的主要致病作用是整合结合蛋白并损害其他剪接的调控 (Hum. Mol. Genet. (2006) 15 (suppl 2): R162-R169)。遗传学家特别关注强直性肌营养不良症 [肌营养不良症 (DM)], 因为它可产生极为广泛的临床特征。部分列出的内容包括肌肉消瘦, 白内障, 胰岛素抵抗, 睾丸萎缩, 心脏传导减慢, 皮肤肿瘤以及对认知的影响。DM的经典形式, 现在称为DM 1型 (DM1), 是由DMPK (编码胞质蛋白激酶的基因) 的3'-非翻译区 (UTR) 中CTG重复序列的扩增引起的。

[1264] 先天免疫系统主要通过识别被感染细胞内部的病毒核酸来检测病毒感染, 这被称为DNA或RNA感应。体外RNA感应测定可用于检测特定的RNA底物。RNA靶向效应子蛋白可以例如用于活细胞中基于RNA的感测。应用的实例是通过感测例如疾病特异性RNA的诊断方法。本发明的靶向RNA的效应子蛋白可以进一步用于抗病毒活性, 特别是针对RNA病毒。可以

使用对所选病毒RNA序列有选择性的合适的引导RNA,将效应子蛋白靶向病毒 RNA。特别地,效应子蛋白可以是切割RNA的活性核酸酶,例如单链RNA。预期本发明的酶系统靶向RNA和上述RNA的治疗剂量为约0.1至约 2mg/kg,该剂量可以在监测到的应答下依次给药,并且如果需要的话可以重复给药,最高至约40mg/kg。每位患者7至10剂。有利地,在治疗方案期间从每个患者收集样品,以确定治疗的有效性。例如,可以分离并定量RNA 样品以确定表达是否降低或改善。这种诊断在本领域技术人员的能力范围内。

[1265] 在一些实施例中,所述疾病是由G→A或C→T点突变或病原性SNP引起的。在一些实施例中,所述疾病由T→C或A→G点突变或病原性SNP引起。例如,所述疾病可以是癌症,血友病,β地中海贫血,马凡氏综合症和维斯科特-奥尔德里奇综合症。

[1266] 示例性疗法

[1267] 本发明还考虑了将CRISPR-Cas系统和本文所述的碱基编辑器用于多种疾病和病症的治疗。在一些实施方式中,本文所述的发明涉及一种治疗方法,其中通过CRISPR或碱基编辑器离体编辑细胞以调节至少一个基因,随后将经编辑的细胞施用于需要其的患者。在一些实施方式中,编辑涉及敲入、敲除或敲低细胞中至少一种靶基因的表达。在特定实施方式中,编辑将可包含一个或多个外显子和内含子或天然或合成内含子的外源,基因,小基因或序列插入靶基因的基因座,热点基因座,安全港基因座中。可以引入新基因或遗传元件而不破坏相邻基因的表达或调控,或不通过插入或缺失进行校正的 DNA序列中编码目标基因调控元件的一个或多个突变的基因组位置。在一些实施方式中,编辑包括在靶细胞中的核酸(例如基因组DNA)中引入一个或多个点突变。

[1268] 在实施方式中,所述治疗是针对器官的疾病/病症,包括肝病、眼病、肌肉病、心脏病、血液病、脑病、肾脏病、或者可以包括自身免疫性疾病、中枢神经系统疾病的治疗、癌症和其他增生性疾病、神经退行性疾病、炎性疾病、代谢性疾病、肌肉骨骼疾病等。

[1269] 特定疾病/疾病包括软骨发育不全、色盲、酸性麦芽糖酶缺乏症、肾上腺髓质营养不良症、心功能不全综合征、α-1抗胰蛋白酶缺乏症、α地中海贫血、雄激素不敏感综合征、精子综合征、致心律失常性右心室、发育不良、共济失调毛细血管扩张症、β贫血综合征、蓝色橡胶泡痣综合征、canavan 病、慢性肉芽肿性疾病(CGD)、cri du chat综合征、囊性纤维化、真皮病、外胚层发育不良、范可尼贫血、骨化性骨增生、渐进性X综合征、半乳糖症、全身型神经节苷脂贮积病(例如GM1)、血色素沉着病、β-珠蛋白(HbC)第六密码子中的血红蛋白C突变、血友病、亨廷顿氏病、霍勒氏综合症、低血磷症、克莱氏症候群、克拉伯斯病、朗格-基迪翁氏综合症、白细胞营养不良、长QT综合征、马凡氏综合症、莫比乌斯氏综合症、黏多糖贮积症(MPS)、指甲髌骨综合症、肾原性肾病血吸虫病、神经纤维瘤病、奈曼-皮克病、成骨不全症、卟啉症、Prader-Willi综合症、早衰症、变形杆菌综合症、视网膜母细胞瘤、Rett综合症、Rubinstein-Taybi综合症、Sanfilippo综合症、严重合并免疫缺陷症(SCID)、Shwachman综合症、镰状细胞疾病(镰状细胞性贫血)、史密斯·马格尼斯综合症、斯蒂克勒综合症、Tay-Sachs病、血小板缺乏症半径(TAR)综合症、特里格·柯林斯综合症、三染色体疾病、结节性硬化症、特纳综合症、尿素循环紊乱、冯·希佩尔·朗道病、Waardenburg综合症、威廉斯氏综合症、威尔逊病和Wiskott-Aldrich综合症。

[1270] 在一些实施方式中,该疾病与肿瘤抗原的表达相关,例如,与该肿瘤抗原的表达相关的增生性疾病,癌前状态,癌症或与非癌相关的适应症,其在一些实施方式中可以包含靶

标。选自B2M,CD247,CD3D,CD3E,CD3G, TRAC,TRBC1,TRBC2,HLA-A,HLA-B,HLA-C,DCK,CD52,FKBP1A,CIITA,NLRC5,RFXANK,RFX5,RFXAP或NR3C1,HAVCR2,LAG3, PDCD1,PD-L2,CTLA4,CEACAM(CEACAM-1,CEACAM-3和/或 CEACAM-5),VISTA,BTLA,TIGIT,LAIR1,CD160、2B4,CD80,CD86, B7-H3(CD113),B7-H4(VTCN1),HVEM(TNFRSF14或CD107),KIR,A2aR, MHC I类,MHC II类,GAL9,腺苷和TGFβ,或PTPN11 DCK,CD52,NR3C1, LILRB1,CD19;CD123;CD22;CD30;CD171;CS-1(也称为CD2子集1, CRACC,SLAMF7,CD319和19A24);C型凝集素样分子-1(CLL-1或 CLECL1);CD33;表皮生长因子受体变体III(EGFRvIII);神经节苷脂G2(GD2);神经节苷脂GD3(aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDG1cp(1-1)Cer);TNF受体家族成员B细胞成熟(BCMA);Tn抗原((Tn Ag)或(GalNAcα-Ser/Thr));前列腺特异性膜抗原(PSMA);受体酪氨酸激酶样孤儿受体1(ROR1);Fms样酪氨酸激酶3(FLT3);肿瘤相关糖蛋白72(TAG72);CD38;CD44v6;癌胚抗原(CEA);上皮细胞粘附分子(EPCAM);B7H3(CD276);套件(CD117);白介素13受体亚基α-2(IL-13Ra2或CD213A2);间皮素;白介素11受体α(IL-11Ra);前列腺干细胞抗原(PSCA);丝氨酸蛋白酶21(Testisin或 PRSS21);血管内皮生长因子受体2(VEGFR2);Lewis(Y)抗原;CD24;血小板衍生的生长因子受体β(PDGFR-β);阶段特异性胚胎抗原4(SSEA-4);CD20;叶酸受体α;受体酪氨酸蛋白激酶ERBB2(Her2/neu);n激酶ERBB2(Her2/neu);粘蛋白1,细胞表面缔合(MUC1);表皮生长因子受体(EGFR);神经细胞粘附分子(NCAM);采购前列腺酸磷酸酶(PAP);延伸因子2突变(ELF2M);以弗林 B2;成纤维细胞激活蛋白α(FAP);胰岛素样生长因子1受体(IGF-1受体),碳酸酐酶IX(CAIX);蛋白酶体(前体、巨蛋白因子)亚基,β型,9(LMP2);糖蛋白100(gp100);癌基因融合蛋白,由断点簇区域(BCR)和Abelson鼠白血病病毒病毒癌基因同源物1(Abl)(bcr-abl)组成;酪氨酸激酶ephrin A型受体2(EphA2);岩藻糖基GM1;唾液酸化的路易斯粘附分子(sLe);神经节苷脂 GM3(aNeu5Ac(2-3)bDGalp(1-4)bDG1cp(1-1)Cer);转谷氨酰胺酶5(TGS5);高分子量黑素瘤相关抗原(HMWMAA);邻乙酰基GD2神经节苷脂(OAcGD2);叶酸受体β;肿瘤内皮标记物1(TEM1/CD248);肿瘤内皮标志物7相关(TEM7R);claudin 6(CLDN6);甲状腺刺激激素受体(TSHR);G蛋白偶联受体C类5组,成员D(GPRC5D);X染色体开放阅读框61(CXORF61);CD97;CD179a;间变性淋巴瘤激酶(ALK);聚唾液酸;胎盘特异性1(PLAC1);globoH糖神经酰胺(GloboH)的六糖部分;乳腺分化抗原(NY-BR-1);uroplakin 2(UPK2);甲型肝炎病毒细胞受体1(HAVCR1);肾上腺素受体β3(ADRB3);pannexin 3(PANX3);G蛋白偶联受体20(GPR20);淋巴细胞抗原6复合体,基因座K 9(LY6K);嗅觉受体51E2(OR51E2);TCRγ备用阅读框蛋白(TARP);Wilms肿瘤蛋白(WT1);癌症/睾丸抗原1(NY-ESO-1);癌症/睾丸抗原2(LAGE-1a);黑色素瘤相关抗原1(MAGE-A1);位于12p染色体上的ETS易位变异基因 6(ETV6-AML);精子蛋白17(SPA17);X抗原家族,成员1A(XAGE1);血管生成素结合细胞表面受体2(Tie 2);黑色素瘤睾丸抗原-1(MAD-CT-1);黑色素瘤睾丸抗原2(MAD-CT-2);Fos相关抗原1;肿瘤蛋白p53(p53);p53突变体;前列腺素生存端粒酶前列腺癌肿瘤抗原1(PCTA-1或Galectin 8),被T细胞 1(MelanA或MART1)识别的黑色素瘤抗原;大鼠肉瘤(Ras)突变体;人端粒酶逆转录酶(hTERT);肉瘤易位断点;黑色素瘤凋亡抑制剂(ML-IAP);ERG(跨膜蛋白酶,丝氨酸2(TMPRSS2)ETS融合基因);N-乙酰氨基葡萄糖氨基转移酶V(NA17);配对盒蛋白Pax-3(PAX3);雄激素受体;细胞周期蛋白B1;v-myc禽骨髓细胞瘤病毒致癌基因神经母细胞瘤同源物(MYCN);Ras Homolog家庭成员C(RhoC);酪氨酸激酶相关蛋白2(TRP-2);细胞色素P450 1B1(CYP1B1);CCCTC结合因子

(锌指蛋白)样(BORIS或印迹部位调节剂的兄弟),T细胞3识别的鳞状细胞癌抗原(SART3);配对盒蛋白Pax-5(PAX5);proacrosin结合蛋白sp32(OY- TES1);淋巴细胞特异性蛋白酪氨酸激酶(LCK);激酶锚蛋白4(AKAP-4);滑膜肉瘤,X断点2(SSX2);先进糖基化终产物受体(RAGE-1);肾无处不在1(RU1);肾无处不在2(RU2);豆uma人乳头瘤病毒E6(HPV E6);人乳头瘤病毒E7(HPV E7);肠道羧基酯酶;热激蛋白70-2突变(mut hsp70-2);CD79a;CD79b;CD72;白细胞相关的免疫球蛋白样受体1(LAIR1);IgA受体的Fc片段(FCAR或CD89);白细胞免疫球蛋白样受体亚家族A成员2(LILRA2);CD300分子样家族成员f(CD300LF);C型凝集素结构域家族12成员A(CLEC12A);骨髓基质细胞抗原2(BST2);含有EGF样模块的类粘蛋白样激素受体样2(EMR2);淋巴细胞抗原75(LY75);Glypican-3(GPC3);Fc受体样5(FCRLS);免疫球蛋白λ样多肽1(IGLL1),CD19,BCMA,CD70,G6PC,肌营养不良蛋白,包括通过缺失或切除修饰外显子51,DMPK,CFTR(囊性纤维化跨膜电导调节剂)。在实施方式中,靶标包括CD70或CD33的敲入和B2M的敲除。在实施方式中,靶标包含或不具有其他靶标基因的TRAC和B2M或TRAC B2M和PD1的敲除。在某些实施方式中,所述疾病是囊性纤维化,其靶向SCNN1A基因,例如非编码或编码区,例如启动子区,或转录序列,例如内含子或外显子序列,靶向敲入CFTR内含子2内的序列,例如可引入编码CFTR外显子3-27的CFTR序列;CFTR内含子10中的序列和序列,其中可引入编码CFTR外显子11-27的序列。

[1271] 在实施方式中,该疾病是异色性白细胞营养不良,并且靶标是芳基硫酸酯酶A,该疾病是维斯科特-奥尔德里奇综合症,并且该靶标是维斯科特-奥尔德里奇综合症蛋白,该疾病是肾上腺白质营养不良,并且该靶标是ATP结合盒DI,该疾病是人类免疫缺陷病毒,靶标是5-CC趋化因子受体或CXCR4基因,疾病是β地中海贫血,靶标是血红蛋白β亚基,疾病是X连锁严重结合ID受体亚基γ,靶标是间白蛋白-2受体亚单位γ,疾病是多系统性溶酶体贮积病的胱氨酸病,靶标是胱氨酸,疾病是Diamond-Blackfan贫血,靶标是核糖体蛋白S19,疾病是范可尼贫血,靶标是范可尼贫血补充组(例如FNACA)(FNACB,FANCC,FANCD1,FANCD2,FANCE,FANCF,RAD51C),该疾病是Shwachman-Bodian-Diamond Bodian-Diamond综合征,而目标是Shwachman综合征e基因,该疾病是高雪氏病,目标是葡萄糖脑苷脂酶,该疾病是A血友病,目标是抗血友病因子OR因子VIII,圣诞节因子,丝氨酸蛋白酶,B血友病因子IX,该疾病是腺苷脱氨酶缺乏症(ADA-SCID),目标是腺苷脱氨酶,疾病是GM1神经节苷脂酶,目标是β-半乳糖苷酶,疾病是II型糖原贮积病,庞贝病,疾病是酸性麦芽糖酶缺乏酸,目标是α-葡萄糖苷酶,该疾病是Niemann-Pick病,与SMPD1相关(鞘磷脂磷酸二酯酶1或A和B型)酸,目标是鞘磷脂酶,该疾病是Krabbe病,球状细胞白细胞营养不良,目标是半乳糖苷酸神经酰胺酶或半乳糖苷神经酰胺脂酶,目标是半乳糖苷,人类白细胞抗原DR-15,DQ-6,该疾病是多发性硬化症(MS)DRB1,该疾病是单纯疱疹病毒1或2,目标是敲低RS1,RL2和/或LAT基因中的一个、两个或三个。在实施方式中,该疾病是HPV相关的癌症,其治疗包括编辑的细胞,该细胞包含结合分子,例如TCR或其抗原结合片段,以及抗体和其抗原结合片段,例如识别或结合人乳头瘤病毒的抗体。该疾病可以是乙肝,其目标是PreC、C、X、PreS1、PreS2、S、P和/或SP基因中的一个或多个。

[1272] 在实施方式中,免疫疾病是严重的联合免疫缺陷(SCID),Omenn综合征,并且在一方面,靶标是重组激活基因1(RAG1)或白介素7受体(IL7R)。在特定实施方式中,该疾病是运甲状腺素蛋白淀粉样变性(ATTR),家族性淀粉样蛋白性心肌病,并且在一方面,靶标是TTR

基因,包括TTR基因中的一个或多个突变。在具体实施方式中,该疾病是A-1抗胰蛋白酶缺乏症(AATD)或牵连A-1抗胰蛋白酶的另一种疾病,例如GvHD、器官移植排斥反应、糖尿病、肝病、COPD、肺气肿和囊性纤维化,目标是SERPINA1。

[1273] 在一些实施方式中,所述疾病是原发性高草酸尿症,在某些实施方式中,所述靶标包含乳酸脱氢酶A(LDHA)和羟酸氧化酶1(HAO 1)中的一种或多种。在实施方式中,所述疾病是原发性高草酸尿症1型(ph1)和其他丙氨酸-乙醛酸氨基转移酶(agxt)基因相关的病症或失调,例如腺癌、慢性酒精中毒、阿尔茨海默氏病、库利氏贫血、动脉瘤、焦虑症、哮喘、恶性乳腺肿瘤、皮肤恶性肿瘤、肾细胞癌、心血管疾病、子宫颈恶性肿瘤、冠状动脉硬化症、冠心病、糖尿病、糖尿病、非糖尿病、非胰岛素依赖型、糖尿病性肾病、子痫、湿疹、亚急性心内膜炎、胶质母细胞瘤、II型糖原贮积病、感觉神经性听力减退(疾病)、肝炎、甲型肝炎、乙型肝炎、同型半胱氨酸尿症、遗传性感觉自主神经病1型、醛固酮增多症、高胆固醇血症、高草酸尿症、原发性高草酸尿症、高血压病、肾结石、肾脏疾病、慢性肾功能衰竭、平滑肌弓形瘤、代谢性疾病、先天性代谢错误、二尖瓣脱垂综合征、心肌梗塞、肿瘤转移、肾病综合征、肥胖症、卵巢疾病、牙周炎、多囊卵巢综合征、肾脏衰竭、成人呼吸窘迫综合征、视网膜疾病、脑血管病综合征、病毒性肝炎、牙齿脱落、卵巢早衰、原发性高血压、左心室肥大、偏头痛、皮肤黑色素瘤、高血压心脏病、慢性肾小球肾炎、先兆性偏头痛、继发性高血压、急性心肌梗死、主动脉粥样硬化、过敏性哮喘、主动脉粥样硬化、过敏性哮喘皮母细胞瘤、肺恶性肿瘤、原发性高草酸尿症I型、原发性高草酸尿症2型、炎性乳腺癌、子宫颈癌、再狭窄、出血性溃疡、婴儿广义糖原贮积病、肾结石症、慢性肾移植排斥反应、尿石症、皮肤刺伤、代谢综合征X、垫神经性高血压、颈动脉粥样硬化、癌变、乳腺癌、肺癌、肾病、微白蛋白尿、家族性视网膜母细胞瘤、收缩性心力衰竭缺血性中风、左心室收缩功能不全、马尾神经节副神经节瘤、多发性肝癌、草酸钙肾结石症、Ablepharon-Macrostomia综合征、冠状动脉疾病、肝癌、慢性肾脏疾病5期、变应性鼻炎(疾病)、Crigler Najjar 综合征2型和缺血性脑血管意外。在某些实施方式中,治疗针对肝脏。在实施方式中,该基因是AGXT,其细胞遗传学位置为2q37.3,并且基因组坐标在前链的染色体2上的位置240,868,479至240,880,502。

[1274] 治疗还可以针对与胶原蛋白vii $\alpha$ 1链(col17a1)基因相关的疾病或病症,例如皮肤恶性肿瘤、鳞状细胞癌、结肠直肠肿瘤、克罗恩病、大疱性表皮松解症、腹股沟疝、瘙痒症、精神分裂症、皮肤病、遗传性皮肤病、畸胎瘤、Cockayne-Touraine病、表皮松解性大疱性肌营养不良症、结节性表皮松解性大疱、Hallopeau-Siemens病、大疱性皮肤病、胼胝体发育不良、性肌营养不良、水疱性口炎、表皮松解性大疱性皮炎指甲的皮肤和畸形、少年性肌阵挛性癫痫、食道鳞状细胞癌、金德勒皮肤癌、大疱性表皮松解性大疱性营养不良性表皮松解症(局部性)、局限性隐性营养不良性表皮松解性大疱性大疱病、全身性隐性营养不良、表皮松解性大疱、皮肤乳腺肿瘤、表皮松解性单纯性趾甲营养不良、新生儿的短暂性大疱性皮肤松解、常染色体隐性表皮松解性大肌萎缩变体和常染色体隐性表皮松解性大隐性肌萎缩。

[1275] 在实施方式中,该疾病是靶向髓母细胞瘤I(WTI)和HLA表达细胞的急性髓细胞性白血病(AML)。在实施方式中,所述疗法是T细胞疗法,如本文其他地方所述,其包括具有WTI特异性TCR的工程化T细胞。在某些实施方式中,靶标是AML中的CD157。

[1276] 在实施方式中,该疾病是血液疾病。在某些实施方式中,疾病是血友病,在一方面,靶标是因子XI。在其他实施方式中,所述疾病是血红蛋白病,例如镰状细胞病,镰状细胞性

状,血红蛋白C病,血红蛋白C性状,血红蛋白 S/C病,血红蛋白D病,血红蛋白E病,地中海贫血,与血红蛋白相关的疾病氧亲和力增加,与血红蛋白相关的疾病,氧亲和力降低,不稳定的血红蛋白疾病,高铁血红蛋白血症。也可以治疗止血和X因子和XII因子缺乏症。在一些实施方式中,靶标是BCL11A基因(例如人BCL11a基因),BCL11a 增强子(例如人BCL11a增强子)或HFPFH区(例如人HFPFH区), $\beta$ 球蛋白,胎儿血红蛋白,球蛋白基因(例如,HBG1,HBG2或HBG1和HBG2),BCL11A 基因的类胡萝卜素特异性增强子(BCL11Ae)或其组合。

[1277] 在实施方式中,靶基因座可以是RAC、TRBC1、TRBC2、CD3E、CD3G、CD3D、B2M、CIITA、CD247、HLA-A、HLA-B、HLA-C、DCK、CD52、FKBP1A、NLRC5、RFXANK、RFX5、RFXAP、NR3C1、CD274、HAVCR2、LAG3、PDCD1、PD-L2、HCF2、PAI、TFPI、PLAT、PLAU、PLG、RPOZ、F7、F8、F9、F2、F5、F7、F10、F11、F12、F13A1、F13B、STAT1、FOXP3、IL2RG、DCLRE1C、ICOS、MHC2TA、GALNS、HGSNAT、ARSB、RFXAP、CD20、CD81、TNFRSF13B、SEC23B、PKLR、IFNG、SPTB、SPTA、SLC4A1、CPOEPB42、VFW、SERPINCA1、CTLA4、CEACAM(例如CEACAM-1、CEACAM-3和/或 CEACAM-5)、VISTA、BTLA、TIGIT、LAIR1、CD160、2B4、CD80、CD86、B7-H3(CD113)、B7-H4(VTCN1)、HVEM(TNFRSF14或CD107)、KIR、A2aR、I类MHC、II类MHC、GAL9、腺苷和TGF $\beta$ 、PTPN11及其组合。在实施方式中,基因组核酸序列内的靶序列在Chr 1: 5,250,094-5,250,237,-链,hg38;和-hg38。Chr1 1:5,255,022-5,255,164,-链, HG38;非缺失HFPFH区;曲线1:5,249,833至曲线1:5,250,237-链,hg38;曲线1:5,254,738至曲线1:5,255,164,-链,hg38;Chr1 1:5,249,833-5,249,927,-链,hg3;Chr1 1:5,254,738-5,254,851,-链,HG38;Chr1 1:5,250,139-5,250,237,-链,hg38。

[1278] 在实施方式中,该疾病与高胆固醇有关,并且提供了胆固醇的调节。在一些实施方式中,调节是通过靶PCSK9中的修饰来实现的。可牵涉PCSK9 并因此成为本文所述系统和方法目标的其他疾病包括无脂蛋白血症、腺瘤、动脉硬化、动脉粥样硬化、心血管疾病、胆石症、冠状动脉硬化、冠心病、非胰岛素依赖型糖尿病、高胆固醇血症、家族性高胆固醇血症、高胰岛素血症、高脂血症、家族性合并高脂血症、低 $\beta$ 脂蛋白血症、慢性肾功能衰竭、肝病、肝脏肿瘤、黑色素瘤、心肌梗死、嗜睡症、肿瘤转移、肾母细胞瘤、肥胖症、皮脂瘤、肥胖症、黄瘤病、周围血管疾病、心肌缺血、血脂异常、葡萄糖耐量受损、黄瘤、多基因高胆固醇血症、继发于肝脏的恶性肿瘤、痴呆、超重、丙型肝炎、慢性、颈动脉粥样硬化、高脂蛋白血症H型、颅内颅内动脉粥样硬化、缺血性中风、急性冠脉综合征、主动脉钙化、心血管疾病、Iib型高脂蛋白血症、周围动脉疾病、II型家族性醛固酮过多症、家族性低脂蛋白血症、常染色体隐性高胆固醇血症、常染色体显性胆固醇血症、常染色体显性胆固醇血症事故和动脉硬化性心血管疾病NOS。在实施方式中,治疗可以靶向肝脏,即PCSK9活性的主要位置。

[1279] 在实施方式中,所述疾病或病症是Hyper IGM综合征或以缺陷的CD40 信号转导为特征的病症。在某些实施方式中,CD40L外显子的插入用于恢复适当的CD40信号传导和B细胞类别开关重组。在特定实施方式中,在例如 T细胞或造血干细胞(HSC)的细胞中,在CD40L基因的外显子2-5中的一个或多个上对CD40配体(CD40L)进行编辑。

[1280] 在实施方式中,该疾病是缺乏铁蛋白的先天性肌营养不良症(mdcmd)和其他层粘连蛋白 $\alpha$ 2(1ama2)基因相关的病症或失调。该治疗可以针对肌肉,例如骨骼肌、平滑肌和/或心肌。在某些实施方式中,靶标是层粘连蛋白, $\alpha$ 2(LAMA2),其也可以称为层粘连蛋白-12亚基 $\alpha$ 、层粘连蛋白-2亚基、层粘连蛋白4亚基 $\alpha$ 3、肌球蛋白重链、层粘连蛋白M链、LMM、先天性

肌营养不良症和肌球蛋白。LAMA2的细胞遗传学位置为6q22.33,基因组坐标位于前链第6号染色体上的128,883,141-129,516,563位置。在实施方式中,所治疗的疾病可以是缺乏肌球蛋白的先天性肌营养不良症(MDCMD)、肌萎缩性侧索硬化症、膀胱肿瘤、炭疽病、大肠癌、挛缩症、囊肿、杜氏肌营养不良症、疲劳、远视、肾血管性高血压、黑色素瘤、精神发育迟滞、肌病、肌肉营养不良、近视、肌炎、神经肌肉疾病、周围神经病变、屈光不正、精神分裂症、严重智力低下(IQ 20-34)、甲状腺肿瘤、烟草使用障碍、严重合并免疫缺陷、滑膜癌(疾病)、肿瘤进展、草莓痣、皮肤变性、小牙畸形、沃克-沃堡先天性肌营养不良、慢性牙周炎、白细胞性脑病、认知障碍、福山型先天性肌营养不良、硬化性肌营养不良肌肉营养不良、神经病、肌肉眼脑疾病、肢体肌肉萎缩症麻疹、腰带、先天性肌营养不良(疾病)、肌肉纤维化、癌症复发、耐药性癫痫、呼吸衰竭、粘液样囊肿、呼吸异常、肌营养不良、先天性肌营养不良、先天性肌营养不良、结直肠癌、部分LAMA2缺陷导致的先天性肌营养不良和常染色体显性颅干骺端发育不良。

[1281] 在某些实施方式中,靶标是AAVS1(PPP1R12C)、ALB基因、Angpt13基因、ApoC3基因、ASGR2基因、CCR5基因、FIX(F9)基因、G6PC基因、Gys2基因、HGD基因、Lp(a)基因、Pcsk9基因、Serpinal基因、TF基因和TTR基因。评估将HDR/NHEJ介导的cDNA敲入第一个外显子的效率可以利用cDNA敲入“安全港”位点,例如:单链或双链DNA与以下区域之一具有同源臂,例如:ApoC3(chr11:116829908-116833071),Angpt13(chr1:62,597,487-62,606,305),Serpinal(chr14:94376747-94390692),Lp(a)(chr6:160531483-160664259),Pcsk9(chr1:55,039,475-55,064,852),FIX(chrX:139,530,736-139,563,458),ALB(chr4:73,404,254-73,421,411),TTR(chr18:31,591,766-31,599,023),TF(chr3:133,661,997-133,779,005),G6PC(chr17:42,900,796-42,914,432),Gys2(chr12:21,536,188-21,604,857),AAVS1(PPP1R12C)(chr19:55,090,912-55,117,599),HGD(chr3:120,628,167-120,682,570),CCR5(chr3:46,370,854-46,376,206)或ASGR2(chr17:7,101,322-7,114,310)。

[1282] 一方面,靶标是可溶的超氧化物歧化酶1(SOD1),其可帮助治疗与基因相关的疾病或病症。在特定实施方式中,所述疾病或病症与SOD1相关,并且可以是例如腺癌、白蛋白尿、慢性酒精中毒、阿尔茨海默氏病、健忘症、淀粉样变性病、肌萎缩性侧索硬化症、贫血、自身免疫性溶血性贫血、镰状细胞性贫血、贫血、焦虑症、主动脉疾病、动脉硬化、类风湿性关节炎、窒息新生儿、哮喘、动脉粥样硬化、自闭症、自身免疫性疾病、巴雷特食管、白塞综合征、膀胱恶性肿瘤、乳腺恶性肿瘤、脑肿瘤、恶性肿瘤结肠癌、支气管癌、非小细胞肺癌、鳞状细胞癌、移行细胞癌、心血管疾病、颈动脉血栓形成、赘生性细胞转化、脑梗死、脑缺血、短暂性脑缺血发作、炭黑-玛丽-牙病、霍乱、结肠炎、大肠癌、冠状动脉硬化症、冠心病缓解、新型隐球菌感染、耳聋、生命中断、脱水失调、老年性痴呆、抑郁症、接触性皮炎、糖尿病、糖尿病、实验性糖尿病、胰岛素依赖型糖尿病、非胰岛素依赖型糖尿病、糖尿病性肾病、糖尿病性视网膜病、唐氏综合症、侏儒症、水肿、日本脑炎、毒性表皮坏死、颞叶癫痫症、皮疹、肌肉束、酒精性脂肪肝、胎儿发育迟缓、纤维肌痛、纤维肉瘤、破伤风X线型胶质瘤、头痛、部分听力下降、心脏骤停、心力衰竭、房间隔缺损、蠕虫病、血色素沉着症、溶血(疾病)、慢性肝炎、HIV感染、亨廷顿病、高胆固醇血症、高血糖症、增生、高血压疾病、甲状腺功能亢进症、低蛋白血症、低血压、自然低温、甲状腺功能低下、免疫缺陷综合症、免疫系统疾病、炎症、炎症性肠病、流行性感冒、肠道疾病、缺血、Kearns-Sayre综合症、圆锥角膜、肾结石、肾脏疾病、急性

肾功能衰竭、慢性肾功能衰竭、多囊炎、白血病、髓样白血病、急性早幼粒细胞白血病、肝硬化、肝病、肝肿瘤、锁定综合征、慢性阻塞性气道疾病、肺肿瘤、系统性红斑狼疮、非霍奇金淋巴瘤、马查多-约瑟夫病、疟疾、恶性肿瘤胃、动物乳腺肿瘤、马凡氏综合症、脑膜脊髓球囊扩张、精神发育迟缓、二尖瓣狭窄、获得性牙齿氟中毒、运动障碍、多发性硬化症、肌肉僵硬、肌肉痉挛、肌肉萎缩、脊髓性肌肉萎缩、肌病、真菌病、心肌梗塞再灌注损伤、坏死、肾病、肾病综合征、神经变性神经系统疾病、神经痛、神经母细胞瘤、神经瘤、神经肌肉疾病、肥胖症、职业病、眼高血压、少精子症、变性多关节炎、骨质疏松症、卵巢癌、疼痛、胰腺炎、Papillon-Lefevre病、麻痹、帕金森病、苯丙氨酸尿毒症疾病、先兆子痫、前列腺肿瘤、蛋白质缺乏症、蛋白尿、牛皮癣、肺纤维化、肾动脉阻塞、再灌注损伤、视网膜变性、视网膜疾病、视网膜母细胞瘤、血吸虫病、曼氏血吸虫病、曼氏血吸虫病、精神分裂症、羊搔痒症、年龄脊髓受压、脑血管意外、蛛网膜下腔出血、进行性核上性麻痹、破伤风、三体性、特纳综合征、单极抑郁症、荨麻疹、白癜风、声带麻痹、肠扭转、体重增加、HMN (遗传性运动神经病) 近端I型、运动神经元疾病、神经原纤维变性 (形态逻辑异常)、烧灼感、冷漠、情绪波动、滑膜囊肿、白内障、偏头痛、坐骨神经病、感觉神经病、皮肤萎缩、肌肉无力、食管癌、舌面部颊运动障碍、特发性肺动脉高压、先兆性偏头痛、混合性传导-感觉神经性听力减退、缺铁性贫血、营养不良、朊病毒疾病、线粒体肌病、MELAS综合征、慢性进行性眼外肌麻痹、全身性麻痹、早衰综合征、颤动、精神病症状、记忆力减退、肌肉退化、神经系统症状、胃出血、胰腺癌、脑部疾病、肝纤维化、肺恶性肿瘤、与年龄有关的黄斑变性、帕金森氏病、疾病进展、低血钾症、细胞色素C氧化酶缺乏症、原发性震颤、家族性运动神经元疾病、下运动神经元疾病、变性性脊髓病、糖尿病多发性神经病、肝和肝内胆道癌波斯湾综合症、老年斑、萎缩性、额颞痴呆、语义性痴呆、偏头痛、认知障碍、肝脏恶性肿瘤、胰腺恶性肿瘤、前列腺恶性肿瘤、纯自主神经衰竭、运动症状、痉挛、痴呆、神经退行性疾病、关岛慢性丙型肝炎、肌萎缩性侧索硬化症、四肢僵硬、多系统疾病、头皮脱落、前列腺癌、肝肺综合征、桥本病、进行性肿瘤病、乳腺癌、末期疾病、肺癌、迟发性运动障碍、淋巴结继发性恶性肿瘤、结肠癌、胃癌、中枢神经母细胞瘤、胸主动脉夹层动脉瘤、糖尿病性黄斑水肿、微白蛋白尿、大脑中动脉闭塞、大脑中动脉梗塞、上运动神经元体征、额颞叶大叶变性、记忆力减退、古典苯丙酮酸血症步态障碍、2型脊髓小脑性共济失调、路易体病、肌肉萎缩、脊髓小球、21号染色体单一性、血小板增多症、皮肤上的斑点、药物诱发的肝损伤、遗传性脑白质萎缩症、脑缺血、卵巢肿瘤、颅骨病变、大血管病变、持续性肺动脉高压、卵巢恶性肿瘤、黏液样囊肿、玻璃疣、肉瘤、体重减轻、重度抑郁症、轻度认知障碍、退化性疾病、部分三体症、心血管疾病、听力障碍、认知障碍、输尿管结石、乳腺疾病肿瘤、大肠癌、慢性肾脏病、最小变化肾病综合征、非赘生性疾病、X连锁球脑脊髓萎缩、乳腺X线密度、正常张力性青光眼易感性、白癜风相关的多种自身免疫性疾病易感性1 (发现)、肌萎缩性侧索硬化症和/或额颞叶性痴呆1、肌萎缩性侧索硬化症1、偶发性肌萎缩性侧索硬化症、单体性肌萎缩症、冠状动脉疾病、转化的偏头痛、返流、尿道上皮癌、运动障碍、肝癌、蛋白歧义失调、T43早幼粒细胞白血病、体重增加不利事件、线粒体细胞病变、特发性肺动脉高压、进行性cGVHD、感染、GRN相关额颞叶痴呆、线粒体病理和听力减退。

[1283] 在特定的实施方式中,该疾病与基因ATXN1、ATXN2或ATXN3相关,其可以靶向治疗。在一些实施方式中,靶向位于ATXN1的外显子8、ATXN2 的外显子1或ATXN3的外显子10的CAG重复区。在实施方式中,该疾病是脊髓小脑共济失调3 (sca3)、sca1或sca2和其他相关疾病,

例如先天性异常、阿尔茨海默氏病、肌萎缩性侧索硬化症、共济失调、共济失调毛细血管扩张症、小脑共济失调、小脑疾病、舞蹈病、Cho囊性纤维化、精神抑郁症、抑郁症、肌张力障碍、食道肿瘤、外斜视、心脏骤停、亨廷顿病、马查多-约瑟夫病、运动障碍、肌肉营养不良、肌强直性营养不良、嗜睡症、神经退行性变、神经母细胞瘤、帕金森病、无周围神经病变腿综合征、视网膜变性、色素性视网膜炎、精神分裂症、胆怯综合征、睡眠障碍、遗传性痉挛性截瘫、血栓栓塞、硬人综合征、脊髓小脑共济失调、食道癌、多发性神经病、热感、紫杉醇征、肌肉抽搐、肌肉抽搐神经系统症状、脑萎缩、帕金森病、蛋白质 S 缺乏症、Cer小脑变性、家族性淀粉样神经病、葡萄牙型、痉挛综合征、垂直眼球震颤、眼球震颤终位、抗凝血酶III缺乏症、萎缩症、复杂的遗传性痉挛性截瘫、多系统萎缩、帕卢迪洛斯变性、肌张力障碍、纯正自主神经功能衰竭、血栓形成缺乏、先天性肌强直性营养不良、运动症状、神经病、神经退行性疾病、食道恶性肿瘤、视力障碍、激活的C抵抗力、绝症、肌强直、中枢神经母细胞瘤、失眠、阑尾共济失调、嗜睡症-多发性综合征、Machado-Joseph病I、Machado-Joseph疾病II型、Machado-Joseph疾病III型、齿状核红核苍白球丘脑下部核萎缩、步态共济失调、脊髓小脑共济失调1型、脊髓小脑共济失调2型、脊髓小脑共济失调6型(失调)、脊髓小脑共济失调7型、肌肉脊髓小球型、基因组不稳定性、情节性共济失调2型(疾病)、X线脊髓萎缩症、脆性X震颤/共济失调综合征、因激活的C蛋白抗性导致的血栓形成性疾病(失调)、肌萎缩性侧索硬化症、神经元核内包涵体疾病、遗传性抗凝血酶II缺乏症和迟发性帕金森病。

[1284] 在实施方式中,该疾病与肿瘤抗原-癌症或非癌症相关的适应症的表达有关,例如急性淋巴白血病、弥漫性大B细胞淋巴瘤、滤泡性淋巴瘤、慢性淋巴细胞性白血病、霍奇金淋巴瘤、非霍奇金淋巴瘤。在实施方式中,靶标可以是TET2内含子,TET2内含子-外显子连接,chr4基因组区域内的序列。

[1285] 在实施方式中,可以治疗神经退行性疾病。在特定实施方式中,靶标是突触核蛋白(SNCA)。在某些实施方式中,所治疗的疾病是与疼痛有关的疾病,包括先天性疼痛不敏感、压迫性神经病、阵发性极端疼痛疾病、高级房室传导阻滞、小纤维神经病和家族性发作性疼痛综合征2。在某些实施方式中,靶标是钠通道、电压门控、X型 $\alpha$ 子单元(SCN10A)。

[1286] 在某些实施方式中,造血干细胞和祖干细胞被编辑,包括敲入。在特定的实施方式中,敲入用于治疗溶酶体贮积病,糖原贮积病,粘多糖或多糖或蛋白质分泌将改善该疾病的任何疾病。在一实施方式中,该疾病是镰状细胞病(SCD)。在另一个实施方式中,该疾病是 $\beta$ -地中海贫血。

[1287] 在某些实施方式中,T细胞或NK细胞用于癌症治疗,并可包括包含重组受体(例如CAR)和一种或多种选自CCR7<sup>+</sup>、4-1BB<sup>+</sup>(CD137<sup>+</sup>)、TIM3<sup>+</sup>、CD27<sup>+</sup>、CD62L<sup>+</sup>、CD127<sup>+</sup>、CD45RA<sup>+</sup>、D45RO<sup>-</sup>、t-bet1<sup>w</sup>、IL-7Ra<sup>+</sup>、CD95<sup>+</sup>、IL-2RP<sup>+</sup>、CXCR3<sup>+</sup>或LFA-1<sup>+</sup>的表型标志物的T细胞。在某些实施方式中,用于癌免疫疗法的T细胞的编辑包括改变一个或多个T细胞表达的基因,例如FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的一个或多个。在一些实施方式中,编辑包括引入到CBLB靶位点中或附近的改变,以减少T细胞中的CBLB基因表达以治疗增生性疾病,并且可以包括在一个或多个CBLB靶位点处的较大插入或缺失。TGFB2靶序列的T细胞编辑可以例如位于TGFB2基因的外显子3、4或5中,并用于癌症和淋巴瘤治疗。

[1288] 可以编辑用于移植的细胞,并且可以包括对细胞的一种或多种免疫原性基因(例如HLA基因)进行等位基因特异性修饰,例如HLA-A、HLA-B、HLA-C、HLA-DRB1、HLA-DRB3/4/

5、HLA-DQ和HLA-DP MiHA以及任何其他I类或II类MHC基因或基因座,其中可能包括将一个或多个匹配的受体HLA等位基因递送至原始位置找到一个或多个不匹配的供体HLA等位基因,并且可以包括将一个或多个匹配的受体HLA等位基因插入“安全港”基因座。在一个实施方式中,该方法进一步包括在基因中引入用于体内选择的化疗抗性基因。

[1289] 方法和系统可以靶向肌营养不良症肌强直蛋白激酶(DMPK)进行编辑,在特定实施方式中,靶标是DMPK基因的3'非翻译区(UTR)中的CTG三核苷酸重复。与DMPK相关的疾病或疾病包括动脉粥样硬化、无精子症、肥厚型心肌病、乳糜泻、先天性染色体疾病、糖尿病、局灶性肾小球硬化症、亨廷顿病、性腺机能减退、肌肉萎缩、肌病、肌营养不良、肌强直、肌萎缩性肌营养不良、轻瘫、精神分裂症、白内障、脊髓小脑性共济失调、肌肉无力、肾上腺皮质营养不良、中心核肌病、间质纤维化、强直性肌营养不良、精神异常、X连锁性Charcot-Marie-Tooth病、先天性强直性营养不良、双侧性白内障先天性纤维类型失调、肌强直性肌营养不良、多系统障碍、3-甲基谷氨酸酸尿症3型、心脏事件、心源性晕厥、先天性结构性肌病、精神障碍、肾上腺神经病、肌营养不良症2和智力障碍。

[1290] 在实施方式中,该疾病是先天性代谢错误。该疾病可以选自碳水化合物代谢紊乱(糖原贮积病、G6PD缺乏症),氨基酸代谢紊乱(苯酮尿症、枫糖浆尿病、戊二酸血症1型),尿素循环障碍或尿素循环缺陷(氨基甲酰基磷酸合成酶I缺乏症),有机酸代谢紊乱(链蛋白尿症,2-羟基戊二酸尿症),脂肪酸氧化/线粒体代谢紊乱(中链酰基辅酶A脱氢酶缺乏症),卟啉代谢紊乱(急性间歇性卟啉症),嘌呤紊乱/嘧啶代谢(Lesch-Nyhan综合征),类固醇代谢障碍(类脂性先天性肾上腺增生,先天性肾上腺增生),线粒体功能障碍(Kearns-Sayre综合征),过氧化物酶体功能障碍(Zellweger综合征)或溶酶体贮积症(高雪氏病、尼曼-皮克病)。

[1291] 在实施方式中,靶标可包含重组激活基因1(RAG1)、BCL11A、PCSK9、层粘连蛋白、 $\alpha 2$ ( $\text{lama2}$ )、ATXN3、丙氨酸-乙醛酸转氨酶(AGXT)、VII型胶原VII链(COL7a1)、脊髓小脑共济失调1型蛋白(ATXN1)、类血管生成素3(ANGPTL3)、Frataxin(FXN)、超氧化物歧化酶1、可溶性(SOD1)、突触核蛋白、 $\alpha$ (SNCA)、钠通道、电压门控、X型A亚基(SCN10A)、脊髓小脑共济失调2型蛋白(ATXN2)、肌营养不良症肌强直蛋白激酶(DMPK)、11号染色体上的 $\beta$ 球蛋白基因座、中链脂肪酸的酰基辅酶A脱氢酶(ACADM)、长链3-羟基辅酶A脱氢酶长链脂肪酸(HADHA)、超长链脂肪酸的酰基辅酶A脱氢酶(ACADVL)、载脂蛋白C3(APOCIII)、运甲状腺素蛋白(TTR)、血管生成素样4(ANGPTL4)、钠电压门控通道A亚基9(SCN9A)、白介素7受体(IL7R)、葡萄糖6磷酸酶、催化(G6PC)、遗传性血色素沉着病(HFE)、SERPINA1、C9ORF72、 $\beta$ -珠蛋白、肌营养不良蛋白、 $\gamma$ -珠蛋白。

[1292] 在某些实施方式中,所述疾病或病症与载脂蛋白C3(APOCIII)有关,所述载脂蛋白C3可以被靶向用于编辑。在实施方式中,所述疾病或病症可以是血脂异常、2型高脂蛋白血症、狼疮肾炎、威尔姆斯肿瘤5、病态肥胖和生精、青光眼、糖尿病性视网膜病、关节炎、肾功能不全胆汁淤积综合症、认知障碍、对心肌梗塞的反应改变、G对甘油三酸酯生物合成过程、肾功能不全、慢性、高脂血症、慢性肾脏衰竭、载脂蛋白C-III缺乏症、冠心病、新生儿糖尿病、新生儿、先天性甲状腺功能减退、高胆固醇血症常染色体显性3、高脂蛋白血症、III型高脂血症的积极调节疾病、肾动脉阻塞、代谢综合征X、高脂血症、家族性合并症、胰岛素抵抗、短暂性婴儿高甘油三酯血症、糖尿病性肾病、糖尿病(1型)、有或没有眼部异常的肾病综合

征5和肾综合征出血热。

[1293] 在某些实施方式中,靶标是血管生成素样4 (ANGPTL4)。可以治疗的与 ANGPTL4相关的疾病或紊乱包括ANGPTL4与血脂异常、血浆甘油三酯水平低、血管生成调节剂和调节肿瘤发生以及严重的糖尿病性视网膜病、增生性糖尿病视网膜病变和非增生性糖尿病视网膜病变相关。

[1294] 在实施方式中,编辑可用于治疗脂肪酸疾病。在某些实施方式中,靶标是ACADM, HADHA, ACADVL中的一种或多种。在实施方式中,靶向编辑是选自中链脂肪酸的酰基辅酶A脱氢酶 (ACADM) 基因,长链脂肪酸的长链3-羟基辅酶A脱氢酶 (HADHA) 基因和超长链脂肪酸 (ACADVL) 基因的酰基辅酶A脱氢酶。一方面,该疾病是中链酰基辅酶A脱氢酶缺乏症 (MCADD),长链3-羟基辅酶A脱氢酶缺乏症 (LCHADD) 和/或极长链酰基辅酶A脱氢酶缺乏症 (VLCADD)。

[1295] 免疫正交直系同源物

[1296] 在一些实施方式中,当需要在受试者中表达或施用CRISPR酶时,可以通过向受试者顺序表达或施用CRISPR酶的免疫正交直系同源物来降低 CRISPR酶的免疫原性。如本文所用,术语“免疫正交直系同源物”是指具有相似或基本相同的功能或活性,但是与彼此产生的免疫反应没有或具有低交叉反应性的直系同源蛋白质。在一些实施方式中,此类直系同源物的顺序表达或施用引起低或无继发免疫应答。免疫正交直系同源物可以避免被抗体(例如在直系同源物表达或施用之前宿主中存在的抗体)中和。表达直系同源物的细胞可以避免被宿主的免疫系统清除(例如被激活的CTL清除)。在一些实例中,来自不同物种的CRISPR酶直系同源物可以是免疫正交直系同源物。

[1297] 免疫正交直系同源物可以通过分析一组候选直系同源物的序列,结构和 /或免疫原性来鉴定。在示例方法中,可以通过以下方法来鉴定一组免疫正交直系同源物:a)比较一组候选直系同源物(例如,来自不同物种的直系同源物)的序列,以鉴定具有低序列相似性或无序列相似性的候选子集;b)评估候选者子集中各成员之间的免疫重叠,以识别没有免疫重叠或免疫重叠低的候选者。在某些情况下,可以通过确定候选直系同源物与宿主的MHC(例如I型和/或MHC II型)之间的结合(例如亲和力)来评估候选者之间的免疫重叠。替代地或另外地,可以通过确定候选直系同源物的B细胞表位来评估候选者之间的免疫重叠。在一个实例中,可以使用Moreno AM等于2018年1月10日在线发表于doi:doi.org/10.1101/245985上的BioRxiv中描述的方法鉴定免疫正交直系同源物。

[1298] 患者特异性筛查方法

[1299] 靶向RNA的核酸靶向系统可用于筛选患者或患者样品中是否存在特定 RNA。

[1300] 转录物检测方法

[1301] 本发明的效应子蛋白和系统可用于特异性检测细胞或其他样品中的 RNA。在存在目标RNA靶标的情况下,引导依赖性CRISPR-Cas核酸酶活性可能伴随着针对附带靶标的非特异性RNA酶活性。为了利用RNase活性,所需的只是可检测地裂解的报告底物。例如,报道分子可以包含RNA,其在一端标记有荧光报道分子 (fluor),而在另一端标记有淬灭剂。在没有 CRISPR-Cas RNase活性的情况下,淬灭剂的物理接近性将荧光从荧光抑制到低水平。当通过存在感兴趣的RNA靶标和合适的引导RNA激活 CRISPR-Cas靶标特异性切割时,含RNA的报告分子被非特异性切割,并且荧光和淬灭剂在空间上分离。当被适当波长的光激发时,这

会导致荧光发出可检测的信号。在一种示例性测定方法中,将CRISPR-Cas效应子,目标靶标特异性引导RNA和报道分子添加到细胞样品中。荧光增加表明存在感兴趣的RNA靶标。在另一示例性方法中,提供了一种检测阵列。阵列的每个位置均配有CRISPR-Cas效应子,报道分子和目标靶标特异性引导RNA。取决于要进行的测定,在阵列的每个位置处的目标靶标特异性引导RNA可以相同,不同或组合。例如,当需要测试单个来源样品中的一个或多个目标时,可能会提供不同的目标感兴趣的特定引导RNA。可能在每个位置提供相同的目标特异性靶引导RNA,例如,当需要为同一目标测试多个样品时。

[1302] 在某些实施方式中,CRISPR-Cas在体外系统或细胞中瞬时或稳定地提供或表达,并靶向或触发非特异性切割细胞核酸。在一个实施方式中,将CRISPR-Cas工程化以敲低ssDNA,例如病毒ssDNA。在另一个实施方式中,将CRISPR-Cas工程化以敲低RNA。可以设计该系统,使得敲除依赖于细胞或体外系统中存在的靶DNA,或通过向该系统或细胞中添加靶核酸来触发。

[1303] 在一个实施方式中,将CRISPR-Cas系统工程化以非特异性地切割细胞的亚群中的RNA,所述细胞的亚群可通过异常DNA序列的存在而区分,例如,异常DNA的切割可能是不完全的或无效的。在一个非限制性实例中,靶向存在于癌细胞中并驱动细胞转化的DNA易位。经历染色体DNA修复的细胞亚群可以存活,而非特异性的核糖核酸酶活性则有利地导致潜在存活者的细胞死亡。

[1304] 应用的额外方面

[1305] 本发明在例如基因治疗、药物筛选、疾病诊断和预后方面具有广泛的应用。

[1306] 术语“多核苷酸”,“核苷酸”,“核苷酸序列”,“核酸”和“寡核苷酸”可互换使用。它们是指任何长度的核苷酸的聚合形式,即脱氧核糖核苷酸或核糖核苷酸,或其类似物。多核苷酸可以具有任何三维结构,并且可以执行任何已知或未知的功能。以下是多核苷酸的非限制性实例:基因或基因片段的编码或非编码区、从连锁分析中定义的基因座(基因座)、外显子、内含子、信使RNA(mRNA)、转移RNA、核糖体RNA、短干扰RNA(siRNA)、短发夹RNA(shRNA)、微小RNA(miRNA)、核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离DNA、任何序列的分离RNA、核酸探针和底物。该术语还包括具有合成主链的核酸样结构,参见例如Eckstein, 1991;Baserga等,1992;Milligan,1993;WO 97/03211;WO 96/39154;Mata, 1997;Strauss-Soukup,1997;和Samstag,1996。多核苷酸可以包含一个或多个修饰的核苷酸,例如甲基化的核苷酸和核苷酸类似物。如果存在的话,可以在聚合物组装之前或之后对核苷酸结构进行修饰。核苷酸的序列可以被非核苷酸组分打断。多核苷酸可在聚合后进一步修饰,例如通过与标记组分缀合。如本文所用,术语“野生型”是本领域技术人员理解的术语,是指与突变体或变体形式不同的自然存在的生物,菌株,基因或特征的典型形式。“野生型”可以是基线。如本文所使用的,术语“变体”应被理解为表示具有与自然界中所发生的模式不同的样式的品质的展现。术语“非自然发生”或“工程化”可互换使用,表示人为介入。当提及核酸分子或多肽时,该术语是指该核酸分子或多肽至少基本上不含至少一种其他组分,它们在自然界中是自然缔合的,以及在自然界中发现的。“互补性”是指核酸通过传统的Watson-Crick碱基配对或其他非传统类型与另一核酸序列形成氢键的能力。互补性百分数表示可以与第二个核酸序列(例如10个中的5个、6个、7个、8个、9个、10个)形成氢键(例如Watson-Crick碱基配对)的核酸分子中的残基百分比是50%,60%,70%,

80%、90%和100%互补)。“完全互补”是指核酸序列的所有连续残基将与第二核酸序列中相同数目的连续残基氢键键合。如本文所用,“基本上互补”是指在8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50个或更多个核苷酸的区域上至少60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%的互补度,或指在严格条件下杂交的两个核酸。如本文所用,用于杂交的“严格条件”是指与靶序列具有互补性的核酸主要与靶序列杂交而基本上不与非靶序列杂交的条件。严格条件通常是序列依赖性的,并且取决于许多因素而变化。通常,序列越长,序列与其靶序列特异性杂交的温度越高。严格条件的非限制性实例详细描述于Tijssen(1993), *Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter* “Overview of principles of hybridization and the strategy of nucleic acid probe assay”, Elsevier, N.Y. 之中。当提及多核苷酸序列时,那么也设想了互补或部分互补的序列。这些优选能够在高度严格的条件下与参考序列杂交。通常,为了最大化杂交速率,选择相对严格的杂交条件:比热熔点 ( $T_m$ ) 低约20至25°C。 $T_m$ 是在定义的离子强度和pH下溶液中50%的特定靶序列与完美互补的探针杂交时的温度。通常,为了要求杂交序列具有至少约85%的核苷酸互补性,将高度严格的洗涤条件选择为比 $T_m$ 低约5至15°C。为了要求杂交序列的至少约70%的核苷酸互补性,将中等严格的洗涤条件选择为比 $T_m$ 低约15至30°C。高度允许(非常低的严格性)洗涤条件可能比 $T_m$ 低50°C,从而允许杂交序列之间的高度错配。本领域技术人员将认识到,杂交和洗涤阶段中的其他物理和化学参数也可以被改变以从靶序列和探针序列之间的特定同源性水平影响可检测到的杂交信号的结果。优选的高度严格条件包括在42°C下于50%甲酰胺、5xSSC和1%SDS中孵育,或在65°C在5xSSC和1%SDS中孵育,并在0.2xSSC和0.1%SDS中洗涤。65°C。“杂交”是指一种或多种多核苷酸反应形成复合物的反应,该复合物通过核苷酸残基的碱基之间的氢键而稳定化。氢键可以通过Watson Crick碱基配对, Hoogsteen结合或以任何其他序列特异性方式发生。该复合物可以包括形成双链体结构的两条链,形成多链复合物的三条或更多条链,一条自杂交链,或这些的任何组合。杂交反应可以构成更广泛过程中的一个步骤,例如PCR的启动或酶对多核苷酸的切割。能够与给定序列杂交的序列称为给定序列的“补体”。如本文所用,术语“基因组基因座”或“基因座”(多个基因座)是基因或DNA序列在染色体上的特定位置。“基因”是指编码多肽或RNA链的DNA或RNA片段,该片段具有在生物体中发挥功能的功能,因此是生物体中遗传的分子单元。为了本发明的目的,可以认为基因包括调节基因产物产生的区域,无论这些调节序列是否与编码和/或转录序列相邻。因此,基因包括但不限于启动子序列,终止子,翻译调节序列,如核糖体结合位点和内部核糖体进入位点,增强子,沉默子,绝缘子,边界元件,复制起点,基质附着位点和基因座控制区域。如本文所用,“基因组基因座的表达”或“基因表达”是将来自基因的信息用于功能性基因产物的合成的过程。基因表达的产物通常是蛋白质,但是在非蛋白质编码基因(例如rRNA基因或tRNA基因)中,产物是功能性RNA。基因表达的过程被所有已知的生命-真核生物(包括多细胞生物),原核生物(细菌和古细菌)和病毒所利用,以生成能够存活的功能性产物。如本文所用,基因或核酸的“表达”不仅涵盖细胞基因表达,而且涵盖在克隆系统中以及在任何其他情况下核酸的转录和翻译。如本文所用,“表达”还指从DNA模板转录多核苷酸的过程(例如转化为mRNA或其他RNA转录物)和/或随后将转录的mRNA翻译为肽,多肽的过程。或蛋白质。转录

物和编码的多肽可以统称为“基因产物”。如果多核苷酸衍生自基因组DNA,则表达可包括在真核细胞中剪接mRNA。术语“多肽”,“肽”和“蛋白质”在本文可互换使用,是指任何长度的氨基酸的聚合物。该聚合物可以是直链或支链的,它可以包含修饰的氨基酸,并且可以被非氨基酸中断。该术语还涵盖已被修饰的氨基酸聚合物。例如,二硫键的形成,糖基化,脂化,乙酰化,磷酸化或任何其他操作,例如与标记组分的缀合。如本文所用,术语“氨基酸”包括天然和/或非天然或合成的氨基酸,包括甘氨酸和D或L旋光异构体,以及氨基酸类似物和拟肽。如本文所用,术语“结构域”或“蛋白质结构域”是指可以独立于蛋白质链的其余部分而存在和起作用的蛋白质序列的一部分。如本发明的方面所述,序列同一性与序列同源性有关。同源性比较可以通过眼睛进行,或更通常地,借助于容易获得的序列比较程序进行。这些可商购的计算机程序可以计算两个或更多个序列之间的同源性百分比(%),还可以计算两个或更多个氨基酸或核酸序列共有的序列同一性。

[1307] [1]如本文所用,术语“野生型”是本领域技术人员理解的术语,是指与突变体或变体形式不同的自然存在的生物,菌株,基因或特征的典型形式。“野生型”可以是基线。

[1308] 如本文所使用的,术语“变体”应被理解为表示具有与自然界中所发生的模式不同的样式的品质的展现。术语“非自然发生”或“工程化”可互换使用,表示人为介入。当提及核酸分子或多肽时,该术语是指该核酸分子或多肽至少基本上不含至少一种其他组分,它们在自然界中是自然缔合的,以及在自然界中发现的。在所有方面和实施例中,无论它们是否包括这些术语,应理解,优选地,其可以是任选的,因此优选地包括或不优选地不包括。此外,术语“非天然存在的”和“工程化的”可以互换使用,因此可以单独使用或组合使用,并且一个或另一个可以代替两者的提及。特别地,“工程化的”代替“非自然发生的”或“非自然发生的和/或工程化的”是优选的。

[1309] 序列同源性可以通过本领域已知的许多计算机程序中的任何一种来生成,例如BLAST或FASTA等。用于进行这种比对的合适的计算机程序是 GCG Wisconsin Bestfit软件包(美国威斯康星大学;Devereux)。等人,1984,核酸研究12:387)。除可以执行序列比较之外的其他软件的示例包括但不限于BLAST软件包(请参见Ausubel等,1999同上,第18章),FASTA(Atschul 等,1990,J.Mol.Biol.,403-410)和GENEWORKS比较工具套件。BLAST 和FASTA均可用于离线和在线搜索(请参见Ausubel等,1999同上,第7-58 至7-60页)。但是,最好使用GCG Bestfit程序。可以在连续序列上计算百分比(%)序列同源性,即,一个序列与另一序列比对,一个序列中的每个氨基酸或核苷酸直接与另一序列中相应的氨基酸或核苷酸进行比较,一个残基为一个时间。这称为“空位”对齐。通常,仅在相对短数量的残基上进行这种无缺口的比对。尽管这是一种非常简单且一致的方法,但是它没有考虑到例如在其他相同的序列对中,一个插入或缺失可能导致以下氨基酸残基失配,从而潜在地导致当进行整体比对时,同源性百分比大大降低。因此,大多数序列比较方法被设计为产生最佳比对,该比对考虑了可能的插入和缺失而不会过度惩罚总体同源性或同一性得分。这是通过在序列比对中插入“缺口”以尝试最大化局部同源性或同一性来实现的。但是,这些更复杂的方法将“空位罚分”分配给比对中出现的每个空位,以便对于相同数目的相同氨基酸,以尽可能少的空位进行序列比对-反映了两个比较序列之间更高的相关性-可能会获得比许多差距更高的分数。通常使用“亲和缺口成本”,其对于缺口的存在收取相对较高的成本,并对缺口中的每个后续残基收取较小的损失。这是最常用的差距评分系统。当然,高空位罚分

可以产生具有较少空位的优化比对。大多数比对程序都允许修改空位罚分。但是,在使用此类软件进行序列比较时,最好使用默认值。例如,当使用GCG Wisconsin Bestfit软件包时,氨基酸序列的默认空位罚分对于空位是-12,对于每个延伸是-4。因此,最大同源性百分比的计算首先需要考虑缺口罚分的情况下产生最佳比对。用于进行这种比对的合适的计算机程序是GCG Wisconsin Bestfit软件包(Devereux 等,1984Nuc.Acids Research 12p387)。除了可以进行序列比较之外的其他软件的实例包括但不限于BLAST程序包(参见Ausubel等人,1999Short Protocols in Molecular Biology,第四版-第18章),FASTA(Altschul等,1990J. Mol.Biol.403-410)和GENEWORKS比较工具套件。BLAST和FASTA都可用于脱机和在线搜索(参见Ausubel等,1999,Short Protocols in Molecular Biology,pages 7-58至7-60)。但是,对于某些应用程序,最好使用GCG Bestfit 程序。还提供了一种称为BLAST 2Sequences的新工具来比较蛋白质和核苷酸序列(参见FEMS Microbiol Lett.1999 174 (2):247-50;FEMS Microbiol Lett. 1999 177 (1):187-8和美国国家卫生研究院网站上的国家生物技术中心信息网站)。尽管可以根据同一性来衡量最终的同源性百分比,但比对过程本身通常不是基于全有或全无对比较。取而代之的是,通常使用缩放的相似性得分矩阵,该矩阵基于化学相似性或进化距离为每个成对比较分配得分。通常使用的这种矩阵的一个示例是BLOSUM62矩阵-BLAST程序套件的默认矩阵。如果提供的话,GCG威斯康星州程序通常使用公用默认值或自定义符号比较表(有关更多详细信息,请参见用户手册)。对于某些应用程序,最好使用GCG 包的公共默认值,或者在其他软件的情况下,使用默认矩阵,例如BLOSUM62。可替代地,可以基于类似于CLUSTAL(Higgins DG&Sharp PM (1988),Gene 73 (1),237-244)的算法,使用DNASISTM(Hitachi Software)中的多重比对特征来计算同源性百分比。一旦软件产生了最佳比对,就可以计算同源性百分比,最好是序列同一性百分比。该软件通常将其作为序列比较的一部分,并生成数值结果。该序列还可以具有氨基酸残基的缺失,插入或取代,其产生沉默变化并产生功能上等同的物质。可以根据氨基酸特性(例如残基的极性,电荷,溶解度,疏水性,亲水性和/或两亲性质)的相似性来进行有意的氨基酸取代,因此将氨基酸一起分组功能组。氨基酸可仅基于其侧链的性质而分组在一起。但是,也包含突变数据会更有用。出于结构原因,由此衍生的氨基酸组可能是保守的。这些集合可以以维恩图的形式描述(Livingstone C.D.和Barton G.J.(1993)“Protein sequence alignments:a strategy for the hierarchical analysis of residue conservation”Comput.Appl.Biosci.9: 745-756)(Taylor W.R.(1986)“The classification of amino acid conservation”J. Theor.Biol.119;205-218)。可以例如根据表7制备保守性,该表7描述了公认的氨基酸的维恩图分组。

[1310] 术语“受试者”、“个体”和“患者”在本文中可互换使用,是指脊椎动物,优选为哺乳动物,更优选为人。哺乳动物包括但不限于鼠类,猿猴,人类,农场动物,运动动物和宠物。还包括体内获得或体外培养的生物实体的组织,细胞及其后代。

[1311] 术语“治疗剂”,“可治疗剂”或“治疗剂”可互换使用,并且是指在施用于受试者时赋予一些有益作用的分子或化合物。有益效果包括启用诊断确定;改善疾病,症状,病症或病理状况;减少或预防疾病,症状,病症或病状的发作;并且通常可以抵抗疾病,症状,疾病或病理状况。如本文所用,“治疗”或“疗法”或“缓解”或“改善”可互换使用。这些术语是指获得有益或期望结果的方法,包括但不限于治疗益处和/或预防益处。治疗益处是指对一种或

多种治疗中的疾病,病症或症状的任何与治疗有关的改善或作用。为了预防有益,可以将所述组合物施用于有发展为特定疾病,病症或症状的风险的受试者,或施用于报告疾病的一种或多种生理症状的受试者,即使该疾病,病症或症状也是如此。可能尚未显现。术语“有效量”或“治疗有效量”是指足以实现有益或期望结果的试剂的量。治疗有效量可以根据以下一项或多项而变化:受试者和所治疗的疾病状况,受试者的体重和年龄,疾病状况的严重程度,给药方式等,可以通过以下方式容易地确定:本领域普通技术之一。该术语还适用于将提供图像以通过本文所述的任何一种成像方法进行检测的剂量。具体剂量可以根据以下一项或多项而变化:所选的特定试剂,要遵循的给药方案,是否与其他化合物联合给药,给药时间,要成像的组织以及体内的物理递送系统。它是随身携带的。

[1312] 除非另有说明,否则本发明的实施采用免疫学,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的常规技术,它们在本领域技术范围内。参见 Sambrook, Fritsch 和 Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F.M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow 和 Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney 编辑 (1987))。本发明的几个方面涉及包含一个或多个载体,或这样的载体的载体系统。可设计载体以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如大肠杆菌,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。或者,重组表达载体可以在体外转录和翻译,例如使用 T7 启动子调节序列和 T7 聚合酶。本发明的实施方式包括可能包含同源取代的序列(多核苷酸或多肽)(取代和置换在本文中均用于表示现有氨基酸残基或核苷酸与替代残基或核苷酸的互换),即在氨基酸的情况下,类似的取代,如碱性的对碱性,酸性的对酸性,极性的对极性等,也可以发生非同源的取代,即从一类残基到另一类残基,或者涉及包括非天然氨基酸,例如鸟氨酸(以下称为 Z),二氨基丁酸鸟氨酸(以下称为 B),正亮氨酸鸟氨酸(以下称为 O),吡啶基丙氨酸,噻吩丙氨酸,萘基丙氨酸和苯基甘氨酸。变体氨基酸序列可以包括合适的间隔基,其可以插入序列的任何两个氨基酸残基之间,除了氨基酸间隔基例如甘氨酸或  $\beta$ -丙氨酸残基外,所述烷基还包括烷基例如甲基,乙基或丙基。本领域技术人员可以很好地理解另一种形式的变异,其涉及类肽形式的一个或多个氨基酸残基的存在。为避免疑问,“类肽形式”用于表示变异的氨基酸残基,其中  $\alpha$ -碳取代基位于残基的氮原子上,而不是  $\alpha$ -碳。类肽形式的肽的制备方法是本领域已知的,例如 Simon RJ 等, PNAS (1992) 89 (20), 9367-9371, 和 Horwell DC, Trends Biotechnol. (1995) 13 (4), 132-134。同源性建模:其他 CRISPR-Cas 直系同源物中的相应残基可通过 Zhang 等, 2012 (Nature; 490 (7421): 556-60) 和 Chen 等, 2015 (PLoS Comput Biol; 11 (5): e1004248) 的方法(计算蛋白-蛋白相互作用 (PPI) 的方法,其用于预测域-基序界面介导的相互作用) 鉴定。PrePPI(预测 PPI) 是一种基于结构的 PPI 预测方法,它使用贝叶斯统计框架将结构证据与非结构证据结合在一起。该方法涉及获取一对查询蛋白,并使用结构比对来鉴定与其实验确定的结构或同源性模型相对应的结构代表。通过考虑整体和局部

几何关系,结构对齐还可用于识别近邻和远邻结构邻居。每当结构代表的两个邻居形成蛋白质数据库中报告的复合物时,这就会定义一个用于对两个查询蛋白质之间的相互作用进行建模的模板。通过将代表性结构叠加在模板中其对应的结构邻居上来创建复合物的模型。这种方法在Dey等,2013 (Prot Sci; 22:359-66)中进一步描述。

[1313] 为了本发明的目的,扩增是指使用能够以合理的保真度复制靶序列的引物和聚合酶的任何方法。可以通过天然或重组DNA聚合酶(例如TaqGold™, T7 DNA聚合酶,大肠杆菌DNA聚合酶的Klenow片段和逆转录酶)进行扩增。优选的扩增方法是PCR。在某些方面,本发明涉及载体。本文所用的“载体”是允许或促进实体从一种环境转移到另一种环境的工具。它是一个复制子,例如质粒、噬菌体或粘粒,可以在其中插入另一个DNA片段,以实现所插入片段的复制。通常,当与适当的控制元件结合时,载体能够复制。通常,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。载体包括但不限于单链,双链或部分双链的核酸分子。核酸分子,其包含一个或多个自由端,无自由端(例如环状);包含DNA、RNA或两者的核酸分子;和本领域已知的其他多核苷酸变体。载体的一种类型是“质粒”,其是指环状双链DNA环,例如通过标准分子克隆技术,可以在其中插入其他DNA片段。载体的另一种类型是病毒载体,其中在载体中存在病毒来源的DNA或RNA序列以包装成病毒(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(AAV))。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)。在导入宿主细胞后,将其他载体(例如,非流行的哺乳动物载体)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,某些载体能够引导与其可操作连接的基因的表达。这样的载体在本文中称为“表达载体”。在重组DNA技术中有用的通用表达载体通常是质粒的形式。重组表达载体可以以适合于在宿主细胞中表达的形式包含本发明的核酸,这意味着重组表达载体包括一种或多种调节元件,其可以根据宿主选择用于表达的细胞,其可操作地连接至要表达的核酸序列。在重组表达载体内,“可操作地连接”是指所关注的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/翻译中)与调控元件连接。载体导入宿主细胞后,该系统或宿主细胞中)。关于重组和克隆方法,提到了2004年9月2日公开的美国专利申请10/815,730,作为US 2004-0171156 A1,其内容通过引用整体并入本文。本发明的方面涉及用于引导RNA和野生型,经修饰或突变的CRISPR效应子蛋白/酶(例如Cas13效应子蛋白)的双顺反子载体。优选双顺反子表达载体引导RNA和野生型,经修饰或突变的CRISPR效应子蛋白/酶(例如Cas13效应子蛋白)。通常,特别是在该实施方式和野生型中,修饰或突变的CRISPR效应子蛋白/酶(例如Cas13效应子蛋白)优选由CBh启动子驱动。RNA可以优选地由Pol III启动子,例如由U6启动子驱动。理想情况下,将两者结合在一起。

[1314] 在一些实施方式中,提供了引导RNA或crRNA中的环。这可以是茎环或四环。该环优选是GAAA,但不限于该序列或实际上仅是4bp的长度。实际上,用于发夹结构的优选的环形成序列的长度为四个核苷酸,并且最优选具有序列GAAA。但是,可以使用更长或更短的循环序列,也可以使用其他序列。序列优选包括核苷酸三联体(例如AAA)和另外的核苷酸(例如C或G)。环形成序列的实例包括CAAA和AAAG。

[1315] 在实施本文公开的任何方法中,可以通过本领域已知的一种或多种方法将合适的载体引入细胞或胚胎,所述方法包括但不限于显微注射,电穿孔,声纳穿孔,生物弹药,磷酸

钙介导的转染,阳离子转染,脂质体转染,树状大分子转染,热休克转染,核转染,磁转染,脂质转染,转染,光学转染,专有试剂对核酸的吸收以及通过脂质体、免疫脂质体、病毒体或人工病毒体的递送。在一些方法中,通过显微注射将载体引入胚胎。可以将一种或多种载体微注射到胚胎的核或细胞质中。在一些方法中,可以通过核转染将一种或多种载体引入细胞。

[1316] 可设计载体以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如大肠杆菌,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。或者,重组表达载体可以在体外转录和翻译,例如使用T7启动子调节序列和T7聚合酶。

[1317] 可以在原核或原核细胞中引入和繁殖载体。在一些实施方式中,原核生物用于扩增将要引入真核细胞的载体的拷贝或作为中间载体,以产生要引入真核细胞的载体(例如将质粒作为病毒的一部分扩增)。向量包装系统)。在一些实施方式中,原核生物用于扩增载体的拷贝并表达一种或多种核酸,从而提供一种或多种蛋白质的来源以递送至宿主细胞或宿主生物。蛋白质在原核生物中的表达最常见的是在大肠杆菌中与含有引导融合蛋白或非融合蛋白表达的组成型或诱导型启动子的载体一起进行。融合载体将许多氨基酸添加到其中编码的蛋白质上,例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,例如:(i)增加重组蛋白的表达;(ii)增加重组蛋白的溶解度;(iii)通过在亲和纯化中充当配体来帮助重组蛋白的纯化。通常,在融合表达载体中,在融合部分和重组蛋白的连接处引入蛋白水解切割位点,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。这样的酶及其同源识别序列包括因子Xa,凝血酶和肠激酶。融合表达载体的实例包括融合谷胱甘肽S-的pGEX(Pharmacia Biotech Inc;Smith和Johnson,1988.Gene 67:31-40),pMAL(New England Biolabs,Beverly,Mass.)和pRIT5(Pharmacia,Piscataway,NJ)。转移酶(GST)、麦芽糖E结合蛋白或蛋白A分别转移至目标重组蛋白。合适的诱导型非融合大肠杆菌表达载体的实例包括 pTrc(Amrann等,(1988)Gene 69:301-315)和pET 11d(Studier等,GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY 185,Academic Press,San Diego,Calif.(1990)60-89)。在一些实施方式中,载体是酵母表达载体。在酵母酿酒酵母中表达的载体的实例包括pYepSec1(Baldari等,1987.EMBO J.6:229-234),pMFa(Kuijan和Herskowitz,1982.Cell 30:933-943),pJRY88(Schultz等,1987.Gene 54:113-123),pYES2(Invitrogen Corporation,San Diego,Calif.)和picZ(InVitrogen Corp,San Diego,Calif.)。在一些实施方式中,载体使用杆状病毒表达载体驱动昆虫细胞中的蛋白质表达。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,1983.Mol.Cell.Biol.3:2156-2165)和pVL系列(Lucklow和Summers,1989.Virology 170:31-39)。在一些实施方式中,载体能够使用哺乳动物表达载体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,et al.,1987.EMBO J.6:187-195)和pMT2PC(Kaufman等,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节元件提供。例如,常用的启动子衍生自多瘤、腺病毒2、巨细胞病毒、猿猴病毒40以及本文公开的和本领域已知的其他启动子。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,MOLECULAR CLONING:A LABORATORY MANUAL.,第二版,Cold Spring Harbor Laboratory,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y., 1989的第16章和第17章。在一些实施方式中,重组哺乳动物表达载体能够引导核酸优先在特定细胞类型中表达(例如组织特异性调节元件用于表达核酸)。组织特异性调节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(肝特异性;Pinkert等,1987.Genes Dev.1: 268-277),淋巴样特异性启动子(Calame和Eaton,1988.Adv.Am.Chem.Soc., 95:1877-1959)。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(Baneiji等,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,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)。还包括发育调节的启动子,例如鼠hox启动子(Kessel和Gruss, 1990.Science 249:374-379)和 $\alpha$ -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,其内容通过引用整体并入本文。本发明的其他实施方式可以涉及病毒载体的使用,关于病毒载体的提及是美国专利申请13/092,085,其内容通过引用整体并入本文。组织特异性调节元件在本领域中是已知的,并且在这方面,提及美国专利7,776,321,其内容通过引用整体并入本文。

[1318] 在一些实施方式中,调节元件可操作地连接至或编码CRISPR Cas系统或复合物的一个或多个元件,以驱动CRISPR系统的一个或多个元件的表达。通常,CRISPRs(聚簇的规则间隔的短回文重复序列),也称为SPIDR(空间穿插的正向重复序列),构成通常对特定细菌物种特异的DNA基因座家族。CRISPR基因座包含在大肠杆菌中识别的不同种类的散布的短序列重复序列(SSR)(Ishino等,J.Bacteriol.,169:5429-5433[1987];和Nakata等,J.Bacteriol., 171:3553-3556[1989])和相关基因。类似的散布的SSR已经在地中海嗜盐杆菌、化脓性链球菌、鱼腥藻和结核分枝杆菌中得到鉴定(参见Groenen等,Mol.Microbiol.,10:1057-1065[1993];Hoe等,Emerg.Infect.Dis.,5:254-263 [1999];Masepohl等,Biochim.Biophys.Acta 1307:26-30[1996];和Mojica等,Mol.Microbiol., 17:85-93[1995])。CRISPR基因座通常与其他SSR的不同之处在于重复序列的结构,这些重复序列被称为短规则间隔重复序列(SRSR)(Janssen等,OMICS J.Integ.Biol.,6:23-33 [2002];和Mojica等,Mol. Microbiol.,36:244-246[2000])。通常,重复序列是短簇,出现在簇中,簇中的簇被具有恒定长度的独特插入序列规则间隔(Mojica等,[2000],同上)。尽管菌株之间的重复序列高度保守,但是散布的重复序列的数目和间隔区的序列通常因菌株而异(van Embden等,J.Bacteriol.,182:2393-2401[2000])。已在40多个原核生物中鉴定了CRISPR基因座(参见例如Jansen等,Mol. Microbiol.,43:1565-1575[2002];和Mojica等,[2005]),包括但不限于气杆菌、嗜热菌、硫磺杆菌、古生球菌、盐杆菌、甲烷菌、甲烷球菌、甲烷菌、甲烷菌属、热球菌、Picrophilus、嗜热菌、棒状杆菌、分枝杆菌、链霉菌、阿基菲克斯氏菌、卟啉单胞菌、支原体、芽孢杆菌、芽孢杆菌、芽孢杆菌、芽孢杆菌、嗜热杆菌属梭菌、海藻、染色体、奈瑟菌、亚硝酸盐单胞菌、脱硫弧菌、土杆菌、粘球菌、弯曲杆菌、沃林氏菌、不动杆菌、欧文氏菌、大肠埃希氏菌、军团菌、甲基球菌、巴斯德氏菌、光细菌、沙门氏菌、

黄单胞菌、嗜热菌、耶尔森氏菌、曲霉。通常,本申请中使用的“RNA靶向系统”统指涉及RNA靶向的CRISPR相关13 (Cas13) 基因的表达或引导RNA的活性的转录物和其他元件(本文也称为“Cas13”)。(如效应子蛋白),包括编码靶向 RNA的Cas(效应子)蛋白和引导RNA(或crRNA序列)的序列,参考本文讨论的突变的CRISPR-Cas。通常,RNA靶向系统的特征在于促进在靶序列的位点形成RNA靶向复合物的元件。在形成靶向RNA的复合物的上下文中,“靶序列”是指引导序列(或引导或crRNA)被设计成具有互补性的RNA序列,其中靶序列与引导之间的杂交RNA促进RNA靶向复合物的形成。如果存在足够的互补性以引起杂交并促进RNA靶向复合物的形成,则不一定需要完全互补。在一些实施方式中,靶序列位于细胞的核或细胞质中。在一些实施方式中,靶序列可以在真核细胞的细胞器内。可以用于重组为包含靶序列的靶基因座的序列或模板被称为“编辑模板”或“编辑RNA”或“编辑序列”。在本发明的方面,外源模板RNA可以被称为编辑模板。在本发明的一方面,重组是同源重组。通常,引导序列是与靶多核苷酸序列具有足够互补性以与靶序列杂交并且将核酸靶向复合物直接序列特异性结合至靶序列的任何多核苷酸序列。在一些实施方式中,当使用合适的比对算法最佳比对时,引导序列与其对应的靶序列之间的互补程度为约或大于约50%,60%,75%,80%,85%,90%,95%,97.5%,99%或更高。可以使用任何合适的用于比对序列的算法来确定最佳比对,其非限制性实例包括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,ELAND(Illumina,圣地亚哥,加利福尼亚),SOAP(可在 [soap.genomics.org.cn](http://soap.genomics.org.cn)上获得)和Maq(可在[maq.sourceforge.net](http://maq.sourceforge.net)上获得)。在一些实施方式中,引导序列为约或大于约5、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、75或更多个核苷酸的长度。在一些实施方式中,引导序列的长度小于约75、50、45、40、35、30、25、20、15、12或更少个核苷酸。引导序列引导RNA靶向复合物与靶序列的序列特异性结合的能力可以通过任何合适的测定法来评估。模板多核苷酸可以具有任何合适的长度,例如约或大于约10、15、20、25、50、75、100、150、200、500、1000或更多个核苷酸的长度。在一些实施方式中,模板多核苷酸与包含靶序列的多核苷酸的一部分互补。当最佳比对时,模板多核苷酸可能与靶序列的一个或多个核苷酸重叠(例如,约或大于约1、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100或更多个核苷酸)。在一些实施方式中,当模板序列和包含靶序列的多核苷酸最佳比对时,模板多核苷酸的最接近核苷酸在距靶序列约1、5、10、15、20、25、50、75、100、200、300、400、500、1000、5000、10000或更多个核苷酸内。在一些实施方式中,靶向RNA的效应子蛋白是包含一个或多个异源蛋白结构域的融合蛋白的一部分(例如,除了靶向核酸的效应子蛋白外,大约或多于大约1、2、3、4、5、6、7、8、9、10或更多个结构域)。在一些实施方式中,CRISPR Cas效应子蛋白/酶是包含一个或多个异源蛋白结构域的融合蛋白的一部分(例如,除了CRISPR Cas酶外,大约或多于大约1、2、3、4、5、6、7、8、9、10或更多个结构域)。可以与效应子蛋白融合的蛋白结构域的实例包括但不限于表位标签,报道基因序列和具有以下一种或多种活性的蛋白结构域:甲基化酶活性,脱甲基酶活性,转录激活活性,转录抑制活性,转录释放因子活性,组蛋白修饰活性,RNA裂解活性和核酸结合活性。表位标签的非限制性实例包括组氨酸(His)标签,V5标签,FLAG标签,流感血凝素(HA)标签,Myc标签,VSV-G标签和硫氧还蛋白(Trx)标签。报道基因的实例包括但不限于谷胱甘肽-S-转移酶(GST),辣根过氧化物酶(HRP),氯霉素乙酰基转移

酶(CAT) $\beta$ -半乳糖苷酶, $\beta$ -葡糖醛酸糖苷酶,荧光素酶,绿色荧光蛋白(GFP),HcRed,DsRed,青色荧光蛋白(CFP),黄色荧光蛋白(YFP)和自发荧光蛋白,包括蓝色荧光蛋白(BFP)。可以将靶向核酸的效应子蛋白与编码结合DNA分子或结合其他细胞分子的蛋白质或蛋白质片段的基因序列融合,所述蛋白质或蛋白质片段包括但不限于麦芽糖结合蛋白(MBP),S-标签, Lex DNA结合结构域(DBD)融合,GAL4 DNA结合域融合和单纯疱疹病毒(HSV)BP16蛋白融合。可以形成包含核酸靶向效应子蛋白的融合蛋白的一部分的另外的结构域描述于US20110059502中,其通过引用并入本文。在一些实施方式中,使用标记的核酸靶向效应子蛋白来鉴定靶序列的位置。在一些实施方式中,CRISPR Cas酶可以形成诱导系统的组分。该系统的可诱导性质将允许使用一种能量来时空控制基因编辑或基因表达。能量的形式可以包括但不限于电磁辐射,声能,化学能和热能。诱导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)或光诱导系统(植物色素,LOV域或隐色色素)。在一个实施方式中,CRISPR CRISPR-Cas酶可以是光诱导转录效应子(LITE)的一部分,以序列特异性方式引导转录活性的变化。灯的组分可以包括CRISPR酶,光响应性细胞色素异二聚体(例如来自拟南芥)和转录激活/抑制域。可诱导的DNA结合蛋白的进一步实例及其使用方法在US 61/736465和US 61/721,283 以及WO 2014/018423和US8889418,US8895308,US20140186919, US20140242700,US20140273234,US20140335620, WO2014093635中提供,其通过引用整体并入本文。在一些方面,本发明提供了包括将一种或多种多核苷酸例如本文所述的一种或多种载体,其一种或多种转录物和/或从其转录的一种或多种蛋白质递送至宿主细胞的方法。在一些方面,本发明进一步提供了通过此类方法产生的细胞,以及包含此类细胞或由此类细胞产生的生物(例如动物、植物或真菌)。在一些实施方式中,与导向RNA或crRNA结合(并且任选地与之结合)的靶向RNA的效应子蛋白被递送至细胞。常规的基于病毒和非病毒的基因转移方法可用于在哺乳动物细胞或靶组织中引入核酸。此类方法可用于向培养中或宿主生物中的细胞施用编码RNA靶向系统组分的核酸。非病毒载体递送系统包括DNA质粒, RNA(例如本文所述载体的转录物),裸核酸以及与诸如脂质体的递送载体复合的核酸。病毒载体递送系统包括DNA和RNA病毒,它们在递送至细胞后具有游离基因组或整合的基因组。有关基因治疗程序的评论,请参见Anderson, Science 256:808-813(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, Current Topics in Microbiology and Immunology, Doerfler和Böhm(编辑)(1995);和Yu等, Gene Therapy 1:13-26(1994)。核酸的非病毒递送方法包括脂质转染,核转染,显微注射,生物弹药,病毒体,脂质体,免疫脂质体,聚阳离子或脂质:核酸偶联物,裸露的DNA,人工病毒体和试剂增强的DNA吸收。脂质转染描述于例如美国专利号5,049,386、4,946,787和4,897,355中)和脂转染试剂在市场上出售(例如Transfectam<sup>TM</sup>和Lipofectin<sup>TM</sup>)。适用于多核苷酸的有效受体识别脂质转染的阳离子脂质和中性脂质包括Felgner, WO 91/17424和WO 91/16024中的那些。可以递送至细胞(例如体外或离体施用)或靶组织(例如体内施用)。

[1319] 本文所述的核酸靶向系统、载体系统、载体和组合物可用于多种核酸靶向应用:改

变或修饰基因产物(例如蛋白质)的合成,核酸切割、核酸编辑、核酸剪接,靶核酸的运输、靶核酸的追踪、靶核酸的分离、靶核酸的可视化等。

[1320] 示例性递送方法

[1321] 通过本公开和本领域的知识,TALEs、CRISPR-Cas系统或其组分或其核酸分子(包括例如HDR模板)或编码或提供其组分的核酸分子可通过本文中概括和详细描述递送系统递送。

[1322] 载体递送,例如质粒、病毒递送:CRISPR酶和/或任何本发明的RNA,例如指导RNA,可以使用任何合适的载体递送,例如质粒或病毒载体,例如腺相关病毒(AAV)、慢病毒、腺病毒或其他病毒载体类型,或其组合。效应蛋白和一种或多种指导RNA可被包装到一种或多种载体中,例如质粒或病毒载体。在一些实施方式中,载体(例如质粒或病毒载体)通过例如肌肉内注射递送至感兴趣的组织,而在其他时间递送是通过静脉内、经皮、鼻内、口腔、粘膜或其他递送方法。这样的递送可通过单剂或多剂进行。本领域技术人员理解,根据多种因素,例如载体选择、靶细胞、生物或组织、待治疗对象的一般状况、所寻求的转化/修饰程度、施用途径、施用模式、所寻求的转化/修饰类型等,本文中待递送的实际剂量可能有很大差异。

[1323] 该剂量可进一步包含例如载体(水、盐水、乙醇、甘油、乳糖、蔗糖、磷酸钙、明胶、葡聚糖、琼胶、果胶、花生油、麻油等)、稀释剂、药学上可接受的载体(例如磷酸盐缓冲盐水)、药学上可接受的赋型剂和/或本领域已知的其他化合物。该剂量可进一步包含一种或多种药学上可接受的盐,例如无机酸盐,例如盐酸盐、氢溴酸盐、磷酸盐、硫酸盐等;和有机酸的盐,例如乙酸盐、丙酸盐、丙二酸盐、苯甲酸盐等。此外,辅助物质,例如湿润剂或乳化剂、pH缓冲物质、凝胶或胶凝材料、香料、着色剂、微球、聚合物、悬浮剂等。也可以存在于此。另外,一种或多种其他常规药物成分,例如防腐剂、湿润剂、助悬剂、表面活性剂、抗氧化剂、抗结块剂、填充剂、螯合剂、涂层剂、化学稳定剂等也可以存在,尤其是当剂型为可重构形式时。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、乙基香兰素、甘油、苯酚、对氯苯酚、明胶、白蛋白及其组合。REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub.Co., N.J. 1991) 对药学上可接受的辅料进行了详尽的讨论,其通过引用并入本文。

[1324] 在本发明的一个实施方式中,通过腺病毒进行递送,其可以是包含至少 $1 \times 10^5$ 个腺病毒载体颗粒(也称为颗粒单位,pu)的单次增强剂量。在本文的实施方式中,剂量优选为至少约 $1 \times 10^6$ 个颗粒(例如,约 $1 \times 10^6$ - $1 \times 10^{12}$ 个颗粒),更优选至少约 $1 \times 10^7$ 个颗粒,更优选至少约 $1 \times 10^8$ 个颗粒(例如,约 $1 \times 10^8$ - $1 \times 10^{11}$ 个颗粒或约 $1 \times 10^8$ - $1 \times 10^{12}$ 个颗粒),最优选至少约 $1 \times 10^9$ 个颗粒(例如,约 $1 \times 10^9$ - $1 \times 10^{10}$ 个颗粒或约 $1 \times 10^9$ - $1 \times 10^{12}$ 个颗粒),或甚至至少约 $1 \times 10^{10}$ 个颗粒(例如约 $1 \times 10^{10}$ - $1 \times 10^{12}$ 个颗粒)的腺病毒载体。或者,剂量包含不超过约 $1 \times 10^{14}$ 个颗粒,优选不超过约 $1 \times 10^{13}$ 个颗粒,甚至更优选不超过约 $1 \times 10^{12}$ 个颗粒,甚至更优选不超过约 $1 \times 10^{11}$ 个颗粒,最优选不超过约 $1 \times 10^{10}$ 个颗粒(例如,不超过约 $1 \times 10^9$ 个颗粒)。因此,该剂量可包含单剂量的腺病毒载体,其具有例如约 $1 \times 10^6$ 个颗粒单位(pu)、约 $2 \times 10^6$ pu、约 $4 \times 10^6$ pu、约 $1 \times 10^7$ pu、约 $2 \times 10^7$ pu、约 $4 \times 10^7$ pu、约 $1 \times 10^8$ pu、约 $2 \times 10^8$ pu、约 $4 \times 10^8$ pu、约 $1 \times 10^9$ pu、约 $2 \times 10^9$ pu、约 $4 \times 10^9$ pu、约 $1 \times 10^{10}$ pu、约 $2 \times 10^{10}$ pu、约 $4 \times 10^{10}$ pu、约 $1 \times 10^{11}$ pu、约 $2 \times 10^{11}$ pu、约 $4 \times 10^{11}$ pu、约 $1 \times 10^{12}$ pu、约 $2 \times 10^{12}$ pu或约 $4 \times 10^{12}$ pu的腺病毒载体。参见例如Nabel等于2013年6月4日授予的美国专利第8,454,972B2号中的腺病毒载

体,通过引用并入本文,以及其第29栏第 36-58行的剂量。在本发明的一个实施方式中,通过多剂量递送腺病毒。

[1325] 在本文的一个实施方式中,递送通过AAV进行。据信,将AAV体内递送至人体的治疗有效剂量为约20至约50ml含约 $1 \times 10^{10}$ 至约 $1 \times 10^{10}$ 功能性AAV/ml溶液的盐溶液。可调整剂量以平衡治疗效果和任何副作用。在本发明的实施方式中,AAV剂量通常在约 $1 \times 10^5$ 至 $1 \times 10^{50}$ 个基因组AAV、约 $1 \times 10^8$ 至 $1 \times 10^{20}$ 个基因组AAV、约 $1 \times 10^{10}$ 至约 $1 \times 10^{16}$ 个基因组或约 $1 \times 10^{11}$ 至约 $1 \times 10^{16}$ 个基因组AAV的浓度范围内。人的剂量约为 $1 \times 10^{13}$ 个基因组AAV。这样的浓度可以以约0.001ml至约100ml、约 0.05至约50ml或约10ml至约25ml载体溶液的形式递送。本领域普通技术人员可以通过建立剂量响应曲线的常规试验容易地确定其他有效剂量。例如,参见Hajjar等于2013年3月26日授予的美国专利第 8,404,658B2号第27栏第45-60行。

[1326] 在本发明的一个实施方式中,通过质粒进行递送。在这样的质粒组合中,剂量应足以引发反应。例如,质粒组合中质粒DNA的合适数量可以为每70kg个体约0.1至约2mg,或约1 $\mu$ g至约10 $\mu$ g。本发明的质粒通常包含(i)启动子;(ii)编码可操作地连接至所述启动子的核酸靶向 CRISPR酶的序列;(iii)可选标志物;(iv)复制起点;和(v)位于(ii)下游并可操作地连接至(ii)的转录终止子。该质粒也可以编码CRISPR复合物的 RNA组分,但其中这些中的一种或多种也可以在不同的载体上编码。

[1327] 本文的剂量基于平均70kg个体。施用频率在医学或兽医学从业人员(例如,医生、兽医)或本领域技术人员的范围内。还应注意的是,实验中使用的小鼠通常约为20g,并且从小鼠实验中可以放大至70kg个体。

[1328] 在一些实施方式中,本发明的RNA分子以脂质体或脂质体制剂等形式递送,并且可以通过本领域技术人员公的方法制备。这样的方面描述于例如,第5,593,972号、第5,589,466号和第5,580,859号美国专利中,通过引用并入本文。已经开发了专门用于增强和改善向哺乳动物细胞递送siRNA的递送系统(例如,参见Shen et al FEBS Let.2003,539:111-114; Xia et al.,Nat.Biotech.2002,20:1006-1010;Reich et al.,Mol.Vision.2003,9:210-216;Sorensen et al.,J.Mol.Biol.2003,327:761-766;Lewis et al.,Nat.Gen.2002,32:107-108and Simeoni et al.,NAR 2003,31,11: 2717-2724),并可适用于本发明。siRNA最近已成功用于抑制灵长目动物的基因表达(例如参见.Tolentino et al.,Retina 24(4):660,其也可适用于本发明。

[1329] 事实上,RNA递送是一种有效的体内递送方法。可以使用脂质体或颗粒将核酸靶向的Cas蛋白和指导RNA递送(例如,HR修复模板)入细胞。因此,本发明的核酸靶向的CRISPR-Cas蛋白的递送和/或指导RNA 或crRNA的递送可以是RNA形式并通过微泡、脂质体或颗粒进行。例如,CRISPR-Cas mRNA和指导RNA或crRNA可被包装成脂质体颗粒用于体内递送。脂质体转染试剂(例如Life Technologies的lipofectamine)和市场上的其他试剂可以有效地将RNA分子递送至肝脏。

[1330] RNA递送的方式还优选包括通过纳米颗粒(Cho,S.,Goldberg,M., Son,S.,Xu,Q., Yang,F.,Mei,Y.,Bogatyrev,S.,Langer,R.and Anderson, D.,Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells,Advanced Functional Materials,19:3112-3118,2010)或外泌体(exosome)(Schroeder, A.,Levins,C.,Cortez,C.,Langer,R.,and Anderson, D.,Lipid-based

nanotherapeutics for siRNA delivery, *Journal of Internal Medicine*, 267:9-21, 2010, PMID:20059641) 递送RNA。事实上,外泌体已被证明对递送siRNA特别有用,该系统与RNA靶向系统有些相似之处。例如El-Andaloussi S, et al. (“Exosome-mediated delivery of siRNA in vitro and in vivo.” *Nat Protoc.* 2012Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012Nov 15) 描述外泌体如何成为跨越不同生物屏障进行药物递送的有前景的工具,并可用于在体外和体内递送 siRNA。他们的方法是通过转染表达载体产生靶向外泌体,该表达载体包含与肽配体融合的外泌体蛋白。然后从转染的细胞上清液中纯化并鉴定外泌体,然后将RNA装载到外泌体中。根据本发明的递送或施用可以用外泌体进行,特别是但不限于脑。维生素E( $\alpha$ -生育酚)可与核靶向的Cas 蛋白缀合,并与HDL(HDL)一起递送至大脑,例如,与Uno et al. (*HUMAN GENE THERAPY* 22:711-719 (June 2011))类似的方式将短干扰 RNA(siRNA)递送至大脑。通过渗透微型泵(1007D型;Alzet,Cupertino, CA)灌注小鼠,其填充有磷酸盐缓冲液(PBS)或游离TocsiBACE or Toc-siBACE/HDL,并与Brain Infusion Kit 3(Alzet)连接。将脑注入套管放置在正中线上前凶后方约0.5毫米处,用于注入第三脑室背侧。Uno 等发现,通过相同的ICV注入方法,仅3nmol的Toc-siRNA与HDL就能诱导靶标降低相当程度。在本发明中,可考虑将类似剂量的核酸靶向效应蛋白与 $\alpha$ -生育酚缀合并与靶向大脑的HDL共同施用于人类,例如,可考虑将约3nmol至约3 $\mu$ mol的核酸靶向效应蛋白用于大脑。Zou等 (*HUMAN GENE THERAPY* 22:465-475 (2011年4月))描述了一种以慢病毒介导的递送针对PKC  $\gamma$  的短发卡RNA的方法,用于在大鼠脊髓中进行体内基因沉默。Zou等通过鞘内导管施用约10 $\mu$ l重组慢病毒,其滴度为  $1 \times 10^9$ 转导单位(TU)/ml。在本发明中,用于人类,可以考虑在靶向大脑的慢病毒载体中表达类似剂量的核酸靶向效应蛋白,例如,可以考虑在  $1 \times 10^9$ 转导单位(TU)/ml的慢病毒中靶向大脑的约10-50ml的核酸靶向效应蛋白。

[1331] 就向大脑的局部传递而言,这可以通过多种方式实现。例如,材料可以通过例如注射的方式进行层内(intrastratially)递送。注射可以通过开颅术进行立体定向。

[1332] 一般的包装和启动子

[1333] 将编码核酸分子(例如DNA)的RNA靶向效应蛋白(CRISPR-Cas蛋白)包装到载体(例如病毒载体)中以在体内介导基因组修饰的方法包括:

[1334] 单一病毒载体:

[1335] 包含两个或更多个表达盒的载体:

[1336] 启动子-编码核酸靶向效应蛋白的核酸分子-终止子

[1337] 启动子-指导RNA1-终止子

[1338] 启动子-指导RNA(N)-终止子(不超过载体的大小限制) 双病毒载体:

[1339] 载体1包含一个表达盒,用于驱动靶向RNA的效应蛋白(CRISPR-Cas)的表达

[1340] 启动子-编码RNA靶向效应(CRISPR-Cas)蛋白的核酸分子-终止子

[1341] 载体2包含一个或多个表达盒,用于驱动一个或多个指导RNA 或crRNA的表达

[1342] 启动子-指导RNA1或crRNA1-终止子

[1343] 启动子-指导RNA1(N)或crRNA1(N)-终止子(不超过载体的大小限制)。

[1344] 用于驱动编码RNA靶向效应蛋白的核酸分子的表达的启动子可包括AAV·ITR可作为启动子:这有利于消除对另外的启动子元件(其可占据载体中的空间)的需求。释放的另

外的空间可用于驱动另外元件的表达 (gRNA等)。此外, ITR活性相对较弱, 因此可用于降低因核酸靶向效应蛋白过度表达而产生的潜在毒性。对于泛表达, 可以使用启动子: CMV、CAG、CBh、PGK、SV40、铁蛋白重链或轻链等。对于大脑或其他中枢神经系统的表达, 可以使用启动子: Synapsin I用于所有神经元, CaMKII alpha用于兴奋性神经元, GAD67或GAD65或VGAT用于 $\gamma$ -氨基丁酸能神经元等。对于肝脏表达, 可使用白蛋白启动子。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, 可使用IFNbeta 或CD45。成骨细胞可以使用OG-2。用于驱动指导RNA的启动子可包括: Pol III启动子, 例如U6或H1; Pol II启动子和内含子盒, 用于表达指导 RNA或crRNA。

[1345] 腺相关病毒 (AAV)

[1346] CRISPR-Cas和一种或多种指导RNA或crRNA可使用腺相关病毒 (AAV)、慢病毒、腺病毒或其他质粒或病毒载体类型递送, 特别是使用来自例如美国专利第8, 454, 972号 (腺病毒制剂、剂量)、第8, 404, 658号 (AAV制剂、剂量) 和第5, 846, 946号 (DNA质粒制剂、剂量) 以及来自关于慢病毒、AAV和腺病毒临床试验的临床试验和出版物的制剂和剂量。例如, 就AAV而言, 施用途径、制剂和剂量可与美国专利第8, 454, 972号和涉及AAV的临床试验相同。就腺病毒而言, 其施用途径、制剂和剂量可参见美国专利第8, 404, 658号和涉及腺病毒的临床试验。就质粒递送而言, 施用途径、制剂和剂量可参见美国专利5, 846, 946和涉及质粒的临床研究。剂量可以基于平均70千克的个体 (例如, 成年男性), 也可以根据患者、受试者、不同体重和种类的哺乳动物进行调整。根据通常的因素, 包括患者或受试者的年龄、性别、一般健康状况、其他状况以及所治疗的特定状况或症状, 施用频率属于医疗或兽医从业人员 (例如医生、兽医) 的范围。病毒载体可以被注射到感兴趣的组织中。对于细胞类型特异性的基因组修饰, 可通过细胞类型特异性的启动子驱动靶向RNA的效应蛋白 (CRISPR-Cas效应蛋白) 的表达。例如, 肝脏特异性表达可能使用白蛋白启动子, 神经元特异性表达 (例如, 用于靶向中枢神经系统疾病) 可能使用Synapsin I启动子。就体内递送而言, AAV优于其他病毒载体, 原因如下: 低毒性 (这可能是由于纯化方法不需要对可激活免疫反应的细胞颗粒超离心) 和因其未整合入宿主基因组而导致低概率的插入突变。

[1347] AAV的包装限制为4.5或4.75Kb。这意味着, 靶向RNA的效应蛋白 (CRISPR-Cas效应蛋白) 的编码序列以及启动子和转录终止子都必须适合同一病毒载体。就AAV而言, AAV可以是AAV1、AAV2、AAV5或其任何组合。可以根据待靶向的细胞相关的AAV选择AAV; 例如, 可以选择AAV血清型1、2、5或杂合衣壳AAV1、AAV2、AAV5或其任何组合, 用于靶向脑或神经细胞; 并且可以选择AAV4用于靶向心脏组织。AAV8可用于递送至肝脏。本文中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. et al, J.Virol. 82: 5887-5911 (2008)) 如下:

[1348] 表9

[1349]

| 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  |
| 未成熟DC  | 2500 | 100 | ND  | ND  | 222 | 2857 | ND  | ND  |
| 成熟DC   | 2222 | 100 | ND  | ND  | 333 | 3333 | ND  | ND  |

[1350] 慢病毒

[1351] 慢病毒是一种复杂的反转录病毒,能够在分裂细胞和分裂后细胞中感染和表达其基因。最常见的慢病毒是人类免疫缺陷病毒(HIV),它利用其他病毒的包膜糖蛋白来针对广泛的细胞类型。慢病毒的制备方法如下。在克隆pCasES10(其包含慢病毒转移质粒主链)后,在转染前一天,将低代HEK293FT(p=5)接种于T-75烧瓶中达到50%融合在10%胎牛血清(无抗生素)的DMEM。20小时后,将培养基更换为OptiMEM(无血清)培养基,4小时后进行转染。用10 $\mu$ g慢病毒转移质粒(pCasES10)和下列包装质粒:(5 $\mu$ g pMD2.G)(VSV-g假型)和7.5 $\mu$ g psPAX2(gag/pol/rev/tat)转染细胞。在4mL OptiMEM中用阳离子脂质递送剂(50 $\mu$ L Lipofectamine 2000和100 $\mu$ L Plus试剂)进行转染。6小时后,用10%胎牛血清将培养基更换为无抗生素的DMEM。这些方法在细胞培养过程中使用血清,但优选无血清方法。

[1352] 慢病毒可按如下方式纯化。48小时后收集病毒上清液。首先清除上清液中的碎片,并通过0.45 $\mu$ m低蛋白结合(pdf)过滤器过滤。然后在超速离心机中以24,000rpm的速度旋转2小时。将病毒颗粒在50 $\mu$ L的DMEM溶液中于4 $^{\circ}$ C下悬浮过夜。然后将其等分并立即在-80 $^{\circ}$ C冷冻。

[1353] 在另一个实施方式中,也考虑了基于马传染性贫血病毒(EIAV)的最小非灵长目慢病毒载体,尤其是用于眼睛基因治疗(例如参见Balagaan, J Gene Med 2006;8:275-285)。在另一个实施方式中,还考虑了 **RetinoStat<sup>®</sup>**,一种基于马传染性贫血病毒的慢病毒基因治疗载体,其表达通过视网膜下注射递送的血管抑制蛋白内皮抑制素和血管抑制素,用于治疗网络形式的年龄相关性黄斑变性(参见,例如,Binley et al., HUMAN GENE THERAPY 23:980-991(September 2012)),并且该载体被修饰用于本发明的核酸靶向系统。

[1354] 在另一个实施方式中,可以使用和/或适应于本发明的核酸靶向系统改造自失活的慢病毒载体,其包含靶向由HIV tat/rev共享的共同外显子的siRNA、核仁定位TAR诱饵和抗CCR5特异性锤头核酶(例如参见 DiGiusto et al.(2010) Sci Transl Med 2:36ra43)。每kg患者体重至少可收集 $2.5 \times 10^6$ 个CD34+细胞,并在含有2 $\mu$ mol/L谷氨酰胺、干细胞因子(100 ng/ml)、Flt-3配体(Flt-3L)(100ng/ml)和血小板生成素(10ng/ml)(CellGenix)的X-VIVO 15培养基(Lonza)中以 $2 \times 10^6$ 个细胞/ml的密度预刺激16至20小时。预刺激的细胞可以在涂有纤维连接蛋白(25mg/cm<sup>2</sup>)的75-cm<sup>2</sup>组织培养瓶中以5的多重感染用慢病毒转导16至24小时(RetroNectin,Takara Bio Inc.)。

[1355] 慢病毒载体已被公开用于治疗帕金森氏病,例如参见美国专利公开号

20120295960和美国专利7303910和7351585。还公开了用于治疗眼睛疾病的慢病毒载体,例如参见美国专利公开号20060281180, 20090007284,20110117189;US20090017543;US20070054961, US20100317109。慢病毒载体也被公开用于递送至大脑,例如参见美国专利公开号US20110293571;US20110293571,US20040013648,US20070025970,US20090111106和US专利号7259015。

[1356] RNA递送

[1357] RNA递送:靶向核酸的CRISPR-Cas蛋白和/或指导RNA也可以以RNA的形式递送。可使用包含以下元件的PCR盒合成mRNA:T7\_启动子-kozak序列(GCCACC)-效应蛋白-来自 $\beta$ 珠蛋白的3' UTR-polyA尾(一串120或更多的腺嘌呤)。该盒可用于T7聚合酶的转录。也可以使用体外转录从含有T7\_promoter-GG-指导RNA或crRNA序列的盒中转录指导RNA或crRNA。

[1358] 颗粒递送系统和/或制剂:

[1359] 已知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。一般而言,颗粒被定义为一个物体,它作为一个完整的单元在其传输和性质方面发挥作用。颗粒根据直径进一步分类。粗颗粒涵盖在2,500至10,000纳米之间的范围。细微粒的尺寸在100至2500纳米之间。超细颗粒或纳米颗粒的尺寸通常在1至100纳米之间。100纳米限制的基础是,将颗粒与散装材料区分开来的新性质通常在低于100纳米的临界长度范围内形成。

[1360] 如本文所用,颗粒递送系统/制剂被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具有小于100微米( $\mu\text{m}$ )的最大尺寸(例如直径)的任何实体。在一些实施方式中,本发明的颗粒具有小于10 $\mu\text{m}$ 的最大尺寸。在一些实施方式中,本发明的颗粒具有小于2000 纳米(nm)的最大尺寸。在一些实施方式中,本发明的颗粒具有小于1000 纳米(nm)的最大尺寸。在一些实施方式中,本发明的颗粒具有小于900 纳米、800纳米、700纳米、600纳米、500纳米、400纳米、300纳米、200纳米或100纳米的最大尺寸。典型地,本发明的颗粒具有500纳米或更小的最大尺寸(例如,直径)。在一些实施方式中,本发明的颗粒具有250纳米或更小的最大尺寸(例如,直径)。在一些实施方式中,本发明的颗粒具有200纳米或更小的最大尺寸(例如,直径)。在一些实施方式中,本发明的颗粒具有150纳米或更小的最大尺寸(例如,直径)。在一些实施方式中,本发明的颗粒具有100纳米或更小的最大尺寸(例如,直径)。在本发明的一些实施方式中使用更小的颗粒,例如最大尺寸为50纳米或更小的颗粒。在一些实施方式中,本发明的颗粒具有25纳米至200纳米范围内的最大尺寸。

[1361] 使用各种不同的技术进行颗粒表征(包括例如表征形态、尺寸等)。常用的技术有电子显微术、原子力显微术、动态光散射术(DLS)、X射线光子光谱术、粉末X射线衍射术、傅里叶变换红外光谱术、基质辅助激光解吸附/离子化飞行时间质谱术、紫外-可见光谱术、双偏振干涉术和核磁共振。可对天然颗粒(即预载)或货物装载后(本文货物指例如CRISPR-Cas系统的一种或多种成分,例如CRISPR-Cas酶或mRNA或指导RNA或其任何组合,并可包括另外的载体和/或赋形剂)进行表征(尺寸测量),以提供最佳尺寸的颗粒用于本发明的任何体外、离体和/或体内应用。在某些优选实施方式中,颗粒尺寸(例如直径)表征基于使用动态激光散射(DLS)的测量。提及美国专利第8,709,843号;美国专利号6,007,845;美国专利号5,855,913;美国专利号5,985,309;美国专利第5,543,158号专利;以及James E. Dahlman和Carmen Barnes等于2014年5月11日在线发布的出版物Nature

Nanotechnology (2014) (doi:10.1038/nano.2014.84), 内容涉及颗粒、制备和使用及其测量的方法。另见Dahlman等,“Orthogonal gene control with a catalytically active Cas9 nuclease,”*Nature Biotechnology* 33,1159-1161 (November,2015)。

[1362] 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、乳状或胶态颗粒。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、微胶团、微泡、外泌体或基因枪,均可作为本发明范围内的颗粒递送系统提供。

[1363] 颗粒

[1364] CRISPR-Cas mRNA和指导RNA或crRNA可使用颗粒或脂包膜同时递送;例如,本发明的CRISPR酶和RNA(例如,作为复合物)可通过 Dahlman等、W02015089419 A2和其中引用的文件,例如7C1(参见James E.Dahlman和Carmen Barnes等2014年5月11日在线出版的*Nature Nanotechnology* (2014) (doi:10.1038/nano.2014.84)中的颗粒递送,例如,包含脂质或类脂的递送颗粒,和亲水聚合物,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1,2-二十四烷酰基-sn-甘油-3-磷酸胆碱(DMPC)和/或其中亲水性聚合物包括乙二醇或PEG(PEG);和/或其中颗粒进一步包含胆固醇(例如,来自制剂1的颗粒=DOTAP 100,DMPC 0,PEG 0,胆固醇0;制剂编号2= DOTAP 90,DMPC 0,PEG 10,胆固醇0;制剂编号3=DOTAP 90,DMPC 0,PEG 5,胆固醇5),其中使用高效的多步骤方法形成颗粒,其中首先将效应蛋白和RNA例如以1:1的摩尔比,例如在室温下混合在一起,例如持续30分钟,例如在无菌的、无核酸酶的1X PBS中;分别将适用于制剂的DOTAP,DMPC,PEG和胆固醇溶解在酒精(例如100%乙醇)中;然后将两种溶液混合在一起,形成含有复合物的颗粒。CRISPR-Cas 效应蛋白mRNA和指导RNA可以使用颗粒或脂质包膜同时递送。该 Dahlman等的技术可以应用在本发明中。可以使用环氧化物修饰的脂质聚合物将本发明的核酸靶向系统递送至肺、心血管或肾细胞,但是,本领域技术人员可以使该系统适应于递送至其他靶器官。设想的剂量范围为约0.05至约0.6mg/kg。还设想了几天或几周的剂量,总剂量为约2 mg/kg。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.*2011Jun 6;8(3):774-87.doi: 10.1021/mp100390w.Epub 2011Apr 1)描述了具有可被生物降解的核-壳结构的颗粒,该颗粒具有被磷脂双层壳包围的聚(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。

[1365] 在一个实施方式中,考虑了基于自组装生物粘附性聚合物的颗粒,该颗粒可用于经口递送肽、经静脉递送肽和经鼻递送肽,这些均可用于脑。也考虑了其他实施方式,例如疏水性药物的口服吸收和眼内递送。分子包膜技术涉及受保护并递送至疾病部位的工程化聚合物包膜(例如,参见Mazza,M.et al.*ACS Nano*,2013.7(2):1016-1026;Siew,A.,et al.*Mol Pharm*,2012.9(1):14-28;Lalatsa,A.,et al.*J Contr Rel*,2012.161(2):523-36;Lalatsa,A.,et al.,*Mol Pharm*,2012.9(6):1665-80;Lalatsa,A.,et al.*Mol Pharm*,2012.9(6):1764-74;Garrett,N.L.,et al.*J Biophotonics*,2012. 5(5-6):458-68;Garrett,N.L.,et al.*J Raman Spect*,2012.43(5):681-688; Ahmad,S.,et al.*J Royal Soc Interface* 2010.7:S423-33;Uchegbu,I.F. *Expert Opin Drug Deliv*,2006.3(5):

629-40; Qu, X., et al. *Biomacromolecules*, 2006.7 (12) :3452-9 and Uchegbu, I.F., et al. *Int J Pharm*, 2001. 224:185-199)。根据靶组织的不同, 预计剂量为约5mg/kg, 单次或多次施用。

[1366] 关于颗粒, 参见 Alabi et al., *Proc Natl Acad Sci U S A*. 2013 Aug 6; 110 (32) : 12881-6; Zhang et al., *Adv Mater*. 2013 Sep 6; 25 (33) : 4641-5; Jiang et al., *Nano Lett*. 2013 Mar 13; 13 (3) : 1059-64; Karagiannis et al., *ACS Nano*. 2012 Oct 23; 6 (10) : 8484-7; Whitehead et al., *ACS Nano*. 2012 Aug 28; 6 (8) : 6922-9 和 Lee et al., *Nat Nanotechnol*. 2012 Jun 3; 7 (6) : 389-93。

[1367] 美国专利申请 20110293703 涉及脂烯醚萜类化合物, 也特别适用于多聚核苷酸的施用, 其可用于递送本发明的核酸靶向系统。在一个方面, 氨基醇类脂质化合物与待递送至细胞或受试者的试剂组合, 以形成微粒、纳米颗粒、脂质体或胶束。待由颗粒、脂质体或胶束递送的试剂可以是气体、液体或固体的形式, 并且试剂可以是多核苷酸、蛋白质、肽或小分子。氨基醇类脂质化合物可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、蛋白质、脂质等结合以形成颗粒。然后, 这些颗粒可任选地与药物赋形剂结合形成药物组合物。美国专利公开号 20110293703 也提供了制备氨基醇类脂质化合物的方法。允许一种或多种当量的胺与一种或多种当量的环氧化物封端的化合物在合适的条件下反应, 以形成本发明的氨基醇类脂质化合物。在某些实施方式中, 胺的所有氨基与环氧化物封端的化合物完全反应形成叔胺。在其他实施方式中, 胺的所有氨基未与环氧化物封端的化合物完全反应以形成叔胺, 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, 或可与另一亲电试剂(例如不同的环氧化物封端的化合物)反应。如本领域技术人员将理解的, 胺与少于过量的环氧化物封端的化合物反应将产生具有不同尾数的多种不同的氨基醇类脂质化合物。某些胺可以用两个环氧化物衍生的化合物尾完全功能化, 而其他分子不会用环氧化物衍生的化合物尾完全功能化。例如, 二胺或多胺可包括一种、两种、三种或四种衍生自环氧化物的化合物, 该化合物从分子的各种氨基部分脱落, 产生伯、仲和叔胺。在某些实施方式中, 并非所有的氨基都被完全功能化。在某些实施方式中, 使用两种相同类型的环氧化物封端的化合物。在其他实施方式中, 使用两种或更多种不同的环氧化物封端的化合物。氨基醇类脂质化合物的合成在有或无溶剂的情况下进行, 并且合成可在 30-100°C 的较高温度下进行, 优选在约 50-90°C 下进行。制备的氨基醇类脂质化合物可任选地被纯化。例如, 可以对氨基醇类脂质化合物的混合物进行纯化, 得到具有特定数量环氧化物衍生化合物尾的氨基醇类脂质化合物。或者可以对混合物进行纯化以产生特定的立体或区域异构体。氨基醇类脂质化合物也可以使用烷基卤(例如甲基碘)或其他烷基化试剂进行烷基化, 和/或它们可以被酰化。

[1368] 美国专利公开号 20110293703 也提供了通过本发明方法制备的氨基醇类脂质化合物的库。可以使用高通量技术(包括液体处理器、机器人、微量滴定板、计算机等)制备和/或筛选这些氨基醇类脂质化合物。在某些实施方式中, 对氨基醇类脂质化合物进行筛选, 以确定其将多核苷酸或其他试剂(例如, 蛋白、肽、小分子)转入细胞的能力。美国专利公开号 20130302401 涉及一类使用组合聚合制备的聚( $\beta$ -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(例如用于医疗器械或植入体的膜或多层膜的涂层)、添加剂、材料、赋形剂、非生物污垢剂、微图案形成剂和细胞包裹剂。当用作表面涂层时, 这些

PBAA根据其化学结构在体外和体内引发不同程度的炎症,取决于他们的化学结构。这类材料的化学多样性使我们能够在体外识别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后,这些涂层减少了炎症细胞的募集,并减少了纤维变性。这些聚合物可用于形成用于细胞包裹的聚电解质复合胶囊。本发明还可具有许多其他生物应用,例如抗微生物涂层、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核酸靶向系统。

[1369] 在另一个实施方式中,考虑了脂质纳米颗粒(LNP)。抗甲状腺素运载蛋白小干扰RNA已被包裹在脂质纳米颗粒中并被递送至人体(例如参见Coelho等,N Engl J Med 2013; 369:819-29),并且这种系统可适用于本发明的核酸靶向系统。每千克体重静脉注射约0.01至约1毫克的剂量是可以考虑的。可考虑使用降低输注相关反应风险的药物,例如地塞米松、乙酰氨基酚、苯海拉明或西替利嗪以及雷尼替丁。多次剂量也可考虑每4周分五次施用,每次约0.3毫克/千克。LNP已被证明能高效地将siRNA递送至肝脏(例如,参见Tabernero et al.,Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470),因此可用于将编码靶向核酸的效应蛋白的RNA递送至肝脏。可考虑每两周使用约四剂6mg/kg的LNP。Tabernero等证明,在以0.7mg/kg剂量施用的前2个周期的LNP后,观察到肿瘤消退,并且在6个周期结束时,患者获得了部分响应,淋巴结转移完全消退,并且肝脏肿瘤显著缩小。该患者在接受40剂后获得完全缓解,并在接受超过26个月的剂量后完成治疗。两名具有肾癌和肝外疾病的患者(包括肾脏,肺和淋巴结)(先前用VEGF途径抑制剂治疗后续展)所有部位疾病稳定约8至12个月,而PNET和肝转移患者继续扩展研究18个月(36剂),病情稳定。但是,必须考虑LNP的电荷。阳离子脂质与带负电荷的脂质结合以诱导促进细胞内递送的非双层结构。由于带电荷的LNP在静脉内注射后迅速从循环中清除,因此开发出pKa 值低于7的可电离的阳离子脂质(请参阅例如Rosin et al.,Molecular Therapy,vol.19,no.12,pages 1286-2200, Dec.2011)。可以将带负电荷的聚合物(例如RNA)以低pH值(例如pH 4)加载到LNP中,在低pH值下可电离脂质显示正电荷。但是,在生理pH值下,LNP的表面电荷低,循环时间更长。四种可离子化的阳离子脂质已得到关注,分别是1,2-二亚油基-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)。已经显示,包含这些脂质的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,其效力根据采用因子 VII基因沉默模型的DLinKC2-DMA>DLinKDMA>DLinDMA>> DLinDAP系列而变化(参见,例如Rosin et al.,Molecular Therapy,vol.19,no.12,pages 1286-2200,Dec.2011)。可以考虑1μg/ml剂量LNP或与在 LNP中或与LNP相关的CRISPR-Cas RNA,特别是对于含有 DLinKC2-DMA的制剂。

[1370] LNP和CRISPR-Cas胶囊的制备可使用和/或改造自Rosin et al., Molecular Therapy,vol.19,no.12,pages 1286-2200,Dec.2011。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-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PEG2000)琥珀酰]-1,2-二肉豆蔻酰基-sn-乙二醇(PEG-S-DMG)和R-3-[(ω-甲氧基-聚(乙二醇)2000)氨基甲酰基]-1,2-二肉豆蔻基氧丙基-3-胺(PEG-C-DOMG)可以由Tekmira Pharmaceuticals(加拿大温哥华)提供

或合成;胆固醇可以从 Sigma (圣路易斯,密苏里州) 购买。特定的核酸靶向复合物 (CRISPR-Cas) RNA 可以封装在含有 DLinDAP, DLinDMA, DLinK-DMA 和 DLinKC2-DMA (阳离子脂质: DSPC:CHOL:PEGS-DMG 或 PEG-C-DMG 于摩尔比为 40:10:40:10) 的 LNP 中。在需要时添加 0.2% SP-DiOC18 (Invitrogen, 加拿大伯灵顿) 可以评估细胞的吸收, 细胞内传递和生物分布。通过将包含阳离子脂质: DSPC:胆固醇:PEG-c-DMG (40:10:40:10 摩尔比) 的脂质混合物溶解在乙醇中至最终脂质浓度为 10mmol/L 进行封装。可以将这种脂质的乙醇溶液逐滴添加至 pH 4.0 的 50mmol/l 柠檬酸盐中以形成多层囊泡, 以产生 30% 乙醇体积/体积的最终浓度。在使用挤出机 (Northern Lipids, 加拿大温哥华) 通过两个堆叠的 80nm Nuclepore 聚碳酸酯滤膜挤出多层囊泡后, 可能会形成大的单层囊泡。通过将溶于含有 30% 乙醇体积/体积的 50mmol/l 柠檬酸, pH 4.0 的浓度为 2mg/ml 的 RNA 逐滴滴加到挤出的预制大单层囊泡中, 并在 31°C 孵育 30 分钟并持续混合至最终的 RNA/脂质重量比为 0.06/1wt/wt 来实现包封。使用 Spectra/Por 2 再生纤维素透析膜, 通过针对 pH 7.4 的磷酸盐缓冲盐水 (PBS) 透析 16 小时, 以去除乙醇并中和制剂缓冲液。粒径分布可通过使用 NICOMP 370 粒度仪, 囊泡/强度模式和高斯拟合 (Nicomp Particle Sizing, 加利福尼亚州圣巴巴拉) 的动态光散射来确定。所有三个 LNP 系统的粒径都可以是直径约 70nm。RNA 封装效率可以通过使用 VivaPureD MiniH 色谱柱 (Sartorius Stedim Biotech) 从透析前后收集的样品中除去游离 RNA 来确定。可以从洗脱的颗粒中提取包封的 RNA, 并在 260nm 处定量。通过使用来自 Wako Chemicals USA (Richmond, VA) 的胆固醇 E 酶法测定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, PEG 化脂质体或 LNP 同样适用于递送核酸靶向系统或其组分。可以使用/和/或改造自 Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011。可以在含有 DLinKC2-DMA, DSPC 和胆固醇的乙醇中以 50:10:38.5 摩尔比制备脂质预混溶液 (总脂质浓度为 20.4mg/ml)。乙酸钠可以以 0.75:1 的摩尔比 (乙酸钠:DLinKC2-DMA) 添加到脂质预混合物中。随后可以通过在剧烈搅拌下将混合物与 1.85 体积的柠檬酸盐缓冲液 (10mmol/l, pH 3.0) 合并来水化脂质, 从而在含有 35% 乙醇的水性缓冲液中自发形成脂质体。脂质体溶液可以在 37°C 下温育, 以便随时间增加粒径。在孵育过程中的不同时间可以取出等分试样, 以通过动态光散射 (Zetasizer Nano ZS, Malvern Instruments, 伍斯特郡, 英国) 研究脂质体大小的变化。一旦达到所需的粒径, 就可以将 PEG 脂质水溶液 (储备液 = 10mg/ml PEG-DMG 在 35% (体积/体积) 乙醇中的溶液) 添加到脂质体混合物中, 以使最终 PEG 摩尔浓度为 3.5% 总脂质。加入 PEG-脂质后, 脂质体应达到其大小, 从而有效地抑制进一步的生长。然后将 RNA 以大约 1:10 (wt:wt) 的 RNA 与总脂质比率添加到空脂质体中, 然后在 37°C 下孵育 30 分钟以形成负载的 LNP。随后可以将混合物在 PBS 中透析过夜, 并用 0.45- $\mu$ m 注射器过滤器过滤。

[1371] 球形核酸 (SNA<sup>TM</sup>) 构建体和其他颗粒 (尤其是金颗粒) 也被视为将核酸靶向系统递送至预期靶标的手段。大量数据表明, 基于核酸功能化金颗粒的 AuraSense Therapeutics' Spherical Nucleic Acid (SNA<sup>TM</sup>) 构建体是有用的。

[1372] 可结合本发明的教导使用的文献包括 Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80,

Mirkin, *Nanomedicine* 2012 7:635-638 Zhang et al., *J. Am. Chem. Soc.* 2012 134:16488-1691, Weintraub, *Nature* 2013 495:S14-S16, Choi et al., *Proc. Natl. Acad. Sci. USA.* 2013 110(19):7625-7630, Jensen et al., *Sci. Transl. Med.* 5, 209ra152 (2013) and Mirkin, et al., *Small*, 10:186-192.

[1373] 具有RNA的自组装颗粒可以用聚乙烯亚胺(PEI)构建,该PEI与连接在PEG(PEG)远端的Arg-Gly-Asp(RGD)肽配体PEG化。例如,该系统已被用作靶向表达整合素的肿瘤新生血管并递送抑制血管内皮生长因子受体-2(VEGF R2)表达的siRNA的手段,从而实现肿瘤血管生成(参见,例如Schiffelers et al., *Nucleic Acids Research*, 2004, Vol. 32, No. 19)。纳米颗粒可通过混合等体积的阳离子聚合物和核酸的水溶液来制备,以使可电离的氮(聚合物)与磷酸盐(核酸)的净摩尔过量在2至6的范围内。阳离子聚合物和核酸之间的静电相互作用导致形成平均颗粒尺寸分布约为100 nm的聚复合物,因此在此称为纳米复合物。预计约100至200mg的核酸靶向复合物RNA的剂量用于在Schiffelers等的自组装颗粒中递送。

[1374] Bartlett等(*PNAS*, September 25, 2007, vol. 104, no. 39)的纳米复合物也可应用于本发明。Bartlett等的纳米复合物通过混合等体积的阳离子聚合物和核酸的水溶液来制备,使可电离的氮(聚合物)与磷酸盐(核酸)的净摩尔过量在2至6的范围内。阳离子聚合物和核酸之间的静电相互作用导致形成平均颗粒尺寸分布约为100纳米的聚复合物,因此在此称为纳米复合物。Bartlett等的DOTA-siRNA的合成如下:1,4,7,10-四氮杂环十二烷-1,4,7,10-四乙酸单(N-羟基丁二酰亚胺酯)(DOTA-NHSester)由Macrocyclics(得克萨斯州达拉斯)订购。将在碳酸缓冲液(pH 9)中具有100倍摩尔过量DOTA-NHS-酯的胺修饰的RNA有义链加入微量离心管。内容物在室温下搅拌反应4小时。DOTA-RNA正义缀合物经乙醇沉淀,重新悬浮于水中,并与未修饰的反义链退火,产生DOTA-siRNA。所有液体均经Chelex-100(Bio-Rad, Hercules, CA)预处理,以去除痕量金属污染物。Tf靶向和非靶向的siRNA颗粒可通过使用含环糊精的聚阳离子形成。通常,颗粒在水中以3(+/-)的电荷比和0.5g/升的siRNA浓度形成。靶向颗粒表面1%的金刚烷-PEG分子被Tf(金刚烷-PEG-Tf)改性。将颗粒悬浮在5%(重量/体积)的葡萄糖载体溶液中用于注射。

[1375] Davis等(*Nature*, Vol 464, 15 April 2010)进行了一项使用靶向微粒递送系统的RNA临床试验(临床试验注册号NCT00689065)。在21天周期的第1天、第3天、第8天和第10天,通过30分钟的静脉注射,给予对标准护理治疗无效的实体癌症患者一定剂量的靶向颗粒。所述颗粒包含、基本上包含或由合成递送系统组成,所述合成递送系统包含:(1)线性的、基于环糊精的聚合物(CDP)、(2)显示在纳米颗粒外部的人转铁蛋白蛋白(TF)靶向配体,以与癌细胞表面的TF受体(TFR)结合,(3)用于促进生物液体中纳米颗粒稳定性的亲水性聚合物(PEG),和(4)旨在降低RRM2表达的siRNA(临床中使用的序列先前称为siR2B+5)。众所周知,TFR在恶性细胞中的表达上调,而RRM2是公认的抗癌靶标。在对非人类灵长目动物的多剂量研究中,这些颗粒(临床版本称为CALAA-01)被证明具有良好的耐受性。尽管单个慢性髓细胞性白血病患者通过脂质体递送接受了siRNA,但Davis等的临床试验是首次采用靶向递送系统递送siRNA并治疗实体癌症患者的人体试验。为了确定靶向施用系统是否能够向人体肿瘤有效地递送功能性siRNA, Davis等调查了来自三个不同施用队列的三名患者的活组织检查;患者A、B和C均患有转移性黑色素瘤,分别接受了18、24和30mg m-2siRNA CALAA-01剂量。对于本发明的核酸靶向系统,也可以考虑类似的剂量。本发明的递送可通过包含线

性环糊精基聚合物(CDP)、显示在颗粒外部以与癌细胞表面上的TF受体(TFR) 结合的人转铁蛋白蛋白(TF) 靶向配体和/或亲水性聚合物(例如用于促进生物液体中颗粒稳定性的PEG)的颗粒来实现。

[1376] 就本发明而言,优选具有一种或多种RNA靶向复合物的成分,例如核酸靶向效应物(CRISPR-Cas)蛋白或其mRNA,或使用颗粒或脂包膜递送的指导RNA或crRNA。其他递送系统或载体可与本发明的颗粒方面结合使用。本发明包含的颗粒可以不同形式提供,例如固体颗粒(例如银、金、铁、钛等金属)、非金属、脂基固体、聚合物、颗粒悬浮液或其组合。可以制备金属、介电和半导体颗粒以及混合结构(例如核-壳颗粒)。由半导体材料制成的颗粒如果足够小(通常小于10纳米)以至于发生电子能级量化,也可以被标记为量子点。这种纳米颗粒在生物医学应用中用作药物载体或成像剂,并且可适用于本发明的类似目的。

[1377] 已经制造了半固体和软颗粒,并且在本发明的范围内。半固体性质的原型颗粒是脂质体。各种类型的脂质体颗粒目前在临床上用作抗癌药和疫苗的递送系统。一半为亲水性,另一半为疏水性的颗粒称为Janus 颗粒,对稳定乳浊液特别有效。它们可以在水/油界面和担任固体表面活性剂上自组装。

[1378] 美国专利第8,709,843号(通过引用并入本文)提供了一种用于将含治疗剂的颗粒定向递送至组织、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。美国专利第6,007,845号提供了具有多嵌段共聚物核的颗粒,该多嵌段共聚物通过将多功能化合物与一种或多种疏水聚合物和一种或多种亲水聚合物共价连接而形成,并且该颗粒包含生物活性材料。美国专利第5,855,913号(通过引用并入本文)提供了一种颗粒组合物,其具有振实密度小于 $0.4\text{g}/\text{cm}^3$ 、平均直径在5微米至30微米之间的空气动力学轻颗粒,在其表面结合了表面活性剂,用于将药物递送至肺系统。美国专利第5,985,309号(通过引用并入本文)提供了包含表面活性剂和/或带正电或带负电的治疗或诊断试剂的亲水或疏水复合物以及带相反电荷的带电分子的颗粒,用于递送至肺系统。美国专利第5,543,158号(通过引用并入本文)提供了可生物降解的可注射颗粒,其具有包含生物活性材料的固体核和表面聚(亚烷基二醇)部分。WO2012135025(也称为US20120251560)(通过引用并入本文)描述了共轭聚乙烯亚胺(PEI)聚合物和共轭氮杂大环(统称为“共轭脂聚体”或“脂聚体”)。在某些实施方式中,可以预见,本文引用的文件中的这样的方法和材料,例如,共轭的脂聚体可以在核酸靶向系统的上下文中使用,以实现体外、离体和体内基因组扰乱,从而改变基因表达,包括蛋白质表达的调节。

[1379] 外泌体

[1380] 外泌体是转运RNA和蛋白质的内源性纳米囊泡,可将RNA递送至大脑和其他靶器官。为降低免疫原性,Alvarez-Erviti等(2011,Nat Biotechnol 29:341)使用自衍生树状细胞进行外泌体生产。通过对树状细胞进行工程化,使其表达Lamp2b(一种融合于神经元特异性RVG肽的外泌体膜蛋白),从而实现了对大脑的定位。用电穿孔法将纯化的外泌体负载外源性RNA。通过静脉注射以RVG靶向的外泌体,将GAPDH siRNA 特异性递送至大脑中的神经元、小胶质细胞和少突胶质细胞,从而导致特定的基因敲除。预先暴露于RVG外泌体不会减弱敲除作用,也未观察到其他组织的非特异性吸收。外泌体介导的siRNA递送的治疗潜力通过 BACE1(阿尔茨海默病的治疗靶标)的强mRNA(60%)和蛋白(62%)敲除得到证实。

[1381] 为了获得免疫惰性外泌体的池,Alvarez-Erviti等从近交系C57BL/6 小鼠中收获

了具有同质主要组织相容性复合物(MHC)单体型的骨髓。由于未成熟的树状细胞产生大量不含T细胞激活剂(如MHC-II和CD86)的外泌体,Alvarez-Erviti等选择了含有粒细胞/巨噬细胞集落刺激因子(GM-CSF)的树状细胞培养7天。所产生的外泌体在物理上是均一的,通过颗粒追踪分析(NTA)和电子显微术确定,其尺寸分布在直径80nm处达到峰值。Alvarez-Erviti等每 $10^6$ 个细胞获得6-12 $\mu$ g外泌体(基于蛋白质浓度测量)。接下来,Alvarez-Erviti等研究了使用适用于纳米尺度应用的电穿孔方案将外源货物装载到经修饰的外泌体的可能性。由于纳米尺度的膜颗粒电穿孔没有很好的表征,非特异性Cy5标记的RNA被用于电穿孔方案的优化。在超速离心和外泌体裂解后,测定包裹的RNA的量。在400伏和125华氏度电穿孔导致最大的RNA保留,并用于所有后续实验。Alvarez-Erviti等将包裹在150 $\mu$ g RVG外泌体中的150 $\mu$ g每一种 BACE1 siRNA给予正常的C57BL/6小鼠,并将敲除效率与四个对照进行比较:未处理的小鼠、仅注射了RVG外泌体的小鼠、注射了与活体内阳离子脂质体试剂复合的BACE1 siRNA的小鼠和注射了与RVG-9R复合的 BACE1 siRNA的小鼠,该RVG肽与9个D-argines缀合,并与siRNA静电结合。在施用后3d对皮质组织样品进行分析,观察到经siRNA-RVG-9R处理的和经siRNA RVG外泌体处理的小鼠的显著蛋白敲除(分别为45%, $P < 0.05$ 和62%, $P < 0.01$ ),这是由于BACE1 mRNA水平显著降低(分别为66% [+或-]15%, $P < 0.001$ 和61% [+或-]13%, $P < 0.01$ )。此外,申请人证明,在接受RVG外泌体治疗的动物中,淀粉样蛋白1-42的总水平显著降低(55%, $P < 0.05$ ),该淀粉样蛋白1-42是老年痴呆症中淀粉样蛋白斑的主要成分。观察到的下降幅度大于正常小鼠脑室注射BACE1抑制剂后表现出的 $\beta$ -淀粉样蛋白1-40的下降幅度。Alvarez-Erviti等对BACE1切割产物进行了5'快速扩增,这为RNAi介导的siRNA敲除提供了证据。最后,Alvarez-Erviti等通过评估IL-6、IP-10、TNF $\alpha$ 和IFN- $\alpha$ 的血清浓度,研究了RNA-RVG外泌体是否在体内诱导免疫反应。在外泌体治疗后,所有细胞因子的无显著变化均与siRNA转染试剂治疗相似,与siRNA-RVG-9R相反,其有效刺激了IL-6的分泌,证实了外泌体治疗的免疫惰性特征。鉴于外泌体仅包裹20%的siRNA,RVG外泌体的递送似乎比RVG-9R递送更有效,因为在没有相应水平的免疫刺激的情况下,相当的mRNA敲除和蛋白质敲除通过少五倍的siRNA取得。本实验展示了RVG外泌体技术的治疗潜力,该技术可能适合于神经退化性疾病相关基因的长期沉默。Alvarez-Erviti等的外泌体递送系统可用于将本发明的核酸靶向系统递送至治疗靶标,尤其是神经退化性疾病。本发明可考虑将约100至1000mg的核酸靶向系统包裹在约100至1000 mg的RVG外泌体中。

[1382] El-and loussi等(Nature Protocols 7,2112-2126(2012))提供了来源于培养细胞的外泌体,用于在体外和体内递送RNA。该方案首先描述了通过转染表达载体产生靶向外泌体,该表达载体包含与肽配体融合的外泌体蛋白。接下来,El-and loussi等解释了如何从转染的细胞上层清液中纯化和鉴定外泌体。接下来,El-and loussi等详述了将RNA装载到外泌体的关键步骤。最后,El-and loussi等概述了如何使用外泌体在体外和体内有效地在小鼠脑中递送RNA。还提供了通过功能性测定和成像评估外泌体介导的RNA递送的预期结果的例子。整个方案耗时3周。根据本发明的递送或施用可以使用自衍生树状细胞产生的外泌体进行。根据本文的教导,这可以用于本发明的实践中。

[1383] 在另一个实施方式中,考虑了Wahlgren et al.(Nucleic Acids Research, 2012,Vol.40,No.17e130)的血浆外泌体。外泌体是由多种细胞类型(包括树状细胞、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30-90纳米)。这些小泡由后期内体向内出芽形成,然后在与质膜融合后释放至细胞外环境。由于外泌体在细胞间自然携带 RNA,这一性质可用于基因治疗,并且从本公开中可用于本发明的实践。通过在900g离心血沉棕黄层20分钟以分离血浆,然后收集细胞上清液,在300g离心10分钟以去除细胞,在16 500g离心30分钟,然后通过0.22 mm过滤器过滤,可以制备来自血浆的外泌体。通过120,000g超速离心 70分钟使外泌体沉淀。根据制造商在RNAi Human/Mouse Starter Kit (Quiagen,Hilden,Germany)中的说明,将siRNA化学转染入外泌体。将 siRNA以2mmol/ml的最终浓度加入100ml的PBS中。在加入HiPerFect 转染试剂后,混合物在室温下孵育10分钟。为了除去过量的胶团,使用醛/硫酸盐乳胶珠重新分离外泌体。核酸靶向系统向外泌体的化学转染可以类似于siRNA进行。外泌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同培养。因此,可以考虑将含有核酸靶向系统的外泌体引入人的单核细胞和淋巴细胞,并自动重新引入人。因此,根据本发明的递送或施用可以使用血浆外泌体进行。

#### [1384] 脂质体

[1385] 根据本发明的递送或施用可以用脂质体进行。脂质体是由围绕内部水室的单层或多层脂双层和相对不渗透的外部亲脂磷脂双层组成的球形囊泡结构。由于脂质体具有生物相容性、无毒性、既能递送亲水性又能递送亲脂性药物分子、保护其货物免受血浆酶的降解、并能将其货物运输通过生物膜和血脑屏障(BBB),因此作为药物递送载体受到了广泛关注(例如,参见Spuch and Navarro,Journal of Drug Delivery,vol.2011,Article ID 469679,12pages,2011.doi:10.1155/2011/469679供审阅)。脂质体可由几种不同类型的脂质制成;然而,磷脂最常用于生成作为药物载体的脂质体。尽管当脂质膜与水溶液混合时,脂质体的形成是自发的,但也可以通过使用均质器、超声波发生器或挤压设备施加振动形式的力来加速形成(例如,参见Spuch and Navarro,Journal of Drug Delivery,vol.2011, Article ID 469679,12pages,2011.doi:10.1155/2011/469679供审阅)。

[1386] 为了改变脂质体的结构和性质,可以在脂质体中加入其他几种添加剂。例如,可以向脂质体混合物中加入胆固醇或鞘磷脂,以帮助稳定脂质体结构并防止脂质体内部货物的渗漏。此外,脂质体由氢化的卵磷脂酰胆碱或卵磷脂酰胆碱、胆固醇和二乙酰磷酸酯制备,并且其平均囊泡尺寸被调整至约50和100nm。(参见,例如,Spuch and Navarro,Journal of Drug Delivery,vol.2011,Article ID 469679,12pages,2011. doi:10.1155/2011/469679供审阅)。脂质体制剂可主要由天然磷脂和脂质组成,例如1,2-二硬脂酰-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。由于该制剂仅由磷脂组成,脂质体制剂遇到了许多挑战,其中之一是血浆的不稳定性。已经进行了多次尝试来克服这些挑战,特别是在操纵脂质膜方面。其中一项尝试的重点是控制胆固醇。向常规制剂中添加胆固醇可减少胶囊化生物活性化合物向血浆中的快速释放,或1,2-二油酰-sn-甘油-3-磷酸乙醇胺(DOPE)可提高稳定性(例如,参见Spuch and Navarro,Journal of Drug Delivery,vol.2011, Article ID 469679,12pages,2011.doi:10.1155/2011/469679供审阅)。在一个特别有利的实施方式中,特洛伊木马脂质体(也称为分子特洛伊木马)是理想的,并且可以在 [cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long](http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long)找到方案。这些颗粒允许在血管内注射后将转基因递送至整个大脑。不受限制地,据信具有结合于

表面的特定抗体的中性脂质颗粒允许通过内吞作用穿过血脑屏障。申请人假设利用特洛伊木马脂质体通过血管内注射将CRISPR-Cas 复合物递送至大脑,这将允许全脑转基因动物无需胚胎操作。约1-5g的 DNA或RNA可被考虑用于脂质体的体内施用。

[1387] 在另一个实施方式中,核酸靶向系统或其组分可以脂质体形式施用,例如稳定的核酸-脂质颗粒(SNALP)(参见,例如,Morrissey et al.,*Nature Biotechnology*,Vol.23, No.8, August 2005)。预计每天静脉注射在SNALP 中靶向的约1、3或5mg/kg/天的特定核酸靶向系统。每日治疗可超过约三天,然后每周治疗约五周。在另一个实施方式中,也考虑了通过静脉注射以约1或2.5mg/kg的剂量施用的特定核酸靶向系统(例如,参见 Zimmerman et al.,*Nature Letters*,Vol.441,4May 2006)。SNALP制剂可包含3-N-[(w-乙氧基聚(乙二醇)2000)氨基甲酰基]-1,2-二嘧啶氧基丙胺(PEG-丙甲氧基)、1,2-二氢氧基-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,其摩尔比为2:40:10:48(参见Zimmerman et al.,*Nature Letters*,Vol.441,4May 2006)。在另一个实施方式中,稳定的核酸-脂质颗粒已被证明是高度血管化的HepG2衍生的肝肿瘤的有效递送分子,但在血管化不良的HCT-116衍生的肝肿瘤中并非如此(参见,例如Li,*Gene Therapy* (2012) 19,775-780)。SNALP脂质体可通过将D-Lin-DMA和PEG-C-DMA与二硬脂酰磷脂酰胆碱(DSPC)、胆固醇和siRNA按25:1的脂质/siRNA比率和48/40/10/2的胆固醇 /D-Lin-DMA/DSPC/PEG-C-DMA的摩尔比率配制而成。生成的SNALP 脂质体的大小约为80-100nm。在另一个实施方式中,SNALP可包含合成的胆固醇(Sigma-Aldrich,St Louis,MO, USA)、二棕榈酰磷脂酰胆碱(美国Alabaster,AL)、3-N-[(w-甲氧基聚(乙二醇)2000)氨基甲酰基]-1,2-二苯乙炔基丙胺,以及阳离子1,2-二氢丙酰氧基-3-N,N-二甲基氨基丙烷(例如参见Geisbert等,*Lancet* 2010375:1896-905)。可以考虑以例如静脉推注的方式施用每次剂量约为2mg/kg总核酸靶向系统。在又一实施方式中,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1,2-二亚油酰氧基-3-(N,N-二甲基)氨基丙烷(DLinDMA)(例如,参见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质量比率。

[1388] RNAi纳米药物的安全性概况已由Alnylam Pharmaceuticals的Barros 和Gollob 审阅(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——一种在低pH值下呈阳离子的可电离脂质(DLinDMA)、一种中性辅助脂质、胆固醇和一种可扩散的PEG脂质。该颗粒的直径约为80nm,在生理pH 值下为中性电荷。在制剂过程中,可电离的脂质在颗粒形成过程中与阴离子RNA发生凝聚。当在酸性越来越强的内体条件下带正电时,可电离的脂质还介导了SNALP与内体膜的融合,从而使RNA释放到胞质中。PEG脂质稳定了颗粒并减少了制剂过程中的聚集,随后提供了中性的亲水性外观,改善了药物动力学性质。迄今为止,已启动了两个使用含RNA的SNALP制剂的临床项目。Tekmira Pharmaceuticals最近完成了一项针对低密度脂蛋白胆固醇升高的成年志愿者的单剂量研究。ApoB主要在肝脏和空肠表达,对VLDL和LDL的组装和分泌至关重要。17名受试者接受了单一剂量的SNALP-ApoB(跨7个剂量水平的剂量递增)。没有证据表明存在肝毒性(根据临床前研究,预计为潜在的剂量限制毒性)。最高剂量的两个受试者中有一个出现了与免疫系统刺激相一致的流感样症状,并决定结束试验。Alnylam

Pharmaceuticals具有类似的先进的ALN-TTR01,其采用上述SNALP技术,并针对突变型和野生型TTR的肝细胞生成治疗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)——均由TTR的常染色体显性突变引起;以及由野生型TTR引起的老年系统性淀粉样变(SSA)。最近在ATTR患者中完成了一项安慰剂对照、单次剂量递增的ALN-TTR01第一阶段试验。对31名患者(23名使用研究药物,8名使用安慰剂)进行了15分钟的静脉注射ALN-TTR01,剂量范围为0.01-1.0mg/kg(基于siRNA)。治疗耐受性良好,肝功能检查无明显增加。23例患者中有3例出现 $\geq 0.4$ mg/kg的输液相关反应;所有受试者均对输注速率减慢作出反应,并且均继续进行研究。在最高剂量为1mg/kg的两名患者中发现了血清细胞因子IL-6、IP-10和IL-1ra的轻微和短暂升高(根据临床前和NHP研究的预期)。在1mg/kg的剂量下观察到血清TTR(ALN-TTR01的预期药效)降低。

[1389] 在另一个实施方式中,可通过将阳离子脂质、DSPC、胆固醇和PEG脂质(例如分别以40:10:40:10的摩尔比)溶解在乙醇中来制备SNALP(参见, Semple et al., Nature Nanotechnology, Volume 28 Number 2 February 2010, pp.172-177)。将脂质混合物加入水性缓冲液(50mM柠檬酸盐, pH 4)中,混合至最终的乙醇和脂质浓度分别为30%(体积/体积)和6.1mg/ml,并在挤出前在22°C平衡2分钟。使用Lipex挤塑机(Northern Lipids)在22°C下通过两个堆叠的80纳米孔径过滤器(Nuclepore)挤出水合脂质,直至通过动态光散射分析确定的囊泡直径达到70-90纳米。这通常需要1-3次通过。在混合的情况下,将siRNA(溶解于含30%乙醇的50mM柠檬酸钠pH 4水溶液中)以约5ml/min的速率加入到预平衡(35°C)的囊泡中。在达到0.06(重量/重量)的最终目标siRNA/脂质比率后,将混合物在35°C下再保温30分钟,以使囊泡重组并包裹siRNA。然后通过透析或切向流渗滤去除乙醇,并用PBS(155mM氯化钠, 3mM Na<sub>2</sub>HPO<sub>4</sub>, 1mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.5)替换外部缓冲液。使用受控的逐步稀释法工艺将siRNA封装在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,以摩尔比为57.1:7.1:34.3:1.4使用。在形成负载颗粒后,将SNALP对PBS进行透析,并在使用前通过0.2  $\mu$ m的过滤器进行过滤消毒。平均颗粒尺寸为75-85纳米,90-95%的siRNA被包裹在脂质颗粒中。用于体内测试的制剂中的最终siRNA/脂质比率约为0.15(重量/重量)。在使用前立即将含有因子VII siRNA的LNP-siRNA系统稀释至适当浓度,并通过尾侧静脉以10ml/kg的总体积静脉内施用。该方法和这些递送系统可外推到本发明的核酸靶向系统。

[1390] 其他脂质

[1391] 其他阳离子脂质,例如氨基脂质2,2-二甲醇-4-二甲胺基乙基-[1,3]-二氧戊环(DLin-KC2-DMA)可用于包裹核酸靶向系统或其自身或编码其的核酸分子,例如类似于siRNA(参见J Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529-8533),因此可用于本发明的实践。可考虑具有以下脂质组成的预成型囊泡:氨基脂质、二硬脂酰磷脂酰胆碱(DSPC)、胆固醇和(R)-2,3-双(十八氧基)丙-1-(甲氧基聚(乙二醇)2000)丙氨基甲酸酯(PEG-脂质),其摩尔比率为40/10/40/10,且FVII siRNA/总脂质比率为约0.05(w/w)。为确保70-90nm范围内的窄颗粒尺寸分布和 $0.11 \pm 0.04$  (n=56)的低多分散指数,在加入指导RNA之前,颗粒可通过80纳米的膜挤出三次。可使用含有高效氨基脂质16的颗粒,其中四种脂质成分16、DSPC、胆固醇和PEG脂质的摩尔比率(50/10/38.5/1.5)可进一步优化至提升体内活性。

[1392] Michael S D Kormann等(Expression of therapeutic proteins after delivery of chemically modified mRNA in mice:Nature Biotechnology, Volume:29, Pages:154-157 (2011))描述了使用脂包膜递送RNA的方法。在本发明中也优选使用脂包膜。

[1393] 在另一个实施方式中,可以用本发明的RNA靶向系统(CRISPR-Cas13复合物,即与crRNA复合的Cas13)或其组分或编码其的核酸分子来制备脂质,以形成脂质纳米颗粒(LNPs)。脂质包括但不限于 DLin-KC2-DMA4、C12-200和二硬脂酰磷脂酰胆碱、胆固醇和PEG-DMG,可以用RNA靶向系统而非siRNA系统进行配制(例如,参见 Novobrantseva, Molecular Therapy-Nucleic Acids (2012) 1,e4; doi:10.1038/mtna.2011.3)使用自发囊泡形成程序。组分摩尔比可以为约 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DLin-KC2-DMA和C12-200脂质颗粒的情况下,最终脂质与siRNA的重量比为约12:1和9:1。制剂的平均颗粒直径约为80纳米,包封率大于90%。可考虑3毫克/千克的剂量。Tekmira在美国和海外拥有约95个专利家族,涉及LNP和LNP制剂的各个方面(例如,参见美国专利号第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和欧洲专利号1766035;1519714;1781593和1664316),所有这些都可以用于和/或适用于本发明。

[1394] 可将RNA靶向系统或其组分或编码其的核酸分子包裹在PLGA微球中递送,例如美国公开申请20130252281和20130245107和 20130244279(转让给Moderna Therapeutics)中进一步描述的,该申请涉及包含可编码蛋白质、蛋白质前体或部分或完全加工形式的蛋白质或蛋白质前体的经修饰的核酸分子的组合物的制剂方面。制剂的摩尔比为 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。融合脂质可能是DSPC。另请参见Schrum et al., Delivery and Formulation of Engineered Nucleic Acids 和美国公开申请20120251618。

[1395] Nanomerics技术解决了生物利用度挑战,涉及广泛的治疗方法,包括低分子量疏水药物、肽和基于核酸的治疗方法(质粒、siRNA、miRNA)。该技术已显示出明显优势的具体施用途径包括口服途径、跨血脑屏障转运、递送至实体肿瘤以及眼睛。参见,例如,Mazza et al.,2013,ACS Nano. 2013Feb 26;7(2):1016-26;Uchegbu and Siew,2013,J Pharm Sci. 102(2):305-10and Lalatsa et al.,2012,J Control Release.2012Jul 20; 161(2):523-36。

[1396] 美国专利公开号20050019923描述了用于向哺乳动物身体递送生物活性分子(例如多聚核苷酸分子、肽和多肽和/或药剂)的阳离子树状大分子。树状大分子适合于将生物活性分子靶向递送至例如肝、脾、肺、肾或心脏(或甚至大脑)。树状大分子是由简单的支化单元逐步制备的合成三维大分子,其性质和功能易于控制和改变。树状大分子是通过向多功能核中重复添加结构单元(分散合成法)或向多功能核中重复添加结构单元(聚合合成法)来合成的,并且每次添加结构单元的三维壳都会导致形成更高代的树状大分子。聚丙烯胺树状大分子从二氨基丁烷核开始,通过向伯胺中双迈克尔加成丙烯腈,然后氢化腈,向其中加入两倍数量的氨基。这导致氨基加倍。聚丙烯胺树状大分子包含100%的可质子化氮和最多64个末端氨基(第5代,DAB 64)。可质子化的基团通常是能够在中性pH值下接受质子的胺基。树状大分子作为基因递送剂的使用主要集中在聚酰胺胺的使用上。以及分别以胺/酰胺或N-P(O<sub>2</sub>)S的混合物作为缀合单元的含磷化合物,但没有关于使用下一代聚丙烯胺树

状大分子进行基因递送。聚丙烯胺树状大分子还被研究为用于药物递送的对pH 敏感的控释系统,以及当其被外周氨基酸基团化学改性时用于包裹客体分子。还研究了聚丙烯胺树状大分子与DNA的细胞毒性和相互作用以及 DAB 64的转染效率。美国专利公开号 20050019923基于如下观察:与先前的报告相反,阳离子树状大分子(例如聚丙烯胺树状大分子)显示出适合用于生物活性分子(例如遗传物质)的靶向递送的性质,例如特定的靶向性和低毒性。另外,阳离子树状大分子的衍生物也显示出适合于生物活性分子的靶向递送的性质。另请参见生物活性聚合物,美国公布的申请 20080267903,该申请公开了“各种聚合物,包括阳离子聚胺聚合物和树状聚合物,显示出具有抗增殖活性,因此可用于治疗以不良细胞增殖为特征的疾病,例如肿瘤和肿瘤、炎症疾病(包括自身免疫疾病)、银屑病和动脉粥样硬化。这样的聚合物可单独用作活性剂,或用作其他治疗剂(例如用于基因治疗的药物分子或核酸)的递送载体。在这种情况下,聚合物自身固有的抗肿瘤活性可以补充待递送药物的活性”。这些专利出版物的公开内容可结合本文的教导用于递送核酸靶向系统或其组分或编码其的核酸分子。

#### [1397] 超电荷蛋白质

[1398] 超电荷蛋白质是一类具有异常高的正或负理论净电荷的工程或天然存在的蛋白质,可用于递送核酸靶向系统或其组分或编码其的核酸分子。超负电荷和超正电荷的蛋白质均表现出抵抗热或化学诱导聚集的显著能力。超正电荷蛋白质也能够穿透哺乳动物细胞。将货物与这些蛋白质(例如质粒DNA、RNA或其他蛋白质)结合,能够使这些大分子在体外和体内功能性地递送至哺乳动物细胞。David Liu实验室于2007年报告了超电荷蛋白质的产生和特性(Lawrence et al.,2007,Journal of the American Chemical Society 129,10110-10112)。

[1399] 将RNA和质粒DNA非病毒递送至哺乳动物细胞对研究和治疗应用均有价值(Akinc et al.,2010,Nat.Biotech.26,561-569)。将纯化的+36GFP 蛋白(或其他超正电荷的蛋白)与RNA在适当的无血清介质中混合,并在加入细胞前进行复合。在此阶段加入血清会抑制超电荷蛋白-RNA复合物的形成,降低治疗效果。以下方案已被发现对多种细胞系有效(McNaughton et al.,2009,Proc.Natl.Acad.Sci.USA 106,6111-6116)。但是,应进行改变蛋白质和RNA剂量的试点实验,以优化特定细胞系的程序。(1) 治疗前一天,在48孔平板中每孔平板 $1 \times 10^5$ 个细胞。(2) 处理当天,将纯化的+36GFP蛋白在无血清介质中稀释至最终浓度200nM。将RNA 添加至最终浓度50nM。涡旋混合并在室温下培养10分钟。(3) 培养过程中,从细胞中吸取培养液,用PBS清洗一次。(4) 培养+36GFP和RNA 后,向细胞中加入蛋白质-RNA复合物。(5) 将细胞与复合物在37℃培养 4小时。(6) 培养后,吸取培养液,用20U/mL肝素PBS洗涤三次。根据活性检测,将细胞与含血清的培养基再培养48小时或更长时间。(7) 通过免疫印迹、qPCR、表型分析或其他适当的方法分析细胞。

[1400] +36GFP被发现在一系列细胞中是有效的质粒递送试剂。另请参见,例如McNaughton et al.,Proc.Natl.Acad.Sci.USA 106,6111-6116(2009); Cronican et al.,ACS Chemical Biology 5,747-752(2010);Cronican et al., Chemistry&Biology 18,833-838(2011);Thompson et al.,Methods in Enzymology 503,293-319(2012); Thompson,D.B.,et al.,Chemistry& Biology 19(7),831-843(2012)。超电荷蛋白的方法可用于和/或适用于递送本发明的RNA靶向系统或其组分或编码其的核酸分子。

#### [1401] 细胞渗透肽 (CPP)

[1402] 在另一个实施方式中,细胞渗透肽 (CPP) 被考虑用于CRISPR Cas系统的递送。CPP是促进细胞吸收各种分子货物(从纳米颗粒到小化学分子和大DNA片段)的短肽。本文中使用的术语“货物”包括但不限于由治疗剂、诊断探针、肽、核酸、反义寡核苷酸、质粒、蛋白质、颗粒(包括纳米颗粒)、脂质体、发色团、小分子和放射性材料组成的组。在本发明的方面,货物还可以包括CRISPR Cas系统的任何组件或整个功能性 CRISPR Cas系统。本发明的方面进一步提供了用于将所需货物递送至受试者的方法,包括:(a) 制备包含本发明的细胞渗透肽和所需货物的复合物,和(b) 经口、关节内、腹膜内、鞘内、动脉内、鼻内、实质内、皮下、肌内、静脉内、皮肤内、直肠内或局部给予受试者该复合物。货物通过共价键的化学连接或非共价相互作用与肽结合。CPP的功能是将货物运送到细胞内,这一过程通常通过将货物运送到活的哺乳动物细胞的内体而发生。细胞渗透肽具有不同的大小、氨基酸序列和电荷,但所有CPP 都有一个独特的特征,即能够转运质膜并促进各种分子货物向胞质或细胞器的递送。CPP转运可分为三种主要的进入机制:直接渗透入膜、胞吞作用介导的进入和通过形成过渡结构的转运。CPP作为药物递送剂已在医学中发现了许多应用,用于治疗不同疾病,包括癌症和病毒抑制剂,以及用于细胞标记的对比剂。后者的例子包括作为GFP、MRI相差剂或量子点的载体。CPP作为体外和体内递送载体在研究和医学中具有巨大潜力。CPP的氨基酸组成通常包含相对丰富的带正电的氨基酸,例如赖氨酸或精氨酸,或者包含极性/带电的氨基酸和非极性、疏水的氨基酸的交替模式的序列。这两类结构分别称为聚阳离子或两亲性结构。第三类 CPP是仅含非极性残基、净电荷低或具有对细胞吸收至关重要的疏水氨基酸基团的疏水肽。最初发现的CPP之一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),发现其在培养过程中被多种细胞类型有效地从周围介质中吸收。自那时起,已知CPP的数量大幅增加,并且产生了具有更有效的蛋白质转导性质的小分子合成类似物。CPP包括但不限于渗透素、Tat (48-60)、Transportan和(R-AhX-R4) (AhX=氨基己酰)。

[1403] 美国专利8,372,951提供了一种来源于嗜酸细胞阳离子蛋白(ECP)的 CPP,其表现出高细胞渗透效率和低毒性。还提供了将CPP及其货物递送给脊椎动物受试者的方面。美国专利8,575,305;8,614,194和8,044,019 中描述了CPP及其递送的其他方面。CPP可用于递送CRISPR-Cas系统或其组件。“Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA”, by Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor, et al. *Genome Res.* 2014 Apr 2. [Epub ahead of print] 一文也提供了CPP可用于递送CRISPR-Cas系统或其组件,通过引用并入本文,其中证明了用CPP缀合的重组Cas9蛋白和CPP复合的指导RNA 处理会导致人细胞系的内源性基因破坏。在本文中,Cas9蛋白通过硫醚键与CPP缀合,而指导RNA与CPP复合,形成凝聚的带正电的颗粒。研究表明,用经修饰的Cas9和指导RNA同时和顺序处理人细胞,包括胚胎干细胞、皮肤成纤维细胞、HEK293T细胞、HeLa细胞和胚胎癌细胞,可导致有效的基因破坏,减少与质粒转染相关的非靶突变。CPP递送可用于本发明的实践。

#### [1404] 植入式设备

[1405] 在另一个实施方式中,还考虑了可植入装置用于递送RNA靶向系统或其组分或编码其的核酸分子。例如,美国专利公开20110195123公开了一种可植入的医疗装置,其在局部和长时间内洗脱药物,包括该装置的几种类型、实施的治疗模式和植入方法。该装置包括

用作装置主体的聚合物基底,例如基质,和药物,在某些情况下还包括额外的支架材料,例如金属或额外的聚合物,以及提升可见性和成像材料。植入式递送装置可有利于提供局部和长时间的释放,其中药物被直接释放至患病区域的细胞外基质(ECM),例如肿瘤、炎症、退化或用于症状性目的,或用于受损的平滑肌细胞,或用于预防。如上所述,一种药物是RNA,该系统可用于/和/或适用于本发明的核酸靶向系统。在一些实施方式中,植入的模式是现有的植入程序,这些植入程序目前被开发并用于其他治疗,包括近距离放射治疗和针活组织检查。在这种情况下,本发明中描述的新植入体的尺寸与原始植入体相似。通常在同一治疗过程中植入一些装置。美国专利公开20110195123提供了一种可植入或可插入的药物递送系统,包括适用于例如腹腔的腔和/或药物递送系统未被固定或连接的任何其他类型的施用的系统,包括生物稳定和/或可降解和/或生物可吸收的聚合物基质,其可以例如可选地为基质。应当注意,术语“插入”也包括植入。药物递送系统优选实施为美国专利公开20110195123中所述的“Loder”。所述一种或多种聚合物是生物相容的,包含试剂和/或多种试剂,能够以受控的速率释放试剂,其中在一些实施方式中,聚合物基质(例如基质)的总体积可选地且优选地不大于允许达到试剂治疗水平的最大体积。作为非限制性示例,该体积优选在 $0.1\text{m}^3$ 至 $1000\text{mm}^3$ 的范围内,这是试剂负载的体积所要求的。Loder可以可选地更大,例如当与尺寸由功能决定的装置结合时,例如但不限于膝关节、宫内或宫颈环等。在一些实施方式中,药物递送系统(用于递送组合物)被设计成优选使用可降解的聚合物,其中主要的释放机制是整体侵蚀;或者在一些实施方式中,使用不可降解或缓慢降解的聚合物,其中主要的释放机制是扩散而非整体侵蚀,使得外部部分起到膜的作用,其内部部分起到药物储库的作用,其实际上不受周围环境的长期影响(例如从约一周至约几个月)。也可选择使用具有不同释放机制的不同聚合物的组合。在整个药物释放期间的重要期间,表面的浓度梯度优选有效地保持不变,因此扩散速率有效地保持不变(称为“零模式”扩散)。术语“恒定”是指扩散速率优选保持在治疗有效性的较低阈值之上,但其仍可选择性地以初始突发为特征和/或可能波动,例如增加或减少至一定程度。扩散速率优选在较长时间内保持恒定,并且可以认为其在一定水平上保持恒定,以优化治疗有效期,例如有效沉默期。药物递送系统可选地且优选地被设计为保护基于核苷酸的治疗剂免于降解,无论是化学性质的还是由于受试者体内的酶和其他因素的攻击。美国专利公开20110195123的药物递送系统可选地与在装置植入时和/或植入后通过非侵入性和/或最小侵入性的激活和/或加速/减速方法操作的感测和/或激活设备相关联,例如可选地包括但不限于加热和冷却、激光束和超声波,包括聚焦超声波和/或射频(射频)方法或装置。根据美国专利公开20110195123的一些实施方式,用于局部递送的部位可选择性地包括以细胞高度异常增殖和抑制细胞凋亡为特征的靶部位,包括肿瘤、包括自身免疫疾病状态在内的活动和/或慢性炎症和感染、包括肌肉和神经组织在内的退化组织、慢性疼痛、退化部位、骨折部位和用于促进组织再生的其他伤口部位,以及受损的心脏、平滑肌和横纹肌。用于植入组合物的部位或靶部位优选具有足够小的半径、面积和/或体积,以用于靶局部递送。例如,靶部位的直径可选地在约0.1毫米至约5厘米的范围内。优选选择靶部位的位置以获得最大的治疗效果。例如,药物递送系统的组成(可选地具有如上所述的用于植入的装置)可选地且优选地植入肿瘤环境内或附近,或与其相关的血液供应。例如,组合物(可选地与装置一起)可选地通过乳头植入到胰腺、前列腺、乳房、肝脏内或其附近、血管系统内等。靶位置可选地从包括以下、基本上包括以下或由以

下(仅作为非限制性示例,因为可选地身体内的任何部位可能适合植入Loder)组成的组中选择:1.位于退化部位的大脑,如基底核、白质和灰质的帕金森病或老年痴呆症;2.脊柱,如肌萎缩性侧索硬化(ALS);3.预防HPV感染的子宫颈;4.活动期和慢性炎性关节;5.真皮,如银屑病;6.具有镇痛作用的交感和感觉神经部位;7.骨内植入;8.急性和慢性感染部位;9.阴道内;10.内耳-听觉系统、内耳曲径、前庭系统;11.气管内;12.心脏内;冠状动脉、心外膜;13.膀胱;14.胆系;15.实质组织,包括但不限于肾、肝、脾;16.淋巴结;17.唾液腺;18.牙龈;19.关节内(进入关节);20.眼内;21.脑组织;22.脑室;23.空腔,包括腹腔(例如但不限于卵巢癌);24.食管内和25.直肠内。

[1406] 任选地,系统(例如包含组合物的装置)的插入与在靶部位和该部位附近向外基质注射材料相关联,以影响局部pH和/或温度和/或影响药物在外基质、靶部位和该部位附近的扩散和/或药物动力学的其他生物因素。可选地,根据一些实施方式,所述试剂的释放可与在插入之前和/或插入时和/或插入之后通过非侵入性和/或最小侵入性和/或其他激活和/或加速/减速方法(包括激光束、辐射、加热和冷却)和超声波(包括聚焦超声波和/或射频(射频)方法或装置)以及化学激活剂操作的感测和/或激活设备相关联。

[1407] 根据可用于实施本发明的美国专利公开20110195123的实施方式,该药物优选包含一种RNA,例如用于如下所述的乳腺、胰腺、脑、肾、膀胱、肺和前列腺的局部癌症病例。尽管以RNAi为例,但许多药物适用于被包裹在Loder中,并且可与本发明结合使用,只要这样的药物可被包裹在Loder基质(例如基质)中,并且该系统可被用于和/或适于递送本发明的核酸靶向系统。作为具体应用的另一个例子,神经和肌肉退化性疾病因基因表达异常而发展。局部递送RNA可能具有干扰这种异常基因表达的治疗性质。局部递送抗凋亡、抗炎和抗退化药物(包括小药物和大分子)也可以任选地具有治疗作用。在这种情况下,Loder以恒定速率和/或通过单独植入的专用装置申请长期释放。

[1408] 所有这些均可用于和/或适用于本发明的RNA靶向系统。本文讨论的可植入装置技术与本文的教导一起使用,因此通过本公开和本领域的知识,CRISPR-Cas13系统或其复合物或成分或其核酸分子或编码或提供成分可通过可植入装置递送。

[1409] 聚合物基颗粒

[1410] 本文的系统 and 组合物可以使用基于聚合物的颗粒(例如,纳米颗粒)递送。在一些实施方式中,基于聚合物的颗粒可以模拟膜融合的病毒机制。基于聚合物的颗粒可以是流感病毒机器的合成拷贝,并与细胞通过内吞途径(包括形成酸室的过程)吸收的各种类型的核酸(siRNA、miRNA、质粒DNA或shRNA、mRNA)形成转染复合物。后期内体的低pH值起到化学交换机的作用,使颗粒表面疏水并促进膜交叉。一旦进入胞质溶胶,颗粒释放其有效载荷用于细胞活动。这种主动内体逃避技术是安全的,并且由于使用了自然吸收途径,因此最大限度地提高了转染效率。在一些实施方式中,基于聚合物的颗粒可包含烷基化和碳烷基化的支化聚乙烯亚胺。在一些实施方式中,基于聚合物的颗粒为VIROMER,例如 VIROMER RNAi、VIROMER RED、VIROMER mRNA、VIROMER CRISPR。在此递送系统和组合物的示例性方法包括Bawage SS等中所述的方法,合成表达的Cas13a减轻了RNA病毒感染, [www.biorxiv.org/content/10.1101/370460v1.full](http://www.biorxiv.org/content/10.1101/370460v1.full) doi: doi.org/10.1101/370460, **Viomer®** RED,一种用于角蛋白细胞转染的强有力工具。doi:10.13140/RG.2.2.16993.61281, **Viomer®**

Transfection- Factbook 2018:technology,product overview,users'data., doi: 10.13140/RG.2.2.23912.16642。

[1411] 载体

[1412] 在某些方面,本发明涉及载体,例如用于在细胞中递送或引入CRISPR-Cas和/或能够指导CRISPR-Cas至靶位点的RNA(即指导RNA),以及用于繁殖这些成分(例如在原核细胞中)。这里使用的“载体”是允许或促进实体从一个环境转移到另一个环境的工具。它是一种复制子,例如质粒、噬菌体或粘粒,可以在其中插入另一个DNA片段,从而实现插入片段的复制。通常,当载体与适当的控制元件相关联时,载体能够复制。一般而言,术语“载体”指能够运输与其连接的另一种核酸的核酸分子。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、无自由端(例如圆形)的核酸分子;包含DNA、RNA或两者的核酸分子;和本领域已知的其他多聚核苷酸。一种载体类型是“质粒”,指可插入额外DNA片段的环状双链DNA环,例如通过标准分子克隆技术。另一类载体是病毒载体,其中病毒衍生的DNA或RNA序列存在于载体中,用于包装成病毒(例如,反转录病毒、复制缺陷型反转录病毒、腺病毒、复制缺陷型腺病毒和腺相关病毒)。病毒载体还包括病毒携带的用于转染入宿主细胞的多核苷酸。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附加体哺乳动物载体)。其他载体(例如,非附加体哺乳动物载体)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有效连接的基因的表达。该载体在本文中称为“表达载体”。重组DNA 技术中常用的表达载体通常为质粒形式。

[1413] 重组表达载体可包含本发明的核酸,其形式适合于在宿主细胞中表达该核酸,这意味着重组表达载体包含一种或多种调节元件,其可基于待用于表达的宿主细胞进行选择,并与待表达的核酸序列有效连接。在重组表达载体中,“可操作地连接”是指感兴趣的核苷酸序列以允许该核苷酸序列表达的方式(例如,在体外转录/翻译系统中或当载体被引入宿主细胞时在宿主细胞中)连接至调节元件。关于重组和克隆方法,请参见美国专利申请10/815,730,该申请于2004年9月2日公布为US 2004-0171156A1,其全部内容通过引用并入本文。

[1414] 载体可包括调节元件,例如启动子。载体可包含CRISPR-Cas编码序列,和/或单个,但也可能包含至少2、3或8或16或32或48或50个指导RNA(例如,crRNA)编码序列,例如1-2、1-3、1-4、1-5、3-6、3-7、3-8、3-9、3-10、3-8、3-16、3-30、3-32、3-48、3-50个RNA(crRNA)。在单个载体中,每一个RNA(例如,crRNA)可以有一个启动子,有利的是,最多约有16个RNA(例如,crRNA);并且,当单个载体提供16个以上的RNA(例如,crRNA)时,一个或多个启动子可以驱动一个以上的RNA(例如,crRNA)的表达,例如,当有32个RNA(例如,sgRNA或crRNA)时,每个启动子可以驱动两个RNA(例如,sgRNA或crRNA)的表达,当有48个RNA(例如sgRNA或crRNA)时,每个启动子可驱动三个RNA(例如,sgRNA或crRNA)的表达。通过简单的算术和完善的克隆方案以及本公开中的教导,本领域技术人员可以容易地针对适合的示例性载体例如AAV和合适的启动子例如U6启动子,例如U6-sgRNA或-crRNA实施本发明的RNA例如sgRNA或crRNA。例如,AAV的包装限制为~4.7kb。技术人员可以容易地在单个载体中装配约12-16个,例如13个U6-sgRNA 或crRNA盒。可以通过任何合适的方式进行组装,例如用于TALE组装的金门策略([www.genome-engineering.org/taleffectors/](http://www.genome-engineering.org/taleffectors/))。技术人员还可以使用串联指

导策略将U6-sgRNA或-crRNA的数量增加约1.5倍,例如,从12-16,例如13增加至约18-24,例如,19个U6-sgRNA或-crRNA。因此,本领域技术人员可以在单个载体例如AAV载体中容易地达到约 18-24,例如约19个启动子-RNA,例如U6-sgRNA或-crRNA。增加载体中启动子和RNA(例如sgRNA或crRNA)数量的另一种方法是使用单个启动子(例如U6)来表达RNA阵列,例如sgRNA或crRNA由可切割序列分隔。并且用于增加载体中的启动子-RNA例如sgRNA或crRNA的数目的另一种方法是表达在载体中被编码序列或基因的内含子中的可切割序列分开的启动子-RNA例如sgRNA或crRNA的阵列;并且,在这种情况下,使用聚合酶II启动子是有利的,该启动子可以增加表达并能够以组织特异性方式转录长RNA。(例如,参见nar.oxfordjournals.org/content/34/7/e53.short, www.nature.com/mt/journal/v16/n9/abs/mt2008144a.html)。在一个有利的实施方式中,AAV可以包装靶向多达约50个基因的U6串联sgRNA。因此,根据本领域的知识和本公开的教导,本领域技术人员可以容易地制备和使用在控制下或可操作地或功能性连接至一个或多个启动子的表达多个RNA或指导或sgRNA或crRNA的载体,例如单个载体——特别是本文讨论的RNA或指导或sgRNA或crRNA的数量,无需进行任何不适当的实验。

#### [1415] 试剂盒

[1416] 一方面,本发明提供了包含上述方法和组合物中公开的任何一种或多种组分的试剂盒。在一些实施方式中,试剂盒包括本文所述的载体系统或本文所述的CRISPR/Cas系统或复合物的一种或多种成分,例如 crRNA和或/CRISPR-Cas效应蛋白或CRISPR-Cas效应蛋白编码mRNA,以及使用试剂盒的说明。元件可以单独提供或组合提供,并且可以在任何合适的容器中提供,例如小瓶、瓶或管。在一些实施方式中,试剂盒包括一种或多种语言的说明书,例如多种语言的说明书。这样的说明书可特定于本文所述的应用和方法。在一些实施方式中,试剂盒包含一种或多种试剂,用于使用本文所述一种或多种组分的方法。试剂可在任何合适的容器中提供。例如,试剂盒可提供一种或多种反应或储存缓冲液。试剂可以以可用于特定测定的形式提供,或者以在使用前需要添加一种或多种其他成分的形式提供(例如,以浓缩或冷冻干燥的形式)。缓冲液可以是任何缓冲液,包括但不限于碳酸钠缓冲液、碳酸氢钠缓冲液、硼酸盐缓冲液、Tris缓冲液、MOPS缓冲液、HEPES缓冲液及其组合。在一些实施方式中,缓冲液是碱性的。在一些实施方式中,缓冲液的pH值为约7至约10。在一些实施方式中,试剂盒包含一种或多种对应于插入载体的指导序列的寡核苷酸,以便可操作地连接指导序列或crRNA序列和调节元件。在一些实施方式中,试剂盒包含同源重组模板多核苷酸。在一些实施方式中,试剂盒包含一种或多种载体和/或一种或多种本文所述的多核苷酸。该试剂盒可有利地允许提供本发明系统的所有组分。

[1417] 本申请还提供了以下编号的陈述中所述的方面和实施方式:

[1418] 1. 一种工程化CRISPR-Cas蛋白,其包含一个或多个HEPN结构域,以及还包含一个或多个修饰的氨基酸,其中所述氨基酸:

[1419] a. 与指导RNA相互作用,所述指导RNA与所述工程化CRISPR-Cas 蛋白形成复合物;

[1420] b. 在所述工程化CRISPR-Cas蛋白的HEPN活性位点、结构域间接头结构域、盖结构域、螺旋结构域1、螺旋结构域2或桥螺旋结构域中;或者

[1421] c. 其组合。

[1422] 2. 根据陈述1所述的工程化CRISPR-Cas蛋白,其中所述HEPN结构域包含RxxxxH基

序。

[1423] 3. 根据陈述1或2所述的工程化CRISPR-Cas蛋白,其中所述RxxxxH 基序包含R {N/H/K} X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H序列。

[1424] 4. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中:

[1425] X<sub>1</sub>是R、S、D、E、Q、N、G或Y,

[1426] X<sub>2</sub>独立地是I、S、T、V或L,和

[1427] X<sub>3</sub>独立地是L、F、N、Y、V、I、S、D、E或A。

[1428] 5. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白是VI型CRISPR-Cas蛋白。

[1429] 6. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述VI型CRISPR-Cas蛋白是Cas13。

[1430] 7. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述VI型CRISPR-Cas蛋白是Cas13a、Cas13b、Cas13c或Cas13d。

[1431] 8. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、A656、V795、A796、W842、K871、E873、R874、R1068、N1069或H1073。

[1432] 9. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294、E400、R56、N157、H161、H452、N455、K484、N486、G566、H567、W842、K871、E873、R874、R1068、N1069或H1073。

[1433] 10. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R53、K943、R1041、Y164、R285、R287、K292、E296、N297、Q646、N647、R402、K393、N653、N652、R482、N480、D396、E397、D398、E399、K294或E400。

[1434] 11. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: K393、R402、N482、T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、

K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或H1073。

[1435] 12. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N482、H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、R56、N157、H161、R1068、N1069或H1073。

[1436] 13. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

[1437] 14. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

[1438] 15. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

[1439] 16. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在桥螺旋结构域中包含对应于PbCas13b的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:W842、K846、K870、E873或R877。

[1440] 17. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N480、N482、N652或N653。

[1441] 18. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N480或N482。

[1442] 19. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N480或N482。

[1443] 20. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:N652或N653。

[1444] 21. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:N652或N653。

[1445] 22. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

[1446] 23. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

[1447] 24. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于

PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、S757、N756或K741。

[1448] 25. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于PbCas13b的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、S757、N756或K741。

[1449] 26. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或N756。

[1450] 27. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、S757或N756。

[1451] 28. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、V795、A796、R791、G566、S757或N756。

[1452] 29. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、V795、A796、R791、G566、S757或N756。

[1453] 30. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846或R874。

[1454] 31. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在桥螺旋结构域中包含对应于PbCas13b的桥螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846或R874。

[1455] 32. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500 或G566。

[1456] 33. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-2中包含对应于PbCas13b的螺旋结构域1-2中的下述氨基酸的氨基酸的一个或多个突变:H567、H500或G566。

[1457] 34. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。

[1458] 35. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、S757或N756。

[1459] 36. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R762、V795、A796、R791、S757或N756。

[1460] 37. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:R762、V795、A796、R791、S757或N756。

[1461] 38. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、K590、R638或K741。

[1462] 39. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、A656、K655、N652、K590、R638 或K741。

[1463] 40. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:T405、H407、N486、K484、N480、H452、N455或K457。

[1464] 41. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:T405、H407、N486、K484、N480、H452、N455或 K457。

[1465] 42. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或K741。

[1466] 43. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于PbCas13b的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、S757、N756或K741。

[1467] 44. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。

[1468] 45. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、S757或N756。

[1469] 46. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、R791、G566、S757或N756。

[1470] 47. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于PbCas13b的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H567、H500、R762、R791、G566、S757或N756。

[1471] 48. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。

[1472] 49. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中

包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、S757或N756。

[1473] 50. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R762、R791、S757或N756。

[1474] 51. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于PbCas13b的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:R762、R791、S757或N756。

[1475] 52. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、K590、R638或K741。

[1476] 53. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于PbCas13b的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、K590、R638或K741。

[1477] 54. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:H407、N486、K484、N480、H452、N455或K457。

[1478] 55. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:H407、N486、K484、N480、H452、N455或K457。

[1479] 56. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R56、N157、H161、R1068、N1069或H1073。

[1480] 57. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于PbCas13b的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R56、N157、H161、R1068、N1069或H1073。

[1481] 58. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R56、N157或H161。

[1482] 59. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于PbCas13b的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R56、N157或H161。

[1483] 60. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:R1068、N1069或H1073。

[1484] 61. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于PbCas13b的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:R1068、N1069或H1073。

[1485] 62. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。

[1486] 63. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含

对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、T405、H407、N486、K484、N480、H452、N455或K457。

[1487] 64. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。

[1488] 65. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、H407、N486、K484、N480、H452、N455或K457。

[1489] 66. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：T405、H407、S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

[1490] 67. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：H407、S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

[1491] 68. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

[1492] 69. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457或K741。

[1493] 70. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：N486、K484、N480、H452、N455或K457。

[1494] 71. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变：N486、K484、N480、H452、N455或K457。

[1495] 72. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、N486、K484、N480、H452、N455或K457。

[1496] 73. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其在盖结构域中包含对应于PbCas13b的盖结构域中的下述氨基酸的氨基酸的一个或多个突变：K393、R402、N482、N486、K484、N480、H452、N455 或K457。

[1497] 74. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白，其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变：S658、N653、A656、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、V795、A796、R791、G566、

K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

[1498] 75. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于PbCas13b的下述氨基酸的氨基酸的一个或多个突变:S658、N653、K655、N652、H567、N455、H500、K871、K857、K870、W842、E873、R877、K846、R874、R762、R791、G566、K590、R638、H452、S757、N756、N486、K484、N480、K457、K741、K393、R402或N482。

[1499] 76. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943或R1041。

[1500] 77. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53或Y164。

[1501] 78. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K943或R1041。

[1502] 79. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943或R1041。

[1503] 80. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R53或Y164。

[1504] 81. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:K943或R1041。

[1505] 82. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

[1506] 83. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、R56、N157或H161。

[1507] 84. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

[1508] 85. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

[1509] 86. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、R56、N157或H161。

[1510] 87. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

[1511] 88. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b(PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943或R1041。

[1512] 89. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b(PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183或K193。

[1513] 90. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b(PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:K943或R1041。

[1514] 91. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943或R1041。

[1515] 92. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183或K193。

[1516] 93. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:K943 or R1041。

[1517] 94. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b(PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

[1518] 95. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b(PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、R56、N157 或H161。

[1519] 96. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b(PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

[1520] 97. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、K943、R1041、R56、N157、H161、R1068、N1069或H1073。

[1521] 98. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K183、K193、R56、N157或H161。

[1522] 99. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的HEPN结构域2中的下述

氨基酸的氨基酸的一个或多个突变:K943、R1041、R1068、N1069或H1073。

[1523] 100. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K183或K193。

[1524] 101. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:K183或K193。

[1525] 102. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943或R1041。

[1526] 103. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、Y164、K943或R1041。

[1527] 104. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53、K943或R1041;优选地R53A、R53K、R53D或R53E;K943A、K943R、K943D或K943E;或者R1041A、R1041K、R1041D或R1041E。

[1528] 105. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53、K943或R1041;优选地R53A、R53K、R53D或R53E;K943A、K943R、K943D或K943E;或者R1041A、R1041K、R1041D或R1041E。

[1529] 106. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 Y164的氨基酸的突变,优选地Y164A、Y164F或Y164W。

[1530] 107. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域1的氨基酸Y164的氨基酸的突变,优选地Y164A、Y164F或Y164W。

[1531] 108. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398或E399。

[1532] 109. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:T405、H407、K457、D434、K431、R402、K393、R482、N480、D396、E397、D398 或E399。

[1533] 110. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 H407的氨基酸的突变,优选地H407Y、H407W或H407F。

[1534] 111. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:

R402、K393、R482、N480、D396、E397、D398或E399。

[1535] 112. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:R402、K393、R482、N480、D396、E397、D398或E399。

[1536] 113. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K457、D434或K431。

[1537] 114. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K457、D434 或K431。

[1538] 115. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。

[1539] 116. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变:H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、Q646、N647、N653或N652。

[1540] 117. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1541] 118. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1542] 119. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762或 R791。

[1543] 120. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1中的下述氨基酸的氨基酸的一个或多个突变:H500、K570、N756、S757、R762或R791。

[1544] 121. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1545] 122. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在桥螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的桥螺旋结构域中的下

述氨基酸的氨基酸的一个或多个突变:K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1546] 123. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H500或K570。

[1547] 124. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-2中的下述氨基酸的氨基酸的一个或多个突变:H500或K570。

[1548] 125. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1549] 126. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870、R877、K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1550] 127. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762或R791。

[1551] 128. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762或R791。

[1552] 129. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870或R877。

[1553] 130. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:N756、S757、R762、R791、K846、K857、K870或R877。

[1554] 131. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1555] 132. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域1-3中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域1-3中的下述氨基酸的氨基酸的一个或多个突变:K826、K828、K829、R824、R830、Q831、K835、K836或R838。

[1556] 133. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653或N652。

[1557] 134. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741、K744、R600、K607、K612、R614、K617、R618、Q646、N647、N653或N652。

[1558] 135. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:Q646或N647。

[1559] 136. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:Q646或N647。

[1560] 137. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:N653或N652。

[1561] 138. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:N653或N652。

[1562] 139. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741或K744。

[1563] 140. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:K590、N634、R638、N652、N653、K655、S658、K741或K744。

[1564] 141. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:R600、K607、K612、R614、K617或 R618。

[1565] 142. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:R600、K607、K612、R614、K617或R618。

[1566] 143. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:R285、R287、K292、E296、N297或 K294。

[1567] 144. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 IDL结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的IDL结构域中的下述氨基酸的氨基酸的一个或多个突变: R285、R287、K292、E296、N297或K294。

[1568] 145. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的下述氨基酸的氨基酸的一个或多个突变:R285、K292、E296或N297。

[1569] 146. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 IDL结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b)的IDL结构域中的下述氨基

酸的氨基酸的一个或多个突变：R285、K292、E296或N297。

[1570] 147. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: T405、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870、R877、K183、K193、R600、K607、K612、R614、K617、K826、K828、K829、R824、R830、Q831、K835、K836、R838、R618、D434、K431、R285、R287、K292、E296、N297、Q646、N647或K294。

[1571] 148. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: R402、K393、N653、N652、R482、N480、D396、E397、D398或E399。

[1572] 149. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: R53、K655、R762或R1041; 优选地 R53A或R53D; K655A; R762A; 或者R1041E或R1041D。

[1573] 150. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: N297、E296、K292或R285; 优选地 N297A、E296A、K292A或R285A。

[1574] 151. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 IDL结构域(的中央通道)中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的IDL结构域(的中央通道)中的下述氨基酸的氨基酸的一个或多个突变: N297、E296、K292或R285; 优选地N297A、E296A、K292A或R285A。

[1575] 152. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: Q831、K836、R838、N652、N653、R830、K655或R762; 优选地Q831A、K836A、R838A、N652A、N653A、R830A、K655A或R762A。

[1576] 153. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: N652、N653、R830、K655 or R762; 优选地N652A、N653A、R830A、K655A或R762A。

[1577] 154. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变: K655或R762; 优选地K655A或 R762A。

[1578] 155. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变: Q831、K836、R838、N652、N653、R830、K655或R762; 优选地Q831A、K836A、R838A、N652A、N653A、R830A、K655A或R762A。

[1579] 156. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域中包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域中的下述氨基酸的氨基酸的一个或多个突变: N652、N653、R830、K655或R762; 优选地N652A、N653A、R830A、K655A或R762A。

[1580] 157. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2中

包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2中的下述氨基酸的氨基酸的一个或多个突变:K655或R762;优选地K655A或R762A。

[1581] 158. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R614、K607、K193、K183或R600;优选地R614A、K607A、K193A、K183A或R600A。

[1582] 159. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在螺旋结构域2的反式亚基环中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的螺旋结构域2的反式亚基环中的下述氨基酸的氨基酸的一个或多个突变:Q646或N647;优选地Q646A或N647A。

[1583] 160. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R53或R1041;优选地R53A或R53D 或R1041E或R1041D。

[1584] 161. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R53或R1041;优选地R53A或R53D或R1041E或R1041D。

[1585] 162. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K457、D397、E398、D399、E400、T405、H407或D434;优选地D397A、E398A、D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。

[1586] 163. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在盖结构域中包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的盖结构域中的下述氨基酸的氨基酸的一个或多个突变:K457、D397、E398、D399、E400、T405、H407或D434;优选地D397A、E398A、D399A、E400A、T405A、H407A、H407W、H407Y、H407F或D434A。

[1587] 164. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸:氨基酸46-57、73-79、152-164、1036-1046和1064-1074。

[1588] 165. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R156、N157、H161、R1068、N1069 和H1073。

[1589] 166. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R285、R287、K292、K294、E296和 N297。

[1590] 167. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:K826、K828、K829、R824、R830、Q831、K835、K836和R838。

[1591] 168. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的下述氨基酸的氨基酸的一个或多个突变:T405、H407、K457、H500、K570、K590、N634、R638、N652、N653、K655、S658、K741、K744、N756、S757、R762、R791、K846、K857、K870和R877。

- [1592] 169. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 T405的氨基酸的突变。
- [1593] 170. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 H407的氨基酸的突变。
- [1594] 171. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K457的氨基酸的突变。
- [1595] 172. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 H500的氨基酸的突变。
- [1596] 173. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K570的氨基酸的突变。
- [1597] 174. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K590的氨基酸的突变。
- [1598] 175. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸N634的氨基酸的突变。
- [1599] 176. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R638的氨基酸的突变。
- [1600] 177. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N652的氨基酸的突变。
- [1601] 178. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N653的氨基酸的突变。
- [1602] 179. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K655的氨基酸的突变。
- [1603] 180. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 S658的氨基酸的突变。
- [1604] 181. 根据前述陈述中任一项中任一项55所述的工程化 CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸K741的氨基酸的突变。
- [1605] 182. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K744的氨基酸的突变。
- [1606] 183. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N756的氨基酸的突变。
- [1607] 184. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 S757的氨基酸的突变。
- [1608] 185. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R762的氨基酸的突变。
- [1609] 186. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R791的氨基酸的突变。
- [1610] 187. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K846的氨基酸的突变。

- [1611] 188. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K857的氨基酸的突变。
- [1612] 189. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K870的氨基酸的突变。
- [1613] 190. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R877的氨基酸的突变。
- [1614] 191. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K183的氨基酸的突变。
- [1615] 192. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K193的氨基酸的突变。
- [1616] 193. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R600的氨基酸的突变。
- [1617] 194. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K607的氨基酸的突变。
- [1618] 195. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K612的氨基酸的突变。
- [1619] 196. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R614的氨基酸的突变。
- [1620] 197. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K617的氨基酸的突变。
- [1621] 198. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K826的氨基酸的突变。
- [1622] 199. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K828的氨基酸的突变。
- [1623] 200. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K829的氨基酸的突变。
- [1624] 201. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R824的氨基酸的突变。
- [1625] 202. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R830的氨基酸的突变。
- [1626] 203. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 Q831的氨基酸的突变。
- [1627] 204. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K835的氨基酸的突变。
- [1628] 205. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K836的氨基酸的突变。
- [1629] 206. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R838的氨基酸的突变。
- [1630] 207. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷

沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R618 的氨基酸的突变。

[1631] 208. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 D434 的氨基酸的突变。

[1632] 209. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K431 的氨基酸的突变。

[1633] 210. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R53 的氨基酸的突变。

[1634] 211. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K943 的氨基酸的突变。

[1635] 212. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R1041 的氨基酸的突变。

[1636] 213. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 Y164 的氨基酸的突变。

[1637] 214. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R285 的氨基酸的突变。

[1638] 215. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R287 的氨基酸的突变。

[1639] 216. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K292 的氨基酸的突变。

[1640] 217. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 E296 的氨基酸的突变。

[1641] 218. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N297 的氨基酸的突变。

[1642] 219. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 Q646 的氨基酸的突变。

[1643] 220. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N647 的氨基酸的突变。

[1644] 221. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R402 的氨基酸的突变。

[1645] 222. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K393 的氨基酸的突变。

[1646] 223. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N653 的氨基酸的突变。

[1647] 224. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N652 的氨基酸的突变。

[1648] 225. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R482 的氨基酸的突变。

[1649] 226. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N480 的氨基酸的突变。

- [1650] 227. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 D396的氨基酸的突变。
- [1651] 228. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 E397的氨基酸的突变。
- [1652] 229. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 D398的氨基酸的突变。
- [1653] 230. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 E399的氨基酸的突变。
- [1654] 231. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K294的氨基酸的突变。
- [1655] 232. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 E400的氨基酸的突变。
- [1656] 233. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R56的氨基酸的突变。
- [1657] 234. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N157的氨基酸的突变。
- [1658] 235. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 H161的氨基酸的突变。
- [1659] 236. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 H452的氨基酸的突变。
- [1660] 237. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N455的氨基酸的突变。
- [1661] 238. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K484的氨基酸的突变。
- [1662] 239. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N486的氨基酸的突变。
- [1663] 240. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 G566的氨基酸的突变。
- [1664] 241. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 H567的氨基酸的突变。
- [1665] 242. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 A656的氨基酸的突变。
- [1666] 243. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 V795的氨基酸的突变。
- [1667] 244. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 A796的氨基酸的突变。
- [1668] 245. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌(*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 W842的氨基酸的突变。
- [1669] 246. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷

沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 K871 的氨基酸的突变。

[1670] 247. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 E873 的氨基酸的突变。

[1671] 248. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R874 的氨基酸的突变。

[1672] 249. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 R1068 的氨基酸的突变。

[1673] 250. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 N1069 的氨基酸的突变。

[1674] 251. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于颊普雷沃菌 (*Prevotella buccae*) Cas13b (PbCas13b) 的氨基酸 H1073 的氨基酸的突变。

[1675] 252. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌 (*Leptotrichia shahii*) Cas13a (LshCas13a) 的下述氨基酸的氨基酸的一个或多个突变:R597、N598、H602、R1278、N1279 或 H1283。

[1676] 253. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌 (*Leptotrichia shahii*) Cas13a (LshCas13a) 的下述氨基酸的氨基酸的一个或多个突变:R597、N598、H602、R1278、N1279 或 H1283。

[1677] 254. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中包含对应于沙氏纤毛菌 (*Leptotrichia shahii*) Cas13a (LshCas13a) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R597、N598、H602、R1278、N1279或H1283。

[1678] 255. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌 (*Leptotrichia shahii*) Cas13a (LshCas13a) 的下述氨基酸的氨基酸的一个或多个突变:R597、N598或H602。

[1679] 256. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域1中包含对应于沙氏纤毛菌 (*Leptotrichia shahii*) Cas13a (LshCas13a) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R597、N598或H602。

[1680] 257. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于沙氏纤毛菌 (*Leptotrichia shahii*) Cas13a (LshCas13a) 的下述氨基酸的氨基酸的一个或多个突变:R1278、N1279或H1283。

[1681] 258. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域2中包含对应于沙氏纤毛菌 (*Leptotrichia shahii*) Cas13a (LshCas13a) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:R1278、N1279或H1283。

[1682] 259. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R146、H151、R1116或H1121。

[1683] 260. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R146、H151、R1116或H1121。

[1684] 261. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在 HEPN结构域中

包含对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:R146、H151、R1116或H1121。

[1685] 262. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R146或H151。

[1686] 263. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的HEPN结构域1中的下述氨基酸的氨基酸的一个或多个突变:R146或H151。

[1687] 264. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的下述氨基酸的氨基酸的一个或多个突变:R1116或H1121。

[1688] 265. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于喉管卟啉单胞菌 (*Porphyromonas gulae*) Cas13b (PguCas13b) 的HEPN结构域2中的下述氨基酸的氨基酸的一个或多个突变:R1116或H1121。

[1689] 266. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种 (*Prevotella sp.*) P5-125 Cas13b (PspCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H133或H1058。

[1690] 267. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种 (*Prevotella sp.*) P5-125 Cas13b (PspCas13b) 的下述氨基酸的氨基酸的一个或多个突变:H133或H1058。

[1691] 268. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域中包含对应于普雷沃菌属种 (*Prevotella sp.*) P5-125 Cas13b (PspCas13b) 的HEPN结构域中的下述氨基酸的氨基酸的一个或多个突变:H133或H1058。

[1692] 269. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种 (*Prevotella sp.*) P5-125 Cas13b (PspCas13b) 的氨基酸H133的氨基酸突变。

[1693] 270. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域1中包含对应于普雷沃菌属种 (*Prevotella sp.*) P5-125 Cas13b (PspCas13b) 的HEPN结构域1中的氨基酸H133的氨基酸突变。

[1694] 271. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其包含对应于普雷沃菌属种 (*Prevotella sp.*) P5-125 Cas13b (PspCas13b) 的氨基酸H1058的氨基酸突变。

[1695] 272. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其在HEPN结构域2中包含对应于普雷沃菌属种 (*Prevotella sp.*) P5-125 Cas13b (PspCas13b) 的HEPN结构域2中的氨基酸H1058的氨基酸突变。

[1696] 273. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为A、P或V,优选地A。

[1697] 274. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为疏水性氨基酸。

[1698] 275. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为芳香族氨基酸。

[1699] 276. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为带电荷的氨基酸。

[1700] 277. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为带正电的氨基酸。

[1701] 278. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为带负电的氨基酸。

[1702] 279. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为极性氨基酸。

[1703] 280. 根据陈述8至272中任一项所述的工程化CRISPR-Cas蛋白,其中所述氨基酸被突变为脂肪族氨基酸。

[1704] 281. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是或源于以下的种:别样杆菌属 (*Alistipes*)、厌氧唾液杆菌 (*Anaerosalibacter*)、拟杆菌属 (*Bacteroides*)、拟杆菌门 (*Bacteroidetes*)、伯杰菌属 (*Bergeyella*)、布劳特氏菌属 (*Blautia*)、丁酸弧菌属 (*Butyrivibrio*)、二氧化碳噬纤维菌属 (*Capnocytophaga*)、肉食杆菌属 (*Carnobacterium*)、绿曲挠丝状菌属 (*Chloroflexus*)、金黄杆菌属 (*Chryseobacterium*)、梭菌属 (*Clostridium*)、*Demequina*、优杆菌科 (*Eubacteriaceae*)、优杆菌属 (*Eubacterium*)、黄杆菌属 (*Flavobacterium*)、梭杆菌属 (*Fusobacterium*)、*Herbinix*、*Insolitispirillum*、毛螺菌科 (*Lachnospiraceae*)、纤毛菌属 (*Leptotrichia*)、李斯特氏菌属 (*Listeria*)、香味菌属 (*Myroides*)、杆状杆菌 (*Paludibacter*)、细菌杆菌 (*Phaeodactylibacter*)、紫单胞菌科 (*Porphyromonadaceae*)、卟啉单胞菌属 (*Porphyromonas*)、普雷沃菌属 (*Prevotella*)、假丁酸弧菌属 (*Pseudobutyrvibrio*)、冷弯菌属 (*Psychroflexus*)、赖兴氏杆菌 (*Reichenbachiella*)、红杆菌属 (*Rhodobacter*)、里氏杆菌属 (*Riemerella*)、*Sinomicrobium*、海旋菌属 (*Thalassospira*)、瘤胃球菌属 (*Ruminococcus*)；优选地,沙氏纤毛菌 (*Leptotrichia shahii*)、斯氏李斯特氏菌 (*Listeria seeligeri*)、毛螺科菌细菌 (*Lachnospiraceae bacterium*) (如Lb MA2020、Lb NK4A179、Lb NK4A144)、*Clostridium aminophilum* (如Ca DSM 10710)、鸡肉杆菌 (*Carnobacterium gallinarum*) (如Cg DSM 4847)、丙酸杆菌 (*Paludibacter propionicigenes*) (如Pp WB4)、魏斯特斯蒂芬氏李斯特氏菌 (*Listeria weihenstephanensis*) (如Lw FSL R9-0317)、李斯特氏菌科细菌 (*Listeriaceae bacterium*) (如Lb FSL M6-0635)、瓦氏细单胞菌 (*Leptotrichia wadei*) (如Lw F0279)、荚膜红杆菌 (*Rhodobacter capsulatus*) (如Rc SB 1003、Rc R121、Rc DE442)、口腔纤毛菌 (*Leptotrichia buccalis*) (如Lb C-1013-b)、*Herbinix hemicellulosilytica*、优杆菌科细菌 (*Eubacteriaceae bacterium*) (如Eb CHKCI004)、布劳特氏菌属种 (*Blautia sp.*) Marseille-P2398、纤毛菌属种 (*Leptotrichia sp.*) 口腔分类单元879str.F0557、聚集绿屈挠菌 (*Chloroflexus aggregans*)、*Demequina aurantiaca*、海旋菌种 (*Thalassospira sp.*) TSL5-1、假丁酸弧菌属种 (*Pseudobutyrvibrio sp.*) OR37、丁酸弧菌属种 (*Butyrivibrio sp.*) YAB3001、纤毛菌属种 (*Leptotrichia sp.*) Marseille-P3007、*Bacteroides ihuae*、紫单胞菌科细菌 (*Porphyromonadaceae bacterium*) (如Pb KH3CP3RA)、*Listeria riparia*、*Insolitispirillum peregrinum*、别样杆菌属种 (*Alistipes sp.*) ZOR0009、酿脓拟杆菌

(*Bacteroides pyogenes*) (如 Bp F0041)、拟杆菌属细菌 (*Bacteroidetes bacterium*) (如 Bb GWA2\_31\_9)、动物溃疡伯格菌 (*Bergeyella zoohelcum*) (如Bz ATCC 43767)、犬咬二氧化碳噬纤维菌 (*Capnocytophaga canimorsus*)、希诺地米碳酸噬胞菌 (*Capnocytophaga cynodegmi*)、*Chryseobacterium carnipullorum*、济州岛金黄杆菌 (*Chryseobacterium jejuense*)、嗜鳃黄杆菌 (*Chryseobacterium ureilyticum*)、嗜分支黄杆菌 (*Flavobacterium branchiophilum*)、柱状黄杆菌 (*Flavobacterium columnare*)、黄杆菌属种 (*Flavobacterium sp.*) 316、*Myroides odoratimimus* (如Mo CCUG 10230、Mo CCUG 12901、Mo CCUG 3837)、*Paludibacter propionicigenes*、*Phaeodactylibacter xiamenensis*、牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) (如Pg F0185、Pg F0568、Pg JCVI SC001、Pg W4087)、马铃薯红单胞菌 (*Porphyromonas gulae*)、卟啉单胞菌属种 (*Porphyromonas sp.*) COT-0520H4946、树蛙普雷沃菌 (*Prevotella aurantiaca*)、颊普雷沃菌 (*Prevotella buccae*) (如Pb ATCC 33574)、*Prevotella falsenii*、中间普雷沃菌 (*Prevotella intermedia*) (如Pi 17、Pi ZT)、*Prevotella pallens* (如Pp ATCC 700821)、*Prevotella pleuritidis*、解糖普雷沃菌 (*Prevotella saccharolytica*) (如Ps F0055)、普雷沃菌属种 (*Prevotella sp.*) MA2016、普雷沃菌属种 (*Prevotella sp.*) MSX73、普雷沃菌属种 (*Prevotella sp.*) P4-76、普雷沃菌属种 (*Prevotella sp.*) P5-119、普雷沃菌属种 (*Prevotella sp.*) P5-125、普雷沃菌属种 (*Prevotella sp.*) P5-60、扭曲冷弯曲菌 (*Psychroflexus torquis*)、*Reichenbachiella agariperforans*、鸭疫里默氏杆菌 (*Riemerella anatipestifer*)、*Sinomicrobium oceani*、坏死梭杆菌 (*Fusobacterium necrophorum*) (如Fn subsp.funduliforme ATCC 51357、Fn DJ-2、Fn BFTR-1、Fn subsp.Funduliforme)、坏疽梭杆菌 (*Fusobacterium perfoetens*) (如Fp ATCC 29250)、溃疡梭杆菌 (*Fusobacterium ulcerans*) (如Fu ATCC 49185)、厌氧唾液杆菌属种 (*Anaerosalibacter sp.*) ND1、产亚硝酸真杆菌 (*Eubacterium siraeum*)、黄化瘤胃球菌 (*Ruminococcus flavefaciens*) (如Rfx XPD3002) 或白色瘤胃球菌 (*Ruminococcus albus*)。

[1705] 282. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是Cas13a蛋白。

[1706] 283. 根据陈述282所述的工程化CRISPR-Cas蛋白,其中所述Cas13a 蛋白是或源于以下的种:拟杆菌属 (*Bacteroides*)、布劳特氏菌属 (*Blautia*)、丁酸弧菌属 (*Butyrivibrio*)、肉食杆菌属 (*Carnobacterium*)、绿曲挠丝状菌属 (*Chloroflexus*)、梭菌属 (*Clostridium*)、*Demequina*、优杆菌属 (*Eubacterium*)、*Herbinix*、*Insolitinspirillum*、毛螺菌科 (*Lachnospiraceae*)、纤毛菌属 (*Leptotrichia*)、李斯特氏菌属 (*Listeria*)、杆状杆菌 (*Paludibacter*)、紫单胞菌科 (*Porphyromonadaceae*)、假丁酸弧菌属 (*Pseudobutyrvibrio*)、红杆菌属 (*Rhodobacter*) 或海旋菌属 (*Thalassospira*); 优选地,沙氏纤毛菌 (*Leptotrichia shahii*)、斯氏李斯特氏菌 (*Listeria seeligeri*)、毛螺科菌细菌 (*Lachnospiraceae bacterium*) (如Lb MA2020、Lb NK4A179、Lb NK4A144)、*Clostridium aminophilum* (如Ca DSM 10710)、鸡肉杆菌 (*Carnobacterium gallinarum*) (如Cg DSM 4847)、丙酸杆菌 (*Paludibacter propionicigenes*) (如Pp WB4)、魏斯特斯蒂芬氏李斯特氏菌 (*Listeria weihenstephanensis*) (如Lw FSL R9-0317)、李斯特氏菌科细菌 (*Listeriaceae bacterium*) (如Lb FSL M6-0635)、瓦氏细单胞菌 (*Leptotrichia wadei*)

(如Lw F0279)、荚膜红杆菌 (*Rhodobacter capsulatus*) (如Rc SB 1003、Rc R121、Rc DE442)、口腔纤毛菌 (*Leptotrichia buccalis*) (如Lb C-1013-b)、*Herbinix hemicellulosilytica*、优杆菌科细菌 (*Eubacteriaceae bacterium*) (如Eb CHKCI004)、布劳特氏菌属种 (*Blautia sp.*) Marseille-P2398、纤毛菌属种 (*Leptotrichia sp.*) 口腔分类单元879str.F0557、聚集绿屈扰菌 (*Chloroflexus aggregans*)、*Demequina aurantiaca*、海旋菌种 (*Thalassospira sp.*) TSL5-1、假丁酸弧菌属种 (*Pseudobutyrvibrio sp.*) OR37、丁酸弧菌属种 (*Butyrvibrio sp.*) YAB3001、纤毛菌属种 (*Leptotrichia sp.*) Marseille-P3007、*Bacteroides ihuae*、紫单孢菌科细菌 (*Porphyromonadaceae bacterium*) (如Pb KH3CP3RA)、*Listeria riparia*或*Insolitispirillum peregrinum*。

[1707] 284. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是Cas13b蛋白。

[1708] 285. 根据陈述284所述的工程化CRISPR-Cas蛋白,其中所述Cas13b 蛋白是或源于以下的种:别样杆菌属 (*Alistipes*)、拟杆菌属 (*Bacteroides*)、拟杆菌门 (*Bacteroidetes*)、伯杰菌属 (*Bergeyella*)、二氧化碳噬纤维菌属 (*Capnocytophaga*)、金黄杆菌属 (*Chryseobacterium*)、黄杆菌属 (*Flavobacterium*)、香味菌属 (*Myroides*)、杆状杆菌 (*Paludibacter*)、细菌杆菌 (*Phaeodactylibacter*)、卟啉单胞菌属 (*Porphyromonas*)、普雷沃菌属 (*Prevotella*)、冷弯菌属 (*Psychroflexus*)、赖兴氏杆菌 (*Reichenbachiella*)、里氏杆菌属 (*Riemerella*) 或*Sinomicrobium*; 优选地,别样杆菌属种 (*Alistipes sp.*) ZOR0009、酿脓拟杆菌 (*Bacteroides pyogenes*) (如Bp F0041)、拟杆菌属细菌 (*Bacteroidetes bacterium*) (如 Bb GWA2\_31\_9)、动物溃疡伯格菌 (*Bergeyella zoohelcum*) (如Bz ATCC 43767)、犬咬二氧化碳噬纤维菌 (*Capnocytophaga canimorsus*)、希诺地米碳酸噬胞菌 (*Capnocytophaga cynodegmi*)、*Chryseobacterium carnipullorum*、济州岛金黄杆菌 (*Chryseobacterium jejuense*)、嗜鳃黄杆菌 (*Chryseobacterium ureilyticum*)、嗜分支黄杆菌 (*Flavobacterium branchiophilum*)、柱状黄杆菌 (*Flavobacterium columnare*)、黄杆菌属种 (*Flavobacterium sp.*) 316、*Myroides odoratimimus* (如Mo CCUG 10230、Mo CCUG 12901、Mo CCUG 3837)、丙酸杆菌 (*Paludibacter propionicigenes*)、*Phaeodactylibacter xiamenensis*、牙龈卟啉单胞菌 (*Porphyromonas gingivalis*) (如Pg F0185、Pg F0568、Pg JCVI SC001、Pg W4087)、马铃薯红单胞菌 (*Porphyromonas gulae*)、卟啉单胞菌属种 (*Porphyromonas sp.*) COT-0520H4946、树蛙普雷沃菌 (*Prevotella aurantiaca*)、颊普雷沃菌 (*Prevotella buccae*) (如Pb ATCC 33574)、*Prevotella falsenii*、中间普雷沃菌 (*Prevotella intermedia*) (如Pi 17、Pi ZT)、*Prevotella pallens* (如Pp ATCC 700821)、*Prevotella pleuritidis*、解糖普雷沃菌 (*Prevotella saccharolytica*) (如Ps F0055)、普雷沃菌属种 (*Prevotella sp.*) MA2016、普雷沃菌属种 (*Prevotella sp.*) MSX73、普雷沃菌属种 (*Prevotella sp.*) P4-76、普雷沃菌属种 (*Prevotella sp.*) P5-119、普雷沃菌属种 (*Prevotella sp.*) P5-125、普雷沃菌属种 (*Prevotella sp.*) P5-60、扭曲冷弯菌 (*Psychroflexus torquis*)、*Reichenbachiella agariperforans*、鸭疫里默氏杆菌 (*Riemerella anatipestifer*) 或 *Sinomicrobium oceani*。

[1709] 286. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是Cas13c蛋白。

[1710] 287. 根据陈述286所述的工程化CRISPR-Cas蛋白,其中所述Cas13c 蛋白是或源于以下的种:梭杆菌属 (*Fusobacterium*) 或厌氧唾液杆菌 (*Anaerosalibacter*); 优选地, 坏死梭杆菌 (*Fusobacterium necrophorum*) (如Fn subsp. *funduliforme* ATCC 51357、Fn DJ-2、Fn BFTR-1、Fn subsp. *Funduliforme*)、坏疽梭杆菌 (*Fusobacterium perfoetens*) (如Fp ATCC 29250)、溃疡梭杆菌 (*Fusobacterium ulcerans*) (如Fu ATCC 49185) 或厌氧唾液杆菌属种 (*Anaerosalibacter* sp.) ND1。

[1711] 288. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中所述Cas13蛋白是Cas13d蛋白。

[1712] 289. 根据陈述288所述的工程化CRISPR-Cas蛋白,其中所述Cas13d 蛋白是或源于以下的种:优杆菌属 (*Eubacterium*) 或瘤胃球菌属 (*Ruminococcus*), 优选地, 产亚硝酸真杆菌 (*Eubacterium siraeum*)、黄化瘤胃球菌 (*Ruminococcus flavefaciens*) (如Rfx XPD3002) 或白色瘤胃球菌 (*Ruminococcus albus*)。

[1713] 290. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的催化活性增加。

[1714] 291. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的催化活性降低。

[1715] 292. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的 gRNA结合增加。

[1716] 293. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的 gRNA结合降低。

[1717] 294. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的特异性增加。

[1718] 295. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的特异性降低。

[1719] 296. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的稳定性增加。

[1720] 297. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白的稳定性降低。

[1721] 298. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其还包含使催化活性失活的一个或多个突变。

[1722] 299. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的脱靶结合增加。

[1723] 300. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的脱靶结合降低。

[1724] 301. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的靶标结合增加。

[1725] 302. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,所述CRISPR-Cas蛋白的靶标结合降低。

[1726] 303. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型

CRISPR-Cas蛋白相比,所述工程化CRISPR-Cas蛋白具有更高蛋白酶活性或多核苷酸结合能力。

[1727] 304. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其中与对应的野生型CRISPR-Cas蛋白相比,PFS识别被改变。

[1728] 305. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其还包含功能性异源结构域。

[1729] 306. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其还包含NLS。

[1730] 307. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白,其还包含NES。

[1731] 308. 一种工程化CRISPR-Cas蛋白,其包含一个或多个HEPN结构域,且长度小于1000个氨基酸。

[1732] 309. 根据陈述308所述的工程化CRISPR-Cas蛋白,其中所述蛋白的尺寸小于950个、小于900个、小于850个、小于800个、或小于750个氨基酸。

[1733] 310. 根据陈述308或309所述的工程化CRISPR-Cas蛋白,其中所述HEPN结构域包含RxxxxH基序。

[1734] 311. 根据陈述310所述的工程化CRISPR-Cas蛋白,其中所述 RxxxxH基序包含R[N/H/K]X<sub>1</sub>X<sub>2</sub>X<sub>3</sub>H序列。

[1735] 312. 根据陈述311所述的工程化CRISPR-Cas蛋白,其中:

[1736] X<sub>1</sub>是R、S、D、E、Q、N、G或Y,

[1737] X<sub>2</sub>独立地是I、S、T、V或L,和

[1738] X<sub>3</sub>独立地是L、F、N、Y、V、I、S、D、E或A。

[1739] 313. 根据陈述308-313中任一项所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白是VI型CRISPR Cas蛋白。

[1740] 314. 根据陈述313所述的工程化CRISPR-Cas蛋白,其中所述VI型 CRISPR Cas蛋白是Cas13a、Cas13b、Cas13c或Cas13d。

[1741] 315. 根据陈述308-315中任一项所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白与功能性结构域相关联。

[1742] 316. 根据陈述308-316中任一项所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白包含与陈述[1386]57-[1386]329中任一项所述的突变等价的一个或多个突变。

[1743] 317. 根据陈述316所述的工程化CRISPR-Cas蛋白,其中所述 CRISPR-Cas蛋白包含在螺旋结构域中的一个或多个突变。

[1744] 318. 根据陈述308-318中任一项所述的工程化CRISPR-Cas蛋白,其中所述CRISPR-Cas蛋白是死形式或具有切口酶活性。

[1745] 319. 一种多核苷酸,其编码陈述1至318中任一项所述的工程化 CRISPR-Cas蛋白。

[1746] 320. 根据陈述319所述的多核苷酸,其是密码子优化的。

[1747] 321. 一种CRISPR-Cas系统,其包含陈述1至[1386]367中任一项所述的工程化CRISPR-Cas蛋白或者陈述318或319所述的多核苷酸,以及能够与所述工程化CRISPR-Cas蛋白形成复合物并能够与靶核酸序列杂交并且引导所述复合物与所述靶核酸序列的序列特异性结合的核苷酸组分。

[1748] 322. 一种载体系统,其包含一个或多个载体,所述一个或多个载体包含编码陈述

321所述的工程化CRISPR-Cas蛋白的组分的一个或多个多核苷酸分子。

[1749] 323.一种修饰靶核酸的方法,其包括:在包含所述靶核酸的细胞或生物体中引入根据陈述1至318中任一项所述的工程化CRISPR-Cas蛋白,根据陈述319或320所述的多核酸,根据陈述321所述的CRISPR-Cas系统,或者根据陈述322所述的载体或载体系统,使得所述工程化CRISPR-Cas蛋白修饰在所述细胞或生物体中的所述靶核酸。

[1750] 324.根据陈述[1386]372所述的方法,其中所述工程化CRISPR-Cas系统是通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪、可植入装置或陈述322所述的载体系统递送引入。

[1751] 325.根据陈述323或324所述的方法,其中所述工程化CRISPR-cas蛋白与一个或多个功能性结构域相关联。

[1752] 326.根据陈述323-325中任一项所述的方法,其中所述靶核酸包含基因组基因座,并且所述工程化CRISPR-Cas蛋白改变在所述基因组基因座处编码的基因产物或所述基因产物的表达。

[1753] 327.根据陈述323-326中任一项所述的方法,其中所述靶核酸是DNA或RNA,且其中所述靶核酸中的一个或多个核苷酸被碱基编辑。

[1754] 328.根据陈述323-327中任一项所述的方法,其中所述靶核酸是DNA或RNA,且其中所述靶核酸被切割。

[1755] 329.根据陈述328所述的方法,其中所述工程化CRISPR-Cas蛋白还切割非靶核酸。

[1756] 330.根据陈述328或329所述的方法,其还包括使活性可视化,以及任选地,使用可检测标记。

[1757] 331.根据陈述328-330中任一项所述的方法,其还包括检测所述CRISPR-Cas系统的一个或多个组分与所述靶核酸的结合。

[1758] 332.根据陈述328-331中任一项所述的方法,其中所述细胞或生物体是原核细胞或生物体。

[1759] 333.根据陈述328-332中任一项所述的方法,其中所述细胞或生物体是动物细胞或生物体。

[1760] 334.根据陈述328-333中任一项所述的方法,其中所述细胞或生物体是植物细胞或生物体。

[1761] 335.一种用于检测样品中的靶核酸的方法,其包括:

[1762] 将样品与下述接触:

[1763] 陈述1-318中任一项所述的工程化CRISPR-Cas蛋白;

[1764] 至少一种指导多核苷酸,其包含能够结合至所述靶核酸并被设计为与所述工程化CRISPR-Cas形成复合物的指导序列;和

[1765] 包含非靶序列的基于RNA的掩蔽构建体;

[1766] 其中所述工程化CRISPR-Cas蛋白表现出附带RNase活性并切割所述检测构建体的所述非靶序列;和

[1767] 检测来自所述非靶序列的切割的信号,从而检测在所述样品中的所述靶核酸。

[1768] 336.根据陈述335所述的方法,其还包括将所述样品与用于扩增所述靶核酸的试剂接触。

- [1769] 337. 根据陈述336所述的方法,其中用于扩增的所述试剂包括等温扩增反应试剂。
- [1770] 338. 根据陈述337所述的方法,其中所述等温扩增试剂包括基于核酸序列的扩增、重组酶聚合酶扩增、环介导的等温扩增、链置换扩增、解旋酶依赖性扩增或切口酶扩增试剂。
- [1771] 339. 根据陈述335-338中任一项所述的方法,其中所述靶核酸是 DNA分子,并且所述方法还包括将所述靶DNA分子与包含RNA聚合酶位点的引物和RNA聚合酶接触。
- [1772] 340. 根据陈述335-339中任一项所述的方法,其中所述掩蔽构建体:
- [1773] 抑制可检测阳性信号的产生直至所述掩蔽构建体被切割或失活,或者
- [1774] 掩蔽可检测阳性信号或产生可检测阴性信号直至所述掩蔽构建体被切割或失活。
- [1775] 341. 根据陈述335-340中任一项所述的方法,其中所述掩蔽构建体包含:
- [1776] a. 沉默RNA,其抑制由报告构建体编码的基因产物的产生,其中当表达时所述基因产物产生所述可检测阳性信号;
- [1777] b. 核酶,其产生所述可检测阴性信号,并且其中当所述核酶失活时产生所述阳性可检测信号;
- [1778] c. 核酶,其将底物转化为第一颜色,并且其中当所述核酶失活时所述底物转化为第二颜色;
- [1779] d. 适体和/或包含多核苷酸栓系的抑制剂;
- [1780] e. 多核苷酸,其附着有可检测配体和掩蔽组分;
- [1781] f. 通过桥分子保持在聚集体中的纳米颗粒,其中至少一部分的所述桥分子包含多核苷酸,并且其中当所述纳米颗粒被释放在溶液中时所述溶液经历色移;
- [1782] g. 通过连接分子连接至一个或多个淬灭剂分子的量子点或荧光团,其中至少一部分的所述连接分子包含多核苷酸;
- [1783] h. 与嵌入剂复合的多核苷酸,其中在切割所述多核苷酸后所述嵌入剂改变吸光度;或者
- [1784] 1. 通过多核苷酸栓系的两个荧光团,其当从所述多核苷酸释放时经历荧光移位。
- [1785] 342. 根据陈述341所述的方法,其中所述适体
- [1786] a. 包含螯合酶的多核苷酸栓系的抑制剂,其中在通过作用于底物而从所述适体或多核苷酸栓系的抑制剂释放后所述酶产生可检测信号;或者
- [1787] b. 是抑制性适体,其抑制酶并阻止酶催化底物产生可检测信号,或其中所述多核苷酸栓系的抑制剂抑制酶并阻止酶催化底物产生可检测信号;或者
- [1788] c. 螯合一对试剂,所述试剂当从所述适体释放时组合以产生可检测信号。
- [1789] 343. 根据陈述341或342所述的方法,其中所述纳米颗粒是胶体金属。
- [1790] 344. 根据陈述335-343中任一项所述的方法,其中所述至少一个指导多核苷酸包含错配。
- [1791] 345. 根据陈述344所述的方法,其中所述错配是在所述一个或多个指导序列上的单核苷酸改变的上游或下游。
- [1792] 346. 一种细胞或生物体,其包含根据陈述1-318中任一项所述的工程化CRISPR-Cas蛋白,根据陈述319或320所述的多核酸,根据陈述 321所述的CRISPR-Cas系统或者根据陈述322所述的载体或载体系统。 347. 一种工程化腺苷脱氨酶,其包含一个或多个突变,其

中所述工程化腺苷脱氨酶具有胞苷脱氨酶活性。

[1793] 348. 根据陈述347所述的工程化腺苷脱氨酶,其中所述工程化腺苷脱氨酶具有腺苷脱氨酶活性。

[1794] 349. 根据陈述347或348所述的工程化腺苷脱氨酶,其中所述工程化腺苷脱氨酶是融合蛋白的一部分。

[1795] 350. 根据陈述349所述的工程化腺苷脱氨酶,其中所述融合蛋白包含功能性结构域。

[1796] 351. 根据陈述350所述的工程化腺苷脱氨酶,其中所述功能性结构域能够引导所述工程化腺苷脱氨酶结合至靶核酸。

[1797] 352. 根据陈述350或351所述的工程化腺苷脱氨酶,其中所述功能性结构域是陈述1至318中任一项所述的CRISPR-Cas蛋白。

[1798] 353. 根据陈述352所述的工程化腺苷脱氨酶,其中所述CRISPR-Cas 蛋白是死形式的CRISPR-Cas蛋白或CRISPR-Cas切口酶蛋白。

[1799] 354. 根据陈述347-353中任一项所述的工程化腺苷脱氨酶,其中基于hADAR2-D的氨基酸序列位置,所述一个或多个突变包含:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T,以及在同源ADAR蛋白中的相应突变。

[1800] 355. 根据陈述347-354中任一项所述的工程化腺苷脱氨酶,其中基于hADAR2-D的氨基酸序列位置,所述一个或多个突变包含:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E和S661T,以及在同源ADAR蛋白中的相应突变。

[1801] 356. 一种多核苷酸,其编码陈述347-355中任一项所述的工程化腺苷脱氨酶或其催化结构域。

[1802] 357. 一种载体,其包含陈述356所述的多核苷酸。

[1803] 358. 一种药物组合物,其包含配制成用于通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪或可植入装置递送的陈述347-355 中任一项所述的工程化腺苷脱氨酶或其催化结构域。

[1804] 359. 一种工程化细胞,其表达陈述347-355中任一项所述的工程化腺苷脱氨酶或其催化结构域。

[1805] 360. 根据陈述359所述的工程化细胞,其中所述细胞瞬时表达所述工程化腺苷脱氨酶或其催化结构域。

[1806] 361. 根据陈述359或360所述的工程化细胞,其中所述细胞非瞬时表达所述工程化腺苷脱氨酶或其催化结构域。

[1807] 362. 一种用于在靶核酸中修饰核苷酸的工程化、非天然存在的系统,其包含:

[1808] a) 死CRISPR-Cas或CRISPR-Cas切口酶蛋白,或者编码所述死Cas 或Cas切口酶蛋白的核苷酸序列;

[1809] b) 指导分子,其包含与靶序列杂交并设计为与所述死CRISPR-Cas 或CRISPR-Cas切口酶蛋白形成复合物的指导序列;和

[1810] c) 核苷酸脱氨酶蛋白或其催化结构域,或者编码所述核苷酸脱氨酶蛋白或其催化

结构域的核苷酸序列,其中所述核苷酸脱氨酶蛋白或其催化结构域共价或非共价连接至所述死CRISPR-Cas或CRISPR-Cas切口酶蛋白,或者所述指导分子适于在递送后与其连接。

[1811] 363. 根据陈述362所述的系统,其中所述腺苷脱氨酶蛋白或其催化结构域基于hADAR2-D的氨基酸序列位置包含突变E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E、S661T中的一个或多个,以及在同源ADAR蛋白中的相应突变。

[1812] 364. 根据陈述362或363所述的系统,其中所述腺苷脱氨酶蛋白或其催化结构域基于hADAR2-D的氨基酸序列位置包含突变:E488Q、V351G、S486A、T375S、S370C、P462A、N597I、L332I、I398V、K350I、M383L、D619G、S582T、V440I、S495N、K418E和S661T,以及在同源ADAR蛋白中的相应突变。

[1813] 365. 根据陈述362-364中任一项所述的系统,其中所述CRISPR-Cas蛋白是Cas9、Cas12、Cas13、Cas14、CasX或CasY。

[1814] 366. 根据陈述362-365中任一项所述的系统,其中所述CRISPR-Cas蛋白是Cas13b。

[1815] 367. 根据陈述362-366中任一项所述的系统,其中所述CRISPR-Cas蛋白是Cas13b-t1、Cas13b-t2或Cas13b-t3。

[1816] 368. 根据陈述362-367中任一项所述的系统,其中所述CRISPR-Cas是根据陈述1-318中任一项所述的工程化CRISPR-Cas蛋白。

[1817] 369. 一种用于在靶核酸中修饰核苷酸的方法,其包括:

[1818] 将陈述347-355中任一项所述的工程化腺苷脱氨酶或陈述362-368中任一项所述的系统递送至所述靶核酸,其中所述脱氨酶使在所述靶核酸上的一个或多个靶基因座处的核苷酸脱氨基。

[1819] 370. 根据陈述369所述的方法,其中所述核苷酸脱氨酶蛋白或其催化结构域已被修饰以增加对DNA-RNA异源双链体的活性。

[1820] 371. 根据陈述369或370所述的方法,其中所述核苷酸脱氨酶蛋白或其催化结构域已被修饰以降低脱靶效应。

[1821] 372. 根据陈述369-371中任一项所述的方法,其中所述靶核酸在细胞内。

[1822] 373. 根据陈述372所述的方法,其中所述细胞是真核细胞。

[1823] 374. 根据陈述372或373所述的方法,其中所述细胞是非人动物细胞。

[1824] 375. 根据陈述372至374中任一项所述的方法,其中所述细胞是人细胞。

[1825] 376. 根据陈述372至375中任一项所述的方法,其中所述细胞是植物细胞。

[1826] 377. 根据陈述369至376中任一项所述的方法,其中所述靶核酸在动物内。

[1827] 378. 根据陈述369至377中任一项所述的方法,其中所述靶核酸在植物内。

[1828] 379. 根据陈述369至378中任一项所述的方法,其中所述靶核酸在体外包含在DNA分子中。

[1829] 380. 根据陈述369至377中任一项所述的方法,其中所述工程化腺苷脱氨酶或者所述系统的一种或多种组分作为核糖核蛋白复合物递送到所述细胞中。

[1830] 381. 根据陈述380所述的方法,其中所述工程化腺苷脱氨酶或者所述系统的一种或多种组分通过一种或多种颗粒、一种或多种囊泡或者一种或多种病毒载体递送。

- [1831] 382. 根据陈述381所述的方法,其中所述一种或多种颗粒包含脂质、糖、金属或蛋白。
- [1832] 383. 根据陈述381或382所述的方法,其中所述一种或多种颗粒包括脂质纳米颗粒。
- [1833] 384. 根据陈述381-383中任一项所述的方法,其中所述一种或多种囊泡包括外泌体或脂质体。
- [1834] 385. 根据陈述381-384中任一项所述的方法,其中所述一种或多种病毒载体包括一种或多种腺病毒载体,一种或多种慢病毒载体或者一种或多种腺相关病毒载体。
- [1835] 386. 根据陈述369-385中任一项所述的方法,其中所述方法通过在感兴趣的基因组基因座处操纵一个或多个靶序列来修饰细胞、细胞系或生物体。
- [1836] 387. 根据陈述386所述的方法,其中在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用治疗由G→A或C→T点突变或病原性SNP 引起的疾病。
- [1837] 388. 根据陈述387所述的方法,其中所述疾病选自癌症、血友病、β地中海贫血,马凡氏综合征和维斯科特-奥尔德里奇综合征。
- [1838] 389. 根据陈述386、387或388所述的方法,其中在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用治疗由T→C或A→G点突变或病原性SNP引起的疾病。
- [1839] 390. 根据陈述389所述的方法,其中在所述感兴趣的靶基因座处的所述核苷酸的所述脱氨基作用使在所述靶基因座处的靶基因失活。
- [1840] 391. 根据陈述380-390中任一项所述的方法,其中通过脂质体、纳米颗粒、外泌体、微囊泡、核酸纳米组件、基因枪、可植入装置或陈述302所述的载体系统递送所述工程化腺苷脱氨酶或者所述系统的一种或多种组分。
- [1841] 392. 根据陈述369-392中任一项所述的方法,其中所述核苷酸的修饰改变在所述靶基因座处编码的基因产物或所述基因产物的表达。
- [1842] 393. 根据陈述347-355中任一项所述的工程化腺苷脱氨酶或陈述 362-368中任一项所述的系统,其中所述腺苷蛋白或其催化结构域基于 hADAR2-D的氨基酸序列位置包含在S375上的突变,以及在同源ADAR 蛋白中的相应突变。
- [1843] 394. 根据陈述393所述的工程化腺苷脱氨酶或系统,其中所述在 S375上的突变是S375N。
- [1844] 395. 根据前述陈述中任一项所述的工程化CRISPR-Cas蛋白或工程化腺苷脱氨酶在制备用于治疗疾病的药物中的用途。
- [1845] 396. 一种药制剂,其包含前述陈述中任一项所述的工程化 CRISPR-Cas蛋白或工程化腺苷脱氨酶用作药物。
- [1846] 本发明在以下实施例中进一步描述,这些实施例不限制权利要求中描述的本发明的范围。
- [1847] 实施例1-与crRNA复合的Cas13b的晶体结构
- [1848] 方法
- [1849] 蛋白纯化用于结晶
- [1850] 在购自New England Biolabs的化学感受态BL21 DE3细胞中,用基于pET28的、具有在N端融合的twin-strep-sumo标签的载体中表达 PbuCas13b。将具有该表达质粒的细胞

在37℃生长至OD 0.2,然后将温度调至21℃。持续生长至OD 0.6,然后用5μM IPTG诱导。培养物生长 18-20小时,然后以5000rpm离心收获细胞并于-80℃冷冻。将冷冻细胞糊在补充有核酸酶 (benzonase) 和溶菌酶 (lysozyme) 的缓冲液A (500 mM氯化钠,50mM Hepes pH 7.5,2mM DTT)中匀浆。细胞在20000psi 下的微流化器中经过两次破碎,并通过10000rpm离心将细胞碎片从可溶性级分中分离出来。使可溶性级分通过Streptactin树脂 (GE life sciences),并用10倍柱体积的缓冲液A、然后是10倍柱体积的洗涤缓冲液 (1M氯化钠,50mM Hepes 7.5,2mM DTT),最后是10倍柱体积的裂解缓冲液 (400mM氯化钠,20mM Hepes 7.5,2mM DTT) 洗涤。通过添加5mM脱硫生物素 (desthiobiotin) (Sigma) 从树脂洗脱 PbuCas13b,然后在补充20mM DTT后,用sumo蛋白酶切割过夜。在切割后,该蛋白通过肝素柱,浓缩至500 μL,并通过在存储缓冲液 (500mM 氯化钠,10mM Hepes pH 7.0,2mM DTT) 中平衡的superdex 200柱 (GE life sciences)。合并峰级分并浓缩至至少20mg/ml。除了在每种缓冲液中补充5mM DTT外,类似地纯化硒代甲硫氨酸蛋白。用Pierce试剂 (Thermo) 定量蛋白。

[1851] 结晶和数据收集

[1852] 将RNA底物以2:1摩尔比添加到PbuCas13b蛋白中,并在透析缓冲液 (50mM氯化钠,10mM Hepes 7.0,2mM TCEP) 中透析7小时。用透析缓冲液将复合PbuCas13b+RNA稀释至10mg/ml,并在不同的滴比 (drop ratio) 下,在165mM柠檬酸钠pH 4.6,5.5%PEG6000和2mM TCEP下,通过悬挂滴液蒸汽扩散在20度设置。棒状晶体生长过夜并在1-2个月内达到最大尺寸。将晶体从滴液转移至冷冻稳定缓冲液 (140mM 柠檬酸钠pH 4.6,5%PEG6000,35%PEG400) 中,浸泡长达24小时,然后在液氮中快速冷冻。在补充5mM TCEP的相似条件下生长用于定相 (phasing) 的硒晶体。

[1853] 来自PbuCas13b和指导RNA的晶体的原生衍射数据是在高级光子源,Argonne国家实验室在光束23-ID-B/D上收集的,而异常数据是在金刚石光源在光束I04上收集的。使用准直 (23-ID) 或聚焦 (金刚石) 到 20微米的小光束,沿着晶体的长度收集多个数据集。在0.97934 (峰值), 0.97958 (拐点) 和0.97204 (远端) 埃波长处收集到异常数据集。使用 XDS (1,2) 处理衍射数据,并在autoPROC工具箱 (4) 中实施无目标缩放 (sclaed in aimless) (3)。统计总结在下表10中。

[1854] 表10

| 数据名                         | PbuCas13b-Se-峰               | PbuCas13b-Se-变形              | PbuCas13b-Se-远端              | PbuCas13b-原始                |
|-----------------------------|------------------------------|------------------------------|------------------------------|-----------------------------|
| 结构中的配体                      | 柠檬酸盐                         | 柠檬酸盐                         | 柠檬酸盐                         | 柠檬酸盐                        |
| 数据收集                        |                              |                              |                              |                             |
| 空间组                         | $P2_12_1$                    | $P2_12_1$                    | $P2_12_1$                    | $P2_12_1$                   |
| 细胞尺寸                        |                              |                              |                              |                             |
| a, b, c (Å)                 | 90.82, 124.65, 140.73        | 90.86, 124.72, 140.77        | 90.88, 124.76, 140.79        | 90.86, 125.03, 140.57       |
| $\alpha, \beta, \gamma$ (°) | 90, 90, 90                   | 90, 90, 90                   | 90, 90, 90                   | 90, 90, 90                  |
| 波长(Å)                       | 0.97934                      | 0.97958                      | 0.97204                      | 1.03320                     |
| 分辨率(Å)                      | 140.73 - 2.32 (2.47 - 2.32)* | 140.77 - 2.35 (2.53 - 2.35)* | 140.79 - 2.40 (2.62 - 2.40)* | 93.42 - 1.97 (2.07 - 1.97)* |
| 独特反射                        | 59325                        | 55600                        | 50274                        | 111373                      |
| $R_{int}$                   | 0.203 (1.732)*               | 0.207 (1.688)*               | 0.214 (1.741)*               | 0.245 (2.754)*              |
| I( $\sigma$ (I))            | 10.2 (1.4)*                  | 10.4 (1.5)*                  | 10.1 (1.5)*                  | 10.4 (1.8)*                 |
| CC1/2                       | 0.996 (0.7)*                 | 0.996 (0.710)*               | 0.996 (0.706)*               | 0.995 (0.572)*              |
| 完整度(%)                      | 94.2 (55.8)*                 | 94.0 (56.6)*                 | 93.8 (52.0)*                 | 97.8 (99.2)                 |
| 冗余度                         | 13.6 (12.3)*                 | 13.5 (12.5)*                 | 13.5 (12.7)*                 | 13.3 (14.1)*                |
| 细化                          |                              |                              |                              |                             |
| $R_{int}/R_{free}$ **       |                              |                              |                              | 0.170/0.2023                |
| 原子数                         |                              |                              |                              | 9111                        |
| 蛋白质                         |                              |                              |                              | 41                          |
| 配体                          |                              |                              |                              | 657                         |
| 水                           |                              |                              |                              |                             |
| B因子(Å <sup>2</sup> )        |                              |                              |                              |                             |
| 蛋白质                         |                              |                              |                              | 39.06                       |
| 配体                          |                              |                              |                              | 58.91                       |
| 水                           |                              |                              |                              | 40.20                       |
| R.m.s 偏差                    |                              |                              |                              |                             |
| 键长(Å)                       |                              |                              |                              | 0.005                       |
| 键角(°)                       |                              |                              |                              | 0.742                       |
| 拉马钱德兰分析 <sup>†</sup> (%)    |                              |                              |                              |                             |
| 有利的                         |                              |                              |                              | 97.01                       |
| 允许的                         |                              |                              |                              | 2.70                        |
| 异常的                         |                              |                              |                              | 0.29                        |

[1856] \*最高分辨率的外壳显示在括号中。

[1857] \*\*Rfree是使用5%的数据计算得出的。

[1858] #使用MolProbity程序(1)计算Ramachandran图中的二面角分布。

[1859] 参考文献:1.V.B.Chen等,MolProbity:all-atom structure validation for macromolecular crystallography.Acta Crystallographica Section D-Biological Crystallography 66,12-21(2010)。

[1860] 结构解析

[1861] PbuCas13b的晶体结构通过使用硒作为反常散射的多波长反常衍射(MAD)进行解析。用phenix.autosol(5,6)测定并细化27个SeMet位点的位置。通过使用具有优值(merit)0.35的图的3.5Å分辨率实验地图的phenix.autobuild(7)构建了局部模型。使用硒实验地图完成了Coot(8,9)中的手动重建和phenix.refine(10-12)中的细化的循环。使用PHENIX中的反射文件编辑器,将无R标志和实验阶段从硒数据传输至高分辨率原始数据。这些反射被用于在Coot中重建和在phenix.refine中细化的进一步循环。异常差异地图用于确保正确注册。phenix.refine中的细化使用了TLS(平移、振动和转动),以及位置和单独B因子细化。柠檬酸盐约束是由phenix.elbow产生(13)。最终模型包含一条多肽链、一条RNA核苷酸链、两个柠檬酸盐分子、一个四乙二醇(PG4)分子、两个C1原子和657个水分子。图是用PyMol软件(14)创建的。

[1862] 结构分析

[1863] 使用DSSR(15)分析RNA结构。使用Consurf服务器(16)完成到结构的蛋白质保守映射。使用PDBSUM Web服务器分析蛋白二级结构(17)。APBS作为PyMol可视化程序的一部分用于计算静电(18)。

[1864] 蛋白质比对

[1865] 如在Geneious(19)中实施的,使用ClustalW或Muscle对Cas13b酶进行比对。使用Jukes-Cantor距离模型生成邻接树。在成功地将HEPN结构域活性位点残基与其他家族成

员进行匹配的树亚组上进行结构分析的保守比对(图14-16)。

#### [1866] 凝胶过滤实验

[1867] 指导序列复合物的形成:在室温下,在100 $\mu$ L缓冲液(125mM NaCl, 10mM HEPES pH 7.0, 2mM TCEP)中,将100 $\mu$ g PbuCas13b与两摩尔当量的指导RNA一起孵育20分钟。指导序列-靶标复合物的形成:如上所述,将100 $\mu$ g PbuCas13b与两摩尔当量的指导RNA共孵育20分钟。然后将两摩尔当量的靶RNA加入到溶液中,并将混合物在室温下再孵育 20分钟(共100 $\mu$ L, 125mM NaCl, 10mM HEPES pH 7.0, 2mM TCEP)。类似地, Apo蛋白在125mM NaCl, 10mM HEPES pH 7.0, 2mM TCEP 的缓冲液中同样稀释至1 $\mu$ g/ $\mu$ L。将样品从2mL毛细管注射到GE Superdex 200Increase 10/300GL柱上,并用500mM NaCl, 10mM HEPES pH 7.0, 2mM DTT缓冲液运行。

#### [1868] ThermoFluor融化测定

[1869] 方案改编自(20)。在终浓度为50mM NaCl, 10mM HEPES pH 7.0, 6.25x SYPRO™橙色染料的溶液中, (如上制备的apo、指导序列、或指导序列-靶标复合物)将样品制备成具有1 $\mu$ g PbuCas13b的20 $\mu$ L最终体积。对于MgCl<sub>2</sub>切割和结合实验,将终浓度6mM的Mg<sup>2+</sup>加入到上述缓冲混合物中。对于用非互补RNA进行的对照实验,将2摩尔当量的 RNA与蛋白复合物一起孵育。在Roche LightCycler 480II上一式三份进行融化实验。

#### [1870] 受限的蛋白水解

[1871] 将10 $\mu$ g PbuCas13b与crRNA或crRNA及靶标一起在室温下孵育30 分钟。加入400 $\mu$ g蛋白酶(胰蛋白酶、糜蛋白酶(Chemotrypsin)或胃蛋白酶),并将混合物在37 $^{\circ}$ C下孵育5分钟,然后快速放置在冰上2分钟,再添加SDS上样缓冲液并在4-12%丙烯酰胺凝胶上运行。

#### [1872] PbuCas13b前crRNA加工突变体的蛋白表达和纯化

[1873] 使用来自TwinStrep-SUMO-PbuCas13b表达质粒的PIPE定点诱变克隆产生在每个推定的crRNA加工催化残基处的丙氨酸突变体,并将其转化到BL21 (DE3) pLysE大肠杆菌细胞中。对于每个突变体,在补充了100  $\mu$ g/mL氨苄青霉素的2L Terrific Broth培养基(12g/L胰蛋白胨, 24g/L 酵母提取物, 9.4g/L K<sub>2</sub>HPO<sub>4</sub>, 2.2g/L KH<sub>2</sub>PO<sub>4</sub>)中接种15mL过夜起始培养物并生长至OD<sub>600</sub>为0.4-0.6。加入0.5mM IPTG诱导蛋白表达,并在21 $^{\circ}$ C以250RPM振荡速度进行16小时。5000RPM离心10分钟收集细胞,并将糊状物直接用于蛋白纯化(总细胞糊10-20g)。对于裂解,通过在4 $^{\circ}$ C搅拌下将细菌糊状物重悬于50mL裂解缓冲液(50mM Tris-HCl pH 7.5, 500mM NaCl, 1mM DTT)中,其中补充了50mg溶菌酶, 1片蛋白酶抑制剂(cOmplete, 无EDTA, Roche Diagnostics Corporation)和500U核酸酶(Sigma)。将悬浮液在25000psi下通过 LM20微流化器,并通过在10000RPM, 4 $^{\circ}$ C下离心1小时来使裂解物澄清。将裂解物与2mL StrepTactin超流动树脂(Oiagen)在旋转振荡器上 4 $^{\circ}$ C孵育2小时。将与蛋白结合的树脂用10mL裂解缓冲液洗涤三次,然后在20mL IGEPAL裂解缓冲液(0.2% IGEPAL)中加入50 $\mu$ L SUMO蛋白酶(内部)。SUMO标签的切割和天然蛋白的释放在Econo柱色谱柱中于4 $^{\circ}$ C在台式摇床上轻轻混合下过夜进行。收集经切割的蛋白作为流通液,用5mL裂解缓冲液洗涤三次并在SDS-PAGE凝胶上检查。

[1874] 用不含盐的离子交换缓冲液A(50mM Tris-HCl pH 7.5, 1mM DTT)将蛋白质稀释两倍,以使NaCl起始浓度为250mM。然后将蛋白质上样到5mL肝素HP色谱柱(GE Healthcare Life Sciences),并在250mM至 1M的NaCl梯度上洗脱。通过SDS-PAGE凝胶和考马斯亮蓝染

色分析洗脱的蛋白级分(约700mM),合并并使用50MWCO离心滤器(Amicon)浓缩至1mL。将浓缩的蛋白上样到预先平衡的尺寸排阻柱上,并使用含有50mM Tris-HCl pH 7.5,500mM NaCl,2mM DTT的S200缓冲液洗脱。通过SDS-PAGE凝胶和考马斯亮蓝染色分析单分散蛋白级分,然后浓缩并进行缓冲液交换至蛋白存储缓冲液(600mM NaCl,50mM Tris-HCl pH 7.5,1mM DTT)中。

[1875] 前crRNA加工测定

[1876] 订购了用于前crRNA加工和核酸酶测定的RNA作为Ultrasomers (IDT),并使用HiScribe T7快速高产RNA合成试剂盒(New England Biolabs)进行体外转录。用AmpureXP RNA清洁珠纯化RNA,并保存在-20°C以备将来使用。为了测试前crRNA加工,将WT和突变蛋白与前crRNA孵育,蛋白质相对RNA为四倍摩尔过量。前crRNA加工是在Cas13b crRNA加工缓冲液(10mM TrisHCl pH 7.5,50mM NaCl,0.5mM MgCl<sub>2</sub>,20U SUPERase(ThermoFisher Scientific),0.1%BSA)中于37°C进行30分钟,通过加入2x TBE-Urea凝胶上样缓冲液停止,95°C下变性5分钟。立即将样品置于冰上10分钟,然后将其在1x TBE缓冲液中的15% TBE-Urea凝胶上以200V运行40分钟。在1x TBE中的1x Sybr Gold 上进行凝胶染色15分钟,在BioRad gel doc系统上成像。

[1877] 前crRNA突变体的荧光附带RNA切割测定

[1878] 在核酸酶测定缓冲液(20mM HEPES,60mM NaCl,6mM MgCl<sub>2</sub>,pH 6.8)和0.5μL鼠RNase抑制剂(New England Biolabs)和125nM poly-U 均聚物RNA传感器(Trilink)中,以等摩尔比的PbuCas13b或PbuCas13b 突变体、crRNA和RNA靶标一式四份进行检测测定。将样品在装备了FAM过滤器组的荧光读板器上于37°C孵育3小时。每隔5分钟记录一次测量,并将数据标准化到第一个时间点。

[1879] 切割片段文库

[1880] 为了绘制Cas14切割产物的图谱,如上所述,用LwCas13a和 PbuCas13b,它们各自的crRNA和靶标RNA或对照进行体外切割反应。切割进行5分钟或30分钟,并使用RNA寡聚清洁和浓缩试剂盒(Zymo research)进行纯化。根据NEB Multiplex小分子RNA测序试剂盒在Illumina NextSeq 500仪器上制备了小RNA测序文库。

[1881] 设计和克隆用于RNA编辑的哺乳动物构建体

[1882] 通过将HEPN结构域的催化位点中的两个精氨酸和两个组氨酸残基突变为丙氨酸(R146A/H151A/R1116A/H1121A),使PguCas13b催化失活(dPguCas13b)。将这些催化失活的Cas13b Gibson克隆到pcDNA-CMV 载体主链中,该主链包含通过GS接头(21)与Cas13b的C末端融合的ADAR2(E488Q)的脱氨酶结构域。为了产生截短形式,设计了引物以PCR扩增从C末端截短60bp(20个氨基酸)逐渐至900bp的dCas13b(共15个截短),并将这些截短的Cas13b基因Gibson克隆到上述pcDNA-CMV-ADAR2主链中。使用金门克隆法将靶向Cluc的指导RNA克隆到哺乳动物表达载体中,该载体包含该直向同源物在间隔序列目的位点的3'端处的正向重复序列,在U6启动子下。

[1883] 使用的荧光素酶报告基因是Cox等(2017)用来测量RNA编辑的CMV-Cluc(W85X)EF1α-Gluc双荧光素酶报告基因(21)。该报告基因载体将功能性Gluc表达为标准化对照,但是由于添加了W85X预终止位点,因此存在缺陷性Gluc。

[1884] 哺乳动物细胞培养

[1885] 哺乳动物细胞培养实验是在HEK293FT细胞系(American Type Culture Collection(ATCC))中进行的,其在含高葡萄糖、丙酮酸钠和 GlutaMAX(Thermo Fisher Scientific)、另外还补充了1×青霉素-链霉素(Thermo Fisher Scientific)和10%胎牛血清(VWR Seradigm)的 Dulbecco's Modified Eagle培养基中生长。

[1886] 所有转染均在96孔板上用Lipofectamine 2000(Thermo Fisher Scientific)进行。转染前16-18小时,将细胞以约20000细胞/孔铺板以确保转染时达到90%的融合度。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)混合至25 $\mu$ l。分别地将 24.5 $\mu$ l的Opti-MEM与0.5 $\mu$ l的Lipofectamine 2000混合。然后将质粒和脂质体转染溶液混合并移液到细胞上。

[1887] 哺乳动物中的RNA敲除

[1888] 为了用报道基因构建体评估哺乳动物中的RNA靶向,将150ng的 Cas13构建体与300ng的指导序列表达质粒和45ng的双荧光素酶报告基因构建体一起共转染。转染后48小时,收获含有分泌的荧光素酶的培养基,并在具有注射方案的读板仪(Biotek Synergy H4)上用BioLux Cypridinia和Biolum Gaussia荧光素酶测定试剂盒(New England Biolabs)测量活性。将来自被靶向的Gluc的信号相对于来自未被靶向的Gluc的信号进行标准化,随后将PbCas13b突变体荧光素酶信号的实验相对于仅使用指导序列荧光素酶信号的实验进行标准化(三个生物学重复的平均值)。进行的所有重复都是生物学重复。

[1889] 哺乳动物细胞中的REPAIR编辑

[1890] 为了评估哺乳动物细胞中的REPAIR活性,申请人转染了150ng REPAIR载体、300ng指导序列表达质粒和45ng RNA编辑报告基因。然后申请人在48小时后收获了含分泌的荧光素酶的培养基,并在 Dulbecco磷酸盐缓冲液(PBS)中以1:10的比例稀释该培养基(10 $\mu$ l培养基至90 $\mu$ l PBS中)。申请人在具有注射方案的读板仪(Biotek Synergy Neo2)上使用BioLux Cypridinia和Biolum Gaussia荧光素酶测定试剂盒(New England Biolabs)测量了荧光素酶活性。进行的所有重复都是生物学重复。

[1891] 参考文献:

[1892] 1.W.Kabsch,Integration,scaling,space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr 66,133-144(2010).

[1893] 2.W.Kabsch,Xds.Acta Crystallogr D Biol Crystallogr 66,125-132(2010).

[1894] 3.P.R.Evans,G.N.Murshudov,How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr 69,1204-1214(2013).

[1895] 4.C.Vonrhein等,Data processing and analysis with the autoPROC toolbox.Acta Crystallogr D Biol Crystallogr 67,293-302(2011).

[1896] 5.P.D.Adams等,PHENIX:a comprehensive Python-based system for macromolecular structure solution.Acta Crystallogr D Biol Crystallogr 66,213-221(2010).

[1897] 6.T.C.Terwilliger等,Decision-making in structure solution using Bayesian estimates of map quality:the PHENIX AutoSol wizard.Acta Crystallogr D Biol Crystallogr 65,582-601(2009).

[1898] 7.T.C.Terwilliger等,Iterative model building,structure refinement and

density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr* 64, 61-69 (2008).

[1899] 8. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66, 486-501 (2010).

[1900] 9. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132 (2004).

[1901] 10. P. V. Afonine 等, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* 68, 352-367 (2012).

[1902] 11. N. Echols 等, Automated identification of elemental ions in macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 70, 1104-1114 (2014).

[1903] 12. P. H. Zwart 等, Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426, 419-435 (2008).

[1904] 13. N. W. Moriarty, R. W. Grosse-Kunstleve, P. D. Adams, electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr* 65, 1074-1080 (2009).

[1905] 14. The PyMOL Molecular Graphics System, Version 2.0 **Schrödinger**, LLC.

[1906] 15. X. J. Lu, H. J. Bussemaker, W. K. Olson, DSSR: an integrated software tool for dissecting the spatial structure of RNA. *Nucleic Acids Res* 43, e142 (2015).

[1907] 16. H. Ashkenazy 等, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* 44, W344-350 (2016).

[1908] 17. T. A. de Beer, K. Berka, J. M. Thornton, R. A. Laskowski, PDBsum additions. *Nucleic Acids Res* 42, D292-296 (2014).

[1909] 18. E. Jurrus 等, Improvements to the APBS biomolecular solvation software suite. *Protein Sci* 27, 112-128 (2018).

[1910] 19. L. A. Ripma, M. G. Simpson, K. Hasenstab-Lehman, Geneious! Simplified genome skimming methods for phylogenetic systematic studies: A case study in *Oreocarya* (Boraginaceae). *Appl Plant Sci* 2, (2014).

[1911] 20. K. Huynh, C. L. Partch, Analysis of protein stability and ligand interactions by thermal shift assay. *Curr Protoc Protein Sci* 79, 28 29 21-14 (2015).

[1912] 21. D. B. T. Cox 等, RNA editing with CRISPR-Cas13. *Science* 358, 1019-1027 (2017).

[1913] 结果

[1914] VI型CRISPR-Cas系统包含可编程的单效应子RNA指导的Rnase, 其包括Cas13b(它是四个已知的VI型亚型家族成员之一)。Cas13b在其线性结构域结构和CRISPR RNA (crRNA) 结构方面在这些蛋白家族中是独一无二的。申请人报道了结合至crRNA的颊普雷沃菌

Cas13b (PbuCas13b) 在1.97埃分辨率下的晶体结构。该结构表明,该指导RNA 在Cas13b内通过介导核酸酶活性的直接和间接相互作用的网络而进行配位。申请人鉴定了用于crRNA加工的第二个活性位点并显示该位点中关键残基的突变消除了加工活性。申请人还发现HEPN2核酸酶结构域对于RNA靶向是非必需的,并建立了用Cas13b进行RNA靶向的结构导向工程化的基础。

[1915] 在此申请人报道了与crRNA手柄(handle)和部分间隔子复合的来自颊普雷沃菌的Cas13b(PbuCas13b)在1.97埃分辨率下的结构。我们的结构揭示了Cas13b核酸酶的整体结构以及crRNA识别和切割的分子基础。

[1916] 申请人解析了与36-核苷酸正向重复序列和短5-核苷酸间隔子复合的PbuCas13b的晶体结构(图1)。类似于其他2类CRISPR效应物,PbuCas13b的整体形状是双叶的(13-19)。在该结构中有五个结构域很明显:两个HEPN结构域(HEPN1和HEPN2),两个主要螺旋结构域(螺旋-1和螺旋-2),以及用两个 $\beta$ 发夹覆盖crRNA的3'末端的结构域(盖结构域)(图1、图18)。为了鉴定与蛋白质数据库中其他结构域的相似性,使用DALI服务器查询完整的PbuCas13b结构以及分离的结构域(15)。HEPN1与LshCas13a的HEPN2结构域匹配。

[1917] 两个HEPN结构域都主要是 $\alpha$ 螺旋:HEPN1由十二个线性连接的 $\alpha$ 螺旋组成,在螺旋之间具有柔性环。HEPN2由九个 $\alpha$ 螺旋、几个短 $\beta$ 链和在末端带有带电荷残基的 $\beta$ 发夹组成,其指向活性部位口袋。HEPN2靠在HEPN1上,使得活性位点残基(R156、N157、H161和R1068、N1069、H1073)组装成经典HEPN活性位点,尽管在线性蛋白的N-和C-末端处(图1)(3、17、18、20)。HEPN1结构域通过高度保守的结构域间接头(IDL)连接到螺旋-1结构域,该接头穿过大的、带正电的内部通道的中心(图1)。将IDL的保守残基(R285、K292、E296)突变为丙氨酸降低了PbuCas13b通过切割荧光素酶mRNA来干扰哺乳动物细胞中荧光素酶表达的能力,证明了在一般核酸酶活性中的作用(图5A、C)。

[1918] 螺旋-1被螺旋-2和盖结构域线性拆分为三个片段。螺旋-1使大量的糖磷酸和核碱基与正向重复RNA接触(图2、图3)。螺旋-1还与HEPN1/2和盖结构域两者有少量的界面接触。盖结构域混合了 $\alpha$ 和 $\beta$ 二级结构,并用两个带电荷的 $\beta$ 发夹覆盖了正向重复RNA的3'游离端。两个 $\beta$ 发夹中较长的一个穿过RNA环以接触螺旋-1结构域,从而在游离RNA末端上形成了盖。来自盖结构域的带正电残基指向贯穿蛋白复合物的中心的大型中央通道(图1、图6C)。在盖结构域的无序环(K431至T438)和两个HEPN结构域之间形成了从外部溶剂渗透到内部通道的带正电荷的侧通道(图6A)。螺旋-2结构域由十一个 $\alpha$ -螺旋组成,并通过其与螺旋-1的连接包裹在正向重复RNA的主体下。螺旋-2与HEPN1结构域广泛连接,并与盖结构域的延伸 $\beta$ -发夹少量接触。第二个带正电的侧通道在螺旋-1和螺旋-2之间,为crRNA提供了大量的溶剂可及性(图6B)。所有结构域、IDL和crRNA形成该大型中央通道,其内部衬有带正电的残基(图6C)。

[1919] 与ADAR脱氨酶结构域融合的核酸酶死亡Cas13b用于REPAIR以实现靶向RNA碱基编辑(11)。AAV介导的递送常用于基因治疗,但REPAIR超过了AAV的载货能力的尺寸限制(11、21)。申请人先前显示了颊普雷沃菌P5-125(PspCas13b)的C端截短不降低REPAIR活性。申请人还使用了来自喉管卟啉单胞菌的另一种Cas13b直向同源物(PguCas13b),其与PbuCas13b相反,在哺乳动物细胞中稳定表达并显示出高活性(11)。基于PbuCas13b和PguCas13b之间的比对,申请人进行截短以移除HEPN2结构域,将其融合至ADAR,并测试其用REPAIR系统进

行碱基编辑的能力。出人意料的是,不仅是这些截短的突变体保留了RNA靶向,而且有些在RNA编辑方面显著更加有效率(图7)。

[1920] Cas13b已显示在crRNA有各种长度、间隔子范围为30至84个核苷酸的REPAIR中有效地发挥作用(11)。所有RNA碱基的明确的密度使得正向重复RNA的完整模型得以建立。该结构表明Cas13b通过广泛糖-磷酸和核碱基相互作用来识别正向重复序列(图2和3)。该正向重复序列大部分埋在两个螺旋结构域和盖结构域之间,但略微从螺旋-1突出,这说明了Cas13b如何能够利用替代的、较长的crRNA。整个crRNA结构是变形的A型双链体,其包含茎(碱基G(-1)-G(-4),C(-33)-C(-36))、环(C(-5)-U(-8),A(-29)-A(-32))、茎(U(-9)-U(-14),A(-23)-A(-28))、凸起(C(-15),G(-21))和发夹环(U(-16)-U(-20))构造(图2和3)。螺旋-1和螺旋-2介导crRNA发夹与盖结构域一起的直接和间接识别,该盖结构域覆盖了3'游离端。

[1921] C(-8)、U(-20)和A(-29)这三个碱基从RNA的主体中翻转出来。T754的主链羰基通过与碱基N4胺相互作用而使翻转出来的、高度保守的C(-8)碱基稳定,使该碱基保留在螺旋-1和螺旋-2结构域中的高度保守残基(Y540、566-571、K751、753-761)的疏水口袋中。该碱基翻转通过C(-8)N3<sup>+</sup>和U(-7)的糖(O2<sup>-</sup>)之间的相互作用而进一步稳定。如热稳定性测定中测量的,将C(-8)改为G或U降低核酸酶活性,并使蛋白质-RNA复合物去稳定(图3)。U(-20)在Cas13b正向重复序列中也是绝对保守的,并通过完全保守的残基来配位,最著名的是与核碱基O2接触的R762,和插入在G(-21)和U(-20)之间的R874,从而维持碱基在外并与U(-20)糖O4<sup>-</sup>接触。将R762突变为丙氨酸大大减少哺乳动物细胞中的RNA干扰(图5A)。将U(-20)突变为G降低核酸酶活性(图3)。与C(-8)和U(-20)相反,A(-29)在Cas13b正向重复序列中不保守,并且是不通过任何氨基酸来配位的核碱基。相反,A(-29)参与与G(-26)和C(-11)的多重碱基配对(图8F)(22)。A(-29)可以耐受向任何其他碱基的身份变化,但是突变为G略微降低一般核酸酶活性(图9)。影响一般核酸酶活性的碱基身份变化也降低了Cas13+crRNA复合物的热稳定性。与该观察结果一致,申请人发现将U(-27)和G(-10)之间的摆动碱基配对变为沃森-克里克碱基配对增加了一般核酸酶活性(图2D)。但是,将A(-32)变为G(这也产生沃森克里克碱基配对)降低了稳定性并降低了RNase活性(图2D)。

[1922] 发夹环被来自螺旋-2结构域内的高度保守残基的蛋白质相互作用的网络所识别(图3)。K870与来自U(-16)和U(-19)两者的O4配位,其在发夹转折处间接地将U(-17)翻转到溶剂中,无可见的残基接触。W842与U(-18)的核碱基堆叠,同时与K846一起与磷酸盐主链相互作用。R877和E873通过与碱基N3和O2位置的相互作用而进一步稳定了U(-18)。R874和R762分别通过糖O4<sup>-</sup>和碱基O2<sup>-</sup>相互作用来稳定U20位置。

[1923] crRNA的发夹环末端(-1至-4和-33至-36)呈螺旋状,并通过碱基和主链相互作用的联合而被识别(图3)。值得注意的是,N653和N652与U(-2)和C(-36)进行关键性小沟直接接触,并配位发夹的5'和3'末端。这些碱基身份的破坏或N653或N652突变为丙氨酸,在体外和哺乳动物干扰试验中均显著降低了Cas13b活性(图2E、图5)。C(-33)通过N756经由核碱基O2和糖O2<sup>-</sup>配位,将该C改变为A或G消除了一般RNase活性且降低了蛋白质稳定性(图2D、图9)。

[1924] RNA发夹末端(核苷酸-17至-20)通过广泛的磷酸主链氢键键合和碱基相互作用而稳定(图2、3)。将U(-18)突变为G消除了一般核酸酶活性。对于U(-19)或U(-20)观察到相同

的结果,但其他碱基是耐受性的,这表明G 06或N2核碱基原子破坏了核酸酶活性(图9)。

[1925] 结晶的RNA底物包括5碱基的间隔子序列(U1-G5),尽管在密度上只有该正向重复的5' 第一个核苷酸可见。该RNA正向重复的5' 端和间隔子的第一个碱基由来自螺旋-2的残基支持,并指向盖和HEPN结构域之间的中央通道中并指向侧通道(图3)。U(1)不受碱基特异性接触的配位,但位于盖结构域的带净正电口袋中。间隔子U(1)附近带电荷和芳香族氨基酸的突变对一般核酸酶活性影响很小,这表明这些残基对间隔子 RNA的配位不存在或不是必需的(图5H)。

[1926] 一些2类CRISPR系统将长前crRNA加工成成熟crRNA(3、7、11、23)。Cas13b已显示在3' 端处加工其自身的crRNA(3)。许多高度保守的残基与该RNA的3' 末端接触或在其附近,并可能形成第二个非HEPN核酸酶位点。为了测试第二个核酸酶位点,申请人突变了在该3' RNA末端附近的四个保守残基,并测试了这些突变体的crRNA加工和靶向激活核酸酶活性(图2)。K393当突变为丙氨酸时消除了RNA加工,但保留了靶向核酸酶活性,从而确认了负责crRNA加工的盖结构域中的第二个核酸酶位点的位置(图2、6、10)。R482A轻微影响了crRNA加工,但显著影响了一般核酸酶活性。这可能是由于使crRNA稳定的重要性(图2)。

[1927] 该解析的间隔子核苷酸指向HEPN叶并指向带正电的通道中。但是,该通道的大小不足以容纳RNA双链体,这表明Cas13b响应于靶标结合而采取了开放构象。申请人使用热变性测定法测量了apo、指导序列、和指导序列+靶标RNA复合物中Cas13b构象的变化。靶标结合的Cas13b 相对于仅有指导序列的Cas13b采用较不稳定的构象,但在非靶标RNA 存在的情况下没有观察到这种变化(图11)。有限的蛋白水解得到相似的结果;指导序列+靶标结合的复合物比仅有指导序列的复合物具有更低的蛋白酶抗性。

[1928] 尽管该晶体的不对称单元中只有一个分子,但来自一个单体的环与另一个形成反式接触,从而在活性位点配位了来自结晶缓冲液的结合柠檬酸盐。为了测试该反式亚基(trans-subunit)接触是否是功能性的,以及PbuCas13b是否通过该环而反式协同作用,申请人对该环的尖端处的残基(Q646和N647)进行了突变以观察它们是否会影响活性。每个突变均降低了哺乳动物细胞中的RNA干扰,启示了一般核酸酶活性的反式亚基调节的可能性(图5F)。

[1929] 最后,申请人将Cas13b与LshCas13a的结构进行了比较(图4)(17)。除了这些家族成员之间的一般功能相似性外,核酸酶之间有结构相似性,尤其是在HEPN结构域和活性位点结构中(图4B、C)。但是,SAS搜索提供了与(以前称为LbCpf1)的晶体结构的匹配,并突出了Cas13b 内的桥螺旋样亚结构域(24)。尽管该结构域在Cas13b家族中保守性很差,但它似乎是与介导必需核酸接触的Cas12a共同的结构特征(图5D、13)。考虑到Cas13b和Cas12a之间的根本差异,申请人假定该桥螺旋会聚地出现且并不表明这两种蛋白质的共同祖先。尽管如此,申请人将此特征称为桥螺旋,以与其他2类效应物的命名法保持一致(1、14)。

[1930] 下表11列出了产生和测试的示例性PbCas13b突变体。

[1931] 表11:测试RNA干扰的突变的列表。突变和来自三个生物学重复的平均标准化荧光值的列表。

[1932]

| 突变    | 指导序列 1 标准化 RLU | 指导序列 2 标准化 RLU |
|-------|----------------|----------------|
| R53A  | 0.747858       | 0.618255       |
| R53K  | 0.533437       | 0.415443       |
| R53D  | 0.708809       | 0.656473       |
| R53E  | 0.653859       | 0.505983       |
| Y164A | 0.560344       | 0.423418       |
| Y164F | 0.555361       | 0.419603       |
| Y164W | 0.578905       | 0.411809       |
| K183A | 0.611807       | 0.434156       |
| K193A | 0.637075       | 0.435537       |
| R285A | 0.621679       | 0.473138       |
| K292A | 0.709821       | 0.47966        |
| E296A | 0.753062       | 0.402674       |
| N297A | 0.697938       | 0.407599       |
| T405A | 0.668786       | 0.366621       |
| H407A | 0.541401       | 0.358297       |
| H407Y | 0.503637       | 0.335036       |
| H407W | 0.528546       | 0.359063       |
| H407F | 0.495551       | 0.341844       |
| K457A | 0.632984       | 0.441894       |

|        |          |          |
|--------|----------|----------|
| H500A  | 0.549885 | 0.34935  |
| K570A  | 0.575468 | 0.362485 |
| K590A  | 0.587262 | 0.383572 |
| R600A  | 0.565624 | 0.417064 |
| K607A  | 0.687397 | 0.430726 |
| R614A  | 0.744806 | 0.450827 |
| N634A  | 0.661617 | 0.386325 |
| R638A  | 0.696471 | 0.410163 |
| Q646A  | 0.675146 | 0.372062 |
| N647A  | 0.677884 | 0.400227 |
| N652A  | 0.665943 | 0.406755 |
| N653A  | 0.650794 | 0.384887 |
| K655A  | 0.900461 | 0.58631  |
| S658A  | 0.625349 | 0.363679 |
| K741A  | 0.648908 | 0.401644 |
| K744A  | 0.651516 | 0.42232  |
| N756A  | 0.650638 | 0.447333 |
| S757A  | 0.618393 | 0.402225 |
| R762A  | 0.862666 | 0.577076 |
| R791A  | 0.644193 | 0.444169 |
| K826A  | 0.600621 | 0.395086 |
| K828A  | 0.619022 | 0.416127 |
| K829A  | 0.593882 | 0.404272 |
| K846A  | 0.576794 | 0.407463 |
| K857A  | 0.595231 | 0.40528  |
| R877A  | 0.683229 | 0.461362 |
| K943A  | 0.573508 | 0.394379 |
| K943R  | 0.60418  | 0.403167 |
| K943D  | 0.69041  | 0.386955 |
| K943E  | 0.702192 | 0.372508 |
| R1041A | 0.662393 | 0.371243 |
| R1041K | 0.629986 | 0.376016 |
| R1041D | 0.813623 | 0.680736 |
| R1041E | 0.842593 | 0.484389 |
| D397A  | 0.63295  | 0.376658 |
| E398A  | 0.557761 | 0.365275 |
| D399A  | 0.590279 | 0.368724 |
| E400A  | 0.560351 | 0.349016 |
| D434A  | 0.524497 | 0.364659 |
| R618A  | 0.611197 | 0.401871 |
| R830A  | 0.663284 | 0.405163 |
| Q831A  | 0.548391 | 0.351777 |
| K835A  | 0.503864 | 0.373504 |
| K836A  | 0.503571 | 0.374439 |
| R838A  | 0.549749 | 0.372399 |
| WT     | 0.563282 | 0.335038 |

[1933]

[1934] PbuCas13b的结构为VI型蛋白家族的结构多样性提供了新的信息,并凸显了Cas13a和b之间的差异和相似性。申请人显示了crRNA识别和加工的结构基础,并揭示了核

酸酶活性在指导RNA和蛋白两者中的关键调节剂。基于PbuCas13b的结构,申请人能够生成REPAIR平台的较小变体,其保持碱基编辑效率并可以打包到AVV中。我们的数据表明在靶标识别过程中发生了主结构域重构。来自PbuCas13b的结构洞见使得能够进行合理工程化以改善RNA靶向特异性、碱基编辑和核酸检测的功能性(11、12、25、26)。

[1935] 实施例2

[1936] 图19显示了pymol文件,其显示了Cas13b的活性位点中的配位核苷酸的位置。这是基于与U核苷酸复合的RNaseL的晶体结构的结构比对。该比对将该核苷酸置于Cas13b的活性位点内,并揭示了可能的残基相互作用。图中标注了涉及碱基特异性的环。

[1937] 实施例3

[1938] RNA环可以被延伸。延伸的RNA指导环可以添加功能性RNA基序。图20显示了示例性RNA环延伸。

[1939] 实施例4

[1940] 图21显示了示例性融合点,核苷酸脱氨酶通过该融合点连接至 Cas13b。该融合点可以是Cas13b上的一个或多个氨基酸。例如,融合点可以是氨基酸411-429、114-124、197-241和607-624中的一个或多个。在一个实例中,该等氨基酸存在于颊普雷沃菌Cas13b中。

[1941] 实施例5

[1942] 使用酵母筛选来筛选影响ADAR活性的ADAR中的突变。该筛选进行了多轮。每轮筛选产生一组候选突变。然后在哺乳动物细胞中验证所述候选突变。表现最好的突变被加入到突变的最新版中并进行重新筛选。下表显示了在10轮中筛选出的突变。在第n轮中鉴定的突变体称为“RESCUE v<sub>n</sub>-1”。如本文所述,RESCUE是指将腺苷脱氨酶活性转化为胞苷脱氨酶活性的突变。

[1943] 表12

| RESCUE 轮次 | ADAR 突变  | 质粒      |
|-----------|--|---------|
| RESCUEv0  | E488Q  | pAB0048 |
| RESCUEv1  | E488Q, V351G   | pAB0359 |
| RESCUEv2  | E488Q, V351G, S486A  | pAB1188 |
| 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 |
| RESCUEv7  | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I               | pAB1194 |
| RESCUEv8  | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V        | pAB1220 |
| RESCUEv9  | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I | pAB1327 |

[1946] 进行了RESCUE v<sub>9</sub>的突变的筛选(图22)。在T翻转指导序列(图 23)和C翻转指导序列(图24)上验证了RESCUEv<sub>9</sub>的作用。使用 RESCUEv<sub>9</sub>,实现对T、A和C基序的至少约60%编辑,对G基序的25%编辑。用内源性靶向(与T翻转指导序列)测试了RESCUEv<sub>9</sub>的性能(图

25)。

[1947] 进行了RESCUE v10的突变的筛选(图26)。

[1948] 测试了30-bp指导序列的C翻转(图27)。

[1949] 用RESCUE v1至v8进行了Cas13b6和Cas13b12之间的比较。Gluc/Cluc结果显示在图28中,编辑分数结果显示在图29中,以及 RESCUEv8对内源性靶向的作用(T-翻转)显示在图30中。

[1950] 比较了RESCUE对碱基转化(C至U和A至I活性)的作用(图31)。测试了CCN 3' 基序靶向(图32)。

[1951] 实施例6

[1952] 产生(图33A)并测试(图33B)了通过接头与ADAR融合的具有各种死Cas13b(包括dCas13b)的构建体。该等构建体还具有N端标签(HIVNES)。进行了该N端标签和接头的测序(图34)。

[1953] 进行了脱靶的定量(图35)。测试了脱靶编辑(图36)。测试了用(GGS)2/Q507R靶向的内源基因(图37)。进行了(GGS)2/Q507R上的突变的eGFP筛选(图38和39)。

[1954] 产生(图40A)并测试(图40B)了具有为Cas13b截短的dCas13b的构建体。该等构建体还具有N端标签(NES/NLS)。产生了多重中靶/脱靶指导序列用于筛选(图41)。

[1955] 实施例7

[1956] 使用酵母筛选来筛选影响ADAR活性的ADAR中的突变。筛选进行了多轮。每轮筛选产生一组候选突变。然后在哺乳动物细胞中验证所述候选突变。表现最好的突变被加入到突变的最新版中并进行重新筛选。下表显示了在10轮中筛选出的突变。在第n轮中鉴定的突变体称为“RESCUE vn-1”。如本文所述,RESCUE是指将腺苷脱氨酶活性转化为胞苷脱氨酶活性的突变。

[1957] 表13

|        | RESCUE<br>轮次 | ADAR 突变   | 质粒      |
|--------|--------------|---|---------|
| [1958] | RESCUEv0     | E488Q   | pAB0048 |
|        | RESCUEv1     | E488Q, V351G  | pAB0359 |
|        | RESCUEv2     | E488Q, V351G, S486A   | pAB1188 |
|        | RESCUEv3     | E488Q, V351G, S486A, T375S  | pAB0642 |
|        | RESCUEv4     | E488Q, V351G, S486A, T375S, S370C,  | pAB1072 |
|        | RESCUEv5     | E488Q, V351G, S486A, T375S, S370C, P462A                                    | pAB1135 |
| [1959] | RESCUEv6     | E488Q, V351G, S486A, T375S, S370C, P462A, N597I                             | pAB1146 |
|        | RESCUEv7     | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I                      | pAB1194 |
|        | RESCUEv8     | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V               | pAB1220 |
|        | RESCUEv9     | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I        | pAB1327 |
|        | RESCUEv10    | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L | pAB1411 |

[1960] 对RESCUEv10进行了多轮验证(图42A-42E)。通过下一代测序(NGS)分析了RESCUEv10(图43)。鉴定出改善特异性的突变(图44)。测试了RESCUE对内源性靶向(C翻转和T翻转)的作用(图45)。

[1961] 使用RESCUE靶向β-连环蛋白。图46显示了使用RESCUE v6和v9 靶向β-连环蛋白。图47显示了新的β-连环蛋白分泌的Gluc/Cluc报道体。图48显示了RESCUEv10靶向β-连环蛋白的结果。

[1962] RESCUE也可用于靶向其他基因。图49显示了RESCUEv10靶向 ApoE4。

[1963] 实施例8

[1964] 本实施例显示了使用RESCUE碱基编辑β-连环蛋白以增加β-连环蛋白的稳定性,以改善HUVEC在营养缺陷型培养基中的增殖和存活。

[1965] HUVEC在营养丰富的培养基中生长。用含有RESCUE构建体的腺病毒转化细胞。该RESCUE靶向β-连环蛋白并产生S37A突变。转化的细胞以低汇合度传入营养缺陷型培养基中。使用细胞计数试剂盒测量细胞增殖和存活率。

[1966] 实施例9

[1967] 本实施例显示了在HepG2细胞中碱基编辑丝氨酸蛋白酶PCSK9。该碱基编辑通过在PCSK9上诱导患者衍生的突变来调节HepG2细胞中的低密度脂蛋白(LDL)胆固醇更新。

[1968] 使用各种转染试剂将GFP表达构建体转染到HepG2。选择产生最佳 GFP表达的最佳转染试剂用于转染RESCUE构建体。使用具有在5', 5、7、9、11处的靶位点的30bp指导序列转染RESCUE构建体。通过RESCUE 产生PCSK9中的一个或多个突变。示例性突变显示于图50。

[1969] 进行RT-PCR和测序以确定表现最佳的指导序列。胞质LDL被荧光标记,并通过细胞成像测量胞质LDL的细胞更新。使用ELISA和/或免疫沉淀监测PCSK9分泌。

[1970] 实施例10

[1971] 本实施例列出了与Cas13b-t有关的信息和数据。Cas13b-t1、Cas13b-t2 和Cas13b-t3的各自大小列于表14。

[1972] 表14

| 命名键(Naming Key) | 大小    |
|-----------------|-------|
| Cas13b-t1       | 804aa |
| Cas13b-t2       | 802aa |
| Cas13b-t3       | 775aa |

[1974] Cas13b-t1、Cas13b-t2和Cas13b-t3的氨基酸序列如下所示:

[1975] Cas13b-t1

[1976] mndkstwqlklhrivrwsflrrqrvgcdishhdfilvrrsgiknmefenikktsnkevysieqyege  
kkwcfaivl nra qtnleenpklf eqtltrfekimqdwfneetkkliyeeenk vkee iqiaaserlknlnryf  
shylhapdclifnrndt iriimekayeks rfeakkkq qedisiefpelfeedkitsagvffvsvffierrflnr  
lmgvvggfrktegeynitrqvf skyclkdsysvqaqhdavmfrdilgylsrvpteyqhikltrkrsqdlser  
ktdkfilfalkyledyglkdladytac farskikrenedtketdgnkhkfhrekpveihfdkekqdqfyikrnn  
vilkaqqkqgsnvfrmgvyelkylvllsllg kaeaiqridryisslkkqlpyldkisneeiqksinflprfvr  
srlgllqvddkrlktrleyvkakwtdkkegsrklel hrkgrdilryinercdrplskeynnilkfivnkdfag  
fyneleelkrtrrrldkniqlsghtlnalhervcdl vlqe lgslqsenlkeyiglipkeekevtfrekvdri

eqpvvykgflryeffkedkksfarlveeaiktkwsdfdiplgeeyyn ipsldrfdtrnkkyetlamdrclclmma  
rqqylrlneklaekaqhiywkkedgreviifkfqnpkeqkksfsirsildy tkmyvmdpflsrlweyfipke  
akeidyhkhyaafdkytnlqkegidailklegriierrkikpaknyiefqeimnrs gynndqqvalkrvrnall  
hynlnferehlkrfygvvkregiekkwsliv (SEQ ID NO:272)

[1977] Cas13b-t2

[1978] mqvenikkgsqgmysieqyegakkwcfaivlrnraqtnlqgnpkfletltrferirkedwfdqetkk  
liyakeqneve eeiqkaadeklrdlrnyfshyfhtpdcliftqndpvriimekayekarfeqakkeqedisief  
gelfeengritsagvvf fasffaerrflnrlmgvqgftrtegeykitrdvfstyclrdsysvktphdhdavmfrd  
ilgylsrvpseyqrikesqmr setqlserktdkfilfalnyledygledladytafartrikreqdentdcke  
qkphrkkprveihferaegdpfyikhn nvilrtqkkgaqtyifrmgvvelkylvllslgkgaevkridryvhs  
lnqlphiekksteeiegyvrflprfvrlshlg llgvddekkikarvdyvkakwlekkksrelqlhrkgrdilry  
inercerplnideynrilellvtkhldgfyreleek ktrridknivcnlsrhksvnlhekvedlvvqeleslg  
reelkeyvglipkeekevsfeektdrvvkqpviykgflrnef fresrksfarlveeavrekgevydvpplggye  
ivsltdfdkdnkrlyetlamdrlllmiaqyhslsnkelakraqqie wkkedgeeviiftlknpaqpeqscsvrf  
slrdytklyvmdaeflarlcedyflpkdeeqidyhrlytqgmnrlynlqreg ieaiielekktigpeqprppkny  
ipfseimdksayneddqalrrvrnallhhnlfnaradfkrfcgimkregiekrwsl av (SEQ ID NO:273)

[1979] Cas13b-t3

[1980] maqvskqtskkrelsideyqgarkwcfiafnkalvnrkdnglfvesllrhkyskhdydedtral  
ikcstqaanaka ealrnyfshyrhspgcltftaedelrtimerayeraifecrrreteeiiefpslfegdritta  
gvvffvsffverrvldr lygavsglkknegqykltrkalsmyclksrftkawdkrvllfrdilaqlgripaeay  
eyhgeqgdkkrandnegtnpk rhkdkfiefalhyleaqhseicfgrrhivreeagagdehkkhrtkgkvvdf  
kkdedqsyyisknnvivridknagprs yrmglnelkylvllslqgkddaiaklyryrqhvenildvkvtdkdn  
hvflprfvleqhgigrkafkqidgrvkvhrg vwekkkaatnemtlhekardilqyvnenctrsfnpgeynrllv  
clvgkdvenfqaglkrlqlaeridgrvysifaqtsti nemhqvvcdqilnrlcrigdqlkydyvglkkdeidyk  
qkvawfkehisirrgflrkkfwydskkgfaklveehlesggg qrdvldkkyhidaidgrfeganpalyetlard  
rlclmmaqyflgsvrkelgnkiwsnsdielpvegsvgneksivsv sdygklyvddaeflgriceyfmphkg  
kiryhtvyekgfrayndlqkkcveavlafeekvkvakkmsekegahyidfre ilaqtmckeaektavnkvrcaff  
hhhlkfvideglfsdvmkkygiekewkfpvk (SEQ ID NO:274)

[1981] Cas13b-t1、Cas13b-t2和Cas13b-t3的基因座显示于图54A-54C。基因座的序列显示如下：

[1982] Cas13b-t1基因座

[1983] agctgtcccgtgagcttattaacaagcattaccgctaaatccccgcggactggttggtttcagctt  
cgtgaatgcaa caacaaaaggcctgtcgaaagcacaatccccggtggtgcatagaaatccaggactttgcctt  
cgagggttttattggtt gccttctttgctgtggcgccatcccccaatcagaaagctgcgatagctttctgcgactg  
cctcggcatctttgggaccgga gcgtttgctcagaaatgccgtgatggtttaccgttaagctggtatccggcag  
cgaagatgtcagtcfaatcttcaaagc caaatgcacttgccagataaagcttaatgcttctgggaccaaattat  
cctttggcaggtgctcgatttcaggtatagcg gtatcatcgtgaacggccaggttcgtgggaatcccccttgca  
cttccgccattgccgcaaacagctcatccgattcggc gaagccgaccagctcgatataatattggccgtgcgcaa  
gataaacgcattactggttttgtatgcaaattgcatatccg gcaggttctcaacttcgggccttttttgcacgc

tgtaaaccgagaatgcgtttctggttctggccatatcaaagatatag agctccatcaccaggttttcatccgcct  
 ggcttacaatctctgggtggacaattttataaaaccagcgtcgatataaag gggggccttgccgttaatctttt  
 cgtaaagattttcggtggtgtagacttcaatttctgaaagcgttttgaatccgtaag gcagaagaaaagtcaggt  
 ctttcttttggtttggcatctgctttatgaataccccaacggcgataagtaagagaatcgct aataagcagatgc  
 ctataacagattcgagacgttttgcccggttggtaccgaaccataaccaactccagtaatgacaa attacttg  
 actttataaccgggctggattataatttttggcggtgttgctgtcaaccccaaatgctacaggtgaaaaagg cga  
 agatagatttctaacgaggttgacaaaagcaggtcagggcgtgttataataggttgctaaagtaaaaaggagactgaa  
 atgattgaatatgcacaatatttggggtttggacgccgggcccccttgaaattgctgttattgcgattgtcgctc  
 ttct gatattcggcagacggctgctgaaatcgcccgcaacgtaggcaagacctgactgaattcaagaaggggc  
 ttcacgagg ccaaggagaccaaggacgaattgggtggatgatgtccgggaagtcaaggatgatgtggtaagagagg  
 cgaaggatgccgcc gggtgaatgaagaggatacaatgggctctgattgattattgataaaggggaactaatcac  
 tgagaacaattgtcaatca ttaatcaacaatcaatattgaagatccgcctgtggcggaatcaatttttaagatgg  
 gcgatacaaagaagaaagaggacc tccttgattccactatgagctctgggcgaccacctgaggaattgccggatgc  
 ggctgattcgccgcctgggtgggctggcg ttagctcttattatctgtctgatcttcggcaagctgctgatatacat  
 ttattcaaaaaccttacgttctgtgatgggtga agaggctactctgaagacgcttgccccggcccaagggatta  
 acagctacgtaaaaatagccttggctctcaggcttgatat tctcatcgccctgggtcttctaccagttatggatgt  
 tcgtggctgcaggactctatcctaataaaaaagatatgtgtat gtagcagtaccttttccggtggtattatttg  
 ttgccggagctttgttttcatctttgtagtggcagaagtgtctcttgc tttcttaataaaggtcgacaggtggc  
 tcggactggaaccgcactggactttcccgaagtatgtgacctttgtaaccacc tgatgctggtatttgggtgtg  
 cgtttcagaccccgatagctattttcttttgaacaagacaggtctggtttcagtccag gcgttacggcggtcaa  
 gaaaatatgtactgctacttatcgttgtagtagcagctatggcgactccgcctgatgtggttcc tcaagtaacac  
 tggcgataaccgttgatgtgctgtttgaattaggcatactgctgagttactttgcagaactaaaaaga gaaagt  
 cgaaaaacaaccagtgataagccgacaatccccagctttcccagtaccgactacttgtttcttccggcctggt t  
 tttatttctgcaatcgagcgactaagaaatcttcaaaggcgttaaactccttccataccgtggcacagttaatggt  
 ttt ggctttgttatctattacggtgtatccatagtcggtaaccgcaatgccgagttttcgggctcatttttagac  
 atttgcat ctatgccgccggcagcgtgaaggtttttcggagctaattgagtattcagcataaatgttgaacgg  
 ttttgccaatgcg ggtactatgatgttgatgctaacgttgataaatacaaatgtgatggctccctcccatagggcc  
 tgtcggcctggactatat cgcaggagccgtcagggcagccgggaaccaggcagacgtagttgatttatgtcttgc  
 tgatgacctgcaagactctcc agggctatttctgctacgcacagccccgcaattgggtgggggtctcttttcgcaa  
 tgtggacgattcttctggccaagcgcc cggtggttctgctccccgacctggctgacactatccgtacgatacgaag  
 tatgacggatgcaccaattgtagttggcggcgt tggcttttccatttttccgagcgaatcgctcgaatataccgg  
 cgctgactttgggattcggggcgacggagagcaggcaa tagtttctacttcttaatcagctgcagcggccggaacg  
 gcttgaacgcatagatgggttagtccggcggcgacggagtt attcacagcaaccgaccagcgtggcctgcacc  
 gctttctttgcgaccgaacgtgatgctgattgataacctcgcttactt caaaaaaggagggcagtggtgtgga  
 gaccaaaccgggctgtaaccgccgatgcctatattgtgccgacctggctta agggctcggcagtcaggccgag  
 ggccccgtcggaggtcgccgatgaggtccagctcttaataaggcaagggaatagaagta ttgcatttgtgcgactc  
 tgagttcaacatctctcaaagccacgcctatgcggtctgcgaagagttcagccgtcgtcatt tgcgaaaaaggt  
 gcgctggtacacatatatggcggtggtgccattcgatgccgagcttgccggggctatgagcagagcgg gctgtgt  
 cggtatcgactttaccggcgactctgcgtgccccatcaattctaaagacctatgccagcggcatcataaagaa ga

ccttgccctcggcgggtgcgctttgtgccgtgctaacggcataacggttatgatagacctgctgtttggcggccccgggt  
ga aacgccggaacggctgcagagacaatagatttcattaagcaaattgacccggattgctgcaggggctccgctc  
ggtataa gaatctaccccggcaccgaaatggcccgaatagtggaacgaaggcccaccggaacgaaccggaac  
gttcaccgaaag tacgaggggctgtggattttcttcaaccaacttactatatactgaagccctcggtagcag  
ccggccgggcttatcaa ggatttgatttcggcagatgaaagattctttgagccgatgccggaatagccccggag  
gctctaaaaagtagccagtcca ccgaccacaattacaatgataataccgaactttagaagcaatcagcaaaggt  
gcacgcggggcatattgggatatactg cgcaagcttcgctgcgactaagcagcttatggtagtagatgattccc  
cctgcgggagattggcccgaatcctgaggaat ttgtagaagcggatgcaatgttgatTTTTGGGtAAAAcgg  
gggcagggggatttggccccggtttgaggattccga gaagcccaccgtagggatctccgctcccttagggata  
aattcgcttcgagtttgaaattggccccggtttgaggattc cgagaagctcaccgtagtgatctccgcttcgc  
ttcggctttgtttgggtttgtttccccgcgctctgcgaagtggttca ttttcataatcctttataacataaagt  
ttacgttcattttgggctttcggcaaattgggtttgaattgggtttgtttt ttgactgcgaaatcatcttttt  
ttctgtaaaccctttgttataagagagtttacattcattttgggcatttagtaaattgg gtttgattggctttgaat  
tgggtttgttttcaccaagtgtccaattggatttattttcataatcctttgtattatatgga tttacgttcattt  
gagcatccagaaaattggctttgtttgcataaaaagggtgattttgtagaggactctttacagttg tagagggc  
aagttagttaagagtgcgctaaagtgcctaaagtgaactaaagttggattctcgattctcgatagcgtata gcg  
tatttcacggttattcaccattcattaaggaataaatttgatttaggctgctggccccctccggcgattagtaaattg  
ttctcggcggcaaaaacaacgcgctctataattgggcgaacatgcacgtttgagtcgaaaattgggtgctttcttga  
cagg ataaacaggagtaactcgttgtagaaaaggagtaaaatttttttcaattttccgatttttaggttccaac  
tacctgcac ttttgattgaaaaatcacaatgtcttgcctattttaacgcagtttttcgctcgaacgtcagcgaa  
ctaggaataatagggc gatttctgggggaaaacaataaaaaatgcacaaaagtgacaaaaaacggccaaaaaagt  
gctttttttggctgccttt acccgtgagatgatttaccacaaccttctctgctattcctatgcaagtttgctca  
gggctggtgtgaatactataaaaa tttgtgctgtaatcactccacaaatcggaggcttcttcagcgtggaattc  
tggaggccaaaatgaaatacgtgtaate accccacaaatcggaggcttcttcagcttactacctctcaaatcg  
cccaactatacgtgtaatcaccacaaatcgg aggcttcttcagctcgaagtcccgtccacgcacaaagttt  
gagctgtaatcaccacaaatcggaggcttcttcagca tgagcttttggttgctggatagccagctgtaat  
caccacaaatcggaggcttcttcagcacaacggttcaaca aggtcgaagaactagctgtaatcaccacaa  
aatcggaggcttcttcagcttctgcggagtctttcgcgggtgttcaaat gctgtaatcaccacaaatcggagg  
cttcttcagcctatcctttataacatttttctatataagatttacaatacaaaa acccagcaaaaactgacttc  
ttcttttgaaatcatgccgtattataacacttttttactatcaaaagaccactttttt ctattccttctctttt  
cacgacccatagaatctcttcagatgttccctctcaaaattgagattatagtgcaaaaagcga tttcgcacccg  
ctttaaagcaacctgttgatcattattataaccgcttctattcattatctcctgaaattctatataatt ttttgc  
tggtttaatctttcttctgataatecttcttcaagcttttagtattgcatcgattcctcttttttgaaggt t  
tgtatattttgtcgaacgcccttgcatagtgttatggtagtctatctttttgcttcttttgggataaaaatatcc  
caa agtctgcttaaaaaatcaggatcgtccattacatacatctttgtataatccaagatcgaaaagcgtatcgaa  
aaactctt cttttgctcttttggattttggaatttgaaaataatcacttctctgccatcttcttcttccaatag  
atatgctgtgect tttctgcaagttttcgttcaatctgagataatattgccttgccatcataaggcaaagctctg  
tccattgccagtgttca tatagcttctgtttgttctgtcaaatcgatcaagagatgggatgttataatactct  
tcaccaagaggaatatcaaaatc cgaccactttgtcttaattgcttcttcaacaagtctggcaaaactctttttg

tcttctttgaagaattcgtatctcaaaa atcccttataaacaaccggctgttccaaaatcctatctaccttttct  
 ctaaaagttacctctttttcttcttttaggtatc agcccaatatattccttgagattctccgattgcaaactgccc  
 agttcttgtagaaccaaatacacatccctttcatgaag tgcattgagcgttgatgcccggaaagcttctggata  
 atatTTTTgtctaatcgctcgtggtcctttcagttcttcaagtt cattataaaaatccggcgaagtctttgttact  
 ataaactttaaaatattattatattccttctgctaagtggcctatcg catcgctcgttgatatactgagtata  
 tcccttcccttttcgatgtagttcaagcttctcgatccctcttttttatccgt ccacttggccttaacatattcc  
 aatcgagcttttaaccttttctcatcatcaacctgtaaaagacctagctttgaacgta cgaatcttggaaaggaag  
 tttatagatttttgaatctcctcattacttattttatctaaataaggcaactgcttctttaa ctactaatatag  
 cggtaaatcttttgaattgctctctcggcttttcccaatagactcaaaagaacaagatatttaagttc ataaact  
 cccatcctgaatacgttggactgtccacctttttttgagccttcagaataacattatttctgtttaataataaa at  
 tggctcttgcctctctttgtcaaaatgaatctcgactaccggcttttccctgtgaaatttgtgtttgttaccatctg  
 tc tctttcgtatcttctgttctccttttaattttacttcttgcaaaacatgctgtgtagctgccaatccttaa  
 gtccata atcctcaagatatttctgcaataatgaacttctcctcttcttcttctcctcaactgatcctggc  
 ttctcttttcgag ttagtttgatgctgatatactcagtggaactcgggacaggtatccgagaatatccctga  
 acataactgcatcatgg tctgctcctgaaccgaataactatccttaagacaatatttggaaaaacttgcctgtg  
 ttatattatattcaccctctgt ttttctaaaccttggacatatcccattaagcgatttaaaaatcttctttcaa  
 taaaaaatgagacaaagaatactacac ctgctgatgttatcttattcttcttcttcaaaactctgaaattcaa  
 tcgaaatatcttcttgttgtttttcttctcct tcaaaacggctcttttctgatgctttttccataattatcctta  
 tgggtgtcatttctgattgaatatcaggcagtcaggcgc gtgaagataatgtgagaaataattccttaattcttta  
 gtctttcactggccgctatttgaatttcttcttttactttgt tttcctcttcttttccataaatcagtttttttg  
 tttcctcattaaaccaatcctgtttcatgatttttcaaatcttgtga agtgtttctcaataactttggatttt  
 cctcctaaatttgtttgtgctctattaagaactattgcaaaacaccactttt ttctcctcatattgctcgatag  
 aatacacttctttattgcttgtttttttatattttcaaacctcatatttttaatcc cegatcttcttactaata  
 taaagtcaaagtggtgcaaatatcgaccctaccctctgcctgcgcaggaaagaccaacgg acgattcgatgca  
 gtttgagctgccaggtgcttttatcattcaagggtaaaatagcagaaaagccttaatgtgtcaagg ggatttta  
 gatttactatttccaatttacgattttggattgagattgcttccggcctaaaagacaggcctcgcaatgacce cct  
 tagagttgaaagcactctaaacaagggggcaggcgggggaatatcgaatatcgaatctgaatgtccaatgtcgaag  
 tgaactgcgcgggaatgacaggtcggcagatttatttaattctgtggccagatccctccgcttccacactgac  
 gtgg acttccggtcgggatgacactgggggtgtgccattgctccgctcg (SEQ ID NO:275)

[1984] Cas13b-t2基因座

[1985] agccgagtcgatggttagctaaagtgaaacgacaagcgtggttatacggagataagttgatgagcactggt  
 tatccatgagga cgaagtagattcgatggattggttatatggaattagatcgaaacgtatacctatgaagctcac  
 agtcaaataccaatcgg gatagaaatgcggcgcgcgcaaccttaggcaaaagccttggctgtttcagcgtttccgc  
 tttacgtgccggttttagccttc accatagtcaccctttccgcaagcctccctgcgatcccggacagtcggatttc  
 ccaaatccggttctggtctcggcctta tttgtcattttctggataaaggccttctgtacaatttgagacttaag  
 tgetagctcacttgcacccataattgtacag tttaccagtatcctcgttccgagagtcctatggcattcgttcca  
 gttcgggtgcttgatgcacatgccttactcagaacca ccgagtaccagagcccccttgcaggcgtgggcgcta  
 cccactacctggatgactttgaaagtcacctcagaagacatt actcttcttcatagctcatacggactcatgac  
 ccagaccaaatccctccaacgtcttggttttcccttgtacgttagg tctttgcaggttctcggcagtcctctgct

gggaaacggcccttcccacattatctctgcaatccttgtataggtgcaagg acccttacccegcagcgtccctt  
 cgggtgcccttgcccgtttcttcccgaaggactgcggtctcacctcaggatttaaagg ttcgacacgccaattatc  
 cgctcgcaatgcaacttcaacaacggggcaaatttcggggctgcagtcattccataacgttca agctcctatacct  
 gctatgccctgcggttgcaaccaccactgagcatatatgagctcagggcagccgggcccgtttacacc acgcatcg  
 cccgatggttaccattccgagatgtggcatcgctacgtgcctgaatcgggcaactggcacgacgggactt tca  
 cccgctggattgcagccttgtcggtgctccaaatccctgttgccacaaaaatcttctttgaggcatccacgttacg  
 acgtgtcggccacgcttcgtagcttatctgaacagcctttcgacttcacgatgattataatggacatcgtgaatc  
 tagc tatgtcatggtcaatcgtagtgtggccaggggcccggccaatcgagtatacttgataaagtgttcatgaagc  
 tgtattctt taatctcccgaagagtatgcttgaacgttttaagaaatgacagatagtggtgaagtgttctgaaaac  
 gggcccgggagggc aagacgtgagtacaggtacagtgaaatggtttaatgcaagaaggggatacggttttattgt  
 ccccgatgatggcggagat gatttatttgttccagcttcggacattaacacagaggactatgcatcgcgagatta  
 ttaaggtcggcaacgacatggttg cggaccatatcaatcataagggcttggataatcgcaaggccaatttgcgag  
 cggcgacgattgctcagaatgctgtgaac cggcagcgcaaaagaagcggatttatgggctagtgtggaataagc  
 agatgaggaaatggcgtgttaataatcagtcacga gggcacgtgcaggcatatcggtacttcgatgatgaggttg  
 aagcggcgaaggcgcacgaccgggcagcgcgaaaaatatac acggagagttcgcgagtttgaatttcacgcgttaa  
 gccacatcacagcaggtccgactatggcggacgcagcaatctta agcatatttggctgcgcaatgtgttgcgcgg  
 gttcctgcttggggcgagctatcgaggtgtaatcaccacacaaatcggg ggcttctccagcgcgctaaaagtga  
 taagaattttagatgcgccgtaatcaccctcaaatcgggggcttctccagcgc tgaccgaattgataaaaccaa  
 gagagcgtgtaatcaccacacaaatcgggggcttctccagctgtacgacaaatcataa cagaatatttgaagct  
 gcaatcaccacacaaatcgggggcttctccagcaaaatgagacaccacgcttgacgtcactgtg ctgtaatcacc  
 ccacaaatcgggggcttctccagcttcgagctatatctggctcggctctgatttggctgtaatcaccaca caaatc  
 gggggcttctccagcactggcttagcaagttcctttgggcgtttcgtgtaatcaccacacaaatcgggggctt c  
 tccagctaatcgaagatgagaccgaagactatcactgctgtaatcaccacacaaatcgggggcttctccagctgat  
 tgg gaaagcactccttacgcacgagagctgtaatcaccacacaaatcgggggcttctccagcacatcctcgataa  
 tacgttat ctcgattggattacaacagaaaaatcactgaaaaataccagggtttttggtgcaatgcgcacacatt  
 agaacctgttttc aactgctaaagaccagcgtttttcaatcccttcccttttcataattccacagaacctctta  
 aaatctgccctggcaaaa ttaaggttatgatgcaaaagcgcgttttcgcacacgtcggagagctttctggctcatct  
 tcattgtaggcgcttttatccat tatctcgctaaatgggatgtagttctttggaggtcttggctgctctggaccg  
 atagtctttttctcaagctcgagtatgg ctcaattccttccctttgcaggtttgtgtatctgttcatccttgc  
 gtataaagcctatggtagtcgatttgttcttcg tcttttgcaaaaaataatcgaaagtcgggcaaaaaactcc  
 gcgtcgtccatcacatagagtttcgtataatccctcag cgagaaccgtaccgaacaactctgctccggctgtgcc  
 ggattcttcaaggtgaaaaataattacttcttcgccatcctctt tcttccactcgatttctgtgacctcttggca  
 agctctttgttaagactaagatgatattgccttgcgatcatcagcaaa agcctgtccattgccaggttttcatac  
 agtctcttattgtctttatcaaacgtatcaagtacacgatttctgtaatactc cccccagagggaacatcataa  
 acctctcctttttccctcaccgcttcttcaacaagcctcgcaaaactctttctgcttt ctctgaagaattcattc  
 ctcaaaaaatcctttataaataaccggctgtttcacaacctgtcctctttttcttcaaatgac acctctttttct  
 tctttgggtatcagccaacatattccttcagttcttctctgccttaggctttcaagttcttgcacgac caaatcg  
 cacaccttttctgtgcagcgcattgacgctttttatgcctggaagattgcacacgatgttcttatctatccgctc tg  
 gtcttcttcaattcttcaagctctcggtaaaaccgctcgaggtgcttagtgacaaaagctccaaaatacggttat

at tcategatgttcagcggcctctcacaccgctcattgatataacctcagaatatcccgtccttttcgatggagct  
 gaagctc cctcgacttttctttttctccaaccacttggccttaacataatcaactcgcgccttgatctttttt  
 catcatcaacc ctaagagaccagatgggaacgcacaaacctcggaagaaatctcacgtatccttcaatctctt  
 ccgtgcttttctctct atgtgaggcaactggttgcgcaagctatgaacataacctgtcgattcttttgactgcct  
 ctgctccttttcttaataagct cagtagaacaagatatttaagctcgtagacgcccacctgaatatataggttt  
 gggcgccttttctctgagttcgcagaa tgacgttattgtgtttgatataaaatgggtctccttcggctctctcaa  
 aatgaatctcgactctcggcttcttctgtga gtttctgctccttgccatctgtattttcgtcctgctcccgt  
 taatcctcgttctggcaaaacatgctgtgtagtctgc caaatcctccagcccgaatcctcaagatagttcagcg  
 caaacaatatgaacttgccttctttctgcttaactggg tttcgttcgcatttgcgattctttgatacgt  
 gatacgaactcactgggaactcgtgacaaataccccagaatatcccg aacatgaccgcacatgatccggcgct  
 taaccgaataactgtccctaagacaatatgtcgaaaaaacgtcccgcgttat tttatattccccctctgtacgcg  
 taaaccctgaacatatcccattaaccgatttaggaaccttctctcagcaaaaaatg acgcgaaaaataccacac  
 ctgctgatgttatcctgccgttctcttgaacaactccccaaattcaatcgaaaatatcttc tgttcttttttg  
 cctgttcaaaacgcgccttttctgacgtttttccataattatcctgaccgggtcattttgggtgaa taccaggc  
 agtcaggcgtatgaaaatagtcgagaaataatcctcaaatctctaagcttttcatcageccgtttttgaa ttt  
 cctcctctacttctgttttcttctgttttgcataagatcagtttttctgttctggtcaaccaatcttctttctg  
 atcctttcgaatcgtgtcagcgtttctcgaacaacttcggattccccctgcaaatgttttgcgcctattaagca  
 ctat cgcaaaacaccacttcttgccccctcatattgctcgcgatagaatacatcttggctgcttcttcttga  
 tattttcaa cctgcataatctcagactctcccaattgttgtttttcccatttttgttgaagtccccgaatgtcag  
 tctattgggccagc tgagtcaaccacaaggcacaatgtacatacagctctcgagtcatttcgagaagactttccg  
 ctcgcccgataagataagc tttgagtatctcacgggtggaccgccagagataatccacatctcgtatccgggt  
 aagctatccggcataaattcgtgc ttagtgaatcgtgtttcgtgttgatacggctcccggctgcattcacttttc  
 acggcagagaatatcgcaaaaataaggcaa cagtcaaaggaaaaagggtaaaaatggtgaaatagatgagcagca  
 gtgaattgttggcaagcaagccgcaaatgaat ccttcggccacgctc (SEQ ID NO:276)

[1986] Cas13b-t3基因座

[1987] tatccaaaatgtggtttgaattcaagaatcaacgctttattccttaaaaagggcggtgcgatggaaa  
 aagaaccagaaa catccgtgcaatcggcgtcgggacacaatatggatatcccgattgactggtcggtaacctcac  
 gctatttcgaagatgaa gatacgtgatgcaggtggtgggatatttgcgtgaagactctccgcagaccgtccggg  
 accttgccaaggctatacagac gcaaatatcccaggatgttcaattgcacgctcacagcctgaaggagcctcgg  
 ctcttatcggggccgaacatctgccc aaagagcctggcggcttgaatacgcgccagagaaaaacacggcgg  
 cgtttgaggcgtggttgacgagacaaaggcc gagttcgacaagctgatgtcgttctttaccgcgccgattgga  
 ttgaagcagcaaaagaacgccactgcaacaggcaaca ggccgagcaggtatgaaacatcttttgaaaagaaggc  
 gatggaatgagtggtggttctcattttgatcattgatgat gacaggatggttacagacaagttggagaagatc  
 agcggcgccaaggctgcaaagaaaaggttcagcctggcaggcgtttt ctcaaaggcgctgaagcccttttattt  
 gcaggcgtgctaccgcttgcacgggcaggggacagaaccgcaatcaggat taccatcagtttcttattccat  
 taacctcgtttttctctcgttctttttcttcttcttctggttttctcagcgttgggc tgtctttttgcccgtttt  
 gtatagttgtcgcgtaaatgtcaatgagtgcggcttttagtttttcgggccagttgcggtt ttcaaaagcgcac  
 acgagcggatcgcgccttgtttcatccagttatgaagccggccctgcacatcttctgatttcgctcc tttgctct  
 gcggaatcgataagggtgttgaggcagtcggggtcatttctgagatcatagaactcctcggccgcgcgatat ctg

aacatcttgacgcgctgtgcgggcgaacttattcgttggagcggcctcgaccatcgcttcatggtaaggccctcgtt  
gttgttttcggtaccagaatctgccgtcgcccacggattgaagatatagccgaagcgttatctctggacgcaccgc  
atcg ggacagcgtctccgccggctttcatgtctatctgcgtaaagaccacgctcgcgtccggattgcttttcgct  
ttcaacagc cccaggaaagaggaaccgtcaagccccctgggtatgccagaccgaccgcttcgagcaccgtcggg  
aagaagtcgatccc tgagataaagtgcgccttatcgacggcgcctgcttttaccatttgccgccaacgaacgatc  
cacggcgtccgcgtgctgg caagataggcgttgcatTTTTGCAAACGGTATGGCGATGCCGTTGTCTGGAGAGGAAC  
atcacaagcgtattctcctcgaag cccgactccttcagggcctgcaacgcttGCCGAAGGTATCGTCGAGTCGG  
cggacggagttgagatagcagctcagttc ctgccgaacgcccggcaggtcgagacaaaaccgggaaccgcaacc  
tcategggcttatacgtctttgaaggttcctttg ccccttgattggcttgccgccgatatgatacgggagatgc  
ggatcgtgcgagttgaccataaagtagaagggttattc tcgcggcgacacctcgccagaaactccttgacgtaa  
actgtaatagagttccggatcgggcggcggcagttccttct ggtcatgcacaaaatcccatttgtaatecgc  
atggggcgttgagtgccccaccttgccgagaataccggtaagatagccg gcatccctcagcgtctgcatgacagt  
catcacaaggcggcgcceaatcctgcgcccttagaaaatcacgacgattcat cattgtccccactaatcctta  
ttgttcttctcaagataccccgacaatttctgcatttgccgatacaggccgccgggaca tatcagtatagccgca  
aaccttgaaaatatcaacctcccggaatataacgtcgacttccaaccagatcgccaatccaga ataagaaaaca  
aagcaaacgcttcaaattcgtttaaccccagggttcgctgaggttcgtaaacaccatctcgatgtac atcggg  
attcaaattcgttgagccccageccttcttggtgctcttgcttcggcaagaaacgctgtaatcaccacaaaatc  
ggggctgctccagcatcgccaagacgggcaatgccgctttgaggctgtaatcaccacaaaatcgggggctgctcc  
agc tgatttcgagtttcgatgcttcggacagggtgtaatcaccacaaaatcgggggctgctccagcactcct  
tatggaga aggagcttatcgtgtcgtgtaatcaccacaaaatcgggggctgctccagcttattccttccatca  
tcccgacagcagt gggctgtaatcaccacaaaatcgggggctgctccagcccacttctgtaaccattttactcg  
caaacgcttataacgaaa acactttcaaaaaccataccaacgctcctcatttaacaggaaacttccactcctttt  
caattccatatttcttcataaca tcaactaacaacccaaattcatctatcacaactttaaatgatgatggaaaa  
acgctctacgcaccttattcacggcgggt cttctcgcctctttacacattgtttgtgccagtatctcacgaaat  
caatataatgcgcccccttctctcgtcctatct ttttggtttgacaaccttctctcaaacgccagcaccgcct  
cgacacatttctctgcagatcattatagccctaaac ccttttctgtaaacgtatgataccgtatcttccctt  
ttctgtcggcataaagtactcacatatccgcccagaactc agcgtcatccaacacatataacttgccgtaat  
cactcactgagaagacgatgctttttcgttaccactgagccctcca cgggcaactcgatgctatcattcgacc  
acacaattttattacccaattccttgcgtagactcccaggaagtattgcgcc atcatcagacacaaacggctctc  
gcgccagcgtttcatacaaggtggttagcaccctcgaatcgcccaatcgcatcaat atgataatactttttat  
ccagcccacgctccctctgtccgccgcttttccaaatgctcttccacaagcttcgcaate ctttcttgctgt  
cataccagaacttcttgcgcaagaaacccctgcggatagaaatatgctccttgaacctgcaaccttc tcttgt  
aatctatttcatccttcttccaagccccacataatcgttagagcttctgatcggcattcggcaaagtctgtt ga  
gaatctgatcacacaccacctgatgcatctcgttttattgtggaggtctgcgcaaaaattgaatataccgcccgtc  
ga ttcgctcggccagttgcaggcgttccagtcgccctgaaaattctcaacatccttgccaaccagacacaccag  
cagccgg ttgtactcggcgggattgaaagacctcgtgcaattttcatttacgtattgaagaatgtcccgcgctt  
ctcgtgaagtgt catctcgttggtcggccttcttcttttccacacccctcgaacatgctttactctgcgctc  
tattctttgcttaaaag ctttctgccaatcccatgttctcagcacaacatcgggcaggaagacgtgattatc  
cttatctgtgaccttcaactaca tccagaatgttctccacatgctgccgataacctgtacagttttgcaatcgcatc

gtcgccctttccctgaaggctaagcaa tacaaggtatttcaattcggttaagccccatgcgataactccgaggccc  
ggcattcttatcaatcctgacgataacattgt tcttactgatatagtatgactgatcttctgtcttttttgaaaa  
gtcgacaactaccttgcccttggctctgtgctttttg tgttcgctgcctgccccggcctcctccctgacaatgtg  
tcgccgcccgaagcatatctcactgtgttgctgctccagata atgcagtgcaaactcgatgaacttgtctttatg  
gcgtttcggattcgtccctcattgtcgtttgctcttttcttgctgc cctgctctccgtggtagtattcatacgc  
ctccgcagggatgctccaagctgcgcgagtatatccctgaaaagcagcacg cgtttgcaccagccttctgtgaa  
acgactgtctttcaggcaatacatcgaaagcgccttccgagtcagcttgtactgtcc ttcgtttttcttaagccc  
acttaccgcaccgtacaaacgatccagcaccgcccgttcaacaaagaacgaaacgaaaaaca caacccccgccgt  
agtgatccggctgccttcgaacaggtgggaaactcgatgatcacttcagtttcgctctcctgcat tcaaagat  
cgcccgtcatacgcctttccatgattgtccgcaactcatcttctgctgtaaatgtcagacacccgggga atg  
tcgatagtgaggagaaatagtttcttaacgcctcggccttcgattggccgcttgtgtgctacacttgatcaaagcgc  
gtgtatcctcatcgtaccagtcgtgctttgaatacttttcatggcgtaacagcgcactcgacaaaaagcccgtcgtt  
ctta tctcgattcacaagagccttgttgaaggcaatcgtaaaacaccatttccgagcacttgatattcatcgat  
agacaactc tctcttttctgaagtctgctttgacacttgcgccattgagcacctcccattccagatttttagtgcg  
atctttacctcatg cctccacaacactcccagcgcacaaacggttgagcaaaagcaaaatagccgcaggcgggctcc  
gtcgaatccgtaatcctaa tttctaacttccaatcatctaaaccgcccgaaccgatttgtcaacaaaaacca  
catcaatccgcagatggccgcaga taaccgcagatattgcaactaatccaccaacccaaaacctctgttccatc  
tgcgccctctgcgaaatctgcggacagct tttttttctgtgcccttcatgtcttctgtggtgaatttcatttaaca  
tttgacaaatatcaaacggcatggtataatgcgt tgcgtatttaaggacaaagcaacacccaaaacagggggagt  
aaaaaacctgttccatccaaaaagaatcgcaggccgcag gcctgccacctatgattaacctcggcttttcagcca  
aggatgctccccacacccaaacaagcgaacgaaccgtgcgcca agctaagctggtgcaattcagcaggtgtaat  
cctgcccggtcaaaggttagccgcccggcgggaatgaacatgtacgtat aaggaggcaacaaat (SEQ ID NO:  
277)

[1988] 在Cas13b-t基因座上的更详细的序列和特征显示在图55A-55C中。

[1989] Cas13b-t1、Cas13b-t2和Cas13b-t3与其他Cas13b直向同源物的比对显示于图56中。在图56中，序列#6是Cas13b-t1，序列#1是Cas13b-t2，和序列#2是Cas13b-t3。其他序列是Cas13b直向同源物。

[1990] Cas 13b-t与来自Alistipes sp.ZOR0009的Cas13b相似 (Cas13b4, NCBI accession WP\_047447904)。使用由GenScript合成到用于哺乳动物表达的pcDNA3.1 (+) 主链中的人密码子优化蛋白 (通过GeneArt算法进行密码子优化)。通过两个指导RNA与非靶向对照在HEK293FT中测试了高斯荧光素酶的敲低。RanCas13b (B6) 用作阳性对照。将荧光素酶值相对于非靶向对照进行标准化，如果未敲低，值为~1。在此测量中发现了一些噪声，因此某些值略高于1，但在可接受的范围内可归因于噪声。Cas13b-t1、Cas13b-t2和Cas13b-t3在哺乳动物细胞中的Gluc敲低分别显示在图51-53。表15、16和17分别列出了Cas13b-t1、Cas13b-t2和Cas13b-t3 的指导RNA键。

[1991] 表15. 指导RNA键-Cas13b-t1

| 键 # | 正向重复序列  | DR 是间隔序列的 5'或 3'? |
|-----|---|-------------------|
| 1   | GCTGTAATCACCCACAAATCGGAGGCTTCTTCAGC<br>(SEQ ID NO:278)  | 3'                |
| 2   | GCTGTAATCACTCCACAAATCGGAGGCTTCTTCAGC<br>(SEQ ID NO:279) | 3'                |
| 3   | GCTGAAGAAGCCTCCGATTTGTGGGGTGATTACAGC<br>(SEQ ID NO:280) | 3'                |
| 4   | GCTGAAGAAGCCTCCGATTTGTGGAGTGATTACAGC<br>(SEQ ID NO:281) | 3'                |
| 5   | GCTGTAATCACCCACAAATCGGAGGCTTCTTCAGC<br>(SEQ ID NO:282)  | 5'                |
| 6   | GCTGTAATCACTCCACAAATCGGAGGCTTCTTCAGC<br>(SEQ ID NO:283) | 5'                |
| 7   | GCTGAAGAAGCCTCCGATTTGTGGGGTGATTACAGC<br>(SEQ ID NO:284) | 5'                |
| 8   | GCTGAAGAAGCCTCCGATTTGTGGAGTGATTACAGC<br>(SEQ ID NO:285) | 5'                |

[1993] 表16. 指导RNA键-Cas13b-t2

| 键 # | 正向重复序列                              | DR 是间隔序列的 5'或 3'? |
|-----|-------------------------------------|-------------------|
| 1   | GCTGTAATCACCCACAAATCGGGGGCTTCTCCAGC | 3'                |

|          |   |    |
|----------|---|----|
|          | (SEQ ID NO:286)   |    |
| 2        | GCTGCAATCACCCACAAATCGGGGGCTTCTCCAGC<br>(SEQ ID NO:287)  | 3' |
| 3        | GCCGTAATCACCCCTCAAATCGGGGGCTTCTCCAGC<br>(SEQ ID NO:288) | 3' |
| 4        | GGTGAATCACCCACAAATCGGGGGCTTCTCCAGC<br>(SEQ ID NO:289)   | 3' |
| 5        | GCTGGAGAAGCCCCGATTTGTGGGGTGATTACAGC<br>(SEQ ID NO:290)  | 3' |
| 6        | GCTGGAGAAGCCCCGATTTGTGGGGTGATTGCAGC<br>(SEQ ID NO:291)  | 3' |
| 7        | GCTGGAGAAGCCCCGATTTGAGGGGTGATTACGGC<br>(SEQ ID NO:292)  | 3' |
| 8        | GCTGGAGAAGCCCCGATTTGTGGGGTGATTACACC<br>(SEQ ID NO:293)  | 3' |
| [1995] 9 | GCTGTAATCACCCACAAATCGGGGGCTTCTCCAGC<br>(SEQ ID NO:294)  | 5' |
| 10       | GCTGCAATCACCCACAAATCGGGGGCTTCTCCAGC<br>(SEQ ID NO:295)  | 5' |
| 11       | GCCGTAATCACCCCTCAAATCGGGGGCTTCTCCAGC<br>(SEQ ID NO:296) | 5' |
| 12       | GGTGAATCACCCACAAATCGGGGGCTTCTCCAGC<br>(SEQ ID NO:297)   | 5' |
| 13       | GCTGGAGAAGCCCCGATTTGTGGGGTGATTACAGC<br>(SEQ ID NO:298)  | 5' |
| 14       | GCTGGAGAAGCCCCGATTTGTGGGGTGATTGCAGC<br>(SEQ ID NO:299)  | 5' |
| 15       | GCTGGAGAAGCCCCGATTTGAGGGGTGATTACGGC<br>(SEQ ID NO:300)  | 5' |
| 16       | GCTGGAGAAGCCCCGATTTGTGGGGTGATTACACC<br>(SEQ ID NO:301)  | 5' |

[1996] 表17. 指导RNA键-Cas13b-t3

| 键 #      | 正向重复序列   | DR 是间隔序列的 5'或 3'? |
|----------|--|-------------------|
| 1        | GCTGTAATCACCCACAAATCGGGGGCTGCTCCAGC<br>(SEQ ID NO:302) | 3'                |
| [1997] 2 | GCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGC<br>(SEQ ID NO:303) | 3'                |
| 3        | GCTGTAATCACCCACAAATCGGGGGCTGCTCCAGC<br>(SEQ ID NO:304) | 5'                |
| 4        | GCTGGAGCAGCCCCGATTTGTGGGGTGATTACAGC<br>(SEQ ID NO:305) | 5'                |

## [1998] 实施例11

[1999] 本实施例总结了RESCUE第1-12轮的结果(见图57-68)。测试的额外表型包括PCSK9、Stat3、IRS1和TFEB。PCSK9显示克隆改善了启动子。Stat3显示在位点上的~10%编辑。信号转导的抑制将用荧光素酶报告基因进行测试。对于IRS1,将在转移至前脂肪细胞前测试合成位点的靶向。对于TFEB,可以将靶向设计为引起转录因子->自噬的易位。此外,还

将测试一组12个内源性磷酸靶标和48个合成靶标。酵母中的筛选将在V11背景上以S22P继续进行。在V12上为V13筛选顶部命中，并将评估新一轮次的酵母命中。将评估荧光素酶上的几百个额外筛选结果，并验证Ade2编辑的特异性筛选。基因混编还将测试针对文库复杂性和不同酵母报道基因进行测试。

[2000] 实施例12

[2001] 本实施例列出了与Cas13b-t有关的信息和数据。

[2002] 通过两个指导RNA在HEK293FT细胞中测试了高斯荧光素酶的敲低。RanCas13b (B6) 用作阳性对照。荧光素酶值相对于非靶向对照进行标准化。一些值高于1，但在一定范围内可归因于噪声。如果没有敲低，值是约1。死亡形式具有在两个已鉴定的HEPN结构域中均突变为丙氨酸的精氨酸和组氨酸残基两者。

[2003] 该实验中使用的间隔子序列示于下表18中。

[2004] 表18

| 名称     | 间隔子序列   |
|--------|---|
| 指导序列 1 | GGGCATTGGCTTCCATCTCTTTGAGCACCT<br>(SEQ ID NO:306) |
| 指导序列 1 | GGAATGTCGACGATCGCCTCGCCTATGCCG<br>(SEQ ID NO:307) |
| 非靶向    | GTAATGCCTGGCTTGTCGACGCATAGTCTG<br>(SEQ ID NO:308) |

[2006] 用于Gluc敲低的死的和活的微小直向同源物的比较显示在图69中。

[2007] 测试了通过RNA编辑来恢复功能性海萤荧光素酶 (W85X) 的能力。

[2008] 错配距离表示从正向重复序列的5'端到规定期望编辑位点的A:C错配的距离。除非另有说明，间隔子序列均为30bp。B6间隔子是30bp，错配距离为22。REPAIRv1、v2间隔子是50bp，错配距离为34(如公布的)。该微小直向同源物构建了 HIVNES-GS-dRanCas13bt-(GGG)2-huADAR2dd (E488Q)。

[2009] 阳性对照构建体如下：

[2010] B6构建体：HIVNES-GS-dRanCas13b (B6) - (GGG)2-huADAR2dd (E488Q)

[2011] REPAIRv1构建体：dPspCas13b (B12) -GS-HIVNES-GS-huADAR2dd (E488Q)

[2012] REPAIRv2构建体：dPspCas13b (B12) -GS-HIVNES-GS-huADAR2dd (E488Q/T375G)

[2013] 关于Cas13b-t1的数据显示于图70中，关于Cas13b-t3的数据显示于图71中。指导序列、非靶向比较显示于图72中。可以进行全转录组测序以进行详细的特异性和活性分析。

[2014] 实施例13

[2015] 可编程RNA编辑提供了基因组编辑的替代方法，具有安全性和靶向灵活性的优点。利用VI型可编程RNA指导RNase CRISPR-Cas13进行RNA编辑的方法，通过将融合的ADAR2的腺苷脱氨酶活性引导至目标转录物，允许腺苷向肌苷的特异性转化。在此，申请人通过直接使ADAR2进化为具有胞苷脱氨酶活性，将RNA编辑能力扩展到额外的碱基转化，催化活性提高1000倍以上。该系统称为“用于特异性C至U交换的RNA编辑 (RESCUE)”，缺乏严格的序列约束，以高效率编辑内源性转录，并执行C到U和A到I的多重编辑。申请人进行了额外的合理诱变以产生RESCUE的高度特异性变体，其中A至I脱靶降低10倍以上，其保留了有效的C至U中靶活性。申请人在本文中显示了RESCUE改变细胞中磷酸化信号传导通路并调节STAT激活

和细胞生长的能力。RESCUE 通过使得能够校正额外突变并调节更多的蛋白质残基,扩展了RNA编辑工具箱,可广泛应用于生物医学研究和治疗。

[2016] 细胞中的核酸的可编程修饰在基础研究和治疗中,尤其是在遗传疾病的治疗中,具有许多应用。随着基于CRISPR核酸酶的工具的开发,包括Cas9和Cpf1/Cas12a,DNA编辑(通常通过双链断裂(DSB)的生成来刺激内源性DNA修复通路,例如非同源末端结合(NHEJ)或同源定向修复(HDR))已变得广泛可用。但是,特定编辑的引入(包括单个碱基变化)依赖于HDR并且在许多细胞型中效率低下。此外,脱靶切割或DNA损伤反应的可能性带来潜在的安全风险。规避DSB形成的DNA编辑(例如碱基编辑)提供了可行的替代方式,尽管它们可受到测序约束的限制,例如在期望编辑位点附近对原型间隔子邻近基序(PAM)的要求,并且有显著的脱靶。但是,通过RNA碱基编辑来暂时控制核酸编辑避免了这些问题中的许多,并有许多应用,包括细胞信号传导、蛋白质稳定性或其他翻译后修饰残基的调节。

[2017] RNA碱基编辑为DNA碱基编辑器提供了一种替代方法,其利用作用于酶的RNA(ADAR)家族的腺苷脱氨酶使腺苷发生特异性水解脱氨基化为肌苷,肌苷在翻译和剪切方面功能等同于鸟苷。已经开发了多种RNA编辑技术,其将ADAR或高活性变体的活性引导至靶标转录物,包括用于可编程A至I(G)替换的RNA编辑(REPAIR),其使用RNA指导的、RNA靶向的CRISPR酶Cas13。尽管这些技术可以有效地将A转换为I(G),但其他碱基变化仍然难以获得,从而阻止了编辑与翻译后修饰相关的多样化疾病相关突变和功能残基。通过水解脱氨基化活性进行的胞苷到尿苷的编辑将打开靶向空间并提供多种新型的残基变化。但是,许多胞苷脱氨酶,例如载脂蛋白B mRNA编辑酶,催化多肽样(APOBEC)酶家族,仅可在单链底物上操作,并且将在APOBEC结合位点附近使胞嘧啶中的许多脱氨基。

[2018] 在此,申请人利用腺苷脱氨酶ADAR2的特征。使用ADAR2,REPAIR 允许通过使用指导RNA形成双链RNA底物来进行精确编辑,其将人ADAR2催化脱氨酶结构域(ADAR2dd[E488Q])活性的高活性突变体引导至通过引入的错配选择的单个腺苷。申请人进行了用于胞苷脱氨基化的ADAR2dd的进化,以将这种精确度水平赋予胞苷碱基转化。申请人使用组合的合理诱变和定向进化方案来将ADAR2dd的胞苷脱氨基化活性迭代地提高超过1000倍。融合至来自鸭疫里默氏菌(*Rimerella anatipestifer*)的Cas13b直向同源物(RanCas13b)的该突变体ADAR2dd 允许对哺乳动物细胞中的报告基因和内源转录物两者进行用于特异性C至U交换的RNA编辑(RESCUE)。最后,申请人通过合理诱变将RESCUE 的特异性提高了10倍以上,并展示了通过用RESCUE进行的C至U编辑而进行的蛋白质信号传导和细胞生长的表型调节。

[2019] 为了产生能够产生可编程C至U修饰的Cas13b指导核苷脱氨酶,申请人对RanCas13b-ADAR2dd融合开始了一系列工程化步骤(图73A-73G)。通过在涉及靶向碱基的结合的残基处的饱和诱变来选择初始突变。评价突变体的C至U编辑和高斯荧光素酶(Gluc)突变体(C82R)催化活性的恢复(图77A)。三轮合理工程化产生了对TCG基序具有~15%编辑的构建体(RESCUEv3)(图73B)。由于周围的基序强烈决定了A至I编辑的RNA编辑效率,因此申请人测试了荧光素酶突变体的活性的恢复,该突变体具有在Gluc C82R位点的所有四个可能的5'碱基和在Gluc L77P突变处的两个3'基序(图77B),发现与这些其他基序的活性的适度增加。为了加快进一步的改进,申请人开始了在ADAR2dd蛋白中的定向进化以确定额外候选突变用于增加RESCUE的活性。

[2020] 为了针对C至U活性进行选择,申请人基于GFP荧光的恢复或His 营养缺陷型选择基因的原养型逆转设计了一套酵母报告基因测定法(图73A,有关所有筛选和所得突变请参见表19)。使用类似的方法,也可以进行作用于RNA (CDAR) 的胞苷脱氨酶的定向进化。

[2021] 表19

|        | RESCUE 版本号 | 突变            | 筛选方法                    |
|--------|------------|---------------|-------------------------|
|        | RESCUEv0   | ADAR2 + E488Q | 来自 Kuttan 和 Bass 的超活性变体 |
| [2022] | RESCUEv1   | v0+V351G      | 合理诱变                    |
|        | RESCUEv2   | v1+S486A      | 合理诱变                    |
|        | RESCUEv3   | v2+T375S      | 合理诱变                    |
|        | RESCUEv4   | v3+S370C      | Y66H EGFP               |
|        | RESCUEv5   | v4+P462A      | P196L HIS               |
|        | RESCUEv6   | v5+N597I      | P196L HIS               |
|        | RESCUEv7   | v6+L332I      | P196L HIS               |
|        | RESCUEv8   | v7+I398V      | P196L HIS               |
|        | RESCUEv9   | v8+K350I      | P196L HIS               |
| [2023] | RESCUEv10  | v9+M383L      | P196L HIS               |
|        | RESCUEv11  | v10+D619G     | S22P HIS                |
|        | RESCUEv12  | v11+S582T     | S22P HIS                |
|        | RESCUEv13  | v12+V440I     | S22P HIS                |
|        | RESCUEv14  | v13+S495N     | P196L HIS               |
|        | RESCUEv15  | v14+K418E     | P196L HIS               |
|        | RESCUEv16  | v15+S661T     | S22P HIS                |

[2024] 对分别用于GFP和His还原的FACS分类培养物或存活菌落进行测序,选择ADAR2dd结构域中的单个突变,将其引入先前RESCUE版本并评估荧光素酶或CTNNB1编辑报告基因构建体在哺乳动物细胞中的活性。这些轮次的进化以最终构建体RESCUEv16告终,导致所测试的所有六个基序中活性的稳定增加并降低了编辑和恢复荧光素酶活性所需的 RESCUE和指导质粒剂量(图73C、73D、78、79A-79B、80)。另外, RESCUEv16在直接5' 和3' 碱基的16种可能基序组合中的12种上实现了超过20%的编辑,并且具有最佳的C或U碱基翻转(图73E和81)。申请人比较了我们的RESCUE版本与PspCas13b和RanCas13b的融合,发现它们具有同等的活性(图82)。当REPAIR使用50nt的指导序列时, RESCUEv16用30nt的指导RNA最优地编辑了TCG构建体,该指导RNA 具有从靶标的5' 端起的靶向碱基翻转26碱基对(图83)。

[2025] 为了验证来自酵母系统中定向进化管线的改进,申请人测试了多个 RESCUE迭代的酵母中活性和生化活性两者。在酵母中检测EGFP和His 恢复两者,申请人发现RESCUE的

最近版本在这两个靶标上都更有效地执行C至U编辑(图84A-84D)。将RESCUE构建体引入纯化的hADAR2dd 蛋白中的生化表征揭示了RESCUE突变改善了在体外底物上C到U编辑的动力学(图85A-85B)。

[2026] 此外,申请人在不存在Cas13b构建体的情况下测定了C至U活性。申请人将RESCUEv16突变引入到ADAR2脱氨酶结构域或全长ADAR2 蛋白两者中。申请人发现,与ADARdd、全长ADAR或不存在蛋白时相比,对于完整的RESCUEv16构建体,荧光素酶活性的编辑和恢复在所有5' 基序上均显著更高(图73F和86A),而且,尽管某些指导位置在使用全长ADAR时实现几乎20%的编辑(图86B-86D),与RESCUE相比最大效率明显降低,这证明RanCas13b融合对于其功能是必需的。RESCUEv16中16个突变的位置使它们位于ADAR2dd的整个结构中(图 73G),指示引入的残基与催化口袋的直接相互作用以及远距离变构效应两者。

[2027] 随着RESCUE进化以对报告基因构建体具有活性,申请人评估了 RESCUE可以在HEK293FT细胞中对内源转录物发挥多大的作用。申请人测试了一组具有可变错配位置的指导RNA,其靶向9个基因中的24 个不同位点(图74A和87A-87C),尤其选择这些基因中的位点以具有不同的5' 碱基身份以询问不同基序上的脱氨基化活性。申请人发现,在所有测试的位点,RESCUEv16实现了在~5%-35%之间的编辑率,并且理想的错配位置或碱基翻转是位点相关性的。此外,RESCUEv16在多个内源位点上的性能均优于所有其他版本,并且需要比早期版本更少的剂量(图74B和88)。为了更好地评估RESCUEv16与治疗的相关性,申请人设计了一系列22个200bp的靶标,以对来自ClinVar的疾病相关突变进行模型编辑(参见表20)。

[2028] 表20:疾病相关突变的疾病信息

| 候选   | 基因            | 疾病                           |
|--|---------------|------------------------------|
| NM_000071.2(CBS):c.325T>C (p.Cys109Arg)            | <i>CBS</i>    | 胸主动脉瘤和主动脉夹层                  |
| NM_000141.4(FGFR2):c.799T>C (p.Ser267Pro)          | <i>FGFR2</i>  | Pfeiffer 综合征/Crouzon 综合征/胃肿瘤 |
| NM_000551.3(VHL):c.473T>C (p.Leu158Pro)            | <i>VHL</i>    | Von Hippel-Lindau 综合征        |
| [2029] NM_002474.2(MYH11):c.3791T>C (p.Leu1264Pro) | <i>MYH11</i>  | 主动脉瘤、家族性胸 4/胸主动脉瘤和主动脉夹层      |
| NM_000018.3(ACADVL):c.848T>C (p.Val283Ala)         | <i>ACADVL</i> | 超长链酰基辅酶 A 脱氢酶缺乏症             |
| NM_002397.4(MEF2C):c.2T>C (p.Met1Thr)              | <i>MEF2C</i>  | 智力低下、定型运动、癫痫和/或脑畸形           |
| NM_002834.4(PTPN11):c.853T>C (p.Phe285Leu)         | <i>PTPN11</i> | 努南综合征                        |

|        |  |               |  |
|--------|--|---------------|--|
|        | NM_005609.3(PYGM):c.2392T>C (p.Trp798Arg)      | <i>PYGM</i>   | 糖原贮积病, V 型   |
|        | NM_001256850.1(TTN):c.90211T>C (p.Cys30071Arg) | <i>TTN</i>    | 肢带型肌营养不良, 2J 型/远端肌病<br>Markesbery-Griggs 型/遗传性肌病伴早期呼吸衰竭/早发性肌病, 伴致命性心脏病/家族性肥厚型心脏病 9 |
|        | NM_005633.3(SOS1):c.806T>C (p.Met269Thr)       | <i>SOS1</i>   | 努南综合征 4/努南综合征  |
|        | NM_015559.2(SETBP1):c.2612T>C (p.Ile871Thr)    | <i>SETBP1</i> | Schinzel-Giedion 综合征   |
|        | NM_004572.3(PKP2):c.2386T>C (p.Cys796Arg)      | <i>PKP2</i>   | 心律失常性右室心肌病, 9 型  |
|        | NM_000138.4(FBN1):c.4222T>C (p.Cys1408Arg)     | <i>FBN1</i>   | 马凡氏综合征   |
|        | NM_000375.2(UROS):c.217T>C (p.Cys73Arg)        | <i>UROS</i>   | 先天性红细胞生成性卟啉症   |
| [2030] | NM_014139.2(SCN11A):c.1187T>C (p.Leu396Pro)    | <i>SCN11A</i> | 未提供/神经病变, 遗传性感觉和自主神经, VII 型  |
|        | NM_000152.4(GAA):c.1655T>C (p.Leu552Pro)       | <i>GAA</i>    | 糖原贮积病, II 型  |
|        | NM_020630.4(RET):c.1858T>C (p.Cys620Arg)       | <i>RET</i>    | 多发性内分泌肿瘤 2a 型/多发性内分泌肿瘤 2 型/MEN2A 和 FMTC  |
|        | NM_000016.5(ACADM):c.199T>C (p.Tyr67His)       | <i>ACADM</i>  | 中链酰基辅酶 A 脱氢酶缺乏症  |
|        | NM_014874.3(MFN2):c.227T>C (p.Leu76Pro)        | <i>MFN2</i>   | 腓骨肌萎缩症 2A2A 型  |
|        | NM_000341.3(SLC3A1):c.1400T>C (p.Met467Thr)    | <i>SLC3A1</i> | 胱氨酸尿   |
|        | NM_000431.3(MVK):c.803T>C (p.Ile268Thr)        | <i>MVK</i>    | 甲羟戊酸尿症/伴发高热的高免疫球蛋白 D   |
|        | NM_004004.5(GJB2):c.29T>C (p.Trp77Arg)         | <i>GJB2</i>   | 耳聋, 常染色体隐性遗传 1A /耳聋, 常染色体显性 3a /非综合征性听力损失和耳聋                                       |
|        | NM_000041.4(APOE):c.388T>C (p.Cys130Arg)       | <i>APOE</i>   | 阿尔茨海默氏病 2  |
| [2031] | NM_000041.4(APOE):c.595T>C (p.Cys176Arg)       | <i>APOE</i>   | 阿尔茨海默氏病 2  |

[2032] RESCUEv16能够以~1%-42%的效率编辑这些位点(图74C和89)。申请人进一步测试了在ApoE4等位基因上的治疗应用,其使阿尔兹海默病的ris10倍升高约10倍,并涉及到两个胞嘧啶单核苷酸多态性,其需要转化为胸腺嘧啶以产生保护性ApoE2等位基因。申请人在来自ApoE4等位基因的表达合成片段上测试了RESCUEv16,并发现该系统在两个位点上实现了约5%和12%的编辑(图90)。

[2033] 由于RESCUEv16保留了腺苷脱氨酶活性,因此Cas13b的天然前 crRNA加工活性使多重腺嘌呤和胞嘧啶脱氨基化成为可能。通过在相同的CTNNB1转录物中递送RESCUEv16以及靶向腺嘌呤和胞嘧啶的前 crRNA(图74D),申请人发现RESCUEv16能够编辑同一群体中的靶向残基两者,分别以~15%和5%的速率将腺嘌呤转化为肌苷和将胞嘧啶转化为尿苷(图74E)。此外,申请人发现当编辑Gluc和内源基因时,在指导双链体内发生了被靶向的胞嘧啶附近A至I脱靶(图91A-91C)。为了消除这些脱靶,申请人在脱靶腺苷对面的指导序列中引入了不利的鸟嘌呤错配(图74F)。这种方法显著减少了Gluc和KRAS两者上的脱靶编辑,同时最小程度地干扰中靶编辑(图74G)。

[2034] 在指导双链体窗口中观察到的A至I脱靶表明RESCUEv16可能在整个转录组中具有明显的脱靶腺苷脱氨酶活性。使用全转录组RNA测序分析脱靶,申请人发现尽管RESCUEv16在Gluc转录物上有约80%的C至U编辑(图75A),其因此具有188C至U脱靶和1695A至I脱靶,与REPAIRv1的A至I脱靶是可比较的,其具有24C至U脱靶和2214A至I脱靶(图75A、75B)。为了提高RESCUEv16的特异性,申请人在与RNA靶标相互作用的残基处进行了合理诱变(图75C),从而产生具有如通过荧光素酶报告基因测试的降低的A至I脱靶活性和高C至U中靶脱氨基化活性的多个RESCUEv16突变体(图75D)。RESCUEv16上的最高特异性突变体S375A(RESCUEv16S)保持约76%中靶C至U编辑(图75E),但仅有103C至U脱靶和139A至I脱靶,腺嘌呤脱氨基化脱靶的数量减少约10倍(图75E、75F)。尽管RESCUEv16S的脱靶编辑减少,但它仍保持了显著的中靶A至I编辑活性(图92A-92D)。申请人重新评估了RESCUEv16S在前一组内源位点上的效力,发现它在许多位点处保留了与RESCUEv16类似的活性,在大量位点上表现优于RESCUEv16(图93A-93C和94A)。此外,在指导双链体窗口内,RESCUEv16S特异性高得多,显著减少了在许多局部脱靶位点处的编辑(图93C、94B-94E)。

[2035] RESCUEv16的胞苷和腺苷脱氨基化活性允许通过错义突变来调节翻译后修饰,例如磷酸化底物丝氨酸和酪氨酸。STAT3和STAT1是转录因子,其在通过JAK/STAT通路进行信号转导中起重要作用,并通常被细胞因子和生长因子激活。为了展示通过RNA编辑进行的信号传导调节,申请人通过使用RESCUEv16编辑STAT3上的磷酸化位点Y705和S727以及STAT1上的Y701和S727来改变STAT通路的激活(图76A)。在HEK293FT细胞中,申请人分别观察到Y705和S727 STAT3位点的8%和9%编辑,以及Y701和S727 STAT1位点的11%和7%的编辑(图76B)。使用针对STAT激活的荧光素酶报告基因,这些编辑导致STAT3和STAT1活性的16%-27%抑制(图76C)。

[2036] 如同JAK/STAT通路,Wnt通路可以通过组成蛋白(最有名的是 $\beta$ -连环蛋白)的磷酸化来调节。 $\beta$ -连环蛋白上的磷酸化残基,例如S33和S37,促进泛素化和降解。Wnt信号传导阻断残基磷酸化并稳定 $\beta$ -连环蛋白,从而允许该蛋白与转录因子(如LEF和TCF1/2/3)接合,从而促进靶基因的表达,并导致细胞增殖增加。申请人测试了一组针对已知参与 $\beta$ -连环蛋白的磷酸化的残基的指导序列,并发现编辑水平在5%-28%之间(图76F),从而导致如通

过TCF/LEF依赖性荧光素酶报告基因测量的,β-连环蛋白的至多5倍激活(图76G)。相应地,用靶向磷酸化位点的 RESCUEv16转染的细胞在最激活β-连环蛋白的条件下(靶向T41I转化)导致细胞生长增加40%(图76H)。

[2037] RESCUEv16是一种可编程的碱基编辑工具,其能够将RNA中的胞苷精确转化为尿苷。使用定向进化,申请人证明了腺苷脱氨酶可以被放宽以接受其他碱基,从而产生一种新型胞苷脱氨基化机制,该机制可以通过碱基翻转来编辑双链RNA。申请人已经能够将ADAR2dd的胞苷脱氨酶活性提高1000倍,从而导致对内源转录物高达40%的编辑。可以进行进一步的进化轮次以进一步增强活性。RESCUE的胞苷脱氨基化活性的更大可靶向氨基酸空间增加了翻译后修饰(如磷酸化、糖基化和甲基化位点)的可能调节以及更好地靶向常见催化残基(图95A-95B)。此外,胞苷脱氨基化活性允许通过RNA编辑和产生保护性等位基因(例如 ApoE2)来扩展与疾病相关突变的靶向性。总而言之,RESCUE扩展了具有新碱基编辑功能的RNA靶向工具包,从而允许更好地建模和治疗遗传疾病。

[2038] RESCUE v16S能够有效编辑内源基因(图96)。RESCUE v16S保持了一些A至I活性(图97)。RESCUE v16被用于靶向STAT以减少 INFα/IL6诱导(图98)。RESCUE靶向诱导细胞生长(图99A-99B)。

[2039] 材料与amp;方法

[2040] 酵母构建体的设计和克隆

[2041] 为了在酵母中表达dRanCas13b-hADAR2dd构建体,通过修饰 pML104(Addgene# 67638)将该融合蛋白克隆到pRSII426主链中pGAL 启动子的下游。为了提高表达,在融合蛋白之间克隆GS接头,并针对酵母对ADAR2dd进行了密码子优化。通过Gibson克隆引入与RESCUE 迭代相对应的额外密码子突变。

[2042] 对用于测试酵母中活性的靶向质粒进行工程化以用于荧光筛选 (GFP)和营养缺陷选择筛选(His)两者。将所有靶向质粒克隆到 pYES3/CT主链(Thermo Scientific)中。所有质粒包含用于RESCUE的 RanCas13b指导盒,其表达由ADH1启动子驱动,以及侧接HH和HDV核酶的间隔子和DR序列[*cite ng and dean*]。克隆了间隔子被金门位点取代的构建体以便利模块指导序列克隆。

[2043] 为了生成C至U RNA编辑活性的GFP指示剂,将Y66H绿色至蓝色突变引入由TEF启动子驱动的酵母密码子优化的EGFP(yeGFP)中。成功的C至U RNA编辑恢复该构建体的绿色荧光。通过测试HIS3中的保守残基当突变为可通过RNA编辑来挽救的残基时的活性丧失来生成用于C至U编辑的His报告基因。将生成失活HIS3的突变在其天然HIS3 启动子下克隆到pYES3/CT主链中的HIS3基因中。

[2044] 用于酵母筛选的诱变文库的生成

[2045] 为了生成用于在酵母系统中筛选突变的诱变文库,使用Genemorph II (Agilent Technologies)对hADAR2脱氨酶结构域进行了突变,用于通过2倍稀释系列,差异在于74ng-9.4μg的模板输入的8个50mL反应的易错PCR。扩增后,将反应物合并,在去离子水中以1:4稀释,并加载到含有溴化乙锭的2%凝胶中。使用MinElute PCR纯化试剂盒(Qiagen) 纯化提取的样品,然后在37°C下用DpnI(Thermo Fisher Scientific)处理 2h以去除残留的模板质粒,然后进行凝胶和MinElute纯化。

[2046] 通过用Kf1I、RruI和Eco72I(Thermo Fisher Scientific)消化7μg模板质粒1小时

来生成主链。用MinElute PCR纯化试剂盒对消化物进行凝胶纯化,并在30 $\mu$ L的预加温水中洗脱。

[2047] 使用Gibson Assembly (New England Biosciences) 连接经纯化的PCR 插入序列和经消化的主链,具体而言,将456ng PCR插入序列和800ng 主链消化物在80 $\mu$ L Gibson反应中运行1小时。用异丙醇沉淀将产物浓缩,重悬于12 $\mu$ L TE-EF再溶解缓冲液 (Macherey-Nagel) 中,并在300 r.p.m振荡的同时加热至50 $^{\circ}$ C 5分钟。将50 $\mu$ L Endura电感受态细胞 (Lucigen) 在冰上解冻10分钟,然后加入2 $\mu$ L重悬的Gibson产物。遵循最佳Endura设置 (1.0mm吸收池、10 $\mu$ F、600欧姆和1800伏),使用GenePulser Xcell (Bio-Rad) 对混合物进行电穿孔。将来自每个电穿孔的样品回收到1mL的回收培养基 (Lucigen) 中,并在37 $^{\circ}$ C下孵育1小时同时在300r.p.m摇动。每个诱变文库进行两次电穿孔。将回收的培养物铺板在大型预加温100 $\mu$ g/mL氨苄青霉素板上。制备系列稀释物以测定每个文库的c.f.u.。将板在37 $^{\circ}$ C下孵育16小时,并使用Nucleobond Xtra Maxi Kit (Macherey-Nagel) 收获。

#### [2048] 酵母中诱变文库的转化

[2049] 如前所述进行大规模酵母转化。简而言之,将含有Y66H EGFP或 HIS3报告质粒的群落放入300mL-Trp 2%葡萄糖选择培养基中并在30 $^{\circ}$ C生长过夜。生长后,测定细胞的OD600,并将2.5e9细胞加入到500mL 预加温2 $\times$ YPAD中,并在30 $^{\circ}$ C孵育4小时。洗涤细胞沉淀多次,然后重悬于36mL转化混合物中,该混合物包含24mL PEG 3350 (50%w/v)、3.6mL的1.0M乙酸锂、5mL的2.0mg/mL的变性单链载体鲑鱼精子DNA (ThermoFisher Scientific)、2.9mL水、和500 $\mu$ L的1 $\mu$ g/ $\mu$ L质粒文库。在42 $^{\circ}$ C下孵育60分钟后,将细胞沉淀重悬于750mL的-Ura/-Trp 2%葡萄糖选择培养基中,并生长过夜直至培养物达到OD600为5-6。此时,将6mL的培养物接种到250mL的2%棉子糖-Ura/-Trp选择培养基并孵育直至OD600为0.5-1。加入27mL的30%半乳糖来诱导培养物,并在 30 $^{\circ}$ C下孵育过夜12-14小时。然后对细胞进行细胞分选或铺板到选择板上,如下所述。

#### [2050] 酵母文库的荧光细胞分选

[2051] 诱导后,通过与阴性非诱导和非靶向指导序列对照相比,对EGFP 荧光进行门控,在SH800S细胞分选仪上对细胞进行分选。在将1亿个细胞分选到2%葡萄糖-Ura/-Trp选择培养基中后,申请人将分选的细胞孵育过夜,然后在其OD600为5-6时将其接种到2%棉子糖-Ura/-Trp选择培养基中。然后当OD600在0.5-1之间时诱导细胞,并孵育过夜12-14 小时,然后再次分选。进行分选直至分选了1-2千万个细胞。迭代生长和分选又重复了2-3次,分选细胞的每次迭代均收集质粒并通过Illumina NextSeq下一代测序进行测序以确定突变体在每一轮选择中的存在。单独订购最富集的突变体,并按如下所述进行克隆用于哺乳动物验证测试。

#### [2052] 酵母文库的His生长选择

[2053] 诱导后,将细胞文库铺板在2%棉子糖/3%半乳糖-Ura/-Trp/-His选择板上。随着群落生长,将它们挑到水中并在2%棉子糖/3%半乳糖 -Ura/-Trp/-His选择板上划线。条纹生长过夜后,对每个条纹进行菌落PCR 并对ADAR2催化结构域以及His基因进行桑格测序以检查重组和DNA 诱变。单独订购突变并按如下所述进行克隆用于哺乳动物验证测试。

#### [2054] 设计和克隆用于RNA编辑的哺乳动物构建体

[2055] 通过在HEPN结构域的催化位点处的组氨酸至丙氨酸和精氨酸至丙氨酸突变

(R142A/H147A/R1039A/H1044A)使RanCas13b催化失活(dRanCas13b)。合成脱氨酶结构域和ADAR2并进行PCR扩增以Gibson克隆到pcDNA-CMV载体主链中,并通过GS-mapkNES-GS(GSSLQKKLEELGGS(SEQ ID NO:309))接头在C端处与dRanCas13b融合。将用于改变胞嘧啶脱氨酶活性或特异性的ADAR2脱氨酶结构域中的突变通过Gibson克隆引入到dRanCas13b-GS-mapkNES-GS-ADAR2dd主链中。表25列出了ADAR2dd中引入的所有用于进化C至U编辑的突变。

[2056] 为了在不同Cas13b直向同源物之间进行比较,通过Gibson克隆到REPAIR构建体dPspCas13b-GS-HIVNES-GS-ADAR2dd上,在dRanCas13b主链上测试的突变被转移至dPspCas13b融合载体。为了单独测试不含dRanCas13b的ADAR2dd和全长ADAR2,申请人使用Gibson克隆将所有突变加入到具有ADAR2dd或全长ADAR2的pcDNA-CMV载体主链上,之前已克隆以测试REPAIR。

[2057] 通过筛选先前报道的荧光素酶报告质粒中Gluc中的潜在突变,生成了用于测量C至U RNA编辑活性的荧光素酶报告载体。该报告载体将功能性Cluc表达为标准化对照,但是由于添加了突变体(C82R或L77P)而导致了缺陷型Gluc。为了测试RESCUE编辑基序偏好,申请人克隆了Gluc的密码子82(AAX CXC)处的胞嘧啶周围的每个可能基序。从M50 Super 8x TOPFlash(Addgene#12456)和M50 Super 8x FOPFlash(Addgene #12457)生成了用于测试CTNNB1编辑效率的分泌的荧光素酶报告载体。通过Gibson克隆,在TCF/LEF应答元件(TOPFlash)或模拟结合位点(TOPFlash)的控制下,将最初的萤火虫荧光素酶替换为分泌型高斯荧光素酶。克隆进了由CMV启动子驱动表达的额外的海萤荧光素酶充当转染对照。表22列出了所有的哺乳动物质粒。

[2058] 哺乳动物细胞中RESCUE版本的选择

[2059] 选择表现与RESCUE的现有版本表现相当或更好的突变,用于在6个荧光素酶报告基因的整个组上进行筛选。对于RESCUE v4至v10的选择,首先在TCG基序上筛选候选突变;使用GCG基序作为初筛分离了RESCUE v11。使用对内源性CTNNB1的T41I残基的编辑进行的初筛,在哺乳动物细胞中验证了RESCUE v12至v14的选择,导致 $\beta$ -连环蛋白通路激活(其用通路活性的发光报告基因进行了谱分析),以及通过对Gluc的L77P CCT基序的活性选择了RESCUE v15和v16。表25列出了用于产生它们的所有轮次和酵母筛选。

[2060] 克隆致病性U>C突变用于测定RESCUE活性

[2061] 为了产生用于测试REPAIR活性的疾病相关突变,选择了ClinVar中定义的与疾病发病机理相关的23个U>C突变(独立性地分组为一组22个基因和ApoE)。从Integrated DNA Technologies订购了选择的靶标作为突变位点周围200-bp的区域,并在Efl $\alpha$ 启动子下克隆到mScarlet的下游。

[2062] 用于RESCUE的指导序列克隆

[2063] 为了表达用于RESCUE的哺乳动物指导RNA,使用先前描述的具有前面是在U6表达下的金门受体位点的RanCas13b正向重复序列的构建体。通过金门克隆将各个指导序列克隆到该表达主链中。为了确定用于选择位点的最佳指导序列,测试了C和U翻转两者,以及在最常见的最佳指导序列范围(错配距离为~24)附近的平铺指导序列。

[2064] 表21-26列出了用于RESCUE试验的指导序列、所有酵母质粒以及在酵母实验中使用的所有靶向指导序列。

[2065] 表21:用于荧光素酶编辑的指导序列

| 名称               | 被靶向的基因 | 基序  | 碱基翻转/<br>间隔子长度/位置 | 密码子变化 | 间隔子序列   | 备注         |
|------------------|--------|-----|-------------------|-------|---|------------|
| UCG 靶向指导序列       | Gluc   | UCG | C/30/26           | C82R  | gugcCauugauguggacaggcagauca<br>aga (SEQ ID NO:310)  | No 5'<br>G |
| GCG 靶向指导序列       | Gluc   | GCG | U/30/20           | C82R  | Guuggcgugcucugauguggacaggcag<br>gag (SEQ ID NO:311) |            |
| ACG 靶向指导序列       | Gluc   | ACG | C/30/28           | C82R  | Ggccuuugauguggacaggcagauca<br>gaca (SEQ ID NO:312)  |            |
| CCG 靶向指导序列       | Gluc   | CCG | C/30/26           | C82R  | Gugccguugauguggacaggcagauca<br>aga (SEQ ID NO:313)  | No 5'<br>G |
| CCU 靶向           | Gluc   | CCU | C/30/26           | L77P  | Gggaacggcagaucaagacagcccgug<br>gca (SEQ ID NO:314)  |            |
| CCA 靶向           | Gluc   | CCA | C/30/26           | L77P  | Gggauuggcagaucaagacagcccgug<br>gca (SEQ ID NO:315)  |            |
| 基序指导序列 UCU, 翻转 U | Gluc   | UCU | U/30/26           | L82F  | gugaUauugauguggacaggcagauca<br>aga (SEQ ID NO:316)  |            |
| 基序指导序列 UCG, 翻转 U | Gluc   | UCG | U/30/26           | C82R  | gugcUauugauguggacaggcagauca<br>aga (SEQ ID NO:317)  |            |
| 基序指导序列 UCG, 翻转 U | Gluc   | UCC | U/30/26           | P82S  | guggUauugauguggacaggcagauca<br>aga (SEQ ID NO:318)  |            |
| 基序指导序列 UCA, 翻转 U | Gluc   | UCA | U/30/26           | H82Y  | guguUauugauguggacaggcagauca<br>aga (SEQ ID NO:319)  |            |

[2066]

[2067]

|                  |      |     |         |      |   |  |
|------------------|------|-----|---------|------|---|--|
| 基序指导序列 ACU, 翻转 U | Gluc | ACU | U/30/26 | L82F | gugaUuuugauguggacaggcagauc<br>aga (SEQ ID NO:320) |  |
| 基序指导序列 ACG, 翻转 U | Gluc | ACG | U/30/26 | C82R | gugcUuuugauguggacaggcagauc<br>aga (SEQ ID NO:321) |  |
| 基序指导序列 ACC, 翻转 U | Gluc | ACC | U/30/26 | P82S | guggUuuugauguggacaggcagauc<br>aga (SEQ ID NO:322) |  |
| 基序指导序列 ACA, 翻转 U | Gluc | ACA | U/30/26 | H82Y | guguUuuugauguggacaggcagauc<br>aga (SEQ ID NO:323) |  |
| 基序指导序列 GCU, 翻转 U | Gluc | GCU | U/30/26 | L82F | gugaUcuugauguggacaggcagauc<br>aga (SEQ ID NO:324) |  |
| 基序指导序列 GCG, 翻转 U | Gluc | GCG | U/30/26 | C82R | gugcUcuugauguggacaggcagauc<br>aga (SEQ ID NO:325) |  |
| 基序指导序列 GCC, 翻转 U | Gluc | GCC | U/30/26 | P82S | guggUcuugauguggacaggcagauc<br>aga (SEQ ID NO:326) |  |
| 基序指导序列 GCA, 翻转 U | Gluc | GCA | U/30/26 | H82Y | guguUcuugauguggacaggcagauc<br>aga (SEQ ID NO:327) |  |
| 基序指导序列 CCU, 翻转 U | Gluc | CCU | U/30/26 | L82F | gugaUguugauguggacaggcagauc<br>aga (SEQ ID NO:328) |  |
| 基序指导序列 CCG, 翻转 U | Gluc | CCG | U/30/26 | C82R | gugcUguugauguggacaggcagauc<br>aga (SEQ ID NO:329) |  |
| 基序指导序列 CCC, 翻转 U | Gluc | CCC | U/30/26 | P82S | guggUguugauguggacaggcagauc<br>aga (SEQ ID NO:330) |  |
| 基序指导序列 CCA, 翻转 U | Gluc | CCA | U/30/26 | H82Y | guguUguugauguggacaggcagauc<br>aga (SEQ ID NO:331) |  |

[2068]

|                  |      |     |         |      |  |
|------------------|------|-----|---------|------|--|
| 基序指导序列 UCU, 翻转 C | Gluc | UCU | C/30/26 | L82F | gugaCauugaugugggacaggcagauc<br>aga (SEQ ID NO:332) |
| 基序指导序列 UCC, 翻转 C | Gluc | UCC | C/30/26 | P82S | guggCauugaugugggacaggcagauc<br>aga (SEQ ID NO:333) |
| 基序指导序列 UCA, 翻转 C | Gluc | UCA | C/30/26 | H82Y | guguCauugaugugggacaggcagauc<br>aga (SEQ ID NO:334) |
| 基序指导序列 ACU, 翻转 C | Gluc | ACU | C/30/26 | L82F | gugaCuuugaugugggacaggcagauc<br>aga (SEQ ID NO:335) |
| 基序指导序列 ACG, 翻转 C | Gluc | ACG | C/30/26 | C82R | gugcCuuugaugugggacaggcagauc<br>aga (SEQ ID NO:336) |
| 基序指导序列 ACC, 翻转 C | Gluc | ACC | C/30/26 | P82S | guggCuuugaugugggacaggcagauc<br>aga (SEQ ID NO:337) |
| 基序指导序列 ACA, 翻转 C | Gluc | ACA | C/30/26 | H82Y | guguCuuugaugugggacaggcagauc<br>aga (SEQ ID NO:338) |
| 基序指导序列 GCU, 翻转 C | Gluc | GCU | C/30/26 | L82F | gugaCcuugaugugggacaggcagauc<br>aga (SEQ ID NO:339) |
| 基序指导序列 GCG, 翻转 C | Gluc | GCG | C/30/26 | C82R | gugcCcuugaugugggacaggcagauc<br>aga (SEQ ID NO:340) |
| 基序指导序列 GCC, 翻转 C | Gluc | GCC | C/30/26 | P82S | guggCcuugaugugggacaggcagauc<br>aga (SEQ ID NO:341) |
| 基序指导序列 GCA, 翻转 C | Gluc | GCA | C/30/26 | H82Y | guguCcuugaugugggacaggcagauc<br>aga (SEQ ID NO:342) |
| 基序指导序列 CCU, 翻转 C | Gluc | CCU | C/30/26 | L82F | gugaCguugaugugggacaggcagauc<br>aga (SEQ ID NO:343) |

[2069]

|                                |      |     |         |      |   |             |
|--------------------------------|------|-----|---------|------|---|-------------|
| 基序指导序列 CCC, 翻转 C               | Gluc | CCC | C/30/26 | P82S | guggCguugauguggacaggcagauca (SEQ ID NO:344)                       |             |
| 基序指导序列 CCA, 翻转 C               | Gluc | CCA | C/30/26 | H82Y | guguCguugauguggacaggcagauca (SEQ ID NO:345)                       |             |
| 非靶向指导序列                        | N/A  | N/A | N/A     | N/A  | Guaaugccuggcuugcagcagcaugucug (SEQ ID NO:346)                     |             |
| 具有脱靶 A-G 错配 1 的 Gluc 特异性指导序列   | Gluc | UCG | C/30/26 | C82R | ggugcuaGugauguggacaggcagauca (SEQ ID NO:347)                      | 额外的 G 增加特异性 |
| 具有脱靶 A-G 错配 2 的 Gluc 特异性指导序列   | Gluc | UCG | C/30/26 | C82R | ggugcuaUGauguggacaggcagauca (SEQ ID NO:348)                       | 额外的 G 增加特异性 |
| 具有脱靶 A-G 错配 3 的 Gluc 特异性指导序列   | Gluc | UCG | C/30/26 | C82R | ggugcuaauugaugGggacaggcagauca (SEQ ID NO:349)                     | 额外的 G 增加特异性 |
| 具有脱靶 A-G 组合 1+2 的 Gluc 特异性指导序列 | Gluc | UCG | C/30/26 | C82R | ggugcuaGGauguggacaggcagauca (SEQ ID NO:350)                       | 额外的 G 增加特异性 |
| 具有脱靶 A-G 所有组合的 Gluc 特异性指导序列    | Gluc | UCG | C/30/26 | C82R | ggugcuaGGaugGggacaggcagauca (SEQ ID NO:351)                       | 额外的 G 增加特异性 |
| A 至 I REPAIR 指导序列              | Cluc | TAG | C/50/34 | *85W | Ggccccugugcggacuccuugucgcuucguaggugggcagcguccuggg (SEQ ID NO:352) |             |
| 平铺指导序列                         | Gluc | UCG | U/30/30 | C82R | Guaaugauguggacaggcagauca (SEQ ID NO:353)                          |             |

[2070]

|                     |      |     |         |      |   |  |
|---------------------|------|-----|---------|------|---|--|
| 30 翻转 30U           |      |     |         |      |   |  |
| 平铺指导序列<br>30 翻转 28U | Gluc | UCG | U/30/28 | C82R | Ggcuaaugaugugggacaggcagauca<br>gaca (SEQ ID NO:354) |  |
| 平铺指导序列<br>30 翻转 26U | Gluc | UCG | U/30/26 | C82R | Ggugcuaaugaugugggacaggcagau<br>caga (SEQ ID NO:355) |  |
| 平铺指导序列<br>30 翻转 24U | Gluc | UCG | U/30/24 | C82R | Ggcgugcuaaugaugugggacaggcag<br>auca (SEQ ID NO:356) |  |
| 平铺指导序列<br>30 翻转 22U | Gluc | UCG | U/30/22 | C82R | Gggcgugcuaaugaugugggacaggc<br>agau (SEQ ID NO:357)  |  |
| 平铺指导序列<br>30 翻转 20U | Gluc | UCG | U/30/20 | C82R | Guuggcgugcuaaugaugugggacag<br>gcag (SEQ ID NO:358)  |  |
| 平铺指导序列<br>30 翻转 18U | Gluc | UCG | U/30/18 | C82R | Gucuuggcgugcuaaugaugugggac<br>aggc (SEQ ID NO:359)  |  |
| 平铺指导序列<br>30 翻转 16U | Gluc | UCG | U/30/16 | C82R | Gcaucuuggcgugcuaaugauguggg<br>acag (SEQ ID NO:360)  |  |
| 平铺指导序列<br>30 翻转 14U | Gluc | UCG | U/30/14 | C82R | Guucaucuuggcgugcuaaugaugug<br>ggac (SEQ ID NO:361)  |  |
| 平铺指导序列<br>30 翻转 12U | Gluc | UCG | U/30/12 | C82R | Gucuucaucuuggcgugcuaaugaug<br>uggg (SEQ ID NO:362)  |  |
| 平铺指导序列<br>30 翻转 10U | Gluc | UCG | U/30/10 | C82R | Gcuucaucuuggcgugcuaaugaug<br>ugug (SEQ ID NO:363)   |  |
| 平铺指导序列<br>30 翻转 8U  | Gluc | UCG | U/30/8  | C82R | Gaacucaucuuggcgugcuaaugaug<br>gaug (SEQ ID NO:364)  |  |
| 平铺指导序列              | Gluc | UCG | U/30/6  | C82R | Gugaacuucuucuuggcgugcua<br>uuga (SEQ ID NO:365)     |  |

|        |                     |      |     |         |      |   |
|--------|---------------------|------|-----|---------|------|---|
|        | 30 翻转 6U            |      |     |         |      |   |
|        | 平铺指导序列<br>30 翻转 4U  | Gluc | UCG | U/30/4  | C82R | Ggaugaacuucuaucuuugggcguc<br>uauu (SEQ ID NO:366)                       |
|        | 平铺指导序列<br>30 翻转 2U  | Gluc | UCG | U/30/2  | C82R | Ggggaugaacuucuaucuuugggcg<br>gcua (SEQ ID NO:367)                       |
|        | 平铺指导序列<br>50 翻转 50U | Gluc | UCG | U/50/50 | C82R | Guaaugauggggacagcagauca<br>cagcccccugggcagccgcuuc<br>(SEQ ID NO:368)    |
|        | 平铺指导序列<br>50 翻转 48U | Gluc | UCG | U/50/48 | C82R | Ggcuaaugauggggacagcagau<br>cagcccccugggcagccgcuu<br>(SEQ ID NO:369)     |
| [2071] | 平铺指导序列<br>50 翻转 46U | Gluc | UCG | U/50/46 | C82R | Ggucuaaugauggggacagcagau<br>cagcccccugggcagccgca<br>(SEQ ID NO:370)     |
|        | 平铺指导序列<br>50 翻转 44U | Gluc | UCG | U/50/44 | C82R | Ggcgucuaaugauggggacagcag<br>aucagcccccugggcagcca<br>(SEQ ID NO:371)     |
|        | 平铺指导序列<br>50 翻转 42U | Gluc | UCG | U/50/42 | C82R | Ggggcgucuaaugauggggacagc<br>agaucagcccccugggcagc<br>(SEQ ID NO:372)     |
|        | 平铺指导序列<br>50 翻转 40U | Gluc | UCG | U/50/40 | C82R | Guuggcgucuaaugauggggacag<br>gcagaucagcccccugggcagca<br>(SEQ ID NO:373)  |
|        | 平铺指导序列<br>50 翻转 38U | Gluc | UCG | U/50/38 | C82R | Gucuugggcgucuaaugauggggac<br>aggcagaucagcccccugggcag<br>(SEQ ID NO:374) |

[2072] 表22:用于内源基因编辑的指导序列

| 名称                                       | 被靶向的基因 | 基序  | 碱基翻转/位置 | 密码子变化 | 间隔子序列   |
|--|--------|-----|---------|-------|---|
| [2073] S33F_CTNNB1_30bp_guide_30_9C_flip | CTNNB1 | UCU | 22      | S33F  | GGGAUCCACAGUCCAGGU<br>AAGACUGUUGCU (SEQ ID<br>NO:375) |

[2074]

|                                     |        |     |    |      |  |
|-------------------------------------|--------|-----|----|------|--|
| H36Y_CTNNB1_30bp_guide_30_9 T flip  | CTNNB1 | CCA | 22 | H36Y | GACCAGAAUUGAUUCCAGAGUCCAGGUAAGA (SEQ ID NO:376)  |
| S37F_CTNNB1_30bp_guide_30_9 T flip  | CTNNB1 | UCU | 22 | S37F | GUGGCACCAUAAUGGAUUC CAGAGUCCAGGU (SEQ ID NO:377) |
| T41I_CTNNB1_30bp_guide_30_11 T flip | CTNNB1 | ACC | 20 | T41I | GAGGAGCUGUGUUAGUGGC ACCAGAAUGGAU (SEQ ID NO:378) |
| P44L_CTNNB1_30bp_guide_30_9 C flip  | CTNNB1 | CCU | 22 | P44L | GUCAGAGAACGAGCUGUGG UAGUGGCACCAG (SEQ ID NO:379) |
| P44S_CTNNB1_30bp_guide_30_11 T flip | CTNNB1 | UCU | 20 | P44S | GCUCAGAGAAGUAGCUGUG GUAGUGGCACCA (SEQ ID NO:380) |
| S45F_CTNNB1_30bp_guide_30_11 C flip | CTNNB1 | UCU | 20 | S45F | GACCACUCAGACAAGGAGC UGUGGUAGUGGC (SEQ ID NO:381) |
| TCG_KRAS_30bp_guide_30_7 T flip     | KRAS   | UCG | 24 | L56L | GUGUGUCUAGAAUUCCAA GAGACAGGUUUC (SEQ ID NO:382)  |
| ACG_KRAS_30bp_guide_30_11 C flip    | KRAS   | ACG | 20 | D30D | GGAUCAUUAUCCUCCACAA AAUGAUUCUGAA (SEQ ID NO:383) |
| GCG_KRAS_30bp_guide_30_11 T flip    | KRAS   | GCG | 20 | G13G | GUCUUGCCUACUCCACCAG CUCCAACUACCA (SEQ ID NO:384) |
| CCT_KRAS_30bp_guide_30_11 C flip    | KRAS   | CCU | 20 | A18A | GGUAUCGUCAACGCACUCU UGCCUACGCCAC (SEQ ID NO:385) |
| TCG_PPIB_30bp_guide_30_11 T flip    | PPIB   | UCG | 20 | I18I | GCGGACCCCGCUAUGAGGG CGGCGCAAGGA (SEQ ID NO:386)  |
| ACG_PPIB_30bp_guide_30_11 C flip    | PPIB   | ACG | 24 | R7C  | GAUAUCCUCCACAAAUG AUUCUGAAUAG (SEQ ID NO:387)    |

[2075]

|                                     |         |     |    |       |  |
|-------------------------------------|---------|-----|----|-------|--|
| 30_7 C flip                         |         |     |    |       | NO:387)  |
| GCG_PPIB_30bp_guide_30_11 T flip    | PPIB    | GCG | 20 | A19V  | GGACGGACCCCUUGAUGAGGGCGGCGCAAG (SEQ ID NO:388)   |
| CCG_PPIB_30bp_guide_30_11 C flip    | PPIB    | CCG | 20 | S21S  | GGGAAGAAGACCGACCCCGCGAUGAGGGCGG (SEQ ID NO:389)  |
| TCG_SMARCA4_30bp_guide_30_9 T flip  | SMARCA4 | UCG | 22 | S85L  | GGGUCGUCCUACAUGCCCUUCUCAUGCAUGG (SEQ ID NO:390)  |
| ACG_SMARCA4_30bp_guide_30_11 T flip | SMARCA4 | ACG | 20 | D86D  | GAGCGCGGGUCUCCGACAUCCCUUCUCAU (SEQ ID NO:391)    |
| GCG_SMARCA4_30bp_guide_30_11 C flip | SMARCA4 | GCG | 20 | R89C  | GUGGUUGUAGCCCGGUCGUCCGACAUGCCC (SEQ ID NO:392)   |
| CCG_SMARCA4_30bp_guide_30_11 T flip | SMARCA4 | CCG | 20 | P88L  | GGUUGUAGCGCUGGUCGUCGCGACAUGCCCUU (SEQ ID NO:393) |
| NRAS_C-flip_guide_30_11             | NRAS    | UCC | 20 | I21I  | GGGAUUAGCUGCAUUGUCAGUGCGUUUCC (SEQ ID NO:394)    |
| NKFB1_T-flip_guide_30_11            | NFKB    | ACC | 20 | P33S  | GGCCAUCUGUGUUUGAAUACUUCUGGAUUA (SEQ ID NO:395)   |
| EZH2_T-flip_guide_30_11             | EZH2    | UCA | 20 | F32F  | GCAGCUCGUCUUAACCUCUUGAGCUGUCUCA (SEQ ID NO:396)  |
| NF2_T-flip_guide_30_7               | NF2     | ACG | 24 | T21M  | GGUGAACUUCUUGGGUUGCUUCCUCUUGAGA (SEQ ID NO:397)  |
| RAF1_T-flip_guide_30_7              | RAF1    | UCC | 24 | P30S  | GUUGUAGUAGAGAUGCAGCUGGAGCCAUCAA (SEQ ID NO:398)  |
| STAT3 Y705C-flip_50_17              | STAT3   | UAC | 34 | Y705C | GAAACUUGGUCUUCAGGCAUGGGCAGCGCUACCUUGG            |

[2076]

|   |        |     |    |       |  |
|---|--------|-----|----|-------|--|
|   |        |     |    |       | UCAGCUUCAGGAU (SEQ ID NO:399)  |
| STAT3 S727<br>C-flip_30_9                       | STAT3  | UCC | 22 | S727F | GUGCGGGGGCACAUCGGCA<br>GGUCAUUGGUAU (SEQ ID<br>NO:400)                       |
| STAT1 Y701<br>C-flip_50_17                      | STAT1  | UAU | 34 | Y705C | GCAACUCAGUCUUGAUACA<br>UCCAGUUCUUUAGGGCCA<br>UCAAGUCCAUUG (SEQ<br>ID NO:401) |
| STAT1 S727<br>C-flip_30_9                       | STAT1  | UCC | 22 | S727F | GCCUCAGGACACAUGGGGA<br>GCAGGUUGUCUG (SEQ ID<br>NO:402)                       |
| S33F_CTNN<br>B1_30bp_gui<br>de_U-flip_30<br>_7  | CTNNB1 | UCU | 24 | S33F  | GAUCCAAGUCCAGGUA<br>GACUGUUGCUGC (SEQ ID<br>NO:403)                          |
| S33F_CTNN<br>B1_30bp_gui<br>de_U-flip_30<br>_9  | CTNNB1 | UCU | 22 | S33F  | GGGAUCCAAGUCCAGGU<br>AAGACUGUUGCU (SEQ ID<br>NO:404)                         |
| S33F_CTNN<br>B1_30bp_gui<br>de_U-flip_30<br>_11 | CTNNB1 | UCU | 20 | S33F  | GAUGGAUCCAAGUCCAG<br>GUAAGACUGUUG (SEQ ID<br>NO:405)                         |
| S33F_CTNN<br>B1_30bp_gui<br>de_U-flip_30<br>_13 | CTNNB1 | UCU | 18 | S33F  | GGAAUGGAUCCAAGUCC<br>AGGUAAGACUGU (SEQ ID<br>NO:406)                         |
| H36Y_CTNN<br>B1_30bp_g<br>uide_U-flip_<br>30_7  | CTNNB1 | CCA | 24 | H36Y  | GCAGAAUUGAUCCAGAGU<br>CCAGGUAAGACU (SEQ ID<br>NO:407)                        |
| H36Y_CTNN<br>B1_30bp_g<br>uide_U-flip_<br>30_9  | CTNNB1 | CCA | 22 | H36Y  | GACCAGAAUUGAUCCAGA<br>GUCCAGGUAAGA (SEQ ID<br>NO:408)                        |
| H36Y_CTNN<br>B1_30bp_g<br>uide_U-flip_<br>30_11 | CTNNB1 | CCA | 20 | H36Y  | GGCACCAGAAUUGAUCCA<br>GAGUCCAGGUA (SEQ ID<br>NO:409)                         |
| H36Y_CTNN<br>B1_30bp_g                          | CTNNB1 | CCA | 18 | H36Y  | GUGGCACCAGAAUUGAUUC<br>CAGAUCCAGGU (SEQ ID                                   |

[2077]

|                                     |        |     |    |      |  |
|-------------------------------------|--------|-----|----|------|--|
| uide_U-flip_30_13                   |        |     |    |      | NO:410)  |
| S37F_CTNNB1_30bp_guide_U-flip_30_7  | CTNNB1 | UCU | 24 | S37F | GGCACCAUAAUGGAUCCA<br>GAGUCCAGGUAA (SEQ ID<br>NO:411)  |
| S37F_CTNNB1_30bp_guide_U-flip_30_9  | CTNNB1 | UCU | 22 | S37F | GUGGCACCAUAAUGGAUUC<br>CAGAGUCCAGGU (SEQ ID<br>NO:412) |
| S37F_CTNNB1_30bp_guide_U-flip_30_11 | CTNNB1 | UCU | 20 | S37F | GAGUGGCACCAUAAUGGAU<br>UCCAGAGUCCAG (SEQ ID<br>NO:413) |
| S37F_CTNNB1_30bp_guide_U-flip_30_13 | CTNNB1 | UCU | 18 | S37F | GGUAGUGGCACCAUAAUGG<br>AUUCCAGAGUCC (SEQ ID<br>NO:414) |
| T41I_CTNNB1_30bp_guide_U-flip_30_7  | CTNNB1 | ACC | 24 | T41I | GGCUGUGUUAGUGGCACCA<br>GAAUGGAUCCA (SEQ ID<br>NO:415)  |
| T41I_CTNNB1_30bp_guide_U-flip_30_9  | CTNNB1 | ACC | 22 | T41I | GGAGCUGUGUUAGUGGCAC<br>CAGAAUGGAUUC (SEQ ID<br>NO:416) |
| T41I_CTNNB1_30bp_guide_U-flip_30_11 | CTNNB1 | ACC | 20 | T41I | GAGGAGCUGUGUUAGUGGC<br>ACCAGAAUGGAU (SEQ ID<br>NO:417) |
| T41I_CTNNB1_30bp_guide_U-flip_30_13 | CTNNB1 | ACC | 18 | T41I | GGAAGGAGCUGUGUUAGUG<br>GCACCAGAAUGG (SEQ ID<br>NO:418) |
| P44L_CTNNB1_30bp_guide_U-flip_30_7  | CTNNB1 | CCU | 24 | P44L | GAGAGAAUGAGCUGUGGUA<br>GUGGCACCAGAA (SEQ ID<br>NO:419) |
| P44L_CTNNB1_30bp_guide_U-flip_30_9  | CTNNB1 | CCU | 22 | P44L | GUCAGAGAAUGAGCUGUGG<br>UAGUGGCACCAG (SEQ ID<br>NO:420) |

[2078]

|   |        |     |    |      |  |
|---|--------|-----|----|------|--|
| P44L_CTNN<br>B1_30bp_guide_U-flip_30_11 | CTNNB1 | CCU | 20 | P44L | GACUCAGAGAAUGAGCUGUGGUAGUGGCACC (SEQ ID NO:421)  |
| P44L_CTNN<br>B1_30bp_guide_U-flip_30_13 | CTNNB1 | CCU | 18 | P44L | GCCACUCAGAGAAUGAGCUGUGGUAGUGGCA (SEQ ID NO:422)  |
| P44S_CTNN<br>B1_30bp_guide_U-flip_30_7  | CTNNB1 | UCU | 24 | P44S | GGAGAAGUAGCUGUGGUAGUGGCACCAGAAU (SEQ ID NO:423)  |
| P44S_CTNN<br>B1_30bp_guide_U-flip_30_9  | CTNNB1 | UCU | 22 | P44S | GCAGAGAAGUAGCUGUGGUAGUGGCACCAGA (SEQ ID NO:424)  |
| P44S_CTNN<br>B1_30bp_guide_U-flip_30_11 | CTNNB1 | UCU | 20 | P44S | GCUCAGAGAAGUAGCUGUGGUAGUGGCACCA (SEQ ID NO:425)  |
| P44S_CTNN<br>B1_30bp_guide_U-flip_30_13 | CTNNB1 | UCU | 18 | P44S | GCACUCAGAGAAGUAGCUGUGGUAGUGGCAC (SEQ ID NO:426)  |
| S45F_CTNN<br>B1_30bp_guide_U-flip_30_7  | CTNNB1 | UCU | 24 | S45F | GCUCAGAUAAAGGAGCUGUGGUAGUGGCACCA (SEQ ID NO:427) |
| S45F_CTNN<br>B1_30bp_guide_U-flip_30_9  | CTNNB1 | UCU | 22 | S45F | GCACUCAGAUAAAGGAGCUGUGGUAGUGGCAC (SEQ ID NO:428) |
| S45F_CTNN<br>B1_30bp_guide_U-flip_30_11 | CTNNB1 | UCU | 20 | S45F | GACCACUCAGAUAAAGGAGCUGUGGUAGUGGC (SEQ ID NO:429) |
| S45F_CTNN<br>B1_30bp_guide_U-flip_30_13 | CTNNB1 | UCU | 18 | S45F | GUUACCACUCAGAUAAAGGAGCUGUGGUAGUG (SEQ ID NO:430) |
| TCG_KRAS_30bp_guide                     | KRAS   | UCG | 24 | L56L | GUGUGUCUAGAAUUCCAA GAGACAGGUUUC (SEQ ID NO:431)  |

[2079]

|  |      |     |    |      |  |
|--|------|-----|----|------|--|
| _U-flip_30_7                                 |      |     |    |      | NO:431)  |
| TCG_KRAS<br>_30bp_guide<br>_U-flip_30_9      | KRAS | UCG | 22 | L56L | GGCUGUGUCUAGAAUAUCC<br>AAGAGACAGGUU (SEQ ID<br>NO:432) |
| TCG_KRAS<br>_30bp_guide<br>_U-flip_30_1<br>1 | KRAS | UCG | 20 | L56L | GCUGCUGUGUCUAGAAUAU<br>CCAAGAGACAGG (SEQ ID<br>NO:433) |
| TCG_KRAS<br>_30bp_guide<br>_U-flip_30_1<br>3 | KRAS | UCG | 18 | L56L | GACCUGCUGUGUCUAGAAU<br>AUCCAAGAGACA (SEQ ID<br>NO:434) |
| ACG_KRAS<br>_30bp_guide<br>_U-flip_30_7      | KRAS | ACG | 24 | D30D | GAUAUUCUCCACAAAUG<br>AUUCUGAAUAG (SEQ ID<br>NO:435)    |
| ACG_KRAS<br>_30bp_guide<br>_U-flip_30_9      | KRAS | ACG | 22 | D30D | GUCAUAUUCUCCACAAA<br>UGAUUCUGAAU (SEQ ID<br>NO:436)    |
| ACG_KRAS<br>_30bp_guide<br>_U-flip_30_1<br>1 | KRAS | ACG | 20 | D30D | GGAUCAUAUUCUCCACAA<br>AAUGAUUCUGAA (SEQ ID<br>NO:437 ) |
| ACG_KRAS<br>_30bp_guide<br>_U-flip_30_1<br>3 | KRAS | ACG | 18 | D30D | GUGGAUCAUAUUCUCCAC<br>AAAUGAUUCUG (SEQ ID<br>NO:438)   |
| GCG_KRAS<br>_30bp_guide<br>_U-flip_30_7      | KRAS | GCG | 24 | G13G | GGCCUACUCCACCAGCUCCA<br>ACUACCACAAG (SEQ ID<br>NO:439) |
| GCG_KRAS<br>_30bp_guide<br>_U-flip_30_9      | KRAS | GCG | 22 | G13G | GUUGCCUACUCCACCAGCU<br>CCAACUACCACA (SEQ ID<br>NO:440) |
| GCG_KRAS<br>_30bp_guide<br>_U-flip_30_1<br>1 | KRAS | GCG | 20 | G13G | GUCUUGCCUACUCCACCAG<br>CUCCAACUACCA (SEQ ID<br>NO:441) |
| GCG_KRAS<br>_30bp_guide<br>_U-flip_30_1<br>3 | KRAS | GCG | 18 | G13G | GACUCUUGCCUACUCCACC<br>AGCUCCAACUAC (SEQ ID<br>NO:442) |

[2080]

|                                  |      |     |    |      |  |
|----------------------------------|------|-----|----|------|--|
| CCT_KRAS_30bp_guide_U-flip_30_7  | KRAS | CCU | 24 | A18A | GCGUCA AUGCACUCUUGCCUACGCCACCAGC (SEQ ID NO:443) |
| CCT_KRAS_30bp_guide_U-flip_30_9  | KRAS | CCU | 22 | A18A | GAUCGUCAAUGCACUCUUGCCUACGCCACCA (SEQ ID NO:444)  |
| CCT_KRAS_30bp_guide_U-flip_30_11 | KRAS | CCU | 20 | A18A | GGUAUCGUCAAUGCACUCUUGCCUACGCCAC (SEQ ID NO:445)  |
| CCT_KRAS_30bp_guide_U-flip_30_13 | KRAS | CCU | 18 | A18A | GCUGUAUCGUCAAUGCACUCUUGCCUACGCC (SEQ ID NO:446)  |
| TCG_PPIB_30bp_guide_U-flip_30_7  | PPIB | UCG | 24 | I18I | GCCCCGCUAUGAGGGCGGC GGCAAGGAGCAC (SEQ ID NO:447) |
| TCG_PPIB_30bp_guide_U-flip_30_9  | PPIB | UCG | 22 | I18I | GGACCCCGCUAUGAGGGCG GCGCAAGGAGC (SEQ ID NO:448)  |
| TCG_PPIB_30bp_guide_U-flip_30_11 | PPIB | UCG | 20 | I18I | GCGGACCCCGCUAUGAGGG CGGCGCAAGGA (SEQ ID NO:449)  |
| TCG_PPIB_30bp_guide_U-flip_30_13 | PPIB | UCG | 18 | I18I | GGACGGACCCCGCUAUGAG GGCGGCGCAAG (SEQ ID NO:450)  |
| ACG_PPIB_30bp_guide_U-flip_30_7  | PPIB | ACG | 24 | R7C  | GUGUUGCUUUCGGAGAGGC GCAGCAUCCACA (SEQ ID NO:451) |
| ACG_PPIB_30bp_guide_U-flip_30_9  | PPIB | ACG | 22 | R7C  | GCAUGUUGCUUUCGGAGAG GCGCAGCAUCCA (SEQ ID NO:452) |
| ACG_PPIB_30bp_guide_U-flip_30_11 | PPIB | ACG | 20 | R7C  | GUUCAUGUUGCUUUCGGAG AGGCGCAGCAUC (SEQ ID NO:453) |
| ACG_PPIB_30bp_guide_U-flip_30_13 | PPIB | ACG | 18 | R7C  | GCCUUCAUGUUGCUUUCGG AGAGGCGCAGCA (SEQ ID NO:454) |
| GCG_PPIB_30bp_guide_U-flip_30_7  | PPIB | GCG | 24 | A19V | GGACCCCUCAUGAGGGCG GCGCAAGGAGC (SEQ ID NO:455)   |

[2081]

|                                     |         |     |    |      |  |
|-------------------------------------|---------|-----|----|------|--|
| U-flip_30_7                         |         |     |    |      | NO:455)  |
| GCG_PPIB_30bp_guide_U-flip_30_9     | PPIB    | GCG | 22 | A19V | GCGGACCCUCGAUGAGGGCGGCGCAAGGA (SEQ ID NO:456)    |
| GCG_PPIB_30bp_guide_U-flip_30_11    | PPIB    | GCG | 20 | A19V | GGACGGACCCUCGAUGAGGGCGGCGCAAG (SEQ ID NO:457)    |
| GCG_PPIB_30bp_guide_U-flip_30_13    | PPIB    | GCG | 18 | A19V | GAAGACGGACCCUCGAUGAGGGCGGCGCA (SEQ ID NO:458)    |
| CCG_PPIB_30bp_guide_U-flip_30_7     | PPIB    | CCG | 24 | S21S | GGAAGACUGACCCCGCAUGAGGGCGGCGG (SEQ ID NO:459)    |
| CCG_PPIB_30bp_guide_U-flip_30_9     | PPIB    | CCG | 22 | S21S | GAAGAAGACUGACCCCGCAUGAGGGCGGCGG (SEQ ID NO:460)  |
| CCG_PPIB_30bp_guide_U-flip_30_11    | PPIB    | CCG | 20 | S21S | GGGAAGAAGACUGACCCCGCAUGAGGGCGG (SEQ ID NO:461)   |
| CCG_PPIB_30bp_guide_U-flip_30_13    | PPIB    | CCG | 18 | S21S | GCAGGAAGAAGACUGACCCCGCAUGAGGGCGG (SEQ ID NO:462) |
| TCG_SMARCA4_30bp_guide_U-flip_30_7  | SMARCA4 | UCG | 24 | S85L | GUCGUCCUACAUGCCCUUCAUGCAUGGAC (SEQ ID NO:463)    |
| TCG_SMARCA4_30bp_guide_U-flip_30_9  | SMARCA4 | UCG | 22 | S85L | GGGUCGUCCUACAUGCCCUUCAUGCAUGG (SEQ ID NO:464)    |
| TCG_SMARCA4_30bp_guide_U-flip_30_11 | SMARCA4 | UCG | 20 | S85L | GCGGGUCGUCCUACAUGCCCUUCAUGCAU (SEQ ID NO:465)    |
| TCG_SMARCA4_30bp_guide_U-flip_30_13 | SMARCA4 | UCG | 18 | S85L | GCGCGGGUCGUCCUACAUGCCCUUCAUGC (SEQ ID NO:466)    |

[2082]

|   |             |     |    |      |  |
|---|-------------|-----|----|------|--|
| ACG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_7  | SMARCA<br>4 | ACG | 24 | D86D | GCGGGUCUCCGACAUGCC<br>CUUCUCAUGCAU (SEQ ID<br>NO:467)  |
| ACG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_9  | SMARCA<br>4 | ACG | 22 | D86D | GCGCGGGUCUCCGACAUG<br>CCCUUCUCAUGC (SEQ ID<br>NO:468)  |
| ACG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_11 | SMARCA<br>4 | ACG | 20 | D86D | GAGCGCGGGUCUCCGACA<br>UGCCCUUCUCAU (SEQ ID<br>NO:469)  |
| ACG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_13 | SMARCA<br>4 | ACG | 18 | D86D | GGUAGCGGGUCUCCGA<br>CAUGCCCUUCUC (SEQ ID<br>NO:470)    |
| GCG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_7  | SMARCA<br>4 | GCG | 24 | R89C | GUGUAGCUCGGGUCGUCCG<br>ACAUGCCCUUCU (SEQ ID<br>NO:471) |
| GCG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_9  | SMARCA<br>4 | GCG | 22 | R89C | GGUUGUAGCUCGGGUCGUC<br>CGACAUGCCCUU (SEQ ID<br>NO:472) |
| GCG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_11 | SMARCA<br>4 | GCG | 20 | R89C | GUGGUUGUAGCUCGGGUCG<br>UCCGACAUGCCC (SEQ ID<br>NO:473) |
| GCG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_13 | SMARCA<br>4 | GCG | 18 | R89C | GUCUGGUUGUAGCUCGGGU<br>CGUCCGACAUGC (SEQ ID<br>NO:474) |
| CCG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_7  | SMARCA<br>4 | CCG | 24 | P88L | GUAGCGCUGGUCGUCCGAC<br>AUGCCCUUCUCA (SEQ ID<br>NO:475) |
| CCG_SMAR<br>CA4_30bp_g<br>uide_U-flip_<br>30_9  | SMARCA<br>4 | CCG | 22 | P88L | GUGUAGCGCUGGUCGUCCG<br>ACAUGCCCUUCU (SEQ ID<br>NO:476) |
| CCG_SMAR<br>CA4_30bp_g                          | SMARCA<br>4 | CCG | 20 | P88L | GGUUGUAGCGCUGGUCGUC<br>CGACAUGCCCUU (SEQ ID            |

[2083]

|                                     |         |     |    |      |   |
|-------------------------------------|---------|-----|----|------|---|
| uide_U-flip_30_11                   |         |     |    |      | NO:477)   |
| CCG_SMARCA4_30bp_guide_U-flip_30_13 | SMARCA4 | CCG | 18 | P88L | GUGGUUGUAGCGCUGGUCG UCCGACAUGCCC (SEQ ID NO:478)    |
| S33F_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | UCU | 24 | S33F | GAUCCACAGUCCAGGUAA GACUGUUGCUGC (SEQ ID NO:479)     |
| S33F_CTNNB1_30bp_C-flip_guide_30_9  | CTNNB1  | UCU | 22 | S33F | GGGAUCCACAGUCCAGGU AAGACUGUUGCU (SEQ ID NO:480)     |
| S33F_CTNNB1_30bp_C-flip_guide_30_11 | CTNNB1  | UCU | 20 | S33F | GAUGGAUCCACAGUCCAG GUAAGACUGUUG (SEQ ID NO:481)     |
| S33F_CTNNB1_30bp_C-flip_guide_30_13 | CTNNB1  | UCU | 18 | S33F | GGAAUGGAUCCACAGUCC AGGUAAGACUGU (SEQ ID NO:482)     |
| H36Y_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | CCA | 24 | H36Y | GCAGAAUCGAUCCAGAGU CCAGGUAAGACU (SEQ ID NO:483)     |
| H36Y_CTNNB1_30bp_C-flip_guide_30_9  | CTNNB1  | CCA | 22 | H36Y | GACCAGAAUCGAUCCAGAGU CCAGGUAAGA (SEQ ID NO:484)     |
| H36Y_CTNNB1_30bp_C-flip_guide_30_11 | CTNNB1  | CCA | 20 | H36Y | GGCACCAGAAUCGAUCCAGAGU CCAGGUAAGA (SEQ ID NO:485)   |
| H36Y_CTNNB1_30bp_C-flip_guide_30_13 | CTNNB1  | CCA | 18 | H36Y | GUGGCACCAGAAUCGAUCCAGAGU CCAGGUAAGA (SEQ ID NO:486) |
| S37F_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | UCU | 24 | S37F | GGCACCACAAUGGAUCCAGAGU CCAGGUAAGA (SEQ ID NO:487)   |

[2084]

|   |        |     |    |      |  |
|---|--------|-----|----|------|--|
| S37F_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_9  | CTNNB1 | UCU | 22 | S37F | GUGGCACCACAAUGGAUUC<br>CAGAGUCCAGGU (SEQ ID<br>NO:488) |
| S37F_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_11 | CTNNB1 | UCU | 20 | S37F | GAGUGGCACCACAAUGGAU<br>UCCAGAGUCCAG (SEQ ID<br>NO:489) |
| S37F_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_13 | CTNNB1 | UCU | 18 | S37F | GGUAGUGGCACCACAAUGG<br>AUCCAGAGUCC (SEQ ID<br>NO:490)  |
| T41I_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_7  | CTNNB1 | ACC | 24 | T41I | GGCUGUGCUAGUGGCACCA<br>GAAUGGAUCCA (SEQ ID<br>NO:491)  |
| T41I_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_9  | CTNNB1 | ACC | 22 | T41I | GGAGCUGUGCUAGUGGCAC<br>CAGAAUGGAUUC (SEQ ID<br>NO:492) |
| T41I_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_11 | CTNNB1 | ACC | 20 | T41I | GAGGAGCUGUGCUAGUGGC<br>ACCAGAAUGGAU (SEQ ID<br>NO:493) |
| T41I_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_13 | CTNNB1 | ACC | 18 | T41I | GGAAGGAGCUGUGCUAGUG<br>GCACCAGAAUGG (SEQ ID<br>NO:494) |
| P44L_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_7  | CTNNB1 | CCU | 24 | P44L | GAGAGAACGAGCUGUGGUA<br>GUGGCACCAGAA (SEQ ID<br>NO:495) |
| P44L_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_9  | CTNNB1 | CCU | 22 | P44L | GUCAGAGAACGAGCUGUGG<br>UAGUGGCACCAG (SEQ ID<br>NO:496) |
| P44L_CTNN<br>B1_30bp_C-f<br>lip_guide_30<br>_11 | CTNNB1 | CCU | 20 | P44L | GACUCAGAGAACGAGCUGU<br>GGUAGUGGCACC (SEQ ID<br>NO:497) |
| P44L_CTNN<br>B1_30bp_C-f                        | CTNNB1 | CCU | 18 | P44L | GCCACUCAGAGAACGAGCU<br>GUGGUAGUGGCA (SEQ ID            |

[2085]

|   |        |     |    |      |  |
|---|--------|-----|----|------|--|
| lip_guide_30_13                             |        |     |    |      | NO:498)  |
| P44S_CTNN<br>B1_30bp_C-f<br>lip_guide_30_7  | CTNNB1 | UCU | 24 | P44S | GGAGAAGCAGCUGUGGUAG<br>UGGCACCAGAAU (SEQ ID<br>NO:499) |
| P44S_CTNN<br>B1_30bp_C-f<br>lip_guide_30_9  | CTNNB1 | UCU | 22 | P44S | GCAGAGAAGCAGCUGUGGU<br>AGUGGCACCAGA (SEQ ID<br>NO:500) |
| P44S_CTNN<br>B1_30bp_C-f<br>lip_guide_30_11 | CTNNB1 | UCU | 20 | P44S | GCUCAGAGAAGCAGCUGUG<br>GUAGUGGCACCA (SEQ ID<br>NO:501) |
| P44S_CTNN<br>B1_30bp_C-f<br>lip_guide_30_13 | CTNNB1 | UCU | 18 | P44S | GCACUCAGAGAAGCAGCUG<br>UGGUAGUGGCAC (SEQ ID<br>NO:502) |
| S45F_CTNN<br>B1_30bp_C-f<br>lip_guide_30_7  | CTNNB1 | UCU | 24 | S45F | GCUCAGACAAGGAGCUGUG<br>GUAGUGGCACCA (SEQ ID<br>NO:503) |
| S45F_CTNN<br>B1_30bp_C-f<br>lip_guide_30_9  | CTNNB1 | UCU | 22 | S45F | GCACUCAGACAAGGAGCUG<br>UGGUAGUGGCAC (SEQ ID<br>NO:504) |
| S45F_CTNN<br>B1_30bp_C-f<br>lip_guide_30_11 | CTNNB1 | UCU | 20 | S45F | GACCACUCAGACAAGGAGC<br>UGUGGUAGUGGC (SEQ ID<br>NO:505) |
| S45F_CTNN<br>B1_30bp_C-f<br>lip_guide_30_13 | CTNNB1 | UCU | 18 | S45F | GUUACCACUCAGACAAGGA<br>GCUGUGGUAGUG (SEQ ID<br>NO:506) |
| TCG_KRAS<br>_30bp_C-flip<br>_guide_30_7     | KRAS   | UCG | 24 | L56L | GUGUGUCCAGAAUAUCCAA<br>GAGACAGGUUC (SEQ ID<br>NO:507)  |
| TCG_KRAS<br>_30bp_C-flip<br>_guide_30_9     | KRAS   | UCG | 22 | L56L | GGCUGUGUCCAGAAUAUCC<br>AAGAGACAGGUU (SEQ ID<br>NO:508) |

[2086]

|  |      |     |    |      |  |
|--|------|-----|----|------|--|
| TCG_KRAS<br>_30bp_C-flip<br>_guide_30_1<br>1 | KRAS | UCG | 20 | L56L | GCUGCUGUGUCCAGAAU<br>CCAAGAGACAGG (SEQ ID<br>NO:509)   |
| TCG_KRAS<br>_30bp_C-flip<br>_guide_30_1<br>3 | KRAS | UCG | 18 | L56L | GACCUGCUGUGUCCAGAAU<br>AUCCAAGAGACA (SEQ ID<br>NO:510) |
| ACG_KRAS<br>_30bp_C-flip<br>_guide_30_7      | KRAS | ACG | 24 | D30D | GAUAAUCCUCCACAAAUG<br>AUUCUGAAUAG (SEQ ID<br>NO:511)   |
| ACG_KRAS<br>_30bp_C-flip<br>_guide_30_9      | KRAS | ACG | 22 | D30D | GUCAUAAUCCUCCACAAA<br>UGAUUCUGAAU (SEQ ID<br>NO:512)   |
| ACG_KRAS<br>_30bp_C-flip<br>_guide_30_1<br>1 | KRAS | ACG | 20 | D30D | GGAUCAUAAUCCUCCACAA<br>AAUGAUUCUGAA (SEQ ID<br>NO:513) |
| ACG_KRAS<br>_30bp_C-flip<br>_guide_30_1<br>3 | KRAS | ACG | 18 | D30D | GUGGAUCAUAAUCCUCCAC<br>AAAUGAUUCUG (SEQ ID<br>NO:514)  |
| GCG_KRAS<br>_30bp_C-flip<br>_guide_30_7      | KRAS | GCG | 24 | G13G | GGCCUACCCACCAGCUCCA<br>ACUACCACAAG (SEQ ID<br>NO:515)  |
| GCG_KRAS<br>_30bp_C-flip<br>_guide_30_9      | KRAS | GCG | 22 | G13G | GUUGCCUACCCACCAGCUC<br>CAACUACCACA (SEQ ID<br>NO:516)  |
| GCG_KRAS<br>_30bp_C-flip<br>_guide_30_1<br>1 | KRAS | GCG | 20 | G13G | GUCUUGCCUACCCACCAGC<br>UCCAACUACCA (SEQ ID<br>NO:517)  |
| GCG_KRAS<br>_30bp_C-flip<br>_guide_30_1<br>3 | KRAS | GCG | 18 | G13G | GACUCUUGCCUACCCACCA<br>GCUCCAACUAC (SEQ ID<br>NO:518)  |
| CCT_KRAS<br>_30bp_C-flip<br>_guide_30_7      | KRAS | CCU | 24 | A18A | GCGUCAACGCACUCUUGCC<br>UACGCCACCAGC (SEQ ID<br>NO:519) |
| CCT_KRAS<br>_30bp_C-flip                     | KRAS | CCU | 22 | A18A | GAUCGUCAACGCACUCUUG<br>CCUACGCCACCA (SEQ ID            |

[2087]

|                                  |      |     |    |      |   |
|----------------------------------|------|-----|----|------|---|
| _guide_30_9                      |      |     |    |      | NO:520)   |
| CCT_KRAS_30bp_C-flip_guide_30_11 | KRAS | CCU | 20 | A18A | GGUAUCGUCAACGCACUCUUGCCUACGCCAC (SEQ ID NO:521)   |
| CCT_KRAS_30bp_C-flip_guide_30_13 | KRAS | CCU | 18 | A18A | GCUGUAUCGUCAACGCACUCUUGCCUACGCC (SEQ ID NO:522)   |
| TCG_PPIB_30bp_C-flip_guide_30_7  | PPIB | UCG | 24 | I18I | GCCCCGCCAUGAGGGCGGGCGCAAGGAGCAC (SEQ ID NO:523)   |
| TCG_PPIB_30bp_C-flip_guide_30_9  | PPIB | UCG | 22 | I18I | GGACCCCGCCAUGAGGGCGGCGCAAGGAGC (SEQ ID NO:524)    |
| TCG_PPIB_30bp_C-flip_guide_30_11 | PPIB | UCG | 20 | I18I | GCGGACCCCGCCAUGAGGGCGGCGGCAAGGA (SEQ ID NO:525)   |
| TCG_PPIB_30bp_C-flip_guide_30_13 | PPIB | UCG | 18 | I18I | GGACGGACCCCGCCAUGAGGGCGGCGGCGCAAG (SEQ ID NO:526) |
| ACG_PPIB_30bp_C-flip_guide_30_7  | PPIB | ACG | 24 | R7C  | GUGUUGCCUUCGGAGAGGCGCAGCAUCCACA (SEQ ID NO:527)   |
| ACG_PPIB_30bp_C-flip_guide_30_9  | PPIB | ACG | 22 | R7C  | GCAUGUUGCCUUCGGAGAGGCGCAGCAUCCA (SEQ ID NO:528)   |
| ACG_PPIB_30bp_C-flip_guide_30_11 | PPIB | ACG | 20 | R7C  | GUUCAUGUUGCCUUCGGAGAGGCGCAGCAUC (SEQ ID NO:529)   |
| ACG_PPIB_30bp_C-flip_guide_30_13 | PPIB | ACG | 18 | R7C  | GCCUUCAUGUUGCCUUCGGAGAGGCGCAGCA (SEQ ID NO:530)   |
| GCG_PPIB_30bp_C-flip_guide_30_7  | PPIB | GCG | 24 | A19V | GGACCCCCCGAUGAGGGCGGCGGCAAGGAGC (SEQ ID NO:531)   |
| GCG_PPIB_30bp_C-flip_guide_30_9  | PPIB | GCG | 22 | A19V | GCGGACCCCCCGAUGAGGGCGGCGGCAAGGA (SEQ ID NO:532)   |

[2088]

|                                     |         |     |    |      |   |
|-------------------------------------|---------|-----|----|------|---|
| GCG_PPIB_30bp_C-flip_guide_30_11    | PPIB    | GCG | 20 | A19V | GGACGGACCCCCGAUGAGGGCGGCGCAAG (SEQ ID NO:533)   |
| GCG_PPIB_30bp_C-flip_guide_30_13    | PPIB    | GCG | 18 | A19V | GAAGACGGACCCCCGAUGAGGGCGGCGCA (SEQ ID NO:534)   |
| CCG_PPIB_30bp_C-flip_guide_30_7     | PPIB    | CCG | 24 | S21S | GGAAGACCGACCCCCGGAUGAGGGCGGCGG (SEQ ID NO:535)  |
| CCG_PPIB_30bp_C-flip_guide_30_9     | PPIB    | CCG | 22 | S21S | GAAGAAGACCGACCCCCGGAUGAGGGCGGCG (SEQ ID NO:536) |
| CCG_PPIB_30bp_C-flip_guide_30_11    | PPIB    | CCG | 20 | S21S | GGGAAGAAGACCGACCCCCGGAUGAGGGCGG (SEQ ID NO:537) |
| CCG_PPIB_30bp_C-flip_guide_30_13    | PPIB    | CCG | 18 | S21S | GCAGGAAGAAGACCGACCCCGGAUGAGGGCG (SEQ ID NO:538) |
| TCG_SMARCA4_30bp_C-flip_guide_30_7  | SMARCA4 | UCG | 24 | S85L | GUCGUCCACAUGCCCUUCAUGCAUGGAC (SEQ ID NO:539)    |
| TCG_SMARCA4_30bp_C-flip_guide_30_9  | SMARCA4 | UCG | 22 | S85L | GGGUCGUCCACAUGCCCUUCAUGCAUGG (SEQ ID NO:540)    |
| TCG_SMARCA4_30bp_C-flip_guide_30_11 | SMARCA4 | UCG | 20 | S85L | GCGGGUCGUCCACAUGCCCUUCAUGCAU (SEQ ID NO:541)    |
| TCG_SMARCA4_30bp_C-flip_guide_30_13 | SMARCA4 | UCG | 18 | S85L | GCGCGGGUCGUCCACAUGCCCUUCAUGCAU (SEQ ID NO:542)  |
| ACG_SMARCA4_30bp_C-flip_guide_30_7  | SMARCA4 | ACG | 24 | D86D | GCGGGUCCUCCGACAUGCCCUUCAUGCAU (SEQ ID NO:543)   |
| ACG_SMARCA4_30bp_C-flip_guide_30_9  | SMARCA4 | ACG | 22 | D86D | GCGCGGGUCCUCCGACAUGCCCUUCAUGCAU (SEQ ID NO:544) |

[2089]

|                                     |         |     |    |      |   |
|-------------------------------------|---------|-----|----|------|---|
| 0_9                                 |         |     |    |      |   |
| ACG_SMARCA4_30bp_C-flip_guide_30_11 | SMARCA4 | ACG | 20 | D86D | GAGCGCGGGUCCCGACAUGCCCUUCUCAU (SEQ ID NO:545)   |
| ACG_SMARCA4_30bp_C-flip_guide_30_13 | SMARCA4 | ACG | 18 | D86D | GGUAGCGCGGGUCCCGACAUGCCCUUCUC (SEQ ID NO:546)   |
| GCG_SMARCA4_30bp_C-flip_guide_30_7  | SMARCA4 | GCG | 24 | R89C | GUGUAGCCCGGGUCGUCCGACAUGCCCUUCU (SEQ ID NO:547) |
| GCG_SMARCA4_30bp_C-flip_guide_30_9  | SMARCA4 | GCG | 22 | R89C | GGUUGUAGCCCGGGUCGUCGACAUGCCCUU (SEQ ID NO:548)  |
| GCG_SMARCA4_30bp_C-flip_guide_30_11 | SMARCA4 | GCG | 20 | R89C | GUGGUUGUAGCCCGGGUCGUCCGACAUGCCC (SEQ ID NO:549) |
| GCG_SMARCA4_30bp_C-flip_guide_30_13 | SMARCA4 | GCG | 18 | R89C | GUCUGGUUGUAGCCCGGGUCGUCCGACAUGC (SEQ ID NO:550) |
| CCG_SMARCA4_30bp_C-flip_guide_30_7  | SMARCA4 | CCG | 24 | P88L | GUAGCGCCGGUCGUCCGACAUGCCCUUCA (SEQ ID NO:551)   |
| CCG_SMARCA4_30bp_C-flip_guide_30_9  | SMARCA4 | CCG | 22 | P88L | GUGUAGCGCCGGUCGUCCGACAUGCCCUUCU (SEQ ID NO:552) |
| CCG_SMARCA4_30bp_C-flip_guide_30_11 | SMARCA4 | CCG | 20 | P88L | GGUUGUAGCGCCGGUCGUCGACAUGCCCUU (SEQ ID NO:553)  |
| CCG_SMARCA4_30bp_C-flip_guide_30_13 | SMARCA4 | CCG | 18 | P88L | GUGGUUGUAGCGCCGGUCGUCCGACAUGCCC (SEQ ID NO:554) |

[2090]

|                                       |       |     |    |      |  |
|---------------------------------------|-------|-----|----|------|--|
| NRAS_30bp<br>_C-flip_guide<br>_30_3   | NRAS  | UCC | 28 | I21I | GUGCAUUGUCAGUGCGCUU<br>UUCCCAACACCA (SEQ ID<br>NO:555) |
| NRAS_30bp<br>_C-flip_guide<br>_30_5   | NRAS  | UCC | 26 | I21I | GGCUGCAUUGUCAGUGCGC<br>UUUCCCAACAC (SEQ ID<br>NO:556)  |
| NRAS_30bp<br>_C-flip_guide<br>_30_7   | NRAS  | UCC | 24 | I21I | GUAGCUGCAUUGUCAGUGC<br>GCUUUCCCAAC (SEQ ID<br>NO:557)  |
| NRAS_30bp<br>_C-flip_guide<br>_30_9   | NRAS  | UCC | 22 | I21I | GAUUAGCUGCAUUGUCAGU<br>GCGCUUUCCCA (SEQ ID<br>NO:558)  |
| NRAS_30bp<br>_C-flip_guide<br>_30_11  | NRAS  | UCC | 20 | I21I | GGGAUUAGCUGCAUUGUCA<br>GUGCGCUUUUCC (SEQ ID<br>NO:559) |
| NKFB1_30b<br>p_C-flip_gui<br>de_30_3  | NKFB1 | ACC | 28 | P33S | GUGCUUGAAAUACUUCUGG<br>AUUAAAUAUUGU (SEQ ID<br>NO:560) |
| NKFB1_30b<br>p_C-flip_gui<br>de_30_5  | NKFB1 | ACC | 26 | P33S | GUGUGCUUGAAAUACUUCU<br>GGAUUAAAUAU (SEQ ID<br>NO:561)  |
| NKFB1_30b<br>p_C-flip_gui<br>de_30_7  | NKFB1 | ACC | 24 | P33S | GUCUGUGCUUGAAAUACU<br>CUGGAUAAAUA (SEQ ID<br>NO:562)   |
| NKFB1_30b<br>p_C-flip_gui<br>de_30_9  | NKFB1 | ACC | 22 | P33S | GCAUCUGUGCUUGAAAUAC<br>UUCUGGAUAAA (SEQ ID<br>NO:563)  |
| NKFB1_30b<br>p_C-flip_gui<br>de_30_11 | NKFB1 | ACC | 20 | P33S | GGCCAUCUGUGCUUGAAAU<br>ACUUCUGGAUUA (SEQ ID<br>NO:564) |
| EZH2_30bp<br>C-flip_guide_<br>30_3    | EZH2  | UCA | 28 | F32F | GCUCAACCUCUUGAGCUGU<br>CUCAGUCGAUG (SEQ ID<br>NO:565)  |
| EZH2_30bp<br>C-flip_guide_<br>30_5    | EZH2  | UCA | 26 | F32F | GGUCUCAACCUCUUGAGCU<br>GUCUCAGUCGA (SEQ ID<br>NO:566)  |
| EZH2_30bp<br>C-flip_guide_<br>30_7    | EZH2  | UCA | 24 | F32F | GUCGUCUCAACCUCUUGAG<br>CUGUCUCAGUCG (SEQ ID<br>NO:567) |

[2091]

|                              |      |     |    |      |  |
|------------------------------|------|-----|----|------|--|
| EZH2_30bp_C-flip_guide_30_9  | EZH2 | UCA | 22 | F32F | GGCUCGUCUCAACCUCUUG<br>AGCUGUCUCAGU (SEQ ID<br>NO:568) |
| EZH2_30bp_C-flip_guide_30_11 | EZH2 | UCA | 20 | F32F | GCAGCUCGUCUCAACCUCU<br>UGAGCUGUCUCA (SEQ ID<br>NO:569) |
| NF2_30bp_C-flip_guide_30_3   | NF2  | ACG | 28 | T21M | GACCUCUUGGGUUGCUUCC<br>UCUUGAGAGAGC (SEQ ID<br>NO:570) |
| NF2_30bp_C-flip_guide_30_5   | NF2  | ACG | 26 | T21M | GGAACCUCUUGGGUUGCUU<br>CCUCUUGAGAGA (SEQ ID<br>NO:571) |
| NF2_30bp_C-flip_guide_30_7   | NF2  | ACG | 24 | T21M | GGUGAACCUCUUGGGUUGC<br>UUCUCUUGAGA (SEQ ID<br>NO:572)  |
| NF2_30bp_C-flip_guide_30_9   | NF2  | ACG | 22 | T21M | GCGGUGAACCUCUUGGGUU<br>GCUUCCUCUUGA (SEQ ID<br>NO:573) |
| NF2_30bp_C-flip_guide_30_11  | NF2  | ACG | 20 | T21M | GCACGGUGAACCUCUUGGG<br>UUGCUUCCUCU (SEQ ID<br>NO:574)  |
| RAF1_30bp_C-flip_guide_30_3  | RAF1 | UCC | 28 | P30S | GAGCAGAGAUGCAGCUGGA<br>GCCAUCAAACAC (SEQ ID<br>NO:575) |
| RAF1_30bp_C-flip_guide_30_5  | RAF1 | UCC | 26 | P30S | GGUAGCAGAGAUGCAGCUG<br>GAGCCAUCAAAC (SEQ ID<br>NO:576) |
| RAF1_30bp_C-flip_guide_30_7  | RAF1 | UCC | 24 | P30S | GUUGUAGCAGAGAUGCAGC<br>UGGAGCCAUCAA (SEQ ID<br>NO:577) |
| RAF1_30bp_C-flip_guide_30_9  | RAF1 | UCC | 22 | P30S | GUAUUGUAGCAGAGAUGCA<br>GCUGGAGCCAU (SEQ ID<br>NO:578)  |
| RAF1_30bp_C-flip_guide_30_11 | RAF1 | UCC | 20 | P30S | GACUAUUGUAGCAGAGAUG<br>CAGCUGGAGCCA (SEQ ID<br>NO:579) |
| NRAS_30bp_T-flip_guide_30_3  | NRAS | UCC | 28 | I21I | GUGUAUUGUCAGUGCGCUU<br>UCCCAACACCA (SEQ ID<br>NO:580)  |

[2092]

|                                       |       |     |    |      |  |
|---------------------------------------|-------|-----|----|------|--|
| NRAS_30bp<br>_T-flip_guide<br>_30_5   | NRAS  | UCC | 26 | I21I | GGCUGUAUUGUCAGUGCGC<br>UUUCCCAACAC (SEQ ID<br>NO:581)  |
| NRAS_30bp<br>_T-flip_guide<br>_30_7   | NRAS  | UCC | 24 | I21I | GUAGCUGUAUUGUCAGUGC<br>GCUUUCCCAAC (SEQ ID<br>NO:582)  |
| NRAS_30bp<br>_T-flip_guide<br>_30_9   | NRAS  | UCC | 22 | I21I | GAUUAGCUGUAUUGUCAGU<br>GCGCUUUCCCA (SEQ ID<br>NO:583)  |
| NRAS_30bp<br>_T-flip_guide<br>_30_11  | NRAS  | UCC | 20 | I21I | GGGAUUAGCUGUAUUGUCA<br>GUGCGCUUUUCC (SEQ ID<br>NO:584) |
| NKFB1_30b<br>p_T-flip_guid<br>e_30_3  | NKFB1 | ACC | 28 | P33S | GUGUUUGAAAUACUUCUGG<br>AUUAAAUAUUGU (SEQ ID<br>NO:585) |
| NKFB1_30b<br>p_T-flip_guid<br>e_30_5  | NKFB1 | ACC | 26 | P33S | GUGUGUUUGAAAUACUUCU<br>GGAUAAAUAU (SEQ ID<br>NO:586)   |
| NKFB1_30b<br>p_T-flip_guid<br>e_30_7  | NKFB1 | ACC | 24 | P33S | GUCUGUGUUUGAAAUACUU<br>CUGGAUAAAUA (SEQ ID<br>NO:587)  |
| NKFB1_30b<br>p_T-flip_guid<br>e_30_9  | NKFB1 | ACC | 22 | P33S | GCAUCUGUGUUUGAAAUAC<br>UUCUGGAUAAA (SEQ ID<br>NO:588)  |
| NKFB1_30b<br>p_T-flip_guid<br>e_30_11 | NKFB1 | ACC | 20 | P33S | GGCCAUCUGUGUUUGAAAU<br>ACUUCUGGAUUA (SEQ ID<br>NO:589) |
| EZH2_30bp<br>T-flip_guide_<br>30_3    | EZH2  | UCA | 28 | F32F | GCUUAACCUCUUGAGCUGU<br>CUCAGUCGAUG (SEQ ID<br>NO:590)  |
| EZH2_30bp<br>T-flip_guide_<br>30_5    | EZH2  | UCA | 26 | F32F | GGUCUUAACCUCUUGAGCU<br>GUCUCAGUCGA (SEQ ID<br>NO:591)  |
| EZH2_30bp<br>T-flip_guide_<br>30_7    | EZH2  | UCA | 24 | F32F | GUCGUCUUAACCUCUUGAG<br>CUGUCUCAGUCG (SEQ ID<br>NO:592) |
| EZH2_30bp<br>T-flip_guide_<br>30_9    | EZH2  | UCA | 22 | F32F | GGCUCGUCUUAACCUCUUG<br>AGCUGUCUCAGU (SEQ ID<br>NO:593) |

|        |                              |      |     |    |      |  |
|--------|------------------------------|------|-----|----|------|--|
| [2093] | EZH2_30bp_T-flip_guide_30_11 | EZH2 | UCA | 20 | F32F | GCAGCUCGUCUUAACCUCU<br>UGAGCUGUCUCA (SEQ ID<br>NO:594) |
|        | NF2_30bp_T-flip_guide_30_3   | NF2  | ACG | 28 | T21M | GACUUCUUGGGUUGCUUCC<br>UCUUGAGAGAGC (SEQ ID<br>NO:595) |
|        | NF2_30bp_T-flip_guide_30_5   | NF2  | ACG | 26 | T21M | GGAACUUCUUGGGUUGCUU<br>CCUCUUGAGAGA (SEQ ID<br>NO:596) |
|        | NF2_30bp_T-flip_guide_30_7   | NF2  | ACG | 24 | T21M | GGUGAACUUCUUGGGUUGC<br>UCCUCUUGAGA (SEQ ID<br>NO:597)  |
|        | NF2_30bp_T-flip_guide_30_9   | NF2  | ACG | 22 | T21M | GCGGUGAACUUCUUGGGUU<br>GCUUCCUCUUGA (SEQ ID<br>NO:598) |
|        | NF2_30bp_T-flip_guide_30_11  | NF2  | ACG | 20 | T21M | GCACGGUGAACUUCUUGGG<br>UUGCUUCCUCUU (SEQ ID<br>NO:599) |
|        | RAF1_30bp_T-flip_guide_30_3  | RAF1 | UCC | 28 | P30S | GAGUAGAGAUGCAGCUGGA<br>GCCAUCAAAC (SEQ ID<br>NO:600)   |
|        | RAF1_30bp_T-flip_guide_30_5  | RAF1 | UCC | 26 | P30S | GGUAGUAGAGAUGCAGCUG<br>GAGCCAUCAAAC (SEQ ID<br>NO:601) |
|        | RAF1_30bp_T-flip_guide_30_7  | RAF1 | UCC | 24 | P30S | GUUGUAGUAGAGAUGCAGC<br>UGGAGCCAUCAA (SEQ ID<br>NO:602) |
|        | RAF1_30bp_T-flip_guide_30_9  | RAF1 | UCC | 22 | P30S | GUAUUGUAGUAGAGAUGC<br>GCUGGAGCCAU (SEQ ID<br>NO:603)   |
|        | RAF1_30bp_T-flip_guide_30_11 | RAF1 | UCC | 20 | P30S | GACUAUUGUAGUAGAGAUG<br>CAGCUGGAGCCA (SEQ ID<br>NO:604) |

[2094] 表23:用于合成靶标编辑的指导序列

| 名称                              | 被靶向的基因 | 基序  | 碱基翻转/位置 | 密码子变化 | 间隔子序列  |
|---------------------------------|--------|-----|---------|-------|--|
| [2095] NM_000016.5_C-flip_guide | ACADM  | ACA | C/7     | H67Y  | GUAUCAUCUUCUGCAGCCA<br>CUGGGAUGAUUU (SEQ ID<br>NO:605) |

[2096]

|                                     |        |     |     |             |  |
|-------------------------------------|--------|-----|-----|-------------|--|
| NM_000018.3<br>_C-flip_guide        | ACADVL | GCG | C/9 | A283V       | GUCUCCACCCCAAAGCUG<br>UGAUCUUCUCCU (SEQ ID<br>NO:606)  |
| NM_000071.2<br>_C-flip_guide        | CBS    | GCG | C/9 | R109C       | GGAACUCACCCUUGGCCAA<br>GAGCUCACACUU (SEQ ID<br>NO:607) |
| NM_000138.4<br>_C-flip_guide        | FBN1   | GCG | C/5 | R1408<br>C  | GGAGCCCUCAUCAAGGUCU<br>GUACAAGUGAAG (SEQ ID<br>NO:608) |
| NM_000141.4<br>_C-flip_guide        | FGFR2  | CCC | C/7 | P267S       | GCUGUGGCGGCAUUUGCCG<br>GCAGUCCGGCUU (SEQ ID<br>NO:609) |
| NM_000152.4<br>_C-flip_guide        | GAA    | CCC | C/7 | P552L       | GGCCUGGCGGGUCCCCCA<br>ACCACCCAGGC (SEQ ID<br>NO:610)   |
| NM_000341.3<br>_C-flip_guide        | SLC3A1 | ACG | C/7 | T467M       | GAGAAGCCUGUUAUCACG<br>UUGACAUACUGA (SEQ ID<br>NO:611)  |
| NM_000375.2<br>_C-flip_guide        | UROS   | ACG | C/9 | R73C        | GCUCAAACCUAACUCUGC<br>UGCUUCCACUGC (SEQ ID<br>NO:612)  |
| NM_000431.3<br>_C-flip_guide        | MVK    | ACA | C/9 | T268I       | GUGGCAUCUCUUGAGGUCA<br>GGAGGGGGGCCA (SEQ ID<br>NO:613) |
| NM_000551.3<br>_C-flip_guide        | VHL    | CCG | C/7 | P158L       | GUCUUCCGAGUAUACACU<br>GGCAGUGUGAUA (SEQ ID<br>NO:614)  |
| NM_00125685<br>0.1_C-flip_gu<br>ide | TTN    | ACG | C/9 | R30071<br>C | GCUUUCCACCUGGGCCAGG<br>GGAAUCAAGCAC (SEQ ID<br>NO:615) |
| NM_002397.4<br>_C-flip_guide        | MEF2C  | ACG | C/9 | T1M         | GUUCUCCCCUAGUCCCCG<br>UUUUUCUUCUCU (SEQ ID<br>NO:616)  |
| NM_002474.2<br>_C-flip_guide        | MYH11  | CCG | C/9 | P1264L      | GUGGACUGCCGCUCUGCA<br>CCUGCGCCUCCA (SEQ ID<br>NO:617)  |
| NM_002834.4<br>_C-flip_guide        | PTPN11 | CCU | C/9 | L285F       | GAUGAUAACGGGCAGGAU<br>GUUUUUUAUAUCU (SEQ ID<br>NO:618) |
| NM_004004.5<br>_C-flip_guide        | GJB2   | ACG | C/5 | R77W        | GGCCCCUAGCCGAUGUGG<br>GAGAUGGGGAAG (SEQ ID<br>NO:619)  |

[2097]

|  |         |     |     |       |   |
|--|---------|-----|-----|-------|---|
| NM_004572.3<br>_C-flip_guide             | PKP2    | CCG | C/9 | R796C | GUGUGUAACCGGCAGAGGC<br>UGUAGUUCAAU (SEQ ID<br>NO:620)   |
| NM_005609.3<br>_C-flip_guide             | PYGM    | GCG | C/9 | R798W | GCCGCGUCCCCUCUCUUGG<br>GUUCUUGUACAA (SEQ ID<br>NO:621)  |
| NM_005633.3<br>_C-flip_guide             | SOS1    | ACG | C/9 | T269M | GCAUCUGUCCUUUCUACUG<br>UAUCUUCUAUUAU (SEQ ID<br>NO:622) |
| NM_014139.2<br>_C-flip_guide             | SCN11A  | CCG | C/9 | P396L | GCAACAGCCCGGUUAAGU<br>UAAUCAGGUAGA (SEQ ID<br>NO:623)   |
| NM_014874.3<br>_C-flip_guide             | MFN2    | CCG | C/9 | P76L  | GCAACAGCCCGGUUAAGU<br>UAAUCAGGUAGA (SEQ ID<br>NO:624)   |
| NM_015559.2<br>_C-flip_guide             | SETBP1  | ACU | C/7 | T871I | GGUCCACUGCCGUGUCG<br>CUGGGGAUCGUC (SEQ ID<br>NO:625)    |
| NM_020630.4<br>_C-flip_guide             | RET     | CCG | C/5 | R620C | GUCGCCGAAGCACUUCUCC<br>UCCUCAGGGAAG (SEQ ID<br>NO:626)  |
| NM_000016.5<br>_F_30bp_C-flip_guide_30_5 | ACADM   | ACA | C/5 | H67Y  | GUCAUCUUCUGCAGCCACU<br>GGGAUGAUUUC (SEQ ID<br>NO:627)   |
| NM_000016.5<br>_F_30bp_C-flip_guide_30_7 | ACADM   | ACA | C/7 | H67Y  | GUAUCAUCUUCUGCAGCCA<br>CUGGGAUGAUUU (SEQ ID<br>NO:628)  |
| NM_000016.5<br>_F_30bp_C-flip_guide_30_9 | ACADM   | ACA | C/9 | H67Y  | GUUUAUCAUCUUCUGCAGC<br>CACUGGGAUGAU (SEQ ID<br>NO:629)  |
| NM_000018.3<br>_F_30bp_C-flip_guide_30_5 | ACADV L | GCG | C/5 | A283V | GCACCCAAAAGCUGUGAU<br>CUUCUCCUUCAC (SEQ ID<br>NO:630)   |
| NM_000018.3<br>_F_30bp_C-flip_guide_30_7 | ACADV L | GCG | C/7 | A283V | GUCCACCCAAAAGCUGUG<br>AUCUUCUCCUUC (SEQ ID<br>NO:631)   |
| NM_000018.3<br>_F_30bp_C-flip_guide_30_9 | ACADV L | GCG | C/9 | A283V | GUCUCCACCCAAAAGCUG<br>UGAUCUUCUCCU (SEQ ID<br>NO:632)   |
| NM_000071.2<br>_F_30bp_C-flip_guide_30_5 | CBS     | GCG | C/5 | R109C | GUCACCCUUGGCCAAGAGC<br>UCACACUUCAGG (SEQ ID<br>NO:633)  |

[2098]

|  |        |     |     |            |  |
|--|--------|-----|-----|------------|--|
| p_guide_30_5                                 |        |     |     |            | NO:633)  |
| NM_000071.2<br>_F_30bp_C-fli<br>p_guide_30_7 | CBS    | GCG | C/7 | R109C      | GACUCACCCUUGGCCAAGA<br>GCUCACACUUCA (SEQ ID<br>NO:634) |
| NM_000071.2<br>_F_30bp_C-fli<br>p_guide_30_9 | CBS    | GCG | C/9 | R109C      | GGAACUCACCCUUGGCCAA<br>GAGCUCACACUU (SEQ ID<br>NO:635) |
| NM_000138.4<br>_F_30bp_C-fli<br>p_guide_30_5 | FBN1   | GCG | C/5 | R1408<br>C | GGAGCCCUCAUCAAGGUCU<br>GUACAAGUGAAG (SEQ ID<br>NO:636) |
| NM_000138.4<br>_F_30bp_C-fli<br>p_guide_30_7 | FBN1   | GCG | C/7 | R1408<br>C | GCAGAGCCCUCAUCAAGGU<br>CUGUACAAGUGA (SEQ ID<br>NO:637) |
| NM_000138.4<br>_F_30bp_C-fli<br>p_guide_30_9 | FBN1   | GCG | C/9 | R1408<br>C | GCUCAGAGCCCUCAUCAAG<br>GUCUGUACAAGU (SEQ ID<br>NO:638) |
| NM_000141.4<br>_F_30bp_C-fli<br>p_guide_30_5 | FGFR2  | CCC | C/5 | P267S      | GGUGGCGGCAUUUGCCGGC<br>AGUCCGGCUUGG (SEQ ID<br>NO:639) |
| NM_000141.4<br>_F_30bp_C-fli<br>p_guide_30_7 | FGFR2  | CCC | C/7 | P267S      | GCUGUGGCGGCAUUUGCCG<br>GCAGUCCGGCUU (SEQ ID<br>NO:640) |
| NM_000141.4<br>_F_30bp_C-fli<br>p_guide_30_9 | FGFR2  | CCC | C/9 | P267S      | GCACUGUGGCGGCAUUUGC<br>CGGCAGUCCGGC (SEQ ID<br>NO:641) |
| NM_000152.4<br>_F_30bp_C-fli<br>p_guide_30_5 | GAA    | CCC | C/5 | P552L      | GCUGGCGGGUCCCCCAAC<br>CACCCAGGCAC (SEQ ID<br>NO:642)   |
| NM_000152.4<br>_F_30bp_C-fli<br>p_guide_30_7 | GAA    | CCC | C/7 | P552L      | GGCCUGGCGGGUCCCCCA<br>ACCACCCAGGC (SEQ ID<br>NO:643)   |
| NM_000152.4<br>_F_30bp_C-fli<br>p_guide_30_9 | GAA    | CCC | C/9 | P552L      | GCCGCCUGGCGGGUCCCCC<br>AACCACCCAG (SEQ ID<br>NO:644)   |
| NM_000341.3<br>_F_30bp_C-fli<br>p_guide_30_5 | SLC3A1 | ACG | C/5 | T467M      | GAAGCCUGUUAUCACGUU<br>GACAUACUGAUU (SEQ ID<br>NO:645)  |
| NM_000341.3<br>_F_30bp_C-fli                 | SLC3A1 | ACG | C/7 | T467M      | GAGAAGCCUGUUAUCACG<br>UUGACAUACUGA (SEQ ID             |

[2099]

|   |        |     |     |             |  |
|---|--------|-----|-----|-------------|--|
| p_guide_30_7                                |        |     |     |             | NO:646)  |
| NM_000341.3<br>_F_30bp_C-flip_guide_30_9    | SLC3A1 | ACG | C/9 | T467M       | GAAAGAAGCCUGUUCAUCA<br>CGUUGACAUACU (SEQ ID<br>NO:647) |
| NM_000375.2<br>_F_30bp_C-flip_guide_30_5    | UROS   | ACG | C/5 | R73C        | GAAACCUAACUCUGCUGCU<br>UCCACUGCUCUG (SEQ ID<br>NO:648) |
| NM_000375.2<br>_F_30bp_C-flip_guide_30_7    | UROS   | ACG | C/7 | R73C        | GCCAAACCUAACUCUGCUG<br>CUUCCACUGCUC (SEQ ID<br>NO:649) |
| NM_000375.2<br>_F_30bp_C-flip_guide_30_9    | UROS   | ACG | C/9 | R73C        | GCUCCAAACCUAACUCUGC<br>UGCUUCCACUGC (SEQ ID<br>NO:650) |
| NM_000431.3<br>_F_30bp_C-flip_guide_30_5    | MVK    | ACA | C/5 | T268I       | GAUCUCUUGAGGUCAGGAG<br>GGGGGCCACGAU (SEQ ID<br>NO:651) |
| NM_000431.3<br>_F_30bp_C-flip_guide_30_7    | MVK    | ACA | C/7 | T268I       | GGCAUCUCUUGAGGUCAGG<br>AGGGGGGCCACG (SEQ ID<br>NO:652) |
| NM_000431.3<br>_F_30bp_C-flip_guide_30_9    | MVK    | ACA | C/9 | T268I       | GUGGCAUCUCUUGAGGUCA<br>GGAGGGGGGCCA (SEQ ID<br>NO:653) |
| NM_000551.3<br>_F_30bp_C-flip_guide_30_5    | VHL    | CCG | C/5 | P158L       | GUUCCGAGUAUACACUGG<br>CAGUGUGAUUU (SEQ ID<br>NO:654)   |
| NM_000551.3<br>_F_30bp_C-flip_guide_30_7    | VHL    | CCG | C/7 | P158L       | GUCUUCCGAGUAUACACU<br>GGCAGUGUGAUA (SEQ ID<br>NO:655)  |
| NM_000551.3<br>_F_30bp_C-flip_guide_30_9    | VHL    | CCG | C/9 | P158L       | GGCUCUUCCGAGUAUACA<br>CUGGCAGUGUGA (SEQ ID<br>NO:656)  |
| NM_00125685<br>0.1_F_30bp_C-flip_guide_30_5 | TTN    | ACG | C/5 | R30071<br>C | GCCACCUGGGCCAGGGGAA<br>UCAAGCACUUUG (SEQ ID<br>NO:657) |
| NM_00125685<br>0.1_F_30bp_C-flip_guide_30_7 | TTN    | ACG | C/7 | R30071<br>C | GUUCCACCUGGGCCAGGGG<br>AAUCAAGCACUU (SEQ ID<br>NO:658) |

[2100]

|   |        |     |     |             |   |
|---|--------|-----|-----|-------------|---|
| NM_00125685<br>0.1_F_30bp_C<br>-flip_guide_30<br>_9 | TTN    | ACG | C/9 | R30071<br>C | GCUUCCACCUGGGCCAGG<br>GGAAUCAAGCAC (SEQ ID<br>NO:659)   |
| NM_002397.4<br>_F_30bp_C-fl<br>p_guide_30_5         | MEF2C  | ACG | C/5 | T1M         | GCCCCUAGUCCCCGUUUU<br>UCUUCUCUCUCU (SEQ ID<br>NO:660)   |
| NM_002397.4<br>_F_30bp_C-fl<br>p_guide_30_7         | MEF2C  | ACG | C/7 | T1M         | GCUCUUUAGUCCCCGUU<br>UUUCUUCUCUCU (SEQ ID<br>NO:661)    |
| NM_002397.4<br>_F_30bp_C-fl<br>p_guide_30_9         | MEF2C  | ACG | C/9 | T1M         | GUUCUUUUUAGUCCCCG<br>UUUUUCUUCUCU (SEQ ID<br>NO:662)    |
| NM_002474.2<br>_F_30bp_C-fl<br>p_guide_30_5         | MYH11  | CCG | C/5 | P1264L      | GCUGCCGCUCCUGCACCUG<br>CGCCUCCAGCUU (SEQ ID<br>NO:663)  |
| NM_002474.2<br>_F_30bp_C-fl<br>p_guide_30_7         | MYH11  | CCG | C/7 | P1264L      | GGACUGCCGCUCCUGCACC<br>UGCGCCUCCAGC (SEQ ID<br>NO:664)  |
| NM_002474.2<br>_F_30bp_C-fl<br>p_guide_30_9         | MYH11  | CCG | C/9 | P1264L      | GUGGACUGCCGCUCCUGCA<br>CCUGCGCCUCCA (SEQ ID<br>NO:665)  |
| NM_002834.4<br>_F_30bp_C-fl<br>p_guide_30_5         | PTPN11 | CCU | C/5 | L285F       | GUCAACGGGCAGGAUGUUU<br>UUAUAUCUAUUU (SEQ ID<br>NO:666)  |
| NM_002834.4<br>_F_30bp_C-fl<br>p_guide_30_7         | PTPN11 | CCU | C/7 | L285F       | GGAUCAACGGGCAGGAUGU<br>UUUUUAUAUCAU (SEQ ID<br>NO:667)  |
| NM_002834.4<br>_F_30bp_C-fl<br>p_guide_30_9         | PTPN11 | CCU | C/9 | L285F       | GAUGAUCAACGGGCAGGAU<br>GUUUUUUAUAUCU (SEQ ID<br>NO:668) |
| NM_004004.5<br>_F_30bp_C-fl<br>p_guide_30_5         | GJB2   | ACG | C/5 | R77W        | GGCCCCUAGCCGGAUGUGG<br>GAGAUGGGGAAG (SEQ ID<br>NO:669)  |
| NM_004004.5<br>_F_30bp_C-fl<br>p_guide_30_7         | GJB2   | ACG | C/7 | R77W        | GGGGCCCCUAGCCGGAUGU<br>GGGAGAUGGGGA (SEQ ID<br>NO:670)  |
| NM_004004.5<br>_F_30bp_C-fl<br>p_guide_30_9         | GJB2   | ACG | C/9 | R77W        | GCAGGGCCCCUAGCCGGAU<br>GUGGGAGAUGGG (SEQ ID<br>NO:671)  |

[2101]

|  |        |     |     |       |  |
|--|--------|-----|-----|-------|--|
| NM_004572.3<br>_F_30bp_C-flip_guide_30_5 | PKP2   | CCG | C/5 | R796C | GUAACCGGCAGAGGCUGUA<br>GUUUCAAUGAGA (SEQ ID<br>NO:672) |
| NM_004572.3<br>_F_30bp_C-flip_guide_30_7 | PKP2   | CCG | C/7 | R796C | GUGUAACCGGCAGAGGCUG<br>UAGUUCAAUGA (SEQ ID<br>NO:673)  |
| NM_004572.3<br>_F_30bp_C-flip_guide_30_9 | PKP2   | CCG | C/9 | R796C | GUGUGUAACCGGCAGAGGC<br>UGUAGUUCAAU (SEQ ID<br>NO:674)  |
| NM_005609.3<br>_F_30bp_C-flip_guide_30_5 | PYGM   | GCG | C/5 | R798W | GGUCCCCUCUCUUGGGUUC<br>UUGUACAAGGCG (SEQ ID<br>NO:675) |
| NM_005609.3<br>_F_30bp_C-flip_guide_30_7 | PYGM   | GCG | C/7 | R798W | GGCGUCCCCUCUCUUGGGU<br>UCUUGUACAAGG (SEQ ID<br>NO:676) |
| NM_005609.3<br>_F_30bp_C-flip_guide_30_9 | PYGM   | GCG | C/9 | R798W | GCCGCGUCCCCUCUCUUGG<br>GUUCUUGUACAA (SEQ ID<br>NO:677) |
| NM_005633.3<br>_F_30bp_C-flip_guide_30_5 | SOS1   | ACG | C/5 | T269M | GUGUCCUUUCUACUGUAUC<br>UUCUAUAUGGCC (SEQ ID<br>NO:678) |
| NM_005633.3<br>_F_30bp_C-flip_guide_30_7 | SOS1   | ACG | C/7 | T269M | GUCUGUCCUUUCUACUGUA<br>UCUUCUAUAUGG (SEQ ID<br>NO:679) |
| NM_005633.3<br>_F_30bp_C-flip_guide_30_9 | SOS1   | ACG | C/9 | T269M | GCAUCUGUCCUUUCUACUG<br>UAUCUUCUAUAU (SEQ ID<br>NO:680) |
| NM_014139.2<br>_F_30bp_C-flip_guide_30_5 | SCN11A | CCG | C/5 | P396L | GAGCCCGGUUAAGUUAU<br>CAGGUAGAAGGA (SEQ ID<br>NO:681)   |
| NM_014139.2<br>_F_30bp_C-flip_guide_30_7 | SCN11A | CCG | C/7 | P396L | GACAGCCCGGUUAAGUUA<br>AUCAGGUAGAAG (SEQ ID<br>NO:682)  |
| NM_014139.2<br>_F_30bp_C-flip_guide_30_9 | SCN11A | CCG | C/9 | P396L | GCAACAGCCCGGUUAAGU<br>UAAUCAGGUAGA (SEQ ID<br>NO:683)  |
| NM_014874.3<br>_F_30bp_C-flip_guide_30_5 | MFN2   | CCG | C/5 | P76L  | GGUCCGAACCUUGUUCUUC<br>UGUGGUAACGGG (SEQ ID<br>NO:684) |

[2102]

|   |        |     |      |       |  |
|---|--------|-----|------|-------|--|
| NM_014874.3<br>_F_30bp_C-flip<br>p_guide_30_7 | MFN2   | CCG | C/7  | P76L  | GACGUCCGAACCGUUCU<br>UCUGUGGUAACG (SEQ ID<br>NO:685)   |
| NM_014874.3<br>_F_30bp_C-flip<br>p_guide_30_9 | MFN2   | CCG | C/9  | P76L  | GUGACGUCCGAACCGUU<br>CUUCUGUGGUA (SEQ ID<br>NO:686)    |
| NM_015559.2<br>_F_30bp_C-flip<br>p_guide_30_5 | SETBP1 | ACU | C/5  | T871I | GCCCACUGCCGUGUCGU<br>GGGGAUCGUCUC (SEQ ID<br>NO:687)   |
| NM_015559.2<br>_F_30bp_C-flip<br>p_guide_30_7 | SETBP1 | ACU | C/7  | T871I | GGUCCACUGCCGUGUCG<br>CUGGGGAUCGUC (SEQ ID<br>NO:688)   |
| NM_015559.2<br>_F_30bp_C-flip<br>p_guide_30_9 | SETBP1 | ACU | C/9  | T871I | GCUGUCCACUGCCGUGU<br>CGCUGGGGAUCG (SEQ ID<br>NO:689)   |
| NM_020630.4<br>_F_30bp_C-flip<br>p_guide_30_5 | RET    | CCG | C/5  | R620C | GUCGCCGAAGCACUUCUCC<br>UCCUCAGGGAAG (SEQ ID<br>NO:690) |
| NM_020630.4<br>_F_30bp_C-flip<br>p_guide_30_7 | RET    | CCG | C/7  | R620C | GGCUCGCCGAAGCACUUCU<br>CCUCCUCAGGGA (SEQ ID<br>NO:691) |
| NM_020630.4<br>_F_30bp_C-flip<br>p_guide_30_9 | RET    | CCG | C/9  | R620C | GGGGCUCGCCGAAGCACUU<br>CUCCUCCUCAGG (SEQ ID<br>NO:692) |
| ApoE4<br>rs429358 C<br>flip 30                | APOE   | GCG | C/30 | C130R | Gccacgucccauguccgcccagcgg<br>(SEQ ID NO:693)           |
| ApoE4<br>rs429358 C<br>flip 28                | APOE   | GCG | C/28 | C130R | Ggcccacgucccauguccgcccagc<br>cg (SEQ ID NO:694)        |
| ApoE4<br>rs429358 C<br>flip 26                | APOE   | GCG | C/26 | C130R | Gccgcccacgucccauguccgccc<br>gc (SEQ ID NO:695)         |
| ApoE4<br>rs429358 C<br>flip 24                | APOE   | GCG | C/24 | C130R | Gggcgcccacgucccauguccgccc<br>ca (SEQ ID NO:696)        |
| ApoE4<br>rs429358 C<br>flip 22                | APOE   | GCG | C/22 | C130R | Ggcggcccacgucccauguccgccc<br>cc (SEQ ID NO:697)        |

[2103]

|                                |      |     |      |       |  |
|--------------------------------|------|-----|------|-------|--|
| ApoE4<br>rs429358 C<br>flip 20 | APOE | GCG | C/20 | C130R | Gaggcggcggcccacgucccaugcgcg<br>(SEQ ID NO:698)       |
| ApoE4<br>rs429358 C<br>flip 18 | APOE | GCG | C/18 | C130R | Gccaggcggcggcccacgucccaugcgcg<br>(SEQ ID NO:699)     |
| ApoE4<br>rs429358 C<br>flip 16 | APOE | GCG | C/16 | C130R | Gcaccaggcggcggcccacgucccauguc<br>(SEQ ID NO:700)     |
| ApoE4 rs7412<br>C flip 30      | APOE | GCG | C/30 | C176R | Gccuucgucagggucaucggcaucggaggga<br>(SEQ ID NO:701)   |
| ApoE4 rs7412<br>C flip 28      | APOE | GCG | C/28 | C176R | Ggcccucgucagggucaucggcaucggcgag<br>(SEQ ID NO:702)   |
| ApoE4 rs7412<br>C flip 26      | APOE | GCG | C/26 | C176R | Gaggccuucgucagggucaucggcaucgcgg<br>(SEQ ID NO:703)   |
| ApoE4 rs7412<br>C flip 24      | APOE | GCG | C/24 | C176R | Gccaggccuucgucagggucaucggcaucgcgc<br>(SEQ ID NO:704) |
| ApoE4 rs7412<br>C flip 22      | APOE | GCG | C/22 | C176R | Gugccaggccuucgucagggucaucggcauc<br>(SEQ ID NO:705)   |
| ApoE4 rs7412<br>C flip 20      | APOE | GCG | C/20 | C176R | Gacugccaggccuucgucagggucaucggca<br>(SEQ ID NO:706)   |
| ApoE4 rs7412<br>C flip 18      | APOE | GCG | C/18 | C176R | Gacacugccaggccuucgucagggucaucggg<br>(SEQ ID NO:707)  |
| ApoE4 rs7412<br>C flip 16      | APOE | GCG | C/16 | C176R | Gguacacugccaggccuucgucagggucauc<br>(SEQ ID NO:708)   |

[2104] 表24:哺乳动物质粒和图谱

[2105]

| 质粒     | 描述  | Benchling 链接  |
|--------|---|---|
| pC0043 | PspCas13b crRNA 主链  | <a href="https://benchling.com/s/seq-OH6nMmZCZn930BWqcFNa">benchling.com/s/seq-OH6nMmZCZn930BWqcFNa</a> |
| pC0076 | CMV-dCas13b6-mapkNES-GS-dADAR2 E488Q  | <a href="https://benchling.com/s/seq-BulRvsrtwP4aEJtTqYM2">benchling.com/s/seq-BulRvsrtwP4aEJtTqYM2</a> |
| pC0077 | pCMV-dCas13b6-mapkNES-GS-dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V) RESCUEv8 | <a href="https://benchling.com/s/seq-gQ13PMPLkcO6OceAfmPC">benchling.com/s/seq-gQ13PMPLkcO6OceAfmPC</a> |
| pC0078 | pCMV-dCas13b6-mapkNES-GS-dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V) RESCUEv8 | <a href="https://benchling.com/s/seq-19Ytwwh0i0vSIbyXYZ95">benchling.com/s/seq-19Ytwwh0i0vSIbyXYZ95</a> |

[2106]

|        |   |  |
|--------|---|--|
|        | 8V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) V16   |  |
| pC0079 | pCMV-dCas13b6-mapkNES-GS-dADAR2(E488Q/V351G/S486A/T375A/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) V16S       | <a href="http://benchling.com/s/seq-WX6VnavLS6JaaZ54XAOx">benchling.com/s/seq-WX6VnavLS6JaaZ54XAOx</a> |
| pC0080 | pCMV-dCas13b12-HIVNES-GS-dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) V16        | <a href="http://benchling.com/s/seq-GQqPCRE916KnEfHksQem">benchling.com/s/seq-GQqPCRE916KnEfHksQem</a> |
| pC0081 | pCMV-dCas13b12-HIVNES-GS-dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T/S375A) V16S | <a href="http://benchling.com/s/seq-qjBEAXZgupeRXBa8ablS">benchling.com/s/seq-qjBEAXZgupeRXBa8ablS</a> |
| pC0082 | CMV- Cluciferase - polyA EF1a-G-荧光素酶(C82R)-polyA C 至 U 报告基因 TCG 基序  | <a href="http://benchling.com/s/seq-Qjsg3Yx0r1Hs77GT58BI">benchling.com/s/seq-Qjsg3Yx0r1Hs77GT58BI</a> |
| pC0083 | CMV-Cluciferase- polyA EF1a-G-荧光素酶(C82R)-polyA C 至 U 报告基因 GCG 基序  | <a href="http://benchling.com/s/seq-Z8zWu3LdetcuYHAFGnpe">benchling.com/s/seq-Z8zWu3LdetcuYHAFGnpe</a> |
| pC0084 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(C82R)-polyA C 至 U 报告基因 ACG 基序   | <a href="http://benchling.com/s/seq-G2Iag618NBQAXqbJnou5">benchling.com/s/seq-G2Iag618NBQAXqbJnou5</a> |
| pC0085 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(C82R)-polyA C 至 U 报告基因 CCG 基序   | <a href="http://benchling.com/s/seq-alkwhNUsFTg80TVmpquP">benchling.com/s/seq-alkwhNUsFTg80TVmpquP</a> |
| pC0086 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(L77P)-polyA C 至 U 报告基因 CCA 基序   | <a href="http://benchling.com/s/seq-1J8Fm6vtF7GS676Q7pwS">benchling.com/s/seq-1J8Fm6vtF7GS676Q7pwS</a> |
| pC0087 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(L77P)-polyA C 至 U 报告基因 CCT 基序   | <a href="http://benchling.com/s/seq-5MMokwvx0Ajq6ML2sjjZ">benchling.com/s/seq-5MMokwvx0Ajq6ML2sjjZ</a> |
| pC0088 | pCMV-ADAR2dd(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) V16                           | <a href="http://benchling.com/s/seq-YISAybq2YnuclVwYDy95">benchling.com/s/seq-YISAybq2YnuclVwYDy95</a> |
| pC0089 | pCMV-ADAR2 全长(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) V16                          | <a href="http://benchling.com/s/seq-95ZpoHj9GhQFzlu3m6cb">benchling.com/s/seq-95ZpoHj9GhQFzlu3m6cb</a> |
| pC0090 | 具有 Gluc/Cluc 的 $\beta$ 连环蛋白报告基因 M50 Super 8x (TCF/LEF 结合位点) TOPFlash  | <a href="http://benchling.com/s/seq-jPxZnxs3wSeKZhgTTDBu">benchling.com/s/seq-jPxZnxs3wSeKZhgTTDBu</a> |
| pC0091 | 具有 Gluc/Cluc 的 $\beta$ 连环蛋白报告基因对照 M51 Super 8x (突变 TCF/LEF 结合位点) FOPFlash   | <a href="http://benchling.com/s/seq-130b6c9baCfw8R3ITgSR">benchling.com/s/seq-130b6c9baCfw8R3ITgSR</a> |

[2107] 表25:酵母质粒和图谱

| 质粒            | 描述  | Benchling 链接  |
|---------------|---|---|
| pC0092        | pGAL-dCas13b6-GS-dADAR2 [RESCUE v0 酵母]  | <a href="https://benchling.com/s/seq-w112aOHR2gSc4P2aQ7VY">benchling.com/s/seq-w112aOHR2gSc4P2aQ7VY</a> |
| pC0093        | pGAL-dCas13b6-GS-dADAR2(V351/S486A/T375S) [RESCUE v3 酵母]  | <a href="https://benchling.com/s/seq-saQngvNf6i3GhSGF0H3I">benchling.com/s/seq-saQngvNf6i3GhSGF0H3I</a> |
| pC0094        | pGAL-dCas13b6-GS-dADAR2(V351G/S486A/T375S/S370C/P462A/L332I) [RESCUE v7 酵母]   | <a href="https://benchling.com/s/seq-GIJ7BnpV3Vd3XtKilxdm">benchling.com/s/seq-GIJ7BnpV3Vd3XtKilxdm</a> |
| pC0095        | pGAL-dCas13b6-GS-dADAR2(V351/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418/ES661T) [RESCUE v16 酵母] | <a href="https://benchling.com/s/seq-yRnAMlwozJk5r6LmOCgG">benchling.com/s/seq-yRnAMlwozJk5r6LmOCgG</a> |
| pC0096        | pYES3/CT pADH1-HH-靶向-B6_DR--HDV-space-ADH1_终止子 His (P196L) [酵母靶向 His P196L]   | <a href="https://benchling.com/s/seq-Xs2fVVMn4FwwQ79zDDEo">benchling.com/s/seq-Xs2fVVMn4FwwQ79zDDEo</a> |
| [2108] pC0097 | pYES3/CT pADH1-HH-金门-BsmBi-BsmBI_DR--HDV-space-ADH1_终止子 His (P196L) [酵母靶向 His P196L NT]   | <a href="https://benchling.com/s/seq-UM9NjG7Jkk0GFe9MowGQ">benchling.com/s/seq-UM9NjG7Jkk0GFe9MowGQ</a> |
| pC0098        | pYES3/CT pADH1-HH-Guide-B6_DR--HDV-space-ADH1_终止子 His S129P [酵母靶向 His S129P]  | <a href="https://benchling.com/s/seq-EefJl5brql13fm0B5Qc5">benchling.com/s/seq-EefJl5brql13fm0B5Qc5</a> |
| pC0099        | pYES3/CT pADH1-HH-Golden-gate-BsmBi-BsmBI_DR--HDV-space-ADH1_终止子 His 基序 S129P [酵母靶向 His S129P NT]                                       | <a href="https://benchling.com/s/seq-bt7gOlrp8OuOoV3YJWZG">benchling.com/s/seq-bt7gOlrp8OuOoV3YJWZG</a> |
| pC0100        | pYES3/CT pADH1-HH-Y66H-靶向-B6-DR-HDV-ADH1-term ATG-ycGFP Y66H [酵母靶向 GFP Y66H]  | <a href="https://benchling.com/s/seq-hiMELqTYPT9y0nOAKEq2">benchling.com/s/seq-hiMELqTYPT9y0nOAKEq2</a> |
| pC0101        | pYES3/CT pADH1-HH-金门-BsmBi-BsmBI_-B6_DR-HDV-ADH1-term ATG-ycGFP Y66H 报告基因[酵母靶向 GFP Y66H NT]   | <a href="https://benchling.com/s/seq-OCWlvnjeKYwSbG8GELTQ">benchling.com/s/seq-OCWlvnjeKYwSbG8GELTQ</a> |

[2109] 表26: 用于酵母靶向的指导序列

| 名称                  | 被靶向的基因 | 基序  | 碱基翻转/间隔子长度/位置 | 密码子变化 | 间隔子长度   |
|---------------------|--------|-----|---------------|-------|---|
| [2110] His L196P 靶向 | HIS    | CCU | U/50/34       | L196P | Ucuuauaggcaaccgaugagccuugaacgcacucacacacggugaugau (SEQ ID |

|        |              |      |     |         |  |
|--------|--------------|------|-----|---------|--|
|        |              |      |     |         | NO:709)  |
| [2111] | His S129P 靶向 | HIS  | UCC | C/30/26 | S129P<br>Gcuugcaagugccucaacca<br>aaggcgcaau (SEQ<br>ID NO:710)                       |
|        | His Y66H 靶向  | EGFP | UCA | U/50/34 | Y66H<br>Aaacauugaacaccuuagu<br>uaaaguagugacuaagguug<br>gccauggaac (SEQ ID<br>NO:711) |

[2112] 哺乳动物细胞培养

[2113] 除非另有说明,否则哺乳动物细胞培养实验在HEK293FT细胞系 (American Type Culture Collection (ATCC)) 中进行,在含有葡萄糖、丙酮酸钠和GlutaMAX (Thermo Fisher Scientific),并补充有1×青霉素-链霉素 (Thermo Fisher Scientific) 和10%胎牛血清 (VWR Seradigm) 的Dulbecco 改良Eagle培养基中生长。细胞保持在低于80%的汇合度。

[2114] 除非另有说明,否则所有转染均用Lipofectamine 2000 (Thermo Fisher Scientific) 在涂覆聚D-赖氨酸 (BD Biocoat) 的96孔板上进行。转染前 16小时,将细胞以约20,000个细胞/孔铺板以确保转染时达到90%的融合度。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) 混合至总计25μl。分别地将24.5μl的Opti-MEM与 0.5μl的Lipofectamine 2000混合。然后将质粒和脂质体转染溶液合并并孵育5分钟,然后将它们移液到细胞上。

[2115] 哺乳动物细胞中的RESCUE编辑

[2116] 为了评估哺乳动物细胞中的RESCUE活性,申请人转染了150ng RESCUE载体、300ng 指导序列表达质粒,和当使用报告基因时 (荧光素酶、STAT活性或β连环蛋白活性) 的40ng RNA编辑报告基因。48 小时后,从细胞中收获RNA并使用先前所述的方法 (33) 用基因特异性逆转录引物进行逆转录。然后使用NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs) 将提取的cDNA进行两轮PCR,以添加Illumina 接头和样品条码。然后将文库在Illumina NextSeq或MiSeq上进行下一代测序。然后在测序窗口内在所有腺苷处评估RNA编辑率。

[2117] 在靶向荧光素酶报告基因用于RNA编辑的实验中,申请人还在RNA 收获之前收获具有分泌的荧光素酶的培养基。申请人在具有注射方案的读板仪 (Biotek Synergy Neo2) 上使用海萤和高斯荧光素酶测定试剂盒 (Targeting Systems) 测量了荧光素酶活性。进行的所有重复都是生物学重复。

[2118] 在RESCUE质粒的输入量发生变化的实验中,除非另有说明,否则通过用表达CMV驱动的mScarlet的填充质粒替换RESCUE表达质粒,使质粒的总量保持恒定。在指导质粒的输入量发生变化的实验中,通过非靶向指导序列的替换使质粒总量保持恒定 (“具有填充质粒”),或不保持恒定 (“不具有填充质粒”);在该实验中,没有用于RESCUE质粒的填充质粒。

[2119] ADAR2dd上RESCUE突变的生化表征

[2120] 为了评估包含RESCUE突变的hADAR2脱氨酶结构域的动力学活性,将多次迭代克隆到包含URA3基因的 pGAL-His6-TwinStrep-SUMO-hADAR2dd主链中。将该质粒转化到BCY123感受态酵母细胞中。简而言之,将冷冻细胞在37℃水浴中解冻 15-30秒。将每种条件

下的10 $\mu$ L细胞在微量离心机中以13,000g离心2分钟,然后去除上清液。对每个构建体制备的转化混合物包含以50%w/v制备的260 $\mu$ L PEG 3350、50 $\mu$ L变性鲑鱼精子(Thermo Fisher Scientific)、36 $\mu$ L 1M乙酸锂和14 $\mu$ L DI H<sub>2</sub>O中的750ng质粒。将酵母沉淀物用转化混合物重悬,并在42 $^{\circ}$ C水浴中温育30分钟,然后以13,000g离心30秒,然后去除上清液。然后将沉淀重悬于1mL DI H<sub>2</sub>O中,并取50 $\mu$ L与1mL DI H<sub>2</sub>O混合。随后,将200 $\mu$ L铺板到减去尿嘧啶的基本葡萄糖平板上进行原养型选择。

[2121] 将平板在30 $^{\circ}$ C下孵育48小时,然后将单个菌落接种到10mL的补充有右旋糖的酵母基本培养基中。这包含酵母脱落补充剂Y2001 (1.39 g/L)、不含氨基酸的酵母氮基础(6.7g/L)、腺嘌呤半硫酸盐(0.022g/L)、组氨酸(0.076g/L)、亮氨酸(0.38g/L)、色氨酸(0.076g/L)和右旋糖(20g/L)。培养物生长过夜,然后将全部10mL接种到100mL基本培养基/右旋糖培养物中。生长8小时后,将每种构建体接种到两个2L烧瓶中,其中含有补充有20g棉子糖(VWR)的1L基本培养基。将这些生长过夜,并在收获前用100mL的30%半乳糖诱导8小时。将培养物在Beckman Coulter Avanti J-E离心机上以6,000r.p.m旋下20分钟,并将沉淀物保存在-80 $^{\circ}$ C下。

[2122] 纯化方法

[2123] 全转录组测序以评估ADAR编辑特异性

[2124] 为了分析整个转录组中的脱靶RNA编辑位点,申请人使用RNeasy Plus Miniprep试剂盒(Oiagen)从转染后48小时的细胞中收获总RNA。然后使用NEBNext Poly(A)mRNA磁性分离模块(NEB)富集mRNA级分,然后使用NEBNext Ultra RNA Library Prep Kit for Illumina(NEB)制备此RNA用于测序。然后将文库在Illumina NextSeq上测序并加载,使得每个样品有至少500万个读取结果。

[2125] 靶向和全转录组实验的RNA编辑分析

[2126] 使用自定义工作流程: [portal.firecloud.org/#methods/m/rna\\_editing\\_final\\_workflow/rna\\_editing\\_final\\_workflow/1](https://portal.firecloud.org/#methods/m/rna_editing_final_workflow/rna_editing_final_workflow/1),在FireCloud计算框架([software.broadinstitute.org/firecloud/](https://software.broadinstitute.org/firecloud/))上进行全转录组编辑RNA测序数据的分析。为了进行分析,除非另有说明,否则序列文件被随机降采样至500万个读取结果。使用添加了Gluc和Cluc序列的RefSeq GRCh38组件生成了索引,并使用Bowtie/RSEM版本1.3.0对读取结果进行比对和定量。然后使用REDitools(35,36)对比对BAM进行RNA编辑位点的分选和分析,参数如下:-t 8-e-d-l-U[AG或TC或CT或者GA]-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重要编辑均视为SNP或转染伪影,并从脱靶分析中滤除。如果通过Benjamini Hochberg校正进行的多个假设校正后,Fisher精密检验得出小于0.05的p值,并且3个生物学重复中至少有2个识别了该编辑位点,则认为脱靶是显著的。相对于最大可能重叠,计算样品之间的编辑的重叠,相当于两个样本之间的编辑的较少次数。重叠编辑位点的百分比计算为共享编辑位点的数量除以两个样品的最小编辑数量再乘以100。使用Kaviar(37)方法对已知SNP位置进行额外一层过滤以识别SNP。

[2127] 差异基因表达分析

[2128] Stat表型测定

[2129] 如上所述使用lipofectamine 2000,用RESCUE质粒、靶向STAT3和STAT1上的残基的指导质粒、以及用于STAT3(Qiagen Cignal STAT3 Reporter)和STAT1信号传导(Qiagen

Cignal GAS Reporter) 的荧光素酶报告基因转染细胞,并孵育48小时。48小时后,使用 Dual-Glo 荧光素酶测定法 (Promega) 来测量细胞中的萤火虫和海肾荧光素酶活性。将萤火虫信号相对于海肾信号进行标准化以测量STAT3和STAT1的相对激活。

#### [2130] β-连环蛋白表型测定

[2131] 转染前24小时将细胞铺板在含有防止细胞在孔中心生长的核的细胞迁移板上。24小时后,如上所述使用lipofectamine 2000,用RESCUE质粒、靶向β-连环蛋白上的残基的指导质粒、以及用于β-连环蛋白激活的荧光素酶报告基因 (Qiagen TCF/LEF Cignal Reporter) 转染细胞并孵育。24小时后,去除中心核以允许细胞向孔中心生长。再孵育24小时后,测定培养基的Gluc和Cluc荧光素酶信号。计算Gluc与Cluc的相对比率,以确定条件之间的相对β连环蛋白激活。在第3天,将细胞与CellTracker Green CMFDA Dye (ThermoFisher Scientific) 一起孵育10分钟,然后用培养基洗涤。每天使用荧光对细胞成像以测量细胞生长。使用ImageJ软件通过计算中心生长区域中荧光的总面积来测量细胞向孔中心区域中的生长。使用自动宏通过以下命令处理图像:

[2132] //ImageJ macro for calculating cellular area

[2133] run("8-bit");

[2134] run("Auto Local Threshold", "method=Bernsen radius=15 parameter\_1=0parameter\_2=0white");

[2135] setAutoThreshold("Default dark");

[2136] run("Measure");

#### [2137] 参考文献

[2138] 1.S.Shmakov等,Discovery and Functional Characterization of Diverse Class 2CRISPR-Cas Systems.Mol Cell 60,385-397 (2015) .

[2139] 2.S.Shmakov等,Diversity and evolution of class 2CRISPR-Cas systems.Nat Rev Microbiol 15,169-182 (2017) .

[2140] 3.A.A.Smargon等,Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28.Mol Cell 65,618-630e617 (2017) .

[2141] 4.0.0.Abudayyeh等,C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.Science 353,aaf5573 (2016) .

[2142] 5.S.Konermann等,Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors.Cell 173,665-676e614 (2018) .

[2143] 6.W.X.Yan等,Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein.Mol Cell 70,327-339e325 (2018) .

[2144] 7.A.East-Seletsky等,Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection.Nature 538,270-273 (2016) .

[2145] 8.J.S.Gootenberg等,Nucleic acid detection with CRISPR-Cas13a/C2c2.Science 356,438-442 (2017) .

[2146] 9.0.0.Abudayyeh等,RNA targeting with CRISPR-Cas13.Nature 550,280-284

(2017) .

[2147] 10.A.East-Seletsky,M.R.O'Connell,D.Burstein,G.J.Knott,J.A. Doudna,RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes.Mol Cell 66,373-383e373 (2017) .

[2148] 11.D.B.T.Cox et al.,RNA editing with CRISPR-Cas13.Science 358, 1019-1027 (2017) .

[2149] 12.J.S.Gootenberg等,Multiplexed and portable nucleic acid detection platform with Cas13,Cas12a,and Csm6.Science 360,439-444 (2018) .

[2150] 13.H.Nishimasu等,Crystal structure of Cas9 in complex with guide RNA and target DNA.Cell 156,935-949 (2014) .

[2151] 14.T.Yamano等,Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA.Cell 165,949-962 (2016) .

[2152] 15.L.Holm,L.M.Laakso,Dali server update.Nucleic Acids Res 44, W351-355 (2016) .

[2153] 16.H.Yang,P.Gao,K.R.Rajashankar,D.J.Patel,PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. Cell 167,1814-1828e1812 (2016) .

[2154] 17.L.Liu等,Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities.Cell 168,121-134e112 (2017) .

[2155] 18.L.Liu等,The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a.Cell 170,714-726e710 (2017) .

[2156] 19.G.J.Knott等,Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme.Nat Struct Mol Biol 24,825-833 (2017) .

[2157] 20.N.F.Sheppard,C.V.Glover,3rd,R.M.Terns,M.P.Terns,The CRISPR-associated Csx1 protein of *Pyrococcus furiosus* is an adenosine-specific endoribonuclease.RNA 22,216-224 (2016) .

[2158] 21.Z.Wu,H.Yang,P.Colosi,Effect of genome size on AAV vector packaging.Mol Ther 18,80-86 (2010) .

[2159] 22.X.J.Lu,H.J.Bussemaker,W.K.Olson,DSSR:an integrated software tool for dissecting the spatial structure of RNA.Nucleic Acids Res 43,e142 (2015) .

[2160] 23.I.Fonfara,H.Richter,M.Bratovic,A.Le Rhun,E.Charpentier,The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA.Nature 532,517-521 (2016) .

[2161] 24.D.Milburn,R.A.Laskowski,J.M.Thornton,Sequences annotated by structure:a tool to facilitate the use of structural information in sequence analysis.Protein Eng 11,855-859 (1998) .

[2162] 25.I.M.Slaymaker et al.,Rationally engineered Cas9 nucleases with improved specificity.Science 351,84-88 (2016) .

[2163] 26.L.Gao等,Engineered Cpf1 variants with altered PAM specificities.

Nat Biotechnol 35,789-792 (2017) .

[2164] 实施例14-腺嘌呤脱氨酶ADAR2到胞嘧啶脱氨酶的转化用于可编程 RNA编辑

[2165] 可编程RNA编辑可以实现RNA信息的可逆编码用于研究和疾病治疗。在此,本实施例通过将ADAR2定向进化为胞苷脱氨酶,显示了C至U RNA编辑器,称为“用于特异性C至U交换的RNA编辑 (RESCUE)”。RESCUE使通过RNA编辑可靶向的致病突变的数量翻倍,并使得能够调节与磷酸化信号传导相关的残基(例如苏氨酸和丝氨酸)。申请人应用RESCUE来驱动 $\beta$ -连环蛋白激活和细胞生长。此外,RESCUE保留了A至I编辑活性,通过使用定制的指导RNA使得能够进行C至U和A至I 多重编辑。

[2166] 总而言之,本实施例显示具有融合至CRISPR-Cas13的定向进化 ADAR2的可编程胞苷至尿苷RNA编辑扩展了RNA编辑工具箱。

[2167] 申请人先前开发了称为REPAIR(用于可编程A至I(G)置换的RNA 编辑)的RNA碱基编辑技术,其使用RNA靶向CRISPR效应子Cas13(1-6) 将ADAR2的催化结构域引导至特定RNA转录物以以单碱基精确度实现腺嘌呤至肌苷的转化(7)。用于将胞苷转化为尿苷的精确RNA编辑的技术将极大地扩展可处理疾病突变的范围,并允许通过翻译后修饰位点的改变来调节细胞中的信号传导通路(图107A)。

[2168] 尽管已利用能够催化C至U转化的天然酶进行DNA碱基编辑(16、17),但它们仅在单链底物上操作(18),在基因组和转录组上均显示脱靶(19-21),并且在窗口内使多个碱基脱氨基。在本实施例中,申请人采用合成方法来进化ADAR2的腺嘌呤脱氨酶结构域(ADAR2dd),其天然作用于双链RNA底物并优先将与胞苷错配对的靶腺嘌呤脱氨基为胞苷脱氨酶。申请人将这种进化的胞苷脱氨酶与dCas13融合,以开发哺乳动物细胞中的用于特异性C至U交换的可编程RNA编辑 (RESCUE) (图107B),申请人用其编辑STAT和 $\beta$ -连环蛋白的磷酸化信号传导并调节细胞生长。最后,申请人展示了通过RESCUE进行A至I和C至U 碱基的多重转化,并通过合理诱变将RESCUE特异性提高了10倍以上,从而生成了高特异性和精确性的C至U RNA编辑工具。

[2169] 基于显示出其催化核心之间的显著结构同源性的 大肠杆菌胞苷脱氨酶和人ADAR2dd的比较(22)(图107B),申请人选择了与RNA底物接触的ADAR2dd的残基(23),用于在与来自鸭疫里默氏菌的催化失活CAS13b直向同源物(dRanCas13b)融合的ADAR2dd上进行了三轮合理诱变,产生了具有15%编辑活性的RESCUE第3轮(RESCUEr3)(图103A-103B、108、109A-109B)。然后申请人开始在ADAR2dd上进行定向进化以鉴定增加酵母中RESCUE活性的额外候选突变。

[2170] 16轮进化以最终构建体RESCUEr16(以下简称RESCUE)告终,导致所有被测基序中增加的胞苷脱氨酶活性,在紧邻的5'和3'碱基的16种可能基序组合中的12种上具有高于20%的编辑(图103C、110、111A-111C、112、113A-113E)。申请人还表征了稳健活性所必需的指导序列特征,发现使用30-nt指导序列,RESCUE在靶标碱基中具有C或U碱基翻转时具有最佳活性(图103C、114A-114C、115)。此外,由于dRanCas13b和来自Prevotella sp.P5-125的催化失活Cas13b直向同源物(dPspCas13b)是等价的,因此最终的RESCUE构建体使用dRanCas13b(图116)。

[2171] RESCUE中的16个突变分布在ADAR2dd的整个结构中(图117A),指示进化的残基与催化口袋内的RNA靶标的直接相互作用以及间接作用两者(图117B)。这些突变使得能够适

合腺苷或胞苷,因为RESCUE 能够进行腺苷和胞苷两者的脱氨基化(图108A-108D)。申请人通过将突变体单独地添加到REPAIR或从RESCUE去除它们来评估每个突变体的作用(图119A-119D)。申请人发现,催化核心中的突变(V351G、K350I)以及与RNA靶标接触的突变(S486A、S495N)对于RESCUE 活性很重要。纯化的ADAR2dd上的RESCUE突变的生化表征显示对 dsDNA、ssDNA或DNA-RNA异源双链体无活性,进化后的突变在体外改善了dsRNA底物上的C至U编辑的动力学(图120A-120D)。

[2172] 由于ADAR2已在无Cas13的其他RNA编辑平台上使用(8、9、11、13),因此申请人在不存在Cas13融合的情况下测定了C至U的活性。申请人将RESCUE突变引入到哺乳动物细胞中与指导RNA一起的 ADAR2dd或全长ADAR2蛋白中,并测定了这些构建体恢复荧光素酶活性的能力,发现完整的RESCUE构建体(包括指导RNA正向重复)对于腺苷和胞苷脱氨基化活性都是必需的(图103D、图121A-121D、122A-122C、123A-123C)。为了在替代RNA编辑系统中测试C至U编辑,该系统依赖于MS2-ADAR2dd融合(24)的募集或用RNA指导进行的全长ADAR2募集(11、24),申请人将RESCUE突变引入到这些构建体中并发现与基于Cas13b的RESCUE相比,编辑效率显著降低(图124A-124F)。

[2173] 申请人接下来通过细胞群的批量测序评估了RESCUE在HEK293FT 细胞中对于内源转录物的效率。申请人测试了在9个基因中的24个不同位点以及来自ClinVar的24个合成疾病相关突变靶标的多种指导序列设计,发现编辑率高达42%(图103E、图125A-125C、126A-126B、127A-127B、130;表28)。在所测试的指导序列中(表29-31),申请人发现了多种指导序列设计规则,其中最明显的是涉及基序的特征(优选5' U或A)和指导序列错配位置。

[2174] 为了展示通过翻译后修饰位点的RNA编辑来控制信号通路,申请人通过调节关键磷酸化残基来改变STAT和Wnt/ $\beta$ -连环蛋白通路的激活(图104A、129A-129F)。对 $\beta$ -连环蛋白上的磷酸化残基(例如S33、S37和 T41)进行突变抑制了泛素化和降解,从而允许蛋白质与LEF和TCF1/2/3 等转录因子接合并导致细胞增殖增加(25)(图104B)。申请人测试了一组在已知被磷酸化的残基处靶向 $\beta$ -连环蛋白转录物(CTNNB1)的指导序列并观察到5%至28%之间的编辑水平(图104C),导致Wnt/ $\beta$ -连环蛋白信号传导的多达5倍激活(图104D)以及在HEK293FT(图104E-104F)和人脐静脉内皮细胞(HUVECs)(图130A-130B)中增加的细胞生长。由于使用RESCUE的治疗应用可受益于用于病毒递送的较短构建体,因此申请人还用dRanCas13b的C端截短评估了RESCUE活性并发现相似或改善的脱氨酶活性(图131)。

[2175] 由于RESCUE保留了腺苷脱氨酶活性(图118A-118D),因此 Cas13b(4)的天然前crRNA加工活性使得能够进行多重腺苷和胞苷脱氨基化。通过递送RESCUE以及靶向CTNNB1转录物中的腺苷和胞苷的前 crRNA(图105A),申请人发现RESCUE可以分别以~15%和5%的比率编辑被靶向的残基S33F和T41A两者(图105B)。但是,在这些实验以及单重试验中,申请人发现被靶向的胞苷附近的A至I脱靶(图132A-132C、133A-133D)。为了消除这些脱靶,申请人在脱靶腺苷对面的指导序列中引入了不利的鸟嘌呤错配(图105C),显著减少了脱靶编辑,同时最小程度地干扰中靶编辑(图105D)。

[2176] 申请人用全转录组RNA测序对脱靶进行了谱分析,发现虽然 RESCUE在Gluc转录物上有~80%C至U编辑(图106A),但其有188C至U脱靶和1,695A至I脱靶,与REPAIRv1(7)的A至I脱靶是可比较的(图108A、108B)。为了提高RESCUE的特异性,申请人对ADAR2dd 在与RNA靶标相互作用的残基处进行了合理诱变(图106C),从而产生多个RESCUE突变体,其如通过

荧光素酶报告基因(图106D)和RNA 测序(图106E-106G)测量的,具有降低的A至I脱靶活性和高C至U 中靶脱氨活性。RESCUE上的最高特异性突变体S375A(以下简称 RESCUE-S)保持约76%中靶C至U编辑(图106E),但仅有103C至 U脱靶和139A至I脱靶,腺嘌呤脱氨化脱靶的数量减少约10倍(图 106E-106G),具有减少的错义突变和差异调控的转录物(图134A-134F、135A-135C、136A-136B、137A-137D)。申请人还发现RESCUE-S在许多内源位点保留了与RESCUE类似的C至U活性,甚至超过其在一些位点处(图138A-138C、139A-139C、140A)具有在局部指导序列窗口内更高的特异性(图138C、140B-140E)。

[2177] RESCUE是一种可编程的碱基编辑工具,其能够将RNA中的胞苷精确转化为尿苷。使用定向进化,申请人展示了腺苷脱氨酶可以被放宽以接受其他碱基,从而产生了可以编辑dsRNA的新型胞苷脱氨化机制。尽管在本研究中,申请人利用了Cas13的RNA指导靶向机制,但其他 RNA靶向机制(8-15、24)也可以与进化的ADAR2dd突变体类似地结合以在RNA转录物上实现精确的胞苷脱氨化。RESCUE的胞苷脱氨化活性的较大可靶向氨基酸密码子空间使得能够调节更多的翻译后修饰(例如磷酸化、糖基化和甲基化)以及扩大对常见催化残基的靶向性(图 107和141)。此外,胞苷脱氨酶介导的RNA编辑允许对疾病相关突变的额外靶向,并产生保护性等位基因,例如ApoE2。总而言之,RESCEU 扩展了具有新碱基编辑功能的RNA靶向工具包,从而允许扩展对遗传疾病的建模和潜在治疗。

[2178] 参考文献

[2179] 1.0.0.Abudayyeh等,C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.Science 353,aaf5573 (2016) .

[2180] 2.C.Cassidy-Amstutz等,Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin.Biochemistry 55,3461-3468 (2016) .

[2181] 3.S.Shmakov等,Discovery and Functional Characterization of Diverse Class 2CRISPR-Cas Systems.Mol Cell 60,385-397 (2015) .

[2182] 4.A.A.Smargon等,Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28.Mol Cell 65,618-630e617 (2017) .

[2183] 5.A.East-Seletsky等,Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection.Nature 538,270-273 (2016) .

[2184] 6.0.0.Abudayyeh等,RNA targeting with CRISPR-Cas13.Nature 550, 280-284 (2017) .

[2185] 7.D.B.T.Cox等,RNA editing with CRISPR-Cas13.Science 358, 1019-1027 (2017) .

[2186] 8.T.Merkle等,Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides.Nat Biotechnol 37,133-138 (2019) .

[2187] 9.P.Vogel等,Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs.Nat Methods 15,535-538 (2018) .

[2188] 10.M.Fukuda等,Construction of a guide-RNA for site-directed RNA mutagenesis utilizing intracellular A-to-I RNA editing.Sci Rep 7,41478 (2017) .

[2189] 11.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) .

[2190] 12.M.F.Montiel-Gonzalez,I.C.Vallecillo-Viejo,J.J.Rosenthal,An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* 44,e157 (2016) .

[2191] 13.P.Vogel,M.F.Schneider,J.Wettengel,T.Stafforst,Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA. *Angew Chem Int Ed Engl* 53,6267-6271 (2014) .

[2192] 14.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) .

[2193] 15.H.A.Rees,D.R.Liu,Base editing:precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet* 19,770-788 (2018) .

[2194] 16.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) .

[2195] 17.K.Nishida等,Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353,(2016) .

[2196] 18.J.D.Salter,R.P.Bennett,H.C.Smith,The APOBEC Protein Family: United by Structure,Divergent in Function. *Trends Biochem Sci* 41,578-594 (2016) .

[2197] 19.S.Jin等,Cytosine,but not adenine,base editors induce genome-wide off-target mutations in rice. *Science*, (2019) .

[2198] 20.E.Zuo等,Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*, (2019) .

[2199] 21.J.Grünwald等,Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature*, (2019) .

[2200] 22.M.R.Macbeth等,Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 309,1534-1539 (2005) .

[2201] 23.M.M.Matthews等,Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nature structural & molecular biology* 23,426-433 (2016) .

[2202] 24.D.Katrekar等,In vivo RNA editing of point mutations via RNA-guided adenosine deaminases. *Nat Methods* 16,239-242 (2019) .

[2203] 25.B.T.MacDonald,K.Tamai,X.He,Wnt/beta-catenin signaling: components, mechanisms,and diseases. *Dev Cell* 17,9-26 (2009) .

[2204] 26.M.K.Chee,S.B.Haase,New and Redesigned pRS Plasmid Shuttle Vectors for Genetic Manipulation of *Saccharomyces cerevisiae*. *G3 (Bethesda)* 2, 515-526 (2012) .

[2205] 27.M.F.Laughery等,New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*.*Yeast* 32,711-720 (2015).

[2206] 28.M.R.Macbeth,B.L.Bass,Large-scale overexpression and purification of ADARs from *Saccharomyces cerevisiae* for biophysical and biochemical studies.*Methods Enzymol* 424,319-310 (2007).

[2207] 29.H.Ng,N.Dean,Dramatic Improvement of CRISPR/Cas9 Editing in *Candida albicans* by Increased Single Guide RNA Expression.*mSphere* 2, (2017).

[2208] 30.R.Heim,D.C.Prasher,R.Y.Tsien,Wavelength mutations and posttranslational autoxidation of green fluorescent protein.*Proc Natl Acad Sci U S A* 91,12501-12504 (1994).

[2209] 31.Y.Wang,P.A.Beal,Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method.*Nucleic Acids Res* 44,9872-9880 (2016).

[2210] 32.R.D.Gietz,R.H.Schiestl,Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method.*Nat Protoc* 2,38-41 (2007).

[2211] 33.M.T.Veeman,D.C.Slusarski,A.Kaykas,S.H.Louie,R.T.Moon, Zebrafish prickle,a modulator of noncanonical Wnt/Fz signaling,regulates gastrulation movements.*Curr Biol* 13,680-685 (2003).

[2212] 材料与方法

[2213] 酵母构建体的设计和克隆

[2214] 为了在酵母中表达dRanCas13b-hADAR2dd构建体,通过修饰pML104 (Addgene# 67638) (27) 将该融合蛋白克隆到pRSII426主链 (26) 中pGAL启动子的下游。为了提高表达,在该融合蛋白之间克隆GS 接头,并针对酵母对ADAR2dd进行了密码子优化 (28)。通过Gibson 克隆引入对应于RESCUE的轮次的额外密码子突变。

[2215] 对用于测试酵母中活性的靶向质粒进行工程化以用于荧光筛选 (GFP) 和营养缺陷选择筛选 (His)。将所有靶向质粒克隆到pYES3/CT 主链 (Thermo Scientific)。所有质粒包含用于RESCUE的RanCas13b指导盒,其表达由ADH1启动子驱动,以及侧接HH和HDV核酶的间隔子和DR序列 (29)。克隆了用金门位点替代间隔子的构建体以便利模块指导序列克隆。

[2216] 为了产生C至U RNA编辑活性的GFP指示剂,将Y66H绿色至蓝色突变 (30) 引入由TEF启动子驱动的酵母密码子优化的EGFP (yeGFP) (31) 中。成功的C至U RNA编辑恢复该构建体的绿色荧光。通过测试 HIS3中的保守残基当突变为可通过RNA编辑来挽救的残基时的活性丧失来生成用于C至U编辑的His报告基因 (图128)。将生成失活HIS3 的突变在其天然HIS3启动子下克隆到pYES3/CT主链中的HIS3基因中。

[2217] 表33列出了所有酵母质粒,表34列出了酵母实验中使用的所有靶向指导序列。

[2218] RESCUE定向进化

[2219] 为了针对酵母中的C至U活性进行选择,申请人基于GFP荧光的恢复或HIS营养缺陷型选择基因的原养型逆转设计了一组酵母报告基因检测方法。对在没有组氨酸的情况下存活的GFP阳性培养物或菌落进行测序,选择了ADAR2dd结构域中的个体突变,将其引入先前RESCUE候选轮中并使用各种报告基因构建体评估其在哺乳动物细胞中的活性。在优化UCG

荧光素酶位点 (C82R) 上的荧光素酶活性11轮后, 申请人转向在CTNNB1转录物上的T41位点处进行了两轮优化, 然后又在Gluc 转录物上的CCU位点 (L77P) 处进行了两轮优化。在最后一轮中, 申请人测试了荧光素酶突变体的活性的恢复, 该荧光素酶突变体具有在Gluc C82R位点的所有四个可能5'碱基 (UCG、ACG、CCG和GCG) 以及在 Gluc L77P突变处的两个额外基序 (CCU和CCA), 发现了在这些基序下的活性增加 (图103B、图110)。为了在我们的酵母系统中进一步验证来自定向进化管线的我们的RESCUEr版本, 申请人测试了多个 RESCUEr迭代在酵母和体外试验两者中的活性 (图113A-113E和 120A-120D)。在酵母中检测EGFP和His恢复两者, 申请人发现RESCUEr 的最新版本可以在两个靶标上更有效地执行C至U编辑 (图113A-113E)。在每轮酵母筛选后, 在一系列哺乳动物报告基因上评估了居首位的突变体以验证活性并选择居首位的突变体用于下一轮酵母筛选。表27列出了所有筛选和所得突变。

#### [2220] 用于酵母筛选的诱变文库的生成

[2221] 为了生成用于在酵母系统中筛选突变的诱变文库, 使用Genemorph II (Agilent Technologies) 对hADAR2脱氨酶结构域进行了突变, 用于通过两倍稀释系列, 在模板输入范围为74ng-9.4 $\mu$ g的8个50 $\mu$ L反应中进行易错PCR。扩增后, 将反应物合并, 在去离子水中以1:4稀释, 并加载到含有溴化乙锭的2%凝胶中。使用MinElute PCR纯化试剂盒 (Qiagen) 纯化提取的样品, 然后在37 $^{\circ}$ C下用DpnI (Thermo Fisher Scientific) 处理 2h以去除残留的模板质粒, 然后进行凝胶和MinElute纯化。通过用KfII、RruI和Eco72I (Thermo Fisher Scientific) 消化7 $\mu$ g模板质粒1小时来生成主链。用MinElute PCR纯化试剂盒对消化物进行凝胶纯化, 并在30 $\mu$ L 的预加温水中洗脱。

[2222] 使用Gibson Assembly (New England Biosciences) 组装经纯化的PCR 插入序列和经消化的主链, 将456ng PCR插入序列和800ng主链消化物在80 $\mu$ L反应中孵育1小时。用异丙醇沉淀将产物沉淀, 并通过加热至 50 $^{\circ}$ C 5分钟重悬于12 $\mu$ L Tris-EDTA缓冲液中。将50 $\mu$ L Endura电感受态细胞 (Lucigen) 在冰上解冻10分钟, 然后加入2 $\mu$ L重悬的Gibson产物。遵循最佳Endura设置 (1.0mm吸收池、10 $\mu$ F、600欧姆和1800伏), 使用GenePulser Xcell (Bio-Rad) 对混合物进行电穿孔。将来自每个电穿孔的样品回收到1mL的回收培养基 (Lucigen) 中, 并在37 $^{\circ}$ C下孵育1 小时同时在300r.p.m摇动。每个诱变文库进行两次电穿孔。将回收的培养物铺板在大型预加温100 $\mu$ g/mL氨苄青霉素板上, 并将板在37 $^{\circ}$ C下孵育16小时, 然后使用Nucleobond Xtra Maxi Kit (Macherey-Nagel) 收获。

#### [2223] 酵母中诱变文库的转化

[2224] 所有酵母实验均使用INVScI (ThermoFisher Scientific) 进行。如前所述进行大规模酵母转化 (32)。简而言之, 将含有Y66H EGFP或HIS3 报告质粒的菌落挑到300mL-Trp2%葡萄糖选择培养基中并在30 $^{\circ}$ C生长过夜。生长后, 测定细胞的OD600, 并将2.5e9细胞加入到500mL预加温的2 $\times$ YPAD中, 并在30 $^{\circ}$ C孵育4小时。洗涤细胞沉淀多次, 然后重悬于36mL转化混合物中, 该混合物包含24mL PEG 3350 (50%w/v)、3.6mL的1.0M乙酸锂、5mL的2.0mg/mL的变性单链载体鲑鱼精子DNA (ThermoFisher Scientific)、2.9mL水、和500 $\mu$ L的1 $\mu$ g/ $\mu$ L质粒文库。在42 $^{\circ}$ C下孵育60分钟后, 将细胞沉淀重悬于750mL的-Ura/-Trp 2%葡萄糖选择培养基中, 并生长过夜直至培养物达到OD600为5-6。此时, 将6mL的培养物接种到250mL的2%棉子糖-Ura/-Trp选择培养基并孵育直至OD600为0.5-1。加入27mL的30%半乳糖来诱导培养物, 并在 30 $^{\circ}$ C下孵育过夜12-14小时。然后对细胞进行细胞分选或铺板到选

择板上,如下所述。任何涉及单个突变体的验证实验都以类似的方式进行转化,但是使用上述大规模转化的按比例缩小版本。

#### [2225] 酵母文库的荧光细胞分选

[2226] 诱导后,通过与阴性非诱导和非靶向指导序列对照相比对EGFP荧光进行门控,在SH800S细胞分选仪上对细胞进行分选。在将1亿个细胞分选到2%葡萄糖-Ura/-Trp选择培养基中后,将分选的细胞孵育过夜,然后在OD600为5-6以1:40稀释到2%棉子糖-Ura/-Trp选择培养基中。将细胞放回振荡器中,在OD600在0.5-1之间用半乳糖诱导,孵育过夜 12-14小时,然后再次分选。进行分选直至分选了1-2千万个细胞。迭代生长和分选又重复了2-3次,用Zymoprep Yeast Plasmid Miniprep II (Zymo)收获分选细胞的每次迭代的质粒。通过Illumina NextSeq NGS 对质粒的Adar2dd区进行PCR扩增和测序以确定突变体在每一轮选择中的存在。单独订购最富集的突变体,并按如下所述进行克隆用于哺乳动物验证测试。

#### [2227] 酵母文库的His生长选择

[2228] 诱导后,将细胞文库铺板在2%棉子糖/3%半乳糖-Ura/-Trp/-His选择板上。随着菌落生长,将它们挑到水中并在2%棉子糖/3%半乳糖-Ura/-Trp/-His选择板上划线。条纹生长过夜后,对每个条纹进行菌落PCR 并对ADAR2催化结构域以及His基因进行桑格测序以检查重组和DNA 诱变。单独订购突变并按如下所述进行克隆用于哺乳动物验证测试。

#### [2229] 设计和克隆用于RNA编辑的哺乳动物构建体

[2230] 通过在HEPN结构域的催化位点处的组氨酸至丙氨酸和精氨酸至丙氨酸突变(R142A/H147A/R1039A/H1044A)使RanCas13b催化失活(dRanCas13b)。合成脱氨酶结构域和ADAR2并进行PCR扩增以Gibson 克隆到pcDNA-CMV载体主链中,并通过GS-mapkNES-GS(GSSLQKKLEELG(S)SEQ ID NO:779))接头在C端处与dRanCas13b 融合。将用于改变胞嘧啶脱氨酶活性或特异性的ADAR2脱氨酶结构域中的突变体通过Gibson克隆引入到dRanCas13b-GS-mapkNES-GS-ADAR2dd主链中。表27列出了ADAR2dd 中引入的所有用于进化C至U编辑的突变。

[2231] 为了在不同的Cas13b直向同源物之间进行比较,通过Gibson克隆到REPAIR构建体dPspCas13b-GS-HIVNES-GS-ADAR2dd上(7),在 dRanCas13b主链上测试的突变被转移至dPspCas13b融合载体。为了单独测试不含dRanCas13b的ADAR2dd和全长ADAR2,申请人使用Gibson 克隆将所有突变添加到具有ADAR2dd或全长ADAR2的pcDNA-CMV载体主链上,之前已克隆以测试REPAIR(7)。通过筛选先前报道的荧光素酶报告质粒中Gluc中的潜在突变,生成了用于测量C至U RNA编辑活性的荧光素酶报告载体(7)。该报告载体将功能性Cluc表达为标准化对照,但是由于添加了突变体(C82R或L77P)而导致了缺陷型Gluc。为了测试RESCUE编辑基序偏好,申请人克隆了在Gluc的密码子82(AAX CXC)处的胞嘧啶周围的每个可能基序。评价了突变体的C82R的C至U编辑和催化活性的恢复(33)。由于周围的基序强烈地决定了用于A至I编辑的RNA编辑效率(7),申请人最初靶向UCG位点,因为5' U和3' G是ADAR2dd最佳活性的优选侧翼碱基。从M50 Super 8x TOPFlash(Addgene#12456)和M50 Super 8x FOPFlash(Addgene#12457)(34)生成了用于测试CTNNB1编辑效率的分泌的荧光素酶报道载体。通过Gibson克隆,在TCF/LEF应答元件(TOPFlash)或模拟结合位点(TOPFlash)的控制下,将最初的萤火虫荧光素酶替换为分泌型高斯荧光素酶。克隆进了由CMV启动子驱动表达的额外的海萤荧光素酶充当转染对照。表32列出了所有哺乳动物质粒。

[2232] 哺乳动物细胞中候选轮次的选择

[2233] 选择表现与现有候选轮次相当或更好的突变,用于在6个荧光素酶报告基因的整个组中进行筛选。对于RESCUEr4至RESCUEr10的选择,首先在TCG基序上筛选候选突变;使用GCG基序作为初筛分离了候选轮次RESCUEr11。使用对内源性CTNNB1的T41I残基的编辑进行的初筛,在哺乳动物细胞中验证了候选轮次RESCUEr12至RESCUEr14的选择,导致 $\beta$ -连环蛋白通路激活(其用通路活性的发光报告基因进行了谱分析),以及通过对Gluc的L77P CCT基序的活性来选择候选轮次 RESCUEr15和RESCUEr16。表27列出了所有轮次和用于产生它们的酵母筛选。

[2234] 克隆致病性U>C突变用于测定RESCUE活性

[2235] 为了产生用于测试REPAIR活性的疾病相关突变,选择了ClinVar 中定义的与疾病发病机理相关的23个U>C突变(独立分组为一组22个基因和ApoE)。从Integrated DNA Technologies订购了选择的靶标作为突变位点周围的200-bp区域,并在Efl $\alpha$ 启动子下克隆到mScarlet的下游。

[2236] 用于RESCUE的指导序列克隆

[2237] 为了表达用于RESCUE的哺乳动物指导RNA,使用先前描述的具有前面是在U6表达下的金门受体位点的RanCas13b正向重复序列的构建体(7)。通过金门克隆将个体指导序列克隆到该表达主链中。为了确定用于选择位点的最佳指导序列,测试了C和U翻转两者以及在最常见的最佳指导序列范围(错配距离为 $\sim$ 24)附近的平铺指导序列。表29-31列出了用于RESCUE实验的指导序列。

[2238] 哺乳动物细胞培养

[2239] 除非另有说明,否则哺乳动物细胞培养实验是在HEK293FT细胞系(American Type Culture Collection(ATCC))中进行,在含葡萄糖、丙酮酸钠和GlutaMAX(Thermo Fisher Scientific)、补充了1 $\times$ 青霉素-链霉素(Thermo Fisher Scientific)和10%胎牛血清(VWR Seradigm)的Dulbecco's Modified Eagle培养基中生长。细胞保持在低于80%的汇合度。

[2240] 除非另有说明,否则所有转染均用Lipofectamine 2000(Thermo Fisher Scientific)在涂覆聚D-赖氨酸(BD Biocoat)的96孔板上进行。转染前16小时,将细胞以约20,000个细胞/孔铺板以确保转染时90%的融合度。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)混合至总计25 $\mu$ l。分别地将24.5 $\mu$ l的Opti-MEM与0.5 $\mu$ l的Lipofectamine 2000混合。然后将质粒和脂质体转染溶液合并并孵育5分钟,然后将它们移液到细胞上。

[2241] 将HUVEC细胞(Lonza)在内皮生长培养基-2(Lonza)中在Nunc Collagen I Coated EasYFlasks(Thermo Fisher Scientific)上培养。细胞保持在低于80%的汇合度。HUVEC转染用Lipofectamine LTX(Thermo Fisher Scientific)在胶原I(BD Biocoat)涂覆的96孔板中进行。转染前16小时,将细胞以约5,000个细胞/孔接种。转染前立即用新鲜的EGM-2替换培养基。对于板上的每个孔,将转染质粒与1 $\mu$ l Plus试剂和Opti-MEM 混合至总计25 $\mu$ l。分别地将24.7 $\mu$ l的Opti-MEM与0.3 $\mu$ l的Lipofectamine LTX混合。然后将质粒和LTX溶液合并并孵育25分钟,然后将它们移液到细胞上。4小时后,用PBS洗涤细胞并用新鲜的EGM-2替换培养基。

#### [2242] 哺乳动物细胞中的RESCUE编辑

[2243] 为了评估哺乳动物细胞中的RESCUE活性,申请人转染了150ng RESCUE载体、300ng 指导表达质粒,和当使用报告基因时(荧光素酶、STAT活性或 $\beta$ 连环蛋白活性)40ng RNA编辑报告基因。48小时后,从细胞中收获RNA并使用先前所述的方法(33)用基因特异性逆转录引物进行逆转录。然后使用NEBNext High-Fidelity 2X PCR Master Mix(New England Biolabs)将提取的cDNA进行两轮PCR,以添加Illumina接头和样品条码。然后将文库在Illumina NextSeq或MiSeq上进行下一代测序。然后评估在测序窗口内的所有腺苷处的RNA编辑率。

[2244] 在靶向荧光素酶报告基因用于RNA编辑的实验中,申请人还在RNA收获之前收获具有分泌的荧光素酶的培养基。申请人在具有注射方案的读板仪(Biotek Synergy Neo2)上使用海萤和高斯荧光素酶测定试剂盒(Targeting Systems)测量荧光素酶活性。进行的所有重复都是生物学重复。

[2245] 在RESCUE质粒的输入量发生变化的实验中,除非另有说明,否则通过用表达CMV驱动的mScarlet的填充质粒替换RESCUE表达质粒,使质粒的总量保持恒定。在指导质粒的输入量发生变化的实验中,通过非靶向指导序列的替换使质粒的总量保持恒定(“具有填充质粒”),或不保持恒定(“不具有填充质粒”);在该实验中,没有用于RESCUE质粒的填充质粒。

#### [2246] RESCUE指导序列设计的考虑

[2247] 申请人测试了一组具有不同错配位置的指导RNA,其靶向9个基因中的24个不同位点(表103E、125A-125C),特别选择了不同的5'碱基身份来询问不同基序上的脱氨基化活性。申请人发现,在所有测试的位点上,RESCUE实现高达35%的编辑率,并且理想的错配位置或碱基翻转(C或U)是位点依赖性的。此外,RESCUE在多个内源位点上的性能均优于前面所有轮次的突变体并且需要比早期版本更少的转染质粒(图126A-126B)。为了更好地评估RESCUE与治疗的相关性,申请人设计了一系列24个靶标,以对来自ClinVar的疾病相关突变的编辑进行建模(参见表28),并发现通过批量测序测得的高达42%的编辑率(图129A-129B),包括阿尔兹海默症风险相关的ApoE4等位基因(图128)。

[2248] 在分析了本文中的所有指导序列后,申请人发现最佳的指导序列设计在靶标位点之间有所不同。申请人建议针对每个新靶标位点测试各种指导序列设计,包括C翻转和U翻转两者以及不同的错配位置。待测的设计的实例将包括30nt的指导序列,其具有C或U翻转和在以下位置的错配:28、26、24、22和20。总而言之,申请人发现任何侧接U或A的胞苷位点均具有强大的编辑活性。具有5' C或G的位点将以较低的效率进行编辑。

#### [2249] ADAR2dd上RESCUE突变的生化表征

[2250] 为了评估包含RESCUE突变的hADAR2脱氨酶结构域的动力学活性,将多次迭代克隆到包含URA3基因的pGAL-His6-TwinStrep-SUMO-hADAR2dd主链中。将质粒转化到BCY123感受态酵母细胞中(10)。简而言之,将冷冻细胞在37°C水浴中解冻15-30秒。将每种条件下的10 $\mu$ L细胞在微量离心机中以13,000g离心2分钟,然后去除上清液。为每个构建体制备的转化混合物包含以50%w/v制备的260 $\mu$ L PEG 3350、50 $\mu$ L变性鲑鱼精子(Thermo Fisher Scientific)、36 $\mu$ L 1M乙酸锂和14 $\mu$ L DI H<sub>2</sub>O中的750ng质粒。将酵母沉淀物用转化混合物重悬,并在42°C水浴中温育30分钟,然后以13,000离心30秒,然后去除上清液。然后将沉淀重悬于1mL DI H<sub>2</sub>O中,并取50 $\mu$ L至1mL DI H<sub>2</sub>O中混合。随后,将200 $\mu$ L铺板到减去尿嘧啶的

基本葡萄糖平板上进行原养型选择。

[2251] 将平板在30℃下孵育48小时,然后将单个菌落接种到10mL的补充有右旋糖(20g/L)的酵母基本培养基中。基本培养基用酵母脱落补充剂Y2001(1.39g/L)、不含氨基酸的酵母氮基础(6.7g/L)、腺嘌呤半硫酸盐(0.022g/L)、组氨酸(0.076g/L)、亮氨酸(0.38g/L)和色氨酸(0.076g/L)制备。培养物生长过夜,然后将全部10mL培养物接种到100mL基本培养基/右旋糖培养物中。生长8小时后,将每种构建体接种到两个2L烧瓶中,其含有补充有20g 棉子糖(VWR)的1L基本培养基。将这些生长过夜,并通过添加溶于200mL基本培养基的30g 半乳糖诱导;然后在收获前将培养物再培养8小时。将培养物在Beckman Coulter Avanti J-E离心机上以5,000r.p.m旋下20分钟,所得沉淀物保存在-80℃下。

[2252] 根据Macbeth和Bass(28)中描述的方案修改了不同的RESCUE候选hADAR2脱氨酶结构域的蛋白质纯化。简而言之,将5-10g冷冻酵母沉淀重悬于补充有一片无EDTA的微型cOmplete ULTRA蛋白酶抑制剂(Sigma)的50mL裂解缓冲液(20mM TrisHCl pH 8、5%甘油、750mM NaCl、1mMβ-巯基乙醇、0.01% Triton-X)中。使悬浮液在25,000psi下通过LM20微流化器七次,并通过在9500RPM下离心80分钟来沉淀细胞碎片。倾析出澄清的裂解物,并与1mL StrepTactin超流动树脂(Qiagen)一起孵育2.5小时,在4℃下使用旋转振荡器轻轻摇动。将悬浮液添加到用裂解缓冲液预平衡的Econo柱色谱柱中,并用40mL裂解缓冲液洗涤树脂。三次后续洗涤(每次40mL)降低了盐浓度(500mM, 250mM、然后100mM NaCl)。通过在台式摇床上在20mL补充了100μg SUMO蛋白酶(内部)的裂解缓冲液中轻轻摇动过夜,将蛋白质从树脂上切割下来。收集流出物并与用裂解缓冲液从树脂洗涤的3x 5mL洗涤物合并。将包含经切割的蛋白的全部级分上样至5mL肝素HP阳离子交换柱(GE Healthcare Life Sciences),并在100mM至1M的NaCl梯度上洗脱(缓冲液20mM Tris-HCl pH 8、5%甘油、1mMβ-巯基乙醇,具有相应的NaCl浓度)。检查级分的纯度,并使用SDS-PAGE和考马斯亮蓝染色进行分析,合并含蛋白质的级分并使用10MWC0离心过滤器(Amicon)浓缩。通过考马斯染色和针对BSA连续稀释液(从1mg/mL开始)的SDS-PAGE电泳来测定每种蛋白质的mg/mL浓度。使用ImageLab软件(BioRad Image Lab Software 6.0.1)定量条带,并通过BSA标准品的线性回归进行内插来估算浓度。

[2253] 将具有DNA手柄的ssRNA和DAN寡核苷酸(Integrated DNA Technologies)在1x双链体缓冲液(HEPES 30mM pH 7.5、K+乙酸盐100 mM)中于85℃退火5分钟,缓慢升温至4℃,然后使用Oligo Clean& Concentrator(Zymo)纯化,用Nanodrop定量,并标准化为100ng/μL。

[2254] 如前所述(23)进行体外测定,但稍有改动。在25nM RNA底物、50nM ADAR蛋白和0.16U/μL RNase抑制剂,以及15.6mM NaCl在1x 测定缓冲液(17mM TrisHCl pH 7.5、5%甘油、1.6mM EDTA、0.003% NP-40、0.5mM TCEP)中建立测定。将20μL反应物(三个技术重复)在30℃下孵育一定范围的时间点(0、5、10、30和60分钟)。加入10μL的0.5% SDS溶液(至总浓度为0.166% SDS)淬灭反应,并在95℃变性5分钟。

[2255] 使用RNA XP清洁珠(Beckman Coulter)从反应混合物中纯化RNA,其中磁珠和异丙醇与样品体积的比率分别为10:3和3:1。根据制造商说明书使用qScript Flex cDNA试剂盒,带有修改,对经纯化的RNA进行逆转录。具体而言,将12.85μL纯化RNA与2μL GSP增强子和0.15μL的100μM RT引物组合,通过涡旋混合并在65℃下孵育5分钟,然后进入42℃维持。

此时,向每个反应中加入4 $\mu$ L qScript flex reaction mastermix (5x)和1 $\mu$ L qScript RT,并通过移液混合,然后在42 $^{\circ}$ C孵育1小时,然后在85 $^{\circ}$ C加热5分钟。用两轮PCR扩增以添加Illumina接头和条码以制备cDNA用于测序,并在Illumina NextSeq上测序。测定在测序窗口内的所有胞苷(C至U活性)和腺苷(A至I活性)处的体外RNA编辑率。

#### [2256] 全转录组测序以评估ADAR编辑特异性

[2257] 为了分析转录组中的脱靶RNA编辑位点,转染后48小时,使用 RNeasy Plus Miniprep试剂盒(Qiagen)从细胞中收集总RNA。然后使用 NEBNext Poly(A) mRNA磁性分离模块(NEB)富集mRNA级分,然后使用NEBNext Ultra RNA Library Prep Kit for Illumina(NEB)制备此RNA用于测序。然后将文库在Illumina NextSeq上测序并加载,使得有每个样品至少500万个读取结果。

#### [2258] 靶向和全转录组实验的RNA编辑分析

[2259] 使用开发用于此公布的自定义工作流程: [portal.firecloud.org/#methods/mrna\\_editing\\_final\\_workflow/rna\\_editing\\_final\\_workflow/1](https://portal.firecloud.org/#methods/mrna_editing_final_workflow/rna_editing_final_workflow/1),在FireCloud计算框架([software.broadinstitute.org/firecloud/](https://software.broadinstitute.org/firecloud/))上进行全转录组编辑RNA测序数据的分析。

[2260] 为了进行分析,除非另有说明,序列文件被随机降采样至500万个读取结果。使用添加了Gluc和Cluc序列的RefSeq GRCh38组件生成了索引,并使用Bowtie/RSEM版本1.3.0对读取结果进行比对和定量。然后使用REDitools(35,36)对比对BAM进行RNA编辑位点的分选和分析,参数如下:-t 8-e-d-l-U[AG或TC或CT或者GA]-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或转染伪影,并从脱靶分析中滤除。如果通过Benjamini Hochberg 校正进行多个假设校正后,Fisher精密检验得出小于0.05的p值,并且 3个生物学重复中至少有2个识别了编辑位点,则认为脱靶是显著的。相对于最大可能重叠,计算样品之间的编辑重叠,相当于两个样本之间的编辑的较少次数。重叠编辑位点的百分比计算为共享编辑位点的数量除以两个样品的最小编辑数量再乘以100。使用Kaviar(37)方法对已知 SNP位置进行额外一层过滤以识别SNP。

#### [2261] RNA编辑的差异基因表达分析

[2262] 基于人hg38 UCSC基因组和RefSeq转录组创建Bowtie指数。接着,运行RSEM v1.3.157,命令行选项为"--estimate-rspd-bowtie-chunkmbs 512 -paired-end",以使用Bowtie将成对末端读取结果直接与该索引比对,并基于比对结果估计每百万转录物(TPM)中的表达水平。为了分析转录组变化,如果RESCUE或GFP对照条件的平均TPM大于1,则认为检测到了转录物。进行了学生t检验以识别p值通过0.01FDR校正的差异表达同种型。

#### [2263] Stat表型分析

[2264] 如上所述使用lipofectamine 2000,用RESCUE质粒、靶向STAT3 和STAT1上的残基的指导质粒、以及用于STAT3(Qiagen Cignal STAT3 Reporter)和STAT1信号传导(Qiagen Cignal GAS Reporter)的荧光素酶报告基因转染细胞并孵育48小时。48小时后,使用Dual-Glo荧光素酶测定法(Promega)来测量细胞中的萤火虫和海肾荧光素酶活性。将萤火虫信号相对于海肾信号进行标准化以测量STAT3和STAT1的相对激活。

#### [2265] $\beta$ -连环蛋白表型测定

[2266] 转染前24小时将细胞铺板在含有防止细胞在孔中心生长的核的细胞迁移板上。24小时后,如上所述使用lipofectamine 2000,用RESCUE质粒、靶向 $\beta$ -连环蛋白上的残基的指

导质粒、以及用于 $\beta$ -连环蛋白激活的荧光素酶报告基因(Qiagen TCF/LEF Signal Reporter)转染细胞并孵育。24小时后,去除中心核以允许细胞向孔中心生长。再孵育24小时后,测定培养基中的Gluc和Cluc荧光素酶信号。计算Gluc与Cluc的相对比率,以确定条件之间的相对 $\beta$ 连环蛋白激活。在第3天,将细胞与 CellTracker Green CMFDA Dye (ThermoFisher Scientific)一起孵育10分钟,然后用培养基洗涤。每天使用荧光对细胞成像以测量细胞生长。使用ImageJ软件通过计算中心生长区域中荧光的总面积来测量细胞向孔中心区域中的细胞生长。使用自动宏通过以下命令处理图像:

[2267] //ImageJ macro for calculating cellular area run("8-bit");

[2268] run("Auto Local Threshold","method=Bernsen radius=15 parameter\_1=0parameter\_2=0white");

[2269] setAutoThreshold("Default dark");run("Measure");

[2270] 连环蛋白迁移测定(HUVEC)

[2271] 转染前16小时,将HUVEC铺板在胶原I涂覆的细胞迁移板上。在上述转染方案中使用了100ng包含RESCUE构建体和指导序列的单一载体。24小时后,去除中心核并用补充有氢化可的松、hFGF-B、FBS、抗坏血酸、GA-1000和来自EGM-2补充包(Lonza)的肝素的内皮基础培养基2(Lonza)代替培养基。在第3天,将细胞与在EBM-2中稀释的 CellTracker Green CMFDA Dye一起孵育10分钟,然后用培养基洗涤。每天使用荧光对细胞成像。使用ImageJ软件通过手动概括和定量每个孔中的无细胞面积来测量细胞生长。

[2272] 表27.RESCUE进化表

[2273]

| RESCUE<br>候选轮次 | 突变               | 产生候选突变的筛选<br>方法            | 哺乳动物编辑靶标 |
|----------------|------------------|----------------------------|----------|
| r0 (REPAIR)    | ADAR2 +<br>E488Q | 来自 Kuttan 和 Bass<br>的超活性变体 | N/A      |

|        |              |           |           |   |
|--------|--------------|-----------|-----------|---|
|        | r1           | r0+V351G  | 合理诱变      | Gluc 转录物上的 CCG 位点 (C82R)  |
|        | r2           | r1+S486A  | 合理诱变      | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r3           | r2+T375S  | 合理诱变      | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r4           | r3+S370C  | Y66H EGFP | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r5           | r4+P462A  | P196L HIS | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r6           | r5+N597I  | P196L HIS | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r7           | r6+L332I  | P196L HIS | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r8           | r7+I398V  | P196L HIS | Gluc 转录物上的 UCG 位点 (C82R)  |
| [2274] | r9           | r8+K350I  | P196L HIS | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r10          | r9+M383L  | P196L HIS | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r11          | r10+D619G | S22P HIS  | Gluc 转录物上的 UCG 位点 (C82R)  |
|        | r12          | r11+S582T | S22P HIS  | 内源 $\beta$ -连环蛋白上的 T41I   |
|        | r13          | r12+V440I | S22P HIS  | 内源 $\beta$ -连环蛋白上的 T41I   |
|        | r14          | r13+S495N | P196L HIS | Gluc 转录物上的 CCT 位点 (L77P)  |
|        | r15          | r14+K418E | P196L HIS | Gluc 转录物上的 CCT 位点 (L77P)  |
|        | r16 (RESCUE) | r15+S661T | S22P HIS  | Gluc 转录物上的 UCG、ACG、GCG、CGC 位点 (C82R), Gluc 转录物上的 CCT 和 CCA 位点 (L77P), 以及内源 $\beta$ -连环蛋白上的 T41I |

[2275] 表28疾病相关突变的疾病信息

| 候选   | 基因            | 疾病                           |
|--|---------------|------------------------------|
| NM_000071.2(CBS):c.325T>C (p.Cys109Arg)            | <i>CBS</i>    | 胸主动脉瘤和主动脉夹层                  |
| NM_000141.4(FGFR2):c.799T>C (p.Ser267Pro)          | <i>FGFR2</i>  | Pfeiffer 综合征/Crouzon 综合征/胃肿瘤 |
| NM_000551.3(VHL):c.473T>C (p.Leu158Pro)            | <i>VHL</i>    | Von Hippel-Lindau 综合征        |
| [2276] NM_002474.2(MYH11):c.3791T>C (p.Leu1264Pro) | <i>MYH11</i>  | 主动脉瘤、家族性胸 4/胸主动脉瘤和主动脉夹层      |
| NM_000018.3(ACADVL):c.848T>C (p.Val283Ala)         | <i>ACADVL</i> | 超长链酰基辅酶 A 脱氢酶缺乏症             |
| NM_002397.4(MEF2C):c.2T>C (p.Met1Thr)              | <i>MEF2C</i>  | 智力低下、定型运动、癫痫和/或脑畸形           |
| NM_002834.4(PTPN11):                               | <i>PTPN11</i> | 努南综合征                        |

|        |  |               |  |
|--------|--|---------------|--|
|        | c.853T>C (p.Phe285Leu)                         |               |  |
|        | NM_005609.3(PYGM):c.2392T>C (p.Trp798Arg)      | <i>PYGM</i>   | 糖原贮积病, V 型   |
|        | NM_001256850.1(TTN):c.90211T>C (p.Cys30071Arg) | <i>TTN</i>    | 肢带型肌营养不良, 2J 型/远端肌病<br>Markesbery-Griggs 型/遗传性肌病伴早期呼吸衰竭/早发性肌病, 伴致命性心肌病/家族性肥厚型心肌病 9 |
|        | NM_005633.3(SOS1):c.806T>C (p.Met269Thr)       | <i>SOS1</i>   | 努南综合征 4/努南综合征  |
|        | NM_015559.2(SETBP1):c.2612T>C (p.Ile871Thr)    | <i>SETBP1</i> | Schinzel-Giedion 综合征   |
|        | NM_004572.3(PKP2):c.2386T>C (p.Cys796Arg)      | <i>PKP2</i>   | 心律失常性右室心肌病, 9 型  |
|        | NM_000138.4(FBN1):c.4222T>C (p.Cys1408Arg)     | <i>FBN1</i>   | 马凡氏综合征   |
|        | NM_000375.2(UROS):c.217T>C (p.Cys73Arg)        | <i>UROS</i>   | 先天性红细胞生成性卟啉症   |
| [2277] | NM_014139.2(SCN11A):c.1187T>C (p.Leu396Pro)    | <i>SCN11A</i> | 未提供/神经病变, 遗传性感觉和自主神经, VII 型  |
|        | NM_000152.4(GAA):c.1655T>C (p.Leu552Pro)       | <i>GAA</i>    | 糖原贮积病, II 型  |
|        | NM_020630.4(RET):c.1858T>C (p.Cys620Arg)       | <i>RET</i>    | 多发性内分泌肿瘤 2a 型/多发性内分泌肿瘤 2 型/MEN2A 和 FMTC  |
|        | NM_000016.5(ACADM):c.199T>C (p.Tyr67His)       | <i>ACADM</i>  | 中链酰基辅酶 A 脱氢酶缺乏症  |
|        | NM_014874.3(MFN2):c.227T>C (p.Leu76Pro)        | <i>MFN2</i>   | 腓骨肌萎缩症 2A2A 型  |
|        | NM_000341.3(SLC3A1):c.1400T>C (p.Met467Thr)    | <i>SLC3A1</i> | 胱氨酸尿   |
|        | NM_000431.3(MVK):c.803T>C (p.Ile268Thr)        | <i>MVK</i>    | 甲羟戊酸尿症/伴发高热的高免疫球蛋白 D   |
|        | NM_004004.5(GJB2):c.229T>C (p.Trp77Arg)        | <i>GJB2</i>   | 耳聋, 常染色体隐性遗传 1A /耳聋, 常染色体显性 3a /非综合征性听力损失和耳聋                                       |
|        | NM_000041.4(APOE):c.388T>C (p.Cys130Arg)       | <i>APOE</i>   | 阿尔茨海默氏病 2  |
|        | NM_000041.4(APOE):c.595T>C (p.Cys176Arg)       | <i>APOE</i>   | 阿尔茨海默氏病 2  |

[2278] 表29用于荧光素酶编辑的指导序列

[2279]

| 名称                      | 被靶向的基因 | REPAIR/RESCUE | 基序  | 碱基翻转变位置 | 密码子变化 | 间隔子序列  | 备注     | 编辑百分比 (第一圈) | 第一圈  |
|-------------------------|--------|---------------|-----|---------|-------|--|--------|-------------|------|
| UCG 靶向指导序列              | Gluc   | RESCUE        | UCG | C/30/26 | C82R  | gugcCauugaugugggacaggcagauca (SEQ ID NO: 780)    | 无 5' G | 67.493      | I03B |
| GCG 靶向指导序列              | Gluc   | RESCUE        | GCG | U/30/20 | C82R  | guuggcgugcucuaugaugugggacaggcag (SEQ ID NO: 781) |        | 45.475      | I03B |
| ACG 靶向指导序列              | Gluc   | RESCUE        | ACG | C/30/28 | C82R  | ggccuuugaugugggacaggcagauca (SEQ ID NO: 782)     |        | 64.464      | I03B |
| CCG 靶向指导序列              | Gluc   | RESCUE        | CCG | C/30/26 | C82R  | gugccguugaugugggacaggcagauca (SEQ ID NO: 783)    | 无 5' G | 62.947      | I03B |
| CCU 靶向                  | Gluc   | RESCUE        | CCU | C/30/26 | L77P  | gggaacggcagaucaagaccgccugugca (SEQ ID NO: 784)   |        | 3.800       | I03B |
| CCA 靶向                  | Gluc   | RESCUE        | CCA | C/30/26 | L77P  | gggaugggcagaucaagaccgccugugca (SEQ ID NO: 785)   |        | 4.509       | I03B |
| 基序指导序列 UCU, 翻转 U        | Gluc   | RESCUE        | UCU | U/30/26 | L82F  | gugaUauugaugugggacaggcagauca (SEQ ID NO: 786)    |        | 46.611      | I03C |
| 基序指导序列 UCG, 翻转 U        | Gluc   | RESCUE        | UCG | U/30/26 | C82R  | gugcUauugaugugggacaggcagauca (SEQ ID NO: 787)    |        | 57.945      | I03C |
| 基序指导序列 UCC, 翻转 U        | Gluc   | RESCUE        | UCC | U/30/26 | P82S  | guggUauugaugugggacaggcagauca (SEQ ID NO: 788)    |        | 57.165      | I03C |
| 基序指导序列 UCA, 翻转 U        | Gluc   | RESCUE        | UCA | U/30/26 | H82Y  | guguUauugaugugggacaggcagauca (SEQ ID NO: 789)    |        | 49.256      | I03C |
| 基序指导序列 ACU, 翻转 U        | Gluc   | RESCUE        | ACU | U/30/26 | L82F  | gugaUuuugaugugggacaggcagauca (SEQ ID NO: 790)    |        | 44.241      | I03C |
| 基序指导序列 ACG, 翻转 U        | Gluc   | RESCUE        | ACG | U/30/26 | C82R  | gugcUuuugaugugggacaggcagauca (SEQ ID NO: 791)    |        | 60.722      | I03C |
| Motif guide ACC, flip U | Gluc   | RESCUE        | ACC | U/30/26 | P82S  | guggUuuugaugugggacaggcagauca (SEQ ID NO: 792)    |        | 58.056      | I03C |
| 基序指导序列 ACA, 翻转 U        | Gluc   | RESCUE        | ACA | U/30/26 | H82Y  | gugaUuuugaugugggacaggcagauca (SEQ ID NO: 793)    |        | 40.921      | I03C |
| 基序指导序列 GCU, 翻转 U        | Gluc   | RESCUE        | GCU | U/30/26 | L82F  | gugaUcuugaugugggacaggcagauca (SEQ ID NO: 794)    |        | 4.603       | I03C |
| 基序指导序列 GCG, 翻转 U        | Gluc   | RESCUE        | GCG | U/30/26 | C82R  | gugcUcuugaugugggacaggcagauca (SEQ ID NO: 795)    |        | 43.507      | I03C |
| 基序指导序列 GCC, 翻转 U        | Gluc   | RESCUE        | GCC | U/30/26 | P82S  | guggUcuugaugugggacaggcagauca (SEQ ID NO: 796)    |        | 11.006      | I03C |
| 基序指导序列 GCA, 翻转 U        | Gluc   | RESCUE        | GCA | U/30/26 | H82Y  | guguUcuugaugugggacaggcagauca (SEQ ID NO: 797)    |        | 4.239       | I03C |
| 基序指导序列 CCU, 翻转 U        | Gluc   | RESCUE        | CCU | U/30/26 | L82F  | gugaUguugaugugggacaggcagauca (SEQ ID NO: 798)    |        | 11.808      | I03C |
| 基序指导序列 CCG, 翻转 U        | Gluc   | RESCUE        | CCG | U/30/26 | C82R  | gugcUguugaugugggacaggcagauca (SEQ ID NO: 799)    |        | 51.692      | I03C |
| 基序指导序列 CCC, 翻转 U        | Gluc   | RESCUE        | CCC | U/30/26 | P82S  | guggUguugaugugggacaggcagauca (SEQ ID NO: 800)    |        | 28.402      | I03C |
| 基序指导序列 CCA, 翻转 U        | Gluc   | RESCUE        | CCA | U/30/26 | H82Y  | guguUguugaugugggacaggcagauca (SEQ ID NO: 801)    |        | 7.597       | I03C |
| 基序指导序列 UCU, 翻转 C        | Gluc   | RESCUE        | UCU | C/30/26 | L82F  | gugaCauugaugugggacaggcagauca (SEQ ID NO: 802)    |        | 49.430      | I03C |
| 基序指导序列 UCC, 翻转 C        | Gluc   | RESCUE        | UCC | C/30/26 | P82S  | guggCauugaugugggacaggcagauca (SEQ ID NO: 803)    |        | 59.973      | I03C |
| 基序指导序列 UCA, 翻转 C        | Gluc   | RESCUE        | UCA | C/30/26 | H82Y  | guguCauugaugugggacaggcagauca (SEQ ID NO: 804)    |        | 48.343      | I03C |
| 基序指导序列 ACU, 翻转 C        | Gluc   | RESCUE        | ACU | C/30/26 | L82F  | gugaCuugaugugggacaggcagauca (SEQ ID NO: 805)     |        | 47.840      | I03C |
| 基序指导序列 ACG, 翻转 C        | Gluc   | RESCUE        | ACG | C/30/26 | C82R  | gugcCuugaugugggacaggcagauca (SEQ ID NO: 806)     |        | 70.120      | I03C |
| 基序指导序列 ACC, 翻转 C        | Gluc   | RESCUE        | ACC | C/30/26 | P82S  | guggCuugaugugggacaggcagauca (SEQ ID NO: 807)     |        | 58.779      | I03C |
| 基序指导序列 ACA, 翻转 C        | Gluc   | RESCUE        | ACA | C/30/26 | H82Y  | guguCuugaugugggacaggcagauca (SEQ ID NO: 808)     |        | 45.594      | I03C |
| 基序指导序列 GCU, 翻转 C        | Gluc   | RESCUE        | GCU | C/30/26 | L82F  | gugaCcuugaugugggacaggcagauca (SEQ ID NO: 809)    |        | 3.652       | I03C |
| 基序指导序列 GCG, 翻转 C        | Gluc   | RESCUE        | GCG | C/30/26 | C82R  | gugcCcuugaugugggacaggcagauca (SEQ ID NO: 810)    |        | 37.719      | I03C |
| 基序指导序列 GCC, 翻转 C        | Gluc   | RESCUE        | GCC | C/30/26 | P82S  | guggCcuugaugugggacaggcagauca (SEQ ID NO: 811)    |        | 34.488      | I03C |
| 基序指导序列 GCA, 翻转 C        | Gluc   | RESCUE        | GCA | C/30/26 | H82Y  | guguCcuugaugugggacaggcagauca (SEQ ID NO: 812)    |        | 2.944       | I03C |
| 基序指导序列 CCU, 翻转 C        | Gluc   | RESCUE        | CCU | C/30/26 | L82F  | gugaCguugaugugggacaggcagauca (SEQ ID NO: 813)    |        | 16.647      | I03C |
| 基序指导序列 CCC, 翻转 C        | Gluc   | RESCUE        | CCC | C/30/26 | P82S  | guggCguugaugugggacaggcagauca (SEQ ID NO: 814)    |        | 48.269      | I03C |

[2280]

|                                |      |        |     |         |      |  |             |        |      |
|--------------------------------|------|--------|-----|---------|------|--|-------------|--------|------|
| 基序指导序列<br>CCA, 翻转 C            | Gluc | RESCUE | CCA | C/30/26 | H82Y | guguCguugauguggagcagcagauca<br>(SEQ ID NO: 815)                              |             | 12.670 | 103C |
| 非靶向指导序列                        | N/A  | N/A    | N/A | N/A     | N/A  | guaugccugcguugcagcagcaugucug<br>(SEQ ID NO: 816)                             |             | N/A    | 104C |
| 具有脱靶 A-G 错配 1 的 Gluc 特异性指导序列   | Gluc | RESCUE | UCG | C/30/26 | C82R | ggugcuaGugauguggagcagcagauca<br>(SEQ ID NO: 817)                             | 额外的 G 增加特异性 | 10.836 | 105D |
| 具有脱靶 A-G 错配 2 的 Gluc 特异性指导序列   | Gluc | RESCUE | UCG | C/30/26 | C82R | ggugcuaUGauguggagcagcagauca<br>(SEQ ID NO: 818)                              | 额外的 G 增加特异性 | 16.037 | 105D |
| 具有脱靶 A-G 错配 3 的 Gluc 特异性指导序列   | Gluc | RESCUE | UCG | C/30/26 | C82R | ggugcuaUaugUGggagcagcagauca<br>(SEQ ID NO: 819)                              | 额外的 G 增加特异性 | 29.555 | 105D |
| 具有脱靶 A-G 组合 1+2 的 Gluc 特异性指导序列 | Gluc | RESCUE | UCG | C/30/26 | C82R | ggugcuaGGauguggagcagcagauca<br>(SEQ ID NO: 820)                              | 额外的 G 增加特异性 | 1.533  | 105D |
| 具有脱靶 A-G 所有组合的 Gluc 特异性指导序列    | Gluc | RESCUE | UCG | C/30/26 | C82R | ggugcuaGGaugUGggagcagcagauca<br>(SEQ ID NO: 821)                             | 额外的 G 增加特异性 | 0.412  | 105D |
| A 至 I REPAIR 指导序列              | Cluc | REPAIR | TAG | C/50/34 | *85W | ggccucugcggcagcucucgucucugaugg<br>uggcagcucuccgg (SEQ ID NO: 822)            |             | N/A    | 106A |
| 平铺指导序列 30<br>翻转 30 U           | Gluc | RESCUE | UCG | U/30/30 | C82R | guaugauguggagcagcagcagcagcagc<br>(SEQ ID NO: 823)                            |             | 6.327  | 115  |
| 平铺指导序列 30<br>翻转 28 U           | Gluc | RESCUE | UCG | U/30/28 | C82R | ggcuaugauguggagcagcagcagcagcagc<br>(SEQ ID NO: 824)                          |             | 45.029 | 115  |
| 平铺指导序列 30<br>翻转 26 U           | Gluc | RESCUE | UCG | U/30/26 | C82R | ggugcuaugauguggagcagcagcagcagc<br>(SEQ ID NO: 825)                           |             | 54.433 | 115  |
| 平铺指导序列 30<br>翻转 24 U           | Gluc | RESCUE | UCG | U/30/24 | C82R | ggcugcuaugauguggagcagcagcagcagc<br>(SEQ ID NO: 826)                          |             | 51.454 | 115  |
| 平铺指导序列 30<br>翻转 22 U           | Gluc | RESCUE | UCG | U/30/22 | C82R | ggggcugcuaugauguggagcagcagcagc<br>(SEQ ID NO: 827)                           |             | 28.375 | 115  |
| 平铺指导序列 30<br>翻转 20 U           | Gluc | RESCUE | UCG | U/30/20 | C82R | gumggcugcuaugauguggagcagcagcagc<br>(SEQ ID NO: 828)                          |             | 34.460 | 115  |
| 平铺指导序列 30<br>翻转 18 U           | Gluc | RESCUE | UCG | U/30/18 | C82R | gucumggcugcuaugauguggagcagcagc<br>(SEQ ID NO: 829)                           |             | 24.148 | 115  |
| 平铺指导序列 30<br>翻转 16 U           | Gluc | RESCUE | UCG | U/30/16 | C82R | gcuaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 830)                           |             | 12.372 | 115  |
| 平铺指导序列 30<br>翻转 14 U           | Gluc | RESCUE | UCG | U/30/14 | C82R | gucuaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 831)                          |             | 2.008  | 115  |
| 平铺指导序列 30<br>翻转 12 U           | Gluc | RESCUE | UCG | U/30/12 | C82R | gucuaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 832)                          |             | 4.807  | 115  |
| 平铺指导序列 30<br>翻转 10 U           | Gluc | RESCUE | UCG | U/30/10 | C82R | gcuaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 833)                           |             | 6.679  | 115  |
| 平铺指导序列 30<br>翻转 8U             | Gluc | RESCUE | UCG | U/30/8  | C82R | gacuaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 834)                          |             | 0.930  | 115  |
| 平铺指导序列 30<br>翻转 6 U            | Gluc | RESCUE | UCG | U/30/6  | C82R | guguaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 835)                          |             | 22.763 | 115  |
| 平铺指导序列 30<br>翻转 4 U            | Gluc | RESCUE | UCG | U/30/4  | C82R | gguaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 836)                           |             | 0.633  | 115  |
| 平铺指导序列 30<br>翻转 2U             | Gluc | RESCUE | UCG | U/30/2  | C82R | gggguaucumggcugcuaugauguggagcagc<br>(SEQ ID NO: 837)                         |             | 0.266  | 115  |
| 平铺指导序列 50<br>翻转 50U            | Gluc | RESCUE | UCG | U/50/50 | C82R | guaugauguggagcagcagcagcagcagcagc<br>cugggagcagcagcagc (SEQ ID NO: 838)       |             | 1.263  | 115  |
| 平铺指导序列 50<br>翻转 48U            | Gluc | RESCUE | UCG | U/50/48 | C82R | ggcuaugauguggagcagcagcagcagcagc<br>ugggagcagcagc (SEQ ID NO: 839)            |             | 24.879 | 115  |
| 平铺指导序列 50<br>翻转 46U            | Gluc | RESCUE | UCG | U/50/46 | C82R | ggugcuaugauguggagcagcagcagcagcagc<br>ccugggagcagc (SEQ ID NO: 840)           |             | 21.993 | 115  |
| 平铺指导序列 50<br>翻转 44U            | Gluc | RESCUE | UCG | U/50/44 | C82R | ggcugcuaugauguggagcagcagcagcagcagc<br>cccgaggagcagc (SEQ ID NO: 841)         |             | 25.736 | 115  |
| 平铺指导序列 50<br>翻转 42U            | Gluc | RESCUE | UCG | U/50/42 | C82R | ggggcugcuaugauguggagcagcagcagcagc<br>agcccgaggagc (SEQ ID NO: 842)           |             | 27.579 | 115  |
| 平铺指导序列 50<br>翻转 40U            | Gluc | RESCUE | UCG | U/50/40 | C82R | gumggcugcuaugauguggagcagcagcagc<br>gagcccgaggagc (SEQ ID NO: 843)            |             | 27.775 | 115  |
| 平铺指导序列 50<br>翻转 38U            | Gluc | RESCUE | UCG | U/50/38 | C82R | gucumggcugcuaugauguggagcagcagcagc<br>cagagcccgaggagc (SEQ ID NO: 844)        |             | 13.260 | 115  |
| 平铺指导序列 50<br>翻转 36U            | Gluc | RESCUE | UCG | U/50/36 | C82R | gcuaucumggcugcuaugauguggagcagcagc<br>aucagagcccgaggagc (SEQ ID NO: 845)      |             | 9.892  | 115  |
| 平铺指导序列 50<br>翻转 34U            | Gluc | RESCUE | UCG | U/50/34 | C82R | gucuaucumggcugcuaugauguggagcagcagc<br>agaucagagcccgaggagc (SEQ ID NO: 846)   |             | 19.186 | 115  |
| 平铺指导序列 50<br>翻转 32U            | Gluc | RESCUE | UCG | U/50/32 | C82R | gucuaucumggcugcuaugauguggagcagcagc<br>gcagaucagagcccgaggagc (SEQ ID NO: 847) |             | 22.932 | 115  |
| 平铺指导序列 50<br>翻转 30U            | Gluc | RESCUE | UCG | U/50/30 | C82R | gcuaucumggcugcuaugauguggagcagcagc<br>agcagaucagagcccgaggagc (SEQ ID NO: 848) |             | 12.794 | 115  |
| 平铺指导序列 50<br>翻转 28U            | Gluc | RESCUE | UCG | U/50/28 | C82R | gacuaucumggcugcuaugauguggagcagcagc<br>acagcagcagcagc (SEQ ID NO: 849)        |             | 33.367 | 115  |

[2281]

|                     |      |        |     |         |      |   |        |           |
|---------------------|------|--------|-----|---------|------|---|--------|-----------|
| 平铺指导序列 50<br>翻转 26U | Gluc | RESCUE | UCG | U/50/26 | C82R | gugaacuucaucacuuugggcugucuaauugaug<br>ggacagcagauca (SEQ ID NO: 850)    | 32.651 | 115       |
| 平铺指导序列 50<br>翻转 24U | Gluc | RESCUE | UCG | U/50/24 | C82R | ggaugaacuucaucacuuugggcugucuaauugaug<br>uggacagcagauca (SEQ ID NO: 851) | 22.201 | 115       |
| 平铺指导序列 50<br>翻转 22U | Gluc | RESCUE | UCG | U/50/22 | C82R | gggaugaacuucaucacuuugggcugucuaauuga<br>tugggacagcagau (SEQ ID NO: 852)  | 12.607 | 115       |
| 平铺指导序列 50<br>翻转 20U | Gluc | RESCUE | UCG | U/50/20 | C82R | gcuuggaagaacuucaucacuuugggcugucuaau<br>gauguggacagcag (SEQ ID NO: 853)  | 17.727 | 115       |
| 平铺指导序列 50<br>翻转 18U | Gluc | RESCUE | UCG | U/50/18 | C82R | gucuggaugaacuucaucacuuugggcugucua<br>uugauguggacagc (SEQ ID NO: 854)    | 11.842 | 115       |
| 平铺指导序列 50<br>翻转 16U | Gluc | RESCUE | UCG | U/50/16 | C82R | gcuucuggaagaacuucaucacuuugggcuguc<br>uaugauguggacag (SEQ ID NO: 855)    | 9.368  | 115       |
| 平铺指导序列 50<br>翻转 14U | Gluc | RESCUE | UCG | U/50/14 | C82R | gagcugcuggaagaacuucaucacuuugggcug<br>gcauugauguggac (SEQ ID NO: 856)    | 2.637  | 115       |
| 平铺指导序列 50<br>翻转 12U | Gluc | RESCUE | UCG | U/50/12 | C82R | ggcagcugcuggaagaacuucaucacuuugggcg<br>ugcuaauugaug (SEQ ID NO: 857)     | 35.033 | 115       |
| 平铺指导序列 50<br>翻转 10U | Gluc | RESCUE | UCG | U/50/10 | C82R | gugcagcugcuggaagaacuucaucacuuuggg<br>gcgcauugaug (SEQ ID NO: 858)       | 10.675 | 115       |
| 平铺指导序列 50<br>翻转 8U  | Gluc | RESCUE | UCG | U/50/8  | C82R | gugugcagcugcuggaagaacuucaucacuu<br>ggcgucuaugaug (SEQ ID NO: 859)       | 1.730  | 115       |
| 平铺指导序列 50<br>翻转 6U  | Gluc | RESCUE | UCG | U/50/6  | C82R | ggugugcagcugcuggaagaacuucaucacuu<br>uggcgugcuauga (SEQ ID NO: 860)      | 2.249  | 115       |
| 平铺指导序列 50<br>翻转 4U  | Gluc | RESCUE | UCG | U/50/4  | C82R | gugcugugcagcugcuggaagaacuucaucacuu<br>ucuuuggcgucua (SEQ ID NO: 861)    | 0.438  | 115       |
| 平铺指导序列 50<br>翻转 2U  | Gluc | RESCUE | UCG | U/50/2  | C82R | gcugaugugcagcugcuggaagaacuucaucuu<br>caucuuuggcgucua (SEQ ID NO: 862)   | 0.293  | 115       |
| 基序指导序列<br>UCU, 翻转 A | Gluc | RESCUE | UCU | A/30/26 | L82F | gugaAauugauguggacagcagauca (SEQ ID NO: 863)                             | 0.084  | 114A-114C |
| 基序指导序列<br>UCG, 翻转 A | Gluc | RESCUE | UCG | A/30/26 | C82R | gugcAauugauguggacagcagauca (SEQ ID NO: 864)                             | 0.399  | 114A-114C |
| 基序指导序列<br>UCC, 翻转 A | Gluc | RESCUE | UCC | A/30/26 | P82S | guggAauugauguggacagcagauca (SEQ ID NO: 865)                             | 0.210  | 114A-114C |
| 基序指导序列<br>UCA, 翻转 A | Gluc | RESCUE | UCA | A/30/26 | H82Y | guguAauugauguggacagcagauca (SEQ ID NO: 866)                             | 0.077  | 114A-114C |
| 基序指导序列<br>ACU, 翻转 A | Gluc | RESCUE | ACU | A/30/26 | L82F | gugaAuuugauguggacagcagauca (SEQ ID NO: 867)                             | 0.097  | 114A-114C |
| 基序指导序列<br>ACG, 翻转 A | Gluc | RESCUE | ACG | A/30/26 | C82R | gugcAuuugauguggacagcagauca (SEQ ID NO: 868)                             | 0.399  | 114A-114C |
| 基序指导序列<br>ACC, 翻转 A | Gluc | RESCUE | ACC | A/30/26 | P82S | guggAuuugauguggacagcagauca (SEQ ID NO: 869)                             | 0.138  | 114A-114C |
| 基序指导序列<br>ACA, 翻转 A | Gluc | RESCUE | ACA | A/30/26 | H82Y | guguAuuugauguggacagcagauca (SEQ ID NO: 870)                             | 0.036  | 114A-114C |
| 基序指导序列<br>GCU, 翻转 A | Gluc | RESCUE | GCU | A/30/26 | L82F | gugaAcuugauguggacagcagauca (SEQ ID NO: 871)                             | 0.057  | 114A-114C |
| 基序指导序列<br>GCG, 翻转 A | Gluc | RESCUE | GCG | A/30/26 | C82R | gugcAcuugauguggacagcagauca (SEQ ID NO: 872)                             | 0.029  | 114A-114C |
| 基序指导序列<br>GCC, 翻转 A | Gluc | RESCUE | GCC | A/30/26 | P82S | guggAcuugauguggacagcagauca (SEQ ID NO: 873)                             | 0.023  | 114A-114C |
| 基序指导序列<br>GCA, 翻转 A | Gluc | RESCUE | GCA | A/30/26 | H82Y | guguAcuugauguggacagcagauca (SEQ ID NO: 874)                             | 0.022  | 114A-114C |
| 基序指导序列<br>CCU, 翻转 A | Gluc | RESCUE | CCU | A/30/26 | L82F | gugaAguugauguggacagcagauca (SEQ ID NO: 875)                             | 0.055  | 114A-114C |
| 基序指导序列<br>CCG, 翻转 A | Gluc | RESCUE | CCG | A/30/26 | C82R | gugcAguugauguggacagcagauca (SEQ ID NO: 876)                             | 0.066  | 114A-114C |
| 基序指导序列<br>CCC, 翻转 A | Gluc | RESCUE | CCC | A/30/26 | P82S | guggAguugauguggacagcagauca (SEQ ID NO: 877)                             | 0.016  | 114A-114C |
| 基序指导序列<br>CCA, 翻转 A | Gluc | RESCUE | CCA | A/30/26 | H82Y | guguAguugauguggacagcagauca (SEQ ID NO: 878)                             | 0.022  | 114A-114C |
| 基序指导序列<br>UCU, 翻转 G | Gluc | RESCUE | UCU | G/30/26 | L82F | gugaGauugauguggacagcagauca (SEQ ID NO: 879)                             | 0.058  | 114A-114C |
| 基序指导序列<br>UCG, 翻转 G | Gluc | RESCUE | UCG | G/30/26 | C82R | gugcGauugauguggacagcagauca (SEQ ID NO: 880)                             | 0.094  | 114A-114C |
| 基序指导序列<br>UCC, 翻转 G | Gluc | RESCUE | UCC | G/30/26 | P82S | guggGauugauguggacagcagauca (SEQ ID NO: 881)                             | 0.022  | 114A-114C |
| 基序指导序列<br>UCA, 翻转 G | Gluc | RESCUE | UCA | G/30/26 | H82Y | guguGauugauguggacagcagauca (SEQ ID NO: 882)                             | 0.026  | 114A-114C |
| 基序指导序列<br>ACU, 翻转 G | Gluc | RESCUE | ACU | G/30/26 | L82F | gugaGuuugauguggacagcagauca (SEQ ID NO: 883)                             | 0.053  | 114A-114C |
| 基序指导序列<br>ACG, 翻转 G | Gluc | RESCUE | ACG | G/30/26 | C82R | gugcGuuugauguggacagcagauca (SEQ ID NO: 884)                             | 0.035  | 114A-114C |
| 基序指导序列<br>ACC, 翻转 G | Gluc | RESCUE | ACC | G/30/26 | P82S | guggGuuugauguggacagcagauca (SEQ ID NO: 885)                             | 0.017  | 114A-114C |
| 基序指导序列<br>ACA, 翻转 G | Gluc | RESCUE | ACA | G/30/26 | H82Y | guguGuuugauguggacagcagauca (SEQ ID NO: 886)                             | 0.030  | 114A-114C |
| 基序指导序列              | Gluc | RESCUE | GCU | G/30/26 | L82F | gugaGcuugauguggacagcagauca (SEQ ID NO: 887)                             | 0.053  | 114A-114C |

[2282]

|                  |      |        |     |         |      |                               |                  |  |        |            |
|------------------|------|--------|-----|---------|------|-------------------------------|------------------|--|--------|------------|
| GCU, 翻转 G        |      |        |     |         |      |                               |                  |  |        |            |
| 基序指导序列 GCG, 翻转 G | Gluc | RESCUE | GCG | G/30/26 | C82R | gugcGcuugaugugggacagcagauca   | (SEQ ID NO: 888) |  | 0.026  | 114A-114C  |
| 基序指导序列 GCC, 翻转 G | Gluc | RESCUE | GCC | G/30/26 | P82S | guggGcuugaugugggacagcagauca   | (SEQ ID NO: 889) |  | 0.018  | 114A-114C  |
| 基序指导序列 GCA, 翻转 G | Gluc | RESCUE | GCA | G/30/26 | H82Y | guguGcuugaugugggacagcagauca   | (SEQ ID NO: 890) |  | 0.018  | 114A-114C  |
| 基序指导序列 CCU, 翻转 G | Gluc | RESCUE | CCU | G/30/26 | L82F | gugaGcuugaugugggacagcagauca   | (SEQ ID NO: 891) |  | 0.049  | 114A-114C  |
| 基序指导序列 CCG, 翻转 G | Gluc | RESCUE | CCG | G/30/26 | C82R | gugcGcuugaugugggacagcagauca   | (SEQ ID NO: 892) |  | 0.064  | 114A-114C  |
| 基序指导序列 CCC, 翻转 G | Gluc | RESCUE | CCC | G/30/26 | P82S | guggGcuugaugugggacagcagauca   | (SEQ ID NO: 893) |  | 0.011  | 114A-114C  |
| 基序指导序列 CCA, 翻转 G | Gluc | RESCUE | CCA | G/30/26 | H82Y | guguGcuugaugugggacagcagauca   | (SEQ ID NO: 894) |  | 0.017  | 114A-114C  |
| 平铺指导序列 30 翻转 30C | Gluc | RESCUE | UCG | C/30/30 | C82R | gugugaugugggacagcagauca       | (SEQ ID NO: 895) |  | 0.164  | 122A-1224C |
| 平铺指导序列 30 翻转 28C | Gluc | RESCUE | UCG | C/30/28 | C82R | ggcugaugugggacagcagauca       | (SEQ ID NO: 896) |  | 37.368 | 122A-122C  |
| 平铺指导序列 30 翻转 26C | Gluc | RESCUE | UCG | C/30/26 | C82R | ggugcugaugugggacagcagauca     | (SEQ ID NO: 897) |  | 44.775 | 122A-122C  |
| 平铺指导序列 30 翻转 24C | Gluc | RESCUE | UCG | C/30/24 | C82R | ggcgugcugaugugggacagcagauca   | (SEQ ID NO: 898) |  | 26.988 | 122A-122C  |
| 平铺指导序列 30 翻转 22C | Gluc | RESCUE | UCG | C/30/22 | C82R | ggggcgugcugaugugggacagcagauca | (SEQ ID NO: 899) |  | 16.710 | 122A-122C  |
| 平铺指导序列 30 翻转 20C | Gluc | RESCUE | UCG | C/30/20 | C82R | gungggcgugcugaugugggacagcag   | (SEQ ID NO: 900) |  | 29.288 | 122A-122C  |
| 平铺指导序列 30 翻转 18C | Gluc | RESCUE | UCG | C/30/18 | C82R | gucugggcgugcugaugugggacagc    | (SEQ ID NO: 901) |  | 18.125 | 122A-122C  |
| 平铺指导序列 30 翻转 16C | Gluc | RESCUE | UCG | C/30/16 | C82R | gcucugggcgugcugaugugggacag    | (SEQ ID NO: 902) |  | 1.532  | 122A-122C  |
| 平铺指导序列 30 翻转 14C | Gluc | RESCUE | UCG | C/30/14 | C82R | gucaucugggcgugcugaugugggac    | (SEQ ID NO: 903) |  | 0.184  | 122A-122C  |
| 平铺指导序列 30 翻转 12C | Gluc | RESCUE | UCG | C/30/12 | C82R | gucucucugggcgugcugauguggg     | (SEQ ID NO: 904) |  | 0.341  | 122A-122C  |
| 平铺指导序列 30 翻转 10C | Gluc | RESCUE | UCG | C/30/10 | C82R | gcuucucucugggcgugcugaugug     | (SEQ ID NO: 905) |  | 0.275  | 122A-122C  |
| 平铺指导序列 30 翻转 8C  | Gluc | RESCUE | UCG | C/30/8  | C82R | gaucucucucugggcgugcugaug      | (SEQ ID NO: 906) |  | 0.054  | 122A-122C  |
| 平铺指导序列 30 翻转 6C  | Gluc | RESCUE | UCG | C/30/6  | C82R | gugaucucucucugggcgugcugaug    | (SEQ ID NO: 907) |  | 0.437  | 122A-122C  |
| 平铺指导序列 30 翻转 4C  | Gluc | RESCUE | UCG | C/30/4  | C82R | ggaucucucucugggcgugcuguu      | (SEQ ID NO: 908) |  | 0.226  | 122A-122C  |
| 平铺指导序列 30 翻转 2C  | Gluc | RESCUE | UCG | C/30/2  | C82R | ggggaucucucucugggcgugcug      | (SEQ ID NO: 909) |  | 0.040  | 122A-122C  |

[2283] 表30用于内源基因编辑的指导序列

[2284]

| 名称                                  | 靶向的基因  | REPAIR/RESCUE | 基序  | 碱基翻转变化 | 密码子变化 | 间隔子序列   | 编辑百分比 (第一图) | 第一图  |
|-------------------------------------|--------|---------------|-----|--------|-------|---|-------------|------|
| S33F_CTNNB1_30bp_guide_30_9 C flip  | CTNNB1 | RESCUE        | UCU | C/22   | S33F  | GCGAUUCCACAGUCCAGGU<br>AAGACUGUUGCU (SEQ ID NO:910) | 11.245      | 103E |
| H36Y_CTNNB1_30bp_guide_30_9 U flip  | CTNNB1 | RESCUE        | CCA | U/22   | H36Y  | GACCAGAAUUGAUUCCAGA<br>GUCCAGGUAAGA (SEQ ID NO:911) | 2.995       | 103E |
| S37F_CTNNB1_30bp_guide_30_9 U flip  | CTNNB1 | RESCUE        | UCU | U/22   | S37F  | GUGGCACCAUAAUGGAUUC<br>CAGAGUCCAGGU (SEQ ID NO:912) | 17.616      | 103E |
| T41I_CTNNB1_30bp_guide_30_11 U flip | CTNNB1 | RESCUE        | ACC | U/20   | T41I  | GAGGAGCUGUGUAGUGG<br>CACCAGAAUGGAU (SEQ ID NO:913)  | 15.711      | 103E |
| P44L_CTNNB1_30bp_guide_30_9 C flip  | CTNNB1 | RESCUE        | CCU | C/22   | P44L  | GUCAGAGAACGAGCUGUG<br>GUAGUGGCACCAG (SEQ ID NO:914) | 8.599       | 103E |
| P44S_CTNNB1_30bp_guide_30_11 U flip | CTNNB1 | RESCUE        | UCU | U/20   | P44S  | GCUCAGAGAAGUAGCUGU<br>GGUAGUGGCACCA (SEQ ID NO:915) | 22.839      | 103E |
| S45F_CTNNB1_30bp_guide_30_11 C flip | CTNNB1 | RESCUE        | UCU | C/20   | S45F  | GACCACUCAGACAAGGAGC<br>UGUGGUAGUGGC (SEQ ID NO:916) | 12.412      | 103E |
| TCG_KRAS_30bp_guide_30_7 U flip     | KRAS   | RESCUE        | UCG | U/24   | L56L  | GUGUGUCUAGAAUCCA<br>AGAGACAGGUUC (SEQ ID NO:917)    | 18.405      | 103E |
| ACG_KRAS_30bp_guide_30_11 C flip    | KRAS   | RESCUE        | ACG | C/20   | D30D  | GGAUCAUUAUCCUCCACAA<br>AAUGAUUCUGAA (SEQ ID NO:918) | 34.013      | 103E |

[2285]

|                                     |                |        |     |      |      |   |        |      |
|-------------------------------------|----------------|--------|-----|------|------|---|--------|------|
| GCG_KRAS_30bp_guide_30_11 U flip    | <i>KRAS</i>    | RESCUE | GCG | U/20 | G13G | GUCUUGCCUACUCCACCAG<br>CUCCAACUACCA (SEQ ID<br>NO:919)  | 2.180  | 103E |
| CCT_KRAS_30bp_guide_30_11 C flip    | <i>KRAS</i>    | RESCUE | CCU | C/20 | A18A | GGUAUCGUACAACGCACUCU<br>UGCCUACGCCAC (SEQ ID<br>NO:920) | 9.465  | 103E |
| TCG_PPIB_30bp_guide_30_11 U flip    | <i>PPIB</i>    | RESCUE | UCG | U/20 | I18I | GCGGACCCCGCUAUGAGGG<br>CGGCGGCAAGGA (SEQ ID<br>NO:921)  | 10.340 | 103E |
| ACG_PPIB_30bp_guide_30_7 C flip     | <i>PPIB</i>    | RESCUE | ACG | C/24 | R7C  | GAUAUCCUCCACAAAAUG<br>AUUCUGAAUAG (SEQ ID<br>NO:922)    | 34.213 | 103E |
| GCG_PPIB_30bp_guide_30_11 U flip    | <i>PPIB</i>    | RESCUE | GCG | U/20 | A19V | GGACGGACCCUCGAUGAG<br>GGCGGCGCAAG (SEQ ID<br>NO:923)    | 4.778  | 103E |
| CCG_PPIB_30bp_guide_30_11 C flip    | <i>PPIB</i>    | RESCUE | CCG | C/20 | S21S | GGGAAGAAGACCGACCCCG<br>CGAUGAGGGCGG (SEQ ID<br>NO:924)  | 6.101  | 103E |
| TCG_SMARCA4_30bp_guide_30_9 U flip  | <i>SMARCA4</i> | RESCUE | UCG | U/22 | S85L | GGGUCGUCUACAUGCCCU<br>UCUCAUGCAUGG (SEQ ID<br>NO:925)   | 6.807  | 103E |
| ACG_SMARCA4_30bp_guide_30_11 U flip | <i>SMARCA4</i> | RESCUE | ACG | U/20 | D86D | GAGCGCGGUCUCCGACA<br>UGCCUUCUCAU (SEQ ID<br>NO:926)     | 6.943  | 103E |
| GCG_SMARCA4_30bp_guide_30_11 C flip | <i>SMARCA4</i> | RESCUE | GCG | C/20 | R89C | GUGGUUGAGCCCGGGUC<br>GUCCGACAUGCC (SEQ ID<br>NO:927)    | 2.277  | 103E |
| CCG_SMARCA4_30bp_guide_30_11 U flip | <i>SMARCA4</i> | RESCUE | CCG | U/20 | P88L | GGUUGUAGCGUGGUCGU<br>CCGACAUGCCUU (SEQ ID<br>NO:928)    | 4.819  | 103E |
| NRAS_C-flip_guide_30_11             | <i>NRAS</i>    | RESCUE | UCC | C/20 | I21I | GGGAUAGCUGCAUUGC<br>AGUGCGUUUCC (SEQ ID<br>NO:929)      | 21.529 | 103E |
| NKFB1_U-flip_guide_30_11            | <i>NFKB</i>    | RESCUE | ACC | U/20 | P33S | GGCAUCUGUGUUUGAAA<br>UACUUCUGGAUUA (SEQ ID<br>NO:930)   | 24.376 | 103E |
| EZH2_U-flip_guide_30_11             | <i>EZH2</i>    | RESCUE | UCA | U/20 | F32F | GCAGCUCGUCUUAACCUCU<br>UGAGCUGUCUCA (SEQ ID<br>NO:931)  | 15.855 | 103E |
| NF2_U-flip_guide_30_7               | <i>NF2</i>     | RESCUE | ACG | U/24 | T21M | GGUGAACUUCUUGGGUUG<br>CUUCCUUGAGA (SEQ ID<br>NO:932)    | 24.904 | 103E |
| RAF1_U-flip_guide_30_7              | <i>RAF1</i>    | RESCUE | UCC | U/24 | P30S | GUUGUAGUAGAGAUGCAG<br>CUGGAGCCAUCAA (SEQ ID<br>NO:933)  | 20.867 | 103E |
| S33F_CTNNB1_30bp_guide_U-flip_30_7  | <i>CTNNB1</i>  | RESCUE | UCU | U/24 | S33F | GAUCCAUAGUCCAGGUA<br>GACUGUUGCUCG (SEQ ID<br>NO:934)    | 9.227  | 125A |
| S33F_CTNNB1_30bp_guide_U-flip_30_9  | <i>CTNNB1</i>  | RESCUE | UCU | U/22 | S33F | GGGAUCCAUAGUCCAGGU<br>AAGACUGUUGU (SEQ ID<br>NO:935)    | 11.245 | 125A |
| S33F_CTNNB1_30bp_guide_U-flip_30_11 | <i>CTNNB1</i>  | RESCUE | UCU | U/20 | S33F | GAUGGAUCCAUAGUCCAG<br>GUAAGACUGUUG (SEQ ID<br>NO:936)   | 7.081  | 125A |
| S33F_CTNNB1_30bp_guide_U-flip_30_13 | <i>CTNNB1</i>  | RESCUE | UCU | U/18 | S33F | GGAAUGGAUCCAUAGUCC<br>AGGUAAGACUGU (SEQ ID<br>NO:937)   | 9.782  | 125A |
| H36Y_CTNNB1_30bp_guide_U-flip_30_7  | <i>CTNNB1</i>  | RESCUE | CCA | U/24 | H36Y | GCAGAAUUGAUCCAGAG<br>UCCAGGUAAGACU (SEQ ID<br>NO:938)   | 1.310  | 125A |
| H36Y_CTNNB1_30bp_guide_U-flip_30_9  | <i>CTNNB1</i>  | RESCUE | CCA | U/22 | H36Y | GACCAGAAUUGAUCCAGA<br>GUCCAGGUAAGA (SEQ ID<br>NO:939)   | 2.995  | 125A |
| H36Y_CTNNB1_30bp_guide_U-flip_30_11 | <i>CTNNB1</i>  | RESCUE | CCA | U/20 | H36Y | GGCACCAGAAUUGAUCCA<br>GAGUCCAGGUAA (SEQ ID<br>NO:940)   | 0.918  | 125A |
| H36Y_CTNNB1_30bp_guide_U-flip_30_13 | <i>CTNNB1</i>  | RESCUE | CCA | U/18 | H36Y | GUGGCACCAGAAUUGAUCC<br>CAGAGUCCAGGU (SEQ ID<br>NO:941)  | 1.061  | 125A |
| S37F_CTNNB1_30bp_guide_U-flip_30_7  | <i>CTNNB1</i>  | RESCUE | UCU | U/24 | S37F | GGCACCAGAAUUGAUCCA<br>GAGUCCAGGUAA (SEQ ID<br>NO:942)   | 17.616 | 125A |
| S37F_CTNNB1_30bp_guide_U-flip_30_9  | <i>CTNNB1</i>  | RESCUE | UCU | U/22 | S37F | GUGGCACCAGAAUUGAUCC<br>CAGAGUCCAGGU (SEQ ID<br>NO:943)  | 10.901 | 125A |
| S37F_CTNNB1_30bp_guide_U-flip_30_11 | <i>CTNNB1</i>  | RESCUE | UCU | U/20 | S37F | GAGUGGCACCAGAAUUGGA<br>UCCAGAGUCCAG (SEQ ID<br>NO:944)  | 8.898  | 125A |
| S37F_CTNNB1_30bp_guide_U-flip_30_13 | <i>CTNNB1</i>  | RESCUE | UCU | U/18 | S37F | GGUAGUGGCACCAGAAUUG<br>GAUCCAGAGUCC (SEQ ID<br>NO:945)  | 11.718 | 125A |
| T41I_CTNNB1_30bp_guide_U-flip_30_7  | <i>CTNNB1</i>  | RESCUE | ACC | U/24 | T41I | GGCUGUGUAGUCCACC<br>AGAAUGGAUCCA (SEQ ID<br>NO:946)     | 4.936  | 125A |

[2286]

|                                     |        |        |     |      |      |  |        |      |
|-------------------------------------|--------|--------|-----|------|------|--|--------|------|
| T41I_CTNNB1_30bp_guide_U-flip_30_9  | CTNNB1 | RESCUE | ACC | U/22 | T41I | GGAGCUGUGUUAGUGGCA<br>CCAGAAUGGAUUC (SEQ ID<br>NO:947) | 14.554 | 125A |
| T41I_CTNNB1_30bp_guide_U-flip_30_11 | CTNNB1 | RESCUE | ACC | U/20 | T41I | GAGGAGCUGUGUUAGUGG<br>CACCAGAAUGGAU (SEQ ID<br>NO:948) | 14.890 | 125A |
| T41I_CTNNB1_30bp_guide_U-flip_30_13 | CTNNB1 | RESCUE | ACC | U/18 | T41I | GGAAGGAGCUGUGUUAGU<br>GGCACCAGAAUUG (SEQ ID<br>NO:949) | 15.711 | 125A |
| P44L_CTNNB1_30bp_guide_U-flip_30_7  | CTNNB1 | RESCUE | CCU | U/24 | P44L | GAGAGAAUGAGCUGUGGU<br>AGUGGCACCAGAA (SEQ ID<br>NO:950) | 3.767  | 125A |
| P44L_CTNNB1_30bp_guide_U-flip_30_9  | CTNNB1 | RESCUE | CCU | U/22 | P44L | GUCAGAGAAUGAGCUGUG<br>GUAGUGGCACCAG (SEQ ID<br>NO:951) | 6.569  | 125A |
| P44L_CTNNB1_30bp_guide_U-flip_30_11 | CTNNB1 | RESCUE | CCU | U/20 | P44L | GACUCAGAGAAUGAGCUG<br>UGGUAGUGGCACC (SEQ ID<br>NO:952) | 8.599  | 125A |
| P44L_CTNNB1_30bp_guide_U-flip_30_13 | CTNNB1 | RESCUE | CCU | U/18 | P44L | GCCACUCAGAGAAUGAGCU<br>GUGGUAGUGGCA (SEQ ID<br>NO:953) | 2.435  | 125A |
| P44S_CTNNB1_30bp_guide_U-flip_30_7  | CTNNB1 | RESCUE | UCU | U/24 | P44S | GGAGAAGUAGCUGUGGUA<br>GUGGCACCAGAAU (SEQ ID<br>NO:954) | 16.371 | 125A |
| P44S_CTNNB1_30bp_guide_U-flip_30_9  | CTNNB1 | RESCUE | UCU | U/22 | P44S | GCAGAGAAGUAGCUGUGG<br>UAGUGGCACCAGA (SEQ ID<br>NO:955) | 22.090 | 125A |
| P44S_CTNNB1_30bp_guide_U-flip_30_11 | CTNNB1 | RESCUE | UCU | U/20 | P44S | GCUCAGAGAAGUAGCUGU<br>GGUAGUGGCACCA (SEQ ID<br>NO:956) | 22.839 | 125A |
| P44S_CTNNB1_30bp_guide_U-flip_30_13 | CTNNB1 | RESCUE | UCU | U/18 | P44S | GCACUCAGAGAAGUAGCUG<br>UGGUAGUGGCAC (SEQ ID<br>NO:957) | 15.900 | 125A |
| S45F_CTNNB1_30bp_guide_U-flip_30_7  | CTNNB1 | RESCUE | UCU | U/24 | S45F | GCUCAGAAUAGGAGCUGU<br>GGUAGUGGCACCA (SEQ ID<br>NO:958) | 7.049  | 125A |
| S45F_CTNNB1_30bp_guide_U-flip_30_9  | CTNNB1 | RESCUE | UCU | U/22 | S45F | GCACUCAGAAUAGGAGCUG<br>UGGUAGUGGCAC (SEQ ID<br>NO:959) | 9.828  | 125A |
| S45F_CTNNB1_30bp_guide_U-flip_30_11 | CTNNB1 | RESCUE | UCU | U/20 | S45F | GACCACUCAGAAUAGGAGC<br>UGUGGUAGUGGC (SEQ ID<br>NO:960) | 12.412 | 125A |
| S45F_CTNNB1_30bp_guide_U-flip_30_13 | CTNNB1 | RESCUE | UCU | U/18 | S45F | GUUACCACUCAGAAUAGGA<br>GCUGUGGUAGUG (SEQ ID<br>NO:961) | 9.093  | 125A |
| TCG_KRAS_30bp_guide_U-flip_30_7     | KRAS   | RESCUE | UCG | U/24 | L56L | GUGUGUCUAGAAUAUCCA<br>AGAGACAGGUUC (SEQ ID<br>NO:962)  | 18.707 | 125A |
| TCG_KRAS_30bp_guide_U-flip_30_9     | KRAS   | RESCUE | UCG | U/22 | L56L | GGCUGUGUCUAGAAUAUCC<br>AAGAGACAGGUU (SEQ ID<br>NO:963) | 18.405 | 125A |
| TCG_KRAS_30bp_guide_U-flip_30_11    | KRAS   | RESCUE | UCG | U/20 | L56L | GCUCUGUGUCUAGAAUA<br>UCCAAGAGACAGG (SEQ ID<br>NO:964)  | 15.533 | 125A |
| TCG_KRAS_30bp_guide_U-flip_30_13    | KRAS   | RESCUE | UCG | U/18 | L56L | GACCUGCUGUGUCUAGAAU<br>AUCCAAGAGACA (SEQ ID<br>NO:965) | 15.119 | 125A |
| ACG_KRAS_30bp_guide_U-flip_30_7     | KRAS   | RESCUE | ACG | U/24 | D30D | GAUUAUUCUCCACAAAUG<br>AUUCUGAAUAG (SEQ ID<br>NO:966)   | 21.288 | 125A |
| ACG_KRAS_30bp_guide_U-flip_30_9     | KRAS   | RESCUE | ACG | U/22 | D30D | GUCAUAUUCUCCACAAAA<br>UGAUUCUGAAU (SEQ ID<br>NO:967)   | 24.011 | 125A |
| ACG_KRAS_30bp_guide_U-flip_30_11    | KRAS   | RESCUE | ACG | U/20 | D30D | GGAUCAUAUUCUCCACAA<br>AAUGAUUCUGAA (SEQ ID<br>NO:968)  | 34.013 | 125A |
| ACG_KRAS_30bp_guide_U-flip_30_13    | KRAS   | RESCUE | ACG | U/18 | D30D | GUGGAUCAUAUUCUCCAC<br>AAAUGAUUCUG (SEQ ID<br>NO:969)   | 22.047 | 125A |
| GCG_KRAS_30bp_guide_U-flip_30_7     | KRAS   | RESCUE | GCG | U/24 | G13G | GGCCUACUCCACCAGCUC<br>AACUACCACAAG (SEQ ID<br>NO:970)  | 0.476  | 125A |
| GCG_KRAS_30bp_guide_U-flip_30_9     | KRAS   | RESCUE | GCG | U/22 | G13G | GUUGCCUACUCCACCAGCU<br>CCAACUACCACA (SEQ ID<br>NO:971) | 1.735  | 125A |
| GCG_KRAS_30bp_guide_U-flip_30_11    | KRAS   | RESCUE | GCG | U/20 | G13G | GUCUUGCCUACUCCACCAG<br>CUCCAACUACCA (SEQ ID<br>NO:972) | 2.180  | 125A |
| GCG_KRAS_30bp_guide_U-flip_30_13    | KRAS   | RESCUE | GCG | U/18 | G13G | GACUCUUGCCUACUCCACC<br>AGCUCCAACUAC (SEQ ID<br>NO:973) | 0.559  | 125A |
| CCT_KRAS_30bp_guide_U-flip_30_7     | KRAS   | RESCUE | CCU | U/24 | A18A | GCGUCAUAGCACUCUUGCC<br>UACGCCACCAGC (SEQ ID<br>NO:974) | 1.694  | 125A |

[2287]

|                                     |                |        |     |      |      |   |        |      |
|-------------------------------------|----------------|--------|-----|------|------|---|--------|------|
| CCT_KRAS_30bp_guide_U-flip_30_9     | <i>KRAS</i>    | RESCUE | CCU | U/22 | A18A | GAUCGUCAAUGCACUCUUG<br>CCUACGCCACCA (SEQ ID<br>NO:975)  | 6.043  | 125A |
| CCT_KRAS_30bp_guide_U-flip_30_11    | <i>KRAS</i>    | RESCUE | CCU | U/20 | A18A | GGUAUCGUCAAUGCACUCU<br>UGCCUACGCCAC (SEQ ID<br>NO:976)  | 9.465  | 125A |
| CCT_KRAS_30bp_guide_U-flip_30_13    | <i>KRAS</i>    | RESCUE | CCU | U/18 | A18A | GCUGUAUCGUCAAUGCACU<br>CUUGCCUACGCC (SEQ ID<br>NO:977)  | 3.147  | 125A |
| TCG_PPIB_30bp_guide_U-flip_30_7     | <i>PPIB</i>    | RESCUE | UCG | U/24 | I18I | GCCCCGUUAUGAGGGCGGC<br>GGCAAGGAGCAC (SEQ ID<br>NO:978)  | 5.536  | 125A |
| TCG_PPIB_30bp_guide_U-flip_30_9     | <i>PPIB</i>    | RESCUE | UCG | U/22 | I18I | GGACCCCGCUAUGAGGGCG<br>GCGGCAAGGAGC (SEQ ID<br>NO:979)  | 8.914  | 125A |
| TCG_PPIB_30bp_guide_U-flip_30_11    | <i>PPIB</i>    | RESCUE | UCG | U/20 | I18I | GCGGACCCCGCUAUGAGGG<br>CGGCGGCAAGGA (SEQ ID<br>NO:980)  | 10.340 | 125A |
| TCG_PPIB_30bp_guide_U-flip_30_13    | <i>PPIB</i>    | RESCUE | UCG | U/18 | I18I | GGACGGACCCCGCUAUGAG<br>GGCGCGGCAAG (SEQ ID<br>NO:981)   | 8.663  | 125A |
| ACG_PPIB_30bp_guide_U-flip_30_7     | <i>PPIB</i>    | RESCUE | ACG | U/24 | R7C  | GUGUUGCUUUCGGAGAGG<br>CGCAGCAUCCACA (SEQ ID<br>NO:982)  | 34.213 | 125A |
| ACG_PPIB_30bp_guide_U-flip_30_9     | <i>PPIB</i>    | RESCUE | ACG | U/22 | R7C  | GCAUGUUGCUUUCGGAGA<br>GGCGCAGCAUCA (SEQ ID<br>NO:983)   | 31.652 | 125A |
| ACG_PPIB_30bp_guide_U-flip_30_11    | <i>PPIB</i>    | RESCUE | ACG | U/20 | R7C  | GUUCAUGUUGCUUUCGGA<br>GAGGCGCAGCAUC' (SEQ ID<br>NO:984) | 26.969 | 125A |
| ACG_PPIB_30bp_guide_U-flip_30_13    | <i>PPIB</i>    | RESCUE | ACG | U/18 | R7C  | GCCUUCAGUUGCUUUCGG<br>AGAGGCGCAGCA (SEQ ID<br>NO:985)   | 22.539 | 125A |
| GCG_PPIB_30bp_guide_U-flip_30_7     | <i>PPIB</i>    | RESCUE | GCG | U/24 | A19V | GGACCCUCGUAUGAGGGCG<br>GCGGCAAGGAGC (SEQ ID<br>NO:986)  | 1.030  | 125A |
| GCG_PPIB_30bp_guide_U-flip_30_9     | <i>PPIB</i>    | RESCUE | GCG | U/22 | A19V | GCGGACCCUCGUAUGAGGG<br>CGGCGGCAAGGA (SEQ ID<br>NO:987)  | 4.819  | 125A |
| GCG_PPIB_30bp_guide_U-flip_30_11    | <i>PPIB</i>    | RESCUE | GCG | U/20 | A19V | GGACGGACCCUCGUAUGAG<br>GGCGCGGCAAG (SEQ ID<br>NO:988)   | 4.778  | 125A |
| GCG_PPIB_30bp_guide_U-flip_30_13    | <i>PPIB</i>    | RESCUE | GCG | U/18 | A19V | GAAGACGGACCCUCGAUG<br>AGGCGCGCGCA (SEQ ID<br>NO:989)    | 1.115  | 125A |
| CCG_PPIB_30bp_guide_U-flip_30_7     | <i>PPIB</i>    | RESCUE | CCG | U/24 | S21S | GGAAGACUGACCCCGCAU<br>GAGGGCGGCGGC (SEQ ID<br>NO:990)   | 3.150  | 125A |
| CCG_PPIB_30bp_guide_U-flip_30_9     | <i>PPIB</i>    | RESCUE | CCG | U/22 | S21S | GAAGAAGACUGACCCCGG<br>AUGAGGGCGGCG (SEQ ID<br>NO:991)   | 2.659  | 125A |
| CCG_PPIB_30bp_guide_U-flip_30_11    | <i>PPIB</i>    | RESCUE | CCG | U/20 | S21S | GGGAAGAAGACUGACCCCG<br>CGAUGAGGGCGG (SEQ ID<br>NO:992)  | 6.101  | 125A |
| CCG_PPIB_30bp_guide_U-flip_30_13    | <i>PPIB</i>    | RESCUE | CCG | U/18 | S21S | GCAGGAAGAAGACUGACCC<br>CGCAUGAGGGC (SEQ ID<br>NO:993)   | 4.372  | 125A |
| TCG_SMARCA4_30bp_guide_U-flip_30_7  | <i>SMARCA4</i> | RESCUE | UCG | U/24 | S85L | GUCGUCCUACAUGCCCUUC<br>UCAUGCAUGGAC (SEQ ID<br>NO:994)  | 5.692  | 125A |
| TCG_SMARCA4_30bp_guide_U-flip_30_9  | <i>SMARCA4</i> | RESCUE | UCG | U/22 | S85L | GGGUCGUCCUACAUGCCCU<br>UCUCAUGGAC (SEQ ID<br>NO:995)    | 6.807  | 125A |
| TCG_SMARCA4_30bp_guide_U-flip_30_11 | <i>SMARCA4</i> | RESCUE | UCG | U/20 | S85L | GCGGGUCGUCCUACAUGCC<br>CUUCUCAUGCAU (SEQ ID<br>NO:996)  | 3.724  | 125A |
| TCG_SMARCA4_30bp_guide_U-flip_30_13 | <i>SMARCA4</i> | RESCUE | UCG | U/18 | S85L | GCGGGGUCGUCCUACAUG<br>CCUUCUCAUGC (SEQ ID<br>NO:997)    | 2.274  | 125A |
| ACG_SMARCA4_30bp_guide_U-flip_30_7  | <i>SMARCA4</i> | RESCUE | ACG | U/24 | D86D | GCGGGUCUUCGACAUGCC<br>CUUCUCAUGCAU (SEQ ID<br>NO:998)   | 3.689  | 125A |
| ACG_SMARCA4_30bp_guide_U-flip_30_9  | <i>SMARCA4</i> | RESCUE | ACG | U/22 | D86D | GCGGGGUCUUCGACAUG<br>CCUUCUCAUGC (SEQ ID<br>NO:999)     | 4.868  | 125A |
| ACG_SMARCA4_30bp_guide_U-flip_30_11 | <i>SMARCA4</i> | RESCUE | ACG | U/20 | D86D | GAGCGGGUCUUCGACA<br>UGCCUUCUCAU (SEQ ID<br>NO:1000)     | 6.943  | 125A |
| ACG_SMARCA4_30bp_guide_U-flip_30_13 | <i>SMARCA4</i> | RESCUE | ACG | U/18 | D86D | GGUAGCGGGUCUUCGACA<br>CAUGCCUUCUC (SEQ ID<br>NO:1001)   | 5.785  | 125A |
| GCG_SMARCA4_30bp_guide_U-flip_30_7  | <i>SMARCA4</i> | RESCUE | GCG | U/24 | R89C | GUGUAGCUGGGGUCGUCG<br>ACAUGCCUUCU (SEQ ID<br>NO:1002)   | 0.642  | 125A |

[2288]

|                                     |         |        |     |      |      |   |        |      |
|-------------------------------------|---------|--------|-----|------|------|---|--------|------|
| GCG_SMARCA4_30bp_guide_U-flip_30_9  | SMARCA4 | RESCUE | GCG | U/22 | R89C | GGUUGUAGCUCGGGUCGU<br>CCGACAUGCCUU (SEQ ID<br>NO:1003)  | 1.808  | 125A |
| GCG_SMARCA4_30bp_guide_U-flip_30_11 | SMARCA4 | RESCUE | GCG | U/20 | R89C | GUGGUUGUAGCUCGGGUC<br>GUCCGACAUGCCC (SEQ ID<br>NO:1004) | 2.277  | 125A |
| GCG_SMARCA4_30bp_guide_U-flip_30_13 | SMARCA4 | RESCUE | GCG | U/18 | R89C | GUCUGGUUGUAGCUCGGG<br>UCGUCGACAUGC (SEQ ID<br>NO:1005)  | 1.323  | 125A |
| CCG_SMARCA4_30bp_guide_U-flip_30_7  | SMARCA4 | RESCUE | CCG | U/24 | P88L | GUAGCGCUGGUCGUCGAC<br>AUGCCCUUCA (SEQ ID<br>NO:1006)    | 4.412  | 125A |
| CCG_SMARCA4_30bp_guide_U-flip_30_9  | SMARCA4 | RESCUE | CCG | U/22 | P88L | GUGUAGCGCUGGUCGUCG<br>ACAUGCCCUUCA (SEQ ID<br>NO:1007)  | 2.911  | 125A |
| CCG_SMARCA4_30bp_guide_U-flip_30_11 | SMARCA4 | RESCUE | CCG | U/20 | P88L | GGUUGUAGCGCUGGUCGU<br>CCGACAUGCCUU (SEQ ID<br>NO:1008)  | 4.819  | 125A |
| CCG_SMARCA4_30bp_guide_U-flip_30_13 | SMARCA4 | RESCUE | CCG | U/18 | P88L | GUGGUUGUAGCGCUGGUC<br>GUCCGACAUGCCC (SEQ ID<br>NO:1009) | 0.841  | 125A |
| S33F_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | RESCUE | UCU | C/24 | S33F | GAUCCACAGUCCAGGUA<br>GACUGUUGCUGC (SEQ ID<br>NO:1010)   | 7.351  | 125B |
| S33F_CTNNB1_30bp_C-flip_guide_30_9  | CTNNB1  | RESCUE | UCU | C/22 | S33F | GGGAUCCACAGUCCAGGU<br>AAGACUGUUGCU (SEQ ID<br>NO:1011)  | 8.783  | 125B |
| S33F_CTNNB1_30bp_C-flip_guide_30_11 | CTNNB1  | RESCUE | UCU | C/20 | S33F | GGAUGAUCCACAGUCCAG<br>GUAAGACUGUUG (SEQ ID<br>NO:1012)  | 6.063  | 125B |
| S33F_CTNNB1_30bp_C-flip_guide_30_13 | CTNNB1  | RESCUE | UCU | C/18 | S33F | GGAAUGGAUCCACAGUCC<br>AGGUAAGACUGU (SEQ ID<br>NO:1013)  | 7.893  | 125B |
| H36Y_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | RESCUE | CCA | C/24 | H36Y | GCAGAAUCGAUCCAGAGU<br>CCAGUAAGACU (SEQ ID<br>NO:1014)   | 0.406  | 125B |
| H36Y_CTNNB1_30bp_C-flip_guide_30_9  | CTNNB1  | RESCUE | CCA | C/22 | H36Y | GACCAGAAUCGAUCCAGA<br>GUCCAGGUAAGA (SEQ ID<br>NO:1015)  | 1.178  | 125B |
| H36Y_CTNNB1_30bp_C-flip_guide_30_11 | CTNNB1  | RESCUE | CCA | C/20 | H36Y | GGCACCAGAAUCGAUCC<br>GAGUCCAGGUA (SEQ ID<br>NO:1016)    | 0.369  | 125B |
| H36Y_CTNNB1_30bp_C-flip_guide_30_13 | CTNNB1  | RESCUE | CCA | C/18 | H36Y | GUGGCACCAGAAUCGAUCC<br>CAGAGUCCAGGU (SEQ ID<br>NO:1017) | 0.589  | 125B |
| S37F_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | RESCUE | UCU | C/24 | S37F | GGCACCACAUGGAUCC<br>GAGUCCAGGUA (SEQ ID<br>NO:1018)     | 7.132  | 125B |
| S37F_CTNNB1_30bp_C-flip_guide_30_9  | CTNNB1  | RESCUE | UCU | C/22 | S37F | GUGGCACCACAUGGAUCC<br>CAGAGUCCAGGU (SEQ ID<br>NO:1019)  | 8.465  | 125B |
| S37F_CTNNB1_30bp_C-flip_guide_30_11 | CTNNB1  | RESCUE | UCU | C/20 | S37F | GAGUGGCACCACAUGGAU<br>UCCAGAGUCCAG (SEQ ID<br>NO:1020)  | 9.422  | 125B |
| S37F_CTNNB1_30bp_C-flip_guide_30_13 | CTNNB1  | RESCUE | UCU | C/18 | S37F | GGUAGUGGCACCACAUGG<br>AUCCAGAGUCC (SEQ ID<br>NO:1021)   | 9.734  | 125B |
| T41I_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | RESCUE | ACC | C/24 | T41I | GGCUGUGCUAGUGGCACCA<br>GAAUGGAUCCA (SEQ ID<br>NO:1022)  | 5.252  | 125B |
| T41I_CTNNB1_30bp_C-flip_guide_30_9  | CTNNB1  | RESCUE | ACC | C/22 | T41I | GGAGCUGUGCUAGUGGCA<br>CCAGAAUGGAUCC (SEQ ID<br>NO:1023) | 12.765 | 125B |
| T41I_CTNNB1_30bp_C-flip_guide_30_11 | CTNNB1  | RESCUE | ACC | C/20 | T41I | GAGGAGCUGUGCUAGUGG<br>CACCAGAAUGGAU (SEQ ID<br>NO:1024) | 14.196 | 125B |
| T41I_CTNNB1_30bp_C-flip_guide_30_13 | CTNNB1  | RESCUE | ACC | C/18 | T41I | GGAAGGAGCUGUGCUAGU<br>GGCACCAGAAUGG (SEQ ID<br>NO:1025) | 11.364 | 125B |
| P44L_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | RESCUE | CCU | C/24 | P44L | GAGAGAACGAGCUGUGGU<br>AGUGGCACCAGAA (SEQ ID<br>NO:1026) | 4.636  | 125B |
| P44L_CTNNB1_30bp_C-flip_guide_30_9  | CTNNB1  | RESCUE | CCU | C/22 | P44L | GUCAGAGAACGAGCUGUG<br>GUAGUGGCACCAG (SEQ ID<br>NO:1027) | 8.413  | 125B |
| P44L_CTNNB1_30bp_C-flip_guide_30_11 | CTNNB1  | RESCUE | CCU | C/20 | P44L | GACUCAGAGAACGAGCUGU<br>GGUAGUGGCACC (SEQ ID<br>NO:1028) | 8.276  | 125B |
| P44L_CTNNB1_30bp_C-flip_guide_30_13 | CTNNB1  | RESCUE | CCU | C/18 | P44L | GCCACUCAGAGAACGAGCU<br>GUGGUAGUGGCA (SEQ ID<br>NO:1029) | 1.775  | 125B |
| P44S_CTNNB1_30bp_C-flip_guide_30_7  | CTNNB1  | RESCUE | UCU | C/24 | P44S | GGAGAAGCAGCUGUGGUA<br>GUGGCACCAGAAU (SEQ ID<br>NO:1030) | 15.256 | 125B |

[2289]

|                                      |        |        |     |      |      |   |        |      |
|--------------------------------------|--------|--------|-----|------|------|---|--------|------|
| P44S_CTNNB1_30bp_C-flip_guide_3_0_9  | CTNNB1 | RESCUE | UCU | C/22 | P44S | GCAGAGAAGCAGCUGUGG<br>UAGUGGCACCAGA (SEQ ID<br>NO:1031) | 14.639 | 125B |
| P44S_CTNNB1_30bp_C-flip_guide_3_0_11 | CTNNB1 | RESCUE | UCU | C/20 | P44S | GCUCAGAGAAGCAGCUGUG<br>GUAGUGGCACCA (SEQ ID<br>NO:1032) | 13.489 | 125B |
| P44S_CTNNB1_30bp_C-flip_guide_3_0_13 | CTNNB1 | RESCUE | UCU | C/18 | P44S | GCACUCAGAGAAGCAGCUG<br>UGGUAGUGGCAC (SEQ ID<br>NO:1033) | 13.906 | 125B |
| S45F_CTNNB1_30bp_C-flip_guide_3_0_7  | CTNNB1 | RESCUE | UCU | C/24 | S45F | GCUCAGACAAGGAGCUGUG<br>GUAGUGGCACCA (SEQ ID<br>NO:1034) | 6.550  | 125B |
| S45F_CTNNB1_30bp_C-flip_guide_3_0_9  | CTNNB1 | RESCUE | UCU | C/22 | S45F | GCACUCAGACAAGGAGCUG<br>UGGUAGUGGCAC (SEQ ID<br>NO:1035) | 8.816  | 125B |
| S45F_CTNNB1_30bp_C-flip_guide_3_0_11 | CTNNB1 | RESCUE | UCU | C/20 | S45F | GACCACUCAGACAAGGAGC<br>UGUGGUAGUGGC (SEQ ID<br>NO:1036) | 11.980 | 125B |
| S45F_CTNNB1_30bp_C-flip_guide_3_0_13 | CTNNB1 | RESCUE | UCU | C/18 | S45F | GUUACCACUCAGACAAGGA<br>GCUGUGGUAGUG (SEQ ID<br>NO:1037) | 7.397  | 125B |
| TCG_KRAS_30bp_C-flip_guide_30_7      | KRAS   | RESCUE | UCG | C/24 | L56L | GUGUGUCCAGAAUUCCAA<br>GAGACAGGUUUC (SEQ ID<br>NO:1038)  | 13.024 | 125B |
| TCG_KRAS_30bp_C-flip_guide_30_9      | KRAS   | RESCUE | UCG | C/22 | L56L | GGCUGUGUCCAGAAUUC<br>AAGAGACAGGUU (SEQ ID<br>NO:1039)   | 8.716  | 125B |
| TCG_KRAS_30bp_C-flip_guide_30_11     | KRAS   | RESCUE | UCG | C/20 | L56L | GCUGCUGUGUCCAGAAU<br>CCAAGAGACAGG (SEQ ID<br>NO:1040)   | 10.067 | 125B |
| TCG_KRAS_30bp_C-flip_guide_30_13     | KRAS   | RESCUE | UCG | C/18 | L56L | GACCUGCUGUGUCCAGAA<br>AUCCAAGAGACA (SEQ ID<br>NO:1041)  | 6.656  | 125B |
| ACG_KRAS_30bp_C-flip_guide_30_7      | KRAS   | RESCUE | ACG | C/24 | D30D | GAUAUCCUCCACAAAUG<br>AUUCUGAAUUG (SEQ ID<br>NO:1042)    | 13.954 | 125B |
| ACG_KRAS_30bp_C-flip_guide_30_9      | KRAS   | RESCUE | ACG | C/22 | D30D | GUCAUAUCCUCCACAAA<br>UGAUUCUGAAU (SEQ ID<br>NO:1043)    | 12.962 | 125B |
| ACG_KRAS_30bp_C-flip_guide_30_11     | KRAS   | RESCUE | ACG | C/20 | D30D | GGAUCAUAUCCUCCACAA<br>AAUGAUUCUGAA (SEQ ID<br>NO:1044)  | 23.138 | 125B |
| ACG_KRAS_30bp_C-flip_guide_30_13     | KRAS   | RESCUE | ACG | C/18 | D30D | GUGGAUCAUAUCCUCCAC<br>AAAUGAUUCUG (SEQ ID<br>NO:1045)   | 16.629 | 125B |
| GCG_KRAS_30bp_C-flip_guide_30_7      | KRAS   | RESCUE | GCG | C/24 | G13G | GGCCUACCCACCAGCUCC<br>AACUACCACAAG (SEQ ID<br>NO:1046)  | 0.394  | 125B |
| GCG_KRAS_30bp_C-flip_guide_30_9      | KRAS   | RESCUE | GCG | C/22 | G13G | GUUGCCUACCCACCAGCU<br>CCAACUACCACA (SEQ ID<br>NO:1047)  | 0.886  | 125B |
| GCG_KRAS_30bp_C-flip_guide_30_11     | KRAS   | RESCUE | GCG | C/20 | G13G | GUCUUGCCUACCCACCAG<br>CUCCAACUACCA (SEQ ID<br>NO:1048)  | 0.950  | 125B |
| GCG_KRAS_30bp_C-flip_guide_30_13     | KRAS   | RESCUE | GCG | C/18 | G13G | GACUCUUGCCUACCCACC<br>AGCUCCAACUAC (SEQ ID<br>NO:1049)  | 0.311  | 125B |
| CCT_KRAS_30bp_C-flip_guide_30_7      | KRAS   | RESCUE | CCU | C/24 | A18A | GCGUCAACGCACUCUUGCC<br>UACGCCACCAGC (SEQ ID<br>NO:1050) | 1.407  | 125B |
| CCT_KRAS_30bp_C-flip_guide_30_9      | KRAS   | RESCUE | CCU | C/22 | A18A | GAUCGUAACGCACUCUUG<br>CCUACGCCACCA (SEQ ID<br>NO:1051)  | 5.431  | 125B |
| CCT_KRAS_30bp_C-flip_guide_30_11     | KRAS   | RESCUE | CCU | C/20 | A18A | GGUAUCGUAACGCACUCU<br>UGCCUACGCCAC (SEQ ID<br>NO:1052)  | 7.769  | 125B |
| CCT_KRAS_30bp_C-flip_guide_30_13     | KRAS   | RESCUE | CCU | C/18 | A18A | GCUGUAUCGUAACGCACU<br>CUUGCCUACGCC (SEQ ID<br>NO:1053)  | 1.685  | 125B |
| TCG_PPIB_30bp_C-flip_guide_30_7      | PPIB   | RESCUE | UCG | C/24 | I18I | GCCCCGCAUGAGGGCGGC<br>GGCAAGGAGCAC (SEQ ID<br>NO:1054)  | 4.128  | 125B |
| TCG_PPIB_30bp_C-flip_guide_30_9      | PPIB   | RESCUE | UCG | C/22 | I18I | GGACCCCGCAUGAGGGCG<br>GCGCAAGGAGC (SEQ ID<br>NO:1055)   | 6.324  | 125B |
| TCG_PPIB_30bp_C-flip_guide_30_11     | PPIB   | RESCUE | UCG | C/20 | I18I | GCGGACCCCGCAUGAGGG<br>CGGCGCAAGGA (SEQ ID<br>NO:1056)   | 6.386  | 125B |
| TCG_PPIB_30bp_C-flip_guide_30_13     | PPIB   | RESCUE | UCG | C/18 | I18I | GGACGGACCCCGCAUGAG<br>GGCGGCGCAAG (SEQ ID<br>NO:1057)   | 5.348  | 125B |
| ACG_PPIB_30bp_C-flip_guide_30_7      | PPIB   | RESCUE | ACG | C/24 | R7C  | GUGUUGCCUUCGGAGAGG<br>CGCAGCAUCCACA (SEQ ID<br>NO:1058) | 28.575 | 125B |

[2290]

|                                     |                |        |     |      |      |   |        |      |
|-------------------------------------|----------------|--------|-----|------|------|---|--------|------|
| ACG_PPIB_30bp_C-flip_guide_30_9     | <i>PPIB</i>    | RESCUE | ACG | C/22 | R7C  | GCAUGUUGCCUUCGGAGA<br>GGCGCAGCAUCCA (SEQ ID<br>NO:1059) | 24.992 | 125B |
| ACG_PPIB_30bp_C-flip_guide_30_11    | <i>PPIB</i>    | RESCUE | ACG | C/20 | R7C  | GUUCAUGUUGCCUUCGGA<br>GAGGCGCAGCAUC (SEQ ID<br>NO:1060) | 24.160 | 125B |
| ACG_PPIB_30bp_C-flip_guide_30_13    | <i>PPIB</i>    | RESCUE | ACG | C/18 | R7C  | GCCUUAUGUUGCCUUCGG<br>AGAGGCGCAGCA (SEQ ID<br>NO:1061)  | 5.582  | 125B |
| GCG_PPIB_30bp_C-flip_guide_30_7     | <i>PPIB</i>    | RESCUE | GCG | C/24 | A19V | GGACCCCGCAUGAGGGCG<br>GCGGCAAGGAGC (SEQ ID<br>NO:1062)  | 0.584  | 125B |
| GCG_PPIB_30bp_C-flip_guide_30_9     | <i>PPIB</i>    | RESCUE | GCG | C/22 | A19V | GCGGACCCCGCAUGAGGG<br>CGGCGGCAAGGA (SEQ ID<br>NO:1063)  | 5.708  | 125B |
| GCG_PPIB_30bp_C-flip_guide_30_11    | <i>PPIB</i>    | RESCUE | GCG | C/20 | A19V | GGACGGACCCCGCAUGAG<br>GGCGGGGCAAG (SEQ ID<br>NO:1064)   | 3.700  | 125B |
| GCG_PPIB_30bp_C-flip_guide_30_13    | <i>PPIB</i>    | RESCUE | GCG | C/18 | A19V | GAAGACGGACCCCGCAUG<br>AGGCGGCGGCA (SEQ ID<br>NO:1065)   | 0.667  | 125B |
| CCG_PPIB_30bp_C-flip_guide_30_7     | <i>PPIB</i>    | RESCUE | CCG | C/24 | S21S | GGAAGACGGACCCCGCAU<br>GAGGGCGGCGG (SEQ ID<br>NO:1066)   | 3.719  | 125B |
| CCG_PPIB_30bp_C-flip_guide_30_9     | <i>PPIB</i>    | RESCUE | CCG | C/22 | S21S | GAAGAAGACGGACCCCGG<br>AUGAGGGCGGCG (SEQ ID<br>NO:1067)  | 4.255  | 125B |
| CCG_PPIB_30bp_C-flip_guide_30_11    | <i>PPIB</i>    | RESCUE | CCG | C/20 | S21S | GGGAAGAAGACGGACCCG<br>CGAUGAGGGCGG (SEQ ID<br>NO:1068)  | 6.843  | 125B |
| CCG_PPIB_30bp_C-flip_guide_30_13    | <i>PPIB</i>    | RESCUE | CCG | C/18 | S21S | GCAGGAAGAAGACGGACCC<br>CGCAUGAGGGCG (SEQ ID<br>NO:1069) | 4.263  | 125B |
| TCG_SMARCA4_30bp_C-flip_guide_30_7  | <i>SMARCA4</i> | RESCUE | UCG | C/24 | S85L | GUCGUCCCACAUGCCUUC<br>UCAUGCAUGGAC (SEQ ID<br>NO:1070)  | 4.081  | 125B |
| TCG_SMARCA4_30bp_C-flip_guide_30_9  | <i>SMARCA4</i> | RESCUE | UCG | C/22 | S85L | GGGUCGUCCCACAUGCCU<br>UCUCAUGCAUGG (SEQ ID<br>NO:1071)  | 3.132  | 125B |
| TCG_SMARCA4_30bp_C-flip_guide_30_11 | <i>SMARCA4</i> | RESCUE | UCG | C/20 | S85L | GCGGGUCCUCCCACAUGCC<br>CUUCUCAUGCAU (SEQ ID<br>NO:1072) | 0.918  | 125B |
| TCG_SMARCA4_30bp_C-flip_guide_30_13 | <i>SMARCA4</i> | RESCUE | UCG | C/18 | S85L | GCGCGGGUCCUCCCACAUG<br>CCUUCUCAUGC (SEQ ID<br>NO:1073)  | 1.830  | 125B |
| ACG_SMARCA4_30bp_C-flip_guide_30_7  | <i>SMARCA4</i> | RESCUE | ACG | C/24 | D86D | GCGGGUCCUCCGACAUGCC<br>CUUCUCAUGCAU (SEQ ID<br>NO:1074) | 2.425  | 125B |
| ACG_SMARCA4_30bp_C-flip_guide_30_9  | <i>SMARCA4</i> | RESCUE | ACG | C/22 | D86D | GCGCGGUCCUCCGACAUG<br>CCUUCUCAUGC (SEQ ID<br>NO:1075)   | 3.392  | 125B |
| ACG_SMARCA4_30bp_C-flip_guide_30_11 | <i>SMARCA4</i> | RESCUE | ACG | C/20 | D86D | GAGCGGGUCCUCCGACA<br>UGCCUUCUCAU (SEQ ID<br>NO:1076)    | 4.485  | 125B |
| ACG_SMARCA4_30bp_C-flip_guide_30_13 | <i>SMARCA4</i> | RESCUE | ACG | C/18 | D86D | GGUAGCGGGUCCUCCGA<br>CAUGCCUUCUC (SEQ ID<br>NO:1077)    | 3.071  | 125B |
| GCG_SMARCA4_30bp_C-flip_guide_30_7  | <i>SMARCA4</i> | RESCUE | GCG | C/24 | R89C | GUGUAGCCGGGUCGUCCG<br>ACAUGCCUUCU (SEQ ID<br>NO:1078)   | 0.225  | 125B |
| GCG_SMARCA4_30bp_C-flip_guide_30_9  | <i>SMARCA4</i> | RESCUE | GCG | C/22 | R89C | GGUUGUAGCCGGGUCGUC<br>CGACAUGCCUUCU (SEQ ID<br>NO:1079) | 1.026  | 125B |
| GCG_SMARCA4_30bp_C-flip_guide_30_11 | <i>SMARCA4</i> | RESCUE | GCG | C/20 | R89C | GUGGUUGUAGCCGGGUC<br>GUCCGACAUGCC (SEQ ID<br>NO:1080)   | 1.737  | 125B |
| GCG_SMARCA4_30bp_C-flip_guide_30_13 | <i>SMARCA4</i> | RESCUE | GCG | C/18 | R89C | GUCUGGUUGUAGCCGGG<br>UCGUCCGACAUGC (SEQ ID<br>NO:1081)  | 0.603  | 125B |
| CCG_SMARCA4_30bp_C-flip_guide_30_7  | <i>SMARCA4</i> | RESCUE | CCG | C/24 | P88L | GUAGCGCGGUCGUCCGAC<br>AUGCCUUCUCA (SEQ ID<br>NO:1082)   | 2.895  | 125B |
| CCG_SMARCA4_30bp_C-flip_guide_30_9  | <i>SMARCA4</i> | RESCUE | CCG | C/22 | P88L | GUGUAGCGCGGUCGUCCG<br>ACAUGCCUUCU (SEQ ID<br>NO:1083)   | 2.766  | 125B |
| CCG_SMARCA4_30bp_C-flip_guide_30_11 | <i>SMARCA4</i> | RESCUE | CCG | C/20 | P88L | GGUUGUAGCGCGGUCGUC<br>CGACAUGCCUUCU (SEQ ID<br>NO:1084) | 4.845  | 125B |
| CCG_SMARCA4_30bp_C-flip_guide_30_13 | <i>SMARCA4</i> | RESCUE | CCG | C/18 | P88L | GUGGUUGUAGCGCGGUC<br>GUCCGACAUGCC (SEQ ID<br>NO:1085)   | 0.684  | 125B |
| NRAS_30bp_C-flip_guide_30_3         | <i>NRAS</i>    | RESCUE | UCC | C/28 | I2II | GUGCAUUGUCAGUGCGCU<br>UUCCCAACACCA (SEQ ID<br>NO:1086)  | 2.714  | 125C |

[2291]

|                               |              |        |     |      |      |  |        |      |
|-------------------------------|--------------|--------|-----|------|------|--|--------|------|
| NRAS_30bp_C-flip_guide_30_5   | <i>NRAS</i>  | RESCUE | UCC | C/26 | I2II | GGCUGCAUUGUCAGUGCGC<br>UUUCCCAACAC (SEQ ID<br>NO:1087)   | 8.839  | 125C |
| NRAS_30bp_C-flip_guide_30_7   | <i>NRAS</i>  | RESCUE | UCC | C/24 | I2II | GUAGCUGCAUUGUCAGUG<br>CGCUUUCCCAAC (SEQ ID<br>NO:1088)   | 10.690 | 125C |
| NRAS_30bp_C-flip_guide_30_9   | <i>NRAS</i>  | RESCUE | UCC | C/22 | I2II | GAUUAGCUGCAUUGUCAG<br>UGCGCUUUCCCA (SEQ ID<br>NO:1089)   | 13.278 | 125C |
| NRAS_30bp_C-flip_guide_30_11  | <i>NRAS</i>  | RESCUE | UCC | C/20 | I2II | GGGAUAGCUGCAUUGUC<br>AGUGCGCUUUCC (SEQ ID<br>NO:1090)    | 18.858 | 125C |
| NKFB1_30bp_C-flip_guide_30_3  | <i>NKFB1</i> | RESCUE | ACC | C/28 | P33S | GUGCUUGAAAACUUCUG<br>GAUAAAUAUUGU (SEQ ID<br>NO:1091)    | 16.138 | 125C |
| NKFB1_30bp_C-flip_guide_30_5  | <i>NKFB1</i> | RESCUE | ACC | C/26 | P33S | GUGUGCUUGAAAACUUC<br>UGGAUAAAUAUU (SEQ ID<br>NO:1092)    | 9.580  | 125C |
| NKFB1_30bp_C-flip_guide_30_7  | <i>NKFB1</i> | RESCUE | ACC | C/24 | P33S | GUCUGUGCUUGAAAACU<br>UCUGGAUAAAUA (SEQ ID<br>NO:1093)    | 14.701 | 125C |
| NKFB1_30bp_C-flip_guide_30_9  | <i>NKFB1</i> | RESCUE | ACC | C/22 | P33S | GCAUCUGUGCUUGAAAAC<br>UUCUGGAUAAA (SEQ ID<br>NO:1094)    | 13.808 | 125C |
| NKFB1_30bp_C-flip_guide_30_11 | <i>NKFB1</i> | RESCUE | ACC | C/20 | P33S | GGCCAUCUGUGCUUGAAA<br>ACUUCUGGAUUA (SEQ ID<br>NO:1095)   | 21.529 | 125C |
| EZH2_30bp_C-flip_guide_30_3   | <i>EZH2</i>  | RESCUE | UCA | C/28 | F32F | GCUCAACCCUUGAGCUGU<br>CUCAGUCGCAUG (SEQ ID<br>NO:1096)   | 2.696  | 125C |
| EZH2_30bp_C-flip_guide_30_5   | <i>EZH2</i>  | RESCUE | UCA | C/26 | F32F | GGUCUCAACCCUUGAGCU<br>GUCUCAGUCGCA (SEQ ID<br>NO:1097)   | 0.106  | 125C |
| EZH2_30bp_C-flip_guide_30_7   | <i>EZH2</i>  | RESCUE | UCA | C/24 | F32F | GUCGUCUCAACCCUUGAG<br>CUGUCUCAGUCG (SEQ ID<br>NO:1098)   | 11.539 | 125C |
| EZH2_30bp_C-flip_guide_30_9   | <i>EZH2</i>  | RESCUE | UCA | C/22 | F32F | GGCUCGUCUCAACCCUUG<br>AGCUGUCUCAGU (SEQ ID<br>NO:1099)   | 10.710 | 125C |
| EZH2_30bp_C-flip_guide_30_11  | <i>EZH2</i>  | RESCUE | UCA | C/20 | F32F | GCAGCUCGUCUCAACCCU<br>UGAGCUGUCUCA (SEQ ID<br>NO:1100)   | 15.855 | 125C |
| NF2_30bp_C-flip_guide_30_3    | <i>NF2</i>   | RESCUE | ACG | C/28 | T21M | GACCUCUUGGGUUGCUUC<br>UCUUGAGAGAGC (SEQ ID<br>NO:1101)   | 8.318  | 125C |
| NF2_30bp_C-flip_guide_30_5    | <i>NF2</i>   | RESCUE | ACG | C/26 | T21M | GGAACCCUUGGGUUGCU<br>UCCUCUUGAGAGA (SEQ ID<br>NO:1102)   | 18.846 | 125C |
| NF2_30bp_C-flip_guide_30_7    | <i>NF2</i>   | RESCUE | ACG | C/24 | T21M | GGUGAACCCUUGGGUUG<br>CUUCCUCUUGAGA (SEQ ID<br>NO:1103)   | 18.617 | 125C |
| NF2_30bp_C-flip_guide_30_9    | <i>NF2</i>   | RESCUE | ACG | C/22 | T21M | GCGGUGAACCCUUGGGU<br>UGCUUCCUCUUGA (SEQ ID<br>NO:1104)   | 24.904 | 125C |
| NF2_30bp_C-flip_guide_30_11   | <i>NF2</i>   | RESCUE | ACG | C/20 | T21M | GCACGGUGAACCCUUGGG<br>UUGCUUCCUCU (SEQ ID<br>NO:1105)    | 14.735 | 125C |
| RAF1_30bp_C-flip_guide_30_3   | <i>RAF1</i>  | RESCUE | UCC | C/28 | P30S | GAGCAGAGAGCAGCUGG<br>AGCCAACAACAC' (SEQ ID<br>NO:1106)   | 8.257  | 125C |
| RAF1_30bp_C-flip_guide_30_5   | <i>RAF1</i>  | RESCUE | UCC | C/26 | P30S | GGUAGCAGAGAGCAGCU<br>GGAGCAUCAAAC (SEQ ID<br>NO:1107)    | 14.435 | 125C |
| RAF1_30bp_C-flip_guide_30_7   | <i>RAF1</i>  | RESCUE | UCC | C/24 | P30S | GUUGUAGCAGAGAGCAG<br>CUGGAGCCAUAAA (SEQ ID<br>NO:1108)   | 20.867 | 125C |
| RAF1_30bp_C-flip_guide_30_9   | <i>RAF1</i>  | RESCUE | UCC | C/22 | P30S | GUAUUGUAGCAGAGAGC<br>AGCUGGAGCCAUC' (SEQ ID<br>NO:1109)  | 13.821 | 125C |
| RAF1_30bp_C-flip_guide_30_11  | <i>RAF1</i>  | RESCUE | UCC | C/20 | P30S | GACUUAUUGUAGCAGAGAU<br>GCAGCUGGAGCCA (SEQ ID<br>NO:1110) | 16.277 | 125C |
| NRAS_30bp_U-flip_guide_30_3   | <i>NRAS</i>  | RESCUE | UCC | U/28 | I2II | GUGUAUUGUCAGUGCGCU<br>UUUCCCAACACCA (SEQ ID<br>NO:1111)  | 3.500  | 125C |
| NRAS_30bp_U-flip_guide_30_5   | <i>NRAS</i>  | RESCUE | UCC | U/26 | I2II | GGCUGUAUUGUCAGUGCG<br>CUUUCCCAACAC' (SEQ ID<br>NO:1112)  | 8.805  | 125C |
| NRAS_30bp_U-flip_guide_30_7   | <i>NRAS</i>  | RESCUE | UCC | U/24 | I2II | GUAGCUGUAUUGUCAGUG<br>CGCUUUCCCAAC (SEQ ID<br>NO:1113)   | 12.296 | 125C |
| NRAS_30bp_U-flip_guide_30_9   | <i>NRAS</i>  | RESCUE | UCC | U/22 | I2II | GAUAGCUGUAUUGUCAG<br>UGCGCUUUCCCA (SEQ ID<br>NO:1114)    | 12.521 | 125C |

[2292]

|                               |              |        |     |      |       |  |        |      |
|-------------------------------|--------------|--------|-----|------|-------|--|--------|------|
| NRAS_30bp_U-flip_guide_30_11  | <i>NRAS</i>  | RESCUE | UCC | U/20 | I2II  | GGGAUUAGCUGUAUUGUC<br>AGUGCGCUUUUCC (SEQ ID<br>NO:1115)  | 18.399 | 125C |
| NKFB1_30bp_U-flip_guide_30_3  | <i>NKFB1</i> | RESCUE | ACC | U/28 | P33S  | GUGUUUGAAAACUUCUG<br>GAUUAUUUUUUGU (SEQ ID<br>NO:1116)   | 14.277 | 125C |
| NKFB1_30bp_U-flip_guide_30_5  | <i>NKFB1</i> | RESCUE | ACC | U/26 | P33S  | GUGUGUUUGAAAACUUC<br>UGGAUUAAAUAUU (SEQ ID<br>NO:1117)   | 10.928 | 125C |
| NKFB1_30bp_U-flip_guide_30_7  | <i>NKFB1</i> | RESCUE | ACC | U/24 | P33S  | GUCUGUUUGAAAACU<br>UCUGGAUUAAAUA (SEQ ID<br>NO:1118)     | 18.012 | 125C |
| NKFB1_30bp_U-flip_guide_30_9  | <i>NKFB1</i> | RESCUE | ACC | U/22 | P33S  | GCAUCUGUUUGAAA<br>CUUCUGGAUUAAA (SEQ ID<br>NO:1119)      | 21.468 | 125C |
| NKFB1_30bp_U-flip_guide_30_11 | <i>NKFB1</i> | RESCUE | ACC | U/20 | P33S  | GGCCAUCUGUUUGAAA<br>UACUUCUGGAUU (SEQ ID<br>NO:1120)     | 24.376 | 125C |
| EZH2_30bp_U-flip_guide_30_3   | <i>EZH2</i>  | RESCUE | UCA | U/28 | F32F  | GCUUAACCUUGAGCUGU<br>CUCAGUCGCAUG (SEQ ID<br>NO:1121)    | 9.307  | 125C |
| EZH2_30bp_U-flip_guide_30_5   | <i>EZH2</i>  | RESCUE | UCA | U/26 | F32F  | GGUCUUAACCUUGAGCU<br>GUCUCAGUCGCA (SEQ ID<br>NO:1122)    | 9.393  | 125C |
| EZH2_30bp_U-flip_guide_30_7   | <i>EZH2</i>  | RESCUE | UCA | U/24 | F32F  | GUCGUCUUAACCUUGAG<br>CUGUCUCAGUCG (SEQ ID<br>NO:1123)    | 8.525  | 125C |
| EZH2_30bp_U-flip_guide_30_9   | <i>EZH2</i>  | RESCUE | UCA | U/22 | F32F  | GGCUCGUCUUAACCUUG<br>AGCUCUCAGUC (SEQ ID<br>NO:1124)     | 6.976  | 125C |
| EZH2_30bp_U-flip_guide_30_11  | <i>EZH2</i>  | RESCUE | UCA | U/20 | F32F  | GCAGCUCGUCUUAACCU<br>UGAGCUCUCUCA (SEQ ID<br>NO:1125)    | 9.534  | 125C |
| NF2_30bp_U-flip_guide_30_3    | <i>NF2</i>   | RESCUE | ACG | U/28 | T21M  | GACUUCUUGGGUUGCUCC<br>UCUUGAGAGAGC (SEQ ID<br>NO:1126)   | 7.253  | 125C |
| NF2_30bp_U-flip_guide_30_5    | <i>NF2</i>   | RESCUE | ACG | U/26 | T21M  | GGAAUCUUCUUGGGUUGCU<br>UCCUCUUGAGAGA (SEQ ID<br>NO:1127) | 16.618 | 125C |
| NF2_30bp_U-flip_guide_30_7    | <i>NF2</i>   | RESCUE | ACG | U/24 | T21M  | GGUGAACUUCUUGGGUUG<br>CUUCUCUUGAGAGA (SEQ ID<br>NO:1128) | 15.696 | 125C |
| NF2_30bp_U-flip_guide_30_9    | <i>NF2</i>   | RESCUE | ACG | U/22 | T21M  | GCGGUGAACUUCUUGGGU<br>UGCUUCUCUUGA (SEQ ID<br>NO:1129)   | 18.984 | 125C |
| NF2_30bp_U-flip_guide_30_11   | <i>NF2</i>   | RESCUE | ACG | U/20 | T21M  | GCACGGUGAACUUCUUGG<br>GUUGCUUCUCUU (SEQ ID<br>NO:1130)   | 16.393 | 125C |
| RAF1_30bp_U-flip_guide_30_3   | <i>RAF1</i>  | RESCUE | UCC | U/28 | P30S  | GAGUAGAGAGCAGCUGG<br>AGCCAUAACAC' (SEQ ID<br>NO:1131)    | 9.450  | 125C |
| RAF1_30bp_U-flip_guide_30_5   | <i>RAF1</i>  | RESCUE | UCC | U/26 | P30S  | GGUAGUAGAGAGCAGC<br>GGAGCCAUAAC (SEQ ID<br>NO:1132)      | 15.056 | 125C |
| RAF1_30bp_U-flip_guide_30_7   | <i>RAF1</i>  | RESCUE | UCC | U/24 | P30S  | GUUGUAGUAGAGAGCAG<br>CUGGAGCCAUA (SEQ ID<br>NO:1133)     | 20.868 | 125C |
| RAF1_30bp_U-flip_guide_30_9   | <i>RAF1</i>  | RESCUE | UCC | U/22 | P30S  | GUAUUGUAGUAGAGAGC<br>AGCUGGAGCAUC' (SEQ ID<br>NO:1134)   | 15.518 | 125C |
| RAF1_30bp_U-flip_guide_30_11  | <i>RAF1</i>  | RESCUE | UCC | U/20 | P30S  | GACUUAUUGUAGAGAG<br>GCAGCUGGAGCCA (SEQ ID<br>NO:1135)    | 20.003 | 125C |
| STAT3 S727 C-flip             | <i>STAT3</i> | RESCUE | UCC | C/22 | S727F | GUGCGGGGCACAUCGGCA<br>GGUCAUUGGUU (SEQ ID<br>NO:1136)    | 6.127  | 129B |
| STAT3 S727 U-flip             | <i>STAT3</i> | RESCUE | UCC | U/22 | S727F | GUGCGGGGUACAUCGGC<br>AGGUCAAUGGUU (SEQ ID<br>NO:1137)    | 3.132  | 129B |
| STAT3 S727 G-flip             | <i>STAT3</i> | RESCUE | UCC | G/22 | S727F | GUGCGGGGGACAUCGGC<br>AGGUCAAUGGUU (SEQ ID<br>NO:1138)    | 0.154  | 129B |
| STAT3 S727 A-flip             | <i>STAT3</i> | RESCUE | UCC | A/22 | S727F | GUGCGGGGAACAUCGGCA<br>GGUCAUUGGUU (SEQ ID<br>NO:1139)    | 0.156  | 129B |
| STAT1 S727 C-flip             | <i>STAT1</i> | RESCUE | UCC | C/22 | S727F | GCCUCAGGACACAUGGGGA<br>GCAGGUUGUCUG (SEQ ID<br>NO:1140)  | 1.141  | 129D |
| STAT1 S727 U-flip             | <i>STAT1</i> | RESCUE | UCC | U/22 | S727F | GCCUCAGGAUACAUGGGGA<br>GCAGGUUGUCUG (SEQ ID<br>NO:1141)  | 0.801  | 129D |
| STAT1 S727 G-flip             | <i>STAT1</i> | RESCUE | UCC | G/22 | S727F | GCCUCAGGAGACAUGGGGA<br>GCAGGUUGUCUG (SEQ ID<br>NO:1142)  | 0.223  | 129D |

[2293]

|                   |       |        |     |      |       |   |       |      |
|-------------------|-------|--------|-----|------|-------|---|-------|------|
| STAT1_S727_A-flip | STAT1 | RESCUE | UCC | A/22 | S727F | GCCUCAGGAAACAUGGGGA<br>GCAGGUUGUCUG (SEQ ID<br>NO:1143)                       | 0.193 | 129D |
| STAT1_Y701_C-flip | STAT1 | REPAIR | UAU | C/34 | Y701C | GCAACUCAGUCUUGAUACA<br>UCCAGUUCUUUAGGGCCAU<br>CAAGUCCAUUG (SEQ ID<br>NO:1144) | 1.289 | 129E |
| STAT1_Y701_U-flip | STAT1 | REPAIR | UAU | U/34 | Y701C | GCAACUCAGUCUUGAUUA<br>UCCAGUUCUUUAGGGCCAU<br>CAAGUCCAUUG (SEQ ID<br>NO:1145)  | 0.050 | 129E |
| STAT1_Y701_G-flip | STAT1 | REPAIR | UAU | G/34 | Y701C | GCAACUCAGUCUUGAUAGA<br>UCCAGUUCUUUAGGGCCAU<br>CAAGUCCAUUG (SEQ ID<br>NO:1146) | 0.011 | 129E |
| STAT3_Y705_C-flip | STAT3 | REPAIR | UAC | C/34 | Y705C | GAAACUUGGUCUUCAGGCA<br>UGGGCAGCGCUACCGGGU<br>CAGCUUCAGGAU (SEQ ID<br>NO:1147) | 5.080 | 139B |
| STAT3_S727_30_5   | STAT3 | RESCUE | UCC | C/26 | S727F | GGGGACAUCGGCAGGUCA<br>AUGGUAUUGCU (SEQ ID<br>NO:1148)                         | 4.842 | 139C |
| STAT3_S727_30_7   | STAT3 | RESCUE | UCC | C/24 | S727F | CGGGGGACAUCGGCAGG<br>UCAAGGUAUUG (SEQ ID<br>NO:1149)                          | 2.990 | 139C |
| STAT3_S727_30_11  | STAT3 | RESCUE | UCC | C/20 | S727F | AGUCGGGGGACAUCGG<br>CAGGCUAAUGGU (SEQ ID<br>NO:1150)                          | 6.786 | 139C |
| STAT1_S727_30_5   | STAT1 | RESCUE | UCC | C/26 | S727F | AGGAGACAUGGGGAGCAG<br>GUUGUCUGUGGU (SEQ ID<br>NO:1151)                        | 2.128 | 139C |
| STAT1_S727_30_7   | STAT1 | RESCUE | UCC | C/24 | S727F | UCAGGAGACAUGGGGAGC<br>AGGUUGUCUGUG (SEQ ID<br>NO:1152)                        | 3.215 | 139C |
| STAT1_S727_30_11  | STAT1 | RESCUE | UCC | C/20 | S727F | CUCCUCAGGAGACAUGGG<br>AGCAGGUUGUC (SEQ ID<br>NO:1153)                         | 2.146 | 139C |

[2294] 表31用于合成靶标编辑的指导序列

[2295]

| 名称                          | 被靶向的基因 | REPAIR/<br>RESCUE | 基序  | 碱基翻<br>转/位<br>置 | 密码子<br>变化 | 间隔子序列   | 编辑百分比<br>(第一圈) | 第一圈  |
|-----------------------------|--------|-------------------|-----|-----------------|-----------|---|----------------|------|
| NM_000016.5_C-flip_guide    | ACADM  | RESCUE            | ACA | C/7             | H67Y      | GUAUCAUCUUCUGCAGCC<br>ACUGGGAUGAUU (SEQ<br>ID NO:1154)  | 7.500          | 127A |
| NM_000018.3_C-flip_guide    | ACADVL | RESCUE            | GCG | C/9             | A283V     | GUCUCCACCCAAAAGCU<br>GUGAUCUUCUCU (SEQ<br>ID NO:1155)   | 6.274          | 127A |
| NM_000071.2_C-flip_guide    | CBS    | RESCUE            | GCG | C/9             | R109C     | GGAACUCACCCUUGGCA<br>AGAGCUCACACUU (SEQ<br>ID NO:1156)  | 6.140          | 127A |
| NM_000138.4_C-flip_guide    | FBN1   | RESCUE            | GCG | C/5             | R1408C    | GGAGCCCUCAUCAAGGUC<br>UGUACAAGUGAAG (SEQ<br>ID NO:1157) | 13.751         | 127A |
| NM_000141.4_C-flip_guide    | FGFR2  | RESCUE            | CCC | C/7             | P267S     | GCUGUGCGGCAUUGCC<br>GGCAGUCCGGCUU (SEQ<br>ID NO:1158)   | 16.131         | 127A |
| NM_000152.4_C-flip_guide    | GAA    | RESCUE            | CCC | C/7             | P552L     | GGCCUGGCGGUCCCCC<br>AACCACCCAGGC (SEQ ID<br>NO:1159)    | 2.644          | 127A |
| NM_000341.3_C-flip_guide    | SLC3A1 | RESCUE            | ACG | C/7             | T467M     | GAGAAGCCUGUUAUCAC<br>GUUGACAUACUGA (SEQ<br>ID NO:1160)  | 28.198         | 127A |
| NM_000375.2_C-flip_guide    | UROS   | RESCUE            | ACG | C/9             | R73C      | GCUCAAACCUAACUCUG<br>CUGCUUCCACUGC (SEQ<br>ID NO:1161)  | 13.632         | 127A |
| NM_000431.3_C-flip_guide    | MVK    | RESCUE            | ACA | C/9             | T268I     | GUGGCAUCUCUUGAGGUC<br>AGGAGGGGGCCA (SEQ<br>ID NO:1162)  | 13.775         | 127A |
| NM_000551.3_C-flip_guide    | VHL    | RESCUE            | CCG | C/7             | P158L     | GUCUUUCCGUAUACAC<br>UGGCAGUGUGAUA (SEQ<br>ID NO:1163)   | 22.466         | 127A |
| NM_001256850.1_C-flip_guide | TTN    | RESCUE            | ACG | C/9             | R30071C   | GCUUCCACCUGGGCCAG<br>GGGAAUCAAGCAC (SEQ<br>ID NO:1164)  | 23.207         | 127A |
| NM_002397.4_C-flip_guide    | MEF2C  | RESCUE            | ACG | C/9             | TIM       | GUUCUCCCUUAGUCCCC<br>GUUUUUCUUCUCU (SEQ<br>ID NO:1165)  | 2.313          | 127A |
| NM_002474.2_C-flip_guide    | MYH11  | RESCUE            | CCG | C/9             | P1264L    | GUGGACUGCCGCUCCUG<br>ACCUGCGCCUCA (SEQ ID<br>NO:1166)   | 34.770         | 127A |

[2296]

|                                      |               |        |     |     |        |  |        |       |
|--------------------------------------|---------------|--------|-----|-----|--------|--|--------|-------|
| NM_002834.4_C-flip_guide             | <i>PTPN11</i> | RESCUE | CCU | C/9 | L285F  | GAUGAUCAACGGGCAGGA<br>UGUUUUUAUUAUCU (SEQ<br>ID NO:1167) | 0.711  | 127A  |
| NM_004004.5_C-flip_guide             | <i>GJB2</i>   | RESCUE | ACG | C/5 | R77W   | GGCCCUUAGCCGGAUGUG<br>GGAGAUUGGGAAG (SEQ<br>ID NO:1168)  | 5.347  | 1127A |
| NM_004572.3_C-flip_guide             | <i>PKP2</i>   | RESCUE | CCG | C/9 | R796C  | GUGUGUAACGGCAGAGG<br>CUGUAGUUCAAU (SEQ<br>ID NO:1169)    | 20.024 | 127A  |
| NM_005609.3_C-flip_guide             | <i>PYGM</i>   | RESCUE | GCG | C/9 | R798W  | GCCGCGUCCUCUCUUG<br>GGUUCUUGUACAA (SEQ<br>ID NO:1170)    | 5.946  | 127A  |
| NM_005633.3_C-flip_guide             | <i>SOS1</i>   | RESCUE | ACG | C/9 | T269M  | GCAUCUGCCUUUCUACU<br>GUAUCUUCUUAU (SEQ<br>ID NO:1171)    | 39.582 | 127A  |
| NM_014139.2_C-flip_guide             | <i>SCN11A</i> | RESCUE | CCG | C/9 | P396L  | GCAACAGCCCGGUUAAG<br>UUAUCAGGUAGA (SEQ<br>ID NO:1172)    | 32.565 | 127A  |
| NM_014874.3_C-flip_guide             | <i>MFN2</i>   | RESCUE | CCG | C/9 | P76L   | GCAACAGCCCGGUUAAG<br>UUAUCAGGUAGA (SEQ<br>ID NO:1173)    | 20.769 | 127A  |
| NM_015559.2_C-flip_guide             | <i>SETBP1</i> | RESCUE | ACU | C/7 | T87II  | GGUCCACUGCCGUGUC<br>GCUGGGGAUCGUC (SEQ<br>ID NO:1174)    | 20.136 | 127A  |
| NM_020630.4_C-flip_guide             | <i>RET</i>    | RESCUE | CCG | C/5 | R620C  | GUCGCCGAAGCACUUCUC<br>CUCCUCAGGGAAG (SEQ<br>ID NO:1175)  | 15.762 | 127A  |
| NM_000016.5_F_30bp_C-flip_guide_30_5 | <i>ACADM</i>  | RESCUE | ACA | C/5 | H67Y   | GUCAUCUUCGCAGCCAC<br>UGGGAUGAUUCC (SEQ<br>ID NO:1176)    | 2.900  | 127B  |
| NM_000016.5_F_30bp_C-flip_guide_30_7 | <i>ACADM</i>  | RESCUE | ACA | C/7 | H67Y   | GUAUCAUCUUCGAGCC<br>ACUGGGAUGAUU (SEQ<br>ID NO:1177)     | 7.500  | 127B  |
| NM_000016.5_F_30bp_C-flip_guide_30_9 | <i>ACADM</i>  | RESCUE | ACA | C/9 | H67Y   | GUUUAUCAUCUUCGAG<br>CCACUGGGAUGAU (SEQ<br>ID NO:1178)    | 6.477  | 127B  |
| NM_000018.3_F_30bp_C-flip_guide_30_5 | <i>ACADVL</i> | RESCUE | GCG | C/5 | A283V  | GCACCCCAAAGCUGUGA<br>UCUUCUCCUUCAC (SEQ<br>ID NO:1178)   | 4.474  | 127B  |
| NM_000018.3_F_30bp_C-flip_guide_30_7 | <i>ACADVL</i> | RESCUE | GCG | C/7 | A283V  | GUCCACCCCAAAGCUGU<br>GAUCUUCUCCUUC (SEQ<br>ID NO:1179)   | 5.357  | 127B  |
| NM_000018.3_F_30bp_C-flip_guide_30_9 | <i>ACADVL</i> | RESCUE | GCG | C/9 | A283V  | GUCUCCACCCCAAAGCU<br>GUGAUUCUCCUUC (SEQ<br>ID NO:1180)   | 6.274  | 127B  |
| NM_000071.2_F_30bp_C-flip_guide_30_5 | <i>CBS</i>    | RESCUE | GCG | C/5 | R109C  | GUCACCCUUGGCCAAGAG<br>CUCACACUUCAGG (SEQ<br>ID NO:1181)  | 1.269  | 127B  |
| NM_000071.2_F_30bp_C-flip_guide_30_7 | <i>CBS</i>    | RESCUE | GCG | C/7 | R109C  | GACUCACCCUUGGCCAAG<br>AGCUCACACUUA (SEQ<br>ID NO:1182)   | 2.346  | 127B  |
| NM_000071.2_F_30bp_C-flip_guide_30_9 | <i>CBS</i>    | RESCUE | GCG | C/9 | R109C  | GGAACUCACCCUUGGCCA<br>AGAGCUCACACUU (SEQ<br>ID NO:1183)  | 6.140  | 127B  |
| NM_000138.4_F_30bp_C-flip_guide_30_5 | <i>FBN1</i>   | RESCUE | GCG | C/5 | R1408C | GGAGCCCUCAUCAAGGUC<br>UGUACAAGUGAAG (SEQ<br>ID NO:1184)  | 13.751 | 127B  |
| NM_000138.4_F_30bp_C-flip_guide_30_7 | <i>FBN1</i>   | RESCUE | GCG | C/7 | R1408C | GCAGAGCCCUCAUCAAGG<br>UCUGUACAAGUGA (SEQ<br>ID NO:1185)  | 12.998 | 127B  |
| NM_000138.4_F_30bp_C-flip_guide_30_9 | <i>FBN1</i>   | RESCUE | GCG | C/9 | R1408C | GCUCAGAGCCCUCAUCA<br>GGUCUGUACAAGU (SEQ<br>ID NO:1186)   | 10.845 | 127B  |
| NM_000141.4_F_30bp_C-flip_guide_30_5 | <i>FGFR2</i>  | RESCUE | CCC | C/5 | P267S  | GGUGGCGGCAUUUGCCGG<br>CAGUCCGGCUUGG (SEQ<br>ID NO:1187)  | 5.882  | 127B  |
| NM_000141.4_F_30bp_C-flip_guide_30_7 | <i>FGFR2</i>  | RESCUE | CCC | C/7 | P267S  | GCUGUGGCGGCAUUUGCC<br>GGCAGUCCGGCUU (SEQ<br>ID NO:1188)  | 16.131 | 127B  |
| NM_000141.4_F_30bp_C-flip_guide_30_9 | <i>FGFR2</i>  | RESCUE | CCC | C/9 | P267S  | GCACUGUGGCGGCAUUUG<br>CCGGCAGUCCGGC (SEQ<br>ID NO:1189)  | 13.994 | 127B  |
| NM_000152.4_F_30bp_C-flip_guide_30_5 | <i>GAA</i>    | RESCUE | CCC | C/5 | P552L  | GCUGGCGGUCUCCCCCAA<br>CCACCCAGGCAC (SEQ ID<br>NO:1190)   | 2.316  | 127B  |
| NM_000152.4_F_30bp_C-flip_guide_30_7 | <i>GAA</i>    | RESCUE | CCC | C/7 | P552L  | GGCCUGGCGGUCUCCCCC<br>AACCACCCAGGC (SEQ ID<br>NO:1191)   | 2.644  | 127B  |
| NM_000152.4_F_30bp_C-flip_guide_30_9 | <i>GAA</i>    | RESCUE | CCC | C/9 | P552L  | GCCGCCUGGCGGUCUCCC<br>CCACCCAGGC (SEQ ID<br>NO:1192)     | 0.168  | 127B  |
| NM_000341.3_F_30bp_C-flip_guide_30_5 | <i>SLC3A1</i> | RESCUE | ACG | C/5 | T467M  | GAAGCCUGUUAUCACGU<br>UGACAUCUGAUU (SEQ<br>ID NO:1193)    | 26.164 | 127B  |

[2297]

|   |               |        |     |     |         |  |        |      |
|---|---------------|--------|-----|-----|---------|--|--------|------|
| NM_000341.3_F_30bp_C-flip_guide_30_7    | <i>SLC3A1</i> | RESCUE | ACG | C/7 | T467M   | GAGAAGCCUGUUCACACGUUGACAUACUGA (SEQ ID NO:1194)    | 28.198 | 127B |
| NM_000341.3_F_30bp_C-flip_guide_30_9    | <i>SLC3A1</i> | RESCUE | ACG | C/9 | T467M   | GAAAAGAAGCCUGUUCACACGUUGACAUACUGA (SEQ ID NO:1195) | 24.276 | 127B |
| NM_000375.2_F_30bp_C-flip_guide_30_5    | <i>UROS</i>   | RESCUE | ACG | C/5 | R73C    | GAAACCUAACUCUGCUGCUUCCACUGCUCUG (SEQ ID NO:1196)   | 8.539  | 127B |
| NM_000375.2_F_30bp_C-flip_guide_30_7    | <i>UROS</i>   | RESCUE | ACG | C/7 | R73C    | GCCAAACCUAACUCUGCUGCUUCCACUGCUC (SEQ ID NO:1197)   | 7.367  | 127B |
| NM_000375.2_F_30bp_C-flip_guide_30_9    | <i>UROS</i>   | RESCUE | ACG | C/9 | R73C    | GCUCCAAACCUAACUCUGCUGCUUCCACUGCUC (SEQ ID NO:1198) | 13.632 | 127B |
| NM_000431.3_F_30bp_C-flip_guide_30_5    | <i>MVK</i>    | RESCUE | ACA | C/5 | T268I   | GAUCUCUUGAGGUCAGGAGGGGGCCACGAU (SEQ ID NO:1199)    | 2.858  | 127B |
| NM_000431.3_F_30bp_C-flip_guide_30_7    | <i>MVK</i>    | RESCUE | ACA | C/7 | T268I   | GGCAUCUCUUGAGGUCAGGAGGGGGCCACG (SEQ ID NO:1200)    | 13.008 | 127B |
| NM_000431.3_F_30bp_C-flip_guide_30_9    | <i>MVK</i>    | RESCUE | ACA | C/9 | T268I   | GUGGCAUCUCUUGAGGUCAGGAGGGGGGCCA (SEQ ID NO:1201)   | 13.775 | 127B |
| NM_000551.3_F_30bp_C-flip_guide_30_5    | <i>VHL</i>    | RESCUE | CCG | C/5 | P158L   | GUUCCGAGUAUACACUGGCAGUGUGAUUU (SEQ ID NO:1202)     | 21.328 | 127B |
| NM_000551.3_F_30bp_C-flip_guide_30_7    | <i>VHL</i>    | RESCUE | CCG | C/7 | P158L   | GUCUUCCGAGUAUACACUGGCAGUGUGUAU (SEQ ID NO:1203)    | 22.466 | 127B |
| NM_000551.3_F_30bp_C-flip_guide_30_9    | <i>VHL</i>    | RESCUE | CCG | C/9 | P158L   | GGCUCUUCCGAGUAUACACUGGCAGUGUGUA (SEQ ID NO:1204)   | 20.534 | 127B |
| NM_001256850.1_F_30bp_C-flip_guide_30_5 | <i>TTN</i>    | RESCUE | ACG | C/5 | R30071C | GCCACCUAGGGCCAGGGGAUCAAGCACUUUG (SEQ ID NO:1205)   | 6.388  | 127B |
| NM_001256850.1_F_30bp_C-flip_guide_30_7 | <i>TTN</i>    | RESCUE | ACG | C/7 | R30071C | GUUCCACCUAGGGCCAGGGGAAUCAAGCACUU (SEQ ID NO:1206)  | 13.115 | 127B |
| NM_001256850.1_F_30bp_C-flip_guide_30_9 | <i>TTN</i>    | RESCUE | ACG | C/9 | R30071C | GCUUCCACCUAGGGCCAGGGGAAUCAAGCAC (SEQ ID NO:1207)   | 23.207 | 127B |
| NM_002397.4_F_30bp_C-flip_guide_30_5    | <i>MEF2C</i>  | RESCUE | ACG | C/5 | TIM     | GCCCCUAGUCCCGUUUUUUCUCUCUCUCU (SEQ ID NO:1208)     | 0.769  | 127B |
| NM_002397.4_F_30bp_C-flip_guide_30_7    | <i>MEF2C</i>  | RESCUE | ACG | C/7 | TIM     | GCUCUUUAGUCCCGUUUUUUCUCUCUCUCU (SEQ ID NO:1209)    | 1.270  | 127B |
| NM_002397.4_F_30bp_C-flip_guide_30_9    | <i>MEF2C</i>  | RESCUE | ACG | C/9 | TIM     | GUUCUUUAGUCCCGUUUUUUCUCUCUCUCU (SEQ ID NO:1210)    | 2.313  | 127B |
| NM_002474.2_F_30bp_C-flip_guide_30_5    | <i>MYH11</i>  | RESCUE | CCG | C/5 | P1264L  | GCUGCCGCUCCUGCACGCGCCUCCAGUU (SEQ ID NO:1211)      | 18.278 | 127B |
| NM_002474.2_F_30bp_C-flip_guide_30_7    | <i>MYH11</i>  | RESCUE | CCG | C/7 | P1264L  | GGACUGCCGCUCCUGCACGCGCCUCCAGC (SEQ ID NO:1212)     | 25.663 | 127B |
| NM_002474.2_F_30bp_C-flip_guide_30_9    | <i>MYH11</i>  | RESCUE | CCG | C/9 | P1264L  | GUGGACUGCCGCUCCUGCACGCGCCUCCAGC (SEQ ID NO:1213)   | 34.770 | 127B |
| NM_002834.4_F_30bp_C-flip_guide_30_5    | <i>PTPN11</i> | RESCUE | CCU | C/5 | L285F   | GUCAACGGGCAGGAUGUUUUUAUAUAUU (SEQ ID NO:1214)      | 0.520  | 127B |
| NM_002834.4_F_30bp_C-flip_guide_30_7    | <i>PTPN11</i> | RESCUE | CCU | C/7 | L285F   | GGAUCAACGGGCAGGAUGUUUUUAUAUAUU (SEQ ID NO:1215)    | 0.562  | 127B |
| NM_002834.4_F_30bp_C-flip_guide_30_9    | <i>PTPN11</i> | RESCUE | CCU | C/9 | L285F   | GAUGAUAACGGGCAGGAUGUUUUUAUAUU (SEQ ID NO:1216)     | 0.711  | 127B |
| NM_004004.5_F_30bp_C-flip_guide_30_5    | <i>GJB2</i>   | RESCUE | ACG | C/5 | R77W    | GGCCCCUAGCCGGAUGUGGAGAUAGGGGAAG (SEQ ID NO:1217)   | 5.347  | 127B |
| NM_004004.5_F_30bp_C-flip_guide_30_7    | <i>GJB2</i>   | RESCUE | ACG | C/7 | R77W    | GGGGCCCCUAGCCGGAUGUGGAGAUAGGGGA (SEQ ID NO:1218)   | 4.763  | 127B |
| NM_004004.5_F_30bp_C-flip_guide_30_9    | <i>GJB2</i>   | RESCUE | ACG | C/9 | R77W    | GCAGGGCCCCUAGCCGGAUGUGGAGAUAGGG (SEQ ID NO:1219)   | 5.188  | 127B |
| NM_004572.3_F_30bp_C-flip_guide_30_5    | <i>PKP2</i>   | RESCUE | CCG | C/5 | R796C   | GUAACGGCAGAGGCUGUAGUUCAAUGAGA (SEQ ID NO:1220)     | 12.583 | 127B |
| NM_004572.3_F_30bp_C-flip_guide_30_7    | <i>PKP2</i>   | RESCUE | CCG | C/7 | R796C   | GUGUAACGGCAGAGGCUGUAGUUCAAUGA (SEQ ID NO:1221)     | 13.950 | 127B |

[2298]

|                                      |        |        |     |      |       |   |        |      |
|--------------------------------------|--------|--------|-----|------|-------|---|--------|------|
| NM_004572.3_F_30bp_C-flip_guide_30_9 | PKP2   | RESCUE | CCG | C/9  | R796C | GUGUGUAACCGGCAGAGG<br>CUGUAGUUCAAU (SEQ ID NO:1222)   | 20.024 | 127B |
| NM_005609.3_F_30bp_C-flip_guide_30_5 | PYGM   | RESCUE | GCG | C/5  | R798W | GGUCCCCUCUCUUGGGUU<br>CUUGUACAAGGCG (SEQ ID NO:1223)  | 4.495  | 127B |
| NM_005609.3_F_30bp_C-flip_guide_30_7 | PYGM   | RESCUE | GCG | C/7  | R798W | GGCGUCCCCUCUCUUGGG<br>UUCUUGUACAAGG (SEQ ID NO:1224)  | 2.128  | 127B |
| NM_005609.3_F_30bp_C-flip_guide_30_9 | PYGM   | RESCUE | GCG | C/9  | R798W | GCCGCGUCCCCUCUCUUG<br>GGUUCUUGUACAA (SEQ ID NO:1225)  | 5.946  | 127B |
| NM_005633.3_F_30bp_C-flip_guide_30_5 | SOS1   | RESCUE | ACG | C/5  | T269M | GUGUCCUUUCUACUGUAU<br>CUUCUAUAUGGCC (SEQ ID NO:1226)  | 12.149 | 127B |
| NM_005633.3_F_30bp_C-flip_guide_30_7 | SOS1   | RESCUE | ACG | C/7  | T269M | GUCUGUCCUUUCUACUGU<br>AUCUUCUAUAUGG (SEQ ID NO:1227)  | 36.068 | 127B |
| NM_005633.3_F_30bp_C-flip_guide_30_9 | SOS1   | RESCUE | ACG | C/9  | T269M | GCAUCUGUCCUUUCUACU<br>GUAUCUUCUAUAU (SEQ ID NO:1228)  | 39.582 | 127B |
| NM_014139.2_F_30bp_C-flip_guide_30_5 | SCN11A | RESCUE | CCG | C/5  | P396L | GAGCCCGGGUUAAGUAAA<br>UCAGGUAAGAAGGA (SEQ ID NO:1229) | 22.989 | 127B |
| NM_014139.2_F_30bp_C-flip_guide_30_7 | SCN11A | RESCUE | CCG | C/7  | P396L | GACAGCCCGGUUAAGUU<br>AAUCAGGUAAGAAG (SEQ ID NO:1230)  | 16.783 | 127B |
| NM_014139.2_F_30bp_C-flip_guide_30_9 | SCN11A | RESCUE | CCG | C/9  | P396L | GCAACAGCCCGGUUAAG<br>UUAUCAGGUAAGA (SEQ ID NO:1231)   | 32.565 | 127B |
| NM_014874.3_F_30bp_C-flip_guide_30_5 | MFN2   | RESCUE | CCG | C/5  | P76L  | GGUCCCGAACCGUUCUU<br>CUGUGGUAACGGG (SEQ ID NO:1232)   | 7.822  | 127B |
| NM_014874.3_F_30bp_C-flip_guide_30_7 | MFN2   | RESCUE | CCG | C/7  | P76L  | GACGUCCCGAACCGUUC<br>UUCUGUGGUAACG (SEQ ID NO:1233)   | 9.585  | 127B |
| NM_014874.3_F_30bp_C-flip_guide_30_9 | MFN2   | RESCUE | CCG | C/9  | P76L  | GUGACGUCCCGAACCGU<br>UCUUCUGUGGUA (SEQ ID NO:1234)    | 20.769 | 127B |
| NM_015559.2_F_30bp_C-flip_guide_30_5 | SETBP1 | RESCUE | ACU | C/5  | T87II | GCCACUGCGCUGUCG<br>UGGGGAUCGUCUC (SEQ ID NO:1235)     | 4.665  | 127B |
| NM_015559.2_F_30bp_C-flip_guide_30_7 | SETBP1 | RESCUE | ACU | C/7  | T87II | GGUCCACUGCGCUGUC<br>GCUGGGGAUCGUC (SEQ ID NO:1236)    | 20.136 | 127B |
| NM_015559.2_F_30bp_C-flip_guide_30_9 | SETBP1 | RESCUE | ACU | C/9  | T87II | GCUGUCCACUGCGCUG<br>UCGUGGGGAUCG (SEQ ID NO:1237)     | 9.130  | 127B |
| NM_020630.4_F_30bp_C-flip_guide_30_5 | RET    | RESCUE | CCG | C/5  | R620C | GUCGCGGAAGCACUUC<br>CUCCUCAGGGAAG (SEQ ID NO:1238)    | 15.762 | 127B |
| NM_020630.4_F_30bp_C-flip_guide_30_7 | RET    | RESCUE | CCG | C/7  | R620C | GGCUCGCGGAAGCACU<br>UCCUCCUCAGGGA (SEQ ID NO:1239)    | 12.758 | 127B |
| NM_020630.4_F_30bp_C-flip_guide_30_9 | RET    | RESCUE | CCG | C/9  | R620C | GGGUCGCGGAAGCACU<br>UCUCCUCCUCAGG (SEQ ID NO:1240)    | 15.652 | 127B |
| ApoE4 rs429358 C flip 30             | APOE   | RESCUE | GCG | C/30 | C130R | gccacgucuccaungcgcgcc<br>cggg (SEQ ID NO:1241)        | 0.040  | 128  |
| ApoE4 rs429358 C flip 28             | APOE   | RESCUE | GCG | C/28 | C130R | ggcccacgucuccaungcgc<br>gccg (SEQ ID NO:1242)         | 0.199  | 128  |
| ApoE4 rs429358 C flip 26             | APOE   | RESCUE | GCG | C/26 | C130R | gcccacgucuccaungcgc<br>cagg (SEQ ID NO:1243)          | 0.534  | 128  |
| ApoE4 rs429358 C flip 24             | APOE   | RESCUE | GCG | C/24 | C130R | ggcgcccacgucuccaungc<br>ccca (SEQ ID NO:1244)         | 0.601  | 128  |
| ApoE4 rs429358 C flip 22             | APOE   | RESCUE | GCG | C/22 | C130R | ggcgcccacgucuccaungc<br>cgcc (SEQ ID NO:1245)         | 0.305  | 128  |
| ApoE4 rs429358 C flip 20             | APOE   | RESCUE | GCG | C/20 | C130R | gagcgcccacgucuccaung<br>cgcc (SEQ ID NO:1246)         | 0.251  | 128  |
| ApoE4 rs429358 C flip 18             | APOE   | RESCUE | GCG | C/18 | C130R | gccagcgcccacgucucca<br>ugcc (SEQ ID NO:1247)          | 0.098  | 128  |
| ApoE4 rs429358 C flip 16             | APOE   | RESCUE | GCG | C/16 | C130R | gcaccagcgcccacgucucc<br>augc (SEQ ID NO:1248)         | 0.066  | 128  |
| ApoE4 rs429358 U flip 30             | APOE   | RESCUE | GCG | U/30 | C130R | gucacgucuccaungc<br>gcgccacggg (SEQ ID NO:1249)       | 0.037  | 128  |
| ApoE4 rs429358 U flip 28             | APOE   | RESCUE | GCG | U/28 | C130R | ggcucacgucuccaungcgc<br>gccg (SEQ ID NO:1250)         | 0.831  | 128  |
| ApoE4 rs429358 U flip 26             | APOE   | RESCUE | GCG | U/26 | C130R | gcccacgucuccaungcgc<br>cagg (SEQ ID NO:1251)          | 0.787  | 128  |

[2299]

|                          |      |        |     |      |       |  |        |     |
|--------------------------|------|--------|-----|------|-------|--|--------|-----|
| ApoE4 rs429358 U flip 24 | APOE | RESCUE | GCG | U/24 | C130R | gggccgcucacgucuccaugucecgccca (SEQ ID NO:1252)         | 5.382  | 128 |
| ApoE4 rs429358 U flip 22 | APOE | RESCUE | GCG | U/22 | C130R | ggcgccgcucacgucuccaugucecgcc (SEQ ID NO:1253)          | 0.913  | 128 |
| ApoE4 rs429358 U flip 20 | APOE | RESCUE | GCG | U/20 | C130R | gaggcgccgcucacgucuccaugucecgcc (SEQ ID NO:1254)        | 0.654  | 128 |
| ApoE4 rs429358 U flip 18 | APOE | RESCUE | GCG | U/18 | C130R | ggcaggcgccgcucacgucuccaugucecgcc (SEQ ID NO:1255)      | 0.281  | 128 |
| ApoE4 rs429358 U flip 16 | APOE | RESCUE | GCG | U/16 | C130R | gcaccaggcgccgcucacgucuccaugucecgcc (SEQ ID NO:1256)    | 0.236  | 128 |
| ApoE4 rs7412 C flip 30   | APOE | RESCUE | GCG | C/30 | C176R | gccucucgaggcaucggcaucggcaggagga (SEQ ID NO:1257)       | 0.079  | 128 |
| ApoE4 rs7412 C flip 28   | APOE | RESCUE | GCG | C/28 | C176R | ggccucucgaggcaucggcaucggcaggagga (SEQ ID NO:1258)      | 2.900  | 128 |
| ApoE4 rs7412 C flip 26   | APOE | RESCUE | GCG | C/26 | C176R | gaggccucucgaggcaucggcaucggcaggagga (SEQ ID NO:1259)    | 0.908  | 128 |
| ApoE4 rs7412 C flip 24   | APOE | RESCUE | GCG | C/24 | C176R | ggcggccucucgaggcaucggcaucggcaggagga (SEQ ID NO:1260)   | 2.397  | 128 |
| ApoE4 rs7412 C flip 22   | APOE | RESCUE | GCG | C/22 | C176R | gugccaggccucucgaggcaucggcaggcauc (SEQ ID NO:1261)      | 3.436  | 128 |
| ApoE4 rs7412 C flip 20   | APOE | RESCUE | GCG | C/20 | C176R | gacucgaggccucucgaggcaucggcaggagga (SEQ ID NO:1262)     | 5.725  | 128 |
| ApoE4 rs7412 C flip 18   | APOE | RESCUE | GCG | C/18 | C176R | gacacugccaggccucucgaggcaucggcaggagga (SEQ ID NO:1263)  | 2.987  | 128 |
| ApoE4 rs7412 C flip 16   | APOE | RESCUE | GCG | C/16 | C176R | gguaacugccaggccucucgaggcaucggcaggagga (SEQ ID NO:1264) | 0.407  | 128 |
| ApoE4 rs7412 U flip 30   | APOE | RESCUE | GCG | U/30 | C176R | gucucucgaggcaucggcaucggcaggagga (SEQ ID NO:1265)       | 0.125  | 128 |
| ApoE4 rs7412 U flip 28   | APOE | RESCUE | GCG | U/28 | C176R | ggcucucgaggcaucggcaucggcaggagga (SEQ ID NO:1266)       | 3.633  | 128 |
| ApoE4 rs7412 U flip 26   | APOE | RESCUE | GCG | U/26 | C176R | gaggcucucgaggcaucggcaucggcaggagga (SEQ ID NO:1267)     | 1.087  | 128 |
| ApoE4 rs7412 U flip 24   | APOE | RESCUE | GCG | U/24 | C176R | ggcaggcucucgaggcaucggcaucggcaggagga (SEQ ID NO:1268)   | 3.305  | 128 |
| ApoE4 rs7412 U flip 22   | APOE | RESCUE | GCG | U/22 | C176R | gugccaggccucucgaggcaucggcaggcauc (SEQ ID NO:1269)      | 6.810  | 128 |
| ApoE4 rs7412 U flip 20   | APOE | RESCUE | GCG | U/20 | C176R | gacucgaggccucucgaggcaucggcaggagga (SEQ ID NO:1270)     | 10.902 | 128 |
| ApoE4 rs7412 U flip 18   | APOE | RESCUE | GCG | U/18 | C176R | gacacugccaggccucucgaggcaucggcaggagga (SEQ ID NO:1271)  | 9.357  | 128 |
| ApoE4 rs7412 U flip 16   | APOE | RESCUE | GCG | U/16 | C176R | gguaacugccaggccucucgaggcaucggcaggagga (SEQ ID NO:1272) | 0.643  | 128 |

[2300] 表32哺乳动物物质粒和图谱

[2301]

| 质粒     | 描述  | Benchling 链接  |
|--------|---|---|
| pC0043 | PspCas13b crRNA 主链  | <a href="https://benchling.com/s/s eq-OH6nMmZCZn930BW qcFNa">benchling.com/s/s eq-OH6nMmZCZn930BW qcFNa</a>   |
| pC0076 | CMV-dRanCas13b-mapkNES-GS-dADAR2 E488Q  | <a href="https://benchling.com/s/s eq-BulRvsrtwP4aEJtTqYM 2">benchling.com/s/s eq-BulRvsrtwP4aEJtTqYM 2</a>   |
| pC0077 | pCMV-dRanCas13b-mapkNES-GS-dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V) r8   | <a href="https://benchling.com/s/s eq-gQ13PMPPLkcO6OceAf mpC">benchling.com/s/s eq-gQ13PMPPLkcO6OceAf mpC</a> |
| pC0078 | pCMV-dRanCas13b-mapkNES-GS-dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) r16      | <a href="https://benchling.com/s/s eq-19Ytwwh0i0vSIbyXYZ 95">benchling.com/s/s eq-19Ytwwh0i0vSIbyXYZ 95</a>   |
| pC0079 | pCMV-dRanCas13b-mapkNES-GS-dADAR2(E488Q/V351G/S486A/T375A/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) RESCUE-S | <a href="https://benchling.com/s/s eq-WX6VnavLS6JaaZ54X AOx">benchling.com/s/s eq-WX6VnavLS6JaaZ54X AOx</a>   |
| pC0080 | pCMV-dCas13b12-HIVNES-GS-   | <a href="https://benchling.com/s/s eq-">benchling.com/s/s eq-</a>   |

[2302]

|        |   |  |
|--------|---|--|
|        | dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) RESCUE                                  | <a href="#">GQqPCRE9I6KnEfHks Qem</a>                      |
| pC0081 | pCMV-dCas13b12-HIVNES-GS-dADAR2(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T/S375A) RESCUE-S | <a href="#">benchling.com/s/s eq-qjbEAXZgupeRXBa8ab 1S</a> |
| pC0082 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(C82R)- polyA C 至 U 报告基因 TCG 基序  | <a href="#">benchling.com/s/s eq-Qjsg3Yx0r1Hs77GT58 BI</a> |
| pC0083 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(C82R)- polyA C 至 U 报告基因 GCG 基序  | <a href="#">benchling.com/s/s eq-Z8zWu3LdetcuYHAFGn pe</a> |
| pC0084 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(C82R)- polyA C 至 U 报告基因 ACG 基序  | <a href="#">benchling.com/s/s eq-G2Iag6I8NBQAXqbJno u5</a> |
| pC0085 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(C82R)- polyA C 至 U 报告基因 CCG 基序  | <a href="#">benchling.com/s/s eq-alkwhNUsFTg80TVmp quP</a> |
| pC0086 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(L77P)- polyA C 至 U 报告基因 CCA 基序  | <a href="#">benchling.com/s/s eq-IJ8Fm6vtF7GS676Q7p wS</a> |
| pC0087 | CMV-Cluciferase-polyA EF1a-G-荧光素酶(L77P)- polyA C 至 U 报告基因 CCT 基序  | <a href="#">benchling.com/s/s eq-5MMokwvx0Ajq6ML2s jjZ</a> |
| pC0088 | pCMV- ADAR2dd(E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) r16                              | <a href="#">benchling.com/s/s eq-YISAybq2YnuclVwYD y95</a> |
| pC0089 | pCMV-ADAR2 全长 (E488Q/V351G/S486A/T375S/S370C/P462A/N597I/L332I/I398V/K350I/M383L/D619G/S582T/V440I/S495N/K418E/S661T) r16                             | <a href="#">benchling.com/s/s eq-95ZpoHj9GhQFzlu3m6 cb</a> |
| pC0090 | 具有 Gluc/Cluc 的 β 连环蛋白报告基因 M50 Super 8x (TCF/LEF binding sites) TOPFlash   | <a href="#">benchling.com/s/s eq-jPxZnxs3wSeKZhgTTD Bu</a> |
| pC0091 | 具有 Gluc/Cluc 的 β 连环蛋白报告基因对照 M51 Super 8x (mutated TCF/LEF binding sites) FOPFlash   | <a href="#">benchling.com/s/s eq-130b6c9baCfw8R3ITgS R</a> |

[2303]

表33酵母质粒和图谱

[2304]

| 描述   | Benchling 链接  |
|--|---|
| pGAL-dRanCas13b-GS-dADAR2 [RESCUEr0 酵母]  | <a href="#">benchling.com/s/seq-w112aOHR 2gSe4P2aQ7VY</a> |
| pGAL-dRanCas13b-GS-dADAR2(V351/S486A/T375S) [RESCUEr3 酵母]                          | <a href="#">benchling.com/s/seq-saQngvNf6 i3GhSGF0H3I</a> |
| pGAL-dRanCas13b-GS-dADAR2(V351G/S486A/T375S/S370C/P462A/L332I) [RESCUEr7 酵母]       | <a href="#">benchling.com/s/seq-GIJ7BnpV 3Vd3XtKiIxdm</a> |
| pYES3/CT pADH1-HH-靶向-RanCas13b_DR--HDV-space-ADH1_终止子 His (P196L) [酵母靶向 His P196L] | <a href="#">benchling.com/s/seq-Xs2ffVMn4 FwwQ79zDDEo</a> |

|        |   |   |
|--------|---|---|
| [2305] | <p>pYES3/CT pADH1-HH-金门-BsmBi-Bsmbl_RanCas13b_DR--HDV-space-ADH1_终止子 His (P196L) [酵母靶向 His P196L NT]</p> <p>pYES3/CT pADH1-HH-Guide-RanCas13b_DR--HDV-space-ADH1_终止子 His S129P [酵母靶向 His S129P]</p> <p>pYES3/CT pADH1-HH-金门-BsmBi-Bsmbl_RanCas13b_DR--HDV-space-ADH1_终止子 His 基序 S129P [酵母靶向 His S129P NT]</p> <p>pYES3/CT pADH1-HH-Y66H-靶向-RanCas13b-DR-HDV-ADH1-term ATG-yeGFP Y66H [酵母靶向 GFP Y66H]</p> <p>pYES3/CT pADH1-HH-金门-BsmBi-Bsmbl_-RanCas13b_DR-HDV-ADH1-term ATG-yeGFP Y66H 报告基因[酵母靶向 GFP Y66H NT]</p> <p>pYES3/CT pADH1-HH-靶向-RanCas13b_DR--HDV-space-ADH1_终止子 His (S22P) [酵母靶向 30/26 His S22P]</p> <p>pYES3/CT pADH1-HH-靶向-RanCas13b_DR--HDV-space-ADH1_终止子 His (S22P) [酵母靶向 30/24 His S22P]</p> <p>pYES3/CT pADH1-HH-靶向-RanCas13b_DR--HDV-space-ADH1_终止子 His (S22P) [酵母靶向 30/22 His S22P]</p> <p>pYES3/CT pADH1-HH-靶向-RanCas13b_DR--HDV-space-ADH1_终止子 His (S22P) [酵母靶向 30/20 His S22P]</p> <p>pYES3/CT pADH1-HH-金门-BsmBi-Bsmbl_RanCas13b_DR--HDV-space-ADH1_终止子 His (S22P) [酵母靶向 His S22P NT]</p> | <p><a href="http://benchling.com/s/seq-UM9NjG7JKK0GFe9MowGo">benchling.com/s/seq-UM9NjG7JKK0GFe9MowGo</a></p> <p><a href="http://benchling.com/s/seq-EefJI5brqll3fm0B5Qc5">benchling.com/s/seq-EefJI5brqll3fm0B5Qc5</a></p> <p><a href="http://benchling.com/s/seq-bt7gOlrp8OuOoV3YJWZG">benchling.com/s/seq-bt7gOlrp8OuOoV3YJWZG</a></p> <p><a href="http://benchling.com/s/seq-hiMELqTYPT9y0nOAKEq2">benchling.com/s/seq-hiMELqTYPT9y0nOAKEq2</a></p> <p><a href="http://benchling.com/s/seq-OCWlvnjeKYwSbG8GELTQ">benchling.com/s/seq-OCWlvnjeKYwSbG8GELTQ</a></p> <p><a href="http://benchling.com/s/seq-ziOgQXpXGZwot9NDkFJf">benchling.com/s/seq-ziOgQXpXGZwot9NDkFJf</a></p> <p><a href="http://benchling.com/s/seq-Ni9S7NsmGwWEYQM7K1EF">benchling.com/s/seq-Ni9S7NsmGwWEYQM7K1EF</a></p> <p><a href="http://benchling.com/s/seq-yW539UdpUtm9kZbLafaJ">benchling.com/s/seq-yW539UdpUtm9kZbLafaJ</a></p> <p><a href="http://benchling.com/s/seq-z37Sri5Pds8UofSHRtGe">benchling.com/s/seq-z37Sri5Pds8UofSHRtGe</a></p> <p><a href="http://benchling.com/s/seq-6HoWi69XrcLL4nW2ya0V">benchling.com/s/seq-6HoWi69XrcLL4nW2ya0V</a></p> |
|--------|---|---|

[2306] 表34用于酵母靶向的指导序列

| 名称                 | 被靶向的基因 | 基序  | 碱基翻转/间隔长度/位置 | 密码子变化 | 间隔子序列   | 第一图  |
|--------------------|--------|-----|--------------|-------|---|------|
| [2307] His Y66H 靶向 | EGFP   | UCA | U/50/34      | Y66H  | aaacauugaacacc<br>auuaguuaaaguag<br>ugacuaagguuggc<br>cauggaac (SEQ ID NO:1273) | 113A |
| His L196P 靶向       | HIS    | CCU | U/50/34      | L196P | ucuuauggcaaccg<br>caugagccuugaac<br>gcacucacuaacg<br>gugaugau (SEQ ID NO:1274)  | 113C |
| His S129P 靶向       | HIS    | UCC | C/30/26      | S129P | gcuugcaagugccu<br>cauccaaaggcgca auu<br>(SEQ ID NO:1275)                        | 113D |

|        |                      |     |     |         |      |   |      |
|--------|----------------------|-----|-----|---------|------|---|------|
| [2308] | His S22P 靶向<br>30/26 | HIS | UCC | U/30/26 | S22P | aauguaaucgcaau<br>cugaauucugguuu ca<br>(SEQ ID NO:1276) | 113E |
|        | His S22P 靶向<br>30/24 | HIS | UCC | U/30/24 | S22P | uuuauguaaucgca<br>aucugaauucuggu uu<br>(SEQ ID NO:1277) | 113E |
|        | His S22P 靶向<br>30/22 | HIS | UCC | U/30/22 | S22P | cuuaaugguaaucg<br>caaucugaauucug gu<br>(SEQ ID NO:1278) | 113E |
|        | His S22P 靶向<br>30/20 | HIS | UCC | U/30/20 | S22P | ccuuuaaugguaau<br>cgcaaucugaauuc u<br>(SEQ ID NO:1279)  | 113E |

## [2309] 参考文献

- [2310] 1.0.0.Abudayyeh等,C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector.Science 353,aaf5573 (2016) .
- [2311] 2.C.Cassidy-Amstutz等,Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin.Biochemistry 55,3461-3468 (2016) .
- [2312] 3.S.Shmakov等,Discovery and Functional Characterization of Diverse Class 2CRISPR-Cas Systems.Mol Cell 60,385-397 (2015) .
- [2313] 4.A.A.Smargon等,Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28.Mol Cell 65,618-630e617 (2017) .
- [2314] 5.A.East-Seletsky等,Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection.Nature 538,270-273 (2016) .
- [2315] 6.0.0.Abudayyeh等,RNA targeting with CRISPR-Cas13.Nature 550, 280-284 (2017) .
- [2316] 7.D.B.T.Cox等,RNA editing with CRISPR-Cas13.Science 358, 1019-1027 (2017) .
- [2317] 8.T.Merkle等,Precise RNA editing by recruiting endogenous ADARs with antisense oligonucleotides.Nat Biotechnol 37,133-138 (2019) .
- [2318] 9.P.Vogel等,Efficient and precise editing of endogenous transcripts with SNAP-tagged ADARs.Nat Methods 15,535-538 (2018) .
- [2319] 10.M.Fukuda等,Construction of a guide-RNA for site-directed RNA mutagenesis utilising intracellular A-to-I RNA editing.Sci Rep 7,41478 (2017) .
- [2320] 11.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) .
- [2321] 12.M.F.Montiel-Gonzalez,I.C.Vallecillo-Viejo,J.J.Rosenthal,An efficient system for selectively altering genetic information within mRNAs. Nucleic Acids Res 44,e157 (2016) .
- [2322] 13.P.Vogel,M.F.Schneider,J.Wettengel,T.Stafforst,Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA.Angew Chem Int Ed Engl 53,6267-6271 (2014) .

- [2323] 14.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) .
- [2324] 15.H.A.Rees,D.R.Liu,Base editing:precision chemistry on the genome and transcriptome of living cells.Nat Rev Genet 19,770-788 (2018) .
- [2325] 16.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) .
- [2326] 17.K.Nishida等,Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems.Science 353, (2016) .
- [2327] 18.J.D.Salter,R.P.Bennett,H.C.Smith,The APOBEC Protein Family: United by Structure,Divergent in Function.Trends Biochem Sci 41,578-594 (2016) .
- [2328] 19.S.Jin等,Cytosine,but not adenine,base editors induce genome-wide off-target mutations in rice.Science, (2019) .
- [2329] 20.E.Zuo等,Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos.Science, (2019) .
- [2330] 21.J.Grünwald等,Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors.Nature, (2019) .
- [2331] 22.M.R.Macbeth等,Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing.Science 309,1534-1539 (2005) .
- [2332] 23.M.M.Matthews等,Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity.Nature structural& molecular biology 23,426-433 (2016) .
- [2333] 24.D.Katrekar等,In vivo RNA editing of point mutations via RNA-guided adenosine deaminases.Nat Methods 16,239-242 (2019) .
- [2334] 25.B.T.MacDonald,K.Tamai,X.He,Wnt/beta-catenin signaling: components, mechanisms,and diseases.Dev Cell 17,9-26 (2009) .
- [2335] 26.M.K.Chee,S.B.Haase,New and Redesigned pRS Plasmid Shuttle Vectors for Genetic Manipulation of Saccharomyces cerevisiae.G3 (Bethesda) 2, 515-526 (2012) .
- [2336] 27.M.F.Laughery等,New vectors for simple and streamlined CRISPR-Cas9 genome editing in Saccharomyces cerevisiae.Yeast 32,711-720 (2015) .
- [2337] 28.M.R.Macbeth,B.L.Bass,Large-scale overexpression and purification of ADARs from Saccharomyces cerevisiae for biophysical and biochemical studies.Methods Enzymol 424,319-331 (2007) .
- [2338] 29.H.Ng,N.Dean,Dramatic Improvement of CRISPR/Cas9 Editing in Candida albicans by Increased Single Guide RNA Expression.mSphere 2, (2017) .
- [2339] 30.R.Heim,D.C.Prasher,R.Y.Tsien,Wavelength mutations and

posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A* 91,12501-12504(1994) .

[2340] 31.Y.Wang,P.A.Beal,Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method. *Nucleic Acids Res* 44,9872-9880 (2016) .

[2341] 32.R.D.Gietz,R.H.Schiestl, Large-scale high-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* 2,38-41 (2007) .

[2342] 33.S.B.Kim,H.Suzuki,M.Sato,H.Tao, Superluminescent variants of marine luciferases for bioassays. *Anal Chem* 83,8732-8740 (2011) .

[2343] 34.M.T.Veeman,D.C.Slusarski,A.Kaykas,S.H.Louie,R.T.Moon, Zebrafish prickle,a modulator of noncanonical Wnt/Fz signaling,regulates gastrulation movements. *Curr Biol* 13,680-685 (2003) .

[2344] 本发明的所描述的方法、药物组合物和试剂盒的各种修改和变化对本领域技术人员将是显而易见的,而不脱离本发明的范围和精神。尽管已经结合特定实施方式描述了本发明,但是应当理解能够进行进一步的修改,并且所要求保护的发明不应不适当地限于这样的特定实施方式。实际上,对于本领域技术人员显而易见的用于实施本发明的所描述的方式的各种修改都旨在在本发明的范围内。本申请旨在覆盖总体上遵循本发明的原理的本发明的任何变型,使用或修改,并且包括这种与本公开的偏离均落入本发明所属领域内的已知惯例中,并且可以在阐述之前适用于本文的必要特征。

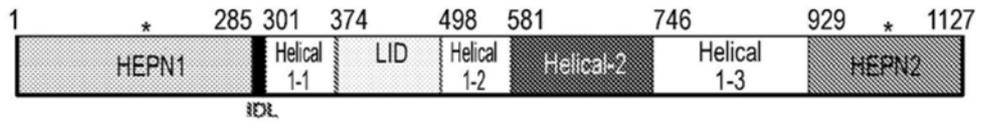


图1A

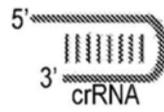


图1B

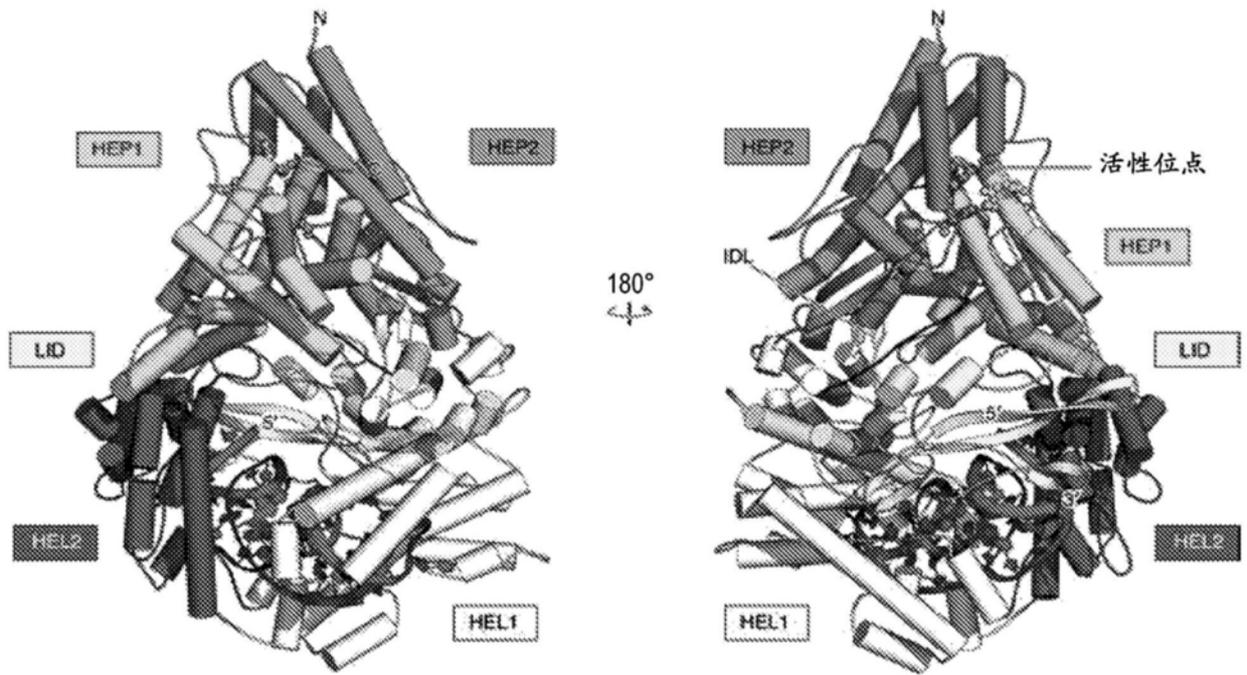


图1C

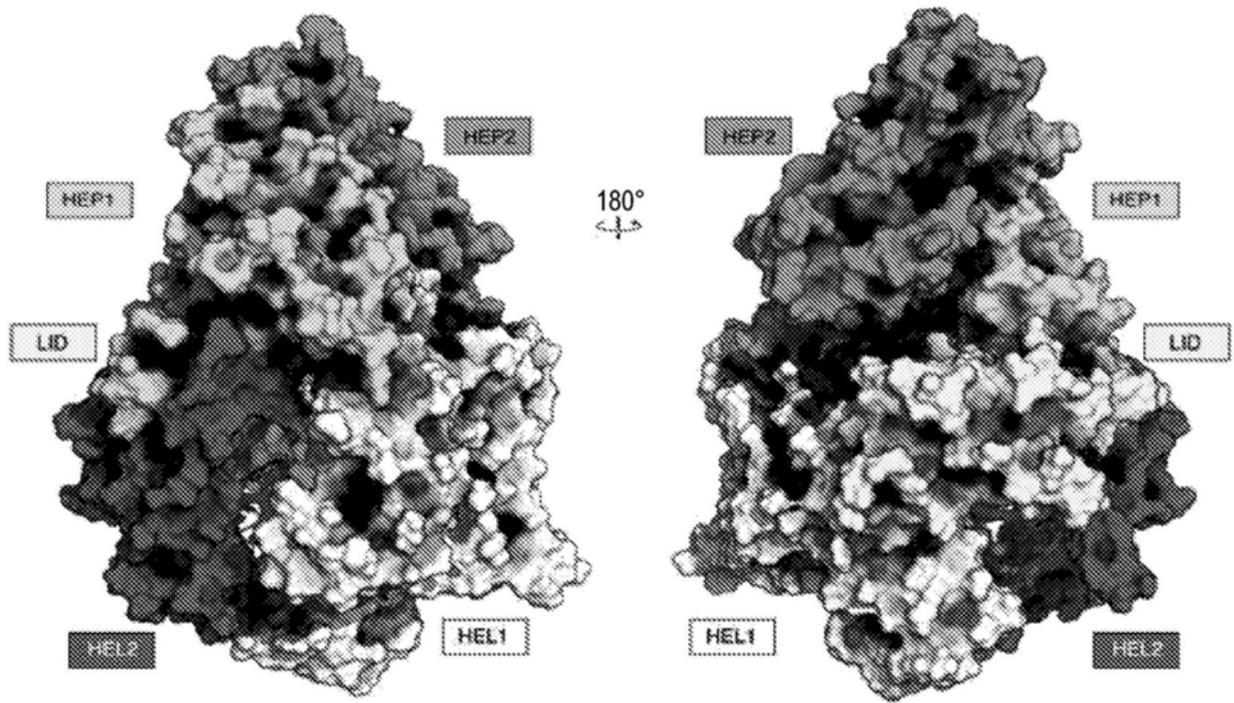


图1D



图2A

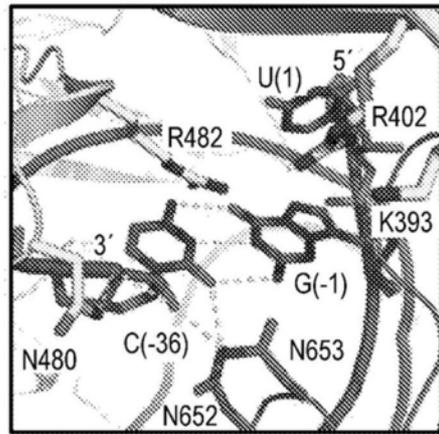


图2B

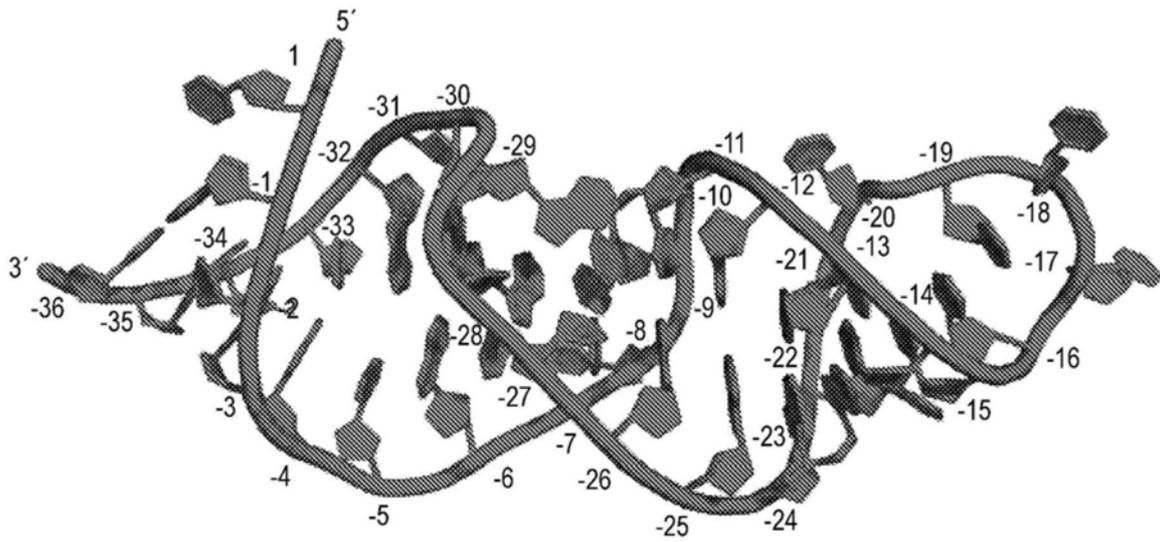


图2C

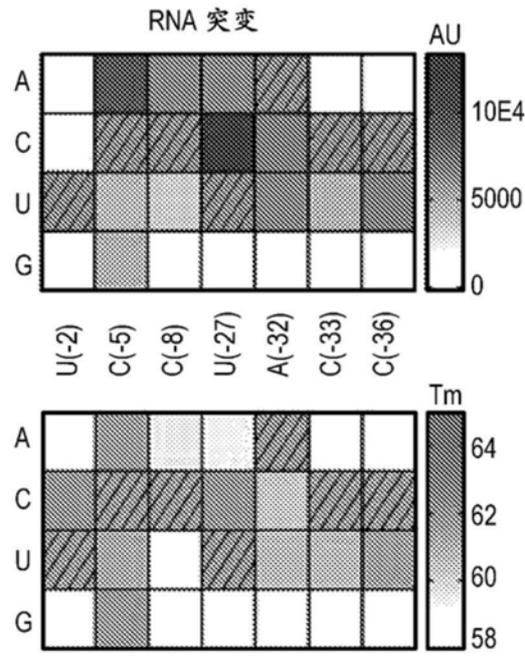


图2D

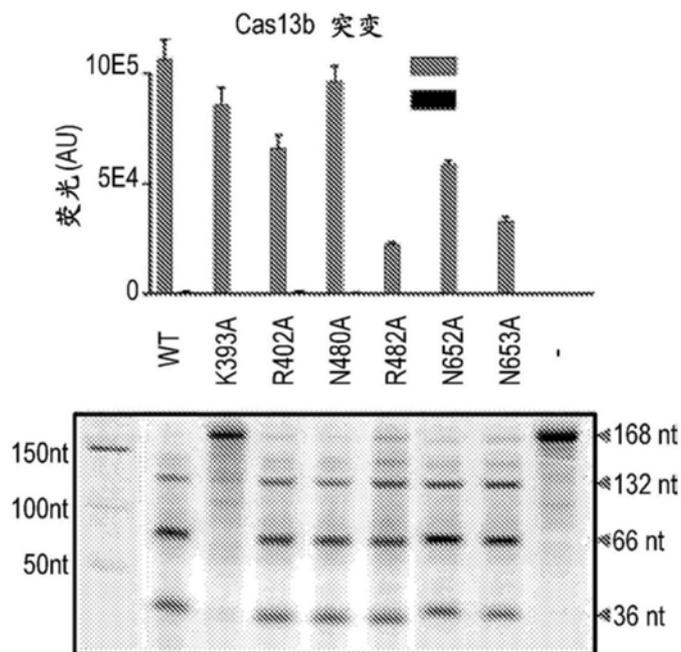


图2E

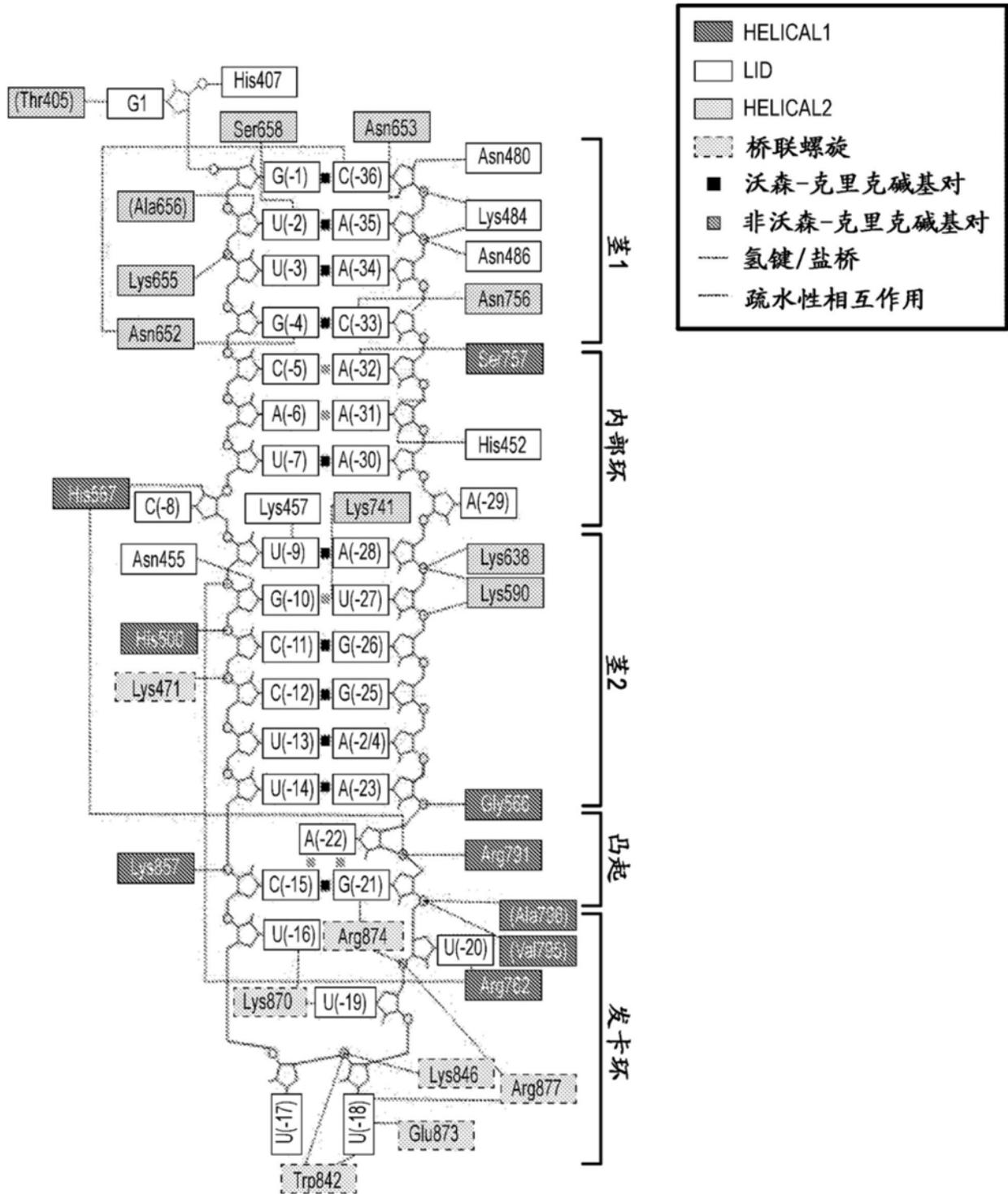


图3

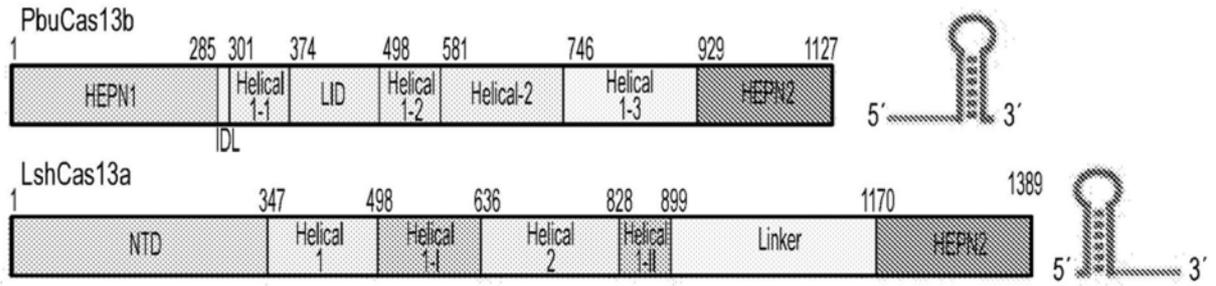


图4A

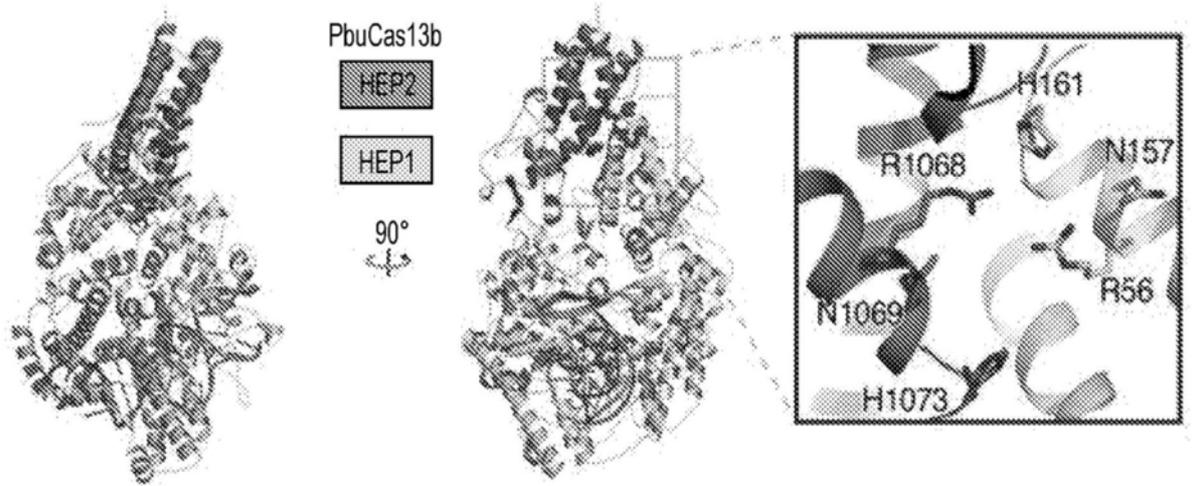


图4B

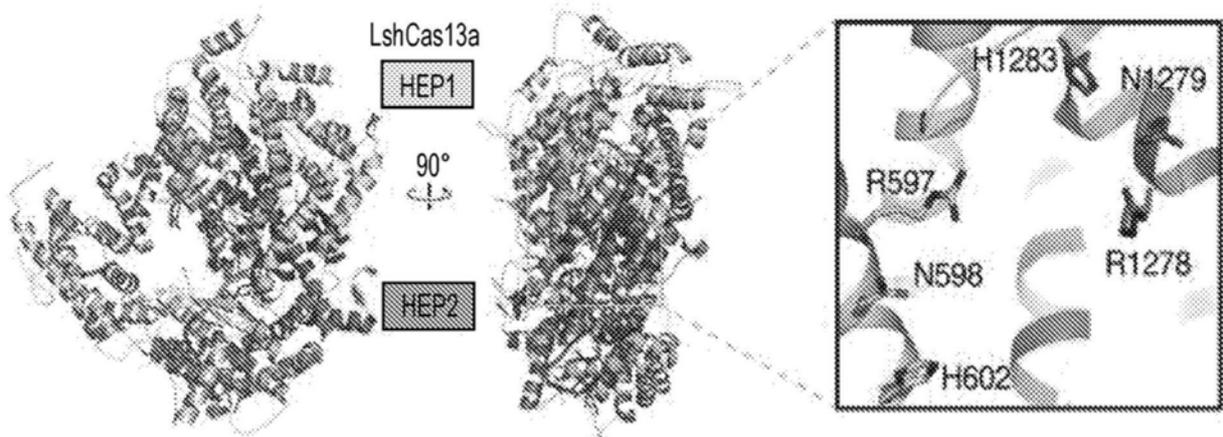


图4C

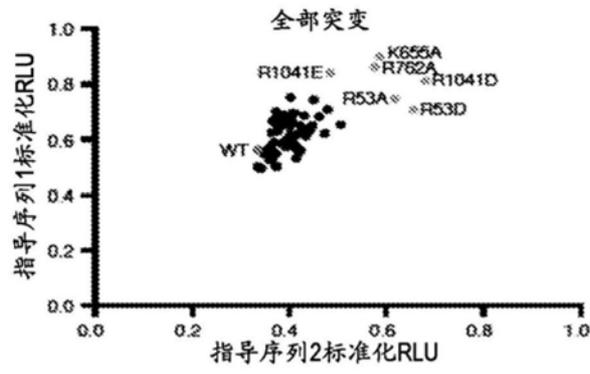


图5A

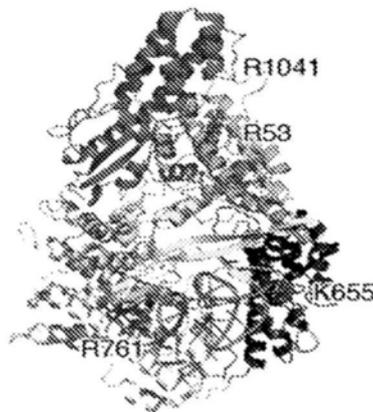


图5B

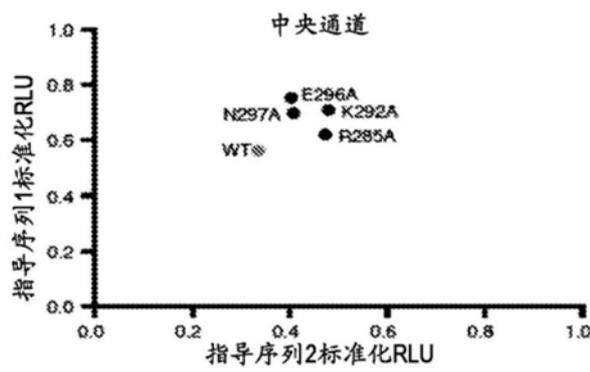


图5C

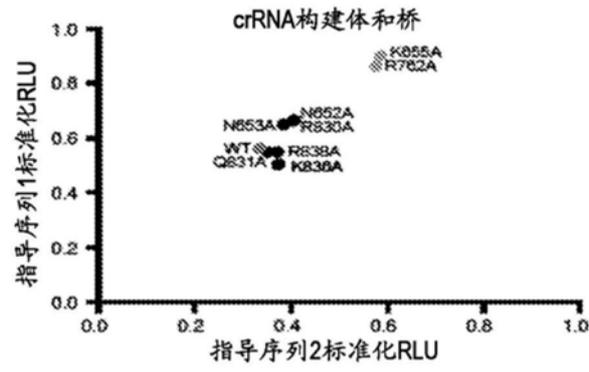


图5D

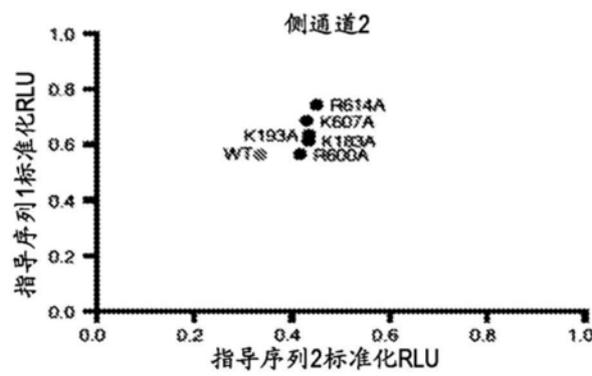


图5E

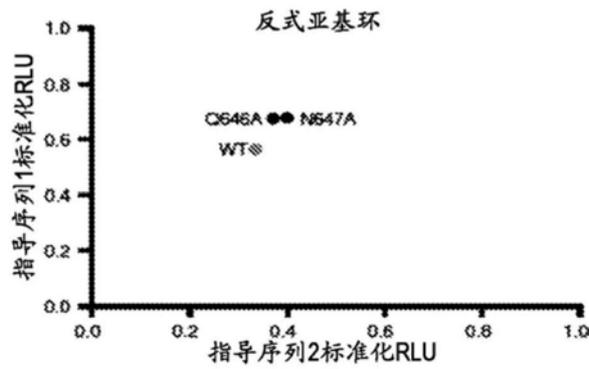


图5F

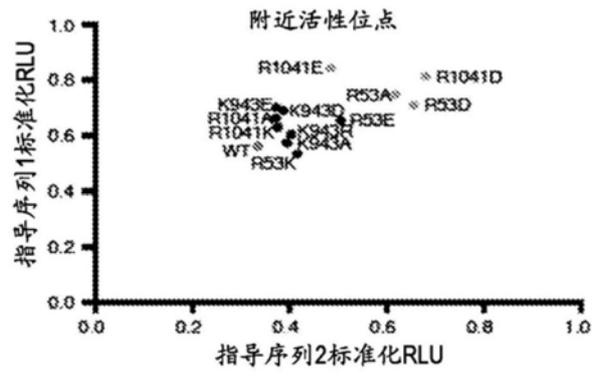


图5G

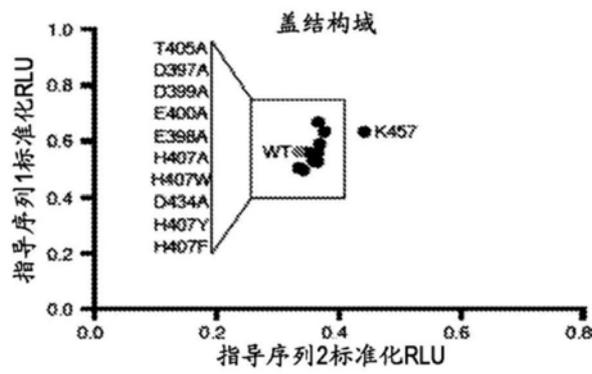


图5H

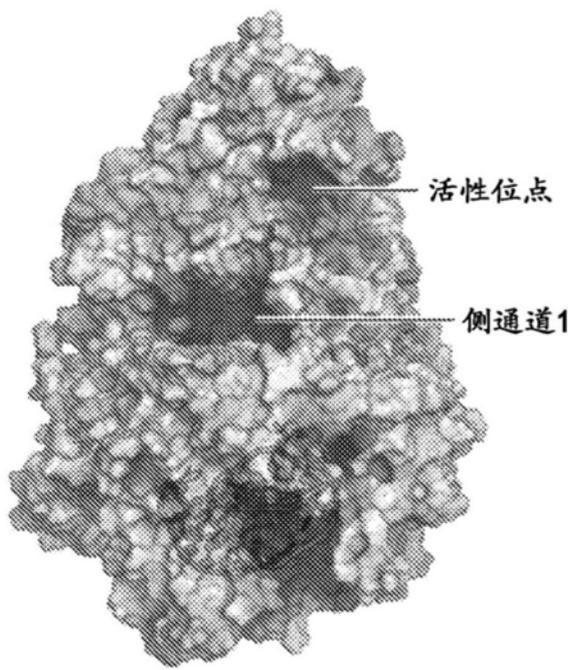


图6A

180°  
↕

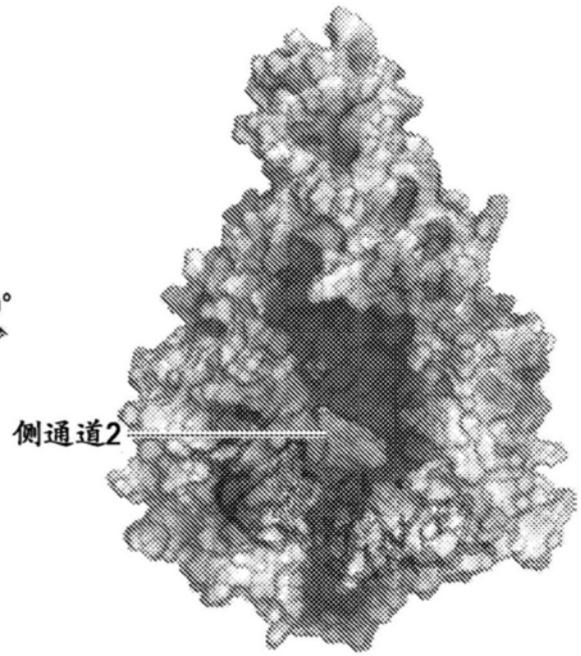


图6B

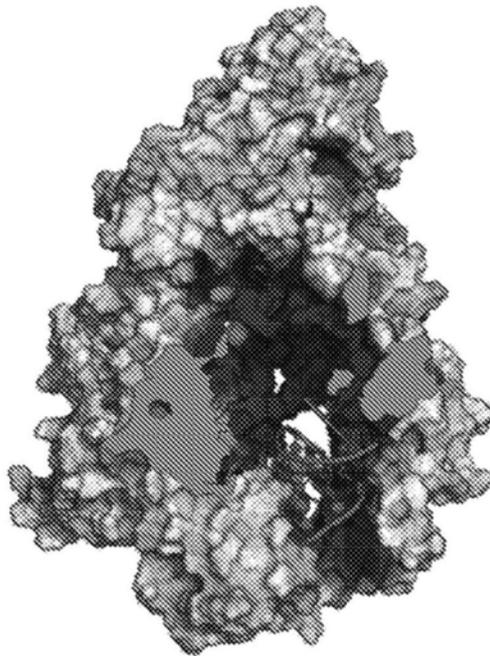


图6C

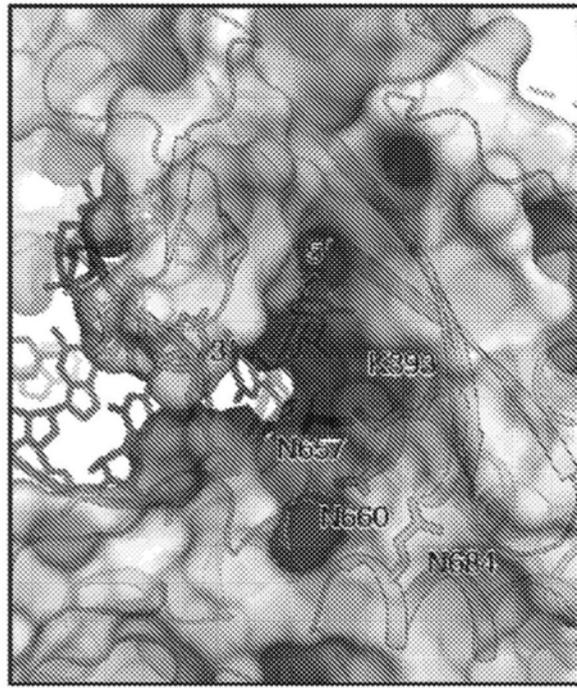


图6D

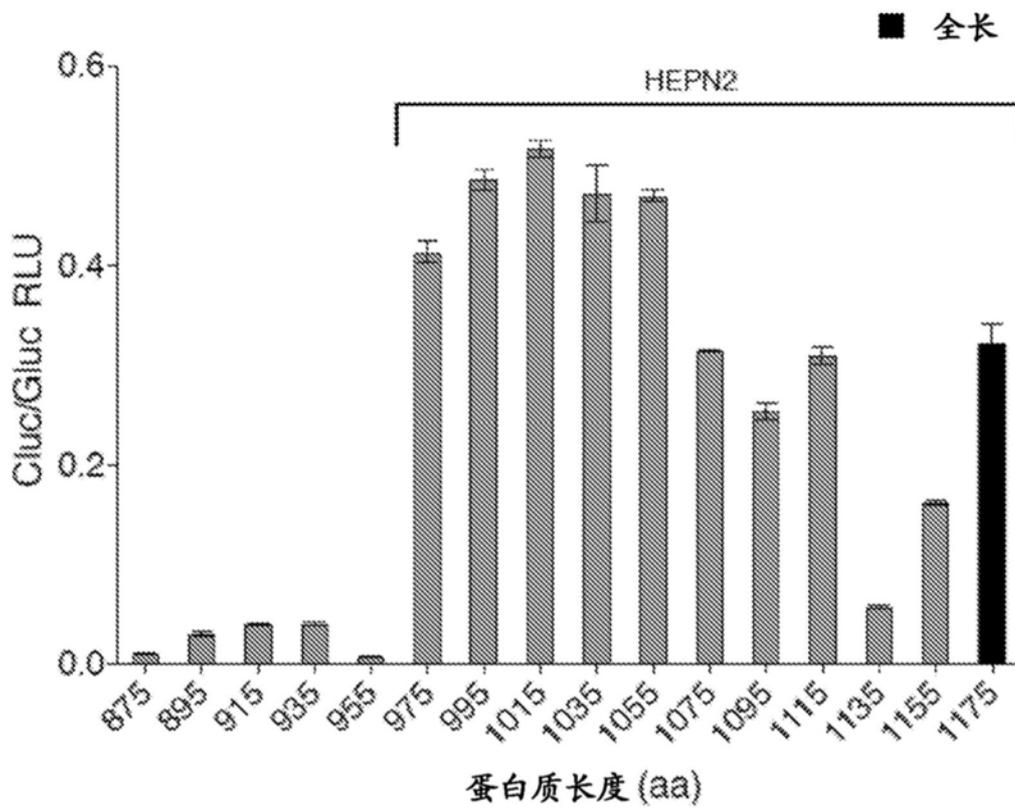


图7



D

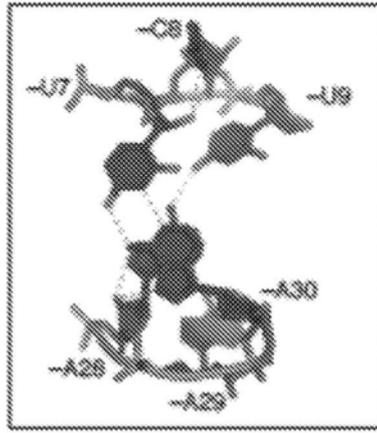


图8D

E

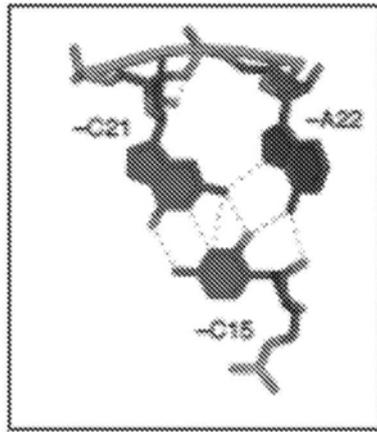


图8E

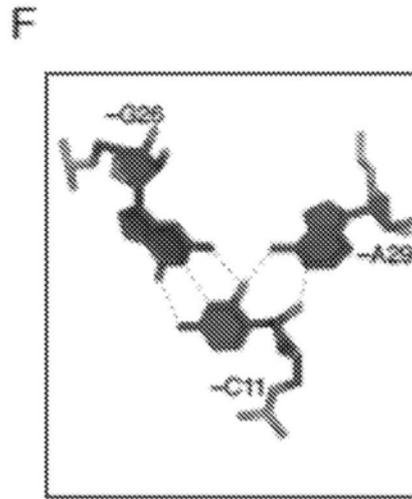


图8F

G

|    |         |      |     |          |        |          |
|----|---------|------|-----|----------|--------|----------|
|    | *****   | ** * | *   | *****    | *****  | ***** *  |
| Pb | GUUGCAU | UG   | CUU | UUUUUGAA | AGGUA  | AAAACAAC |
| Pi | GUUGCAU | UG   | CUG | UGUUUGCA | AGGUA  | AAAACAAC |
| Ra | GUUGGGA | UG   | UCU | ACUUUGA  | AGGGUA | UUCACAAC |
| Pa | GUUGUAU | UG   | CUU | UGUUUGA  | AAGGUA | AAAACAAC |
| Ps | GUUGUGU | UA   | CUC | UUUUUGAG | AGGUA  | AAAACAG  |
| Pg | GUUGGAU | UA   | CCU | UAUUUGA  | AGGGUA | CACAAC   |
|    | -1      | -10  |     | -20      |        | -30      |

图8G

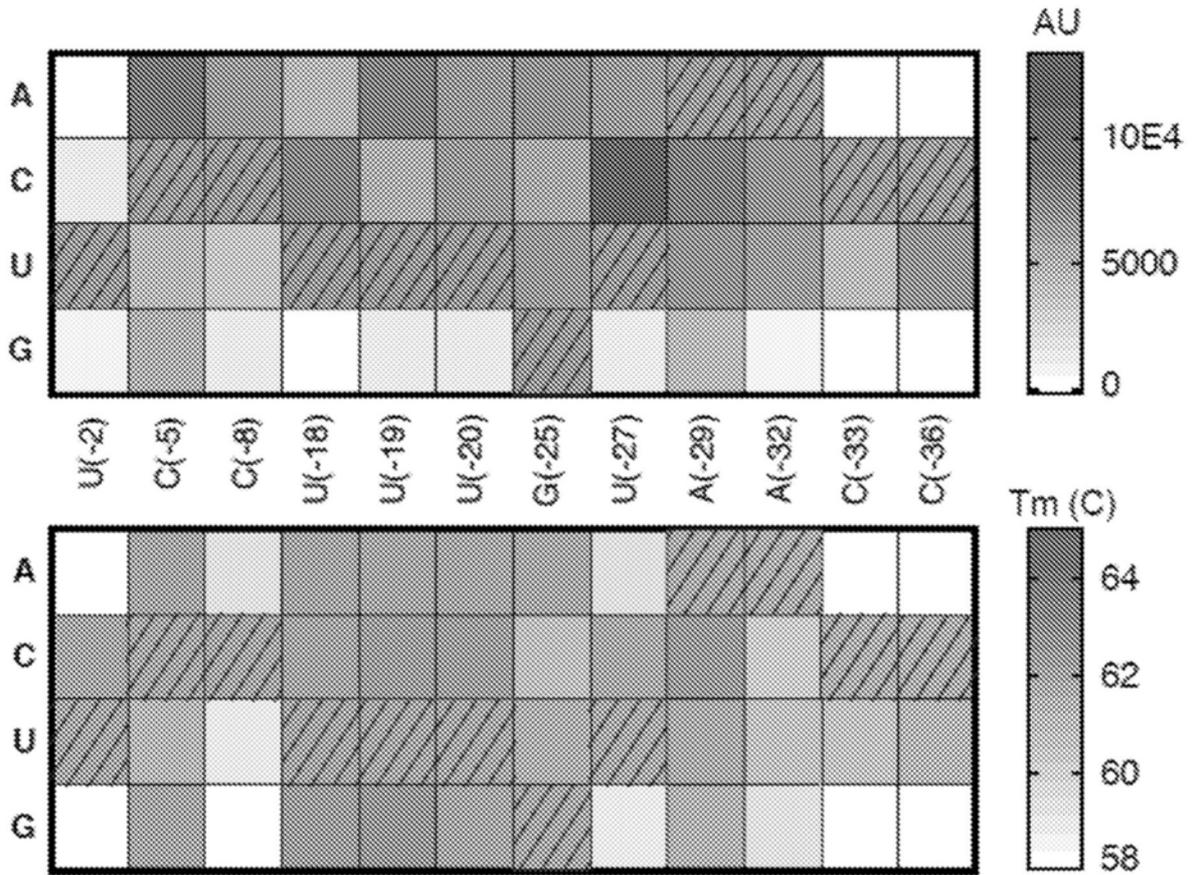


图9

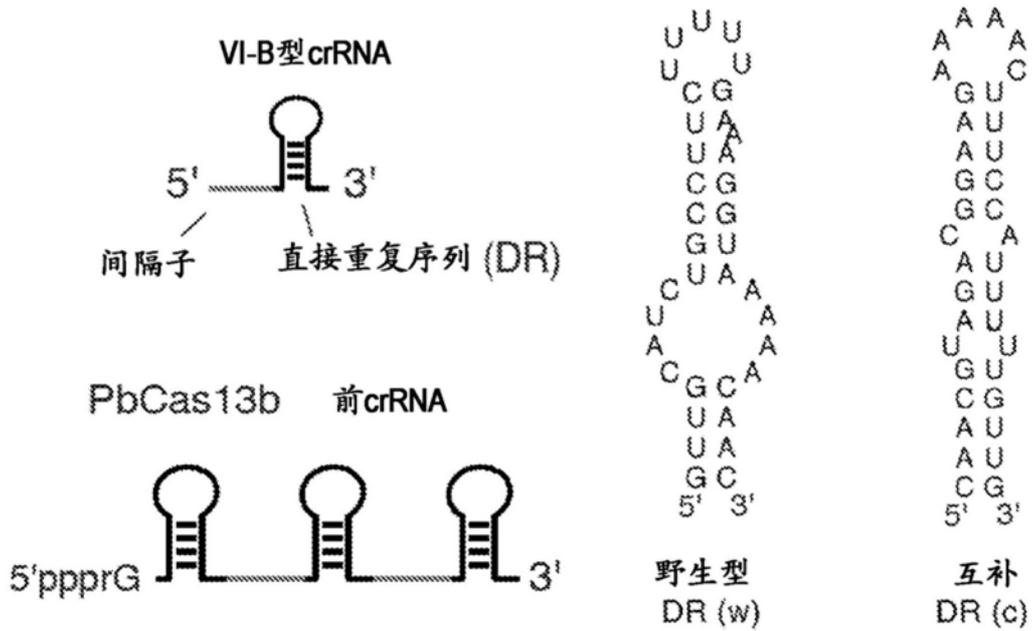


图10A

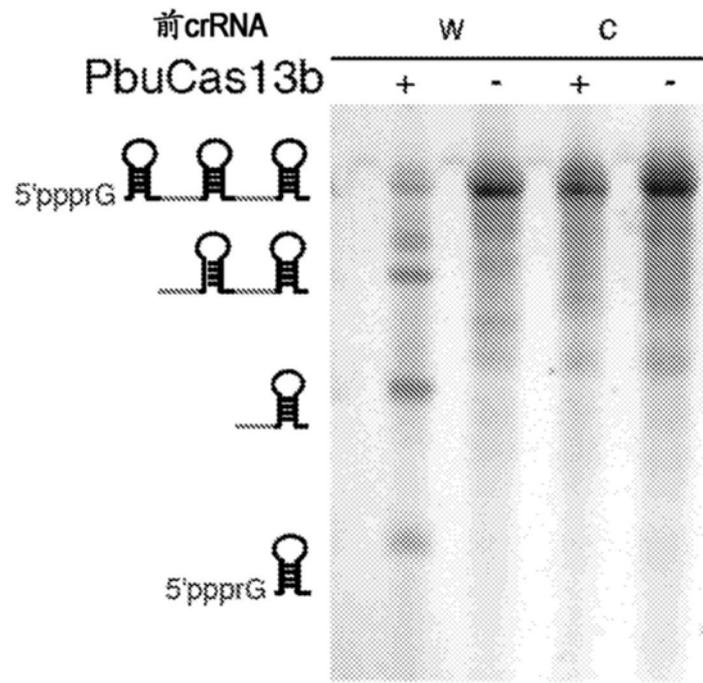


图10B

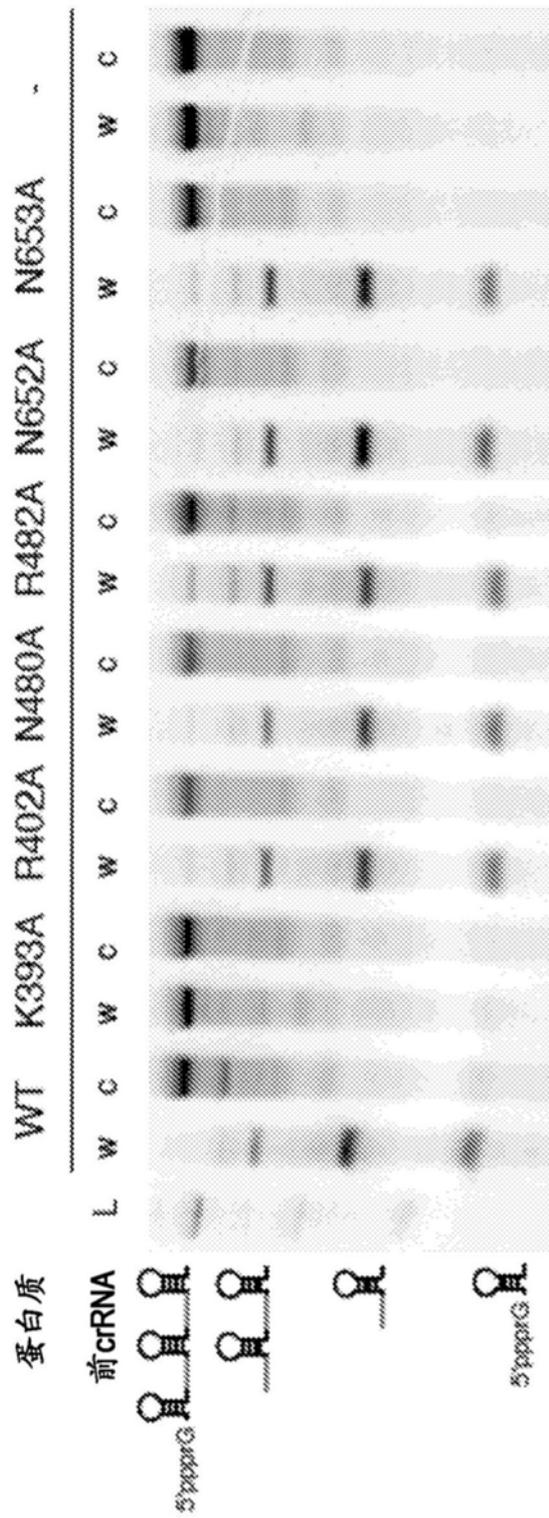


图10C

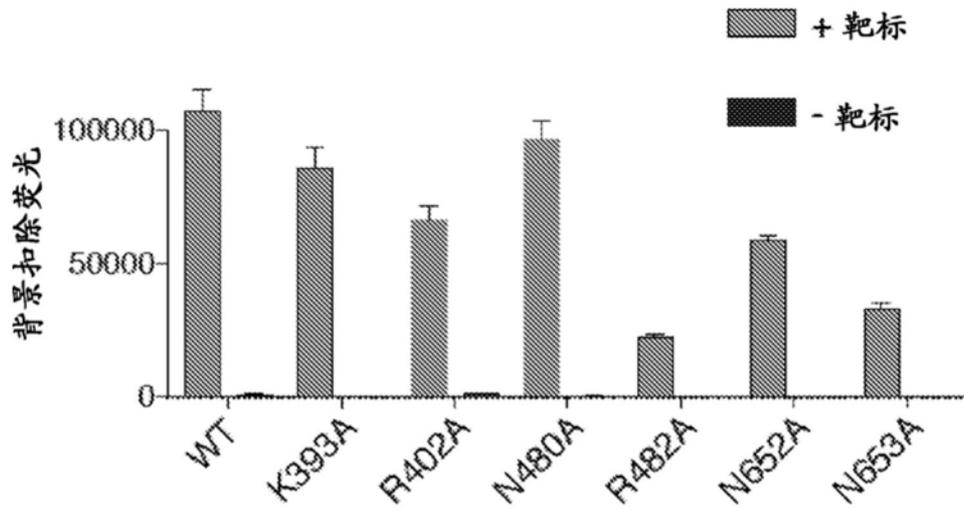


图10D

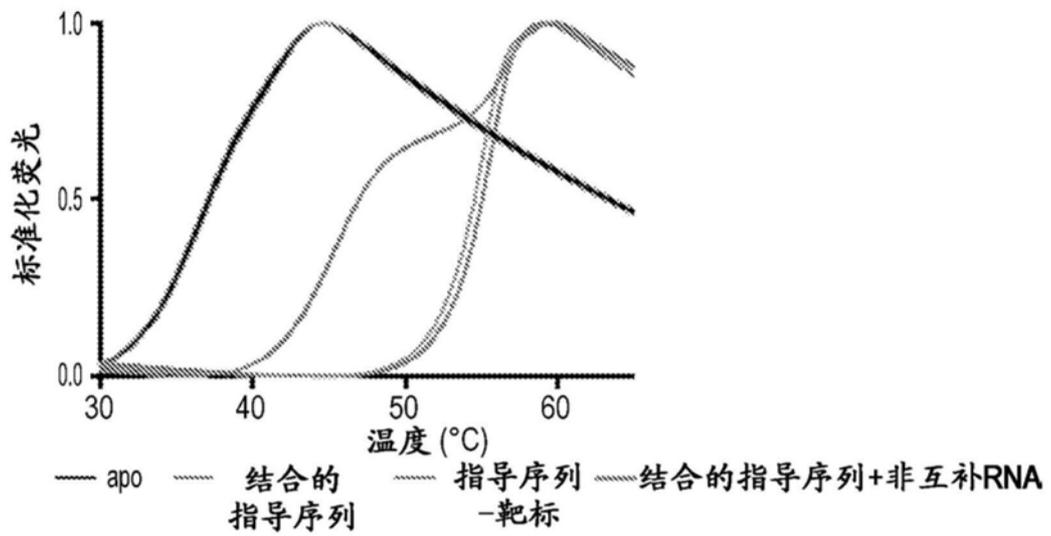


图11A

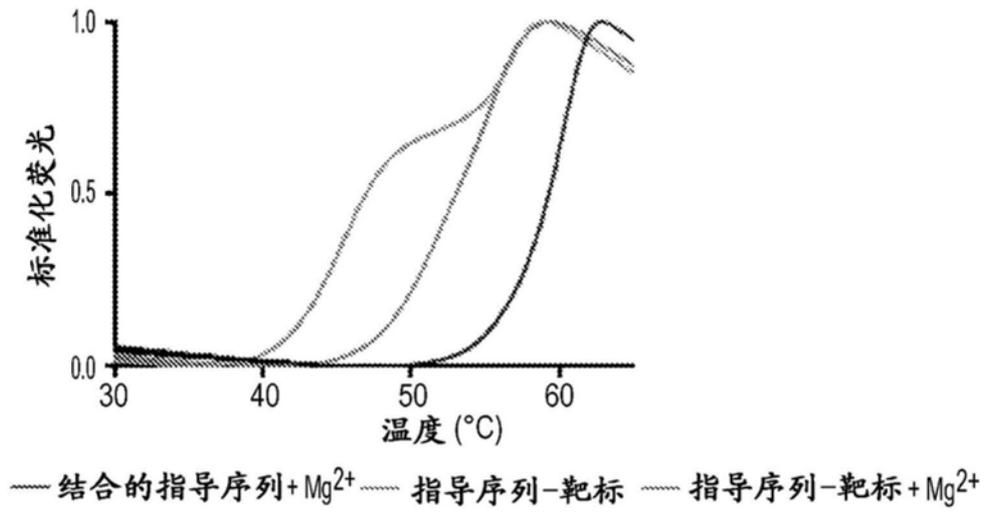


图11B

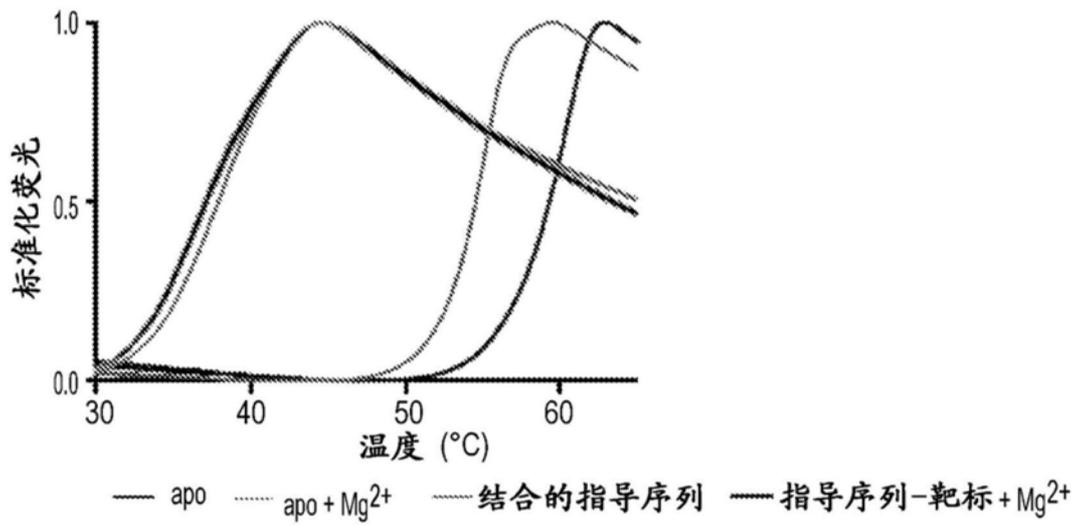


图11C

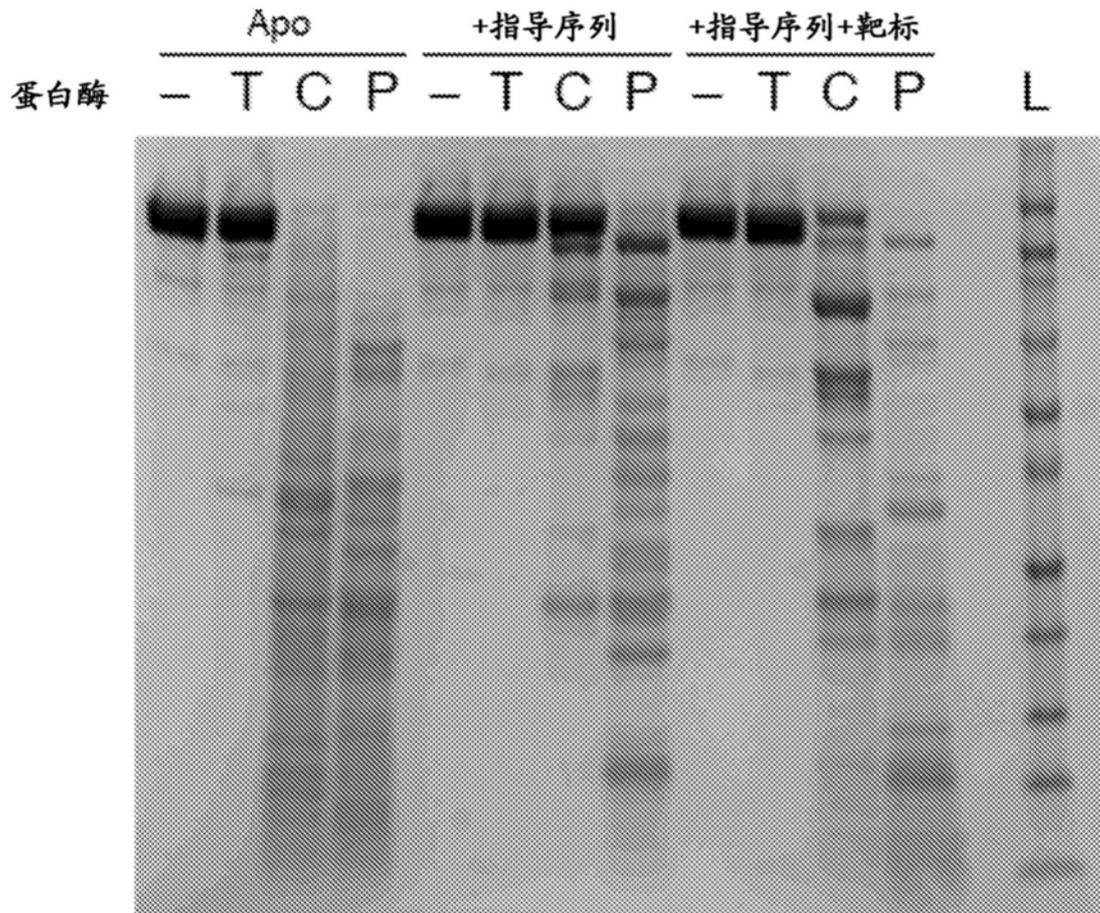


图12



图13A

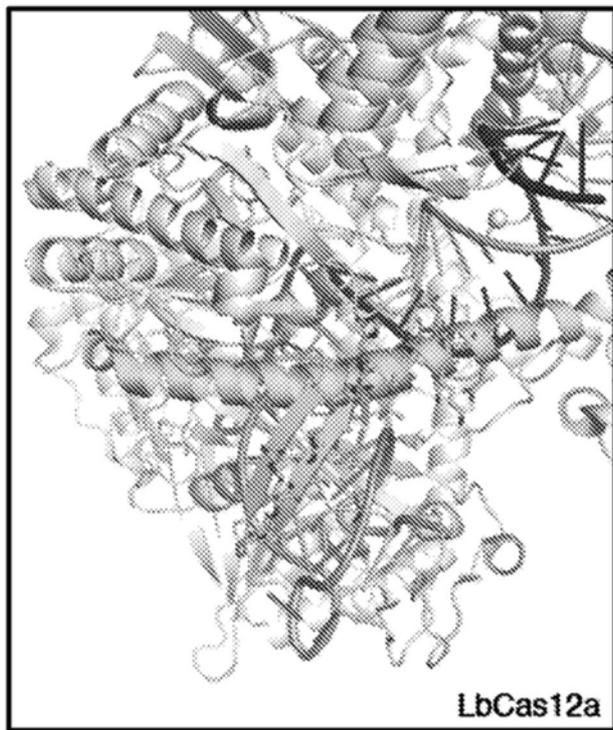


图13B



图13C

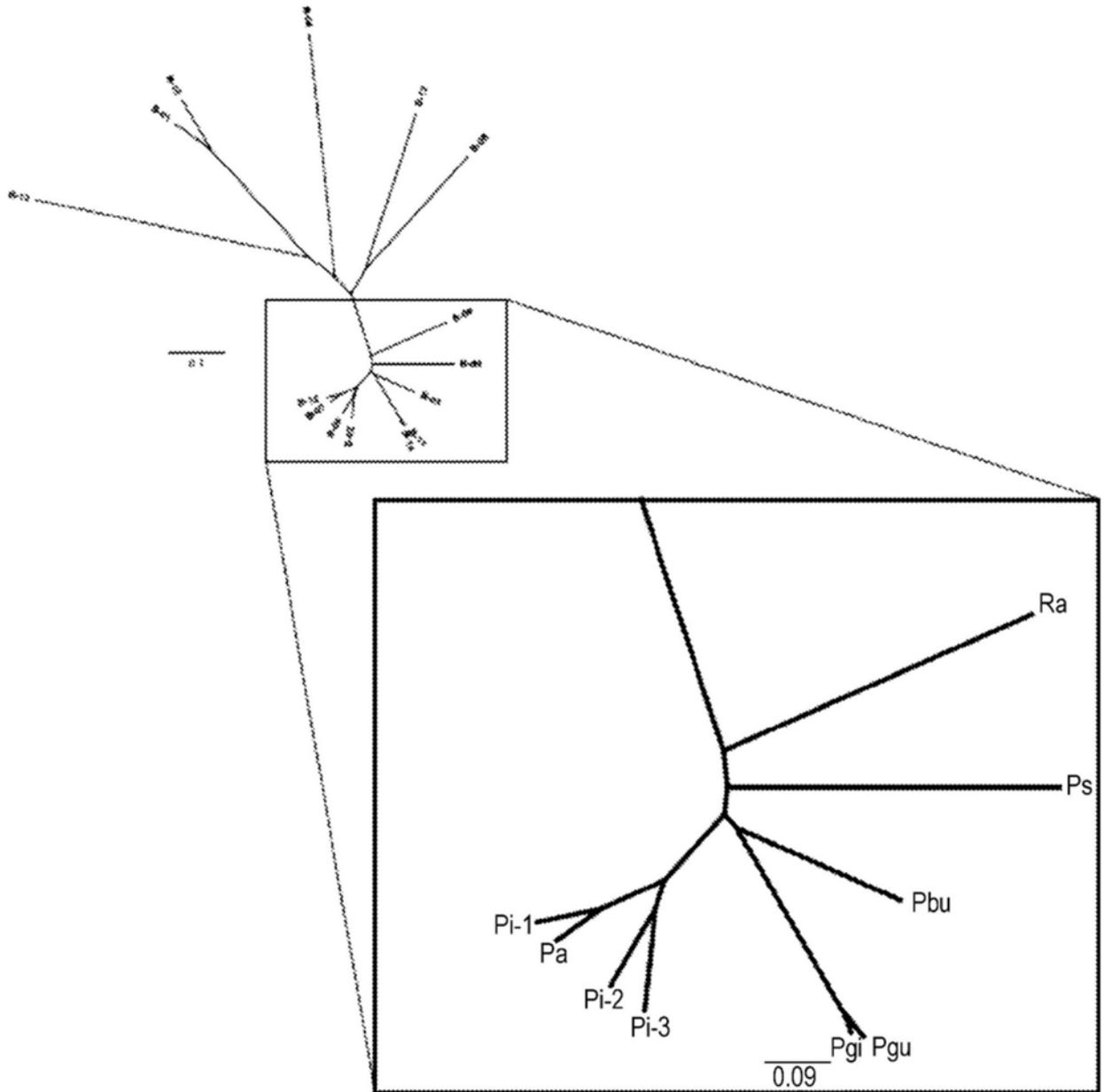


图14

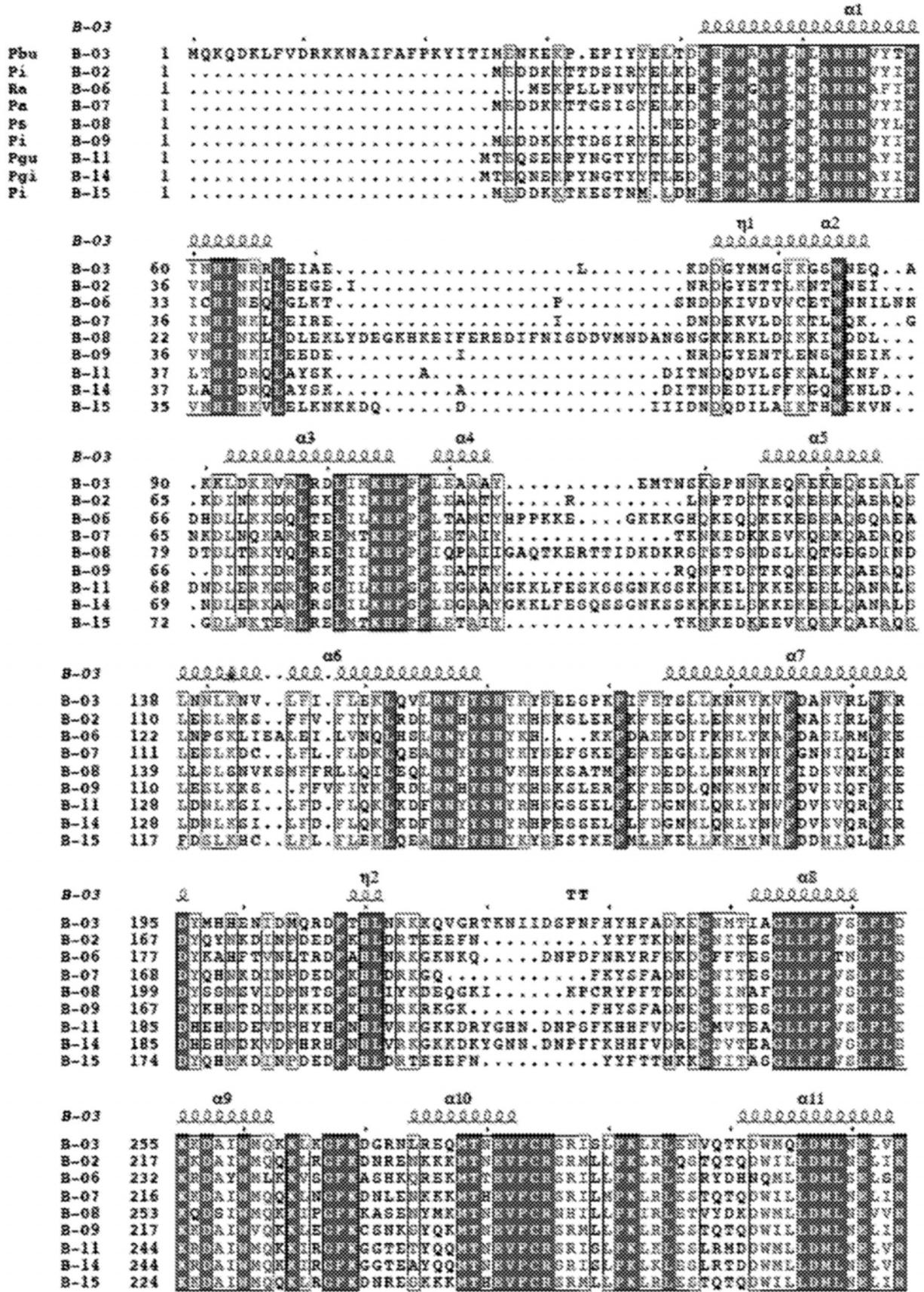


图15

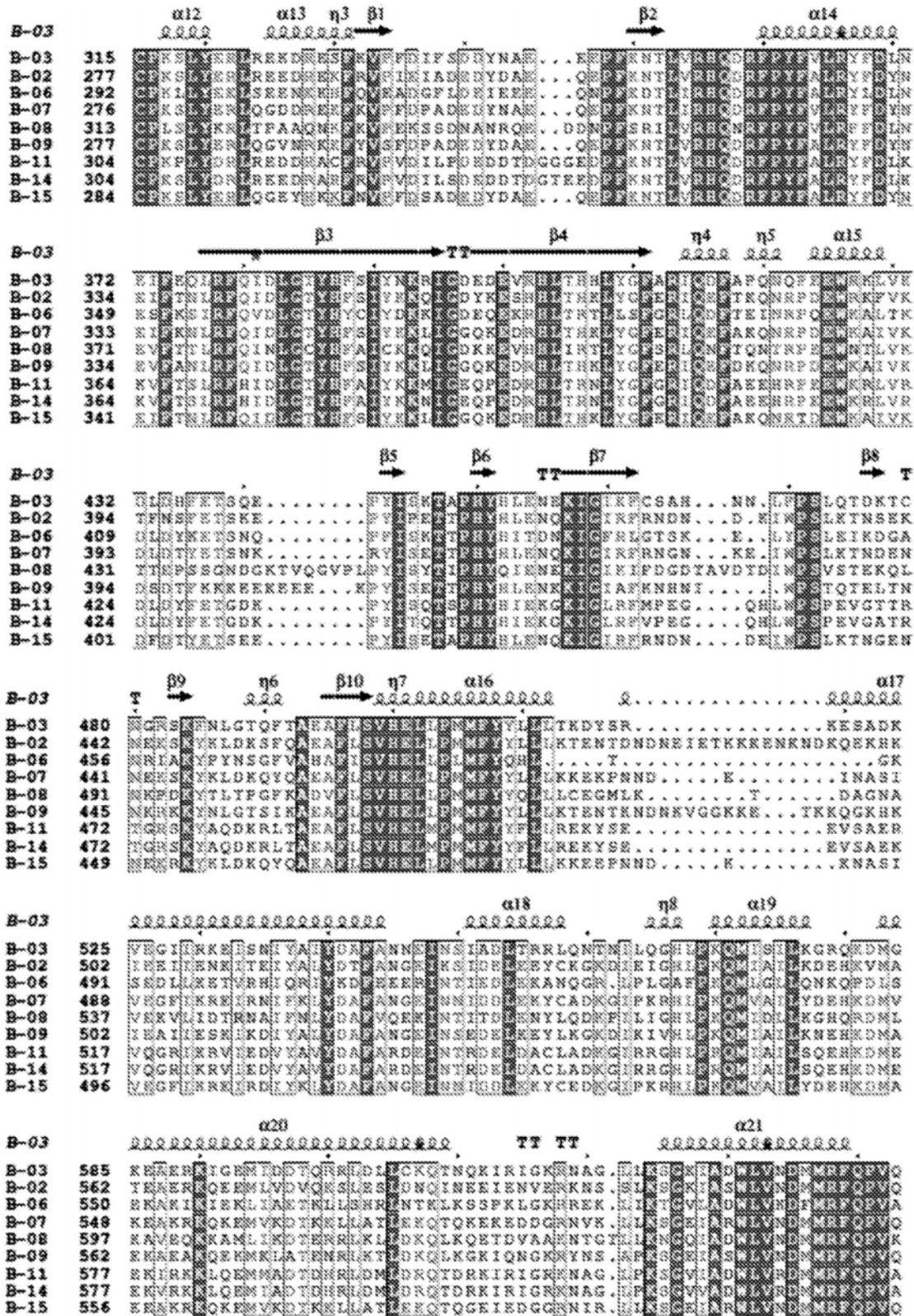


图15(续)

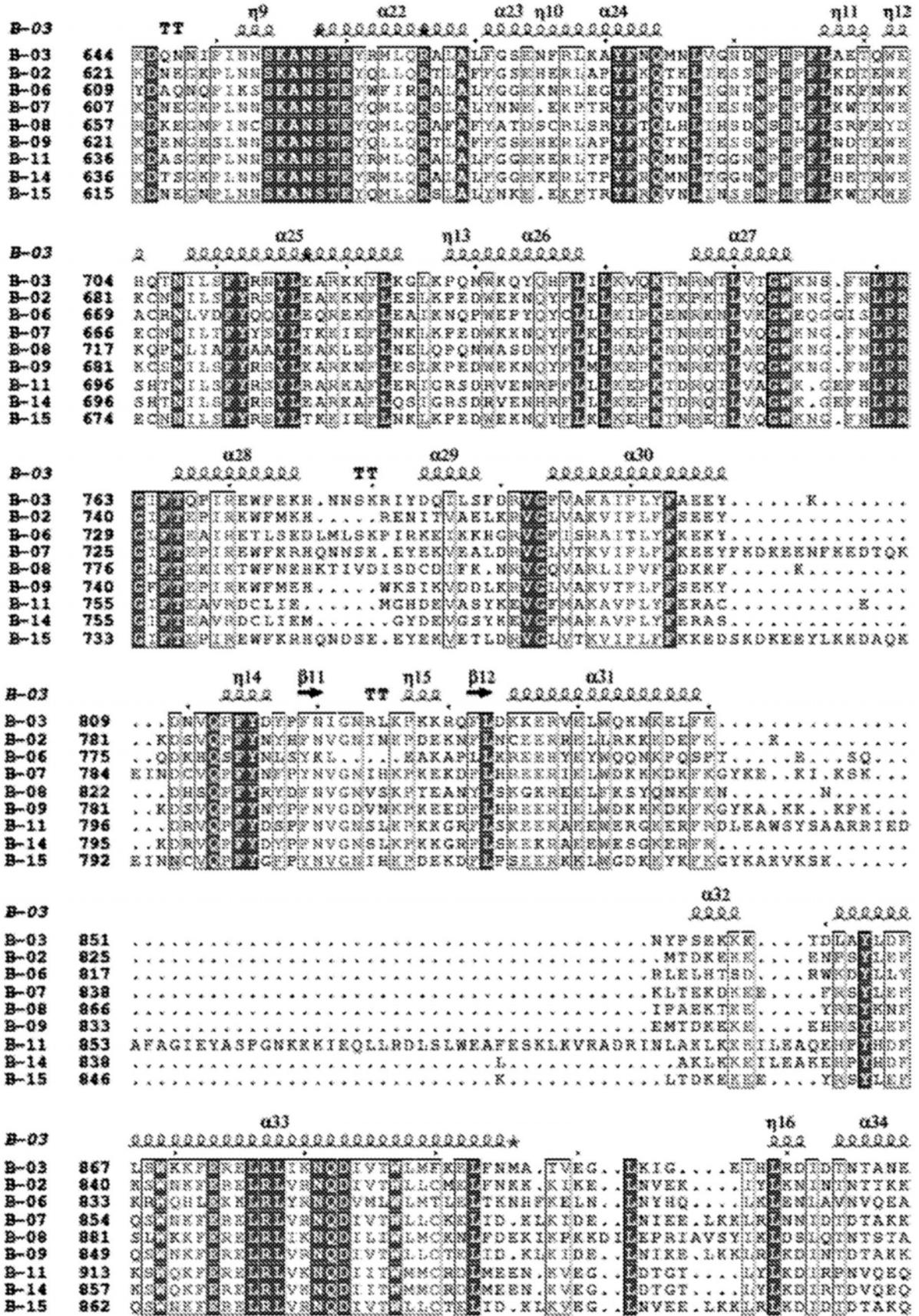


图15(续)

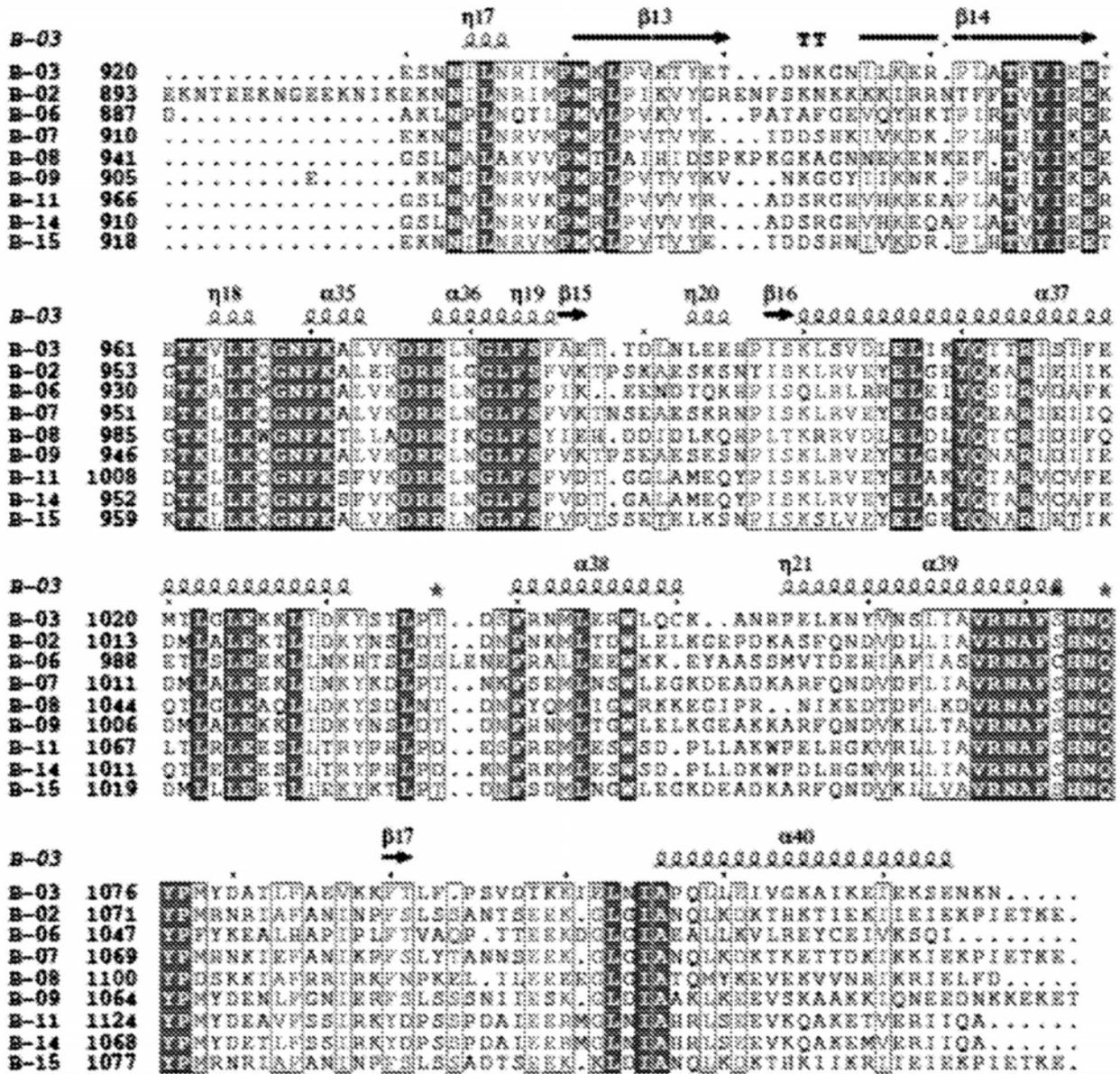


图15(续)

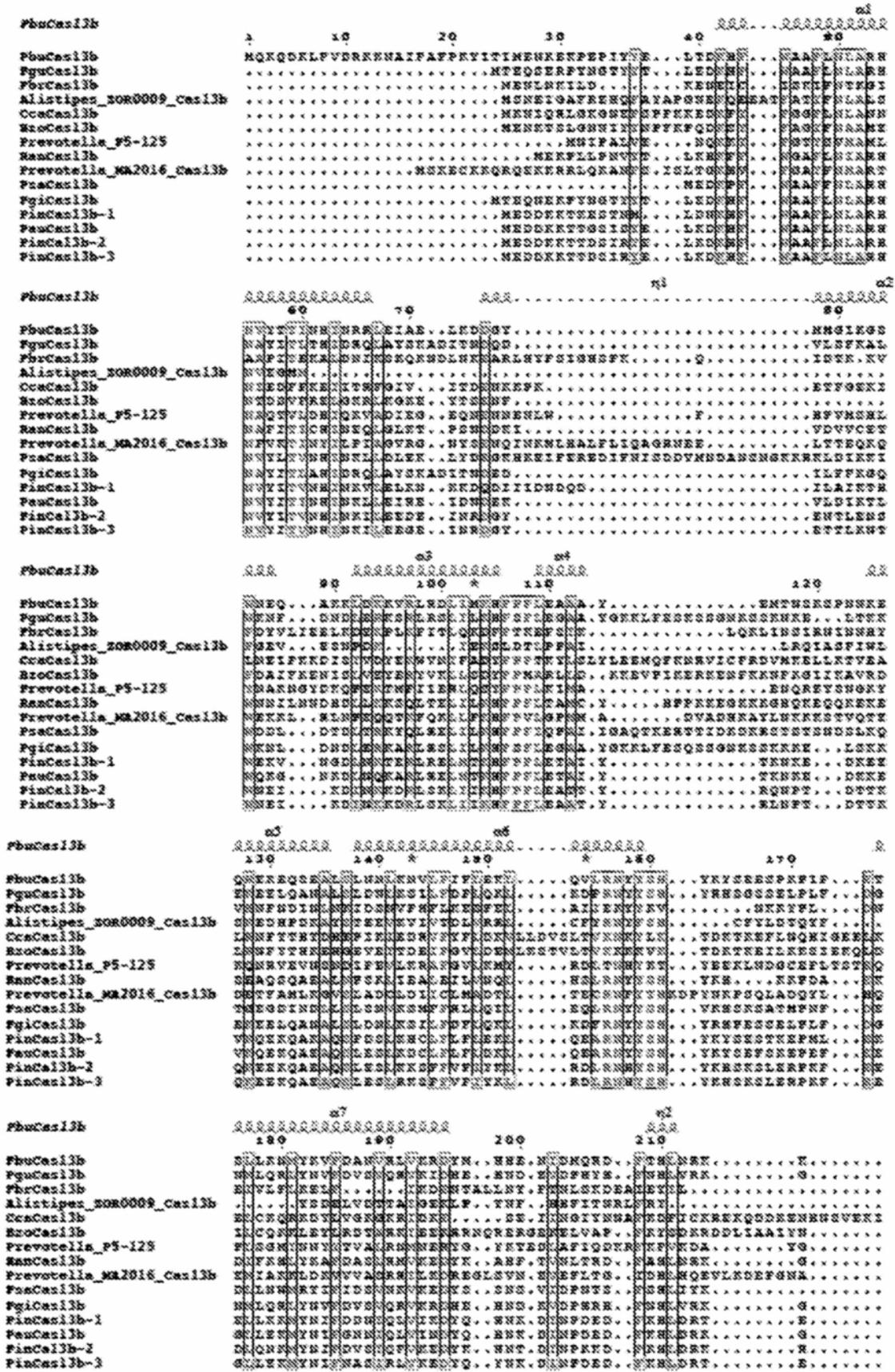


图16

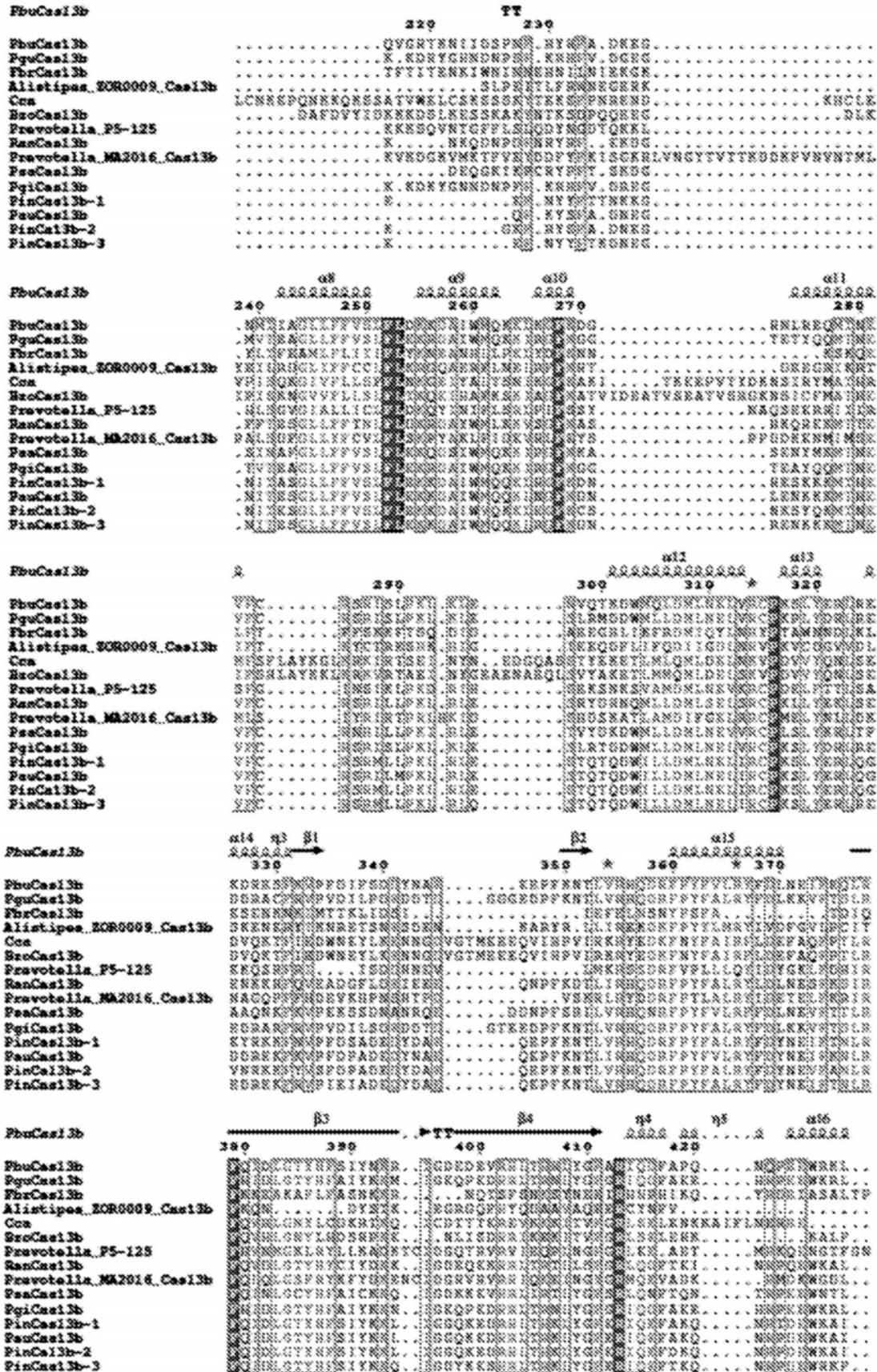


图16(续)

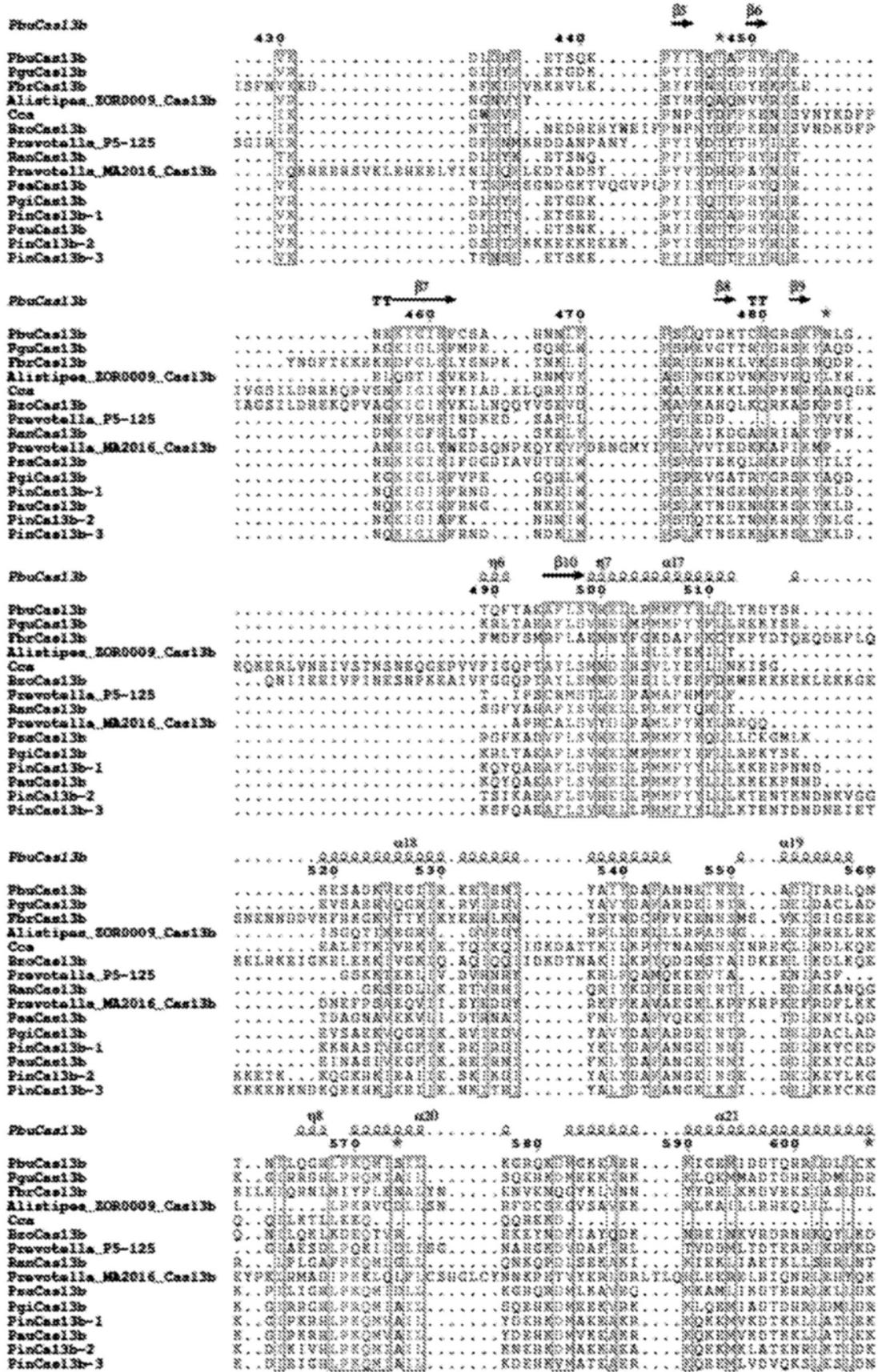


图16(续)

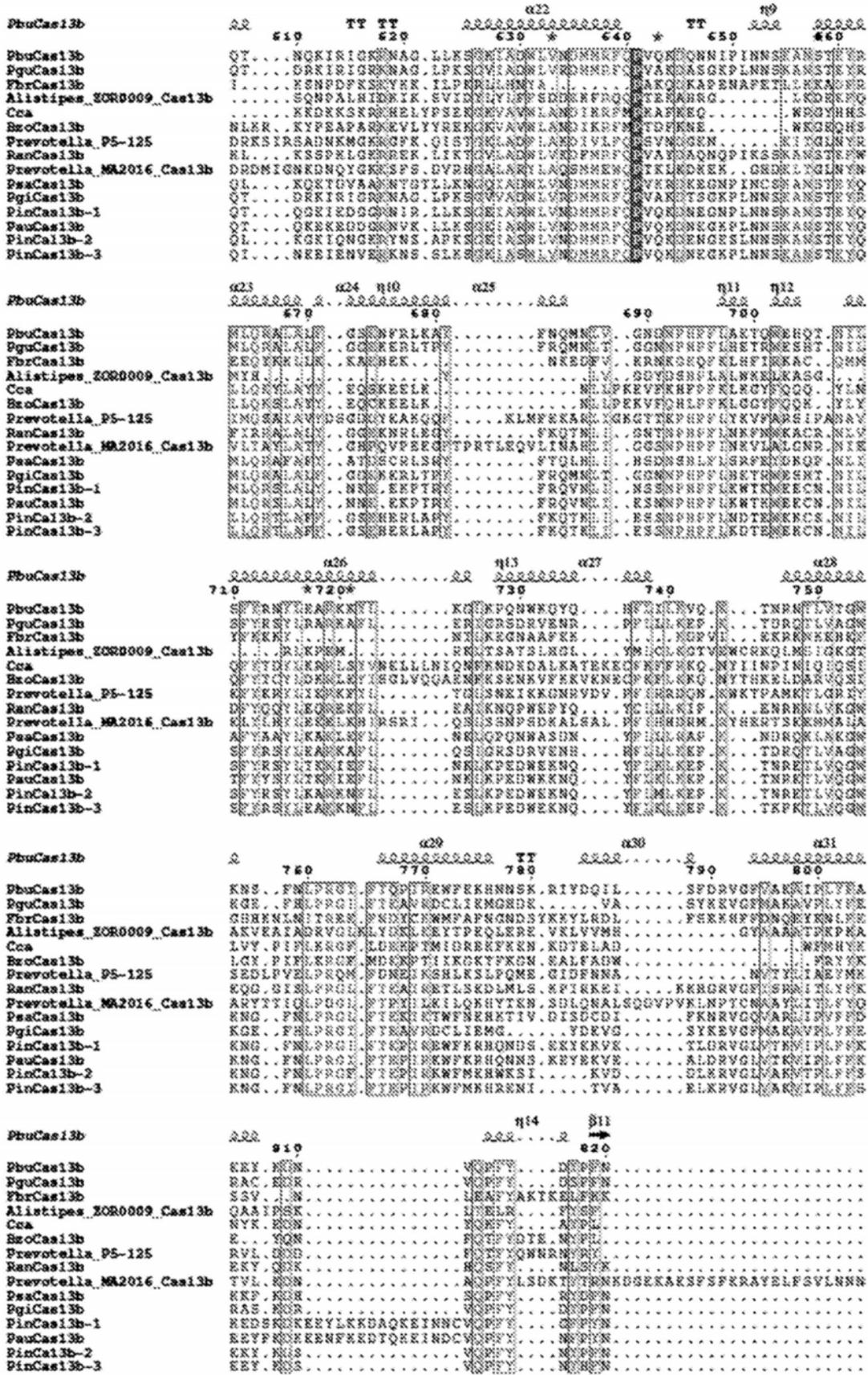


图16(续)

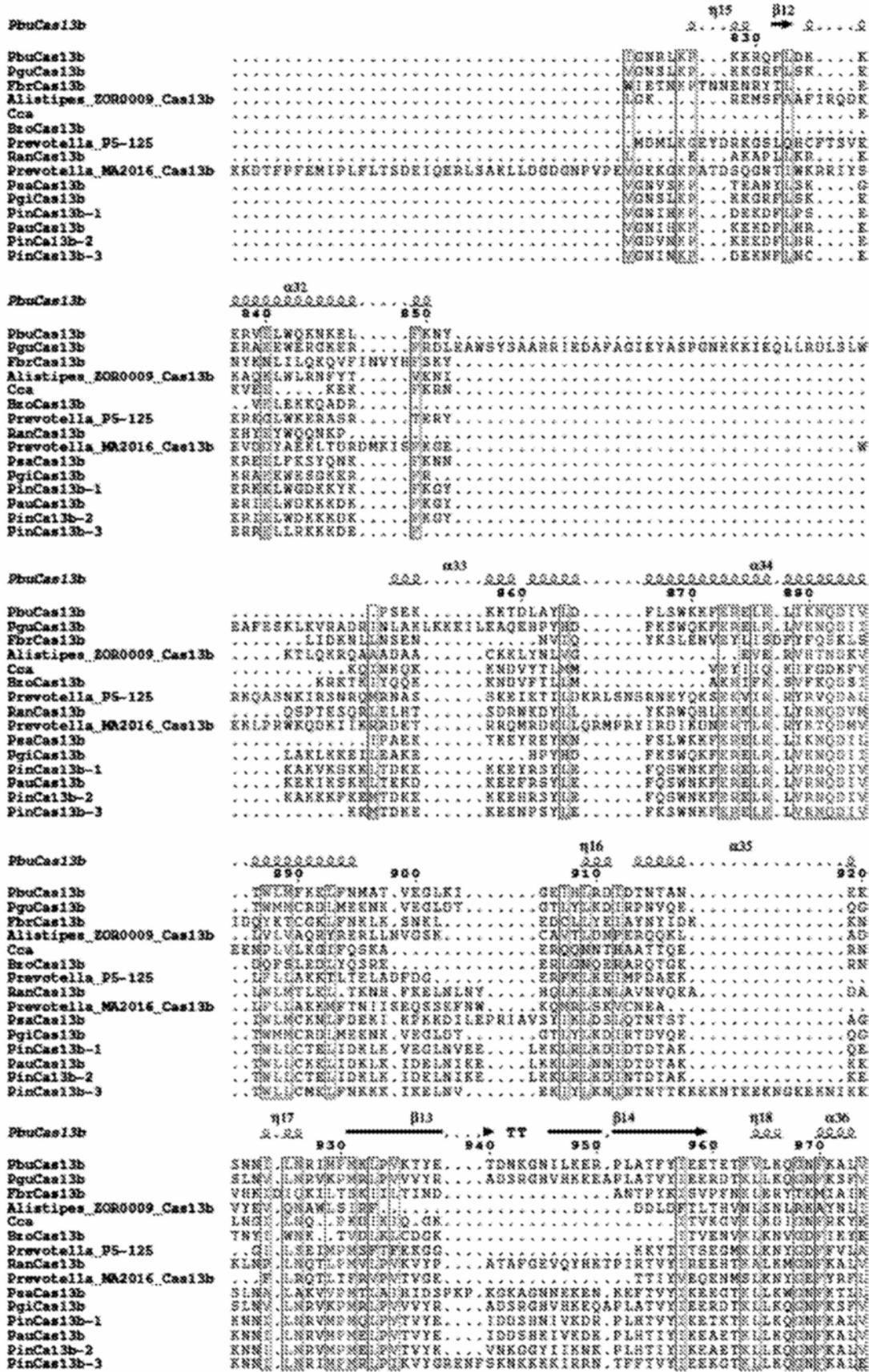


图16(续)

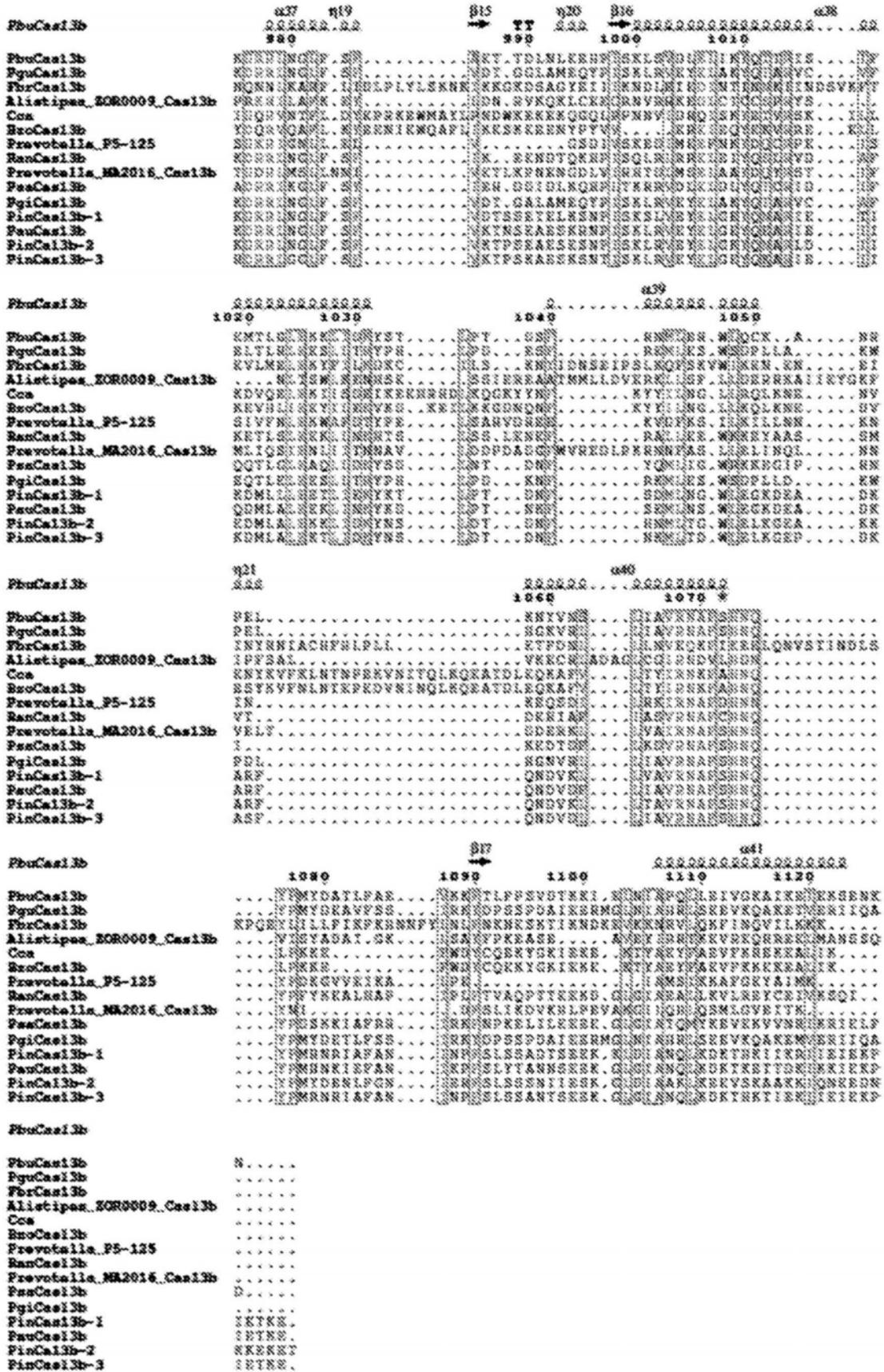


图16(续)

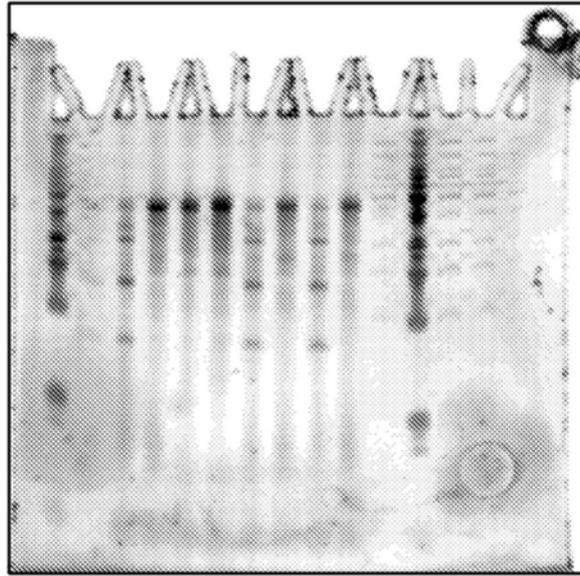


图17A

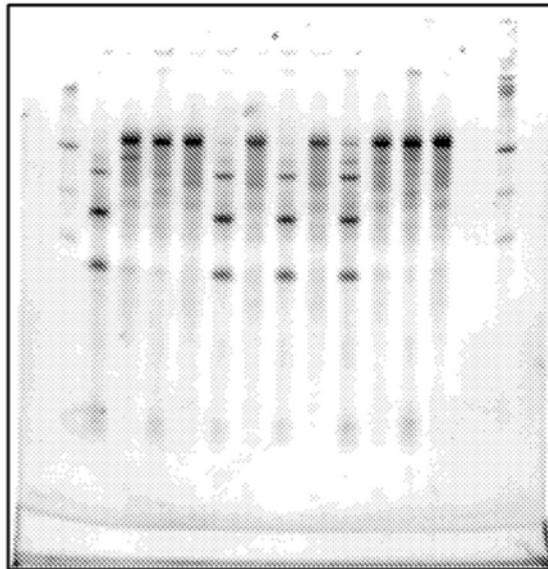


图17B

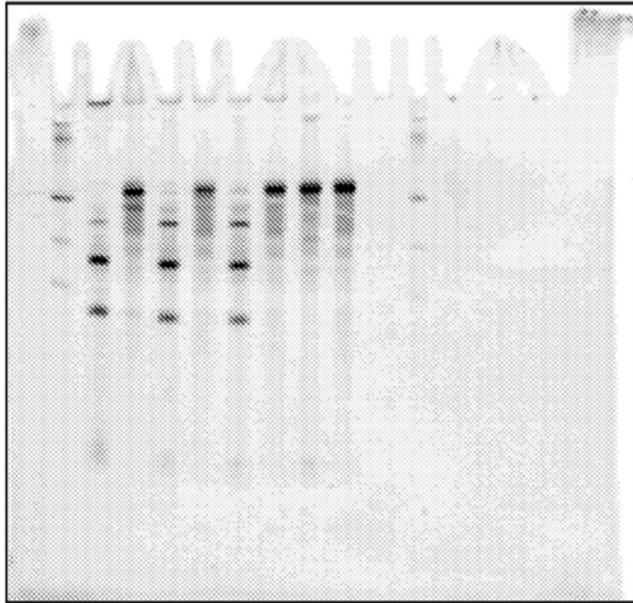


图17C

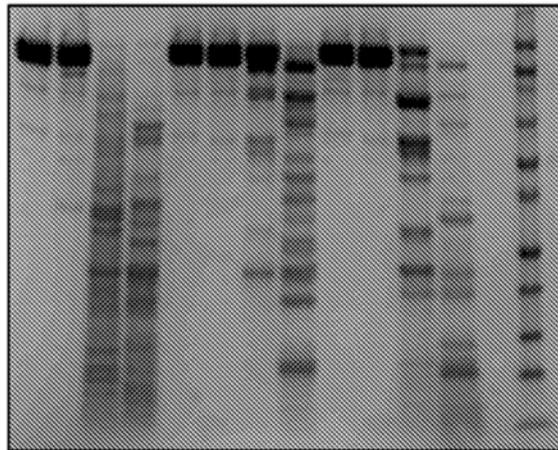


图17D

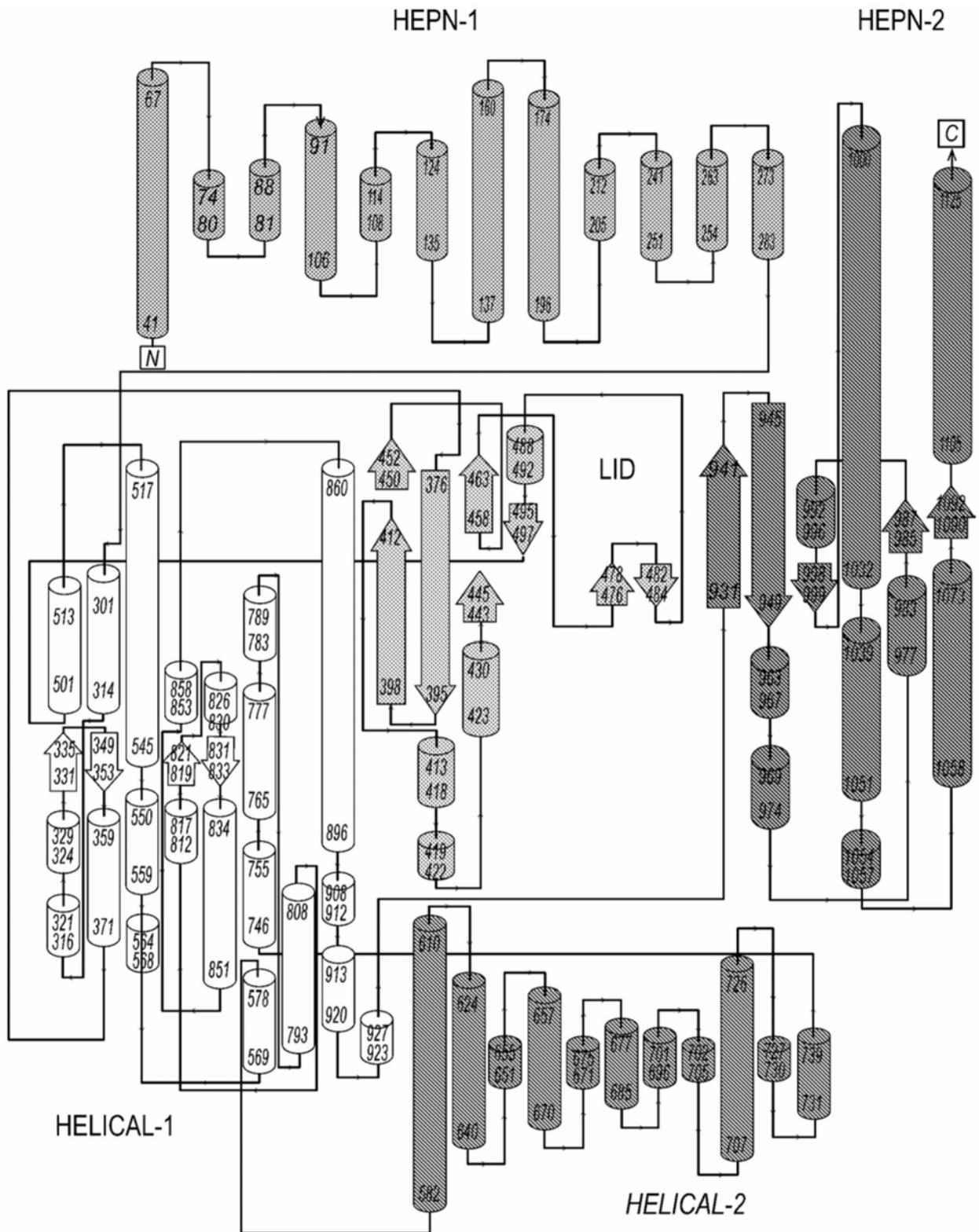


图18

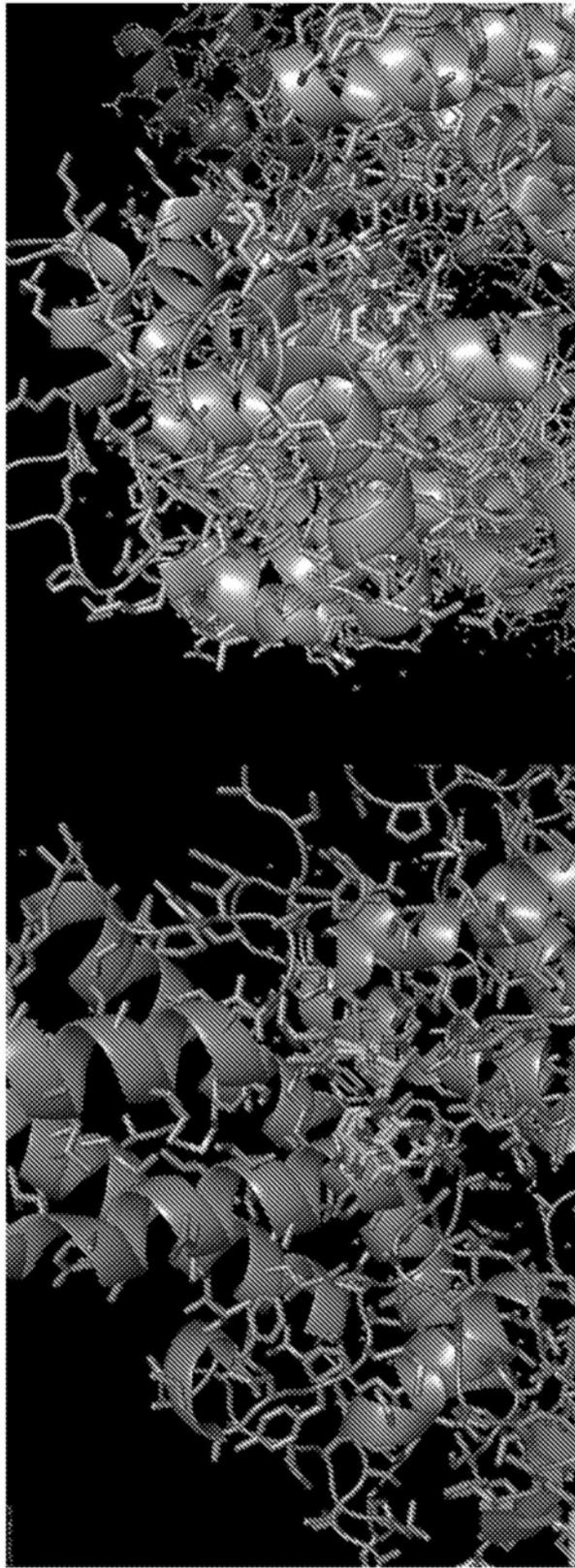


图19

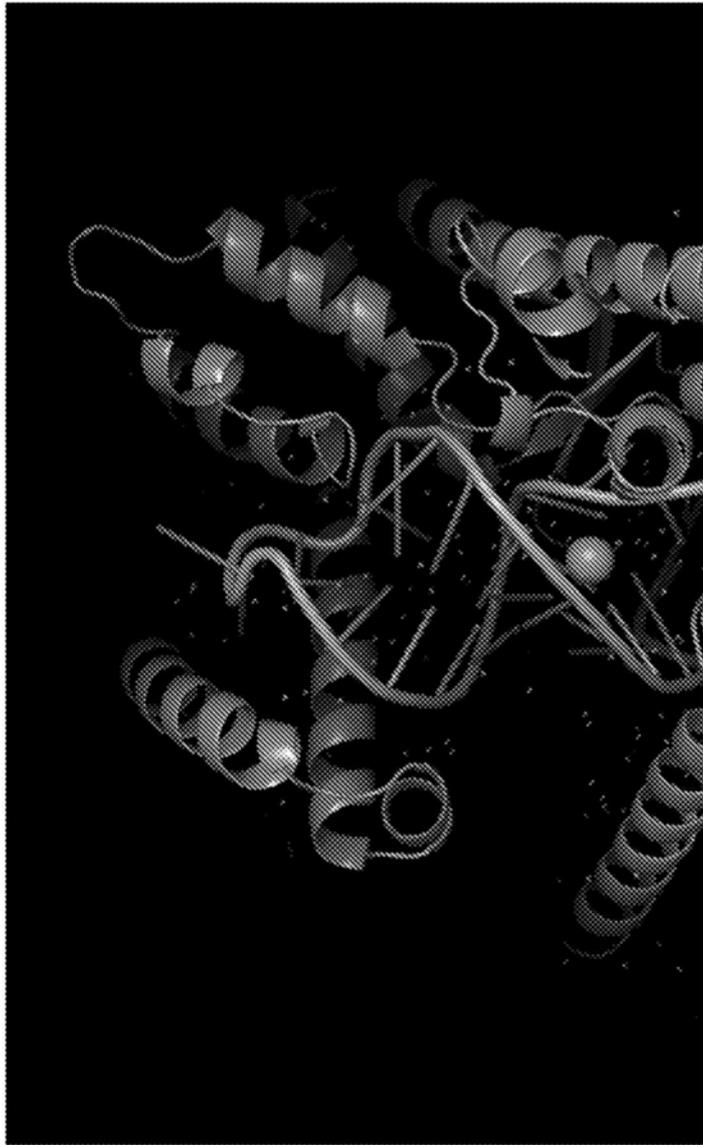


图20

Cast3b  
催化结构域呈洋红色和青色  
红色虚线是靶标RNA  
黑色虚线是间隔子RNA

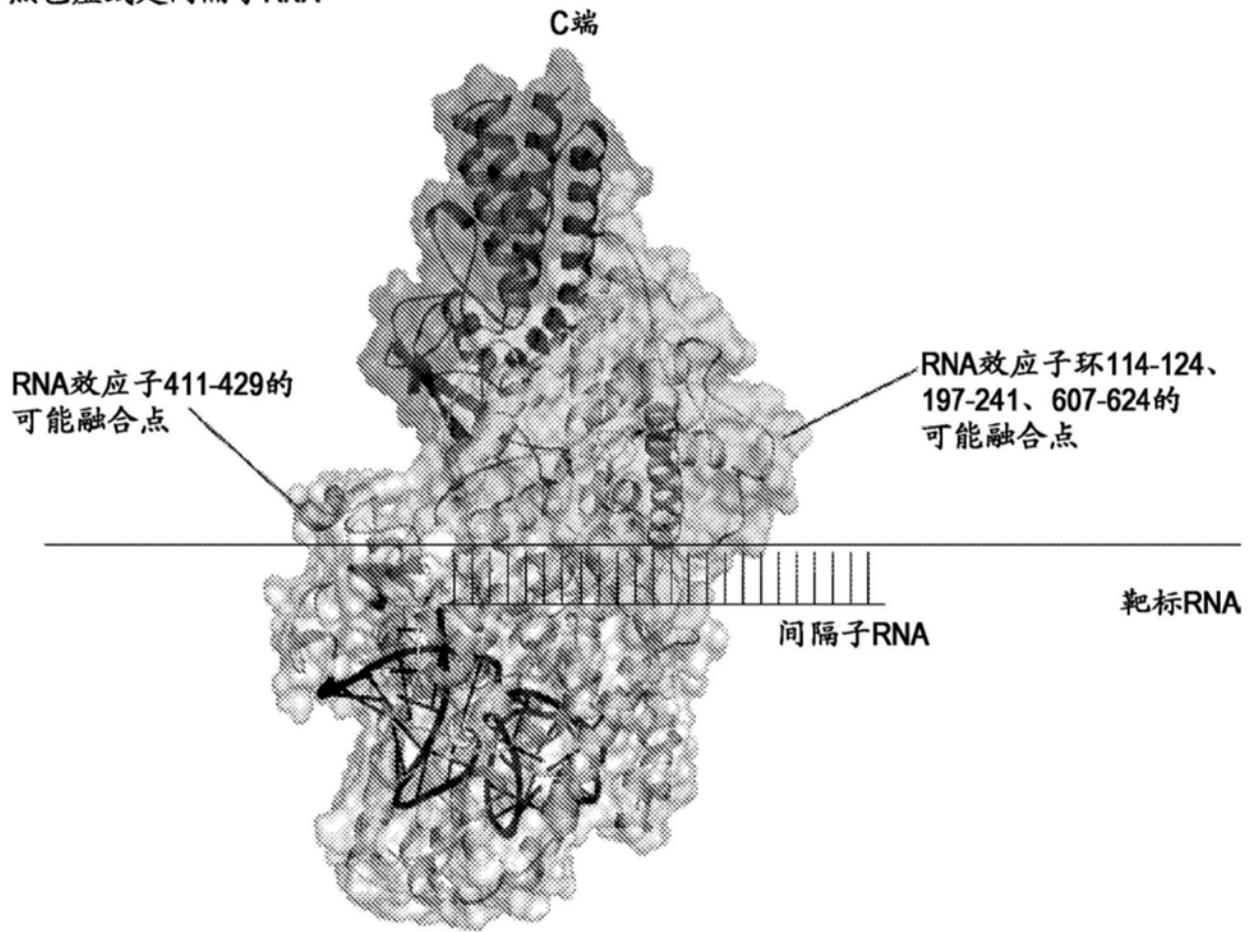


图21



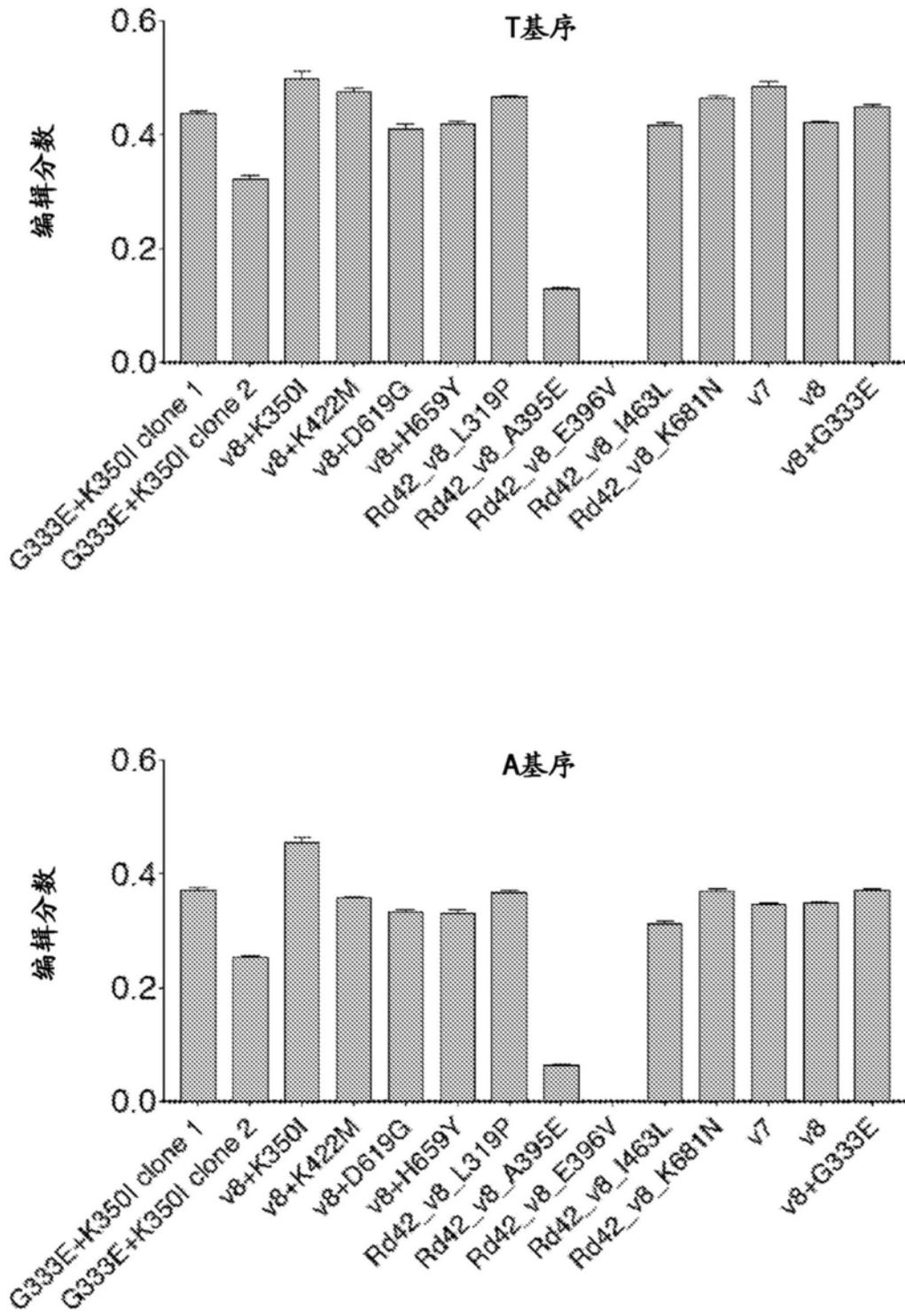


图23

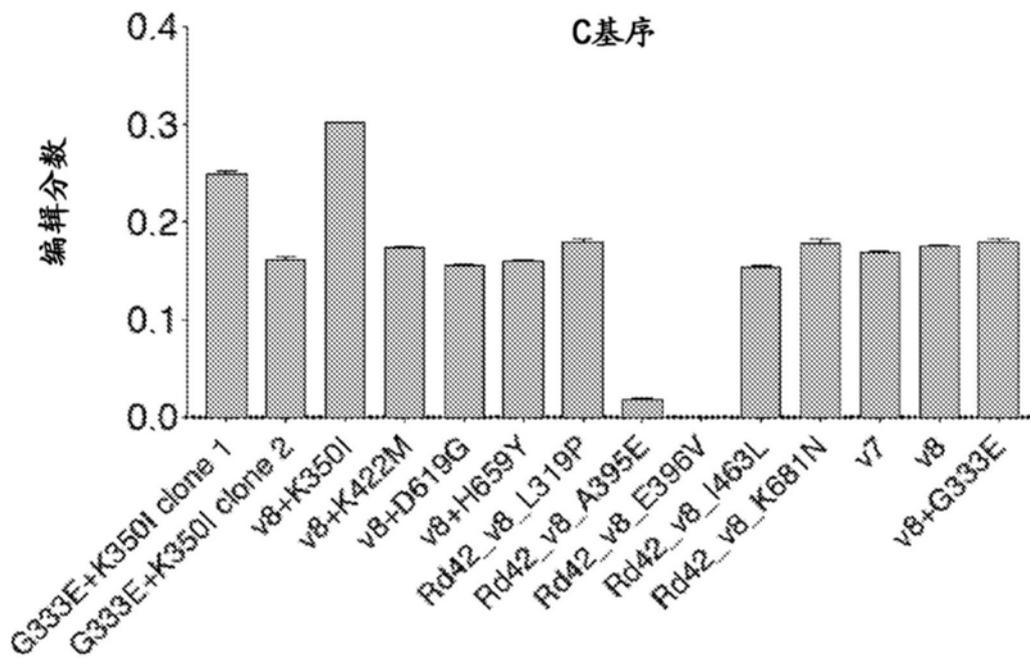
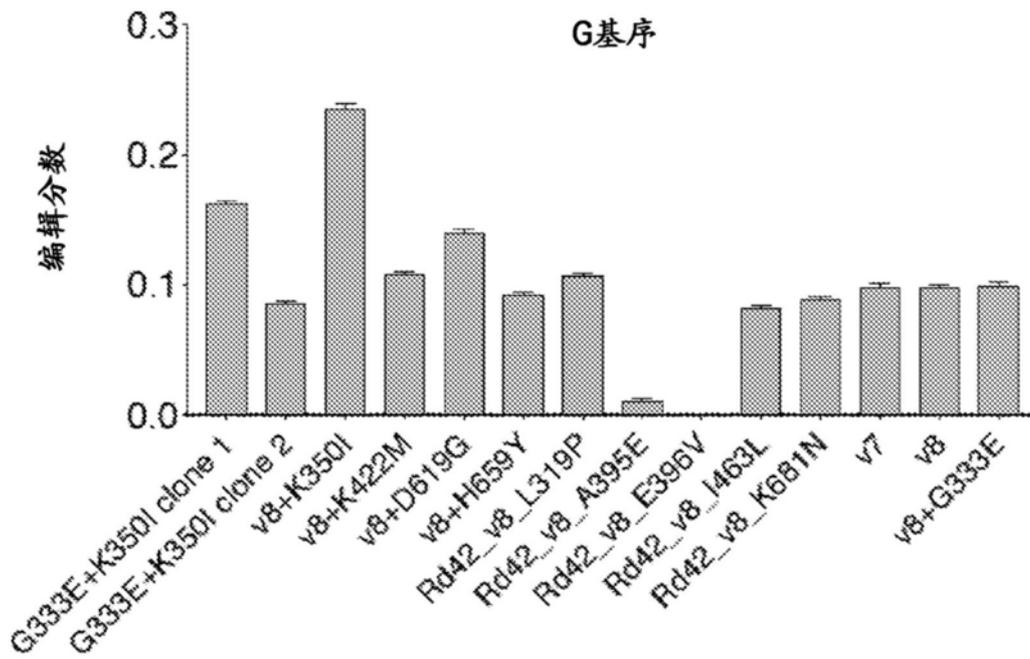


图23 (续)

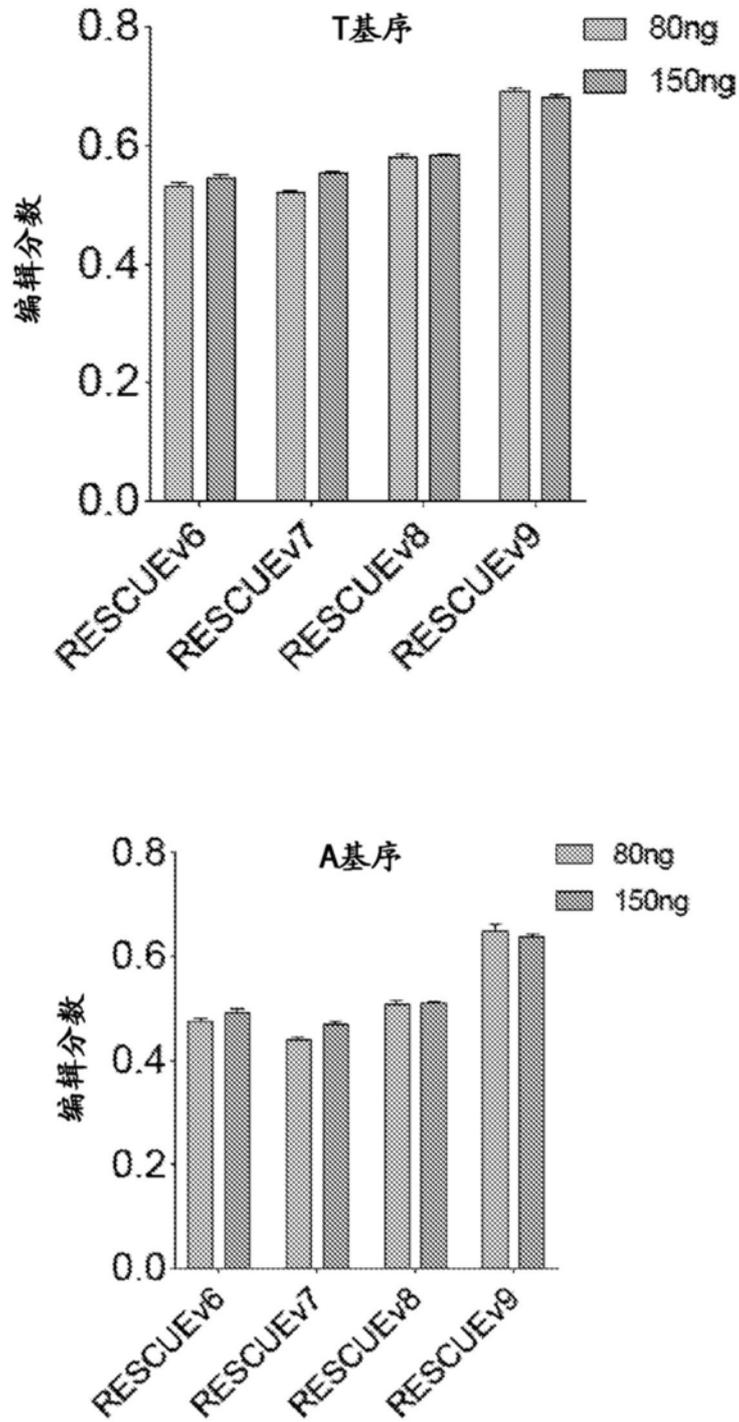


图24

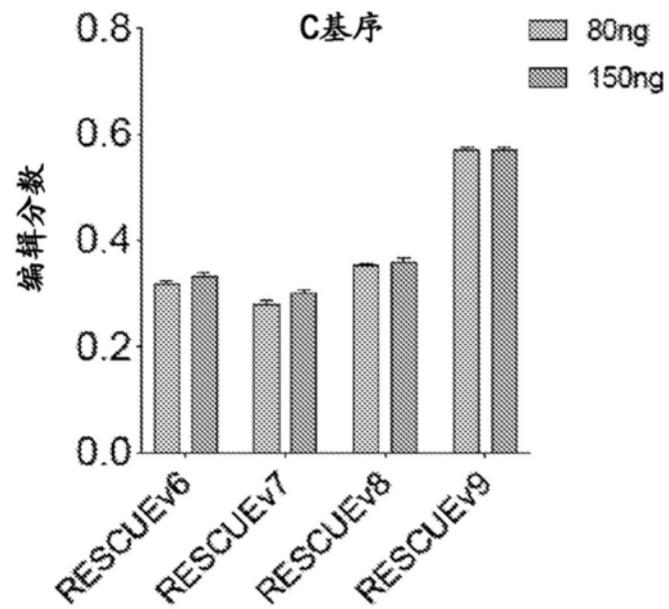
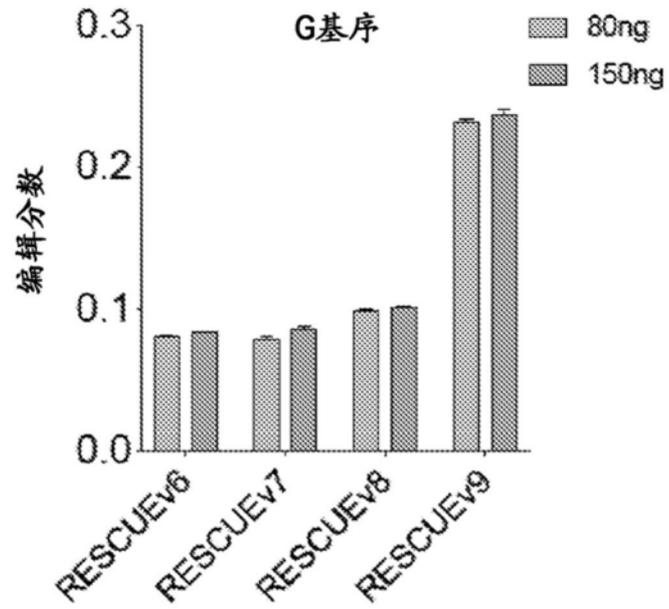


图24(续)

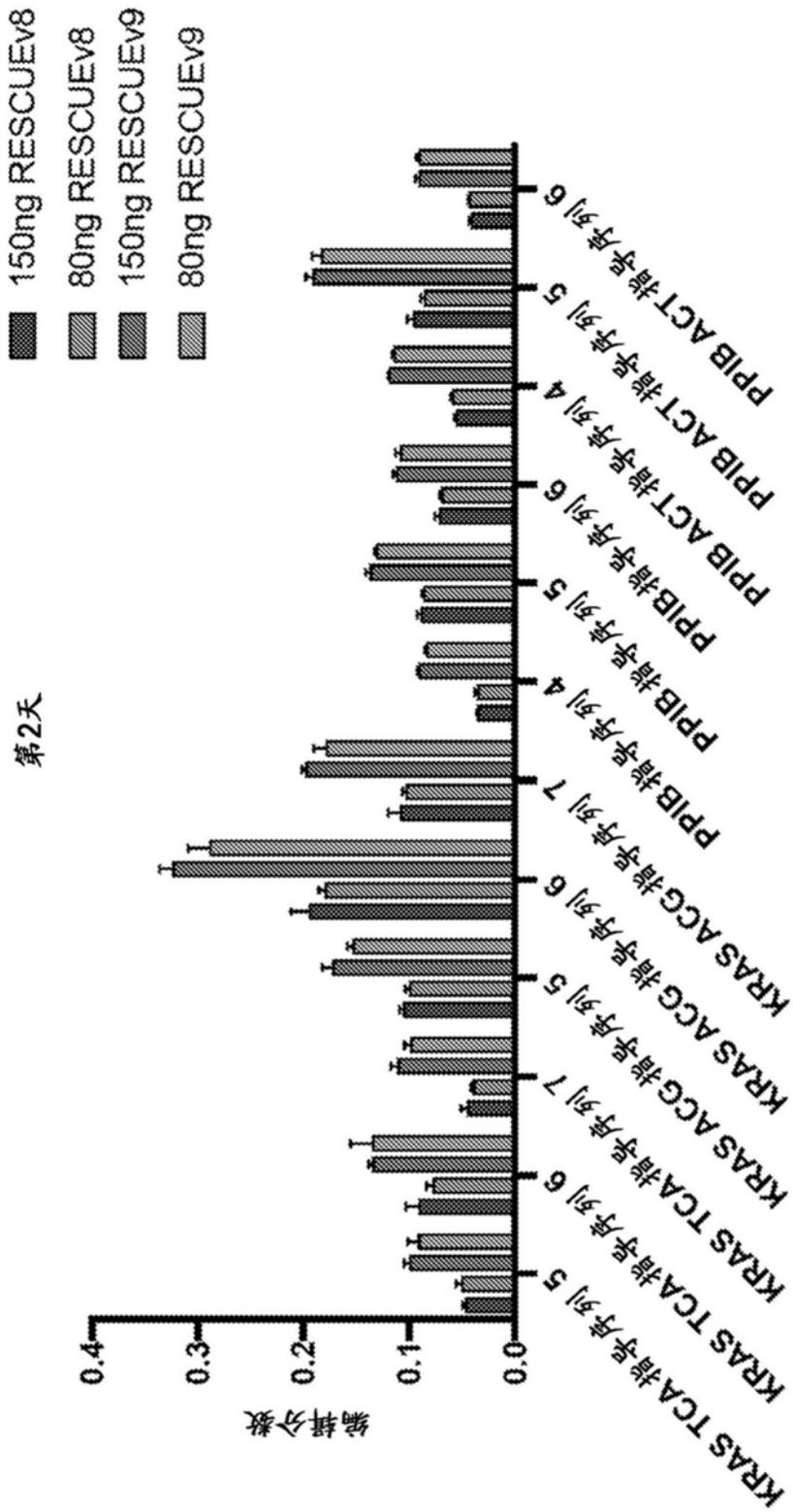


图25



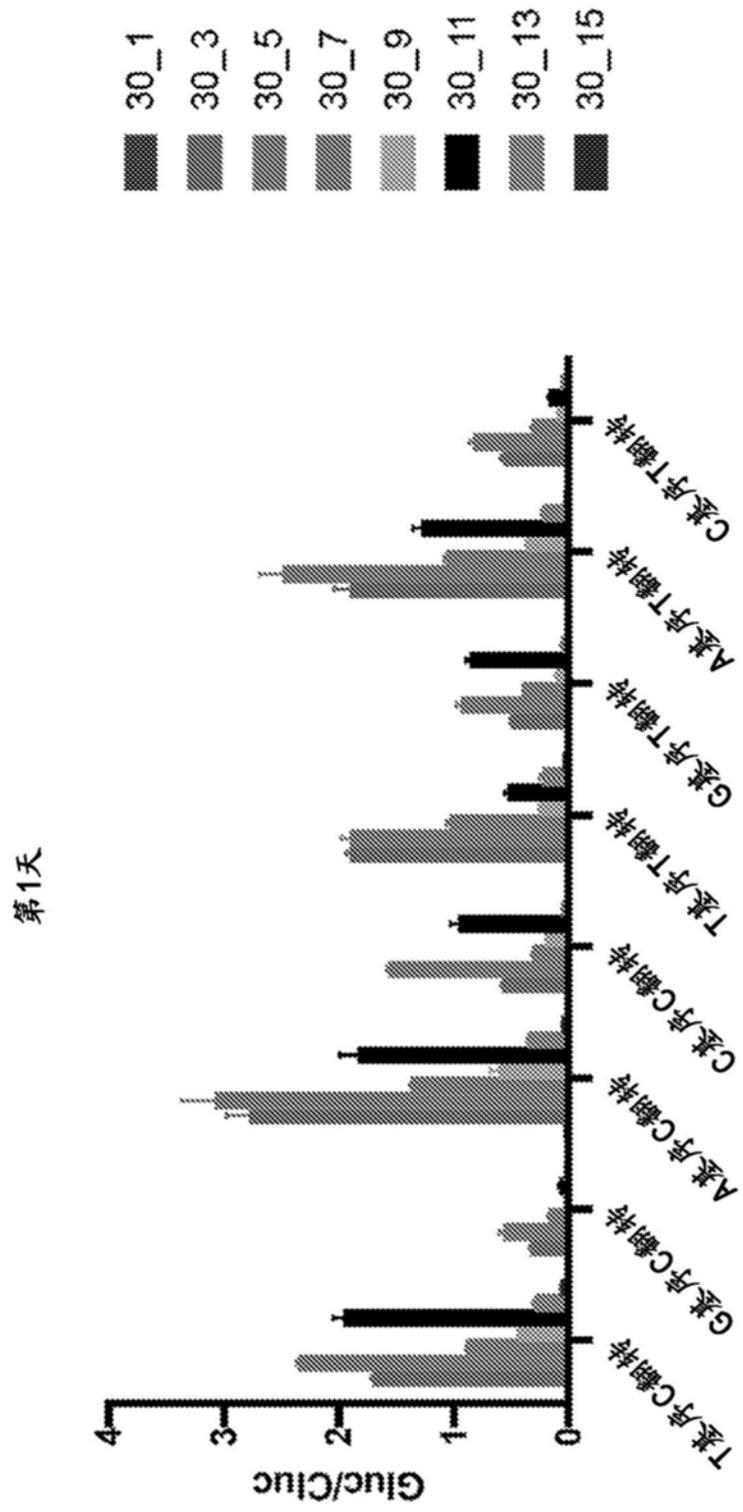


图27

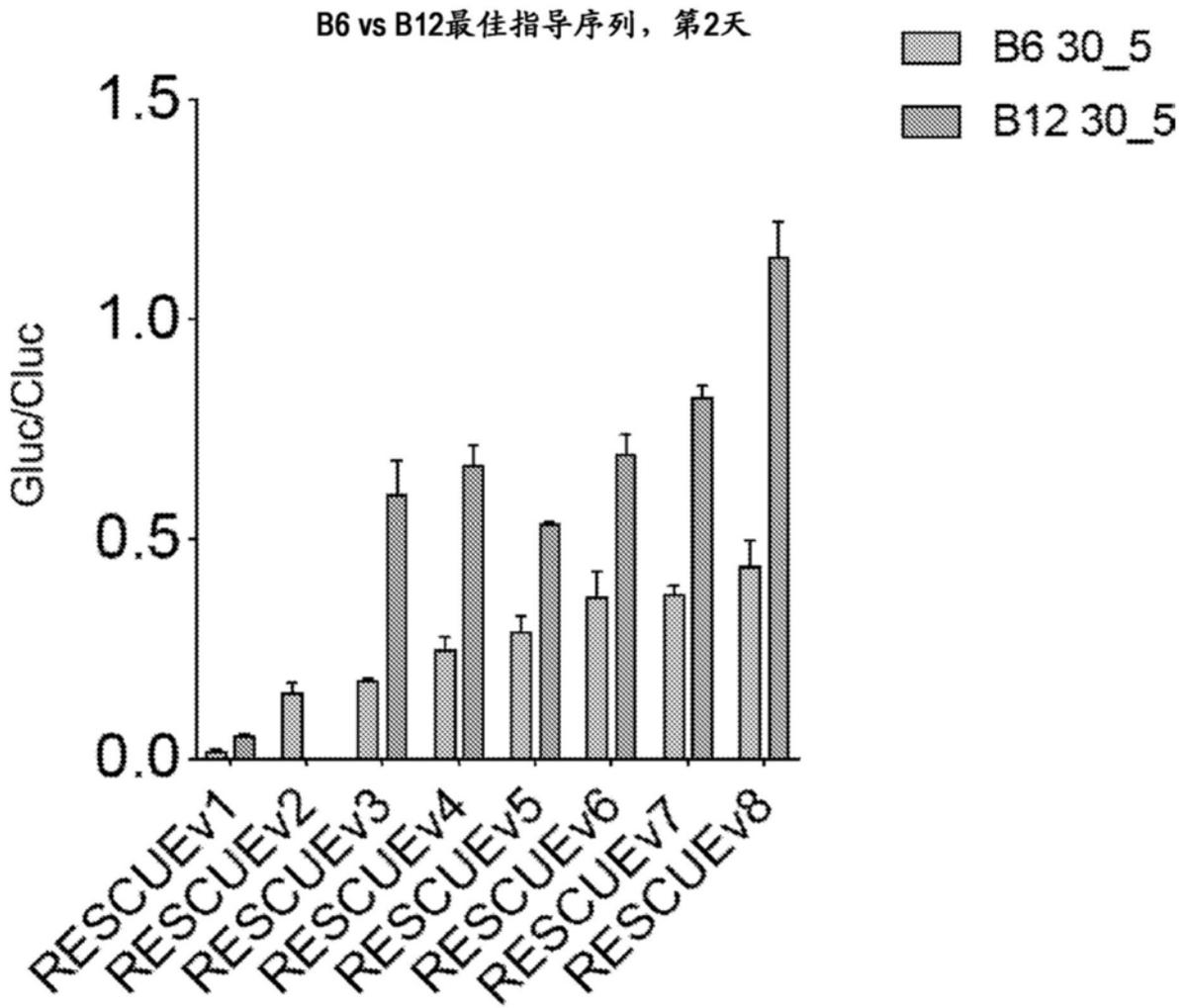


图28

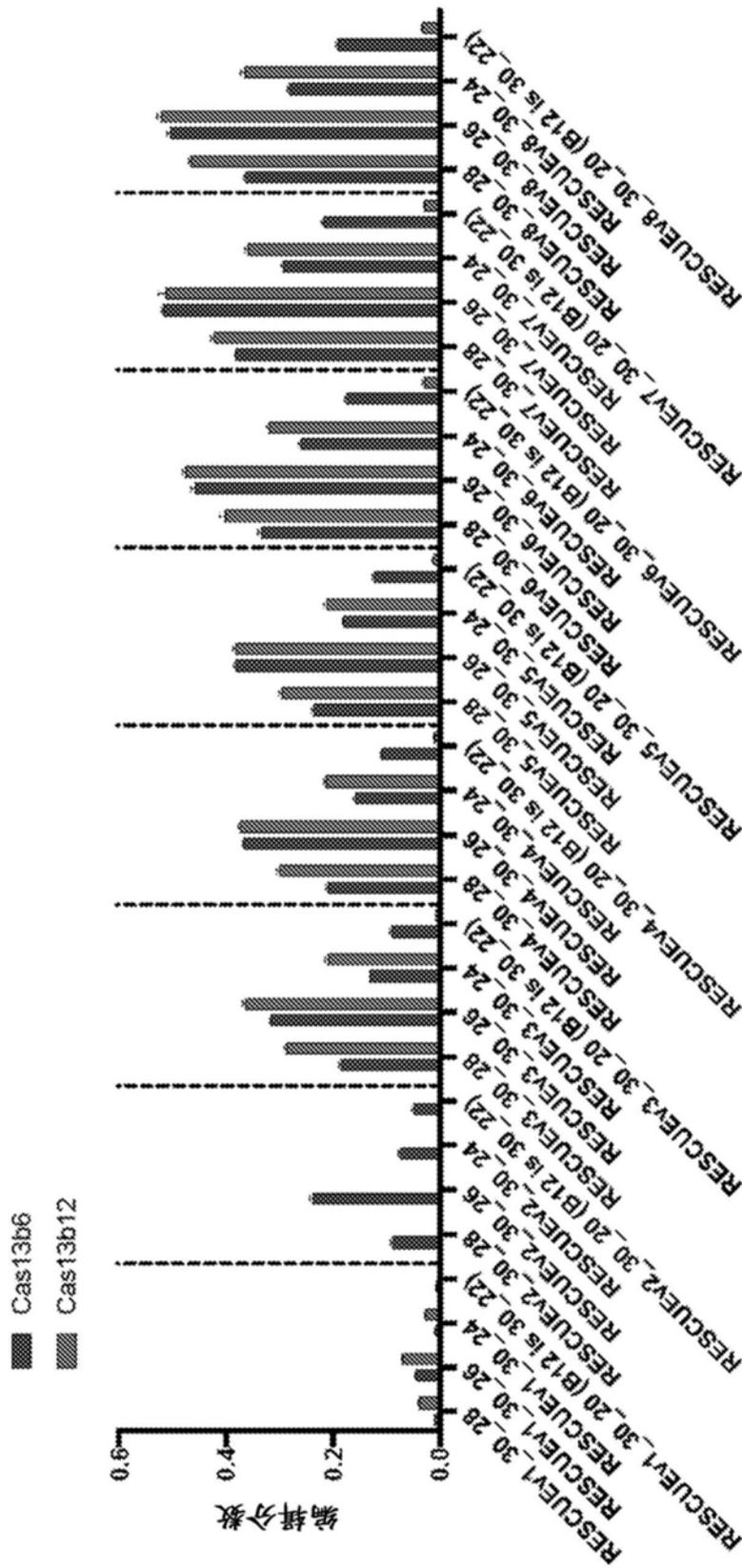


图29

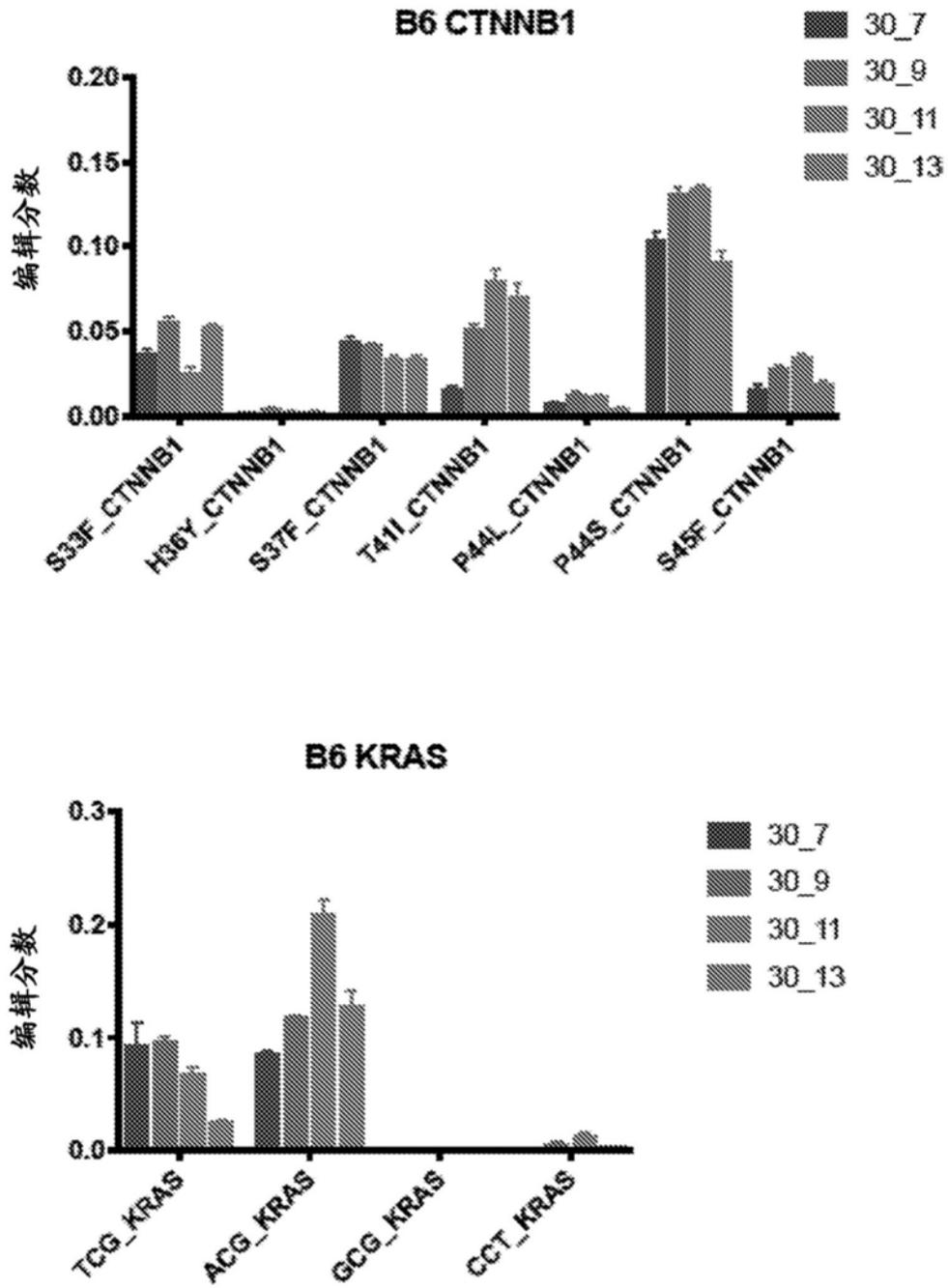


图30

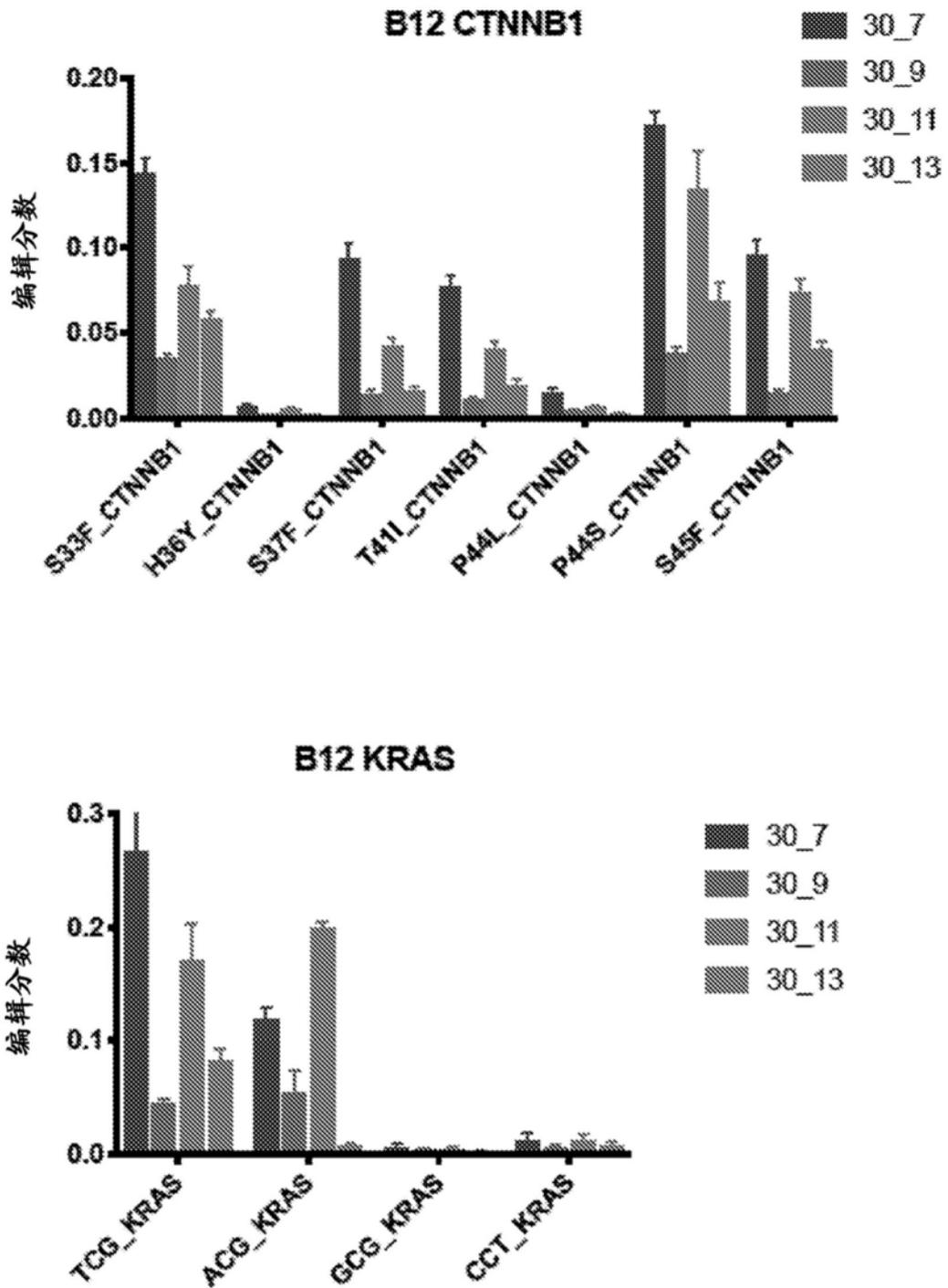


图30(续)

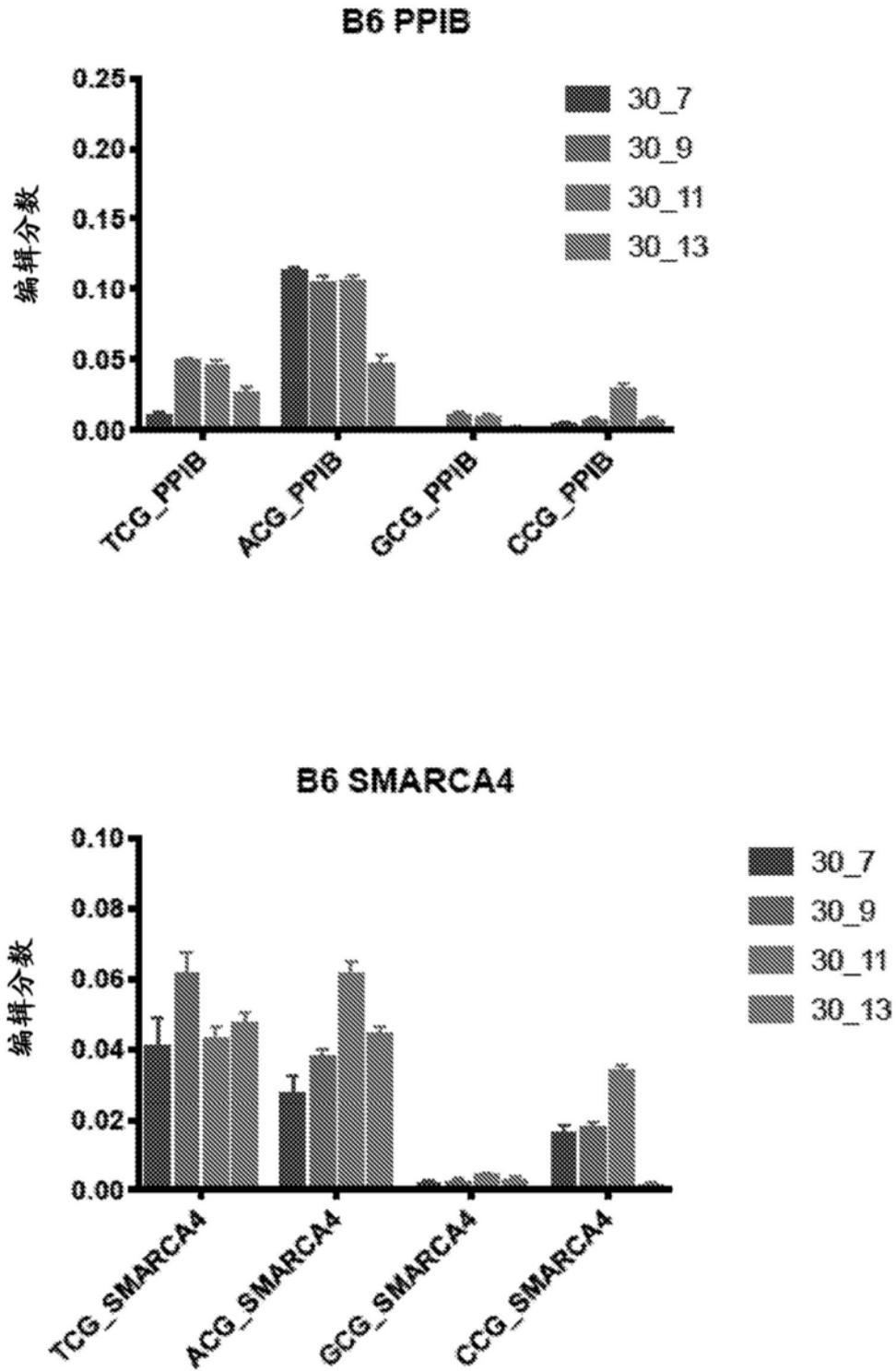


图30(续)

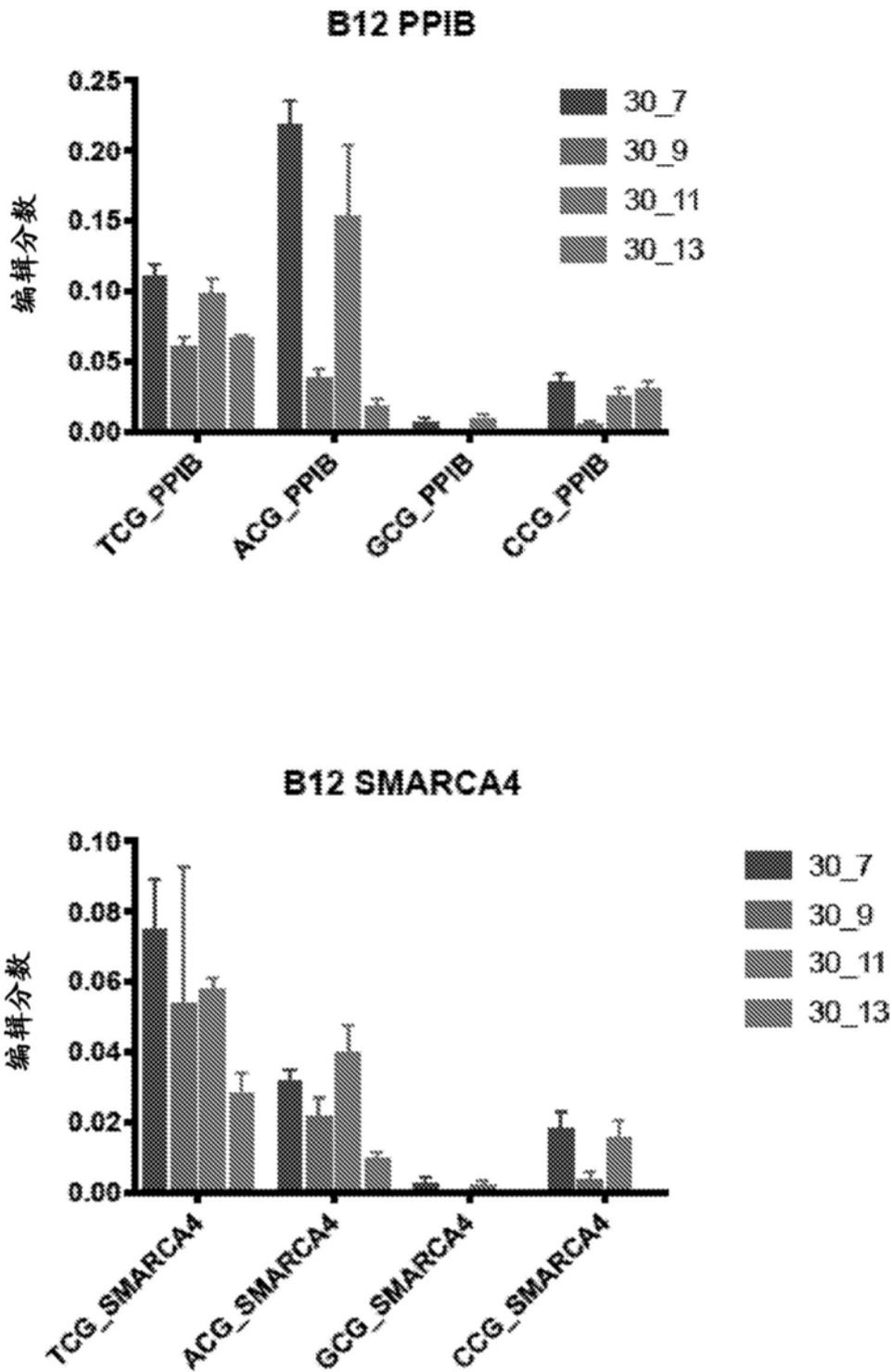


图30(续)

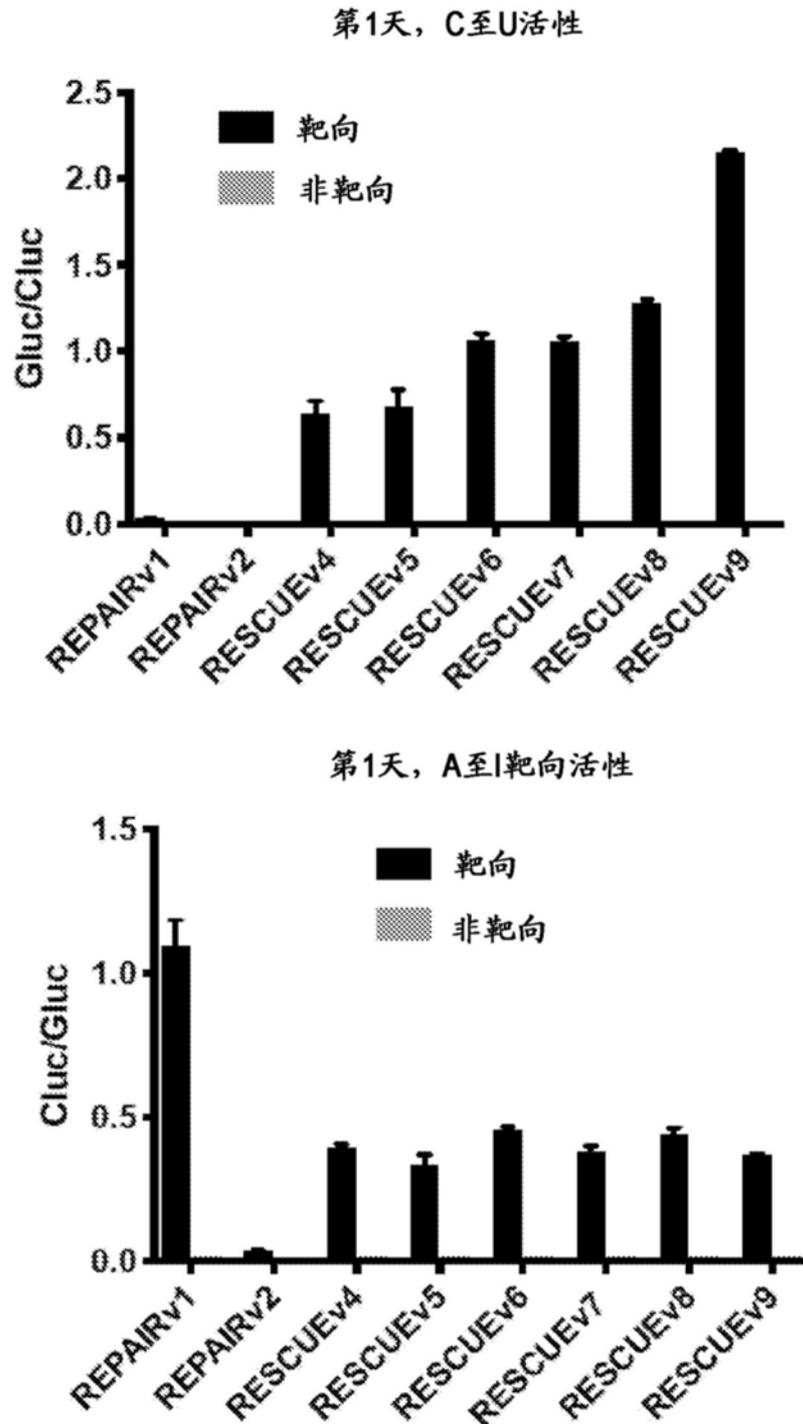


图31

第1天, A至I活性, 仅非靶向

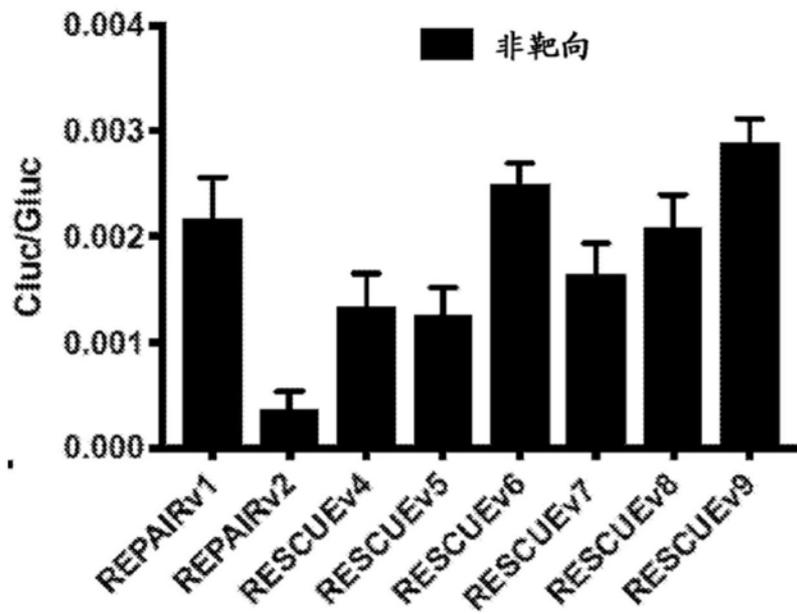


图31 (续)

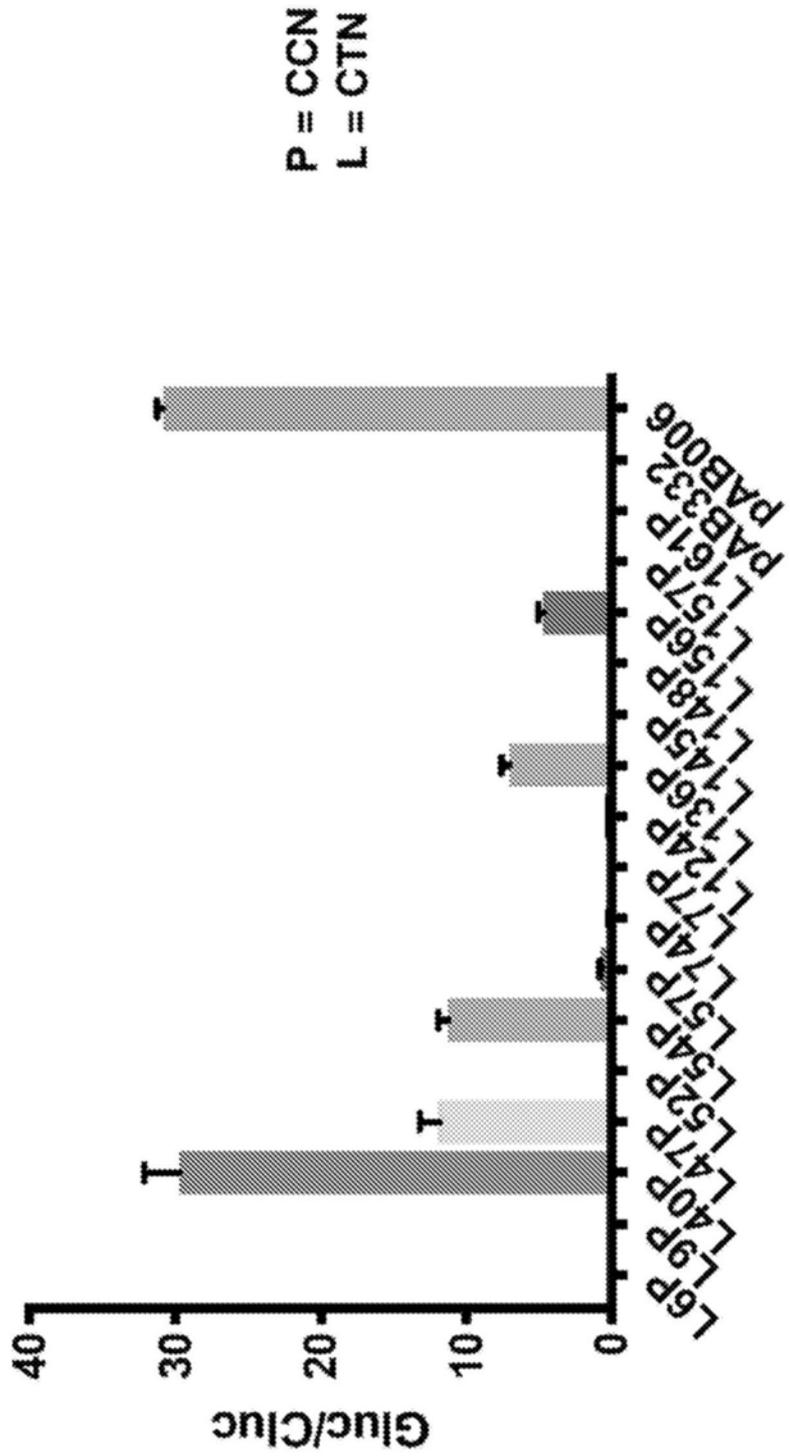


图32

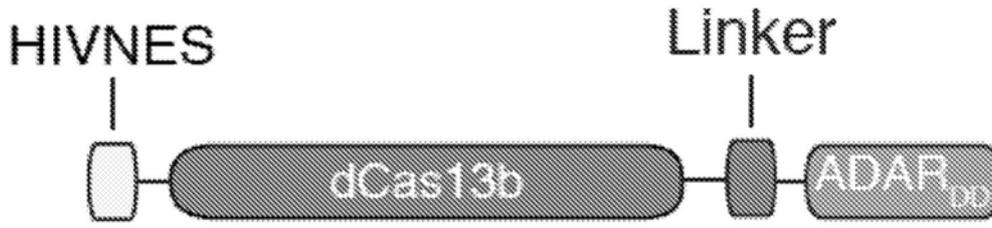


图33A

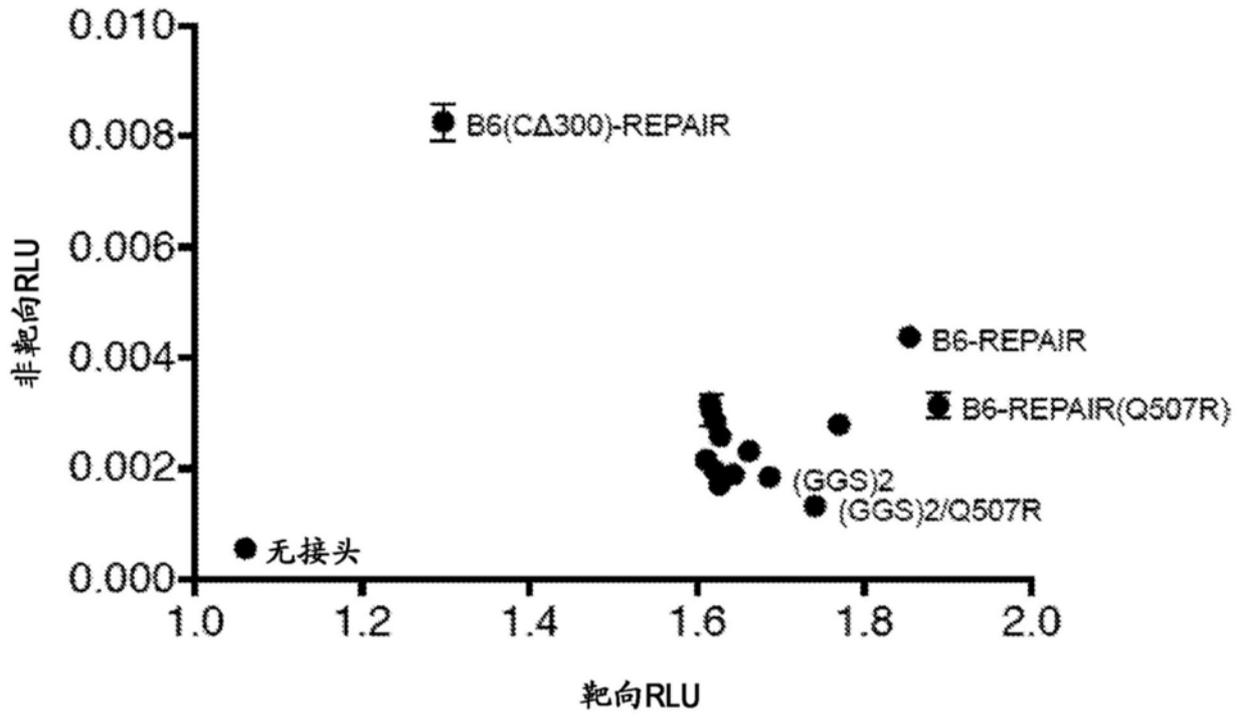


图33B

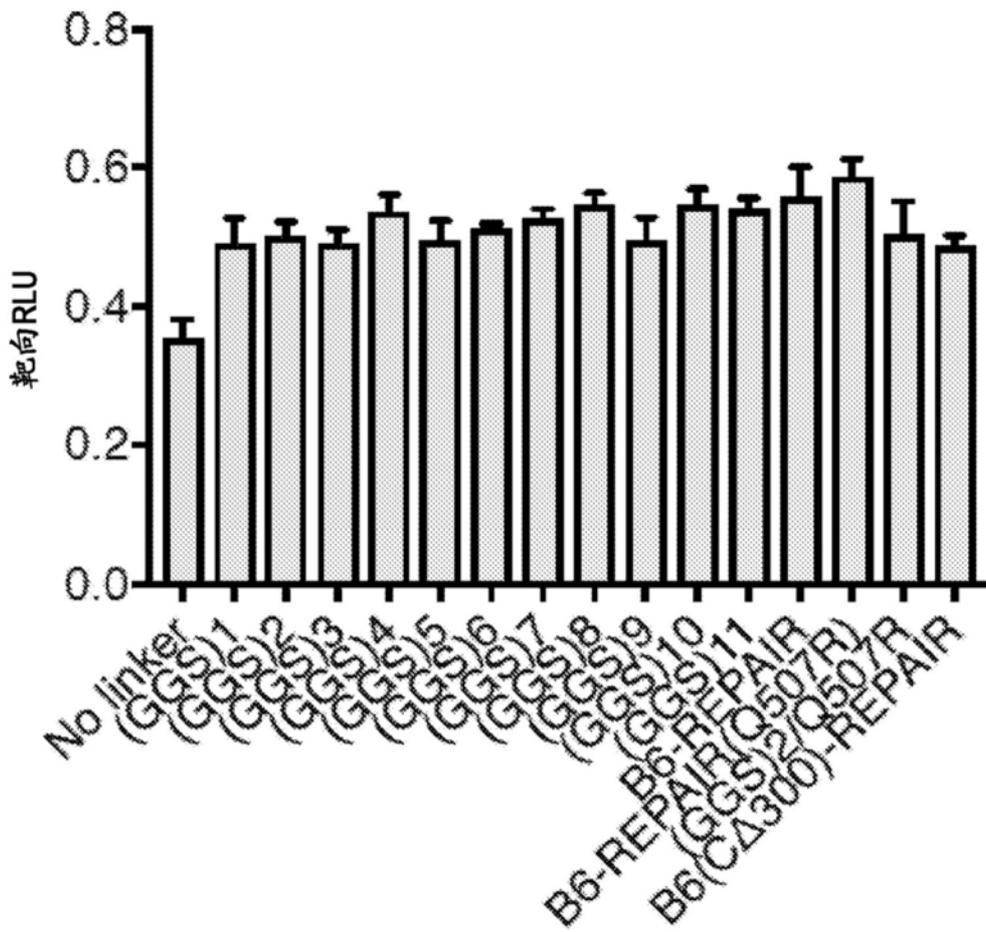


图34

**Cluc W85X mRNA**

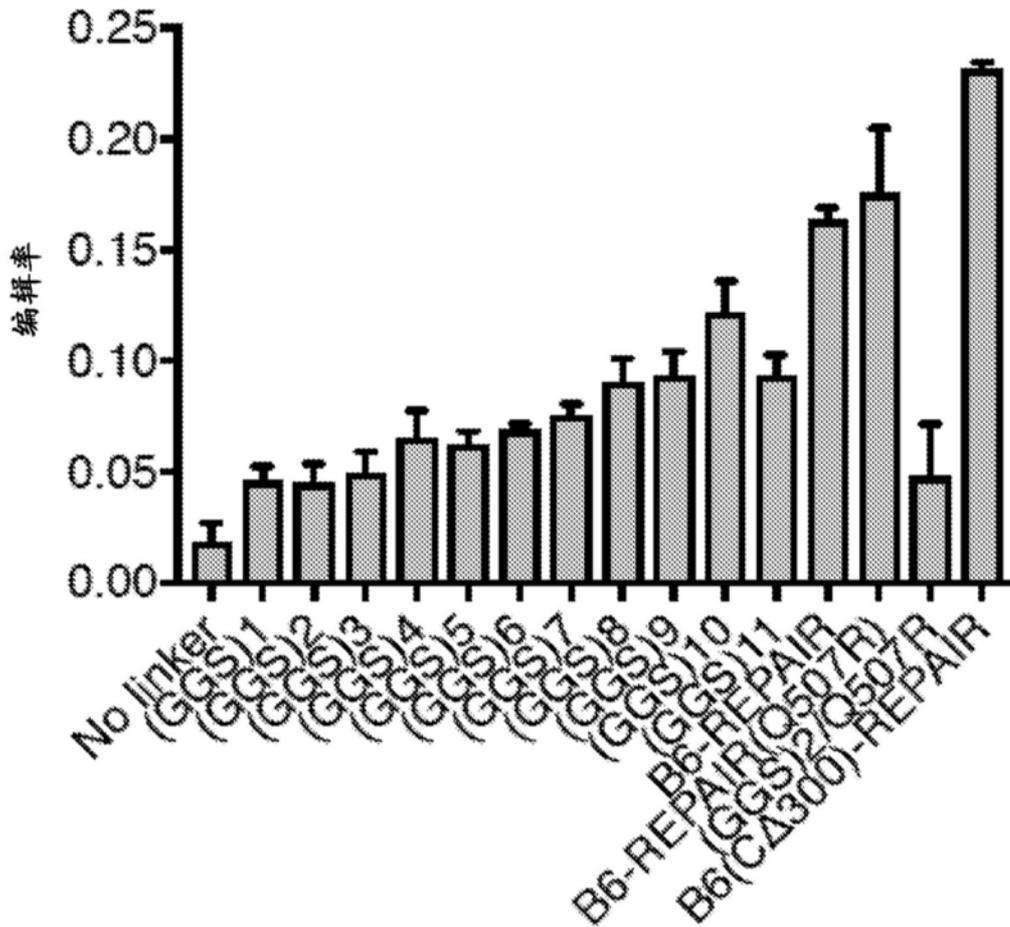
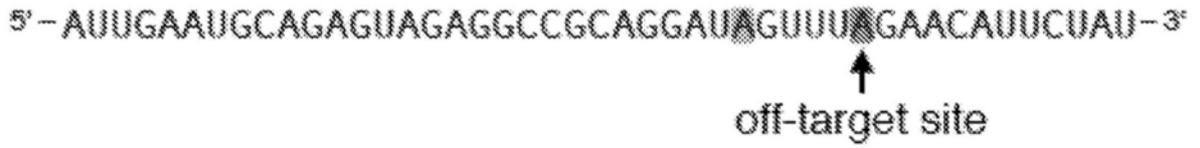


图35

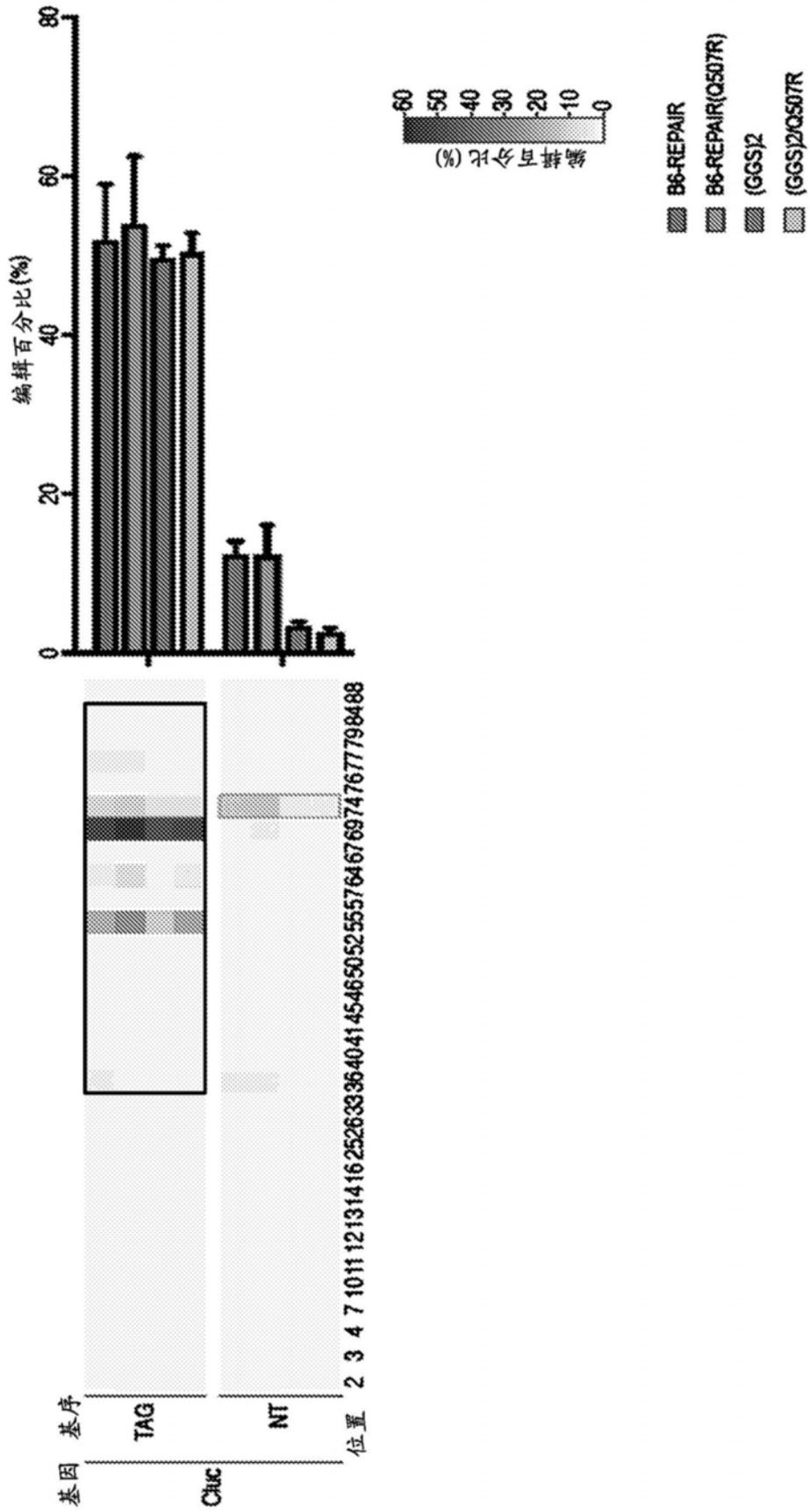


图36

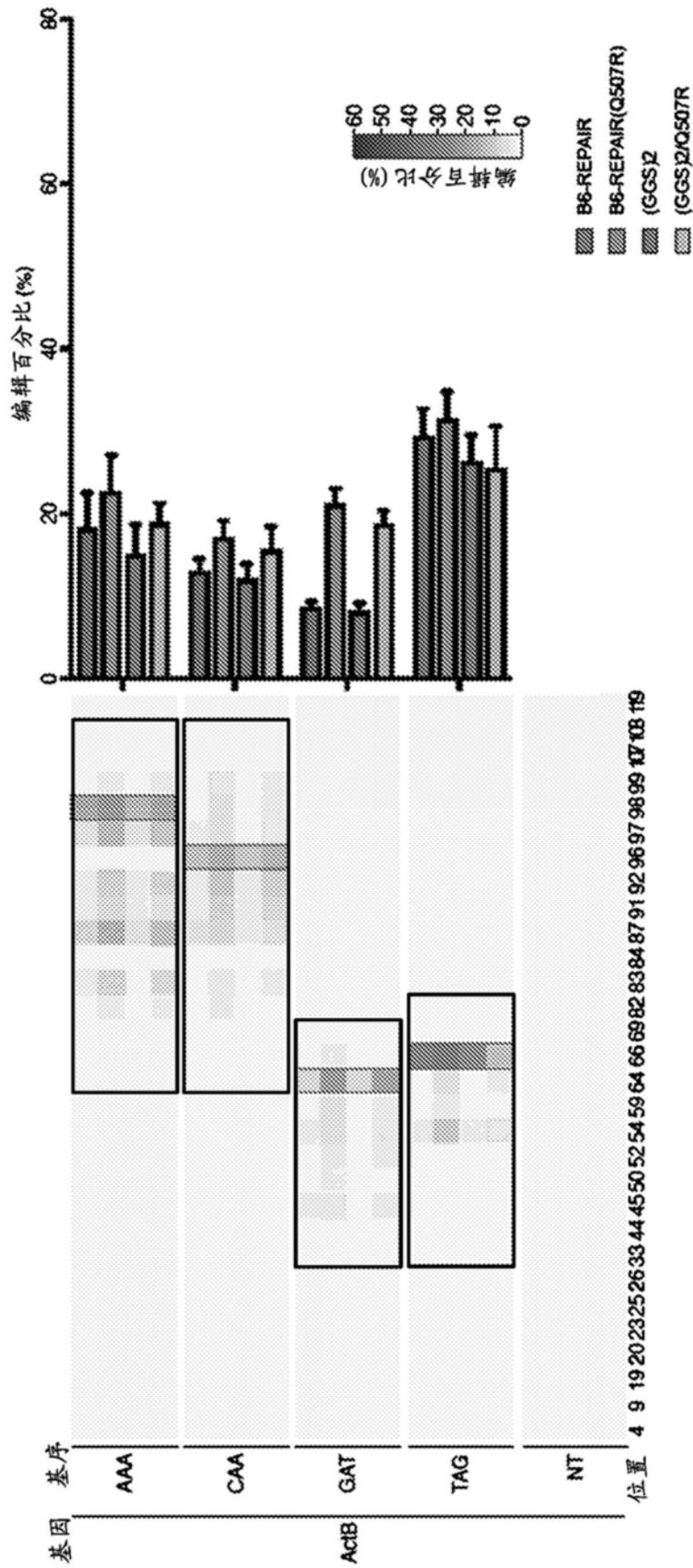


图37

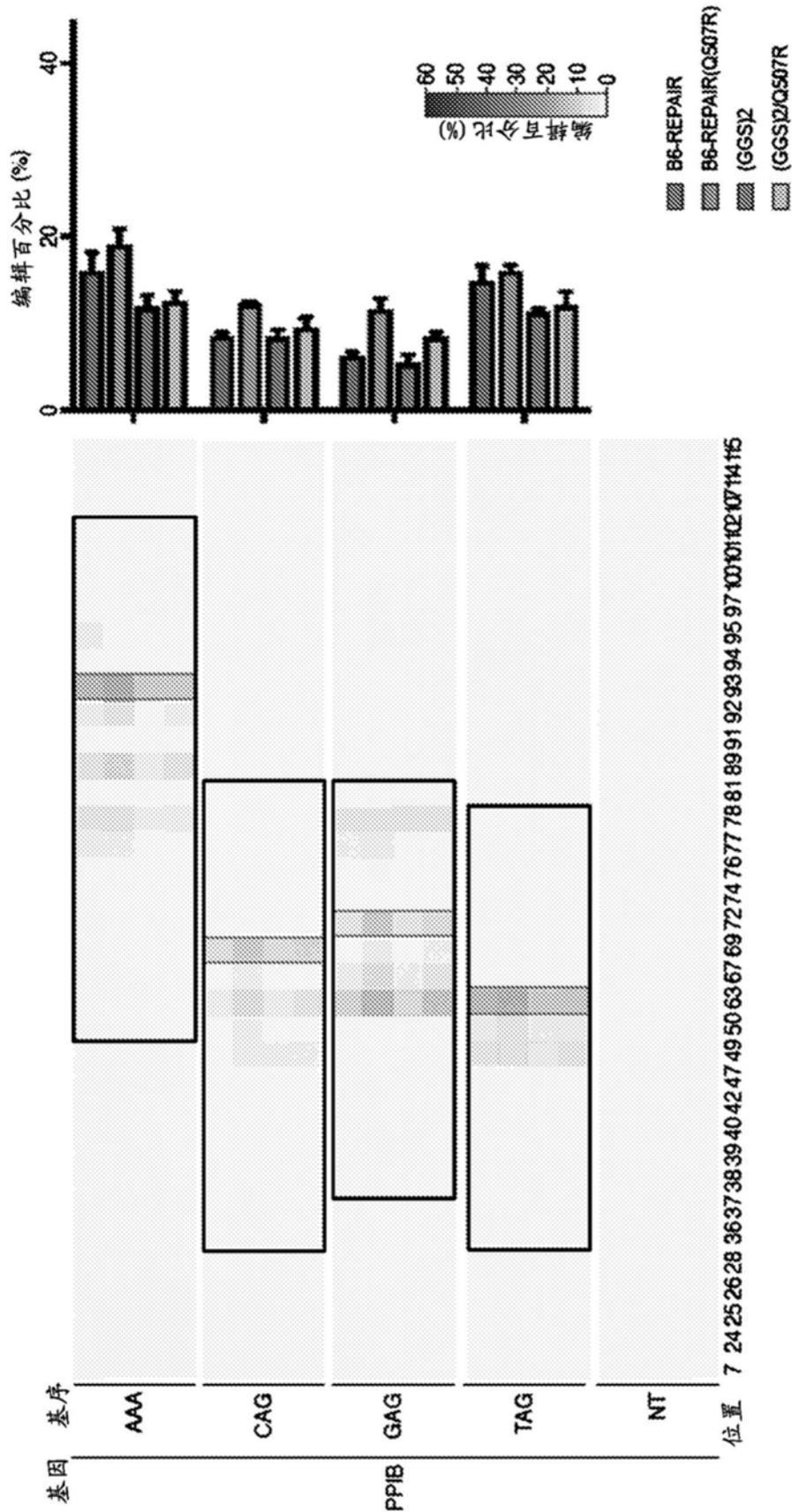


图37 (续)

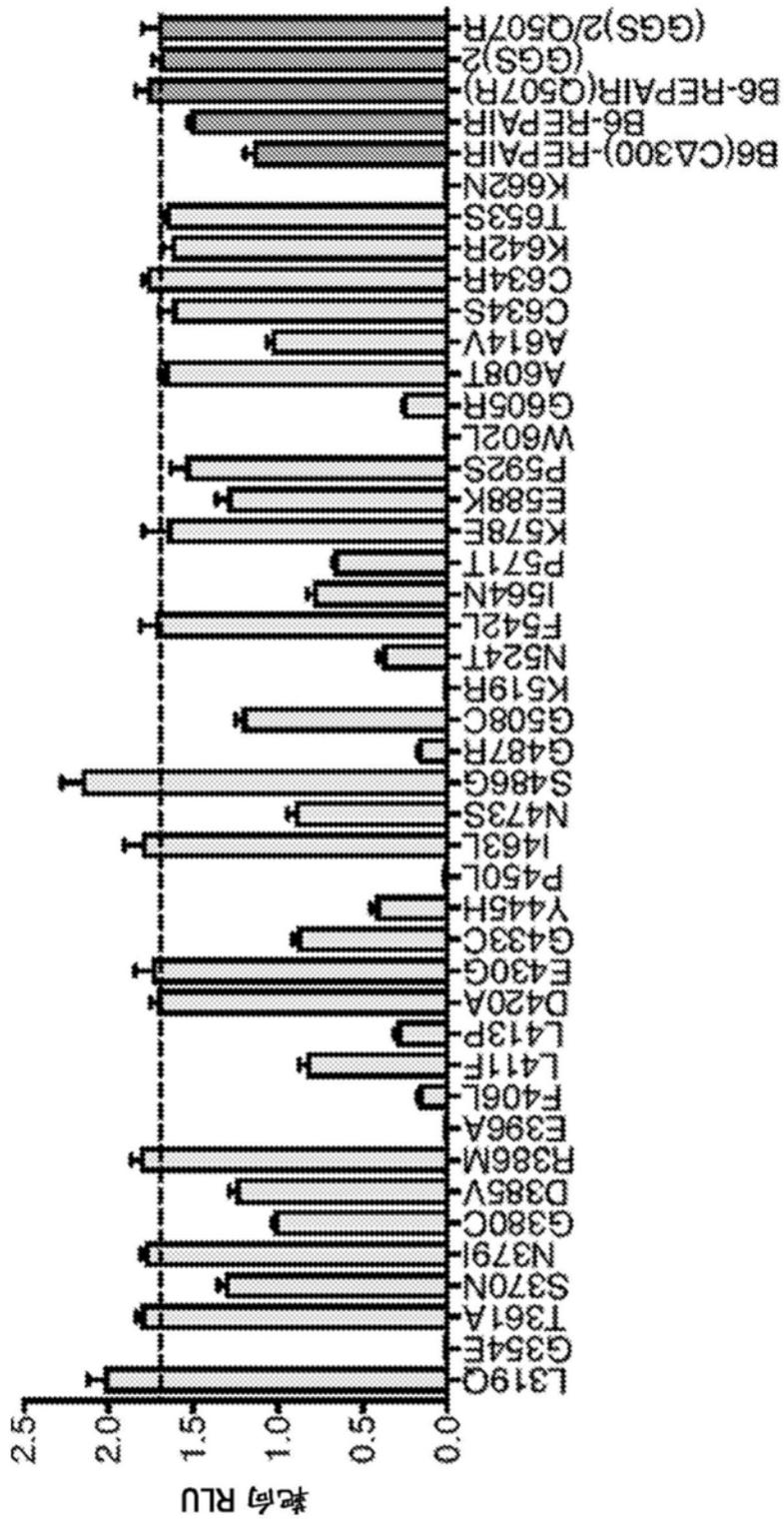


图38

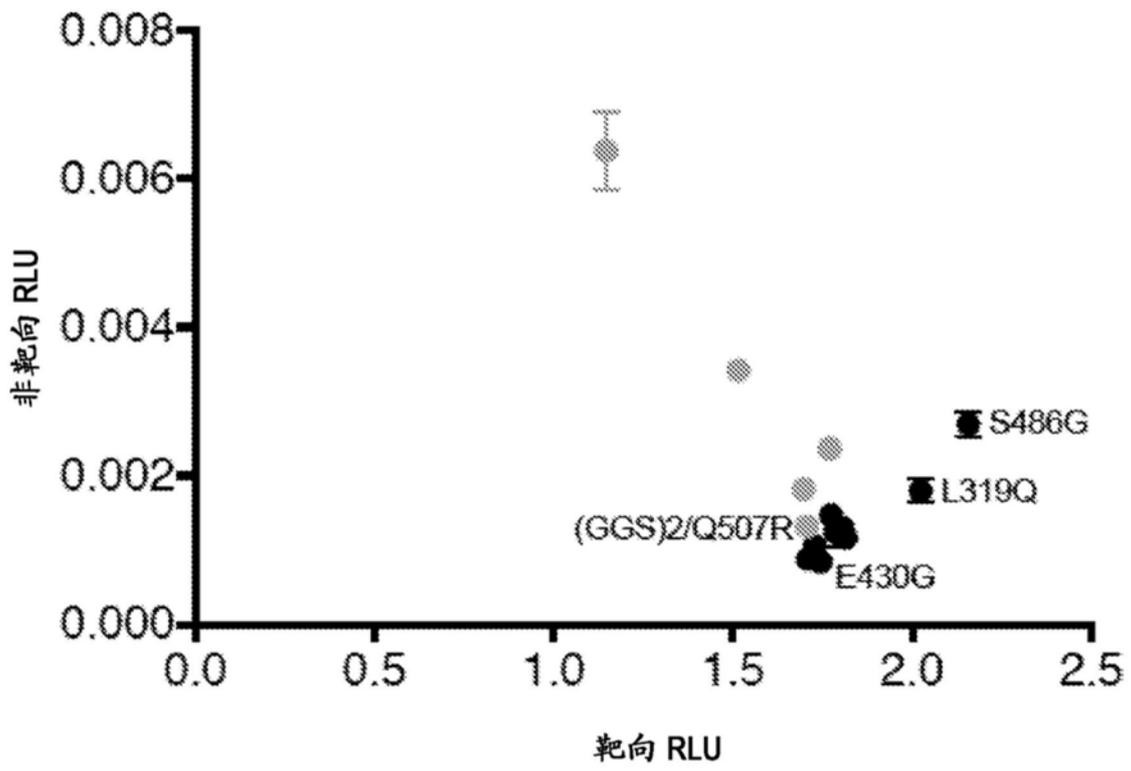


图39



图40A

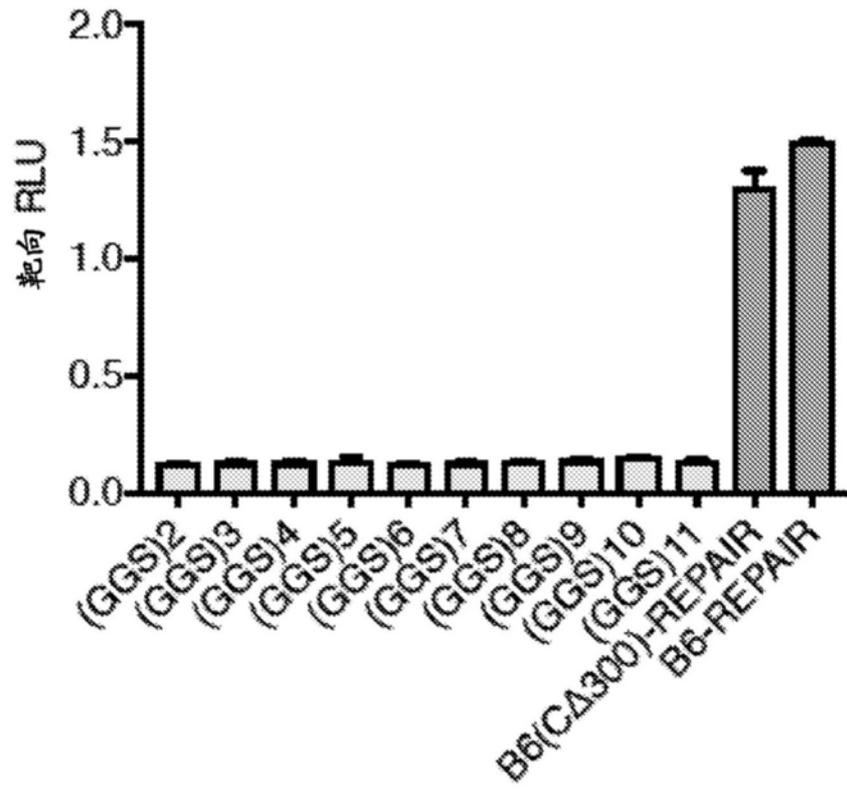


图40B

**ADE2 W185X: 靶向**

5'-CGGAAGCTTTGGAGTACTGAAGGATCGTCCCTTTGTACGCCGAAAAATGGCACCATTIACIAAGAATTAGCAGTCATG-3'  
W185X

**URA3 A116A: 脱靶**

5'-AATACAGTCAAATTGCGACTCTGCCGGGTGTATACAGAAATAGCAGAAATGGCAGACATTACGAAATGCACACGGGTGGT-3'  
W115X A116

图41



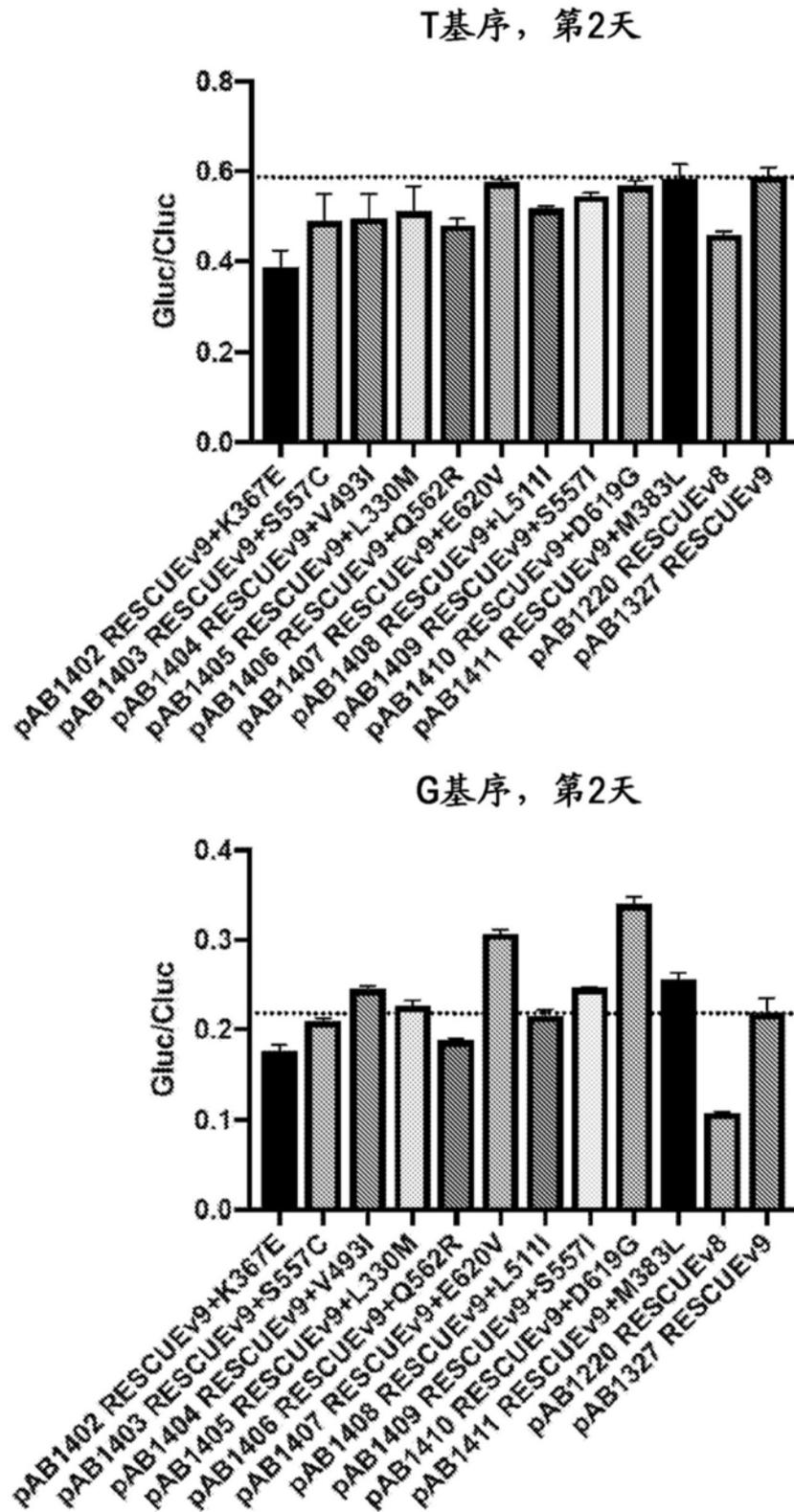


图42A (续)

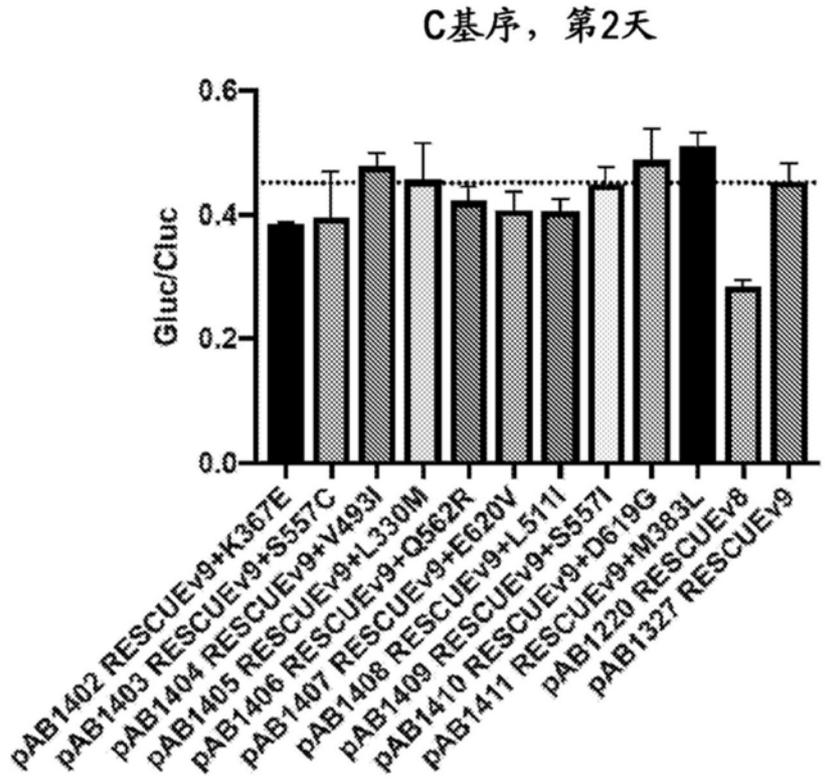
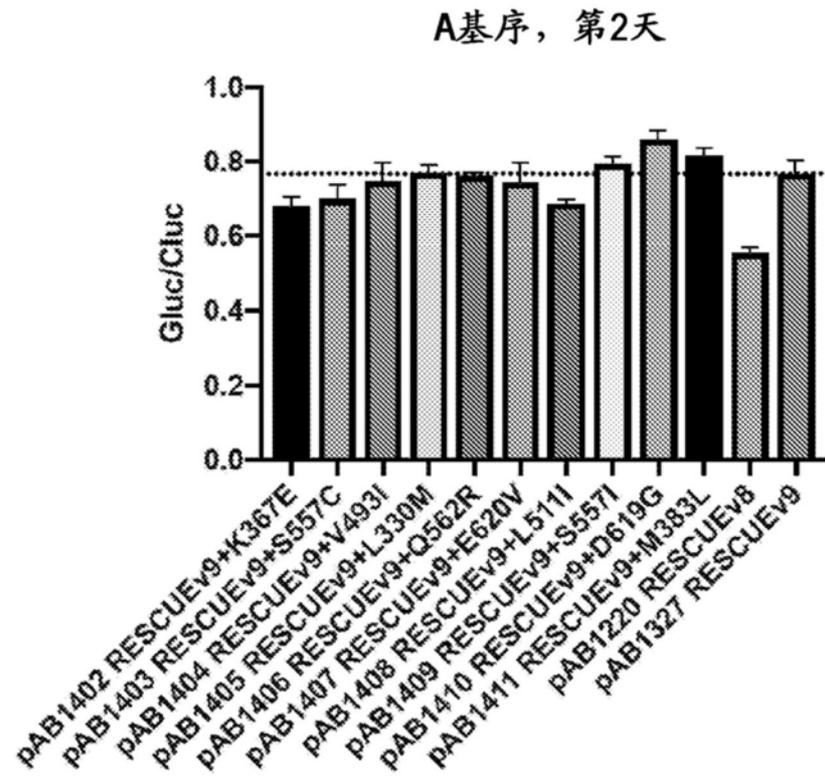


图42A (续)

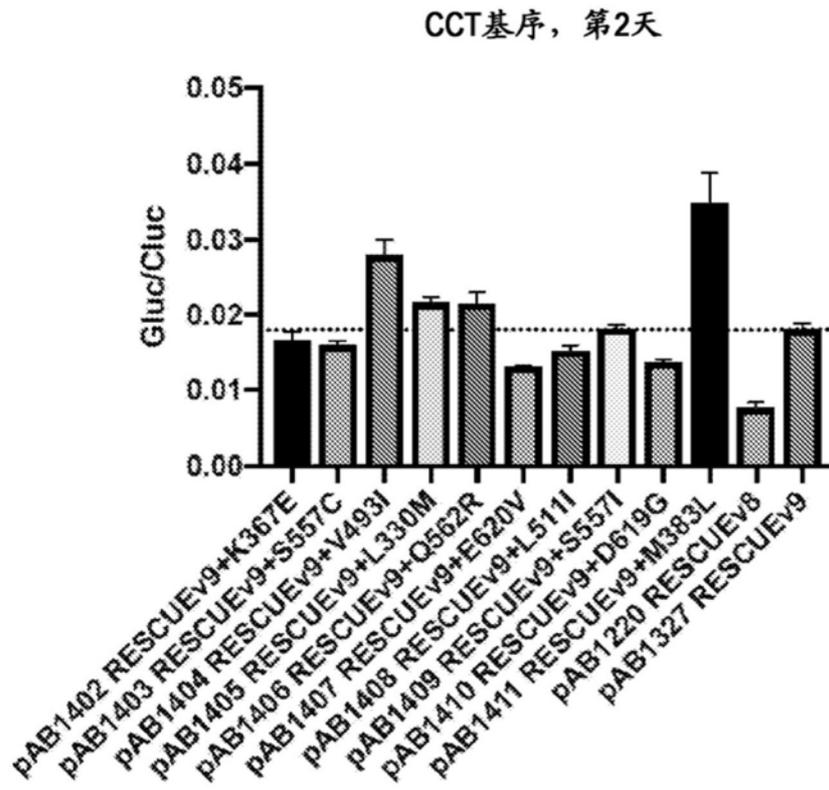


图42A (续)

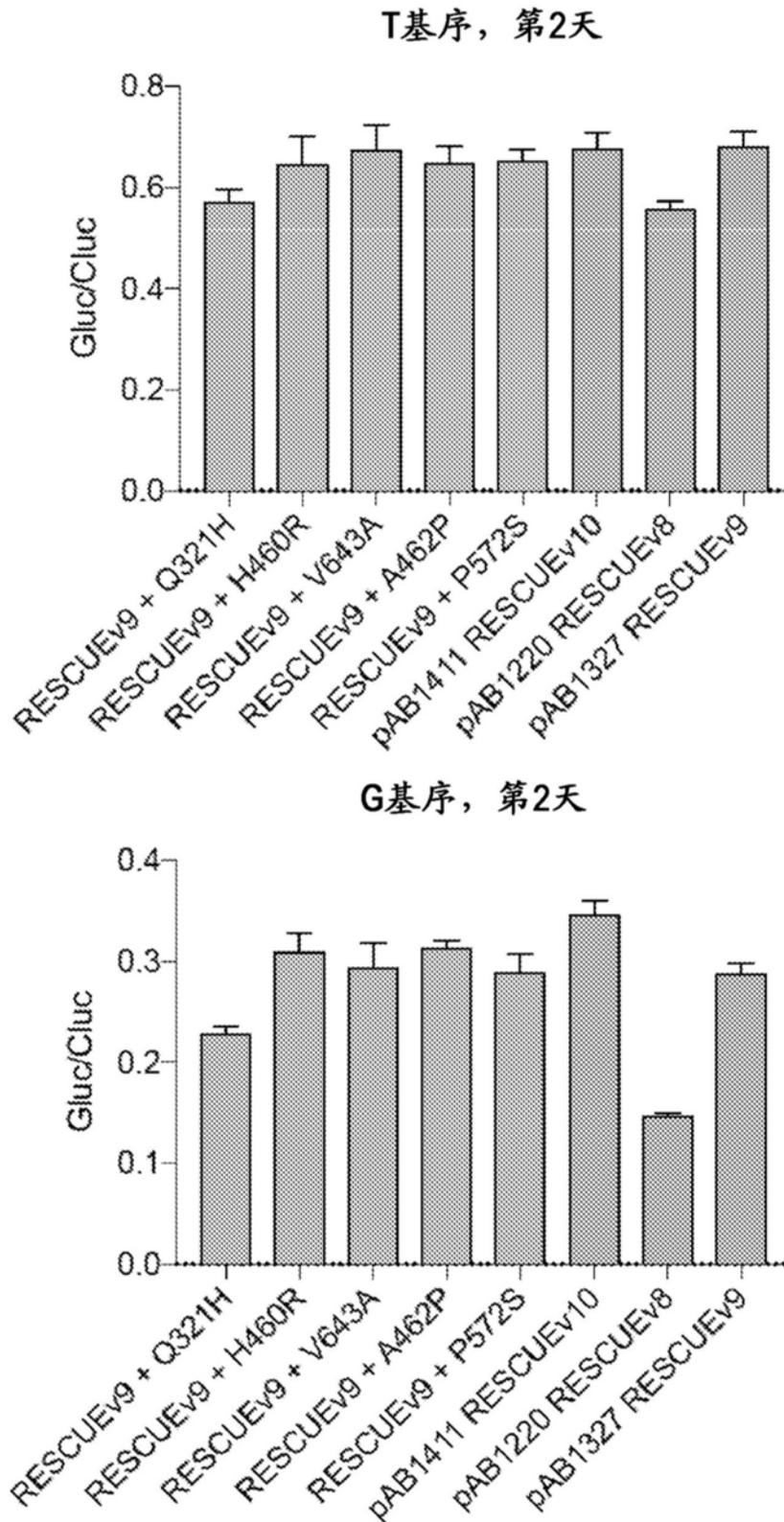


图42B

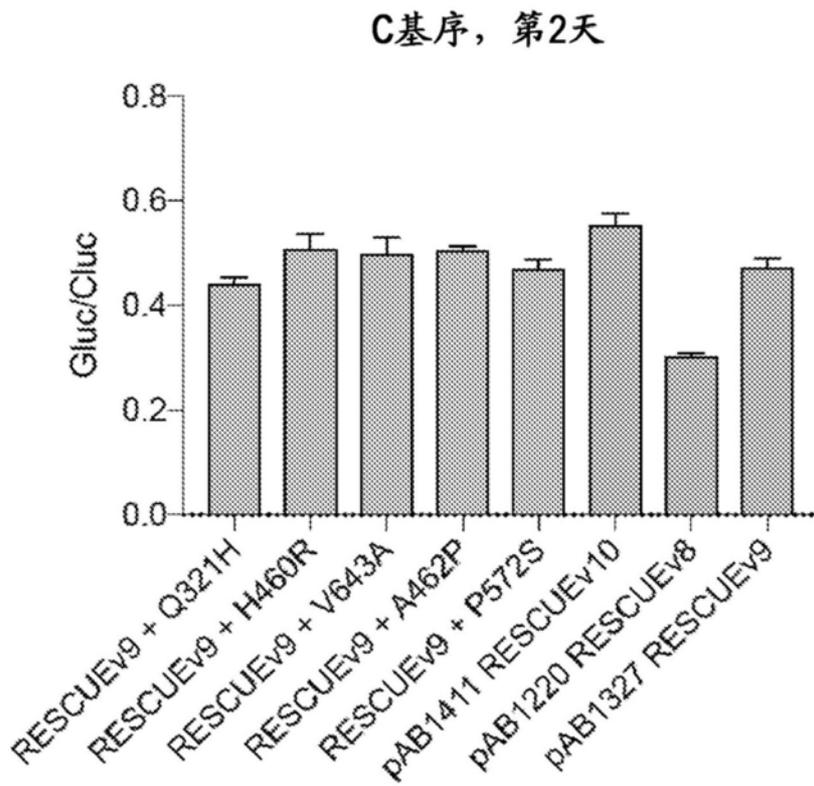
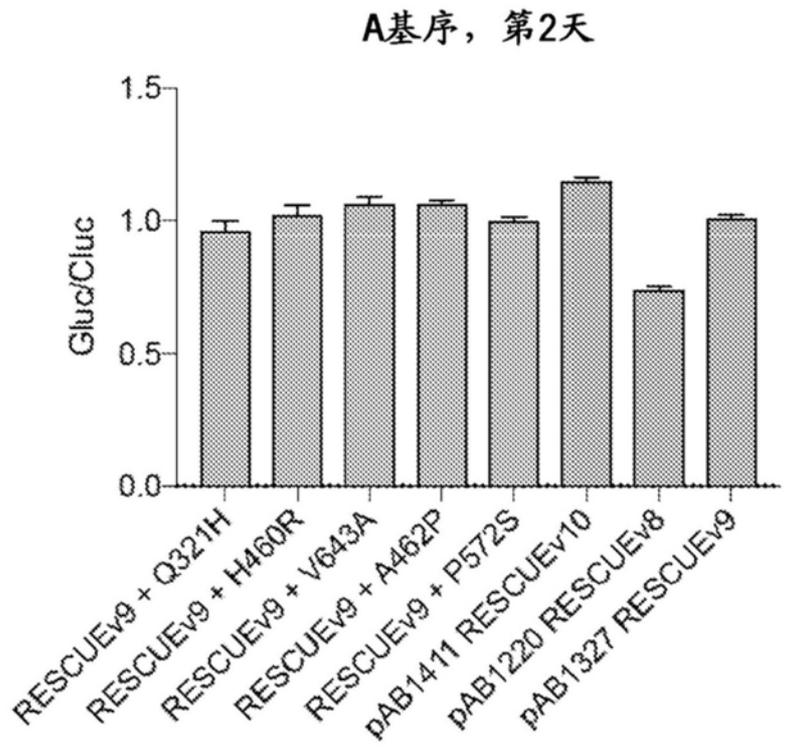


图42B(续)

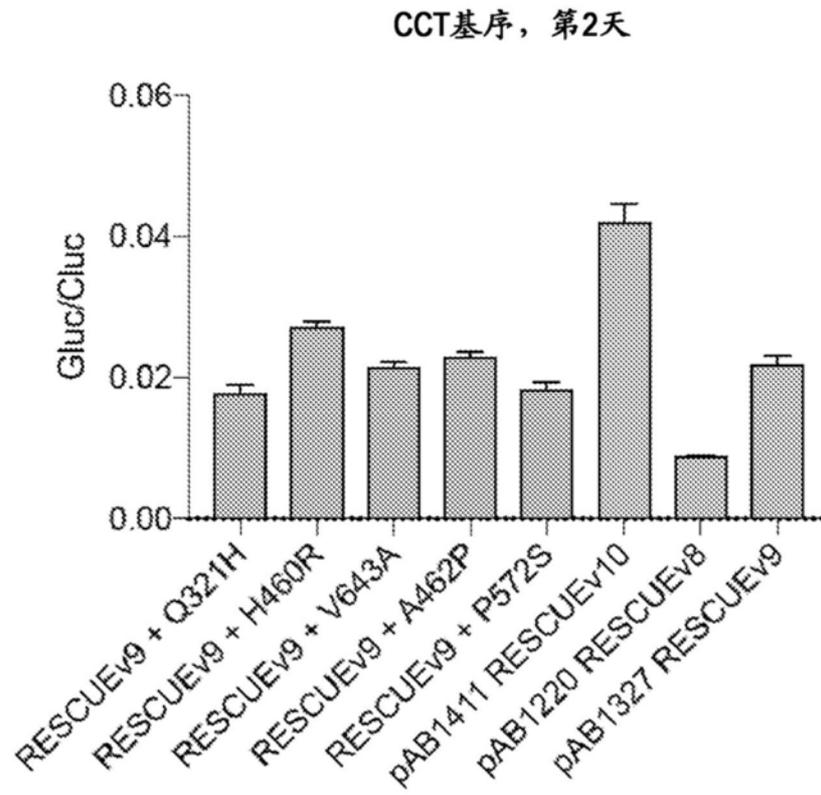


图42B (续)



第2天第59轮, CCG基序

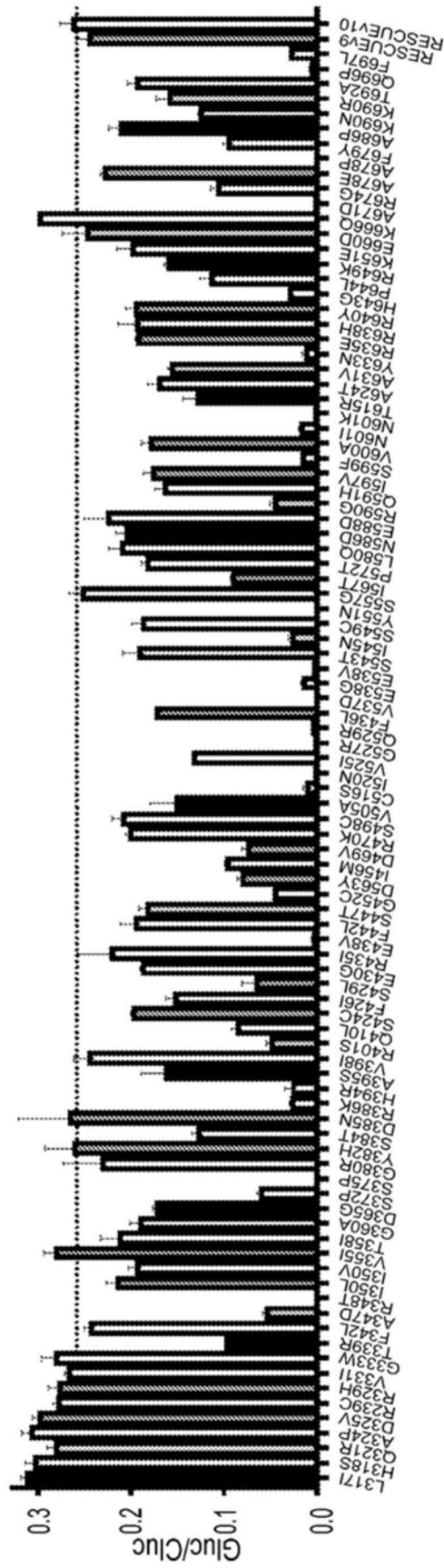


图42D

第1天第61轮

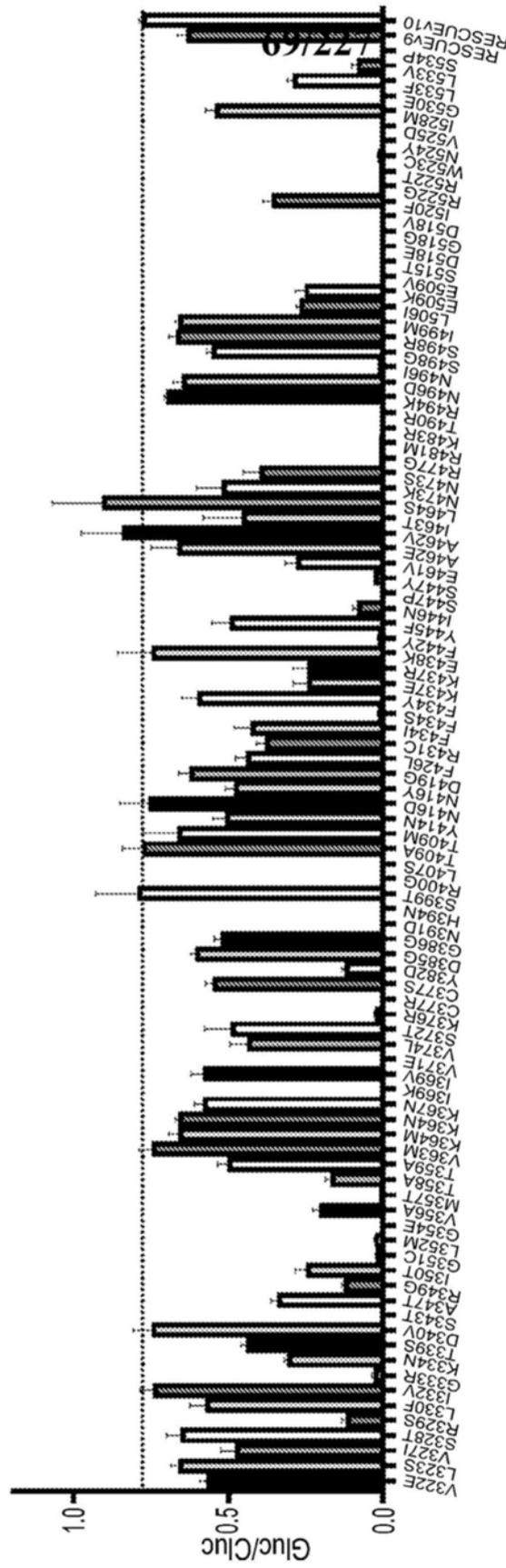


图42E

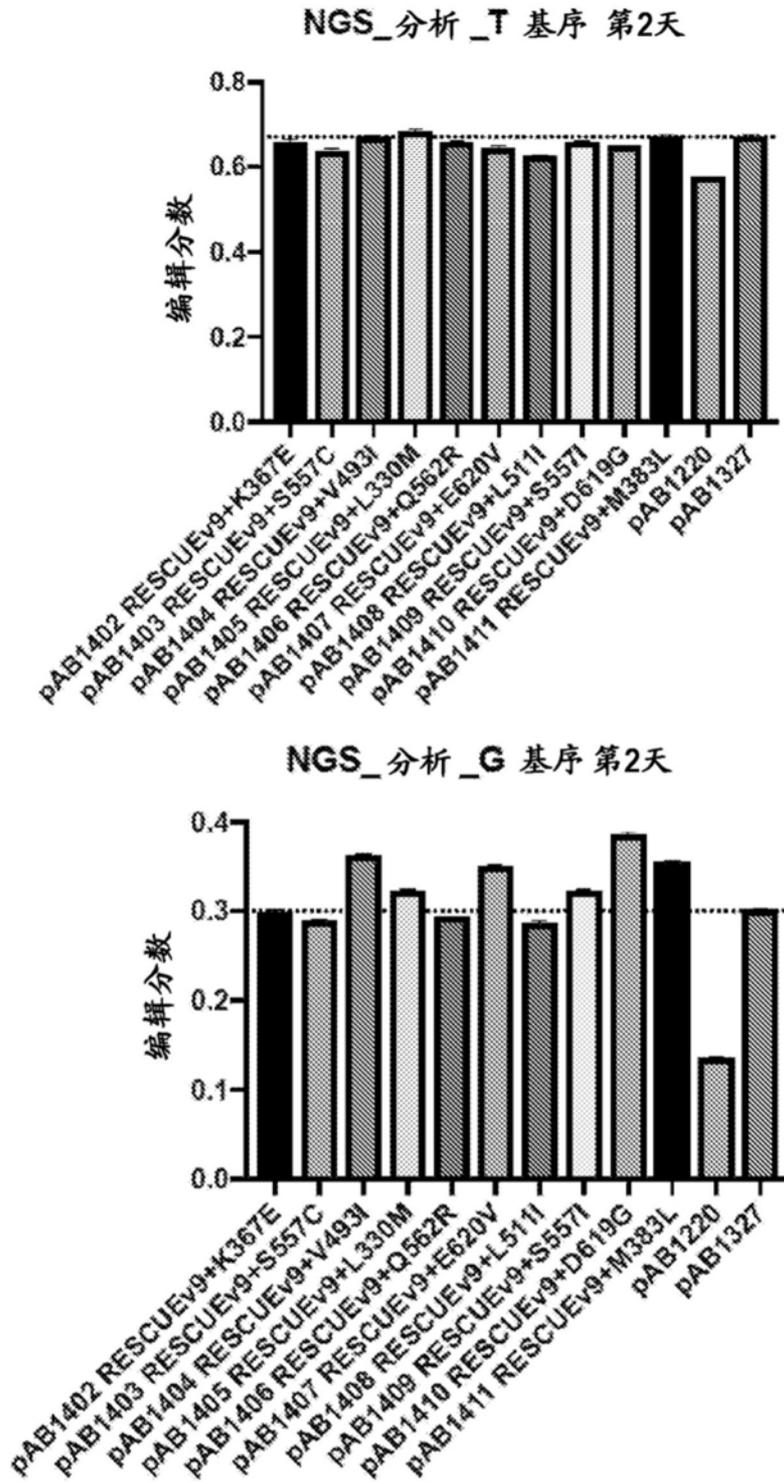


图43

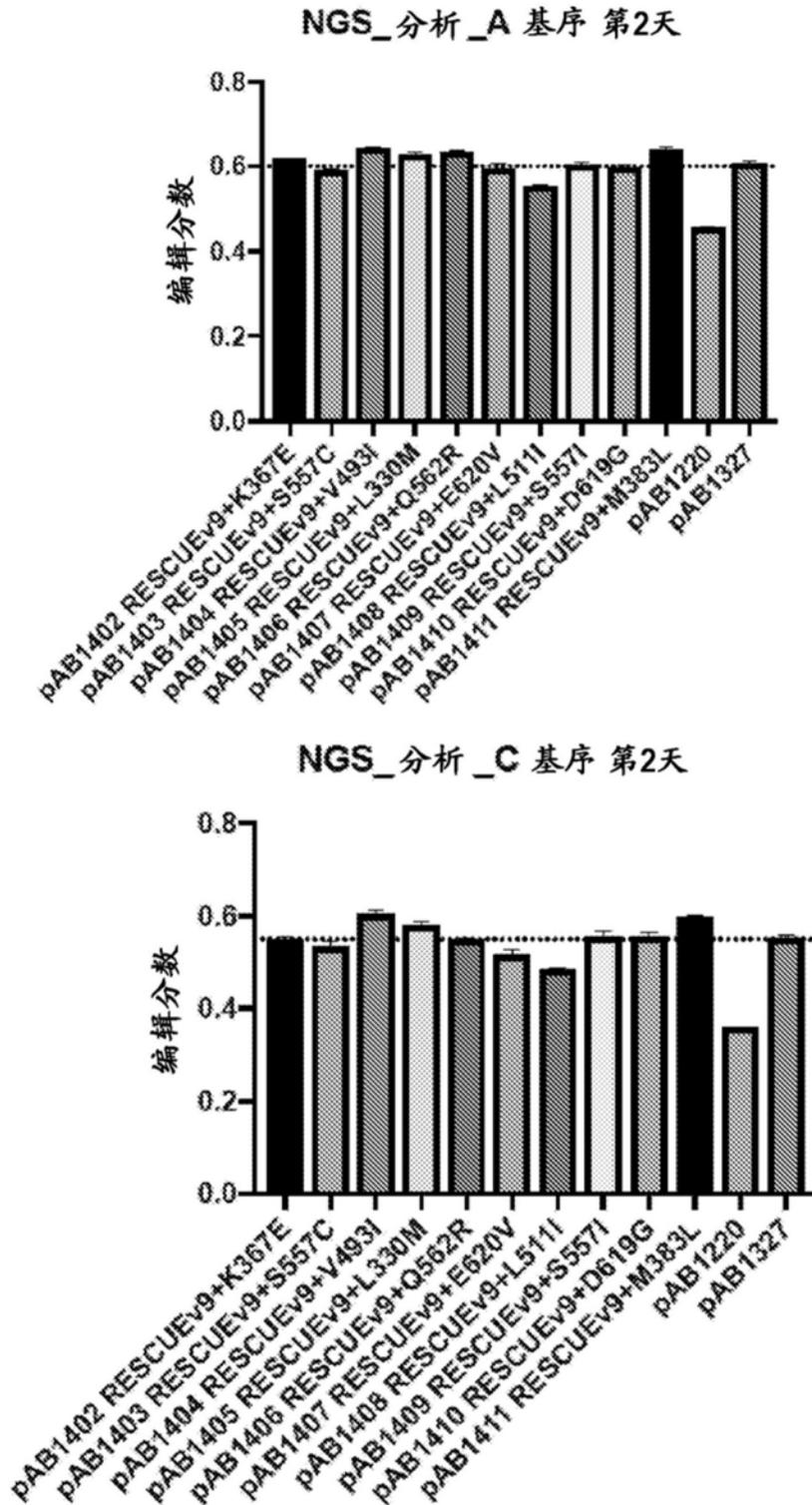


图43 (续)

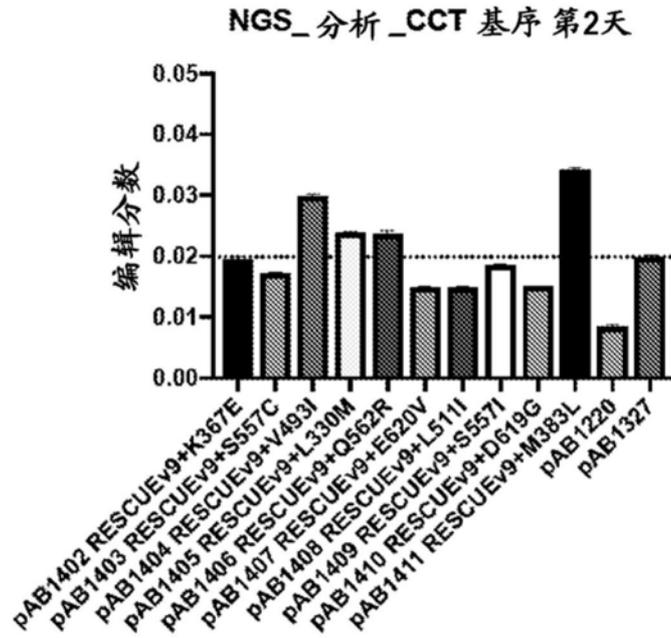


图43 (续)

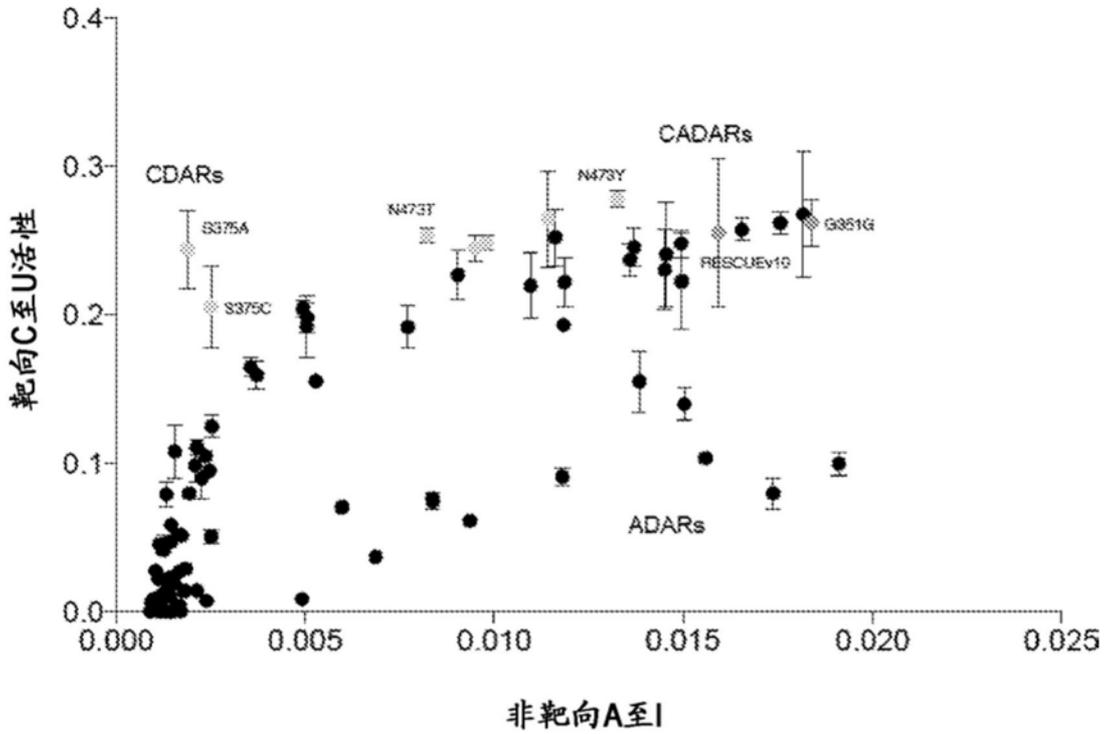


图44

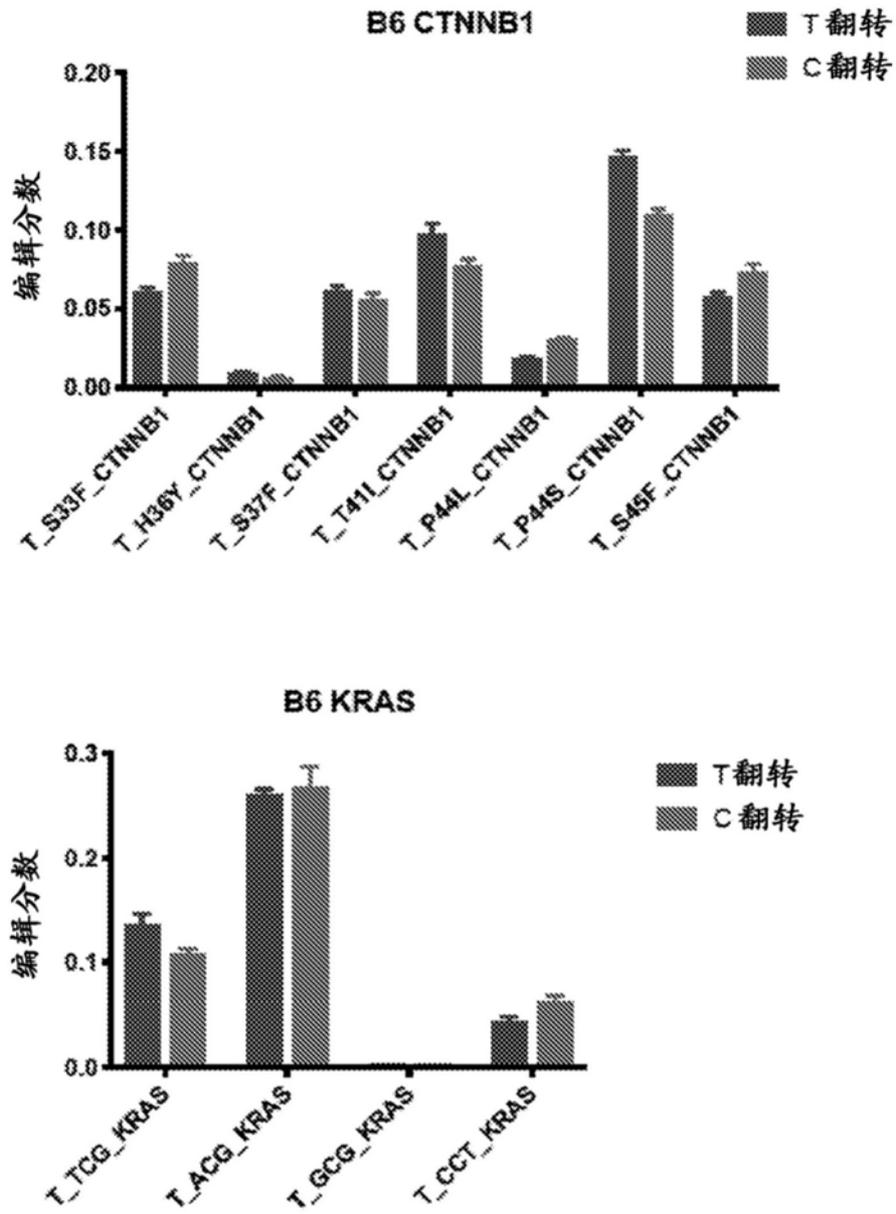


图45

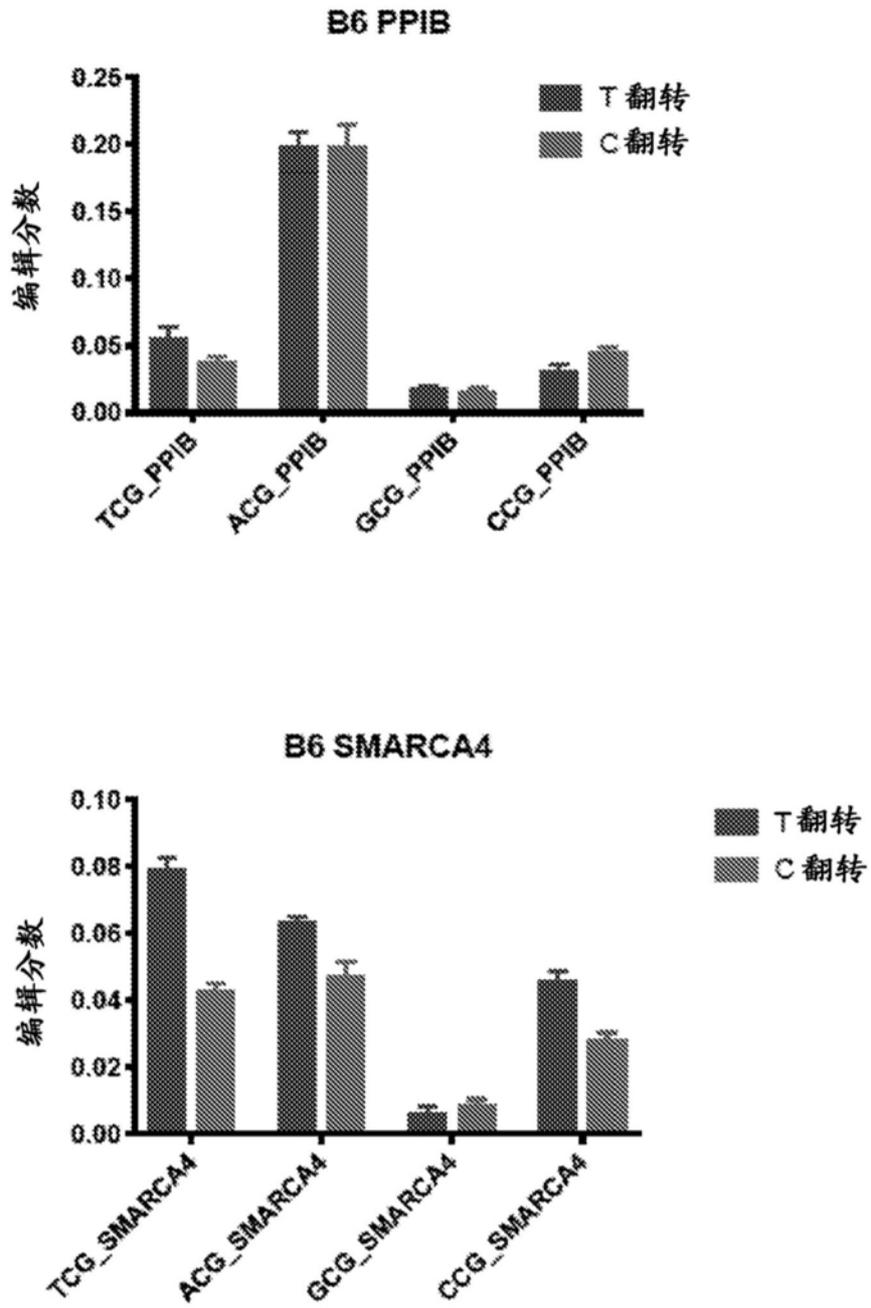


图45 (续)

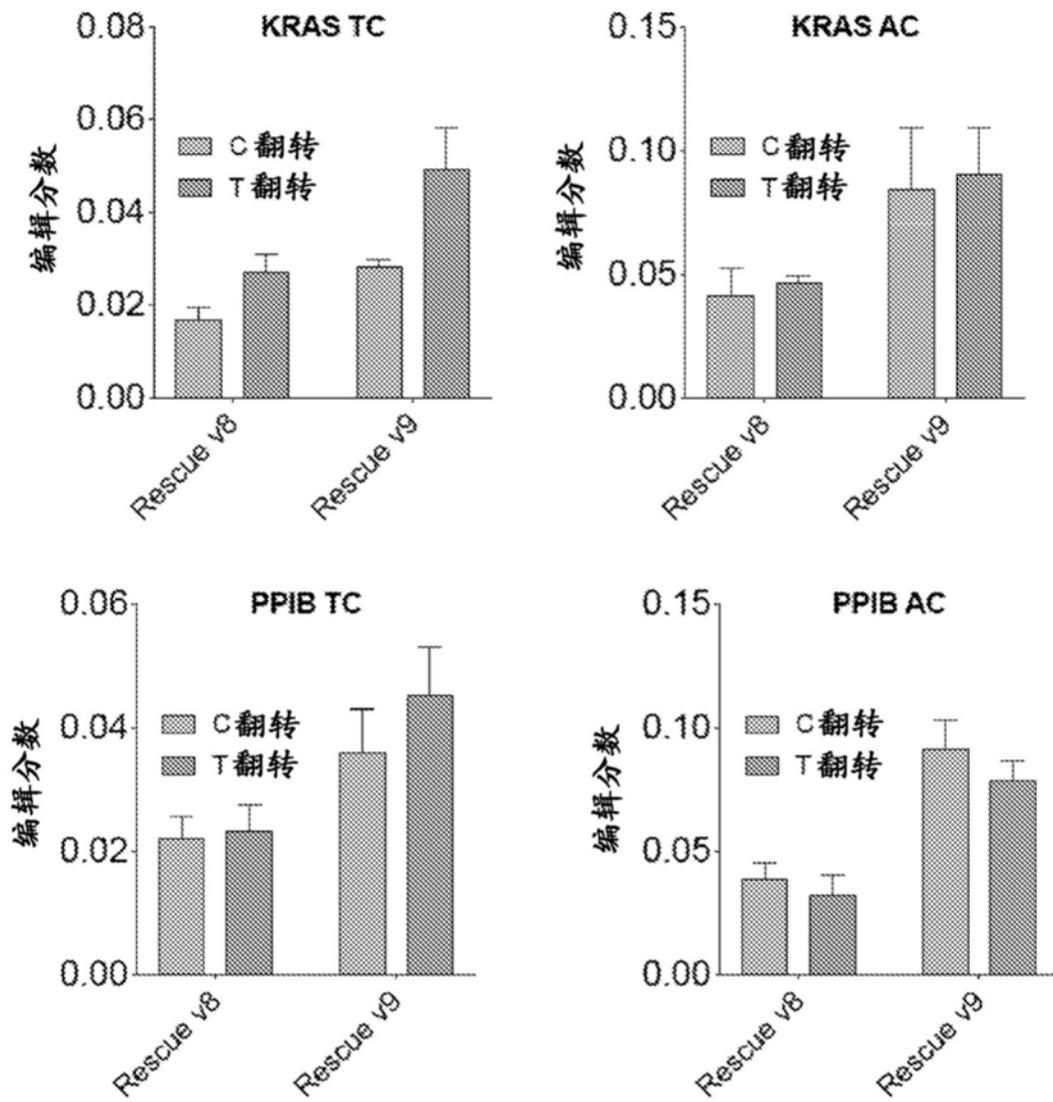


图45 (续)

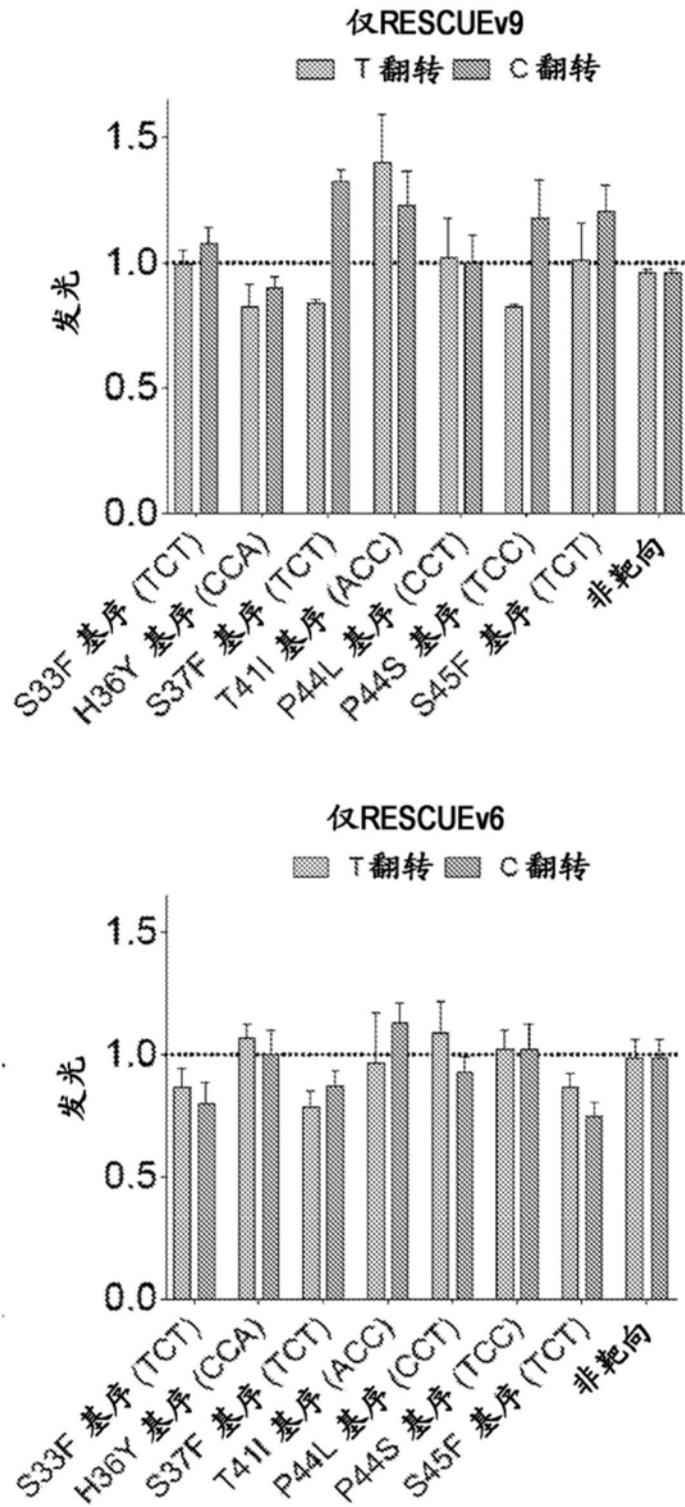


图46

|     |        |
|-----|--------|
|     | 编辑率    |
| R平方 | 0.5095 |

RESCUEv9编辑相对于荧光素酶C翻转

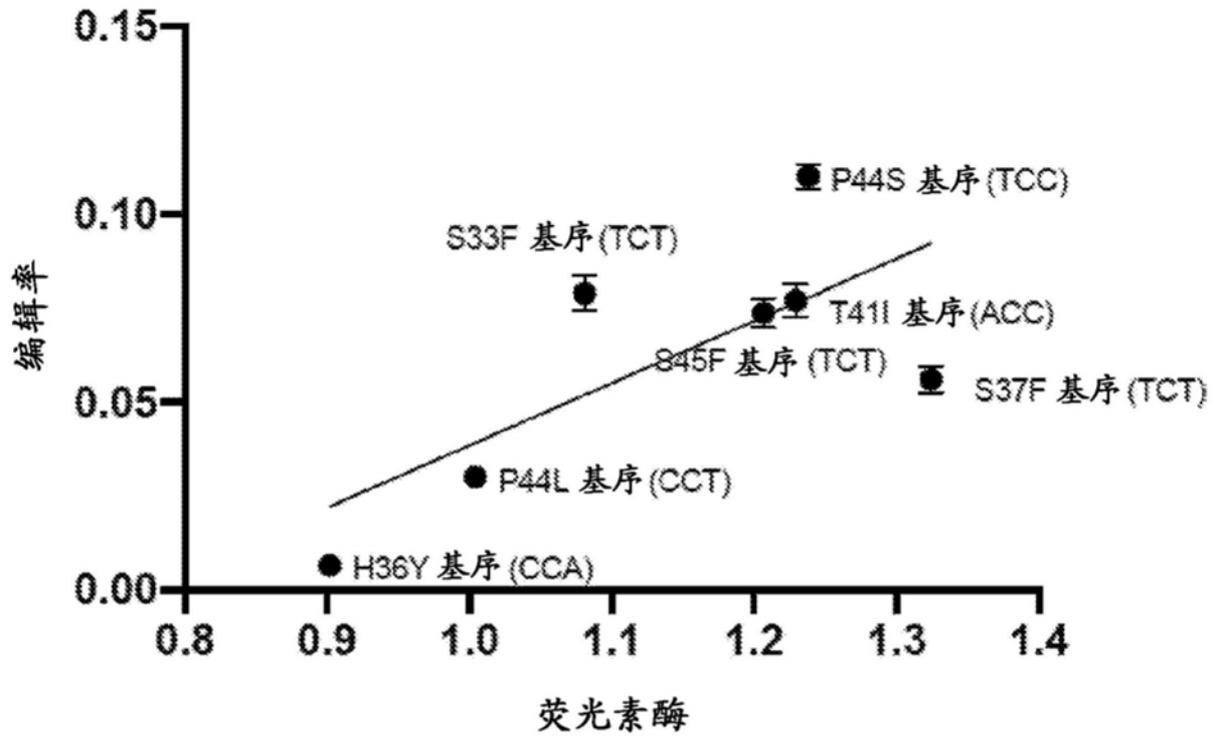


图46 (续)

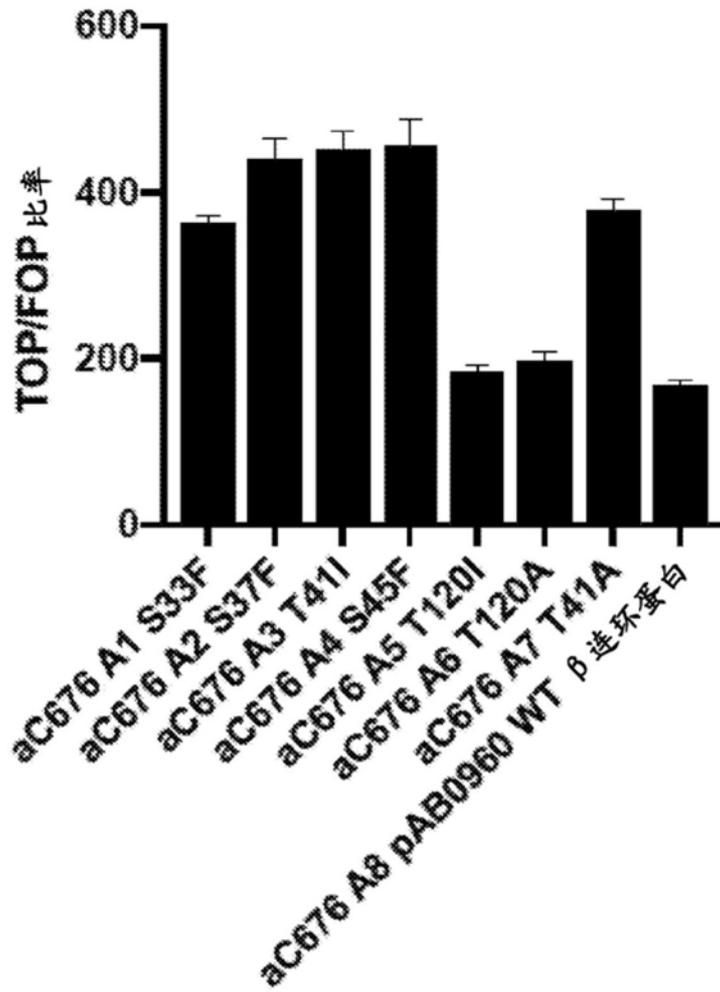


图47

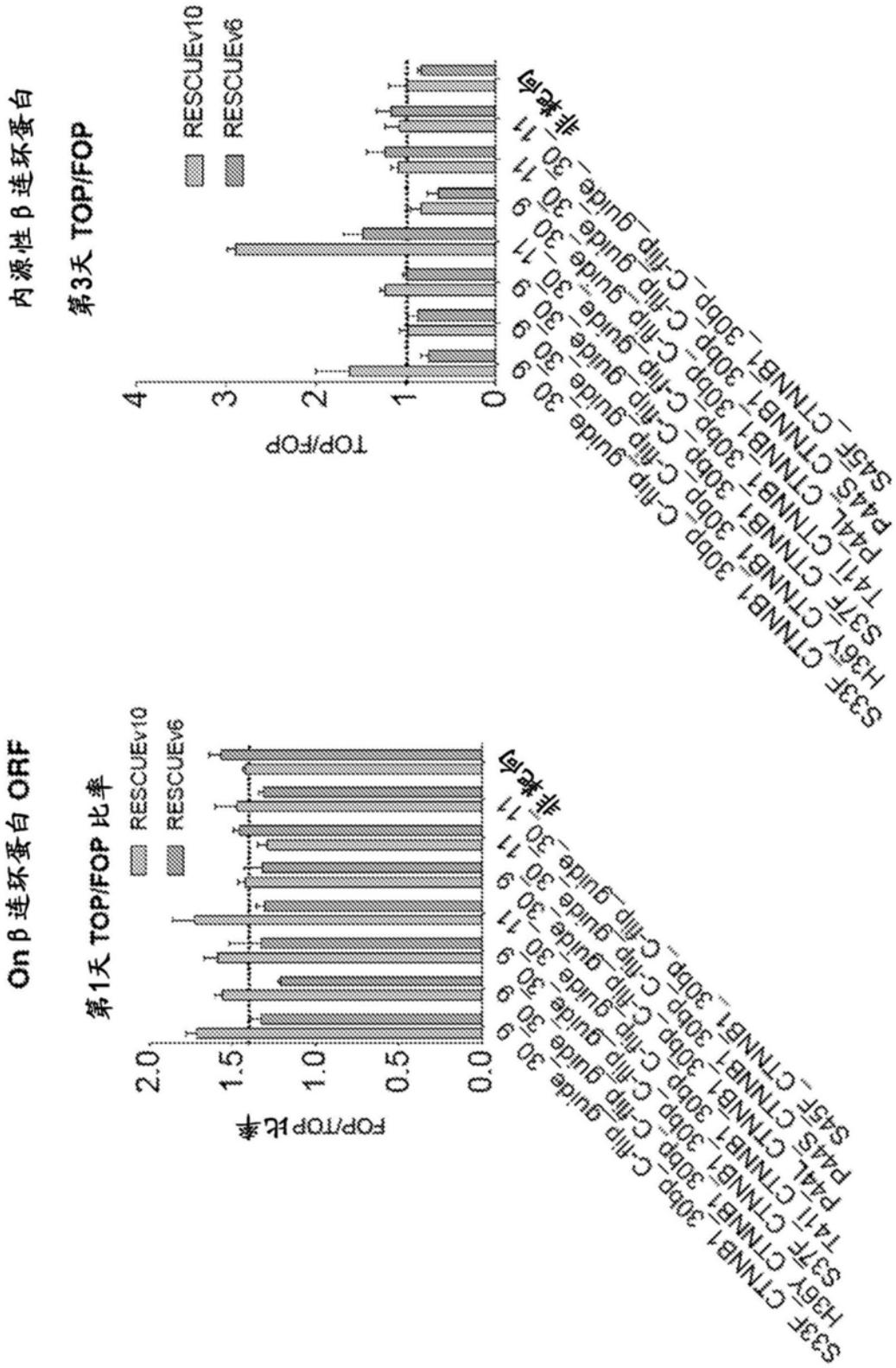


图48

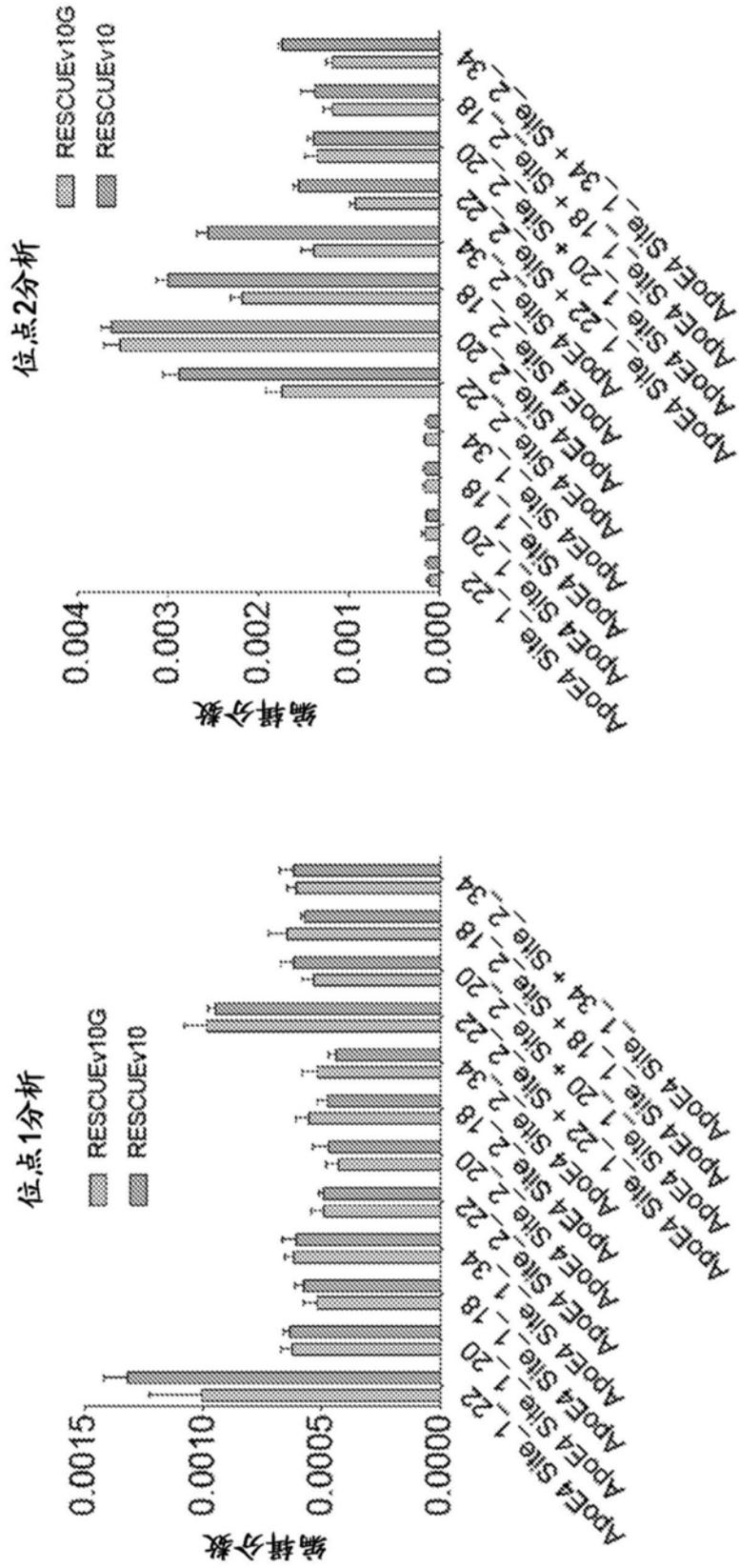


图49

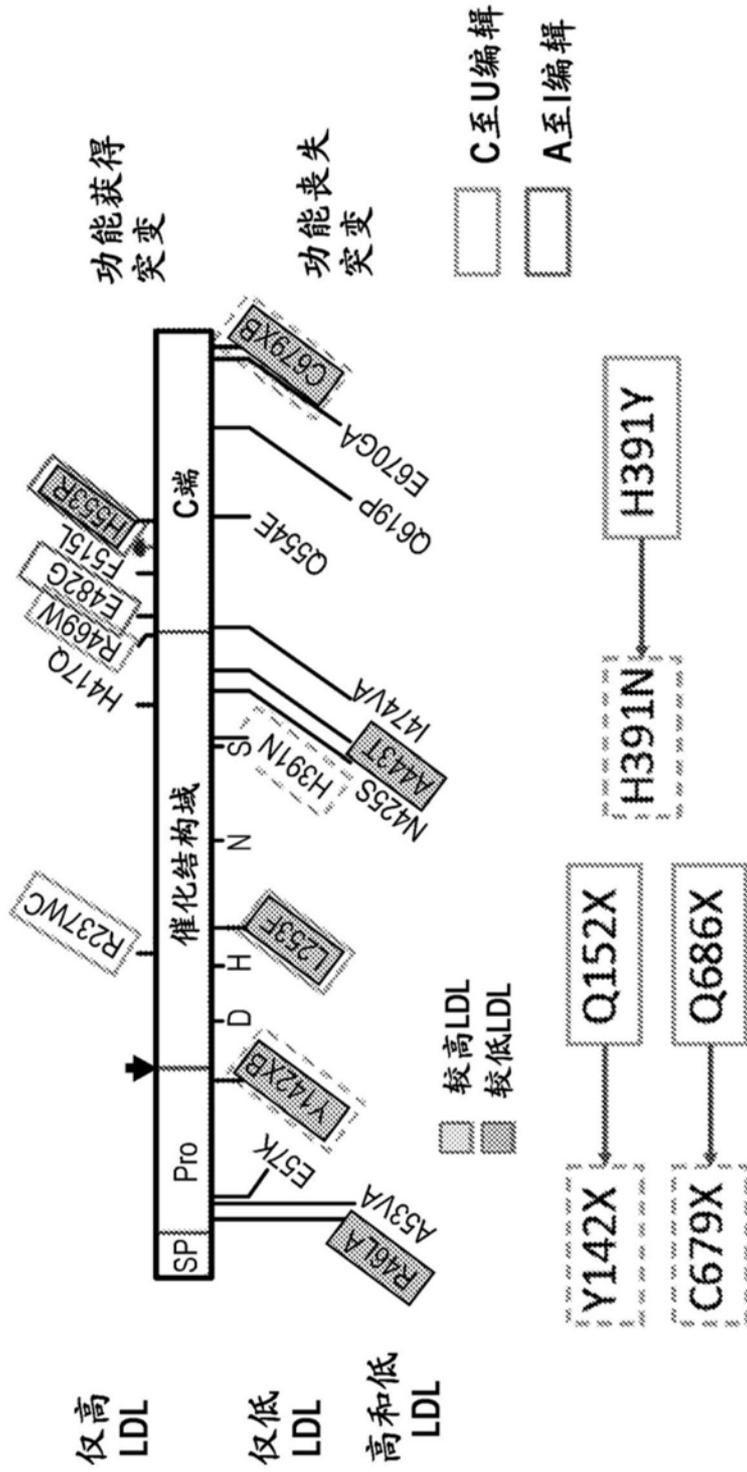


图50

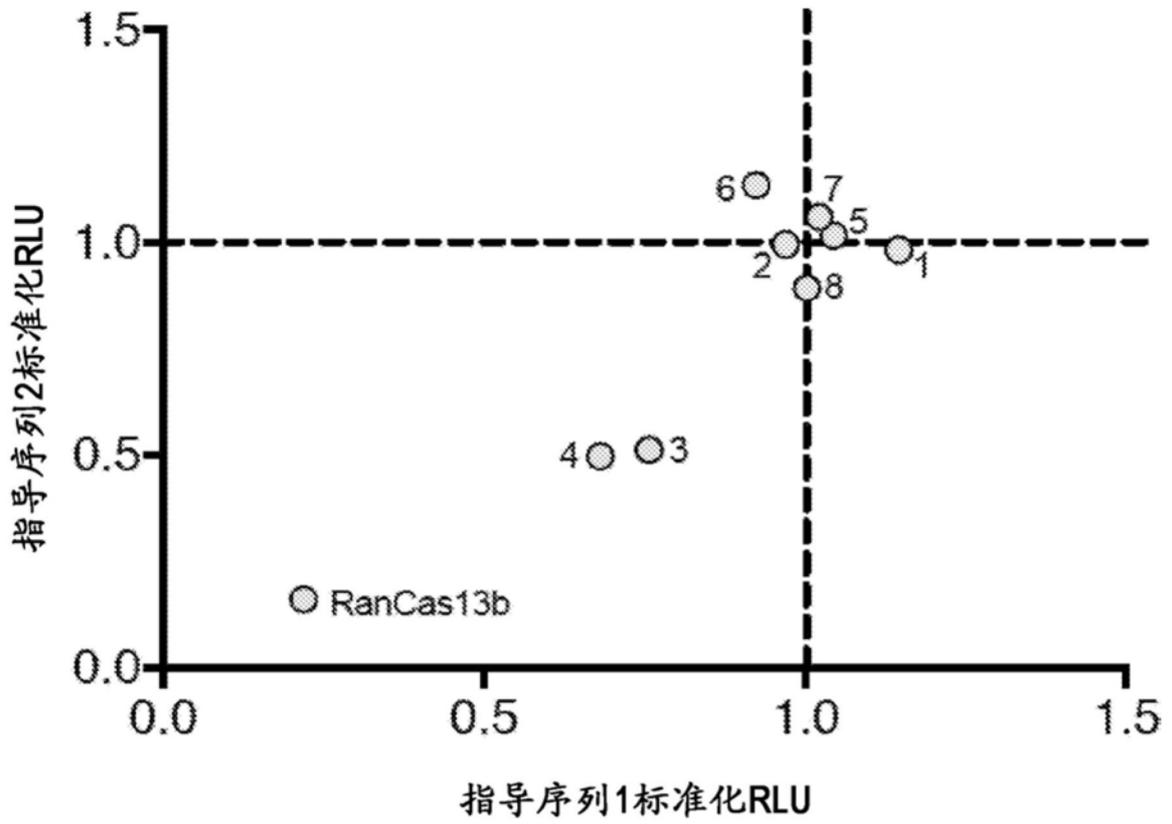


图51

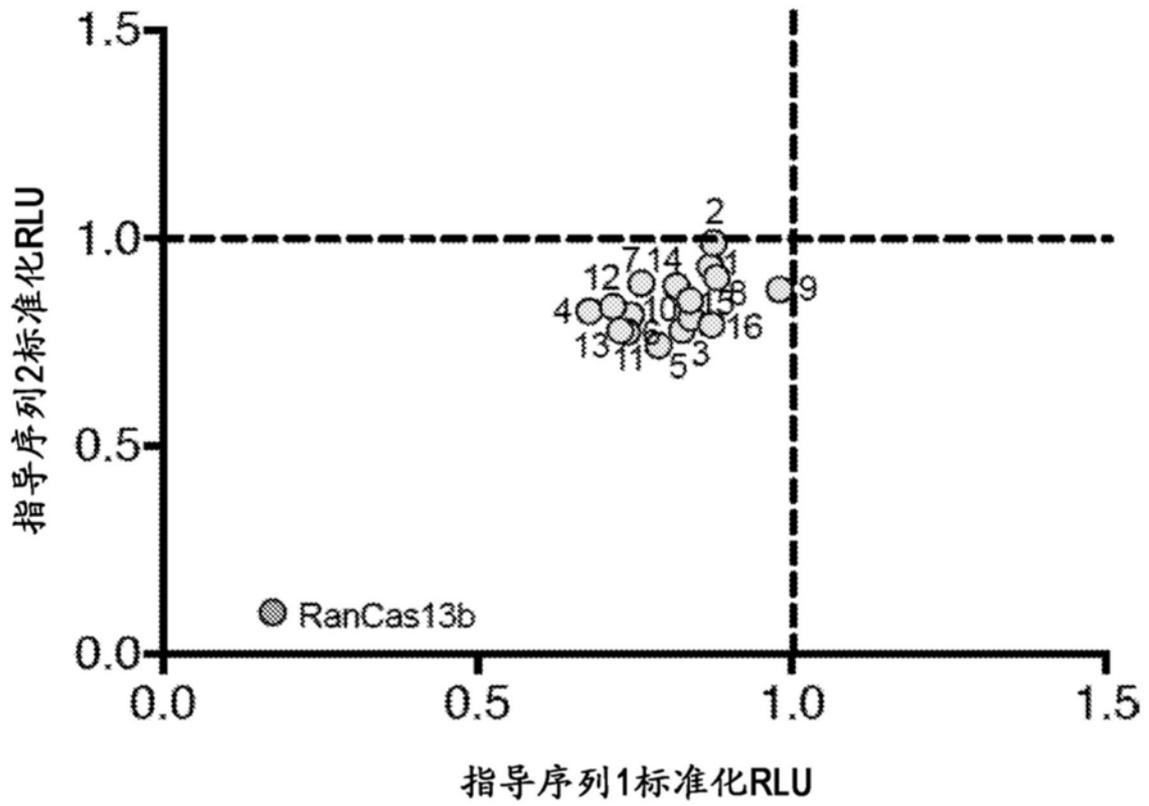


图52

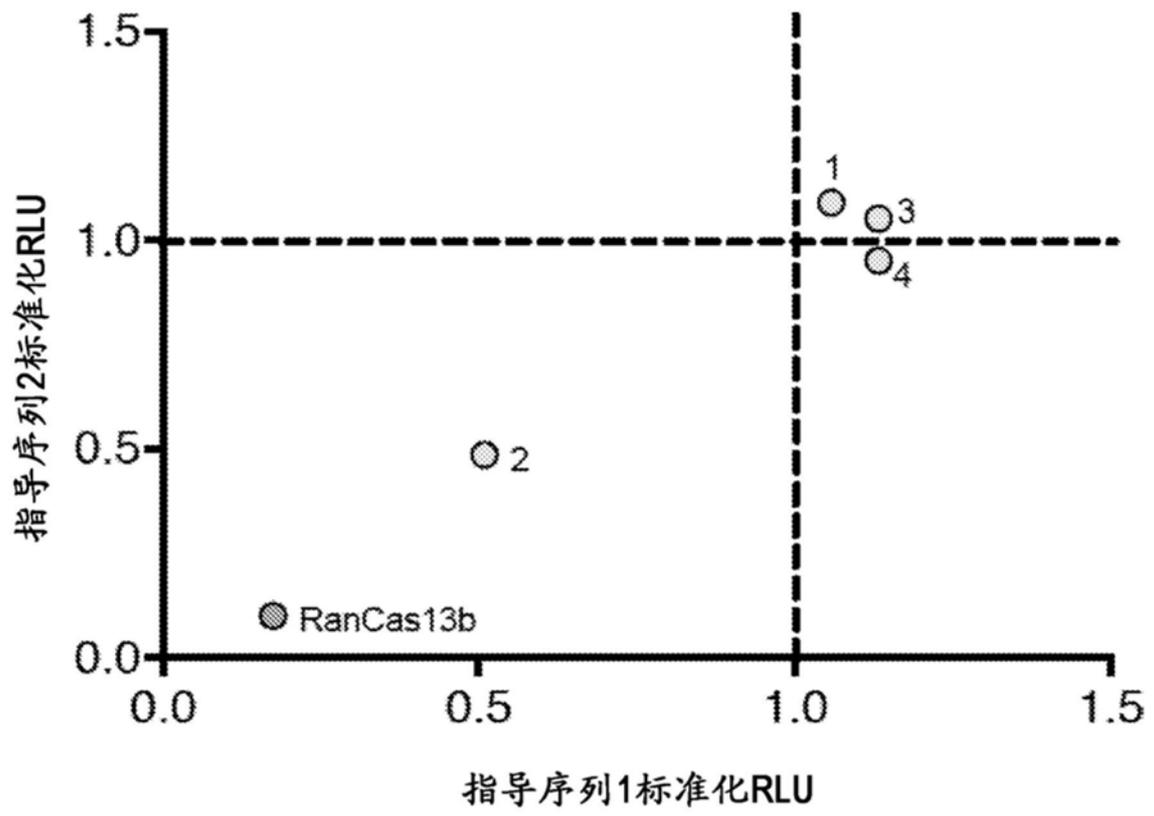


图53

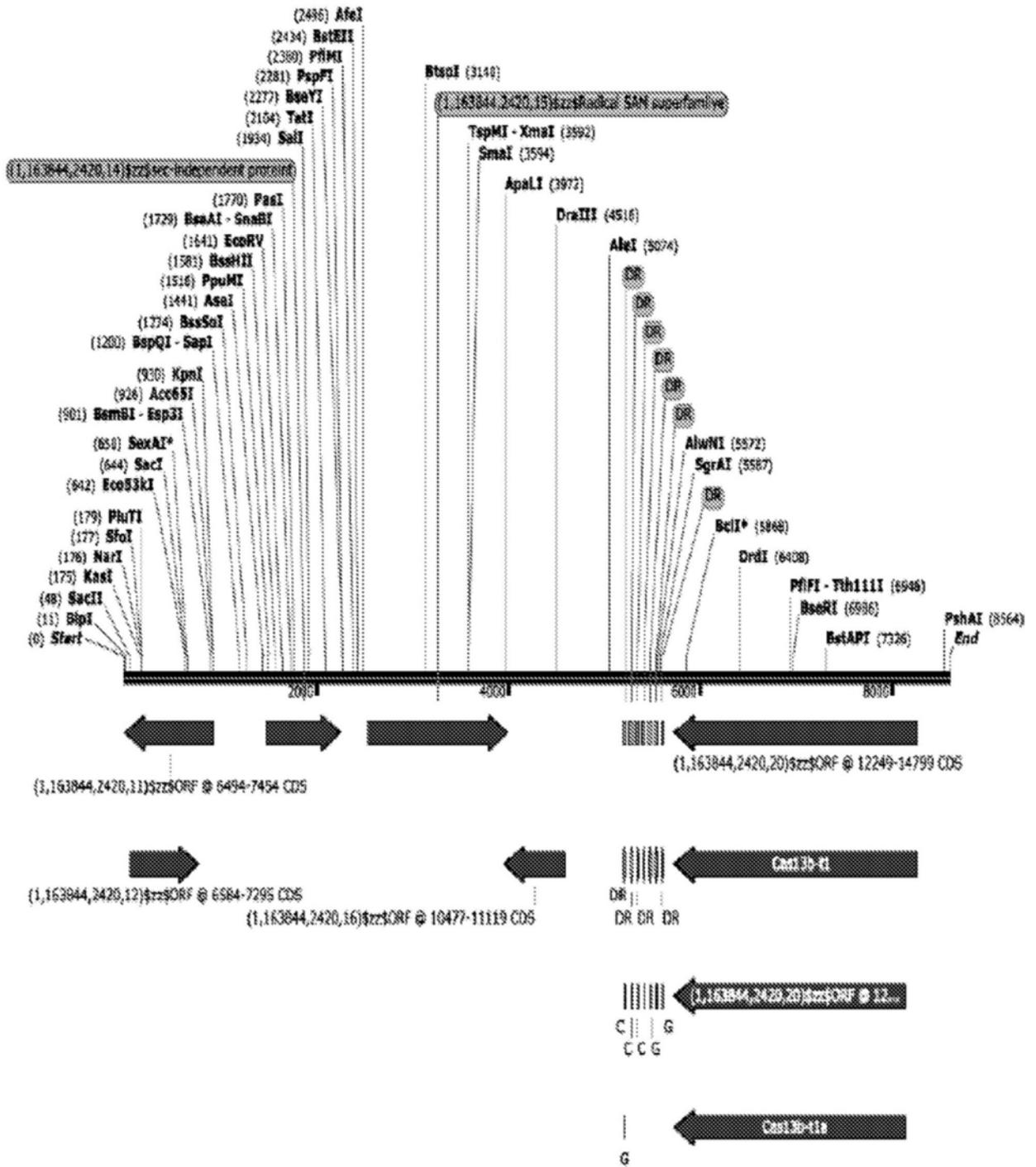


图54A





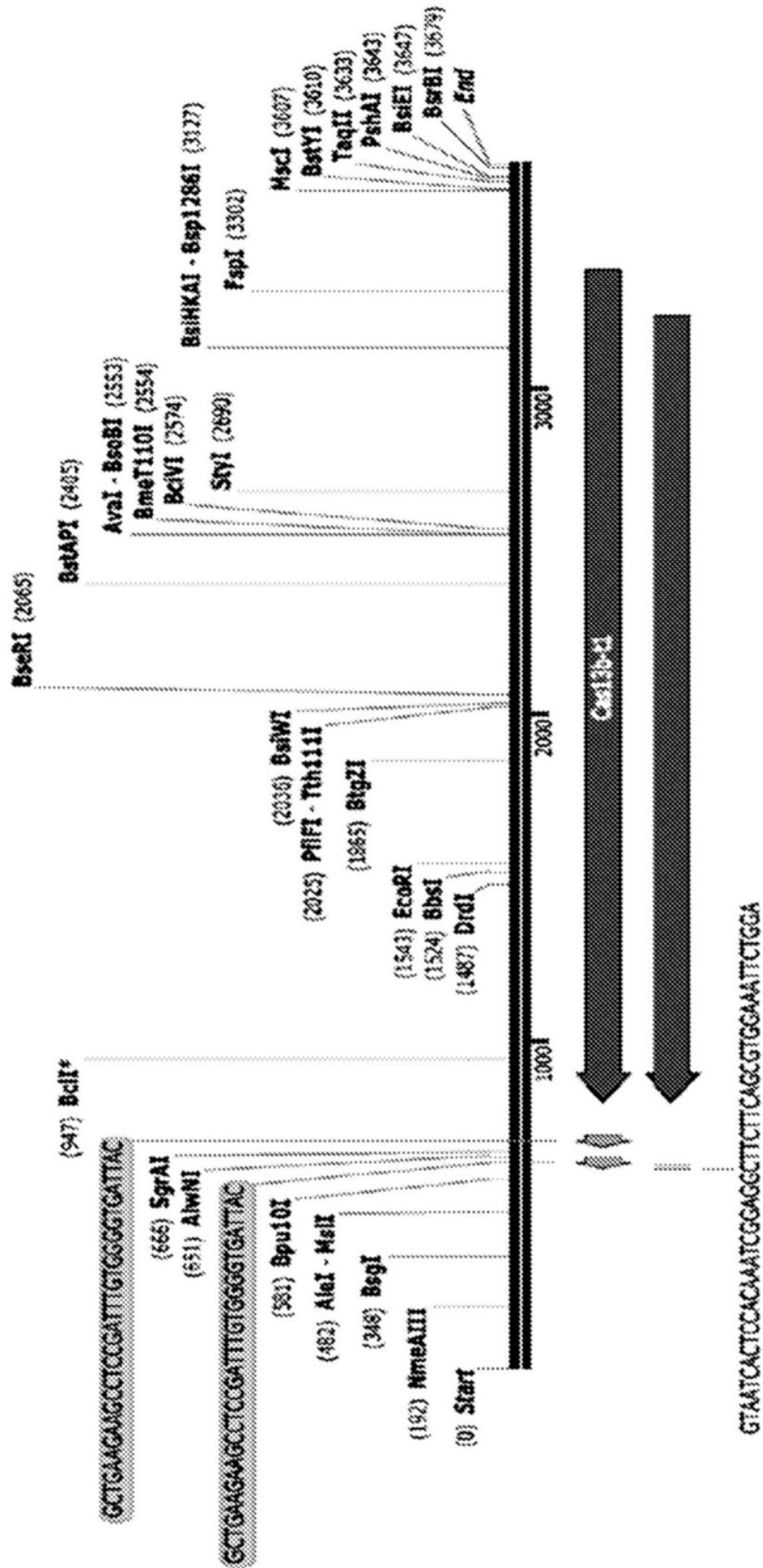


图55A



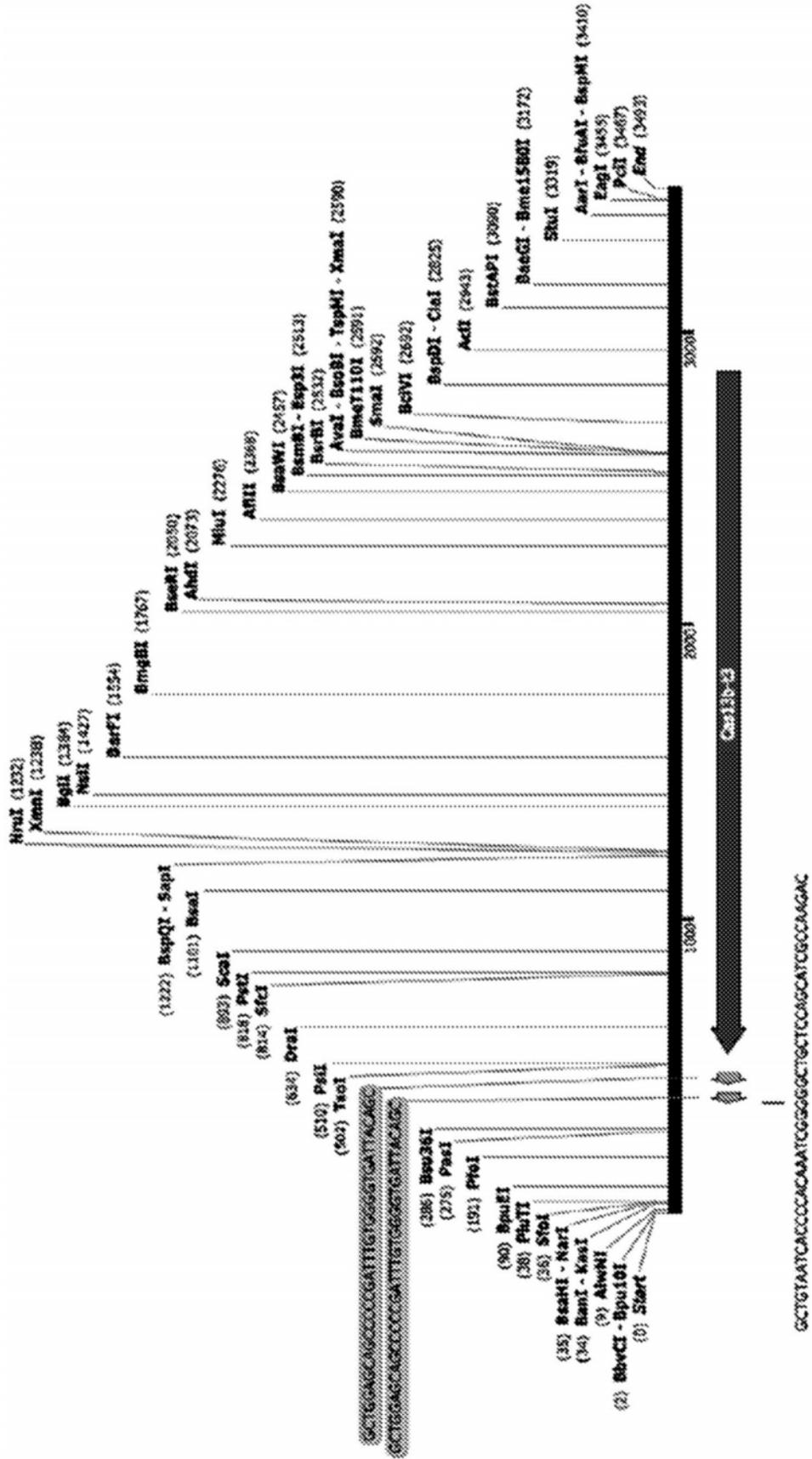


图55C

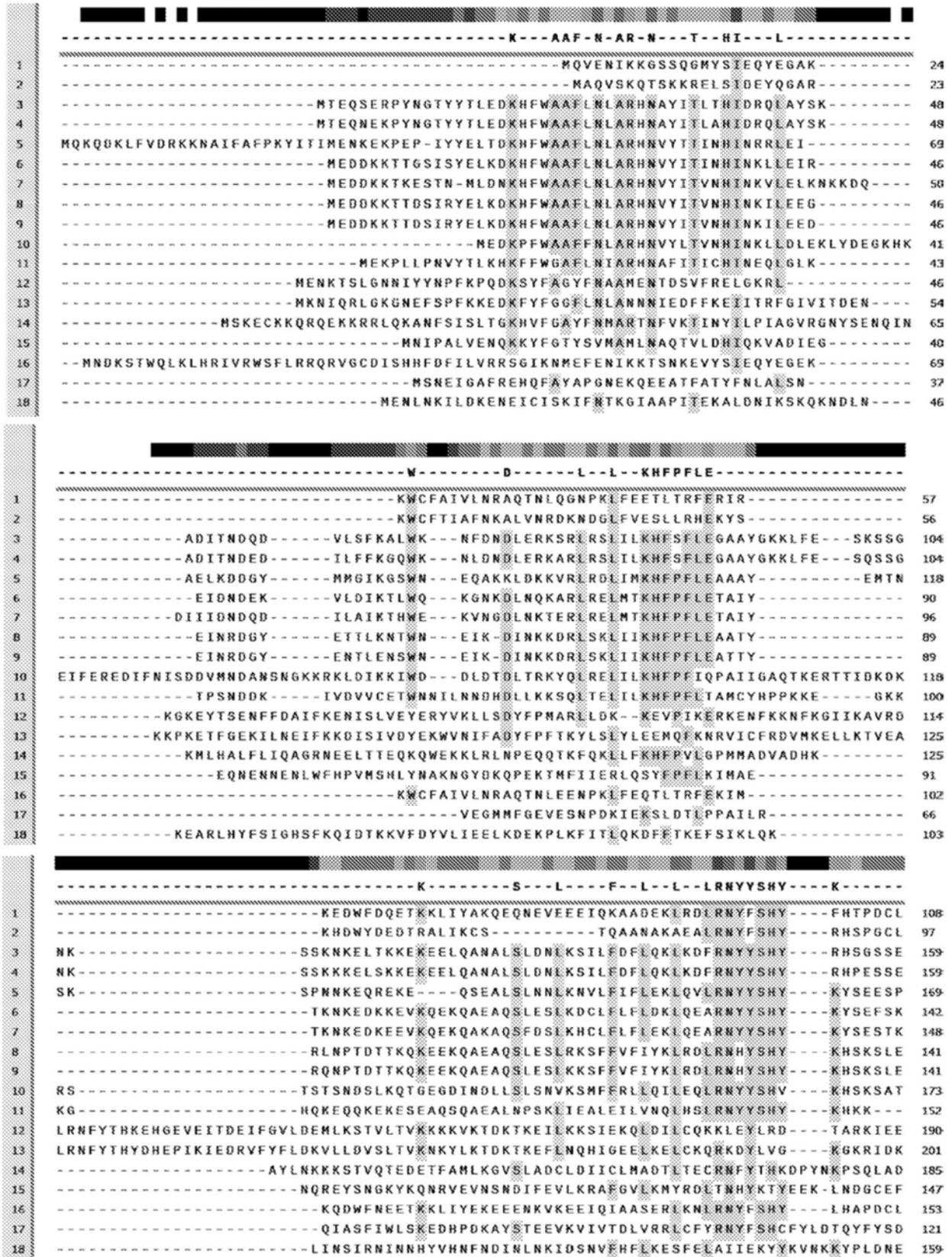


图56

| -----F-----D-----I-----F-HL-RK-----            |   |
|--|---|
| 1  | I FTQNDPVRIIMEKAYEKARFEQA-KKEQEDISIEFGELFEENGR 152                                      |
| 2  | TFTAEDELRTIMERAYERAIFECR-RRE-TEVIIIEFPSLFE-GDR 139                                      |
| 3  | LPLFDGNMLQRLYNVFDVSVQRVK-IDHEHNDEVDPHYHFNHLVRKGGKDRYG-----HNDNP 216                     |
| 4  | LPLFDGNMLQRLYNVFDVSVQRVK-RDHEHNKVDPHRHFNHLVRKGGKDKYG-----NNDNP 216                      |
| 5  | KPIFETSLLNMYKVF DANVRLVK-RDYMHHENIDMQRDFTHLRNKKQVGRTK-----NIIDSP 227                    |
| 6  | EPEFEGLLEKMYNIFGNNIQLVI-NDYQHNKIDINPDEDKHLDR-----KGG 189                                |
| 7  | EPMLEKELLKMYNIFDDNIQLVI-KDYQHNKIDINPDEDKHLDR-----EEE 196                                |
| 8  | RPKFEGLLEKMYNIFNASIRLVK-EDYQYNKIDINPDEDKHLORT-----EEE 189                               |
| 9  | RPKFEEDLQNKMYNIFDVSIIQFVK-EDYKHNIDINPKKDFKHLDRK-----RKG 189                             |
| 10   | MPNFDEDLLNWMRYIFIDSVNKVK-EDYSSNSVIDPNTSFSLIYKDEQG-----KIK 225                           |
| 11   | -PDAEKDIFKHLYKAFDASIRMVK-EDYKAHFTVNLTRDFAHNRKKGKNKQ-----DNP 204                         |
| 12   | KR--RNQRERGEKELVAPFKYSD-KRDDLIAAIYNDADFVYIDKKKDS-----LKES 239                           |
| 13   | ESEIINGIYNNAFKDFICKREKQD-DKENHNSVEKILCNKEPQNKQKSSATV-----WELCSK 259                     |
| 14   | QYLHQEMIAKKLDKVVVASRRILKDREGLSVNEVEFLTGDHHLHQEVLEKDFEGNAKVKDGGKVMKTFVEYDDFFYFKISGKR 265 |
| 15   | LTSTEQPLSGMINNYTYVALRNMNERYGYKTEGLAFIQDKRKFVKDAYGKKK-----SQV 203                        |
| 16   | IFNRNDTIRIIMEKAYEKSRFEAK-KKQQEDISIEFPELFEEEDK----- 197                                  |
| 17   | ELVDTTAIGEKLPPYFHHFITNRLFRYSLPEITLFR--WNEGERK----- 164                                  |
| 18   | IVLFLKELFIKQENTALLNYFTNLSKDEAIEYILTFTITENKIWNIN-----N 207                               |
| -----IT--GLLFFVSLFLEK--A-----K--GFK-----T--    |   |
| 1  | -----ITSAGVVFVFFASFFAERRFLNRLMGVYQGFTRTEGEYK-----ITR 192                                |
| 2  | -----ITTAGVVFVFFVFFVERVLDRLYGAVSGLKKNEGQYK-----LTR 179                                  |
| 3  | SFKHHFVDGE-----GMVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKGGTETYQ-----QMTN 269                     |
| 4  | FFKHHFVDRE-----GTVTEAGLLFFVSLFLEKRDAIWMQKKIRGFKGGTEAYQ-----QMTN 269                     |
| 5  | NFHYHFADKE-----GNMTIAGLLFFVSLFLDKKDAIWMQKKLKGFKDGRNLR-----QMTN 280                      |
| 6  | FKYSFADNE-----GNITESGLLFFVSLFLEKDAIWMQKLNKGFKNLENKK-----KMTN 241                        |
| 7  | FNYYFTTNKK-----GNITASGLLFFVSLFLEKDAIWMQKLRGFKDNRESKK-----KMTN 249                       |
| 8  | FNYYFTKQNE-----GNITESGLLFFVSLFLEKDAIWMQKLRGFKDNRENKK-----KMTN 242                       |
| 9  | KFHYSFADNE-----GNITESGLLFFVSLFLEKDAIWMQKLEGFKCSNKSQY-----KMTN 242                       |
| 10   | PCRYPFTSKD-----GSINAFGLLFFVSLFLEKQDSIWNQKKIPGFKKASENYM-----KMTN 278                     |
| 11   | DFNRRYFEKD-----GFFTESGLLFFFTNLFLDKRDAIWMKLVVSGFKASHKQRE-----KMTT 257                    |
| 12   | SKAKYNTKSDPQQEEGD--LKIPISKNGVVFLLSIFLTKQEIHAFKSXIAGFKATVIDEATVSEATVSHGKNSICFNATH 317    |
| 13   | SSKYTEKSFNRENDKHCEVPISQKIVFLLSFFLNKGEIYALTSNKGFKAKITKEE----PVTYDKNSIRYMAH 334           |
| 14   | LVNGYTVTTKDDKPVNVNTMLPALSDFGLLYEFCVIFLSKPYAKLFIDEVRLFYSPFDK-----ENMMS 331               |
| 15   | NTGFFLSLQDYG--DTQKHLHLSGVGIALLICLFLDKQYINIFLSRLPIFSSYNAQSE-----ERRII 266                |
| 16   | -----ITSAGVVFVFFVFFIERRFLNRLMGVYQGFTRKTEGEYN-----ITR 237                                |
| 17   | -----YEILRDSLIFFCCLFLKRGQERFLNELRFFKRTDEEGR-----IKR 206                                 |
| 18   | EHNILNIEKG-----KYLTFEAMLELITIFLYKNEANHLLPKLYDFKMKNSKQE-----LF 258                       |
| EVF-----P--K-----LDML-EL-RCP--LY--L-----F----- |   |
| 1  | DVFSTYCLRDSYSVKTP-----DHDVAMFRDILGYLSRVPSSESYQRIKESQ--MRSETQ--LSE 247                   |
| 2  | KALSNYCLKDSRFTKAW-----DKRVLLFRDILAQLGRIPAEAYEYHGEQGDKKRANDNEGTPK 240                    |
| 3  | EVFCRSRISLP--KLKLES-----LRMDDWMLLDMLNELVRCPPKLYDRLEDDRACFRVPVDILPDED 331                |
| 4  | EVFCRSRISLP--KLKLES-----LRTDDWMLLDMLNELVRCPKSLYDRLEEDRARFRVPVDILSDED 331                |
| 5  | EVFCRSRISLP--KLKLEN-----VQTKDWMQLDMLNELVRCPKSLYERLREKDRSEFKVPFDIFSDDY 342               |
| 6  | EVFCRSRILMP--KLRLES-----TQTQDWILLDMLNELIRCPKSLYERLQGDREKFKVPFDPADEDY 303                |
| 7  | EVFCRSRMLLP--KLRLES-----TQTQDWILLDMLNELIRCPKSLYERLQGEYRKKFNVPFOSADEY 311                |
| 8  | EVFCRSRMLLP--KLRLQS-----TQTQDWILLDMLNELIRCPKSLYERLREEDREKFRVPIEIAEDY 304                |
| 9  | EVFCRSRMLLP--KLRLES-----TQTQDWILLDMLNELIRCPKSLYERLQGVNRKKFYVSDPADEDY 304                |
| 10   | EVFCRNHILLP--KIRLET-----VYDKDWMMLDMLNEVVRCPKSLYKRLTPAAQNKFKVPEKSSDNAN 340               |
| 11   | EVFCRSRILLP--KLRLES-----RYDHNQMLLDMLNELSRCPKLIYEKLSSEENKHKFQVEADGFLDEI 319              |
| 12   | EIFSHLAYKKL--KRKVRTAEINYGEAENAEQLSVYAKETLMMQMLDELSKVPDVVYQNLSEDVQKTFIEDWNEYLKEN 394     |
| 13   | RMFSFLAYKGL--KRKIRTSEINY--NEDGQASSTYEKETLMLQMLDELNKPDPVVYQNLSEDVQKTFIEDWNEYLKEN 409     |
| 14   | EMLSIYRIRTPRLHKIDS--HDSKATLAMDIFGILRRCPMELYNLIDKNAGQPFHDEVKHPNSH 394                    |
| 15   | RSFGINSIKLP--KDRHS-----EKSNSVAMDMLEVYKRCPELFTTLSAEKQSRERISDDHNEVL 328                   |
| 16   | QVFSKYCLKDSYSVQAQ-----DHDVAMFRDILGYLSRVPTIYQHIKLTR--KRSQDQ--LSE 292                     |
| 17   | TIFTKYCTRES--HKHIGIE--EQDFLIFQDIIGDINRVPKVCDGGVVDLSKENERYIKNRETSNES 268                 |
| 18   | TFSKKFTSQDIDAEEG-----HLIKFRDNIQYLNHYPTAWNNDLKLESENKKNIMTKLIDSI 317                      |

图56(续)

|    |   |     |
|----|---|-----|
|    | -----F--L-R-----DRFPYFALRY-D-----F--LRFQ--LQ-YH--Y-K--I-----RH-----L-GF-R         |     |
| 1  | RKT-----DKFILFALNYLEDYGLEDLADYTACFARTRIKREQD-----ENTDQKEQKPHRKK-----              | 300 |
| 2  | RHK-----DKFIEFALHYLEAQ-----HSEICFGRRHIVREEAG-----AGDEHKHRTK-----                  | 285 |
| 3  | DTDG--GGEDPFKNTLVR--HQDRFPYFALRYFDLKKVFTSLRFHIDIGTYHFAIYKKN--IGEQPEDRHLTRNIYGFGR  | 405 |
| 4  | DTDG--TEEDPFKNTLVR--HQDRFPYFALRYFDLKKVFTSLRFHIDIGTYHFAIYKKN--IGEQPEDRHLTRNIYGFGR  | 405 |
| 5  | NAE----EPPFKNTLVR--HQDRFPYFVLRVFDLNEIFEQLRFQIDLGTYHFSIYNKR--IGDEDEVRLHLLHLYGFAR   | 413 |
| 6  | NAE----QEPFKNTLIR--HQDRFPYFVLRVFDYNEIFKNLRFQIDLGTYHFSIYKKN--IGGQKEDRHLTHKLYGFER   | 374 |
| 7  | DAE----QEPFKNTLVR--HQDRFPYFALRYFDYNEIFNLRVQIDLGTYHFSIYKKN--IGGQKEDRHLTHKLYGFER    | 382 |
| 8  | DAE----QEPFKNTLVR--HQDRFPYFALRYFDYNEIFNLRVQIDLGTYHFSIYKKN--IGDYKESHLLTHKLYGFER    | 375 |
| 9  | DAE----QEPFKNTLVR--HQDRFPYFALRYFDYNEVFANLRVQIDLGTYHFSIYKKN--IGGQKEDRHLTHKLYGFER   | 375 |
| 10 | RQE----DDNPFSSRIIVR--HQDRFPYFVLRVFDLNEVFNLRVQIDLGTYHFAICKKQ--IGDKKEVHLLIRTLYGFGR  | 412 |
| 11 | EEE----QNPFKDTLIR--HQDRFPYFALRYDLNESFKSIRFQVLDGTYHYCIYOKK--IGDEQEKRLHTRTLISLGR    | 390 |
| 12 | NGDVGTMEEEQVIHPVIRKRYEDKFNFAIRFLDEFAQFPTLRFQVHLGNYLHDSRPKE---NLISDRRIKEKITYVGR    | 470 |
| 13 | NGDVGTMEEEQVIHPVIRKRYEDKFNFAIRFLDEFAQFPTLRFQVHLGNYLHDSRPKE---ICDITTEVEYKKKITYVGR  | 487 |
| 14 | TPD-----VSKRLRYDDRFPTLALRYIDETELFKRIRFQLQSGFRYKFKYOKENCIDGRVVRRIQKEINGYGR         | 463 |
| 15 | MKR-----SSDRFVPLLQYIDYQKGLDHIRFVHMGGKRLRYLLKADKTCIDGQTRVRIEQPIINGEGR              | 391 |
| 16 | RKT-----DKFILFALKYLEDYGLKDLADYTACFARSKIKRENED--TKETDGNKHKFHREK-----               | 347 |
| 17 | DEN----KARYRLIR--EKDKFPYYLMRYIVDFGLPCITFKQNDYSTKEGRGQFHYQDAVAQEERCYNFVVRNG-       | 339 |
| 18 | IEFELNSNYPSPFATDIQFKKEAKAELFASNKKRNQTSFSNKSYNNEIRHNPNIKQYRDEIASALTPISFNVKEDKFKIFV | 397 |
|    | -----Q-----E-----PYI--T-PHY--E--KIG-----  |     |
| 1  | -----PRVEIHFERAEGDP-----FYIKHNNVILRTQ-----  | 327 |
| 2  | -----GKVVVDFSKDEDQS-----YYISKNNVIVRID-----  | 313 |
| 3  | IQDFAEEHRPEE---WKRLVRLDYFETGDK-----PYISQTSPHYHIKGGKIGLRFM---PEG                   | 458 |
| 4  | IQDFAEEHRPEE---WKRLVRLDYFETGDK-----PYITQITPHYHIKGGKIGLRFV---PEG                   | 458 |
| 5  | IQDFAPQNPQEE---WRKLVKDLDFHETSQE---PYISKIAPHYHLENEKIGIKFC---SAH                    | 466 |
| 6  | IQEFAKQNRPEE---WKAIVKDLDTYETSNNK-----RYISETTPHYHLENQKIGIRFR---NGN                 | 427 |
| 7  | IQEFAKQNRPEE---WKAIVKDFDTYETSEE---PYISETAPHYHLENQKIGIRFR---NDN                    | 435 |
| 8  | IQEFKQNRPEE---WRKFVKTNSFETSKE---PYIPEITPHYHLENQKIGIRFR---NDN                      | 428 |
| 9  | IQEFKQNRPEE---WKAIVKDSDTFKKKEEKEE---KPYISETTPHYHLENKKGIAFK---NHN                  | 433 |
| 10 | LQNFQTNRPEE---WNTLVKTEPSSGNDGKTVQG---VPLPYISYIIPHYQIENEKIGIKIFDGDGTAVD            | 477 |
| 11 | LQDFTEINRPQE---WKALTKDLDYKETSNNQ---PFISKIIPHYHITDNKIGIFRLG---TS                   | 442 |
| 12 | LSELEHKK-----ALFIKNTETNEDREH-----YWEIFPNPNYDFPKENISVNDK---D                       | 516 |
| 13 | LSELENKK-----AIFLNEREEIKG-----WEVFPNPSYDFPKENISVNYK---D                           | 529 |
| 14 | MQEVADKRMOKW---GDLIQKREERSVKLEHEELYINLDQFLEDTADSTPYVTDRRPAINIHANRIGLYWEDSQNPQK    | 538 |
| 15 | LEEAETMRKQENGTFGNSGIRIRDFENKRRDANPAN---YPYIVDTYTHYIILENNKVFIND---IND              | 452 |
| 16 | -----PVVEIHFDKQKQDQ-----FYIKRNNVILKAQ-----  | 374 |
| 17 | -----NVYYSYMPQAQNVVR-----ISELQGTISVEELRNMVYAS-----                                | 374 |
| 18 | KKHVLEEYFPNSIG---YEKFLYNDFTKEKE-----DFGLKLYSNPKTNKLIERIDN-----                    | 448 |
|    | -----S-----KY-----AE--LSV-EL--M-FY--L-----E--I                                    |     |
| 1  | -----KGAQTYIFRMGVYELKYVLLSLGKG-----AEAV   | 359 |
| 2  | -----KNAGPR--SYRMGLNELKYVLLSLQKG-----DDAI   | 344 |
| 3  | QHLWP--SPEVGTTRTGRSKYAQDKRLTAEAFISVHELPMHMFYYFLREKYSEEV---SAERVQGR                | 521 |
| 4  | QHLWP--SPEVGATRTGRSKYAQDKRLTAEAFISVHELPMHMFYYFLREKYSEEV---SAEKVQGR                | 521 |
| 5  | NNLFP--SLQTDKTCNGRSKFNLTGQFTAFAFISVHELLPMHMFYYLLTKDYSRKE-----SADKVEGII            | 529 |
| 6  | KEIWP--SLKTNDENNEKSKYKLDKQYQAEAFISVHELLPMHMFYYLLKKEKPNND---EINASIVEGFI            | 492 |
| 7  | DEIWP--SLKTNGENNEKSKYKLDKQYQAEAFISVHELLPMHMFYYLLKKEKPNND---KKNASIVEGFI            | 500 |
| 8  | DKIWP--SLKTNSEKNEKSKYKLDKSFQALAFISVHELLPMHMFYYLLKKTENTDNDNEIETKKKENKNDKQEKHKIEEII | 506 |
| 9  | --IWP--STQTELTNNKRRKYNLGTSIKAEAFISVHELLPMHMFYYLLKKTENTKNDKNGVGGK---ETKKQKHKKIFAI  | 506 |
| 10 | TDIWP--SVSTEKQLNPKDKYTLTPGFKADVLSVHELLPMHMFYYQLLCEGMLKT-----DAGNAVEKVL            | 541 |
| 11 | KELYP--SLEIKDGANRIAKYPYNSGFVAHAFISVHELLPLNFYQHLTGKS-----EDLL                      | 495 |
| 12 | FPIAG--SILDREKQPVAGKIGIKVYKLLNQYVSEVDKAVKAHLKQ---RKASKPSIQN---IIEEIVPINESNPKEAIV  | 589 |
| 13 | FPIVG--SILDREKQPVSNKIGIRVKIADQLR--EIDKAIKEKKLRNPKNRKANQDEKQKERLVNEIVSTNSNEQGEVV   | 606 |
| 14 | YKVFDENGMYIPELVVTEDEKAPKMPAPRCALSVYDLPANLFFYEYLRQDNEFP---SAEQVI                   | 600 |
| 15 | -----KEDSAPLLPVIEDDRYVVKTIIPSCRNSTLEIPANAFHMFIFGSK-----KTEKLI                     | 502 |
| 16 | -----KGGQSNVFRMGVYELKYVLLSLGKA-----EEAI   | 406 |
| 17 | -----INGKDVNKSVEQYLYHLHLLYEKIIITISGQTIKE-----G                                    | 409 |
| 18 | -----HKLVKSHGRNQDRFMDFSRFLAENNYFGKDAFFKCYKFKYDTEQDEFLQ-----SNENDDVKFH             | 509 |

图56(续)

```

-----L-----I-----LPK-----L-----D-----K-----T
1 KRIDRYVHSLRNQLPHIEKKSTEEIEG-----YVRFLLPRFV--RSHLGGLLVGDEKKIKARVDY 416
2 AKLYRYRQHVENILDVVYKVTDKD-----NHVFLPRFV--LEQHGIGRKAFAKQRIDGRVKH 397
3 KRVIEDVYAVYDAFARDEINTRDELDAACL-----ADKGIIRRGHLPRQMIAILS--QEHKDMEEKIRKKLQENMADT 590
4 KRVIEDVYAVYDAFARDEINTRDELDAACL-----ADKGIIRRGHLPRQMIAILS--QEHKDMEEKVRRKKLQEMIAADT 590
5 RKEISNIYAIYDAFANNEINSIADLTRRL-----QNTNLLQGHLPKQMHISILK--GRQKDMGKEAERKIGEMIDDI 598
6 KREIRNIFKLYDAFANGEINNIDDEKYC-----ADKGIPKRHLPKQMVAILY--DEHKDMVKEAKRKKQKEMVKDT 561
7 KREIRDIYKLYDAFANGEINNIDDEKYC-----EDKGIPKRHLPKQMVAILY--DEHKDMAEEAKRKKQKEMVKDT 569
8 ENKITEIYALYDTFANGEIKSIDELEEYC-----KGGDIEIGHLPKQMIAILK--DEHKVMATEAERKQEEMLVDV 575
9 ESKIKDIYALYDAFANGEINSEDEIKEYL-----KGGDIKIVHLPKQMIAILK--NEHKDMAEKAEAKQKEMKLAIT 575
10 IDTRNAIFNLYDAFVQEKINTITDLENYL-----QDKPILIGHLPKQMIIDLLK--GHQRDMLKAVEQKKAMLIKDT 610
11 KETVRHIQRIYKDFEERINTIEDIEKAN-----QGR-LPLGAFPKQMLGLLQ--NKQPDLSEKAKIKIEKLI AET 563
12 FIGQPTAYLSMNDIHSILYEFFDKWEKKKLEKKGEKELRKEIGKELEKKIVGKIQ--AQIQQIIOKDINAIAKIPYQD 667
13 FIGQPTAYLSMNDIHSVLYEFL-----INKTISGEALETKIVEKIE--TQIKQIIGKQATTKILKPYTN 667
14 IEYEDDYRKFVKAVAEGKLPFKRPFKEFR---DFLKKKEYPKLRHADIPKQLFLC--SHGLCYNNKPETVYERLDRLI 674
15 VDVHNRKYLRFQAMQKEEVTAENIASFG-----IAESDLPQKILDLSGNAHGKQVDVAFIRLTVDDMLDTI 568
16 KETVRHIQRIYKDFEERINTIEDIEKAN-----SINFLLPRFV--RSRLGLLVQDEKRLKTRLEY 463
17 RVDVEDYRPLLDKLLRPASNGEELR-----RELKLLP--KRVCDLLSNRFDCSEGVSAVE 464
18 KGKVTTYIKYEEHLKNYSYWDPCPFVEENN---SMSVKISIGSEEEKILKIQRNLMYFLENALYNENVENQGYKLVNNY 584

```

```

---L---L-----L-K-G---A---L---D-M-FQP-----D---P-N---H
1 VKAKWLEKKEKSRELQLHR-----KGRDILRYIN-----ERCERPLNIDEYN 458
2 VRGVWEKKAATNEMTLHE-----KARDILQVYN-----ENCTRSFNPGEYN 439
3 DHRLDMLDRQTDKIRIGR-----KNA-GLPKSGVIADWLVRDMMRFQPVAK-----DASGKPLNNSKAN 649
4 DHRLDMLDRQTDKIRIGR-----KNA-GLPKSGVVADWLVRDMMRFQPVAK-----DTSGKPLNNSKAN 649
5 QRRLDMLCKQTNQKIRIGK-----RNA-GLLKSGKIADWLVDNMMRFQPVQK-----DQNNIFINNSKAN 657
6 KKLATLEKQTQKEKEDDG-----RNV-KLLKSGEIARWLVNDMMRFQPVQK-----DNEGKPLNNSKAN 620
7 KKLATLEKQTQGEIEDGG-----RNI-RLIKSGEIARWLVNDMMRFQPVQK-----DNEGKPLNNSKAN 628
8 QKSLESLDNQINEEIEINVE-----RKN-SSLKSGKIASWLVNDMMRFQPVQK-----DNEGKPLNNSKAN 634
9 ENRKLTLQKQKQKQKQK-----RYN-SAPKSGEIASWLVNDMMRFQPVQK-----DENGESLNSKAN 634
10 ERRLKLDKQLKQEDVAA-----KNTGTLLKNGQIADWLVDNMMRFQPVKR-----DKEGNPINCSCAN 670
11 KLLSHRLNFKLKSSPKLGG-----RRE-KLTKTGVLDWLKDFMRFQPVAY-----DAQNQPIKSSKAN 622
12 GNSTAIKDEKLIKDLKQEQ-----NILQKLDKQEQTVREKEYNDFIAYQDKNREINKVRDRNHKQYLKQNLK 733
13 ANSNSINREKLLRDLEQEQ-----QILKTLLEEQQQREK-----DKKDKK-----707
14 LQHLEERELHIQNRLEHYQKDRDMIGNKDNQYGGKKSFSQVHRGALARYLAQSNMEWQPTKLDKDEKGDHKLTLGLNYNVLT 754
15 ERRIKRFKDDKRSIRSADN-----KMGKRGFKQISTGKLADFLAKDIVLQFQSVN-----DGENKITGLNYR 630
16 VKAKWTDKKEGSRKLELHR-----KGRDILRYIN-----ERCERPLNIDEYN 505
17 KRLKAILLRHEQLLLSQNP-----ALHIDKIKSVIDYLYLFFSODEK-----FRQPTEKAHRGLK 520
18 YRELKQDVEESIASLDLIK-----SNPDFKSKYKILPKRLLHNYAPAKQ-----DKAPENAFETLLK 642

```

```

--E---R---A---E---L---F---LI---NPHPF-----N---FY---YL-----L-----
1 RILELLVTKHLGGFYRELEELKTRRIDKNIVCNLSR-----HKSVNALHEKVCIDLVVQEELES-----516
2 RLLVCLVGVKQVFNFAQGLKRLQLAERIDGRVYSIFAQ-----TSTINEMHQVVCQQLNRLCR-----497
3 STEYRMLQRALALFGGKERLTPYFRQMNLTG6NNPHPF---LHETRWEHSTNLSFYRSYLARKAFLEK-----717
4 STEYRMLQRALALFGGKERLTPYFRQMNLTG6NNPHPF---LHETRWEHSTNLSFYRSYLEARKAFLEK-----717
5 STEYRMLQRALALFGSENFRLKAYFNQMNLVGN6NNPHPF---LAETQWEHQTNLSFYRNYLARKKYLQK-----725
6 STEYQMLQSLALYNNI-EKPTRYFRQVNLIESNNPHPF---LKWTWEECNILTFYYSYLTKKIEFLNK-----687
7 STEYQMLQSLALYNNI-EKPTRYFRQVNLIESNNPHPF---LKWTWEECNILTFYRSYLTKKIEFLNK-----695
8 STEYQLLQRTLAFFGSEHERLAPYFKQTKLIESSNPHPF---LKQTEWEKCNILTFYRSYLEAKKNFLES-----702
9 STEYQLLQRTLAFFGSEHERLAPYFKQTKLIESSNPHPF---LNDTEWEKCSNLSFYRSYLKARKNFLES-----702
10 STEYQMLQRAFAFYATOSCRISRYFTQLHLIHS6NSHLF---LSRFYEDKQPNLIAFYAAYLAKLEFLNE-----738
11 STEFWFIRRALALYGG6KNRLEGYFKQTNLIGN6NNPHPF---LNKFNWACRNLDVDFYQYLEQREKFLA-----690
12 RKYPEAPARKEVLYYREKGVAVWLAND--IKRFMPTDF---KNEWKGEQHSLLQKSLAYYEQCKEELKNL-----799
13 -----SKRKHELYPSEKGVAVWLAND--IKRFMPKAF---KEQWRGYYHSSLLQKYLAYYEQCKEELKNL-----767
14 AYLATYGHPPVPEEGFPRTLEQVLINAHLIG6SNPHPF---INKVLALGNRIEELYLHYLEEELKHISRIQSLSSNP 831
15 IMQSAIAVYDSGDDYEAQQFKLHFEKARLIGKGTTEPHFPFLYKVFARSIPANAVEFYERYLIERKFYLTG-----701
16 NILKFIVNKDFAGFYNELEELKTRRRLDKNIIQKLSG-----HTTLNALHERVCDLVLQELGS-----563
17 DEEFQMYHYLVGDYDHPALWKELEASGRKLP6MRKLT-----SATSLHGLYMLCKGTVEWCRC-----581
18 KADFREEQYKLLKKAHEKKNKEDYKRNK6KQFKLHFIRKACQMMYFKEKYNTLKEGNAAF6KDPVIEKR-----714

```

图56(续)

|    |  |      |
|----|--|------|
|    | -----L-----L-----LV---K---LP-G-FT-----V----- |      |
| 1  | -----L-----L-----LV---K---LP-G-FT-----V----- | 547  |
| 2  | -----L-----L-----LV---K---LP-G-FT-----V----- | 526  |
| 3  | -----L-----L-----LV---K---LP-G-FT-----V----- | 786  |
| 4  | -----L-----L-----LV---K---LP-G-FT-----V----- | 786  |
| 5  | -----L-----L-----LV---K---LP-G-FT-----V----- | 799  |
| 6  | -----L-----L-----LV---K---LP-G-FT-----V----- | 761  |
| 7  | -----L-----L-----LV---K---LP-G-FT-----V----- | 769  |
| 8  | -----L-----L-----LV---K---LP-G-FT-----V----- | 772  |
| 9  | -----L-----L-----LV---K---LP-G-FT-----V----- | 772  |
| 10 | -----L-----L-----LV---K---LP-G-FT-----V----- | 812  |
| 11 | -----L-----L-----LV---K---LP-G-FT-----V----- | 766  |
| 12 | -----L-----L-----LV---K---LP-G-FT-----V----- | 874  |
| 13 | -----L-----L-----LV---K---LP-G-FT-----V----- | 842  |
| 14 | -----L-----L-----LV---K---LP-G-FT-----V----- | 909  |
| 15 | -----L-----L-----LV---K---LP-G-FT-----V----- | 773  |
| 16 | -----L-----L-----LV---K---LP-G-FT-----V----- | 594  |
| 17 | -----L-----L-----LV---K---LP-G-FT-----V----- | 629  |
| 18 | -----L-----L-----LV---K---LP-G-FT-----V----- | 790  |
|    | -----D---QPFFY---F-----K-----F---EE---W----- |      |
| 1  | -----D---QPFFY---F-----K-----F---EE---W----- | 588  |
| 2  | -----D---QPFFY---F-----K-----F---EE---W----- | 567  |
| 3  | -----D---QPFFY---F-----K-----F---EE---W----- | 846  |
| 4  | -----D---QPFFY---F-----K-----F---EE---W----- | 837  |
| 5  | -----D---QPFFY---F-----K-----F---EE---W----- | 851  |
| 6  | -----D---QPFFY---F-----K-----F---EE---W----- | 832  |
| 7  | -----D---QPFFY---F-----K-----F---EE---W----- | 840  |
| 8  | -----D---QPFFY---F-----K-----F---EE---W----- | 820  |
| 9  | -----D---QPFFY---F-----K-----F---EE---W----- | 827  |
| 10 | -----D---QPFFY---F-----K-----F---EE---W----- | 864  |
| 11 | -----D---QPFFY---F-----K-----F---EE---W----- | 813  |
| 12 | -----D---QPFFY---F-----K-----F---EE---W----- | 923  |
| 13 | -----D---QPFFY---F-----K-----F---EE---W----- | 893  |
| 14 | -----D---QPFFY---F-----K-----F---EE---W----- | 978  |
| 15 | -----D---QPFFY---F-----K-----F---EE---W----- | 834  |
| 16 | -----D---QPFFY---F-----K-----F---EE---W----- | 635  |
| 17 | -----D---QPFFY---F-----K-----F---EE---W----- | 685  |
| 18 | -----D---QPFFY---F-----K-----F---EE---W----- | 851  |
|    | -----W-K-E-EL-----                           |      |
| 1  | -----W-K-E-EL-----                           | 609  |
| 2  | -----W-K-E-EL-----                           | 588  |
| 3  | -----W-K-E-EL-----                           | 924  |
| 4  | -----W-K-E-EL-----                           | 868  |
| 5  | -----W-K-E-EL-----                           | 878  |
| 6  | -----W-K-E-EL-----                           | 865  |
| 7  | -----W-K-E-EL-----                           | 873  |
| 8  | -----W-K-E-EL-----                           | 851  |
| 9  | -----W-K-E-EL-----                           | 860  |
| 10 | -----W-K-E-EL-----                           | 892  |
| 11 | -----W-K-E-EL-----                           | 844  |
| 12 | -----W-K-E-EL-----                           | 954  |
| 13 | -----W-K-E-EL-----                           | 921  |
| 14 | -----W-K-E-EL-----                           | 1056 |
| 15 | -----W-K-E-EL-----                           | 867  |
| 16 | -----W-K-E-EL-----                           | 656  |
| 17 | -----W-K-E-EL-----                           | 712  |
| 18 | -----W-K-E-EL-----                           | 882  |

图56(续)

|    |   |      |
|----|---|------|
|    | -----D-----M-----L-----L-----N-----L-----   |      |
| 1  | TLAMDRLLMIARQYHLSLNKELAKRAQ-----QIEWKKEDGEEVII                                    | 651  |
| 2  | TLARDRLCLMMAQYFLGSRKELGN-----KIVWSNDS-----IE                                      | 623  |
| 3  | VKNQDIITWMMCRDLMEENKVEGLDTG-----TLYLKDIPRNVQEQGS-----LNVLNRVKPMRIPVV              | 982  |
| 4  | VKNQDIITWMMCRDLMEENKVEGLDTG-----TLYLKDIRTQVQEQGS-----LNVLNRVKPMRIPVV              | 926  |
| 5  | IKNQDIVTWLHFKELFNMATVEGLKIG-----EIHRLRIDTNTANES-----NNILNRVMPHQLPVT               | 936  |
| 6  | VRNQDIVTWLCKEIDKLIKIDELNIEE-----LKKLRLENIOTDTAKKEK-----NNILNRVMPHQLPVT            | 926  |
| 7  | VRNQDIVTWLCTEIDKLIKVEGLNVEE-----LKKLRLEKIDTDTAKQEK-----NNILNRVMPHQLPVT            | 934  |
| 8  | VRNQDIVTWLCCMEIFNKKKIKELN-----VEKIYIKNINTNTTKEKNTTEKNKEEKNKEKNILNRVMPHQLPIK       | 924  |
| 9  | VRNQDIVTWLCTEIDKLIKIDELNIKE-----LKKLRLEKIDTDTAKKEK-----NNILNRVMPHQLPVT            | 921  |
| 10 | IKNQDILWLWCKNLFDEKIKPKKDILEPRIAVSYIKLDSLQNTSTAGS-----LNALAKVVPHTLAIH              | 957  |
| 11 | YRNQDVNLWMLTLELTKN-HFKELNLN-----YHQLKLELAVNVQEQADAK-----LNPLNQTLPMLPVT            | 904  |
| 12 | QKKNQVFTLLMAKHIFKSVFKQDSIDQ-----FSLEDLYQSREERLG-----NQRARQGTGERNTNY               | 1011 |
| 13 | QKKNQVYTLHMVEYIIQKIFGDKFVEEN-----PLVLKGFQSKAERQQ-----NNTHAATTQERNLNG              | 980  |
| 14 | IKARQETRRQMRDELQRMPPRYIRDIKDNERTLRRYKTDQMVFLFLAEKMFN-----IISEQSSEFNWQKMRISK       | 1127 |
| 15 | RLSNRSRNEYQKSEKVIIRRYRVQDALLFL-----LAKKTELADDFGERFK-----LKEIMPDAEKGIILSE          | 928  |
| 16 | TLAMDRRLCLMMAQYFLGSRKELGN-----KIVWSNDS-----IE                                     | 698  |
| 17 | VHTNDKVLVLAQRYRERLLNVGSKCAVT-----LDNPERQKQLADVVEVQN-----AWLSIRFDDLDFTLT           | 774  |
| 18 | NKLEDCLLYEIAYNYIOKKNVHKIDIQKILTSKIILTINDANTPYKISVPPFN-----KLERYTEMIATKNQNNIKAR    | 954  |
|    | -----T-YI-----KLL-----GNF-----D-RL-----L-----                                     |      |
| 1  | FTLKN-----PAQPEQSCSVRFSLRDYTKLYVMDDAEFLARLCDYFLPKDEEQIDYHRLYT-----                | 707  |
| 2  | LPVEG-----SVGNEK-----SIVFSVSDYSKLYVLDBAEFLGRICEYFMPHEKGIKIRYHTVYE-----            | 677  |
| 3  | VYRAD-----SRGHVHKEEAPLAVYIEERDTKLLKQGNFKSFVKQRRRLNGLFSFVDTG-GLAMEQYPIISK-----     | 1047 |
| 4  | VYRAD-----SRGHVHKEQAPLATVYIEERDTKLLKQGNFKSFVKQRRRLNGLFSFVDTG-ALAMEQYPIISK-----    | 991  |
| 5  | TYETD-----NKGNIKER-PLATFYIEETETKVLKQGNFKALVKQRRRLNGLFSFAETT-DLNLEEHPISK-----      | 1000 |
| 6  | VYEID-----DSHKIVKDK-PLHTYIYKEAETKLLKQGNFKALVKQRRRLNGLFSFVKTNSEAESKRNPISK-----     | 991  |
| 7  | VYEID-----DSHNIVKDR-PLHTYIYKEAETKLLKQGNFKALVKQRRRLNGLFSFVDTSSSETELKSNPISK-----    | 999  |
| 8  | VYGRNFSXNKKKIRRN-TFFTVYIEEKGTKLLKQGNFKALERDRRLGGLFSFVKTSPKAESKSNPISK-----         | 993  |
| 9  | VYKVN-----KGGYIKNK-PLHTYIYKEAETKLLKQGNFKALVKQRRRLNGLFSFVKTSPSEAESNPISK-----       | 986  |
| 10 | IDSPKP-KGKAGNNEKEN-KEFTVYIKEEGTKLLKWSNFKTLADRRIKGLFSYIEHD-DIDLKQHPLTK-----        | 1024 |
| 11 | VYPAT-----AFGEVQYHKTPIRTVYIREEHTKALKMGNFKALVKQRRRLNGLFSFIKEE--NDTQKHPISQ-----     | 968  |
| 12 | IWNKT-----VDLKLCDGKITVENVKLNVDI IKYEYDQRVQAFKYEENIEWQAFLIKESKEEENYP-----          | 1075 |
| 13 | ILNQP-----KDIKIQG-KITVKGVKLKDISENFKYIEIDQVNTFLDYEPKEMAYLPNDWKEKEKQGGQPPN          | 1048 |
| 14 | VCNEAFLRQTLTRFVPVTVGETIYVEQENMSLKNYDEYRFLTDRLMSLNNIVETLKPENGDVLR-----H            | 1198 |
| 15 | INPNS-----FTFEKGGKYYITSEGMLKNYDFVFLASQKIGNLLELVGSDIVSKED-----                     | 983  |
| 16 | FKFQN-----PKEQKKSFSIRFSIILDYTKHYVMDDPEFLSRLWEYFIPKEAKEIDYHKHYA-----               | 754  |
| 17 | HVNLS-----NLRKAYNLIPRKHILAFKEYLDNRVKQKLCCECRNVRRKEDLCTCCSPRYSN-----L              | 832  |
| 18 | FLIDLPLYLSKNKIKKQKDSAGYEIITKNDLEIEDINTINNKIINDSVKFTFVLEMELEKYFILKDKCILSKN-----YID | 1029 |
|    | -----EL-Y-R-----L-LE-L-----L-----   |      |
| 1  | ---QGMNRYTNLQ-REGIEAILELEKKTIGPEQP-----RPPKNYIPFSEIMDK                            | 752  |
| 2  | ---KGFRAYNLDLQ-KKCVEAVLAFEEKVVKAKKMS-----KEGAHYIDFREILAQ                          | 724  |
| 3  | LRVEYELAKYQTAH-VCVFELTLREESLLTRYPHLP-----DESFRMLESWSDP-LLAKWPELHGK                | 1108 |
| 4  | LRVEYELAKYQTAH-VCAFEQTLREESLLTRYPHLP-----DKNFRKMLLESWSDP-LLDKWPDLHGK              | 1052 |
| 5  | LSVDLELILKYQTR-ISIFEMTLGLEKLLDKYSTLP-----TDSFRNMLERWLQ--CKANRPELKNY               | 1060 |
| 6  | LRVEYELGEYQEAR-IEIIQDMLALEEKLINKYKDL-----TNKFSEMLNSWLEGGKDEADKARFQND              | 1053 |
| 7  | SLVEYELGEYQNAH-IETIKDMLLEETLIEKYKTL-----TDNFSOHLNGWLEGGKDEADKARFQND               | 1061 |
| 8  | LRVEYELGEYQKAR-IEIIKDMLALEKTLIDKYNLSD-----TDNFKNLTDLWLELKGEPDKASFQND              | 1055 |
| 9  | LRVEYELGKYQNAH-LDIIEDMLALEKLLDKYNLSD-----TDNFMNMLTGWLELKGAKKARFQND                | 1048 |
| 10 | RRVDLELDLYQTCR-IDIFQQLSLEAQLLDKYSDL-----TDNFIYQMLIGWRKKEGIPR--NIKED               | 1084 |
| 11 | LRLRRELEIYQSLR-VDAFKETLSLEKLLNKHTSLSS-----LENEFRALLEEWKKEYAASS-NVTDEH             | 1031 |
| 12 | YVVERIEQYKVRREELLKEVHLIEEYILEKVKDK-----ILKKGDNQNFKYIILNGLLKQLKNEQVSEYKVFNLNTEPED  | 1153 |
| 13 | NVIDRQISKYETVRSKILLKDVQEELEKIIISDEIKEEHRDLKQGGYNYFKYIILNGLLRQLKNENVENYKVKLNTNPEK  | 1128 |
| 14 | TDLMSLAAYDQYR-STIFNLIQSIENLIITNNAVD-----DPDADGFWVREBLPKR                          | 1250 |
| 15 | --IMEEFNKYDQCR-PEISSIVFNLEKWFDTYPELS-----ARVDREK                                  | 1025 |
| 16 | --RAFDKYTNLQ-KEGIDAILKLEGRITERRKI-----KPAKNYIEFQEIMNR                             | 799  |
| 17 | TSWLKENHSESSIEREAATMMLLOVERKLLSFLDER-----RKAIIEYKGFIPFSAL                         | 885  |
| 18 | NSEIPSKQFSKVIKENENEIINRYNIACHFHLPLET-----FDNLLNVEYQKFIKEE                         | 1084 |

图56(续)

|    | V-----A---NQY-----I-----L-----   |      |
|----|--|------|
| 1  | SAYNEDDQKALRRVRNALLHHNLN FARADFKRFGIMKREGIEKRWSLAV-----                      | 802  |
| 2  | TMCKEAEKTAVNKVRRAFFHHHLKFVDEFGFLSDVMKKYGIKEWKFPVK-----                       | 775  |
| 3  | VRLLI AVRNAF SHNQYPMYDEAVFSSIRKYDPSSPDAIEERMGNIAHRLSEEVKQAKETVERIIQA-----    | 1175 |
| 4  | VRLLI AVRNAF SHNQYPMYDETLFSSIRKYDPSSPDAIEERMGNIAHRLSEEVKQAKEMVERIIQA-----    | 1119 |
| 5  | VNSLI AVRNAF SHNQYPMYDATLFAEVKKFT-LFPSVDTKKIELNIAPQLLEIVGKAIKEIEKSENKN-----  | 1127 |
| 6  | VDFLI AVRNAF SHNQYPMHNKIEFANIKPFSLYTANNSEEK-GLGIANQLKDKTKETTDKIKKIEKPIETKE-  | 1124 |
| 7  | VKLLV AVRNAF SHNQYPMRNRIAFANINPFSLSADTSEEK-KLGIANQLKDKTKHKIIEIEIEKPIETKE-    | 1132 |
| 8  | VDLLI AVRNAF SHNQYPMRNRIAFANINPFSLSANTSEEK-GLGIANQLKDKTKHKIEKIEIEIEKPIETKE-  | 1126 |
| 9  | VKLLT AVRNAF SHNQYPMYDENLFGNIERFSLSSNIIESK-GLDIAAKLKEEVSKAAKKIQNEEDNKKEKET   | 1120 |
| 10 | TDFLX DV RNAF SHNQY PDSKKIAFRRIKFNPKELILEEEE-GLGIATQMYKEVEKVVNRICKRIELFD---- | 1151 |
| 11 | IAFIASVRNAFCHNQYPFYKEALHAPIPLFTVAQPTTEEKD-GLGIAEALLKVLREYCEIVKSQI-----       | 1095 |
| 12 | VNINQLKQEATDLEQKAFVLT YIRNKFAHNQLPKKEFWDYCQEKYKIEKEKTYAEYFAEVFKKEKEALIK--    | 1224 |
| 13 | VNITQLKQEATDLEQKAFVLT YIRNKFAHNQLPKKEFWDYCQEKYKIEKEKTYAEYFAEVFKREKEALIK--    | 1199 |
| 14 | NNFASLLELINQLNNVELTDDERKLLVAIRNAF SHNSYNIDFSLIKDVKHLPEVAKGILQHLQSM LGVEITK-  | 1322 |
| 15 | VDFKSILKILLNNKNINKEQSDILRKIRNAFDHNNYPDKGVVEIKALPEIAMS IKKAFGEYAIMK-----      | 1090 |
| 16 | SGYNDQQVALKRVRNALLHYNLNFEREHLKRFYGVV KREGIEKKWSLIV-----                      | 849  |
| 17 | VKECRLADAGLCGIRNDVLHONVISYADAIGKLSAYFPKEASEAVEYIRRTKEVREQRREELMANSSQ-----    | 953  |
| 18 | LQNVSTINDLSKPQEYLILLFIKFKHNNFYLNLFNKNESKTIKNDKEVKKNRVLQKFINQVILKKK-----      | 1150 |

图56 (续)

| RESCUE Round | ADAR 突变   | 质粒      |
|--------------|---|---------|
| RESCUEv0     | E488Q   | pAB0048 |
| RESCUEv1     | E488Q, V351G  | pAB0359 |
| RESCUEv2     | E488Q, V351G, S486A   | pAB1188 |
| 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 |
| RESCUEv7     | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I                                    | pAB1194 |
| RESCUEv8     | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V                             | pAB1220 |
| RESCUEv9     | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I                      | pAB1327 |
| RESCUEv10    | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L               | pAB1411 |
| RESCUEv11    | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G        | pAB1491 |
| RESCUEv12    | E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T | pAB1553 |

图57

β 连环蛋白比率

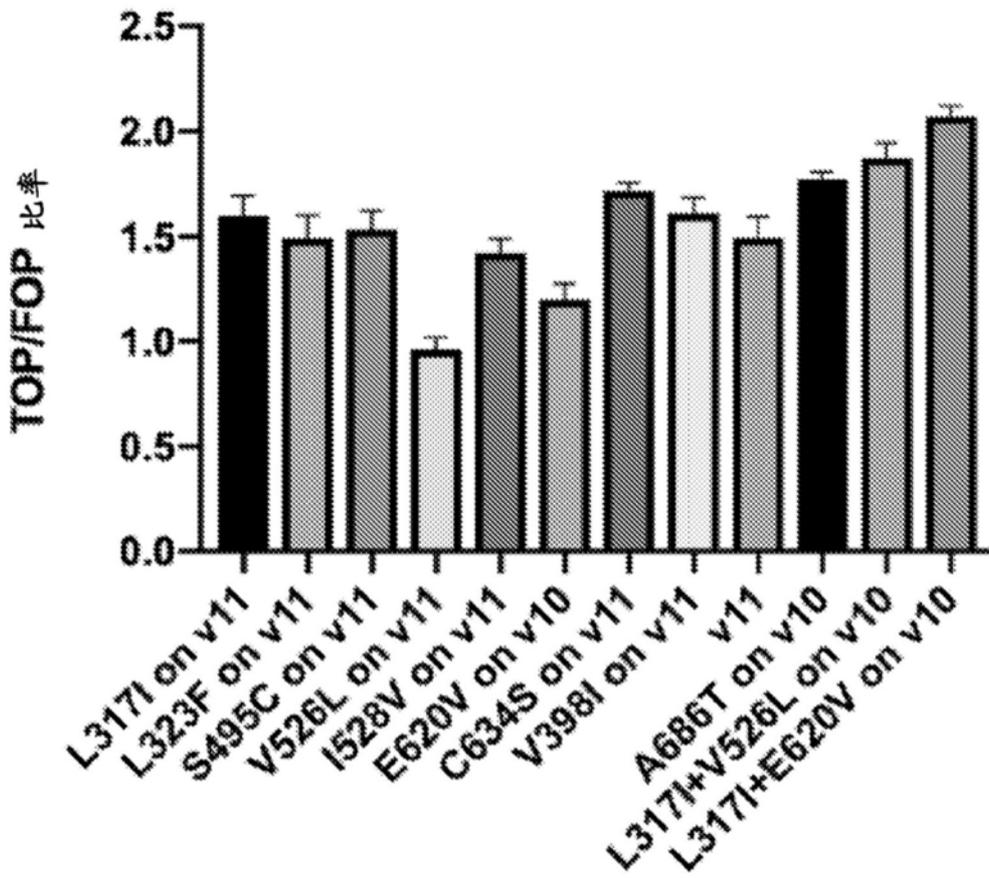


图58

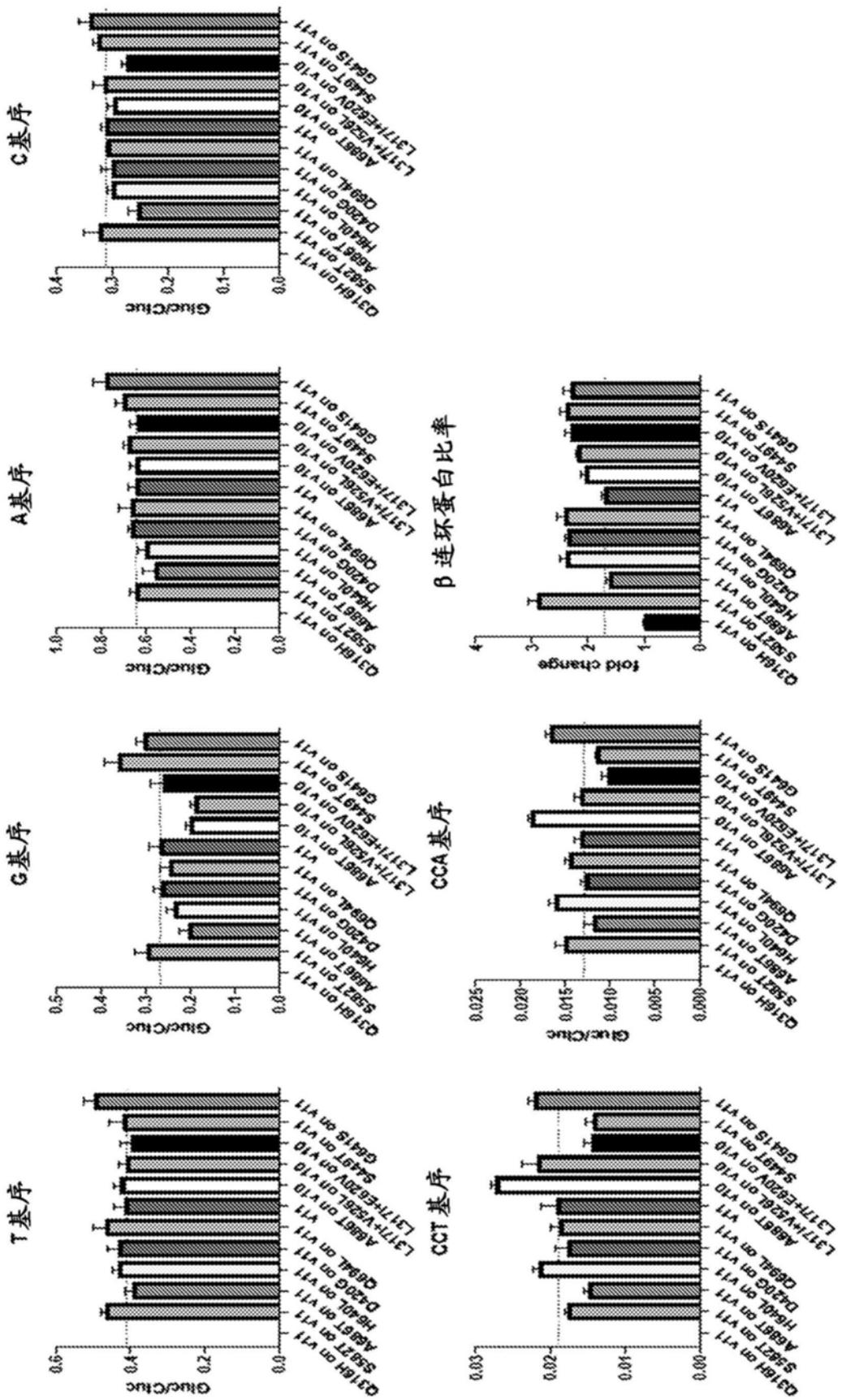


图59

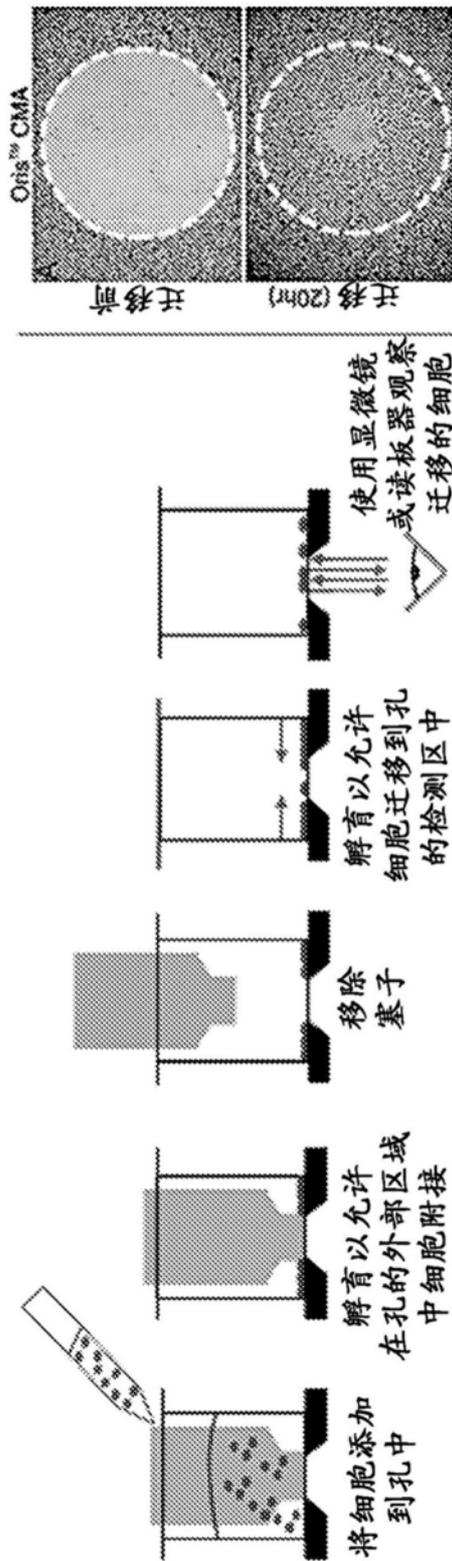


图60

### HEK293FT 生长

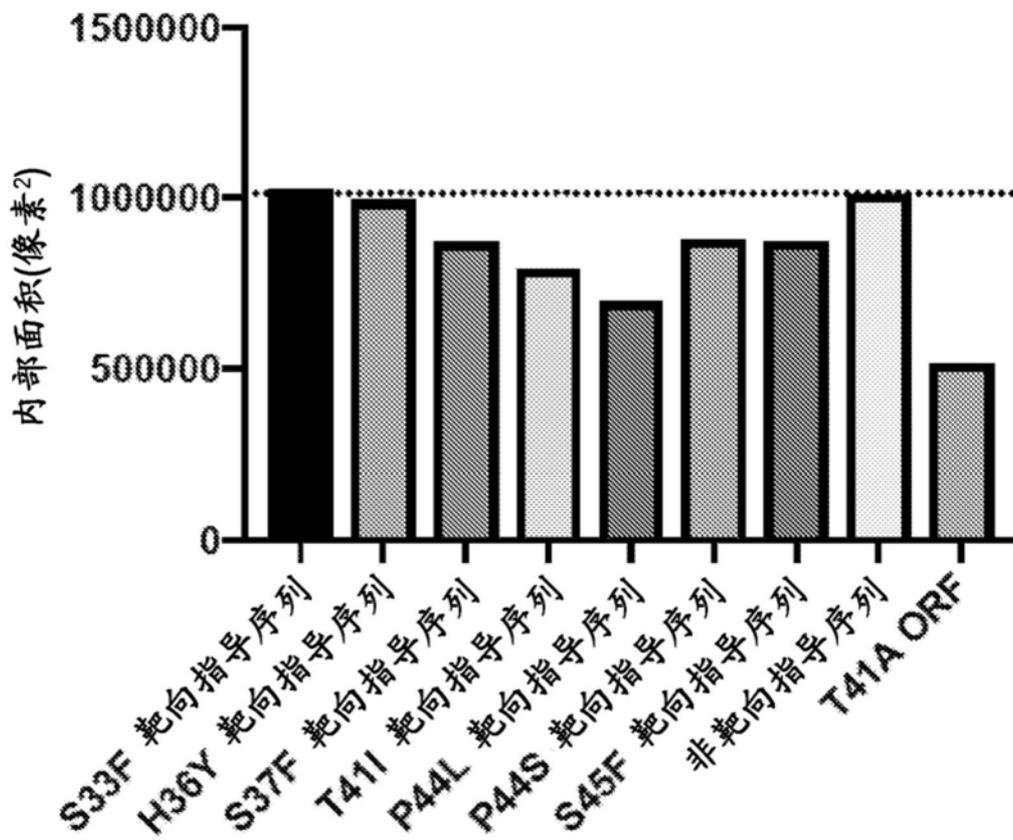


图61

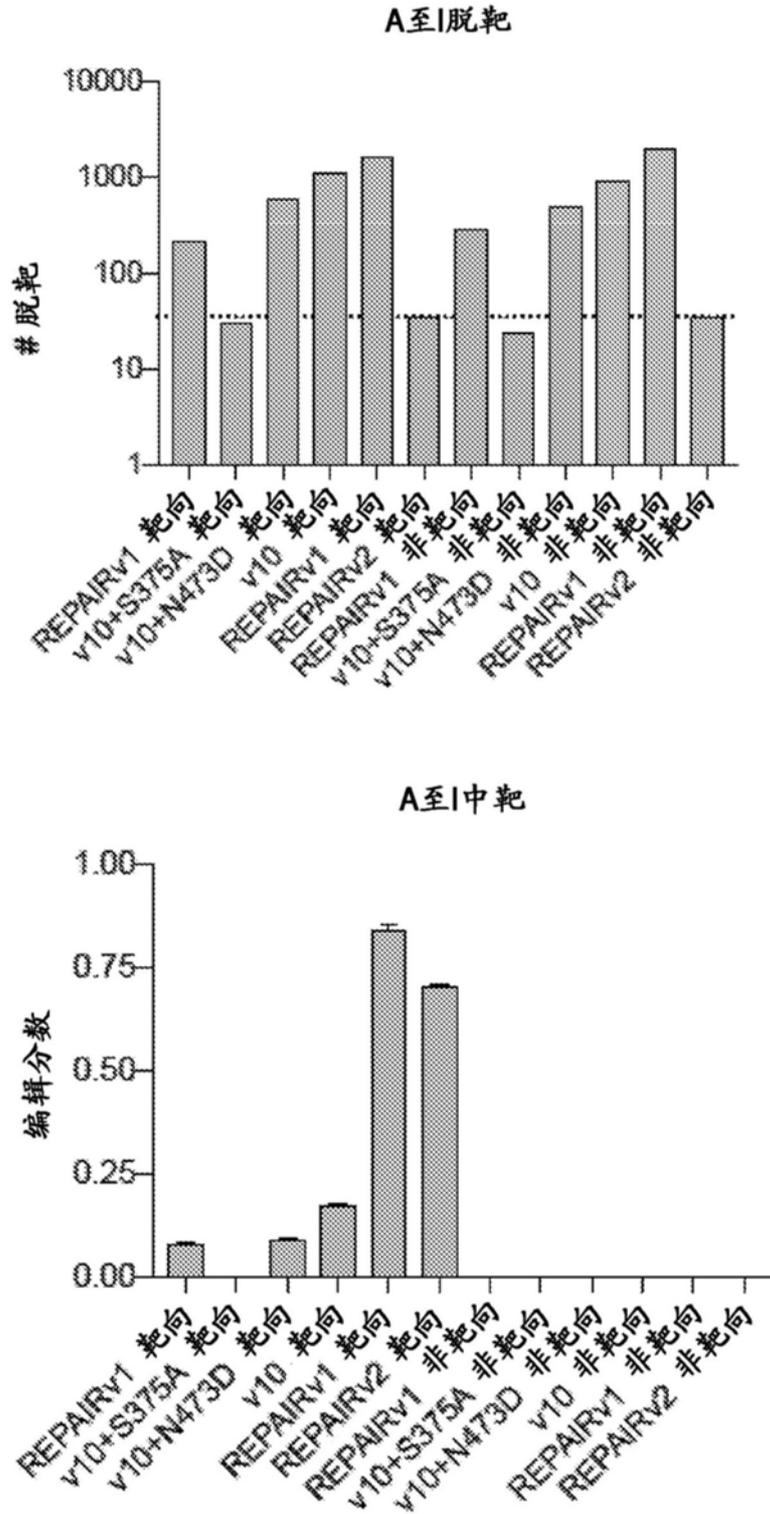


图62

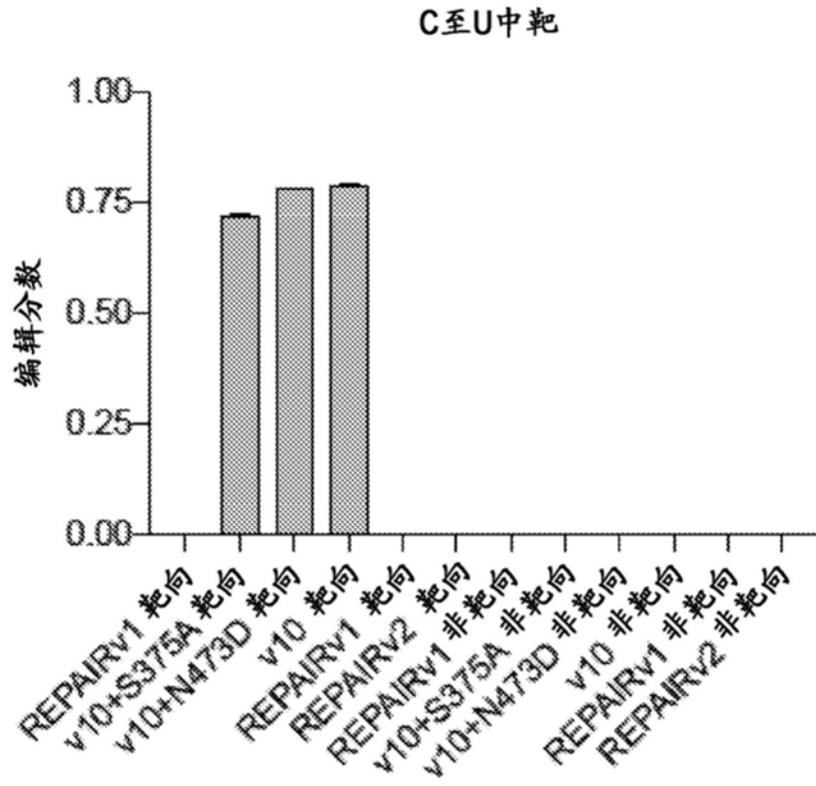


图62(续)

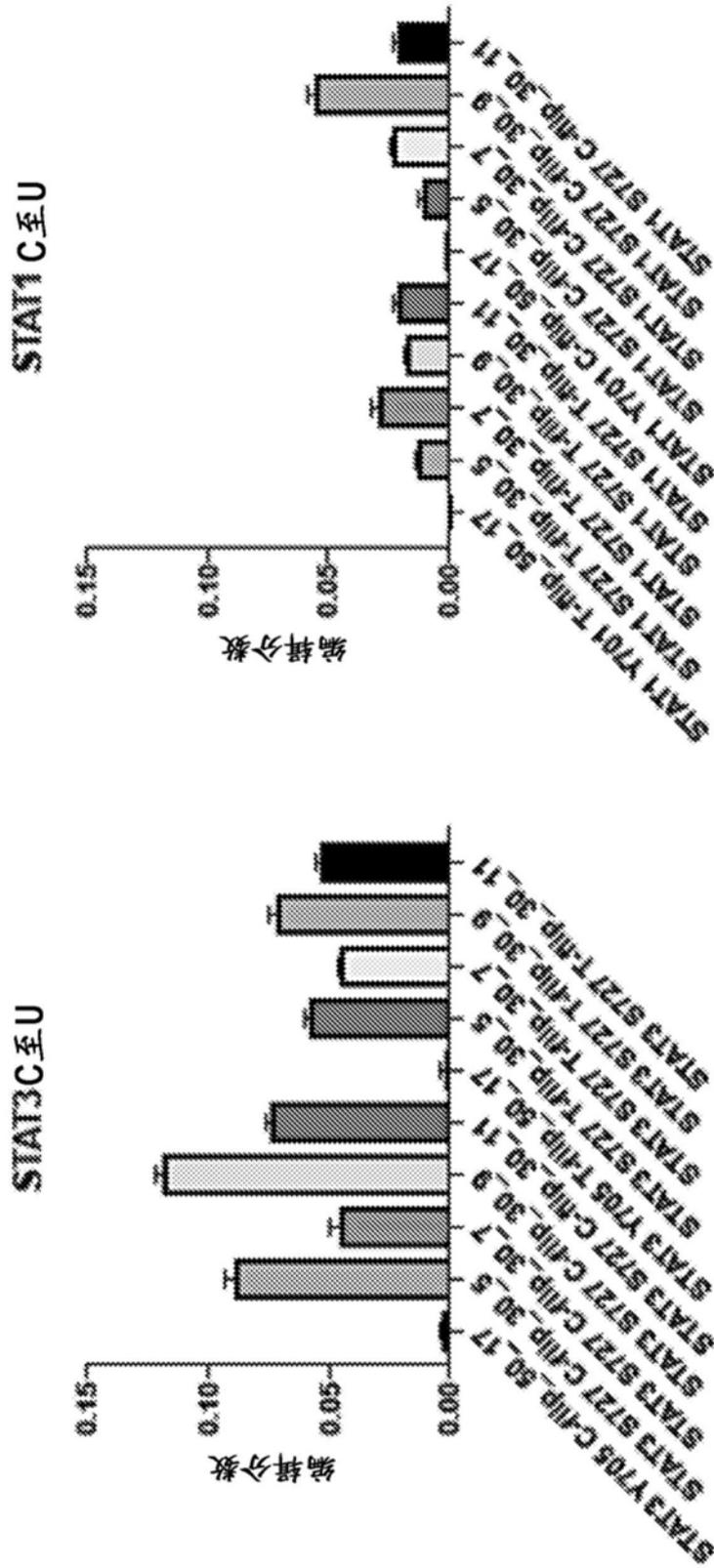


图63



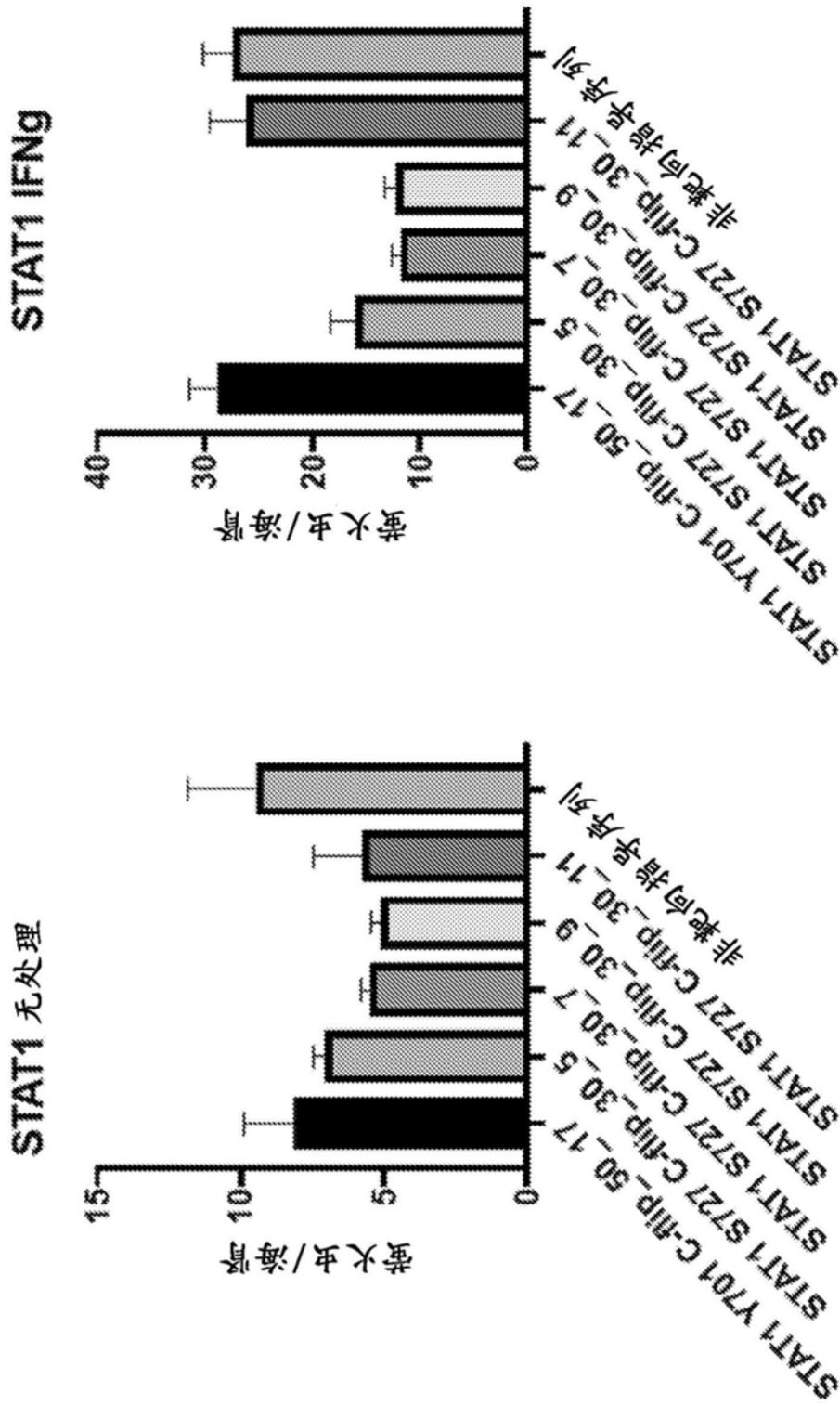


图64

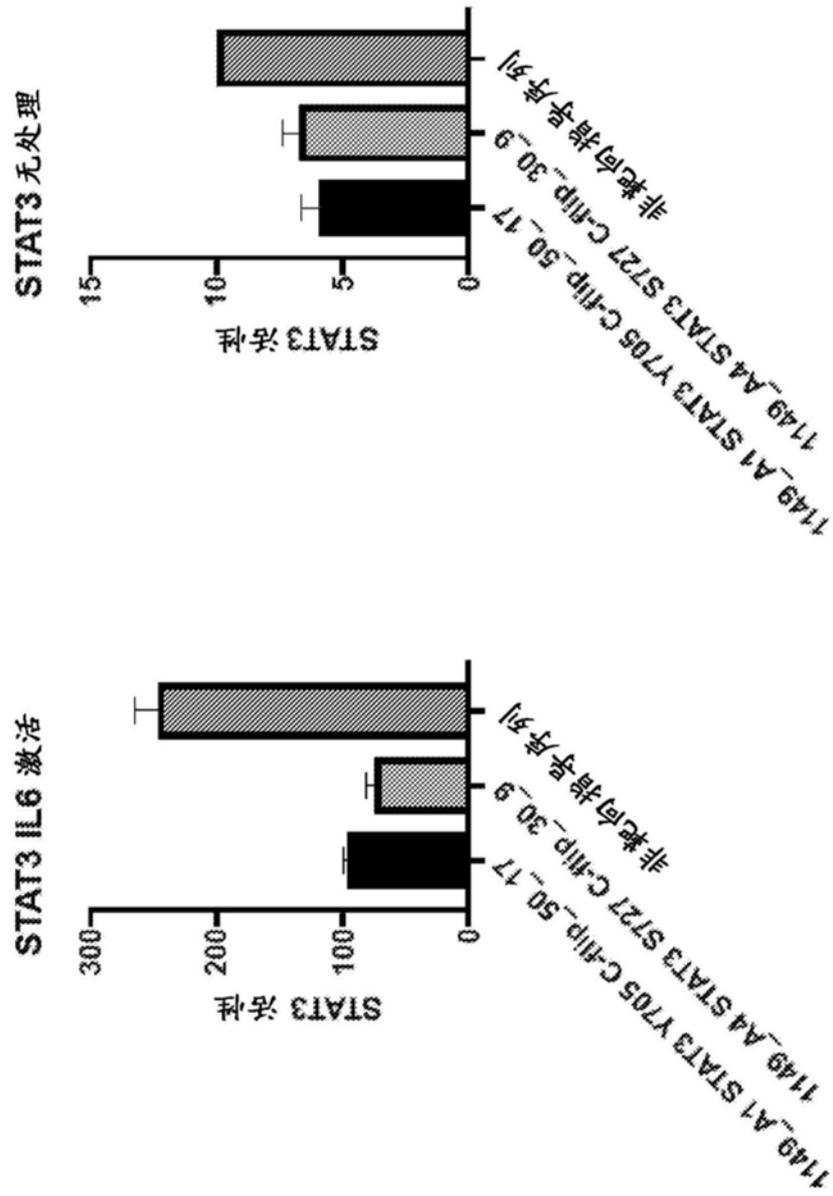


图65

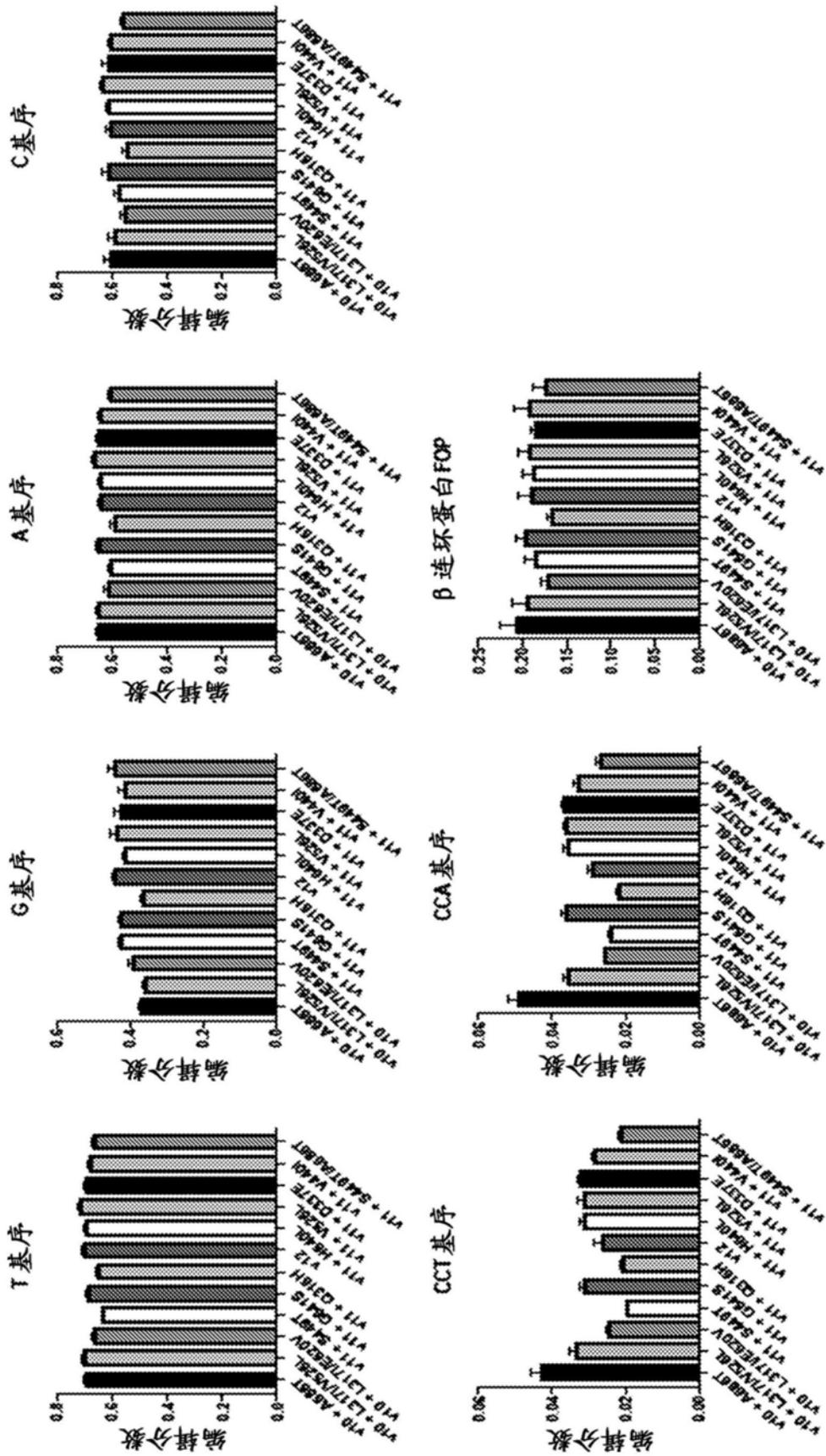


图66

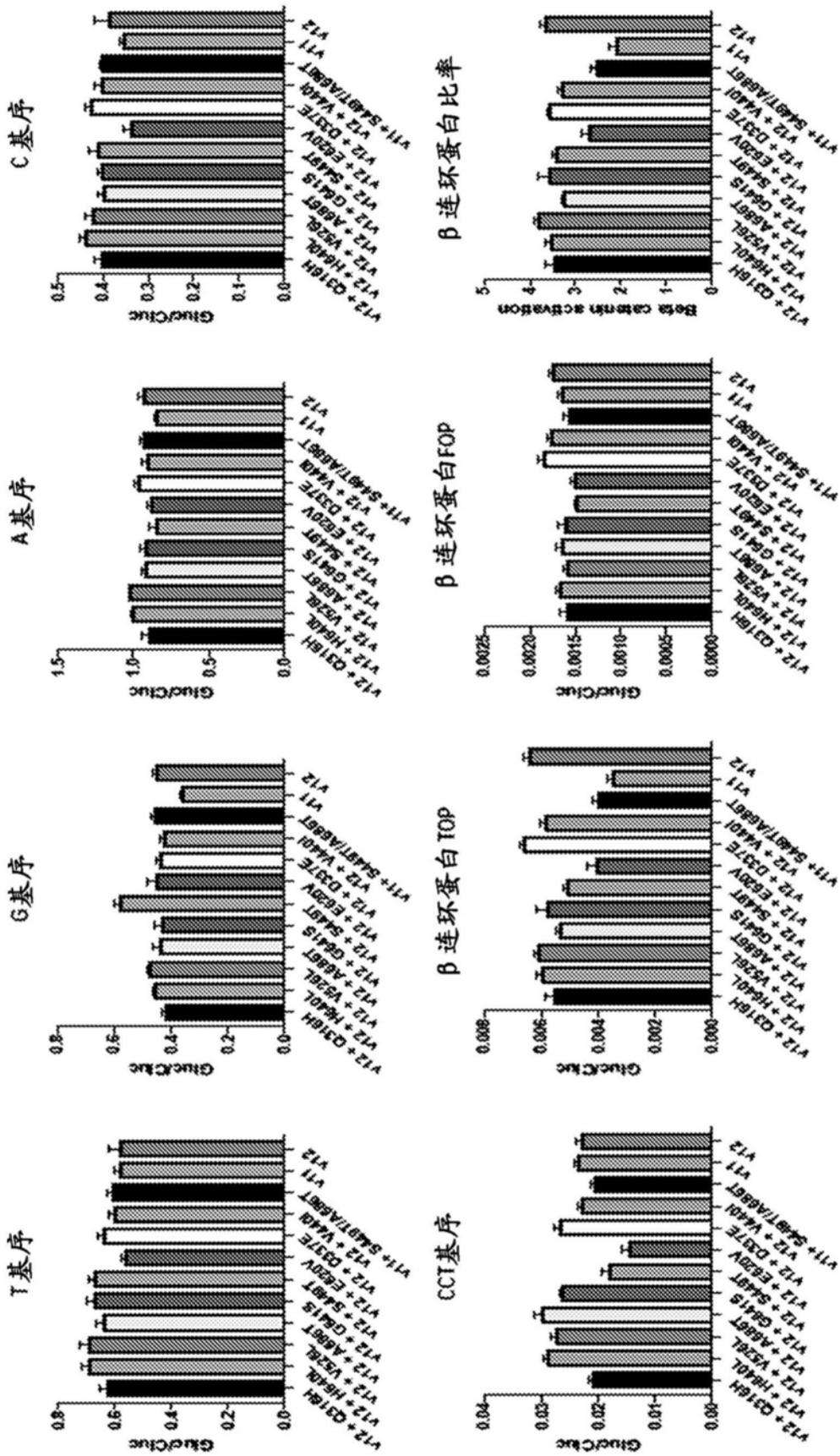


图67

第2天，HEK293FT CTNNB1迁移，具有细胞追踪绿色

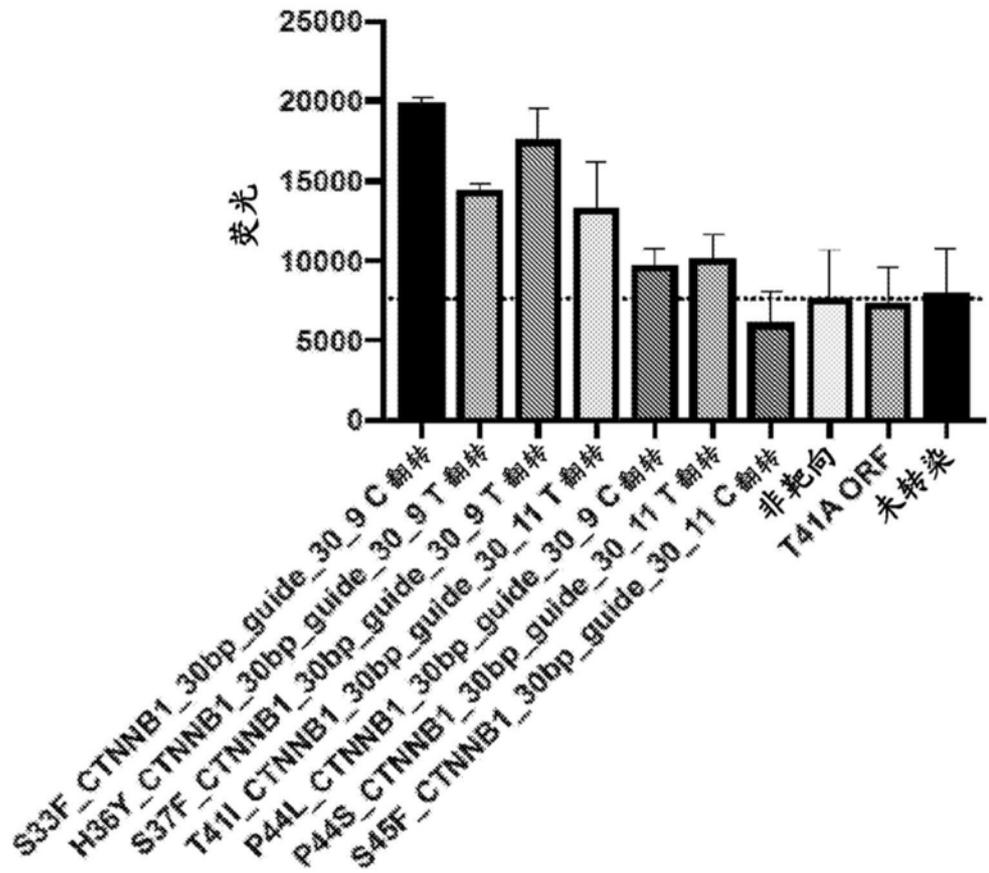


图68

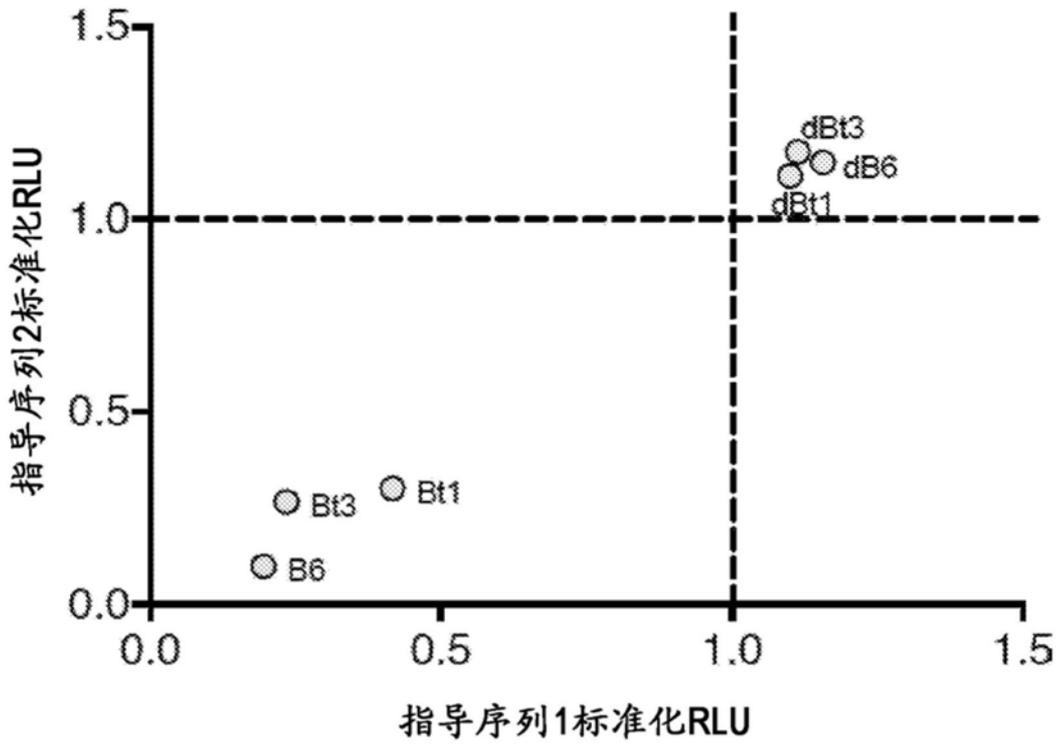


图69

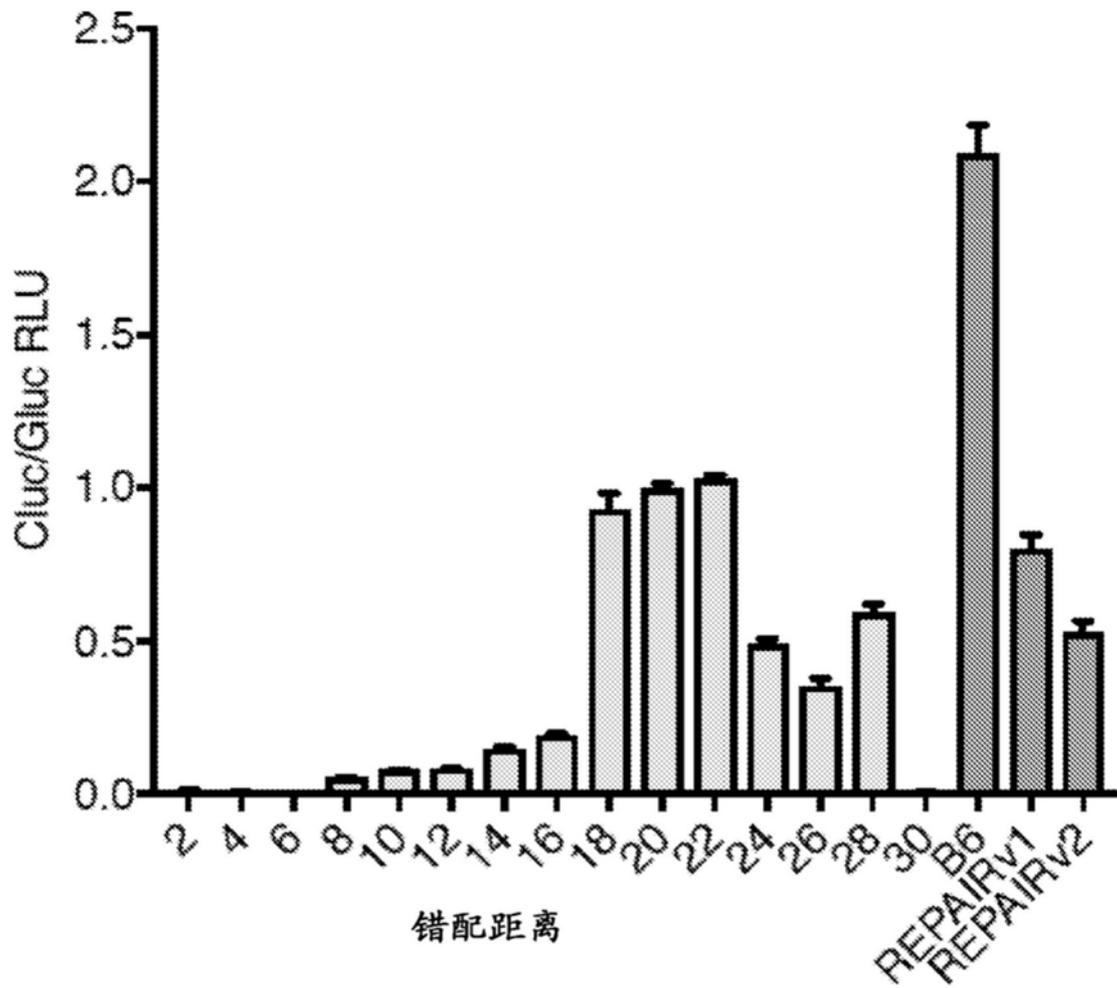


图70

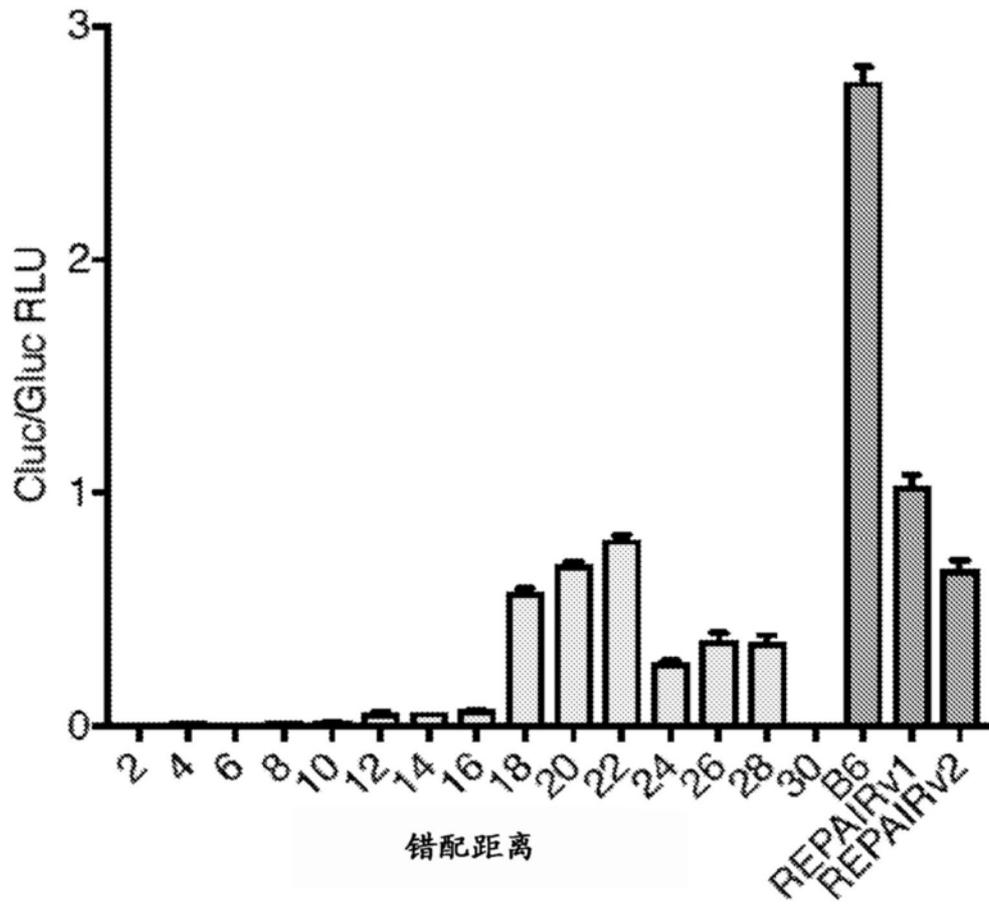


图71

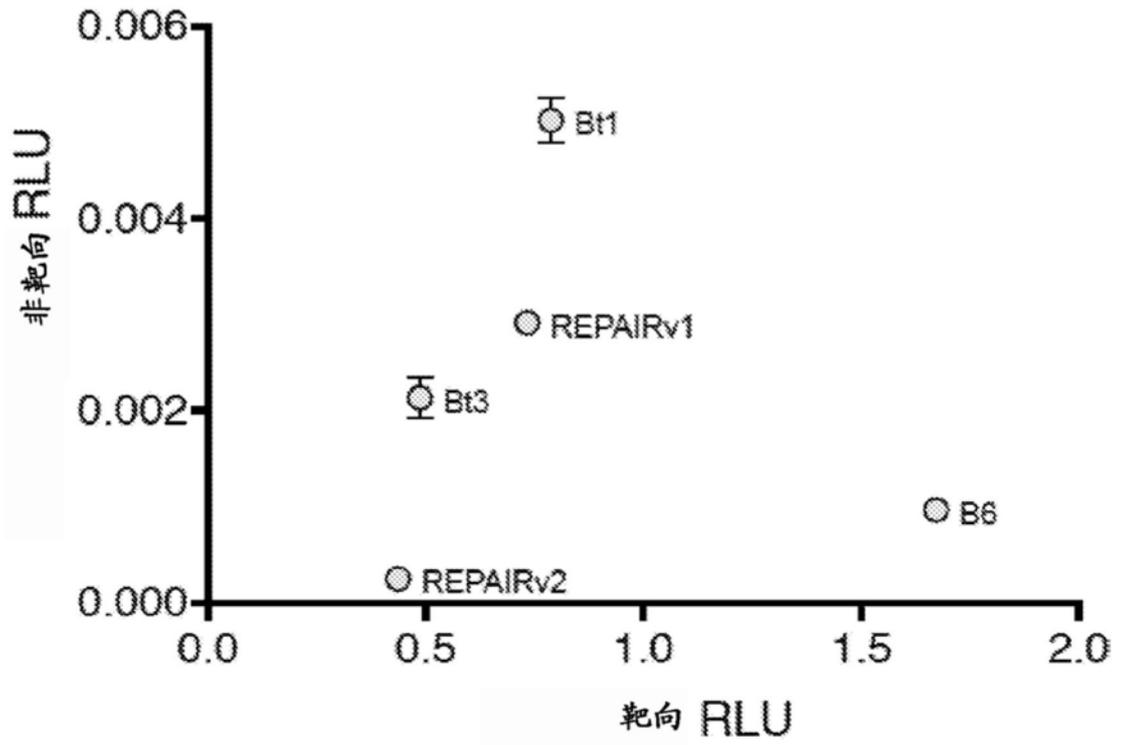


图72



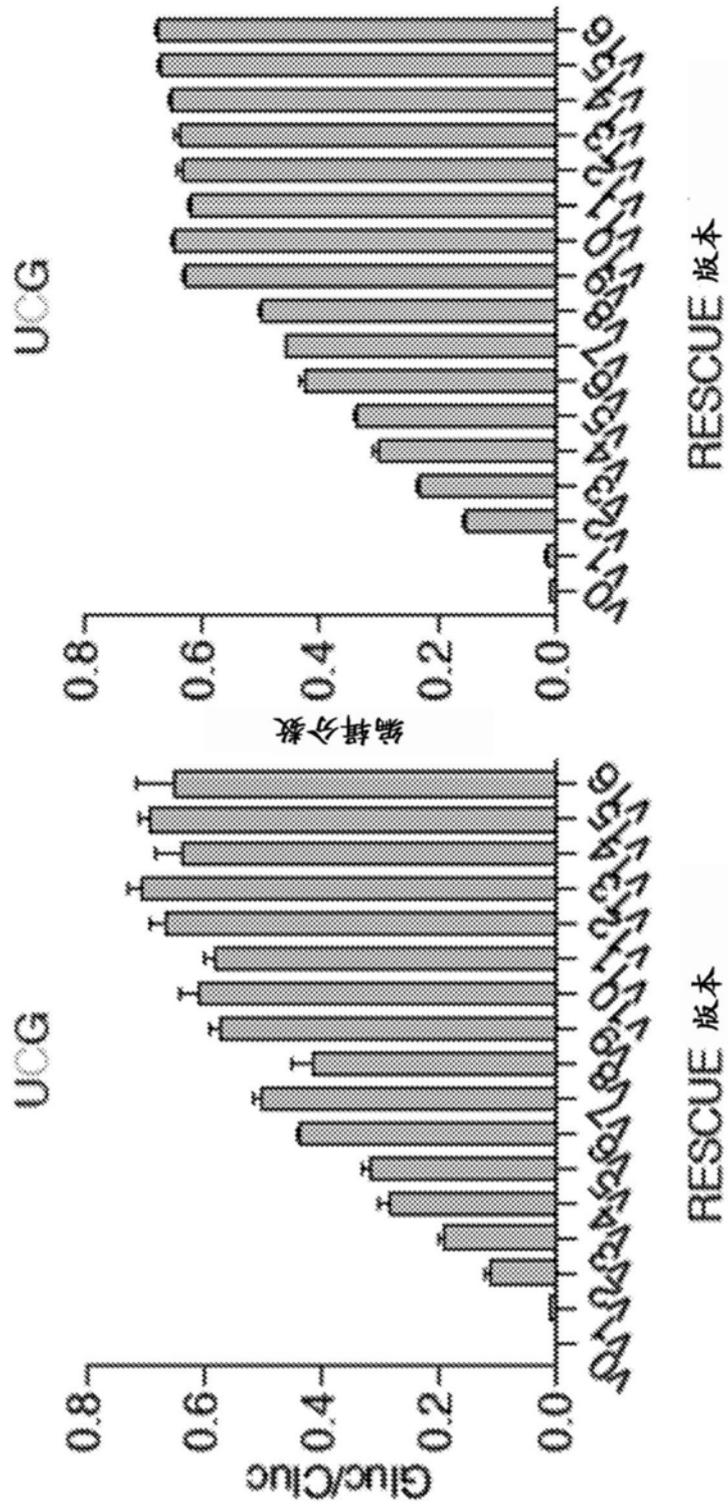


图73B

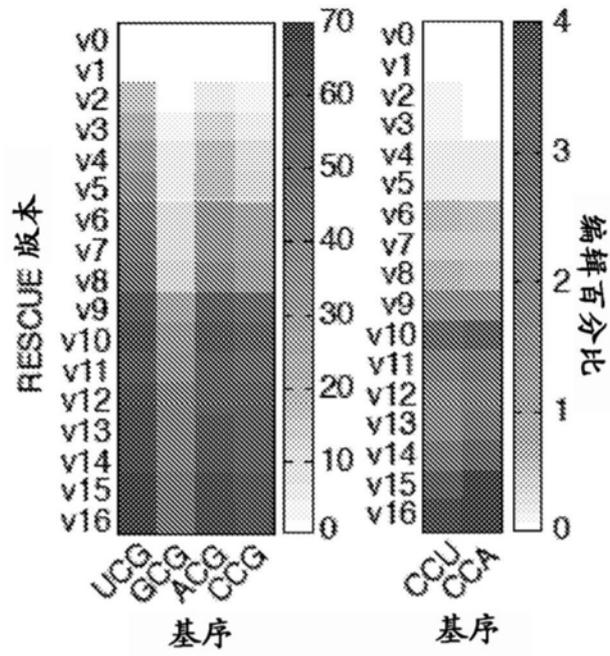


图73C

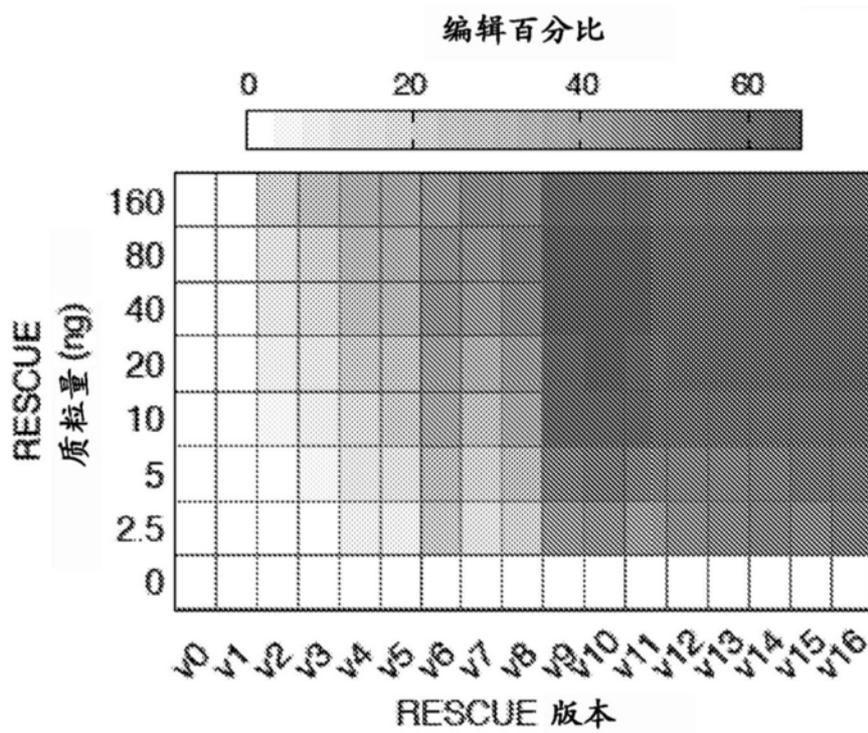


图73D

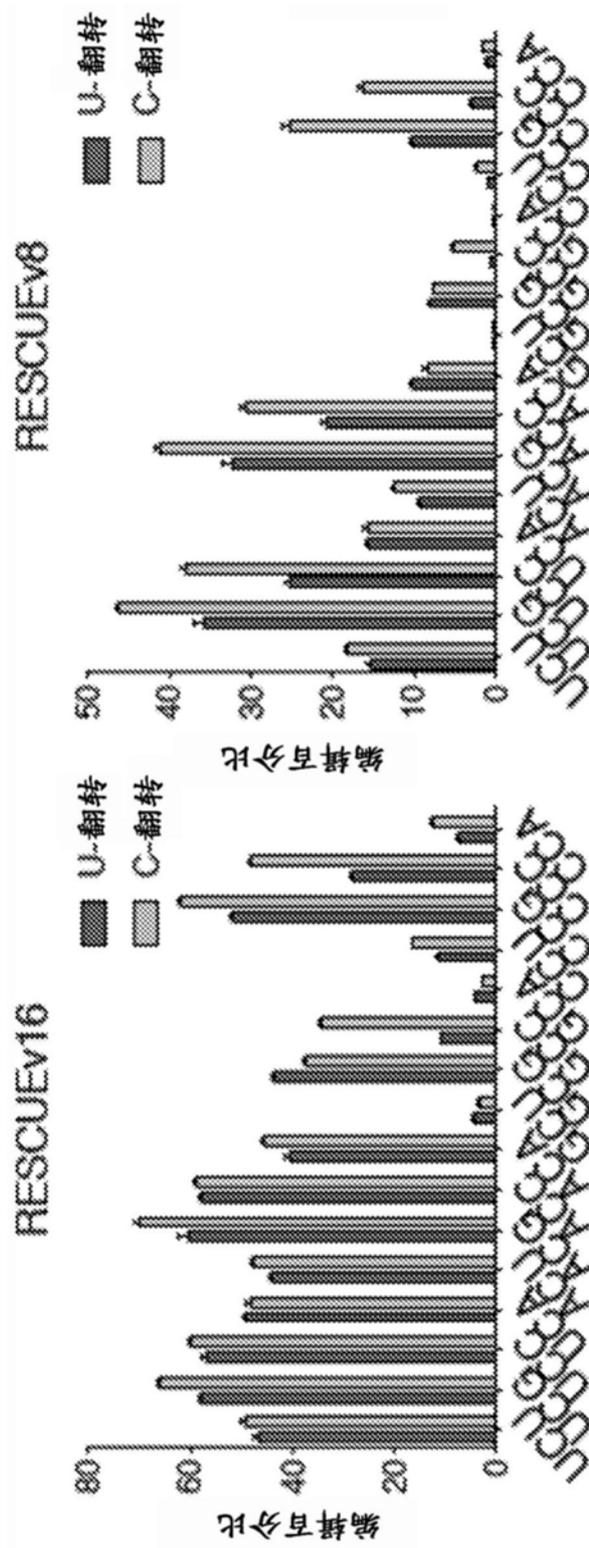


图73E

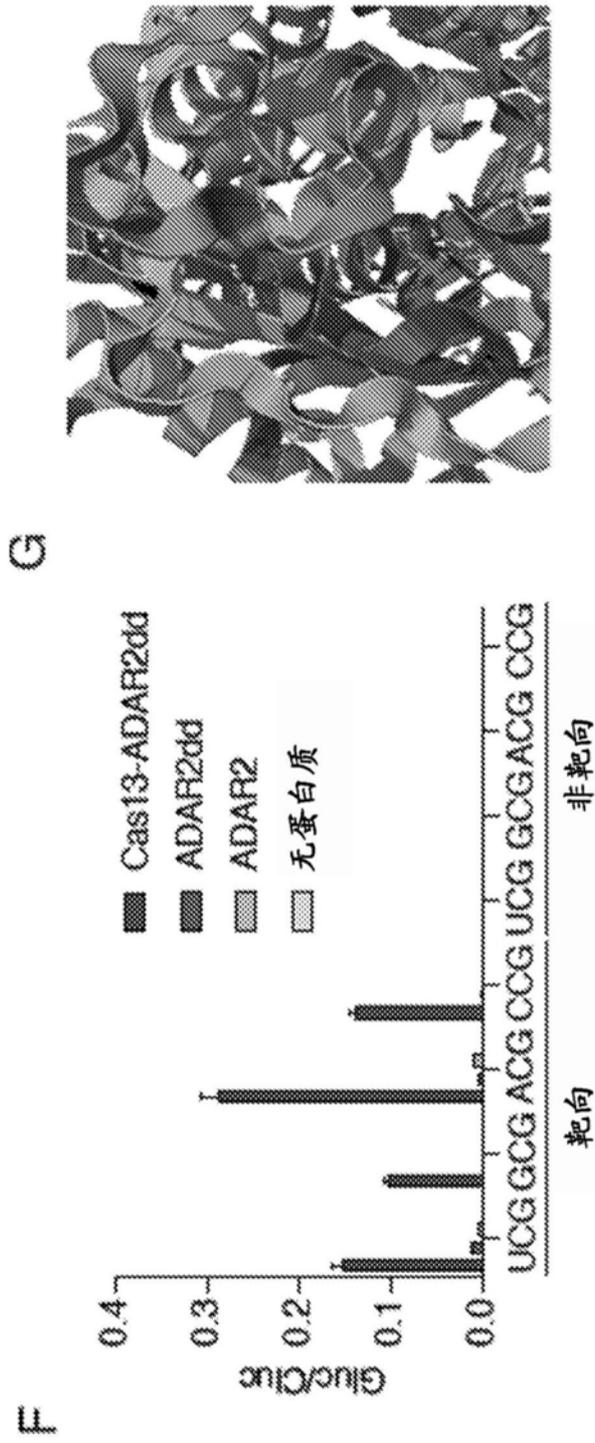


图73F图73G

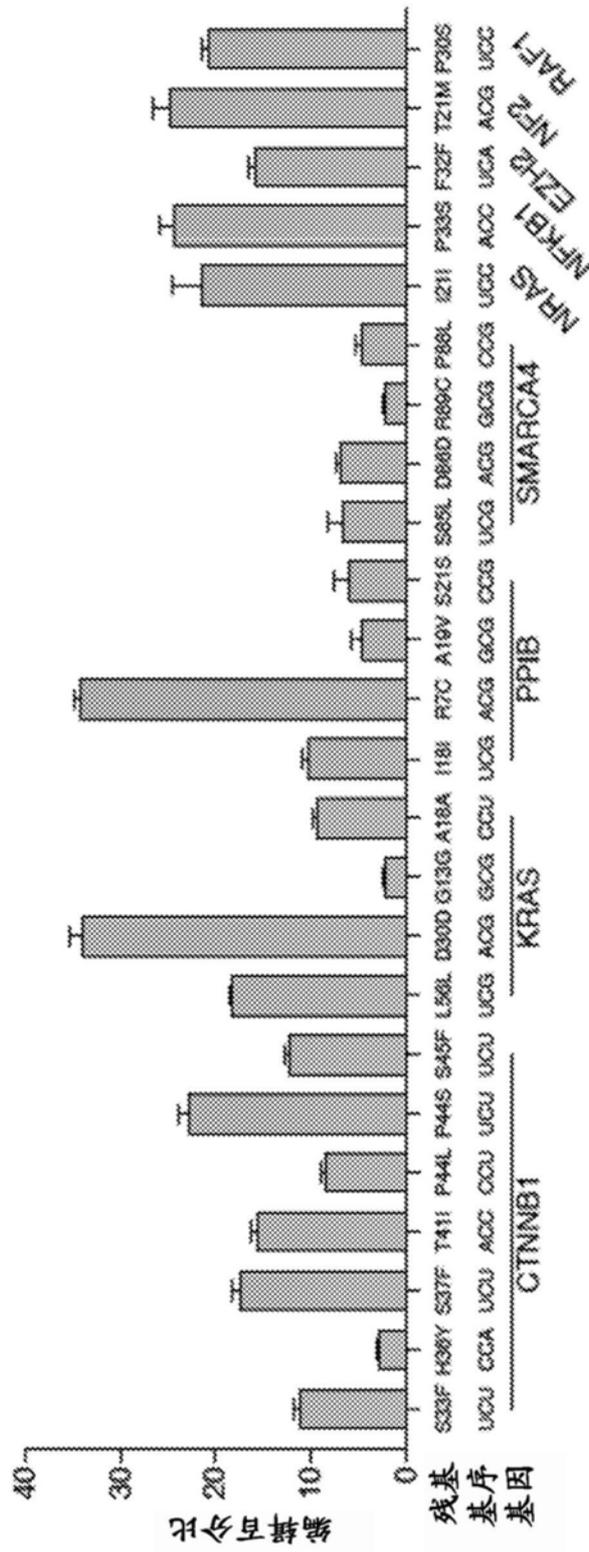


图74A

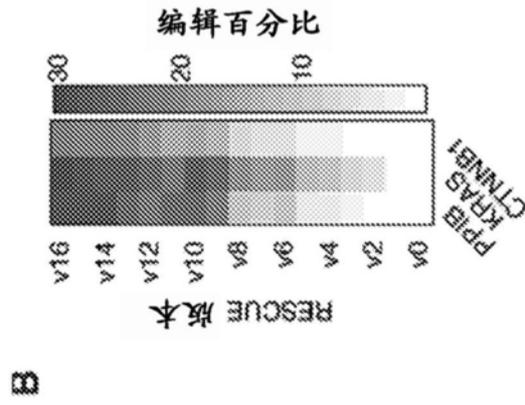


图74B



图74C

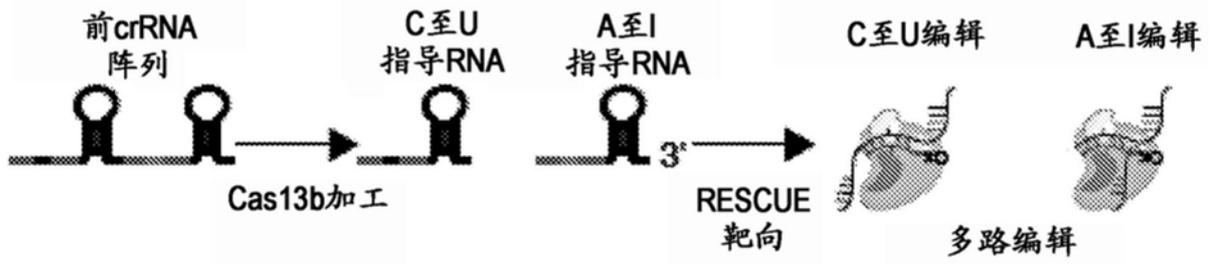


图74D

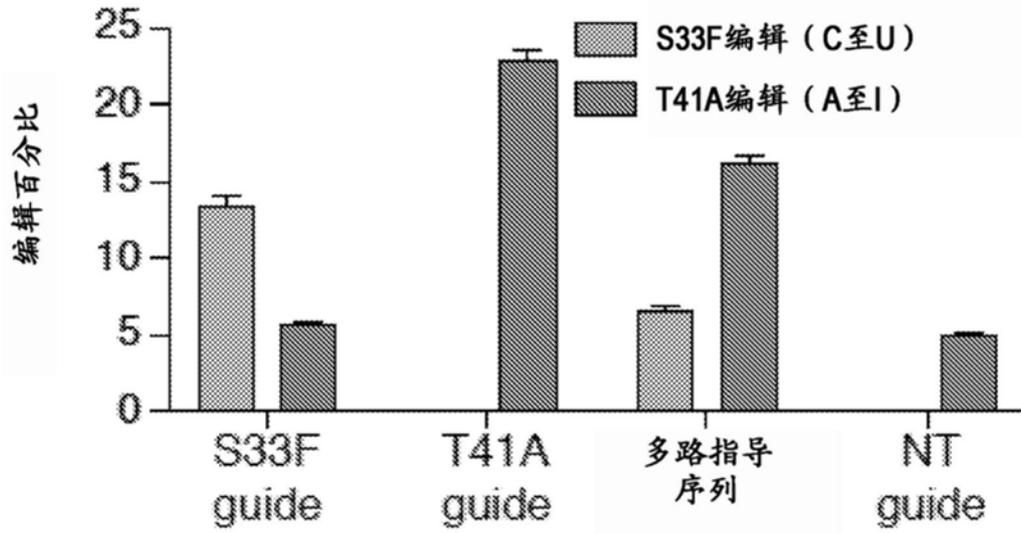


图74E

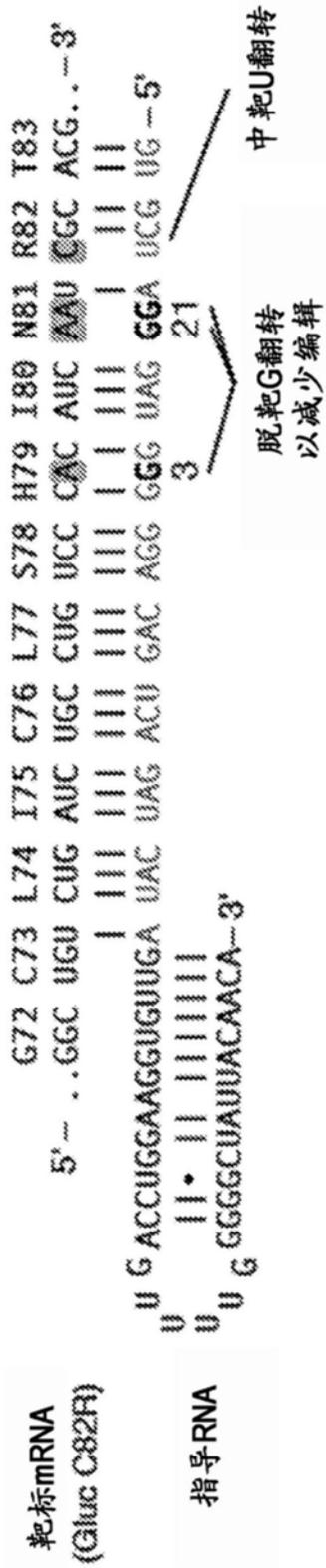


图74F

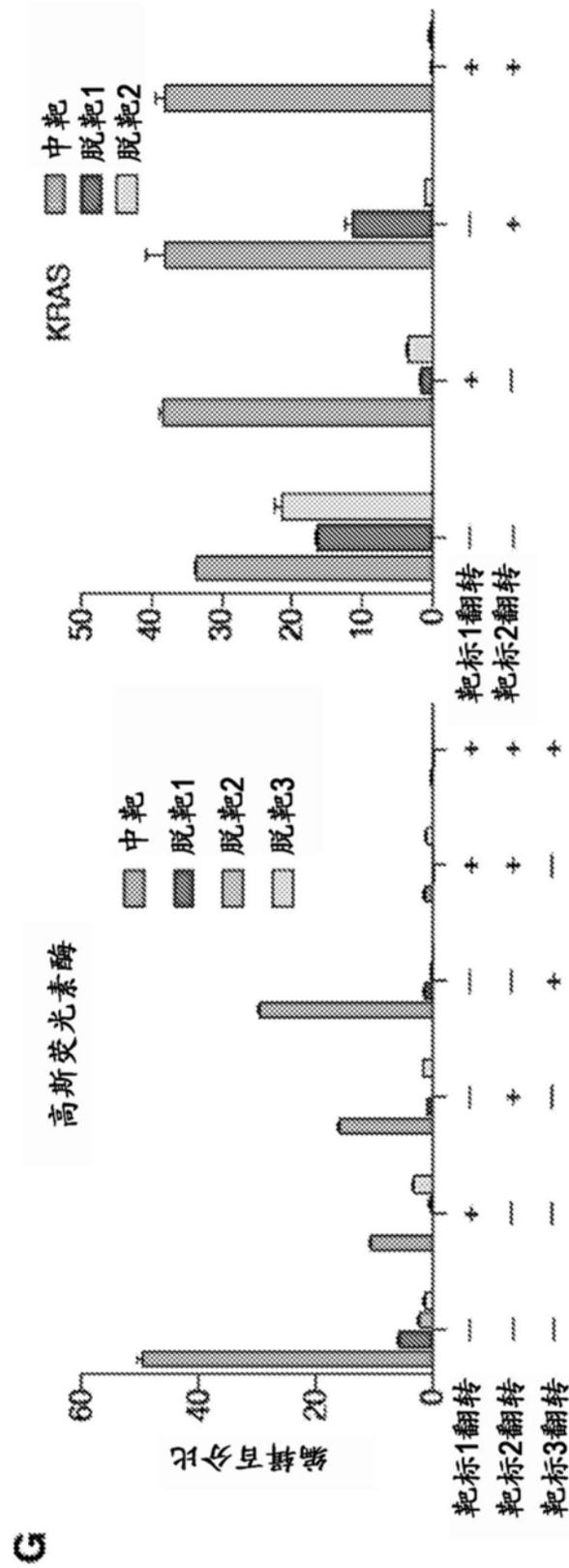


图74G

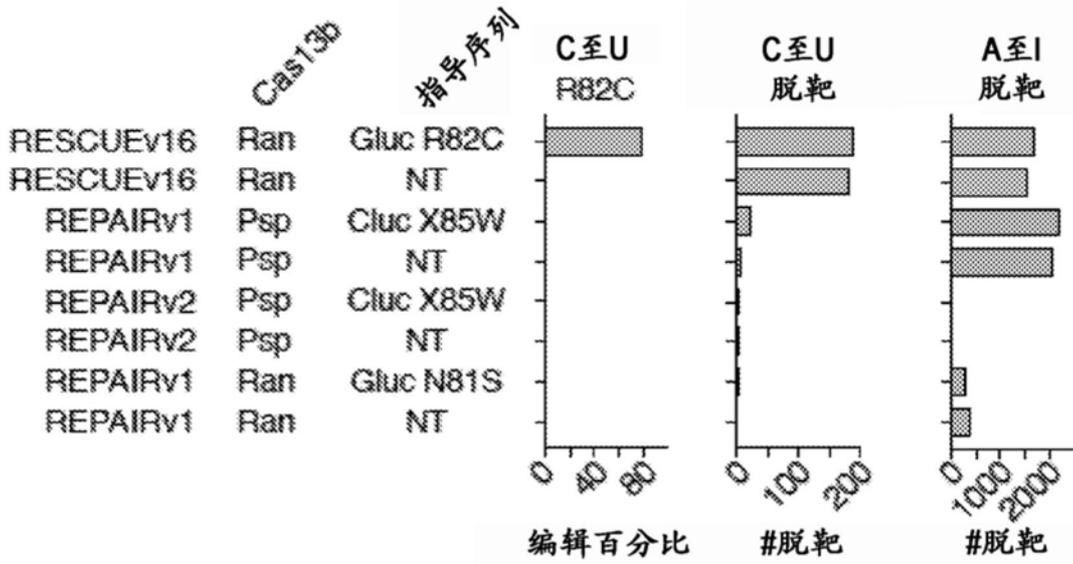


图75A

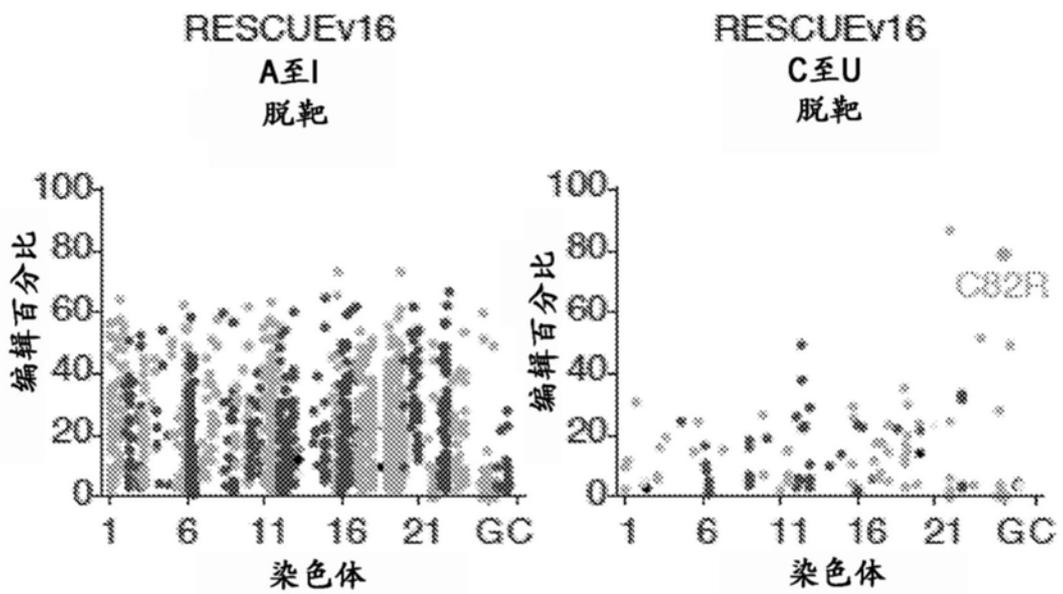


图75B



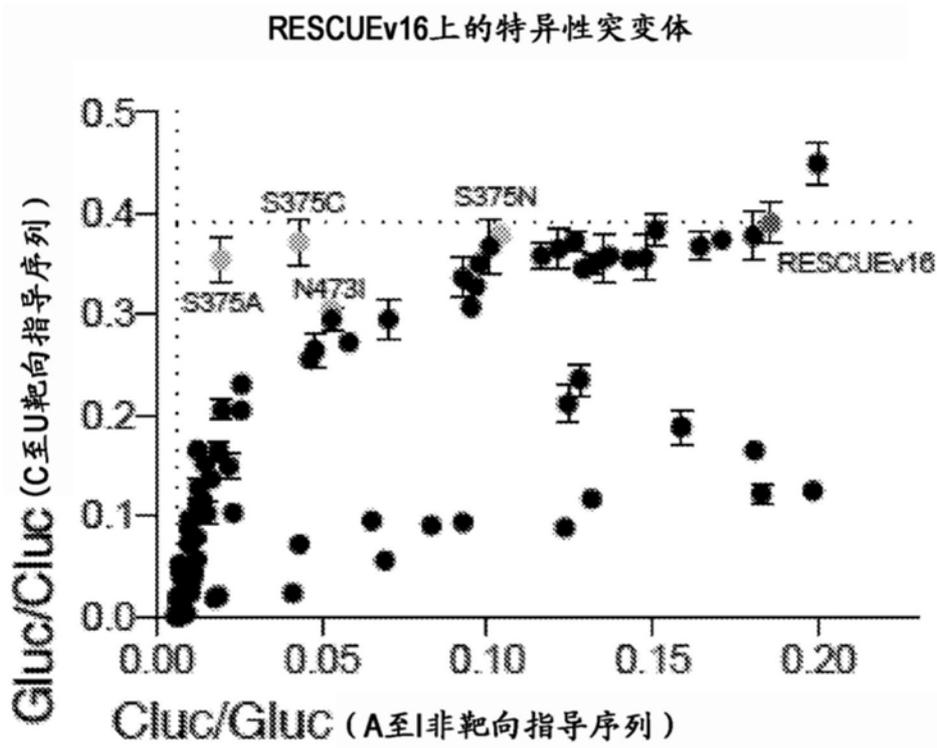


图75D

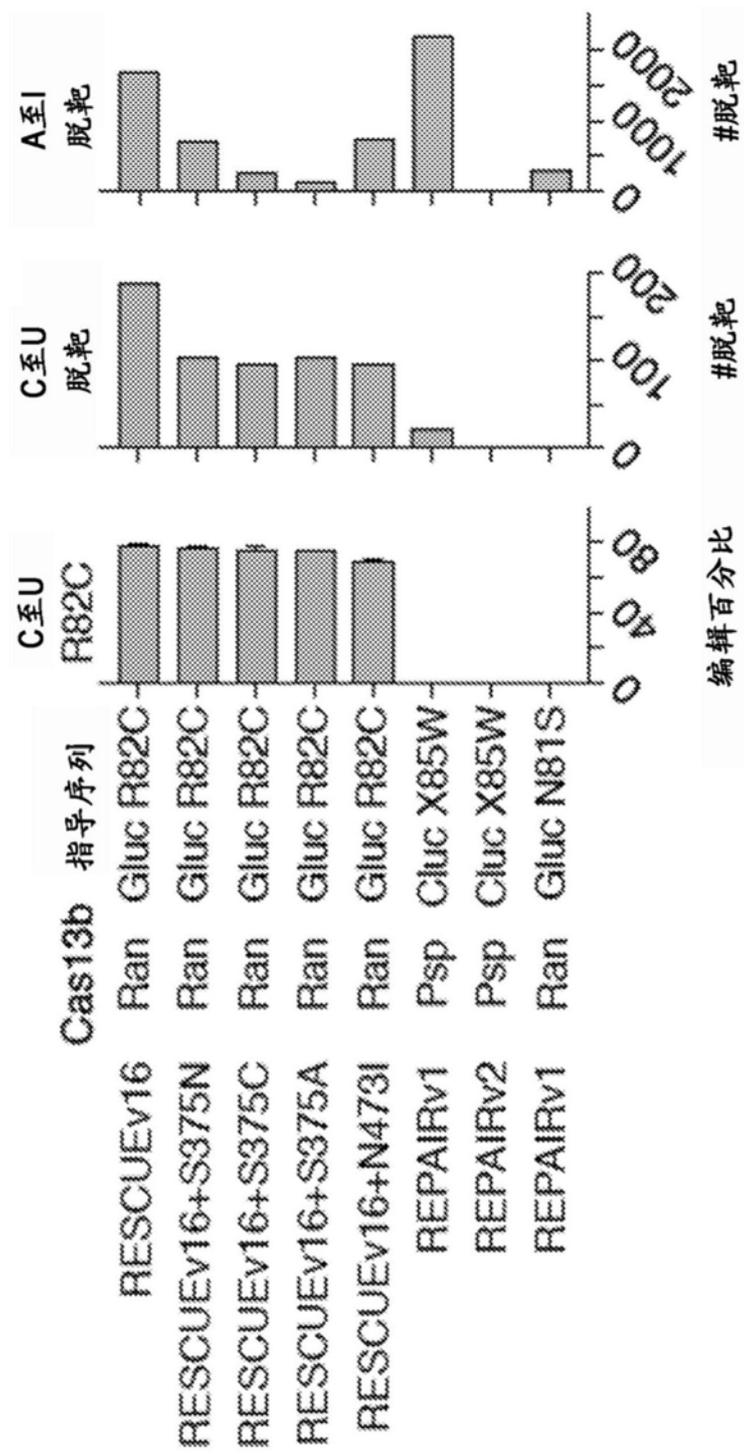


图75E

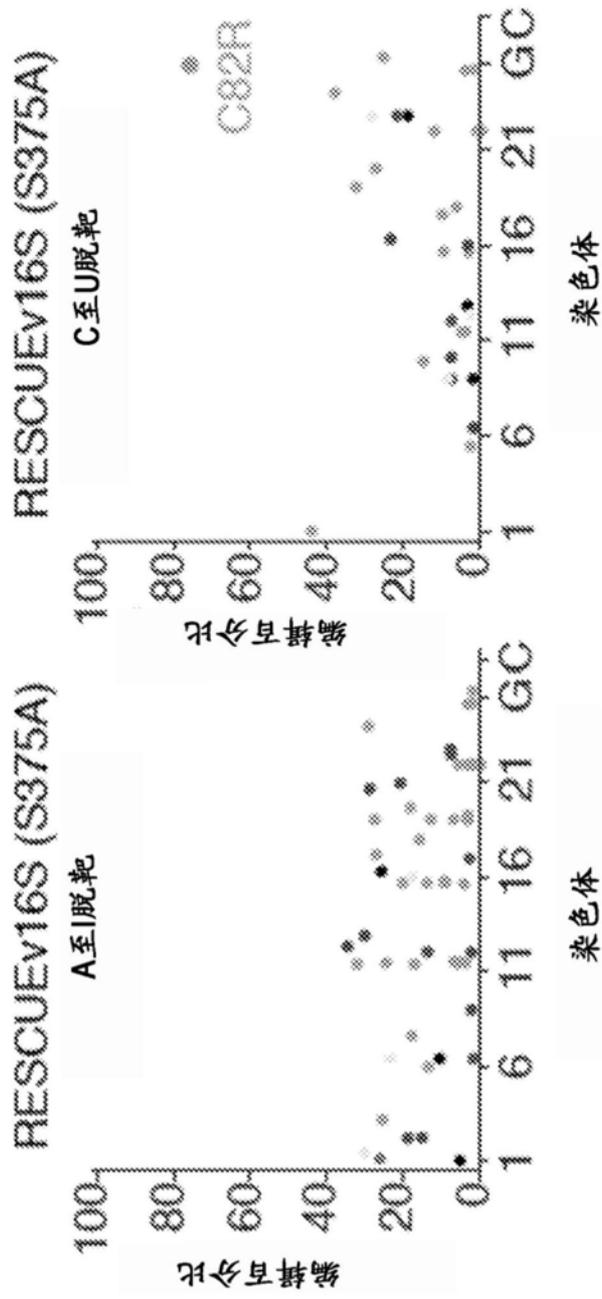


图75F



图76A

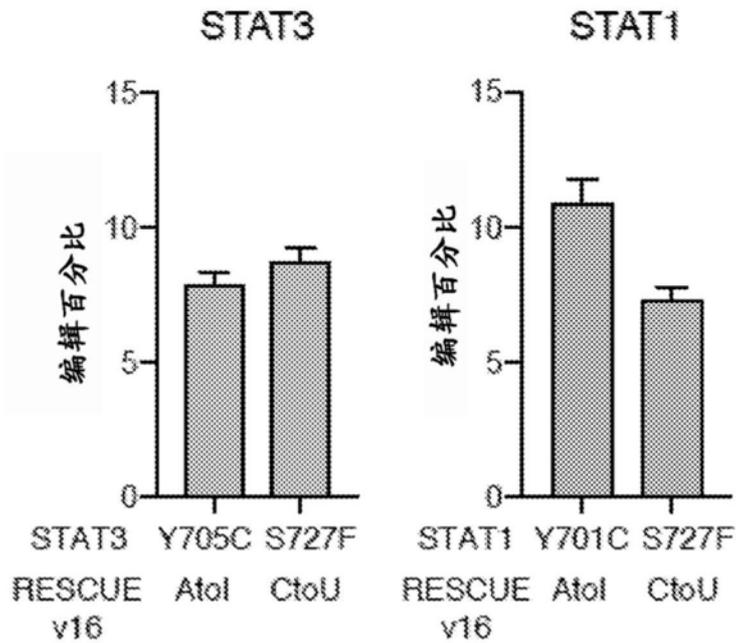


图76B

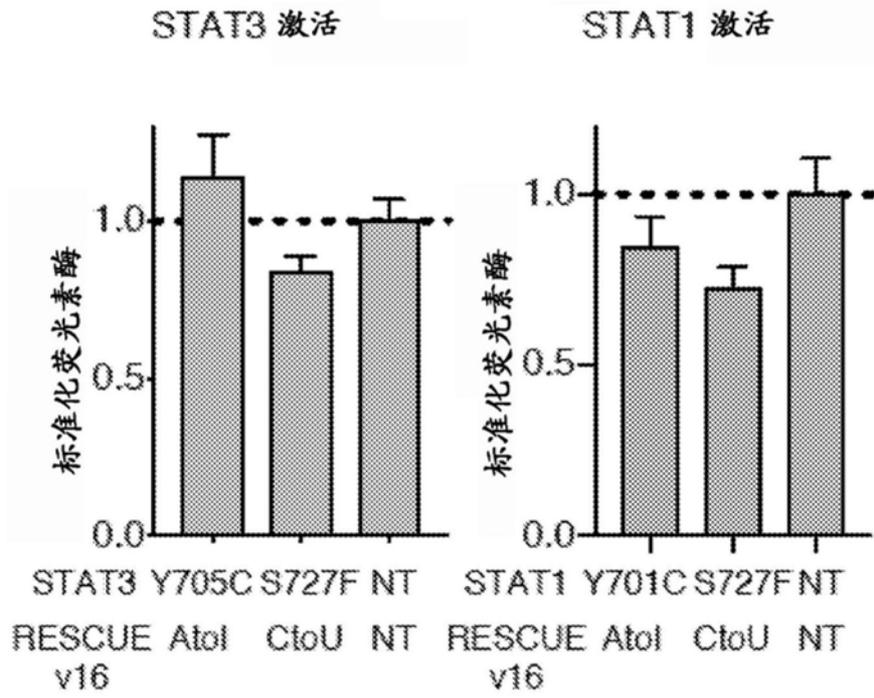


图76C

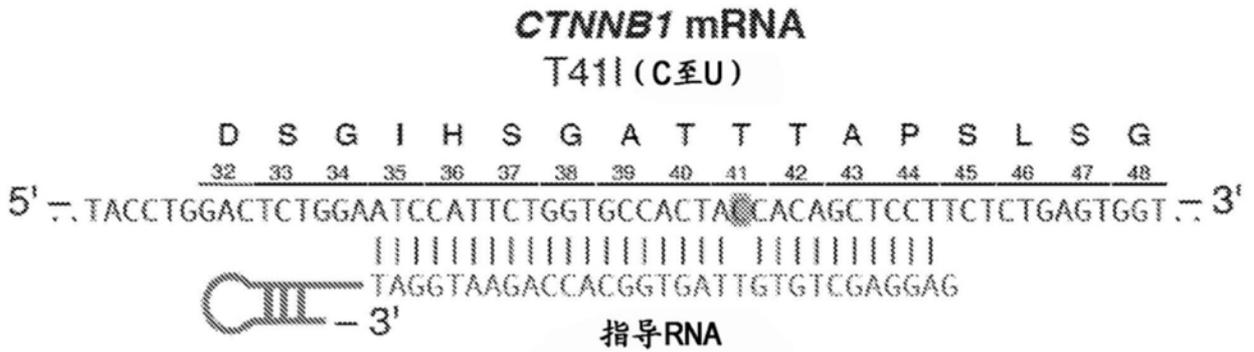


图76D

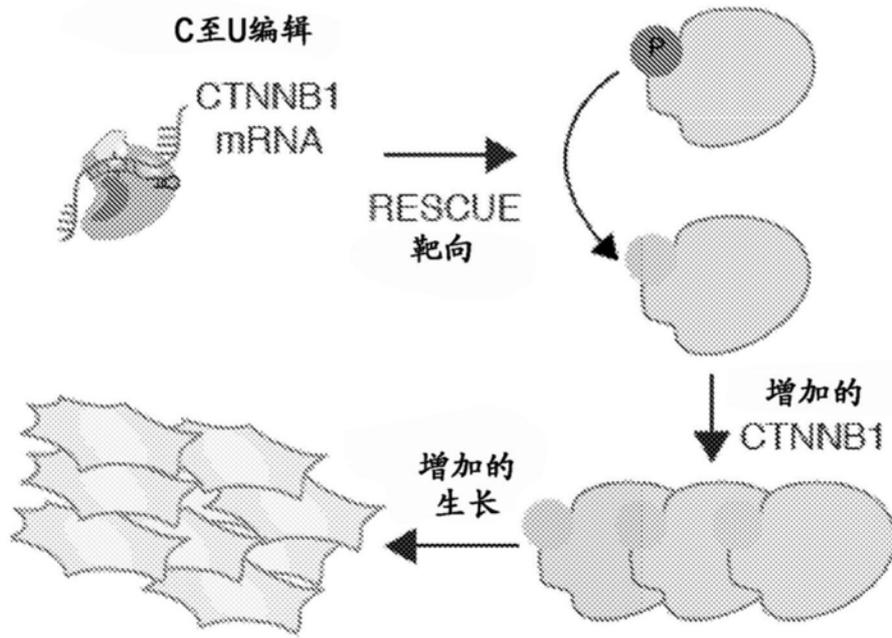


图76E

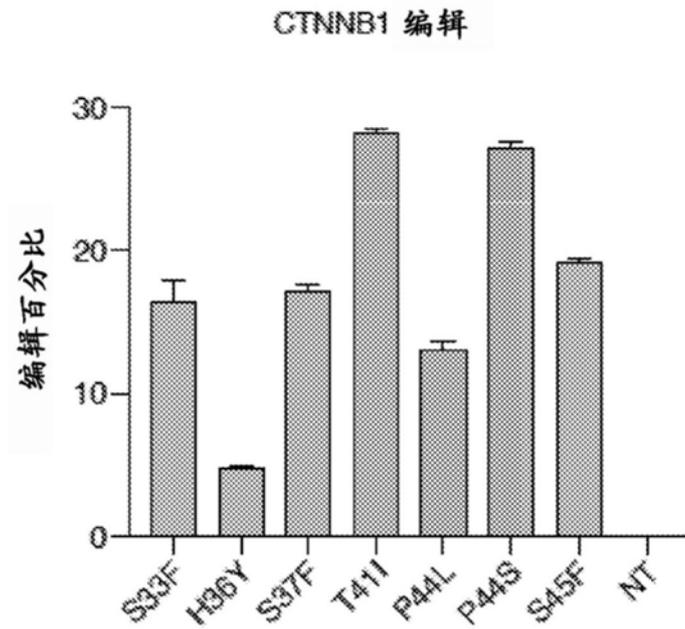


图76F

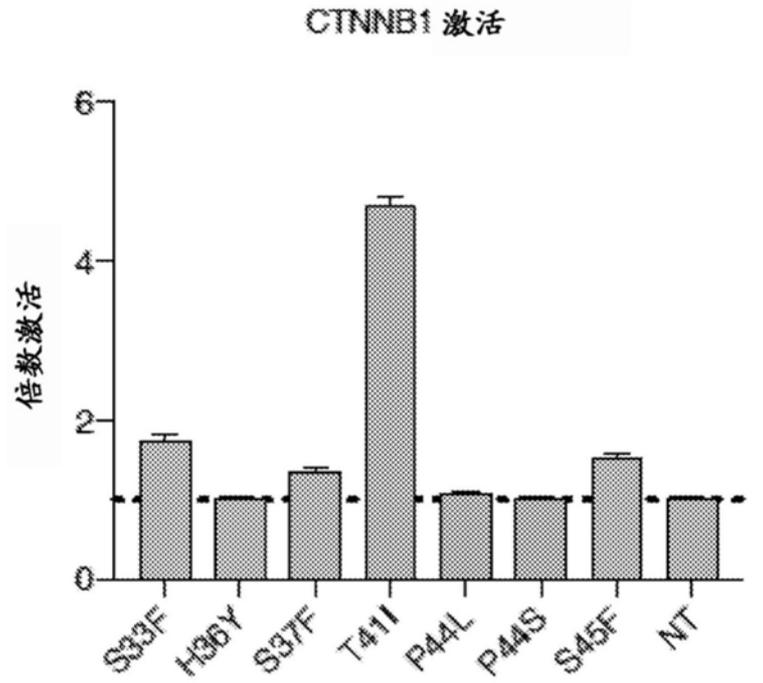


图76G

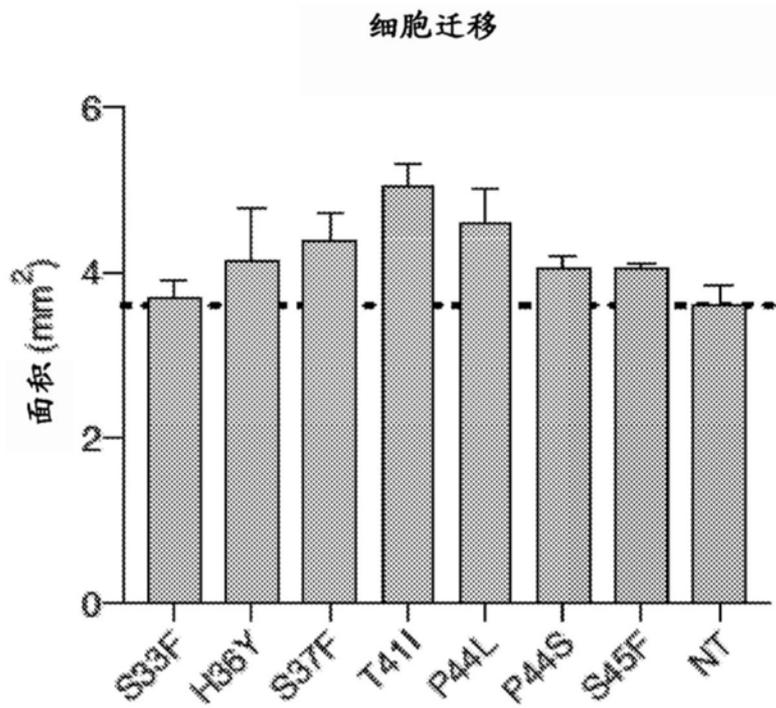


图76H

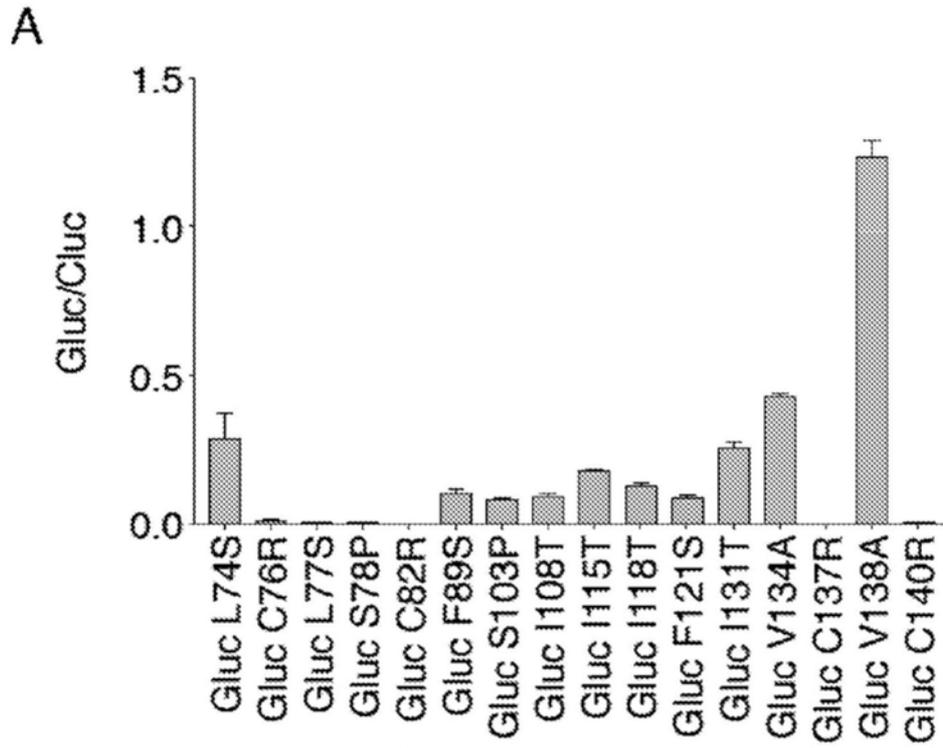


图77A

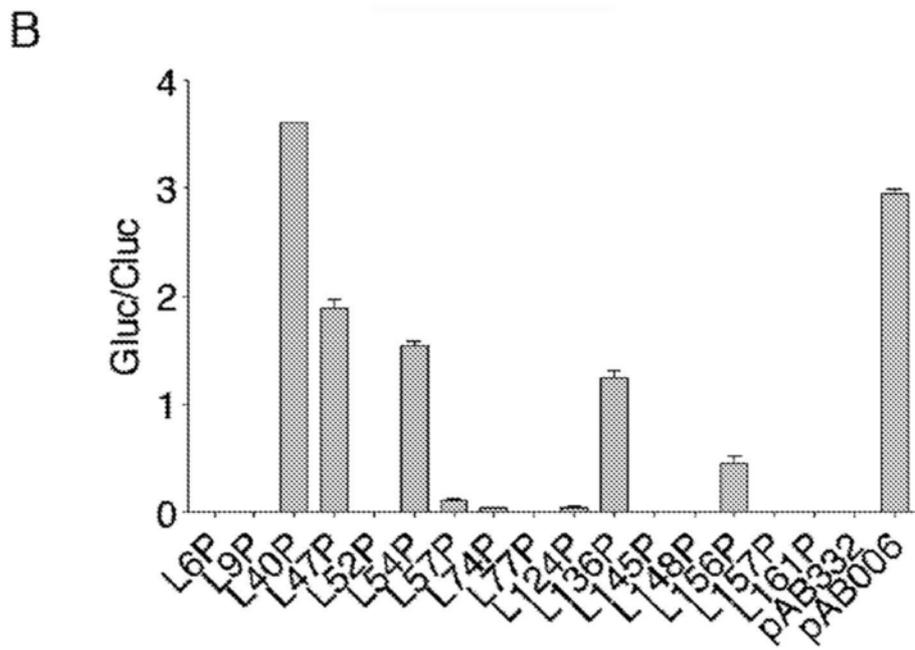


图77B

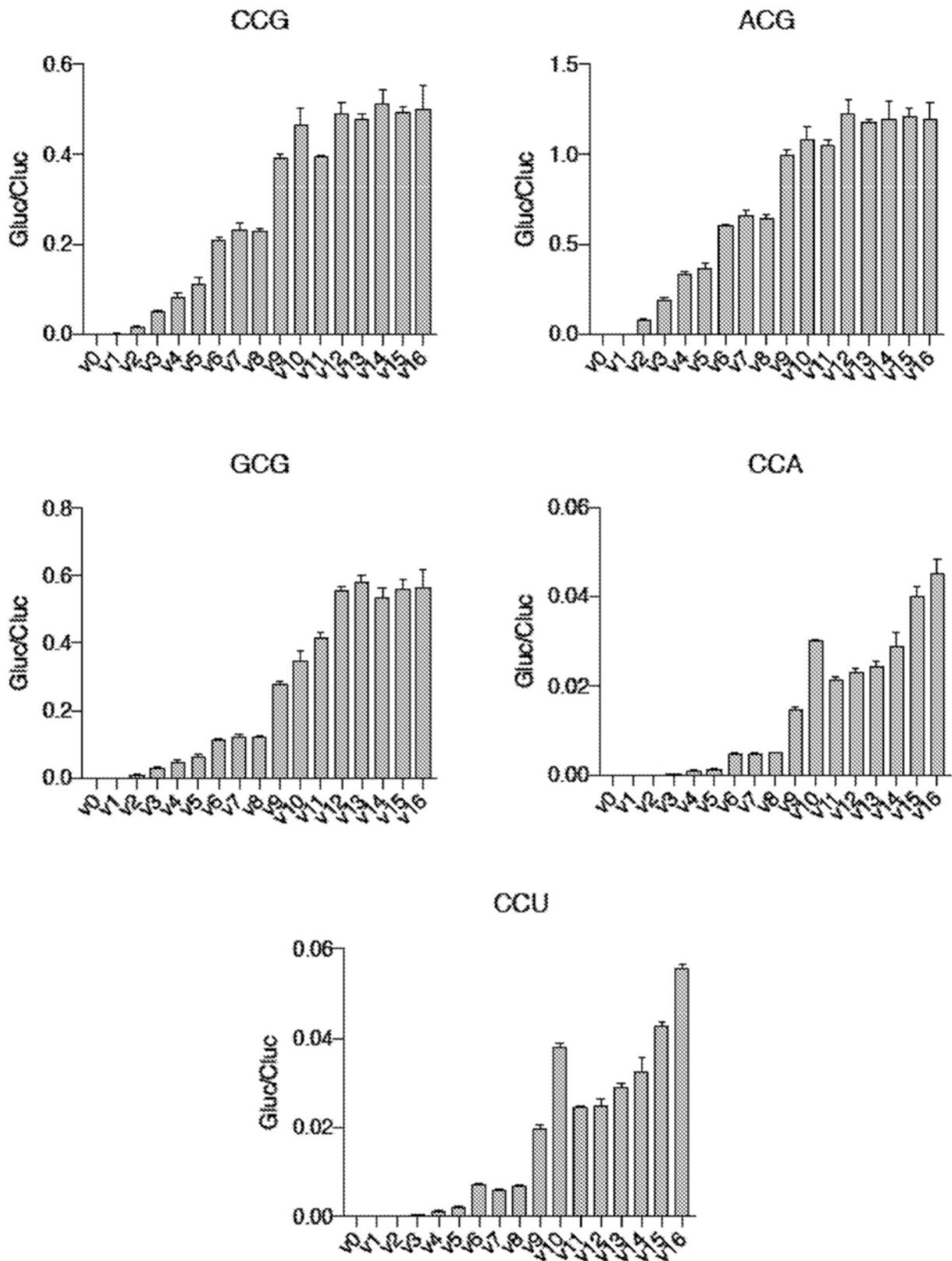


图78

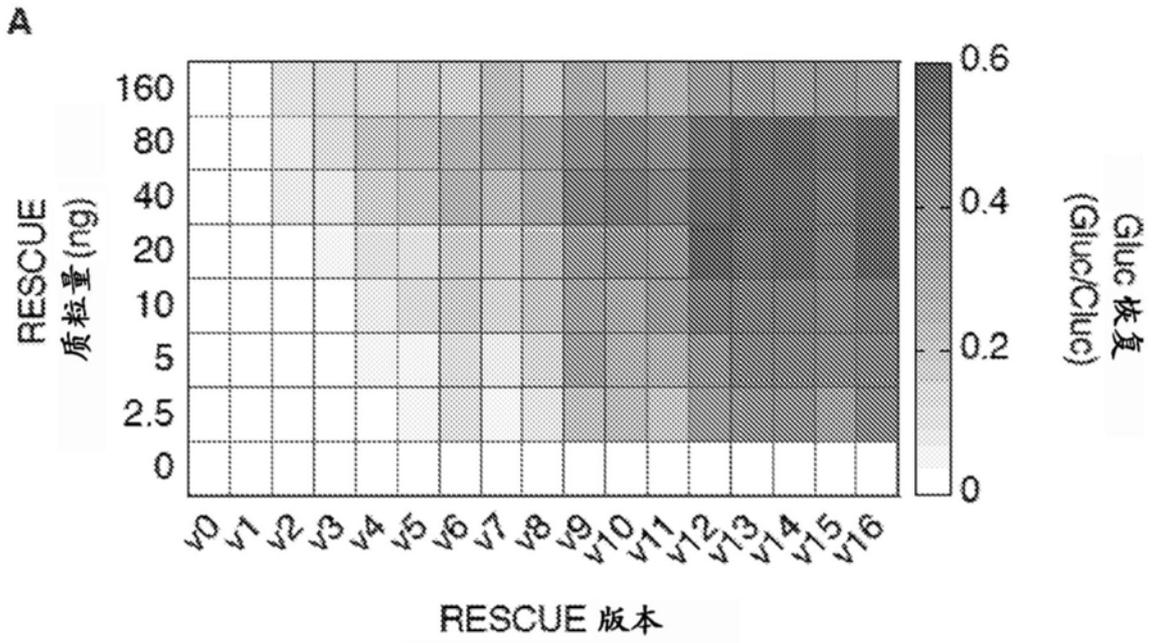


图79A

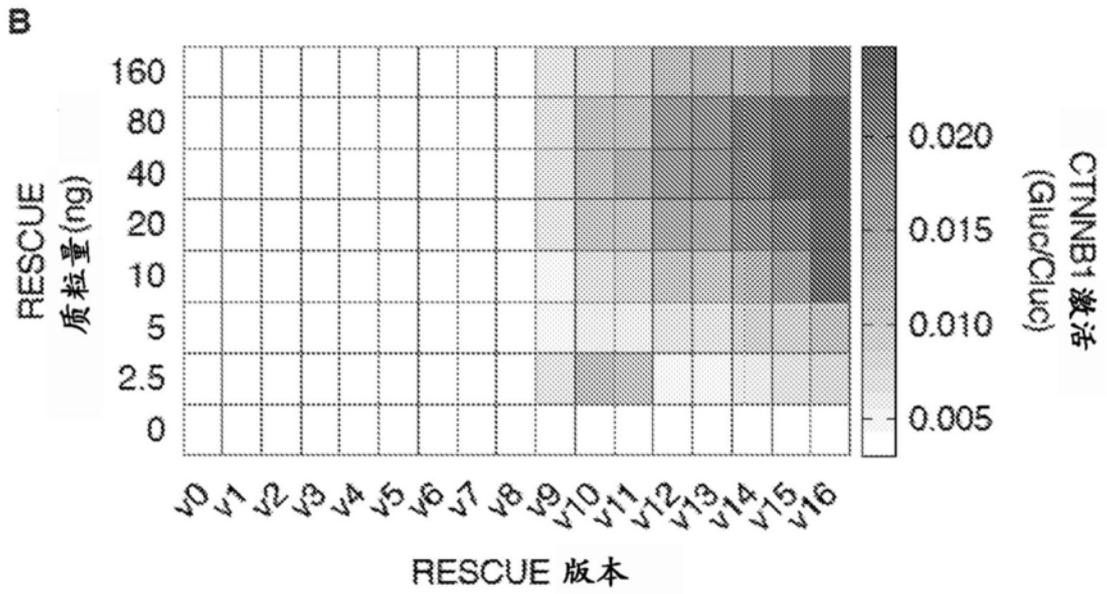


图79B

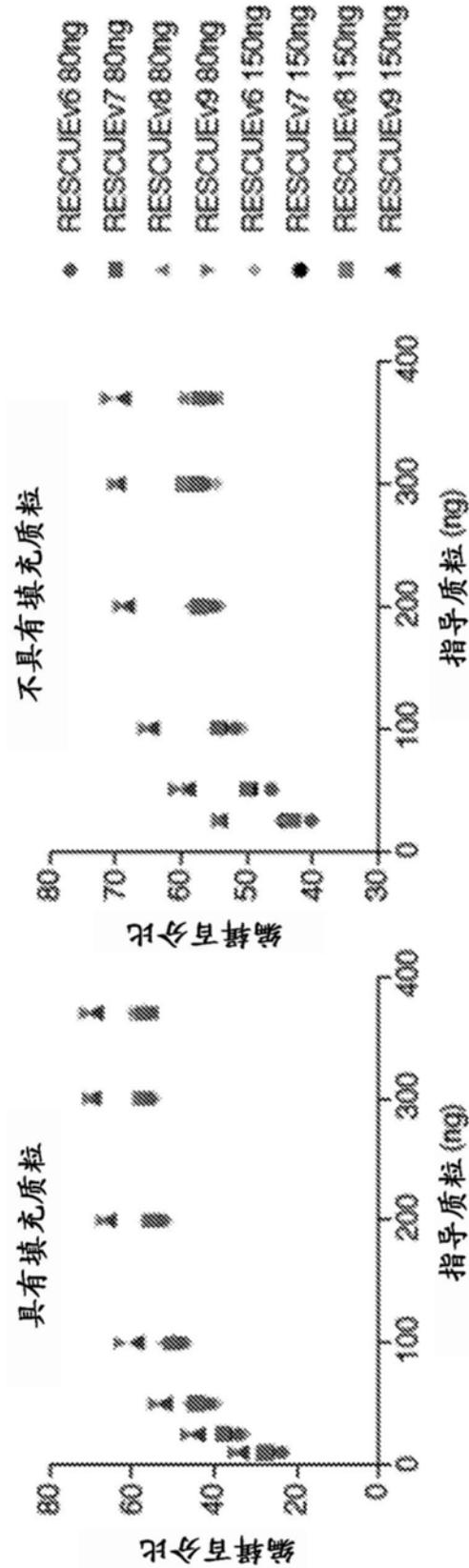


图80



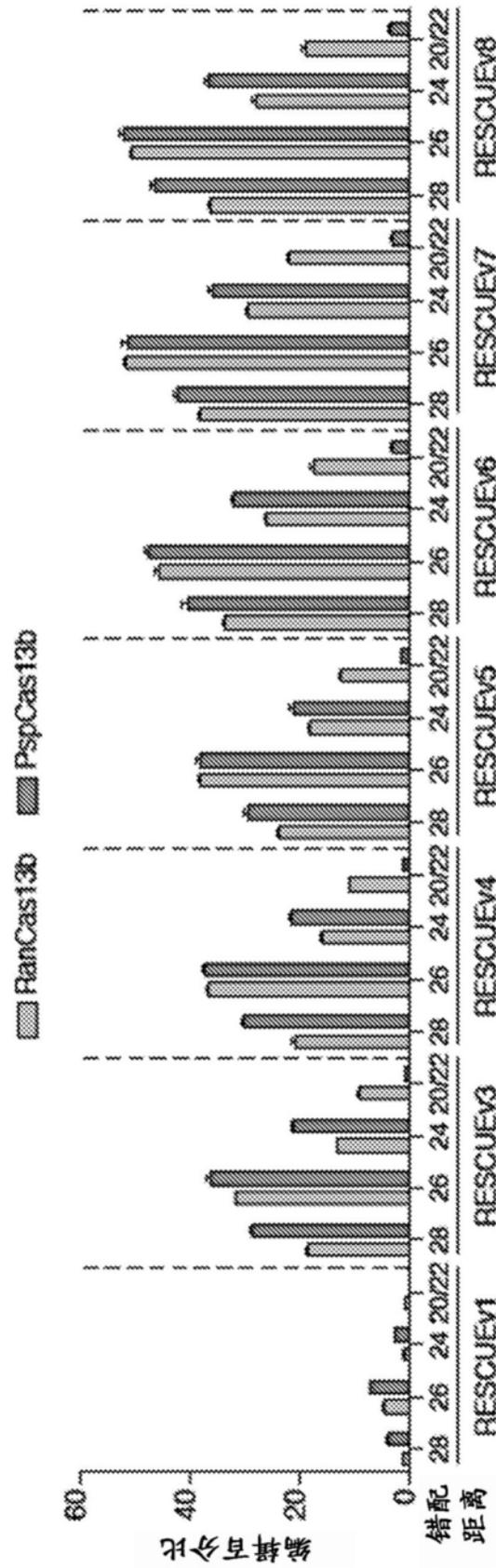


图82

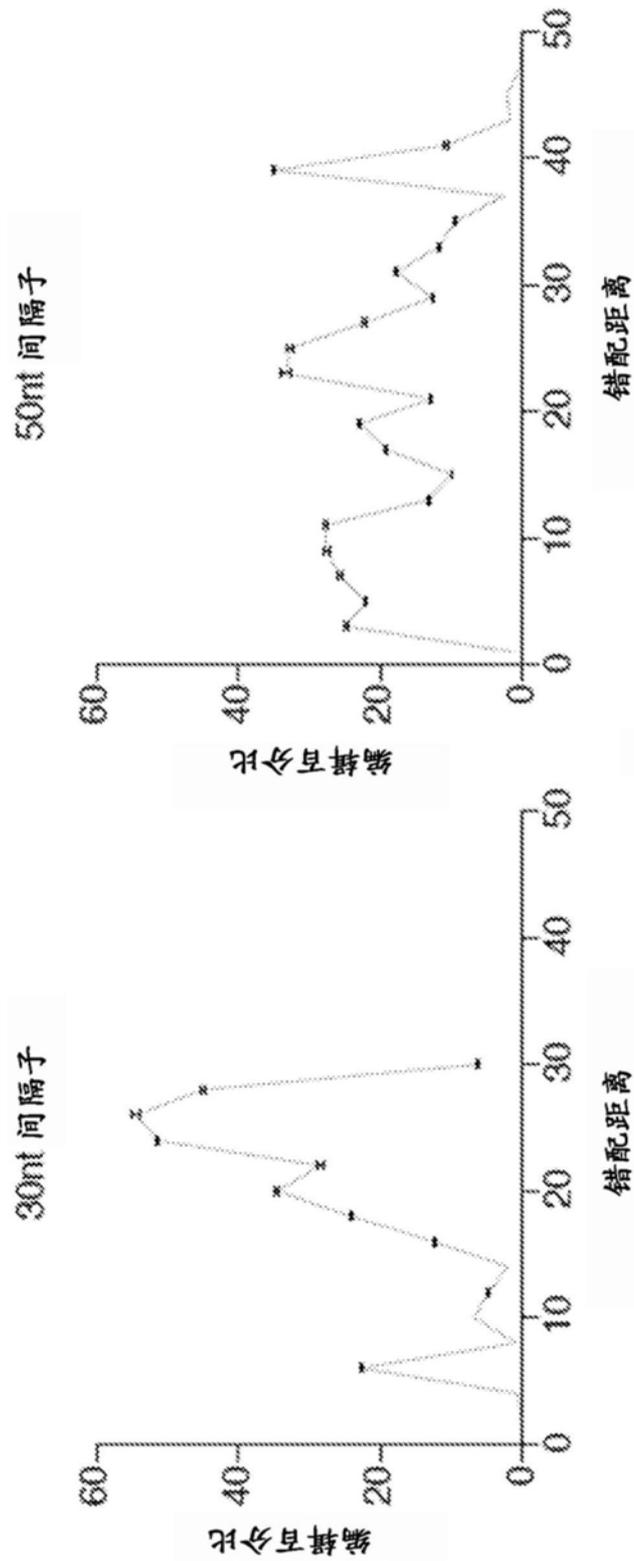


图83

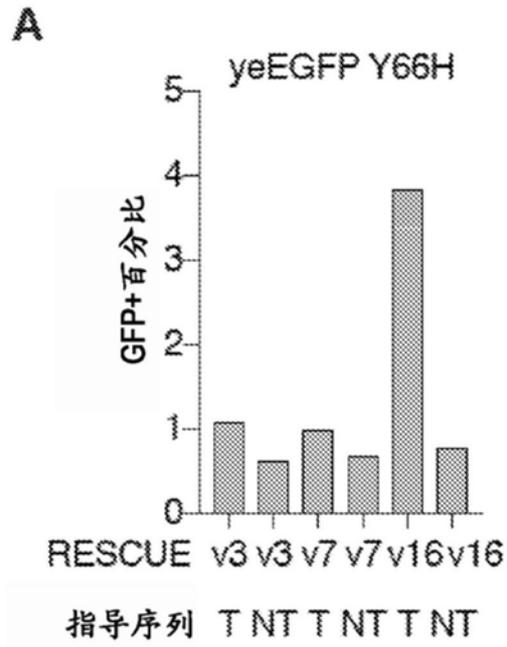


图84A

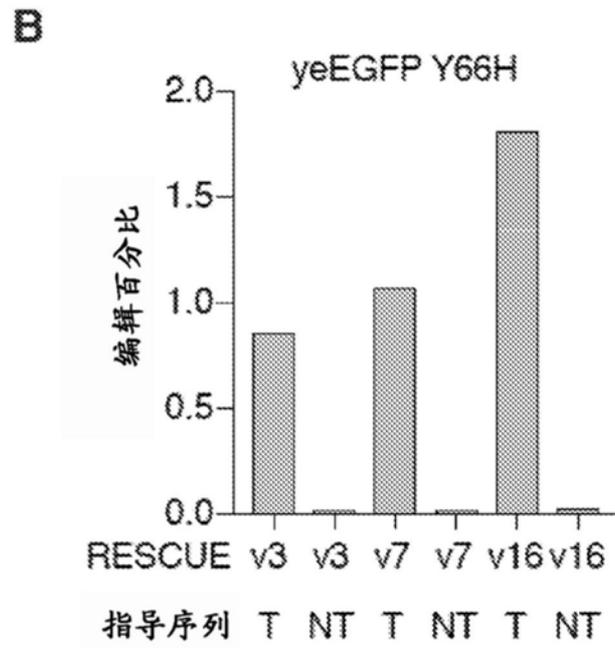


图84B

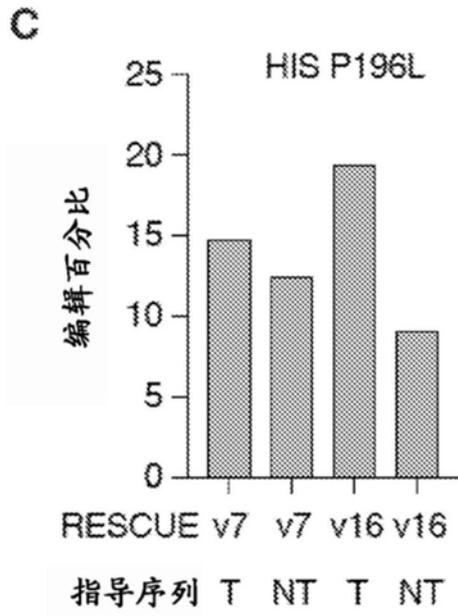


图84C

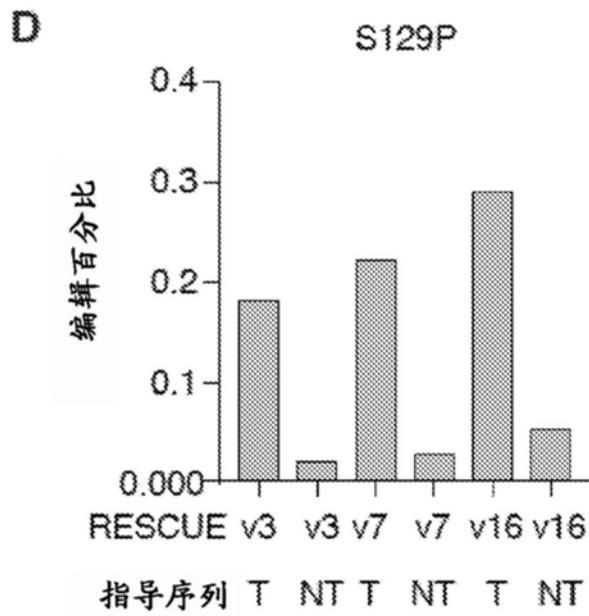


图84D

FIG.

A

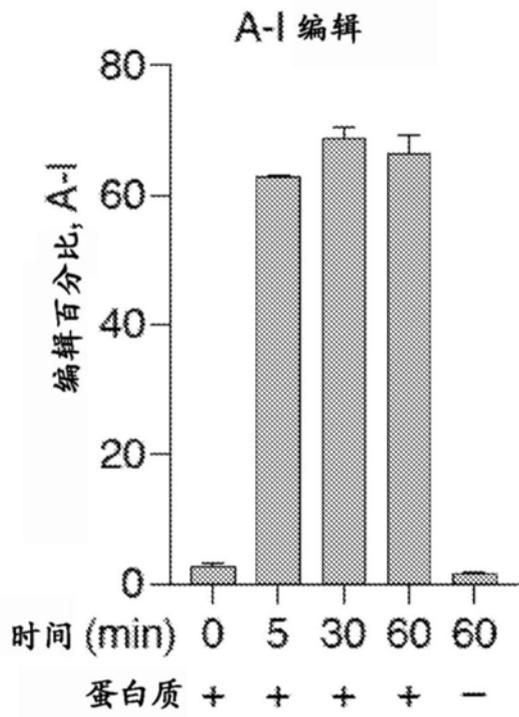


图85A

B

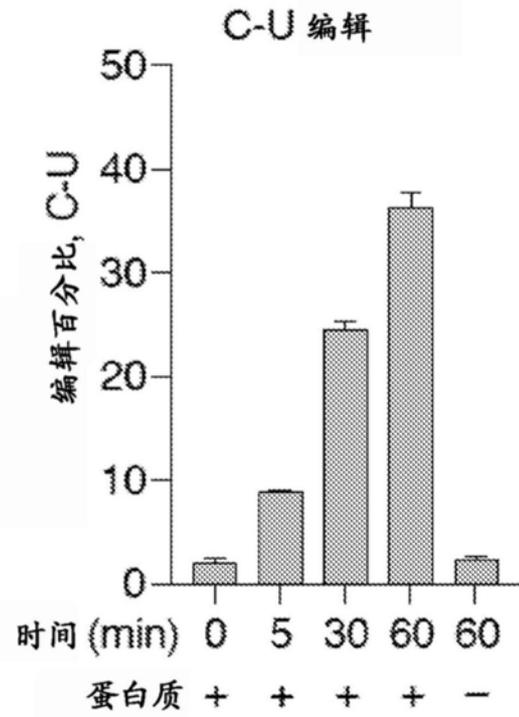


图85B

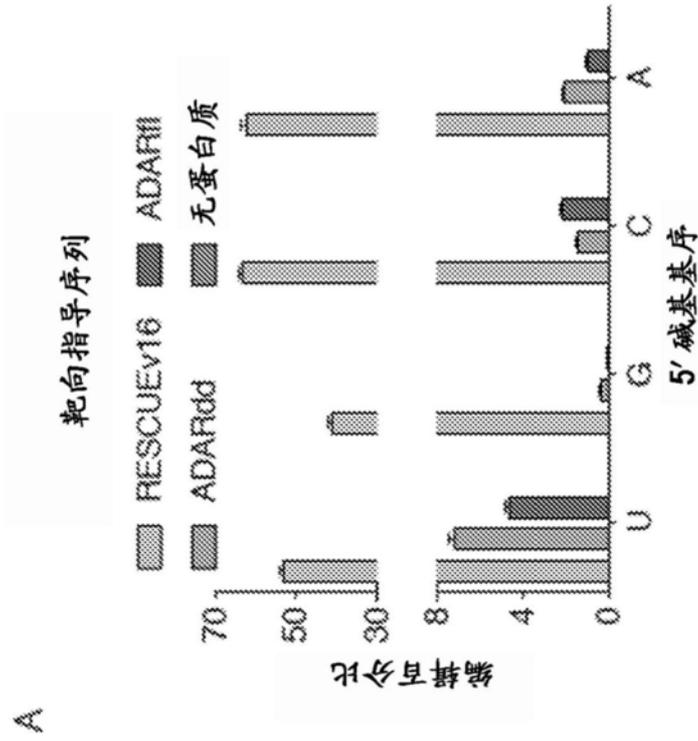


图86A

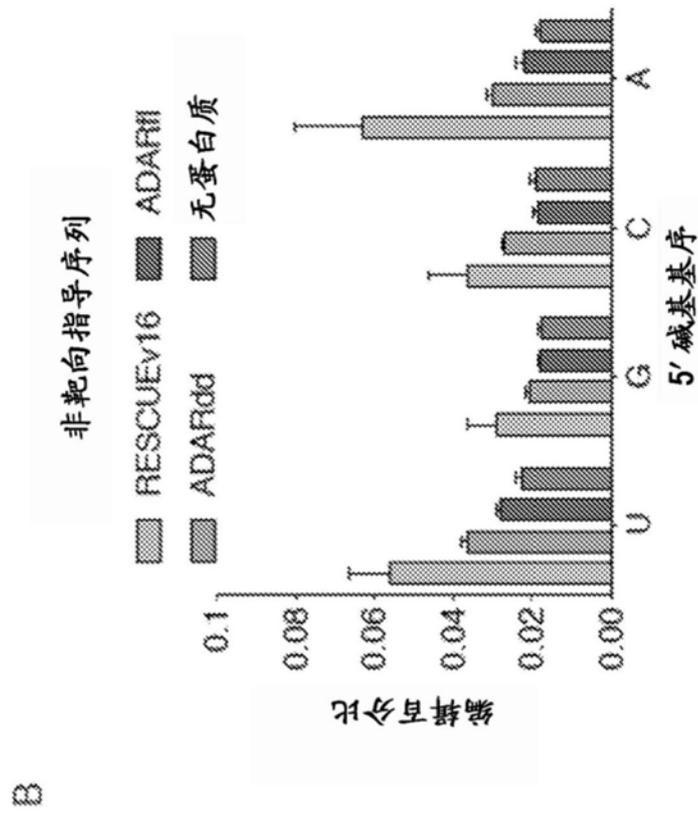


图86B

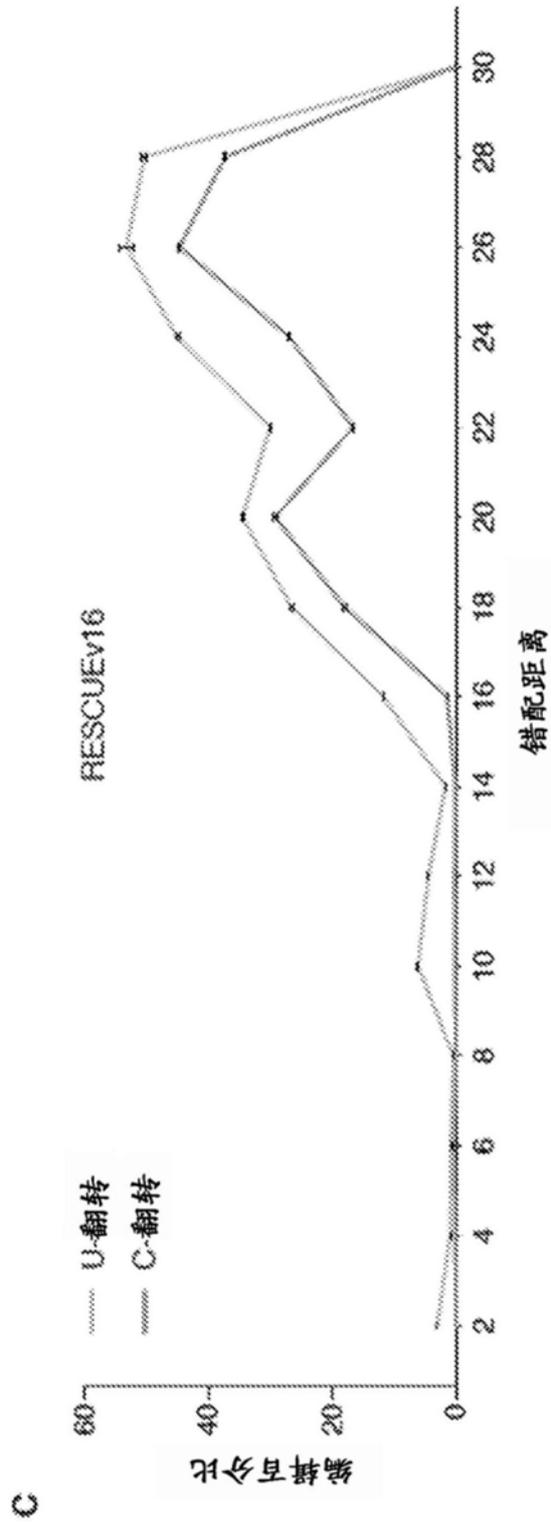


图86C

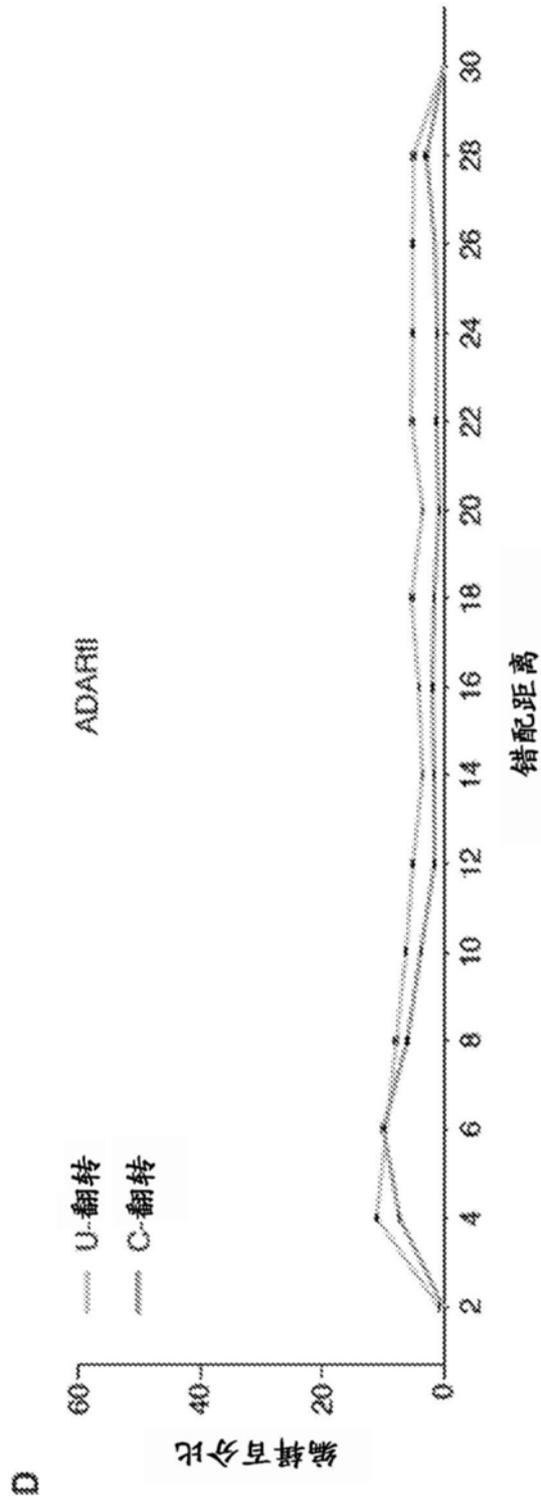


图86D

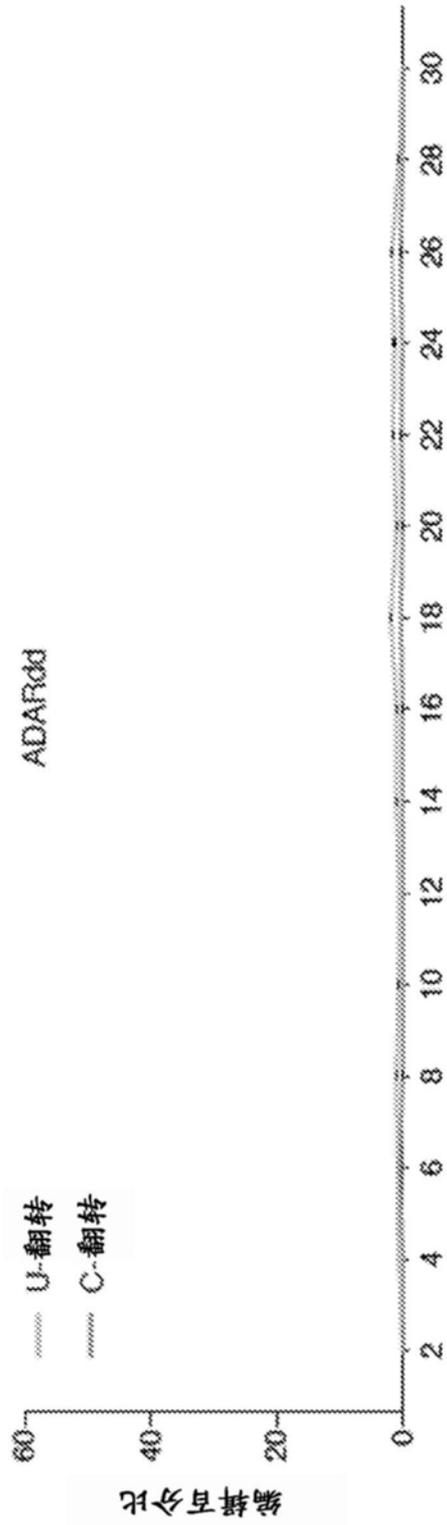


图86E

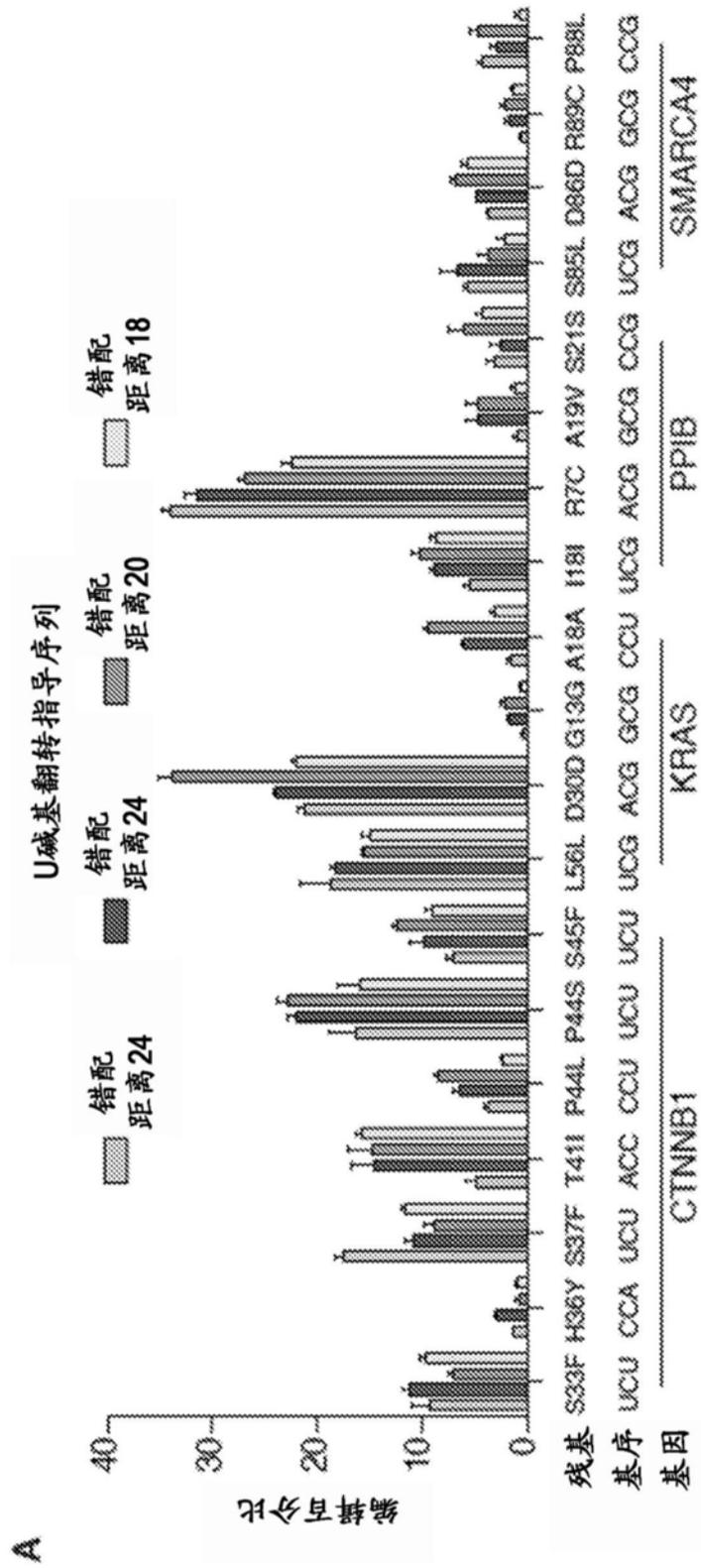


图87A

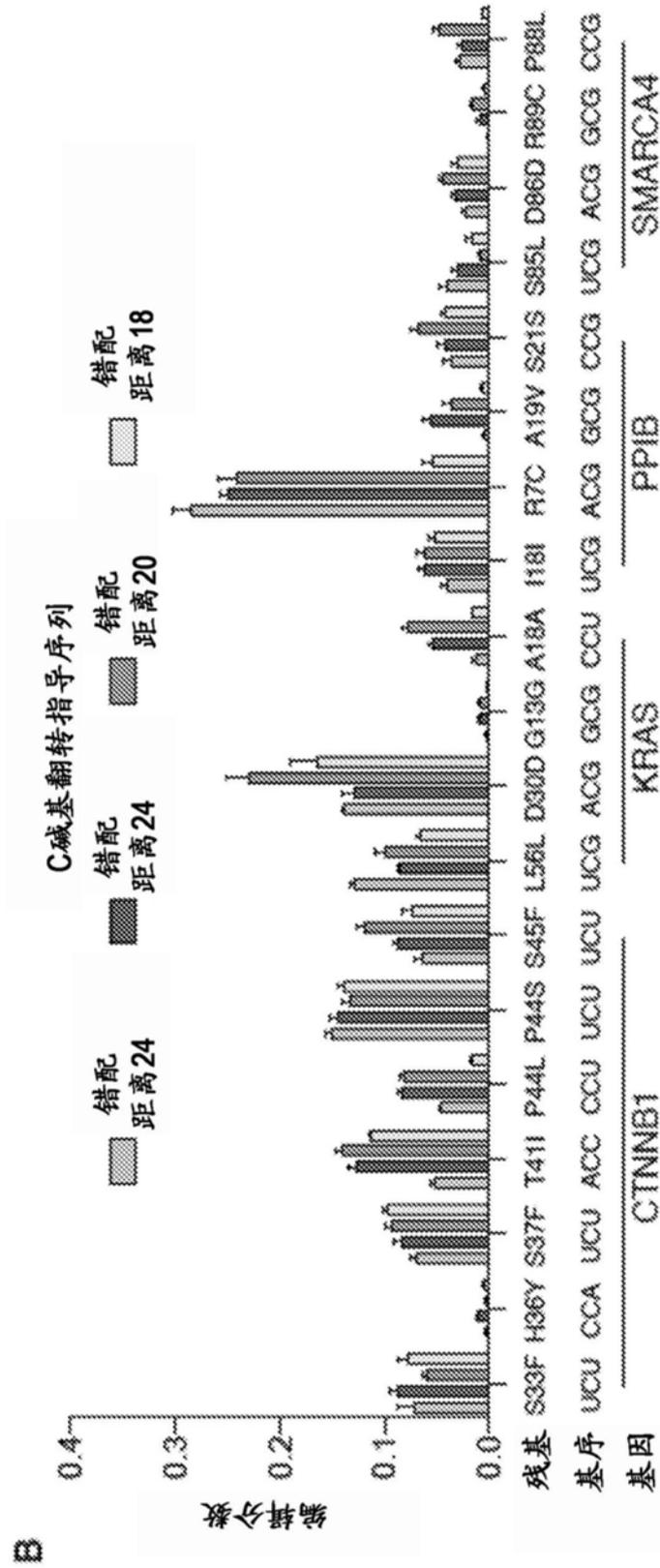


图87B

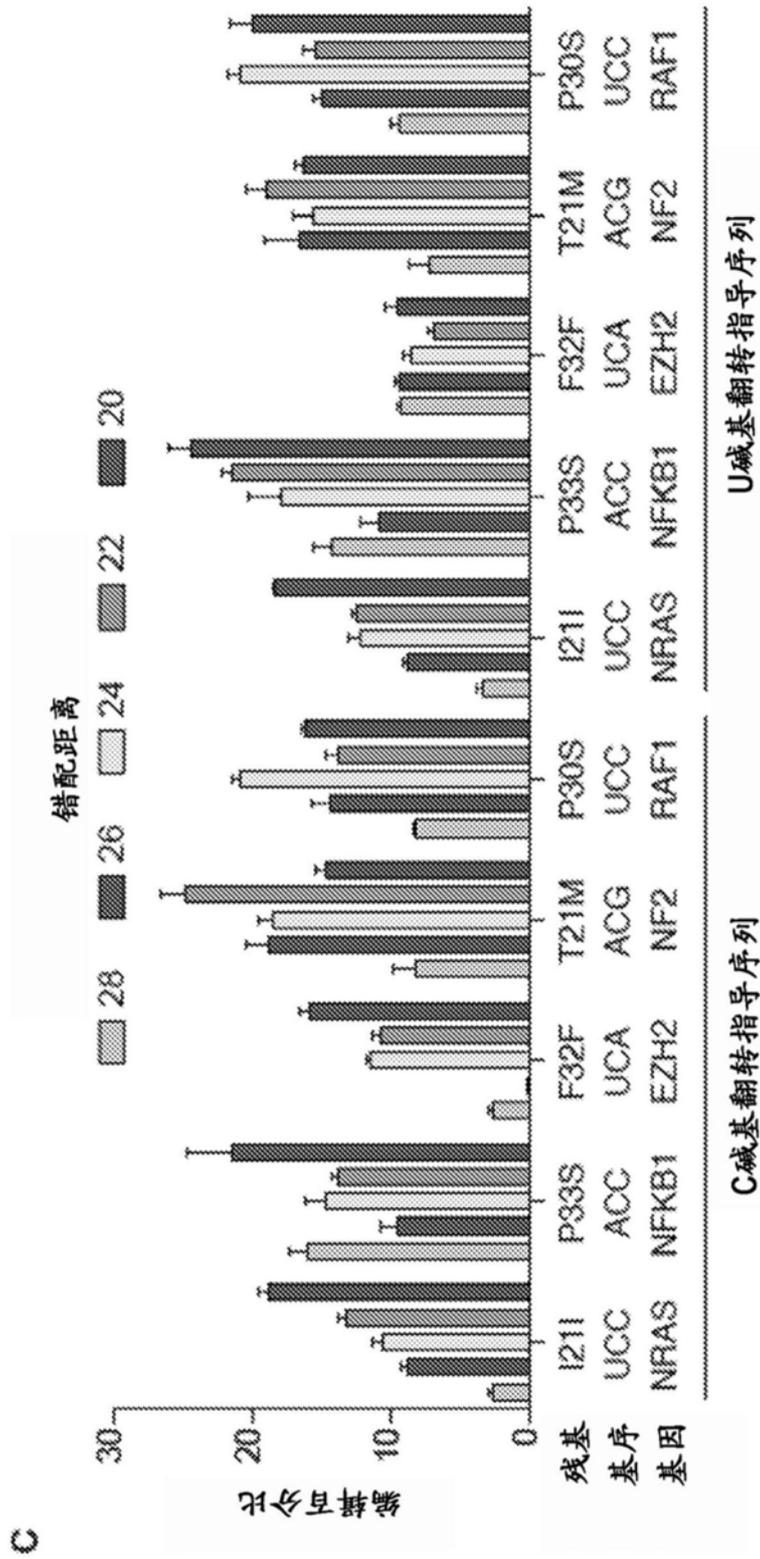


图87C

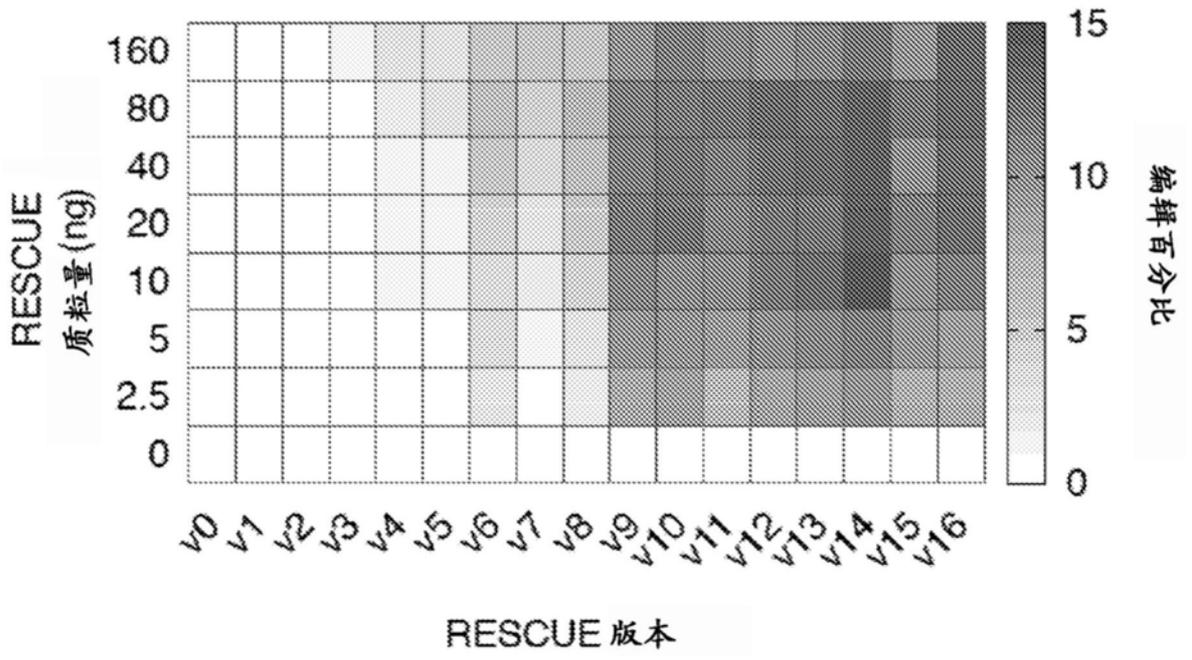


图88

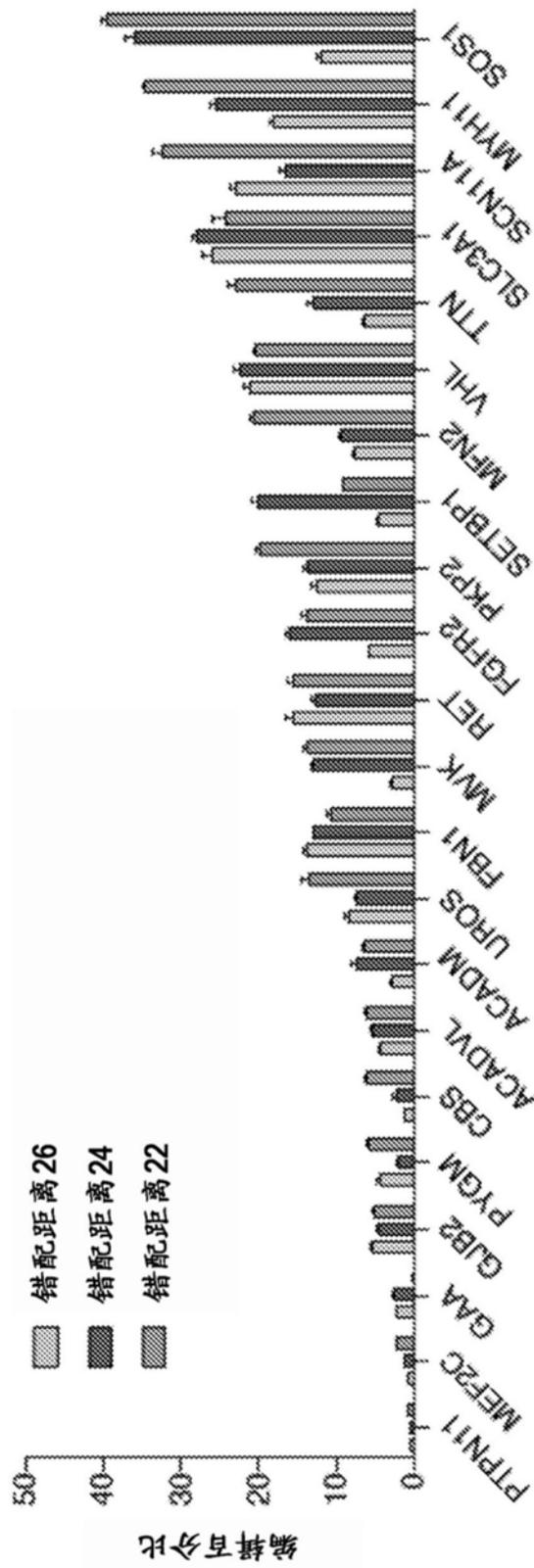


图89



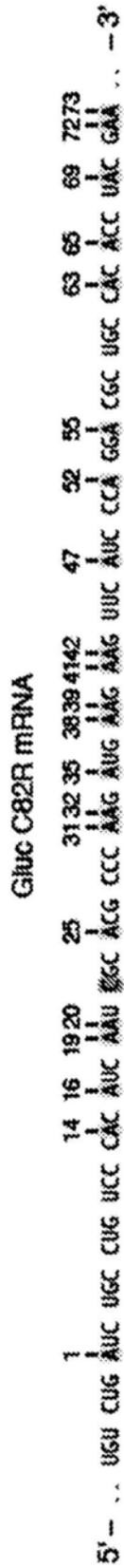


图91A

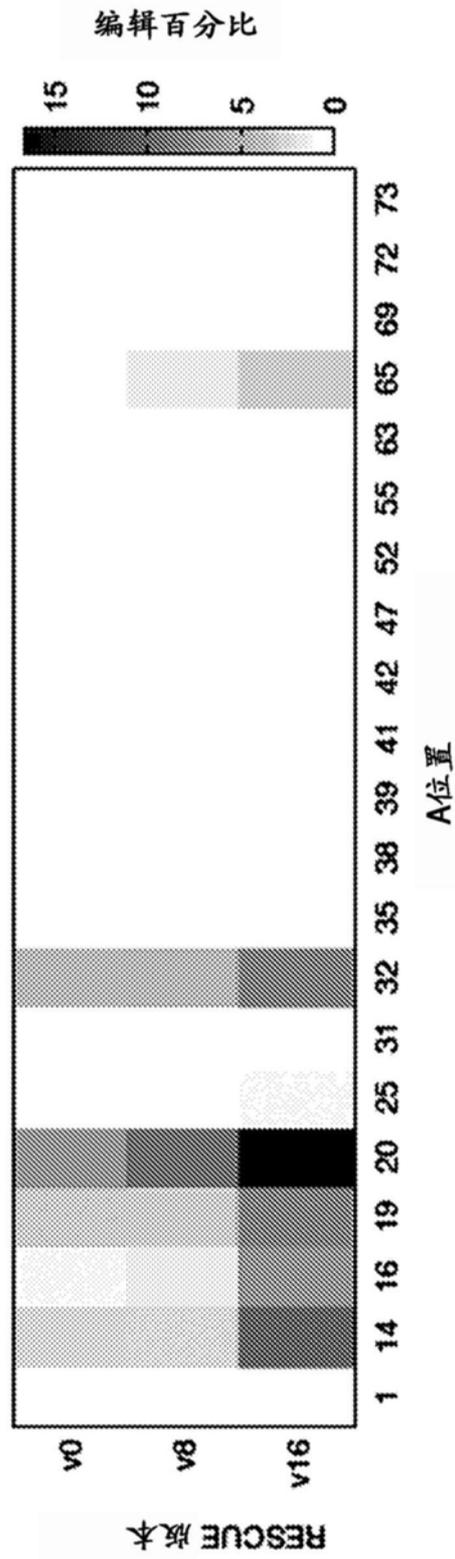


图91B

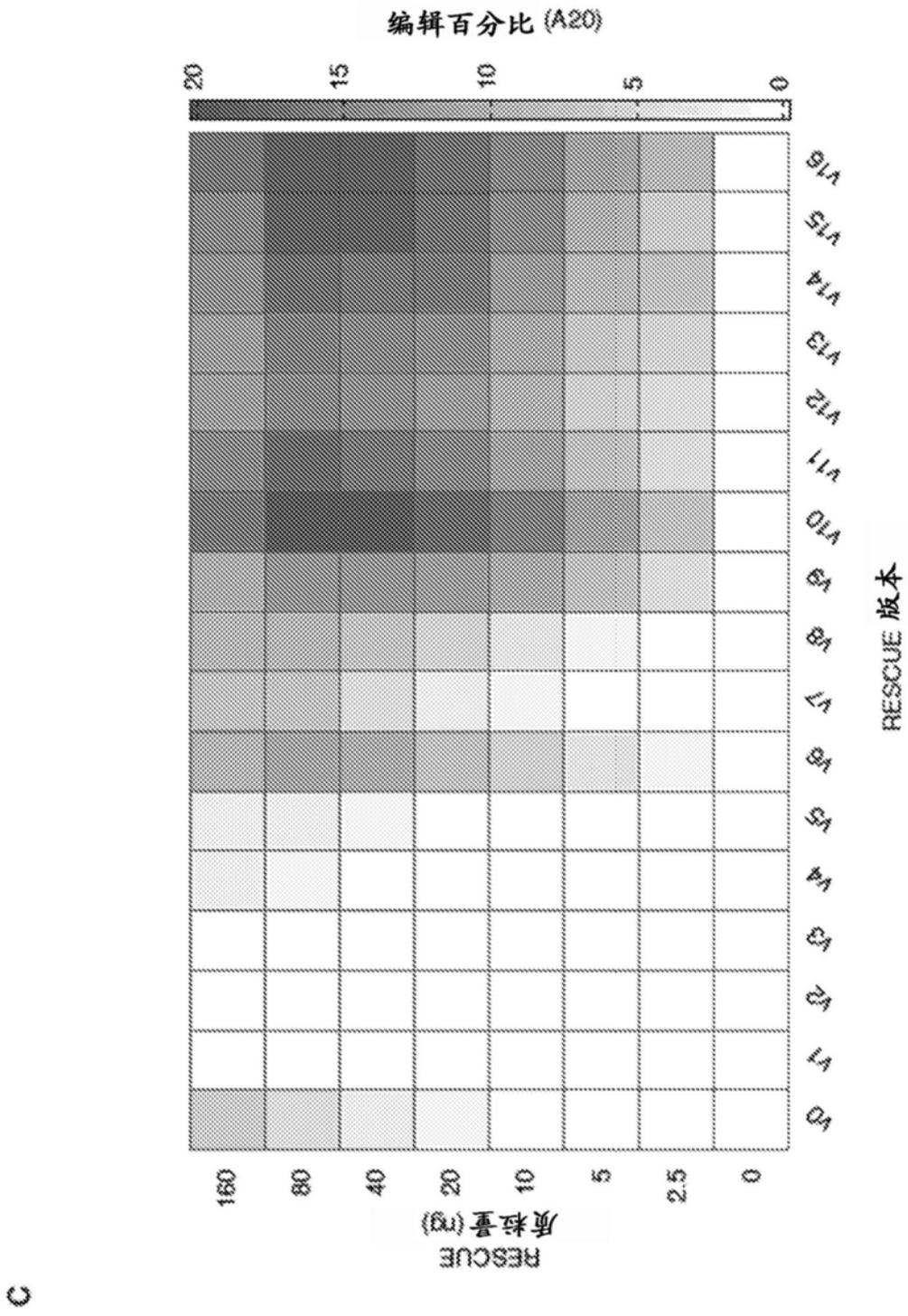


图91C

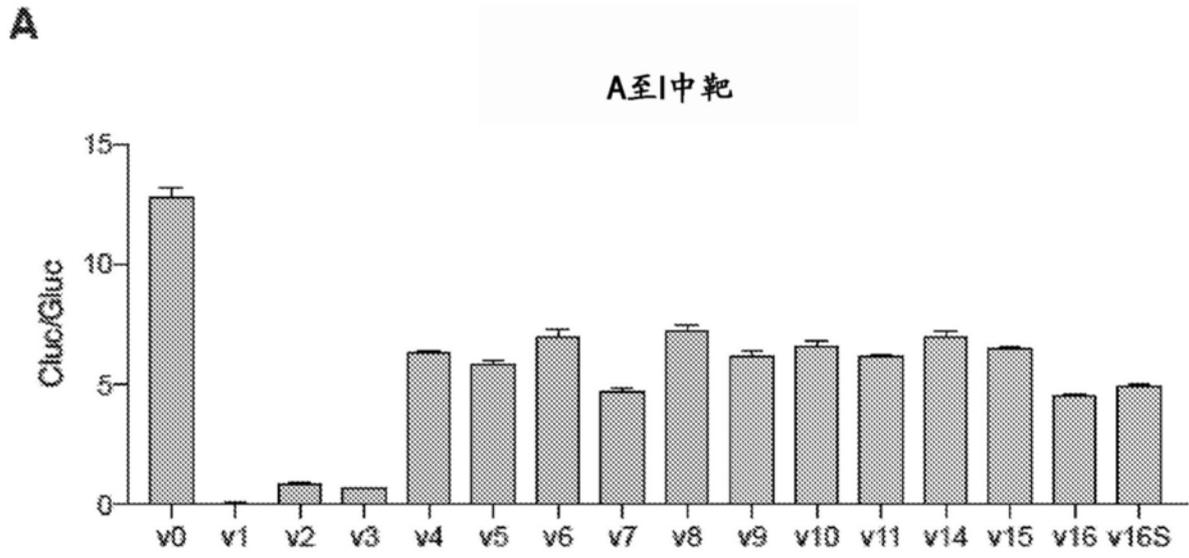


图92A

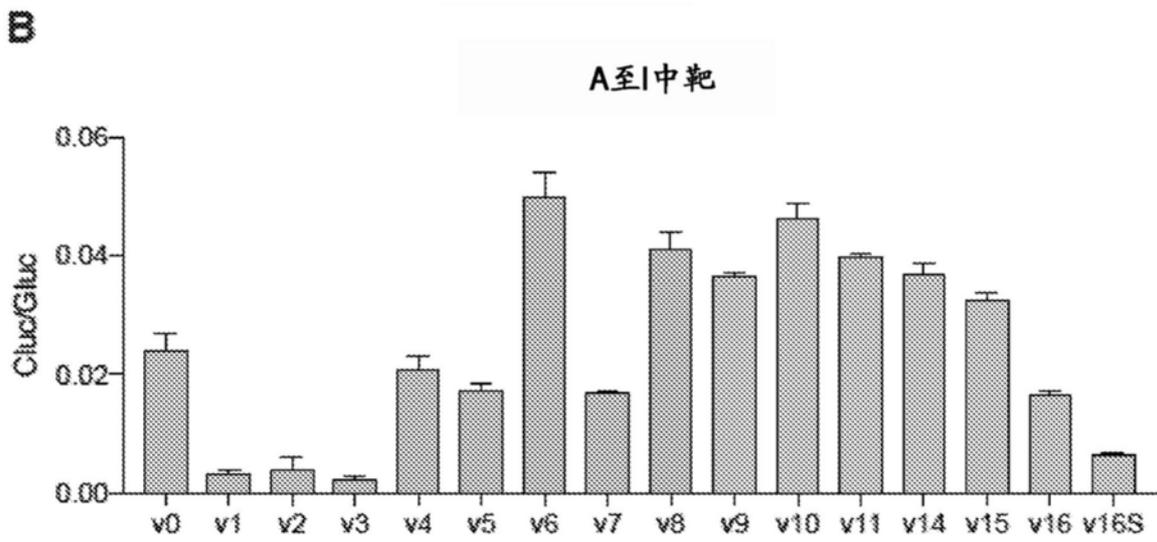


图92B

C

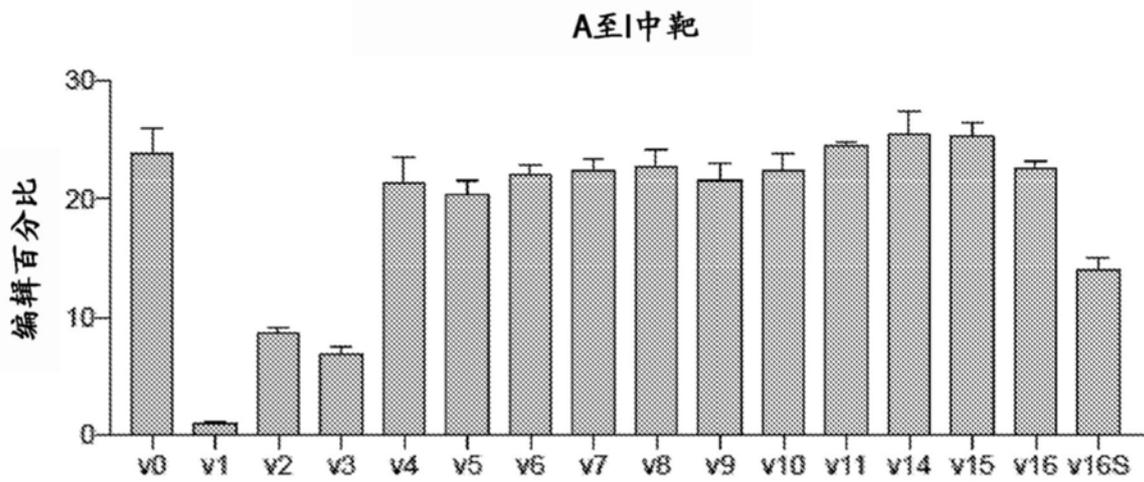


图92C

D

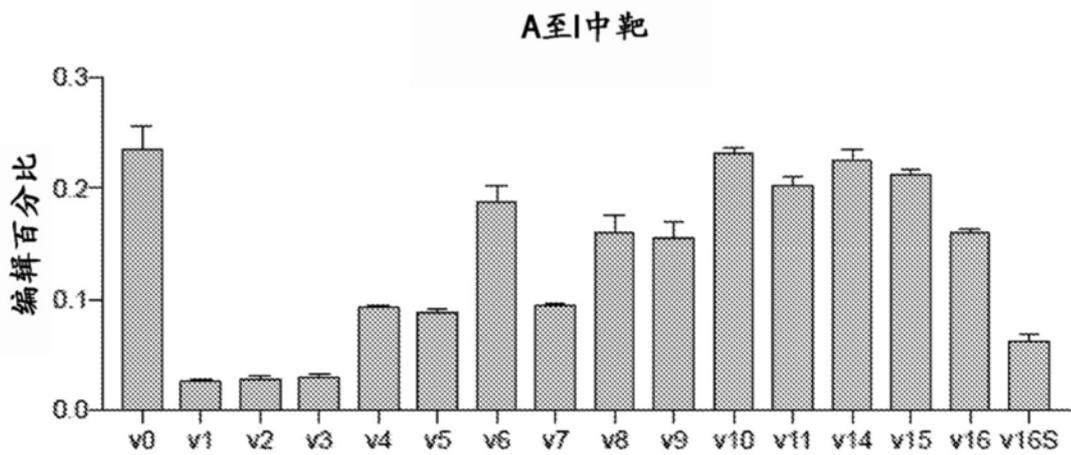


图92D

CTNNB1 T41I



图93A

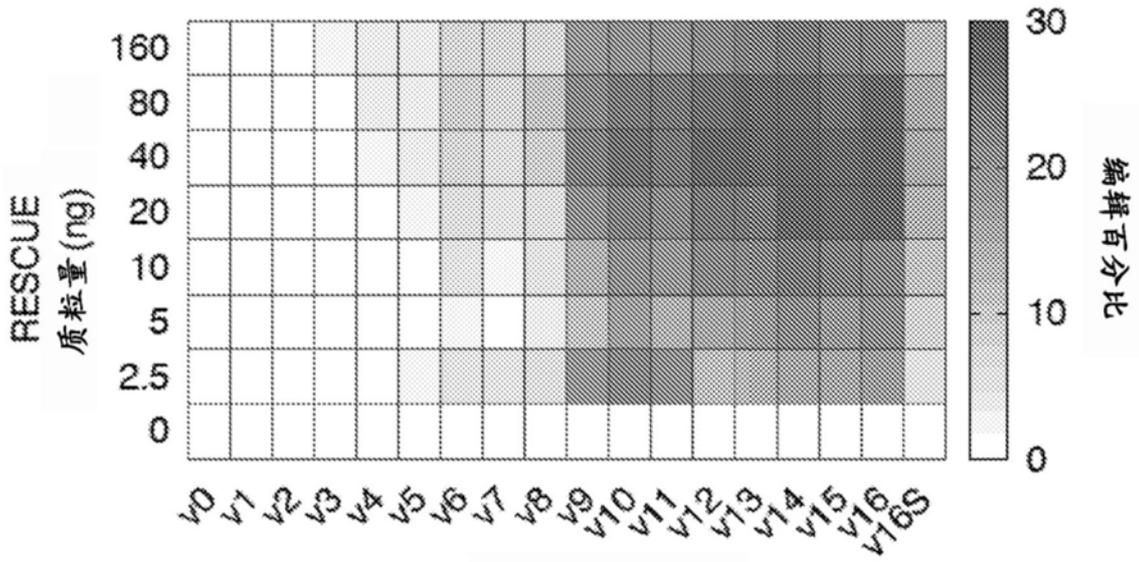


图93B

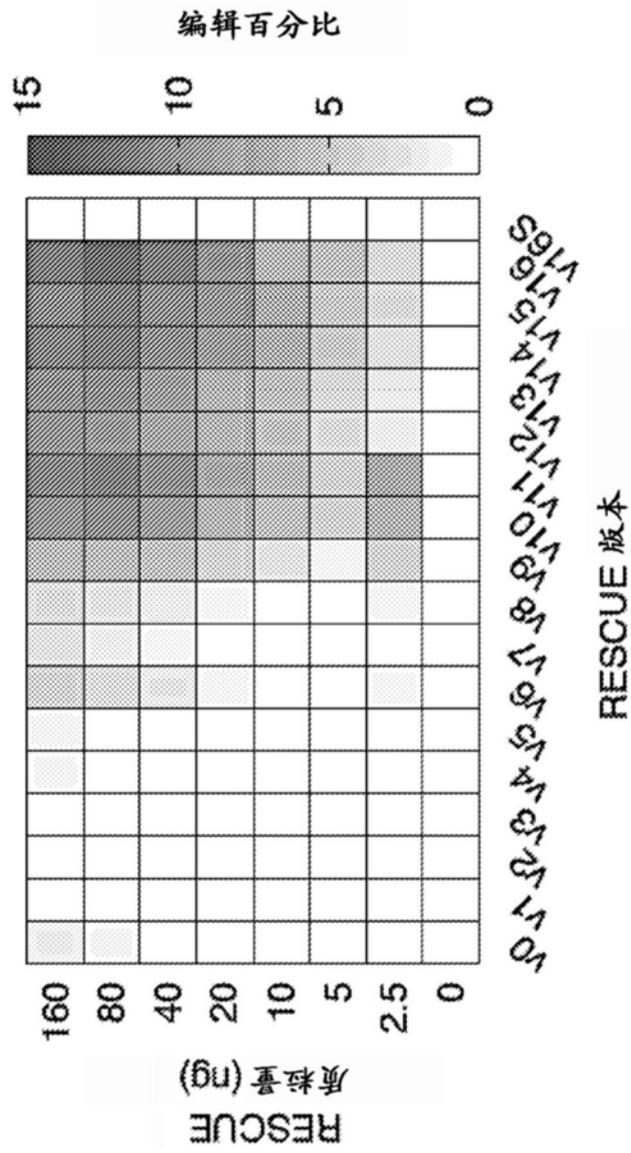


图93C



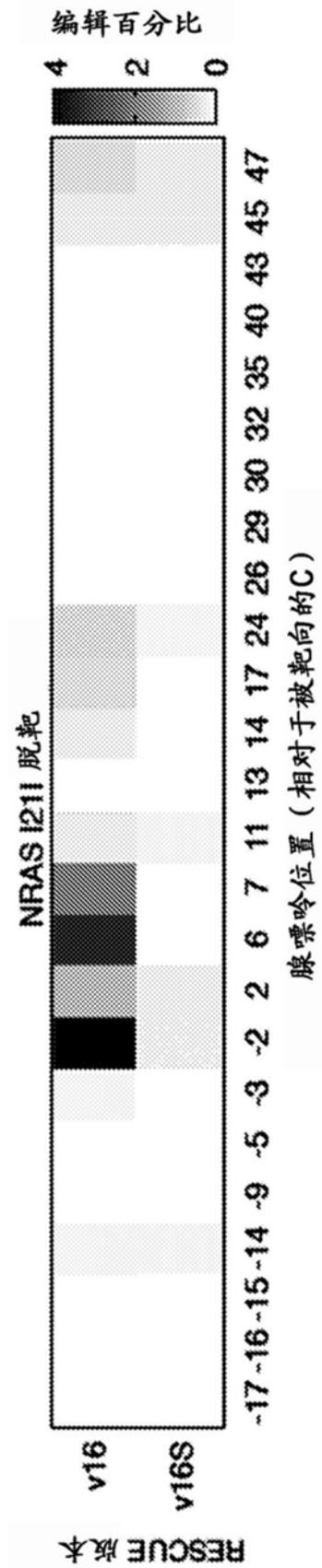


图94B

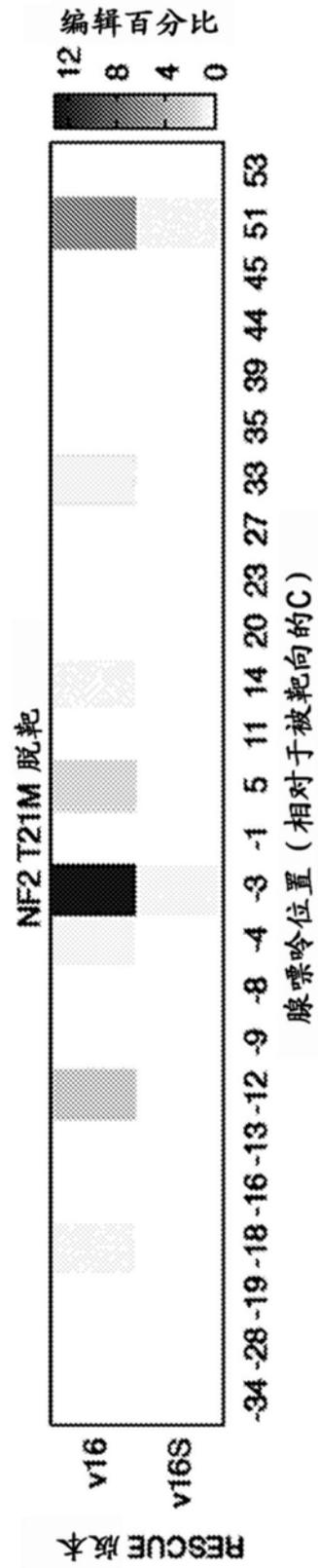


图94C

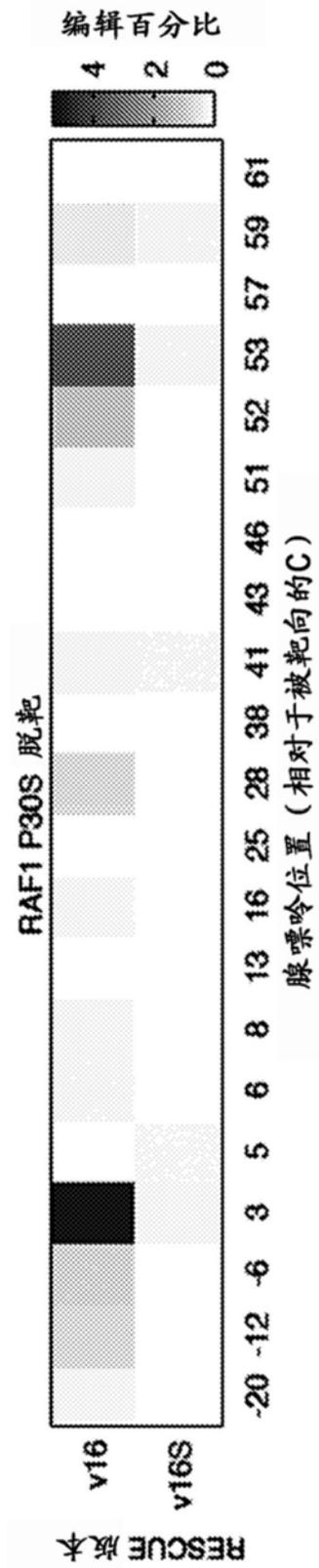


图94D



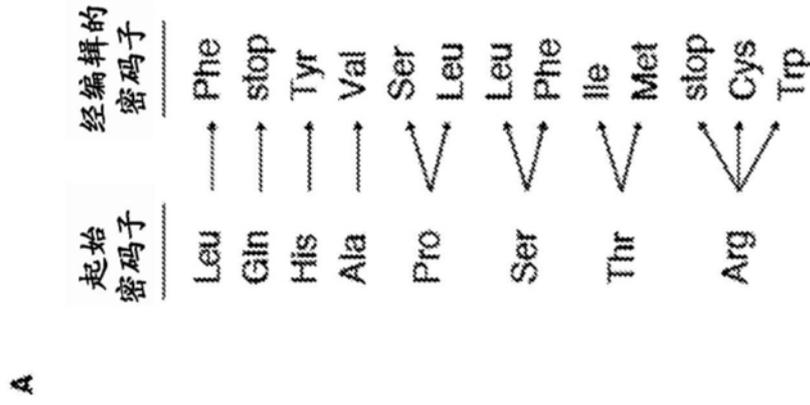


图95A

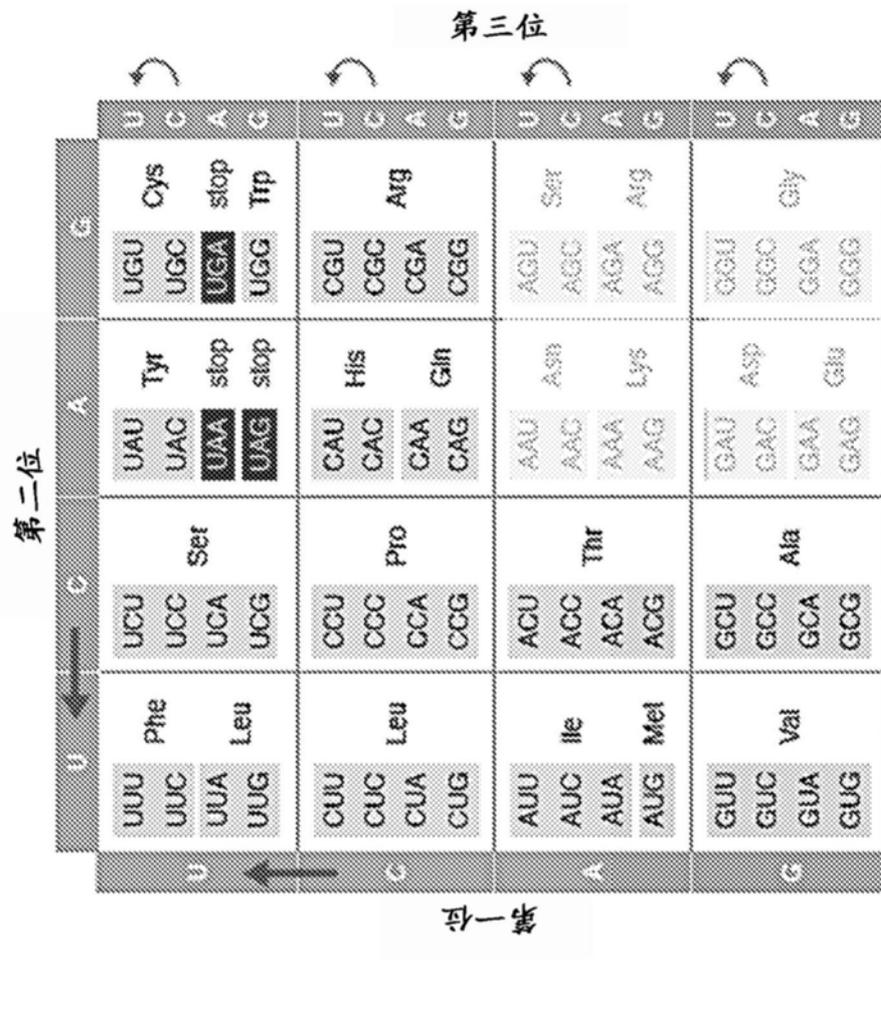


图95B



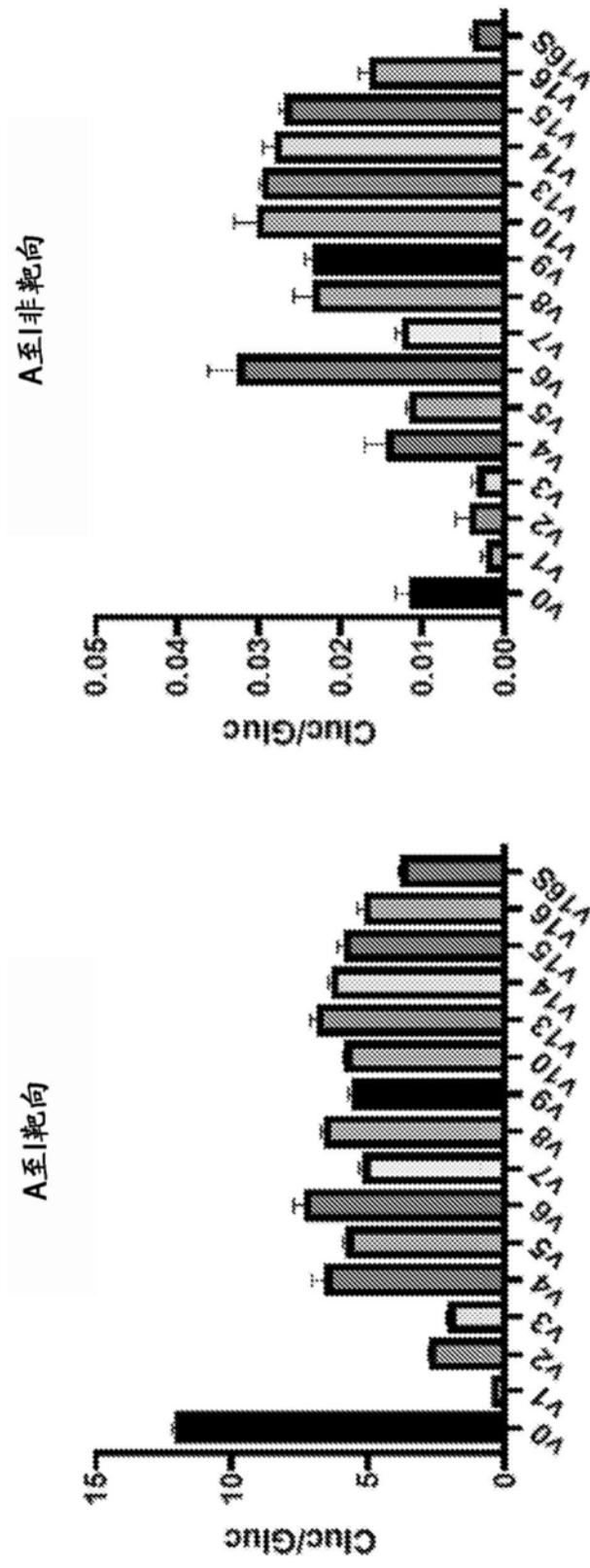


图97

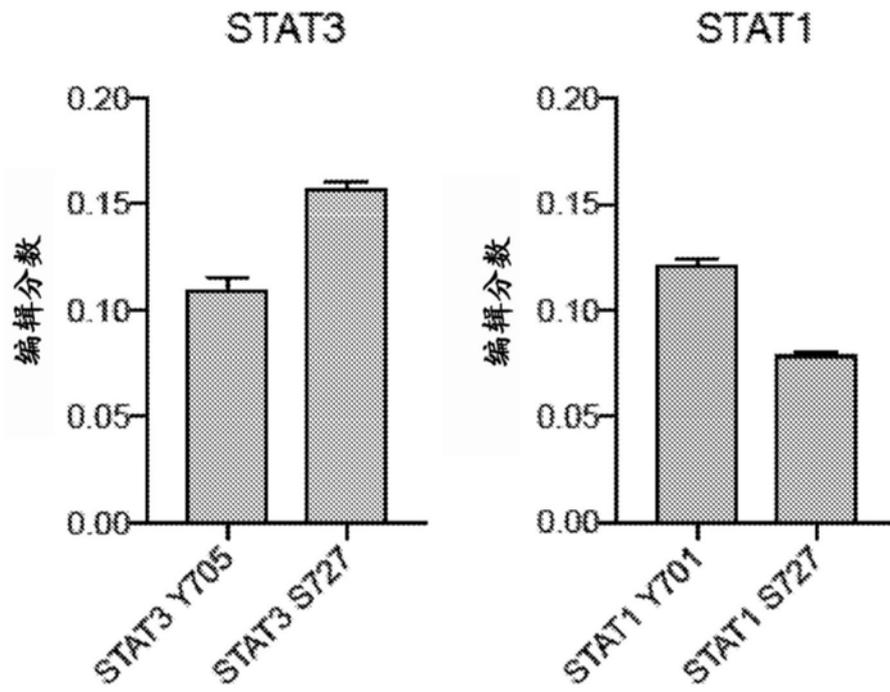


图98

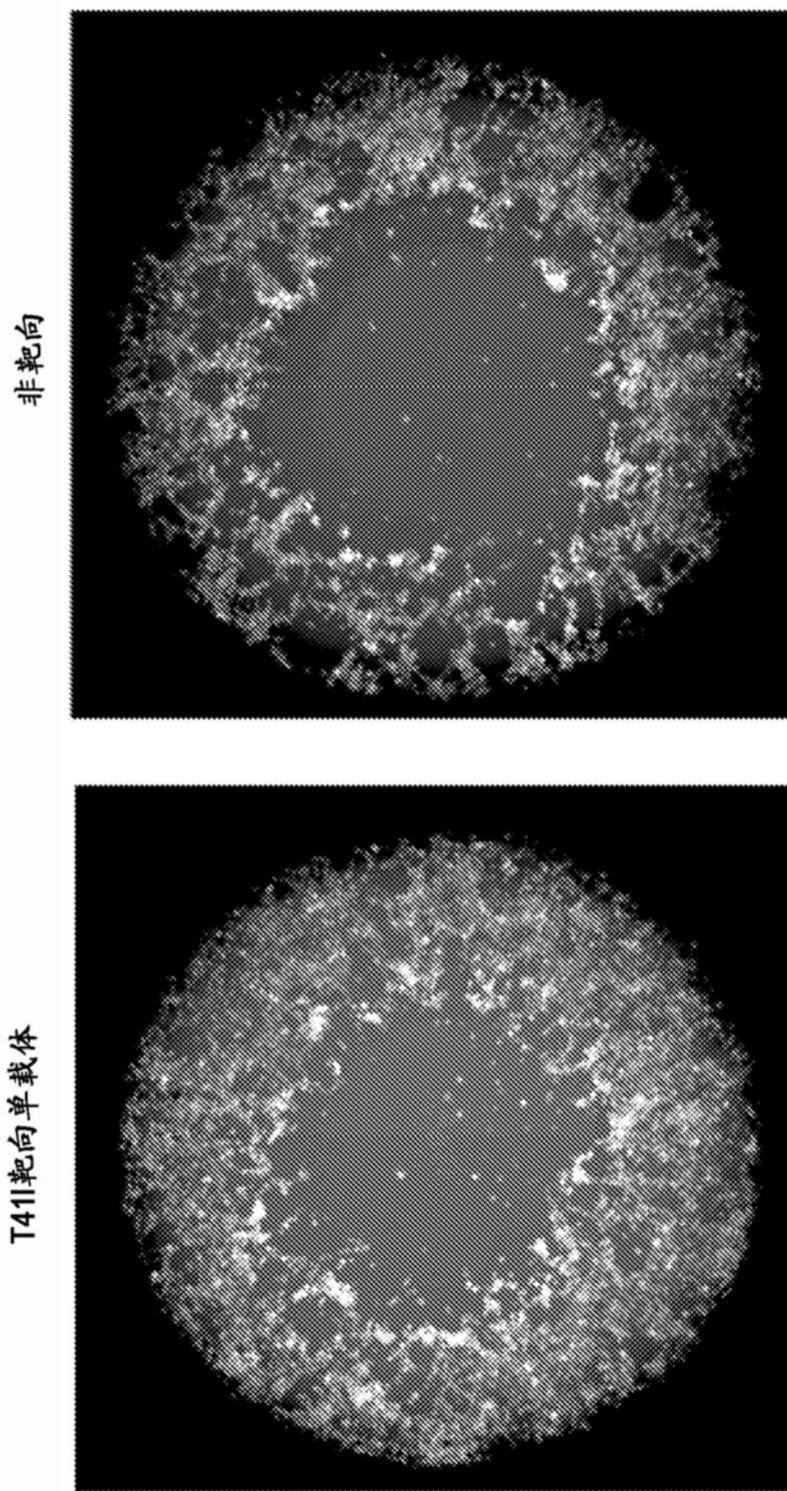


图99A

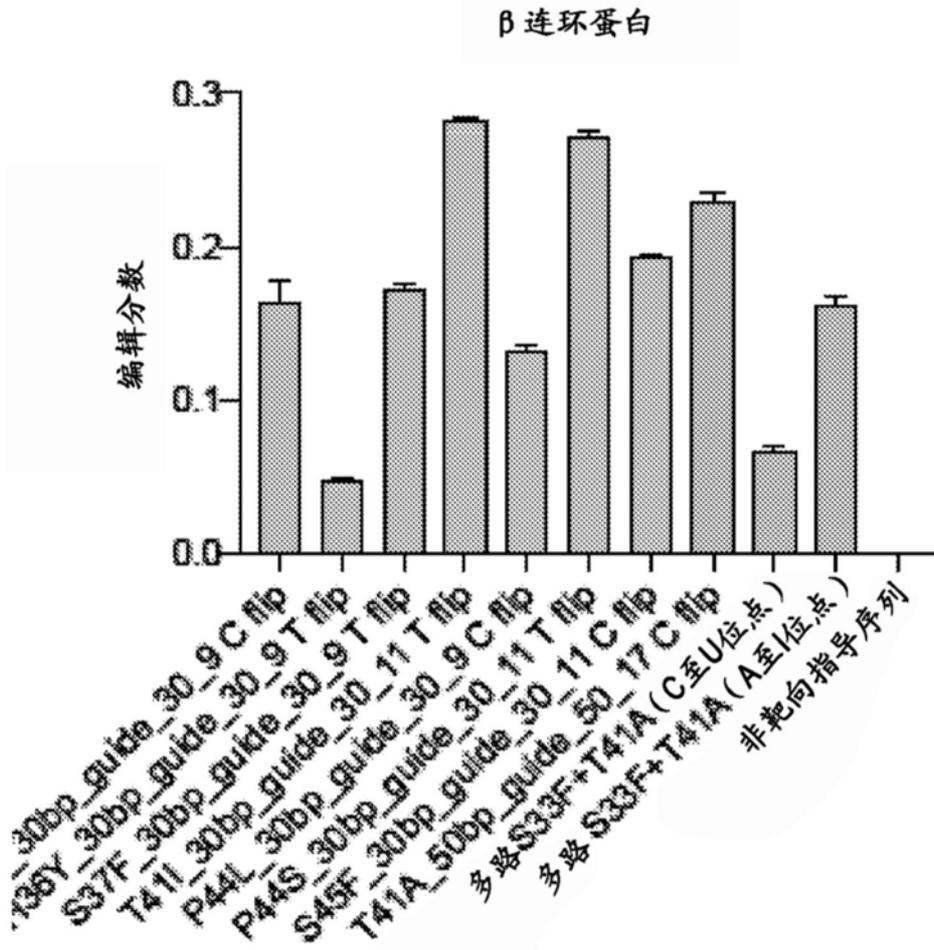


图99B

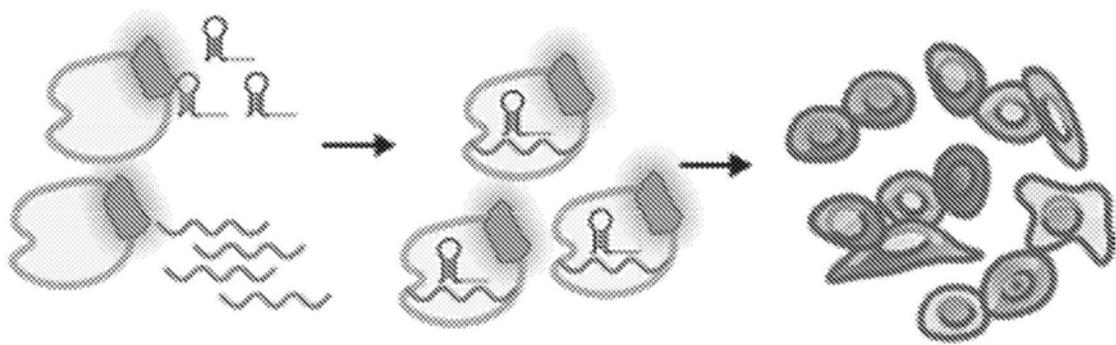


图100

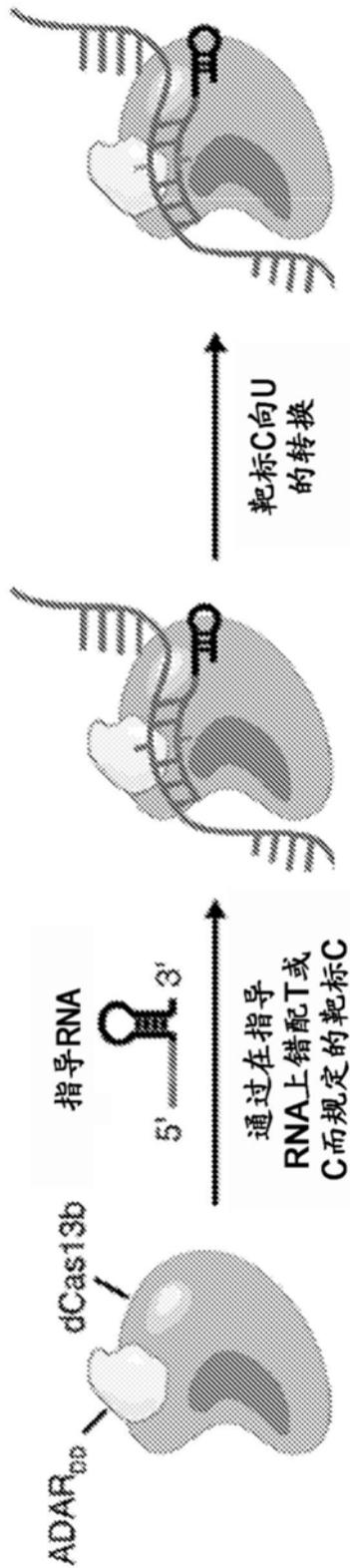


图101

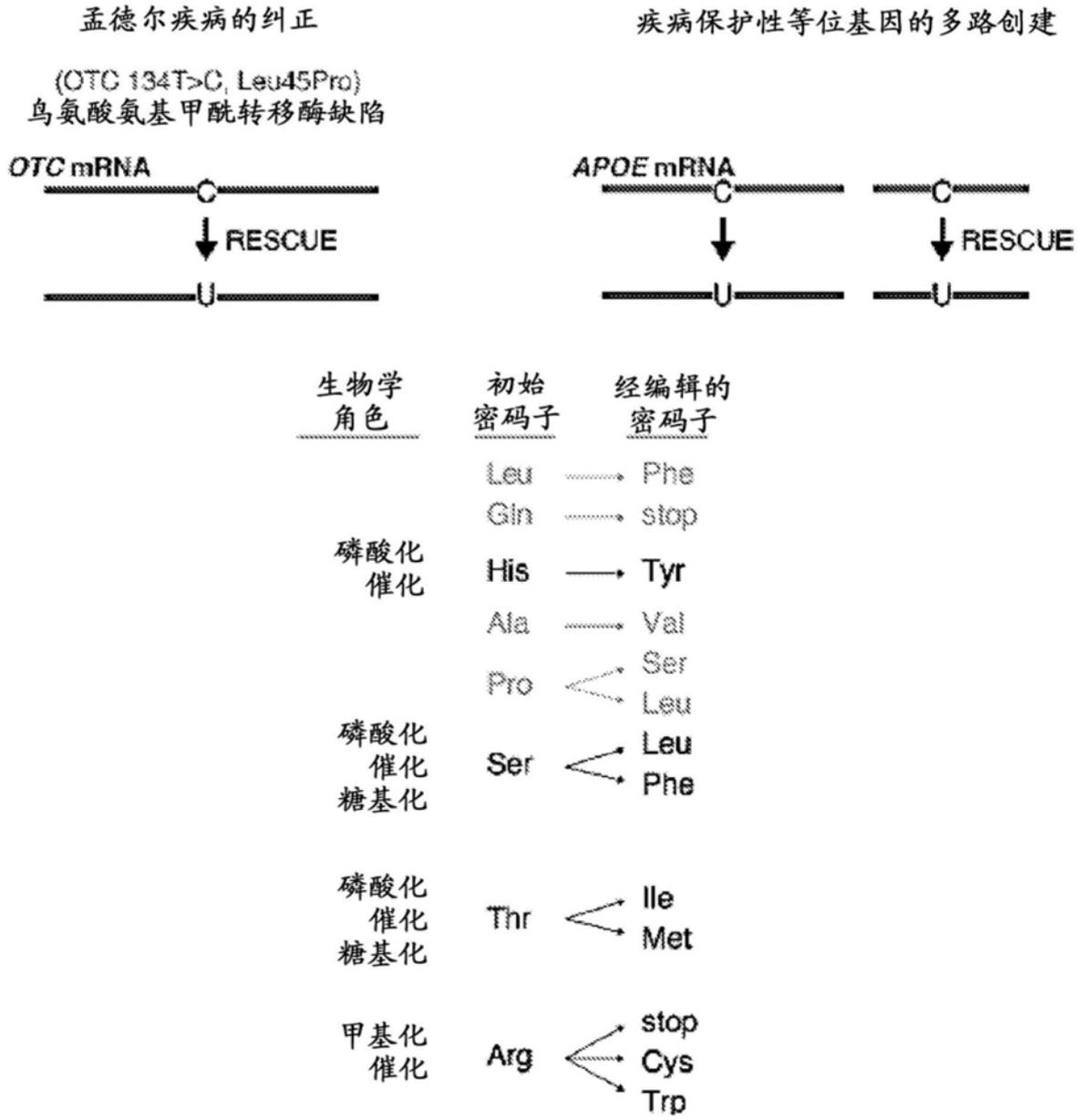


图102

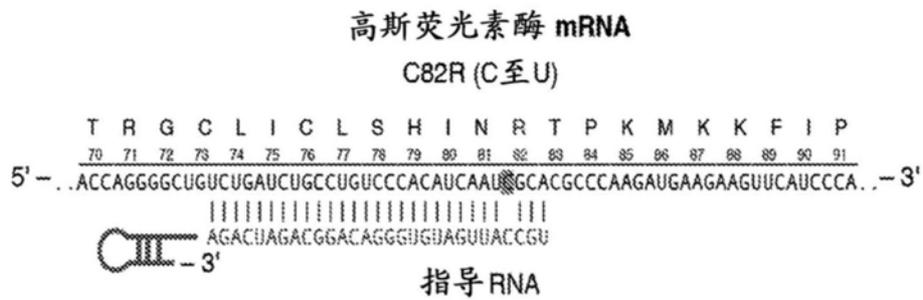


图103A

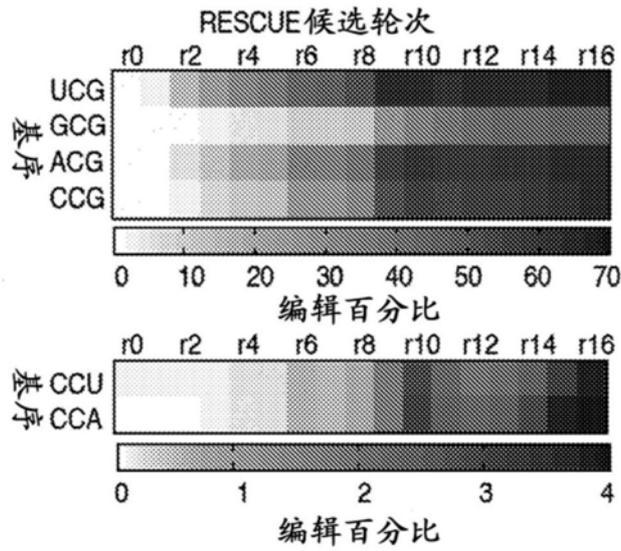


图103B

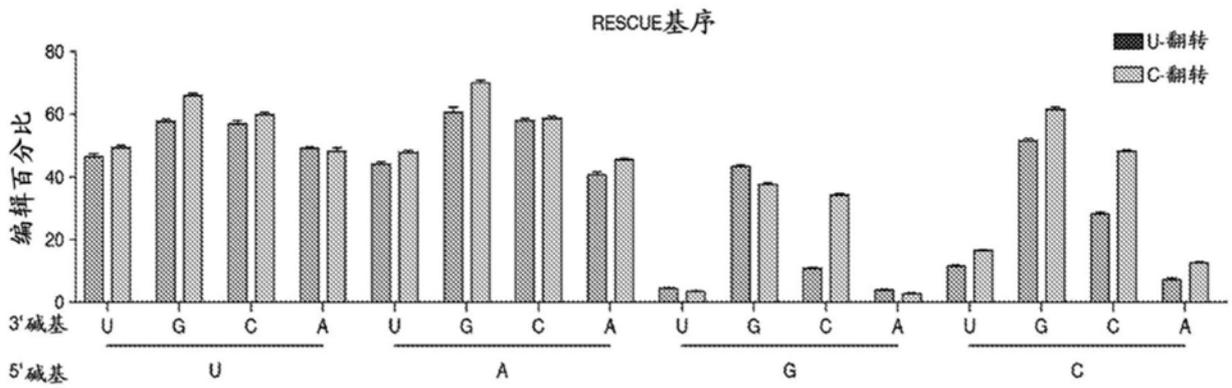


图103C

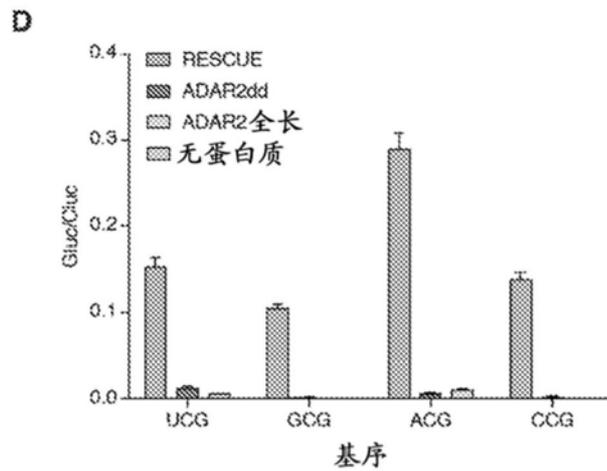


图103D

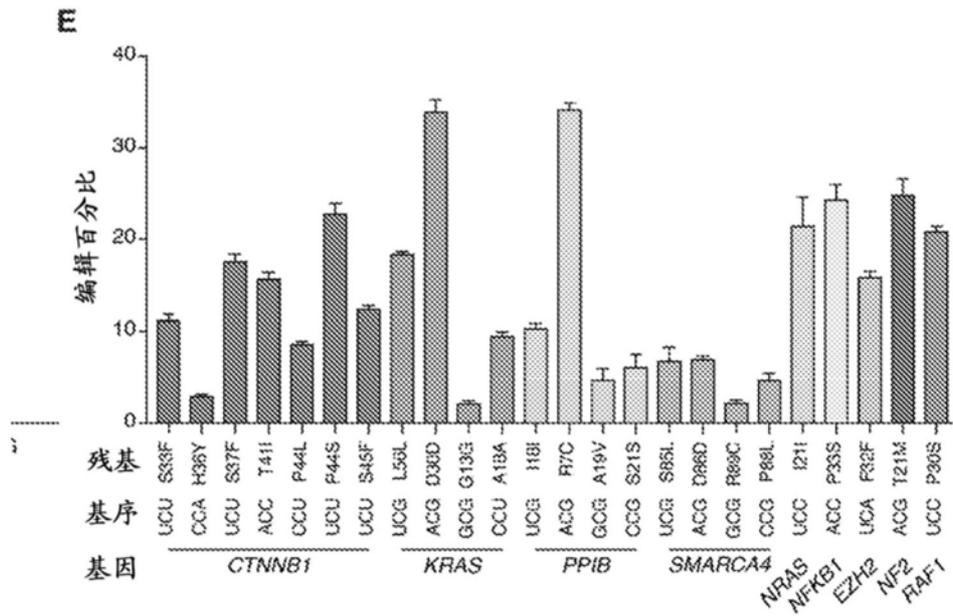


图103E

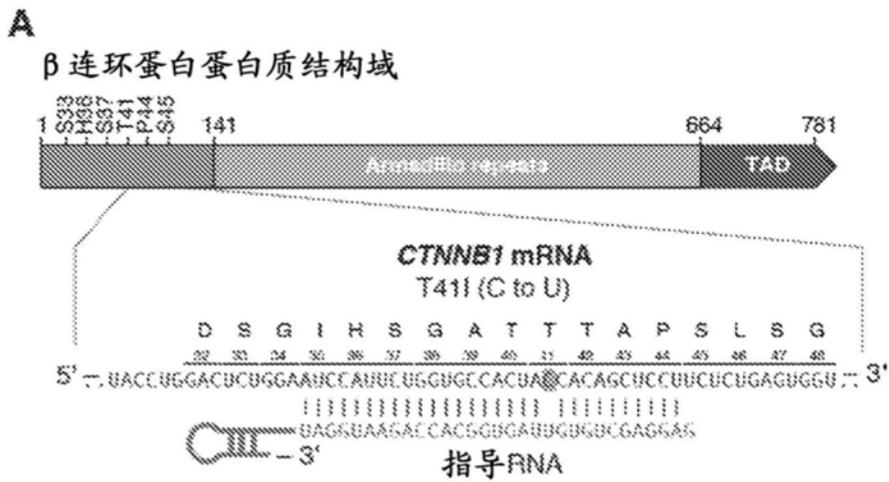


图104A

**B**

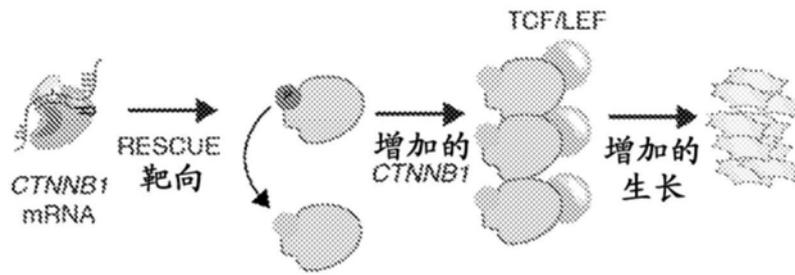


图104B

**C**

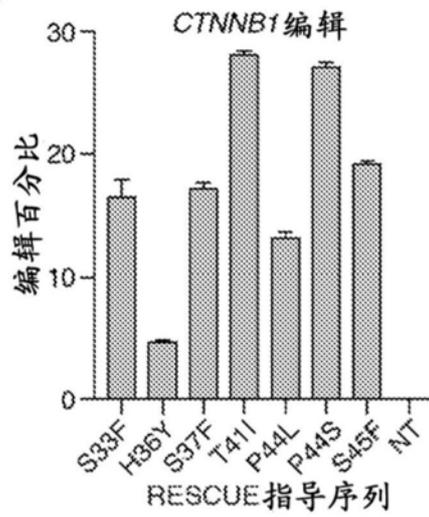


图104C

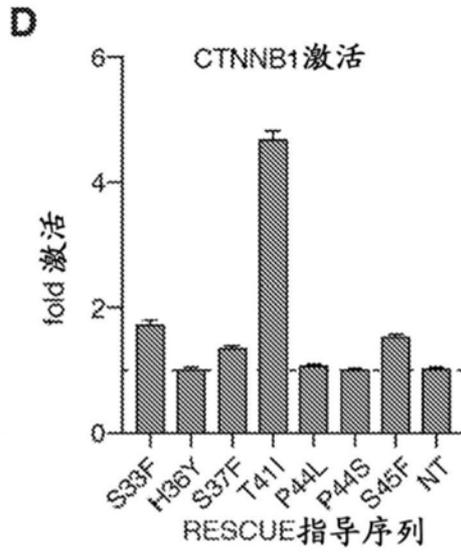


图104D

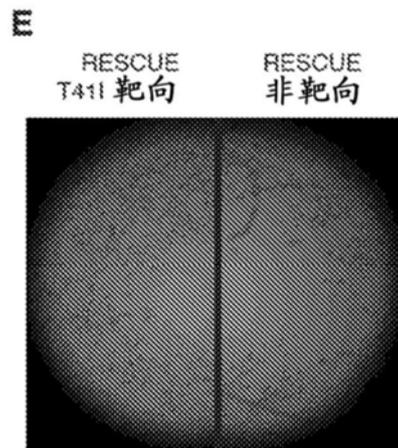


图104E

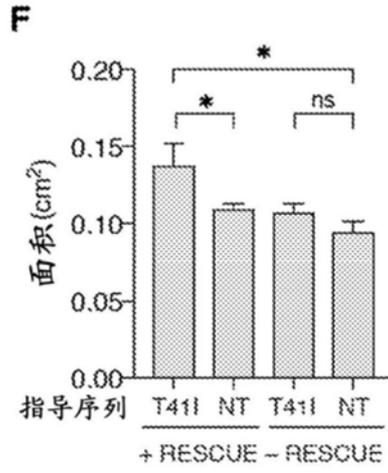


图104F

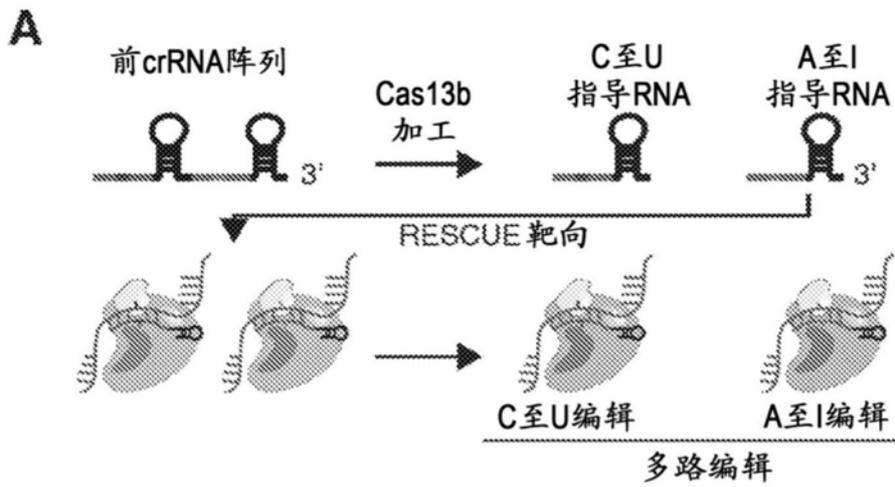


图105A

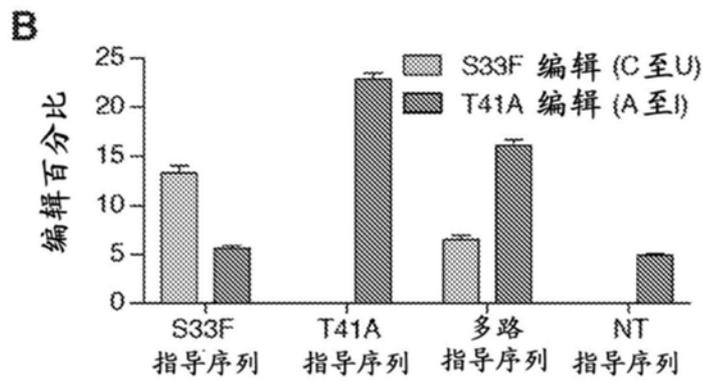


图105B

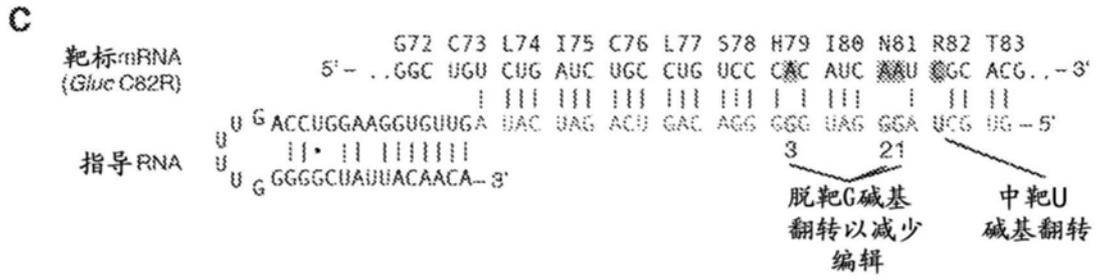


图105C

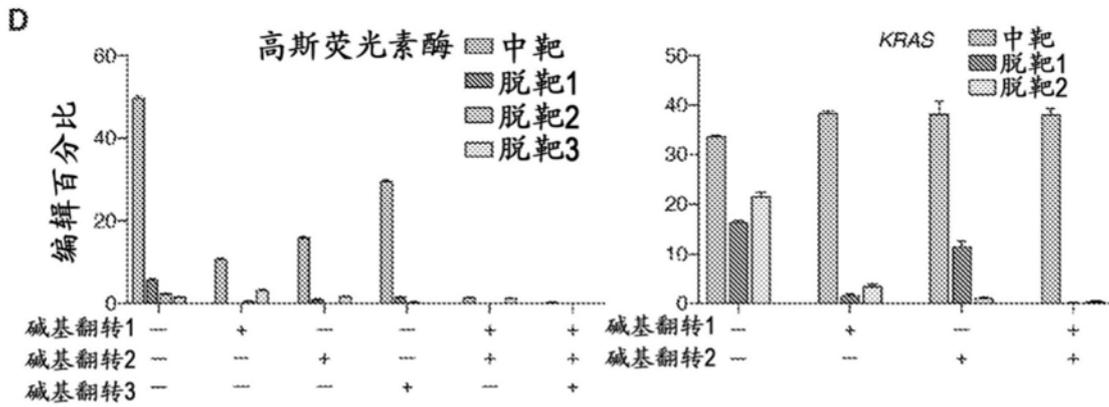


图105D

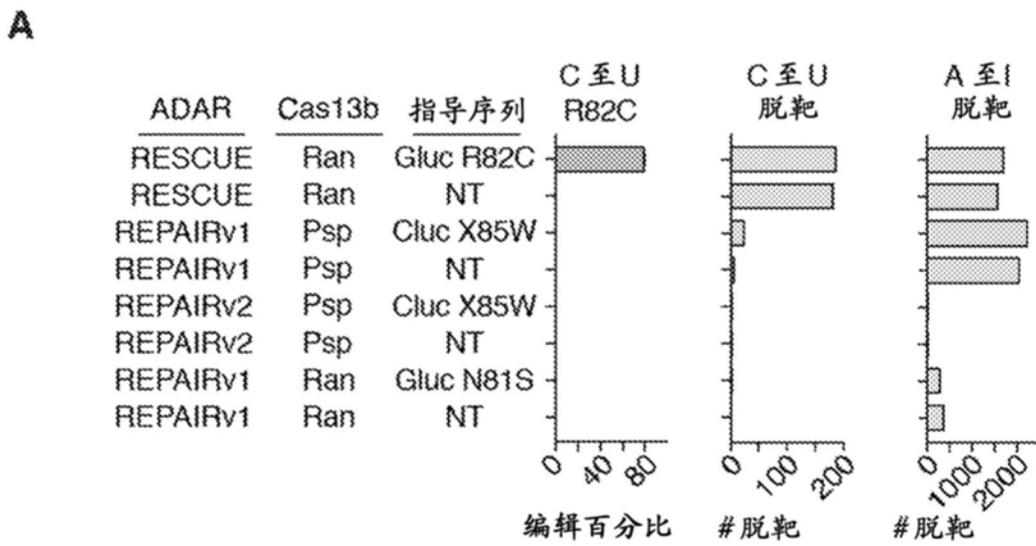


图106A

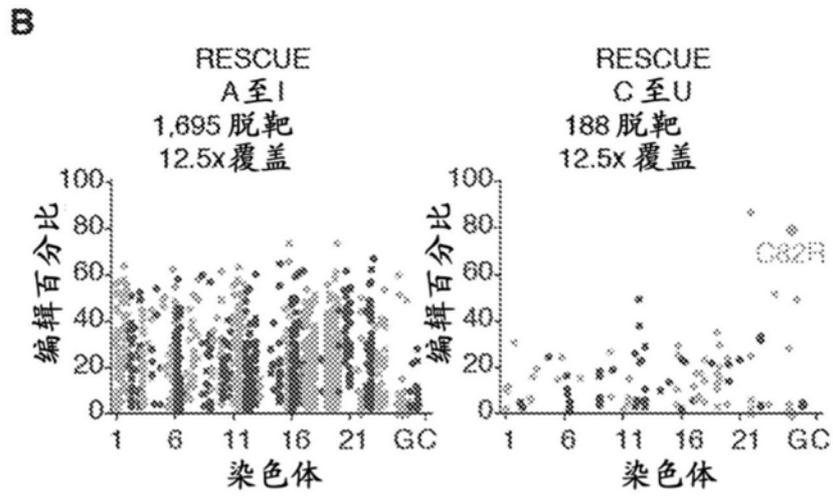


图106B

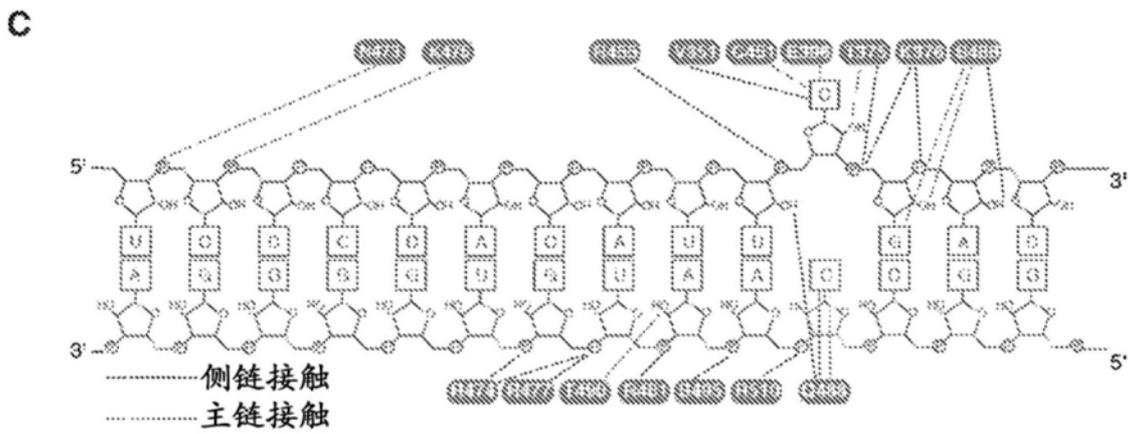


图106C

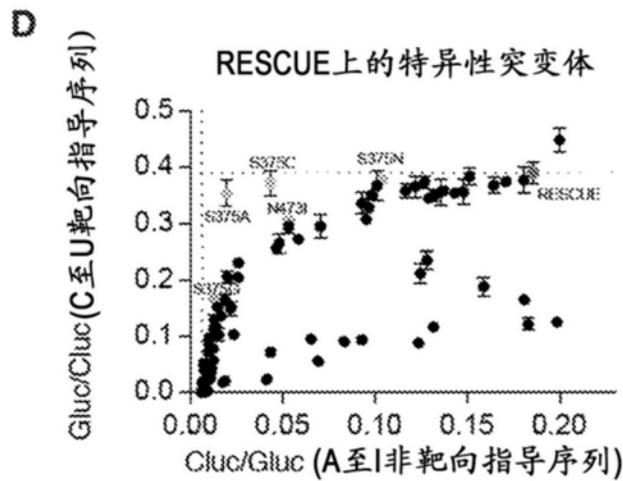


图106D

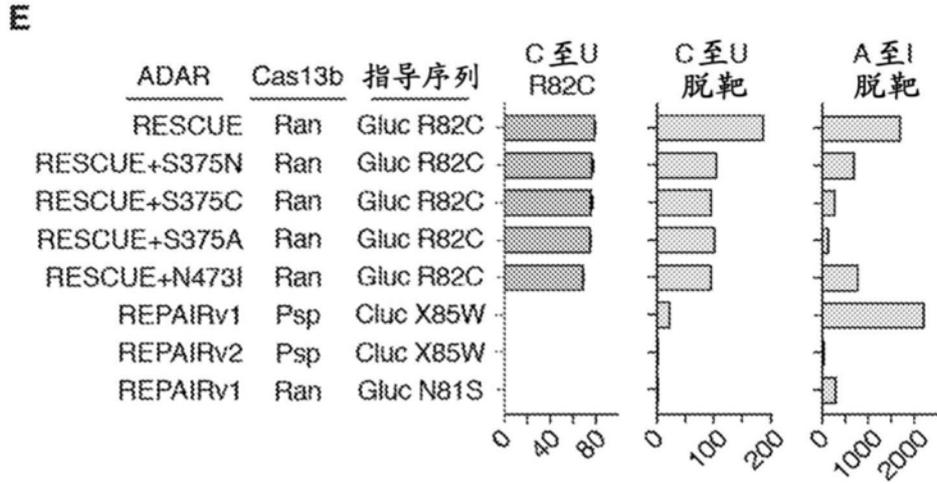


图106E

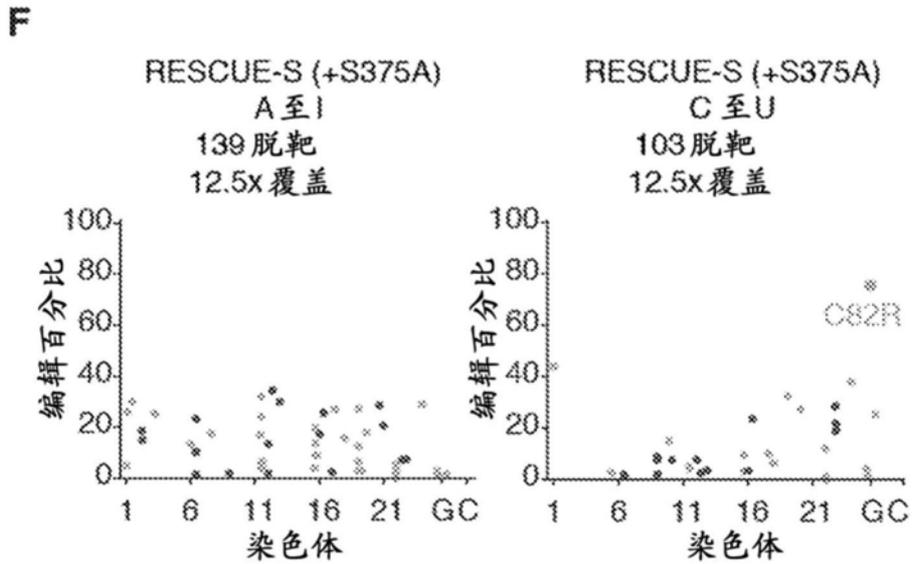


图106F



## 采用RESCUE的可能氨基酸转换

| 初始<br>密码子 | 最终<br>密码子                | 潜在<br>功能         |
|-----------|--------------------------|------------------|
| Leu       | → Phe                    |                  |
| Gln       | → stop                   |                  |
| His       | → Tyr                    | 磷酸化<br>催化        |
| Ala       | → Val                    |                  |
| Pro       | ↘ Ser<br>↗ Leu           |                  |
| Ser       | ↘ Leu<br>↗ Phe           | 磷酸化<br>催化<br>糖基化 |
| Thr       | ↘ Ile<br>↗ Met           | 磷酸化<br>催化<br>糖基化 |
| Arg       | ↘ stop<br>↗ Cys<br>↘ Trp | 甲基化<br>催化        |

图107A

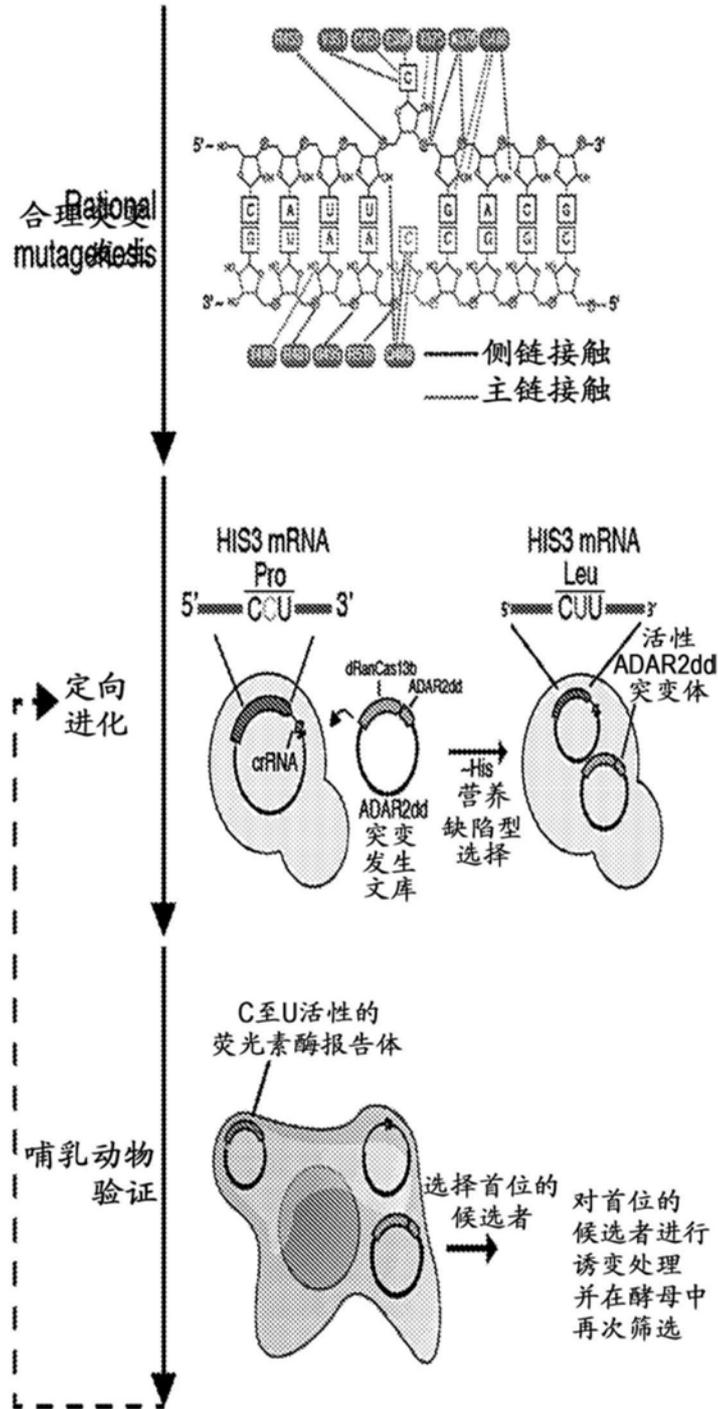


图107B

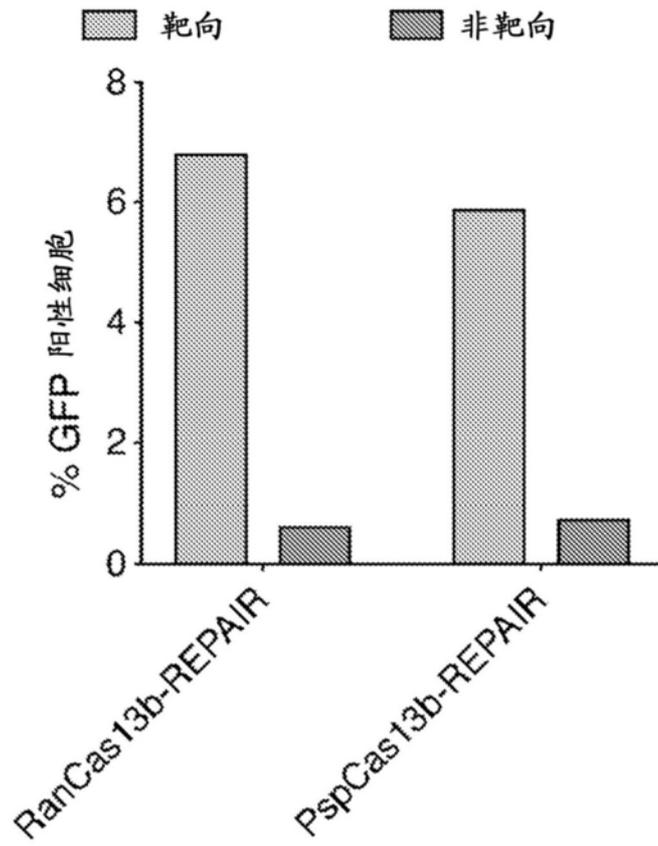


图108

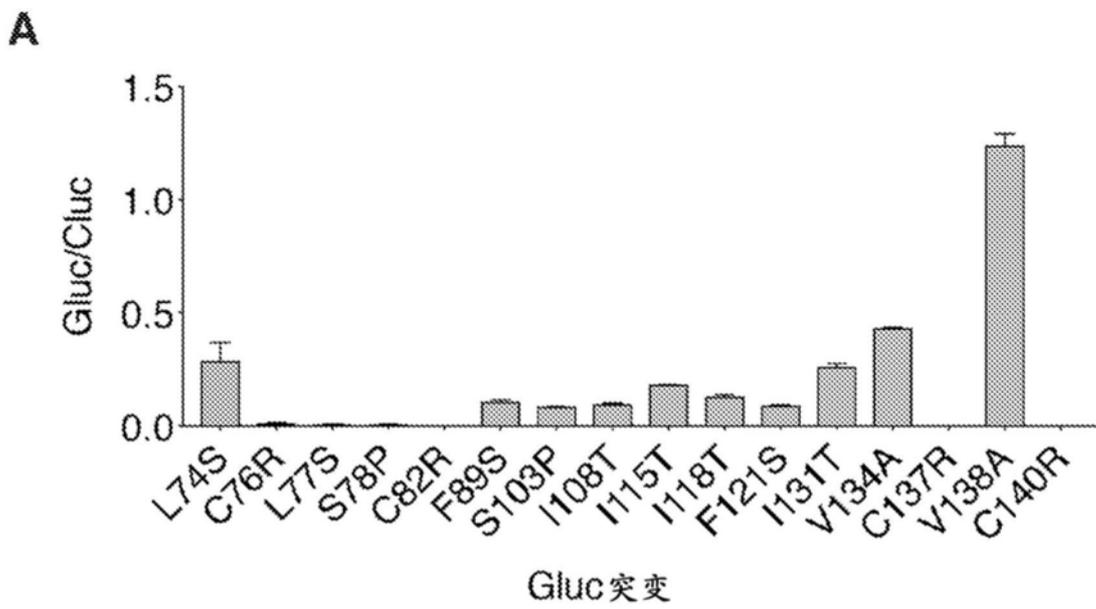


图109A

**B**

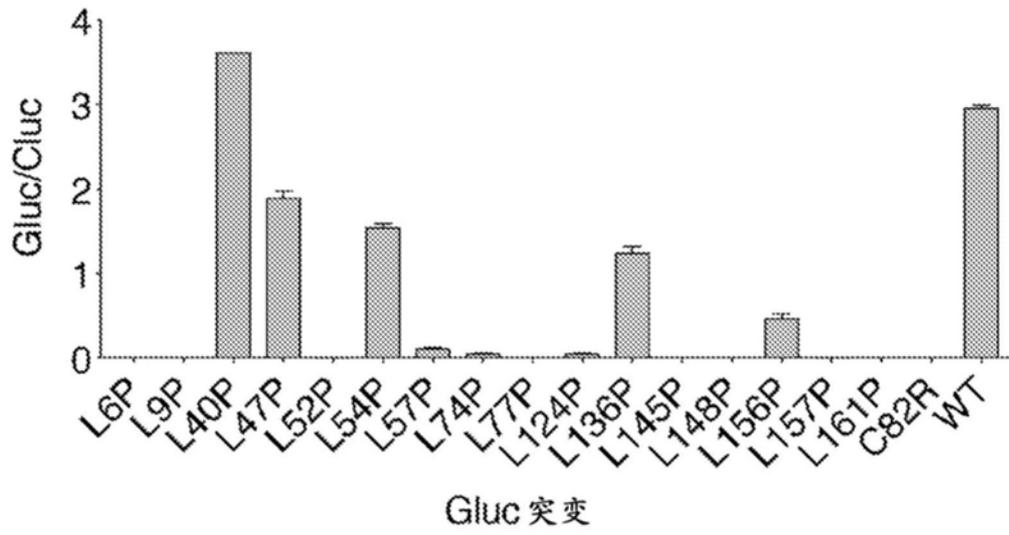


图109B

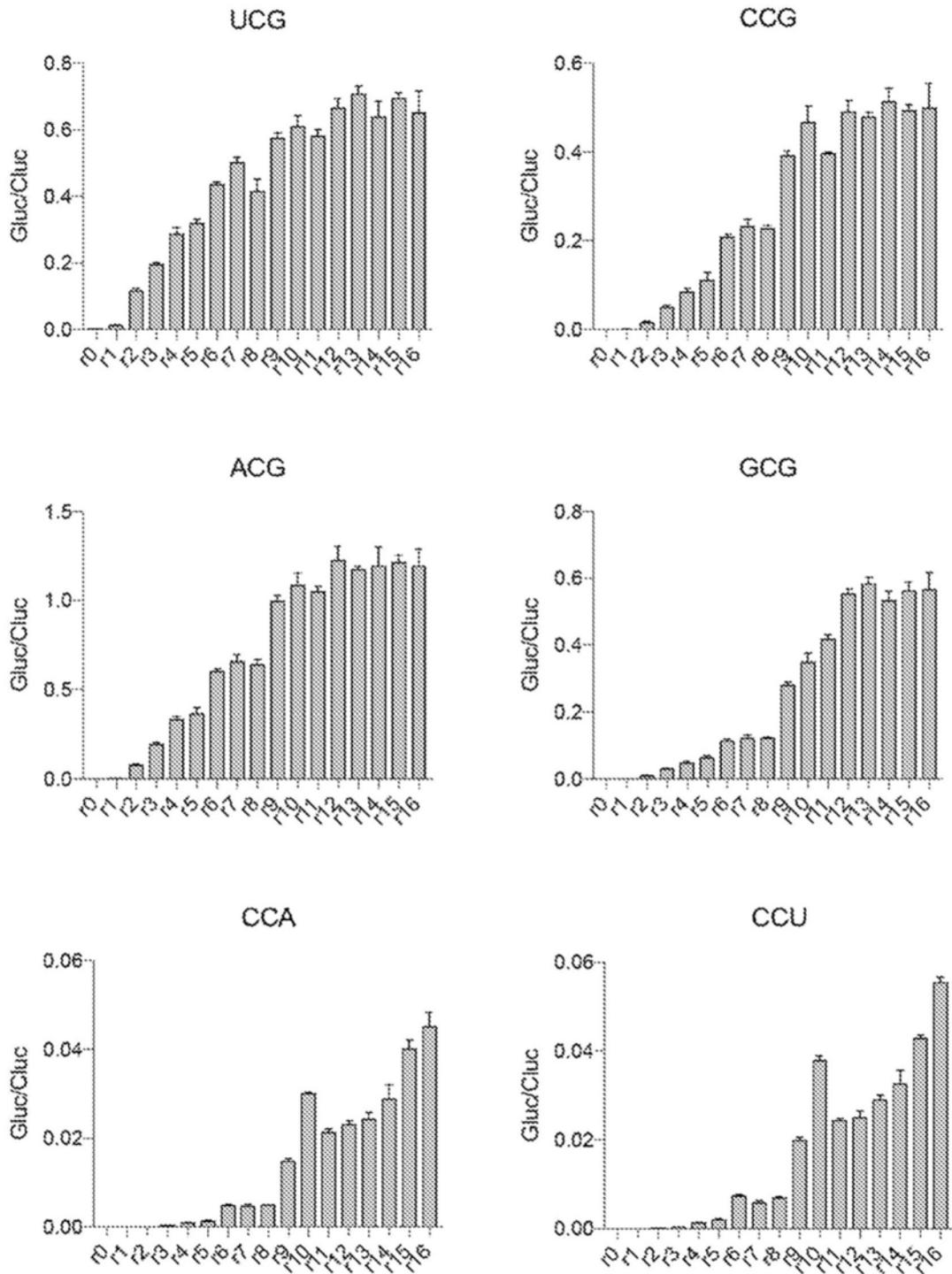


图110

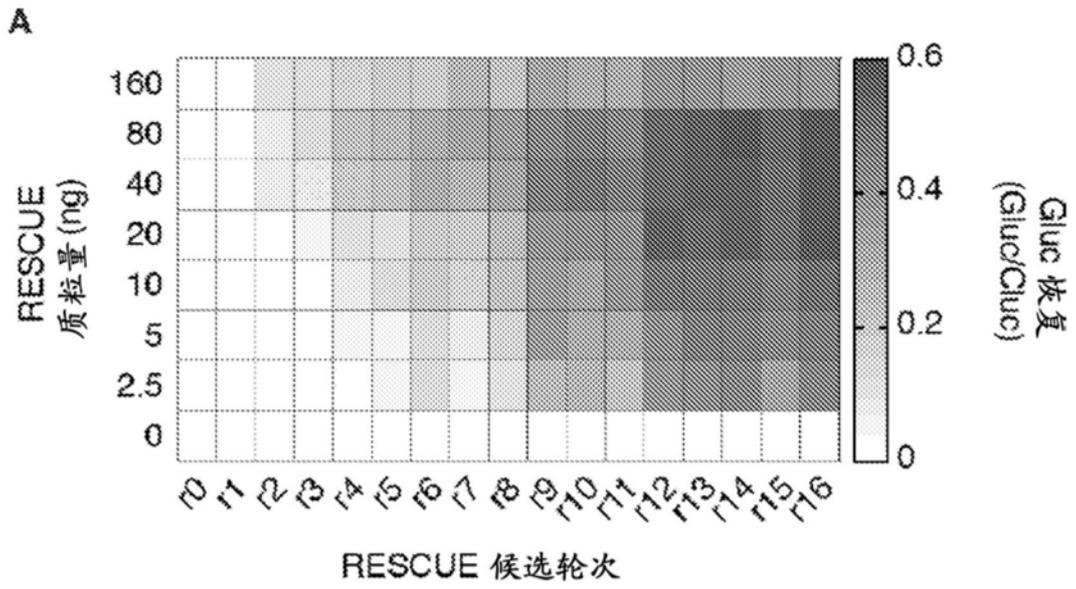


图111A

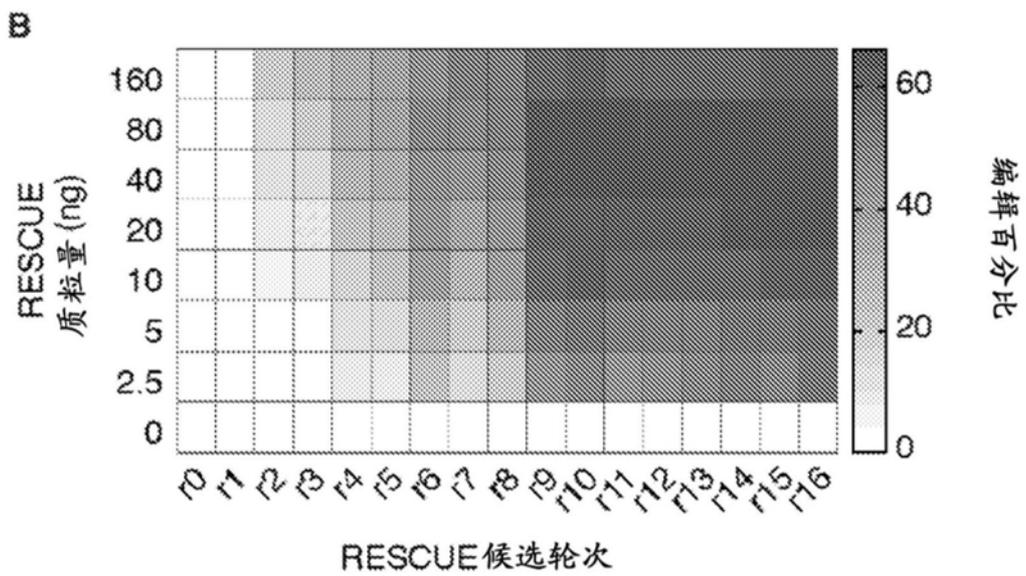


图111B

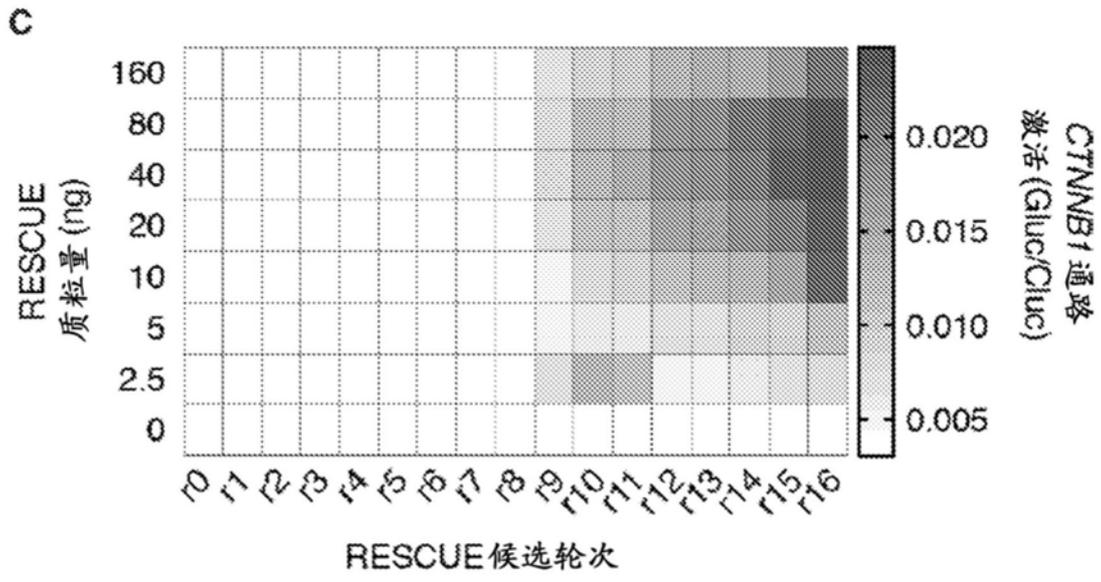


图111C

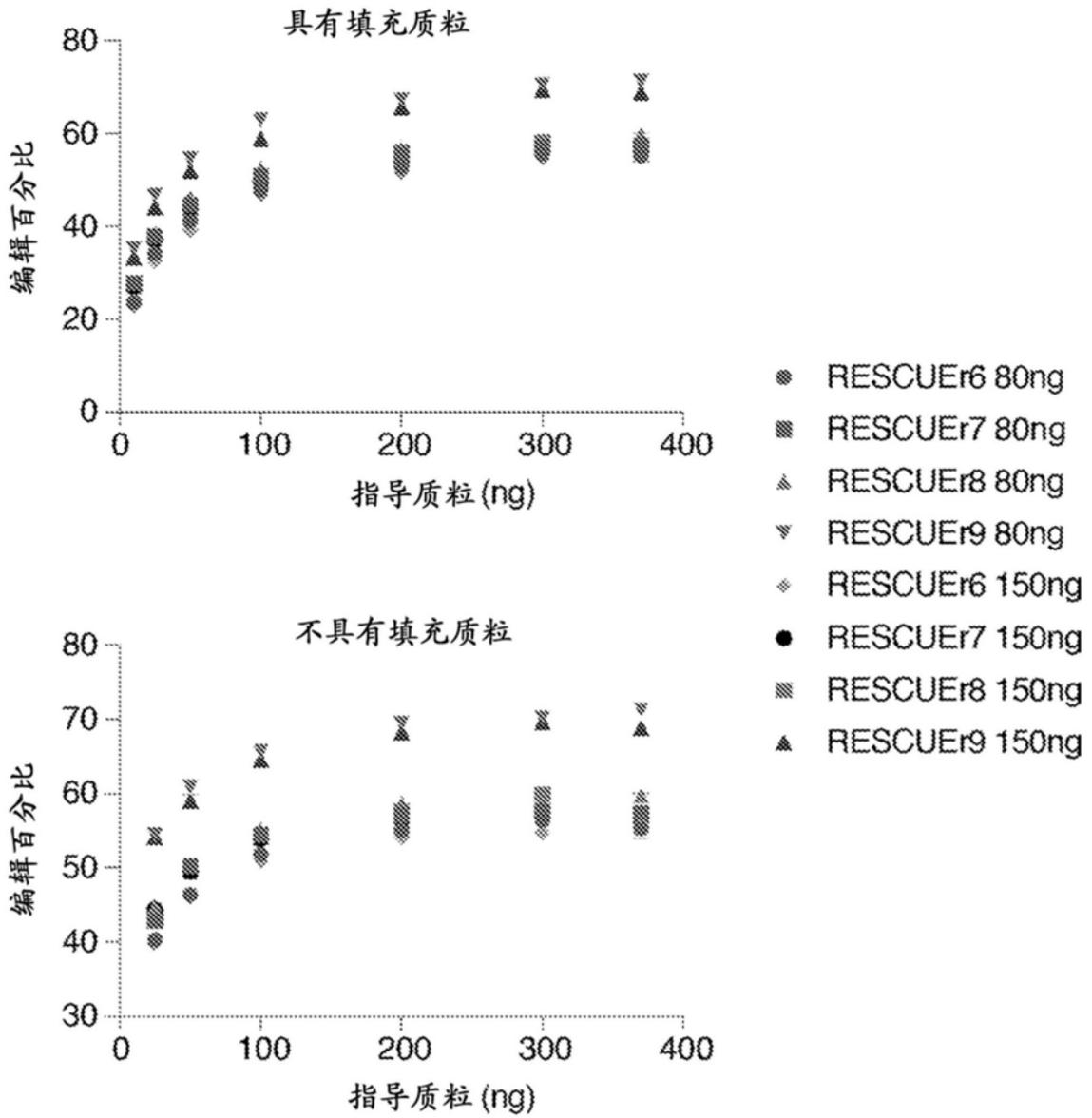


图112

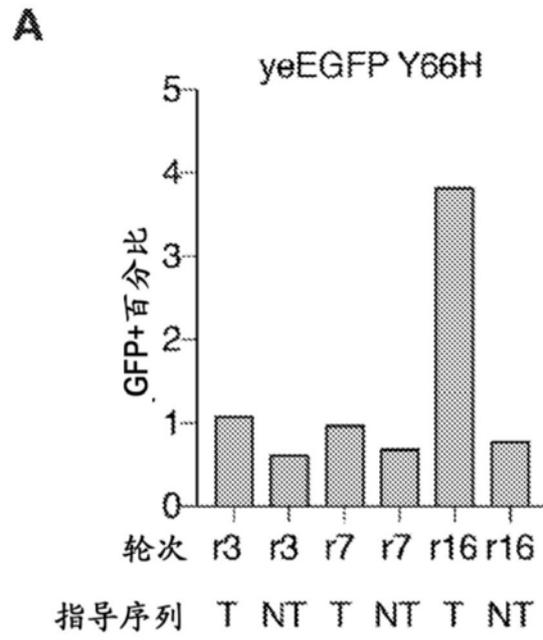


图113A

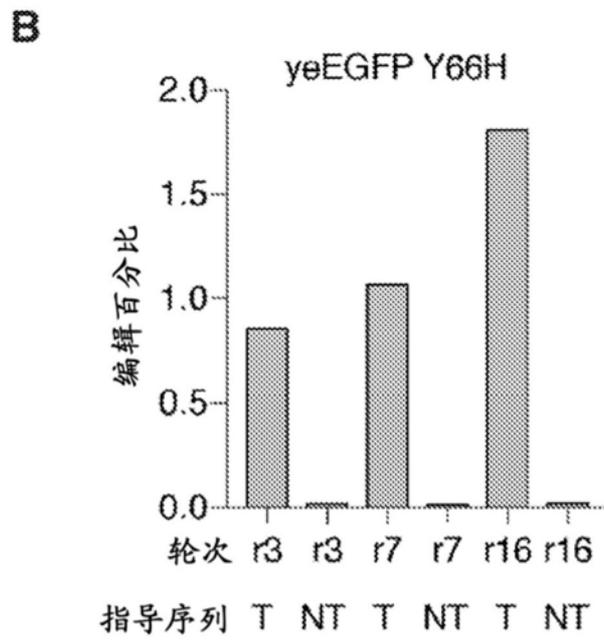


图113B

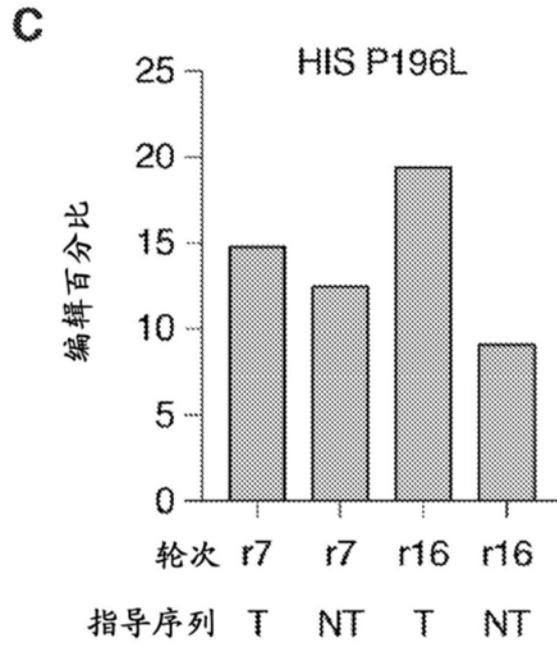


图113C

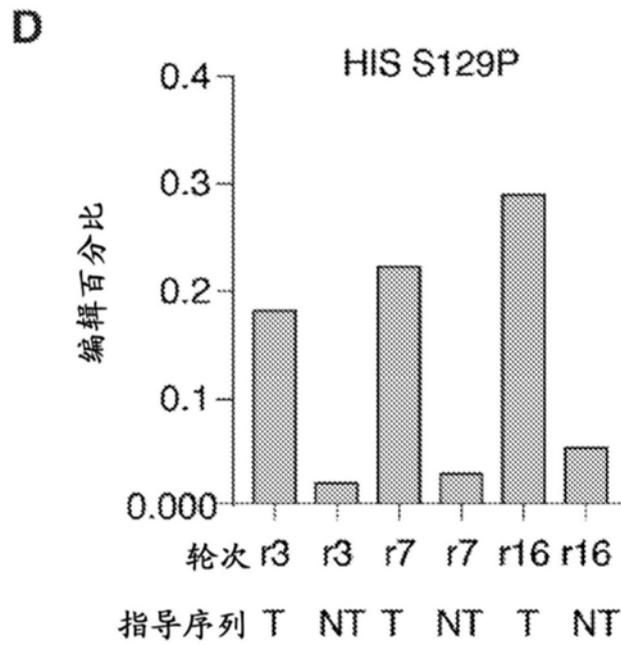


图113D

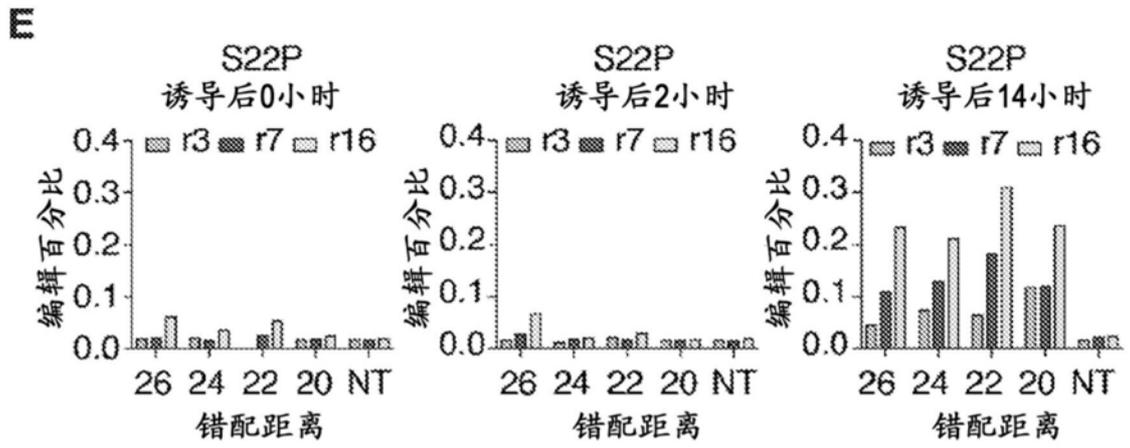


图113E

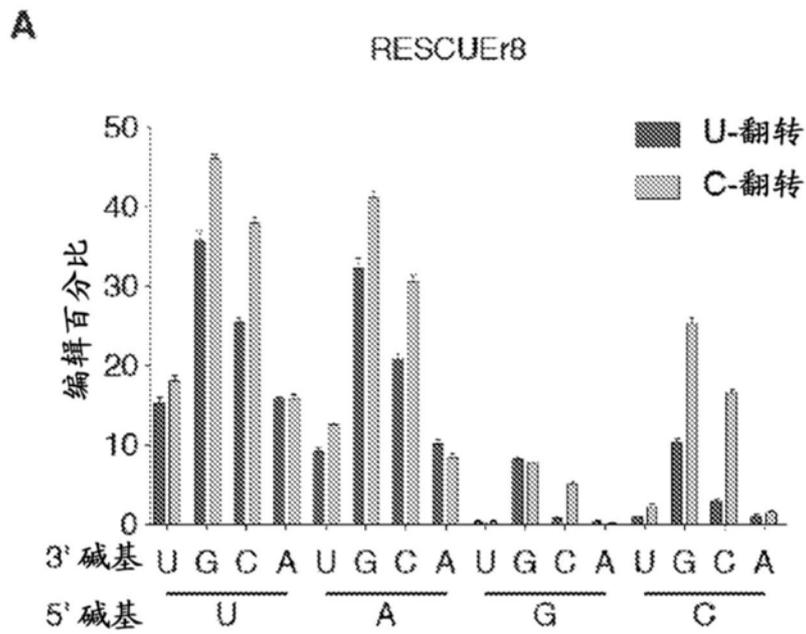


图114A

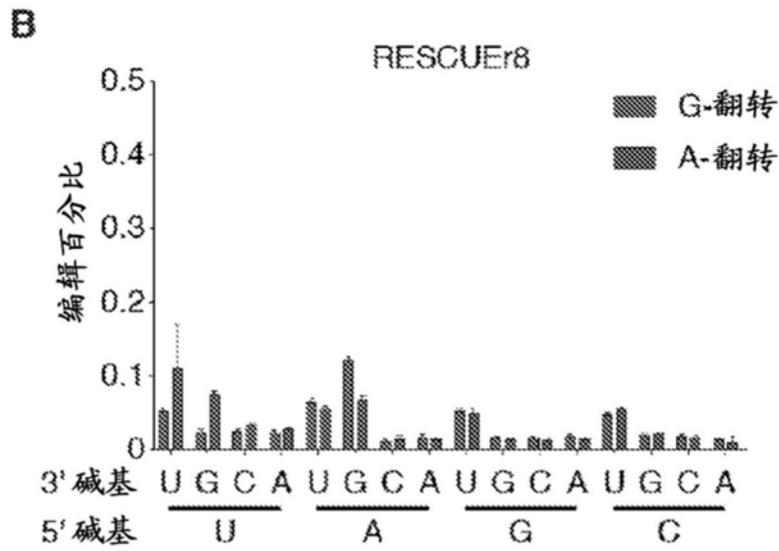


图114B

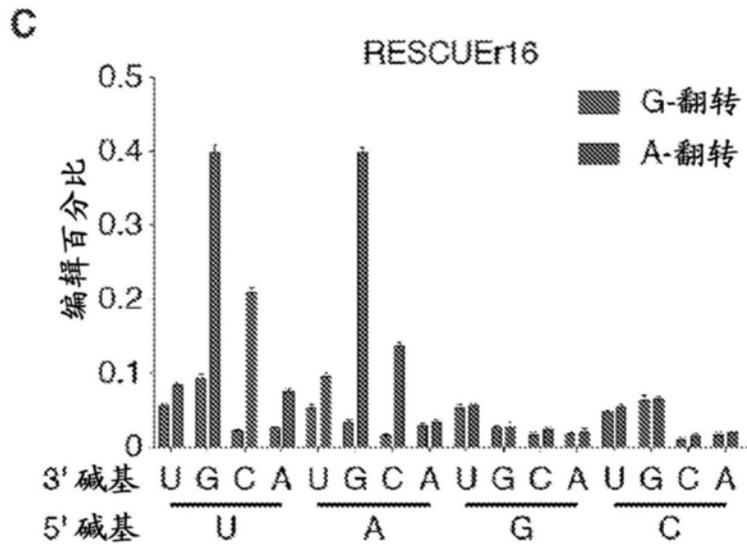


图114C

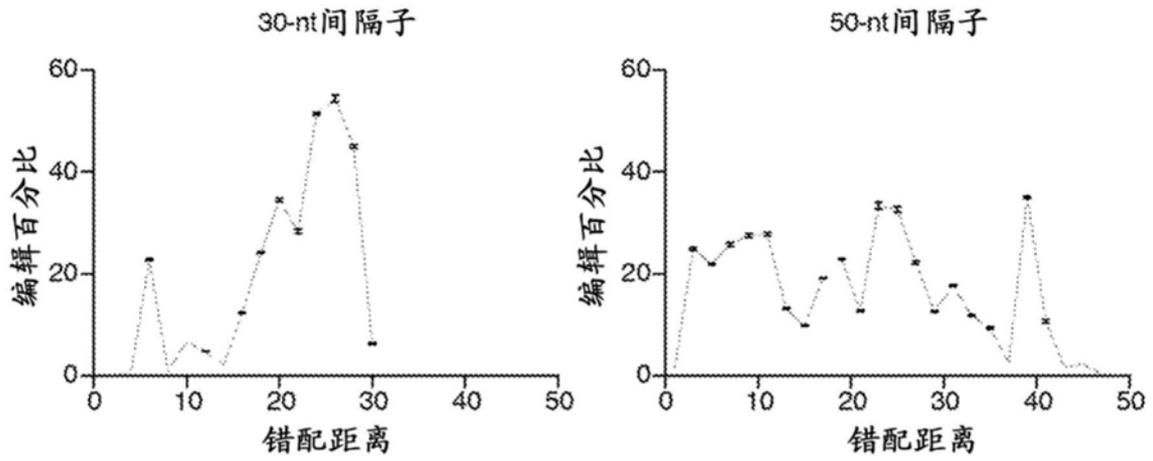


图115

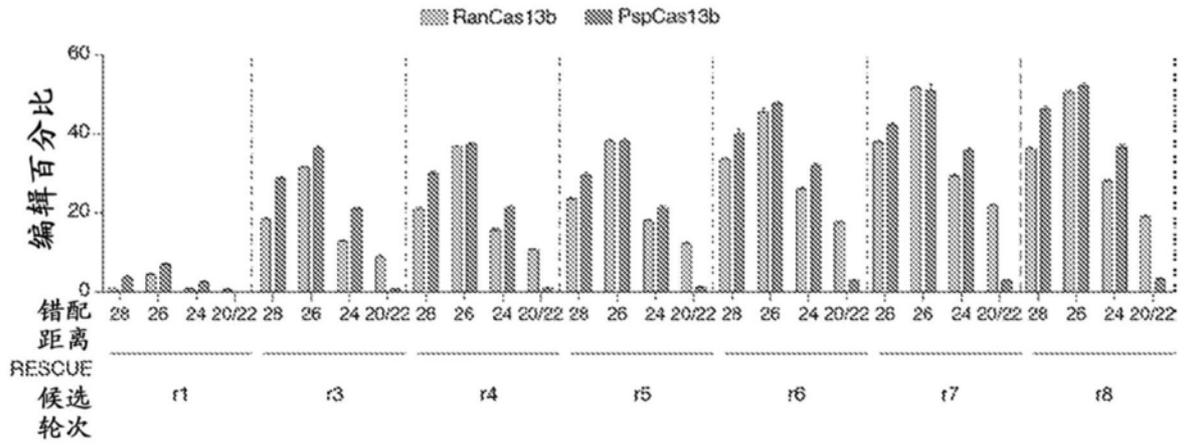


图116

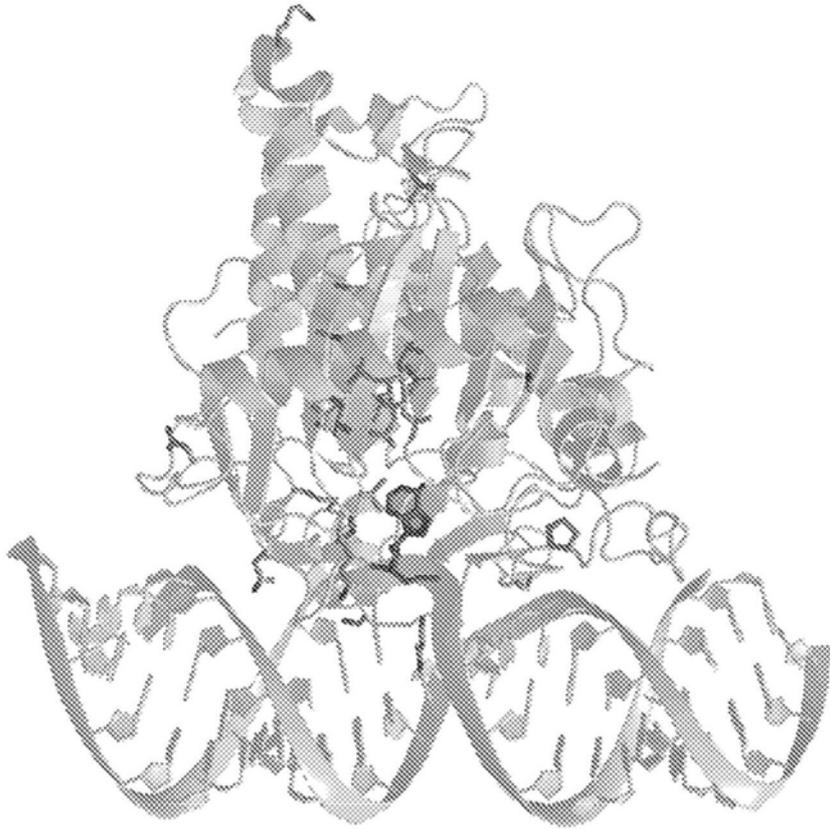


图117A

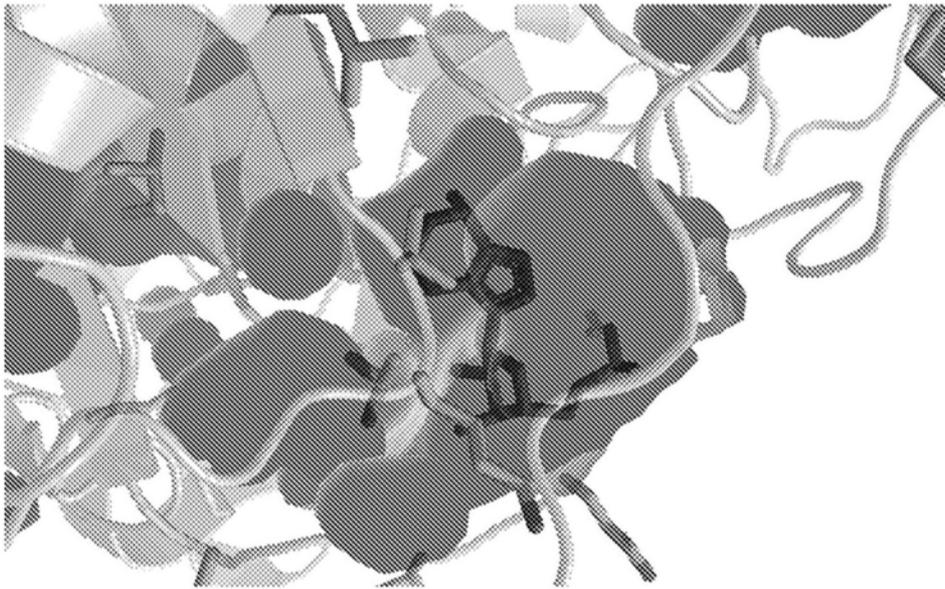


图117B

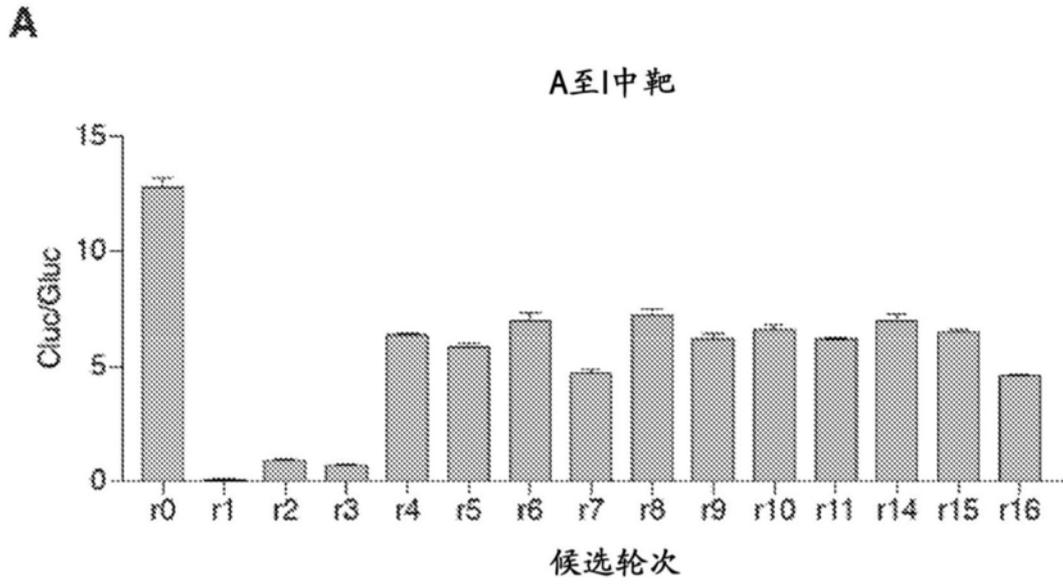


图118A

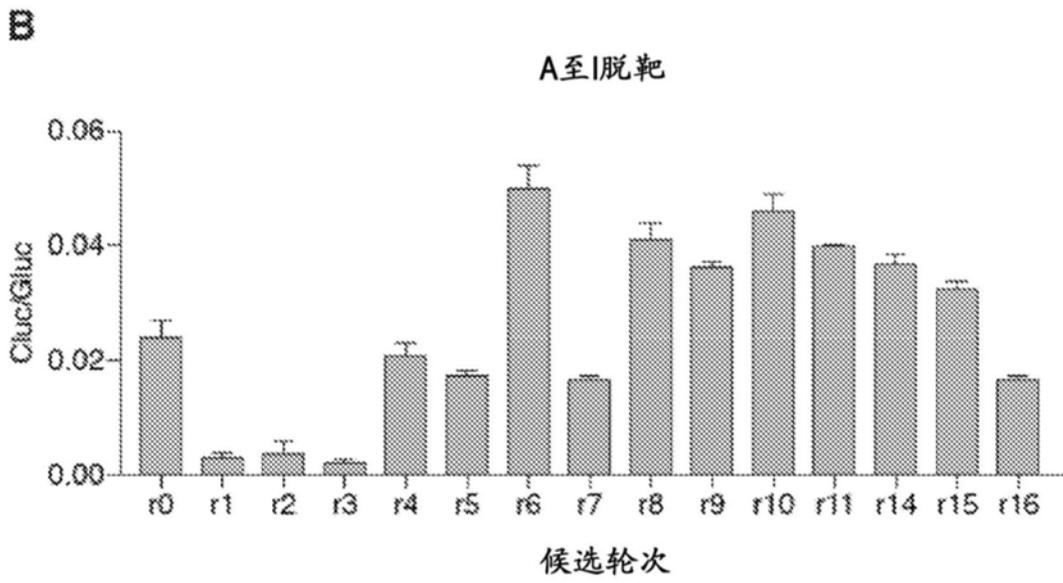


图118B

C

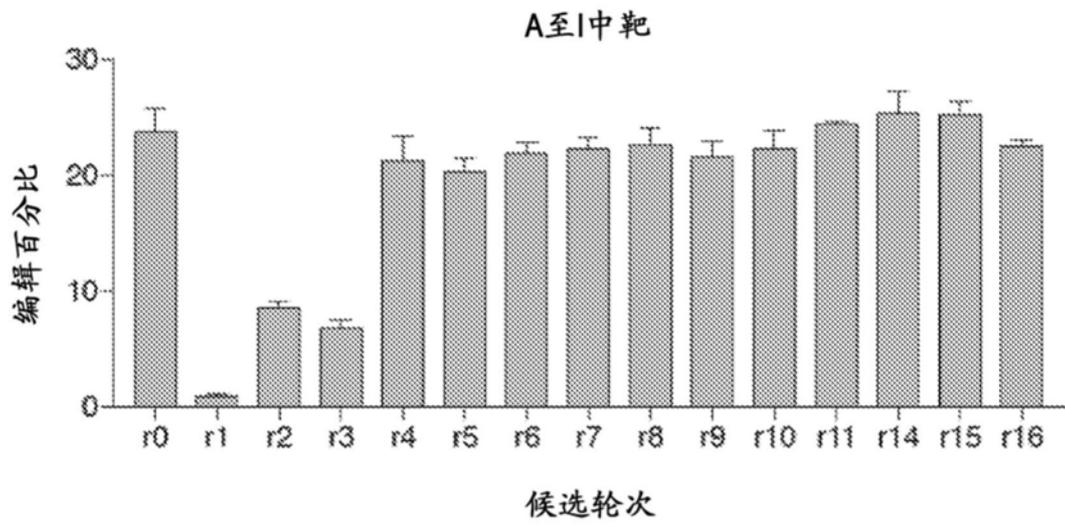


图118C

D

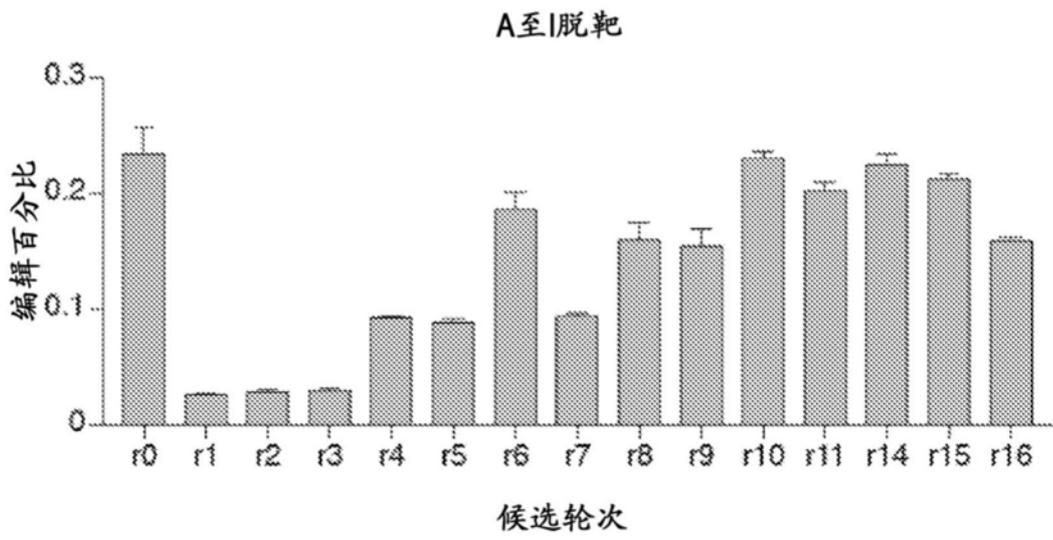


图118D

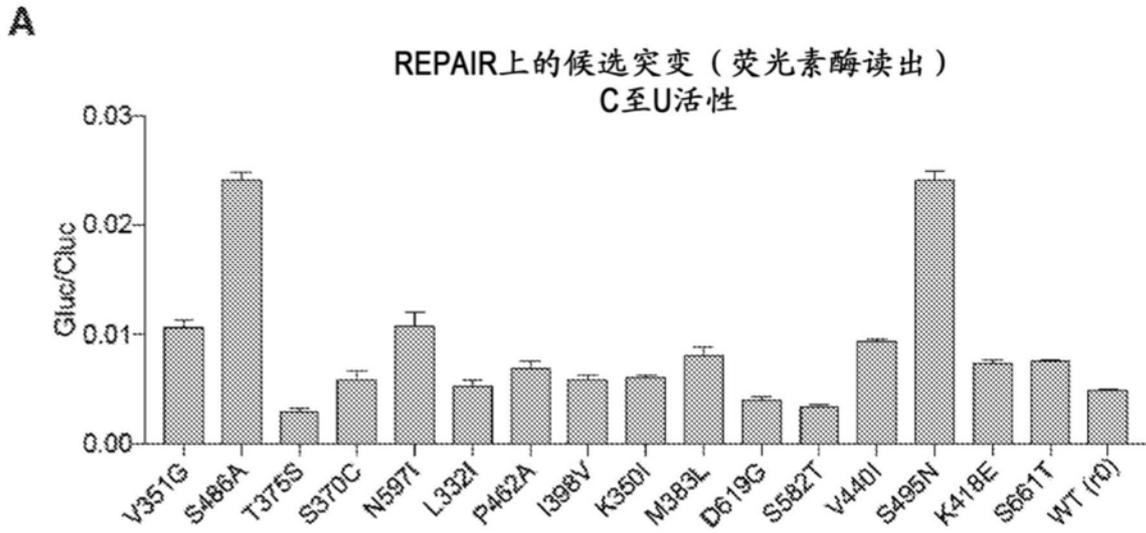


图119A

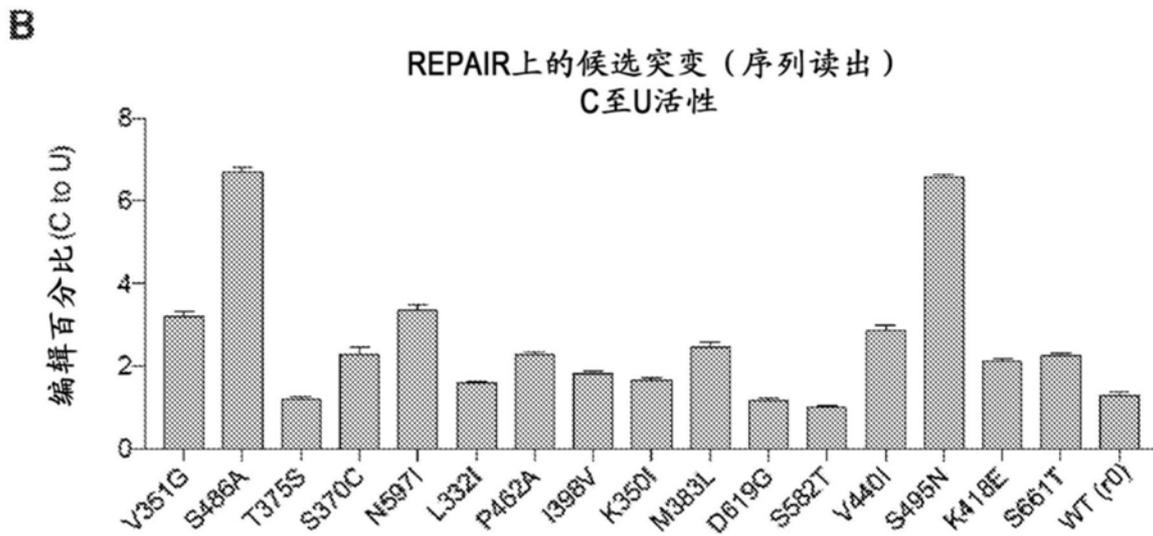


图119B

C

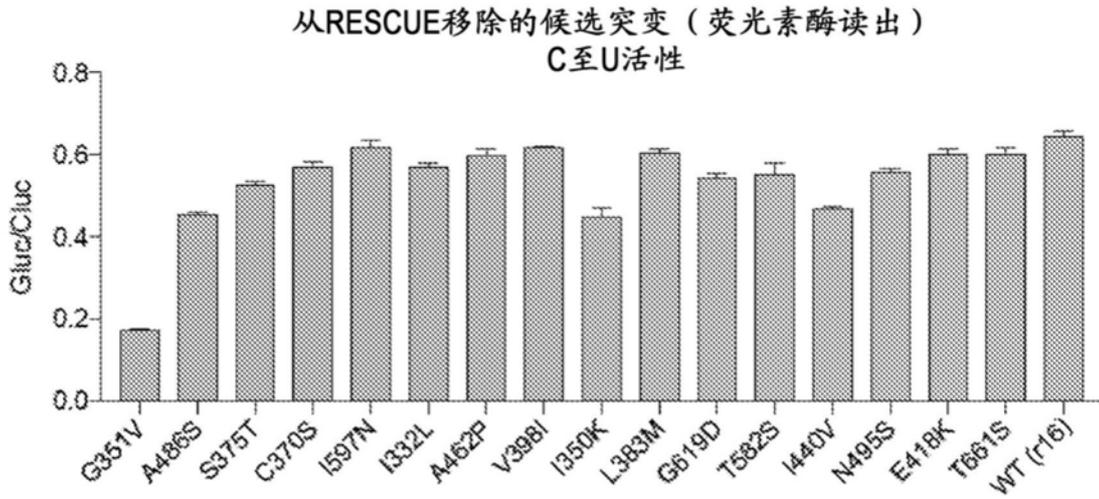


图119C

D

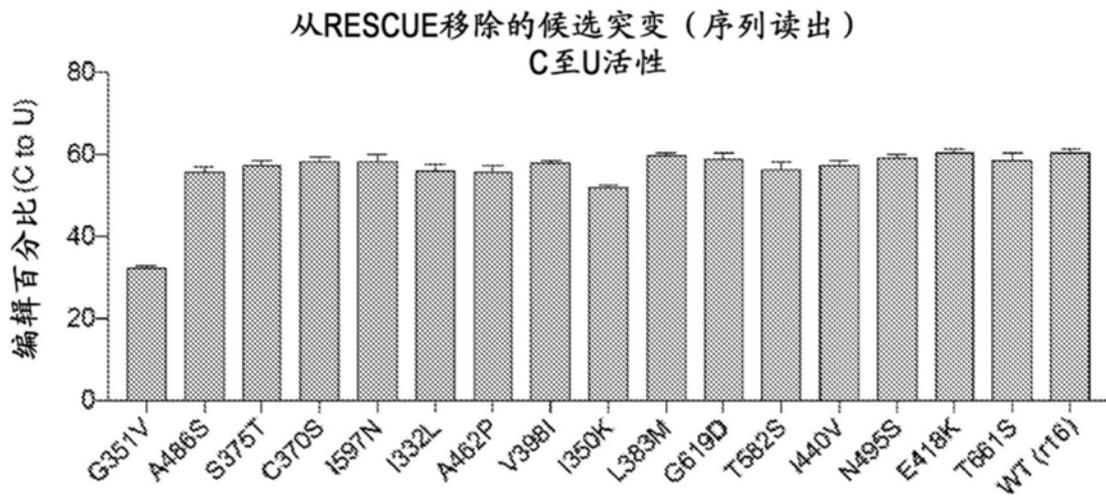


图119D

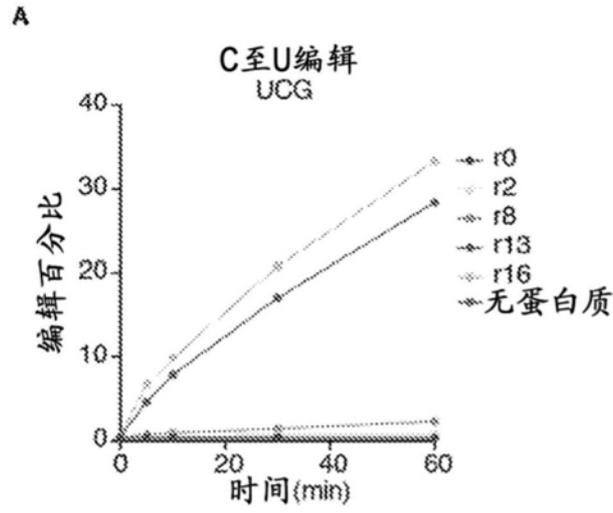


图120A

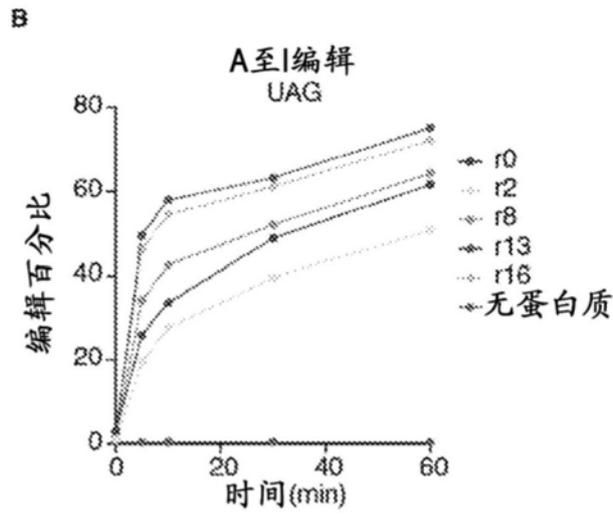


图120B

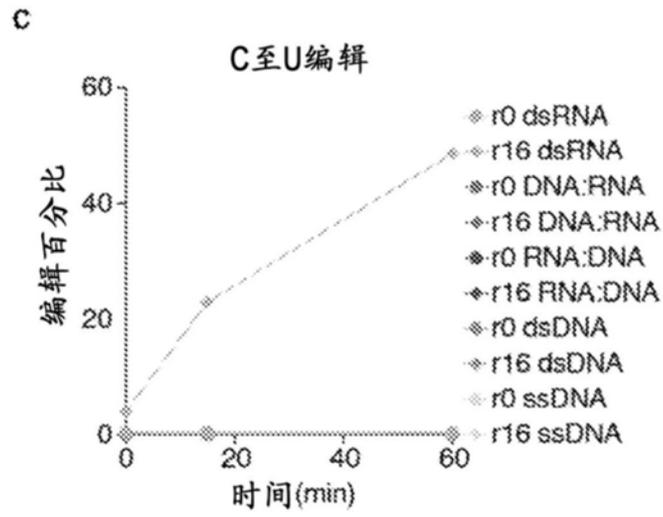


图120C

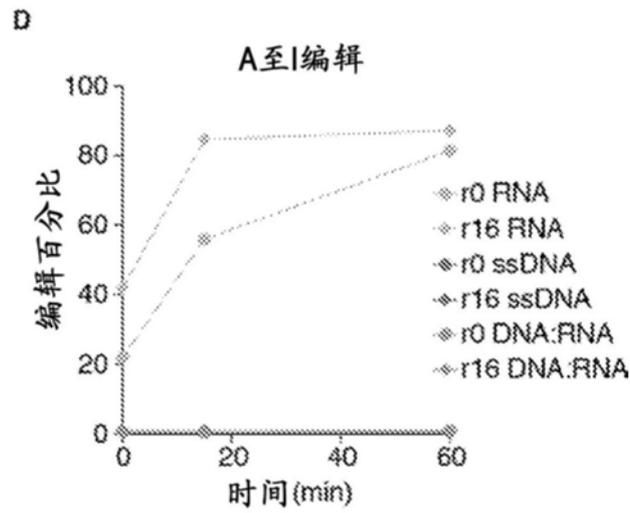


图120D

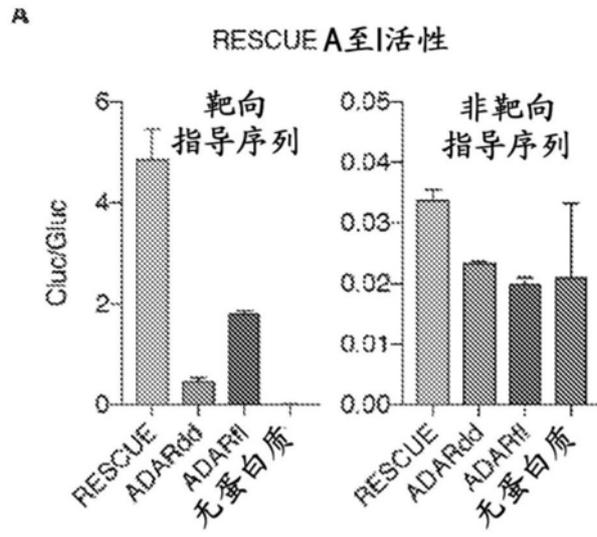


图121A

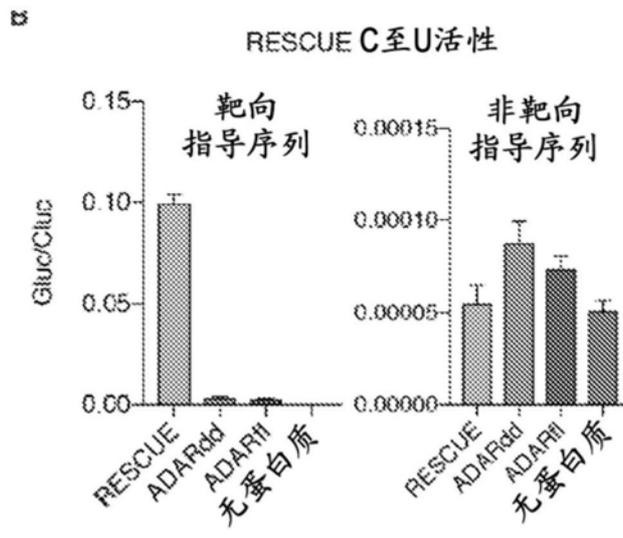


图121B

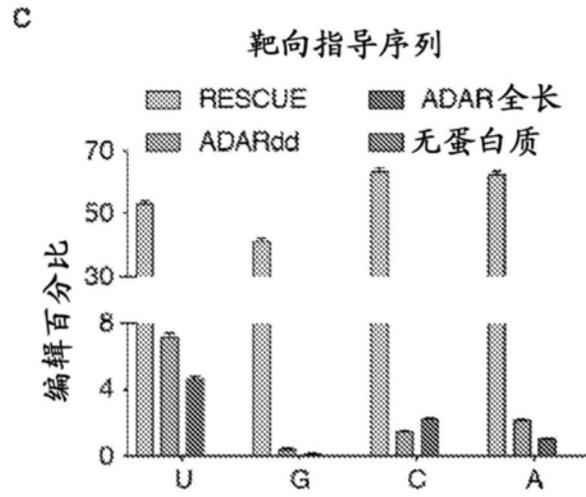


图121C

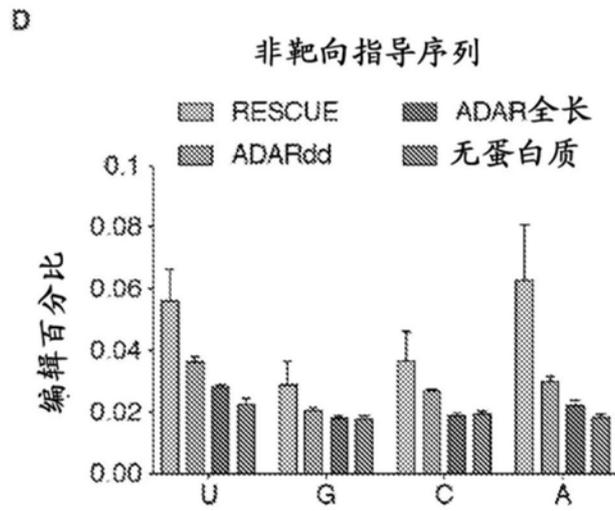


图121D

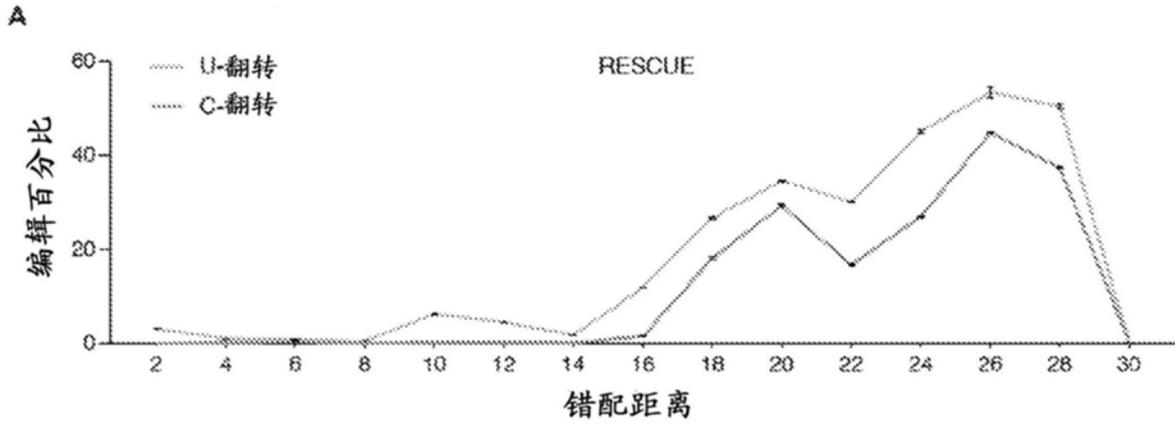


图122A

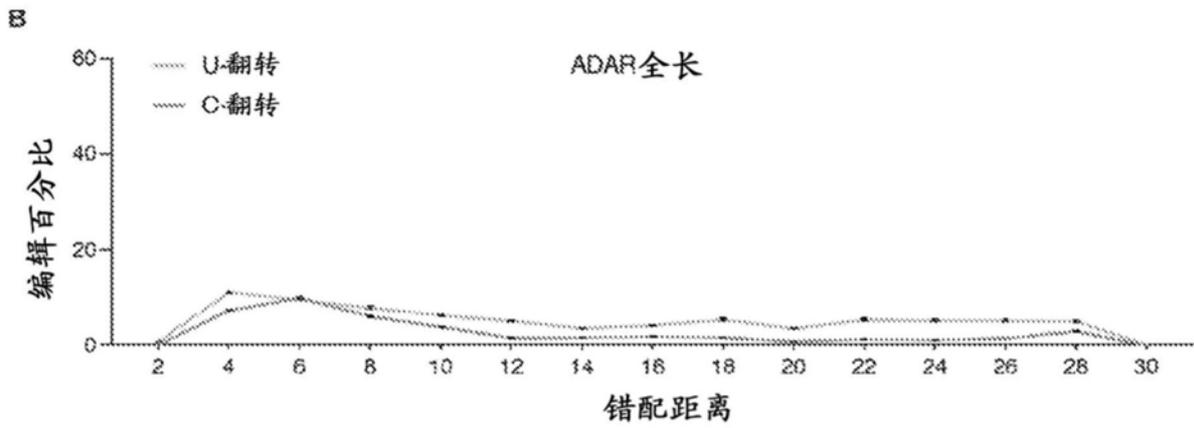


图122B

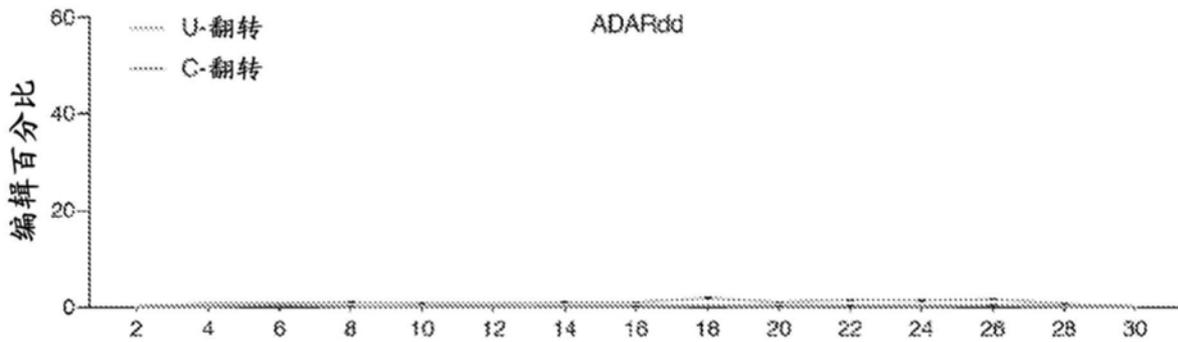


图122C

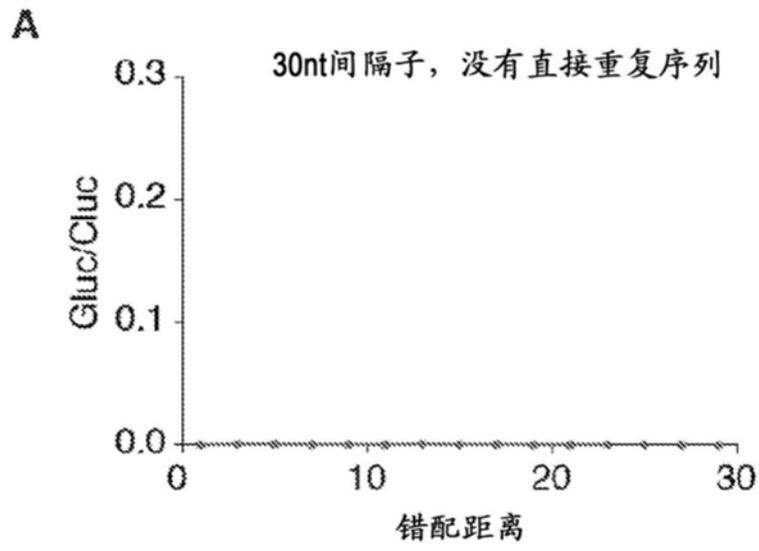


图123A

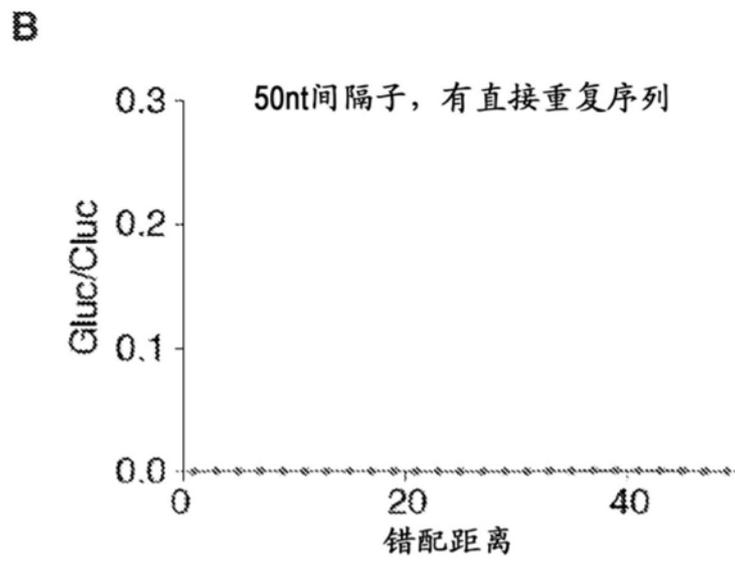


图123B

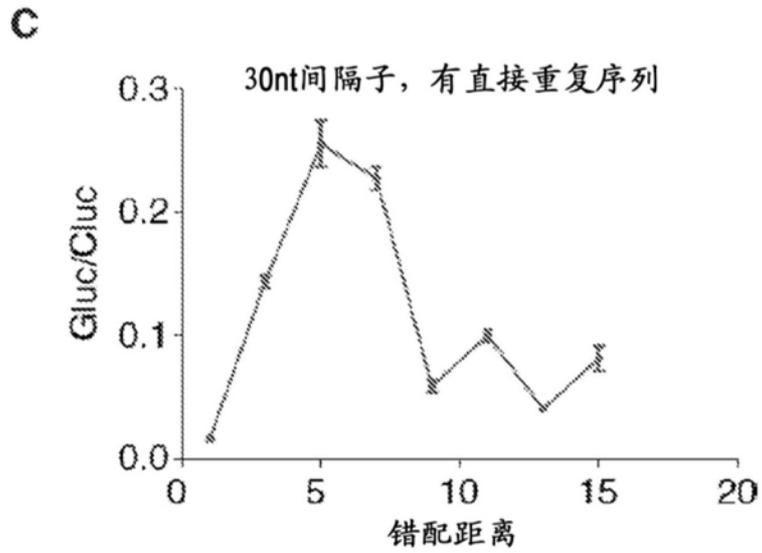


图123C

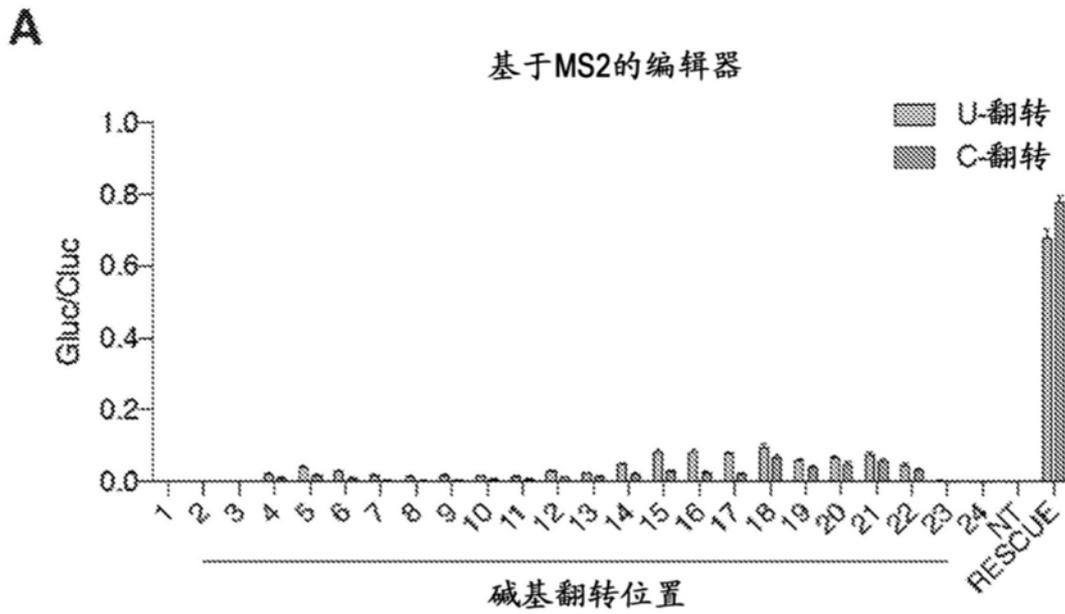


图124A

**B**

基于MS2的编辑器

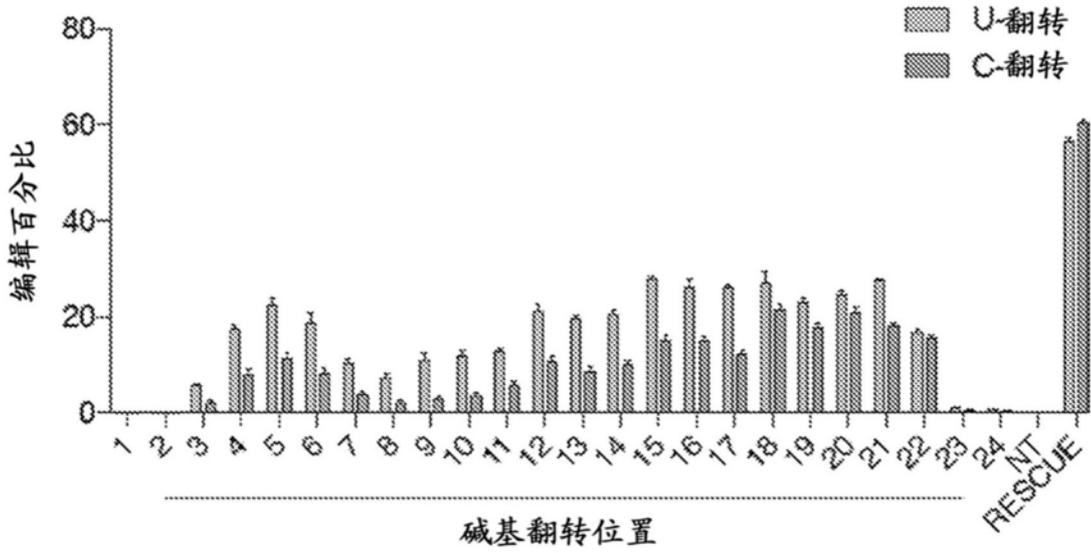


图124B

**C**

基于adRNA的编辑器

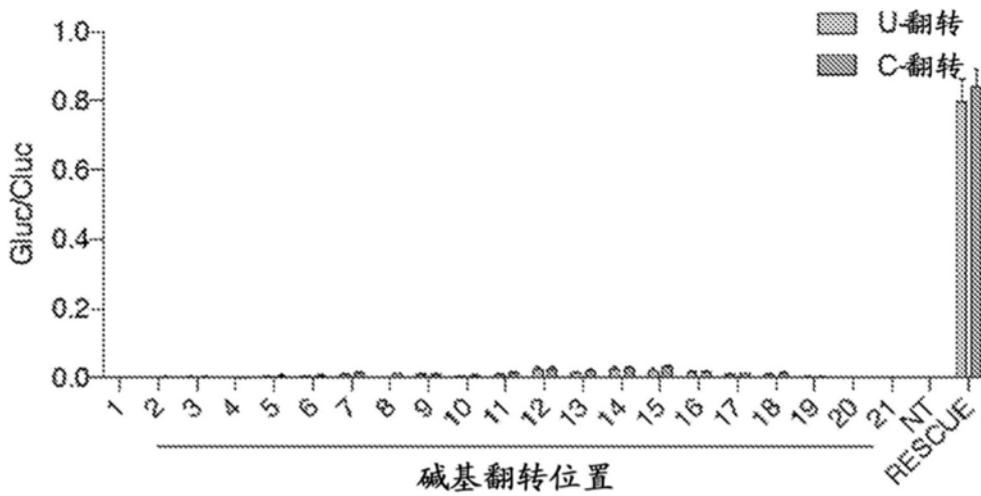


图124C

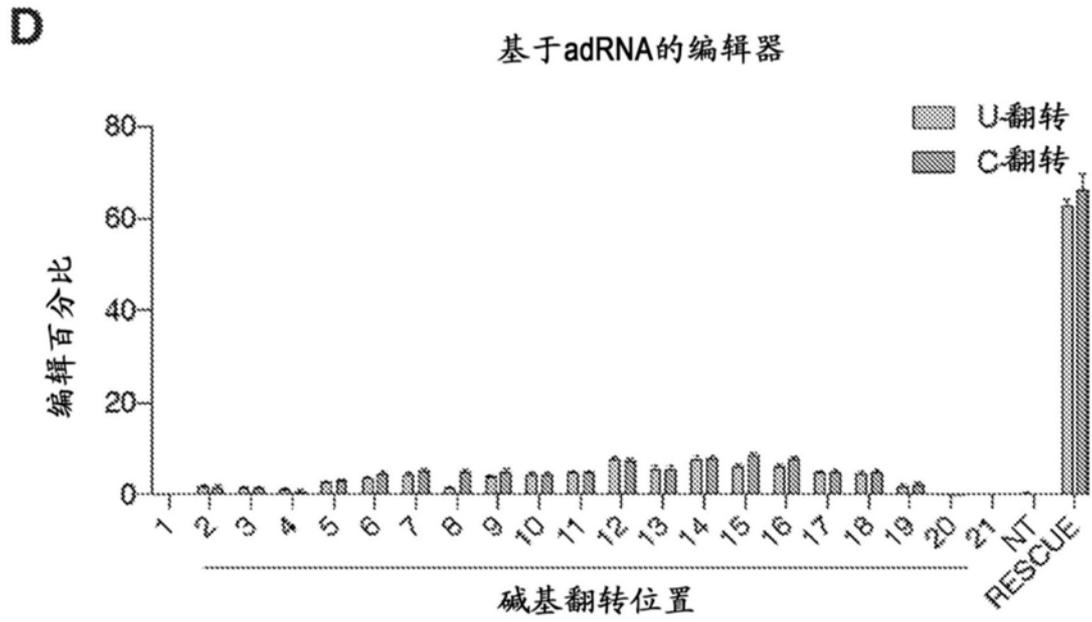


图124D

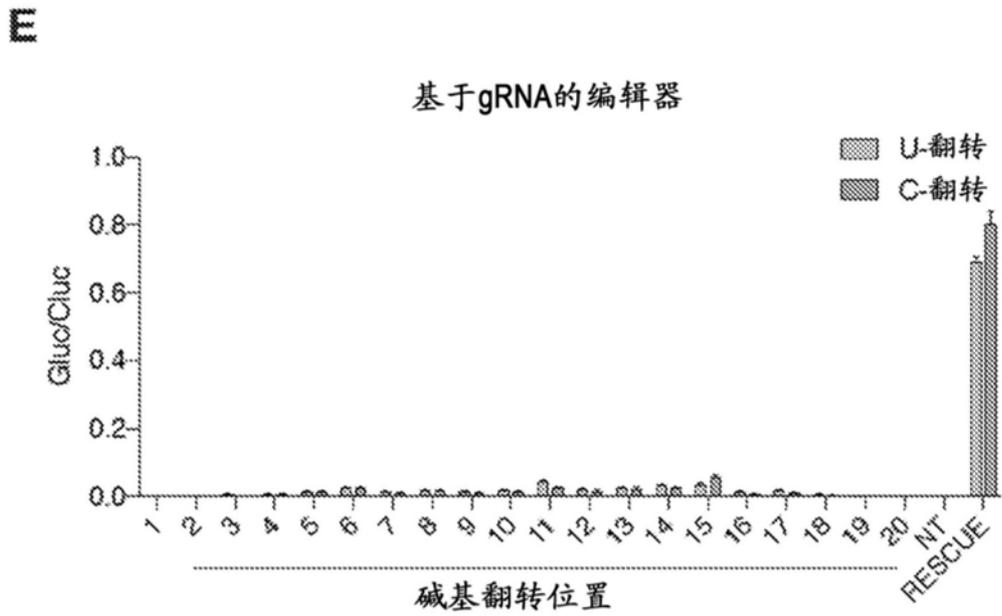


图124E

F

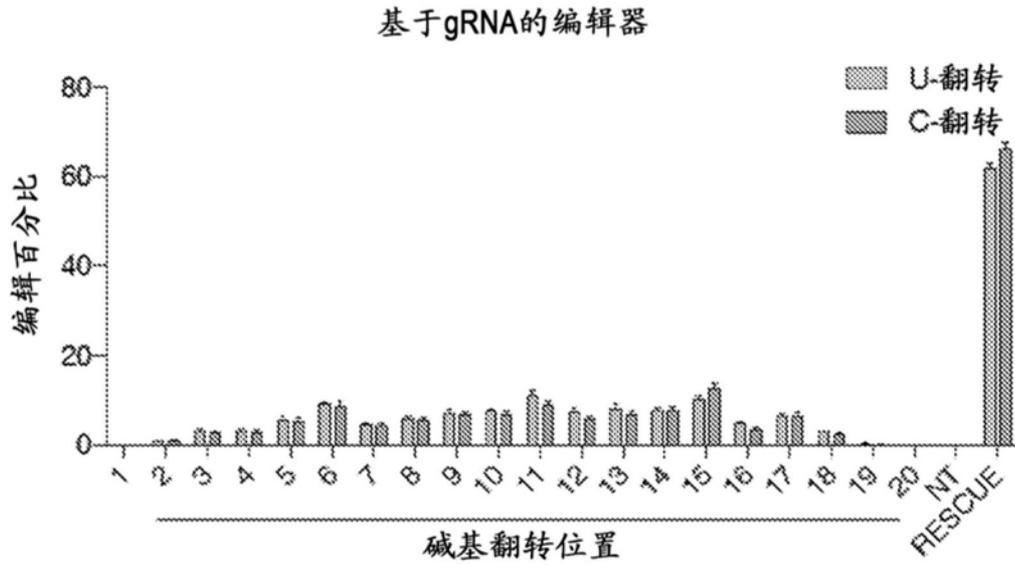


图124F

A

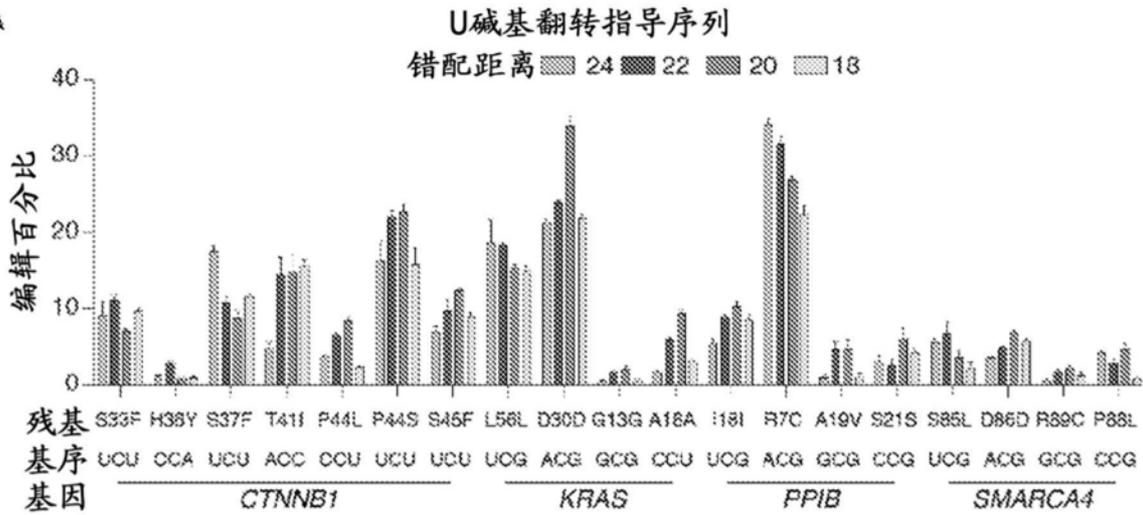


图125A

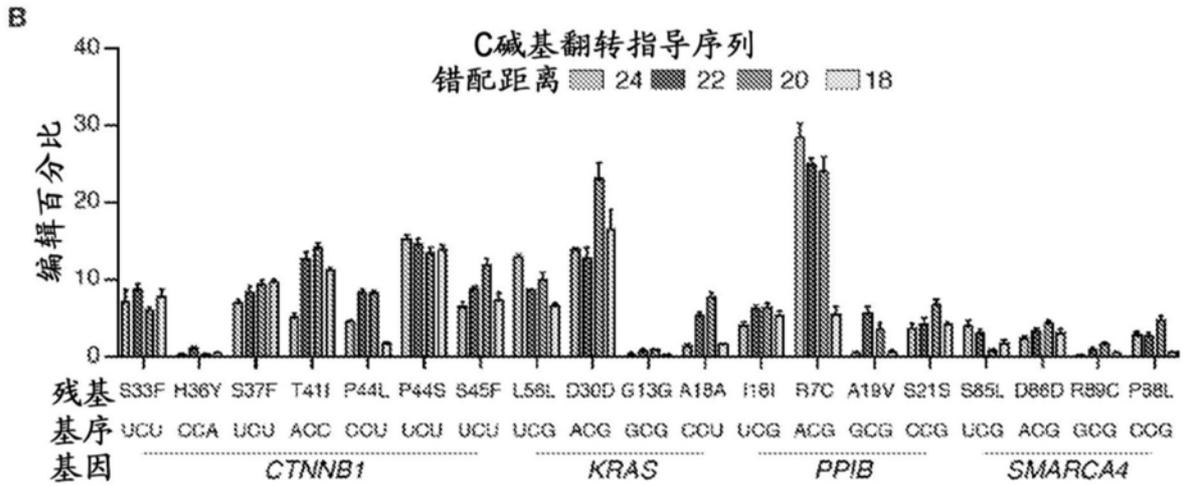


图125B

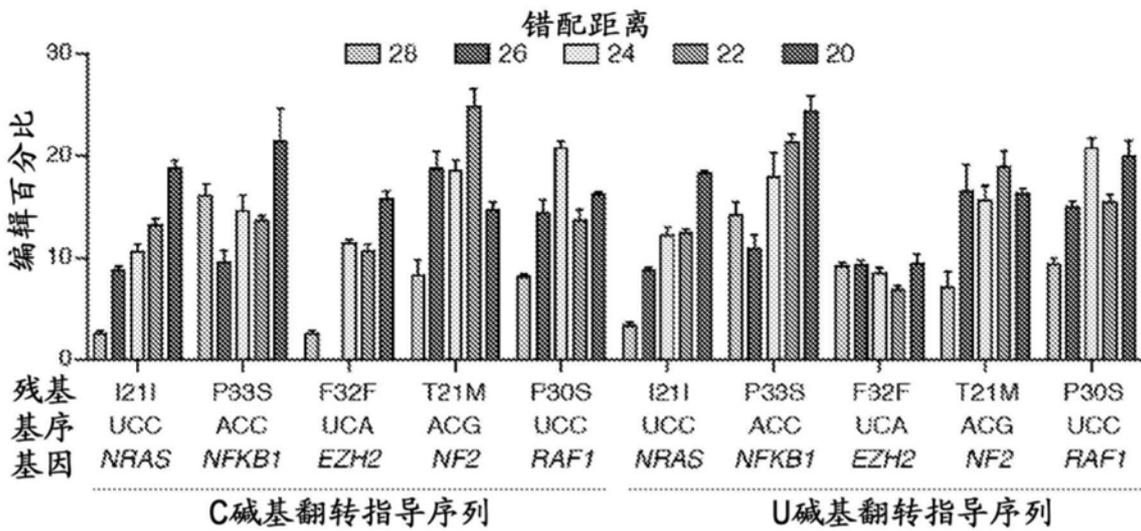


图125C

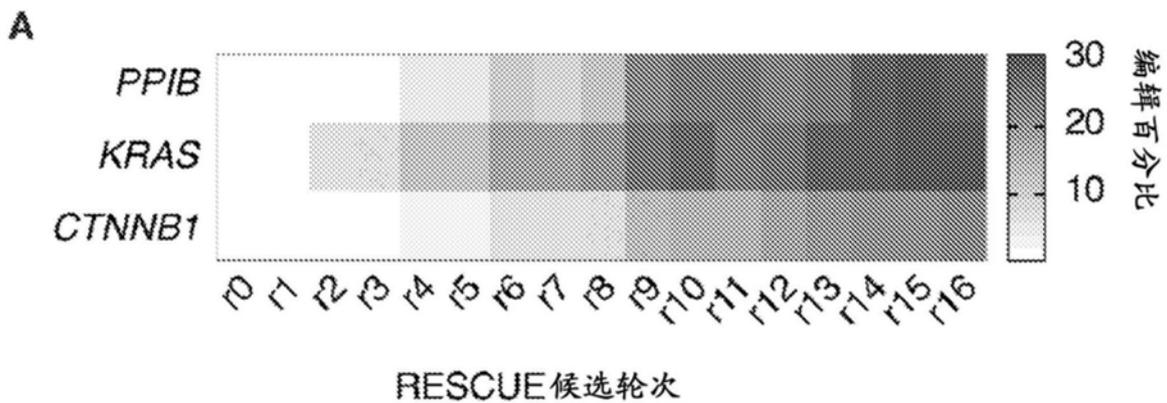


图126A

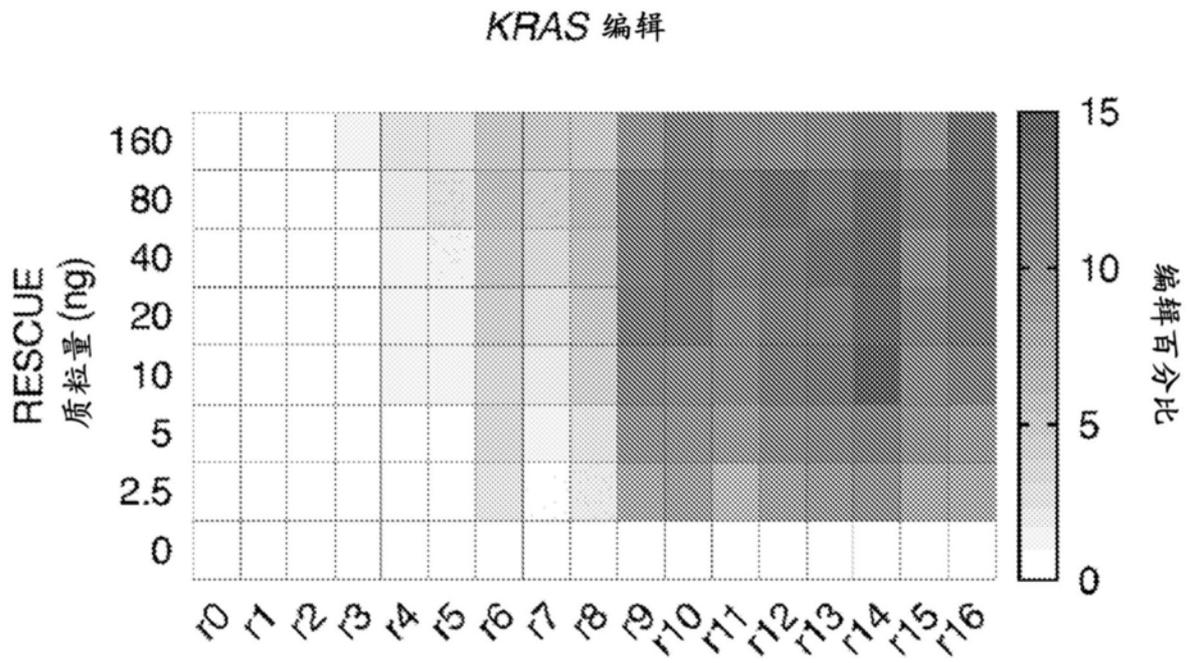


图126B

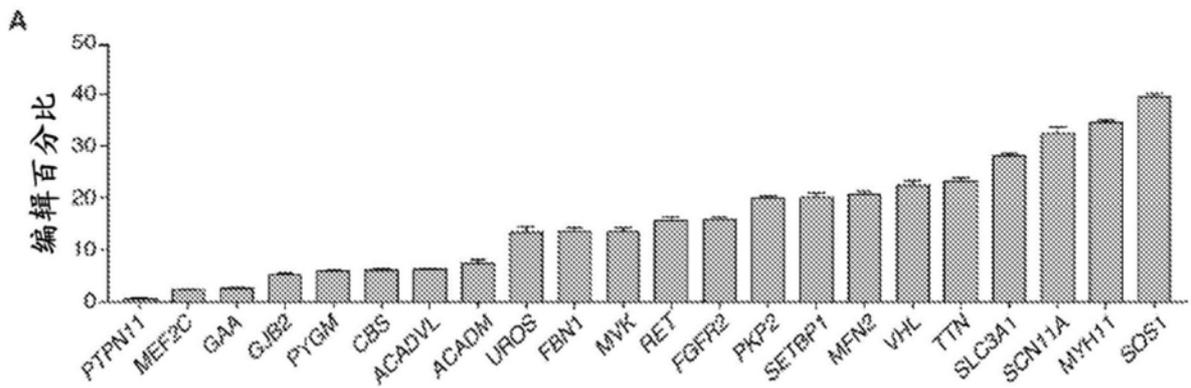


图127A

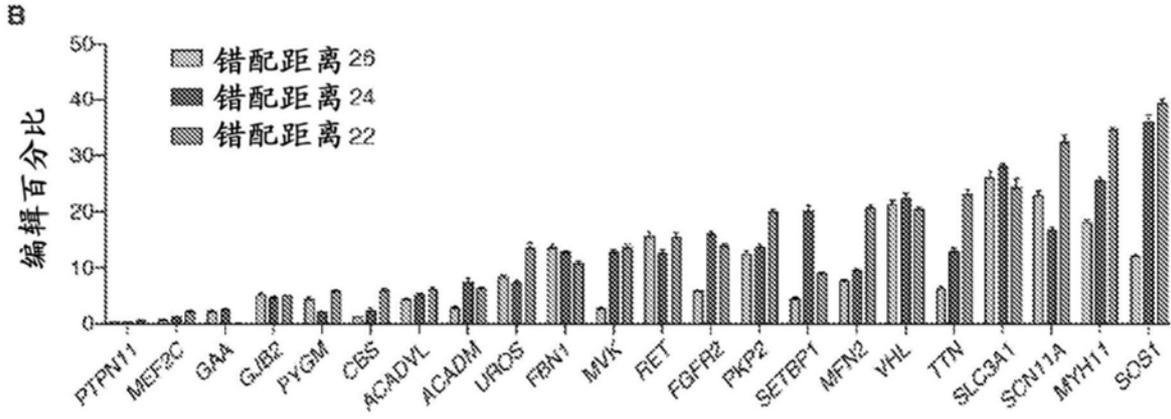


图127B

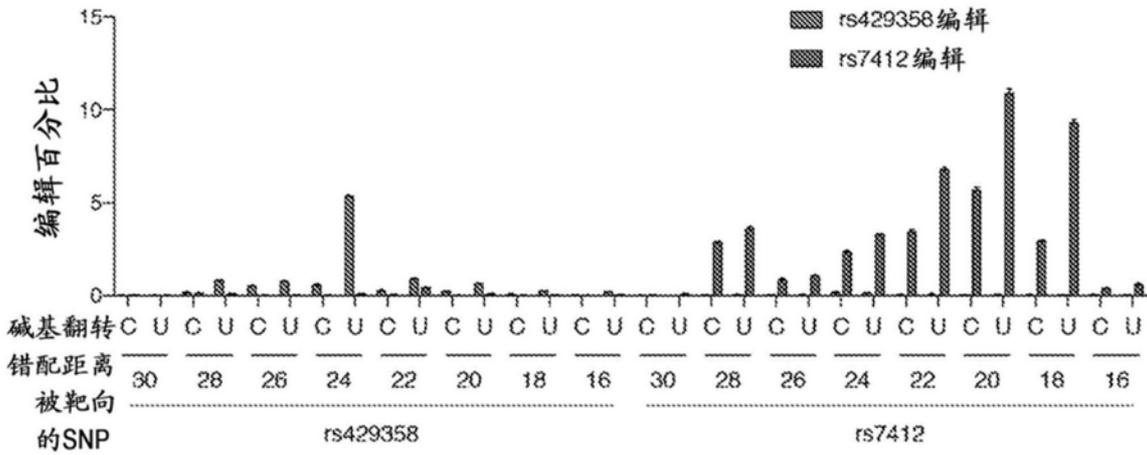


图128

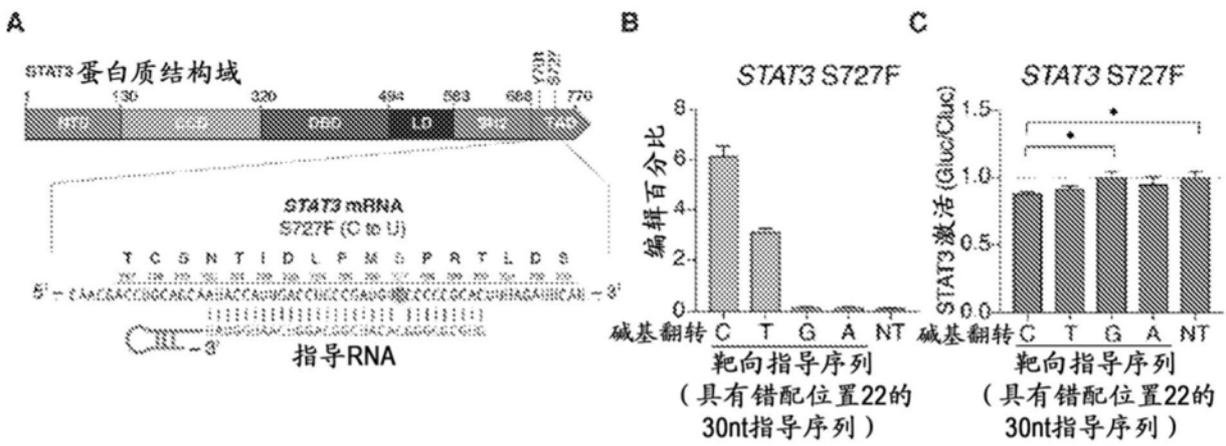


图129A

图129B

图129C

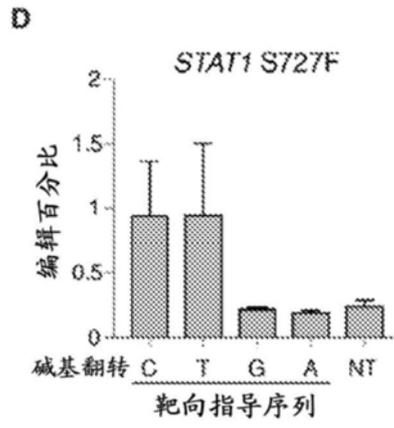


图129D

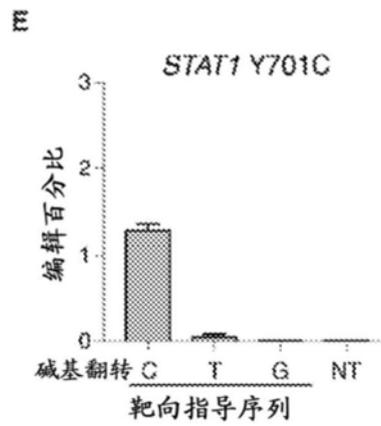


图129E

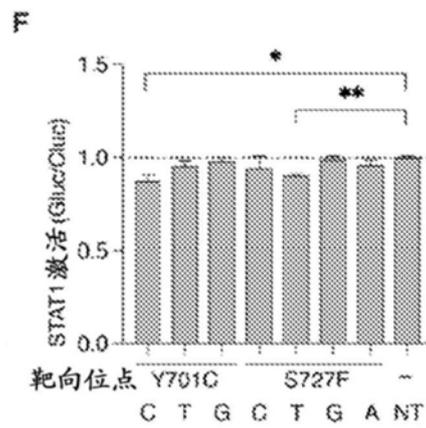


图129F

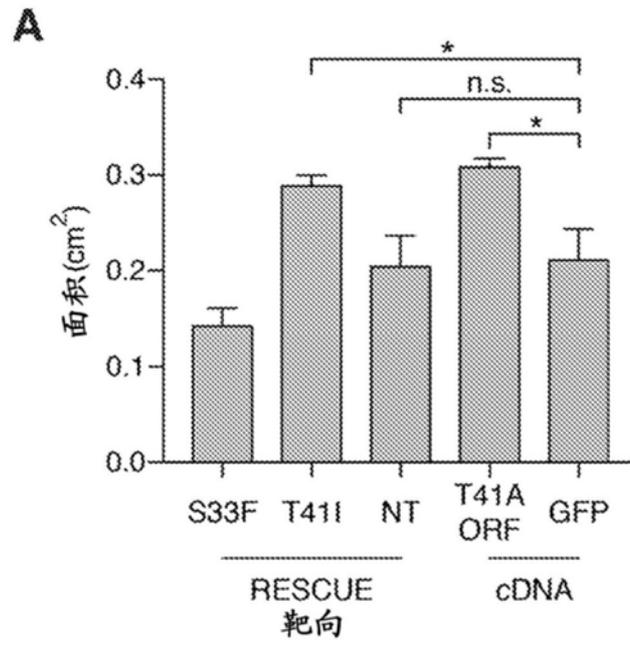


图130A

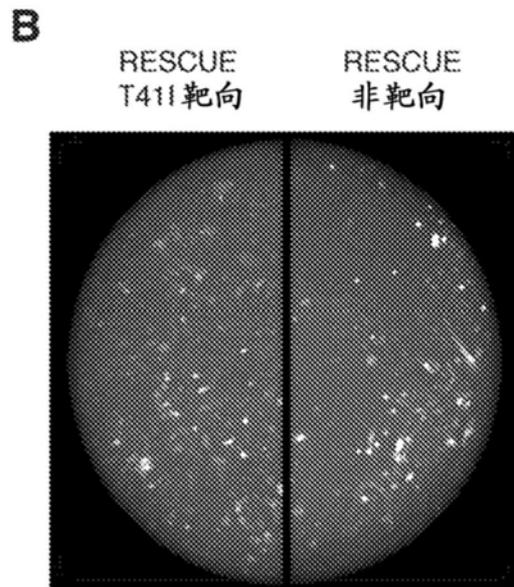


图130B

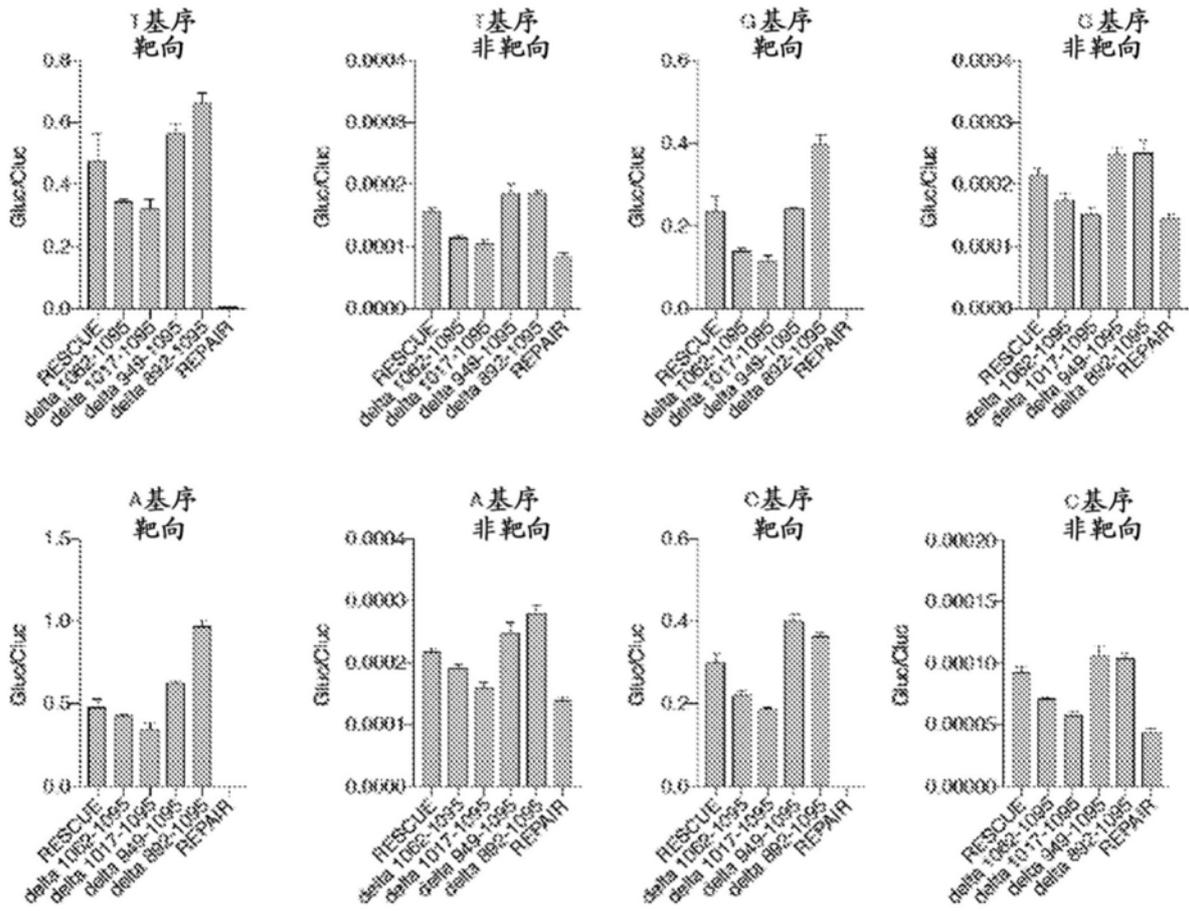


图131

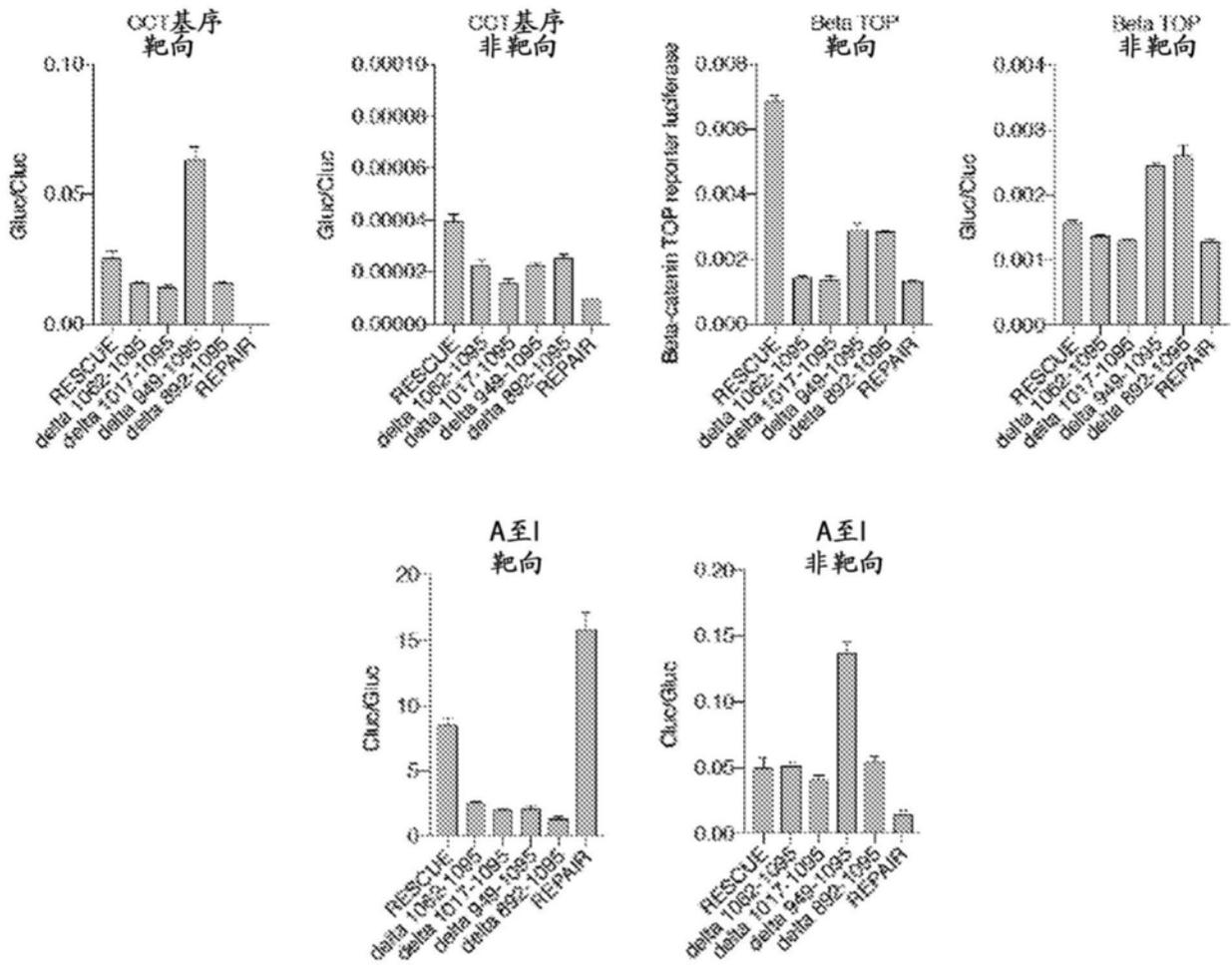


图131 (续)

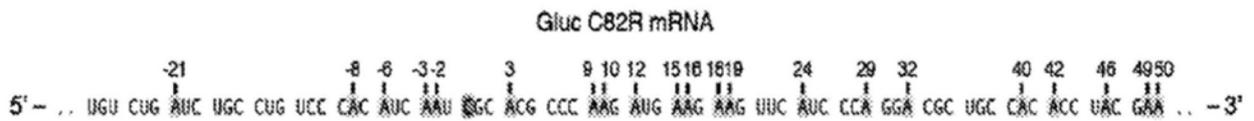


图132A

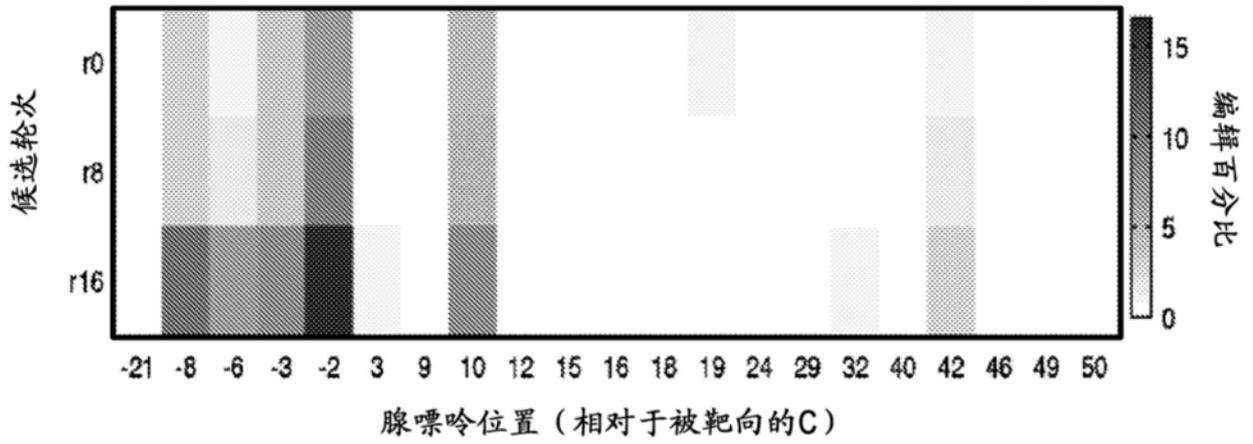


图132B

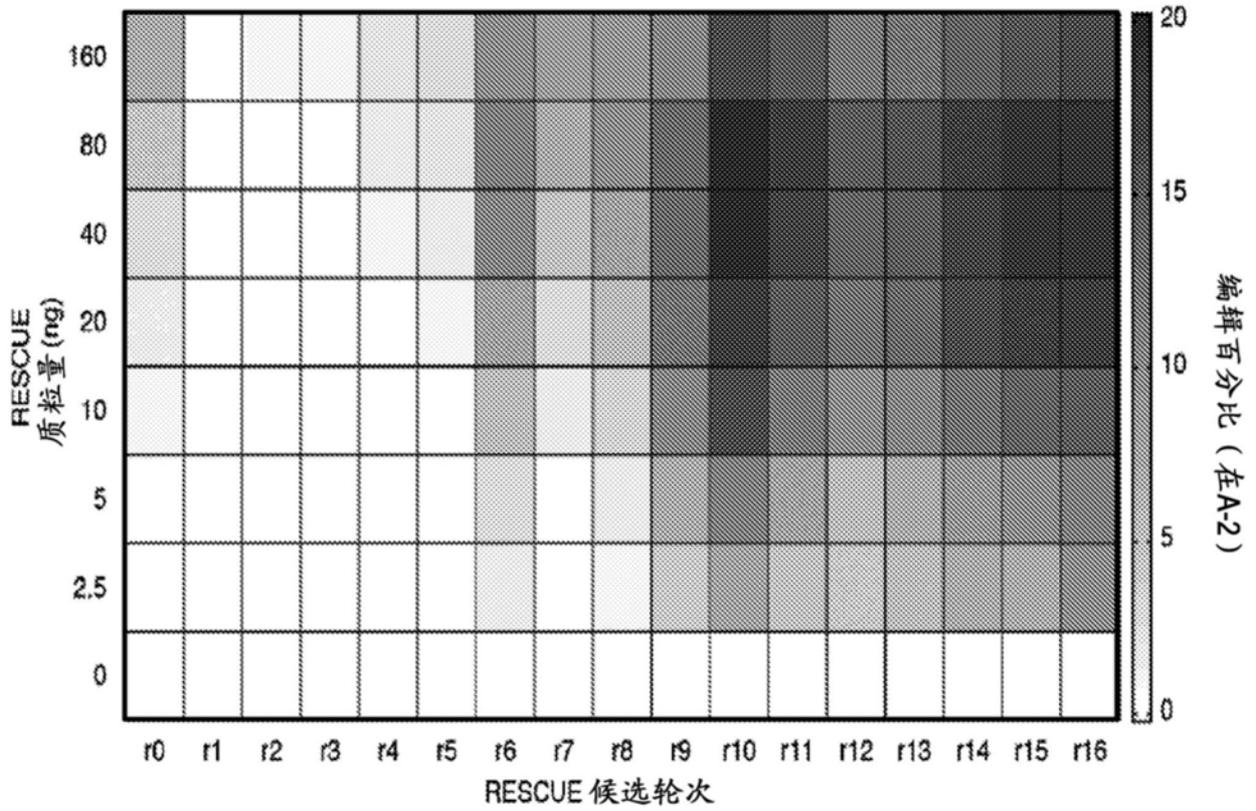


图132C



图133A

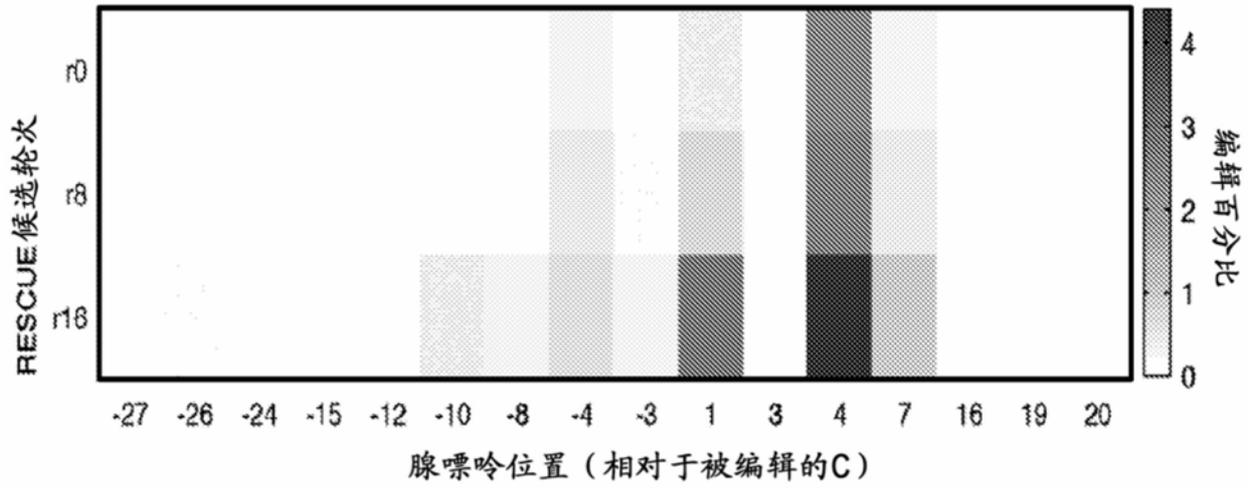


图133B

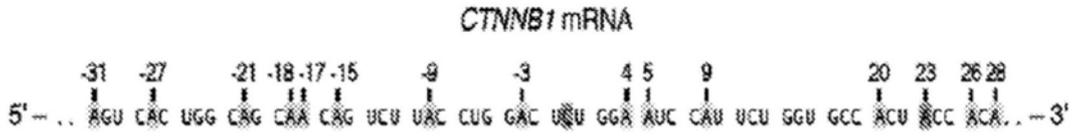


图133C

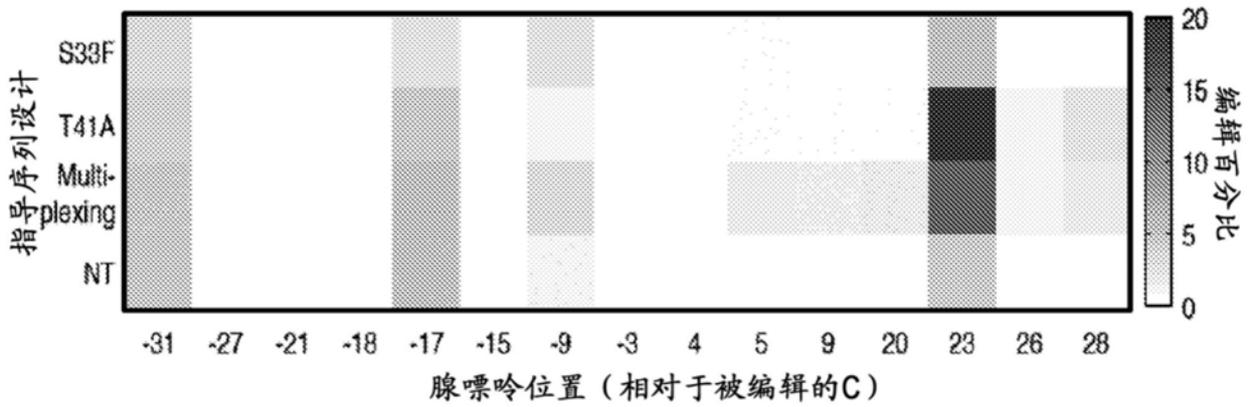


图133D

A

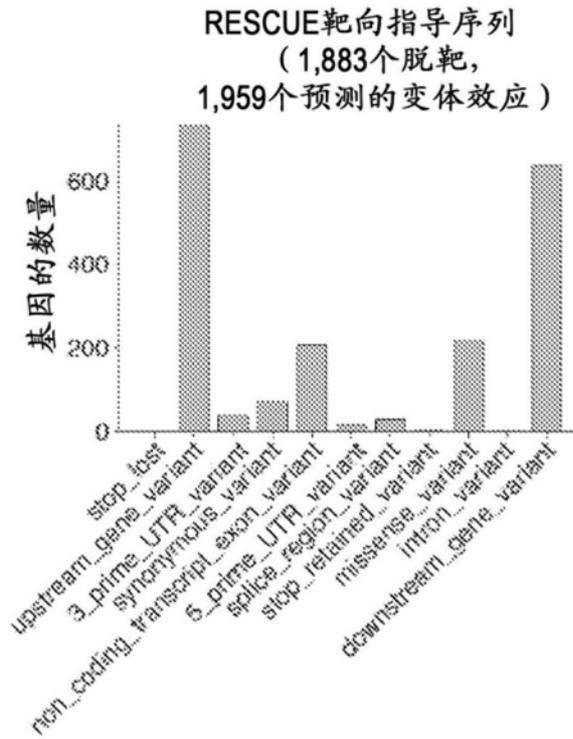


图134A

B

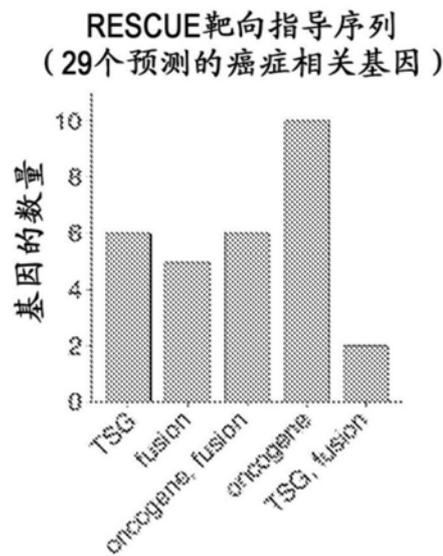


图134B

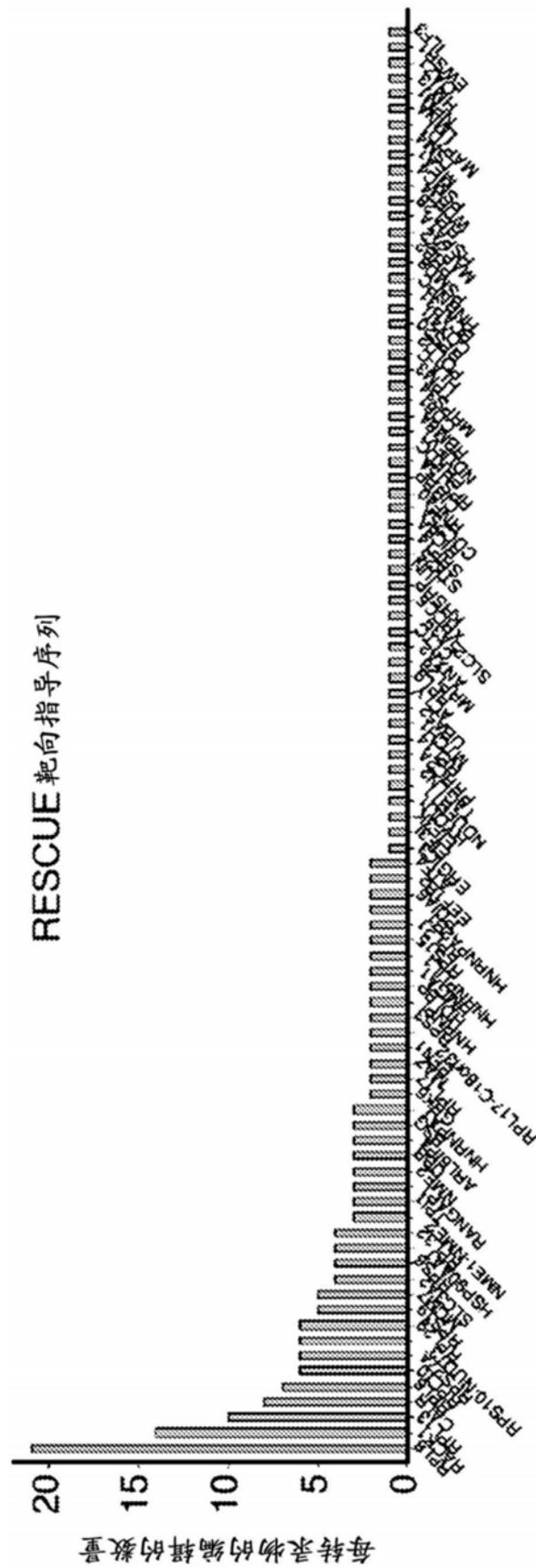


图134C

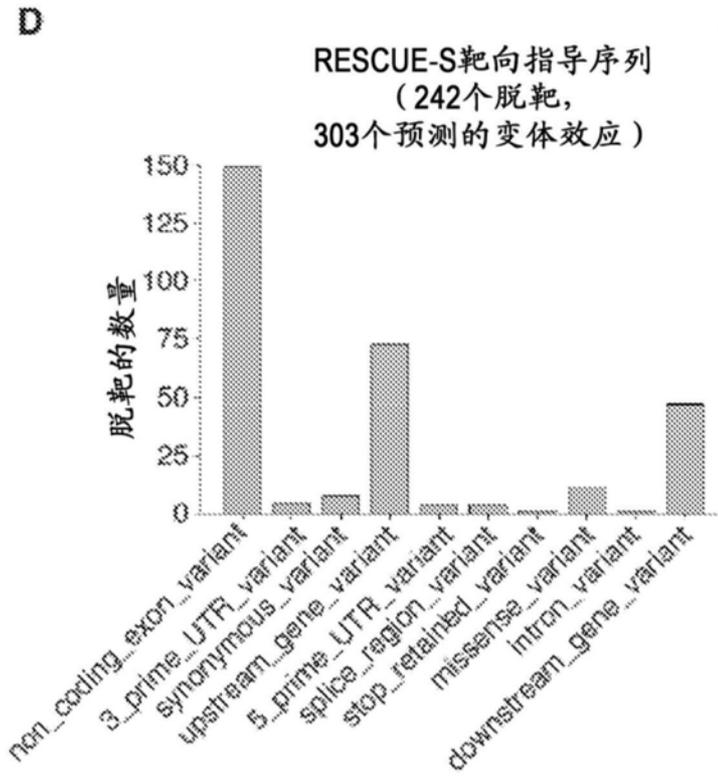


图134D

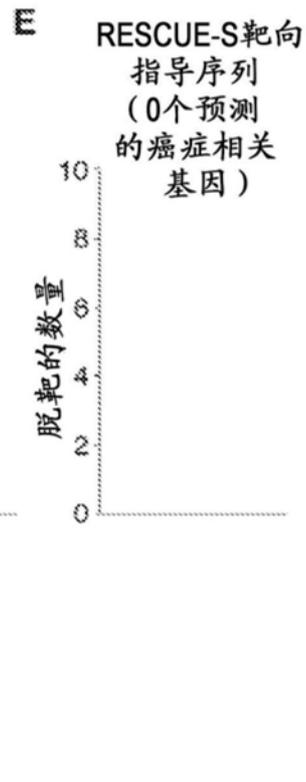


图134E

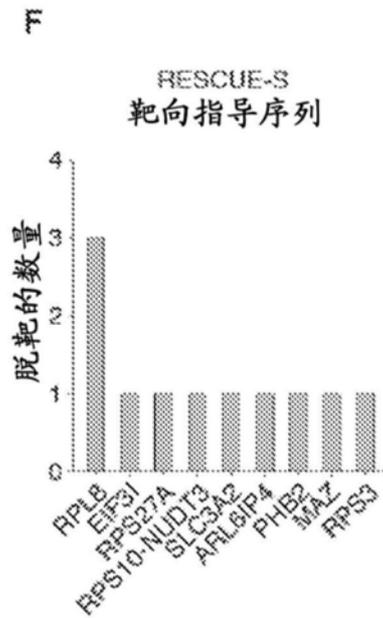


图134F

A

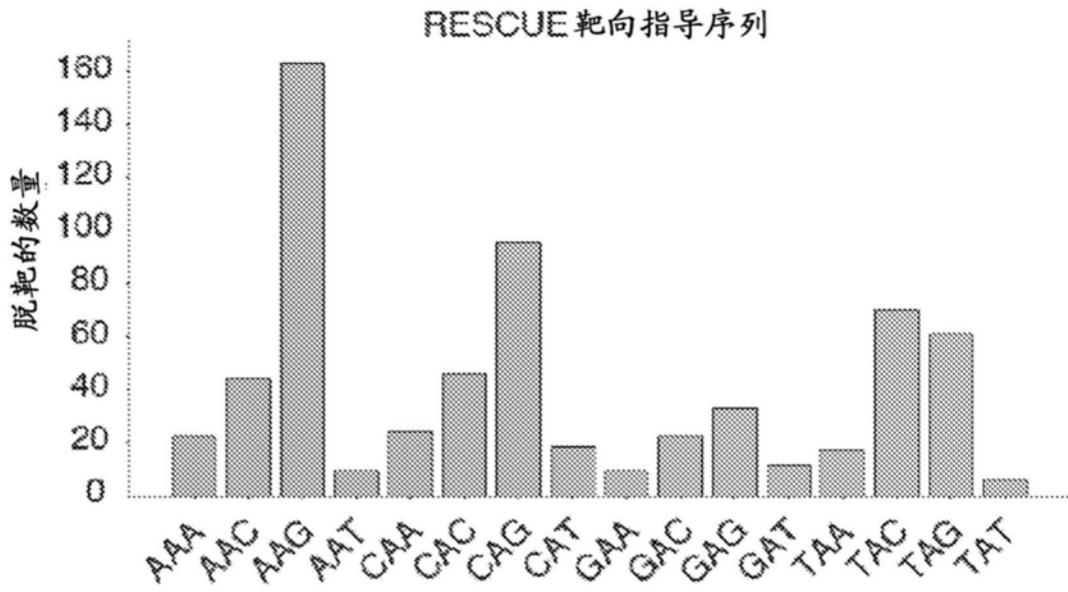


图135A

B

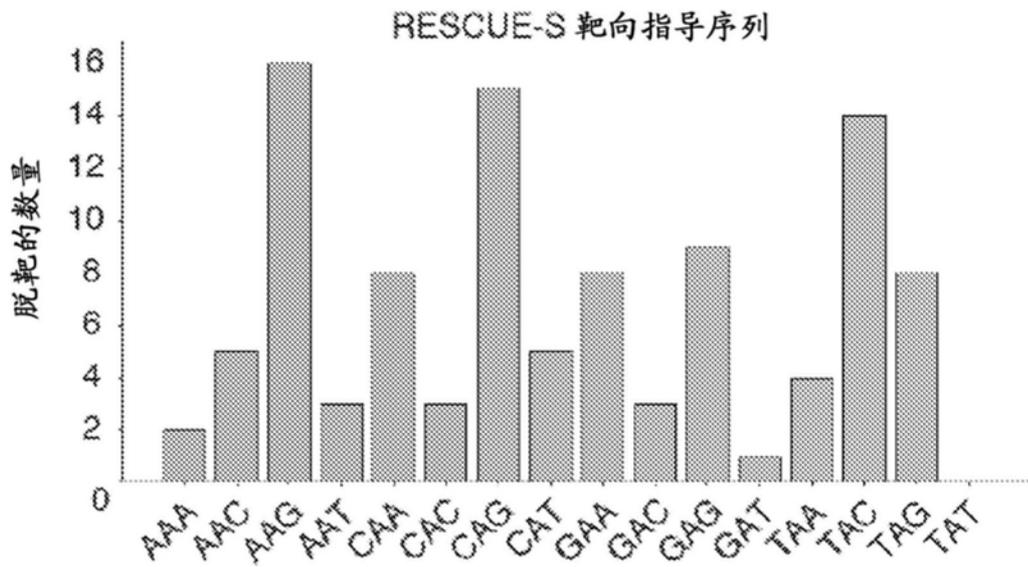


图135B

C

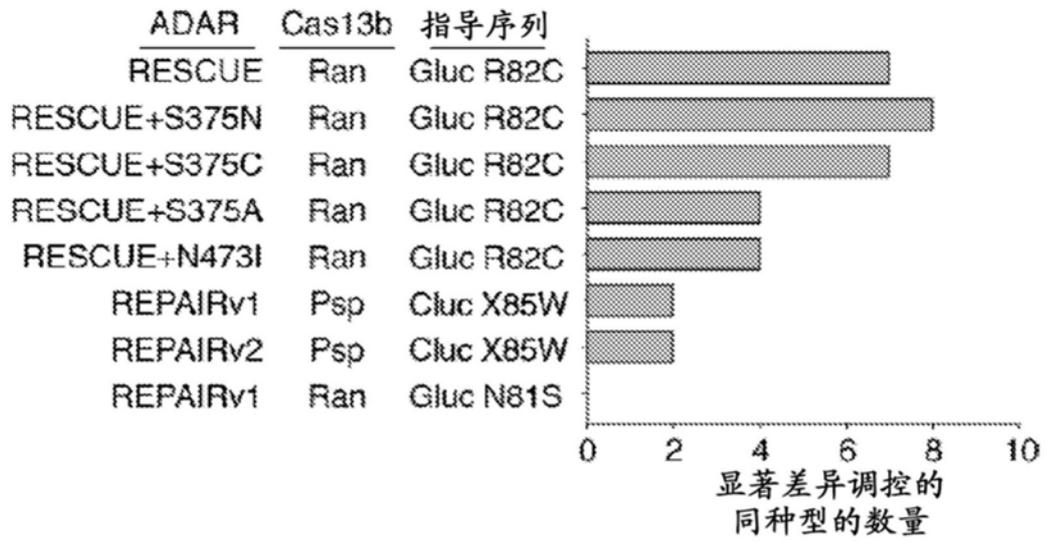


图135C

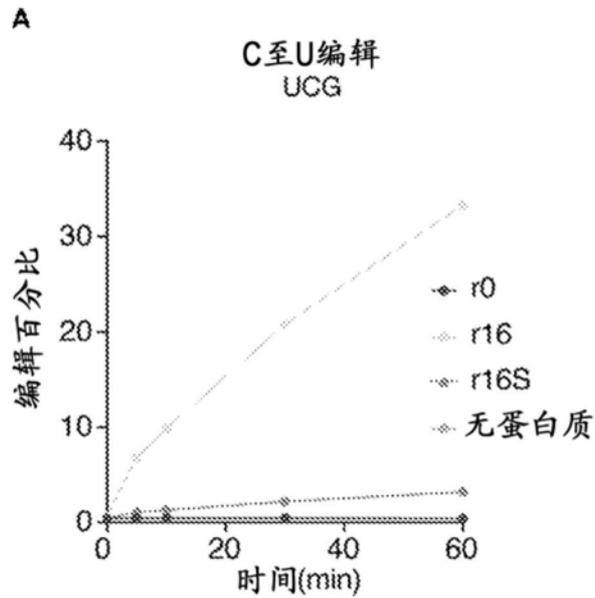


图136A

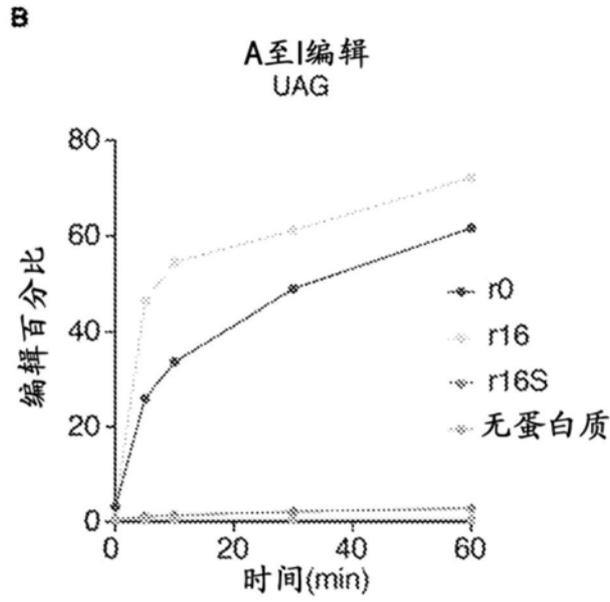


图136B

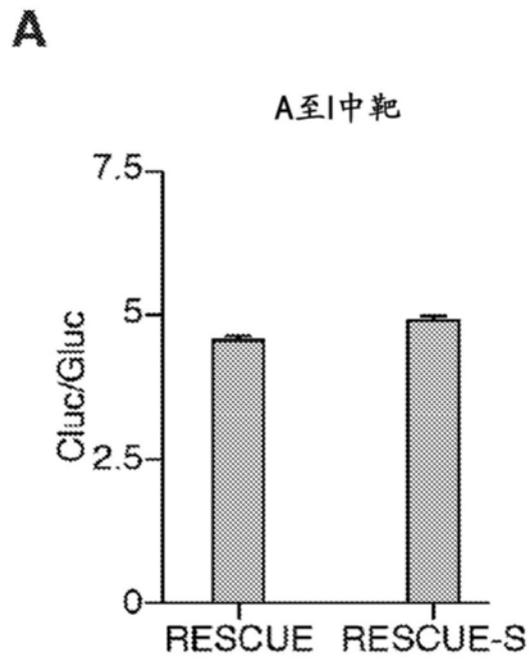


图137A

**B**

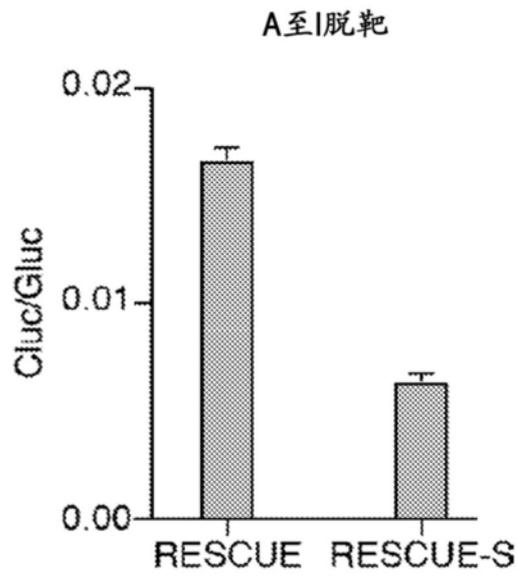


图137B

**C**

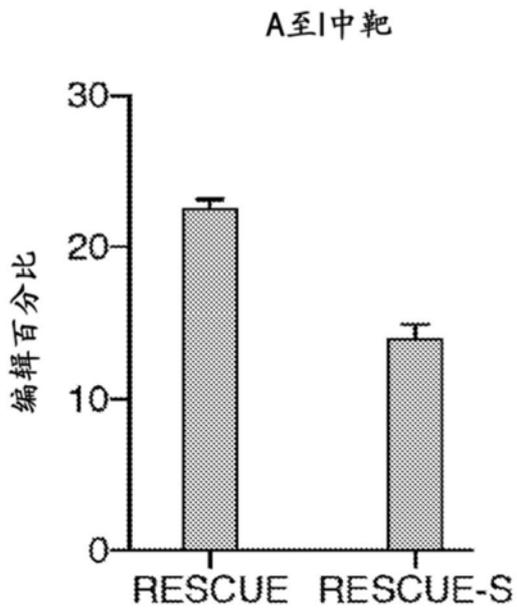


图137C

**D**

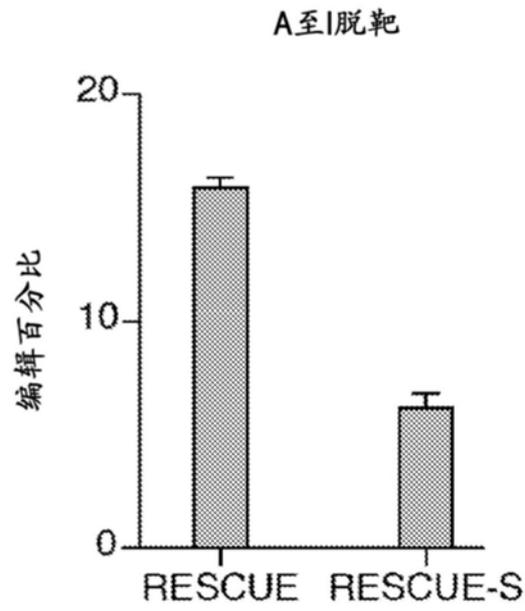


图137D

**A**

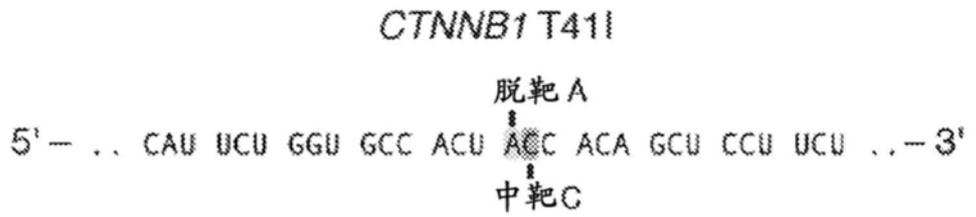


图138A

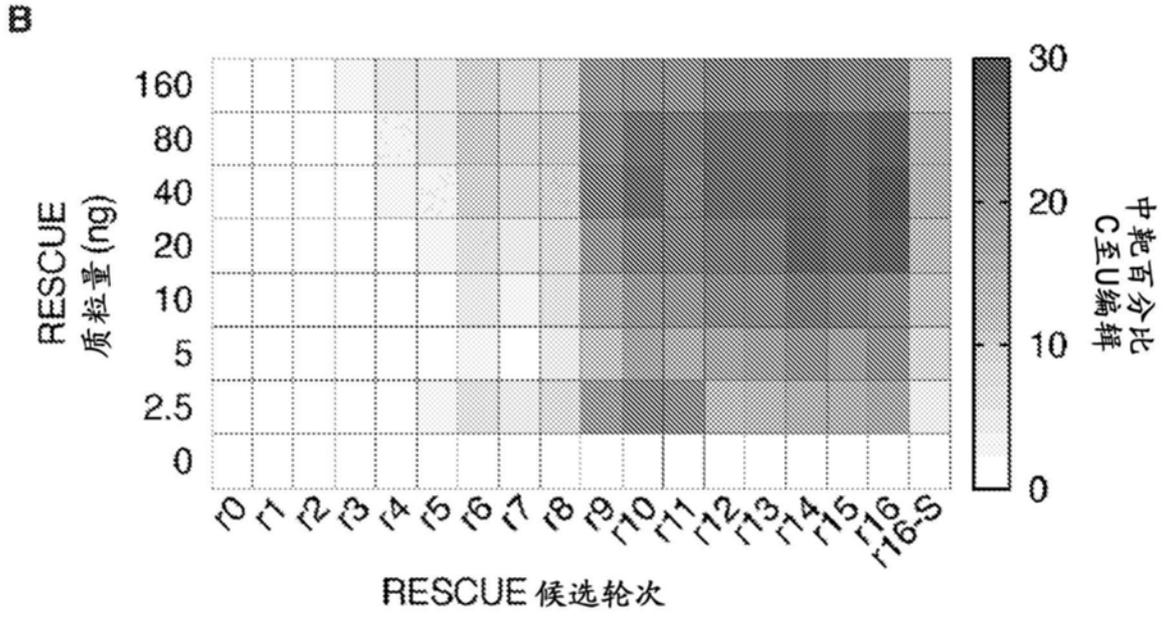


图138B

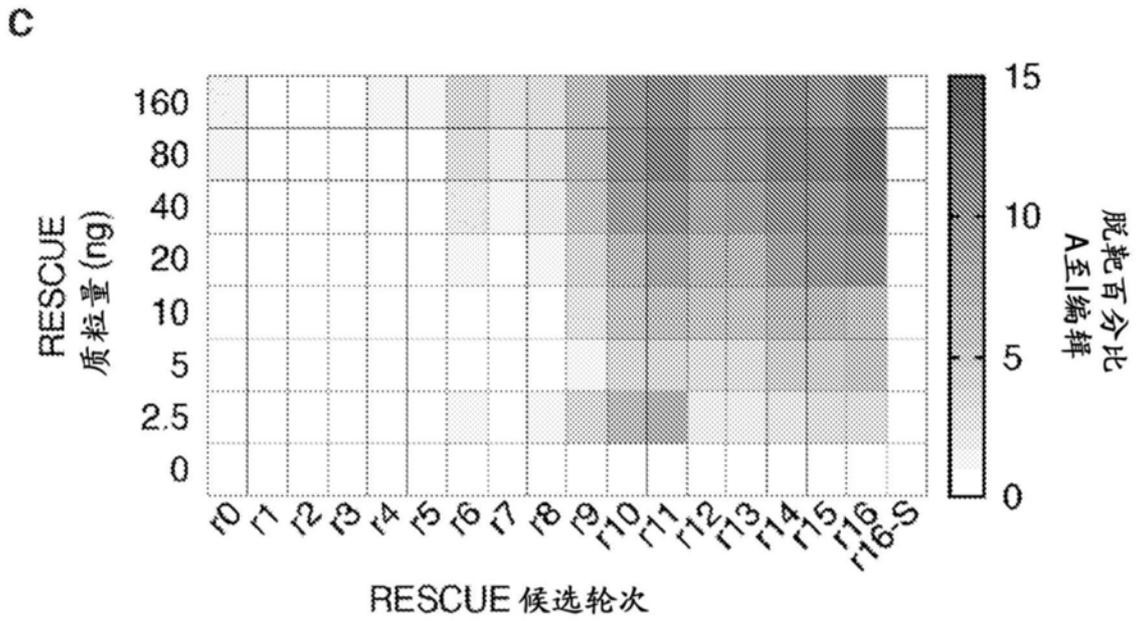


图138C

**A**

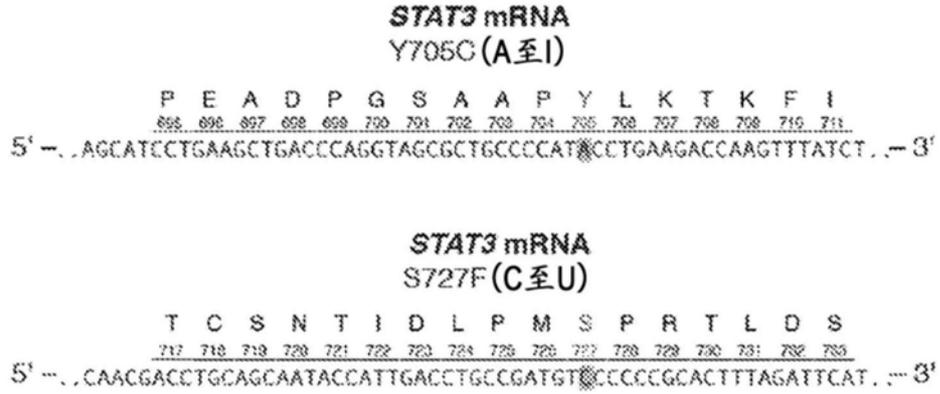


图139A

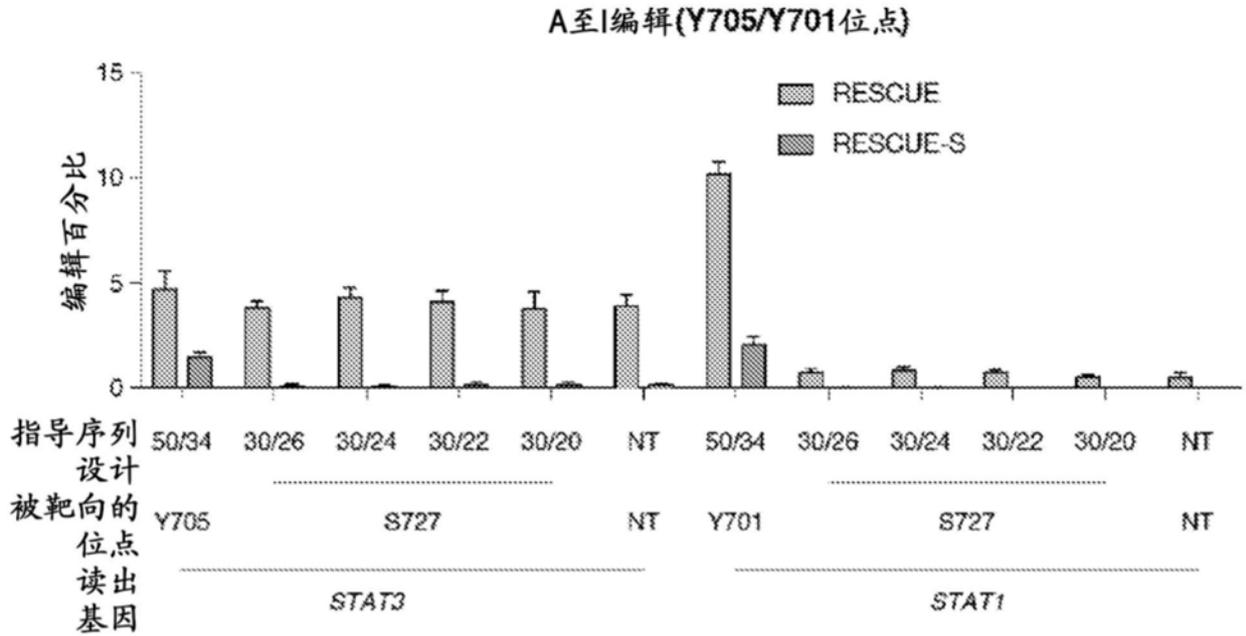


图139B

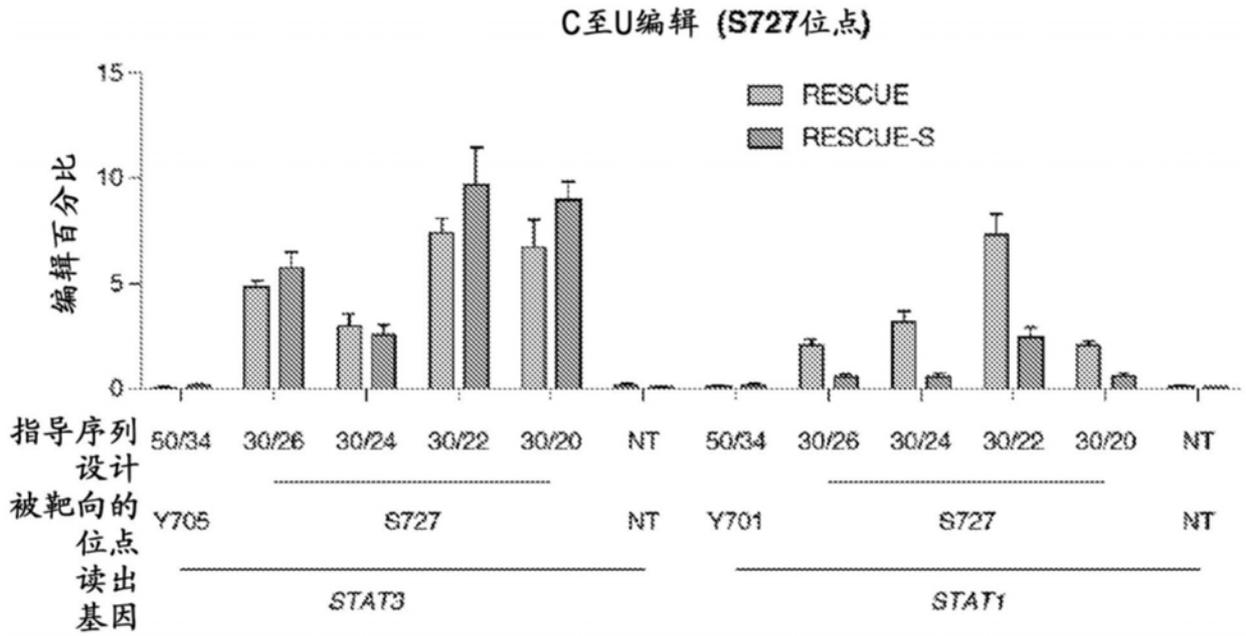


图139C

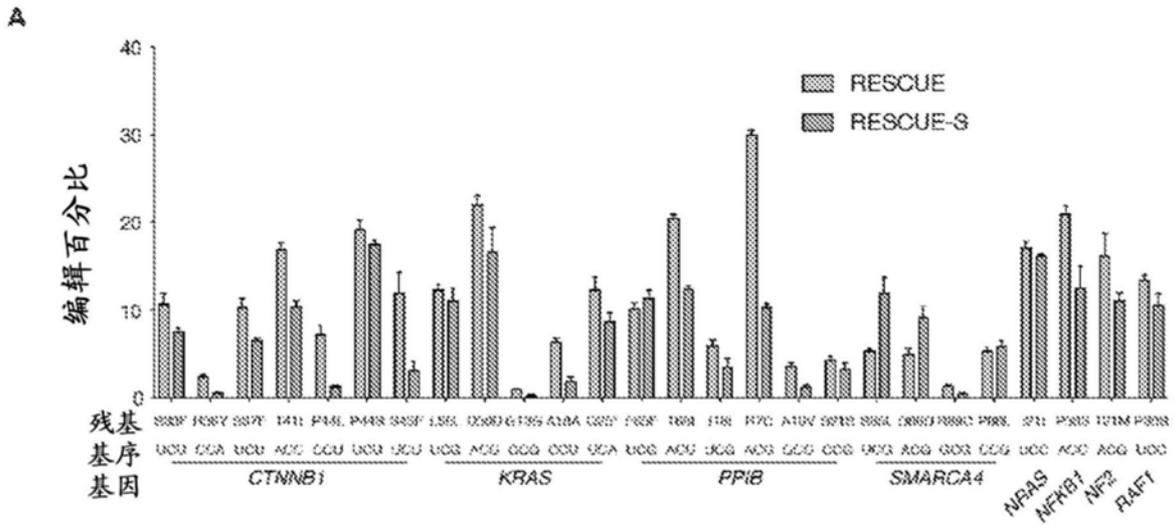


图140A

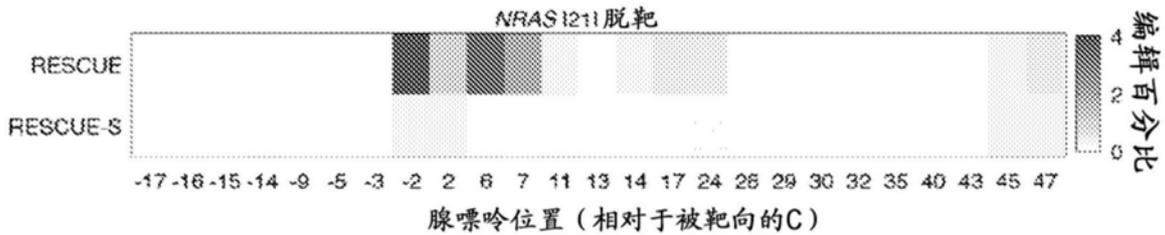


图140B

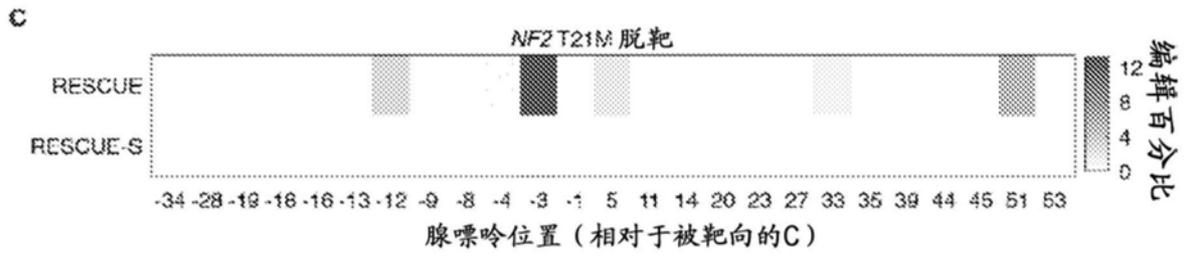


图140C



图140D

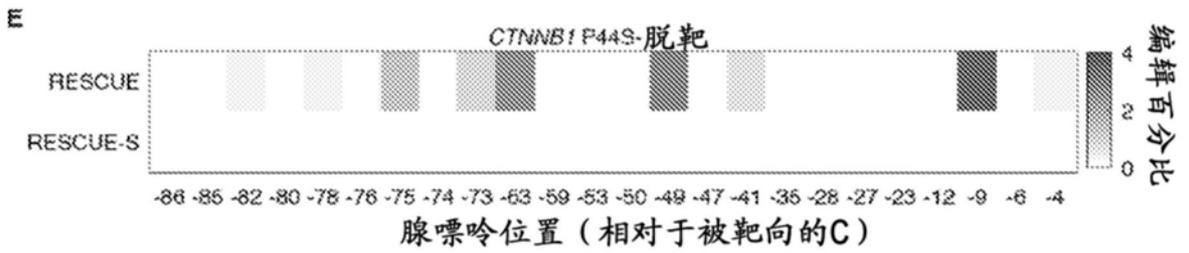


图140E

|     |   | 第二位                              |                              |  |                                       |   |   |
|-----|---|----------------------------------|------------------------------|--|---------------------------------------|---|---|
|     |   | U                                | C                            | A                                      | G                                     |   |   |
| 第一位 | U | UUU Phe<br>UUC<br>UUA Leu<br>UUG | UCU Ser<br>UCC<br>UCA<br>UCG | UAU Tyr<br>UAC<br>UAA stop<br>UAG stop | UGU Cys<br>UGC<br>UGA stop<br>UGG Trp | U | ↻ |
|     | C | CUU Leu<br>CUC<br>CUA<br>CUG     | CCU Pro<br>CCC<br>CCA<br>CCG | CAU His<br>CAC<br>CAA Gln<br>CAG       | CGU Arg<br>CGC<br>CGA<br>CGG          | U | ↻ |
|     | A | AUU Ile<br>AUC<br>AUA<br>AUG Met | ACU Thr<br>ACC<br>ACA<br>ACG | AAU Asn<br>AAC<br>AAA Lys<br>AAG       | AGU Ser<br>AGC<br>AGA Arg<br>AGG      | U | ↻ |
|     | G | GUU Val<br>GUC<br>GUA<br>GUG     | GCU Ala<br>GCC<br>GCA<br>GCG | GAU Asp<br>GAC<br>GAA Glu<br>GAG       | GGU Gly<br>GGC<br>GGA<br>GGG          | U | ↻ |

图141