(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109804066 A
(43)申请公布日 2019.05.24

(21)申请号 201780062152.6

(22)申请日 2017.08.09

(30)优先权数据
62/372,755 2016.08.09 US
62/456,048 2017.02.07 US

(85)PCT国际申请进入国家阶段日
2019.04.08

(86)PCT国际申请的申请数据
PCT/US2017/046144 2017.08.09

(87)PCT国际申请的公布数据
WO2018/031683 EN 2018.02.15

(71)申请人 哈佛大学的校长及成员们
地址 美国马萨诸塞州

(72)发明人 D.R.刘 B.柴金德 J.L.贝森

(74)专利代理机构 北京市柳沈律师事务所 11105
代理人 张文辉

(51)Int.Cl.
C12N 9/22(2006.01)
C12N 15/11(2006.01)
A61K 48/00(2006.01)

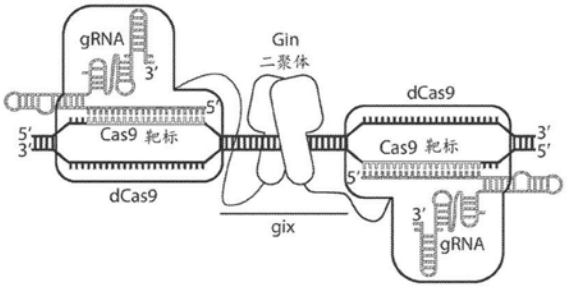
权利要求书7页 说明书128页
序列表285页 附图20页

(54)发明名称

可编程CAS9-重组酶融合蛋白及其用途

(57)摘要

本公开的一些方面提供了融合蛋白，其包含引导核苷酸序列-可编程DNA结合蛋白域(例如，Cas9的核酸酶无活性变体，例如dCas9)、任选的接头和重组酶催化域(例如，酪氨酸重组酶催化域或丝氨酸重组酶催化域，例如Gin重组酶催化域)。该融合蛋白可以重组含有侧翼为引导RNA指定序列的最小重组酶核心位点的DNA位点。本公开代表朝向不依赖于内源性细胞机制或细胞状态的在未修饰的细胞中的可编程无痕基因组编辑的步骤。



CN 109804066 A

1.融合蛋白,其包含:

(i)引导核苷酸序列-可编程DNA结合蛋白域;

(ii)接头;和

(iii)重组酶催化域。

2.权利要求1的融合蛋白,其中所述引导核苷酸序列-可编程DNA结合蛋白域选自下组:核酸酶无活性的Cas9(dCas9)域、核酸酶无活性的Cpf1域、核酸酶无活性的Argonaute域,及其变体。

3.权利要求1的融合蛋白,其中所述引导核苷酸序列-可编程DNA结合蛋白域是核酸酶无活性的Cas9(dCas9)域。

4.权利要求2或3的融合蛋白,其中所述dCas9域的氨基酸序列包含对应于SEQ ID NO:1中D10A或H840A突变的突变。

5.权利要求2-4中任一项的融合蛋白,其中所述dCas9域的氨基酸序列包含对应于SEQ ID NO:1中D10A突变的突变和对应于SEQ ID NO:1中H840A突变的突变。

6.权利要求2-5中任一项的融合蛋白,其中所述dCas9域的氨基酸序列进一步包含对应于SEQ ID NO:1中缺少的N-末端甲硫氨酸的突变。

7.权利要求2或3的融合蛋白,其中所述dCas9域的氨基酸序列包含SEQ ID NO:712。

8.权利要求2或3的融合蛋白,其中所述dCas9域的氨基酸序列与SEQ ID NO:712具有95%或更高的序列一致性。

9.权利要求2或3的融合蛋白,其中所述dCas9域的氨基酸序列与SEQ ID NO:712具有96%、97%、98%、99%或更高的序列一致性。

10.前述权利要求中任一项的融合蛋白,其中所述重组酶催化域是丝氨酸重组酶催化域或酪氨酸重组酶催化域。

11.权利要求10的融合蛋白,其中所述丝氨酸重组酶催化域或酪氨酸重组酶催化域选自Gin、Sin、Tn3、Hin、β、γδ、PhiC31、Cre或FLP重组酶催化域。

12.权利要求10或权利要求11的融合蛋白,其中所述Gin重组酶催化域的氨基酸序列包含选自下组的一个或多个突变:SEQ ID NO:713中的H106Y、I127L、I136R或G137F突变。

13.权利要求11或权利要求12的融合蛋白,其中所述Gin重组酶催化域的氨基酸序列包含SEQ ID NO:713。

14.权利要求10或权利要求11的融合蛋白,其中所述Cre重组酶的氨基酸序列是截短的。

15.权利要求11或权利要求14的融合蛋白,其中所述酪氨酸重组酶催化域是所述Cre重组酶的25kDa羧基-末端域。

16.权利要求14的融合蛋白,其中所述Cre重组酶以氨基酸R118、A127、E138或R154开始。

17.权利要求1-16中任一项的融合蛋白,其中所述重组酶的氨基酸序列已得以进一步突变。

18.前述权利要求中任一项的融合蛋白,其中所述重组酶催化域是经演化的重组酶催化域。

19.权利要求1-18中任一项的融合蛋白,其中所述接头具有约33埃至约81埃的长度。

20.权利要求1-19中任一项的融合蛋白,其中所述接头是肽接头。

21.权利要求20的融合蛋白,其中所述肽接头包含XTEN接头(SGSETPGTSESATPES(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:9))、包含三肽GGS的一个或多个重复的氨基酸序列,或任何以下氨基酸序列:VPFLLEPDNINGKTC(SEQ ID NO:10)、GSAGSAAGSGEF(SEQ ID NO:11)、SIVAQLSRPDPA(SEQ ID NO:12)、MKIIEQLPSA(SEQ ID NO:13)、VRHKLKRVGS(SEQ ID NO:14)、GHGTGSTGSGSS(SEQ ID NO:15)、MSRPDPA(SEQ ID NO:16)或GGSM(SEQ ID NO:17)。

22.权利要求20或权利要求21的融合蛋白,其中所述肽接头包含所述三肽GGS的一个或多个重复。

23.权利要求20-22中任一项的融合蛋白,其中所述肽接头包含所述三肽GGS的六至十个重复。

24.权利要求20-23中任一项的融合蛋白,其中所述肽接头包含所述三肽GGS的八个重复。

25.权利要求20-24中任一项的融合蛋白,其中所述肽接头长约18至约27个氨基酸。

26.权利要求20-25中任一项的融合蛋白,其中所述肽接头长24个氨基酸。

27.权利要求20-26中任一项的融合蛋白,其中所述肽接头具有氨基酸序列GGSGGSGGSGGSGGSGGSGGSGGS(SEQ ID NO:183)。

28.权利要求1-19中任一项的融合蛋白,其中所述接头是非肽接头。

29.权利要求28的融合蛋白,其中所述非肽接头包含聚乙二醇(PEG)、聚丙二醇(PPG)、共聚(乙烯/丙烯)二醇、聚氧乙烯(POE)、聚氨酯、聚磷腈、多糖、右旋糖酐、聚乙烯醇、聚乙烯吡咯烷酮、聚乙烯乙醚、聚丙烯酰胺、聚丙烯酸酯、聚氰基丙烯酸酯、脂质聚合物、甲壳质、透明质酸、肝素或烷基接头。

30.权利要求29的融合蛋白,其中所述烷基接头具有式—NH—(CH2)s—C(O)—,其中s可以是1-100的任何整数,包括端点。

31.权利要求30的融合蛋白,其中s是1-20的整数,包括端点。

32.权利要求1-31中任一项的融合蛋白,其进一步包含核定位信号(NLS)域。

33.权利要求32的融合蛋白,其中所述NLS域经由一个或多个第二接头与所述引导核苷酸序列-可编程DNA结合蛋白域或所述Gin重组酶催化域结合。

34.权利要求33的融合蛋白,其中所述第二接头是肽接头。

35.权利要求33或权利要求34的融合蛋白,其中所述第二接头包含XTEN接头(SGSETPGTSESATPES(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:9))、包含一个或多个三肽GGS的重复的氨基酸序列,或任何以下氨基酸序列:VPFLLEPDNINGKTC(SEQ ID NO:10)、GSAGSAAGSGEF(SEQ ID NO:11)、SIVAQLSRPDPA(SEQ ID NO:12)、MKIIEQLPSA(SEQ ID NO:13)、VRHKLKRVGS(SEQ ID NO:14)、GHGTGSTGSGSS(SEQ ID NO:15)、MSRPDPA(SEQ ID NO:16)或GGSM(SEQ ID NO:17)。

36.权利要求33-35中任一项的融合蛋白,其中所述第二接头包含所述三肽GGS的一个或多个重复。

37.权利要求33-36中任一项的融合蛋白,其中所述第二接头包含所述三肽GGS的一至五个重复。

38.权利要求33-37中任一项的融合蛋白,其中所述第二接头包含所述三肽GGS的一个重复。

39.权利要求33-38中任一项的融合蛋白,其中所述第二接头具有序列GGS。

40.权利要求33的融合蛋白,其中所述第二接头是非肽接头。

41.权利要求40的融合蛋白,其中所述非肽接头包含聚乙二醇(PEG)、聚丙二醇(PPG)、共聚(乙烯/丙烯)二醇、聚氧乙烯(POE)、聚氨酯、聚磷腈、多糖、右旋糖酐、聚乙烯醇、聚乙烯吡咯烷酮、聚乙烯乙醚、聚丙烯酰胺、聚丙烯酸酯、聚氰基丙烯酸酯、脂质聚合物、甲壳质、透明质酸、肝素或烷基接头。

42.权利要求41的融合蛋白,其中所述烷基接头具有式—NH—(CH2)s—C(O)—,其中s可以是1-100的任何整数,包括端点。

43.权利要求42的融合蛋白,其中s是1-20的任何整数,包括端点。

44.权利要求32-43中任一项的融合蛋白,其中所述融合蛋白包含结构NH_2-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-COOH。

45.权利要求32-44中任一项的融合蛋白,其中所述融合蛋白包含结构NH_2-[Gin重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-COOH。

46.权利要求32-39、44或45中任一项的融合蛋白,其中所述融合蛋白包含SEQ ID NO:719中所示的氨基酸序列。

47.前述权利要求中任一项的融合蛋白,其进一步包含一个或多个亲和标签。

48.权利要求47的融合蛋白,其中所述亲和标签选自下组:FLAG标签、多组氨酸(多His)标签、多精氨酸(多Arg)标签、Myc标签和HA标签。

49.权利要求47或权利要求48的融合蛋白,其中所述亲和标签是FLAG标签。

50.权利要求47-49中任一项的融合蛋白,其中所述FLAG标签具有序列PKKKRKV(SEQ ID NO:702)。

51.权利要求47-50中任一项的融合蛋白,其中所述一个或多个亲和标签经由一个或多个第三接头与所述引导核苷酸序列-可编程DNA结合蛋白域、所述重组酶催化域或所述NLS域结合。

52.权利要求51的融合蛋白,其中所述第三接头是肽接头。

53.权利要求51或权利要求52的融合蛋白,其中所述第三接头包含XTEN接头,(SGSETPGTSESATPES(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:9))、包含三肽GGS的一个或多个重复的氨基酸序列,或任何以下氨基酸序列:VPFLLEPDNINGKTC(SEQ ID NO:10)、GSAGSAAGSGEF(SEQ ID NO:11)、SIVAQLSRPDPA(SEQ ID NO:12)、MKIIEQLPSA(SEQ ID NO:13)、VRHKLKRVGS(SEQ ID NO:14)、GHGTGSTGSGSS(SEQ ID NO:15)、MSRPDPA(SEQ ID NO:16)或GGSM(SEQ ID NO:17)。

54.权利要求51-53中任一项的融合蛋白,其中所述第三接头包含所述三肽GGS的一个或多个重复。

55.权利要求51-54中任一项的融合蛋白,其中所述第三接头包含所述三肽GGS的一至五个重复。

56.权利要求51-55中任一项的融合蛋白，其中所述第三接头包含所述三肽GGS的一个重复。

57.权利要求51-56中任一项的融合蛋白，其中所述第三接头具有序列GGS。

58.权利要求51的融合蛋白，其中所述第三接头是非肽接头。

59.权利要求58的融合蛋白，其中所述非肽接头包含聚乙二醇(PEG)、聚丙二醇(PPG)、共聚(乙烯/丙烯)二醇、聚氧乙烯(POE)、聚氨酯、聚磷腈、多糖、右旋糖酐、聚乙烯醇、聚乙烯吡咯烷酮、聚乙烯乙醚、聚丙烯酰胺、聚丙烯酸酯、聚氰基丙烯酸酯、脂质聚合物、甲壳质、透明质酸、肝素或烷基接头。

60.权利要求59的融合蛋白，其中所述烷基接头具有式—NH—(CH2)s—C(O)—，其中s可以是1-100的任何整数，包括端点。

61.权利要求60的融合蛋白，其中s可以是1-20的任何整数。

62.权利要求47-61中任一项的融合蛋白，其中所述融合蛋白包含结构NH_2-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[亲和标签]-COOH、NH_2-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH、NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH、NH_2-[亲和标签]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-COOH、NH_2-[亲和标签]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-COOH或NH_2-[亲和标签]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-COOH。

63.权利要求47-57中任一项的融合蛋白，其中所述融合蛋白与SEQ ID NO:185中所示的氨基酸序列具有大于99％的序列一致性。

64.权利要求47-57中任一项的融合蛋白，其中所述融合蛋白与SEQ ID NO:185中所示的氨基酸序列具有大于90％或95％的序列一致性。

65.权利要求47-57、63或64中任一项的融合蛋白，其中所述融合蛋白具有SEQ ID NO:185中所示的氨基酸序列。

66.前述权利要求中任一项的融合蛋白，其中所述引导核苷酸序列-可编程DNA结合蛋白域与引导RNA(gRNA)结合。

67.权利要求66的融合蛋白的二聚体。

68.权利要求67的融合蛋白的二聚体，其中所述二聚体与DNA分子结合。

69.权利要求68的融合蛋白的二聚体，其中所述二聚体的每个融合蛋白与所述DNA分子的相同链结合。

70.权利要求68的融合蛋白的二聚体，其中所述二聚体的每个融合蛋白与所述DNA分子的相对链结合。

71.权利要求67-70中任一项的融合蛋白的二聚体,其中所述二聚体的gRNA与重组酶位点旁侧的gRNA结合位点杂交。

72.权利要求71的方法,其中所述重组酶位点包含res、gix、hix、six、resH、LoxP、FTR或att核心或相关核心序列。

73.权利要求71或权利要求72的融合蛋白的二聚体,其中所述重组酶位点包含gix核心或gix相关核心序列。

74.权利要求72或权利要求73的二聚体,其中所述gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为3至7个碱基对。

75.权利要求72-74中任一项的二聚体,其中所述gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为5至6个碱基对。

76.权利要求67-75中任一项的融合蛋白的二聚体,其中第一二聚体与第二二聚体结合,从而形成所述融合蛋白的四聚体。

77.权利要求66的融合蛋白的四聚体。

78.权利要求77的融合蛋白的四聚体,其中所述四聚体与DNA分子结合。

79.权利要求68的融合物的四聚体,其中每个二聚体与DNA的相对链结合。

80.权利要求68的融合蛋白的四聚体,其中每个二聚体与DNA的相同链结合。

81.用于两个DNA分子之间的位点特异性重组的方法,其包括:

(a)使第一DNA与第一融合蛋白接触,其中所述引导核苷酸序列-可编程DNA结合蛋白域结合第一gRNA,所述第一gRNA与所述第一DNA的第一区域杂交;

(b)使所述第一DNA与第二融合蛋白接触,其中所述第二融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第二gRNA,所述第二gRNA与所述第一DNA的第二区域杂交;

(c)使第二DNA与第三融合蛋白接触,其中所述第三融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第三gRNA,所述第三gRNA与所述第二DNA的第一区域杂交;和

(d)使所述第二DNA与第四融合蛋白接触,其中所述第四融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第四gRNA,所述第四gRNA与所述第二DNA的第二区域杂交;

其中在使得所述DNA重组的条件下,步骤(a)-(d)中所述融合蛋白的结合导致所述融合蛋白的重组酶催化域的四聚化,并且其中所述第一、第二、第三和/或第四融合蛋白是权利要求1-63中任一项的融合蛋白。

82.权利要求81的方法,其中所述第一和第二DNA分子具有不同的序列。

83.权利要求81的方法,其中步骤(a)和(b)的所述gRNA与所述第一DNA的相对链杂交,并且步骤(c)和(d)的所述gRNA与所述第二DNA的相对链杂交。

84.权利要求81-83中任一项的方法,其中步骤(a)和(b)的所述gRNA;和/或步骤(c)和(d)的所述gRNA与它们各自的DNA的区域杂交,所述区域相隔不超过100个碱基对。

85.权利要求84的方法,其中步骤(a)和(b)的所述gRNA;和/或步骤(c)和(d)的所述gRNA与它们各自的DNA的区域杂交,所述区域相隔不超过10、不超过15、不超过20、不超过25、不超过30、不超过40、不超过50、不超过60、不超过70、不超过80或不超过90个碱基对。

86.权利要求81-85中任一项的方法,其中步骤(a)和(b)的所述gRNA;和/或步骤(c)和(d)的所述gRNA在重组酶位点旁侧的gRNA结合位点处与它们各自的DNA的区域杂交。

87.权利要求86的方法,其中所述重组酶位点包含res、gix、hix、six、resH、LoxP、FTR或

att核心或相关核心序列。

88.权利要求86或权利要求87的方法，其中所述重组酶位点包含gix核心或gix相关核心序列。

89.权利要求87或权利要求88的方法，其中所述gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为3至7个碱基对。

90.权利要求87-89中任一项的方法，其中所述gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为5至6个碱基对。

91.用于单一DNA分子的两个区域之间的位点特异性重组的方法，其包括：

(a)使所述DNA与第一融合蛋白接触，其中所述引导核苷酸序列-可编程DNA结合蛋白域结合第一gRNA，所述第一gRNA与所述DNA的第一区域杂交；

(b)使所述DNA与第二融合蛋白接触，其中所述第二融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第二gRNA，所述第二gRNA与所述DNA的第二区域杂交；

(c)使所述DNA与第三融合蛋白接触，其中所述第三融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第三gRNA，所述第三gRNA与所述DNA的第三区域杂交；和

(d)使所述DNA与第四融合蛋白接触，其中所述第四融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第四gRNA，所述第四gRNA与所述DNA的第四区域杂交；

其中在使得所述DNA重组的条件下，步骤(a)-(d)中所述融合蛋白的结合导致所述融合蛋白的重组酶催化域的四聚化，并且其中所述第一、第二、第三和/或第四融合蛋白是权利要求1-63中任一项的融合蛋白。

92.权利要求91的方法，其中重组的所述单一DNA分子的两个区域具有不同的序列。

93.权利要求91或权利要求92的方法，其中所述重组导致所述DNA分子的区域的缺失。

94.权利要求93的方法，其中缺失的所述DNA分子的区域易于在减数分裂中发生交换事件。

95.权利要求91-94中任一项的方法，其中步骤(a)-(d)的所述第一和第二gRNA与所述DNA的相同链杂交，并且步骤(a)-(d)的所述第三和第四gRNA与所述DNA的相对链杂交。

96.权利要求91-95中任一项的方法，其中步骤(a)和(b)的所述gRNA与所述DNA的区域杂交，所述区域相隔不超过100个碱基对，并且步骤(c)和(d)的所述gRNA与所述DNA的区域杂交，所述区域相隔不超过100个碱基对。

97.权利要求96的方法，其中步骤(a)和(b)的所述gRNA与所述DNA的区域杂交，所述区域相隔不超过50、不超过60、不超过70、不超过80或不超过90个碱基对，并且步骤(c)和(d)的所述gRNA与所述DNA的区域杂交，所述区域相隔不超过10、不超过15、不超过20、不超过25、不超过30、不超过40、不超过50、不超过60、不超过70、不超过80或不超过90个碱基对。

98.权利要求91-97中任一项的方法，其中步骤(a)和(b)的所述gRNA；和/或步骤(c)和(d)的所述gRNA与重组酶位点旁侧的gRNA结合位点杂交。

99.权利要求98的方法，其中所述重组酶位点包含res、gix、hix、six、resH、LoxP、FTR或att核心或相关核心序列。

100.权利要求98或权利要求99的方法，其中所述重组酶位点包含gix核心或gix相关核心序列。

101.权利要求99或权利要求100的方法，其中所述gix核心或gix相关核心序列与至少

一个gRNA结合位点之间的距离为3至7个碱基对。

102.权利要求99-101中任一项的方法，其中所述gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为5至6个碱基对。

103.权利要求81-102中任一项的方法，其中所述DNA在细胞中。

104.权利要求103的方法，其中所述细胞是真核细胞。

105.权利要求103的方法，其中所述细胞是原核细胞。

106.权利要求104或权利要求105的方法，其中所述细胞在受试者中。

107.权利要求106的方法，其中所述受试者是人。

108.多核苷酸，其编码权利要求1-63中任一项的融合蛋白。

109.载体，其包含权利要求108的多核苷酸。

110.用于重组蛋白表达的载体，其包含编码权利要求1-63中任一项的融合蛋白的多核苷酸。

111.细胞，其包含用于表达权利要求1-63中任一项的融合蛋白的遗传构建体。

112.试剂盒，其包含权利要求1-63中任一项的融合蛋白。

113.试剂盒，其包含编码权利要求1-63中任一项的融合蛋白的多核苷酸。

114.试剂盒，其包含用于重组蛋白表达的载体，其中所述载体包含编码权利要求1-63中任一项的融合蛋白的多核苷酸。

115.试剂盒，其包含含有用于表达权利要求1-63中任一项的融合蛋白的遗传构建体的细胞。

116.权利要求112-115中任一项的试剂盒，其进一步包含一个或多个gRNA和/或用于表达一个或多个gRNA的载体。

可编程CAS9-重组酶融合蛋白及其用途

[0001] 相关申请

[0002] 根据35U.S.C.§119(e),本申请要求2016年8月9日提交的美国临时专利申请U.S.S.N.62/372,755和2017年2月7日提交的美国临时专利申请U.S.S.N.62/456,048的优先权,其每一个通过引用并入本文。

[0003] 政府资助

[0004] 本发明是在国立卫生研究院(National Institutes of Health)授予的拨款号R01EB022376和R35GM118062下得到政府支持完成的。政府拥有本发明的某些权利。

[0005] 发明背景

[0006] 高效、可编程且位点特异性同源重组仍然是遗传学和基因组编辑的长期目标。将重组导向感兴趣的基因座的早期尝试依赖于用与靶基因座同源的长侧翼序列转染供体DNA。这种策略受到非常低的效率阻碍,并因此需要严格的选择来识别整合体。最近的努力利用了双链DNA断裂(DSB)诱导同源性定向修复(HDR)的能力。归巢内切核酸酶和后来的可编程内切核酸酶如锌指核酸酶、TALE核酸酶、Cas9和fCas9已用于引入靶定的DSB并在供体DNA存在下诱导HDR。然而,在大多数有丝分裂后细胞中,DSB诱导的HDR强烈下调并且通常效率低。此外,通过易错修复途径(例如非同源末端连接(NHEJ)或单链退火(SSA))修复DSB以比HDR更高的频率导致DSB位点处的核苷酸的随机插入或缺失(插入/缺失)。若细胞经受迫使细胞周期同步的条件或者若NHEJ中涉及的酶得以抑制,则可以提高HDR的效率。但是,此类情况可导致许多随机且不可预测的事件,从而限制了潜在的应用。本公开提供了融合蛋白,其可以重组含有侧翼为引导RNA指定序列的最小重组酶核心位点的DNA位点,并且代表朝向不依赖于内源性细胞机制或细胞状态的在未修饰的细胞中的可编程无痕(scarless)基因组编辑的步骤。

[0007] 发明概述

[0008] 本公开描述了融合蛋白的开发,所述融合蛋白包含引导核苷酸序列-可编程DNA结合蛋白域、任选的接头和重组酶催化域(例如,丝氨酸重组酶催化域如Gin重组酶催化域、酪氨酸重组酶催化域,或任何演化的重组酶催化域)。该融合蛋白在侧翼为两个引导RNA指定的DNA序列的最小gix核心重组酶位点(NNNNAAASSWWSSTTTNNNN,SEQ ID NO:19)上起作用。由所述融合蛋白介导的重组依赖于这两个引导RNA,导致不同引导核苷酸:融合蛋白复合物之间的正交性,并且在培养的人细胞中在与人基因组中发现的DNA序列匹配的DNA序列上有效地发挥功能。本公开的融合蛋白还可以直接在人细胞(例如,培养的人细胞)的基因组上起作用,催化位于相隔约14千碱基之间的两个recCas9假位点(psuedosite)之间的缺失、插入、倒位、易位或重组。这项工作提供了工程化的酶,其可以在未修饰的基因组中以用户定义的单碱基对分辨率催化基因插入、缺失、倒位或染色体易位。

[0009] 在一个方面,本发明提供融合蛋白,其包含:(i)引导核苷酸序列-可编程DNA结合蛋白域;(ii)任选的接头;和(iii)重组酶催化域,如任何丝氨酸重组酶催化域(包括但不限于Gin、Sin、Tn3、Hin、β、γδ或PhiC31重组酶催化域)、任何酪氨酸重组酶域(包括但不限于Cre或FLP重组酶催化域),或任何演化的重组酶催化域。

[0010] 引导核苷酸序列-可编程DNA结合蛋白域可以选自下组：核酸酶无活性的Cas9(dCas9)域、核酸酶无活性的Cpf1域、核酸酶无活性的Argonaute域，及其变体。在某些实施方案中，引导核苷酸序列-可编程DNA结合蛋白域是核酸酶无活性的Cas9(dCas9)域。在某些实施方案中，dCas9域的氨基酸序列包含对应于SEQ ID NO:1中的D10A和/或H840A突变的突变。在另一个实施方案中，dCas9域的氨基酸序列包含对应于SEQ ID NO:1中的D10A突变的突变和对应于SEQ ID NO:1中的H840A突变的突变。在另一个实施方案中，dCas9域的氨基酸序列进一步不包括SEQ ID NO:1中所示的N-末端甲硫氨酸。在某些实施方案中，dCas9域的氨基酸序列包含SEQ ID NO:712。在一个实施方案中，dCas9域的氨基酸序列与SEQ ID NO:712具有大于95％的序列一致性。在一个实施方案中，dCas9域的氨基酸序列与SEQ ID NO:712具有大于96、97、98、99％或更高的序列一致性。在一些实施方案中，重组酶催化域是丝氨酸重组酶催化域或酪氨酸重组酶催化域。

[0011] 在一个实施方案中，重组酶催化域的氨基酸序列是Gin重组酶催化域。在一些实施方案中，Gin重组酶催化域包含对应于选自下组的一个或多个突变的突变：SEQ ID NO:713中的H106Y、I127L、I136R和/或G137F突变。在实施方案中，Gin重组酶催化域的氨基酸序列包含对应于选自下组的两个或更多个突变的突变：SEQ ID NO:713中的I127L、I136R和/或G137F突变。在实施方案中，Gin重组酶催化域的氨基酸序列包含对应于SEQ ID NO:713中的I127L、I136R和G137F突变的突变。在另一个实施方案中，Gin重组酶的氨基酸序列已得以进一步突变。在具体的实施方案中，Gin重组酶催化域的氨基酸序列包含SEQ ID NO:713。

[0012] 在另一个实施方案中，重组酶催化域的氨基酸序列是Hin重组酶、β重组酶、Sin重组酶、Tn3重组酶、γδ重组酶、Cre重组酶；FLP重组酶；或phiC31重组酶催化域。

[0013] 在一个实施方案中，Cre重组酶的氨基酸序列是截短的。在另一个实施方案中，酪氨酸重组酶催化域是Cre重组酶的25kDa羧基端域。在另一个实施方案中，Cre重组酶以氨基酸R118、A127、E138或R154开始(每种情况下其之前都是甲硫氨酸)。在一个实施方案中，重组酶的氨基酸序列已得以进一步突变。在某些实施方案中，重组酶催化域是经演化的重组酶催化域。在一些实施方案中，重组酶的氨基酸序列已得以进一步突变。

[0014] 在一些实施方案中，接头(例如第一、第二或第三接头)可以具有约0埃至约81埃的长度。接头通常具有约33埃至约81埃的长度。接头可以是肽的、非肽的或这两种类型的接头的组合。在某些实施方案中，接头是肽接头。在某些实施方案中，肽接头包含XTEN接头SGSETPGTSESATPES(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:9)，包含三肽GGS的一个或多个重复的氨基酸序列，或任何以下氨基酸序列：VPFLLEPDNINGKTC(SEQ ID NO:10)、GSAGSAAGSGEF(SEQ ID NO:11)、SIVAQLSRPDPA(SEQ ID NO:12)、MKIIEQLPSA(SEQ ID NO:13)、VRHKLKRVGS(SEQ ID NO:14)、GHGTGSTGSGSS(SEQ ID NO:15)、MSRPDPA(SEQ ID NO:16)或GGSM(SEQ ID NO:17)。在另一个实施方案中，肽接头包含三肽GGS的一个或多个重复。在一个实施方案中，肽接头包含三肽GGS的一至五个重复。在另一个实施方案中，肽接头包含三肽GGS的六至十个重复。在具体的实施方案中，肽接头包含三肽GGS的八个重复。在另一个实施方案中，肽接头长约18至约27个氨基酸。在某些实施方案中，肽接头长24个氨基酸。在某些实施方案中，肽接头具有氨基酸序列GGSGGSGGSGGSGGSGGSGGSGGS(SEQ ID NO:183)。

[0015] 在某些实施方案中，接头是非肽接头。在某些实施方案中，非肽接头包含聚乙二醇

(PEG)、聚丙二醇(PPG)、共聚(乙烯/丙烯)二醇、聚氧乙烯(POE)、聚氨酯、聚磷腈、多糖、右旋糖酐、聚乙烯醇、聚乙烯吡咯烷酮、聚乙烯乙醚、聚丙烯酰胺、聚丙烯酸酯、聚氰基丙烯酸酯、脂质聚合物、甲壳质、透明质酸、肝素或烷基接头。在某些实施方案中，烷基接头具有式—NH—$(CH_2)_s$—C(O)—，其中s是1-100之间的任何整数，包括端点。在某些实施方案中，s是1-20的任何整数，包括端点。

[0016] 在另一个实施方案中，融合蛋白进一步包含核定位信号(NLS)域。在某些实施方案中，NLS域经由一个或多个第二接头与引导核苷酸序列-可编程DNA结合蛋白域或重组酶催化域结合。

[0017] 在一个实施方案中，融合蛋白包含结构NH_2-[重组酶催化域]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的第二接头序列]-[NLS域]-COOH。在某些实施方案中，融合蛋白与SEQ ID NO:719中所示的氨基酸序列具有大于85％、90％、95％、98％或99％的序列一致性。在具体的实施方案中，融合蛋白包含SEQ ID NO:719中所示的氨基酸序列。在一个实施方案中，融合蛋白由SEQ ID NO:719中所示的氨基酸序列组成。

[0018] 在另一个实施方案中，融合蛋白进一步包含一个或多个亲和标签。在一个实施方案中，亲和标签选自下组:FLAG标签、多组氨酸(多His)标签、多精氨酸(多Arg)标签、Myc标签和HA标签。在实施方案中，亲和标签是FLAG标签。在具体的实施方案中，FLAG标签具有序列PKKKRKV(SEQ ID NO:702)。在另一个实施方案中，一个或多个亲和标签经由一个或多个第三接头与引导核苷酸序列-可编程DNA结合蛋白域、重组酶催化域或NLS域结合。在某些实施方案中，第三接头是肽接头。

[0019] 本文描述的融合蛋白的元件可以不做限制以任何顺序。在一些实施方案中，融合蛋白具有结构NH_2-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH、NH_2-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH或NH_2-[二分叉(bifurcated)或环状置换(circularly permuted)的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。

[0020] 在一些实施方案中，融合蛋白具有结构NH_2-[任选的亲和标签]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-COOH、NH_2-[任选的亲和标签]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-COOH或NH_2-[任选的亲和标签]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-COOH。

[0021] 在某些实施方案中，融合蛋白与SEQ ID NO:185中所示的氨基酸序列具有大于85％、90％、95％、98％或99％的序列一致性。在具体的实施方案中，融合蛋白具有SEQ ID NO:185中所示的氨基酸序列。在某些实施方案中，融合蛋白的重组酶催化域与SEQ ID NO:

185的氨基酸1-142中所示的氨基酸序列(其与SEQ ID NO:713中所示的序列相同)具有大于85%、90%、95%、98%或99%的序列一致性。在某些实施方案中,dCas9域与SEQ ID NO:185的氨基酸167-1533中所示的氨基酸序列(其与SEQ ID NO:712中所示的序列相同)具有大于90%、95%或99%的序列一致性。在某些实施方案中,本公开的融合蛋白与SEQ ID NO:185的氨基酸1-1544中所示的氨基酸序列(其与SEQ ID NO:719中所示的序列相同)具有大于90%、95%或99%的序列一致性。在一个实施方案中,融合蛋白与引导RNA(gRNA)结合。

[0022] 在一个方面,本公开提供了本文所述的融合蛋白的二聚体。在某些实施方案中,二聚体与靶DNA分子结合。在某些实施方案中,二聚体的每个融合蛋白与靶DNA分子的相同链结合。在某些实施方案中,二聚体的每个融合蛋白与靶DNA分子的相对链结合。在某些实施方案中,二聚体的gRNA与靶DNA分子的重组酶位点侧翼的gRNA结合位点杂交。在某些实施方案中,重组酶位点包含res、gix、hix、six、resH、LoxP、FTR或att核心或相关核心序列。在某些实施方案中,重组酶位点包含gix核心或gix相关核心序列。在进一步的实施方案中,gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为3至7个碱基对。在某些实施方案中,gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为5至6个碱基对。

[0023] 在某些实施方案中,第一二聚体与第二二聚体结合,从而形成融合蛋白的四聚体。在一个方面,本公开提供了本文所述的融合蛋白的四聚体。在某些实施方案中,四聚体与靶DNA分子结合。在某些实施方案中,每个二聚体与DNA的相对链结合。在其他实施方案中,每个二聚体与DNA的相同链结合。

[0024] 在另一个方面,本公开提供了用于两个DNA分子之间的位点特异性重组的方法,其包括:(a)使第一DNA与第一融合蛋白接触,其中引导核苷酸序列-可编程DNA结合蛋白域结合第一gRNA,所述第一gRNA与第一DNA的第一区域杂交;(b)使第一DNA与第二融合蛋白接触,其中第二融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第二gRNA,所述第二gRNA与第一DNA的第二区域杂交;(c)使第二DNA与第三融合蛋白接触,其中第三融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第三gRNA,所述第三gRNA与第二DNA的第一区域杂交;和(d)使第二DNA与第四融合蛋白接触,其中第四融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第四gRNA,所述第四gRNA与第二DNA的第二区域杂交;其中在使得DNA重组的条件下,步骤(a)-(d)中融合蛋白的结合导致融合蛋白的重组酶催化域的四聚化,并且其中第一、第二、第三和/或第四融合蛋白是本文所述的任何融合蛋白。

[0025] 在一个实施方案中,第一和第二DNA分子具有不同的序列。在另一个实施方案中,步骤(a)和(b)的gRNA与第一DNA的相对链杂交,并且步骤(c)和(d)的gRNA与第二DNA的相对链杂交。在另一个实施方案中,其中步骤(a)和(b)的gRNA;和/或步骤(c)和(d)的gRNA与它们各自的DNA的区域杂交,所述区域相隔不超过10、不超过15、不超过20、不超过25、不超过30、不超过40、不超过50、不超过60、不超过70、不超过80、不超过90或不超过100个碱基对。在某些实施方案中,步骤(a)和(b)的gRNA;和/或步骤(c)和(d)的gRNA在重组酶位点侧翼的gRNA结合位点处与它们各自的DNA的区域杂交(参见例如图1D)。在某些实施方案中,重组酶位点包含res、gix、hix、six、resH、LoxP、FTR或att核心或相关核心序列。在某些实施方案中,重组酶位点包含gix核心或gix相关核心序列。在某些实施方案中,gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为3至7个碱基对。在某些实施方案中,gix核心

或gix相关核心序列与至少一个gRNA结合位点之间的距离为5至6个碱基对。

[0026] 本文提供的用于位点特异性重组的方法也可以与单一DNA分子一起使用。在一个方面，本公开提供了用于单一DNA分子的两个区域之间的位点特异性重组的方法，其包括：(a)使DNA与第一融合蛋白接触，其中引导核苷酸序列-可编程DNA结合蛋白域结合第一gRNA，所述第一gRNA与DNA的第一区域杂交；(b)使DNA与第二融合蛋白接触，其中第二融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第二gRNA，所述第二gRNA与DNA的第二区域杂交；(c)使DNA与第三融合蛋白接触，其中第三融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第三gRNA，所述第三gRNA与DNA的第三区域杂交；和(d)使DNA与第四融合蛋白接触，其中第四融合蛋白的引导核苷酸序列-可编程DNA结合蛋白域结合第四gRNA，所述第四gRNA与DNA的第四区域杂交；其中在使得DNA重组的条件下，步骤(a)-(d)中融合蛋白的结合导致融合蛋白的重组酶催化域的四聚化，并且其中第一、第二、第三和/或第四融合蛋白是所述的任何融合蛋白。

[0027] 在某些实施方案中，重组的单一DNA分子的两个区域具有不同的序列。在另一个实施方案中，重组导致DNA分子的区域的缺失。在具体的实施方案中，缺失的DNA分子的区域易于在减数分裂中发生交换事件。在一个实施方案中，步骤(a)-(d)的第一和第二gRNA与DNA的相同链杂交，并且步骤(a)-(d)的第三和第四gRNA与DNA的相对链杂交。在另一个实施方案中，步骤(a)和(b)的gRNA与DNA的区域杂交，所述区域相隔不超过50、不超过60、不超过70、不超过80、不超过90或不超过100个碱基对，并且步骤(c)和(d)的gRNA与DNA的区域杂交，所述区域相隔不超过10、不超过15、不超过20、不超过25、不超过30、不超过40、不超过50、不超过60、不超过70、不超过80、不超过90或不超过100个碱基对。在某些实施方案中，步骤(a)和(b)的gRNA；和/或步骤(c)和(d)的gRNA与重组酶位点侧翼的gRNA结合位点杂交。在某些实施方案中，重组酶位点包含res、gix、hix、six、resH、LoxP、FTR或att核心或相关核心序列。在一个实施方案中，重组酶位点包含gix核心或gix相关核心序列。在某些实施方案中，gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为3至7个碱基对。在某些实施方案中，gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离为5至6个碱基对。

[0028] 本文所述的DNA可以在细胞中。在某些实施方案中，细胞是真核细胞。在某些实施方案中，细胞是植物细胞。在某些实施方案中，细胞是原核细胞。在一些实施方案中，细胞可以是哺乳动物细胞。在一些实施方案中，细胞可以是人细胞。在某些实施方案中，细胞在受试者中。在一些实施方案中，受试者可以是哺乳动物。在某些实施方案中，受试者是人。在某些实施方案中，细胞可以是植物细胞。

[0029] 在一个方面，本公开提供了编码本文所公开的任何融合蛋白的多核苷酸。在某些实施方案中，本公开提供了包含编码本文所公开的任何融合蛋白的多核苷酸的载体。

[0030] 在另一个方面，本公开提供了包含用于表达本文所公开的任何融合蛋白的遗传构建体的细胞。

[0031] 在一个方面，本公开提供了包含本文所公开的任何融合蛋白的试剂盒。在另一个方面，本公开提供了包含编码本文所公开的任何融合蛋白的多核苷酸的试剂盒。在另一个方面，本公开提供了包含用于重组蛋白表达的载体的试剂盒，其中所述载体包含编码本文所公开的任何融合蛋白的多核苷酸。在另一个方面，本公开提供了包含细胞的试剂盒，所述

细胞包含用于表达本文所公开的任何融合蛋白的遗传构建体。在一个实施方案中，试剂盒进一步包含一个或多个gRNA和/或用于表达一个或多个gRNA的载体。

[0032] 如下所述，在某些实施方案的详细描述中阐述了本发明的某些实施方案的细节。根据定义、实施例、附图和权利要求，本发明的其他特征、目的和优点将显而易见。

[0033] 附图简述

[0034] 图1A-1D。实验设置的概述。用(图1A)在hU6启动子的控制下的引导RNA表达载体、(图1B)在CMV启动子的控制下的recCas9表达载体和(图1C)recCas9报告质粒转染细胞。这些组分的共转染导致在靶位点处重新组装引导RNA编程的recCas9(图1D)。这将介导polyA终止子的缺失，允许GFP的转录。引导RNA表达载体和引导RNA序列缩写为gRNA。

[0035] 图2A-2F。融合接头长度和靶位点间隔区变体的优化。在这些实验中使用单一靶引导RNA表达载体pHU6-NT1或非靶载体pHU6-BC74。序列可以在表6-9中找到。(图2A)显示靶位点的一部分，其中引导RNA靶位点为黑色带有虚线下划线并且gix核心序列位点为黑色。假gix位点的任一侧的5'和3'序列是相同的，但是是反向的，并且被pHU6-NT1识别。将gix假位点与5'和3'结合位点分开的碱基对间隔区的数量分别由X和Y表示。该图分别描绘了SEQ ID NO:700和703。(图2B)Z表示将Ginβ与dCas9连接的GGS重复的数量。当X＝Y时，对于连接Gin催化域和dCas9域的接头(图2C)$(GGS)_2$(SEQ ID NO:182)、(图2D)$(GGS)_5$(SEQ ID NO:701)和(图2E)$(GGS)_8$(SEQ ID NO:183)，评估recCas9活性。(图2F)测定recCas9在由不均匀碱基对间隔区(X≠Y)组成的靶位点上的活性；包括X＝Y＝6用于比较。所有实验一式三份进行，并从这些实验中减去背景荧光。eGFP阳性细胞的百分比仅是那些转染的细胞的(即，表达组成型表达的iRFP基因)，并且每个实验记录至少6,000个活事件(live event)。引导RNA表达载体和引导RNA序列缩写为“gRNA”。值和误差条分别代表三个独立生物学重复的平均值和标准偏差。

[0036] 图3A-3B。正向和反向引导RNA对recCas9活性的依赖性。(图3A)在PCDH15内发现的序列替换了图1A-1D中测试的靶位点。在假gix核心位点的5'和3'侧两者上，引导RNA可以靶向两个偏移序列。该图分别描绘了SEQ ID NO:704-705。(图3B)recCas9活性通过与所有四个引导RNA表达载体对和个别引导RNA载体与脱靶(O.T.)引导RNA载体共转染recCas9表达载体和报告质粒来测量。脱靶正向物和反向物含有分别靶向CLTA和VEGF的引导RNA序列。还显示了用报告质粒转染但未用靶引导RNA转染的对照实验。还显示了与不同引导RNA表达载体共转染但不与recCas9表达载体共转染的报告质粒的结果。所有实验一式四份进行，并且不从这些实验中减去背景荧光。eGFP阳性细胞的百分比仅是那些转染的细胞的(即，表达组成型表达的iRFP基因)，并且每个实验记录至少6,000个活事件。引导RNA表达载体和引导RNA序列缩写为gRNA。值和误差条分别代表四个独立生物学重复的平均值和标准偏差。

[0037] 图4A-4D。recCas9可以靶向与人基因组中的序列相同的多个序列。(图4A)图1A-1D中所示的靶位点被人基因组内发现的序列替换。有关序列参见表6。将recCas9表达载体与引导RNA载体对和报告质粒的所有组合共转化。脱靶引导RNA载体也与recCas9表达载体和报告质粒共转化，并含有靶向CLTA和VEGF的引导RNA序列(参见例如Guilinger et al., Fusion of catalytically inactive Cas9to FokI nuclease improves the specificity of genome modification.Nature biotechnology,(2014)，其全部内容在此通过引用并入)。eGFP阳性细胞的百分比反映了经转染的(iRFP阳性)细胞的百分比。每个实

验至少记录6,000个活事件。值和误差条分别代表至少三个独立的生物学重复的平均值和标准偏差。(图4B)再次进行转染实验,用SpecR替换recCas9表达载体和pUC中的抗性标志物。共转染和温育后,提取附加型DNA(episomal DNA),转化到大肠杆菌中并选择羧苄青霉素抗性。然后对菌落进行测序以确定(图4C)重组质粒与完全完整质粒的比率。(图4D)分离自转染细胞的附加型提取物的测序数据。列和行代表转染条件。每个单元格显示重组的质粒的百分比和比率。显示的值反映了两个独立的生物学重复的平均值和标准偏差。平均值和每个重复之间的平均差值显示为误差。引导RNA表达载体和引导RNA序列缩写为gRNA。

[0038] 图5A-5D。recCas9介导培养的人细胞中基因组DNA的引导RNA-和recCas9-依赖性缺失。(图5A)示意图显示位于染色体12的FAM19A2基因座的内含子区域内的预测的recCas9靶位点和用于巢式PCR的引物的位置。该图分别从上到下和从左到右描绘了SEQ ID NO:706-709。(图5B)来自用指定表达载体转染的细胞的模板的巢式基因组PCR的代表性结果(n=3个生物学重复;NTC=无模板对照)。星号表示1.3-kb预测的初级PCR产物的位置。箭头表示二级PCR后预测的缺失产物。这两个图来自相同的凝胶,但经切割以去除空白泳道。(图5C)对用所有四种gRNA表达载体转染的细胞的巢式基因组PCR产生的PCR产物的Sanger测序,并且recCas9表达载体与预测的重组后产物匹配。该图分别从上到下描绘了SEQ ID NO:710和711。(图5D)通过有限稀释巢式PCR测定的FAM19A2基因座的估计的最小缺失效率。显示的值反映了三次重复的平均值和标准偏差。

[0039] 图6。报告质粒构建。Golden Gate组装用于构建本工作中描述的报告质粒。所有组装均以共同质粒pCALNL-EGFP-Esp3I开始,该质粒衍生自pCALNL-EGFP并包含Esp3I限制性位点。显示的片段侧翼为Esp3I位点。Esp3I消化产生一系列相容的、独特的4碱基对5'突出端,以便按所示顺序进行组装。为了组装靶位点,将Esp3I(ThermoFisher Scientific,Waltham,MA)和五个片段加入到单个反应管中以允许Esp3I消化和T7连接的迭代循环。然后用Plasmid-Safe-ATP依赖性DNA酶(Epicentre,Madison,WI)消化反应以减少背景。通过菌落PCR分析菌落以鉴定与预期的全长5部分组装产物匹配的PCR产物;然后将来自这些菌落的质粒送去进行sanger测序。对于图4中所示的基因组报告物,将片段1和2以及片段4和5组合成编码整个靶位点的两个gBlock(IDT,Coralville,IA)片段(图中未显示)。然后如上所述完成组装。构建细节可以在支持材料的方法中找到。用于产生片段的寡核苷酸和gBLOCK可以在表2中找到。

[0040] 图7A和7B。演化以靶向人基因组的Rosa基因座中称为"36C6"的位点的Cre重组酶与dCas9融合。然后将该融合物用于以引导RNA依赖性方式重组含有Rosa靶位点的基于质粒的报告物。图7A证明了使用野生型Cre和36C6的接头优化的结果。显示靶向其关联报告物的GinB构建体用于参考。显示的1x 2x、5x和8x接头是接头中GGS重复的数量。图7B显示了回复分析的结果,其证明了对与dCas9融合的36C6进行突变可以影响嵌合融合物的相对引导依赖性。显示靶向其关联报告物的GinB构建体用于参考。GGS-36C6:1x GGS接头;2GGS-36C6(使用接头SEQ ID NO:181):2x GGS接头(使用接头SEQ ID NO:181)。

[0041] 图8。鉴定了人基因组中Rosa26位点侧翼的PAM,其可以支持dCas9结合(参见上图)。设计引导RNA和质粒报告物以测试内源性前间隔区是否可以支持dCas9-36C6活性。显示靶向gix报告物的GinB构建体用于参考。混合:Cas9和36C6之间所有5种接头变体的等份混合物。序列对应于SEQ ID NO:769(核苷酸序列)和770(氨基酸序列)。

[0042] 图9A-9B。Cre重组酶的各种测试的截短的位置显示在图9A中。与dCas9融合的Cre重组酶的截短的变体显示出可察觉的重组酶活性以及对Lox质粒报告物系统中引导RNA的存在的严格依赖性(图9B)。与dCas9融合的野生型Cre显示为阳性对照。

[0043] 定义

[0044] 如本文所用,除非上下文另外明确指出,否则单数形式“一种”、“一个”和“该/所述”包括单数和复数。因此,例如,提及“试剂”包括单一试剂和多个此类试剂。

[0045] 非限制性、示例性RNA-可编程DNA结合蛋白包括Cas9核酸酶、Cas9切口酶、核酸酶无活性的Cas9(dCas9)、CasX、CasY、Cpf1、C2c1、C2c2、C2C3和Argonaute。术语“Cas9”或“Cas9域”是指包含Cas9蛋白或其片段的RNA引导的核酸酶(例如,包含Cas9的活性、无活性或部分活性的DNA切割域,和/或Cas9的gRNA结合域的蛋白质)。Cas9具有两个切割域,其切割特定的DNA链(例如有义链和反义链)。可以产生切割任一条链的Cas9切口酶(包括但不限于spCas9的D10A和H840A)。可以不受限制地在本文所述的融合蛋白和方法中使用Cas9域(例如,核酸酶活性Cas9、核酸酶无活性的Cas9或Cas9切口酶)。此外,本文所述的任何引导核苷酸序列-可编程DNA结合蛋白可以用作切口酶。

[0046] Cas9核酸酶有时也称为casn1核酸酶或CRISPR(聚簇规则间隔短回文重复)相关核酸酶。CRISPR是适应性免疫系统,其提供针对移动遗传元件(病毒、转座元件和接合质粒)的保护。CRISPR簇含有间隔区,与先前的移动元件互补的序列,并靶向侵入核酸。CRISPR簇得以转录并加工成CRISPR RNA(crRNA)。在II型CRISPR系统中,对pre-crRNA的正确加工需要反式编码的小RNA(tracrRNA)、内源性核糖核酸酶3(rnc)和Cas9蛋白。tracrRNA充当用于pre-crRNA的核糖核酸酶3辅助加工的引导。随后,Cas9/crRNA/tracrRNA以内切核水解方式切割与间隔区互补的线性或环状dsDNA靶标。首先以内切核水解方式切割不与crRNA互补的靶链,然后以3'-5'外切核水解方式修剪(trim)。在自然界中,DNA结合和切割通常需要蛋白质和这两种RNA。然而,单一引导RNA(“sgRNA”或简称“gNRA”)可以经工程化以将crRNA和tracrRNA两者的方面并入单一RNA种类中。参见例如Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science337:816-821(2012),其全部内容通过引用并入本文。Cas9识别CRISPR重复序列中的短基序(PAM或前间隔区相邻基序),以帮助区分自我与非自我。Cas9核酸酶序列和结构是本领域技术人员熟知的(参见例如“Complete genome sequence of an M1strain of Streptococcus pyogenes.”Ferretti et al.,J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);“CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.”Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011);和“A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.”Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),其各自的全部内容通过引用并入本文)。已经在各种物种中描述了Cas9直系同源物,包括但不限于酿脓链球菌(S.pyogenes)和嗜热链球菌(S.thermophilus)。基于本公开,其他合适的Cas9核酸酶和序

列对于本领域技术人员将是显而易见的,并且此类Cas9核酸酶和序列包括来自Chylinski, Rhun,and Charpentier,“The tracrRNA and Cas9families of type II CRISPR-Cas immunity systems”(2013)RNA Biology 10:5,726-737中公开的生物体和基因座的Cas9序列;其全部内容通过引用并入本文。在一些实施方案中,Cas9核酸酶具有无活性的(例如失活的)DNA切割域,也就是说,Cas9是切口的酶。作为一个实例,Cas9核酸酶(例如Cas9切口酶)可以切割与gRNA结合的DNA链。作为另一个实例,Cas9核酸酶(例如Cas9切口酶)可以切割不与gRNA结合的DNA链。在另一个实施方案中,任何引导核苷酸序列-可编程DNA结合蛋白可以具有无活性的(例如失活的)DNA切割域,即,引导核苷酸序列-可编程DNA结合蛋白是切口酶。作为一个实例,引导核苷酸序列-可编程DNA结合蛋白可以切割与gRNA结合的DNA链。作为另一个实例,引导核苷酸序列-可编程DNA结合蛋白可以切割不与gRNA结合的DNA链。

[0047] 另外的示例性Cas9序列可以在2017年4月27日公开的题为“Evolved Cas9Proteins for Gene Editing”的国际公开号WO/2017/070633中找到。

[0048] 核酸酶失活的Cas9蛋白可以互换地称为“dCas9”蛋白(相当于核酸酶-“死亡的”Cas9)。在一些实施方案中,dCas9对应于,或部分或全部包含如下以SEQ ID NO:1所示的氨基酸。在一些实施方案中,提供了dCas9的变体(例如SEQ ID NO:1的变体)。例如,在一些实施方案中,提供了具有除D10A和H840A之外的突变的dCas9变体,其例如导致核酸酶失活的Cas9(dCas9)。举例来说,此类突变包括D10和H840处的其他氨基酸取代,或Cas9的核酸酶域内的其他取代(例如,HNH核酸酶亚域和/或RuvC1亚域中的取代)。在一些实施方案中,提供了dCas9的变体或同源物(例如SEQ ID NO:1的变体),其与SEQ ID NO:10至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%。在一些实施方案中,提供了dCas9的变体(例如SEQ ID NO:1的变体),其具有比SEQ ID NO:1短或长约5个氨基酸、约10个氨基酸、约15个氨基酸、约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多的氨基酸序列。

[0049] dCas9(D10A和H840A):

[0050] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE

IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD(SEQ ID NO:1)

[0051] 用于生成具有无活性的DNA切割域的Cas9蛋白(或其片段)的方法是已知的(参见例如Jinek et al.,Science.337:816-821(2012);Qi et al.,“Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression”(2013)Cell.28;152(5):1173-83,其各自的全部内容通过引用并入本文)。例如,已知Cas9的DNA切割域包括两个亚域,即HNH核酸酶亚域和RuvC1亚域。HNH亚域切割与gRNA互补的链,而RuvC1亚域切割非互补链。这些亚域内的突变可以沉默Cas9的核酸酶活性。例如,突变D10A和H840A完全使酿脓链球菌Cas9的核酸酶活性失活(参见例如,Jinek et al.,Science.337:816-821(2012);Qi et al.,Cell.28;152(5):1173-83(2013))。在一些实施方案中,提供了包含Cas9的片段的蛋白质。例如,在一些实施方案中,蛋白质包含两个Cas9域的一个:(1)Cas9的gRNA结合域;或(2)Cas9的DNA切割域。在一些实施方案中,包含Cas9或其片段的蛋白质称为“Cas9变体”。Cas9变体与Cas9或其片段共享同源性。例如,Cas9变体与野生型Cas9至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%。在一些实施方案中,Cas9变体包含Cas9的片段(例如,gRNA结合域或DNA切割域),使得该片段与野生型Cas9的相应的片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%。在一些实施方案中,野生型Cas9对应与来自酿脓链球菌(Streptococcus pyogenes)的Cas9(NCBI参考序列:NC_017053.1,SEQ ID NO:2(核苷酸);SEQ ID NO:3(氨基酸))。在一些实施方案中,Cas9域包含与野生型Cas9至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,与野生型Cas9相比,Cas9域包含具有1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、21、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50个或更多个突变的氨基酸序列。在一些实施方案中,与野生型Cas9相比,Cas9域包含具有至少10、至少15、至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少150、至少200、至少250、至少300、至少350、至少400、至少500、至少600、至少700、至少800、至少900、至少1000、至少1100或至少1200个相同的连续氨基酸残基的氨基酸序列。在一些实施方案中,Cas9变体包含Cas9的片段(例如,gRNA结合域或DNA切割域),使得该片段与野生型Cas9的相应的片段至少约70%相同、至少约80%相同、至少约90%相同、至少约95%相同、至少约96%相同、至少约97%相同、至少约98%相同、至少约99%相同、至少约99.5%相同或至少约99.9%相同。在一些实施方案中,片段是相应的野生型Cas9的氨基酸长度的至少30%、至少35%、至少40%、至少45%、至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%相同、至少96%、至少97%、至少98%、至少99%或至少99.5%。

[0052] 在一些实施方案中，片段的长度为至少100个氨基酸。在一些实施方案中，片段的长度为至少100、150、200、250、300、350、400、450、500、550、600、650、700、750、800、850、900、950、1000、1050、1100、1150、1200、1250或1300个氨基酸。

[0053] ATGGATAAGAAATACTCAATAGGCTTAGATATCGGCACAAATAGCGTCGGATGGGCGGTGATCACTGA
TGATTATAAGGTTCCGTCTAAAAAGTTCAAGGTTCTGGGAAATACAGACCGCCACAGTATCAAAAAAAATCTTATA
GGGGCTCTTTTATTTGGCAGTGGAGAGACAGCGGAAGCGACTCGTCTCAAACGGACAGCTCGTAGAAGGTATACAC
GTCGGAAGAATCGTATTTGTTATCTACAGGAGATTTTTTCAAATGAGATGGCGAAAGTAGATGATAGTTTCTTTCA
TCGACTTGAAGAGTCTTTTTTGGTGGAAGAAGACAAGAAGCATGAACGTCATCCTATTTTTGGAAATATAGTAGAT
GAAGTTGCTTATCATGAGAAATATCCAACTATCTATCATCTGCGAAAAAAATTGGCAGATTCTACTGATAAAGCGG
ATTTGCGCTTAATCTATTTGGCCTTAGCGCATATGATTAAGTTTCGTGGTCATTTTTTGATTGAGGGAGATTTAAA
TCCTGATAATAGTGATGTGGACAAACTATTTATCCAGTTGGTACAAATCTACAATCAATTATTTGAAGAAAACCCT
ATTAACGCAAGTAGAGTAGATGCTAAAGCGATTCTTTCTGCACGATTGAGTAAATCAAGACGATTAGAAAATCTCA
TTGCTCAGCTCCCCGGTGAGAAGAGAAATGGCTTGTTTGGGAATCTCATTGCTTTGTCATTGGGATTGACCCCTAA
TTTTAAATCAAATTTTGATTTGGCAGAAGATGCTAAATTACAGCTTTCAAAAGATACTTACGATGATGATTTAGAT
AATTTATTGGCGCAAATTGGAGATCAATATGCTGATTTGTTTTTGGCAGCTAAGAATTTATCAGATGCTATTTTAC
TTTCAGATATCCTAAGAGTAAATAGTGAAATAACTAAGGCTCCCCTATCAGCTTCAATGATTAAGCGCTACGATGA
ACATCATCAAGACTTGACTCTTTTAAAAGCTTTAGTTCGACAACAACTTCCAGAAAAGTATAAAGAAATCTTTTTT
GATCAATCAAAAAACGGATATGCAGGTTATATTGATGGGGGAGCTAGCCAAGAAGAATTTTATAAATTTATCAAAC
CAATTTTAGAAAAAATGGATGGTACTGAGGAATTATTGGTGAAACTAAATCGTGAAGATTTGCTGCGCAAGCAACG
GACCTTTGACAACGGCTCTATTCCCCATCAAATTCACTTGGGTGAGCTGCATGCTATTTTGAGAAGACAAGAAGAC
TTTTATCCATTTTTAAAAGACAATCGTGAGAAGATTGAAAAAATCTTGACTTTTCGAATTCCTTATTATGTTGGTC
CATTGGCGCGTGGCAATAGTCGTTTTGCATGGATGACTCGGAAGTCTGAAGAAACAATTACCCCATGGAATTTTGA
AGAAGTTGTCGATAAAGGTGCTTCAGCTCAATCATTTATTGAACGCATGACAAACTTTGATAAAAATCTTCCAAAT
GAAAAAGTACTACCAAAACATAGTTTGCTTTATGAGTATTTTACGGTTTATAACGAATTGACAAAGGTCAAATATG
TTACTGAGGGAATGCGAAAACCAGCATTTCTTTCAGGTGAACAGAAGAAAGCCATTGTTGATTTACTCTTCAAAAC
AAATCGAAAAGTAACCGTTAAGCAATTAAAAGAAGATTATTTCAAAAAAATAGAATGTTTTGATAGTGTTGAAATT
TCAGGAGTTGAAGATAGATTTAATGCTTCATTAGGCGCCTACCATGATTTGCTAAAAATTATTAAAGATAAAGATT
TTTTGGATAATGAAGAAAATGAAGATATCTTAGAGGATATTGTTTTAACATTGACCTTATTTGAAGATAGGGGGAT
GATTGAGGAAAGACTTAAAACATATGCTCACCTCTTTGATGATAAGGTGATGAAACAGCTTAAACGTCGCCGTTAT
ACTGGTTGGGGACGTTTGTCTCGAAAATTGATTAATGGTATTAGGGATAAGCAATCTGGCAAAACAATATTAGATT
TTTTGAAATCAGATGGTTTTGCCAATCGCAATTTTATGCAGCTGATCCATGATGATAGTTTGACATTTAAAGAAGA
TATTCAAAAAGCACAGGTGTCTGGACAAGGCCATAGTTTACATGAACAGATTGCTAACTTAGCTGGCAGTCCTGCT
ATTAAAAAAGGTATTTTACAGACTGTAAAAATTGTTGATGAACTGGTCAAAGTAATGGGGCATAAGCCAGAAAATA
TCGTTATTGAAATGGCACGTGAAAATCAGACAACTCAAAAGGGCCAGAAAAATTCGCGAGAGCGTATGAAACGAAT
CGAAGAAGGTATCAAAGAATTAGGAAGTCAGATTCTTAAAGAGCATCCTGTTGAAAATACTCAATTGCAAAATGAA
AAGCTCTATCTCTATTATCTACAAAATGGAAGAGACATGTATGTGGACCAAGAATTAGATATTAATCGTTTAAGTG
ATTATGATGTCGATCACATTGTTCCACAAAGTTTCATTAAAGACGATTCAATAGACAATAAGGTACTAACGCGTTC
TGATAAAAATCGTGGTAAATCGGATAACGTTCCAAGTGAAGAAGTAGTCAAAAAGATGAAAAACTATTGGAGACAA
CTTCTAAACGCCAAGTTAATCACTCAACGTAAGTTTGATAATTTAACGAAAGCTGAACGTGGAGGTTTGAGTGAAC

TTGATAAAGCTGGTTTTATCAAACGCCAATTGGTTGAAACTCGCCAAATCACTAAGCATGTGGCACAAATTTTGGA
TAGTCGCATGAATACTAAATACGATGAAAATGATAAACTTATTCGAGAGGTTAAAGTGATTACCTTAAAATCTAAA
TTAGTTTCTGACTTCCGAAAAGATTTCCAATTCTATAAAGTACGTGAGATTAACAATTACCATCATGCCCATGATG
CGTATCTAAATGCCGTCGTTGGAACTGCTTTGATTAAGAAATATCCAAAACTTGAATCGGAGTTTGTCTATGGTGA
TTATAAAGTTTATGATGTTCGTAAAATGATTGCTAAGTCTGAGCAAGAAATAGGCAAAGCAACCGCAAAATATTTC
TTTTACTCTAATATCATGAACTTCTTCAAAACAGAAATTACACTTGCAAATGGAGAGATTCGCAAACGCCCTCTAA
TCGAAACTAATGGGGAAACTGGAGAAATTGTCTGGGATAAAGGGCGAGATTTTGCCACAGTGCGCAAAGTATTGTC
CATGCCCCAAGTCAATATTGTCAAGAAAACAGAAGTACAGACAGGCGGATTCTCCAAGGAGTCAATTTTACCAAAA
AGAAATTCGGACAAGCTTATTGCTCGTAAAAAAGACTGGGATCCAAAAAAATATGGTGGTTTTGATAGTCCAACGG
TAGCTTATTCAGTCCTAGTGGTTGCTAAGGTGGAAAAAGGGAAATCGAAGAAGTTAAAATCCGTTAAAGAGTTACT
AGGGATCACAATTATGGAAAGAAGTTCCTTTGAAAAAAATCCGATTGACTTTTTAGAAGCTAAAGGATATAAGGAA
GTTAAAAAAGACTTAATCATTAAACTACCTAAATATAGTCTTTTTGAGTTAGAAAACGGTCGTAAACGGATGCTGG
CTAGTGCCGGAGAATTACAAAAAGGAAATGAGCTGGCTCTGCCAAGCAAATATGTGAATTTTTTATATTTAGCTAG
TCATTATGAAAAGTTGAAGGGTAGTCCAGAAGATAACGAACAAAAACAATTGTTTGTGGAGCAGCATAAGCATTAT
TTAGATGAGATTATTGAGCAAATCAGTGAATTTTCTAAGCGTGTTATTTTAGCAGATGCCAATTTAGATAAAGTTC
TTAGTGCATATAACAAACATAGAGACAAACCAATACGTGAACAAGCAGAAAATATTATTCATTTATTTACGTTGAC
GAATCTTGGAGCTCCCGCTGCTTTTAAATATTTTGATACAACAATTGATCGTAAACGATATACGTCTACAAAAGAA
GTTTTAGATGCCACTCTTATCCATCAATCCATCACTGGTCTTTATGAAACACGCATTGATTTGAGTCAGCTAGGAG
GTGACTGA(SEQ ID NO:2)

[0054] MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLAD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSR
RLENLIAQLPGEKRNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGHSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMG
HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD
INRLSDYDVDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER
GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY
HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI
RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG
FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG
RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA
NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID
LSQLGGD(SEQ ID NO:3)

[0055] 在一些实施方案中,野生型Cas9对应于,或包含SEQ ID NO:4(核苷酸)和/或SEQ

ID NO:5(氨基酸)。

[0056] ATGGATAAAAAGTATTCTATTGGTTTAGACATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGA
TGAATACAAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATC
GGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGAGAAGGTATACAC
GTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCA
CCGTTTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGAT
GAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGG
ACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAA
TCCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCT
ATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGA
TCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAA
TTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGAC
AATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCC
TATCTGACATACTGAGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGA
ACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTCTTT
GATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAAC
CCATATTAGAGAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCG
GACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCAGGAGGAT
TTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGAC
CCCTGGCCCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGA
GGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAATTTACCGAAC
GAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATG
TCACTGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGAC
CAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATC
TCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACT
TCCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAAT
GATTGAGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTAT
ACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCGATT
TTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGA
TATACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCC
ATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAA
ACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAAGAG
AATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAAC
GAGAAACTTTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTAT
CTGATTACGACGTCGATCACATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGCTTACACG
CTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGG
CAGCTCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTG
AACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGTTGCACAGATACT
AGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCA

AAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACG
ACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGG
TGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAAATAC
TTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTT
TAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTT
GTCCATGCCCCAAGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCA
AAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAGCCCTA
CAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATT
ATTGGGGATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAG
GAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAAAACGGATGT
TGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGC
GTCCCATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACAT
TATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCCAATCTGGACAAAG
TATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCT
TACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAG
GAGGTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTG
GGGGTGACGGATCCCCCAAGAAGAAGAGGAAAGTCTCGAGCGACTACAAAGACCATGACGGTGATTATAAAGATCA
TGACATCGATTACAAGGATGACGATGACAAGGCTGCAGGA(SEQ ID NO:4)

[0057] MDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTA
RRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVD
STDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSR
RLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNL
SDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEF
YKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRI
PYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNEL
TKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKI
IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSG
KTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMG
RHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQEL
DINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAE
RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINN
YHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGE
IRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYG
GFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELEN
GRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILAD
ANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRI
DLSQLGGD(SEQ ID NO:5)

[0058] 在一些实施方案中,Cas9是指来自以下的Cas9:溃疡棒杆菌(Corynebacterium ulcerans)(NCBI Refs:NC_015683.1,NC_017317.1);白喉棒杆菌(Corynebacterium

diphtheria)(NCBI Refs:NC_016782.1,NC_016786.1);Spiroplasma syrphidicola(NCBI Ref:NC_021284.1);间型普雷沃氏菌(Prevotella intermedia)(NCBI Ref:NC_017861.1);台湾螺原体(Spiroplasma taiwanense)(NCBI Ref:NC_021846.1);海豚链球菌(Streptococcus iniae)(NCBI Ref:NC_021314.1);波罗的海贝尔氏菌(Belliella baltica)(NCBI Ref:NC_018010.1);Psychroflexus torquisI(NCBI Ref:NC_018721.1);嗜热链球菌(Streptococcus thermophilus)(NCBI Ref:YP_820832.1)、无害李斯特氏菌(Listeria innocua)(NCBI Ref:NP_472073.1)、空肠弯曲杆菌(Campylobacter jejuni)(NCBI Ref:YP_002344900.1)或脑膜炎奈瑟氏球菌(Neisseria meningitidis)(NCBI Ref:YP_002342100.1)或者是指来自任何其他生物体的Cas9。

[0059] Cas9识别靶DNA序列中CRISPR重复序列中的短基序(PAM基序)。如本文所用,“PAM基序”或“前间隔区相邻基序”是指在CRISPR细菌适应性免疫系统中直接在由Cas9核酸酶靶向的DNA序列之后的DNA序列。PAM是入侵病毒或质粒的组分,但不是细菌CRISPR基因座的组分。自然地,若其后没有PAM序列,则Cas9不会成功结合或切割靶DNA序列。PAM是靶向组分(在细菌基因组中未发现),其区分细菌自我与非自我DNA,从而防止CRISPR基因座被Cas9核酸酶活性靶向和破坏。

[0060] 野生型酿脓链球菌(Streptococcus pyogenes)Cas9识别规范PAM序列(例如来自嗜热链球菌(Streptococcus thermophiles)、金黄色葡萄球菌、脑膜炎奈瑟氏球菌或齿垢密螺旋体(Treponema denticolaor)的Cas9)及其Cas9变体已在本领域中描述为具有不同的或更放松的PAM需求。通常,Cas9蛋白,例如来自酿脓链球菌的Cas9(spCas9),需要规范的NGG PAM序列来结合特定的核酸区域,其中“NGG”中的“N”是腺嘌呤(A)、胸腺嘧啶(T)、鸟嘌呤(G)或胞嘧啶(C),并且G是鸟嘌呤。这可以限制在基因组内编辑期望的碱基的能力。在一些实施方案中,本文提供的碱基编辑融合蛋白需要定位于精确的位置处,例如,其中靶碱基在4碱基区域(例如“脱氨基作用窗口”)内,其在PAM的上游的约15个碱基。参见Komor,A.C.,et al.,“Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage”Nature 533,420-424(2016),其全部内容在此通过引用并入。在一些实施方案中,脱氨基作用窗口在2、3、4、5、6、7、8、9或10碱基区域内。在一些实施方案中,脱氨基作用窗口在PAM的上游的5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个碱基。因此,在一些实施方案中,本文提供的任何融合蛋白可以含有能够结合不含规范的(例如,NGG)PAM序列的核苷酸序列的Cas9域。本领域中已经描述了结合非规范PAM序列的Cas9域,并且其对于熟练技术人员而言将是显而易见的。例如,结合非规范PAM序列的Cas9域已经描述于Kleinstiver,B.P.,et al.,“Engineered CRISPR-Cas9nucleases with altered PAM specificities”Nature 523,481-485(2015);和Kleinstiver,B.P.,et al.,“Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9by modifying PAM recognition”Nature Biotechnology 33,1293-1298(2015);每篇的全部内容在此通过引用并入。还参见:Klenstiver et al.,Nature 529,490-495,2016;Ran et al.,Nature,Apr 9;520(7546):186-191,2015;Hou et al.,Proc Natl Acad Sci U S A,110(39):15644-9,2014;Prykhozhij et al.,PLoS One,10(3):e0119372,2015;Zetsche et al.,Cell 163,759-771,2015;Gao et al.,Nature Biotechnology,doi:10.1038/nbt.3547,2016;Want et al.,Nature 461,754-761,2009;

Chavez et al.,doi:dx dot doi dot org/10.1101/058974;Fagerlund et al.,Genome Biol.2015;16:25,2015;Zetsche et al.,Cell,163,759-771,2015;和Swarts et al.,Nat Struct Mol Biol,21(9):743-53,2014,其每一个的全部内容通过引用并入本文。

[0061] 因此,本公开的引导核苷酸序列-可编程DNA结合蛋白可以识别多种PAM序列,包括但不限于:NGG、NGAN(SEQ ID NO:741)、NGNG(SEQ ID NO:742)、NGAG(SEQ ID NO:743)、NGCG(SEQ ID NO:744)、NNGRRT(SEQ ID NO:745)、NGRRN(SEQ ID NO:746)、NNNRRT(SEQ ID NO:747)、NNNGATT(SEQ ID NO:748)、NNAGAAW(SEQ ID NO:749)、NAAAC(SEQ ID NO:750)、TTN、TTTN(SEQ ID NO:751)和YTN,其中Y是嘧啶,并且N是任何核碱基。

[0062] 具有不同的PAM特异性的RNA-可编程DNA结合蛋白的一个实例是来自普雷沃氏菌(Prevotella)和弗朗西斯菌(Francisella)1(Cpf1)的聚簇规则间隔短回文重复。与Cas9类似,Cpf1也是2类CRISPR效应物。已经显示,Cpf1介导了强大的DNA干扰,其具有与Cas9不同的特征。Cpf1是缺乏tracrRNA的单一RNA引导的内切核酸酶,并且它利用富含T的前间隔区相邻基序(TTN、TTTN(SEQ ID NO:751)或YTN)。此外,Cpf1经由交错的DNA双链断裂切割DNA。在16种Cpf1家族蛋白中,来自氨基酸球菌(Acidaminococcus)和毛螺菌(Lachnospiraceae)的两种酶显示在人细胞中具有有效的基因组编辑活性。

[0063] 本文还提供了核酸酶无活性的Cpf1(dCpf1)变体,其可以用作RNA-可编程DNA结合蛋白域。Cpf1蛋白具有RuvC样内切核酸酶域,其类似于Cas9的RuvC域,但不具有HNH内切核酸酶域,并且Cpf1的N端不具有Cas9的alpha螺旋识别叶(lobe)。它在Zetsche et al.,Cell,163,759-771,2015(其全部内容通过引用并入本文)中显示,Cpf1的RuvC样域负责切割两条DNA链并且RuvC样域的失活使Cpf1核酸酶活性失活。例如,对应于新凶手弗朗西斯菌(Francisella novicida)Cpf1(SEQ ID NO:714)中的D917A、E1006A或D1255A的突变使Cpf1核酸酶活性失活。在一些实施方案中,本公开的dCpf1包含对应于SEQ ID NO:714中D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。应当理解,可以根据本公开使用使Cpf1的RuvC域失活的任何突变,例如取代突变、缺失或插入。

[0064] 在一些实施方案中,本公开的引导核苷酸序列-可编程DNA结合蛋白域对PAM序列没有需求。此类引导核苷酸序列-可编程DNA结合蛋白的一个实例可以是来自格氏嗜盐碱杆菌(Natronobacterium gregoryi)的Argonaute蛋白(NgAgo)。NgAgo是ssDNA引导的内切核酸酶。NgAgo结合约24个核苷酸的5′磷酸化ssDNA(gDNA),以将其引导至其靶位点,并将在gDNA位点处产生DNA双链断裂。与Cas9相比,NgAgo-gDNA系统不需要前间隔区相邻基序(PAM)。使用核酸酶无活性的NgAgo(dNgAgo)可以极大地扩展可以靶向的密码子。NgAgo的表征和使用已经描述于Gao et al.,Nat Biotechnol.Epub 2016May 2.PubMed PMID:27136078;Swarts et al.,Nature.507(7491)(2014):258-61;和Swarts et al.,Nucleic Acids Res.43(10)(2015):5120-9,其每个的全部内容通过引用并入本文。格氏嗜盐碱杆菌Argonaute的序列提供于SEQ ID NO:718中。

[0065] 本文还提供了具有放松的PAM需求的Cas9变体(无PAM(PAMless)的Cas9)。与如SEQ ID NO:1提供的酿脓链球菌Cas9相比,无PAM的Cas9对靶序列表现出增加的活性,所述靶序列在其3'端处不包含规范的PAM(NGG),例如增加的活性为至少5倍、至少10倍、至少50倍、至少100倍、至少500倍、至少1,000倍、至少5,000倍、至少10,000倍、至少50,000倍、至少100,

000倍、至少500,000倍或至少1,000,000倍。因此，本公开的dCas9或Cas9切口酶可以进一步包含放松PAM需求的突变，例如，对应于SEQ ID NO:1中A262T、K294R、S409I、E480K、E543D、M694I或E1219V的突变。

[0066] 应当理解，另外的Cas9蛋白(例如，核酸酶死亡的Cas9(dCas9)、Cas9切口酶(nCas9)或核酸酶活性Cas9)，包括其变体和同源物，都在本公开的范围内。示例性Cas9蛋白包括但不限于下文提供的那些。在一些实施方案中，Cas9蛋白是核酸酶死亡的Cas9(dCas9)。在一些实施方案中，dCas9包含如下所示的氨基酸序列。在一些实施方案中，Cas9蛋白是Cas9切口酶(nCas9)。在一些实施方案中，nCas9包含如下所示的氨基酸序列。在一些实施方案中，Cas9蛋白是核酸酶活性Cas9。在一些实施方案中，核酸酶活性Cas9包含如下所示的氨基酸序列。

[0067] 示例性的催化无活性的Cas9(dCas9):DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:752)

[0068] 示例性的Cas9切口酶(nCas9):DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANL

AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVEN
TQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKM
KNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKV
ITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGK
ATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSK
ESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLE
AKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFV
EQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKR
YTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:753)

[0069] 示例性的催化活性Cas9:DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI
GALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVD
EVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENP
INASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD
NLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFF
DQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQED
FYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPN
EKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEI
SGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRY
TGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPA
IKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQN
EKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWR
QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
KLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKY
FFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP
KRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYK
EVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKH
YLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTK
EVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:754)

[0070] 在一些实施方案中,Cas9是指来自构成单细胞原核微生物的域和界的古生菌(arehaea)(例如纳古生菌(nanoarchaea))的Cas9。在一些实施方案中,Cas9是指CasX或CasY,其已经描述于例如Burstein et al.,“New CRISPR-Cas systems from uncultivated microbes.”Cell Res.2017Feb 21.doi:10.1038/cr.2017.21,其全部内容在此通过引用并入。使用基因组分辨的宏基因组学,鉴定了许多CRISPR-Cas系统,包括在生命的古生菌域中首次报告的Cas9。这种趋异的Cas9蛋白在研究很少的纳古生菌中作为活性CRISPR-Cas系统的一部分发现。在细菌中,发现了两个以前未知的系统,CRISPR-CasX和CRISPR-CasY,它们是迄今发现的最紧凑的系统之一。在一些实施方案中,Cas9是指CasX或CasX的变体。在一些实施方案中,Cas9是指CasY或CasY的变体。应当理解,其他RNA引导的DNA结合蛋白可以用作引导核苷酸序列-可编程DNA结合蛋白,并且在本公开的范围内。

[0071] 在一些实施方案中,本文提供的任何融合蛋白的引导核苷酸序列-可编程DNA结合

蛋白域可以是CasX或CasY蛋白。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白域是CasX蛋白。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白域是CasY蛋白。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白域包含与天然存在的CasX或CasY蛋白至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白域是天然存在的CasX或CasY蛋白。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白域包含与本文所述的示例性CasX或CasY蛋白的任一个至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白域包含本文所述的示例性CasX或CasY蛋白的任一个的氨基酸序列。应当理解，根据本公开也可以使用来自其他细菌物种的CasX和CasY。

[0072] CasX(uniprot.org/uniprot/F0NN87;uniprot.org/uniprot/F0NH53)

[0073] >tr|F0NN87|F0NN87_SULIH CRISPR相关的Casx蛋白OS＝冰岛硫化叶菌(Sulfolobus islandicus)(菌株HVE10/4)GN＝SiH_0402PE＝4SV＝1

[0074] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKG
EEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYEFGRSPGMVERT
RRVKLEVEPHYLIIAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVV
SSVTNPNVSVVRIYTISDAVGQNPTTINGGFSIDLTKLLEKRYLLSERLEAIARNALSISSNMRERYIVLANYIYE
YLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG(SEQ ID NO:755)

[0075] >tr|F0NH53|F0NH53_SULIR CRISPR相关的蛋白质,Casx OS＝冰岛硫化叶菌(菌株REY15A)GN＝SiRe_0771PE＝4SV＝1

[0076] MEVPLYNIFGDNYIIQVATEAENSTIYNNKVEIDDEELRNVLNLAYKIAKNNEDAAAERRGKAKKKKG
EEGETTTSNIILPLSGNDKNPWTETLKCYNFPTTVALSEVFKNFSQVKECEEVSAPSFVKPEFYKFGRSPGMVERT
RRVKLEVEPHYLIMAAAGWVLTRLGKAKVSEGDYVGVNVFTPTRGILYSLIQNVNGIVPGIKPETAFGLWIARKVV
SSVTNPNVSVVSIYTISDAVGQNPTTINGGFSIDLTKLLEKRDLLSERLEAIARNALSISSNMRERYIVLANYIYE
YLTGSKRLEDLLYFANRDLIMNLNSDDGKVRDLKLISAYVNGELIRGEG(SEQ ID NO:756)

[0077] CasY(ncbi.nlm.nih.gov/protein/APG80656.1)

[0078] >APG80656.1CRISPR相关的蛋白质CasY[未培养的Parcubacteria组细菌]MSKRHP
RISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYVGLYGLSNFDDLYNAEKRNEEKV
YSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYELTKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLK
KDIIDCFKAEYRERHKDQCNKLADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFD
TVNNNRNRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELKKAMMDITDAW
RGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNYINQTVKIKEDLKGHKKDLKKAKEMIN
RFGESDTKEEAVVSSLLESIEKIVPDDSADDEKPDIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKK
KKPKKRKKKSDAEDEKETIDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSL
KNSFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQSRSRKSAAIDKNRVR
LPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKTALALLLAVTETQLDISALDFVENGTVKDFM
KTRDGNLVLEGRFLEMFSQSIVFSELRGLAGLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAF
LDLAPAEFATSLEPESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKTLG

RGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTVALEPVSGSERVFVSQPFT IFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQNFISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIA RIRESLVHSLRNRIHHLALKHKAKIVYELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASE ISASYTSQFCGACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDFCDKHHI SKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKNIKVLGQMKKI(SEQ ID NO: 757)

[0079] 术语“缀合”是指两个实体的缔合，例如两个分子，诸如两个蛋白质、两个域(例如，结合域和切割域)，或蛋白质和试剂，例如蛋白质结合域和小分子的缔合。在一些方面，缔合在蛋白质(例如，RNA-可编程核酸酶)和核酸(例如，引导RNA)之间。缔合可以是例如经由直接或间接的(例如，经由接头)共价连接。在一些实施方案中，缔合是共价的。在一些实施方案中，两个分子经由连接两个分子的接头缀合。例如，在两个蛋白质彼此缀合(例如，经工程化的核酸酶的结合域和切割域)以形成蛋白质融合物的一些实施方案中，这两个蛋白质可以经由多肽接头(例如，将一个蛋白质的C端连接到另一个蛋白质的N端的氨基酸序列)缀合。

[0080] 如本文在核酸序列的背景下使用的术语“共有序列”是指表示在多个相似序列中的每个位置处发现的最频繁的核苷酸残基的经计算的序列。通常，通过序列比对确定共有序列，其中相似的序列彼此比较并计算相似的序列基序。在重组酶靶位点序列的背景下，在一些实施方案中，重组酶靶位点的共有序列可以是通过给定的重组酶最频繁地结合或以最高亲和力结合的序列。

[0081] 如本文所用的术语“工程化的”是指已人为设计、生产、制备、合成和/或制造的蛋白质分子、核酸、复合物、物质或实体。因此，工程化的产品是天然不存在的产品。

[0082] 如本文所用，术语“有效量”是指足以引起期望的生物学反应的生物活性剂的量。在一些实施方案中，重组酶的有效量可以指足以诱导在由重组酶特异性结合并重组的靶位点处的重组的重组酶的量。如熟练技术人员将理解的，试剂，例如核酸酶、重组酶、杂合蛋白、融合蛋白、蛋白质二聚体、蛋白质(或蛋白质二聚体)和多核苷酸的复合物，或多核苷酸的有效量可以随各种因素而变化，诸如例如随期望的生物学反应、特定等位基因、基因组、靶位点、靶定的细胞或组织和使用的试剂而变化。

[0083] 如本文所用，“引导核苷酸序列-可编程DNA结合蛋白”是指能够结合DNA的蛋白质、多肽或域，并且与其靶DNA序列的结合由引导核苷酸序列介导。“引导核苷酸”可以是RNA或DNA分子(例如，单链DNA或ssDNA分子)，其与靶序列互补并且可以将DNA结合蛋白引导至靶序列。因此，引导核苷酸序列-可编程DNA结合蛋白可以是RNA-可编程DNA结合蛋白，或ssDNA-可编程DNA结合蛋白。“可编程”意指DNA结合蛋白可以被编程以结合引导核苷酸靶标的任何DNA序列。本文提及的引导核苷酸序列-可编程DNA结合蛋白可以是本领域已知的任何引导核苷酸序列-可编程DNA结合蛋白而没有限制，其包括但不限于，二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白。术语“环状置换”是指其中蛋白质中氨基酸的顺序已经改变，导致具有改变的连接性(connectivity)但相似的(整体)三维形状的蛋白质结构的蛋白质。当原始的n和c末端氨基酸经由肽键连接时形成环状置换；然后肽序列在肽序列内的另一个位置中断裂，产生新的n和c端。环状置换可以通过许多过程发生，包括演化事件、翻译后修饰或人工工程化突变。例如，环状置换可以用于改善蛋白质的催化活性或热稳

定性。环状置换的引导核苷酸序列-可编程DNA结合蛋白可与本文所述的任何实施方案一起使用。术语“二分叉”通常是指分成两部分的单体蛋白质。通常，两个部分都是单体蛋白质的功能所需要的。二分叉的蛋白质可以或不可以自身二聚化以重建功能性蛋白质。二分叉可以通过许多过程发生，包括演化事件、翻译后修饰或人工工程化突变。当与二分叉域融合时，其他蛋白质域可以用于强制二分叉蛋白质的重新组装。在一些情况下，其相互作用依赖于小分子的蛋白质域可以与每个二分叉域融合，导致二分叉蛋白质的小分子调节的二聚化。

[0084] 如本文所用，术语“同源的”是本领域理解的术语，其是指在核苷酸和/或氨基酸序列的水平上高度相关的核酸或多肽。彼此同源的核酸或多肽称为“同源物”。两个序列之间的同源性可以通过本领域技术人员已知的序列比对方法确定。根据本发明，若两个序列对于至少20、至少30、至少40、至少50、至少60、至少70、至少80、至少90、至少100、至少120、至少150或至少200个氨基酸的至少一段(stretch)，至少约50-60％相同，例如在一个或另一个序列中包含的所有残基的至少约50-60％中共享相同的残基(例如氨基酸残基)、至少约70％相同、至少约80％相同、至少约85％相同、至少约90％相同、至少约95％相同、至少约98％相同、至少约99％相同、至少约99.5％相同或至少约99.9％相同，则认为它们是同源的。

[0085] 如本文所用，术语“序列一致性”或“序列一致性百分比”可以分别指给定DNA或蛋白质中与参照序列相同的核酸或氨基酸残基的百分比。参见例如:Christopher M.Holman, Protein Similarity Score:A Simplified Version of the BLAST Score as a Superior Alternative to Percent Identity for Claiming Genuses of Related Protein Sequences,21SANTA CLARA COMPUTER&HIGH TECH.L.J.55,60(2004)，以其全部内容通过引用并入本文。

[0086] 如本文所用，术语“接头”是指连接两个分子或部分，例如融合蛋白的两个域，诸如例如核酸酶无活性的Cas9域和核酸编辑域(例如腺苷脱氨酶)的键(例如共价键)、化学基团或分子。在一些实施方案中，接头连接RNA-可编程核酸酶的gRNA结合域，包括Cas9核酸酶域和核酸编辑蛋白的催化域。在一些实施方案中，接头连接dCas9和核酸编辑蛋白。通常，接头位于两个基团、分子或其他部分之间或侧翼有两个基团、分子或其他部分，并且经由共价键与每一个连接，从而连接两者。在一些实施方案中，接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施方案中，接头是有机分子、基团、聚合物或化学部分。在一些实施方案中，接头的长度为5-100个氨基酸，例如长度为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、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-150或150-200个氨基酸。也考虑了更长或更短的接头。在一些实施方案中，接头包含氨基酸序列SGSETPGTSESATPES(SEQ ID NO:7)，其也可以称为XTEN接头。在一些实施方案中，接头包含氨基酸序列SGGS(SEQ ID NO:758)。在一些实施方案中，接头包含$(SGGS)_n$(SEQ ID NO:758)、$(GGGS)_n$(SEQ ID NO:759)、$(GGGGS)_n$(SEQ ID NO:722)、$(G)_n$、$(EAAAK)_n$(SEQ ID NO:723)、$(GGS)_n$或$(XP)_n$基序，或这些中任何的组合，其中n独立地是1和30之间的整数，并且其中X是任何氨基酸。在一些实施方案中，n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。

[0087] 如本文所用，术语“突变”是指序列(例如核酸或氨基酸序列)内的残基用另一个残

基取代或序列内一个或多个残基的缺失或插入。本文通常通过鉴定初始残基，随后是序列内残基的位置和新取代的残基的身份来描述突变。用于产生本文提供的氨基酸取代(突变)的各种方法在本领域中是熟知的，并且由例如Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。

[0088] 术语“核定位序列”或“NLS”是指促进蛋白质输入细胞核(例如通过核转运)的氨基酸序列。核定位序列是本领域已知的，并且对于熟练技术人员是显而易见的。例如，NLS序列描述于2001年11月23日提交的Plank等人的国际PCT申请PCT/EP2000/011690，2001年5月31日公布为WO/2001/038547，其内容通过引用并入本文，用于其对示例性的核定位序列的公开内容。在一些实施方案中，NLS包含氨基酸序列PKKKRKV(SEQ ID NO:702)或MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(SEQ ID NO:761)。

[0089] 如本文所用，术语“核酸酶”是指能够切割连接核酸分子中两个核苷酸残基的磷酸二酯键的试剂，例如蛋白质。在一些实施方案中，“核酸酶”是指具有无活性的DNA切割域，使得核酸酶不能切割磷酸二酯键的蛋白质。在一些实施方案中，核酸酶是蛋白质，例如，可以结合核酸分子并切割连接核酸分子内的核苷酸残基的磷酸二酯键的酶。核酸酶可以是切割多核苷酸链内的磷酸二酯键的内切核酸酶，或切割在多核苷酸链的末端处的磷酸二酯键的外切核酸酶。在一些实施方案中，核酸酶是结合和/或切割特定核苷酸序列(其在本文中也称为“识别序列”、“核酸酶靶位点”或“靶位点”)内的特定磷酸二酯键的位点特异性核酸酶。在一些实施方案中，核酸酶是RNA引导的(即，RNA-可编程)核酸酶，其与具有与靶位点互补的序列的RNA(例如，引导RNA，“gRNA”)缔合(例如，结合)，从而提供核酸酶的序列特异性。在一些实施方案中，核酸酶识别单链靶位点，而在其他实施方案中，核酸酶识别双链靶位点，例如双链DNA靶位点。许多天然存在的核酸酶(例如许多天然存在的DNA限制性核酸酶)的靶位点是本领域技术人员熟知的。核酸酶蛋白质通常包含介导蛋白质与核酸底物相互作用并且在一些情况下，还特异性结合靶位点的“结合域”，以及催化核酸骨架内的磷酸二酯键的切割的“切割域”。在一些实施方案中，核酸酶蛋白质可以以单体形式结合和切割核酸分子，而在其他实施方案中，核酸酶蛋白质必须二聚化或多聚化以切割靶核酸分子。天然存在的核酸酶的结合域和切割域，以及可以融合以产生结合特定靶位点的核酸酶的模块结合域和切割域是本领域技术人员熟知的。例如，引导核苷酸序列-可编程DNA结合蛋白(诸如RNA-可编程核酸酶(例如Cas9))或具有无活性DNA切割域的Cas9蛋白的结合域可以用作结合域(例如，结合gRNA以直接结合靶位点)以特异性结合所期望的靶位点，并与切割域融合或缀合。

[0090] 如本文所用，术语“核酸”和“核酸分子”是指包含核碱基和酸性部分(例如核苷、核苷酸或核苷酸的聚合物)的化合物。通常，聚合核酸，例如包含三个或更多个核苷酸的核酸分子是线性分子，其中相邻的核苷酸经由磷酸二酯连接彼此连接。在一些实施方案中，“核酸”是指个别的核酸残基(例如核苷酸和/或核苷)。在一些实施方案中，“核酸”是指包含三个或更多个个别核苷酸残基的寡核苷酸链。如本文所用，术语“寡核苷酸”和“多核苷酸”可以可互换地使用以指核苷酸的聚合物(例如，至少三个核苷酸的串)。在一些实施方案中，“核酸”涵盖RNA以及单链和/或双链DNA。核酸可以是天然存在的，例如在基因组、转录物、mRNA、tRNA、rRNA、siRNA、snRNA、质粒、粘粒、染色体、染色单体或其他天然存在的核酸分子的背景下。另一方面，核酸分子可以是非天然存在的分子，例如重组DNA或RNA、人工染色体、

工程化的基因组或其片段，或合成的DNA、RNA、DNA/RNA杂交体，或包括非天然存在的核苷酸或核苷。此外，术语“核酸”、“DNA”、“RNA”和/或相似术语包括核酸类似物，即具有除磷酸二酯主链之外的类似物。核酸可以从天然来源纯化，使用重组表达系统产生并任选地纯化，化学合成等。在适当的情况下，例如在化学合成分子的情况下，核酸可以包含核苷类似物，例如具有化学修饰的碱基或糖、和主链修饰的类似物。除非另有说明，核酸序列以5’至3’方向呈现。在一些实施方案中，核酸是或包含天然核苷(例如腺苷、胸苷、鸟苷、胞苷、尿苷、脱氧腺苷、脱氧胸苷、脱氧鸟苷和脱氧胞苷)；核苷类似物(例如2-氨基腺苷、2-硫代胸苷、肌苷、吡咯并嘧啶、3-甲基腺苷、5-甲基胞苷、2-氨基腺苷、C5-溴尿苷、C5-氟尿苷、C5-碘尿苷、C5-丙炔基-尿苷、C5-丙炔基-胞苷、C5-甲基胞苷、2-氨基腺苷、7-脱氮腺苷、7-脱氮鸟苷、8-氧代腺苷、8-氧代鸟苷、O(6)-甲基鸟嘌呤和2-硫代胞苷)；化学修饰的碱基；生物修饰的碱基(例如甲基化碱基)；插入的碱基；修饰的糖(例如2’-氟核糖、核糖、2’-脱氧核糖、阿拉伯糖和己糖)；和/或修饰的磷酸基团(例如硫代磷酸酯和5’-N-亚磷酰胺连接)。

[0091] 术语“正交”是指最低限度相互作用(若有的话)的生物组分。若gRNA引导的recCas9蛋白不与其他潜在的重组酶位点相互作用或最低限度地相互作用，则含有不同gRNA结合位点的重组酶靶位点是正交的。术语“正交性”指的是系统组分可以独立变化而不影响其他组分的效能的构思。复合物的gRNA引导的性质使得与recCas9蛋白复合的该组gRNA分子能够仅在gRNA引导的位点处引导重组酶活性。通过该组gRNA分子对靶定重组酶位点上的酶活性的完全或近完全的依赖性来证明该系统的正交性。

[0092] 如本文所用，术语“药物组合物”是指可以在治疗和/或预防疾病或病症的背景下施用于受试者的组合物。在一些实施方案中，药物组合物包含活性成分，例如与Cas9蛋白融合的重组酶，或其片段(或编码此类融合物的核酸)和任选的药学上可接受的赋形剂。在一些实施方案中，药物组合物包含发明的Cas9变体/融合(例如fCas9)蛋白和适于将Cas9变体/融合蛋白靶向到靶核酸的gRNA。在一些实施方案中，靶核酸是基因。在一些实施方案中，靶核酸是与疾病相关的等位基因，其中等位基因通过Cas9变体/融合蛋白的作用而被切割。在一些实施方案中，等位基因是CLTA基因、VEGF基因、PCDH15基因或FAM19A2基因的等位基因。参见例如实施例。

[0093] 如本文所用，术语“增殖性疾病”是指其中细胞或组织稳态受到干扰，使得细胞或细胞群表现出异常升高的增殖速率的任何疾病。增殖性疾病包括过度增殖性疾病，如新生前期增生性状况和新生性疾病。新生性疾病的特征是细胞的异常增殖，并包括良性和恶性新生物两者。恶性新生物也称为癌症。在一些实施方案中，本文提供的组合物和方法可用于治疗增殖性疾病。例如，在一些实施方案中，药物组合物包含Cas9(例如，fCas9)蛋白和适合于将Cas9蛋白靶向到VEGF等位基因的gRNA，其中等位基因通过Cas9蛋白的作用而失活。参见例如实施例。

[0094] 术语“蛋白质”、“肽”和“多肽”在本文中可互换使用，并且是指通过肽(酰胺)键连接在一起的氨基酸残基的聚合物。该术语是指具有任何大小、结构或功能的蛋白质、肽或多肽。通常，蛋白质、肽或多肽将是至少三个氨基酸长。蛋白质、肽或多肽可以指个别的蛋白质或蛋白质的集合。蛋白质、肽或多肽中的一个或多个氨基酸可以被修饰，例如通过添加化学实体如碳水化合物基团、羟基、磷酸基团、法呢基、异法呢基、脂肪酸基团，用于缀合、官能化或其他修饰的接头等。蛋白质、肽或多肽也可以是单个分子或者可以是多分子复合物。蛋白

质、肽或多肽可以仅仅是天然存在的蛋白质或肽的片段。蛋白质、肽或多肽可以是天然存在的、重组的或合成的，或其任何组合。如本文所用，术语“融合蛋白”是指包含来自至少两种不同蛋白质的蛋白质域的杂合多肽。一种蛋白质可以位于融合蛋白的氨基端(N端)部分或羧基端(C端)蛋白质，从而分别形成“氨基端融合蛋白”或“羧基端融合蛋白”。本文提供的任何蛋白质可以通过本领域已知的任何方法产生。例如，本文提供的蛋白质可以经由重组蛋白质表达和纯化产生，其特别适用于包含肽接头的融合蛋白。用于重组蛋白质表达和纯化的方法是熟知的，并且包括Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))描述的那些，其全部内容通过引用并入本文。本文提及的特定融合蛋白是recCas9，一种通过将催化无活性的dCas9融合至重组酶的催化域创建的RNA编程的小丝氨酸重组酶，其能够在哺乳动物细胞中发挥功能。

[0095] 如本文所讨论的“假gix”位点或“gix假位点”是类似于Gix重组酶的天然DNA识别序列的特定的假回文核心DNA序列。参见例如N.D.F.Grindley,K.L.Whiteson,P.A.Rice,Mechanisms of site-specific recombination.Annu Rev Biochem 75,567-605(2006)，以其全部内容通过引用并入本文。类似地，“假hix”或“hix假位点”；“假six”或“six假位点”；“假resH”或“resH假位点”；“假res”或“res假位点”；“假LoxP”或“LoxP假位点”；“假att”或“att假位点”；“假FTR”或“FTR假位点”是类似于Hin重组酶、β重组酶、Sin重组酶、Tn3或γδ重组酶、Cre重组酶、λ噬菌体整合酶或FLP重组酶的天然DNA识别序列的特定的假回文核心DNA序列。

[0096] 术语“RNA-可编程核酸酶”和“RNA引导的核酸酶”在本文中可互换使用，并且是指与一个或多个不是切割靶标的RNA形成复合物(例如，结合或缔合)的核酸酶。在一些实施方案中，当与RNA形成复合物时，RNA-可编程核酸酶可以称为核酸酶:RNA复合物。通常，结合的RNA称为引导RNA(gRNA)。gRNA可以作为两个或更多个RNA的复合物或者作为单个RNA分子存在。作为单个RNA分子存在的gRNA可以称为单引导RNA(sgRNA)，尽管“gRNA”可互换使用以指作为单个分子或作为两个或更多个分子的复合物存在的引导RNA。通常，作为单一RNA种类存在的gRNA包含两个域：(1)与靶核酸共享同源性(例如，并引导Cas9复合物与靶物的结合)的域；和(2)结合Cas9蛋白的域。在一些实施方案中，域(2)对应已知为tracrRNA的序列，并且包含茎-环结构。例如，在一些实施方案中，域(2)与Jinek et al.,Science 337:816-821(2012)的图1E中描绘的tracrRNA同源，其全部内容通过引用并入本文。gRNA的其他实例(例如包括域2的那些)可以在2013年9月6日提交的题为“Switchable Cas9Nucleases and Uses Thereof”的美国临时专利申请U.S.S.N.61/874,682；2013年9月6号提交的题为“Delivery System For Functional Nucleases”的美国临时专利申请U.S.S.N.61/874,746；2013年3月15日提交的题为“Methods and Compositions for RNA-Directed Target DNA Modification and for RNA-Directed Modulation of Transcription”的PCT申请WO 2013/176722；和2013年3月20日提交的题为“RNA-Directed DNA Cleavage by the Cas9-crRNA Complex”的PCT申请WO 2013/142578中找到，每篇的全部内容通过引用以其整体并入本文。本文提供了gRNA的其他实例。参见例如实施例。在一些实施方案中，gRNA包含域(1)和(2)中的两个或更多个，并且可以称为“延伸的gRNA”。例如，延伸的gRNA将例如结合两个或更多个Cas9蛋白并在两个或更多个不同区域处结合靶核酸，如本文所述。gRNA包含与靶

位点互补的核苷酸序列,其介导核酸酶/RNA复合物与所述靶位点的结合,提供了核酸酶:RNA复合物的序列特异性。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白是RNA-可编程核酸酶如(CRISPR相关系统)(CRISPR-associated system)Cas9内切核酸酶,例如来自酿脓链球菌的Cas9(Csn1)(参见例如"Complete genome sequence of an M1strain of Streptococcus pyogenes."Ferretti J.J.,McShan W.M.,Ajdic D.J.,Savic D.J.,Savic G.,Lyon K.,Primeaux C.,Sezate S.,Suvorov A.N.,Kenton S.,Lai H.S.,Lin S.P.,Qian Y.,Jia H.G.,Najar F.Z.,Ren Q.,Zhu H.,Song L.,White J.,Yuan X.,Clifton S.W.,Roe B.A.,McLaughlin R.E.,Proc.Natl.Acad.Sci.U.S.A.98:4658-4663(2001);"CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III."Deltcheva E.,Chylinski K.,Sharma C.M.,Gonzales K.,Chao Y.,Pirzada Z.A.,Eckert M.R.,Vogel J.,Charpentier E.,Nature 471:602-607(2011);和"A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity."Jinek M.,Chylinski K.,Fonfara I.,Hauer M.,Doudna J.A.,Charpentier E.Science 337:816-821(2012),每篇的全部内容通过引用并入本文。

[0097] 因为RNA-可编程核酸酶(例如Cas9)使用RNA:DNA杂交来确定靶DNA切割位点,所以这些蛋白质原则上能够切割由引导RNA规定的任何序列。使用RNA-可编程核酸酶例如Cas9进行位点特异性切割(例如,以修饰基因组)的方法是本领域已知的(参见例如Cong,L.et al.Multiplex genome engineering using CRISPR/Cas systems.Science 339,819-823(2013);Mali,P.et al.RNA-guided human genome engineering via Cas9.Science 339,823-826(2013);Hwang,W.Y.et al.Efficient genome editing in zebrafish using a CRISPR-Cas system.Nature biotechnology 31,227-229(2013);Jinek,M.et al.RNA-programmed genome editing in human cells.eLife 2,e00471(2013);Dicarlo,J.E.et al.Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.Nucleic acids research(2013);Jiang,W.et al.RNA-guided editing of bacterial genomes using CRISPR-Cas systems.Nature biotechnology 31,233-239(2013);每篇的全部内容通过引用并入本文)。

[0098] 如本文所用,术语"重组酶"是指介导重组酶识别序列之间DNA的重组的位点特异性酶,所述重组导致重组酶识别序列之间的DNA片段的切除、整合、倒位或交换(例如,易位)。重组酶可以分为两个不同的家族:丝氨酸重组酶(例如,解离酶和转化酶)和酪氨酸重组酶(例如整合酶)。丝氨酸重组酶的实例包括但不限于,Hin、Gin、Tn3、β-six、CinH、ParA、γδ、Bxb1、ΦC31、TP901、TG1、φBT1、R4、φRV1、φFC1、MR11、A118、U153和gp29。酪氨酸重组酶的实例包括但不限于,Cre、FLP、R、Lambda、HK101、HK022和pSAM2。本文提及的Gin重组酶可以是本领域已知的任何Gin重组酶,其包括但不限于T.Gaj et al.,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic Acids Research 41,3937-3946(2013)中提出的Gin重组酶,其通过引用以其整体并入本文。在某些实施方案中,Gin重组酶催化域与SEQ ID NO:713中所示的氨基酸序列具有大于85%、90%、95%、98%或99%的序列一致性。在另一个实施方案中,Gin重组酶催化域的氨基酸序列包含对应于H106Y,和/或I127L,和/或I136R和/或G137F的突变。在另一个实施方案中,Gin重组酶催化域的氨基酸序列包含对

应于H106Y、I127L、I136R和G137F的突变。在进一步的实施方案中,Gin重组酶的氨基酸序列已得以进一步突变。在具体的实施方案中,Gin重组酶催化域的氨基酸序列包含SEQ ID NO:713。Gin重组酶与gix靶位点(本文也称为“gix核心”、“最小gix核心”或“gix相关核心”序列)结合。最小gix核心重组酶位点是NNNNAAASSWWSSTTTNNNN(SEQ ID NO:19),其中N定义为任何氨基酸,W是A或T,并且S是G或C。gix靶位点可以包括本领域已知的任何其他突变。在某些实施方案中,gix靶位点与SEQ ID NO:19中所示的氨基酸序列具有大于90%、95%或99%的序列一致性。gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离可以是1至10个碱基对、3至7个碱基对、5至7个碱基对或5至6个碱基对。gix核心或gix相关核心序列与至少一个gRNA结合位点之间的距离可以是1、2、3、4、5、6、7、8、9或10个碱基对。

[0099] 丝氨酸和酪氨酸重组酶的名称源自重组酶用来攻击DNA,并在链交换期间与DNA共价连接的保守亲核氨基酸残基。重组酶具有许多应用,包括基因敲除/敲入的创建和基因治疗应用。参见例如Brown et al.,“Serine recombinases as tools for genome engineering.”Methods.2011;53(4):372-9;Hirano et al.,“Site-specific recombinases as tools for heterologous gene integration.”Appl.Microbiol.Biotechnol.2011;92(2):227-39;Chavez and Calos,“Therapeutic applications of theΦC31integrase system.”Curr.Gene Ther.2011;11(5):375-81;Turan and Bode,“Site-specific recombinases:from tag-and-target-to tag-and-exchange-based genomic modifications.”FASEB J.2011;25(12):4088-107;Venken and Bellen,“Genome-wide manipulations of Drosophila melanogaster with transposons,Flp recombinase,andΦC31integrase.”Methods Mol.Biol.2012;859:203-28;Murphy,“Phage recombinases and their applications.”Adv.Virus Res.2012;83:367-414;Zhang et al.,“Conditional gene manipulation:Creating a new biological era.”J.Zhejiang Univ.Sci.B.2012;13(7):511-24;Karpenshif and Bernstein,“From yeast to mammals:recent advances in genetic control of homologous recombination.”DNA Repair(Amst).2012;1;11(10):781-8;每篇的全部内容以其整体在此通过引用并入。本文提供的重组酶不意味着是可以用于本发明的实施方案的重组酶的排他性实例。通过挖掘新的正交重组酶的数据库或设计具有限定的DNA特异性的合成重组酶,可以扩展本发明的方法和组合物(参见例如,Groth et al.,“Phage integrases:biology and applications.”J.Mol.Biol.2004;335,667-678;Gordley et al.,“Synthesis of programmable integrases.”Proc.Natl.Acad.Sci.U S A.2009;106,5053-5058;每篇的全部内容以其整体在此通过引用并入)。

[0100] 可用于本文所述方法和组合物的重组酶的其他实例是本领域技术人员已知的,并且预期发现或产生的任何新重组酶能够用于本发明的不同实施方案中。在一些实施方案中,重组酶的催化域与核酸酶失活的RNA-可编程核酸酶(例如,dCas9或其片段)融合,使得重组酶域不包含核酸结合域或不能结合随后导致酶促催化的靶核酸(例如,重组酶域经工程化改造以使其不具有特异性DNA结合活性)。缺乏DNA结合活性的部分的重组酶和独立于辅助蛋白起作用的重组酶以及其工程化方法是已知的,并且包括Klippel et al.,“Isolation and characterisation of unusual gin mutants.”EMBO J.1988;7:3983-3989:Burke et al.,“Activating mutations of Tn3resolvase marking interfaces

important in recombination catalysis and its regulation.Mol Microbiol.2004;51:937-948;Olorunniji et al.,“Synapsis and catalysis by activated Tn3resolvase mutants.”Nucleic Acids Res.2008;36:7181-7191;Rowland et al.,“Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome.”Mol Microbiol.2009;74:282-298;Akopian et al.,“Chimeric recombinases with designed DNA sequence recognition.”Proc Natl Acad Sci USA.2003;100:8688-8691;Gordley et al.,“Evolution of programmable zinc finger-recombinases with activity in human cells.J Mol Biol.2007;367:802-813;Gordley et al.,“Synthesis of programmable integrases.”Proc Natl Acad Sci USA.2009;106:5053-5058;Arnold et al.,“Mutants of Tn3resolvase which do not require accessory binding sites for recombination activity.”EMBO J.1999;18:1407-1414;Gaj et al.,“Structure-guided reprogramming of serine recombinase DNA sequence specificity.”Proc Natl Acad Sci USA.2011;108(2):498-503;和Proudfoot et al.,“Zinc finger recombinases with adaptable DNA sequence specificity.”PLoS One.2011;6(4):e19537描述的那些;每篇的全部内容在此通过引用并入。例如,解离酶-转化酶组的丝氨酸重组酶(例如Tn3和γδ解离酶以及Hin和Gin转化酶)具有模块化结构,它们具有部分自主的催化和DNA结合域(参见例如,Grindley et al.,“Mechanism of site-specific recombination.”Ann Rev Biochem.2006;75:567-605,其全部内容通过引用并入)。因此,这些重组酶的催化域适合于与如本文所述的核酸酶失活的RNA-可编程核酸酶(例如,dCas9或其片段)重组,例如,在分离不需要任何辅助因素(例如DNA结合活性)的“激活性”重组酶突变体之后(参见例如,Klippel et al.,“Isolation and characterisation of unusual gin mutants.”EMBO J.1988;7:3983-3989:Burke et al.,“Activating mutations of Tn3resolvase marking interfaces important in recombination catalysis and its regulation.Mol Microbiol.2004;51:937-948;Olorunniji et al.,“Synapsis and catalysis by activated Tn3resolvase mutants.”Nucleic Acids Res.2008;36:7181-7191;Rowland et al.,“Regulatory mutations in Sin recombinase support a structure-based model of the synaptosome.”Mol Microbiol.2009;74:282-298;Akopian et al.,“Chimeric recombinases with designed DNA sequence recognition.”Proc Natl Acad Sci USA.2003;100:8688-8691)。

[0101] 另外,许多其他具有N端催化域和C端DNA结合域的天然丝氨酸重组酶是已知的(例如,phiC31整合酶、TnpX转座酶、IS607转座酶),并且它们的催化域可以增选用于工程化如本文所述的可编程位点特异性重组酶(参见例如,Smith et al.,“Diversity in the serine recombinases.”Mol Microbiol.2002;44:299-307,其全部内容通过引用并入)。类似地,酪氨酸重组酶(例如Cre、λ整合酶)的核心催化域是已知的,并且可以类似地增选以工程化如本文所述的可编程位点特异性重组酶(参见例如Guo et al.,“Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse.”Nature.1997;389:40-46;Hartung et al.,“Cre mutants with altered DNA binding properties.”J Biol Chem 1998;273:22884-22891;Shaikh et al.,“Chimeras of the Flp and Cre recombinases:Tests of the mode of cleavage by Flp and Cre.J Mol

Biol.2000;302:27-48;Rongrong et al.,“Effect of deletion mutation on the recombination activity of Cre recombinase.”Acta Biochim Pol.2005;52:541-544;Kilbride et al.,“Determinants of product topology in a hybrid Cre-Tn3resolvase site-specific recombination system.”J Mol Biol.2006;355:185-195;Warren et al.,“A chimeric cre recombinase with regulated directionality.”Proc Natl Acad Sci USA.2008 105:18278-18283;Van Duyne,“Teaching Cre to follow directions.”Proc Natl Acad Sci USA.2009Jan 6;106(1):4-5;Numrych et al.,“A comparison of the effects of single-base and triple-base changes in the integrase arm-type binding sites on the site-specific recombination of bacteriophageλ.”Nucleic Acids Res.1990;18:3953-3959;Tirumalai et al.,“The recognition of core-type DNA sites byλintegrase.”J Mol Biol.1998;279:513-527;Aihara et al.,“A conformational switch controls the DNA cleavage activity ofλ integrase.”Mol Cell.2003;12:187-198;Biswas et al.,“A structural basis for allosteric control of DNA recombination byλintegrase.”Nature.2005;435:1059-1066;和Warren et al.,“Mutations in the amino-terminal domain ofλ-integrase have differential effects on integrative and excisive recombination.”Mol Microbiol.2005;55:1104-1112;每篇的全部内容通过引用并入)。

[0102] 在核酸修饰(例如,基因组修饰)的背景下,术语“重组”用于指通过重组酶蛋白质(例如,本文提供的本发明的重组酶融合蛋白)的作用修饰两个或更多个核酸分子或单个核酸分子的两个或更多个区域的过程。重组尤其可以导致核酸的插入、倒位、切除或易位,例如在一个或多个核酸分子之中或之间。

[0103] 如本文中在蛋白质或核酸的背景中使用,术语“重组体(recombinant)”是指自然界中不存在,但是作为人工程化的产物的蛋白质或核酸。例如,在一些实施方案中,重组蛋白质或核酸分子包含相比于任何天然存在的序列包含至少一个、至少两个、至少三个、至少四个、至少五个、至少六个或至少七个突变的氨基酸或核苷酸序列。

[0104] 如本文所用,术语“受试者”是指个体生物体,例如个体哺乳动物。在一些实施方案中,受试者是人。在一些实施方案中,受试者是非人哺乳动物。在一些实施方案中,受试者是非人灵长类动物。在一些实施方案中,受试者是啮齿动物。在一些实施方案中,受试者是绵羊、山羊、牛、猫或狗。在一些实施方案中,受试者是脊椎动物、两栖动物、爬行动物、鱼、昆虫、苍蝇或线虫。在一些实施方案中,受试者是研究动物。在一些实施方案中,受试者是经遗传工程化的,例如基因遗传化的非人受试者。受试者可以是任何一个性别和处于任何发展阶段的。在一些实施方案中,受试者是经遗传工程化的,例如基因遗传化的非人受试者。受试者可以是任何一个性别和处于任何发展阶段的。

[0105] 如本文在核酸酶的背景中使用,术语“靶核酸”和“靶基因组”分别指包含给定核酸酶的至少一个靶位点的核酸分子或基因组。在包含(核酸酶失活的)RNA-可编程核酸酶和重组酶域的融合物的背景中,“靶核酸”和“靶基因组”分别指一个或多个包含至少一个靶位点的核酸分子或基因组。在一些实施方案中,靶核酸包含至少两个、至少三个、至少四个、至少五个、至少六个、至少七个或至少八个靶位点。在一些实施方案中,靶核酸包含四个靶位点。

[0106] 术语“靶位点”是指由重组酶(例如本文提供的dCas9-重组酶融合蛋白)结合并重

组(例如在靶位点处或附近)的核酸分子内的序列。靶位点可以是单链或双链的。例如,在一些实施方案中,四个重组酶单体协调以重组靶核酸,每个单体与由gRNA引导的(核酸酶失活的)Cas9蛋白融合。在此类实例中,每个Cas9域由不同的gRNA引导以结合靶核酸,因此靶核酸包含四个靶位点,每个位点通过分开的dCas9-重组酶融合物靶向(从而协调重组靶核酸的四个重组酶单体)。对于RNA引导的核酸酶失活的Cas9(或其gRNA结合域)和本发明的Cas9的融合物,在一些实施方案中,靶位点可以是17-20个碱基对加上3个碱基对PAM(例如,NNN,其中N独立地代表任何核苷酸)。通常,PAM的第一个核苷酸可以是任何核苷酸,而两个下游核苷酸是根据特定的RNA引导的核酸酶指定的。用于RNA引导的核酸酶(例如Cas9)的示例性靶位点(例如,包含PAM)是本领域技术人员已知的并且包括但不限于NNG、NGN、NAG和NGG,其中每个N独立地是任何核苷酸。另外,来自不同物种(例如嗜热链球菌而不是酿脓链球菌)的Cas9核酸酶识别包含序列NGGNG(SEQ ID NO:763)的PAM。另外的PAM序列是已知的,包括但不限于NNAGAAW(SEQ ID NO:749)和NAAR(SEQ ID NO:771)(参见例如,Esvelt and Wang, Molecular Systems Biology,9:641(2013),其全部内容通过引用并入本文)。在一些方面,RNA引导的核酸酶(诸如例如Cas9)的靶位点可以包含结构[N_Z]-[PAM],其中每个N独立地是任何核苷酸,并且z是1和50之间的整数,包括端点。在一些实施方案中,z是至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少25、至少30、至少35、至少40、至少45或至少50。在一些实施方案中,z是5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50。在一些实施方案中,z是20。在某些实施方案中,可以使用“无PAM的”RNA引导的核酸酶(例如,无PAM的Cas9)或如本文进一步描述的具有放松PAM需求的RNA引导的核酸酶。在一些实施方案中,“靶位点”还可以指核酸分子内被核酸酶结合但未切割的序列。例如,本文描述的某些实施方案提供包含无活性(或失活)Cas9DNA切割域的蛋白质。此类蛋白质(例如,当还包括Cas9RNA结合域时)能够结合由gRNA指定的靶位点;然而,因为DNA切割位点被灭活,所以靶位点不被特定蛋白质切割。在一些实施方案中,此类蛋白质与重组酶(或重组酶的催化域)缀合、融合或结合,所述重组酶介导靶核酸的重组。在一些实施方案中,实际上切割或重组的序列将取决于介导核酸分子的切割或重组的蛋白质(例如,重组酶)或分子,并且在一些情况下,例如可与结合失活的Cas9蛋白的接近性或距离有关。

[0107] 如本文所用,术语“转录激活物样效应物”(TALE)是指包含DNA结合域的细菌蛋白质,所述DNA结合域含有包含高度可变的双氨基酸基序(重复可变双残基,RVD)的高度保守的33-34个氨基酸的序列。RVD基序决定了对核酸序列的结合特异性,并且可以根据本领域技术人员已知的方法进行工程化以特异性结合所期望的DNA序列(参见例如,Miller, Jeffrey;et.al.(February 2011).“A TALE nuclease architecture for efficient genome editing”.Nature Biotechnology 29(2):143-8;Zhang,Feng;et.al.(February 2011).“Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription”Nature Biotechnology 29(2):149-53;Geiβler, R.;Scholze,H.;Hahn,S.;Streubel,J.;Bonas,U.;Behrens,S.E.;Boch,J.(2011),Shiu, Shin-Han.ed.“Transcriptional Activators of Human Genes with Programmable DNA-Specificity”.PLoS ONE 6(5):e19509;Boch,Jens(February 2011).“TALEs of genome

targeting".Nature Biotechnology 29(2):135-6;Boch,Jens;et.al.(December 2009)."Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors".Science 326(5959):1509-12;和Moscou,Matthew J.;Adam J.Bogdanove(December 2009)."A Simple Cipher Governs DNA Recognition by TAL Effectors"Science 326(5959):1501;其每篇的全部内容通过引用并入本文)。氨基酸序列和DNA识别之间的简单关系允许通过选择含有适当RVD的重复区段的组合来工程化改造特定的DNA结合域。

[0108] 如本文所用,术语"转录激活物样元件核酸酶"(TALEN)是指包含转录激活物样效应物DNA结合域至DNA切割域,例如FokI域的人工核酸酶。已经报告了许多用于产生工程化TALE构建体的模块化组装方案(参见例如,Zhang,Feng;et.al.(February 2011)."Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription".Nature Biotechnology 29(2):149-53;Geiβler,R.;Scholze,H.;Hahn,S.;Streubel,J.;Bonas,U.;Behrens,S.E.;Boch,J.(2011),Shiu,Shin-Han.ed."Transcriptional Activators of Human Genes with Programmable DNA-Specificity".PLoS ONE 6(5):e19509;Cermak,T.;Doyle,E.L.;Christian,M.;Wang,L.;Zhang,Y.;Schmidt,C.;Baller,J.A.;Somia,N.V.et al.(2011)."Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting".Nucleic Acids Research;Morbitzer,R.;Elsaesser,J.;Hausner,J.;Lahaye,T.(2011)."Assembly of custom TALE-type DNA binding domains by modular cloning".Nucleic Acids Research;Li,T.;Huang,S.;Zhao,X.;Wright,D.A.;Carpenter,S.;Spalding,M.H.;Weeks,D.P.;Yang,B.(2011)."Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes".Nucleic Acids Research.;Weber,E.;Gruetzner,R.;Werner,S.;Engler,C.;Marillonnet,S.(2011).Bendahmane,Mohammed.ed."Assembly of Designer TAL Effectors by Golden Gate Cloning".PLoS ONE 6(5):e19722;其每篇的全部内容通过引用并入本文)。

[0109] 术语"治疗/处理"是指如本文所述旨在逆转、缓解疾病或病症或其一种或多种症状、延迟疾病或病症或其一种或多种症状的发作或抑制疾病或病症或其一种或多种症状进展的临床干预。如本文所用,术语"治疗/处理"是指如本文所述旨在逆转、缓解疾病或病症或其一种或多种症状、延迟疾病或病症或其一种或多种症状的发作或抑制疾病或病症或其一种或多种症状进展的临床干预。在一些实施方案中,可以在一种或多种症状已经得以形成之后和/或疾病已经得到诊断之后施用治疗。在其他实施方案中,可以在没有症状的情况下施用治疗,例如用于预防或延迟症状的发作或抑制疾病的发作或进展。例如,可以在症状发作之前(例如,鉴于症状的历史和/或鉴于遗传或其他易感性因素)施用治疗于易感个体。治疗也可以在症状消退后继续进行,例如以预防或延迟其复发。

[0110] 术语"载体"是指包含一种或多种本发明的重组多核苷酸的多核苷酸,例如编码本文提供的Cas9蛋白(或其融合物)和/或gRNA的那些。载体包括但不限于质粒、病毒载体、粘粒、人工染色体和噬菌粒。载体可以能够在宿主细胞中复制,并且可以进一步通过一个或多个内切核酸酶限制性位点来表征,在该位点处可以切割载体并且可以在其中插入所期望的核酸序列。载体可以含有一种或多种适用于鉴定和/或选择已经用或尚未用载体转化或用

载体进行基因组修饰的细胞的标志物序列。标志物包括例如编码增加或降低对抗生素(例如,卡那霉素、氨苄青霉素)或其他化合物的抗性或敏感性的蛋白质的基因,编码其活性可通过本领域已知的标准测定法检测的酶(例如,β-半乳糖苷酶、碱性磷酸酶或萤光素酶)的基因,以及可见地影响转化或转染细胞、宿主、菌落或噬斑的表型的基因。适用于转化本发明所囊括的宿主细胞(例如大肠杆菌、哺乳动物细胞如CHO细胞、昆虫细胞等)的任何载体,例如属于pUC系列、pGEM系列、pET系列、pBAD系列、pTET系列或pGEX系列的载体。在一些实施方案中,载体适合于转化宿主细胞用于重组蛋白质生产。用于选择和工程化用于表达蛋白质(例如,本文提供的蛋白质)的载体和宿主细胞、转化细胞和表达/纯化重组蛋白质的方法是本领域熟知的,并且由例如Green and Sambrook,Molecular Cloning:A Laboratory Manual(4th ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.(2012))提供。

[0111] 如本文所用,术语“锌指”是指以折叠和一个或多个稳定化折叠的锌离子的配位为特征的小核酸结合蛋白结构基序。锌指涵盖极其多种不同的蛋白质结构(参见例如Klug A,Rhodes D(1987).“Zinc fingers:a novel protein fold for nucleic acid recognition”.Cold Spring Harb.Symp.Quant.Biol.52:473-82,其全部内容通过引用并入本文)。可以设计锌指以结合核苷酸的特定序列,并且可以设计包含一系列锌指的融合物的锌指阵列以结合实际上任何所期望的靶序列。此类锌指阵列可以形成蛋白质(例如核酸酶)的结合域,例如,在与核酸切割域缀合的情况下。不同类型的锌指基序是本领域技术人员已知的,包括但不限于Cys_2His_2、Gag knuckle、Treble clef、锌带(Zinc ribbon)、Zn_2/Cys_6和TAZ2域样基序(参见例如Krishna SS,Majumdar I,Grishin NV(January 2003).“Structural classification of zinc fingers:survey and summary”.Nucleic Acids Res.31(2):532-50)。通常,单一锌指基序结合核酸分子的3或4个核苷酸。因此,包含2个锌指基序的锌指域可以结合6-8个核苷酸,包含3个锌指基序的锌指域可以结合9-12个核苷酸,包含4个锌指基序的锌指域可以结合12-16个核苷酸,等等。可以采用任何合适的蛋白质工程化技术来改变锌指的DNA结合特异性和/或设计新的锌指融合物以结合长度为3至30个核苷酸的实际任何所期望的靶序列(参见例如Pabo CO,Peisach E,Grant RA(2001).“Design and selection of novel cys2His2Zinc finger proteins”.Annual Review of Biochemistry 70:313-340;Jamieson AC,Miller JC,Pabo CO(2003).“Drug discovery with engineered zinc-finger proteins”.Nature Reviews Drug Discovery 2(5):361-368;和Liu Q,Segal DJ,Ghiara JB,Barbas CF(May 1997).“Design of polydactyl zinc-finger proteins for unique addressing within complex genomes”.Proc.Natl.Acad.Sci.U.S.A.94(11);其每篇的全部内容通过引用并入本文)。工程化锌指阵列和切割核酸的蛋白域之间的融合可以用于产生“锌指核酸酶”。锌指核酸酶通常包含结合核酸分子内特定靶位点的锌指域和切割由结合域结合的靶位点内或附近的核酸分子的核酸切割域。典型的工程化锌指核酸酶包含具有3和6个之间的个别锌指基序的结合域和长度范围为9个碱基对至18个碱基对的结合靶位点。在期望结合和切割给定基因组中独特的靶位点的情况下,较长的靶位点特别有吸引力。

[0112] 如本文所用,术语“锌指核酸酶”是指包含与包含锌指阵列的结合域缀合的核酸切割域的核酸酶。在一些实施方案中,切割域是II型限制性内切核酸酶FokI的切割域。可以设

计锌指核酸酶以靶向给定核酸分子中的实际任何所期望的序列用于切割，并且设计锌指结合域以在复杂基因组的背景下结合独特位点的可能性允许活细胞中靶向切割单个基因组位点，例如，以实现治疗价值的靶向基因组改变。由于非同源DNA修复途径的易错特性，将双链断裂靶向到期望的基因组基因座可以用于将移码突变引入基因的编码序列中。可以通过本领域技术人员熟知的方法产生锌指核酸酶以靶向感兴趣的位点。例如，可以通过组合已知特异性的个别锌指基序来设计具有所期望的特异性的锌指结合域。与DNA结合的锌指蛋白Zif268的结构已经为该领域的大量工作提供了信息，并且已经描述了为64个可能的碱基对三联体中的每一个获得锌指，然后混合和匹配这些模块化锌指以设计具有任何所期望的序列特异性的蛋白质的构思(Pavletich NP,Pabo CO(May 1991).“Zinc finger-DNA recognition:crystal structure of a Zif268-DNA complex at 2.1A”.Science 252(5007):809-17，其全部内容并入本文)。在一些实施方案中，将各自识别3碱基对DNA序列的单独的锌指组合以产生3-、4-、5-或6-指阵列，其识别长度范围为9个碱基对至18个碱基对的靶位点。在一些实施方案中，考虑了更长的阵列。在其他实施方案中，组合识别6-8个核苷酸的2指模块以产生4-、6-或8-锌指阵列。在一些实施方案中，采用细菌或噬菌体展示来开发识别所期望的核酸序列的锌指域，例如，长度为3-30bp的所期望的核酸酶靶位点。在一些实施方案中，锌指核酸酶包含经由接头(例如多肽接头)彼此融合或以其他方式彼此缀合的锌指结合域和切割域。接头的长度决定了切口与由锌指域结合的核酸序列的距离。若使用较短的接头，则切割域将使核酸切割得更接近结合的核酸序列，而较长的接头将导致切口和结合的核酸序列之间更大的距离。在一些实施方案中，锌指核酸酶的切割域必须二聚化以切割结合的核酸。在一些此类实施方案中，二聚体是两个单体的异源二聚体，每个单体包含不同的锌指结合域。例如，在一些实施方案中，二聚体可以包含一个包含与FokI切割域缀合的锌指域A的单体，和一个包含与FokI切割域缀合的锌指域B的单体。在该非限制性实例中，锌指域A结合靶位点的一侧上的核酸序列，锌指域B结合靶位点的另一侧上的核酸序列，并且二聚化FokI域切割锌指域结合位点之间的核酸。

[0113] 发明详述

[0114] 从以下实施例将更全面地理解本发明的这些和其他实施方案的功能和优点。以下实施例旨在说明本发明的益处并描述特定实施方案，但并不旨在举例说明本发明的全部范围。因此，应当理解，实施例不意味着限制本发明的范围。

[0115] 引导核苷酸序列-可编程DNA结合蛋白

[0116] 本文描述的融合蛋白和方法可以使用任何可编程的DNA结合域。

[0117] 在一些实施方案中，可编程DNA结合蛋白域包含锌指核酸酶(ZFN)或转录激活物样效应域(TALE)的DNA结合域。在一些实施方案中，可编程DNA结合蛋白域可以由引导核苷酸序列编程，并因此称为“引导核苷酸序列-可编程DNA结合蛋白域”。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白是核酸酶无活性的Cas9，或dCas9。如本文所用，dCas9涵盖Cas9，其在其核酸酶活性中完全无活性，或在其核酸酶活性中部分无活性(例如，Cas9切口酶)。因此，在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白是Cas9切口酶。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白是核酸酶无活性的Cpf1。在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白是核酸酶无活性的Argonaute。

[0118] 在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白是dCas9域。在一些实施

方案中，引导核苷酸序列-可编程DNA结合蛋白是Cas9切口酶。在一些实施方案中，dCas9域包含SEQ ID NO:2或SEQ ID NO:3的氨基酸序列。在一些实施方案中，dCas9域包含与本文提供的Cas9域的任一个至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列，并且包含对应于SEQ ID NO:1中D10X(X是除D以外的任何氨基酸)和/或H840X(X是除H以外的任何氨基酸)的突变。在一些实施方案中，dCas9域包含与本文提供的Cas9域的任一个至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列，并且包含对应于SEQ ID NO:1中D10A和/或H840A的突变。在一些实施方案中，Cas9切口酶包含与本文提供的Cas9域的任一个至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列，并且包含对应于SEQ ID NO:1中D10X(X是除以外的任何氨基酸D)的突变和对应于SEQ ID NO:1中位置840的位置处的组氨酸。在一些实施方案中，Cas9切口酶包含与本文提供的Cas9域的任一个至少60％、至少65％、至少70％、至少75％、至少80％、至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同，并且包含对应于SEQ ID NO:1中D10A的突变和对应于SEQ ID NO:1中位置840的位置处的组氨酸的氨基酸序列。在一些实施方案中，提供了dCas9或Cas9切口酶的变体或同源物(例如，分别是SEQ ID NO:2或SEQ ID NO:3的变体)，其分别与SEQ ID NO:2或SEQ ID NO:3至少约70％相同、至少约80％相同、至少约90％相同、至少约95％相同、至少约98％相同、至少约99％相同、至少约99.5％相同或至少约99.9％相同，并且包含对应于SEQ ID NO:1中D10A和/或H840A的突变。在一些实施方案中，提供了Cas9的变体(例如，SEQ ID NO:2的变体)，其具有比SEQ ID NO:2短或长约5个氨基酸、约10个氨基酸、约15个氨基酸、约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多个的氨基酸序列，条件是dCas9变体包含对应于SEQ ID NO:1中的D10A和/或H840A的突变。在一些实施方案中，提供了Cas9切口酶的变体(例如，SEQ ID NO:3的变体)，其具有比SEQ ID NO:3短或长约5个氨基酸、约10个氨基酸、约15个氨基酸、约20个氨基酸、约25个氨基酸、约30个氨基酸、约40个氨基酸、约50个氨基酸、约75个氨基酸、约100个氨基酸或更多个的氨基酸序列，条件是dCas9变体包含对应于D10A的突变并且包含对应于SEQ ID NO:1中位置840的位置处的组氨酸。

[0119] 基于本公开和本领域的知识，另外的合适的核酸酶无活性的dCas9域对于本领域技术人员而言将是显而易见的，并且在本公开的范围内。此类另外的示例性合适的核酸酶无活性的Cas9域包括但不限于SEQ ID NO:1中的D10A/H840A、D10A/D839A/H840A、D10A/D839A/H840A/N863A突变体域(参见例如，Prashant et al.,Nature Biotechnology.2013;31(9):833-838，其通过引用并入本文)，或K603R(参见例如，Chavez et al.,Nature Methods 12,326-328,2015，其通过引用并入本文)。

[0120] 在一些实施方案中，本文所述的核碱基编辑器包含与野生型Cas9域相比，具有降低的Cas9域与DNA的糖-磷酸主链之间的静电相互作用的Cas9域。在一些实施方案中，Cas9域包含降低Cas9域与DNA的糖-磷酸主链之间的缔合的一个或多个突变。在一些实施方案

中,本文所述的核碱基编辑器包含dCas9(例如,具有SEQ ID NO:1中D10A和H840A突变)或Cas9切口酶(例如,具有SEQ ID NO:1中D10A突变),其中dCas9或Cas9切口酶进一步包含SEQ ID NO:10中提供的氨基酸序列的N497X、R661X、Q695X和/或Q926X突变,或SEQ ID NO:11-260中提供的任何氨基酸序列中相应的突变的一个或多个,其中X是任何氨基酸。在一些实施方案中,本文所述的核碱基编辑器包含dCas9(例如,具有SEQ ID NO:1中D10A和H840A突变)或Cas9切口酶(例如,具有SEQ ID NO:1中D10A突变),其中dCas9或Cas9切口酶进一步包含SEQ ID NO:10中提供的氨基酸序列的N497A、R661A、Q695A和/或Q926A突变,或SEQ ID NO:11-260中提供的任何氨基酸序列中相应的突变的一个或多个。在一些实施方案中,Cas9域(例如本文提供的任何核碱基编辑器的Cas9域)包含如SEQ ID NO:720中所示的氨基酸序列。在一些实施方案中,核碱基编辑器包含如SEQ ID NO:721中所示的氨基酸序列。具有高保真度的Cas9域在本领域中是已知的,并且对于本领域技术人员而言是显而易见的。例如,具有高保真度的Cas9域已经描述于Kleinstiver,B.P.,et al.“High-fidelity CRISPR-Cas9nucleases with no detectable genome-wide off-target effects.”Nature 529,490-495(2016);和Slaymaker,I.M.,et al.“Rationally engineered Cas9nucleases with improved specificity.”Science 351,84-88(2015)中;其每一个的全部内容通过引用并入本文。

[0121] 具有降低的Cas9与DNA主链之间的静电相互作用的Cas9变体DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:720)

[0122] 高保真度核碱基编辑器

[0123] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPP

HILWATGLKSGSETPGTSESATPESDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAY
HEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINAS
GVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLA
QIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSK
NGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPF
LKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTAFDKNLPNEKVL
PKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVE
DRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWG
ALSRKLINGIRDKQSGKTILDFLKSDGFANRNFMALIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG
ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLY
LYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLN
AKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRAITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVS
DFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNS
DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDE
IIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLD
ATLIHQSITGLYETRIDLSQLGGD(SEQ ID NO:721)

[0124] Cas9蛋白识别靶DNA序列内的短基序(PAM基序),其是Cas9-DNA相互作用所需要的,但不是由与引导RNA核苷酸序列的互补性确定的。如本文所用,“PAM基序”或“前间隔区相邻基序”是指直接在与引导RNA寡核苷酸序列互补的DNA序列后5’-或3’-的DNA序列。若其随后没有合适的PAM序列,Cas9将不会成功结合、切割或切口靶DNA序列。不希望受任何特定理论的束缚,Cas9酶中的特定氨基酸残基负责与PAM的碱基相互作用并确定PAM特异性。因此,这些残基或附近残基的变化导致不同或放松的PAM特异性。改变或放松PAM特异性可以转变Cas9可以结合的位置,如基于本公开对于本领域技术人员显而易见的。

[0125] 野生型酿脓链球菌Cas9识别规范PAM序列(5’-NGG-3’)。其他Cas9核酸酶(例如来自嗜热链球菌(Streptococcus thermophiles)、金黄色葡萄球菌、脑膜炎奈瑟氏球菌(Neisseria meningitidis)或齿垢密螺旋体(Treponema denticolaor)的Cas9)及其Cas9变体已在本领域中描述为具有不同的或更放松的PAM需求。例如,在Kleinstiver et al.,Nature 523,481-485,2015;Klenstiver et al.,Nature 529,490-495,2016;Ran et al.,Nature,Apr 9;520(7546):186-191,2015;Kleinstiver et al.,Nat Biotechnol,33(12):1293-1298,2015;Hou et al.,Proc Natl Acad Sci U S A,110(39):15644-9,2014;Prykhozhij et al.,PLoS One,10(3):e0119372,2015;Zetsche et al.,Cell 163,759-771,2015;Gao et al.,Nature Biotechnology,doi:10.1038/nbt.3547,2016;Want et al.,Nature 461,754-761,2009;Chavez et al.,doi:dx.doi dot org/10.1101/058974;Fagerlund et al.,Genome Biol.2015;16:25,2015;Zetsche et al.,Cell,163,759-771,2015;和Swarts et al.,Nat Struct Mol Biol,21(9):743-53,2014中,其每一个通过引用并入本文。

[0126] 因此，本公开的引导核苷酸序列-可编程DNA结合蛋白可以识别多种PAM序列，包括但不限于在由引导RNA确定的DNA序列的3'或5'末端上的PAM序列。例如，序列可以是：NGG、NGAN(SEQ ID NO:741)、NGNG(SEQ ID NO:742)、NGAG(SEQ ID NO:743)、NGCG(SEQ ID NO:744)、NNGRRT(SEQ ID NO:745)、NGRRN(SEQ ID NO:746)、NNNRRT(SEQ ID NO:747)、NNNGATT(SEQ ID NO:748)、NNAGAAW(SEQ ID NO:749)、NAAAC(SEQ ID NO:750)、TTN、TTTN(SEQ ID NO:751)和YTN，其中Y是嘧啶，R是嘌呤，并且N是任何核碱基。

[0127] 本公开的一些方面提供了RNA-可编程DNA结合蛋白，其可以用于将蛋白质(例如碱基编辑器)引导至特定核酸(例如DNA或RNA)序列。核酸可编程DNA结合蛋白包括但不限于Cas9(例如dCas9和nCas9)、CasX、CasY、Cpf1、C2c1、C2c2、C2C3和Argonaute。具有不同的PAM特异性的RNA-可编程DNA结合蛋白的一个实例是来自普雷沃氏菌(Prevotella)和弗朗西斯菌(Francisella)1(Cpf1)的聚簇规则间隔短回文重复。与Cas9类似，Cpf1也是2类CRISPR效应物。已经显示，Cpf1介导了强大的DNA干扰，其具有与Cas9不同的特征。Cpf1是缺乏tracrRNA的单个RNA引导的内切核酸酶，并且它可以利用富含T的前间隔区相邻基序(例如TTN、TTTN(SEQ ID NO:751)或YTN)，其在由引导RNA确定的DNA序列的5'端上。此外，Cpf1经由交错的DNA双链断裂切割DNA。在16种Cpf1家族蛋白中，来自氨基酸球菌(Acidaminococcus)和毛螺菌(Lachnospiraceae)的两种酶显示在人细胞中具有有效的基因组编辑活性。Cpf1蛋白是本领域已知的并且先前已有描述，例如Yamano et al.,“Crystal structure of Cpflin complex with guide RNA and target DNA.”Cell(165)2016,p.949-962;其全部内容在此通过引用并入。

[0128] 在本组合物和方法中也有用的是核酸酶无活性的Cpf1(dCpf1)变体，其可以用作引导核苷酸序列-可编程DNA结合蛋白域。Cpf1蛋白具有RuvC样内切核酸酶域，其类似于Cas9的RuvC域，但不具有HNH内切核酸酶域，并且Cpf1的N端不具有Cas9的alfa螺旋识别叶(lobe)。它在Zetsche et al.,Cell,163,759-771,2015(其通过引用并入本文)中显示，Cpf1的RuvC样域负责切割两条DNA链并且RuvC样域的失活使Cpf1核酸酶活性失活。例如，对应于新凶手弗朗西斯菌(Francisella novicida)Cpf1(SEQ ID NO:714)中的D917A、E1006A或D1255A的突变使Cpf1核酸酶活性失活。在一些实施方案中，本公开的dCpf1包含对应于SEQ ID NO:714中D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。在其他实施方案中，本公开的Cpf1切口酶可以包含对应于SEQ ID NO:714中D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。有用于本公开的实施方案的Cpf1切口酶可以包含其他突变和/或本领域已知的其他突变。应当理解，可以根据本公开使用使Cpf1的RuvC域完全或部分失活的任何突变，例如取代突变、缺失或插入，以及Cpf1的这些突变可以产生例如dCpf1或Cpf1切口酶。

[0129] 因此，在一些实施方案中，引导核苷酸序列-可编程DNA结合蛋白是核酸酶无活性的Cpf1(dCpf1)。在一些实施方案中，dCpf1包含SEQ ID NO:714-717的任一个的氨基酸序列。在一些实施方案中，dCpf1包含与SEQ ID NO:714-717的任一个至少85％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或至少99.5％相同的氨基酸序列，并且包含对应于SEQ ID NO:714中D917A、E1006A、D1255A、D917A/E1006A、D917A/D1255A、E1006A/D1255A或D917A/E1006A/D1255A的突变。根

据本公开,来自其他细菌物种的Cpf1也可以用作dCpf1或Cpf1切口酶。

[0130] 野生型新凶手弗朗西斯菌Cpf1(SEQ ID NO:714)(D917、E1006和D1255是粗体且加下划线的)

[0131]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0132] 新凶手弗朗西斯菌Cpf1 D917A(SEQ ID NO:715)

[0133]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSI**A**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0134] 新凶手弗朗西斯菌Cpf1 E1006A(SEQ ID NO:716)

[0135]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**A**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0136] 新凶手弗朗西斯菌Cpf1 D1255A(SEQ ID NO:717)

[0137]

MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQF
FIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKN
LFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGF
HENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELT
FDIDYKTSEVNQRVFSLDEVFEIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYI
NLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFK
TVEEKSIKETLSLLFDDLKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVIGTAVLEYITQ
QIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEEFNKHRDIDKQCRFEEILANFA
AIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNNLLHKLKIF
HISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEKFKLNFENST
LANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKL
LPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFID
FYKQSISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGK
LYLFQIYNKDFSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKI
THPAKEAIANKNKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINL
LLKEKANDVHILSI**D**RGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIE
KDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVF**E**DLNFGFKRGRFKV
EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYV
PAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDKA
AKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAIC
GESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
AANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN

[0138] 除了Cas9和Cpf1之外，Shmakov et al.,“Discovery and Functional Characterization of Diverse Class 2 CRISPR Cas Systems”,Mol.Cell,2015 Nov 5;60(3):385-397已经描述了三种不同的2类CRISPR-Cas系统(C2c1、C2c2和C2c3)，其全部内容在此通过引用并入。两个系统(C2c1和C2c3)的效应物含有与Cpf1相关的RuvC样内切核酸酶域。第三个系统，C2c2含有具有两个预测的HEPN RNA酶域的效应物。与C2c1产生CRISPR RNA不同，成熟CRISPR RNA的产生是不依赖tracrRNA的。C2c1依赖于CRISPR RNA和tracrRNA两者用于DNA切割。已显示细菌性C2c2对于CRISPR RNA成熟具有独特的RNA酶活性，不同于其RNA激活的单链RNA降解活性。这些RNA酶功能彼此不同，并且与Cpf1的CRISPR RNA加工行为不同。参见例如，East-Seletsky,et al.,“Two distinct RNase activities of CRISPR-C2c2enable guide-RNA processing and RNA detection”,Nature,2016Oct 13;538(7624):270-273，其全部内容在此通过引用并入。Leptotrichia shahii中C2c2的体外生化分析已显示，C2c2由单个CRISPR RNA引导，并且可以编程以切割携带互补前间隔区的ssRNA靶标。两个保守的HEPN域中的催化残基介导切割。催化残基中的突变产生催化无活性的RNA结合蛋白。参见例如Abudayyeh et al.,“C2c2is a single-component programmable RNA-guided RNA-targeting CRISPR effector”,Science,2016Aug5;353(6299)，其全部内容在此通过引用并入。

[0139] 已经报告了与嵌合单分子引导RNA(sgRNA)复合的酸土脂环酸芽孢杆菌(Alicyclobaccillus acidoterrastris)C2c1(AacC2c1)的晶体结构。参见例如，Liu et al.,“C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”,Mol.Cell,2017Jan 19;65(2):310-322，其全部内容在此通过引用并入。还已经报告了在与靶DNA结合的酸土脂环酸芽孢杆菌C2c1中作为三元复合物的晶体结构。参见例如，Yang et

al.,“PAM-dependent Target DNA Recognition and Cleavage by C2C1CRISPR-Cas endonuclease”,Cell,2016Dec 15;167(7):1814-1828,其全部内容在此通过引用并入。具有靶DNA链和非靶DNA链两者的AacC2c1的催化能力构象已被独立地捕获定位在单个RuvC催化袋内,其中C2c1介导的切割导致靶DNA的交错的七核苷酸断裂。C2c1三元复合物与先前鉴定的Cas9和Cpf1对应物之间的结构比较证明了CRISPR-Cas9系统使用的机制的多样性。

[0140] 在一些实施方案中,本文提供的任何融合蛋白的引导核苷酸序列-可编程DNA结合蛋白可以是C2c1、C2c2或C2c3蛋白。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白是C2c1蛋白。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白是C2c2蛋白。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白是C2c3蛋白。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白包含与天然存在的C2c1、C2c2或C2c3蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白是天然存在的C2c1、C2c2或C2c3蛋白。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白包含本文所述的任何C2c1、C2c2或C2c3蛋白至少85%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%、至少99%或至少99.5%相同的氨基酸序列。在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白包含本文所述的C2c1、C2c2或C2c3蛋白的任一个的氨基酸序列。应当理解,根据本公开也可以使用来自其他细菌物种的C2c1、C2c2或C2c3。

[0141] C2c1(uniprot.org/uniprot/T0D7A2#)

[0142] sp|T0D7A2|C2C1_ALIAG CRISPR相关的内切核酸酶C2c1OS=酸土脂环酸芽孢杆菌(菌株ATCC 49025/DSM 3922/CIP 106132/NCIMB 13137/GD3B)GN=c2c1PE=1SV=1

[0143] MAVKSIKVKLRLDDMPEIRAGLWKLHKEVNAGVRYYTEWLSLLRQENLYRRSPNGDGEQECDKTAEEC
KAELLERLRARQVENGHRGPAGSDDELLQLARQLYELLVPQAIGAKGDAQQIARKFLSPLADKDAVGGLGIAKAGN
KPRWVRMREAGEPGWEEEKEKAETRKSADRTADVLRALADFGLKPLMRVYTDSEMSSVEWKPLRKGQAVRTWDRDM
FQQAIERMMSWESWNQRVGQEYAKLVEQKNRFEQKNFVGQEHLVHLVNQLQQDMKEASPGLESKEQTAHYVTGRAL
RGSDKVFEKWGKLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQALWREDASFLTRYAVYNSILRKLN
HAKMFATFTLPDATAHPIWTRFDKLGGNLHQYTFLFNEFGERRHAIRFHKLLKVENGVAREVDDVTVPISMSEQLD
NLLPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRRDQLAHMHRRRGARDVYLNVSVRVQSQSEARGERRPPYAA
VFRLVGDNHRAFVHFDKLSDYLAEHPDDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSKGRVPF
FFPIKGNDNLVAVHERSQLLKLPGETESKDLRAIREERQRTLRQLRTQLAYLRLLVRCGSEDVGRRERSWAKLIEQ
PVDAANHMTPDWREAFENELQKLKSLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAKDV
VGGNSIEQIEYLERQYKFLKSWSFFGKVSGQVIRAEKGSRFAITLREHIDHAKEDRLKKLADRIIMEALGYVYALD
ERGKGKWVAKYPPCQLILLEELSEYQFNNDRPPSENNQLMQWSHRGVFQELINQAQVHDLLVGTMYAAFSSRFDAR
TGAPGIRCRRVPARCTQEHNPEPFPWWLNKFVVEHTLDACPLRADDLIPTGEGEIFVSPFSAEEGDFHQIHADLNA
AQNLQQRLWSDFDISQIRLRCDWGEVDGELVLIPRLTGKRTADSYSNKVFYTNTGVTYYERERGKKRRKVFAQEKL
SEEEAELLVEADEAREKSVVLMRDPSGIINRGNWTRQKEFWSMVNQRIEGYLVKQIRSRVPLQDSACENTGDI(SEQ
ID NO:762)

[0144] C2c2(uniprot.org/uniprot/P0D0C6)

[0145] >sp|P0D0C6|C2C2_LEPSD CRISPR相关的内切核糖核酸酶C2c2OS=Leptotrichia

shahii(菌株DSM 19757/CCUG 47503/CIP 107916/JCM 16776/LB37)GN＝c2c2PE＝1SV＝1

[0146] MGNLFGHKRWYEVRDKKDFKIKRKVKVKRNYDGNKYILNINENNNKEKIDNNKFIRKYINYKKNDNIL
KEFTRKFHAGNILFKLKGKEGIIRIENNDDFLETEEVVLYIEAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKI
EIKRQENEEEIEIDIRDEYTNKTLNDCSIILRIIENDELETKKSIYEIFKNINMSLYKIIEKIIENETEKVFENRY
YEEHLREKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYLNVGGDKKKSKNKKMLVEKILNINVDLTVEDIAD
FVIKELEFWNITKRIEKVKKVNNEFLEKRRNRTYIKSYVLLDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIE
KILAEFKIDELIKKLEKELKKGNCDTEIFGIFKKHYKVNFDSKKFSKKSDEEKELYKIIYRYLKGRIEKILVNEQK
VRLKKMEKIEIEKILNESILSEKILKRVKQYTLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKEELDLELITFFAS
TNMELNKIFSRENINNDENIDFFGGDREKNYVLDKKILNSKIKIIRDLDFIDNKNNITNNFIRKFTKIGTNERNRI
LHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSKALNLDVVFKDKKNIITKINDIKISEENNNDIKYLPSFSKV
LPEILNLYRNNPKNEPFDTIETEKIVLNALIYVNKELYKKLILEDDLEENESKNIFLQELKKTLGNIDEIDENIIE
NYYKNAQISASKGNNKAIKKYQKKVIECYIGYLRKNYEELFDFSDFKMNIQEIKKQIKDINDNKTYERITVKTSDK
TIVINDDFEYIISIFALLNSNAVINKIRNRFFATSVWLNTSEYQNIIDILDEIMQLNTLRNECITENWNLNLEEFI
QKMKEIEKDFDDFKIQTKKEIFNNYYEDIKNNILTEFKDDINGCDVLEKKLEKIVIFDDETKFEIDKKSNILQDEQ
RKLSNINKKDLKKKVDQYIKDKDQEIKSKILCRIIFNSDFLKKYKKEIDNLIEDMESENENKFQEIYYPKERKNEL
YIYKKNLFLNIGNPNFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEIDAILKNLNDKLNGYSKEYKEKYIKK
LKENDDFFAKNIQNKNYKSFEKDYNRVSEYKKIRDLVEFNYLNKIESYLIDINWKLAIQMARFERDMHYIVNGLRE
LGIIKLSGYNTGISRAYPKRNGSDGFYTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKPENESIRNYISHF
YIVRNPFADYSIAEQIDRVSNLLSYSTRYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNNDILERLMKPKKVSV
LELESYNSDYIKNLIIELLTKIENTNDTL(SEQ ID NO:764)

[0147] 在一些实施方案中,本公开的引导核苷酸序列-可编程DNA结合蛋白对PAM序列没有需求。此类引导核苷酸序列-可编程DNA结合蛋白的一个实例可以是来自格氏嗜盐碱杆菌(Natronobacterium gregoryi)的Argonaute蛋白(NgAgo)。NgAgo是ssDNA引导的内切核酸酶。NgAgo结合约24个核苷酸的5′磷酸化ssDNA(gDNA),以将其引导至其靶位点,并将在gDNA位点处产生DNA双链断裂。与Cas9相比,NgAgo-gDNA系统不需要前间隔区相邻基序(PAM)。使用核酸酶无活性的NgAgo(dNgAgo)可以极大地扩展可以靶向的密码子。NgAgo的表征和使用已经描述于Gao et al.,Nat Biotechnol.,2016Jul;34(7):768-73.PubMed PMID:27136078;Swarts et al.,Nature.507(7491)(2014):258-61;和Swarts et al.,Nucleic Acids Res.43(10)(2015):5120-9,其每一个通过引用并入本文。格氏嗜盐碱杆菌Argonaute的序列提供于SEQ ID NO:718中。

[0148] 野生型格氏嗜盐碱杆菌Argonaute(SEQ ID NO:718)

[0149] MTVIDLDSTTTADELTSGHTYDISVTLTGVYDNTDEQHPRMSLAFEQDNGERRYITLWKNTTPKDVFTY
DYATGSTYIFTNIDYEVKDGYENLTATYQTTVENATAQEVGTTDEDETFAGGEPLDHHLDDALNETPDDAETESDSG
HVMTSFASRDQLPEWTLHTYTLTATDGAKTDTEYARRTLAYTVRQELYTDHDAAPVATDGLMLLTPEPLGETPLDLD
CGVRVEADETRTLDYTTAKDRLLARELVEEGLKRSLWDDYLVRGIDEVLSKEPVLTCDEFDLHERYDLSVEVGHSGR
AYLHINFRHRFVPKLTLADIDDDNIYPGLRVKTTYRPRRGHIVWGLRDECATDSLNTLGNQSVVAYHRNNQTPINTD
LLDAIEAADRRVVETRRQGHGDDAVSFPQELLAVEPNTHQIKQFASDGFHQQARSKTRLSASRCSEKAQAFAERLDP
VRLNGSTVEFSSEFFTGNNEQQLRLLYENGESVLTFRDGARGAHPDETFSKGIVNPPESFEVAVVLPEQQADTCKAQ
WDTMADLLNQAGAPPTRSETVQYDAFSSPESISLNVAGAIDPSEVDAAFVVLPPDQEGFADLASPTETYDELKKALA

NMGIYSQMAYFDRFRDAKIFYTRNVALGLLAAAGGVAFTTEHAMPGDADMFIGIDVSRSYPEDGASGQINIAATATA
VYKDGTILGHSSTRPQLGEKLQSTDVRDIMKNAILGYQQVTGESPTHIVIHRDGFMNEDLDPATEFLNEQGVEYDIV
EIRKQPQTRLLAVSDVQYDTPVKSIAAINQNEPRATVATFGAPEYLATRDGGGLPRPIQIERVAGETDIETLTRQVY
LLSQSHIQVHNSTARLPITTAYADQASTHATKGYLVQTGAFESNVGFL

[0150] 本文还提供了具有放松的PAM需求的Cas9变体(无PAM的Cas9)。与如SEQ ID NO:1提供的酿脓链球菌Cas9相比,无PAM的Cas9对靶序列表现出增加的活性,所述靶序列在其3'端处不包括规范的PAM(例如,NGG)序列,例如活性增加至少5倍、至少10倍、至少50倍、至少100倍、至少500倍、至少1,000倍、至少5,000倍、至少10,000倍、至少50,000倍、至少100,000倍、至少500,000倍或至少1,000,000倍。此类具有放松的PAM需求的Cas9变体描述于2015年10月23日提交的美国临时申请,USSN 62/245,828;2016年1月15日提交的62/279,346;2016年3月22日提交的62/311,763;2016年4月13日提交的62/322,178;和2016年6月30日提交的62/357,332,其每一个通过引用并入本文。在一些实施方案中,可用于本公开的dCas9或Cas9切口酶可以进一步包含放松PAM需求的突变,例如,对应于SEQ ID NO:1中A262T、K294R、S409I、E480K、E543D、M694I或E1219V的突变。

[0151] 本文讨论的一般结构中使用的"-"可以表示存在任选的接头。如本文所用,术语"接头"是指连接两个分子或部分,例如融合蛋白的两个域,诸如例如引导核苷酸序列-可编程DNA结合蛋白域和重组酶催化域的化学基团或分子。通常,接头位于两个基团、分子或其他部分之间或侧翼有两个基团、分子或其他部分,并且经由共价键与每一个连接,从而连接两者。在一些实施方案中,接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施方案中,接头是有机分子、基团、聚合物或化学部分。在一些实施方案中,接头的长度为5-100个氨基酸,例如长度为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、30-35、35-40、40-45、45-50、50-60、60-70、70-80、80-90、90-100、100-150或150-200个氨基酸。也考虑了更长或更短的接头。接头可以是本领域已知的任何形式。例如,街头可以是来自网站诸如www[dot]ibi[dot]vu[dot]nl/programs/linkerdbwww/或来自www[dot]ibi[dot]vu[dot]nl/programs/linkerdbwww/src/database.txt的接头。接头也可以是非结构化的、结构化的、螺旋形的或延伸的。

[0152] 在一些实施方案中,引导核苷酸序列-可编程DNA结合蛋白域和重组酶催化域经由接头彼此融合。可以采用引导核苷酸序列-可编程DNA结合蛋白域和重组酶催化域之间的各种接头长度和柔性度(例如,范围为从形式(GGGS)n(SEQ ID NO:759)、(GGGGS)n(SEQ ID NO:722)、(GGS)n和(G)n的柔性接头至形式(EAAAK)n(SEQ ID NO:723)、SGSETPGTSESATPES(SEQ ID NO:724)的更刚性接头(参见例如Guilinger et al.,Nat.Biotechnol.2014;32(6):577-82;其全部内容通过引用并入本文)、(XP)n或任何这些的组合,其中X是任何氨基酸并且n独立地是1至30之间的整数,以便实现对于特定应用的活性的最佳长度。在一些实施方案中,n独立地是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,或者若存在超过一个接头或超过一个接头基序,其任何组合。在一些实施方案中,接头包含(GGS)n基序,其中n是1、2、3、4、5、6、7、8、9、10、11、12、13、14或15。在一些实施方案中,接头包含(GGS)n基序,其中n是1、3或7。在一些实施方案中,接头包含XTEN接头。XTEN接头可以具有序列SGSETPGTSESATPES(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:9)。在一些实施方案中,接头包含选自下

组的氨基酸序列，其包括但不限于AGVF(SEQ ID NO:772)、GFLG(SEQ ID NO:773)、FK、AL、ALAL(SEQ ID NO:774)和ALALA(SEQ ID NO:775)。在一些实施方案中，合适的接头基序和配置包括描述于Chen et al.,Fusion protein linkers:property,design and functionality.Adv Drug Deliv Rev.2013;65(10):1357-69(其通过引用并入本文)中的那些。在一些实施方案中，接头可以包含任何以下氨基酸序列:VPFLLEPDNINGKTC(SEQ ID NO:10)、GSAGSAAGSGEF(SEQ ID NO:11)、SIVAQLSRPDPA(SEQ ID NO:12)、MKIIEQLPSA(SEQ ID NO:13)、VRHKLKRVGS(SEQ ID NO:14)、GHGTGSTGSGSS(SEQ ID NO:15)、MSRPDPA(SEQ ID NO:16)、GSAGSAAGSGEF(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)、SGSETPGTSESATPEGGSGGS(SEQ ID NO:9)和GGSM(SEQ ID NO:17)。

[0153] 基于本公开，另外的合适的接头序列对于本领域技术人员而言将是显而易见的。在某些实施方案中，接头可以具有约33埃至约81埃的长度。在另一个实施方案中，接头可以具有约54埃至约81埃的长度。在进一步的实施方案中，接头可以具有约63至约81埃的长度。在另一个实施方案中，接头可以具有约65埃至约75埃的长度。在一些实施方案中，接头可以具有约1.20kDa至约1.85kDa的重量。在某些实施方案中，接头可以具有约1.40kDa至约1.85kDa的重量。在某些实施方案中，接头可以具有约1.60kDa至约1.7kDa的重量。在一些实施方案中，接头是一个氨基酸或多个氨基酸(例如肽或蛋白质)。在一些实施方案中，接头是有机分子、基团、聚合物或化学部分。在一些实施方案中，接头是肽接头。在一些实施方案中，肽接头是具有至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少15、至少20、至少25、至少30、至少40、至少50或更多个氨基酸的任何段的氨基酸。在某些实施方案中，肽接头为18至27个氨基酸长。在具体的实施方案中，肽接头为24个氨基酸长。在一些实施方案中，肽接头包含三肽Gly-Gly-Ser的重复，例如包含序列$(GGS)_n$，其中n表示至少1、2、3、4、5、6、7、8、9、10或更多个重复。在一些实施方案中，接头包含序列$(GGS)_6$(SEQ ID NO:6)。在一些实施方案中，肽接头是16个残基的“XTEN”接头或其变体(参见例如，实施例;和Schellenberger et al.A recombinant polypeptide extends the in vivo half-life of peptides and proteins in a tunable manner.Nat.Biotechnol.27,1186-1190(2009))。在一些实施方案中，XTEN接头包含序列SGSETPGTSESATPES(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:9)。在一些实施方案中，肽接头选自VPFLLEPDNINGKTC(SEQ ID NO:10)、GSAGSAAGSGEF(SEQ ID NO:11)、SIVAQLSRPDPA(SEQ ID NO:12)、MKIIEQLPSA(SEQ ID NO:13)、VRHKLKRVGS(SEQ ID NO:14)、GHGTGSTGSGSS(SEQ ID NO:15)、MSRPDPA(SEQ ID NO:16);或GGSM(SEQ ID NO:17)。在一些实施方案中，接头是非肽接头。在某些实施方案中，非肽接头包含以下的一个或多个:聚乙二醇(PEG)、聚丙二醇(PPG)、共聚(乙烯/丙烯)二醇、聚氧乙烯(POE)、聚氨酯、聚磷腈、多糖、右旋糖酐、聚乙烯醇、聚乙烯吡咯烷酮、聚乙烯乙醚、聚丙烯酰胺、聚丙烯酸酯、聚氰基丙烯酸酯、脂质聚合物、甲壳质、透明质酸、肝素或烷基接头。在一个实施方案中，烷基接头具有式—NH—$(CH_2)_s$—C(O)—，其中s可以是任何整数。在进一步的实施方案中，s可以是1-20的任何整数。

[0154] 重组酶催化域

[0155] 用于本公开的组合物和方法的重组酶催化域可以来自任何重组酶。用于所公开的方法和组合物的合适的重组酶催化域可以获得自例如但不限于酪氨酸重组酶和丝氨酸重

组酶。本文提供的一些示例性合适的重组酶包括，例如但不限于，Gin重组酶(作用于gix位点)、Hin重组酶(作用于hix位点)、β重组酶(作用于6个位点)、Sin重组酶(作用于resH位点)、Tn3重组酶(作用于res位点)、γδ重组酶(作用于res位点)、来自噬菌体P1的Cre重组酶(作用于LoxP位点)；真菌来源的FLP重组酶(作用于FTR位点)；和phiC31整合酶(作用于att位点)。示例性合适的重组酶的非限制性序列可以在下面找到。

[0156] Cre重组酶序列

[0157] MSNLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWAAWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRPSDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRNLAFLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTKLVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGAKDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLDSETGAMVRLLEDGD(SEQID NO:725)

[0158] FLP重组酶

[0159] MPQFGILCKTPPKVLVRQFVERFERPSGEKIALCAAELTYLCWMITHNGTAIKRATFMSYNTIISNSLSFDIVNKSLQFKYKTQKATILEASLKKLIPAWEFTIIPYYGQKHQSDITDIVSSLQLQFESSEEADKGNSHSKKMLKALLSEGESIWEITEKILNSFEYTSRFTKTKTLYQFLFLATFINCGRFSDIKNVDPKSFKLVQNKYLGVIIQCLVTETKTSVSRHIYFFSARGRIDPLVYLDEFLRNSEPVLKRVNRTGNSSSNKQEYQLLKDNLVRSYNKALKKNAPYSIFAIKNGPKSHIGRHLMTSFLSMKGLTELTNVVGNWSDKRASAVARTTYTHQITAIPDHYFALVSRYYAYDPISKEMIALKDETNPIEEWQHIEQLKGSAEGSIRYPAWNGIISQEVLDYLSSYINRRI(SEQ ID NO:726)

[0160] γδ重组酶(Gamma Delta解离酶)

[0161] MRLFGYARVSTSQQSLDIQVRALKDAGVKANRIFTDKASGSSSDRKGLDLLRMKVEEGDVILVKKLDRLGRDTADMIQLIKEFDAQGVSIRFIDDGISTDGEMGKMVVTILSAVAQAERQRILERTNEGRQEAMAKGVVFGRKR(SEQ ID NO:727)

[0162] γδ重组酶(E124Q突变)

[0163] MRLFGYARVSTSQQSLDIQVRALKDAGVKANRIFTDKASGSSSDRKGLDLLRMKVEEGDVILVKKLDRLGRDTADMIQLIKEFDAQGVSIRFIDDGISTDGEMGKMVVTILSAVAQAERQRILQRTNEGRQEAMAKGVVFGRKR(SEQ ID NO:728)

[0164] γδ重组酶(E102Y/E124Q突变)

[0165] MRLFGYARVSTSQQSLDIQVRALKDAGVKANRIFTDKASGSSSDRKGLDLLRMKVEEGDVILVKKLDRLGRDTADMIQLIKEFDAQGVSIRFIDDGISTDGYMGKMVVTILSAVAQAERQRILQRTNEGRQEAMAKGVVFGRKR(SEQ ID NO:729)

[0166] β重组酶

[0167] MAKIGYARVSSKEQNLDRQLQALQGVSKVFSDKLSGQSVERPQLQAMLNYIREGDIVVVTELDRLGRNNKELTELMNAIQQKGATLEVLDLPSMNGIEDENLRRLINNLVIELYKYQAESERKRIKERQAQGIEIAKSKGKFKGRQH(SEQ ID NO:730)

[0168] β重组酶(N95D突变)

[0169] MAKIGYARVSSKEQNLDRQLQALQGVSKVFSDKLSGQSVERPQLQAMLNYIREGDIVVVTELDRLGRNNKELTELMNAIQQKGATLEVLDLPSMDGIEDENLRRLINNLVIELYKYQAESERKRIKERQAQGIEIAKSKGKFKGRQH(SEQ ID NO:731)

[0170] Sin重组酶

[0171] MIIGYARVSSLDQNLERQLENLKTFGAEKIFTEKQSGKSIENRPILQKALNFVRMGDRFIVESIDRLG RNYNEVIHTVNYLKDKEVQLMITSLPMMNEVIGNPLLDKFMKDLIIQILAMVSEQERNESKRRQAQGIQVAKEKGV YKGRPL(SEQ ID NO:732)

[0172] Sin重组酶(Q87R/Q115R突变)

[0173] MIIGYARVSSLDQNLERQLENLKTFGAEKIFTEKQSGKSIENRPILQKALNFVRMGDRFIVESIDRLG RNYNEVIHTVNYLKDKEVRLMITSLPMMNEVIGNPLLDKFMKDLIIRILAMVSEQERNESKRRQAQGIQVAKEKGV YKGRPL(SEQ ID NO:733)

[0174] Tn3重组酶

[0175] MRLFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVKEGDVILVKKLDR LGRDTADMLQLIKEFDAQGVAVRFIDDGISTDGDMGQMVVTILSAVAQAERRRILERTNEGRQEAKLKGIKFGRRR (SEQ ID NO:734)

[0176] Tn3重组酶(G70S/D102Y、E124Q突变)

[0177] MRLFGYARVSTSQQSLDLQVRALKDAGVKANRIFTDKASGSSTDREGLDLLRMKVKEGDVILVKKLDR LSRDTADMLQLIKEFDAQGVAVRFIDDGISTDGYMGQMVVTILSAVAQAERRRILQRTNEGRQEAKLKGIKFGRRR (SEQ ID NO:735)

[0178] Hin重组酶

[0179] MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPGLKRALKYVNKGDTLVVWKLDRLG RSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFHVMSALAEMERELIVERTLAGLAAARAQGRLGGRPV (SEQ ID NO:736)

[0180] Hin重组酶(H107Y突变)

[0181] MATIGYIRVSTIDQNIDLQRNALTSANCDRIFEDRISGKIANRPGLKRALKYVNKGDTLVVWKLDRLG RSVKNLVALISELHERGAHFHSLTDSIDTSSAMGRFFFYVMSALAEMERELIVERTLAGLAAARAQGRLGGRPV (SEQ ID NO:737)

[0182] PhiC31重组酶

[0183] MDTYAGAYDRQSRERENSSAASPATQRSANEDKAADLQREVERDGGRFRFVGHFSEAPGTSAFGTAER PEFERILNECRAGRLNMIIVYDVSRFSRLKVMDAIPIVSELLALGVTIVSTQEGVFRQGNVMDLIHLIMRLDASHK ESSLKSAKILDTKNLQRELGGYVGGKAPYGFELVSETKEITRNGRMVNVVINKLAHSTTPLTGPFEFEPDVIRWWW REIKTHKHLPFKPGSQAAIHPGSITGLCKRMDADAVPTRGETIGKKTASSAWDPATVMRILRDPRIAGFAAEVIYK KKPDGTPTTKIEGYRIQRDPITLRPVELDCGPIIEPAEWYELQAWLDGRGRGKGLSRGQAILSAMDKLYCECGAVM TSKRGEESIKDSYRCRRRKVVDPSAPGQHEGTCNVSMAALDKFVAERIFNKIRHAEGDEETLALLWEAARRFGKLT EAPEKSGERANLVAERADALNALEELYEDRAAGAYDGPVGRKHFRKQQAALTLRQQGAEERLAELEAAEAPKLPLD QWFPEDADADPTGPKSWWGRASVDDKRVFVGLFVDKIVVTKSTTGRGQGTPIEKRASITWAKPPTDDDEDDAQDGT EDVAATGA(SEQ ID NO:738)

[0184] 与所公开的组合物和方法一起使用的重组酶还可以包括其他突变。该公开的一些方面提供了包含与本文所讨论的重组酶序列的序列至少70％、至少80％、至少90％、至少95％或至少97％相同的氨基酸序列的重组酶，其中与本文所讨论的重组酶序列的序列相比，重组酶的氨基酸序列包含至少一个突变。在一些实施方案中，与本文所讨论的重组酶序列的序列相比，重组酶的氨基酸序列包含至少2个、至少3个、至少4个、至少5个、至少6个、至少7个、至少8个、至少9个、至少10个、至少11个、至少12个、至少13个、至少14个或至少15个

突变。

[0185] 例如，γδ重组酶可以包含来自列表的一个或多个突变:R2A、E56K、G101S、E102Y、M103I或E124Q。在一个实施方案中，γδ重组酶可以包含E102Y突变、E124Q突变或E102Y和E124Q突变两者。在另一个实施方案中，β重组酶可以包含一个或多个突变，其包括但不限于N95D。参见例如，Sirk et al.，“Expanding the zinc-finger recombinase repertoire: directed evolution and mutational analysis of serine recombinase specificity determinants”Nucl Acids Res(2014)42(7):4755-4766。在另一个实施方案中，Sin重组酶可以具有一个或多个突变，其包括但不限于:Q87R、Q115R或Q87R和Q115R。在另一个实施方案中，Tn3重组酶可以具有一个或多个突变，其包括但不限于:G70S、D102Y、E124Q及其任何组合。在另一个实施方案中，Hin重组酶可以具有一个或多个突变，其包括但不限于:H107Y。在另一个实施方案中，Sin重组酶可以具有一个或多个突变，其包括但不限于:H107Y。与所公开的组合物和方法一起使用的任何重组酶催化域可以与天然(或野生型)氨基酸序列具有大于85％、90％、95％、98％或99％的序列一致性。例如，在某些实施方案中，Gin重组酶催化域与SEQ ID NO:713中所示的氨基酸序列具有大于85％、90％、95％、98％或99％的序列一致性。在另一个实施方案中，Gin重组酶催化域的氨基酸序列包含对应于H106Y，和/或I127L，和/或I136R和/或G137F的突变。在另一个实施方案中，Gin重组酶催化域的氨基酸序列包含对应于H106Y、I127L、I136R和G137F的突变。在进一步的实施方案中，Gin重组酶的氨基酸序列已得以进一步突变。在具体的实施方案中，Gin重组酶催化域的氨基酸序列包含SEQ ID NO:713。

[0186] 用于本公开的组合物和方法的重组酶催化域可以来自演化的重组酶。如本文所用，术语“演化的重组酶”是指已经改变(例如，通过突变)以识别非天然DNA靶序列的重组酶。

[0187] 可以演化的合适的重组酶包括例如但不限于酪氨酸重组酶和丝氨酸重组酶(例如，本文讨论的任何重组酶)。可以通过本文提供的方法和策略演化的一些示例性合适的重组酶包括例如但不限于，Gin重组酶(作用于gix位点)、Hin重组酶(作用于hix位点)、β重组酶(作用于6个位点)、Sin重组酶(作用于resH位点)、Tn3重组酶(作用于res位点)、γδ重组酶(作用于res位点)、来自噬菌体P1的Cre重组酶(作用于LoxP位点)；λ噬菌体整合酶(作用于att位点)；真菌来源的FLP重组酶(作用于FTR位点)；phiC31整合酶；Dre重组酶，BxB1；和原核β-重组酶。

[0188] 例如，可以改变与本公开的组合物和方法一起使用的演化的重组酶以与非规范重组酶靶序列相互作用(例如，结合和重组)。作为非限制性实例，非规范重组酶靶序列可以是天然存在的，诸如例如哺乳动物基因组中“安全港(safe harbor)”基因组基因座内的序列，例如已知对基因修饰耐受而没有任何不期望的影响的基因组基因座。靶向此类序列的重组酶允许例如在特定基因组位置处靶向插入核酸构建体，而不需要常规的时间和劳动密集型基因靶向规程，例如经由同源重组技术。此外，本文提供的定向演化策略可以用于演化具有改变的活性谱的重组酶，例如，有利于核酸序列的整合而不是该序列的切除的重组酶，或反之亦然。

[0189] 与其野生型对应物相比，演化的重组酶表现出改变的靶序列偏好，可以用于靶向用于重组酶活性的实际任何靶序列。因此，演化的重组酶可以用于修饰例如细胞或受试者

的基因组内的任何序列。因为重组酶可以影响异源核酸分子插入到靶核酸分子中、从核酸分子中切除核酸序列、倒位或替换核酸序列，所以本文提供的技术能够以多种方式(例如，整合、缺失、倒位、交换核酸序列)有效修饰基因组靶标。

[0190] 来自与本公开的方法和组合物一起使用的演化的重组酶的催化域包含与野生型重组酶的序列至少70％、至少80％、至少90％、至少95％或至少97％相同的氨基酸序列，其中与野生型重组酶的序列相比，演化的重组酶的氨基酸序列包含至少一个突变，并且其中演化的重组酶识别DNA重组酶靶序列，其与规范重组酶靶序列相差至少一个核苷酸。在一些实施方案中，演化的重组酶识别DNA重组酶靶序列，其与规范重组酶靶序列(例如，res、gix、hix、six、resH、LoxP、FTR或att核心或相关核心序列)相差至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20至少25或至少30个核苷酸。在一些实施方案中，演化的重组酶识别DNA重组酶靶序列，其与规范重组酶靶序列相差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个核苷酸。

[0191] 在一些实施方案中，仅重组酶的一部分用于本文所述的融合蛋白和方法中。作为非限制性实施方案，仅重组酶的C端部分可以用于本文所述的融合蛋白和方法中。在具体的实施方案中，Cre重组酶的25kDa羧基末端域可以用于组合物和方法中。参见例如，Hoess et al，“DNA Specificity of the Cre Recombinase Resides in the 25kDa Carboxyl Domain of the Protein，”J.Mol.Bio.1990Dec 20，216(4):873-82，其出于所有目的通过引用并入本文。Cre重组酶的25kDa羧基末端域是从蛋白质的R118延伸至羧基末端的部分。在一些实施方案中，用于本发明融合蛋白和方法的Cre重组酶的25kDa羧基末端域可以与Cre重组酶的规范25kDa羧基末端域相差至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19或至少20个氨基酸。在一些实施方案中，用于本发明融合蛋白和方法的Cre重组酶的25kDa羧基末端域可以与Cre重组酶的规范25kDa羧基末端域相差1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19或20个氨基酸。在某些实施方案中，仅Cre重组酶的25kDa羧基末端域的一部分可以用于本文所述的融合蛋白和方法中。例如，使用的Cre重组酶的部分可以是R130至蛋白质的羧基末端、T140至蛋白质的羧基末端、E150至蛋白质的羧基末端、N160至蛋白质的羧基末端、T170至蛋白质的羧基末端、I180至蛋白质的羧基末端、G190至蛋白质的羧基末端、T200至蛋白质的羧基末端、E210至蛋白质的羧基末端、L220至蛋白质的羧基末端、V230至蛋白质的羧基末端、C240至蛋白质的羧基末端、P250至蛋白质的羧基末端、A260至蛋白质的羧基末端、R270至蛋白质的羧基末端、G280至蛋白质的羧基末端、S290至蛋白质的羧基末端、A300至蛋白质的羧基末端或M310至蛋白质的羧基末端。作为另一组非限制性实例，使用的Cre重组酶的部分可以是R118-E340、R118-S330、R118-I320、R118-M310、R118-A300、R118-S290、R118-G280、R118-R270、R118-A260、R118-P250、R118-C240、R118-V230、R118-L220或R118-E210。作为进一步的一组非限制性实例，使用的Cre重组酶的部分可以是R118-E210、G190-R270、E210-S290、P250-M310或R270至蛋白质的羧基末端。

[0192] 在一些实施方案中，本文所述的融合蛋白和方法中使用的Cre重组酶可以在任何位置截短。在具体的实施方案中，本文所述的融合蛋白和方法中使用的Cre重组酶可以截短，使得它以氨基酸R118、A127、E138或R154开始)(每种情况下其之前都是甲硫氨酸)。在另

一组非限制性实施方案中，本文所述的融合蛋白和方法中使用的Cre重组酶可以在R118、A127、E138或R154的10个氨基酸、9个氨基酸、8个氨基酸、7个氨基酸、6个氨基酸、5个氨基酸、4个氨基酸、3个氨基酸、2个氨基酸或1个氨基酸内截短。

[0193] 在一些实施方案中，重组酶靶序列为10-50个核苷酸之间长。在一些实施方案中，重组酶是Cre重组酶、Hin重组酶或FLP重组酶。在一些实施方案中，规范重组酶靶序列是LoxP位点(5′-ATAACTTCGTATA GCATACAT TATACGAAGTTAT-3′(SEQ ID NO:739)。在一些实施方案中，规范重组酶靶序列是FRT位点(5'-GAAGTTCCTATTCTCTAGAAA GTATAGGAACTTC-3′)(SEQ ID NO:740)。在一些实施方案中，与野生型重组酶的序列相比，演化的重组酶的氨基酸序列包含至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14或至少15个突变。在一些实施方案中，演化的重组酶识别包含左半位点、间隔区序列和右半位点的DNA重组酶靶序列，并且其中左半位点不是右半位点的回文序列。

[0194] 在一些实施方案中，演化的重组酶识别包含天然存在的序列的DNA重组酶靶序列。在一些实施方案中，演化的重组酶识别包含在哺乳动物的基因组中的DNA重组酶靶序列。在一些实施方案中，演化的重组酶识别包含在人的基因组中的DNA重组酶靶序列。在一些实施方案中，演化的重组酶识别在哺乳动物的基因组中仅发生一次的DNA重组酶靶序列。在一些实施方案中，演化的重组酶识别哺乳动物的基因组中的DNA重组酶靶序列，其与基因组中的任何其他位点相差至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14或至少15个核苷酸。在一些实施方案中，演化的重组酶识别位于安全港基因组基因座中的DNA重组酶靶序列。在一些实施方案中，安全港基因组基因座是Rosa26基因座。在一些实施方案中，演化的重组酶识别位于与疾病或病症相关的基因组基因座中的DNA重组酶靶序列。

[0195] 在某些实施方案中，演化的重组酶可以靶向人基因组的Rosa基因座中的位点(例如36C6)。此类重组酶的非限制性组可以在例如，2017年1月26日公开的国际PCT公开WO 2017/015545A1，题为“Evolution of Site Specific Recombinases，”中找到，其为此目的通过引用并入本文。在一些实施方案中，与野生型重组酶的序列相比，演化的重组酶的氨基酸序列包含至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14或至少15个突变。编码36C6的核苷酸序列以粗体显示如下；编码GGS接头的那些以斜体显示；编码dCas9接头的那些为黑色；编码FLAG

[0196] 标签和NLS的那些分别用下划线和小写字母表示。

[0197] dCas9-36C6(核苷酸)(SEQ ID NO:765)

[0198]

ATGTCCAACCTCCTTACCGTCCACCAGAATCTCCCTGCCCTTCCGGTGGATGC
CACCTCTGATGAAGTGCGAAAAAACCTGATGGATATGTTTCGCGATAGGCAAG
CTTTTTCTGAACACACGTGGAAGATGCTCCTGTCAGTGTGTAGAAGCTGGGCA
GCTTGGTGCAAGTTGAACAACCGAAAATGGTTTCCTGCCGAACCCGAAGATGT
GAGAGACTACCTCCTCTACCTGCAGGCTCGAGGGCTCGCCGTGAAAACAATCC
AACAACACTTGGGTCAGCTCAACATGCTGCACAGGAGATCTGGGCTGCCCCGG
CCGAGTGACTCTAATGCCGTTAGTCTCGTAATGCGGCGCATTCGCAAAGAGAA
TGTGGATGCTGGAGAACGGGCGAAACAGGCACTGGCTTTTGAACGGACCGAC
TTCGATCAGGTGCGGAGTCTTATGGAGAATAGTGACAGATGCCAGGACATTCG
GAACCTTGCATTCCTGGGTATCGCGTATAATACCCTGCTGAGAATCGCTGAGA
TCGCCAGAATCAGGGTAAAGGATATTTCTCGAACGGACGGGGGACGGATGTTG
ATTCATATCGGTCGCACTAAAACACTTGTGAGTACCGCCGGGGTAGAGAAAGC
CCTGAGCCTTGGAGTTACTAAACTGGTGGAGCGGTGGATTAGCGTGTCCGGCG
TGGCGGATGACCCAAACAATTACTTGTTTTGTAGGGTGCGGAAAAATGGTGTA
GCCGCTCCATCCGCTACCTCACAGTTGAGTACACGCGCGTTGGAGGGGATTTT
CGAAGCCACACATCGCTTGATCTACGGCGCCAAGGACGATTCAGGCCAGCGAT
ATCTTGCCTGGAGCGGGCATAGTGCCCGGGTGGGTGCCGCCCGAGACATGGC
AAGGGCTGGCGTGTCAATTCCTGAAATCATGCAGGCCGGCGGGTGGACCAAC
GTGAACATTGTGATGAACTATATCCGGAACCTGGATAGCGAGACCGGAGCAAT
GGTCAGACTGCTTGAGGATGGCGAC*GGTGGATCCGGAGGGTCCGGAGGTAGTGGC*
*GGCAGCGGTGGTTCAGGTGGCAGCGGAGGGTCAGGAGGCTCT*GATAAAAAGTATTCT
ATTGGTTTAGCTATCGGCACTAATTCCGTTGGATGGGCTGTCATAACCGATGAATAC
AAAGTACCTTCAAAGAAATTTAAGGTGTTGGGGAACACAGACCGTCATTCGATTAA
AAAGAATCTTATCGGTGCCCTCCTATTCGATAGTGGCGAAACGGCAGAGGCGACTC
GCCTGAAACGAACCGCTCGGAGAAGGTATACACGTCGCAAGAACCGAATATGTTAC
TTACAAGAAATTTTTAGCAATGAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGT
TTGGAAGAGTCCTTCCTTGTCGAAGAGGACAAGAAACATGAACGGCACCCCATCTT
TGGAAACATAGTAGATGAGGTGGCATATCATGAAAAGTACCCAACGATTTATCACC
TCAGAAAAAAGCTAGTTGACTCAACTGATAAAGCGGACCTGAGGTTAATCTACTTG
GCTCTTGCCCATATGATAAAGTTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAAT
CCGGACAACTCGGATGTCGACAAACTGTTCATCCAGTTAGTACAAACCTATAATCA
GTTGTTTGAAGAGAACCCTATAAATGCAAGTGGCGTGGATGCGAAGGCTATTCTTA
GCGCCCGCCTCTCTAAATCCCGACGGCTAGAAAACCTGATCGCACAATTACCCGGA
GAGAAGAAAAATGGGTTGTTCGGTAACCTTATAGCGCTCTCACTAGGCCTGACACC
AAATTTTAAGTCGAACTTCGACTTAGCTGAAGATGCCAAATTGCAGCTTAGTAAGG
ACACGTACGATGACGATCTCGACAATCTACTGGCACAAATTGGAGATCAGTATGCG
GACTTATTTTTGGCTGCCAAAAACCTTAGCGATGCAATCCTCCTATCTGACATACTG
AGAGTTAATACTGAGATTACCAAGGCGCCGTTATCCGCTTCAATGATCAAAAGGTA
CGATGAACATCACCAAGACTTGACACTTCTCAAGGCCCTAGTCCGTCAGCAACTGC
CTGAGAAATATAAGGAAATATTCTTTGATCAGTCGAAAAACGGGTACGCAGGTTAT
ATTGACGGCGGAGCGAGTCAAGAGGAATTCTACAAGTTTATCAAACCCATATTAGA
GAAGATGGATGGGACGGAAGAGTTGCTTGTAAAACTCAATCGCGAAGATCTACTGC
GAAAGCAGCGGACTTTCGACAACGGTAGCATTCCACATCAAATCCACTTAGGCGAA
TTGCATGCTATACTTAGAAGGCAGGAGGATTTTTATCCGTTCCTCAAAGACAATCGT
GAAAAGATTGAGAAAATCCTAACCTTTCGCATACCTTACTATGTGGGACCCCTGGC
CCGAGGGAACTCTCGGTTCGCATGGATGACAAGAAAGTCCGAAGAAACGATTACTC
CATGGAATTTTGAGGAAGTTGTCGATAAAGGTGCGTCAGCTCAATCGTTCATCGAG

[0199]

AGGATGACCAACTTTGACAAGAATTTACCGAACGAAAAAGTATTGCCTAAGCACAG
TTTACTTTACGAGTATTTCACAGTGTACAATGAACTCACGAAAGTTAAGTATGTCAC
TGAGGGCATGCGTAAACCCGCCTTTCTAAGCGGAGAACAGAAGAAAGCAATAGTA
GATCTGTTATTCAAGACCAACCGCAAAGTGACAGTTAAGCAATTGAAAGAGGACTA
CTTTAAGAAAATTGAATGCTTCGATTCTGTCGAGATCTCCGGGGTAGAAGATCGATT
TAATGCGTCACTTGGTACGTATCATGACCTCCTAAAGATAATTAAAGATAAGGACTT
CCTGGATAACGAAGAGAATGAAGATATCTTAGAAGATATAGTGTTGACTCTTACCC
TCTTTGAAGATCGGGAAATGATTGAGGAAAGACTAAAAACATACGCTCACCTGTTC
GACGATAAGGTTATGAAACAGTTAAAGAGGCGTCGCTATACGGGCTGGGGACGATT
GTCGCGGAAACTTATCAACGGGATAAGAGACAAGCAAAGTGGTAAAACTATTCTCG
ATTTTCTAAAGAGCGACGGCTTCGCCAATAGGAACTTTATGCAGCTGATCCATGATG
ACTCTTTAACCTTCAAAGAGGATATACAAAAGGCACAGGTTTCCGGACAAGGGGAC
TCATTGCACGAACATATTGCGAATCTTGCTGGTTCGCCAGCCATCAAAAAGGGCAT
ACTCCAGACAGTCAAAGTAGTGGATGAGCTAGTTAAGGTCATGGGACGTCACAAAC
CGGAAAACATTGTAATCGAGATGGCACGCGAAAATCAAACGACTCAGAAGGGGCA
AAAAAACAGTCGAGAGCGGATGAAGAGAATAGAAGAGGGTATTAAAGAACTGGGC
AGCCAGATCTTAAAGGAGCATCCTGTGGAAAATACCCAATTGCAGAACGAGAAACT
TTACCTCTATTACCTACAAAATGGAAGGGACATGTATGTTGATCAGGAACTGGACA
TAAACCGTTTATCTGATTACGACGTCGATGCCATTGTACCCCAATCCTTTTTGAAGG
ACGATTCAATCGACAATAAAGTGCTTACACGCTCGGATAAGAACCGAGGGAAAAGT
GACAATGTTCCAAGCGAGGAAGTCGTAAAGAAAATGAAGAACTATTGGCGGCAGC
TCCTAAATGCGAAACTGATAACGCAAAGAAAGTTCGATAACTTAACTAAAGCTGAG
AGGGGTGGCTTGTCTGAACTTGACAAGGCCGGATTTATTAAACGTCAGCTCGTGGA
AACCCGCCAAATCACAAAGCATGTTGCACAGATACTAGATTCCCGAATGAATACGA
AATACGACGAGAACGATAAGCTGATTCGGGAAGTCAAAGTAATCACTTTAAAGTCA
AAATTGGTGTCGGACTTCAGAAAGGATTTTCAATTCTATAAAGTTAGGGAGATAAA
TAACTACCACCATGCGCACGACGCTTATCTTAATGCCGTCGTAGGGACCGCACTCAT
TAAGAAATACCCGAAGCTAGAAAGTGAGTTTGTGTATGGTGATTACAAAGTTTATG
ACGTCCGTAAGATGATCGCGAAAAGCGAACAGGAGATAGGCAAGGCTACAGCCAA
ATACTTCTTTTATTCTAACATTATGAATTTCTTTAAGACGGAAATCACTCTGGCAAA
CGGAGAGATACGCAAACGACCTTTAATTGAAACCAATGGGGAGACAGGTGAAATC
GTATGGGATAAGGGCCGGGACTTCGCGACGGTGAGAAAAGTTTTGTCCATGCCCCA
AGTCAACATAGTAAAGAAAACTGAGGTGCAGACCGGAGGGTTTTCAAAGGAATCG
ATTCTTCCAAAAAGGAATAGTGATAAGCTCATCGCTCGTAAAAAGGACTGGGACCC
GAAAAAGTACGGTGGCTTCGATAGCCCTACAGTTGCCTATTCTGTCCTAGTAGTGGC
AAAAGTTGAGAAGGGAAAATCCAAGAAACTGAAGTCAGTCAAAGAATTATTGGGG
ATAACGATTATGGAGCGCTCGTCTTTTGAAAAGAACCCCATCGACTTCCTTGAGGCG
AAAGGTTACAAGGAAGTAAAAAAGGATCTCATAATTAAACTACCAAAGTATAGTCT
GTTTGAGTTAGAAAATGGCCGAAAACGGATGTTGGCTAGCGCCGGAGAGCTTCAAA
AGGGGAACGAACTCGCACTACCGTCTAAATACGTGAATTTCCTGTATTTAGCGTCCC
ATTACGAGAAGTTGAAAGGTTCACCTGAAGATAACGAACAGAAGCAACTTTTTGTT
GAGCAGCACAAACATTATCTCGACGAAATCATAGAGCAAATTTCGGAATTCAGTAA
GAGAGTCATCCTAGCTGATGCCAATCTGGACAAAGTATTAAGCGCATACAACAAGC
ACAGGGATAAACCCATACGTGAGCAGGCGGAAAATATTATCCATTTGTTTACTCTT
ACCAACCTCGGCGCTCCAGCCGCATTCAAGTATTTTGACACAACGATAGATCGCAA
ACGATACACTTCTACCAAGGAGGTGCTAGACGCGACACTGATTCACCAATCCATCA
CGGGATTATATGAAACTCGGATAGATTTGTCACAGCTTGGGGGTGAC*GGTGGCTCC*
GATTATAAGGATGATGACGACAAG*GGAGGTTCC*ccaaagaagaaaaggaaggtcTGA

[0200] dCas9-36C6 (氨基酸) (SEQ ID NO: 766)

[0201]

MSNLLTVHQNLPALPVDATSDEVRKNLMDMFRDRQAFSEHTWKMLLSVCRSWA
AWCKLNNRKWFPAEPEDVRDYLLYLQARGLAVKTIQQHLGQLNMLHRRSGLPRP
SDSNAVSLVMRRIRKENVDAGERAKQALAFERTDFDQVRSLMENSDRCQDIRNLA
FLGIAYNTLLRIAEIARIRVKDISRTDGGRMLIHIGRTKTLVSTAGVEKALSLGVTK
LVERWISVSGVADDPNNYLFCRVRKNGVAAPSATSQLSTRALEGIFEATHRLIYGA
KDDSGQRYLAWSGHSARVGAARDMARAGVSIPEIMQAGGWTNVNIVMNYIRNLD
SETGAMVRLLEDGD*GGSGGSGGSGGSGGSGGSGGSGGS*DKKYSIGLAIGTNSVGWAVI
TDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRIC
YLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRK
KLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEEN
PINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSA
SMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPI
LEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNRE
KIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNF
DKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTN
RKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDIL
EDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQS
GKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKK
GILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGS
QILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSI
DNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLS
ELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRK
DFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKS
EQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRK
VLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL
VVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLF
ELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQH
KHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAA
FKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD*GGS*DYKDDDDK*GGS*p
kkkrkv 终止

[0202] 本公开的一些方面提供了演化的重组酶(例如Cre重组酶),其包含与本文讨论的重组酶序列(例如Cre重组酶)的序列至少70%、至少80%、至少90%、至少95%或至少97%相同的氨基酸序列,其中与本文讨论的重组酶(例如Cre重组酶)的序列相比,重组酶(例如Cre重组酶)的氨基酸序列包含至少一个突变,并且其中重组酶(例如Cre重组酶)识别与规范LoxP位点5′-ATAACTTCGTATA GCATACAT TATACGAAGTTAT-3′(SEQ ID NO:739)相差至少一个核苷酸的DNA重组酶靶序列。

[0203] 在一些实施方案中,与本文讨论的重组酶(例如Cre重组酶)序列的序列相比,演化的重组酶(例如Cre重组酶)的氨基酸序列包含至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14或至少15个突变,并且识别与规范靶位点(例如LoxP位点)相差至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14或至少15个核苷酸的DNA重组酶靶序列。

[0204] 在一些实施方案中,演化的Cre重组酶识别包含左半位点、间隔区序列和右半位点的DNA重组酶靶序列,并且其中左半位点不是右半位点的回文序列。在一些实施方案中,演化的Cre重组酶识别包含天然存在的序列的DNA重组酶靶序列。在一些实施方案中,演化的

Cre重组酶识别包含在哺乳动物的基因组中的DNA重组酶靶序列。

[0205] 在一些实施方案中，演化的Cre重组酶识别包含在人的基因组中的DNA重组酶靶序列。在一些实施方案中，演化的Cre重组酶识别在哺乳动物的基因组中仅包含一次的DNA重组酶靶序列。在一些实施方案中，演化的Cre重组酶识别哺乳动物的基因组中的DNA重组酶靶序列，其与基因组中的任何其他位点相差至少1、至少2、至少3、至少4、至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14或至少15个核苷酸。在一些实施方案中，演化的Cre重组酶识别位于安全港基因组基因座中的DNA重组酶靶序列。在一些实施方案中，安全港基因组基因座是Rosa26基因座。在一些实施方案中，演化的Cre重组酶识别位于与疾病或病症相关的基因组基因座中的DNA重组酶靶序列。

[0206] 与本发明方法和组合物一起使用的另外的演化的重组酶(及其制备方法)可以在例如美国专利申请号15/216,844中找到，其通过引用并入本文。

[0207] 对于提供重组酶催化域或演化的重组酶催化域两者，另外的合适的重组酶对于本领域技术人员而言将是显而易见的，并且此类合适的重组酶包括但不限于描述于Hirano et al.,Site-specific recombinases as tools for heterologous gene integration.Appl Microbiol Biotechnol.2011Oct;92(2):227-39;Fogg et al.,New applications for phage integrases.J Mol Biol.2014Jul 29;426(15):2703;Brown et al.,Serine recombinases as tools for genome engineering.Methods.2011Apr;53(4):372-9;Smith et al.,Site-specific recombination by phiC31integrase and other large serine recombinases.Biochem Soc Trans.2010Apr;38(2):388-94;Grindley et al.,Mechanisms of site-specific recombination.Annu Rev Biochem.2006;75:567-605;Smith et al.,Diversity in the serine recombinases.Mol Microbiol.2002Apr;44(2):299-307;Grainge et al.,The integrase family of recombinase:organization and function of the active site.Mol Microbiol.1999Aug;33(3):449-56;Gopaul et al.,Structure and mechanism in site-specific recombination.Curr Opin Struct Biol.1999Feb;9(1):14-20;Cox et al.,Conditional gene expression in the mouse inner ear using Cre-loxP.J Assoc Res Otolaryngol.2012Jun;13(3):295-322;Birling et al.,Site-specific recombinases for manipulation of the mouse genome.Methods Mol Biol.2009;561:245-63;和Mishina M,Sakimura K.Conditional gene targeting on the pure C57BL/6genetic background.Neurosci Res.2007Jun;58(2):105-12中的那些;其每一个的全部内容通过引用并入本文。

[0208] 融合蛋白的结构

[0209] 本公开的融合蛋白可以是本文所述的元件的任何组合和顺序。示例性融合蛋白包括但不限于任何以下结构：NH_2-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中，融合蛋白具有结构NH_2-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中，融合蛋白具有结构NH_2-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-[任

选的接头序列]-[亲和标签]-COOH。在另一个实施方案中,融合蛋白具有结构NH_2-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。

[0210] 在另一个实施方案中,融合蛋白具有结构NH_2-[重组酶催化域]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[亲和标签]-COOH、NH_2-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH、NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH、NH_2-[亲和标签]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[任选的接头序列]-[NLS域]-COOH、NH_2-[亲和标签]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-COOH或NH_2-[亲和标签]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-COOH。

[0211] 在另一个实施方案中,融合蛋白具有结构:NH_2-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[亲和标签]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-[接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。

[0212] 在另一个实施方案中,融合蛋白具有结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接

头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。

[0213] 在另一个实施方案中,融合蛋白具有结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。

[0214] 在另一个实施方案中,融合蛋白具有结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。

[0215] 在另一个实施方案中,融合蛋白具有结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[任选的亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[任选的

亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[亲和标签]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-[接头序列]-[NLS域]-[接头序列]-[亲和标签]-COOH。

[0216] 在一个实施方案中,融合蛋白具有结构NH_2-[任选的亲和标签]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[任选的亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[亲和标签]-[接头序列]-[NLS域]-[接头序列]-[重组酶催化域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-COOH。

[0217] 在一个实施方案中,融合蛋白具有结构NH_2-[任选的亲和标签]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[任选的亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-COOH。在一个实施方案中,融合蛋白包含结构NH_2-[亲和标签]-[接头序列]-[NLS域]-[接头序列]-[引导核苷酸序列-可编程DNA结合蛋白域]-[接头序列]-[重组酶催化域]-COOH。

[0218] 在另一个实施方案中,融合蛋白具有结构NH_2-[任选的亲和标签]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[任选的亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中,融合蛋白包含结构NH_2-[亲和标签]-[接头序列]-[NLS域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导

核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。

[0219] 在另一个实施方案中，融合蛋白具有结构NH_2-[任选的亲和标签]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[任选的亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[亲和标签]-[接头序列]-[NLS域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH.

[0220] 在另一个实施方案中，融合蛋白具有结构NH_2-[任选的亲和标签]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[任选的亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[亲和标签]-[接头序列]-[NLS域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[任选的接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。

[0221] 在另一个实施方案中，融合蛋白具有结构NH_2-[任选的亲和标签]-[任选的接头序列]-[任选的NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[任选的亲和标签]-[任选的接头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[亲和标签]-[任选的接

头序列]-[NLS域]-[任选的接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。在另一个实施方案中，融合蛋白包含结构NH_2-[亲和标签]-[接头序列]-[NLS域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的N端部分]-[接头序列]-[重组酶催化域]-[接头序列]-[二分叉或环状置换的引导核苷酸序列-可编程DNA结合蛋白域的C端部分]-COOH。

[0222] 融合蛋白可以进一步包含一个或多个亲和标签。本文提供的合适的亲和标签包括但不限于生物素羧化酶载体蛋白(BCCP)标签、myc标签、钙调蛋白标签、FLAG标签、血凝素(HA)标签、多组氨酸标签(也称为组氨酸标签或His标签)、多精氨酸(多Arg)标签、麦芽糖结合蛋白(MBP)-标签、nus标签、谷胱甘肽-S-转移酶(GST)标签、绿色荧光蛋白(GFP)标签、硫氧还蛋白标签、S标签、Softag(例如，Softag 1、Softag 3)、strep标签、生物素连接酶标签、FlAsH标签、V5标签和SBP标签。另外的合适的序列对于本领域技术人员而言将是显而易见的。FLAG标签可以具有序列PKKKRKV(SEQ ID NO:702)。一个或多个亲和标签经由一个或多个第三接头与引导核苷酸序列-可编程DNA结合蛋白域、重组酶催化域或NLS域结合。第三接头可以是本文所述的任何肽接头。例如，第三接头可以是肽接头。

[0223] 作为非限制性组的实例，第三接头可以包含XTEN接头SGSETPGTSESATPES(SEQ ID NO:7)、SGSETPGTSESA(SEQ ID NO:8)或SGSETPGTSESATPEGGSGGS(SEQ ID NO:9)，包含三肽GGS的一个或多个重复的氨基酸序列，或任何以下氨基酸序列:VPFLLEPDNINGKTC(SEQ ID NO:10)、GSAGSAAGSGEF(SEQ ID NO:11)、SIVAQLSRPDPA(SEQ ID NO:12)、MKIIEQLPSA(SEQ ID NO:13)、VRHKLKRVGS(SEQ ID NO:14)、GHGTGSTGSGSS(SEQ ID NO:15)、MSRPDPA(SEQ ID NO:16)或GGSM(SEQ ID NO:17)。在某些实施方案中，第三接头包含三肽GGS的一个或多个重复。在实施方案中，第三接头包含三肽GGS的一个至五个重复。在实施方案中，第三接头包含三肽GGS的一个重复。在具体的实施方案中，第三接头具有序列GGS。

[0224] 第三接头也可以是非肽接头。在某些实施方案中，非肽接头包含聚乙二醇(PEG)、聚丙二醇(PPG)、共聚(乙烯/丙烯)二醇、聚氧乙烯(POE)、聚氨酯、聚磷腈、多糖、右旋糖酐、聚乙烯醇、聚乙烯吡咯烷酮、聚乙烯乙醚、聚丙烯酰胺、聚丙烯酸酯、聚氰基丙烯酸酯、脂质聚合物、甲壳质、透明质酸、肝素或烷基接头。在其他实施方案中，烷基接头具有式—NH—$(CH_2)_s$—C(O)—，其中s可以是1和100之间的任何整数，包括端点。在具体的实施方案中，s是1和20之间的任何整数，包括端点。

[0225] 本公开的融合蛋白与SEQ ID NO:185的氨基酸1-1544中所示的氨基酸序列(其与SEQ ID NO:719中所示的序列相同)具有大于90％、95％或99％的序列一致性。

[0226] MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLAAARNKGRRFGRPPKGGSGGSGGSGGSGGSGGSGGSGGSDKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKD

NREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH
SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRF
NASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLS
RKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQ
TVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKL
ITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFR
KDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIM
NFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKL
IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLI
IKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIE
QISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATL
IHQSITGLYETRIDLSQLGGDGGSDYKDDDDK终止(SEQ ID NO:719)

[0227] 在二聚化(或多聚化)的蛋白质,诸如例如核酸酶失活的Cas9(或Cas9gRNA结合域)与重组酶(或重组酶的催化域)之间的融合的背景下,靶位点通常包含左半位点(由一种蛋白质结合)、右半位点(由第二种蛋白质结合),以及在其中进行重组的半位点之间的间隔区序列。在一些实施方案中,重组左半位点或右半位点(而非间隔区序列)。在其他实施方案中,重组间隔区序列。该结构([左半位点]-[间隔区序列]-[右半位点])在本文中称为LSR结构。在一些实施方案中,左半位点和/或右半位点对应于RNA引导的靶位点(例如,Cas9靶位点)。在一些实施方案中,半位点中的任一个或两个比例如Cas9靶向的典型区域更短或更长,例如短于或长于20个核苷酸。在一些实施方案中,左半位点和右半位点包含不同的核酸序列。在一些实施方案中,间隔区序列长度为至少5、至少6、至少7、至少8、至少9、至少10、至少11、至少12、至少13、至少14、至少15、至少16、至少17、至少18、至少19、至少20、至少25、至少30、至少35、至少40、至少45、至少50、至少60、至少70、至少80、至少90、至少100、至少125、至少150、至少175、至少200或至少250bp。在一些实施方案中,间隔区序列为约15bp和约25bp之间长。在一些实施方案中,间隔区序列为约15bp长。在一些实施方案中,间隔区序列为约25bp长。

实施例

[0228] 实施例1:可编程Cas9-丝氨酸重组酶融合蛋白,其在哺乳动物细胞中的DNA序列上起作用

[0229] 材料和方法

[0230] 寡核苷酸和PCR

[0231] 所有寡核苷酸购自Integrated DNA Technologies(IDT,Coralville,CA)并列于表1-5中。除非另有说明,否则酶购自New England Biolabs(Ipswich,MA)。Plasmid Safe ATP依赖性DNA酶购自Epicenter(Madison,WI)。将所有组装的载体转化到One Shot Mach1-T1噬菌体抗性化学感受态细胞(Fisher Scientific,Waltham,MA)。除非另有说明,否则所有PCR反应均使用Q5热启动高保真(Hot Start High-Fidelity)2X Master Mix进行。Phusion聚合酶用于环状聚合酶延伸克隆(CPEC)组装。

[0232] 表1:用于gRNA构建的寡核苷酸

[0233]

寡核苷酸名称	序列	SEQ ID NO:
R.pHU6.TSS(-1).univ	GGTGTTTCGTCCTTTCCACAAG	20
F.非靶	GCACACTAGTTAGGGATAACAGTTTTAGAGCTAGAAATAGC	21
F.Chr10-1	GCCCATGACCCTTCTCCTCTGTTTTAGAGCTAGAAATAGC	22
F.Chr10-1-rev	GCTCAGGGCCTGTGATGGGAGGTTTTAGAGCTAGAAATAGC	23
F.Chr10-2	GGCCCATGACCCTTCTCCTCGTTTTAGAGCTAGAAATAGC	24
F.Chr10-2rev	GCCTCAGGGCCTGTGATGGGAGTTTTAGAGCTAGAAATAGC	25
F.Centromere_Chr_1_5_19-gRNA-for	GACTTGAAACACTCTTTTTCGTTTTAGAGCTAGAAATAGC	26
F.Centromere_Chr_1_5_19-gRNA-rev	GAGTTGAAGACACACAACACAGTTTTAGAGCTAGAAATAGC	27
F.Ch5_155183064-gRNA-for	GGAACTCATGTGATTAACTGGTTTTAGAGCTAGAAATAGC	28
F.Ch5_155183064-gRNA-rev-1	GTCTACCTCTCATGAGCCGGTGTTTTAGAGCTAGAAATAGC	29
F.Ch5_169395198-gRNA-for	GTTTCCCGCAGGATGTGGGATGTTTTAGAGCTAGAAATAGC	30
F.Ch5_169395198-gRNA-rev	GCCTGGGGATTTATGTTCTTAGTTTTAGAGCTAGAAATAGC	31
F.Ch12_62418577-gRNA-for	GAAATAGCACAATGAATGGAAGTTTTAGAGCTAGAAATAGC	32
F.Ch12_62418577-gRNA-rev	GACTTTTTGGGGGAGAGGGAGGTTTTAGAGCTAGAAATAGC	33
F.Ch13_102010574-gRNA-for	GGAGACTTAAGTCCAAAACCGTTTTAGAGCTAGAAATAGC	34
F.Ch13_102010574-gRNA-rev	GTCAGCTATGATCACTTCCCTGTTTTAGAGCTAGAAATAGC	35

[0234] 表2:用于报告物构建的寡核苷酸和gBlock

[0235]

构建体名称	序列	SEQ ID NO:
1-0bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAGGTCTGTAAACCGAGGTGAGACGG	36

[0236]

1-0bp-rev	CCGTCTCACCTCGGTTTACAGACCTCTGTTTGG GAAAATTGGGGACGCCGAGACGA	37
1-1bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTtCTGTAAACCGAGGTGAGACGG	38
1-1bp-rev	CCGTCTCACCTCGGTTTACAGaACCTCTGTTTG GGAAAATTGGGGACGCCGAGACGA	39
1-2bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTatCTGTAAACCGAGGTGAGACGG	40
1-2bp-rev	CCGTCTCACCTCGGTTTACAGatACCTCTGTTTG GGAAAATTGGGGACGCCGAGACGA	41
1-3bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTaatCTGTAAACCGAGGTGAGACGG	42
1-3bp-rev	CCGTCTCACCTCGGTTTACAGattACCTCTGTTTG GGAAAATTGGGGACGCCGAGACGA	43
1-4bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTaaatCTGTAAACCGAGGTGAGACGG	44
1-4bp-rev	CCGTCTCACCTCGGTTTACAGatttACCTCTGTTT GGGAAAATTGGGGACGCCGAGACGA	45
1-5bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTgaaatCTGTAAACCGAGGTGAGACGG	46
1-5bp-rev	CCGTCTCACCTCGGTTTACAGatttcACCTCTGTTT GGGAAAATTGGGGACGCCGAGACGA	47
1-6bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTcgaaatCTGTAAACCGAGGTGAGACGG	48
1-6bp-rev	CCGTCTCACCTCGGTTTACAGatttcgACCTCTGTT TGGGAAAATTGGGGACGCCGAGACGA	49
1-7bp-for	TCGTCTCGGCGTCCCCAATTTTCCCAAACAGAG GTtcgaaatCTGTAAACCGAGGTGAGACGG	50
1-7bp-rev	CCGTCTCACCTCGGTTTACAGatttcgaACCTCTGT TTGGGAAAATTGGGGACGCCGAGACGA	51
2-0bp-for	TCGTCTCGGAGGTTTTGGAACCTCTGTTTGGGA AAATTGGGGAGTCTGAGACGG	52
2-0bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTTCCAAAACCTCCGAGACGA	53
2-1bp-for	TCGTCTCGGAGGTTTTGGACACCTCTGTTTGGG AAAATTGGGGAGTCTGAGACGG	54
2-1bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTGTCCAAAACCTCCGAGACGA	55
2-2bp-for	TCGTCTCGGAGGTTTTGGACTACCTCTGTTTGG GAAAATTGGGGAGTCTGAGACGG	56
2-2bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTAGTCCAAAACCTCCGAGACGA	57
2-3bp-for	TCGTCTCGGAGGTTTTGGACTTACCTCTGTTTG GGAAAATTGGGGAGTCTGAGACGG	58
2-3bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTAAGTCCAAAACCTCCGAGACGA	59
2-4bp-for	TCGTCTCGGAGGTTTTGGACTTAACCTCTGTTT GGGAAAATTGGGGAGTCTGAGACGG	60
2-4bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTTAAGTCCAAAACCTCCGAGACGA	61

[0237]

2-5bp-for	TCGTCTCGGAGGTTTTGGACTTAGACCTCTGTT TGGGAAAATTGGGGAGTCTGAGACGG	62
2-5bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTCTAAGTCCAAAACCTCCGAGACGA	63
2-6bp-for	TCGTCTCGGAGGTTTTGGACTTAGCACCTCTGT TTGGGAAAATTGGGGAGTCTGAGACGG	64
2-6bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTGCTAAGTCCAAAACCTCCGAGACGA	65
2-7bp-for	TCGTCTCGGAGGTTTTGGACTTAGCTACCTCTG TTTGGGAAAATTGGGGAGTCTGAGACGG	66
2-7bp-rev	CCGTCTCAGACTCCCCAATTTTCCCAAACAGAG GTAGCTAAGTCCAAAACCTCCGAGACGA	67
4-0bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTCTGTAAACCGATGAGACGG	68
4-0bp-rev	CCGTCTCATCGGTTTACAGACCTCTGTTTGGGA AAATTGGGGGTGCAGAGACGA	69
4-1bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTtCTGTAAACCGATGAGACGG	70
4-1bp-rev	CCGTCTCATCGGTTTACAGaACCTCTGTTTGGG AAAATTGGGGGTGCAGAGACGA	71
4-2bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTatCTGTAAACCGATGAGACGG	72
4-2bp-rev	CCGTCTCATCGGTTTACAGatACCTCTGTTTGGG AAAATTGGGGGTGCAGAGACGA	73
4-3bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTaatCTGTAAACCGATGAGACGG	74
4-3bp-rev	CCGTCTCATCGGTTTACAGattACCTCTGTTTGGG AAAATTGGGGGTGCAGAGACGA	75
4-4bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTaaatCTGTAAACCGATGAGACGG	76
4-4bp-rev	CCGTCTCATCGGTTTACAGatttACCTCTGTTTGG GAAAATTGGGGGTGCAGAGACGA	77
4-5bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTgaaatCTGTAAACCGATGAGACGG	78
4-5bp-rev	CCGTCTCATCGGTTTACAGatttcACCTCTGTTTGG GAAAATTGGGGGTGCAGAGACGA	79
4-6bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTcgaaatCTGTAAACCGATGAGACGG	80
4-6bp-rev	CCGTCTCATCGGTTTACAGatttcgACCTCTGTTTG GGAAAATTGGGGGTGCAGAGACGA	81
4-7bp-for	TCGTCTCTGCACCCCCAATTTTCCCAAACAGAG GTtcgaaatCTGTAAACCGATGAGACGG	82
4-7bp-rev	CCGTCTCATCGGTTTACAGatttcgaACCTCTGTTT GGGAAAATTGGGGGTGCAGAGACGA	83
5-0bp-for	TCGTCTCGCCGAGGTTTTGGAACCTCTGTTTGG GAAAATTGGGGCTCGTGAGACGG	84
5-0bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTTCCAAAACCTCGGCGAGACGA	85
5-1bp-for	TCGTCTCGCCGAGGTTTTGGACACCTCTGTTTG GGAAAATTGGGGCTCGTGAGACGG	86

[0238]

5-1bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTGTCCAAAACCTCGGCGAGACGA	87
5-2bp-for	TCGTCTCGCCGAGGTTTTGGACTACCTCTGTTT GGGAAAATTGGGGCTCGTGAGACGG	88
5-2bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTAGTCCAAAACCTCGGCGAGACGA	89
5-3bp-for	TCGTCTCGCCGAGGTTTTGGACTTACCTCTGTT TGGGAAAATTGGGGCTCGTGAGACGG	90
5-3bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTAAGTCCAAAACCTCGGCGAGACGA	91
5-4bp-for	TCGTCTCGCCGAGGTTTTGGACTTAACCTCTGT TTGGGAAAATTGGGGCTCGTGAGACGG	92
5-4bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTTAAGTCCAAAACCTCGGCGAGACGA	93
5-5bp-for	TCGTCTCGCCGAGGTTTTGGACTTAGACCTCTG TTTGGGAAAATTGGGGCTCGTGAGACGG	94
5-5bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTCTAAGTCCAAAACCTCGGCGAGACGA	95
5-6bp-for	TCGTCTCGCCGAGGTTTTGGACTTAGCACCTCT GTTTGGGAAAATTGGGGCTCGTGAGACGG	96
5-6bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTGCTAAGTCCAAAACCTCGGCGAGACGA	97
5-7bp-for	TCGTCTCGCCGAGGTTTTGGACTTAGCTACCTC TGTTTGGGAAAATTGGGGCTCGTGAGACGG	98
5-7bp-rev	CCGTCTCACGAGCCCCAATTTTCCCAAACAGA GGTAGCTAAGTCCAAAACCTCGGCGAGACGA	99
1-Chr10--54913298-54913376-for	TCGTCTCGGCGTCCCCTCCCATCACAGGCCCTG AGGTTTAAGAGAAAACCTGAGACGG	100
1-Chr10-54913298-54913376-rev	CCGTCTCAGGTTTTCTCTTAAACCTCAGGGCCT GTGATGGGAGGGGACGCCGAGACGA	101
2-Chr10--54913298-54913376-for	TCGTCTCGAACCATGGTTTTGTGGGCCAGGCCC ATGACCCTTCTCCTCTGGGAGTCTGAGACGG	102
2-Chr10--54913298-54913376-rev	CCGTCTCAGACTCCCAGAGGAGAAGGGTCATG GGCCTGGCCCACAAAACCATGGTTCGAGACGA	103
4-Chr10-54913298-54913376-for	TCGTCTCTGCACCCCCTCCCATCACAGGCCCTG AGGTTTAAGAGAAAACCATTGAGACGG	104
4-Chr10-54913298-54913376-rev	CCGTCTCAATGGTTTTCTCTTAAACCTCAGGGC CTGTGATGGGAGGGGGTGCAGAGACGA	105
5-Chr10-54913298-54913376-for	TCGTCTCGCCATGGTTTTGTGGGCCAGGCCCAT GACCCTTCTCCTCTGGGCTCGTGAGACGG	106
5-Chr10-54913298-54913376-rev	CCGTCTCACGAGCCCAGAGGAGAAGGGTCATG GGCCTGGCCCACAAAACCATGGCGAGACGA	107
3-for	ATCCGTCTCCAGTCGAGTCGGATTTGATCTGAT CAAGAGACAG	108
3-rev	AACCGTCTCGGTGCGTTCGGATTTGATCCAGAC ATGATAAGATAC	109
Esp3I-插入-for	/Phos/CGCGTTGAGACGCTGCCATCCGTCTCGC	110
Esp3I-插入-rev	/Phos/TCGAGCGAGACGGATGGCAGCGTCTCAA	111
着丝粒_Chr_1_5_19-1_2*	GTTGTTCGTCTCGGCGTCCTTGTGTTGTGTGTCT TCAACTCACAGAGTTAAACGATGCTTTACACA GAGTAGACTTGAAACACTCTTTTTCTGGAGTCT	112

[0239]

	GAGACGGTTCTGTTTTGGTGTGATTAGTTAT	
着丝粒_Chr_1_5_19-4_5*	GTTGGTCGTCTCTGCACCCTTGTGTTGTGTGTC TTCAACTCACAGAGTTAAACGATGCTTTACACA GAGTAGACTTGAAACACTCTTTTTCTGGCTCGT GAGACGGTTCTGTTTTGGTGTGATTAGTTAT	113
Ch5_155183064-155183141-1_2*	GTTGTTCGTCTCGGCGTCCCACCGGCTCATGAG AGGTAGAGCTAAGGTCCAAACCTAGGTTTATC TGAGACCGGAACTCATGTGATTAACTGTGGAG TCTGAGACGGTTCTGTTTTGGTGTGATTAGTTA T	114
Ch5_155183064-155183141-4_5*	GTTGGTCGTCTCTGCACCCCACCGGCTCATGAG AGGTAGAGCTAAGGTCCAAACCTAGGTTTATC TGAGACCGGAACTCATGTGATTAACTGTGGCT CGTGAGACGGTTCTGTTTTGGTGTGATTAGTTA T	115
Ch5_169395198-169395274-1_2*	GTTGTTCGTCTCGGCGTCCTTAAGAACATAAAT CCCCAGGAATTCACAGAAACCTTGGTTTGAGC TTTGGATTTCCCGCAGGATGTGGGATAGGAGT CTGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	116
Ch5_169395198-169395274-4_5*	GTTGGTCGTCTCTGCACCCTTAAGAACATAAAT CCCCAGGAATTCACAGAAACCTTGGTTTGAGC TTTGGATTTCCCGCAGGATGTGGGATAGGCTCG TGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	117
Ch12_62418577-62418652-1_2*	GTTGTTCGTCTCGGCGTCCACTCCCTCTCCCCC AAAAAGTAAAGGTAGAAAACCAAGGTTTACAG GCAACAAATAGCACAATGAATGGAATGGAGTC TGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	118
Ch12_62418577-62418652-4_5*	GTTGGTCGTCTCTGCACCCACTCCCTCTCCCCC AAAAAGTAAAGGTAGAAAACCAAGGTTTACAG GCAACAAATAGCACAATGAATGGAATGGCTCG TGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	119
chr13_102010574-102010650-1_2*	GTTGTTCGTCTCGGCGTCCTAGGGAAGTGATCA TAGCTGAGTTTCTGGAAAAACCTAGGTTTTAAA GTTGAGGAGACTTAAGTCCAAAACCTGGAGTC TGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	120
chr13_102010574-102010650-4_5*	GTTGGTCGTCTCTGCACCCTAGGGAAGTGATCA TAGCTGAGTTTCTGGAAAAACCTAGGTTTTAAA GTTGAGGAGACTTAAGTCCAAAACCTGGCTCG TGAGACGGTTCTGTTTTGGTGTGATTAGTTAT	121
将寡核苷酸序列退火以产生图1中所示的片段。名称对应于片段编号(1、2、4或5)，并然后对应于将Cas9结合位点与gix核心位点分开的碱基对间隔区核苷酸的数量。 *如支持材料文件中的方法中所述的双链gBlock。		

[0240] 表3:用于recCas9构建的寡核苷酸

[0241]

寡核苷酸名称	序列	SEQ ID NO:

[0242]

1GGS-连接-for_BamHI	TTCATCGGATCCGATAAAAAGTATTCTATTGGTTTAGCTATCGGCAC	122
5GGS-连接-for_BamHI	TTCATCGGATCCGGTGGTTCAGGTGGCAGCGGAG	123
8GGS-连接-for_BamHI	TTCATCGGATCCGGAGGGTCCGGAGGTAGTGGCGGCAGCGGTGGTTCAGGTGGCAGCGGAG	124
Cas9-rev-FLAG-NLS-AgeI	AATAACCGGTTCAGACCTTCCTTTTCTTCTTTGGGGAACCTCCCTTGTCGTCATCATCCTTATAATCGGAGCCACCGTCACCCCCAAGCTGTGACAAATC	125
1GGS-rev-BamHI	TGATAAGGATCCACCCTTTGGTGGTCTTCCAAACCGCC	126
2GGS-rev-BamH	TGATAAGGATCCACCGCTACCACCCTTTGGTGGTCTTC	127
Gin-for_NotI	AGATCCGCGGCCGCTAATAC	128
Esp3I-for-质粒	TTGAGTcgtctcTATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG	129
Esp3I-rev-质粒	CTGGAAcgtctcACTGTCAGACCAAGTTTACTCATATATACTTTAGATTG	130
spec-Esp3I-for	GGTGTGcgtctcTACAGTTATTTGCCGACTACCTTGGTGATCTCGC	131
spec-Esp3I-rev	ACACCAcgtctcTGTATGAGGGAAGCGGTGATCGCC	132
cpec组装-for-质粒	CATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG	133
cpec 组装-rev-质粒	CTGTCAGACCAAGTTTACTCATATATACTTTAGATTG	134
cpec 组装-for-spec	CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTTGCCGACTACCTTGGTGATCTCG	135
cpec 组装-for-spec2	CAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTATTTGCCGACTACCTTGGTGATCTCG	136
cpec 组装-rev-spec	CCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATG	137

[0243] 表4:定制测序寡核苷酸

[0244]

寡核苷酸名称	序列	SEQ ID NO:
Fwd CMV	CGCAAATGGGCGGTAGGCGTG	138
Cas9coRevE1	CCGTGATGGATTGGTGAATC	139
Cas9coRevE2	CCCATACGATTTCACCTGTC	140
Cas9coRevE3	GGGTATTTTCCACAGGATGC	141
Cas9coRevE4	CTTAGAAAGGCGGGTTTACG	142
Cas9coRevE5	CTTACTAAGCTGCAATTTGG	143
Cas9coRevE6	TGTATTCATCGGTTATGACAG	144
bGH_PArev seq1	CAGGGTCAAGGAAGGCACG	145

[0245]

pHU6-gRNA_for	GTTCCGCGCACATTTCC	146
pHU6-gRNA_rev	GCGGAGCCTATGGAAAAAC	147
pCALNL-for1	GCCTTCTTCTTTTTCCTACAGC	148
pCALNL-for2	CGCATCGAGCGAGCAC	149

[0246] 表5:基因组PCR引物

[0247]

寡核苷酸名称	序列	SEQ ID NO:
FAM19A2-F1	TCAAGTAGCAAAAGAAGTAGGAGTCAG	150
FAM19A2-F2	TTAGATGCATTCGTGCTTGAAG	151
FAM19A2-C1	TTAATTTCTGCTGCTAGAACTAAATCTGG	152
FAM19A2-R1	GGGAAGAAAACTGGATGGAGAATG	153
FAM19A2-R2	CATAAATGACCTAGTGGAGCTG	154
FAM19A2-C2	TGGTTATTTTGCCCATTAGTTGATGC	155

[0248] 报告物构建

[0249] 使用五件式Golden Gate组装构建下述的报告物。片段1-5侧翼为Esp3I位点;Esp3I消化产生互补的5'突出端,指定片段组装的顺序(图6)。通过退火表5中列出的正向和反向互补寡核苷酸产生片段1、2、4和5。通过将20μl的分子级水中的10μl的每种寡核苷酸(100μM)混合,在95℃温育3分钟,并以-0.1℃/秒的速率将温度降至16℃来使片段退火。用引物3-for和3-rev通过PCR扩增含有kanR和PolyA终止密码子的区域产生片段3。这些引物还在该序列的5'和3'末端附加了Esp3I。

[0250] 将退火的片段1、2、4和5稀释12,000倍并将0.625μl的每个片段添加至含有以下的混合物:

[0251] 1)40-50ng片段3

[0252] 2)100ng pCALNL EGFP-Esp3I

[0253] 3)1μL Tango缓冲液(10X)

[0254] 4)1μL DTT(10mM)

[0255] 5)1μL ATP(10mM)

[0256] 6)0.25uL T7连接酶(3,000U/μL)

[0257] 7)0.75uL Esp3I(10U/μL)

[0258] 8)H_2O至10μL

[0259] 将反应在编程进行20个循环(37℃5分钟,20℃)的热循环仪中温育。

[0260] 在完成GoldenGate反应后,将7μL的每个反应物与1μL的ATP(10mM)、1μL的10X Plasmid Safe ATP依赖性DNA酶缓冲液(10X)和1μL的Plasmid Safe ATP依赖性DNA酶(10U/μL)(Epicentre,Madison,WI)混合以去除线性DNA并减少背景。将DNA酶消化物在37℃温育30分钟,并在70℃加热杀死30分钟。将每个反应物的一半(5μL)转化到Mach1-T1细胞中。通过菌落PCR分析菌落并测序。

[0261] 针对图4中使用的报告物修改了方案。使用编码PolyA终止子的5'或3'的靶位点的两个gBlock代替片段1、2、4和5。将这些gBlock(10ng)加入到MMX中,其循环10次(37℃5分钟,20℃)并如上所述继续进行。

[0262] 质粒

[0263] 除非另有说明,否则使用QIAquick凝胶提取试剂盒(Qiagen,Valencia,CA)从琼脂糖凝胶分离DNA片段并使用DNA Clean&Concentrator-5(Zymo Research,Irvine,CA)或Qiaquick PCR纯化试剂盒(Qiagen,Valencia,CA)进一步纯化。使用上面列出的试剂盒之一分离不需要凝胶纯化的PCR片段。

[0264] 使用pCALNL-GFP亚克隆载体pCALNL-EGFP-Esp3I克隆所有recCas9报告物质粒,并基于先前描述的pCALNL-GFP载体(Matsuda and Cepko,Controlled expression of transgenes introduced by in vivo electroporation.Proceedings of the National Academy of Sciences of the United States of America 104,1027-1032(2007),其通过引用并入本文)。为了产生pCALNL-EGFP-Esp3I,用XhoI和MluI消化pCALNL-GFP载体并凝胶纯化以除去loxP位点、卡那霉素抗性标志物和poly-A终止子。退火的寡核苷酸形成EspI插入物,其含有反向的Esp3I位点以及XhoI和MluI相容的突出端;将该插入物连接到XhoI和MluI消化的质粒中并转化。

[0265] 用退火的寡聚物和含有相容的Esp3I突出端的PCR产物通过Golden Gate组装产生pCALNL-GFP recCas9报告物质粒。如前所述用Esp3I(ThermoFisher Scientific,Waltham,MA)建立并进行Golden Gate反应(Sanjana et al.,A transcription activator-like effector toolbox for genome engineering.Nature protocols 7,171-192(2012),其全部内容在此通过引用并入)。图6概述了用于报告物组装的一般组装方案和相关引物以及所有recCas9靶位点的序列分别列于表2和6中。侧翼为两个recCas9靶位点的含有KanR(粗体和下划线)和PolyA终止子(斜体和下划线)的代表性DNA序列如下所示。显示的靶位点均为PAM_NT1-0bp-gix_核心-0bp-NT1_PAM(见表6)。原相邻间隔区基序(protoadjacent spacer motif)(PAM)以粗体显示。碱基对间隔区是小写的。gix位点或gix相关位点以斜体显示,并且dCas9结合位点以下划线标出。对于图4的测定中使用的基因组报告物质粒,在卡那霉素抗性标志物中观察到G至T颠换,在下面的序列中由G/T表示。这存在于该图中使用的所有报告物中,并且预计不会影响结果,因为它远离PolyA终止子和recCas9靶位点。

[0266]

ACGCGT**CCC**CAATTTTCCCAAACAGAGGT*CTGTAAACCGAGGTTTTGGA*ACCTCTGTT
TGGGAAAATTG**GGG**AGTCGAGTCGGATTTGATCTGATCAAGAGACAGGATGAGGA
TCGTTTCGC**ATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGG**
GTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGA
TGCCGCCGTGTTCCGGCTGTCAG/TCGCAGGGGCGCCCGGTTCTTTTTGTCAAG
ACCGACCTGTCCGGTGCCCTGAATGAACTGCAGGACGAGGCAGCGCGGCTAT
CGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACT
GAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCC
TGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATG
CGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAA
ACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGG
ATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAG
GCTCAAGGCGCGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGAT
GCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGA
CTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACC
CGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCT
TTACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGA
CGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGACCAAGCGACGCCCA
ACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCG
GAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTG
GAGTTCTTCGCCCACCCCATCGATAAC*TTGTTTATTGCAGCTTATAATGGTTACAAATAA*
AGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTT
*GTCCAAACTCATCAATGTATCTTATC*ATGTCTGGATCAAATCCGAACGCAC**CCC**CAATT
TTCCCAAACAGAGGT*CTGTAAACCGAGGTTTTGGA*ACCTCTGTTTGGGAAAATTG**GG**
GCTCGAG (SEQ ID NO: 156)

[0267] 表6:用于报告物测定法的靶位点序列的列表

[0268]

靶位点名称	序列	SEQ ID NO:
PAM_NT1-0bp-gix_核心-0bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTt*CTGTAAACCGAG GTTTTGG*AACCTCTGTTTGGGAAAATTG**GGG**	157
PAM_NT1-1bp-gix_核心-1bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTt*CTGTAAACCGAG GTTTTGG*cAACCTCTGTTTGGGAAAATTG**GGG**	158
PAM_NT1-2bp-gix_核心-2bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTat*CTGTAAACCGA GGTTTTGG*ctAACCTCTGTTTGGGAAAATTG**GGG**	159
PAM_NT1-3bp-gix_核心-3bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTaat*CTGTAAACCG AGGTTTTGG*cttAACCTCTGTTTGGGAAAATTG**GGG**	160
PAM_NT1-4bp-gix_	**CCC**CAATTTTCCCAAACAGAGGTaaat*CTGTAAACCG*	161

[0269]

核心-4bp-NT1_PAM	*AGGTTTTGG*cttaAACCTCTGTTTGGGAAAATTG**GGG**	
PAM_NT1-5bp-gix_ 核心-5bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTgaaat*CTGTAAACC* *GAGGTTTTGG*cttagAACCTCTGTTTGGGAAAATTG**G** **GG**	162
PAM_NT1-6bp-gix_ 核心-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC* *CGAGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG **GGG**	163
PAM_NT1-7bp-gix_ 核心-7bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTtcgaaat*CTGTAAAC* *CGAGGTTTTGG*cttagctAACCTCTGTTTGGGAAAATT G**GGG**	164
PAM_NT1-6bp-gix_ 核心-0bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTtcgaaat*CTGTAAAC* *CGAGGTTTTGG*AACCTCTGTTTGGGAAAATTG**GGG**	165
PAM_NT1-6bp-gix_ 核心-1bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTtcgaaat*CTGTAAAC* *CGAGGTTTTGG*cAACCTCTGTTTGGGAAAATTG**GG** **G**	166
PAM_NT1-6bp-gix_ 核心-2bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC* *CGAGGTTTTGG*ctAACCTCTGTTTGGGAAAATTG**GG** **G**	167
PAM_NT1-6bp-gix_ 核心-4bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC* *CGAGGTTTTGG*cttaAACCTCTGTTTGGGAAAATTG**G** **GG**	168
PAM_NT1-6bp-gix_ 核心-5bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTcgaaat*CTGTAAAC* *CGAGGTTTTGG*cttagAACCTCTGTTTGGGAAAATTG **GGG**	169
PAM_NT1-0bp-gix_ 核心-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGT*CTGTAAACCGAG* *GTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**GGG**	170
PAM_NT1-1bp-gix_ 核心-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTt*CTGTAAACCGAG* *GTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**GGG**	171
PAM_NT1-2bp-gix_ 核心-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTat*CTGTAAACCGA* *GGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**GG** **G**	172
PAM_NT1-3bp-gix_ 核心-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTaat*CTGTAAACCG* *AGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**G** **GG**	173
PAM_NT1-4bp-gix_ 核心-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTaaat*CTGTAAACCG* *AGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**G** **GG**	174
PAM_NT1-5bp-gix_ 核心-6bp-NT1_PAM	**CCC**CAATTTTCCCAAACAGAGGTgaaat*CTGTAAACC* *GAGGTTTTGG*cttagcAACCTCTGTTTGGGAAAATTG**G** **GG**	175
染色体 _10-54913298-54913376*	**CCC**CTCCCATCACAGGCCCTGAGgtttaa*GAGAAAAC* *CATGGTTTTGTG*ggccagGCCCATGACCCTTCTCCTCT **GGG**	176
着丝粒_染色体 s_1_5_19	**CCT**TGTGTTGTGTGTCTTCAACTcacag*AGTTAAACGA* *TGCTTTACAC*agagtaGACTTGAAACACTCTTTTTC**TG** **G**	177
染色体 _5_155183064-155183141 (位点1)	**CCA**CCGGCTCATGAGAGGTAGAGctaag*GTCCAAAC* *CTAGGTTTATCT*gagaccGGAACTCATGTGATTAACTG **TGG**	178
染色体	**CCT**TAAGAACATAAATCCCCAGGaattc*ACAGAAACC*	179

[0270]

_5_169395198-169395274 (位点2)	*TTGGTTTGAGC*tttggaTTTCCCGCAGGATGTGGGAT**AGG**	
染色体_12_62418577-62418652	**CCA**CTCCCTCTCCCCCAAAAAGTaaagg*TAGAAAACCAAGGTTTACAG*gcaacAAATAGCACAATGAATGGAA**TGG**	180
染色体_13_102010574-102010650 (FGF14)	**CCT**AGGGAAGTGATCATAGCTGAgtttct*GGAAAAACCTAGGTTTTAAA*gttgaGGAGACTTAAGTCCAAAACC**TGG**	181

原相邻间隔区基序(PAM)以粗体显示。碱基对间隔区是小写的。gix位点或gix相关位点用斜体显示，dCas9结合位点用下划线标出。

*染色体_10报告物在gix位点的5'和3'末端含有两个重叠的PAM位点和dCas9结合位点。

[0271] 通过用寡核苷酸1GGS-rev-BamHI或2GGS-rev-BamHI(使用接头SEQ ID NO:182)和Gin-for-NotI对编码演化的、过度激活的Gin变体(Ginβ)(Gaj et al.,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic acids research 41,3937-3946(2013),其全部内容在此通过引用并入)的gBlock进行PCR扩增来构建含有recCas9基因的质粒。用BamHI和NotI消化PCR片段,纯化并连接到先前描述的表达载体(Addgene质粒43861)(参见例如,Fu et al.,High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.Nature biotechnology 31,822-826(2013),其全部内容在此通过引用并入)中以产生亚克隆载体pGin-1GGS和pGIN-2GGS(使用接头SEQ ID NO:182)。与Cas9-rev-FLAG-NLS-AgeI一起使用寡核苷酸1GGS-连接-for-BamHI、5GGS-连接-for-BamHI(使用接头SEQ ID NO:701)或8GGS-连接-for-BamHI(使用接头SEQ ID NO:183)构建编码具有1、5或8个GGS接头的Cas9-FLAG-NLS的PCR片段(见表3)。对于编码GGS氨基酸接头的DNA序列,见表7。用BamHI和AgeI消化PCR片段和亚克隆质粒并连接以产生质粒pGinβ-2xGGS-dCas9-FLAG-NLS(使用接头SEQ ID NO:182)、pGinβ-5xGGS-dCas9-FLAG-NLS(使用接头SEQ ID NO:701)和pGinβ-8xGGS-dCas9-FLAG-NLS(使用接头SEQ ID NO:183)。对于pGinβ-8xGGS-dCas9-FLAG-NLS(即recCas9)的DNA和氨基酸序列,参见下文。编码Ginβ的序列以粗体显示;编码GGS接头的序列以斜体显示;编码dCas9接头的序列为黑色;编码FLAG标记和NLS的序列分别用下划线和小写字母表示。

[0272]

ATGCTCATTGGCTACGTGCGCGTCTCAACTAACGACCAGAATACCGATCTTCA
GAGGAACGCACTGGTTTGTGCAGGCTGCGAACAGATTTTCGAGGACAAACTCA
GCGGGACACGGACGGACAGACCTGGCCTCAAGCGAGCACTCAAGAGGCTGCA
GAAAGGAGACACTCTGGTGGTCTGGAAATTGGACCGCCTGGGTCGAAGCATG
AAGCATCTCATTTCTCTGGTTGGCGAACTGCGAGAAAGGGGGATCAACTTTCG
AAGTCTGACGGATTCCATAGATACAAGCAGCCCCATGGGCCGGTTCTTCTTCT
ACGTGATGGGTGCACTGGCTGAAATGGAAAGAGAACTCATTATAGAGCGAACC
ATGGCAGGGCTTGCGGCTGCCAGGAATAAAGGCAGGCGGTTTGGAAGACCAC
CAAAG*GGTGGATCCGGAGGGTCCGGAGGTAGTGGCGGCAGCGGTGGTTCAGGTGGCA*
*GCGGAGGGTCAGGAGGCTCT*GATAAAAAGTATTCTATTGGTTTAGCTATCGGCACTA
ATTCCGTTGGATGGGCTGTCATAACCGATGAATACAAAGTACCTTCAAAGAAATTT
AAGGTGTTGGGGAACACAGACCGTCATTCGATTAAAAAGAATCTTATCGGTGCCCT
CCTATTCGATAGTGGCGAAACGGCAGAGGCGACTCGCCTGAAACGAACCGCTCGGA
GAAGGTATACACGTCGCAAGAACCGAATATGTTACTTACAAGAAATTTTTAGCAAT
GAGATGGCCAAAGTTGACGATTCTTTCTTTCACCGTTTGGAAGAGTCCTTCCTTGTC
GAAGAGGACAAGAAACATGAACGGCACCCCATCTTTGGAAACATAGTAGATGAGG
TGGCATATCATGAAAAGTACCCAACGATTTATCACCTCAGAAAAAAGCTAGTTGAC
TCAACTGATAAAGCGGACCTGAGGTTAATCTACTTGGCTCTTGCCCATATGATAAAG
TTCCGTGGGCACTTTCTCATTGAGGGTGATCTAAATCCGGACAACTCGGATGTCGAC
AAACTGTTCATCCAGTTAGTACAAACCTATAATCAGTTGTTTGAAGAGAACCCTATA
AATGCAAGTGGCGTGGATGCGAAGGCTATTCTTAGCGCCCGCCTCTCTAAATCCCG
ACGGCTAGAAAACCTGATCGCACAATTACCCGGAGAGAAGAAAAATGGGTTGTTCG
GTAACCTTATAGCGCTCTCACTAGGCCTGACACCAAATTTTAAGTCGAACTTCGACT
TAGCTGAAGATGCCAAATTGCAGCTTAGTAAGGACACGTACGATGACGATCTCGAC
AATCTACTGGCACAAATTGGAGATCAGTATGCGGACTTATTTTTGGCTGCCAAAAA
CCTTAGCGATGCAATCCTCCTATCTGACATACTGAGAGTTAATACTGAGATTACCAA
GGCGCCGTTATCCGCTTCAATGATCAAAAGGTACGATGAACATCACCAAGACTTGA
CACTTCTCAAGGCCCTAGTCCGTCAGCAACTGCCTGAGAAATATAAGGAAATATTC
TTTGATCAGTCGAAAAACGGGTACGCAGGTTATATTGACGGCGGAGCGAGTCAAGA
GGAATTCTACAAGTTTATCAAACCCATATTAGAGAAGATGGATGGGACGGAAGAGT
TGCTTGTAAAACTCAATCGCGAAGATCTACTGCGAAAGCAGCGGACTTTCGACAAC
GGTAGCATTCCACATCAAATCCACTTAGGCGAATTGCATGCTATACTTAGAAGGCA
GGAGGATTTTTATCCGTTCCTCAAAGACAATCGTGAAAAGATTGAGAAAATCCTAA
CCTTTCGCATACCTTACTATGTGGGACCCCTGGCCCGAGGGAACTCTCGGTTCGCAT
GGATGACAAGAAAGTCCGAAGAAACGATTACTCCATGGAATTTTGAGGAAGTTGTC
GATAAAGGTGCGTCAGCTCAATCGTTCATCGAGAGGATGACCAACTTTGACAAGAA
TTTACCGAACGAAAAAGTATTGCCTAAGCACAGTTTACTTTACGAGTATTTCACAGT
GTACAATGAACTCACGAAAGTTAAGTATGTCACTGAGGGCATGCGTAAACCCGCCT
TTCTAAGCGGAGAACAGAAGAAAGCAATAGTAGATCTGTTATTCAAGACCAACCGC
AAAGTGACAGTTAAGCAATTGAAAGAGGACTACTTTAAGAAAATTGAATGCTTCGA
TTCTGTCGAGATCTCCGGGGTAGAAGATCGATTTAATGCGTCACTTGGTACGTATCA
TGACCTCCTAAAGATAATTAAAGATAAGGACTTCCTGGATAACGAAGAGAATGAAG
ATATCTTAGAAGATATAGTGTTGACTCTTACCCTCTTTGAAGATCGGGAAATGATTG
AGGAAAGACTAAAAACATACGCTCACCTGTTCGACGATAAGGTTATGAAACAGTTA
AAGAGGCGTCGCTATACGGGCTGGGGACGATTGTCGCGGAAACTTATCAACGGGAT
AAGAGACAAGCAAAGTGGTAAAACTATTCTCGATTTTCTAAAGAGCGACGGCTTCG
CCAATAGGAACTTTATGCAGCTGATCCATGATGACTCTTTAACCTTCAAAGAGGATA
TACAAAAGGCACAGGTTTCCGGACAAGGGGACTCATTGCACGAACATATTGCGAAT
CTTGCTGGTTCGCCAGCCATCAAAAAGGGCATACTCCAGACAGTCAAAGTAGTGGA
TGAGCTAGTTAAGGTCATGGGACGTCACAAACCGGAAAACATTGTAATCGAGATGG

[0273]

CACGCGAAAATCAAACGACTCAGAAGGGGCAAAAAAACAGTCGAGAGCGGATGAA
GAGAATAGAAGAGGGTATTAAAGAACTGGGCAGCCAGATCTTAAAGGAGCATCCT
GTGGAAAATACCCAATTGCAGAACGAGAAACTTTACCTCTATTACCTACAAAATGG
AAGGGACATGTATGTTGATCAGGAACTGGACATAAACCGTTTATCTGATTACGACG
TCGATGCCATTGTACCCCAATCCTTTTTGAAGGACGATTCAATCGACAATAAAGTGC
TTACACGCTCGGATAAGAACCGAGGGAAAAGTGACAATGTTCCAAGCGAGGAAGT
CGTAAAGAAAATGAAGAACTATTGGCGGCAGCTCCTAAATGCGAAACTGATAACGC
AAAGAAAGTTCGATAACTTAACTAAAGCTGAGAGGGGTGGCTTGTCTGAACTTGAC
AAGGCCGGATTTATTAAACGTCAGCTCGTGGAAACCCGCCAAATCACAAAGCATGT
TGCACAGATACTAGATTCCCGAATGAATACGAAATACGACGAGAACGATAAGCTGA
TTCGGGAAGTCAAAGTAATCACTTTAAAGTCAAAATTGGTGTCGGACTTCAGAAAG
GATTTTCAATTCTATAAAGTTAGGGAGATAAATAACTACCACCATGCGCACGACGC
TTATCTTAATGCCGTCGTAGGGACCGCACTCATTAAGAAATACCCGAAGCTAGAAA
GTGAGTTTGTGTATGGTGATTACAAAGTTTATGACGTCCGTAAGATGATCGCGAAA
AGCGAACAGGAGATAGGCAAGGCTACAGCCAAATACTTCTTTTATTCTAACATTAT
GAATTTCTTTAAGACGGAAATCACTCTGGCAAACGGAGAGATACGCAAACGACCTT
TAATTGAAACCAATGGGGAGACAGGTGAAATCGTATGGGATAAGGGCCGGGACTT
CGCGACGGTGAGAAAAGTTTTGTCCATGCCCCAAGTCAACATAGTAAAGAAAACTG
AGGTGCAGACCGGAGGGTTTTCAAAGGAATCGATTCTTCCAAAAAGGAATAGTGAT
AAGCTCATCGCTCGTAAAAAGGACTGGGACCCGAAAAAGTACGGTGGCTTCGATAG
CCCTACAGTTGCCTATTCTGTCCTAGTAGTGGCAAAAGTTGAGAAGGGAAAATCCA
AGAAACTGAAGTCAGTCAAAGAATTATTGGGGATAACGATTATGGAGCGCTCGTCT
TTTGAAAAGAACCCCATCGACTTCCTTGAGGCGAAAGGTTACAAGGAAGTAAAAAA
GGATCTCATAATTAAACTACCAAAGTATAGTCTGTTTGAGTTAGAAAATGGCCGAA
AACGGATGTTGGCTAGCGCCGGAGAGCTTCAAAAGGGGAACGAACTCGCACTACC
GTCTAAATACGTGAATTTCCTGTATTTAGCGTCCCATTACGAGAAGTTGAAAGGTTC
ACCTGAAGATAACGAACAGAAGCAACTTTTTGTTGAGCAGCACAAACATTATCTCG
ACGAAATCATAGAGCAAATTTCGGAATTCAGTAAGAGAGTCATCCTAGCTGATGCC
AATCTGGACAAAGTATTAAGCGCATACAACAAGCACAGGGATAAACCCATACGTG
AGCAGGCGGAAAATATTATCCATTTGTTTACTCTTACCAACCTCGGCGCTCCAGCCG
CATTCAAGTATTTTGACACAACGATAGATCGCAAACGATACACTTCTACCAAGGAG
GTGCTAGACGCGACACTGATTCACCAATCCATCACGGGATTATATGAAACTCGGAT
AGATTTGTCACAGCTTGGGGGTGAC*GGTGGCTCC*GATTATAAGGATGATGACGACA
AG*GGAGGTTCC*ccaaagaagaaaaggaaggtcTGA (SEQ ID NO: 184)

MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKG
DTLVVWKLDRLGRSMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGAL
AEMERELIIERTMAGLAAARNKGRRFGRPPK*GGSGGSGGSGGSGGSGGSGGSGGS*DK
KYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATR
LKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIV
DEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVD
KLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLI
ALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAIL
LSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDL
LKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYT
GWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQ
GDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQ
KNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRL

[0274]

SDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAK
LITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKL
IREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEF
VYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGE
TGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDW
DPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAK
GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYE
KLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIR
EQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQ
LGGD*GGS*DYKDDDDK*GGS*pkkkrkv 终止 (SEQ ID NO: 185)

[0275] Gin重组酶催化域(其是SEQ ID NO:185的氨基酸1-142)与SEQ ID NO:713的序列相同。dCas9域(其是SEQ ID NO:185的氨基酸167-1533)与SEQ ID NO:712的序列相同。

[0276] MLIGYVRVSTNDQNTDLQRNALVCAGCEQIFEDKLSGTRTDRPGLKRALKRLQKGDTLVVWKLDRLGR SMKHLISLVGELRERGINFRSLTDSIDTSSPMGRFFFYVMGALAEMERELIIERTMAGLAAARNKGRRFGRPPK (SEQ ID NO:713)

[0277] DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTAR RRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDS TDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRR LENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLS DAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFY KFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIP YYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELT KVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGK TILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGR HKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELD INRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAER GGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEI RKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGG FDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENG RKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADA NLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRID LSQLGGD(SEQ ID NO:712)

[0278] 表7:编码GGS接头的DNA序列

[0279]

GGS 接头	SEQ ID NO:	用于GGS接头的DNA序列	SEQ ID NO:

[0280]

2XGGS	182	GGTGGTAGCGGTGGATCC	186
5XGGS	701	GGTGGATCCGGTGGTTCAGGTGGCAGCGGAGGGTCAGGAGGCTCT	187
8XGGS	183	GGTGGATCCGGAGGGTCCGGAGGTAGTGGCGGCAGCGGTGGTTCAGGTGGCAGCGGAGGGTCAGGAGGCTCT	188

[0281] 对于质粒测序实验,通过用PCR片段的golden gate克隆,用SpecR替换pGinβ-8xGGS-dCas9-FLAG-NLS(使用接头SEQ ID NO:183)中的AmpR基因。通过用Esp3I-for-质粒和Esp3I-rev-质粒进行PCR,在位于AmpR基因侧翼的位点处将Esp3I位点引入pGinβ-8xGGS-dCas9-FLAG-NLS(使用接头SEQ ID NO:183)质粒中。引物spec-Esp3I-for和spec-Esp3I-rev用于扩增SpecR标志物以及引入Esp3I位点和Esp3I产生的突出端,所述突出端与由Esp3I切割的质粒PCR产物产生的突出端相容。按照用于产生如本文所述的报告物质粒的方案对两个片段进行golden gate组装。

[0282] pHU6-NT1引导RNA表达载体基于先前描述的pFYF1328(Fu et al.,High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.Nature biotechnology 31,822-826(2013),其全部内容在此通过引用并入),其被改变以靶向细菌萤光素酶基因LuxAB内的区域。通过用通用引物R.pHU6.TSS(-1).univ和编码独特引导RNA序列的引物(表1)进行整个载体的PCR扩增来产生引导RNA表达载体。表8中给出了引导RNA序列的列表。这些引物用T4多核苷酸激酶磷酸化。将PCR反应产物和线性引导RNA表达载体平端连接并转化。用于初始优化的引导RNA表达载体、脱靶对照引导RNA序列和靶向染色体10基因座的那些包含AmpR。本研究中所述的所有其他质粒含有specR以促进测序实验。最初经由CPEC将壮观霉素抗性引入到引导RNA表达载体中,基本上如所述(Quan et al.,Circular polymerase extension cloning of complex gene libraries and pathways.PloS one 4,e6441(2009);和Hillson(2010),vol.2015,pp.CPEC protocol;其每篇通过引用并入本文),并且然后如上所述通过载体的PCR扩增构建引导RNA质粒。将反应物在37℃与40U的DpnI温育过夜,纯化并转化。通过用寡核苷酸cpec-组装-for-spec2和cpec组装-rev进行引导RNA表达载体的PCR扩增产生CPEC的片段。通过经由寡核苷酸cpec-组装-for-spec和cpec-组装-rev-spec对SpecR基因进行PCR扩增产生specR片段。pUC19(ThermoFisher Scientific,Waltham,MA)进行了类似的修饰。

[0283] 表8:gRNA序列的列表

[0284]

gRNA名称	gRNA序列	SEQ ID NO:
中靶_gRNA	ACCTCTGTTTGGGAAAATTG	189
非靶_gRNA	gCACACTAGTTAGGGATAACA	190
染色体_10-54913298-54913376_gRNA-rev-5	gCCTCAGGGCCTGTGATGGGA	191
染色体_10-54913298-54913376_gRNA-rev-6	gCTCAGGGCCTGTGATGGGAG	192
染色体_10-54913298-54913376_gRNA-for-5	GGCCCATGACCCTTCTCCTC	193
染色体_10-54913298-54913376_gRNA-for-6	GCCCATGACCCTTCTCCTCT	194
着丝粒_染色体s_1_5_19-gRNA-for	GACTTGAAACACTCTTTTTC	195
着丝粒_染色体s_1_5_19-gRNA-rev	gAGTTGAAGACACACAACACA	196
染色体_5_155183064-155183141_(位点1)_gRNA-for	GGAACTCATGTGATTAACTG	197
染色体_5_155183064-155183141_(位点1)_gRNA-rev	gTCTACCTCTCATGAGCCGGT	198
染色体_5_169395198-169395274_(位点2)_gRNA-for	gTTTCCCGCAGGATGTGGGAT	199
染色体_5_169395198-169395274_(位点2)_gRNA-rev	gCCTGGGGATTTATGTTCTTA	200
染色体_12_62418577-62418652_gRNA-for	gAAATAGCACAATGAATGGAA	201
染色体_12_62418577-62418652_gRNA-rev	gACTTTTTGGGGGAGAGGGAG	202
染色体_13_102010574-102010650_(FGF14)_gRNA-for	GGAGACTTAAGTCCAAAACC	203
染色体_13_102010574-102010650_(FGF14)_gRNA-rev	gTCAGCTATGATCACTTCCCT	204
脱靶-for (CLTA)	GCAGATGTAGTGTTTCCACA	205
脱靶-rev(VEGF)	GGGTGGGGGGAGTTTGCTCC	206
染色体_12_62098359-62098434_(FAM19A2)_gRNA-rev	gATATCCGTTTATCAGTGTCA	207
染色体_12_62098359-62098434_(FAM19A2)_gRNA-for	gTTCCTAAGCTTGGGCTGCAG	208
染色体_12_62112591-62112668_(FAM19A2)_gRNA-rev	gCCTAAAAGTGACTGGGAGAA	209

[0285]

染色体_12_62112591-62112668_(FAM19A2)_gRNA-for	gCACAGTCCCATATTTCTTGG	210

[0286] 细胞培养和转染

[0287] HEK293T细胞购自美国典型培养物保藏中心(American Type Culture Collection)(ATCC,Manassas,VA)。将细胞在补充有10％胎牛血清(FBS,Life Technologies,Carlsbad,CA)的Dulbecco改良的Eagle培养基(DMEM)+GlutaMAX-1(4.5g/LD葡萄糖+110mg/mL丙酮酸钠)中培养。将细胞在37℃,5％CO_2在潮湿的培养箱中培养。

[0288] 用于转染的质粒从PureYield Plasmid Miniprep System(Promega,Madison,WI)中分离。转染前一晚,在48孔胶原处理的平板(Corning,Corning,NY)中以每孔3×10^5个细胞的密度接种HEK293T细胞。在25μL的Opti-MEM(ThermoFisher Scientific,Waltham,MA)中制备转染反应。对于每次转染,将45ng的每种引导RNA表达载体、9ng的报告物质粒、9ng的piRFP670-N1(Addgene Plasmid 45457)和160ng的recCas9表达载体混合,与Opti-MEM(ThermoFisher Scientific,Waltham,MA)中的0.8μL liplofectamine 2000组合并加入各孔中。

[0289] 流式细胞术

[0290] 转染后60-72小时后,用磷酸盐缓冲盐水洗涤细胞,并用50μL的0.05％胰蛋白酶-EDTA(Life Technologies,Carlsbad,CA)在37℃收获5-10分钟。将细胞在250μL培养基中稀释,并在BD Fortessa分析仪上运行。使用635nm激光激发iRFP荧光,并使用670/30带通滤光器收集发射。使用488nM激光激发EGFP并用505长通和530/30带通滤光器获得发射荧光。在FlowJo软件上分析数据,门控活的且经转染的事件(表达iRFP)。以从至少6,000个活事件门控的转染细胞的百分比测量阳性GFP表达细胞。对于优化实验,通过测量在用报告物质粒和pUC(没有recCas9或引导RNA表达载体)共转染时产生eGFP的转染细胞的百分比来确定测定背景。然后,当与recCas9和中靶或非靶引导RNA表达载体共转染报告物质粒时从观察到的eGFP阳性细胞的百分比中减去该背景。

[0291] 基因组靶位点的鉴定

[0292] 使用Bioconductor(使用R统计程序的开源生物信息学包)进行搜索适当的靶位点(Fu et al.,High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.Nature biotechnology 31,822-826(2013),其全部内容在此通过引用并入)。由Genome Reference Consortium发布的人参考基因组的最新版本(GRCh38)用于搜索与Cas9的PAM需求和如文本中描述的演化的gix序列两者相匹配的位点。在将基因组加载到R中的情况下,每个搜索模式都表示为Biostring(允许字符串匹配和操作的R中的容器)。当使用GRCh38参考组装进行搜索时,使用所述参数扫描整个基因组的两条DNA链揭示了人基因组中大约450个潜在靶标(表9)。

[0293] 表9:计算机中鉴定的recCas9基因组靶标

[0294]

染色体	开始	结束	序列	模式ID	SEQ ID NO:
chr1	34169027	34169103	CCTTTAGTGAAAAGTAGACAGCTCTGAATATGAAAGGTAGGTTTTCATTTCTGGGAAAGAGACGCCAAGTGATGTGG	2	211
chr1	51006703	51006780	CCTCCAATAAATATGGGACTATGTGGAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACGGGAAGAATGG	1	212
chr1	89229373	89229450	CCATTCTGCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTAGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	213
chr1	115638077	115638154	CCATTCTCCCCGTCACTTTCAGGTACAACAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	214
chr1	122552402	122552478	CCTTGTAGTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGACTTGAAACACTCTTGTTGTGG	2	215
chr1	122609874	122609950	CCTTGTAGTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCATACTTGAAACACTCTTTTTGTGG	2	216
chr1	122668677	122668753	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGACTTGAAACACTCTTTTTGTGG	2	217
chr1	123422419	123422495	CCTTGTGTTGTGTTTATTCAACTCACAGAGTTAAACGATCCTTTACA	2	218

[0295]

			CAGAGCAGACTTGAAATACTCTT TTTGTGG		
chr1	123648614	123648690	CCTTGTAGTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCATACTTGAAACACTCTT TTTGTGG	2	219
chr1	123806335	123806411	CCTTGTATTGTGAGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT TTTGTGG	2	220
chr1	124078228	124078304	CCTTGTGTTGTGTGTCTTCAACTC ACAGAGTTAAACGATGCTTTACA CAGAGTAGACTTGAAACACTCTT TTTCTGG	2	221
chr1	124231074	124231150	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGTAACACTCTT TTTGTGG	2	222
chr1	124232435	124232511	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACGTGAAACACTCTT TTTGTGG	2	223
chr1	124344781	124344857	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT TTTGTGG	2	224
chr1	124435716	124435792	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGGAGACTTGTAACACTCTT TTTGTGG	2	225
chr1	158677186	158677262	CCTGAGGTTTTCCAGGTTTTAAA AGGAAACCTAAAGGTAGGTTTAG CATTAAGTGTCTTGAAGTTTATTT TAAAAGG	2	226
chr1	167629479	167629554	CCAAAATTCCCACAAAACCGAAT GCATCAGTCAAAGCAAGGTTTGA AGAAAAGATTTACCACTTCAGGG AGCTTGG	4	227
chr1	167783428	167783504	CCTTTTCTGGATATCGTTGATGCT CTGTATGCAAAAGGTAGGTTTTT GGGTTATGTTGTTAAACAGTGAT TGAATGG	3	228
chr1	169409367	169409444	CCTCCAAGAAATATGGAACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAG AGAGAATGG	1	229
chr1	174145346	174145423	CCTCCAAGAAATATGGGACTATG TGAGAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGATGG GGAGAATGG	1	230
chr1	1837501	1837502	CCATTCTCCCCATCGCTTTCAGGT	1	231

[0296]

	68	45	ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTTCCATATTCT TTGGAGG		
chr1	200801540	200801617	CCATTCTCCCCATCACTTTCAGGT GTACCGATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	232
chr1	207589936	207590013	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACGG GGAGAATGG	1	233
chr1	209768370	209768445	CCTTCAGGGCAGAAACAGCTCTA CTAGCAGAGAAAGCAAGCTTTCA ATATTGTGCAATACAAAAACGAG AGCAGGG	4	234
chr1	218652378	218652455	CCATTCTCCTCATCTCCTTCTGGT ACTCCAATCAAACGTAGGTTTGG TCTTTTCTCATAGTCTCATATTTC TTGGAGG	1	235
chr1	222147250	222147327	CCTCCAAGACATATAGGACTATG TGAAAATACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAG GGAGTATGG	1	236
chr1	245870710	245870785	CCTGCCAGATACCAGTAGTCACT GTGAATTACAAAGCTACGTTTCT TCCATAGGGAAAGTTTGGAGTCC AGCCAGG	4	237
chr2	2376037	2376114	CCATTCTCCCTGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	238
chr2	4119629	4119706	CCATTCTCCCCACCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGTAGG	1	239
chr2	4909047	4909124	CCTAACCAGAAACTAACTAATAG ATATGGGCAGAAAGCATCCTTTC ACTTTTGTTCTGGGAGAGGGAAG AAGCAAAGG	1	240
chr2	28984877	28984953	CCATTTTGGGGAGGCCTTGATGG GAAGCTGGAAAAGGAAGCTTTCC TCCCAGTCCTGCTGAAGGCCTTG CCAGCTGG	2	241
chr2	31755833	31755910	CCTCCAAGAAACACAGGACTATG TGAAAAGATCAAACCTACGTTTG ATTGGTGTTCCTGAAAGTGATGG GGAGAATGG	1	242
chr2	39829583	39829660	CCATTCTCTTCATGACTTTCAGGT ACACCATTGAAACGTAGGTTTGG TCTTTTCACATTGTCCCATATTTC TTGGAGG	1	243

[0297]

chr2	60205947	60206024	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCGTATTTCTTGGTGG	1	244
chr2	79082362	79082439	CCATTCTCCCTGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGGGG	1	245
chr2	79082362	79082438	CCATTCTCCCTGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGGG	3	246
chr2	108430915	108430992	CCTCCAAGAAATATGAGATTATATGAAAAGACCAAACCTACGTTTGATTGGTGTACTTTAAAGTGACGGGGAGAATGG	1	247
chr2	115893685	115893762	CCATTCTCCCCGTCATTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCAAATTTCTTGGAGG	1	248
chr2	119620068	119620145	CCCCCAAGAAATGTGGGACTATATGAAAAGACCAAACCTACGTTTGACTGGTGTACCTAAAAGTGATGGGGAGAATGG	1	249
chr2	119620069	119620145	CCCCAAGAAATGTGGGACTATATGAAAAGACCAAACCTACGTTTGACTGGTGTACCTAAAAGTGATGGGGAGAATGG	2	250
chr2	128495068	128495144	CCCATTGGTGCTGACCAGATGGTGAAGGAGGCAAAGGTTGCTTTGAATGACTGTGCTCTGGGGTGAGCCAGGCCTGG	2	251
chr2	133133559	133133634	CCCTTTACAGAGGTGAGCTTTGTTATTAGTAAAAAGGTAGGTTTCCCTGTTTTTCTGAAGAAAAGCTGTGAGTGGG	4	252
chr2	134174983	134175060	CCACTGCCCATTGACAGAGTGGCGAGGTGGGTGAAACCTTGCTTTCCTCCTGGCCCATGGGCAGGGTGGGGCTGTGGG	1	253
chr2	134174983	134175059	CCACTGCCCATTGACAGAGTGGCGAGGTGGGTGAAACCTTGCTTTCCTCCTGGCCCATGGGCAGGGTGGGGCTGTGG	3	254
chr2	138069945	138070022	CCATTCTCCCTGTCACTTTTAGATACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATGTTTCTTGGAGG	1	255
chr2	138797420	138797496	CCTCCAAGAAATATCAACTGTGTGAAAAGACGAAACCTACGTTTGATTAATGTACCTGAAAGTGACAGG	2	256

[0298]

			GAGAATGG		
chr2	145212434	145212511	CCATTCTCCCATTAACTTTCAAGTACACCAATCAAAGGTAGGTTTGGTGTTTTCCCATAGTCCCGTATTTCTTGGAGG	1	257
chr2	147837842	147837919	CCTTTTCATCATGCCCCTTTCACTTTAAGGTGAAAACCTTGCTTTACATGTCAGAGAAAAGAAGAGCCCTCAGCTGGG	1	258
chr2	147837842	147837918	CCTTTTCATCATGCCCCTTTCACTTTAAGGTGAAAACCTTGCTTTACATGTCAGAGAAAAGAAGAGCCCTCAGCTGG	3	259
chr2	154152540	154152617	CCATTCACCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	260
chr2	157705943	157706019	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATGGTGTACCCGAAAGTGACAGGGAGAATGG	3	261
chr2	158361152	158361229	CCACCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATAGGTATACCTGAAAGTGACAGGGAGAATGG	1	262
chr2	161461006	161461083	CCATTCTCCCCATCACTTTCAGGTGCACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	263
chr2	179077376	179077453	CCCTCAAGAAATATGAGACTATGTGAAAAGACCAAACCTACGTTTGACTGGTATACCTGAAAGTGACAGGGAGAATGG	1	264
chr2	179077377	179077453	CCTCAAGAAATATGAGACTATGTGAAAAGACCAAACCTACGTTTGACTGGTATACCTGAAAGTGACAGGGAGAATGG	2	265
chr2	181090699	181090776	CCTCCAACAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACGGGGATAATGG	1	266
chr2	182331957	182332034	CCATTCTCTCCCTCACTTTCAAGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCTTATATTTCTTGGCGG	1	267
chr2	183620562	183620638	CCATTCTCCCTGTCACTGTCAGTACACCAATCAAACGTAGGTTTGGTCTCTTCACATAGTCCCATATTTCTTGGAGG	2	268
chr2	207345927	207346003	CCTCCAAGAAATATGGGACTATGTGAACAGACCAAACCTACGTTTG	3	269

[0299]

			ATTGGTGTACCTGAAAGTGATGG CAGAATGG		
chr2	216652047	216652123	CCACCATGCCTGGCCACCACACA TTTTTTTCTAAAGCTTGGTTTTGG CCACAGTGAGAGTTTCTTGGGCT GTCAGGG	2	270
chr2	216652047	216652122	CCACCATGCCTGGCCACCACACA TTTTTTTCTAAAGCTTGGTTTTGG CCACAGTGAGAGTTTCTTGGGCT GTCAGG	4	271
chr2	223780040	223780116	CCCACTAGGTGGCGATATCTGAG GGTCCAATGAAACCATGCTTTTT ACTCAGATCTTCCACTAACCACC TCCCCCGG	2	272
chr2	224486595	224486672	CCTCTAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ACTGGTGTACCTGAAAGTGACGG GGAGAATGG	1	273
chr2	230526902	230526979	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTAGTGTACCTGAAAGTGACGG GGAGAATGG	1	274
chr2	232036127	232036204	CCATTCTCCCTGTCACTTTCAGGT ACATCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	275
chr3	4072812	4072889	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ACTGGTGTACCTGAAAGGGATGG GGAGAATGG	1	276
chr3	9261677	9261754	CCCCCAAGAAATATGAGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAG GGAGAATGG	1	277
chr3	9261678	9261754	CCCCAAGAAATATGAGACTATGT GAAAAGACCAAACCTACGTTTGA TTGGTGTACCTGAAAGTGACAGG GAGAATGG	2	278
chr3	16732146	16732223	CCTCTAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTAACTGAAAGTGACAG GGAGAATGG	1	279
chr3	17450712	17450789	CCTCCAAGAAATATGCGCCTATG TGAAAAGACCAAACCTACGTTTG ATTGGTATACCTGAAAGTGATGG AGAGAATGG	1	280
chr3	21559769	21559846	CCATTCTCCCTGTCACTTTGAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATATTCGCATATTTC TTGGAGG	1	281
chr3	2341665	2341673	CCATTCTCCCCGTCACTTTCAGGT	1	282

[0300]

	8	5	ACACCAACCAAACGTTGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG		
chr3	2998401 9	2998409 6	CCATTCTCCCTGTCACTTTCCAGT ACACCAGTCAAACGTAGGTTTGG TCTTTTCACATACTCCCATATTTC TTGGAGG	1	283
chr3	3826955 1	3826962 7	CCTGGCCTAATTTTTAATTCTTAG TTTGACTTAAACCTTGCTTTTAGT GTGATGGCGACAAAAGCTGAGCT GAAAGG	2	284
chr3	4051521 3	4051528 8	CCAGTGCTTTTTGGTTTTAAAGG CAAGCCTCCAAACCTTCCTTTCTC CTGGATGCTGTGGTGGTTGCCAT GCATGG	4	285
chr3	4923361 2	4923368 7	CCCAACTCCTGCGAGAAGTAGCT CACCATGACAAAGCTACCTTTGC TTTTATCGTTTTGCAAAACAAAA AAGGGGG	4	286
chr3	6629289 4	6629297 1	CCATTCTCCCCGTCACTTTGAGGT GTGCCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCTATATTTC TTGGAGG	1	287
chr3	6754149 3	6754157 0	CCTCCAAAAAATATGGGACTACG TAAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAACTGACAG GGAGAATGG	1	288
chr3	8227301 1	8227308 8	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTTCCATATTTC TTGGAGG	1	289
chr3	9868334 9	9868342 6	CCTACAAGATATATGGGACTATG TGAAAAGACCAAACCTACGTTTT ACTGGTGTGCCTGAAACTGACGG GGAGAATGG	1	290
chr3	1019236 53	1019237 30	CCATTCTCTCTGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	291
chr3	1145334 67	1145335 44	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTC ATTGGTGTACCTGAAAGTGATAG GGAGAATGG	1	292
chr3	1326076 02	1326076 79	CCTCCAAAAAATATGGGATGATG TGAAAAGACCAAACCTAGGTTTG ACTGGTGTACCTGAAAATGATGG GGAGAATGG	1	293
chr3	1375451 76	1375452 53	CCTCCAAGAAATATGAGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAG GGAGAATGG	1	294

[0301]

chr3	137655679	137655756	CCTCCAAGAAATATGGGACTACGTGAAAAGATCAAACCTACGTTTGATTGTTGTACCTGAAAGTGATGGGGAGAATGG	1	295
chr3	137662040	137662117	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGTTGTACCTGAAAGTGATGGGGAGAATGG	1	296
chr3	142133796	142133873	CCTCAAAAGTGTTCTGGTTTTGTTTTGTTTTTTAAACCATGGTTTTACCTCTGGCTTAGTGGGACTAAAAATAGGAGG	1	297
chr3	146726949	146727026	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGACTGGTGTACCTGAAAGTGATGGGGAAAATGG	1	298
chr3	152421096	152421173	CCTCCAAGAAATATGGGACTGTGTGTAAAGACCAAACCTACGTTTGATTGGTGTACCTCAAAGTGATGGGGAGAATGG	1	299
chr3	170620247	170620324	CCATTCTCCCCATCACATTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	300
chr3	181166873	181166949	CCCCTGGAAAAGTTGGAGCATCACAGGAAAAGCAAACCAACCTTTTTTCTCCCCTAGGTAAACTGGGGAGCCAGGGG	3	301
chr3	181166874	181166949	CCCTGGAAAAGTTGGAGCATCACAGGAAAAGCAAACCAACCTTTTTTCTCCCCTAGGTAAACTGGGGAGCCAGGGG	4	302
chr4	6604233	6604309	CCTTCCCCAGTTGCAGCAGACAAGAGTCTCGAAAAGCTTGCTTTGGTTGCTGCAGTGGATGGGTTGGTAGGCACAGG	2	303
chr4	6626269	6626344	CCCCCACCTCCCAAGCTGCTGGCTTCTCGAATAAAGCTACCTTTCCTTTTACCAAAACTTGTCTCTCGAATGTCGG	4	304
chr4	8155396	8155472	CCTTGGCCCTGGACAGCTGCTTTTCCTTCCCTAAACCTTGGTTTCCCCCTTTGTGCAGGTGGGTGGGTTTGGGCTGG	2	305
chr4	10386803	10386880	CCTCTTCTAGTGAACCCATGGGGTTACCAAGGGAAAGCAACCTTTTGATAAATATTCCCATCTTTTTATGTTGTCTGG	1	306
chr4	20701579	20701656	CCACTTGAAAGGGTTACCAAGGATAAGATTTTTAAAGCTTGCTTTCACAAACAACTCATGCTCCAGGCTT	1	307

[0302]

			GTCAGTGG		
chr4	29594286	29594363	CCTTTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGATCTTTTCACATAGTCCCATATTTCTTGGAGG	1	308
chr4	53668422	53668499	CCATTCTCCCCATCAATTTCAGTTACACCAATGAAACGTAGGTTTGGCCTTTTCACATAGTCCCATATTTCTTAGAGG	1	309
chr4	74914802	74914879	CCATTCTCCCTGTCACTCTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCATATAGTCCCATATTTCTTGGAGG	1	310
chr4	75332783	75332859	CCTCCAAGAAAATTGGGACTATGTGAAAAAACCAAACCTACGTTTGATTGATGTACCTGAAAGTGACAGGAGAATGG	3	311
chr4	88123643	88123720	CCTTCAAGAAATATGGGACTATGTGAAAGGACAAAACCTACGTTTTATTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	312
chr4	89567192	89567269	CCATTCTCCCCATCACTTTCAGGTACGCTAATCAAACGTAGGTTTGATCTTTTCACATAGTCTTATATTTCTTGGAGG	1	313
chr4	93556577	93556654	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGACTGGTGTACCTCAATGTGACAGGGAGAATGG	1	314
chr4	100266379	100266456	CCATTCTCCCTGTCACTTTTAGGTACACCAATCAAACGTACGTTTGGTCTTTTCACATAGACCCATATTTCTTGGAGG	1	315
chr4	103486234	103486311	CCTTCAAGAAATATGGGACTGTGTGAAAAGACCAAAGCTAGGTTTGATTGGTGTACCTGAAAGTGATGGGGAGAATGG	1	316
chr4	105923129	105923204	CCTACTATTCACAGAGTAATGCAGTTTGCTGAAAAGGTTGGTTTTTGCTGACCTCTGAGAGCTCACATTACAGTGG	4	317
chr4	106874711	106874788	CCATTCTCTCTGTCACTTTCTGGTACACCAATCAAACGTAGGTTTGCTCTTTTCACATAATCCCATATTTATTGAAGG	1	318
chr4	115805791	115805867	CCATAACATGTATTTGCTGGTGCTAGACTCTCCAAAGCTAGGTTTCTTTCTACAACAATGGCTGGAAGTCTTCTTGG	3	319
chr4	122033277	122033354	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGG	1	320

[0303]

			TCTTCTCACACAGTCCCATATTTC TTGGAGG		
chr4	129125132	129125209	CCATTCTTCCCATTACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCACATTTC TTGGAGG	1	321
chr4	135472562	135472639	CCATTCTCCCCCTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATTGTCCCATATTTC TTGGAGG	1	322
chr4	138507099	138507176	CCATTCTCCCCAGCACTTACAGG TACACCAATCAAACGTAGGTTTG GTCATTTCACATAGTCCCATATTT CTTGGAGG	1	323
chr4	144249093	144249170	CCATTCTCCCTGTCACTTTCAGGT ACAGCAATCAAACGTAGGTTTGG TCTTTTCACATGGTCCCATATTTC TTGGAGG	1	324
chr4	144436406	144436483	CCTCCAAGAAATATGAGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACGG GGAAGATGG	1	325
chr4	154110259	154110336	CCTCCAAGAAATATGAGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAG GGAGAATGG	1	326
chr4	154893438	154893515	CCTCCAAGAGATATGAGACTATG TAAATAGACCAAACCTACCTTTG ATTGGTGTACGTGAAAGTGACAG GAAGAATGG	1	327
chr4	161116854	161116931	CCATTCTCCCCATCACTTTCAGGT ACACCAACCAAACGTAGGTTTGG TCTTTTCACATAGTCTCATATTTC TTGGAGG	1	328
chr4	165140748	165140823	CCTCCATTGACTACTCCTTATCAT TGGCTAGAAAACCTACCTTTCAA CCAGTTTCTAAGGCCAAGAAACT TGGAGG	4	329
chr4	181928508	181928585	CCACCAAGAAATATGGGACTACG TGAAAAGACCAAACCTACGTTTG ATGGGTGTGCCTGAAAGTGACGG GAAGAATGG	1	330
chr4	187521958	187522035	CCTCCAAGAAATAAGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAGGTGACAG GGAGAATGG	1	331
chr5	12675639	12675715	CCAAAGGGCCTTTGTGATTCTAC TTTGTAATATAAAGGATGGTTTC TTACTACGGTTGGTGTCCTTGCA GGAGTGGG	3	332
chr5	2927180	2927188	CCTCCAAGAAATATGGGACTATG	1	333

[0304]

	4	1	TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGATGG GGAGAATGG		
chr5	35352660	35352737	CCATTCTCCCCGTTACTTTCAGGT ACACCAATAAAACCTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	334
chr5	38723235	38723310	CCCATATCTCTGGCAAGGGCAGC TCTCTGGCTAAACCAAGCTTTCC TGTAGAGCTTGAGTTCCAAGGCA GCGTTGG	4	335
chr5	47358339	47358415	CCTTGTAGTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT GTTGTGG	2	336
chr5	47415811	47415887	CCTTGTAGTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCATACTTGAAACACTCTT TTTGTGG	2	337
chr5	47474614	47474690	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT TTTGTGG	2	338
chr5	48228356	48228432	CCTTGTGTTGTGTTTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAATACTCTT TTTGTGG	2	339
chr5	48454551	48454627	CCTTGTAGTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCATACTTGAAACACTCTT TTTGTGG	2	340
chr5	48612272	48612348	CCTTGTATTGTGAGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT TTTGTGG	2	341
chr5	48884165	48884241	CCTTGTGTTGTGTGTCTTCAACTC ACAGAGTTAAACGATGCTTTACA CAGAGTAGACTTGAAACACTCTT TTTCTGG	2	342
chr5	49037011	49037087	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGTAACACTCTT TTTGTGG	2	343
chr5	49038372	49038448	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACGTGAAACACTCTT TTTGTGG	2	344
chr5	49150718	49150794	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT TTTGTGG	2	345

[0305]

chr5	49241653	49241729	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGGAGACTTGTAACACTCTTTTTGTGG	2	346
chr5	88582714	88582790	CCTTTTCATAAGAAGAAAATCGACTCATCATTGAAACCAAGCTTTGGTACAATTTCATTGATGTTTCCAGAAGCAGG	3	347
chr5	93497156	93497231	CCCATAGACTATGATAGAAACAAAATAACCCAAAAGCTAGCTTTCTGATTGAGTTTCCATAAATGCAATGTGAAGG	4	348
chr5	94295029	94295105	CCATTCACTTGTCACTTTCTGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCTCATATTTCTTGGAGG	2	349
chr5	94956746	94956823	CCTCCAAGAAATATGGGACTCTGTAAAGAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGAAGGGGAGAATGG	1	350
chr5	106003488	106003565	CCATTCTCCCCGTCATTTTCAGGTACACCAATCAAACCTAGGTTTGGTCTTTTTACATAGTCCCATATTTCTTGGAGG	1	351
chr5	118727905	118727982	CCTCCACGAAACATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	352
chr5	132156032	132156109	CCAATTTCCCCCTCACTTTCAGATACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTTCCATATTTCCTGGAGG	1	353
chr5	152037951	152038028	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATATTCCCATATGTCTTGGAGG	1	354
chr5	155183064	155183141	CCCACCGGCTCATGAGAGGTAGAGCTAAGGTCCAAACCTAGGTTTATCTGAGACCGGAACTCATGTGATTAACTGTGG	1	355
chr5	155183065	155183141	CCACCGGCTCATGAGAGGTAGAGCTAAGGTCCAAACCTAGGTTTATCTGAGACCGGAACTCATGTGATTAACTGTGG	2	356
chr5	163148211	163148288	CCTTCAAGAAATATGGGACTATGTGAAGAGACCAAACCTACGTTTGATTGGTGTAGCCAAAAGTGATGGGGAAAATGG	1	357
chr5	165889537	165889614	CCTCAGATTAGATTTACTTGCAAAGAGACATTTAAAGGATCGTTTTGATACTATTTTGAAAGTACTATA	1	358

[0306]

			CAAAGATGG		
chr5	169395198	169395274	CCTTAAGAACATAAATCCCCAGGAATTCACAGAAACCTTGGTTTGAGCTTTGGATTTCCCGCAGGATGTGGGATAGG	2	359
chr5	171021380	171021457	CCATTCTCTCTGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCTCATAGTCCCATATTTCTTGGAGG	1	360
chr5	173059898	173059973	CCATTTACCATCATTCTCTGTCATGGCAGGTGAAAGCAAGCTTTTATATAGACAATGTTCTACTTAGTTTACAGGG	4	361
chr5	174102359	174102435	CCCAAAGTTAATTTTACTCTTTTTCTGAATCAAAAGGAACCTTTCCTCCATGAGAAGAATCCTGCCATATTTCTAGG	2	362
chr5	180927811	180927888	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGCTATACATGAAAGTGACGGGGAGAATGG	1	363
chr6	1752363	1752440	CCTTCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACCTTTGATTGGTGTACCTGAAAGTGATGGGAAGAATGG	1	364
chr6	20595279	20595356	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATAGTTCTTGGAGG	1	365
chr6	234313710	234314447	CCATTCTCCCCGTCACTTTCAGGGACAACAATCAAACGTAGGTTTGGCCTTTGCACATAGTCTTATATTTCTTGGAGG	1	366
chr6	29190624	29190701	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	367
chr6	61533266	61533343	CCTCCAAAAAATATGGGACTATGTGAGAAGACCAAACCTACGTTTTATTAGTGTACCTCAAAGTGACAGGGAGGATGG	1	368
chr6	101052764	101052841	CCATTCTCCCCATCACTTTCAGGTACACCAATGAAACGTAGGTTTGGCCTTTTCACATAGTTTCATATTTCTTGGAGG	1	369
chr6	117176355	117176432	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGATGGGGAGAATGG	1	370
chr6	1177470	1177471	CCTACAAGAAATATGGAACTTGT	2	371

[0307]

	73	49	AAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACGGGGAGAATGG		
chr6	118422508	118422585	CCTCCAAGAAATATGGGACAATGTGAAAAGGCCAAAGCTACGTTTGATTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	372
chr6	122035019	122035096	CCTTTCAAACTTAGAGGTAAACAAAAGTCCTGAAAACCTAGGTTTGACCATAAGTTGGGACCATACGAGCATAGAAGG	1	373
chr6	134445210	134445287	CCAAAAATAAAAAAAAAATTGACTTATAAGTAAGAAAGGTTCGTTTTCTCACATTCAGAAAGAGAACCCACATGTTGGG	1	374
chr6	134445210	134445286	CCAAAAATAAAAAAAAAATTGACTTATAAGTAAGAAAGGTTCGTTTTCTCACATTCAGAAAGAGAACCCACATGTTGG	3	375
chr6	135154944	135155021	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	376
chr6	137889995	137890072	CCATTCTCCCCGTCACTTTCAGGTACACCAATCAAACGTTGGTTTAGTCTATTCACATAGTCCCATATTTCTTGGAGG	1	377
chr6	143993904	143993981	CCGAAAAGAATAAGACTATCAGCTGAAGTCTTAAAACGATCCTTTGGCCCCCAGTACTCTATATGCAGGATAGAAAGG	1	378
chr6	152610473	152610549	CCTACAAAAATAGGGGACTATGTGATAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGATGGGGAGAATGG	2	379
chr6	160372604	160372681	CCATTCTACCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGCCTTTTCATATAGTCTCATATTTCTTGGAGG	1	380
chr6	169352478	169352555	CCATTCTCCCCATCACTTTCTGGTATACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTAGAGG	1	381
chr6_GL000251v2_alt	677196	677273	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	382
chr6_GL000252v2_alt	456242	456319	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	383

[0308]

chr6_GL000253v2_alt	456202	456279	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	384
chr6_GL000254v2_alt	456371	456448	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	385
chr6_GL000255v2_alt	456225	456302	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	386
chr6_GL000256v2_alt	500011	500088	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	387
chr7	5256551	5256627	CCACCACACCCAGCCTTATGGGATGGTTTTCAAAAGCATCCTTTTTTAGAAGTGGATTCTGATATATAATCGGATGG	2	388
chr7	7392583	7392660	CCATTCTCAATGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	389
chr7	8737741	8737818	CCATTCTCTCTGTCACTTTCAGGTACACCAGTCAAAGGTAGGTTTGTTTTATTCACACGTTCACATATTTCTTGGAGG	1	390
chr7	11352226	11352303	CCATTCGCCCCATCACTTTCAGGTACACTAGTAAAACGTAGGTTTGGTCTTTTCACATAGTTCCATATTTCTTGGAGG	1	391
chr7	15519145	15519222	CCTCCAAGAAATATGGGACTATGTGAAGAGATCAAACCTAGGTTTGATTGTTGTACCTGAAAGTGATAAGAAGAATGG	1	392
chr7	19228341	19228418	CCTCCAATAAATATGGGGCTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	393
chr7	23778445	23778522	CCCTTTTCCCTGTCACTTTCAGGTACACCAGTCAAACGTAGGTTTGGTCTTTTCACATAGTCGAATATTTCTTCAAGG	1	394
chr7	23778446	23778522	CCTTTTCCCTGTCACTTTCAGGTACACCAGTCAAACGTAGGTTTGGTCTTTTCACATAGTCGAATATTTCTTCAAGG	2	395
chr7	26769065	26769142	CCATTCTCCCTGTCACTTTCAGGTACACTAATCAAACGTAGGTTTGGTGTATTCACACAGTCCCATATTTC	1	396

[0309]

			TTGGAGG		
chr7	42864035	42864112	CCATTCTTCCTGTCACTTTCAGGTATACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATGTTTCTTGGAGG	1	397
chr7	46498923	46499000	CCTCCAAGAAATATGAGACTATATGAAAATACCAAACCTACGTTTGATTGGTGTACCTGAAAGAGACAGGGAGAATGG	1	398
chr7	51535360	51535437	CCATTCTCCCTATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCATGTAGTCCCATATTTCTTGGAGG	1	399
chr7	51927106	51927183	CCATTCTGCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	400
chr7	56976942	56977018	CCGTCCGATTATATATCAGAATCTACTTCTAAAAAAGGATGCTTTTGAAAACCATCCCATAAGGCTGGGTGTGGTGG	3	401
chr7	80021598	80021675	CCTACAAGGAATATAGGACTATGTGAAAATACCAAACCTACGTTTCACTGCTGTACCTGAAGGTGACAGGGAGAATGG	1	402
chr7	89673853	89673930	CCATTCTCCCCATCATTTCCAGGTAAACCAATCAAAGGTAGGTTTGGTCATTTCACATAGTCCCATATTTCTTGGAGG	1	403
chr7	103404790	103404867	CCATTCTCCCCGTCACTTTCAGGTACACCAGTCAAACGTAGGTTTGGTCTTTTCACACAGTCCCATATTTCCTGGAGG	1	404
chr7	113053651	113053728	CCATTCTCCCCATCACTTTCAGGTACAGCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	405
chr7	125765204	125765279	CCACTACAGATTCTTGGGTCAAGATGTGTGCAAAAGGATGCTTTAGGGTGATGGATATGAGTGGGATGAAATGAGG	4	406
chr7	128042158	128042234	CCTGAAAAAAAACCCTGCCAGCCAGCAACTCTGAAAGGATGCTTTGTGTGAGTGAGCAGTGTCTGAGATGGACAGGG	3	407
chr7	130637332	130637409	CCATTCTCCCCATCACTTTCAGGTACGCCAATCAAACGTAGGTTTGGTCTTTTGACATAGTCCCATATTTCTTGGAGG	1	408
chr7	136983050	136983127	CCGTTCTCCCCATCACTTTTAGGTACACCAATCAAACGTAGGTTTGG	1	409

[0310]

			TCTTTTCACATAGTCTCATATTTC TTGGAGG		
chr7	143579507	143579584	CCATTCTCCTGGTCACTTTCAGGT ATACCAATCAAACGTAGGTTTGG TCTTTTCATGTAGTCCCATATTTC TTGGAGG	1	410
chr7	143749881	143749958	CCTCCAAGAAATATGGGACTACA TGAAAAGACCAAACCTACGTTTG ATTGGTATACCTGAAAGTGACCA GGAGAATGG	1	411
chr8	2338364	2338441	CCTCCAAGAACTATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACGG GGAGAATGG	1	412
chr8	2383289	2383366	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATAGTTC TTGGAGG	1	413
chr8	8414568	8414645	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACAGAGTCCCATATTTC TTGGAGG	1	414
chr8	24163142	24163219	CCATTCTCCCCGTCACTTTCATGT ACACCAAGCAAACGTAGGTTTGA TCTTTCCACATAGTCCCGTGTTTC TTGGAGG	1	415
chr8	34299051	34299128	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACTTGAAAGTGACAG GGAGAATGG	1	416
chr8	40965485	40965562	CCTCCAAGAAATATGGGACTATG TGAAAAGACAAAACCTACGTTTC ACTGGTGTACCTGAAAGTGACAG GGAGGATGG	1	417
chr8	48371659	48371735	CCCCCACCTTTTAAAAACATGCA TACATACGGAAACGTTGCTTTCT GCACGATTTCATTTTAATGGAAC AGAACAGG	2	418
chr8	82534960	82535037	CCATTTCCCCTGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTATCATATTTC TTGGAGG	1	419
chr8	109217624	109217700	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TGGAGG	3	420
chr8	134790285	134790361	CCTTTTGTTAAAGTAATAGAATT CTGCTTCTTAAAGGAACCTTTCA GGCAAGATGGTGGTTAGAGCACC TAAATGGG	2	421
chr8	1347902	1347903	CCTTTTGTTAAAGTAATAGAATT	4	422

[0311]

	85	60	CTGCTTCTTAAAGGAACCTTTCA GGCAAGATGGTGGTTAGAGCACC TAAATGG		
chr8_KI270821v1_alt	519635	519712	CCTCCAAGAACTATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACGG GGAGAATGG	1	423
chr8_KI270821v1_alt	564557	564634	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG CCTTTTCACATAGTCCCATAGTTC TTGGAGG	1	424
chr9	14951207	14951283	CCTCCAAGAAATATGGGACTGGT GAAAAGACCAAACCTACGTTTGA CTGGTGTACCTGAAAGTGACGGG GAGACTGG	2	425
chr9	23249218	23249295	CCTCCAAGAAACATGGGAATGTG TGAAAAGACCAAACCTACGTTTG ATTGGCGTACCTGAAAGTGACGG GGAGTATGG	1	426
chr9	26278896	26278973	CCTCCAAGAAATATGGGACTGTG TGAAAAGACCAAACCTACGTTTG ATTGGTATACCTGAAAGTGACAG AGAGAATGG	1	427
chr9	27323237	27323314	CCATTCTCCCCTTCACTATCAGGT ACACCAATCAAACGTAGGTTTAG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	428
chr9	31517993	31518070	CCATTCTCCCCGTCACTTTCAGAT ACACCAGTCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	429
chr9	39694860	39694937	CCATCTTACTTTGTACTACACTGT TCTTTAGAGAAAGCTTCCTTTTG GAGACCAACCAGGACTCCTTAGA AGCAGAGG	1	430
chr9	42451132	42451209	CCATCTTACTTTGTACTACACTGT TCTTTAGAGAAAGCTTCCTTTTG GAGACCAACCAGGACTCCTTAGA AGCAGAGG	1	431
chr9	60776573	60776650	CCTCTGCTTCTAAGGAGTCCTGG TTGGTCTCCAAAAGGAAGCTTTC TCTAAAGAACAGTGTAGTACAAA GTAAGATGG	1	432
chr9	62647482	62647559	CCTCTGCTTCTAAGGAGTCCTGG TTGGTCTCCAAAAGGAAGCTTTC TCTAAAGAACAGTGTAGTACAAA GTAAGATGG	1	433
chr9	66682030	66682107	CCTCTGCTTCTAAGGAGTCCTGG TTGGTCTCCAAAAGGAAGCTTTC TCTAAAGAACAGTGTAGTACAAA GTAAGATGG	1	434

[0312]

chr9	82264427	82264503	CCACCACTGTGCCTGGCCATTTTCACTATTCTTAAAGGAAGCTTTGGTTTACAAAGGTTTGCTACTGTACTTCCAGG	3	435
chr9	84042684	84042761	CCATTCTCCCTGTCACTTTCAGGTACACCATTCAAACGTAGGTTTGGTCTTTTCTCATAGTCCCATATTTCTTGGAGG	1	436
chr9	95256012	95256089	CCTCCAAGAAATTCGGGACTATGTGAAAAGACAAAACCTACGTTTAATTGGTGTGTGGTGTACCTGAAAGTGACAAGG	1	437
chr9	101816988	101817065	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACCAGAAGAATGG	1	438
chr9	135842327	135842403	CCTCCAAGAAATATGGGACTATGTGAAAAGCCCAAACCTACGTTTGACTGATGTACCTAAAGTGACGGGGAGAATGG	3	439
chr9	136910865	136910940	CCCGCACTGTGAGCTTGGCCGAGTGCTGTCTGAAAGCATCCTTTCCCTTCACCTGGAGACTGGAGCGCCATAGAGG	4	440
chr10	13710312	13710389	CCTGTCTCCCCCATTCCATGCAAAATAAAACACAAACCAAGCTTTGCTTTAAGTGCTCCCTGATGCAGTTCAGCGTGG	1	441
chr10	18938129	18938206	CCATTCTTCCCGTCACATTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCCCATAGTCCCATATTTCTTAGAGG	1	442
chr10	22712838	22712914	CCCCCTGCTCAGCTTGGGGAAGAAAAATACAAAAACGATGCTTTTAGGCATTTTAAACAACTTCACTACATTGAGGG	2	443
chr10	22712838	22712913	CCCCCTGCTCAGCTTGGGGAAGAAAAATACAAAAACGATGCTTTTAGGCATTTTAAACAACTTCACTACATTGAGG	4	444
chr10	40160932	40161009	CCTTTGTGTTGTGTGTATTCAACTCACAGAGTGAAACCTTCCTTTATTCAGAGCAGTTTTGAAACACTCTTTTTGTGG	1	445
chr10	40390136	40390213	CCTTTGTGTTGTGTGTATTCAACTCACAGAGTGAAACCTTCCTTTATTCAGAGCAGTTTTGAAAAACACTTTTTGTGG	1	446
chr10	40409152	40409229	CCTTTGTGTTGTGTGTATTCAACTCACAGAGTGAAACCTTCCTTTATTCAGAGCAGTTTTGAAAAACTCT	1	447

[0313]

			TTTTGTGG		
chr10	40433940	40434017	CCTTTGTGTTGTGTGTATTCAACTCACAGAGTGAAACCTTCCTTTATTCAGAGCAGTTTTGAAACACTCTTTTTGTGG	1	448
chr10	40588155	40588232	CCTTTGTGTTGTGTGTATTCAACTCACAGAGTGAAACCTTCCTTTATTCAGAGCAGTTTTGAAATACTCTTTTTGTGG	1	449
chr10	41146207	41146284	CCTTTGTGTTGTGTGTATTCAACTCACAGAGTGAAACCTTCCTTTATTCAGAGCAGTTTTGAAACACTCTTTTTGTGG	1	450
chr10	43835183	43835260	CCATTCTCCCTGTCACTTTCAAGTACACCAATCAAACCTAGGTTTGGTCTTTTCACATAGTTCCATATTTCTTGGAGG	1	451
chr10	54913222	54913299	CCCCTCCCATCACAGGCCCTGAGGTTTAAGAGAAAACCATGGTTTTGTGGGCCAGGCCCATGACCCTTCTCCTCTGGG	1	452
chr10	54913222	54913298	CCCCTCCCATCACAGGCCCTGAGGTTTAAGAGAAAACCATGGTTTTGTGGGCCAGGCCCATGACCCTTCTCCTCTGG	3	453
chr10	54913223	54913299	CCCTCCCATCACAGGCCCTGAGGTTTAAGAGAAAACCATGGTTTTGTGGGCCAGGCCCATGACCCTTCTCCTCTGGG	2	454
chr10	54913223	54913298	CCCTCCCATCACAGGCCCTGAGGTTTAAGAGAAAACCATGGTTTTGTGGGCCAGGCCCATGACCCTTCTCCTCTGG	4	455
chr10	58035951	58036028	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTCATCTTTTCACATAGTCCCACGGTTTTTGGAGG	1	456
chr10	58677525	58677602	CCTCCAAGATATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAATTGATGGGGAGAATGG	1	457
chr10	84021390	84021467	CCTCCAAGAAATATGGGACTGTGTGAAAAGAACAAACCTACGTTTGATTGGTGTACGTGAAAGTGATGGGGAGAATGG	1	458
chr10	91442692	91442769	CCATTCCTCCCGTCACTTTCAGATACACCAAAAAAACGTAGGTTTGGTCTCTTCACATAGTCCCACATTTCTTGGAGG	1	459
chr10	91446848	91446925	CCTCCAAGAAATGTGGGACTATGTGAAGAGACCAAACCTACGTTTT	1	460

[0314]

			TTTGGTGTATCTGAAAGTGACGG GAGGAATGG		
chr10	116928784	116928860	CCTCCAAGGGGAATCTGAGTTCT CTGAAGACAAAAAGCATGGTTTC TTTTCTTCTGTATTTCTTATTGTTT CCTAGG	3	461
chr10	116937771	116937848	CCATTCTCCCTATCACTTTCCAGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	462
chr11	31182070	31182147	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTATACTTGAAATTGACAA GGAGAATGG	1	463
chr11	34739273	34739350	CCTCCAAGAAATATGGGACTATG TGGAAAGACCAAACCTACGTTTG ACTGGTGTACCTGAAAGTGATGG GGAGAATGG	1	464
chr11	86646529	86646606	CCTCTAAGAAATATGGGACTATG TGAAGAGATGAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACGA GGAGAATGG	1	465
chr11	90469791	90469867	CCCTCGTATACTACATGCTATAG TCAAAGCAGTAAACCTTCCTTTC CTTAAGCAGACCACACTCTTTCA TGCCTGGG	3	466
chr11	90469792	90469867	CCTCGTATACTACATGCTATAGT CAAAGCAGTAAACCTTCCTTTCC TTAAGCAGACCACACTCTTTCAT GCCTGGG	4	467
chr11	92429985	92430062	CCATTCTCCCCATCACTTTCAGGT ATACTAATCAAAGGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC ATGGAGG	1	468
chr11	102818498	102818574	CCATTCCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	2	469
chr11	120765065	120765142	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTTG TCTTTTCTTATAGTCCCATATTTC TTGGAGG	1	470
chr11	123131901	123131978	CCACTGCACCTGACCAAGATCCT TAATTTTTCTAAACCTACGTTTAT CATCTATAAAATGAGCCATCTTT TCACATGG	1	471
chr11	129468520	129468597	CCTCCGAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGTTGTACCTGAAAGTGACAG GGAGAATGG	1	472
chr11	1312723	1312724	CCATTCTCCCCATCACTTTTAGGT	1	473

[0315]

	61	38	ACACCAATCAAACGTAGGTTTGG TCCTTTTGCATAGACCCATATTTC TTGGAGG		
chr11	132761415	132761492	CCATTTTCCCCGTCAGTTTCATAT ACACCTATCAAACGTAGGTTTAC TGTTTTCACATAGTCCCTTATTTC TTGGAGG	1	474
chr12	22367416	22367493	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACCTTTG ATTGGTGTACCTGAAAGTGACGG GCAGGATGG	1	475
chr12	33146384	33146461	CCATTCTTCTCGTCATTTTCAAGT ACACCAATCAAACGTAGGTTTGG TCTTTTCGCATAGTCCCATATTTC TTGGAGG	1	476
chr12	33198476	33198553	CCATTCTTCTCGTCACTTTCAAGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	477
chr12	46038332	46038409	CCTCCAAGAAATATAGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACTTGAAAGTGACAG GGAGAATGG	1	478
chr12	60236126	60236203	CCTCCAAGAAATGTGGAACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAG GGAGAATGG	1	479
chr12	62098359	62098434	CCCTGACACTGATAAACGGATAT GAAGAGAAAAAAGCTAGGTTTTC GCTGGAATTCCTAAGCTTGGGCT GCAGTGG	4	480
chr12	62112591	62112668	CCCTTCTCCCAGTCACTTTTAGGT ACACCAATGAAACGTAGGTTTGG TCTTTTCACACAGTCCCATATTTC TTGGAGG	1	481
chr12	62112592	62112668	CCTTCTCCCAGTCACTTTTAGGTA CACCAATGAAACGTAGGTTTGGT CTTTTCACACAGTCCCATATTTCT TGGAGG	2	482
chr12	62418577	62418652	CCACTCCCTCTCCCCCAAAAAGT AAAGGTAGAAAACCAAGGTTTA CAGGCAACAAATAGCACAATGA ATGGAATGG	4	483
chr12	71732311	71732388	CCAAACCCGCATCGCACACCCTG TGAGGGGGACAAAGGAACCTTTC CGTTCCAACATCAAGGTTGTTTT GACCCAAGG	1	484
chr12	78047816	78047893	CCATTCTTTCTGTCACTTTCAGGT ATACCAGTCAAACCTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	485

[0316]

chr12	81480016	81480093	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	486
chr12	96840231	96840307	CCACACGGTAGAGGATAAACTAGGTGGATTCTCAAAGCAACCTTTGAAATAATCTATGCAGTTTTTCTGGGTACTGG	3	487
chr12	99187165	99187242	CCACCAAGAAACATGGGACTATGTGAAAAGACCAAACCTACGTTTGGTTGGTGTACCTGGAAGTGACGGGGAGAGTGG	1	488
chr12	107860841	107860918	CCTCCAAGAAATATGGGACCATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	489
chr12	110882809	110882885	CCTGTAAAAAGGTCACATGGTCAGGTGTGCCTAAACGATCCTTTTATTTATTTATTTATTTATTTTTAAGAAACAGG	2	490
chr12	119063321	119063397	CCAGCCCCAAAATGTCAGGGGCTTAGAACAACAAAGGTTCCTTTTCATGTTTATACTACATGTTTGTCATGGGCTGG	2	491
chr13	35320704	35320781	CCGTTTTCCCCATCACTTTCAGGTACACCAGTCAAACGTAGGTTTGGTCTTTTCACATGGTCCCACATTTCTTGGAGG	1	492
chr13	53133477	53133554	CCTGGAATAGCTTTCCTGACTGTCTGACTTCAAAAACCTTGGTTTGACCACTTCGTCTATATCATGAGGAAGGACTGG	1	493
chr13	53184880	53184956	CCCTACTCTGAACCTACCTTGATAAAGCCTAGAAAACCAAGCTTTGACAAGATTTGACAAGAGATGGAATTTGGAGG	3	494
chr13	53184881	53184956	CCTACTCTGAACCTACCTTGATAAAGCCTAGAAAACCAAGCTTTGACAAGATTTGACAAGAGATGGAATTTGGAGG	4	495
chr13	57896962	57897038	CCCTTATAAAACTGAAAACTTTAACCTTTTTTAAAGCATGCTTTTGAATAAATTCTTTTATTACAAAAAAGACCAGG	2	496
chr13	62610100	62610177	CCATTCTCCCTGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACGTAGTCCCATATTTCTTGGAGG	1	497
chr13	77004382	77004458	CCCTTTATTATCCAAGTGGTTTCCTGCTCTTCAAACCTTCCTTTCAAAATTTTGTCTCCTACTTAAAACAA	2	498

[0317]

			GTTAGG		
chr13	81646075	81646151	CCTTCTGTTGAGACCTACTGCTAAGAAAACAAAAAAGGTTCCTTTCAAATATTATTGTGAATCAATAATGTACCTGG	3	499
chr13	83755854	83755931	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTCATTGATGGACCTGAAAGTGATGGGGAGAATGG	1	500
chr13	89719199	89719275	CCATTCTCCCTTCACTTTCAGTTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	2	501
chr13	102010574	102010650	CCTAGGGAAGTGATCATAGCTGAGTTTCTGGAAAAACCTAGGTTTTAAAGTTGAGGAGACTTAAGTCCAAAACCTGG	3	502
chr13_KI270841v1_alt	124240	124316	CCATTCTCCCTTCACTTTCAGTTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	2	503
chr14	25980646	25980723	CCTCCAAGAAATATGGGACTATGTGAAAAGACTAAACCTACGTTTGATTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	504
chr14	35842786	35842863	CCATTCTCCCTGTCACTTTCAGGTATGCCAGTCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTCCTTGGAGG	1	505
chr14	42646400	42646477	CCTCCAAGAAATATGGGACTATGTAAAAAGACGAAACCTACGTTTGATTGGTGTACTTAAAAGTGACGAGGAGAATGG	1	506
chr14	49063242	49063319	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTTGTGTACCTGAAAGTGATGGGGAGAATGG	1	507
chr14	49130379	49130456	CCATTCTCCCCGTCACTTTCAGGCACACCAATCAAACGTAGGTTTAGTCTTTTCACATAGTCCCATATTTCTTAGAGG	1	508
chr14	51352342	51352418	CCTTAATGCATTCATATTTCATATTTTAAATAAAACCATGGTTTCCCACAGAGTGACTTCTACTCTAAGAAATGGGG	2	509
chr14	51352342	51352417	CCTTAATGCATTCATATTTCATATTTTAAATAAAACCATGGTTTCCCACAGAGTGACTTCTACTCTAAGAAATGGG	4	510
chr14	60835842	60835919	CCGTTCTTTCCGTCACTTTCAGGTACACCAGTCAAACGTAGGTTTGG	1	511

[0318]

			TCTTTTCACATAGTCCCATATTTC TTGGAGG		
chr14	66529072	66529148	CCATTCTCCCCATCACTTTCATGT ACACCAATCAAACGTAGGTTTGG TCTTTGTTAACATAGTCCCATATT TCTTGG	3	512
chr14	79210873	79210949	CCCTATAAAGCTTAGAGAAACAC AGGGCTCTTTAAACGATCCTTTTT CTCTTTTCTGTTTTAAATTTCATC ACTTGG	3	513
chr14	79210874	79210949	CCTATAAAGCTTAGAGAAACACA GGGCTCTTTAAACGATCCTTTTTC TCTTTTCTGTTTTAAATTTCATCA CTTGG	4	514
chr14	85371541	85371618	CCATTCTCCCCATCACTTTCAGGT ACACTAATCAAAGGTAGGTTTGG TCTTTTCACATGGTCCTATATTTC TTGGAGG	1	515
chr14	92918713	92918790	CCCCATAGCACGATCACATGGGA CATTCAGGGGAAAGCAACCTTTT CCAGGAAGGAAAACCCAATGCT GGGACCCAGG	1	516
chr14	92918714	92918790	CCCATAGCACGATCACATGGGAC ATTCAGGGGAAAGCAACCTTTTC CAGGAAGGAAAACCCAATGCTG GGACCCAGG	2	517
chr14	103386821	103386897	CCCTTTCAGCGCTCACAGGCTAT GGTTTTATAAAAGGAACCTTTGA TTTTGTTCATGTGAAACTACAAA ATGCCAGG	2	518
chr14_KI270847v1_alt	33275	33352	CCCCATAGCACGATCACATGGGA CATTCAGGGGAAAGCAACCTTTT CCAGGAAGGAAAACCCAATGCT GGGACCCAGG	1	519
chr14_KI270847v1_alt	33276	33352	CCCATAGCACGATCACATGGGAC ATTCAGGGGAAAGCAACCTTTTC CAGGAAGGAAAACCCAATGCTG GGACCCAGG	2	520
chr15	20630566	20630643	CCTCCAAGAAATATTGGAGTATG TGATAAGACCAAACCTTCGTTTG ACTGGTGTACCTGAAAGTGATGG GGAGAATGG	1	521
chr15	21675103	21675180	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	522
chr15	22117571	22117648	CCATTCTCCCCGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	523
chr15	2236974	2236982	CCATTCTCCCCATCACTTTCAGGT	1	524

[0319]

	4	1	ACACCAGTCAAACGAAGGTTTGGTCTTATCACATACTCCAATATTTCTTGGAGG		
chr15	42302832	42302909	CCTCCAAGATATATGGGACTATGTGAAAAGGCCAAACCTACCTTTGATTGATACACCTGAAAATGACAGGGAGAATGG	1	525
chr15	49967601	49967678	CCTCCAAGAAATATGCGACTATGTGAAAAGACCAAACCTACGTTTCATTGGTGTACCTGAAAGTGATGGGGAGAATGG	1	526
chr15	83964501	83964577	CCTCCAAGAAATATGGGACTATGTGGAAAGACCAAACCTACGTTTGTTTGGTGTACCTGAAAGTGAGGGGAGAATGG	3	527
chr15	87261388	87261465	CCATTCTCCTCATCACTTTCAAGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCTTATATTTCTTGGAGG	1	528
chr15_KI270727v1_random	409348	409425	CCATTCTCCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	529
chr15_KI270851v1_alt	14235	14312	CCATTCTCCCCATCACTTTCAGGTACACCAGTCAAACGAAGGTTTGGTCTTATCACATACTCCAATATTTCTTGGAGG	1	530
chr15_KI270852v1_alt	440099	440176	CCATTCTCCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	531
chr16	22123671	22123748	CCAGCAGAAGAATCTGGGGCACAGTCTGTGAAAAAAGGTACCTTTCTTAAGCAGGGTTCTTATCCTTCATGGGTCTGG	1	532
chr16	25557623	25557700	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGTTGTACCTGAAAGTGAGGGGGAGAATGG	1	533
chr16	36427179	36427255	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	534
chr16	36476450	36476526	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	535
chr16	36512469	36512545	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	536

[0320]

chr16	36520964	36521040	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACACAGCAGATTTGAAACACTGTTTTTCTGG	2	537
chr16	36524704	36524780	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	538
chr16	36566812	36566888	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	539
chr16	36573603	36573679	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	540
chr16	36667694	36667770	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	541
chr16	36677320	36677396	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	542
chr16	36683096	36683172	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	543
chr16	36691251	36691327	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	544
chr16	36710951	36711027	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	545
chr16	36750364	36750440	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	546
chr16	36791455	36791531	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACACAGCAGATTTGAAACACTGTTTTTCTGG	2	547
chr16	36856683	36856759	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	548
chr16	36926655	36926731	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTT	2	549

[0321]

			TTTCTGG		
chr16	3693175 2	3693182 8	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	550
chr16	3694805 8	3694813 4	CCTTGTGTTGTGTGTATTCAACTC ACCGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	551
chr16	3697454 1	3697461 7	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	552
chr16	3698133 1	3698140 7	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	553
chr16	3699083 9	3699091 5	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	554
chr16	3702107 5	3702115 1	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	555
chr16	3704281 2	3704288 8	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	556
chr16	3708597 1	3708604 7	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	557
chr16	3712946 2	3712953 8	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	558
chr16	3714611 0	3714618 6	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CACAGCAGATTTGAAACACTGTT TTTCTGG	2	559
chr16	3715730 9	3715738 5	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	560
chr16	3718311 8	3718319 4	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	561
chr16	3719092 4	3719100 0	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA	2	562

[0322]

			CAGAGCAGATTTGAAACACTGTT TTTCTGG		
chr16	3722180 8	3722188 4	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	563
chr16	3725950 1	3725957 7	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	564
chr16	3727240 9	3727248 5	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	565
chr16	3728192 3	3728199 9	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGTAACACTGTT TTTCTGG	2	566
chr16	3734647 2	3734654 8	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	567
chr16	3735700 0	3735707 6	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	568
chr16	3737330 1	3737337 7	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	569
chr16	3741949 8	3741957 4	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	570
chr16	3743071 4	3743079 0	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	571
chr16	3745584 5	3745592 1	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	572
chr16	3745855 8	3745863 4	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	573
chr16	3748612 7	3748620 3	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	574
chr16	3752518	3752525	CCTTGTGTTGTGTGTATTCAACTC	2	575

[0323]

	3	9	ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTGTGG		
chr16	3753673 5	3753681 1	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	576
chr16	3755473 0	3755480 6	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	577
chr16	3757578 4	3757586 0	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	578
chr16	3757748 3	3757755 9	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	579
chr16	3758359 8	3758367 4	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	580
chr16	3769636 8	3769644 4	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTCCA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	581
chr16	3770452 4	3770460 0	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	582
chr16	3770622 3	3770629 9	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	583
chr16	3770894 1	3770901 7	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	584
chr16	3776362 2	3776369 8	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	585
chr16	3777211 5	3777219 1	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	586
chr16	3779181 5	3779189 1	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGATTTGAAACACTGTT TTTCTGG	2	587

[0324]

chr16	37796229	37796305	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	588
chr16	37797928	37798004	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	589
chr16	37843453	37843529	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	590
chr16	37848548	37848624	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	591
chr16	37864846	37864922	CCTTGTGTTGTGTGTATTCAACTCACCGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	592
chr16	37902550	37902626	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	593
chr16	37907307	37907383	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	594
chr16	37928033	37928109	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	595
chr16	37959262	37959338	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	596
chr16	37964355	37964431	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	597
chr16	37974881	37974957	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAAAACTGTTTTTCTGG	2	598
chr16	37987789	37987865	CCTTGTGTTGTGTGTATTCAACTCACAGAGTAAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	599
chr16	37994586	37994662	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTT	2	600

[0325]

			TTTGTGG		
chr16	38006479	38006555	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	601
chr16	38011567	38011643	CCTTGTGTTGTGTGTATTTAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	602
chr16	38040096	38040172	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	603
chr16	38041456	38041532	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	604
chr16	38062179	38062255	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	605
chr16	38102937	38103013	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	606
chr16	38128412	38128488	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	607
chr16	38131809	38131885	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	608
chr16	38144723	38144799	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	609
chr16	38168845	38168921	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	610
chr16	38209287	38209363	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	611
chr16	38210986	38211062	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGATTTGAAACACTGTTTTTCTGG	2	612
chr16	38229667	38229743	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACA	2	613

[0326]

			CAGAGCAGATTTGAAACACTGTTTTTCTGG		
chr16	47424037	47424114	CCATTCTCCCTATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	614
chr16	60730549	60730625	CCTCGTCACTGCCAGATTTTGTGGCTACCAGCAAAGGATCGTTTTAAGCTGCAACTCAGGAAATTGAGAAAATATGG	2	615
chr16	72545014	72545091	CCTCCAAGAAATATGGGACTATGTGAAAAAACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	616
chr16	81945503	81945579	CCCTGTGTTCTTTTATACTAAAACAAGCCAGCAAACCAACCTTTGAGATGTGTTGCCTTAAACATTACTGAATGGGG	2	617
chr16	81945503	81945578	CCCTGTGTTCTTTTATACTAAAACAAGCCAGCAAACCAACCTTTGAGATGTGTTGCCTTAAACATTACTGAATGGG	4	618
chr17	16474024	16474100	CCGAGAAACGGCTTTAGCAACAAATAAATATCAAAAGGATGCTTTCTCTTCAGAATAATCTAAAGTAAGTTGGGAGG	3	619
chr17	34438512	34438589	CCATGTTACTCCGGATAAGGACAGCAAAGGAGGAAAGGAACCTTTTCTGGGCCACCAGAAGGATGAGCTTGGGCTTGG	1	620
chr17	43690782	43690859	CCCAGGGATATGCTGGCCACGGGGAGGAGCCGGAAACCAACCTTTGTGTCACTGTGTAGTGACAAGTGCCTTTGGAGG	1	621
chr17	43690783	43690859	CCAGGGATATGCTGGCCACGGGGAGGAGCCGGAAACCAACCTTTGTGTCACTGTGTAGTGACAAGTGCCTTTGGAGG	2	622
chr17	69156298	69156375	CCTTAGGGACCCATAATGGCCACAACCAGGAGAAAAGCAAGCTTTGATGCTTAAACACTACTTACAGACATGTACAGG	1	623
chr17	74595228	74595305	CCTGCCTCTGTTCCTCCTTCCTGATGGTGGCGGAAAGGATGCTTTTGCCAGATCAACAGTCACACACAACACACCAGG	1	624
chr17	83191644	83191721	CCTGACTCCAGCCCTCCTTGACAAGGTCTCCGTAAAGCATGCTTTCTCTTAGGGACCCTCAGAGGGAGGCTTGGTGGG	1	625
chr17	8319164	8319172	CCTGACTCCAGCCCTCCTTGACA	3	626

[0327]

	4	0	AGGTCTCCGTAAAGCATGCTTTC TCTTAGGGACCCTCAGAGGGAGG CTTGGTGG		
chr18	3513522 4	3513530 0	CCTTATTTGGAATGTGACAAGAC CCATTTGTTTAAACCTTGGTTTTT ATGCAGAAAGAAAAGGAAGGCT GCAGTGGG	3	627
chr18	3891886 1	3891893 8	CCATTCTCCCTGTCACTTTCAGGT ACACTAATCAAACGTAGGTTTGC TGTTTTTACATAGGCTCATATTTC TTGGAGG	1	628
chr18	4547658 9	4547666 6	CCATTCTCCCCATCACTTTCAGGT ACACCAGTCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	629
chr18	4864082 1	4864089 6	CCTGTTTGTTATTTTAGCTAATGT CAAAAAGAAAACCTTGCTTTTTC TGAACCCTTTCAGAGGCAGAAAG TGGGGG	4	630
chr18	7109673 2	7109680 8	CCATTTTCCCCACCACTTTCACGT ACAGCAATCAAACGTAGGTTTGG TCTTTTCACTAGTCCCATATTTCT TGGAGG	3	631
chr19	2495784 4	2495792 0	CCTTGTAGTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT GTTGTGG	2	632
chr19	2501531 6	2501539 2	CCTTGTAGTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCATACTTGAAACACTCTT TTTGTGG	2	633
chr19	2507411 9	2507419 5	CCTTGTGTTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT TTTGTGG	2	634
chr19	2582786 1	2582793 7	CCTTGTGTTGTGTTTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAATACTCTT TTTGTGG	2	635
chr19	2605405 6	2605413 2	CCTTGTAGTGTGTGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCATACTTGAAACACTCTT TTTGTGG	2	636
chr19	2621177 7	2621185 3	CCTTGTATTGTGAGTATTCAACTC ACAGAGTTAAACGATCCTTTACA CAGAGCAGACTTGAAACACTCTT TTTGTGG	2	637
chr19	2648367 0	2648374 6	CCTTGTGTTGTGTGTCTTCAACTC ACAGAGTTAAACGATGCTTTACA CAGAGTAGACTTGAAACACTCTT TTTCTGG	2	638

[0328]

chr19	26636516	26636592	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGACTTGTAACACTCTTTTTGTGG	2	639
chr19	26637877	26637953	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGACGTGAAACACTCTTTTTGTGG	2	640
chr19	26750223	26750299	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGCAGACTTGAAACACTCTTTTTGTGG	2	641
chr19	26841158	26841234	CCTTGTGTTGTGTGTATTCAACTCACAGAGTTAAACGATCCTTTACACAGAGGAGACTTGTAACACTCTTTTTGTGG	2	642
chr19	28517220	28517297	CCAGGAAAAAATTTAAACTTTCTTAACTTGATAAAAGGTAGCTTTCAAAACCTACAATAAATAACATACTTAGAGTGG	1	643
chr19	34566821	34566898	CCATTCTCCTCGTCACTTTCAGGTACACCAAACAAACGTAGGTTTGGTCTTTTTACGTAGTCCCATATTTCTTGGAGG	1	644
chr19	52261770	52261847	CCCTCTTGAAGTTAGGGAAGTAGCATTTAAGGGAAACGTAGCTTTACTATTAAGAATTTCAAACAGCACTTGTCAGGG	1	645
chr19	52261770	52261846	CCCTCTTGAAGTTAGGGAAGTAGCATTTAAGGGAAACGTAGCTTTACTATTAAGAATTTCAAACAGCACTTGTCAGG	3	646
chr19	52261771	52261847	CCTCTTGAAGTTAGGGAAGTAGCATTTAAGGGAAACGTAGCTTTACTATTAAGAATTTCAAACAGCACTTGTCAGGG	2	647
chr19	52261771	52261846	CCTCTTGAAGTTAGGGAAGTAGCATTTAAGGGAAACGTAGCTTTACTATTAAGAATTTCAAACAGCACTTGTCAGG	4	648
chr20	11151392	11151469	CCATTCTCCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATATTCCCATATTTCTTGGAGG	1	649
chr20	14027067	14027143	CCATTCTCCCTTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTTTTGGAGG	2	650
chr20	50615399	50615476	CCTATAGTCTCAGTTACTTGGGAGGCTGAGGTAAAAGGATCGTTTGAGCCCAGGAGGTGGAGGTTGCA	1	651

[0329]

			GTGAGCCGGG		
chr20	50615399	50615475	CCTATAGTCTCAGTTACTTGGGAGGCTGAGGTAAAAGGATCGTTTGAGCCCAGGAGGTGGAGGTTGCAGTGAGCCGG	3	652
chr20	60909414	60909490	CCTTTCCCAACTCTGCTATTGCCCCCACATCCTAAAGGAACCTTTCTTTTTTTATATATTTTATTTTAAGTTCCAGG	3	653
chr21	16226086	16226163	CCTCCAAGAAATATGGAACTATGTGAAAAGACCAAACCTACGTTTGATTGACGTACCTGAAAGTGACAGGGAGAATGG	1	654
chr21	17835234	17835309	CCTCTTCTGAAAGCATTGATAATCAACATTTTAAACGTAGCTTTTCCCCATATTGCTAGGAAGGCTCATTCCCGGG	4	655
chr21	19425636	19425713	CCTCCAAGAAATATGGGACTATGTGAAAAGGCCAAACCTACGTTTGATTGCTGTACCCGAGAGTGACGGGGAGAATGG	1	656
chr21	32220958	32221035	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGATGGGGAGAATGG	1	657
chr21	34335877	34335953	CCCGGGGCCTGGGTGCCCAGTGCCAGTGGTCAGAAAGGTTGCTTTGGTGTTTTTCATTGTTAGTGAGACAGAGATGG	3	658
chr21	34335878	34335953	CCGGGGCCTGGGTGCCCAGTGCCAGTGGTCAGAAAGGTTGCTTTGGTGTTTTTCATTGTTAGTGAGACAGAGATGG	4	659
chr21	36315276	36315353	CCATTCTCCCCATCATTTTCAGGTACACCAATCAAACGTAGGTTTGATCTTTTCACATAGCCCCATATTTCTTGGAGG	1	660
chr21	41547952	41548028	CCACCAGCACTTCTGTTAGAAGTTGCAGCAGAGAAAGGATCCTTTAGGCACATCTCCCAGATCCTTGCGAAGAGGGG	3	661
chr22	18973194	18973271	CCTGTGCCAGGGTCCTTCCACTGGGACTGGCAGAAACGTAGGTTTGCATGGAGTGAGAAGCAGGGGAGAGGTTGAGGG	1	662
chr22	18973194	18973270	CCTGTGCCAGGGTCCTTCCACTGGGACTGGCAGAAACGTAGGTTTGCATGGAGTGAGAAGCAGGGGAGAGGTTGAGG	3	663
chr22	20265462	20265539	CCCTCAGCCTCTCCCCTGCTTCTCACTCCATGCAAACCTACGTTTCT	1	664

[0330]

			GCCAGTCCCAGCAGAAGGACCCTGGCACGGG		
chr22	20265462	20265538	CCCTCAGCCTCTCCCCTGCTTCTCACTCCATGCAAACCTACGTTTCTGCCAGTCCCAGCAGAAGGACCCTGGCACGG	3	665
chr22	20265463	20265539	CCTCAGCCTCTCCCCTGCTTCTCACTCCATGCAAACCTACGTTTCTGCCAGTCCCAGCAGAAGGACCCTGGCACGGG	2	666
chr22	20265463	20265538	CCTCAGCCTCTCCCCTGCTTCTCACTCCATGCAAACCTACGTTTCTGCCAGTCCCAGCAGAAGGACCCTGGCACGG	4	667
chrX	27300998	27301075	CCTCCAAGAAATATGGGGCTATGTGAAAAGACCAAACCTACCTTTGATTGGTGTATCTGAAAGTGACGGGGAGAATGG	1	668
chrX	28456666	28456743	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTTGTGTACCTGAAAGTGATGGGGAGAATGG	1	669
chrX	35634985	35635062	CCATTCTCCCCGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCTCATTGTCCCATATTTCTTGGAGG	1	670
chrX	39460148	39460223	CCCATCAAGAGCGGTTGTGCATGGCAACAGTAAAAGGATGGTTTGTTACACTAGTACAAAAAGAGGTGGCCAGAGG	4	671
chrX	43926403	43926480	CCATTCTCTCTGTCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	672
chrX	44254600	44254677	CCTCCAAGAAATACGGGACTATGTGAAAAGACCAAACGTACGTTTGATTGGTGTACCTGAAAGTGATAGGGAGAATGG	1	673
chrX	46088602	46088679	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGTGTACCTGAAAGTGACTGGGAGAATGG	1	674
chrX	50222874	50222951	CCATTCTCCCTGTCACTTTCAGGTACACGAATCAAACGTAGGTTTCATCTTTTCACATAGTCCCATATTTCTTAGAGG	1	675
chrX	57416835	57416911	CCATTCTCTCTGTCACTTTCTGGTACACCAATCAAACGTAGGTTTGGTCTTTTCACATAGTTTCACATATTTCTTGG	3	676
chrX	5785646	5785654	CCTCCAAGAAATATGGGACTATG	1	677

[0331]

	6	3	TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAA GGAAAATGG		
chrX	6270247 9	6270255 6	CCTGAAAAACATTGTTTCCAACC TGGTAAATCAAAAGGAAGGTTTA ACTTTGTTAGATAAGTCCACATA TCACCAAGG	1	678
chrX	6306712 9	6306720 6	CCTCCAAGAAATGTGGGACTATG GGAAAAGACCAAACCTACCTTTG TTTGGTGTACCTGAAAGTGACGG GGAGAAAGG	1	679
chrX	6493625 0	6493632 7	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTC ATTGGTGTACCTGAAAGTGATGG GTAGAATGG	1	680
chrX	6672009 9	6672017 6	CCTACAAGAAATATGGGACTATG GGAAAAGACCAAACCTACGTTTG ATTGGTACACTGGAAAGTGACAG GGATAATGG	1	681
chrX	6852908 6	6852916 3	CCATTCTCCCTGTCACTTTCTGGT ACACCAATCAAAGGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	682
chrX	7389399 4	7389407 1	CCTCCAAGAAATATGGGACTATG TGAAAAGACCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGATGG GGAGAATGG	1	683
chrX	7572320 1	7572327 8	CCATTCTCTTTGTCACTTTCAGGT ATACCAATCAAACGTTGGTTTGG TCTTTTTGCATAGTCCCATATTTT GTGGAGG	1	684
chrX	7581565 9	7581573 6	CCTCCAAGAAATATGAGACTATG TGAAAAGACCAAACCTACGTTTG ATTAGTGTACCTGAAAATGATGG GGAGAATGG	1	685
chrX	8096710 3	8096718 0	CCATTCTTTCTGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	686
chrX	8993642 5	8993650 2	CCATTCTCCCTGTCACTTTCAGGT ACACCAATCAAACGTAGGTTTGT TCTTTTCACATAGTCCCATATTTC TTGGAGG	1	687
chrX	9103876 8	9103884 5	CCATTATCCCCATCACTTTCAGGT ACACCAATCAAACGTAGGTTTGG TTTTTTCACATAGTTCAATATTTC TTTGAGG	1	688
chrX	9147127 1	9147134 8	CCTCCAAGAAATATGGGACTATC TGAAAAGATCAAACCTACGTTTG ATTGGTGTACCTGAAAGTGACAG GGAGAATGG	1	689

[0332]

chrX	96428180	96428257	CCTTTCTCCCCATCACTTTCAGGTACACCAATCAAACGTAGGTTTGGTCTTTTCATATAGTCCCATATTTCTTGGAGG	1	690
chrX	100268291	100268368	CCTCCAAGAAATATGGGACTATGTGCAAAGATCAAACCTACGTTTGATTGCTGTACCTGAAAGTGATGGGGAGAATGG	1	691
chrX	105811046	105811123	CCATTCTCCCCATCACTTTCAGGTACACCAGTCAAACGTAGGTTTGGTCTTTTCACATAATCCCATATTTCTTGGAGG	1	692
chrX	115673065	115673141	CCTCCAAGAAGTATGGGACCATGGAAAAGATCAAACCTACGTTTGACTGGTGTACCTGAAAGTGACTGGGAGAATGG	2	693
chrX	117269846	117269923	CCTCCAAGAAATATGGGACTATGTGAAAAGACCAAACCTACGTTTGATTGGAGTACTTGAAAATGACAGGGATAATGG	1	694
chrX	139191369	139191445	CCTTTAAAGACATGCTCTTTGTGCCAGAAATTCAAAGGTTGCTTTTATGTCCAGTGGGGTGGAGGGAGGAAGCTCGG	3	695
chrX	147988614	147988691	CCATTCTCCCCGTCACTTTCAGGGACCTCAATCAAACGTAGGTTTTGTCTTTTCACATAGTCCCATATTTCTTGGAGG	1	696
chrX	155321041	155321118	CCTCCAAGAAATATAGGACTATGTGAAAAGACCAAACCTACGTTTGACTGGTGTACCTGAAAGTGACAGGGAGAATGG	1	697
chrY	15109391	15109468	CCATTCTCCCCATCACTTTCAGGTACACCAATCAAAGGTAGGTTTGGTCTTTTCACATAGTCCGATATTTCCTGCAGG	1	698

通过搜索$CCX_{(30-31)}$-AAASSWWSSTTT-$X_{(30-31)}$-GG (SEQ ID NO: 699)鉴定染色体位点，其中W是T或A并且S是G或C。模式1是$CCX_{(31)}$-AAASSWWSSTTT-$X_{(31)}$-GG (SEQ ID NO: 699)，2是$CCX_{(30)}$-AAASSWWSSTTT-$X_{(31)}$-GG (SEQ ID NO: 699)，3是$CCX_{(31)}$-AAASSWWSSTTT-$X_{(30)}$-GG (SEQ ID NO: 699)，并且4是$CCX_{(30)}$-AAASSWWSSTTT-$X_{(30)}$-GG (SEQ ID NO: 699)。仅显示+链并且开始和结束对应于染色体(GRCh38)中的第一个和最后一个碱基对或适用时的备用组装。

[0333] DNA测序

[0334] 如上述一式六份进行293T细胞的转染并温育72小时。收获细胞并合并重复。使用涉及碱性裂解和旋转柱纯化的改良HIRT提取来提取附加型DNA，基本上如所述(Quan et al.,Circular polymerase extension cloning of complex gene libraries and pathways.PloS one 4,e6441(2009)；和Hillson(2010),vol.2015,pp.CPEC protocol;其每篇的全部内容在此通过引用并入)。简而言之，收获后，将HEK293T细胞在500μL的冰冷的PBS中洗涤，重悬于250μL GTE缓冲液(50mM葡萄糖，25mM Tris-HCl，10mM EDTA并且pH8.0)

中,在室温温育5分钟,并用200μL裂解缓冲液(200mM NaOH,1%十二烷基硫酸钠)在冰上裂解5分钟。用150μL的乙酸钾溶液(5M乙酸盐,3M钾,pH 6.7)中和裂解。通过在21,130g离心15分钟沉淀细胞碎片,并将裂解物应用于Econospin旋转柱(Epoch Life Science,Missouri City,TX)。用750μL洗涤缓冲液(Omega Bio-tek,Norcross,GA)洗涤柱两次,并在45μLTE缓冲液,pH8.0中洗脱。

[0335] 用RecBCD(10U)按照制造商的说明将分离的附加型DNA在37℃消化2小时,并用MinElute Reaction Cleanup试剂盒(Qiagen,Valencia,CA)纯化到10μLEB中。用5μL的附加型提取物转化Mach1-T1化学感受态细胞,并在选择羧苄青霉素抗性的琼脂糖平板(含有50μg/mL羧苄青霉素)上铺板。用引物pCALNL-for-1对个别菌落进行测序以确定重组率。测序读段揭示了“左”完整的非重组recCas9位点、预期的重组产物、具有小插入/缺失的“左”非重组位点的罕见实例、或大缺失产物的一个情况。

[0336] 对recCas9催化基因组缺失的分析

[0337] 在24孔胶原处理的平板中以每孔6×10^5个细胞的密度接种HEK293T细胞并生长过夜(Corning,Corning,NY)。在Opti-MEM(ThermoFisher Scientific,Waltham,MA)中使转染反应物达到100μL的终体积。对于每次转染,将90ng的每种引导RNA表达载体、20ng的pmaxGFP(Lonza,Allendale,NJ)和320ng的recCas9表达载体与Opti-MEM(ThermoFisher Scientific,Waltham,MA)中2μL Lipofectamine2000组合并加入各孔中。48小时后,收获细胞并在BD FACS AriaIIIu细胞分选仪上分选GFP转染对照。使用100μm喷嘴在纯度模式下分选细胞,并通过与未转染的细胞比较确定背景荧光。在PBS中在冰上收集分选的细胞,沉淀并用冷PBS洗涤两次。使用E.Z.N.A.组织DNA试剂盒(Omega Bio-Tek,Norcross,GA)收获基因组DNA并在100μL Eb中洗脱。使用在Tecan Infinite M1000Pro荧光酶标仪(plate reader)上测量的Quant-iT PicoGreen dsDNA试剂盒(ThermoFisher Scientific,Waltham,MA)定量基因组DNA。

[0338] 使用补充有3%DMSO的Q5热启动聚合酶2xMaster Mix进行巢式PCR,并用HyClone水,分子生物学级(GE Life Sciences,Logan,UT)稀释。使用引物对FAM19A2-F1和FAM19A2-R1(表5)用20ng的基因组DNA作为模板以25uL规模进行初级PCR。初级PCR条件如下:98℃1分钟,35个循环(98℃10秒、59℃30秒、72℃30秒),72℃1分钟。使用引物FAM19A2-F2和FAM19A2-R2,将1:50稀释的初级PCR用作次级PCR的模板。次级PCR条件如下:98℃1分钟,30个循环(98℃10秒、59℃20秒、72℃20秒),72℃1分钟。通过在TAE中的1%琼脂糖凝胶上与1Kb Plus DNA梯(ladder)(ThermoFisher Scientific,Waltham,MA)一起电泳来分析DNA。在Qiagen Minelute柱(Valencia,CA)上使用制造商的方案纯化待Sanger测序的材料。来自3个生物学重复的模板DNA用于三个独立的基因组巢式PCR。

[0339] 考虑到一组完整的人染色体重约3.6pg($3.3\cdot10^9$bp$\times1\cdot10^{-1}$,计算检出限。因此,用20ng的基因组DNA模板接种的PCR反应物含有约5500组染色体。

[0340] 为了定量基因组缺失,对于3个生物学重复中的每一个,一式三份进使用上述条件行巢式PCR。使用两倍稀释系列的基因组DNA作为模板,从未稀释的储液开始(对于样品1,47.17ng/uL;对于样品2,75.96ng/uL;并且对于样品3,22.83ng/uL)以减少潜在的移液误差来源。假定可以观察到缺失PCR产物的最低DNA浓度含有每个总基因组DNA的单一缺失产物。

[0341] 可以推断给定量的模板DNA中存在的基因组的数量,并因此可以确定FAM19A2基因

座处recCas9的最小缺失效率的估计值。例如，以20ng基因组DNA模板开始，以两倍稀释系列为例。巢式PCR后，只有接种20ng的孔产生正确的PCR产物。每个基因组3.6pg，该PCR含有大约5500个基因组，并且由于必须存在至少一个重组基因组，因此最小缺失效率为5500分之一或0.018％。

[0342] 使用基因组模板的有限稀释来定量基因组DNA的水平，因为使用定量PCR(qPCR)来确定基因组编辑的绝对水平将需要一组PCR条件，其仅明确地且仅特异性地扩增自重组后基因组DNA。如图5B中所示，使用基因组DNA作为模板的初级PCR导致大约2.5kb脱靶带作为优势种类；需要使用巢式引物进行第二轮的PCR以揭示引导RNA-和recCas9依赖性基因组编辑。

[0343] 结果

[0344] 将Gin重组酶融合至dCas9

[0345] 最近已经证明dCas9的N端可以融合至FokI核酸酶催化域，产生二聚体dCas9-FokI融合物，其切割侧翼为两个引导RNA指定序列的DNA位点(参见例如，Guilinger et al., Fusion of catalytically inactive Cas9to FokI nuclease improves the specificity of genome modification.Nature biotechnology,(2014)；Tsai et al., Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.Nature biotechnology,(2014)；其每篇的全部内容在此通过引用并入)。使用相同的融合方向将dCas9与Ginβ(Barbas及其同事先前演化的二聚体Gin转化酶的高活性催化域)连接(Gaj et al.,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic acids research 41,3937-3946(2013)，其全部内容在此通过引用并入)。Ginβ混杂地重组与天然核心序列CTGTAAACCGAGGTTTTGGA(SEQ ID NO:700)相关的几个20-bp核心“gix”序列(Gaj et al.,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic acids research 41,3937-3946(2013)；Klippel et al.,The DNA Invertase Gin of Phage Mu-Formation of a Covalent Complex with DNA Via a Phosphoserine at Amino-Acid Position-9.Embo Journal 7,1229-1237(1988)；Mertens et al.,Site-specific recombination in bacteriophage Mu:characterization of binding sites for the DNA invertase Gin.The EMBO journal 7,1219-1227(1988)；Plasterk et al.,DNA inversions in the chromosome of Escherichia coli and in bacteriophage Mu:relationship to other site-specific recombination systems.Proceedings of the National Academy of Sciences of the United States of America 80,5355-5358(1983)；其每篇的全部内容在此通过引用并入)。引导RNA将recCas9二聚体定位于侧翼为两个引导RNA指定的序列的gix位点，使得Ginβ域能够以引导RNA编程的方式催化DNA重组(图1D)。

[0346] 为了测定所得的dCas9-Ginβ(recCas9)融合物，构建了含有两个recCas9靶位点的报告物质粒，两个recCas9靶位点位于阻断EGFP转录的poly-A终止子的侧翼(图1A-1C)。每个recCas9靶位点由侧翼为与引导RNA前间隔区序列匹配的位点的gix核心假位点组成。重组酶介导的缺失去除了终止子，恢复了EGFP的转录。用该报告物质粒、转录引导RNA的质粒和产生候选dCas9-Ginβ融合蛋白的质粒共转染HEK293T细胞，并使用表现出EGFP荧光的细

胞级分评估每种融合构建体的相对活性。

[0347] 影响recCas9组分的结构的参数，包括核心gix位点和引导RNA结合位点之间的间距(0至7bp)，以及dCas9和Ginβ部分之间的接头长度((GGS)$_2$(SEQ ID NO:182)、(GGS)$_5$(SEQ ID NO:701)或(GGS)$_8$(SEQ ID NO:183))是变化的(图2A-2F)。大多数融合结构导致无观察到的引导RNA依赖性EGFP表达(图1C-1D)。然而，当存在匹配但不错配的引导RNA时，一个含有8个GGS重复的接头和3-6个碱基对间隔区的融合构建体导致约1％的重组(图2E-2F)。当5-6个碱基对将dCas9结合位点与核心分开时，重组活性始终较高(图2F)。这些结果共同揭示了dCas9和Ginβ之间的特异性融合结构可以在人细胞中与间隔区侧翼的gix相关的核心位点处产生引导RNA依赖性重组活性。8xGGS接头融合构建体称为“recCas9”。

[0348] 用recCas9靶向人基因组中发现的DNA序列

[0349] 观察到的活性的低水平可以由次优的引导RNA序列或核心gix序列引起，与表明引导RNA:Cas9结合的效率呈序列依赖性的之前的报告(参见例如，Xu et al.,Sequence determinants of improved CRISPR sgRNA design.Genome research 25,1147-1157(2015)，其全部内容在此通过引用并入)一致。此外，尽管目前的优化是用天然gix核心序列进行的(参见例如，Klippel et al.,The DNA Invertase Gin of Phage Mu-Formation of a Covalent Complex with DNA Via a Phosphoserine at Amino-Acid Position-9.Embo Journal 7,1229-1237(1988);Mertens et al.,Site-specific recombination in bacteriophage Mu:characterization of binding sites for the DNA invertase Gin.The EMBO journal 7,1219-1227(1988);Plasterk et al.,DNA inversions in the chromosome of Escherichia coli and in bacteriophage Mu:relationship to other site-specific recombination systems.Proceedings of the National Academy of Sciences of the United States of America 80,5355-5358(1983);其每篇的全部内容在此通过引用并入)，但是一些研究已表明锌指-Gin或TALE-Gin融合对略微改变的核心位点是有活性的，并且在某些情况下更具活性。参见例如，Gordley et al.,3rd,Synthesis of programmable integrases.Proceedings of the National Academy of Sciences of the United States of America 106,5053-5058(2009);Gersbach et al.,Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase.Nucleic acids research 39,7868-7878(2011);Mercer et al.,Chimeric TALE recombinases with programmable DNA sequence specificity.Nucleic acids research 40,11163-11172(2012);Gaj et al.,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic acids research 41,3937-3946(2013);Gordley et al.,3rd,Evolution of programmable zinc finger-recombinases with activity in human cells.J Mol Biol 367,802-813(2007);Gersbach et al.,3rd,Directed evolution of recombinase specificity by split gene reassembly.Nucleic acids research 38,4198-4206(2010);和Gaj et al.,Structure-guided reprogramming of serine recombinase DNA sequence specificity.Proceedings of the National Academy of Sciences of the United States of America 108,498-503(2011);其每篇的全部内容在此通过引用并入)。因此，靶向在人基因组内发现的序列以测试未修饰的人基因组序列是否能够被recCas9靶

向并测试是否改变引导RNA和核心序列将增加recCas9活性。

[0350] 为了鉴定潜在的靶位点，使用了表征演化的Gin变体的先前发现(参见例如，Gaj et al.,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic acids research41,3937-3946(2013)，其全部内容在此通过引用并入)以及上述观察结果。使用该信息，在人基因组中搜索含有$CCN_{(30-31)}$-AAASSWWSSTTT-$N_{(30-31)}$-GG(SEQ ID NO:699)的位点，其中W是A或T，S是G或C，N是任何核苷酸。$N_{(30-31)}$包括以下各项的N:NGG前间隔区相邻基序(PAM)、20碱基对Cas9结合位点、Cas9和gix位点之间5-6碱基对间隔以及gix核心位点的四个最外碱基对。预先确定gix核心位点的内部12个碱基对(AAASSWWSSTTT，SEQ ID NO:699)对Ginβ活性是重要的(参见例如，Gaj et al.,Nucleic acids research 41,3937-3946(2013)。

[0351] 该搜索揭示了人基因组中约450个此类基因座(表9)。产生报告物构建体，其含有与PCDH15中发现的这些基因组基因座之一相同的序列，并且然后构建引导RNA表达载体以将recCas9导向该序列(图3A)。这些载体编码两对引导RNA，其每对含有与PCDH15假gix位点侧翼的5'和3'区域匹配的间隔区序列。报告物质粒、这些侧翼引导RNA表达载体的组合和recCas9表达载体的共转染导致在11％-13％的转染细胞中的EGFP表达(图3B)，表明与图2中所示结果相比，活性提高了10倍以上。这些发现表明，recCas9靶序列的更明智的选择可以导致与人基因组中发现的DNA序列匹配的DNA序列处的重组效率显著提高。

[0352] 接下来，确定是否需要这两个引导RNA序列引起recCas9介导的缺失。用靶向PCDH15假gix核心位点的5'或3'侧翼序列的引导RNA载体的仅一个、PCDH15报告物质粒和recCas9表达载体共转染HEK293T细胞。这些共转染导致2.5-3％EGFP表达(图3B)。靶向性引导RNA的仅一个和recCas9的表达时观察到的低活性水平可以是由过度激活的gix单体形成二聚体的倾向引起的(参见例如，Gaj et al.,Enhancing the Specificity of Recombinase-Mediated Genome Engineering through Dimer Interface Redesign.J Am Chem Soc 136,5047-5056(2014)，其全部内容在此通过引用并入)；瞬时二聚化可以偶尔允许单一前间隔区序列将二聚体定位于靶位点。当使用脱靶引导RNA载体时或当用pUC替换recCas9载体时，未检测到高于背景的活性(图3B)。

[0353] 这些发现表明，通过选择不同的靶位点和匹配引导RNA序列，recCas9活性可以比初始实验中观察到的适度活性显著增加。与原始靶序列相比，观察到PCDH15位点上的活性增加大于10倍(比较图3B和图2F)。此外，最大重组活性取决于两个引导RNA和recCas9的存在。

[0354] recCas9的正交性

[0355] 接下来，测试recCas9是否可以以正交方式靶向在人基因组中发现的序列匹配的多个分开的基因座。基于其作为基因组整合的安全港基因座的潜在用途，或者在一种情况下基于其在遗传疾病中涉及的基因内的位置，选择人基因组中recCas9靶位点的子集。

[0356] 为了鉴定这些位点，搜索ENSEMBL(版本81)以鉴定哪些预测的recCas9靶位点落入注释的基因内(参见例如，Cunningham et al.,Ensembl 2015.Nucleic acids research 43,D662-669(2015)，其全部内容在此通过引用并入)。一个此类位点落入FGF14的内含子区域。据信FGF14内的突变会引起脊髓小脑性共济失调(spinocerebellar ataxia)27(SCA 27)(参见例如，van Swieten et al.,A mutation in the fibroblast growth factor

14gene is associated with autosomal dominant cerebellar ataxia[corrected].Am J Hum Genet 72,191-199(2003);Brusse et al.,Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14gene(SCA27):A new phenotype.Mov Disord 21,396-401(2006);Choquet et al.,A novel frameshift mutation in FGF14causes an autosomal dominant episodic ataxia.Neurogenetics 16,233-236(2015);Coebergh et al.,A new variable phenotype in spinocerebellar ataxia 27(SCA 27)caused by a deletion in the FGF14gene.Eur J Paediatr Neurol 18,413-415(2014);Shimojima et al.,Spinocerebellar ataxias type 27derived from a disruption of the fibroblast growth factor 14gene with mimicking phenotype of paroxysmal non-kinesigenic dyskinesia.Brain Dev 34,230-233(2012);其每篇的全部内容通过引用并入本文)。最后,手动询问未落入基因内的预测的recCas9靶位点的级分,以确定一些序列是否落入安全港基因座内。使用ENSEMBL中的注释鉴定了基因组靶标,该靶标与Bushman及其同事描述的安全港基因座的五个标准的大多数相匹配(Cunningham et al.,Ensembl 2015.Nucleic acids research 43,D662-669(2015);和Sadelain et al.,Safe harbours for the integration of new DNA in the human genome.Nat Rev Cancer 12,51-58(2012),其每篇的全部内容通过引用并入本文)。构建了五种报告物和含有与基因组中序列相同的序列的相应的引导RNA载体对。为了评估recCas9在用不同的引导RNA编程时的正交性,测试了五种引导RNA对与五种报告物的所有组合。

[0357] 报告物、引导RNA质粒和recCas9表达载体的共转染显示,所测试的五种报告物中的三种导致与recCas9介导的重组一致的显著水平的EGFP阳性细胞。该EGFP表达严格依赖于recCas9表达载体和与报告物构建体上的靶位点序列匹配的引导RNA质粒的共转染(图4A)。当与非关联报告物质粒共转染时,与关联报告物质粒和recCas9载体共转染时引起重组的相同的引导RNA对不能介导重组(图4A)。这些结果证明recCas9活性是正交的,并且当用与侧翼序列匹配的一对引导RNA编程时,仅催化gix相关核心位点处的重组。当未用表达recCas9和引导RNA的载体转染报告物质粒时,没有观察到高于测定的背景水平的重组酶活性。

[0358] recCas9产物的特性

[0359] 表征recCas9介导的报告物质粒重组的产物以证实EGFP表达是recCas9介导的poly-A终止子序列除去的结果。在与recCas9表达载体和与产生关联或非关联引导RNA对的质粒共转染后,对报告物质粒进行染色体5位点1、染色体12和染色体13(FGF14基因座)的测序。温育72小时后,提取附加型DNA(如上所述)并转化到大肠杆菌中以分离报告物质粒。对含有报告物质粒的单菌落进行测序(图4B)。

[0360] 预期个别菌落含有未修饰的或重组的报告物质粒(图4C)。对于每个生物学重复,对用从每个转染条件分离的报告物质粒转化的平均97个菌落进行测序。若报告物质粒先前与关联引导RNA质粒和recCas9表达载体共转染,则仅观察到重组质粒(图4D)。在两个单独的实验中,重组质粒的百分比范围从染色体5中的位点1的12%到染色体13中的FGF14基因座的平均32%。因此,测序数据与图4A中更早的流式细胞术分析一致。重组质粒的绝对水平略高于EGFP阳性细胞的百分比(图4)。这种差异可能是因为流式细胞术测定没有报告当在单一细胞中存在多个拷贝的报告物质粒时可发生的多个重组事件而出现;甚至单一重组事

件可以导致EGFP荧光。结果,EGFP阳性细胞的百分比可以对应于重组报告物质粒的实际百分比的下限。或者,差异可以反映质粒大小和转化效率之间的负相关性(参见例如,Hanahan,Studies on transformation of Escherichia coli with plasmids.J Mol Biol 166,557-580(1983),其全部内容在此通过引用并入);重组质粒约为5,700个碱基对,并且可以转化得略好于完整质粒(其约为6,900个碱基对)。

[0361] 由于已经报告锌指重组酶在重组酶核心位点连接处引起突变(参见例如,e.g.,Gaj et al.,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic acids research 41,3937-3946(2013),其全部内容在此通过引用并入),因此测试了此类诱变是否自recCas9处理发生。在报告物构建体中,recCas9应通过首先切割这两个gix核心位点的中心二核苷酸并然后将两个核心彼此再连接来删除kanR和poly-A终止子(图4C)。因此,重组产物应当是由"左"靶位点的第一半和"右"靶位点的第二半组成的单一重组位点。错误或不完整的反应可以导致其他产物。引人注目的是,所检查的所有134个重组序列都含有预期的重组产物。此外,来自两组分开的转染实验的总共2,317个测序读段显示仅三个含有在其他情况下非重组质粒上潜在的缺失产物的测序读段。

[0362] 在用pUC对照转染并且缺少两个recCas9靶位点以及polyA终止子的染色体12报告物质粒中观察到这些含有缺失的读段之一。该产物归因于在转染、分离或随后操作期间发生的DNA损伤。因为当与报告物和关联引导RNA表达载体共转染时recCas9仅可以定位于序列,所以更相关的度量可以是测量当与关联引导RNA载体和recCas9表达载体共转染报告物质粒时观察到的缺失产物的总数。从与染色体5-位点1报告物和关联引导RNA的共转染测序的总共185个质粒中观察到单一插入/缺失。类似地,在用关联引导RNA和recCas9表达载体转染后,从来自染色体12报告物的204个质粒中观察到一个插入/缺失。值得注意的是,在202个测序读段中,在关联引导RNA和recCas9共转染后,从染色体13报告物中未观察到插入/缺失,尽管导致观察到的最高重组水平。这些观察结果共同表明recCas9主要介导无错重组。

[0363] 总之,这些结果建立了recCas9可以在最小的交叉反应性或副产物形成的情况下靶向人基因组内发现的多个位点。仅在存在关联引导RNA序列和recCas9的情况下,底物经历有效重组,在人细胞中产生干净的重组产物,并且通常不导致核心位点连接处的突变或产物诸如由细胞DNA修复产生的插入/缺失。

[0364] RecCas9介导的基因组缺失

[0365] 最后,研究了recCas9是否能够直接在培养的人细胞的基因组DNA上操作。使用人基因组中潜在的recCas9识别位点的列表(表9),搜索在由recCas9靶向的情况下将产生通过PCR可检测的染色体缺失事件的位点对。设计引导RNA表达载体以将recCas9引导至最接近染色体5-位点1或染色体13(FGF14基因座)的那些recCas9位点,即均显示在瞬时转染测定中重组的位点(图4)。新的靶位点范围为染色体5-位点1的上游约3至23Mbp和下游7至10Mbp,以及染色体13-FGF14位点的上游的12至44Mbp。与这些新的引导RNA对之每种和用于染色体5-位点1或染色体13-FGF14的经验证的引导RNA对共转染recCas9表达载体,但通过基因组PCR未观察到染色体缺失的证据。

[0366] 若recCas9靶位点在基因组上彼此更接近,则认为基因组缺失可能更有效。鉴定了

在FAM19A2的内含子区域内以14.2kb分开的两个recCas9位点;这些位点还含有相同的二核苷酸核心,这应当有助于缺失。FAM19A2是五个密切相关的TAFA家族基因之一,其编码认为在免疫和神经细胞中具有调节作用的小的分泌蛋白(参见例如Parker et al.,Admixture mapping identifies a quantitative trait locus associated with FEV1/FVC in the COPDGene Study.Genet Epidemiol 38,652-659(2014),其全部内容在此通过引用并入)。位于FAM19A2的内含子序列中的小核苷酸多态性在全基因组关联研究中与系统性红斑狼疮(SLE)和慢性阻塞性肺病(COPD)的风险升高有关(参见例如,Parker et al.,Admixture mapping identifies a quantitative trait locus associated with FEV1/FVC in the COPDGene Study.Genet Epidemiol 38,652-659(2014),其全部内容在此通过引用并入);因此,该基因的内含子区域的缺失可以了解这些疾病的原因。将四个引导RNA序列克隆到表达载体中,所述表达载体设计以介导这两个FAM19A2位点之间的recCas9缺失。与recCas9表达载体共转染表达这些引导RNA的载体(图5A)。RecCas9介导的两个位点之间的重组应当导致14.2kb居间区域的缺失。实际上,使用位于两个FAM19A2recCas9靶标侧翼的基因特异性引物,通过巢式PCR检测到该缺失事件。仅在从用recCas9和所有四种引导RNA表达载体共转染的细胞分离的基因组DNA中观察到与recCas9介导的缺失一致的预期PCR产物(图5B)。在没有单独的引导RNA表达载体的上游或下游对的情况下,在没有recCas9表达质粒的情况下转染的细胞的基因组DNA中,或对于未转染的对照细胞的基因组DNA,未检测到缺失PCR产物(图5B)。这些巢式PCR产物的估计检测限为每5,500个染色体拷贝约1个缺失事件。分离对应于预测的基因组缺失的415-bp PCR产物并测序。测序证实PCR产物与自重组酶介导的基因组缺失预期的预测连接相匹配,并且不含任何提示NHEJ的插入或缺失(图5C)。

[0367] 在基因组模板的连续稀释液上使用巢式PCR估计最小基因组缺失效率的下限(参见上文或例如Sykes et al.,Quantitation of targets for PCR by use of limiting dilution.Biotechniques 13,444-449(1992),其全部内容在此通过引用并入,以获得更多细节)。产生recCas9特异性巢式PCR产物的给定量的基因组DNA必须含有至少一个经编辑的染色体。为了建立该recCas9介导的基因组缺失事件的下限,在基因组DNA(从用recCas9和四种FAM19A2引导RNA表达载体转染的细胞中分离)的连续稀释液上进行巢式PCR,以确定导致可检测的缺失产物的基因组模板DNA的最低浓度。这些实验揭示了0.023±0.017%(三个生物学重复的平均值)的缺失效率的下限(图5D),表明recCas9介导的基因组缺失以至少该效率进行。未转染细胞的基因组DNA的巢式PCR不产生产物,估计检测限为<0.0072%重组。

[0368] 其他替代重组酶的使用

[0369] 称为“36C6”的演化为靶向人基因组的Rosa基因座中位点的Cre重组酶与dCas9融合。然后,将该融合物用于以引导RNA依赖性方式重组含有Rosa靶位点的基于质粒的报告物。图7A证明了使用野生型Cre和36C6的接头优化的结果。显示的1x 2x、5x和8x接头是接头中GGS重复的数量。回复分析证明了对与dCas9融合的36C6进行突变可以影响嵌合融合物的相对引导依赖性(图7B)。回复用其未突变的氨基酸标记。例如,已经突变为M的位置306在进行测定之前回复为I。靶向其关联报告物的GinB构建体用作图7A和7B中所示实验数据的对照。中靶引导是chr13-102010574引导(质粒BC165和166)。显示的缩写是GGS-36C6:dCas9-GGS-36C6;2GGS-36C6(使用接头SEQ ID NO:182):sdCas9-GGSGGS-36C6(使用接头SEQ ID NO:182)。

[0370] 用于36C6和所有变体转染的靶序列如下所示:(引导物-斜体;Rosa位点-粗体):

[0371] CCTAGGGAAGTGATCATAGCTGAGTTTCTATCTCATGGTTTATGCTAAACTATATGTTGACATGTTGAGGAGACTTAAGTCCAAAACCTGG(SEQ ID NO:760)

[0372] 在图7A、7B、8、9A和9B中,GinB的中靶引导是chr13-102010574引导(质粒BC165和166)。图7A、7B、8、9A和9B中的所有脱靶引导由chr12-62418577引导(BC163和BC164)构成。

[0373] 鉴定了人基因组中Rosa26位点侧翼的PAM,其可以支持dCas9结合(图8,上图)。然后设计引导RNA和质粒报告物以测试内源性前间隔区是否可以支持dCas9-36C6活性。靶向其关联报告物的GinB构建体用作对照。见图8。混合:Cas9和36C6之间所有5种接头变体的等份混合物。对于hRosa,靶序列(包括引导RNA靶标)如下:(引导-斜体;Rosa位点-粗体)

[0374] CCTGAAATAATGCAAGTGTAGAATAACTTTTTAAAATCTCATGGTTTATGCTAAACTATATGTTGACATAAGAGTGGTGATAAGGCAACAGTAGG(SEQ ID NO:767)

[0375] 用于hRosa的中靶引导质粒与其他gRNA表达质粒相同,只是前间隔区被上文所示的那些替换(图8)。

[0376] dCas9-Cre重组酶融合物的几个测试的Cre截短显示在图9A中。与dCas9融合的Cre重组酶的截短的变体显示出可察觉的重组酶活性以及对Lox质粒报告物系统中引导RNA的存在的严格依赖(图9B)。截短的变体用截短的Cre开始的残基标记。图9A和9B中所示的所有融合蛋白的接头是8xGGS。与dCas9融合的野生型Cre用作阳性对照。用于36C6和所有变体转染的靶序列如下所示:(引导-斜体;Rosa位点-粗体):

[0377] CCTAGGGAAGTGATCATAGCTGAGTTTCTATCTCATGGTTTATGCTAAACTATATGTTGACATGTTGAGGAGACTTAAGTCCAAAACCTGG(SEQ ID NO:768)

[0378] 使用的中靶引导是chr13-102010574引导(质粒BC165和166),并且脱靶引导是chr12-62418577引导(BC163和BC164)。

[0379] 参考文献

[0380] 1.J.A.Doudna,E.Charpentier,Genome editing.The new frontier of genome engineering with CRISPR-Cas9.Science346,1258096(2014).

[0381] 2.M.R.Capecchi,Altering the genome by homologous recombination.Science244,1288-1292(1989).

[0382] 3.K.R.Thomas,K.R.Folger,M.R.Capecchi,High frequency targeting of genes to specific sites in the mammalian genome.Cell44,419-428(1986).

[0383] 4.A.Choulika,A.Perrin,B.Dujon,J.F.Nicolas,Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae.Mol Cell Biol15,1968-1973(1995).

[0384] 5.D.Carroll,Progress and prospects:zinc-finger nucleases as gene therapy agents.Gene Ther15,1463-1468(2008).

[0385] 6.J.C.Miller et al.,A TALE nuclease architecture for efficient genome editing.Nature biotechnology29,143-U149(2011).

[0386] 7.J.K.Joung,J.D.Sander,TALENs:a widely applicable technology for targeted genome editing.Nat Rev Mol Cell Biol14,49-55(2013).

[0387] 8.P.Mali et al.,RNA-guided human genome engineering via

Cas9.Science339,823-826(2013).

[0388] 9.L.Cong et al.,Multiplex genome engineering using CRISPR/Cas systems.Science339,819-823(2013).

[0389] 10.J.P.Guilinger,D.B.Thompson,D.R.Liu,Fusion of catalytically inactive Cas9to FokI nuclease improves the specificity of genome modification.Nature biotechnology,(2014).

[0390] 11.S.Q.Tsai et al.,Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing.Nature biotechnology,(2014).

[0391] 12.H.Fung,D.M.Weinstock,Repair at single targeted DNA double-strand breaks in pluripotent and differentiated human cells.PloS one6,e20514(2011).

[0392] 13.W.D.Heyer,K.T.Ehmsen,J.Liu,Regulation of homologous recombination in eukaryotes.Annu Rev Genet44,113-139(2010).

[0393] 14.D.Branzei,M.Foiani,Regulation of DNA repair throughout the cell cycle.Nat Rev Mol Cell Bio9,297-308(2008).

[0394] 15.V.T.Chu et al.,Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.Nature biotechnology,(2015).

[0395] 16.T.Maruyama et al.,Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.Nature biotechnology,(2015).

[0396] 17.S.Lin,B.T.Staahl,R.K.Alla,J.A.Doudna,Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.eLife3,e04766(2014).

[0397] 18.S.Turan,C.Zehe,J.Kuehle,J.H.Qiao,J.Bode,Recombinase-mediated cassette exchange(RMCE)-A rapidly-expanding toolbox for targeted genomic modifications.Gene515,1-27(2013).

[0398] 19.T.Gaj,S.J.Sirk,C.F.Barbas,Expanding the Scope of Site-Specific Recombinases for Genetic and Metabolic Engineering.Biotechnology and bioengineering111,1-15(2014).

[0399] 20.N.D.F.Grindley,K.L.Whiteson,P.A.Rice,Mechanisms of site-specific recombination.Annu Rev Biochem75,567-605(2006).

[0400] 21.C.R.Sclimenti,B.Thyagarajan,M.P.Calos,Directed evolution of a recombinase for improved genomic integration at a native human sequence.Nucleic acids research29,5044-5051(2001).

[0401] 22.R.Shah,F.Li,E.Voziyanova,Y.Voziyanov,Target-specific variants of Flp recombinase mediate genome engineering reactions in mammalian cells.The FEBS journal282,3323-3333(2015).

[0402] 23.J.Karpinski et al.,Directed evolution of a recombinase that excises the provirus of most HIV-1 primary isolates with high

specificity.Nature biotechnology,(2016).

[0403] 24.F.Buchholz,A.F.Stewart,Alteration of Cre recombinase site specificity by substrate-linked protein evolution.Nature biotechnology19,1047-1052(2001).

[0404] 25.B.Thyagarajan,E.C.Olivares,R.P.Hollis,D.S.Ginsburg,M.P.Calos,Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase.Mol Cell Biol21,3926-3934(2001).

[0405] 26.B.Thyagarajan,M.J.Guimaraes,A.C.Groth,M.P.Calos,Mammalian genomes contain active recombinase recognition sites.Gene244,47-54(2000).

[0406] 27.A.Akopian,J.He,M.R.Boocock,W.M.Stark,Chimeric recombinases with designed DNA sequence recognition.Proceedings of the National Academy of Sciences of the United States of America100,8688-8691(2003).

[0407] 28.R.M.Gordley,C.A.Gersbach,C.F.Barbas,3rd,Synthesis of programmable integrases.Proceedings of the National Academy of Sciences of the United States of America106,5053-5058(2009).

[0408] 29.M.M.Prorocic et al.,Zinc-finger recombinase activities in vitro.Nucleic acids research39,9316-9328(2011).

[0409] 30.C.A.Gersbach,T.Gaj,R.M.Gordley,A.C.Mercer,C.F.Barbas,Targeted plasmid integration into the human genome by an engineered zinc-finger recombinase.Nucleic acids research39,7868-7878(2011).

[0410] 31.A.C.Mercer,T.Gaj,R.P.Fuller,C.F.Barbas,Chimeric TALE recombinases with programmable DNA sequence specificity.Nucleic acids research40,11163-11172(2012).

[0411] 32.T.Matsuda,C.L.Cepko,Controlled expression of transgenes introduced by in vivo electroporation.Proceedings of the National Academy of Sciences of the United States of America104,1027-1032(2007).

[0412] 33.N.E.Sanjana et al.,A transcription activator-like effector toolbox for genome engineering.Nature protocols7,171-192(2012).

[0413] 34.T.Gaj,A.C.Mercer,S.J.Sirk,H.L.Smith,C.F.Barbas,A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells.Nucleic acids research41,3937-3946(2013).

[0414] 35.Y.Fu et al.,High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.Nature biotechnology31,822-826(2013).

[0415] 36.J.Quan,J.Tian,Circular polymerase extension cloning of complex gene libraries and pathways.PloS one4,e6441(2009).

[0416] 37.N.Hillson.(2010),vol.2015,pp.CPEC protocol.

[0417] 38.R.C.Gentleman et al.,Bioconductor:open software development for computational biology and bioinformatics.Genome Biol5,R80(2004).

[0418] 39.K.Motmans,S.Thirion,J.Raus,C.Vandevyver,Isolation and

quantification of episomal expression vectors in human T cells.Biotechniques23,1044-1046(1997).

[0419] 40.B.Hirt,Selective extraction of polyoma DNA from infected mouse cell cultures.J Mol Biol26,365-369(1967).

[0420] 41.A.Klippel,G.Mertens,T.Patschinsky,R.Kahmann,The DNA Invertase Gin of Phage Mu-Formation of a Covalent Complex with DNA Via a Phosphoserine at Amino-Acid Position-9.Embo Journal7,1229-1237(1988).

[0421] 42.G.Mertens et al.,Site-specific recombination in bacteriophage Mu: characterization of binding sites for the DNA invertase Gin.The EMBO journal7,1219-1227(1988).

[0422] 43.R.H.Plasterk,A.Brinkman,P.van de Putte,DNA inversions in the chromosome of Escherichia coli and in bacteriophage Mu:relationship to other site-specific recombination systems.Proceedings of the National Academy of Sciences of the United States of America80,5355-5358(1983).

[0423] 44.H.Xu et al.,Sequence determinants of improved CRISPR sgRNA design.Genome research25,1147-1157(2015).

[0424] 45.R.M.Gordley,J.D.Smith,T.Graslund,C.F.Barbas,3rd,Evolution of programmable zinc finger-recombinases with activity in human cells.J Mol Biol367,802-813(2007).

[0425] 46.C.A.Gersbach,T.Gaj,R.M.Gordley,C.F.Barbas,3rd,Directed evolution of recombinase specificity by split gene reassembly.Nucleic acids research38, 4198-4206(2010).

[0426] 47.T.Gaj,A.C.Mercer,C.A.Gersbach,R.M.Gordley,C.F.Barbas,Structure-guided reprogramming of serine recombinase DNA sequence specificity.Proceedings of the National Academy of Sciences of the United States of America108,498-503(2011).

[0427] 48.T.Gaj et al.,Enhancing the Specificity of Recombinase-Mediated Genome Engineering through Dimer Interface Redesign.J Am Chem Soc136,5047-5056(2014).

[0428] 49.F.Cunningham et al.,Ensembl 2015.Nucleic acids research43,D662-669 (2015).

[0429] 50.J.C.van Swieten et al.,A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia[corrected].Am J Hum Genet72,191-199(2003).

[0430] 51.E.Brusse et al.,Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14 gene(SCA27):A new phenotype.Mov Disord21, 396-401(2006).

[0431] 52.K.Choquet,R.La Piana,B.Brais,A novel frameshift mutation in FGF14causes an autosomal dominant episodic ataxia.Neurogenetics16,233-236

(2015).

[0432] 53.J.A.Coebergh et al.,A new variable phenotype in spinocerebellar ataxia 27(SCA 27)caused by a deletion in the FGF14 gene.Eur J Paediatr Neurol18,413-415(2014).

[0433] 54.K.Shimojima et al.,Spinocerebellar ataxias type 27 derived from a disruption of the fibroblast growth factor 14 gene with mimicking phenotype of paroxysmal non-kinesigenic dyskinesia.Brain Dev34,230-233(2012).

[0434] 55.M.Sadelain,E.P.Papapetrou,F.D.Bushman,Safe harbours for the integration of new DNA in the human genome.Nat Rev Cancer12,51-58(2012).

[0435] 56.D.Hanahan,Studies on transformation of Escherichia coli with plasmids.J Mol Biol166,557-580(1983).

[0436] 57.M.M.Parker et al.,Admixture mapping identifies a quantitative trait locus associated with FEV1/FVC in the COPDGene Study.Genet Epidemiol38,652-659(2014).

[0437] 58.P.J.Sykes et al.,Quantitation of targets for PCR by use of limiting dilution.Biotechniques13,444-449(1992).

[0438] 59.A.Rath,R.Hromas,A.De Benedetti,Fidelity of end joining in mammalian episomes and the impact of Metnase on joint processing.BMC Mol Biol15,6(2014).

[0439] 60.P.Rebuzzini et al.,New mammalian cellular systems to study mutations introduced at the break site by non-homologous end-joining.DNA Repair(Amst)4,546-555(2005).

[0440] 61.J.Smith,C.Baldeyron,I.De Oliveira,M.Sala-Trepat,D.Papadopoulo,The influence of DNA double-strand break structure on end-joining in human cells.Nucleic acids research29,4783-4792(2001).

[0441] 62.S.Turan et al.,Recombinase-mediated cassette exchange(RMCE):traditional concepts and current challenges.J Mol Biol407,193-221(2011).

[0442] 63.S.J.Sirk,T.Gaj,A.Jonsson,A.C.Mercer,C.F.Barbas,Expanding the zinc-finger recombinase repertoire:directed evolution and mutational analysis of serine recombinase specificity determinants.Nucleic acids research42,4755-4766(2014).

[0443] 64.B.P.Kleinstiver et al.,Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition.Nature biotechnology33,1293-1298(2015).

[0444] 65.B.P.Kleinstiver et al.,Engineered CRISPR-Cas9 nucleases with altered PAM specificities.Nature523,481-U249(2015).

[0445] 66.K.M.Esvelt et al.,Orthogonal Cas9 proteins for RNA-guided gene regulation and editing.Nature methods10,1116-1121(2013).

[0446] 67.B.Zetsche et al.,Cpf1 Is a Single RNA-Guided Endonuclease of a

Class 2CRISPR-Cas System.Cell163,759-771(2015).

[0447] 68.K.Dormiani et al.,Long-term and efficient expression of human beta-globin gene in a hematopoietic cell line using a new site-specific integrating non-viral system.Gene Ther22,663-674(2015).

[0448] 69.E.Wijnker,H.de Jong,Managing meiotic recombination in plant breeding.Trends in plant science13,640-646(2008).

[0449] 70.J.F.Petolino,V.Srivastava,H.Daniell,Editing Plant Genomes:a new era of crop improvement.Plant Biotechnol J14,435-436(2016).

[0450] 等同实施方案和范围

[0451] 本领域技术人员将认识到或能够仅仅使用常规的实验确定本文所述的本发明具体实施方案的许多等同实施方案。本发明的范围不意图限于以上说明书，而是如所附权利要求中所述。

[0452] 在权利要求中，诸如“一种”、“一个”和“该”的冠词可以表示一个或超出一个，除非相反地指出或者从上下文中显而易见。若一个、超出一个或所有组成员在给定产物或过程中存在、使用或以其他方式相关，则认为在组中的一个或多个成员之间包括“或”的权利要求或说明书是满足的，除非另有说明或从上下文中显而易见。本发明包括实施方案，其中组的恰好一个成员在给定产物或过程中存在、使用或以其他方式相关。本发明还包括实施方案，其中超出一个或所有组成员在给定产物或过程中存在、使用或以其他方式相关。

[0453] 此外，应理解，本发明涵盖所有变型、组合和置换，其中来自一个或多个权利要求或来自说明书的相关部分的一个或多个限制、元素、条款、描述性术语等被引入另一个权利要求中。例如，可以修改依赖于另一个权利要求的任何权利要求以包括在依赖于相同基本权利要求的任何其他权利要求中找到的一个或多个限制。此外，在权利要求叙述组合物的情况下，应当理解包括将组合物用于本文公开的任何目的的方法，并且包括根据本文公开的任何制备方法或本领域中已知的其他方法制备组合物的方法，除非另有说明或者除非本领域普通技术人员明白会出现矛盾或不一致。

[0454] 在将元素呈现为列表(例如，以马库什群组格式)的情况下，应当理解，还公开了元素的每个子群，并且可以从群组中移除任何元素。还应注意，术语“包含”旨在是开放的并且允许包含另外的元素或步骤。应当理解，通常，在本发明或本发明的方面称为包含特定元素、特征、步骤等的情况下，本发明或本发明的方面的某些实施方案由此类元素、特征、步骤等组成，或基本上由之组成。出于简化的目的，那些实施方案未在本文中用同样的词语具体阐述。因此，对于包含一个或多个元素、特征、步骤等的本发明的每个实施方案，本发明还提供了由这些元素、特征、步骤等组成或基本上由之组成的实施方案。

[0455] 在给出范围的情况下，端点包括在内。此外，应当理解，除非另有说明或从上下文和/或本领域普通技术人员的理解中明显看出，否则表示为范围的值可以假定在本发明的不同实施方案中的所述范围内的任何特定值，至该范围下限的单位的十分之一，除非上下文另有明确规定。还应当理解，除非另有说明或从上下文和/或本领域普通技术人员的理解中明显看出，否则表示为范围的值可以假定给定范围内的任何子范围，其中子范围的端点表示为与范围的下限的单位的十分之一相同的精度。

[0456] 此外，应当理解，本发明的任何具体实施方案可以明确地从任何一个或多个权利

要求中排除。在给出范围的情况下，该范围内的任何值可以明确地从任何一个或多个权利要求中排除。本发明的组合物和/或方法的任何实施方案、元素、特征、应用或方面可以从任何一个或多个权利要求中排除。出于简洁的目的，本文未明确阐述其中排除一个或多个元素、特征、目的或方面的所有实施方案。

[0457] 本文提及的所有出版物、专利和序列数据库条目，包括上面列出的那些项，通过引用整体并入本文，如同每个单独的出版物或专利被具体和单独地指出通过引用并入。在冲突的情况下，以本申请(包括本文中的任何定义)为准。

序列表

<110> 哈佛大学的校长及成员们

<120> 可编程CAS9-重组酶融合蛋白及其用途

<130> H0824.70243W000

<140> Not Yet Assigned

<141> 2017-08-09

<150> US 62/456,048

<151> 2017-02-07

<150> US 62/372,755

<151> 2016-08-09

<160> 775

<170> PatentIn version 3.5

<210> 1

<211> 1368

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 1

```
Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1               5                  10                  15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
            20                  25                  30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
        35                  40                  45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
    50                  55                  60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
```

145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815

Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830

Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1085 1090 1095
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1100 1105 1110
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1115 1120 1125
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1130 1135 1140
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1145 1150 1155
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1160 1165 1170
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1175 1180 1185
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1190 1195 1200
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1205 1210 1215
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1220 1225 1230
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1235 1240 1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1250 1255 1260
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1265 1270 1275
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365
<210> 2

<211> 4104

<212> DNA

<213> 酿脓链球菌

<400> 2

atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60
atcactgatg attataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120
cacagtatca aaaaaaatct tataggggct cttttatttg gcagtggaga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300
cttgaagagt ctttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggcag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480
atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tatttatcca gttggtacaa atctacaatc aattatttga agaaaaccct 600
attaacgcaa gtagagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga gaaatggctt gtttgggaat 720
ctcattgctt tgtcattggg attgacccct aattttaaat caaattttga tttggcagaa 780
gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840
caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900
ttactttcag atatcctaag agtaaatagt gaaataacta aggctcccct atcagcttca 960
atgattaagc gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320
gaaaaaatct tgacttttcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380
cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa 1440
gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt 1560
tataacgaat tgacaaaggt caaatatgtt actgagggaa tgcgaaaacc agcatttctt 1620
tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt 1740
tcaggagttg aagatagatt taatgcttca ttaggcgcct accatgattt gctaaaaatt 1800
attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt 1860
ttaacattga ccttatttga agataggggg atgattgagg aaagacttaa aacatatgct 1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980
cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040
gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100

agtttgacat ttaaagaaga tattcaaaaa gcacaggtgt ctggacaagg ccatagttta 2160
catgaacaga ttgctaactt agctggcagt cctgctatta aaaaaggtat tttacagact 2220
gtaaaaattg ttgatgaact ggtcaaagta atggggcata agccagaaaa tatcgttatt 2280
gaaatggcac gtgaaaatca gacaactcaa aagggccaga aaaattcgcg agagcgtatg 2340
aaacgaatcg aagaaggtat caaagaatta ggaagtcaga ttcttaaaga gcatcctgtt 2400
gaaaatactc aattgcaaaa tgaaaagctc tatctctatt atctacaaaa tggaagagac 2460
atgtatgtgg accaagaatt agatattaat cgtttaagtg attatgatgt cgatcacatt 2520
gttccacaaa gtttcattaa agacgattca atagacaata aggtactaac gcgttctgat 2580
aaaaatcgtg gtaaatcgga taacgttcca agtgaagaag tagtcaaaaa gatgaaaaac 2640
tattggagac aacttctaaa cgccaagtta atcactcaac gtaagtttga taatttaacg 2700
aaagctgaac gtggaggttt gagtgaactt gataaagctg gttttatcaa acgccaattg 2760
gttgaaactc gccaaatcac taagcatgtg gcacaaattt tggatagtcg catgaatact 2820
aaatacgatg aaaatgataa acttattcga gaggttaaag tgattacctt aaaatctaaa 2880
ttagtttctg acttccgaaa agatttccaa ttctataaag tacgtgagat taacaattac 2940
catcatgccc atgatgcgta tctaaatgcc gtcgttggaa ctgctttgat taagaaatat 3000
ccaaaacttg aatcggagtt tgtctatggt gattataaag tttatgatgt tcgtaaaatg 3060
attgctaagt ctgagcaaga aataggcaaa gcaaccgcaa aatatttctt ttactctaat 3120
atcatgaact tcttcaaaac agaaattaca cttgcaaatg gagagattcg caaacgccct 3180
ctaatcgaaa ctaatgggga aactggagaa attgtctggg ataaagggcg agattttgcc 3240
acagtgcgca aagtattgtc catgccccaa gtcaatattg tcaagaaaac agaagtacag 3300
acaggcggat tctccaagga gtcaatttta ccaaaaagaa attcggacaa gcttattgct 3360
cgtaaaaaag actgggatcc aaaaaaatat ggtggttttg atagtccaac ggtagcttat 3420
tcagtcctag tggttgctaa ggtggaaaaa gggaaatcga agaagttaaa atccgttaaa 3480
gagttactag ggatcacaat tatggaaaga agttcctttg aaaaaaatcc gattgacttt 3540
ttagaagcta aaggatataa ggaagttaaa aaagacttaa tcattaaact acctaaatat 3600
agtctttttg agttagaaaa cggtcgtaaa cggatgctgg ctagtgccgg agaattacaa 3660
aaaggaaatg agctggctct gccaagcaaa tatgtgaatt ttttatattt agctagtcat 3720
tatgaaaagt tgaagggtag tccagaagat aacgaacaaa aacaattgtt tgtggagcag 3780
cataagcatt atttagatga gattattgag caaatcagtg aattttctaa gcgtgttatt 3840
ttagcagatg ccaatttaga taaagttctt agtgcatata acaaacatag agacaaacca 3900
atacgtgaac aagcagaaaa tattattcat ttatttacgt tgacgaatct tggagctccc 3960
gctgctttta aatattttga tacaacaatt gatcgtaaac gatatacgtc tacaaaagaa 4020
gttttagatg ccactcttat ccatcaatcc atcactggtc tttatgaaac acgcattgat 4080
ttgagtcagc taggaggtga ctga 4104

<210> 3
<211> 1367
<212> PRT
<213> 酿脓链球菌
<400> 3

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser

305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590
Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705 710 715 720
His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035

Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050

Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080

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

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 4
<211> 4212
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 4

atggataaaa agtattctat tggtttagac atcggcacta attccgttgg atgggctgtc 60
ataaccgatg aatacaaagt accttcaaag aaatttaagg tgttggggaa cacagaccgt 120
cattcgatta aaaagaatct tatcggtgcc ctcctattcg atagtggcga aacggcagag 180
gcgactcgcc tgaaacgaac cgctcggaga aggtatacac gtcgcaagaa ccgaatatgt 240
tacttacaag aaatttttag caatgagatg gccaaagttg acgattcttt ctttcaccgt 300
ttggaagagt ccttccttgt cgaagaggac aagaaacatg aacggcaccc catctttgga 360
aacatagtag atgaggtggc atatcatgaa aagtacccaa cgatttatca cctcagaaaa 420
aagctagttg actcaactga taaagcggac ctgaggttaa tctacttggc tcttgcccat 480
atgataaagt tccgtgggca ctttctcatt gagggtgatc taaatccgga caactcggat 540
gtcgacaaac tgttcatcca gttagtacaa acctataatc agttgtttga agagaaccct 600
ataaatgcaa gtggcgtgga tgcgaaggct attcttagcg cccgcctctc taaatcccga 660
cggctagaaa acctgatcgc acaattaccc ggagagaaga aaaatgggtt gttcggtaac 720
cttatagcgc tctcactagg cctgacacca aattttaagt cgaacttcga cttagctgaa 780
gatgccaaat tgcagcttag taaggacacg tacgatgacg atctcgacaa tctactggca 840

caaattggag atcagtatgc ggacttattt ttggctgcca aaaaccttag cgatgcaatc 900
ctcctatctg acatactgag agttaatact gagattacca aggcgccgtt atccgcttca 960
atgatcaaaa ggtacgatga acatcaccaa gacttgacac ttctcaaggc cctagtccgt 1020
cagcaactgc ctgagaaata taaggaaata ttctttgatc agtcgaaaaa cgggtacgca 1080
ggttatattg acggcggagc gagtcaagag gaattctaca agtttatcaa acccatatta 1140
gagaagatgg atgggacgga agagttgctt gtaaaactca atcgcgaaga tctactgcga 1200
aagcagcgga ctttcgacaa cggtagcatt ccacatcaaa tccacttagg cgaattgcat 1260
gctatactta gaaggcagga ggatttttat ccgttcctca aagacaatcg tgaaaagatt 1320
gagaaaatcc taacctttcg cataccttac tatgtgggac ccctggcccg agggaactct 1380
cggttcgcat ggatgacaag aaagtccgaa gaaacgatta ctccatggaa ttttgaggaa 1440
gttgtcgata aaggtgcgtc agctcaatcg ttcatcgaga ggatgaccaa ctttgacaag 1500
aatttaccga acgaaaaagt attgcctaag cacagtttac tttacgagta tttcacagtg 1560
tacaatgaac tcacgaaagt taagtatgtc actgagggca tgcgtaaacc cgcctttcta 1620
agcggagaac agaagaaagc aatagtagat ctgttattca agaccaaccg caaagtgaca 1680
gttaagcaat tgaaagagga ctactttaag aaaattgaat gcttcgattc tgtcgagatc 1740
tccggggtag aagatcgatt taatgcgtca cttggtacgt atcatgacct cctaaagata 1800
attaaagata aggacttcct ggataacgaa gagaatgaag atatcttaga agatatagtg 1860
ttgactctta ccctctttga agatcgggaa atgattgagg aaagactaaa aacatacgct 1920
cacctgttcg acgataaggt tatgaaacag ttaaagaggc gtcgctatac gggctgggga 1980
cgattgtcgc ggaaacttat caacgggata agagacaagc aaagtggtaa aactattctc 2040
gattttctaa agagcgacgg cttcgccaat aggaacttta tgcagctgat ccatgatgac 2100
tctttaacct tcaaagagga tatacaaaag gcacaggttt ccggacaagg ggactcattg 2160
cacgaacata ttgcgaatct tgctggttcg ccagccatca aaaagggcat actccagaca 2220
gtcaaagtag tggatgagct agttaaggtc atgggacgtc acaaaccgga aaacattgta 2280
atcgagatgg cacgcgaaaa tcaaacgact cagaaggggc aaaaaaacag tcgagagcgg 2340
atgaagagaa tagaagaggg tattaaagaa ctgggcagcc agatcttaaa ggagcatcct 2400
gtggaaaata cccaattgca gaacgagaaa ctttacctct attacctaca aaatggaagg 2460
gacatgtatg ttgatcagga actggacata aaccgtttat ctgattacga cgtcgatcac 2520
attgtacccc aatccttttt gaaggacgat tcaatcgaca ataaagtgct tacacgctcg 2580
gataagaacc gagggaaaag tgacaatgtt ccaagcgagg aagtcgtaaa gaaaatgaag 2640
aactattggc ggcagctcct aaatgcgaaa ctgataacgc aaagaaagtt cgataactta 2700
actaaagctg agaggggtgg cttgtctgaa cttgacaagg ccggatttat taaacgtcag 2760
ctcgtggaaa cccgccaaat cacaaagcat gttgcacaga tactagattc ccgaatgaat 2820
acgaaatacg acgagaacga taagctgatt cgggaagtca aagtaatcac tttaaagtca 2880
aaattggtgt cggacttcag aaaggatttt caattctata aagttaggga gataaataac 2940
taccaccatg cgcacgacgc ttatcttaat gccgtcgtag ggaccgcact cattaagaaa 3000
tacccgaagc tagaaagtga gtttgtgtat ggtgattaca aagtttatga cgtccgtaag 3060
atgatcgcga aaagcgaaca ggagataggc aaggctacag ccaaatactt cttttattct 3120
aacattatga atttctttaa gacggaaatc actctggcaa acggagagat acgcaaacga 3180

cctttaattg aaaccaatgg ggagacaggt gaaatcgtat gggataaggg ccgggacttc 3240
gcgacggtga gaaaagtttt gtccatgccc caagtcaaca tagtaaagaa aactgaggtg 3300
cagaccggag ggttttcaaa ggaatcgatt cttccaaaaa ggaatagtga taagctcatc 3360
gctcgtaaaa aggactggga cccgaaaaag tacggtggct tcgatagccc tacagttgcc 3420
tattctgtcc tagtagtggc aaaagttgag aagggaaaat ccaagaaact gaagtcagtc 3480
aaagaattat tggggataac gattatggag cgctcgtctt ttgaaaagaa ccccatcgac 3540
ttccttgagg cgaaaggtta caaggaagta aaaaaggatc tcataattaa actaccaaag 3600
tatagtctgt ttgagttaga aaatggccga aaacggatgt tggctagcgc cggagagctt 3660
caaaagggga acgaactcgc actaccgtct aaatacgtga atttcctgta tttagcgtcc 3720
cattacgaga agttgaaagg ttcacctgaa gataacgaac agaagcaact ttttgttgag 3780
cagcacaaac attatctcga cgaaatcata gagcaaattt cggaattcag taagagagtc 3840
atcctagctg atgccaatct ggacaaagta ttaagcgcat acaacaagca cagggataaa 3900
cccatacgtg agcaggcgga aaatattatc catttgttta ctcttaccaa cctcggcgct 3960
ccagccgcat tcaagtattt tgacacaacg atagatcgca aacgatacac ttctaccaag 4020
gaggtgctag acgcgacact gattcaccaa tccatcacgg gattatatga aactcggata 4080
gatttgtcac agcttggggg tgacggatcc cccaagaaga agaggaaagt ctcgagcgac 4140
tacaaagacc atgacggtga ttataaagat catgacatcg attacaagga tgacgatgac 4200
aaggctgcag ga 4212

<210> 5
<211> 1368
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 5

Met Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305 310 315 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
325 330 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
340 345 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
355 360 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
370 375 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385 390 395 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
405 410 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe

420 425 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020
Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

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

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser

1235 1240 1245

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 6
<211> 18
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 6

Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1 5 10 15
Gly Ser

<210> 7
<211> 16
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 7

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

<210> 8
<211> 12
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 8

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

<210> 9
<211> 21
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 9

Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser Ala Thr Pro Glu Gly
1 5 10 15
Gly Ser Gly Gly Ser
20

<210> 10
<211> 15
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 10
Val Pro Phe Leu Leu Glu Pro Asp Asn Ile Asn Gly Lys Thr Cys
1 5 10 15

<210> 11
<211> 12
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 11
Gly Ser Ala Gly Ser Ala Ala Gly Ser Gly Glu Phe
1 5 10

<210> 12
<211> 12
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 12
Ser Ile Val Ala Gln Leu Ser Arg Pro Asp Pro Ala
1 5 10

<210> 13
<211> 10
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 13
Met Lys Ile Ile Glu Gln Leu Pro Ser Ala

1 5 10

<210> 14

<211> 10

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 14

Val Arg His Lys Leu Lys Arg Val Gly Ser

1 5 10

<210> 15

<211> 12

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 15

Gly His Gly Thr Gly Ser Thr Gly Ser Gly Ser Ser

1 5 10

<210> 16

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 16

Met Ser Arg Pro Asp Pro Ala

1 5

<210> 17

<211> 4

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 17

Gly Gly Ser Met

1

<210> 18

<400> 18

000

<210> 19

<211> 17

<212> RNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<220>

<221> misc_feature

<222> (1)..(4)

<223> n是a, c, g或u

<220>

<221> misc_feature

<222> (8)..(9)

<223> s是g或c

<220>

<221> misc_feature

<222> (10)..(11)

<223> w是a, t或u

<220>

<221> misc_feature

<222> (12)..(13)

<223> s是g或c

<220>

<221> misc_feature

<222> (14)..(17)

<223> n是a, c, g或u

<400> 19

nnnnaaassw wssnnnn 17

<210> 20

<211> 22

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 20

ggtgtttcgt cctttccaca ag 22

<210> 21

<211> 41

<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 21
gcacactagt tagggataac agttttagag ctagaaatag c 41
<210> 22
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 22
gcccatgacc cttctcctct gttttagagc tagaaatagc 40
<210> 23
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 23
gctcagggcc tgtgatggga ggttttagag ctagaaatag c 41
<210> 24
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 24
ggcccatgac ccttctcctc gttttagagc tagaaatagc 40
<210> 25
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 25
gcctcagggc ctgtgatggg agttttagag ctagaaatag c 41
<210> 26

<211> 40
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 26
gacttgaaac actcttttc gttttagagc tagaaatagc 40
<210> 27
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 27
gagttgaaga cacacaacac agttttagag ctagaaatag c 41
<210> 28
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 28
ggaactcatg tgattaactg gttttagagc tagaaatagc 40
<210> 29
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 29
gtctacctct catgagccgg tgttttagag ctagaaatag c 41
<210> 30
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 30
gtttcccgca ggatgtggga tgttttagag ctagaaatag c 41

<210> 31
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 31
gcctggggat ttatgttctt agttttagag ctagaaatag c 41
<210> 32
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 32
gaaatagcac aatgaatgga agttttagag ctagaaatag c 41
<210> 33
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 33
gacttttgg gggagaggga ggttttagag ctagaaatag c 41
<210> 34
<211> 40
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 34
ggagacttaa gtccaaaacc gttttagagc tagaaatagc 40
<210> 35
<211> 41
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 35

gtcagctatg atcacttccc tgttttagag ctagaaatag c 41

<210> 36

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 36

tcgtctcggc gtccccaatt ttcccaaaca gaggtctgta aaccgaggtg agacgg 56

<210> 37

<211> 56

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 37

ccgtctcacc tcggtttaca gacctctgtt tgggaaaatt ggggacgccg agacga 56

<210> 38

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 38

tcgtctcggc gtccccaatt ttcccaaaca gaggttctgt aaaccgaggt gagacgg 57

<210> 39

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 39

ccgtctcacc tcggtttaca gaacctctgt ttgggaaaat tggggacgcc gagacga 57

<210> 40

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 40
tcgtctcggc gtccccaatt ttcccaaaca gaggtatctg taaaccgagg tgagacgg 58
<210> 41
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 41
ccgtctcacc tcggtttaca gatacctctg tttgggaaaa ttggggacgc cgagacga 58
<210> 42
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 42
tcgtctcggc gtccccaatt ttcccaaaca gaggtaatct gtaaaccgag gtgagacgg 59
<210> 43
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 43
ccgtctcacc tcggtttaca gattacctct gtttgggaaa attggggacg ccgagacga 59
<210> 44
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 44
tcgtctcggc gtccccaatt ttcccaaaca gaggtaaatc tgtaaaccga ggtgagacgg 60
<210> 45
<211> 60
<212> DNA
<213> 人工序列
<220>

<223> 合成多核苷酸
<400> 45
ccgtctcacc tcggtttaca gatttacctc tgtttgggaa aattggggac gccgagacga 60
<210> 46
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 46
tcgtctcggc gtccccaatt ttcccaaaca gaggtgaaat ctgtaaaccg aggtgagacg 60
g 61
<210> 47
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 47
ccgtctcacc tcggtttaca gatttcacct ctgtttggga aaattgggga cgccgagacg 60
a 61
<210> 48
<211> 62
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 48
tcgtctcggc gtccccaatt ttcccaaaca gaggtcgaaa tctgtaaacc gaggtgagac 60
gg 62
<210> 49
<211> 62
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 49
ccgtctcacc tcggtttaca gatttcgacc tctgtttggg aaaattgggg acgccgagac 60
ga 62

<210> 50
<211> 63
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 50
tcgtctcggc gtccccaatt ttcccaaaca gaggttcgaa atctgtaaac cgaggtgaga 60
cgg 63
<210> 51
<211> 63
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 51
ccgtctcacc tcggtttaca gatttcgaac ctctgtttgg gaaaattggg gacgccgaga 60
cga 63
<210> 52
<211> 54
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 52
tcgtctcgga ggttttggaa cctctgtttg ggaaaattgg ggagtctgag acgg 54
<210> 53
<211> 54
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 53
ccgtctcaga ctccccaatt ttcccaaaca gaggttccaa aacctccgag acga 54
<210> 54
<211> 55
<212> DNA
<213> 人工序列
<220>

<223> 合成多核苷酸
<400> 54
tcgtctcgga ggttttggac acctctgttt gggaaaattg gggagtctga gacgg 55
<210> 55
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 55
ccgtctcaga ctccccaatt ttcccaaaca gaggtgtcca aaacctccga gacga 55
<210> 56
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 56
tcgtctcgga ggttttggac tacctctgtt tgggaaaatt ggggagtctg agacgg 56
<210> 57
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 57
ccgtctcaga ctccccaatt ttcccaaaca gaggtagtcc aaaacctccg agacga 56
<210> 58
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 58
tcgtctcgga ggttttggac ttacctctgt ttgggaaaat tggggagtct gagacgg 57
<210> 59
<211> 57
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 59
ccgtctcaga ctccccaatt ttcccaaaca gaggtaagtc caaaacctcc gagacga 57
<210> 60
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 60
tcgtctcgga ggttttggac ttaacctctg tttgggaaaa ttggggagtc tgagacgg 58
<210> 61
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 61
ccgtctcaga ctccccaatt ttcccaaaca gaggttaagt ccaaaacctc cgagacga 58
<210> 62
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 62
tcgtctcgga ggttttggac ttagacctct gtttgggaaa attggggagt ctgagacgg 59
<210> 63
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 63
ccgtctcaga ctccccaatt ttcccaaaca gaggtctaag tccaaaacct ccgagacga 59
<210> 64
<211> 60
<212> DNA

<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 64
tcgtctcgga ggttttggac ttagcacctc tgtttgggaa aattggggag tctgagacgg 60
<210> 65
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 65
ccgtctcaga ctccccaatt ttcccaaaca gaggtgctaa gtccaaaacc tccgagacga 60
<210> 66
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 66
tcgtctcgga ggttttggac ttagctacct ctgtttggga aaattgggga gtctgagacg 60
g 61
<210> 67
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 67
ccgtctcaga ctccccaatt ttcccaaaca gaggtagcta agtccaaaac ctccgagacg 60
a 61
<210> 68
<211> 54
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 68
tcgtctctgc acccccaatt ttcccaaaca gaggtctgta aaccgatgag acgg 54

<210> 69
<211> 54
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 69
ccgtctcatc ggtttacaga cctctgtttg ggaaaattgg gggtgcagag acga 54
<210> 70
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 70
tcgtctctgc acccccaatt ttcccaaaca gaggttctgt aaaccgatga gacgg 55
<210> 71
<211> 55
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 71
ccgtctcatc ggtttacaga acctctgttt gggaaaattg ggggtgcaga gacga 55
<210> 72
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 72
tcgtctctgc acccccaatt ttcccaaaca gaggtatctg taaaccgatg agacgg 56
<210> 73
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 73

ccgtctcatc ggtttacaga tacctctgtt tgggaaaatt gggggtgcag agacga 56

<210> 74

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 74

tcgtctctgc acccccaatt ttcccaaaca gaggtaatct gtaaaccgat gagacgg 57

<210> 75

<211> 57

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 75

ccgtctcatc ggtttacaga ttacctctgt ttgggaaaat tgggggtgca gagacga 57

<210> 76

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 76

tcgtctctgc acccccaatt ttcccaaaca gaggtaaatc tgtaaaccga tgagacgg 58

<210> 77

<211> 58

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 77

ccgtctcatc ggtttacaga tttacctctg tttgggaaaa ttgggggtgc agagacga 58

<210> 78

<211> 59

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 78
tcgtctctgc acccccaatt ttcccaaaca gaggtgaaat ctgtaaaccg atgagacgg 59
<210> 79
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 79
ccgtctcatc ggtttacaga tttcacctct gtttgggaaa attgggggtg cagagacga 59
<210> 80
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 80
tcgtctctgc acccccaatt ttcccaaaca gaggtcgaaa tctgtaaacc gatgagacgg 60
<210> 81
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 81
ccgtctcatc ggtttacaga tttcgacctc tgtttgggaa aattgggggt gcagagacga 60
<210> 82
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 82
tcgtctctgc acccccaatt ttcccaaaca gaggttcgaa atctgtaaac cgatgagacg 60
g 61
<210> 83
<211> 61
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 83
ccgtctcatc ggtttacaga tttcgaacct ctgtttggga aaattggggg tgcagagacg 60
a 61
<210> 84
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 84
tcgtctcgcc gaggttttgg aacctctgtt tgggaaaatt ggggctcgtg agacgg 56
<210> 85
<211> 56
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 85
ccgtctcacg agccccaatt ttcccaaaca gaggttccaa aacctcggcg agacga 56
<210> 86
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 86
tcgtctcgcc gaggttttgg acacctctgt ttgggaaaat tggggctcgt gagacgg 57
<210> 87
<211> 57
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 87
ccgtctcacg agccccaatt ttcccaaaca gaggtgtcca aaacctcggc gagacga 57
<210> 88
<211> 58

<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 88
tcgtctcgcc gaggttttgg actacctctg tttgggaaaa ttggggctcg tgagacgg 58
<210> 89
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 89
ccgtctcacg agccccaatt ttcccaaaca gaggtagtcc aaaacctcgg cgagacga 58
<210> 90
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 90
tcgtctcgcc gaggttttgg acttacctct gtttgggaaa attggggctc gtgagacgg 59
<210> 91
<211> 59
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 91
ccgtctcacg agccccaatt ttcccaaaca gaggtaagtc caaaacctcg gcgagacga 59
<210> 92
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 92
tcgtctcgcc gaggttttgg acttaacctc tgtttgggaa aattggggct cgtgagacgg 60
<210> 93

<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 93
ccgtctcacg agccccaatt ttcccaaaca gaggttaagt ccaaaacctc ggcgagacga 60
<210> 94
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 94
tcgtctcgcc gaggttttgg acttagacct ctgtttggga aaattggggc tcgtgagacg 60
g 61
<210> 95
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 95
ccgtctcacg agccccaatt ttcccaaaca gaggtctaag tccaaaacct cggcgagacg 60
a 61
<210> 96
<211> 62
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 96
tcgtctcgcc gaggttttgg acttagcacc tctgtttggg aaaattgggg ctcgtgagac 60
gg 62
<210> 97
<211> 62
<212> DNA
<213> 人工序列
<220>

<223> 合成多核苷酸
<400> 97
ccgtctcacg agccccaatt ttcccaaaca gaggtgctaa gtccaaaacc tcggcgagac 60
ga 62
<210> 98
<211> 63
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 98
tcgtctcgcc gaggttttgg acttagctac ctctgtttgg gaaaattggg gctcgtgaga 60
cgg 63
<210> 99
<211> 63
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 99
ccgtctcacg agccccaatt ttcccaaaca gaggtagcta agtccaaaac ctcggcgaga 60
cga 63
<210> 100
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 100
tcgtctcggc gtcccctccc atcacaggcc ctgaggttta agagaaaacc tgagacgg 58
<210> 101
<211> 58
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 101
ccgtctcagg ttttctctta aacctcaggg cctgtgatgg gaggggacgc cgagacga 58
<210> 102

<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 102
tcgtctcgaa ccatggtttt gtgggccagg cccatgaccc ttctcctctg ggagtctgag 60
acgg 64
<210> 103
<211> 64
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 103
ccgtctcaga ctcccagagg agaagggtca tgggcctggc ccacaaaacc atggttcgag 60
acga 64
<210> 104
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 104
tcgtctctgc acccccctccc atcacaggcc ctgaggttta agagaaaacc attgagacgg 60
<210> 105
<211> 60
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 105
ccgtctcaat ggttttctct taaacctcag ggcctgtgat gggaggggt gcagagacga 60
<210> 106
<211> 62
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸

<400> 106
tcgtctcgcc atggttttgt gggccaggcc catgaccctt ctcctctggg ctcgtgagac 60
gg 62
<210> 107
<211> 62
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 107
ccgtctcacg agcccagagg agaagggtca tgggcctggc ccacaaaacc atggcgagac 60
ga 62
<210> 108
<211> 43
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 108
atccgtctcc agtcgagtcg gatttgatct gatcaagaga cag 43
<210> 109
<211> 45
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 109
aaccgtctcg gtgcgttcgg atttgatcca gacatgataa gatac 45
<210> 110
<211> 30
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 110
hscgcgttga gacgctgcca tccgtctcgc 30
<210> 111
<211> 30
<212> DNA

<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 111
hstcgagcga gacggatggc agcgtctcaa 30
<210> 112
<211> 130
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 112
gttgttcgtc tcggcgtcct tgtgttgtgt gtcttcaact cacagagtta aacgatgctt 60
tacacagagt agacttgaaa cactcttttt ctggagtctg agacggttct gttttggtgt 120
gattagttat 130
<210> 113
<211> 130
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 113
gttggtcgtc tctgcaccct tgtgttgtgt gtcttcaact cacagagtta aacgatgctt 60
tacacagagt agacttgaaa cactcttttt ctggctcgtg agacggttct gttttggtgt 120
gattagttat 130
<210> 114
<211> 131
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 114
gttgttcgtc tcggcgtccc accggctcat gagaggtaga gctaaggtcc aaacctaggt 60
ttatctgaga ccggaactca tgtgattaac tgtggagtct gagacggttc tgttttggtg 120
tgattagtta t 131
<210> 115
<211> 131
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 115
gttggtcgtc tctgcacccc accggctcat gagaggtaga gctaaggtcc aaacctaggt 60
ttatctgaga ccggaactca tgtgattaac tgtggctcgt gagacggttc tgttttggtg 120
tgattagtta t 131
<210> 116
<211> 130
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 116
gttgttcgtc tcggcgtcct taagaacata aatccccagg aattcacaga aaccttggtt 60
tgagctttgg atttcccgca ggatgtggga taggagtctg agacggttct gttttggtgt 120
gattagttat 130
<210> 117
<211> 130
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 117
gttggtcgtc tctgcaccct taagaacata aatccccagg aattcacaga aaccttggtt 60
tgagctttgg atttcccgca ggatgtggga taggctcgtg agacggttct gttttggtgt 120
gattagttat 130
<210> 118
<211> 129
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 118
gttgttcgtc tcggcgtcca ctccctctcc cccaaaaagt aaaggtagaa aaccaaggtt 60
tacaggcaac aaatagcaca atgaatggaa tggagtctga gacggttctg ttttggtgtg 120
attagttat 129
<210> 119
<211> 129
<212> DNA

<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 119
gttggtcgtc tctgcaccca ctccctctcc cccaaaaagt aaaggtagaa aaccaaggtt 60
tacaggcaac aaatagcaca atgaatggaa tggctcgtga gacggttctg ttttggtgtg 120
attagttat 129
<210> 120
<211> 130
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 120
gttgttcgtc tcggcgtcct agggaagtga tcatagctga gtttctggaa aaacctaggt 60
tttaaagttg aggagactta agtccaaaac ctggagtctg agacggttct gttttggtgt 120
gattagttat 130
<210> 121
<211> 130
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 121
gttggtcgtc tctgcaccct agggaagtga tcatagctga gtttctggaa aaacctaggt 60
tttaaagttg aggagactta agtccaaaac ctggctcgtg agacggttct gttttggtgt 120
gattagttat 130
<210> 122
<211> 47
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 122
ttcatcggat ccgataaaaa gtattctatt ggtttagcta tcggcac 47
<210> 123
<211> 34
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 123
ttcatcggat ccggtggttc aggtggcagc ggag 34
<210> 124
<211> 61
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 124
ttcatcggat ccggagggtc cggaggtagt ggcggcagcg gtggttcagg tggcagcgga 60
g 61
<210> 125
<211> 100
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 125
aataaccggt tcagaccttc cttttcttct ttggggaacc tcccttgtcg tcatcatcct 60
tataatcgga gccaccgtca cccccaagct gtgacaaatc 100
<210> 126
<211> 38
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 126
tgataaggat ccacccttttg gtggtcttcc aaaccgcc 38
<210> 127
<211> 38
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 127
tgataaggat ccaccgctac cacccttttgg tggtcttc 38
<210> 128

<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 128
agatccgcgg ccgctaatac 20
<210> 129
<211> 54
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 129
ttgagtcgtc tctatactct tccttttca atattattga agcatttatc aggg 54
<210> 130
<211> 50
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 130
ctggaacgtc tcactgtcag accaagttta ctcatatata ctttagattg 50
<210> 131
<211> 46
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 131
ggtgtgcgtc tctacagtta tttgccgact accttggtga tctcgc 46
<210> 132
<211> 36
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 132
acaccacgtc tctgtatgag ggaagcggtg atcgcc 36

<210> 133
<211> 42
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 133
catactcttc ctttttcaat attattgaag catttatcag gg 42
<210> 134
<211> 37
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 134
ctgtcagacc aagtttactc atatatactt tagattg 37
<210> 135
<211> 62
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 135
caatctaaag tatatatgag taaacttggt ctgacagttt gccgactacc ttggtgatct 60
cg 62
<210> 136
<211> 65
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 136
caatctaaag tatatatgag taaacttggt ctgacagtta tttgccgact accttggtga 60
tctcg 65
<210> 137
<211> 42
<212> DNA
<213> 人工序列
<220>

<223> 合成多核苷酸
<400> 137
ccctgataaa tgcttcaata atattgaaaa aggaagagta tg 42
<210> 138
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 138
cgcaaatggg cggtaggcgt g 21
<210> 139
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 139
ccgtgatgga ttggtgaatc 20
<210> 140
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 140
cccatacgat ttcacctgtc 20
<210> 141
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 141
gggtattttc cacaggatgc 20
<210> 142
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 142
cttagaaagg cgggtttacg 20
<210> 143
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 143
cttactaagc tgcaatttgg 20
<210> 144
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 144
tgtattcatc ggttatgaca g 21
<210> 145
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 145
cagggtcaag gaaggcacg 19
<210> 146
<211> 17
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 146
gttccgcgca catttcc 17
<210> 147
<211> 19
<212> DNA

<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 147
gcggagccta tggaaaaac 19
<210> 148
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 148
gccttcttct ttttcctaca gc 22
<210> 149
<211> 16
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 149
cgcatcgagc gagcac 16
<210> 150
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 150
tcaagtagca aaagaagtag gagtcag 27
<210> 151
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 151
ttagatgcat tcgtgcttga ag 22
<210> 152
<211> 29

<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 152
ttaatttctg ctgctagaac taaatctgg 29
<210> 153
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 153
gggaagaaaa ctggatggag aatg 24
<210> 154
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 154
cataaatgac ctagtggagc tg 22
<210> 155
<211> 26
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 155
tggttatttt gcccattagt tgatgc 26
<210> 156
<211> 1318
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 156
acgcgtcccc aattttccca aacagaggtc tgtaaaccga ggttttggaa cctctgtttg 60
ggaaaattgg ggagtcgagt cggatttgat ctgatcaaga gacaggatga ggatcgtttc 120

gcatgattga acaagatgga ttgcacgcag gttctccggc cgcttgggtg gagaggctat 180
tcggctatga ctgggcacaa cagacaatcg gctgctctga tgccgccgtg ttccggctgt 240
cagtcgcagg ggcgcccggt tctttttgtc aagaccgacc tgtccggtgc cctgaatgaa 300
ctgcaggacg aggcagcgcg gctatcgtgg ctggccacga cgggcgttcc ttgcgcagct 360
gtgctcgacg ttgtcactga agcgggaagg gactggctgc tattgggcga agtgccgggg 420
caggatctcc tgtcatctca ccttgctcct gccgagaaag tatccatcat ggctgatgca 480
atgcggcggc tgcatacgct tgatccggct acctgcccat tcgaccacca agcgaaacat 540
cgcatcgagc gagcacgtac tcggatggaa gccggtcttg tcgatcagga tgatctggac 600
gaagagcatc aggggctcgc gccagccgaa ctgttcgcca ggctcaaggc gcgcatgccc 660
gacggcgagg atctcgtcgt gacccatggc gatgcctgct tgccgaatat catggtggaa 720
aatggccgct tttctggatt catcgactgt ggccggctgg gtgtggcgga ccgctatcag 780
gacatagcgt tggctacccg tgatattgct gaagagcttg gcggcgaatg ggctgaccgc 840
ttcctcgtgc tttacggtat cgccgctccc gattcgcagc gcatcgcctt ctatcgcctt 900
cttgacgagt tcttctgagc gggactctgg ggttcgaaat gaccgaccaa gcgacgccca 960
acctgccatc acgagatttc gattccaccg ccgccttcta tgaaaggttg ggcttcggaa 1020
tcgttttccg ggacgccggc tggatgatcc tccagcgcgg ggatctcatg ctggagttct 1080
tcgcccaccc catcgataac ttgtttattg cagcttataa tggttacaaa taaagcaata 1140
gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt ggtttgtcca 1200
aactcatcaa tgtatcttat catgtctgga tcaaatccga acgcaccccc aattttccca 1260
aacagaggtc tgtaaaccga ggttttggaa cctctgtttg ggaaaattgg ggctcgag 1318

<210> 157
<211> 67
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 157
ccccaatttt cccaaacaga ggttctgtaa accgaggttt tggaacctct gtttgggaaa 60
attgggg 67

<210> 158
<211> 68
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 158
ccccaatttt cccaaacaga ggttctgtaa accgaggttt tggcaacctc tgtttgggaa 60
aattgggg 68

<210> 159

<211> 70
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 159
ccccaatttt cccaaacaga ggtatctgta aaccgaggtt ttggctaacc tctgtttggg 60
aaaattgggg 70
<210> 160
<211> 72
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 160
ccccaatttt cccaaacaga ggtaatctgt aaaccgaggt tttggcttaa cctctgtttg 60
ggaaaattgg gg 72
<210> 161
<211> 74
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 161
ccccaatttt cccaaacaga ggtaaatctg taaaccgagg ttttggctta aacctctgtt 60
tgggaaaatt gggg 74
<210> 162
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 162
ccccaatttt cccaaacaga ggtgaaatct gtaaaccgag gttttggctt agaacctctg 60
tttgggaaaa ttgggg 76
<210> 163
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 163
ccccaatttt cccaaacaga ggtcgaaatc tgtaaaccga ggttttggct tagcaacctc 60
tgtttgggaa aattgggg 78
<210> 164
<211> 80
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 164
ccccaatttt cccaaacaga ggttcgaaat ctgtaaaccg aggttttggc ttagctaacc 60
tctgtttggg aaaattgggg 80
<210> 165
<211> 73
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 165
ccccaatttt cccaaacaga ggttcgaaat ctgtaaaccg aggttttgga acctctgttt 60
gggaaaattg ggg 73
<210> 166
<211> 74
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 166
ccccaatttt cccaaacaga ggttcgaaat ctgtaaaccg aggttttggc aacctctgtt 60
tgggaaaatt gggg 74
<210> 167
<211> 74
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 167

ccccaatttt cccaaacaga ggtcgaaatc tgtaaaccga ggttttggct aacctctgtt 60
tgggaaaatt gggg 74
<210> 168
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 168
ccccaatttt cccaaacaga ggtcgaaatc tgtaaaccga ggttttggct taaacctctg 60
tttgggaaaa ttgggg 76
<210> 169
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 169
ccccaatttt cccaaacaga ggtcgaaatc tgtaaaccga ggttttggct tagaacctct 60
gtttgggaaa attgggg 77
<210> 170
<211> 72
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 170
ccccaatttt cccaaacaga ggtctgtaaa ccgaggtttt ggcttagcaa cctctgtttg 60
ggaaaattgg gg 72
<210> 171
<211> 73
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 171
ccccaatttt cccaaacaga ggttctgtaa accgaggttt tggcttagca acctctgttt 60
gggaaaattg ggg 73
<210> 172

<211> 74
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 172
ccccaatttt cccaaacaga ggtatctgta aaccgaggtt ttggcttagc aacctctgtt 60
tgggaaaatt gggg 74
<210> 173
<211> 75
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 173
ccccaatttt cccaaacaga ggtaatctgt aaaccgaggt tttggcttag caacctctgt 60
ttgggaaaat tgggg 75
<210> 174
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 174
ccccaatttt cccaaacaga ggtaaatctg taaaccgagg ttttggctta gcaacctctg 60
tttgggaaaa ttgggg 76
<210> 175
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 175
ccccaatttt cccaaacaga ggtgaaatct gtaaaccgag gttttggctt agcaacctct 60
gtttgggaaa attgggg 77
<210> 176
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 176
cccctcccat cacaggccct gaggtttaag agaaaaccat ggttttgtgg gccaggccca 60
tgacccttct cctctggg 78
<210> 177
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 177
ccttgtgttg tgtgtcttca actcacagag ttaaacgatg ctttacacag agtagacttg 60
aaacactctt tttctgg 77
<210> 178
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 178
ccaccggctc atgagaggta gagctaaggt ccaaacctag gtttatctga gaccggaact 60
catgtgatta actgtgg 77
<210> 179
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 179
ccttaagaac ataaatcccc aggaattcac agaaaccttg gtttgagctt tggatttccc 60
gcaggatgtg ggatagg 77
<210> 180
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 180

ccactccctc tcccccaaaa agtaaaggta gaaaaccaag gtttacaggc aacaaatagc 60
acaatgaatg gaatgg 76
<210> 181
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 181
cctagggaag tgatcatagc tgagtttctg gaaaaaccta ggttttaaag ttgaggagac 60
ttaagtccaa aacctgg 77
<210> 182
<211> 6
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 182
Gly Gly Ser Gly Gly Ser
1 5
<210> 183
<211> 24
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 183
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly
1 5 10 15
Gly Ser Gly Gly Ser Gly Gly Ser
20
<210> 184
<211> 4665
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 184
atgctcattg gctacgtgcg cgtctcaact aacgaccaga ataccgatct tcagaggaac 60

gcactggttt gtgcaggctg cgaacagatt ttcgaggaca aactcagcgg gacacggacg 120
gacagacctg gcctcaagcg agcactcaag aggctgcaga aaggagacac tctggtggtc 180
tggaaattgg accgcctggg tcgaagcatg aagcatctca tttctctggt tggcgaactg 240
cgagaaaggg ggatcaactt tcgaagtctg acggattcca tagatacaag cagccccatg 300
ggccggttct tcttctacgt gatgggtgca ctggctgaaa tggaaagaga actcattata 360
gagcgaacca tggcagggct tgcggctgcc aggaataaag gcaggcggtt tggaagacca 420
ccaaagggtg gatccggagg gtccggaggt agtggcggca gcggtggttc aggtggcagc 480
ggagggtcag gaggctctga taaaaagtat tctattggtt tagctatcgg cactaattcc 540
gttggatggg ctgtcataac cgatgaatac aaagtacctt caaagaaatt taaggtgttg 600
gggaacacag accgtcattc gattaaaaag aatcttatcg gtgccctcct attcgatagt 660
ggcgaaacgg cagaggcgac tcgcctgaaa cgaaccgctc ggagaaggta tacacgtcgc 720
aagaaccgaa tatgttactt acaagaaatt tttagcaatg agatggccaa agttgacgat 780
tctttctttc accgtttgga agagtccttc cttgtcgaag aggacaagaa acatgaacgg 840
caccccatct ttggaaacat agtagatgag gtggcatatc atgaaaagta cccaacgatt 900
tatcacctca gaaaaaagct agttgactca actgataaag cggacctgag gttaatctac 960
ttggctcttg cccatatgat aaagttccgt gggcactttc tcattgaggg tgatctaaat 1020
ccggacaact cggatgtcga caaactgttc atccagttag tacaaaccta taatcagttg 1080
tttgaagaga accctataaa tgcaagtggc gtggatgcga aggctattct tagcgcccgc 1140
ctctctaaat cccgacggct agaaaacctg atcgcacaat tacccggaga gaagaaaaat 1200
gggttgttcg gtaaccttat agcgctctca ctaggcctga caccaaattt taagtcgaac 1260
ttcgacttag ctgaagatgc caaattgcag cttagtaagg acacgtacga tgacgatctc 1320
gacaatctac tggcacaaat tggagatcag tatgcggact tatttttggc tgccaaaaac 1380
cttagcgatg caatcctcct atctgacata ctgagagtta atactgagat taccaaggcg 1440
ccgttatccg cttcaatgat caaaaggtac gatgaacatc accaagactt gacacttctc 1500
aaggccctag tccgtcagca actgcctgag aaatataagg aaatattctt tgatcagtcg 1560
aaaaacgggt acgcaggtta tattgacggc ggagcgagtc aagaggaatt ctacaagttt 1620
atcaaaccca tattagagaa gatggatggg acggaagagt tgcttgtaaa actcaatcgc 1680
gaagatctac tgcgaaagca gcggactttc gacaacggta gcattccaca tcaaatccac 1740
ttaggcgaat tgcatgctat acttagaagg caggaggatt tttatccgtt cctcaaagac 1800
aatcgtgaaa agattgagaa aatcctaacc tttcgcatac cttactatgt gggacccctg 1860
gcccgaggga actctcggtt cgcatggatg acaagaaagt ccgaagaaac gattactcca 1920
tggaattttg aggaagttgt cgataaaggt gcgtcagctc aatcgttcat cgagaggatg 1980
accaactttg acaagaattt accgaacgaa aaagtattgc ctaagcacag tttactttac 2040
gagtatttca cagtgtacaa tgaactcacg aaagttaagt atgtcactga gggcatgcgt 2100
aaacccgcct ttctaagcgg agaacagaag aaagcaatag tagatctgtt attcaagacc 2160
aaccgcaaag tgacagttaa gcaattgaaa gaggactact ttaagaaaat tgaatgcttc 2220
gattctgtcg agatctccgg ggtagaagat cgatttaatg cgtcacttgg tacgtatcat 2280
gacctcctaa agataattaa agataaggac ttcctggata acgaagagaa tgaagatatc 2340
ttagaagata tagtgttgac tcttaccctc tttgaagatc gggaaatgat tgaggaaaga 2400

ctaaaaacat acgctcacct gttcgacgat aaggttatga aacagttaaa gaggcgtcgc 2460
tatacgggct ggggacgatt gtcgcggaaa cttatcaacg ggataagaga caagcaaagt 2520
ggtaaaacta ttctcgattt tctaaagagc gacggcttcg ccaataggaa ctttatgcag 2580
ctgatccatg atgactcttt aaccttcaaa gaggatatac aaaaggcaca ggtttccgga 2640
caaggggact cattgcacga acatattgcg aatcttgctg gttcgccagc catcaaaaag 2700
ggcatactcc agacagtcaa agtagtggat gagctagtta aggtcatggg acgtcacaaa 2760
ccggaaaaca ttgtaatcga gatggcacgc gaaaatcaaa cgactcagaa ggggcaaaaa 2820
aacagtcgag agcggatgaa gagaatagaa gagggtatta aagaactggg cagccagatc 2880
ttaaaggagc atcctgtgga aaatacccaa ttgcagaacg agaaacttta cctctattac 2940
ctacaaaatg gaagggacat gtatgttgat caggaactgg acataaaccg tttatctgat 3000
tacgacgtcg atgccattgt accccaatcc tttttgaagg acgattcaat cgacaataaa 3060
gtgcttacac gctcggataa gaaccgaggg aaaagtgaca atgttccaag cgaggaagtc 3120
gtaaagaaaa tgaagaacta ttggcggcag ctcctaaatg cgaaactgat aacgcaaaga 3180
aagttcgata acttaactaa agctgagagg ggtggcttgt ctgaacttga caaggccgga 3240
tttattaaac gtcagctcgt ggaaacccgc caaatcacaa agcatgttgc acagatacta 3300
gattcccgaa tgaatacgaa atacgacgag aacgataagc tgattcggga agtcaaagta 3360
atcactttaa agtcaaaatt ggtgtcggac ttcagaaagg attttcaatt ctataaagtt 3420
agggagataa ataactacca ccatgcgcac gacgcttatc ttaatgccgt cgtagggacc 3480
gcactcatta agaaataccc gaagctagaa agtgagtttg tgtatggtga ttacaaagtt 3540
tatgacgtcc gtaagatgat cgcgaaaagc gaacaggaga taggcaaggc tacagccaaa 3600
tacttctttt attctaacat tatgaatttc tttaagacgg aaatcactct ggcaaacgga 3660
gagatacgca aacgaccttt aattgaaacc aatggggaga caggtgaaat cgtatgggat 3720
aagggccggg acttcgcgac ggtgagaaaa gttttgtcca tgccccaagt caacatagta 3780
aagaaaactg aggtgcagac cggagggttt tcaaaggaat cgattcttcc aaaaaggaat 3840
agtgataagc tcatcgctcg taaaaaggac tgggacccga aaaagtacgg tggcttcgat 3900
agccctacag ttgcctattc tgtcctagta gtggcaaaag ttgagaaggg aaaatccaag 3960
aaactgaagt cagtcaaaga attattgggg ataacgatta tggagcgctc gtcttttgaa 4020
aagaacccca tcgacttcct tgaggcgaaa ggttacaagg aagtaaaaaa ggatctcata 4080
attaaactac caaagtatag tctgtttgag ttagaaaatg gccgaaaacg gatgttggct 4140
agcgccggag agcttcaaaa ggggaacgaa ctcgcactac cgtctaaata cgtgaatttc 4200
ctgtatttag cgtcccatta cgagaagttg aaaggttcac ctgaagataa cgaacagaag 4260
caacttttg ttgagcagca caaacattat ctcgacgaaa tcatagagca aatttcggaa 4320
ttcagtaaga gagtcatcct agctgatgcc aatctggaca aagtattaag cgcatacaac 4380
aagcacaggg ataaacccat acgtgagcag gcggaaaata ttatccattt gtttactctt 4440
accaacctcg gcgctccagc cgcattcaag tattttgaca caacgataga tcgcaaacga 4500
tacacttcta ccaaggaggt gctagacgcg acactgattc accaatccat cacgggatta 4560
tatgaaactc ggatagattt gtcacagctt gggggtgacg gtggctccga ttataaggat 4620
gatgacgaca agggaggttc cccaaagaag aaaaggaagg tctga 4665

<210> 185

<211> 1554
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 185

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15
Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45
Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60
Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125
Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Gly Gly
130 135 140
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
145 150 155 160
Gly Gly Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile
165 170 175
Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val
180 185 190
Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile
195 200 205
Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala
210 215 220
Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
225 230 235 240
Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala
245 250 255
Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val

260 265 270

Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val

275 280 285

Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg

290 295 300

Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr

305 310 315 320

Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu

325 330 335

Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln

340 345 350

Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala

355 360 365

Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser

370 375 380

Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn

385 390 395 400

Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn

405 410 415

Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser

420 425 430

Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly

435 440 445

Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala

450 455 460

Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala

465 470 475 480

Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp

485 490 495

Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr

500 505 510

Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile

515 520 525

Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile

530 535 540

Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg

545 550 555 560

Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro

565 570 575

His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu
580 585 590
Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile
595 600 605
Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn
610 615 620
Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro
625 630 635 640
Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe
645 650 655
Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val
660 665 670
Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu
675 680 685
Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe
690 695 700
Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr
705 710 715 720
Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys
725 730 735
Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe
740 745 750
Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp
755 760 765
Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile
770 775 780
Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg
785 790 795 800
Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu
805 810 815
Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile
820 825 830
Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu
835 840 845
Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp
850 855 860
Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly
865 870 875 880
Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro

885 890 895
Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu
900 905 910
Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met
915 920 925
Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
930 935 940
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile
945 950 955 960
Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu
965 970 975
Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu
980 985 990
Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro
995 1000 1005
Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr
1010 1015 1020
Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
1025 1030 1035
Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn
1040 1045 1050
Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala
1055 1060 1065
Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys
1070 1075 1080
Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln
1085 1090 1095
Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys
1100 1105 1110
Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val
1115 1120 1125
Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile
1130 1135 1140
Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
1145 1150 1155
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
1160 1165 1170
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1175 1180 1185

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1190 1195 1200
Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1205 1210 1215
Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1220 1225 1230
Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1235 1240 1245
Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1250 1255 1260
Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1265 1270 1275
Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1280 1285 1290
Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1295 1300 1305
Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1310 1315 1320
Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1325 1330 1335
Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1340 1345 1350
Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1355 1360 1365
Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1370 1375 1380
Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1385 1390 1395
Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1400 1405 1410
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1415 1420 1425
His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1430 1435 1440
Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1445 1450 1455
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1460 1465 1470
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1475 1480 1485
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1490 1495 1500
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1505 1510 1515
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1520 1525 1530
Gly Gly Ser Asp Tyr Lys Asp Asp Asp Asp Lys Gly Gly Ser Pro
1535 1540 1545
Lys Lys Lys Arg Lys Val
1550

<210> 186
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 186
ggtggtagcg gtggatcc 18
<210> 187
<211> 45
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 187
ggtggatccg gtggttcagg tggcagcgga gggtcaggag gctct 45
<210> 188
<211> 72
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 188
ggtggatccg gagggtccgg aggtagtggc ggcagcggtg gttcaggtgg cagcggaggg 60
tcaggaggct ct 72
<210> 189
<211> 20
<212> DNA

<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 189
acctctgttt gggaaaattg 20
<210> 190
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 190
gcacactagt tagggataac a 21
<210> 191
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 191
gcctcagggc ctgtgatggg a 21
<210> 192
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 192
gctcagggcc tgtgatggga g 21
<210> 193
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 193
ggcccatgac ccttctcctc 20
<210> 194
<211> 20

<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 194
gcccatgacc cttctcctct 20
<210> 195
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 195
gacttgaaac actctttttc 20
<210> 196
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 196
gagttgaaga cacacaacac a 21
<210> 197
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 197
ggaactcatg tgattaactg 20
<210> 198
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 198
gtctacctct catgagccgg t 21
<210> 199

<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 199
gtttcccgca ggatgtggga t 21
<210> 200
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 200
gcctggggat ttatgttctt a 21
<210> 201
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 201
gaaatagcac aatgaatgga a 21
<210> 202
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 202
gacttttgg gggagaggga g 21
<210> 203
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 203
ggagacttaa gtccaaaacc 20

<210> 204
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 204
gtcagctatg atcacttccc t 21
<210> 205
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 205
gcagatgtag tgtttccaca 20
<210> 206
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 206
gggtgggggg agtttgctcc 20
<210> 207
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 207
gatatccgtt tatcagtgtc a 21
<210> 208
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 208

gttcctaagc ttgggctgca g 21

<210> 209

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 209

gcctaaaagt gactgggaga a 21

<210> 210

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 210

gcacagtccc atatttcttg g 21

<210> 211

<211> 77

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 211

cctttagtga aaagtagaca gctctgaata tgaaaggtag gttttcattt ctgggaaaga 60

gacgccaagt gatgtgg 77

<210> 212

<211> 78

<212> DNA

<213> 人工序列

<220>

<223> 合成多核苷酸

<400> 212

cctccaataa atatgggact atgtggaaag accaaaccta cgtttgattg gtgtacctga 60

aagtgacggg aagaatgg 78

<210> 213

<211> 78

<212> DNA

<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 213
ccattctgcc cgtcactttc aggtacacca atcaaacgta ggtttagtct tttcacatag 60
tcccatattt cttggagg 78
<210> 214
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 214
ccattctccc cgtcactttc aggtacaaca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 215
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 215
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt gttgtgg 77
<210> 216
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 216
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcatacttg 60
aaacactctt tttgtgg 77
<210> 217
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 217

ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 218
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 218
ccttgtgttg tgtttattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaatactctt tttgtgg 77
<210> 219
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 219
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcatacttg 60
aaacactctt tttgtgg 77
<210> 220
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 220
ccttgtattg tgagtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 221
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 221
ccttgtgttg tgtgtcttca actcacagag ttaaacgatg ctttacacag agtagacttg 60
aaacactctt tttctgg 77
<210> 222

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 222
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
taacactctt tttgtgg 77
<210> 223
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 223
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacgtg 60
aaacactctt tttgtgg 77
<210> 224
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 224
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 225
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 225
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag aggagacttg 60
taacactctt tttgtgg 77
<210> 226
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 226
cctgaggttt tccaggtttt aaaaggaaac ctaaaggtag gtttagcatt aagtgtcttg 60
aagtttattt taaaagg 77
<210> 227
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 227
ccaaaattcc cacaaaaccg aatgcatcag tcaaagcaag gtttgaagaa aagatttacc 60
acttcaggga gcttgg 76
<210> 228
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 228
ccttttctgg atatcgttga tgctctgtat gcaaaaggta ggtttttggg ttatgttgtt 60
aaacagtgat tgaatgg 77
<210> 229
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 229
cctccaagaa atatggaact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacaga gagaatgg 78
<210> 230
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 230

cctccaagaa atatgggact atgtgagaag accaaaccta cgtttgattg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 231
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 231
ccattctccc catcgctttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
ttccatattc tttggagg 78
<210> 232
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 232
ccattctccc catcactttc aggtgtaccg atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 233
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 233
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacggg gagaatgg 78
<210> 234
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 234
ccttcagggc agaaacagct ctactagcag agaaagcaag ctttcaatat tgtgcaatac 60
aaaaacgaga gcaggg 76
<210> 235

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 235
ccattctcct catctccttc tggtactcca atcaaacgta ggtttggtct tttctcatag 60
tctcatattt cttggagg 78
<210> 236
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 236
cctccaagac atataggact atgtgaaaat accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagtatgg 78
<210> 237
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 237
cctgccagat accagtagtc actgtgaatt acaaagctac gtttcttcca tagggaaagt 60
ttggagtcca gccagg 76
<210> 238
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 238
ccattctccc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 239
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 239
ccattctccc caccactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttgtagg 78
<210> 240
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 240
cctaaccaga aactaactaa tagatatggg cagaaagcat cctttcactt ttgttctggg 60
agagggaaga agcaaagg 78
<210> 241
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 241
ccattttggg gaggccttga tgggaagctg gaaaaggaag ctttcctccc agtcctgctg 60
aaggccttgc cagctgg 77
<210> 242
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 242
cctccaagaa acacaggact atgtgaaaag atcaaaccta cgtttgattg gtgttcctga 60
aagtgatggg gagaatgg 78
<210> 243
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 243

ccattctctt catgactttc aggtacacca ttgaaacgta ggtttggtct tttcacattg 60
tcccatattt cttggagg 78
<210> 244
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 244
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccgtattt cttggtgg 78
<210> 245
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 245
ccattctccc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggggg 78
<210> 246
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 246
ccattctccc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttgggg 77
<210> 247
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 247
cctccaagaa atatgagatt atatgaaaag accaaaccta cgtttgattg gtgtacttta 60
aagtgacggg gagaatgg 78
<210> 248

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 248
ccattctccc cgtcattttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccaaattt cttggagg 78
<210> 249
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 249
cccccaagaa atgtgggact atatgaaaag accaaaccta cgtttgactg gtgtacctaa 60
aagtgatggg gagaatgg 78
<210> 250
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 250
ccccaagaaa tgtgggacta tatgaaaaga ccaaacctac gtttgactgg tgtacctaaa 60
agtgatgggg agaatgg 77
<210> 251
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 251
cccattggtg ctgaccagat ggtgaaggag gcaaaggttg ctttgaatga ctgtgctctg 60
gggtgagcca ggcctgg 77
<210> 252
<211> 76
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 252
ccctttacag aggtgagctt tgttattagt aaaaaggtag gtttccctgt ttttctgaag 60
aaaagctgtg agtggg 76
<210> 253
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 253
ccactgccca ttgacagagt ggcgaggtgg gtgaaacctt gctttcctcc tggcccatgg 60
gcagggtggg gctgtggg 78
<210> 254
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 254
ccactgccca ttgacagagt ggcgaggtgg gtgaaacctt gctttcctcc tggcccatgg 60
gcagggtggg gctgtgg 77
<210> 255
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 255
ccattctccc tgtcactttt agatacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatgttt cttggagg 78
<210> 256
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 256

cctccaagaa atatcaactg tgtgaaaaga cgaaacctac gtttgattaa tgtacctgaa 60
agtgacaggg agaatgg 77
<210> 257
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 257
ccattctccc attaactttc aagtacacca atcaaaggta ggtttggtgt tttcccatag 60
tcccgtattt cttggagg 78
<210> 258
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 258
ccttttcatc atgccccttt cactttaagg tgaaaacctt gctttacatg tcagagaaaa 60
gaagagccct cagctggg 78
<210> 259
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 259
ccttttcatc atgccccttt cactttaagg tgaaaacctt gctttacatg tcagagaaaa 60
gaagagccct cagctgg 77
<210> 260
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 260
ccattcaccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 261

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 261
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgatgg tgtacccgaa 60
agtgacaggg agaatgg 77
<210> 262
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 262
ccaccaagaa atatgggact atgtgaaaag accaaaccta cgtttgatag gtatacctga 60
aagtgacagg gagaatgg 78
<210> 263
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 263
ccattctccc catcactttc aggtgcacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 264
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 264
ccctcaagaa atatgagact atgtgaaaag accaaaccta cgtttgactg gtatacctga 60
aagtgacagg gagaatgg 78
<210> 265
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 265
cctcaagaaa tatgagacta tgtgaaaaga ccaaacctac gtttgactgg tatacctgaa 60
agtgacaggg agaatgg 77
<210> 266
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 266
cctccaacaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacggg gataatgg 78
<210> 267
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 267
ccattctctc cctcactttc aagtacacca atcaaacgta ggtttggtct tttcacatag 60
tcttatattt cttggcgg 78
<210> 268
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 268
ccattctccc tgtcactgtc agtacaccaa tcaaacgtag gtttggtctc ttcacatagt 60
cccatatttc ttggagg 77
<210> 269
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 269

cctccaagaa atatgggact atgtgaacag accaaaccta cgtttgattg gtgtacctga 60
aagtgatggc agaatgg 77
<210> 270
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 270
ccaccatgcc tggccaccac acatttttt ctaaagcttg gttttggcca cagtgagagt 60
ttcttgggct gtcaggg 77
<210> 271
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 271
ccaccatgcc tggccaccac acatttttt ctaaagcttg gttttggcca cagtgagagt 60
ttcttgggct gtcagg 76
<210> 272
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 272
cccactaggt ggcgatatct gagggtccaa tgaaaccatg ctttttactc agatcttcca 60
ctaaccacct cccccgg 77
<210> 273
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 273
cctctaagaa atatgggact atgtgaaaag accaaaccta cgtttgactg gtgtacctga 60
aagtgacggg gagaatgg 78
<210> 274

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 274
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgatta gtgtacctga 60
aagtgacggg gagaatgg 78
<210> 275
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 275
ccattctccc tgtcactttc aggtacatca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 276
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 276
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgactg gtgtacctga 60
aagggatggg gagaatgg 78
<210> 277
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 277
cccccaagaa atatgagact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 278
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 278
ccccaagaaa tatgagacta tgtgaaaaga ccaaacctac gtttgattgg tgtacctgaa 60
agtgacaggg agaatgg 77
<210> 279
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 279
cctctaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtaactga 60
aagtgacagg gagaatgg 78
<210> 280
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 280
cctccaagaa atatgcgcct atgtgaaaag accaaaccta cgtttgattg gtatacctga 60
aagtgatgga gagaatgg 78
<210> 281
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 281
ccattctccc tgtcactttg aggtacacca atcaaacgta ggtttggtct tttcacatat 60
tcgcatattt cttggagg 78
<210> 282
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 282

ccattctccc cgtcactttc aggtacacca accaaacgtt ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 283
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 283
ccattctccc tgtcactttc cagtacacca gtcaaacgta ggtttggtct tttcacatac 60
tcccatattt cttggagg 78
<210> 284
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 284
cctggcctaa tttttaattc ttagtttgac ttaaaccttg cttttagtgt gatggcgaca 60
aaagctgagc tgaaagg 77
<210> 285
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 285
ccagtgcttt ttggttttaa aggcaagcct ccaaaccttc ctttctcctg gatgctgtgg 60
tggttgccat gcatgg 76
<210> 286
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 286
cccaactcct gcgagaagta gctcaccatg acaaagctac ctttgctttt atcgttttgc 60
aaaacaaaaa aggggg 76
<210> 287

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 287
ccattctccc cgtcactttg aggtgtgcca atcaaacgta ggtttggtct tttcacatag 60
tcctatattt cttggagg 78
<210> 288
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 288
cctccaaaaa atatgggact acgtaaaaag accaaaccta cgtttgattg gtgtacctga 60
aactgacagg gagaatgg 78
<210> 289
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 289
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
ttccatattt cttggagg 78
<210> 290
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 290
cctacaagat atatgggact atgtgaaaag accaaaccta cgttttactg gtgtgcctga 60
aactgacggg gagaatgg 78
<210> 291
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 291
ccattctctc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 292
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 292
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttcattg gtgtacctga 60
aagtgatagg gagaatgg 78
<210> 293
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 293
cctccaaaaa atatgggatg atgtgaaaag accaaaccta ggtttgactg gtgtacctga 60
aaatgatggg gagaatgg 78
<210> 294
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 294
cctccaagaa atatgagact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 295
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 295

cctccaagaa atatgggact acgtgaaaag atcaaaccta cgtttgattg ttgtacctga 60
aagtgatggg gagaatgg 78
<210> 296
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 296
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg ttgtacctga 60
aagtgatggg gagaatgg 78
<210> 297
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 297
cctcaaaagt gttctggttt tgttttgttt tttaaaccat ggttttacct ctggcttagt 60
gggactaaaa ataggagg 78
<210> 298
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 298
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgactg gtgtacctga 60
aagtgatggg gaaaatgg 78
<210> 299
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 299
cctccaagaa atatgggact gtgtgtaaag accaaaccta cgtttgattg gtgtacctca 60
aagtgatggg gagaatgg 78
<210> 300

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 300
ccattctccc catcacattc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 301
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 301
cccctggaaa agttggagca tcacaggaaa agcaaaccaa ccttttttct cccctaggta 60
aactggggag ccagggg 77
<210> 302
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 302
ccctggaaaa gttggagcat cacaggaaaa gcaaaccaac ctttttttctc ccctaggtaa 60
actggggagc cagggg 76
<210> 303
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 303
ccttccccag ttgcagcaga caagagtctc gaaaagcttg ctttggttgc tgcagtggat 60
gggttggtag gcacagg 77
<210> 304
<211> 76
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 304
cccccacctc ccaagctgct ggcttctcga ataaagctac ctttccttt accaaaactt 60
gtctctcgaa tgtcgg 76
<210> 305
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 305
ccttggccct ggacagctgc ttttccttcc ctaaaccttg gtttcccct ttgtgcaggt 60
gggtgggttt gggctgg 77
<210> 306
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 306
cctcttctag tgaacccatg gggttaccaa gggaaagcaa ccttttgata aatattccca 60
tctttttatg ttgtctgg 78
<210> 307
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 307
ccacttgaaa gggttaccaa ggataagatt tttaaagctt gctttcacaa acaactcatg 60
ctccaggctt gtcagtgg 78
<210> 308
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 308

cctttctccc catcactttc aggtacacca atcaaacgta ggtttgatct tttcacatag 60
tcccatattt cttggagg 78
<210> 309
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 309
ccattctccc catcaatttc agttacacca atgaaacgta ggtttggcct tttcacatag 60
tcccatattt cttagagg 78
<210> 310
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 310
ccattctccc tgtcactctc aggtacacca atcaaacgta ggtttggtct tttcatatag 60
tcccatattt cttggagg 78
<210> 311
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 311
cctccaagaa aattgggact atgtgaaaaa accaaaccta cgtttgattg atgtacctga 60
aagtgacagg agaatgg 77
<210> 312
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 312
ccttcaagaa atatgggact atgtgaaagg acaaaaccta cgtttattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 313

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 313
ccattctccc catcactttc aggtacgcta atcaaacgta ggtttgatct tttcacatag 60
tcttatattt cttggagg 78
<210> 314
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 314
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgactg gtgtacctca 60
atgtgacagg gagaatgg 78
<210> 315
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 315
ccattctccc tgtcactttt aggtacacca atcaaacgta cgtttggtct tttcacatag 60
acccatattt cttggagg 78
<210> 316
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 316
ccttcaagaa atatgggact gtgtgaaaag accaaagcta ggtttgattg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 317
<211> 76
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 317
cctactattc acagagtaat gcagtttgct gaaaaggttg gtttttgctg acctctgaga 60
gctcacatta cagtgg 76
<210> 318
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 318
ccattctctc tgtcactttc tggtacacca atcaaacgta ggtttgctct tttcacataa 60
tcccatattt attgaagg 78
<210> 319
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 319
ccataacatg tatttgctgg tgctagactc tccaaagcta ggtttctttc tacaacaatg 60
gctggaagtc ttcttgg 77
<210> 320
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 320
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tctcacacag 60
tcccatattt cttggagg 78
<210> 321
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 321

ccattcttcc cattactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccacattt cttggagg 78
<210> 322
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 322
ccattctccc cctcactttc aggtacacca atcaaacgta ggtttggtct tttcacattg 60
tcccatattt cttggagg 78
<210> 323
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 323
ccattctccc cagcacttac aggtacacca atcaaacgta ggtttggtca tttcacatag 60
tcccatattt cttggagg 78
<210> 324
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 324
ccattctccc tgtcactttc aggtacagca atcaaacgta ggtttggtct tttcacatgg 60
tcccatattt cttggagg 78
<210> 325
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 325
cctccaagaa atatgagact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacggg gaagatgg 78
<210> 326

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 326
cctccaagaa atatgagact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 327
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 327
cctccaagag atatgagact atgtaaatag accaaaccta cctttgattg gtgtacgtga 60
aagtgacagg aagaatgg 78
<210> 328
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 328
ccattctccc catcactttc aggtacacca accaaacgta ggtttggtct tttcacatag 60
tctcatattt cttggagg 78
<210> 329
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 329
cctccattga ctactcctta tcattggcta gaaaacctac ctttcaacca gtttctaagg 60
ccaagaaact tggagg 76
<210> 330
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 330
ccaccaagaa atatgggact acgtgaaaag accaaaccta cgtttgatgg gtgtgcctga 60
aagtgacggg aagaatgg 78
<210> 331
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 331
cctccaagaa ataagggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aggtgacagg gagaatgg 78
<210> 332
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 332
ccaaagggcc tttgtgattc tactttgtaa tataaaggat ggtttcttac tacggttggt 60
gtccttgcag gagtggg 77
<210> 333
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 333
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 334
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 334

ccattctccc cgttactttc aggtacacca ataaaaccta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 335
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 335
cccatatctc tggcaagggc agctctctgg ctaaaccaag ctttcctgta gagcttgagt 60
tccaaggcag cgttgg 76
<210> 336
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 336
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt gttgtgg 77
<210> 337
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 337
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcatacttg 60
aaacactctt tttgtgg 77
<210> 338
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 338
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 339

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 339
ccttgtgttg tgtttattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaatactctt tttgtgg 77
<210> 340
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 340
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcatacttg 60
aaacactctt tttgtgg 77
<210> 341
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 341
ccttgtattg tgagtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 342
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 342
ccttgtgttg tgtgtcttca actcacagag ttaaacgatg ctttacacag agtagacttg 60
aaacactctt tttctgg 77
<210> 343
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 343
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
taacactctt tttgtgg 77
<210> 344
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 344
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacgtg 60
aaacactctt tttgtgg 77
<210> 345
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 345
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 346
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 346
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag aggagacttg 60
taacactctt tttgtgg 77
<210> 347
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 347

cctttttcata agaagaaaat cgactcatca ttgaaaccaa gctttggtac aatttcattg 60
atgtttccag aagcagg 77
<210> 348
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 348
cccatagact atgatagaaa caaaataacc caaaagctag ctttctgatt gagtttccat 60
aaatgcaatg tgaagg 76
<210> 349
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 349
ccattcactt gtcactttct ggtacaccaa tcaaacgtag gtttggtctt ttcacatagt 60
ctcatatttc ttggagg 77
<210> 350
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 350
cctccaagaa atatgggact ctgtaaagag accaaaccta cgtttgattg gtgtacctga 60
aagtgaaggg gagaatgg 78
<210> 351
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 351
ccattctccc cgtcattttc aggtacacca atcaaaccta ggtttggtct ttttacatag 60
tcccatattt cttggagg 78
<210> 352

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 352
cctccacgaa acatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 353
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 353
ccaatttccc cctcactttc agatacacca atcaaacgta ggtttggtct tttcacatag 60
ttccatattt cctggagg 78
<210> 354
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 354
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatat 60
tcccatatgt cttggagg 78
<210> 355
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 355
cccaccggct catgagaggt agagctaagg tccaaaccta ggtttatctg agaccggaac 60
tcatgtgatt aactgtgg 78
<210> 356
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 356
ccaccggctc atgagaggta gagctaaggt ccaaacctag gtttatctga gaccggaact 60
catgtgatta actgtgg 77
<210> 357
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 357
ccttcaagaa atatgggact atgtgaagag accaaaccta cgtttgattg gtgtagccaa 60
aagtgatggg gaaaatgg 78
<210> 358
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 358
cctcagatta gatttacttg caaagagaca tttaaaggat cgttttgata ctattttgaa 60
agtactatac aaagatgg 78
<210> 359
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 359
ccttaagaac ataaatcccc aggaattcac agaaaccttg gtttgagctt tggatttccc 60
gcaggatgtg ggatagg 77
<210> 360
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 360

ccattctctc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttctcatag 60
tcccatattt cttggagg 78
<210> 361
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 361
ccatttacca tcattctctg tcatggcagg tgaaagcaag cttttatata gacaatgttc 60
tacttagttt acaggg 76
<210> 362
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 362
cccaaagtta attttactct ttttctgaat caaaaggaac ctttcctcca tgagaagaat 60
cctgccatat ttctagg 77
<210> 363
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 363
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg ctatacatga 60
aagtgacggg gagaatgg 78
<210> 364
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 364
ccttcaagaa atatgggact atgtgaaaag accaaaccta cctttgattg gtgtacctga 60
aagtgatggg aagaatgg 78
<210> 365

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 365
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatagtt cttggagg 78
<210> 366
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 366
ccattctccc cgtcactttc agggacaaca atcaaacgta ggtttggcct ttgcacatag 60
tcttatattt cttggagg 78
<210> 367
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 367
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 368
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 368
cctccaaaaa atatgggact atgtgagaag accaaaccta cgtttatta gtgtacctca 60
aagtgacagg gaggatgg 78
<210> 369
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 369
ccattctccc catcactttc aggtacacca atgaaacgta ggtttggcct tttcacatag 60
tttcatattt cttggagg 78
<210> 370
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 370
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 371
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 371
cctacaagaa atatggaact tgtaaaaaga ccaaacctac gtttgattgg tgtacctgaa 60
agtgacgggg agaatgg 77
<210> 372
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 372
cctccaagaa atatgggaca atgtgaaaag gccaaagcta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 373
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 373

cctttcaaac ttagaggtaa acaaaagtcc tgaaaaccta ggtttgacca taagttggga 60
ccatacgagc atagaagg 78
<210> 374
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 374
ccaaaaataa aaaaaaattg acttataagt aagaaaggtt cgttttctca cattcagaaa 60
gagaacccac atgttggg 78
<210> 375
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 375
ccaaaaataa aaaaaaattg acttataagt aagaaaggtt cgttttctca cattcagaaa 60
gagaacccac atgttgg 77
<210> 376
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 376
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 377
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 377
ccattctccc cgtcactttc aggtacacca atcaaacgtt ggtttagtct attcacatag 60
tcccatattt cttggagg 78
<210> 378

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 378
ccgaaaagaa taagactatc agctgaagtc ttaaaacgat cctttggccc ccagtactct 60
atatgcagga tagaaagg 78
<210> 379
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 379
cctacaaaaa taggggacta tgtgataaga ccaaacctac gtttgattgg tgtacctgaa 60
agtgatgggg agaatgg 77
<210> 380
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 380
ccattctacc catcactttc aggtacacca atcaaacgta ggtttggcct tttcatatag 60
tctcatattt cttggagg 78
<210> 381
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 381
ccattctccc catcactttc tggtatacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttagagg 78
<210> 382
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 382
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 383
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 383
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 384
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 384
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 385
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 385
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 386
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 386

ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 387
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 387
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 388
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 388
ccaccacacc cagccttatg ggatggtttt caaaagcatc ctttttaga agtggattct 60
gatatataat cggatgg 77
<210> 389
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 389
ccattctcaa tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 390
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 390
ccattctctc tgtcactttc aggtacacca gtcaaaggta ggtttgtttt attcacacgt 60
tcacatattt cttggagg 78
<210> 391

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 391
ccattcgccc catcactttc aggtacacta gtaaaacgta ggtttggtct tttcacatag 60
ttccatattt cttggagg 78
<210> 392
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 392
cctccaagaa atatgggact atgtgaagag atcaaaccta ggtttgattg ttgtacctga 60
aagtgataag aagaatgg 78
<210> 393
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 393
cctccaataa atatggggct atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 394
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 394
ccctttccc tgtcactttc aggtacacca gtcaaacgta ggtttggtct tttcacatag 60
tcgaatattt cttcaagg 78
<210> 395
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 395
cctttccct gtcactttca ggtacaccag tcaaacgtag gtttggtctt ttcacatagt 60
cgaatatttc ttcaagg 77
<210> 396
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 396
ccattctccc tgtcactttc aggtacacta atcaaacgta ggtttggtgt attcacacag 60
tcccatattt cttggagg 78
<210> 397
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 397
ccattcttcc tgtcactttc aggtatacca atcaaacgta ggtttggtct tttcacatag 60
tcccatgttt cttggagg 78
<210> 398
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 398
cctccaagaa atatgagact atatgaaaat accaaaccta cgtttgattg gtgtacctga 60
aagagacagg gagaatgg 78
<210> 399
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 399

ccattctccc tatcactttc aggtacacca atcaaacgta ggtttggtct tttcatgtag 60
tcccatattt cttggagg 78
<210> 400
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 400
ccattctgcc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 401
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 401
ccgtccgatt atatatcaga atctacttct aaaaaaggat gcttttgaaa accatcccat 60
aaggctgggt gtggtgg 77
<210> 402
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 402
cctacaagga atataggact atgtgaaaat accaaaccta cgtttcactg ctgtacctga 60
aggtgacagg gagaatgg 78
<210> 403
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 403
ccattctccc catcatttcc aggtaaacca atcaaaggta ggtttggtca tttcacatag 60
tcccatattt cttggagg 78
<210> 404

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 404
ccattctccc cgtcactttc aggtacacca gtcaaacgta ggtttggtct tttcacacag 60
tcccatattt cctggagg 78
<210> 405
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 405
ccattctccc catcactttc aggtacagca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 406
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 406
ccactacaga ttcttgggtc aagatgtgtg caaaaggatg ctttagggtg atggatatga 60
gtgggatgaa atgagg 76
<210> 407
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 407
cctgaaaaaa aaccctgcca gccagcaact ctgaaaggat gctttgtgtg agtgagcagt 60
gtctgagatg gacaggg 77
<210> 408
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 408
ccattctccc catcactttc aggtacgcca atcaaacgta ggtttggtct tttgacatag 60
tcccatattt cttggagg 78
<210> 409
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 409
ccgttctccc catcactttt aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tctcatattt cttggagg 78
<210> 410
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 410
ccattctcct ggtcactttc aggtatacca atcaaacgta ggtttggtct tttcatgtag 60
tcccatattt cttggagg 78
<210> 411
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 411
cctccaagaa atatgggact acatgaaaag accaaaccta cgtttgattg gtatacctga 60
aagtgaccag gagaatgg 78
<210> 412
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 412

cctccaagaa ctatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacggg gagaatgg 78
<210> 413
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 413
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatagtt cttggagg 78
<210> 414
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 414
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacagag 60
tcccatattt cttggagg 78
<210> 415
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 415
ccattctccc cgtcactttc atgtacacca agcaaacgta ggtttgatct ttccacatag 60
tcccgtgttt cttggagg 78
<210> 416
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 416
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacttga 60
aagtgacagg gagaatgg 78
<210> 417

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 417
cctccaagaa atatgggact atgtgaaaag acaaaaccta cgtttcactg gtgtacctga 60
aagtgacagg gaggatgg 78
<210> 418
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 418
cccccacctt ttaaaaacat gcatacatac ggaaacgttg ctttctgcac gatttcattt 60
taatggaaca gaacagg 77
<210> 419
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 419
ccatttcccc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tatcatattt cttggagg 78
<210> 420
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 420
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt ctggagg 77
<210> 421
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 421
cctttttgtta aagtaataga attctgcttc ttaaaggaac ctttcaggca agatggtggt 60
tagagcacct aaatggg 77
<210> 422
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 422
cctttttgtta aagtaataga attctgcttc ttaaaggaac ctttcaggca agatggtggt 60
tagagcacct aaatgg 76
<210> 423
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 423
cctccaagaa ctatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacggg gagaatgg 78
<210> 424
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 424
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggcct tttcacatag 60
tcccatagtt cttggagg 78
<210> 425
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 425

cctccaagaa atatgggact ggtgaaaaga ccaaacctac gtttgactgg tgtacctgaa 60
agtgacgggg agactgg 77
<210> 426
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 426
cctccaagaa acatgggaat gtgtgaaaag accaaaccta cgtttgattg gcgtacctga 60
aagtgacggg gagtatgg 78
<210> 427
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 427
cctccaagaa atatgggact gtgtgaaaag accaaaccta cgtttgattg gtatacctga 60
aagtgacaga gagaatgg 78
<210> 428
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 428
ccattctccc cttcactatc aggtacacca atcaaacgta ggtttagtct tttcacatag 60
tcccatattt cttggagg 78
<210> 429
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 429
ccattctccc cgtcactttc agatacacca gtcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 430

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 430
ccatcttact ttgtactaca ctgttcttta gagaaagctt ccttttggag accaaccagg 60
actccttaga agcagagg 78
<210> 431
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 431
ccatcttact ttgtactaca ctgttcttta gagaaagctt ccttttggag accaaccagg 60
actccttaga agcagagg 78
<210> 432
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 432
cctctgcttc taaggagtcc tggttggtct ccaaaaggaa gctttctcta aagaacagtg 60
tagtacaaag taagatgg 78
<210> 433
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 433
cctctgcttc taaggagtcc tggttggtct ccaaaaggaa gctttctcta aagaacagtg 60
tagtacaaag taagatgg 78
<210> 434
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 434
cctctgcttc taaggagtcc tggttggtct ccaaaaggaa gctttctcta aagaacagtg 60
tagtacaaag taagatgg 78
<210> 435
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 435
ccaccactgt gcctggccat tttcactatt cttaaaggaa gctttggttt acaaaggttt 60
gctactgtac ttccagg 77
<210> 436
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 436
ccattctccc tgtcactttc aggtacacca ttcaaacgta ggtttggtct tttctcatag 60
tcccatattt cttggagg 78
<210> 437
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 437
cctccaagaa attcgggact atgtgaaaag acaaaaccta cgtttaattg gtgtgtggtg 60
tacctgaaag tgacaagg 78
<210> 438
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 438

cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgaccag aagaatgg 78
<210> 439
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 439
cctccaagaa atatgggact atgtgaaaag cccaaaccta cgtttgactg atgtacctaa 60
agtgacgggg agaatgg 77
<210> 440
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 440
cccgcactgt gagcttggcc gagtgctgtc tgaaagcatc ctttcccttc acctggagac 60
tggagcgcca tagagg 76
<210> 441
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 441
cctgtctccc ccattccatg caaaataaaa cacaaaccaa gctttgcttt aagtgctccc 60
tgatgcagtt cagcgtgg 78
<210> 442
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 442
ccattcttcc cgtcacattc aggtacacca atcaaacgta ggtttggtct tttcccatag 60
tcccatattt cttagagg 78
<210> 443

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 443
ccccctgctc agcttgggga agaaaaatac aaaaacgatg cttttaggca ttttaaacaa 60
cttcactaca ttgaggg 77
<210> 444
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 444
ccccctgctc agcttgggga agaaaaatac aaaaacgatg cttttaggca ttttaaacaa 60
cttcactaca ttgagg 76
<210> 445
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 445
cctttgtgtt gtgtgtattc aactcacaga gtgaaacctt cctttattca gagcagtttt 60
gaaacactct ttttgtgg 78
<210> 446
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 446
cctttgtgtt gtgtgtattc aactcacaga gtgaaacctt cctttattca gagcagtttt 60
gaaaaacact ttttgtgg 78
<210> 447
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 447
cctttgtgtt gtgtgtattc aactcacaga gtgaaacctt cctttattca gagcagtttt 60
gaaaaactct ttttgtgg 78
<210> 448
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 448
cctttgtgtt gtgtgtattc aactcacaga gtgaaacctt cctttattca gagcagtttt 60
gaaacactct ttttgtgg 78
<210> 449
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 449
cctttgtgtt gtgtgtattc aactcacaga gtgaaacctt cctttattca gagcagtttt 60
gaaatactct ttttgtgg 78
<210> 450
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 450
cctttgtgtt gtgtgtattc aactcacaga gtgaaacctt cctttattca gagcagtttt 60
gaaacactct ttttgtgg 78
<210> 451
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 451

ccattctccc tgtcactttc aagtacacca atcaaaccta ggtttggtct tttcacatag 60
ttccatattt cttggagg 78
<210> 452
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 452
cccctcccat cacaggccct gaggtttaag agaaaaccat ggttttgtgg gccaggccca 60
tgacccttct cctctggg 78
<210> 453
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 453
cccctcccat cacaggccct gaggtttaag agaaaaccat ggttttgtgg gccaggccca 60
tgacccttct cctctgg 77
<210> 454
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 454
ccctcccatc acaggccctg aggtttaaga gaaaaccatg gttttgtggg ccaggcccat 60
gacccttctc ctctggg 77
<210> 455
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 455
ccctcccatc acaggccctg aggtttaaga gaaaaccatg gttttgtggg ccaggcccat 60
gacccttctc ctctgg 76
<210> 456

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 456
ccattctccc catcactttc aggtacacca atcaaacgta ggtttcatct tttcacatag 60
tcccacggtt tttggagg 78
<210> 457
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 457
cctccaagat atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aattgatggg gagaatgg 78
<210> 458
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 458
cctccaagaa atatgggact gtgtgaaaag aacaaaccta cgtttgattg gtgtacgtga 60
aagtgatggg gagaatgg 78
<210> 459
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 459
ccattcctcc cgtcactttc agatacacca aaaaaacgta ggtttggtct cttcacatag 60
tcccacattt cttggagg 78
<210> 460
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 460
cctccaagaa atgtgggact atgtgaagag accaaaccta cgtttttttg gtgtatctga 60
aagtgacggg aggaatgg 78
<210> 461
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 461
cctccaaggg gaatctgagt tctctgaaga caaaaagcat ggtttctttt cttctgtatt 60
tcttattgtt tcctagg 77
<210> 462
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 462
ccattctccc tatcactttc cagtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 463
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 463
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtatacttga 60
aattgacaag gagaatgg 78
<210> 464
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 464

cctccaagaa atatgggact atgtggaaag accaaaccta cgtttgactg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 465
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 465
cctctaagaa atatgggact atgtgaagag atgaaaccta cgtttgattg gtgtacctga 60
aagtgacgag gagaatgg 78
<210> 466
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 466
ccctcgtata ctacatgcta tagtcaaagc agtaaacctt cctttcctta agcagaccac 60
actctttcat gcctggg 77
<210> 467
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 467
cctcgtatac tacatgctat agtcaaagca gtaaaccttc ctttccttaa gcagaccaca 60
ctctttcatg cctggg 76
<210> 468
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 468
ccattctccc catcactttc aggtatacta atcaaaggta ggtttggtct tttcacatag 60
tcccatattt catggagg 78
<210> 469

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 469
ccattccccc gtcactttca ggtacaccaa tcaaacgtag gtttggtctt ttcacatagt 60
cccatatttc ttggagg 77
<210> 470
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 470
ccattctccc cgtcactttc aggtacacca atcaaacgta ggttttgtct tttcttatag 60
tcccatattt cttggagg 78
<210> 471
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 471
ccactgcacc tgaccaagat ccttaatttt tctaaaccta cgtttatcat ctataaaatg 60
agccatcttt tcacatgg 78
<210> 472
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 472
cctccgagaa atatgggact atgtgaaaag accaaaccta cgtttgattg ttgtacctga 60
aagtgacagg gagaatgg 78
<210> 473
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 473
ccattctccc catcactttt aggtacacca atcaaacgta ggtttggtcc ttttgcatag 60
acccatattt cttggagg 78
<210> 474
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 474
ccattttccc cgtcagtttc atatacacct atcaaacgta ggtttactgt tttcacatag 60
tcccttattt cttggagg 78
<210> 475
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 475
cctccaagaa atatgggact atgtgaaaag accaaaccta cctttgattg gtgtacctga 60
aagtgacggg caggatgg 78
<210> 476
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 476
ccattcttct cgtcattttc aagtacacca atcaaacgta ggtttggtct tttcgcatag 60
tcccatattt cttggagg 78
<210> 477
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 477

ccattcttct cgtcactttc aagtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 478
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 478
cctccaagaa atataggact atgtgaaaag accaaaccta cgtttgattg gtgtacttga 60
aagtgacagg gagaatgg 78
<210> 479
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 479
cctccaagaa atgtggaact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 480
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 480
ccctgacact gataaacgga tatgaagaga aaaaagctag gttttcgctg gaattcctaa 60
gcttgggctg cagtgg 76
<210> 481
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 481
cccttctccc agtcactttt aggtacacca atgaaacgta ggtttggtct tttcacacag 60
tcccatattt cttggagg 78
<210> 482

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 482
ccttctccca gtcactttta ggtacaccaa tgaaacgtag gtttggtctt ttcacacagt 60
cccatatttc ttggagg 77
<210> 483
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 483
ccactccctc tcccccaaaa agtaaaggta gaaaaccaag gtttacaggc aacaaatagc 60
acaatgaatg gaatgg 76
<210> 484
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 484
ccaaacccgc atcgcacacc ctgtgagggg gacaaaggaa cctttccgtt ccaacatcaa 60
ggttgttttg acccaagg 78
<210> 485
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 485
ccattctttc tgtcactttc aggtatacca gtcaaaccta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 486
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 486
ccattctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 487
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 487
ccacacggta gaggataaac taggtggatt ctcaaagcaa cctttgaaat aatctatgca 60
gtttttctgg gtactgg 77
<210> 488
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 488
ccaccaagaa acatgggact atgtgaaaag accaaaccta cgtttggttg gtgtacctgg 60
aagtgacggg gagagtgg 78
<210> 489
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 489
cctccaagaa atatgggacc atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 490
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 490

cctgtaaaaa ggtcacatgg tcaggtgtgc ctaaacgatc cttttattta tttatttatt 60
tatttttaag aaacagg 77
<210> 491
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 491
ccagccccaa aatgtcaggg gcttagaaca acaaaggttc cttttcatgt ttatactaca 60
tgtttgtcat gggctgg 77
<210> 492
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 492
ccgttttccc catcactttc aggtacacca gtcaaacgta ggtttggtct tttcacatgg 60
tcccacattt cttggagg 78
<210> 493
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 493
cctggaatag ctttcctgac tgtctgactt caaaaacctt ggtttgacca cttcgtctat 60
atcatgagga aggactgg 78
<210> 494
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 494
ccctactctg aacctacctt gataaagcct agaaaaccaa gctttgacaa gatttgacaa 60
gagatggaat ttggagg 77
<210> 495

<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 495
cctactctga acctaccttg ataaagccta gaaaaccaag ctttgacaag atttgacaag 60
agatggaatt tggagg 76
<210> 496
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 496
cccttataaa actgaaaact ttaacctttt ttaaagcatg cttttgaata aattctttta 60
ttacaaaaaa gaccagg 77
<210> 497
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 497
ccattctccc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacgtag 60
tcccatattt cttggagg 78
<210> 498
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 498
ccctttatta tccaagtggt ttcctgctct tcaaaccttc ctttcaaaat tttgtctcct 60
acttaaaaca agttagg 77
<210> 499
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 499
ccttctgttg agacctactg ctaagaaaac aaaaaaggtt cctttcaaat attattgtga 60
atcaataatg tacctgg 77
<210> 500
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 500
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttcattg atggacctga 60
aagtgatggg gagaatgg 78
<210> 501
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 501
ccattctccc ttcactttca gttacaccaa tcaaacgtag gtttggtctt ttcacatagt 60
cccatatttc ttggagg 77
<210> 502
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 502
cctagggaag tgatcatagc tgagtttctg gaaaaaccta ggttttaaag ttgaggagac 60
ttaagtccaa aacctgg 77
<210> 503
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 503

ccattctccc ttcactttca gttacaccaa tcaaacgtag gtttggtctt ttcacatagt 60
cccatatttc ttggagg 77
<210> 504
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 504
cctccaagaa atatgggact atgtgaaaag actaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 505
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 505
ccattctccc tgtcactttc aggtatgcca gtcaaacgta ggtttggtct tttcacatag 60
tcccatattc cttggagg 78
<210> 506
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 506
cctccaagaa atatgggact atgtaaaaag acgaaaccta cgtttgattg gtgtacttaa 60
aagtgacgag gagaatgg 78
<210> 507
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 507
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattt gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 508

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 508
ccattctccc cgtcactttc aggcacacca atcaaacgta ggtttagtct tttcacatag 60
tcccatattt cttagagg 78
<210> 509
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 509
ccttaatgca ttcatatttc atattttaaa taaaaccatg gtttcccaca gagtgacttc 60
tactctaaga aatgggg 77
<210> 510
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 510
ccttaatgca ttcatatttc atattttaaa taaaaccatg gtttcccaca gagtgacttc 60
tactctaaga aatggg 76
<210> 511
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 511
ccgttctttc cgtcactttc aggtacacca gtcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 512
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 512
ccattctccc catcactttc atgtacacca atcaaacgta ggtttggtct ttgttaacat 60
agtcccatat ttcttgg 77
<210> 513
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 513
ccctataaag cttagagaaa cacagggctc tttaaacgat ccttttctc ttttctgttt 60
taaatttcat cacttgg 77
<210> 514
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 514
cctataaagc ttagagaaac acagggctct ttaaacgatc ctttttctct tttctgtttt 60
aaatttcatc acttgg 76
<210> 515
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 515
ccattctccc catcactttc aggtacacta atcaaaggta ggtttggtct tttcacatgg 60
tcctatattt cttggagg 78
<210> 516
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 516

ccccatagca cgatcacatg ggacattcag gggaaagcaa ccttttccag gaaggaaaac 60
ccaatgctgg gacccagg 78
<210> 517
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 517
cccatagcac gatcacatgg gacattcagg ggaaagcaac cttttccagg aaggaaaacc 60
caatgctggg acccagg 77
<210> 518
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 518
ccctttcagc gctcacaggc tatggtttta taaaaggaac ctttgatttt gttcatgtga 60
aactacaaaa tgccagg 77
<210> 519
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 519
ccccatagca cgatcacatg ggacattcag gggaaagcaa ccttttccag gaaggaaaac 60
ccaatgctgg gacccagg 78
<210> 520
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 520
cccatagcac gatcacatgg gacattcagg ggaaagcaac cttttccagg aaggaaaacc 60
caatgctggg acccagg 77
<210> 521

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 521
cctccaagaa atattggagt atgtgataag accaaacctt cgtttgactg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 522
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 522
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 523
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 523
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 524
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 524
ccattctccc catcactttc aggtacacca gtcaaacgaa ggtttggtct tatcacatac 60
tccaatattt cttggagg 78
<210> 525
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 525
cctccaagat atatgggact atgtgaaaag gccaaaccta cctttgattg atacacctga 60
aaatgacagg gagaatgg 78
<210> 526
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 526
cctccaagaa atatgcgact atgtgaaaag accaaaccta cgtttcattg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 527
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 527
cctccaagaa atatgggact atgtggaaag accaaaccta cgtttgtttg gtgtacctga 60
aagtgagggg agaatgg 77
<210> 528
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 528
ccattctcct catcactttc aagtacacca atcaaacgta ggtttggtct tttcacatag 60
tcttatattt cttggagg 78
<210> 529
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 529

ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 530
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 530
ccattctccc catcactttc aggtacacca gtcaaacgaa ggtttggtct tatcacatac 60
tccaatattt cttggagg 78
<210> 531
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 531
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 532
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 532
ccagcagaag aatctggggc acagtctgtg aaaaaaggta cctttcttaa gcagggttct 60
tatccttcat gggtctgg 78
<210> 533
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 533
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg ttgtacctga 60
aagtgagggg gagaatgg 78
<210> 534

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 534
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 535
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 535
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 536
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 536
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 537
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 537
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacac agcagatttg 60
aaacactgtt tttctgg 77
<210> 538
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 538
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 539
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 539
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 540
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 540
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 541
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 541
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 542
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 542

ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 543
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 543
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 544
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 544
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 545
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 545
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 546
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 546
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 547

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 547
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacac agcagatttg 60
aaacactgtt tttctgg 77
<210> 548
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 548
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 549
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 549
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 550
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 550
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 551
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 551
ccttgtgttg tgtgtattca actcaccgag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 552
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 552
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 553
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 553
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 554
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 554
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 555
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 555

ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 556
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 556
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 557
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 557
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 558
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 558
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 559
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 559
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacac agcagatttg 60
aaacactgtt tttctgg 77
<210> 560

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 560
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 561
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 561
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 562
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 562
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 563
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 563
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 564
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 564
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 565
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 565
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 566
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 566
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
taacactgtt tttctgg 77
<210> 567
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 567
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 568
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 568

ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 569
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 569
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 570
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 570
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 571
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 571
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 572
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 572
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 573

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 573
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 574
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 574
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 575
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 575
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttgtgg 77
<210> 576
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 576
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 577
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 577
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 578
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 578
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 579
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 579
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 580
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 580
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 581
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 581

ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttccacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 582
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 582
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 583
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 583
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 584
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 584
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 585
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 585
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 586

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 586
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 587
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 587
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 588
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 588
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 589
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 589
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 590
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 590
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 591
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 591
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 592
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 592
ccttgtgttg tgtgtattca actcaccgag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 593
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 593
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 594
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 594

ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 595
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 595
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 596
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 596
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 597
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 597
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 598
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 598
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaaaactgtt tttctgg 77
<210> 599

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 599
ccttgtgttg tgtgtattca actcacagag taaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 600
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 600
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttgtgg 77
<210> 601
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 601
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 602
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 602
ccttgtgttg tgtgtattta actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 603
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 603
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 604
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 604
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 605
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 605
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 606
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 606
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 607
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 607

ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 608
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 608
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 609
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 609
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 610
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 610
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 611
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 611
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 612

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 612
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 613
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 613
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagatttg 60
aaacactgtt tttctgg 77
<210> 614
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 614
ccattctccc tatcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 615
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 615
cctcgtcact gccagatttt gtggctacca gcaaaggatc gttttaagct gcaactcagg 60
aaattgagaa aatatgg 77
<210> 616
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 616
cctccaagaa atatgggact atgtgaaaaa accaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 617
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 617
ccctgtgttc ttttatacta aaacaagcca gcaaaccaac ctttgagatg tgttgcctta 60
aacattactg aatgggg 77
<210> 618
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 618
ccctgtgttc ttttatacta aaacaagcca gcaaaccaac ctttgagatg tgttgcctta 60
aacattactg aatggg 76
<210> 619
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 619
ccgagaaacg gctttagcaa caaataaata tcaaaaggat gctttctctt cagaataatc 60
taaagtaagt tgggagg 77
<210> 620
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 620

ccatgttact ccggataagg acagcaaagg aggaaaggaa ccttttctgg gccaccagaa 60
ggatgagctt gggcttgg 78
<210> 621
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 621
cccagggata tgctggccac ggggaggagc cggaaaccaa cctttgtgtc actgtgtagt 60
gacaagtgcc tttggagg 78
<210> 622
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 622
ccagggatat gctggccacg gggaggagcc ggaaaccaac ctttgtgtca ctgtgtagtg 60
acaagtgcct ttggagg 77
<210> 623
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 623
ccttagggac ccataatggc cacaaccagg agaaaagcaa gctttgatgc ttaaacacta 60
cttacagaca tgtacagg 78
<210> 624
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 624
cctgcctctg ttcctccttc ctgatggtgg cggaaaggat gcttttgcca gatcaacagt 60
cacacacaac acaccagg 78
<210> 625

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 625
cctgactcca gccctccttg acaaggtctc cgtaaagcat gctttctctt agggaccctc 60
agagggaggc ttggtggg 78
<210> 626
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 626
cctgactcca gccctccttg acaaggtctc cgtaaagcat gctttctctt agggaccctc 60
agagggaggc ttggtgg 77
<210> 627
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 627
ccttatttgg aatgtgacaa gacccatttg tttaaacctt ggttttatg cagaaagaaa 60
aggaaggctg cagtggg 77
<210> 628
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 628
ccattctccc tgtcactttc aggtacacta atcaaacgta ggtttgctgt ttttacatag 60
gctcatattt cttggagg 78
<210> 629
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 629
ccattctccc catcactttc aggtacacca gtcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 630
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 630
cctgtttgtt attttagcta atgtcaaaaa gaaaaccttg ctttttctga acccttttcag 60
aggcagaaag tggggg 76
<210> 631
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 631
ccattttccc caccactttc acgtacagca atcaaacgta ggtttggtct tttcactagt 60
cccatatttc ttggagg 77
<210> 632
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 632
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt gttgtgg 77
<210> 633
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 633

ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcatacttg 60
aaacactctt tttgtgg 77
<210> 634
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 634
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 635
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 635
ccttgtgttg tgtttattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaatactctt tttgtgg 77
<210> 636
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 636
ccttgtagtg tgtgtattca actcacagag ttaaacgatc ctttacacag agcatacttg 60
aaacactctt tttgtgg 77
<210> 637
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 637
ccttgtattg tgagtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 638

<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 638
ccttgtgttg tgtgtcttca actcacagag ttaaacgatg ctttacacag agtagacttg 60
aaacactctt tttctgg 77
<210> 639
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 639
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
taacactctt tttgtgg 77
<210> 640
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 640
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacgtg 60
aaacactctt tttgtgg 77
<210> 641
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 641
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag agcagacttg 60
aaacactctt tttgtgg 77
<210> 642
<211> 77
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 642
ccttgtgttg tgtgtattca actcacagag ttaaacgatc ctttacacag aggagacttg 60
taacactctt tttgtgg 77
<210> 643
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 643
ccaggaaaaa atttaaactt tcttaacttg ataaaaggta gctttcaaaa cctacaataa 60
ataacatact tagagtgg 78
<210> 644
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 644
ccattctcct cgtcactttc aggtacacca aacaaacgta ggtttggtct ttttacgtag 60
tcccatattt cttggagg 78
<210> 645
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 645
ccctcttgaa gttagggaag tagcatttaa gggaaacgta gctttactat taagaatttc 60
aaacagcact tgtcaggg 78
<210> 646
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 646

ccctcttgaa gttagggaag tagcatttaa gggaaacgta gctttactat taagaatttc 60
aaacagcact tgtcagg 77
<210> 647
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 647
cctcttgaag ttagggaagt agcatttaag ggaaacgtag ctttactatt aagaatttca 60
aacagcactt gtcaggg 77
<210> 648
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 648
cctcttgaag ttagggaagt agcatttaag ggaaacgtag ctttactatt aagaatttca 60
aacagcactt gtcagg 76
<210> 649
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 649
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatat 60
tcccatattt cttggagg 78
<210> 650
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 650
ccattctccc ttcactttca ggtacaccaa tcaaacgtag gtttggtctt ttcacatagt 60
cccatatttt ttggagg 77
<210> 651

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 651
cctatagtct cagttacttg ggaggctgag gtaaaaggat cgtttgagcc caggaggtgg 60
aggttgcagt gagccggg 78
<210> 652
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 652
cctatagtct cagttacttg ggaggctgag gtaaaaggat cgtttgagcc caggaggtgg 60
aggttgcagt gagccgg 77
<210> 653
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 653
cctttcccaa ctctgctatt gcccccacat cctaaaggaa cctttctttt tttatatatt 60
ttattttaag ttccagg 77
<210> 654
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 654
cctccaagaa atatggaact atgtgaaaag accaaaccta cgtttgattg acgtacctga 60
aagtgacagg gagaatgg 78
<210> 655
<211> 76
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 655
cctcttctga aagcattgat aatcaacatt ttaaacgtag cttttcccca tattgctagg 60
aaggctcatt cccggg 76
<210> 656
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 656
cctccaagaa atatgggact atgtgaaaag gccaaaccta cgtttgattg ctgtacccga 60
gagtgacggg gagaatgg 78
<210> 657
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 657
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 658
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 658
cccggggcct gggtgcccag tgccagtggt cagaaaggtt gctttggtgt ttttcattgt 60
tagtgagaca gagatgg 77
<210> 659
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 659

ccggggcctg ggtgcccagt gccagtggtc agaaaggttg ctttggtgtt tttcattgtt 60
agtgagacag agatgg 76
<210> 660
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 660
ccattctccc catcattttc aggtacacca atcaaacgta ggtttgatct tttcacatag 60
ccccatattt cttggagg 78
<210> 661
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 661
ccaccagcac ttctgttaga agttgcagca gagaaaggat cctttaggca catctcccag 60
atccttgcga agagggg 77
<210> 662
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 662
cctgtgccag ggtccttcca ctgggactgg cagaaacgta ggtttgcatg gagtgagaag 60
caggggagag gttgaggg 78
<210> 663
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 663
cctgtgccag ggtccttcca ctgggactgg cagaaacgta ggtttgcatg gagtgagaag 60
caggggagag gttgagg 77
<210> 664

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 664
ccctcagcct ctcccctgct tctcactcca tgcaaaccta cgtttctgcc agtcccagca 60
gaaggaccct ggcacggg 78
<210> 665
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 665
ccctcagcct ctcccctgct tctcactcca tgcaaaccta cgtttctgcc agtcccagca 60
gaaggaccct ggcacgg 77
<210> 666
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 666
cctcagcctc tcccctgctt ctcactccat gcaaacctac gtttctgcca gtcccagcag 60
aaggaccctg gcacggg 77
<210> 667
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 667
cctcagcctc tcccctgctt ctcactccat gcaaacctac gtttctgcca gtcccagcag 60
aaggaccctg gcacgg 76
<210> 668
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 668
cctccaagaa atatggggct atgtgaaaag accaaaccta cctttgattg gtgtatctga 60
aagtgacggg gagaatgg 78
<210> 669
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 669
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattt gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 670
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 670
ccattctccc cgtcactttc aggtacacca atcaaacgta ggtttggtct tttctcattg 60
tcccatattt cttggagg 78
<210> 671
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 671
cccatcaaga gcggttgtgc atggcaacag taaaaggatg gtttgttaca ctagtacaaa 60
aagaggtggc cagagg 76
<210> 672
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 672

ccattctctc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 673
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 673
cctccaagaa atacgggact atgtgaaaag accaaacgta cgtttgattg gtgtacctga 60
aagtgatagg gagaatgg 78
<210> 674
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 674
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgactgg gagaatgg 78
<210> 675
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 675
ccattctccc tgtcactttc aggtacacga atcaaacgta ggtttcatct tttcacatag 60
tcccatattt cttagagg 78
<210> 676
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 676
ccattctctc tgtcactttc tggtacacca atcaaacgta ggtttggtct tttcacatag 60
tttcacatat ttcttgg 77
<210> 677

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 677
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgacaag gaaaatgg 78
<210> 678
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 678
cctgaaaaac attgtttcca acctggtaaa tcaaaaggaa ggtttaactt tgttagataa 60
gtccacatat caccaagg 78
<210> 679
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 679
cctccaagaa atgtgggact atgggaaaag accaaaccta cctttgtttg gtgtacctga 60
aagtgacggg gagaaagg 78
<210> 680
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 680
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttcattg gtgtacctga 60
aagtgatggg tagaatgg 78
<210> 681
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 681
cctacaagaa atatgggact atgggaaaag accaaaccta cgtttgattg gtacactgga 60
aagtgacagg gataatgg 78
<210> 682
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 682
ccattctccc tgtcactttc tggtacacca atcaaaggta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 683
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 683
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gtgtacctga 60
aagtgatggg gagaatgg 78
<210> 684
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 684
ccattctctt tgtcactttc aggtatacca atcaaacgtt ggtttggtct ttttgcatag 60
tcccatattt tgtggagg 78
<210> 685
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 685

cctccaagaa atatgagact atgtgaaaag accaaaccta cgtttgatta gtgtacctga 60
aaatgatggg gagaatgg 78
<210> 686
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 686
ccattctttc tgtcactttc aggtacacca atcaaacgta ggtttggtct tttcacatag 60
tcccatattt cttggagg 78
<210> 687
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 687
ccattctccc tgtcactttc aggtacacca atcaaacgta ggtttgttct tttcacatag 60
tcccatattt cttggagg 78
<210> 688
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 688
ccattatccc catcactttc aggtacacca atcaaacgta ggtttggttt tttcacatag 60
ttcaatattt ctttgagg 78
<210> 689
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 689
cctccaagaa atatgggact atctgaaaag atcaaaccta cgtttgattg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 690

<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 690
cctttctccc catcactttc aggtacacca atcaaacgta ggtttggtct tttcatatag 60
tcccatattt cttggagg 78
<210> 691
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 691
cctccaagaa atatgggact atgtgcaaag atcaaaccta cgtttgattg ctgtacctga 60
aagtgatggg gagaatgg 78
<210> 692
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 692
ccattctccc catcactttc aggtacacca gtcaaacgta ggtttggtct tttcacataa 60
tcccatattt cttggagg 78
<210> 693
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 693
cctccaagaa gtatgggacc atggaaaaga tcaaacctac gtttgactgg tgtacctgaa 60
agtgactggg agaatgg 77
<210> 694
<211> 78
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 694
cctccaagaa atatgggact atgtgaaaag accaaaccta cgtttgattg gagtacttga 60
aaatgacagg gataatgg 78
<210> 695
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 695
cctttaaaga catgctcttt gtgccagaaa ttcaaaggtt gcttttatgt ccagtggggt 60
ggagggagga agctcgg 77
<210> 696
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 696
ccattctccc cgtcactttc agggacctca atcaaacgta ggttttgtct tttcacatag 60
tcccatattt cttggagg 78
<210> 697
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 697
cctccaagaa atataggact atgtgaaaag accaaaccta cgtttgactg gtgtacctga 60
aagtgacagg gagaatgg 78
<210> 698
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 698

ccattctccc catcactttc aggtacacca atcaaaggta ggtttggtct tttcacatag 60
tccgatattt cctgcagg 78
<210> 699
<211> 12
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<220>
<221> misc_feature
<222> (4)..(5)
<223> s是g或c
<220>
<221> misc_feature
<222> (6)..(7)
<223> w是a, t或u
<220>
<221> misc_feature
<222> (8)..(9)
<223> s是g或c
<400> 699
aaasswwsst tt 12
<210> 700
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 700
ctgtaaaccg aggttttgga 20
<210> 701
<211> 15
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 701
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
1 5 10 15

<210> 702
<211> 7
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 702
Pro Lys Lys Lys Arg Lys Val
1 5
<210> 703
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 703
tccaaaacct cggtttacag 20
<210> 704
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 704
cccctcccat cacaggccct gaggtttaag agaaaaccat ggttttgtgg gccaggccca 60
tgacccttct cctctggg 78
<210> 705
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 705
cccagaggag aagggtcatg ggcctggccc acaaaaccat ggttttctct taaacctcag 60
ggcctgtgat gggagggg 78
<210> 706
<211> 76
<212> DNA
<213> 人工序列

<220>
<223> 合成多核苷酸
<400> 706
ccctgacact gataaacgga tatgaagaga aaaaagctag gttttcgctg gaattcctaa 60
gcttgggctg cagtgg 76
<210> 707
<211> 76
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 707
ccactgcagc ccaagcttag gaattccagc gaaaacctag ctttttctc ttcatatccg 60
tttatcagag tcaggg 76
<210> 708
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 708
cccttctccc agtcactttt aggtacacca atgaaacgta ggtttggtct tttcacacag 60
tcccatattt cttggagg 78
<210> 709
<211> 78
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 709
cctccaagaa atatgggact gtgtgaaaag accaaaccta cgtttcattg gtgtacctaa 60
aagtgactgg gagaaggg 78
<210> 710
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 710

ccctgacact gataaacgga tatgaagaga aaaaagctag gtttggtctt ttcacacagt 60
cccatatttc ttggagg 77
<210> 711
<211> 77
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 711
cctccaagaa atatgggact gtgtgaaaag accaaaccta gctttttttct cttcatatcc 60
gtttatcaga gtcaggg 77
<210> 712
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 712

Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Ala	Ile	Gly	Thr	Asn	Ser	Val	Gly
1				5				10					15		
Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro	Ser	Lys	Lys	Phe	Lys
		20					25					30			
Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	Lys	Asn	Leu	Ile	Gly
	35					40					45				
Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala	Thr	Arg	Leu	Lys
50					55					60					
Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg	Ile	Cys	Tyr
65				70					75					80	
Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp	Asp	Ser	Phe
			85					90					95		
Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp	Lys	Lys	His
		100					105					110			
Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val	Ala	Tyr	His
	115					120					125				
Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu	Val	Asp	Ser
130					135					140					
Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu	Ala	His	Met
145				150					155					160	
Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu	Asn	Pro	Asp

165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn

180 185 190

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

195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu

210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu

225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp

245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp

260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu

275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile

290 295 300

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

305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala

325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp

340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln

355 360 365

Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly

370 375 380

Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys

385 390 395 400

Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly

405 410 415

Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu

420 425 430

Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

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

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 713
<211> 142
<212> PRT

<213> 人工序列
<220>
<223> 合成多肽
<400> 713

```
Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1               5                  10                  15
Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
            20                  25                  30
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
        35                  40                  45
Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
    50                  55                  60
Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65                  70                  75                  80
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
                85                  90                  95
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
            100                 105                 110
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
        115                 120                 125
Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys
    130                 135                 140
```

<210> 714
<211> 1300
<212> PRT
<213> 新凶手弗朗西斯菌
<400> 714

```
Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1               5                  10                  15
Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
            20                  25                  30
Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
        35                  40                  45
Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
    50                  55                  60
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65                  70                  75                  80
Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
                85                  90                  95
```

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala

405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290

Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 715

<211> 1300

<212> PRT

<213> 新凶手弗朗西斯菌

<400> 715

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15
Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30
Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45
Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80
Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95
Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285

Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480
Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile

675 680 685

Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln

690 695 700

Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe

705 710 715 720

Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp

725 730 735

Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu

740 745 750

Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn

755 760 765

Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr

770 775 780

Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800

Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn

805 810 815

Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr

820 825 830

Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala

835 840 845

Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu

850 855 860

Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe

865 870 875 880

His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe

885 890 895

Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His

900 905 910

Ile Leu Ser Ile Ala Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095
Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu

1280 1285 1290

Phe Val Gln Asn Arg Asn Asn

1295 1300

<210> 716

<211> 1300

<212> PRT

<213> 新凶手弗朗西斯菌

<400> 716

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr

1 5 10 15

Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys

20 25 30

Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys

35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu

50 55 60

Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser

65 70 75 80

Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys

85 90 95

Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr

100 105 110

Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile

115 120 125

Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln

130 135 140

Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr

145 150 155 160

Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr

165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys
355 360 365
Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln
370 375 380
Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr
385 390 395 400
Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala
405 410 415
Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn
420 425 430
Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala
435 440 445
Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn
450 455 460
Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala
465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys
485 490 495
Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys
500 505 510
Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp
515 520 525
Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His
530 535 540
Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His
545 550 555 560
Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val
565 570 575
Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser
580 585 590
Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly
595 600 605
Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys
610 615 620
Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile
625 630 635 640
Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys
645 650 655
Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val
660 665 670
Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg

785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile
980 985 990
Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Ala Asp Leu
995 1000 1005
Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val
1010 1015 1020
Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu
1025 1030 1035
Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg
1040 1045 1050
Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly
1055 1060 1065
Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser
1070 1075 1080
Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys
1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp
1100 1105 1110
Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe
1115 1120 1125
Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr
1130 1135 1140
Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp
1145 1150 1155
Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu
1160 1165 1170
Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly
1175 1180 1185
Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe
1190 1195 1200
Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg
1205 1210 1215
Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val
1220 1225 1230
Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys
1235 1240 1245
Asn Met Pro Gln Asp Ala Asp Ala Asn Gly Ala Tyr His Ile Gly
1250 1255 1260
Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu
1265 1270 1275
Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 717
<211> 1300
<212> PRT
<213> 新凶手弗朗西斯菌
<400> 717

Met Ser Ile Tyr Gln Glu Phe Val Asn Lys Tyr Ser Leu Ser Lys Thr
1 5 10 15
Leu Arg Phe Glu Leu Ile Pro Gln Gly Lys Thr Leu Glu Asn Ile Lys
20 25 30
Ala Arg Gly Leu Ile Leu Asp Asp Glu Lys Arg Ala Lys Asp Tyr Lys
35 40 45

Lys Ala Lys Gln Ile Ile Asp Lys Tyr His Gln Phe Phe Ile Glu Glu
50 55 60
Ile Leu Ser Ser Val Cys Ile Ser Glu Asp Leu Leu Gln Asn Tyr Ser
65 70 75 80
Asp Val Tyr Phe Lys Leu Lys Lys Ser Asp Asp Asp Asn Leu Gln Lys
85 90 95
Asp Phe Lys Ser Ala Lys Asp Thr Ile Lys Lys Gln Ile Ser Glu Tyr
100 105 110
Ile Lys Asp Ser Glu Lys Phe Lys Asn Leu Phe Asn Gln Asn Leu Ile
115 120 125
Asp Ala Lys Lys Gly Gln Glu Ser Asp Leu Ile Leu Trp Leu Lys Gln
130 135 140
Ser Lys Asp Asn Gly Ile Glu Leu Phe Lys Ala Asn Ser Asp Ile Thr
145 150 155 160
Asp Ile Asp Glu Ala Leu Glu Ile Ile Lys Ser Phe Lys Gly Trp Thr
165 170 175
Thr Tyr Phe Lys Gly Phe His Glu Asn Arg Lys Asn Val Tyr Ser Ser
180 185 190
Asn Asp Ile Pro Thr Ser Ile Ile Tyr Arg Ile Val Asp Asp Asn Leu
195 200 205
Pro Lys Phe Leu Glu Asn Lys Ala Lys Tyr Glu Ser Leu Lys Asp Lys
210 215 220
Ala Pro Glu Ala Ile Asn Tyr Glu Gln Ile Lys Lys Asp Leu Ala Glu
225 230 235 240
Glu Leu Thr Phe Asp Ile Asp Tyr Lys Thr Ser Glu Val Asn Gln Arg
245 250 255
Val Phe Ser Leu Asp Glu Val Phe Glu Ile Ala Asn Phe Asn Asn Tyr
260 265 270
Leu Asn Gln Ser Gly Ile Thr Lys Phe Asn Thr Ile Ile Gly Gly Lys
275 280 285
Phe Val Asn Gly Glu Asn Thr Lys Arg Lys Gly Ile Asn Glu Tyr Ile
290 295 300
Asn Leu Tyr Ser Gln Gln Ile Asn Asp Lys Thr Leu Lys Lys Tyr Lys
305 310 315 320
Met Ser Val Leu Phe Lys Gln Ile Leu Ser Asp Thr Glu Ser Lys Ser
325 330 335
Phe Val Ile Asp Lys Leu Glu Asp Asp Ser Asp Val Val Thr Thr Met
340 345 350
Gln Ser Phe Tyr Glu Gln Ile Ala Ala Phe Lys Thr Val Glu Glu Lys

355 360 365

Ser Ile Lys Glu Thr Leu Ser Leu Leu Phe Asp Asp Leu Lys Ala Gln

370 375 380

Lys Leu Asp Leu Ser Lys Ile Tyr Phe Lys Asn Asp Lys Ser Leu Thr

385 390 395 400

Asp Leu Ser Gln Gln Val Phe Asp Asp Tyr Ser Val Ile Gly Thr Ala

405 410 415

Val Leu Glu Tyr Ile Thr Gln Gln Ile Ala Pro Lys Asn Leu Asp Asn

420 425 430

Pro Ser Lys Lys Glu Gln Glu Leu Ile Ala Lys Lys Thr Glu Lys Ala

435 440 445

Lys Tyr Leu Ser Leu Glu Thr Ile Lys Leu Ala Leu Glu Glu Phe Asn

450 455 460

Lys His Arg Asp Ile Asp Lys Gln Cys Arg Phe Glu Glu Ile Leu Ala

465 470 475 480

Asn Phe Ala Ala Ile Pro Met Ile Phe Asp Glu Ile Ala Gln Asn Lys

485 490 495

Asp Asn Leu Ala Gln Ile Ser Ile Lys Tyr Gln Asn Gln Gly Lys Lys

500 505 510

Asp Leu Leu Gln Ala Ser Ala Glu Asp Asp Val Lys Ala Ile Lys Asp

515 520 525

Leu Leu Asp Gln Thr Asn Asn Leu Leu His Lys Leu Lys Ile Phe His

530 535 540

Ile Ser Gln Ser Glu Asp Lys Ala Asn Ile Leu Asp Lys Asp Glu His

545 550 555 560

Phe Tyr Leu Val Phe Glu Glu Cys Tyr Phe Glu Leu Ala Asn Ile Val

565 570 575

Pro Leu Tyr Asn Lys Ile Arg Asn Tyr Ile Thr Gln Lys Pro Tyr Ser

580 585 590

Asp Glu Lys Phe Lys Leu Asn Phe Glu Asn Ser Thr Leu Ala Asn Gly

595 600 605

Trp Asp Lys Asn Lys Glu Pro Asp Asn Thr Ala Ile Leu Phe Ile Lys

610 615 620

Asp Asp Lys Tyr Tyr Leu Gly Val Met Asn Lys Lys Asn Asn Lys Ile

625 630 635 640

Phe Asp Asp Lys Ala Ile Lys Glu Asn Lys Gly Glu Gly Tyr Lys Lys

645 650 655

Ile Val Tyr Lys Leu Leu Pro Gly Ala Asn Lys Met Leu Pro Lys Val

660 665 670

Phe Phe Ser Ala Lys Ser Ile Lys Phe Tyr Asn Pro Ser Glu Asp Ile
675 680 685
Leu Arg Ile Arg Asn His Ser Thr His Thr Lys Asn Gly Ser Pro Gln
690 695 700
Lys Gly Tyr Glu Lys Phe Glu Phe Asn Ile Glu Asp Cys Arg Lys Phe
705 710 715 720
Ile Asp Phe Tyr Lys Gln Ser Ile Ser Lys His Pro Glu Trp Lys Asp
725 730 735
Phe Gly Phe Arg Phe Ser Asp Thr Gln Arg Tyr Asn Ser Ile Asp Glu
740 745 750
Phe Tyr Arg Glu Val Glu Asn Gln Gly Tyr Lys Leu Thr Phe Glu Asn
755 760 765
Ile Ser Glu Ser Tyr Ile Asp Ser Val Val Asn Gln Gly Lys Leu Tyr
770 775 780
Leu Phe Gln Ile Tyr Asn Lys Asp Phe Ser Ala Tyr Ser Lys Gly Arg
785 790 795 800
Pro Asn Leu His Thr Leu Tyr Trp Lys Ala Leu Phe Asp Glu Arg Asn
805 810 815
Leu Gln Asp Val Val Tyr Lys Leu Asn Gly Glu Ala Glu Leu Phe Tyr
820 825 830
Arg Lys Gln Ser Ile Pro Lys Lys Ile Thr His Pro Ala Lys Glu Ala
835 840 845
Ile Ala Asn Lys Asn Lys Asp Asn Pro Lys Lys Glu Ser Val Phe Glu
850 855 860
Tyr Asp Leu Ile Lys Asp Lys Arg Phe Thr Glu Asp Lys Phe Phe Phe
865 870 875 880
His Cys Pro Ile Thr Ile Asn Phe Lys Ser Ser Gly Ala Asn Lys Phe
885 890 895
Asn Asp Glu Ile Asn Leu Leu Leu Lys Glu Lys Ala Asn Asp Val His
900 905 910
Ile Leu Ser Ile Asp Arg Gly Glu Arg His Leu Ala Tyr Tyr Thr Leu
915 920 925
Val Asp Gly Lys Gly Asn Ile Ile Lys Gln Asp Thr Phe Asn Ile Ile
930 935 940
Gly Asn Asp Arg Met Lys Thr Asn Tyr His Asp Lys Leu Ala Ala Ile
945 950 955 960
Glu Lys Asp Arg Asp Ser Ala Arg Lys Asp Trp Lys Lys Ile Asn Asn
965 970 975
Ile Lys Glu Met Lys Glu Gly Tyr Leu Ser Gln Val Val His Glu Ile

980 985 990

Ala Lys Leu Val Ile Glu Tyr Asn Ala Ile Val Val Phe Glu Asp Leu

995 1000 1005

Asn Phe Gly Phe Lys Arg Gly Arg Phe Lys Val Glu Lys Gln Val

1010 1015 1020

Tyr Gln Lys Leu Glu Lys Met Leu Ile Glu Lys Leu Asn Tyr Leu

1025 1030 1035

Val Phe Lys Asp Asn Glu Phe Asp Lys Thr Gly Gly Val Leu Arg

1040 1045 1050

Ala Tyr Gln Leu Thr Ala Pro Phe Glu Thr Phe Lys Lys Met Gly

1055 1060 1065

Lys Gln Thr Gly Ile Ile Tyr Tyr Val Pro Ala Gly Phe Thr Ser

1070 1075 1080

Lys Ile Cys Pro Val Thr Gly Phe Val Asn Gln Leu Tyr Pro Lys

1085 1090 1095

Tyr Glu Ser Val Ser Lys Ser Gln Glu Phe Phe Ser Lys Phe Asp

1100 1105 1110

Lys Ile Cys Tyr Asn Leu Asp Lys Gly Tyr Phe Glu Phe Ser Phe

1115 1120 1125

Asp Tyr Lys Asn Phe Gly Asp Lys Ala Ala Lys Gly Lys Trp Thr

1130 1135 1140

Ile Ala Ser Phe Gly Ser Arg Leu Ile Asn Phe Arg Asn Ser Asp

1145 1150 1155

Lys Asn His Asn Trp Asp Thr Arg Glu Val Tyr Pro Thr Lys Glu

1160 1165 1170

Leu Glu Lys Leu Leu Lys Asp Tyr Ser Ile Glu Tyr Gly His Gly

1175 1180 1185

Glu Cys Ile Lys Ala Ala Ile Cys Gly Glu Ser Asp Lys Lys Phe

1190 1195 1200

Phe Ala Lys Leu Thr Ser Val Leu Asn Thr Ile Leu Gln Met Arg

1205 1210 1215

Asn Ser Lys Thr Gly Thr Glu Leu Asp Tyr Leu Ile Ser Pro Val

1220 1225 1230

Ala Asp Val Asn Gly Asn Phe Phe Asp Ser Arg Gln Ala Pro Lys

1235 1240 1245

Asn Met Pro Gln Asp Ala Ala Ala Asn Gly Ala Tyr His Ile Gly

1250 1255 1260

Leu Lys Gly Leu Met Leu Leu Gly Arg Ile Lys Asn Asn Gln Glu

1265 1270 1275

Gly Lys Lys Leu Asn Leu Val Ile Lys Asn Glu Glu Tyr Phe Glu
1280 1285 1290
Phe Val Gln Asn Arg Asn Asn
1295 1300

<210> 718
<211> 887
<212> PRT
<213> 格氏嗜盐碱杆菌
<400> 718

Met Thr Val Ile Asp Leu Asp Ser Thr Thr Thr Ala Asp Glu Leu Thr
1 5 10 15
Ser Gly His Thr Tyr Asp Ile Ser Val Thr Leu Thr Gly Val Tyr Asp
20 25 30
Asn Thr Asp Glu Gln His Pro Arg Met Ser Leu Ala Phe Glu Gln Asp
35 40 45
Asn Gly Glu Arg Arg Tyr Ile Thr Leu Trp Lys Asn Thr Thr Pro Lys
50 55 60
Asp Val Phe Thr Tyr Asp Tyr Ala Thr Gly Ser Thr Tyr Ile Phe Thr
65 70 75 80
Asn Ile Asp Tyr Glu Val Lys Asp Gly Tyr Glu Asn Leu Thr Ala Thr
85 90 95
Tyr Gln Thr Thr Val Glu Asn Ala Thr Ala Gln Glu Val Gly Thr Thr
100 105 110
Asp Glu Asp Glu Thr Phe Ala Gly Gly Glu Pro Leu Asp His His Leu
115 120 125
Asp Asp Ala Leu Asn Glu Thr Pro Asp Asp Ala Glu Thr Glu Ser Asp
130 135 140
Ser Gly His Val Met Thr Ser Phe Ala Ser Arg Asp Gln Leu Pro Glu
145 150 155 160
Trp Thr Leu His Thr Tyr Thr Leu Thr Ala Thr Asp Gly Ala Lys Thr
165 170 175
Asp Thr Glu Tyr Ala Arg Arg Thr Leu Ala Tyr Thr Val Arg Gln Glu
180 185 190
Leu Tyr Thr Asp His Asp Ala Ala Pro Val Ala Thr Asp Gly Leu Met
195 200 205
Leu Leu Thr Pro Glu Pro Leu Gly Glu Thr Pro Leu Asp Leu Asp Cys
210 215 220
Gly Val Arg Val Glu Ala Asp Glu Thr Arg Thr Leu Asp Tyr Thr Thr
225 230 235 240

Ala Lys Asp Arg Leu Leu Ala Arg Glu Leu Val Glu Glu Gly Leu Lys
245 250 255
Arg Ser Leu Trp Asp Asp Tyr Leu Val Arg Gly Ile Asp Glu Val Leu
260 265 270
Ser Lys Glu Pro Val Leu Thr Cys Asp Glu Phe Asp Leu His Glu Arg
275 280 285
Tyr Asp Leu Ser Val Glu Val Gly His Ser Gly Arg Ala Tyr Leu His
290 295 300
Ile Asn Phe Arg His Arg Phe Val Pro Lys Leu Thr Leu Ala Asp Ile
305 310 315 320
Asp Asp Asp Asn Ile Tyr Pro Gly Leu Arg Val Lys Thr Thr Tyr Arg
325 330 335
Pro Arg Arg Gly His Ile Val Trp Gly Leu Arg Asp Glu Cys Ala Thr
340 345 350
Asp Ser Leu Asn Thr Leu Gly Asn Gln Ser Val Val Ala Tyr His Arg
355 360 365
Asn Asn Gln Thr Pro Ile Asn Thr Asp Leu Leu Asp Ala Ile Glu Ala
370 375 380
Ala Asp Arg Arg Val Val Glu Thr Arg Arg Gln Gly His Gly Asp Asp
385 390 395 400
Ala Val Ser Phe Pro Gln Glu Leu Leu Ala Val Glu Pro Asn Thr His
405 410 415
Gln Ile Lys Gln Phe Ala Ser Asp Gly Phe His Gln Gln Ala Arg Ser
420 425 430
Lys Thr Arg Leu Ser Ala Ser Arg Cys Ser Glu Lys Ala Gln Ala Phe
435 440 445
Ala Glu Arg Leu Asp Pro Val Arg Leu Asn Gly Ser Thr Val Glu Phe
450 455 460
Ser Ser Glu Phe Phe Thr Gly Asn Asn Glu Gln Gln Leu Arg Leu Leu
465 470 475 480
Tyr Glu Asn Gly Glu Ser Val Leu Thr Phe Arg Asp Gly Ala Arg Gly
485 490 495
Ala His Pro Asp Glu Thr Phe Ser Lys Gly Ile Val Asn Pro Pro Glu
500 505 510
Ser Phe Glu Val Ala Val Val Leu Pro Glu Gln Gln Ala Asp Thr Cys
515 520 525
Lys Ala Gln Trp Asp Thr Met Ala Asp Leu Leu Asn Gln Ala Gly Ala
530 535 540
Pro Pro Thr Arg Ser Glu Thr Val Gln Tyr Asp Ala Phe Ser Ser Pro

545 550 555 560
Glu Ser Ile Ser Leu Asn Val Ala Gly Ala Ile Asp Pro Ser Glu Val
565 570 575
Asp Ala Ala Phe Val Val Leu Pro Pro Asp Gln Glu Gly Phe Ala Asp
580 585 590
Leu Ala Ser Pro Thr Glu Thr Tyr Asp Glu Leu Lys Lys Ala Leu Ala
595 600 605
Asn Met Gly Ile Tyr Ser Gln Met Ala Tyr Phe Asp Arg Phe Arg Asp
610 615 620
Ala Lys Ile Phe Tyr Thr Arg Asn Val Ala Leu Gly Leu Leu Ala Ala
625 630 635 640
Ala Gly Gly Val Ala Phe Thr Thr Glu His Ala Met Pro Gly Asp Ala
645 650 655
Asp Met Phe Ile Gly Ile Asp Val Ser Arg Ser Tyr Pro Glu Asp Gly
660 665 670
Ala Ser Gly Gln Ile Asn Ile Ala Ala Thr Ala Thr Ala Val Tyr Lys
675 680 685
Asp Gly Thr Ile Leu Gly His Ser Ser Thr Arg Pro Gln Leu Gly Glu
690 695 700
Lys Leu Gln Ser Thr Asp Val Arg Asp Ile Met Lys Asn Ala Ile Leu
705 710 715 720
Gly Tyr Gln Gln Val Thr Gly Glu Ser Pro Thr His Ile Val Ile His
725 730 735
Arg Asp Gly Phe Met Asn Glu Asp Leu Asp Pro Ala Thr Glu Phe Leu
740 745 750
Asn Glu Gln Gly Val Glu Tyr Asp Ile Val Glu Ile Arg Lys Gln Pro
755 760 765
Gln Thr Arg Leu Leu Ala Val Ser Asp Val Gln Tyr Asp Thr Pro Val
770 775 780
Lys Ser Ile Ala Ala Ile Asn Gln Asn Glu Pro Arg Ala Thr Val Ala
785 790 795 800
Thr Phe Gly Ala Pro Glu Tyr Leu Ala Thr Arg Asp Gly Gly Gly Leu
805 810 815
Pro Arg Pro Ile Gln Ile Glu Arg Val Ala Gly Glu Thr Asp Ile Glu
820 825 830
Thr Leu Thr Arg Gln Val Tyr Leu Leu Ser Gln Ser His Ile Gln Val
835 840 845
His Asn Ser Thr Ala Arg Leu Pro Ile Thr Thr Ala Tyr Ala Asp Gln
850 855 860

Ala Ser Thr His Ala Thr Lys Gly Tyr Leu Val Gln Thr Gly Ala Phe
865 870 875 880
Glu Ser Asn Val Gly Phe Leu
885

<210> 719
<211> 1544
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 719

Met Leu Ile Gly Tyr Val Arg Val Ser Thr Asn Asp Gln Asn Thr Asp
1 5 10 15
Leu Gln Arg Asn Ala Leu Val Cys Ala Gly Cys Glu Gln Ile Phe Glu
20 25 30
Asp Lys Leu Ser Gly Thr Arg Thr Asp Arg Pro Gly Leu Lys Arg Ala
35 40 45
Leu Lys Arg Leu Gln Lys Gly Asp Thr Leu Val Val Trp Lys Leu Asp
50 55 60
Arg Leu Gly Arg Ser Met Lys His Leu Ile Ser Leu Val Gly Glu Leu
65 70 75 80
Arg Glu Arg Gly Ile Asn Phe Arg Ser Leu Thr Asp Ser Ile Asp Thr
85 90 95
Ser Ser Pro Met Gly Arg Phe Phe Phe Tyr Val Met Gly Ala Leu Ala
100 105 110
Glu Met Glu Arg Glu Leu Ile Ile Glu Arg Thr Met Ala Gly Leu Ala
115 120 125
Ala Ala Arg Asn Lys Gly Arg Arg Phe Gly Arg Pro Pro Lys Gly Gly
130 135 140
Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser
145 150 155 160
Gly Gly Ser Gly Gly Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile
165 170 175
Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val
180 185 190
Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile
195 200 205
Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala
210 215 220

Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg
225 230 235 240
Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala
245 250 255
Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val
260 265 270
Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val
275 280 285
Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg
290 295 300
Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr
305 310 315 320
Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu
325 330 335
Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln
340 345 350
Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala
355 360 365
Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser
370 375 380
Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn
385 390 395 400
Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn
405 410 415
Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser
420 425 430
Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly
435 440 445
Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala
450 455 460
Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala
465 470 475 480
Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp
485 490 495
Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr
500 505 510
Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile
515 520 525
Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile

530 535 540

Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg
545 550 555 560
Glu Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro
565 570 575
His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu
580 585 590
Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile
595 600 605
Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn
610 615 620
Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro
625 630 635 640
Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe
645 650 655
Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val
660 665 670
Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu
675 680 685
Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe
690 695 700
Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr
705 710 715 720
Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys
725 730 735
Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe
740 745 750
Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp
755 760 765
Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile
770 775 780
Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg
785 790 795 800
Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu
805 810 815
Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile
820 825 830
Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu
835 840 845

Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp
850 855 860
Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly
865 870 875 880
Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro
885 890 895
Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu
900 905 910
Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met
915 920 925
Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
930 935 940
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile
945 950 955 960
Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu
965 970 975
Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu
980 985 990
Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro
995 1000 1005
Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr
1010 1015 1020
Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu
1025 1030 1035
Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn
1040 1045 1050
Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala
1055 1060 1065
Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys
1070 1075 1080
Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln
1085 1090 1095
Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys
1100 1105 1110
Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys Leu Val
1115 1120 1125
Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu Ile
1130 1135 1140
Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val

1145 1150 1155

Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
1160 1165 1170

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1175 1180 1185

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1190 1195 1200

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1205 1210 1215

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
1220 1225 1230

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
1235 1240 1245

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
1250 1255 1260

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
1265 1270 1275

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
1280 1285 1290

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
1295 1300 1305

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
1310 1315 1320

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
1325 1330 1335

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
1340 1345 1350

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu
1355 1360 1365

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly
1370 1375 1380

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val
1385 1390 1395

Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
1400 1405 1410

Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys
1415 1420 1425

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
1430 1435 1440

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala
1445 1450 1455
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1460 1465 1470
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1475 1480 1485
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1490 1495 1500
Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1505 1510 1515
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1520 1525 1530
Gly Gly Ser Asp Tyr Lys Asp Asp Asp Asp Lys
1535 1540

<210> 720
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 720

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met

450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Ala
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Ala Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Ala Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Ala Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 721
<211> 1612
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 721

Met Ser Ser Glu Thr Gly Pro Val Ala Val Asp Pro Thr Leu Arg Arg
1 5 10 15
Arg Ile Glu Pro His Glu Phe Glu Val Phe Phe Asp Pro Arg Glu Leu
20 25 30
Arg Lys Glu Thr Cys Leu Leu Tyr Glu Ile Asn Trp Gly Gly Arg His
35 40 45
Ser Ile Trp Arg His Thr Ser Gln Asn Thr Asn Lys His Val Glu Val
50 55 60
Asn Phe Ile Glu Lys Phe Thr Thr Glu Arg Tyr Phe Cys Pro Asn Thr
65 70 75 80
Arg Cys Ser Ile Thr Trp Phe Leu Ser Trp Ser Pro Cys Gly Glu Cys
85 90 95
Ser Arg Ala Ile Thr Glu Phe Leu Ser Arg Tyr Pro His Val Thr Leu
100 105 110
Phe Ile Tyr Ile Ala Arg Leu Tyr His His Ala Asp Pro Arg Asn Arg
115 120 125
Gln Gly Leu Arg Asp Leu Ile Ser Ser Gly Val Thr Ile Gln Ile Met
130 135 140
Thr Glu Gln Glu Ser Gly Tyr Cys Trp Arg Asn Phe Val Asn Tyr Ser
145 150 155 160
Pro Ser Asn Glu Ala His Trp Pro Arg Tyr Pro His Leu Trp Val Arg
165 170 175
Leu Tyr Val Leu Glu Leu Tyr Cys Ile Ile Leu Gly Leu Pro Pro Cys
180 185 190
Leu Asn Ile Leu Arg Arg Lys Gln Pro Gln Leu Thr Phe Phe Thr Ile
195 200 205
Ala Leu Gln Ser Cys His Tyr Gln Arg Leu Pro Pro His Ile Leu Trp
210 215 220
Ala Thr Gly Leu Lys Ser Gly Ser Glu Thr Pro Gly Thr Ser Glu Ser
225 230 235 240
Ala Thr Pro Glu Ser Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly
245 250 255

Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro
260 265 270
Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys
275 280 285
Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu
290 295 300
Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys
305 310 315 320
Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys
325 330 335
Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu
340 345 350
Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp
355 360 365
Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys
370 375 380
Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu
385 390 395 400
Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly
405 410 415
Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu
420 425 430
Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser
435 440 445
Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg
450 455 460
Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly
465 470 475 480
Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe
485 490 495
Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys
500 505 510
Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp
515 520 525
Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile
530 535 540
Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro
545 550 555 560
Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu

565 570 575
Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys
580 585 590
Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp
595 600 605
Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu
610 615 620
Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu
625 630 635 640
Asp Leu Leu Arg Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His
645 650 655
Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp
660 665 670
Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu
675 680 685
Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser
690 695 700
Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp
705 710 715 720
Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile
725 730 735
Glu Arg Met Thr Ala Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu
740 745 750
Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu
755 760 765
Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu
770 775 780
Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn
785 790 795 800
Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile
805 810 815
Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn
820 825 830
Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys
835 840 845
Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val
850 855 860
Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu
865 870 875 880

Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys
885 890 895
Arg Arg Arg Tyr Thr Gly Trp Gly Ala Leu Ser Arg Lys Leu Ile Asn
900 905 910
Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys
915 920 925
Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Ala Leu Ile His Asp Asp
930 935 940
Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln
945 950 955 960
Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
965 970 975
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val
980 985 990
Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala
995 1000 1005
Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu
1010 1015 1020
Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
1025 1030 1035
Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu
1040 1045 1050
Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val
1055 1060 1065
Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp
1070 1075 1080
His Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn
1085 1090 1095
Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn
1100 1105 1110
Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr Trp Arg
1115 1120 1125
Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe Asp Asn
1130 1135 1140
Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys Ala
1145 1150 1155
Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Ala Ile Thr Lys
1160 1165 1170
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp

1175 1180 1185

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys

1190 1195 1200

Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys

1205 1210 1215

Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu

1220 1225 1230

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu

1235 1240 1245

Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg

1250 1255 1260

Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala

1265 1270 1275

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu

1280 1285 1290

Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu

1295 1300 1305

Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp

1310 1315 1320

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

1325 1330 1335

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

1340 1345 1350

Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys

1355 1360 1365

Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val

1370 1375 1380

Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser

1385 1390 1395

Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met

1400 1405 1410

Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala

1415 1420 1425

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro

1430 1435 1440

Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu

1445 1450 1455

Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro

1460 1465 1470

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys
1475 1480 1485
Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val
1490 1495 1500
Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1505 1510 1515
Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys
1520 1525 1530
Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu
1535 1540 1545
Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly
1550 1555 1560
Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys
1565 1570 1575
Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His
1580 1585 1590
Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln
1595 1600 1605
Leu Gly Gly Asp
1610

<210> 722
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 722
Gly Gly Gly Gly Ser
1 5

<210> 723
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 723
Glu Ala Ala Ala Lys
1 5

<210> 724

<211> 16
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 724

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

<210> 725
<211> 343
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 725

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1 5 10 15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
20 25 30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
35 40 45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
50 55 60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
65 70 75 80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
85 90 95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
100 105 110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
115 120 125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
130 135 140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
145 150 155 160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
165 170 175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
180 185 190

Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
195 200 205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
210 215 220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
225 230 235 240
Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
245 250 255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
260 265 270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
275 280 285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
290 295 300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
305 310 315 320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
325 330 335
Arg Leu Leu Glu Asp Gly Asp
340

<210> 726
<211> 423
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 726

Met Pro Gln Phe Gly Ile Leu Cys Lys Thr Pro Pro Lys Val Leu Val
1 5 10 15
Arg Gln Phe Val Glu Arg Phe Glu Arg Pro Ser Gly Glu Lys Ile Ala
20 25 30
Leu Cys Ala Ala Glu Leu Thr Tyr Leu Cys Trp Met Ile Thr His Asn
35 40 45
Gly Thr Ala Ile Lys Arg Ala Thr Phe Met Ser Tyr Asn Thr Ile Ile
50 55 60
Ser Asn Ser Leu Ser Phe Asp Ile Val Asn Lys Ser Leu Gln Phe Lys
65 70 75 80
Tyr Lys Thr Gln Lys Ala Thr Ile Leu Glu Ala Ser Leu Lys Lys Leu
85 90 95

Ile Pro Ala Trp Glu Phe Thr Ile Ile Pro Tyr Tyr Gly Gln Lys His
100 105 110
Gln Ser Asp Ile Thr Asp Ile Val Ser Ser Leu Gln Leu Gln Phe Glu
115 120 125
Ser Ser Glu Glu Ala Asp Lys Gly Asn Ser His Ser Lys Lys Met Leu
130 135 140
Lys Ala Leu Leu Ser Glu Gly Glu Ser Ile Trp Glu Ile Thr Glu Lys
145 150 155 160
Ile Leu Asn Ser Phe Glu Tyr Thr Ser Arg Phe Thr Lys Thr Lys Thr
165 170 175
Leu Tyr Gln Phe Leu Phe Leu Ala Thr Phe Ile Asn Cys Gly Arg Phe
180 185 190
Ser Asp Ile Lys Asn Val Asp Pro Lys Ser Phe Lys Leu Val Gln Asn
195 200 205
Lys Tyr Leu Gly Val Ile Ile Gln Cys Leu Val Thr Glu Thr Lys Thr
210 215 220
Ser Val Ser Arg His Ile Tyr Phe Phe Ser Ala Arg Gly Arg Ile Asp
225 230 235 240
Pro Leu Val Tyr Leu Asp Glu Phe Leu Arg Asn Ser Glu Pro Val Leu
245 250 255
Lys Arg Val Asn Arg Thr Gly Asn Ser Ser Ser Asn Lys Gln Glu Tyr
260 265 270
Gln Leu Leu Lys Asp Asn Leu Val Arg Ser Tyr Asn Lys Ala Leu Lys
275 280 285
Lys Asn Ala Pro Tyr Ser Ile Phe Ala Ile Lys Asn Gly Pro Lys Ser
290 295 300
His Ile Gly Arg His Leu Met Thr Ser Phe Leu Ser Met Lys Gly Leu
305 310 315 320
Thr Glu Leu Thr Asn Val Val Gly Asn Trp Ser Asp Lys Arg Ala Ser
325 330 335
Ala Val Ala Arg Thr Thr Tyr Thr His Gln Ile Thr Ala Ile Pro Asp
340 345 350
His Tyr Phe Ala Leu Val Ser Arg Tyr Tyr Ala Tyr Asp Pro Ile Ser
355 360 365
Lys Glu Met Ile Ala Leu Lys Asp Glu Thr Asn Pro Ile Glu Glu Trp
370 375 380
Gln His Ile Glu Gln Leu Lys Gly Ser Ala Glu Gly Ser Ile Arg Tyr
385 390 395 400
Pro Ala Trp Asn Gly Ile Ile Ser Gln Glu Val Leu Asp Tyr Leu Ser

405 410 415

Ser Tyr Ile Asn Arg Arg Ile
420

<210> 727
<211> 144
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 727

Met Arg Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
1 5 10 15
Asp Ile Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg
20 25 30
Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Ser Asp Arg Lys Gly Leu
35 40 45
Asp Leu Leu Arg Met Lys Val Glu Glu Gly Asp Val Ile Leu Val Lys
50 55 60
Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Ile Gln Leu Ile
65 70 75 80
Lys Glu Phe Asp Ala Gln Gly Val Ser Ile Arg Phe Ile Asp Asp Gly
85 90 95
Ile Ser Thr Asp Gly Glu Met Gly Lys Met Val Val Thr Ile Leu Ser
100 105 110
Ala Val Ala Gln Ala Glu Arg Gln Arg Ile Leu Glu Arg Thr Asn Glu
115 120 125
Gly Arg Gln Glu Ala Met Ala Lys Gly Val Val Phe Gly Arg Lys Arg
130 135 140

<210> 728
<211> 144
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 728

Met Arg Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
1 5 10 15
Asp Ile Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg
20 25 30

Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Ser Asp Arg Lys Gly Leu
35 40 45
Asp Leu Leu Arg Met Lys Val Glu Glu Gly Asp Val Ile Leu Val Lys
50 55 60
Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Ile Gln Leu Ile
65 70 75 80
Lys Glu Phe Asp Ala Gln Gly Val Ser Ile Arg Phe Ile Asp Asp Gly
85 90 95
Ile Ser Thr Asp Gly Glu Met Gly Lys Met Val Val Thr Ile Leu Ser
100 105 110
Ala Val Ala Gln Ala Glu Arg Gln Arg Ile Leu Glu Arg Thr Asn Glu
115 120 125
Gly Arg Gln Glu Ala Met Ala Lys Gly Val Val Phe Gly Arg Lys Arg
130 135 140

<210> 729
<211> 144
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 729

Met Arg Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
1 5 10 15
Asp Ile Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg
20 25 30
Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Ser Asp Arg Lys Gly Leu
35 40 45
Asp Leu Leu Arg Met Lys Val Glu Glu Gly Asp Val Ile Leu Val Lys
50 55 60
Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Ile Gln Leu Ile
65 70 75 80
Lys Glu Phe Asp Ala Gln Gly Val Ser Ile Arg Phe Ile Asp Asp Gly
85 90 95
Ile Ser Thr Asp Gly Tyr Met Gly Lys Met Val Val Thr Ile Leu Ser
100 105 110
Ala Val Ala Gln Ala Glu Arg Gln Arg Ile Leu Gln Arg Thr Asn Glu
115 120 125
Gly Arg Gln Glu Ala Met Ala Lys Gly Val Val Phe Gly Arg Lys Arg
130 135 140

<210> 730
<211> 147
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 730

Met Ala Lys Ile Gly Tyr Ala Arg Val Ser Ser Lys Glu Gln Asn Leu
1 5 10 15
Asp Arg Gln Leu Gln Ala Leu Gln Gly Val Ser Lys Val Phe Ser Asp
20 25 30
Lys Leu Ser Gly Gln Ser Val Glu Arg Pro Gln Leu Gln Ala Met Leu
35 40 45
Asn Tyr Ile Arg Glu Gly Asp Ile Val Val Val Thr Glu Leu Asp Arg
50 55 60
Leu Gly Arg Asn Asn Lys Glu Leu Thr Glu Leu Met Asn Ala Ile Gln
65 70 75 80
Gln Lys Gly Ala Thr Leu Glu Val Leu Asp Leu Pro Ser Met Asn Gly
85 90 95
Ile Glu Asp Glu Asn Leu Arg Arg Leu Ile Asn Asn Leu Val Ile Glu
100 105 110
Leu Tyr Lys Tyr Gln Ala Glu Ser Glu Arg Lys Arg Ile Lys Glu Arg
115 120 125
Gln Ala Gln Gly Ile Glu Ile Ala Lys Ser Lys Gly Lys Phe Lys Gly
130 135 140
Arg Gln His
145

<210> 731
<211> 147
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 731

Met Ala Lys Ile Gly Tyr Ala Arg Val Ser Ser Lys Glu Gln Asn Leu
1 5 10 15
Asp Arg Gln Leu Gln Ala Leu Gln Gly Val Ser Lys Val Phe Ser Asp
20 25 30
Lys Leu Ser Gly Gln Ser Val Glu Arg Pro Gln Leu Gln Ala Met Leu

35 40 45

Asn Tyr Ile Arg Glu Gly Asp Ile Val Val Val Thr Glu Leu Asp Arg
50 55 60

Leu Gly Arg Asn Asn Lys Glu Leu Thr Glu Leu Met Asn Ala Ile Gln
65 70 75 80

Gln Lys Gly Ala Thr Leu Glu Val Leu Asp Leu Pro Ser Met Asp Gly
85 90 95

Ile Glu Asp Glu Asn Leu Arg Arg Leu Ile Asn Asn Leu Val Ile Glu
100 105 110

Leu Tyr Lys Tyr Gln Ala Glu Ser Glu Arg Lys Arg Ile Lys Glu Arg
115 120 125

Gln Ala Gln Gly Ile Glu Ile Ala Lys Ser Lys Gly Lys Phe Lys Gly
130 135 140

Arg Gln His
145

<210> 732
<211> 150
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 732

Met Ile Ile Gly Tyr Ala Arg Val Ser Ser Leu Asp Gln Asn Leu Glu
1 5 10 15

Arg Gln Leu Glu Asn Leu Lys Thr Phe Gly Ala Glu Lys Ile Phe Thr
20 25 30

Glu Lys Gln Ser Gly Lys Ser Ile Glu Asn Arg Pro Ile Leu Gln Lys
35 40 45

Ala Leu Asn Phe Val Arg Met Gly Asp Arg Phe Ile Val Glu Ser Ile
50 55 60

Asp Arg Leu Gly Arg Asn Tyr Asn Glu Val Ile His Thr Val Asn Tyr
65 70 75 80

Leu Lys Asp Lys Glu Val Gln Leu Met Ile Thr Ser Leu Pro Met Met
85 90 95

Asn Glu Val Ile Gly Asn Pro Leu Leu Asp Lys Phe Met Lys Asp Leu
100 105 110

Ile Ile Gln Ile Leu Ala Met Val Ser Glu Gln Glu Arg Asn Glu Ser
115 120 125

Lys Arg Arg Gln Ala Gln Gly Ile Gln Val Ala Lys Glu Lys Gly Val

130 135 140
Tyr Lys Gly Arg Pro Leu
145 150
<210> 733
<211> 150
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 733
Met Ile Ile Gly Tyr Ala Arg Val Ser Ser Leu Asp Gln Asn Leu Glu
1 5 10 15
Arg Gln Leu Glu Asn Leu Lys Thr Phe Gly Ala Glu Lys Ile Phe Thr
20 25 30
Glu Lys Gln Ser Gly Lys Ser Ile Glu Asn Arg Pro Ile Leu Gln Lys
35 40 45
Ala Leu Asn Phe Val Arg Met Gly Asp Arg Phe Ile Val Glu Ser Ile
50 55 60
Asp Arg Leu Gly Arg Asn Tyr Asn Glu Val Ile His Thr Val Asn Tyr
65 70 75 80
Leu Lys Asp Lys Glu Val Arg Leu Met Ile Thr Ser Leu Pro Met Met
85 90 95
Asn Glu Val Ile Gly Asn Pro Leu Leu Asp Lys Phe Met Lys Asp Leu
100 105 110
Ile Ile Arg Ile Leu Ala Met Val Ser Glu Gln Glu Arg Asn Glu Ser
115 120 125
Lys Arg Arg Gln Ala Gln Gly Ile Gln Val Ala Lys Glu Lys Gly Val
130 135 140
Tyr Lys Gly Arg Pro Leu
145 150
<210> 734
<211> 144
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 734
Met Arg Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
1 5 10 15

Asp Leu Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg
20 25 30
Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Thr Asp Arg Glu Gly Leu
35 40 45
Asp Leu Leu Arg Met Lys Val Lys Glu Gly Asp Val Ile Leu Val Lys
50 55 60
Lys Leu Asp Arg Leu Gly Arg Asp Thr Ala Asp Met Leu Gln Leu Ile
65 70 75 80
Lys Glu Phe Asp Ala Gln Gly Val Ala Val Arg Phe Ile Asp Asp Gly
85 90 95
Ile Ser Thr Asp Gly Asp Met Gly Gln Met Val Val Thr Ile Leu Ser
100 105 110
Ala Val Ala Gln Ala Glu Arg Arg Arg Ile Leu Glu Arg Thr Asn Glu
115 120 125
Gly Arg Gln Glu Ala Lys Leu Lys Gly Ile Lys Phe Gly Arg Arg Arg
130 135 140

<210> 735
<211> 144
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 735

Met Arg Leu Phe Gly Tyr Ala Arg Val Ser Thr Ser Gln Gln Ser Leu
1 5 10 15
Asp Leu Gln Val Arg Ala Leu Lys Asp Ala Gly Val Lys Ala Asn Arg
20 25 30
Ile Phe Thr Asp Lys Ala Ser Gly Ser Ser Thr Asp Arg Glu Gly Leu
35 40 45
Asp Leu Leu Arg Met Lys Val Lys Glu Gly Asp Val Ile Leu Val Lys
50 55 60
Lys Leu Asp Arg Leu Ser Arg Asp Thr Ala Asp Met Leu Gln Leu Ile
65 70 75 80
Lys Glu Phe Asp Ala Gln Gly Val Ala Val Arg Phe Ile Asp Asp Gly
85 90 95
Ile Ser Thr Asp Gly Tyr Met Gly Gln Met Val Val Thr Ile Leu Ser
100 105 110
Ala Val Ala Gln Ala Glu Arg Arg Arg Ile Leu Gln Arg Thr Asn Glu
115 120 125

Gly Arg Gln Glu Ala Lys Leu Lys Gly Ile Lys Phe Gly Arg Arg Arg
130 135 140

<210> 736
<211> 142
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 736

Met Ala Thr Ile Gly Tyr Ile Arg Val Ser Thr Ile Asp Gln Asn Ile
1 5 10 15
Asp Leu Gln Arg Asn Ala Leu Thr Ser Ala Asn Cys Asp Arg Ile Phe
20 25 30
Glu Asp Arg Ile Ser Gly Lys Ile Ala Asn Arg Pro Gly Leu Lys Arg
35 40 45
Ala Leu Lys Tyr Val Asn Lys Gly Asp Thr Leu Val Val Trp Lys Leu
50 55 60
Asp Arg Leu Gly Arg Ser Val Lys Asn Leu Val Ala Leu Ile Ser Glu
65 70 75 80
Leu His Glu Arg Gly Ala His Phe His Ser Leu Thr Asp Ser Ile Asp
85 90 95
Thr Ser Ser Ala Met Gly Arg Phe Phe Phe His Val Met Ser Ala Leu
100 105 110
Ala Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Leu Ala Gly Leu
115 120 125
Ala Ala Ala Arg Ala Gln Gly Arg Leu Gly Gly Arg Pro Val
130 135 140

<210> 737
<211> 142
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 737

Met Ala Thr Ile Gly Tyr Ile Arg Val Ser Thr Ile Asp Gln Asn Ile
1 5 10 15
Asp Leu Gln Arg Asn Ala Leu Thr Ser Ala Asn Cys Asp Arg Ile Phe
20 25 30
Glu Asp Arg Ile Ser Gly Lys Ile Ala Asn Arg Pro Gly Leu Lys Arg

35 40 45

Ala Leu Lys Tyr Val Asn Lys Gly Asp Thr Leu Val Val Trp Lys Leu
50 55 60

Asp Arg Leu Gly Arg Ser Val Lys Asn Leu Val Ala Leu Ile Ser Glu
65 70 75 80

Leu His Glu Arg Gly Ala His Phe His Ser Leu Thr Asp Ser Ile Asp
85 90 95

Thr Ser Ser Ala Met Gly Arg Phe Phe Phe Tyr Val Met Ser Ala Leu
100 105 110

Ala Glu Met Glu Arg Glu Leu Ile Val Glu Arg Thr Leu Ala Gly Leu
115 120 125

Ala Ala Ala Arg Ala Gln Gly Arg Leu Gly Gly Arg Pro Val
130 135 140

<210> 738

<211> 608

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 738

Met Asp Thr Tyr Ala Gly Ala Tyr Asp Arg Gln Ser Arg Glu Arg Glu
1 5 10 15

Asn Ser Ser Ala Ala Ser Pro Ala Thr Gln Arg Ser Ala Asn Glu Asp
20 25 30

Lys Ala Ala Asp Leu Gln Arg Glu Val Glu Arg Asp Gly Gly Arg Phe
35 40 45

Arg Phe Val Gly His Phe Ser Glu Ala Pro Gly Thr Ser Ala Phe Gly
50 55 60

Thr Ala Glu Arg Pro Glu Phe Glu Arg Ile Leu Asn Glu Cys Arg Ala
65 70 75 80

Gly Arg Leu Asn Met Ile Ile Val Tyr Asp Val Ser Arg Phe Ser Arg
85 90 95

Leu Lys Val Met Asp Ala Ile Pro Ile Val Ser Glu Leu Leu Ala Leu
100 105 110

Gly Val Thr Ile Val Ser Thr Gln Glu Gly Val Phe Arg Gln Gly Asn
115 120 125

Val Met Asp Leu Ile His Leu Ile Met Arg Leu Asp Ala Ser His Lys
130 135 140

Glu Ser Ser Leu Lys Ser Ala Lys Ile Leu Asp Thr Lys Asn Leu Gln

145 150 155 160
Arg Glu Leu Gly Gly Tyr Val Gly Gly Lys Ala Pro Tyr Gly Phe Glu
165 170 175
Leu Val Ser Glu Thr Lys Glu Ile Thr Arg Asn Gly Arg Met Val Asn
180 185 190
Val Val Ile Asn Lys Leu Ala His Ser Thr Thr Pro Leu Thr Gly Pro
195 200 205
Phe Glu Phe Glu Pro Asp Val Ile Arg Trp Trp Trp Arg Glu Ile Lys
210 215 220
Thr His Lys His Leu Pro Phe Lys Pro Gly Ser Gln Ala Ala Ile His
225 230 235 240
Pro Gly Ser Ile Thr Gly Leu Cys Lys Arg Met Asp Ala Asp Ala Val
245 250 255
Pro Thr Arg Gly Glu Thr Ile Gly Lys Lys Thr Ala Ser Ser Ala Trp
260 265 270
Asp Pro Ala Thr Val Met Arg Ile Leu Arg Asp Pro Arg Ile Ala Gly
275 280 285
Phe Ala Ala Glu Val Ile Tyr Lys Lys Lys Pro Asp Gly Thr Pro Thr
290 295 300
Thr Lys Ile Glu Gly Tyr Arg Ile Gln Arg Asp Pro Ile Thr Leu Arg
305 310 315 320
Pro Val Glu Leu Asp Cys Gly Pro Ile Ile Glu Pro Ala Glu Trp Tyr
325 330 335
Glu Leu Gln Ala Trp Leu Asp Gly Arg Gly Arg Gly Lys Gly Leu Ser
340 345 350
Arg Gly Gln Ala Ile Leu Ser Ala Met Asp Lys Leu Tyr Cys Glu Cys
355 360 365
Gly Ala Val Met Thr Ser Lys Arg Gly Glu Glu Ser Ile Lys Asp Ser
370 375 380
Tyr Arg Cys Arg Arg Arg Lys Val Val Asp Pro Ser Ala Pro Gly Gln
385 390 395 400
His Glu Gly Thr Cys Asn Val Ser Met Ala Ala Leu Asp Lys Phe Val
405 410 415
Ala Glu Arg Ile Phe Asn Lys Ile Arg His Ala Glu Gly Asp Glu Glu
420 425 430
Thr Leu Ala Leu Leu Trp Glu Ala Ala Arg Arg Phe Gly Lys Leu Thr
435 440 445
Glu Ala Pro Glu Lys Ser Gly Glu Arg Ala Asn Leu Val Ala Glu Arg
450 455 460

Ala Asp Ala Leu Asn Ala Leu Glu Glu Leu Tyr Glu Asp Arg Ala Ala
465 470 475 480
Gly Ala Tyr Asp Gly Pro Val Gly Arg Lys His Phe Arg Lys Gln Gln
485 490 495
Ala Ala Leu Thr Leu Arg Gln Gln Gly Ala Glu Glu Arg Leu Ala Glu
500 505 510
Leu Glu Ala Ala Glu Ala Pro Lys Leu Pro Leu Asp Gln Trp Phe Pro
515 520 525
Glu Asp Ala Asp Ala Asp Pro Thr Gly Pro Lys Ser Trp Trp Gly Arg
530 535 540
Ala Ser Val Asp Asp Lys Arg Val Phe Val Gly Leu Phe Val Asp Lys
545 550 555 560
Ile Val Val Thr Lys Ser Thr Thr Gly Arg Gly Gln Gly Thr Pro Ile
565 570 575
Glu Lys Arg Ala Ser Ile Thr Trp Ala Lys Pro Pro Thr Asp Asp Asp
580 585 590
Glu Asp Asp Ala Gln Asp Gly Thr Glu Asp Val Ala Ala Thr Gly Ala
595 600 605

<210> 739
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 739
ataacttcgt atagcataca ttatacgaag ttat 34
<210> 740
<211> 34
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 740
gaagttccta ttctctagaa agtataggaa cttc 34
<210> 741
<211> 4
<212> PRT
<213> 人工序列
<220>

<223> 合成多肽
<400> 741
Asn Gly Ala Asn
1
<210> 742
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 742
Asn Gly Asn Gly
1
<210> 743
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 743
Asn Gly Ala Gly
1
<210> 744
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 744
Asn Gly Cys Gly
1
<210> 745
<211> 6
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 745
Asn Asn Gly Arg Arg Thr

1 5

<210> 746

<211> 5

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 746

Asn Gly Arg Arg Asn

1 5

<210> 747

<211> 6

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 747

Asn Asn Asn Arg Arg Thr

1 5

<210> 748

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 748

Asn Asn Asn Gly Ala Thr Thr

1 5

<210> 749

<211> 7

<212> PRT

<213> 人工序列

<220>

<223> 合成多肽

<400> 749

Asn Asn Ala Gly Ala Ala Trp

1 5

<210> 750

<211> 5

<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 750
Asn Ala Ala Ala Cys
1 5
<210> 751
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 751
Thr Thr Thr Asn
1
<210> 752
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 752
Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240
Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro

435 440 445

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

450 455 460

Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val

465 470 475 480

Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn

485 490 495

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

500 505 510

Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr

515 520 525

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

530 535 540

Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560

Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser

565 570 575

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

580 585 590

Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn

595 600 605

Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu

610 615 620

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

625 630 635 640

Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr

645 650 655

Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

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

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr

1055 1060 1065

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg

1070 1075 1080

Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu

1085 1090 1095

Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg

1100 1105 1110

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys

1115 1120 1125

Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu

1130 1135 1140

Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser

1145 1150 1155

Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170

Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu

1175 1180 1185

Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe

1190 1195 1200

Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu

1205 1210 1215

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn

1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His

1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr

1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile

1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe

1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly

1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 753
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 753

Asp Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly
35 40 45
Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60
Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80
Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95
Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110
Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125
Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140
Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160
Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175
Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190
Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205
Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220
Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255
Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270
Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285
Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300
Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320
Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335
Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350
Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val

545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys
660 665 670
Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala
675 680 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe Lys
690 695 700
Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His
705 710 715 720
Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile
725 730 735
Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg
740 745 750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
755 760 765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
770 775 780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785 790 795 800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
805 810 815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
820 825 830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp
835 840 845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865 870 875 880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
885 890 895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
900 905 910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
915 920 925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
930 935 940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945 950 955 960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
965 970 975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe

1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 754
<211> 1367
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 754

Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val Gly
1 5 10 15
Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys
20 25 30
Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly

35 40 45

Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys
50 55 60

Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr
65 70 75 80

Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe
85 90 95

Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His
100 105 110

Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His
115 120 125

Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser
130 135 140

Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met
145 150 155 160

Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp
165 170 175

Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn
180 185 190

Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys
195 200 205

Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu
210 215 220

Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu
225 230 235 240

Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp
245 250 255

Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp
260 265 270

Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu
275 280 285

Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile
290 295 300

Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met
305 310 315 320

Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala
325 330 335

Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp
340 345 350

Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln
355 360 365
Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly
370 375 380
Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys
385 390 395 400
Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly
405 410 415
Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu
420 425 430
Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro
435 440 445
Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met
450 455 460
Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val
465 470 475 480
Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn
485 490 495
Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu
500 505 510
Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr
515 520 525
Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys
530 535 540
Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val
545 550 555 560
Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser
565 570 575
Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr
580 585 590
Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn
595 600 605
Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu
610 615 620
Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His
625 630 635 640
Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
645 650 655
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp Lys

660 665 670

Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe Ala

675 680 685

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

690 695 700

Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu His

705 710 715 720

Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly Ile

725 730 735

Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly Arg

740 745 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

755 760 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu

770 775 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val

785 790 795 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln

805 810 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu

820 825 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys Asp

835 840 845

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

850 855 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn

865 870 875 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

885 890 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys

900 905 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys

915 920 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu

930 935 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys

945 950 955 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu

965 970 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
980 985 990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val
995 1000 1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
1025 1030 1035
Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
1040 1045 1050
Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
1055 1060 1065
Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
1070 1075 1080
Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
1085 1090 1095
Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
1100 1105 1110
Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
1115 1120 1125
Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
1130 1135 1140
Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
1145 1150 1155
Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
1160 1165 1170
Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
1175 1180 1185
Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
1190 1195 1200
Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
1205 1210 1215
Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230
Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245
Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

1265 1270 1275
Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290
Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305
Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 755
<211> 345
<212> PRT
<213> 冰岛硫化叶菌
<400> 755

Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln
1 5 10 15
Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu
20 25 30
Ile Asp Asp Glu Glu Leu Arg Asn Val Leu Asn Leu Ala Tyr Lys Ile
35 40 45
Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys
50 55 60
Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65 70 75 80
Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys
85 90 95
Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
100 105 110
Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys
115 120 125
Pro Glu Phe Tyr Glu Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
130 135 140
Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Ile Ala Ala
145 150 155 160
Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly

165 170 175
Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
180 185 190
Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
195 200 205
Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
210 215 220
Asn Pro Asn Val Ser Val Val Arg Ile Tyr Thr Ile Ser Asp Ala Val
225 230 235 240
Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr
245 250 255
Lys Leu Leu Glu Lys Arg Tyr Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270
Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
275 280 285
Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
290 295 300
Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305 310 315 320
Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
325 330 335
Asn Gly Glu Leu Ile Arg Gly Glu Gly
340 345

<210> 756
<211> 345
<212> PRT
<213> 冰岛硫化叶菌
<400> 756
Met Glu Val Pro Leu Tyr Asn Ile Phe Gly Asp Asn Tyr Ile Ile Gln
1 5 10 15
Val Ala Thr Glu Ala Glu Asn Ser Thr Ile Tyr Asn Asn Lys Val Glu
20 25 30
Ile Asp Asp Glu Glu Leu Arg Asn Val Leu Asn Leu Ala Tyr Lys Ile
35 40 45
Ala Lys Asn Asn Glu Asp Ala Ala Ala Glu Arg Arg Gly Lys Ala Lys
50 55 60
Lys Lys Lys Gly Glu Glu Gly Glu Thr Thr Thr Ser Asn Ile Ile Leu
65 70 75 80
Pro Leu Ser Gly Asn Asp Lys Asn Pro Trp Thr Glu Thr Leu Lys Cys

85 90 95
Tyr Asn Phe Pro Thr Thr Val Ala Leu Ser Glu Val Phe Lys Asn Phe
100 105 110
Ser Gln Val Lys Glu Cys Glu Glu Val Ser Ala Pro Ser Phe Val Lys
115 120 125
Pro Glu Phe Tyr Lys Phe Gly Arg Ser Pro Gly Met Val Glu Arg Thr
130 135 140
Arg Arg Val Lys Leu Glu Val Glu Pro His Tyr Leu Ile Met Ala Ala
145 150 155 160
Ala Gly Trp Val Leu Thr Arg Leu Gly Lys Ala Lys Val Ser Glu Gly
165 170 175
Asp Tyr Val Gly Val Asn Val Phe Thr Pro Thr Arg Gly Ile Leu Tyr
180 185 190
Ser Leu Ile Gln Asn Val Asn Gly Ile Val Pro Gly Ile Lys Pro Glu
195 200 205
Thr Ala Phe Gly Leu Trp Ile Ala Arg Lys Val Val Ser Ser Val Thr
210 215 220
Asn Pro Asn Val Ser Val Val Ser Ile Tyr Thr Ile Ser Asp Ala Val
225 230 235 240
Gly Gln Asn Pro Thr Thr Ile Asn Gly Gly Phe Ser Ile Asp Leu Thr
245 250 255
Lys Leu Leu Glu Lys Arg Asp Leu Leu Ser Glu Arg Leu Glu Ala Ile
260 265 270
Ala Arg Asn Ala Leu Ser Ile Ser Ser Asn Met Arg Glu Arg Tyr Ile
275 280 285
Val Leu Ala Asn Tyr Ile Tyr Glu Tyr Leu Thr Gly Ser Lys Arg Leu
290 295 300
Glu Asp Leu Leu Tyr Phe Ala Asn Arg Asp Leu Ile Met Asn Leu Asn
305 310 315 320
Ser Asp Asp Gly Lys Val Arg Asp Leu Lys Leu Ile Ser Ala Tyr Val
325 330 335
Asn Gly Glu Leu Ile Arg Gly Glu Gly
340 345

<210> 757
<211> 1210
<212> PRT
<213> Parcubacteria
<400> 757
Met Ser Lys Arg His Pro Arg Ile Ser Gly Val Lys Gly Tyr Arg Leu

1 5 10 15
His Ala Gln Arg Leu Glu Tyr Thr Gly Lys Ser Gly Ala Met Arg Thr
20 25 30
Ile Lys Tyr Pro Leu Tyr Ser Ser Pro Ser Gly Gly Arg Thr Val Pro
35 40 45
Arg Glu Ile Val Ser Ala Ile Asn Asp Asp Tyr Val Gly Leu Tyr Gly
50 55 60
Leu Ser Asn Phe Asp Asp Leu Tyr Asn Ala Glu Lys Arg Asn Glu Glu
65 70 75 80
Lys Val Tyr Ser Val Leu Asp Phe Trp Tyr Asp Cys Val Gln Tyr Gly
85 90 95
Ala Val Phe Ser Tyr Thr Ala Pro Gly Leu Leu Lys Asn Val Ala Glu
100 105 110
Val Arg Gly Gly Ser Tyr Glu Leu Thr Lys Thr Leu Lys Gly Ser His
115 120 125
Leu Tyr Asp Glu Leu Gln Ile Asp Lys Val Ile Lys Phe Leu Asn Lys
130 135 140
Lys Glu Ile Ser Arg Ala Asn Gly Ser Leu Asp Lys Leu Lys Lys Asp
145 150 155 160
Ile Ile Asp Cys Phe Lys Ala Glu Tyr Arg Glu Arg His Lys Asp Gln
165 170 175
Cys Asn Lys Leu Ala Asp Asp Ile Lys Asn Ala Lys Lys Asp Ala Gly
180 185 190
Ala Ser Leu Gly Glu Arg Gln Lys Lys Leu Phe Arg Asp Phe Phe Gly
195 200 205
Ile Ser Glu Gln Ser Glu Asn Asp Lys Pro Ser Phe Thr Asn Pro Leu
210 215 220
Asn Leu Thr Cys Cys Leu Leu Pro Phe Asp Thr Val Asn Asn Asn Arg
225 230 235 240
Asn Arg Gly Glu Val Leu Phe Asn Lys Leu Lys Glu Tyr Ala Gln Lys
245 250 255
Leu Asp Lys Asn Glu Gly Ser Leu Glu Met Trp Glu Tyr Ile Gly Ile
260 265 270
Gly Asn Ser Gly Thr Ala Phe Ser Asn Phe Leu Gly Glu Gly Phe Leu
275 280 285
Gly Arg Leu Arg Glu Asn Lys Ile Thr Glu Leu Lys Lys Ala Met Met
290 295 300
Asp Ile Thr Asp Ala Trp Arg Gly Gln Glu Gln Glu Glu Glu Leu Glu
305 310 315 320

Lys Arg Leu Arg Ile Leu Ala Ala Leu Thr Ile Lys Leu Arg Glu Pro
325 330 335
Lys Phe Asp Asn His Trp Gly Gly Tyr Arg Ser Asp Ile Asn Gly Lys
340 345 350
Leu Ser Ser Trp Leu Gln Asn Tyr Ile Asn Gln Thr Val Lys Ile Lys
355 360 365
Glu Asp Leu Lys Gly His Lys Lys Asp Leu Lys Lys Ala Lys Glu Met
370 375 380
Ile Asn Arg Phe Gly Glu Ser Asp Thr Lys Glu Glu Ala Val Val Ser
385 390 395 400
Ser Leu Leu Glu Ser Ile Glu Lys Ile Val Pro Asp Asp Ser Ala Asp
405 410 415
Asp Glu Lys Pro Asp Ile Pro Ala Ile Ala Ile Tyr Arg Arg Phe Leu
420 425 430
Ser Asp Gly Arg Leu Thr Leu Asn Arg Phe Val Gln Arg Glu Asp Val
435 440 445
Gln Glu Ala Leu Ile Lys Glu Arg Leu Glu Ala Glu Lys Lys Lys Lys
450 455 460
Pro Lys Lys Arg Lys Lys Lys Ser Asp Ala Glu Asp Glu Lys Glu Thr
465 470 475 480
Ile Asp Phe Lys Glu Leu Phe Pro His Leu Ala Lys Pro Leu Lys Leu
485 490 495
Val Pro Asn Phe Tyr Gly Asp Ser Lys Arg Glu Leu Tyr Lys Lys Tyr
500 505 510
Lys Asn Ala Ala Ile Tyr Thr Asp Ala Leu Trp Lys Ala Val Glu Lys
515 520 525
Ile Tyr Lys Ser Ala Phe Ser Ser Ser Leu Lys Asn Ser Phe Phe Asp
530 535 540
Thr Asp Phe Asp Lys Asp Phe Phe Ile Lys Arg Leu Gln Lys Ile Phe
545 550 555 560
Ser Val Tyr Arg Arg Phe Asn Thr Asp Lys Trp Lys Pro Ile Val Lys
565 570 575
Asn Ser Phe Ala Pro Tyr Cys Asp Ile Val Ser Leu Ala Glu Asn Glu
580 585 590
Val Leu Tyr Lys Pro Lys Gln Ser Arg Ser Arg Lys Ser Ala Ala Ile
595 600 605
Asp Lys Asn Arg Val Arg Leu Pro Ser Thr Glu Asn Ile Ala Lys Ala
610 615 620
Gly Ile Ala Leu Ala Arg Glu Leu Ser Val Ala Gly Phe Asp Trp Lys

625 630 635 640
Asp Leu Leu Lys Lys Glu Glu His Glu Glu Tyr Ile Asp Leu Ile Glu
645 650 655
Leu His Lys Thr Ala Leu Ala Leu Leu Leu Ala Val Thr Glu Thr Gln
660 665 670
Leu Asp Ile Ser Ala Leu Asp Phe Val Glu Asn Gly Thr Val Lys Asp
675 680 685
Phe Met Lys Thr Arg Asp Gly Asn Leu Val Leu Glu Gly Arg Phe Leu
690 695 700
Glu Met Phe Ser Gln Ser Ile Val Phe Ser Glu Leu Arg Gly Leu Ala
705 710 715 720
Gly Leu Met Ser Arg Lys Glu Phe Ile Thr Arg Ser Ala Ile Gln Thr
725 730 735
Met Asn Gly Lys Gln Ala Glu Leu Leu Tyr Ile Pro His Glu Phe Gln
740 745 750
Ser Ala Lys Ile Thr Thr Pro Lys Glu Met Ser Arg Ala Phe Leu Asp
755 760 765
Leu Ala Pro Ala Glu Phe Ala Thr Ser Leu Glu Pro Glu Ser Leu Ser
770 775 780
Glu Lys Ser Leu Leu Lys Leu Lys Gln Met Arg Tyr Tyr Pro His Tyr
785 790 795 800
Phe Gly Tyr Glu Leu Thr Arg Thr Gly Gln Gly Ile Asp Gly Gly Val
805 810 815
Ala Glu Asn Ala Leu Arg Leu Glu Lys Ser Pro Val Lys Lys Arg Glu
820 825 830
Ile Lys Cys Lys Gln Tyr Lys Thr Leu Gly Arg Gly Gln Asn Lys Ile
835 840 845
Val Leu Tyr Val Arg Ser Ser Tyr Tyr Gln Thr Gln Phe Leu Glu Trp
850 855 860
Phe Leu His Arg Pro Lys Asn Val Gln Thr Asp Val Ala Val Ser Gly
865 870 875 880
Ser Phe Leu Ile Asp Glu Lys Lys Val Lys Thr Arg Trp Asn Tyr Asp
885 890 895
Ala Leu Thr Val Ala Leu Glu Pro Val Ser Gly Ser Glu Arg Val Phe
900 905 910
Val Ser Gln Pro Phe Thr Ile Phe Pro Glu Lys Ser Ala Glu Glu Glu
915 920 925
Gly Gln Arg Tyr Leu Gly Ile Asp Ile Gly Glu Tyr Gly Ile Ala Tyr
930 935 940

Thr Ala Leu Glu Ile Thr Gly Asp Ser Ala Lys Ile Leu Asp Gln Asn
945 950 955 960
Phe Ile Ser Asp Pro Gln Leu Lys Thr Leu Arg Glu Glu Val Lys Gly
965 970 975
Leu Lys Leu Asp Gln Arg Arg Gly Thr Phe Ala Met Pro Ser Thr Lys
980 985 990
Ile Ala Arg Ile Arg Glu Ser Leu Val His Ser Leu Arg Asn Arg Ile
995 1000 1005
His His Leu Ala Leu Lys His Lys Ala Lys Ile Val Tyr Glu Leu
1010 1015 1020
Glu Val Ser Arg Phe Glu Glu Gly Lys Gln Lys Ile Lys Lys Val
1025 1030 1035
Tyr Ala Thr Leu Lys Lys Ala Asp Val Tyr Ser Glu Ile Asp Ala
1040 1045 1050
Asp Lys Asn Leu Gln Thr Thr Val Trp Gly Lys Leu Ala Val Ala
1055 1060 1065
Ser Glu Ile Ser Ala Ser Tyr Thr Ser Gln Phe Cys Gly Ala Cys
1070 1075 1080
Lys Lys Leu Trp Arg Ala Glu Met Gln Val Asp Glu Thr Ile Thr
1085 1090 1095
Thr Gln Glu Leu Ile Gly Thr Val Arg Val Ile Lys Gly Gly Thr
1100 1105 1110
Leu Ile Asp Ala Ile Lys Asp Phe Met Arg Pro Pro Ile Phe Asp
1115 1120 1125
Glu Asn Asp Thr Pro Phe Pro Lys Tyr Arg Asp Phe Cys Asp Lys
1130 1135 1140
His His Ile Ser Lys Lys Met Arg Gly Asn Ser Cys Leu Phe Ile
1145 1150 1155
Cys Pro Phe Cys Arg Ala Asn Ala Asp Ala Asp Ile Gln Ala Ser
1160 1165 1170
Gln Thr Ile Ala Leu Leu Arg Tyr Val Lys Glu Glu Lys Lys Val
1175 1180 1185
Glu Asp Tyr Phe Glu Arg Phe Arg Lys Leu Lys Asn Ile Lys Val
1190 1195 1200
Leu Gly Gln Met Lys Lys Ile
1205 1210

<210> 758
<211> 4
<212> PRT

<213> 人工序列
<220>
<223> 合成多肽
<400> 758
Ser Gly Gly Ser
1
<210> 759
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 759
Gly Gly Gly Ser
1
<210> 760
<211> 91
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 760
cctagggaag tgatcatagc tgagtttcta tctcatggtt tatgctaaac tatatgttga 60
catgttgagg agacttaagt ccaaaacctg g 91
<210> 761
<211> 30
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 761
Met Asp Ser Leu Leu Met Asn Arg Arg Lys Phe Leu Tyr Gln Phe Lys
1 5 10 15
Asn Val Arg Trp Ala Lys Gly Arg Arg Glu Thr Tyr Leu Cys
20 25 30
<210> 762
<211> 1129
<212> PRT
<213> 酸土脂环酸芽孢杆菌

<400> 762

Met Ala Val Lys Ser Ile Lys Val Lys Leu Arg Leu Asp Asp Met Pro
1 5 10 15
Glu Ile Arg Ala Gly Leu Trp Lys Leu His Lys Glu Val Asn Ala Gly
20 25 30
Val Arg Tyr Tyr Thr Glu Trp Leu Ser Leu Leu Arg Gln Glu Asn Leu
35 40 45
Tyr Arg Arg Ser Pro Asn Gly Asp Gly Glu Gln Glu Cys Asp Lys Thr
50 55 60
Ala Glu Glu Cys Lys Ala Glu Leu Leu Glu Arg Leu Arg Ala Arg Gln
65 70 75 80
Val Glu Asn Gly His Arg Gly Pro Ala Gly Ser Asp Asp Glu Leu Leu
85 90 95
Gln Leu Ala Arg Gln Leu Tyr Glu Leu Leu Val Pro Gln Ala Ile Gly
100 105 110
Ala Lys Gly Asp Ala Gln Gln Ile Ala Arg Lys Phe Leu Ser Pro Leu
115 120 125
Ala Asp Lys Asp Ala Val Gly Gly Leu Gly Ile Ala Lys Ala Gly Asn
130 135 140
Lys Pro Arg Trp Val Arg Met Arg Glu Ala Gly Glu Pro Gly Trp Glu
145 150 155 160
Glu Glu Lys Glu Lys Ala Glu Thr Arg Lys Ser Ala Asp Arg Thr Ala
165 170 175
Asp Val Leu Arg Ala Leu Ala Asp Phe Gly Leu Lys Pro Leu Met Arg
180 185 190
Val Tyr Thr Asp Ser Glu Met Ser Ser Val Glu Trp Lys Pro Leu Arg
195 200 205
Lys Gly Gln Ala Val Arg Thr Trp Asp Arg Asp Met Phe Gln Gln Ala
210 215 220
Ile Glu Arg Met Met Ser Trp Glu Ser Trp Asn Gln Arg Val Gly Gln
225 230 235 240
Glu Tyr Ala Lys Leu Val Glu Gln Lys Asn Arg Phe Glu Gln Lys Asn
245 250 255
Phe Val Gly Gln Glu His Leu Val His Leu Val Asn Gln Leu Gln Gln
260 265 270
Asp Met Lys Glu Ala Ser Pro Gly Leu Glu Ser Lys Glu Gln Thr Ala
275 280 285
His Tyr Val Thr Gly Arg Ala Leu Arg Gly Ser Asp Lys Val Phe Glu
290 295 300

Lys Trp Gly Lys Leu Ala Pro Asp Ala Pro Phe Asp Leu Tyr Asp Ala
305 310 315 320
Glu Ile Lys Asn Val Gln Arg Arg Asn Thr Arg Arg Phe Gly Ser His
325 330 335
Asp Leu Phe Ala Lys Leu Ala Glu Pro Glu Tyr Gln Ala Leu Trp Arg
340 345 350
Glu Asp Ala Ser Phe Leu Thr Arg Tyr Ala Val Tyr Asn Ser Ile Leu
355 360 365
Arg Lys Leu Asn His Ala Lys Met Phe Ala Thr Phe Thr Leu Pro Asp
370 375 380
Ala Thr Ala His Pro Ile Trp Thr Arg Phe Asp Lys Leu Gly Gly Asn
385 390 395 400
Leu His Gln Tyr Thr Phe Leu Phe Asn Glu Phe Gly Glu Arg Arg His
405 410 415
Ala Ile Arg Phe His Lys Leu Leu Lys Val Glu Asn Gly Val Ala Arg
420 425 430
Glu Val Asp Asp Val Thr Val Pro Ile Ser Met Ser Glu Gln Leu Asp
435 440 445
Asn Leu Leu Pro Arg Asp Pro Asn Glu Pro Ile Ala Leu Tyr Phe Arg
450 455 460
Asp Tyr Gly Ala Glu Gln His Phe Thr Gly Glu Phe Gly Gly Ala Lys
465 470 475 480
Ile Gln Cys Arg Arg Asp Gln Leu Ala His Met His Arg Arg Arg Gly
485 490 495
Ala Arg Asp Val Tyr Leu Asn Val Ser Val Arg Val Gln Ser Gln Ser
500 505 510
Glu Ala Arg Gly Glu Arg Arg Pro Pro Tyr Ala Ala Val Phe Arg Leu
515 520 525
Val Gly Asp Asn His Arg Ala Phe Val His Phe Asp Lys Leu Ser Asp
530 535 540
Tyr Leu Ala Glu His Pro Asp Asp Gly Lys Leu Gly Ser Glu Gly Leu
545 550 555 560
Leu Ser Gly Leu Arg Val Met Ser Val Asp Leu Gly Leu Arg Thr Ser
565 570 575
Ala Ser Ile Ser Val Phe Arg Val Ala Arg Lys Asp Glu Leu Lys Pro
580 585 590
Asn Ser Lys Gly Arg Val Pro Phe Phe Phe Pro Ile Lys Gly Asn Asp
595 600 605
Asn Leu Val Ala Val His Glu Arg Ser Gln Leu Leu Lys Leu Pro Gly

610 615 620
Glu Thr Glu Ser Lys Asp Leu Arg Ala Ile Arg Glu Glu Arg Gln Arg
625 630 635 640
Thr Leu Arg Gln Leu Arg Thr Gln Leu Ala Tyr Leu Arg Leu Leu Val
645 650 655
Arg Cys Gly Ser Glu Asp Val Gly Arg Arg Glu Arg Ser Trp Ala Lys
660 665 670
Leu Ile Glu Gln Pro Val Asp Ala Ala Asn His Met Thr Pro Asp Trp
675 680 685
Arg Glu Ala Phe Glu Asn Glu Leu Gln Lys Leu Lys Ser Leu His Gly
690 695 700
Ile Cys Ser Asp Lys Glu Trp Met Asp Ala Val Tyr Glu Ser Val Arg
705 710 715 720
Arg Val Trp Arg His Met Gly Lys Gln Val Arg Asp Trp Arg Lys Asp
725 730 735
Val Arg Ser Gly Glu Arg Pro Lys Ile Arg Gly Tyr Ala Lys Asp Val
740 745 750
Val Gly Gly Asn Ser Ile Glu Gln Ile Glu Tyr Leu Glu Arg Gln Tyr
755 760 765
Lys Phe Leu Lys Ser Trp Ser Phe Phe Gly Lys Val Ser Gly Gln Val
770 775 780
Ile Arg Ala Glu Lys Gly Ser Arg Phe Ala Ile Thr Leu Arg Glu His
785 790 795 800
Ile Asp His Ala Lys Glu Asp Arg Leu Lys Lys Leu Ala Asp Arg Ile
805 810 815
Ile Met Glu Ala Leu Gly Tyr Val Tyr Ala Leu Asp Glu Arg Gly Lys
820 825 830
Gly Lys Trp Val Ala Lys Tyr Pro Pro Cys Gln Leu Ile Leu Leu Glu
835 840 845
Glu Leu Ser Glu Tyr Gln Phe Asn Asn Asp Arg Pro Pro Ser Glu Asn
850 855 860
Asn Gln Leu Met Gln Trp Ser His Arg Gly Val Phe Gln Glu Leu Ile
865 870 875 880
Asn Gln Ala Gln Val His Asp Leu Leu Val Gly Thr Met Tyr Ala Ala
885 890 895
Phe Ser Ser Arg Phe Asp Ala Arg Thr Gly Ala Pro Gly Ile Arg Cys
900 905 910
Arg Arg Val Pro Ala Arg Cys Thr Gln Glu His Asn Pro Glu Pro Phe
915 920 925

Pro Trp Trp Leu Asn Lys Phe Val Val Glu His Thr Leu Asp Ala Cys
930 935 940
Pro Leu Arg Ala Asp Asp Leu Ile Pro Thr Gly Glu Gly Glu Ile Phe
945 950 955 960
Val Ser Pro Phe Ser Ala Glu Glu Gly Asp Phe His Gln Ile His Ala
965 970 975
Asp Leu Asn Ala Ala Gln Asn Leu Gln Gln Arg Leu Trp Ser Asp Phe
980 985 990
Asp Ile Ser Gln Ile Arg Leu Arg Cys Asp Trp Gly Glu Val Asp Gly
995 1000 1005
Glu Leu Val Leu Ile Pro Arg Leu Thr Gly Lys Arg Thr Ala Asp
1010 1015 1020
Ser Tyr Ser Asn Lys Val Phe Tyr Thr Asn Thr Gly Val Thr Tyr
1025 1030 1035
Tyr Glu Arg Glu Arg Gly Lys Lys Arg Arg Lys Val Phe Ala Gln
1040 1045 1050
Glu Lys Leu Ser Glu Glu Glu Ala Glu Leu Leu Val Glu Ala Asp
1055 1060 1065
Glu Ala Arg Glu Lys Ser Val Val Leu Met Arg Asp Pro Ser Gly
1070 1075 1080
Ile Ile Asn Arg Gly Asn Trp Thr Arg Gln Lys Glu Phe Trp Ser
1085 1090 1095
Met Val Asn Gln Arg Ile Glu Gly Tyr Leu Val Lys Gln Ile Arg
1100 1105 1110
Ser Arg Val Pro Leu Gln Asp Ser Ala Cys Glu Asn Thr Gly Asp
1115 1120 1125
Ile

<210> 763
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 763
Asn Gly Gly Asn Gly
1 5

<210> 764
<211> 1389
<212> PRT

<213> Leptotrichia shahii

<400> 764

Met Gly Asn Leu Phe Gly His Lys Arg Trp Tyr Glu Val Arg Asp Lys
1 5 10 15
Lys Asp Phe Lys Ile Lys Arg Lys Val Lys Val Lys Arg Asn Tyr Asp
20 25 30
Gly Asn Lys Tyr Ile Leu Asn Ile Asn Glu Asn Asn Asn Lys Glu Lys
35 40 45
Ile Asp Asn Asn Lys Phe Ile Arg Lys Tyr Ile Asn Tyr Lys Lys Asn
50 55 60
Asp Asn Ile Leu Lys Glu Phe Thr Arg Lys Phe His Ala Gly Asn Ile
65 70 75 80
Leu Phe Lys Leu Lys Gly Lys Glu Gly Ile Ile Arg Ile Glu Asn Asn
85 90 95
Asp Asp Phe Leu Glu Thr Glu Glu Val Val Leu Tyr Ile Glu Ala Tyr
100 105 110
Gly Lys Ser Glu Lys Leu Lys Ala Leu Gly Ile Thr Lys Lys Lys Ile
115 120 125
Ile Asp Glu Ala Ile Arg Gln Gly Ile Thr Lys Asp Asp Lys Lys Ile
130 135 140
Glu Ile Lys Arg Gln Glu Asn Glu Glu Glu Ile Glu Ile Asp Ile Arg
145 150 155 160
Asp Glu Tyr Thr Asn Lys Thr Leu Asn Asp Cys Ser Ile Ile Leu Arg
165 170 175
Ile Ile Glu Asn Asp Glu Leu Glu Thr Lys Lys Ser Ile Tyr Glu Ile
180 185 190
Phe Lys Asn Ile Asn Met Ser Leu Tyr Lys Ile Ile Glu Lys Ile Ile
195 200 205
Glu Asn Glu Thr Glu Lys Val Phe Glu Asn Arg Tyr Tyr Glu Glu His
210 215 220
Leu Arg Glu Lys Leu Leu Lys Asp Asp Lys Ile Asp Val Ile Leu Thr
225 230 235 240
Asn Phe Met Glu Ile Arg Glu Lys Ile Lys Ser Asn Leu Glu Ile Leu
245 250 255
Gly Phe Val Lys Phe Tyr Leu Asn Val Gly Gly Asp Lys Lys Lys Ser
260 265 270
Lys Asn Lys Lys Met Leu Val Glu Lys Ile Leu Asn Ile Asn Val Asp
275 280 285
Leu Thr Val Glu Asp Ile Ala Asp Phe Val Ile Lys Glu Leu Glu Phe

290 295 300

Trp Asn Ile Thr Lys Arg Ile Glu Lys Val Lys Lys Val Asn Asn Glu
305 310 315 320
Phe Leu Glu Lys Arg Arg Asn Arg Thr Tyr Ile Lys Ser Tyr Val Leu
325 330 335
Leu Asp Lys His Glu Lys Phe Lys Ile Glu Arg Glu Asn Lys Lys Asp
340 345 350
Lys Ile Val Lys Phe Phe Val Glu Asn Ile Lys Asn Asn Ser Ile Lys
355 360 365
Glu Lys Ile Glu Lys Ile Leu Ala Glu Phe Lys Ile Asp Glu Leu Ile
370 375 380
Lys Lys Leu Glu Lys Glu Leu Lys Lys Gly Asn Cys Asp Thr Glu Ile
385 390 395 400
Phe Gly Ile Phe Lys Lys His Tyr Lys Val Asn Phe Asp Ser Lys Lys
405 410 415
Phe Ser Lys Lys Ser Asp Glu Glu Lys Glu Leu Tyr Lys Ile Ile Tyr
420 425 430
Arg Tyr Leu Lys Gly Arg Ile Glu Lys Ile Leu Val Asn Glu Gln Lys
435 440 445
Val Arg Leu Lys Lys Met Glu Lys Ile Glu Ile Glu Lys Ile Leu Asn
450 455 460
Glu Ser Ile Leu Ser Glu Lys Ile Leu Lys Arg Val Lys Gln Tyr Thr
465 470 475 480
Leu Glu His Ile Met Tyr Leu Gly Lys Leu Arg His Asn Asp Ile Asp
485 490 495
Met Thr Thr Val Asn Thr Asp Asp Phe Ser Arg Leu His Ala Lys Glu
500 505 510
Glu Leu Asp Leu Glu Leu Ile Thr Phe Phe Ala Ser Thr Asn Met Glu
515 520 525
Leu Asn Lys Ile Phe Ser Arg Glu Asn Ile Asn Asn Asp Glu Asn Ile
530 535 540
Asp Phe Phe Gly Gly Asp Arg Glu Lys Asn Tyr Val Leu Asp Lys Lys
545 550 555 560
Ile Leu Asn Ser Lys Ile Lys Ile Ile Arg Asp Leu Asp Phe Ile Asp
565 570 575
Asn Lys Asn Asn Ile Thr Asn Asn Phe Ile Arg Lys Phe Thr Lys Ile
580 585 590
Gly Thr Asn Glu Arg Asn Arg Ile Leu His Ala Ile Ser Lys Glu Arg
595 600 605

Asp Leu Gln Gly Thr Gln Asp Asp Tyr Asn Lys Val Ile Asn Ile Ile
610 615 620
Gln Asn Leu Lys Ile Ser Asp Glu Glu Val Ser Lys Ala Leu Asn Leu
625 630 635 640
Asp Val Val Phe Lys Asp Lys Lys Asn Ile Ile Thr Lys Ile Asn Asp
645 650 655
Ile Lys Ile Ser Glu Glu Asn Asn Asn Asp Ile Lys Tyr Leu Pro Ser
660 665 670
Phe Ser Lys Val Leu Pro Glu Ile Leu Asn Leu Tyr Arg Asn Asn Pro
675 680 685
Lys Asn Glu Pro Phe Asp Thr Ile Glu Thr Glu Lys Ile Val Leu Asn
690 695 700
Ala Leu Ile Tyr Val Asn Lys Glu Leu Tyr Lys Lys Leu Ile Leu Glu
705 710 715 720
Asp Asp Leu Glu Glu Asn Glu Ser Lys Asn Ile Phe Leu Gln Glu Leu
725 730 735
Lys Lys Thr Leu Gly Asn Ile Asp Glu Ile Asp Glu Asn Ile Ile Glu
740 745 750
Asn Tyr Tyr Lys Asn Ala Gln Ile Ser Ala Ser Lys Gly Asn Asn Lys
755 760 765
Ala Ile Lys Lys Tyr Gln Lys Lys Val Ile Glu Cys Tyr Ile Gly Tyr
770 775 780
Leu Arg Lys Asn Tyr Glu Glu Leu Phe Asp Phe Ser Asp Phe Lys Met
785 790 795 800
Asn Ile Gln Glu Ile Lys Lys Gln Ile Lys Asp Ile Asn Asp Asn Lys
805 810 815
Thr Tyr Glu Arg Ile Thr Val Lys Thr Ser Asp Lys Thr Ile Val Ile
820 825 830
Asn Asp Asp Phe Glu Tyr Ile Ile Ser Ile Phe Ala Leu Leu Asn Ser
835 840 845
Asn Ala Val Ile Asn Lys Ile Arg Asn Arg Phe Phe Ala Thr Ser Val
850 855 860
Trp Leu Asn Thr Ser Glu Tyr Gln Asn Ile Ile Asp Ile Leu Asp Glu
865 870 875 880
Ile Met Gln Leu Asn Thr Leu Arg Asn Glu Cys Ile Thr Glu Asn Trp
885 890 895
Asn Leu Asn Leu Glu Glu Phe Ile Gln Lys Met Lys Glu Ile Glu Lys
900 905 910
Asp Phe Asp Asp Phe Lys Ile Gln Thr Lys Lys Glu Ile Phe Asn Asn

915 920 925

Tyr Tyr Glu Asp Ile Lys Asn Asn Ile Leu Thr Glu Phe Lys Asp Asp

930 935 940

Ile Asn Gly Cys Asp Val Leu Glu Lys Lys Leu Glu Lys Ile Val Ile

945 950 955 960

Phe Asp Asp Glu Thr Lys Phe Glu Ile Asp Lys Lys Ser Asn Ile Leu

965 970 975

Gln Asp Glu Gln Arg Lys Leu Ser Asn Ile Asn Lys Lys Asp Leu Lys

980 985 990

Lys Lys Val Asp Gln Tyr Ile Lys Asp Lys Asp Gln Glu Ile Lys Ser

995 1000 1005

Lys Ile Leu Cys Arg Ile Ile Phe Asn Ser Asp Phe Leu Lys Lys

1010 1015 1020

Tyr Lys Lys Glu Ile Asp Asn Leu Ile Glu Asp Met Glu Ser Glu

1025 1030 1035

Asn Glu Asn Lys Phe Gln Glu Ile Tyr Tyr Pro Lys Glu Arg Lys

1040 1045 1050

Asn Glu Leu Tyr Ile Tyr Lys Lys Asn Leu Phe Leu Asn Ile Gly

1055 1060 1065

Asn Pro Asn Phe Asp Lys Ile Tyr Gly Leu Ile Ser Asn Asp Ile

1070 1075 1080

Lys Met Ala Asp Ala Lys Phe Leu Phe Asn Ile Asp Gly Lys Asn

1085 1090 1095

Ile Arg Lys Asn Lys Ile Ser Glu Ile Asp Ala Ile Leu Lys Asn

1100 1105 1110

Leu Asn Asp Lys Leu Asn Gly Tyr Ser Lys Glu Tyr Lys Glu Lys

1115 1120 1125

Tyr Ile Lys Lys Leu Lys Glu Asn Asp Asp Phe Phe Ala Lys Asn

1130 1135 1140

Ile Gln Asn Lys Asn Tyr Lys Ser Phe Glu Lys Asp Tyr Asn Arg

1145 1150 1155

Val Ser Glu Tyr Lys Lys Ile Arg Asp Leu Val Glu Phe Asn Tyr

1160 1165 1170

Leu Asn Lys Ile Glu Ser Tyr Leu Ile Asp Ile Asn Trp Lys Leu

1175 1180 1185

Ala Ile Gln Met Ala Arg Phe Glu Arg Asp Met His Tyr Ile Val

1190 1195 1200

Asn Gly Leu Arg Glu Leu Gly Ile Ile Lys Leu Ser Gly Tyr Asn

1205 1210 1215

Thr Gly Ile Ser Arg Ala Tyr Pro Lys Arg Asn Gly Ser Asp Gly
1220 1225 1230
Phe Tyr Thr Thr Thr Ala Tyr Tyr Lys Phe Phe Asp Glu Glu Ser
1235 1240 1245
Tyr Lys Lys Phe Glu Lys Ile Cys Tyr Gly Phe Gly Ile Asp Leu
1250 1255 1260
Ser Glu Asn Ser Glu Ile Asn Lys Pro Glu Asn Glu Ser Ile Arg
1265 1270 1275
Asn Tyr Ile Ser His Phe Tyr Ile Val Arg Asn Pro Phe Ala Asp
1280 1285 1290
Tyr Ser Ile Ala Glu Gln Ile Asp Arg Val Ser Asn Leu Leu Ser
1295 1300 1305
Tyr Ser Thr Arg Tyr Asn Asn Ser Thr Tyr Ala Ser Val Phe Glu
1310 1315 1320
Val Phe Lys Lys Asp Val Asn Leu Asp Tyr Asp Glu Leu Lys Lys
1325 1330 1335
Lys Phe Lys Leu Ile Gly Asn Asn Asp Ile Leu Glu Arg Leu Met
1340 1345 1350
Lys Pro Lys Lys Val Ser Val Leu Glu Leu Glu Ser Tyr Asn Ser
1355 1360 1365
Asp Tyr Ile Lys Asn Leu Ile Ile Glu Leu Leu Thr Lys Ile Glu
1370 1375 1380
Asn Thr Asn Asp Thr Leu
1385

<210> 765
<211> 5268
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 765

atgtccaacc tccttaccgt ccaccagaat ctccctgccc ttccggtgga tgccacctct 60
gatgaagtgc gaaaaaacct gatggatatg tttcgcgata ggcaagcttt ttctgaacac 120
acgtggaaga tgctcctgtc agtgtgtaga agctgggcag cttggtgcaa gttgaacaac 180
cgaaaatggt ttcctgccga acccgaagat gtgagagact acctcctcta cctgcaggct 240
cgagggctcg ccgtgaaaac aatccaacaa cacttgggtc agctcaacat gctgcacagg 300
agatctgggc tgccccggcc gagtgactct aatgccgtta gtctcgtaat gcggcgcatt 360
cgcaaagaga atgtggatgc tggagaacgg gcgaaacagg cactggcttt tgaacggacc 420
gacttcgatc aggtgcggag tcttatggag aatagtgaca gatgccagga cattcggaac 480

cttgcattcc tgggtatcgc gtataatacc ctgctgagaa tcgctgagat cgccagaatc 540
agggtaaagg atatttctcg aacggacggg ggacggatgt tgattcatat cggtcgcact 600
aaaacacttg tgagtaccgc cggggtagag aaagccctga gccttggagt tactaaactg 660
gtggagcggt ggattagcgt gtccggcgtg gcggatgacc caaacaatta cttgttttgt 720
agggtgcgga aaaatggtgt agccgctcca tccgctacct cacagttgag tacacgcgcg 780
ttggagggga ttttcgaagc cacacatcgc ttgatctacg gcgccaagga cgattcaggc 840
cagcgatatc ttgcctggag cgggcatagt gcccgggtgg gtgccgcccg agacatggca 900
agggctggcg tgtcaattcc tgaaatcatg caggccggcg ggtggaccaa cgtgaacatt 960
gtgatgaact atatccggaa cctggatagc gagaccggag caatggtcag actgcttgag 1020
gatggcgacg gtggatccgg agggtccgga ggtagtggcg gcagcggtgg ttcaggtggc 1080
agcggagggt caggaggctc tgataaaaag tattctattg gtttagctat cggcactaat 1140
tccgttggat gggctgtcat aaccgatgaa tacaaagtac cttcaaagaa atttaaggtg 1200
ttggggaaca cagaccgtca ttcgattaaa aagaatctta tcggtgccct cctattcgat 1260
agtggcgaaa cggcagaggc gactcgcctg aaacgaaccg ctcggagaag gtatacacgt 1320
cgcaagaacc gaatatgtta cttacaagaa atttttagca atgagatggc caaagttgac 1380
gattctttct ttcaccgttt ggaagagtcc ttccttgtcg aagaggacaa gaaacatgaa 1440
cggcacccca tctttggaaa catagtagat gaggtggcat atcatgaaaa gtacccaacg 1500
atttatcacc tcagaaaaaa gctagttgac tcaactgata aagcggacct gaggttaatc 1560
tacttggctc ttgcccatat gataaagttc cgtgggcact ttctcattga gggtgatcta 1620
aatccggaca actcggatgt cgacaaactg ttcatccagt tagtacaaac ctataatcag 1680
ttgtttgaag agaaccctat aaatgcaagt ggcgtggatg cgaaggctat tcttagcgcc 1740
cgcctctcta aatcccgacg gctagaaaac ctgatcgcac aattacccgg agagaagaaa 1800
aatgggttgt tcggtaacct tatagcgctc tcactaggcc tgacaccaaa ttttaagtcg 1860
aacttcgact tagctgaaga tgccaaattg cagcttagta aggacacgta cgatgacgat 1920
ctcgacaatc tactggcaca aattggagat cagtatgcgg acttattttt ggctgccaaa 1980
aaccttagcg atgcaatcct cctatctgac atactgagag ttaatactga gattaccaag 2040
gcgccgttat ccgcttcaat gatcaaaagg tacgatgaac atcaccaaga cttgacactt 2100
ctcaaggccc tagtccgtca gcaactgcct gagaaatata aggaaatatt ctttgatcag 2160
tcgaaaaacg ggtacgcagg ttatattgac ggcggagcga gtcaagagga attctacaag 2220
tttatcaaac ccatattaga gaagatggat gggacggaag agttgcttgt aaaactcaat 2280
cgcgaagatc tactgcgaaa gcagcggact ttcgacaacg gtagcattcc acatcaaatc 2340
cacttaggcg aattgcatgc tatacttaga aggcaggagg atttttatcc gttcctcaaa 2400
gacaatcgtg aaaagattga gaaaatccta acctttcgca taccttacta tgtgggaccc 2460
ctggcccgag ggaactctcg gttcgcatgg atgacaagaa agtccgaaga aacgattact 2520
ccatggaatt ttgaggaagt tgtcgataaa ggtgcgtcag ctcaatcgtt catcgagagg 2580
atgaccaact ttgacaagaa tttaccgaac gaaaaagtat tgcctaagca cagtttactt 2640
tacgagtatt tcacagtgta caatgaactc acgaaagtta agtatgtcac tgagggcatg 2700
cgtaaacccg cctttctaag cggagaacag aagaaagcaa tagtagatct gttattcaag 2760
accaaccgca aagtgacagt taagcaattg aaagaggact actttaagaa aattgaatgc 2820

ttcgattctg tcgagatctc cggggtagaa gatcgattta atgcgtcact tggtacgtat 2880
catgacctcc taaagataat taaagataag gacttcctgg ataacgaaga gaatgaagat 2940
atcttagaag atatagtgtt gactcttacc ctctttgaag atcgggaaat gattgaggaa 3000
agactaaaaa catacgctca cctgttcgac gataaggtta tgaaacagtt aaagaggcgt 3060
cgctatacgg gctggggacg attgtcgcgg aaacttatca acgggataag agacaagcaa 3120
agtggtaaaa ctattctcga ttttctaaag agcgacggct tcgccaatag gaactttatg 3180
cagctgatcc atgatgactc tttaaccttc aaagaggata tacaaaaggc acaggtttcc 3240
ggacaagggg actcattgca cgaacatatt gcgaatcttg ctggttcgcc agccatcaaa 3300
aagggcatac tccagacagt caaagtagtg gatgagctag ttaaggtcat gggacgtcac 3360
aaaccggaaa acattgtaat cgagatggca cgcgaaaatc aaacgactca gaaggggcaa 3420
aaaaacagtc gagagcggat gaagagaata gaagagggta ttaaagaact gggcagccag 3480
atcttaaagg agcatcctgt ggaaaatacc caattgcaga acgagaaact ttacctctat 3540
tacctacaaa atggaaggga catgtatgtt gatcaggaac tggacataaa ccgtttatct 3600
gattacgacg tcgatgccat tgtaccccaa tccttttga aggacgattc aatcgacaat 3660
aaagtgctta cacgctcgga taagaaccga gggaaaagtg acaatgttcc aagcgaggaa 3720
gtcgtaaaga aaatgaagaa ctattggcgg cagctcctaa atgcgaaact gataacgcaa 3780
agaaagttcg ataacttaac taaagctgag aggggtggct tgtctgaact tgacaaggcc 3840
ggatttatta aacgtcagct cgtggaaacc cgccaaatca caaagcatgt tgcacagata 3900
ctagattccc gaatgaatac gaaatacgac gagaacgata agctgattcg ggaagtcaaa 3960
gtaatcactt taaagtcaaa attggtgtcg gacttcagaa aggattttca attctataaa 4020
gttagggaga taaataacta ccaccatgcg cacgacgctt atcttaatgc cgtcgtaggg 4080
accgcactca ttaagaaata cccgaagcta gaaagtgagt ttgtgtatgg tgattacaaa 4140
gtttatgacg tccgtaagat gatcgcgaaa agcgaacagg agataggcaa ggctacagcc 4200
aaatacttct tttattctaa cattatgaat ttctttaaga cggaaatcac tctggcaaac 4260
ggagagatac gcaaacgacc tttaattgaa accaatgggg agacaggtga aatcgtatgg 4320
gataagggcc gggacttcgc gacggtgaga aaagttttgt ccatgcccca agtcaacata 4380
gtaaagaaaa ctgaggtgca gaccggaggg ttttcaaagg aatcgattct tccaaaaagg 4440
aatagtgata agctcatcgc tcgtaaaaag gactgggacc cgaaaaagta cggtggcttc 4500
gatagcccta cagttgccta ttctgtccta gtagtggcaa aagttgagaa gggaaaatcc 4560
aagaaactga agtcagtcaa agaattattg gggataacga ttatggagcg ctcgtctttt 4620
gaaaagaacc ccatcgactt ccttgaggcg aaaggttaca aggaagtaaa aaaggatctc 4680
ataattaaac taccaaagta tagtctgttt gagttagaaa atggccgaaa acggatgttg 4740
gctagcgccg gagagcttca aaaggggaac gaactcgcac taccgtctaa atacgtgaat 4800
ttcctgtatt tagcgtccca ttacgagaag ttgaaaggtt cacctgaaga taacgaacag 4860
aagcaacttt ttgttgagca gcacaaacat tatctcgacg aaatcataga gcaaatttcg 4920
gaattcagta agagagtcat cctagctgat gccaatctgg acaaagtatt aagcgcatac 4980
aacaagcaca gggataaacc catacgtgag caggcggaaa atattatcca tttgtttact 5040
cttaccaacc tcggcgctcc agccgcattc aagtattttg acacaacgat agatcgcaaa 5100
cgatacactt ctaccaagga ggtgctagac gcgacactga ttcaccaatc catcacggga 5160

ttatatgaaa ctcggataga tttgtcacag cttgggggtg acggtggctc cgattataag 5220
gatgatgacg acaagggagg ttccccaaag aagaaaagga aggtctga 5268

<210> 766
<211> 1755
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 766

Met Ser Asn Leu Leu Thr Val His Gln Asn Leu Pro Ala Leu Pro Val
1 5 10 15
Asp Ala Thr Ser Asp Glu Val Arg Lys Asn Leu Met Asp Met Phe Arg
20 25 30
Asp Arg Gln Ala Phe Ser Glu His Thr Trp Lys Met Leu Leu Ser Val
35 40 45
Cys Arg Ser Trp Ala Ala Trp Cys Lys Leu Asn Asn Arg Lys Trp Phe
50 55 60
Pro Ala Glu Pro Glu Asp Val Arg Asp Tyr Leu Leu Tyr Leu Gln Ala
65 70 75 80
Arg Gly Leu Ala Val Lys Thr Ile Gln Gln His Leu Gly Gln Leu Asn
85 90 95
Met Leu His Arg Arg Ser Gly Leu Pro Arg Pro Ser Asp Ser Asn Ala
100 105 110
Val Ser Leu Val Met Arg Arg Ile Arg Lys Glu Asn Val Asp Ala Gly
115 120 125
Glu Arg Ala Lys Gln Ala Leu Ala Phe Glu Arg Thr Asp Phe Asp Gln
130 135 140
Val Arg Ser Leu Met Glu Asn Ser Asp Arg Cys Gln Asp Ile Arg Asn
145 150 155 160
Leu Ala Phe Leu Gly Ile Ala Tyr Asn Thr Leu Leu Arg Ile Ala Glu
165 170 175
Ile Ala Arg Ile Arg Val Lys Asp Ile Ser Arg Thr Asp Gly Gly Arg
180 185 190
Met Leu Ile His Ile Gly Arg Thr Lys Thr Leu Val Ser Thr Ala Gly
195 200 205
Val Glu Lys Ala Leu Ser Leu Gly Val Thr Lys Leu Val Glu Arg Trp
210 215 220
Ile Ser Val Ser Gly Val Ala Asp Asp Pro Asn Asn Tyr Leu Phe Cys
225 230 235 240

Arg Val Arg Lys Asn Gly Val Ala Ala Pro Ser Ala Thr Ser Gln Leu
245 250 255
Ser Thr Arg Ala Leu Glu Gly Ile Phe Glu Ala Thr His Arg Leu Ile
260 265 270
Tyr Gly Ala Lys Asp Asp Ser Gly Gln Arg Tyr Leu Ala Trp Ser Gly
275 280 285
His Ser Ala Arg Val Gly Ala Ala Arg Asp Met Ala Arg Ala Gly Val
290 295 300
Ser Ile Pro Glu Ile Met Gln Ala Gly Gly Trp Thr Asn Val Asn Ile
305 310 315 320
Val Met Asn Tyr Ile Arg Asn Leu Asp Ser Glu Thr Gly Ala Met Val
325 330 335
Arg Leu Leu Glu Asp Gly Asp Gly Gly Ser Gly Gly Ser Gly Gly Ser
340 345 350
Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Gly Gly Ser Asp
355 360 365
Lys Lys Tyr Ser Ile Gly Leu Ala Ile Gly Thr Asn Ser Val Gly Trp
370 375 380
Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe Lys Val
385 390 395 400
Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile Gly Ala
405 410 415
Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu Lys Arg
420 425 430
Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys Tyr Leu
435 440 445
Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser Phe Phe
450 455 460
His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys His Glu
465 470 475 480
Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr His Glu
485 490 495
Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp Ser Thr
500 505 510
Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His Met Ile
515 520 525
Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro Asp Asn
530 535 540
Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr Asn Gln

545 550 555 560
Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala Lys Ala
565 570 575
Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn Leu Ile
580 585 590
Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn Leu Ile
595 600 605
Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe Asp Leu
610 615 620
Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp Asp Asp
625 630 635 640
Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp Leu Phe
645 650 655
Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp Ile Leu
660 665 670
Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser Met Ile
675 680 685
Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys Ala Leu
690 695 700
Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe Asp Gln
705 710 715 720
Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser Gln Glu
725 730 735
Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp Gly Thr
740 745 750
Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg Lys Gln
755 760 765
Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu Gly Glu
770 775 780
Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe Leu Lys
785 790 795 800
Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile Pro Tyr
805 810 815
Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp Met Thr
820 825 830
Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu Val Val
835 840 845
Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr Asn Phe
850 855 860

Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser Leu Leu
865 870 875 880
Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys Tyr Val
885 890 895
Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln Lys Lys
900 905 910
Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr Val Lys
915 920 925
Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp Ser Val
930 935 940
Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly Thr Tyr
945 950 955 960
His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp Asn Glu
965 970 975
Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr Leu Phe
980 985 990
Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala His Leu
995 1000 1005
Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr Thr
1010 1015 1020
Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
1025 1030 1035
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly
1040 1045 1050
Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu
1055 1060 1065
Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly
1070 1075 1080
Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala
1085 1090 1095
Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu
1100 1105 1110
Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
1115 1120 1125
Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser
1130 1135 1140
Arg Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly
1145 1150 1155
Ser Gln Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln

1160 1165 1170

Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met

1175 1180 1185

Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp

1190 1195 1200

Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile

1205 1210 1215

Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly Lys Ser

1220 1225 1230

Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn Tyr

1235 1240 1245

Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe

1250 1255 1260

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp

1265 1270 1275

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile

1280 1285 1290

Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys

1295 1300 1305

Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr

1310 1315 1320

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe

1325 1330 1335

Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala

1340 1345 1350

Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro

1355 1360 1365

Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp

1370 1375 1380

Val Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala

1385 1390 1395

Thr Ala Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys

1400 1405 1410

Thr Glu Ile Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu

1415 1420 1425

Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly

1430 1435 1440

Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser Met Pro Gln Val

1445 1450 1455

Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly Phe Ser Lys
1460 1465 1470
Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile Ala Arg
1475 1480 1485
Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser Pro
1490 1495 1500
Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1505 1510 1515
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr
1520 1525 1530
Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
1535 1540 1545
Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
1550 1555 1560
Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg
1565 1570 1575
Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala
1580 1585 1590
Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1595 1600 1605
Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu
1610 1615 1620
Phe Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln
1625 1630 1635
Ile Ser Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu
1640 1645 1650
Asp Lys Val Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile
1655 1660 1665
Arg Glu Gln Ala Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn
1670 1675 1680
Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp
1685 1690 1695
Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu
1700 1705 1710
Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu
1715 1720 1725
Ser Gln Leu Gly Gly Asp Gly Gly Ser Asp Tyr Lys Asp Asp Asp
1730 1735 1740
Asp Lys Gly Gly Ser Pro Lys Lys Lys Arg Lys Val

1745 1750 1755

<210> 767
<211> 95
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 767
cctgaaataa tgcaagtgta gaataacttt ttaaaatctc atggtttatg ctaaactata 60
tgttgacata agagtggtga taaggcaaca gtagg 95
<210> 768
<211> 91
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 768
cctagggaag tgatcatagc tgagtttcta tctcatggtt tatgctaaac tatatgttga 60
catgttgagg agacttaagt ccaaaacctg g 91
<210> 769
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 合成多核苷酸
<400> 769
atgtcctgaa ataatgcaag tgtagaataa ctttttaaaa tctcatggtt tatgctaaac 60
tatatgttga cataagagtg gtgataaggc aacagtaggt aaaag 105
<210> 770
<211> 35
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<220>
<221> MISC_FEATURE
<222> (3)..(3)
<223> Xaa是终止密码子
<220>

<221> MISC_FEATURE
<222> (10)..(10)
<223> Xaa是终止密码子
<220>
<221> MISC_FEATURE
<222> (25)..(25)
<223> Xaa是终止密码子
<220>
<221> MISC_FEATURE
<222> (28)..(29)
<223> Xaa是终止密码子
<220>
<221> MISC_FEATURE
<222> (34)..(34)
<223> Xaa是终止密码子
<400> 770

```
Met Ser Xaa Asn Asn Ala Ser Val Glu Xaa Leu Phe Lys Ile Ser Trp
1               5                  10                  15
Phe Met Leu Asn Tyr Met Leu Thr Xaa Glu Trp Xaa Xaa Gly Asn Ser
            20                  25                  30
Arg Xaa Lys
        35
```

<210> 771
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 771

```
Asn Ala Ala Arg
1
```

<210> 772
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 772

```
Ala Gly Val Phe
```

1
<210> 773
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 773
Gly Phe Leu Gly
1
<210> 774
<211> 4
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 774
Ala Leu Ala Leu
1
<210> 775
<211> 5
<212> PRT
<213> 人工序列
<220>
<223> 合成多肽
<400> 775
Ala Leu Ala Leu Ala
1 5

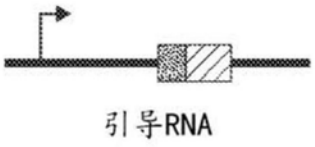


图1A

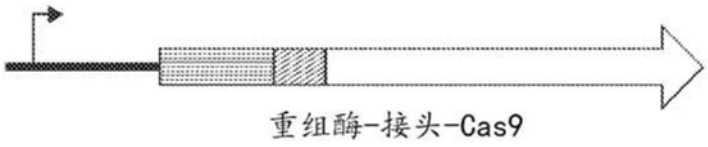


图1B

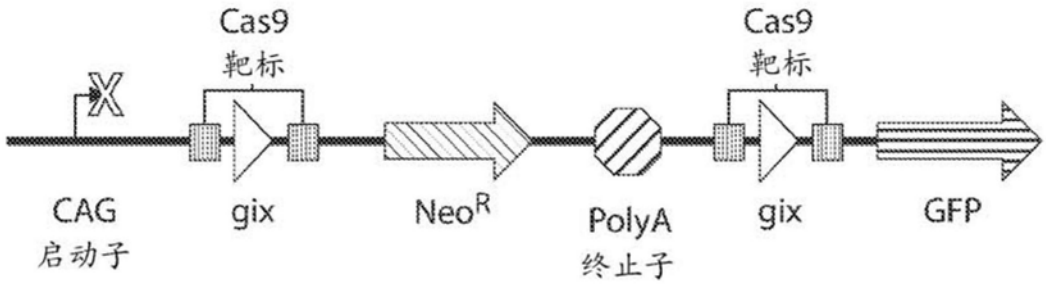


图1C

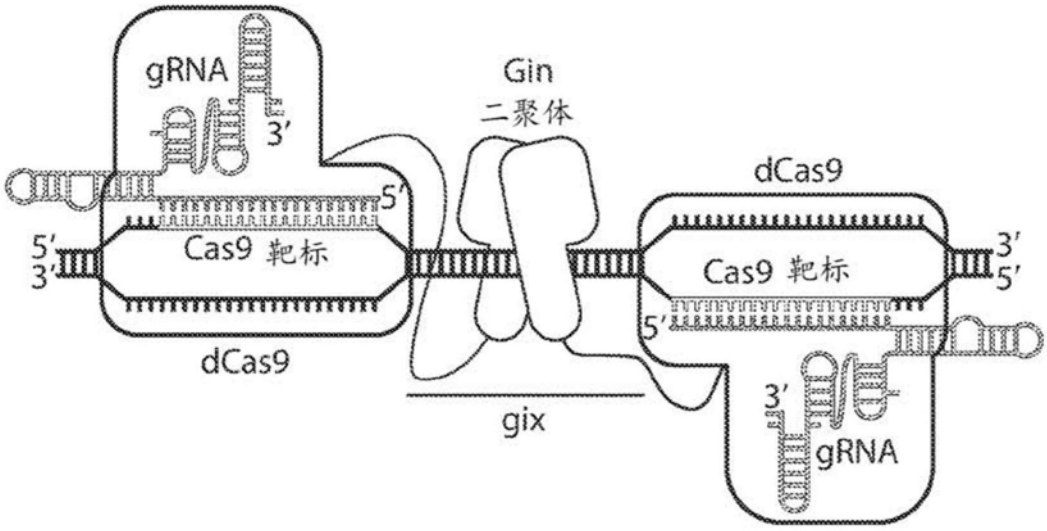


图1D

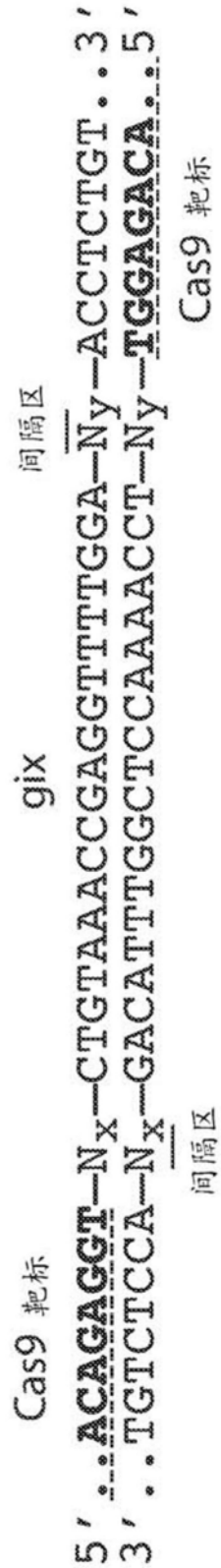

Cas9 靶标
gix
间隔区
5′ ACAGAGGT—Nx—CTGTAAACCGAGGTTTTGGA—Ny—ACCTCTGT 3′
3′ TGTCTCCA—Nx—GACATTTGGCTCCAAAACCT—Ny—TGGAGACA 5′
间隔区
Cas9 靶标

图2A

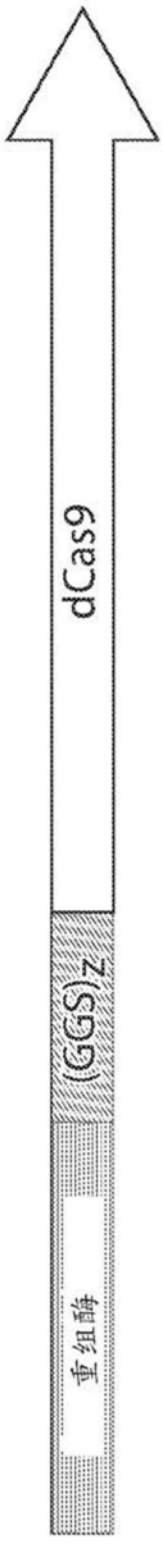


图2B

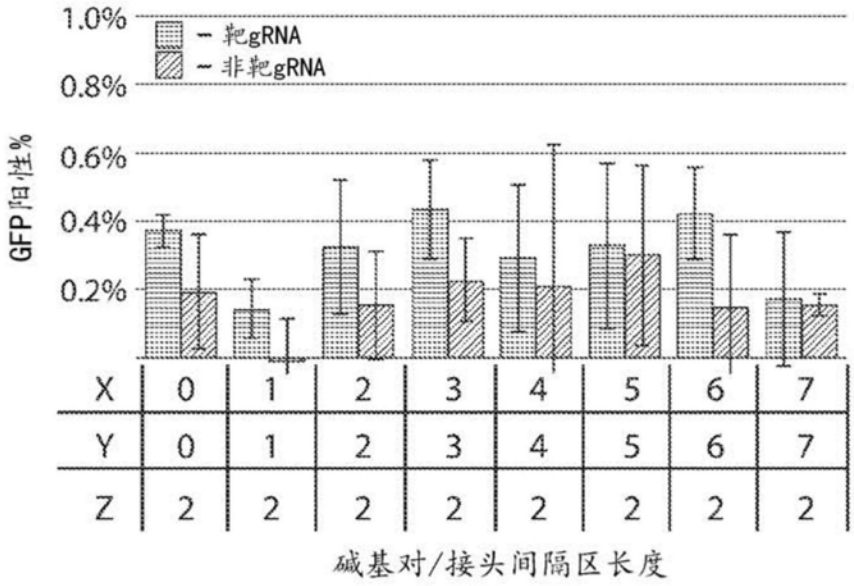

1.0%
0.8%
0.6%
0.4%
0.2%
GFP阳性%
靶gRNA
非靶gRNA
X 0 1 2 3 4 5 6 7
Y 0 1 2 3 4 5 6 7
Z 2 2 2 2 2 2 2 2
碱基对/接头间隔区长度

图2C

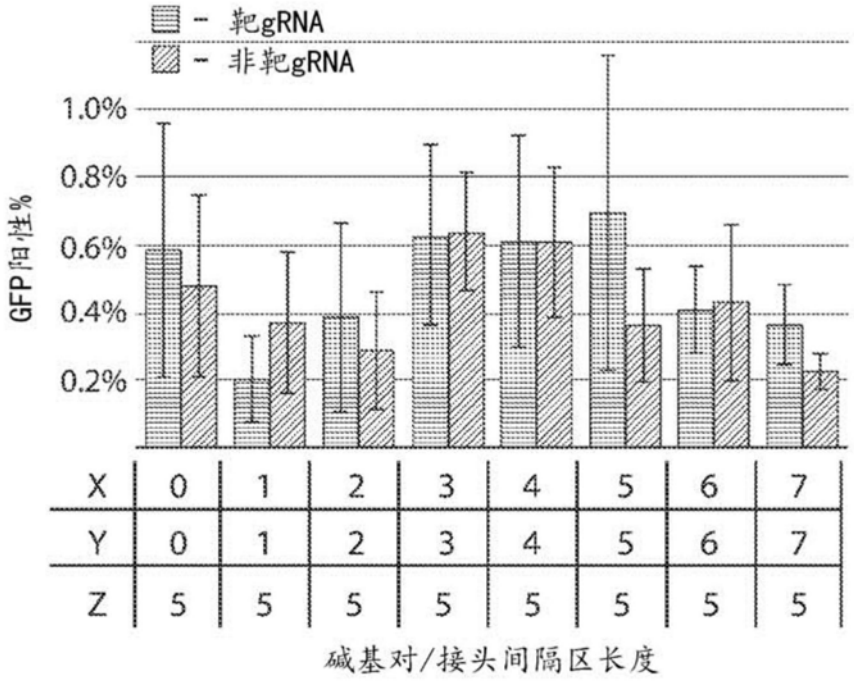

靶gRNA
非靶gRNA
1.0%
0.8%
0.6%
0.4%
0.2%
GFP阳性%
X 0 1 2 3 4 5 6 7
Y 0 1 2 3 4 5 6 7
Z 5 5 5 5 5 5 5 5
碱基对/接头间隔区长度

图2D

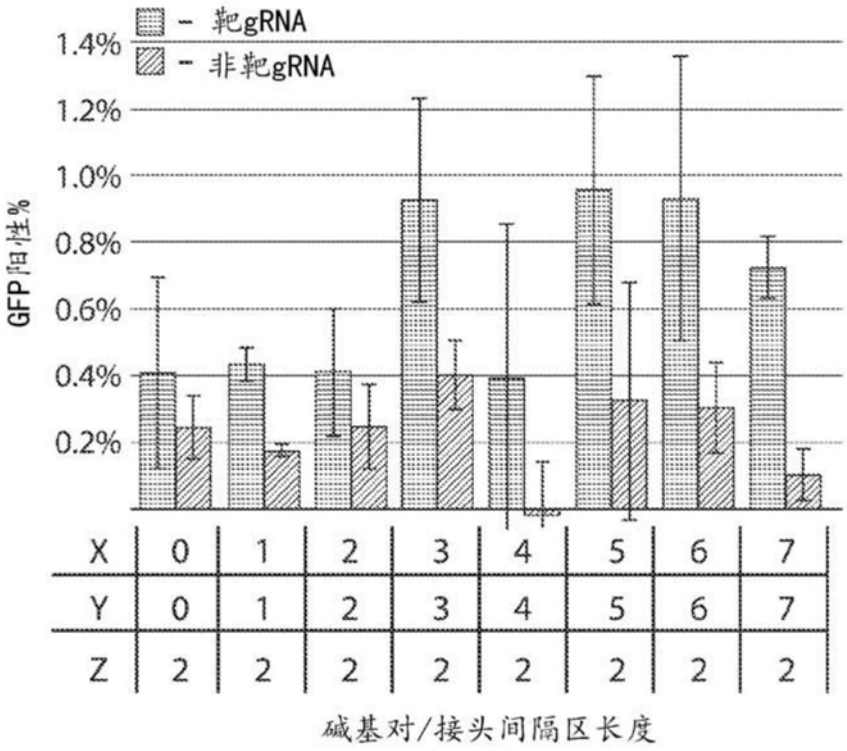

- 靶gRNA
- 非靶gRNA
GFP阳性%
1.4%
1.2%
1.0%
0.8%
0.6%
0.4%
0.2%
X 0 1 2 3 4 5 6 7
Y 0 1 2 3 4 5 6 7
Z 2 2 2 2 2 2 2 2
碱基对/接头间隔区长度

图2E

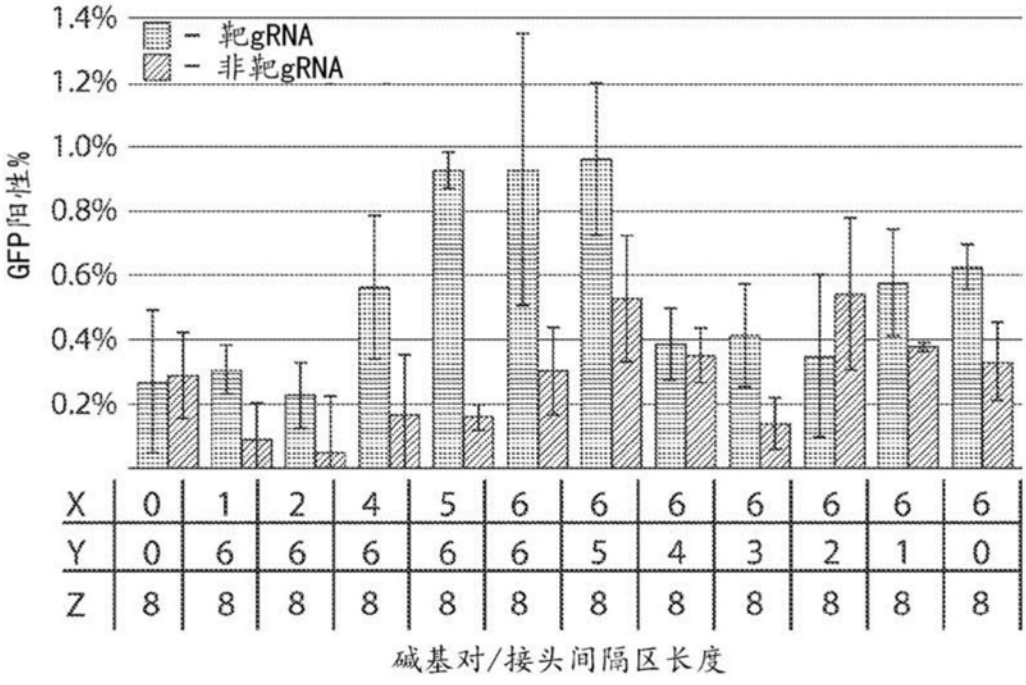

- 靶gRNA
- 非靶gRNA
GFP阳性%
1.4%
1.2%
1.0%
0.8%
0.6%
0.4%
0.2%
X 0 1 2 4 5 6 6 6 6 6 6 6
Y 0 6 6 6 6 6 5 4 3 2 1 0
Z 8 8 8 8 8 8 8 8 8 8 8 8
碱基对/接头间隔区长度

图2F

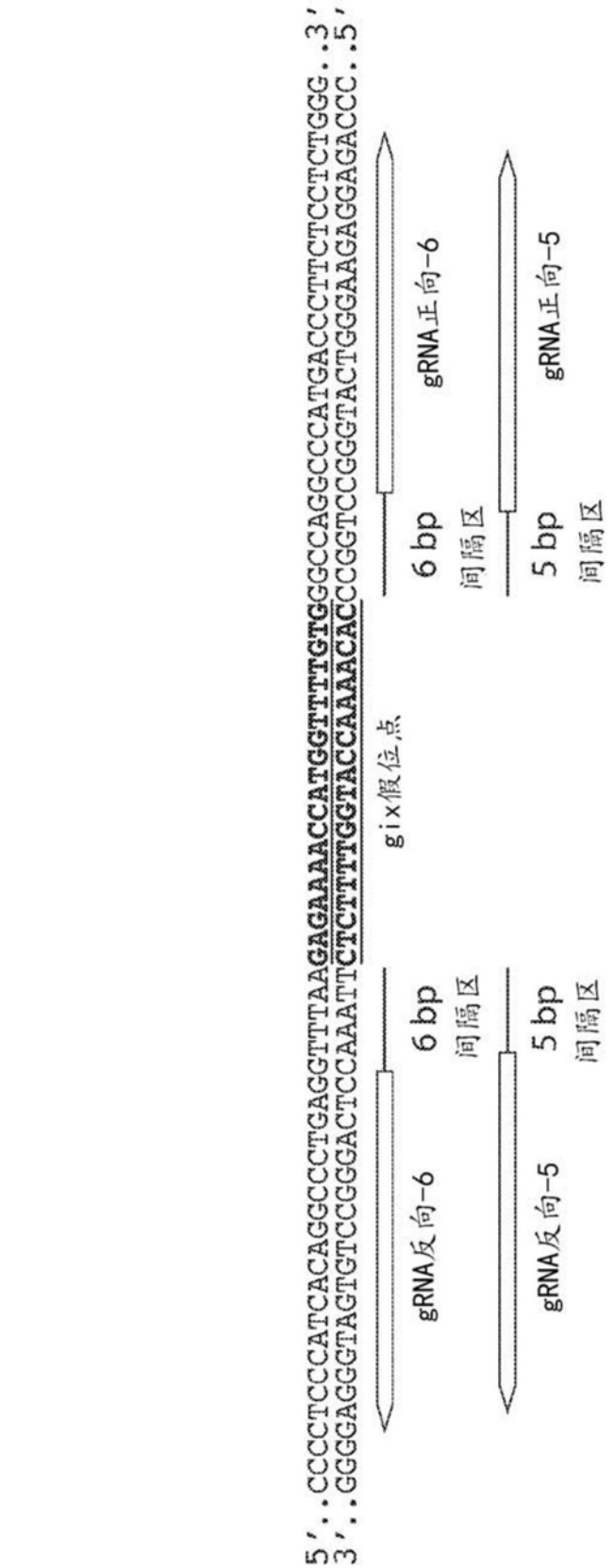

5′..CCCCTCCCATCACAGGCCCTGAGGTTTAAGAGAAAACCATGGTTTTTGTGGGCCAGGCCCATGACCCTTCTCCTCTGGG..3′
3′..GGGGAGGGTAGTGTCCGGGACTCCCAAATTCTCTTTTTGGTACCAAAACACCCGGTCCGGGTACTGGGAAGAGGACCCC..5′
gix假位点
gRNA反向-6
6 bp
间隔区
gRNA正向-6
6 bp
间隔区
gRNA反向-5
5 bp
间隔区
gRNA正向-5
5 bp
间隔区

图3A

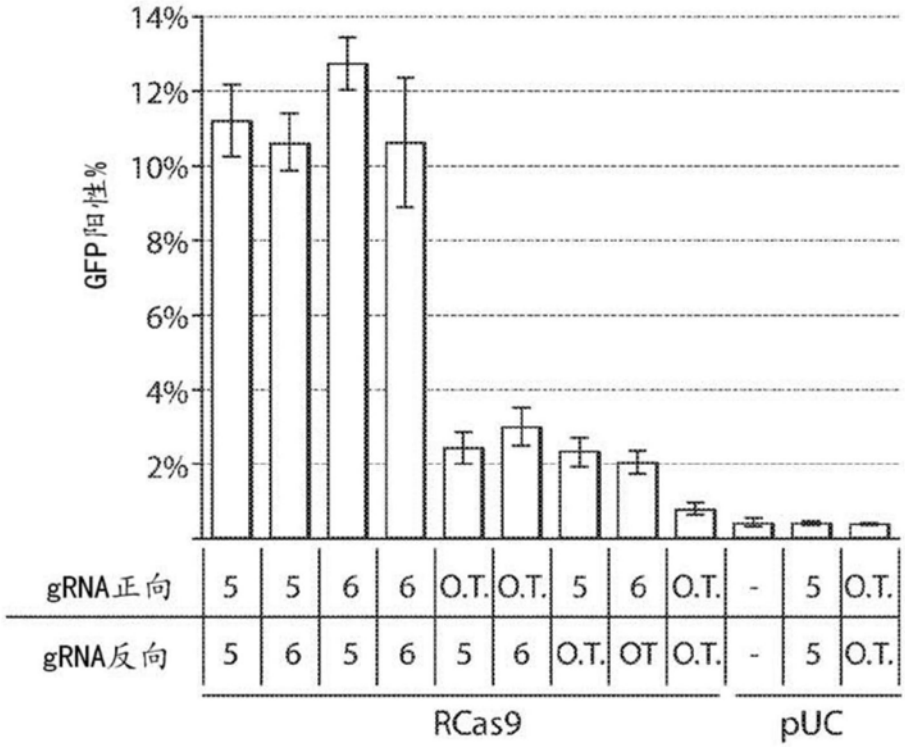

14%
12%
10%
8%
6%
4%
2%
GFP阳性%
gRNA正向 5 5 6 6 O.T. O.T. 5 6 O.T. - 5 O.T.
gRNA反向 5 6 5 6 5 6 O.T. OT O.T. - 5 O.T.
RCas9
pUC

图3B

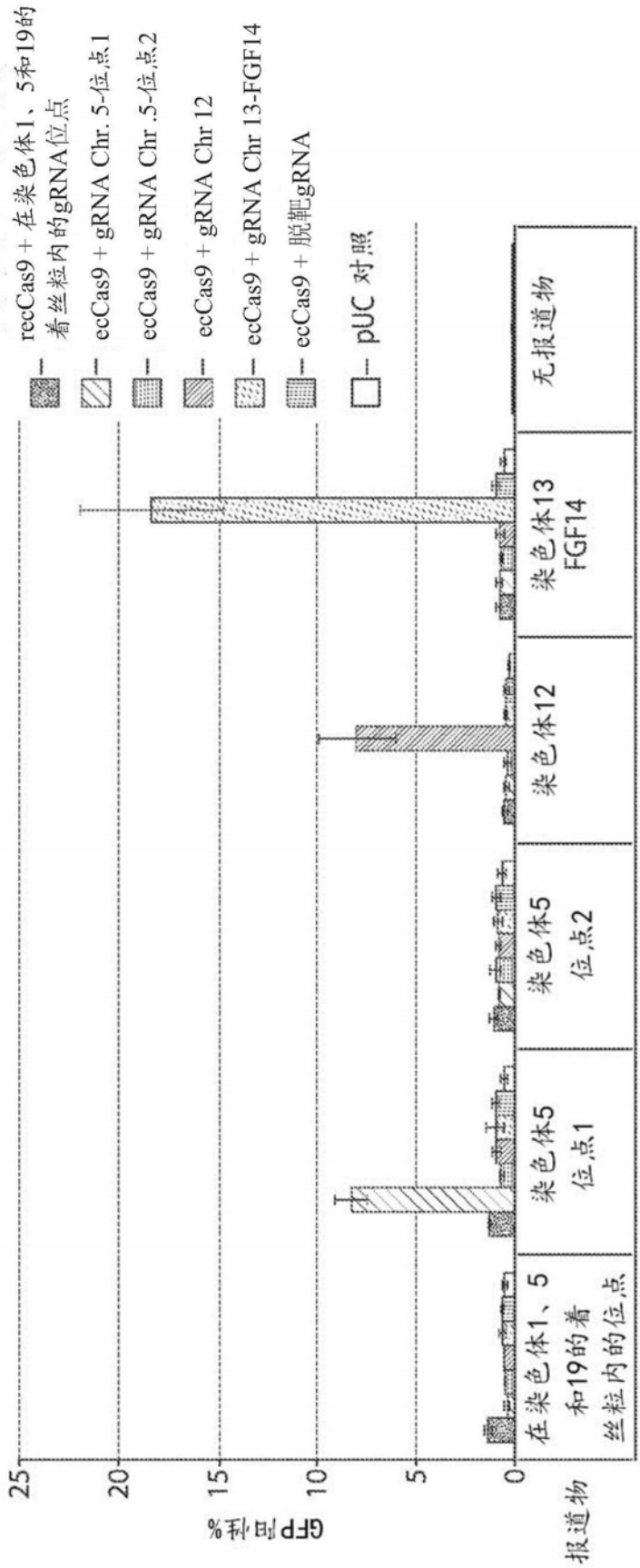

recCas9 + 在染色体1、5和19的着丝粒内的gRNA位点
ecCas9 + gRNA Chr. 5-位点1
ecCas9 + gRNA Chr .5-位点2
ecCas9 + gRNA Chr 12
ecCas9 + gRNA Chr 13-FGF14
ecCas9 + 脱靶gRNA
pUC 对照
GFP阳性%
25
20
15
10
5
0
报道物
在染色体1、5和19的着丝粒内的位点
染色体5 位点1
染色体5 位点2
染色体12
染色体13 FGF14
无报道物

图4A


1) 在293t细胞中共转
染质粒并温育3天
2) 质粒提取、RecBCD
消化和转化到
大肠杆菌中；选择Amp^R
3) 菌落的测序
gRNA
载体
RCas9
载体
报道物
Spec^R
Spec^R
Spec^R
Amp^R

图4B

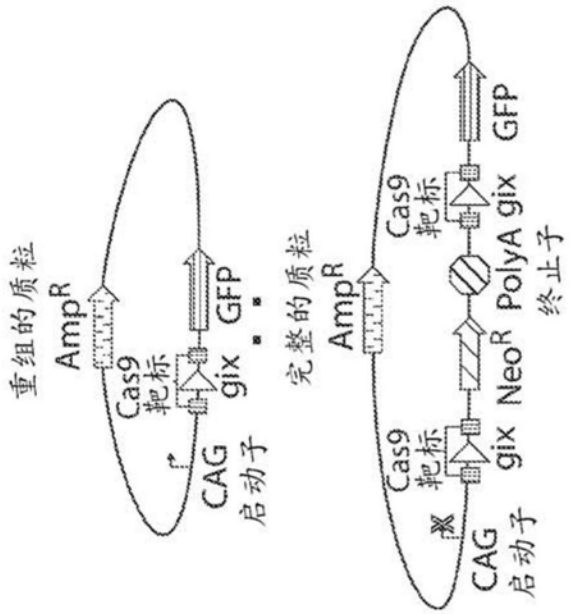

重组的质粒
Amp^R
Cas9
靶标
CAG
启动子
gix
GFP
完整的质粒
Amp^R
Cas9
靶标
CAG
启动子
gix
Neo^R
PolyA
终止子
gix
GFP

图4C

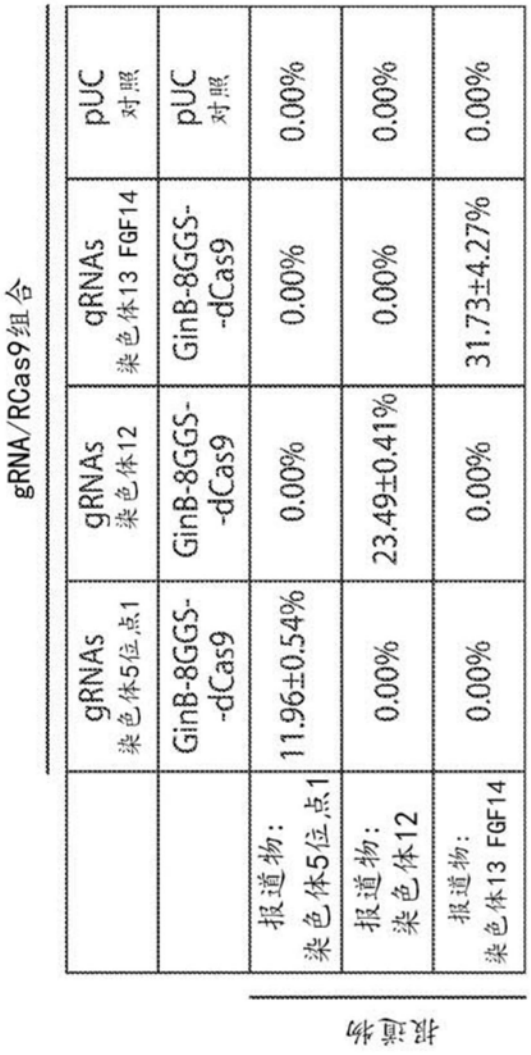

gRNA/RCas9组合
gRNAs 染色体5位点1
gRNAs 染色体12
gRNAs 染色体13 FGF14
pUC 对照
GinB-8GGS--dCas9
GinB-8GGS--dCas9
GinB-8GGS--dCas9
pUC 对照
报道物：染色体5位点1
11.96±0.54%
0.00%
0.00%
0.00%
报道物：染色体12
0.00%
23.49±0.41%
0.00%
0.00%
报道物：染色体13 FGF14
0.00%
0.00%
31.73±4.27%
0.00%
报道物

图4D

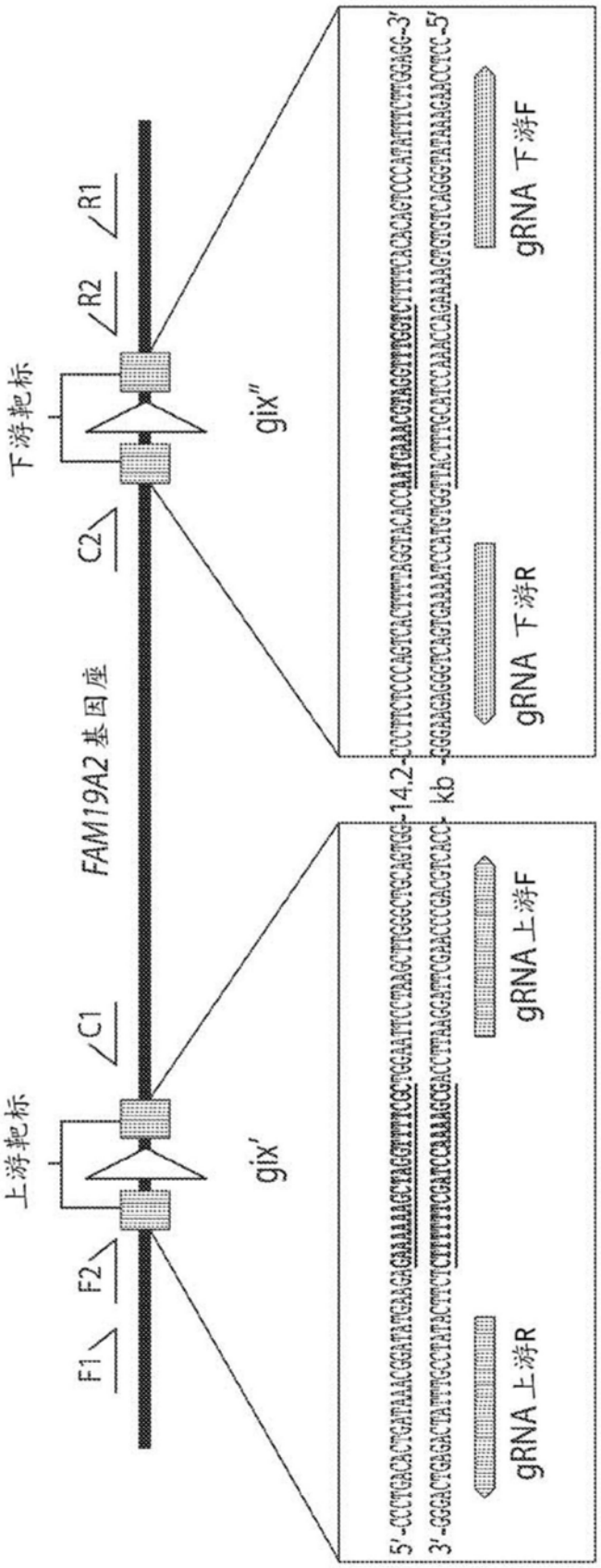

上游靶标
下游靶标
F1
F2
C1
C2
R1
R2
FAM19A2 基因座
gix′
gix″
gRNA 上游R
gRNA 上游F
gRNA 下游R
gRNA 下游F
14.2 kb

图5A

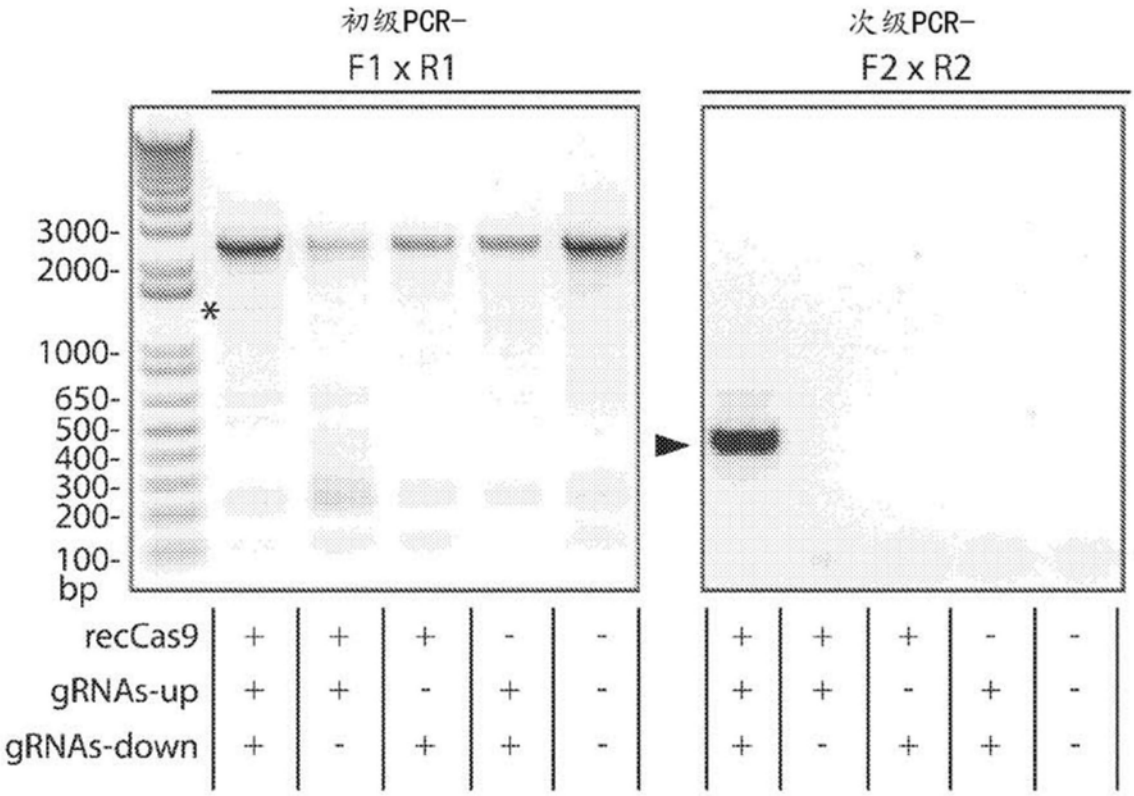

初级PCR-
F1 x R1
次级PCR-
F2 x R2
3000-
2000-
1000-
650-
500-
400-
300-
200-
100-
bp
*
recCas9 + + + - - + + + - -
gRNAs-up + + - + - + + - + -
gRNAs-down + - + + - + - + + -

图5B

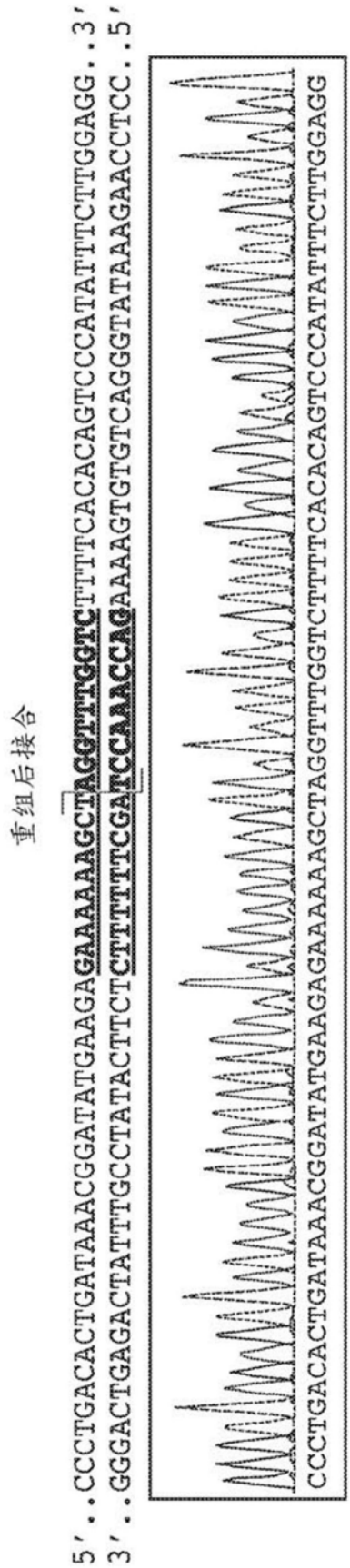

重组后接合
5′..CCCTGACACTGATAAACGGATATGAAGAGAGAAAAAGCTAGGTTTGGTCTTTTCACACAGTCCCATATTTCTTGGAGG..3′
3′..GGGACTGAGACTATTTGCCTATACTTCTCTTTTTTTCGATCCAAACCAGAAAAGTGTGTCAGGGTATAAAGAACCTCC..5′

图5C

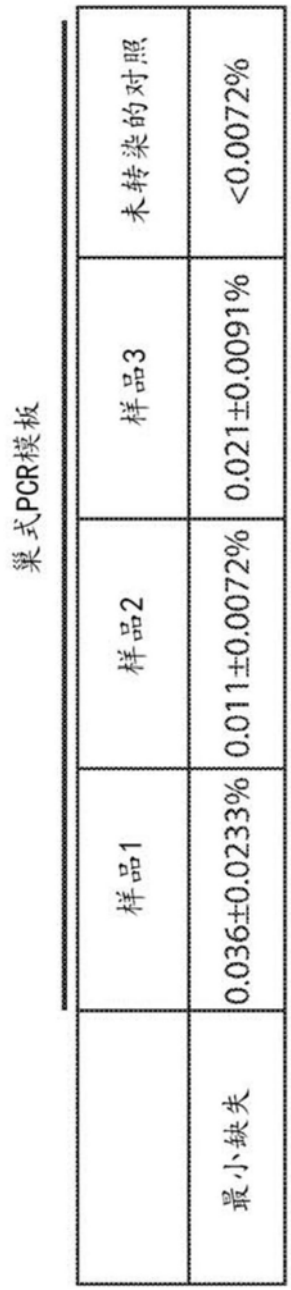

巢式PCR模板
样品1
样品2
样品3
未转染的对照
最小缺失
0.036±0.0233%
0.011±0.0072%
0.021±0.0091%
<0.0072%

图5D

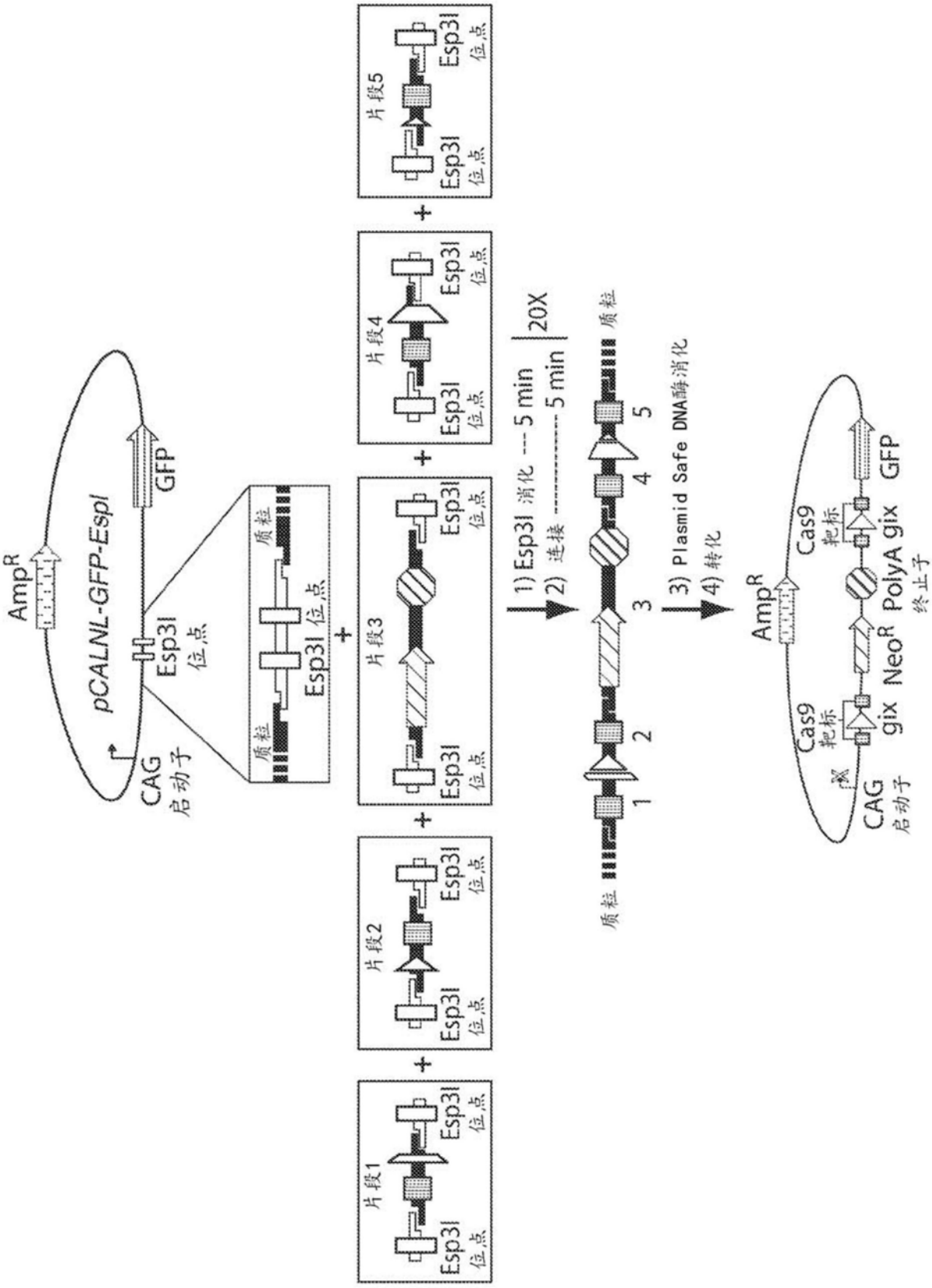

Amp^R
pCALNL-GFP-EspI
GFP
CAG
启动子
Esp3I
位点
质粒
片段1
片段2
片段3
片段4
片段5
1) Esp3I 消化 --- 5 min
2) 连接 ---------- 5 min
20X
3) Plasmid Safe DNA酶消化
4) 转化
Cas9
靶标
gix
Neo^R
PolyA
终止子

图6

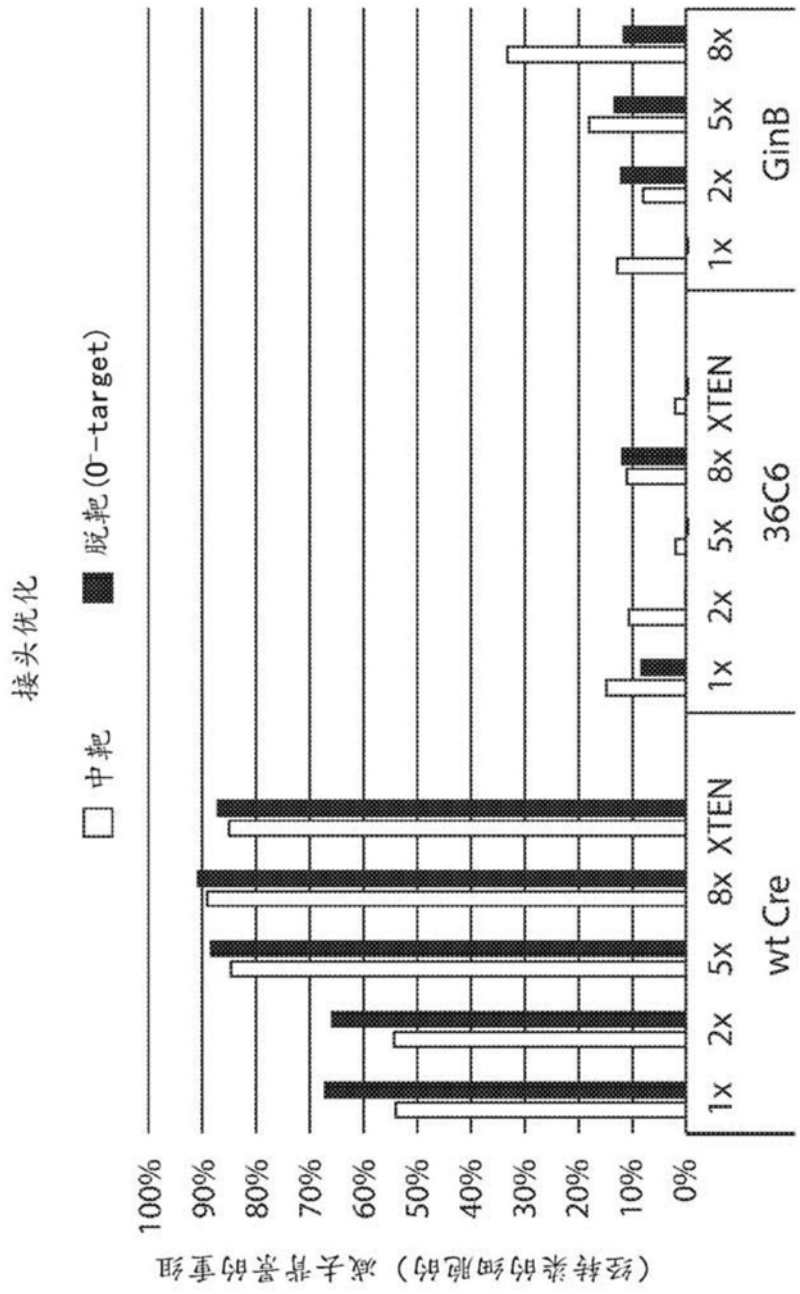

接头优化
□ 中靶
■ 脱靶(0-target)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1x 2x 5x 8x XTEN
wt Cre
1x 2x 5x 8x XTEN
36C6
1x 2x 5x 8x
GinB
(经转染的细胞的）减去背景的重组

图7A

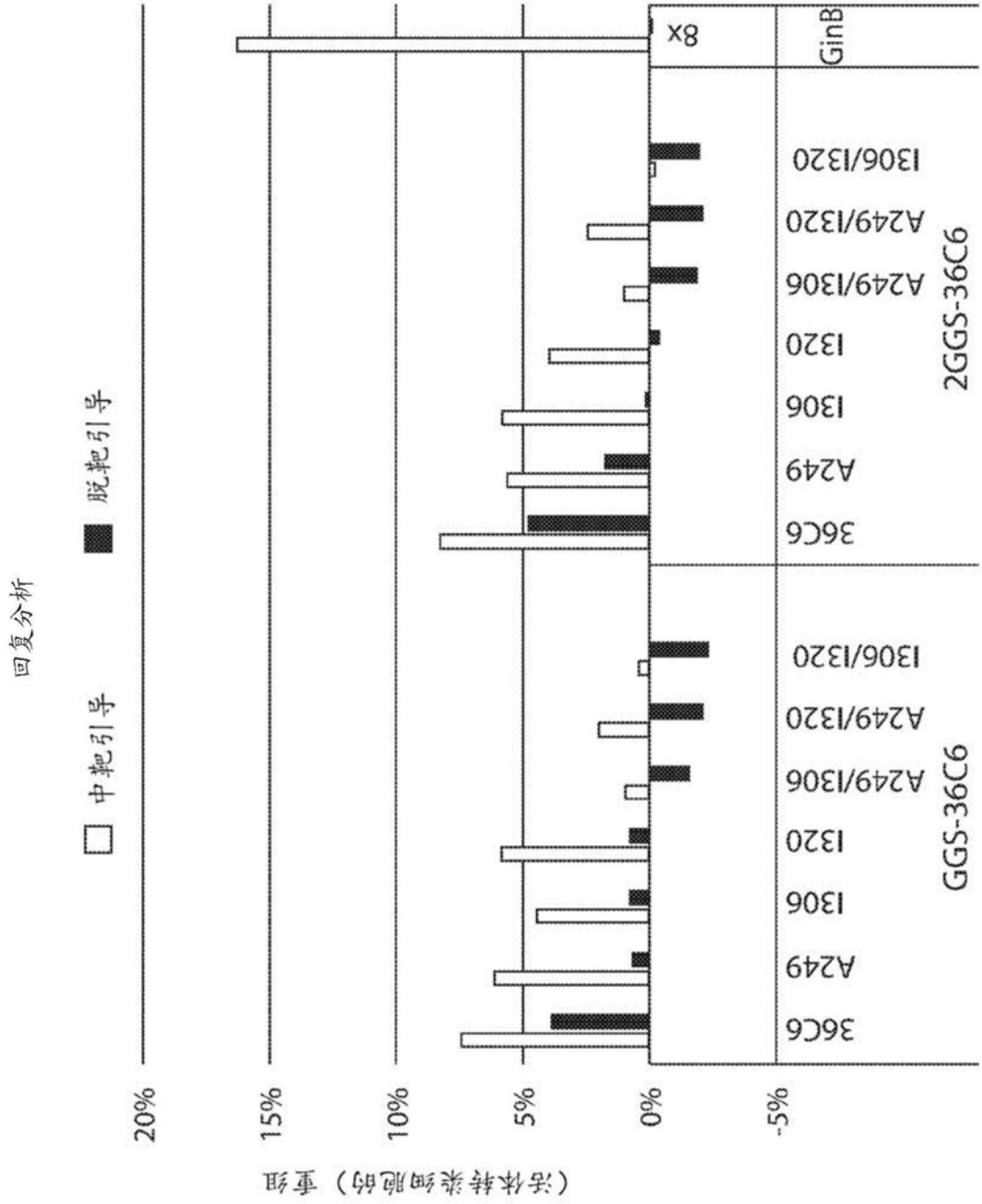

回复分析
□ 中靶引导
■ 脱靶引导
20%
15%
10%
5%
0%
-5%
(活体转染细胞的) 重组
GGS-36C6
36C6
A249
1306
1320
A249/1306
A249/1320
1306/1320
2GGS-36C6
36C6
A249
1306
1320
A249/1306
A249/1320
1306/1320
GinB
8x

图7B

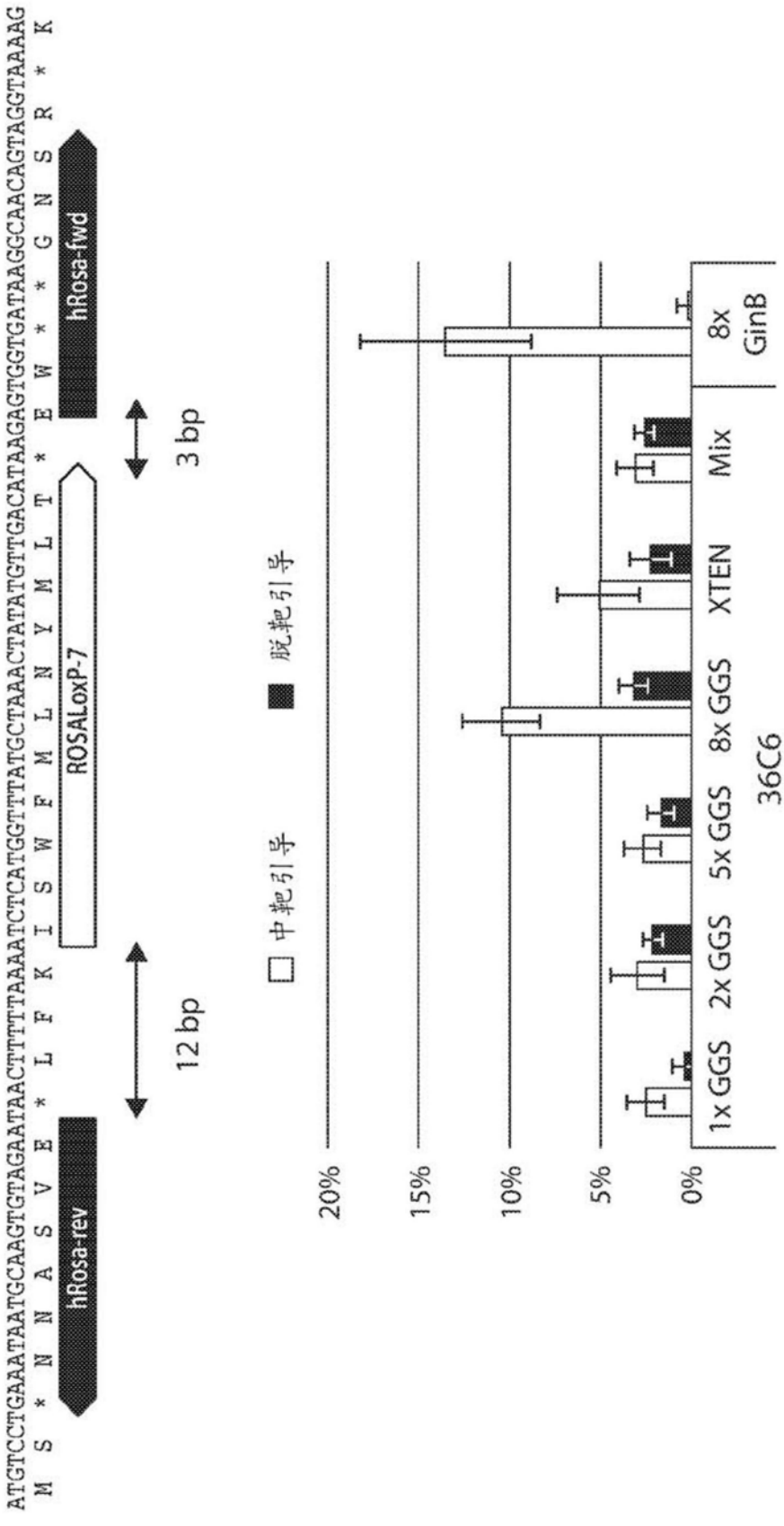

ATGTCCTGAAATAATGCAAGTGTAGAATAACTTTTTAAAATCTCATGGTTTATGCTAAACTATATGTTGACATAAGAGTGGTGATAAGGCAACAGTAGGTAAAAG
M S * N N A S V E * L F K I S W F M L N Y M L T * E W * * G N S R * K
hRosa-rev
ROSALoxP-7
hRosa-fwd
12 bp
3 bp
□ 中靶引导
■ 脱靶引导
20%
15%
10%
5%
0%
1x GGS
2x GGS
5x GGS
8x GGS
XTEN
Mix
8x GinB
36C6

图8

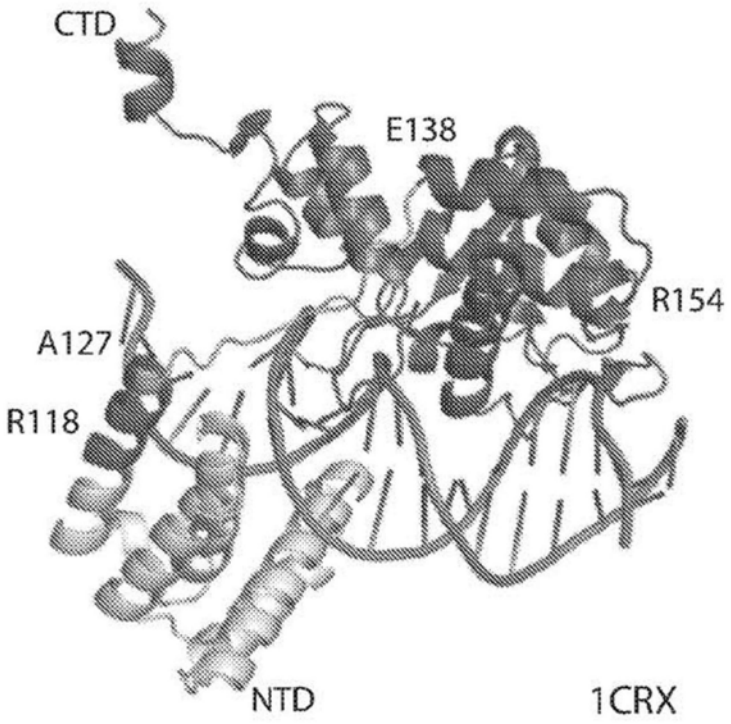

CTD
E138
R154
A127
R118
NTD
1CRX

图9A

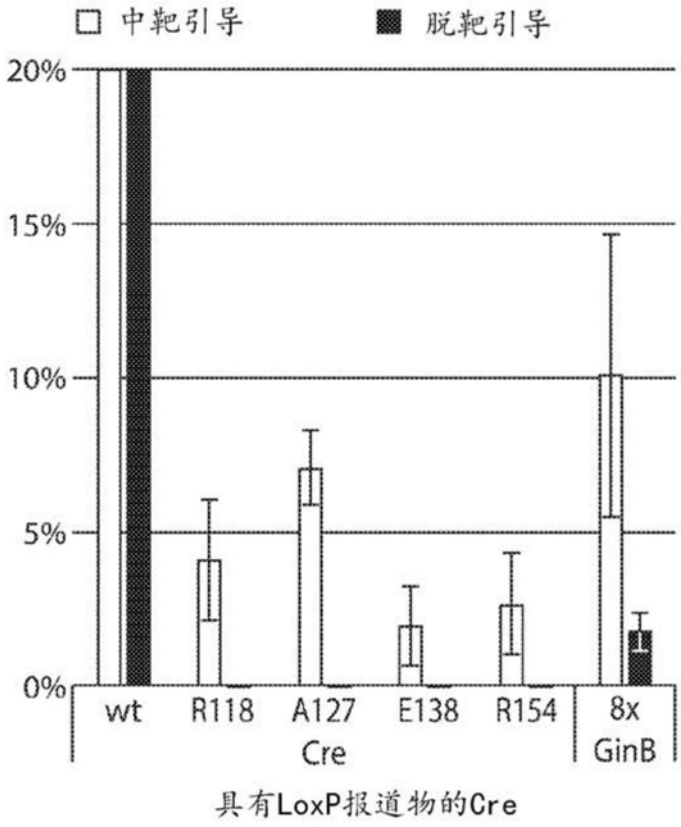

中靶引导
脱靶引导
20%
15%
10%
5%
0%
wt
R118
A127
E138
R154
8x GinB
Cre
具有LoxP报道物的Cre

图9B