(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 110872583 A
(43)申请公布日 2020.03.10

(21)申请号 201910837608.2

(22)申请日 2013.12.12

(30)优先权数据
61/736,527 2012.12.12 US
61/748,427 2013.01.02 US
61/758,468 2013.01.30 US
61/769,046 2013.02.25 US
61/791,409 2013.03.15 US
61/802,174 2013.03.15 US
61/806,375 2013.03.28 US
61/814,263 2013.04.20 US
61/819,803 2013.05.06 US
61/828,130 2013.05.28 US
61/836,123 2013.06.17 US
61/835,931 2013.06.17 US
61/847,537 2013.07.17 US

(62)分案原申请数据
201380072752.2 2013.12.12

(71)申请人 布罗德研究所有限公司
地址 美国马萨诸塞州
申请人 麻省理工学院
哈佛大学校长及研究员协会

(72)发明人 F.张 M.海登雷希 F.蓝
L.斯维奇

(74)专利代理机构 北京市金杜律师事务所 11256
代理人 陈文平 袁元

(51)Int.Cl.
C12N 9/22(2006.01)
C12N 15/113(2010.01)
C12N 15/864(2006.01)
C12N 15/867(2006.01)
C12N 5/10(2006.01)
C12N 15/90(2006.01)
A61K 31/7088(2006.01)
A61P 35/00(2006.01)
A61P 35/02(2006.01)
A61P 29/00(2006.01)
A61P 37/02(2006.01)
A61P 3/00(2006.01)
A61P 1/16(2006.01)
A61P 25/18(2006.01)
A61P 11/00(2006.01)
A61P 13/12(2006.01)
A61P 21/00(2006.01)
A61P 19/08(2006.01)
A61P 25/00(2006.01)
A61P 3/06(2006.01)

(续)

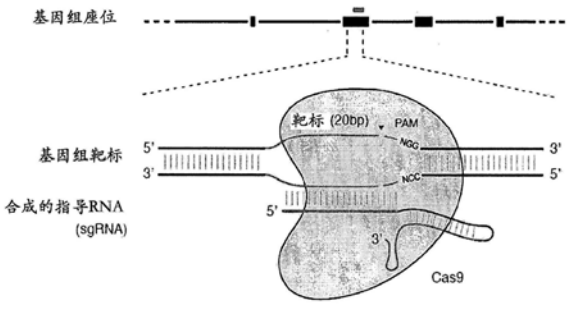
权利要求书11页 说明书237页
序列表179页 附图107页

(54)发明名称

用于序列操纵和治疗应用的系统、方法和组合物的递送、工程化和优化

(57)摘要

本发明提供了用于操纵序列和/或靶序列的活性的系统、方法、和组合物的递送、工程化和优化。提供了递送系统和被靶向作为用于递送的部位的组织或器官。还提供了其中的一些对CRISPR复合物的一种或多种组分进行编码的载体和载体系统、以及用于设计和使用这样的载体的方法。还提供了指导在真核细胞中形成CRISPR复合物的方法，以确保对于靶标识别的增强特异性和避免毒性，并且以编辑或修饰在感兴趣基因组座位中的靶位点以便改变或改善疾病或病症的状态。



CN 110872583 A

[转续页]

[接上页]

(51)Int.Cl.

A61P 7/00(2006.01)
A61P 31/12(2006.01)
A61P 31/14(2006.01)
A61P 31/18(2006.01)
A61P 9/00(2006.01)
A61P 19/10(2006.01)
A61P 25/28(2006.01)
A61P 25/14(2006.01)
A61P 25/16(2006.01)
A61P 27/02(2006.01)

1.一种工程化组合物,其包含:

Cas9蛋白;

CRISPR-Cas系统指导RNA,其包含:

(a)指导序列,所述指导序列能够在真核细胞中杂交到与原型间隔子邻近基序(PAM)相邻的靶序列上并指导CRISPR复合物与所述靶序列的序列特异性结合,和

(b)tracr配对序列,所述tracr配对序列能够与tracr序列杂交;和

包含所述tracr序列的tracrRNA。

2.一种工程化组合物,其包含:

Cas9蛋白;和

CRISPR-Cas系统嵌合RNA(chiRNA),其包含:

(a)指导序列,所述指导序列能够在真核细胞中杂交到与原型间隔子邻近基序(PAM)相邻的靶序列上并指导CRISPR复合物与所述靶序列的序列特异性结合,

(b)tracr配对序列,所述tracr配对序列能够与tracr序列杂交,和

(c)所述tracr序列,其具有30个或更多个核苷酸,

其中(a)、(b)和(c)以5'至3'方向排列。

3.一种工程化组合物,其包含一个或多个载体,所述一个或多个载体包含与编码Cas9蛋白的核苷酸序列、CRISPR-Cas系统指导RNA和包含tracr序列的tracrRNA可操作地连接的一个或多个调节元件,所述CRISPR-Cas系统指导RNA包含:

(a)指导序列,所述指导序列能够在真核细胞中杂交到与原型间隔子邻近基序(PAM)相邻的靶序列上并指导CRISPR复合物与所述靶序列的序列特异性结合,和

(b)tracr配对序列,所述tracr配对序列能够与所述tracr序列杂交。

4.一种工程化组合物,其包含一个或多个载体,所述一个或多个载体包含与编码Cas9蛋白和CRISPR-Cas系统嵌合RNA(chiRNA)的一个或多个核苷酸序列可操作地连接的一个或多个调节元件,所述chiRNA包含:

(a)指导序列,所述指导序列能够在真核细胞中杂交到与原型间隔子邻近基序(PAM)相邻的靶序列上并指导CRISPR复合物与所述靶序列的序列特异性结合,

(b)tracr配对序列,所述tracr配对序列能够与tracr序列杂交,和

(c)所述tracr序列,其具有30个或更多个核苷酸,

其中(a)、(b)和(c)以5'至3'方向排列。

5.根据权利要求1-4中任一项所述的组合物,其还包含修复模板。

6.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白与一个或多个核定位序列(NLS)连接或融合。

7.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白与两个或更多个核定位序列(NLS)连接或融合。

8.根据权利要求6所述的组合物,其中至少一个NLS位于所述Cas9蛋白的氨基端处或附近,并且至少一个NLS位于所述Cas9蛋白的羧基端处或附近。

9.根据权利要求3或4所述的组合物,其中所述编码Cas9蛋白的核苷酸序列经密码子优化以在真核细胞中表达。

10.根据权利要求1-4中任一项所述的组合物,其中所述指导序列的长度为至少15个核

苷酸。

11.根据权利要求1-4中任一项所述的组合物,其中所述tracr序列的长度为至少40个核苷酸。

12.根据权利要求1-4中任一项所述的组合物,其中所述tracr序列的长度为至少50个核苷酸。

13.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白是核酸酶。

14.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白是在RuvC结构域中具有突变或在HNH结构域中具有突变的切口酶。

15.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白是在RuvC结构域中具有突变且在HNH结构域中具有突变的无催化活性的Cas9。

16.根据权利要求14所述的组合物,其中参照化脓链球菌Cas9的氨基酸编号,所述Cas9蛋白包含选自D10A,H840A,N854A和N863A的一个或多个突变。

17.根据权利要求3或4所述的组合物,其中所有组分都位于一个载体上。

18.根据权利要求3或4所述的组合物,其中所述载体是双顺反子载体。

19.根据权利要求3或4所述的组合物,其中所述一个或多个载体是病毒载体。

20.根据权利要求1或2所述的组合物,其中所述Cas9蛋白质是翻译自mRNA。

21.根据权利要求1或2所述的组合物,其中所述Cas9蛋白以蛋白质的形式递送。

22.根据权利要求19所述的组合物,其中所述病毒载体是腺相关病毒(AAV)载体。

23.根据权利要求19所述的组合物,其中所述病毒载体是慢病毒载体。

24.根据权利要求1-4中任一项所述的组合物,其中所述组合物配置用于单一或共同递送。

25.根据权利要求1-4中任一项所述的组合物,其中所述组合物配置用于单次施用。

26.根据权利要求1-4中任一项所述的组合物,其中所述组合物配置为适于用颗粒递送的递送组合物。

27.根据权利要求26所述的组合物,其中所述颗粒是脂质纳米颗粒(LNP)。

28.根据权利要求1-4中任一项所述的组合物,其中所述组合物配置为适于通过电穿孔方法递送的递送组合物。

29.根据权利要求3或4所述的组合物,其中至少一个调节元件是用于普遍表达或组织特异性表达的启动子。

30.根据权利要求29所述的组合物,其中所述启动子与编码所述Cas9蛋白的核苷酸序列可操作地连接。

31.根据权利要求1-4中任一项所述的组合物,其中提供两种或更多种不同的指导序列用于多元靶向。

32.根据权利要求1-4中任一项所述的组合物,其配制用于单剂量施用。

33.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白与至少一个功能结构域连接或融合。

34.根据权利要求33所述的组合物,其中所述功能结构域具有甲基酶活性,脱甲基酶活性,转录激活活性,转录阻抑活性,转录释放因子活性,组蛋白修饰活性,RNA切割活性,核定位活性或核酸结合活性。

35.根据权利要求33所述的组合物,其中所述功能结构域是转录激活结构域或转录阻抑结构域。

36.根据权利要求33所述的组合物,其中所述功能结构域是KRAB或VP64。

37.根据权利要求1-4中任一项所述的组合物,其中所述工程化组合物包含或编码两种或更多种chiRNA或指导RNA,其各自能够在所述真核细胞中与不同靶序列杂交。

38.根据权利要求37所述的组合物,其中所述Cas9蛋白是切口酶,并且所述切口酶与所述两种或更多种chiRNA或指导RNA组合能够产生具有限定的突出端的双链DNA断裂。

39.根据权利要求1-4中任一项所述的组合物,其中所述工程化组合物还包含重组模板。

40.根据权利要求39所述的组合物,其中所述重组模板通过同源定向修复(HDR)插入。

41.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白衍生自链球菌或葡萄球菌。

42.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白是化脓链球菌Cas9。

43.根据权利要求1-4中任一项所述的组合物,其中所述Cas9蛋白是金黄色葡萄球菌Cas9。

44.根据权利要求1-4中任一项所述的工程化组合物在制备用于治疗疾病或障碍的药物中的用途。

45.根据权利要求44所述的用途,其中所述治疗导致一个或多个细胞中的遗传变化。

46.根据权利要求44所述的用途,其中所述治疗导致一种或多种缺陷基因型的修正。

47.根据权利要求44所述的用途,其中所述治疗导致改善的表型。

48.根据权利要求44所述的用途,其中所述治疗包括治疗或抑制由感兴趣的基因组基因座中的基因缺陷引起的病症。

49.根据权利要求44所述的用途,其中所述Cas9蛋白衍生自链球菌或葡萄球菌。

50.根据权利要求44所述的用途,其中所述Cas9蛋白是化脓链球菌Cas9。

51.根据权利要求44所述的用途,其中所述Cas9蛋白是金黄色葡萄球菌Cas9。

52.根据权利要求44所述的用途,其中所述治疗包括离体疗法。

53.根据权利要求44所述的用途,其中所述治疗包括体内疗法。

54.根据权利要求44所述的用途,其中所述治疗包括施用所述工程化组合物或所述药物以治疗疾病或障碍。

55.根据权利要求44所述的用途,其中所述治疗包括将所述工程化组合物或所述药物施用于人。

56.根据权利要求44所述的用途,其中所述治疗包括治疗细胞失调和肿瘤疾病或障碍,炎症和免疫相关疾病或障碍,代谢疾病或障碍,肝疾病或障碍,肺疾病或障碍,肾疾病或障碍,蛋白质疾病或障碍,肌肉疾病或障碍,骨骼疾病或障碍,神经疾病或障碍,或神经元疾病或障碍。

57.根据权利要求44所述的用途,其中所述疾病或障碍与胆固醇有关。

58.根据权利要求44所述的用途,其中所述组合物靶向枯草杆菌蛋白酶丝氨酸蛋白酶家族中的基因。

59.根据权利要求58所述的用途,其中所述枯草杆菌蛋白酶丝氨酸蛋白酶家族中的所

述基因是前蛋白转化酶枯草杆菌蛋白酶/kexin 9型(PCSK9)。

60.根据权利要求59所述的用途,其中所述指导RNA或嵌合RNA靶向PCSK9基因组基因座。

61.根据权利要求60所述的用途,其中所述复合物切口或切割PCSK9基因组基因座的DNA。

62.根据权利要求57所述的用途,其中导致或引起高胆固醇血症的突变被校正。

63.根据权利要求62所述的用途,其中靶向导致高胆固醇血症或引起高胆固醇血症相关症状的基因的所述组合物被静脉内递送。

64.根据权利要求57所述的用途,其中所述组合物通过颗粒来静脉内递送。

65.根据权利要求64所述的用途,其中所述颗粒是脂质颗粒。

66.根据权利要求65所述的用途,其中所述脂质颗粒是LNP。

67.根据权利要求64所述的用途,其中所述颗粒是AAV颗粒。

68.根据权利要求67所述的用途,其中所述AAV颗粒包含用于靶向肝脏的AAV8。

69.根据权利要求44所述的用途,其中所述疾病或障碍是肺的。

70.根据权利要求69所述的用途,其中所述疾病或障碍是囊性纤维化。

71.根据权利要求70所述的用途,其中所述组合物包含含有F5Q8残基的修复模板。

72.根据权利要求70所述的用途,其中所述指导RNA或嵌合RNA靶向CFTR基因组基因座。

73.根据权利要求72所述的用途,其中所述复合物切口或切割CFTR基因组基因座的DNA。

74.根据权利要求72所述的用途,其中导致囊性纤维化或引起囊性纤维化相关症状的缺失被修正。

75.根据权利要求72所述的用途,其中靶向导致囊性纤维化或引起囊性纤维化相关症状的基因的所述组合物通过雾化递送来递送。

76.根据权利要求75所述的用途,其中通过雾化递送来递送的所述组合物是颗粒。

77.根据权利要求76所述的用途,其中所述颗粒是脂质颗粒。

78.根据权利要求77所述的用途,其中所述脂质颗粒是LNP。

79.根据权利要求76所述的用途,其中所述颗粒是AAV颗粒。

80.根据权利要求79所述的用途,其中所述AAV颗粒包含用于靶向肺上皮的AAV1,AAV5,AAV6或AAV9。

81.根据权利要求44所述的用途,其中所述疾病或障碍是蛋白质来源的淀粉样变性。

82.根据权利要求81所述的用途,其中所述疾病或障碍是α-1抗胰蛋白酶(ATT)缺陷。

83.根据权利要求82所述的用途,其中所述组合物靶向ATT家族中的基因。

84.根据权利要求82所述的用途,其中所述指导RNA或嵌合RNA靶向ATT基因组基因座。

85.根据权利要求85所述的用途,其中所述复合物切口或切割ATT基因组基因座的DNA。

86.根据权利要求82所述的用途,其中导致或引起ATT缺陷的突变被校正。

87.根据权利要求81所述的用途,其中所述疾病或障碍是TTR淀粉样变性(ATTR)。

88.根据权利要求87所述的用途,其中所述指导RNA或嵌合RNA靶向TTR基因组基因座,并且其中所述复合物切口或切割TTR基因组基因座的DNA。

89.根据权利要求81所述的用途,其中导致与由TTR中的常染色体显性突变引起的家族

性淀粉样变性多发性神经病(FAP)和家族性淀粉样变性心肌病(FAC)及由野生型TTR引起的老年系统性淀粉样变性(SSA)相关的表型的突变被校正。

90.根据权利要求81所述的用途,其中靶向导致淀粉样变性疾病的基因的所述组合物被静脉内递送。

91.根据权利要求81所述的用途,其中靶向导致淀粉样变性疾病的基因的所述组合物通过雾化递送来递送。

92.根据权利要求81所述的用途,其中靶向导致淀粉样变性疾病的基因的所述组合物通过颗粒来递送。

93.根据权利要求92所述的用途,其中所述颗粒是脂质颗粒。

94.根据权利要求93所述的用途,其中所述脂质颗粒是LNP。

95.根据权利要求92所述的用途,其中所述颗粒是AAV颗粒。

96.根据权利要求44所述的用途,其中所述疾病或障碍是眼部遗传疾病。

97.根据权利要求96所述的用途,其中所述眼部遗传疾病是视网膜疾病。

98.根据权利要求97所述的用途,其中所述视网膜疾病是遗传性视网膜疾病。

99.根据权利要求96所述的用途,其中所述眼部遗传疾病是LCA。

100.根据权利要求96所述的用途,其中所述眼部遗传疾病是年龄相关性黄斑变性。

101.根据权利要求96所述的用途,其中所述眼部遗传疾病是Usher综合征。

102.根据权利要求96所述的用途,其中所述工程化组合物是多元金黄色葡萄球菌Cas9(SaCas9)酶系统,其进一步包含多个指导序列和单个tracr序列。

103.根据权利要求44所述的用途,其中所述疾病或障碍是肝疾病或障碍。

104.根据权利要求103所述的用途,其中所述肝疾病或障碍是由受试者的基因组基因座中的肝靶序列的缺陷引起,所述肝靶序列的缺陷在肝细胞中表达。

105.根据权利要求104所述的用途,其中所述肝靶序列表达α-1抗胰蛋白酶(A1AT),载脂蛋白B(ApoB)或低密度脂蛋白20受体(LDL-R)。

106.根据权利要求44所述的用途,其中所述疾病或障碍是代谢疾病或障碍。

107.根据权利要求106所述的用途,其中所述代谢疾病或障碍由受试者的基因组基因座中的糖原贮积相关序列的缺陷引起。

108.根据权利要求44所述的用途,其中所述疾病或障碍是血液或凝血疾病或障碍。

109.根据权利要求108所述的用途,用于治疗或抑制由受试者的基因组基因座中的BCL11A相关序列的缺陷引起的病症。

110.根据权利要求108所述的用途,用于治疗或抑制由受试者的基因组基因座中的血红蛋白相关序列的缺陷引起的病症。

111.根据权利要求38所述的用途,其中所述疾病或障碍是病毒性疾病或障碍。

112.根据权利要求111所述的用途,用于治疗或抑制由受试者的基因组基因座中的丙型肝炎病毒(HCV)相关序列的缺陷引起的病症。

113.根据权利要求111所述的用途,用于治疗或抑制由受试者的基因组基因座中的乙型肝炎病毒(HBV)相关序列的缺陷引起的病症。

114.根据权利要求109所述的用途,用于治疗或抑制由受试者的基因组基因座中的人免疫缺陷病毒(HIV)相关序列的缺陷引起的病症。

115.根据权利要求114所述的用途,其中所述HIV是HIV-1。

116.根据权利要求114所述的用途,其中所述治疗引入保护性突变或序列。

117.根据权利要求116所述的用途,其中所述保护性突变或序列是CCR5基因座中的突变以预防HIV感染。

118.根据权利要求114所述的用途,其中所述治疗靶向病毒的一种或多种变体。

119.根据权利要求118所述的用途,其中靶标是HIV-1变体中的一种或多种。

120.根据权利要求119所述的用途,其中所述HIV-1变体中的一种或多种包含以下序列中的一个或多个:CACTGCTTAAGCCTCGCTCGAGG;TCACCAGCAATATTCGCTCGAGG;CACCAGCAATATTCCGCTCGAGG;TAGCAACAGACATACGCTCGAGG;GGGCAGTAGTAATACGCTCGAGG;和CCAATTCCCATACATTATTGTAC。

121.根据权利要求111所述的用途,其中所述用途是破坏一个或多个病毒片段或基因。

122.根据权利要求44所述的用途,其中所述疾病或障碍是三核苷酸重复障碍。

123.根据权利要求122所述的用途,其中所述三核苷酸重复障碍是多聚谷氨酰胺(polyQ)障碍或非多聚谷氨酰胺障碍。

124.根据权利要求44所述的用途,其中所述疾病或障碍是杜兴氏肌肉营养不良症(DMD)。

125.根据权利要求44所述的用途,其中所述疾病或障碍是脊髓性肌萎缩症(SMA)。

126.根据权利要求44所述的用途,其中所述疾病或障碍是心脏疾病或失调。

127.根据权利要求126所述的用途,其中所述AAV载体包含用于靶向心脏组织的AAV4。

128.根据权利要求44所述的用途,其中所述药物包含用所述组合物离体修饰的干细胞或iPS细胞。

129.根据权利要求128所述的用途,其中所述干细胞是HSC。

130.根据权利要求128所述的用途,其中所述离体治疗在干细胞或iPS细胞上进行。

131.根据权利要求44所述的用途,其中至少一个调节元件是用于普遍表达或组织特异性表达的启动子。

132.根据权利要求131所述的用途,其中所述启动子用于普遍表达,并且包括CMV启动子,CAG启动子,CBh启动子,PGK启动子,SV40启动子或铁蛋白重链或轻链启动子。

133.根据权利要求131所述的用途,其中所述启动子用于组织特异性表达,并且指导在包含肌肉,神经元,骨,皮肤,血液,肝脏,胰腺或淋巴细胞的组织中的表达。

134.根据权利要求44所述的用途,其中所述AAV载体包括AAV1,AAV2,AAV3,AAV4,AAV5,AAV6,AAV8或AAV9或其任何组合。

135.根据权利要求44所述的用途,其中所述AAV载体包含用于靶向大脑或神经元细胞的AAV血清型1、2、5或杂合或衣壳AAV1、AAV2、AAV5或其任何组合。

136.根据权利要求44所述的用途,其中免疫细胞被分离,用权利要求1至4中任一项所述的组合物转导,并重新引入受试者。

137.根据权利要求136所述的用途,其中所述免疫细胞被自体重新引入。

138.根据权利要求136所述的用途,其中所述免疫细胞是单核细胞或淋巴细胞。

139.根据权利要求136所述的用途,其中转导的免疫细胞用于治疗癌症。

140.根据权利要求32所述的组合物,其中所述启动子用于普遍表达,并且包括CMV启动

子,CAG启动子,CBh启动子,PGK启动子,SV40启动子或铁蛋白重链或轻链启动子。

141.根据权利要求32所述的组合物,其中所述启动子用于组织特异性表达,并且指导在包含肌肉,神经元,骨,皮肤,血液,肝脏,胰腺或淋巴细胞的组织中的表达。

142.根据权利要求44所述的用途,其中所述疾病或障碍是选自以下的血液和凝血疾病或障碍:贫血,裸淋巴细胞综合征,出血障碍,因子H和因子H样1,因子V和因子VIII,因子VII缺乏,因子X缺乏,因子XI缺乏,因子XII缺乏,因子XIIIA缺乏,因子XIIIB缺乏,范科尼贫血,噬血细胞淋巴组织细胞增生障碍,血友病A,血友病B,出血性障碍,白细胞缺乏或障碍,镰状细胞贫血,或地中海贫血。

143.根据权利要求56所述的用途,其中所述疾病或障碍是选自B细胞非霍奇金淋巴瘤或白血病的细胞调节失常和肿瘤疾病或障碍。

144.根据权利要求56所述的用途,其中所述疾病或障碍是选自AIDS,自身免疫性淋巴增生综合征,联合免疫缺陷,HIV-1感染,HIV易感性或感染,免疫缺陷,炎症或重症联合免疫缺陷(SCID)的炎症和免疫相关疾病或障碍。

145.根据权利要求56所述的用途,其中所述疾病或障碍是选自以下的代谢疾病或障碍,肝疾病或障碍,肾疾病或障碍,或蛋白质疾病或障碍:淀粉样蛋白神经病变,淀粉样变性,TTR淀粉样变性,囊性纤维化,糖原贮积病,肝腺瘤,肝衰竭,神经障碍,肝脂肪酶缺乏,肝母细胞瘤,癌症或癌,髓质囊肾病,苯丙酮尿症,多囊肾或肝病。

146.根据权利要求145所述的用途,其中所述TTR淀粉样变性是ATTR。

147.根据权利要求56所述的用途,其中所述疾病或障碍是选自肌营养不良,骨质疏松症或肌肉萎缩的肌肉或骨骼疾病或障碍。

148.根据权利要求56所述的用途,其中所述疾病或障碍是选自以下的神经或神经元疾病或障碍:ALS;阿尔茨海默病,自闭症,脆性X综合征,亨廷顿病和亨廷顿病样障碍,帕金森病,精神分裂症,分泌酶相关障碍,三核苷酸重复障碍,肯尼迪病,弗里德赖希共济失调,马查多-约瑟夫病,脊髓小脑共济失调,强直性肌营养不良,齿状核红核苍白球丘脑底核萎缩(DRPLA)疾病,或creb BP整体不稳定性。

149.根据权利要求44所述的用途,其中所述疾病或失调是选自年龄相关性黄斑变性,角膜混浊和营养不良,角膜先天性扁平,青光眼,莱伯先天性黑朦,色素性视网膜炎,或黄斑营养不良的眼部疾病或障碍。

150.根据权利要求44所述的用途,其中所述治疗包括治疗有丝分裂后障碍。

151.根据权利要求150所述的用途,其中所述有丝分裂后障碍包括神经元障碍。

152.根据权利要求44所述的用途,其中所述治疗包括:治疗代谢障碍,淀粉样变性或蛋白聚积相关疾病,由基因突变或易位引起的细胞转化,基因突变的负显性效应,或潜伏病毒感染;或靶向包含SERPINA1(丝氨酸蛋白酶抑制蛋白肽酶抑制剂,进化枝A(α-1抗蛋白酶,抗胰蛋白酶),成员1)的染色体序列。

153.一种真核细胞,其包含由以下形成的CRISPR复合物:

(A)包含至少一个核定位序列(NLS)的Cas9蛋白,和CRISPR-Cas系统嵌合RNA(chiRNA),所述嵌合RNA包含(a)能够与真核细胞的基因组中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr序列杂交的tracr配对序列,和(c)所述tracr序列,其中(a)、(b)和(c)以5'至3'方向排列,

或

(B)包含至少一个NLS的Cas9蛋白,CRISPR-Cas系统指导RNA,和包含所述tracr序列的tracrRNA,所述指导RNA包含(a)能够与真核细胞的基因组中与PAM相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,和(b)能够与tracr序列杂交的tracr配对序列。

154.一种工程化非人多细胞生物体,其包含权利要求153所述的真核细胞。

155.一种工程化Cas9蛋白,其与至少两个核定位序列(NLS)融合,其中至少一个NLS位于所述Cas9蛋白的氨基端处或附近,并且至少一个NLS位于所述Cas9蛋白的羧基端处或附近。

156.一种工程化CRISPR-Cas系统嵌合RNA(chiRNA),其包含:(a)能够与真核细胞的基因组中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr序列杂交的tracr配对序列,和(c)所述tracr序列,所述tracr序列的长度为至少40个核苷酸。

157.一种工程化CRISPR-Cas系统嵌合RNA(chiRNA),其包含:(a)能够与真核细胞的基因组中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr序列杂交的tracr配对序列,和(c)所述tracr序列,所述tracr序列的长度为至少50个核苷酸。

158.一种工程化CRISPR-Cas系统嵌合RNA(chiRNA),其能够与嗜热链球菌Cas9形成CRISPR复合物,所述嵌合RNA包含:(a)能够与靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr序列杂交的tracr配对序列,和(c)所述tracr序列,所述tracr序列衍生自嗜热链球菌tracrRNA。

159.一种工程化CRISPR-Cas系统嵌合RNA(chiRNA),其能够与空肠弯曲杆菌Cas9形成CRISPR复合物,所述嵌合RNA包含:(a)能够与靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr序列杂交的tracr配对序列,和(c)所述tracr序列,所述tracr序列衍生自空肠弯曲杆菌tracrRNA。

160.一种工程化载体,其编码:

I.包含至少一个核定位序列(NLS)的Cas9蛋白,和CRISPR-Cas系统嵌合RNA(chiRNA),所述嵌合RNA包含(a)能够与真核细胞的基因组中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr序列杂交的tracr配对序列,和(c)所述tracr序列,其中(a)、(b)和(c)以5'至3'方向排列,

或

II.包含至少一个NLS的Cas9蛋白,CRISPR-Cas系统指导RNA,和包含tracr序列的tracrRNA,所述指导RNA包含(a)能够与真核细胞的基因组中与PAM相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,和(b)能够与所述tracr序列杂交的tracr配对序列。

161.一种工程化病毒载体,其编码:

I.包含至少一个核定位序列(NLS)的Cas9蛋白,和CRISPR-Cas系统嵌合RNA(chiRNA),所述嵌合RNA包含(a)能够与真核细胞的基因组中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr

序列杂交的tracr配对序列,和(c)所述tracr序列,其中(a)、(b)和(c)以5'至3'方向排列,或

II.包含至少一个NLS的Cas9蛋白,CRISPR-Cas系统指导RNA,和包含tracr序列的tracrRNA,所述指导RNA包含(a)能够与真核细胞的基因组中与PAM相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,和(b)能够与所述tracr序列杂交的tracr配对序列。

162.一种工程化AAV载体,其编码:

I.包含至少一个核定位序列(NLS)的Cas9蛋白,和CRISPR-Cas系统嵌合RNA(chiRNA),所述嵌合RNA包含(a)能够与真核细胞的基因组中与原型间隔子邻近基序(PAM)相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,(b)能够与tracr序列杂交的tracr配对序列,和(c)所述tracr序列,其中(a)、(b)和(c)以5'至3'方向排列,或

II.包含至少一个NLS的Cas9蛋白,CRISPR-Cas系统指导RNA,和包含tracr序列的tracrRNA,所述指导RNA包含(a)能够与真核细胞的基因组中与PAM相邻的靶序列杂交并指导CRISPR复合物与所述靶序列的序列特异性结合的指导序列,和(b)能够与所述tracr序列杂交的tracr配对序列。

163.一种用于选择真核细胞中核酸序列内用于被CRISPR复合物靶向的候选靶序列的计算机系统,所述系统包括:存储单元,其配置为接收和/或存储所述核酸序列;和单独或组合的一个或多个处理器,其编程为(i)在所述核酸序列内定位CRISPR基序序列,和(ii)选择与所述定位的CRISPR基序序列相邻的序列作为所述CRISPR复合物与其结合的所述候选靶序列。

164.一种计算机可读介质,其包含在由一个或多个处理器执行时实施选择真核细胞中核酸序列内用于被CRISPR复合物靶向的候选靶序列的方法的代码,所述方法包括:(a)定位所述核酸序列内的CRISPR基序序列,和(b)选择与所述定位的CRISPR基序列相邻的序列作为所述CRISPR复合物与其结合的所述候选靶序列。

165.一种工程化组合物,其包含:

Cas9蛋白;和

指导体,所述指导体包括:

(a)能够杂交到与原型间隔子邻近基序相邻的靶序列上的工程化指导序列,

(b)能够与tracr序列杂交的tracr配对序列,和

(c)所述tracr序列。

166.一种工程化组合物,其包含一个或多个载体,所述一个或多个载体包含编码Cas9蛋白的一个或多个核苷酸序列,和编码指导体的一个或多个核苷酸序列,所述指导体包括:

(a)能够杂交到与原型间隔子邻近基序相邻的靶序列上的工程化指导序列,

(b)能够与tracr序列杂交的tracr配对序列,和

(c)所述tracr序列。

167.根据权利要求165或166所述的组合物,其中所述Cas9蛋白质翻译自mRNA。

168.根据权利要求165或166所述的组合物,其中所述Cas9蛋白以蛋白质的形式递送。

169.根据权利要求165或166所述的组合物,其中(a)、(b)和(c)以5'至3'方向排列。

170.根据权利要求165或166所述的组合物，其中(b)和(c)用插入核苷酸彼此融合或共价连接。

171.根据权利要求165或166所述的组合物，其中所述Cas9蛋白衍生自链球菌或葡萄球菌。

172.根据权利要求171所述的组合物，其中所述Cas9蛋白是化脓链球菌Cas9。

173.根据权利要求165或166所述的组合物，其还包含模板。

174.根据权利要求165或166所述的组合物，其中所述Cas9蛋白与一个或多个NLS连接或融合。

175.根据权利要求166所述的组合物，其中编码所述Cas9蛋白的所述核苷酸序列经密码子优化以在真核细胞中表达。

176.根据权利要求170所述的组合物，其中所述tracr序列的长度为至少40个核苷酸。

177.根据权利要求165或166所述的组合物，其中所述组合物配置为适于通过电穿孔方法递送的递送组合物。

178.根据权利要求165或166所述的组合物，其中提供两种或更多种不同的指导序列用于多元靶向。

179.根据权利要求165或166所述的组合物，其中所述Cas9蛋白与至少一个功能结构域连接或融合。

180.根据权利要求179所述的组合物，其中所述功能结构域是转录激活结构域或转录阻抑结构域。

181.根据权利要求165或166所述的组合物，其中所述工程化组合物包含或编码两个或更多个指导体，其各自能够与所述真核细胞中的不同靶序列杂交。

182.根据权利要求165或166所述的组合物，其中所述工程化组合物还包含重组模板。

183.根据权利要求182所述的组合物，其中所述重组模板通过同源定向修复(HDR)插入。

184.一种修饰一个或多个靶序列的方法，其包括使细胞或细胞群与权利要求165或166所述的组合物接触，其中所述指导序列指导与所述细胞或细胞群中的靶序列的序列特异性结合，由此所述靶序列被修饰。

185.根据权利要求184所述的方法，其中所述一个或多个靶序列的修饰导致：

a.包含一种或多种基因产物的改变的表达水平的细胞或细胞群；

b.产生和/或分泌内源或非内源生物产物或化学化合物的细胞或细胞群。

186.一种工程化细胞，其包含根据权利要求184所述的方法修饰的一种或多种修饰基因。

187.根据权利要求186所述的工程化细胞，其中所述细胞产生和/或分泌内源或非内源生物产物或化学化合物。

188.根据权利要求187所述的工程化细胞，其中所述生物产物是蛋白质或RNA。

189.一种细胞系，其包含权利要求186所述的工程化细胞及其后代。

190.一种植物或动物模型，其包含权利要求186所述的工程化细胞及其后代。

191.一种组合物，其包含权利要求186所述的工程化细胞。

192.根据权利要求191所述的组合物，其配制用作治疗剂。

193.一种由权利要求186所述的工程化细胞产生的生物产物或化学化合物。

194.一种工程化细胞或其后代,所述细胞通过使用工程化CRISPR-Cas复合物来工程化。

195.根据权利要求194所述的细胞或其后代,其是分离的。

196.根据权利要求194所述的细胞或其后代,其进一步用作治疗剂。

197.根据权利要求194所述的细胞或其后代,从其分离产物。

198.一种由权利要求194所述的细胞或其后代产生的产物。

199.根据权利要求198所述的产物,其中所述产物是蛋白质或RNA。

200.根据权利要求199所述的产物,其中所述蛋白质包含突变。

201.一种用于治疗疾病或障碍的药物组合物,其包含权利要求194所述的细胞或其后代。

202.根据权利要求201所述的药物组合物,其中所述治疗导致一个或多个细胞中的遗传变化。

203.根据权利要求201所述的药物组合物,其中所述治疗导致一种或多种缺陷基因型的修正。

204.根据权利要求201所述的药物组合物,其中所述治疗导致改善的表型。

205.根据权利要求194所述的细胞或其后代,其中所述细胞包含从包含所述靶序列的基因表达的蛋白质中的突变。

206.根据权利要求205所述的细胞或其后代,其中所述细胞包含含有所述靶序列的基因组区域的缺失。

207.根据权利要求205所述的细胞或其后代,其中所述细胞包含通过同源定向修复对外源序列的整合。

208.根据权利要求205所述的细胞或其后代,其中所述细胞包含与所述靶序列相关的基因的减少的转录。

209.根据权利要求205所述的细胞或其后代,其中所述细胞包含与所述靶序列相关的基因的增加的转录。

210.根据权利要求198所述的产物,其中所述产物是由模板提供的突变的蛋白质或产物。

用于序列操纵和治疗应用的系统、方法和组合物的递送、工程化和优化

[0001] 本申请是申请日为2013年12月12日、申请号为2013800727522、发明 名称为“用于序列操纵和治疗应用的系统、方法和组合物的递送、工程化和 优化”的发明专利申请的分案申请。

[0002] 相关应用和引用参考

[0003] 本申请要求下列美国临时专利申请号的权益和优先权:2012年12月12 日提交的61/736,527;2013年1月2日提交的61/748,427;2013年1月30日 提交的61/758,468,2013年2月25日提交的61/769,046;2013年3月15日提 交的61/791,409和61/802,174,2013年3月28日提交的61/806,375;2013年 4月20日提交的61/814,263;2013年5月6日提交的61/819,803;2013年5 月28日提交的61/828,130;2013年6月17日提交的61/835,931和61/836,123以及2013年7月17日提交的61/847,537。

[0004] 还参考了下列美国临时专利申请号:2013年3月15日提交的 61/799,800;2013年6月17日提交的61/835,931、61/835,936、61/836,127、 61/836,101、61/836,080和61/835,973;2013年8月5日提交的61/862,468和 61/862,355;2013年8月28日提交的61/871,301;2013年9月25日提交的 61/960,777;以及2013年10月28日提交的61/961,980。

[0005] 前述申请、以及在其中或在它们的审查程序期间引用的所有文献或 (“申请引用文献”)以及在这些申请引用文献中引用或参考的所有文献、以及在本文中引用或参考的所有文献(“本文引用的文献”)、在本文中引 用的文献中引用或参考的所有文献,连同针对在本文中提及或通过引用结合 在本文中的任何文献中的任何产品的任何制造商的说明书、说明、产品规 格、和产品表,特此通过引用并入本文,并且可以在本发明的实践中采用。更具体地说,所有参考的文献均通过引用并入本文,其程度如同每个单独的 文献被确切地并单独地指明通过引用而并入本文。

发明领域

[0006] 本发明总体上涉及用于控制涉及序列靶向的基因表达的系统、方法、和 组合物的递送、工程化、优化和治疗应用,该序列靶向例如涉及规律间隔成 簇短回文重复序列(CRISPR)及其组分的基因组干扰或基因编辑。

[0007] 关于联邦资助研究的声明

[0008] 本发明是在美国国立卫生研究院颁发的NIH先锋奖(Pioneer Award)(1DP1MH100706)的政府支持下完成的。美国政府享有本发明的某些权 利。

[0009] 发明背景

[0010] 在基因组测序技术和分析方法中的最新进展显著加速了对与范围广泛的 生物功能和疾病相关联的遗传因子进行编目和映射的能力。精确的基因组靶 向技术对于通过允许个体遗传元件的选择性干扰而使得因果性遗传变异的统 性逆向工程成为可能、以及推进合成生物学、生物技术应用、和医学应用是 需要的。虽然基因组编辑技术,如设计师锌指、转录激活子样效应因子 (TALE)、或归巢大范围核酸酶(homing meganuclease)对于产

生靶向的基 因组干扰是可得的,但是仍然需要负担得起的、易于建立的、可扩展的、并 且便于靶向真核基因组内的多个位置的新的基因组工程技术。

[0011] 发明概述

[0012] 该CRISPR-Cas系统不要求产生针对靶特异性序列的定制蛋白,相反, 通过识别特异性DNA靶的短RNA分子可以对单个Cas酶进行程序设计。将 该CRISPR-Cas系统添加到基因组测序技术和分析方法的组库中可以显著使该 方法论简化并且提高对与范围广泛的生物功能和疾病相关联的遗传因子进行 编目和映射的能力。为了有效而无有害作用地利用CRISPR-Cas系统用于基因 组编辑,了解这些基因组工程工具的工程化、优化和细胞类型/组织/器官特异 性递送等方面是至关重要的,这些方面是所要求的本发明方面。

[0013] 对于用于一系列广泛的应用的核酸序列靶向的替代性的稳健的系统和技 术存在着迫切需要。本发明的多个方面着手解决这种需要并且提供了相关优 点。示例性CRISPR复合物包括与一种指导序列复合的CRISPR酶,该指导序 列与靶多核苷酸内的靶序列杂交。该指导序列与一种tracr配对序列连接,该 tracr配对序列进而与一种tracr序列杂交。

[0014] 在一方面,本发明提供了使用CRISPR-Cas系统的一个或多个元件的方 法。本发明的CRISPR复合物提供了用于修饰靶多核苷酸的有效手段。本发 明的CRISPR复合物具有多种多样的实用性,包括修饰(例如,缺失、插 入、转位、失活、激活)各种组织和器官中的多种细胞类型中的靶多核苷 酸。正因为如此,本发明的CRISPR复合物在例如基因或基因组编辑、基因 治疗、药物发现、药物筛选、疾病诊断、以及预后中具有广阔的应用谱。

[0015] 本发明的方面涉及在具备具有优化活性的指导RNA的CRISPR-Cas9系 统中的、具有改进的靶向特异性的、在长度上小于野生型Cas9酶的Cas9酶 和编码它的核酸分子、和嵌合的Cas9酶、以及改进Cas9酶的靶特异性或设 计CRISPR-Cas9系统的方法,所述方法包括设计或制备具有优化活性的指导RNA和/或选择或制备具有比野生型Cas9更小的尺寸或长度的Cas9酶,由此 将编码它的核酸包装到更高级的递送载体(因为在该递送载体中对它的编码 少于野生型Cas9)中,和/或产生嵌合Cas9酶。

[0016] 还提供了本发明的序列、载体、酶或系统在医学中的用途。还提供了它 们在基因或基因组编辑中的用途。

[0017] 在本发明的另外方面,Cas9酶可以包含一个或多个突变,并且可以用作 具有或不具有与功能结构域融合的通用DNA结合蛋白。这些突变可以是人工 引入的突变或获得性和丢失性功能突变。这些突变可以包括但不限于分别在 RuvC和HNH催化结构域中的催化结构域(D10和H840)之一中的突变。已 经表征了其他的突变。在本发明的一个方面,该转录激活结构域可以是 VP64。在本发明的其他方面,该转录阻抑结构域可以是KRAB或SID4X。本 发明的其他方面涉及融合到结构域上的突变的Cas9酶,这些结构域包括但不 限于,转录激活剂、阻抑物、重组酶、转座酶、组蛋白重塑剂(histone remodeler)、脱甲基酶、DNA甲基转移酶、隐花色素、光可诱导/可控制结构 域或化学可诱导/可控制结构域。

[0018] 在一个另外的实施例中,本发明提供了产生突变体tracrRNA和允许增强 这些RNA在细胞中的性能的同向重复序列或突变体嵌合指导序列的方法。本 发明的方面还提供了所述序列的选择。

[0019] 本发明的方面还提供了简化CRISPR复合物的组分的克隆和递送的方 法。在本发明的优选实施例中,适合的启动子,如U6启动子,与DNA寡核 苷酸一起扩增并且添加到指导

RNA上。然后将生成的PCR产物转染到细胞 中以驱动指导RNA的表达。本发明的方面还涉及体外转录的或从合成公司订 购并且直接转染的指导RNA。

[0020] 在一个方面，本发明提供了通过使用更具活性的聚合酶来提高活性的方 法。在一个优选的实施例中，这些指导RNA在T7启动子控制下的表达是由 该细胞中的T7聚合酶的表达驱动的。在一个有利的实施例中，该细胞是真核 细胞。在一个优选的实施例中，该真核细胞是人类细胞。在一个更优选的实 施例中，该人类细胞是患者特异性细胞。

[0021] 在一个方面，本发明提供了用于降低Cas酶的毒性的方法。在某些方 面，该Cas酶是如本文描述的任何Cas9，例如任何天然存在的细菌Cas9以及 任何嵌合体、突变体、同源物或直向同源物。在一个优选的实施例中，该 Cas9以mRNA的形式递送到该细胞中。这允许该酶的瞬时表达，由此降低毒 性。在另一个优选实施例中，本发明还提供了在诱导型启动子的控制下表达 Cas9的方法、以及在其中使用的构建体。

[0022] 在另一方面，本发明提供了用于改进CRISPR-Cas系统的体内应用的方 法。在该优选实施例中，该Cas酶是本文描述的野生型Cas9或任何修改版 本，包括任何天然存在的细菌Cas9以及任何嵌合体、突变体、同源物或直向 同源物。本发明的一个有利方面提供了易于被包装到病毒载体以便递送的 Cas9同源物的选择。Cas9直向同源物典型地共享3-4个RuvC结构域和一个 HNH结构域的通用结构。最5'的RuvC结构域切割非互补链，而HNH结构域切割互补链。所有符号都是就指导序列而言的。

[0023] 在5'RuvC结构域中的催化残基是通过该感兴趣的Cas9与其他Cas9直向 同源物(来自化脓链球菌II型CRISPR位点、嗜热链球菌CRISPR位点1、嗜 热链球菌CRISPR位点3、以及凶手弗朗西斯菌(Francisciella novicida)II型 CRISPR位点)的同源性比较而鉴定的，并且保守性Asp残基(D10)被突变 为丙氨酸以将Cas9转化成互补链切口酶。类似地，在HNH结构域中的保守 性His和Asn残基被突变为丙氨酸以将Cas9转化为非互补链切口酶。在一些 实施例中，这两组突变都可以进行，将Cas9转化为非切割酶。

[0024] 在一些实施例中，该CRISPR酶是I型或III型CRISPR酶，优选II型 CRISPR酶。这种II型CRISPR酶可以是任何Cas酶。优选的Cas酶可以被鉴 定为Cas9，因为这可以是指与来自II型CRISPR系统的具有多个核酸酶结构 域的最大核酸酶共享同源性的酶的通用类别。最优选地，该Cas9酶来自或衍 生自spCas9或saCas9。关于衍生，申请人意指该衍生的酶在很大程度上是基 于与野生型酶具有高度序列同源性的含义，但是如本文所述它在某些方面已 经被突变(修饰)。

[0025] 应当理解的是，术语Cas和CRISPR酶在本文中通常可互换地使用，除 非另外说明。如上提及的，在本文中使用的许多残基编号是指来自化脓链球 菌中的II型CRISPR位点的Cas9酶。然而，应当理解的是，本发明包括更多 的来自其他微生物物种的Cas9s，如SpCas9、SaCas9、St1Cas9等等。在本文 中提供了另外的实例。熟练的技术人员将能够通过比较相关氨基酸序列而确 定在除了SpCas9之外的Cas9酶中的适当的相应残基。因此，在特异性氨基酸置换是指使用SpCas9编号的情况下，那么，除非上下文清楚说明，这并不 预期是指其他Cas9酶，本披露预期涵盖在其他Cas9酶中的相应修饰。

[0026] 本文中提供了在这种情形下针对人类进行优化(即，针对在人类中表达 进行优化)的密码子优化序列的实例，参见SaCas9人类密码子优化序列。在 这被优化的同时，应当理解其他实例是可能的并且针对宿主物种的密码子优 化是已知的。

[0027] 在另外的实施例中,本发明提供了通过产生嵌合Cas9蛋白而增强Cas9 的功能的方法。嵌合Cas9蛋白嵌合Cas9可以是含有来自一种以上的天然存 在的Cas9的片段的新Cas9。这些方法可包括使一种Cas9同源物的N末端片 段与另一种Cas9同源物的C末端片段融合。这些方法还允许选择这些嵌合 Cas9蛋白所展示的新特性。

[0028] 应当理解的是,在本发明的方法中,在该生物是一种动物或植物的情况 下,该修饰可以离体地或在体外例如在细胞培养物中进行,在一些情况下不 在体内进行。在其他实施例中,可以在体内进行。

[0029] 在一个方面,本发明提供了一种通过操纵在感兴趣的基因组座位中的靶 序列来修饰一种生物或非人类生物的方法,该方法包括:

[0030] 递送一种非天然存在的或工程化的组合物,该组合物包含:

[0031] A)-I.一种CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列,其中该多 核苷酸序列包含:

[0032] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

[0033] (b)一种tracr配对序列,和

[0034] (c)一种tracr序列,以及

[0035] II.对包含至少一个或多个核定位序列的CRISPR酶进行编码的多核苷酸序 列,

[0036] 其中(a)、(b)和(c)以5'到3'方向排列,

[0037] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列指导CRISPR复合物与该靶序列的序列特异性结合,并且

[0038] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的指导序列、和(2) 杂交到该tracr序列上的tracr配对序列复合的CRISPR酶,并且该编码 CRISPR酶的多核苷酸序列是DNA或RNA,

[0039] 或者

[0040] (B)I.多核苷酸,其包含:

[0041] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,和

[0042] (b)至少一个或多个tracr配对序列,

[0043] II.一种编码CRISPR酶的多核苷酸序列,以及

[0044] III.一种包含tracr序列的多核苷酸序列,

[0045] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列指导CRISPR复合物与该靶序列的序列特异性结合,并且

[0046] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的指导序列、和(2) 杂交到该tracr序列上的tracr配对序列复合的CRISPR酶,并且该编码 CRISPR酶的多核苷酸序列是DNA或RNA。

[0047] 编码CRISPR酶的多核苷酸序列、指导序列、tracr配对序列或tracr序列 的任一者或全部可以是RNA。对编码CRISPR酶的序列、指导序列、tracr配 对序列或tracr序列进行编码的多核苷酸可以是RNA并且可以经由脂质体、 纳米粒子、外泌体、微囊泡、或基因枪进行递送。

[0048] 应当理解的是,在提及作为RNA并且被认为'包含'这样一种tracr配 对序列的特征的多核苷酸的情况下,该RNA序列包括该特征。在该多核苷酸 是DNA并且被认为包含这样

一种tracr配对序列的特征的情况下，该DNA序 列被或者可以被转录成包括该讨论中的特征的RNA。在该特征是蛋白质的情 况下，如CRISPR酶，所提及的该DNA或RNA序列被或者可以被翻译(以 及在DNA首先被转录的情况下)。

[0049] 因此，在某些实施例中，本发明提供了一种通过操纵在感兴趣的基因组 座位中的靶序列来修饰一种生物(例如包括人类的哺乳动物或非人类哺乳动 物或生物)的方法，该方法包括递送一种包含病毒或质粒载体系统的非天然 存在的或工程化的组合物，该载体系统包含可操作地编码用于其表达的组合 物的一种或多种病毒或质粒载体，其中该组合物包括：(A)一种非天然存在 的或工程化的组合物，该组合物包含一种载体系统，该载体系统包含一种或 多种载体，该一种或多种载体包含：I.一种第一调节元件，该第一调节元件可 操作地连接到一个CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列上， 其中该多核苷酸序列包含(a)一个能够杂交到真核细胞中的靶序列上的指导 序列，(b)一种tracr配对序列，和(c)一种tracr序列，以及II.一种第二调 节元件，该第二调节元件可操作地连接到编码CRISPR酶的酶编码序列上， 该CRISPR酶包含至少一个或多个核定位序列(或者任选地至少一个或多个 核定位序列，因为在一些实施例中可能不涉及NLS)，其中(a)、(b)和 (c)以5’到3’方向排列，其中组分I和II位于该系统的相同或不同载体上， 其中在转录时，该tracr配对序列杂交到该tracr序列上，并且该指导序列指导 CRISPR复合物与该靶序列的序列特异性结合，并且其中该CRISPR复合物包 含与(1)杂交到该靶序列上的指导序列、和(2)杂交到该tracr序列上的 tracr配对序列复合的CRISPR酶，或者(B)一种非天然存在的或工程化的组 合物，该组合物包含一种载体系统，该载体系统含有一种或多种载体，该一种或多种载体包含I.一种第一调节元件，该第一调节元件可操作地连接到 (a)能够杂交到真核细胞中的靶序列上的指导序列、和(b)至少一个或多 个tracr配对序列，II.一种第二调节元件，该第二调节元件可操作地连接到编 码CRISPR酶的酶编码序列上，以及III.一种第三调节元件，该第三调节元件 可操作地连接到tracr序列上，其中组分I、II和III位于该系统的相同或不同 载体上，其中在转录时，该tracr配对序列杂交到该tracr序列上，并且该指导 序列指导CRISPR复合物与该靶序列的序列特异性结合，并且其中该CRISPR 复合物包含与(1)杂交到该靶序列上的指导序列、和(2)杂交到该tracr序 列上的tracr配对序列复合的CRISPR酶。在一些实施例中，组分I、II和III 位于相同载体上。在其他实施例中，组分I和II位于相同载体上，而组分III 位于另一种载体上。在其他实施例中，组分I和III位于相同载体上，而组分 II位于另一种载体上。在其他实施例中，组分II和III位于相同载体上，而组 分I位于另一种载体上。在其他实施例中，组分I、II和III各自位于不同的载 体上。本发明还提供了如本文所述的病毒或质粒载体系统。

[0050] 优选地，该载体可以是病毒载体，如慢病毒或杆状病毒或优选地腺病毒/ 腺病毒相关病毒载体，但是其他递送手段也是已知的(如酵母系统、微囊 泡、基因枪/将载体附接到金纳米粒子上的手段)并且被提供。在一些实施例 中，可以经由脂质体、纳米粒子、外泌体、微囊泡、或基因枪递送一种或多 种病毒载体或质粒载体。

[0051] 通过操纵靶序列，申请人还打算靶序列的表观遗传操纵。这可以是靶序 列的染色质状态的操纵，如借助于修饰靶序列的甲基化状态(即，甲基化或 甲基化模式或CpG岛的添加或去除)、组蛋白修饰，从而增加或降低靶序列 的可及性，或者借助于3D折叠。

[0052] 应当理解的是，在提及一种通过操纵在感兴趣的基因组座位中的靶序列 来修饰

一种生物或哺乳动物(包括人类或非人类哺乳动物或生物)的方法的 情况下,这可适用于作为整体的生物(或哺乳动物)或者仅仅是来自这种生 物的单个细胞或细胞群(如果该生物是多细胞的话)。在人类的情况下,例 如,申请人特别地设想到单个细胞或细胞群,并且这些细胞可以优选地进行 离体修饰,进而重新引入。在这种情况下,活组织检查或其他组织或生物流 体样品可以必要的。在这方面,干细胞也是特别优选的。但是,当然还设想 了体内实施例。

[0053] 在某些实施例中,本发明提供了一种治疗或抑制对其有需要的受试者 (例如哺乳动物或人类)或非人类受试者(例如,哺乳动物)中感兴趣的基 因组座位中的靶序列缺陷所引起的病症的方法,该方法包括通过操纵该靶序 列而修饰该受试者或非人类受试者,并且其中该病症对于借助于包括提供治 疗的靶序列操纵的治疗或抑制是敏感的,该治疗包括:递送包含AAV或慢病 毒载体系统的非天然存在的或工程化的组合物,该载体系统含有一种或多种 AAV或慢病毒载体,所述载体可操作地编码用于其表达的组合物,其中在表 达时该靶序列由该非天然存在的或工程化的组合物操纵,其中该非天然存在 的或工程化的组合物包括:(A)一种非天然存在的或工程化的组合物,该组 合物包含一种载体系统,该载体系统包含一种或多种载体,该一种或多种载 体包含:I.一种第一调节元件,该第一调节元件可操作地连接到一个CRISPR- Cas系统嵌合RNA(chiRNA)多核苷酸序列上,其中该多核苷酸序列包含 (a)一个能够杂交到真核细胞中的靶序列上的指导序列,(b)一种tracr配 对序列,和(c)一种tracr序列,以及II.一种第二调节元件,该第二调节元 件可操作地连接到编码CRISPR酶的酶编码序列上,该CRISPR酶包含至少一 个或多个核定位序列(或者任选地至少一个或多个核定位序列,因为在一些 实施例中可能不涉及NLS),其中(a)、(b)和(c)以5'到3'方向排列, 其中组分I和II位于该系统的相同或不同载体上,其中在转录时,该tracr配 对序列杂交到该tracr序列上,并且该指导序列指导CRISPR复合物与该靶序 列的序列特异性结合,并且其中该CRISPR复合物包含与(1)杂交到该靶序 列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列复合的 CRISPR酶,或者(B)一种非天然存在的或工程化的组合物,该组合物包含 一种载体系统,该载体系统含有一种或多种载体,该一种或多种载体包含I. 一种第一调节元件,该第一调节元件可操作地连接到(a)能够杂交到真核细胞中的靶序列上的指导序列、和(b)至少一个或多个tracr配对序列,II.一种 第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶的酶编码序 列上,以及III.一种第三调节元件,该第三调节元件可操作地连接到tracr序列 上,其中组分I、II和III位于该系统的相同或不同载体上,其中在转录时, 该tracr配对序列杂交到该tracr序列上,并且该指导序列指导CRISPR复合物 与该靶序列的序列特异性结合,并且其中该CRISPR复合物包含与(1)杂交 到该靶序列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列复 合的CRISPR酶。在一些实施例中,组分I、II和III位于相同载体上。在其他 实施例中,组分I和II位于相同载体上,而组分III位于另一种载体上。在其 他实施例中,组分I和III位于相同载体上,而组分II位于另一种载体上。在 其他实施例中,组分II和III位于相同载体上,而组分I位于另一种载体上。 在其他实施例中,组分I、II和III各自位于不同的载体上。本发明还提供了 如本文所述的病毒(例如,AAV或慢病毒)载体系统,并且可以是如本文所 述的载体系统的一部分。

[0054] 本发明的一些方法可以包括诱导表达。在本发明的一些方法中,该生物 或受试者

是真核生物(包括人类在内的哺乳动物)或非人类真核生物或非人 类动物或非人类哺乳动物。在一些实施例中,该生物或受试者是非人类动 物,并且可以是节肢动物例如昆虫,或者可以是线虫。在本发明的一些方法 中,该生物或受试者是植物。在本发明的一些方法中,该生物或受试者是哺 乳动物或非人类哺乳动物。非人类哺乳动物可以是例如啮齿动物(优选小鼠 或大鼠)、有蹄动物、或灵长动物。在本发明的一些方法中,该生物或受试 者是包括微藻在内的藻类,或者是真菌。在本发明的一些方法中,该病毒载 体是AAV或慢病毒,并且可以是如本文所述的载体系统的一部分。在本发明 的一些方法中,该CRISPR酶是Cas9。在本发明的一些方法中,该指导序列 的表达是在T7启动子控制下并且是由T7聚合酶的表达驱动的。

[0055] 在一些实施例中,本发明包括一种递送CRISPR酶的方法,该方法包括 向细胞递送编码该CRISPR酶的mRNA。在本发明的这些方法的一些中,该 CRISPR酶是Cas9。

[0056] 本发明还提供了制备本发明的载体系统尤其是如本文所述的病毒载体系 统的方法。在一些实施例中,本发明包括一种制备本发明的AAV的方法,该 方法包括将含有或其基本组成为编码AAV的一种或多种核酸分子的一种或多 种质粒转染到感染AAV的细胞中,并且提供对于AAV的复制和包装必须的 AAV rep和/或cap。在一些实施例中,对于AAV的复制和包装必须的AAV rep和/或cap是通过用一种或多种辅助质粒或一种或多种辅助病毒转染这些细 胞而提供的。在一些实施例中,该辅助病毒是痘病毒、腺病毒、疱疹病毒或 杆状病毒。在一些实施例中,该痘病毒是牛痘病毒。在一些实施例中,这些 细胞是哺乳动物细胞。并且在一些实施例中,这些细胞是昆虫细胞,并且该 辅助病毒是杆状病毒。在其他实施例中,该病毒是慢病毒。

[0057] 在植物中,病原体常常是宿主特异性的。例如,番茄尖镰孢菌番茄专化 型(Fusarium oxysporum f.sp.lycopersici)引起番茄枯萎病,而且只攻击番 茄,并且香石竹尖镰孢禾柄锈菌小麦专化型(F.oxysporum f.dianthii Puccinia graminis f.sp.tritici)只攻击小麦。植物具有抵抗大多数病原体的现有的和诱 导性的防御。跨植物代的突变和重组事件导致产生敏感性的遗传变异性,特 别是当病原体以比植物更高的频率繁殖时。在植物中可以存在非宿主抗性, 例如,该宿主和病原体是不相容的。还可以存在水平抗性,例如典型地由许 多基因控制的针对所有种的病原体的部分抗性,以及垂直抗性,例如典型地 由很少的基因控制的针对病原体的某些种而不是其他种的完全抗性。在基因 对基因水平中,植物和病原体一起演化,并且在一者中的遗传变化使在另一 者中的变化平衡。因此,使用自然变异,育种者针对产率、质量、均匀性、 耐性、抗性将最有用的基因进行结合。抗性基因的来源包括天然或外来品 种、传家宝品种(Heirloom Varieties)、近缘野生植物、以及诱导的突变,例 如用诱变剂处理植物材料。利用本发明,向植物育种者提供了一种新的诱导 突变的工具。因此,本领域的技术人员可以分析抗性基因的来源的基因组, 并且在具有所希望的特征或性状的品种方面,采用本发明来诱导抗性基因的 发生,这样具有比先前的诱变剂更好的精确性,并且因此加速并改良植物育 种计划。

[0058] 本发明进一步包括本发明的组合物或其CRISPR酶(包括或可替代地是 编码CRISPR酶的mRNA),其用于在医学或治疗中使用。在一些实施例 中,本发明包括根据本发明的组合物或其CRISPR酶(包括或可替代地是编 码CRISPR酶的mRNA),其用于在根据本发明的方法中使用。在一些实施 例中,本发明提供了本发明的组合物或其CRISPR酶(包括或可

替代地是编 码CRISPR酶的mRNA)在离体基因或基因组编辑中的用途。在某些实施例 中,本发明包括本发明的组合物或其CRISPR酶(包括或可替代地是编码 CRISPR酶的mRNA)在用于离体基因或基因组编辑中或在根据本发明的方法 中使用的药剂的制造中的用途。在一些实施例中,本发明包括本发明的组合 物或其CRISPR酶(包括或可替代地是编码CRISPR酶的mRNA),其中该靶 序列在其3'端的侧翼为PAM(原型间隔子邻近基序)序列,该PAM序列包 含5'-基序,尤其是在Cas9是(或衍生自)化脓链球菌或金黄色葡萄球菌 Cas9的情况下。例如,适合的PAM是分别针对SpCas9或SaCas9酶(或衍生 的酶)的5'-NRG或5'-NNGRR(其中N是任何核苷酸),如下文提及的。

[0059] 应当理解的是,SpCas9或SaCas9是来自或衍生自化脓链球菌或金黄色 葡萄球菌Cas9的那些。

[0060] 本发明的方面包括提高CRISPR酶例如Cas9介导的基因靶向的特异性，以及降低经由CRISPR酶例如Cas9的脱靶修饰的可能性。在一些实施例中，本发明包括通过借助于操纵在细胞中的感兴趣基因组座位中的DNA双链体的 相对链上的第一和第二靶序列而将脱靶修饰降低到最低限度来修饰一种生物 或非人类生物的方法,该方法包括递送一种非天然存在的或工程化的组合 物,该组合物包含:

[0061] I.一种第一CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列,其中 该第一多核苷酸序列包含:

[0062] (a)一个第一指导序列,该第一指导序列能够杂交到该第一靶序列上,

[0063] (b)一个第一tracr配对序列,和

[0064] (c)一个第一tracr序列,

[0065] II.一种第二CRISPR-Cas系统chiRNA多核苷酸序列,其中该第二多核苷 酸序列包含:

[0066] (a)一个第二指导序列,该第二指导序列能够杂交到该第二靶序列上,

[0067] (b)一个第二tracr配对序列,和

[0068] (c)一个第二tracr序列,以及

[0069] III.一种对包含至少一个或多个核定位序列并且包含一个或多个突变的 CRISPR酶进行编码的多核苷酸序列,其中(a)、(b)和(c)以5'到3'方 向排列,其中在转录时,该第一和第二tracr配对序列分别杂交到该第一和第 二tracr序列上,并且该第一和第二指导序列分别指导第一和第二CRISPR复 合物与该第一和第二靶序列的序列特异性结合,其中该第一CRISPR复合物 包含与(1)杂交到该第一靶序列上的第一指导序列、和(2)杂交到该第一 tracr序列上的第一tracr配对序列复合的CRISPR酶,其中该第二CRISPR复 合物包含与(1)杂交到该第二靶序列上的第二指导序列、和(2)杂交到该 第二tracr序列上的第二tracr配对序列复合的CRISPR酶,其中编码CRISPR 酶的多核苷酸序列是DNA或RNA,并且其中该第一指导序列指导邻近该第 一靶序列的DNA双链体的一条链的切割,并且该第二指导序列指导邻近该第 二靶序列的另一条链的切割,从而诱导双链的断裂,由此通过将脱靶修饰降 低到最低限度而修饰该生物或非人类生物。

[0070] 在本发明的一些方法中,编码CRISPR酶的多核苷酸序列、该第一和第 二指导序列、该第一和第二tracr配对序列或该第一和第二tracr序列中的任一 者或全部是RNA。在本发明另外的实施例中,编码CRISPR酶的编码序列的 多核苷酸、该第一和第二指导序列、

该第一和第二tracr配对序列或该第一和 第二tracr序列是RNA，并且是经由脂质体、纳米粒子、外泌体、微囊泡、或 基因枪递送的。在本发明的某些实施例中，该第一和第二tracr配对序列共享 100％的一致性和/或该第一和第二tracr序列共享100％的一致性。在一些实施 例中，这些多核苷酸可以被包含在含有一种或多种载体的载体系统中。在本 发明的优选实施例中，该CRISPR酶是一种Cas9酶，例如SpCas9。在本发明 的一个方面，该CRISPR酶包含在催化结构域中的一个或多个突变，其中该 一个或多个突变选自下组，该组由以下各项组成：D10A、E762A、H840A、 N854A、N863A和D986A。在一个高度优选的实施例中，该CRISPR酶具有 D10A突变。在优选的实施例中，该第一CRISPR酶具有一个或多个突变，使 得该酶是一种互补链切口酶，并且该第二CRISPR酶具有一个或多个突变， 使得该酶是一种非互补链切口酶。可替代地，该第一酶可以是一种非互补链 切口酶，而该第二酶可以是一种互补链切口酶。

[0071] 在本发明的优选方法中，指导邻近该第一靶序列的DNA双链体的一条链 的切割的第一指导序列以及指导邻近该第二靶序列的另一条链的切割的第二 指导序列导致一个5’突出端。在本发明的实施例中，该5’突出端具有至多 200个碱基对、优选至多100个碱基对、或更优选至多50个碱基对。在本发 明的实施例中，该5’突出端具有至少26个碱基对、优选至少30个碱基对、或更优选34-50个碱基对。

[0072] 在一些实施例中，本发明包括通过借助于操纵在细胞中的感兴趣基因组 座位中的DNA双链体的相对链上的第一和第二靶序列而将脱靶修饰降低到最 低限度来修饰一种生物或非人类生物的方法，该方法包括递送一种非天然存 在的或工程化的组合物，该组合物包含一种载体系统，该载体系统包含一种 或多种载体，该一种或多种载体包含

[0073] I.一种第一调节元件，该第一调节元件可操作地连接至

[0074] (a)一个第一指导序列，该第一指导序列能够杂交到该第一靶序列上，和

[0075] (b)至少一个或多个tracr配对序列，

[0076] II.一种第二调节元件，该第二调节元件可操作地连接至

[0077] (a)一个第二指导序列，该第二指导序列能够杂交到该第二靶序列上，和

[0078] (b)至少一个或多个tracr配对序列，

[0079] III.一种第三调节元件，该第三调节元件可操作地连接至编码CRISPR酶 的酶编码序列，以及

[0080] IV.一个第四调节元件，该第四调节元件可操作地连接至一种tracr序列，

[0081] 其中组分I、II、III和IV位于该系统的相同或不同载体上，在转录时， 该tracr配对序列杂交到该tracr序列上，并且该第一和第二指导序列分别指导 第一和第二CRISPR复合物与该第一和第二靶序列的序列特异性结合，其中 该第一CRISPR复合物包含与(1)杂交到该第一靶序列上的第一指导序列、和(2)杂交到该tracr序列上的tracr配对序列复合的CRISPR酶，其中该第 二CRISPR复合物包含与(1)杂交到该第二靶序列上的第二指导序列、和 (2)杂交到该tracr序列上的tracr配对序列复合的CRISPR酶，其中编码 CRISPR酶的多核苷酸序列是DNA或RNA，并且其中该第一指导序列指导邻 近该第一靶序列的DNA双链体的一条链的切割，并且该第二指导序列指导邻 近该第二靶序列的另一条链的切割，从而诱导双链的断裂，由此通过将脱靶 修饰降低到最低限度而修饰该生物或非人类生物。

[0082] 本发明还提供了如本文所述的载体系统。该系统可以包含一个、二个、 三个或四

个不同的载体。组分I、II、III和IV因此可以位于一个、二个、三 个或四个不同的载体上，并且本文设想了对于这些组分的可能位置的所有组 合，例如：组分I、II、III和IV可以位于相同的载体上；组分I、II、III和IV 可以各自位于不同的载体上；组分I、II、III和IV可以位于总共二个或三个 不同的载体上，其中设想了所有的位置组合，等等。

[0083] 在本发明的一些方法中，编码CRISPR酶的多核苷酸序列、该第一和第 二指导序列、该第一和第二tracr配对序列或该第一和第二tracr序列中的任一 者或全部是RNA。在本发明另外的实施例中，该第一和第二tracr配对序列共 享100％的一致性和/或该第一和第二tracr序列共享100％的一致性。在本发明 的优选实施例中，该CRISPR酶是一种Cas9酶，例如SpCas9。在本发明的一 个方面，该CRISPR酶包含在催化结构域中的一个或多个突变，其中该一个 或多个突变选自下组，该组由以下各项组成：D10A、E762A、H840A、 N854A、N863A和D986A。在一个高度优选的实施例中，该CRISPR酶具有 D10A突变。在优选的实施例中，该第一CRISPR酶具有一个或多个突变，使 得该酶是一种互补链切口酶，并且该第二CRISPR酶具有一个或多个突变， 使得该酶是一种非互补链切口酶。可替代地，该第一酶可以是一种非互补链 切口酶，而该第二酶可以是一种互补链切口酶。在本发明的一个另外的实施 例中，这些病毒载体的一种或多种可以经由脂质体、纳米粒子、外泌体、微 囊泡、或基因枪进行递送。

[0084] 在本发明的优选方法中，指导邻近该第一靶序列的DNA双链体的一条链 的切割的第一指导序列以及指导邻近该第二靶序列的另一条链的切割的第二 指导序列产生一个5’突出端。在本发明的实施例中，该5’突出端具有至多 200个碱基对、优选至多100个碱基对、或更优选至多50个碱基对。在本发 明的实施例中，该5’突出端具有至少26个碱基对、优选至少30个碱基对、或更优选34-50个碱基对。

[0085] 在一些实施例中，本发明包括一种通过借助于向含有和表达编码感兴趣 的基因产物的双链DNA分子的细胞中引入一种工程化的、非天然存在的 CRISPR-Cas系统将脱靶修饰降低至最低限度而修饰感兴趣的基因组的方法， 该CRISPR-Cas系统包含具有一个或多个突变的Cas蛋白和两个分别靶向该 DNA分子的第一链和第二链的指导RNA，由此这些指导RNA靶向编码该基 因产物的DNA分子，并且该Cas蛋白使编码该基因产物的DNA分子的第一链和第二链的每一者产生切口，由此改变该基因产物的表达；并且，其中该 Cas蛋白和这两个指导RNA并不一起天然存在。

[0086] 在本发明的优选方法中，该Cas蛋白使编码该基因产物的DNA分子的第 一链和第二链的每一者产生切口导致一个5’突出端。在本发明的实施例中， 该5’突出端具有至多200个碱基对、优选至多100个碱基对、或更优选至多 50个碱基对。在本发明的实施例中，该5’突出端具有至少26个碱基对、优选 至少30个碱基对、或更优选34-50个碱基对。

[0087] 本发明的实施例还包括包含融合到tracr配对序列和tracr序列上的指导序 列的指导RNA。在本发明的一个方面，该Cas蛋白经密码子优化以便在真核 细胞，优选哺乳动物细胞或人类细胞中中表达。在本发明另外的实施例中， 该Cas蛋白是一种II型CRISPR-Cas蛋白，例如一种Cas9蛋白。在一个高度 优选的实施例中，该Cas蛋白是一种Cas9蛋白，例如SpCas9。在本发明的方 面中，该Cas蛋白具有一个或多个选自下组的突变，该组由以下各项组成： D10A、E762A、H840A、N854A、N863A和D986A。在一个高度优选的实施 例中，该Cas蛋白具有D10A突变。

[0088] 本发明的方面涉及被减少的基因产物或被进一步引入到编码基因产物的 DNA分子中的模板多核苷酸或通过允许两个5'突出端重新退火并连接而被精 确切断的间插序列的表达、或被改变的基因产物的活性或功能、或被增加的 基因产物的表达。在本发明的一个实施例中,该基因产物是一种蛋白质。

[0089] 本发明还包括一种工程化的、非天然存在的CRISPR-Cas系统,该系统 包含具有一个或多个突变的Cas蛋白和两个分别靶向在细胞中编码基因产物 的双链DNA分子的第一链和第二链的指导RNA,由此这些指导RNA靶向编 码该基因产物的DNA分子,并且该Cas蛋白使编码该基因产物的DNA分子 的第一链和第二链的每一者产生切口,由此改变该基因产物的表达;并且, 其中该Cas蛋白和这两个指导RNA并不一起天然存在。

[0090] 在本发明的方面中,这些指导RNA可包含融合到tracr配对序列和tracr 序列上的指导序列。在本发明的一个实施例中,该Cas蛋白是一种II型 CRISPR-Cas蛋白。在本发明的一个方面,该Cas蛋白经密码子优化以便在真 核细胞,优选哺乳动物细胞或人类细胞中中表达。在本发明另外的实施例 中,该Cas蛋白是一种II型CRISPR-Cas蛋白,例如一种Cas9蛋白。在一个 高度优选的实施例中,该Cas蛋白是一种Cas9蛋白,例如SpCas9。在本发明 的方面中,该Cas蛋白具有一个或多个选自下组的突变,该组由以下各项组 成:D10A、E762A、H840A、N854A、N863A和D986A。在一个高度优选的 实施例中,该Cas蛋白具有D10A突变。

[0091] 本发明的方面涉及被减少的基因产物或被进一步引入到编码基因产物的 DNA分子中的模板多核苷酸或通过允许两个5'突出端重新退火并连接而被精 确切断的间插序列的表达、或被改变的基因产物的活性或功能、或被增加的 基因产物的表达。在本发明的一个实施例中,该基因产物是一种蛋白质。

[0092] 本发明还包括包含一种或多种载体的工程化的、非天然存在的载体系 统,所述载体包含:

[0093] a)一种第一调节元件,该第一调节元件可操作地连接到两个CRISPR-Cas系 统指导RNA的每一者上,这两个指导RNA分别靶向编码基因产物的双链 DNA分子的第一链和第二链,

[0094] b)一种第二调节元件,该第二调节元件可操作地连接至Cas蛋白,

[0095] 其中组分(a)和(b)位于该系统的相同或不同载体上,由此这些指导RNA 靶向编码该基因产物的DNA分子,并且该Cas蛋白使编码该基因产物的 DNA分子的第一链和第二链的每一者产生切口,由此改变该基因产物的表 达;并且,其中该Cas蛋白和这两个指导RNA并不一起天然存在。

[0096] 在本发明的方面中,这些指导RNA可包含融合到tracr配对序列和tracr 序列上的指导序列。在本发明的一个实施例中,该Cas蛋白是一种II型 CRISPR-Cas蛋白。在本发明的一个方面,该Cas蛋白经密码子优化以便在真 核细胞,优选哺乳动物细胞或人类细胞中表达。在本发明另外的实施例中, 该Cas蛋白是一种II型CRISPR-Cas蛋白,例如一种Cas9蛋白。在一个高度 优选的实施例中,该Cas蛋白是一种Cas9蛋白,例如SpCas9。在本发明的方面中,该Cas蛋白具有一个或多个选自下组的突变,该组由以下各项组成: D10A、E762A、H840A、N854A、N863A和D986A。在一个高度优选的实施 例中,该Cas蛋白具有D10A突变。

[0097] 本发明的方面涉及被减少的基因产物或被进一步引入到编码基因产物的 DNA分子中的模板多核苷酸或通过允许两个5'突出端重新退火并连接而被精 确切断的间插序列

的表达、或被改变的基因产物的活性或功能、或被增加的 基因产物的表达。在本发明的一个实施例中，该基因产物是一种蛋白质。在 本发明的优选实施例中，该系统的这些载体是病毒载体。一个另外的实施例 中，该系统的这些载体经由脂质体、纳米粒子、外泌体、微囊泡、或基因枪 进行递送。

[0098] 在一个方面，本发明提供了修饰在一种真核细胞中的靶多核苷酸的方 法。在一些实施例中，该方法包括允许一种CRISPR复合物结合到该靶多核 苷酸上以实施所述靶多核苷酸的切割，由此修饰该靶多核苷酸，其中该 CRISPR复合物包括与杂交到在所述靶多核苷酸内的一个靶序列上的指导序列 复合的CRISPR酶，其中所述指导序列连接到一种tracr配对序列上，该tracr 配对序列进而杂交到一种tracr序列上。在一些实施例中，所述切割包括通过 所述CRISPR酶切割在该靶序列位置的一条或两条链。在一些实施例中，所 述切割导致靶基因的转录降低。在一些实施例中，该方法进一步包括通过与 外源模板多核苷酸同源重组修复所述切割的靶多核苷酸，其中所述修复导致 一种突变，包括所述靶多核苷酸的一个或多个核苷酸的插入、缺失、或取 代。在一些实施例中，所述突变导致在从包含该靶序列的基因表达的蛋白质 中的一个或多个氨基酸改变。在一些实施例中，该方法进一步包括将一种或 多种载体递送到所述真核细胞，其中该一种或多种载体驱动下列一者或多者 的表达：该CRISPR酶、连接到该tracr配对序列上的指导序列、以及该tracr 序列。在一些实施例中，所述载体被递送到受试者内的真核细胞中。在一些 实施例中，所述修饰发生在细胞培养物中的所述真核细胞中。在一些实施例 中，该方法进一步包括在所述修饰之前从受试者中分离所述真核细胞。在一 些实施例中，该方法进一步包括使所述真核细胞和/或从中衍生的细胞返回到 所述受试者中。

[0099] 在一个方面，本发明提供了修饰一种多核苷酸在真核细胞中的表达的方 法。在一些实施例中，该方法包括允许一种CRISPR复合物结合到该多核苷 酸上，这样使得所述结合导致所述多核苷酸的表达增加或降低；其中该 CRISPR复合物包括与杂交到在所述多核苷酸内的一个靶序列上的指导序列复 合的CRISPR酶，其中所述指导序列连接到一种tracr配对序列上，该tracr配 对序列进而杂交到一种tracr序列上。在一些实施例中，该方法进一步包括将 一种或多种载体递送到所述真核细胞，其中该一种或多种载体驱动下列一者 或多者的表达：该CRISPR酶、连接到该tracr配对序列上的指导序列、以及 该tracr序列。

[0100] 在一个方面，本发明提供了一种产生包含突变的疾病基因的模式真核细 胞的方法。在一些实施例中，疾病基因是与患有或产生一种疾病的风险的增 加相关联的任何基因。在一些实施例中，该方法包括(a)将一种或多种载体 引入真核细胞中，其中该一种或多种载体驱动下列一者或多者的表达：CRISPR酶、连接到tracr配对序列上的指导序列、以及tracr序列；以及(b) 允许一种CRISPR复合物结合到一个靶多核苷酸上以实施在所述疾病基因内 的该靶多核苷酸的切割，其中该CRISPR复合物包含与(1)杂交到该靶多核 苷酸内的该靶序列上的指导序列、和(2)杂交到该tracr上的该tracr配对序 列复合的CRISPR酶，由此产生包含突变的疾病基因的模式真核细胞。在一 些实施例中，所述切割包括通过所述CRISPR酶切割在该靶序列位置的一条 或两条链。在一些实施例中，所述切割导致靶基因的转录降低。在一些实施 例中，该方法进一步包括通过与外源模板多核苷酸同源重组修复所述切割的 靶多核苷酸，其中所述修复导致一种突变，包括所述靶多核苷酸的一个或多 个核苷酸的插入、缺失、或取代。在一些实施例中，所述突变导致在从包含 该靶序列的基因表

达的蛋白质中的一个或多个氨基酸改变。

[0101] 在一个方面,本发明提供了一种通过在一个或多个原核细胞中的基因中 引入一个或多个突变来选择一个或多个原核细胞的方法,该方法包括:将一 种或多种载体引入一种或多种原核细胞中,其中该一种或多种载体驱动下列 一者或多者的表达:CRISPR酶、连接到tracr配对序列上的指导序列、tracr 序列、以及编辑模板;其中该编辑模板包含消除CRISPR酶切割的一个或多 个突变;允许该编辑模板与在有待选择的一个或多个细胞中的靶多核苷酸同 源重组;允许CRISPR复合物结合到靶多核苷酸上以实施在所述基因内的靶多核苷酸的切割,其中该CRISPR复合物包含与(1)杂交到该靶多核苷酸内 的靶序列上的指导序列、和(2)杂交到该tracr序列上的tracr配对序列复合 的CRISPR酶,其中该CRISPR复合物到该靶多核苷酸上的结合诱导细胞死 亡,由此允许选择其中已经引入一个或多个突变的一个或多个原核细胞。在 一个优选的实施例中,该CRISPR酶是Cas9。在本发明的另一方面,该有待 选择的细胞可以是真核细胞。本发明的方面允许选择特异细胞,而不需要选择标记或可能包括反选择系统的两步法。

[0102] 在一个方面,本发明提供了用于修饰在一种真核细胞中的靶多核苷酸的 方法。在一些实施例中,该方法包括允许一种CRISPR复合物结合到该靶多 核苷酸上以实施所述靶多核苷酸的切割,由此修饰该靶多核苷酸,其中该 CRISPR复合物包括与杂交到在所述靶多核苷酸内的一个靶序列上的指导序列 复合的CRISPR酶,其中所述指导序列连接到一种tracr配对序列上,该tracr 配对序列进而杂交到一种tracr序列上。

[0103] 在其他实施例中,本发明提供了一种修饰多核苷酸在真核细胞中的表达 的方法。该方法包括通过使用结合到多核苷酸上的CRISPR复合物增加或降 低靶多核苷酸的表达。

[0104] 在希望的情况下,为了实施在细胞中的表达的修饰,将包含tracr序列、连接到该tracr配对序列上的指导序列、编码CRISPR酶的序列的一种或多种 载体递送到细胞。在一些方法中,该一种或多种载体包含可操作地连接到编 码所述CRISPR酶的酶编码序列上的一种调节元件,所述CRISPR酶包含核定 位序列;和可操作地连接到tracr配对序列的一种调节元件以及用于在该tracr 配对序列的上游插入一种指导序列的一个或多个插入位点。当表达时,该指 导序列指导CRISPR复合物与细胞中的一个靶序列的序列特异性结合。典型地,该CRISPR复合物包含与(1)杂交到该靶序列上的指导序列、和(2) 杂交到该tracr序列上的tracr配对序列复合的CRISPR酶。

[0105] 在一些方法中,可以使靶多核苷酸失活以实施细胞中的表达的修饰。例 如,当CRISPR复合物结合到细胞中的一个靶序列上时,该靶多核苷酸失 活,这样使得该序列不被转录,该编码蛋白不被产生,或者该序列不会像野 生型序列一样起作用。例如,可以使蛋白质或微小RNA编码序列失活,这样 使得该蛋白质不被产生。

[0106] 在某些实施例中,该CRISPR酶包含一个或多个选自下组的突变,该组 由以下各项组成:D10A、E762A、H840A、N854A、N863A或D986A和/或 该一个或多个突变在该CRISPR酶的RuvC1或HNH结构域中或者为如本文 另外论述的突变。在一些实施例中,该CRISPR酶具有一个或多个在催化结 构域中的突变,其中在转录时,该tracr配对序列杂交到该tracr序列上,并且 该指导序列指导CRISPR复合物与该靶序列的序列特异性结合,并且其中该 酶进一步包含一个功能结构域。在一些实施例中,该功能结构域是一个转录 激活结构域,优选VP64。在一些实施例中,该功能结构域一个转录阻抑结构 域,优选KRAB。在一些实施例中,

该转录阻抑结构域是SID、或SID的多 联体(例如SID4X)。在一些实施例中,该功能结构域是一个表观遗传修饰 结构域,从而提供了一种表观遗传修饰酶。在一些实施例中,该功能结构域 是一种激活结构域,其可以是P65激活结构域。

[0107] 在一些实施例中,该CRISPR酶是I型或III型CRISPR酶,但是优选II 型CRISPR酶。这种II型CRISPR酶可以是任何Cas酶。一种Cas酶可以被 鉴定为Cas9,因为这可以是指与来自II型CRISPR系统的具有多个核酸酶结 构域的最大核酸酶共享同源性的酶的通用类别。最优选地,该Cas9酶来自或 衍生自spCas9或saCas9。关于衍生,申请人意指该衍生的酶在很大程度上是 基于与野生型酶具有高度序列同源性的含义,但是如本文所述它在某些方面 已经被突变(修饰)。

[0108] 应当理解的是,术语Cas和CRISPR酶在本文中通常可互换地使用,除 非另外说明。如上提及的,在本文中使用的许多残基编号是指来自化脓链球 菌中的II型CRISPR座位的Cas9酶。然而,应当理解的是,本发明包括更多 的来自其他微生物物种的Cas9s,如SpCas9、SaCa9、St1Cas9等等。

[0109] 本文中提供了在这种情形下针对人类进行优化(即,针对在人类中表达 进行优化)的密码子优化序列的实例,参见SaCas9人类密码子优化序列。在 这被优化的同时,应当理解其他实例是可能的并且针对宿主物种的密码子优 化是已知的。

[0110] 优选地,递送是以载体形式进行的,该载体可以是病毒载体,如慢病毒 或杆状病毒或优选地腺病毒/腺病毒相关病毒载体,但是其他递送手段也是已 知的(如酵母系统、微囊泡、基因枪/将载体附接到金纳米粒子上的手段)并 且被提供。一种载体可能不仅是指病毒或酵母系统(例如,在感兴趣的核酸 可被可操作地连接到一个启动子上并且在其控制下(就表达而言,从而最终 提供加工的RNA)的情况下),而且指导核酸递送到宿主细胞中。虽然在本 文的方法中该载体可以是病毒载体并且AAV在这里是有利的,但是可以采用 如本文论述的其他病毒载体,如慢病毒。例如,杆状病毒可以用于昆虫细胞 中的表达。这些昆虫细胞进而可用于产生大量的其他载体,如适合于本发明 递送的AAV或慢病毒载体。还设想了一种递送本发明CRISPR酶的方法,该 方法包括向细胞递送编码该CRISPR酶的mRNA。应当理解的是,在某些实 施例中,该CRISPR酶被截短,和/或由少于一千个氨基酸或少于四千个氨基 酸组成,和/或是一种核酸酶或切口酶,和/或是经密码子优化的,和/或包含 一个或多个突变,和/或包含嵌合CRISPR酶,和/或如本文论述的其他选项。AAV和慢病毒载体是优选的。

[0111] 在某些实施例中,该靶序列在其3'端的侧翼或下游为适合于CRISPR酶 (典型地Cas,并且尤其是Cas9)的PAM。

[0112] 例如,适合的PAM是分别针对SpCas9或SaCas9酶(或衍生的酶)的 5'-NRG或5'-NNGRR。

[0113] 应当理解的是,SpCas9或SaCas9是来自或衍生自化脓链球菌或金黄色 葡萄球菌Cas9的那些。

[0114] 因此,本发明的目的在于,在本发明中不涵盖任何先前已知的产品、制 造该产品的过程、或使用该产品的方法,使得申请人保留和特此公开放弃任 何先前已知的产品、过程、或方法的权利。进一步指出的是,在本发明的范 围之内,本发明并不旨在涵盖任何产品、过程、或该产品的制造或使用该产 品的方法,其不符合USPTO(35 U.S.C.§112,第一段)

或EPO(EPC的第 83条)的书面说明和可实施性要求,使得申请人保留和特此公开放弃任何先 前描述的产品、制造该产品的过程、或使用该产品的方法的权利。

[0115] 应当指出的是,在本披露中尤其是在权利要求书和/或段落中,术语如 “包含(comprises)”、“包含(comprised)”、“包含(comprising)”等 等可具有在美国专利法中属于它的含义;例如,它们可以表示“包括 (includes)”、“包括(included)”、“包括(including)”等等;并且这 些术语如“基本上由……组成(consisting essentially of)”和“基本上由…… 组成(consists essentially of)”具有在美国专利法中归于它们的含义,例如, 它们允许未被清楚叙述的要素,但是排除在现有技术中发现或者影响本发明 的基本或新颖特征的要素。

[0116] 这些和其他实施例披露于以下详细说明中,或者据其显而易见并且由其 涵盖。

[0117] 本发明还涉及以下项目:

[0118] 1.一种通过操纵在感兴趣的基因组座位中的靶序列来修饰一种生物或 非人类生物的方法,该方法包括

[0119] 递送一种非天然存在的或工程化的组合物,该组合物包含:

[0120] A)-I.一种CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列,其中该多 核苷酸序列包含:

[0121] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

[0122] (b)一种tracr配对序列,和

[0123] (c)一种tracr序列,以及

[0124] II.一种编码CRISPR酶的多核苷酸序列,该CRISPR酶任选地包含至少 一个或多个核定位序列,

[0125] 其中(a)、(b)和(c)以5’到3’方向排列,

[0126] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0127] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,并且该编 码CRISPR酶的多核苷酸序列是DNA或RNA,

[0128] 或者

[0129] (B)I.多核苷酸,其包含:

[0130] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,和 (b)至少一个或多个tracr配对序列,

[0131] II.一种编码CRISPR酶的多核苷酸序列,以及

[0132] III.一种包含tracr序列的多核苷酸序列,

[0133] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0134] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,并且该编 码CRISPR酶的多核苷酸序列是DNA或RNA。

[0135] 2.如项目1所述的方法,其中该编码CRISPR酶的多核苷酸序列、指 导序列、tracr

配对序列或tracr序列的任一者或全部可以是RNA。

[0136] 3.如项目1或2所述的方法,其中对该编码CRISPR酶的序列、该指 导序列、tracr配对序列或tracr序列进行编码的多核苷酸可以是RNA并且可 以经由脂质体、纳米粒子、外泌体、微囊泡、或基因枪进行递送。

[0137] 4.如项目1到3中任一项所述的方法,其中这些多核苷酸被包含在含 有一种或多种载体的载体系统中。

[0138] 5.一种通过操纵在感兴趣的基因组座位中的靶序列来修饰一种生物或 非人类生物的方法,该方法包括

[0139] 递送一种包含病毒载体系统的非天然存在的或工程化的组合物,该病毒 载体系统包含一种或多种病毒载体,该一种或多种病毒载体可操作地编码一 种用于其表达的组合物,其中该非天然存在的或工程化的组合物包括:

[0140] (A)一种非天然存在的或工程化的组合物,该组合物包含一种载体系 统,该载体系统包含一种或多种载体,该一种或多种载体包含

[0141] I.一种第一调节元件,该第一调节元件可操作地连接到一种CRISPR-Cas 系统嵌合RNA(chiRNA)多核苷酸序列上,其中该多核苷酸序列包含

[0142] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

[0143] (b)一种tracr配对序列,和

[0144] (c)一种tracr序列,以及

[0145] II.一种第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶 的酶编码序列上,该CRISPR酶任选地包含至少一个或多个核定位序列,

[0146] 其中(a)、(b)和(c)以5’到3’方向排列,

[0147] 其中组分I和II位于该系统的相同或不同载体上,

[0148] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0149] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,

[0150] 或者

[0151] (B)一种非天然存在的或工程化的组合物,该组合物包含一种载体系 统,该载体系统包含一种或多种载体,该一种或多种载体包含

[0152] I.一种第一调节元件,该第一调节元件可操作地连接至

[0153] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,和

[0154] (b)至少一个或多个tracr配对序列,

[0155] II.一种第二调节元件,该第二调节元件可操作地连接至编码CRISPR酶 的酶编码序列,以及

[0156] III.一种第三调节元件,该第三调节元件可操作地连接至一种tracr序列,

[0157] 其中组分I、II和III位于该系统的相同或不同载体上,

[0158] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0159] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到

该tracr序列上的该tracr配对序列复合的CRISPR酶。

[0160] 6.如项目5所述的方法,其中这些病毒载体的一种或多种可以经由脂 质体、纳米粒子、外泌体、微囊泡、或基因枪进行递送。

[0161] 7.一种在对其有需要的受试者或非人类受试者中治疗或抑制由感兴趣 基因组座位内的靶序列的缺陷引起的病症的方法,该方法包括通过操纵该靶 序列来修饰该受试者或非人类受试者,并且其中该病症对于通过操纵该靶序 列的治疗或抑制是敏感的,该方法包括提供包括以下项的治疗:

[0162] 递送一种包含AAV或慢病毒载体系统的非天然存在的或工程化的组合 物,该载体系统包含一种或多种AAV或慢病毒载体,该一种或多种AAV或 慢病毒载体可操作地编码用于其表达的组合物,其中在表达时该靶序列由该 非天然存在的或工程化的组合物操纵,其中该非天然存在的或工程化的组合 物包括:

[0163] (A)一种非天然存在的或工程化的组合物,该组合物包含一种载体系 统,该载体系统包含一种或多种载体,该一种或多种载体包含

[0164] I.一种第一调节元件,该第一调节元件可操作地连接到一种CRISPR-Cas 系统嵌合RNA(chiRNA)多核苷酸序列上,其中该多核苷酸序列包含

[0165] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

[0166] (b)一种tracr配对序列,和

[0167] (c)一种tracr序列,以及

[0168] II.一种第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶 的酶编码序列上,该CRISPR酶包含至少一个或多个核定位序列,

[0169] 其中(a)、(b)和(c)以5’到3’方向排列,

[0170] 其中组分I和II位于该系统的相同或不同载体上,

[0171] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0172] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,

[0173] 或者

[0174] (B)一种非天然存在的或工程化的组合物,该组合物包含一种载体系 统,该载体系统包含一种或多种载体,该一种或多种载体包含

[0175] I.一种第一调节元件,该第一调节元件可操作地连接至

[0176] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,和

[0177] (b)至少一个或多个tracr配对序列,

[0178] II.一种第二调节元件,该第二调节元件可操作地连接至编码CRISPR酶 的酶编码序列,以及

[0179] III.一种第三调节元件,该第三调节元件可操作地连接至一种tracr序列,

[0180] 其中组分I、II和III位于该系统的相同或不同载体上,

[0181] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0182] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到

该tracr序列上的该tracr配对序列复合的CRISPR酶。

[0183] 8.如以上项目中任一项所述的方法,其中该方法是在体外、和/或离体 地进行的。

[0184] 9.如以上项目中任一项所述的方法,包括诱导表达。

[0185] 10.如以上项目中任一项所述的方法,其中该生物或受试者是一种真核 生物。

[0186] 11.如项目10所述的方法,其中该生物或受试者是一种非人类真核生 物。

[0187] 12.如项目1到11中任一项所述的方法,其中该生物或受试者是一种哺 乳动物或非人类哺乳动物。

[0188] 13.如项目4到8中任一项所述的方法,其中该病毒载体是一种AAV或 慢病毒载体。

[0189] 14.根据以上项目中任一项所述的方法,其中该CRISPR酶是一种 Cas9。

[0190] 15.根据以上项目中任一项所述的方法,其中该指导序列的表达是在T7 启动子控制下并且是由T7聚合酶的表达驱动的。

[0191] 16.一种递送如以上项目中任一项所述的CRISPR酶的方法,该方法包 括向一种细胞递送编码该CRISPR酶的mRNA。

[0192] 17.如项目1到16中任一项所述的方法,其中通过将编码该CRISPR酶 的mRNA递送到该细胞而将编码该CRISPR酶的该多核苷酸或酶编码序列递 送至该细胞。

[0193] 18.一种制备如项目7所述的AAV或慢病毒载体的方法,该方法包括 将含有编码该AAV或慢病毒的一种或多种核酸分子或基本上由该一种或多种 核酸分子组成的一种或多种质粒转染到AAV感染的或慢病毒感染的细胞中, 并且提供对于该AAV或慢病毒的复制和包装必须的AAV AAV或慢病毒rep 和/或cap和/或辅助核酸分子。

[0194] 19.一种制备用于在项目7所述的方法中使用的AAV或慢病毒载体的 方法,该方法包括将含有编码该AAV或慢病毒的一种或多种核酸分子或基本 上由该一种或多种核酸分子组成的一种或多种质粒转染到AAV感染的或慢病 毒感染的细胞中,并且提供对于该AAV或慢病毒的复制和包装必须的AAV AAV或慢病毒rep和/或cap和/或辅助核酸分子。

[0195] 20.如项目18或19所述的方法,其中对于该AAV或慢病毒的复制和包 装必须的AAV或慢病毒rep和/或cap是通过用一种或多种辅助质粒或一种或 多种辅助病毒转染这些细胞而提供的。

[0196] 21.如项目20所述的方法,其中该辅助病毒是一种痘病毒、腺病毒、慢 病毒、疱疹病毒或杆状病毒。

[0197] 22.如项目21所述的方法,其中该痘病毒是牛痘病毒。

[0198] 23.如项目18到22中任一项所述的方法,其中这些细胞是哺乳动物细 胞。

[0199] 24.如项目18到22中任一项所述的方法,其中这些细胞是昆虫细胞并 且该辅助病毒(如果存在的话)是杆状病毒。

[0200] 25.如项目1到15中任一项所述的方法,其中该靶序列在其3'端的侧翼 为或在其之后为5'-NRG(其中N是任何核苷酸)、或者其中该CRISPR酶是 (或来源于)属于下组的属,该组由以下各项组成:棒状杆菌属、萨特氏菌 属、军团杆菌属、密螺旋体属、产线菌属、优杆菌属、链球菌属、乳杆菌 属、支原体属、拟杆菌属、Flaviivola、黄杆菌属、Sphaerochaeta、固氮螺菌 属、葡糖醋杆菌属、奈瑟菌属、罗氏菌属、细小棒菌属(Parvibaculum)、葡 萄球菌属、Nitratifractor、支原体属以及弯曲杆菌属。

[0201] 26.一种如项目1-25中任一项中定义的组合物,用于在医药中或在疗法 中使用。

[0202] 27.一种如项目1-25中任一项定义的组合物,用于在一种通过操纵感兴 趣的基因组座位中的靶序列来修饰生物或非人类生物的方法中使用、或者在 一种治疗或抑制由感兴趣基因组座位中的靶序列的缺陷引起的病症的方法中 使用。

[0203] 28.如项目1-25中任一项中定义的组合物在离体基因或基因组编辑中的 用途。

[0204] 29.一种如项目1-25中任一项中定义的组合物在制造用于离体基因或基 因组编辑的药剂中、或用于在一种通过操纵感兴趣的基因组座位中的靶序列 来修饰生物或非人类生物的方法中使用、或者在一种治疗或抑制由感兴趣基 因组座位中的靶序列的缺陷引起的病症的方法中使用的用途。

[0205] 30.一种组合物,其包含:

[0206] A)-I.一种CRISPR-Cas系统嵌合RNA(chiRNA)多核苷酸序列,其中该多 核苷酸序列包含:

[0207] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,

[0208] (b)一种tracr配对序列,和

[0209] (c)一种tracr序列,以及

[0210] II.一种编码CRISPR酶的多核苷酸序列,该CRISPR酶任选地包含至少 一个或多个核定位序列,

[0211] 其中(a)、(b)和(c)以5'到3'方向排列,

[0212] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0213] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,并且该编 码CRISPR酶的多核苷酸序列是DNA或RNA,

[0214] 或者

[0215] (B)I.多核苷酸,其包含:

[0216] (a)一种指导序列,该指导序列能够杂交到真核细胞中的靶序列上,和

[0217] (b)至少一个或多个tracr配对序列,

[0218] II.一种编码CRISPR酶的多核苷酸序列,以及

[0219] III.一种包含tracr序列的多核苷酸序列,

[0220] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列 引导CRISPR复合物与该靶序列的序列特异性结合,并且

[0221] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的该指导序列、和 (2)杂交到该tracr序列上的该tracr配对序列复合的CRISPR酶,并且该编 码CRISPR酶的多核苷酸序列是DNA或RNA;

[0222] 用于在医药或疗法中使用;或者用于在一种通过操纵感兴趣的基因组座 位中的靶序列来修饰生物或非人类生物的方法中使用;或者用于在一种治疗 或抑制由感兴趣基因组座位中的靶序列的缺陷引起的病症的方法中使用;或 者用于在离体基因或基因组编辑中使用。

[0223] 31.如项目30所述的组合物,其中这些多核苷酸被包含在含有一种或多种载 体的载体系统中。

[0224] 附图简要说明

[0225] 本发明的新颖特征在所附权利要求书中具体提出。通过参考对说明性实 施例进行叙述的以下详细说明,将获得对本发明的特征和优点的更好理解, 在这些实施例中利用了本发明的原理,并且在这些附图中:

[0226] 图1显示了该CRISPR系统的示意模型。来自化脓链球菌(黄色)的 Cas9核酸酶经由合成的指导RNA(sgRNA)而被靶向基因组DNA,该指导 RNA由20个nt的指导序列(蓝色)和支架(红色)组成。与DNA靶(蓝 色)的指导序列碱基对,直接地在必要的5'-NGG原型间隔子邻近基序 (PAM;红紫色)的上游,并且Cas9在该PAM(红色三角形)的上游约3 bp介导了双链断裂(DSB)。

[0227] 图2A-F显示了一个示例性CRISPR系统、一个可能的作用机制、一个针 对在真核细胞中的表达的示例性适应性改变、以及评估核定位和CRISPR活 性的试验的结果。

[0228] 图3A-D显示了针对示例性靶标的SpCas9特异性的评估结果。

[0229] 图4A-G显示了一个示例性载体系统以及它在指导真核细胞中的同源重 组中的使用结果。

[0230] 图5提供了原型间隔子序列的表,并且总结了基于示例性脓链球菌和嗜 热链球菌CRISPR系统设计的原型间隔子靶标的修饰效率结果,这些CRISPR 系统具有针对人类和小鼠基因组中的座位的相应PAM。用Cas9以及pre- crRNA/tracrRNA或嵌合RNA转染细胞,在转染72小时之后进行分析。基于 来自指示细胞系的检验员(Surveyor)分析结果,计算了插入缺失(indel)百 分比(对于所有原型间隔子靶标N=3,误差为S.E.M.,N.D.表示使用Surveyor测定未检测到,并且N.T.表示在本研究中未检测)。

[0231] 图6A-C显示了对于Cas9介导的基因靶向而言的不同tracrRNA转录本的 比较。

[0232] 图7显示了对于双链断裂诱导的微插入和微缺失的检测的surveyor核酸 酶测定的示意图。

[0233] 图8A-B显示了用于CRISPR系统元件在真核细胞中的表达的示例性双顺 反子表达载体。

[0234] 图9A-C显示了在人类基因组中在邻近的化脓链球菌SF370座位1PAM (NGG)(图9A)和嗜热链球菌LMD9座位2PAM(NNAGAAW)(图 9B)之间的距离、以及对于每个PAM就染色体(Chr)而言的距离(图9C) 的直方图。

[0235] 图10A-D显示了一个示例性CRISPR系统、一个针对在真核细胞中的表 达的示例性适应性改变、以及评估CRISPR活性的试验的结果。

[0236] 图11A-C显示了用于靶向哺乳动物细胞中的基因组座位的CRISPR系统 的示例性操纵。

[0237] 图12A-B显示了在哺乳动物中的crRNA加工的Northern印迹分析的结 果。

[0238] 图13A-B显示了在人PVALB和小鼠Th座位中的原型间隔子的示例性选 择。

[0239] 图14显示了在人EMX1座位中的嗜热链球菌CRISPR系统的示例性原型 间隔子和相应PAM序列靶标。

[0240] 图15提供了用于Surveyor、RFLP、基因组测序、以及Northern印迹分 析的引物和探针序列的表。

[0241] 图16A-C显示了具有嵌合RNA的CRISPR系统的示例性操纵以及在真核 细胞中针对

系统活性的SURVEYOR分析的结果。

[0242] 图17A-B显示了在真核细胞中的CRISPR系统活性的SURVEYOR分析 的结果的图示。

[0243] 图18显示了使用UCSC基因组浏览器进行的在人类基因组中的一些化脓 链球菌Cas9靶位点的示例性可视化。

[0244] 图19A-D显示了系统发生分析的环状描绘,该系统发生分析揭示五个 Cas9家族,这五个家族包括三组大的Cas9(约1400个氨基酸)和两组小的 Cas9(约1100个氨基酸)。

[0245] 图20A-F显示了系统发生分析的线性描绘,该系统发生分析揭示五个Cas9 家族,这五个家族包括三组大的Cas9(约1400个氨基酸)和两组个小的 Cas9(约1100个氨基酸)。

[0246] 图21A-D显示了经由同源重组的基因组编辑。(a)SpCas9切口酶的示 意图,具有在RuvC I催化结构域中的D10A突变。(b)示意图,表示使用有 义或反义单链寡核苷酸作为修复模板在人EMX1座位处的同源重组(HR)。以上红色箭头表示sgRNA切割位点;用于基因分型的PCR引物(表J和K) 表示为右图中的箭头。(c)由HR修饰的区域的序列。d对于野生型(wt) 和切口酶(D10A)SpCas9-介导的在EMX1靶标1座位处的插入缺失而言的 SURVEYOR分析(n=3)。箭头指示预期片段大小的位置。

[0247] 图22A-B显示了用于SpCas9的单个载体设计。

[0248] 图23显示了代表Cas9直向同源物的长度分布的曲线图。

[0249] 图24A-M显示了在突变点位于SpCas9基因内的情况下的序列。

[0250] 图25A显示了条件型Cas9、Rosa26靶向载体图谱。

[0251] 图25B显示了组成型Cas9、Rosa26靶向载体图谱。

[0252] 图26显示了在组成型和条件型Cas9构建体中的重要元件的示意图。

[0253] 图27显示了递送和体内小鼠脑Cas9表达数据。

[0254] 图28显示了Cas9和嵌合RNA到细胞中的RNA递送(A)GFP报告基 因作为DNA或mRNA到Neuro-2A细胞中的递送。(B)针对作为RNA的 Icam2基因的Cas9和嵌合RNA的递送导致测试的两个间隔子之一的切割。 (C)针对作为RNA的F7基因的Cas9和嵌合RNA的递送导致测试的两个间 隔子之一的切割。

[0255] 图29显示DNA双链断裂(DSB)修复如何促进基因编辑。在易错非同 源末端连接(NHEJ)途径中,DSB的末端由内源DNA修复机构加工并重新 连接在一起,可导致在连接位点的随机插入缺失(indel)突变。发生在基因 编码区内的插入缺失突变可产生移码和提前终止密码子,导致基因敲除。可 替代地,处于质粒或单链寡脱氧核苷酸(ssODN)形式的修复模板可以被提 供为利用同源定向修复(HDR)途径,该途径允许高保真和精确编辑。

[0256] 图30A-C显示了在HEK和HUES9细胞中的HDR的预期结果。(a)具 有同源臂的靶向质粒或ssODN(有义或反义)可以用来编辑在靶基因组座位 处由Cas9(红色三角形)切割的序列。为了分析HDR的效率,我们在靶座 位中引入了HindIII位点(红条),用在同源区外部退火的引物进行PCR扩 增。用HindIII消化该PCR产物揭示了HDR事件的存在。(b)相对于感兴趣基因组的有义方向或反义方向(s或a)定向的ssODN可以与Cas9组合使 用,以实现有效的在该靶座位处的HDR介导的编辑。在该修饰(红条)的任 一侧面上,推荐40bp、并且优选90bp的最小同源区。(c)使用野生型 Cas9和Cas9切口酶(D10A)两者,显示了在EMX1座位中的HDR上的 ssODN的作用的实例。每个ssODN含有90bp的同源臂,这些同源臂侧接两 个限制性位点的12-bp插入片物。

[0257] 图31A-C显示了囊性纤维化δF508突变的修复策略。

[0258] 图32A-B:A显示了在FXN内含子1中的GAA重复扩展的示意图并且 B显示了所采用的使用该CRISPR/Cas系统切除该GAA扩展区的策略的示意 图。

[0259] 图33显示了Tet1-3和Dnmt1、3a和3b座位的有效的SpCas9介导的靶 向的筛选。通过使用不同的gRNA,在来自转染的N2A细胞的DNA上的 Surveyor测定证明了有效的DNA切割。

[0260] 图34显示了使用AAV1/2递送系统中的2载体系统进行的多元基因组靶 向策略。Tet1-3和Dnmt1、3a和3b gRNA在U6启动子的控制之下。GFP- KASH在人突触蛋白启动子的控制之下。限制性侧面显示了经由亚克隆的简 单gRNA置换策略。显示了侧翼为两个核定位信号(NLS)的、HA标记的 SpCas9。两种载体都以1:1比率通过AAV1/2病毒递送到脑中。

[0261] 图35显示了使用Surveyor测定进行的多元DNMT靶向载体#1的功能性 的验证。用DNMT靶向载体#1(+)和SpCas9编码载体共转染N2A细胞，以便检测SpCas9介导的DNMT基因家族座位的切割。阴性对照是仅有gRNA (-)。收获细胞用于DNA纯化,并且在转染后48小时进行下游加工。

[0262] 图36显示了使用Surveyor测定进行的多元DNMT靶向载体#2的功能性 的验证。用DNMT靶向载体#1(+)和SpCas9编码载体共转染N2A细胞，以便检测SpCas9介导的DNMT基因家族座位的切割。阴性对照是仅有gRNA (-)。收获细胞用于DNA纯化,并且在转染后48小时进行下游加工。

[0263] 图37显示了用于体内HA-SpCas9表达的短启动子和短polyA版本的示 意性概述。从L-ITR到R-ITR的编码区的大小显示在右边。

[0264] 图38显示了用于体内HA-SaCas9表达的短启动子和短polyA版本的示意 性概述。从L-ITR到R-ITR的编码区的大小显示在右边。

[0265] 图39显示了SpCas9和SaCas9在N2A细胞中的表达。在不同的短启动 子的控制之下并且具有短polyA(spA)序列的、HA标记的SpCas9和SaCas9 版本的代表性Western印迹。微管蛋白是加载对照。mCherry(mCh)是一种 转染对照。收获细胞并进一步加工,以便在转染后48小时进行Western印 迹。

[0266] 图40显示了有效的SaCas9介导的Tet3基因座位的靶向的筛选。通过使 用具有NNGGGT PUM序列的不同的gRNA,在来自转染的N2A细胞的DNA 上的Surveyor测定证明了有效的DNA切割。GFP转染的细胞和仅仅表达 SaCas9的细胞是对照。

[0267] 图41显示了HA-SaCas9在小鼠脑中的表达。用在人突触蛋白启动子控制 之下驱动HA-SaCas9表达的病毒注射到动物的齿状脑回中。在手术后2周将 动物处死。使用兔单克隆抗体C29F4(细胞信号传导)检测HA标签。用 DAPI染料将细胞核染成蓝色。

[0268] 图42显示了SpCas9和SaCas9在转导后7天的培养物中的皮层初级神经 元中的表达。在不同的启动子的控制之下并且具有bgh或短polyA(spA)序 列的、HA标记的SpCas9和SaCas9版本的代表性Western印迹。微管蛋白是 加载对照。

[0269] 图43显示了在用携带具有不同启动子的SpCas9和多元gRNA构建体 (显示在DNMT的最后一个小图上的实例)的AAV1粒子转导之后7天的初 级皮层神经元的LIVE/DEAD染色。将AAV转导之后的神经元与未转导的对 照神经元进行比较。红色细胞核指示渗透化处理的死细胞(小图的第二条 线)。活细胞标记为绿色(小图的第三条线)。

[0270] 图44显示了用携带具有不同启动子的SaCas9的AAV1粒子转导之后7 天的初级皮层神经元的LIVE/DEAD染色。红色细胞核指示渗透化处理的死细 胞(小图的第二条线)。活细胞标记为绿色(小图的第三条线)。

[0271] 图45显示了在用携带用于TET和DNMT基因座位的SpCas9和gRNA 多元体的AAV1病毒转导之后的神经元的形态学比较。未转导的神经元作为 对照示出。

[0272] 图46显示了在初级皮层神经元中使用Surveyor测定进行的多元DNMT 靶向载体#1的功能性的验证。用DNMT靶向载体#1和具有不同启动子的 SpCas9病毒共转导细胞,以便检测SpCas9介导的DNMT基因家族座位的切 割。

[0273] 图47显示了在脑中SpCas9切割的体内效率。用携带多元靶向DNMT家 族基因座位的gRNA的AAV1/2病毒连同在2个不同的启动子(小鼠Mecp2 和大鼠Map1b)控制之下的SpCas9病毒注射小鼠。在注射两周之后提取脑组 织,并且基于由来自gRNA多元构建体的突触蛋白启动子驱动的GFP表达, 使用FACS制备和分选细胞核。在gDNA提取之后,进行Surveyor测定。+指 示GFP阳性细胞核,且-为对照,来自相同动物的GFP阴性细胞核。在凝胶上的数字指示评估的SpCas9效率。

[0274] 图48显示了GFP-KASH标记的、来海马神经元的细胞核的纯化。用 GFP与KASH蛋白跨膜结构域的融合体标记细胞核膜的外层核膜(ONM)。 在立体定向手术和AAV1/2注射之后一周GFP在脑中强表达。进行密度梯度 离心步骤纯化来自完整脑的细胞核。显示了纯化的细胞核。通过 Vybrant® DyeCycle™ Ruby染色的染色质染色显示为红色,GFP标记的细胞核为绿色。 GFP+和GFP-细胞核的代表性FACS分布图(红紫色:Vybrant® DyeCycle™ Ruby染色,绿色:GFP)。

[0275] 图49显示了在小鼠脑中SpCas9切割的效率。用携带多元靶向TET家族 基因座位的gRNA的AAV1/2病毒连同在2个不同的启动子(小鼠Mecp2和 大鼠Map1b)控制之下的SpCas9病毒注射小鼠。在注射三周之后提取脑组 织,基于由来自gRNA多元构建体的突触蛋白启动子驱动的GFP表达,使用 FACS制备和分选细胞核。在gDNA提取之后,进行Surveyor测定。+指示 GFP阳性细胞核,-为对照,来自相同动物的GFP阴性细胞核。在凝胶上的 数字指示评估的SpCas9效率。

[0276] 图50显示了在培养物中的皮层神经元中的GFP-KASH表达。用携带靶 向TET基因座位的gRNA多元构建体的AAV1病毒转导神经元。由于KASH 结构域定位,最强信号位于细胞核周围。

[0277] 图51显示了(上)在指导RNA对之间的间隔的清单(如由两个PAM序 列的排列模式指示)。当与SpCas9(D10A)切口酶一起使用时,仅有满足模 式1、2、3、4的指导RNA对展现出插入缺失。(下)凝胶图像显示, SpCas9(D10A)与满足模式1、2、3、4的指导RNA对的组合导致靶位点中 的插入缺失的形成。

[0278] 图52显示了用来产生U6-指导RNA表达盒的U6反向引物序列的清单。 每个引物需要与U6正向引物“gcactgagggcctatttcccatgattc”配对,以便产生含 有U6和所希望的指导RNA的扩增子。

[0279] 图53显示了来自人Emx1座位的基因组序列图谱,显示了在图33中列 出的24个模式的位置。

[0280] 图54显示了(右侧)指示当由不同指导RNA对靶向的Cas9切口酶切割 之后存在可变5’突出端时在靶位点处的插入缺失信息的凝胶图像。(左侧) 为指示在右侧凝胶上的泳道数目和不同参数的表,这些参数包括鉴定所使用 的指导RNA对和在由Cas9切口酶切割之后存在的5’突出端的长度。

[0281] 图55显示了来自人Emx1座位的基因组序列图谱,其显示了导致图54 (右侧)的凝胶模式并且在实例35中进一步描述的不同的指导RNA对的位 置。

[0282] 图56显示了在注射编码Mecp2-HA-SpCas9和具有Syn-KASH-GFP的 3xgRNA-TETS的病毒之后8周在背侧和腹侧海马中的HA-SpCas9的染色。

[0283] 图57显示在3xgRNA病毒注射之后8周的Syn_GFP-KASH表达对于神 经元是特异性的(NeuN阳性细胞)而对于神经胶质细胞不是特异性的 (GFAP阳性)。

[0284] 图58显示了在齿状脑回(腹侧或背侧部分)中CRISPR介导的、TET和 DNMT的KD之后5周进行的行为试验,其显示出焦虑水平的增加和学习缺 陷。A)在高架十字迷宫试验期间在开放臂中花费的时间。B)旷场试验,测 量了在舞台的中心花费的时间与在角落中的时间。C)新物体识别试验,在熟 悉阶段之后3小时测量了结果。D)巴恩斯迷宫;在训练的3天之内找到逃跑 路线的效率。E)巴恩斯迷宫结果。F)在场景性恐惧条件化期间的木僵行 为。G)在场景性恐惧条件化期间的首次木僵发作的潜伏期。H)对于TET KD和DNMT KD的痕迹恐惧条件化结果(I)。对照-用SpCas9病毒和没有 gRNA的GFP-KASH构建体注射动物。TET-用SpCas9和编码针对Tet1、Tet2和Tet3的gRNA的构建体两者注射动物。DNMT-用SpCas9和编码针对 Dnmt1、Dnmt3a和Dnmt3b的gRNA的构建体两者注射动物。

[0285] 图59显示了在Mecp_SpCas9病毒注射之后8周与仅仅用Mecp2_SpCas9 病毒注射的对照动物相比较的在脑中的Tet座位的切割效率。

[0286] 图60显示了在Mecp_SpCas9病毒注射之后8周与仅仅用Mecp2_SpCas9 病毒注射的对照动物相比较的在脑中的Dnmt座位的切割效率。

[0287] 图61显示了在用编码Mecp2_SpCas9的病毒和靶向Dnmt座位的gRNA 进行脑立体定向注射之后8周在脑中的Dnmt3a染色。下部小图显示了在上部 小图中指示的ROI的放大。

[0288] 图62显示了在注射病毒之后4周在背侧海马中的Syn_HA-SaCas9的染 色。第一栏显示了仅仅用Sa-Cas9注射的动物,中间栏为用SaCas9和针对 TET座位的gRNA两者注射的动物,并且右栏代表仅仅用编码gRNA的病毒 注射的动物。SaCas9核定位取决于gRNA的存在。

[0289] 图63显示了在N2a细胞中的SpCas9。A)靶向载体与SpCas9表达载 体。B)在不同的启动子控制之下表达HA标记的SpCas9的N2a细胞的 Western印迹分析。C)Dnmt座位的切割效率。D)证明Dnmt3a的有效敲低 的Western印迹分析。e)Tet座位的切割效率。

[0290] 图64显示了在初级神经元中的SpCas9。A)在本研究中使用的SpCas9 克隆策略的示意性概述。用于有效包装到AAV递送系统中的短启动子和短 polyA。B)组合的多元靶向和核包膜标记策略的示意性概述。C)显示在 rMap1b和mMecp2启动子以及bGH和spA信号的控制之下的、HA标记的 SpCas9的表达的Western印迹分析。D)证明SpCas9和GFP-KASH在初级神 经元中的共表达的免疫细胞化学。SpCas9在mMecp2启动子控制之下在神经 元(Map1b,NeuN)中表达但是在星形神经胶质细胞(GFAP)中不表达。

[0291] 图65显示了在初级神经元中Dnmt3a的敲低。A)证明在用多元靶向载 体和mMecp2-

SpCas9靶向之后Dnmt3a的有效敲低的免疫细胞化学。B)在对 照和靶向的神经元中的Dnmt3a抗体染色的定量。C)证明Dnmt3a蛋白水平 降低的Western印迹分析。D)Western印迹分析的定量,证明在混合初级神 经元培养物(神经元和星形神经胶质)中大约75％的Dnmt3a蛋白水平的总体 敲低。

[0292] 图66显示了体内Dnmt3a的敲低。A)在Mecp_SpCas9病毒注射之后8 周与仅仅用Mecp2_SpCas9病毒注射的对照动物相比较的在脑中的Dnmt座位 的切割效率。B)Western印迹分析,显示在使用FACS进行细胞核分选之后 与对照细胞核(RubyDye阳性)相比较在靶向的神经元核(KASH-GFP阳 性)中Dnmt3a蛋白水平降低。

[0293] 图67显示了SaCas9在初级神经元中的表达。A)使用人突触蛋白启动子 和bGH信号的SaCas9表达载体的大小。B)SaCas9在初级神经元(NeuN) 中表达但是在星形神经胶质(GFAP)中不表达。C)SaCas9在gRNA不存在 下的核外定位。C'显示在C)中的SaCas9阳性神经元的更高放大倍数。D) SaCas9在gRNA存在下的核定位。D')显示在D)中的SaCas9阳性神经元的 更高放大倍数。E)证明HA标记的SaCas9和GFP-KASH的表达的Western 印迹分析。F)在AAV感染之后1周的Dnmt座位切割效率。

[0294] 图68显示了依赖于gRNA的SaCas9的核定位。A)共聚焦成像分析, 证明在初级神经元中在gRNA不存在下SaCas9的核外定位。B)SaCas9在 gRNA存在下的核定位。C)共聚焦图片A)的线扫描分析,显示在gRNA不 存在下SaCas9的核外定位(红色,SaCas9信号;蓝色,DAPI信号;绿色, GFP-KASH信号)。D)共聚焦图片B)的线扫描分析,显示在gRNA存在下 SaCas9的核定位(红色,SaCas9信号;蓝色,DAPI信号;绿色,GFP- KASH信号)。E)SaCas9和SpCas9在不具有(-)和具有(+)gRNA条件 下在N2a细胞中的亚细胞定位。在细胞质部分(微管蛋白阳性)中在250kDa 处的SaCas9信号,指示SaCas9在细胞质中二聚化。在gRNA存在下,可见 SaCas9蛋白移动到核部分(Sun2阳性)中。在100kDa处的SaCas9信号指示 SaCas9同聚体的gRNA依赖性形成和向细胞核中转运。相比之下,SpCas9主 要以同聚体存在,并且它的核定位不依赖于gRNA。

[0295] 图69显示了AAV-Sa-Cas9载体、肝脏特异性AAV-Sa-Cas9载体、以及 可替代的AAV-Sa-Cas9载体。

[0296] 图70显示了关于优化的CMV-SaCas9-NLS-U6-sgRNA载体(最后一次提 交的载体设计)的数据;新的数据比较了N'-末端与C'-末端标记的SaCas9, 并且显示了使用C'-末端NLS标记的切割效率提高。

[0297] 图71显示了示出通过新的Pcsk9靶标产生的插入缺失的SURVEYOR图 像。

[0298] 图72显示了SaCas9特异性:基于2个标准预测了全基因组脱靶位点 (GWOT):它们相对于预期SaCas9靶含有4个或更少的错配碱基,并且对 于SaCas9具有最小的限制性PAM(NNGRR)。用在靶位点(EMX1: TAGGGTTAGGGGCCCCAGGC)处具有它们的相应sgRNA的SpCas9或SaCas9转染HEK 293FT细胞,该靶位点具有CGGGGT作为PAM,使得它可 以被SpCas9(CGG)或被SaCas9(CGGGGT)切割。从细胞中收获DNA并 且通过Illumina测序分析在中靶和41个预测的脱靶座位处的插入缺失(遵循 来自徐(Hsu)等人的《自然生物技术》(Nature Biotech)2013方案和由大 卫·斯科特(David Scott)和乔什·温斯坦(Josh Weinstein)开发的数据分析流 程)。

[0299] 图73显示SaCas9在某些座位处比SpCas9可具有更高水平的脱靶活性。

[0300] 本文的这些图仅仅是出于说明性的目的，并且不一定是按比例绘制的。

[0301] 本发明的详细说明

[0302] 本发明涉及用于控制涉及序列靶向的基因表达的系统、方法、和组合物 的工程化和优化，该序列靶向例如涉及CRISPR-Cas系统及其组分的基因组干 扰或基因编辑。在有利的实施例中，该Cas酶是Cas9。

[0303] 本发明方法的一个优点在于，该CRISPR系统避免了脱靶结合及其所产 生的副作用。这是通过使用被安排为具有针对靶DNA的高度序列特异性的系 统而实现的。

[0304] Cas9

[0305] Cas9优化可以用来增强功能或者用来开发新的功能，人们可以产生嵌合 Cas9蛋白。申请人已经产生的例子提供在实例6中。可以通过组合来自不同 的Cas9同源物的片段而制成嵌合Cas9蛋白。例如，两种示例性嵌合Cas9蛋 白来自本文描述的Cas9。例如，申请人使St1Cas9的N-末端(来自这种蛋白 质的片段是粗体的)与SpCas9的C-末端融合。制造嵌合Cas9的益处包括下 列任一项或全部：毒性降低；在真核细胞中改进表达；特异性增强；蛋白质 分子量降低，例如，通过组合来自不同Cas9同源物的最小结构域使蛋白质更 小；和/或改变PAM序列要求。

[0306] 该Cas9可以用作通用的DNA结合蛋白。例如，如实例7中所示，通过 使负责切割该DNA靶的两条链的两个催化结构域(D10和H840)突变，申 请人使用Cas9作为通用的DNA结合蛋白。为了上调在靶座位处的基因转 录，申请人使转录激活结构结构域(VP64)融合到Cas9上。其他的转录激活 结构域是已知的。如实例17中所示，转录激活是可能的。还如实例17中所 示，使用结合到靶基因序列上的Cas9阻抑物(DNA结合域)，基因阻抑 (在β-连环蛋白基因的这种情况下)是可能的，由此阻抑其活性。

[0307] Cas9以及一种或多种指导RNA可以使用腺相关病毒(AAV)、慢病 毒、腺病毒或其他质粒或病毒载体类型进行递送，尤其是，使用来自以下文 献的配方和剂量：例如，美国专利号8,454,972(针对腺病毒的配方、剂 量)、8,404,658(针对AAV的配方、剂量)和5,846,946(针对DNA质粒的 配方、剂量)以及来自临床试验和关于涉及慢病毒、AAV、和腺病毒的临床试验的出版物。例如，对于AAV，给药途径、配方和剂量可以如美国专利号 8,454,972并且如涉及AAV的临床试验。对于腺病毒，给药途径、配方和剂 量可以如美国专利号8,404,658并且如涉及腺病毒的临床试验。对于质粒递 送，给药途径、配方和剂量可以如美国专利号5,846,946并且如涉及质粒的临 床试验。剂量可以基于或推断为平均70kg的个体，并且可以针对患者、受试 者、不同重量和物种的哺乳动物进行调整。给药频率在医学或兽医学从业者 (例如，医师、兽医师)的范围之内，其取决于常规因素，包括患者或受试 者的年龄、性别、一般健康状况、其他状况以及着手解决的特殊状况或症 状。

[0308] 可以将病毒载体注射到感兴趣的组织中。对于细胞类型特异性基因组修 饰，Cas9的表达可以由细胞类型特异性启动子驱动。例如，肝脏特异性表达 可以使用白蛋白启动子，而神经元特异性表达可以使用突触蛋白I启动子。

[0309] 转基因动物和植物

[0310] 还提供了转基因动物。优选实例包括含有Cas9(就编码Cas9的多核苷 酸或该蛋白质自身而言)的动物。小鼠、大鼠和兔是优选的。为了用如本文 示例的这些构建体产生转基因小鼠，人们可以将纯的线性DNA注射到来自假 孕雌性(例如，CB56雌性)的合子的前核中。

然后对首建者小鼠(founder) 进行鉴定、基因分型、并且与CB57回交。然后,例如通过桑格(Sanger)测 序将这些构建体克隆和任选地进行证实。在例如一个或多个基因在模型中被敲除的情况下,设想了敲除。然而,还设想了敲入(单独或以组合方式)。产生了一种示例性敲入Cas9小鼠,并且这是示例性的,但是Cas9敲入是优 选的。为了产生Cas9敲入小鼠,人们可以将相同的组成型或条件型构建体靶 向到Rosa26座位上,如本文所述(图25A-B和26)。可以修饰针对靶向 Rosa座位的、转让给桑加莫生物科学公司(Sangamo BioSciences,Inc.)的美 国专利公开号20120017290和20110265198的方法,以便利用本发明的 CRISPR Cas系统。在另一个实施例中,也可以修饰针对靶向Rosa座位的、转让给Cellectis公司的美国专利公开号20130236946的方法,以便利用本发 明的CRISPR Cas系统。

[0311] 条件型Cas9小鼠的实用性:申请人已经在293细胞中显示,可以通过与 Cre共表达而激活Cas9条件型表达构建体。申请人还显示,当表达Cre时, 正确靶向的R1 mESC可以具有活性Cas9。因为在Cas9之后为P2A肽切割序 列并且然后为EGFP,申请人通过观察EGFP而鉴定成功表达。申请人已经表 明在mESCs中的Cas9激活。这个相同的观念在于,什么使条件型Cas9小鼠 如此有用。申请人可以将其条件型Cas9小鼠与普遍表达Cre的小鼠(ACTB- Cre系)杂交并且可以得到在每个细胞中表达Cas9的小鼠。应当仅仅采用嵌 合的RNA的递送,以便诱导在胚胎小鼠或成年小鼠中的基因组编辑。有趣的 是,如果条件型Cas9小鼠与在组织特异的启动子下表达Cre的小鼠杂交,在 也表达Cre的组织中将会仅有Cas9。通过将嵌合的RNA递送到相同的组织 中,此途径可以用来仅仅在精确的组织中编辑基因组。

[0312] 如上提及的,还提供了转基因动物,正如转基因植物一样,尤其是作物 和藻类。这些转基因植物可用于在除了提供疾病模型之外的应用中。通过表 达例如比通常在野生型中可见的更高的蛋白质、碳水化合物、营养素或维生 素水平,这些应用可包括食物或饲料生产。在这方面,转基因植物,尤其是 豆类和块茎类,以及动物,尤其是哺乳动物,如家畜(奶牛、绵羊、山羊和 猪),而且还有禽类和食用昆虫,是优选的。

[0313] 转基因藻类或其他植物,如油菜,可以在植物油或生物燃料例如像醇类 (尤其是甲醇和乙醇)的生产中特别有用。这些可以被工程化为表达或过表 达高水平的油或醇类,以供在油或生物燃料行业中使用。

[0314] 腺相关病毒(AAV)

[0315] 就体内递送而言,AAV相比于其他病毒载体是有利的,这是由于几个原 因:

[0316] 低毒性(这可以是由于纯化方法不需要细胞粒子的超速离心所致,而超 速离心可能激活免疫反应)

[0317] 引起插入诱变的低概率,原因在于它未整合到宿主基因组中。

[0318] AAV具有4.5或4.75Kb的包装限制。这意味着Cas9以及启动子和转录 终止子必须都配合在同一个病毒载体中。大于4.5或4.75Kb的构建体将导致 病毒产生的显著降低。SpCas9是相当大的,其基因自身超过4.1Kb,使其难 于包装到AAV中。因此本发明的实施例包括利用更短的Cas9同源物。例 如:

[0319]

物种	Cas9 大小
白喉棒状杆菌（Corynebacter diphtheriae）	3252
凸腹真杆菌（Eubacterium ventriosum）	3321
巴氏链球菌（Streptococcus pasteurianus）	3390
香肠乳杆菌（Lactobacillus farciminis）	3378
Sphaerochaeta globus	3537
固氮螺菌属 B510	3504
重氮营养葡糖酸醋杆菌（Gluconacetobacter diazotrophicus）	3150
灰色奈瑟菌（Neisseria cinerea）	3246
肠道罗氏菌（Roseburia intestinalis）	3420
食清洁剂细小棒菌（Parvibaculum lavamentivorans）	3111
金黄色葡萄球菌	3159
Nitratifractor salsuginis DSM 16511	3396
红嘴鸥弯曲杆菌（Campylobacter lari） CF89-12	3009
嗜热链球菌 LMD-9	3396

[0320] 因此这些物种总体上是优选的Cas9物种。申请人已经示出了递送和体内 小鼠脑Cas9表达数据。

[0321] 介导体内基因组修饰的、将Cas9编码核酸分子例如DNA包装到病毒载 体中的两种方式是优选的：

[0322] 为了实现NHEJ介导的基因敲除：

[0323] 单病毒载体：

[0324] 含有两个或更多个表达盒的载体：

[0325] 启动子-Cas9编码核酸分子-终止子

[0326] 启动子-gRNA1-终止子

[0327] 启动子-gRNA2-终止子

[0328] 启动子-gRNA(N)-终止子(一直到载体的大小限制)

[0329] 双病毒载体：

[0330] 含有一个用于驱动Cas9表达的表达盒的载体1

[0331] 启动子-Cas9编码核酸分子-终止子

[0332] 含有一个或多个用于驱动一个或多个指导RNA表达的表达盒的载体2

[0333] 启动子-gRNA1-终止子

[0334] 启动子-gRNA(N)-终止子(一直到载体的大小限制)

[0335] 为了介导同源定向修复。除了上述单和双病毒载体途径之外，另外的载 体可以用

来递送同源定向修复模板。

[0336] 用来驱动Cas9编码核酸分子表达的启动子包括：

[0337] AAV ITR可以用作一种启动子：这对于消除另外的启动子元件(可在载 体中占用空间)的需要是有利的。空出来的另外的空间可以用来驱动另外的 元件的表达(gRNA，等等)。同样，ITR活性是较弱的，因此可以用来降低 由于Cas9的过表达所致的毒性。

[0338] 对于遍存表达，可以使用启动子：CMV、CAG、CBh、PGK、SV40、铁 蛋白重链或轻链，等等。

[0339] 对于脑表达，可以使用启动子：用于所有神经元的突触蛋白I、用于兴奋 性神经元的CaMKIIα、用于GABA能神经元的GAD67或GAD65或VGAT，等等。

[0340] 对于肝脏表达，可以使用白蛋白启动子。

[0341] 对于肺表达，可以使用SP-B。

[0342] 对于内皮细胞，可以使用ICAM。

[0343] 对于造血细胞，可以使用IFNβ或CD45。

[0344] 对于成骨细胞，可以使用OG-2。

[0345] 用来驱动指导RNA的启动子可以包括：

[0346] Pol III启动子，如U6或H1

[0347] 使用Pol II启动子和内含子盒来表达gRNA

[0348] 关于AAV，AAV可以是AAV1、AAV2、AAV5或任何其组合。可以相 对于有待被靶向的细胞而选择AAV；例如，可以选择用于靶向脑或神经元细 胞的AAV血清型1、2、5或杂交体或衣壳AAV1、AAV2、AAV5或其任何 组合；并且可以选择用于靶向心脏组织的AAV4。AAV8可用于递送到肝脏。以上启动子和载体是单独优选的。

[0349] RNA递送也是一种有用的体内递送方法。图27显示了递送和体内小鼠 脑Cas9表达数据。有可能使用脂质体或纳米粒子将Cas9和gRNA(并且，例 如，HR修复模板)递送到细胞中。因此，CRISPR酶如Cas9的递送和/或本 发明的RNA的递送可以处于RNA的形式并且经由微囊泡、脂质体或纳米粒 子来进行。例如，Cas9 mRNA和gRNA可以被包装到脂质体粒子中用于体内 递送。脂质体转染试剂例如来自生命技术公司(Life Technologies)的lipofectamine以及市售的其他试剂可以有效地将RNA分子递送到肝脏中。

[0350] 提高NHEJ或HR效率也有助于递送。优选的是，NHEJ效率通过共表达 末端加工酶(co-expressing end-processing enzyme)如Trex2而增强(杜米特 拉切(Dumitrache)等人《遗传学》(Genetics)2011年8月;188(4): 787-797)。优选的是，HR效率通过瞬时抑制NHEJ机构如Ku70和Ku86而 增加。HR效率也可以通过共表达原核或真核同源重组酶如RecBCD、RecA 而增加。

[0351] 本文描述了各种递送手段，并且进一步在这个部分中进行论述。

[0352] 病毒递送：该CRISPR酶，例如Cas9，和/或任何本发明的RNA，例如 指导RNA，可以使用腺相关病毒(AAV)、慢病毒、腺病毒或其他病毒载体 类型、或其组合进行递送。Cas9以及一种或多种指导RNA可以包装到一种或 多种病毒载体中。在一些实施例中，病毒载体可以，例如，通过肌肉注射递 送到感兴趣组织中，而有时经由静脉内、经皮、鼻内、经口、粘膜、或其他 递送方法进行病毒递送。这样的递送可以经由单剂量或多剂量来进行。本领 域技术人员理解的是，本文有待递送的实际剂量可以在很大程度上取决于多 种因素而变化，如载

体选择、靶细胞、生物、或组织、有待治疗的受试者的 一般状况、所寻求的转化/修饰的程度、给药途径、给药方式、所寻求的转化/ 修饰的类型,等等。

[0353] 这样的剂量可以进一步含有,例如,载体(水、盐水、乙醇、甘油、乳 糖、蔗糖、磷酸钙、明胶、葡聚糖、琼脂、果胶、花生油、芝麻油,等 等)、稀释剂、药学上可接受的载体(例如,磷酸盐缓冲盐水)、药学上可 接受的赋形剂、增强抗原性的佐剂、免疫刺激化合物或分子、和/或本领域已 知的其他化合物。本文的佐剂可以含有抗原被吸附到其上的矿物质(明矾、氢氧化铝、磷酸铝)的悬浮液;或油包水乳剂,其中抗原溶液被乳化在油中 (MF-59,弗氏不完全佐剂),有时包含有杀死的分枝杆菌(弗氏完全佐剂) 以进一步增强抗原性(抑制抗原的降解和/或引起巨噬细胞的移入)。佐剂还 包括免疫刺激分子,如细胞因子、共刺激分子,例如,免疫刺激DNA或 RNA分子,如CpG寡核苷酸。这样一种剂量配方可由本领域技术人员容易地 确定。该剂型可以进一步含有一种或多种药学上可接受的盐,例如像,无机 酸盐如盐酸盐、氢溴酸盐、磷酸盐、硫酸盐、等等;以及有机酸盐,如乙酸 盐、丙酸盐、丙二酸盐、苯甲酸盐等等。另外,本文也可以存在辅助物质, 如润湿剂或乳化剂、pH缓冲物质、凝胶或胶凝材料、调味剂、着色剂、微球 体、聚合物、悬浮剂等等。另外,也可以存在一种或多种其他常规药用成 分,如防腐剂、保湿剂、悬浮剂、表面活性剂、抗氧化剂、抗结剂、填充 剂、螯合剂、包衣剂、化学稳定剂等等,尤其是该剂型是可复原形式时。适 合的示例性成分包括微晶纤维素、羧甲基纤维素钠、聚山梨酯80、苯乙醇、 三氯叔丁醇、山梨酸钾、抗坏血酸、二氧化硫、没食子酸丙酯、对羟基苯甲 酸酯、乙基香兰素、甘油、苯酚、对氯酚、明胶、白蛋白和它们的组合。药 学上可接受的赋形剂的彻底论述可获自《雷明顿药物科学》(REMINGTON'S PHARMACEUTICAL SCIENCES)(马克出版公司,纽约,1991),通过引 用将其并入本文。

[0354] 在本文的一个实施例中,递送是经由腺病毒进行的,其可以是含有至少 1x 10^5个腺病毒载体粒子(也称为粒子单位,pu)的单次加强剂量。在本文 的一个实施例中,该剂量优选地是该腺病毒载体的至少约1x 10^6个粒子(例 如,约1x 10^6-1x 10^{12}个粒子),更优选至少约1x 10^7个粒子,更优选至少 约1x 10^8个粒子(例如,约1x 10^8-1x 10^{11}个粒子或约1x 10^8-1x 10^{12}个粒 子),并且最优选至少约1x 10^0个粒子(例如,约1*x 10^9-1x 10^{10}个粒子或约1x 10^9-1x 10^{12}个粒子),或者甚至至少约1x 10^{10}个粒子(例如,约1x 10^{10}-1x 10^{12}个粒子)。可替代地,该剂量包含不多于约1x 10^{14}个粒子,优 选不多于约1x 10^{13}个粒子,甚至更优选不多于约1x 10^{12}个粒子,甚至更优 选不多于约1x 10^{11}个粒子,并且最优选不多于约1x 10^{10}个粒子(例如,不 多于约1x 10^9个粒子)。因此,该剂量可以含有单剂量的腺病毒载体,其具 有例如、约1x 10^6粒子单位(pu)、约2x 10^6pu、约4x 10^6pu、约1x 10^7 pu、约2x 10^7pu、约4x 10^7pu、约1x 10^8pu、约2x 10^8pu、约4x 10^8pu、 约1x 10^9pu、约2x 10^9pu、约4x 10^9pu、约1x 10^{10}pu、约2x 10^{10}pu、约 4x 10^{10}pu、约1x 10^{11}pu、约2x 10^{11}pu、约4x 10^{11}pu、约1x 10^{12}pu、约2 x 10^{12}pu、或约4x 10^{12}pu的腺病毒载体。参见,例如,在2013年6月4日 授权的授予纳贝尔(Nabel)等人的美国专利号8,454,972 B2中的腺病毒载体 (通过引用并入本文)以及在其第29栏第36-58行的剂量。在本文的一个实 施例中,该腺病毒是经由多剂量递送的。

[0355] 在本文的一个实施例中,该递送是经由AAV进行的。用于针对人类的 AAV的体内递送的治疗有效剂量被认为处于含有从约1x 10^{10}到约1x 10^{10}个 功能AAV/ml溶液的从约20到约50ml的盐水溶液的范围内。该剂量可以调 整以便使治疗益处相对于任何副作用的平

衡。在本文的一个实施例中，AAV 剂量大致处于从约1x 10^5到1x 10^{50}个基因组AAV、从约1x 10^8到1x 10^{20}个基因组AAV、从约1x 10^{10}到约1x 10^{16}个基因组、或约1x 10^{11}到约1x 10^{16}个基因组AAV的浓度范围内。人类剂量可以是约1x 10^{13}个基因组 AAV。这样的浓度能以从约0.001ml到约100ml、约0.05到约50ml、或约 10到约25ml的载体溶液进行递送。通过建立剂量反应曲线的常规试验，本 领域普通技术人员可以容易地确立其他有效剂量。参见，例如，2013年3月 26日授权的授予哈加(Hajjar)等人的美国专利号8,404,658 B2，在第27栏 第45-60行。

[0356] 在本文的一个实施例中，该递送是经由质粒进行的。在这样的质粒组合 物中，该剂量应当是足以引出反应的质粒的量。例如，在质粒组合物中的质 粒DNA的适当量可以是从约0.1到约2mg，或从约1μg到约10μg。

[0357] 本文的剂量是基于平均70kg的个体。给药频率在医学或兽医学从业者 (例如医师、兽医师)或本领域熟练的科学家的范围之内。

[0358] 慢病毒

[0359] 慢病毒是复杂的反转录病毒，其具有在有丝分裂细胞和有丝分裂后细胞 两者中感染并表达其基因的能力。最为人熟知的慢病毒是人类免疫缺陷病毒 (HIV)，其利其他病毒的包膜糖蛋白来靶向广泛范围的细胞类型。

[0360] 慢病毒可以制备如下。在克隆pCasES10(含有慢病毒转移质粒骨架)之 后，将处于低传代数(p＝5)的HEK293FT接种在T-75烧瓶中，直到在转染 之前的一天在具有10％胎牛血清而没有抗生素的DMEM中50％汇合。在20 小时之后，培养基更换为OptiMEM(无血清)培养基，并且在4小时后进行 转染。将细胞用10μg的慢病毒转移质粒(pCasES10)和下列包装质粒转 染:5μg的pMD2.G(VSV-g假型)、和7.5μg的psPAX2 (gag/pol/rev/tat)。在具有阳离子脂质递送剂(50μL的Lipofectamine 2000 和100μl的Plus试剂)的4mL OptiMEM中进行转染。在6小时之后，培养 基更换为具有10％胎牛血清的无抗生素的DMEM。

[0361] 慢病毒可以如下纯化。在48小时后收获病毒上清液。首先清除上清液的 碎片并通过0.45μm的低蛋白结合(PVDF)滤膜进行过滤。然后将它们在超 速离心机中以24,000rpm旋转2小时。将病毒沉淀重新悬浮在50μl的 DMEM中在4C过夜。然后将它们等分，并且立即在-80C冷冻。

[0362] 在另一个实施例中，还考虑了基于马传染性贫血病毒(EIAV)的最小非 灵长类慢病毒载体，尤其是对于眼基因治疗而言(参见，例如，巴鲁谷安 (Balagaan)，《基因医学杂志》(J Gene Med)2006;8:275-285，2005年 11月21日在《威利在线期刊》(Wiley InterScience) (www.interscience.wiley.com).DOI:10.1002/jgm.845)在线公开。在另一个 实施例中，还考虑了RetinoStat®，一种经由视网膜下注射用于治疗湿型的年 龄相关性黄斑变性的、表达血管生成抑制蛋白(内皮抑素和血管抑素)的基 于马传染性贫血病毒的慢病毒基因治疗载体(参见，例如，宾利(Binley)等 人，《人类基因治疗》(HUMAN GENE THERAPY)23:980-991(2012年9 月))，其可以经修饰用于本发明的CRISPR-Cas系统。

[0363] 在另一个实施例中，自灭活慢病毒载体可以用于和/或适合于本发明的 CRISPR-Cas系统，该自灭活慢病毒载体具有靶向由HIV tat/rev共享的共有外 显子的siRNA、核仁定位TAR诱饵、和抗CCR5特异性锤头状核酶(参见， 例如，迪吉斯托(DiGiusto)等人(2010)《科学转化医学》(Sci Transl Med)2:36ra43)。可以收集最少2.5×10^6个CD34+细胞/每千

克患者体重，并且以2×10^{6}个细胞/ml的密度在X-VIVO 15培养基中(龙沙公司 (Lonza))预刺激16到20个小时，该培养基含有2mM L-谷氨酰胺、干细 胞因子(100ng/ml)、Flt-3配体(Flt-3L)(100ng/ml)、和促血小板生成 素(10ng/ml)(CellGenix公司)。可以用慢病毒以感染复数5在75-cm^{2}的 包被有纤连蛋白(25mg/cm^2)(重组人纤维蛋白片段(RetroNectin)，宝生 物工程株式会社(Takara Bio Inc.))的组织培养瓶中转导预刺激的细胞16到 24小时。

[0364] 慢病毒载体还披露于帕金森病的治疗中，参见，例如，美国专利公开号20120295960以及美国专利号7303910和7351585。慢病毒载体还披露于眼病 的治疗中，参见，例如，美国专利公开号20060281180、20090007284、US 20110117189;US 20090017543;US 20070054961、US 20100317109。还披露 了将慢病毒载体递送到脑，参见，例如，美国专利公开号US 20110293571; US 20110293571、US 20040013648、US 20070025970、US 20090111106和美 国专利号US 7259015。

[0365] RNA递送

[0366] RNA递送：该CRISPR酶，例如Cas9，和/或任何本发明的RNA，例如 指导RNA，也可以按RNA的形式递送。可以使用体外转录产生Cas9 mRNA。例如，可以使用含有下列元件的PCR盒来合成Cas9 mRNA：来自β 珠蛋白-polyA尾(一串120个或更多的腺嘌呤)的T7_启动子-科扎克 (kozak)序列(GCCACC)-Cas9-3’UTR。该盒可以用于经由T7聚合酶的 转录。也可以使用体外转录从含有T7_启动子-GG-指导RNA序列的盒来转录 指导RNA。

[0367] 为了增强表达并且降低毒性，可以使用假-U或5-甲基-C修饰该CRISPR 酶和/或指导RNA。

[0368] mRNA递送方法用于当前的肝脏递送是特别有希望的。具体地说，AAV8对于递送到肝脏是特别优选的。

[0369] 纳米粒子

[0370] 可以使用纳米粒子或脂质包膜同时递送CRISPR酶mRNA和指导RNA。

[0371] 例如，苏X(Su X)、弗里克J(Fricke J)、卡瓦纳DG(Kavanagh DG)、欧文DJ(Irvine DJ)(“使用脂质包封的pH响应性聚合物纳米粒子 的体外和体内mRNA递送”(“In vitro and in vivo mRNA delivery using lipid- enveloped pH-responsive polymer nanoparticles”)《分子药理学》(Mol Pharm.)2011年6月6;8(3):774-87.doi:10.1021/mp100390w。2011年4月1 日电子版描述了生物可降解的核-壳结构纳米粒子，其具有由磷脂双层壳包封 的聚(β-氨基酯)(PBAE)核。这些被开发为用于体内mRNA递送。该pH 响应性PBAE被选择为促进内体破裂，而该脂质表面层被选择为将聚阳离子 核的毒性降低到最低限度。因此，这些对于递送本发明的RNA是优选的。

[0372] 在一个实施例中，考虑了基于自组装生物粘附聚合物的纳米粒子，其可 应用于肽的经口递送、肽的静脉内递送以及肽的鼻递送，均递送到脑。其他 实施例，还考虑了例如疏水性药物的经口吸收和眼部递送。分子包膜技术涉 及被保护并递送到疾病部位的工程化聚合物包膜(参见，例如，马萨 (Mazza)M.等人《ACS纳米》(ACSNano)，2013.7(2):1016-1026;秀 (Siew)A.等人《分子药理学》(Mol Pharm)，2012.9(1):14-28;拉拉特萨(Lalatsa)A.等人《控释杂志》(J Contr Rel)，2012.161(2):523-36;拉拉 特萨(Lalatsa)A.等人，《分子药理学》(Mol Pharm)，2012.9(6):1665- 80;拉拉特萨(Lalatsa)A.等人《分子药理学》(Mol Pharm)，2012.9(6): 1764-74;加勒特(Garrett)N.L.等人《生物光电杂志》

(J Biophotonics)，2012.5(5-6):458-68;加勒特(Garrett)N.L.等人《拉曼光谱杂志》(J Raman Spect)，2012.43(5):681-688;阿哈默德(Ahmad)S.等人《皇家学会界面杂 志》(J Royal Soc Interface)2010.7:S423-33;乌切克布(Uchegbu)I.F.《药 物递送专家评论》(Expert Opin Drug Deliv)，2006.3(5):629-40;曲(Qu) X.等人《生物大分子》(Biomacromolecules)，2006.7(12):3452-9和乌切克 布(Uchegbu)I.F.等人《国际药学杂志》(Int J Pharm)，2001.224:185- 199)。考虑了约5mg/kg的剂量，呈单剂量或多剂量形式，这取决于靶组 织。

[0373] 在一个实施例中，可以使用由丹·安德森实验室(Dan Anderson's lab)在 MIT开发的可将RNA递送到癌细胞以便使肿瘤生长停止的纳米粒子/并且或 使其适用于本发明的CRISPR Cas系统。具体地说，安德森实验室开发了用于 新生物材料和纳米制剂的合成、纯化、表征、和配制的全自动化组合系统。参见，例如，阿拉比(Alabi)等人，《美国国家科学院院刊》(Proc Natl Acad Sci U S A.)2013年8月6日;110(32):12881-6;张(Zhang)等人，《先进材料》(Adv Mater.)2013年9月6日;25(33):4641-5;江(Jiang) 等人，《纳米通讯》(Nano Lett.)2013年3月13日;13(3):1059-64;卡拉 吉安尼斯(Karagiannis)等人，《ACS纳米》(ACS Nano.)2012年10月23 日;6(10):8484-7;怀特海德(Whitehead)等人，《ACS纳米》(ACS Nano.)2012年8月28日;6(8):6922-9和李(Lee)等人，《自然纳米技术》 (Nat Nanotechnol.)2012年6月3日;7(6):389-93。

[0374] 美国专利申请20110293703涉及类脂质化合物，这些化合物在多核苷酸 的给药中也是特别有用的，其可适用于递送本发明的CRISPR Cas系统。在一 个方面，氨基醇类脂质化合物与有待递送到细胞或受试者的药剂结合而形成 微粒子、纳米粒子、脂质体、或胶束。有待通过粒子、脂质体、或胶束递送 的药剂可以处于气体、液体、或固体的形式，并且该药剂可以是一种多核苷 酸、蛋白质、肽、或小分子。这些氨基醇类脂质化合物可以与其他氨基醇类 脂质化合物、聚合物(合成的或天然的)、表面活性剂、胆固醇、碳水化合 物、蛋白质、脂质、等等形成粒子。然后这些粒子可以任选地与药用赋形剂 结合而形成药物组合物。

[0375] 美国专利公开号0110293703也提供了制备氨基醇类脂质化合物的方法。使胺的一种或多种等效物与环氧化物封端化合物的一种或多种等效物在适当 条件下反应而形成本发明的氨基醇类脂质化合物。在某些实施例中，胺的所 有氨基基团与环氧化物封端化合物充分反应而形成叔胺。在其他实施例中， 胺的所有氨基基团未与环氧化物封端化合物完全反应形成叔胺，由此生成在 氨基醇类脂质化合物中的伯胺或仲胺。这些伯胺或仲胺照原样留下或者可以 与另一种亲电体如一种不同的环氧化物封端化合物反应。正如本领域技术人 员将理解的，胺与未过量的环氧化物封端化合物反应将产生多种不同的具有 不同数目的尾部的氨基醇类脂质化合物。某些胺类可以用两个环氧化物衍生 的化合物尾部将其完全功能化，而其他分子用环氧化物衍生的化合物尾部将 不会被完全功能化。例如，二胺或多胺可包括离开该分子的不同氨基部分的 一个、二个、三个、或四个环氧化物衍生的化合物尾部，从而产生伯胺、仲 胺、和叔胺。在某些实施例中，并不是所有氨基基团都被完全功能化。在某 些实施例中，使用相同类型的环氧化物封端化合物中的两种。在其他实施例中，使用两种或更多种不同的环氧化物封端化合物。氨基醇类脂质化合物的 合成是用或不用溶剂进行的，并且该合成可以在范围从30-100C，优选在大致 50-90C的较高温度下进行。任选地，可以将制备的氨基醇类脂质化合物纯 化。例如，可以纯化氨基醇类脂质化合物的

混合物而产生具有特定数目的、环氧化物衍生的化合物尾部的氨基醇类脂质化合物。或者,该混合物可以被 纯化而产生特定的立体异构体或区域异构体。也可以使用烷基卤化物(例 如,碘甲烷)或其他烷化剂将这些氨基醇类脂质化合物烷化,和/或它们可以 被酰化。

[0376] 美国专利公开号0110293703还提供了通过发明方法制备的氨基醇类脂质 化合物的文库。使用涉及液体处理器、机器人、微量滴定板、计算机、等等 的高通量技术,可以制备和/或筛选这些氨基醇类脂质化合物。在某些实施例 中,筛选了这些氨基醇类脂质化合物的将多核苷酸或其他药剂(例如,蛋白 质、肽、小分子)转染到细胞中的能力。

[0377] 美国专利公开号20130302401涉及已经使用组合聚合制备的一类聚(β- 氨基醇)(PBAAs)。这些发明的PBAA可以在生物技术和生物医学应用中 用作涂层(如用于医疗装置或植入物的薄膜或多层薄膜的涂层)、添加剂、 材料、赋形剂、生物防污剂(non-biofouling agent)、微图像化剂 (micropatterning agent)、以及细胞封装剂(cellular encapsulation agent)。 当用作表面涂层时,这些PBAA在体外和体内均引出不同水平的炎症,这取 决于它们的化学结构。这类材料的巨大化学多样性允许我们鉴定出体外抑制 巨噬细胞激活的聚合物涂层。此外,在羧化聚苯乙烯微粒的皮下移植之后, 这些涂层减少了炎症细胞的募集,并且减轻了纤维化。这些聚合物可以用来 形成用于细胞封装的聚电解质复合物胶囊。本发明还可具有许多其他的生物 应用,如抗微生物涂层、DNA或siRNA递送、以及干细胞组织工程。美国专 利公开号20130302401的传授内容可以适用于本发明的CRISPR Cas系统。

[0378] 在另一个实施例中,考虑了脂质纳米粒子(LNP)。具体地说,封装在 脂质纳米粒子中的抗转甲状腺素蛋白小干扰RNA(参见,例如,科尔贺 (Coelho)等人,《新英格兰医学杂志》(N Engl J Med)2013;369:819- 29)可以适用于本发明的CRISPR Cas系统。考虑到静脉内给予约0.01到约1 mg/kg体重的剂量。考虑了降低输注相关反应的风险的用药,如考虑到地塞米 松、对乙酰氨基酚(acetampinophen)、苯海拉明或西替利嗪、以及雷尼替 丁。考虑了约0.3mg/千克的多剂量,每4周一次,五个剂量。

[0379] LNP已经显示在将siRNA递送到肝脏中是高度有效的(参见,例如,塔 韦内罗(Tabernero)等人,《癌症发现》(Cancer Discovery),2013年4 月,第3卷,第4期,第363-470页),并且因此被考虑用于递送CRISPR Cas到肝脏。考虑了6mg/kg的LNP的约四个剂量的用量,每两周一次。塔韦 内罗(Tabernero)等人证明,在以0.7mg/kg给予LNP的前2个周期之后, 观察到肿瘤消退,并且在6个周期结束之后,患者已经实现了部分应答,具 有淋巴结转移完全消退以及肝脏肿瘤的显著萎缩。在此患者中给予40个剂量 之后获得完全应答,在接受经过26个月的剂量之后其保持缓解和完全治疗。具有RCC和在用VEGF途径抑制剂进行的在先治疗之后进展的包括肾脏、肺、以及淋巴结在内的肝外部位疾病的两位患者在所有部位的疾病都保持稳 定大约8到12个月,并且一位具有PNET和肝转移的患者继续在18个月(36个剂量)的延伸研究中保持疾病稳定。

[0380] 然而,必须将LNP的电荷考虑在内。当阳离子脂质与带负电的脂质结合 时,诱导促进细胞内递送的非双层结构。由于带电荷的LNP在静脉注射之后 迅速从循环中清除,开发了具有低于7的pKa值的可电离阳离子脂质(参 见,例如,罗辛(Rosin)等人,《分子治疗》(Molecular Therapy),第19 卷,第12期,第1286-2200页,2011年12月)。带负电荷的聚合物如 siRNA寡核苷酸可以低pH值(例如,pH 4)加载到LNP中,在此pH时可电 离脂质展示出正

电荷。然而，在生理学pH值时，LNP展现出与更长的循环时 间相容的低表面电荷。已经关注了四种可电离阳离子脂质，即1,2-二亚油酰 基-3-二甲基铵-丙烷(DLinDAP)、1,2-二亚油基氧基-3-N,N-二甲基氨基丙烷 (DLinDMA)、1,2-二亚油基氧基-酮基-N,N-二甲基-3-氨基丙烷 (DLinKDMA)、以及1,2-二亚油基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLinKC2-DMA)。已经表明，含有这些脂质的LNP siRNA系统在体内肝 细胞中展现出显著不同的基因沉默特性，具有根据采用因子VII基因沉默模 型的DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP系列而变化的潜 能(参见，例如，罗辛(Rosin)等人，《分子治疗》(Molecular Therapy)，第19卷，第12期，第1286-2200页，2011年12月)。可以考虑 1μg/ml水平的剂量，尤其是对于含有DLinKC2-DMA的配制品而言。

[0381] LNP的制备和CRISPR Cas封装可以使用/和或改编自罗辛(Rosin)等人 的《分子治疗》(Molecular Therapy)，第19卷，第12期，第1286-2200 页，2011年12月)。阳离子脂质1,2-二亚油酰基-3-二甲基铵-丙烷 (DLinDAP)、1,2-二亚油基氧基-3-N,N-二甲基氨基丙烷(DLinDMA)、 1,2-二亚油基氧基酮基-N,N-二甲基-3-氨基丙烷(DLinK-DMA)、1,2-二亚油基-4-(2-二甲基氨基乙基)-[1,3]-二氧戊环(DLinKC2-DMA)、(3-o-[2″-(甲氧 基聚乙二醇2000)琥珀酰]-1,2-二肉豆蔻酰基-sn-二肉豆蔻酰基(PEG-S- DMG)、以及R-3-[(ω-甲氧基-聚(乙二醇)2000)氨甲酰]-1,2-二肉豆蔻酰氧基丙 基-3-胺(PEG-C-DOMG)可以由Tekmira制药公司(Tekmira Pharmaceuticals)(温哥华，加拿大)提供或者合成。胆固醇可以购自西格玛 公司(圣路易斯，密苏里州)。特异性CRISPR Cas RNA可以封装在含有DLinDAP、DLinDMA、DLinK-DMA、和DLinKC2-DMA的LNP中(阳离子 脂质:DSPC:CHOL:PEGS-DMG或PEG-C-DOMG的摩尔比为40:10:40: 10)。在需要时，可以结合0.2%SP-DiOC18(英杰公司，伯灵顿，加拿大) 来评估细胞摄取、细胞内递送、和生物分布。通过将由阳离子脂质:DSPC:胆固醇:PEG-c-DOMG(40:10:40:10摩尔比)组成的混合物溶解在乙醇中 直到最终脂质浓度为10mmol/l来进行封装。可以将脂质的这种乙醇溶液逐滴 加入到pH 4.0的50mmol/l柠檬酸盐中而形成多层囊泡，从而产生30% (vol/vol)乙醇的最终浓度。在使用挤出机(北方脂质公司(Northern Lipids)，温哥华，加拿大)通过两个重叠的80nm Nuclepore聚碳酸酯滤膜 挤出多层囊泡之后，可以形成大的单层囊泡。可以通过如下步骤实现封装：将溶解在含有30%乙醇(vol/vol)的pH 4.0的50mmol/l柠檬酸盐中的2 mg/ml的RNA逐滴加入到挤出的形成的大单层囊泡中，并且在31℃孵育30 分钟，伴随持续混合直到最终的RNA/脂质重量比为0.06/1(wt/wt)。通过使 用Spectra/Por 2再生纤维素透析膜在pH为7.4的磷酸盐缓冲盐水(PBS)中 透析16小时进行乙醇的去除以及配制缓冲液的中和。使用NICOMP 370型粒 径分析仪、囊泡/强度模式、以及高斯拟合，通过动态光散射，可以测定纳米 粒子粒径分布(Nicomp粒径分析仪公司，圣巴巴拉市，加利福尼亚州)。对 于所有三个LNP系统的粒径可以是约70nm的直径。可以通过使用 VivaPureD MiniH柱(赛多利斯斯泰迪生物技术公司(Sartorius Stedim Biotech))从分析前后收集的样品中去除游离siRNA来确定siRNA封装效 率。从洗脱的纳米粒子提取封装的RNA并且将其在260nm定量。通过使用 来自美国瓦克化学公司(Wako Chemicals USA)(里士满(Richmond)，弗 吉尼亚州)的胆固醇E酶法测定法测量囊泡中的胆固醇含量，确定了siRNA 与脂质的比率。

[0382] 大LNP的制备可以使用/和或改编自罗辛(Rosin)等人的《分子治疗》 (Molecular Therapy)，第19卷，第12期，第1286-2200页，2011年12 月。可以在含有50:10:38.5的摩尔比

的DLinKC2-DMA、DSPC、和胆固醇 的乙醇中制备脂质预混物溶液(20.4mg/ml总脂质浓度)。可以按照0.75:1 的摩尔比(乙酸钠:DLinKC2-DMA)将乙酸钠添加到脂质预混物中。随后可以通过将该混合物与1.85倍体积的柠檬酸盐缓冲液(10mmol/l,pH 3.0)在 剧烈搅拌下合并而使脂质水合,从而导致在含有35%的在水性缓冲液中的自 发脂质体形成。可以在37℃孵育该脂质体溶液以允许粒径的时间依赖性增 加。可以通过动态光散射(纳米粒径电位分析仪(Zetasizer Nano ZS),马尔 文仪器公司(Malvern Instruments),乌斯特郡(Worcestershire),英国)在 孵育的不同时间去除等分试样以研究脂质体大小的变化。一旦实现所希望的 粒径,可以将水性PEG脂质溶液(储备溶液=在35%(vol/vol)乙醇中的10mg/ml PEG-DMG)添加到该脂质体混合物中,以便产生3.5%总脂质的最终 PEG摩尔浓度。在添加PEG-脂质之后,这些脂质体应该其大小,有效抑制进 一步生长。然后以大约1:10(wt:wt)的siRNA与总脂质比率将RNA添加 到空脂质体,然后在37℃孵育30分钟以形成加载的LNP。随后将该混合物 在PBS中透析过夜,并且用0.45-μm的注射器过滤器。

[0383] 球形核酸(SNA™)构建体和其他纳米粒子(尤其是金纳米粒子)也被 考虑为将CRISPR/Cas系统递送到预期靶标的手段。重要数据表明,基于核酸 功能化的金纳米粒子的AuraSense治疗性球形核酸(SNA™)构建体优于替代 平台,其基于多个关键成功因素,如:

[0384] 高体内稳定性。由于它们的密集加载,大多数负荷(DNA或siRNA)保 持接合到细胞内部的构建体上,从而赋予核酸稳定性和对酶促降解的抗性。

[0385] 可递送性。对于所研究的所有细胞类型(例如,神经元、肿瘤细胞系, 等等),这些构建体证明了在不需要载体或转染剂时99%的转染效率。

[0386] 治疗靶向性。这些构建体的独特靶结合亲和力和特异性允许对于匹配的 靶序列的精密特异性(即,限制性脱靶效应)。

[0387] 优良效能。这些构建体显著胜过一流的常规转染试剂(Lipofectamine 2000和细胞转染剂(Cytofectin))。

[0388] 低毒性。这些构建体可以进入多种培养的细胞、原代细胞、和组织中, 没有明显的毒性。

[0389] 没有显著的免疫反应。这些构建体在全局基因表达上引起最小的变化, 正如通过全基因组微阵列研究和细胞因子特异性蛋白质分析所测量。

[0390] 化学可修整性。任何数目的单个或组合药剂(例如,蛋白质、肽、小分 子)可以用来修整这些构建体的表面。

[0391] 对于基于核酸治疗剂的这个平台可以适用于许多疾病状态,包括炎症和 感染性疾病、癌症、皮肤病症和心血管疾病。

[0392] 可引用的文献包括:卡特勒(Cutler)人,《美国化学会志》(J.Am. Chem.Soc.)2011 133:9254-9257,郝(Hao)等人,Small.2011 7:3158- 3162,张(Zhang)等人,《ACS纳米》(ACS Nano.)2011 5:6962-6970, 卡特勒(Cutler)等人,《美国化学会志》(J.Am.Chem.Soc.)2012 134: 1376-1391,杨(Young)等人,《纳米通讯》(Nano Lett.)2012 12:3867- 71,郑(Zheng)等人,《美国国家科学院院刊》(Proc.Natl.Acad.Sci. USA.)2012 109:11975-80,米尔金(Mirkin),《纳米医学》 (Nanomedicine)2012 7:635-638张(Zhang)等人,《美国化学会志》(J. Am.Chem.Soc.)2012 134:16488-1691,因特劳布(Weintraub),《自然》 (Nature)2013 495:S14-S16,崔(Choi)等人,《美国国家科学院院刊》 (Proc.Natl.Acad.Sci.USA.)2013

110(19):7625-7630,詹森(Jensen)等 人,《科学转化医学》(Sci Transl Med)5,209ra152(2013)以及米尔金 (Mirkin)等人,Small,doi.org/10.1002/smll.201302143。

[0393] 具有siRNA的自组装纳米粒子可以用被聚乙二醇化的聚乙烯亚胺 (PEI)构建,其中Arg-Gly-Asp(RGD)肽配体附接在聚乙二醇(PEG)的 远端,例如,作为靶向表达整合素的肿瘤新血管系统的手段和用来递送抑制 血管内皮生长因子受体2(VEGF R2)表达以及由此抑制肿瘤血管形成的 siRNA的手段(参见,例如,施弗勒斯(Schiffelers)等人,《核酸研究》(Nucleic Acids Research),2004,第32卷,第19期)。通过将等体积的阳 离子聚合物水溶液和核酸水溶液混合从而产生在2到6范围上的可电离氮(聚合物)相比磷酸盐(核酸)的净摩尔过量,从而制备纳米束。在阳离子 聚合物与核酸之间的静电相互作用导致聚复合物的形成,其具有约100nm的 平均粒径分布,此后称为纳米束。设想了CRISPR Cas的约100到200mg的 剂量,用于施弗勒斯(Schiffelers)等人的自组装纳米粒子中的递送。

[0394] 巴特利特(Bartlett)等人的纳米束(PNAS,2007年9月25日,第104 卷,第39期)也可以应用于本发明。通过将等体积的阳离子聚合物水溶液和 核酸水溶液混合从而产生在2到6范围上的可电离氮(聚合物)相比磷酸盐 (核酸)的净摩尔过量,从而制备巴特利特(Bartlett)等人的纳米束。在阳 离子聚合物与核酸之间的静电相互作用导致聚复合物的形成,其具有约100 nm的平均粒径分布,此后称为纳米束。巴特利特(Bartlett)等人的DOTA- siRNA合成如下:1,4,7,10-四氮杂环十二烷-1,4,7,10-四乙酸单(N-羟基琥珀酰 亚胺酯)(DOTA-NHSester)订购自Macrocyclics公司(达拉斯(Dallas),德 克萨斯州)。将碳酸盐缓冲液(pH 9)中的具有100倍摩尔过量的DOTA- NHS-酯的胺修饰的RNA有义链加入到微量离心管中。通过在室温搅拌4小 时使这些内容物反应。将该DOTA-RNA有义缀合物用乙醇沉淀,重新悬浮在 水中,并且退火到未修饰的反义链上而产生DOTA-siRNA。所有液体用Chelex-100(伯乐公司(Bio-rad),赫库斯(Hercules),加利福尼亚州)预 处理,以便去除微量金属污染物。通过使用含有环糊精的聚阳离子可以形成 Tf靶向和非靶向的siRNA纳米粒子。典型地,纳米粒子以3(+/-)的进料比 和0.5g/升的siRNA浓度在水中形成。用Tf(金刚烷-PEG-Tf)修饰在靶向纳 米粒子表面上的百分之一的金刚烷-PEG分子。将纳米粒子悬浮在用于注射的 5%(wt/vol)的葡萄糖载体溶液中。

[0395] 戴维斯(Davis)等人(《自然》,第464卷,2010年4月15日)使用 靶向的纳米粒子递送系统进行了siRNA临床试验(临床试验登记号NCT 00689065)。在21天周期的第1、3、8和10天通过30分钟的静脉输注对患 有标准护理治疗难治的实体癌的患者给予靶向纳米粒子剂量。这些纳米粒子 包含合成递送系统,该系统含有:(1)一种线性的、基于环糊精的聚合物(CDP),(2)一种展示在纳米粒子的外部上的用于接合癌细胞表面上的 TF受体(TFR)的人转铁蛋白蛋白(TF)靶向配体,(3)一种亲水聚合物 (用来促进在生物流体中的纳米粒子稳定性的聚乙二醇(PEG)),以及 (4)被设计为降低RRM2(在临床中使用的序列,先前表示为siR2B+5)表 达的siRNA。TFR已经久为所知在恶性细胞中被下调,并且RRM2是一种确 立的抗癌靶标。已经显示这些纳米粒子(临床版本表示为CALAA-01)在非 人类灵长动物中的多剂量研究中良好耐受。虽然已经通过脂质体递送向患有 慢性粒细胞白血病的单一患者给予了siRNA,但是戴维斯等人的临床试验是 初期人类试验,该试验用一个靶向的递送系统全身性地递送siRNA并且治疗 患有实体癌的患者。为了确定该靶向的递送系统是否能够将功能性siRNA有 效递送到人类肿瘤,戴维斯等人研究了来自三个不同的剂量组群的三位患

者 的活组织检查;患者A、B和C均患有转移性黑素瘤并且分别接受了18、24 和30mg m^{-2}siRNA的CALAA-01剂量。还可以针对本发明的CRISPR Cas系 统考虑相似的剂量。用含有一种线性的基于环糊精的聚合物(CDP)、一种 展示在纳米粒子的外部上的用于接合癌细胞表面上的TF受体(TFR)的人转 铁蛋白(TF)靶向配体和/或一种亲水聚合物(例如,用来促进在生物流体中 的纳米粒子稳定性的聚乙二醇(PEG))的纳米粒子,可以实现本发明的递 送。

[0396] 外泌体

[0397] 外泌体是可以向小鼠脑中递送短干扰(si)RNA的转运RNA和蛋白质的内 源纳米囊泡。为了降低免疫原性,阿尔瓦雷斯-尔维蒂(Alvarez-Erviti)等人 (2011,《自然生物技术》29:341)使用了用于外泌体产生的自我衍生的树 突细胞。通过将树突细胞工程化为表达Lamp2b(一种外泌体膜蛋白,融合到 神经元特异性RVG肽3上)而实现了靶向。通过电穿孔使纯化的外泌体加载 外源siRNA。静脉注射的RVG靶向的外泌体特异性地将GAPDH siRNA递送 到脑中的神经元、小胶质细胞、少突神经胶质细胞,导致特异性的基因敲 除。预暴露于RVG外泌体未减弱敲低,并且在其他组织中未观察到非特异性 摄取。通过BACE1的强的mRNA(60%)和蛋白质(62%)敲低证明了外泌 体介导的siRNA递送的治疗潜能,BACE1是一种阿尔茨海默病中的治疗靶 标。

[0398] 为了获得免疫惰性的外泌体池,阿尔瓦雷斯-尔维蒂(Alvarez-Erviti)等 人收获了来自具有同源主要组织相容性复合体(MHC)单体型的近交 C57BL/6小鼠的骨髓。由于未成熟树突细胞产生大量的缺乏T细胞激活剂如 MHC-II和CD86的外泌体,阿尔瓦雷斯-尔维蒂(Alvarez-Erviti)等人选择具 有粒细胞/巨噬细胞集落刺激因子(GM-CSF)的树突细胞持续7天。次日, 使用良好建立的超速离心方案从培养上清从纯化外泌体。这些产生的外泌体 在物理上是同质的,具有直径为80nm的粒径分布峰,正如通过纳米粒子跟 踪分析(NTA)和电子显微镜检查所测定。阿尔瓦雷斯-尔维蒂(Alvarez- Erviti)等人获得了6-12μg的外泌体(基于蛋白质浓度测量的)/每10^6个细 胞。

[0399] 其次,阿尔瓦雷斯-尔维蒂(Alvarez-Erviti)等人研究了使用适用于纳米 级应用的穿孔方案给修饰的外泌体加载外源负荷的可能性。由于电穿孔对于 纳米级的膜粒子尚未良好表征,使用非特异性Cy5标记的siRNA用于电穿孔 方案的经验优化。在外泌体超速离心和溶解之后测定了封装的siRNA的量。 在400V和125μF的电穿孔导致siRNA的最好保留并且用于所有的后续实 验。

[0400] 阿尔瓦雷斯-尔维蒂(Alvarez-Erviti)向正常C57BL/6小鼠给予被包封在 150μg的RVG外泌体中的150μg的每种BACE1 siRNA并且将敲低效率与四 个对照进行比较:未处理的小鼠、仅仅用RVG外泌体注射的小鼠,用与一种 体内阳离子脂质体试剂络合的BACE1 siRNA注射的小鼠、以及用与RVG-9R 络合的BACE1 siRNA注射的小鼠,该RVG肽与静电结合到siRNA上的9个 D-精氨酸缀合。在给药之后3天,分析皮层组织样品,并且在siRNA-RVG-9R处理的和siRNARVG外泌体处理的小鼠中均观察到显著的蛋白质敲低 (45%,$P<0.05$,相对于62%,$P<0.01$),这是由于显著的BACE1 mRNA 水平降低(分别为66%[+或-]15%,$P<0.001$以及61%[+或-]13%,$P< 0.01$)。而且,申请人证明了在RVG-外泌体处理的动物中在总[β]-淀粉样蛋 白1-42水平上的显著降低(55%,$P<0.05$),其中β淀粉样蛋白为一种在阿尔茨海默病理学中的淀粉样斑块的主要成分。所观察到的降低大于在心室内 注射BACE1抑制剂之后的正常小鼠中证明的β淀粉样蛋白1-40降低。阿尔 瓦雷斯-尔维蒂(Alvarez-

Erviti)在BACE1切割产物上进行了5'-cDNA末端 快速扩增(RACE),其提供了经由siRNA的RNAi-介导的敲低的证据。

[0401] 最后,阿尔瓦雷斯-尔维蒂(Alvarez-Erviti)等人通过评定IL-6、IP-10、TNFα和IFN-α血清浓度研究了siRNA-RVG外泌体是否诱导了体内免疫反 应。在siRNA-RVG外泌体处理之后,类似于与有力地刺激IL-6分泌的 siRNA-RVG-9R相反的siRNA转染试剂处理,登记了在所有细胞因子上的非 显著性变化,证实了该外泌体处理的免疫惰性属性。假定外泌体仅仅封装 20%的siRNA,用RVG-外泌体递送比RVG-9R递送显得更有效,因为用少五 倍的siRNA实现了相当的mRNA敲低和更好的蛋白质敲低,而没有相应水平 的免疫刺激。这个实验证明了RVG-外泌体技术的治疗潜力,其潜在地适合于 与神经退行性疾病相关的基因的长期沉默。阿尔瓦雷斯-尔维蒂(Alvarez- Erviti)等人的外泌体递送系统可以用于将本发明的CRISPR-Cas系统递送至 治疗靶标,尤其是神经退行性疾病。对于本发明可以考虑封装在约100到 1000mg的RVG外泌体中的约100到1000mg的CRISPR Cas的剂量。

[0402] 艾尔·安达卢西(El-Andaloussi)等人(《自然实验手册》(Nature Protocols)7,2112-2126(2012))披露了可以如何利用来源于培养的细胞的 外泌体用于体外和体内递送siRNA。这个方案首先描述了通过转染一种包含 与肽配体融合的外泌体蛋白的表达载体的靶向外泌体的产生。其次,艾尔·安 达卢西(El-Andaloussi)等人解释了如何纯化和表征来自转染的细胞上清液的 外泌体。接着,艾尔·安达卢西(El-Andaloussi)等人详述了将siRNA加载到 外泌体中的关键步骤。最后,艾尔·安达卢西(El-Andaloussi)等人概述了如何使用外泌体有效体外递送siRNA以及体内递送到小鼠脑中。还提供了预期 结果的实例,其中外泌体介导的siRNA递送通过功能分析和成像而评估。整 个方案进行约3周。根据本发明的递送或给药可以使用从自我衍生的树突细 胞产生的外泌体来进行。

[0403] 在另一个实施例中,考虑了瓦尔格伦(Wahlgren)等人的血浆外泌体 (《核酸研究》(Nucleic Acids Research),2012年,第40卷,第17期, e130)。外泌体是由包括树突细胞(DC)、B细胞、T细胞、肥大细胞、上 皮细胞和肿瘤细胞在内的许多细胞类型产生的纳米囊泡(30-90nm大小)。 这些囊泡通过晚期核内体的向内出芽而形成,并且然后在与质膜融合后释放 到细胞外环境中。由于外泌体天然地在细胞之间运送RNA,这种特性在基因 治疗中可能有用。

[0404] 来自血浆的外泌体通过如下制备:在900g离心血沉棕黄层持续20分钟 以便分离血浆,之后收获细胞上清液,在300g离心10分钟以便去除细胞, 并且在16 500g离心30分钟,之后通过0.22mm过滤器进行过滤。通过在 120 000g离心70分钟使外泌体沉淀。根据在RNAi人/小鼠启动试剂盒 (Quiagen,希尔顿(Hilden),德国)中的制造商的说明进行siRNA到外泌 体中的化学转染。siRNA以终浓度2mmol/ml添加到100ml PBS中。在加入 HiPerFect转染试剂之后,将该混合物在室温下孵育10分钟。为了去除过量的 胶束,使用醛/硫酸盐乳胶珠再分离外泌体。可以类似于siRNA进行CRISPR Cas到外泌体中的化学转染。外泌体可以与从健康供体的外周血中分离的单核 细胞和淋巴细胞共培养。因此,可以考虑的是,可以将含有CRISPR Cas的外 泌体引入人单核细胞和淋巴细胞中并且以自体方式再引入。因此,可以使用 血浆外泌体进行根据本发明的递送或给药。

[0405] 脂质体

[0406] 可以用脂质体进行根据本发明的递送或给药。脂质体是球形囊泡结构, 其组成为

围绕内部水性区室的单层或多层脂质双层以及相对不可渗透的外部 亲脂性磷脂双层。脂质体作为药物递送载体受到了相当的重视,因为它们是 生物相容、无毒的,可以递送亲水和亲脂性药物分子,包护它们的负荷物免 于被血浆酶降解,并且运送它们的负荷跨过生物膜和血脑屏障(BBB)(对 于评述,参见,例如,斯普奇(Spuch)和纳瓦罗(Navarro)《药物递送杂 志》(Journal of Drug Delivery),2011卷,文献标识码469679,第12页, 2011.doi:10.1155/2011/469679)。

[0407] 可以从几种不同类型的脂质制造脂质体;然而,磷脂最常用来产生作为 药物载体的脂质体。虽然当脂质膜与一种水性溶液混合时脂质体形成是自发 的,但也可通过使用一个均质机、超声发生器、或挤出设备以振摇的形式施 加力使其加速(对于评述,参见,例如,斯普奇(Spuch)和纳瓦罗 (Navarro)《药物递送杂志》(Journal of Drug Delivery),2011卷,文献标 识码469679,第12页,2011.doi:10.1155/2011/469679)。

[0408] 可以将几种其他的添加剂添加到脂质体,以便修饰其结构和特性。例 如,可以将胆固醇或鞘磷脂添加到脂质体混合物中,以便帮助稳定脂质体结 构并且防止脂质体内部负荷物的泄漏。此外,从氢化卵磷脂酰胆碱或卵磷脂 酰胆碱、胆固醇、和磷酸二鲸蜡脂制备脂质体,并且脂质体的平均囊泡大小 被调整到约50至100nm。(对于评述,参见,例如,斯普奇(Spuch)和纳 瓦罗(Navarro)《药物递送杂志》(Journal of Drug Delivery),2011卷,文献标识码469679,第12页,2011.doi:10.1155/2011/469679)。

[0409] 常规的脂质体配制品主要由天然磷脂和脂质如1,2-二硬脂酰-sn-甘油基-3- 磷脂酰胆碱(DSPC)、鞘磷脂、卵磷脂酰胆碱和单唾液酸神经节苷酯构成。由于这种配制品仅仅由磷脂组成,脂质体配制品已经遇到了许多挑战,其中 之一是在血浆中的不稳定性。已经作出战胜这些挑战的若干尝试,特别是在 脂质膜的处理方面。这些尝试之一集中于胆固醇的处理。将胆固醇添加到常 规配制品中减缓了封装的生物活性化合物到血浆中的迅速释放,或者添加 1,2-二油酰基-sn-甘油基-3-磷酸乙醇胺(DOPE)增加稳定性(对于评述,参见,例如,斯普奇(Spuch)和纳瓦罗(Navarro)《药物递送杂志》(Journal of Drug Delivery),2011卷,文献标识码469679,第12页,2011. doi:10.1155/2011/469679)。

[0410] 在一个特别有利的实施例中,特洛伊木马(Trojan Horse)脂质体(也称 为分子木马)是令人希望的并且方案可见于 http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long。这些粒子允许转 基因在血管内注射之后递送到整个脑。不受到限制,表面结合有特异性抗体 的中性脂质粒子允许经由胞吞作用跨过血脑屏障。申请人假定利用特洛伊木 马脂质体将核酸酶的CRISPR家族经由血管内注射递送到脑,这将允许全脑 转基因动物,而不需要胚胎操作。对于在脂质体中的体内给药,可以考虑约 1-5g的DNA。

[0411] 在另一个实施例中,该CRISPR Cas系统可以在脂质体中给药,如一种稳 定的核酸-脂质粒子(SNALP)(参见,例如,莫里西(Morrissey)等人,《自然生物技术》(Nature Biotechnology),第23卷,第8期,2005年8 月)。考虑了约1、3或5mg/kg/天的靶向SNALP中的特异性CRISPR Cas的 每日静脉注射。日治疗可以经过约三天,然后每周治疗持续约五周。在另一 个实施例中,还考虑了通过以约1或2.5mg/kg的剂量静脉注射给药的封装有 特异性CRISPR Cas的SNALP(参见,例如,齐默尔曼(Zimmerman)等 人,《自然通讯》(Nature Letters),第441卷,2006年5月4日)。该 SNALP配制品可含有以2:40:10:48的摩尔百分比的脂质3-N-[(w甲氧基 聚(乙二醇)2000)氨甲酰]-1,2-二肉豆蔻氧基-丙胺(PEG-C-DMA)、1,

2-二亚油 基氧基-N,N-二甲基-3-氨基丙烷(DLinDMA)、1,2-二硬脂酰-sn-甘油基-3-磷 酸胆碱(DSPC)和胆固醇,(参见,例如,齐默尔曼(Zimmerman)等人,《自然通讯》(Nature Letters),第441卷,2006年5月4日)。

[0412] 在另一个实施例中,已经证明稳定的核酸-脂质粒子(SNALP)将分子有 效递送到高度血管化的HepG2-衍生的肝脏肿瘤,但是不递送到血管化不良的 HCT-116衍生的肝脏肿瘤(参见,例如,李(Li),《基因治疗》(Gene Therapy)(2012)19,775-780)。可以通过如下制备这些SNALP脂质体:使用25:1的脂质/siRNA比率和48/40/10/2的胆固醇/D-Lin-DMA/DSPC/PEG- C-DMA的摩尔比,用二硬脂酰磷脂酰胆碱(DSPC)、胆固醇和siRNA配制 D-Lin-DMA和PEG-C-DMA。生成的SNALP脂质体在大小上为约80-100 nm。

[0413] 在又另一个实施例中,SNALP可以包含合成胆固醇(西格玛-奥德里奇 公司(Sigma-Aldrich),圣路易斯,密苏里州,美国)、二棕榈酰磷脂酰胆 碱(Avanti Polar Lipids公司,阿拉巴斯特(Alabaster),阿拉巴马州,美 国)、3-N-[(w-甲氧基聚(乙二醇)2000)氨甲酰]-1,2-二肉豆蔻氧基丙基胺、和 阳离子的1,2-二亚油基氧基-3-N,N二甲基氨基丙烷(参见,例如,盖斯伯特 (Geisbert)等人,《柳叶刀》(Lancet)2010;375:1896-905)。例如可以 考虑静脉推注给予约2mg/kg总CRISPR Cas/剂的剂量。

[0414] 在又另一个实施例中,SNALP可以包含合成胆固醇(西格玛-奥德里奇 公司(Sigma-Aldrich))、1,2-二硬脂酰基-sn-甘油基-3-磷酸胆碱(DSPC; Avanti Polar Lipids公司)、PEG-cDMA、和1,2-二亚油基氧基-3-(N;N-二甲基) 氨基丙烷(DLinDMA)(参见,例如,贾奇(Judge),《临床研究杂志》 (J.Clin.Invest.)119:661-673(2009))。用于体内研究的配制品可包含约 9:1的最终脂质/RNA质量比。

[0415] 已经由阿尔尼拉姆制药公司(Alnylam Pharmaceuticals)的巴洛斯 (Barros)和格罗布(Gollob)评论了RNAi纳米药物的安全性(参见,例 如,《先进药物递送评论》(Advanced Drug Delivery Reviews)64(2012) 1730-1737)。稳定的核酸脂质粒子(SNALP)由四种不同的脂质构成—一种 在低pH时为阳离子的可电离脂质(DLinDMA)、一种中性辅助脂质、胆固 醇、和一种可扩散的聚乙二醇(PEG)-脂质。该粒子直径大约为80nm并且 在生理pH时是电中性的。在配制过程中,该可电离脂质用于在粒子形成过程 中使脂质与阴离子siRNA缩合。当在渐增的酸性内体条件下带正电荷时,该 可电离的脂质还介导了SNALP与内体膜的融合,从而使得能够将siRNA释 放到细胞质中。该PEG-脂质在配制过程中使该粒子稳定并且减少聚集,随后 提供中性的亲水性外部,以改进药代动力学特性。

[0416] 到目前为止,已经使用SNALPsiRNA配制品开始两个临床项目。Tekmira制药公司最近完成了SNALP-ApoB在具有增高的LDL胆固醇的成年 志愿者中的I期单剂量研究。ApoB主要在肝脏和空肠中表达,并且对于 VLDL和LDL的组装和分泌是必需的。十七位受试者接受了SNALP-ApoB的 单剂量(跨7个剂量水平的剂量递增)。没有肝脏毒性(预期为基于临床前研究的潜在剂量限制性毒性)的证据。处于最高剂量的(两位中的)一位受 试者经历了与免疫系统刺激一致的流感样症状,于是做出结束该试验的决 定。

[0417] 阿尔尼拉姆制药公司(Alnylam Pharmaceuticals)已经类似地推出了 ALN-TTR01,其采用上述SNALP技术并且靶向突变体和野生型TTR的肝细 胞产生,从而治疗TTR淀粉样变性(ATTR)。已经描述了三个ATTR综合 征:家族性淀粉样多发性神经病变(FAP)和家族性淀粉样心肌病(FAC)— 两者均由TTR中的常染色体显性突变引起;以及由野生型TTR引

起的老年全 身性淀粉样变性(SSA)。最近在具有ATTR的患者中完成了ALN-TTR01的 安慰剂对照单剂量递增I期试验。向31位患者(23位用研究药物,8位用安 慰剂)在0.01到1.0mg/kg(基于siRNA)的剂量范围内以15分钟静脉内输 注给予ALN-TTR01。治疗耐受良好,其中在肝功能试验中没有显著增加。在 ≥0.4mg/kg时在23位患者的3位中注意到输注相关反应;对所有患者均做出 减慢输注速率的反应并且所有患者继续参与研究。在处于1mg/kg的最高剂量 (如根据临床前和NHP研究预期的)的两位患者中注意到血清细胞因子IL- 6、IP-10和IL-1ra的最小与瞬时升高。在1mg/kg时观察到ALN-TTR01的预 期药理效应,即,血清TTR的降低。

[0418] 在又另一个实施例中,可以通过将阳离子脂质、DSPC、胆固醇以及 PEG-脂质以40:10:40:10的摩尔比分别溶解在乙醇中来制作SNALP(参 见,森普尔(Semple)等人,《自然生物技术》(Nature Niotechnology), 第28卷,第2期,2010年2月,第172-177页)。将该脂质混合物添加到水 性缓冲液中(50mM柠檬酸盐,pH 4),混合至最终的乙醇和脂质浓度分别 为30%(vol/vol)和6.1mg/ml,允许在22℃平衡2分钟,然后挤出。使用 Lipex挤出仪(北方脂质公司(Northern Lipids)),在22℃将水合脂质通过 两层80nm孔径大小的过滤器(Nuclepore)直到获得70-90nm直径的囊泡时 为止,正如通过动态光散射分析测定的。这大致需要1-3次通过。将该 siRNA(溶解在50mM柠檬酸盐中,pH为4的含有30%乙醇的水溶液)以约 5ml/min的速率在混合下添加到预平衡的(35℃)囊泡中。在达到0.06 (wt/wt)的最终靶siRNA/脂质比率之后,将该混合物在35℃另外孵育30分 钟,以允许囊泡重组和该siRNA的封装。然后去除乙醇并且通过透析或切向 流渗滤用PBS(155mM NaCl、3mM Na2HPO4、1mM KH2PO4,pH 7.5)替 换外部缓冲液。使用控制的逐步稀释法工艺将siRNA封装在SNALP中。KC2-SNALP的脂质组分为以57.1:7.1:34.3:1.4的摩尔比使用的DLin-KC2- DMA(阳离子脂质)、二棕榈酰磷脂酰胆碱(DPPC;Avanti Polar Lipids公 司)、合成胆固醇(西格玛公司)和PEG-C-DMA。在形成加载的粒子后, 将SNALP在PBS中透析并且在使用之前通过0.2μm的滤膜消毒过滤。平均 粒径为75-85nm,并且将90%-95%的siRNA封装在脂质粒子内。用于体内测 试的在配制品中的最终siRNA/脂质比率是大约0.15(wt/wt)。在使用之前即 刻将含有因子VII siRNA的LNP-siRNA系统在无菌PBS中稀释到适当浓度, 并且通过侧尾静脉以10ml/kg的总体积静脉内给药。这种方法可以类推到本 发明的CRISPR Cas系统。

[0419] 其他脂质

[0420] 其他阳离子脂质,如氨基脂质2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧 戊环(DLin-KC2-DMA)可以类似于SiRNA地用来封装CRISPR Cas(参 见,例如,加雅拉曼(Jayaraman),《德国应用化学》(Angew.Chem.Int. Ed.)2012,51,8529-8533)。可以考虑具有下列脂质组成的预成型囊泡:分 别处于摩尔比40/10/40/10的氨基脂质、二硬脂酰卵磷脂(DSPC)、胆固醇和 (R)-2,3-双(十八烷氧基)丙基-1-(甲氧基聚(乙二醇)2000)丙基碳酸酯(PEG-脂 质),以及大约0.05(w/w)的FVII siRNA/总脂质比率。为了确保在70-90 nm范围内的窄粒径分布以及0.11_0.04(n=56)的低多分散性指数,可以在 添加CRISPR Cas RNA之前将粒子通过80nm的膜挤出达三次。可以使用含 有高度有效的氨基脂质16的粒子,其中四种脂质组分16、DSPC、胆固醇和 PEG-脂质的摩尔比(50/10/38.5/1.5)可以被进一步优化,以增强体内活性。

[0421] 迈克尔S D科尔曼(Michael S D Kormann)等人(“在小鼠中递送化学 修饰的mRNA

之后治疗蛋白的表达)("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice):《自然生物技术》(Nature Biotechnology),第29卷,第154-157页,(2011)2011年1月9日在线公 开)描述了脂质包膜用于递送RNA的用途。脂质包膜的用途在本发明中也是 优选的。

[0422] 在另一个实施例中,脂质可以与本发明的CRISPR Cas系统配制在一起而 形成脂质纳米粒子(LNP)。脂质包括但不限于,DLin-KC2-DMA4、C12- 200和辅助脂质二硬脂酰磷脂酰胆碱、胆固醇和PEG-DMG,可以使用自发囊泡形成程序将其与CRISPR Cas而不是siRNA一起配制(参见,例如, Novobrantseva,《分子治疗-核酸》(Molecular Therapy-Nucleic Acids) (2012)1,e4;doi:10.1038/mtna.2011.3)。组分摩尔比可以是约 50/10/38.5/1.5(DLin-KC2-DMA或C12-200/二硬脂酰磷脂酰胆碱/胆固醇 /PEG-DMG)。在DLin-KC2-DMA和C12-200脂质纳米粒子(LNP)的情况 下,最终脂质:siRNA重量比可以分别是约12:1和9:1。配制品可以具有约80nm的平均粒径,具有>90%的包覆效率。可以考虑3mg/kg的剂量。

[0423] Tekmira公司在美国和国外具有一组针对LNP和LNP配制品的不同方面 的大约95个同族专利(参见,例如,美国专利号7,982,027;7,799,565; 8,058,069;8,283,333;7,901,708;7,745,651;7,803,397;8,101,741; 8,188,263;7,915,399;8,236,943和7,838,658以及欧洲专利号1766035; 1519714;1781593和1664316),所有这些专利均可用于和/或适用于本发 明。

[0424] 该CRISPR Cas系统可以封装在PLGA微球中进行递送,例如进一步描 述于美国公开申请20130252281和20130245107以及20130244279(转让给 Moderna Therapeutics公司)中,其涉及包含修饰的核酸分子的组合物的配制 品方面,所述核酸分子可以编码一种蛋白质、一种蛋白质前体、或该蛋白质 或蛋白质前体的部分或完全加工形式。该配制品具有50:10:38.5:1.5-3.0 (阳离子脂质:融合脂质:胆固醇:PEG脂质)的摩尔比。该PEG脂质可以选 自,但不限于PEG-c-DOMG、PEG-DMG。该融合脂质可以是DSPC。还参 见,施鲁姆(Schrum)等人,"工程化核酸的递送和配制"(Delivery and Formulation of Engineered Nucleic Acids),美国公开申请20120251618。

[0425] Nanomerics公司的技术着手解决针对广泛治疗学的生物利用度挑战,包 括基于低分子量疏水药物、肽以及核酸(质粒、siRNA、miRNA)的治疗 学。该技术已经证明了明显优势的特异性的给药途径包括口服途径、跨血脑 屏障的运送、向实体瘤以及眼部的递送。参见,例如,马萨(Mazza)等人, 2013,《ACS纳米》(ACS Nano.)2013年2月26日;7(2):1016-26;乌切 克布(Uchegbu)和秀(Siew),2013,《制药科学杂志》(J Pharm Sci.) 102(2):305-10和拉拉特萨(Lalatsa)等人,2012,《控释杂志》(J Control Release),2012年7月20日;161(2):523-36。

[0426] 美国专利公开号20050019923描述了用于向哺乳动物身体递送生物活性 分子例如多核苷酸分子、肽和多肽和/或药剂的阳离子树状聚合物。这些树状 聚合物适合于将生物活性分子的递送靶向到例如肝、脾、肺、肾或心脏。树 状聚合物是从简单的支化单体单元以逐步方式制备的3维大分子,其性质和 功能性可以容易地进行控制和改变。经由向多功能核(发散式合成法)或朝 向多功能核(收敛式合成法)重复加成结构单元(building blocks)而合成树 状聚合物,并且结构单元的3维壳的各次加成导致更高级别的树状聚合物的 形成。聚丙烯亚胺树状聚合物从二氨基丁烷核开始,通过对伯胺的丙烯腈的 双迈克

尔加成反应向其上添加两倍数目的氨基基团，继之为腈的氢化。这导 致氨基基团的加倍。聚丙烯亚胺树状聚合物含有100％的可质子化氮以及高达 64个末端氨基基团(5级，DAB 64)。可质子化基团常常为能够在中性pH 时接受质子的胺基。树状聚合物作为基因递送剂的用途在很大程度上集中于 聚酰胺-胺和含磷化合物的用途，其中胺/酰胺的混合物或N--P(O_2)S分别作为 缀合单元，没有报道关于更低级别的聚丙烯亚胺树状聚合物用于基因递送的 用途的工作。还研究了作为pH敏感的控制释放系统的聚丙烯亚胺树状聚合 物，其用于药物递送以及当被外周氨基酸基团化学修饰时用于它们的客体分 子的封装。还研究了聚丙烯亚胺树状聚合物的细胞毒性和与DNA的相互作用 以及DAB 64的转染效力。

[0427] 美国专利公开号20050019923是基于与早期报道相反的观察：阳离子树 状聚合物例如聚丙烯亚胺树状聚合物展示出适当的特性，如，特异性靶向和 低毒性，其用于在生物活性分子如基因材料的靶向递送中使用。另外，阳离 子树状聚合物的衍生物也展示出适当的用于生物活性分子的靶向递送的特 性。还参见，《生物活性聚合物》(Bioactive Polymers)、美国公开申请 20080267903，其披露了“不同的聚合物，包括阳离子聚胺聚合物和树枝状聚 合物，其显示具有抗增殖活性，并且因此可用于治疗其特征为不希望的细胞增殖的失调，如新生物和肿瘤、炎性失调(包括自身免疫性失调)、银屑病 和动脉粥样硬化。这些聚合物可以作为活性剂单独使用、或者作为其他治疗 剂(如药物分子或用于基因治疗的核酸)的递送载体。在这样的情况下，这 些聚合物自身固有的抗肿瘤活性可以补足有待递送的药剂的活性。

[0428] 超电荷蛋白

[0429] 超电荷蛋白是一类具有非常高的正或负的理论净电荷的工程化的或天然 存在的蛋白质。超负电荷蛋白和超正电荷蛋白两者都展示出显著的抵抗热诱 导或化学诱导的聚集的能力。超正电荷蛋白还能够渗透哺乳动物细胞。使负 荷物与这些蛋白质结合，如质粒DNA、siRNA、或其他蛋白质，可使得这些 大分子到体外和体内的哺乳动物细胞中的功能递送成为可能。刘大卫实验室 (David Liu's lab)在2007年报道了超电荷蛋白的建立和表征(劳伦斯 (Lawrence)等人，2007，《美国化学会志》(Journal of the American Chemical Society)129，10110-10112)。

[0430] siRNA和质粒DNA到哺乳动物细胞中的非病毒递送对于研究和治疗应用 都是有价值的(阿肯克(Akinc)等人，2010，《自然生物技术》(Nat. Biotech.)26，561-569)。纯化的+36 GFP蛋白(或其他超正电荷蛋白)与 siRNA在适当的无血清培养基中混合并且允许在添加到细胞中之前复合。在 这个阶段的血清的包含将会抑制超电荷蛋白-siRNA复合物的形成和降低治疗 效果。已经发现以下方案对于多种细胞系是有效的(麦克诺顿(McNaughton)等人，2009，《美国国家科学院院刊》 (Proc.Natl.Acad.Sci.USA)106，6111-6116)。然而，应当进行改变蛋白质 和siRNA剂量的先导实验来优化用于特异性细胞系的程序。

[0431] (1)在治疗前一天，以1x 10^5个细胞/孔铺板于48孔板中。

[0432] (2)在治疗当天，在无血清的培养基中稀释纯化的+36 GFP蛋白直到终 浓度200nM。添加siRNA到50nM的终浓度。涡旋混合并且在室温孵育10 分钟。

[0433] (3)在孵育过程中，抽吸培养基离开细胞并且再次用PBS洗涤。

[0434] (4)在孵育+36 GFP和siRNA之后，向细胞加入蛋白质-siRNA复合 物。

[0435] (5)将细胞与复合物在37C孵育4小时。

[0436] (6)在孵育之后,抽出培养基并且用20U/mL的肝素PBS洗涤三次。用含血清培养基另外孵育细胞48小时或更长,这取决于用于敲低的试验。

[0437] (7)通过免疫印迹、qPCR、表型分析、或其他适当的方法分析细胞。

[0438] 刘大卫实验室(David Liu's lab)已经进一步发现+36 GFP在一些列细胞 中是一种有效的质粒递送试剂。由于质粒DNA是一种比siRNA大的负荷 物,有效复合质粒需要成比例地更大的+36 GFP蛋白。为了有效质粒递送, 申请人已经开发了一种带有C末端HA2肽标签的+36 GFP变体,这种肽是一 种已知的从流感病毒血凝素蛋白衍生的内体破坏肽。下列方案在多种细胞中 是有效的,但是如上所述,建议针对特异性细胞系和递送应用优化质粒DNA 的超电荷蛋白的剂量。

[0439] (1)在治疗前一天,以1x 10^5/孔铺板于48孔板中。

[0440] (2)在治疗当天,在无血清的培养基中稀释纯化的þ36 GFP蛋白直到终 浓度2mM。加入1mg的质粒DNA。涡旋混合并且在室温孵育10分钟。

[0441] (3)在孵育过程中,抽吸培养基离开细胞并且再次用PBS洗涤。

[0442] (4)在孵育þ36 GFP和质粒DNA之后,向细胞轻轻加入蛋白质-DNA 复合物。

[0443] (5)将细胞与复合物在37孵育4小时。

[0444] (6)在孵育之后,抽出培养基并且用PBS洗涤。在含血清培养基中孵 育细胞,并且另外孵育24-48小时。

[0445] (7)在适当时分析质粒递送(例如,通过质粒驱动的基因表达)。

[0446] 还参见,例如,麦克诺顿(McNaughton)等人,《美国国家科学院院 刊》(Proc.Natl.Acad.Sci.USA)106,6111-6116(2009);克罗尼肯 (Cronican)等人,《ACS化学生物学》(ACS Chemical Biology)5,747- 752(2010);克罗尼肯(Cronican)等人,《化学与生物学》(Chemistry& Biology)18,833-838(2011);汤普森(Thompson)等人,《酶学方法》(Methods in Enzymology)503,293-319(2012);汤普森(Thompson) D.B.,等人,《化学与生物学》(Chemistry&Biology)19(7),831-843 (2012)。超电荷蛋白的这些方法可以使用和/或适用于本发明的CRISPR Cas 系统的递送。

[0447] 可植入装置

[0448] 在另一个实施例中,还考虑了用于CRISPR Cas系统的递送的可植入装 置。例如,美国专利公开20110195123披露了一种可植入医用装置,其局部 地并且在延长的时期内洗脱药物,包括几种类型的这种装置、实施的治疗方 式和植入方法。该装置包含聚合物基材,例如,用作装置主体的基质、以及 药物,并且在一些情况下包含另外的支架材料,如金属或另外的聚合物,以 及增强可见度和成像的材料。药物的选择是基于局部地并且在延长的时期内 释放药物的优点,其中药物直接释放到患病区域的细胞外基质(ECM),所 述患病区域如肿瘤、炎症、退化,或用于针对症状的目的,或者释放到受损 的平滑肌细胞,或者用于预防。一种药物是基RNA干扰(RNAi)的基因沉 默药物,包括但不限于si RNA、sh RNA、或反义RNA/DNA、核酶和核苷类 似物。因此,这个系统可以用于和/或适用于本发明的CRISPR Cas系统。在 一些实施例中,植入方式为用于包括近距离放射疗法和针吸活组织检查在内 的其他治疗的、当今开发和使用的现有植入程序。在这样的情况下,在本发 明中描述的新植入物的尺寸类似于初始植入物。典型地在相同的治疗程序中 植物很少的装置。

[0449] 正如在美国专利公开20110195123中所述，提供了一种药物递送可植入 或可插入系统，包括适用于空腔例如腹腔和/或其中药物递送系统未被锚定或 附接的任何其他类型的给药，其包含生物稳定的和/或可降解的和/或生物可吸 收的聚合物基材，其可以例如任选地是一种基质。应当指出的是术语“插 入”也包括植入。该药物递送系统优选地被实施为如美国专利公开 20110195123中描述的“装填器(Loder)”。

[0450] 聚合物或多种聚合物是生物相容的，其结合一种药剂和/或多种药剂，使 得药剂以控制的速率释放，其中该聚合物基材如基质的总体积，例如在一些 实施例中是任选地并且优选地不大于容许达到该药剂的治疗水平的最大体 积。作为一个非限制性实例，这样的体积优选在0.1m^3到1000mm^3的范围 内，正如该药剂负荷的体积所要求的。该装填器任选地是更大的，例如当结 合有一个其大小由功能性决定的装置时，例如而不限于，膝关节、宫内节育 环或子宫颈环等等。

[0451] 在一些实施例中，该药物递送系统(用于递送该组合物)被设计为优选 地采用可降解聚合物，其中主要释放机制是本体溶蚀(Bulk erosion)；或者 在一些实施例中，使用了不可降解的、或缓慢降解的聚合物，其中主要释放 机制是扩散而不是本体溶蚀，使得外部部分用作一种膜，而其内部部分用作 一种储药池，该储药池在一个延长的时期内(例如从约一周到约几个月)实 际上不受环境的影响。还可以任选地使用具有不同释放机制的不同聚合物的 组合。在总药物释放期的重要时段期间，在表面处的浓度梯度优选地被维持有效地恒定，并且因此扩散速率是有效地恒定的(称为“零模式”扩散)。关于术语“恒定”，它意指优选地维持在治疗效果的低阈值以上的扩散速 率，但是可以仍然任选地具有初期突释的特征和/或波动，例如增加和降低到 一定程度。该扩散速率优选地被如此维持一个延长的时期，并且它被考虑为 相对于一定的水平是恒定的，以便优化治疗有效期，例如该有效沉默期。

[0452] 该药物递送系统任选地并且优选地被设计为保护基于核苷酸的治疗剂免 于降解，而不论化学性质或由于受试者体内的酶和其他因素的攻击。

[0453] 如描述于美国专利公开20110195123中的该药物递送系统任选地与传感 和/或激活器具相关联，这些器具通过激活和/或加速/减速的无创和/或微创方 法在该装置的植入之时和/或之后被操作，例如任选地包括但不限于热力加热 和冷却、激光束和超声波，包括聚焦超声和/或RF(射频)方法或装置。

[0454] 根据美国专利公开20110195123的以下实施例，用于局部递送的部位可 以任选地包括其特征为高度异常的细胞增殖、受到抑制的细胞凋亡的靶部 位，包括肿瘤、活动性和或慢性炎症和感染，包括自身免疫性疾病状态、退 化组织(包括肌肉和神经组织)、慢性疼痛、退行性部位，以及用于增强组 织再生的骨折位置以及其他伤口位置，以及损伤的心肌、平滑肌和横纹肌。用于局部递送的部位还任选地可包括能够进行预防活动(包括怀孕)、预防 感染和衰老的部位。

[0455] 用于植入该组合物的部位、或靶部位，优选地其特征为用于靶向局部递 送的足够小的半径、面积和/或体积。例如，该靶部位任选地具有在从约0.1 mm到约5cm范围内的直径。

[0456] 该靶部位的位置优选地针对最大治疗效力而选择。例如，该药物递送系 统的组合物(任选地与如上所述的用于植入的装置一起)任选地并且优选地 被植入在肿瘤环境或与

之相关的血供之内或者附近。

[0457] 例如该组合物(任选地与该装置一起)任选地植入在胰脏、前列腺、乳 房、肝脏之内或附近,通过乳头,在血管系统之内,等等。

[0458] 靶位置任选地选自下组,该组由以下各项组成(仅仅作为非限制性实 例,因为任选地在身体内的任何部位可以适合于植入一个装填器):1.在退行 性部位的脑,像在帕金森病或阿尔茨海默病中的基底神经节、白质和灰质;2. 如在肌萎缩性脊髓侧索硬化症(ALS)的情况下的脊柱;3.预防HPV感染的 子宫颈;4.活动性或慢性炎性关节;5.在银屑病情况下的真皮;6.用于止痛作 用的交感神经部位和感觉神经部位;7.骨内植入;8.急性和慢性感染部位;9. 阴道内;10.耳内--听觉系统、内耳的迷路、前庭系统;11.气管内;12.心内; 冠状动脉、心外膜;13.膀胱;14.胆道系统;15.实质组织,包括且不限于肾 脏、肝脏、脾脏;16.淋巴结;17.唾液腺;18.牙龈;19.关节内(进入关 节);20.眼内;21.脑组织;22.脑室;23.空腔,包括腹腔(例如但不限于, 卵巢癌);24.食管内以及25.直肠内。

[0459] 任选地,该系统(例如含有该组合物的装置)的插入与向在该靶部位和 该部位附近的ECM注射材料有关,从而影响该靶部位和这个部位附近的 ECM中的局部pH和/或温度和/或影响该药物的扩散和/或药物动力学的其他 生物因素。

[0460] 任选地,根据一些实施例,所述药剂的释放可以与传感和/或激活器具相 关联,这些器具通过激活和/或加速/减速的无创和/或微创方法和/或别的方法 在插入之前和/或之时和/或之后被操作,所述方法包括激光束、放射、热力加 热和冷却、和超声波,包括聚焦超声和/或RF(射频)方法或装置、以及化学 激活剂。

[0461] 根据美国专利公开20110195123的其他实施例,该药物优选地包括基因 沉默性生物RNAi药物,例如,对于局限性癌症情况,在乳房、胰脏、脑、 肾脏、膀胱、肺、以及前列腺中,如下文所述。而且,除了siRNA之外的许 多药物可适用于封装在装填器中,并且可以与本发明相关联,只要这样的药 物可以被封装在装填器基材(例如像一种基质)中即可。这样的药物包括当 今通过除了本发明之外的方法递送的经批准的药物,包括用于真菌感染的两 性霉素B;例如用于骨髓炎中的抗生素;止痛剂,如麻醉药;例如在阿尔茨 海默病或帕金森病中的抗退行性剂,在背痛的情况其在植入脊柱附近的一个 装填器中。这样的系统可以用于和/或适用于递送本发明的CRISPR Cas系 统。

[0462] 例如,对于特殊应用,如预防平滑肌细胞(在支架置放程序中受损并且 因此而倾向增殖)的生长和再生长,该药物可以任选地是使平滑肌细胞沉默 (包括H19沉默)的siRNA,或者为选自下组的药物,该组由以下各项组 成:紫杉醇、雷帕霉素和雷帕霉素类似物。在这样的情况下,该装填器优选 地是一种药物洗脱支架(DES),其以恒定速率延长释放,或者是一个与该 支架关联的单独植入的专用装置。这都可以用于和/或适用于本发明的 CRISPR Cas系统。

[0463] 作为特殊应用的另一个实例,神经肌肉退行性疾病由于异常基因表达而 发生。沉默RNA的局部递送可以具有干扰这样的异常基因表达的治疗特性。包括小分子药物和大分子在内的抗凋亡、抗炎和抗退行性药物的局部递送也 可以任选地是治疗性的。在这样的情况下,该装填器用于以恒定速率和/或通 过单独植入的专用装置延长释放。这都可以用于和/或适用于本发明的 CRISPR Cas系统。

[0464] 作为特殊应用的又另一个实例,用基因修饰剂治疗精神和认知障碍。用 沉默RNA

的基因敲低是一个治疗选项。向中枢神经系统部位局部递送基于核 苷酸的药剂的装填器是对于精神和认知障碍的治疗选项,这些精神和认知障 碍包括但不限于,精神病、双极性疾病、神经性障碍和行为疾病(behavioral maladies)。这些装填器也可以在特定脑部位进行植入时局部递送包括小分子 药物和大分子的药物。这都可以用于和/或适用于本发明的CRISPR Cas系 统。

[0465] 作为特殊应用的另一个实例,在局部部位的先天性和/或适应性免疫介质 的沉默使得能够预防器官移植排斥。用植入到移植器官和/或植入部位的装填 器局部递送沉默RNA和免疫调节试剂致使经由排斥性免疫细胞(如针对移植 器官而被激活的CD8)的局部免疫抑制。这都可以用于和/或适用于本发明的 CRISPR Cas系统。

[0466] 作为特殊应用的另一个实例,包括VEGF和血管生成素及其他在内的血 管生长因子对于新血管形成是必需的。这些因子、肽、肽模拟物的局部递送 或抑制它们的抑制物是一种重要的治疗模式;使阻抑物沉默以及用装填器局 部递送刺激血管发生的这些因子、肽、大分子和小分子药物对于周围血管疾 病、全身性血管疾病和心血管疾病是治疗性的。

[0467] 插入的方法,如植入,可以任选地已用于其他类型的组织植入和/或用于 插入和/或用于组织取样,任选地在这样的方法中没有修改,或者可替代地任 选地仅仅具有非重点修改。这样的方法任选地包括但不限于,近距离放射治 疗方法、活组织检查、用和/或不用超声的内窥镜检查如ERCP、进入脑组织 的立体定向法、腹腔镜检查,包括用腹腔镜进入关节、腹腔器官、膀胱壁和 体腔的植入。

[0468] CRISPR酶mRNA和指导RNA

[0469] 也可以单独递送CRISPR酶mRNA和指导RNA。可以在该指导RNA在 给出时间以待CRISPR酶表达之前递送CRISPR酶mRNA。可以在给予指导 RNA之前1-12小时(优选约2-6小时)给予CRISPR酶mRNA。

[0470] 可替代地,可以一起给予CRISPR酶mRNA和指导RNA。有利地,可以 在初始给予CRISPR酶mRNA+指导RNA之后1-12小时(优选约2-6小时) 给予指导RNA的第二加强剂量。

[0471] 为了实现基因组修饰的最有效水平,另外给予CRISPR酶mRNA和/或指 导RNA可以是有用的。

[0472] 为了将毒性和脱靶效应降低到最低限度,重要的是控制所递送的 CRISPR酶mRNA和指导RNA的浓度。CRISPR酶mRNA和指导RNA的最 佳浓度可以通过在细胞或动物模型中测试不同的浓度并且使用深度测序分析 在潜在的脱靶基因组座位处的修饰的范围而确定。例如,对于靶向人类基因 组的EMX1基因中的5'-GAGTCCGAGCAGAAGAAGAA-3'的指导序列,可以使用深度测序来评定在下列两个脱靶位点处的修饰水平,1:5'-GAGTCCTAGCAGGAGAAGAA-3'和2:5'-GAGTCTAAGCAGAAGAAGAA- 3'。对于体内递送,应当选择产生最高的中靶修饰水平同时将脱靶修饰水平 降低到最低限度的浓度。

[0473] 可替代地,为了将毒性水平和脱靶效应降低到最低限度,可以用一对靶 向感兴趣部位的指导RNA来递送CRISPR酶切口酶mRNA(例如,具有 D10A修饰的化脓链球菌Cas9)。这两个指导RNA需要如下间隔开。分别处 于红色(单下划线)和蓝色(双下划线)的指导序列(这些实例是基于化脓 链球菌Cas9需要的PAM)。

[0474]

突出端长度（bp）	指导RNA设计（颜色编码的指导序列和PAM）
14	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
13	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
12	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
11	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
10	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
9	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
8	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
7	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
6	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
5	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-5’
4	5’-NNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN-3’ 3’-NNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNN-

[0475]

3	NNCCNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN
2	NGGNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN
1	NCCNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN
钝的	GGNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN
1	CCNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN
2	GGNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN
3	CCNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN
4	NNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN
5	NNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN
6	NNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN
7	NNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN
8	NNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN
12	NNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN
13	NNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNN
14	NNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNN-3’

[0476]

15	3’- NNNNNNNNNNNNNNNNNNGGNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
16	5’- NNNNNNNNNNNNNNNNNNCCNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
17	3’- NNNNNNNNNNNNNNNNNNGGNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’- NNNNNNNNNNNNNNNNNNCCNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
	3’- NNNNNNNNNNNNNNNNNNGGNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’- NNNNNNNNNNNNNNNNNNCCNNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
	3’- NNNNNNNNNNNNNNNNNNGGNNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’- NNNNNNNNNNNNNNNNNNCCNNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
	3’- NNNNNNNNNNNNNNNNNNGGNNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’- NNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
	3’- NNNNNNNNNNNNNNNNNNGGNNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’- NNNNNNNNNNNNNNNNNNCCNNNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
	3’- NNNNNNNNNNNNNNNNNNGGNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’- NNNNNNNNNNNNNNNNNNCCNNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
	3’- NNNNNNNNNNNNNNNNNNGGNNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’- NNNNNNNNNNNNNNNNNNCCNNGGNNNNNNNNNNNNNNNNNNNNNNNNN-3’
	3’- NNNNNNNNNNNNNNNNNNGGNNCCNNNNNNNNNNNNNNNNNNNNNNNNN-5’
	5’-

[0477]

	NNNNNNNNNNNNNNNNNNCCNGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNGGNCCNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNCCGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNGGCCNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNCCGGNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNNCCNGGNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNNCCNNGGNNNNNNN NNNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNNCCNNNGGNNNNNN NNNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNNCCNNNNGGNNNNN NNNNNNNNNNNNNNNNNNNNN-5’ 5’- NNNNNNNNNNNNNNNNNNNCGGNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNN-3’ 3’- NNNNNNNNNNNNNNNNNNNNCCNNNNNGGNNNNN

[0478]

	NNNNNNNNNNNNNNNNNNNNNNN-5’

[0479] 该系统的进一步询问向申请人提供了5’突出端的证据(参见,例如,拉 恩(Ran)等人,《细胞》(Cell)2013年9月12日;154(6):1380-9和2013 年8月28日提交的美国临时专利申请序列号61/871,301)。申请人进一步鉴 定了涉及当与两个指导RNA结合时通过Cas9切口酶突变体的有效切割的参 数,并且这些参数包括但不限于该5’突出端的长度。在本发明的实施例中, 该5’突出端具有至多200个碱基对、优选至多100个碱基对、或更优选至多 50个碱基对。在本发明的实施例中,该5’突出端具有至少26个碱基对、优选 至少30个碱基对、或更优选34-50个碱基对或1-34个碱基对。在本发明的其 他优选方法中,指导邻近该第一靶序列的DNA双链体的一条链的切割的第一 指导序列以及指导邻近该第二靶序列的另一条链的切割的第二指导序列产生 一个钝切口或一个3’突出端。在本发明的实施例中,该3’突出端具有至多 150、100或25个碱基对,或至少15、10或1个碱基对。在优选的实施例 中,该3’突出端具有1-100个碱基对。

[0480] 本发明的方面涉及被减少的基因产物或被进一步引入到编码基因产物的 DNA分子中的模板多核苷酸或通过允许两个5’突出端重新退火并连接而被精 确切断的间插序列的表达、或被改变的基因产物的活性或功能、或被增加的 基因产物的表达。在本发明的一个实施例中,该基因产物是一种蛋白质。

[0481] 只有产生在这些指导序列(大于-8bp的偏移)之间具有小于8bp重叠的 5’突出端的sgRNA对才能介导可检测的插入缺失信息。重要的是,当与野生 型Cas9配对时,在这些分析中使用的每个指导对于有效诱导插入缺失是重要 的,表明这些指导对的相对位置在预测双切割活性中是最重要的参数。

[0482] 由于Cas9n和Cas9H840A切割DNA的相对链,对于一个给定的sgRNA 对而言,用Cas9H840A取代Cas9n会导致该突出端类型的倒转。例如,将产 生具有Cas9n的5’突出端的一对sgRNA原则上会相反地产生相应的3’突出 端。因此,导致具有Cas9n的3’突出端产生的sgRNA对可以与Cas9H840A 一起使用,以产生一个5’突出端。出人意料地,申请人利用被设计为产生5’ 和3’突出端两者(偏移范围从-278到+58bp)的一组sgRNA对测试了Cas9H840A,但是不能观察插入缺失信息。可能需要进一步的工作来鉴定用 于sgRNA配对的必要的设计规则,以允许经由Cas9H840A进行双切割。

[0483] 肝脏,前蛋白转化酶枯草杆菌蛋白酶Kexin9(PCSK9)

[0484] 前蛋白转化酶枯草杆菌蛋白酶Kexin9(PCSK9)是枯草杆菌蛋白酶丝氨 酸蛋白酶

家族的一个成员。PCSK9主要由肝脏表达并且对于肝细胞LDL受体 表达的下调是关键的。血浆中的LDL-C水平在获得PCSK9功能突变的人体 中高度上升,这些人被归类为患有严重的高胆固醇血症。因此,PCSK9是针 对CRISPR的有吸引力的靶标。可以将PCS9K靶向的CRISPR配制在脂质粒 子中并且例如以约15、45、90、150、250和400μg/kg静脉内给予(参见, 例如, http://www.alnylam.com/capella/wp-content/uploads/2013/08/ALN-PCS02- 001-Protocol-Lancet.pdf)。

[0485] 贝利(Bailey)等人的《分子医学杂志》(J Mol Med(Berl),1999年 1月;77(1):244-9)披露了借助于离体体细胞基因治疗的胰岛素递送,其涉及 从糖尿病患者移除非B细胞体细胞(例如成纤维细胞),并且将其进行体外 遗传改变以产生和分泌胰岛素。这些细胞可在培养物中生长,并且将其作为 胰岛素替代来源输回到供体。可以在植入之前评估以这种方式修饰的细胞, 并且将储备原液冷冻保存。通过使用患者自身的细胞,该程序将避免免疫抑 制的需要,并且克服了组织供应的问题,同时避免了细胞破坏的重现。离体 体细胞基因治疗要求经得起多重转染并且服从受控增殖的、可取得的且稳健 的细胞类型。与使用非B细胞体细胞相关联的特殊问题包括胰岛素原到胰岛 素的加工、以及葡萄糖刺激的胰岛素原合成和调节性胰岛素释放的敏感性的 赋予。使用成纤维细胞、垂体细胞、肾(COS)细胞和卵巢(CHO)细胞的 初步研究表明,可能遇到这些挑战,并且离体体细胞基因治疗提供了针对胰 岛素替代治疗的可行途径。贝利(Bailey)等人的系统可以使用和/或适用于 递送到肝脏的本发明的CRISPR Cas系统。

[0486] 萨托(Sato)等人的方法(《自然生物技术》(Nature Biotechnology), 第26卷,2008年4月第4期,第431-442页)可以适用于递送到肝脏的本发 明的CRISPR Cas系统。萨托(Sato)等人发现,在具有另外的致死性二甲基 亚硝胺诱导的肝硬化小鼠中,用荷siRNA的偶联维生素A的脂质体治疗以剂 量依赖性方式和持续时间依赖性方式几乎完全解决了肝纤维化并且延长了存 活期。含有摩尔比为4:3:3的作为阳离子脂质的O,O'-双十四酰-N-(a-三甲基 铵基乙酰基)二乙醇胺氯化物(DC-6-14)、胆固醇和二油酰基磷脂酰乙醇胺 的阳离子脂质体(Lipotrust)(其显示对于在含血清条件下的体外和体内基因 递送具有高转染效率)购自(北海道系统科学株式会社(Hokkaido System Science))。使用冷冻干燥的空脂质体方法制造这些脂质体,并且通过在使 用之前在涡旋之下将双蒸馏水(DDW)添加到该冻干的脂质混合物中配制为 1mM的浓度(DC-16-4)。为了制备VA偶联的脂质体,通过在1.5ml的管 中在25 1C下将溶解在DMSO中的200nmol的维生素A(视黄醇,西格玛公 司)与脂质体悬浮液(作为DC-16-4100nmol)混合。为了制备携带 siRNAgp46的VA偶联的脂质体(VA-lip-siRNAgp46),在25C搅拌下将 siRNAgp46溶液(580pmol/ml,在DDW中)添加到该视黄醇偶联的脂质体 溶液中。siRNA与DC-16-4的比率为1:11.5(mol/mol)并且siRNA与脂质 体的比率(wt/wt)为1:1。使用微分区系统(VIVASPIN 2浓缩器30,000 MWCO PES,VIVASCIENCE公司),将任何未被脂质体吸收的游离维生素 A或siRNA从脂质体制剂中分离。将该脂质体悬浮液添加到过滤器并且在 1,500g在25 1C离心5分钟,进行3次。收集各个级分,并且将截留在过滤器 中的物质用PBS复原,以实现所希望的用于体外或体内使用的剂量。每隔一 天向大鼠给予0.75mg/kg siRNA,注射三次。通过将如萨托(Sato)等人所述 的在脂质体中的约0.5到1mg/kg的CRISPR Cas RNA递送给人类,萨托 (Sato)等人的系统可用于和/或适用于递送到肝脏的本发明的CRISPR Cas系 统。

[0487] 针对用于递送siRNA到体外和体内肝细胞两者的载体,已被罗泽马 (Rozema)等人命名为siRNA动态多聚缀合物(siRNA Dynamic PolyConjugates)的罗泽马(Rozema)等人的方法(PNAS,2007年8月7 日,第104卷,第32期)也可应用于本发明。动态多聚缀合物技术的关键特 征包括膜活性聚合物、可逆地掩蔽这种聚合物的活性直到它到达内体的酸性 环境时为止的能力、以及使这种修饰的聚合物和它的siRNA负荷物在简单的 低压静脉注射之后特异性靶向体内肝细胞的能力。通过使5'胺修饰的siRNA 与1个重量当量(wt eq)的N-琥珀酰亚胺基-S-乙酰硫代乙酸酯(SATA)试 剂(皮尔斯(Pierce)公司)和0.36wt eq的在水中的$NaHCO_3$在4℃反应16 h合成SATA修饰的siRNA。然后通过加入9倍体积的乙醇并且在80℃孵育 2h使修饰的siRNA沉淀。将该沉淀重新悬浮在1X siRNA缓冲液中 (Dharmacon公司)并且通过在260-nm波长处测量吸光度进行定量。通过添 加1.5wt%SMPT(皮尔斯(Pierce)公司)修饰PBAVE(30mg/ml,在5 mM TAPS中,pH 9)。在1小时的孵育之后,将0.8mg的SMPT-PBAVE添加到400μl的含有5mM TAPS的等渗葡萄糖溶液中(pH 9)。向此溶液中添 加50μg的SATA修饰的siRNA。对于其中[PBAVE]是恒定的剂量反应实 验,添加不同量的siRNA。然后将该混合物孵育16小时。然后向此溶液中加 入5.6mg的Hepes游离碱,接着加入3.7mg的CDM-NAG与1.9mg的CDM- PEG的混合物。然后在注射之前将该溶液在室温下孵育至少1小时。从经由 使用草酰氯产生的酰基氯合成CDM-PEG和CDM-NAG。向该酰基氯加入1.1 摩尔当量的聚乙二醇单甲醚(平均分子量450)以产生CDM-PEG或(氨基乙 氧基)乙氧基-2-(乙酰氨基)-2-脱氧-β-D-吡喃葡萄糖苷,从而产生CDM- NAG。通过使用反相HPLC以0.1%TFA水/乙腈梯度纯化终产物。向小鼠递 送约25到50μg的siRNA。例如,通过设想用于递送到人类的约50到约200 mg剂量的CRISPR Cas,罗泽马(Rozema)等人的系统可应用于递送到肝脏 的本发明的CRISPR Cas系统。

[0488] 骨

[0489] 奥克斯(Oakes)和利伯曼(Lieberman)《临床矫形学及相关学科研 究》(Clin Orthop Relat Res.)2000年10月;(379增刊):S101-12)论述 了向骨的基因递送。通过将基因转移到在特定解剖部位的细胞中,可以使用 持续一段时间的生理学剂量的生长因子的骨诱导特性来促进更显著的愈合反 应。该特定的解剖部位、骨的质量、以及软组织包膜,影响区域性基因治疗 的靶细胞的选择。在骨诱导载体中递送到治疗部位的基因治疗载体产生了有 希望的结果。若干研究者在动物模型中使用离体和体内区域性基因治疗显示 出令人兴奋的结果。这样的系统可以用于和/或适用于递送到骨的本发明的 CRISPR Cas系统。

[0490] 脑

[0491] 用于脑的递送选项包括将处于DNA或RNA形式的CRISPR酶和指导 RNA封装到脂质体中并且结合到用于跨血脑屏障(BBB)递送的分子特洛伊 木马上。分子特洛伊木马已经显示对于将B-gal表达载体递送到非人类灵长动 物的脑中是有效的。可以使用相同的途径来递送含有CRISPR酶和指导RNA 的递送载体。例如,夏CF(Xia CF)和博阿多RJ(Boado RJ)、帕德瑞吉 WM(Pardridge WM)"使用亲和素-生物素技术经由人胰岛素受体的siRNA 的抗体介导的靶向"("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology."《分子药理学》(Mol Pharm.) 2009年5月-6月;6(3):747-51.doi:10.1021/mp800194)描述了如何将短干扰 RNA(siRNA)递送到培养物中以及体内的细胞,并且可以与受体特异性单 克隆抗体(mAb)和亲和素-生物素技术联合

使用。作者还报道,就亲和素-生 物素技术而言,由于在靶向mAb与siRNA之间的键是稳定的,在静脉给予该 靶向的siRNA之后,体内观察到在远处部位(例如脑)的RNAi效应。

[0492] 张(Zhang)等人《分子治疗学》(Mol Ther.)2003年1月;7(1):11- 8.))描述了编码报道分子如萤光素酶的表达质粒如何被包封在一种“人工病 毒”的内部,该“人工病毒”包含85nm的聚乙二醇化的免疫脂质体,其在 体内与一种针对人胰岛素受体(HIR)的单克隆抗体(MAb)一起靶向到恒 河猴脑。该HIRMAb使得携带外源基因的脂质体在静脉注射之后经历跨血脑 屏障的转胞吞作用和跨神经元质膜的胞吞作用。与大鼠比较,在恒河猴的脑 中的萤光素酶基因表达水平高50倍。通过组织化学和共聚焦显微镜检查都证 明了β-半乳糖苷酶基因在灵长动物脑中的广泛神经元表达。作者指明,这种 途径使得在24小时内的可逆成年转基因可行。因此,免疫脂质体的使用是优 选的。这些可以与抗体结合靶向特定的组织或细胞表面蛋白。

[0493] 其他递送手段或RNA也是优选的,如经由纳米粒子(卓(Cho)S.、金 伯格(Goldberg)M.、松(Son)S.、许(Xu)Q.、杨(Yang)F.、梅 (Mei)Y.、博加特廖夫(Bogatyrev)S.、朗格(Langer)R.和安德森 (Anderson)D.,“用于将小干扰RNA递送到内皮细胞的脂质样纳米粒子” (Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells),《先进功能材料》(Advanced Functional Materials),19:3112-3118,2010)或外泌体(施罗德(Schroeder)A.、莱文斯(Levins)C.、科 特斯(Cortez)C.、朗格(Langer)R.、和安德森(Anderson)D.,“用于 siRNA递送的基于脂质的纳米治疗剂” (Lipid-based nanotherapeutics for siRNA delivery),《内科学杂志》(Journal of Internal Medicine),267:9-21, 2010,PMID:20059641)。

[0494] 事实上,已经表明外泌体在递送siRNA中特别有用,其为与CRISPR系 统有一些相似之处的系统。例如,艾尔·安达卢西(El-Andaloussi)等人“外 泌体诱导的体外和体内siRNA递送”(“Exosome-mediated delivery of siRNA in vitro and in vivo.”)《自然-实验手册》(Nat Protoc.)2012年12月;7(12): 2112-26.doi:10.1038/nprot.2012.131。电子版2012年11月15日描述了外泌体 对于跨不同的生物屏障的药物递送是有希望的工具并且可以用于体外和体内 递送siRNA。其途径在于通过转染一种包含与肽配体融合的外泌体蛋白的表 达载体产生靶向外泌体。然后将这些外泌体纯化并且从转染的细胞上清液中 表征,然后将siRNA加载到外泌体中。根据本发明的递送或给药可以用外泌 体进行,尤其是但不限于脑。

[0495] 维生素E(α-生育酚)可以与CRISPR Cas结合并且连同高密度脂蛋白 (HDL)一起递送到脑,例如以与乌诺(Uno)等人完成的用于将短干扰 RNA(siRNA)递送到脑的类似方式《人类基因治疗》(HUMAN GENE THERAPY)22:711-719(2011年6月)。经由用磷酸盐缓冲盐水(PBS)或 游离TocsiBACE或Toc-siBACE/HDL充满的并且与脑灌注试剂盒3(Brain Infusion Kit 3)(Alzet公司)连接的微渗透压泵(型号1007D;Alzet公司, 库比蒂诺(Cupertino),CA)灌注小鼠。将一根脑灌注导管置于在正中线的 前囟的后方约0.5mm,用于灌注到背侧第三脑室中。乌诺(Uno)等人发 现,通过相同的ICV灌注方法,少至3nmol的Toc-siRNA与HDL可诱导相 当程度的靶减少。可以考虑结合至α-生育酚的并且与HDL共同给予靶向脑的 相似剂量的CRISPR Cas用于本发明,例如,可以考虑靶向脑的约3nmol到 约3μmol的CRISPR Cas。

[0496] 邹(Zou)等人(《人类基因治疗》(HUMAN GENE THERAPY 22: 465-475(2011年4月))描述了靶向PKCγ的短发夹RNA的慢病毒介导的 递送方法,其用于在大鼠脊髓中的体内基因沉默。邹(Zou)等人通过鞘内导 管给予约10μl的具有1x 10^9个转导单位(TU)/ml的滴度的重组慢病毒。在 本发明中可以考虑相似剂量的在靶向脑的慢病毒载体中表达的CRISPR Cas用 于人类,例如,可以考虑靶向脑的在具有1x 10^9个转导单位(TU)/ml的滴 度的慢病毒中的约10-50ml的CRISPR Cas。

[0497] 靶向缺失、治疗应用

[0498] 基因的靶向缺失是优选的。例子示例于实例18中。因此优选的是涉及胆 固醇生物合成、脂肪酸生物合成、以及其他代谢障碍的基因,编码涉及淀粉 样蛋白以及其他疾病的错误折叠蛋白的基因,导致细胞转化的癌基因,潜伏 病毒基因,以及导致负显性疾病(dominant-negative disorder)、其他失调的 基因。在这里作为示例,申请人提出使用病毒或纳米粒子递送系统向对其有 需要的受试者或患者的肝脏、脑、眼、上皮、造血、或另一种组织进行 CRISPR-Cas系统的基因递送,该受试者或患者具有代谢障碍、淀粉样变性、 蛋白聚积相关疾病、由于基因突变和基因异位所致的细胞转化、基因突变的 负显性作用、潜伏病毒感染、以及其他相关症状。

[0499] CRISPR-Cas系统的治疗应用包括青光眼、淀粉样变性、和亨廷顿病。这 些示例于实例20中,并且在其中描述的特征是单独地或以组合方式优选的。

[0500] 可以通过本发明治疗的多聚谷氨酰胺扩增疾病(polyglutamine expansion disease)的另一个实例包括脊髓小脑共济失调1型(SCA1)。在小脑内注射 后,表达短发夹RNA的重组腺相关病毒(AAV)载体深刻改进了动作协调、 恢复了小脑形态并且溶解了在SCA1小鼠的浦肯野细胞中的特征性脊髓小脑 失调症蛋白1包涵体(参见,例如,夏(Xia)等人,《自然医学》(Nature Medicine),第10卷,第8期,2004年8月)。具体地说,AAV1和AAV5 载体是优选的,并且约1x 10^{12}个载体基因组/ml的AAV滴度是令人希望的。

[0501] 作为一个实例,可以治疗或预防由HIV-1引起的慢性感染。为了实现这 个目标,可以产生靶向绝大多数HIV-1基因组的CRISPR-Cas指导RNA,同 时考虑HIV-1株变体,以使覆盖面和效力最大。通过宿主免疫系统的常规腺 病毒或慢病毒介导的感染,可以实现CRISPR-Cas系统的递送。取决于途径, 宿主免疫细胞可以是a)分离的、用CRISPR-Cas转导的、选择的、并且重新 引入该宿主中的或b)通过全身递送CRISPR-Cas系统体内转导的。第一种途 径允许产生抗性免疫群体,而第二种更可能靶向宿主体内的潜伏病毒储库。 这在实例部分中更详细地论述。

[0502] 在另一个实例中,转让给桑加莫生物科学公司(Sangamo BioSciences, Inc.)的美国专利公开号20130171732涉及抗HIV转基因到该基因组中的插 入,其方法可应用于本发明的CRISPR Cas系统。在另一个实施例中,该 CXCR4基因可以被靶向并且转让给桑加莫生物科学公司(Sangamo BioSciences,Inc.)的美国专利公开号20100291048的TALE系统可以被修饰 为本发明的CRISPR Cas系统。转让给桑加莫生物科学公司(Sangamo BioSciences,Inc.)美国专利公开号20130137104和和20130122591以转让给 Cellectis公司的美国专利公开号20100146651的方法通常可应用于转基因表 达,因为它涉及修饰用于增加基因修饰的频率的次黄嘌呤鸟嘌呤磷酸核糖转 移酶(HPRT)座位。

[0503] 还可以设想的是,本发明产生一个基因敲除细胞文库。每个细胞可以有 单基因被

敲除。这被示例于实例23中。

[0504] 人们可以制作ES细胞的文库,其中每个细胞有单基因被敲除,并且ES 细胞的整个文库将有每个单基因被敲除。此文库对于筛选细胞过程以及疾病 的基因功能是有用的。为了制作这个细胞文库,人们可以将由诱导型启动子 (例如,多西环素诱导型启动子)驱动的Cas9整合到ES细胞中。另外,人 们可以将靶向特异性基因的单个指导RNA整合到ES细胞中。为了制作ES 细胞文库,人们可以简单地将ES细胞与编码靶向人类基因组中的每个基因的指导RNA的基因的文库混合。人们可以首先将单个BxB1 attB位点引入到人 ES细胞的AAVS1座位中。然后可以使用BxB1整合酶来促进单独的指导 RNA基因整合到AAVS1座位中的BxB1 attB位点中。为了促进整合,每个指 导RNA基因可以被包含在携带单个attP位点的质粒上。这样,BxB1将使基 因组中的attB位点与含有指导RNA的质粒上的attP位点重组。为了生成该细 胞文库,人们可以采用具有整合的单个指导RNA的细胞的文库,并且诱导 Cas9表达。在诱导之后,Cas9介导由该指导RNA指定的位点处的双链断 裂。

[0505] 蛋白质治疗剂的长期给药可以引出针对该特殊蛋白质的不可接受的免疫 反应。蛋白质药物的免疫原性可以归因于少数免疫显性辅助性T淋巴细胞 (HTL)表位。降低包含在这些蛋白质内的这些HTL表位的MHC结合亲和 力可以产生具有更低免疫原性的药物(腾格里S(Tangri S)等人“具有降低 免疫原性的合理工程化的治疗蛋白”(“Rationally engineered therapeutic proteins with reduced immunogenicity”《免疫学杂志》(J Immunol.)2005年3 月15日;174(6):3187-96。)在本发明中,可以具体地遵循首先由腾格里(Tangri)等人针对促红细胞生成素提出并且随后开发的方法来降低该 CRISPR酶的免疫原性。因此,定向演化或合理设计可以用来降低在宿主物种 (人类或其他物种)中的CRISPR酶(例如Cas9)的免疫原性。

[0506] 在实例28中,申请人使用了3种感兴趣的指导RNA,并且这些指导 RNA能够将仅仅发生在一个小的细胞亚群中的有效体内DNA切割可视化。实质上,申请人在这里示出的是靶向性体内切割。具体地说,这提供了也可 以在高级生物(如哺乳动物)中实现特异性靶向的概念证明。还强调的多元 方面在于,可以同时使用多重指导序列(即,分开的靶标)(从共同递送的 意义上说)。换言之,申请人使用了多重途径,其中几种不同的序列被同时 而独立地靶向。

[0507] 适合的用于产生AAV(本发明的一种优选的载体)的方案的实例提供在 实例34中。

[0508] 三核苷酸重复障碍是优选的有待治疗的病症。这些也示例于本文中。

[0509] 例如,美国专利公开号20110016540描述了使用锌指核酸酶基因修饰与 三核苷酸重复扩增障碍相关的细胞、动物以及蛋白质。三核苷酸重复扩增障 碍是复杂的、进行性的疾病,其涉及发育神经生物学并且常常影响认知以及 感觉运动功能。

[0510] 三核苷酸重复扩增蛋白是一套多样化的蛋白质,这些蛋白质与发生三核 苷酸重复扩增障碍的易感性、三核苷酸重复扩增障碍的存在、三核苷酸重复 扩增障碍的严重性或其任何组合相关联。三核苷酸重复扩增障碍被分为由重 复类型决定的两个类别。最常见的重复是三联体CAG,当出现在一个基因的 编码区中时,其编码氨基酸谷氨酰胺(Q)。因此,这些障碍被称为多聚谷氨 酰胺(polyQ)障碍并且包括下列疾病:亨廷顿病(HD);脊延髓肌萎缩症 (SBMA);脊髓小脑性共济失调(SCA型1、2、3、6、7、和17);以及 齿状核红核苍白球丘脑下部核萎缩(DRPLA)。其余的三核苷酸重复扩增障 碍不涉及CAG三联体,或者该CAG三联体

不在该基因的编码区中,并且因 此被称为非多聚谷氨酰胺障碍。非多聚谷氨酰胺障碍包括脆性X综合征 (FRAXA);脆性XE精神发育迟滞(FRAXE);弗里德赖希共济失调 (FRDA);肌强直性营养不良(DM);和脊髓小脑性共济失调(SCA型 8、和12)。

[0511] 与三核苷酸重复扩增障碍相关联的蛋白质典型地是基于三核苷酸重复扩 增障碍相关性蛋白质与三核苷酸重复扩增障碍的实验性关联而选择的。例 如,相对于没有三核苷酸重复扩增障碍的群体,在具有三核苷酸重复扩增障 碍的群体中,与三核苷酸重复扩增障碍相关的蛋白质的产生率或循环浓度可 以升高或降低。可以使用蛋白质组技术评估蛋白质水平差异,包括但不限于 蛋白质印迹、免疫组织化学染色、酶联免疫吸附测定(ELISA)以及质谱 法。可替代地,可以使用基因组技术通过获得编码这些蛋白质的基因的基因 表达谱而鉴定与三核苷酸重复扩增障碍相关的蛋白质,包括但不限于DNA微 阵列分析、基因表达系列分析(SAGE)以及定量实时聚合酶链式反应(Q- PCR)。

[0512] 与三核苷酸重复扩增障碍相关联的蛋白质的非限制性实例包括AR(雄激 素受体)、FMR1(脆性x精神发育迟滞1)、HTT(亨廷丁)、DMPK(肌 强直性营养不良蛋白激酶)、FXN(线粒体型共济失调蛋白)、ATXN2(脊 髓小脑共济失调蛋白2)、ATN1(萎缩蛋白1)、FEN1(片段结构特异内切 核酸酶1)、TNRC6A(含有6A的三核苷酸重复)、PABPN1(多聚(A)结 合蛋白,核1)、JPH3(亲联蛋白3)、MED15(介体复合物亚基15)、 ATXN1(脊髓小脑共济失调蛋白1)、ATXN3(脊髓小脑共济失调蛋白 3)、TBP(TATA盒结合蛋白)、CACNA1A(钙通道,电压依赖性,P/Q 型、α1A亚基)、ATXN80S(ATXN8相反链(非蛋白质编码))、PPP2R2B (蛋白磷酸酶2,调节亚基B,β)、ATXN7(脊髓小脑共济失调蛋白7)、 TNRC6B(含有6B的三核苷酸重复)、TNRC6C(含有6C的三核苷酸重 复)、CELF3(CUGBP、Elav样家族成员3)、MAB21L1(mab-21-样1(秀 丽隐杆线虫))、MSH2(MutS同源物2,结肠癌,无息肉病型1(大肠杆 菌))、TMEM185A(跨膜蛋白185A)、SIX5(SIX同源框5)、CNPY3 (冠层3同源物(斑马鱼))、FRAXE(脆性部位,叶酸型,罕见型,fra (X)(q28)E)、GNB2(鸟嘌呤核苷酸结合蛋白(G蛋白)、β多肽 2)、RPL14(核糖体蛋白L14)、ATXN8(脊髓小脑共济失调蛋白8)、 INSR(胰岛素受体)、TTR(转甲状腺素蛋白)、EP400(E1A结合蛋白 p400)、GIGYF2(GRB10相互作用GYF蛋白2)、OGG1(8-氧鸟嘌呤 DNA糖基化酶)、STC1(斯钙素1)、CNDP1(肌肽二肽酶1(金属肽酶 M20家族))、C10orf2(染色体10开放阅读框2)、MAML3智者基因样3 (果蝇)、DKC1(先天性角化不全症1,角化不良蛋白)、PAXIP1(PAX 相互作用(具有转录激活域)蛋白1)、CASK(钙/钙调蛋白依赖性丝氨酸蛋 白激酶(MAGUK家族)、MAPT(微管相关蛋白tau)、SP1(Sp1转录因 子)、POLG(聚合酶(DNA指导的),γ)、AFF2(AF4/FMR2家族,成员 2)、THBS1(凝血酶敏感蛋白1)、TP53(肿瘤蛋白p53)、ESR1(雌激素 受体1)、CGGBP1(CGG三联体重复结合蛋白1)、ABT1(基本转录激活 因子1)、KLK3(激肽释放酶相关肽酶3)、PRNP(朊病毒蛋白)、JUN (jun癌基因)、KCNN3(钾中间/小电导钙激活通道,亚家族N,成员 3)、BAX(BCL2相关X蛋白)、FRAXA(脆性部位,叶酸型,罕见型, fra(X)(q27.3)A(巨睾丸,精神发育迟滞))、KBTBD10(Kelch重复和 BTB(POZ)域包含蛋白10)、MBNL1(盲肌样(果蝇))、RAD51 (RAD51同源物(RecA同源物,大肠杆菌)(酿酒酵母))、NCOA3(核 受体共激活因子3)、ERDA1(扩展重复结构域,CAG/CTG 1)、TSC1(结 节性硬化1)、COMP(软骨寡聚基质蛋白)、GCLC(谷氨酰半胱氨酸连接 酶,催化亚基)、RRAD(Ras相关关联糖尿病)、MSH3(mutS同源物3 (大肠杆菌))、DRD2(多巴胺受体D2)、CD44(CD44分子(印度血型))、CTCF(CCCTC结合因子(锌指蛋白))、CCND1(细胞周期蛋白 D1)、CLSPN(扣蛋白同源物

(非洲爪蟾))、MEF2A(肌细胞增强因子 2A)、PTPRU(蛋白酪氨酸磷酸酶,受体型,U)、GAPDH(3-磷酸甘油醛 脱氢酶)、TRIM22(三基序蛋白22)、WT1(维尔姆斯瘤1)、AHR(芳香 烃受体)、GPX1(谷胱甘肽过氧化物酶1)、TPMT(硫嘌呤甲基转移 酶)、NDP(诺里病(假神经胶质瘤))、ARX(无芒相关同源框)、 MUS81(MUS81内切核酸酶同源物(酿酒酵母))、TYR(酪氨酸酶(眼皮肤白化病IA))、EGR1(早期生长反应蛋白1)、UNG(尿嘧啶DNA糖基 化酶)、NUMBL(麻木同源物(果蝇)样)、FABP2(脂肪酸结合蛋白2, 肠)、EN2(锯齿状同源框2)、CRYGC(晶状体蛋白,γC)、SRP14(信 号识别粒子14kDa(同源Alu RNA结合蛋白))、CRYGB(晶状体蛋白,γ B)、PDCD1(程序性细胞死亡1)、HOXA1(同源框A1)、ATXN2L(脊 髓小脑共济失调蛋白2样)、PMS2(PMS2减数分裂后分离增加2样蛋白 (酿酒酵母))、GLA(半乳糖苷酶,α)、CBL(Cas-Br-M(鼠)热带逆转 录病毒转化序列)、FTH1(铁蛋白,重多肽1)、IL12RB2(白细胞介素12 受体,β2)、OTX2(正小齿同源框2)、HOXA5(同源框A5)、POLG2 (聚合酶(DNA指导的),γ2,辅助亚基)、DLX2(末端减少同源框2)、 SIRPA(信号调节蛋白α)、OTX1(正小齿同源框1)、AHRR(芳香烃受体 阻抑物)、MANF(中脑星形胶质细胞衍生神经营养因子)、TMEM158(跨 膜蛋白158(基因/假基因))、以及ENSG00000078687。

[0513] 与三核苷酸重复扩增障碍相关联的优选蛋白质包括HTT(亨廷丁)、AR (雄激素受体)、FXN(线粒体型共济失调蛋白)、Atxn3(脊髓小脑共济失 调蛋白)、Atxn1(脊髓小脑共济失调蛋白)、Atxn2(脊髓小脑共济失调蛋 白)、Atxn7(脊髓小脑共济失调蛋白)、Atxn10(脊髓小脑共济失调蛋 白)、DMPK(肌强直性营养不良蛋白激酶)、Atn1(萎缩蛋白1)、CBP (creb结合蛋白)、VLDLR(极低密度脂蛋白受体)、及其任何组合。

[0514] 根据另一个方面,提供了一种用于治疗具有在CFTR基因中的突变的受 试者的基因治疗方法,并且该方法包括任选地经由一种生物相容性药用载体 向受试者的细胞给予治疗有效量的CRISPR-Cas基因治疗粒子。优选地,该靶 DNA包含突变δF508。通常,优选的是该突变被修复为野生型。在这种情况 下,该突变是三个核苷酸的缺失,这些核苷酸包括在位置508处的针对苯丙 氨酸(F)的密码子。因此,在这种情况下的修复要求将缺失密码子重新引入 该突变体中。

[0515] 为了实施这个基因修复策略,优选的是将一个腺病毒/AAV载体系统引入 该宿主细胞、多个细胞或患者中。优选地,该系统包含Cas9(或Cas9切口 酶)和指导RNA连同腺病毒/AAV载体系统,其包含含有F508残基的同源性 修复模板。这可以经由较早论述的递送方法之一引入受试者。该CRISPR-Cas 系统可以由该CFTRδ508嵌合指导RNA指导。它靶向有待切割或切割的 CFTR基因组座位的特定位点。在切割之后,将修复模板经由同源重组插入切割位点中,从而修正导致囊性纤维化或引起囊性纤维化相关症状的缺失。可 以采用用来指导RNA递送并提供CRISPR系统的系统性引入的这个策略来靶 向基因突变,以编辑或另外地操纵如那些在表B中的引起代谢障碍、肝脏、肾和蛋白质病和障碍的基因。

[0516] 基因组编辑

[0517] 本发明的CRISPR/Cas9系统可以用来校正基因突变,先前使用TALEN 和ZFN尝试了这些突变,但是成功有限。例如,WO 2013163628 A2,《突变 基因的基因校正》(Genetic Correction of Mutated Genes),杜克大学的公开 申请,描述了例如校正一种框移突变的努力,该框移突变引起提前终止密码 子和可经由核酸酶介导的非同源末端连接进行校正的截短基因产物,该基因 产物如引起迪谢内肌营养不良(“DMD”)的那些,迪谢内肌营养不

良是一 种隐性遗传的、致命的、X连锁疾病，其导致由于肌营养不良蛋白基因突变 所致的肌肉变性。引起DMD的大多数肌营养不良蛋白突变是破坏该阅读框并 且引起肌营养不良蛋白基因的提前翻译终止的外显子缺失。肌营养不良蛋白 是一种细胞质蛋白，其提供负责调节肌细胞完整性和功能的细胞膜肌营养不 良蛋白聚糖复合物的结构稳定性。如在此可互换地使用的肌营养不良蛋白基 因或“DMD基因”是在座位Xp21处的2.2兆碱基。初级转录测量了约2,400 kb，其中成熟mRNA为约14kb。79个外显子编码超过3500个氨基酸的蛋白 质。在DMD患者中，外显子51常常接近破坏框的缺失并且已经在临床试验 中被靶向基于寡核苷酸的外显子跳跃。对于外显子51跳跃化合物依替利森 (eteplirsen)的临床试验最近报道了跨48周的显著功能益处，相比于基线具 有47％的肌营养不良蛋白阳性纤维。外显子51的突变理想地适合于经由基于 NHEJ的基因组编辑的持久校正。

[0518] 涉及从人类肌营养不良蛋白基因(DMD)切割靶序列的大范围核酸酶变 体的转让给Cellectis公司的美国专利公开号20130145487的方法也可以针对 本发明的CRISPR Cas系统进行修改。

[0519] 血液

[0520] 本发明还考虑了向血液递送该CRISPR-Cas系统。

[0521] 瓦尔格伦(Wahlgren)等人的血浆外泌体(《核酸研究》(Nucleic Acids Research)，2012年，第40卷，第17期，e130)被较早描述并且利用为向血 液递送该CRISPR Cas系统。

[0522] 还考虑本发明的CRISPR Cas系统用来治疗血红蛋白病，如地中海贫血和 镰状细胞病。参见，例如，可以被本发明的CRISPR Cas系统靶向的潜在靶标 的国际专利公开号WO 2013/126794。

[0523] 转让给Cellectis公司的美国专利公开号20110225664、20110091441、20100229252、20090271881和20090222937，涉及CREI变体，其中两个I- CreI单体的至少一个具有至少两个取代，一个在分别位于从I-CreI的位置26 到40以及44到77的LAGLIDADG核心域的两个功能性亚结构域的每一个 中，能够从人白细胞介素2受体γ链(IL2RG)基因切割DNA靶序列的所述 变体也称为普通细胞因子受体γ链基因或γC基因。在美国专利公开号 20110225664、20110091441、20100229252、20090271881和20090222937中 鉴定的靶序列可以用于本发明的CRISPR Cas系统。

[0524] 由于在淋巴细胞T成熟方面的缺陷所致的严重联合免疫缺陷(SCID)总 是与淋巴细胞B的功能缺陷相关联(卡瓦扎纳-卡尔沃(Cavazzana-Calvo)等 人，《医学年鉴》(Annu.Rev.Med.)，2005，56，585-602；费舍尔 (Fischer)等人，《免疫学评论》(Immunol.Rev.)，2005，203，98- 109)。总发病率估计为每75 000个出生人数有1个。患有未经治疗的SCID 的患者遭受多重机会性微生物感染，并且通常不能活过一年。SCID可以通过同种异体造血干细胞(来自家族供体)转移进行治疗。对于供体而言的组织 相容性可在很大程度上变化。在腺苷脱氨酶(ADA)缺陷的情况下，SCID之 一形成，患者可以通过注射酶重组腺苷脱氨酶进行治疗。

[0525] 由于已经显示ADA基因在SCID患者中突变(吉布勒特(Giblett)等 人，《柳叶刀》(Lancet)，1972，2，1067-1069)，已经鉴定了几种其他的 涉及SCID的基因(卡瓦扎纳-卡尔沃(Cavazzana-Calvo)等人，《医学年 鉴》(Annu.Rev.Med.)，2005，56，585-602；费舍尔

(Fischer)等人，《免疫学评论》(Immunol.Rev.)，2005，203，98-109)。对于SCID有四 种主要原因：(i)最常见的SCID形式，SCID-X1(X连锁的SCID或X- SCID)，是由IL2RG基因的突变引起的，导致成熟T淋巴细胞和NK细胞的 缺乏。IL2RG编码γC蛋白(野口(Noguchi)等人，《细胞》(Cell)， 1993，73，147-157)，其为至少五种白细胞介素受体复合物的一种共有组分。这些受体通过JAK3激酶激活几种靶标(马基(Macchi)等人，《自然》 (Nature)，1995，377，65-68)，其失活导致与γC失活相同的综合征； (ii)ADA基因的突变导致嘌呤代谢缺陷，这对于淋巴细胞前体是致命的， 进而导致B、T和NK细胞的准缺乏；(iii)V(D)J重组在免疫球蛋白和T淋 巴细胞受体(TCR)的成熟中是一个重要步骤。在涉及这个过程的三种基因重组活化基因1和2(RAG1和RAG2)以及阿蒂米斯(Artemis)中的突变导 致成熟T和B淋巴细胞的缺乏；以及(iv)还报道了在涉及T细胞特异性信 号传导的其他基因如CD45中的突变，虽然它们代表少数情形(卡瓦扎纳-卡 尔沃(Cavazzana-Calvo)等人，《医学年鉴》(Annu.Rev.Med.)，2005， 56，585-602；费舍尔(Fischer)等人，《免疫学评论》(Immunol.Rev.)， 2005，203，98-109)。

[0526] 出于两个主要原因，自从它们的遗传基础已经鉴定以来，不同的SCID 形式已经成为基因治疗方法的范例(费舍尔(Fischer)等人，《免疫学评 论》(Immunol.Rev.)，2005，203，98-109)。首先，是由于在所有血液疾 病中可以设想离体治疗。造血干细胞(HSC)可以从骨髓恢复，并且在较少 的细胞分裂中保持了它们的多能性。因此，可以将它们进行体外处理，然后 重新注射到患者中，它们入驻骨髓中。其次，由于淋巴细胞的成熟在SCID患 者中受损，经校正的细胞具有选择性优势。因此，少量的校正细胞可恢复功 能性免疫系统。这种假设借助于下列各项确认数次(i)与SCID患者中的突 变的逆转相关联的免疫功能的部分恢复(赫希霍恩(Hirschhorn)等人，《自 然-遗传学》(Nat.Genet.)，1996，13，290-295；斯蒂芬(Stephan)等人，《新英格兰医学杂志》(N.Engl.J.Med.)，1996，335，1563-1567；布索(Bousso)等人，《美国国家科学院院刊》(Proc.Natl.，Acad.Sci.USA)， 2000，97，274-278；和田(Wada)等人，《美国国家科学院院刊》，2001， 98，8697-8702；西小森(Nishikomori)等人，《血液》(Blood)，2004， 103，4565-4572)，(ii)在造血细胞中的体外SCID-X1缺陷的校正(坎多蒂 (Candotti)等人，《血液》，1996，87，3097-3102；卡瓦扎纳-卡尔沃 (Cavazzana-Calvo)等人，《血液》，1996，《血液》，88，3901-3909；泰 勒(Taylor)等人，《血液》，1996，87，3103-3107；哈辛-贝(Hacein- Bey)等人，《血液》，1998，92，4090-4097)，(iii)SCID-X1的校正 (苏戴斯(Soudais)等人，《血液》，2000，95，3071-3077；蔡(Tsai)等 人，《血液》，2002，100，72-79)，JAK-3(邦廷(Bunting)等人，《自然 医学》(Nat.Med.)，1998，4，58-64；邦廷等人，《人类基因治疗》 (Hum.Gene Ther.)，2000，11，2353-2364)和RAG2(耶茨(Yates)等 人，《血液》，2002，100，3942-3949)在动物模型中的体内缺陷以及(iv) 基因治疗临床试验的结果(卡瓦扎纳-卡尔沃等人，《科学》，2000，288， 669-672；艾尔迪(Aiuti)等人，《自然医学》(Nat.Med.)，2002；8，423- 425；加斯帕(Gaspar)等人，《柳叶刀》(Lancet)，2004，364，2181-2187)。

[0527] 转让给儿童医疗中心有限公司(Children's Medical Center Corporation) 和哈弗大学(President and Fellows of Harvard College)的美国专利公开号20110182867涉及经由BCL11A表达或活性抑制剂(如RNAi和抗体)在造血 祖细胞中调节胎儿血红蛋白(HbF)表达的方法和用途。在美国专利公开号20110182867中披露的靶标，如

BCL11A,可以被本发明的CRISPR Cas系统 靶向,以便调节胎儿血红蛋白表达。对于另外的BCL11A靶标,还参见鲍尔 (Bauer)等人(《科学》(Science)2013年10月11日:第342卷,第6155期,第253-257页)和许(Xu)等人(《科学》2011年11月18日:第 334卷,第6058期,第993-996页)。

[0528] 耳

[0529] 本发明还考虑了向一只或两只耳递送该CRISPR-Cas系统。

[0530] 研究者正在调查基因治疗是否可以用来辅助当前的耳聋治疗—即,耳蜗 植入物。耳聋常常由毛细胞的缺失或损害引起,其不能将信号传递到听觉神 经元。在这样的情况下,可以使用耳蜗植入物对声音做出响应并且将电信号 传输到神经细胞。但是,由于受损的毛细胞释放释放更少的生长因子,这些 神经元常常退化并从耳蜗缩回。

[0531] 美国专利申请20120328580描述了例如使用一个注射器(例如单剂量注 射器)将一种药物组合物注射到耳中(例如,耳部施用),如注射到耳蜗的 腔(luminae)中(例如,中阶、前庭阶、和鼓阶)。例如,可以通过鼓室内 注射(例如,进入中耳)、和/或注射到外耳、中耳、和/或内耳给予本文描述 的一种或多种化合物。这样的方法在本领域中常规使用,例如,用于将甾体 和抗生素给予到人耳中。例如,可以通过耳的圆窗或通过耳蜗囊进行注射。其他内耳施用方法是本领域已知的(参见,例如,索尔特(Salt)和普兰特克 (Plontke),《今日药物发现》(Drug Discovery Today),10:1299-1306, 2005)。

[0532] 在另一种给药方式中,可以经由一根导管或泵原位施用该药物组合物。导管或泵可以,例如,将药物组合物导向到耳蜗腔或耳的圆窗和/或结肠腔 中。适合于将本文所述的一种或多种化合物施用到耳(例如,人耳)中的示 例性药物递送设备和方法由麦肯纳(McKenna)等人,(美国公开号 2006/0030837)和雅各布森(Jacobsen)等人,(美国专利号7,206,639)描 述。在一些实施例中,可以在外科手术过程中将导管或泵定位在例如患者的耳中(例如,外耳、中耳和/或内耳)。在一些实施例中,可以将导管或泵定 位在例如患者的耳中(例如,外耳、中耳和/或内耳),而不需要外科手术。

[0533] 可替代地或另外地,本文所述的一种或多种化合物可以与一种佩戴在外 耳中的机械装置(如一种耳蜗植入物或助听器)联合使用。适合于与本发明 一起使用的示例性耳蜗植入物由埃奇(Edge)等人(美国公开号 2007/0093878)描述。

[0534] 在一些实施例中,上述给药方式可以按照任何顺序组合并且可以是同时 的或者散布的。

[0535] 可替代地或另外地,本发明可以根据食品药品监督管理局批准的任何方 法给药,例如,如描述于《CDER数据标准手册》(CDER Data Standards Manual),版本号004(在fda.give/cder/dsm/DRG/drg00301.htm可获得)。

[0536] 一般而言,在美国专利申请20120328580中描述的这些细胞治疗方法可 以用来体外促进细胞向或朝向内耳的成熟细胞类型(例如,毛细胞)的完全 或部分分化。从这样的方法得到的细胞然后可以移植或植入到需要这种治疗 的患者中。下面描述了实践这些方法所需的细胞培养方法,包括用于鉴定和 选择适当细胞类型的方法、用于促进选定细胞的完全或部分分化的方法、用 于鉴定完全或部分分化的细胞类型的方法、以及用于植入完全或部分分化的 细胞的方法。

[0537] 适合于在本发明中使用的细胞包括但不限于,当与本文所述的一种或多 种化合

物例如体外接触时能够完全或部分分化成内耳的成熟细胞，例如，毛 细胞(例如，内耳和/或外耳毛细胞)的细胞。能够分化成毛细胞的示例性细 胞包括但不限于干细胞(例如，内耳干细胞、成体干细胞、骨髓源性干细 胞、胚胎干细胞、间充质干细胞、皮肤干细胞、iPS细胞、和脂肪来源干细 胞)、祖细胞(例如，内耳祖细胞)、支持细胞(例如，戴特斯细胞 (Deiters' cell)、柱细胞、内指状细胞、顶盖细胞和汉森细胞)、和/或生殖 细胞。干细胞用于替代内耳感觉细胞的用途描述于Li等人(美国公开号 2005/0287127)和李(Li)等人(美国专利序列号11/953,797)。骨髓源性干 细胞用于替代内耳感觉细胞的用途描述于埃奇(Edge)等人的PCT/US 2007/084654。iPS细胞描述于，例如，高桥(Takahashi)等人，《细胞》 (Cell)，第131卷，第5次发行，第861-872页(2007)；高桥 (Takahashi)和山中(Yamanaka)，《细胞》126，663-76(2006)；冲田 (Okita)等人，《自然》(Nature)448，260-262(2007)；余(Yu)J.等 人，《科学》(Science)318(5858):1917-1920(2007)；中川(Nakagawa) 等人，《自然生物技术》(Nat.Biotechnol.)26:101-106(2008)；以及扎瑞 斯(Zaehres)和肖勒(Scholer)，《细胞》131(5):834-835(2007)。

[0538] 可以通过分析(例如，定性或定量)一种或多种组织特异型基因的存在 来鉴定这种适合的细胞。例如，可以通过检测一种或多种组织特异型基因的 蛋白质产物来检测基因表达。蛋白质检测技术包括使用针对适当抗原的抗体 将蛋白质染色(例如，使用细胞提取物或全细胞)。在这种情况下，该适当 抗原是该组织特异型基因表达的蛋白质产物。虽然在原则上可以标记第一抗 体(即，结合该抗原的抗体)，更普遍的是(并且改进可视化)使用针对该 第一抗体的第二抗体(例如，一种抗IgG)。这种第二抗体与荧光染料、或 适当的用于比色反应的酶、或金珠(用于电子显微镜检查)、或者与生物素- 亲和素系统结合，使得该一级抗体进而该抗原的位置可以被识别。

[0539] 通过将药物组合物直接应用于外耳，本发明的CRISPR Cas分子可以递送 到耳，其中组合物从美国公开申请20110142917修改而来。在一些实施例 中，该药物组合物应用于耳道。递送至耳还可以称为听觉或耳递送。

[0540] 在一些实施例中，本发明的RNA分子在脂质体或阳离子脂质体配制品等 等中进行递送并且可以通过本领域技术人员熟知的方法进行制备。这样的方 法描述于，例如，美国专利号5,593,972、5,589,466、和5,580,859中，将其通 过引用并入本文。

[0541] 已经开发了专门针对增强和改进递送siRNA到哺乳动物细胞中的递送系 统，(参见，例如，沈(Shen)等人《欧洲生化学会联合会快报》(FEBS Let.)2003，539:111-114；夏(Xia)等人，《自然生物技术》(Nat. Biotech.)2002，20:1006-1010；赖希(Reich)等人，《分子视界》 (Mol.Vision.)2003，9:210-216；索伦森(Sorensen)等人，《分子生物学 杂志》(J.Mol.Biol.)2003，327:761-766；路易斯(Lewis)等人，《自然遗 传学》(Nat.Gen.)2002，32:107-108以及西梅奥尼(Simeoni)等人，NAR 2003，31，11:2717-2724)并且可以将其应用于本发明。最近，siRNA已经 成功用于抑制灵长动物中的基因表达(参见例如托伦蒂诺(Tolentino)等 人，《视网膜》(Retina)24(4):660，它也可应用于本发明。

[0542] 齐(Qi)等人披露了通过可应用于本发明的CRISPR Cas系统的新蛋白质 递送技术经由完整圆窗到内耳的有效siRNA转染方法(参见，例如，齐 (Qi)等人，《基因治疗》(Gene Therapy)(2013)，1-9)。具体地说， 可将Cy3标记的siRNA经由完整圆窗渗透转染到内耳的细胞中(包括内和外 毛细胞、壶腹嵴、椭圆囊斑和球囊斑)的TAT双链RNA结合结构域(TAT-

DRBD)成功用于体内递送双链siRNA,用于治疗各种内耳疾病和保留听觉 功能。可以考虑约40μl的10mM RNA作为施用至耳的剂量。

[0543] 根据瑞嘉利(Rejali)等人(《听力研究》(Hear Res.)2007年6月; 228(1-2):180-7),耳蜗植入物的功能借助于良好保留螺旋神经节神经元而得 以改善,这些神经元是经由植入物电刺激的靶标,并且先前已经表明脑源性 神经营养因子(BDNF)在实验性变聋的耳中增强了螺旋神经节的存活。瑞嘉 利(Rejali)等人测试了耳蜗植入物电极的改良设计,该电极包括被具有 BDNF基因插入片段的病毒载体转导的成纤维细胞的涂层。为了完成这种类 型的离体基因转移,瑞嘉利(Rejali)等人用具有BDNF基因盒插入片段的腺 病毒转导豚鼠,并且确定这些细胞分泌BDNF,进而将BDNF分泌细胞经由 琼脂糖凝胶附着在该耳蜗植入物电极上,并且将该电极植入鼓阶中。瑞嘉利 (Rejali)等人确定,该BDNF表达电极与对照电极相比在植入48天之后能 够保留显著更多的在耳蜗底回中的螺旋神经节神经元,并且证明了耳蜗植入 物治疗与用于增强螺旋神经节神经元存活的离体基因转移相结合的可行性。这样一种系统可以应用于递送到耳的本发明的CRISPR Cas系统。

[0544] 穆科吉(Mukherjea)等人(《抗氧化剂与氧化还原信号》(Antioxidants &Redox Signaling),第13卷,第5期,2010)证明,使用短干扰(si)RNA敲 低NOX3消除了顺铂的耳毒性,正如保护OHC免于损害以及在听觉脑干反应 (ABR)中降低的阈值位移所证明。将不同剂量的siNOX3(0.3、0.6、和0.9 μg)施用至大鼠并且通过实时RT-PCR评估NOX3表达。当与序列打乱的 (scrambled)siRNA的经鼓膜施用或未治疗的耳蜗比较时,使用的最低剂量 的NOX3 siRNA(0.3μg)未显示对NOX3 mRNA的任何抑制。然而,与序列 打乱的对照siRNA相比,给予更高剂量的NOX3 siRNA(0.6和0.9μg)降低 了NOX3表达。这样一种系统系统可应用于经鼓膜给药的本发明的CRISPR Cas系统,其中向人类给药的CRISPR Cas的剂量为约2mg到约4mg。

[0545] 荣格(Jung)等人(《分子治疗》(Molecular Therapy),第21卷,第 4期,834-841,2013年4月)证明,在椭圆囊中的Hes5水平在应用siRNA 之后降低,并且在这些椭圆囊中的毛细胞的数目显著大于对照处理之后的数 目。这些数据表明,siRNA技术可以用于诱导内耳中的修复和再生,并且 Notch信号通路对于特异性基因表达抑制是一个潜在有用的靶标。荣格 (Jung)等人将通过将无菌生理盐水加入冻干siRNA而制备的8μg的Hes5 siRNA以2μl体积注射到耳的前庭上皮。这样一种系统系统可应用于向耳的 前庭上皮给药的本发明的CRISPR Cas系统,其中向人类给药的CRISPR Cas 的剂量为约1到约30mg。

[0546] 眼

[0547] 本发明还考虑了向一只或两只眼耳递送该CRISPR-Cas系统。

[0548] 在本发明的又另一个方面,该CRISPR-Cas系统可以用来矫正几种基因 突变引起的眼部缺陷,其进一步描述于《眼的遗传疾病》(Genetic Diseases of the Eye),第二次编辑,由伊莱亚斯(Elias)I.特拉布勒西(Traboulsi)编 辑,牛津大学出版社,2012年。

[0549] 为了向眼部给药,慢病毒载体,尤其是马传染性贫血病毒(EIAV)是特 别优选的。

[0550] 在另一个实施例中,还考虑了基于马传染性贫血病毒(EIAV)的最小非 灵长类慢病毒载体,尤其是对于眼基因治疗而言(参见,例如,巴鲁谷安 (Balagaan),《基因医学杂志》(J Gene Med)2006;8:275-285,2005年 11月21日在《威利在线期刊》(Wiley InterScience) (www.interscience.wiley.com),DOI:10.1002/jgm.845)在线公开。考虑

这些 载体具有驱动靶基因表达的巨细胞病毒(CMV)启动子。前房内、视网膜 下、眼内和玻璃体内注射菌均予以考虑(参见,例如,巴鲁谷安 (Balagaan),《基因医学杂志》(J Gene Med)2006;8:275-285,2005年 11月21日在《威利在线期刊》(Wiley InterScience)(www.interscience.wiley.com),DOI:10.1002/jgm.845)在线公开。可以在手 术显微镜的辅助下进行眼内注射。为了视网膜下和玻璃体内注射,通过轻轻 指压可以使眼睛突出,并且使用接触镜系统看见眼底,该接触镜系统包含在 角膜上的一滴耦合介质溶液,角膜用玻璃显微镜载玻片盖玻片覆盖。为了视 网膜下注射,安装在5-μl的汉密尔顿注射器上的10-mm 34号针的尖端可以在 直接可视化之下穿过巩膜赤道部上部朝向后极切向行进,直到该针的孔在视 网膜下空间中可见时为止。然而,可以注射2μl的载体悬浮液以产生上部泡 状视网膜脱离,从而证实视网膜下载体给药。这种方法建立了一种自我愈合 的巩膜切开术,允许载体悬浮液保留在视网膜下空间中,直到它在该操作的 48小时之内被RPE吸收为止。可以在半球下方重复这个操作以产生下部视网 膜脱离。这种技术导致大约70%的感觉神经性视网膜和RPE暴露于该载体悬 浮液。为了玻璃体内注射,针尖可以在角巩膜缘后方1mm穿过巩膜并且将2 μl的载体悬浮液注射到玻璃体腔中。为了前房内注射,针尖可以通过角巩膜 缘穿刺朝向中央角膜行进,并且可以注射2μl的载体悬浮液。为了前房内注 射,针尖可以通过角巩膜缘穿刺朝向中央角膜行进,并且可以注射2μl的载 体悬浮液。可以1.0-1.4×10^{10}或1.0-1.4×10^{9}个转导单位(TU)/ml的滴度注 射这些载体。

[0551] 在另一个实施例中,还考虑了RetinoStat®,一种经由视网膜下注射用于 治疗湿型的年龄相关性黄斑变性的表达血管生成抑制蛋白(内皮抑素和血管 抑素)的基于马传染性贫血病毒的慢病毒基因治疗载体(参见,例如,宾利 (Binley)等人,《人类基因治疗》(HUMAN GENE THERAPY)23:980- 991(2012年9月))。这样一种载体可以改良为本发明的CRISPR-Cas系 统。每只眼可以用RetinoStat®以1.1x 10^{5}个转导单位/眼(TU/眼)以总体积100μl进行治疗。

[0552] 在另一个实施例中,对于递送至眼可以考虑E1-、部分E3-、E4-缺失的 腺病毒载体。对患有晚期新生血管性年龄相关性黄斑变性(AMD)的二十八 位患者给予表达人色素上皮源性因子(AdPEDF.11)的E1-、部分E3-、E4-缺 失的腺病毒载体的单次玻璃体内注射(参见,例如,坎波基亚罗 (Campochiaro)等人,《人类基因治疗》(Human Gene Therapy)17:167- 176(2006年2月))。研究了从10^{6}到$10^{9.5}$个粒子单位(PU)范围的剂 量,没有与AdPEDF.11有关的严重不良事件且没有剂量限制性毒性(参见, 例如,坎波基亚罗(Campochiaro)等人,《人类基因治疗》(Human Gene Therapy)17:167-176(2006年2月))。腺病毒载体介导的眼部基因转移显 得是一种可行的用于治疗眼部病症的方法,并且可应用于该CRISPR Cas系 统。

[0553] 在另一个实施例中,RXi制药公司(RXi Pharmaceuticals)的sd-rxRNA® 系统可以用于和/或适用于将CRISPR Cas递送至眼。在这个系统中,3μg的 sd-rxRNA的单次玻璃体内给药导致PPIB mRNA水平的序列特异性降低,持 续14天。该sd-rxRNA®系统可以应用于本发明的CRISPR Cas系统,考虑向 人类给药的剂量为约3到20mg的CRISPR。

[0554] 米林顿-华德(Millington-Ward)等人(《分子治疗》(Molecular Therapy),第19卷,第4期,642-649,2011年4月)描述了腺相关病毒 (AAV)载体,其用来递送基于RNA干扰

(RNAi)的视紫红质抑制剂和由 于在RNAi靶位点上的简并位置处的核苷酸改变而抵抗抑制的密码子修饰的 视紫红质代替基因。由米林顿-华德(Millington-Ward)等人将6.0x 10^8vp或 1.8x 10^{10}vp AAV的注射剂经视网膜下注射到眼中。米林顿-华德(Millington-Ward)等人的AAV载体可以应用于本发明的CRISPR Cas系统,考虑向人类 给药的剂量为约2x 10^{11}到约6x 10^{13}vp。

[0555] 戴尔卡拉(Dalkara)等人(《科学转化医学》(Sci Transl Med)5, 189ra76(2013))还涉及用于形成一种AAV载体的体内定向演化,该AAV 载体在将野生型缺陷型基因无损伤地注射到眼的玻璃体液中之后递送至整个 视网膜。戴尔卡拉(Dalkara)描述了7聚体肽展示文库和通过从AAV1、2、 4、5、6、8、和9的cap基因的DNA改组构建的AAV文库。将这些rcAAV文库和在CAG或Rho启动子之下表达GFP的rAAV载体进行包装,并且通 过定量PCR获得抗脱氧核糖核酸酶的基因组滴度。汇集这些文库,并且进行 两轮演化,每轮演化包括初始文库多样化继之以三个体内选择步骤。在每个 这样的步骤中,用2ml的碘克沙醇纯化的磷酸盐缓冲盐水(PBS)透析的文 库以约1×10^{12}vg/ml的基因组滴度对P30 rho-GFP小鼠进行玻璃体内注射。戴尔卡拉(Dalkara)等人的AAV载体可以应用于本发明的CRISPR Cas系 统,考虑向人类给药的剂量为约1x 10^{15}到约1x 10^{16}vg/ml。

[0556] 在另一个实施例中,该视紫红质基因可以被靶向用于治疗色素性视网膜 炎(RP),其中转让给桑加莫生物科学公司(Sangamo BioSciences,Inc.)的 美国专利公开号20120204282的系统可以依照本发明的CRISPR Cas系统进行 修饰。

[0557] 在另一个实施例中,针对从人类视紫红质基因切割靶序列的方法的转让 给Cellectis公司的美国专利公开号20130183282的方法也可以针对本发明的 CRISPR Cas系统进行修改。

[0558] 转让给台湾研究院(Academia Sinica)的美国专利公开号20130202678 涉及用于治疗视网膜病变和视力威胁性眼科疾病(sight-threatening ophthalmologic disorders)的方法,这些方法涉及向眼的视网膜下或玻璃体内 空间中递送Puf-A基因(在眼组织的视网膜神经节和色素细胞中表达并且展 示出独特的抗凋亡活性)。具体地说,令人希望的靶标是zgc:193933、prdm1a、spata2、tex10、rbb4、ddx3、zp2.2、Blimp-1和HtrA2,均可以由本 发明的CRISPR Cas系统靶向。

[0559] 吴(Wu)(《细胞—干细胞》(Cell Stem Cell),13:659-62,2013) 设计了一种指导RNA,其将Cas9导向到单个碱基突变,该突变在小鼠中引 起白内障,其中它诱导DNA切割。然后,在突变小鼠中,使用针对接合子修 复机制给予的另一种野生型等位基因或寡核苷酸(oligos)校正断裂的等位基 因的序列并且校正引起白内障的基因缺陷。

[0560] 美国专利公开号20120159653描述了使用锌指核酸酶基因修饰与黄斑变 性(MD)相关的细胞、动物以及蛋白质。黄斑变性(MD)是老年人视力损 害的主要原因,但也是儿童疾病如斯塔加特病(Stargardt disease)、索斯比 基底营养不良(Sorsby fundus)、和儿童致死性神经变性疾病(在如婴儿期一 般年幼的年龄就开始)的标志性症状。由于对视网膜的损害,黄斑变性导致 视野中心(黄斑)的视力丧失。当前存在的动物模型并未概括该疾病的主要 标志,因为它是在人类中观察到的。包含编码与MD相关的蛋白质的突变基 因的可得到的动物模型也产生高度可变的表型,使得针对人类疾病和治疗开 发的翻译是有问题的。

[0561] 美国专利公开号20120159653的一个方面包括编辑任何编码与MD相关 的蛋白质的染色体序列,这些序列可以应用于本发明的CRISPR Cas系统。典 型地,基于与MD相关的蛋白质与MD病症的实验性关联,选择与MD相关 的蛋白质。例如,相对于缺乏MD病症的群体,在具有MD病症的群体中, 与MD相关的蛋白质的产生率或循环浓度可以升高或降低。可以使用蛋白质 组技术评估蛋白质水平差异,包括但不限于蛋白质印迹、免疫组织化学染 色、酶联免疫吸附测定(ELISA)以及质谱法。可替代地,可以使用基因组 技术通过获得编码这些蛋白质的基因的基因表达谱而鉴定与MD相关的蛋白 质,包括但不限于DNA微阵列分析、基因表达系列分析(SAGE)以及定量 实时聚合酶链式反应(Q-PCR)。

[0562] 通过非限制性实例的方式,与MD相关的蛋白质包括但不限以下蛋白 质:(ABCA4)ATP结合盒,亚家族A(ABC1),成员4、ACHM1全色盲 (视杆单色色盲)1、ApoE,载脂蛋白E(ApoE)、C1QTNF5(CTRP5), C1q和肿瘤坏死因子相关蛋白5(C1QTNF5)、C2补体,补体2(C2)、C3 补体,补体(C3)、CCL2,趋化因子(C-C基序)配体2(CCL2)、 CCR2,趋化因子(C-C基序)受体2(CCR2)、CD36分化抗原簇36、 CFB,补体受体B、CFH,补体因子CFH H、CFHR1,补体因子H相关1、 CFHR3,补体因子H相关3、CNGB3环核苷酸门控通道β3、CP血浆铜蓝蛋 白(CP)、CRP,C反应蛋白(CRP)、CST3半胱氨酸蛋白酶抑制剂C或半 胱氨酸蛋白酶抑制剂3(CST3)、CTSD,组织蛋白酶D(CTSD)、CX3CR1,趋化因子(C-X3-C基序)受体1、ELOVL4,超长链脂肪酸延伸 4、ERCC6,切除修复交叉互补啮齿动物修复缺陷,互补群6、FBLN5,抗衰 老蛋白-5,FBLN5,抗衰老蛋白5、FBLN6,抗衰老蛋白6FSCN2聚束蛋白 (FSCN2)、HMCN1,半椎蛋白1,HMCN1,半椎蛋白1、HTRA1,HtrA 丝氨酸肽酶1(HTRA1),HTRA1、HtrA丝氨酸肽酶1、IL-6,白细胞介素 6、IL-8,白细胞介素8、LOC387715、假定蛋白、LEKHA1、含血小板白细 胞C激酶底物同源性域家族A成员1(PLEKHA1)、PROM1,普罗敏蛋白1 (PROM1或CD133)、PRPH2,外周蛋白-2RPGR色素性视网膜炎GTP酶 调节剂、SERPING1,丝氨酸蛋白酶抑制剂肽酶抑制剂,进化枝G,成员1 (C1-抑制剂)、TCOF1,糖蜜TIMP3金属蛋白酶抑制剂3(TIMP3)、 TLR3 Toll样受体3。

[0563] 与MD相关的其染色体序列被编辑的蛋白质的一致性可以并且应该变 化。在优选的实施例中,与其染色体序列被编辑的MD相关的蛋白质可以是 ATP结合盒、由ABCR基因编码的亚家族A(ABC1)成员4蛋白 (ABCA4)、由APOE基因编码的载脂蛋白E蛋白(APOE)、由CCL2基 因编码的趋化因子(C-C基序)配体2蛋白(CCL2)、由CCR2基因编码的 趋化因子(C-C基序)受体2蛋白(CCR2)、由CP基因编码的血浆铜蓝蛋 白蛋白(CP)、由CTSD基因编码的组织蛋白酶D蛋白(CTSD)、或由 TIMP3基因编码的金属蛋白酶抑制剂3蛋白(TIMP3)。在一个示例性实施 例中,该遗传修饰的动物是大鼠,并且与编码与MD相关的蛋白质的编辑的 染色体序列可以是:(ABCA4)ATP结合盒NM_000350亚家族A (ABC1),成员4、APOE,载脂蛋白ENM_138828(APOE)、CCL2,趋 化因子(C-C NM_031530基序)配体2(CCL2)、CCR2趋化因子(C-C NM_021866基序)受体2(CCR2)、CP血浆铜蓝蛋白(CP)NM_012532、 CTSD组织蛋白酶D(CTSD)NM_134334、TIMP3金属蛋白酶NM_012886 抑制剂3(TIMP3)。该动物或细胞可以包括1、2、3、4、5、6、7个或更多 个破坏的编码与MD相关的蛋白质的染色体序列以及零、1、2、3、4、5、 6、7个或更多个编码破坏的与MD相关的蛋白质的染色体整合序列。

[0564] 编辑的或整合的染色体序列可以被修饰为编码一种改变的与MD相关的 蛋白质。MD相关染色体序列中的几种突变已经与MD相关。在与MD相关 的染色体序列中的突变的非限制性实例包括可引起MD的那些,包括在 ABCR蛋白中的:E471K(即,在位置471的谷氨酸被

改变为赖氨酸)、R1129L(即,在位置1129处的精氨酸被改变为亮氨酸)、T1428M(即,在 位置1428处的苏氨酸被改变为甲硫氨酸)、R1517S(即,在位置1517处的 精氨酸被改变为丝氨酸)、I1562T(即,在位置1562处的异亮氨酸被改变为 苏氨酸)、以及G1578R(即,在位置1578处的甘氨酸被改变为精氨酸); 在CCR2蛋白中的:V64I(即,在位置192处的缬氨酸被改变为异亮氨 酸);在CP蛋白质中的:G969B(即,在位置969处的甘氨酸被改变为天冬 酰胺或天冬氨酸);在TIMP3蛋白中的:S156C(即,在位置156处的丝氨 酸被改变为半胱氨酸)、G166C(即,在位置166处的甘氨酸被改变为半胱 氨酸)、G167C(即,在位置167处的甘氨酸被改变为半胱氨酸)、Y168C (即,在位置168处的酪氨酸被改变为半胱氨酸)、S170C(即,在位置170处的丝氨酸被改变为半胱氨酸)、Y172C(即,在位置172处的酪氨酸被改 变为半胱氨酸)以及S181C(即,在位置181处的丝氨酸被改变为半胱氨 酸)。MD相关基因和疾病的遗传变体的其他关联在本领域是已知的。

[0565] 心脏

[0566] 本发明还考虑了向心脏递送该CRISPR-Cas系统。对于心脏,心肌热带 腺相关病毒(AAVM)优选的,尤其是在心脏中显示出优先基因转移的 AAVM41(参见,例如,林-扬加(Lin-Yanga)等人,PNAS,3月10日, 2009年,第106卷,第10期)。给药可以是全身性的或局部的。对于全身给 药可以考虑约1-10x 10^{14}个载体基因组的剂量。还参见,例如,埃拉里奥(Eulalio)等人(2012)《自然》(Nature)492:376和索马森达拉姆 (Somasuntharam)等人(2013)《生物材料》(Biomaterials)34:7790。

[0567] 例如,美国专利公开号20110023139描述了使用锌指核酸酶基因修饰与 心血管疾病相关的细胞、动物以及蛋白质。心血管疾病通常包括高血压、心 脏病发作、心力衰竭、以及中风和TIA。涉及心血管疾病的任何染色体序列 或由涉及心血管疾病的任何染色体序列编码的蛋白质都可以在本披露中描述 的方法中加以利用。典型地,基于心血管相关蛋白与心血管疾病的发展的实 验性关联,选择心血管相关蛋白。例如,相对于缺乏心血管障碍的群体,在 具有心血管障碍的群体中,心血管相关蛋白的产生率或循环浓度可以升高或 降低。可以使用蛋白质组技术评估蛋白质水平差异,包括但不限于蛋白质印 迹、免疫组织化学染色、酶联免疫吸附测定(ELISA)以及质谱法。可替代 地,可以使用基因组技术通过获得编码这些蛋白质的基因的基因表达谱而鉴 定心血管相关蛋白,包括但不限于DNA微阵列分析、基因表达系列分析 (SAGE)以及定量实时聚合酶链式反应(Q-PCR)。

[0568] 作为举例,该染色体序列可以包括但不限于,IL1B(白细胞介素1, β)、XDH(黄嘌呤脱氢酶)、TP53(肿瘤蛋白p53)、PTGIS(前列腺素12 (前列环素)合酶)、MB(肌红蛋白)、IL4(白细胞介素4)、ANGPT1 (血管生成素1)、ABCG8(ATP结合盒,亚家族G(白),成员8)、 CTSK(组织蛋白酶K)、PTGIR(前列腺素12(前列环素)受体(IP))、 KCNJ11(内向整流钾通道,亚家族J,成员11)、INS(胰岛素)、CRP(C 反应蛋白,正五聚蛋白相关的)、PDGFRB(血小板源生长因子受体,β多 肽)、CCNA2(细胞周期蛋白A2)、PDGFB(血小板源生长因子β多肽 (猴肉瘤病毒(v-sis)癌基因同源物))、KCNJ5(内向整流钾通道,亚家 族J,成员5)、KCNN3(钾中间小电导钙激活通道,亚家族N,成员3)、 CAPN10(卡配因10)、PTGES(前列腺素E合酶)、ADRA2B(肾上腺素 能,α-2B-,受体)、ABCG5(ATP结合盒,亚家族G(WHITE)、成员 5)、PRDX2(过氧化物氧化还原酶2)、CAPN5(卡配因5)、PARP14 (聚(ADP-核糖)聚合酶家族,成员14)、MEX3C(mex-3同源物C(秀丽隐 杆线虫))、ACE血管紧张素I转化酶(肽基二肽酶A)1)、TNF(肿瘤坏 死因子

(TNF超家族,成员2))、IL6(白细胞介素6(干扰素,β2))、STN(抑制素)、SERPINE1(丝氨酸蛋白酶抑制蛋白肽酶抑制剂,进化枝E (微管连接蛋白,纤溶酶原激活物抑制剂1型)、成员1)、ALB(白蛋 白)、ADIPOQ(脂联素,含C1Q和胶原蛋白域)、APOB(载脂蛋白B(包 括Ag(x)抗原))、APOE(载脂蛋白E)、LEP(瘦素)、MTHFR(5,10-亚 甲基四氢叶酸还原酶(NADPH))、APOA1(载脂蛋白A-I)、EDN1(内 皮素1)、NPPB(利钠肽前体B)、NOS3(一氧化氮合酶3(内皮细胞))、PPARG(过氧化物酶体增殖物活化受体γ)、PLAT(纤溶酶原激活 物,组织)、PTGS2(前列腺素内过氧化物合酶2(前列腺素G/H合酶和环 加氧酶))、CETP(胆固醇酯转移蛋白,血浆)、AGTR1(血管紧张素II受 体,1型)、HMGCR(3-羟基-3-甲基戊二酸单酰辅酶A还原酶)、IGF1(胰 岛素样生长因子1(生长素调节素C))、SELE(选择素E)、REN(肾 素)、PPARA(过氧化物酶体增殖物活化受体α)、PON1(对氧磷酶1)、KNG1(激肽原1)、CCL2(趋化因子(C-C基序)配体2)、LPL(脂蛋白 脂酶)、VWF(冯·维勒布兰德因子)、F2(凝血因子II(凝血酶))、ICAM1(细胞间粘附分子1)、TGFB1(转化生长因子,β1)、NPPA(利钠 肽前体A)、IL10(白细胞介素10)、EPO(促红细胞生成素)、SOD1(超 氧化物歧化酶1,可溶性)、VCAM1(血管细胞粘附分子1)、IFNG(干扰 素,γ)、LPA(脂蛋白,Lp(a))、MPO(髓过氧化物酶)、ESR1(雌激素 受体1)、MAPK1(丝裂原活化蛋白激酶1)、HP(触珠蛋白)、F3(凝血 因子III(促凝血酶原激酶,组织因子))、CST3(半胱氨酸蛋白酶抑制剂 C)、COG2(寡聚高尔基复合体成分2)、MMP9(基质金属肽酶9(明胶酶 B,92kDa明胶酶,92kDa I型V胶原酶))、SERPINC1(丝氨酸蛋白酶抑 制蛋白肽酶抑制剂,进化枝C(抗凝血酶)、成员1)、F8(凝血因子VIII, 促凝血成分)、HMOX1(血红素加氧酶(解环)1)、APOC3(载脂蛋白C- III)、IL8(白细胞介素8)、PROK1(前动力蛋白1)、CBS(胱硫醚-β-合 酶)、NOS2(一氧化氮合酶2,诱导型)、TLR4(toll样受体4)、SELP (选择素P(颗粒膜蛋白140kDa,抗原CD62))、ABCA1(ATP结合盒, 亚家族A(ABC1)、成员1)、AGT(血管紧张素原(丝氨酸蛋白酶抑制蛋 白肽酶抑制剂,进化枝A,成员8))、LDLR(低密度脂蛋白受体)、GPT(谷丙转氨酶(丙氨酸转氨酶))、VEGFA(血管内皮生长因子A)、NR3C2(细胞核受体亚家族3,型C,成员2)、IL18(白细胞介素18(干扰 素-γ-诱导因子))、NOS1(一氧化氮合酶1(神经元的))、NR3C1(细胞 核受体亚家族3,C组,成员1(糖皮质激素受体))、FGB(纤维蛋白原β 链)、HGF(肝细胞生长因子(肝细胞生长因子A;分散因子))、IL1A (白细胞介素1,α)、RETN(抵抗素)、AKT1(v-akt鼠科胸腺瘤病毒癌基 因同源物1)、LIPC(脂肪酶,肝脏的)、HSPD1(热休克60kDa蛋白1 (伴侣蛋白))、MAPK14(丝裂原活化蛋白激酶14)、SPP1(分泌磷蛋白 1)、ITGB3(整合素,β3(血小板糖蛋白111a,抗原CD61))、CAT(过 氧化氢酶)、UTS2(尾加压素2)、THBD(血栓调节蛋白)、F10(凝血因 子X)、CP(血浆铜蓝蛋白(亚铁氧化酶))、TNFRSF11B(肿瘤坏死因子 受体超家族,成员11b)、EDNRA(内皮素A型受体)、EGFR(表皮生长 因子受体(红白血病病毒(v-erb-b)癌基因同源物,鸟类的))、MMP2(基 质金属肽酶2(明胶酶A,72kDa明胶酶,72kDa I型V胶原酶))、PLG (纤维蛋白溶酶原)、NPY(神经肽Y)、RHOD(ras同源物基因家族,成 员D)、MAPK8(丝裂原活化蛋白激酶8)、MYC(V-Myc骨髓细胞瘤病毒 癌基因同源物(鸟类的))、FN1(纤连蛋白1)、CMA1(糜酶1,肥大细 胞)、PLAU(纤溶酶原激活物,尿激酶)、GNB3(鸟嘌呤核苷酸结合蛋白 (G蛋白)、β多肽3)、ADRB2(肾上腺素能,β-2-,受体,表面)、APOA5(载脂蛋白A-V)、SOD2(超氧化物歧化酶2,线粒体的)、F5(凝 血因子V(促凝血球蛋白原,不稳定因子))、VDR(维生素D(1,25-二羟 维生素D3)受体)、ALOX5(花生四烯酸盐5-脂氧合酶)、HLA-DRB1(主 要组织相容性复合物,I类I,DRβ1)、PARP1(聚(ADP-核糖)聚合酶1)、CD40LG

(CD40配体)、PON2(对氧磷酶2)、AGER(晚期糖基化终末产 物特异性受体)、IRS1(胰岛素受体底物1)、PTGS1(前列腺素内过氧化物 合酶1(前列腺素G/H合酶和环加氧酶))、ECE1(内皮素转化酶1)、F7 (凝血因子VII(血清凝血酶原转变加速因子))、URN(白细胞介素1受体 拮抗剂)、EPHX2(环氧化物水解酶2,细胞质的)、IGFBP1(胰岛素样生 长因子结合蛋白1)、MAPK10(丝裂原活化蛋白激酶10)、FAS(Fas (TNF受体超家族,成员6))、ABCB1(ATP结合盒,亚家族B (MDR/TAP),成员1)、JUN(jun癌基因)、IGFBP3(胰岛素样生长因 子结合蛋白3)、CD14(CD14分子)、PDE5A(磷酸二酯酶5A,cGMP特 异性)、AGTR2(血管紧张素II受体,2型)、CD40(CD40分子,TNF受 体超家族成员5)、LCAT(卵磷脂胆甾醇酰基转移酶)、CCR5(趋化因子(C-C基序)受体5)、MMP1(基质金属肽酶1(间质胶原酶))、TIMP1 (TIMP金属肽酶抑制剂1)、ADM(肾上腺髓质素)、DYT10(肌张力障碍 10)、STAT3(信号传导和转录激活因子3(急性期反应因子))、MMP3 (基质金属肽酶3(基质溶解素1,前白明胶酶))、ELN(弹性蛋白)、USF1(上游转录因子1)、CFH(补体因子H)、HSPA4(热休克70kDa蛋 白4)、MMP12(基质金属肽酶12(巨噬细胞弹性蛋白酶))、MME(膜金 属肽链内切酶)、F2R(凝血因子II(凝血酶)受体)、SELL(选择素 L)、CTSB(组织蛋白酶B)、ANXA5(膜联蛋白A5)、ADRB1(肾上腺 素能,β-1-,受体)、CYBA(细胞色素b-245,α多肽)、FGA(纤维蛋白原 α链)、GGT1(γ-谷氨酰转肽酶1)、LIPG(脂肪酶,内皮的)、HIF1A (缺氧诱导因子1,α亚基(碱性-螺旋-环-螺旋转录因子))、CXCR4(趋化 因子(C-X-C基序)受体4)、PROC(蛋白C(凝血因子Va和VIIIa抑制因 子)、SCARB1(清道夫受体B类,成员1)、CD79A(CD79a分子,免疫球 蛋白相关α)、PLTP(磷脂转移蛋白)、ADD1(内收蛋白1(α))、FGG (纤维蛋白原γ链)、SAA1(血清淀粉样蛋白A1)、KCNH2(电压门控钾 离子通道,亚家族H(触角电位相关)、成员2)、DPP4(二肽基肽酶 4)、G6PD(6-磷酸葡萄糖脱氢酶)、NPR1(钠尿肽受体A/鸟苷酸环化酶A (心房钠尿肽受体A))、VTN(玻连蛋白)、KIAA0101(KIAA0101)、FOS(FBJ鼠科骨肉瘤病毒癌基因同源物)、TLR2(toll样受体2)、PPIG (肽基脯氨酰异构酶G(亲环素G))、IL1R1(白细胞介素1受体,I 型)、AR(雄激素受体)、CYP1A1(细胞色素P450,家族1,亚家族A, 多肽1)、SERPINA1(丝氨酸蛋白酶抑制蛋白肽酶抑制剂,进化枝A(α-1 抗蛋白酶,抗胰蛋白酶),成员1)、MTR(5-甲基四氢叶酸高半胱胺酸甲基 转移酶)、RBP4(视黄醇结合蛋白4,血浆)、APOA4(载脂蛋白A-IV)、CDKN2A(细胞周期蛋白依赖性激酶抑制剂2A(黑素瘤p16,抑制 CDK4))、FGF2(成纤维细胞生长因子2(碱性))、EDNRB(内皮素受体B型)、ITGA2(整合素,α2(CD49B,VLA-2受体α2亚基))、CABIN1(钙调神经磷酸酶结合蛋白1)、SHBG(性激素结合球蛋白)、HMGB1(高迁移率族1)、HSP90B2P(热休克蛋白90kDaβ(Grp94),成 员2(假基因))、CYP3A4(细胞色素P450,家族3,亚家族A,多肽 4)、GJA1(间隙连接蛋白,α1,43kDa)、CAV1(小窝蛋白1,细胞质膜 微囊蛋白,22kDa)、ESR2(雌激素受体2(ERβ))、LTA(淋巴毒素α (TNF超家族,成员1))、GDF15(生长分化因子15)、BDNF(脑源性神 经营养因子)、CYP2D6(细胞色素P450,家族2,亚家族D,多肽6)、NGF(神经生长因子(β多肽))、SP1(Sp1转录因子)、TGIF1(TGFB-诱 导因子同源框1)、SRC(v-src肉瘤(施密特-鲁平(Schmidt-Ruppin)A-2)病 毒癌基因同源物(鸟类的))、EGF(表皮生长因子(β-抑胃素))、PIK3CG(磷酸肌醇-3-激酶,催化的,γ多肽)、HLA-A(主要组织相容性复 合物,I类,A)、KCNQ1(电压门控钾通道,KQT样亚家族,成员1)、CNR1(大麻素受体1(脑))、FBN1(微纤维蛋白1)、CHKA(胆碱激酶 α)、BEST1(卵黄状黄斑病蛋白1)、APP(淀粉样蛋白β(A4)前体蛋 白)、CTNNB1(连环蛋白(钙粘着蛋白关联蛋白),β1,88kDa)、IL2 (白细胞介素2)、CD36(CD36分子

(凝血酶敏感蛋白受体))、PRKAB1 (蛋白激酶,AMP活化的,β1非催化亚基)、TPO(甲状腺过氧化物酶)、ALDH7A1(醛脱氢酶7家族,成员A1)、CX3CR1(趋化因子(C-X3-C基 序)受体1)、TH(酪氨酸羟化酶)、F9(凝血因子IX)、GH1(生长激素 1)、TF(转铁蛋白)、HFE(血色素沉着病)、IL17A(白细胞介素 17A)、PTEN(磷酸酯酶与张力蛋白同源物)、GSTM1(谷胱甘肽S-转移酶 μ1)、DMD(肌营养不良蛋白)、GATA4(GATA结合蛋白4)、F13A1 (凝血因子XIII,A1多肽)、TTR(转甲状腺素蛋白)、FABP4(脂肪酸结 合蛋白4,脂肪细胞)、PON3(对氧磷酶3)、APOC1(载脂蛋白C-I)、 INSR(胰岛素受体)、TNFRSF1B(肿瘤坏死因子受体超家族,成员1B)、 HTR2A(5-羟色胺(血清素)受体2A)、CSF3(集落刺激因子3(粒细 胞))、CYP2C9(细胞色素P450,家族2,亚家族C,多肽9)、TXN(硫 氧还蛋白)、CYP11B2(细胞色素P450,家族11,亚家族B,多肽2)、PTH(甲状旁腺素、CSF2(集落刺激因子2(粒细胞-巨噬细胞))、KDR (激酶插入结构域受体受体(a I型II受体酪氨酸激酶))、PLA2G2A(磷脂 酶A2,型IIA(血小板,滑液))、B2M(β-2-微球蛋白)、THBS1(凝血 酶敏感蛋白1)、GCG(胰高血糖素)、RHOA(ras同源物基因家族,成员 A)、ALDH2(醛脱氢酶2家族(线粒体的))、TCF7L2(转录因子7样2 (T细胞特异性HMG盒))、BDKRB2(缓激肽受体B2)、NFE2L2(红细 胞衍生核因子2样蛋白)、NOTCH1(Notch同源物1,易位相关的(果 蝇))、UGT1A1(UDP葡糖醛酸基转移酶1家族,多肽A1)、IFNA1(干 扰素,α1)、PPARD(过氧化物酶体增殖物活化受体δ)、SIRT1(长寿蛋白 (沉默交配I型信息调控2同源物)1(酿酒酵母))、GNRH1(促性腺素释 放激素1(黄体生成素释放激素))、PAPPA(妊娠相关血浆蛋白A,冠毛 素1)、ARR3(抑制蛋白3,视网膜的(X-抑制蛋白))、NPPC(利钠肽前 体C)、AHSP(α血红蛋白稳定蛋白)、PTK2(PTK2蛋白酪氨酸激酶 2)、IL13(白细胞介素13)、MTOR(雷帕霉素机械靶(丝氨酸/苏氨酸激 酶))、ITGB2(整合素,β2(补体成分3受体3和4亚基))、GSTT1(谷 胱甘肽S-转移酶θ1)、IL6ST(白细胞介素6信号传导因子(gp130,抑瘤素 M受体))、CPB2(羧肽酶B2(血浆))、CYP1A2(细胞色素P450,家 族1,亚家族A,多肽2)、HNF4A(肝细胞核因子4,α)、SLC6A4(溶质 载体家族6(神经递质转运蛋白,血清素),成员4)、PLA2G6(磷脂酶 A2,型VI(细胞溶质的,钙依赖性))、TNFSF11(肿瘤坏死因子(配体) 超家族,成员11)、SLC8A1(溶质载体家族8(钠/钙交换蛋白),成员 1)、F2RL1(凝血因子II(凝血酶)受体样1)、AKR1A1(醛酮还原酶家 族1,成员A1(醛还原酶))、ALDH9A1(醛脱氢酶9家族,成员A1)、 BGLAP(骨γ-羧谷氨酸(gla)蛋白)、MTTP(微粒体甘油三酯转移蛋 白)、MTRR(5-甲基四氢叶酸-高半胱氨酸甲基转移酶还原酶)、SULT1A3 (磺基转移酶家族,细胞溶质的,1A,酚优选,成员3)、RAGE(肾肿瘤抗 原)、C4B(补体成分4B(奇都血型)、P2RY12(嘌呤能受体P2Y,G-蛋白 偶联的,12)、RNLS(肾酶,FAD依赖性胺氧化酶)、CREB1(cAMP应答 元件结合蛋白1)、POMC(阿黑皮素原)、RAC1(ras相关C3肉毒毒素底 物1(rho家族,小GTP结合蛋白Rac1))、LMNA(核纤层蛋白NC)、 CD59(CD59分子,补体调节蛋白)、SCN5A(钠通道,电压门控,V型,α 亚基)、CYP1B1(细胞色素P450,家族1,亚家族B,多肽1)、MIF(巨 噬细胞游走抑制因子(糖基化抑制因子))、MMP13(基质金属肽酶13(胶 原酶3))、TIMP2(TIMP金属肽酶抑制剂2)、CYP19A1(细胞色素 P450,家族19,亚家族A,多肽1)、CYP21A2(细胞色素P450,家族21, 亚家族A,多肽2)、PTPN22(蛋白酪氨酸磷酸酶,非受体型22(淋巴 样))、MYH14(肌球蛋白,重链14,非肌肉)、MBL2(甘露糖结合凝集 素(蛋白C)2,可溶性(调理素缺陷))、SELPLG(选择素P配体)、 AOC3(胺氧化酶,含铜3(血管粘附蛋白1))、CTSL1(组织蛋白酶 L1)、PCNA(增殖细胞核抗原)、IGF2(胰岛素样生长因子2(生长素调节 素A))、ITGB1(整合素,β1(纤连蛋白受体,β多肽,抗原CD29包括 MDF2,

MSK12))、CAST(钙蛋白酶抑制蛋白)、CXCL12(趋化因子 (C-X-C基序)配体12(基质细胞衍生因子1))、IGHE(免疫球蛋白恒定 区ε)、KCNE1(电压门控钾通道,Isk相关家族,成员1)、TFRC(转铁蛋 白受体(p90,CD71))、COL1A1(胶原,I型,α1)、COL1A2(胶原,I 型,α2)、IL2RB(白细胞介素2受体,β)、PLA2G10(磷脂酶A2,型 X)、ANGPT2(血管生成素2)、PROCR(蛋白C受体,内皮的 (EPCR))、NOX4(NADPH氧化酶4)、HAMP(海帕西啶抗微生物 肽)、PTPN11(蛋白酪氨酸磷酸酶,非受体1型1)、SLC2A1(溶质载体家 族2(易化葡萄糖转运蛋白),成员1)、IL2RA(白细胞介素2受体, α)、CCL5(趋化因子(C-C基序)配体5)、IRF1(干扰素调节因子1)、 CFLAR(CASP8和FADD样凋亡调节因子)、CALCA(降钙素相关多肽 α)、EIF4E(真核翻译起始因子4E)、GSTP1(谷胱甘肽S-转移酶pi 1)、JAK2(Janus激酶2)、CYP3A5(细胞色素P450,家族3,亚家族A,多肽 5)、HSPG2(类肝素硫酸蛋白聚糖2)、CCL3(趋化因子(C-C基序)配体 3)、MYD88(髓性分化原发反应基因(88))、VIP(血管活性肠肽)、 SOAT1(甾醇O-酰基转移酶1)、ADRBK1(肾上腺素能,β,受体激酶 1)、NR4A2(细胞核受体亚家族4,型A,成员2)、MMP8(基质金属肽酶8(中性白细胞胶原酶))、NPR2(钠尿肽受体B/鸟苷酸环化酶B(心房 钠尿肽受体B))、GCH1(GTP环化水解酶1)、EPRS(谷氨酰-脯氨酰- tRNA合成酶)、PPARGC1A(过氧化物酶体增殖物活化受体γ,共激活剂1 α)、F12(凝血因子XII(哈格曼因子))、PECAM1(血小板/内皮细胞粘附分子)、CCL4(趋化因子(C-C基序)配体4)、SERPINA3(丝氨酸蛋白 酶抑制蛋白肽酶抑制剂,进化枝A(α-1抗蛋白酶,抗胰蛋白酶),成员 3)、CASR(钙传感受体)、GJA5(间隙连接蛋白,α5,40kDa)、FABP2 (脂肪酸结合蛋白2,肠)、TTF2(转录终止因子,RNA聚合酶II)、 PROS1(蛋白S(α))、CTF1(心脏营养素1)、SGCB(肌聚糖,β(43 kDa肌营养不良蛋白相关糖蛋白))、YME1L1(YME1样1(酿酒酵 母))、CAMP(卡色力西丁抗微生物肽)、ZC3H12A(含锌指CCCH型12A)、AKR1B1(醛酮还原酶家族1,成员B1(醛糖还原酶))、DES(结 蛋白)、MMP7(基质金属肽酶7(基质溶解因子,子宫的))、AHR(芳香 烃受体)、CSF1(集落刺激因子1(巨噬细胞))、HDAC9(组蛋白去乙酰 化酶9)、CTGF(结缔组织生长因子)、KCNMA1(大电导钙激活钾通道,亚家族M,α成员1)、UGT1A(UDP葡糖醛酸基转移酶1家族,多肽A复 合体座位)、PRKCA(蛋白激酶C,α)、COMT(儿茶酚-β-甲基转移 酶)、S100B(S100 calcium结合蛋白B)、EGR1(早期生长反应蛋白1)、 PRL(催乳素)、IL15(白细胞介素15)、DRD4(多巴胺受体D4)、 CAMK2G(钙-钙调蛋白依赖性蛋白激酶IIγ)、SLC22A2(溶质载体家族22 (有机阳离子转运蛋白),成员2)、CCL11(趋化因子(C-C基序)配体 11)、PGF(B321胎盘生长因子)、THPO(血小板生成素)、GP6(糖蛋白 VI(血小板))、TACR1(速激肽受体1)、NTS(神经降压肽)、HNF1A (HNF1同源框A)、SST(生长抑素)、KCND1(电压门控钾通道,Shal相 关亚家族,成员1)、LOC646627(磷脂酶抑制剂)、TBXAS1(血栓烷A合 酶1(血小板))、CYP2J2(细胞色素P450,家族2,亚家族J,多肽2)、TBXA2R(血栓烷A2受体)、ADH1C(醇脱氢酶1C(I类),γ多肽)、 ALOX12(花生四烯酸盐12-脂氧合酶)、AHSG(α-2-HS-糖蛋白)、BHMT (甜菜碱同型半胱氨酸甲基转移酶)、GJA4(间隙连接蛋白,α4,37kDa)、 SLC25A4(溶质载体家族25(线粒体载体;腺嘌呤核苷酸转运蛋白),成员4)、ACLY(ATP柠檬酸裂合酶)、ALOX5AP(花生四烯酸盐5-脂氧合酶- 活化蛋白)、NUMA1(核有丝分裂器蛋白1)、CYP27B1(细胞色素P450, 家族27,亚家族B,多肽1)、CYSLTR2(半胱氨酰白三烯受体2)、SOD3 (超氧化物歧化酶3,细胞外的)、LTC4S(白三烯C4合酶)、UCN(尿皮 质素)、GHRL(胃促生长素/肥胖抑制素前体肽)、APOC2(载脂蛋白C- II)、CLEC4A(C型凝集素结构域家族4,成员A)、KBTBD10(Kelch重 复和BTB(POZ)域包含蛋白)、TNC(腱生蛋白C)、TYMS

(胸苷酸合成 酶)、SHC1(SHC(含Src同源物2域)转化蛋白1)、LRP1(低密度脂蛋 白受体相关蛋白1)、SOCS3(细胞因子信号传导抑制因子3)、ADH1B (醇脱氢酶1B(I类),β多肽)、KLK3(激肽释放酶相关肽酶3)、 HSD11B1(羟基固醇(11-β)脱氢酶1)、VKORC1(生素K环氧化物还原酶复合体,亚基1)、SERPINB2(丝氨酸蛋白酶抑制蛋白肽酶抑制剂,进化 枝B(卵清蛋白),成员2)、TNS1(张力蛋白1)、RNF19A(环指蛋白 9A)、EPOR(促红细胞生成素受体)、ITGAM(整合素,αM(补体成分3 受体3亚基))、PITX2(配对样同源域2)、MAPK7(丝裂原活化蛋白激 酶7)、FCGR3A(IgG的Fc片段,低亲和力111a,受体(CD16a))、 LEPR(瘦素受体)、ENG(内皮糖蛋白)、GPX1(谷胱甘肽过氧化酶1)、 GOT2(谷草转氨酶2,线粒体(天冬氨酸氨基转移酶2))、HRH1(组胺受 体H1)、NR112(细胞核受体亚家族1,型I,成员2)、CRH(促肾上腺皮 质素释放激素)、HTR1A(5-羟色胺(血清素)受体1A)、VDAC1(电压 依赖性阴离子通道1)、HPSE(类肝素酶)、SFTPD(表面活性蛋白D)、 TAP2(转运蛋白2,ATP结合盒,亚家族B(MDR/TAP))、RNF123(环 指蛋白123)、PTK2B(PTK2B蛋白酪氨酸激酶2β)、NTRK2(神经营养酪 氨酸激酶,受体,2型)、IL6R(白细胞介素6受体)、ACHE(乙酰胆碱酯 酶(Yt血型))、GLP1R(胰高血糖素样肽1受体)、GHR(生长激素受 体)、GSR(谷胱甘肽还原酶)、NQO1(NAD(P)H脱氢酶,醌1)、NR5A1 (细胞核受体亚家族5,型A,成员1)、GJB2(间隙连接蛋白,β2,26 kDa)、SLC9A1(溶质载体家族9(钠/氢交换体)、成员1)、MAOA(单 胺氧化酶A)、PCSK9(前蛋白转化酶枯草杆菌蛋白酶/kexin 9型)、 FCGR2A(IgG的Fc片段,低亲和力IIa,受体(CD32))、SERPINF1(丝 氨酸蛋白酶抑制蛋白肽酶抑制剂,进化枝F(α-2抗纤维蛋白溶酶,色素上皮 衍生因子),成员1)、EDN3(内皮素3)、DHFR(二氢叶酸还原酶)、 GAS6(生长停滞特异蛋白6)、SMPD1(鞘磷脂磷酸二酯酶1,酸溶酶 体)、UCP2(解偶联蛋白2(线粒体的,质子载体))、TFAP2A(转录因 子AP-2α(激活增强子结合蛋白2α))、C4BPA(补体成分4结合蛋白, α)、SERPINF2(丝氨酸蛋白酶抑制蛋白肽酶抑制剂,进化枝F(α-2抗纤维 蛋白溶酶,色素上皮衍生因子),成员2)、TYMP(胸苛酸磷酸化酶)、ALPP(碱性磷酸酶,胎盘的(Regan同工酶))、CXCR2(趋化因子(C-X- C基序)受体2)、SLC39A3(溶质载体家族39(锌转运蛋白)、成员3)、 ABCG2(ATP结合盒,亚家族G(WHITE)、成员2)、ADA(腺苷脱氨 酶)、JAK3(Janus激酶3)、HSPA1A(热休克70kDa蛋白1A)、FASN (脂肪酸合酶)、FGF1(成纤维细胞生长因子1(酸性))、F11(凝血因子 XI)、ATP7A(ATP酶,Cu++转运的,α多肽)、CR1(补体成分(3b/4b) 受体1(Knops血型))、GFAP(胶质细胞原纤维酸性蛋白)、ROCK1 (Rho相关,含卷曲螺旋蛋白激酶1)、MECP2(甲基CpG结合蛋白2(雷 特综合征))、MYLK(肌球蛋白轻链激酶)、BCHE(丁酰胆碱酯酶)、 LIPE(脂肪酶,激素敏感的)、PRDX5(过氧化物氧化还原酶5)、 ADORA1(腺苷A1受体)、WRN(维尔纳综合征,RecQ解旋酶样)、 CXCR3(趋化因子(C-X-C基序)受体3)、CD81(CD81分子)、SMAD7 (SMAD家族成员7)、LAMC2(层粘连蛋白,γ2)、MAP3K5(丝裂原活 化蛋白激酶激酶激酶5)、CHGA(嗜铬粒蛋白A(甲状旁腺分泌蛋白1))、IAPP(胰岛淀粉样蛋白多肽)、RHO(视紫红质)、ENPP1(外核苷 酸焦磷酸酶/磷酸二酯酶1)、PTHLH(甲状旁腺激素样激素)、NRG1(神 经调节蛋白1)、VEGFC(血管内皮生长因子C)、ENPEP(谷氨酰基氨肽 酶(氨基肽酶A))、CEBPB(CCAAT/增强子结合蛋白(C/EBP),β)、 NAGLU(N-乙酰氨基葡糖苷酶,α-)、F2RL3(凝血因子II(凝血酶)受体 样3)、CX3CL1(趋化因子(C-X3-C基序)配体1)、BDKRB1(缓激肽受 体B1)、ADAMTS13(具有凝血酶敏感蛋白1型基序的ADAM金属肽酶, 13)、ELANE(弹性蛋白酶,嗜中性粒细胞表达的)、ENPP2(外核苷酸焦 磷酸酶/磷酸二酯酶2)、CISH(细胞因子诱导的含SH2的蛋白)、GAST (胃泌素)、MYOC(肌纤蛋白,

小梁网可诱导糖皮质激素应答)、ATP1A2 (ATP酶,Na+/K+转运的,α2多肽)、NF1(神经纤维瘤蛋白1)、GJB1 (间隙连接蛋白,β1,32kDa)、MEF2A(肌细胞增强因子2A)、VCL(纽 蛋白)、BMPR2(骨形态发生蛋白受体,II型(丝氨酸/苏氨酸激酶))、TUBB(微管蛋白,β)、CDC42(细胞分裂周期42(GTP结合蛋白,25 kDa))、KRT18(角蛋白18)、HSF1(热休克转录因子1)、MYB(v- myb成髓细胞瘤病毒癌基因同源物(鸟类))、PRKAA2(蛋白激酶,AMP 活化的,α2催化亚基)、ROCK2(Rho关联含卷曲螺旋蛋白激酶2)、TFPI (组织因子途径抑制物(脂蛋白相关凝血抑制剂))、PRKG1(蛋白激酶, cGMP依赖性,I型)、BMP2(骨形态发生蛋白2)、CTNND1(连环蛋白 (钙粘着蛋白关联蛋白),δ1)、CTH(胱硫醚酶(胱硫醚γ-裂解酶))、CTSS(组织蛋白酶S)、VAV2(vav 2鸟苷酸交换因子)、NPY2R(神经肽 Y受体Y2)、IGFBP2(胰岛素样生长因子结合蛋白2,36kDa)、CD28 (CD28分子)、GSTA1(谷胱甘肽S-转移酶α1)、PPIA(肽基脯氨酰异构酶A(亲环素A))、APOH(载脂蛋白H(β-2-糖蛋白I))、S100A8 (S100钙结合蛋白A8)、IL11(白细胞介素11)、ALOX15(花生四烯酸盐 15-脂氧合酶)、FBLN1(腓骨蛋白1)、NR1H3(细胞核受体亚家族1,型 H,成员3)、SCD(硬脂酰基-辅酶A去饱和酶(Δ-9-去饱和酶))、GIP (抑胃多肽)、CHGB(嗜铬粒蛋白B(分泌粒蛋白1))、PRKCB(蛋白激 酶C,β)、SRD5A1(类固醇-5-α还原酶α多肽1(3-氧代-5α-类固醇δ4-脱 氢酶α1))、HSD11B2(羟基固醇(11-β)脱氢酶2)、CALCRL(降钙素受体 样)、GALNT(UDP-N-乙酰基-α-D-半乳糖胺:多肽N-乙酰半乳糖胺基转移 酶2(GalNAc-T2))、ANGPTL4(血管生成素样4)、KCNN4(钾中间/小 电导钙激活通道,亚家族N,成员4)、PIK3C2A(磷酸肌醇-3-激酶,2类, α多肽)、HBEGF(肝素结合EGF样生长因子)、CYP7A1(细胞色素 P450,家族7,亚家族A,多肽1)、HLA-DRB5(主要组织相容性复合物, I类I,DRβ5)、BNIP3(BCL2/腺病毒E1B 19kDa相互作用蛋白3)、GCKR(葡糖激酶(己糖激酶4)调节蛋白)、S100A12(S100钙结合蛋白 A12)、PADI4(肽基精氨酸脱亚氨酶,I型V)、HSPA14(热休克70kDa 蛋白14)、CXCR1(趋化因子(C-X-C基序)受体1)、H19(H19,母系印 记表达转录物(非蛋白质编码))、KRTAP19-3(角蛋白关联蛋白19-3)、IDDM2(胰岛素依赖型糖尿病2)、RAC2(ras相关C3肉毒毒素底物2(rho 家族,小GTP结合蛋白Rac2))、RYR1(兰尼碱受体1(骨骼))、CLOCK(clock同源物(小鼠))、NGFR(神经生长因子受体(TNFR超家 族,成员16))、DBH(多巴胺β-羟化酶(多巴胺β-单加氧酶))、CHRNA4(胆碱能受体,烟碱的,α4)、CACNA1C(钙通道,电压依赖性, L型,α1C亚基)、PRKAG2(蛋白激酶,AMP激活的,γ2非催化亚基)、CHAT(胆碱乙酰转移酶)、PTGDS(前列腺素D2合酶21kDa(脑))、NR1H2(细胞核受体亚家族1,型H,成员2)、TEK(TEK酪氨酸激酶,内 皮的)、VEGFB(血管内皮生长因子B)、MEF2C(肌细胞增强因子 2C)、MAPKAPK2(丝裂原活化的蛋白激酶活化的蛋白激酶2)、TNFRSF11A(肿瘤坏死因子受体超家族,成员11a,NFKB激活剂)、HSPA9(热休克70kDa蛋白9(致死蛋白))、CYSLTR1(半胱胺酰白三烯 受体1)、MAT1A(甲硫氨酸腺苷转移酶I,α)、OPRL1(阿片受体样 1)、IMPA1(肌醇(肌肉)-1(或4)-单磷酸酶1)、CLCN2(氯通道2)、DLD(二氢硫辛酰胺脱氢酶)、PSMA6(蛋白酶体(前体,巨蛋白因子)亚 基,α型,6)、PSMB8(蛋白酶体(前体,巨蛋白因子)亚基,β型,8(大 型多功能肽酶7))、CHI3L1(壳多糖酶3样1(软骨糖蛋白-39))、ALDH1B1(醛脱氢酶1家族,成员B1)、PARP2(聚(ADP-核糖)聚合酶2)、STAR(类固醇生成性急性期调节蛋白)、LBP(脂多糖结合蛋白)、ABCC6(ATP结合盒,亚家族C(CFTR/MRP),成员6)、RGS2(G蛋白 信号传导调节因子2,24kDa)、EFNB2(肝配蛋白-B2)、GJB6(间隙连接 蛋白,β6,30kDa)、APOA2(载脂蛋白A-II)、AMPD1(腺苷单磷酸脱氨 酶1)、DYSF(迪斯弗林(dysferlin),肢带型肌营养不良2B(常染色

体隐 性))、FDFT1(法呢酰二磷酸酯法呢酰基转移酶1)、EDN2(内皮素 2)、CCR6(趋化因子(C-C基序)受体6)、GJB3(间隙连接蛋白,β3, 31kDa)、IL1RL1(白细胞介素1受体样1)、ENTPD1(外核苷三磷酸二磷 酸水解酶1)、BBS4(巴-比二氏综合征(Bardet-Biedl syndrome)4)、 CELSR2(钙粘着蛋白,EGF LAG七经G型受体2(火烈鸟同源物,果 蝇))、F11R(F11受体)、RAPGEF3(Rap鸟苷酸交换因子(GEF)3)、 HYAL1(透明质酸葡糖胺酶1)、ZNF259(锌指蛋白259)、ATOX1 (ATX1抗氧化剂蛋白1同源物(酵母))、ATF6(活化转录因子6)、 KHK(已酮糖激酶(果糖激酶))、SAT1(亚精胺/精胺N1-乙酰转移酶 1)、GGH(γ-谷氨酰水解酶(结合酶,叶酰聚γ谷氨酰水解酶))、TIMP4 (TIMP金属肽酶抑制剂4)、SLC4A4(溶质载体家族4,碳酸氢钠协同转运 蛋白,成员4)、PDE2A(磷酸二酯酶2A,cGMP刺激的)、PDE3B(磷酸 二酯酶3B,cGMP抑制的)、FADS1(脂肪酸去饱和酶1)、FADS2(脂肪 酸去饱和酶2)、TMSB4X(胸腺素β4,X连锁的)、TXNIP(硫氧还蛋白 相互作用蛋白)、LIMS1(LIM和衰老细胞抗原样域1)、RHOB(ras同源 物基因家族,成员B)、LY96(淋巴细胞抗原96)、FOXO1(叉头框 O1)、PNPLA2(含Patatin样磷脂酶域2)、TRH(促甲状腺激素释放激 素)、GJC1(间隙连接蛋白,γ1,45kDa)、SLC17A5(溶质载体家族17 (阴离子/糖转运蛋白),成员5)、FTO(脂肪量和肥胖相关)、GJD2(间 隙连接蛋白,δ2,36kDa)、PSRC1(富含脯氨酸/丝氨酸卷曲螺旋蛋白 1)、CASP12(半胱天冬酶12(基因/假基因))、GPBAR1(G蛋白耦联胆 汁酸受体1)、PXK(含PX域丝氨酸/苏氨酸激酶)、IL33(白细胞介素 33)、TRIB1(tribbles同源物1(果蝇))、PBX4(前B细胞白血病同源框 4)、NUPR1(核蛋白,转录调节子,1)、15-Sep(15kDa硒蛋白)、CILP2 (软骨中间层蛋白2)、TERC(端粒酶RNA组分)、GGT2(γ-谷氨酰转肽 酶2)、MT-CO1(线粒体编码细胞色素c氧化酶I)、以及UOX(尿酸氧化 酶,假基因)。

[0569] 在另一个实施例中,该染色体序列可以进一步选自Pon1(对氧磷酶 1)、LDLR(LDL受体)、ApoE(载脂蛋白E)、Apo B-100(载脂蛋白B- 100)、ApoA(载脂蛋白(a))、ApoA1(载脂蛋白A1)、CBS(胱硫醚 (Cystathione)B-合酶)、糖蛋白IIb/IIb、MTHRF(5,10-亚甲基四氢叶酸还 原酶(NADPH)、及其组合。在一次迭代中,这些染色体序列和由涉及心血 管疾病的染色体序列编码的蛋白质可以选自Cacna1C、Sod1、Pten、Ppar (α)、Apo E、瘦素、及其组合。

[0570] 肾脏

[0571] 本发明还考虑了向肾脏递送该CRISPR-Cas系统。诱导治疗性核酸的细 胞摄取的递送策略包括物理力或载体系统,如基于病毒、脂质或复合体的递 送、或纳米载体。根据具有较低可能的临床相关性的最初应用,当以全身性 流体动力高压注射将核酸寻址(addressed)于肾细胞时,许多种基因治疗性 病毒和非病毒载体已经被应用于体内靶向不同的动物肾脏疾病模型中的转录 后事件(Csaba Révész和Péter Hamar(2011)。“靶向肾脏中的RNA的递送 方法”(Delivery Methods to Target RNAs in the Kidney),《基因治疗应用》 (Gene Therapy Applications),康春生教授(Prof.Chunsheng Kang)(编 辑),ISBN:978-953-307-541-9,印天科技(InTech),可获自:http://www.intechopen.com/books/gene-therapy-applications/delivery-methods-to- target-rnas-in-the-kidney)。针对肾脏的递送方法总结如下:

[0572]

递送方法	载体	靶RNA	疾病	模型	功能分析	作者
流体动力的/脂质	TransIT体内基因递送系统，DOTAP	p85α	急性肾损伤	缺血再灌注	摄取，生物分布	拉尔森（Larson）等人，《外科学》（Surgery），（2007年8月），第142卷，第2期，第（262-269）页
流体动力的/脂质	Lipofectamine 2000	Fas	急性肾损伤	缺血再灌注	血尿素氮、Fas免疫组织化学、凋亡、组织学评分	哈玛（Hamar）等人，《国家科学院院刊》（Proc Natl Acad Sci），（2004年10月），第101卷，第41期，第（14883-14888）页
流体动力的	不适用	凋亡级联元件	急性肾损伤	缺血再灌注	不适用	郑（Zheng）等人，《美国病理学杂志》（Am J Pathol），（2008年10月），第173卷，第4期，第（973-980）页

[0573]

递送方法	载体	靶RNA	疾病	模型	功能分析	作者
流体动力的	不适用	核因子k-b（NF kB）	急性肾损伤	缺血再灌注	不适用	冯（Feng）等人，《移植》（Transplantation），（2009年5月），第87卷，第9期，第（1283-1289）页
流体动力的/病毒	Lipofectamine 2000	拮抗凋亡转录因子（AATF）	急性肾损伤	缺血再灌注	凋亡、氧化应激、半胱天冬酶激活、膜脂过氧化作用	谢（Xie）和郭（Guo），《美国肾脏病学会杂志》（Am Soc Nephrol），（2006年12月），第17卷，第12期，第（3336-3346）页
流体动力的	pBAsimU6 Neo/TransIT-EE流体动力递送系统	格里莫林（Gremlin）	糖尿病肾病	链脲佐菌素诱导的糖尿病	蛋白尿、血清肌酸酐、肾小球和肾小管直径、胶原IV型/BMP7表达	Q.张（Q. Zhang）等人，《《公共科学图书馆-综合》》（PloS ONE），（2010年7月），第5卷，第7期，第（1-13）页
病毒/脂质	pSUPER载体/Lipofectamine	TGF-βII型受体	肾间质纤维化	单侧输尿管梗阻	α-SMA表达、胶原含量、	樋引（Kushibikia）等人，《控释杂志》（J Controlled Release），（2005年7月），第105卷，第3期，第（318-331）页
病毒	腺相关病毒2	盐皮质激素受体	高血压引起的肾损害	冷诱导高血压	血压、血清白蛋白、血清尿素氮、血清肌酸酐、肾重、尿钠	王（Wang）等人，《基因治疗》（Gene Therapy），（2006年7月），第13卷，第14期，第（1097-1103）页

[0574]

递送方法	载体	靶RNA	疾病	模型	功能分析	作者
流体动力的/病毒	pU6 载体	萤光素酶	不适用	不适用	摄取	小林（Kobayashi）等人，《药理学与实验治疗学杂志》（Journal of Pharmacology and Experimental Therapeutics），（2004 年 2 月），第 308 卷，第 2 期，第（688-693）页
脂质	脂蛋白类、白蛋白	apoB1、apoM	不适用	不适用	摄取、与脂蛋白类和白蛋白的结合亲和力	沃尔夫鲁姆（Wolfrum）等人，《自然生物技术》（Nature Biotechnology），（2007 年 9 月），第 25 卷，第 10 期，第（1149-1157）页
脂质	Lipofectamine 2000	p53	急性肾损伤	缺血和顺铂诱导的急性损伤	组织学评分、凋亡	莫利托里斯（Molitoris）等人，《美国肾脏病学会杂志》（J Am Soc Nephrol），（2009 年 8 月），第 20 卷，第 8 期，第（1754-1764）页
脂质	DOTAP/DOPE、DOTAP/DOPE/DOPE-PEG2000	COX-2	乳腺腺癌	荷MDA-MB-231 乳腺癌异种移植物小鼠	细胞活力、摄取	米哈伊洛瓦（Mikhaylova）等人，《癌症基因治疗》（Cancer Gene Therapy），（2011 年 3 月），第 16 卷，第 3 期，第（217-226）页

[0575]

递送方法	载体	靶RNA	疾病	模型	功能分析	作者
脂质	胆固醇	12/15-脂氧合酶	糖尿病肾病	链脲佐菌素诱导的糖尿病	白蛋白尿、尿肌酸酐、组织学、I型和IV型胶原、TGF-β、纤连蛋白、纤溶酶原激活物抑制剂1	袁（Yuan）等人，《美国生理学-肾脏生理学杂志》（Am J Physiol Renal Physiol），（2008年6月），第295卷，第（F605-F617）页
脂质	Lipofectamine 2000	线粒体膜44（TIM44）	糖尿病肾病	链脲佐菌素诱导的糖尿病	细胞增殖和凋亡、组织学、ROS、Mn-SOD和谷胱甘肽过氧化物酶的线粒体输入、细胞膜极化	Y.张（Y. Zhang）等人，《美国肾脏病学会杂志》（J Am Soc Nephrol），（2006年4月），第17卷，第4期，第（1090-1101）页
流体动力的/脂质	脂蛋白体	RLIP76	肾癌	荷Caki-2肾癌异种移植物小鼠	摄取	辛格尔（Singhal）等人，《癌症研究》（Cancer Res），（2009年5月），第69卷，第10期，第（4244-4251）页
聚合物	聚乙二醇化PEI	萤光素酶pGL3	不适用	不适用	摄取、生物分布、红细胞聚集	马莱克（Malek）等人，《毒理学与应用药理学》（Toxicology and Applied Pharmacology），（2009年4月），第236卷，第1期，第（97-108）页

[0576]

递送方法	载体	靶RNA	疾病	模型	功能分析	作者
聚合物	聚乙二醇化聚-L-赖氨酸	MAPK1	狼疮性肾小球肾炎	肾小球肾炎	蛋白尿、肾小球硬化、TGF-β、纤连蛋白、纤溶酶原激活物抑制剂1	清水（Shimizu）等人，《美国肾脏病学会杂志》（J Am Soc Nephrology），（2010年4月），第21卷，第4期，第（622-633）页
聚合物/纳米粒子	透明质酸/量子点/PEI	VEGF	肾癌/黑素瘤	B16F1黑素瘤荷瘤小鼠	生物分布、细胞毒性、肿瘤体积、胞吞作用	蒋（Jiang）等人，《分子药物学》（Molecular Pharmaceutics），（2009年5月-6月），第6卷，第3期，第（727-737）页
聚合物/纳米粒子	聚乙二醇化聚已酸内酯纳米纤维	GAPDH	不适用	不适用	细胞活力、摄取	曹（Cao）等人，《控释杂志》（J Controlled Release），（2010年6月），第144卷，第2期，第（203-212）页
适体	镜像适配体mNOX-E36	CC趋化因子配体2	肾小球硬化	单侧肾切除后的小鼠	尿蛋白、尿肌酸酐、组织病理学、肾小球滤过率、巨噬细胞计数、血清Ccl2、Mac-2+、Ki-67+	尼尼卡克（Ninichuk）等人，《美国病理学杂志》（Am J Pathol），（2008年3月），第172卷，第3期，第（628-637）页

[0577]

递送方法	载体	靶RNA	疾病	模型	功能分析	作者
适体	适体NOX-F37	加压素(AVP)	充血性心力衰竭	不适用	与D-AVP的结合亲和力、AVP信号传导的抑制、尿渗透压和钠浓度、	普尔施克(Purschke)等人，《国家科学院院刊》(Proc Natl Acad Sci)，(2006年3月)，第103卷，第13期，第(5173-5178)页

[0578] 袁(Yuan)等人(《美国生理学-肾脏生理学杂志》(Am J Physiol Renal Physiol) 295:F605-F617,2008)研究了在注射链脲佐菌素的1型糖尿病小鼠 模型中体内递送靶向花生四烯酸代谢的12/15-脂氧合酶(12/15-LO)途径的 小干扰RNA(siRNA)是否能够改善肾损伤和糖尿病肾病(DN)。为了实现 更好的体内通路以及在肾脏中的siRNA表达,袁(Yuan)等人使用与胆固醇 结合的双链12/15-LO siRNA寡核苷酸。将约400μg的siRNA皮下注射到小鼠中。袁(Yuang)等人的方法可应用于本发明的CRISPR Cas系统,其中考 虑将1-2g的与胆固醇结合的CRISPR Cas皮下注射至人,以便递送到肾脏。

[0579] 莫利托里斯(Molitoris)等人(《美国肾脏病学会杂志》(J Am Soc Nephrol)20: 1754-1764,2009年)采用了近端小管细胞(PTC)作为在肾脏 内的寡核苷酸再吸收部位,以便检验靶向p53(凋亡途径中的一种关键蛋白 质)的siRNA的效力,从而预防肾损伤。在缺血性损伤之后4小时静脉内注 射针对p53的裸露的合成siRNA最大限度地保护了PTC和肾功能。莫利托里 斯(Molitoris)等人的数据表明,在静脉内给药之后siRNA被快速递送到近端小管细胞。为了剂量反应分析,用在相同的四个时间点给予的0.33、1、 3、或5mg/kg剂量的siP53注射大鼠,分别产生1.32、4、12、和20mg/kg的 累积剂量。与PBS处理的缺血对照大鼠相比较,所有检验的siRNA剂量在第 一天产生了SCr降低作用,其中更高的剂量在经过大约五天中更有效。12和 20mg/kg的累积剂量提供了最好的保护作用。莫利托里斯(Molitoris)等人的 方法可应用于本发明的CRISPR Cas系统,其中对于人类考虑递送到肾脏的 12和20mg/kg累积剂量。

[0580] 汤普森(Thompson)等人(《核酸治疗学》(Nucleic Acid Therapeutics),第22卷,第4期,2012年)报道了合成的小干扰RNA I5NP 在啮齿类和非人类灵长动物中静脉内给药之后的毒理学和药代动力学特性。I5NP被设计为经由RNA干扰(RNAi)途径起作用以便暂时抑制促凋亡蛋白 p53的表达并且被开发为保护细胞免于诸如急性肾损伤之类的急性缺血/再灌 注损伤,急性缺血/再灌注损伤可能在重大心脏手术以及在肾移植后可能出现 的移植功能延迟中发生。在啮齿类中的800mg/kg I5NP以及在非人类灵长动 物中的1,000mg/kg I5NP的剂量对于引出不良作用是需要的,在猴中被分离 为导向对血液的作用,包括补体的亚临床激活和凝血时间的轻度增加。在大 鼠中,用I5NP的大鼠类似物未观察到另外的不良作用,表明这些作用很可能 代表合成RNA双链体的类作用,而不是与I5NP的预期药理学活性有关的毒 性。总之,这些数据支持用于在急性缺血/再灌注损伤之后保留肾功能的I5NP的静脉内给药的临床测试。在猴子中的无明显损害作用水平(NOAEL)是 500mg/kg。在猴中,

在以高达25mg/kg的剂量水平静脉内给药之后未观察到 对心血管、呼吸、以及神经系统参数的影响。因此,对于向人的肾脏静脉内 给予CRISPR Cas可以考虑相似的剂量。

[0581] 清水(Shimizu)等人(《美国肾脏病学会杂志》(J Am Soc Nephrol) 21:622-633,2010)开发了一种经由基于聚(乙二醇)-聚(L-赖氨酸)的载体将 siRNA靶向递送到肾小球的系统。该siRNA/纳米载体复合物在直径上大约为 10到20nm,该大小将允许它移动跨过有窗孔的内皮细胞而接近肾小球膜。在腹膜内注射荧光标记的siRNA/纳米载体复合物之后,清水(Shimizu)等人 在血液循环中检测到siRNA,持续一段延长的时间。在肾小球肾炎的小鼠模 型中,丝裂原活化蛋白激酶1(MAPK1)siRNA/纳米载体复合物的重复腹膜 内给药抑制了肾小球MAPK1 mRNA和蛋白质的表达。为了研究siRNA累 积,向BALB-c小鼠给予与PIC纳米载体复合的Cy5标记的siRNA(0.5ml, 5nmol的siRNA含量)、裸露的Cy5标记的siRNA(0.5ml,5nmol)、或封 装在HVJ-E中的Cy5标记的siRNA(0.5ml,5nmol的siRNA含量)。清水(Shimizu)等人的方法可应用于本发明的CRISPR Cas系统,其中考虑将约 10-20μmol的与纳米载体复合的CRISPR Cas以约1-2升腹膜内注射至人并且 递送到肾脏。

[0582] 肺

[0583] 本发明还考虑了向一侧或两侧肺部递送该CRISPR-Cas系统。

[0584] 虽然基于AAV-2的载体最初被提议用于向CF气道的CFTR递送,但其 他血清型如AAV-1、AAV-5、AAV-6、和AAV-9在肺上皮细胞的多种模型中 也展示出提高的基因转移效率(参见,例如,李(Li)等人,《分子治疗》 (Molecular Therapy),第17卷,第12期,2067-2077,2009年12月)。在 体外转导人气道上皮细胞上,AAV-1被证明比AAV-2和AAV-5更有效约 100倍,5虽然AAV-1体内转导的鼠类气管气道上皮具有与AAV-5相等的效 率。其他研究已经表明,在针对体外人气道上皮(HAE)的基因递送上, AAV-5比AAV-2更有效50倍,并且在体内小鼠肺气道上皮中显著更有效。还已经表明,在体外人气道上皮细胞中以及在体内鼠类气道中,AAV-6比AAV-2更有效。8更为近期的分离物,AAV-9,显示出在体内鼠类鼻和肺泡 上皮中展示了比AAV-5更好的基因转移效率,其中检测出基因表达持续超过 9个月,表明AAV可使得体内长期基因表达成为可能,这是对于CFTR基因 递送载体而言的理想特性。此外,证明了AAV-9可以被再次给予至鼠类的肺 部,而不丧失CFTR表达并且具有最低限度的免疫后果。可以在CF和非CF HAE培养物的顶面用100μl的AAV载体接种,维持数小时(参见,例如,李(Li)等人,《分子治疗》(Molecular Therapy),第17卷,第12期, 2067-2077,2009年12月)。MOI可以从1×10^3到4×10^5个载体基因组/细 胞而变化,这取决于病毒浓度和这些实验的目的。以上引用的载体被考虑用 于本发明的递送和/或给药。

[0585] 萨莫拉(Zamora)等人(《美国呼吸道与危重护理学杂志》(Am J Respir Crit Care Med)第183卷,第531-538页,2011年)报道了针对人类感 染性疾病治疗的RNA干扰治疗法的应用以及抗病毒药物在呼吸道合胞病毒 (RSV)感染的肺移植受体中的随机试验。萨莫拉(Zamora)等人进行了一 项在具有RSV呼吸道感染的LTX受体中的随机化、双盲、安慰剂对照的试 验。容许患者接受针对RSV的护理标准。每天给予雾化的ALN-RSV01(0.6 mg/kg)或安慰剂,持续3天。这项研究证明,可以安全地向具有RSV感染 的LTX受体给予靶向RSV的RNAi治疗剂。ALN-RSV01的三个每日剂量并 不导致任何呼吸道症状的恶化或肺功能的损害,并且未展示任何出全身性致 炎作用,如细胞因子或CRP的诱导。在吸入之后,药代动力学仅仅显示出低 的短暂的全身性暴露,与临床前动物数据一致,表明静脉内或通过吸入给予 的

ALN-RSV01通过外切核酸酶介导的消化和肾脏排泄而从循环中迅速清除。萨莫拉(Zamora)等人的方法可以应用于本发明的CRISPR Cas系统,并且雾 化的CRISPR Cas,例如以0.6mg/kg的剂量,可以被考虑用于本发明。

[0586] 对于CFTRδ508嵌合的指导RNA的实例,参见实例22,其使用腺相关 病毒(AAV)粒子证明了CRISPR-Cas系统在对其有需要的受试者或患者的 气道中的基因转移或基因递送,所述受试者或患者患有囊性纤维化或囊性纤 维化(CF)相关症状。具体地说,他们示例了囊性纤维化δF508突变的修复 策略。这种策略类型可应用于所有生物。特别提到CF,适合的患者可以包 括:人类、非人类灵长动物、犬、猫、牛、马和其他家养动物。在这种情况 下,申请人利用包含Cas9酶的CRISPR-Cas系统来靶向δF508或其他CFTR 诱导的突变。

[0587] 在这种情况下,这些经治疗的受试者的每一侧肺接受支气管内递送的药 学上有效量的雾化AAV载体系统,同时自然地呼吸。像这样,通常对于 AAV递送而言,雾化递送是优选的。腺病毒或AAV粒子可以用于递送。每 一者都可操作地连接至一个或多个调节序列上的适合的基因构建体可以被克 隆到递送载体中。在这种情况下,提供下列构建体作为实例:用于Cas9的 Cbh或EF1a启动子、用于嵌合指导RNA的U6或H1启动子:一种优选的安 排是使用靶向嵌合指导的CFTRδ508、用于δF508突变的修复模板以及密码 子优化的Cas9酶(优选的Cas9是具有核酸酶或切口酶活性的那些),所述 酶具有任选地一个或多个核定位信号或序列(NLS),例如,两(2)个 NLS。还设想了没有NLS的构建体。

[0588] 为了鉴定该Cas9靶位点,申请人分析了人CFTR基因组座位并且鉴定了 该Cas9靶位点。优选地,一般而言并且在这种CF的情况下,该PAM可以含 有一个NGG或一个NNAGAAW基序。

[0589] 因此,在CF的情况下,本方法包括操纵在感兴趣的基因组座位中的靶序 列,该方法包括:

[0590] 递送一种包含病毒载体系统的非天然存在的或工程化的组合物,该病毒载体 系统包含一种或多种病毒载体,该一种或多种病毒载体可操作地编码用于其 表达的组合物,其中该非天然存在的或工程化的组合物包括:

[0591] 一种非天然存在的或工程化的组合物,该组合物包含一种载体系统,该载体 系统包含一种或多种载体,该一种或多种载体包含

[0592] I.一种第一调节元件,该第一调节元件可操作地连接到CRISPR-Cas系统嵌合 RNA(chiRNA)多核苷酸序列上,其中该第一多核苷酸序列包含

[0593] (a)一种指导序列,该指导序列能够杂交到适合的哺乳动物细胞中的CF靶 序列上,

[0594] (b)一种tracr配对序列,和

[0595] (c)一种tracr序列,以及

[0596] II.一种第二调节元件,该第二调节元件可操作地连接到编码CRISPR酶的酶 编码序列上,该CRISPR酶包含至少一个或多个核定位序列,

[0597] 其中(a)、(b)和(c)以5'到3'方向排列,

[0598] 其中组分I和II位于该系统的相同或不同载体上,

[0599] 其中在转录时,该tracr配对序列杂交到该tracr序列上,并且该指导序列指导CRISPR复合物与该靶序列的序列特异性结合,并且

[0600] 其中该CRISPR复合物包含与(1)杂交到该靶序列上的指导序列、和(2) 杂交到该tracr序列上的tracr配对序列复合的CRISPR酶。就CF而言,优选 的靶DNA序列包含该CFTRδ508突变。上文描述了优选的PAM。优选的 CRISPR酶是任何Cas(在本文中描述,但是特别地描述于实例22中)。

[0601] 对于CF的替代物包括任何基因缺陷并且这些实例是众所周知的。本发明 的另一个优选方法或用途是用于校正EMP2A和EMP2B基因的缺陷,这些基 因已经被鉴定为与拉福拉病(Lafora disease)相关。

[0602] 在一些实施例中,一个“指导序列”可以与“指导RNA”不同。一种指 导序列可以是指在该指导RNA内的大约20bp的序列,其指定该靶位点。

[0603] 在一些实施例中,该Cas9是(或衍生自)SpCas9。在这样的实施例中, 优选的突变是在SpCas9的任何或所有的10、762、840、854、863和/或986 位置或在其他Cas9的相应位置(其可以例如通过标准序列比较工具进行确 定)。具体地说,在SpCas9中的任何或所有下列突变是优选的:D10A、 E762A、H840A、N854A、N863A和/或D986A;并且还设想了这些置换氨基酸的任何一种的保守性取代。在其他Cas9中的相应位置处的相同取代(或这 些突变的保守性取代)也是优选的。特别优选的是在SpCas9中的D10和 H840。然而,在其他Cas9中,相应于SpCas9 D10和H840的残基也是优选 的。这些是有利的,因为它们提供了切口酶活性。这样的突变可以应用于本 发明的所有方面,不仅是治疗CF。

[0604] 施万克(Schwank)等人(《细胞—干细胞》(Cell Stem Cell),13: 653-58,2013)使用了CRISPR/Cas9来校正在人类干细胞中的与囊性纤维化 相关联的缺陷。该研究组的目标是离子通道囊性纤维化跨膜传导受体 (CFTR)的基因。在CFTR中的缺失引起该蛋白质在囊性纤维化患者中的错 误折叠。使用从患有囊性纤维化的两个儿童的细胞样品中发育而来的培养的 肠干细胞,施万克(Schwank)等人能够利用CRISPR连同含有有待插入的修 补序列的供体质粒来校正该缺陷。然后这些研究者使这些细胞生长成肠“类 器官”或微型肠,并且显示它们正常地起作用。在这种情况下,约一半的克 隆类器官经历正确的基因修正。

[0605] 肌肉

[0606] 本发明还考虑了向肌肉递送该CRISPR-Cas系统。

[0607] Bortolanza等人(《分子治疗》(Molecular Therapy)第19卷,第11 期,2055-2064,2011年11月)表明,在面肩肱型肌营养不良(FSHD)发作 之后的FRG1小鼠中,干扰RNA表达盒的全身性递送导致剂量依赖性长期 FRG1敲低,而没有毒性征象。Bortolanza等人发现,单次静脉内注射5×10^{12} vg的rAAV6-sh1FRG1挽救了FRG1小鼠的肌肉组织病理学和肌肉功能。详细 而言,使用一个25号的泰尔茂(Terumo)注射器将200μl的含有2×10^{12}或 5×10^{12}vg的在生理溶液中的载体在注射到尾静脉中。Bortolanza等人的方法 可以应用于表达CRISPR Cas的AAV,并且将其以约2×10^{15}或2×10^{16}vg载 体的剂量注射到人体内。

[0608] Dumonceaux等人(《分子治疗》(Molecular Therapy)第18卷,第5 期,881-887,2010年5月)使用针对肌肉生长抑制素受体AcvRIIb mRNA (sh-AcvRIIb)的RNA干扰技术抑制了肌肉生长抑制素途径。由矢量化 (vectorized)U7外显子跳跃技术(U7-DYS)介导准肌营养不良蛋白(quasi- dystrophin)的恢复。将携带单独的sh-AcvrIIb构建体、单独的U7-DYS构建 体、或这两种构建体的组合的腺相关载体注射到营养不良mdx小鼠的胫骨前 肌(TA)肌肉中。以10^{11}个AAV病毒基因组进行注射。Dumonceaux等人的 方法可以应用于表达

CRISPR Cas的AAV,并且将其以例如约10^{14}到10^{15}vg 载体的剂量注射到人体内。

[0609] 木内(Kinouchi)等人(《基因治疗》(Gene Therapy)(2008)15，1126-1130)报道了通过未经化学修饰的siRNA与缺端胶原(ATCOL)的纳 米粒子形成的到正常或患病小鼠骨骼肌的体内siRNA递送的有效性。靶向肌 肉生长抑制素(一种骨骼肌生长的负调节剂)的siRNA的ATCOL介导的在 小鼠骨骼肌中的局部应用或者经由静脉内在应用之后几周之内引起肌肉质量 的显著增加。这些结果暗示siRNA的ATCOL介导的应用是一种强大的用于 包括肌肉萎缩在内的疾病的未来治疗用途的工具。根据制造商的说明,将 Mst-siRNA(终浓度，10mM)与ATCOL(对于局部给药的终浓度,0.5％) (AteloGene,高研株式会社(Kohken),东京,日本)混合。在通过戊巴比 妥钠(25mg/kg,腹膜内注射)麻醉小鼠(20周大,雄性C57BL/6)之后，将Mst-siRNA/ATCOL复合物注射到咬肌和股二头肌中。木内(Kinouchi)等 人的方法可以应用于CRISPR Cas,并且例如将其以40μM溶液的约500到 1000ml的剂量注射到人体内。

[0610] 哈格斯特龙(Hagstrom)等人(《分子治疗》(Molecular Therapy)第 10卷,第2期,2004年8月)描述了一种使得能够将核酸有效且可重复地递 送到遍及哺乳动物四肢肌肉的肌细胞(肌纤维)的血管内、非病毒方法。该 程序涉及注射裸质粒DNA或siRNA到暂时由止血带或血压袖带分离的肢体 的远端静脉中。通过以足够的体积将其迅速注射,促进了向肌纤维的核酸递 送,使得该核酸溶液能够溢出到肌肉组织中。在小动物和大动物中都以最低毒性实现了在骨骼肌中的高水平转基因表达。还获得了向四肢肌肉递送 siRNA的证据。为了将质粒DNA静脉注射到恒河猴中,将一个三通旋塞连接 到各自加载有单个注射器的两个注射器泵(型号PHD 2000;哈佛仪器公司 (Harvard Instruments))上。在罂粟碱注射五分钟之后,以1.7或2.0ml/s的 速率注射pDNA(15.5到25.7mg,在40-100ml盐水中)。对于本发明的表 达CRISPR Cas的质粒DNA,这可以按比例增加,对于人类,注射在800到 2000ml盐水中的约300到500mg。对于到大鼠中的腺病毒载体注射,注射在 3ml的生理盐水溶液中的(NSS)2x 10^9个感染粒子。对于本发明的表达 CRISPR Cas的腺病毒载体,这可以按比例增加,对于人类,注射在10升 NSS中的约1x 10^{13}个感染粒子。对于siRNA,以12.5μg的siRNA注射到大 鼠的大隐静脉中,并且以750μg的siRNA注射到灵长动物的大隐静脉中。对 于本发明的CRISPR Cas,这可以按比例增加,例如,向人的大隐静脉注射约 15到约50mg。

[0611] 皮肤

[0612] 本发明还考虑了向皮肤递送该CRISPR-Cas系统。

[0613] 希克森(Hickerson)等人(《分子治疗-核酸》(Molecular Therapy- Nucleic Acids)(2013)2,e129)涉及一种用于向人类和鼠类皮肤递送自我 递送(sd)-siRNA的机动化的微针阵列皮肤递送装置。将基于siRNA的皮肤治 疗剂转化到临床的主要挑战是有效递送系统的开发。在多种皮肤递送技术中 已经投入了实质性的努力,但是成功有限。在其中用siRNA治疗皮肤的临床 研究中,与皮下针注射相关的剧烈疼痛预先排除了试验中额外患者的纳入，这凸显了对于改进的、更为“患者友好的”(即,很少或没有疼痛)递送方 法的需要。微针代表行包括siRNA在内的大的带电负荷物跨越一级屏障、角 质层进行递送的有效途径,并且通常被认为比常规皮下针疼痛更少。机动化 的“冲压型”微针装置,包括由希克森(Hickerson)等人使用的机动化的微 针阵列(MMNA)装置,已经显示在无毛小鼠研究中是安全的并且引起很少 的或没有疼痛,其证据为:(i)在美容业中广泛使用以及(ii)其中几

乎所 有志愿者发现使用该装置比流感疫苗针剂(flushot)疼痛少得多的有限测试，表明使用这种装置的siRNA递送将产生比使用皮下针注射的先前临床试验中 所体验的少得多的疼痛。该MMNA装置(作为Triple-M或Tri-M由韩国首尔 的Bomtech电子有限公司在市场上销售)适用于将siRNA递送到小鼠和人类 皮肤。将sd-siRNA溶液(高达300μl的0.1mg/ml RNA)引入被设定为0.1 mm深度的一次性Tri-M针盒(Bomtech)的腔室中。为了处理皮肤，在处理 之前将未鉴定的皮肤(在外科手术之后立即获得)手动拉伸并且钉在软木平 台上。使用具有28号0.5英寸针头的胰岛素注射器进行所有皮内注射。该 MMNA装置和希克森(Hickerson)等人的方法可以用于和/或适用于，例 如，以高达300μl的0.1mg/ml CRISPR Cas的剂量向皮肤递送本发明的 CRISPR Cas。

[0614] 里奇曼(Leachman)等人(《分子治疗》(Molecular Therapy)，第18 卷，第2期，442-446，2010年2月)涉及利用基于针对皮肤的第一短干扰 RNA(siRNA)的治疗剂用于治疗罕见的皮肤病症先天性厚甲(PC)的Ib期 临床试验，先天性厚甲为一种常染色体显性综合征，其包括致残性的掌跖角 化病。这种siRNA，称为TD101，特异性地并且有力地靶向角蛋白6a(K6a)N171K突变体mRNA，而不影响野生型K6a mRNA。剂量递增计划 提出如下：

[0615]

周	剂量编号	天	体积(ml)	TD101的浓度 (mg/ml)	总剂量 TD101 (mg)
1	1-2	1-7	0.1	1.0	0.10
2	3-4	8-14	0.25	1.0	0.25
3	5-6	15-21	0.50	1.0	0.50
4	7-8	22-28	1.0	1.0	1.0
5	9-10	29-35	1.5	1.0	1.5
6	11-12	36-42	2.0	1.0	2.0
7	13-14	43-49	2.0	1.5	3.0
8	15-16	50-56	2.0	2.0	4.0
9	17-18	57-63	2.0	2.5	5.0
10	19-20	64-70	2.0	3.0	6.0
11	21-22	71-77	2.0	3.5	7.0
12	23-24	78-84	2.0	4.0	8.0
13	25-26	85-91	2.0	4.5	9.0
14	27-28	92-98	2.0	5.0	10.0
15	29-30	99-105	2.0	6.0	12.0
16	31-32	106-112	2.0	7.0	14.0
17	33	113-119	2.0	8.5	17.0

[0616] 最初，向对称的胼胝给予0.1ml的TD101或单独的载体(没有钙或镁的 杜尔贝科磷酸盐缓冲盐水)的1.0mg/ml溶液。完成六个渐增的剂量体积，对 于增加没有不良反应：每次注射0.1、0.25、0.5、1.0、1.5、和2.0ml的1.0 mg/ml TD101溶液。由于最高计划体积(2.0ml)被良好耐受，然后每周将 TD101的浓度从1mg/ml增加到8.5mg/ml的终浓度。对于给予特异性地并且 有力地靶向角蛋白6a(K6a)N171K突变体mRNA的CRISPR Cas考虑了相 似的剂量。

[0617] 郑(Zheng)等人(PNAS，7月24日，2012年，第109卷，第30期， 11975-11980)表明，球形核酸纳米粒子结合物(SNA-NC)，由高度定向 的、共价固定的siRNA的致密壳围绕的金核，在应用之后数小时之内在体外 自由地穿透几乎100％的角化细胞、小鼠皮肤、以及人表皮。郑(Zheng)等 人证明，在人类皮肤中单次应用25nM的表皮生长因子受体(EGFR)SNA- NC持续60小时显示出有效的基因敲低。对于向皮肤给予固定在SNA-NC中 的CRISPR Cas可以考虑了相似的剂量。

[0618] 肝炎病毒

[0619] 本发明还可应用于治疗B型肝炎病毒(HBV)。然而，通过例如优化剂 量和序列，该

CRISPR Cas系统必须适应于避免RNAi的缺点,如内源小 RNA途径的过度饱和风险(参见,例如,格林姆(Grimm)等人,《自然》 第441卷,2006年5月26日)。例如,考虑例如约每人1-10x 10^{14}个粒子的 低剂量。

[0620] 在另一个实施例中,针对HBV的该CRISPR Cas系统可以在脂质体中给 药,如一种稳定的核酸-脂质粒子(SNALP)(参见,例如,莫里西 (Morrissey)等人,《自然生物技术》(Nature Biotechnology),第23卷, 第8期,2005年8月)。考虑了约1、3或5mg/kg/天的靶向SNALP中的HBV RNA的CRISPR Cas的每日静脉注射。每天治疗可以经过约三天,然后 每周治疗持续约五周。

[0621] 在另一个实施例中,陈(Chen)等人的系统(《基因治疗》(2007) 14,11-19)可以使用和/或适用于本发明的CRISPR Cas系统。陈(Chen)等 人使用双链腺相关病毒8-假型载体(dsAAV2/8)来递送shRNA。在HBV转 基因小鼠的肝脏中,携带HBV特异性shRNA的dsAAV2/8载体(1x 10^{12}个 载体基因组/小鼠)的单次给药有效抑制了HBV蛋白、mRNA和复制型DNA 的稳定水平,从而导致在循环中的HBV负荷的高达2-3个$\log_{10}$减少。在载体 给药之后,显著的HBV抑制持续了至少120天。shRNA的治疗作用是靶序列 依赖性的并且不涉及干扰素的激活。对于本发明,可以将针对HBV的 CRISPR Cas系统克隆到AAV载体如dsAAV2/8载体中,并且给予至人,例 如,以1x 10^{15}个载体基因组到约1x 10^{16}个载体基因组/人的剂量。

[0622] 在另一个实施例中,伍德尔(Wooddell)等人的方法(《分子治疗》第 21卷,第5期,973-985,2013年5月)可以使用和/或适用于本发明的 CRISPR Cas系统。伍德尔(Wooddell)等人表明,肝细胞靶向的、N-乙酰半 乳糖胺结合的蜂毒肽样肽(NAG-MLP)与靶向凝血因子VII(F7)的嗜肝胆 固醇结合的siRNA(chol-siRNA)的简单共注射导致在小鼠和非人类灵长动 物中的有效F7敲低,而在细胞因子的临床化学或诱导上没有变化。使用 HBV感染的瞬时转基因小鼠模型,伍德尔(Wooddell)等人表明,NAG- MLP与有效的靶向保守HBV序列的chol-siRNA的简单共注射导致病毒 RNA、蛋白质、和病毒DNA的多对数(multilog)阻抑。对于本发明可以设 想例如,约6mg/kg的NAG-MLP和6mg/k的HBV特异性CRISPR Cas的静 脉内共注射。在替代方案中,约3mg/kg的NAG-MLP和3mg/kg的HBV特 异性CRISPR Cas可以在第一天递送,随后在两周之后给予约2-3mg/kg的 NAG-MLP和2-3mg/kg的HBV特异性CRISPR Cas。

[0623] 本发明还可应用于治疗C型肝炎病毒(HCV)。Roelvinki等人的这些方 法(《分子治疗》第20卷,第9期,1737-1749,2012年9月)可以应用于 该CRISPR Cas系统。例如,AAV载体如AAV8可以是考虑的载体,并且可 以考虑例如约1.25×10^{11}到1.25×10^{13}个载体基因组/千克体重(vg/kg)的剂 量。

[0624] 将易于清楚的是,可以按照相似的方式治疗具有其他疾病的宿主。本文 提供了由突变引起的遗传疾病的一些实例,但是更多的疾病是已知的。以上 策略可以应用于这些疾病。

[0625] 亨廷顿病(HD)

[0626] 通过降低HTT(亨廷顿病的致病基因)的表达,RNA干扰(RNAi)对 这种病症提供了治疗潜能(参见,例如,麦克布赖德(McBride)等人,《分 子治疗》(Molecular Therapy)第19卷,第12期,2011年12月,第2152- 2162页),因此申请人假定它可以用于/或适用于该CRISPR-Cas系统。可以 使用一种算法来生成该CRISPR-Cas系统,以便降低反义序列的脱靶可能性。这些CRISPR-Cas序列可以靶向小鼠、恒河猴或人类亨廷丁的外显子52中的 一个

序列并且表达于一种病毒载体如AAV中。可以按照约三次微量注射/半 球(总共六次注射)来注射包括人类在内的动物:第一次在前连合的吻侧1 mm(12μl),并且用1e12 vg/ml的AAV以约1μl/分的速率在第一次注射的 尾侧间隔3mm和6mm进行其余两次注射(分别为12μl和10μl),并且将 针留在适当位置持续另外的5分钟,以允许注射物从针尖扩散。

[0627] 迪费吉里亚(DiFiglia)等人(PNAS,10月23日,2007年,第104 卷,第43期,17204-17209)观察到,向成熟的纹状体中单次给予靶向Htt的 siRNA可以使突变体Htt沉默,减弱神经元病理,并且延迟在急骤起病的HD 的病毒性转基因小鼠模型中观察到的异常行为表现。迪费吉里亚(DiFiglia) 用2μl的Cy3标记的cc-siRNA-Htt或10μM的未结合的siRNA-Htt注射到小 鼠的纹状体内。在本发明中对于人类可以考虑靶向Htt的CRISPR Cas的相似 剂量,例如,可以将约5-10ml的10μM的靶向Htt的CRISPR Cas注射到纹 状体内。

[0628] 在另一个实施例中,布德罗(Boudreau)等人(《分子治疗》 (Molecular Therapy)第17卷,第6期,2009年6月)将5μl的表达htt特 异性RNAi病毒的重组AAV血清型2/1载体(以4x 10^{12}病毒基因组/ml)注 射到纹状体(straiatum)中。在本发明中对于人类可以考虑靶向Htt的 CRISPR Cas的相似剂量,例如,可以将约10-20ml的4x 10^{12}个病毒基因组 /ml靶向Htt的CRISPR Cas注射到纹状体内。

[0629] 在另一个实施例中,可以连续给予靶向HTT的CRISPR Cas(参见,例 如,于(Yu)等人,《细胞》(Cell)150,895-908,8月31日,2012 年)。于(Yu)等人利用递送0.25ml/小时的渗透泵(型号2004)来递送 300mg/天的ss-siRNA或磷酸盐缓冲盐水(PBS)(西格玛奥德里奇公司 (Sigma Aldrich),持续28天,并且使用被设计为递送0.5μl/小时的泵(型 号2002)来递送75mg/天的阳性对照MOE ASO,持续14天。用稀释在无菌 PBS中的ss-siRNA或MOE充满泵(度瑞公司(Durect Corporation),进而在 植入之前将其在37℃孵育24或48(型号2004)小时。用2.5%异氟烷麻醉 小鼠,并且在头盖骨底部做正中切口。使用脑功能区定位导子,将套管植入 右侧脑室并且用乐泰胶(Loctite adhesive)固定。将一根附接到Alzet微渗透压泵的导管附接到该套管上,并且将该泵皮下置于肩胛间区(midscapular area)中。用5.0号尼龙缝线将该切口闭合。在本发明中对于人类可以考虑靶 向Htt的CRISPR Cas的相似剂量,例如,可以给予约500到1000g/天的靶向Htt的CRISPR Cas。

[0630] 在连续输注的另一个实例中,斯泰尔斯(Stiles)等人(《实验神经病 学》(Experimental Neurology)233(2012)463-471)将一根具有钛针尖的脑 实质导管植入到右侧壳核中。将该导管连接到一个皮下植入在腹部的 SynchroMed® II泵(美敦力神经调控部门(Medtronic Neurological),明尼阿 波里斯市,明尼苏达州)上。在以6μL/天输注磷酸盐缓冲盐水7天之后,用 试验品将这些泵重新充满并且编程为连续递送7天。以约0.1到0.5μL/分的 变化输注速率输注约约2.3到11.52mg/d的siRNA。在本发明中对于人类可以考虑靶向Htt的CRISPR Cas的相似剂量,例如,可以给予约20到200mg/天 的靶向Htt的CRISPR Cas。

[0631] 在另一个实施例中,转让给桑加莫(Sangamo)的美国专利公开号 20130253040的这些方法也可以适合于用于治疗亨廷顿病的从塔乐斯 (TALES)到本发明的CRISPR Cas系统。

[0632] 核酸、氨基酸和蛋白质

[0633] 本发明使用核酸来结合靶TDNA序列。这是有利的,因为产生核酸比产 生蛋白质容

易且价廉得多,并且特异性可以根据其中寻求同源性的拉伸长度 而变化。例如多指的复杂3-D复杂定位是不需要的。

[0634] 术语“多核苷酸”、“核苷酸”、“核苷酸序列”、“核酸”和“寡核 苷酸”可互换地使用。它们是指具有任何长度的核苷酸的聚合形式,是脱氧 核糖核苷酸或核糖核苷酸、或其类似物。多核苷酸可具有任何三维结构,并 且可以执行已知或未知的任何功能。以下是多核苷酸的非限制性实例:基因 或基因片段的编码区或非编码区、根据连接分析定义的多个座位(一个座 位)、外显子、内含子、信使RNA(mRNA)、转运RNA、核糖体RNA、 短干扰RNA(siRNA)、短发夹RNA(shRNA)、micro-RNA(miRNA)、 核酶、cDNA、重组多核苷酸、分支多核苷酸、质粒、载体、任何序列的分离 的DNA、任何序列的分离的RNA、核酸探针、和引物。该术语还涵盖具有合 成骨架的核酸样结构,参见,例如,埃克斯坦(Eckstein),1991;巴塞加 (Baserga)等人,1992;米利根(Milligan),1993;WO 97/03211;WO 96/39154;马塔(Mata),1997;施特劳斯-绍库普(Strauss-Soukup), 1997;和扎姆斯塔(Samstag),1996。多核苷酸可以包含一个或多个经修饰 的核苷酸,如甲基化的核苷酸和核苷酸类似物。如果存在,可以在聚合物组装之前或之后进行核苷酸结构的修饰。核苷酸的序列可以被非核苷酸组分中 断。多核苷酸可以在聚合后,如通过与标记的组分缀合来进一步修饰。

[0635] 如本文所用的术语“野生型”是本领域技术人员所理解的术语,并且表 示生物、菌株、基因的典型形式或者当它在自然界存在时区别于突变体或变 体形式的特征。

[0636] 如本文所用的术语“变体”应当被理解为表示具有衍生自在自然界中存 在的模式的性质的展示。

[0637] 术语“非天然存在的”或“工程化的”可互换地使用并且表面人工的参 与。这些术语,当指核酸分子或多肽时,表示该核酸分子或多肽至少基本上 从它们在自然界中或如发现于自然界中的与其结合的至少另一种组分游离出 来。

[0638] “互补性”是指核酸与另一个核酸序列借助于传统的沃森-克里克碱基配 对或其他非传统类型形成一个或多个氢键的能力。互补百分比表示一个核酸 分子中可与一个第二核酸序列形成氢键(例如,沃森-克里克碱基配对)的残 基的百分比(例如,10个之中有5、6、7、8、9、10个即为50%、60%、 70%、80%、90%、和100%互补)。“完全互补”表示一个核酸序列的所有 连续残基与一个第二核酸序列中的相同数目的连续残基形成氢键。如本文使用的“基本上互补”是指在一个具有8、9、10、11、12、13、14、15、16、 17、18、19、20、21、22、23、24、25、30、35、40、45、50个或更多个核 苷酸的区域上至少为60%、65%、70%、75%、80%、85%、90%、95%、 97%、98%、99%、或100%的互补程度,或者是指在严格条件下杂交的两个 核酸。

[0639] 如本文使用的对于杂交的“严格条件”是指与靶序列具有互补性的一个 核酸主要地与该靶序列杂交并且基本上不杂交到非靶序列上的条件。严格条 件通常是序列依赖性的,并且取决于许多因素而变化。一般而言,该序列越 长,则该序列特异性地杂交到其靶序列上的温度就越高。严格条件的非限制 性实例描述于蒂森(Tijssen)(1993)的《生物化学和分子生物学中的实验 室技术-核酸探针杂交》(Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes),第I部分,第二章,“杂交 原理概述和核酸探针分析策略”(“Overview of principles of hybridization and the strategy of nucleic acid probe assay”),爱思唯尔

(Elsevier)，纽约。在提 及一个多核苷酸序列时，那么也设想了互补的或部分互补的序列。能够在高 严格条件下杂交到参考序列上的这些序列是优选的。通常，为了使杂交率最 大化，选择了相对低严格性的杂交条件：低于热熔点(T_m)约20℃到 25℃。T_m是50％的特异性靶序列在具有规定的离子强度和pH的溶液中杂交 到完全互补的探针上的温度。通常，为了要求杂交序列的至少约85％的核苷 酸互补性，高严格洗涤条件被选择为低于T_m约5℃到15℃。为了要求杂交 序列的至少约70％的核苷酸互补性，中等严格洗涤条件被选择为低于T_m约 15℃到30℃。高容许(极低严格性)洗涤条件可以低至在T_m之下50℃，从 而允许在杂交序列之间的高水平错配。本领域技术人员将认识到，在杂交和 洗涤阶段中的其他物理和化学参数也可以改变，从而影响来自在靶序列与探 针序列之间的特定同源性水平的可检测杂交信号的结果。优选的高严格条件 包括在50％甲酰胺、5×SSC、和1％SDS中在42℃孵育，或者在5×SSC和 1％SDS中在65℃孵育，在0.2×SSC和0.1％SDS中在65℃洗涤。

[0640] “杂交”是指其中一个或多个多核苷酸反应形成一种复合物的反应，该 复合物经由这些核苷酸残基之间的碱基的氢键键合而稳定化。氢键键合可以 借助于沃森-克里克碱基配对、Hoogstein结合或以任何其他序列特异性方式而 发生。该复合物可包含形成一个双链体的两条链、形成多链复合物的三条或 多条链、单个自我杂交链、或这些的任何组合。杂交反应可以构成一个更广 泛的过程(如PCR的开始、或经由一种酶的多核苷酸的切割)中的一个步 骤。能够与一个给定序列杂交的序列被称为该给定序列的“互补物”。

[0641] 如本文使用的，术语“基因组座位(locus)”或“座位(locus)”(复 数是座位(loci))是在一个染色体上的一个基因或DNA序列的特定位置。“基因”是指编码多肽或RNA链的DNA或RNA的段(stretch)，其在生物 中发挥功能作用并且因此是活生物体遗传的分子单元。出于本发明的目的，可以考虑包括调节基因产物的产生的区域的基因，而不论这样的调节序列是 否接近编码和/或转录的序列。因此，基因包括而不必限于，启动子序列、终止子、翻译调节序列(如核糖体结合位点和内部核糖体进入位点)、增强 子、沉默子、隔离子、边界元件、复制起点、核基质附着位点和座位控制 区。

[0642] 如本文使用的“基因组座位的表达”或“基因表达”是藉此在功能性基 因产物的合成中使用来自基因的信息的过程。基因表达的产物常常是蛋白 质，但是在非蛋白质编码基因如rRNA基因或tRNA基因中，产物是功能性 RNA。基因表达的过程由所有已知的生物利用-产生功能性产物以便存活的真 核生物(包括多细胞生物)、原核生物(细菌和古细菌)以及病毒。如本文 使用的基因或核酸的“表达”不仅涵盖细胞基因表达，而且涵盖在克隆系统中或在任何其他背景下的一个或多个核酸的转录和翻译。如本文使用的“表 达”是指藉此从DNA模板转录成多核苷酸(如转录成mRNA或其他RNA转 录物)的过程和/或转录的mRNA随后藉此翻译成肽、多肽或蛋白质的过程。转录物和编码的多肽可以总称为“基因产物”。如果多核苷酸来源于基因组 DNA，表达可以包括真核细胞中mRNA的剪接。

[0643] 术语“多肽”、“肽”和“蛋白质”在本文可互换地使用，是指具有任 何长度的氨基酸的聚合物。该聚合物可以是可以是直链或支链的，它可以包 含修饰的氨基酸，并且它可以被非氨基酸中断。这些术语还涵盖已经被修饰 的氨基酸聚合物；这些修饰例如二硫键形成、糖基化、脂化(lipidation)、乙酰化、磷酸化、或任何其他操纵，如与标记组分的结合。如本文使用的术 语“氨基酸”包括天然的和/或非天然的或者合成的氨基酸，包括甘氨酸以及D和L旋光异构体、以及氨基酸类似物和肽模拟物。

[0644] 如本文使用的,术语“结构域”或“蛋白质结构域”是指可以独立于该 蛋白质链的其余部分而存在并且起作用的蛋白质序列的一部分。

[0645] 正如在本发明的多个方面中所描述,序列一致性与序列同源性有关。可 以通过肉眼、更通常地借助于可得的序列比较程序来进行同源性比较。这些 可商购的计算机程序可以计算在两个或更多个序列之间的同源性的百分比 (%)并且还可以计算由两个或更多个氨基酸或核酸序列共享的序列一致 性。在一些优选的实施例中,本文描述的dTALE的封端区具有与本文提供的 封端区氨基酸序列至少95%一致性或共享一致性的序列。

[0646] 可以通过本领域已知的许多计算机程序(例如BLAST或FASTA等等) 来生成序列同源性。用于进行这样的比对的适合的计算机程序是GCG威斯康 星Bestfit软件包(美国威斯康星大学;德弗罗(Devereux)等人,1984,《核酸研究》(Nucleic Acids Research)12:387)。可以进行序列比较的其他 软件的实例包括但不限于,BLAST软件包(参见奥苏贝尔(Ausubel)等人, 1999,出处同上-第18章)、FASTA(阿尔丘尔(Atschul)等人,1990,《分子生物学杂志》(J.Mol.Biol.),403-410)以及GENEWORKS系列比 较工具。BLAST和FASTA均可用于离线和在线搜索(参见奥苏贝尔 (Ausubel)等人,1999,出处同上,7-58页到7-60页)。然而,优选使用 GCG Bestfit程序。

[0647] 可以计算在连续序列上的序列同源性百分比(%),即,将一个序列与 另一个序列比对,并且将一个序列中的每个氨基酸或核苷酸与另一个序列中 的相应氨基酸或核苷酸直接进行比较,一次比较一个残基。这被称为“无空 位”比对。典型地,这样的无空位比对仅仅在较少数目的残基上进行。

[0648] 虽然这是一种非常简单和一致的方法,但是它未能考虑的是,例如,在 其他方面完全相同的序列对中,一个插入或缺失可以引起随后的氨基酸残基 被排除在比对之外,因此当进行全局比对时可能导致在同源性%上的大幅降 低。因此,大多数序列比较方法被设计为产生考虑到可能的插入和缺失的优 化比对,而没有过度地使总体同源性或一致性评分不利。这是通过在序列比 对中插入“空位”来实现的,以便试图将局部同源性或一致性最大化。

[0649] 然而,这些更复杂的方法向出现在该比对中的每个空位分配“空位罚 分”,从而对于相同数目的一致的氨基酸而言,具有尽可能少的空位的序列 比对-反映了在这两个比较的序列之间的更高关联性-可以比具有许多空位者实 现更高的分数。“亲合空位成本”(“Affinity gap costs”)典型地用于对空位 的存在要求相对高的成本并对空位中的每个后续残基施加较小的罚分。这是 最常用的空位评分系统。高空位罚分当然将产生具有更少空位的最佳比对。 大多数比对程序允许修改空位罚分。然而,当使用这种序列比较软件时优选 使用默认值。例如当使用GCG威斯康星Bestfit软件包时,氨基酸序列的默认 空位罚分为对于每个空位的-12,以及对于每个延伸的-4。

[0650] 因此计算最大同源性%首先要求产生最佳比对,考虑空位罚分。用于进 行这样的比对的适合的计算机程序是GCG威斯康星Bestfit软件包(德弗罗 (Devereux)等人,1984,《核酸研究》(Nucleic Acids Research)12 p387)。可以进行序列比较的其他软件的实例包括但不限于,BLAST软件包 (参见奥苏贝尔(Ausubel)等人,1999《精编分子生物学实验指南》(Short Protocols in Molecular Biology),第4次编辑-第18章)、FASTA(阿尔丘尔(Altschul)等人,1990《分子生物学杂志》(J.Mol.Biol)。BLAST和 FASTA均可用于离线和在

线搜索(参见奥苏贝尔(Ausubel)等人,1999,《精编分子生物学实验指南》(Short Protocols in Molecular Biology),7-58 页到7-60页)。然而,对于一些应用,优选使用GCG Bestfit程序。一种新的 工具,称为BLAST 2序列,也可用于比较蛋白质和核苷酸序列(参见《欧洲 微生物学会联合会微生物学快报》(FEMS Microbiol Lett)(FEMS Microbiol Lett.)1999 174(2):247-50;《欧洲微生物学会联合会微生物学快报》(FEMS Microbiol Lett)1999 177(1):187-8以及在国立卫生研究院的网址的国家生物 技术信息中心的网址)。

[0651] 虽然最终同源性%可以按照一致性进行测量,但是比对过程自身典型地 不是基于全或无配对比较(all-or-nothing pair comparison)。相反,通常使用 尺度相似性评分矩阵(scaled similarity score matrix),基于化学相似性或进 化距离对各个成对比较评分。通常使用的这种矩阵的实例为BLOSUM62矩阵 -BLAST系列程序的默认矩阵。GCG威斯康星程序通常使用公开的默认值或 自定义符号比较表(更多细节详见用户手册)。对于一些应用,优选使用 GCG软件包的公共默认值,或在其他软件情况下的默认矩阵,如BLOSUM62。

[0652] 可替代地,可以基于类似于CLUSTAL(希金斯DG(Higgins DG)和夏 普PM(Sharp PM)(1988),《基因》(Gene)73(1),237-244)的一 种算法,使用在DNASIS™(日立软件公司(Hitachi Software))中的多重比 对特征来计算同源性百分比。一旦软件已经产生最佳比对,就有可能计算同 源性%,优选序列一致性%。软件典型地这样进行,作为序列比较的一部分, 并且产生数字结果。

[0653] 这些序列也可以具有氨基酸残基的缺失、插入或取代,其产生沉默变化 并生成功能上等同的物质。可以基于氨基酸特性(如残基的极性、电荷、可 溶性、疏水性、亲水性、和/或两亲性)的相似性做出有意氨基酸取代,并且 因此它在将氨基酸集合成官能团中是有用的。可以仅仅基于氨基酸的侧链特 性将它们集合在一起。然而,包括突变数据同样是更有用的。出于结构原 因,如此衍生的氨基酸的集合很有可能是保守性的。这些集合可以被描述为 文氏图形式(Venn diagram)(利文斯敦C.D.(Livingstone C.D.)和巴顿G.J.(Barton G.J.)(1993)“蛋白质序列比对:用于残基保守些的分级分析的策 略”(“Protein sequence alignments:a strategy for the hierarchical analysis of residue conservation”)《生物科学计算应用》(Comput.Appl.Biosci.)9:745- 756)(泰勒(Taylor)W.R.(1986)“氨基酸保守性的分类”(“The classification of amino acid conservation”)《理论生物学杂志》(J. Theor.Biol.)119;205-218)。例如可以根据下表做出保守性取代,该表描述 了普遍接受的氨基酸的文氏图分组。

[0654]

集合		子集	
疏水性的	FWYHKMILVAGC	芳香族的	FWYH
		脂肪族的	ILV
极性的	WYHKREDCSTNQ	带电荷的	HKRED
		带正电荷的	HKR
		带负电荷的	ED
小的	VCAGSPTND	极小的	AGS

[0655] 本发明的实施例包括可包含同源取代(本文使用取代和置换两者来表示 现有氨基酸残基或核苷酸与替代残基和核苷酸之间的互换)的序列(多核苷 酸或多肽两者),该同源取代,即,在氨基酸的情况下发生同类取代(like- for-like substitution),如碱性对碱性、酸性对酸性、极性对极性。也可以发 生非同源取代,即,从一类残基到另一类残基,或者可替代地涉及包含非天 然氨基酸如鸟氨酸(在下文称为Z)、二氨基丁酸鸟氨酸(在下文称为B)、 正亮氨酸鸟氨酸(在下文称为O)、吡啶基丙氨酸、噻吩基丙氨酸、萘基丙 氨酸和苯基甘氨酸。

[0656] 变体氨基酸序列可以包括适合的可插入在该序列的任何两个氨基酸残基 之间的间隔基团,包括除了氨基酸间隔物(如甘氨酸或β-丙氨酸)之外的烷 基基团,如甲基、乙基或丙基基团。一个另外的变化形式,其涉及处于类肽 形式的一个或多个氨基酸残基的存在,也是本领域技术人员熟知的。为避免 疑义,"该类肽形式"用来指代变体氨基酸残基,其中该α-碳取代基在该残 基的氮原子上,而不是在该α-碳上。用于制备处于该类肽形式的肽的方法是 本领域已知的,例如西蒙RJ(Simon RJ)等人,PNAS(1992)89(20),9367- 9371和豪威尔DC(Horwell DC),《生物技术趋势》(Trends Biotechnol.) (1995)13(4),132-134。

[0657] 除非另有说明,本发明的实践采用免疫学、生物化学、化学、分子生物 学、微生物学、细胞生物学、基因组学和重组DNA的常规技术,这些在本领 域的技能之内。参见萨姆布鲁克(Sambrook)、弗里奇(Fritsch)和马尼亚 蒂斯(Maniatis),《分子克隆:实验室手册》(MOLECULAR CLONING:A LABORATORY MANUAL),第2次编辑(1989);《当代分子生物学实验手册》(CURRENT PROTOCOLS IN MOLECULAR BIOLOGY)(F.M.奥苏 贝尔(F.M.Ausubel)等人编辑,(1987));《酶学方法》(METHODS IN ENZYMOLOGY)系列(学术出版公司):《PCR 2:实用方法》(PCR 2: A PRACTICAL APPROACH)(M.J.麦克弗森(M.J.MacPherson)、B.D.黑 姆斯(B.D.Hames)和G.R.泰勒(G.R.Taylor)编辑(1995))、哈洛 (Harlow)和拉内(Lane)编辑(1988)《抗体:实验室手册》 (ANTIBODIES,A LABORATORY MANUAL),以及《动物细胞培养》(ANIMAL CELL CULTURE)(R.I.弗雷谢尼(R.I.Freshney)编辑 (1987))。

[0658] 载体

[0659] 在一个方面,本发明提供了在CRISPR-Cas系统的工程化和优化中使用 的载体。

[0660] 如本文使用的,"载体"是一种允许或促进一个实体从一个环境转移到 另一个环境中的工具。它是一种复制子,如质粒、噬菌体、或粘粒,另一个 DNA片段可以插入其中,从而引起该插入的片段的复制。通常,当与适当的 控制元件关联时,一种载体能够复制。一般而言,术语"载体"是指一种核酸 分子,其能够运送与其连接的另一种核酸分子。载体包括但不限于,单链、 双链、或部分双链的核酸分子;包括一个或多个自由端、无自由端(例如环状的)的核酸分子;包括DNA、RNA、或两者的核酸分子;以及本领域已知 的其他多种多样的多核苷酸。一种类型的载体是"质粒",其是指其中可以 例如通过标准分子克隆技术插入另外的DNA片段的环状双链DNA环。另一 种类型的载体是病毒载体,其中病毒衍生的DNA或RNA序列存在于用于包 装病毒(例如,逆转录病毒、复制缺陷型逆转录病毒、腺病毒、复制缺陷型 腺病毒、以及腺相关病毒(AAV))的载体中。病毒载体还包含由用于转染 到一种宿主细胞中的病毒携带的多核苷酸。某些载体(例如,具有细菌复制 起点的细菌载体和附加型哺乳动物载体)能够在它们被导入的宿主细胞中自 主复制。其他载体(例如,非附加型哺乳动物载体)在引入宿主细胞后整合 到该宿主细胞的基因组中,并且由此与该宿主基因组一起

复制。而且,某些 载体能够指导它们可操作连接的基因的表达。这样的载体在此被称为“表达 载体”。在重组DNA技术中使用的普通表达载体通常是质粒形式。

[0661] 重组表达载体可包含处于适合于在宿主细胞中的核酸表达的形式的本发 明的核酸,这意味着这些重组表达载体包含基于待用于表达的宿主细胞而选 择的一种或多种调节元件,所述调节元件可操作地连接至待表达的核酸序 列。在重组表达载体内,“可操作地连接”旨在表示感兴趣的核苷酸序列以 一种允许该核苷酸序列的表达的方式被连接至该一种或多种调节元件(例 如,处于一种体外转录/翻译系统中或当该载体被引入到宿主细胞中时,处于 该宿主细胞中)。关于重组和克隆方法,提及2004年9月2日以US 2004-0171156 A1公开的美国专利申请10/815,730,该专利的内容通过引用以其全 文并入本文。

[0662] 本发明的多个方面涉及用于嵌合的RNA与Cas9的双顺反子载体。用于 嵌合的RNA与Cas9的双顺反子表达载体是优选的。一般而言并且特别地, 在这个实施例中,Cas9优选由CBh启动子驱动。嵌合RNA可以优选地由U6 启动子驱动。理想地,将这两者结合。嵌合的指导RNA典型地由20bp指导 序列(N)组成并且这可以接合到tracr序列上(从下链的第一个“U”到该转 录本的结尾)。该tracr序列可以在如指示的不同位置被截短。指导序列和tracr序列被该tracr配对序列隔开,该tracr配对序列可以是 GUUUUAGAGCUA。这之后可以是如所示的环序列GAAA。这两者都是优选 的实例。申请人已经通过SURVEYOR分析证明了在人EMX1和PVALB座位 处的Cas9介导的插入缺失。ChiRNA由其“+n”标志指示,并且crRNA是指指导和tracr序列被表达为分开的转录本的杂交体RNA。贯穿本申请,嵌合 RNA也可以被称为单指导、或合成指导RNA(sgRNA)。该环优选地是 GAAA,但是并并限于这个序列或者实际上在长度上仅仅为4bp。实际上, 用于在发夹结构中使用的优选环形成序列在长度上为四个核苷酸,并且最优 选地具有序列GAAA。然而,可以使用更长或更短的环序列,正如可替代的 序列。这些序列优选地包括三联体(例如,AAA)、和另外的核苷酸(例如 C或G)。环形成序列的实例包括CAAA和AAAG。

[0663] 术语“调节元件”旨在包括启动子、增强子、内部核糖体进入位点 (IRES)、和其他表达控制元件(例如转录终止信号,如多聚腺苷酸化信号 和多聚U序列)。这样的调节序列例如描述于戈德尔(Goeddel),《基因表 达技术:酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185,学术出版社(Academic Press),圣地亚哥(San Diego),加利福尼亚州(1990)中。调节元件包括指导一个核苷酸序列在许 多类型的宿主细胞中的组成型表达的那些序列以及指导该核苷酸序列只在某 些宿主细胞中表达的那些序列(例如,组织特异型调节序列)。组织特异型 启动子可主要指导在感兴趣的期望组织中的表达,所述组织例如肌肉、神经 元、骨、皮肤、血液、特定的器官(例如肝脏、胰腺)、或特殊的细胞类型 (例如淋巴细胞)。调节元件还可以时序依赖性方式(如以细胞周期依赖性 或发育阶段依赖性方式)指导表达,该方式可以是或者可以不是组织或细胞 类型特异性的。在一些实施例中,一个载体包含一个或多个pol III启动子(例如1、2、3、4、5、或更多个pol III启动子)、一个或多个pol II启动子 (例如1、2、3、4、5、或更多个pol II启动子)、一个或多个pol I启动子 (例如1、2、3、4、5、或更多个pol I启动子)、或其组合。pol III启动子的实例包括但不限于U6和H1启动子。pol II启动子的实例包括但不限于逆转录 劳斯肉瘤病毒(RSV)LTR启动子(任选地具有RSV增强子)、巨细胞病毒 (CMV)启动子(任选地具有CMV增强子)[参见,例如,波沙特 (Boshart)等人,《细胞》(Cell)41:521-530(1985)]、SV40启动

子、二 氢叶酸还原酶启动子、β-肌动蛋白启动子、磷酸甘油激酶(PGK)启动子、 和EF1α启动子。还被术语“调节元件”涵盖的是增强子元件,如WPRE; CMV增强子;在HTLV-I的LTR中的R-U5’片段《分子细胞生物学》(Mol. Cell.Biol.,第8(1)卷,第466-472页,1988);SV40增强子;以及在兔β- 珠蛋白的外显子2与3之间的内含子序列(《美国国家科学院院刊》(Proc. Natl.Acad.Sci.USA.),第78(3)卷,第1527-31页,1981)。本领域技术 人员将理解,表达载体的设计可取决于诸如待转化的宿主细胞的选择、所希 望的表达水平等因素。一种载体可以被引入到宿主细胞中而由此产生转录 物、蛋白质、或肽,包括由如本文所述的核酸编码的融合蛋白或肽(例如, 成簇的规律间隔的短回文重复(CRISPR)转录物、蛋白质、酶、其突变体形 式、其融合蛋白,等等)。关于调节序列,提及美国专利申请10/491,026,该 专利的内容通过引用以其全文并入本文。关于启动子,提及PCT公开WO 2011/028929和美国申请12/511,940,这些专利的内容通过引用以其全文并入 本文。

[0664] 载体可以被设计为用于在原核或真核细胞中表达CRISPR转录物(例如 核酸转录物、蛋白质、或酶)。例如,CRISPR转录物可表达于例如大肠杆菌 的细菌细胞、昆虫细胞(使用杆状病毒表达载体)、酵母细胞、或哺乳动物 细胞中。适合的宿主细胞进一步讨论于Goeddel(戈德尔),《基因表达技 术:酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185,学术出版社(Academic Press),圣地亚哥(San Diego),加利福尼亚州(Calif.)(1990年)中。可替代地,重组表达载体 可在体外例如使用T7启动子调节序列和T7聚合酶来转录和翻译。

[0665] 载体可以被引入原核生物或原核细胞中并且在其中增殖。在一些实施例 中,原核生物用来扩增有待引入真核细胞中的载体的多个拷贝,或者作为有 待引入真核细胞中的载体的产生中的中间载体(例如,扩增质粒作为病毒载 体包装系统的一部分)。在一些实施例中,原核生物用来扩增一个载体的多 个拷贝并且表达一种或多种核酸,如提供用于递送到宿主细胞或宿主生物中 的一种或多种蛋白质的来源。蛋白质在原核生物中的表达最经常在大肠杆菌 中用含有指导融合或非融合蛋白的表达的组成型或诱导型启动子的载体来进 行。融合载体将多个氨基酸添加到在其中编码的蛋白质上,如该重组蛋白的 氨基端上。这样的融合载体可以用于一个或多个目的,如:(i)增加重组蛋 白的表达;(ii)增加重组蛋白的溶解性;以及(iii)通过在亲和纯化中充当 配体来辅助重组蛋白的纯化。通常,在融合表达载体中,将蛋白切割位点引 入至融合部分与重组蛋白的接合处以使得能够在纯化融合蛋白之后将重组蛋 白与融合部分分离。这类酶以及它们的同源识别序列包括因子Xa、凝血酶以 及肠激酶。示例性融合表达载体包括pGEX(发玛西亚生物技术有限公司(Pharmacia Biotech Inc);史密斯和约翰逊,1988,《基因》(Gene)67: 31-40)、pMAL(纽英伦生物技术公司(New England Biolabs),贝弗利 (Beverly),马萨诸塞州(Mass.))以及pRIT5(发玛西亚公司,皮斯卡塔 韦(Piscataway),新泽西州(N.J.)),它们分别将谷胱甘肽S-转移酶 (GST)、麦芽糖E结合蛋白或蛋白A融合至靶重组蛋白。

[0666] 适合的诱导型非融合大肠杆菌表达载体的实例包括pTrc(Amrann(阿姆 兰)等人,(1988年)《基因》(Gene)69:301-315)以及pET 11d(Studier (斯图迪尔)等人,《基因表达技术:酶学方法》(GENE EXPRESSION TECHNOLOGY:METHODS IN ENZYMOLOGY)185,学术出版社(Academic Press),圣地亚哥(San Diego),加利福尼亚州(Calif.) (1990年)60-89。

[0667] 在一些实施例中,载体是酵母表达载体。用于在酵母酿酒中表达的载体 的实例包

括pYepSec1(巴尔戴利(Baldari)等人,1987,《欧洲分子生物学 学会杂志》(EMBO J)6:229-234)、pMFa(库尔坚(Kurjan)和赫斯库伍 特兹(Herskowitz),1982,《细胞》(Cell)30:933-943)、pJRY88(舒尔 茨(Schultz)等人,1987,《基因》(Gene)54:113-123)、pYES2(Invitrogen Corporation(英杰公司),San Diego(圣地亚哥),Calif(加利 福尼亚州))、以及picZ(Invitrogen Corp(英杰公司),San Diego(圣地亚 哥),Calif(加利福尼亚州))。

[0668] 在一些实施例中,使用杆状病毒载体,载体驱动昆虫细胞中的蛋白质表 达。可用于在培养的昆虫细胞(例如,SF9细胞)中表达蛋白质的杆状病毒 载体包括pAc系列(史密斯(Smith)等人,1983,《分子细胞生物学》 (Mol.Cell.Biol.)3:2156-2165)和pVL系列(拉克瑙(Lucklow)和萨莫斯 (Summers),1989,《病毒学》(Virology)170:31-39)。

[0669] 在一些实施例中,使用哺乳动物表达载体,载体能够驱动一种或多种序 列在哺乳动物细胞中表达。哺乳动物表达载体的实例包括pCDM8(锡德 (Seed),1987,《自然》(Nature)329:840)和pMT2PC(考夫曼 (Kaufman)等人,1987,《欧洲分子生物学学会杂志》(EMBO J.)6:187- 195)。当用于哺乳动物细胞中时,表达载体的控制功能典型地由一种或多种 调节元件提供。例如,常用的启动子来源于多瘤、腺病毒2、巨细胞病毒、猴 病毒40、以及本文所述和本领域已知的其他病毒。对于用于原核细胞和真核 细胞两者的其他适合的表达系统参见例如萨姆布鲁克(Sambrook)等人的 《分子克隆实验指南》(Molecular Cloning:A Laboratory Manual.)(第2 版)的第16和第17章,冷泉港实验室(Cold Spring Harbor Laboratory),冷 泉港实验室出版社(Cold Spring Harbor Laboratory Press),冷泉港(Cold Spring Harbor),纽约,1989。

[0670] 在一些实施例中,重组哺乳动物表达载体能够指导核酸优先在特定细胞 类型中表达(例如,使用组织特异型调节元件来表达核酸)。组织特异型调 节元件是本领域中已知的。适合的组织特异型启动子的非限制性实例包括白 蛋白启动子(肝脏特异性的;平克特(Pinkert)等人,1987.《基因发育》 (Genes Dev.)1:268-277),淋巴特异性启动子(卡里曼(Calame)和伊顿 (Eaton),1988.《免疫学进展》(Adv.Immunol)43:235-275),特别是T 细胞受体的启动子(维纳图(Winoto)和巴尔迪莫(Baltimore),1989.《欧 洲分子生物学学会杂志》(EMBO J.)8:729-733)和免疫球蛋白(巴纳吉 (Baneiji)等人,1983.《细胞》(Cell)33:729-740;奎恩(Queen)和巴尔 迪莫(Baltimore),1983.《细胞》(Cell)33:741-748),神经元特异性启动 子(例如,神经丝启动子;伯恩(Byrne)和鲁德尔(Ruddle),1989.《美国 科学院院刊》(Proc.Natl.Acad.Sci USA)86:5473-5477),胰腺特异性启动 子(艾德兰德(Edlund)等人,1985.《科学》(Science)230:912-916),以 及乳腺特异性启动子(例如,乳清启动子;美国专利号4,873,316和欧洲申请 公开号264,166)。还涵盖发育型调节启动子,例如,鼠科动物同源框蛋白 (hox)启动子凯赛尔((Kessel)和格鲁斯(Gruss),1990.《科学》(Science)249:374-379)和α-甲胎蛋白启动子(卡姆皮斯(Campes)和蒂尔 曼(Tilghman),1989.《基因发育》(Genes Dev.)3:537-546)。关于这些 原核和真核载体,提及美国专利6,750,059,该专利的内容通过引用以其全文 并入本文。本发明的其他实施例可涉及病毒载体的使用,关于它提及美国专 利申请13/092,085,该专利的内容通过引用以其全文并入本文。组织特异型调 节元件是本领域中已知的,并且在这个方面提及美国专利7,776,321,该专利 的内容通过引用以其全文并入本文。

[0671] 调节元件

[0672] 在一些实施例中,调节元件可操作地连接至CRISPR系统的一个或多个 元件,从而驱动该CRISPR系统的该一个或多个元件的表达。一般而言, CRISPR(规律间隔成簇短回文重复),也称为SPIDR(SPacer间隔开的同向 重复),构成通常对于特定细菌物种而言特异性的DNA基因座的家族。该 CRISPR座位包含在大肠杆菌中被识别的间隔开的短序列重复(SSR)的一个 不同类(石野(Ishino)等人,《细菌学杂志》(J.Bacteriol.),169:5429-5433;和中田(Nakata)等人,《细菌学杂志》(J.Bacteriol.),171:3553- 3556)、以及相关基因。类似的间隔开的SSR已经鉴定于地中海富盐菌 (Haloferax mediterranei)、化脓链球菌、鱼腥藻属、和结核分枝杆菌中(参 见,格鲁恩(Groenen)等人,《分子微生物学》(Mol.Microbiol.),10: 1057-1065;霍(Hoe)等人,《新发感染性疾病》(Emerg.Infect.Dis.),5: 254-263;马斯波尔(Masepohl)等人,《生物化学与生物物理学学报》 (Biochim.Biophys.Acta)1307:26-30;以及莫吉卡(Mojica)等人,《分子 微生物学》,17:85-93)。这些CRISPR座位典型地不同于其他SSR的重复 结构,这些重复已被称为规律间隔的短重复(SRSR)(詹森(Janssen)等 人,《OMICS:整合生物学杂志》(OMICS J.Integ.Biol.),6:23-33;以及 莫吉卡(Mojica)等人,《分子微生物学》(Mol.Microbiol.),36:244- 246)。一般而言,这些重复是以簇存在的短元件,其被具有基本上恒定长度 的独特间插序列规律地间隔开(莫吉卡(Mojica)等人,,同上)。虽然重 复序列在菌株之间是高度保守的,许多间隔开的重复和这些间隔区的序列一 般在菌株与菌株之间不同(冯·埃姆登(van Embden)等人,《细菌学杂志》 (J.Bacteriol.),182:2393-2401)。已经在40种以上的原核生物中鉴定出 CRISPR座位(参见,例如,詹森(Janssen)等人,《分子微生物学》(Mol. Microbiol.),43:1565-1575;以及莫吉卡(Mojica)等人,),包括但不限 于:气火菌属(Aeropyrum)、热棒菌属(Pyrobaculum)、硫化叶菌属 (Sulfolobus)、古球菌属(Archaeoglobus)、盐盒菌属(Halocarcula)、甲 烷杆菌属(Methanobacterium)、甲烷球菌属(Methanococcus)、甲烷八叠 球菌属(Methanosarcina)、甲烷火菌属(Methanopyrus)、焦球菌属 (Pyrococcus)、嗜酸菌属(Picrophilus)、热原体属(Thermoplasma)、棒 状杆菌属(Corynebacterium)、分枝杆菌属(Mycobacterium)、链霉菌属 (Streptomyces)、产水菌属(Aquifex)、紫单胞菌属(Porphyromonas)、 绿菌属(Chlorobium)、栖热菌属(Thermus)、芽孢杆菌属(Bacillus)、利 斯特菌属(Listeria)、葡萄球菌属(Staphylococcus)、梭菌属 (Clostridium)、好热厌氧杆菌属(Thermoanaerobacter)、支原体属 (Mycoplasma)、梭杆菌属(Fusobacterium)、固氮弓菌属(Azarcus)、色 杆菌属(Chromobacterium)、奈瑟菌属(Neisseria)、亚硝化单胞菌属(Nitrosomonas)、脱硫弧菌属(Desulfovibrio)、地杆菌属(Geobacter)、 粘球菌属(Myxococcus)、弯曲杆菌属(Campylobacter)、类杆菌属 (Wolinella)、不动杆菌属(Acinetobacter)、欧文菌属(Erwinia)、埃希菌 属(Escherichia)、军团杆菌属(Legionella)、甲基球菌属 (Methylococcus)、巴斯德菌属(Pasteurella)、发光细菌属(Photobacterium)、沙门菌属(Salmonella)、黄单胞菌属 (Xanthomonas)、耶尔森菌属(Yersinia)、密螺旋体属(Treponema)、和 栖热袍菌属(Thermotoga)。

[0673] 一般而言,“CRISPR系统”统称为转录物和涉及CRISPR相关 (“Cas”)基因的表达或指导其活性的其他元件,包括编码Cas基因的序 列、tracr(反式激活CRISPR)序列(例如tracrRNA或活性部分 tracrRNA)、tracr配对序列(涵盖“同向重复”和在内源CRISPR系统背

景 下的tracrRNA加工的部分同向重复)、指导序列(在内源CRISPR系统背景 下也称为“spacer”)、或来自CRISPR座位的其他序列和转录物。在本发明 的实施例中,术语指导序列和指导RNA可互换地使用。在一些实施例中, CRISPR系统的一个或多个元件来源于I型、II型、或III型CRISPR系统。在 一些实施例中,CRISPR系统的一个或多个元件来源于包含内源CRISPR系统 的特殊生物,如化脓链球菌。一般而言,CRISPR系统的特征为促进在靶序列 的位点处的CRISPR复合物(在内源CRISPR系统的背景下也称为原型间隔 子)的形成的元件。在CRISPR复合物形成的背景下,“靶序列”是指指导 序列被设计为对其具有互补性的序列,其中在靶序列与指导序列之间的杂交 促进CRISPR复合物的形成。一个靶序列可以包含任何多核苷酸,如DNA或 RNA多核苷酸。在一些实施例中,靶序列位于细胞的细胞核或细胞质中。

[0674] 在一些实施例中,可以通过计算机搜索重复基序来鉴定同向重复,其满 足下列标准的任一项或全部:

[0675] 1.发现一个在II型CRISPR座位侧翼的基因组序列的2Kb窗口;

[0676] 2.跨从20到50bp;以及

[0677] 3.以20到50bp间隔开。

[0678] 在一些实施例中,可以使用这些标准中的2条,例如1和2、2和3、或 1和3。在一些实施例中,可以使用所有这3条标准。

[0679] 在一些实施例中,随后可以通过满足下列标准的任一项或全部的序列来 预测候选tracrRNA:

[0680] 1.与同向重复同源的序列(在Geneious中搜索的具有高达18-bp错配的 基序);

[0681] 2.在转录方向上的预测的不依赖Rho的转录终止子的存在;以及

[0682] 3.在tracrRNA与同向重复之间的稳定发夹二级结构。

[0683] 在一些实施例中,可以使用这些标准中的2条,例如1和2、2和3、或 1和3。在一些实施例中,可以使用所有这3条标准。

[0684] 在一些实施例中,嵌合的合成指导RNA(sgRNA)设计可以在同向重复 与tracrRNA之间掺入至少12bp的双链体结构。

[0685] 在本发明的优选实施例中,该CRISPR系统是一个II型CRISPR系统, 并且该Cas酶是Cas9,其催化DNA切割。通过来源于化脓链球菌或Cas9或 任何密切相关的Cas9的酶促作用,产生了在靶位点序列处的双链断裂,所述 靶位点序列杂交到该指导序列的20个核苷酸上并且具有在该靶序列的20个 核苷酸之后的原型间隔子相邻基序(PAM)序列(实例包括NGG/NRG或可 以如本文所述进行确定的PAM)。经由Cas9的对于位点特异性DNA识别和 切割的CRISPR活性是由该指导序列、部分杂交到该指导序列上的tracr序列 以及该PAM序列定义的。该CRISPR系统的更多方面描述于卡吉诺夫 (Karginov)和汉农(Hannon)的“CRISPR系统:在细菌和古细菌中的小 RNA指导的防御”(The CRISPR system:small RNA-guided defence in bacteria and archaea),《分子细胞学杂志》(Mole Cell)2010,1月15日;37(1): 7。

[0686] II型CRISPR座位来自化脓链球菌SF370,该座位含有四个基因Cas9、Cas1、Cas2和Csn1的聚簇以及两个非编码RNA元件tracrRNA和由非重复序 列的短段(间隔子,每个约30bp)间隔开的重复序列(同向重复)的特征性 阵列。在此系统中,以四个连续步骤产生靶

向的DNA双链断裂(DSB)(图 2A)。第一步,从CRISPR座位转录两个非编码RNApre-crRNA阵列和 tracrRNA。第二步,将tracrRNA杂交到pre-crRNA的同向重复上,然后将其 加工成含有单独的间隔子序列的成熟crRNA。第三步,该成熟 crRNA:tracrRNA复合物经由在crRNA的间隔子区与原型间隔子DNA之间形 成异源双链体而将Cas9指向由原型间隔子和对应的PAM组成的DNA靶标。 最后,Cas9介导PAM上游的靶DNA的切割,以在原型间隔子内产生一个 DSB(图2A)。图2B证明了该密码子优化的Cas9的核定位。为了促进转录 精确起始,选择基于RNA聚合酶III的U6启动子,以驱动tracrRNA的表达 (图2C)。类似地,研发基于U6启动子的构建体,以表达包含侧翼为两个 同向重复(DR,还被术语“tracr配对序列”所涵盖;图2C)的单个间隔子 的pre-crRNA阵列。将起始间隔子设计成靶向人EMX1座位(图2C)中的 33-碱基对(bp)靶位点(满足Cas9的NGG识别基序的30-bp原型间隔子加 3-bp CRISPR基序(PAM)序列),该座位是大脑皮层的发育中的一个关键 基因。

[0687] 典型地,在内源CRISPR系统的背景下,CRISPR复合物(包含杂交到靶 序列上并且与一种或多种Cas蛋白复合的指导序列)的形成导致在该靶序列 中或其附近(例如在1、2、3、4、5、6、7、8、9、10、20、50、或更多个碱 基对之内)的一条链或两条链的切割。不希望受到理论的束缚,该tracr序列 (其可以包含或其组成为野生型tracr序列的全部或部分(例如野生型tracr序 列的约或多于约20、26、32、45、48、54、63、67、85个、或更多个核苷 酸))也可以形成CRISPR复合物的一部分,如通过沿着该tracr序列的至少 一部分杂交到与该指导序列可操作地连接的tracr配对序列的全部或部分上。 在一些实施例中,将驱动CRISPR系统的一个或多个元件的表达的一种或多 种载体引入到宿主细胞中,使得该CRISPR系统的这些元件的表达在一个或 多个靶位点指导CRISPR复合物的形成。例如,Cas酶、连接到tracr配对序 列上的指导序列、以及tracr序列可以各自可操作地连接到在分开的载体上的 分开的调节元件上。可替代地,从相同或不同调节元件表达的二个或更多个 元件可以结合在单个载体中,其中提供该CRISPR系统的任何组分的一个或 多个另外的载体不包括在第一载体中。结合在单个载体中的CRISPR系统元 件可以按照任何适合的取向排列,如一个元件相对于第二元件位于5’(“上 游”)或3’(“下游”)。一个元件的编码序列可以位于第二元件的编码序 列的同一条链或相对链上,并且取向为相同或相对的方向。在一些实施例中,单个启动子驱动编码CRISPR酶的转录物以及该指导序列、tracr配对序 列(任选地可操作地连接到该指导序列上)、和嵌入在一个或多个内含子序 列之内(例如,各自在不同的内含子中,两个或更多个在至少一个内含子 中,或者全部都在单个内含子中)的tracr序列中的一者或多者的表达。在一 些实施例中,该CRISPR酶、指导序列、tracr配对序列、和tracr序列可操作 地连接到相同的启动子上并且从其表达。

[0688] 在一些实施例中,一个载体包含一个或多个插入位点,如限制性核酸内 切酶识别序列(也称为“克隆位点”)。在一些实施例中,一个或多个插入 位点(例如,约或多于约1、2、3、4、5、6、7、8、9、10个、或更多个插 入位点)位于一种或多种载体的一个或多个序列元件的上游和/或下游。在一 些实施例中,一个载体包含在tracr配对序列的上游、并且任选地在可操作地 连接到该tracr配对序列上的调节元件的下游的插入位点,使得在将指导序列插入到该插入位点中之后并且在表达时该指导序列指导CRISPR复合物与真 核细胞中的靶序列的序列特异性结合。在一些实施例中,一个载体包含一个 或多个插入位点,每个插入位点位于两个tracr配对序列之间,从而允许在每 个位点插入指导序列。在这样一种安排

中,两个或更多个指导序列可以包含 两个或更多个拷贝的单指导序列、两个或更多个不同的指导序列、或这些的 组合。当使用多个不同的指导序列时,可以使用单个表达构建体使CRISPR 活性靶向到细胞内的多个不同的相应靶序列。例如,单个载体可以包含约或 多于约1、2、3、4、5、6、7、8、9、10、15、20个、或更多个指导序列。 在一些实施例中,可以提供约或多于约1、2、3、4、5、6、7、8、9、10 个、或更多个这样的含有靶序列的载体,并且任选地将其递送到细胞中。

[0689] 在一些实施例中,一个载体包含可操作地连接到编码CRISPR酶(如Cas 蛋白)的酶编码序列上的调节元件。Cas蛋白的非限制性实例包括:Cas1、 Cas1B、Cas2、Cas3、Cas4、Cas5、Cas6、Cas7、Cas8,Cas9(也称为Csn1 和Csx12)、Cas10、Csy1、Csy2、Csy3、Cse1、Cse2、Csc1、Csc2、Csa5、 Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、 Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、 Csx1、Csx15、Csf1、Csf2、Csf3、Csf4、其同源物、或其修饰形式。在一些 实施例中,未修饰的CRISPR酶如Cas9具有DNA切割活性。在一些实施例 中,该CRISPR酶指导在靶序列位置处(例如在该靶序列之内和/或在该靶序列的互补物之内)的一条链或两条链的切割。在一些实施例中,该CRISPR 酶指导距离靶序列的第一个或最后一个核苷酸约1、2、3、4、5、6、7、8、 9、10、15、20、25、50、100、200、500个、或更多个碱基对之内的一条链 或两条链的切割。在一些实施例中,一个载体编码相对于相应的野生型酶被 突变的CRISPR酶,使得该突变的CRISPR酶缺乏切割含有一个靶序列的靶多 核苷酸的一条链和两条链的能力。例如,在来自化脓链球菌的Cas9的RuvC I 催化结构域中的天冬氨酸到丙氨酸取代(D10A)将Cas9从切割两条链的核 酸酶转化成切口酶(切割单条链)。致使Cas9为切口酶的其他突变实例包括 而不限于H840A、N854A、和N863A。作为一个另外的实例,Cas9的两个或 更多个催化结构域(RuvC I、RuvC II、和RuvC III或该HNH结构域)可以 被突变为产生实质上缺乏所有DNA切割活性的突变的Cas9。在一些实施例 中,D10A突变与H840A、N854A、或N863A突变的一者或多者相组合,以 便产生实质上缺乏所有DNA切割活性的Cas9酶。在一些实施例中,当该突 变的CRISPR酶的DNA切割活性比它的非突变形式低约25%、10%、5%、 1%、0.1%、0.01%、或更少时,则该酶被考虑为实质上缺乏所有DNA切割活 性。在该酶不是SpCas9的情况下,可以在相应于SpCas9的位置10、762、 840、854、863和/或986的任何或所有残基处进行突变(其可以例如通过标 准序列比较工具进行确定)。具体地说,在SpCas9中的任何或所有下列突变 是优选的:D10A、E762A、H840A、N854A、N863A和/或D986A;并且还设 想了这些置换氨基酸的任何一种的保守性取代。在其他Cas9中的相应位置处 的相同取代(或这些突变的保守性取代)也是优选的。特别优选的是在 SpCas9中的D10和H840。然而,在其他Cas9中,相应于SpCas9 D10和 H840的残基也是优选的。

[0690] 另外,将SpCas9的RuvC I催化结构域中的天冬氨酸至丙氨酸取代 (D10A)工程化,以将核酸酶转化为切口酶(SpCas9n)(参见例如萨普拉 诺萨克斯(Sapranauskas)等人,2011,《核酸研究》(Nucleic Acis Research),39:9275;加索纳斯(Gasiunas)等人,2012,《美国国家科学 院院刊》(Proc.Natl.Acad.Sci.USA),109:E2579),这样使得带切口的基因组DNA经历高保真同源定向修复(HDR)。Surveyor测定确认到,SpCas9n 在EMX1原型间隔子靶标处不产生插入缺失。靶向EMX1的嵌合crRNA(同 样具有tracrRNA组分)与SpCas9的共表达在靶位点中产生插入缺失,而与 SpCas9n的共表达则不然(n=3)。此外,327个扩增子的

测序未检测到由 SpCas9n诱导的任何插入缺失。选择相同的座位,以通过用靶向EMX1、hSpCas9或hSpCas9n的嵌合的RNA以及用于在原型间隔子附近引入一对限 制酶切位点(HindIII和NheI)的HR模板共转染HEK 293FT细胞而测试 CRISPR介导的HR。

[0691] 本文描述了优选的直向同源物。一种Cas酶可以被鉴定为Cas9,因为这 可以是指与来自II型CRISPR系统的具有多个核酸酶结构域的最大核酸酶共 享同源性的酶的通用类别。最优选地,该Cas9酶来自或衍生自spCas9或 saCas9。关于衍生,申请人意指该衍生的酶在很大程度上是基于与野生型酶 具有高度序列同源性的含义,但是如本文所述它在某些方面已经被突变(修 饰)。

[0692] 应当理解的是,术语Cas和CRISPR酶在本文中通常可互换地使用,除 非另外说明。如上提及的,在本文中使用的许多残基编号是指来自化脓链球 菌中的II型CRISPR位点的Cas9酶。然而,应当理解的是,本发明包括更多 的来自其他微生物物种的Cas9s,如SpCas9、SaCa9、St1Cas9等等。

[0693] 密码子优化

[0694] 本文中提供了在这种情形下针对人类进行优化(即,针对在人类中表达 进行优化)的密码子优化序列的实例,参见SaCas9人类密码子优化序列。在 这被优化的同时,应当理解其他实例是可能的并且针对宿主物种的密码子优 化是已知的。

[0695] 在一些实施例中,编码CRISPR酶的酶编码序列经密码子优化,以便在 特定的细胞如真核细胞中表达。这些真核细胞可以是特定生物的那些或来源 于特定生物,如哺乳动物,包括但不限于人、小鼠、大鼠、兔、狗、或非人 类哺乳动物或灵长动物。在一些实施例中,对于人类或动物而言很可能使得 他们(它们)受苦而没有任何实质性医学益处的用于修饰人类的种系遗传同 一性的方法和/或用于修饰动物的遗传同一性的方法、以及还有作为这样的方 法的结果的动物,可以被排除在外。

[0696] 一般而言,密码子优化是指通过用在宿主细胞的基因中更频繁地或者最 频繁地使用的密码子代替天然序列的至少一个密码子(例如约或多于约1、 2、3、4、5、10、15、20、25、50个、或更多个密码子同时维持该天然氨基 酸序列而修饰一个核酸序列以便增强在感兴趣宿主细胞中的表达的方法。不 同的物种对于具有特定氨基酸的某些密码子展示出特定的偏好。密码子偏好 性(在生物之间的密码子使用的差异)经常与信使RNA(mRNA)的翻译效 率相关,而该翻译效率则被认为依赖于(除其他之外)被翻译的密码子的性 质和特定的转运RNA(tRNA)分子的可用性。细胞内选定的tRNA的优势一 般反映了最频繁用于肽合成的密码子。因此,可以将基因定制为基于密码子 优化在给定生物中的最佳基因表达。密码子利用率表可以容易地获得,例如 在www.kazusa.orjp/codon/(2002年7月9日访问)上可获得的密码子使用数 据库(“Codon Usage Database”)中,并且这些表可以通过不同的方式调整适 用。参见,中村Y.(Nakamura Y.)等人,“从国际DNA序列数据库中制表 的密码子使用:2000年的状态”(Codon usage tabulated from the international DNA sequence databases:status for the year 2000)”《核酸研究》(Nucl.Acids Res.)28:292(2000年)。用于密码子优化特定的数列以便在特定的宿主细 胞中表达的计算机算法也是可得的,如基因制造(Gene Forge)(Aptagen公 司;雅各布斯(Jacobus),PA),也是可得的。在一些实施例中,在编码 CRISPR酶的序列中的一个或多个密码子(例如1、2、3、4、5、10、15、 20、25、50个、或更多个、或所有密码子)相应于对于特定氨基酸最频繁使 用的密码子。

[0697] 核定位序列(NLS)

[0698] 在一些实施例中,一种载体编码包含一个或多个核定位序列(NLS)如 约或多于约1、2、3、4、5、6、7、8、9、10个、或更多个NLS的CRISPR 酶。在一些实施例中,该CRISPR酶包含在或接近于氨基端的约或多于约1、 2、3、4、5、6、7、8、9、10个、或更多个NLS,在或接近于羧基端约或多 于约1、2、3、4、5、6、7、8、9、10个、或更多个NLS,或这些的组合 (例如在氨基端的一个或多个NLS以及在羧基端的一个或多个NLS)。当存 在多于一个NLS时,每一个可以被选择为不依赖于其他NLS,使得在多于一 个拷贝中可以存在单个NLS和/或与在一个或多个拷贝中存在一个或多个其他 NLS相组合。在本发明的一个优选实施例中,该CRISPR酶包含至多6个 NLS。在一些实施例中,当NLS的最近的氨基酸是在从N端或C端沿着该多 肽链的约1、2、3、4、5、10、15、20、25、30、40、50个、或更多个氨基 酸之内时,NLS可以被视为接近该N端或C端。NLS的非限制性实例包括来 源于以下项的NLS序列:SV40病毒大T抗原的NLS,其具有氨基酸序列 PKKKRKV;来自核质蛋白的NLS(例如,具有序列 KRPAATKKAGQAKKKK的核质蛋白二分NLS);c-myc NLS,其具有氨基 酸序列PAAKRVKLD或RQRRNELKRSP;hRNPA1 M9 NLS,其具有序列 NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY;来自输入蛋白-α 的IBB结构域的序列RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV;肌瘤T蛋 白的序列VSRKRPRP和PPKKARED;人p53的序列POPKKKPL;小鼠c-abl IV的序列SALIKKKKKMAP;流感病毒NS1的序列DRLRR和PKQKKRK; 肝炎病毒δ抗原的序列RKLKKKIKKL;小鼠Mx1蛋白的序列REKKKFLKRR;人聚(ADP-核糖)聚合酶的序列 KRKGDEVDGVDEVAKKKSKK;以及类固醇激素受体(人)糖皮质激素的 序列RKCLQAGMNLEARKTKK。

[0699] 一般而言,该一个或多个NLS具有足够的强度,以便在真核细胞的核中 驱动该CRISPR复合物以可检测到的量积聚。一般而言,核定位活性的强度 可以驱动在该CRISPR酶中的NLS的数目、所使用的一个或多个特定的 NLS、或这些因素的组合。可以通过任何适合的技术进行细胞核中积聚的检 测。例如,一种检测标记可以融合到CRISPR酶上,使得细胞内的位置可以 被可视化,如与检测细胞核的位置的手段(例如,对于细胞核特异的染料,如DAPI)相结合。还可以将细胞核从细胞中分离出来,然后可以通过任何适 合的用于检测蛋白质的方法分析其内容物,如免疫组织化学、蛋白质印迹或 酶活性测定。如通过测定CRISPR复合物形成的作用(例如,测定在靶序列 处的DNA切割或突变、或测定由于CRISPR复合物形成和/或CRISPR酶活性 的影响而改变的基因表达活性),与没有暴露于CRISPR酶或复合物、或暴 露于缺乏该一个或多个NLS的CRISPR酶的对照相比较,还可以间接地确定 细胞核中的积聚。

[0700] 指导序列

[0701] 一般而言,指导序列是与靶多核苷酸序列具有足够互补性以便与该靶序 列杂交并且指导CRISPR复合物与该靶序列的序列特异性结合的任何多核苷 酸序列。在一些实施例中,当使用适合的比对算法进行最佳比对是,在指导 序列与其相应的靶序列之间的互补程度是约或多于约50%、60%、75%、 80%、85%、90%、95%、97.5%、99%、或更多。可以使用用于比对序列的任 何适合的算法来确定最佳比对,其非限制性实例包括史密斯-沃特曼(Smith- Waterman)算法、尼德曼-翁施(Needleman-Wunsch)算法、基于伯罗斯-惠 勒变换(Burrows-Wheeler Transform)的算法(例如伯罗斯-惠勒比对工具 (Burrows Wheeler Aligner))、ClustalW、Clustal X、BLAT、Novoalign (Novocraft技术公司;在

www.novocraft.com可获得)、ELAND(亿明达公 司(Illumina),圣地亚哥,加利福尼亚州)、SOAP(在 soap.genomics.org.cn可获得)、以及Maq(在maq.sourceforge.net可获得)。在一些实施例中,指导序列在长度上为约或多于约5、10、11、12、13、14、 15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、 35、40、45、50、75个、或更多个核苷酸。在一些实施例中,指导序列在长 度上为少于约75、50、45、40、35、30、25、20、15、12个、或更少的核苷 酸。可以通过任何适合的测定法来评估指导序列指导CRISPR复合物与靶序 列的序列特异性结合的能力。例如,足以形成CRISPR复合物的CRISPR系统 的组分,包括有待测试的指导序列在内,可以例如通过用编码该CRISPR序 列的组分的载体进行转染而被提供到具有相应靶序列的宿主细胞中,随后通 过如本文所述的测量员测定来评估在该靶序列之内的优先切割。类似地,通 过提供该靶序列、包括有待测试的指导序列在内的CRISPR复合物的组分、和不同于该测试指导序列的对照指导序列,并且比较在该测试指导序列与该 对照指导序列反应之间的靶序列处的结合或切割率,可以在一个试管中评估 靶多核苷酸序列的切割。其他测定法是可能的,并且将由本领域技术人员想 到。

[0702] 指导序列可以被选择为靶向任何靶序列。在一些实施例中,该靶序列是 在细胞的基因组内的序列。示例性靶序列包括在靶基因组中为独特的那些。例如,对于化脓链球菌Cas9,在基因组中的独特靶序列可以包括形式 MMMMMMMMNNNNNNNNNNNNXGG的Cas9靶位点,其中 NNNNNNNNNNNNXGG(N是A、G、T、或C;并且X可以是任何物)在 该基因组中单次出现。在基因组中的独特靶序列可以包括形式 MMMMMMMMMNNNNNNNNNNNNXGG的化脓链球菌Cas9靶位点,其中 NNNNNNNNNNNXGG(N是A、G、T、或C;并且X可以是任何物)在该 基因组中单次出现。对于嗜热链球菌CRISPR1 Cas9,在基因组中的独特靶序 列可以包括形式MMMMMMMMNNNNNNNNNNNNXXAGAAW的Cas9靶 位点,其中NNNNNNNNNNNNXXAGAAW(N是A、G、T、或C;X可以 是任何物;并且W是A或T)在该基因组中单次出现。在基因组中的独特靶 序列可以包括形式MMMMMMMMMNNNNNNNNNNNNXXAGAAW的嗜热链 球菌CRISPR1 Cas9靶位点,其中NNNNNNNNNNNXXAGAAW(N是A、 G、T、或C;X可以是任何物;并且W是A或T)在该基因组中单次出现。对于化脓链球菌Cas9,在基因组中的独特靶序列可以包括形式MMMMMMMMNNNNNNNNNNNNXGGXG的Cas9靶位点,其中 NNNNNNNNNNNNXGGXG(N是A、G、T、或C;并且X可以是任何物) 在该基因组中单次出现。在基因组中的独特靶序列可以包括形式MMMMMMMMMNNNNNNNNNNNNXGGXG的化脓链球菌Cas9靶位点,其 中NNNNNNNNNNNXGGXG(N是A、G、T、或C;并且X可以是任何 物)在该基因组中单次出现。在这些序列的每一个中,“M”可以是A、G、 T、或C,并且在序列鉴定中不必考虑为是独特的。

[0703] 在一些实施例中,指导序列被选择为降低在该指导序列内的二级结构程 度。在一些实施例中,在最佳地折叠时,该指导序列的约或少于约75%、 50%、40%、30%、25%、20%、15%、10%、5%、1%、或更少的核苷酸参与 自我互补碱基配对。可以通过任何适合的多核苷酸折叠算法来确定最佳折 叠。一些算法是基于计算最小吉布斯(Gibbs)自由能。一种这样的算法的实 例是mFold,正如祖克(Zuker)和施蒂格勒(Stiegler)《核酸研究》(Nucleic Acids Res.)9(1981),133-148)所描述。折叠算法的另一个实例 是使用由维也纳大学(University of Vienna)的理论化学研究所(Institute for Theoretical Chemistry)研发的在线网络服务器RNAfold(参见例如A.R.格鲁 伯(A.R.Gruber)等人,2008,《细胞》(Cell)106(1):23-24;以及PA卡 尔(PA Carr)和GM丘奇(GM Church),2009,

《自然生物技术》(Nature Biotechnology)27(12):1151-62)。

[0704] tracr配对序列

[0705] 一般而言,tracr配对序列包括与tracr序列具有足够互补性以促进下列一 项或多项的任何序列:(1)在含有相应tracr序列的细胞中的侧翼为tracr配 对序列的指导序列的切除;以及(2)在靶序列处的CRISPR复合物的形成, 其中该CRISPR复合物包含杂交到tracr序列上的tracr配对序列。通常,互补 程度是就tracr配对序列与tracr序列沿着这两个序列的较短者的长度的最佳比 对而言。可以通过任何适合的比对算法来确定最佳比对,并且可以可以进一 步对二级结构做出解释,如在该tracr序列或tracr配对序列之内的自我互补 性。在一些实施例中,在进行最佳比对时,在该tracr序列与tracr配对序列之 间沿着这两者的较短者的长度的互补程度是约或多于约25%、30%、40%、 50%、60%、70%、80%、90%、95%、97.5%、99%、或更高。在一些实施例 中,该tracr序列在长度上为约或多于约5、6、7、8、9、10、11、12、13、 14、15、16、17、18、19、20、25、30、40、50个、或更多个核苷酸。在一 些实施例中,该tracr序列和tracr配对序列被包含在单个转录物中,使得在这 两者之间的杂交产生具有二级结构(如发夹)的转录物。在本发明的一个实 施例中,该转录物或转录的多核苷酸序列具有至少两个或更多个发夹。在优 选的实施例中,该转录物具有两个、三个、四个或五个发夹。在本发明的一 个另外的实施例中,该转录物具有至多五个发夹。在一个发夹结构中,在该 环的最终"N"和上游的序列5'的部分相应于该tracr配对序列,并且该环的 序列3'的部分相应于该tracr序列。另外的包含指导序列、tracr配对序列、和 tracr序列的单个多核苷酸的非限制性实例如下(列出为5'到3'),其中 "N"代表指导序列的碱基,小写字体的第一区代表tracr配对序列,且小写 字体的第二区代表tracr序列,并且该最后的多聚T序列代表转录终止子: (1) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctaca aagataa ggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT;(2) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggctt catgccg aaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT;(3) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg aaatcaacaccctgtcattttatggcagggtgtTTTTTT;(4) NNNNNNNNNNNNNNNNNNNgtttt a g a g c t a G A A A t a g c a a g t t a a a a t a a g g c t a g t c c g t t a t c a a c t t gaaaaagtggcaccgagtcggtgcTTTTTT;(5) NNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccg ttatcaac ttgaaaaagtgTTTTTTT;和(6) NNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgt tatcaTT TTTTTT。在一些实施例中,序列(1)到(3)与来自嗜热链球菌 CRISPR1的Cas9结合使用。在一些实施例中,序列(4)到(6)与来自化脓 链球菌CRISPR1的Cas9结合使用。在一些实施例中,该tracr序列是一个与 包含该tracr配对序列的转录物分开的转录物。

[0706] 重组模板

[0707] 在一些实施例中,还提供了重组模板。重组模板可以是如本文所述的另 一个载体的组分,其被包含在一个分开的载体中,或者被提供为一个分开的 多核苷酸。在一些实施例中,重组模板被设计为用作在同源重组中的模板, 如在被作为CRISPR复合物的一部分的CRISPR酶切开或切割的靶序列之内或 在其附近。模板多核苷酸可以具有任何适合的长度,如约或多于约10、15、 20、25、50、75、100、150、200、500、1000个、或更多个核苷酸长度。在

一些实施例中，该模板多核苷酸与包含该靶序列的多核苷酸的一部分互补。当进行最佳比对时，模板多核苷酸可能与靶序列的一个或多个核苷酸(例如 约或多于约1、5、10、15、20个、或更多个核苷酸)重叠。在一些实施例 中，当一个模板序列与包含靶序列的多核苷酸进行最佳比对时，该模板多核 苷酸的最近的核苷酸在距离该靶序列的约1、5、10、15、20、25、50、75、 100、200、300、400、500、1000、5000、10000个、或更多个核苷酸之内。

[0708] 融合蛋白

[0709] 在一些实施例中，该CRISPR酶是包含一个或多个异源蛋白结构域(例 如除了该CRISPR酶之外的约或多于约1、2、3、4、5、6、7、8、9、10 个、或更多个结构域)的融合蛋白的一部分。CRISPR酶融合蛋白可以包含任 何其他蛋白质，以及任选地在任何两个结构域之间的连接序列。可以融合到 CRISPR酶上的蛋白质结构域的实例包括但不限于，表位标签、报告基因序 列、以及具有下列活性的一者或多者的蛋白质结构域：甲基酶活性、脱甲基 酶活性、转录激活活性、转录阻抑活性、转录释放因子活性、组蛋白修饰活 性、RNA切割活性和核酸结合活性。表位标签的非限制性实例包括组氨酸 (His)标签、V5标签、FLAG标签、流感病毒血凝素(HA)标签、Myc标 签、VSV-G标签、和硫氧还蛋白(Trx)标签。报告基因的实例包括，但不限 于，谷胱甘肽-S-转移酶(GST)、辣根过氧化物酶(HRP)、氯霉素乙酰转 移酶(CAT)、β-半乳糖苷酶、β-葡糖醛酸糖苷酶、萤光素酶、绿色荧光蛋白 (GFP)、HcRed、DsRed、青荧光蛋白(CFP)、黄色荧光蛋白(YFP)、 以包括蓝色荧光蛋白(BFP)的自发荧光蛋白。CRISPR酶可以融合到编码一 种蛋白质或蛋白质片段的基因序列上，所述蛋白质或蛋白质片段结合DNA分 子或结合其他细胞分子，其包括，但不限于，麦芽糖结合蛋白(MBP)、S- tag、Lex A DNA结合结构域(DBD)融合物、GAL4 DNA结合结构域融合 物、以及单纯疱疹病毒(HSV)BP16蛋白融合物。可以形成包含CRISPR酶 的融合蛋白的一部分的另外的结构域描述于US 20110059502中，通过引用将 其并入本文。在一些实施例中，使用标记的CRISPR酶来鉴定靶序列的位 置。

[0710] 可诱导系统

[0711] 在一些实施例中，CRISPR酶可以形成可诱导系统的组分。该系统的诱导 性质将允许该系统利用一种能量形式对基因编辑或基因表达进行时空控制。 该能量形式可以包括但不限于，电磁辐射、声能、化学能和热能。可诱导系 统的实例包括四环素诱导型启动子(Tet-On或Tet-Off)、小分子双杂交转录 激活系统(FKBP、ABA等))或光可诱导系统(光敏色素、LOV结构域或隐 花色素)。在一个实施例中，该CRISPR酶可以是以序列特异性方式指导转 录活性的变化的光诱导的转录效应因子(LITE)的一部分。光的组分可以包 括CRISPR酶、光响应细胞色素异二聚体(例如，来自拟南芥)、以及转录 激活/抑制结构域。诱导型DNA结合蛋白和关于使用它们的方法的其他实例 提供在US 61/736465和US 61/721,283中，特此通过引用以其全文并入。

[0712] 递送

[0713] 在一些方面，本发明提供了以下方法，这些方法包括向宿主细胞递送一 种或多种多核苷酸，如或如在此描述的一种或多种载体、其一种或多种转录 本和/或一种或自其转录的蛋白。在一些方面，本发明进一步提供了通过这样 的方法产生的细胞以及包括这样的细胞或由这样的细胞产生的动物。在一些 实施例中，将与(并且任选地与其复合)指导序列组合的CRISPR酶递送至 细胞。可以使用常规的病毒和非病毒基的基因转移方法将核酸引

入哺乳动物 细胞或靶组织中。可以使用这样的方法向培养物中或宿主生物中的细胞给予编码CRISPR系统的组分的核酸。非病毒载体递送系统包括DNA质粒、RNA (例如在此描述的载体的转录本)、裸核酸以及与递送赋形剂(如脂质体) 复合的核酸。病毒载体递送系统包括DNA和RNA病毒,在被递送至细胞后 它们具有游离型或整合型基因组。关于基因疗法程序的综述,参见安德森 (Anderson),《科学》(Science)256:808-813(1992);纳贝尔(Nabel) & 费尔格纳(Felgner),TIBTECH 11:211-217(1993);三谷(Mitani)& 卡斯基(Caskey),TIBTECH 11:162-166(1993);狄龙(Dillon), TIBTECH 11:167-175(1993);米勒(Miller),《自然》(Nature)357: 455-460(1992);范·布朗特(Van Brunt),《生物技术》(Biotechnology) 6(10):1149-1154(1988);维涅(Vigne),《恢复神经学和神经科学》(Restorative Neurology and Neuroscience)8:35-36(1995);克雷默 (Kremer)&佩里科德特(Perricaudet),《英国医学公报》(British Medical Bulletin)51(1):31-44(1995);哈塔塔(Haddada)等人,在《微生 物学和免疫学当前主题》(Current Topics in Microbiology and Immunology) 中多尔夫勒(Doerfler)和博姆(Böhm)(编辑)(1995);以及余(Yu) 等人,《基因疗法》(Gene Therapy)1:13-26(1994)。

[0714] 核酸的非病毒递送方法包括脂转染、显微注射、基因枪、病毒颗粒、脂 质体、免疫脂质体、聚阳离子或脂质:核酸共轭物、裸DNA、人工病毒体以 及DNA的试剂增强的摄取。脂转染描述于例如美国专利号5,049,386、4,946,787和4,897,355中并且脂转染试剂是市售的(例如,Transfectam™和 Lipofectin™)。适于多核苷酸的有效的受体识别脂转染的阳离子和中性脂质 包括Felgner(费尔格纳),WO 91/17424;WO 91/16024的那些。递送可以 针对细胞(例如体外或离体给予)或靶组织(例如体内给予)。

[0715] 脂质:核酸复合物(包括靶向的脂质体,如免疫脂质复合物)的制备是 本领域的技术人员熟知的(参见例如,克丽丝特尔(Crystal),《科学》 (Science)270:404-410(1995);布莱泽(Blaese)等人,《癌症基因疗 法》(Cancer Gene Ther.)2:291-297(1995);贝尔(Behr)等人,《生物 共轭化学》(Bioconjugate Chem.)5:382-389(1994);雷米(Remy)等人,《生物共轭化学》5:647-654(1994);高(Gao)等人,《基因疗法》 (Gene Therapy)2:710-722(1995);艾哈迈德(Ahmad)等人,《癌症研 究》(Cancer Res.)52:4817-4820(1992);美国专利号4,186,183、4,217,344、4,235,871、4,261,975、4,485,054、4,501,728、4,774,085、4,837,028以及4,946,787)。

[0716] 使用RNA或DNA病毒基的系统递送核酸利用将病毒靶向体内的特定细 胞并将病毒有效负荷(payload)运至细胞核中的高度进化的过程。可以将病 毒载体直接给予至患者(体内)或可以使用它们在体外处理细胞,并且任选 地,可以将修饰的细胞给予至患者(离体)。常规的病毒基的系统可以包括 用于基因转移的逆转录病毒载体、慢病毒载体、腺病毒载体、腺相关病毒载 体以及单纯疱疹病毒载体。用逆转录病毒、慢病毒和腺相关病毒基因转移方 法整合进宿主基因组中是可能的,这通常导致插入转基因的长期表达。另 外,已经在不同细胞类型和靶组织中观察到高转导效率。

[0717] 可以通过掺入外源包膜蛋白,扩展靶细胞的潜在靶群而改变逆转录病毒 的向性。慢病毒载体是能够转导或感染非分裂细胞并典型地产生较高病毒效 价的逆转录病毒载体。因此,逆转录病毒基因转移系统的选择将依赖于靶组 织。逆转录病毒载体由顺式作用长末端重复组成,这些长末端重复具有包装 多达6-10kb的外源序列的能力。最低量的顺式

作用LTR对于载体的复制和 包装而言是足够的,然后使用这些载体将治疗基因整合进靶细胞中,以提供 永久的转基因表达。广泛使用的逆转录病毒载体包括基于鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猴免疫缺陷病毒(SIV)、人类 免疫缺陷病毒(HIV)及其组合的那些(参见例如,布赫谢尔(Buchscher) 等人,《病毒学杂志》(J.Virol.)66:2731-2739(1992);约翰(Johann) 等人,《病毒学杂志》66:1635-1640(1992);佐姆内尔费尔特(Sommnerfelt)等人,《病毒学》(Virol.)176:58-59(1990);威尔逊 (Wilson)等人,《病毒学杂志》63:2374-2378(1989);米勒(Miller)等 人,《病毒学杂志》65:2220-2224(1991);PCT/US 94/05700)。

[0718] 在另一个实施例中,考虑了科卡尔(Cocal)水泡性病毒包膜假型逆转录 病毒载体粒子(参见例如,转让给弗雷德·哈钦森癌症研究中心(Fred Hutchinson Cancer Research Center)的美国专利公开号20120164118)。科卡 尔病毒属于水泡性病毒属,并且是一种哺乳动物水泡性口炎病原体。科卡尔 病毒最初分离自特立尼达拉岛(Trinidad)的螨(琼克斯(Jonkers)等人,《美国兽医研究杂志》(Am.J.Vet.Res.)25:236-242(1964)),并且已经 在特立尼达拉岛、巴西和阿根廷从昆虫、牛和马体内鉴定出感染。已经从天 然地感染的节肢动物体内分离出感染哺乳动物的水泡性病毒中的许多,表明 它们是媒介传播的。水疱性病毒抗体在居住于这些病毒是地方性的且是实验 室获得的农村地区的人中间是常见的;人类感染通常导致流感样症状。科卡 尔病毒包膜糖蛋白在氨基酸水平上与印第安纳VSV-G(VSV-G Indiana)具有 71.5%一致性,并且水疱性病毒的包膜基因的系统发育比较显示科卡尔病毒在 血清学上与印第安纳VSV-G株不同,但是在水疱性病毒中间与其最密切相 关。琼克斯(Jonkers)等人,《美国兽医研究杂志》(Am.J.Vet.Res.)25: 236-242(1964)和特拉瓦索德罗萨(Travassos da Rosa)等人,《美国热带 药学与卫生学杂志》(Am.J.Tropical Med.&Hygiene)33:999-1006 (1984)。科卡尔水疱性病毒包膜假型逆转录病毒载体粒子可以包括例如慢 病毒、α逆转录病毒、β逆转录病毒、γ逆转录病毒、δ逆转录病毒以及ε逆转 录病毒载体粒子,这些粒子可以包括逆转录病毒Gag、Pol和/或一种或多种辅 助蛋白以及科卡尔水疱性病毒包膜蛋白。在这些实施例的某些方面内,Gag、Pol和辅助蛋白是慢病毒的和/或γ逆转录病毒的。

[0719] 在瞬时表达是优选的应用中,可以使用腺病毒基系统。腺病毒基载体在 许多细胞类型中能够具有非常高的转导效并且无需细胞分裂。用这样的载 体,已经获得了较高的效价和表达水平。可以在相对简单的系统中大量地产 生此载体。

[0720] 还可以使用腺相关病毒("AAV")载体转导具有靶核酸的细胞,例 如,在体外产生核酸和肽,以及用于体内和离体基因疗法程序(参见例如, 韦斯特(West)等人,《病毒学》(Virology)160:38-47(1987);美国专 利号4,797,368;WO 93/24641;科丁(Kotin),《人类基因疗法》(Human Gene Therapy)5:793-801(1994);缪斯科斯卡(Muzyczka),《临床研究杂志》(J.Clin.Invest.)94:1351(1994))。重组AAV载体的构建描述于 多个出版物中,包括美国专利号5,173,414;特拉特斯金(Tratschin)等人, 《分子与细胞生物学》(Mol.Cell.Biol.)5:3251-3260(1985);特拉特斯金 等人,《分子与细胞生物学》4:2072-2081(1984);埃尔莫奈特 (Hermonat)&缪斯科斯卡(Muzyczka),《美国国家科学院院刊》(PNAS)81:6466-6470(1984);以及萨穆尔斯基(Samulski)等人,《病 毒学杂志》(J.Virol.)63:03822-3828(1989)。

[0721] 典型地使用包装细胞形成能够感染宿主细胞的病毒粒子。这样的细胞包 括包装腺病毒的293细胞和包含逆转录病毒的ψ2细胞或PA317细胞。在基因 疗法中使用的病毒载体通常由将核酸载体包装进病毒粒子的细胞系产生。这 些载体典型地包含包装并且随后整合进宿主所需的最低量序列,其他病毒序 列由用于有待表达的一个或多个多核苷酸的表达盒替换。失去的病毒功能典 型地由包装细胞系反式地提供。例如,在基因疗法中使用的AAV载体典型地 仅具有来自AAV基因组的ITR序列,这些序列为包装并整合进宿主基因组所 需。将病毒DNA包装进以下细胞系中,该细胞系包含编码辅助质粒的其他 AAV基因,即rep和cap,但缺少ITR序列。该细胞系还可以被作为辅助者的 腺病毒感染。该辅助病毒促进AAV载体的复制和从辅助质粒表达AAV基 因。由于缺少ITR序列,未以显著的量包装该辅助质粒。可以通过例如与 AAV相比腺病毒更加敏感的热处理减少腺病毒的污染。

[0722] 因此,AAV被认为是用作转导载体的理想候选物。这样的AAV转导载 体可以包括足够的顺式作用功能,以在反式地提供的腺病毒或疱疹病毒或痘 病毒(例如,牛痘病毒)辅助功能的存在下进行复制。可以使用重组AAV (rAAV)将外源基因携带进多种谱系的细胞中。在这些载体中,将AAV cap 和/或rep基因从病毒基因组中缺失并且用选择的DNA区段置换。当前的 AAV载体可以容纳多达插入DNA的4300个碱基。

[0723] 存在多种产生rAAV的方式,并且本发明提供了rAAV以及用于制备 rAAV的方法。例如,一个或多个质粒包含或基本包含希望的病毒构建体,将 其转染进AAV感染的细胞中。此外,将第二或另外的辅助质粒共转染进这些 细胞中,以提供重组病毒构建体的复制和包装必需的AAV rep和/或cap基 因。在这些条件下,AAV的rep和/或cap蛋白反式地作用,以刺激rAAV构 建体的复制和包装。转染后两至三天,收获rAAV。传统地,从细胞中收获 rAAV连同腺病毒。然后,通过热处理使污染的腺病毒失活。在本发明中,有 利地不从细胞自身收获rAAV,而是从细胞上清液收获rAAV。因此,在初始 方面,本发明提供了制备rAAV,并且除前述内容之外,可以通过以下方法制 备rAAV,该方法包括或基本由以下组成:用包含外源DNA(包括供表达的 DNA)和辅助病毒(例如,腺病毒、疱疹病毒、痘病毒如牛痘病毒)的 rAAV感染易感细胞,其中该rAAV缺少功能性cap和/或rep(以及提供该 rAAV缺少的cap和/或rev功能的辅助病毒(例如,腺病毒、疱疹病毒、痘病 毒如牛痘病毒));或用包含外源DNA(包括供表达的DNA)的rAAV感染 易感细胞,其中该重组体缺少功能性cap和/或rep,并且用提供该rAAV缺少 的cap和/或rep功能的质粒转染所述细胞;或用包含外源DNA(包括供表达 的DNA)的rAAV感染易感细胞,其中该重组体缺少功能性cap和/或rep, 其中所述细胞提供该重组体缺少的cap和/或rep功能;或用缺少功能性cap和 /或rep的AAV和用于将外源DNA插入该重组体中这样使得由该重组体表达 该外源DNA并且用于提供rep和/或cap功能的质粒转染易感细胞,由此转染 产生包含该外源DNA(包括供表达的DNA)的缺少功能性cap和/或rep的 rAAV。

[0724] 该rAAV可以来自如在此描述的AAV,并且有利地可以是rAAV1、rAAV2、AAV5或具有杂交体或衣壳的rAAV,该rAAV可以包括AAV1、AAV2、AAV5或其任何组合。可以相对于有待被rAAV靶向的细胞而选择 rAAV的AAV;例如,可以选择用于靶向脑或神经元细胞的AAV血清型1、2、5或杂交体或衣壳AAV1、AAV2、AAV5或其任何组合;并且可以选择用 于靶向心脏组织的AAV4。

[0725] 除293细胞之外,可以在本发明的实践中使用的其他细胞以及这些细胞 的某些体

外AAV血清型的相对感染性(参见格林,D.(Grimm,D.)等人,《病毒学杂志》(J.Virol.)82:5887-5911(2008))如下:

[0726]

细胞系	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
海拉 (HeLa)	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
单核细胞	1111	100	ND	ND	125	1429	ND	ND

[0727]

未成熟 DC	2500	100	ND	ND	222	2857	ND	ND
成熟 DC	2222	100	ND	ND	333	3333	ND	ND

[0728] 本发明提供了包含或基本上由编码CRISPR(规律间隔成簇短回文重 复)系统的外源核酸分子组成的rAAV,例如,多个盒,该多个盒包括或包含 第一盒,该第一盒包括或基本上由一个启动子、一个编码CRISPR相关 (Cas)蛋白(假定的核酸酶或解旋酶蛋白)(例如,Cas9)的核酸分子和一 个终止子组成,以及两个或更多个,有利地直到载体的包装尺寸极限,例如 总共(包括第一盒)五个盒,这些盒包括或基本上由一个启动子、一个编码 指导RNA(gRNA)的核酸分子和一个终止子组成(例如,每个盒示意性地 由启动子-gRNA1-终止子、启动子-gRNA2-终止子...启动子-gRNA(N)-终止子 表示(其中N是可以插入的处于载体的包装尺寸极限的上限的数目)),或 两个或更多个单独的rAAV,每个rAAV包含一个或多于一个CRISPR系统 盒,例如,第一rAAV包含第一盒,该第一盒包括或基本上由一个启动子、一个编码Cas(例如,Cas9)的核酸分子和一个终止子组成,并且第二rAAV 包含多个,四个,盒,这些盒包括或基本上由一个启动子、编码指导RNA (gRNA)的核酸分子和一个终止子组成(例如,每个盒示意性地由启动子- gRNA1-终止子、启动子-gRNA2-终止子...启动子-gRNA(N)-终止子表示(其中 N是可以插入的处于载体的包装尺寸极限的上限的数目))。由于rAAV是 一种DNA病毒,因此涉及AAV或rAAV的在此讨论中的核酸分子有利地是 DNA。在一些实施例中,启动子有利地是人类突触蛋白I启动子(hSyn)。

[0729] 用于将核酸递送至细胞的另外的方法是本领域的技术人员已知的。参见 例如通过引用结合在此的US 20030087817。

[0730] 在一些实施例中,用一个或多个在此描述的载体瞬时地或非瞬时地转染 宿主细胞。在一些实施例中,当细胞天然地出现在受试者体内时将其转染。在一些实施例中,被转染的细胞取自受试者。在一些实施例中,该细胞来源 于取自受试者的细胞,如细胞系。用于

组织培养的多种多样的细胞系在本领 域是已知的。细胞系的实例包括但不限于,C8161、CCRF-CEM、MOLT、 mIMCD-3、NHDF、HeLa-S3、Huh1、Huh4、Huh7、HUVEC、HASMC、 HEKn、HEKa、MiaPaCell、Panc1、PC-3、TF1、CTLL-2、C1R、Rat6、 CV1、RPTE、A10、T24、J82、A375、ARH-77、Calu1、SW480、SW620、 SKOV3、SK-UT、CaCo2、P388D1、SEM-K2、WEHI-231、HB56、TIB55、Jurkat、J45.01、LRMB、Bcl-1、BC-3、IC21、DLD2、Raw264.7、NRK、 NRK-52E、MRC5、MEF、HepG2、海拉B、海拉T4、COS、COS-1、COS- 6、COS-M6A、BS-C-1猴肾上皮细胞、BALB/3T3小鼠胚胎成纤维细胞、3T3 Swiss、3T3-L1、132-d5人类胎儿成纤维细胞;10.1小鼠成纤维细胞、293-T、3T3、721、9L、A2780、A2780ADR、A2780cis、A172、A20、A253、 A431、A-549、ALC、B16、B35、BCP-1细胞、BEAS-2B、bEnd.3、BHK- 21、BR 293、BxPC3、C3H-10T1/2、C6/36、Cal-27、CHO、CHO-7、CHO- IR、CHO-K1、CHO-K2、CHO-T、CHO Dhfr-/-、COR-L23、COR- L23/CPR、COR-L23/5010、COR-L23/R23、COS-7、COV-434、CML T1、 CMT、CT26、D17、DH82、DU145、DuCaP、EL4、EM2、EM3、 EMT6/AR1、EMT6/AR10.0、FM3、H1299、H69、HB54、HB55、HCA2、 HEK-293、HeLa、Hepa1c1c7、HL-60、HMEC、HT-29、Jurkat、JY细胞、 K562细胞、Ku812、KCL22、KG1、KYO1、LNCap、Ma-Mel 1-48、MC- 38、MCF-7、MCF-10A、MDA-MB-231、MDA-MB-468、MDA-MB-435、 MDCK II、MDCK II、MOR/0.2R、MONO-MAC 6、MTD-1A、MyEnd、NCI- H69/CPR、NCI-H69/LX10、NCI-H69/LX20、NCI-H69/LX4、NIH-3T3、 NALM-1、NW-145、OPCN/OPCT细胞系、Peer、PNT-1A/PNT 2、RenCa、 RIN-5F、RMA/RMAS、Saos-2细胞、Sf-9、SkBr3、T2、T-47D、T84、THP1 细胞系、U373、U87、U937、VCaP、维洛(Vero)细胞、WM39、WT-49、 X63、YAC-1、YAR及其转基因变种。细胞系可获得自本领域的技术人员已 知的多种来源(参见例如,美国典型培养物保藏中心(American Type Culture Collection)(ATCC)(马纳萨斯(Manassus),弗吉尼亚州))。在一些实 施例中,使用用一种或多种在此描述的载体转染的细胞建立新的细胞系,该 新的细胞系包括一种或多种载体来源的序列。在一些实施例中,使用用如在 此描述的CRISPR系统的组分转染(如通过用一种或多种载体进行瞬时转染 或用RNA进行转染)且通过CRISPR复合物的活性修饰的细胞建立新的细胞 系,该新的细胞系包括以下细胞,这些细胞包含吸水但是缺少任何其他外源 序列。在一些实施例中,在评估一种或多种测试化合物中使用用一个或多个 在此描述的载体瞬时或非瞬时转染的细胞或来源于这样的细胞的细胞系。

[0731] 在一些实施例中,使用一个或多个在此描述的载体产生非人转基因动物 或转基因植物。在一些实施例中,该转基因动物是一种哺乳动物,如小鼠、大鼠或兔。用于产生转基因动物和植物的方法在本领域是已知的,并且通常 以如在此描述的细胞转染方法开始。

[0732] 在另一个实施例中,可以考虑用具有针阵列的流体递送装置(参见例 如,转让给弗雷德·哈钦森癌症研究中心的美国专利公开号20110230839)将 CRISPR Cas递送至实体组织。用于将流体递送至实体组织的美国专利公开号 20110230839的装置可以包括多个排列为阵列的针;多个储库,每个储库都与 该多个针中的对应的针处于流体联通;以及多个致动器,这些致动器可操作 地耦合至该多个储库中的对应的储库并被配置为控制该储库内的流体压力。 在某些实施例中,该多个致动器中的每个都可以包括多个柱塞中的一个,该 多个柱塞中的每个的第一端都被接收进该多个储库中的对应的储库中,并且 在某些另外的实施例中,该多个柱塞中的柱塞在其对应的第二端被可操作地 耦合在一起,以便可以同时被按下。某些仍另外的实施例可以包括一个柱塞 驱动器,该柱塞驱动器被配置为以选择性地可变速率按下所有该多个柱塞。 在其他实施例中,该多个致动器中的每个都可以包

括多个流体传输线中的一 个,这些流体传输线具有第一和第二端,该多个流体传输线中的每个的第一 端被耦合至该多个储库中的对应的储库。在其他实施例中,该装置可以包括一个流体压力源,并且该多个致动器中的每个在该流体压力源与该多个储库 中的对应的储库之间都包括一个液压耦合器。在另外的实施例中,该流体压 力源可以包括压缩机、真空蓄能器、蠕动泵、主缸、微流体泵以及阀中的至 少一个。在另一个实施例中,该多个针中的每个都可以包括多个沿其长度分 布的端口。

[0733] 修饰靶标

[0734] 在一个方面,本发明提供了修饰真核细胞中的靶多核苷酸的方法,这些 方法可以在体内、离体或在体外。在一些实施例中,该方法包括从人或非人 动物或植物取样细胞或细胞群,并且修饰该细胞或这些细胞。培养可以发生 在离体的任何阶段。该细胞或这些细胞甚至可以被重新引入该非人动物或植 物中。对于重新引入的细胞,特别优选的是这些细胞是干细胞。

[0735] 在一些实施例中,该方法包括允许一种CRISPR复合物结合到该靶多核 苷酸上以实施所述靶多核苷酸的切割,由此修饰该靶多核苷酸,其中该 CRISPR复合物包括与杂交到在所述靶多核苷酸内的一个靶序列上的指导序列 复合的CRISPR酶,其中所述指导序列连接到一种tracr配对序列上,该tracr 配对序列进而杂交到一种tracr序列上。

[0736] 在一个方面,本发明提供了修饰一种多核苷酸在真核细胞中的表达的方 法。在一些实施例中,该方法包括允许一种CRISPR复合物结合到该多核苷 酸上,这样使得所述结合导致所述多核苷酸的表达增加或降低;其中该 CRISPR复合物包括与杂交到在所述多核苷酸内的一个靶序列上的指导序列复 合的CRISPR酶,其中所述指导序列连接到一种tracr配对序列上,该tracr配 对序列进而杂交到一种tracr序列上。类似的考虑因素和条件适用如上文针对 修饰靶多核苷酸的方法。事实上,这些取样、培养和重新引入选择跨本发明 的多个方面而适用。

[0737] 确实,在本发明的任何方面,该CRISPR复合物可以包括与杂交到一个 靶序列上的指导序列复合的CRISPR酶,其中所述指导序列可以连接到一种 tracr配对序列上,该tracr配对序列进而可以杂交到一种tracr序列上。类似的 考虑因素和条件适用如上文针对修饰靶多核苷酸的方法。

[0738] 试剂盒

[0739] 在一个方面,本发明提供了以下试剂盒,这些试剂盒包含披露于以上方 法和组合物中的元件中的任何一个或多个。元件可以单独地或组合地提供, 并且可以被提供于任何适合的容器中,如小瓶、瓶子或管。在一些实施例 中,该试剂盒包括一种或多种语言,例如多于一种语言的说明书。

[0740] 在一些实施例中,试剂盒包括一种或多种用于在利用在此描述的元件中 的一种或多种的方法中使用的试剂。试剂可以被提供于任何适合的容器中。 例如,试剂盒可以提供一种或多种反应或存储缓冲液。可以按在具体测定中 可用的形式或按在使用之前需要添加一种或多种其他组分的形式(例如按浓 缩或冻干形式)提供试剂。缓冲液可以是任何缓冲液,包括但不限于碳酸钠 缓冲液、碳酸氢钠缓冲液、硼酸盐缓冲液、Tris缓冲液、MOPS缓冲液、 HEPES缓冲液及其组合。在一些实施例中,该缓冲液是碱性的。在一些实施 例中,该缓冲液具有从约7至约10的pH。在一些实施例中,该试剂盒包括 一个或多个寡核苷酸,该

一个或多个寡核苷酸对应于一个用于插入进载体中 的指导序列，以便可操作地连接该指导序列和一个调节元件。在一些实施例 中，该试剂盒包括一个同源重组模板多核苷酸。在一些实施例中，该试剂盒 包括在此描述的载体中的一个或多个和/或在此描述的多核苷酸中的一个或多 个。该试剂盒可以有利地允许提供本发明的系统的所有元件。

[0741] CRISPR复合物

[0742] 在一个方面，本发明提供了用于使用CRISPR系统的一个或多个元件的 方法。本发明的CRISPR复合物提供了用于修饰靶多核苷酸的有效手段。本 发明的CRISPR复合物具有多种多样的实用性，包括修饰(例如，缺失、插 入、转位、失活、激活)多种细胞类型中的靶多核苷酸。正因为如此，本发 明的CRISPR复合物在例如基因疗法、药物筛选、疾病诊断以及预后中具有 广阔的应用谱。示例性CRISPR复合物包括与一种指导序列复合的CRISPR 酶，该指导序列与靶多核苷酸内的靶序列杂交。该指导序列与一种tracr配对 序列连接，该tracr配对序列进而与一种tracr序列杂交。

[0743] 在一个实施例中，本发明提供了一种切割靶多核苷酸的方法。该方法包 括使用一种CRISPR复合物修饰靶多核苷酸，该CRISPR复合物结合到该靶多 核苷酸上并且实施所述靶多核苷酸的切割。典型地，当被引入进细胞中时， 本发明的CRISPR复合物在基因组序列中产生一个断裂(例如，单链或双链 断裂)。例如，可以使用该方法切割细胞中的疾病基因。

[0744] 由该CRISPR复合物产生的断裂可以通过修复过程进行修复，如易出错 的非同源末端连接(NHEJ)途径或高保真同源定向修复(HDR)(图 29)。在这些修复过程期间，可以将一个外源多核苷酸模板引入进基因组序 列中。在一些方法中，该HDR过程被用于修饰基因组序列。例如，将一个外 源多核苷酸模板引入进细胞中，该外源多核苷酸模板包括一个有待整合的侧 翼为一个上游序列和一个下游序列的序列。上游和下游序列与染色体中的整合位点的任一侧都具有序列相似性。

[0745] 在希望的情况下，供体多核苷酸可以是DNA，例如DNA质粒、细菌人 工染色体(BAC)、酵母人工染色体(YAC)、病毒载体、一段线性DNA、PCR片段、裸核酸或与递送赋形剂(如脂质体或泊洛沙姆)复合的核酸。

[0746] 该外源多核苷酸模板包括一个有待整合的序列(例如，突变的基因)。供整合的序列可以是一个对细胞而言内源或外源的序列。有待整合的序列的 实例包括编码蛋白质的多核苷酸或非编码RNA(例如，微小RNA)。因此，供整合的序列可以可操作地连接到一个或多个适当的控制序列上。可替代 地，有待整合的序列可以提供一种调节功能。

[0747] 选择外源多核苷酸模板中的上游和下游序列，以促进感兴趣的染色体序 列与供体多核苷酸之间的重组。该上游序列是一个与供整合的靶向位点的上 游的基因组序列具有序列相似性的核酸序列。类似地，该下游序列是一个与 供整合的靶向位点的染色体序列下游具有序列相似性的核酸序列。外源多核 苷酸模板中的上游和下游序列与靶向的基因组序列可以具有75％、80％、 85％、90％、95％或100％序列一致性。优选地，外源多核苷酸模板中的上游 和下游序列与靶向的基因组序列具有约95％、96％、97％、98％、99％或100％序列一致性。在一些方法中，外源多核苷酸模板中的上游和下游序列与 靶向的基因组序列具有约99％或100％序列一致性。

[0748] 上游或下游序列可以包括从约20bp至约2500bp，例如约50、100、 200、300、400、500、600、700、800、900、1000、1100、1200、1300、 1400、1500、1600、1700、1800、1900、2000、

2100、2200、2300、2400或 2500bp。在一些方法中，示例性上游或下游序列具有约200bp至约2000 bp、约600bp至约1000bp或更具体地约700bp至约1000bp。

[0749] 在一些方法中，该外源多核苷酸模板可以进一步包括一种标记。这样的 一种标记使得容易地筛选靶向的整合。适合的标记的实例包括限制性位点、荧光蛋白或选择性标记。可以使用重组技术构建本发明的外源多核苷酸模板 （参见例如，萨姆布鲁克(Sambrook)等人，2001和奥苏贝尔(Ausubel)等 人，1996)。

[0750] 在一种用于通过整合外源多核苷酸模板而修饰靶多核苷酸的示例性方法 中，通过CRISPR复合物将双链断裂引入进基因组序列中，经由同源重组一 个外源多核苷酸模板而修复该断裂，这样使得将该模板整合进基因组中。双 链断裂的存在促进模板的整合。

[0751] 在其他实施例中，本发明提供了一种修饰多核苷酸在真核细胞中的表达 的方法。该方法包括通过使用结合到多核苷酸上的CRISPR复合物增加或降 低靶多核苷酸的表达。

[0752] 在一些方法中，可以使靶多核苷酸失活以实施细胞中的表达的修饰。例 如，当CRISPR复合物结合到细胞中的一个靶序列上时，该靶多核苷酸失 活，这样使得该序列不被转录，该编码蛋白不被产生，或者该序列不会像野 生型序列一样起作用。例如，可以使蛋白质或微小RNA编码序列失活，这样 使得该蛋白质不被产生。

[0753] 在一些方法中，可以使控制序列失活，这样使得它不再作为控制序列起 作用。如在此使用，“控制序列”是指影响核酸序列的转录、翻译或可及性 的任何核酸序列。控制序列的实例包括启动子、转录终止子和增强子，它们 是控制序列。

[0754] 失活的靶序列可以包括缺失突变(即，缺失一个或多个核苷酸)、插入 突变(即，插入一个或多个核苷酸)或无义突变(即，用另一个核苷酸取代 一个单核苷酸，这样使得引入终止密码子)。在一些方法中，靶序列的失活 导致该靶序列的“敲除(knock-out)”。

[0755] 疾病模型

[0756] 可以使用本发明的方法产生可以用作疾病模型的植物、动物或细胞。如 在此使用，“疾病”是指受试者的疾病、障碍或适应症。例如，可以使用本 发明的方法产生以下动物或细胞，该动物或细胞在一个或多个与疾病相关的 核酸序列中包括修饰，或以下植物、动物或细胞，一个或多个与疾病相关的 核酸序列的表达在该植物、动物或细胞中被改变。这样的一个核酸序列可以 编码一个疾病相关蛋白序列或可以是一个疾病相关控制序列。因此，应该理 解的是在本发明的实施例中，植物、受试者、患者、生物或细胞可以是非人 受试者、患者、生物或细胞。因此，本发明提供了由本发明方法产生的植 物、动物或细胞或其子代。该子代可以是产生的植物或动物的克隆或可以通 过与同一物种的其他个体杂交而由有性繁殖产生，以在其后代中基因渗入另 外的令人希望的性状。在多细胞生物(特别是动物或植物)的情况下，该细 胞可以是在体内或离体的。在该细胞处于培养中的情况下，如果满足适当的 培养条件并且优选地，如果该细胞适合地适于此目的(例如，干细胞)，则 可以建立细胞系。还设想了通过本发明产生的细菌细胞系。因此，还设想了 细胞系。

[0757] 在一些方法中，可以使用疾病模型研究突变对动物或细胞的影响以及使 用在疾病研究中常用的措施对疾病的发展和/或进展的影响。可替代地，这样 的一种疾病模型有用于研究药物活性化合物对疾病的影响。

[0758] 在一些方法中，可以使用疾病模型评估潜在的基因疗法策略的效力。也 就是说，可以将疾病相关基因或多核苷酸进行修饰，这样使得疾病发展和/或 进展被抑制或减少。

具体而言，该方法包括修饰一个疾病相关基因或多核苷 酸，这样使得产生一种改变的蛋白质，其结果是，该动物或细胞具有改变的 反应。因此，在一些方法中，可以将基因修饰动物与易患该疾病的动物进行 比较，这样使得可以评估基因疗法事件的效果。

[0759] 在另一个实施例中，本发明提供了一种研发生物活性剂的方法，该生物 活性剂调制与疾病基因相关的细胞信号传导事件。该方法包括使一种测试化 合物与一个细胞接触，该细胞包括一种或多种载体，这一种或多种载体驱动 CRISPR酶、连接到tracr配对序列上的指导序列和tracr序列中的一种或多种 的表达；并且检测读出的变化，该变化指示与例如包含在该细胞中的疾病基 因的突变相关的细胞信号传导事件的减少或增加。

[0760] 可以与用于筛选细胞功能变化的本发明的方法组合构建细胞模型或动物 模型。可以使用这样的一种模型研究由本发明的CRISPR复合物修饰的基因 组序列对感兴趣的细胞功能的影响。例如，可以使用细胞功能模型研究修饰 的基因组序列对细胞内信号传导或细胞外信号传导的影响。可替代地，可以 使用细胞功能模型研究修饰的基因组序列对感知觉的影响。在一些这样的模 型中，修饰一个或多个与该模型中的信号传导生化学途径相关的基因组序 列。

[0761] 已经具体研究了若干疾病模型。这些疾病模型包括新生自闭症风险基因 CHD8、KATNAL2和SCN2A；以及综合征性自闭症(安格曼综合征 (Angelman Syndrome))基因UBE3A。这些基因以及所得的自闭症模型当 然是优选的，但用于显示本发明的跨基因和对应的模型的广阔的适用性。

[0762] 可以通过当将测试模型细胞和对照细胞与候选试剂接触时，测定它们之 间的对应的基因的mRNA水平差异而确定一个或多个与信号传导生化途径相 关的基因组序列的改变的表达。可替代地，通过检测编码的多肽或基因产物 的水平差异而确定与信号传导生化途径相关的序列的差异表达。

[0763] 为了在mRNA转录本或对应的多核苷酸水平上测定试剂诱导的变化，首 先根据本领域的标准方法提取包含在样品中的核酸。例如，可以根据陈述于 萨姆布鲁克(Sambrook)等人(1989)中的程序使用各种水解酶分离mRNA 或遵循由制造商提供的随附说明书通过核酸结合树脂提取mRNA。然后，根 据本领域广泛已知的方法或基于在此示例的方法，通过扩增程序或常规的杂 交测定(例如Northern印迹分析)检测包含在提取的核酸样品中的mRNA。

[0764] 出于本发明的目的，扩增意指利用引物和聚合酶的能够以合理的保真度 复制靶序列的任何方法。可以通过天然或重组DNA聚合酶，如TaqGold™、T7 DNA聚合酶、大肠杆菌DNA聚合酶的Klenow片段以及逆转录酶进行扩 增。一种优选的扩增方法是PCR。具体而言，可以使分离的RNA经受逆转录 测定，该测定与定量聚合酶链式反应(RT-PCR)耦合，以便定量与信号传导 生化途径相关的序列的表达水平。

[0765] 在扩增测定中可以实时地执行基因表达水平的检测。在一个方面，可以 用荧光DNA结合剂直接使扩增产物可视化，这些荧光DNA结合剂包括但不 限于DNA嵌入剂和DNA沟结合剂。因为掺入进双链DNA分子中的嵌入剂 的量典型地与扩增DNA产物的量成比例，可以常规地通过使用本领域的常规 光学系统定量嵌入的染料的荧光而确定扩增产物的量。适于本申请的DNA结 合染料包括SYBR绿、SYBR蓝、DAPI、碘化丙锭、Hoeste、SYBR金、溴 化乙锭、吖啶、原黄素、吖啶橙、吖啶黄、氟香豆素(fluorcoumanin)、玫 瑰树碱、道诺霉素、氯喹、

偏端霉素D、色霉素、乙菲啶(homidium)、光 辉霉素、多吡啶钌、安曲霉素等。

[0766] 在另一个方面,可以在扩增反应中利用其他荧光标记物(如序列特异的 探针),以促进扩增产物的检测和定量。基于探针的定量扩增依赖于希望的 扩增产物的序列特异的检测。它利用荧光的靶标特异的探针(例如, TaqMan®探针),从而导致增加的特异性和敏感性。用于进行基于探针的定 量扩增的方法本领域是确立的并且教导于美国专利号5,210,015中。

[0767] 在又另一个方面,可以使用与以下序列具有序列同源性的杂交探针进行 常规的杂交测定,这些序列与信号传导生化途径相关。典型地,在杂交反应 中,允许探针和与信号传导生化途径相关的序列形成稳定的复合物,这些序 列被包含在来源于测试受试者的生物样品内。本领域的技术人员应该意识到 的是,在将反义核酸用作探针核酸的情况下,提供于样品中的靶多核苷酸被 选择为与该反义核酸的序列互补。相反地,在核苷酸探针是有义核酸的情况 下,该靶多核苷酸被选择为与该有义核酸的序列互补。

[0768] 可以在不同严格度条件下进行杂交。用于实践本发明的适合的杂交条件 是这样的,使得探针和与信号传导生化途径相关的序列之间的识别相互作用 既是充分特异的又是充分稳定的。增加杂交反应的严格度的条件在本领域是 广泛已知且公开的。参见例如,(萨姆布鲁克(Sambrook)等人, (1989);非放射性原位杂交应用手册(Nonradioactive In Situ Hybridization Application Manual),宝灵曼公司(Boehringer Mannheim),第二版)。可 以使用固定在任何固体支持物上的探针进行杂交测定,所述固体支持物包括 但不限于硝化纤维、玻璃、硅以及多种基因阵列。如描述于美国专利号 5,445,934中,在高密度基因芯片上执行优选的杂交测定。

[0769] 对于在杂交测定过程中形成的探针-靶标复合物的常规检测,将核苷酸探 针络合至可检测标记物上。适于在本发明中使用的可检测标记物包括通过光 化学、生物化学、光谱学、免疫化学、电学、光学或化学手段可检测的任何 组合物。多种多样的适当的可检测标记物在本领域是已知的,包括荧光或化 学发光标记物、放射性同位素标记物、酶或其他配体。在优选的实施例中, 可能希望的是利用荧光标记物或酶标签,如地高辛、β-半乳糖苷酶、脲酶、 碱性磷酸酶或过氧化物酶、亲和素/生物素复合物。

[0770] 用于检测或定量杂交强度的检测方法将典型地取决于以上选择的标记 物。例如,可以使用照相底片或感光成像仪检测放射性标记物。可以使用检 测发射光的光检测器检测并定量荧光标记物。典型地通过为酶提供底物并测 量由酶对底物的作用而产生的反应产物来检测酶标记物;并且最后,通过简 单地使着色的标记物可视化而检测比色标记物。

[0771] 还可以通过检查对应的基因产物确定与信号传导生化途径相关的序列的 试剂诱导的表达变化。确定蛋白质水平典型地涉及a)使包含在生物样品中的 蛋白质与特异性结合到与信号传导生化途径相关的蛋白质上的试剂接触;并 且(b)鉴定这样形成的任何试剂:蛋白质复合物。在此实施例的一个方面, 特异性结合与信号传导生化途径相关的蛋白质的该试剂是一种抗体,优选单 克隆抗体。

[0772] 通过在将允许在该试剂和与信号传导生化途径相关的蛋白质之间形成复 合物的条件下,通过使该试剂与来源于测试样品的与信号传导生化途径相关 的蛋白质的样品接触而进行该反应。可以根据本领域中的标准程序直接地或 间接地检测复合物的形成。在直接检测方法中,这些试剂提供有可检测的标 记物并且未反应的试剂可以从复合物中除去;

由此剩余的标记物的量指示形 成的复合物的量。对于这样的方法,优选的是选择即使在严格洗涤条件过程 中仍附接至这些试剂上的标记物。优选的是,该标记物不干扰结合反应。在 替代方案中,间接检测程序可以使用包含用化学方法或酶方法引入的标记物 的试剂。令人希望的标记物通常不干扰所得的试剂:多肽复合物的结合或稳 定性。然而,标记物典型地被设计成是用于有效结合并且因此产生可检测的 信号的抗体可及的。

[0773] 适于检测蛋白质水平的多种多样的标记物在本领域是已知的。非限制性 实例包括放射性同位素、酶、胶体金属、荧光化合物、生物发光化合物以及 化学发光化合物。

[0774] 可以通过标准定量测定定量在结合反应过程中形成的试剂:多肽复合物 的量。如上所示,可以通过仍留在结合位点处的标记物的量直接测量试剂: 多肽复合物的形成。在一个替代方案中,测试与信号传导生化途径相关的蛋 白质与标记的类似物结合特定试剂上的位点的竞争能力。在此竞争测定中, 捕获的标记物的量与存在于测试样品中的与信号传导生化途径相关的蛋白质 序列的量成反比。

[0775] 多种用于蛋白质分析的基于以上概括的总则的技术在本领域是可获得 的。它们包括但不限于放射免疫测定、ELISA(酶联免疫放射测定)、“夹 层”免疫测定、免疫放射测定、原位免疫测定(使用例如,胶体金、酶或放 射性同位素标记物)、蛋白质印迹分析、免疫沉淀测定、免疫荧光测定以及 SDS-PAGE。

[0776] 特异性识别或结合到与信号传导生化途径相关的蛋白质上的抗体对于执 行前述蛋白质分析而言是优选的。在希望的情况下,可以使用识别特定类型 的翻译后修饰(例如,信号传导生化途径诱导的修饰)的抗体。翻译后修饰 包括但不限于糖基化、脂化、乙酰化以及磷酸化。这些抗体可以购自商业供 应商。例如,特异性识别酪氨酸磷酸化蛋白质的抗磷酸酪氨酸抗体可以获得 自多个供应商,包括英杰公司和珀金埃尔默公司(Perkin Elmer)。抗磷酸酪 氨酸抗体在检测响应于ER应激而在其酪氨酸残基上存在差异磷酸化的蛋白质中是特别有用的。这样的蛋白质包括但不限于真核翻译起始因子2α(eIF- 2α)。可替代地,可以通过用展示出希望的翻译后修饰的靶蛋白免疫宿主动 物或抗体产生细胞而使用常规的多克隆或单克隆抗体技术产生这些抗体。

[0777] 在实践主题方法中,可以令人希望的是,在不同身体组织中、在不同细 胞类型中和/或在不同亚细胞结构中辨别与信号传导生化途径相关的蛋白质的 表达谱。可以通过使用能够结合到优先表达于某些组织、细胞类型或亚细胞 结构中的蛋白质标记的组织特异的、细胞特异的或亚细胞结构特异的抗体进 行这些研究。

[0778] 还可以通过检查基因产物相对于对照细胞的活性变化而确定与信号传导 生化途径相关的基因的改变的表达。与信号传导生化途径相关的蛋白质的试 剂诱导的活性变化的测定将取决于处于研究之下的生物活性和/或信号传导途 径。例如,在该蛋白质是一种激酶的情况下,可以通过本领域已知的多种测 定确定它磷酸化一种或多种下游底物的能力的变化。代表性测定包括但不限 于用抗体进行免疫印迹和免疫沉淀,这些抗体是如识别磷酸化蛋白质的抗磷 酸酪氨酸抗体。此外,可以通过高通量化学发光测定检测激酶活性,如 AlphaScreen™(可获得自珀金埃尔默公司)和eTag™测定(陈-辉(Chan- Hui)等人(2003)《临床免疫学》(Clinical Immunology)111:162-174)。

[0779] 在与信号传导生化途径相关的蛋白质是使得细胞内pH条件波动的信号传 导级联的一部分的情况下,可以将pH敏感的分子(如荧光pH染料)用作报 道分子。在与信号传导生

化途径相关的蛋白质是一种离子通道的另一个实例 中,可以监测膜电位和/或细胞内离子浓度的波动。多种商业试剂盒和高通量 装置特别适于快速且稳健地筛选离子通道调节剂。代表性仪器包括FLIPRTM (分子设备公司(Molecular Devices,Inc.))和VIPR(奥罗拉生物科学公司 (Aurora Biosciences)))。这些仪器能够同时检测微板的1000多个样品孔中 的反应,并且能够提供一秒内或甚至一毫秒内的实时测量和功能数据。

[0780] 在实践在此披露的任何方法中,可以经由本领域已知的一种或多种方法 将适合的载体引入进细胞或胚胎中,这些方法包括但不限于,显微注射、电 穿孔、声致穿孔、基因枪、磷酸钙介导的转染、阳离子转染、脂质体转染、 树枝状聚合物转染、热休克转染、核转染、磁转染、脂转染、刺穿转染(impalefection)、光学转染、专有剂增强的核酸摄取以及经由脂质体、免疫 脂质体、病毒颗粒或人工病毒体进行递送。在一些方法中,通过显微注射将该载体引入进胚胎中。可以将这个或这些载体显微注射进胚胎的细胞核或细 胞质中。在一些方法中,通过核转染将这个或这些载体引入进细胞中。

[0781] CRISPR复合物的靶多核苷酸可以是对真核细胞而言内源或外源的任何多 核苷酸。例如,该靶多核苷酸可以是一种驻留在真核细胞的细胞核中的多核 苷酸。该靶多核苷酸可以是一个编码基因产物(例如,蛋白质)的序列或一 个非编码序列(例如,调节多核苷酸或无用DNA)。

[0782] 靶多核苷酸的实例包括与信号传导生化途径相关的序列,例如信号传导 生化途径相关基因或多核苷酸。靶多核苷酸的实例包括疾病相关基因或多核 苷酸。“疾病相关”基因或多核苷酸是指与非疾病对照的组织或细胞相比, 在来源于疾病影响的组织的细胞中以异常水平或以异常形式产生转录或翻译 产物的任何基因或多核苷酸。在改变的表达与疾病的出现和/或进展相关的情 况下,它可以是一个以异常高的水平被表达的基因;它可以是一个以异常低 的水平被表达的基因。疾病相关基因还指具有一个或多个突变或直接负责或 与一个或多个负责疾病的病因学的基因连锁不平衡的遗传变异的基因。转录 的或翻译的产物可以是已知的或未知的,并且可以处于正常或异常水平。

[0783] CRISPR复合物的靶多核苷酸可以是对真核细胞而言内源或外源的任何多 核苷酸。例如,该靶多核苷酸可以是一种驻留在真核细胞的细胞核中的多核 苷酸。该靶多核苷酸可以是一个编码基因产物(例如,蛋白质)的序列或一 个非编码序列(例如,调节多核苷酸或无用DNA)。不希望被理论所束缚, 据信该靶序列应该与PAM(原型间隔子邻近基序)相关;也就是说,与由 CRISPR复合物识别的短序列相关。对PAM的精确序列和长度要求取决于使 用的CRISPR酶而不同,但是PAM典型地是临近原型间隔子(也就是说,靶 序列)的2-5个碱基对序列。在以下实例部分中给出PAM序列的实例,并且 熟练人员将能够鉴定与给定的CRISPR酶一起使用的另外的PAM序列。

[0784] CRISPR复合物的靶多核苷酸可以包括多个疾病相关基因和多核苷酸以及 信号传导生化途径相基因和多核苷酸,如分别提交于2012年12月12日和 2013年1月2日、标题均为用于序列操纵的系统方法和组合物(SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION)的分 别具有博德(Broad)参考号BI-2011/008/WSGR案卷号44063-701.101和BI- 2011/008/WSGR案卷号44063-701.102的美国临时专利申请61/736,527和 61/748,427中所列举,将所有这些申请的内容通过引用以其全文结合在此。

[0785] 靶多核苷酸的实例包括与信号传导生化途径相关的序列,例如信号传导 生化途

径相关基因或多核苷酸。靶多核苷酸的实例包括疾病相关基因或多核 苷酸。“疾病相关”基因或多核苷酸是指与非疾病对照的组织或细胞相比，在来源于疾病影响的组织的细胞中以异常水平或以异常形式产生转录或翻译 产物的任何基因或多核苷酸。在改变的表达与疾病的出现和/或进展相关的情 况下,它可以是一个以异常高的水平被表达的基因;它可以是一个以异常低 的水平被表达的基因。疾病相关基因还指具有一个或多个突变或直接负责或 与一个或多个负责疾病的病因学的基因连锁不平衡的遗传变异的基因。转录 的或翻译的产物可以是已知的或未知的,并且可以处于正常或异常水平。

[0786] 疾病相关基因和多核苷酸的实例列于表A和B中。在万维网上可获得的 疾病具体信息可获得自约翰斯·霍普金斯大学的麦考斯克-纳森遗传医学研究所 (McKusick-Nathans Institute of Genetic Medicine,Johns Hopkins University) (巴尔的摩,马里兰州)和国立医学图书馆的国家生物技术信息中心 (National Center for Biotechnology Information,National Library of Medicine) (贝塞斯达,马里兰州)。信号传导生化途径相关基因和多核苷酸的实例列 于表C中。

[0787] 这些基因和途径的突变可以导致产生不当的蛋白质或以不当的量影响功 能的蛋白质。通过引用而特此结合来自于2012年12月12日提交的美国临时 申请61/736,527的基因、疾病和蛋白质的另外的实例。这样的基因、蛋白质 和途径可以是CRISPR复合物的靶多核苷酸。

[0788] 表A

[0789]

疾病/障碍	一个或多个基因
瘤形成	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPARα; PPAR
	γ; WT1（维耳姆斯瘤（Wilms Tumor））; FGF 受体家族
	成员（5 个成员：1、2、3、4、5）; CDKN2a; APC; RB
	（视网膜母细胞瘤）; MEN1; VHL; BRCA1; BRCA2; AR
	（雄激素受体）; TSG101; IGF; IGF 受体; Igf1（4
	种变体）; Igf2（3 种变体）; Igf 1 受体; Igf 2 受体;
	Bax; Bcl2; 半胱天冬酶家族（9 个成员：
	1、2、3、4、6、7、8、9、12）; Kras; Apc
年龄相关性黄斑	Abcr; Ccl2; Cc2; cp（铜蓝蛋白）; Timp3; 组织蛋白酶 D;
变性	Vldlr; Ccr2
精神分裂症	神经调节蛋白 1（Nrg1）; Erb4（神经调节蛋白受体）;
	复蛋白（Complexin）1（Cplx1）; Tph1 色氨酸羟化酶;

	Tph2
	色氨酸羟化酶 2；轴突蛋白（Neurexin）1；GSK3；GSK3a；
	GSK3b
障碍	5-HTT（Slc6a4）；COMT；DRD（Drd1a）；SLC6A3；DAOA；
	DTNBP1；Dao（Dao1）
三核苷酸重复	HTT（亨廷顿氏 Dx）；SBMA/SMAX1/AR（肯尼迪氏
障碍	Dx）；FXN/X25（弗里德里希共济失调）；ATX3（马查多-
	约瑟夫氏 Dx）；ATXN1 和 ATXN2（脊髓小脑
	共济失调）；DMPK（强直性肌营养不良）；Atrophin-1 和 Atn1
	（DRPLA Dx）；CBP（Creb-BP-整体不稳定性）；VLDLR
	（阿尔茨海默病）；Atxn7；Atxn10
脆性 X 综合征	FMR2；FXR1；FXR2；mGLUR5
分泌酶相关	APH-1（α 和 β）；早老素（Psen1）；呆蛋白
障碍	（Ncstn）；PEN-2
其他	Nos1；Parp1；Nat1；Nat2
朊病毒相关障碍	Prp
ALS	SOD1；ALS2；STEX；FUS；TARDBP；VEGF（VEGF-a；
	VEGF-b；VEGF-c）
药物成瘾	Prkce（酒精）；Drd2；Drd4；ABAT（酒精）；GRIA2；
	Grm5；Grin1；Htr1b；Grin2a；Drd3；Pdyn；Gria1（酒精）
自闭症	Mecp2；BZRAP1；MDGA2；Sema5A；轴突蛋白（Neurexin）1；脆性 X
	（FMR2（AFF2）；FXR1；FXR2；Mglur5）
阿尔茨海默病	E1；CHIP；UCH；UBB；Tau；LRP；PICALM；簇集蛋白；PS1；
	SORL1；CR1；Vldlr；Uba1；Uba3；CHIP28（Aqp1，
	水通道蛋白 1）；Uchl1；Uchl3；APP
炎症	IL-10；IL-1（IL-1a；IL-1b）；IL-13；IL-17（IL-17a（CTLA8）；IL-
	17b；IL-17c；IL-17d；IL-17f）；II-23；Cx3cr1；ptpn22；TNFa；
	IBD 的 NOD2/CARD15；IL-6；IL-12（IL-12a；IL-12b）；
	CTLA4；Cx3cl1
帕金森病	x-突触核蛋白；DJ-1；LRRK2；帕金蛋白（Parkin）；PINK1

[0790]

[0791] 表B：

[0792]

血液和凝血疾病以及障碍	贫血（CDAN1、CDA1、RPS19、DBA、PKLR、PK1、NT5C3、UMPH1、PSN1、RHAG、RH50A、NRAMP2、SPTB、ALAS2、ANH1、ASB、ABCB7、ABC7、ASAT）；裸淋巴细胞综合征（TAPBP、TPSN、TAP2、ABCB3、PSF2、RING11、MHC2TA、C2TA、RFX5、RFXAP、RFX5），出血障碍（TBXA2R、P2RX1、P2X1）；因子H和因子H样1（HF1、CFH、HUS）；因子V和因子VIII（MCFD2）；因子VII缺乏（F7）；因子X缺乏（F10）；因子XI缺乏（F11）；因子XII缺乏（F12、HAF）；因子XIIIA缺乏（F13A1、F13A）；因子XIIIB缺乏（F13B）；范科尼贫血（FANCA、FACA、FA1、FA、FAA、FAAP95、FAAP90、FLJ34064、FANCB、FANCC、FACC、BRCA2、FANCD1、FANCD2、FANCD、FACD、FAD、FANCE、FACE、FANCF、XRCC9、FANCG、BRIP1、BACH1、FANCJ、PHF9、FANCL、FANCM、KIAA1596）；噬血细胞淋巴组织细胞增生障碍（PRF1、HPLH2、UNC13D、MUNC13-4、HPLH3、HLH3、FHL3）；血友病A（F8、F8C、HEMA）；血友病B（F9、HEMB），出血性障碍（PI、ATT、F5）；白细胞缺乏和障碍（ITGB2、CD18、LCAMB、LAD、EIF2B1、EIF2BA、EIF2B2、EIF2B3、EIF2B5、LVWM、CACH、CLE、EIF2B4）；镰状细胞贫血（HBB）；地中海贫血（HBA2、HBB、HBD、LCRB、HBA1）。
细胞调节异常和肿瘤疾病以及障碍	B细胞非霍奇金淋巴瘤（BCL7A、BCL7）；白血病（TAL1、TCL5、SCL、TAL2、FLT3、NBS1、NBS、ZNFN1A1、IK1、LYF1、HOXD4、HOX4B、BCR、CML、PHL、ALL、ARNT、KRAS2、RASK2、GMPS、AF10、ARHGEF12、LARG、KIAA0382、CALM、CLTH、CEBPA、CEBP、CHIC2、BTL、FLT3、KIT、PBT、LPP、NPM1、NUP214、D9S46E、CAN、CAIN、RUNX1、CBFA2、AML1、WHSC1L1、NSD3、FLT3、AF1Q、NPM1、NUMA1、ZNF145、PLZF、PML、MYL、STAT5B、AF10、CALM、CLTH、ARL11、ARLTS1、P2RX7、P2X7、BCR、CML、PHL、ALL、GRAF、NF1、VRNF、WSS、NFNS、PTPN11、

[0793]

	PTP2C、SHP2、NS1、BCL2、CCND1、PRAD1、BCL1、TCRA、GATA1、GF1、ERYF1、NFE1、ABL1、NQO1、DIA4、NMOR1、NUP214、D9S46E、CAN、CAIN）。
炎症和免疫相关疾病以及障碍	AIDS（KIR3DL1、NKAT3、NKB1、AMB11、KIR3DS1、IFNG、CXCL12、SDF1）；自身免疫性淋巴增生综合征（TNFRSF6、APT1、FAS、CD95、ALPS1A）；联合免疫缺陷，（IL2RG、SCIDX1、SCIDX、IMD4）；HIV-1（CCL5、SCYA5、D17S136E、TCP228），HIV 易感性或感染（IL10、CSIF、CMKBR2、CCR2、CMKBR5、CCCKR5（CCR5））；免疫缺陷（CD3E、CD3G、AICDA、AID、HIGM2、TNFRSF5、CD40、UNG、DGU、HIGM4、TNFSF5、CD40LG、HIGM1、IGM、FOXP3、IPEX、AIID、XPID、PIDX、TNFRSF14B、TACI）；炎症（IL-10、IL-1（IL-1a、IL-1b）、IL-13、IL-17（IL-17a（CTLA8）、IL-17b、IL-17c、IL-17d、IL-17f）、II-23、Cx3cr1、ptpn22、TNFa、IBD 的 NOD2/CARD15、IL-6、IL-12（IL-12a、IL-12b）、CTLA4、Cx3cl1）；重症联合免疫缺陷（SCID）（JAK3、JAKL、DCLRE1C、ARTEMIS、SCIDA、RAG1、RAG2、ADA、PTPRC、CD45、LCA、IL7R、CD3D、T3D、IL2RG、SCIDX1、SCIDX、IMD4）。
代谢、肝、肾和蛋白疾病以及障碍	淀粉样蛋白神经病变（TTR、PALB）；淀粉样变性（APOA1、APP、AAA、CVAP、AD1、GSN、FGA、LYZ、TTR、PALB）；肝硬化（KRT18、KRT8、CIRH1A、NAIC、TEX292、KIAA1988）；囊性纤维化（CFTR、ABCC7、CF、MRP7）；糖原贮积病（SLC2A2、GLUT2、G6PC、G6PT、G6PT1、GAA、LAMP2、LAMPB、AGL、GDE、GBE1、GYS2、PYGL、PFKM）；肝腺瘤，142330（TCF1、HNF1A、MODY3），肝衰竭，早期发病以及神经障碍（SCOD1、SCO1），肝脂酶缺乏（LIPC）、肝母细胞癌、癌症（cancer）和癌（carcinomas）（CTNNB1、PDGFRL、PDGRL、PRLTS、AXIN1、AXIN、CTNNB1、TP53、P53、LFS1、IGF2R、MPRI、MET、CASP8、MCH5；髓质囊肾病（UMOD、HNFJ、FJHN、MCKD2、ADMCKD2）；苯丙酮尿症（PAH、PKU1、QDPR、DHPR、PTS）；多囊肾和肝病

[0794]

	（FCYT、PKHD1、ARPKD、PKD1、PKD2、PKD4、PKDTS、PRKCSH、G19P1、PCLD、SEC63）。
肌肉/骨骼疾病以及障碍	贝克肌营养不良（DMD、BMD、MYF6），杜氏肌营养不良（Duchenne Muscular Dystrophy）（DMD、BMD）；埃默里-德赖富斯（Emery-Dreifuss）肌营养不良（LMNA、LMN1、EMD2、FPLD、CMD1A、HGPS、LGMD1B、LMNA、LMN1、EMD2、FPLD、CMD1A）；面肩肱型肌营养不良（FSHMD1A、FSHD1A）；肌营养不良（FKRP、MDC1C、LGMD2I、LAMA2、LAMM、LARGE、KIAA0609、MDC1D、FCMD、TTID、MYOT、CAPN3、CANP3、DYSF、LGMD2B、SGCG、LGMD2C、DMDA1、SCG3、SGCA、ADL、DAG2、LGMD2D、DMDA2、SGCB、LGMD2E、SGCD、SGD、LGMD2F、CMD1L、TCAP、LGMD2G、CMD1N、TRIM32、HT2A、LGMD2H、FKRP、MDC1C、LGMD2I、TTN、CMD1G、TMD、LGMD2J、POMT1、CAV3、LGMD1C、SEPN1、SELN、RSMD1、PLEC1、PLTN、EBS1）；骨硬化病（LRP5、BMND1、LRP7、LR3、OPPG、VBCH2、CLCN7、CLC7、OPTA2、OSTM1、GL、TCIRG1、TIRC7、OC116、OPTB1）；肌萎缩（VAPB、VAPC、ALS8、SMN1、SMA1、SMA2、SMA3、SMA4、BSCL2、SPG17、GARS、SMAD1、CMT2D、HEXB、IGHMBP2、SMUBP2、CATF1、SMARD1）。
神经和神经元疾病以及障碍	ALS（SOD1、ALS2、STEX、FUS、TARDBP、VEGF（VEGF-a、VEGF-b、VEGF-c）；阿尔茨海默病（APP、AAA、CVAP、AD1、APOE、AD2、PSEN2、AD4、STM2、APBB2、FE65L1、NOS3、PLAU、URK、ACE、DCP1、ACE1、MPO、PACIP1、PAXIP1L、PTIP、A2M、BLMH、BMH、PSEN1、AD3）；自闭症（Mecp2、BZRAP1、MDGA2、Sema5A、轴突蛋白 1、GLO1、MECP2、RTT、PPMX、MRX16、MRX79、NLGN3、NLGN4、KIAA1260、AUTSX2）；脆性 X 综合征（FMR2、FXR1、FXR2、mGLUR5）；亨廷顿病和疾病样障碍（HD、IT15、PRNP、PRIP、JPH3、JP3、HDL2、TBP、SCA17）；帕金森病（NR4A2、NURR1、NOT、TINUR、SNCAIP、TBP、

[0795]

	SCA17、SNCA、NACP、PARK1、PARK4、DJ1、PARK7、LRRK2、PARK8、PINK1、PARK6、UCHL1、PARK5、SNCA、NACP、PARK1、PARK4、PRKN、PARK2、PDJ、DBH、NDUFV2）；雷特综合征（Rett syndrome）（MECP2、RTT、PPMX、MRX16、MRX79、CDKL5、STK9、MECP2、RTT、PPMX、MRX16、MRX79、x-突触核蛋白、DJ-1）；精神分裂症（神经调节蛋白 1（Nrg1）、Erb4（神经调节蛋白受体）、复蛋白 1（Cplx1）、Tph1 色氨酸羟化酶、Tph2、色氨酸羟化酶 2、轴突蛋白 1、GSK3、GSK3a、GSK3b、5-HTT（Slc6a4）、COMT、DRD（Drd1a）、SLC6A3、DAOA、DTNBP1、Dao（Dao1））；分泌酶相关障碍（APH-1（α 和 β）、早老素（Psen1）、呆蛋白，（Ncstn）、PEN-2、Nos1、Parp1、Nat1、Nat2）；三核苷酸重复障碍（HTT（亨廷顿氏 Dx）、SBMA/SMAX1/AR（肯尼迪氏 Dx）、FXN/X25（弗里德里希共济失调）、ATX3（马查多-约瑟夫氏 Dx）、ATXN1 和 ATXN2（脊髓小脑共济失调）、DMPK（强直性肌营养不良）、Atrophin-1 和 Atn1（DRPLA Dx）、CBP（Creb-BP -整体不稳定性）、VLDLR（阿尔茨海默病）、Atxn7、Atxn10）。
眼部疾病以及障碍	年龄相关性黄斑变性（Abcr、Ccl2、Cc2、cp（铜蓝蛋白）、Timp3、组织蛋白酶 D、Vldlr、Ccr2）；白内障（CRYAA、CRYA1、CRYBB2、CRYB2、PITX3、BFSP2、CP49、CP47、CRYAA、CRYA1、PAX6、AN2、MGDA、CRYBA1、CRYB1、CRYGC、CRYG3、CCL、LIM2、MP19、CRYGD、CRYG4、BFSP2、CP49、CP47、HSF4、CTM、HSF4、CTM、MIP、AQP0、CRYAB、CRYA2、CTPP2、CRYBB1、CRYGD、CRYG4、CRYBB2、CRYB2、CRYGC、CRYG3、CCL、CRYAA、CRYA1、GJA8、CX50、CAE1、GJA3、CX46、CZP3、CAE3、CCM1、CAM、KRIT1）；角膜混浊和营养不良（APOA1、TGFBI、CSD2、CDGG1、CSD、BIGH3、CDG2、TACSTD2、TROP2、M1S1、VSX1、RINX、PPCD、PPD、KTCN、COL8A2、FECD、PPCD2、PIP5K3、CFD）；角膜先天性扁平（Cornea plana congenital）（KERA、CNA2）；青光眼（MYOC、TIGR、GLC1A、JOAG、GPOA、OPTN、GLC1E、FIP2、

[0796]

	HYPL、NRP、CYP1B1、GLC3A、OPA1、NTG、NPG、CYP1B1、GLC3A）；莱伯（Leber）先天性黑朦（CRB1、RP12、CRX、CORD2、CRD、RPGRIP1、LCA6、CORD9、RPE65、RP20、AIPL1、LCA4、GUCY2D、GUC2D、LCA1、CORD6、RDH12、LCA3）；黄斑营养不良（ELOVL4、ADMD、STGD2、STGD3、RDS、RP7、PRPH2、PRPH、AVMD、AOFMD、VMD2）。

[0797] 表C：

细胞功能	基因
PI3K/AKT信号传导	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;
	HSP90AA1; RPS6KB1
ERK/MAPK信号传导	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
糖皮质激素	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;

[0798]

[0799]

受体	
信号传导	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
轴突指导信号传导	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA
肝配蛋白受体信号传导	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;

[0800]

	AKT3；SGK
肌动蛋白细胞骨架	ACTN4；PRKCE；ITGAM；ROCK1；ITGA5；IRAK1；
信号传导	PRKAA2；EIF2AK2；RAC1；INS；ARHGEF7；GRK6；
	ROCK2；MAPK1；RAC2；PLK1；AKT2；PIK3CA；CDK8；
	PTK2；CFL1；PIK3CB；MYH9；DIAPH1；PIK3C3；MAPK8；
	F2R；MAPK3；SLC9A1；ITGA1；KRAS；RHOA；PRKCD；
	PRKAA1；MAPK9；CDK2；PIM1；PIK3C2A；ITGB7；
	PPP1CC；PXN；VIL2；RAF1；GSN；DYRK1A；ITGB1；
	MAP2K2；PAK4；PIP5K1A；PIK3R1；MAP2K1；PAK3；
	ITGB3；CDC42；APC；ITGA2；TTK；CSNK1A1；CRKL；
	BRAF；VAV3；SGK
亨廷顿病	PRKCE；IGF1；EP300；RCOR1；PRKCZ；HDAC4；TGM2；
信号传导	MAPK1；CAPNS1；AKT2；EGFR；NCOR2；SP1；CAPN2；
	PIK3CA；HDAC5；CREB1；PRKCI；HSPA5；REST；
	GNAQ；PIK3CB；PIK3C3；MAPK8；IGF1R；PRKD1；
	GNB2L1；BCL2L1；CAPN1；MAPK3；CASP8；HDAC2；
	HDAC7A；PRKCD；HDAC11；MAPK9；HDAC9；PIK3C2A；
	HDAC3；TP53；CASP9；CREBBP；AKT1；PIK3R1；
	PDPK1；CASP1；APAF1；FRAP1；CASP2；JUN；BAX；
	ATF4；AKT3；PRKCA；CLTC；SGK；HDAC6；CASP3
凋亡信号传导	PRKCE；ROCK1；BID；IRAK1；PRKAA2；EIF2AK2；BAK1；
	BIRC4；GRK6；MAPK1；CAPNS1；PLK1；AKT2；IKBKB；
	CAPN2；CDK8；FAS；NFKB2；BCL2；MAP3K14；MAPK8；
	BCL2L1；CAPN1；MAPK3；CASP8；KRAS；RELA；
	PRKCD；PRKAA1；MAPK9；CDK2；PIM1；TP53；TNF；
	RAF1；IKBKG；RELB；CASP9；DYRK1A；MAP2K2；
	CHUK；APAF1；MAP2K1；NFKB1；PAK3；LMNA；CASP2；
	BIRC2；TTK；CSNK1A1；BRAF；BAX；PRKCA；SGK；
	CASP3；BIRC3；PARP1
B 细胞受体信号传导	RAC1；PTEN；LYN；ELK1；MAPK1；RAC2；PTPN11；
	AKT2；IKBKB；PIK3CA；CREB1；SYK；NFKB2；CAMK2A；
	MAP3K14；PIK3CB；PIK3C3；MAPK8；BCL2L1；ABL1；

[0801]

	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
白细胞外渗	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
信号传导	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
整合素信号传导	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
急性期应答	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
信号传导	AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL1R1; IL6
PTEN 信号传导	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;

[0802]

	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 信号传导	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
芳烃受体	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;
信号传导	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
异生物质代谢	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
信号传导	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP90AA1
SAPK/JNK 信号传导	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;

[0803]

	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
PPAr/RXR 信号传导	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
NF-KB 信号传导	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6;
	TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL1R1
神经调节蛋白信号传导	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & β 连环蛋白	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
信号传导	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;

	AKT3; SOX2
胰岛素受体信号传导	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1;
	PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3;
	MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1;
	SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN;
	MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK;
	RPS6KB1
IL-6 信号传导	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3;
	MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1;
	MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG;
	RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3;
	MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
肝胆汁淤积	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA;
	RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8;
	PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1;
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8;
	CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4;
	JUN; IL1R1; PRKCA; IL6
IGF-1 信号传导	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; CTGF; RPS6KB1
NRF2 介导的氧化	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
应激反应	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;

[0804]

	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
肝纤维化/肝	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
星形细胞激活	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR 信号传导	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
FcεRI 信号传导	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
G-蛋白偶联	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
受体信号传导	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
磷酸肌醇	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
代谢	MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF 信号	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;

[0805]

[0806]

传导	
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF 信号传导	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;
	VEGFA; AKT3; FOXO1; PRKCA
自然杀伤细胞信号传导	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
细胞周期: G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
检查点调节	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	GSK3B; RBL1; HDAC6
T 细胞受体信号传导	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
死亡受体信号传导	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3

[0807]

FGF 信号传导	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
GM-CSF 信号传导	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
肌萎缩性脊髓侧索	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
硬化症信号传导	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat 信号传导	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1
烟酸和烟酰胺	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
代谢	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
趋化因子信号传导	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 信号传导	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;

[0808]

	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
突触长时程	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
抑制	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
雌激素受体	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
信号传导	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
蛋白质泛素化	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
途径	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
IL-10 信号传导	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
VDR/RXR 激活	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
TGF-β 信号传导	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll 样受体信号传导	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;

[0809]

	NFKB1; TLR2; JUN
p38 MAPK 信号传导	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1;
	SRF; STAT1
神经营养蛋白/TRK 信号传导	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4
FXR/RXR 激活	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
突触长时程	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
增强	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
钙信号传导	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
EGF 信号传导	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
心血管系统	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
中的缺氧信号传导	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 介导的	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
RXR 功能的抑制	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;

[0810]

	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR 激活	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;
	SREBF1; IL1R1; CCL2; IL6; MMP9
淀粉样蛋白处理	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 信号传导	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
细胞周期: G2/M DNA	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
损伤检查点	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
调节	PRKDC; ATM; SFN; CDKN2A
心血管系统	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
中的一氧化氮信号传导	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
嘌呤代谢	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
cAMP 介导的信号传导	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
线粒体功能异常	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Notch 信号传导	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
内质网	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
应激途径	EIF2AK3; CASP3
嘧啶代谢	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;

	NT5E；POLD1；NME1
帕金森病信号传导	UCHL1；MAPK8；MAPK13；MAPK14；CASP9；PARK7；
	PARK2；CASP3
心脏&β肾上腺素能信号传导	GNAS；GNAQ；PPP2R1A；GNB2L1；PPP2CA；PPP1CC；
	PPP2R5C
糖酵解/糖异生	HK2；GCK；GPI；ALDH1A1；PKM2；LDHA；HK1
干扰素信号传导	IRF1；SOCS1；JAK1；JAK2；IFITM1；STAT1；IFIT3
音猬因子信号传导	ARRB2；SMO；GLI2；DYRK1A；GLI1；GSK3B；DYRK1B
甘油磷脂	PLD1；GRN；GPAM；YWHAZ；SPHK1；SPHK2
代谢	
磷脂降解	PRDX6；PLD1；GRN；YWHAZ；SPHK1；SPHK2
色氨酸代谢	SIAH2；PRMT5；NEDD4；ALDH1A1；CYP1B1；SIAH1
赖氨酸降解	SUV39H1；EHMT2；NSD1；SETD7；PPP2R5C
核苷酸切除修复	ERCC5；ERCC4；XPA；XPC；ERCC1
途径	
淀粉和蔗糖	UCHL1；HK2；GCK；GPI；HK1
代谢	
氨基糖代谢	NQO1；HK2；GCK；HK1
花生四烯酸	PRDX6；GRN；YWHAZ；CYP1B1
代谢	
昼夜节律信号传导	CSNK1E；CREB1；ATF4；NR1D1
凝血系统	BDKRB1；F2R；SERPINE1；F3
多巴胺受体	PPP2R1A；PPP2CA；PPP1CC；PPP2R5C
信号传导	
谷胱甘肽代谢	IDH2；GSTP1；ANPEP；IDH1
甘油脂代谢	ALDH1A1；GPAM；SPHK1；SPHK2
亚油酸代谢	PRDX6；GRN；YWHAZ；CYP1B1

[0811]

[0812]

甲硫氨酸代谢	DNMT1；DNMT3B；AHCY；DNMT3A
丙酮酸代谢	GLO1；ALDH1A1；PKM2；LDHA
精氨酸和脯氨酸	ALDH1A1；NOS3；NOS2A
代谢	
类花生酸信号传导	PRDX6；GRN；YWHAZ
果糖和甘露糖	HK2；GCK；HK1
代谢	
半乳糖代谢	HK2；GCK；HK1
芪、香豆素和	PRDX6；PRDX1；TYR
木质素生物合成	
抗原呈递	CALR；B2M
途径	
类固醇的生物合成	NQO1；DHCR7
丁酸代谢	ALDH1A1；NLGN1
柠檬酸循环	IDH2；IDH1
脂肪酸代谢	ALDH1A1；CYP1B1
甘油磷脂	PRDX6；CHKA
代谢	
组氨酸代谢	PRMT5；ALDH1A1
肌醇代谢	ERO1L；APEX1
细胞色素p450	GSTP1；CYP1B1
对异生物质的代谢	
甲烷代谢	PRDX6；PRDX1
苯丙氨酸代谢	PRDX6；PRDX1
丙酸代谢	ALDH1A1；LDHA
含硒氨基酸	PRMT5；AHCY

[0813]

代谢	
鞘脂代谢	SPHK1；SPHK2
氨基膦酸	PRMT5
代谢	
雄激素和雌激素	PRMT5
代谢	
抗坏血酸和Aldarate	ALDH1A1
代谢	
胆汁酸生物合成	ALDH1A1
半胱氨酸代谢	LDHA
脂肪酸生物合成	FASN
谷氨酸受体	GNB2L1
信号传导	
NRF2 介导的氧化	PRDX1
应激反应	
磷酸戊糖	GPI
途径	
戊糖和葡糖醛酸	UCHL1
相互转化	
视黄醇代谢	ALDH1A1
核黄素代谢	TYR
酪氨酸代谢	PRMT5，TYR
泛醌生物合成	PRMT5
缬氨酸、亮氨酸和	ALDH1A1
异亮氨酸降解	
甘氨酸、丝	CHKA

[0814]

氨酸和	
苏氨酸代谢	
赖氨酸降解	ALDH1A1
疼痛/味觉	TRPM5; TRPA1
疼痛	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
线粒体功能	AIF; CytC; SMAC (Diablo) ; Aifm-1; Aifm-2
发育神经学	BMP-4; 脊索素 (Chordin) (Chrd) ; 头发生素 (Nog) ; WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16) ; β-连环蛋白;
	Dkk-1; 卷曲相关蛋白; Otx-2; Gbx2; FGF-8;
	颤蛋白 (Reelin) ; Dab1; unc-86 (Pou4f1 或 Brn3a) ; Numb; Reln

[0815] 本发明的实施例还涉及与敲除基因、扩增基因以及修复与DNA重复不稳 定性和神经障碍相关的具体突变有关的方法和组合物(罗伯特D.•威尔斯 (Robert D.Wells)、芦沢哲夫(Tetsuo Ashizawa),遗传不稳定性与神经疾 病(Genetic Instabilities and Neurological Diseases),第二版,学术出版社 (Academic Press),2011年10月13日-《医学》(Medical))。已经发现 串联重复序列的特定方面对超过二十种人类疾病负责(重复不稳定新见：RNA•DNA杂交体的作用(New insights into repeat instability:role of RNA•DNA hybrids),麦基弗EI(McIvor EI)、波拉克U(Polak U)、纳皮 尔拉拉M(Napierala M),《RNA生物学》(RNA Biol.)2010年9月-10 月;7(5):551-8)。可以利用CRISPR-Cas系统修正基因组不稳定性缺陷。

[0816] 本发明的一个另外的方面涉及利用CRISPR-Cas系统修正EMP2A和 EMP2B基因缺陷,这些基因已经被鉴定为与拉福拉病(Lafora disease)相 关。拉福拉病是一种由可以作为青年期的癫痫发作而开始的进行性肌阵挛性 癫痫表征的常染色体隐性病症。该疾病的少数病可以由尚未鉴定的基因的突 变引起。该疾病引起发作、肌肉痉挛、行走困难、痴呆,并且最终引起死 亡。当前没有疗法被证明有效对抗疾病进展。与癫痫相关的其他遗传异常还 可以靶向CRISPR-Cas系统并且基础遗传学进一步描述于由朱利亚诺阿文济尼(Giuliano Avanzini)、杰弗里L.诺贝尔斯(Jeffrey L.Noebels)编辑的《癫 痫与遗传癫痫遗传学》(Genetics of Epilepsy and Genetic Epilepsies),马里亚 尼儿科神经学基金会(Mariani Foundation Paediatric Neurology):20;2009) 中。

[0817] 转让给加莫生物科技公司(Sangamo BioSciences,Inc.)的美国专利公开 号20110158957的涉及使T细胞受体(TCR)基因失活的方法也可以被修饰 为本发明的CRISPR Cas系统。在另一个实例中,转让给加莫生物科技公司的 美国专利公开号20100311124和转让给Cellectis公司的美国专利公开号 20110225664的方法(两者都涉及使谷氨酰胺合成酶基因表达基因失活)也可 以被修饰为本发明的CRISPR Cas系统。。

[0818] 本发明的若干另外的方面涉及修正与范围广泛的遗传性疾病相关的缺 陷,这些遗传性疾病在专题小节遗传性障碍(Genetic Disorders)下被进一步 描述于国立卫生研究院的网站(网址为 health.nih.gov/topic/GeneticDisorders)。遗传性脑病可以包括但不限于,肾上 腺脑白质营养不良、胼胝体发育不全、艾卡尔迪综合征(Aicardi Syndrome)、阿尔佩斯病(Alpers'Disease)、阿尔茨海默病、巴特综合征 (Barth Syndrome)、巴藤病(Batten Disease)、CADASIL、小脑变性、费 波瑞病(Fabry's Disease)、格斯特曼-施特劳斯纳病(Gerstmann-Straussler- Scheinker Disease)、亨廷顿病以及其他三联体重复障碍、莱氏病(Leigh's Disease)、莱施-奈恩综合征、门克斯病、线粒体肌病以及NINDS空洞脑(Colpocephaly)。这些疾病在小节遗传性脑部障碍(Genetic Brain Disorders)下被进一步描述于国立卫生研究院的网站。

[0819] 在一些实施例中,该病症可以是瘤形成。在一些实施例中,在该病症是 瘤形成的情况下,有待被靶向的基因是列于表A中的那些基因中的任一种 (在这种情况下是PTEN等)。在一些实施例中,该病症可以是年龄相关性 黄斑变性。在一些实施例中,该病症可以是一种精神分裂症障碍。在一些实 施例中,该病症可以是一种三核苷酸重复障碍。在一些实施例中,该病症可 以是脆性X综合征。在一些实施例中,该病症可以是一种分泌酶相关障碍。 在一些实施例中,该病症可以是一种朊病毒相关障碍。在一些实施例中,该 病症可以是ALS。在一些实施例中,该病症可以是一种药物成瘾。在一些实 施例中,该病症可以是自闭症。在一些实施例中,该病症可以是阿尔茨海默 病。在一些实施例中,该病症可以是炎症。在一些实施例中,该病症可以是 帕金森病。

[0820] 例如,美国专利公开号20110023145描述了使用锌指核酸酶基因修饰与 自闭症谱系障碍(ASD)相关的细胞、动物以及蛋白质。自闭症谱系障碍 (ASD)是一组由社会交往和沟通的质量损伤以及行为、兴趣和活动的限制 性重复和刻板性模式表征的障碍。这三种障碍自闭症、阿斯佩格综合征 (AS)和未另行规定的广泛性发育障碍(PDD-NOS)是具有不同的严重程 度、相关智力功能和医疗条件的同一障碍的连续体。ASD是主要由遗传决定 的障碍,具有大约90%遗传力。

[0821] 美国专利公开号20110023145包括编辑任何编码与ASD相关的蛋白质的 染色体序列,这些序列可以被应用于本发明的CRISPR Cas系统。典型地基于 与ASD相关的蛋白质与ASD的发生率或适应症的实验相关性选择与ASD相 关的蛋白质。例如,相对于缺少ASD的群体,在具有ASD的群体中,与 ASD相关的蛋白质的产生率或循环浓度可以升高或降低。可以使用蛋白质组 技术评估蛋白质水平差异,包括但不限于蛋白质印迹、免疫组织化学染色、酶联免疫吸附测定(ELISA)以及质谱法。可替代地,可以使用基因组技术 通过获得编码这些蛋白质的基因的基因表达谱而鉴定与ASD相关的蛋白质, 包括但不限于DNA微阵列分析、基因表达序列分析(SAGE)以及定量实时 聚合酶链式反应(Q-PCR)。

[0822] 可能和与ASD相关的蛋白质相关的疾病状态或障碍的非限制性实例包括 自闭症、阿斯佩格综合征(AS)、未另行规定的广泛性发育障碍(PDD- NOS)、雷特综合征、结节性硬化、苯丙酮尿症、史-伦-奥三氏综合征 (Smith-Lemli-Opitz syndrome)以及脆性X综合征。通过非限制性实例的方 式,与ASD相关的蛋白质包括但不限于以下蛋白质:ATP10C氨基磷脂-MET MET受体转运ATP酶酪氨酸激酶(ATP10C)BZRAP1 MGLUR5(GRM5) 代谢型谷氨酸受体5(MGLUR5)CDH10钙粘素-10 MGLUR6(GRM6)代谢 型谷氨酸受体6(MGLUR6)CDH9钙粘素-9

NLGN1神经连接蛋白-1 CNTN4 接触蛋白-4 NLGN2神经连接蛋白-2 CNTNAP2接触蛋白相关的SEMA5A神经 连接蛋白-3蛋白样2(CNTNAP2)DHCR7 7-脱氢胆固醇NLGN4X神经连接 蛋白-4X-还原酶(DHCR7)连接的DOC2A双重C2样-结构域-NLGN4Y神经 连接蛋白-4Y-包含蛋白α连接的DPP6二肽基NLGN5神经连接蛋白-5氨肽酶 样蛋白6EN2锯齿状2(EN2)NRCAM神经细胞粘附分子(NRCAM) MDGA2脆性X智力迟钝NRXN1轴突蛋白-1 1(MDGA2)FMR2(AFF2) AF4/FMR2家族成员2 OR4M2嗅觉受体(AFF2)4M2 FOXP2叉头框蛋白P2 OR4N4嗅觉受体(FOXP2)4N4 FXR1脆性X智力OXTR催产素受体阻滞，常染色体(OXTR)同系物1(FXR1)FXR2脆性X智力PAH苯丙氨酸阻 滞,常染色体羟化酶(PAH)同系物2(FXR2)GABRA1γ-氨基丁酸PTEN 磷酸酶和受体亚基α-1张力蛋白同源物(GABRA1)(PTEN)GABRA5 GABAA(.γ.-氨基丁酸PTPRZ1受体类型酸)受体α5酪氨酸蛋白亚基 (GABRA5)磷酸酶ζ(PTPRZ1)GABRB1γ-氨基丁酸RELN颤蛋白受体亚 基β-1(GABRB1)GABRB3 GABAA(.γ.-氨基丁酸RPL10 60S核糖体酸)受 体.β.3亚基蛋白L10(GABRB3)GABRG1γ-氨基丁酸SEMA5A脑信号蛋白 (Semaphorin)-5A受体亚基γ-1(SEMA5A)(GABRG1)HIRIP3 HIRA-相 互作用蛋白3 SEZ6L2发作相关6同系物(小鼠)样2 HOXA1同源盒蛋白Hox-A1 SHANK3 SH3和多重(HOXA1)锚蛋白重复结构域3(SHANK3) IL6白介素-6 SHBZRAP1 SH3和多重锚蛋白重复结构域3(SHBZRAP1) LAMB1层粘连蛋白亚基β-1 SLC6A4血清素(LAMB1)转运体(SERT) MAPK3丝裂原激活蛋白TAS2R1味觉受体激酶3类型2成员1 TAS2R1 MAZ Myc相关的锌指TSC1结节性硬化蛋白的蛋白1 MDGA2 MAM结构域包含TSC2结节性硬化糖基磷脂酰肌醇蛋白2锚定物2(MDGA2)MECP2甲基 CpG结合UBE3A泛素蛋白蛋白2(MECP2)连接酶E3A(UBE3A)MECP2 甲基CpG结合WNT2无翅型蛋白2(MECP2)MMTV整合位点家族,成员2 (WNT2)。

[0823] 与ASD相关的其染色体序列被编辑的蛋白质的一致性可以并且将会变 化。在优选实施例中,与ASD相关的其染色体序列被编辑的蛋白质可以是由 BZRAP1基因编码的苯二氮卓受体(外周)相关蛋白1(BZRAP1)、由 AFF2基因编码的AF4/FMR2家族成员2蛋白(AFF2)(亦称MFR2)、由 FXR1基因编码的脆性X智力迟钝常染色体同系物1蛋白(FXR1)、由 FXR2基因编码的脆性X智力迟钝常染色体同系物2蛋白(FXR2),由 MDGA2基因编码的含MAM结构域的糖基磷脂酰肌醇锚定物2蛋白 (MDGA2)、由MECP2基因编码的甲基CpG结合蛋白2(MECP2)、由MGLUR5-1基因编码的代谢型谷氨酸受体5(MGLUR5)(亦称GRM5)、由NRXN1基因编码的轴突蛋白1蛋白或由SEMA5A基因编码的脑信号蛋白- 5A蛋白(SEMA5A)。在一个示例性实施例中,该基因修饰动物是大鼠,并 且编码与ASD相关的蛋白质的编辑的染色体序列如下所列:BZRAP1苯二氮 卓受体XM_002727789,(外周)相关XM_213427,蛋白1(BZRAP1) XM_002724533,XM_001081125 AFF2(FMR2)AF4/FMR2家族成员2 XM_219832,(AFF2)XM_001054673 FXR1脆性X智力NM_001012179迟 钝,常染色体同系物1(FXR1)FXR2脆性X智力NM_001100647迟钝,常 染色体同系物2(FXR2)MDGA2含MAM结构域的NM_199269糖基磷脂酰肌醇锚定物2(MDGA2)MECP2甲基CpG结合NM_022673蛋白2 (MECP2)MGLUR5代谢型谷氨酸NM_017012(GRM5)受体5 (MGLUR5)NRXN1轴突蛋白-1NM_021767SEMA5A脑信号蛋白-5A(SEMA5A)NM_001107659。

[0824] 示例性动物或细胞可以包括一、二、三、四、五、六、七、八或九个或 更多个失活的编码与ASD相关的蛋白质的染色体序列,以及零、一、二、 三、四、五、六、七、八、九个或更多个编码与ASD相关的蛋白质的染色体 整合的序列。编辑的或整合的染色体序列可以被修饰

为编码一种改变的与 ASD相关的蛋白质。与ASD相关的蛋白质的突变的非限制性实例包括轴突蛋 白1中的L18Q突变,其中位置18处的亮氨酸被谷氨酰胺替换;神经连接蛋 白3中的R451C突变,其中位置451处的精氨酸被半胱氨酸替换;神经连接 蛋白4中的R87W突变,其中位置87处的精氨酸被色氨酸替换;以及血清素 转运体中的I425V突变,其中位置425处的异亮氨酸被缬氨酸替换。ASD相 关染色体序列中的许多其他突变和染色体重排已经与ASD相关并且在本领域 是已知的。参见例如,弗赖塔格(Freitag)等人(2010)《欧洲儿童与青少年精神病学》(Eur.Child.Adolesc.Psychiatry)19:169-178,以及布坎 (Bucan)等人(2009)《PLoS遗传学》(PLoS Genetics)5:e1000536,将 其披露通过引用以其全文结合在此。

[0825] 与帕金森病相关的蛋白质的实例包括但不限于α-突触核蛋白、DJ-1、 LRRK2、PINK1、帕金蛋白、UCHL1、Synphilin-1以及NURR1。

[0826] 成瘾相关的蛋白质的实例可以包括例如ABAT。

[0827] 炎症相关蛋白的实例可以包括例如由Ccr2基因编码的单核细胞趋化蛋白 -1(MCP1)、由Ccr5基因编码的C-C趋化因子受体类型5(CCR5)、由 Fcgr2b基因编码的IgG受体IIB(FCGR2b,亦称CD32)或由Fcer1g基因编 码的FcεR1g(FCER1g)蛋白。

[0828] 心血管疾病相关蛋白的实例可以包括例如IL1B(白介素1,β)、XDH (黄嘌呤脱氢酶)、TP53(肿瘤蛋白p53)、PTGIS(前列腺素I2(前列腺环 素)合酶)、MB(肌红蛋白)、IL4(白介素4)、ANGPT1(血管生成素 1)、ABCG8(ATP-结合盒,亚家族G(WHITE),成员8)或CTSK(组织蛋白酶K)。

[0829] 例如,美国专利公开号20110023153描述了使用锌指核酸酶基因修饰与 阿尔茨海默病相关的细胞、动物以及蛋白质。曾经修饰的细胞和动物可以进 一步使用已知的用于研究靶向突变对AD的发展和/或进展的影响的方法,使 用AD研究中常用的措施进行测试-如但不限于,学习和记忆、焦虑、抑郁、 成瘾、感觉运动功能以及测量行为、功能、病理、代谢和生化功能的测定。

[0830] 本披露包括编码与AD相关的蛋白质的任何染色体序列的编辑。典型地 基于AD相关蛋白与AD障碍的实验相关性选择AD相关蛋白。例如,相对于 缺少AD障碍的群体,在具有AD障碍的群体中,AD相关蛋白的产生率或循 环浓度可以升高或降低。可以使用蛋白质组技术评估蛋白质水平差异,包括 但不限于蛋白质印迹、免疫组织化学染色、酶联免疫吸附测定(ELISA)以 及质谱法。可替代地,可以使用基因组技术通过获得编码这些蛋白质的基因的基因表达谱而鉴定AD相关蛋白,包括但不限于DNA微阵列分析、基因表 达序列分析(SAGE)以及定量实时聚合酶链式反应(Q-PCR)。

[0831] 阿尔茨海默病相关蛋白的实例可以包括例如由VLDLR基因编码的极低 密度脂蛋白受体蛋白(VLDLR)、由UBA1基因编码的泛素样改性剂激活酶 1(UBA1)或由UBA3基因编码的NEDD8-激活酶E1催化亚基蛋白(UBE1C)。

[0832] 通过非限制性实例的方式,与AD相关的蛋白质包括但不限于如下所列 的蛋白质:染色体序列编码蛋白ALAS2δ-氨基酮戊酸合酶2(ALAS2) ABCA1 ATP-结合盒转运体(ABCA1) ACE血管紧张素I-转化酶(ACE) APOE载脂蛋白E前体(APOE)APP淀粉样蛋白前体蛋白(APP) AQP1水 通道蛋白1(AQP1)BIN1 Myc盒依赖性相互作用蛋白1或桥接整合蛋白1 (BIN1)BDNF脑源性神经营养因子(BDNF)BTNL8嗜乳脂蛋白样蛋白8 (BTNL8)C10RF49 1号染色体开放阅

读框49 CDH4钙粘素-4 CHRNB2神经 元乙酰胆碱受体亚基β-2 CKLFSF2 CKLF样MARVEL跨膜结构域-包含蛋白2 (CKLFSF2)CLEC4E C型凝集素结构域家族4,成员e(CLEC4E)CLU簇 集蛋白(也称为载脂蛋白J)CR1红细胞补体受体1(CR1,也称为CD35、C3b/C4b受体和免疫粘附受体)CR1L红细胞补体受体1(CR1L)CSF3R粒 细胞集落刺激因子3受体(CSF3R)CST3胱抑素C或胱抑素3 CYP2C细胞 色素P450 2C DAPK1死亡相关蛋白激酶1(DAPK1)ESR1雌激素受体1FCAR IgA受体的Fc片段(FCAR,也称为CD89)FCGR3B IgG的Fc片 段,低亲和性IIIb,受体(FCGR3B或CD16b)FFA2游离脂肪酸受体2 (FFA2)FGA纤维蛋白原(因子I)GAB2 GRB2-相关结合蛋白2(GAB2) GAB2 GRB2-相关结合蛋白2(GAB2)GALP甘丙肽样肽GAPDHS甘油醛-3- 磷酸脱氢酶,精子发生的(GAPDHS)GMPB GMBP HP结合珠蛋白(HP) HTR7 5-羟色胺(血清素)受体7(腺苷酸环化酶偶联的)IDE胰岛素降解酶 IF127 IF127 IFI6干扰素,α-可诱导蛋白6(IFI6)IFIT2具有三角形四肽重复 的干扰素诱导的蛋白2(IFIT2)IL1RN白介素-1受体拮抗剂(IL-1RA) IL8RA白介素8受体,α(IL8RA或CD181)IL8RB白介素8受体,β (IL8RB)JAG1锯齿状1(JAG1)KCNJ15钾内向整流通道,亚家族J,成员 15(KCNJ15)LRP6低密度脂蛋白受体相关蛋白6(LRP6)MAPT微管相关 蛋白tau(MAPT)MARK4 MAP/微管亲和性调节激酶4(MARK4)MPHOSPH1 M期磷蛋白1 MTHFR 5,10-亚甲基四氢叶酸还原酶MX2干扰素 诱导的GTP-结合蛋白Mx2 NBN Nibrin,也称为NBN NCSTN呆蛋白NIACR2 烟酸受体2(NIACR2,也称为GPR109B)NMNAT3烟酰胺核苷酸腺苷酰转 移酶3NTM Neurotrimin(或HNT)ORM1血清类粘蛋白(Orosmucoid)1 (ORM1)或α-1-酸性糖蛋白1 P2RY13 P2Y嘌呤受体13(P2RY13)PBEF1 烟酰胺磷酸核糖转移酶(NAmPRTase或Nampt)也称为前B细胞集落增强因 子1(PBEF1)或内脂素PCK1磷酸烯醇丙酮酸羧激酶PICALM磷脂酰肌醇结 合网格蛋白装配蛋白(PICALM)PLAU尿激酶型纤维蛋白溶酶原激活剂 (PLAU)PLXNC1丛状蛋白(Plexin)C1(PLXNC1)PRNP朊病毒蛋白PSEN1早老素1蛋白(PSEN1)PSEN2早老素2蛋白(PSEN2)PTPRA蛋白 质酪氨酸磷酸酶受体A型蛋白(PTPRA)RALGPS2具有PH结构域和SH3结 合基序的Ral GEF 2(RALGPS2)RGSL2 G-蛋白信号传导样调节剂2 (RGSL2)SELENBP1含硒结合蛋白1(SELNBP1)SLC25A37线粒体铁转 运蛋白(Mitoferrin)-1 SORL1含分拣蛋白相关受体L(DLR类别)A重复的 蛋白质(SORL1)TF转铁蛋白TFAM线粒体转录因子A TNF肿瘤坏死因子 TNFRSF10C肿瘤坏死因子受体超家族成员10C(TNFRSF10C)TNFSF10肿 瘤坏死因子受体超家族,(TRAIL)成员10a(TNFSF10)UBA1泛素样改性 剂激活酶1(UBA1)UBA3 NEDD8-激活酶E1催化亚基蛋白(UBE1C)UBB 泛素B蛋白(UBB)UBQLN1泛醌蛋白(Ubiquilin)-1 UCHL1泛素羧基末端 酯酶L1蛋白(UCHL1)UCHL3泛素羧基末端水解酶同工酶L3蛋白 (UCHL3)VLDLR极低密度脂蛋白受体蛋白(VLDLR)

[0833] 在示例性实施例中,与AD相关的其染色体序列被编辑的蛋白质可以是 由VLDLR基因编码的极低密度脂蛋白受体蛋白(VLDLR)、由UBA1基因 编码的泛素样改性剂激活酶1(UBA1)、由UBA3基因编码的NEDD8-激活 酶E1催化亚基蛋白(UBE1C)、由AQP1基因编码的水通道蛋白1 (AQP1)、由UCHL1基因编码的泛素羧基末端酯酶L1蛋白(UCHL1)、 由UCHL3基因编码的泛素羧基末端水解酶同工酶L3蛋白(UCHL3)、由 UBB基因编码的泛素B蛋白(UBB)、由MAPT基因编码的微管相关蛋白 tau(MAPT)、由PTPRA基因编码的蛋白质酪氨酸磷酸酶受体A型蛋白 (PTPRA)、由PICALM基因编码的磷脂酰肌醇结合网格蛋白装配蛋白 (PICALM)、由CLU基因编码的簇集蛋白(也称为载脂蛋白J)、由 PSEN1基因编码的早老素1蛋白、由PSEN2基因编码的早老素2蛋白、由 SORL1基因编码的含分拣蛋白相关受体L(DLR类别)A重复的蛋

白质 (SORL1)蛋白质、由APP基因编码的淀粉样蛋白前体蛋白(APP)、由 APOE基因编码的载脂蛋白E前体(APOE)或由BDNF基因编码的脑源性神 经营养因子(BDNF)。在一个示例性实施例中,该基因修饰动物是大鼠,并 且编码与AD相关的蛋白质的编辑的染色体序列如下:APP淀粉样蛋白前体 蛋白(APP)NM_019288 AQP1水通道蛋白1(AQP1)NM_012778 BDNF脑 源性神经营养因子NM_012513 CLU簇集蛋白(也称为NM_053021载脂蛋白 J)MAPT微管相关蛋白NM_017212 tau(MAPT)PICALM磷脂酰肌醇结合 NM_053554网格蛋白装配蛋白(PICALM)PSEN1早老素1蛋白(PSEN1) NM_019163 PSEN2早老素2蛋白(PSEN2)NM_031087 PTPRA酪氨酸磷酸 酶NM_012763受体A型蛋白(PTPRA)SORL1含分拣蛋白相关受体L (DLR NM_053519,类别)A重复的XM_001065506,蛋白质(SORL1) XM_217115 UBA1泛素样改性剂激活NM_001014080酶1(UBA1)UBA3 NEDD8-激活酶E1 NM_057205催化亚基蛋白(UBE1C)UBB泛素B蛋白(UBB)NM_138895 UCHL1泛素羧基末端NM_017237酯酶L1蛋白 (UCHL1)UCHL3泛素羧基末端NM_001110165水解酶同工酶L3蛋白 (UCHL3)VLDLR极低密度脂蛋白NM_013155受体蛋白(VLDLR)。

[0834] 该动物或细胞可以包括1、2、3、4、5、6、7、8、9、10、11、12、 13、14、15个或更多个破坏的、编码与AD相关的蛋白质的染色体序列以及 零、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15个或更多个编码 与AD相关的蛋白质的染色体整合序列。

[0835] 编辑的或整合的染色体序列可以被修饰为编码一种改变的与AD相关的 蛋白质。AD相关染色体序列中的许多突变已经与AD相关。例如,APP中的 V7171(即位置717处的缬氨酸变为异亮氨酸)错义突变引起家族性AD。早 老素-1蛋白中的多重突变,如H163R(即位置163处的组氨酸变为精氨 酸)、A246E(即位置246处的丙氨酸变为谷氨酸)、L286V(即位置286处 的亮氨酸变为缬氨酸)以及C410Y(即位置410处的半胱氨酸变为酪氨酸) 引起家族性3型阿尔茨海默病。早老素-2蛋白中的突变,如N141I(即位置 141处的天冬酰胺变为异亮氨酸)、M239V(即位置239处的甲硫氨酸变为缬 氨酸)以及D439A(即位置439处的天冬氨酸变为丙氨酸)引起家族性4型 阿尔茨海默病。AD相关基因和疾病的遗传变体的其他相关性在本领域是已知 的。参见例如,韦林(Waring)等人(2008)《神经病学年鉴》(Arch.Neurol.)65:329-334,将其披露通过引用以其全文结合在此。

[0836] 自闭症谱系障碍相关蛋白的实例可以包括例如由BZRAP1基因编码的苯 二氮卓受体(外周)相关蛋白1(BZRAP1)、由AFF2基因编码的 AF4/FMR2家族成员2蛋白(AFF2)(亦称MFR2)、由FXR1基因编码的 脆性X智力迟钝常染色体同系物1蛋白(FXR1)或由FXR2基因编码的脆性 X智力迟钝常染色体同系物2蛋白(FXR2)。

[0837] 黄斑变性相关蛋白的实例可以包括例如由ABCR基因编码的ATP-结合 盒,亚家族A(ABC1)成员4蛋白(ABCA4)、由APOE基因编码的载脂 蛋白E蛋白(APOE)或由CCL2基因编码的趋化因子(C-C基序)配体2蛋 白(CCL2)。

[0838] 精神分裂症相关蛋白的实例可以包括NRG1、ErbB4、CPLX1、TPH1、 TPH2、NRXN1、GSK3A、BDNF、DISC1、GSK3B及其组合。

[0839] 涉及肿瘤抑制的蛋白质的实例可以包括例如ATM(共济失调性毛细血管 扩张突变的)、ATR(共济失调性毛细血管扩张和Rad3相关的)、EGFR (表皮生长因子受体)、ERBB2(v-erb-b2红白血病病毒癌基因同系物2)、 ERBB3(v-erb-b2红白血病病毒癌基因同系物3)、ERBB4(v-erb-b2红白血 病病毒癌基因同系物4)、Notch 1、Notch2、Notch 3或Notch 4。

[0840] 与分泌酶障碍相关的蛋白质的实例可以包括例如PSENEN(早老素增强 子2同系物(秀丽隐杆线虫))、CTSB(组织蛋白酶B)、PSEN1(早老素 1)、APP(淀粉样蛋白β(A4)前体蛋白)、APH1B(前咽缺陷1同系物B (秀丽隐杆线虫))、PSEN2(早老素2(阿尔茨海默病4))或BACE1(β- 位点APP-切割酶1)。

[0841] 例如,美国专利公开号20110023146描述了使用锌指核酸酶基因修饰与 分泌酶相关障碍相关的细胞、动物以及蛋白质。分泌酶对于将前蛋白加工成 其生物活性形式是必需的。分泌酶途径的各种组分的缺陷促成许多障碍,特 别是具有标志性淀粉样蛋白生成或淀粉样蛋白斑块的那些,如阿尔茨海默病 (AD)。

[0842] 分泌酶障碍以及与这些障碍相关的蛋白质是一套影响众多障碍的易感 性、障碍的存在、障碍的严重性或其任何组合的相异蛋白质。本披露包括编 码与分泌酶障碍相关的蛋白质的任何染色体序列的编辑。典型地基于分泌酶 相关蛋白质与分泌酶障碍的发展的实验相关性选择与分泌酶障碍相关的蛋白 质。例如,相对于没有分泌酶障碍的群体,在具有分泌酶障碍的群体中,与 分泌酶障碍相关的蛋白质的产生率或循环浓度可以升高或降低。可以使用蛋 白质组技术评估蛋白质水平差异,包括但不限于蛋白质印迹、免疫组织化学 染色、酶联免疫吸附测定(ELISA)以及质谱法。可替代地,可以使用基因 组技术通过获得编码这些蛋白质的基因的基因表达谱而鉴定与分泌酶障碍相 关的蛋白质,包括但不限于DNA微阵列分析、基因表达序列分析(SAGE) 以及定量实时聚合酶链式反应(Q-PCR)。

[0843] 通过非限制性实例的方式,与分泌酶障碍相关的蛋白质包括PSENEN (早老素增强子2同系物(秀丽隐杆线虫))、CTSB(组织蛋白酶B)、 PSEN1(早老素1)、APP(淀粉样蛋白β(A4)前体蛋白)、APH1B(前咽 缺陷1同系物B(秀丽隐杆线虫))、PSEN2(早老素2(阿尔茨海默病 4))、BACE1(β-位点APP-切割酶1)、ITM2B(膜内在蛋白2B)、CTSD (组织蛋白酶D)、NOTCH1(Notch同系物1、易位相关(果蝇))、TNF (肿瘤坏死因子(TNF超家族,成员2))、INS(胰岛素)、DYT10(肌张 力障碍10)、ADAM17(ADAM金属肽酶结构域17)、APOE(载脂蛋白 E)、ACE(血管紧张素I转化酶(肽基二肽酶A)1)、STN(他汀)、 TP53(肿瘤蛋白p53)、IL6(白介素6(干扰素,β2))、NGFR(神经生 长因子受体(TNFR超家族,成员16))、IL1B(白介素1,β)、ACHE (乙酰胆碱酯酶(Yt血型))、CTNNB1(连环蛋白(钙粘素相关蛋白),β 1,88kDa)、IGF1(胰岛素样生长因子1(生长调节素C))、IFNG(干扰 素,γ)、NRG1(神经调节蛋白1)、CASP3(半胱天冬酶3,凋亡相关半胱 氨酸肽酶)、MAPK1(丝裂原激活蛋白激酶1)、CDH1(钙粘素1,1型, E-钙粘素(上皮))、APBB1(淀粉样蛋白β(A4)前体蛋白结合,家族B,成员1(Fe65))、HMGCR(3-羟基-3-甲基戊二酰-辅酶A还原酶)、 CREB1(cAMP反应元件结合蛋白1)、PTGS2(前列腺素内过氧化物合酶2 (前列腺素G/H合酶和环氧合酶))、HES1(长毛和裂口增强蛋白(hairy and enhancer of split)1,(果蝇))、CAT(过氧化氢酶)、TGFB1(转化 生长因子,β1)、ENO2(烯醇酶2(γ,神经元的))、ERBB4(v-erb-a红 白血病病毒癌基因同系物4(禽的))、TRAPPC10(运输蛋白粒子复合物 10)、MAOB(单胺氧化酶B)、NGF(神经生长因子(β多肽))、 MMP12(基质金属肽酶12(巨噬细胞弹性蛋白酶))、JAG1(锯齿状1(艾 欧吉勒综合征(Alagille syndrome)))、CD40LG(CD40配体)、PPARG (过氧化物酶体增殖剂激活的受体γ)、FGF2(成纤维细胞生长因子2(碱性 的))、IL3(白介素3(集落刺激因子,多重的))、LRP1(低密度脂蛋白受体相关蛋白1)、NOTCH4(Notch同系物4(果蝇))、MAPK8(丝裂原 激活蛋白激酶8)、PREP(脯氨酰内肽酶)、NOTCH3(Notch同系物3(果 蝇))、PRNP(朊病毒蛋白)、CTSG(组织蛋白酶G)、

EGF(表皮生长 因子(β-尿抑胃素))、REN(肾素)、CD44(CD44分子(印度血 型))、SELP(选择素P(颗粒膜蛋白140kDa,抗原CD62))、GHR(生 长激素体)、ADCYAP1(腺苷酸环化酶激活多肽1(脑垂体的))、INSR (胰岛素受体)、GFAP(胶质酸性蛋白)、MMP3(基质金属肽酶3(溶基 质素1,前明胶酶(progelatinase)))、MAPK10(丝裂原激活蛋白激酶 10)、SP1(Sp1转录因子)、MYC(v-myc骨髓细胞瘤病病毒癌基因同系物 (禽的))、CTSE(组织蛋白酶E)、PPARA(过氧化物酶体增殖剂激活受 体α)、JUN(jun癌基因)、TIMP1(TIMP金属肽酶抑制剂1)、IL5(白介 素5(集落刺激因子,嗜酸性粒细胞))、IL1A(白介素1,α)、MMP9(基 质金属肽酶9(明胶酶B,92kDa明胶酶,92kDa IV型胶原酶))、HTR4 (5-羟色胺(血清素)受体4)、HSPG2(硫酸乙酰肝素蛋白多糖2)、 KRAS(v-Ki-ras2柯尔斯顿(Kirsten)大鼠肉瘤病毒癌基因同系物)、CYCS(细胞色素c,躯体的)、SMG1(SMG1同系物,磷脂酰肌醇3-激酶相关激 酶(秀丽隐杆线虫))、IL1R1(白介素1受体,I型)、PROK1(前动力蛋 白1)、MAPK3(丝裂原激活蛋白激酶3)、NTRK1(神经营养酪氨酸激酶 受体,1型)、IL13(白介素13)、MME(膜金属内肽酶)、TKT(转酮醇酶)、CXCR2(趋化因子(C-X-C基序)受体2)、IGF1R(胰岛素样生长因 子1受体)、RARA(视黄酸受体,α)、CREBBP(CREB结合蛋白)、 PTGS1(前列腺素内过氧化物合酶1(前列腺素G/H合酶和环氧合酶))、 GALT(半乳糖-1-磷酸尿甙基转移酶(uridylyltransferase))、CHRM1(胆碱能受体,毒蕈碱1)、ATXN1(共济失调蛋白1)、PAWR(PRKC,凋亡, WT1,调节剂)、NOTCH2(Notch同系物2(果蝇))、M6PR(甘露糖-6- 磷酸受体(阳离子依赖性的))、CYP46A1(细胞色素P450,家族46、亚家 族A,多肽1)、CSNK1 D(酪蛋白激酶1,δ)、MAPK14(丝裂原激活蛋 白激酶14)、PRG2(蛋白多糖2,骨髓(自然杀伤细胞激活剂,嗜酸性粒细 胞颗粒主要碱性蛋白))、PRKCA(蛋白激酶C,α)、L1 CAM(L1细胞粘 附分子)、CD40(CD40分子,TNF受体超家族成员5)、NR1I2(核受体亚 家族1,I组,成员2)、JAG2(锯齿状2)、CTNND1(连环蛋白(钙粘素 相关蛋白)、δ1)、CDH2(钙粘素2,1型、N-钙粘素(神经元的))、 CMA1(糜酶1、肥大细胞)、SORT1(分拣蛋白1)、DLK1(δ样1同系物 (果蝇))、THEM4(硫酯酶超家族成员4)、JUP(连接斑珠蛋白(junction plakoglobin))、CD46(CD46分子,补体调节蛋白)、CCL11 (趋化因子(C-C基序)配体11)、CAV3(小窝蛋白3)、RNASE3(核糖 核酸酶,RNA酶A家族、3(嗜酸性粒细胞阳离子蛋白))、HSPA8(热休 克70kDa蛋白8)、CASP9(半胱天冬酶9,凋亡相关半胱氨酸肽酶)、 CYP3A4(细胞色素P450,家族3,亚家族A,多肽4)、CCR3(趋化因子 (C-C基序)受体3)、TFAP2A(转录因子AP-2α(激活增强子结合蛋白 2α))、SCP2(固醇载体蛋白2)、CDK4(细胞周期蛋白依赖性激酶4)、 HIF1A(缺氧可诱导的因子1,α亚基(基本螺旋-环-螺旋转录因子))、 TCF7L2(转录因子7样2(T细胞特异的,HMG盒))、IL1R2(白介素1 受体,II型)、B3GALTL(β1,3-半乳糖转移酶样)、MDM2(Mdm2 p53结 合蛋白同系物(小鼠))、RELA(v-rel网状内皮组织增殖病毒癌基因同系物 A(禽的))、CASP7(半胱天冬酶7,凋亡相关半胱氨酸肽酶)、IDE(胰 岛素降解酶)、FABP4(脂肪酸结合蛋白4,脂肪细胞)、CASK(钙/钙调蛋 白依赖性丝氨酸蛋白激酶(MAGUK家族))、ADCYAP1R1(腺苷酸环化酶 激活多肽1(脑垂体的)受体类型I)、ATF4(激活转录因子4(tax-反应增 强子元件B67))、PDGFA(血小板衍生的生长因子α多肽)、C21或f33 (21号染色体开放阅读框33)、SCG5(分泌粒蛋白V(7B2蛋白))、 RNF123(环指蛋白123)、NFKB1(B细胞中的κ轻多肽基因增强子的核因 子1)、ERBB2(v-erb-b2红白血病病毒癌基因同系物2、神经/胶质母细胞瘤 衍生的癌基因同系物(禽的))、CAV1(小窝蛋白1,胞膜窖蛋白,22 kDa)、MMP7(基质金属肽酶7(基质溶解素,子宫的))、TGFA(转化 生长因子,α)、RXRA(类视黄醇X受体,

α)、STX1A(突触融合蛋白1A (脑的))、PSMC4(蛋白酶体(前体,巨蛋白因子)26S亚基,ATP酶,4)、P2RY2(嘌呤能受体P2Y,G蛋白偶联的,2)、TNFRSF21(肿瘤坏死 因子受体超家族,成员21)、DLG1(discs,大的同系物1(果蝇))、 NUMBL(numb同系物(果蝇)样)、SPN(涎福林(sialophorin))、 PLSCR1(磷脂混杂酶1)、UBQLN2(泛醌蛋白2)、UBQLN1(泛醌蛋白1)、PCSK7(前蛋白转化酶枯草杆菌/科信(kexin)类型7)、SPON1(脊 椎蛋白1,细胞外基质蛋白)、SILV(西尔弗(silver)同系物(小鼠))、 QPCT(谷氨酰胺-肽环转移酶)、HESS(长毛和裂口增强蛋白5(果 蝇))、GCC1(GRIP和含卷曲螺旋结构域的1)及其任何组合。

[0844] 该基因修饰动物或细胞可以包括1、2、3、4、5、6、7、8、9、10个或 更多个破坏的、编码与分泌酶障碍相关的蛋白质的染色体序列以及零、1、 2、3、4、5、6、7、8、9、10个或更多个编码破坏的与分泌酶障碍相关的蛋 白质的染色体整合序列。

[0845] 与肌萎缩性侧索硬化相关的蛋白质的实例可以包括SOD1(超氧化物歧 化酶1)、ALS2(肌萎缩性侧索硬化2)、FUS(融合在肉瘤中)、TARDBP (TAR DNA结合蛋白)、VAGFA(血管内皮生长因子A)、VAGFB(血管 内皮生长因子B)以及VAGFC(血管内皮生长因子C)及其任何组合。

[0846] 例如,美国专利公开号20110023144描述了使用锌指核酸酶基因修饰与 肌萎缩性侧索硬化(ALS)疾病相关的细胞、动物以及蛋白质。ALS由涉及 随意运动的脑皮层、脑干和脊髓中的某些神经细胞的逐步稳定退化表征。

[0847] 运动神经元障碍以及与这些障碍相关的蛋白质是一套影响患上运动神经 元障碍的易感性、运动神经元障碍的存在、运动神经元障碍的严重性或其任 何组合的相异蛋白质。本披露包括编码与一种特定的运动神经元障碍ALS疾 病相关的蛋白质的任何染色体序列的编辑。典型地基于ALS相关蛋白与ALS 的实验相关性选择与ALS相关的蛋白质。例如,相对于没有ALS的群体,在 具有ALS的群体中,与ALS相关的蛋白质的产生率或循环浓度可以升高或降 低。可以使用蛋白质组技术评估蛋白质水平差异,包括但不限于蛋白质印 迹、免疫组织化学染色、酶联免疫吸附测定(ELISA)以及质谱法。可替代 地,可以使用基因组技术通过获得编码这些蛋白质的基因的基因表达谱而鉴 定与ALS相关的蛋白质,包括但不限于DNA微阵列分析、基因表达序列分 析(SAGE)以及定量实时聚合酶链式反应(Q-PCR)。

[0848] 通过非限制性实例的方式,与ALS相关的蛋白质包括但不限以下蛋白 质:SOD1超氧化物歧化酶1,ALS3肌萎缩性侧索可溶性硬化3 SETX senataxin ALS5肌萎缩性侧索硬化5 FUS融合在肉瘤中ALS7肌萎缩性侧索硬 化7 ALS2肌萎缩性侧索DPP6二肽基肽酶6硬化2 NEFH神经微丝,重 PTGS1前列腺素-多肽内过氧化物合酶1 SLC1A2溶质载体家族1 TNFRSF10B 肿瘤坏死因子(胶质高亲和性受体超家族,谷氨酸转运体),成员10b成员2 PRPH外周蛋白HSP90AA1热休克蛋白90kDaα(胞质的),类别A成员1 GRIA2谷氨酸受体,IFNG干扰素,γ离子型的,AMPA 2 S100B S100钙结合 FGF2成纤维细胞生长因子2蛋白质B AOX1醛氧化酶1CS柠檬酸合酶 TARDBP TAR DNA结合蛋白TXN硫氧还原蛋白RAPH1 Ras缔合MAP3K5 丝裂原激活蛋白质(RaIGDS/AF-6)和激酶5普列克底物蛋白(pleckstrin)同 源性结构域1NBEAL1蛋白激酶锚定蛋白(neurobeachin)样1GPX1谷胱甘 肽过氧化物酶1 ICA1L胰岛细胞自身抗原RAC1 ras-相关C3肉毒菌1.69kDa 样毒素底物1 MAPT微管相关ITPR2肌醇1,4,5-蛋白tau三磷酸受体,2型 ALS2CR4肌萎缩性侧索GLS谷氨酰胺酶硬化2(青少年)染色体区域,候选 物4 ALS2CR8肌萎缩性侧索CNTFR睫状神经营养因子硬化2(青少年)受 体染色体

区域,候选物8 ALS2CR11肌萎缩性侧索FOLH1叶酸水解酶1硬化 2(青少年)染色体区域,候选物11 FAM117B具有序列P4HB的家族脯氨酰 基4-羟化酶,相似性117,成员Bβ多肽CNTF睫状神经营养因子SQSTM1 死骨片(sequestosome)1STRADB STE20相关激酶NAIP NLR家族,凋亡适 配器β抑制蛋白YWHAQ酪氨酸3-SLC33A1溶质载体家族33单加氧酶/色氨 酸(乙酰-CoA转运体),一种5-单加氧酶成员1激活蛋白,θ多肽TRAK2 运输蛋白,图4图4同系物,SAC1驱动蛋白结合2脂质磷酸酶结构域包含 NIF3L1 NIF3 NGG1相互作用INA互联蛋白(internexin)神经因子3样1中 间丝蛋白,αPARD3B par-3分区COX8A细胞色素c氧化酶缺陷3同系物B 亚基VIIIA CDK15细胞周期蛋白依赖性激酶HECW1 HECT,C2和WW 15结 构域包含E3泛素蛋白连接酶1NOS1一氧化氮合酶1MET met原癌基因 SOD2超氧化物歧化酶2,HSPB1热休克27kDa线粒体蛋白1NEFL神经微 丝,轻CTSB组织蛋白酶B多肽ANG血管生成素,HSPA8热休克70kDa核 糖核酸酶,RNA酶A蛋白8家族,5VAPB VAMP(囊泡-ESR1雌激素受体1 相关膜蛋白)-相关蛋白B和C SNCA突触核蛋白,αHGF肝细胞生长因子 CAT过氧化氢酶ACTB肌动蛋白,βNEFM神经微丝,介质TH酪氨酸羟化 酶多肽BCL2 B-细胞CLL/淋巴瘤2FAS Fas(TNF受体超家族,成员6) CASP3半胱天冬酶3,凋亡-CLU簇集蛋白相关半胱氨酸肽酶SMN1运动神经 元的存活G6PD葡萄糖-6-磷酸1,端粒脱氢酶BAX BCL2-相关X HSF1热休 克转录蛋白因子1 RNF19A环指蛋白19A JUN jun癌基因ALS2CR12肌萎缩 性侧索HSPA5热休克70kDa硬化2(青少年)蛋白5染色体区域,候选物 12 MAPK14丝裂原激活蛋白IL10白介素10激酶14 APEX1 APEX核酸酶TXNRD1硫氧还原蛋白还原酶1(多功能DNA修复酶)1 NOS2一氧化氮合 酶2,TIMP1 TIMP金属肽酶可诱导的抑制剂1 CASP9半胱天冬酶9,相关半 胱氨酸凋亡肽酶的凋亡-XIAP X-连接的抑制剂GLG1高尔基体糖蛋白1 EPO 促红细胞生成素VEGFA血管内皮ELN弹性蛋白生长因子A GDNF胶质细胞 衍生的NFE2L2核因子(红系(erythroid)-神经营养因子衍生的2)样2 SLC6A3溶质载体家族6 HSPA4热休克70kDa(神经递质蛋白4转运体,多 巴胺),成员3 APOE载脂蛋白E PSMB8蛋白酶体(前体,巨蛋白因子)亚 基,β类型,8 DCTN1动力蛋白激活蛋白(dynactin)1 TIMP3 TIMP金属肽 酶抑制剂3 KIFAP3驱动蛋白相关SLC1A1溶质载体家族1蛋白3(神经元/上 皮高亲和性谷氨酸转运体,系统Xag),成员1 SMN2运动神经元的存活CCNC细胞周期蛋白C 2,着丝粒MPP4膜蛋白,STUB1 STIP1同源性和U- 棕榈酰化的4盒包含蛋白1 ALS2淀粉样蛋白β(A4)PRDX6过氧化物酶6 前体蛋白SYP突触素CABIN1钙调磷酸酶结合蛋白1 CASP1半胱天冬酶1, 凋亡-GART磷酸核糖甘氨酰胺相关半胱氨酸甲酰转移酶,肽酶磷酸核糖甘氨 酰胺合成酶,磷酸核糖氨基咪唑合成酶CDK5细胞周期蛋白依赖性激酶5 ATXN3共济失调蛋白3 RTN4浆膜蛋白(reticulon)4 C1QB补体成分1,q亚 成分,B链VEGFC神经生长因子HTT亨廷顿蛋白(huntingtin)受体PARK7 帕金森病7 XDH黄嘌呤脱氢酶GFAP胶质纤维酸性MAP2微管相关蛋白2 CYCS细胞色素c,躯体的FCGR3B IgG的Fc片段,低亲和性IIIb,CCS UBL5(泛素样5超氧化物歧化酶)的铜分子伴侣MMP9基质金属肽酶 SLC18A3溶质载体家族18 9((囊状乙酰胆碱),成员3 TRPM7瞬时受体 HSPB2热休克27kDa潜在的阳离子通道,蛋白2亚家族M,成员7 AKT1 v- akt鼠胸腺瘤DERL1 Der1样结构域家族,病毒癌基因同系物1成员1 CCL2 趋化因子(C--C基序)NGRN突触生长相关蛋白(neugrin),轴突配体2赘生物(outgrowth)相关GSR谷胱甘肽还原酶TPPP3促微管蛋白聚合蛋白家族 成员3 APAF1凋亡肽酶BTBD10 BTB(POZ)结构域激活因子1包含10 GLUD1谷氨酸CXCR4趋化因子(C--X--C基序)脱氢酶1受体4 SLC1A3溶 质载体家族1 FLT1 fms相关酪氨酸(胶质高亲和性谷氨酸转

运体),成员3 激酶1 PON1对氧磷酶1AR雄激素受体LIF白血病抑制因子ERBB3 v-erb-b2 红白血病病毒癌基因同系物3 LGALS1凝集素,半乳糖苷-CD44 CD44分子结 合,可溶的,1 TP53肿瘤蛋白p53 TLR3 toll样受体3 GRIA1谷氨酸受体, GAPDH甘油-3-离子型的,AMPA 1磷酸脱氢酶GRIK1谷氨酸受体,DES结 蛋白离子型的,红藻氨酸(kainate)1 CHAT胆碱乙酰转移酶FLT4 fms相关 酪氨酸激酶4 CHMP2B染色质修饰的BAG1 BCL2相关蛋白2B永生基因(athanogene)MT3金属硫蛋白3 CHRNA4胆碱能受体,烟碱的,α4 GSS谷 胱甘肽合成酶BAK1 BCL2-拮抗剂(antagonist/killer)1 KDR激酶插入结构域 GSTP1谷胱甘肽S-转移酶受体(III型pi 1受体酪氨酸激酶)OGG1 8-氧桥鸟 嘌呤(oxoguanine)DNA IL6白介素6(干扰素,糖基化酶β2)。

[0849] 该动物或细胞可以包括1、2、3、4、5、6、7、8、9、10个或更多个破 坏的、编码与ALS相关的蛋白质的染色体序列以及零、1、2、3、4、5、6、 7、8、9、10个或更多个编码破坏的与ALS相关的蛋白质的染色体整合序 列。与ALS相关的优选蛋白包括SOD1(超氧化物歧化酶1)、ALS2(肌萎 缩性侧索硬化2)、FUS(融合在肉瘤中)、TARDBP(TAR DNA结合蛋 白)、VAGFA(血管内皮生长因子A)、VAGFB(血管内皮生长因子B)以 及VAGFC(血管内皮生长因子C)及其任何组合。

[0850] 与朊病毒病相关的蛋白质的实例可以包括SOD1(超氧化物歧化酶1)、 ALS2(肌萎缩性侧索硬化2)、FUS(融合在肉瘤中)、TARDBP(TAR DNA结合蛋白)、VAGFA(血管内皮生长因子A)、VAGFB(血管内皮生 长因子B)以及VAGFC(血管内皮生长因子C)及其任何组合。

[0851] 与朊病毒病症中的神经退行性疾病相关的蛋白质的实例可以包括例如 A2M(α-2-巨球蛋白)、AATF(凋亡拮抗转录因子)、ACPP(前列腺的酸 性磷酸酶)、ACTA2(肌动蛋白α2平滑肌主动脉)、ADAM22(ADAM金 属肽酶结构域)、ADORA3(腺苷A3受体)或ADRA1D(α-1D肾上腺素能 受体(Alpha-1D adrenergic receptor或Alpha-1D adrenoreceptor))。

[0852] 与免疫缺陷相关的蛋白质的实例可以包括例如A2M[α-2-巨球蛋白]; AANAT[芳烷基胺N-乙酰转移酶];ABCA1[ATP-结合盒,亚家族A (ABC1),成员1];ABCA2[ATP-结合盒,亚家族A(ABC1),成员2]; 或ABCA3[ATP-结合盒,亚家族A(ABC1),成员3]。

[0853] 与三核苷酸重复障碍相关的蛋白质的实例包括例如AR(雄激素受体)、 FMR1(脆性X智力迟钝1)、HTT(亨廷顿蛋白)或DMPK(强直性肌营养 不良蛋白激酶)、FXN(弗氏共济失调蛋白(frataxin))、ATXN2(共济失 调蛋白2)。

[0854] 与神经传递障碍相关的蛋白质的实例包括例如SST(生长抑素)、NOS1 (一氧化氮合酶1(神经元的))、ADRA2A(肾上腺素能的,α-2A-,受 体)、ADRA2C(肾上腺素能的,α-2C-,受体)、TACR1(速激肽受体1) 或HTR2c(5-羟色胺(血清素)受体2C)。

[0855] 神经发育相关序列的实例包括例如A2BP1[共济失调蛋白2-结合蛋白 1]、AADAT[氨基己二酸氨基转移酶],AANAT[芳烷基胺N-乙酰转移酶]、 ABAT[4-氨基丁酸氨基转移酶]、ABCA1[ATP-结合盒,亚家族A (ABC1),成员1]或ABCA13[ATP-结合盒,亚家族A(ABC1),成员 13]。

[0856] 可用本发明系统治疗的优选病症的另外的实例可以选自:艾卡迪-古铁雷 斯综合征(Aicardi-Goutières Syndrome);亚历山大病;阿伦-赫恩登-达德利 综合征(Allan-Herndon-Dudley Syndrome);POLG相关障碍;α-甘露糖苷贮 积症(II和III型);阿尔斯特伦综合征(Alström Syndrome);安格曼;综 合征;共济失调-毛细血管扩张;神经元蜡样-脂褐

质沉积;β-地中海贫血;双 侧视神经萎缩和1型(婴儿)视神经萎缩;视网膜母细胞瘤(双侧的);卡 纳万病(Canavan Disease);脑-眼-面-骨骼综合征(Cerebrooculofacioskeletal Syndrome)1[COFS1];脑腱黄瘤病;科尼利亚迪兰吉综合征(Cornelia de Lange Syndrome);MAPT相关障碍;遗传性朊病毒病;德拉韦综合征(Dravet Syndrome);早期发病家族性阿尔茨海默病;弗里德赖希共济失调 [FRDA];多发性畸形;岩藻糖苷贮积症;福山(Fukuyama)先天性肌营养不 良;半乳糖唾液酸贮积症;戈谢病(Gaucher Disease);有机酸血症;噬血细 胞淋巴组织细胞增生症;早衰症;粘脂贮积症II;婴儿游离唾液酸贮积病;PLA2G6相关神经变性;耶韦尔和朗格-尼尔森综合征(Jervell and Lange- Nielsen Syndrome);结合性大疱性表皮松解;亨廷顿病;克拉伯病(Krabbe Disease)(婴儿的);线粒体DNA相关雷吉综合征(Mitochondrial DNA- Associated Leigh Syndrome)和NARP;莱施-奈恩综合征;LIS1相关无脑回; 洛氏综合征(Lowe Syndrome);槭糖尿病;MECP2复制综合征;ATP7A相 关铜转运障碍;LAMA2相关肌营养不良;芳基硫酸酯酶A缺乏;I、II或III 型粘多糖贮积症;过氧化物酶体生物发生障碍,齐薇格谱系综合征 (Zellweger Syndrome Spectrum);伴随脑铁累积障碍的神经变性;酸性鞘磷 脂酶缺乏;C型尼曼-皮克病;甘氨酸脑病;ARX相关障碍;尿素循环障碍; COL1A1/2相关成骨不全;线粒体DNA缺失综合征;PLP1相关障碍;佩里综 合征(Perry Syndrome);费伦-麦克德米德综合征(Phelan-McDermid Syndrome);II型糖原贮积症(蓬佩病(Pompe Disease))(婴儿的); MAPT相关障碍;MECP2相关障碍;1型肢近端型点状软骨发育不全 (Rhizomelic Chondrodysplasia Punctata Type 1);罗伯茨综合征(Roberts Syndrome);桑德霍夫病(Sandhoff Disease);辛德勒病(Schindler Disease)-1型;腺苷脱氨酶缺乏;史-伦-奥三氏综合征;脊髓性肌萎缩;婴 儿发病脊髓小脑性共济失调;己糖胺酶A缺乏;1型致死性发育不良;胶原 VI型相关障碍;I型乌谢尔综合征(Usher Syndrome);先天性肌营养不良; 沃尔夫-赫奇霍恩综合征(Wolf-Hirschhorn Syndrome);溶酶体酸脂酶缺乏; 以及着色性干皮病。

[0857] 如将是显而易见的,设想的是可以使用本发明系统靶向任何感兴趣的多 核苷酸序列。使用本发明系统进行治疗可能是有用的病症或疾病的一些实例 被包括在上表中并且当前与那些病症相关的基因的实例也被提供于此。然 而,示例的基因并不是穷尽性的。

[0858] 例如,“野生型StCas9”是指来自嗜热链球菌的野生型Cas9,其蛋白质 序列在登录号G3ECR1下给出于SwissProt数据库中。类似地,化脓链球菌 Cas9在登录号Q99ZW2下被包括在SwissProt中。

[0859] 使用CRISPR-Cas系统进行有效且符合成本效益的基因编辑和操纵的能 力将允许快速选择并比较单个和多元遗传操作,以转化这样的基因组,以便 改善生产并增强性状。在此方面,参考美国专利和出版物:美国专利号 6,603,061-农杆菌介导的植物转化法(Agrobacterium-Mediated Plant Transformation Method);美国专利号7,868,149-植物基因组序列及其用途(Plant Genome Sequences and Uses Thereof)以及US 2009/0100536-转具有增 强的农艺性状的基因植物(Transgenic Plants with Enhanced Agronomic Traits),将每者的所有内容和披露通过引用以其全文结合在此。在本发明的实践中,莫雷尔(Morrell)等人“作物基因组学:进展与应用(Crop genomics:advances and applications)”《遗传学自然评论》(Nat Rev Genet.) 2011年12月29日;13(2):85-96的内容和披露也通过引用以其全文结合在 此。

[0860] 实例

[0861] 以下实例出于说明本发明的不同实施例的目的而给出并且并不意在以任 何方式限制本发明。本发明实例连同在此描述的方法目前代表优选的实施 例，是示例性的，并且并不旨在限制本发明的范围。涵盖在如由权利要求书 的范围所定义的本发明的精神内的在此的变化以及其他用途是本领域普通技 术人员可以想到的。

[0862] 实例1:真核细胞的细胞核中的CRISPR复合物活性

[0863] 一个示例性II型CRISPR系统是来自化脓链球菌SF370的II型CRISPR 座位，该座位包含四个基因Cas9、Cas1、Cas2和Csn1的聚簇以及两个非编 码RNA元件tracrRNA和由非重复序列的短段(间隔子，每个约30bp)间隔 开的重复序列(同向重复)的特征性阵列。在此系统中，以四个连续步骤产 生靶向的DNA双链断裂(DSB)(图2A)。第一步，从CRISPR座位转录两个非编码RNA、pre-crRNA阵列和tracrRNA。第二步，将tracrRNA杂交到 pre-crRNA的同向重复上，然后将其加工成包含单独的间隔子序列的成熟 crRNA。第三步，该成熟crRNA:tracrRNA复合物经由在crRNA的间隔子区 与原型间隔子DNA之间形成异源双链体而指导Cas9到由原型间隔子和对应 的PAM组成的DNA靶标。最后，Cas9介导PAM上游的靶DNA的切割，以 在原型间隔子内产生一个DSB(图2A)。此实例描述了一种用于将此RNA 可编程的核酸酶系统适配成指导真核细胞的细胞核中的CRISPR复合物活性 的示例性过程。

[0864] 细胞培养和转染

[0865] 在37℃下，伴随5％CO_2孵育，将人类胚肾(HEK)细胞系HEK 293FT (生命技术公司)维持在补充有10％胎牛血清(海可隆公司(HyClone))、2mM GlutaMAX(生命技术公司)、100U/mL青霉素及100μg/mL链霉素的 杜尔贝科改良伊格尔培养基(Dulbecco's modified Eagle's Medium) (DMEM)中。在37℃下，在5％CO_2下，用补充有5％胎牛血清(海可隆 公司)、2mM GlutaMAX(生命技术公司)、100U/mL青霉素及100μg/mL 链霉素的DMEM维持小鼠neuro2A(N2A)细胞系(ATCC)。

[0866] 在转染前一天，将HEK 293FT或N2A细胞以200,000个细胞/孔的密度 接种到24孔板(康宁公司(Corning))中。遵循制造商推荐的方案，使用 Lipofectamine 2000(生命技术公司)转染细胞。对于24孔板的每个孔，使用 总共800ng的质粒。

[0867] 基因组修饰的Surveyor测定和测序分析

[0868] 如上所述，用质粒DNA转染HEK 293FT或N2A细胞。转染后，在基因 组DNA提取之前，将细胞在37℃下孵育72小时。遵循制造商的方案，使用 QuickExtract DNA提取试剂盒(Epicentre)提取基因组DNA。简言之，将细 胞重悬于QuickExtract溶液中并在65℃下孵育15分钟并且在98℃下孵育10 分钟。将提取的基因组DNA立即加工或存储在-20℃下。

[0869] 对于每个基因而言，PCR扩增围绕CRISPR靶位点的基因组区域，并且 遵循制造商的方案，使用QiaQuick Spin柱(凯杰公司(Qiagen))纯化产 物。将总共400ng的纯化的PCR产物与2μl 10X Taq聚合酶PCR缓冲液 (Enzymatics公司)和超纯水混合，至终体积为20μl，并使其经受重退火过 程，以使得可以形成异源双链体:95℃持续10min，以-2℃/s 95℃降至 85℃，以-0.25℃/s 85℃降至25℃，并且在25℃保持1分钟。重退火后，遵 循制造商推荐的方案，将产物用Surveyor核酸酶和Surveyor增强子S(转基 因组学公司(Transgenomics))处理，并且在4％-20％Novex TBE聚丙烯酰 胺凝胶(生命技术公司)上进行分析。将凝胶用SYBR Gold DNA染色剂(生 命技术公司)染色30分钟并用Gel Doc凝胶成

像系统(伯乐公司(Bio- rad))成像。定量是基于作为切割的DNA的部分的量度的相对条带强度。图7提供了此Surveyor测定的一个示意图。

[0870] 用于检测同源重组的限制性片段长度多态性测定。

[0871] 将HEK 293FT和N2A细胞用质粒DNA转染,并且如上所述,在基因组 DNA提取之前,将其在37℃下孵育72小时。使用引物在同源重组(HR)模 板的同源臂外PCR扩增靶基因组区域。将PCR产物分离在1%琼脂糖凝胶上 并用MinElute GelExtraction试剂盒(凯杰公司)提取。将纯化产物用HindIII (费尔芒斯公司(Fermentas))消化并在6%Novex TBE聚丙烯酰胺凝胶 (生命技术公司)上进行分析。

[0872] RNA二级结构预测和分析

[0873] 使用质心结构预测算法,使用由维也纳大(University of Vienna)的理论 化学研究所(Institute for Theoretical Chemistry)研发的在线网络服务器 RNAfold进行RNA二级结构预测(参见例如A.R.格鲁伯(A.R.Gruber)等 人,2008,《细胞》(Cell)106(1):23-24;以及PA卡尔(PA Carr)和GM 丘奇(GM Church),2009,《自然生物技术》(Nature Biotechnology) 27(12):1151-62)。

[0874] RNA纯化

[0875] 将HEK 293FT细胞如上所述地维持和转染。通过胰酶消化收获细胞,随 后在磷酸盐缓冲生理盐水(PBS)中洗涤。遵循制造商的方案,用TRI试剂 (西格玛公司)提取总细胞RNA。使用Naonodrop(赛默飞世尔科技公司 (Thermo Scientific))对提取的总RNA进行定量并标准化至同一浓度。

[0876] 哺乳动物细胞中的crRNA和tracrRNA表达的Northern印迹分析

[0877] 将RNA与等体积的2X加样缓冲液(Ambion)混合,加热至95℃持续 5min,在冰上冷冻1min,并且然后在将凝胶预跑胶至少30分钟后,加样到 8%变性聚丙烯酰胺凝胶(SequaGel,国家诊断(National Diagnostics))上。将样品以40W极限电泳1.5小时。之后,在室温下,将RNA转移到半干转 移设备(伯乐公司(Bio-rad))中的300mA的Hybond N+膜(通用电气医 疗公司(GE Healthcare))上,持续1.5小时。使用Stratagene UV Crosslinker the Stratalinker(Stratagene)上的自动交联按钮将RNA交联到该膜上。在 42℃下,伴随旋转,将该膜在ULTRAhyb-Oligo杂交缓冲液(Ambion)中预 杂交30min,并且然后添加探针并杂交过夜。探针订购自IDT并将其用具有 T4多核苷酸激酶(新英格兰生物实验室(New England Biolabs)))的[γ-^{32}P] ATP(珀金埃尔默公司)标记。将该膜用预加温的(42℃)2xSSC、0.5% SDS洗涤一次,持续1min,随后在42℃下洗涤两次,持续30分钟。将该膜 在室温下向荧光屏暴露一小时或过夜并且然后用感光成像仪(Typhoon)扫 描。

[0878] 细菌CRISPR系统构建和评价

[0879] 从具有用于吉布森拼接(Gibson Assembly)的侧翼同源臂的化脓链球菌 SF370基因组DNA PCR扩增CRISPR座位元件(包括tracrRNA、Cas9和前 导子)。将两个BsaI类型IIS位点引入在两个同向重复之间,以促进方便地 插入间隔子(图8)。使用Gibson Assembly Master Mix(NEB)将PCR产物 克隆进EcoRV-消化的pACYC184中,在tet启动子的下游。省略其他内源 CRISPR系统元件,Csn2的最后50bp除外。将编码具有互补突出端的间隔子 的寡聚物(整合DNA技术公司(Integrated DNA Technology))克隆进BsaI- 消化的载体pDC000(NEB)中并且然后用T7连接酶(Enzymatics公司)连 接,以产生pCRISPR质粒。包含具有PAM

的间隔子的激发质粒

[0880] 在哺乳动物细胞中的表达(图6A中所示出的表达构建体,其中功能性是 如显示于图6B中的Surveyor测定的结果所确定的)。将转录起始位点标为 +1,并且还指示了转录终止子和通过Northern印迹探测的序列。还通过 Northern印迹确认了加工的tracrRNA的表达。图6C显示了提取自293FT细 胞的总RNA的Northern印迹分析的结果,用携带长或短tracrRNA以及 SpCas9和DR-EMX1(1)-DR的U6表达构建体转染这些细胞。左图和右图分别是来自用或没用SpRNA酶III转染的293FT细胞。U6指示用靶向人类U6 snRNA的探针进行印迹的加样对照。短tracrRNA表达构建体的转染产生丰富 水平的加工形式的tracrRNA(~75bp)。在Northern印迹上检测极低量的长 tracrRNA。

[0881] 为了促进转录精确起始,选择基于RNA聚合酶III的U6启动子,以驱 动tracrRNA的表达(图2C)。类似地,研发基于U6启动子的构建体,以表 达由单个间隔子组成的pre-crRNA阵列,该单个间隔子侧翼为两个同向重复 (DR,还被术语“tracr配对序列”所涵盖;图2C)。将起始间隔子设计成 靶向人EMX1座位(图2C)中的33-碱基对(bp)靶位点(满足Cas9的 NGG识别基序的30-bp原型间隔子加3-bp CRISPR基序(PAM)序列),该 座位是大脑皮层的发育中的一个关键基因。

[0882] 为了测试CRISPR系统(SpCas9、SpRNA酶III、tracrRNA以及pre- crRNA)在哺乳动物细胞中的异源表达是否可以实现哺乳动物染色体的靶向 切割,用CRISPR组分的组合转染HEK 293FT细胞。由于哺乳动物细胞核中 的DSB通过导致形成插入缺失的非同源末端连接(NHEJ)途径而部分地修 复,使用Surveyor测定检测靶EMX1座位处的潜在的切割活性(图7)(参 见例如格斯钦(Guschin)等人,2010,《分子生物学方法》(Methods Mol Biol)649:247)。所有四种CRISPR组分的共转染能够在原型间隔子中诱导 高达5.0%切割(参见图2D)。所有CRISPR组分(减去SpRNA酶III)的共 转染在原型间隔子中也诱导高达4.7%插入缺失,表明可能存在能够辅助 crRNA成熟的内源性哺乳动物RNase,如例如相关的Dicer和Drosha酶。除 去剩余的三种组分中的任何组分消除CRISPR系统的基因组切割活性(图2D)。包含已证实切割活性的靶座位的扩增子的桑格测序:在43个测序的克 隆中,发现5个突变的等位基因(11.6%)。使用多种指导序列的类似实验产 生了高达29%的插入缺失百分比(参见图3-6、10和11)。这些结果定义了 一种在哺乳动物细胞中的有效的CRISPR介导的基因组修饰的三组分系统。为了优化切割效率,申请人还测试了tracrRNA的不同亚型是否影响切割效率 并且发现,在实例系统中,仅仅短(89-bp)转录本形式能够介导人EMX1基因组座位的切割(图6B)。

[0883] 图12提供了哺乳动物细胞中的crRNA加工的另外的Northern印迹分 析。图12A示出了一个示意图,显示了侧翼为两个同向重复(DR-EMX1(1)- DR)的单个间隔子的表达载体。靶向人EMX1座位原型间隔子1(参见图 6)的30bp间隔子和同向重复序列显示于图12A下方的序列中。直线指示其 反向互补系列用于产生供EMX1(1)crRNA检测用的Northern印迹探针的区 域。图12B显示了提取自293FT细胞的总RNA的Northern印迹分析,用携 带DR-EMX1(1)-DR的U6表达构建体转染这些细胞。左图和右图分别是来自 用或没用SpRNA酶III转染的293FT细胞。DR-EMX1(1)-DR仅在存在 SpCas9和短tracrRNA的情况下被加工为成熟crRNA并且不依赖于SpRNA酶 III的存在。检测的来自转染的293FT总RNA的成熟crRNA是~33bp并且比 来自化脓链球菌的39-42bp成熟crRNA短。这些结果证明,CRISPR系统可 以被

移植进真核细胞中并被重新编程,以促进内源哺乳动物靶多核苷酸的切 割。

[0884] 图2示出了描述于此实例中的细菌CRISPR系统。图2A示出了一个示意 图,显示了来自化脓链球菌SF370的CRISPR座位和通过此系统CRISPR介 导的DNA切割的建议机制。加工自同向重复-间隔子阵列的成熟crRNA指导 Cas9到以下基因组靶标,这些靶标由互补原型间隔子和原型间隔子相邻基序 (PAM)组成。靶标-间隔子碱基配对后,Cas9介导靶DNA中的双链断裂。 图2B示出了具有使得可以导入到哺乳动物细胞核中的核定位信号(NLS)的化脓链球菌Cas9(SpCas9)和RNA酶III(SpRNA酶III)的工程化。图2C 示出了由组成型EF1a启动子驱动的SpCas9和SpRNA酶III以及由RNA Pol3 启动子U6驱动的tracrRNA和pre-crRNA阵列(DR-间隔子-DR)的哺乳动物 表达,上述启动子促进转录精确起始和终止。将来自具有令人满意的PAM序 列的人EMX1座位的原型间隔子用作pre-crRNA阵列中的间隔子。图2D示出 了SpCas9介导的次要插入和缺失的Surveyor核酸酶测定。用和不用SpRNA 酶III、tracrRNA以及携带EMX1-靶间隔子的pre-crRNA阵列表达SpCas9。图 2E示出了靶座位与EMX1-靶向crRNA之间的碱基配对的一个示意表示以及一 个显示了与SpCas9切割位点邻近的微小缺失的示例性层析谱。图2F示出了 从43个克隆扩增子的测序分析鉴定出的突变的等位基因,这些扩增子显示出 多种微小插入和缺失。破折号指示缺失的碱基,并且非比对或错配碱基指示 插入或突变。比例尺=10μm。

[0885] 为了进一步简化该三组分系统,改编嵌合的crRNA-tracrRNA杂交体设 计,其中成熟crRNA(包括一种指导序列)可以经由茎-环而融合到部分 tracrRNA上,以模拟天然的crRNA:tracrRNA双链体。为了增加共递送效 率,产生双顺反子表达载体,以驱动嵌合的RNA和SpCas9在转染细胞中的 共表达。平行地,使用该双顺反子载体表达pre-crRNA(DR-指导序列-DR) 与SpCas9,以诱导用分开表达的tracrRNA加工为crRNA(比较图11B的上 图和底图)。图8提供了pre-crRNA阵列(图8A)或嵌合的crRNA(由图 8B中的指导序列插入位点的下游和EF1α启动子的上游的短线表示)与 hSpCas9的双顺反子表达载体的示意图,显示了各种元件的位置和指导序列插 入点。图8B中的围绕指导序列插入位点的位置的展开的序列还显示了部分 DR序列(GTTTAGAGCTA)和部分tracrRNA序列(TAGCAAGTTAAAATAAGGCTAGTCCGTTTTT)。可以使用退火的寡核 苷酸将指导序列插入在BbsI位点之间。寡核苷酸的序列设计显示在图8中的 示意图的下方,其中指示了适当的连接适配子。WPRE表示土拨鼠肝炎病毒 转录后调节元件。通过靶向与上述相同的EMX1座位而测试嵌合的RNA介导 的切割效率。使用扩增子的Surveyor测定和桑格测序两者,申请人确认到嵌 合的RNA设计以大约4.7%的修饰率而促进人EMX1座位的切割(图3)。

[0886] 通过设计靶向人EMX1和PVALB以及小鼠Th座位中的多个位点的嵌合 的RNA,通过靶向人类和小鼠细胞两者中的另外的基因组座位而测试 CRISPR介导的切割在真核细胞中的普遍性。图13示出了人类PVALB(图 13A)和小鼠Th(图13B)座位中的一些另外的靶向原型间隔子的选择。提 供了各自的最后的外显子内的基因座位位和三个原型间隔子的位置的示意 图。加下划线的序列包括30bp的原型间隔子序列和对应于PAM序列的3'端 处的3bp。分别在DNA序列的上方和下方指示了有义链和反义链上的原型间 隔子。人类PVALB和小鼠Th座位分别达到了6.3%和0.75%的修饰率,证明 了CRISPR系统跨多种生物在修饰不同座位方面的广阔适用性(图5)。虽然 对于每个座位而言,使用嵌合的构建体仅伴随三分之一的间隔子检测切割, 但是当使用共表达的pre-crRNA安排时,所有靶序列都被切割,伴随达

到 27％的有效的插入缺失产生(图6和13)。

[0887] 图11提供了可以被重新编程为靶向哺乳动物细胞中的多个基因组座位的 SpCas9的另外的图解。图11A提供了人EMX1座位的示意图，显示了五个原 型间隔子的位置，由加下划线的序列指示。图11B提供了pre-crRNA/trcrRNA 复合物的示意图，显示了pre-crRNA和tracrRNA的同向重复区之间的杂交 (上)，以及嵌合的RNA设计的示意图，包括一个20bp指导序列和由部分 同向重复组成的tracr配对和tracr序列以及以发夹结构杂交的tracrRNA序列 (下)。比较人EMX1座位中的五个原型间隔子处的Cas9介导的切割的效率 的Surveyor测定的结果示于图11C中。使用加工的pre-crRNA/tracrRNA复合 物(crRNA)或嵌合的RNA(chiRNA)靶向每个原型间隔子。

[0888] 由于RNA的二级结构对于分子间相互作用可以是决定性的，使用基于最 小自由能和玻尔兹曼(Boltzmann)加权结构集合的结构预测算法比较基因组 靶向实验中使用的所有指导序列的假定的二级结构(参见例如格鲁贝尔 (Gruber)等人，2008，《核酸研究》(Nucleic Acids Research)，36: W70)。分析揭示到在大多数情况下，在嵌合的crRNA背景下有效的指导序 列基本上不含二级结构基序，而无效的指导序列更可能形成可以阻止与靶原 型间隔子DNA进行碱基配对的内部二级结构。因此，可能的是，当使用嵌合 的crRNA时，间隔子二级结构的变异性可能影响CRISPR介导的干扰的效 率。

[0889] SpCas9的另外的载体设计显示于图22中，示出了掺入连接到指导寡核 苷酸的插入位点的U6启动子和连接到SpCas9编码序列上的Cbh启动子的单 个表达载体。显示于图22b中的载体包括一个连接到H1启动子上的 tracrRNA编码序列。

[0890] 在细菌测定中，所有的间隔子均促进有效的CRISPR干扰(图3C)。这 些结果表明，可能存在影响哺乳动物细胞中的CRISPR活性效率的另外的因 素。

[0891] 为了研究CRISPR介导的切割的特异性，使用一系列具有单点突变的 EMX1-靶向嵌合的crRNA分析指导序列中的单核苷酸突变对哺乳动物基因组 中的原型间隔子切割的影响(图3A)。图3B示出了比较当与不同突变的嵌 合的RNA配对时Cas9的切割效率的Surveyor核酸酶测定的结果。PAM的5’的高达12-bp的单碱基错配基本上消除了由SpCas9进行的基因组切割，而在 离上游位置更远处具有突变的间隔子保留针对原始原型间隔子靶标的活性 (图3B)。除PAM之外，SpCas9在间隔子的最后12-bp内具有单碱基特异 性。此外，CRISPR能够像一对靶向同一EMX1原型间隔子的TALE核酸酶 (TALEN)一样有效地介导基因组切割。图3C提供了一个示意图，显示了 靶向EMX1的TALEN的设计并且图3D显示了比较TALEN和Cas9的效率的 Surveyor凝胶(n＝3)。

[0892] 已经通过易错NHEJ机制建立了一套用于在哺乳动物细胞中实现CRISPR 介导的基因编辑的组分，测试CRISPR刺激同源重组(HR)的能力，同源重 组是一种用于使得基因组中的编辑精确的高保真基因修复途径。野生型 SpCas9能够介导位点特异的DSB，这些DSB可以通过NHEJ和HR两者进行 修复。另外，将SpCas9的RuvC I催化结构域中的天冬氨酸至丙氨酸取代 (D10A)工程化，以将核酸酶转化为切口酶(SpCas9n；示于图4A中)(参 见例如萨普拉诺萨克斯(Sapranausaks)等人，2011，《核酸研究》(Nucleic Acids Research)，39:9275；加索纳斯(Gasiunas)等人，2012，《美国国家 科学院院刊》(Proc.Natl.Acad.Sci.USA)，109:E2579)，这样使得带切口的 基因组DNA经历高保真同源定向修复(HDR)。Surveyor测定确认到， SpCas9n在EMX1原型间隔子靶标处不产生插入缺失。如图4B所示，EMX1- 靶向的嵌合

crRNA与SpCas9的共表达在靶位点中产生插入缺失,而与 SpCas9n的共表达不产生插入缺失(n=3)。此外,327个扩增子的测序未检 测到由SpCas9n诱导的任何插入缺失。选择相同的座位,以通过用靶向 EMX1、hSpCas9或hSpCas9n的嵌合的RNA以及用于在原型间隔子附近引入 一对限制酶切位点(HindIII和NheI)的HR模板共转染HEK 293FT细胞而测 试CRISPR介导的HR。图4C提供了HR策略的示意图,包含重组点的相对 位置和引物退火序列(箭头)。SpCas9和SpCas9n的确催化将HR模板整合 进EMX1座位中。PCR扩增靶区域,随后用HindIII进行限制酶切消化揭示了 对应于期望片段尺寸的切割产物(显示于图4D中的限制性片段长度多态性分 析中的箭头),其中SpCas9和SpCas9n介导类似水平的HR效率。申请人使 用基因组扩增子的桑格测序进一步证实了HR(图4E)。这些结果证明了 CRISPR促进哺乳动物基因组中的靶向基因插入的实用性。鉴于14-bp(来自 间隔子的12-bp和来自PAM的2-bp)特异地靶向野生型SpCas9,切口酶的可 获得性可以显著减少脱靶修饰的可能性,因为单链断裂不是易错NHEJ途径 的底物。

[0893] 构建模拟具有阵列的间隔子的CRISPR座位的天然构造的表达构建体 (图2A),以测试多元序列靶向的可能性。使用编码一对EMX1-和PVALB- 靶向间隔子的单CRISPR阵列,在两个座位处都检测到了有效的切割(图 4F,显示了crRNA阵列的示意设计和显示出切割的有效修饰的检验员印迹两 者)。使用针对由119bp间隔开的EMX1内的两个靶标的间隔子通过共存 DSB还检测到了较大基因组区的靶向缺失,并且检测到了1.6%缺失效率 (182个扩增子中的3个;图4G)。这证明,CRISPR系统可以介导单个基因 组中的多元编辑。

[0894] 实例2:CRISPR系统修饰和替代方案

[0895] 使用RNA编程序列特异的DNA切割的能力定义了一个新类别的针对多 种研究和工业应用的基因组工程化工具。可以进一步改善CRISPR系统的若 干方面,以增加CRISPR靶向的效率和通用性。优化的Cas9活性可以依赖于 处于比存在于哺乳动物细胞核中的游离Mg^{2+}高的水平的游离Mg^{2+}的可获得性 (参见例如季聂克(Jinek)等人,2012,《科学》(Science),337:816), 并且对恰好位于原型间隔子的下游的NGG基序的偏好限制靶向人类基因组中的平均每12-bp的能力(图9,评价了人类染色体序列的正链和负链两者)。这些约束中的一些可以通过探索CRISPR座位跨微生物宏基因组的多样性而 克服(参见例如马卡洛娃(Makarova)等人,2011,《自然微生物学综述》 (Nat Rev Microbiol),9:467)。可以通过类似于实例1中描述的方法将其 他CRISPR座位移植进哺乳动物细胞环境中。例如,图10示出了来自嗜热链 球菌LMD-9的CRISPR 1的用于在哺乳动物细胞中进行异源表达的II型CRISPR系统的适应,以实现CRISPR介导的基因组编辑。图10A提供了来自 嗜热链球菌LMD-9的CRISPR 1的示意图。图10B示出了嗜热链球菌 CRISPR系统的表达系统的设计。使用组成型EF1α启动子表达人类密码子优 化的hStCas9。使用促进转录精确起始的U6启动子表达tracrRNA和crRNA 的成熟版本。示出了来自成熟crRNA和tracrRNA的序列。crRNA序列中的由小写字母“a”指示的单个碱基用于除去作为RNA polIII转录终止子的 polyU序列。图10C提供了一个示意图,显示了靶向人EMX1座位的指导序 列。图10D显示了使用Surveyor测定的靶座位中的hStCas9介导的切割的结 果。RNA指导间隔子1和2分别诱导了14%和6.4%。在这两个原型间隔子位 点处跨生物复制物的切割活性的统计分析还提供于图5中的。图14提供了嗜 热链球菌CRISPR系统的另外的原型间隔子和在人EMX1座位中的对应的 PAM序列靶标示意图。两个原型间隔子序列被高亮并且通过相对于对应的高 亮序列为3’加下划线

而指示满足NNAGAAW基序的对应的PAM序列。两个 原型间隔子均靶向反义链。

[0896] 实例3:样品靶序列选择算法

[0897] 设计一个软件程序,以基于指定的CRISPR酶的希望的指导序列长度和 CRISPR基序序列(PAM),在输入DNA序列的两条链上鉴定候选CRISPR 靶序列。例如,可以通过在输入序列和该输入序列的反向互补体两者上检索 5'-Nx-NGG-3'而鉴定来自化脓链球菌的Cas9的具有PAM序列NGG的靶位 点。同样地,可以通过在输入序列和该输入序列的反向互补体两者上检索5'- Nx-NNAGAAW-3'而鉴定化脓链球菌CRISPR1的Cas9的具有PAM序列 NNAGAAW的靶位点。同样地,可以通过在输入序列和该输入序列的反向互 补体两者上检索5'-Nx-NGGNG-3'而鉴定化脓链球菌CRISPR3的Cas9的具有 PAM序列NGGNG的靶位点。可以通过程序固定或由使用者指定N_x中的值 "x",如20。

[0898] 由于在DNA靶位点的基因组中出现多次可以导致非特异的基因组编辑,因此在鉴定所有潜在的位点后,基于序列在相关参比基因组中出现的次数,该程序将其滤出。对于其序列特异性是通过'种子(seed)'序列(如来自 PAM序列的11-12bp 5',包括PAM序列本身)确定的那些CRISPR酶,过 滤步骤可以基于该种子序列。因此,为了避免另外的基因组座位处的编辑, 基于种子:PAM序列在相关基因组中出现的次数过滤结果。可以允许使用者选择种子序列的长度。处于通过过滤器的目的,还可以允许使用者指定种 子:PAM序列在基因组中出现的次数。默认筛选独特序列。通过改变种子序 列的长度和该序列在基因组中出现的次数两者而改变过滤水平。该程序可以 另外或可替代地通过提供鉴定的一个或多个靶序列的反向互补体而提供与报 告的一个或多个靶序列互补的指导序列的序列。一些靶位点在人类基因组中 的示例性可视化提供于图18中。

[0899] 用于优化序列选择的方法和算法的另外的细节可以发现于通过引用结合 在此的美国申请序列号61/064,798(代理人案号44790.11.2022;博德参考号 BI-2012/084)中。

[0900] 实例4:多种嵌合的crRNA-tracrRNA杂交体的评价

[0901] 此实例描述了针对具有掺入不同长度的野生型tracrRNA序列的tracr序列 的嵌合的RNA(chiRNAs;在单个转录本中包括指导序列、tracr配对序列以 及tracr序列)而获得的结果。图16a示出了嵌合的RNA和Cas9的双顺反子 表达载体的示意图。Cas9由CBh启动子驱动并且嵌合的RNA由U6启动子驱 动。嵌合的指导RNA由接合到tracr序列的20bp指导序列(N)组成(从下 链的第一个"U"到该转录本的结尾),其在如指示的各点被截短。指导和tracr序列由tracr-配对序列GUUUUAGAGCUA,随后是环序列GAAA隔开。人EMX1和PVALB座位处的Cas9介导的插入缺失的SURVEYOR测定的结 果分别示于图16b和16c中。箭头指示预期的SURVEYOR片段。ChiRNA由 其"+n"标志指示,并且crRNA是指指导序列和tracr序列被表达为分开的转 录本的杂交体RNA。一式三份地进行的这些结果的定量示于图17a和17b的 直方图中,分别对应于图16b和16c("N.D."指示未检测到插入缺失)。原 型间隔子ID及其对应的基因组靶标、原型间隔子、PAM序列以及链位置提 供于表D中。将指导序列设计成与整个原型间隔子序列互补(在杂交体系统 中的分开的转录本的情况下),或仅为部分加下划线(在嵌合的RNA的情况 下)。

[0902] 表D:

[0903]

原型间隔子ID	基因组靶标	原型间隔子序列（5'至3'）	PAM	链
1	*EMX1*	GGACATCGATGTCACCTCCAATGACTAGGG	TGG	+
2	*EMX1*	CATTGGAGGTGACATCGATGTCCTCCCCAT	TGG	-
3	*EMX1*	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+
4	*PVALB*	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	+
5	*PVALB*	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	+

[0904] 用于优化指导序列的另外的细节可以发现于通过引用结合在此的美国申 请序列号61/836,127(代理人案号44790.08.2022;博德参考号BI- 2013/004G)中。

[0905] 初始,靶向人HEK 293FT细胞中的EMX1座位内的三个位点。使用 SURVEYOR核酸酶测定评估每种chiRNA的基因组修饰效率,该测定检测由 DNA双链断裂(DSB)及其通过非同源末端连接(NHEJ)DNA损伤修复途 径进行的随后修复造成的突变。被指定为chiRNA(+n)的构建体指示,野生型 tracrRNA的直达第+n个核苷酸被包括在嵌合的RNA构建体中,其中48、54、67和85的值用于n。嵌合的RNA在所有三个EMX1靶位点处都包含长 片段的野生型tracrRNA(chiRNA(+67)和chiRNA(+85))介导的DNA切割, 其中chiRNA(+85)尤其展示出比以分开的转录本形式表达指导序列和tracr序 列的对应的crRNA/tracrRNA杂交体显著更高水平的DNA切割(图16b和 17a)。还使用chiRNA靶向PVALB座位中的两个位点,其不产生使用杂交 体系统(作为分开的转录本而表达的指导序列和tracr序列)的可检测的切 割。chiRNA(+67)和chiRNA(+85)能够在两个PVALB原型间隔子处介导显著 的切割(图16c和17b)。

[0906] 对于EMX1和PVALB座位中的所有五个靶标,伴随增加的tracr序列长 度观察到一致增加的基因组修饰效率。不希望受任何理论的束缚,由 tracrRNA的3'端形成的二级结构可以在增强CRISPR复合物形成率方面发挥 一定作用。

[0907] 实例5:Cas9多样性

[0908] CRISPR-Cas系统是一种由跨细菌和古生菌的相异物种利用的针对侵入型 外源DNA的获得性免疫机制。II型CRISPR-Cas9系统由一套编码负责将外源 DNA"捕获(acquisition)"进CRISPR座位中的蛋白质的基因以及一套编码 负责"执行(execution)"DNA切割机制的蛋白质的基因组成;这些包括 DNA核酸酶(Cas9)、非编码反式激活cr-RNA(tracrRNA)以及侧翼为同向 重复的外源DNA衍生的间隔子阵列(crRNA)。当被Cas9成熟后,tracRNA 和crRNA双链体指导Cas9核酸酶到由间隔子指导序列规定的靶DNA序列, 并且在靶DNA中介导短序列基序附近的DNA中的双链断裂,该导短序列基 序为切割所需并且对于每种CRISPR-Cas系统而言是特异的。II型CRISPR- Cas系统发现在整个细菌界中并且在用于靶标切割的Cas9蛋白序列和尺寸、tracrRNA和crRNA同向重复序列、这些元件的基因组组织以及基序要求方面 高度相异。一个物种可以具有多种相异的CRISPR-Cas系统。

[0909] 申请人评价了来自基于与已知Cas9的序列同源性和与已知亚结构域直向 同源的结构鉴定的细菌物种的207个假定的Cas9,这些亚结构域包括HNH 内切核酸酶结构域和RuvC内切核酸酶结构域[信息来自尤金库宁(Eugene Koonin)和基拉马卡洛娃(Kira Makarova)]。基于此套的蛋白质序列保守的 系统发生分析揭示了五个Cas9家族,包括三组大的Cas9(～1400个氨基酸) 和两组小的Cas9(～1100个氨基酸)(参见图19和20A-F)。

[0910] Cas9和Cas9酶转化为切口酶或DNA结合蛋白的突变及其具有的改变的 功能性的用途的另外的细节可以发现于通过引用结合在此的美国申请序列号 61/836,101和61/835,936(代理人案号分别为44790.09.2022和4790.07.2022 并且博德参考号分别为BI-2013/004E和BI-2013/004F)中。

[0911] 实例6:Cas9直向同源物

[0912] 申请人分析了Cas9直向同源物,以鉴定相关的PAM序列和对应的嵌合 的指导RNA。具有一套展开的PAM提供了跨基因组的更广泛的靶向并且还 显著增加独特靶位点的数目并且提供了在基因组中鉴定具有增加水平的特异 性的新颖的Cas9的潜力。

[0913] 可以通过测试每种Cas9容忍指导RNA及其DNA靶标之间的错配的能力 评价Cas9直向同源物的特异性。例如,已经通过测试指导RNA中的突变对 切割效率的影响表征了SpCas9的特异性。用指导序列与靶DNA之间的单个 或多个错配建立指导RNA文库。基于这些发现,可以基于以下准则选择 SpCas9的靶位点:

[0914] 为了最大化SpCas9编辑具体基因的特异性,应该在感兴趣的座位内选择 一个靶位点,这样使得潜在的'脱靶'基因组序列遵守以下四个约束:首 先,它们的后面不应该是具有的5'-NGG或NAG序列的PAM。其次,它们 与靶序列的整体序列相似性应该是最小化的。第三,最大数目的错配应该位 于脱靶位点的PAM-近侧区内。最后,最大数目的错配应该是连续的或由少于 四个碱基隔开。

[0915] 可以使用类似的方法评价其他Cas9直向同源物的特异性和建立在靶标种 类的基因组内选择特定靶位点的标准。如前所述,基于此套的蛋白质序列保 守的系统发生分析揭示了五个Cas9家族,包括三组大的Cas9(～1400个氨基 酸)和两组小的Cas9(～1100个氨基酸)(参见图19和20A-F)。Cas直向 同源物的另外的细节可以发现于通过引用结合在此的美国申请序列号 61/836,101和61/835,936(代理人案号分别为44790.09.2022和4790.07.2022 并且博德参考号分别为BI-2013/004E和BI-2013/004F)中。

[0916] 实例7:用于简化克隆和递送的方法改进。

[0917] 并非在质粒上编码U6-启动子和指导RNA,申请人用DNA寡核苷酸扩 增了U6启动子,以添加到指导RNA上。可以将生成的PCR产物转染到细胞 中,以驱动指导RNA的表达。

[0918] 允许产生由靶向人EMX1座位的U6-启动子::指导RNA组成的PCR产物 的示例性引物对:

[0919] 正向引物:AAACTCTAGAgagggcctatttcccatgattc

[0920] 反向引物(携带指导RNA,该指导RNA已加下划线):acctctagAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC

[0921] CTTATTTTAACTTGCTATGCTGTTTTGTTTCCAAAACAGCATAGCTCTAAAACCCC TAGTCATTGGAGGTGACGGTGTTTCGTCCTTTCCACaag

[0922] 实例8:用于改善活性的方法改进:

[0923] 并非使用pol3启动子(特别是RNA聚合酶III(例如U6或H1启动子) 在真核细胞中表达指导RNA,申请人在真核细胞中表达T7聚合酶,以使用 T7启动子驱动指导RNA的表达。

[0924] 此系统的一个实例可以涉及引入三段DNA:

[0925] 1.Cas9的表达载体

[0926] 2.T7聚合酶的表达载体

[0927] 3.包含融合到T7启动子上的指导RNA的表达载体

[0928] 实例9:用于减少Cas9的毒性的方法改进:将Cas9以mRNA的形式递送。

[0929] 将Cas9以mRNA的形式递送使得可以在细胞中瞬时表达Cas9,以减少 毒性。例如,可以使用以下引物对扩增人源化SpCas9:

[0930] 正向引物(用以添加到用于体外转录的T7启动子上):TAATACGACTCACTATAGGAAG TGCGCCACCATGGCCCCAAAGAAGAA GCGG

[0931] 反向引物(用以添加到polyA尾上):GGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTttctt aCTTTTCTTTTTTGCCTG GCCG

[0932] 申请人用处于RNA或DNA盒形式的指导RNA将Cas9 mRNA转染进细 胞中,以驱动指导RNA在真核细胞中的表达。

[0933] 实例10:用于减少Cas9的毒性的方法改进:使用诱导型启动子

[0934] 申请人仅当需要进行基因组修饰时才瞬时开启Cas9表达。可诱导系统的 实例包括四环素诱导型启动子(Tet-On或Tet-Off)、小分子双杂交体转录激 活系统(FKBP、ABA等))或光可诱导系统(光敏素、LOV结构域或隐花色 素)。

[0935] 实例11:Cas9系统用于体内应用的改进

[0936] 申请人执行了具有小分子量的Cas9的宏基因组检索。大多数Cas9同系 物相当大。例如,SpCas9的长度为1368aa左右,这对于容易包装进供递送的 病毒载体中而言太大。从存放在GenBank中的序列生成表示出Cas9同系物的 长度分布的图表(图23)。一些序列可能已经被错误注释并且因此每个长度 的准确频率不一定是精确的。尽管如此,但是它提供了对Cas9蛋白分布的模 糊认识并且表明存在较短的Cas9同系物。

[0937] 通过计算分析,申请人发现在细菌菌株弯曲杆菌属中,存在两个具有少 于1000个氨基酸的Cas9蛋白。下文呈现了来自空肠弯曲杆菌的一种Cas9的 序列。以此长度,CjCas9可以被容易地包装进AAV、慢病毒、腺病毒以及其 他用于稳健地递送进原代细胞和递送进动物模型体内的病毒载体中。在本发 明的一个优选实施例中,使用来自金黄色葡萄球菌的Cas9蛋白。

[0938] >空肠弯曲杆菌Cas9(CjCas9)

[0939] MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRL AR

[0940] SARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYE LRFRALN

[0941] ELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSV GEYLYK

[0942] EYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSK KFEEEVL

[0943] SVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLK NTEGIL

[0944] YTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKY KEFIKAL

[0945] GEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHL NISFKAL

[0946] KLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPV VLRAIKE

[0947] YRKVLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDA ELECEK

[0948] LGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSF DDSYMN

[0949] KVLVFTKQNQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKN YKDKEQ

[0950] KNFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHV EAKSGM

[0951] LTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNS AELYAK

[0952] KISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKE EEFYQS

[0953] YGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYT MDFALK

[0954] VLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVY YNAFTSS

[0955] TVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSAL GEVTKA EFRQREDFKK。

[0956] 此CjCas9的假定的tracrRNA元件是：

[0957] TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTG CG GGACTCTGCGGGGTTACA ATCCCCTAAAACCGCTTTTAAAATT

[0958] 同向重复序列是：

[0959] ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC

[0960] CjCas9的嵌合的指导RNA的一个实例是：

[0961] NNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAA AU

[0962] AAAGAGUUUGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCUU UU

[0963] 实例12:Cas9优化

[0964] 用于增强功能或用于开发新的功能,申请人通过组合来自不同Cas9同系 物的片段而产生了嵌合的Cas9蛋白。例如,嵌合的Cas9蛋白的两个实例：

[0965] 例如,申请人将St1Cas9(来自此蛋白的片段为粗体)的N-末端与 SpCas9(来自此蛋白的片段被加下划线)的C-末端融合。

[0966] >St1(N)Sp(C)Cas9

[0967] **MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRR TN RQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTD ELSNEE LFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPG QIQLERY**

QTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQIT DEFINR
YLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYP DEFRAAK
ASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAK LFKYIAK
LLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLD KLAYVLT
LNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSV KLMME
LIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAK SVRQAIK
IVNAAIKEYGDFDNIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI LKEHPVE
NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSID NKVLT

[0968] RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG LSELDK
AGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSD FRKDFQ
FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRK MIAKSEQ
EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF ATVRKV
LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSP TVAYSV
LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLI IKLPKYS
LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE QKQLF
VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIH LFTLTNL
GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0969] >Sp(N)St1(C)Cas9

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLI GA
LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLE ESFLVE
EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA HMIKFRG

[0970] HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKS RRLENL
IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLD NLLAQIG
DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLL KALVRQ

[0971] QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL
NREDLL
RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVG
PLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPK
HSLLYE
YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKE
DYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFE
DREMIE
ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLK
SDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELV
KVMGRHKPENIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAAN
QYNGKA
ELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVD
HILPLSI
TFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRE
SKTLSNK
KKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKID
TKVSVVR
GQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVS
YSEDQL
LDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFN
RKISDATI
YATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLM
YRHDPQT
FEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGP
EIKSLKYY
DSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYA
DLQFEKG
TGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQL
FRFLSRT
MPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISI
YKVRTD VLGNQHIIKNEGDKPKLDF

[0972] 制备嵌合的Cas9的益处包括：

[0973] 减少毒性

[0974] 改善在真核细胞中的表达

[0975] 增强特异性

[0976] 减少蛋白质的分子量,通过组合来自不同Cas9同系物的最小结构域使得 蛋白质更小。

[0977] 改变PAM序列要求

[0978] 实例13:将Cas9作为通用DNA结合蛋白而利用

[0979] 通过使负责切割DNA靶标的两条链的两个催化结构域(D10和H840) 突变,申请人使用Cas9作为通用DNA结合蛋白。为了上调在靶座位的基因 转录,申请人将转录激活结构

域(VP64)融合到Cas9上。申请人推测见到 Cas9-VP64融合蛋白的较强的核定位应该是重要的,因为转录因子激活强度 是在靶标处花费的时间的函数。因此,申请人克隆了一套Cas9-VP64-GFP构 建体,将它们转染进293细胞中并在转染后12小时在荧光显微镜下评估其定位。

[0980] 将相同的构建体克隆为2A-GFP而非直接融合,以便在功能上测试不存 在庞大GFP干扰的构建体。申请人选择用Cas9反式激活因子靶向Sox2座 位,因为这对于细胞重新编程可以是有用的并且该座位已经被验证为TALE- TF介导的转录激活的靶标。对于Sox2座位,申请人选择了转录起始位点 (TSS)附近的八个靶标。每个靶标的长度均为20bp,具有一个相邻的NGG 原型间隔子邻近基序(PAM)。将每个Cas9-VP64构建体与每个PCR产生的 嵌合的crispr RNA(chiRNA)共转染进293细胞中。转染后72小时,使用 RT-qPCR评估转录激活。

[0981] 为了进一步优化转录激活因子,申请人滴定了转染进293细胞中的 chiRNA(Sox2.1和Sox2.5)与Cas9(NLS-VP64-NLS-hSpCas9-NLS-VP64- NLS)的比例,并使用RT-qPCR进行定量。这些结果指示可以将Cas9用作通 用DNA结合结构域,以上调靶座位处的基因转录。

[0982] 申请人设计了第二代构建体。(下表)。

[0983]

pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A,H840A)-NLS
pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A,H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-NLS-hSpCsn1(D10A,H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A,H840A)-NLS
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A,H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-NLS-hSpCsn1(D10A,H840A)

[0984] 申请人使用这些构建体通过RT-qPCR评估转录激活(VP64融合的构建 体)和阻抑(仅有Cas9)。申请人使用抗-His抗体评估每个构建体的细胞定 位,使用Surveyor核酸酶测定评估核酸酶活性,并使用凝胶迁移测定评估 DNA结合亲和性。在本发明的一个优选实施例中,该凝胶迁移测定是EMSA 凝胶迁移测定。

[0985] 实例14:Cas9转基因和基因敲入小鼠

[0986] 为了产生表达Cas9核酸酶的小鼠,申请人提交了两个通用策略转基因和 基因敲入。可以应用这些策略产生任何其他感兴趣的模式生物,例如大鼠。 对于这些通用策略中的每个策略,申请人制备了组成型活性Cas9和条件表达 的Cas9(Cre重组酶依赖性的)。在以下背景下表达组成型活性Cas9核酸 酶:pCAG-NLS-Cas9-NLS-P2A-EGFP-WPRE-bGHpolyA。pCAG是启动子, NLS是核定位信号,P2A是肽切割序列,EGFP是增强型绿色荧光蛋白, WPRE是土拨鼠肝炎病毒转录后调节元件,并且bGHpolyA是牛生长激素 poly-A信号序列(图25A-B)。该条件版本在启动子之后NLS-Cas9-NLS之 前具有一个另外的终止盒元件loxP-SV40 polyA x3-loxP(即pCAG-loxP- SV40polyAx3-loxP-NLS-Cas9-NLS-P2A-EGFP-WPRE-bGHpolyA)。这些重要 的表达元件可以如在图26中可视化。组成型构建体应该在整个发育过程中在 所有细胞类型中进行表达,而条件型构建体将仅当同一细胞正在表达Cre重 组酶时才允许Cas9表达。当Cre处于组织特异的启动子的表达之下时,后一 版本才将允许Cas9的组织特异的表达。此外,可以通过将Cre置于诱导型启 动子(如TET on或off系统)的表达之下而在成年小鼠中诱导Cas9表达。

[0987] Cas9构建体的验证:以三种方式在功能上验证每个质粒:1)瞬时转染 进293个细胞中,随后确认GFP表达;2)瞬时转染进293细胞中,随后使用 识别P2A序列的抗体进行免疫荧光;以及3)瞬时转染,随后是Surveyor核 酸酶测定。取决于感兴趣的细胞,这些293细胞可以是293FT或293T细胞。 在一个优选的实施例中,这些细胞是293FT细胞。条件型和组成型构建体的 Surveyor的结果分别示于凝胶的顶行和底行。在存在和不存在靶向hEMX1座位的嵌合的RNA(嵌合的RNA hEMX1.1)的情况下测试每个构建体。结果 指示,构建体仅在嵌合的RNA的存在下才可以成功地靶向hEMX1座位(并 且在条件型构建体的情况下,是Cre)。将凝胶定量并且将结果呈现为三个样 品的平均的切割效率和标准差。

[0988] 转基因的Cas9小鼠:为了产生具有构建体的转基因小鼠,申请人将纯的 线性DNA注射进来自假怀孕的CB56雌鼠的受精卵的原核中。对首建者小鼠 进行鉴定、基因分型,并与CB57小鼠回交。将构建体成功地克隆并通过桑格 测序进行证实。

[0989] 基因敲入Cas9小鼠:为了产生Cas9基因敲入小鼠,申请人将相同的组 成型和条件型构建体靶向Rosa26座位。申请人通过将每个构建体克隆进具有 以下元件的Rosa26靶向载体中而将其完成:Rosa26短同源臂-组成型/条件型 Cas9表达盒-pPGK-Neo-Rosa26长同源臂-pPGK-DTA。pPGK是赋予对新霉素 的抗性的阳性选择标记Neo、1kb短臂,4.3kb长臂以及由PGK驱动的阴性 选择白喉毒素(DTA)的启动子。

[0990] 将这两种构建体在R1 mESC中电穿孔并且在应用新霉素选择之前,允许 生长2天。挑选存活5-7天的单独的菌落并使其在单独的孔中生长。5-7天 后,收获菌落,将一半冷冻并且将另一半用于基因分型。通过基因组PCR进 行基因分型,其中一个引物在供体质粒(AttpF)内退火并且另一引物在短同 源臂(Rosa26-R)外退火。在针对条件构建体情况下收获的22个菌落中,7 个是阳性的(左侧)。在针对组成型构建体情况下收获的27个菌落中,零个 是阳性的(右侧)。可能的是Cas9在mESC中引起一定水平的毒性并且出于 此原因,不存在阳性克隆。为了对其进行测试,申请人将Cre表达质粒引入 正确靶向的条件型Cas9细胞中并且在多天后在培养物中发现了极低的毒性。 正确靶向的条件型Cas9细胞中的Cas9的减少的拷贝数(1-2个拷贝/个细胞) 足以允许稳定表达和相对地没有细胞毒性。此外,此数据指示,Cas9拷贝数 决定毒性。电穿孔后,每个细胞都应该得到若干拷贝的Cas9并且这可能是为 什么在组成型Cas9构建体的情况下没有发现阳性菌落。这提供了较强的证 据,即利用条件型Cre依赖性策略应该显示出减少的毒性。申请人将正确靶 向的细胞注射进囊胚中并植入雌性小鼠体内。鉴定并回交嵌合体。对首建者 小鼠进行鉴定和基因分型。

[0991] 条件型Cas9小鼠的实用性:申请人已经在293细胞中显示,可以通过与 Cre共表达而激活Cas9条件型表达构建体。申请人还显示,当表达Cre时, 正确靶向的R1 mESC可以具有活性Cas9。因为在Cas9之后为P2A肽切割序 列并且然后为EGFP,申请人通过观察EGFP而鉴定成功表达。这个相同的观 念在于,什么使条件型Cas9小鼠如此有用。申请人可以将其条件型Cas9小 鼠与普遍表达Cre的小鼠(ACTB-Cre系)杂交并且可以得到在每个细胞中表达Cas9的小鼠。应当仅仅采用嵌合的RNA的递送,以便诱导在胚胎小鼠或 成年小鼠中的基因组编辑。有趣的是,如果条件型Cas9小鼠与在组织特异的 启动子下表达Cre的小鼠杂交,在也表达Cre的组织中将会仅有Cas9。通过 将嵌合的RNA递送到相同的组织中,此途径可以用来仅仅在精确的组织中编 辑基因组。

[0992] 实例15:Cas9多样性和嵌合的RNA

[0993] CRISPR-Cas系统是一种由跨细菌和古生菌的相异物种利用的针对侵入型 外源DNA的获得性免疫机制。II型CRISPR-Cas系统由一套编码负责将外源 DNA“捕获”进CRISPR座位中的蛋白质的基因以及一套编码负责“执行” DNA切割机制的蛋白质的基因组成;这些包括DNA核酸酶(Cas9)、非编 码反式激活cr-RNA(tracrRNA)以及侧翼为同向重复的外源DNA衍生的间 隔子阵列(crRNA)。当被Cas9成熟后,tracrRNA和crRNA双链体指导 Cas9核酸酶到由间隔子指导序列规定的靶DNA序列,并且在靶DNA中介导 短序列基序附近的DNA中的双链断裂,该导短序列基序为切割所需并且对于 每种CRISPR-Cas系统而言是特异的。II型CRISPR-Cas系统发现在整个细菌 界中并且在用于靶标切割的Cas9蛋白序列和尺寸、tracrRNA和crRNA同向 重复序列、这些元件的基因组组织以及基序要求方面高度相异。一个物种可 以具有多种相异的CRISPR-Cas系统。

[0994] 申请人评价了来自基于与已知Cas9的序列同源性和与已知亚结构域直向 同源的结构鉴定的细菌种类的207个假定的Cas9,这些亚结构域包括HNH 内切核酸酶结构域和RuvC内切核酸酶结构域[信息来自尤金库宁(Eugene Koonin)和基拉马卡洛娃(Kira Makarova)]。基于此套的蛋白质序列保守的 系统发生分析揭示了五个Cas9家族,包括三组大的Cas9(～1400个氨基酸) 和两组小的Cas9(～1100个氨基酸)(参见图19A-D和20A-F)。

[0995] 申请人还已经使用体外方法优化了Cas9指导RNA。

[0996] 实例16:Cas9突变

[0997] 在此实例中,申请人显示以下突变可以将SpCas9转化为切口酶: D10A、E762A、H840A、N854A、N863A、D986A。

[0998] 申请人提供了显示出序列突变点位于SpCas9基因内何处的序列(图 24A-M)。申请人还显示,这些切口酶仍能够介导同源重组。此外,申请人 显示,具有这些突变的SpCas9(单独地)不诱导双链断裂。

[0999] 所有Cas9直向同源物共享3-4个RuvC结构域和一个HNH结构域的通用 结构。最5’的RuvC结构域切割非互补链,而HNH结构域切割互补链。所有 符号都是就指导序列而言的。

[1000] 在5’RuvC结构域中的催化残基是通过感兴趣的Cas9与其他Cas9直向同 源物(来自化脓链球菌II型CRISPR座位、嗜热链球菌CRISPR座位1、嗜热 链球菌CRISPR座位3以及凶手弗朗西斯菌(Francisciella novicida)II型 CRISPR座位)的同源性比较而鉴定的,并且保守性Asp残基被突变为丙氨酸 以将Cas9转化成互补链切口酶。类似地,在HNH结构域中的保守性His和 Asn残基被突变为丙氨酸以将Cas9转化为非互补链切口酶。

[1001] 实例17:Cas9转录激活和Cas9阻抑物

[1002] Cas9转录激活

[1003] 设计并测试第二代构建体(表1)。使用这些构建体通过RT-qPCR评估 转录激活(VP64融合的构建体)和阻抑(仅有Cas9)。申请人使用抗-His抗 体评估每个构建体的细胞定位,使用Surveyor核酸酶测定评估核酸酶活性, 并使用凝胶迁移测定评估DNA结合亲和性。

[1004] Cas阻抑物

[1005] 先前已经显示,可以将dCas9用作通用DNA结合结构域,以阻抑基因表 达。申请人报告了改进的dCas9设计以及至阻抑物结构域KRAB和SID4x的 dCas9融合。从表1中的建立用于使用Cas9来调制转录的质粒文库中,通过 qPCR在功能上表征了以下阻抑物质粒:

pXRP27、pXRP28、pXRP29、 pXRP48、pXRP49、pXRP50、pXRP51、pXRP52、pXRP53、pXRP56、pXRP58、pXRP59、pXRP61以及pXRP62。

[1006] 将每个dCas9阻抑物质粒都与两个靶向β-连环蛋白基因的编码链的指导RNA共转染。转染后72小时，分离RNA并通过RT-qPCR定量基因表达。内 源对照基因为GAPDH。将两个验证的shRNA用作阳性对照。阴性对照是某 些未用gRNA转染的质粒，这些质粒被表示为“pXRP##对照”。当使用指定 的靶向策略时，质粒pXRP28、pXRP29、pXRP48以及pXRP49可以阻抑β-连 环蛋白基因。这些质粒对应于没有功能性结构域的dCas9(pXRP28和 pXRP28)和融合到SID4x上的dCas9(pXRP48和pXRP49)。

[1007] 另外的工作研究：重复以上实验，靶向不同基因，利用其他gRNA确定 最优靶向位置和多元阻抑。

[1008] 表1

[1009]

pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-$GGGGS_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-$EAAAK_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-$GGGGS_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-$EAAAK_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-$GGGGS_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-$GGGGS_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-$GGGGS_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-$EAAAK_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-$EAAAK_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-$EAAAK_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-$GGGGS_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-$EAAAK_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-$GGGGS_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-$EAAAK_3$接头-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-dCas9-接头-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-dCas9-接头-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-dCas9-接头-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-dCas9-$GGGGGS_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-dCas9-$GGGGGS_3$-NLS-SID4X-gLuc-2A-GFP-WPRE

[1010]

pXRP059-pLenti2-EF1a-dCas9-GGGGGS$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GGGGS$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK$_3$接头-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-Cas9-接头-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-Cas9-接头-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-Cas9-接头-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE

[1011] 实例18:涉及胆固醇生物合成、脂肪酸生物合成以及其他代谢障碍的基因，编码涉及淀粉样蛋白以及其他疾病的错误折叠蛋白的基因,导致细胞转化的 癌基因,潜伏病毒基因,以及导致负显性障碍(在其他障碍中)的基因的靶 向缺失。

[1012] 申请人证明使用病毒或纳米粒子递送系统向对其有需要的受试者或患者 的肝脏、脑、眼、上皮、造血或另一组织进行CRISPR-Cas系统的基因递送，该受试者或患者罹患代谢障碍、淀粉样变性和蛋白聚积相关疾病、由于基因 突变和基因转位所致的细胞转化、

基因突变的负显性作用、潜伏病毒感染以 及其他相关症状。

[1013] 研究设计:罹患代谢障碍、淀粉样变性和蛋白聚积相关疾病的对其有需 要的受试者或患者,包括但不限于人类、非灵长类人类、犬、猫、牛、马、 其他家养动物以及相关哺乳动物。CRISPR-Cas系统由嵌合的指导RNA指导 并靶向有待切割的人类基因组座位的一个特定位点。切割和非同源末端连接 介导的修复之后,移码突变导致基因的敲除。

[1014] 申请人选择靶向涉及上述障碍的基因的指导-RNA,使其以最小脱靶活性 对内源座位特异。可以将两个或更多个指导RNA编码进单个CRISPR阵列 中,以在DNA中诱导同时的双链断裂,从而导致受影响的基因或染色体区域 的微小缺失。

[1015] 基因靶标的鉴定和设计

[1016] 对于每个候选疾病基因,申请人都选择了包括蛋白质编码外显子的感兴 趣DNA序列、包括并侧翼于已知的负显性突变位点的序列、包括并侧翼于病 理重复序列的序列。对于基因敲除途径,离起始密码子最近的早期编码外显 子为实现完全敲除提供了最佳选择并使截短的蛋白质产物保留部分功能的可 能性最小化。

[1017] 申请人分析了紧邻NGG基序(对于SpCas9系统)或NNAGAAW(对于 St1Cas9系统)的5'的所有可能的可靶向的20-bp序列的感兴趣的序列。申请 人选择了用于在基因组中识别独特的单个RNA-指导的Cas9的序列,以使基 于计算算法的脱靶效应最小化,以确定特异性。

[1018] 将指导序列克隆进递送系统中

[1019] 将指导序列合成为双链的20-24bp寡核苷酸。对寡核苷酸进行5'-磷酸化 处理和退火以形成双链体后,取决于递送方法,将寡核苷酸连接进适合的载 体中:

[1020] 基于病毒的递送方法

[1021] 已经在别处描述了基于AAV的载体(PX260、330、334、335)

[1022] 基于慢病毒的载体使用类似的克隆策略,即将指导序列直接连接进单个 载体中,该载体携带一个U6启动子驱动的嵌合的RNA支架和一种EF1a启动 子驱动的Cas9或Cas9切口酶。

[1023] 在别处描述了病毒产生。

[1024] 基于纳米粒子的RNA递送方法

[1025] 1.将指导序列合成为编码T7启动子—指导序列—嵌合的RNA的寡核 苷酸双链。通过PCR方法将一个T7启动子添加到Cas9的5'上。

[1026] 2.将T7驱动的Cas9和指导-嵌合的RNA在体外转录,并且使用商业 试剂盒为Cas9 mRNA进一步加帽和A-添尾(A-tailed)。根据试剂盒说明书 纯化RNA产物。

[1027] 流体动力尾静脉递送方法(用于小鼠)

[1028] 如在上文和在本申请别处所描述的,将指导序列克隆进AAV质粒中。

[1029] 细胞系的体外验证

[1030] 转染

[1031] 1.DNA质粒转染

[1032] 使用基于脂质、化学或电穿孔的方法将携带指导序列的质粒转染进人类 胚肾(HEK293T)或人类胚胎干(hES)细胞、其他相关细胞类型中。对于 HEK293T细胞(~260,000个细胞)的24孔转染,使用Lipofectamine 2000将 总共500ng的DNA转染进每个单孔中。对

于hES细胞的12孔转染,使用 Fugene HD将总共1ug的DNA转染进单孔中。

[1033] 2.RNA转染

[1034] 将如上所述的纯化RNA用于转染进HEK293T细胞中。根据制造商的说 明,可以使用Lipofectamine 2000将1-2ug的RNA转染进～260,000个细胞 中。Cas9和嵌合的RNA的RNA递送显示于图28中。

[1035] 体外插入缺失形成的测定

[1036] 转染后72小时收获细胞并且测定作为双链断裂的指征的插入缺失形成。

[1037] 简言之,使用高保真聚合酶PCR扩增靶序列周围的基因组区域(～400- 600bp扩增子尺寸)。将产物纯化,标准化至相等浓度,并从95℃缓慢退火 至4℃,以允许形成DNA异源双链体。退火后,使用Cel-I酶切割异源双链 体,并且将生成的产物在聚丙烯酰胺凝胶上分离并计算插入缺失效率。

[1038] 动物的体内原理论证

[1039] 递送机制

[1040] 在别处描述了AAV或慢病毒产生。

[1041] 纳米粒子配制品:将RNA混合进纳米粒子配制品中

[1042] 使用商业试剂盒在小鼠中执行DNA质粒的流体动力尾静脉注射。

[1043] 将Cas9和指导序列作为病毒、纳米粒子-包衣的RNA混合物或DNA质 粒而递送,并注射进受试动物体内。一个平行组的对照动物注射无菌盐水、Cas9和GFP或指导序列以及单独的GFP。

[1044] 注射后三周,测试动物的症状的改善并将其处死。针对插入缺失形分析 相关器官系统。表型测定包括HDL、LDL、脂质的血液水平。

[1045] 插入缺失形成的测定

[1046] 使用商业试剂盒从组织中提取DNA;将如针对体外证明所描述的进行插 入缺失测定。

[1047] CRISPR-Cas系统的治疗应用可适合于实现候选疾病基因的组织特异的且 暂时受控的靶向缺失。实例包括涉及胆固醇和脂肪酸代谢、淀粉样蛋白疾 病、负显性疾病、潜伏病毒感染(在其他障碍中)的基因。

[1048] 用于在座位位处引入靶向的插入缺失的单个指导-RNA的实例

[1049]

疾病	基因	间隔子	PAM	机制	参考文献
高胆固醇血症	HMG-CR	GCCAAATTGGACGACCCTCG	CGG	敲除	氟伐他汀：它在高胆固醇血症的管理中的药理学和用途综述（Fluvastatin: a review of its pharmacology and use in the management of hypercholesterolaemia）.（普罗斯科尔 GL（Plosker GL）等人《药物》（Drugs）1996，51(3): 433-459）
高胆固醇血症	SQLE	CGAGGAGACCCCCGTTTCGG	TGG	敲除	非他汀类降胆固醇剂的潜在作用（Potential role of nonstatin cholesterol lowering agents）（特拉帕尼（Trapani）等人《IUBMB 生命》（IUBMB Life），第 63 卷，第 11 期，第 964-971 页，2011 年 11 月）
高血脂症	DGAT1	CCCGCCGCCGCCGTGGCTCG	AGG	敲除	作为抗肥胖和抗糖尿病剂的 DGAT1 抑制剂（DGAT1 inhibitors as anti-obesity and anti-diabetic agents）.（伯奇 AM（Birch AM）等人《药物发现与研发当前观点》（Current Opinion in Drug Discovery & Development）[2010，13(4): 489-496）

[1050]

疾病	基因	间隔子	PAM	机制	参考文献
白血病	BCR-ABL	TGAGCTCTACGAGATCCACA	AGG	敲除	通过 RNA 干扰（RNAi）杀死具有 BCR/ABL 融合基因的白血病细胞（Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi)）（福克斯（Fuchs）等人《癌基因》（Oncogene）2002，21(37): 5716-5724）

[1051] 用于在基因座位位处引入染色体微小缺失的指导-RNA对的实例

	疾病	基因	间隔子	PAM	机制	参考文献
[1052]	高血脂症	PLIN2指导1	CTCAAAATTCATACCGGTTG	TGG	微小缺失	无围脂滴蛋白-2的小鼠受保护免受饮食诱导的肥胖、脂肪炎症和脂肪肝的影响（Perilipin-2 Null Mice are Protected Against Diet-Induced Obesity, Adipose Inflammation and Fatty Liver Disease）（麦克马纳曼JL（McManaman JL）等人《脂质研究杂志》（The Journal of Lipid Research），jlr.M035063.首次公开于2013年2月12日）
	高血脂症	PLIN2指导2	CGTTAAACAACAACCGGACT	TGG	微小缺失	
[1053]	高血脂症	SREBP指导1	TTCACCCCGCGGCGCTGAAT	ggg	微小缺失	小分子桦木醇对SREBP的抑制，改善高血脂症和胰岛素抗性并减少动脉粥样硬化斑块（Inhibition of SREBP by a Small Molecule, Betulin, Improves Hyperlipidemia and Insulin Resistance and Reduces Atherosclerotic Plaques）（唐J（Tang J）等人《细胞代谢》（Cell Metabolism），第13卷，第1期，44-56，2011年1月5日）
	高血脂症	SREBP指导2	ACCACTACCAGTCCGTCCAC	agg	微小缺失	

[1054] 实例19:携带引起疾病的突变的基因的修复的靶向整合;酶缺陷以及其他相 关疾病的重构。

[1055] 研究设计

[1056] I.基因靶标的鉴定和设计

[1057] •描述于实例22中

[1058] II.将指导序列和修复模板克隆进递送系统中

[1059] •以上描述于实例22中

[1060] • 申请人将DNA修复模板克隆为包括具有疾病等位基因的同源 臂以及一个野生型修复模板

[1061] III.细胞系的体外验证

[1062] a.以上转染描述于实例22中;将Cas9、指导RNA和修复模板共 转染进相关细胞类型中。

[1063] b.体外修复的测定

[1064] i.申请人在转染后72小时收获细胞并测定修复

[1065] ii.简言之,申请人使用高保真聚合酶PCR扩增了修复模板周 围的基因组区域。针对突变体等位基因的减少的发生率,申请人对产物进行 了测序。

[1066] IV.动物的体内原理论证

[1067] a.以上递送机制描述于实例22和34中。

[1068] b.体内修复的测定

[1069] i.如描述于体外证明中的,申请人进行了修复测定。

[1070] V.治疗应用

[1071] CRISPR-Cas系统可适合于实现候选疾病基因的组织特异的且暂时受 控的靶向缺失。实例包括涉及胆固醇和脂肪酸代谢、淀粉样蛋白疾病、负显 性疾病、潜伏病毒感染(在其他障碍中)的基因。

[1072] 修复模板的一个单个错义突变的实例:

[1073]

疾病	基因	间隔子	PAM
家族性淀粉样蛋白多神经病	TTR	AGCCTTTCTGAACACATGCA	CGG

[1074]

机制	参考文献
V30M 修复	健康和疾病中的运甲状腺素蛋白突变（Transthyretin mutations in health and disease）（若昂（Joao）等人《人类突变》（Human Mutation），第 5 卷，第 3 期，第 191-196 页，1995）

[1075] V30M

[1076] 等位

[1077] 基因CCTGCCATCAATGTGGCCATGCATGTGTTCAGAAAGGCT

[1078] WT

[1079] 等位

[1080] 基因CCTGCCATCAATGTGGCCGTGCATGTGTTCAGAAAGGCT

[1081] 实例20:CRISPR-Cas系统在青光眼、淀粉样变性和亨廷顿病中的治疗应用

[1082] 青光眼:申请人设计了靶向mycilin(MYOC)基因的第一个外显子的指 导RNA。申请人使用腺病毒载体(Ad5)包装Cas9以及靶向MYOC基因的 指导RNA两者。申请人将腺病毒载体注射进小梁网中,其中的细胞已经牵涉 于青光眼的病理生理学中。申请人最初在携带突变的MYOC基因的小鼠模型 中对其进行了测试,以察看它们是否改善视敏度并降低眼压。人类中的治疗 应用利用类似的策略。

[1083] 淀粉样变性:申请人设计了靶向肝脏中的运甲状腺素蛋白(TTR)基因 的第一个外显子的指导RNA。申请人使用AAV8包装Cas9以及靶向TTR基 因的第一个外显子的指导RNA。

AAV8已经显示具有有效的肝脏靶向并且将 被静脉内给予。可以使用肝脏特异的启动子(如白蛋白启动子)或使用组成 型启动子驱动Cas9。pol3启动子驱动指导RNA。

[1084] 可替代地,申请人利用质粒DNA的流体动力递送来敲除TTR基因。申 请人递送了一个编码Cas9的质粒和靶向TTR的外显子1的指导RNA。

[1085] 作为一种另外的替代途径,申请人给予了RNA的组合(Cas9的mRNA 和指导RNA)。可以使用脂质体(如来自生命技术公司的Invivofectamine) 包装RNA并静脉内递送。为了减少RNA诱导的免疫原性、增加Cas9表达的 水平和指导RNA稳定性,申请人使用5’加帽修饰了Cas9 mRNA。申请人还 将修饰的RNA核苷酸掺入了Cas9 mRNA和指导RNA中,以增加其稳定性并 减少免疫原性(例如TLR的激活)。为了增加效率,申请人给予了多剂量的 病毒、DNA或RNA。

[1086] 亨廷顿病:申请人基于患者的HTT基因中的等位基因特异的突变设计了 指导RNA。例如,在对HTT而言杂合的、具有扩增的CAG重复的患者体 内,申请人鉴定了对于突变HTT等位基因而言独特的的核苷酸序列并将其用 于设计指导RNA。申请人保证突变甲基位于指导RNA的最后9bp内(诸位 申请人已经确定具有区别靶标尺寸与指导RNA之间的单个DNA碱基错配的 能力)。

[1087] 申请人将突变HTT等位基因特异的指导RNA和Cas9包装进AAV9中并 递送进亨廷顿病患者的纹状体中。经由开颅术将病毒立体定向地注射进纹状 体中。已知AAV9有效地转导神经元。申请人使用神经元特异的启动子(如 人类突触蛋白I)驱动Cas9。

[1088] 实例21:CRISPR-Cas系统在HIV中的治疗应用

[1089] 慢性病毒感染是显著的发病率和死亡率的来源。虽然针对这些病毒中的 许多病毒存在有效靶向病毒复制的不同方面的常规的抗病毒疗法,但是由于 “病毒潜伏期”当前治疗形式在本质上通常是非治愈性的。就其本质而言, 病毒潜伏期由病毒生命周期中的未进行活性病毒产生的休眠期表征。在此阶 段过程中,病毒在很大程度上能够逃避免疫监视和常规治疗两者,从而允许 它在宿主体内建立长期存在的病毒储库,从这些病毒储库随后重新激活可以 允许病毒继续繁殖和传播。病毒潜伏期的关键在于通过游离型或前病毒潜伏 而完成的稳定地维持病毒基因组,该游离型或前病毒潜伏分别将病毒基因组 存储在细胞质中或将它整合在宿主基因组中。在不存在预防初次感染的有效 疫苗接种的情况下,由潜伏贮存器表征的慢性病毒感染和溶解活性的发作可 以具有显著的后果:人乳头瘤病毒(HPV)可以导致宫颈癌,丙型肝炎病毒 (HCV)容易诱发肝细胞癌,并且人类免疫缺陷病毒最终摧毁宿主免疫系 统,从而使得对机会性感染易感。因此,这些感染需要终生使用当前可获得 的抗病毒治疗。更复杂的事情是这些病毒基因组中的许多的较高的可突变 性,这导致了对其而言不存在有效疗法的抗性菌株的进化。

[1090] CRISPR-Cas系统是一种能够以可以多元的、序列特异的方式诱导双链 DNA断裂(DSB)的细菌获得性免疫系统并且最近已经在哺乳动物细胞系统 内重构。已经显示,用一个或众多指导-RNA靶向DNA可以分别导致间插序 列的插入缺失和缺失两者。因此,这一新的技术代表了一种手段,通过该手 段可以在单个细胞中以较高效率和特异性完成靶向和多元DNA诱变。结果, 递送直接针对病毒DNA序列的CRISPR-Cas系统甚至在不存在正在进行的病 毒产生的情况下仍可以允许靶向破坏和缺失潜伏的病毒基因组。

[1091] 作为一个实例,被HIV-1慢性感染代表了一个全球性健康问题,其中影 响的个人

为3300万并且年发生率为260万次感染。使用同时靶向病毒复制的 多个方面的高效抗逆转录病毒疗法(HAART)已经允许在很大程度上将HIV 感染作为一种慢性而非晚期病疾进行管理。不进行治疗的话,HIV进展为 AIDS通常在9-10年内发生,从而导致耗竭宿主免疫系统并且机会性感染的 出现通常将在之后不久导致死亡。继发于病毒潜伏期,HAART的中断总是导 致病毒反跳。此外,针对通过可获得的手段无法控制的HIV抗性菌株,甚至 可以在疗法方面选择暂时中断。另外,HAART治疗的费用是显著的:在美国 每年每人$10,000-15,0000。因此,直接靶向HIV基因组而非病毒复制过程的 治疗途径代表了一种手段,通过该手段潜伏的贮存器的根除可以允许治愈性 治疗选择。

[1092] HIV-1靶向的CRISPR-Cas系统的研发和递送代表了一种可与现存的靶向 DNA诱变的手段可区分的独特途径,即具有众多治疗意义的ZFN和 TALEN。CRISPR介导的DSB对HIV-1基因组的靶向破坏和缺失以及与 HAART结合的插入缺失可以允许同时预防活性病毒产生以及宿主体内的潜伏 的病毒储库的耗竭。

[1093] 一旦整合进宿主免疫系统内,CRISPR-Cas系统允许产生HIV-1抗性亚群 体,该亚群体即使在不存在彻底的病毒根除的情况下仍可以允许维持和重构 宿主免疫活性。这可以通过破坏病毒基因组而潜在地预防初次感染,从而预 防病毒产生和整合,代表了一种“疫苗接种”手段。CRISPR-Cas系统的多元 性质允许同时靶向单独细胞内的基因组的多个方面。

[1094] 如在HAART中,通过需要同时获取多个获得性突变而使病毒通过诱变 而进行的逃避最小化。可以同时靶向HIV-1的多种菌株,这使重复感染的机 会最小化并防止随后产生新的重组菌株。核苷酸而非蛋白质介导的CRISPR- Cas系统的序列特异性允许快速产生治疗而无需显著改变递送机制。

[1095] 为了将其完成,申请人产生了靶向绝大多数HIV-1基因组的CRISPR-Cas 指导RNA,同时考虑了HIV-1株变体,以使覆盖面和效力最大。HIV-1亚型 和变体之间的基因组保守的序列分析应该允许靶向基因组的侧翼保守区,目 的是缺失间插病毒序列或诱导将破坏病毒基因功能的移码突变。

[1096] 通过宿主免疫系统的常规腺病毒或慢病毒介导的感染,申请人完成了 CRISPR-Cas系统的递送。取决于途径,宿主免疫细胞可以是a)分离的、用 CRISPR-Cas转导的、选择的、并且重新引入该宿主中的或b)通过全身递送 CRISPR-Cas系统体内转导的。第一种途径允许产生抗性免疫群体,而第二种 更可能靶向宿主体内的潜伏病毒储库。

[1097]

从麦金太尔（Mcintyre）等人改适的潜在的 HIV-1 靶向的间隔子的实例，其产生针对 HIV-1 的 shRNA，优化用于最大覆盖 HIV-1 变体。
CACTGCTTAAGCCTCGCTCGAGG TCACCAGCAATATTCGCTCGAGG CACCAGCAATATTCCGCTCGAGG TAGCAACAGACATACGCTCGAGG GGGCAGTAGTAATACGCTCGAGG CCAATTCCCATACATTATTGTAC

[1098] 实例22:囊性纤维化中的δF508或其他突变的靶向修正

[1099] 本发明的一个方面提供了一种药物组合物,该药物组合物可以包括一种 CRISPR-Cas基因疗法粒子和一种生物相容的药物载体。根据另一个方面,一 种用于治疗在CFTR基

因中具有突变的受试者的基因疗法方法包括向受试者 的细胞给予治疗有效量的CRISPR-Cas基因疗法粒子。

[1100] 此实例展示了使用腺相关病毒(AAV)粒子的CRISPR-Cas系统在对其 有需要的受试者或患者的气道中的基因转移或基因递送,所述受试者或患者 患有囊性纤维化或囊性纤维化相关症状。

[1101] 研究设计:对其有需要的受试者或患者:相关的人类、非灵长类人类、犬、猫、牛、马以及其他家养动物。此研究测试了通过AAV载体对CRISPR- Cas系统进行基因转移的效力。申请人确定了足以进行基因表达的转基因水平 并利用了一种包括靶向δF508或其他CFTR诱导的突变的Cas9酶的CRISPR- Cas系统。

[1102] 这些经治疗的受试者的每一侧肺接受支气管内递送的药学上有效量的雾 化AAV载体系统,同时自然地呼吸。对照受试者接受等量的具有内部对照基 因的假型AAV载体系统。该载体系统可以与一种药学上可接受的生物或相容 的药物载体一起递送。给予载体三周或适当的时间间期后,测试经治疗的受 试者的囊性纤维化相关症状的改善。

[1103] 申请人使用了一种腺病毒或一种AAV粒子。

[1104] 申请人将以下各自可操作地连接到一种或多种调节序列(对于Cas9而言 是Cbh或EF1a启动子,对于嵌合的指导RNA是U6或H1启动子)上的基因 构建体克隆进一种或多种腺病毒或AAV载体或任何其他相容的载体中:靶向 嵌合的指导RNA的CFTRδ508(图31B),δF508突变的修复模板(图31C)以及具有任选地一个或多个核定位信号或序列(NLS)(例如,两个(2)NLS)的密码子优化的Cas9酶。

[1105] Cas9靶位点的鉴定

[1106] 申请人分析了人类CFTR基因组座位并鉴定了Cas9靶位点(图31A)。(PAM可以包含一个NGG或一个NNAGAAW基序)。

[1107] 基因修复策略

[1108] 申请人经由先前讨论的递送方法之一将包括Cas9(或Cas9切口酶)和 指导RNA的腺病毒/AAV载体系统连同包括包含F508残基的同源性修复模板 的腺病毒/AAV载体系统一起引入受试者体内。CRISPR-Cas系统由 CFTRδ508嵌合的指导RNA指导并靶向有待切口或切割的CFTR基因组座位 的一个特定位点。在切割之后,将修复模板经由同源重组插入切割位点中, 从而修正导致囊性纤维化或引起囊性纤维化相关症状的缺失。可以采用用来 指导递送并提供CRISPR系统的系统性引入的此策略来靶向基因突变,以编 辑或另外地操纵如那些在表B中的引起代谢、肝脏、肾和蛋白质疾病和障碍 的基因。

[1109] 实例23:基因敲除细胞文库的产生

[1110] 此实例展示了如何产生其中的每个细胞都具有单基因被敲除的细胞文 库:

[1111] 申请人制备了ES细胞的文库,其中每个细胞都具有单基因被敲除,并且 ES细胞的整个文库将有每个单基因的敲除。此文库对于筛选细胞过程以及疾 病的基因功能是有用的。

[1112] 为了制备此细胞文库,申请人将由诱导型启动子(例如,多西环素诱导 型启动子)驱动的Cas9整合到ES细胞中。另外,申请人将靶向特异的基因 的单个指导RNA整合到ES细胞中。为了制备ES细胞文库,申请人简单地 将ES细胞与编码靶向人类基因组中的每个基因的指导RNA的基因的文库混 合。申请人首先将单个BxB1 attB位点引入人类ES细胞的AAVS1

座位中。然后,申请人使用BxB1整合酶来促进各个的指导RNA基因整合到AAVS1 座位中的BxB1 attB位点中。为了促进整合,每个指导RNA基因都被包含在 携带单个attP位点的质粒上。这样,BxB1将使基因组中的attB位点与含有指 导RNA的质粒上的attP位点重组。

[1113] 为了产生该细胞文库,申请人采用具有整合的单个指导RNA的细胞的文 库,并且诱导Cas9表达。在诱导之后,Cas9介导由该指导RNA指定的位点 处的双链断裂。为了验证此细胞文库的多样性,申请人进行了全外显子组测 序,以保证申请人能够在每个单靶向基因中观察到突变。此细胞文库可以用 于多种应用(包括基于全文库的筛选),或可以被分选进各个细胞克隆中以 促进快速产生具有各个人类基因被敲除的克隆细胞系。

[1114] 实例24:使用Cas9工程化微藻

[1115] 递送Cas9的方法

[1116] 方法1:申请人使用在组成型启动子(如Hsp70A-Rbc S2或β2-微管蛋 白)的控制之下表达Cas9的载体递送Cas9和指导RNA。

[1117] 方法2:申请人使用在组成型启动子(如Hsp70A-Rbc S2或β2-微管蛋 白)的控制之下表达Cas9和T7聚合酶的载体递送Cas9和T7聚合酶。将使 用包含驱动指导RNA的T7启动子的载体递送该指导RNA。

[1118] 方法3:申请人将Cas9 mRNA和体外转录的指导RNA递送到藻类细胞 中。RNA可以在体外转录。Cas9 mRNA将由Cas9的编码区以及来自的Cop1 的3'UTR组成,以保证Cas9 mRNA的稳定。

[1119] 用于同源重组,申请人提供了一个另外的同源定向修复模板。

[1120] 在位于Cop1的3'UTR之后的β-2微管蛋白启动子的控制之下驱动Cas9 的表达的盒的序列。

[1121] TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGA GACGGCTTCCCGGCGCTGC ATGCAACACCGATGATGCTTCGACCCCCCG AAGCTC

[1122] CTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTT TAAATA

[1123] GCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATAT TCAAAC

[1124] ACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCA CTCCGC

[1125] TAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACATGTA CCCATA

[1126] CGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGC GTCCGA

[1127] CAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGG GCCGTG

[1128] ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGC AACACC

[1129] GACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACA GCGGC

[1130] GAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATAC ACCAG

[1131] ACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATG GCCAA

[1132] GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAA GAGGAT

[1133] AAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTG GCCTAC

[1134] CACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGAC AGCACC

[1135] GACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCA AGTTCC

[1136] GGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGT GGACA

[1137] AGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAA CCCCAT

[1138] CAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGC AAGAG

[1139] CAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAA TGGCCT

[1140] GTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAG AGCAAC

[1141] TTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACG ACGAC

[1142] GACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGT TTCTGG

[1143] CCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGT GAACAC

[1144] CGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGAC GAGCA

[1145] CCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCT GAGAAG

[1146] TACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACA TTGACG

[1147] GCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGA AAAGA

[1148] TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGC TGCGGA

[1149] AGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGG AGAGC

[1150] TGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGA CAACCG

[1151] GGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGC CCTCTG

[1152] GCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAA ACCATC

[1153] ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGA GCTTCA

[1154] TCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCT GCCCA

[1155] AGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAA AGTGAA

[1156] ATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCA GAAAAA

[1157] GGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAG CAGCT

[1158] GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC TCCGGC

[1159] GTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGA AAATTA

[1160] TCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGG AAGATA

[1161] TCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACG GCTGA

[1162] AAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCG GCGGA

[1163] GATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCC GGGACA

[1164] AGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGC CAACAG

[1165] AAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGAC ATCCAG

[1166] AAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCC AATCTG

[1167] GCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTG GTGGAC

[1168] GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATC GAAATG

[1169] GCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGA GAGAAT

[1170] GAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAA AGAACA

[1171] CCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTAC CTGCAG

[1172] AATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTG TCCGAC

[1173] TACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCA TCGACA

[1174] ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACG TGCCCT

[1175] CCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGA ACGCCA

[1176] AGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAG GCGGCC

[1177] TGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAA CCCGGC

[1178] AGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTA AGTACG

[1179] ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGT CCAAGC

[1180] TGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGAT CAACAA

[1181] CTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCC CTGATC

[1182] AAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAG GTGTAC

[1183] GACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCT ACCGCC

[1184] AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTAC CCTGG

[1185] CCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAA CCGGGG

[1186] AGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCT GAGCA

[1187] TGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCT TCAGCA

[1188] AAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAA AGAAGG

[1189] ACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTA TTCTGT

[1190] GCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAG TGTGAA

[1191] AGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAA TCCCAT

[1192] CGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGAT CATCAA

[1193] GCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATG CTGGC

[1194] CTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAA TATGT

[1195] GAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCC GAGGAT

[1196] AATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGAC GAGATC

[1197] ATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTA ATCTGG

[1198] ACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAG AGCAGG

[1199] CCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGC CGCCTT

[1200] CAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAA AGAGGT

[1201] GCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACA CGGAT

[1202] CGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGT GGAGGC

[1203] CAGCTAAGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGA GCCCCTC

[1204] CCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGT CAACC

[1205] CGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCA GTTGAT GGTACT

[1206] 在位于Cop1的3'UTR之后的β-2微管蛋白启动子的控制之下驱动T7聚 合酶的表达的盒的序列：

[1207] TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGA GACGGCTTCCCGGCGCTGC ATGCAACACCGATGATGCTTCGACCCCCCG AAGCTC

[1208] CTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTT TAAATA

[1209] GCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATAT TCAAAC

[1210] ACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCA CTCCGC

[1211] TAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACatgcctaa gaagaaga

[1212] ggaaggttaacacgattaacatcgctaagaacgacttctctgacatcgaactggctgctatcccgttc aacactctggctg accatt

[1213] acggtgagcgtttagctcgcgaacagttggcccttgagcatgagtcttacgagatgggtgaagcacgc ttccgcaagat gtttga

[1214] gcgtcaacttaaagctggtgaggttgcggataacgctgccgccaagcctctcatcactaccctactcc ctaagatgattg cacgc

[1215] atcaacgactggtttgaggaagtgaaagctaagcgcggcaagcgcccgacagccttccagttcctgca agaaatcaag ccgga

[1216] agccgtagcgtacatcaccattaagaccactctggcttgcctaaccagtgctgacaatacaaccgttc aggctgtagcaa gcgca

[1217] atcggtcgggccattgaggacgaggctcgcttcggtcgtatccgtgaccttgaagctaagcacttcaa gaaaaacgttg aggaa

[1218] caactcaacaagcgcgtagggcacgtctacaagaaagcatttatgcaagttgtcgaggctgacatgct ctctaagggtct actcg

[1219] gtggcgaggcgtggtcttcgtggcataaggaagactctattcatgtaggagtacgctgcatcgagatg ctcattgagtca accgg

[1220] aatggttagcttacaccgccaaaatgctggcgtagtaggtcaagactctgagactatcgaactcgcac ctgaatacgctg aggct

[1221] atcgcaacccgtgcaggtgcgctggctggcatctctccgatgttccaaccttgcgtagttcctcctaa gccgtggactgg cattac

[1222] tggtggtggctattgggctaacggtcgtcgtcctctggcgctggtgcgtactcacagtaagaaagcac tgatgcgctacg aagac

[1223] gtttacatgcctgaggtgtacaaagcgattaacattgcgcaaaacaccgcatggaaaatcaacaagaa agtcctagcgg tcgcc

[1224] aacgtaatcaccaagtggaagcattgtccggtcgaggacatccctgcgattgagcgtgaagaactccc gatgaaaccg gaaga

[1225] catcgacatgaatcctgaggctctcaccgcgtggaaacgtgctgccgctgctgtgtaccgcaaggaca aggctcgcaa gtctcg

[1226] ccgtatcagccttgagttcatgcttgagcaagccaataagtttgctaaccataaggccatctggttcc cttacaacatggac tggcg

[1227] cggtcgtgtttacgctgtgtcaatgttcaacccgcaaggtaacgatatgaccaaaggactgcttacgc tggcgaaaggta aacca

[1228] atcggtaaggaaggttactactggctgaaaatccacggtgcaaactgtgcgggtgtcgacaaggttcc gttccctgagc gcatca

[1229] agttcattgaggaaaaccacgagaacatcatggcttgcgctaagtctccactggagaacacttggtgg gctgagcaaga ttctcc

[1230] gttctgcttccttgcgttctgctttgagtacgctggggtacagcaccacggcctgagctataactgct ccccttccgctggcg tttgac

[1231] gggtcttgctctggcatccagcacttctccgcgatgctccgagatgaggtaggtggtcgcgcggttaa cttgcttcctagt gaaac

[1232] cgttcaggacatctacgggattgttgctaagaaagtcaacgagattctacaagcagacgcaatcaatg

ggaccgataac gaagta

[1233] gttaccgtgaccgatgagaacactggtgaaatctctgagaaagtcaagctgggcactaaggcactggc tggtcaatggc tggct

[1234] tacggtgttactcgcagtgtgactaagcgttcagtcatgacgctggcttacgggtccaaagagttcgg cttccgtcaacaa gtgct

[1235] ggaagataccattcagccagctattgattccggcaagggtctgatgttcactcagccgaatcaggctg ctggatacatgg ctaag

[1236] ctgatttgggaatctgtgagcgtgacggtggtagctgcggttgaagcaatgaactggcttaagtctgc tgctaagctgctg gctgc

[1237] tgaggtcaaagataagaagactggagagattcttcgcaagcgttgcgctgtgcattgggtaactcctg atggtttccctgt gtggc

[1238] aggaatacaagaagcctattcagacgcgcttgaacctgatgttcctcggtcagttccgcttacagcct accattaacacca acaa

[1239] agatagcgagattgatgcacacaaacaggagtctggtatcgctcctaactttgtacacagccaagacg gtagccaccttc gtaag

[1240] actgtagtgtgggcacacgagaagtacggaatcgaatcttttgcactgattcacgactccttcggtac gattccggctgac gctgc

[1241] gaacctgttcaaagcagtgcgcgaaactatggttgacacatatgagtcttgtgatgtactggctgatt tctacgaccagttc gctga

[1242] ccagttgcacgagtctcaattggacaaaatgccagcacttccggctaaaggtaacttgaacctccgtg acatcttagagtc ggact

[1243] tcgcgttcgcgtaaGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTG AGCCC

[1244] CTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACG GGTCA

[1245] ACCCGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACG TCAGTT GATGGTACT

[1246] 由T7启动子驱动的指导RNA的序列(T7启动子,N表示靶向序列)：gaaatTAATACGA CTCACTATANNNNNNNNNNNNNNNNNNNgttttagagc taGAAAtagcaagttaaaataaggctagtccgttat caacttgaaaaagtggcaccgagtcggtgctttttt

[1247] 基因递送：

[1248] 来自衣藻资源中心(Chlamydomonas Resource Center)的莱茵衣藻株CC- 124和CC-125将被用于电穿孔。电穿孔方案遵循来自GeneArt Chlamydomonas Engineering试剂盒的标准推荐方案。

[1249] 而且,申请人产生组成性地表达Cas9的莱茵衣藻系。这可以通过使用 pChlamy1(使用PvuI线性化)并选择潮霉素抗性菌落而完成。以下是包含 Cas9的pChlamy1的序列。以此方式实现基因敲除,简单地需要递送指导 RNA的RNA。对于同源重组,申请人递送了指导RNA以及线性化同源重组 模板。

[1250] pChlamy1-Cas9：

[1251] TGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGC

[1252] GCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCG CTCATGA

[1253] GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAG TTTTAA

[1254] ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGC TTAATC

[1255] AGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGC CTGACT

[1256] CCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCC AGTGCT

[1257] GCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAA TAAACC

[1258] AGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGC CTCCAT

[1259] CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTT AATAGT

[1260] TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTC GTTTGG

[1261] TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGA TCCCCC

[1262] ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAG AAGTA

[1263] AGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCT CTTACT

[1264] GTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAA GTCATT

[1265] CTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATA CGGGAT

[1266] AATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAAC GTTCTT

[1267] CGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGAT GTAACC

[1268] CACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTT CTGGGT

[1269] GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCG ACACGG

[1270] AAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTA TCAGGG

[1271] TTATTGTCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAG CGTCAG

[1272] ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGC GTAATC

[1273] TGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGC CGGATC

[1274] AAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCA GATACC

[1275] AAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACT CTGTA

[1276] GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGTTGC CAGTGG

[1277] CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGAT AAGGCG

[1278] CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAG CGAACG

[1279] ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCC ACGCTTC

[1280] CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAA CAGGA

[1281] GAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTC CTGTCG

[1282] GGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGG GGGCGG

[1283] AGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCT TTTGCTG

[1284] GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAA CCGTA

[1285] TTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGA GCGCAG

[1286] CGAGTCAGTGAGCGAGGAAGCGGTCGCTGAGGCTTGACATGATTGGTGC GTATGT

[1287] TTGTATGAAGCTACAGGACTGATTTGGCGGGCTATGAGGGCGGGGGAA GCTCTGG

[1288] AAGGGCCGCGATGGGGCGCGCGGCGTCCAGAAGGCGCCATACGGCCCG CTGGCG

[1289] GCACCCATCCGGTATAAAAGCCCGCGACCCCGAACGGTGACCTCCACTT TCAGCG

[1290] ACAAACGAGCACTTATACATACGCGACTATTCTGCCGCTATACATAACC ACTCAG

[1291] CTAGCTTAAGATCCCATCAAGCTTGCATGCCGGGCGCGCCAGAAGGAGC GCAGC

[1292] CAAACCAGGATGATGTTTGATGGGGTATTTGAGCACTTGCAACCCTTAT CCGGAA

[1293] GCCCCCTGGCCCACAAAGGCTAGGCGCCAATGCAAGCAGTTCGCATGCA GCCCCT

[1294] GGAGCGGTGCCCTCCTGATAAACCGGCCAGGGGGCCTATGTTCTTTACT TTTTTAC

[1295] AAGAGAAGTCACTCAACATCTTAAAATGGCCAGGTGAGTCGACGAGCA AGCCCG

[1296] GCGGATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGT GTCGAC

[1297] GAAGGCTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCG TCGCCG

[1298] TTTCCATTTGCAGGAGATTCGAGGTACCATGTACCCATACGATGTTCCAG ATTACG

[1299] CTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACA GCATCG

[1300] GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGA GTACAA

[1301] GGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAG CATCAA

[1302] GAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGA GGCCAC

[1303] CCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCG GATCTG

[1304] CTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAG CTTCTTC

[1305] CACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAG CGGCAC

[1306] CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACC CCACCA

[1307] TCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACC TGCGGC

[1308] TGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTG ATCGA

[1309] GGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCA GCTGGT

[1310] GCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGG CGTGGA

[1311] CGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGA AAATCT

[1312] GATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTG ATTGC

[1313] CCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCC GAGGAT

[1314] GCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAAC CTGCTG

[1315] GCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACC TGTCCG

[1316] ACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAA GGCCCC

[1317] CCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTG ACCCTG

[1318] CTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTT TCTTCG

[1319] ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCC AGGAAG

[1320] AGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGA GGAAC

[1321] TGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCT TCGACA

[1322] ACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCT GCGGCG

[1323] GCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGA GAAGAT

[1324] CCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAAC AGCAGA

[1325] TTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACT TCGAG

[1326] GAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATG ACCAAC

[1327] TTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGC TGTAC

[1328] GAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCG AGGGA

[1329] ATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTG GACCT

[1330] GCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGA CTACTT

[1331] CAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGAT CGGTTC

[1332] AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACA AGGACT

[1333] TCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGAC CCTGA

[1334] CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATG CCCACC

[1335] TGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCG GCTGGG

[1336] GCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCG GCAAGA

[1337] CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCAT GCAGCT

[1338] GATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCA GGTGTC

[1339] CGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGC CCCGC

[1340] CATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGT GAAAGT

[1341] GATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGA GAACCA

[1342] GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGA TCGAAG

[1343] AGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGG AAAACA

[1344] CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCG GGATA

[1345] TGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGT GGACC

[1346] ATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGT GCTGAC

[1347] CAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGA GGTCG

[1348] TGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGA TTACCC

[1349] AGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCG AACTGG

[1350] ATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCA CAAAGC

[1351] ACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGA ATGACA

[1352] AGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTC CGATTT

[1353] CCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCAC CACGCC

[1354] CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGT ACCCTA

[1355] AGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGC GGAAGA

[1356] TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACT TCTTCTA

[1357] CAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGC GAGATC

[1358] CGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTG TGGGAT

[1359] AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAG TGAAT

[1360] ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCT ATCCTG

[1361] CCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGAC CCTAAG

[1362] AAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGG TGGCCA

[1363] AAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGC TGGGG

[1364] ATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTC TGGAAG

[1365] CCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTA AGTACT

[1366] CCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGG CGAAC

[1367] TGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCT GTACC

[1368] TGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGC AGAAAC

[1369] AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCA GATCA

[1370] GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGT GCTGTC

[1371] CGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAA TATCAT

[1372] CCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACT TTGACA

[1373] CCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACG CCACCC

[1374] TGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTC TCAGCT

[1375] GGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAAC ATATGATTCGAATGTCTTTC TTGCGCTATGACACTTCCAGCAAAAGGTAG GGCG

[1376] GGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCG ACCCCC

[1377] CGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGA GCGCTG

[1378] TTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAA AGCCAT

[1379] ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTG TGATCG

[1380] CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCA AACAT

[1381] GACACAAGAATCCCTGTTACTTCTCGACCGTATTGATTCGGATGATTCCT ACGCG

[1382] AGCCTGCGGAACGACCAGGAATTCTGGGAGGTGAGTCGACGAGCAAGC CCGGCG

[1383] GATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTC GACGAA

[1384] GGCTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCG CCGTTT

[1385] CCATTTGCAGCCGCTGGCCCGCCGAGCCCTGGAGGAGCTCGGGCTGCCG GTGCCG

[1386] CCGGTGCTGCGGGTGCCCGGCGAGAGCACCAACCCCGTACTGGTCGGCG AGCCC

[1387] GGCCCGGTGATCAAGCTGTTCGGCGAGCACTGGTGCGGTCCGGAGAGCC TCGCG

[1388] TCGGAGTCGGAGGCGTACGCGGTCCTGGCGGACGCCCCGGTGCCGGTGC CCCGC

[1389] CTCCTCGGCCGCGGCGAGCTGCGGCCCGGCACCGGAGCCTGGCCGTGGC CCTACC

[1390] TGGTGATGAGCCGGATGACCGGCACCACCTGGCGGTCCGCGATGGACG GCACGA

[1391] CCGACCGGAACGCGCTGCTCGCCCTGGCCCGCGAACTCGGCCGGGTGCT CGGCCG

[1392] GCTGCACAGGGTGCCGCTGACCGGGAACACCGTGCTCACCCCCCATTCC GAGGTC

[1393] TTCCCGGAACTGCTGCGGGAACGCCGCGCGGCGACCGTCGAGGACCACC GCGGG

[1394] TGGGGCTACCTCTCGCCCCGGCTGCTGGACCGCCTGGAGGACTGGCTGC CGGACG

[1395] TGGACACGCTGCTGGCCGGCCGCGAACCCCGGTTCGTCCACGGCGACCT GCACGG

[1396] GACCAACATCTTCGTGGACCTGGCCGCGACCGAGGTCACCGGGATCGTC GACTTC

[1397] ACCGACGTCTATGCGGGAGACTCCCGCTACAGCCTGGTGCAACTGCATC TCAACG

[1398] CCTTCCGGGGCGACCGCGAGATCCTGGCCGCGCTGCTCGACGGGGCGCA GTGGA

[1399] AGCGGACCGAGGACTTCGCCCGCGAACTGCTCGCCTTCACCTTCCTGCA CGACTT

[1400] CGAGGTGTTCGAGGAGACCCCGCTGGATCTCTCCGGCTTCACCGATCCG GAGGAA

[1401] CTGGCGCAGTTCCTCTGGGGGCCGCCGGACACCGCCCCCGGCGCCTGAT AAGGAT

[1402] CCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAG TGTGTT

[1403] TGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACCCGAAC AGATTG

[1404] ATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGGTACT

[1405] 对于所有修饰的莱茵衣藻细胞而言，申请人使用了PCR、SURVEYOR 核酸酶测定和DNA测序，以验证成功的修饰。

[1406] 实例25：Cas9靶向多种疾病类型的用途

[1407] 涉及蛋白质编码序列中的突变的疾病：

[1408] 可以通过靶向负显性等位基因而靶向显性障碍。申请人使用Cas9靶向负 显性等位基因中的独特序列并经由NHEJ引入突变。NHEJ诱导的插入缺失可 能能够在负显性等位基因中引入移码突变并消除负显性蛋白。如果该基因是 充分单的(haplo-sufficient),这可以起作用(例如MYOC突变诱导的青光眼 和亨廷顿病)。

[1409] 可以通过修复两个等位基因中的疾病突变而靶向隐性障碍。对于分裂细 胞,申请人使用Cas9在突变位点附近引入双链断裂并使用外源重组模板增加 同源重组率。对于分裂细胞,这可以经由携带互补突出端的外源DNA片段的 NHEJ介导的连接使用催化两个等位基因中的突变序列的替换的多元切口酶活 性而实现。

[1410] 申请人还使用Cas9引入保护性突变(例如,用于预防HIV感染的CCR5 的失活、用于减少胆固醇或将A673T引入APP中以减少阿尔茨海默病的可能 性的PCSK9的失活)。

[1411] 涉及非编码序列的疾病

[1412] 申请人使用Cas9破坏启动子区域中的非编码序列,以改变转录因子结合 位点并改变增强子或阻抑子元件。例如,可以使用Cas9切除造血干细胞中的 Klf1增强子EHS1,以减少BCL11a水平并重新激活分化的红细胞中的胎珠蛋 白基因表达。

[1413] 申请人还使用Cas9破坏5'或3'非翻译区中的功能性基序。例如,用于强 直性肌营养不良的治疗,可以使用Cas9除去DMPK基因中的CTG重复扩 展。

[1414] 实例26:多元切口酶

[1415] 也可以使用在本申请中详细说明的Cas9的优化和教导的方面产生Cas9 切口酶。申请人使用与指导RNA对组合的Cas9切口酶,以产生具有限定的 突出端的DNA双链断裂。当使用两对指导RNA时,可能的是切除间插DNA 片段。如果外源DNA段被这两对指导RNA切割以产生可与基因组DNA相容 的突出端,则可以将该外源DNA片段连接进该基因组DNA中,以替换被切 除的片段。例如,可以将其用于除去亨廷顿蛋白(huntintin)(HTT)基因中 的三核苷酸重复扩展,以治疗亨廷顿病。

[1416] 通过提供了带有较少数目的CAG重复的外源DNA,则可能能够产生带 有相同突出端的DNA片段并可以被连接进HTT基因组座位中并替换被切除 的片段。

[1417]

[1418] 将外源DNA片段连接进基因组中无需同源重组机器并且因此可以在有丝 分裂后细胞(如神经元)中使用此方法。

[1419] 实例27:CRISPR系统的递送

[1420] 可以将Cas9及其嵌合的指导RNA或tracrRNA和crRNA的组合作为 DNA或RNA而递送。递送均作为RNA(正常的或包含碱基或骨架修饰)分 子的Cas9和指导RNA可以用来减少Cas9蛋白在细胞中坚持的时间量。这可 以减少靶细胞中的脱靶切割活性的水平。由于递送作为mRNA的Cas9花费时 间翻译成蛋白质,所以可能有利的是在递送Cas9 mRNA后若干小时

递送指导 RNA,以使可供与Cas9蛋白相互作用的指导RNA的水平最大化。

[1421] 在指导RNA量有限的情况下,可以令人希望的是引入作为mRNA的 Cas9和呈DNA表达盒形式的指导RNA,用驱动指导RNA的表达的启动子。以此方式,可获得的指导RNA的量将经由转录而放大。

[1422] 可以引入多种递送系统,以向宿主细胞中引入Cas9(DNA或RNA)和 指导RNA(DNA或RNA)。这些递送系统包括使用脂质体、病毒载体、电 穿孔、纳米粒子、纳米线(沙莱克(Shalek)等人,《纳米快报》(Nano Letters),2012)、外来体。可以使用分子特洛伊木马脂质体(巴德里智 (Pardridge)等人,《冷泉港草案》(Cold Spring Harb Protoc);2010;doi:10.1101/pdb.prot5407)地递送Cas9和指导RNA跨过血脑屏障。

[1423] 实例28:三核苷酸重复障碍的治疗策略

[1424] 如先前在应用中所提及的,CRISPR复合物的靶多核苷酸可以包括多个疾 病相关基因和多核苷酸并且这些疾病相关基因中的一些可以属于一套称为三 核苷酸重复障碍的遗传障碍(还称为三核苷酸重复扩展障碍、三联体重复扩 展障碍或密码子反复障碍)。

[1425] 这些疾病由以下突变引起,在这些突变中,某些基因的三核苷酸重复超 过了通常在基因中可以变化的正常的稳定的阈值。更多的重复扩展障碍的发 现已经允许基于基础的相似特征将这些障碍分类为多个类别。由特定基因的 蛋白质编码部分中的CAG重复扩展引起的亨廷顿病(HD)和脊髓小脑共济 失调被包括在类别I中。具有倾向于使得它们表型相异的扩展并包括在量值上 通常较小并且还发现于基因的外显子中的扩展的疾病或障碍被包括在类别II 中。类别III包括由比类别I或II大得多的并且通常位于蛋白质编码区外的重 复扩展表征的障碍或疾病。类别III疾病或障碍的实例包括但不限于脆性X综 合征、强直性肌营养不良、两种脊髓小脑共济失调、青少年肌阵挛性癫痫以 及弗里德赖希共济失调。

[1426] 还可以采用类似的治疗策略(像下文针对弗里德赖希共济失调提及的策 略)解决其他三核苷酸重复或扩展障碍。例如,可以使用几乎相同的策略治 疗的另一种三联体重复疾病是强直性肌营养不良1(DM1),其中在3'UTR 中存在扩展的CTG基序。在弗里德赖希共济失调中,该疾病由线粒体型共济 失调蛋白(FXN)的第一个内含子中的GAA三核苷酸的扩展引起。一种使用 CRISPR的治疗策略是从第一个内含子上切除GAA重复。认为扩展的GAA 重复影响DNA结构并且导致募集关闭线粒体型共济失调蛋白(frataxin)基因 的异染色质的形成(图32A)。

[1427] 下文列出了优于其他治疗策略的竞争优势:

[1428] siRNA敲低在此情况下是不适用的,因为疾病归因于线粒体型共济失调 蛋白的减少的表达。病毒基因疗法当前处于探索之中。使用HSV-1基的载体 在动物模型中递送线粒体型共济失调蛋白基因并且已经显示出治疗效果。然 而,基于病毒的线粒体型共济失调蛋白递送的长期效力受若干问题的困扰:首先,难于将线粒体型共济失调蛋白的表达调节为与健康个体中的天然水平 相匹配,并且其次,线粒体型共济失调蛋白长期过量表达导致细胞死亡。

[1429] 可以使用核酸酶切除GAA重复,以恢复健康基因型,但是锌指核酸酶和 TALEN策略需要递送两对高效核酸酶,这对于递送以及核酸酶工程化两者而 言都是困难的(通过ZFN或TALEN有效切除基因组DNA难于实现)。

[1430] 与以上策略相比之下,CRISPR-Cas系统具有清晰的优势。Cas9酶更有效 并且更具多元性,这意味着可以同时设置一个或多个靶标。到目前为止,在 人类细胞中Cas9对基因组DNA的有效切除>30%并且可以高达30%,并且 在未来可以得到改进。此外,相对于某些三核苷酸重复障碍(像亨廷顿病 (HD)),如果在两个等位基因之间存在差异,则可以解决编码区中的三核 苷酸重复。确切地,如果HD患者对于突变的HTT而言是杂合的并且在野生 型与突变的HTT等位基因之间存在核苷酸差异(如SNP),则可以使用Cas9 特异性靶向突变的HTT等位基因。ZFN或TALEN将不具有基于单碱基差异 而区分两个等位基因的能力。

[1431] 在采用使用CRISPR-Cas 9酶解决弗里德赖希共济失调的策略中,申请人 设计了若干侧翼于GAA扩展的指导RNA靶向位点并且选择了最有效且特异 的位点(图32B)。

[1432] 申请人递送了靶向FXN的内含子1的指导RNA连同Cas9的组合,以介 导GAA扩展区域的切除。可以使用AAV9介导将Cas9有效递送进脊髓中。

[1433] 如果将邻近GAA扩展的Alu元件认为是重要的,则对可以被靶向的位点 的数目可以存在约束,但是申请人可以采用避免将其破坏的策略。

[1434] 替代性策略:

[1435] 并非使用Cas9修饰基因组,申请人还可以使用基于Cas9(核酸酶活性 缺陷的)的DNA结合结构域直接激活FXN基因,以将转录激活结构域靶向 FXN基因。申请人可能不得不解决Cas9介导的人工转录激活的稳健性,以保 证当与其他方法相比时它足够稳健(特朗布莱(Tremblay)等人,转录激活 因子样效应蛋白诱导线粒体型共济失调蛋白的表达(Transcription Activator- Like Effector Proteins Induce the Expression of the Frataxin Gene);《人类基因 疗法》(Human Gene Therapy),2012年8月,23(8):883-890)。

[1436] 实例29:用于使用Cas9切口酶使脱靶切割最小化的策略

[1437] 如先前在本申请中所提及的,可以经由一个或多个以下突变将Cas9突 变,以介导单链切割:D10A、E762A和H840A。

[1438] 为了经由NHEJ介导基因基因敲除,申请人使用了Cas9的切口酶版本连 同两个指导RNA。由每个单独的指导RNA产生的脱靶切口可以无需突变而 首先被修复,仅当靶位点彼此相邻时才发生双链断裂(这可以经由NHEJ而 导致突变)。由于通过双切口引入的双链断裂不是钝的,末端加工酶(如 TREX1)的共表达将增加NHEJ活性的水平。

[1439] 处于表格形式的以下靶标列表是针对涉及以下疾病的基因:

[1440] 拉福拉病-靶标GSY1或PPP1R3C(PTG)以减少神经元中的糖原。

[1441] 高胆固醇血症-靶标PCSK9

[1442] 在间隔子(0至3bp)中具有不同数目的核苷酸的靶序列成对列出(L和 R)。每个间隔子都还可以独自地与野生型Cas9一起使用,以在靶座位处引 入双链断裂。

[1443]

GYS1（人类）	GGCC-L	ACCCTTGTTAGCCACCTCCC
	GGCC-R	GAACGCAGTGCTCTTCGAAG
	GGNCC-L	CTCACGCCCTGCTCCGTGTA
	GGNCC-R	GGCGACAACTACTTCCTGGT
	GGNNCC-L	CTCACGCCCTGCTCCGTGTA
	GGNNCC-R	GGGCGACAACTACTTCCTGG
	GGNNNCC-L	CCTCTTCAGGGCCGGGGTGG
	GGNNNCC-R	GAGGACCCAGGTGGAACTGC
PCSK9（人类）	GGCC-L	TCAGCTCCAGGCGGTCCTGG
	GGCC-R	AGCAGCAGCAGCAGTGGCAG
	GGNCC-L	TGGGCACCGTCAGCTCCAGG
	GGNCC-R	CAGCAGTGGCAGCGGCCACC
	GGNNCC-L	ACCTCTCCCCTGGCCCTCAT
	GGNNCC-R	CCAGGACCGCCTGGAGCTGA
	GGNNNCC-L	CCGTCAGCTCCAGGCGGTCC
	GGNNNCC-R	AGCAGCAGCAGCAGTGGCAG
PPP1R3C（PTG）（人类）	GGCC-L	ATGTGCCAAGCAAAGCCTCA
	GGCC-R	TTCGGTCATGCCCGTGGATG
	GGNCC-L	GTCGTTGAAATTCATCGTAC
	GGNCC-R	ACCACCTGTGAAGAGTTTCC
	GGNNCC-L	CGTCGTTGAAATTCATCGTA
	GGNNCC-R	ACCACCTGTGAAGAGTTTCC
Gys1（小鼠）	GGCC-L	GAACGCAGTGCTTTTCGAGG
	GGCC-R	ACCCTTGTTGGCCACCTCCC
	GGNCC-L	GGTGACAACTACTATCTGGT
	GGNCC-R	CTCACACCCTGCTCCGTGTA
	GGNNCC-L	GGGTGACAACTACTATCTGG
	GGNNCC-R	CTCACACCCTGCTCCGTGTA
	GGNNNCC-L	CGAGAACGCAGTGCTTTTCG
	GGNNNCC-R	ACCCTTGTTGGCCACCTCCC
PPP1R3C（PTG）（小鼠）	GGCC-L	ATGAGCCAAGCAAATCCTCA
	GGCC-R	TTCCGTCATGCCCGTGGACA
	GGNCC-L	CTTCGTTGAAAACCATTGTA
	GGNCC-R	CCACCTCTGAAGAGTTTCCT

[1444]

	GGNNCC-L	CTTCGTTGAAAACCATTGTA
	GGNNCC-R	ACCACCTCTGAAGAGTTTCC
	GGNNNCC-L	CTTCCACTCACTCTGCGATT
	GGNNNCC-R	ACCATGTCTCAGTGTCAAGC
PCSK9（小鼠）	GGCC-L	GGCGGCAACAGCGGCAACAG
	GGCC-R	ACTGCTCTGCGTGGCTGCGG
	GGNNCC-L	CCGCAGCCACGCAGAGCAGT
	GGNNCC-R	GCACCTCTCCTCGCCCCGAT

[1445] 用于改善指导RNA的稳定性并增加特异性的替代性策略

[1446] 1.可以经由硫酯键而非磷酸酯键(像在天然RNA中)连接指导RNA 的5’中的核苷酸。硫酯键可以防止指导RNA被内源RNA降解机械熟化。

[1447] 2.指导RNA的指导序列(5’20bp)中的核苷酸可以使用桥接的核酸 (BNA)作为碱

基,以改进结合特异性。

[1448] 实例30:用于快速、多元基因组编辑的CRISPR-Cas

[1449] 本发明的方面涉及以下方案和方法,通过这些方案和方法,可以在靶标 设计后3-4天内测试基因修饰的效率和特异性,并且可以在2-3周内得到修饰 的克隆细胞系。

[1450] 可编程的核酸酶是用于以高精度介导基因组改变的强大技术。可以使用 来自微生物CRISPR获得性免疫系统的RNA-指导的Cas9核酸酶通过简单地 在其指导RNA中指定一个20-nt靶向序列而在真核细胞中促进有效的基因组 编辑。申请人描述了一套用于应用Cas9以在哺乳动物细胞中促进有效的基因 组编辑并产生用于下游功能研究的细胞系的方案。以靶标设计开始,可以在 3-4天内实现有效且特异的基因修饰,并且可以在2-3内得到修饰的克隆细胞 系。

[1451] 工程化生物系统和生物的能力跨基础科学、医药和生物技术拥有巨大的 应用潜力。促进内源基因组座位的精确编辑的可编程的序列特异的内切核酸 酶现在允许在广泛的物种中系统地询问遗传元件和致病性遗传变异,包括先 前用基因方法不易处理的那些。近年来已经出现了多种基因组编辑技术,包 括锌指核酸酶(ZFN)、转录激活因子样效应核酸酶(TALEN)和RNA-指 导的CRISPR-Cas核酸酶系统。前两种技术使用将内切核酸酶催化结构域束缚 为模块化DNA-结合蛋白用于在特定基因组座位处诱导靶向的DNA双链断裂(DSB)的共同策略。相比之下,Cas9是一种通过与靶DNA进行沃森-克里 克碱基配对而由小RNA指导的核酸酶,呈现了一种易于设计、有效且很好地 适于多种细胞类型和生物的高通量和多元基因编辑的系统。在此,申请人描 述了一套用于应用最近研发的Cas9核酸酶在哺乳动物细胞中促进有效的基因 组编辑并产生用于下游功能研究的细胞系的方案。

[1452] 像ZFN和TALEN一样,Cas9通过刺激靶基因组座位处的DSB而促进基 因组编辑。当被Cas9切割后,靶座位经历两种主要途径之一进行DNA损伤 修复,即易错非同源末端连接(NHEJ)或高保真同源定向修复(HDR)途 径。两种途径都可以用于实现希望的编辑结果。

[1453] NHEJ:在不存在修复模板的情况下,NHEJ过程重新连接DSB,这可以 留下一个处于插入缺失突变形式的瘢痕。可以利用此过程实现基因敲除,因 为发生在编码外显子内的插入缺失可以导致移码突变并形成提前终止密码 子。还可以采用多个DSB在基因组中介导较大的缺失。

[1454] HDR:同源定向修复是NHEJ的一种主要的替换性DNA修复途径。尽管 HDR典型地以低于NHEJ的频率发生,但是可以利用它在存在外源引入的修 复模板的情况下在靶座位处产生精确的限定的修饰。该修复模板可以处于类 似于具有侧翼于插入序列的同源臂的常规DNA靶向构建体而设计的双链 DNA的形式,或单链DNA寡核苷酸(ssODN)的形式。后者提供了一种用 于在基因组中进行小编辑的有效且简单的方法,如引入用于探测致病性遗传 变异的单核苷酸突变。不同于NHEJ,HDR通常仅在分裂细胞中有活性并且 其效率取决于细胞类型和状态而变化。

[1455] CRISPR概述:相比之下,CRISPR-Cas系统最低是一个由Cas9核酸酶和 短指导RNA组成的双组分系统。将Cas9重新靶向不同座位或同时编辑多个 基因简单地需要克隆一个不同的20-bp寡核苷酸。尽管Cas9核酸酶的特异性 尚未彻底阐明,CRISPR-Cas系统的简单的沃森-克里克碱基配对似乎比ZFN 或TALEN结构域的碱基配对更可预测。

[1456] II型CRISPR-Cas(规律间隔成簇短回文重复)是一种使用Cas9切割外 源遗传元件

的细菌获得性免疫系统。Cas9由一对非编码RNA指导,即一个可 变的crRNA和一个必需的辅助tracrRNA。包含一个20-nt指导序列的crRNA 经由沃森-克里克碱基配对通过定位靶DNA而确定特异性。在天然细菌系统 中,共转录多种crRNA,以指向针对不同靶标的Cas9。在来源于化脓链球菌 的CRISPR-Cas系统中,靶DNA必须紧接在一个5'-NGG/NRG原型间隔子邻近基序(PAM)之前,该基序可以针对其他CRISPR系统而变化。

[1457] CRISPR-Cas通过异源表达人类密码子优化的Cas9和必需的RNA组分而 在哺乳动物细胞中重构。此外,可以将crRNA和tracrRNA融合以产生嵌合的 合成的指导RNA(sgRNA)。Cas9因此可以通过改变sgRNA内的20-nt指导 序列而重新指向任何感兴趣的靶标。

[1458] 鉴于其易于实施和多种能力,Cas9已经被用于经由NHEJ和HDR两者而 产生携带特异的突变的工程化真核细胞。另外,直接将sgRNA和编码Cas9 的mRNA注射进胚胎中已经使得可以快速产生具有多个修饰的等位基因的转 基因小鼠;这些结果拥有编辑以另外的方式可用基因方法易处理的生物的前 景。

[1459] 在其催化结构域之一中携带破坏的突变的Cas9已经被工程化为切口而非 切割DNA,从而允许单链断裂和优先通过HDR进行修复,从而潜在地减少 来自脱靶DSB的不想要的插入缺失。另外,具有两个突变的DNA切割催化 残基的Cas9突变体已经被适配成使得可以在大肠杆菌中进行转录调节,从而 展示了功能化Cas9用于相异应用的潜力。本发明的某些方面涉及用于多元编 辑人类细胞的Cas9的构建和应用。

[1460] 申请人已经提供了人类密码子优化的核定位序列侧翼的Cas9,以促进真 核基因编辑。申请人描述了设计20-nt指导序列的考虑因素、用于快速构建 sgRNA并对其进行功能验证的方案,并且最后使用Cas9核酸酶在人类胚肾 (HEK-293FT)和人类干细胞(HUES9)系中介导基于NHEJ和HDR的两种 基因组修饰。此方案可以被同样地应用于其他细胞类型和生物。

[1461] sgRNA的靶标选择:在用于基因靶向的20-nt指导序列的选择中存在两 个主要考虑因素:1)靶序列应该在化脓链球菌Cas9的5'-NGG PAM之前, 和2)指导序列应该被选择为使脱靶活性最小化。申请人提供了一种采用感兴 趣的输入序列并鉴定适合的靶位点的在线Cas9靶向设计工具。为了用实验方 法评估每个sgRNA的脱靶修饰,申请人还为每个预期靶标提供了计算预测的 脱靶位点,这些位点是根据申请人对碱基配对错配的身份、位置和分布的作 用的定量特异性分析排列的。

[1462] 关于计算预测的脱靶位点的详细信息如下:

[1463] 脱靶切割活性的考虑因素:类似于其他核酸酶,Cas9可以按减少的频率 切割基因组中的脱靶DNA靶标。给定的指导序列展现出脱靶活性的程度取决 于以下因素的组合,包括酶浓度、利用的特定指导序列的热力学以及类似序 列在靶基因组中的丰度。对于Cas9的常规应用,重要的是考虑使脱靶切割水 平最小化并且还能够检测脱靶切割的存在的方式。

[1464] 最小化脱靶活性:对于在细胞系中应用,申请人推荐了以下两个步骤来 减少脱靶基因组修饰的水平。首先,使用我们的在线CRISPR靶标选择工具 可以计算评估给定的指导序列具有脱靶位点的可能性。通过在基因组中彻底 检索作为与指导序列类似的序列的脱靶序列而进行这些分析。sgRNA与其靶 DNA之间的错配碱基的作用的综合性实验研究揭示到错配耐受是1)位置依 赖性的-指导序列的3'末端的8-14bp比5'碱基更不耐受错配,2)数量依赖性 的-通常超过3个错配是不被容忍的,3)指导序列依赖性的-一些指导序列比 其

他指导序列更不耐受错配,以及4)浓度依赖性的-脱靶切割对转染DNA的 量高度敏感。申请人的靶位点分析网络工具(在网站enome- engineering.org/tools可获得)整合了这些标准,以为靶基因组中的可能的脱靶 位点提供预测。其次,申请人推荐滴定Cas9和sgRNA表达质粒的量,以使 脱靶活性最小化。

[1465] 脱靶活性的检测:使用申请人的CRISPR靶向网络工具可以产生最可能 的脱靶位点以及进行那些位点的SURVEYOR或测序分析的引物的列表。对 于使用Cas9产生的同基因克隆,申请人强烈推荐对这些候选脱靶位点进行测 序,以检查任何不希望的突变。值得注意的是可能在未被包括在预测候选物 列表中的位点中存在脱靶修饰并且应该进行全基因组测序以完全验证不存在 脱靶位点。此外,在于同一基因组内诱导若干DSB的多元测定中,可能存在 较低的转位事件率并且可以使用多种技术(如深度测序)对其进行评价。

[1466] 在线工具为1)制备sgRNA构建体、2)测定靶标修饰效率和3)评估潜 在的脱靶位点处的切割所需的所有寡核苷酸和引物提供了序列。值得注意的 是因为用于表达sgRNA的U6 RNA聚合酶III启动子偏好鸟嘌呤(G)核苷酸 作为其转录本的第一个碱基,所以在20-nt指导序列未以G开始的情况下,在 sgRNA的5'附加一个额外的G。

[1467] sgRNA构建和递送途径:取决于希望的应用,可以将sgRNA作为1)包 含表达盒的PCR扩增子或2)sgRNA表达质粒而递送。基于PCR的sgRNA 递送将默认sgRNA序列附加到用于扩增U6启动子模板的反向PCR引物上。 可以将生成的扩增子与包含Cas9的质粒(PX165)共转染。此方法最适于快 速筛选多个候选sgRNA,因为可以在获得sgRNA-编码引物后仅仅几个小时 进行用于功能测试的细胞转染。因为此简单方法省却了基于质粒的克隆和序 列验证的需要,所以它特别适于测试或共转染大量用于产生较大的敲除文库 或其他规模敏感的应用的sgRNA。注意的是,与基于质粒的sgRNA递送所需 的～20-bp寡核苷酸相比,sgRNA-编码引物超过100-bp。

[1468] sgRNA的表达质粒的构建也是简单且快速的,涉及使用一对部分互补的 寡核苷的单个克隆步骤。在将寡核苷酸对退火后,可以将生成的指导序列插 入带有Cas9和具有sgRNA序列的剩余部分的不变支架两者的质粒(PX330) 中。还可以将转染质粒修饰为使得可以产生用于体内递送的病毒。

[1469] 除基于PCR和质粒的递送方法之外,还可以将Cas9和sgRNA两者作为 RNA而引入进细胞中。

[1470] 修复模板的设计:传统地,靶向的DNA修饰具有包含侧翼于改变位点的 同源臂的基于质粒的供体修复模板的所需用途。每侧的同源臂的长度可以变 化,但是典型地长于500bp。可以使用此方法产生较大的修饰,包括插入报 告基因(如荧光蛋白或抗生素抗性标记)。靶向质粒的设计和构建已经在别 处进行了描述。

[1471] 最近,已经使用单链DNA寡核苷酸(ssODN)代替靶向质粒,用于在限 定的座位内进行短修饰而无需克隆。为了实现较高的HDR效率,ssODN在每 侧包含至少40bp的与靶区域同源的侧翼序列,并且可以相对于靶座位以有义 或反义方向定向。

[1472] 功能测试

[1473] SURVEYOR核酸酶测定:申请人通过SURVEYOR核酸酶测定或PCR 扩增子测序检测了插入缺失突变。申请人的在线CRISPR靶标设计工具为两 个途径提供了推荐的引物。然而,也可以手动地设计SURVEYOR或测序引 物,以从基因组DNA扩增感兴趣的区域并使用NCBI引

物-BLAST避免非特 异的扩增子。SURVEYOR引物应该被设计为在Cas9靶标的任一侧扩增300- 400bp(对于总共600-800bp的扩增子而言),以允许通过凝胶电泳而清晰地 可视化切割条带。为了防止形成过量的引物二聚体,SURVEYOR引物应该被 设计为典型地长度为25-nt之下并且解链温度为～60℃。申请人推荐在 SURVEYOR核酸酶消化过程期间针对特异的PCR扩增子以及不存在非特异 的切割的情况测试每对候选引物。

[1474] 质粒-或ssODN-介导的HDR:可以经由修饰的区域的PCR-扩增和测序 检测HDR。为此目的,PCR引物应该在由同源臂跨越的区域外退火,以避免 错误地检测到残余的修复模板(HDR Fwd和Rev,图30)。对于ssODN介 导的HDR,可以使用SURVEYOR PCR引物。

[1475] 通过测序检测插入缺失或HDR:申请人通过桑格或下一代深度测序 (NGS)检测了靶向的基因组修饰。对于前者,可以使用SURVEYOR或 HDR引物扩增来自修饰的区域的基因组DNA。应该将扩增子亚克隆进供转化 的质粒(如pUC19)中;可以对单独的菌落测序,以揭示克隆基因型。

[1476] 申请人为较短的扩增子设计了下一代测序(NGS)引物,这些扩增子典 型地在100-200bp尺寸范围内。对于检测NHEJ突变,重要的是设计在引发 区和Cas9靶位点之间具有至少10-20bp的引物,以允许检测较长的插入缺 失。申请人为两步PCR方法提供了指南,以为多元深度测序附加条形码适配 子。归因于Illumina平台的通常较低水平的假阳性插入缺失,申请人推荐了 该平台。然后可以通过读取比对程序(如ClustalW、Geneious或简单的序列 分析脚本)进行脱靶分析(先前所述)。

[1477] 材料与试剂

[1478] sgRNA制备:

[1479] 超纯的无DNaseRNase蒸馏水(生命技术公司,目录号10977-023)

[1480] Herculase II融合聚合酶(安捷伦技术公司,目录号600679)

[1481] 关键。标准Taq聚合酶,缺少3'-5'外切核酸酶校正活性,具有较低的保 真性并且可以导致扩增错误。Herculase II是一种以最小优化而产生高产率的 PCR产物的高保真聚合酶(保真性与Pfu相当)。可以替代其他高保真聚合 酶。

[1482] Herculase II反应缓冲液(5x;安捷伦技术公司,包括聚合酶)

[1483] dNTP混合溶液(各自25mM;Enzymatics公司,目录号N205L)

[1484] MgCl2(25mM;赛默飞世尔科技公司,目录号R0971)

[1485] QIAquick凝胶提取试剂盒(凯杰公司,目录号28704)

[1486] QIAprep spin miniprep试剂盒(凯杰公司,目录号27106)

[1487] UltraPure TBE缓冲液(10X;生命技术公司,目录号15581-028)

[1488] SeaKem LE琼脂糖(龙沙公司,目录号50004)

[1489] SYBR Safe DNA染色剂(10,000x;生命技术公司,目录号S33102)

[1490] 1-kb Plus DNA梯(生命技术公司,目录号10787-018)

[1491] TrackIt CyanOrange上样缓冲液(生命技术公司,目录号10482-028)

[1492] 快速消化BbsI(FastDigest BpiI)(费尔芒斯/赛默飞世尔科技公司,目 录号FD1014)

[1493] Fermentas Tango缓冲液(费尔芒斯/赛默飞世尔科技公司,目录号BY5)

[1494] DL-二硫苏糖醇(DTT;费尔芒斯/赛默飞世尔科技公司,目录号 R0862)

[1495] T7 DNA连接酶(Enzymatic公司,目录号L602L)

[1496] 关键的:不替代更常用的T4连接酶。T7连接酶在粘性末端具有比平头 末端高1,000倍的活性并且具有比可商购的浓缩T4连接酶高的整体活性。

[1497] T7 2X Rapid Ligation缓冲液(包括T7 DNA连接酶,Enzymatics公司, 目录号L602L)

[1498] T4多核苷酸激酶(新英格兰生物实验室,目录号M0201S)

[1499] T4 DNA连接酶反应缓冲液(10X;新英格兰生物实验室,目录号 B0202S)

[1500] 腺苷5'-三磷酸(10mM;新英格兰生物实验室,目录号P0756S)

[1501] PlasmidSafe ATP-依赖性DNase(Epicentre,目录号E3101K)

[1502] One Shot Stbl3化学感受态大肠杆菌(E.coli)(生命技术公司,目录号 C7373-03)

[1503] SOC培养基(新英格兰生物实验室,目录号B9020S)

[1504] LB培养基(西格玛公司,目录号L3022)

[1505] LB琼脂培养基(西格玛公司,目录号L2897)

[1506] 氨苄西林,无菌过滤的(100mg ml-1;西格玛公司,目录号A5354)

[1507] 哺乳动物细胞培养:

[1508] HEK293FT细胞(生命技术公司,目录号R700-07)

[1509] 杜尔贝科最低伊格尔培养基(Dulbecco's minimum Eagle's medium) (DMEM,1X,高葡萄糖;生命技术公司,目录号10313-039)

[1510] 杜尔贝科最低伊格尔培养基(DMEM,1X,高葡萄糖,无酚红;生命技 术公司,目录号31053-028)

[1511] 杜尔贝科磷酸盐缓冲盐水(DPBS,1X;生命技术公司,目录号14190- 250)

[1512] 胎牛血清,合格的且热灭活的(生命技术公司,目录号10438-034)

[1513] Opti-MEM I减少的血清培养基(FBS;生命技术公司,目录号11058- 021)

[1514] 青霉素-链霉素(100x;生命技术公司,目录号15140-163)

[1515] TrypLE™Express(1X,无酚红;生命技术公司,目录号12604-013)

[1516] Lipofectamine 2000转染试剂(生命技术公司,目录号11668027)

[1517] Amaxa SF Cell Line 4D-Nucleofector®X Kit S(32 RCT;龙沙公司,目录 号V4XC-2032)

[1518] HUES 9细胞系(哈佛干细胞科学(HARVARD STEM CELL SCIENCE))

[1519] Geltrex无LDEV的减少的生长因子基膜基质(生命技术公司,目录号 A1413201)

[1520] mTeSR1培养基(干细胞技术公司(Stemcell Technologies),目录号 05850)

[1521] Accutase细胞脱离溶液(干细胞技术公司,目录号07920)

[1522] ROCK抑制剂(Y-27632;密理博公司(Millipore),目录号SCM075) Amaxa P3 Primary Cell 4D-Nucleofector®X Kit S(32 RCT;龙沙公司,目录 号V4XP-3032)

[1523] 基因分型分析:

[1524] QuickExtract DNA提取溶液(Epicentre,目录号QE09050)

[1525] 检验员、RFLP分析或测序的PCR引物(参见引物表)

[1526] Herculase II融合聚合酶(安捷伦技术公司,目录号600679)

[1527] 关键。因为Surveyor测定对单碱基错配敏感,所以特别重要的是使用高 保真聚合酶。可以替代其他高保真聚合酶。

[1528] Herculase II反应缓冲液(5x;安捷伦技术公司,包括聚合酶)

[1529] dNTP混合溶液(各自25mM;Enzymatics公司,目录号N205L)

[1530] QIAquick凝胶提取试剂盒(凯杰公司,目录号28704)

[1531] Taq缓冲液(10x;金斯瑞公司(Genscript),目录号B0005)

[1532] 用于标准凝胶电泳的SURVEYOR突变检测试剂盒(转基因组学公司, 目录号706025)

[1533] UltraPure TBE缓冲液(10x;生命技术公司,目录号15581-028)

[1534] SeaKem LE琼脂糖(龙沙公司,目录号50004)

[1535] 4%-20%TBE Gels 1.0mm,15孔(生命技术公司,目录号 EC62255BOX)

[1536] Novex®Hi-Density TBE样品缓冲液(5X;生命技术公司,目录号LC6678)

[1537] SYBR Gold核酸凝胶染色剂(10,000X;生命技术公司,目录号S- 11494)

[1538] 1-kb Plus DNA梯(生命技术公司,目录号10787-018)

[1539] TrackIt CyanOrange上样缓冲液(生命技术公司,目录号10482-028)

[1540] 快速消化HindIII(FastDigest HindIII)(费尔芒斯/赛默飞世尔科技公 司,目录号FD0504)

[1541] 设备

[1542] 过滤的无菌移液管(康宁公司)

[1543] 标准1.5ml微量离心管(艾本德公司(Eppendorf),目录号0030 125.150)

[1544] Axygen 96孔PCR平板(VWR,目录号PCR-96M2-HSC)

[1545] Axygen 8-Strip PCR管(飞世尔科技公司(Fischer Scientific),目录号 14-222-250)

[1546] Falcon管,聚丙烯,15ml(BD福尔肯公司,目录号352097)

[1547] Falcon管,聚丙烯,50ml(BD福尔肯公司,目录号352070)

[1548] 带有细胞滤网帽的圆底管,5ml(BD福尔肯公司,目录号352235)

[1549] 皮氏培养皿(60mm×15mm;BD生物科学公司(BD Biosciences),目 录号351007)

[1550] 组织培养平板(24孔;BD福尔肯公司,目录号353047)

[1551] 组织培养平板(96孔,平底;BD福尔肯公司,目录号353075)

[1552] 组织培养皿(100mm;BD福尔肯公司,353003)

[1553] 带有可编程的温度步进功能的96孔热循环仪(应用生物系统公司Veriti(Applied Biosystems Veriti),目录号4375786)。

[1554] 台式微量离心机5424,5804(艾本德公司)

[1555] 凝胶电泳系统(PowerPac基础电源,伯乐公司(Bio-rad),目录号164- 5050,和Sub-Cell GT系统凝胶托盘,伯乐公司(Bio-rad),目录号170- 4401)

[1556] Novex XCell SureLock Mini-Cell(生命技术公司,目录号EI0001)

[1557] 数字凝胶成像系统(GelDoc EZ,伯乐公司(Bio-rad),目录号170- 8270,和蓝色样品托盘,伯乐公司(Bio-rad),目录号170-8273)

[1558] 蓝光透射仪和橙色过滤器护目镜(SafeImager 2.0;英杰公司,目录号 G6600)

[1559] 凝胶定量软件(伯乐公司(Bio-rad),ImageLab,包括GelDoc EZ,或 可在网站rsbweb.nih.gov/ij/获得的来自国立卫生研究院的开放源ImageJ)UV 分光光度计(NanoDrop 2000c,赛默飞世尔科技公司)

[1560] 试剂准备

[1561] 三硼酸EDTA(TBE)电泳溶液:将TBE缓冲液稀释于蒸馏水中成1X 工作溶液,用于铸造琼脂糖凝胶并用作供凝胶电泳用的缓冲液。可以将缓冲 液存储在室温(18℃-22℃)下持续至少1年。

[1562] • ATP,10mM将10mM ATP分为50-μl等分部分并存储在-20℃ 下,持续长达1年;避免重复的冻融循环。

[1563] • DTT,10mM在蒸馏水中制备10mM DTT溶液并以20-μl等分部分 存储在-70℃下,持续长达2年;用于每个反应,使用一个新的等分部分,因 为DTT容易被氧化。

[1564] • D10培养基用于培养HEK293FT细胞,通过给DMEM补充1X GlutaMAX和10%(vol/vol)胎牛血清而制备D10培养基。如在该方案中所 指示,此培养基还可以补充有1X青霉素-链霉素。可以提前制备D10培养基 并存储在4℃下,持续长达1个月。

[1565] • mTeSR1培养基用于培养人类胚胎干细胞,通过补充5X添加物(包 括mTeSR1基础培养基)和100ug/ml Normocin而制备mTeSR1培养基。

[1566] 程序

[1567] 靶向组分的设计以及在线工具的使用 • 定时1d

[1568] 1|输入靶基因组DNA序列。申请人提供了一种在线Cas9靶向设计工 具,该工具采用感兴趣的输入序列,鉴定并排列适合的靶位点,并计算预测 每个预期靶标的脱靶位点。可替代地,可以通过鉴定正好位于任何5'-NGG的 上游的20-bp序列而手动地选择指导序列。

[1569] 2|订购如通过该在线工具所指定的必需的寡核苷酸和引物。如果手动地 选择位点,则应该设计寡核苷酸和引物。

[1570] sgRNA表达构建体的制备

[1571] 3|为了产生sgRNA表达构建体,可以使用基于PCR或质粒的方案。

[1572] (A)经由PCR扩增 • 定时2h

[1573] (i)申请人制备了稀释的U6 PCR模板。申请人推荐使用PX330作为PCR 模板,但是可以同样地将任何含U6质粒用作PCR模板。申请人用ddH_2O将 模板稀释至浓度为10ng/ul。注意的是,如果将已经包含U6驱动的sgRNA的 质粒或盒用作模板,则需要进行凝胶提取,以保证产物仅包含预期的sgRNA 而不包含从模板带入的痕量sgRNA。

[1574] (ii)申请人制备了稀释的PCR寡核苷酸。将U6-Fwd和U6-sgRNA-Rev引 物在ddH_2O中稀释至终浓度为10uM(将10ul的100uM引物添加至90ul ddH_2O中)。

[1575] (iii)U6-sgRNA PCR反应。申请人按照需要为每个U6-sgRNA-Rev引物和 预混合液设置了以下反应:

[1576]

组分:	量 (ul)	终浓度
Herculase II PCR 缓冲液, 5X	10	1X
dNTP, 100 mM (各自 25 mM)	0.5	1 mM
U6 模板 (PX330)	1	0.2 ng/ul
U6-Fwd 引物	1	0.2 uM
U6-sgRNA-Rev 引物 (可变的)	1	0.2 uM
Herculase II 融合聚合酶	0.5	
蒸馏水	36	
总计	50	

[1577] (iv)申请人使用以下循环条件在来自步骤(iii)的反应上进行了PCR反 应:

[1578]

循环数	变性	退火	延伸
1	95°C, 2 m		
2-31	95°C, 20 s	60°C, 20 s	72°C, 20 s
32			72°C, 3 m

[1579] (v)反应完成后,申请人使产物在凝胶上跑胶,以验证成功的单条带扩 增。用1X SYBR Safe染料在1X TBE缓冲液中铸造2%(wt/vol)琼脂糖凝 胶。使5ul的PCR产物在该凝胶中以15V cm-1跑胶20-30min。成功的扩增 子应该产生单个的370-bp产物并且模板应该是不可见的。应该不必对PCR扩 增子进行凝胶提取。

[1580] (vi)申请人根据制造商的指导,使用QIAquick PCR纯化试剂盒纯化了 PCR产物。将DNA在35ul的缓冲液EB或水中洗脱。可以将纯化的PCR产 物存储在4℃或-20℃下。

[1581] (B)将sgRNA克隆进含Cas9的双顺反子表达载体中•定时3d

[1582] (i)制备sgRNA寡核苷酸插入片段。申请人将每个sgRNA设计的寡核苷 酸的顶链和底链再悬浮至终浓度为100uM。如下使寡核苷酸磷酸化和退火:

[1583]

寡核苷酸 1 (100 uM)		1 ul
寡核苷酸 2 (100 uM)		1 ul
T4 连接缓冲液, 10X		1 ul
T4 PNK	1 ul	
ddH_2O	6 ul	
总计	10 ul	

[1584] (ii)使用以下参数在热循环仪中退火:

[1585] 37℃,持续30m

[1586] 95℃,持续5m

[1587] 以5℃/m降至25℃

[1588] (iii)申请人通过将1ul的寡核苷酸添加至199ul室温ddH_2O中而以1: 200稀释了磷酸化且退火的寡核苷酸。

[1589] (iv)将sgRNA寡核苷酸克隆进PX330中。申请人为每个sgRNA设置了 Golden Gate反应。申请人还推荐设置了一种无插入物,仅PX330作为阴性对 照。

PX330（100 ng）	x ul
来自步骤(iii)的稀释的寡核苷酸双链体	2 ul
Tango缓冲液，10X	2 ul
DTT，10 mM	1 ul
ATP，10 mM	1 ul
快速消化BbsI	1 ul
T7连接酶	0.5 ul
ddH_2O	x ul
总计	20 ul

[1590]

[1591] (v)将Golden Gate反应孵育总共1h：

循环数	条件
1-6	37℃持续5m,21℃持续5m

[1592]

[1593]

[1594] (vi)申请人用PlasmidSafe外切核酸酶处理了Golden Gate反应,以消化任 何残余的线性化DNA。此步骤是任选的但被强烈推荐。

来自步骤4的Golden Gate反应	11 ul
10X PlasmidSafe缓冲液	1.5 ul
ATP，10 mM	1.5 ul
PlasmidSafe外切核酸酶	1 ul
总计	15 ul

[1595]

[1596] (vii)申请人将PlasmidSafe反应在37℃下孵育了30min,随后在70℃下 失活30min。暂停点:完成后,可以将反应冷冻并且稍后继续。环状DNA应 该至少1周是稳定的。

[1597] (viii)转化。申请人根据与细胞一起提供的方案将PlasmidSafe处理的质 粒转化进感受态大肠杆菌菌株中。申请人推荐Stbl3用于迅速转化。简言之，申请人将来自步骤(vii)的5ul的产物添加进20ul的冰冷化学感受态Stbl3细 胞中。然后将其在冰上孵育10m,在42℃下热休克30s,立即返回冰上持 续,添加100ul的SOC培养基,并且将其在包含100ug/ml氨苄西林的LB平 板上铺板,并在37℃下孵育过夜。

[1598] (ix)第2天:申请人检查了平板的菌落生长情况。典型地,在阴性对照平 板上不存在菌落(仅连接了BbsI-消化的PX330,没有退火的sgRNA寡核苷 酸),并且在PX330-sgRNA克隆平板上存在数十至数百个菌落。

[1599] (x)申请人从每个平板中挑选了2-3个菌落,以检查sgRNA的正确插入。申请人使用无菌移液管尖端将单菌落接种进3ml具有100ug/ml氨苄西林的 LB培养基的培养物中。在37℃下孵育并振荡过夜。

[1600] (xi)第3天:申请人根据制造商的说明书使用QiAprep Spin miniprep试剂 盒从过夜培养物中分离出质粒DNA。

[1601] (xii)序列验证CRISPR质粒。申请人通过使用U6-Fwd引物从U6启动子 测序证实了每个菌落的序列。任选的:使用以下引物表中列出的引物对Cas9 基因进行测序。

[1602]

引物	序列（5'至3'）	目的
U6-For	GAGGGCCTATTTCCCATGATTCC	扩增U6-sgRNA
U6-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG CTCTAAAACNNNNNNNNNNNNNNNNNNNCCGGTGTTTCGTCCTTTCCACAAG	扩增U6-sgRNA；N是靶标的反向互补序列
sgRNA-上	CACCGNNNNNNNNNNNNNNNNNN	将sgRNA克隆进PX330中
sgRNA-下	AAACNNNNNNNNNNNNNNNNNNC	将sgRNA克隆进PX330中
U6-*EMX1*-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG CTCTAAAACCCCTAGTCATTGGAGGTGACCGGTGTTTCGTCCTTTCCACAAG	扩增U6-EMX1 sgRNA
EMX1-上	CACCGTCACCTCCAATGACTAGGG	将EMX1 sgRNA克隆进PX330中
EMX1-下	AAACCCCTAGTCATTGGAGGTGAC	将EMX1 sgRNA克隆进PX330中
ssODN-有义	CAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATC GATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTG CTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGAC	EMX1 HDR（有义；插入已加下划线）

[1603]

	TCTGGCCACTCCCT	
ssODN-反义	AGGGAGTGGCCAGAGTCCAGCTTGGGCCCACGC AGGGGCCTGGCCAGCAGCAAGCAGCACTCTGCC CTCGTG GGTTTGTGGTTGCCCACCGCTAGCAAGCTTGTC ATTGGAGGTGACATCGATGTCCTCCCCATTGGC CTGCTTCG TGGCAATGCGCCACCGGTTGATGTGATGGGAGC CCTTCTTCTTCTG	EMX1 HDR（反义：插入已加下划线）
EMX1-SURV-F	CCATCCCCTTCTGTGAATGT	EMX1 SURVEYOR 测定 PCR，测序
EMX1-SURV-R	GGAGATTGGAGACACGGAGA	EMX1 SURVEYOR 测定 PCR，测序
EMX1-HDR-F	GGCTCCCTGGGTTCAAAGTA	EMX1 RFLP 分析 PCR，测序
EMX1-HDR-R	AGAGGGGTCTGGATGTCGTAA	EMX1 RFLP 分析 PCR，测序
pUC19-F	CGCCAGGGTTTTCCCAGTCACGAC	pUC19 多克隆位点 F 引物，用于桑格测序

[1604] 申请人参考了PX330克隆载体序列的测序结果,以检查20bp指导序列 插入在U6启动子与sgRNA支架的剩余部分之间。可以在网站crispr.genome- engineering.org找到处于GenBank载体图谱格式的PX330图谱的细节和序列 (*.gb文件)。

[1605] (任选的)设计ssODN模板·定时3d提前计划

[1606] 3|设计和订购ssODN。有义或反义ssODN可以直接从供应商购买。申请 人推荐设计在任一侧上具有至少40bp和用于优化HDR效率的90bp的同源 臂。在申请人的经验中,反义寡核苷酸具有稍微更高的修饰效率。

[1607] 4|申请人将ssODN ultramer再悬浮并稀释至终浓度为10uM。不将有义 和反义ssODN合并或退火。存储在-20℃下。

[1608] 5|注意对于HDR应用,申请人推荐将sgRNA克隆进PX330质粒中。

[1609] sgRNA的功能验证:细胞培养和转染·定时3-4d

[1610] CRISPR-Cas系统已经被用于多种哺乳动物细胞系中。每个细胞系的条件 可以变化。以下方案详细说明了HEK239FT细胞的转染条件。注意对于 ssODN介导的HDR转染,使用

Amaxa SF Cell Line Nucleofector试剂盒优化 ssODN的递送。将在下节中对其进行描述。

[1611] 7|HEK293FT维持。根据制造商的推荐维持细胞。简言之,申请人在 37℃和5%CO2下将细胞培养在D10培养基(补充有10%胎牛血清的 GlutaMax DMEM)中。

[1612] 8|为了传代,申请人移除培养基并且通过将DPBS轻轻地添加到容器的 侧面而将其冲洗一次,以便移动细胞。申请人将2ml的TrypLE添加至T75 烧瓶中并在37℃下孵育5m。添加10ml的温D10培养基以失活并将其转移 至50ml Falcon管中。申请人通过轻轻地研磨而使细胞解离,并且必要时重新 接种新的烧瓶。申请人典型地使细胞以1:4或1:8的分传比每2-3d传代一 次,绝不允许细胞达到大于70%的汇合。当达到传代数15时,细胞系被重新启动。

[1613] 9|制备供转染用的细胞。在以1.3x 10^5个细胞/孔的接种密度和500ul的 接种体积进行转染之前16-24h,申请人将良好解离的细胞在24孔平板上铺板 在没有抗生素的D10培养基中。根据制造商手册按需要按比例放大或缩小。 建议不要铺比推荐的密度多的细胞,因为这样做可以降低转染效率。

[1614] 10|在转染的当天,细胞的汇合最佳为70%-90%。根据制造商的方案, 可以用Lipofectamine 2000或Amaxa SF Cell Line Nucleofector试剂盒转染细 胞。

[1615] (A)对于克隆进PX330中的sgRNA,申请人转染了500ng的序列已证实 的CRISPR质粒;如果转染超过一种质粒,以等摩尔比例混合并且总共不超 过500ng。

[1616] (B)对于通过PCR扩增的sgRNA,申请人将以下项混合:

[1617] PX165(仅有Cas9) 200ng

[1618] sgRNA扩增子(各自) 40ng

[1619] pUC19 填充总DNA至500ng

[1620] 申请人推荐在技术上一式三份地转染用于可靠地定量并包括转染对照 (例如GFP质粒)以监测转染效率。另外,可以将PX330克隆质粒和/或 sgRNA扩增子作为阴性对照而单独地转染,用于下游的功能测定。

[1621] 11|申请人将Lipofectamine复合物轻轻地添加至细胞中,因为 HEK293FT细胞可以容易地从平板上脱离并导致转染效率较低。

[1622] 12|申请人在转染后24h通过使用荧光显微镜估计对照(例如,GFP)转 染中的荧光细胞的分数而检查了细胞的效率。典型地,超过70%的细胞被转 染。

[1623] 13|申请人给培养基补充了另外500ul的温D10培养基。将D10非常缓 慢地添加至孔的侧面并且不要使用冷的培养基,因为细胞可以容易地脱离 开。

[1624] 14|在收获用于插入缺失分析之前,将细胞在转染后孵育总共48-72h。48h后插入缺失效率未显著增加。

[1625] (任选的)CRISPR质粒和ssODN或靶向质粒的共转染用于进行HR• 定时3-4d

[1626] 15|线性化靶向质粒。如果可能的话,通过在同源臂之一的附近或在任一 同源臂的远端在载体骨架中的限制酶切位点处切割一次而将靶向载体线性 化。

[1627] 16|申请人使少量的线性化质粒在未切割的质粒旁边在0.8%-1%琼脂糖凝 胶上跑胶,以检查成功的线性化。线性化质粒应该跑到超螺旋质粒的上方。

[1628] 17|申请人用QIAQuick PCR纯化试剂盒纯化了线性化质粒。

[1629] 18|制备供转染用的细胞。申请人在T75或T225烧瓶中培养 HEK293FT。在转染当天

之前，为重组的细胞计数订计划。对于Amaxa条状 比色皿格式，每个转染使用2x 10^6个细胞。

[1630] 19|制备供转染用的平板。申请人将1ml的温D10培养基添加进12孔平 板的每个孔中。将平板置于恒温箱中，以保持培养基是温的。

[1631] 20|核转染。申请人根据适于以下步骤的Amaxa SF Cell Line Nucleofector 4D试剂盒制造商的说明书转染HEK293FT细胞。

[1632] a.用于ssODN和CRISPR共转染，将以下DNA预混合在PCR管 中：

[1633] pCRISPR质粒(Cas9+sgRNA) 500ng

[1634] ssODN模板(10uM) 1ul

[1635] b.用于HDR靶向质粒和CRISPR共转染，将以下DNA预混合在 PCR管中：

[1636] CRISPR质粒(Cas9+sgRNA) 500ng

[1637] 线性化靶向质粒 500ng

[1638] 对于转染对照，参见上节。另外，申请人推荐转染单独的ssODN或靶向 质粒作为阴性对照。

[1639] 21|解离为单细胞。申请人移除培养基并用DPBS轻轻地冲洗一次，注意 不要移动细胞。将2ml的TrypLE添加至T75烧瓶中并在37℃下孵育5m。 添加10ml的温D10培养基以失活并将其在50ml Falcon管中轻轻地研磨。推 荐轻轻地研磨细胞并解离为单细胞。较大的团块将降低转染效率。申请人从 悬浮液中取了一个10ul的等分部分并将其稀释于90ul的D10培养基中用于 计数。申请人对细胞进行了计数并计算了转染所需的细胞数目和悬浮液体 积。申请人使用Amaxa Nucleocuvette条典型地每个条件转染2x 10^5个细胞，并且推荐计算比所需多20％的细胞，以调节在随后的吸移步骤中的体积损 失。将所需体积转移至新的Falcon管中。

[1640] 23|申请人使新管以200x g旋降，持续5m。

[1641] 如由Amaxa所推荐的，申请人通过将SF溶液与S1添加物混合而制备了 转染溶液。对于Amaxa条状比色皿，每个转染需要总共20ul的补充SF的溶 液。同样地，申请人推荐计算比所需多20％的体积。

[1642] 25|申请人从来自步骤23的沉淀细胞中完全移除培养基并将其轻轻地再 悬浮于适当体积(20ul/2x 10^5个细胞)的补充S1的SF溶液中。不要将细胞 在SF溶液中留延长的时间段。

[1643] 26|将20ul的再悬浮的细胞吸移进来自步骤20的每个DNA预混合液 中。轻轻地吸移以混合并转移至Nucleocuvette条室中。针对每个转染条件将 其重复。

[1644] 使用由Amaxa推荐的Nucleofector 4D程序CM-130将细胞电穿孔。

[1645] 28|申请人轻轻地并缓慢地将100ul的温D10培养基吸移进每个 Nucleocuvette条室中，并将全部体积转移至来自步骤19的预加温的平板中。 关键.细胞在此阶段是非常脆弱的并且粗鲁的吸移可以引起细胞死亡。孵育24 h。在此时，可以从阳性转染对照中的荧光细胞的分数估计转染效率。核转染 典型地产生大于70％-80％的转染效率。申请人将1ml温D10培养基缓慢地添 加至每个孔中而未移动细胞。将细胞孵育总共72h。

[1646] 人类胚胎干细胞(HUES 9)培养和转染·定时3-4d

[1647] 维持hESC(HUES9)系。申请人常规地用mTesR1培养基将HUES9细 胞系维持在无进

料器的条件中。申请人通过添加5X包括基础培养基和100 ug/ml Normocin的添加物而制备了mTeSR1培养基。申请人制备了进一步补 充有10uM Rock抑制剂的mTeSR1培养基的一个10ml的等分部分。包衣组 织培养平板。将冷的GelTrex 1:100地稀释于冷的DMEM中并包衣100mm 组织培养平板的整个表面。

[1648] 将平板在37℃下置于恒温箱中持续至少30m。在37℃下解冻一小瓶15 ml Falcon管中的细胞,添加5ml的mTeSR1培养基,并在200x g下沉淀5 m。抽出GelTrex包衣并用～1x 106个细胞接种10ml的包含Rock抑制剂的 mTeSR1培养基。转染后24h换为正常的mTeSR1培养基并且每天重新给 料。细胞传代。每天用新鲜的mTeSR1培养基为细胞重新给料并在达到70％ 汇合之前传代。抽出mTeSR1培养基并用DPBS将细胞洗涤一次。通过添加2 ml Accutase并在37℃下孵育3-5m而使细胞解离。向脱离开的细胞中添加10 ml mTeSR1培养基,转移至15ml Falcon管中并轻轻地再悬浮。在具有10uM Rock抑制剂的mTeSR1培养基中在GelTrex包衣的平板上重新铺板。铺板后 24h换为正常的mTeSR1培养基。

[1649] 转染。申请人推荐在使用Amaxa P3 Primary Cell 4-D Nucleofector试剂盒 (龙沙公司)转染之前,在解冻后将细胞培养至少1周。转染之前2h,用新 鲜的培养基为对数期生长细胞重新给料。用accutase解离为单细胞或不超过 10个细胞的小聚簇并轻轻地再悬浮。为核转染所需的细胞数目计数并以200x g旋降,持续5m。完全地移除培养基并再悬浮于推荐体积的补充S1的P3核 转染溶液中。在1X Rock抑制剂的存在下,将电穿孔的细胞轻轻地铺进包衣 的平板中。

[1650] 检查转染是否成功并且在核转染后24h开始每天用普通的mTeSR1培养 基重新给料。典型地,申请人用Amaxa核转染观察到大于70％的转染效率。收获DNA。转染后48-72h,使用accutase使细胞解离并通过添加5x体积的 mTeSR1而失活。将细胞以200x g旋降,持续5m。可以将沉淀的细胞直接 加工,用于使用QuickExtract溶液提取DNA。推荐不要不用accutase而机械 地解离细胞。推荐不要在未灭活accutase的情况下或不按以上推荐的速度使 细胞旋降;这样做可以导致细胞溶解。

[1651] 通过FACS分离克隆细胞系。定时•2-3h亲身实践的;2-3周扩展

[1652] 可以在转染后24h通过FACS或通过连续稀释进行克隆分离。

[1653] 54|制备FACS缓冲液。可以在补充有1X青霉素/链霉素的普通D10培养 基中分选无需使用彩色荧光分选的细胞。如果还需要彩色荧光分选,则用无 酚DMEM或DPBS替代正常的DMEM。补充1X青霉素/链霉素并通过.22 umSteriflip过滤器进行过滤。

[1654] 55|准备96孔平板。申请人向每个孔中添加了100ul补充有1X青霉素/ 链霉素的D10培养基并根据希望的克隆数目的需要准备了平板的数目。

[1655] 56|制备供FACS用的细胞。申请人通过完全吸除培养基并向24孔平板 的每个孔中添加100ul TrypLE而使细胞解离。孵育5m并添加400ul温D10 培养基。

[1656] 57|将再悬浮的细胞转移至15ml Falcon管中并轻轻地研磨20次。建议在 显微镜下进行检查,以保证解离单细胞。

[1657] 58|将细胞以200x g旋降,持续5分钟。

[1658] 59|申请人抽出培养基,并且将细胞再悬浮于200ul的FACS培养基中。

[1659] 60|通过35um目过滤器将细胞过滤进标记的FACS管中。申请人推荐使 用带有细胞滤网帽的BD Falcon 12x 75mm管。将细胞放在冰上直到分选。

[1660] 61|申请人将单细胞分选进从步骤55准备的96孔平板中。申请人推荐在 每个平板上的一个单个指定的孔中,分选100个细胞作为阳性对照。

[1661] 注意。可以将细胞的剩余部分保存和用于以群体水平进行基因分型,以 测量整体修饰效率。

[1662] 62|申请人将细胞返回恒温箱总并允许它们扩增2-3周。分选后5d,添 加100ul的温D10培养基。必要时,每3-5d更换100ul的培养基。

[1663] 63|分选后1周检查菌落的“克隆”外观:从中心点辐射的圆菌落。划出 空的或可能已经接种有双联体或多联体的孔。

[1664] 64|当超过60%的细胞汇合时,申请人准备了一套用于传代的复制平板。将100ul的D10培养基添加至复制平板中的每个孔中。申请人通过上下剧烈 吸移20次而直接将细胞解离。将20%的再悬浮体积铺进准备的复制平板中, 以保存克隆系。之后每2-3d更换培养基并相应地传代。

[1665] 65|使用剩余的80%细胞进行DNA分离和基因分型。

[1666] 任选的:通过稀释分离克隆细胞系。定时•2-3h亲身实践的;2-3周扩 展

[1667] 66|申请人如上所述地将细胞从24孔平板上解离。确保解离成单细胞。 可以使用细胞滤网防止细胞聚集。

[1668] 67|在每个条件下对细胞的数目计数。将每个条件连续稀释于D10培养基 中至终浓度为0.5个细胞/100ul。对于每个96孔平板,申请人推荐稀释至终 计数为12ml的D10中有60个细胞。推荐将细胞数目的精确计数用于适当的 克隆稀释。可以在中间的连续稀释阶段中对细胞重新计数以保证精确度。

[1669] 68|使用多通道移液管将100ul的稀释的细胞吸移进96孔平板的每个孔 中。

[1670] 注意。可以将细胞的剩余部分保存并用于以群体水平进行基因分型,以 测量整体修饰效率。

[1671] 69|铺板后～1周申请人检查了菌落的“克隆”外观:从中心点辐射的圆 菌落。划出可能已经接种有双联体或多联体的孔。

[1672] 70|申请人将细胞返回恒温箱总并允许它们扩增2-3周。如在上节中详细 说明的,根据需要为细胞重新给料。

[1673] 用于CRISPR切割效率的SURVEYOR测定。定时•5-6h

[1674] 在测定转染细胞的切割效率之前,申请人推荐使用以下描述的方案在阴 性(未转染的)对照样品上通过SURVEYOR核酸酶消化步骤测试每个新的 SURVEYOR引物。偶然地,甚至单条带的干净的SURVEYOR PCR产物可以 产生非特异的SURVEYOR核酸酶切割条带并潜在地干扰准确的插入缺失分 析。

[1675] 71|收获细胞的DNA。将细胞解离并以200x g旋降,持续5m。注意.根 据需要在此阶段需要复制平板,以保存转染细胞系。

[1676] 72|完全地抽出上清液。

[1677] 73|申请人根据制造商的说明书使用QuickExtract DNA提取溶液。申请人 典型地为24孔平板的每个孔使用50ul的溶液并为96孔平板的每个孔使用10 ul的溶液。

[1678] 74|申请人用ddH2O将提取的DNA标准化至终浓度为100-200ng/ul。暂 停点:可以将提取的DNA存储在-20℃下,持续若干个月。

[1679] 75|设置SURVEYOR PCR。使用由申请人的在线/计算机算法工具提供的 SURVEYOR引物将以下项预混合：

[1680]

组分:	量（ul）	终浓度
Herculase II PCR 缓冲液，5X	10	1X
dNTP，100 mM（各自 25 mM）	1	1 mM
SURVEYOR Fwd 引物（10 uM）	1	0.2 uM
SURVEYOR Rev 引物（10 uM）	1	0.2 uM
Herculase II 融合聚合酶	1	
$MgCl_2$（25 mM）	2	1 mM
蒸馏水	33	
总计	49（对于每个反应而言）	

[1681] 76|申请人为每个反应添加了100-200ng的来自步骤74的标准化的基因 组DNA模板。

[1682] 77|使用以下循环条件进行PCR反应，持续不超过30个扩增循环：

[1683]

循环数	变性	退火	延伸
1	95°C，2 min		
2-31	95°C，20 s	60°C，20 s	72°C，30 s
32			72°C，3 min

[1684] 78|申请人使2-5ul的PCR产物在1％凝胶上跑胶，以检查单条带产物。尽管将这些PCR条件设计为与大部分SURVEYOR引物对一起工作，但是一 些引物可能需要通过调节模板浓度、$MgCl_2$浓度和/或退火温度而进行另外的 优化。

[1685] 79|申请人使用QIAQuick PCR纯化试剂盒与标准化洗脱液纯化了PCR反 应至20ng/ul。暂停点：可以将纯化的PCR产物存储在-20℃下。

[1686] 80|DNA异源双链体形成。如下设置退火反应：

[1687]

Taq PCR 缓冲液，10X	2 ul
标准化的 DNA（20 ng/ul）	18 ul
总体积	20 ul

[1688] 81|使用以下条件使反应退火：

[1689]

循环数	条件

	1	95°C，10 mn
	2	95°C-85°C，-2°C/s
	3	85°C，1 min
	4	85°C-75°C，-0.3°C/s
	5	75°C，1 min
	6	75°C-65°C，-0.3°C/s
	7	65°C，1 min
[1690]	8	65°C-55°C，-0.3°C/s
	9	55°C，1 min
	10	55°C-45°C，-0.3°C/s
	11	45°C，1 min
	12	45°C-35°C，-0.3°C/s
	13	35°C，1 min
	14	35°C-25°C，-0.3°C/s
	15	25°C，1 min

[1691] 82|SURVEYOR核酸酶S消化。申请人制备了预混合液并将冰上的以下 组分添加至来自步骤81的退火的异源双链体，总终体积为25ul：

[1692]

组分	量（ul）	终浓度
$MgCl_2$ 溶液，0.15 M	2.5	15 mM
ddH_2O	0.5	
SURVEYOR 核酸酶 S	1	1X
SURVEYOR 增强子 S	1	1X
总计	5	

[1693] 83|使孔涡旋并旋降。将反应在42℃下孵育1h。

[1694] 84|任选的：可以添加2ul来自SURVEYOR试剂盒的终止溶液。暂停点. 可以将消化的产物存储在-20℃下用于稍后的分析。

[1695] 85|可视化SURVEYOR反应。可以在2％琼脂糖凝胶上使SURVEYOR 核酸酶消化产物可视化。用于更好地拆分，可以使产物在4％-20％梯度聚丙烯 酰胺TBE凝胶上跑胶。申请人用推荐的加样缓冲液加样了10ul的产物并根 据制造商的说明书在凝胶上跑胶。典型地，申请人跑胶直到溴酚蓝染料已经 迁移到凝胶的底部。同一凝胶上包括DNA梯和阴性对照。

[1696] 86|申请人用稀释于TBE中的1X SYBR Gold染料将凝胶染色。将凝胶轻 轻地摇动15m。

[1697] 87|申请人使用定量成像系统使凝胶成像而未过度暴露这些条带。阴性对 照应该仅具有一个对应于PCR产物的尺寸的条带，但是可以偶尔具有其他尺 寸的非特异的切割条带。如果它们与靶切割条带在尺寸上不同，则这些条带 将不干扰分析。由申请人的在线/计算机算法工具提供的靶切割条带尺寸的总 和应该等于PCR产物的尺寸。

[1698] 88|估计切割强度。申请人使用ImageJ或其他凝胶定量软件定量了积分 强度。

[1699] 89|对于每个泳道,申请人使用以下公式计算了切割的PCR产物的分数（$f_{切割}$）:$f_{切割}$＝(b+c)/(a+b+c),其中a是未消化的PCR产物的积分强度 并且b和c是每个切割产物的积分强度。90|可以使用以下公式基于双链体形 成的二项式概率分布估计切割效率：

[1700] 91| $\text{indel}(\%) = 100 \times (1 - \sqrt{(1 - f_{切割})})$

[1701] 用于评估CRISPR切割效率的桑格测序。定时 • 3d

[1702] 初始步骤与SURVEYOR测定的步骤71-79相同。注意:如果将适当的限 制酶切位点附加至正向和反向引物上,则可以将SURVEYOR引物用于桑格 测序。对于克隆进推荐的pUC19骨架中,可以使用EcoRI用于Fwd引物并使 用HindIII用于Rev引物。

[1703] 92|扩增子消化。如下设置消化反应：

[1704]

组分	量（ul）
快速消化（Fast Digest）缓冲液，10X	3
快速消化 EcoRI	1
快速消化 HindIII	1
标准化的 DNA（20 ng/ul）	10
ddH_2O	15
总体积	30

[1705] 93|pUC19骨架消化。如下设置消化反应：

[1706]

组分	量（ul）
快速消化缓冲液，10X	3
快速消化 EcoRI	1
快速消化 HindIII	1
FastAP 碱性磷酸酶	1
pUC19 载体（200 ng/ul）	5
ddH_2O	20
总体积	30

[1707] 94|申请人使用QIAQuick PCR纯化试剂盒纯化了消化反应。暂停点:可 以将纯化的PCR产物存储在-20℃下。

[1708] 95|申请人如下以1:3载体:插入片段比例将消化的pUC19骨架与桑格 扩增子连接：

[1709]

[1710]

组分	量（ul）
消化的 pUC19	x（50 ng）
消化的插入片段	x（1:3 载体: 插入片段摩尔比）
T7 连接酶	1
2X 快速连接缓冲液（Rapid Ligation Buffer）	10
ddH_2O	x
总体积	20

[1711] 96|转化。申请人根据与细胞一起提供的方案将PlasmidSafe处理的质粒 转化进

感受态大肠杆菌菌株中。申请人推荐Stbl3用于迅速转化。简言之，将 5ul来自步骤95的产物添加进20ul的冰冷的化学感受态Stbl3细胞中，在冰 上孵育10m，在42℃下热休克30s，立即返回冰上持续2m，添加100ul的 SOC培养基，并在包含100ug/ml氨苄西林的LB平板上铺板。将其在37℃ 下孵育过夜。

[1712] 97|第2天：申请人检查了平板的菌落生长情况。典型地，在阴性对照平 板上不存在菌落(仅连接了EcoRI-HindIII消化的pUC19，没有桑格扩增子插 入片段)，并且在pUC19-桑格扩增子克隆平板上存在数十至数百个菌落。

[1713] 98|第3天：申请人根据制造商的说明书使用QIAprep Spin miniprep试剂 盒从过夜培养物中分离出质粒DNA。

[1714] 99|桑格测序。申请人通过使用pUC19-For引物从pUC19骨架测序证实 了每个菌落的序列。申请人参照了预期的基因组DNA序列的测序结果，以检 查Cas9诱导的NHEJ突变的存在。%编辑效率＝(#修饰的克隆)/(#总克隆)。重要的是挑选合理数目的克隆(>24)，以产生准确的修饰效率。

[1715] 微小缺失的基因分型。定时·2-3d亲身实践的；2-3周扩展

[1716] 100|如上所述的，用一对靶向有待缺失的区域的sgRNA转染细胞。

[1717] 101|转染后24h，如上所述的，通过FACS或连续稀释分离克隆系。

[1718] 102|将细胞扩增2-3周。

[1719] 103|如上所述的，申请人使用10ul QuickExtract溶液从克隆系中收获了 DNA并用ddH_2O将基因组DNA标准化至终浓度为50-100ng/ul。

[1720] 104|PCR扩增修饰的区域。如下设置PCR反应：

[1721]

组分：	量 (ul)	终浓度
Herculase II PCR 缓冲液，5X	10	1X
dNTP，100 mM（各自 25 mM）	1	1 mM
Out Fwd 引物（10 uM）	1	0.2 uM
Out Rev 引物（10 uM）	1	0.2 uM
Herculase II 融合聚合酶	1	
MgCl2（25 mM）	2	1 mM
ddH_2O	32	

[1722] 总计 48（对于每个反应而言）

[1723] 注意：如果缺失尺寸超过1kb，用In-Fwd和In-Rev引物设置一套平行的 PCR反应，以筛选野生型等位基因的存在。

[1724] 105|为了筛选倒置，如下设置PCR反应：

[1725]

组分：	量（ul）	终浓度
Herculase II PCR 缓冲液，5X	10	1X
dNTP，100 mM（各自 25 mM）	1	1 mM
外 Fwd 或外-Rev 引物（10 uM）	1	0.2 uM
内 Fwd 或内-Rev 引物（10 uM）	1	0.2 uM
Herculase II 融合聚合酶	1	
$MgCl_2$（25 mM）	2	1 mM
ddH_2O	32	
总计	48（对于每个反应而言）	

[1726] 注意：引物作为外-Fwd+内Fwd或外-Rev+内-Rev而配对。

[1727] 106|申请人为每个反应添加了100-200ng的来自步骤103的标准化的基 因组DNA模板。

[1728] 107|使用以下循环条件进行PCR反应：

[1729]

循环数	变性	退火	延伸
1	95°C，2 min		
2-31	95°C，20 s	60°C，20 s	72°C，30 s
32			72°C，3 m

[1730] 108|申请人使2-5ul的PCR产物在1％-2％凝胶上跑胶，以检查产物。尽 管将这些PCR条件设计为与大部分引物一起工作，但是一些引物可能需要通 过调节模板浓度、$MgCl_2$浓度和/或退火温度而进行另外的优化。

[1731] 经由HDR针对靶向修饰进行基因分型。定时·2-3d，2-3h亲身实践的

[1732] 109|如上所述的，申请人使用QuickExtract溶液收获了DNA并用TE将 基因组DNA标准化至终浓度为100-200ng/ul。

[1733] 110|PCR扩增修饰的区域。如下设置PCR反应：

[1734]

组分：	量（ul）	终浓度
Herculase II PCR 缓冲液，5X	10	1X
dNTP，100 mM（各自 25 mM）	1	1 mM
HDR Fwd 引物（10 uM）	1	0.2 uM
HDR Rev 引物（10 uM）	1	0.2 uM
Herculase II 融合聚合酶	1	
$MgCl_2$（25 mM）	2	1 mM
ddH_2O	33	

[1735] 总计 49（对于每个反应而言）

[1736] 111|申请人为每个反应添加了100-200ng的来自步骤109的基因组DNA 模板并按以下程序跑胶。

[1737]

循环数	变性	退火	延伸
1	95°C，2 min		
2-31	95°C，20 s	60°C，20 s	72°C，*30-60 s/kb*
32			72°C，3 min

[1738] 112|申请人使5ul的PCR产物在0.8%-1%凝胶上跑胶，以检查单条带产 物。引物可能需要通过调节模板浓度、$MgCl_2$浓度和/或退火温度而进行另外 的优化。

[1739] 113|申请人使用QIAQuick PCR纯化试剂盒纯化了PCR反应。

[1740] 114|在HDR实例中，将HindIII限制酶切位点插入EMX1基因中。通过 限制酶切消化PCR扩增子而对其进行检测：

[1741]

组分	量（ul）
纯化的 PCR 扩增子（200-300 ng）	x
F.D.缓冲液，Green	1
HindIII	0.5
ddH2O	x
总计	10

[1742] i.将DNA在37℃下消化10m：

[1743] ii.申请人使10ul具有加样染料的消化的产物在4%-20%梯度聚丙烯酰胺 TBE凝胶上跑胶，直到二甲苯蓝条带已经迁移到凝胶的底部。

[1744] iii.申请人用1X SYBR Gold染料将凝胶染色同时摇动15m。

[1745] iv.如上在SURVEYOR测定章节中所述的，使切割产物成像并对其进行 定量。通过以下公式估计HDR效率：(b+c)/(a+b+c)，其中a是未消化的 HDR PCR产物的积分强度，并且b和c是HindIII-切割片段的积分强度。

[1746] 115|可替代地，可以克隆来自步骤113的纯化的PCR扩增子并使用桑格 测序或NGS进行基因分型。

[1747] 深度测序和脱靶分析 • 定时1-2d

[1748] 在线CRISPR靶标设计工具针对每个鉴定的靶位点产生了候选基因组脱 靶位点。可以通过SURVEYOR核酸酶测定、桑格测序或下一代深度测序进 行这些位点处的脱靶分析。鉴于这些位点中的许多处的修饰率较低或不可检 测的可能性，申请人推荐用Illumina Miseq平台进行深度测序，用于获得较高 的灵敏度和精确度。方案将随测序平台而变化；在此，申请人简要地描述了 一种用于附接测序适配子的融合PCR方法。

[1749] 116|设计深度测序引物。为较短的扩增子设计下一代测序(NGS)引 物，这些扩增子典型地在100-200bp尺寸范围内。可以使用NCBI引物-Blast 手动地设计引物或用在线CRISPR靶标设计工具产生(网站为genome- engineering.org/tools)。

[1750] 117|从Cas9靶向的细胞中收获基因组DNA。用ddH2O将QuickExtract 基因组DNA标准化至100-200ng/ul。

[1751] 118|初始文库制备PCR。使用来自步骤116的NGS引物准备初始文库制 备PCR

[1752]

组分:	量（ul）	终浓度
Herculase II PCR 缓冲液，5X	10	1X
dNTP，100 mM（各自 25 mM）	1	1 mM
NGS Fwd 引物（10 uM）	1	0.2 uM
NGS Rev 引物（10 uM）	1	0.2 uM
Herculase II 融合聚合酶	1	
MgCl2（25 mM）	2	1 mM
ddH2O	33	
总计	49（对于每个反应而言）	

[1753] 119|为每个反应添加100-200ng的标准化的基因组DNA模板。

[1754] 120|使用以下循环条件进行PCR反应，持续不超过20个扩增循环：

[1755]

循环数	变性	退火	延伸
1	95°C，2 min		
2-21	95°C，20 s	60°C，20 s	72°C，15 s
22			72°C，3 min

[1756] 121|使2-5ul的PCR产物在1％凝胶上跑胶，以检查单条带产物。与基 因组DNAPCR一样，NGS引物可能需要通过调节模板浓度、$MgCl_2$浓度和/或 退火温度而进行另外的优化。

[1757] 122|使用QIAQuick PCR纯化试剂盒纯化PCR反应并用洗脱液标准化至 20ng/ul。暂停点：可以将纯化的PCR产物存储在-20℃下。

[1758] 123|Nextera XT DNA样品制备试剂盒。遵循制造商的方案，用独特的条 形码为每个样品产生Miseq测序就绪的文库。

[1759] 124|分析测序数据。可以通过读取比对程序(如ClustalW、Geneious或 简单的序列分析脚本)进行脱靶分析。

[1760] 定时

[1761] 步骤1-2设计并合成sgRNA寡核苷酸和ssODN：1-5d，取决于供应商而 可变

[1762] 步骤3-5构建CRISPR质粒或PCR表达盒：2h至3d

[1763] 步骤6-53转染进细胞系中：3d(1h亲身实践时间)

[1764] 步骤54-70任选地衍生克隆系：1-3周，取决于细胞类型而可变

[1765] 步骤71-91经由SURVEYOR对NHEJ进行功能验证：5-6h

[1766] 步骤92-124经由桑格或下一代深度测序进行基因分型：2-3d(3-4h亲身 实践时间)

[1767] 解决在此的实例涉及的情况

[1768]

情况	解决方案
未扩增出 sgRNA	滴定 U6-模板浓度
SURVEYOR 或 HDR PCR 较脏或没有扩增	滴定 MgCl2；标准化并滴定模板浓度；退火温度梯度；重新设计引物
微小缺失 PCR 中的等位基因的不等扩增	设置分开的 PCR，以检测野生型和缺失等位基因；用类似尺寸的扩增子重新设计引物
在阴性对照平板上存在菌落	增加 BbsI；增加 Golden Gate 反应循环数，用南极磷酸盐处理分开地切割 PX330
没有 sgRNA 序列或序列错误	筛选另外的菌落
脂质体转染率低	检查细胞健康状况和密度；滴定 DNA；添加 GFP 转染对照
核转染率低	检查细胞健康状况和密度；滴定 DNA；悬浮成单细胞
FACS 后聚集或没有细胞	FACS 之前过滤细胞；解离成单细胞；以适当的密度悬浮
在连续稀释中聚集或没有细胞	对细胞进行重新计数；解离成单细胞并通过滤网过滤；检查连续稀释
阴性样品上的 SURVEYOR 背景较高	重新设计引物，以从不同位置引发
凝胶上的 SURVEYOR 结果较脏	纯化 PCR 产物；减少输入 DNA；将 42°C 孵育减少至 30 m
无 SURVEYOR 切割	纯化并标准化 PCR 产物；用 TaqB 缓冲液重新退火；重新设计 sgRNA；在 px330 骨架上验证 Cas9 序列
样品在 TBE 丙烯酰胺凝胶中不下沉	补充 MgCl2 至终浓度为 15 mM 或添加包含甘油的加样缓冲液

[1769] 讨论

[1770] 可以容易地将CRISPR-Cas多元化，以促进同时修饰若干基因并以较高 效率介导染色体微小缺失。申请人使用了两种sgRNA展示以高达68％的效率 同时靶向HEK293FT细胞中的人类GRIN2B和DYRK1A座位。同样地，可以 使用一对sgRNA介导微小缺失，如切除外显子，可以通过PCR在克隆水平上 对微小缺失进行基因分型。注意的是外显子接合的精确位置可以变化。申请 人还展示了使用ssODN和靶向载体在HEK 293FT和HUES9细胞中用Cas9的野生型和切口酶突变体两者介导HDR(图30)。注意的是申请人还不能使 用Cas9切口酶在HUES9细胞中检测到HDR，这可能是归因于在HUES9细 胞中效率较低或修复活性方面的潜在差异。尽管这些值是典型的，但是给定 的sgRNA的切割效率仍存在一些可变性，并且在个别情况下某些sgRNA可 能不工作，究其原因仍是未知的。申请人推荐为每个座位设计两种sgRNA，并在预期的细胞类型中测试其效率。

[1771] 实例31:NLS

[1772] Cas9转录调节剂:申请人着手将Cas9/gRNA CRISPR系统变成通用的 DNA结合系统,在该系统中可以执行超过DNA切割的功能。例如,通过将 一个或多个功能结构域融合到无催化活性的Cas9上,申请人已经给予了新颖 的功能,如转录激活/抑制、甲基化/脱甲基化或染色质修饰。为了完成此目 标,申请人通过将两个核酸酶活性必需的残基D10和H840变为丙氨酸而制 备了无催化活性的Cas9突变体。通过突变这两个残基,Cas9的核酸酶活性被 消除同时维持结合靶DNA的能力。申请人决定聚焦以测试申请人的假设的功 能结构域是转录激活因子VP64和转录阻遏子SID和KRAB。

[1773] Cas9核定位:申请人假设最有效的Cas9转录调节剂将牢固地定位于细胞 核,在细胞核中它将对转录具有最大的影响。此外,细胞质中的任何残余的 Cas9都可能具有不想要的作用。申请人确定野生型Cas9因为不包括多个核定 位信号(NLS)而不定位进细胞核中(尽管CRISPR系统无需具有一个或多个 NLS,但是有利地具有至少一个或多个NLS)。因为需要多个NLS序列,所 以推断难于将Cas9进入细胞核中并且融合到Cas9上的任何另外的结构域都 可能破坏核定位。因此,申请人制备了四种具有不同NLS序列的Cas9-VP64- GFP融合构建体(pXRP02-pLenti2-EF1a-NLS-hSpCsn1(10A,840A)-NLS-VP64- EGFP、pXRP04-pLenti2-EF1a-NLS-hSpCsn1(10A,840A)-NLS-VP64-2A-EGFP- NLS、pXRP06-pLenti2-EF1a-NLS-EGFP-VP64-NLS-hSpCsn1(10A,840A)- NLS、pXRP08-pLenti2-EF1a-NLS-VP64-NLS-hSpCsn1(10A,840A)-NLS-VP64- EGFP-NLS)。将这些构建体克隆进处于人类EF1a启动子的表达之下的lenti 骨架中。还添加了WPRE元件,用于更稳健地表达蛋白质。使用Lipofectame 2000将每个构建体转染进HEK 293FT细胞中并在转染后24小时成像。当融 合蛋白在该融合蛋白的N-末端和C-末端上都具有NLS序列时获得最佳核定 位。观察到的最高核定位出现在具有四个NLS元件的构建体中。

[1774] 为了更稳健地理解NLS元件对Cas9的影响,申请人通过将同一α输入 蛋白NLS序列融合在三个串联重复的看着为零的N-末端或C-末端而制备了 16个Cas9-GFP融合。使用Lipofectame 2000将每个构建体转染进HEK 293FT 细胞中并在转染后24小时成像。值得注意地,NLS元件的数目不与核定位的 范围直接相关。在C-末端上添加NLS比在N-末端上添加对核定位具有更大 的影响。

[1775] Cas9转录激活因子:申请人通过靶向Sox2座位并通过RT-qPCR定量转 录激活而在功能上对Cas9-VP64蛋白进行了测试。选择了八个DNA靶位点来 跨Sox2的启动子。使用Lipofectame 2000将每个构建体转染进HEK 293FT细 胞中并且在转染后72小时,从细胞中提取总RNA。在40ul反应中,将1ug 的RNA反转录成cDNA(qScript Supermix)。将2ul的反应产物添加进单个 的20ul TaqMan测定qPCR反应中。在生物和技术上一式三份地进行每个实验。没有RT对照并且没有模板对照反应显示出无扩增。不显示出强烈的核定 位的构建体pXRP02和pXRP04导致无激活。对于的确显示出较强的核定位的 构建体pXRP08,观察到适度的激活。可以在指导RNA Sox2.4和Sox2.5的情 况下观察到统计学上有义意的激活。

[1776] 实例32:体内小鼠数据

[1777] 材料与试剂

[1778] Herculase II融合聚合酶(安捷伦技术公司,目录号600679)

[1779] 10x NE缓冲液4(NEB,目录号B7004S)

[1780] BsaI HF(NEB,目录号R3535S)

[1781] T7 DNA连接酶(Enzymatic公司,目录号L602L)

[1782] 快速消化缓冲液,10X(赛默飞世尔科技公司,目录号B64)

[1783] 快速消化NotI(赛默飞世尔科技公司,目录号FD0594)

[1784] FastAP碱性磷酸酶(赛默飞世尔科技公司,目录号EF0651)

[1785] Lipofectamine2000(生命技术公司,目录号11668-019)

[1786] 胰蛋白酶(生命技术公司,目录号15400054)

[1787] Forceps#4(西格玛公司,目录号Z168777-1EA)

[1788] Forceps#5(西格玛公司,目录号F6521-1EA)

[1789] 10x汉克氏平衡盐溶液(Hank's Balanced Salt Solution)(西格玛公司,目录 号H4641-500ML)

[1790] 青霉素/链霉素溶液(生命技术公司,目录号P4333)

[1791] Neurobasal(生命技术公司,目录号21103049)

[1792] B27添加物(生命技术公司,目录号17504044)

[1793] L-谷氨酰胺(生命技术公司,目录号25030081)

[1794] 谷氨酸盐(西格玛公司,目录号RES5063G-A7)

[1795] β-巯基乙醇(西格玛公司,目录号M6250-100ML)

[1796] HA兔抗体(细胞信号传导,目录号3724S)

[1797] LIVE/DEAD®细胞成像试剂盒(生命技术公司,目录号R37601)

[1798] 30G世界精密仪器(World Precision Instrument)注射器(世界精密仪器 (World Precision Instruments),目录号NANOFIL)

[1799] 立体定位仪(科夫仪器(Kopf Instruments))

[1800] UltraMicroPump3(世界精密仪器,目录号UMP3-4)

[1801] 蔗糖(西格玛公司,目录号S7903)

[1802] 氯化钙(西格玛公司,目录号C1016)

[1803] 乙酸镁(西格玛公司,目录号M0631)

[1804] Tris-HCl(西格玛公司,目录号T5941)

[1805] EDTA(西格玛公司,目录号E6758)

[1806] NP-40(西格玛公司,目录号NP40)

[1807] 苯甲基磺酰氟(西格玛公司,目录号78830)

[1808] 氯化镁(西格玛公司,目录号M8266)

[1809] 氯化钾(西格玛公司,目录号P9333)

[1810] β-甘油磷酸酯(西格玛公司,目录号G9422)

[1811] 甘油(西格玛公司,目录号G9012)

[1812] Vybrant® DyeCycle™ Ruby Stain(生命技术公司,目录号S4942)

[1813] FACS Aria Flu-act-细胞分选仪(美国剑桥MIT的科赫研究所(Koch Institute)

[1814] DNAeasy Blood&Tissue试剂盒(凯杰公司,目录号69504)

[1815] 程序

[1816] 构建在脑中体内使用的gRNA多元体

[1817] 申请人设计并PCR扩增了靶向小鼠TET和DNMT家族成员的单个的 gRNA(如在此描述的)。在N2a细胞系中评估靶向效率(图33)。为了在 体内获得若干基因的同时修饰,在AAV包装的载体中对有效的gRNA进行多 元化(图34)。为了促进系统效率的进一步分析,申请人向该系统上添加了 与人类突触蛋白I启动子的控制之下的GFP-KASH结构域融合蛋白一致的表 达盒(图34)。此修饰允许在神经元群体中进一步分析系统效率(更详细的 程序参见章节分选细胞核与体内结果)。

[1818] 使用以下引物使用Herculase II融合聚合酶PCR扩增该系统的所有4个部 分:

[1819] 第一U6 Fw:

[1820] gagggtctcgtccttgcggccgcgctagcgagggcctatttcccatgattc

[1821] 第一gRNA Rv:

[1822] ctcggtctcggtAAAAAAgcaccgactcggtgccacttttcaagttgataacggactagc

[1823] cttattttaacttgctaTTTCtagctctaaaacNNNNNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCAC

[1824] 第二U6 Fw:

[1825] gagggtctcTTTaccggtgagggcctatttcccatgattcc

[1826] 第二gRNA Rv:

[1827] ctcggtctcctcAAAAAAgcaccgactcggtgccacttttcaagttgataacggactagc

[1828] cttattttaacttgctaTTTCtagctctaaaacNNNNNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCAC

[1829] 第三U6 Fw:

[1830] gagggtctcTTTgagctcgagggcctatttcccatgattc

[1831] 第三gRNA Rv:

[1832] ctcggtctcgcgtAAAAAAgcaccgactcggtgccacttttcaagttgataacggactag

[1833] ccttattttaacttgctaTTTCtagctctaaaacNNNNNNNNNNNNNNNNNNNGGTGTTTCGTCCTTTCCA

[1834] hSyn_GFP-kash Fw:gagggtctcTTacgcgtgtgtctagac

[1835] hSyn_GFP-kash Rv:

[1836] ctcggtctcAaggaCAGGGAAGGGAGCAGTGGTTCACGCCTGTAATCCCA GCAATTTGGGA GGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTTGC (NNNNNNNNNNNNNNNNNNN是反向互补序列靶向的基因组序列)

[1837] 申请人使用了Golden Gate策略在单步反应中组装了该系统的所有部分 (1:1分子比):

[1838]

第一 U6_gRNA	18 ng
第二 U6_gRNA	18 ng
第三 U6_gRNA	18 ng
Syn_GFP-kash	100 ng
10x NE 缓冲液 4	1.0 μl
10x BSA	1.0 μl
10 mM ATP	1.0 μl
BsaI HF	0.75 μl
T7 连接酶	0.25 μl
ddH_2O	10 μl

循环数	条件
1-50	37°C 持续 5 m，21°C 持续 5 m

[1839] 使用Herculase II融合聚合酶和以下引物PCR扩增Golden Gate反应产 物：

[1840] Fw 5’cctgtccttgcggccgcgctagcgagggcc

[1841] Rv 5’cacgcggccgcaaggacagggaagggagcag

[1842] 使用NotI限制酶切位点将PCR产物克隆进AAV骨架中在ITR序列之 间：

[1843] PCR产物消化：

[1844]

快速消化缓冲液，10X	3 μl
快速消化 NotI	1 μl
DNA	1 μg
ddH_2O	直到 30 μl

[1845] AAV骨架消化：

[1846]

快速消化缓冲液，10X	3 μl
快速消化 NotI	1 μl
FastAP 碱性磷酸酶	1 μl
AAV 骨架	1 μg
ddH_2O	直到 30 μl

[1847] 在37℃下孵育20min后，使用QIAQuick PCR纯化试剂盒纯化样品。如 下以1:3载体:插入片段比例连接标准化的样品：

[1848]

消化的 pUC19	50 ng
消化的插入片段	1：3 载体：插入片段摩尔比
T7 连接酶	1 μl
2X 快速连接缓冲液	5 μl
ddH_2O	直到 10 μl

[1849] 用连接反应产物转化细菌后，申请人用桑格测序确认了获得的克隆。

[1850] 与Cas9构建体共转染后，在N2a细胞中测试阳性DNA克隆(图35和 36)。

[1851] 设计用于AAV递送的新的Cas9构建体

[1852] 尽管其具有独特的功能，AAV递送系统具有包装限制-为了成功地在体 内递送表

达盒,则它的尺寸不得不<4.7kb。为了减少SpCas9表达盒的尺寸 并促进递送,申请人测试了若干改变:不同启动子、较短的polyA信号以及 最后是较小版本的来自金黄色葡萄球菌的Cas9(SaCas9)(图37和38)。 所有测试的启动子先前已经被测试并且公开为在神经元中有活性,包括小鼠 Mecp2(格雷(Gray)等人,2011)、大鼠Map1b和截短的大鼠Map1b(刘(Liu)和费舍尔(Fischer),1996)。替代性合成polyA序列先前也已显示 具有功能(莱维特(Levitt)等人,1989;格雷(Gray)等人,2011)。在用 Lipofectamine 2000转染后,在N2a细胞中表达所有克隆的构建体,并用蛋白 质印迹法进行测试(图39)。

[1853] 在初级神经元中测试AAV多元系统

[1854] 为了确认研发的系统在神经元中的功能性,申请人使用体外初级神经元 培养物。根据先前由班克(Banker)和戈斯林(Goslin)(班克和戈斯林, 1988)公开的方案制备小鼠皮层神经元。

[1855] 第16天,从胚胎中获得神经元细胞。胚胎提取自安乐死的怀孕雌鼠,并 且将头部置于冰冷的HBSS中。然后,用镊子(#4和#5)从头骨中提取大脑 并将其转移至另一个更换的冰冷的HBSS中。在立体显微镜和#5镊子的辅助 下在填充有冰冷的HBSS的皮氏培养皿中进行另外的步。将半球彼此分离并 与脑干分离并清除脑膜。然后,非常谨慎地解剖海马并将其置于填充有冰冷 的HBSS的15ml锥形管中。可以使用在海马解剖后留下的皮层,在除去脑干 残留物和嗅球后使用类似的方案进行另外的细胞分离。将分离的海马用10ml 冰冷的HBSS洗涤三次并通过在37℃下于HBSS中用胰蛋白酶孵育15min而 解离(4ml HBSS,添加10μl 2.5%胰蛋白酶/海马)。胰酶消化后,将海马非 常谨慎地洗涤三次,以用预加热至37℃的HBSS除去任何痕量的胰蛋白酶并 在温HBSS中解离。申请人通常使用1ml移液管尖端在1ml HBSS中解离获 得自10-12个胚胎的细胞并稀释解离的细胞直到4ml。将细胞以250个细胞 /mm2的密度铺板并在37℃和5%CO2下培养长达3周

[1856] HBSS

[1857] 435ml H2O

[1858] 50ml 10x汉克氏平衡盐溶液

[1859] 16.5ml 0.3M HEPES pH 7.3

[1860] 5ml青霉素-链霉素溶液

[1861] 过滤(0.2μm)并存储在4℃下

[1862] 神经元铺板培养基(100ml)

[1863] 97ml Neurobasal

[1864] 2ml B27添加物

[1865] 1ml青霉素-链霉素溶液

[1866] 250μl谷氨酰胺

[1867] 125μl谷氨酸

[1868] 用来自HEK293FT细胞的过滤培养基的浓缩AAV1/2病毒或AAV1病毒 转导神经元,培养4-7天之间并且在转导后在培养物中保存至少一周,以允许 递送的基因表达。

[1869] AAV驱动的系统的表达

[1870] 申请人使用蛋白质印迹法确认了AAV递送后SpCas9和SaCas9在神经元 培养物中的表达(图42)。转导后一周,用β-巯基乙醇在NuPage SDS加样 缓冲液中收集神经元,以使

蛋白质在95℃下变性5min。将样品在SDS PAGE凝胶上分离并转移到PVDF膜上，用于进行WB蛋白质检测。用HA抗 体检测Cas9蛋白。

[1871] 用荧光显微镜确认来自gRNA多元AAV的Syn-GFP-kash的表达(图 50)。

[1872] 毒性

[1873] 为了评估具有CRISPR系统的AAV的毒性，申请人在病毒转导后一周测 试了神经元的整体形态(图45)。另外，申请人用LIVE/DEAD®细胞成像试 剂盒测试了设计的系统的潜在的毒性，该试剂盒允许区分培养物中的活细胞 和死细胞。它是基于细胞内酯酶活性的存在(如通过将非荧光钙黄绿素AM 酶促转化为强烈的绿色荧光钙黄绿素所确定的)。另一方面，该试剂盒的红 色的细胞不通透性组分仅进入具有受损膜的细胞并结合到死细胞中的产荧光 DNA上。两种荧光团都可以容易地用荧光显微镜在活细胞中可视化。AAV驱 动的Cas9蛋白和多元gRNA构建体在初级皮层神经元中的表达耐受性良好并 且无毒(图43和44)，这指示设计的AAV系统适于体内测试。

[1874] 病毒产生

[1875] 根据描述于麦克卢尔(McClure)等人，2011中的方法产生浓缩的病 毒。上清液病毒产生出现在HEK293FT细胞中。

[1876] 脑部手术

[1877] 对于病毒载体注射，通过腹膜内注射用氯胺酮/甲苯噻嗪混合物(氯胺酮 剂量为100mg/kg并且甲苯噻嗪剂量为10mg/kg)麻醉的10-15周龄的雄性 C57BL/6N小鼠。将腹膜内给予Buprenex用作优先购买的镇痛剂(1 mg/kg)。使用耳内定位螺柱和齿条将动物固定在Kopf立体定位仪中，以维 持头骨不动。使用手持电钻，在前囟后面-3.0mm且在海马的CA1区域中的 供注射的侧面3.5mm处钻出一个洞(1-2mm)。使用深度为2.5mm的30G 世界精密仪器注射器，注射总体积为1ul的AAV病毒颗粒溶液。通过流速为 0.5ul/min的世界精密仪器UltraMicroPump3’注射泵监测注射，以防止损伤组 织。当完成注射时，以0.5mm/min的速度缓慢地移除注射针。注射后，用6- 0爱惜良(Ethilon)缝合线将皮肤缝合。在手术后，给动物供应1mL乳酸林 格氏溶液(皮下的)并放在温度受控的(37℃)环境中，直到恢复走动。术 后3周，通过深度麻醉将动物处以安乐死，随后除去组织，用于细胞核分选 或用4％多聚甲醛灌注用于免疫化学。

[1878] 分选细胞核与体内结果

[1879] 申请人设计了一种的方法以用GFP特异地在遗传上标记gRNA靶向的神 经元细胞核，用以标记的细胞核的荧光激活细胞分选术(FACS)以及并 DNA、RNA和核内蛋白的下游加工。为了这个目的，诸位申请人的多元靶向 载体被设计成表达在GFP与小鼠细胞核膜蛋白结构域KASH(斯塔尔DA (Starr DA)，2011，《当代生物学》(Current biology))之间的融合蛋白 和用于靶向特定的感兴趣的基因座位的3个gRNA(图34)。在人类突触蛋 白启动子的控制之下表达GFP-KASH，以特异性标记神经元。融合蛋白GFP- KASH的氨基酸是：

[1880] MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT

[1881] TGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFF KDDGN

[1882] YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK QKNGIK

[1883] VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNE KRDHMV

[1884] LLEFVTAAGITLGMDELYKSGLRSREEEEETDSRMPHLDSPGSSQPRRSFLS RVIRAAL

[1885] PLQLLLLLLLLLACLLPASEDDYSCTQANNFARSFYPMLRYTNGPPPT

[1886] AAV1/2介导的递送进入脑后一周,观察到GFP-KASH的稳健表达。对 于FACS和标记的细胞核的下游加工,术后3周将海马解剖并加工,用于使 用梯度离心步骤进行细胞核纯化。为了这个目的,使用2ml Dounce均质机 (西格玛公司)将组织在320mM蔗糖、5mM CaCl、3mM Mg(Ac)2、10 mM Tris pH 7.8、0.1mM EDTA、0.1%NP40、0.1mM苯甲基磺酰氟 (PMSF)、1mMβ-巯基乙醇中均质化。根据制造商的方案,将匀浆在4℃ 下以3.500rpm以25%至29% Optiprep®梯度离心30min。将细胞核沉淀物再 悬浮于340mM蔗糖、2mM MgCl2、25mM KCl、65mM甘油磷酸酯、5%甘 油、0.1mM PMSF、1mMβ-巯基乙醇中并向标记的细胞核中添加 Vybrant® DyeCycle™ Ruby Stain(生命技术公司)(为DNA提供近红外发射)。使用 Aria Flu-act-细胞分选仪和BDFACS Diva软件通过FACS分选标记的和纯化的 细胞核。最后,使用DNAeasy Blood&Tissue试剂盒(凯杰公司),将分选 的GFP+和GFP-细胞核用于纯化基因组DNA,用于对靶向的基因组区域进行 Surveyor测定分析。可以容易地使用同一途径从靶向的细胞中纯化出细胞核 RNA或蛋白质,用于下游加工。由于2-载体系统(图34),诸位申请人在此 途径中使用有效的Cas9介导的DNA切割预期仅在脑中的一小部分的细胞中 出现(用多元靶向载体和Cas9编码载体两者共转染的细胞)。在此描述的方 法使得诸位申请人可以从表达3种感兴趣的gRNA并且因此应该会经历Cas9 介导的DNA切割的细胞群体中特异地纯化出DNA、RNA和核内蛋白。通过 使用此方法,诸位申请人能够可视化仅发生在一小部分的细胞中的有效的体 内DNA切割。

[1887] 实质上,申请人在此显示的是靶向的体内切割。此外,申请人使用了一 种多元途径,其中同时但是独立地靶向若干不同序列。呈现的系统可以适于 研究脑部病理病症(基因敲除,例如帕金森病)并且还为进一步研发脑中的 基因组编辑工具开辟了一个领域。通过将核酸酶活性替换为基因转录调节剂 或表观遗传调节剂,可以回答关于脑中的基因调节和表观改变不仅在病理病 症而且在生理过程(如学习和记忆形成)中的作用的整个范围的科学问题。 最后,呈现的技术可以应用于更复杂的哺乳动物系统(如灵长类)中,该技 术允许克服当前的技术局限。

[1888] 实例33:模型数据

[1889] 已经具体研究了若干疾病模型。这些疾病模型包括新生(de novo)自闭症 风险基因CHD8、KATNAL2和SCN2A;以及综合征性自闭症(安格曼综合 征(Angelman Syndrome))基因UBE3A。这些基因以及生成的自闭症模型 当然是优选的,但是显示本发明可以被应用于任何基因并且因此任何模型都 是可能的。

[1890] 申请人已经使用Cas9核酸酶在人类胚胎干细胞(hESC)中制备了这些 细胞系。通过用Cbh-Cas9-2A-EGFP和pU6-sgRNA瞬时转染hESC而建立了 这些系。为最经常地靶向相同外显子的每个基因设计了两种sgRNA,在这些 外显子中患者无义(敲除)突变最近已经由自闭症患者的全外显子组测序研 究得到描述。为此项目特异建立了Cas9-2A-EGFP和pU6质粒。

[1891] 实例34:AAV产生系统或方案

[1892] 在此提供了为高通量筛选用途研发并且与其一起工作特别好的AAV产生 系统或方案,但是它在本发明中也具有较广泛的适用性。操纵内源基因表达 提出了各种挑战,因

为表达率取决于许多因素,包括调节元件、mRNA加工 和转录本稳定性。为了克服此挑战,申请人研发了一种用于供递送的基于腺 相关病毒(AAV)的载体。AAV具有一个基于ssDNA的基因组并且因此较 不易受重组影响。

[1893] AAV1/2(血清型AAV1/2,即杂交体或嵌合体AAV1/AAV2衣壳AAV) 肝素纯化的浓缩病毒方案

[1894] 培养基:D10+HEPES

[1895] 500ml瓶DMEM高葡萄糖+Glutamax(GIBCO)

[1896] 50ml Hyclone FBS(热灭活的)(赛默飞世尔)

[1897] 5.5ml HEPES溶液(1M,GIBCO)

[1898] 细胞:低传代HEK293FT(产生病毒时传代<10次,解冻新的传代2-4次的 细胞用于产生病毒,生长直到3-5次传代)

[1899] 转染试剂:聚乙烯亚胺(PEI)“Max”

[1900] 将50mg PEI“Max”溶解于50ml无菌的超纯H20中

[1901] 将pH调节至7.1

[1902] 用0.22um fliptop过滤器过滤

[1903] 用封口膜将管和包裹物密封

[1904] 在-20℃下将等分部分冷冻(用于存储,也可以立即使用)

[1905] 细胞培养

[1906] 在D10+HEPES中培养低传代HEK293FT

[1907] 每天传代1:2与1:2.5之间

[1908] 有利地不允许达到细胞超过85%的汇合

[1909] 对于T75

[1910] -温10ml HBSS(-Mg2+,-Ca2+,GIBCO)+1ml TrypLE Express (GIBCO)/烧瓶至37℃(水浴)

[1911] 完全地抽出培养基

[1912] -轻轻地添加10ml温HBSS(以完全洗掉培养基)

[1913] -添加1ml TrypLE/烧瓶

[1914] -将烧瓶在恒温箱(37℃)中放置1min

[1915] -摇动烧瓶以使细胞脱离开

[1916] -添加9ml D10+HEPES培养基(37℃)

[1917] -上下吸移5次,以产生单细胞悬浮液

[1918] -以1:2-1:2.5(对于T75,是12ml培养基)比例分离(如果细胞生长更缓 慢,则丢弃并解冻新的批次,它们未处于最佳生长)

[1919] -当存在足够的细胞时即转移至T225中(对于易于处理大量的细胞)

[1920] AAV产生(5*15cm培养皿尺度/构建体):

[1921] 将21.5ml培养基中的1千万个细胞铺进15cm培养皿中

[1922] 在37℃下孵育18-22小时

[1923] 在80%汇合处转染是理想的

[1924] 每个平板

[1925] 预温的22ml培养基(D10+HEPES)

[1926] 准备具有DNA混合物的管(使用无内毒素大量质粒提取DNA(endofree maxiprep DNA))：

[1927] 5.2ug感兴趣的质粒载体

[1928] 4.35ug AAV 1血清型质粒

[1929] 4.35ug AAV 2血清型质粒

[1930] 10.4ug pDF6质粒(腺病毒辅助基因)□涡旋至混合

[1931] 添加434uL DMEM(无血清!)

[1932] 添加130ul PEI溶液

[1933] 涡旋5-10秒

[1934] 向预温的培养基添加DNA/DMEM/PEI混合物

[1935] 简单地涡旋至混合

[1936] 将15cm培养皿中的培养基替换为DNA/DMEM/PEI混合物

[1937] 返回37℃恒温箱中

[1938] 收获之前孵育48h(确保培养基没有变得太酸)

[1939] 收获病毒：

[1940] 1.将培养基谨慎地从15cm培养皿中抽出(有利地不要移动细胞)

[1941] 2.向每个平板中添加25ml RT DPBS(英杰公司)并用细胞刮棒轻轻地移除细 胞。将悬浮液收集在50ml管中。

[1942] 3.将细胞在800x g下沉淀10分钟。

[1943] 4.丢弃上清液

[1944] 暂停点:希望的话,将细胞沉淀物冷冻在-80C下

[1945] 5.将沉淀物再悬浮于150mM NaCl、20mM Tris pH 8.0中,每个组织培养平板 使用10ml。

[1946] 6.在dH20中制备10％脱氧胆酸钠的新鲜溶液。向每个组织培养平板中添加1.25ml的此溶液至终浓度为0.5％。添加benzonase核酸酶,至终浓度为50个 单位/ml。将管充分混合。

[1947] 7.在37℃下孵育1小时(水浴)。

[1948] 8.通过在3000x g下离心15min除去细胞碎片。转移至干净的50ml管中并且 保证已经除去所有的细胞碎片,以防止堵塞肝素柱。

[1949] AAV1/2的肝素柱纯化：

[1950] 1.使用蠕动泵设置HiTrap肝素柱,这样使得溶液以每分钟1ml流过该 柱。重要的是保证没有将气泡引入进该肝素柱中。

[1951] 2.使用蠕动泵,用10ml 150mM NaCl、20mM Tris(pH 8.0)平衡该 柱。

[1952] 3.病毒的结合:向柱上施加50ml病毒溶液并允许流过。

[1953] 4.洗涤步骤1:用20ml 100mM NaCl、20mM Tris(pH 8.0)洗涤柱。

[1954] (使用蠕动泵)

[1955] 5.洗涤步骤2:使用3ml或5ml注射器继续用1ml 200mM NaCl、20 mM Tris(pH 8.0)洗涤该柱,随后用1ml 300mM NaCl、20mM Tris(pH 8.0)洗涤。

[1956] 丢弃流过物(flow-through)。

[1957] (在以上病毒溶液流过50min过程中准备具有不同缓冲液的注射器)

[1958] 6.洗脱使用5ml注射器和轻压(流速<1ml/min)通过施加以下项将病 毒从该柱上洗脱下来:

[1959] 1.5ml 400mM NaCl,20mM Tris,pH 8.0

[1960] 3.0ml 450mM NaCl,20mM Tris,pH 8.0

[1961] 1.5ml 500mM NaCl,20mM Tris,pH 8.0

[1962] 将这些收集在15ml离心管中。

[1963] AAV1/2的浓度:

[1964] 1.浓缩步骤1:使用具有100,000分子量截留的Amicon ultra 15ml离心过 滤装置浓缩洗脱的病毒。将柱洗脱物加样进浓缩器中并在2000x g下离心2分 钟(在室温下)。检查浓缩体积-它应该是大约500μl。必要的话,以1min间 隔离心,直到达到正确体积。

[1965] 2.缓冲液交换:向过滤装置中添加1ml无菌DPBS,以1min间隔离心, 直到达到正确体积(500ul)。

[1966] 3.浓缩步骤2:向Amicon Ultra 0.5ml 100K过滤装置中添加500ul浓缩 物。在6000g下离心2min。检查浓缩体积-它应该是大约100μl。必要的 话,以1min间隔离心,直到达到正确体积。

[1967] 4.回收:倒置滤芯并插入新鲜的收集管中。在1000g下离心2min。

[1968] 等分并在-80℃下冷冻 每个注射位点典型地需要1ul,因此推荐小的等分部分(例如5ul)(避免冻 融病毒)。

[1969] 使用qPCR确定DNaseI-抗性GC粒子效价(参见分离方案)

[1970] 材料

[1971] Amicon Ultra,0.5ml,100K;密理博公司;UFC510024

[1972] Amicon Ultra,15ml,100K;密理博公司;UFC910024

[1973] Benzonase核酸酶;西格玛-奥德里奇公司,E1014

[1974] HiTrap肝素柱体;西格玛-奥德里奇公司;54836

[1975] 脱氧胆酸钠;西格玛-奥德里奇公司;D5670

[1976] AAV1上清液产生方案

[1977] 培养基:D10+HEPES

[1978] 500ml瓶DMEM高葡萄糖+Glutamax(英杰公司)

[1979] 50ml Hyclone FBS(热灭活的)(赛默飞世尔)

[1980] 5.5ml HEPES溶液(1M,GIBCO)

[1981] 细胞:低传代HEK293FT(产生病毒时传代<10次)

[1982] 解冻新的传代2-4次的细胞用于产生病毒,生长直到2-5次传代

[1983] 转染试剂:聚乙烯亚胺(PEI)“Max”

[1984] 将50mg PEI“Max”溶解于50ml无菌的超纯H20中

[1985] 将pH调节至7.1

[1986] 用0.22um fliptop过滤器过滤

[1987] 用封口膜将管和包裹物密封

[1988] 在-20℃下将等分部分冷冻(用于存储,也可以立即使用)

[1989] 细胞培养

[1990] 在D10+HEPES中培养低传代HEK293FT每天传代1:2与1:2.5之间

[1991] 有利地一定让细胞达到超过85％的汇合

[1992] 对于T75

[1993] -温10ml HBSS(-Mg2+,-Ca2+,GIBCO)+1ml TrypLE Express (GIBCO)/烧瓶至37℃(水浴)

[1994] -完全地抽出培养基

[1995] -轻轻地添加10ml温HBSS(以完全洗掉培养基)

[1996] -添加1ml TrypLE/烧瓶

[1997] -将烧瓶在恒温箱(37℃)中放置1min

[1998] -摇动烧瓶以使细胞脱离开

[1999] -添加9ml D10+HEPES培养基(37℃)

[2000] -上下吸移5次,以产生单细胞悬浮液

[2001] -以1:2-1:2.5(对于T75,是12ml培养基)比例分离(如果细胞生长更缓 慢,则丢弃并解冻新的批次,它们未处于最佳生长)

[2002] -当存在足够的细胞时即转移至T225中(对于易于处理大量的细胞)

[2003] AAV产生(单个15cm培养皿尺度)

[2004] 将21.5ml培养基中的1千万个细胞铺进15cm培养皿中

[2005] 在37℃下孵育18-22小时

[2006] 在每个平板80％汇合处转染是理想的

[2007] 预温的22ml培养基(D10+HEPES)

[2008] 准备具有DNA混合物的管(使用无内毒素大量质粒提取DNA(endofree maxiprep DNA))：

[2009] 5.2ug感兴趣的质粒载体

[2010] 8.7ug AAV 1血清型质粒

[2011] 10.4ug DF6质粒(腺病毒辅助基因)

[2012] 涡旋至混合

[2013] 添加434uL DMEM(无血清!)添加130ul PEI溶液

[2014] 涡旋5-10秒

[2015] 向预温的培养基添加DNA/DMEM/PEI混合物

[2016] 简单地涡旋至混合

[2017] 将15cm培养皿中的培养基替换为DNA/DMEM/PEI混合物

[2018] 返回37℃恒温箱中

[2019] 收获之前孵育48h(有利地进行监测,以保证培养基没有变得太酸)

[2020] 收获病毒：

[2021] 从15cm培养皿中除去上清液

[2022] 用0.45um过滤器过滤(较低的蛋白质结合)等分并在-80℃下冷冻

[2023] 转导(24孔格式中的原代神经元培养物,5DIV)

[2024] 将有待转导的神经元的每个孔中的完全neurobasal培养基替换为新鲜的neurobasal(通常替换每个孔中的500ul中的400ul)

[2025] 在37℃水浴中解冻AAV上清液

[2026] 在恒温箱中平衡30min

[2027] 向每个孔中添加250ul AAV上清液

[2028] 在37℃下孵育24h

[2029] 移除培养基/上清液并替换为新鲜的完全neurobasal

[2030] 48h后表达开始可见,感染后6-7天左右饱和

[2031] 具有GOI的pAAV质粒的构建体不应该超过4.8kb(包括两个ITRS)。

[2032] 人类密码子优化序列(即针对在人类中的表达而优化的)序列的实例：下面提供了SaCas9：

[2033] ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGC

[2034] TTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAAAGGAACTA CATTCT

[2035] GGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTAT GAAAC

[2036] AAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGT GGAAAA

[2037] CAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAACGAC GGAGAA

[2038] GGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGATTACAACCTGCT GACCG

[2039] ACCATTCTGAGCTGAGTGGAATTAATCCTTATGAAGCCAGGGTGAAAGG CCTGA

[2040] GTCAGAAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGC TAAGC

[2041] GCCGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACG AGCTG

[2042] TCTACAAAGGAACAGATCTCACGCAATAGCAAAGCTCTGGAAGAGAAG TATGTC

[2043] GCAGAGCTGCAGCTGGAACGGCTGAAGAAAGATGGCGAGGTGAGAGGG TCAATT

[2044] AATAGGTTCAAGACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTG AAAGTG

[2045] CAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGATACTTATATCG ACCTGC

[2046] TGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCT TCGGAT

[2047] GGAAAGACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACCT ATTTTCC

[2048] AGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGTACAA CGCCCT

[2049] GAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAACGAGAAACT GGAATA

[2050] CTATGAGAAGTTCCAGATCATCGAAAACGTGTTTAAGCAGAAGAAAAA GCCTAC

[2051] ACTGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAA GGGCTA

[2052] CCGGGTGACAAGCACTGGAAAACCAGAGTTCACCAATCTGAAAGTGTAT CACGA

[2053] TATTAAGGACATCACAGCACGGAAAGAAATCATTGAGAACGCCGAACT GCTGGA

[2054] TCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGACATCCAG GAAGAG

[2055] CTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATT AGTAAT

[2056] CTGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAAGCTATCAATC TGATTC

[2057] TGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCG GCTGAA

[2058] GCTGGTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAAC CACACT

[2059] GGTGGACGATTTCATTCTGTCACCCGTGGTCAAGCGGAGCTTCATCCAG AGCATCAA

[2060] AGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATATCATT ATCGAGCT

[2061] GGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGAT GCAGAAA

[2062] CGAAACCGGCAGACCAATGAACGCATTGAAGAGATTATCCGAACTACC GGGAAAGAGAACGCAAAGT ACCTGATTGAAAAAATCAAGCTGCACGAT ATGCAGGAGGGAAAGT

[2063] GTCTGTATTCTCTGGAGGCCATCCCCCTGGAGGACCTGCTGAACAATCC ATTCAACT

[2064] ACGAGGTCGATCATATTATCCCCAGAAGCGTGTCCTTCGACAATTCCTTT AACAACA

[2065] AGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAGGACTC CTTTCCAG

[2066] TACCTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCA CATTCTG

[2067] AATCTGGCCAAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTAC CTGCTGG

[2068] AAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCG GAATCTG

[2069] GTGGACACAAGATACGCTACTCGCGGCCTGATGAATCTGCTGCGATCCT ATTTCCGG

[2070] GTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTCACAT CTTTTCTG

[2071] AGGCGCAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCAC CATGCCG

[2072] AAGATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAA AAAGCTGG

[2073] ACAAAGCCAAGAAAGTGATGGAGAACCAGATGTTCGAAGAGAAGCAGG CCGAATC

[2074] TATGCCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATCACT CCTCACC

[2075] AGATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCTCACCGGGT GGATAAA

[2076] AAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAAAA GACGATAA

[2077] GGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGAT AATGACA

[2078] AGCTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACC ACCATGAT

[2079] CCTCAGACATATCAGAAACTGAAGCTGATTATGGAGCAGTACGGCGACG AGAA

[2080] GAACCCACTGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAG TATAGCAAAA

[2081] AGGATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGC TGAATGCC

[2082] CATCTGGACATCACAGACGATTACCCTAACAGTCGCAACAAGGTGGTCA AGCTGTC

[2083] ACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTGTATAAA TTTGTGAC

[2084] TGTCAAGAATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAAT AGCAAGT

[2085] GCTACGAAGAGGCTAAAAAGCTGAAAAAGATTAGCAACCAGGCAGAGT TCATCGCC

[2086] TCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATAGGG TCATCGGG

[2087] GTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTGACATCA CTTACCGA

[2088] GAGTATCTGGAAAACATGAATGATAAGCGCCCCCCTCGAATTATCAAAA CAATTGC

[2089] CTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAAC CTGTATGA

[2090] GGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAATT C

[2091] 实例35:使用Cas9切口酶和两种指导RNA最小化脱靶切割

[2092] Cas9是一种可以在20bp RNA指导的帮助下靶向基因组中的特定位置的 RNA-指导的DNA核酸酶。然而,该指导序列可以容忍指导序列与DNA-靶 序列之间的一些错配。当该指导RNA将Cas9靶向具有几个与该指导序列不 同的碱基的脱靶序列时,归因于脱靶切割的可能性,该灵活性是不令人希望 的。对于所有实验应用(基因靶向、作物工程、治疗应用等),重要的是能 够改进Cas9介导的基因靶向的特异性并减少Cas9的脱靶修饰的可能性。

[2093] 申请人研发了一种使用与两种指导RNA组合的Cas9切口酶突变体以促 进基因组中的靶向的双链断裂而无脱靶修饰的方法。可以通过使其切割活性 无效而从Cas9核酸酶

产生Cas9切口酶突变体,这样使得并非切割DNA双链 体的两条链,仅仅切割一条链。可以通过在Cas9核酸酶的一个或多个结构域(例如Ruvc1或HNH)中诱导突变而产生Cas9切口酶。这些突变可以包括 但不限于Cas9催化结构域中的突变,例如在SpCas9中,这些突变可以在位置D10或H840处。这些突变可以包括但不限于SpCas9中的D10A、E762A、H840A、N854A、N863A或D986A,但是可以通过在其他CRISPR 酶或Cas9直向同源物中在对应的位置处诱导突变而产生切口酶。在本发明的 一个最优选的实施例中,该Cas9切口酶突变体是一种具有D10A突变的 SpCas9切口酶。

[2094] 这工作的方式是与Cas9切口酶组合的每个指导RNA都将诱导双螺旋 DNA靶标的靶向的单链断裂。由于每个指导RNA切口一条链,所以最终结 果是双链断裂。此方法消除脱靶突变的原因是因为非常不可能存在与两种指 导序列都具有较高水平的相似性的脱靶位点(20bp+2bp(PAM)=22bp针对 每个指导的特异性,并且两种指导意味着任何脱靶位点将不得不具有接近于 44bp的同源序列)。尽管仍可能的是单独的指导可能具有脱靶,但是那些脱 靶将仅仅是带切口的,这不可能会被诱变的NHEJ过程修复。因此,DNA双 链切口的多元化提供了一种强有力的引入靶向的DNA双链断裂而无脱靶诱变 效应的方式。

[2095] 申请人进行了以下实验,这些实验涉及用编码Cas9(D10A)切口酶的 质粒以及一种或多种指导物的DNA表达盒共转染HEK293FT细胞。申请人使 用Lipofectamine 2000转染了细胞,并且在转染后转染48或72小时收获了细 胞。如在此先前所所描述的,使用SURVEYOR核酸酶测定检测双切口诱导 的NHEJ(图51、52和53)。

[2096] 申请人进一步鉴定了涉及当与两种指导RNA结合时通过Cas9切口酶突 变体的有效切割的参数,并且这些参数包括但不限于该5'突出端的长度。针 对具有至少26个碱基对的5'突出端报道了有效的切割。在本发明的一个优选 实施例中,5'突出端是至少30个碱基对并且更优选至少34个碱基对。长达 200个碱基对的突出端对于切割而言可以是可接受的,然而少于100个碱基对 的5'突出端是优选的并且少于50个个碱基对的5'突出端是最优选的(图54 和55)。

[2097] 实例36:行为方案

[2098] 高架十字迷宫。使用高架十字迷宫评估焦虑样行为,从而探索啮齿动物 在开放区域中具有的先天性恐惧与它们希望探索新颖环境之间的冲突。

[2099] 用于高架十字迷宫测试的仪器由不锈钢制成并且由四个长35cm并且宽5 cm的臂组成(两个开放的没有壁并且两个由15cm高的壁封闭)。迷宫的每 个臂都附接到坚固的金属腿上,这样使得它被抬离地面40cm:

[2100] 将小鼠以多达每笼三只放入12hr光暗循环的随意获得食物和水的房间 中。

[2101] 在光线明亮的房间中在上午9:00与下午6:00之间进行行为测试。在第一 次试验开始前30min,将所有包含小鼠的笼转移至行为测试房间中。

[2102] 将一只小鼠从其笼中取出并放在开放与关闭臂的接点处,使开放臂朝向 与实验者相反的方向并且让小鼠自由地探索迷宫持续5min。在该时间过程 中,记录动物并追踪行为。用Noldus软件自动测量在开放臂中行进的总距 离、在开放臂中花费的时间、在臂之间转换的总次数以及进入开放臂的等待 时间。在5-min测试结束时,将啮齿动物从十字迷宫中移出并放入运输笼 中。放入到其居住笼内。在用另一只小鼠测试之前,将高架十字迷宫谨慎地 清理并用纸巾干燥。

[2103] 恐惧条件化

[2104] 动物:4周龄雄鼠(BL6J/57);每组10只

[2105] 足底电击:2s;0.5mA

[2106] 方案设计三个阶段:

[2107] 1.暴露,允许小鼠适应并熟悉训练室。对于不成对的组:适应2min,随 后呈现20-sec提示音,随后是80-sec间隔,重复六次。对于成对的组,适应 周期为12min。

[2108] 2.训练;为小鼠提供CS和或US刺激物。适应周期-4min。以持续100 sec的试验时间(包括试验间间隔)运行涉及音调(CS)和电击(US)的六 个条件化试验。对于成对的组,每个试验由20-sec“基线”间隔、20-sec音调 呈现、18-sec扫描间隔、2-sec电击和40-sec电击后间隔组成。对于不成对的 组,试验是一样的,除省略音调除外。

[2109] 3.测试;观察小鼠响应每个场景以及响应音调和痕量-CS的木僵。每个阶 段以24-h间隔出现。允许小鼠在训练(熟悉)和测试(改变的;具有一个格 子的平地板,在壁上是bw图形;有香兰素香味)场景下自由运动3min。所 有小鼠接受四个100-sec测试试验。对于所有小鼠,每个试验以20-sec间隔开 始并且20-sec音调呈现之后是60-sec间隔。

[2110] 评分:所有评分都是基于视频而完成。贯穿每个场景测试暴露和每个 100-sec试验时间以5-sec增量观察木僵%。如果动物至少木僵2s,则在每个 时间纠缠(time bin)中将木僵评分为1或0。

[2111] 在柱状图上呈现的数据是7只动物的组的平均值,误差棒表示SEM。按 时间呈现出训练/测试过程和每个点的折线图上的数据是来自6个或4个试验 (分别是训练和测试方案)的7只动物的平均值。误差棒表示SEM。

[2112] 旷场。广泛地使用旷场(Open Field)仪器评估探索行为并对其进行验证 以用于在测量焦虑相关行为中使用。将动物放在方形舞台(40x 40cm)中持 续10min同时记录其行为。为运动的总距离、速度和在舞台中央花费的时间 评分。

[2113] 新颖的物体识别。新颖的物体识别测试(NOR)是一种用于研究啮齿动 物的学习和记忆的流行的测试。它是基于它们与新颖的物体比熟悉的物体相 互作用要多的趋势。通常,首先将动物放在仪器中并允许探索物体。规定的 间隔后,将动物返回该仪器中,该仪器现在包含熟悉的物体和新颖的物体。 通过花费更多的时间与新颖的物体相互作用而区分物体识别。NOR任务是特 别有吸引力的,因为它不需要外部动机、奖励或惩罚,但是需要一点训练或 习惯,并且它可以在相对较短的时间内完成。

[2114] I预训练:

[2115] 1.到达聚居区之后,允许这些动物适应至少3-7d。

[2116] 2.开始实验之前至少3天,将动物每天处理至少3min/天并暴露于运 输惯例。

[2117] 3.熟悉测试环境-实验之前24h,将动物暴露于行为房间中的测试舞台 (40x 40x 35cm,清晰的壁)持续10min。

[2118] II物体识别训练:

[2119] 1.将两个‘完全一样的’有待熟悉的(样品)物品放在该仪器的后面 左侧和右侧角落中

[2120] 2.将动物放在与样品物品相对的壁的中点,这防止了放置动物中的任 何的无意识的偏差,这样使得它更朝向具体侧/物体定向

[2121] 3.实验者正在使用照相机记录动物行为,呆在帘子后面,以便不为动 物充当暗示或以便不向研究中引入无意识的偏差。

[2122] 4.计划的样品-物品暴露时间(5-10min)后,将该动物从仪器中移出 并返回聚居区,持续计划的训练至测试间隔。在动物之间将该仪器用70%乙 醇清理。

[2123] III延迟阶段。常用的训练至测试间隔从1h(用于稳健的物体识别)至 24h而变化。

[2124] IV物体识别测试:

[2125] 1.为了在训练至测试间隔后测试物体识别,将熟悉的样品物品之一放 在该仪器的后面角落中;将新颖的物体放在另一个后面角落中。

[2126] 2.如在步骤II.2中的,将动物放在该仪器中。实验者正在记录动物行 为,呆在帘子后面。

[2127] 3.计划的物品暴露时间(3-5min)后,将动物从仪器中移出并返回聚 居区。在动物之间将该仪器用70%乙醇清理。

[2128] 数据分析:使用记录的视频用于稍后的数据分析:a)定向接触评分(动 物与物体接触的总的持续时间);b)面积内(within area)评分(当动物的 鼻子与该物体接触或在2cm的最小距离内冲着该物体时,该动物与该物体 “相互作用”)。常用的量度包括熟悉物体对新颖物体的时间、得分差值(新颖物体相互作用-熟悉物体相互作用)、判别率(新颖物体相互作用/两个 物体的总的相互作用)。分别通过与新颖物体相互作用比熟悉物体多的时间,正的得分差值或判别率高于0.5来反映这些量度中的物体识别。

[2129] 巴恩斯迷宫。卡罗尔巴恩斯(Carol Barnes)在1979年研发了一种用于 空间学习和记忆的旱地迷宫测试,在该测试中,动物从灯火通明的暴露的圆 形开放平台表面逃出进入位于围绕该平台的周长的18个洞之一之下的小暗凹 室中。尽管最初它是针对大鼠而发明的,但是利用它们优越的发现并逃出小 洞的能力,巴恩斯迷宫(BM)用于评估小鼠的空间记忆已经变得越来越流 行。

[2130] 程序:

[2131] 1.适应。将小鼠放在迷宫中间的圆柱形黑色开始室中。已经过去10s之 后,移除室并且与强光一起打开蜂鸣器。小鼠渐渐地被引导进逃避箱中。一 旦小鼠在该箱内,则关闭蜂鸣器和光。小鼠应该在该逃避箱中呆2min。

[2132] 2.空间获取。将小鼠放在迷宫中间的圆柱形黑色开始室中。10s后,将该 室举起并打开峰鸣器和光。允许小鼠对该迷宫探索3分钟。当小鼠进入目标 通道或已经过去3min后,试验结束。在小鼠进入通道后,立即关闭峰鸣器和 光并允许小鼠在该通道中呆1min。如果小鼠在3分钟内未到达目标,则实验 者将它渐渐其引导进逃避箱中并将小鼠留在其中1min。小鼠呆在居住笼中, 直到下次试验。在接下来的3-4天过程中,将整个试验以15分钟间隔重复4 次。

[2133] 3.参照记忆试验。最后一次训练日后24h,执行探测试验。靶洞必须被 关闭。将动物放在迷宫中间的圆柱形黑色开始室下并且10s后,将该室移 除,打开蜂鸣器和光。90s后,将小鼠从迷宫中移出。完成探测试验,以便 确定该动物是否记住了靶目标位于哪里。测量每个洞中的捅数(误差)以及 到达实质上靶洞的等待时间和路径长度。

[2134] 实例37:SaCas9更新

[2135] 图69显示了AAV-Sa-Cas9载体、肝脏特异性AAV-Sa-Cas9载体、以及 备选的AAV-

Sa-Cas9载体。

[2136] 图70显示了优化的CMV-SaCas9-NLS-U6-sgRNA载体的数据(上次提交 的载体设计);新的数据比较了N'-末端对C'-末端标记的SaCas9并使用C'-末 端NLS标记显示了增强的切割效率。

[2137] 在小鼠Pcsk9基因的外显子4和5中选择了新的靶标(序列和位置参见 下文)。外显子4-5位于Pcsk9 N-末端前结构域区域的下游,当蛋白质成熟时 该区域被蛋白水解切割,并且此下游区域中的插入缺失预期导致蛋白质降 解。Pcsk9涉及LDL-受体的循环和负向调节,并且Pcsk9的损失应该间接导 致血浆胆固醇水平降低。

[2138] 将SgRNA克隆进AAV载体中并在单载体转染的Hepa1-6小鼠肝细胞系 中测试了插入缺失活性。通过Lipofectamine 2000将500ng的载体转染进 200,000个Hepa1-6细胞中并收集DNA用于SURVEYOR测定,显示出Pcsk9 的切割。

[2139] pAAV-CMV-SaCas9-NLS-U6-sgRNA(231:atctcttagataccagcatc)

[2140] pAAV-CMV-SaCas9-NLS-U6-sgRNA(232:tcaatctcccgatgggcacc)

[2141] pAAV-CMV-SaCas9-NLS-U6-sgRNA(233:gcccatcgggagattgaggg)

[2142] pAAV-CMV-SaCas9-NLS-U6-sgRNA(234:acttcaacagcgtgccggag)

[2143] pAAV-CMV-SaCas9-NLS-U6-sgRNA(235:ccgctgaccacacctgccag)

[2144] pAAV-CMV-SaCas9-NLS-U6-sgRNA(236:tggcaggtgtggtcagcggc)

[2145] 图71显示了示出通过新的Pcsk9靶标产生的插入缺失的SURVEYOR图 像。

[2146] 图72显示了SaCas9特异性:基于2个标准预测了全基因组脱靶位点 (GWOT):它们相对于预期SaCas9靶含有4个或更少的错配碱基,并且对 于SaCas9具有最小的限制性PAM(NNGRR)。用在靶位点(EMX1: TAGGGTTAGGGGCCCCAGGC)处具有它们的相应sgRNA的SpCas9或SaCas9转染HEK 293FT细胞,该靶位点具有CGGGGT作为PAM,使得它可 以被SpCas9(CGG)或被SaCas9(CGGGGT)切割。从细胞中收获DNA并 且通过Illumina测序分析在中靶和41个预测的脱靶座位处的插入缺失(遵循 来自徐(Hsu)等人的《自然生物技术》(Nature Biotech)2013方案和由大 卫·斯科特(David Scott)和乔什·温斯坦(Josh Weinstein)开发的数据分析流 程)。

[2147] 图73显示SaCas9在某些座位处比SpCas9可具有更高水平的脱靶活性。

[2148] 虽然已经在此示出并描述了本发明的优选实施例,但是对本领域的普通 技术人员而言应该显而易见是这样的实施例仅以举例方式提供。在不偏离本 发明的情况下众多变化、改变和取代现在将被本领域的普通技术人员想到。应该理解的是,在实践本发明中可以采用在此描述的本发明的实施例的不同 替代方案。

[2149] 参考文献

[2150] 班克G(Banker G)、戈斯林K(Goslin K).神经元细胞培养进展 (Developments in neuronal cell culture).《自然》(Nature).1988年11月10 日;336(6195):185-6。

[2151] 比德尔,V.M.(Bedell,V.M.)等人使用高效率TALEN系统进行体内基 因组编辑(In vivo genome editing using a high-efficiency TALEN system).《自 然》(Nature)491,114-U133(2012)。

[2152] 布哈雅,D.(Bhaya,D.)、戴维森,M.(Davison,M.)&巴郎镐,R. (Barrangou,R.)细菌和古细菌中的CRISPR-Cas系统:用于获得性防御和 调节的通用小RNA(CRISPR-Cas

systems in bacteria and archaea:versatile small RNAs for adaptive defense and regulation).《遗传学年评》(Annu Rev Genet)45,273-297(2011)。

[2153] 博维斯-沃佐维奇,S.(Bobis-Wozowicz,S.)、奥斯克,A.(Osiak, A.)、拉赫曼,S.H.(Rahman,S.H.)&凯思曼,T.(Cathomen,T.)使用锌 指核酸酶的多能干细胞中的靶向基因组编辑(Targeted genome editing in pluripotent stem cells using zinc-finger nucleases).《方法》(Methods)53, 339-346(2011)。

[2154] 博赫,J.(Boch,J.)等人打破TAL-III型效应器的DNA结合特异性的 密码(Breaking the code of DNA binding specificity of TAL-type III effectors).《科学》(Science)326,1509-1512(2009)。

[2155] 伯根哈根,D.F.(Bogenhagen,D.F.)&布朗,D.D.(Brown,D.D.)转 录终止所需的非洲蟾蜍5S DNA中的核苷酸序列(Nucleotide sequences in Xenopus 5S DNA required for transcription termination).《细胞》(Cell)24, 261-270(1981)。

[2156] 布尔特曼,S.(Bultmann,S.)等人通过组合设计者TALE和抑制表观遗 传修饰剂而靶向转录激活沉默的oct4多能性基因(Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers).《核酸研究》(Nucleic Acids Res)40, 5368-5377(2012)。

[2157] 卡尔松,D.F.(Carlson,D.F.)等人家畜中的有效的TALEN介导的基因 敲除(Efficient TALEN-mediated gene knockout in livestock).《美国国家科学 院院刊》(Proc Natl Acad Sci U S A)109,17382-17387(2012)。

[2158] 陈,F.Q.(Chen,F.Q.)等人使用ssDNA寡核苷酸用锌指核酸酶进行高 效基因组编辑(High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases).《自然方法》(Nat Methods)8,753-U796 (2011)。

[2159] 卓,S.W.(Cho,S.W.),金姆,S.(Kim,S.),金姆,J.M.(Kim, J.M.)&金姆,J.S.(Kim,J.S.)在人类细胞中用Cas9 RNA指导的内切核酸 酶进行靶向的基因组工程化(Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease).《自然生物技术》(Nat Biotechnol) 31,230-232(2013)。

[2160] 克里斯汀,M.(Christian,M.)等人用TAL效应器核酸酶靶向DNA双 链断裂(Targeting DNA double-strand breaks with TAL effector nucleases).《遗传学》(Genetics)186,757-761(2010)。

[2161] 丛,L.(Cong,L.)等人使用CRISPR-Cas系统的多重基因组工程化 (Multiplex genome engineering using CRISPR-Cas systems).《科学》 (Science)339,819-823(2013)。

[2162] 德尔特科娃,E.(Deltcheva,E.)等人通过反式编码的小RNA和宿主因 子RNA酶III熟化CRISPR RNA(CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III).《自然》(Nature)471,602-607 (2011)。

[2163] 德沃,H.(Deveau,H.)、加尔诺,J.E.(Garneau,J.E.)&莫万诺,S. (Moineau,S.)CRISPR-Cas系统及其在噬菌体细菌相互作用中的作用 (CRISPR-Cas system and its role in phage-bacteria interactions).《微生物学年 评》(Annu Rev Microbiol)64,475-493(2010)。

[2164] 丁,Q.(Ding,Q.)等人一种用于产生人类干细胞基的疾病模型的 TALEN基因组编辑系统(A TALEN genome-editing system for generating human stem cell-based disease models).《细胞与干细胞》(Cell Stem Cell) 12,238-251(2013)。

[2165] 加尔诺,J.E.(Garneau,J.E.)等人CRISPR-Cas细菌免疫系统切割噬菌 体和质粒DNA(The CRISPR-Cas bacterial immune system cleaves bacteriophage and plasmid DNA).《自然》(Nature)468,67-71(2010)。

[2166] 加斯诺斯,G.(Gasiunas,G.)、巴郎镐,R.(Barrangou,R.)、霍瓦 特,P.(Horvath,P.)&西克斯尼斯(Siksnys,V.)Cas9-crRNA核糖核蛋白 复合体在细菌中针对获得性免疫介导特异的DNA切割(Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria).《美国国家科学院院刊》(Proc Natl Acad Sci U S A)109, E2579-2586(2012)。

[2167] 格尤特思(Geurts,A.M.)等人经由胚胎微注射锌指核酸酶敲除大鼠 (Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases).《科学》 (Science)325,433-433(2009)。

[2168] 格雷SJ(Gray SJ)、福蒂SB(Foti SB)、施瓦兹JW(Schwartz JW)、帕夏博纳(Bachaboina L)、泰勒-布雷克B(Taylor-Blake B)、科尔 曼J(Coleman J)、埃莱尔MD(Ehlers MD)、泽尔卡MJ(Zylka MJ)、麦 克考恩TJ(McCown TJ)、萨穆尔斯基RJ(Samulski RJ).使用自互补载体 优化外周和中枢神经系统中的重组腺相关病毒介导的基因表达的启动子 (Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors).《人类基因治疗》(Hum Gene Ther.)2011年9月;22(9):1143-53.doi:10.1089/hum.2010.245。

[2169] 格斯钦,D.Y.(Guschin,D.Y.)等人一种用于监测内源基因修饰的快速 且通用的测定(A rapid and general assay for monitoring endogenous gene modification).《分子生物学方法》(Methods Mol Biol)649,247-256 (2010)。

[2170] 黑斯蒂,P.(Hasty,P.)、里维拉-佩雷斯,J.(Rivera-Perez,J.)&布 兰得利,A.(Bradley,A.)胚胎干细胞中的基因靶向所需的同源性长度(The length of homology required for gene targeting in embryonic stem cells).《分子 与细胞生物学》(Mol Cell Biol)11,5586-5591(1991)。

[2171] 霍瓦特,P.(Horvath,P.)&巴郎镐,R.(Barrangou,R.)CRISPR- Cas,细菌和古生菌的免疫系统(CRISPR-Cas,the immune system of bacteria and archaea).《科学》(Science)327,167-170(2010)。

[2172] 舒,P.D.(Hsu,P.D.)&张,F.(Zhang,F.)使用靶向的基因组工程技 术剖分神经功能(Dissecting neural function using targeted genome engineering technologies).《ACS化学神经科学》(ACS Chem Neurosci)3,603-610 (2012)。

[2173] 黄,W.Y.(Hwang,W.Y.)等人使用CRISPR-Cas系统在斑马鱼中进行 有效的基因组编辑(Efficient genome editing in zebrafish using a CRISPR-Cas system).《自然生物技术》(Nat Biotechnol)31,227-229(2013)。

[2174] 姜,W.(Jiang,W.)、碧卡尔德,D.(Bikard,D.)、考克斯,D. (Cox,D.)、张,F.

(Zhang,F.)&马拉夫尼L.A.(Marraffini,L.A.)使 用CRISPR-Cas系统的RNA指导的细菌基因组编辑(RNA-guided editing of bacterial genomes using CRISPR-Cas systems).《自然生物技术》(Nat Biotechnol)31,233-239(2013)。

[2175] 季聂克,M.(Jinek,M.)等人获得性免疫中的可编程的双RNA指导的 DNA内切核酸酶(A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity).《科学》(Science)337,816-821(2012)。

[2176] 季聂克,M.(Jinek,M.)等人人类细胞中的RNA编程的基因组编辑 (RNA-programmed genome editing in human cells),eLife 2,e00471 (2013)。

[2177] 卡普利特,M.G.(Kaplitt,M.G.)等人,用携带腺相关病毒(AAV)的 GAD基因对帕金森病进行基因治疗的安全性与可容许性:开放标签的I期试 验(Safety and tolerability of gene therapy with an adeno-associated virus(AAV) borne GAD gene for Parkinson's disease:an open label,phase I trial).《柳叶刀》(Lancet).2007年6月23日;369(9579):2097-105。

[2178] 莱维特N.(Levitt N.)布里格斯D.(Briggs D.)吉尔A.(Gil A.)普劳德 富特N.J.(Proudfoot N.J.)有效的合成poly(A)位点的定义(Definition of an efficient synthetic poly(A)site)《基因与发育》(Genes Dev.)1989;3:1019- 1025。

[2179] 刘D(Liu D)、费舍尔I(Fischer I).两个替代性启动子指导大鼠微管 相关蛋白质1B基因的神经元特异性表达(Two alternative promoters direct neuron-specific expression of the rat microtubule-associated protein 1B gene).(J Neurosci.)1996年8月15日;16(16):5026-36。

[2180] (Lopes,V.S.)等人,使用腺相关病毒用大MYO7A cDNA进行的视网 膜基因治疗(Retinal gene therapy with a large MYO7A cDNA using adeno- assocaited virus).《基因治疗》(Gene Ther),2013年1月24日,doi: 10.1038/gt 2013.3.[Epub优先出版]

[2181] 马赫福兹,M.M.(Mahfouz,M.M.)等人具有新颖的DNA结合特异性 的从头工程化的转录激活因子样效应器(TALE)杂合核酸酶产生双链断裂 (De novo-engineered transcription activator-like effector(TALE)hybrid nuclease with novel DNA binding specificity creates double-strand breaks).《美国国家科 学院院刊》(Proc Natl Acad Sci U S A)108,2623-2628(2011)。

[2182] 马卡洛夫,K.S.(Makarova,K.S.)等人CRISPR-Cas系统的进化与分类 (Evolution and classification of the CRISPR-Cas systems).《自然微生物学综 述》(Nat Rev Microbiol)9,467-477(2011)。

[2183] 玛丽,P.(Mali,P.)等人经由Cas9的RNA指导的人类基因组工程 (RNA-guided human genome engineering via Cas9).《科学》(Science) 339,823-826(2013)。

[2184] 麦克卢尔C(McClure C)、科尔KL(Cole KL)、伍尔夫P(Wulff P)、克卢格曼M(Klugmann M)、默里AJ(Murray AJ).重组腺相关病毒 载体的生产与滴定(Production and titering of recombinant adeno-associated viral vectors).J Vis Exp.2011年11月27日;(57):e3348.doi:10.3791/3348。

[2185] 米凯利斯,L.M.(Michaelis,L.M.),Maud"Die kinetik der invertinwirkung." Biochem.z(1913)。

[2186] 米勒,J.C.(Miller,J.C.)等人用于高度特异的基因组编辑的改进的锌指 核酸酶构造(An improved zinc-finger nuclease architecture for highly specific genome editing).《自然生物技术》(Nat Biotechnol)25,778-785 (2007)。

[2187] 米勒,J.C.(Miller,J.C.)等人用于有效基因组编辑的TALE核酸酶构造 (A TALE nuclease architecture for efficient genome editing).《自然生物技 术》(Nat Biotechnol)29,143-148(2011)。

[2188] 莫斯科,M.J.(Moscou,M.J.)&保达诺夫,A.J.(Bogdanove,A.J.)简 单的密码通过TAL效应器管理DNA识别(A simple cipher governs DNA recognition by TAL effectors).《科学》(Science)326,1501(2009)。波蒂 厄斯,M.H.(Porteus,M.H.)&巴尔的摩,D.(Baltimore,D.)嵌合的核酸 酶刺激人类细胞中的基因靶向(Chimeric nucleases stimulate gene targeting in human cells).《科学》(Science)300,763(2003)。

[2189] 墨索里诺,C.(Mussolino,C.)等人新颖的TALE核酸酶支架允许与较 低的毒性组合的较高的基因组编辑活性(A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity).《核酸研 究》(Nucleic acids research)39,9283-9293(2011)。

[2190] 纳斯瓦尼,A.C.(Nathwani,A.C.)等人,腺病毒相关病毒载体介导的B 型血友病的基因转移(Adenovirus-associated virus vector-mediated gene transfer in hemophilia B).《新英格兰医学杂志》(N Engl J Med.)2011年12月22 日;365(25):2357-65.doi:10.1056/NEJMoa1108046.Epub 2011年12月10 日。

[2191] 奥利韦拉,T.Y.(Oliveira,T.Y.)等人转位捕获测序:一种用于高通量 标绘染色体重排的方法(Translocation capture sequencing:a method for high throughput mapping of chromosomal rearrangements).《免疫学方法杂志》(J Immunol Methods)375,176-181(2012)。

[2192] 佩雷斯,E.E.(Perez,E.E.)等人使用锌指核酸酶通过基因组编辑在 CD4(+)T细胞中建立HIV-1抗性(Establishment of HIV-1resistance in CD4(+) T cells by genome editing using zinc-finger nucleases).《自然生物技术》(Nat Biotechnol)26,808-816(2008)。

[2193] 齐,L.S.(Qi,L.S.)等人将CRISPR再利用为用于基因表达的序列特异 性控制的RNA指导的平台(Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression).《细胞》(Cell)152,1173- 1183(2013)。

[2194] 雷明顿药物科学(REMINGTON'S PHARMACEUTICAL SCIENCES) (马克出版公司(Mack Pub.Co.),新泽西州1991)

[2195] 雷荣,D.(Reyon,D.)等人用于高通量基因组编辑的TALEN的 FLASH装配(FLASH assembly of TALENs for high-throughput genome editing).《自然生物技术》(Nat Biotechnol)30,460-465(2012)。

[2196] 萨利赫-德哈利,N.(Saleh-Gohari,N.)&荷雷德,T.(Helleday,T.) 保守同源重组优先修复人类细胞中的处于细胞周期的S期的DNA双链断裂 (Conservative homologous recombination preferentially repairs DNA double- strand breaks in the S phase of the cell cycle in human cells).《核酸研究》 (Nucleic Acids Res)32,3683-3688

(2004)。

[2197] 桑德尔,J.D.(Sander,J.D.)等人通过背景依赖性装配的无选择性锌指 核酸酶工程(CoDA)(Selection-free zinc-finger-nuclease engineering by context-dependent assembly(CoDA)).《自然方法》(Nat Methods)8,67-69 (2011)。

[2198] 萨亚纳,N.E.(Sanjana,N.E.)等人用于基因组工程的转录激活因子样 效应器工具箱(A transcription activator-like effector toolbox for genome engineering).《自然-实验手册》(Nat Protoc)7,171-192(2012)。

[2199] 萨普拉诺斯卡斯(Sapranauskas,R.)等人嗜热链球菌CRISPR-Cas系统 在大肠杆菌中提供免疫力(The Streptococcus thermophilus CRISPR-Cas system provides immunity in Escherichia coli).《核酸研究》(Nucleic Acids Res) 39,9275-9282(2011)。

[2200] 沈,B.(Shen,B.)等人经由Cas9/RNA介导的基因靶向产生基因修饰的 小鼠(Generation of gene-modified mice via Cas9/RNA-mediated gene targeting).《细胞研究》(Cell Res)23,720-723(2013)。

[2201] 斯米蒂斯,O.(Smithies,O.)、格雷格,R.G.(Gregg,R.G.)、博格 斯,S.S.(Boggs,S.S.)、科拉雷斯基,M.A.(Koralewski,M.A.)&库切尔 拉帕提,R.S.(Kucherlapati,R.S.)提供同源重组将DNA序列插入人类染色 体β-珠蛋白座位中(Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination).《自然》(Nature)317,230- 234(1985)。

[2202] 索德奈尔,F.(Soldner,F.)等人产生在两个早期发作的帕金森蛋白点 突变处排他地相异的同基因多能干细胞(Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations).《细胞》(Cell)146,318-331(2011)。

[2203] (Takasu,Y.)等人使用锌指核酸酶mRNA注射靶向诱变家蚕 (Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection).《昆虫生物化学与分子生物学》(Insect Biochem Molec) 40,759-765(2010)。

[2204] 腾格里S(Tangri S)等人,具有减少的免疫原性的合理工程化的治疗性 蛋白质(Rationally engineered therapeutic proteins with reduced immunogenicity),《免疫学杂志》(J Immunol.)2005年5月15日;174(6): 3187-96。

[2205] 托马斯,K.R.(Thomas,K.R.)、福尔杰,K.R.(Folger,K.R.)&卡佩 基,M.R.(Capecchi,M.R.)将基因高频靶向哺乳动物基因组中的特定位点 (High frequency targeting of genes to specific sites in the mammalian genome).《细胞》(Cell)44,419-428(1986)。

[2206] 杜施,T(Tuschl,T).扩展小RNA干扰(Expanding small RNA interference).《自然生物技术》(Nat Biotechnol)20,446-448(2002)。

[2207] 乌尔诺夫,F.D.(Urnov,F.D.)、雷巴尔,E.J.(Rebar,E.J.)、福尔摩 斯,M.C.(Holmes,M.C.)、张,H.S.(Zhang,H.S.)&格雷戈里,P.D. (Gregory,P.D.)用工程化的锌指核酸酶进行的基因组编辑(Genome editing with engineered zinc finger nucleases).《遗传学自然评论》(Nat Rev Genet) 11,636-646(2010)。

[2208] 沃尔顿,J.(Valton,J.)等人克服转录激活因子样效应器(TALE) DNA结合结构域对胞嘧啶甲基化的敏感性(Overcoming transcription activator- like effector (TALE)DNA binding domain sensitivity to cytosine methylation).《生物化学杂志》(J Biol Chem)287,38427-38432(2012)。

[2209] 王,H.(Wang,H.)等人通过CRISPR-Cas介导的基因组工程化一步产 生携带多个基因突变的小鼠(One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering).《细胞》(Cell)153,910-918(2013)。

[2210] 渡边,T.(Watanabe,T.)等人使用锌指和TAL效应器核酸酶在半变态 昆虫中进行非转基因基因组修饰(Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases).《自然通 讯》(Nat Commun)3(2012)。

[2211] 威尔逊,E.B.(Wilson,E.B.)可能的推断、继承法则与统计推断 (Probable inference,the law of succession,and statistical inference).《美国统计 协会杂志》(J Am Stat Assoc)22,209-212(1927)。

[2212] 伍德,A.J.(Wood,A.J.)等人使用ZFN和TALEN跨物种进行靶向的 基因组编辑(Targeted genome editing across species using ZFNs and TALENs).《科学》(Science)333,307(2011)。

[2213] 吴,S.(Wu,S.)、应,G.X.(Ying,G.X.)、吴,Q.(Wu,Q.)&卡 佩基,M.R.(Capecchi,M.R.)一种用于构建基因靶向载体的方案:产生钙 粘素家族及以外的基因敲除小鼠(A protocol for constructing gene targeting vectors:generating knockout mice for the cadherin family and beyond).《自然-实 验手册》(Nat Protoc)3,1056-1076(2008)。

[2214] 张,F.(Zhang,F.)等人有效构建用于调控哺乳动物转录的序列特异的 TAL效应器(Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription).《自然生物技术》(Nat Biotechnol) 29,149-153(2011)。

序列表

<110> 布罗德研究所有限公司
麻省理工学院
哈佛大学校长及研究院协会

<120> 用于序列操纵和治疗应用的系统、方法和组合物的递送、工程化和优化

<130> 44790.00.2041

<140> 14/104,837
<141> 2013-12-12

<150> 61/847,537
<151> 2013-07-17

<150> 61/836,123
<151> 2013-06-17

<150> 61/835,931
<151> 2013-06-17

[0001]

<150> 61/828,130
<151> 2013-05-28

<150> 61/819,803
<151> 2013-05-06

<150> 61/814,263
<151> 2013-04-20

<150> 61/806,375
<151> 2013-03-28

<150> 61/802,174
<151> 2013-03-15

<150> 61/791,409
<151> 2013-03-15

<150> 61/769,046
<151> 2013-02-25

<150> 61/758,468

<150> 61/748,427
<151> 2013-01-02

<150> 61/736,527
<151> 2012-12-12

<160> 443

<170> PatentIn version 3.5

<210> 1
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 1
[0002] gcactgaggg cctatttccc atgattc 27

<210> 2
<211> 20
<212> DNA
<213> 智人

<400> 2
tagggttagg ggccccaggc 20

<210> 3
<211> 20
<212> DNA
<213> 智人

<400> 3
gagtccgagc agaagaagaa 20

<210> 4
<211> 20
<212> DNA
<213> 智人

<400> 4
gagtcctagc aggagaagaa 20

<210> 5
<211> 20
<212> DNA
<213> 智人

<400> 5
gagtctaagc agaagaagaa 20

<210> 6
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
[0003] <223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(44)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (47)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 6
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnnnggnnnn nnnnnnnnnn 60

<210> 7
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(14)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (17)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 7
nnnnnnnnnn nnnnccnnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 8
<211> 60
<212> RNA
[0004] <213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(43)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (46)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 8
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnnggnnnnn nnnnnnnnnn 60

<210> 9
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(15)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (18)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
[0005] <223> a, c, u, g, 未知的或其他的

<400> 9
nnnnnnnnnn nnnnnccnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 10
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(42)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (45)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 10
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nnggnnnnnn nnnnnnnnnn 60

<210> 11
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(15)
<223> a, c, u, g, 未知的或其他的

[0006] <220>
<221> 修饰的碱基
<222> (18)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 11
nnnnnnnnnn nnnnnccnnn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 12
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(41)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (44)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 12
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn nggnnnnnnn nnnnnnnnnn 60

<210> 13
<211> 60
<212> RNA
<213> 人工序列

[0007]

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(17)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (20)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 13
nnnnnnnnnn nnnnnnnccn nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 14

<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(40)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (43)..(60)
<223> a, c, u, g, 未知的或其他的

[0008]

<400> 14
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnnn ggnnnnnnnn nnnnnnnnnn 60

<210> 15
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(18)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (21)..(38)
<223> a, c, u, g, 未知的或其他的

<220>

<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 15
nnnnnnnnnn nnnnnnnncc nnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 16
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
[0009] <221> 修饰的碱基
<222> (23)..(39)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (42)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 16
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnnng gnnnnnnnnn nnnnnnnnnn 60

<210> 17
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(19)

<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (22)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 17
nnnnnnnnnn nnnnnnnnnc cnnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 18
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0010]

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 18
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 19
<211> 60
<212> RNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 19

[0011] nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 20
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(37)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (40)..(60)

<223> a, c, u, g, 未知的或其他的

<400> 20
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnnggn nnnnnnnnnn nnnnnnnnnn 60

<210> 21
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(21)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (24)..(38)
[0012] <223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 21
nnnnnnnnnn nnnnnnnnnn nccnnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 22
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(36)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (39)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 22
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnnggnn nnnnnnnnnn nnnnnnnnnn 60

<210> 23
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0013] <220>
<221> 修饰的碱基
<222> (1)..(22)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (25)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 23
nnnnnnnnnn nnnnnnnnnn nnccnnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 24
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(35)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (38)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 24
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnnggnnn nnnnnnnnnn nnnnnnnnnn 60

[0014] <210> 25
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(23)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (26)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 25
nnnnnnnnnn nnnnnnnnnn nnnccnnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 26
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(34)
<223> a, c, u, g, 未知的或其他的

[0015]

<220>
<221> 修饰的碱基
<222> (37)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 26
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnnggnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 27
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(24)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基

<222> (27)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 27
nnnnnnnnnn nnnnnnnnnn nnnnccnnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 28
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
[0016] <223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(33)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (36)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 28
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnnggnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 29
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(25)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (28)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 29
nnnnnnnnnn nnnnnnnnnn nnnnnccnnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 30
<211> 60
<212> RNA
[0017] <213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(32)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (35)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 30
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nnggnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 31
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(26)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (29)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
[0018] <223> a, c, u, g, 未知的或其他的

<400> 31
nnnnnnnnnn nnnnnnnnnn nnnnnnccnn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 32
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(31)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (34)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 32
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn nggnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 33
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(27)
<223> a, c, u, g, 未知的或其他的

[0019] <220>
<221> 修饰的碱基
<222> (30)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 33
nnnnnnnnnn nnnnnnnnnn nnnnnnnccn nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 34
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>

<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(30)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (33)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 34
nnnnnnnnnn nnnnnnnnnn ccnnnnnnnn ggnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 35
<211> 60
<212> RNA
<213> 人工序列

[0020]

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(28)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (31)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 35
nnnnnnnnnn nnnnnnnnnn nnnnnnnncc nnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 36

<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(29)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (32)..(60)
<223> a, c, u, g, 未知的或其他的

[0021]

<400> 36
nnnnnnnnnn nnnnnnnnnn ccnnnnnnng gnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 37
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(29)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (32)..(38)
<223> a, c, u, g, 未知的或其他的

<220>

<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 37
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnc cnnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 38
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
[0022] <221> 修饰的碱基
<222> (23)..(28)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (31)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 38
nnnnnnnnnn nnnnnnnnnn ccnnnnnngg nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 39
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(30)

<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (33)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 39
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ccnnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 40
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0023]

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(27)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (30)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 40
nnnnnnnnnn nnnnnnnnnn ccnnnnnggn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 41
<211> 60
<212> RNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(31)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (34)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 41
[0024] nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nccnnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 42
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(26)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (29)..(60)

<223> a, c, u, g, 未知的或其他的

<400> 42
nnnnnnnnnn nnnnnnnnnn ccnnnnggnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 43
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(32)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (35)..(38)
[0025] <223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 43
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnccnnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 44
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(25)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (28)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 44
nnnnnnnnnn nnnnnnnnnn ccnnnggnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 45
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0026] <220>
<221> 修饰的碱基
<222> (1)..(33)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (36)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 45
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnccnnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 46
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(24)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (27)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 46
nnnnnnnnnn nnnnnnnnnn ccnnggnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

[0027] <210> 47
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(34)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (37)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 47
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnccnngg nnnnnnnnnn nnnnnnnnnn 60

<210> 48
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (23)..(23)
<223> a, c, u, g, 未知的或其他的

[0028]

<220>
<221> 修饰的碱基
<222> (26)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 48
nnnnnnnnnn nnnnnnnnnn ccnggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 49
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(35)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基

<222> (38)..(38)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (41)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 49
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnccngg nnnnnnnnnn nnnnnnnnnn 60

<210> 50
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(21)
[0029] <223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (26)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 50
nnnnnnnnnn nnnnnnnnnn nccggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 51
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(35)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (40)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 51
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnccggn nnnnnnnnnn nnnnnnnnnn 60

<210> 52
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(23)
<223> a, c, u, g, 未知的或其他的

[0030] <220>
<221> 修饰的碱基
<222> (26)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 52
nnnnnnnnnn nnnnnnnnnn nnnggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 53
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(33)
<223> a, c, u, g, 未知的或其他的

<220>

<221> 修饰的碱基
<222> (38)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 53
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnggccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 54
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(22)
<223> a, c, u, g, 未知的或其他的

<220>
[0031] <221> 修饰的碱基
<222> (26)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 54
nnnnnnnnnn nnnnnnnnnn nncggnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 55
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(32)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (35)..(35)

<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (38)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 55
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnggnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 56
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(31)
[0032] <223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (34)..(35)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (38)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 56
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nggnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 57
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(30)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (33)..(35)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (38)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 57
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn ggnnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 58
<211> 60
<212> RNA
[0033] <213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(29)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (32)..(35)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (38)..(60)
<223> a, c, u, g, 未知的或其他的

<400> 58
nnnnnnnnnn nnnnnnnnnn nnnnnnnnng gnnnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 59
<211> 60
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(28)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (31)..(35)
<223> a, c, u, g, 未知的或其他的

<220>
<221> 修饰的碱基
<222> (38)..(60)
<223> a, c, u, g, 未知的或其他的

[0034]

<400> 59
nnnnnnnnnn nnnnnnnnnn nnnnnnnngg nnnnnccnnn nnnnnnnnnn nnnnnnnnnn 60

<210> 60
<211> 9
<212> PRT
<213> 未知的

<220>
<223> 未知的的描述：'LAGLIDADG'家族基序肽

<400> 60
Leu Ala Gly Leu Ile Asp Ala Asp Gly
1 5

<210> 61
<211> 12
<212> RNA
<213> 人工序列

<220>

<223> 人工序列的描述：合成寡核苷酸

<400> 61
guuuuagagc ua 12

<210> 62
<211> 7
<212> PRT
<213> 猴病毒 40

<400> 62
Pro Lys Lys Lys Arg Lys Val
1 5

<210> 63
<211> 16
<212> PRT
<213> 未知的

<220>
<223> 未知的的描述：Nucleoplasmin 双向的 NLS 序列

[0035]

<400> 63
Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15

<210> 64
<211> 9
<212> PRT
<213> 未知的

<220>
<223> 未知的的描述：C-myc NLS 序列

<400> 64
Pro Ala Ala Lys Arg Val Lys Leu Asp
1 5

<210> 65
<211> 11
<212> PRT
<213> 未知的

<220>

<223> 未知的的描述：C-myc NLS 序列

<400> 65
Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro
1 5 10

<210> 66
<211> 38
<212> PRT
<213> 智人

<400> 66
Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly
1 5 10 15

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

Arg Asn Gln Gly Gly Tyr
35

[0036] <210> 67
<211> 42
<212> PRT
<213> 未知的

<220>
<223> 未知的的描述：来自输入蛋白-α 序列的 IBB 的序列

<400> 67
Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu
1 5 10 15

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

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

<210> 68
<211> 8
<212> PRT
<213> 未知的

<220>
<223> 未知的的描述：肌瘤 T 蛋白序列

<400> 68
Val Ser Arg Lys Arg Pro Arg Pro
1 5

<210> 69
<211> 8
<212> PRT
<213> 未知的

<220>
<223> 未知的的描述：肌瘤 T 蛋白序列

<400> 69
Pro Pro Lys Lys Ala Arg Glu Asp
1 5

<210> 70
<211> 8
<212> PRT
[0037] <213> 智人

<400> 70
Pro Gln Pro Lys Lys Lys Pro Leu
1 5

<210> 71
<211> 12
<212> PRT
<213> 小鼠

<400> 71
Ser Ala Leu Ile Lys Lys Lys Lys Lys Met Ala Pro
1 5 10

<210> 72
<211> 5
<212> PRT
<213> 流感病毒

<400> 72
Asp Arg Leu Arg Arg

1 5

<210> 73
<211> 7
<212> PRT
<213> 流感病毒

<400> 73
Pro Lys Gln Lys Lys Arg Lys
1 5

<210> 74
<211> 10
<212> PRT
<213> 肝炎D病毒

<400> 74
Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu
1 5 10

[0038]

<210> 75
<211> 10
<212> PRT
<213> 小鼠

<400> 75
Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg
1 5 10

<210> 76
<211> 20
<212> PRT
<213> 智人

<400> 76
Lys Arg Lys Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys
1 5 10 15

Lys Ser Lys Lys
20

<210> 77
<211> 17
<212> PRT

<213> 智人

<400> 77
Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys
1 5 10 15

Lys

<210> 78
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, t or g

[0039]

<220>
<221> 修饰的碱基
<222> (21)..(22)
<223> a, c, t, g, 未知的或其他的

<400> 78
nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> 79
<211> 19
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(12)
<223> a, c, t or g

<220>

<220>
<221> 修饰的碱基
<222> (13)..(14)
<223> a, c, t, g, 未知的或其他的

<400> 79
nnnnnnnnnn nnnnagaaw 19

<210> 80
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, t or g

<220>
<221> 修饰的碱基
<222> (21)..(22)
<223> a, c, t, g, 未知的或其他的

[0040]

<400> 80
nnnnnnnnnn nnnnnnnnnn nnagaaw 27

<210> 81
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(11)
<223> a, c, t or g

<220>
<221> 修饰的碱基
<222> (12)..(13)

<223> a, c, t, g, 未知的或其他的

<400> 81
nnnnnnnnnn nnnagaaw 18

<210> 82
<211> 137
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, t, g, 未知的或其他的

<400> 82
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcaagatt tagaaataaa tcttgcagaa 60

[0041] gctacaaaga taaggcttca tgccgaaatc aacaccctgt cattttatgg cagggtgttt 120

tcgttattta atttttt 137

<210> 83
<211> 123
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, t, g, 未知的或其他的

<400> 83
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgttttcgt tatttaattt 120

ttt 123

<210> 84
<211> 110
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, t, g, 未知的或其他的

<400> 84
nnnnnnnnnn nnnnnnnnnn gtttttgtac tctcagaaat gcagaagcta caaagataag 60

gcttcatgcc gaaatcaaca ccctgtcatt ttatggcagg gtgtttttt 110

<210> 85
<211> 102
<212> DNA
[0042] <213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, t, g, 未知的或其他的

<400> 85
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt ggcaccgagt cggtgctttt tt 102

<210> 86
<211> 88
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, t, g, 未知的或其他的

<400> 86
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcaac ttgaaaaagt gttttttt 88

<210> 87
<211> 76
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
[0043] <223> a, c, t, g, 未知的或其他的

<400> 87
nnnnnnnnnn nnnnnnnnnn gttttagagc tagaaatagc aagttaaaat aaggctagtc 60

cgttatcatt tttttt 76

<210> 88
<211> 12
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 88
gttttagagc ta 12

<210> 89
<211> 31
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 89
tagcaagtta aaataaggct agtccgtttt t 31

<210> 90
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(22)
<223> a, c, t, g, 未知的或其他的

<400> 90
nnnnnnnnnn nnnnnnnnnn nnagaaw 27

[0044]

<210> 91
<211> 33
<212> DNA
<213> 智人

<400> 91
ggacatcgat gtcacctcca atgactaggg tgg 33

<210> 92
<211> 33
<212> DNA
<213> 智人

<400> 92
cattggaggt gacatcgatg tcctccccat tgg 33

<210> 93
<211> 33
<212> DNA
<213> 智人

<400> 93
ggaagggcct gagtccgagc agaagaagaa ggg 33

<210> 94
<211> 33
<212> DNA
<213> 智人

<400> 94
ggtggcgaga ggggccgaga ttgggtgttc agg 33

<210> 95
<211> 33
<212> DNA
<213> 智人

<400> 95
atgcaggagg gtggcgagag gggccgagat tgg 33

<210> 96
<211> 32
<212> DNA
[0045] <213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 96
aaactctaga gagggcctat ttcccatgat tc 32

<210> 97
<211> 153
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 97
acctctagaa aaaaagcacc gactcggtgc actttttca agttgataac ggactagcct 60

tattttaact tgctatgctg ttttgtttcc aaaacagcat agctctaaaa cccctagtca 120

ttggaggtga cggtgtttcg tcctttccac aag 153

<210> 98
<211> 52
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 98
taatacgact cactatagga agtgcgccac catggcccca aagaagaagc gg 52

<210> 99
<211> 60
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 99
[0046] ggtttttttt ttttttttt ttttttttt ttttcttact ttttcttttt tgcctggccg 60

<210> 100
<211> 984
<212> PRT
<213> Campylobacter jejuni

<400> 100
Met Ala Arg Ile Leu Ala Phe Asp Ile Gly Ile Ser Ser Ile Gly Trp
1 5 10 15

Ala Phe Ser Glu Asn Asp Glu Leu Lys Asp Cys Gly Val Arg Ile Phe
20 25 30

Thr Lys Val Glu Asn Pro Lys Thr Gly Glu Ser Leu Ala Leu Pro Arg
35 40 45

Arg Leu Ala Arg Ser Ala Arg Lys Arg Leu Ala Arg Arg Lys Ala Arg
50 55 60

Leu Asn His Leu Lys His Leu Ile Ala Asn Glu Phe Lys Leu Asn Tyr
65 70 75 80

Glu Asp Tyr Gln Ser Phe Asp Glu Ser Leu Ala Lys Ala Tyr Lys Gly
85 90 95

Ser Leu Ile Ser Pro Tyr Glu Leu Arg Phe Arg Ala Leu Asn Glu Leu
100 105 110

Leu Ser Lys Gln Asp Phe Ala Arg Val Ile Leu His Ile Ala Lys Arg
115 120 125

Arg Gly Tyr Asp Asp Ile Lys Asn Ser Asp Asp Lys Glu Lys Gly Ala
130 135 140

Ile Leu Lys Ala Ile Lys Gln Asn Glu Glu Lys Leu Ala Asn Tyr Gln
145 150 155 160

Ser Val Gly Glu Tyr Leu Tyr Lys Glu Tyr Phe Gln Lys Phe Lys Glu
165 170 175

Asn Ser Lys Glu Phe Thr Asn Val Arg Asn Lys Lys Glu Ser Tyr Glu
180 185 190

[0047]

Arg Cys Ile Ala Gln Ser Phe Leu Lys Asp Glu Leu Lys Leu Ile Phe
195 200 205

Lys Lys Gln Arg Glu Phe Gly Phe Ser Phe Ser Lys Lys Phe Glu Glu
210 215 220

Glu Val Leu Ser Val Ala Phe Tyr Lys Arg Ala Leu Lys Asp Phe Ser
225 230 235 240

His Leu Val Gly Asn Cys Ser Phe Phe Thr Asp Glu Lys Arg Ala Pro
245 250 255

Lys Asn Ser Pro Leu Ala Phe Met Phe Val Ala Leu Thr Arg Ile Ile
260 265 270

Asn Leu Leu Asn Asn Leu Lys Asn Thr Glu Gly Ile Leu Tyr Thr Lys
275 280 285

Asp Asp Leu Asn Ala Leu Leu Asn Glu Val Leu Lys Asn Gly Thr Leu
290 295 300

Thr Tyr Lys Gln Thr Lys Lys Leu Leu Gly Leu Ser Asp Asp Tyr Glu
305 310 315 320

Phe Lys Gly Glu Lys Gly Thr Tyr Phe Ile Glu Phe Lys Lys Tyr Lys
325 330 335

Glu Phe Ile Lys Ala Leu Gly Glu His Asn Leu Ser Gln Asp Asp Leu
340 345 350

Asn Glu Ile Ala Lys Asp Ile Thr Leu Ile Lys Asp Glu Ile Lys Leu
355 360 365

Lys Lys Ala Leu Ala Lys Tyr Asp Leu Asn Gln Asn Gln Ile Asp Ser
370 375 380

Leu Ser Lys Leu Glu Phe Lys Asp His Leu Asn Ile Ser Phe Lys Ala
385 390 395 400

Leu Lys Leu Val Thr Pro Leu Met Leu Glu Gly Lys Lys Tyr Asp Glu
405 410 415

Ala Cys Asn Glu Leu Asn Leu Lys Val Ala Ile Asn Glu Asp Lys Lys
420 425 430

[0048]

Asp Phe Leu Pro Ala Phe Asn Glu Thr Tyr Tyr Lys Asp Glu Val Thr
435 440 445

Asn Pro Val Val Leu Arg Ala Ile Lys Glu Tyr Arg Lys Val Leu Asn
450 455 460

Ala Leu Leu Lys Lys Tyr Gly Lys Val His Lys Ile Asn Ile Glu Leu
465 470 475 480

Ala Arg Glu Val Gly Lys Asn His Ser Gln Arg Ala Lys Ile Glu Lys
485 490 495

Glu Gln Asn Glu Asn Tyr Lys Ala Lys Lys Asp Ala Glu Leu Glu Cys
500 505 510

Glu Lys Leu Gly Leu Lys Ile Asn Ser Lys Asn Ile Leu Lys Leu Arg
515 520 525

Leu Phe Lys Glu Gln Lys Glu Phe Cys Ala Tyr Ser Gly Glu Lys Ile
530 535 540

Lys Ile Ser Asp Leu Gln Asp Glu Lys Met Leu Glu Ile Asp His Ile
545 550 555 560

Tyr Pro Tyr Ser Arg Ser Phe Asp Asp Ser Tyr Met Asn Lys Val Leu
565 570 575

Val Phe Thr Lys Gln Asn Gln Glu Lys Leu Asn Gln Thr Pro Phe Glu
580 585 590

Ala Phe Gly Asn Asp Ser Ala Lys Trp Gln Lys Ile Glu Val Leu Ala
595 600 605

Lys Asn Leu Pro Thr Lys Lys Gln Lys Arg Ile Leu Asp Lys Asn Tyr
610 615 620

Lys Asp Lys Glu Gln Lys Asn Phe Lys Asp Arg Asn Leu Asn Asp Thr
625 630 635 640

Arg Tyr Ile Ala Arg Leu Val Leu Asn Tyr Thr Lys Asp Tyr Leu Asp
645 650 655

[0049]

Phe Leu Pro Leu Ser Asp Asp Glu Asn Thr Lys Leu Asn Asp Thr Gln
660 665 670

Lys Gly Ser Lys Val His Val Glu Ala Lys Ser Gly Met Leu Thr Ser
675 680 685

Ala Leu Arg His Thr Trp Gly Phe Ser Ala Lys Asp Arg Asn Asn His
690 695 700

Leu His His Ala Ile Asp Ala Val Ile Ile Ala Tyr Ala Asn Asn Ser
705 710 715 720

Ile Val Lys Ala Phe Ser Asp Phe Lys Lys Glu Gln Glu Ser Asn Ser
725 730 735

Ala Glu Leu Tyr Ala Lys Lys Ile Ser Glu Leu Asp Tyr Lys Asn Lys
740 745 750

Arg Lys Phe Phe Glu Pro Phe Ser Gly Phe Arg Gln Lys Val Leu Asp
755 760 765

Lys Ile Asp Glu Ile Phe Val Ser Lys Pro Glu Arg Lys Lys Pro Ser
770 775 780

Gly Ala Leu His Glu Glu Thr Phe Arg Lys Glu Glu Glu Phe Tyr Gln
785 790 795 800

Ser Tyr Gly Gly Lys Glu Gly Val Leu Lys Ala Leu Glu Leu Gly Lys
805 810 815

Ile Arg Lys Val Asn Gly Lys Ile Val Lys Asn Gly Asp Met Phe Arg
820 825 830

Val Asp Ile Phe Lys His Lys Lys Thr Asn Lys Phe Tyr Ala Val Pro
835 840 845

Ile Tyr Thr Met Asp Phe Ala Leu Lys Val Leu Pro Asn Lys Ala Val
850 855 860

Ala Arg Ser Lys Lys Gly Glu Ile Lys Asp Trp Ile Leu Met Asp Glu
865 870 875 880

Asn Tyr Glu Phe Cys Phe Ser Leu Tyr Lys Asp Ser Leu Ile Leu Ile
885 890 895

[0050] Gln Thr Lys Asp Met Gln Glu Pro Glu Phe Val Tyr Tyr Asn Ala Phe
900 905 910

Thr Ser Ser Thr Val Ser Leu Ile Val Ser Lys His Asp Asn Lys Phe
915 920 925

Glu Thr Leu Ser Lys Asn Gln Lys Ile Leu Phe Lys Asn Ala Asn Glu
930 935 940

Lys Glu Val Ile Ala Lys Ser Ile Gly Ile Gln Asn Leu Lys Val Phe
945 950 955 960

Glu Lys Tyr Ile Val Ser Ala Leu Gly Glu Val Thr Lys Ala Glu Phe
965 970 975

Arg Gln Arg Glu Asp Phe Lys Lys
980

<210> 101
<211> 91
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 101
tataatctca taagaaattt aaaaagggac taaaataaag agtttgcggg actctgcggg 60

gttacaatcc cctaaaaccg cttttaaaat t 91

<210> 102
<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 102
attttaccat aaagaaattt aaaaagggac taaaac 36

[0051] <210> 103
<211> 95
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<400> 103
nnnnnnnnnn nnnnnnnnnn guuuuagucc cgaaagggac uaaaauaaag aguuugcggg 60

acucugcggg guuacaaucc ccuaaaaccg cuuuu 95

<210> 104
<211> 1115
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成多肽

<400> 104

```
Met Ser Asp Leu Val Leu Gly Leu Asp Ile Gly Ile Gly Ser Val Gly
1               5                  10                  15

Val Gly Ile Leu Asn Lys Val Thr Gly Glu Ile Ile His Lys Asn Ser
            20                  25                  30

Arg Ile Phe Pro Ala Ala Gln Ala Glu Asn Asn Leu Val Arg Arg Thr
        35                  40                  45

Asn Arg Gln Gly Arg Arg Leu Ala Arg Arg Lys Lys His Arg Arg Val
    50                  55                  60

Arg Leu Asn Arg Leu Phe Glu Glu Ser Gly Leu Ile Thr Asp Phe Thr
65                  70                  75                  80

Lys Ile Ser Ile Asn Leu Asn Pro Tyr Gln Leu Arg Val Lys Gly Leu
                85                  90                  95
```

[0052]

```
Thr Asp Glu Leu Ser Asn Glu Glu Leu Phe Ile Ala Leu Lys Asn Met
            100                 105                 110

Val Lys His Arg Gly Ile Ser Tyr Leu Asp Asp Ala Ser Asp Asp Gly
        115                 120                 125

Asn Ser Ser Val Gly Asp Tyr Ala Gln Ile Val Lys Glu Asn Ser Lys
    130                 135                 140

Gln Leu Glu Thr Lys Thr Pro Gly Gln Ile Gln Leu Glu Arg Tyr Gln
145                 150                 155                 160

Thr Tyr Gly Gln Leu Arg Gly Asp Phe Thr Val Glu Lys Asp Gly Lys
                165                 170                 175

Lys His Arg Leu Ile Asn Val Phe Pro Thr Ser Ala Tyr Arg Ser Glu
            180                 185                 190

Ala Leu Arg Ile Leu Gln Thr Gln Gln Glu Phe Asn Pro Gln Ile Thr
        195                 200                 205

Asp Glu Phe Ile Asn Arg Tyr Leu Glu Ile Leu Thr Gly Lys Arg Lys
```

210 215 220

Tyr Tyr His Gly Pro Gly Asn Glu Lys Ser Arg Thr Asp Tyr Gly Arg
225 230 235 240

Tyr Arg Thr Ser Gly Glu Thr Leu Asp Asn Ile Phe Gly Ile Leu Ile
245 250 255

Gly Lys Cys Thr Phe Tyr Pro Asp Glu Phe Arg Ala Ala Lys Ala Ser
260 265 270

Tyr Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr
275 280 285

Val Pro Thr Glu Thr Lys Lys Leu Ser Lys Glu Gln Lys Asn Gln Ile
290 295 300

Ile Asn Tyr Val Lys Asn Glu Lys Ala Met Gly Pro Ala Lys Leu Phe
305 310 315 320

Lys Tyr Ile Ala Lys Leu Leu Ser Cys Asp Val Ala Asp Ile Lys Gly
[0053] 325 330 335

Tyr Arg Ile Asp Lys Ser Gly Lys Ala Glu Ile His Thr Phe Glu Ala
340 345 350

Tyr Arg Lys Met Lys Thr Leu Glu Thr Leu Asp Ile Glu Gln Met Asp
355 360 365

Arg Glu Thr Leu Asp Lys Leu Ala Tyr Val Leu Thr Leu Asn Thr Glu
370 375 380

Arg Glu Gly Ile Gln Glu Ala Leu Glu His Glu Phe Ala Asp Gly Ser
385 390 395 400

Phe Ser Gln Lys Gln Val Asp Glu Leu Val Gln Phe Arg Lys Ala Asn
405 410 415

Ser Ser Ile Phe Gly Lys Gly Trp His Asn Phe Ser Val Lys Leu Met
420 425 430

Met Glu Leu Ile Pro Glu Leu Tyr Glu Thr Ser Glu Glu Gln Met Thr
435 440 445

Ile Leu Thr Arg Leu Gly Lys Gln Lys Thr Thr Ser Ser Ser Asn Lys
450 455 460

Thr Lys Tyr Ile Asp Glu Lys Leu Leu Thr Glu Glu Ile Tyr Asn Pro
465 470 475 480

Val Val Ala Lys Ser Val Arg Gln Ala Ile Lys Ile Val Asn Ala Ala
485 490 495

Ile Lys Glu Tyr Gly Asp Phe Asp Asn Ile Val Ile Glu Met Ala Arg
500 505 510

Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met
515 520 525

Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys
530 535 540

Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu
545 550 555 560

[0054] Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp
565 570 575

Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser
580 585 590

Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp
595 600 605

Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys
610 615 620

Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr
625 630 635 640

Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser
645 650 655

Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg
660 665 670

Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr
675 680 685

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

Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr
705 710 715 720

Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu
725 730 735

Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu
740 745 750

Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met
755 760 765

Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe
770 775 780

Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
785 790 795 800

[0055] Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
805 810 815

Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys
820 825 830

Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln
835 840 845

Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp
850 855 860

Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly
865 870 875 880

Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val
885 890 895

Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly
900 905 910

Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe

915 920 925

Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys
930 935 940

Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met
945 950 955 960

Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro
965 970 975

Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu
980 985 990

Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln
995 1000 1005

His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe
1010 1015 1020

Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu
1025 1030 1035

[0056]

Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala
1040 1045 1050

Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro
1055 1060 1065

Ala Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr
1070 1075 1080

Thr Ser Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser
1085 1090 1095

Ile Thr Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly
1100 1105 1110

Gly Asp
1115

<210> 105
<211> 1374
<212> PRT

<213> 人工序列

<220>
<223> 人工序列的描述：合成多肽

<400> 105

[0057]

```
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 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

[0058] 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

[0059] 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 Thr Asn
755 760 765

Glu Asp Asp Glu Lys Lys Ala Ile Gln Lys Ile Gln Lys Ala Asn Lys
770 775 780

[0060] Asp Glu Lys Asp Ala Ala Met Leu Lys Ala Ala Asn Gln Tyr Asn Gly
785 790 795 800

Lys Ala Glu Leu Pro His Ser Val Phe His Gly His Lys Gln Leu Ala
805 810 815

Thr Lys Ile Arg Leu Trp His Gln Gln Gly Glu Arg Cys Leu Tyr Thr
820 825 830

Gly Lys Thr Ile Ser Ile His Asp Leu Ile Asn Asn Ser Asn Gln Phe
835 840 845

Glu Val Asp His Ile Leu Pro Leu Ser Ile Thr Phe Asp Asp Ser Leu
850 855 860

Ala Asn Lys Val Leu Val Tyr Ala Thr Ala Asn Gln Glu Lys Gly Gln
865 870 875 880

Arg Thr Pro Tyr Gln Ala Leu Asp Ser Met Asp Asp Ala Trp Ser Phe
885 890 895

Arg Glu Leu Lys Ala Phe Val Arg Glu Ser Lys Thr Leu Ser Asn Lys

900 905 910

Lys Lys Glu Tyr Leu Leu Thr Glu Glu Asp Ile Ser Lys Phe Asp Val
915 920 925

Arg Lys Lys Phe Ile Glu Arg Asn Leu Val Asp Thr Arg Tyr Ala Ser
930 935 940

Arg Val Val Leu Asn Ala Leu Gln Glu His Phe Arg Ala His Lys Ile
945 950 955 960

Asp Thr Lys Val Ser Val Val Arg Gly Gln Phe Thr Ser Gln Leu Arg
965 970 975

Arg His Trp Gly Ile Glu Lys Thr Arg Asp Thr Tyr His His His Ala
980 985 990

Val Asp Ala Leu Ile Ile Ala Ala Ser Ser Gln Leu Asn Leu Trp Lys
995 1000 1005

Lys Gln Lys Asn Thr Leu Val Ser Tyr Ser Glu Asp Gln Leu Leu
1010 1015 1020

[0061]

Asp Ile Glu Thr Gly Glu Leu Ile Ser Asp Asp Glu Tyr Lys Glu
1025 1030 1035

Ser Val Phe Lys Ala Pro Tyr Gln His Phe Val Asp Thr Leu Lys
1040 1045 1050

Ser Lys Glu Phe Glu Asp Ser Ile Leu Phe Ser Tyr Gln Val Asp
1055 1060 1065

Ser Lys Phe Asn Arg Lys Ile Ser Asp Ala Thr Ile Tyr Ala Thr
1070 1075 1080

Arg Gln Ala Lys Val Gly Lys Asp Lys Ala Asp Glu Thr Tyr Val
1085 1090 1095

Leu Gly Lys Ile Lys Asp Ile Tyr Thr Gln Asp Gly Tyr Asp Ala
1100 1105 1110

Phe Met Lys Ile Tyr Lys Lys Asp Lys Ser Lys Phe Leu Met Tyr
1115 1120 1125

Arg His Asp Pro Gln Thr Phe Glu Lys Val Ile Glu Pro Ile Leu
1130 1135 1140

Glu Asn Tyr Pro Asn Lys Gln Ile Asn Glu Lys Gly Lys Glu Val
1145 1150 1155

Pro Cys Asn Pro Phe Leu Lys Tyr Lys Glu Glu His Gly Tyr Ile
1160 1165 1170

Arg Lys Tyr Ser Lys Lys Gly Asn Gly Pro Glu Ile Lys Ser Leu
1175 1180 1185

Lys Tyr Tyr Asp Ser Lys Leu Gly Asn His Ile Asp Ile Thr Pro
1190 1195 1200

Lys Asp Ser Asn Asn Lys Val Val Leu Gln Ser Val Ser Pro Trp
1205 1210 1215

Arg Ala Asp Val Tyr Phe Asn Lys Thr Thr Gly Lys Tyr Glu Ile
1220 1225 1230

[0062]

Leu Gly Leu Lys Tyr Ala Asp Leu Gln Phe Glu Lys Gly Thr Gly
1235 1240 1245

Thr Tyr Lys Ile Ser Gln Glu Lys Tyr Asn Asp Ile Lys Lys Lys
1250 1255 1260

Glu Gly Val Asp Ser Asp Ser Glu Phe Lys Phe Thr Leu Tyr Lys
1265 1270 1275

Asn Asp Leu Leu Leu Val Lys Asp Thr Glu Thr Lys Glu Gln Gln
1280 1285 1290

Leu Phe Arg Phe Leu Ser Arg Thr Met Pro Lys Gln Lys His Tyr
1295 1300 1305

Val Glu Leu Lys Pro Tyr Asp Lys Gln Lys Phe Glu Gly Gly Glu
1310 1315 1320

Ala Leu Ile Lys Val Leu Gly Asn Val Ala Asn Ser Gly Gln Cys
1325 1330 1335

Lys Lys Gly Leu Gly Lys Ser Asn Ile Ser Ile Tyr Lys Val Arg
1340 1345 1350

Thr Asp Val Leu Gly Asn Gln His Ile Ile Lys Asn Glu Gly Asp
1355 1360 1365

Lys Pro Lys Leu Asp Phe
1370

<210> 106
<211> 15
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 106
Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
1 5 10 15

[0063]

<210> 107
<211> 15
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 107
Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys Glu Ala Ala Ala Lys
1 5 10 15

<210> 108
<211> 18
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 108
Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly Gly Ser Gly Gly Gly Gly
1 5 10 15

Gly Ser

<210> 109
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 109
gccaaattgg acgaccctcg cgg 23

<210> 110
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0064]

<400> 110
cgaggagacc cccgtttcgg tgg 23

<210> 111
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 111
cccgccgccg ccgtggctcg agg 23

<210> 112
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 112
tgagctctac gagatccaca agg 23

<210> 113
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 113
ctcaaaattc ataccggttg tgg 23

<210> 114
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0065]

<400> 114
cgttaaacaa caaccggact tgg 23

<210> 115
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 115
ttcaccccgc ggcgctgaat ggg 23

<210> 116
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 116
accactacca gtccgtccac agg 23

<210> 117
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 117
agcctttctg aacacatgca cgg 23

<210> 118
<211> 39
<212> DNA
<213> 智人

<400> 118
cctgccatca atgtggccat gcatgtgttc agaaaggct 39

[0066]

<210> 119
<211> 39
<212> DNA
<213> 智人

<400> 119
cctgccatca atgtggccgt gcatgtgttc agaaaggct 39

<210> 120
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 120
cactgcttaa gcctcgctcg agg 23

<210> 121
<211> 23

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 121
tcaccagcaa tattcgctcg agg 23

<210> 122
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 122
caccagcaat attccgctcg agg 23

[0067]

<210> 123
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 123
tagcaacaga catacgctcg agg 23

<210> 124
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 124
gggcagtagt aatacgctcg agg 23

<210> 125
<211> 23

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 125
ccaattccca tacattattg tac 23

<210> 126
<211> 4677
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 126
tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg 60

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg 120

[0068]

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga 180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag 240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc 300

acaacccgca aacatgtacc catacgatgt tccagattac gcttcgccga agaaaaagcg 360

caaggtcgaa gcgtccgaca agaagtacag catcggcctg gacatcggca ccaactctgt 420

gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc aagaaattca aggtgctggg 480

caacaccgac cggcacagca tcaagaagaa cctgatcgga gccctgctgt tcgacagcgg 540

cgaaacagcc gaggccaccc ggctgaagag aaccgccaga agaagataca ccagacggaa 600

gaaccggatc tgctatctgc aagagatctt cagcaacgag atggccaagg tggacgacag 660

cttcttccac agactggaag agtccttcct ggtggaagag gataagaagc acgagcggca 720

ccccatcttc ggcaacatcg tggacgaggt ggcctaccac gagaagtacc ccaccatcta 780

ccacctgaga aagaaactgg tggacagcac cgacaaggcc gacctgcggc tgatctatct 840

ggccctggcc cacatgatca agttccgggg ccacttcctg atcgagggcg acctgaaccc 900

cgacaacagc gacgtggaca agctgttcat ccagctggtg cagacctaca accagctgtt 960

cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag gccatcctgt ctgccagact 1020

gagcaagagc agacggctgg aaaatctgat cgcccagctg cccggcgaga agaagaatgg 1080

cctgttcggc aacctgattg ccctgagcct gggcctgacc cccaacttca agagcaactt 1140

cgacctggcc gaggatgcca aactgcagct gagcaaggac acctacgacg acgacctgga 1200

caacctgctg gcccagatcg gcgaccagta cgccgacctg tttctggccg ccaagaacct 1260

gtccgacgcc atcctgctga gcgacatcct gagagtgaac accgagatca ccaaggcccc 1320

cctgagcgcc tctatgatca agagatacga cgagcaccac caggacctga ccctgctgaa 1380

agctctcgtg cggcagcagc tgcctgagaa gtacaaagag attttcttcg accagagcaa 1440

[0069]

gaacggctac gccggctaca ttgacggcgg agccagccag gaagagttct acaagttcat 1500

caagcccatc ctggaaaaga tggacggcac cgaggaactg ctcgtgaagc tgaacagaga 1560

ggacctgctg cggaagcagc ggaccttcga caacggcagc atcccccacc agatccacct 1620

gggagagctg cacgccattc tgcggcggca ggaagatttt tacccattcc tgaaggacaa 1680

ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc tactacgtgg gccctctggc 1740

caggggaaac agcagattcg cctggatgac cagaaagagc gaggaaacca tcacccccctg 1800

gaacttcgag gaagtggtgg acaagggcgc ttccgcccag agcttcatcg agcggatgac 1860

caacttcgat aagaacctgc ccaacgagaa ggtgctgccc aagcacagcc tgctgtacga 1920

gtacttcacc gtgtataacg agctgaccaa agtgaaatac gtgaccgagg gaatgagaaa 1980

gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg gacctgctgt tcaagaccaa 2040

ccggaaagtg accgtgaagc agctgaaaga ggactacttc aagaaaatcg agtgcttcga 2100

ctccgtggaa atctccggcg tggaagatcg gttcaacgcc tccctgggca cataccacga 2160

tctgctgaaa attatcaagg acaaggactt cctggacaat gaggaaaacg aggacattct 2220

ggaagatatc gtgctgaccc tgacactgtt tgaggacaga gagatgatcg aggaacggct 2280

gaaaacctat gcccacctgt tcgacgacaa agtgatgaag cagctgaagc ggcggagata 2340

caccggctgg ggcaggctga gccggaagct gatcaacggc atccgggaca agcagtccgg 2400

caagacaatc ctggatttcc tgaagtccga cggcttcgcc aacagaaact tcatgcagct 2460

gatccacgac gacagcctga cctttaaaga ggacatccag aaagcccagg tgtccggcca 2520

gggcgatagc ctgcacgagc acattgccaa tctggccggc agccccgcca ttaagaaggg 2580

catcctgcag acagtgaagg tggtggacga gctcgtgaaa gtgatgggcc ggcacaagcc 2640

cgagaacatc gtgatcgaaa tggccagaga gaaccagacc acccagaagg gacagaagaa 2700

cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa gagctgggca gccagatcct 2760

[0070]

gaaagaacac cccgtggaaa acacccagct gcagaacgag aagctgtacc tgtactacct 2820

gcagaatggg cgggatatgt acgtggacca ggaactggac atcaaccggc tgtccgacta 2880

cgatgtggac catatcgtgc ctcagagctt tctgaaggac gactccatcg acaacaaggt 2940

gctgaccaga agcgacaaga accggggcaa gagcgacaac gtgccctccg aagaggtcgt 3000

gaagaagatg aagaactact ggcggcagct gctgaacgcc aagctgatta cccagagaaa 3060

gttcgacaat ctgaccaagg ccgagagagg cggcctgagc gaactggata aggccggctt 3120

catcaagaga cagctggtgg aaacccggca gatcacaaag cacgtggcac agatcctgga 3180

ctcccggatg aacactaagt acgacgagaa tgacaagctg atccgggaag tgaaagtgat 3240

caccctgaag tccaagctgg tgtccgattt ccggaaggat ttccagtttt acaaagtgcg 3300

cgagatcaac aactaccacc acgcccacga cgcctacctg aacgccgtcg tgggaaccgc 3360

cctgatcaaa aagtacccta agctggaaag cgagttcgtg tacggcgact acaaggtgta 3420

cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc ggcaaggcta ccgccaagta 3480

cttcttctac agcaacatca tgaacttttt caagaccgag attaccctgg ccaacggcga 3540

gatccggaag cggcctctga tcgagacaaa cggcgaaacc ggggagatcg tgtgggataa 3600

gggccgggat tttgccaccg tgcggaaagt gctgagcatg ccccaagtga atatcgtgaa 3660

aaagaccgag gtgcagacag gcggcttcag caaagagtct atcctgccca agaggaacag 3720

cgataagctg atcgccagaa agaaggactg ggaccctaag aagtacggcg gcttcgacag 3780

ccccaccgtg gcctattctg tgctggtggt ggccaaagtg gaaaagggca agtccaagaa 3840

actgaagagt gtgaaagagc tgctggggat caccatcatg gaaagaagca gcttcgagaa 3900

gaatcccatc gactttctgg aagccaaggg ctacaaagaa gtgaaaaagg acctgatcat 3960

caagctgcct aagtactccc tgttcgagct ggaaaacggc cggaagagaa tgctggcctc 4020

tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc tccaaatatg tgaacttcct 4080

[0071] gtacctggcc agccactatg agaagctgaa gggctccccc gaggataatg agcagaaaca 4140

gctgtttgtg gaacagcaca agcactacct ggacgagatc atcgagcaga tcagcgagtt 4200

ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa gtgctgtccg cctacaacaa 4260

gcaccgggat aagcccatca gagagcaggc cgagaatatc atccacctgt ttaccctgac 4320

caatctggga gcccctgccg ccttcaagta ctttgacacc accatcgacc ggaagaggta 4380

caccagcacc aaagaggtgc tggacgccac cctgatccac cagagcatca ccggcctgta 4440

cgagacacgg atcgacctgt ctcagctggg aggcgacagc cccaagaaga agagaaaggt 4500

ggaggccagc taaggatccg gcaagactgg ccccgcttgg caacgcaaca gtgagcccct 4560

ccctagtgtg tttggggatg tgactatgta ttcgtgtgtt ggccaacggg tcaacccgaa 4620

cagattgata cccgccttgg catttcctgt cagaatgtaa cgtcagttga tggtact 4677

<210> 127
<211> 3150

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 127

[0072]

tctttcttgc gctatgacac ttccagcaaa aggtagggcg ggctgcgaga cggcttcccg 60

gcgctgcatg caacaccgat gatgcttcga ccccccgaag ctccttcggg gctgcatggg 120

cgctccgatg ccgctccagg gcgagcgctg tttaaatagc caggcccccg attgcaaaga 180

cattatagcg agctaccaaa gccatattca aacacctaga tcactaccac ttctacacag 240

gccactcgag cttgtgatcg cactccgcta agggggcgcc tcttcctctt cgtttcagtc 300

acaacccgca aacatgccta agaagaagag gaaggttaac acgattaaca tcgctaagaa 360

cgacttctct gacatcgaac tggctgctat cccgttcaac actctggctg accattacgg 420

tgagcgttta gctcgcgaac agttggccct tgagcatgag tcttacgaga tgggtgaagc 480

acgcttccgc aagatgtttg agcgtcaact taaagctggt gaggttgcgg ataacgctgc 540

cgccaagcct ctcatcacta ccctactccc taagatgatt gcacgcatca acgactggtt 600

tgaggaagtg aaagctaagc gcggcaagcg cccgacagcc ttccagttcc tgcaagaaat 660

caagccggaa gccgtagcgt acatcaccat taagaccact ctggcttgcc taaccagtgc 720

tgacaataca accgttcagg ctgtagcaag cgcaatcggt cgggccattg aggacgaggc 780

tcgcttcggt cgtatccgtg accttgaagc taagcacttc aagaaaaacg ttgaggaaca 840

actcaacaag cgcgtagggc acgtctacaa gaaagcattt atgcaagttg tcgaggctga 900

catgctctct aagggtctac tcggtggcga ggcgtggtct tcgtggcata aggaagactc 960

tattcatgta ggagtacgct gcatcgagat gctcattgag tcaaccggaa tggttagctt 1020

acaccgccaa aatgctggcg tagtaggtca agactctgag actatcgaac tcgcacctga 1080

atacgctgag gctatcgcaa cccgtgcagg tgcgctggct ggcatctctc cgatgttcca 1140

accttgcgta gttcctccta agccgtggac tggcattact ggtggtggct attgggctaa 1200

cggtcgtcgt cctctggcgc tggtgcgtac tcacagtaag aaagcactga tgcgctacga 1260

agacgtttac atgcctgagg tgtacaaagc gattaacatt gcgcaaaaca ccgcatggaa 1320

aatcaacaag aaagtcctag cggtcgccaa cgtaatcacc aagtggaagc attgtccggt 1380

cgaggacatc cctgcgattg agcgtgaaga actcccgatg aaaccggaag acatcgacat 1440

gaatcctgag gctctcaccg cgtggaaacg tgctgccgct gctgtgtacc gcaaggacaa 1500

ggctcgcaag tctcgccgta tcagccttga gttcatgctt gagcaagcca ataagtttgc 1560

taaccataag gccatctggt tcccttacaa catggactgg cgcggtcgtg tttacgctgt 1620

gtcaatgttc aacccgcaag gtaacgatat gaccaaagga ctgcttacgc tggcgaaagg 1680

taaaccaatc ggtaaggaag gttactactg gctgaaaatc cacggtgcaa actgtgcggg 1740

tgtcgacaag gttccgttcc ctgagcgcat caagttcatt gaggaaaacc acgagaacat 1800

[0073]

catggcttgc gctaagtctc cactggagaa cacttggtgg gctgagcaag attctccgtt 1860

ctgcttcctt gcgttctgct ttgagtacgc tggggtacag caccacggcc tgagctataa 1920

ctgctccctt ccgctggcgt ttgacgggtc ttgctctggc atccagcact tctccgcgat 1980

gctccgagat gaggtaggtg gtcgcgcggt taacttgctt cctagtgaaa ccgttcagga 2040

catctacggg attgttgcta agaaagtcaa cgagattcta caagcagacg caatcaatgg 2100

gaccgataac gaagtagtta ccgtgaccga tgagaacact ggtgaaatct ctgagaaagt 2160

caagctgggc actaaggcac tggctggtca atggctggct tacggtgtta ctcgcagtgt 2220

gactaagcgt tcagtcatga cgctggctta cgggtccaaa gagttcggct tccgtcaaca 2280

agtgctggaa gataccattc agccagctat tgattccggc aagggtctga tgttcactca 2340

gccgaatcag gctgctggat acatggctaa gctgatttgg gaatctgtga gcgtgacggt 2400

ggtagctgcg gttgaagcaa tgaactggct taagtctgct gctaagctgc tggctgctga 2460

ggtcaaagat aagaagactg gagagattct tcgcaagcgt tgcgctgtgc attgggtaac 2520

tcctgatggt ttccctgtgt ggcaggaata caagaagcct attcagacgc gcttgaacct 2580

gatgttcctc ggtcagttcc gcttacagcc taccattaac accaacaaag atagcgagat 2640

tgatgcacac aaacaggagt ctggtatcgc tcctaacttt gtacacagcc aagacggtag 2700

ccaccttcgt aagactgtag tgtgggcaca cgagaagtac ggaatcgaat cttttgcact 2760

gattcacgac tccttcggta cgattccggc tgacgctgcg aacctgttca aagcagtgcg 2820

cgaaactatg gttgacacat atgagtcttg tgatgtactg gctgatttct acgaccagtt 2880

cgctgaccag ttgcacgagt ctcaattgga caaaatgcca gcacttccgg ctaaaggtaa 2940

cttgaacctc cgtgacatct tagagtcgga cttcgcgttc gcgtaaggat ccggcaagac 3000

tggccccgct tggcaacgca acagtgagcc cctccctagt gtgtttgggg atgtgactat 3060

gtattcgtgt gttggccaac gggtcaaccc gaacagattg atacccgcct tggcatttcc 3120

[0074]

tgtcagaatg taacgtcagt tgatggtact 3150

<210> 128
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<220>
<221> 修饰的碱基
<222> (23)..(42)
<223> a, c, t, g, 未知的或其他的

<400> 128
gaaattaata cgactcacta tannnnnnnn nnnnnnnnnn nngttttaga gctagaaata 60

gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgctt 120

ttttt 125

<210> 129
<211> 8452
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 129
tgcggtattt cacaccgcat caggtggcac ttttcgggga aatgtgcgcg gaacccctat 60

ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagattat caaaaaggat 120

cttcacctag atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga 180

gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg 240

tctatttcgt tcatccatag ttgcctgact ccccgtcgtg tagataacta cgatacggga 300

gggcttacca tctggcccca gtgctgcaat gataccgcga gacccacgct caccggctcc 360

[0075] agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac 420

tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 480

agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc 540

gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 600

catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttgtca gaagtaagtt 660

ggccgcagtg ttatcactca tggttatggc agcactgcat aattctctta ctgtcatgcc 720

atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 780

tatgcggcga ccgagttgct cttgcccggc gtcaatacgg gataataccg cgccacatag 840

cagaacttta aaagtgctca tcattggaaa acgttcttcg gggcgaaaac tctcaaggat 900

cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 960

atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 1020

aaagggaata agggcgacac ggaaatgttg aatactcata ctcttccttt ttcaatatta 1080

ttgaagcatt tatcagggtt attgtctcat gaccaaaatc ccttaacgtg agttttcgtt 1140

ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct 1200

gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc 1260

ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc 1320

aaatactgtt cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc 1380

gcctacatac ctcgctctgc taatcctgtt accagtggct gttgccagtg gcgataagtc 1440

gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 1500

aacgggggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata 1560

cctacagcgt gagctatgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta 1620

tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc 1680

[0076]

ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 1740

atgctcgtca ggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 1800

cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt 1860

ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 1920

gcgcagcgag tcagtgagcg aggaagcggt cgctgaggct tgacatgatt ggtgcgtatg 1980

tttgtatgaa gctacaggac tgatttggcg ggctatgagg gcgggggaag ctctggaagg 2040

gccgcgatgg ggcgcgcggc gtccagaagg cgccatacgg cccgctggcg gcacccatcc 2100

ggtataaaag cccgcgaccc cgaacggtga cctccacttt cagcgacaaa cgagcactta 2160

tacatacgcg actattctgc cgctatacat aaccactcag ctagcttaag atcccatcaa 2220

gcttgcatgc cgggcgcgcc agaaggagcg cagccaaacc aggatgatgt ttgatggggt 2280

atttgagcac ttgcaaccct tatccggaag ccccctggcc cacaaaggct aggcgccaat 2340

gcaagcagtt cgcatgcagc ccctggagcg gtgccctcct gataaaccgg ccaggggcc 2400

tatgttcttt acttttttac aagagaagtc actcaacatc ttaaaatggc caggtgagtc 2460

gacgagcaag cccggcggat caggcagcgt gcttgcagat ttgacttgca acgcccgcat 2520

tgtgtcgacg aaggcttttg gctcctctgt cgctgtctca agcagcatct aaccctgcgt 2580

cgccgtttcc atttgcagga gattcgaggt accatgtacc catacgatgt tccagattac 2640

gcttcgccga agaaaaagcg caaggtcgaa gcgtccgaca agaagtacag catcggcctg 2700

gacatcggca ccaactctgt gggctgggcc gtgatcaccg acgagtacaa ggtgcccagc 2760

aagaaattca aggtgctggg caacaccgac cggcacagca tcaagaagaa cctgatcgga 2820

gccctgctgt tcgacagcgg cgaaacagcc gaggccaccc ggctgaagag aaccgccaga 2880

agaagataca ccagacggaa gaaccggatc tgctatctgc aagagatctt cagcaacgag 2940

atggccaagg tggacgacag cttcttccac agactggaag agtccttcct ggtggaagag 3000

[0077]

gataagaagc acgagcggca ccccatcttc ggcaacatcg tggacgaggt ggcctaccac 3060

gagaagtacc ccaccatcta ccacctgaga aagaaactgg tggacagcac cgacaaggcc 3120

gacctgcggc tgatctatct ggccctggcc cacatgatca agttccgggg ccacttcctg 3180

atcgagggcg acctgaaccc cgacaacagc gacgtggaca agctgttcat ccagctggtg 3240

cagacctaca accagctgtt cgaggaaaac cccatcaacg ccagcggcgt ggacgccaag 3300

gccatcctgt ctgccagact gagcaagagc agacggctgg aaaatctgat cgcccagctg 3360

cccggcgaga agaagaatgg cctgttcggc aacctgattg ccctgagcct gggcctgacc 3420

cccaacttca agagcaactt cgacctggcc gaggatgcca aactgcagct gagcaaggac 3480

acctacgacg acgacctgga caacctgctg gcccagatcg gcgaccagta cgccgacctg 3540

tttctggccg ccaagaacct gtccgacgcc atcctgctga gcgacatcct gagagtgaac 3600

accgagatca ccaaggcccc cctgagcgcc tctatgatca agagatacga cgagcaccac 3660

caggacctga ccctgctgaa agctctcgtg cggcagcagc tgcctgagaa gtacaaagag 3720

attttcttcg accagagcaa gaacggctac gccggctaca ttgacggcgg agccagccag 3780

gaagagttct acaagttcat caagcccatc ctggaaaaga tggacggcac cgaggaactg 3840

ctcgtgaagc tgaacagaga ggacctgctg cggaagcagc ggaccttcga caacggcagc 3900

atcccccacc agatccacct gggagagctg cacgccattc tgcggcggca ggaagatttt 3960

tacccattcc tgaaggacaa ccgggaaaag atcgagaaga tcctgacctt ccgcatcccc 4020

tactacgtgg gccctctggc caggggaaac agcagattcg cctggatgac cagaaagagc 4080

gaggaaacca tcacccccctg gaacttcgag gaagtggtgg acaagggcgc ttccgcccag 4140

agcttcatcg agcggatgac caacttcgat aagaacctgc ccaacgagaa ggtgctgccc 4200

aagcacagcc tgctgtacga gtacttcacc gtgtataacg agctgaccaa agtgaaatac 4260

gtgaccgagg gaatgagaaa gcccgccttc ctgagcggcg agcagaaaaa ggccatcgtg 4320

[0078]

gacctgctgt tcaagaccaa ccggaaagtg accgtgaagc agctgaaaga ggactacttc 4380

aagaaaatcg agtgcttcga ctccgtggaa atctccggcg tggaagatcg gttcaacgcc 4440

tccctgggca cataccacga tctgctgaaa attatcaagg acaaggactt cctggacaat 4500

gaggaaaacg aggacattct ggaagatatc gtgctgaccc tgacactgtt tgaggacaga 4560

gagatgatcg aggaacggct gaaaacctat gcccacctgt tcgacgacaa agtgatgaag 4620

cagctgaagc ggcggagata caccggctgg ggcaggctga gccggaagct gatcaacggc 4680

atccgggaca agcagtccgg caagacaatc ctggatttcc tgaagtccga cggcttcgcc 4740

aacagaaact tcatgcagct gatccacgac gacagcctga cctttaaaga ggacatccag 4800

aaagcccagg tgtccggcca gggcgatagc ctgcacgagc acattgccaa tctggccggc 4860

agccccgcca ttaagaaggg catcctgcag acagtgaagg tggtggacga gctcgtgaaa 4920

gtgatgggcc ggcacaagcc cgagaacatc gtgatcgaaa tggccagaga gaaccagacc 4980

acccagaagg gacagaagaa cagccgcgag agaatgaagc ggatcgaaga gggcatcaaa 5040

gagctgggca gccagatcct gaaagaacac cccgtggaaa acacccagct gcagaacgag 5100

aagctgtacc tgtactacct gcagaatggg cgggatatgt acgtggacca ggaactggac 5160

atcaaccggc tgtccgacta cgatgtggac catatcgtgc ctcagagctt tctgaaggac 5220

gactccatcg acaacaaggt gctgaccaga agcgacaaga accggggcaa gagcgacaac 5280

gtgccctccg aagaggtcgt gaagaagatg aagaactact ggcggcagct gctgaacgcc 5340

aagctgatta cccagagaaa gttcgacaat ctgaccaagg ccgagagagg cggcctgagc 5400

gaactggata aggccggctt catcaagaga cagctggtgg aaacccggca gatcacaaag 5460

cacgtggcac agatcctgga ctcccggatg aacactaagt acgacgagaa tgacaagctg 5520

atccgggaag tgaaagtgat caccctgaag tccaagctgg tgtccgattt ccggaaggat 5580

ttccagtttt acaaagtgcg cgagatcaac aactaccacc acgcccacga cgcctacctg 5640

[0079]

aacgccgtcg tgggaaccgc cctgatcaaa aagtacccta agctggaaag cgagttcgtg 5700

tacggcgact acaaggtgta cgacgtgcgg aagatgatcg ccaagagcga gcaggaaatc 5760

ggcaaggcta ccgccaagta cttcttctac agcaacatca tgaacttttt caagaccgag 5820

attaccctgg ccaacggcga gatccggaag cggcctctga tcgagacaaa cggcgaaacc 5880

ggggagatcg tgtgggataa gggccgggat tttgccaccg tgcggaaagt gctgagcatg 5940

ccccaagtga atatcgtgaa aaagaccgag gtgcagacag gcggcttcag caaagagtct 6000

atcctgccca agaggaacag cgataagctg atcgccagaa agaaggactg ggaccctaag 6060

aagtacggcg gcttcgacag ccccaccgtg gcctattctg tgctggtggt ggccaaagtg 6120

gaaaagggca agtccaagaa actgaagagt gtgaaagagc tgctggggat caccatcatg 6180

gaaagaagca gcttcgagaa gaatcccatc gactttctgg aagccaaggg ctacaaagaa 6240

gtgaaaaagg acctgatcat caagctgcct aagtactccc tgttcgagct ggaaaacggc 6300

cggaagagaa tgctggcctc tgccggcgaa ctgcagaagg gaaacgaact ggccctgccc 6360

tccaaatatg tgaacttcct gtacctggcc agccactatg agaagctgaa gggctccccc 6420

gaggataatg agcagaaaca gctgtttgtg gaacagcaca agcactacct ggacgagatc 6480

atcgagcaga tcagcgagtt ctccaagaga gtgatcctgg ccgacgctaa tctggacaaa 6540

gtgctgtccg cctacaacaa gcaccgggat aagcccatca gagagcaggc cgagaatatc 6600

atccacctgt ttaccctgac caatctggga gcccctgccg ccttcaagta ctttgacacc 6660

accatcgacc ggaagaggta caccagcacc aaagaggtgc tggacgccac cctgatccac 6720

cagagcatca ccggcctgta cgagacacgg atcgacctgt ctcagctggg aggcgacagc 6780

cccaagaaga agagaaaggt ggaggccagc taacatatga ttcgaatgtc tttcttgcgc 6840

tatgacactt ccagcaaaag gtagggcggg ctgcgagacg gcttcccggc gctgcatgca 6900

acaccgatga tgcttcgacc ccccgaagct ccttcggggc tgcatgggcg ctccgatgcc 6960

[0080]

gctccagggc gagcgctgtt taaatagcca ggcccccgat tgcaaagaca ttatagcgag 7020

ctaccaaagc catattcaaa cacctagatc actaccactt ctacacaggc cactcgagct 7080

tgtgatcgca ctccgctaag ggggcgcctc ttcctcttcg tttcagtcac aacccgcaaa 7140

catgacacaa gaatccctgt tacttctcga ccgtattgat tcggatgatt cctacgcgag 7200

cctgcggaac gaccaggaat tctgggaggt gagtcgacga gcaagcccgg cggatcaggc 7260

agcgtgcttg cagatttgac ttgcaacgcc cgcattgtgt cgacgaaggc ttttggctcc 7320

tctgtcgctg tctcaagcag catctaaccc tgcgtcgccg tttccatttg cagccgctgg 7380

cccgccgagc cctggaggag ctcgggctgc cggtgccgcc ggtgctgcgg gtgcccggcg 7440

agagcaccaa ccccgtactg gtcggcgagc ccggcccggt gatcaagctg ttcggcgagc 7500

actggtgcgg tccggagagc ctcgcgtcgg agtcggaggc gtacgcggtc ctggcggacg 7560

ccccggtgcc ggtgccccgc ctcctcggcc gcggcgagct gcggcccggc accggagcct 7620

ggccgtggcc ctacctggtg atgagccgga tgaccggcac cacctggcgg tccgcgatgg 7680

acggcacgac cgaccggaac gcgctgctcg ccctggcccg cgaactcggc cgggtgctcg 7740

gccggctgca cagggtgccg ctgaccggga acaccgtgct caccccccat tccgaggtct 7800

tcccggaact gctgcgggaa cgccgcgcgg cgaccgtcga ggaccaccgc gggtggggct 7860

acctctcgcc ccggctgctg gaccgcctgg aggactggct gccggacgtg gacacgctgc 7920

tggccggccg cgaaccccgg ttcgtccacg gcgacctgca cgggaccaac atcttcgtgg 7980

acctggccgc gaccgaggtc accgggatcg tcgacttcac cgacgtctat gcgggagact 8040

cccgctacag cctggtgcaa ctgcatctca acgccttccg gggcgaccgc gagatcctgg 8100

ccgcgctgct cgacggggcg cagtggaagc ggaccgagga cttcgcccgc gaactgctcg 8160

ccttcacctt cctgcacgac ttcgaggtgt tcgaggagac cccgctggat ctctccggct 8220

tcaccgatcc ggaggaactg gcgcagttcc tctgggggcc gccggacacc gccccggcg 8280

[0081]

cctgataagg atccggcaag actggccccg cttggcaacg caacagtgag cccctcccta 8340

gtgtgtttgg ggatgtgact atgtattcgt gtgttggcca acgggtcaac ccgaacagat 8400

tgatacccgc cttggcattt cctgtcagaa tgtaacgtca gttgatggta ct 8452

<210> 130
<211> 27
<212> DNA
<213> 智人

<400> 130
ccgtgccggg cggggagacc gccatgg 27

<210> 131
<211> 22
<212> DNA
<213> 智人

<400> 131
ggcccggctg tggctgagga gc 22

<210> 132
<211> 20
<212> DNA
<213> 智人

<400> 132
cggtctcccg cccggcacgg 20

<210> 133
<211> 26
<212> DNA
<213> 智人

<400> 133
gctcctcagc cacagccggg ccgggt 26

<210> 134
<211> 12
<212> DNA
<213> 智人

[0082]
<400> 134
cgaccctgga aa 12

<210> 135
<211> 14
<212> DNA
<213> 智人

<400> 135
ccccgccgcc accc 14

<210> 136
<211> 18
<212> DNA
<213> 智人

<400> 136
tttccagggt cgccatgg 18

<210> 137
<211> 10
<212> DNA
<213> 智人

<400> 137
ggcggcgggg 10

<210> 138
<211> 20
<212> DNA
<213> 智人

<400> 138
accettgtta gccacctccc 20

<210> 139
<211> 20
<212> DNA
<213> 智人

<400> 139
gaacgcagtg ctcttcgaag 20

[0083]

<210> 140
<211> 20
<212> DNA
<213> 智人

<400> 140
ctcacgccct gctccgtgta 20

<210> 141
<211> 20
<212> DNA
<213> 智人

<400> 141
ggcgacaact acttcctggt 20

<210> 142
<211> 20
<212> DNA
<213> 智人

<400> 142
ctcacgccct gctccgtgta 20

<210> 143
<211> 20
<212> DNA
<213> 智人

<400> 143
gggcgacaac tacttcctgg 20

<210> 144
<211> 20
<212> DNA
<213> 智人

<400> 144
cctcttcagg gccggggtgg 20

<210> 145
<211> 20
<212> DNA
<213> 智人

[0084]
<400> 145
gaggacccag gtggaactgc 20

<210> 146
<211> 20
<212> DNA
<213> 智人

<400> 146
tcagctccag gcggtcctgg 20

<210> 147
<211> 20
<212> DNA
<213> 智人

<400> 147
agcagcagca gcagtggcag 20

<210> 148
<211> 20
<212> DNA
<213> 智人

<400> 148
tgggcaccgt cagctccagg 20

<210> 149
<211> 20
<212> DNA
<213> 智人

<400> 149
cagcagtggc agcggccacc 20

<210> 150
<211> 20
<212> DNA
<213> 智人

<400> 150
acctctcccc tggccctcat 20

[0085]

<210> 151
<211> 20
<212> DNA
<213> 智人

<400> 151
ccaggaccgc ctggagctga 20

<210> 152
<211> 20
<212> DNA
<213> 智人

<400> 152
ccgtcagctc caggcggtcc 20

<210> 153
<211> 20
<212> DNA
<213> 智人

<400> 153
agcagcagca gcagtggcag 20

<210> 154
<211> 20
<212> DNA
<213> 智人

<400> 154
atgtgccaag caaagcctca 20

<210> 155
<211> 20
<212> DNA
<213> 智人

<400> 155
ttcggtcatg cccgtggatg 20

<210> 156
<211> 20
<212> DNA
<213> 智人

[0086]
<400> 156
gtcgttgaaa ttcatcgtac 20

<210> 157
<211> 20
<212> DNA
<213> 智人

<400> 157
accacctgtg aagagtttcc 20

<210> 158
<211> 20
<212> DNA
<213> 智人

<400> 158
cgtcgttgaa attcatcgta 20

<210> 159
<211> 20
<212> DNA
<213> 智人

<400> 159
accacctgtg aagagtttcc 20

<210> 160
<211> 20
<212> DNA
<213> 小鼠

<400> 160
gaacgcagtg cttttcgagg 20

<210> 161
<211> 20
<212> DNA
<213> 小鼠

<400> 161
acccttgttg gccacctccc 20

[0087]

<210> 162
<211> 20
<212> DNA
<213> 小鼠

<400> 162
ggtgacaact actatctggt 20

<210> 163
<211> 20
<212> DNA
<213> 小鼠

<400> 163
ctcacaccct gctccgtgta 20

<210> 164
<211> 20
<212> DNA
<213> 小鼠

<400> 164
gggtgacaac tactatctgg 20

<210> 165
<211> 20
<212> DNA
<213> 小鼠

<400> 165
ctcacaccct gctccgtgta 20

<210> 166
<211> 20
<212> DNA
<213> 小鼠

<400> 166
cgagaacgca gtgcttttcg 20

<210> 167
<211> 20
<212> DNA
<213> 小鼠

[0088]
<400> 167
acccttgttg gccacctccc 20

<210> 168
<211> 20
<212> DNA
<213> 小鼠

<400> 168
atgagccaag caaatcctca 20

<210> 169
<211> 20
<212> DNA
<213> 小鼠

<400> 169
ttccgtcatg cccgtggaca 20

<210> 170
<211> 20
<212> DNA
<213> 小鼠

<400> 170
cttcgttgaa aaccattgta 20

<210> 171
<211> 20
<212> DNA
<213> 小鼠

<400> 171
ccacctctga agagtttcct 20

<210> 172
<211> 20
<212> DNA
<213> 小鼠

<400> 172
cttcgttgaa aaccattgta 20

[0089]

<210> 173
<211> 20
<212> DNA
<213> 小鼠

<400> 173
accacctctg aagagtttcc 20

<210> 174
<211> 20
<212> DNA
<213> 小鼠

<400> 174
cttccactca ctctgcgatt 20

<210> 175
<211> 20
<212> DNA
<213> 小鼠

<400> 175
accatgtctc agtgtcaagc 20

<210> 176
<211> 20
<212> DNA
<213> 小鼠

<400> 176
ggcggcaaca gcggcaacag 20

<210> 177
<211> 20
<212> DNA
<213> 小鼠

<400> 177
actgctctgc gtggctgcgg 20

<210> 178
<211> 20
<212> DNA
<213> 小鼠

[0090]
<400> 178
ccgcagccac gcagagcagt 20

<210> 179
<211> 20
<212> DNA
<213> 小鼠

<400> 179
gcacctctcc tcgccccgat 20

<210> 180
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 180
gagggcctat ttcccatgat tcc 23

<210> 181

<211> 126
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<220>
<221> 修饰的碱基
<222> (84)..(102)
<223> a, c, t, g, 未知的或其他的

<400> 181
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacnnnnnnn nnnnnnnnnn nnccggtgtt tcgtcctttc 120

cacaag 126

<210> 182
[0091] <211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<220>
<221> 修饰的碱基
<222> (6)..(24)
<223> a, c, t, g, 未知的或其他的

<400> 182
caccgnnnnn nnnnnnnnnn nnnn 24

<210> 183
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<220>
<221> 修饰的碱基
<222> (5)..(23)
<223> a, c, t, g, 未知的或其他的

<400> 183
aaacnnnnnn nnnnnnnnnn nnnc 24

<210> 184
<211> 126
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 184
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacccctagt cattggaggt gaccggtgtt tcgtcctttc 120

cacaag 126

[0092]

<210> 185
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 185
caccgtcacc tccaatgact aggg 24

<210> 186
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 186
aaacccctag tcattggagg tgac 24

<210> 187
<211> 192
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 187
cagaagaaga agggctccca tcacatcaac cggtggcgca ttgccacgaa gcaggccaat 60

ggggaggaca tcgatgtcac ctccaatgac aagcttgcta gcggtgggca accacaaacc 120

cacgagggca gagtgctgct tgctgctggc caggcccctg cgtgggccca agctggactc 180

tggccactcc ct 192

<210> 188
<211> 192
<212> DNA
<213> 人工序列

[0093]

<220>
<223> 人工序列的描述：合成引物

<400> 188
agggagtggc cagagtccag cttgggccca cgcaggggcc tggccagcag caagcagcac 60

tctgccctcg tgggtttgtg gttgcccacc gctagcaagc ttgtcattgg aggtgacatc 120

gatgtcctcc ccattggcct gcttcgtggc aatgcgccac cggttgatgt gatgggagcc 180

cttcttcttc tg 192

<210> 189
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 189

ccatcccctt ctgtgaatgt 20

<210> 190
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 190
ggagattgga gacacggaga 20

<210> 191
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0094]
<400> 191
ggctccctgg gttcaaagta 20

<210> 192
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 192
agaggggtct ggatgtcgta a 21

<210> 193
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 193

cgccagggtt ttcccagtca cgac 24

<210> 194
<211> 51
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 194
gagggtctcg tccttgcggc cgcgctagcg agggcctatt tcccatgatt c 51

<210> 195
<211> 133
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0095] <220>
<221> 修饰的碱基
<222> (95)..(114)
<223> a, c, t, g, 未知的或其他的

<400> 195
ctcggtctcg gtaaaaaagc accgactcgg tgccactttt tcaagttgat aacggactag 60

ccttatttta acttgctatt tctagctcta aaacnnnnnn nnnnnnnnnn nnnnggtgtt 120

tcgtcctttc cac 133

<210> 196
<211> 41
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 196
gagggtctct ttaccggtga gggcctattt cccatgattc c 41

<210> 197
<211> 133
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<220>
<221> 修饰的碱基
<222> (95)..(114)
<223> a, c, t, g, 未知的或其他的

<400> 197
ctcggtctcc tcaaaaaagc accgactcgg tgccactttt tcaagttgat aacggactag 60

ccttatttta acttgctatt tctagctcta aaacnnnnnn nnnnnnnnnn nnnnggtgtt 120

tcgtcctttc cac 133

<210> 198
[0096] <211> 40
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 198
gagggtctct ttgagctcga gggcctattt cccatgattc 40

<210> 199
<211> 133
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<220>
<221> 修饰的碱基
<222> (96)..(115)
<223> a, c, t, g, 未知的或其他的

<400> 199
ctcggtctcg cgtaaaaaag caccgactcg gtgccacttt ttcaagttga taacggacta 60

gccttatttt aacttgctat ttctagctct aaaacnnnnn nnnnnnnnnn nnnnnggtgt 120

ttcgtccttt cca 133

<210> 200
<211> 27
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 200
gagggtctct tacgcgtgtg tctagac 27

<210> 201
<211> 98
<212> DNA
<213> 人工序列

[0097]

<220>
<223> 人工序列的描述：合成引物

<400> 201
ctcggtctca aggacaggga agggagcagt ggttcacgcc tgtaatccca gcaatttggg 60

aggccaaggt gggtagatca cctgagatta ggagttgc 98

<210> 202
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 202
cctgtccttg cggccgcgct agcgagggcc 30

<210> 203
<211> 31

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 203
cacgcggccg caaggacagg gaagggagca g 31

<210> 204
<211> 327
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成多肽

<400> 204
Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
1 5 10 15

[0098]
Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20 25 30

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

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

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

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

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

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

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

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

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

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

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

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

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

Gly Leu Arg Ser Arg Glu Glu Glu Glu Glu Thr Asp Ser Arg Met Pro
245 250 255

[0099] His Leu Asp Ser Pro Gly Ser Ser Gln Pro Arg Arg Ser Phe Leu Ser
260 265 270

Arg Val Ile Arg Ala Ala Leu Pro Leu Gln Leu Leu Leu Leu Leu Leu
275 280 285

Leu Leu Leu Ala Cys Leu Leu Pro Ala Ser Glu Asp Asp Tyr Ser Cys
290 295 300

Thr Gln Ala Asn Asn Phe Ala Arg Ser Phe Tyr Pro Met Leu Arg Tyr
305 310 315 320

Thr Asn Gly Pro Pro Pro Thr
325

<210> 205
<211> 3243
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 205

accggtgcca ccatgtaccc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc 120

gtggggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg 180

ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc 240

ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac 300

ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc 360

ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc 420

cgaggagtgc ataacgtcaa tgaggtggaa gaggacaccg gcaacgagct gtctacaaag 480

gaacagatct cacgcaatag caaagctctg gaagagaagt atgtcgcaga gctgcagctg 540

gaacggctga agaaagatgg cgaggtgaga gggtcaatta ataggttcaa gacaagcgac 600

[0100] tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag 660

agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca 720

ggagaaggga gccccttcgg atggaaagac atcaaggaat ggtacgagat gctgatggga 780

cattgcacct attttccaga agagctgaga agcgtcaagt acgcttataa cgcagatctg 840

tacaacgccc tgaatgacct gaacaacctg gtcatcacca gggatgaaaa cgagaaactg 900

gaatactatg agaagttcca gatcatcgaa aacgtgttta agcagaagaa aaagcctaca 960

ctgaaacaga ttgctaagga gatcctggtc aacgaagagg acatcaaggg ctaccgggtg 1020

acaagcactg gaaaaccaga gttcaccaat ctgaaagtgt atcacgatat taaggacatc 1080

acagcacgga aagaaatcat tgagaacgcc gaactgctgg atcagattgc taagatcctg 1140

actatctacc agagctccga ggacatccag gaagagctga ctaacctgaa cagcgagctg 1200

acccaggaag agatcgaaca gattagtaat ctgaaggggt acaccggaac acacaacctg 1260

tccctgaaag ctatcaatct gattctggat gagctgtggc atacaaacga caatcagatt 1320

gcaatcttta accggctgaa gctggtccca aaaaaggtgg acctgagtca gcagaaagag 1380

atcccaacca cactggtgga cgatttcatt ctgtcacccg tggtcaagcg gagcttcatc 1440

cagagcatca aagtgatcaa cgccatcatc aagaagtacg gcctgcccaa tgatatcatt 1500

atcgagctgg ctagggagaa gaacagcaag gacgcacaga agatgatcaa tgagatgcag 1560

aaacgaaacc ggcagaccaa tgaacgcatt gaagagatta tccgaactac cgggaaagag 1620

aacgcaaagt acctgattga aaaaatcaag ctgcacgata tgcaggaggg aaagtgtctg 1680

tattctctgg aggccatccc cctggaggac ctgctgaaca atccattcaa ctacgaggtc 1740

gatcatatta tccccagaag cgtgtccttc gacaattcct ttaacaacaa ggtgctggtc 1800

aagcaggaag agaactctaa aaagggcaat aggactcctt tccagtacct gtctagttca 1860

gattccaaga tctcttacga aacctttaaa aagcacattc tgaatctggc caaaggaaag 1920

[0101]

ggccgcatca gcaagaccaa aaaggagtac ctgctggaag agcgggacat caacagattc 1980

tccgtccaga aggattttat taaccggaat ctggtggaca caagatacgc tactcgcggc 2040

ctgatgaatc tgctgcgatc ctatttccgg gtgaacaatc tggatgtgaa agtcaagtcc 2100

atcaacggcg ggttcacatc ttttctgagg cgcaaatgga agtttaaaaa ggagcgcaac 2160

aaagggtaca agcaccatgc cgaagatgct ctgattatcg caaatgccga cttcatcttt 2220

aaggagtgga aaaagctgga caaagccaag aaagtgatgg agaaccagat gttcgaagag 2280

aagcaggccg aatctatgcc cgaaatcgag acagaacagg agtacaagga gattttcatc 2340

actcctcacc agatcaagca tatcaaggat ttcaaggact acaagtactc tcaccgggtg 2400

gataaaaagc ccaacagaga gctgatcaat gacaccctgt atagtacaag aaaagacgat 2460

aaggggaata ccctgattgt gaacaatctg aacggactgt acgacaaaga taatgacaag 2520

ctgaaaaagc tgatcaacaa aagtcccgag aagctgctga tgtaccacca tgatcctcag 2580

acatatcaga aactgaagct gattatggag cagtacggcg acgagaagaa cccactgtat 2640

aagtactatg aagagactgg gaactacctg accaagtata gcaaaaagga taatggcccc 2700

gtgatcaaga agatcaagta ctatgggaac aagctgaatg cccatctgga catcacagac 2760

gattacccta acagtcgcaa caaggtggtc aagctgtcac tgaagccata cagattcgat 2820

gtctatctgg acaacggcgt gtataaattt gtgactgtca agaatctgga tgtcatcaaa 2880

aaggagaact actatgaagt gaatagcaag tgctacgaag aggctaaaaa gctgaaaaag 2940

attagcaacc aggcagagtt catcgcctcc ttttacaaca acgacctgat taagatcaat 3000

ggcgaactgt atagggtcat cggggtgaac aatgatctgc tgaaccgcat tgaagtgaat 3060

atgattgaca tcacttaccg agagtatctg gaaaacatga atgataagcg ccccccctcga 3120

attatcaaaa caattgcctc taagactcag agtatcaaaa agtactcaac cgacattctg 3180

ggaaacctgt atgaggtgaa gagcaaaaag caccctcaga ttatcaaaaa gggctaagaa 3240

[0102] ttc 3243

<210> 206
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 206
atctcttaga taccagcatc 20

<210> 207
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 207

tcaatctccc gatgggcacc 20

<210> 208
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 208
gcccatcggg agattgaggg 20

<210> 209
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0103]
<400> 209
acttcaacag cgtgccggag 20

<210> 210
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 210
ccgctgacca cacctgccag 20

<210> 211
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 211

tggcaggtgt ggtcagcggc 20

<210> 212
<211> 102
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

<400> 212
gttttagagc tatgctgttt tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa 60

gaagaagttt tagagctatg ctgttttgaa tggtcccaaa ac 102

<210> 213
<211> 100
<212> DNA
<213> 智人

<400> 213
[0104] cggaggacaa agtacaaacg gcagaagctg gaggaggaag ggcctgagtc cgagcagaag 60

aagaagggct cccatcacat caaccggtgg cgcattgcca 100

<210> 214
<211> 50
<212> DNA
<213> 智人

<400> 214
agctggagga ggaagggcct gagtccgagc agaagaagaa gggctcccac 50

<210> 215
<211> 30
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 215
gaguccgagc agaagaagaa guuuuagagc 30

<210> 216
<211> 49
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 216
agctggagga ggaagggcct gagtccgagc agaagagaag ggctcccac 49

<210> 217
<211> 53
<212> DNA
<213> 智人

<400> 217
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccatca cat 53

<210> 218
<211> 52
<212> DNA
[0105] <213> 智人

<400> 218
ctggaggagg aagggcctga gtccgagcag aagagaaggg ctcccatcac at 52

<210> 219
<211> 54
<212> DNA
<213> 智人

<400> 219
ctggaggagg aagggcctga gtccgagcag aagaaagaag ggctcccatc acat 54

<210> 220
<211> 50
<212> DNA
<213> 智人

<400> 220
ctggaggagg aagggcctga gtccgagcag aagaagggct cccatcacat 50

<210> 221

<211> 47
<212> DNA
<213> 智人

<400> 221
ctggaggagg aagggcctga gcccgagcag aagggctccc atcacat 47

<210> 222
<211> 48
<212> DNA
<213> 智人

<400> 222
ctggaggagg aagggcctga gtccgagcag aagaagaagg gctcccat 48

<210> 223
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0106]

<400> 223
gaguccgagc agaagaagau 20

<210> 224
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 224
gaguccgagc agaagaagua 20

<210> 225
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 225
gaguccgagc agaagaacaa 20

<210> 226
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 226
gaguccgagc agaagaugaa 20

<210> 227
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0107]

<400> 227
gaguccgagc agaaguagaa 20

<210> 228
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 228
gaguccgagc agaugaagaa 20

<210> 229
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 229
gaguccgagc acaagaagaa 20

<210> 230
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 230
gaguccgagg agaagaagaa 20

<210> 231
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0108]

<400> 231
gaguccgugc agaagaagaa 20

<210> 232
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 232
gagucggagc agaagaagaa 20

<210> 233
<211> 20
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 233
gagaccgagc agaagaagaa 20

<210> 234
<211> 24
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 234
aatgacaagc ttgctagcgg tggg 24

<210> 235
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0109]

<400> 235
aaaacggaag ggcctgagtc cgagcagaag aagaagttt 39

<210> 236
<211> 39
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 236
aaacaggggc cgagattggg tgttcagggc agaggtttt 39

<210> 237
<211> 38
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 237
aaaacggaag ggcctgagtc cgagcagaag aagaagtt 38

<210> 238
<211> 40
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 238
aacggaggga ggggcacaga tgagaaactc agggttttag 40

<210> 239
<211> 38
<212> DNA
<213> 智人

<400> 239
agcccttctt cttctgctcg gactcaggcc cttcctcc 38

[0110]

<210> 240
<211> 40
<212> DNA
<213> 智人

<400> 240
cagggaggga ggggcacaga tgagaaactc aggaggcccc 40

<210> 241
<211> 80
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 241
ggcaatgcgc caccggttga tgtgatggga gcccttctag gaggccccca gagcagccac 60

tggggcctca acactcaggc 80

<210> 242
<211> 33
<212> DNA
<213> 智人

<400> 242
catcgatgtc ctccccattg gcctgcttcg tgg 33

<210> 243
<211> 33
<212> DNA
<213> 智人

<400> 243
ttcgtggcaa tgcgccaccg gttgatgtga tgg 33

<210> 244
<211> 33
<212> DNA
<213> 智人

[0111]
<400> 244
tcgtggcaat gcgccaccgg ttgatgtgat ggg 33

<210> 245
<211> 33
<212> DNA
<213> 智人

<400> 245
tccagcttct gccgtttgta ctttgtcctc cgg 33

<210> 246
<211> 33
<212> DNA
<213> 智人

<400> 246
ggagggaggg gcacagatga gaaactcagg agg 33

<210> 247
<211> 33
<212> DNA
<213> 智人

<400> 247
aggggccgag attgggtgtt cagggcagag agg 33

<210> 248
<211> 33
<212> DNA
<213> 小鼠

<400> 248
caagcactga gtgccattag ctaaatgcat agg 33

<210> 249
<211> 33
<212> DNA
<213> 小鼠

<400> 249
aatgcatagg gtaccaccca caggtgccag ggg 33

[0112]

<210> 250
<211> 33
<212> DNA
<213> 小鼠

<400> 250
acacacatgg gaaagcctct gggccaggaa agg 33

<210> 251
<211> 37
<212> DNA
<213> 智人

<400> 251
ggaggaggta gtatacagaa acacagagaa gtagaat 37

<210> 252
<211> 37
<212> DNA
<213> 智人

<400> 252
agaatgtaga ggagtcacag aaactcagca ctagaaa 37

<210> 253
<211> 98
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 253
ggacgaaaca ccggaaccat tcaaaacagc atagcaagtt aaaataaggc tagtccgtta 60

tcaacttgaa aaagtggcac cgagtcggtg cttttttt 98

<210> 254
<211> 186
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成多核苷酸

[0113]
<400> 254
ggacgaaaca ccggtagtat taagtattgt tttatggctg ataaatttct ttgaatttct 60

ccttgattat ttgttataaa agttataaaa taatcttgtt ggaaccattc aaaacagcat 120

agcaagttaa aataaggcta gtccgttatc aacttgaaaa agtggcaccg agtcggtgct 180

tttttt 186

<210> 255
<211> 95
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 255
gggttttaga gctatgctgt tttgaatggt cccaaaacgg gtcttcgaga agacgtttta 60

gagctatgct gttttgaatg gtcccaaaac ttttt 95

<210> 256

<211> 36
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (5)..(34)
<223> a, c, t, g, 未知的或其他的

<400> 256
aaacnnnnnn nnnnnnnnnn nnnnnnnnnn nnnngt 36

<210> 257
<211> 36
<212> DNA
<213> 人工序列

<220>
[0114] <223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (7)..(36)
<223> a, c, t, g, 未知的或其他的

<400> 257
taaaacnnnn nnnnnnnnnn nnnnnnnnnn nnnnnn 36

<210> 258
<211> 84
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 258
gtggaaagga cgaaacaccg ggtcttcgag aagacctgtt ttagagctag aaatagcaag 60

ttaaaataag gctagtccgt tttt 84

<210> 259
<211> 46
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(19)
<223> a, c, u, g, 未知的或其他的

<400> 259
nnnnnnnnnn nnnnnnnnng uuauuguacu cucaagauuu auuuuu 46

<210> 260
<211> 91
<212> RNA
<213> 人工序列

[0115] <220>
<223> 人工序列的描述：合成寡核苷酸

<400> 260
guuacuuaaa ucuugcagaa gcuacaaaga uaaggcuuca ugccgaaauc aacacccugu 60

cauuuuaugg caggguguuu ucguuauuua a 91

<210> 261
<211> 70
<212> DNA
<213> 智人

<400> 261
ttttctagtg ctgagtttct gtgactcctc tacattctac ttctctgtgt ttctgtatac 60

tacctcctcc 70

<210> 262
<211> 122
<212> DNA
<213> 智人

<400> 262
ggaggaaggg cctgagtccg agcagaagaa gaagggctcc catcacatca accggtggcg 60

cattgccacg aagcaggcca atggggagga catcgatgtc acctccaatg actagggtgg 120

gc 122

<210> 263
<211> 48
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(32)
<223> a, c, u, g, 未知的或其他的

<400> 263
[0116] acnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnguuuuaga gcuaugcu 48

<210> 264
<211> 67
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<223> Description of Combined DNA/RNA Molecule: Synthetic 寡核苷酸

<400> 264
agcauagcaa guuaaaauaa ggctaguccg uuaucaacuu gaaaaagugg caccgagucg 60

gugcuuu 67

<210> 265
<211> 62
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<400> 265
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cg 62

<210> 266
<211> 73
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0117] <400> 266
tgaatggtcc caaaacggaa gggcctgagt ccgagcagaa gaagaagttt tagagctatg 60

ctgttttgaa tgg 73

<210> 267
<211> 69
<212> DNA
<213> 智人

<400> 267
ctggtcttcc acctctctgc cctgaacacc caatctcggc ccctctcgcc accctcctgc 60

atttctgtt 69

<210> 268
<211> 138
<212> DNA
<213> 小鼠

<400> 268
acccaagcac tgagtgccat tagctaaatg catagggtac cacccacagg tgccaggggc 60

ctttcccaaa gttcccagcc ccttctccaa cctttcctgg cccagaggct ttcccatgtg 120

tgtggctgga ccctttga 138

<210> 269
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 269
aaaaccaccc ttctctctgg c 21

<210> 270
<211> 21
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0118]

<400> 270
ggagattgga gacacggaga g 21

<210> 271
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 271
ctggaaagcc aatgcctgac 20

<210> 272
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 272
ggcagcaaac tccttgtcct 20

<210> 273
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 273
gtgctttgca gaggcctacc 20

<210> 274
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0119]

<400> 274
cctggagcgc atgcagtagt 20

<210> 275
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 275
accttctgtg tttccaccat tc 22

<210> 276
<211> 20
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 276
ttggggagtg cacagacttc 20

<210> 277
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成
probe

<400> 277
tagctctaaa acttcttctt ctgctcggac 30

<210> 278
<211> 30
<212> DNA
<213> 人工序列

[0120]
<220>
<223> 人工序列的描述：合成
probe

<400> 278
ctagccttat tttaacttgc tatgctgttt 30

<210> 279
<211> 99
<212> RNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (1)..(20)
<223> a, c, u, g, 未知的或其他的

<400> 279
nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uugaaaaagu ggcaccgagu cggugcuuu 99

<210> 280
<211> 12
<212> DNA
<213> 智人

<400> 280
tagcgggtaa gc 12

<210> 281
<211> 12
<212> DNA
<213> 智人

<400> 281
tcggtgacat gt 12

<210> 282
<211> 12
<212> DNA
<213> 智人

[0121]

<400> 282
actccccgta gg 12

<210> 283
<211> 12
<212> DNA
<213> 智人

<400> 283
actgcgtgtt aa 12

<210> 284
<211> 12
<212> DNA
<213> 智人

<400> 284
acgtcgcctg at 12

<210> 285
<211> 12

<212> DNA
<213> 智人

<400> 285
taggtcgacc ag 12

<210> 286
<211> 12
<212> DNA
<213> 智人

<400> 286
ggcgttaatg at 12

<210> 287
<211> 12
<212> DNA
<213> 智人

<400> 287
tgtcgcatgt ta 12

[0122]

<210> 288
<211> 12
<212> DNA
<213> 智人

<400> 288
atggaaacgc at 12

<210> 289
<211> 12
<212> DNA
<213> 智人

<400> 289
gccgaattcc tc 12

<210> 290
<211> 12
<212> DNA
<213> 智人

<400> 290

gcatggtacg ga 12

<210> 291
<211> 12
<212> DNA
<213> 智人

<400> 291
cggtactctt ac 12

<210> 292
<211> 12
<212> DNA
<213> 智人

<400> 292
gcctgtgccg ta 12

<210> 293
<211> 12
<212> DNA
<213> 智人

[0123]

<400> 293
tacggtaagt cg 12

<210> 294
<211> 12
<212> DNA
<213> 智人

<400> 294
cacgaaatta cc 12

<210> 295
<211> 12
<212> DNA
<213> 智人

<400> 295
aaccaagata cg 12

<210> 296
<211> 12

<210> DNA
<213> 智人

<400> 296
gagtcgatac gc 12

<210> 297
<211> 12
<212> DNA
<213> 智人

<400> 297
gtctcacgat cg 12

<210> 298
<211> 12
<212> DNA
<213> 智人

<400> 298
tcgtcgggtg ca 12

[0124]

<210> 299
<211> 12
<212> DNA
<213> 智人

<400> 299
actccgtagt ga 12

<210> 300
<211> 12
<212> DNA
<213> 智人

<400> 300
caggacgtcc gt 12

<210> 301
<211> 12
<212> DNA
<213> 智人

<400> 301

tcgtatccct ac 12

<210> 302
<211> 12
<212> DNA
<213> 智人

<400> 302
tttcaaggcc gg 12

<210> 303
<211> 12
<212> DNA
<213> 智人

<400> 303
cgccggtgga at 12

<210> 304
<211> 12
<212> DNA
<213> 智人

[0125]

<400> 304
gaacccgtcc ta 12

<210> 305
<211> 12
<212> DNA
<213> 智人

<400> 305
gattcatcag cg 12

<210> 306
<211> 12
<212> DNA
<213> 智人

<400> 306
acaccggtct tc 12

<210> 307
<211> 12

<212> DNA
<213> 智人

<400> 307
atcgtgccct aa 12

<210> 308
<211> 12
<212> DNA
<213> 智人

<400> 308
gcgtcaatgt tc 12

<210> 309
<211> 12
<212> DNA
<213> 智人

<400> 309
ctccgtatct cg 12

[0126]

<210> 310
<211> 12
<212> DNA
<213> 智人

<400> 310
ccgattcctt cg 12

<210> 311
<211> 12
<212> DNA
<213> 智人

<400> 311
tgcgcctcca gt 12

<210> 312
<211> 12
<212> DNA
<213> 智人

<400> 312

taacgtcgga gc 12

<210> 313
<211> 12
<212> DNA
<213> 智人

<400> 313
aaggtcgccc at 12

<210> 314
<211> 12
<212> DNA
<213> 智人

<400> 314
gtcggggact at 12

<210> 315
<211> 12
<212> DNA
<213> 智人

[0127]

<400> 315
ttcgagcgat tt 12

<210> 316
<211> 12
<212> DNA
<213> 智人

<400> 316
tgagtcgtcg ag 12

<210> 317
<211> 12
<212> DNA
<213> 智人

<400> 317
tttacgcaga gg 12

<210> 318
<211> 12

<212> DNA
<213> 智人

<400> 318
aggaagtatc gc 12

<210> 319
<211> 12
<212> DNA
<213> 智人

<400> 319
actcgatacc at 12

<210> 320
<211> 12
<212> DNA
<213> 智人

<400> 320
cgctacatag ca 12

[0128]

<210> 321
<211> 12
<212> DNA
<213> 智人

<400> 321
ttcataaccg gc 12

<210> 322
<211> 12
<212> DNA
<213> 智人

<400> 322
ccaaacggtt aa 12

<210> 323
<211> 12
<212> DNA
<213> 智人

<400> 323

cgattccttc gt 12

<210> 324
<211> 12
<212> DNA
<213> 智人

<400> 324
cgtcatgaat aa 12

<210> 325
<211> 12
<212> DNA
<213> 智人

<400> 325
agtggcgatg ac 12

<210> 326
<211> 12
<212> DNA
<213> 智人

[0129]

<400> 326
cccctacggc ac 12

<210> 327
<211> 12
<212> DNA
<213> 智人

<400> 327
gccaacccgc ac 12

<210> 328
<211> 12
<212> DNA
<213> 智人

<400> 328
tgggacaccg gt 12

<210> 329
<211> 12

<212> DNA
<213> 智人

<400> 329
ttgactgcgg cg 12

<210> 330
<211> 12
<212> DNA
<213> 智人

<400> 330
actatgcgta gg 12

<210> 331
<211> 12
<212> DNA
<213> 智人

<400> 331
tcacccaaag cg 12

[0130]

<210> 332
<211> 12
<212> DNA
<213> 智人

<400> 332
gcaggacgtc cg 12

<210> 333
<211> 12
<212> DNA
<213> 智人

<400> 333
acaccgaaaa cg 12

<210> 334
<211> 12
<212> DNA
<213> 智人

<400> 334

cggtgtattg ag 12

<210> 335
<211> 12
<212> DNA
<213> 智人

<400> 335
cacgaggtat gc 12

<210> 336
<211> 12
<212> DNA
<213> 智人

<400> 336
taaagcgacc cg 12

<210> 337
<211> 12
<212> DNA
<213> 智人

[0131]

<400> 337
cttagtcggc ca 12

<210> 338
<211> 12
<212> DNA
<213> 智人

<400> 338
cgaaaacgtg gc 12

<210> 339
<211> 12
<212> DNA
<213> 智人

<400> 339
cgtgccctga ac 12

<210> 340
<211> 12

<210> DNA
<213> 智人

<400> 340
tttaccatcg aa 12

<210> 341
<211> 12
<212> DNA
<213> 智人

<400> 341
cgtagccatg tt 12

<210> 342
<211> 12
<212> DNA
<213> 智人

<400> 342
cccaaacggt ta 12

[0132]

<210> 343
<211> 12
<212> DNA
<213> 智人

<400> 343
gcgttatcag aa 12

<210> 344
<211> 12
<212> DNA
<213> 智人

<400> 344
tcgatggtaa ac 12

<210> 345
<211> 12
<212> DNA
<213> 智人

<400> 345

cgacttttg ca 12

<210> 346
<211> 12
<212> DNA
<213> 智人

<400> 346
tcgacgactc ac 12

<210> 347
<211> 12
<212> DNA
<213> 智人

<400> 347
acgcgtcaga ta 12

<210> 348
<211> 12
<212> DNA
<213> 智人

[0133]

<400> 348
cgtacggcac ag 12

<210> 349
<211> 12
<212> DNA
<213> 智人

<400> 349
ctatgccgtg ca 12

<210> 350
<211> 12
<212> DNA
<213> 智人

<400> 350
cgcgtcagat at 12

<210> 351
<211> 12

<212> DNA
<213> 智人

<400> 351
aagatcggta gc 12

<210> 352
<211> 12
<212> DNA
<213> 智人

<400> 352
cttcgcaagg ag 12

<210> 353
<211> 12
<212> DNA
<213> 智人

<400> 353
[0134] gtcgtggact ac 12

<210> 354
<211> 12
<212> DNA
<213> 智人

<400> 354
ggtcgtcatc aa 12

<210> 355
<211> 12
<212> DNA
<213> 智人

<400> 355
gttaacagcg tg 12

<210> 356
<211> 12
<212> DNA
<213> 智人

<400> 356
tagctaaccg tt 12

<210> 357
<211> 12
<212> DNA
<213> 智人

<400> 357
agtaaaggcg ct 12

<210> 358
<211> 12
<212> DNA
<213> 智人

<400> 358
ggtaatttcg tg 12

<210> 359
<211> 15
<212> DNA
[0135] <213> 智人

<400> 359
cagaagaaga agggc 15

<210> 360
<211> 51
<212> DNA
<213> 智人

<400> 360
ccaatgggga ggacatcgat gtcacctcca atgactaggg tggtgggcaa c 51

<210> 361
<211> 15
<212> DNA
<213> 智人

<400> 361
ctctggccac tccct 15

<210> 362

<211> 52
<212> DNA
<213> 智人

<400> 362
acatcgatgt cacctccaat gacaagcttg ctagcggtgg gcaaccacaa ac 52

<210> 363
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (6)..(25)
<223> a, c, t, g, 未知的或其他的

<400> 363
[0136] caccgnnnnn nnnnnnnnnn nnnnn 25

<210> 364
<211> 25
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (5)..(24)
<223> a, c, t, g, 未知的或其他的

<400> 364
aaacnnnnnn nnnnnnnnnn nnnnc 25

<210> 365
<211> 54
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 365
aacaccgggt cttcgagaag acctgtttta gagctagaaa tagcaagtta aaat 54

<210> 366
<211> 54
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<400> 366
caaaacgggt cttcgagaag acgttttaga gctatgctgt tttgaatggt ccca 54

<210> 367
<211> 4104
<212> DNA
<213> 智人

[0137]

<220>
<221> CDS
<222> (1)..(4104)

<400> 367
atg gac aag aag tac agc atc ggc ctg gac atc ggc acc aac tct gtg 48
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

ggc tgg gcc gtg atc acc gac gag tac aag gtg ccc agc aag aaa ttc 96
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

aag gtg ctg ggc aac acc gac cgg cac agc atc aag aag aac ctg atc 144
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

gga gcc ctg ctg ttc gac agc ggc gaa aca gcc gag gcc acc cgg ctg 192
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

aag aga acc gcc aga aga aga tac acc aga cgg aag aac cgg atc tgc 240

```
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65                  70                  75                  80

tat ctg caa gag atc ttc agc aac gag atg gcc aag gtg gac gac agc      288
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
                85                  90                  95

ttc ttc cac aga ctg gaa gag tcc ttc ctg gtg gaa gag gat aag aag      336
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
            100                 105                 110

cac gag cgg cac ccc atc ttc ggc aac atc gtg gac gag gtg gcc tac      384
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
        115                 120                 125

cac gag aag tac ccc acc atc tac cac ctg aga aag aaa ctg gtg gac      432
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140

agc acc gac aag gcc gac ctg cgg ctg atc tat ctg gcc ctg gcc cac      480
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160
```

[0138]

```
atg atc aag ttc cgg ggc cac ttc ctg atc gag ggc gac ctg aac ccc      528
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175

gac aac agc gac gtg gac aag ctg ttc atc cag ctg gtg cag acc tac      576
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190

aac cag ctg ttc gag gaa aac ccc atc aac gcc agc ggc gtg gac gcc      624
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205

aag gcc atc ctg tct gcc aga ctg agc aag agc aga cgg ctg gaa aat      672
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

ctg atc gcc cag ctg ccc ggc gag aag aag aat ggc ctg ttc ggc aac      720
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

ctg att gcc ctg agc ctg ggc ctg acc ccc aac ttc aag agc aac ttc      768
```

```
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

gac ctg gcc gag gat gcc aaa ctg cag ctg agc aag gac acc tac gac      816
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

gac gac ctg gac aac ctg ctg gcc cag atc ggc gac cag tac gcc gac      864
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

ctg ttt ctg gcc gcc aag aac ctg tcc gac gcc atc ctg ctg agc gac      912
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

atc ctg aga gtg aac acc gag atc acc aag gcc ccc ctg agc gcc tct      960
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

atg atc aag aga tac gac gag cac cac cag gac ctg acc ctg ctg aaa      1008
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335
```

[0139]

```
gct ctc gtg cgg cag cag ctg cct gag aag tac aaa gag att ttc ttc      1056
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

gac cag agc aag aac ggc tac gcc ggc tac att gac ggc gga gcc agc      1104
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

cag gaa gag ttc tac aag ttc atc aag ccc atc ctg gaa aag atg gac      1152
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380

ggc acc gag gaa ctg ctc gtg aag ctg aac aga gag gac ctg ctg cgg      1200
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

aag cag cgg acc ttc gac aac ggc agc atc ccc cac cag atc cac ctg      1248
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

gga gag ctg cac gcc att ctg cgg cgg cag gaa gat ttt tac cca ttc      1296
```

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
420 425 430

ctg aag gac aac cgg gaa aag atc gag aag atc ctg acc ttc cgc atc 1344
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
435 440 445

ccc tac tac gtg ggc cct ctg gcc agg gga aac agc aga ttc gcc tgg 1392
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
450 455 460

atg acc aga aag agc gag gaa acc atc acc ccc tgg aac ttc gag gaa 1440
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465 470 475 480

gtg gtg gac aag ggc gct tcc gcc cag agc ttc atc gag cgg atg acc 1488
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
485 490 495

aac ttc gat aag aac ctg ccc aac gag aag gtg ctg ccc aag cac agc 1536
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
500 505 510

[0140]

ctg ctg tac gag tac ttc acc gtg tat aac gag ctg acc aaa gtg aaa 1584
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
515 520 525

tac gtg acc gag gga atg aga aag ccc gcc ttc ctg agc ggc gag cag 1632
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
530 535 540

aaa aag gcc atc gtg gac ctg ctg ttc aag acc aac cgg aaa gtg acc 1680
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545 550 555 560

gtg aag cag ctg aaa gag gac tac ttc aag aaa atc gag tgc ttc gac 1728
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

tcc gtg gaa atc tcc ggc gtg gaa gat cgg ttc aac gcc tcc ctg ggc 1776
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
580 585 590

aca tac cac gat ctg ctg aaa att atc aag gac aag gac ttc ctg gac 1824

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

aat gag gaa aac gag gac att ctg gaa gat atc gtg ctg acc ctg aca   1872
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

ctg ttt gag gac aga gag atg atc gag gaa cgg ctg aaa acc tat gcc   1920
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

cac ctg ttc gac gac aaa gtg atg aag cag ctg aag cgg cgg aga tac   1968
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

acc ggc tgg ggc agg ctg agc cgg aag ctg atc aac ggc atc cgg gac   2016
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

aag cag tcc ggc aag aca atc ctg gat ttc ctg aag tcc gac ggc ttc   2064
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685
```

[0141]

```
gcc aac aga aac ttc atg cag ctg atc cac gac gac agc ctg acc ttt   2112
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

aaa gag gac atc cag aaa gcc cag gtg tcc ggc cag ggc gat agc ctg   2160
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705                 710                 715                 720

cac gag cac att gcc aat ctg gcc ggc agc ccc gcc att aag aag ggc   2208
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

atc ctg cag aca gtg aag gtg gtg gac gag ctc gtg aaa gtg atg ggc   2256
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

cgg cac aag ccc gag aac atc gtg atc gcc atg gcc aga gag aac cag   2304
Arg His Lys Pro Glu Asn Ile Val Ile Ala Met Ala Arg Glu Asn Gln
        755                 760                 765

acc acc cag aag gga cag aag aac agc cgc gag aga atg aag cgg atc   2352
```

```
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
    770                 775                 780

gaa gag ggc atc aaa gag ctg ggc agc cag atc ctg aaa gaa cac ccc      2400
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785                 790                 795                 800

gtg gaa aac acc cag ctg cag aac gag aag ctg tac ctg tac tac ctg      2448
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815

cag aat ggg cgg gat atg tac gtg gac cag gaa ctg gac atc aac cgg      2496
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
            820                 825                 830

ctg tcc gac tac gat gtg gac gcc atc gtg cct cag agc ttt ctg aag      2544
Leu Ser Asp Tyr Asp Val Asp Ala Ile Val Pro Gln Ser Phe Leu Lys
        835                 840                 845

gac gac tcc atc gac gcc aag gtg ctg acc aga agc gac aag gcc cgg      2592
Asp Asp Ser Ile Asp Ala Lys Val Leu Thr Arg Ser Asp Lys Ala Arg
    850                 855                 860
```

[0142]

```
ggc aag agc gac aac gtg ccc tcc gaa gag gtc gtg aag aag atg aag      2640
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865                 870                 875                 880

aac tac tgg cgg cag ctg ctg aac gcc aag ctg att acc cag aga aag      2688
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
                885                 890                 895

ttc gac aat ctg acc aag gcc gag aga ggc ggc ctg agc gaa ctg gat      2736
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
            900                 905                 910

aag gcc ggc ttc atc aag aga cag ctg gtg gaa acc cgg cag atc aca      2784
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
        915                 920                 925

aag cac gtg gca cag atc ctg gac tcc cgg atg aac act aag tac gac      2832
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
    930                 935                 940

gag aat gac aag ctg atc cgg gaa gtg aaa gtg atc acc ctg aag tcc      2880
```

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

aag ctg gtg tcc gat ttc cgg aag gat ttc cag ttt tac aaa gtg cgc      2928
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
                965                 970                 975

gag atc aac aac tac cac cac gcc cac gcc gcc tac ctg aac gcc gtc      2976
Glu Ile Asn Asn Tyr His His Ala His Ala Ala Tyr Leu Asn Ala Val
            980                 985                 990

gtg gga acc gcc ctg atc aaa aag  tac cct aag ctg gaa  agc gag ttc    3024
Val Gly Thr Ala Leu Ile Lys Lys  Tyr Pro Lys Leu Glu  Ser Glu Phe
        995                 1000                1005

gtg tac  ggc gac tac aag gtg  tac gac gtg cgg aag  atg atc gcc       3069
Val Tyr  Gly Asp Tyr Lys Val  Tyr Asp Val Arg Lys  Met Ile Ala
    1010                1015                1020

aag agc  gag cag gaa atc ggc  aag gct acc gcc aag  tac ttc ttc       3114
Lys Ser  Glu Gln Glu Ile Gly  Lys Ala Thr Ala Lys  Tyr Phe Phe
    1025                1030                1035

tac agc  aac atc atg aac ttt  ttc aag acc gag att  acc ctg gcc       3159
Tyr Ser  Asn Ile Met Asn Phe  Phe Lys Thr Glu Ile  Thr Leu Ala
    1040                1045                1050

aac ggc  gag atc cgg aag cgg  cct ctg atc gag aca  aac ggc gaa       3204
Asn Gly  Glu Ile Arg Lys Arg  Pro Leu Ile Glu Thr  Asn Gly Glu
    1055                1060                1065

acc ggg  gag atc gtg tgg gat  aag ggc cgg gat ttt  gcc acc gtg       3249
Thr Gly  Glu Ile Val Trp Asp  Lys Gly Arg Asp Phe  Ala Thr Val
    1070                1075                1080

cgg aaa  gtg ctg agc atg ccc  caa gtg aat atc gtg  aaa aag acc       3294
Arg Lys  Val Leu Ser Met Pro  Gln Val Asn Ile Val  Lys Lys Thr
    1085                1090                1095

gag gtg  cag aca ggc ggc ttc  agc aaa gag tct atc  ctg ccc aag       3339
Glu Val  Gln Thr Gly Gly Phe  Ser Lys Glu Ser Ile  Leu Pro Lys
    1100                1105                1110

agg aac  agc gat aag ctg atc  gcc aga aag aag gac  tgg gac cct       3384
```

[0143]

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

aag aag  tac ggc ggc ttc gac  agc ccc acc gtg gcc  tat tct gtg      3429
Lys Lys  Tyr Gly Gly Phe Asp  Ser Pro Thr Val Ala  Tyr Ser Val
    1130                1135                1140

ctg gtg  gtg gcc aaa gtg gaa  aag ggc aag tcc aag  aaa ctg aag      3474
Leu Val  Val Ala Lys Val Glu  Lys Gly Lys Ser Lys  Lys Leu Lys
    1145                1150                1155

agt gtg  aaa gag ctg ctg ggg  atc acc atc atg gaa  aga agc agc      3519
Ser Val  Lys Glu Leu Leu Gly  Ile Thr Ile Met Glu  Arg Ser Ser
    1160                1165                1170

ttc gag  aag aat ccc atc gac  ttt ctg gaa gcc aag  ggc tac aaa      3564
Phe Glu  Lys Asn Pro Ile Asp  Phe Leu Glu Ala Lys  Gly Tyr Lys
    1175                1180                1185

gaa gtg  aaa aag gac ctg atc  atc aag ctg cct aag  tac tcc ctg      3609
Glu Val  Lys Lys Asp Leu Ile  Ile Lys Leu Pro Lys  Tyr Ser Leu
    1190                1195                1200
```

[0144]

```
ttc gag  ctg gaa aac ggc cgg  aag aga atg ctg gcc  tct gcc ggc      3654
Phe Glu  Leu Glu Asn Gly Arg  Lys Arg Met Leu Ala  Ser Ala Gly
    1205                1210                1215

gaa ctg  cag aag gga aac gaa  ctg gcc ctg ccc tcc  aaa tat gtg      3699
Glu Leu  Gln Lys Gly Asn Glu  Leu Ala Leu Pro Ser  Lys Tyr Val
    1220                1225                1230

aac ttc  ctg tac ctg gcc agc  cac tat gag aag ctg  aag ggc tcc      3744
Asn Phe  Leu Tyr Leu Ala Ser  His Tyr Glu Lys Leu  Lys Gly Ser
    1235                1240                1245

ccc gag  gat aat gag cag aaa  cag ctg ttt gtg gaa  cag cac aag      3789
Pro Glu  Asp Asn Glu Gln Lys  Gln Leu Phe Val Glu  Gln His Lys
    1250                1255                1260

cac tac  ctg gac gag atc atc  gag cag atc agc gag  ttc tcc aag      3834
His Tyr  Leu Asp Glu Ile Ile  Glu Gln Ile Ser Glu  Phe Ser Lys
    1265                1270                1275

aga gtg  atc ctg gcc gac gct  aat ctg gac aaa gtg  ctg tcc gcc      3879
```

```
Arg Val  Ile Leu Ala Asp Ala  Asn Leu Asp Lys Val  Leu Ser Ala
    1280                 1285                 1290

tac aac  aag cac cgg gat aag  ccc atc aga gag cag  gcc gag aat      3924
Tyr Asn  Lys His Arg Asp Lys  Pro Ile Arg Glu Gln  Ala Glu Asn
    1295                 1300                 1305

atc atc  cac ctg ttt acc ctg  acc aat ctg gga gcc  cct gcc gcc      3969
Ile Ile  His Leu Phe Thr Leu  Thr Asn Leu Gly Ala  Pro Ala Ala
    1310                 1315                 1320

ttc aag  tac ttt gac acc acc  atc gac cgg aag agg  tac acc agc      4014
Phe Lys  Tyr Phe Asp Thr Thr  Ile Asp Arg Lys Arg  Tyr Thr Ser
    1325                 1330                 1335

acc aaa  gag gtg ctg gac gcc  acc ctg atc cac cag  agc atc acc      4059
Thr Lys  Glu Val Leu Asp Ala  Thr Leu Ile His Gln  Ser Ile Thr
    1340                 1345                 1350

ggc ctg  tac gag aca cgg atc  gac ctg tct cag ctg  gga ggc gac      4104
Gly Leu  Tyr Glu Thr Arg Ile  Asp Leu Ser Gln Leu  Gly Gly Asp
    1355                 1360                 1365
```

[0145]

<210> 368
<211> 1368
<212> PRT
<213> 智人

<400> 368

```
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 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

[0146]

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
[0147] 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

[0148] 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 Ala 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 Ala Lys Val Leu Thr Arg Ser Asp Lys Ala 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

[0149]

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 Ala 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

[0150]

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

[0151]

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> 369
<211> 26
<212> DNA
<213> 智人

<400> 369
tagggttagg ggccccaggc cggggt 26

<210> 370
<211> 6
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成

6xHis tag

<400> 370
His His His His His His
1 5

<210> 371
<211> 57
<212> DNA
<213> 智人

<220>
<221> CDS
<222> (1)..(57)

<400> 371
ggc acc att aaa gaa aat atc att ggt gtt tcc tat gat gaa tat aga 48
Gly Thr Ile Lys Glu Asn Ile Ile Gly Val Ser Tyr Asp Glu Tyr Arg
1 5 10 15

tac aga agc 57
[0152] Tyr Arg Ser

<210> 372
<211> 19
<212> PRT
<213> 智人

<400> 372
Gly Thr Ile Lys Glu Asn Ile Ile Gly Val Ser Tyr Asp Glu Tyr Arg
1 5 10 15

Tyr Arg Ser

<210> 373
<211> 48
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> CDS
<222> (1)..(48)

<400> 373
att aaa gaa aat atc att ggc ttt gtt tcc tat gat gaa tat aga tac 48
Ile Lys Glu Asn Ile Ile Gly Phe Val Ser Tyr Asp Glu Tyr Arg Tyr
1 5 10 15

<210> 374
<211> 16
<212> PRT
<213> 人工序列

<220>
<223> 人工序列的描述：合成肽

<400> 374
Ile Lys Glu Asn Ile Ile Gly Phe Val Ser Tyr Asp Glu Tyr Arg Tyr
1 5 10 15

[0153]
<210> 375
<211> 50
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(48)
<223> a, c, t, g, 未知的或其他的

<400> 375
ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnngg 50

<210> 376
<211> 46
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(44)
<223> a, c, t, g, 未知的或其他的

<400> 376
ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnngg 46

<210> 377
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(40)
<223> a, c, t, g, 未知的或其他的

[0154] <400> 377
ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn gg 42

<210> 378
<211> 38
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(36)
<223> a, c, t, g, 未知的或其他的

<400> 378
ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnngg 38

<210> 379
<211> 34
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(32)
<223> a, c, t, g, 未知的或其他的

<400> 379
ccnnnnnnnn nnnnnnnnnn nnnnnnnnnn nngg 34

<210> 380
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0155]
<220>
<221> 修饰的碱基
<222> (3)..(28)
<223> a, c, t, g, 未知的或其他的

<400> 380
ccnnnnnnnn nnnnnnnnnn nnnnnnnngg 30

<210> 381
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(24)
<223> a, c, t, g, 未知的或其他的

<400> 381
ccnnnnnnnn nnnnnnnnnn nnnngg 26

<210> 382
<211> 22
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(20)
<223> a, c, t, g, 未知的或其他的

<400> 382
ccnnnnnnnn nnnnnnnnnn gg 22

<210> 383
<211> 18
<212> DNA
<213> 人工序列

[0156] <220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(16)
<223> a, c, t, g, 未知的或其他的

<400> 383
ccnnnnnnnn nnnnnngg 18

<210> 384
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(14)

<223> a, c, t, g, 未知的或其他的

<400> 384
ccnnnnnnnn nnnngg 16

<210> 385
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(13)
<223> a, c, t, g, 未知的或其他的

<400> 385
ccnnnnnnnn nnngg 15

[0157]
<210> 386
<211> 14
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(12)
<223> a, c, t, g, 未知的或其他的

<400> 386
ccnnnnnnnn nngg 14

<210> 387
<211> 13
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(11)
<223> a, c, t, g, 未知的或其他的

<400> 387
ccnnnnnnnn ngg 13

<210> 388
<211> 12
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(10)
<223> a, c, t, g, 未知的或其他的

[0158] <400> 388
ccnnnnnnnn gg 12

<210> 389
<211> 11
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(9)
<223> a, c, t, g, 未知的或其他的

<400> 389
ccnnnnnnng g 11

<210> 390
<211> 10
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

<220>
<221> 修饰的碱基
<222> (3)..(8)
<223> a, c, t, g, 未知的或其他的

<400> 390
ccnnnnnngg 10

<210> 391
<211> 12
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成寡核苷酸

[0159] <220>
<221> 修饰的碱基
<222> (3)..(10)
<223> a, c, t, g, 未知的或其他的

<400> 391
ggnnnnnnnn cc 12

<210> 392
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 392
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc 120

acaag 125

<210> 393
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 393
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc 120

acaag 125

<210> 394
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0160]

<400> 394
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc 120

acaag 125

<210> 395
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 395
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc 120

acaag 125

<210> 396
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 396
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgtggcgc attgccacga agcggtgttt cgtcctttcc 120

acaag 125

<210> 397
<211> 125
<212> DNA
<213> 人工序列

[0161]
<220>
<223> 人工序列的描述：合成引物

<400> 397
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccgagggc agagtgctgc ttgggtgttt cgtcctttcc 120

acaag 125

<210> 398
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 398
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgagtccg agcagaagaa gaaggtgttt cgtcctttcc 120

acaag 125

<210> 399
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 399
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc 120

acaag 125

<210> 400
<211> 125
<212> DNA
<213> 人工序列

[0162]

<220>
<223> 人工序列的描述：合成引物

<400> 400
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacagcagaa gaagaagggc tccggtgttt cgtcctttcc 120

acaag 125

<210> 401
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 401
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcacatc aaccggtggc gcaggtgttt cgtcctttcc 120

acaag 125

<210> 402
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 402
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacccctggc ccaggtgaag gtgggtgttt cgtcctttcc 120

acaag 125

<210> 403
<211> 125
<212> DNA
<213> 人工序列

[0163]

<220>
<223> 人工序列的描述：合成引物

<400> 403
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactccctcc ctggcccagg tgaggtgttt cgtcctttcc 120

acaag 125

<210> 404
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 404
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc 120

acaag 125

<210> 405
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 405
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc 120

acaag 125

<210> 406
<211> 125
<212> DNA
[0164] <213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 406
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacggtgaag gtgtggttcc agaggtgttt cgtcctttcc 120

acaag 125

<210> 407
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 407
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc 120

acaag 125

<210> 408
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 408
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacccctggc ccaggtgaag gtgggtgttt cgtcctttcc 120

acaag 125

<210> 409
<211> 125
[0165] <212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 409
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc 120

acaag 125

<210> 410
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 410

aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaggaca aagtacaaac ggcggtgttt cgtcctttcc 120

acaag 125

<210> 411
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 411
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc 120

acaag 125

[0166] <210> 412
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 412
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccaccttc acctgggcca gggggtgttt cgtcctttcc 120

acaag 125

<210> 413
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 413
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacaccctag tcattggagg tgaggtgttt cgtcctttcc 120

acaag 125

<210> 414
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 414
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccagagca gccactgggg cctggtgttt cgtcctttcc 120

acaag 125

[0167]

<210> 415
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 415
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccaccttc acctgggcca gggggtgttt cgtcctttcc 120

acaag 125

<210> 416
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 416
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacccccatt ggcctgcttc gtgggtgttt cgtcctttcc 120

acaag 125

<210> 417
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 417
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacattggcc tgcttcgtgg caaggtgttt cgtcctttcc 120

acaag 125

[0168]

<210> 418
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 418
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcctcct ccagcttctg ccgggtgttt cgtcctttcc 120

acaag 125

<210> 419
<211> 125
<212> DNA
<213> 人工序列

<220>

<223> 人工序列的描述：合成引物

<400> 419
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccctccag cttctgccgt ttgggtgttt cgtcctttcc 120

acaag 125

<210> 420
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 420
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacattggcc tgcttcgtgg caaggtgttt cgtcctttcc 120

[0169]

acaag 125

<210> 421
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 421
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgcagcaa gcagcactct gccggtgttt cgtcctttcc 120

acaag 125

<210> 422
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 422
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacttcttct tctgctcgga ctcggtgttt cgtcctttcc 120

acaag 125

<210> 423
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 423
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

[0170] cttgctattt ctagctctaa aacaccggag gacaaagtac aaaggtgttt cgtcctttcc 120

acaag 125

<210> 424
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 424
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcttctt ctgctcggac tcaggtgttt cgtcctttcc 120

acaag 125

<210> 425
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 425
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgttgatg tgatgggagc cctggtgttt cgtcctttcc 120

acaag 125

<210> 426
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 426
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

[0171] cttgctattt ctagctctaa aacgggccag ggagggaggg gcaggtgttt cgtcctttcc 120

acaag 125

<210> 427
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 427
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc 120

acaag 125

<210> 428
<211> 125
<212> DNA

<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 428
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacccggttc tggaaccaca cctggtgttt cgtcctttcc 120

acaag 125

<210> 429
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 429
[0172] aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcacctg ggccagggag ggaggtgttt cgtcctttcc 120

acaag 125

<210> 430
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 430
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aactcacctg ggccagggag ggaggtgttt cgtcctttcc 120

acaag 125

<210> 431
<211> 125

<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 431
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgttctgg aaccacacct tcaggtgttt cgtcctttcc 120

acaag 125

<210> 432
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0173]
<400> 432
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgggaggg aggggcacag atgggtgttt cgtcctttcc 120

acaag 125

<210> 433
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 433
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgggccag ggagggaggg gcaggtgttt cgtcctttcc 120

acaag 125

<210> 434

<210> 125
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 434
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgttctgg aaccacacct tcaggtgttt cgtcctttcc 120

acaag 125

<210> 435
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0174]

<400> 435
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacaggtgaa ggtgtggttc cagggtgttt cgtcctttcc 120

acaag 125

<210> 436
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 436
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgaaccgg aggacaaagt acaggtgttt cgtcctttcc 120

acaag 125

<210> 437
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

<400> 437
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aaccaaaccc acgagggcag agtggtgttt cgtcctttcc 120

acaag 125

<210> 438
<211> 125
<212> DNA
<213> 人工序列

<220>
<223> 人工序列的描述：合成引物

[0175]

<400> 438
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa 60

cttgctattt ctagctctaa aacgagtttc tcatctgtgc cccggtgttt cgtcctttcc 120

acaag 125

<210> 439
<211> 684
<212> DNA
<213> 智人

<220>
<221> CDS
<222> (1)..(123)

<220>
<221> CDS
<222> (127)..(159)

<220>

<221> CDS
<222> (163)..(399)

<220>
<221> CDS
<222> (403)..(684)

<400> 439

```
aaa acc acc ctt ctc tct ggc cca ctg tgt cct ctt cct gcc ctg cca        48
Lys Thr Thr Leu Leu Ser Gly Pro Leu Cys Pro Leu Pro Ala Leu Pro
1               5                  10                  15

tcc cct tct gtg aat gtt aga ccc atg gga gca gct ggt cag agg gga        96
Ser Pro Ser Val Asn Val Arg Pro Met Gly Ala Ala Gly Gln Arg Gly
            20                  25                  30

ccc cgg cct ggg gcc cct aac cct atg tag cct cag tct tcc cat cag       144
Pro Arg Pro Gly Ala Pro Asn Pro Met     Pro Gln Ser Ser His Gln
        35                  40                  45

gct ctc agc tca gcc tga gtg ttg agg ccc cag tgg ctg ctc tgg ggg       192
Ala Leu Ser Ser Ala     Val Leu Arg Pro Gln Trp Leu Leu Trp Gly
        50                  55                  60

cct cct gag ttt ctc atc tgt gcc cct ccc tcc ctg gcc cag gtg aag       240
Pro Pro Glu Phe Leu Ile Cys Ala Pro Pro Ser Leu Ala Gln Val Lys
        65                  70                  75

gtg tgg ttc cag aac cgg agg aca aag tac aaa cgg cag aag ctg gag       288
Val Trp Phe Gln Asn Arg Arg Thr Lys Tyr Lys Arg Gln Lys Leu Glu
    80                  85                  90

gag gaa ggg cct gag tcc gag cag aag aag aag ggc tcc cat cac atc       336
Glu Glu Gly Pro Glu Ser Glu Gln Lys Lys Lys Gly Ser His His Ile
95                  100                 105                 110

aac cgg tgg cgc att gcc acg aag cag gcc aat ggg gag gac atc gat       384
Asn Arg Trp Arg Ile Ala Thr Lys Gln Ala Asn Gly Glu Asp Ile Asp
                115                 120                 125

gtc acc tcc aat gac tag ggt ggg caa cca caa acc cac gag ggc aga       432
Val Thr Ser Asn Asp     Gly Gly Gln Pro Gln Thr His Glu Gly Arg
            130                 135                 140
```

[0176]

gtg ctg ctt gct gct ggc cag gcc cct gcg tgg gcc caa gct gga ctc 480
Val Leu Leu Ala Ala Gly Gln Ala Pro Ala Trp Ala Gln Ala Gly Leu
145 150 155

tgg cca ctc cct ggc cag gct ttg ggg agg cct gga gtc atg gcc cca 528
Trp Pro Leu Pro Gly Gln Ala Leu Gly Arg Pro Gly Val Met Ala Pro
160 165 170

cag ggc ttg aag ccc ggg gcc gcc att gac aga ggg aca agc aat ggg 576
Gln Gly Leu Lys Pro Gly Ala Ala Ile Asp Arg Gly Thr Ser Asn Gly
175 180 185

ctg gct gag gcc tgg gac cac ttg gcc ttc tcc tcg gag agc ctg cct 624
Leu Ala Glu Ala Trp Asp His Leu Ala Phe Ser Ser Glu Ser Leu Pro
190 195 200 205

gcc tgg gcg ggc ccg ccc gcc acc gca gcc tcc cag ctg ctc tcc gtg 672
Ala Trp Ala Gly Pro Pro Ala Thr Ala Ala Ser Gln Leu Leu Ser Val
210 215 220

tct cca atc tcc 684
[0177] Ser Pro Ile Ser
225

<210> 440
<211> 41
<212> PRT
<213> 智人

<400> 440
Lys Thr Thr Leu Leu Ser Gly Pro Leu Cys Pro Leu Pro Ala Leu Pro
1 5 10 15

Ser Pro Ser Val Asn Val Arg Pro Met Gly Ala Ala Gly Gln Arg Gly
20 25 30

Pro Arg Pro Gly Ala Pro Asn Pro Met
35 40

<210> 441
<211> 11
<212> PRT
<213> 智人

<400> 441
Pro Gln Ser Ser His Gln Ala Leu Ser Ser Ala
1 5 10

<210> 442
<211> 79
<212> PRT
<213> 智人

<400> 442
Val Leu Arg Pro Gln Trp Leu Leu Trp Gly Pro Pro Glu Phe Leu Ile
1 5 10 15

Cys Ala Pro Pro Ser Leu Ala Gln Val Lys Val Trp Phe Gln Asn Arg
20 25 30

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

Glu Gln Lys Lys Lys Gly Ser His His Ile Asn Arg Trp Arg Ile Ala
50 55 60

[0178]

Thr Lys Gln Ala Asn Gly Glu Asp Ile Asp Val Thr Ser Asn Asp
65 70 75

<210> 443
<211> 94
<212> PRT
<213> 智人

<400> 443
Gly Gly Gln Pro Gln Thr His Glu Gly Arg Val Leu Leu Ala Ala Gly
1 5 10 15

Gln Ala Pro Ala Trp Ala Gln Ala Gly Leu Trp Pro Leu Pro Gly Gln
20 25 30

Ala Leu Gly Arg Pro Gly Val Met Ala Pro Gln Gly Leu Lys Pro Gly
35 40 45

Ala Ala Ile Asp Arg Gly Thr Ser Asn Gly Leu Ala Glu Ala Trp Asp
50 55 60

His Leu Ala Phe Ser Ser Glu Ser Leu Pro Ala Trp Ala Gly Pro Pro

65 70 75 80

[0179] Ala Thr Ala Ala Ser Gln Leu Leu Ser Val Ser Pro Ile Ser
85 90

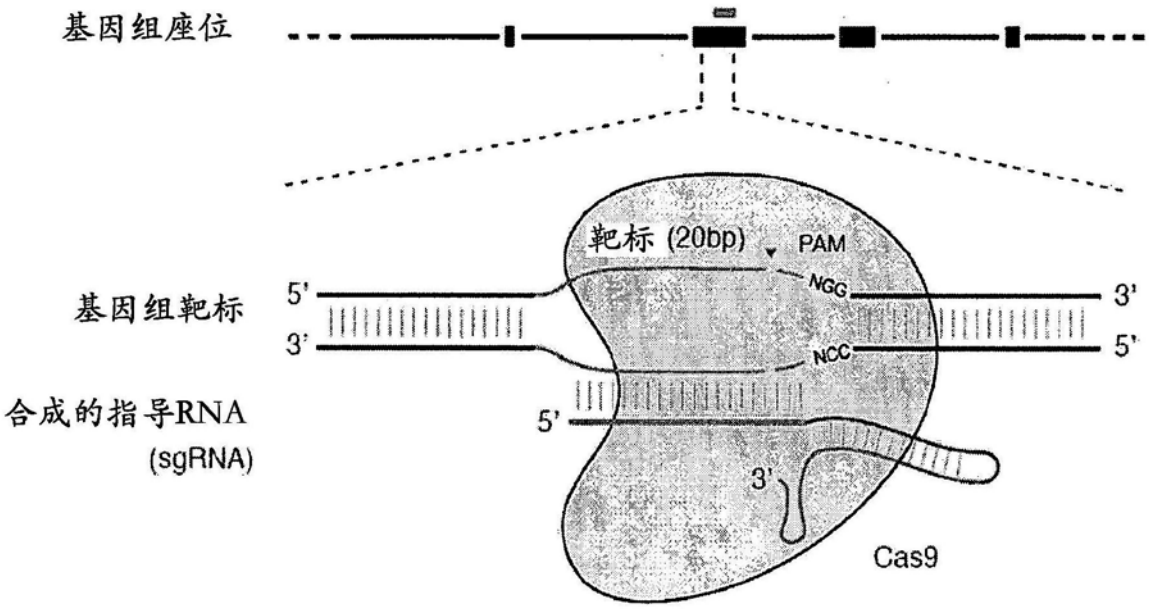


图1

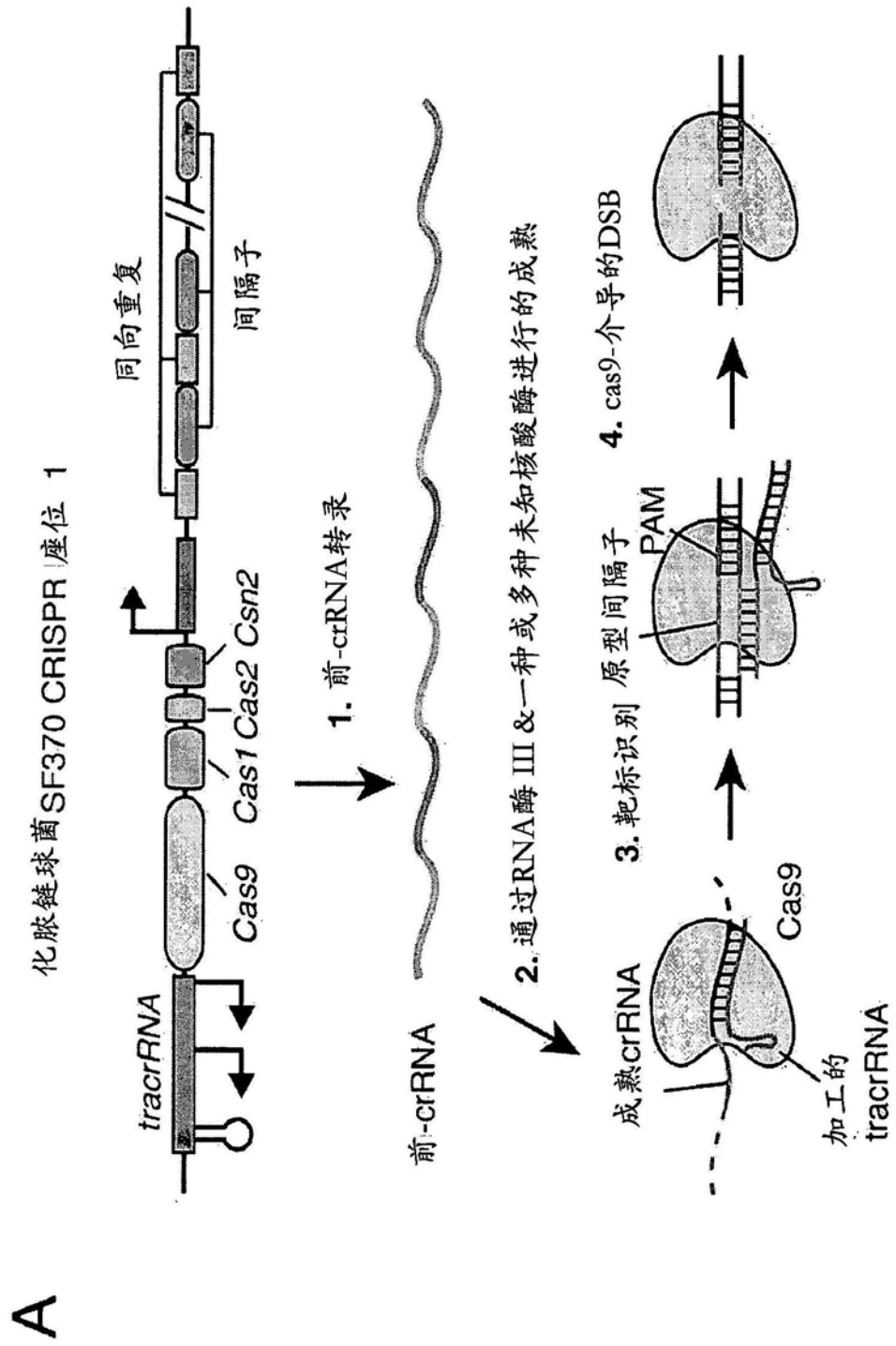

A
化脓链球菌SF370 CRISPR 座位 1
同向重复
间隔子
tracrRNA
Cas9
Cas1
Cas2
Csn2
1. 前-crRNA转录
前-crRNA
2. 通过RNA酶 III &一种或多种未知核酸酶进行的成熟
成熟crRNA
3. 靶标识别
原型间隔子
PAM
4. cas9-介导的DSB
Cas9
加工的
tracrRNA

图2A

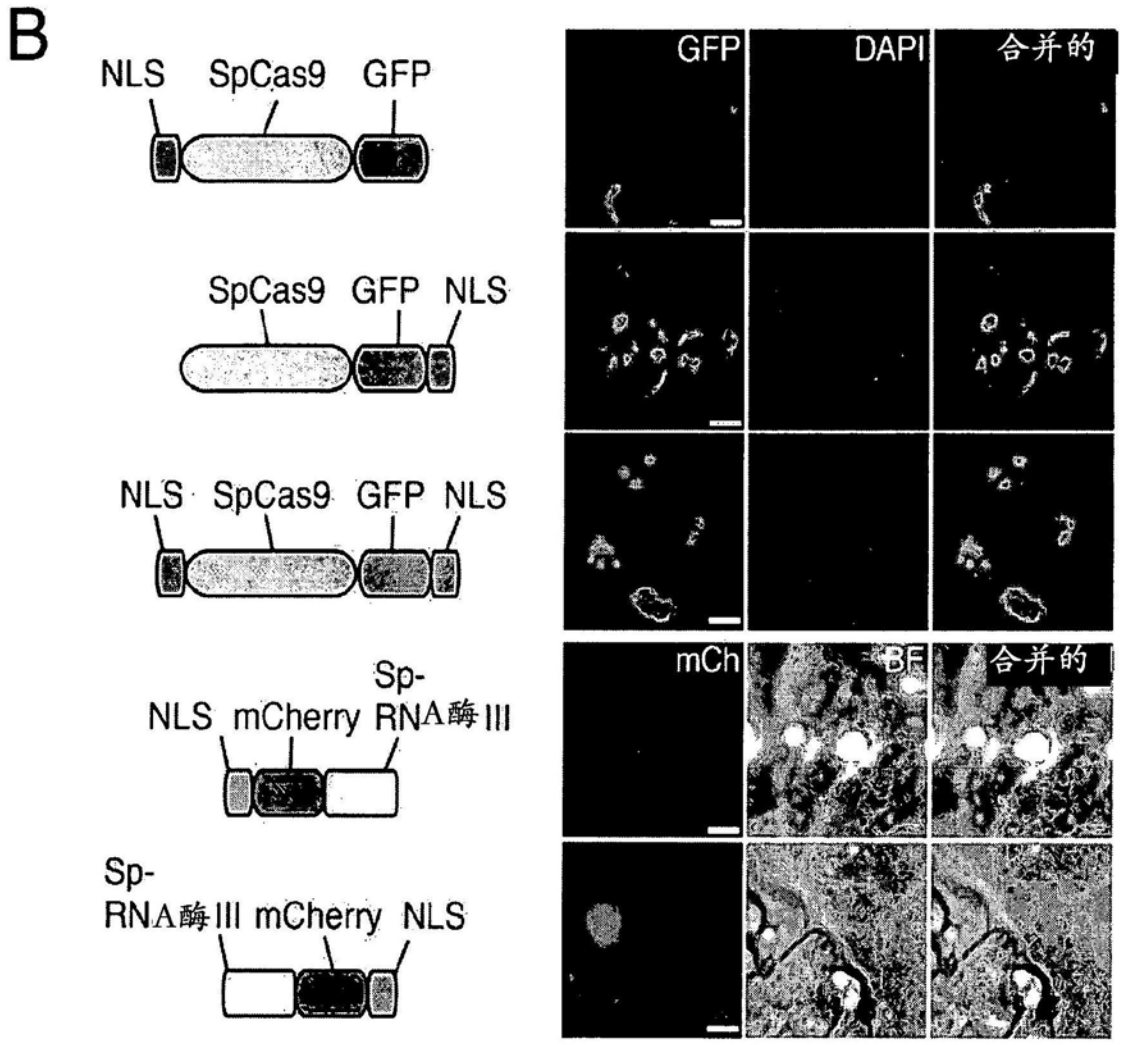

B
NLS SpCas9 GFP
SpCas9 GFP NLS
NLS SpCas9 GFP NLS
NLS mCherry Sp-RNA酶III
Sp-RNA酶III mCherry NLS
GFP
DAPI
合并的
mCh
BF
合并的

图2B

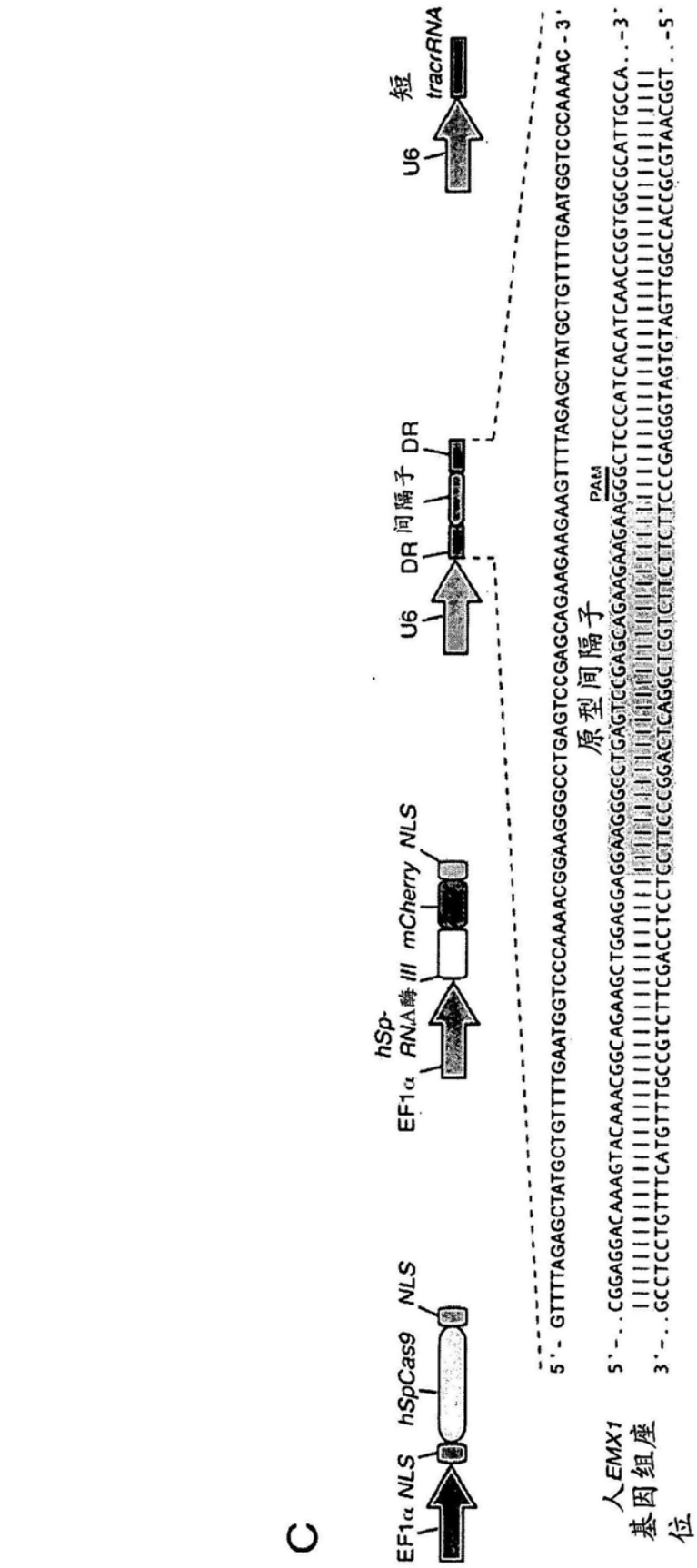

C
EF1α NLS hSpCas9 NLS
hSp-
EF1α RNA酶 III mCherry NLS
U6 DR 间隔子 DR
U6 短 tracrRNA
5'- GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC - 3'
原型间隔子
PAM
人EMX1 基因组座位
5'-..CGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCA..-3'
3'-..GCCTCCTGTTTCATGTTTGCCGTCTTCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTAGTGTAGTTGGCCACCGCGTAACGGT..-5'

图2C

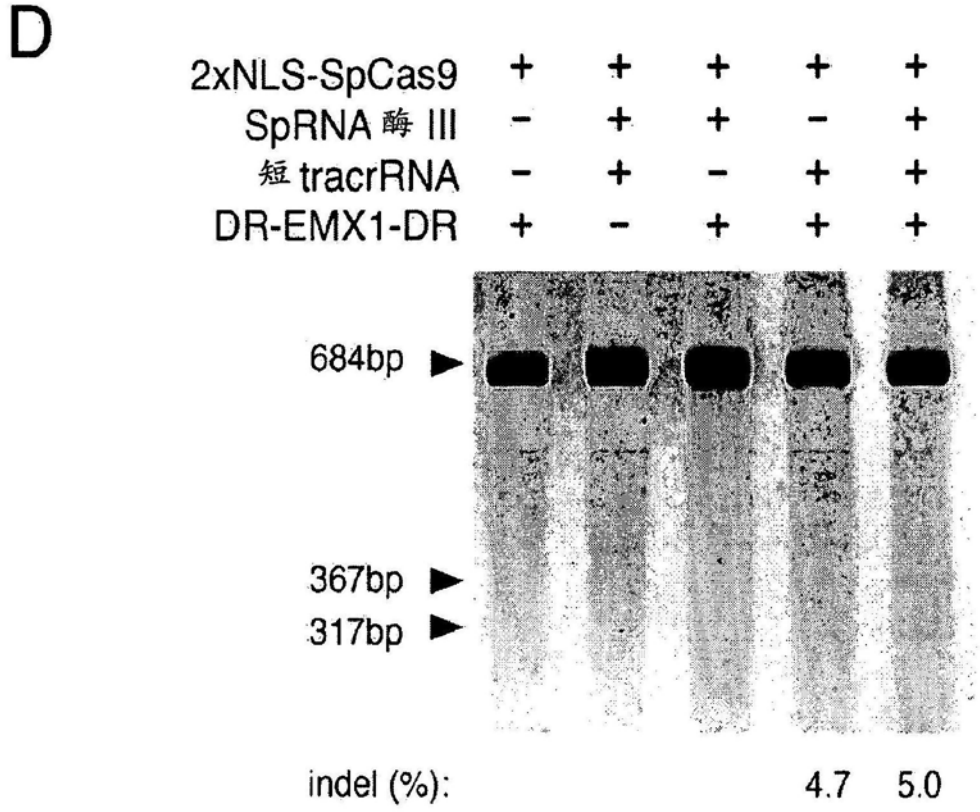

D
2xNLS-SpCas9 + + + + +
SpRNA 酶 III - + + - +
短 tracrRNA - + - + +
DR-EMX1-DR + - + + +
684bp
367bp
317bp
indel (%): 4.7 5.0

图2D

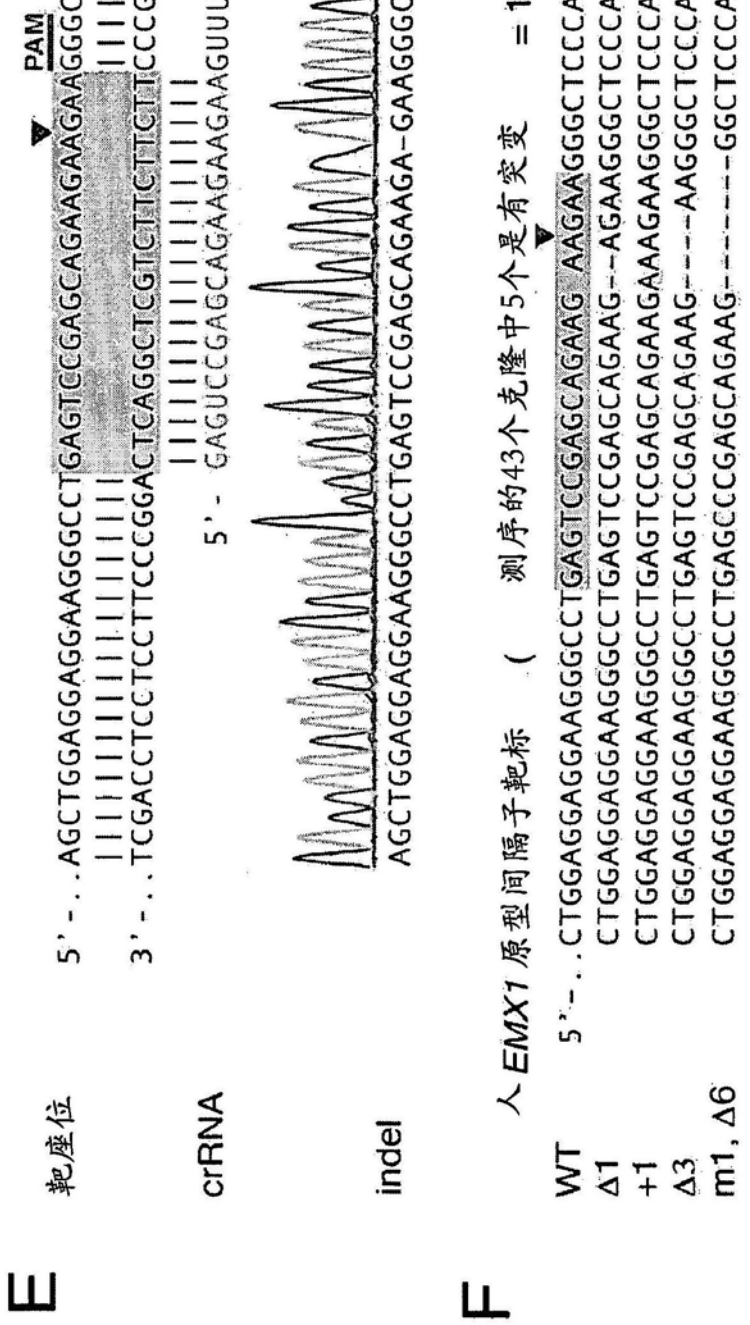

E
靶座位
PAM
5'-...AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGGGCTCCCAC...-3'
3'-...TCGACCTCCTCCTTCCCGGACTCAGGCTCGTCGTCTTCTTCCCGAGGGTG...-5'
crRNA
5'- GAGUCCGAGCAGAAGAAGUUUUAGAGC...-3'
indel
AGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGA-GAAGGGCTCCCAT
F
人 EMX1 原型间隔子靶标 (测序的43个克隆中5个是有突变 = 11.6%)
WT 5'-...CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACAT...-3'
Δ1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG--AGAAGGGCTCCCATCACAT
+1 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACAT
Δ3 CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG---AAGGGCTCCCATCACAT
m1, Δ6 CTGGAGGAGGAAGGGCCTGAGCCCGAGCAGAAG------GGCTCCCATCACAT

图2E-F

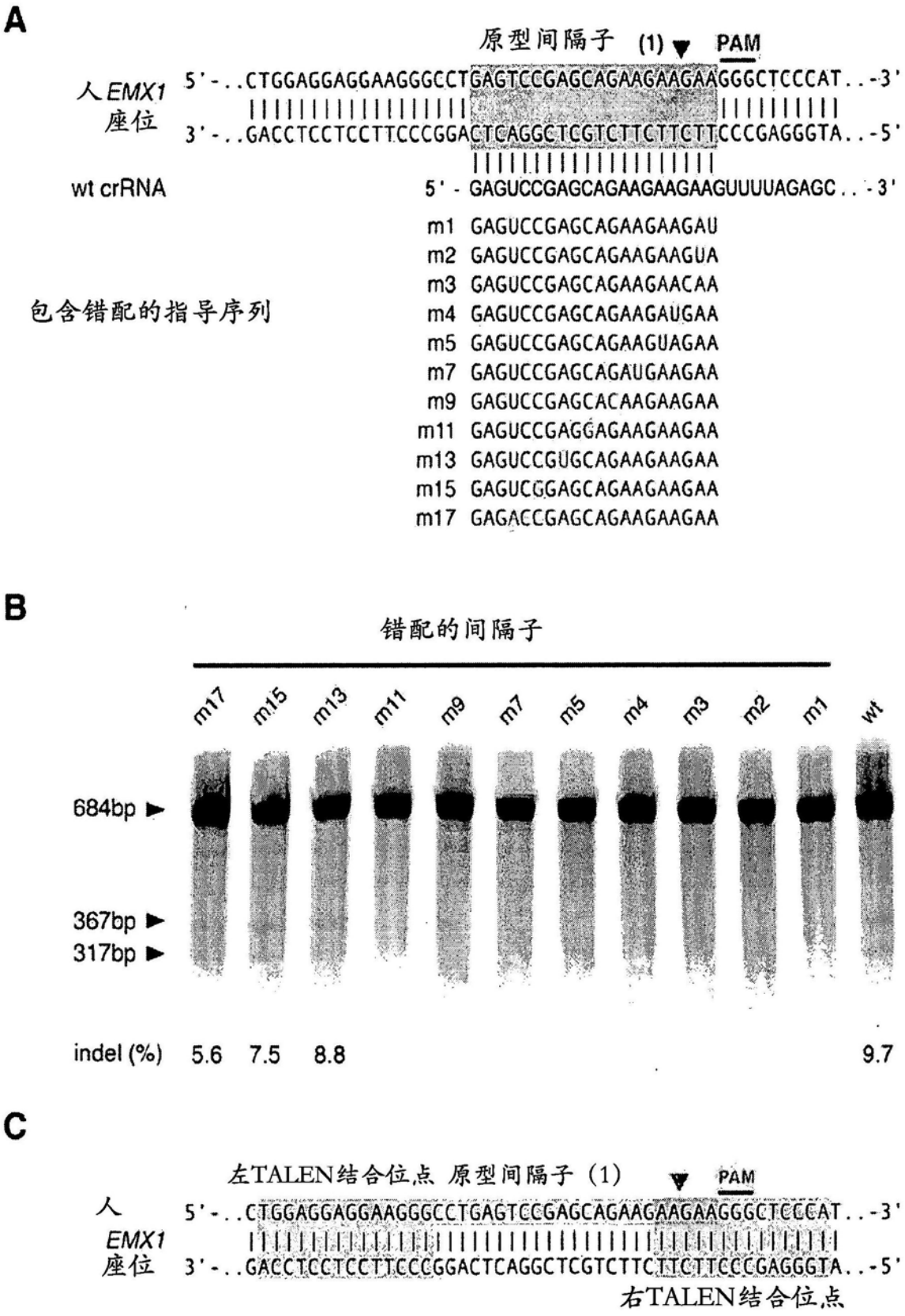

A
原型间隔子 (1) PAM
人EMX1座位
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
wt crRNA 5'-GAGUCCGAGCAGAAGAAGAAGUUUUAGAGC..-3'
包含错配的指导序列
m1 GAGUCCGAGCAGAAGAAGAU
m2 GAGUCCGAGCAGAAGAAGUA
m3 GAGUCCGAGCAGAAGAACAA
m4 GAGUCCGAGCAGAAGAUGAA
m5 GAGUCCGAGCAGAAGUAGAA
m7 GAGUCCGAGCAGAUGAAGAA
m9 GAGUCCGAGCACAAGAAGAA
m11 GAGUCCGAGGAGAAGAAGAA
m13 GAGUCCGUGCAGAAGAAGAA
m15 GAGUCCGAGCAGAAGAAGAA
m17 GAGACCGAGCAGAAGAAGAA
B
错配的间隔子
m17 m15 m13 m11 m9 m7 m5 m4 m3 m2 m1 wt
684bp
367bp
317bp
indel (%) 5.6 7.5 8.8 9.7
C
左TALEN结合位点 原型间隔子 (1) PAM
人EMX1座位
5'-..CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCAT..-3'
3'-..GACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTA..-5'
右TALEN结合位点

图3A-C

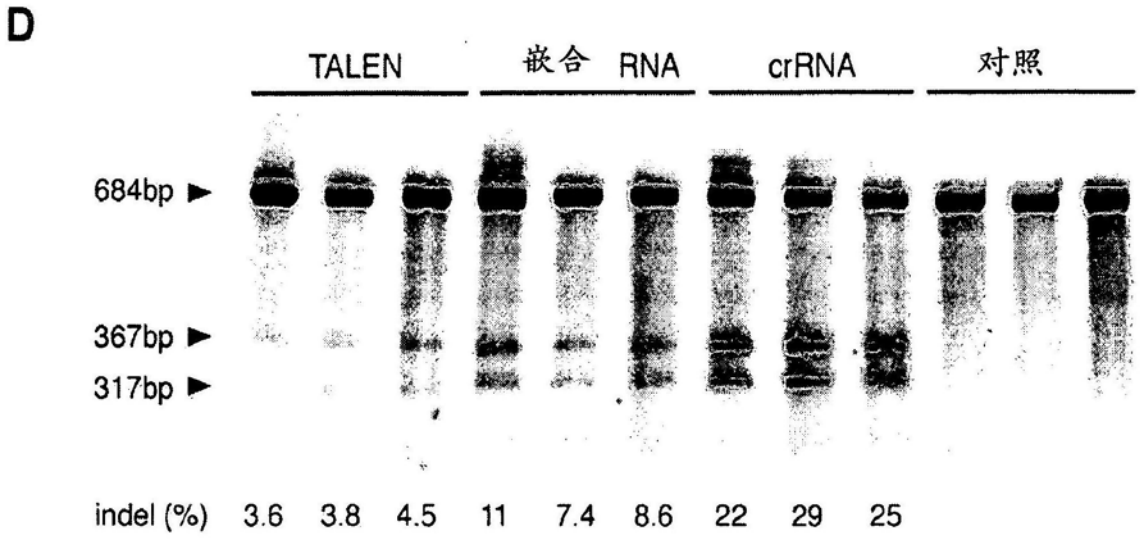

D
TALEN
嵌合 RNA
crRNA
对照
684bp
367bp
317bp
indel (%) 3.6 3.8 4.5 11 7.4 8.6 22 29 25

图3D

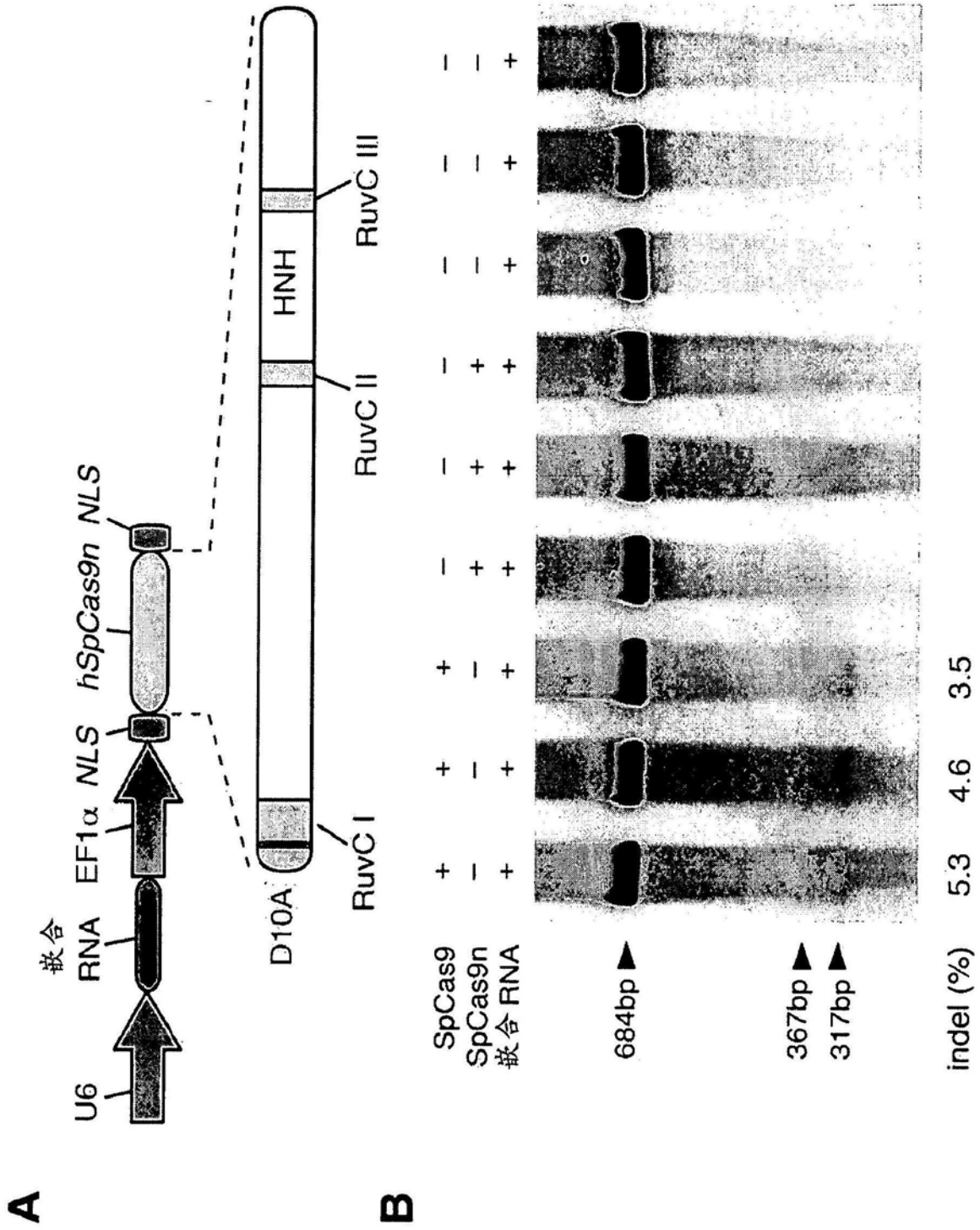

A
U6
嵌合 RNA
EF1α
NLS
hSpCas9n
NLS
D10A
RuvC I
RuvC II
HNH
RuvC III
B
SpCas9
SpCas9n
嵌合 RNA
684bp
367bp
317bp
indel (%)
5.3
4.6
3.5

图4A-B

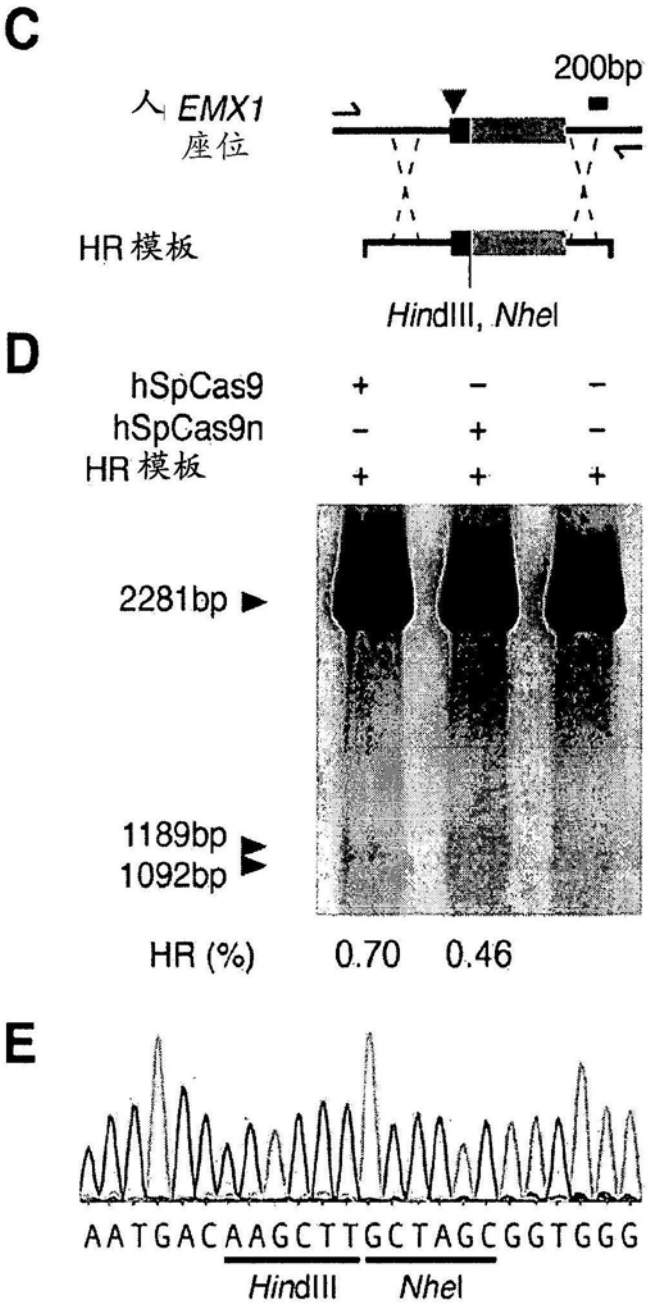

C
200bp
人 EMX1
座位
HR 模板
HindIII, NheI
D
hSpCas9 + − −
hSpCas9n − + −
HR 模板 + + +
2281bp
1189bp
1092bp
HR (%) 0.70 0.46
E
AATGACAAGCTTGCTAGCGGTGGG
HindIII
NheI

图4C-E

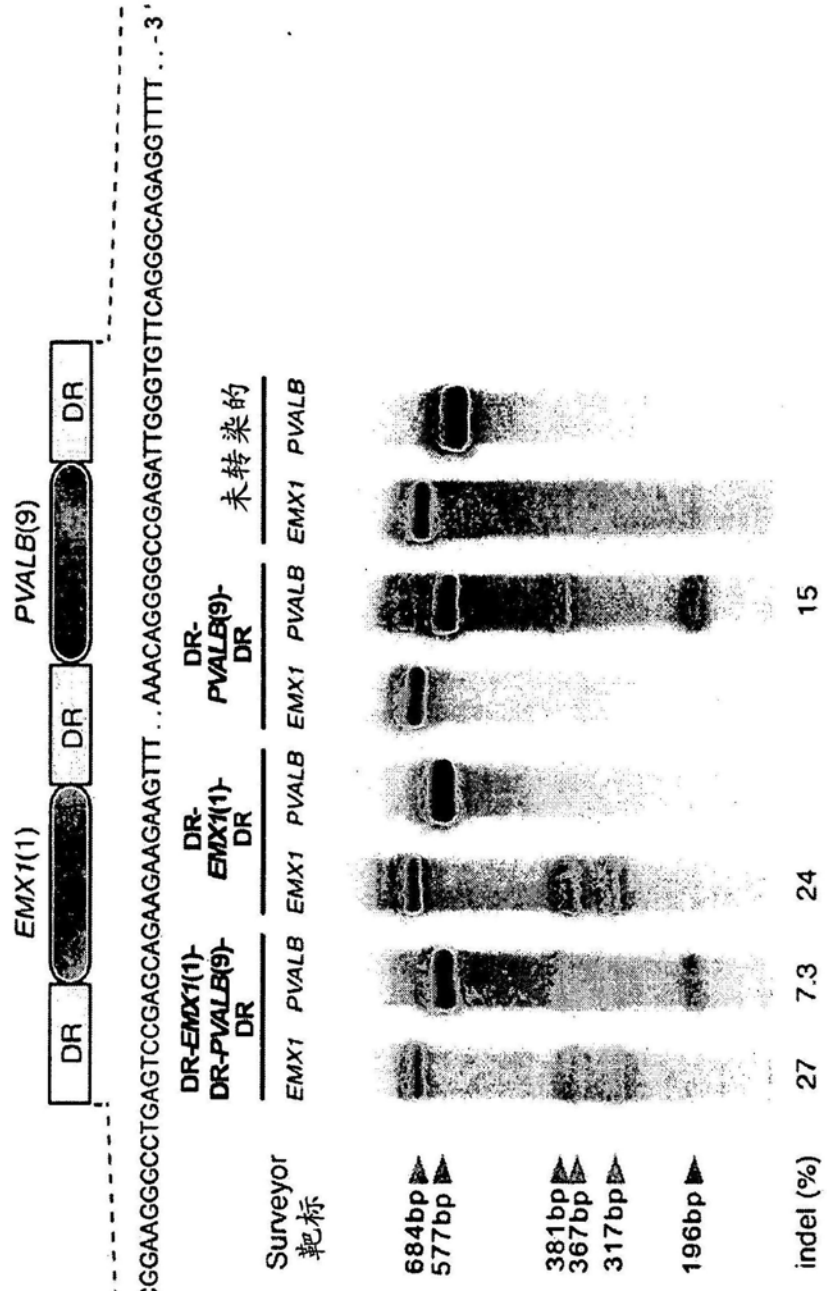

F
crRNA阵列设计
EMX1(1)
PVALB(9)
DR
DR
DR
DR
5'... AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTT ..AAACAGGGGCCGAGATTGGGTGTTCAGGGCAGAGGTTTT ...3'
DR-EMX1(1)-DR-PVALB(9)-DR
DR-EMX1(1)-DR
DR-PVALB(9)-DR
未转染的
EMX1 PVALB EMX1 PVALB EMX1 PVALB EMX1 PVALB
Surveyor 靶标
684bp
577bp
381bp
367bp
317bp
196bp
indel (%)
27
7.3
24
15

图4F

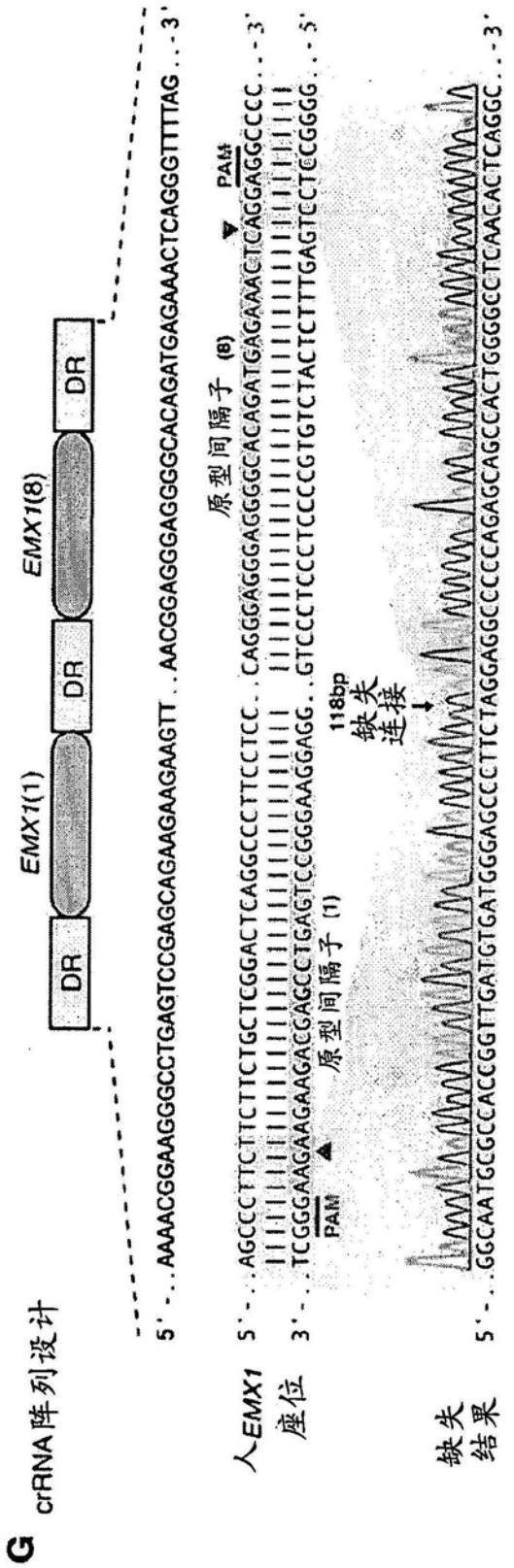

G
crRNA阵列设计
EMX1(1)
EMX1(8)
DR
DR
DR
5'-...AAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTT...AACGGAGGGAGGGGCACAGATGAGAAACTCAGGGTTTTAG...-3'
原型间隔子 (8)
PAM
人EMX1
座位
5'-...AGCCCTTCTTCTTCTGCTCGGACTCAGGCCCTTCCTCC...CAGGGAGGGAGGGGCACAGATGAGAAACTCAGGAGGCCCC...-3'
3'-...TCGGGAAGAAGAAGACGAGCCTGAGTCCGGGAAGGAGG...GTCCCTCCCTCCCCGTGTCTACTCTTTGAGTCCTCCGGGG...-5'
PAM
原型间隔子 (1)
118bp
缺失
连接
缺失
结果
5'-...GGCAATGCGGCCACCGGTTGATGTGATGGGAGCCCTTCTAGGAGGCCCCCAGAGCCAGCCACTGGGGCCTCAACACTCAGGC...-3'

图4G

Cas9	靶标物种	基因	原型间隔子 ID	原型间隔子序列(5' to 3')	PAM	链	测试的细胞系	% indel (前-crRNA + tracrRNA)	% indel (嵌合 RNA)
化脓链球菌 SF370 II型 CRISPR	智人	*EMX1*	1	GGAAGGGCCTGAGTCCGAGCAGAAGAAGAA	GGG	+	293FT	20 ± 1.8	6.7 ± 0.62
		EMX1	2	CATTGGAGGTGACATCGATGTCCTCCCCAT	TGG	–	293FT	2.1 ± 0.31	N.D.
		EMX1	3	GGACATCGATGTCACCTCCAATGACTAGGG	TGG	+	293FT	14 ± 1.1	N.D.
		EMX1	4	CATCGATGTCCTCCCCATTGGCCTGCTTCG	TGG	–	293FT	11 ± 1.7	N.D.
		EMX1	5	TTCGTGGCAATGCGCCACCGGTTGATGTGA	TGG	–	293FT	4.3 ± 0.46	2.1 ± 0.51
		EMX1	6	TCGTGGCAATGCGCCACCGGTTGATGTGAT	GGG	–	293FT	4.0 ± 0.66	0.41 ± 0.25
		EMX1	7	TCCAGCTTCTGCCGTTTGTACTTTGTCCTC	CGG	–	293FT	1.5 ± 0.12	N.D.
		EMX1	8	GGAGGGAGGGGCACAGATGAGAAACTCAGG	AGG	–	293FT	7.8 ± 0.83	2.3 ± 1.2
	智人	*PVALB*	9	AGGGCCGAGATTGGGTGTTCAGGGCAGAG	AGG	+	293FT	21 ± 2.6	6.5 ± 0.32
		PVALB	10	ATGCAGGAGGGTGGCGAGAGGGGCCGAGAT	TGG	+	293FT	N.D.	N.D.
		PVALB	11	GGTGGCGAGAGGGGCCGAGATTGGGTGTTC	AGG	+	293FT	N.D.	N.D.
	小鼠	*Th*	12	CAAGCACTGAGTGCCATTAGCTAAATGCAT	AGG	–	Neuro2A	27 ± 4.3	4.1 ± 2.2
		Th	13	AATGCATAGGGTACCACCCACAGGTGCCAG	GGG	–	Neuro2A	4.8 ± 1.2	N.D.
		Th	14	ACACACATGGGAAAGCCTCTGGGCCAGGAA	AGG	+	Neuro2A	11.3 ± 1.3	N.D.
嗜热链球菌 *LMD-9 CRISPR1*	智人	*EMX1*	15	GGAGGAGGTAGTATACAGAAACACAGAGAA	GTAGAAT	–	293FT	14 ± 0.88	N.T.
		EMX1	16	AGAATGTAGAGGAGTCACAGAAACTCAGCA	CTAGAAA	–	293FT	7.8 ± 0.77	N.T.

图5

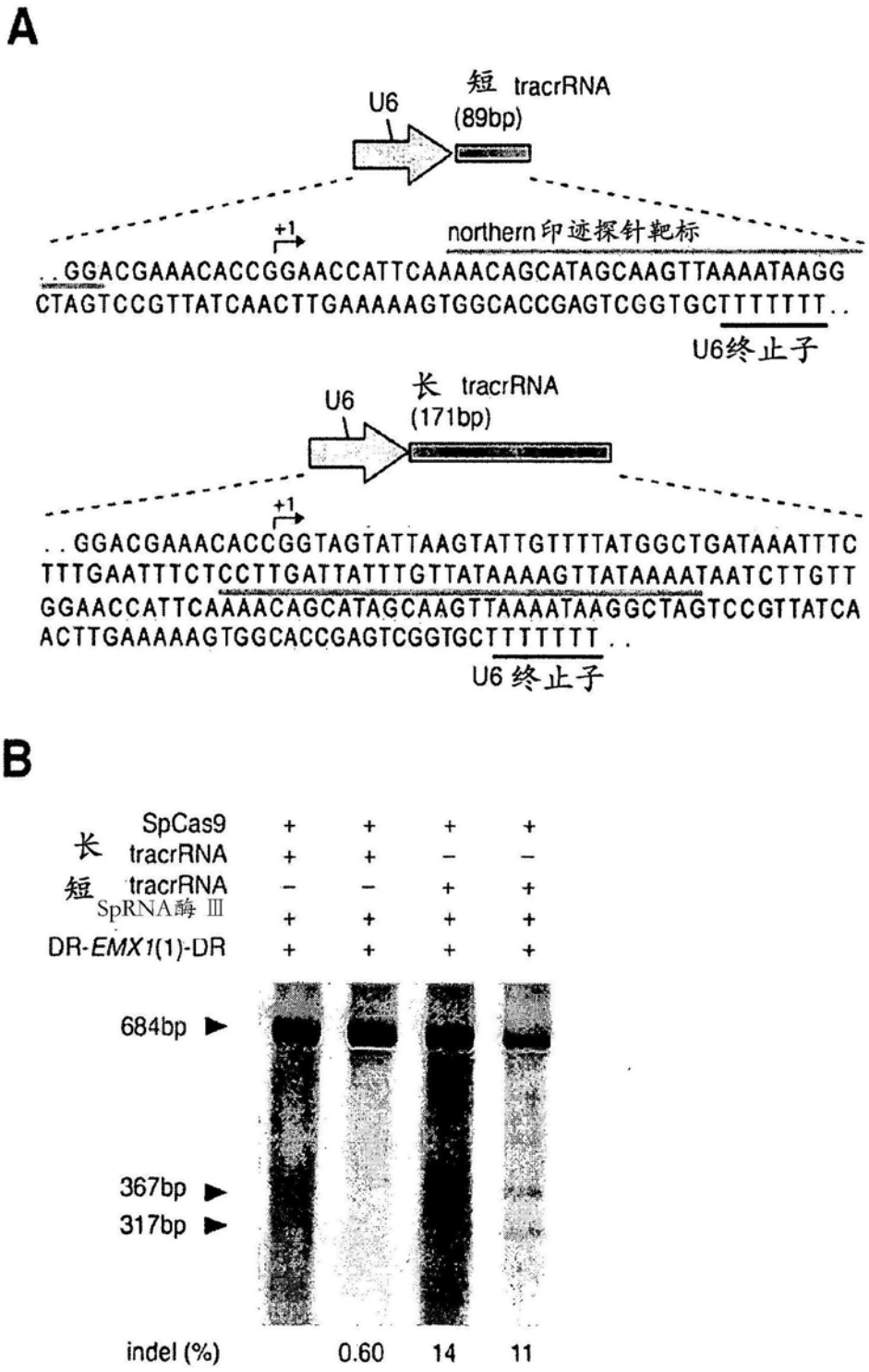

A
U6
短 tracrRNA
(89bp)
+1
northern印迹探针靶标
..GGACGAAACACCGGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGG
CTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT..
U6终止子
U6
长 tracrRNA
(171bp)
+1
..GGACGAAACACCGGTAGTATTAAGTATTGTTTTATGGCTGATAAATTTC
TTTGAATTTCTCCTTGATTATTTGTTATAAAAGTTATAAAATAATCTTGTT
GGAACCATTCAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATCA
ACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTT..
U6 终止子
B
SpCas9 + + + +
长 tracrRNA + + − −
短 tracrRNA − − + +
SpRNA酶 III + + + +
DR-EMX1(1)-DR + + + +
684bp
367bp
317bp
indel (%) 0.60 14 11

图6A-B

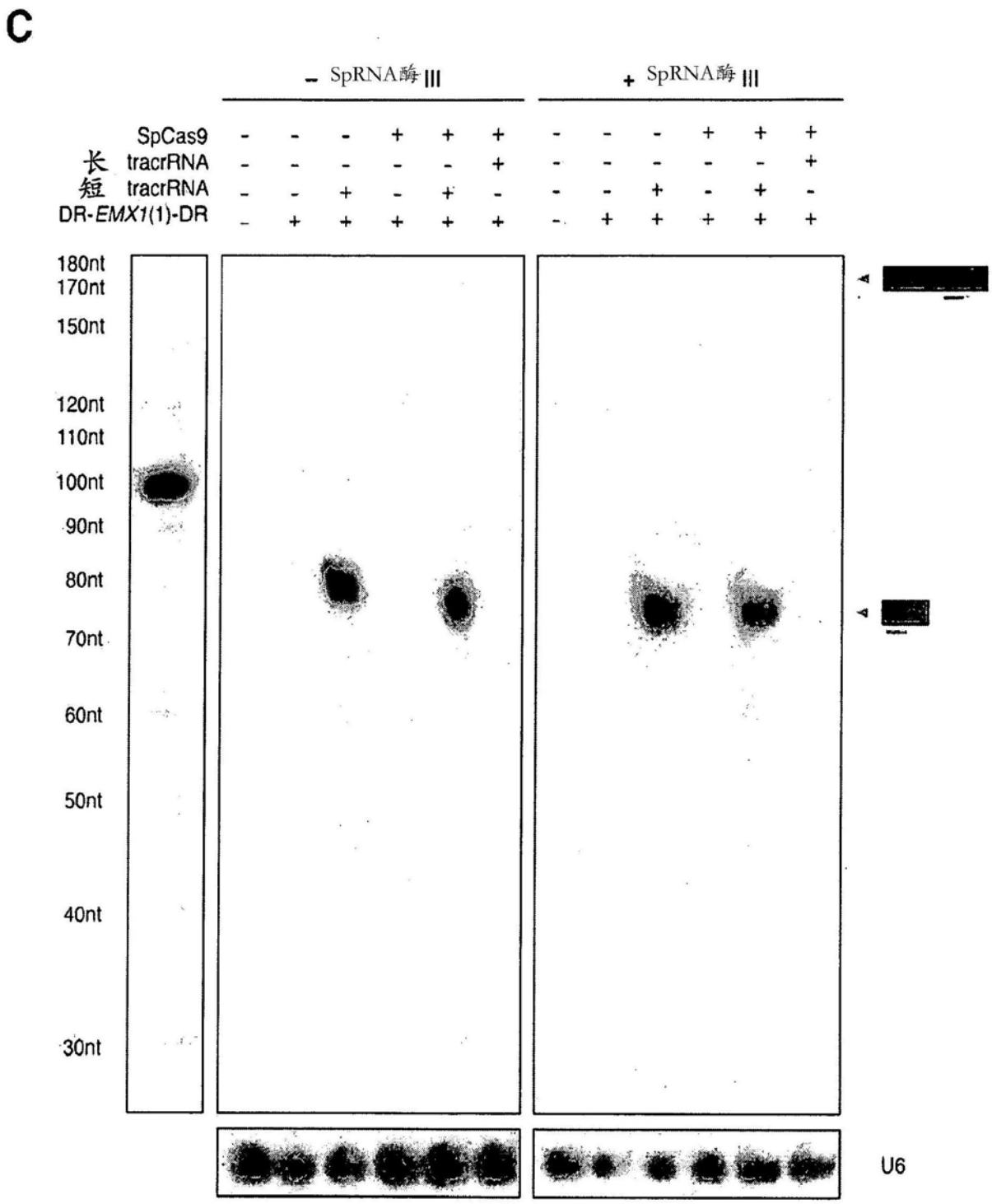

C
- SpRNA酶 III
+ SpRNA酶 III
SpCas9
长 tracrRNA
短 tracrRNA
DR-EMX1(1)-DR
180nt
170nt
150nt
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

图6C

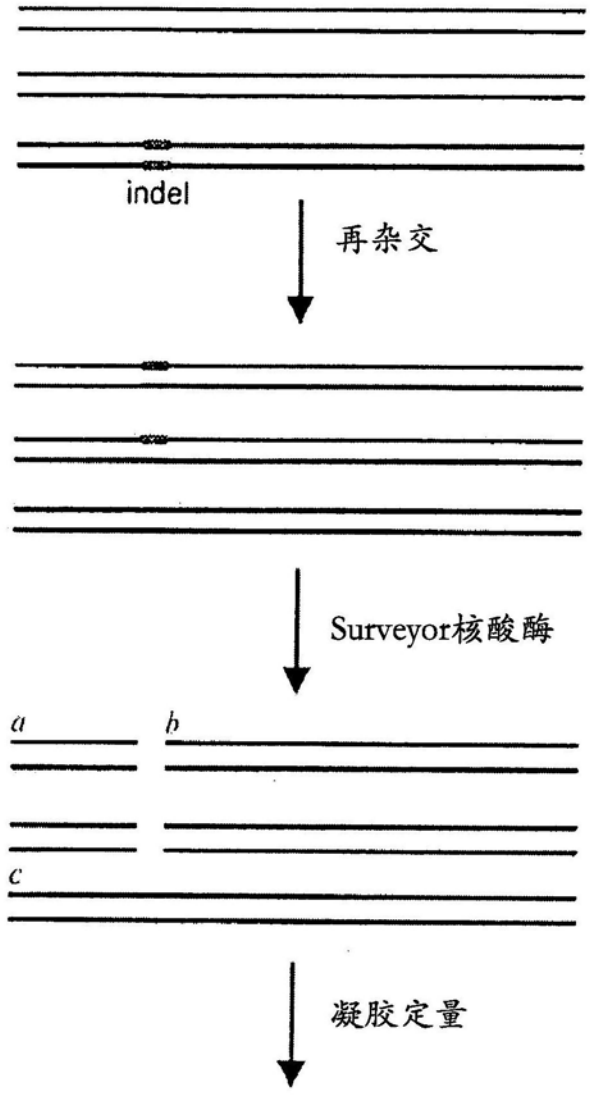


图7

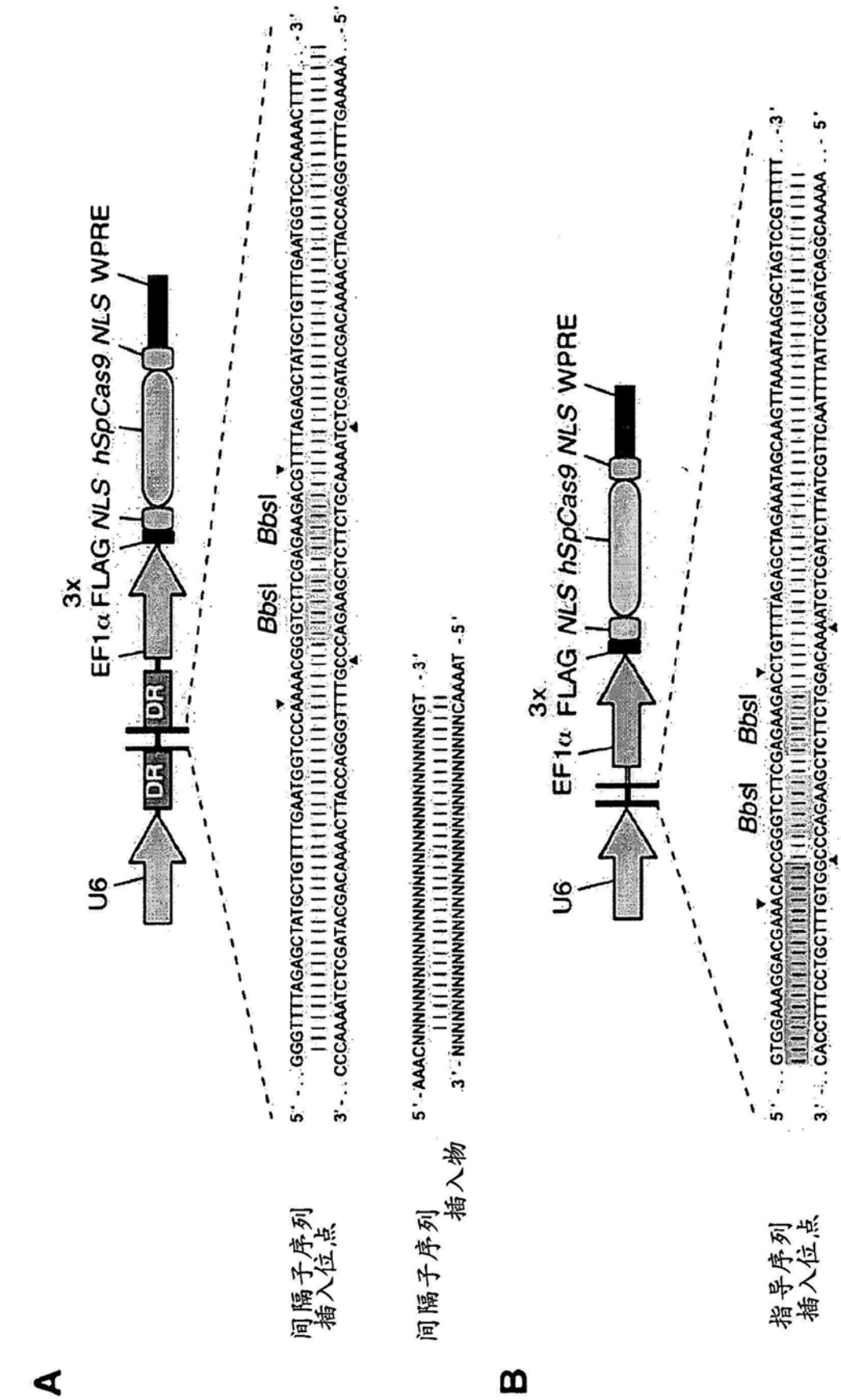
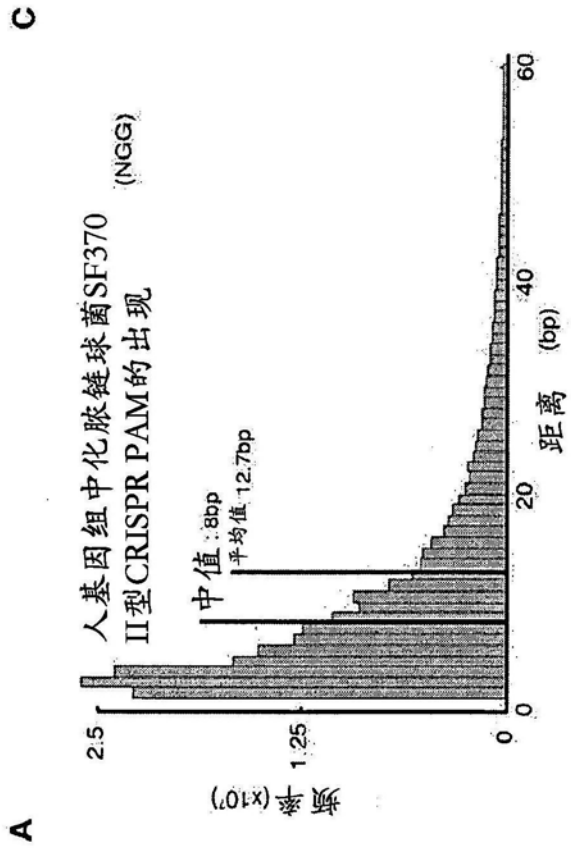
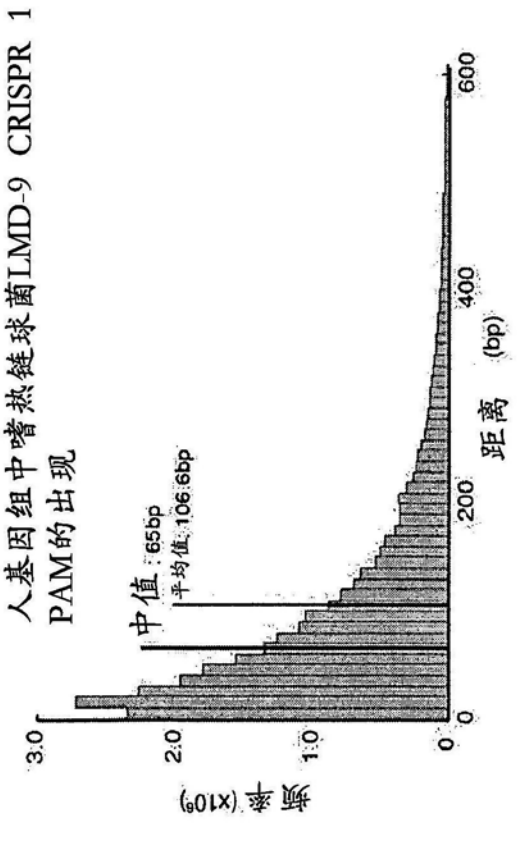

A
U6
DR
DR
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BbsI
BbsI
5' - ...GGGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACTTTT...-3'
3' - ...CCCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGCCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGTTTTGAAAAA...-5'
间隔子序列
插入位点
5' -AAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNGT -3'
3' - NNNNNNNNNNNNNNNNNNNNNNNNNNNNCAAAAT -5'
间隔子序列
插入物
B
U6
3x
EF1α FLAG NLS hSpCas9 NLS WPRE
BbsI
BbsI
5' - ...GTGGAAAGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTT...-3'
3' - ...CACCTTTCCTGCTTTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTATTCCGATCAGGCAAAAA...-5'
指导序列
插入位点
5' - CACCGNNNNNNNNNNNNNNNNNNN -3'
3' - CNNNNNNNNNNNNNNNNNNNCAAA -5'
指导序列
插入物

图8A-B





C

Chr	NGG 中值	NGG 平均值	NNAGAAW 中值	NNAGAAW 平均值
1	7	12.8	67	115.8
2	8	12.7	64	100.8
3	8	13.0	63	98.5
4	9	14.0	61	94.5
5	8	13.1	63	97.9
6	8	13.1	63	98.5
7	8	12.4	64	102.9
8	8	12.8	64	100.9
9	7	13.9	65	120.5
10	7	12.1	66	107.0
11	7	12.0	65	105.8
12	8	12.4	65	103.5
13	8	13.6	62	94.6
14	8	12.0	65	101.5
15	7	11.5	68	107.7
16	7	11.7	74	136.8
17	6	10.3	76	127.9
18	8	13.4	63	101.8
19	6	9.4	82	145.4
20	7	11.1	72	121.8
21	7	13.4	64	111.4
22	6	9.2	85	140.3
X	8	13.2	63	99.0
Y	8	29.2	62	223.7

图9A-C

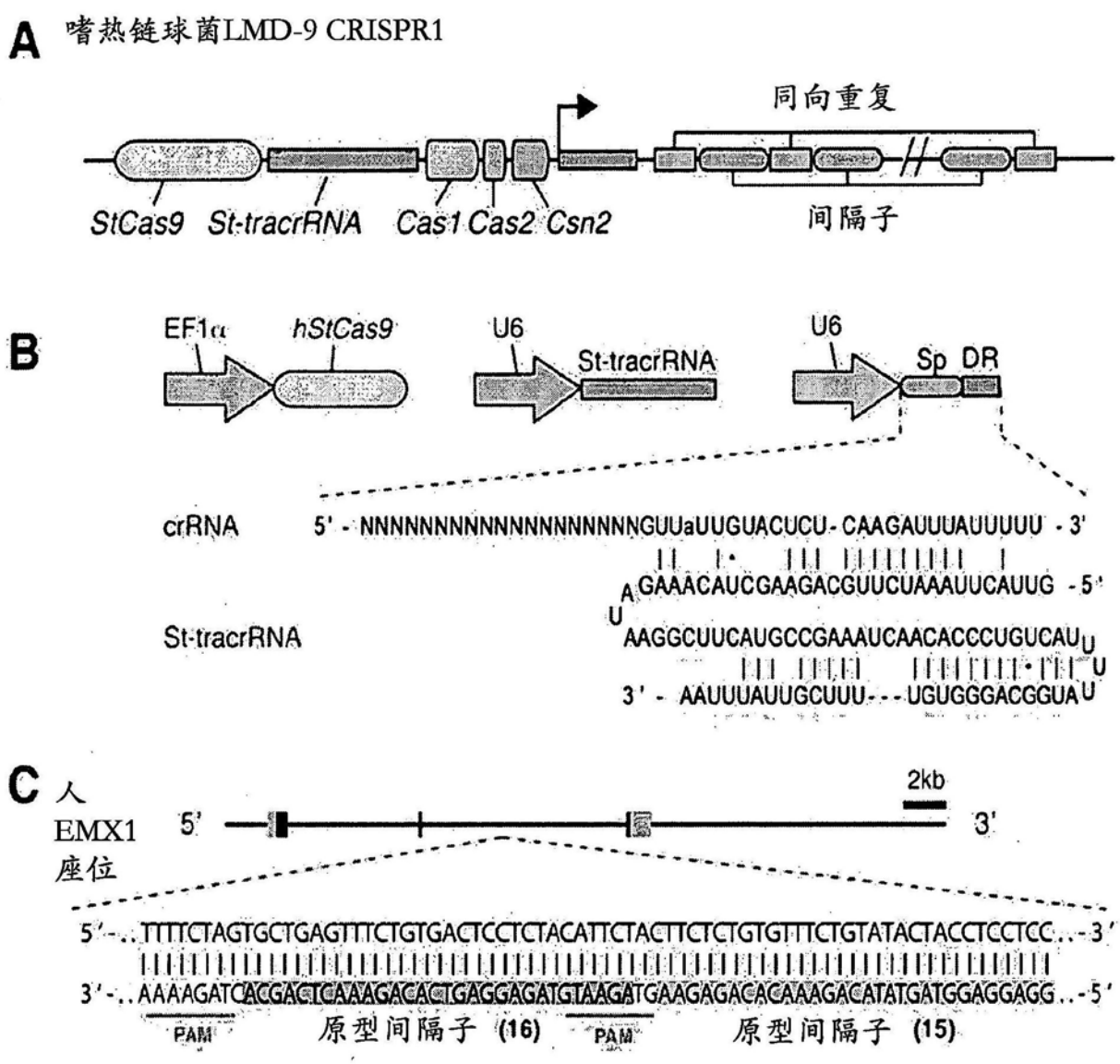

A
嗜热链球菌LMD-9 CRISPR1
同向重复
StCas9
St-tracrRNA
Cas1
Cas2
Csn2
间隔子
B
EF1α
hStCas9
U6
St-tracrRNA
U6
Sp
DR
crRNA
5' - NNNNNNNNNNNNNNNNNNNGUUaUUGUACUCU- CAAGAUUUAUUUUU - 3'
GAAACAUCGAAGACGUUCUAAAUUCAUUG - 5'
St-tracrRNA
AAGGCUUCAUGCCGAAAUCAACACCCUGUCAU
3' - AAUUUAUUGCUUU - - - UGUGGGACGGUA
C
人
EMX1
座位
5'
3'
2kb
5'-..TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC..-3'
3'-..AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG..-5'
PAM
原型间隔子 (16)
PAM
原型间隔子 (15)

图10A-C

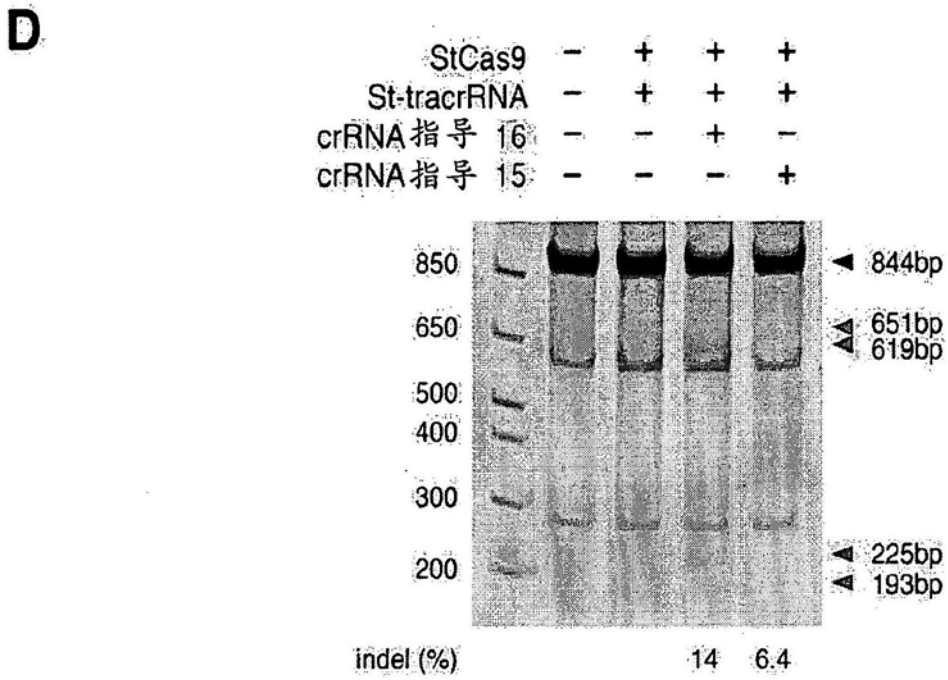

D
StCas9 − + + +
St-tracrRNA − + + +
crRNA指导 16 − − + −
crRNA指导 15 − − − +
850
650
500
400
300
200
844bp
651bp
619bp
225bp
193bp
indel (%)
14
6.4

图10D

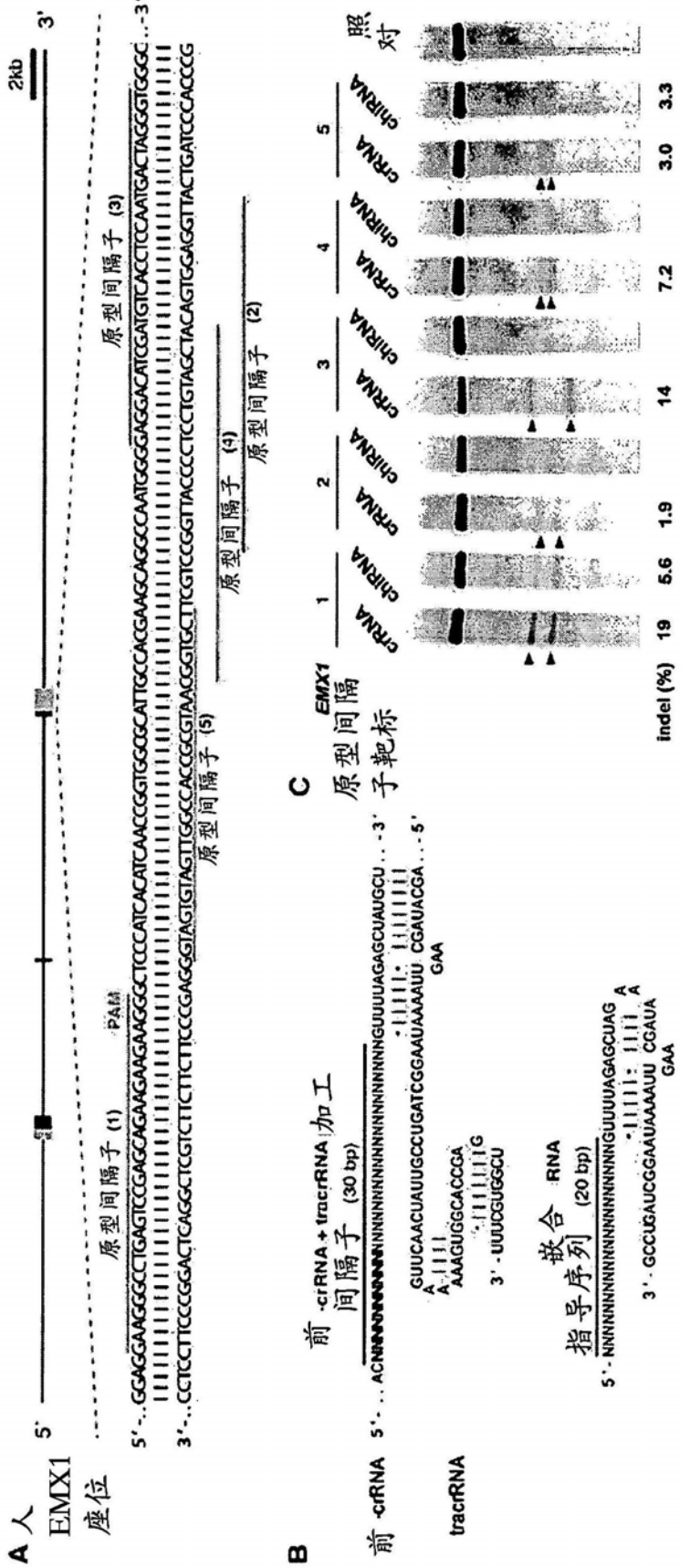

A 人
EMX1
座位
5'
3'
2kb
原型间隔子 (1)
PAM
原型间隔子 (3)
原型间隔子 (4)
原型间隔子 (2)
原型间隔子 (5)
B
前-crRNA
tracrRNA
前-crRNA + tracrRNA 加工
间隔子 (30 bp)
嵌合 RNA
指导序列 (20 bp)
C
EMX1
原型间隔子靶标
crRNA
chiRNA
1
2
3
4
5
照对
indel (%)
19
5.6
1.9
14
7.2
3.0
3.3

图11A-C

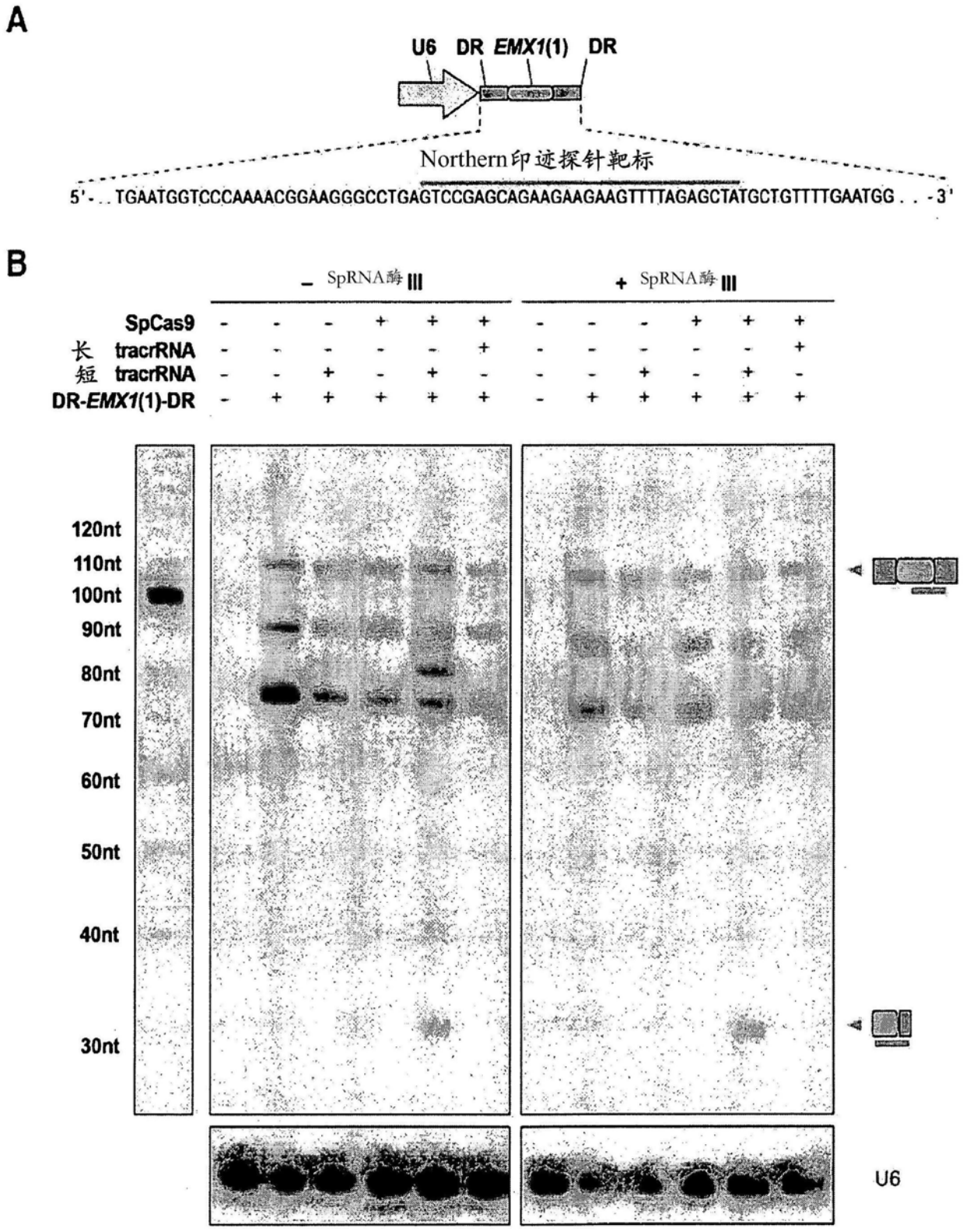

A
U6 DR EMX1(1) DR
Northern印迹探针靶标
5'-..TGAATGGTCCCAAAACGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGG...-3'
B
– SpRNA酶 III
+ SpRNA酶 III
SpCas9
长 tracrRNA
短 tracrRNA
DR-EMX1(1)-DR
120nt
110nt
100nt
90nt
80nt
70nt
60nt
50nt
40nt
30nt
U6

图12A-B

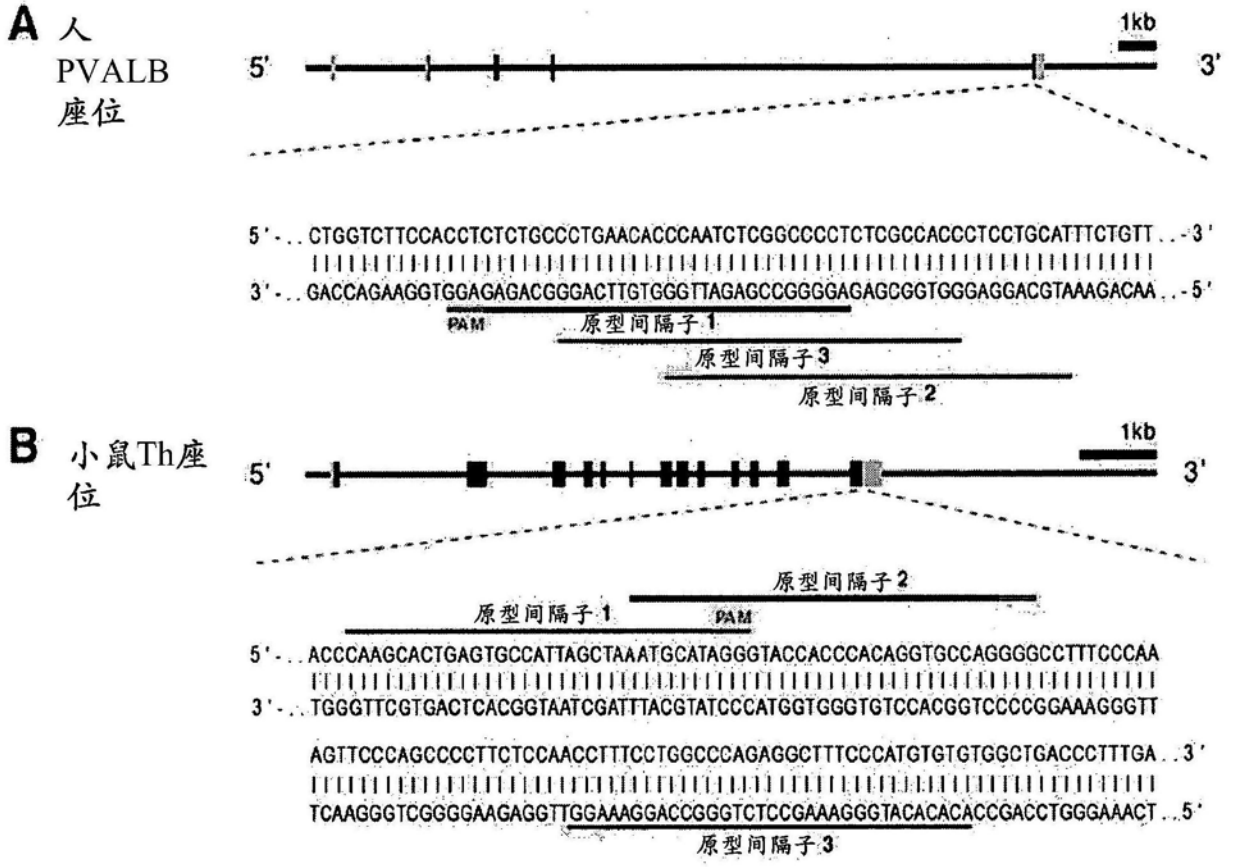

A
人
PVALB
座位
1kb
5'
3'
5'-...CTGGTCTTCCACCTCTCTGCCCTGAACACCCAATCTCGGCCCCTCTCGCCACCCTCCTGCATTTCTGTT...-3'
3'-...GACCAGAAGGTGGAGAGACGGGACTTGTGGGTTAGAGCCGGGGAGAGCGGTGGGAGGACGTAAAGACAA...-5'
PAM
原型间隔子1
原型间隔子3
原型间隔子2
B
小鼠Th座
位
1kb
5'
3'
原型间隔子2
原型间隔子1
PAM
5'-...ACCCAAGCACTGAGTGCCATTAGCTAAATGCATAGGGTACCACCCACAGGTGCCAGGGGCCTTTCCCAA
3'-...TGGGTTCGTGACTCACGGTAATCGATTTACGTATCCCATGGTGGGTGTCCACGGTCCCCGGAAAGGGTT
AGTTCCCAGCCCCTTCTCCAACCTTTCCTGGCCCAGAGGCTTTCCCATGTGTGTGGCTGACCCTTTGA...3'
TCAAGGGTCGGGGAAGAGGTTGGAAAGGACCGGGTCTCCGAAAGGGTACACACACCGACCTGGGAAACT...5'
原型间隔子3

图13A-B

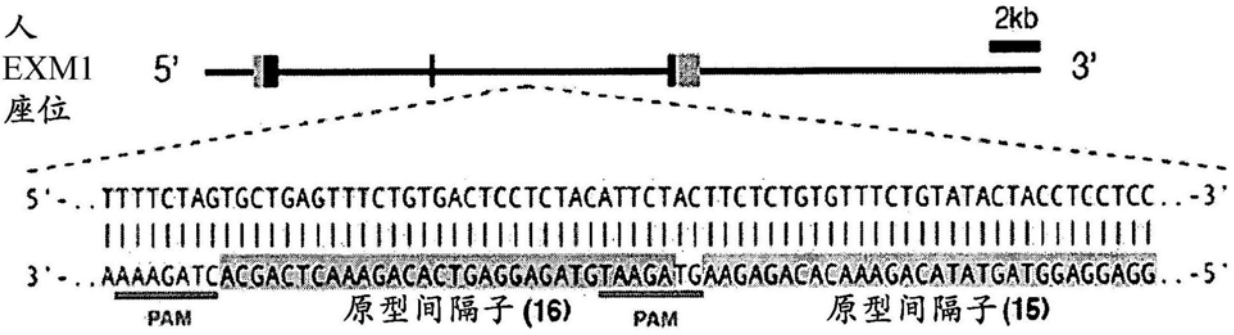

人
EXM1
座位
5'
2kb
3'
5'-...TTTTCTAGTGCTGAGTTTCTGTGACTCCTCTACATTCTACTTCTCTGTGTTTCTGTATACTACCTCCTCC...-3'
3'-...AAAAGATCACGACTCAAAGACACTGAGGAGATGTAAGATGAAGAGACACAAAGACATATGATGGAGGAGG...-5'
PAM
原型间隔子(16)
PAM
原型间隔子(15)

图14

引物名称	测定	基因组靶标	引物序列
Sp-EMX1-F	SURVEYOR测定,测序	*EMX1*	AAAACCACCCTTCTCTCTGGC
Sp-EMX1-R	SURVEYOR测定,测序	*EMX1*	GGAGATTGGAGACACGGAGAG
Sp-PVALB-F	SURVEYOR测定,测序	*PVALB*	CTGGAAAGCCAATGCCTGAC
Sp-PVALB-R	SURVEYOR测定,测序	*PVALB*	GGCAGCAAACTCCTTGTCCT
Sp-Th-F	SURVEYOR测定,测序	*Th*	GTGCTTTGCAGAGGCCTACC
Sp-Th-R	SURVEYOR测定,测序	*Th*	CCTGGAGCGCATGCAGTAGT
St-EMX1-F	SURVEYOR测定,测序	*EMX1*	ACCTTCTGTGTTTCCACCATTC
St-EMX1-R	SURVEYOR测定,测序	*EMX1*	TTGGGGAGTGCACAGACTTC
Sp-EMX1-RFLP-F	RFLP,测序	*EMX1*	GGCTCCCTGGGTTCAAAGTA
Sp-EMX1-RFLP-R	RFLP,测序	*EMX1*	AGAGGGGTCTGGATGTCGTAA
Pb_EMX1_sp1	Northern印迹探针	不适用	TAGCTCTAAAACTTCTTCTTCTGCTCGGAC
Pb_tracrRNA	Northern印迹探针	不适用	CTAGCCTTATTTTAACTTGCTATGCTGTTT

图15

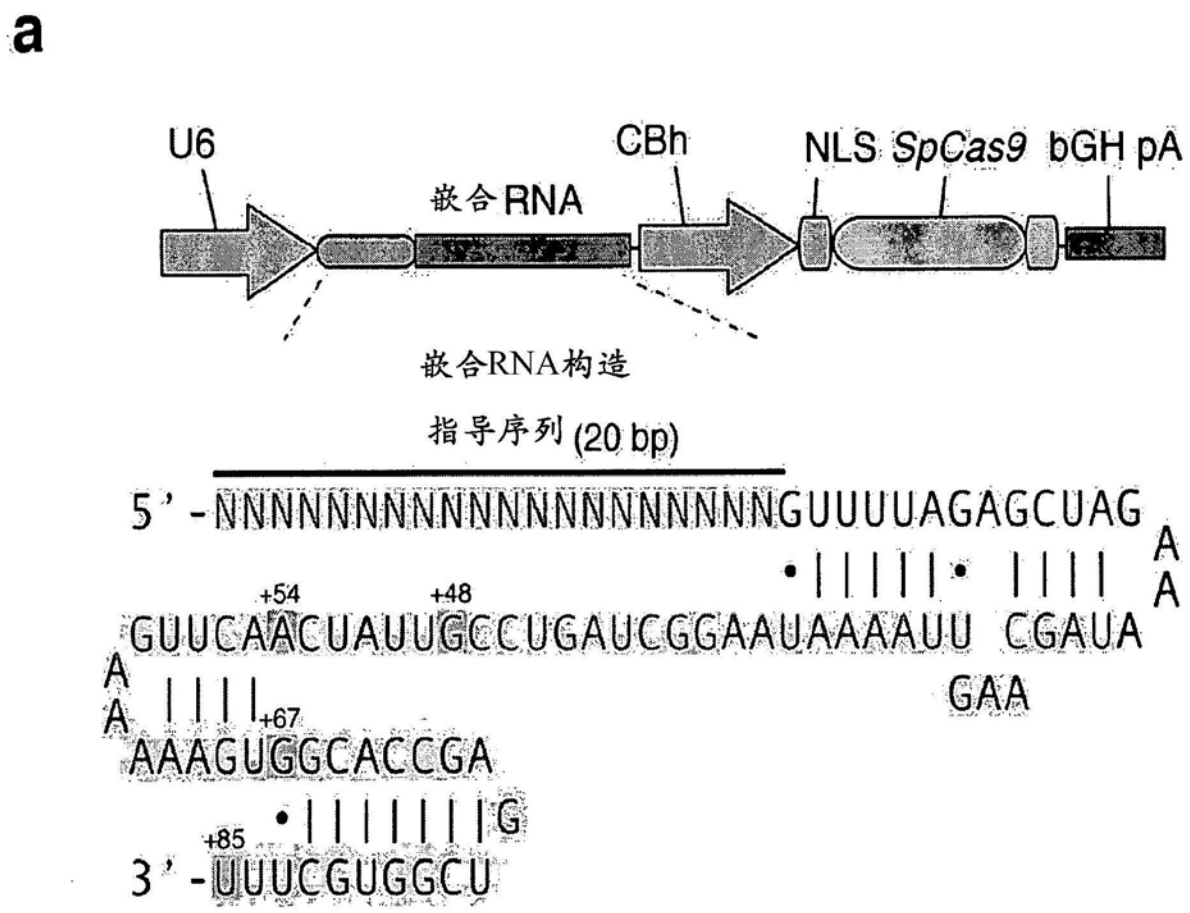

a
U6
嵌合RNA
CBh
NLS SpCas9 bGH pA
嵌合RNA构造
指导序列 (20 bp)
5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
A
A
+54
+48
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
A
A
GAA
+67
AAAGUGGCACCGA
G
+85
3'-UUUCGUGGCU

图16A

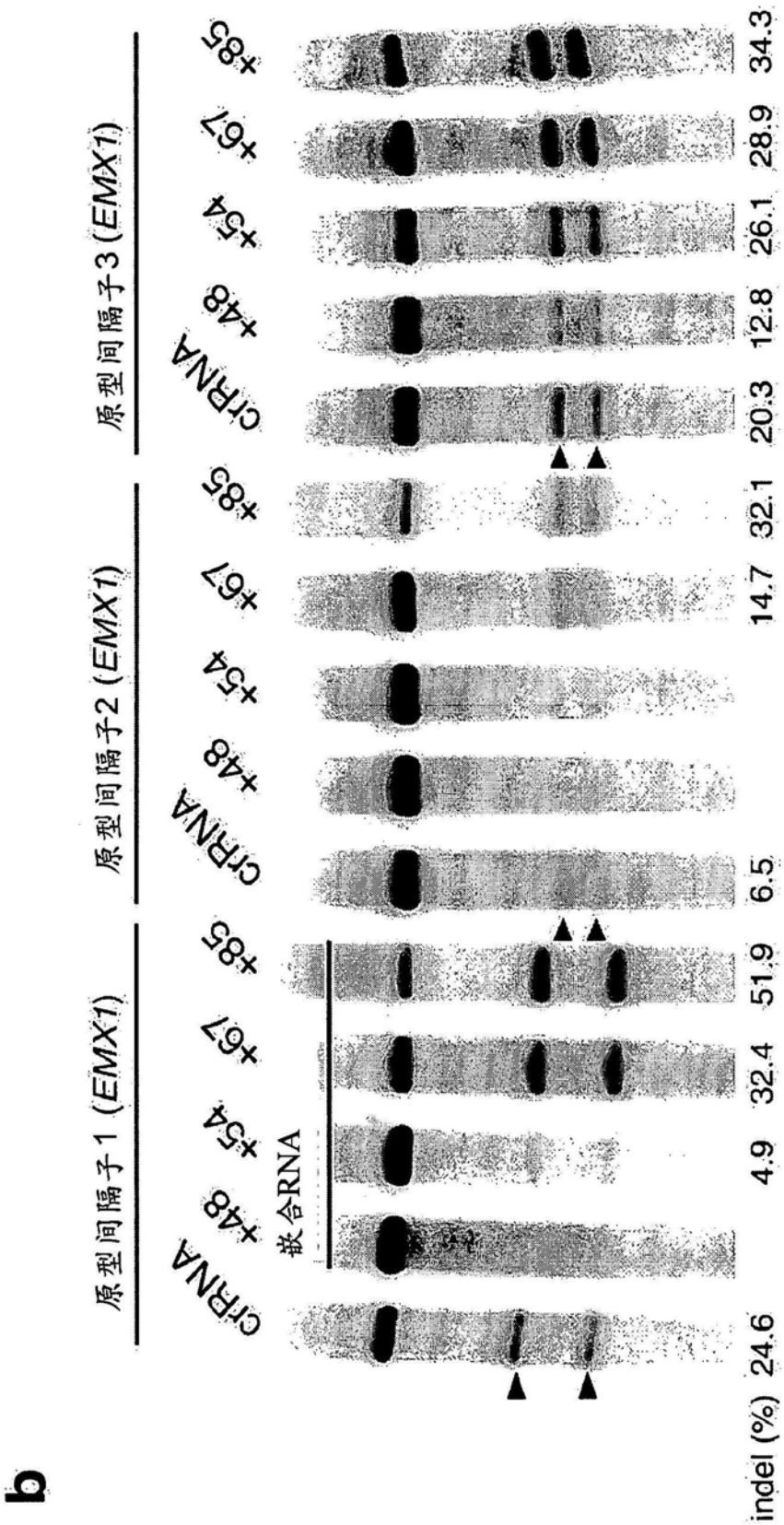

b
原型间隔子1 (EMX1)
原型间隔子2 (EMX1)
原型间隔子3 (EMX1)
crRNA
+48
+54
+67
+85
嵌合RNA
indel (%)
24.6
4.9
32.4
51.9
6.5
14.7
32.1
20.3
12.8
26.1
28.9
34.3

图16B

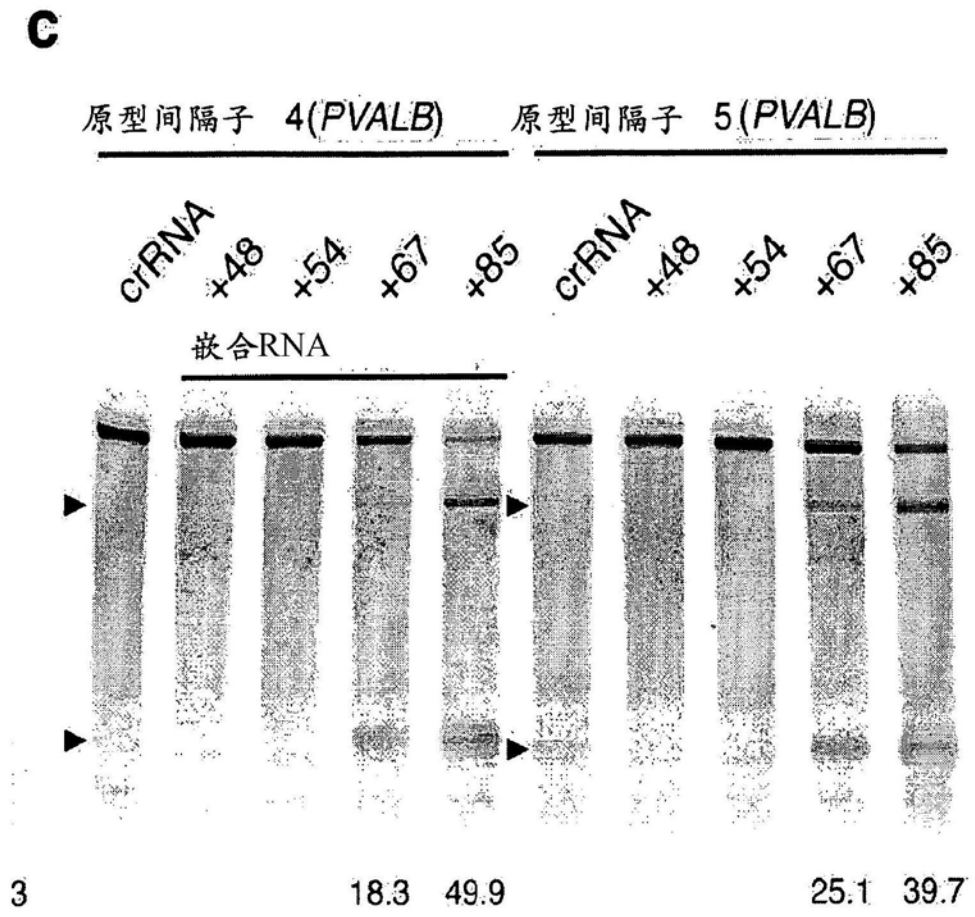

C
原型间隔子 4(PVALB)
原型间隔子 5(PVALB)
crRNA
+48
+54
+67
+85
crRNA
+48
+54
+67
+85
嵌合RNA
3
18.3
49.9
25.1
39.7

图16C

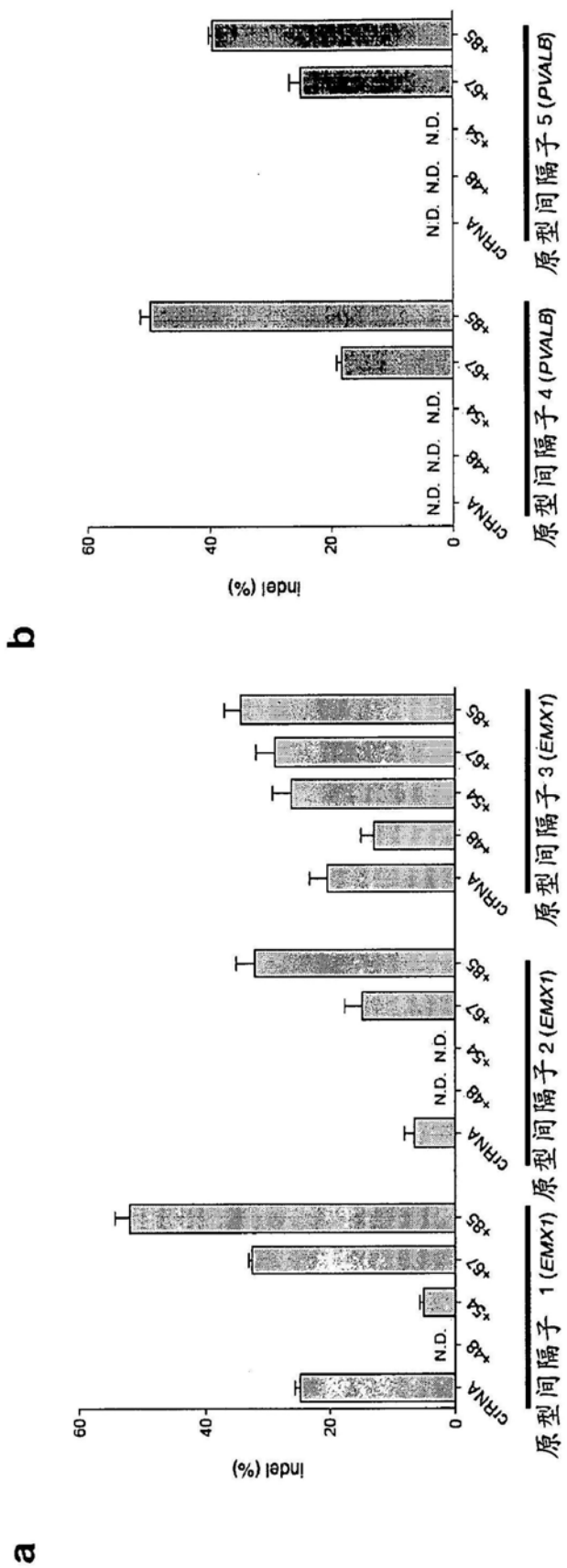

a
b
Indel (%)
0
20
40
60
crRNA
+48
+54
+67
+85
N.D.
原型间隔子 1 (EMX1)
原型间隔子 2 (EMX1)
原型间隔子 3 (EMX1)
原型间隔子 4 (PVALB)
原型间隔子 5 (PVALB)

图17A-B

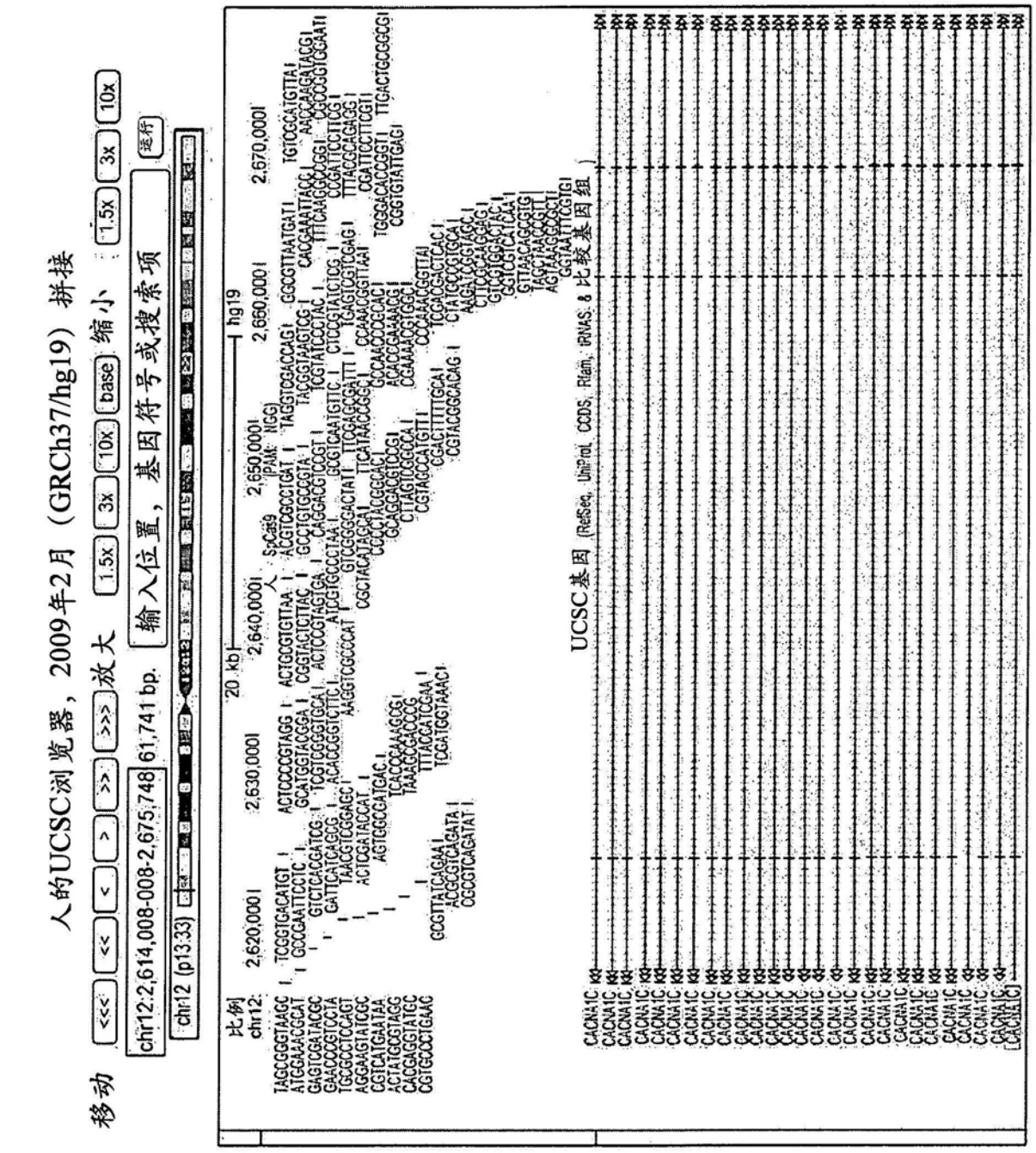
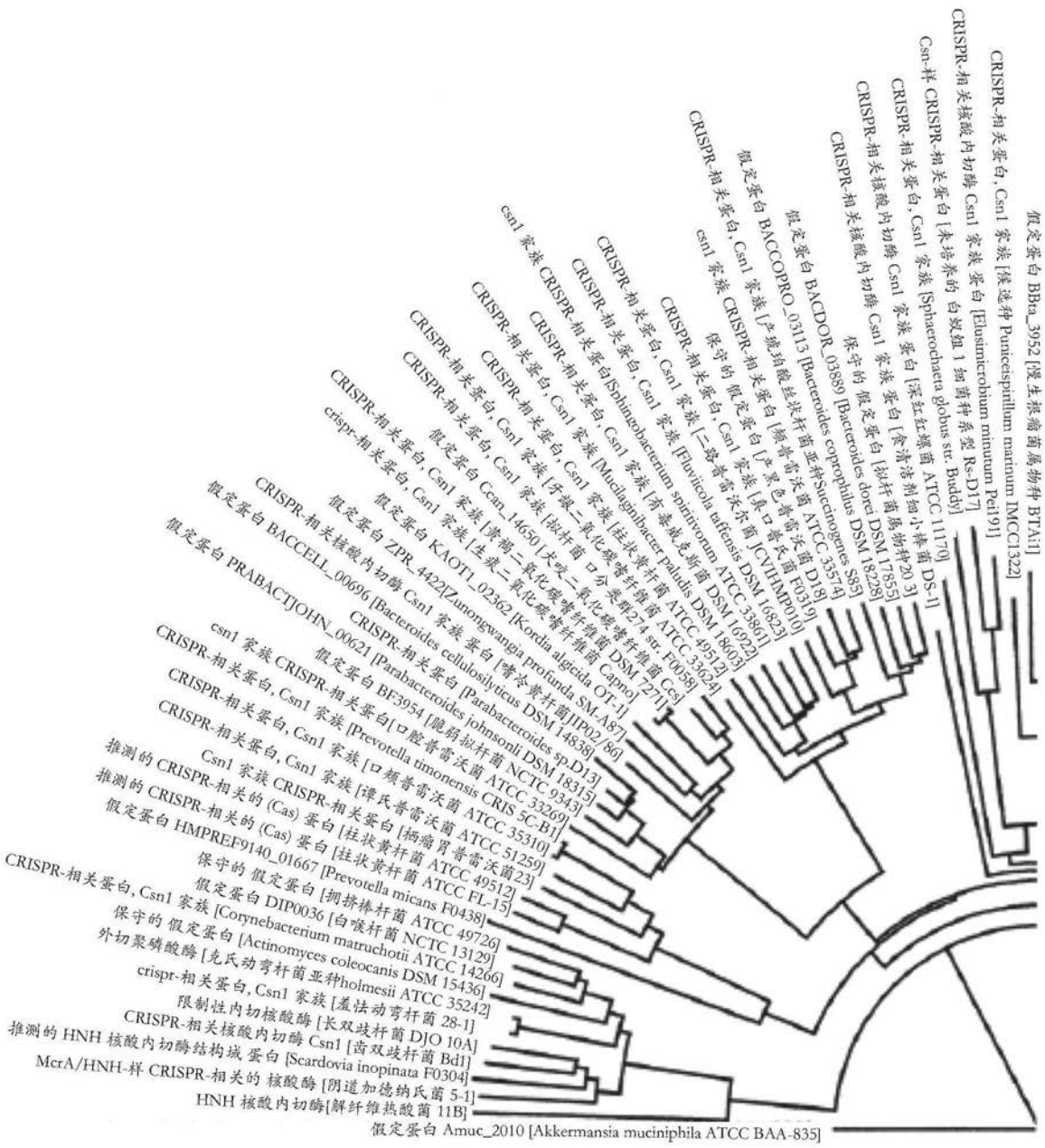
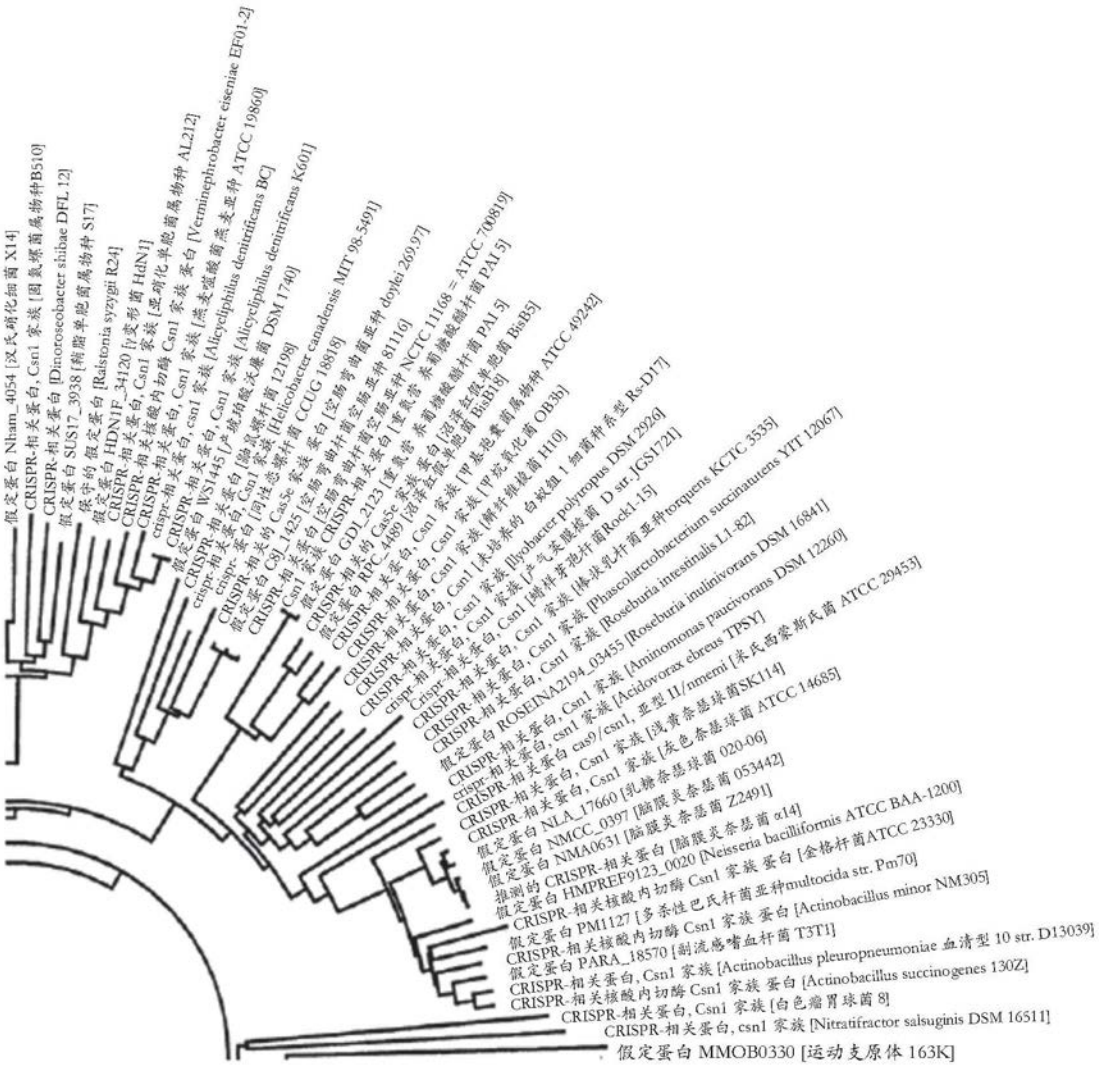

人的UCSC浏览器，2009年2月（GRCh37/hg19）拼接
移动
<<<
<<
<
>
>>
>>>
放大
1.5x
3x
10x
base
缩小
1.5x
3x
10x
chr12:2,614,008-008-2,675,748
61,741 bp.
输入位置，基因符号或搜索项
运行
chr12 (p13.33)
比例
chr12:
20 kb
hg19
2,620,000
2,630,000
2,640,000
2,650,000
2,660,000
2,670,000
SpCas9
PAM: NGG
UCSC基因 (RefSeq, UniProt, CCDS, Rfam, tRNAs & 比较基因组)
CACNA1C

图18


假定蛋白 BBta_3952 [慢生根瘤菌属物种 BTAi1]
CRISPR-相关蛋白, Csn1 家族 [候选种 Puniceispirillum marinum IMCC1322]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [Elusimicrobium minutum Pei91]
Csn-样 CRISPR-相关蛋白 [未培养的白蚁纽 1 细菌种系型 Rs-D17]
CRISPR-相关蛋白, Csn1 家族 [Sphaerochaeta globus str. Buddy]
CRISPR-相关核酸内切酶 Csn1 家族蛋白 [深红红螺菌 ATCC 11170]
保守的 假定蛋白 [拟杆菌属物种 20 3]
CRISPR-相关核酸内切酶 Csn1 家族蛋白 [食清洁剂细小棒菌 DS-1]
假定蛋白 BACDOR_03889 [Bacteroides coprophilus DSM 18228]
假定蛋白 BACCOPRO_03113 [Bacteroides coprosuis DSM 18777]
CRISPR-相关蛋白, Csn1 家族 [琥珀酸丝状杆菌亚种Succinogenes S85]
csn1 家族 CRISPR-相关蛋白 [产琥珀酸丝状杆菌亚种Succinogenes ATCC 35746]
CRISPR-相关蛋白, Csn1 家族 [二路普雷沃尔菌 JCVIHMP010]
CRISPR-相关蛋白, Csn1 家族 [Filifactor alocis ATCC 35896]
保守的 假定蛋白 [产琥珀酸丝状杆菌亚种Succinogenes ATCC 33574]
csn1 家族 CRISPR-相关蛋白 [二路普雷沃尔菌 JCVIHMP010]
CRISPR-相关蛋白, Csn1 家族 [Filifactor alocis ATCC 35896]
CRISPR-相关蛋白, Csn1 家族 [二路普雷沃尔菌 JCVIHMP010]
CRISPR-相关蛋白, Csn1 家族 [Filifactor alocis ATCC 35896


假定蛋白 Nham_4054 [汉氏硝化细菌 X14]
CRISPR-相关蛋白, Csn1 家族 [固氮螺菌属物种BS10]
CRISPR-相关蛋白 [Dinoroseobacter shibae DFL 12]
假定蛋白 SUS17_3938 [稍脂单胞菌属物种 S17]
保守的假定蛋白 [Ralstonia syzygii R24]
假定蛋白 HDN1F_34120 [立克形菌 HdN1]
CRISPR-相关蛋白, Csn1 家族 [恶硝化单胞菌属物种 AL212]
CRISPR-相关核酸内切酶 Csn1 家族 [燕麦强酸菌燕麦亚种 ATCC 19860]
crispr-相关蛋白, csn1 家族 [Verminephrobacter eiseniae EF01-2]
假定蛋白 WS1445 [产琥珀酸沃氏菌 DSM 1740]
CRISPR-相关蛋白, Csn1 家族 [Alicycliphilus denitrificans BC]
cisprc-蛋白 [间性恶螺杆菌 CCUG 18818]
假定蛋白 [弧菌螺杆菌 12198]
CRISPR-相关的 Cas5e 家族 [Helicobacter canadensis MIT 98-5491]
Csn1 家族 [Alicycliphilus denitrificans K601]
假定蛋白 [空肠弯曲杆菌空肠亚种 81116]
CRISPR-相关蛋白 [空肠弯曲杆菌空肠亚种 doylei 269.97]
假定蛋白 GDI_2123 [重氮营养葡糖酸酯杆菌 NCTC 11168 = ATCC 700819]
假定蛋白 RPC_4489 [沼泽红假单胞菌 BisB5]
CRISPR-相关蛋白, Csn1 家族 [重氮营养葡糖酸酯杆菌 PAI 5]
CRISPR-相关蛋白, Csn1 家族 [沼泽红假单胞菌 BisB5]
CRISPR-相关蛋白, Csn1 家族 [甲基思囊菌属物种 ATCC 49242]
CRISPR-相关蛋白, Csn1 家族 [甲烷真化菌 OB3b]
crispr-相关蛋白, Csn1 家族 [未培养的白蚁组 1 细菌种系型 Rs-D17]
CRISPR-相关蛋白, Csn1 家族 [未培养的白蚁组 1 细菌种系型 Rs-D17]
CRISPR-相关蛋白, Csn1 家族 [Hyphomicrobium denitrificans 1NES1]
CRISPR-相关蛋白, Csn1 家族 [Hyphomicrobium denitrificans 1NES1]
CRISPR-相关蛋白, Csn1 家族 [Hyphomicrobium sp. MC

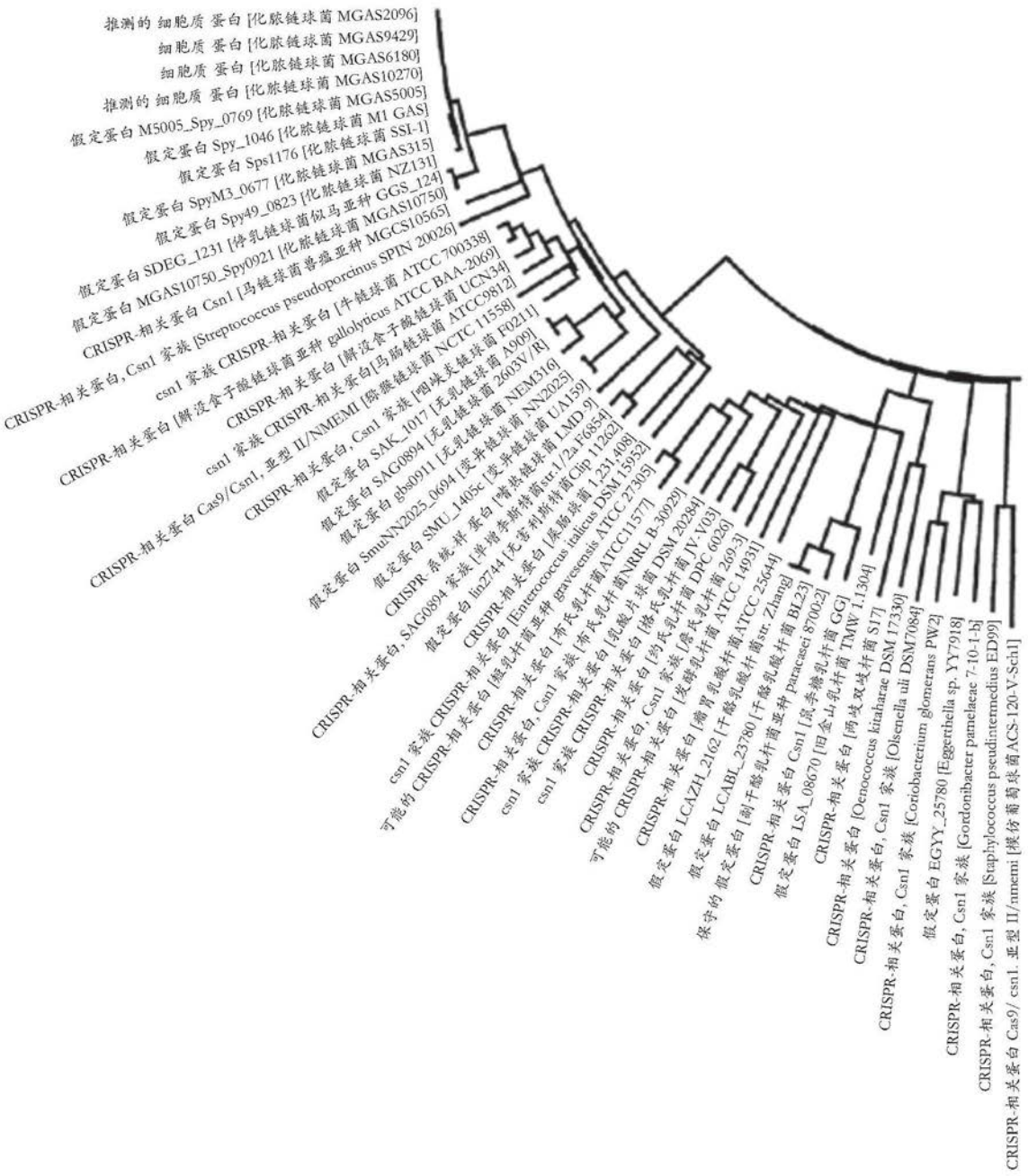

推测的 细胞质 蛋白 [化脓链球菌 MGAS2096]
细胞质 蛋白 [化脓链球菌 MGAS9429]
细胞质 蛋白 [化脓链球菌 MGAS6180]
推测的 细胞质 蛋白 [化脓链球菌 MGAS10270]
假定蛋白 M5005_Spy_0769 [化脓链球菌 MGAS5005]
假定蛋白 Spy_1046 [化脓链球菌 M1 GAS]
假定蛋白 Sps1176 [化脓链球菌 SSI-1]
假定蛋白 SpyM3_0677 [化脓链球菌 MGAS315]
假定蛋白 Spy49_0823 [化脓链球菌 NZ131]
假定蛋白 SDEG_1231 [停乳链球菌似马亚种 GGS_124]
假定蛋白 MGAS10750_Spy0921 [化脓链球菌 MGAS10750]
CRISPR-相关蛋白 Csn1 [马链球菌兽瘟亚种 MGCS10565]
CRISPR-相关蛋白, Csn1 家族 [Streptococcus pseudoporcinus SPIN 20026]
csn1 家族 CRISPR-相关蛋白 [牛链球菌 ATCC 700338]
CRISPR-相关蛋白 [解没食子酸链球菌亚种 gallolyticus ATCC BAA-2069]
CRISPR-相关蛋白 [解没食子酸链球菌 UCN34]
CRISPR-相关蛋白 Cas9/Csn1, 亚型 II/NMEMI [脑膜链球菌 ATCC98121]
csn1 家族 CRISPR-相关蛋白, Csn1 家族 [咽峡炎链球菌 NCTC 11558]
CRISPR-相关蛋白 [马脑链球菌 F0211]
假定蛋白 SAG0894 [无乳链球菌 A909]
假定蛋白 gbs0911 [无乳链球菌 2603V/R]
假定蛋白 SmuNN2025_0694 [变异链球菌 UA159]
假定蛋白 SMU_1405c [变异链球菌 LMD 9]
CRISPR-系统-样 蛋白 [*嗜热链球菌str.1/2a F6854]
假定蛋白 lin2744 [无害利斯特菌 Clip 11262]
CRISPR-相关蛋白 [Enterococcus italicus DSM 15952]
CRISPR-相关蛋白 [屎肠球菌 1_231_408]
CRISPR-相关蛋白 [*肺炎链球菌NRRL B-30929]
CRISPR-相关蛋白 [*肺炎链球菌 DSM 20284]
CRISPR-相关蛋白 [*肺炎链球菌 V-V03]
CRISPR-相关蛋白 [*肺炎链球菌 DPC 6026]
CRISPR-相关蛋白 [*肺炎链球菌 ATCC 1491]
CRISPR-相关蛋白 [*肺炎链球菌 269-3]
CRISPR-相关蛋白 [*肺炎链球菌ATCC 25644]
CRISPR-相关蛋白 [*肺炎链球菌str. Zhang]
CRISPR-相关蛋白 [*肺炎链球菌 BL23]
CRISPR-相关蛋白 [*肺炎链球菌 GG]
CRISPR-相关蛋白 [*肺炎链球菌 TMW 1.1304]
CRISPR-相关蛋白 [*肺炎链球菌 S17]
CRISPR-相关蛋白, Csn1 家族 [Olsenella uli DSM 17330]
CRISPR-相关蛋白, Csn1 家族 [Coriobacterium glomerans PW2]
假定蛋白 EGYY_25780 [Eggerthella sp. YY7918]
CRISPR-相关蛋白, Csn1 家族 [Gordonibacter pamelaeae 7-10-1-b]
CRISPR-相关蛋白, Csn1 家族 [Staphylococcus pseudintermedius ED99]
CRISPR-相关蛋白 Cas9/ csn1. 亚型 II/nmemi [模仿葡萄球菌ACS-120-V-Sch1]

图19C

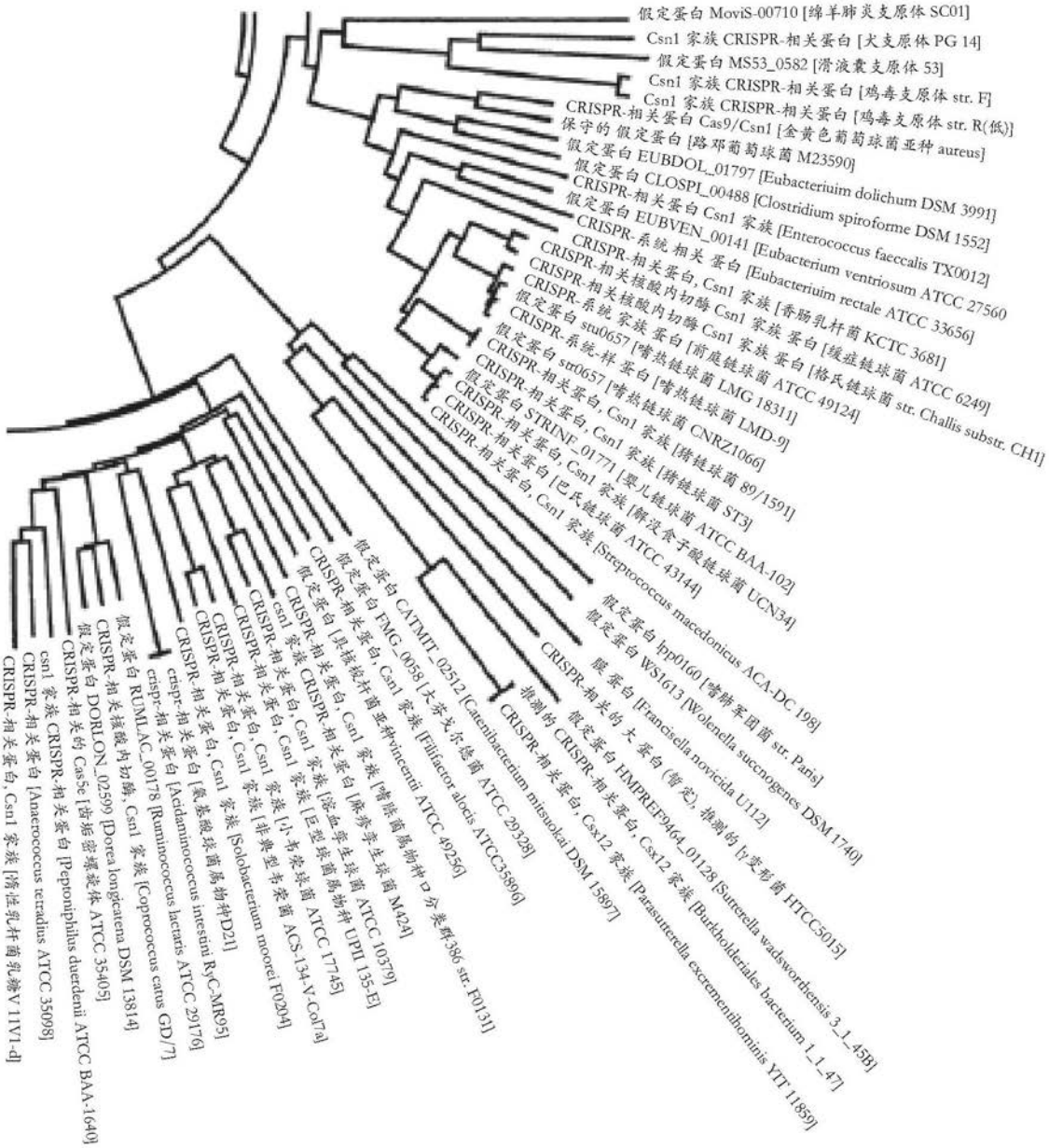

假定蛋白 MoviS-00710 [绵羊肺炎支原体 SC01]
Csn1 家族 CRISPR-相关蛋白 [犬支原体 PG 14]
假定蛋白 MS53_0582 [滑液囊支原体 53]
Csn1 家族 CRISPR-相关蛋白 [鸡毒支原体 str. F]
Csn1 家族 CRISPR-相关蛋白 [鸡毒支原体 str. R(低)]
CRISPR-相关蛋白 Cas9/Csn1 [金黄色葡萄球菌亚种 aureus]
保守的 假定蛋白 [路邓葡萄球菌 M23590]
假定蛋白 EUBDOL_01797 [Eubacteriuim dolichum DSM 3991]
假定蛋白 CLOSPI_00488 [Clostridium spiroforme DSM 1552]
CRISPR-相关蛋白 Csn1 家族 [Enterococcus faeccalis TX0012]
假定蛋白 EUBVEN_00141 [Eubacterium ventriosum ATCC 27560
CRISPR-系统 相关 蛋白 [Eubacteriuim rectale ATCC 33656]
CRISPR-相关蛋白, Csn1 家族 [香肠乳杆菌 KCTC 3681]
CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [缓症链球菌 ATCC 6249]
CRISPR-系统 家族 蛋白 [前庭链球菌 ATCC 49124]
假定蛋白 stu0657 [嗜热链球菌 LMG 18311]
CRISPR-系统-样 蛋白 [嗜热链球菌 LMD-9]
假定蛋白 str0657 [嗜热链球菌 CNRZ1066]
CRISPR-相关蛋白, Csn1 家族 [猪链球菌 89/1591]
假定蛋白 STRINF_01771 [婴儿链球菌 ATCC BAA-102]
CRISPR-相关蛋白, Csn1 家族 [解没食子酸链球菌 UCN34]
CRISPR-相关蛋白, Csn1 家族 [巴氏链球菌 ATCC 43144]
假定蛋白 CATMIT_02512 [Catenibacterium mitsuokai DSM 15897]
假定蛋白 FMG_0058 [大肠支尔德菌 ATCC 25928]
CRISPR-相关蛋白, Csn1 家族 [Filifactor alocis ATCC 49256]
假定蛋白 [具核梭杆菌亚种vincentii ATCC 10379]
CRISPR-相关蛋白, Csn1 家族 [解脲菌属物种 D 分类群386 str. F0131]
csn1 家族 CRISPR-相关蛋白, Csn1 家族 [溶血韦生球菌 M424]
CRISPR-相关蛋白, Csn1 家族 [巨型球菌属物种 UPII 135-E]
CRISPR-相关蛋白, Csn1 家族 [小韦荣球菌 ATCC 17745]
CRISPR-相关蛋白, Csn1 家族 [非典型韦荣菌 ACS-134-V-Col7a]
CRISPR-相关蛋白, Csn1 家族 [Solobacterium moorei F0204]
CRISPR-相关蛋白 [氨基酸球菌属物种D21]
crispr-相关蛋白 [Acidaminococcus intestini RyC-MR95]
假定蛋白 RUMLAC_00178 [Ruminococcus lactaris ATCC 29176]
CRISPR-相关核酸内切酶, Csn1 家族 [Coprococcus catus GD/7]
假定蛋白 DORLON_02599 [Dorea longicatena DSM 13814]
CRISPR-相关的 Cas5e [齿垢密螺旋体 ATCC 35405]
CRISPR-相关蛋白 [Peptoniphilus duerdenii ATCC BAA-1640]
csn1 家族 CRISPR-相关蛋白 [Anaerococcus tetradius ATCC 35098]
CRISPR-相关蛋白, Csn1 家族 [消化乳杆菌乳糖V 11V1-d]
假定蛋白 Ipp0160 [肺炎军团菌 str. Paris]
假定蛋白 WS1613 [Wolenella succnogenes DSM 1740]
腺 蛋白 [Francisella noviecida U112]
假定蛋白 HMPREF9464_01128 [Sutterella wadsworthensis 3_1_45B]
CRISPR-相关的 大 蛋白 (暂定), 拟杆的 [γ变形菌 HTCC5015]
CRISPR-相关蛋白, Csx12 家族 [Burkholderiales bacterium 1_1_47]
假定蛋白 CRISPR-相关蛋白, Csx12 家族 [Parasutterella excrementihominis YIT 11859]

图19D

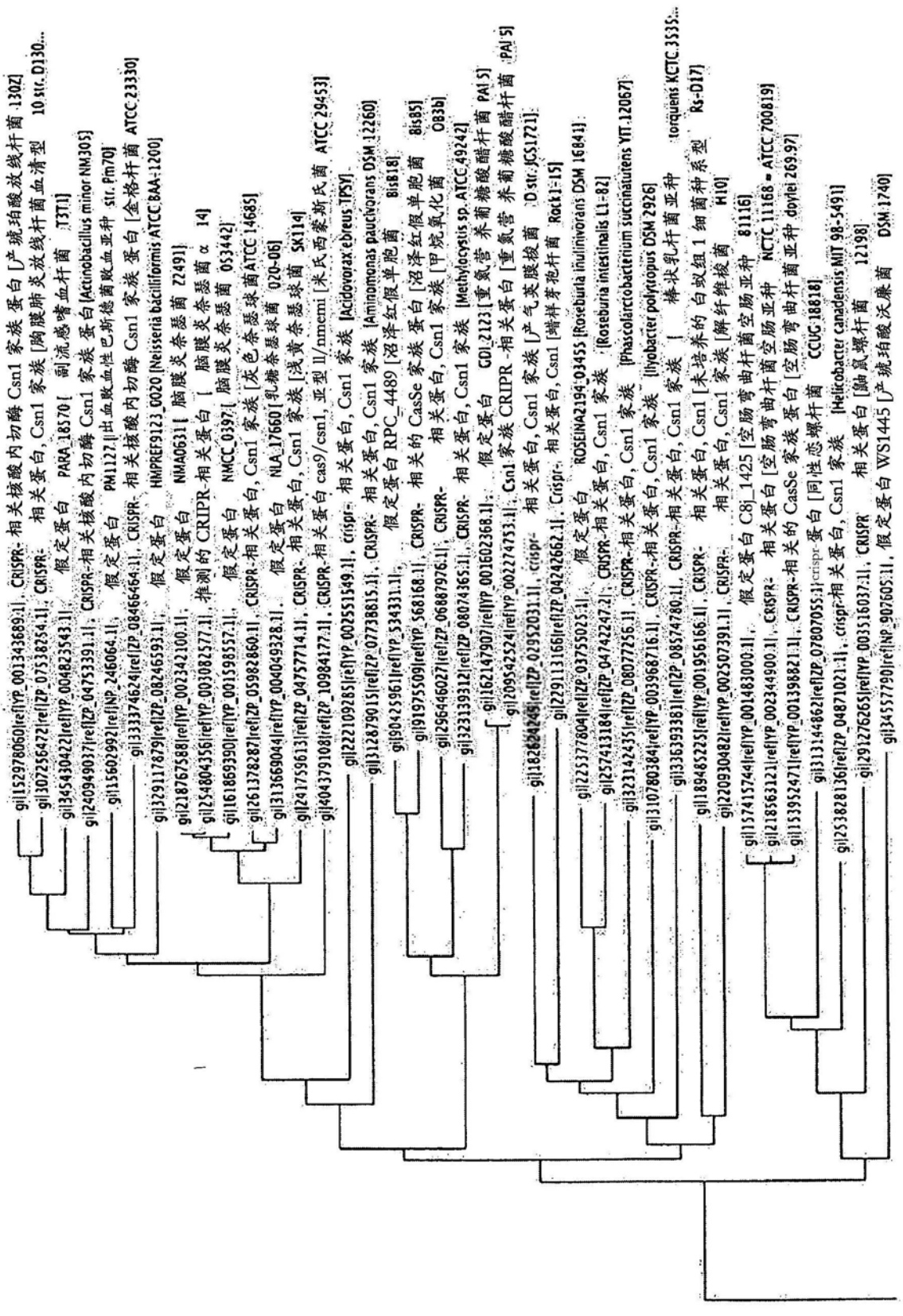

gi|152978060|ref|YP_001343689.1| CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [产琥珀酸放线杆菌 130Z]
gi|307255472|ref|ZP_07538254.1| CRISPR- 相关蛋白, Csn1 家族 [胸膜肺炎放线杆菌血清型 10 str. D130...
gi|345430422|ref|YP_004823543.1| 假定蛋白 PARA_18570 [副流感嗜血杆菌 T3T1]
gi|240949037|ref|ZP_04753391.1| CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [Actinobacillus minor NM305]
gi|115602992|ref|NP_246064.1| 假定蛋白 PM1127.1 出血败血性巴斯德菌败血亚种 str. Pm70]
gi|333374624|ref|ZP_08466464.1| CRISPR-相关核酸内切酶 Csn1 家族 蛋白 [金格杆菌 ATCC 23330]
gi|329117879|ref|ZP_08246593.1| 假定蛋白 HMPREF9123_0020 [Neisseria bacilliformis ATCC BAA-1200]
gi|218767588|ref|YP_002342100.1| 假定蛋白 NMA0631 [脑膜炎奈瑟菌 Z2491]
gi|254804356|ref|YP_003082577.1| 推测的 CRIPR-相关蛋白 [脑膜炎奈瑟菌 α 14]
gi|161869390|ref|YP_001598557.1| 假定蛋白 NMCC_0397 [脑膜炎奈瑟菌 053442]
gi|261378287|ref|ZP_05982860.1| CRISPR-相关蛋白, Csn1 家族 [灰色奈瑟球菌 ATCC 14685]
gi|313669044|ref|YP_004049328.1| 假定蛋白 NLA_17660 [乳糖奈瑟球菌 020-06]
gi|241759613|ref|ZP_04757714.1| CRISPR-相关蛋白, Csn1 家族 [浅黄奈瑟球菌 SK114]
gi|404379108|ref|ZP_10984177.1| CRISPR-相关蛋白 cas9/csn1, 亚型 II/nmemi [米氏西蒙斯氏菌 ATCC 29453]
gi|222109285|ref|YP_002551549.1| crispr-相关蛋白, Csn1 家族 [Acidovorax ebreus TPSY]
gi|311287901 5|ref|ZP_07738815.1| CRISPR-相关蛋白, Csn1 家族 [Aminomonas paucivorans DSM 12260]
gi|90425961|ref|YP_534331.1| 假定蛋白 RPC_4489 [沼泽红假单胞菌 BisB18]
gi|91975509|ref|YP_568168.1| CRISPR-相关的 CasSe 家族蛋白 [沼泽红假单胞菌 BisB5]
gi|296446027|ref|ZP_06887976.1| CRISPR-相关蛋白, Csn1 家族 [甲烷氧化菌 OB3b]
gi|323139312|ref|ZP_08074365.1| CRISPR-相关蛋白, Csn1 家族 [Methylocystis sp. ATCC 49242]
gi|162147907|ref|YP_001602368.1| 假定蛋白 GDI_2123 [重氮营养葡糖酸醋杆菌 PAl 5]
gi|209542524|ref|YP_002274733.1| Csn1 家族 CRIPR-相关蛋白 [重氮营养葡糖酸醋杆菌 (PAl 5]
gi|182624245|ref|ZP_02952031.1| crispr-相关蛋白, Csn1 家族 [产气荚膜梭菌 D str. JGS1721]
gi|229113166|ref|ZP_04242662.1| Crispr-相关蛋白, Csn1 [蜡样芽孢杆菌 Rock1-15]
gi|225377804|ref|ZP_03755025.1| 假定蛋白 ROSEINA2194_03455 [Roseburia inulinivorans DSM 16841]
gi|257413184|ref|ZP_04742247.1| CRISPR-相关蛋白, Csn1 家族 [Roseburia intestinalis L1-82]
gi|323142435|ref|ZP_08077256.1| CRISPR-相关蛋白, Csn1 家族 [Phascolarctobacterium succinatutens YIT 12067]
gi|310780384|ref|YP_003968716.1| CRISPR-相关蛋白, Csn1 家族 [Ilyobacter polytropus DSM 2926]
gi|336393381|ref|ZP_08574780.1| CRISPR-相关蛋白, Csn1 家族 [棒状乳杆菌亚种 torquens KCTC 3535...
gi|189465225|ref|YP_001956166.1| CRISPR-相关蛋白, Csn1 [未培养的白蚁肠 1 细菌种系型 Rs-D17]
gi|220930482|ref|YP_002507391.1| CRISPR-相关蛋白, Csn1 家族 [解纤维梭菌 H10]
gi|157415744|ref|YP_001483000.1| 假定蛋白 C8J_1425 [空肠弯曲杆菌空肠亚种 811161]
gi|218563121|ref|YP_002344900.1| CRISPR-相关蛋白 [空肠弯曲杆菌空肠亚种 NCTC 11168 = ATCC 700819]
gi|153952471|ref|YP_001398821.1| CRISPR-相关的 CasSe 家族蛋白 [空肠弯曲杆菌亚种 doyleі 269.97]
gi|311144862|ref|ZP_07807055.1| crispr-蛋白 [同性恋螺杆菌 CCUG 18818]
gi|253828136|ref|ZP_04871021.1| crispr相关蛋白, Csn1 家族 [Helicobacter canadensis MIT 98-5491]
gi|291276265|ref|YP_003516037.1| CRISPR 相关蛋白 [鼠螺杆菌 12198]
gi|345577901|ref|NP_907605.1| 假定蛋白 WS1445 [产琥珀酸沃康菌 DSM 1740]

图20A

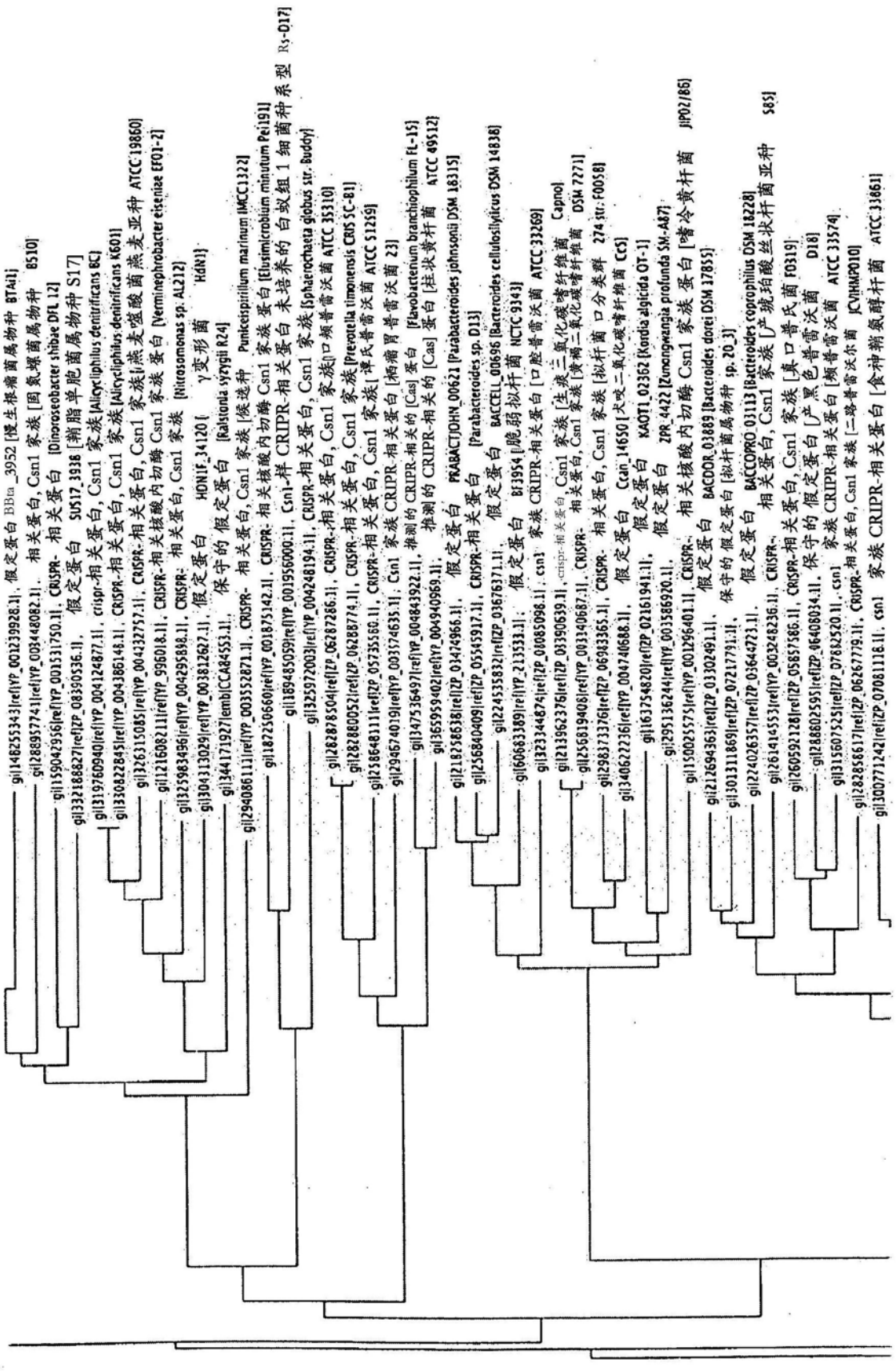


图20B

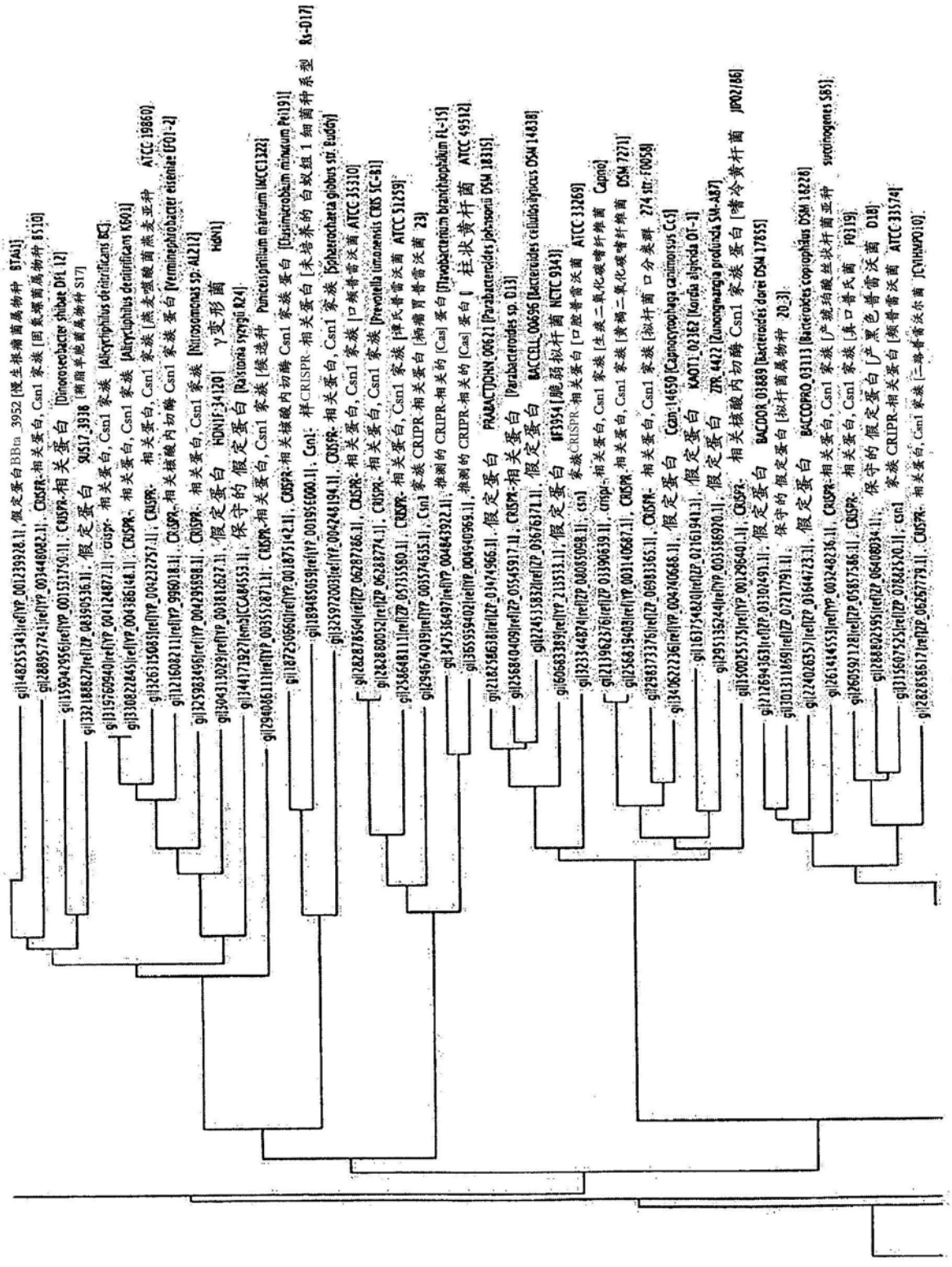


图20C

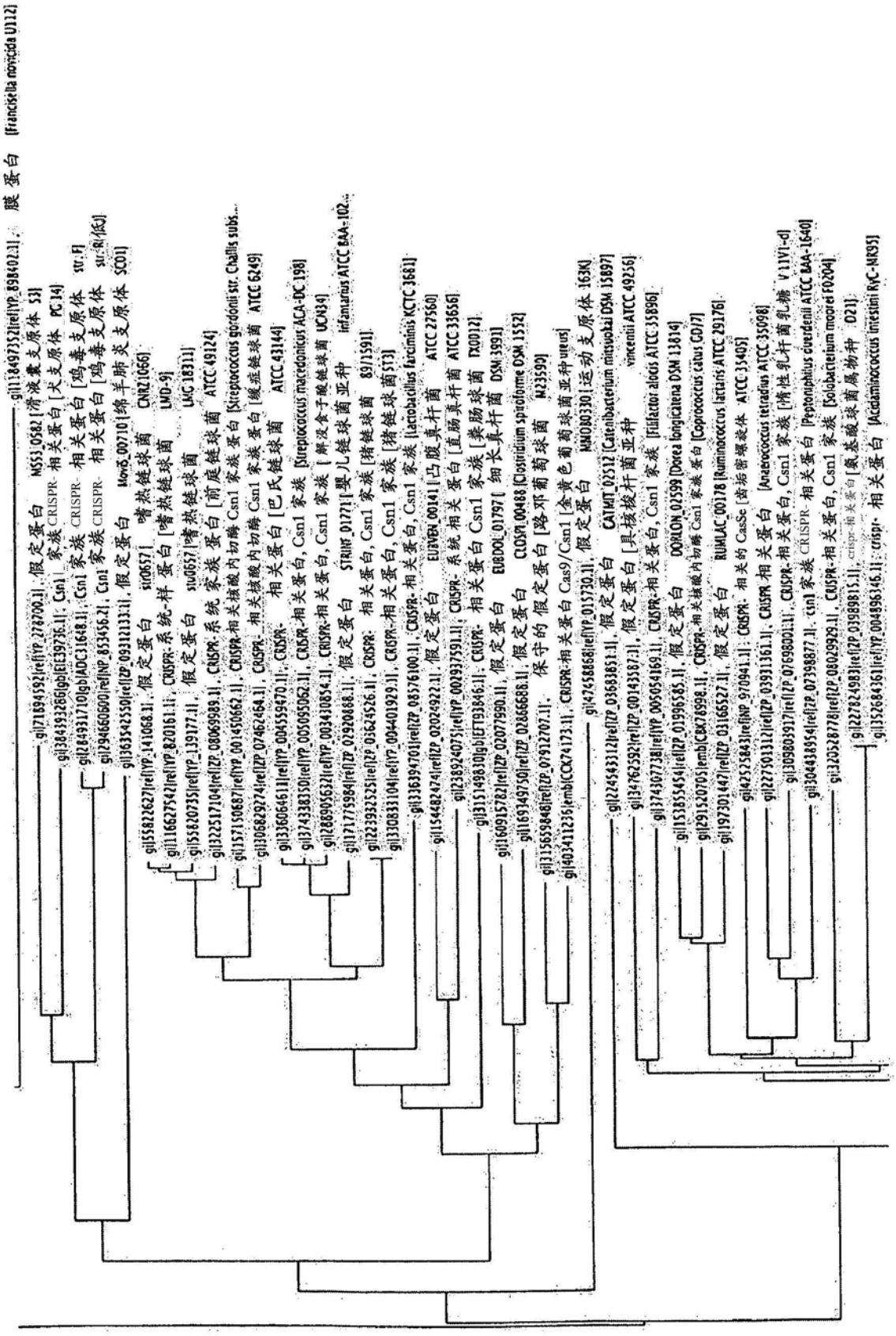


图20D

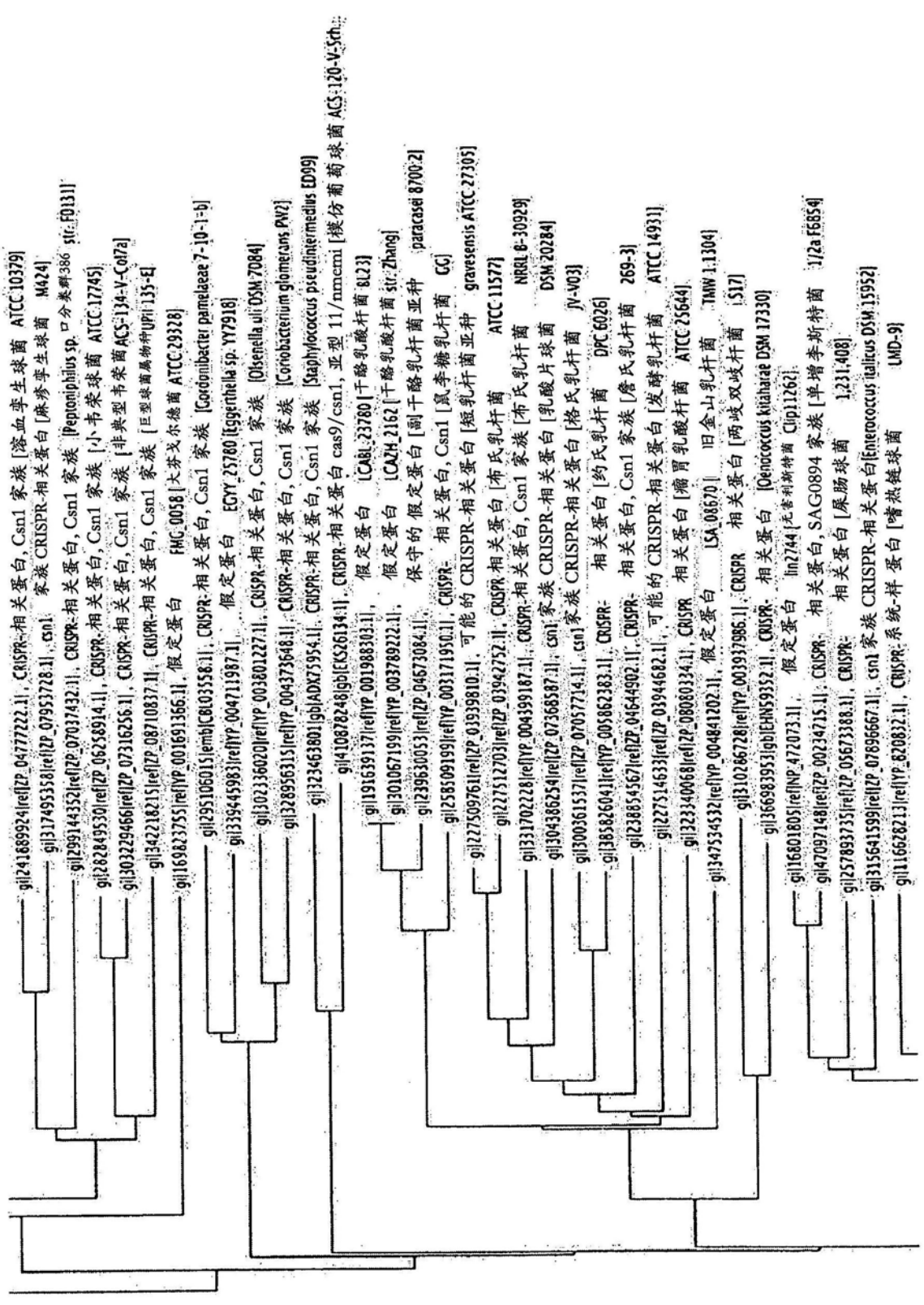

gi|241689924|ref|ZP_04777222.1|, CRISPR-相关蛋白, Csn1 家族 [溶血链球菌 ATCC 10379]
gi|317495358|ref|ZP_07953728.1|, csn1 家族 CRISPR-相关蛋白 [麻疹链球菌 M424]
gi|299144352|ref|ZP_07037432.1|, CRISPR-相关蛋白, Csn1 家族 [Peptoniphilus sp. 口分类群386 str. FD1311]
gi|282849539|ref|ZP_06258914.1|, CRISPR-相关蛋白, Csn1 家族 [小韦荣球菌 ATCC 17745]
gi|303229466|ref|ZP_07316256.1|, CRISPR-相关蛋白, Csn1 家族 [非典型韦荣菌 ACS-134-V-Col7a]
gi|342218215|ref|ZP_08710837.1|, CRISPR-相关蛋白, Csn1 家族 [巨型球菌属物种 UPII 135-E]
gi|169823755|ref|YP_001691366.1|, 假定蛋白 FMG_0058 [大肝戈尔德菌 ATCC 29328]
gi|295106015|emb|CBL03558.1|, CRISPR-相关蛋白, Csn1 家族 [Coriobacteriaceae 7-10-1-b]
gi|339445983|ref|YP_004711987.1|, 假定蛋白 EGYY_25760 [Eggerthella sp. YY7918]
gi|302336020|ref|YP_003801227.1|, CRISPR-相关蛋白, Csn1 家族 [Olsenella uli DSM 7084]
gi|328956315|ref|YP_004373648.1|, CRISPR-相关蛋白, Csn1 家族 [Coriobacterium glomerans PW2]
gi|323463801|gb|ADX75954.1|, CRISPR-相关蛋白, Csn1 家族 [Staphylococcus pseudintermedius ED99]
gi|410878248|gb|CCK26134.1|, CRISPR-相关蛋白 cas9/csn1, 亚型 11/nmemi [模仿葡萄球菌 ACS-120-V-Sch...
gi|191639137|ref|YP_001988303.1|, 假定蛋白 LCABL_23780 [干酪乳杆菌 BL23]
gi|301067199|ref|YP_003789222.1|, 假定蛋白 LCAZH_2162 [干酪乳杆菌 str. Zhang]
gi|239630053|ref|ZP_04573084.1|, 保守的假定蛋白 [副干酪乳杆菌亚种 paracasei 8700:2]
gi|258509199|ref|YP_003171950.1|, CRISPR-相关蛋白, Csn1 [鼠李糖乳杆菌 GG]
gi|227509761|ref|ZP_03939810.1|, 可能的 CRISPR-相关蛋白 [布氏乳杆菌亚种 gravesensis ATCC 27305]
gi|227512703|ref|ZP_03942752.1|, CRISPR 相关蛋白 [布氏乳杆菌 ATCC 11577]
gi|331702228|ref|YP_004399187.1|, CRISPR-相关蛋白, Csn1 家族 [布氏乳杆菌 NRRL B-30929]
gi|304386254|ref|ZP_07368587.1|, csn1家族 CRISPR-相关蛋白 [乳酸片球菌 DSM 20284]
gi|300361537|ref|ZP_07057714.1|, csn1家族 CRISPR-相关蛋白 [格氏乳杆菌 JV-V03]
gi|385826041|ref|YP_005862383.1|, CRISPR-相关蛋白 [约氏乳杆菌 DPC 6026]
gi|238854567|ref|ZP_04644902.1|, CRISPR-相关蛋白, Csn1 家族 [詹氏乳杆菌 269-3]
gi|227514633|ref|ZP_03944682.1|, 可能的 CRISPR-相关蛋白 [发酵乳杆菌 ATCC 14931]
gi|323340068|ref|ZP_08080334.1|, CRISPR 相关蛋白 [瘤胃乳酸杆菌 ATCC 25644]
gi|347534532|ref|YP_004841202.1|, 假定蛋白 LSA_08670 [旧金山乳杆菌 TMW 1.1304]
gi|310286728|ref|YP_003937986.1|, CRISPR 相关蛋白 [两歧双岐杆菌 JCM 1217]
gi|366983953|gb|EHN59352.1|, CRISPR-相关蛋白 [Oenococcus kitaharae DSM 17330]
gi|168018058|ref|NP_472073.1|, 假定蛋白 lin2744 [无害利斯特菌 Clip11262]
gi|470971488|ref|ZP_00234715.1|, CRISPR-相关蛋白, SAG0894 家族 [单增李斯特菌 1/2a F6854]
gi|257893735|ref|ZP_05673388.1|, CRISPR-相关蛋白 [屎肠球菌 1,231.408]
gi|315641599|ref|ZP_07896667.1|, csn1 家族 CRISPR-相关蛋白[Enterococcus italicus DSM 15952]
gi|116628213|ref|YP_820832.1|, CRISPR-系统-样蛋白 [嗜热链球菌 LMD-9]

图20E

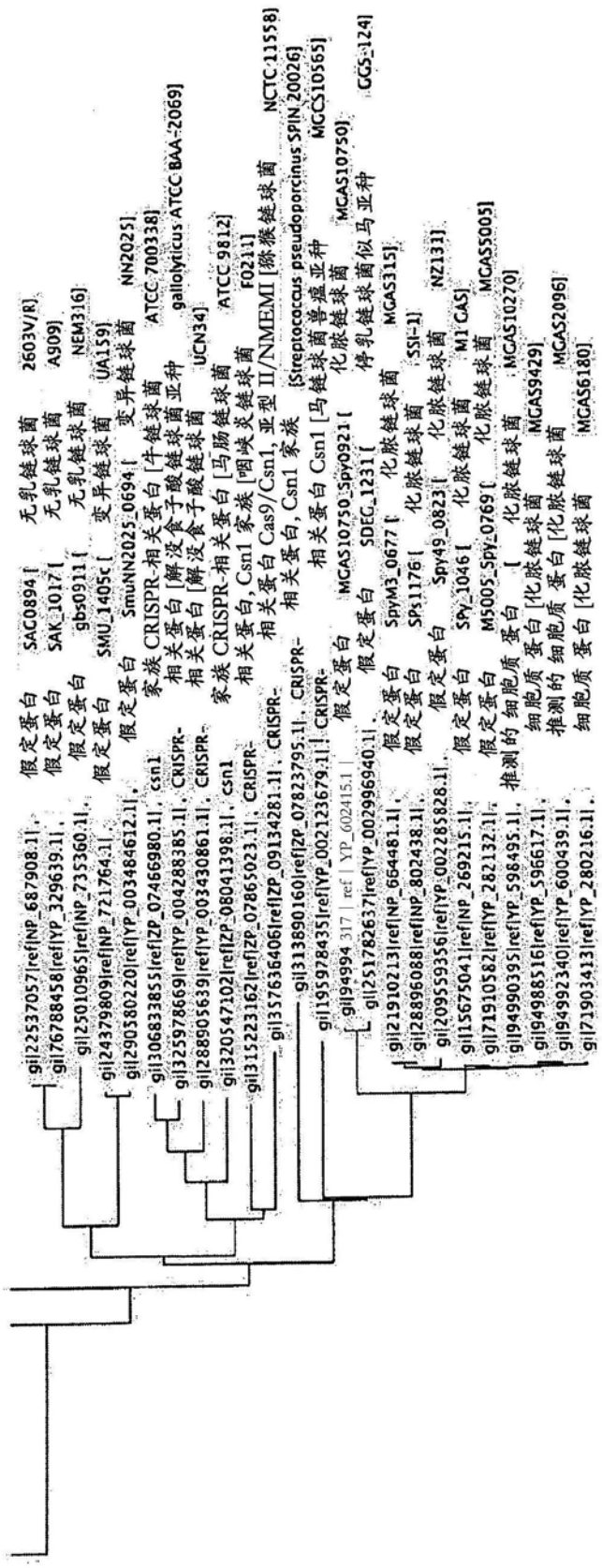

gi|22537057|ref|NP_687908.1|, 假定蛋白 SAG0894 [无乳链球菌 2603V/R]
gi|76788458|ref|YP_329639.1|, 假定蛋白 SAK_1017 [无乳链球菌 A909]
gi|25010965|ref|NP_735360.1|, 假定蛋白 gbs0911 [无乳链球菌 NEM316]
gi|24379809|ref|NP_721764.1|, 假定蛋白 SMU_1405c [变异链球菌 UA159]
gi|29058022O|ref|YP_003484612.1|, 假定蛋白 SmuNN2025_0694 [变异链球菌 NN2025]
gi|30683385 5|ref|ZP_07466980.1|, csn1 家族 CRISPR-相关蛋白 [牛链球菌 ATCC 700338]
gi|32597866 9|ref|YP_004288385.1|, CRISPR-相关蛋白 [解没食子酸链球菌亚种 gallolyticus ATCC BAA-2069]
gi|28895639|ref|YP_003430861.1|, CRISPR-家族 CRISPR-相关蛋白 [解没食子酸链球菌 UCN34]
gi|32054710 2|ref|ZP_08041398.1|, csn1 家族 CRISPR-相关蛋白,Csn1 家族 [马肠链球菌 ATCC 9812]
gi|31522316 2|ref|ZP_07865023.1|, CRISPR-相关蛋白 [咽峡炎链球菌 F0211]
gi|35763640 6|ref|ZP_09134281.1|, CRISPR-相关蛋白 Cas9/Csn1, 亚型 II/NMEMI [猕猴链球菌 NCTC 11558]
gi|31389016 0|ref|ZP_07823795.1|, CRISPR-相关蛋白, Csn1 家族 [Streptococcus pseudoporcinus SPIN 20026]
gi|19597843 5|ref|YP_002123679.1|, CRISPR-相关蛋白 Csn1 [马链球菌兽瘟亚种 MGCS10565]
gi|94994 317 | ref | YP_602415.1 | 假定蛋白 MGAS10750_Spy0921 [化脓链球菌 MGAS10750]
gi|25178263 7|ref|YP_002996940.1|, 假定蛋白 SDEG_1231 [停乳链球菌似马亚种 GGS_124]
gi|21910213|ref|NP_664481.1|, 假定蛋白 SpyM3_0677 [化脓链球菌 MGAS315]
gi|28896088|ref|NP_802438.1|, 假定蛋白 SPs1176 [化脓链球菌 SSI-1]
gi|20955935 6|ref|YP_002285828.1|, 假定蛋白 Spy49_0823 [化脓链球菌 NZ131]
gi|15675041|ref|NP_269215.1|, 假定蛋白 SPy_1046 [化脓链球菌 M1 GAS]
gi|71910582|ref|YP_282132.1|, 假定蛋白 M5005_Spy_0769 [化脓链球菌 MGAS5005]
gi|19490395|ref|YP_598495.1|, 推测的细胞质蛋白 [化脓链球菌 MGAS10270]
gi|19498816|ref|YP_596617.1|, 细胞质蛋白 [化脓链球菌 MGAS9429]
gi|19499234 0|ref|YP_600439.1|, 推测的细胞质蛋白 [化脓链球菌 MGAS2096]
gi|71903413|ref|YP_280216.1|, 细胞质蛋白 [化脓链球菌 MGAS6180]

图20F

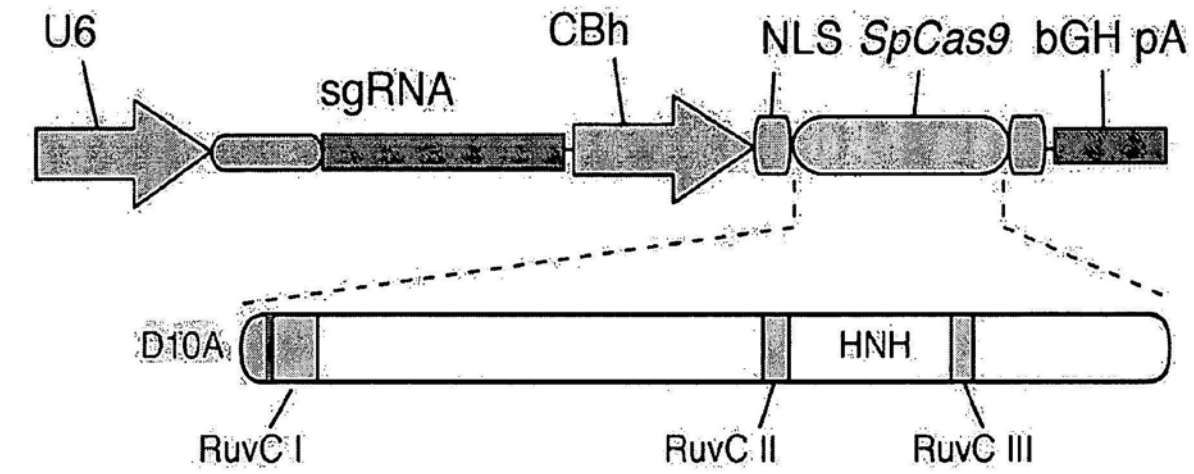

a
U6
sgRNA
CBh
NLS SpCas9 bGH pA
D10A
HNH
RuvC I
RuvC II
RuvC III

图21A

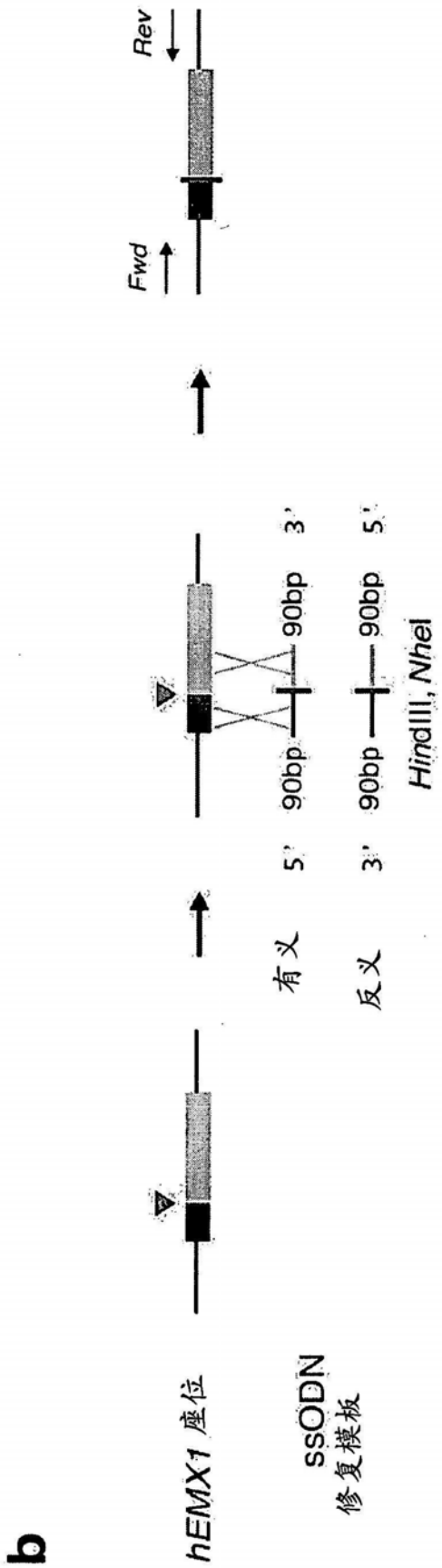

b
hEMX1 座位
ssODN
修复模板
有义
反义
5'
3'
90bp
90bp
90bp
90bp
3'
5'
HindIII, NheI
Fwd
Rev

图21B

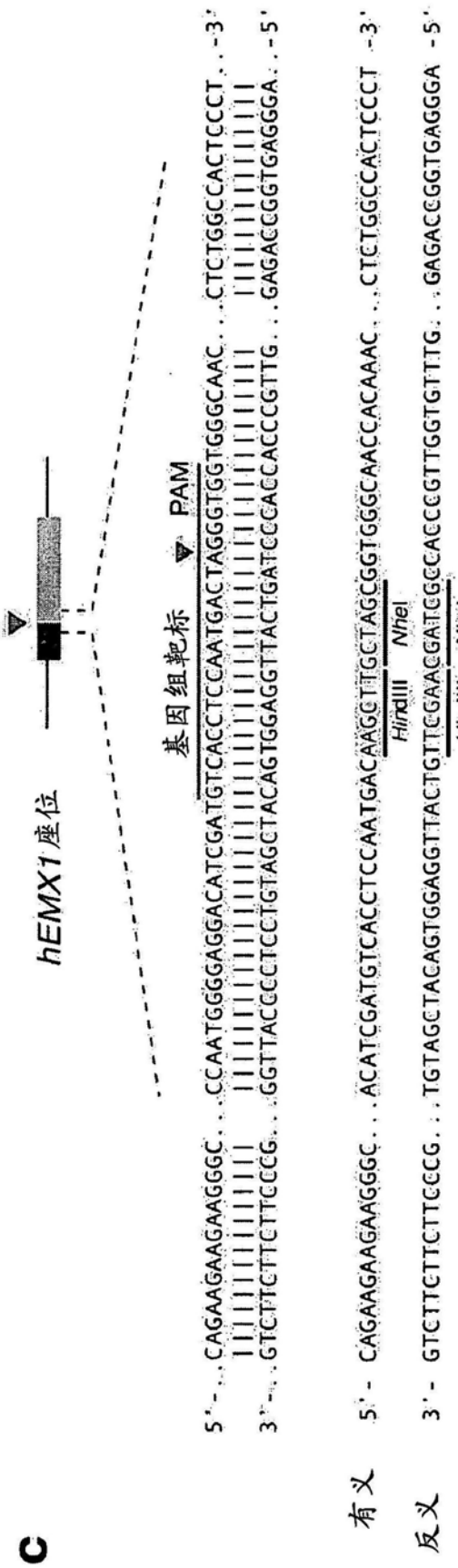

C
hEMX1座位
基因组靶标
PAM
5'-...CAGAAGAAGAAGGGC...CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC...CTCTGGCCACTCCCT...-3'
3'-...GTCTTCTTCTTCCCG...GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG...GAGACCGGTGAGGGA...-5'
有义 5'- CAGAAGAAGAAGGGC...ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACACAAAC...CTCTGGCCACTCCCT -3'
HindIII NheI
反义 3'- GTCTTCTTCTTCCCG...TGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGGCCACCCGTTGGTGTGTTTG...GAGACCGGTGAGGGA -5'
HindIII NheI

图21C

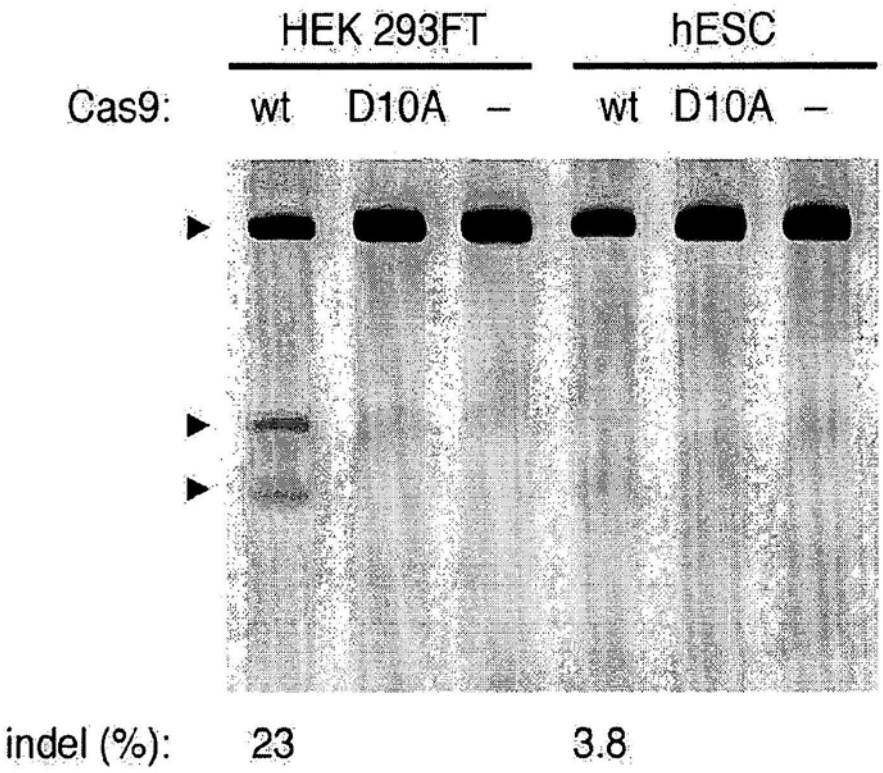


图21D

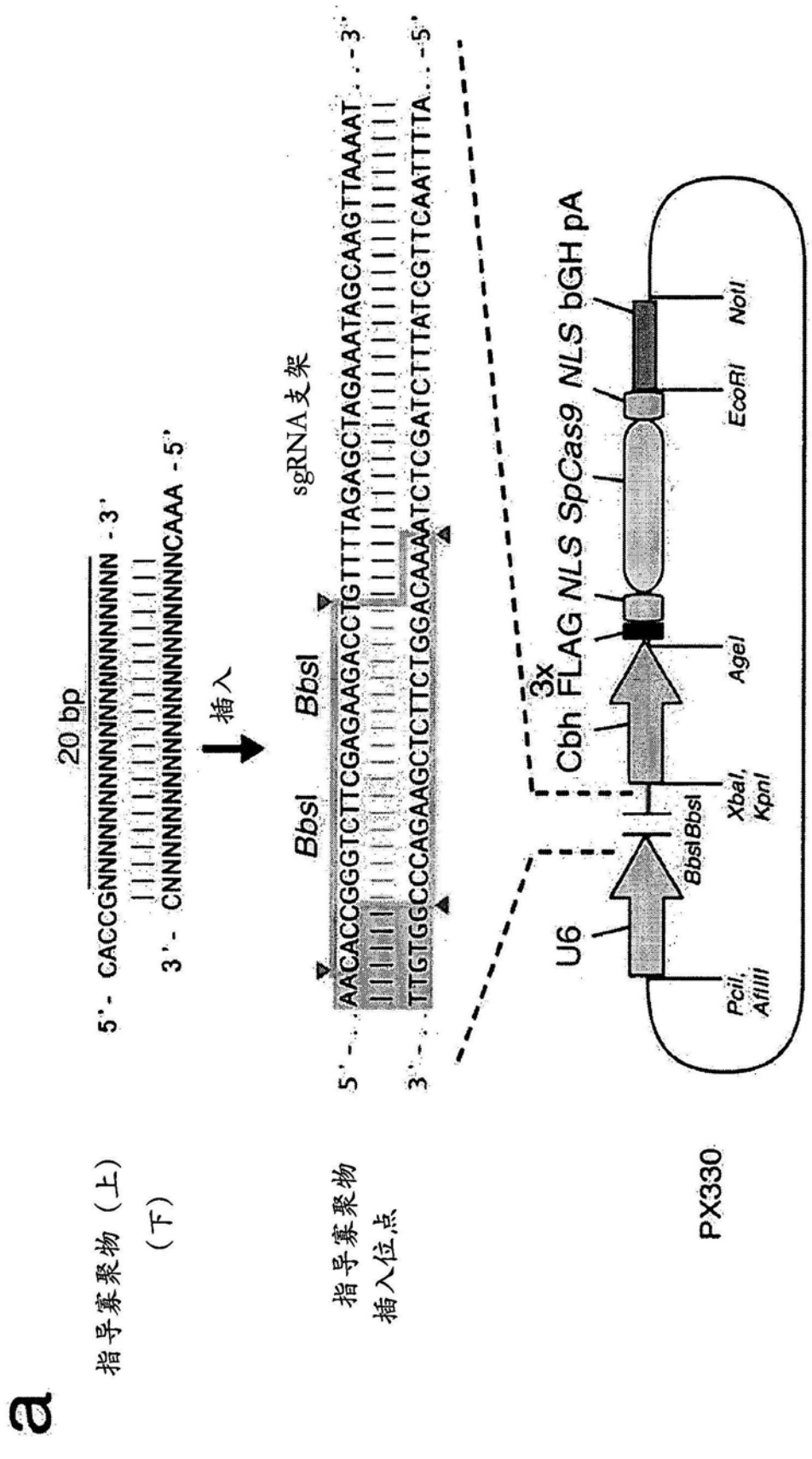

a
指导寡聚物（上）
（下）
20 bp
5'- CACCGNNNNNNNNNNNNNNNNNNN -3'
3'- CNNNNNNNNNNNNNNNNNNNNNCAAA -5'
插入
BbsI
BbsI
sgRNA支架
5'- AACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAAT -3'
3'- TTGTGGCCCAGAAGCTCTTCTGGACAAAATCTCGATCTTTATCGTTCAATTTTA -5'
指导寡聚物
插入位点
U6
3x
Cbh FLAG NLS SpCas9 NLS bGH pA
PciI, AflIII
BbsIBbsI
XbaI, KpnI
AgeI
EcoRI
NotI
PX330

图22A

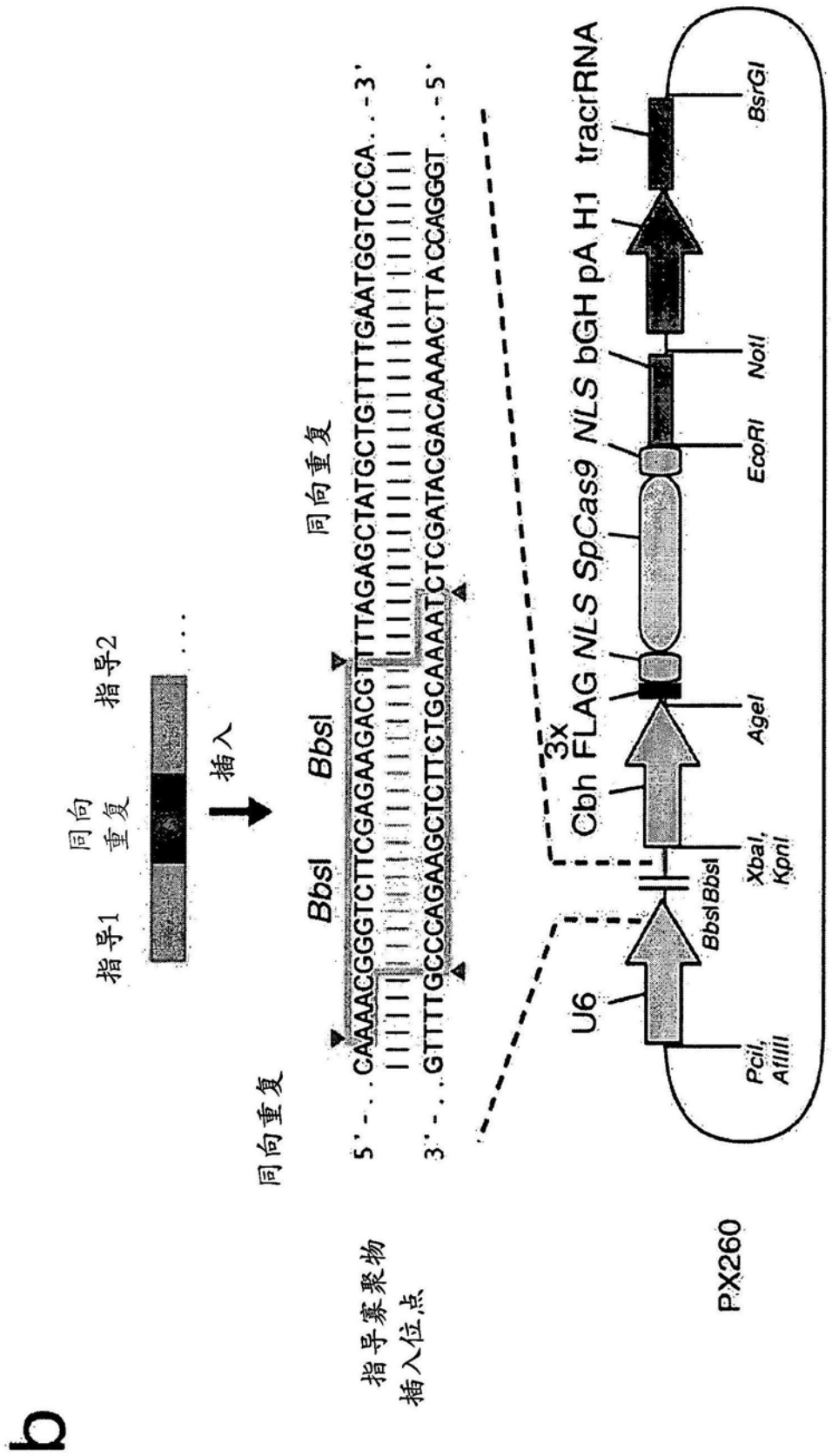

b
指导1
同向
重复
指导2
插入
同向重复
BbsI
BbsI
同向重复
5'-...CAAAACGGGTCTTCGAGAAGACGTTTTAGAGCTATGCTGTTTTGAATGGTCCCA...-3'
3'-...GTTTTGCCCAGAAGCTCTTCTGCAAAATCTCGATACGACAAAACTTACCAGGGT...-5'
指导寡聚物
插入位点
U6
3x
Cbh FLAG NLS SpCas9 NLS bGH pA H1 tracrRNA
PciI,
AflIII
BbsIBbsI
XbaI,
KpnI
AgeI
EcoRI
NotI
BsrGI
PX260

图22B

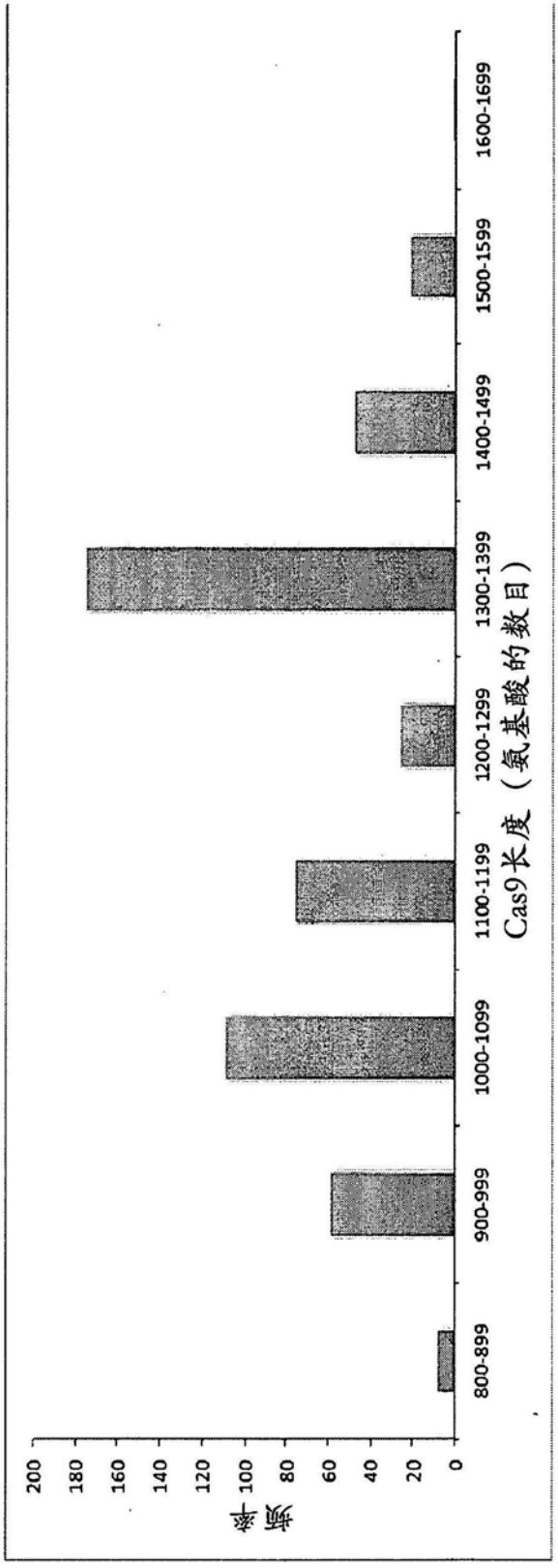

频率
0
20
40
60
80
100
120
140
160
180
200
800-899
900-999
1000-1099
1100-1199
1200-1299
1300-1399
1400-1499
1500-1599
1600-1699
Cas9长度（氨基酸的数目）

图23

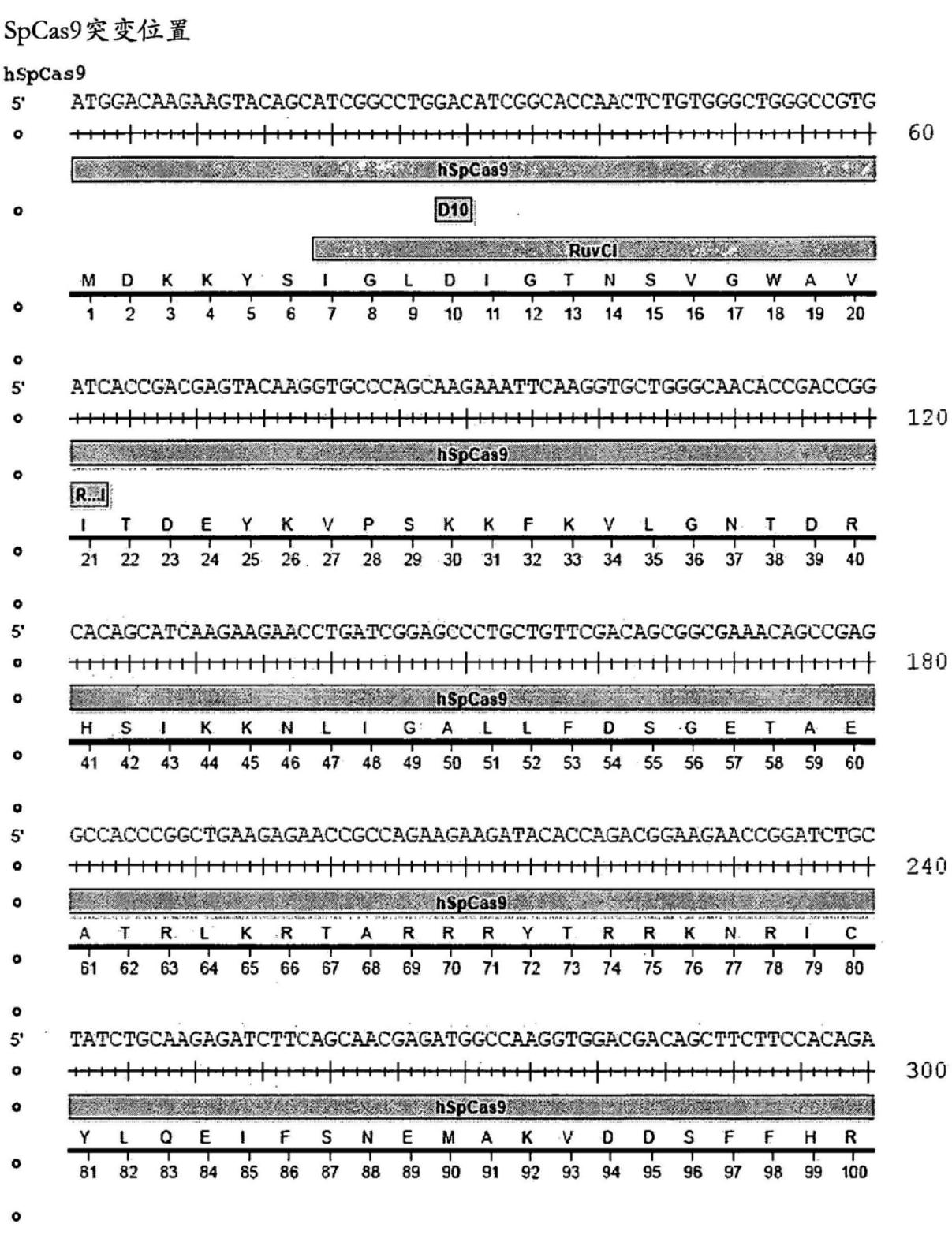

SpCas9突变位置
hSpCas9
5' ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG 60
hSpCas9
D10
RuvCI
M D K K Y S I G L D I G T N S V G W A V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5' ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG 120
hSpCas9
R...I
I T D E Y K V P S K K F K V L G N T D R
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
5' CACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAG 180
hSpCas9
H S I K K N L I G A L L F D S G E T A E
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
5' GCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC 240
hSpCas9
A T R L K R T A R R R Y T R R K N R I C
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
5' TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA 300
hSpCas9
Y L Q E I F S N E M A K V D D S F F H R
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

图24A

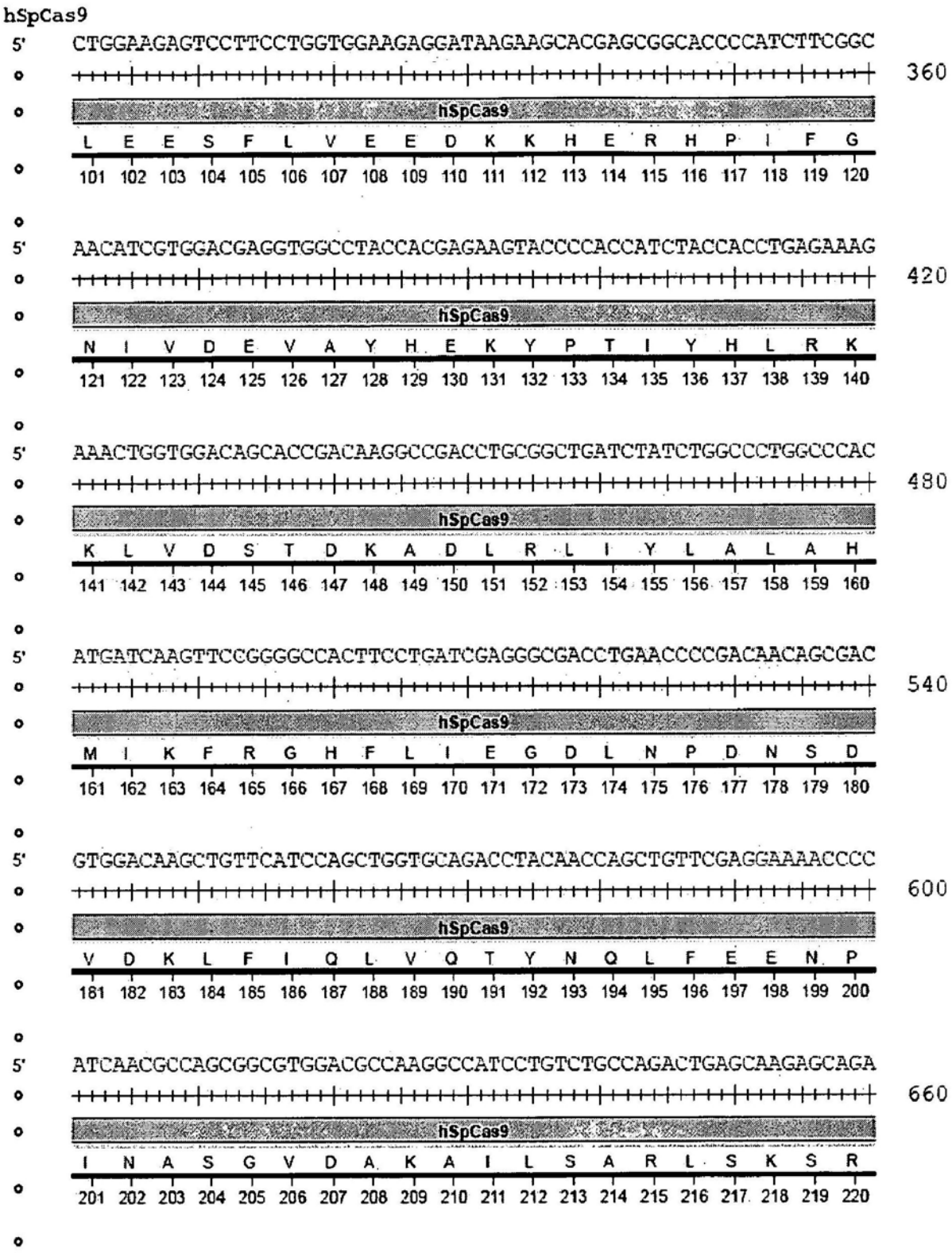

hSpCas9
5' CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC
360
hSpCas9
L E E S F L V E E D K K H E R H P I F G
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
5' AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAG
420
hSpCas9
N I V D E V A Y H E K Y P T I Y H L R K
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
5' AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC
480
hSpCas9
K L V D S T D K A D L R L I Y L A L A H
141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
5' ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
540
hSpCas9
M I K F R G H F L I E G D L N P D N S D
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
5' GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC
600
hSpCas9
V D K L F I Q L V Q T Y N Q L F E E N P
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
5' ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
660
hSpCas9
I N A S G V D A K A I L S A R L S K S R
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220

图24B

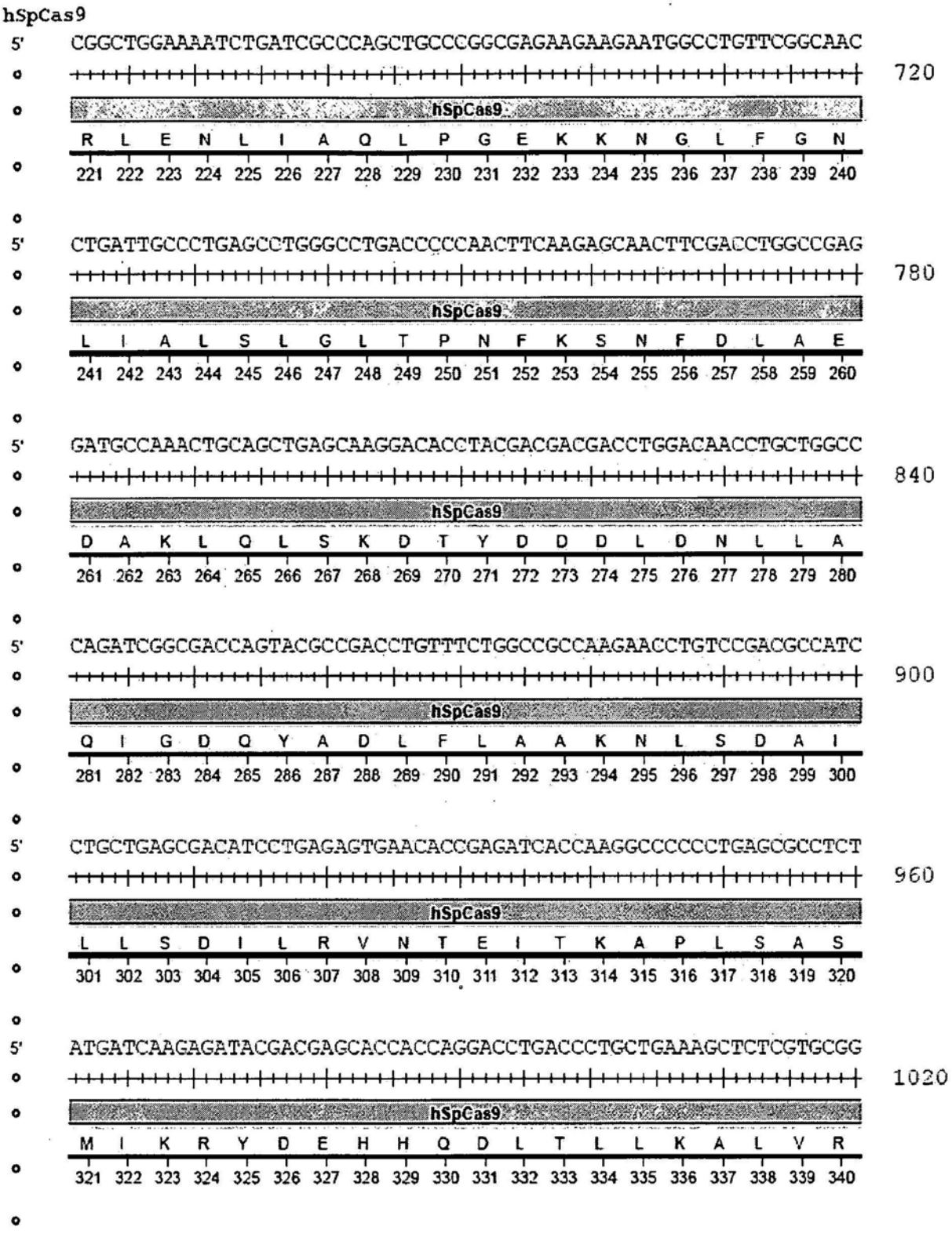

hSpCas9
5' CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC
720
hSpCas9
R L E N L I A Q L P G E K K N G L F G N
221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
5' CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAG
780
hSpCas9
L I A L S L G L T P N F K S N F D L A E
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
5' GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCC
840
hSpCas9
D A K L Q L S K D T Y D D D L D N L L A
261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
5' CAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATC
900
hSpCas9
Q I G D Q Y A D L F L A A K N L S D A I
281 282 283 284 265 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
5' CTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT
960
hSpCas9
L L S D I L R V N T E I T K A P L S A S
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
5' ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGG
1020
hSpCas9
M I K R Y D E H H Q D L T L L K A L V R
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340

图24C

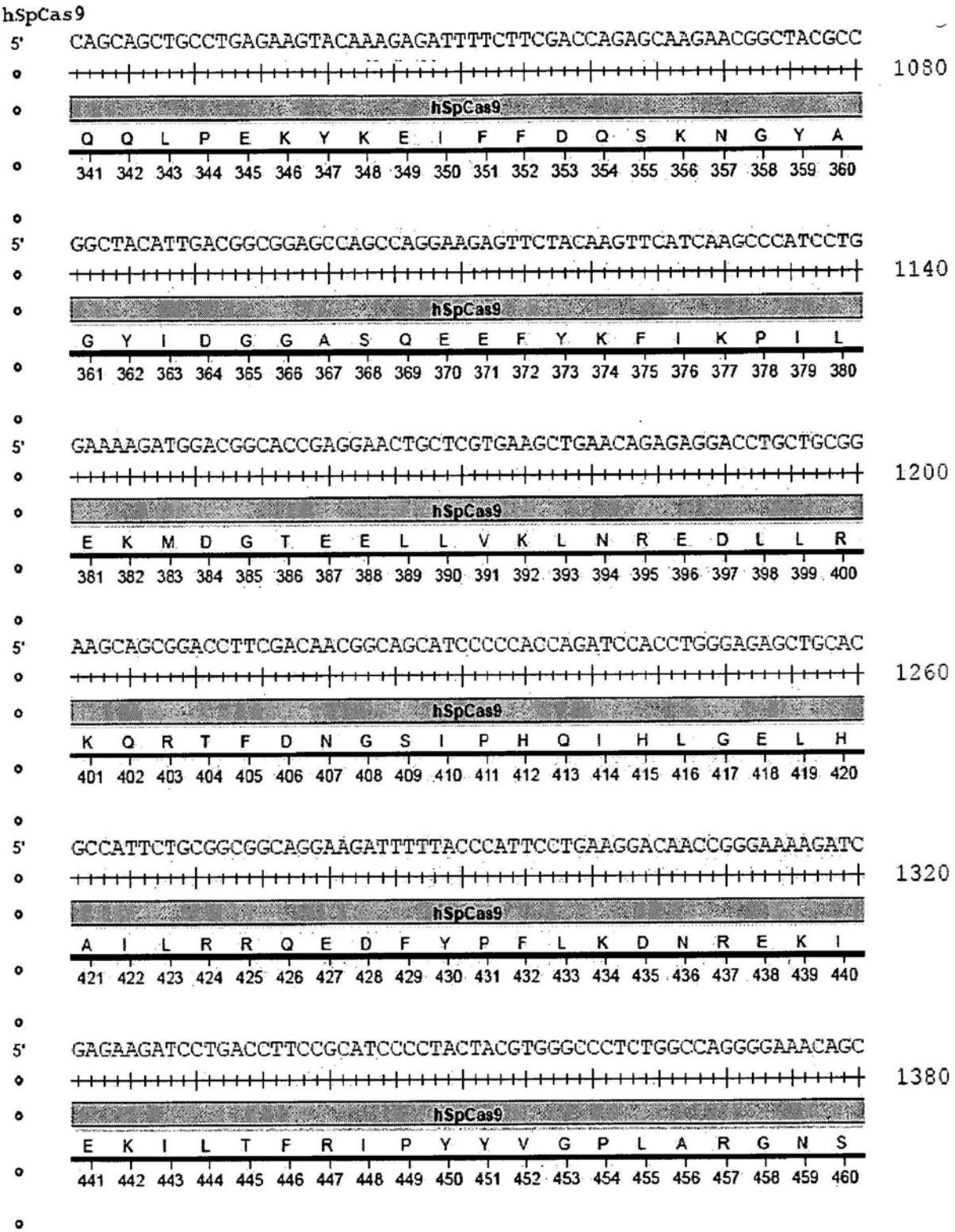

hSpCas9
5' CAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
1080
hSpCas9
Q Q L P E K Y K E I F F D Q S K N G Y A
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
5' GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTG
1140
hSpCas9
G Y I D G G A S Q E E F Y K F I K P I L
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
5' GAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG
1200
hSpCas9
E K M D G T E E L L V K L N R E D L L R
381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
5' AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCAC
1260
hSpCas9
K Q R T F D N G S I P H Q I H L G E L H
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
5' GCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATC
1320
hSpCas9
A I L R R Q E D F Y P F L K D N R E K I
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
5' GAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC
1380
hSpCas9
E K I L T F R I P Y Y V G P L A R G N S
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460

图24D

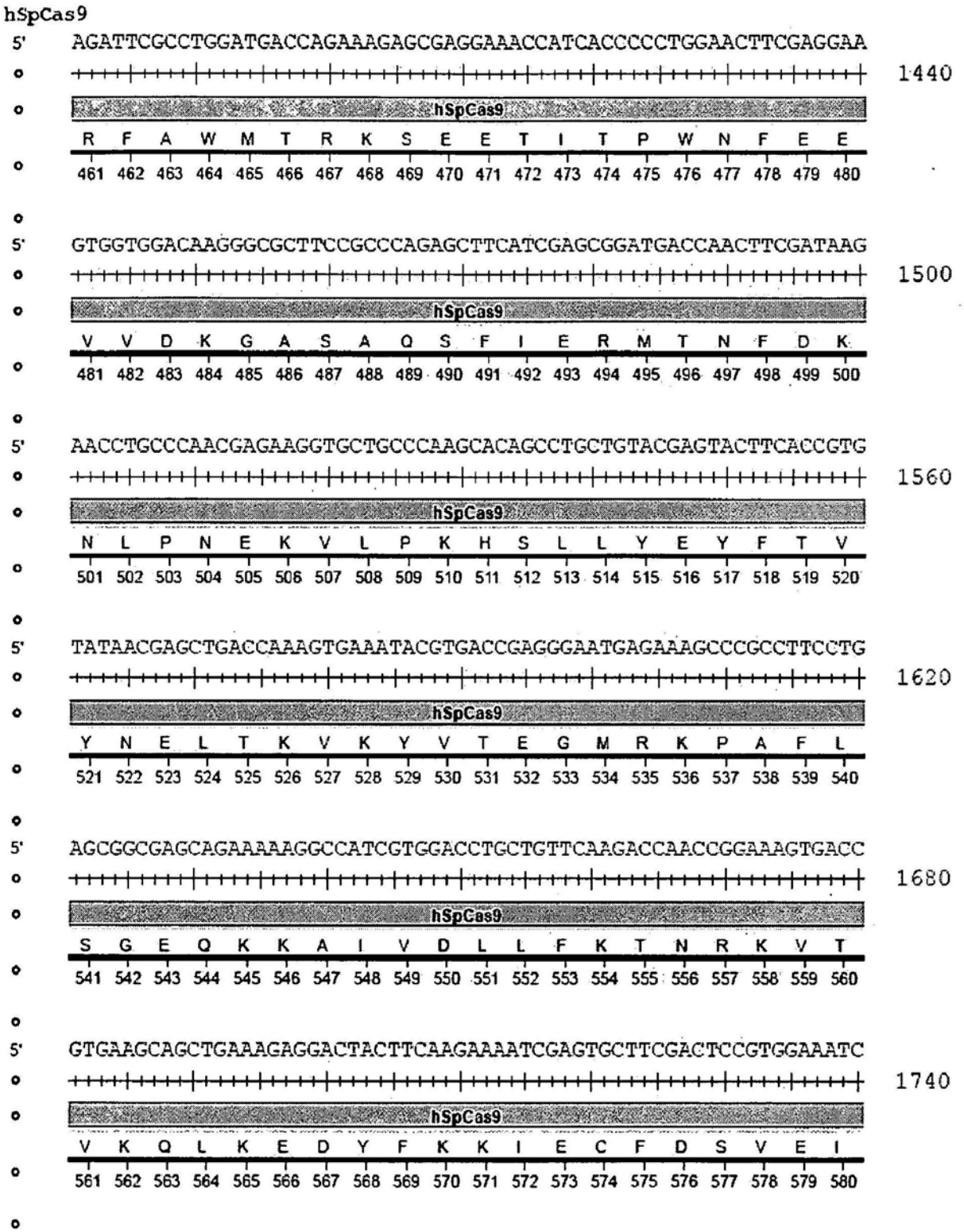

hSpCas9
5' AGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA
1440
hSpCas9
R F A W M T R K S E E T I T P W N F E E
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
5' GTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
1500
hSpCas9
V V D K G A S A Q S F I E R M T N F D K
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
5' AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
1560
hSpCas9
N L P N E K V L P K H S L L Y E Y F T V
501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
5' TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTG
1620
hSpCas9
Y N E L T K V K Y V T E G M R K P A F L
521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
5' AGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
1680
hSpCas9
S G E Q K K A I V D L L F K T N R K V T
541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
5' GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC
1740
hSpCas9
V K Q L K E D Y F K K I E C F D S V E I
561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

图24E

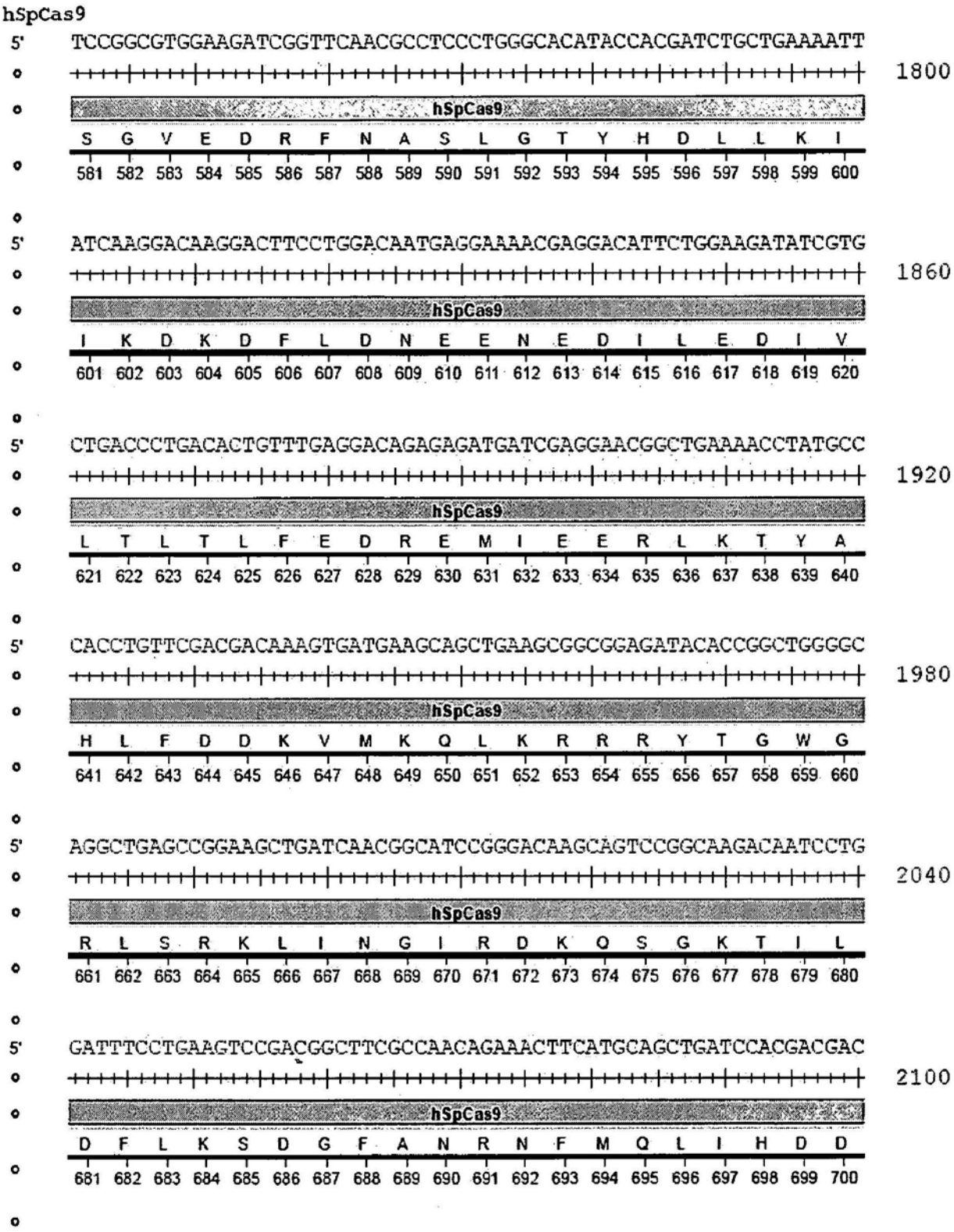

hSpCas9
5' TCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
1800
hSpCas9
S G V E D R F N A S L G T Y H D L L K I
581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
5' ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTG
1860
hSpCas9
I K D K D F L D N E E N E D I L E D I V
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620
5' CTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
1920
hSpCas9
L T L T L F E D R E M I E E R L K T Y A
621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
5' CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC
1980
hSpCas9
H L F D D K V M K Q L K R R R Y T G W G
641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
5' AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
2040
hSpCas9
R L S R K L I N G I R D K Q S G K T I L
661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
5' GATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC
2100
hSpCas9
D F L K S D G F A N R N F M Q L I H D D
681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700

图24F

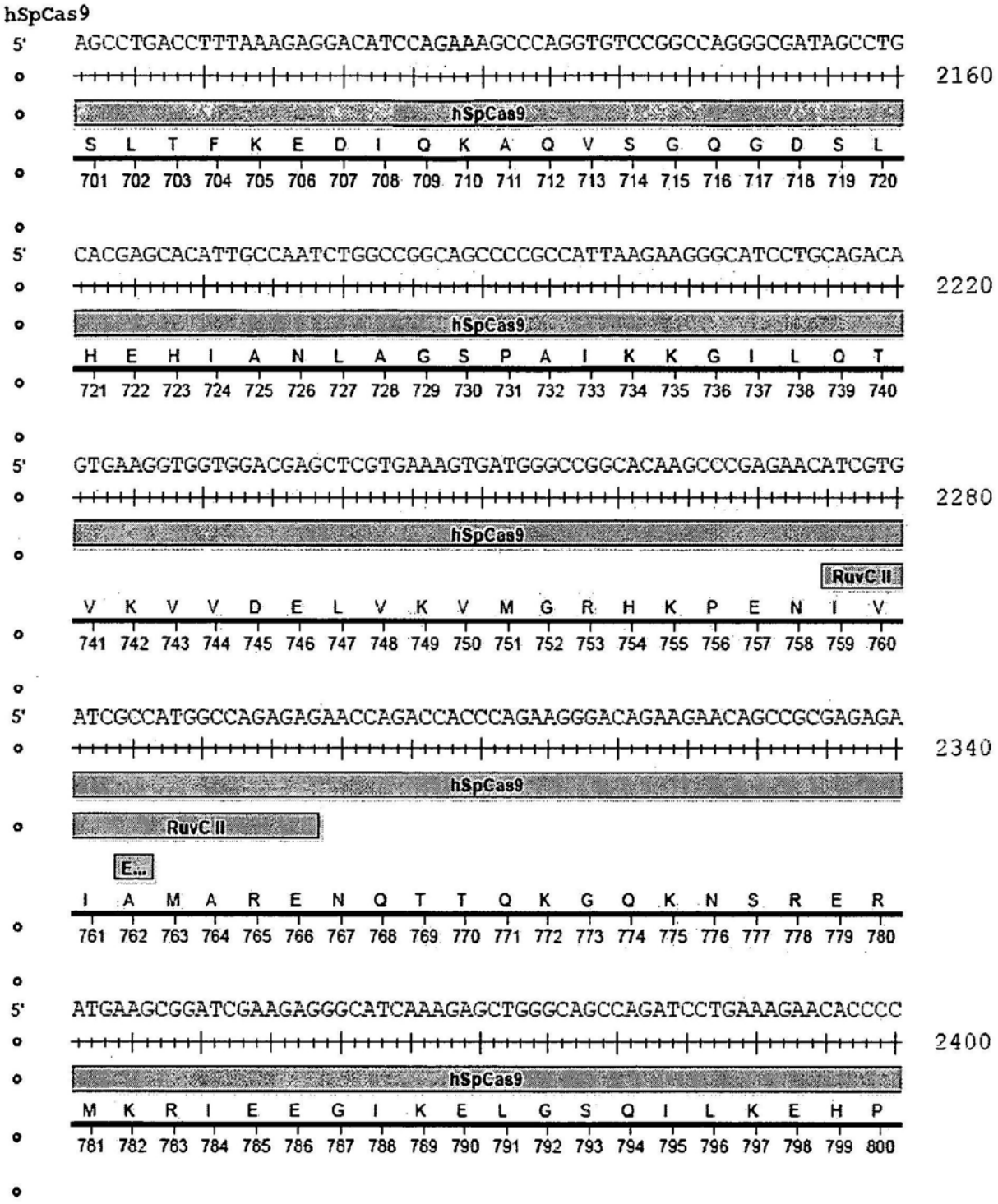

hSpCas9
5' AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
2160
hSpCas9
S L T F K E D I Q K A Q V S G Q G D S L
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720
5' CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACA
2220
hSpCas9
H E H I A N L A G S P A I K K G I L Q T
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
5' GTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTG
2280
hSpCas9
RuvC II
V K V V D E L V K V M G R H K P E N I V
741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
5' ATCGCCATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
2340
hSpCas9
RuvC II
E...
I A M A R E N Q T T Q K G Q K N S R E R
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
5' ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
2400
hSpCas9
M K R I E E G I K E L G S Q I L K E H P
781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800

图24G

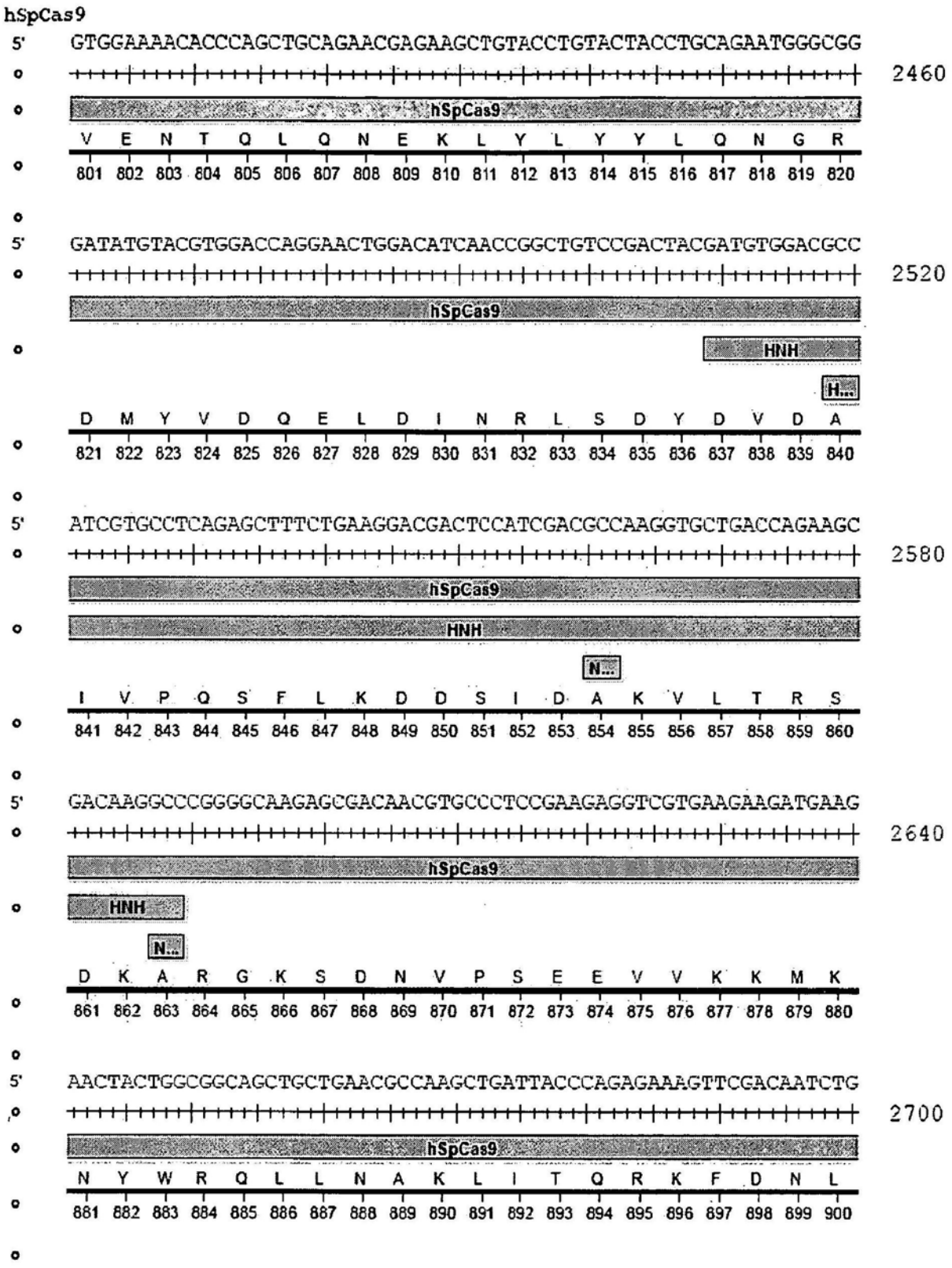

hSpCas9
5' GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGG
2460
hSpCas9
V E N T Q L Q N E K L Y L Y Y L Q N G R
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
5' GATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCC
2520
hSpCas9
HNH
H...
D M Y V D Q E L D I N R L S D Y D V D A
821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
5' ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACGCCAAGGTGCTGACCAGAAGC
2580
hSpCas9
HNH
N...
I V P Q S F L K D D S I D A K V L T R S
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
5' GACAAGGCCCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAG
2640
hSpCas9
HNH
N...
D K A R G K S D N V P S E E V V K K M K
861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
5' AACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTG
2700
hSpCas9
N Y W R Q L L N A K L I T Q R K F D N L
881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900

图24H

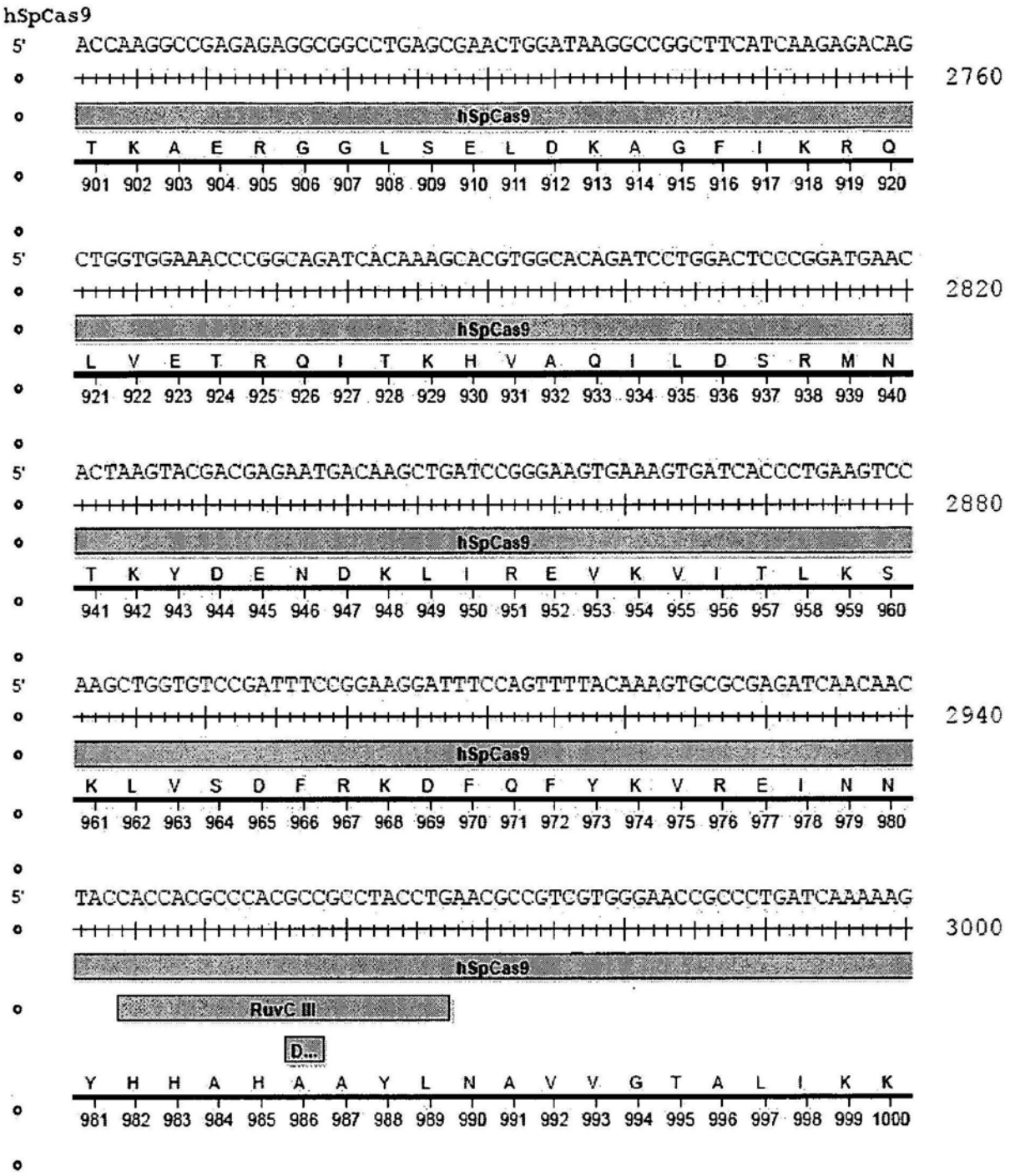

hSpCas9
5' ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
2760
hSpCas9
T K A E R G G L S E L D K A G F I K R Q
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
5' CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAAC
2820
hSpCas9
L V E T R Q I T K H V A Q I L D S R M N
921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
5' ACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCC
2880
hSpCas9
T K Y D E N D K L I R E V K V I T L K S
941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
5' AAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAAC
2940
hSpCas9
K L V S D F R K D F Q F Y K V R E I N N
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
5' TACCACCACGCCCACGCCGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
3000
hSpCas9
RuvC III
D...
Y H H A H A A Y L N A V V G T A L I K K
981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

图24I

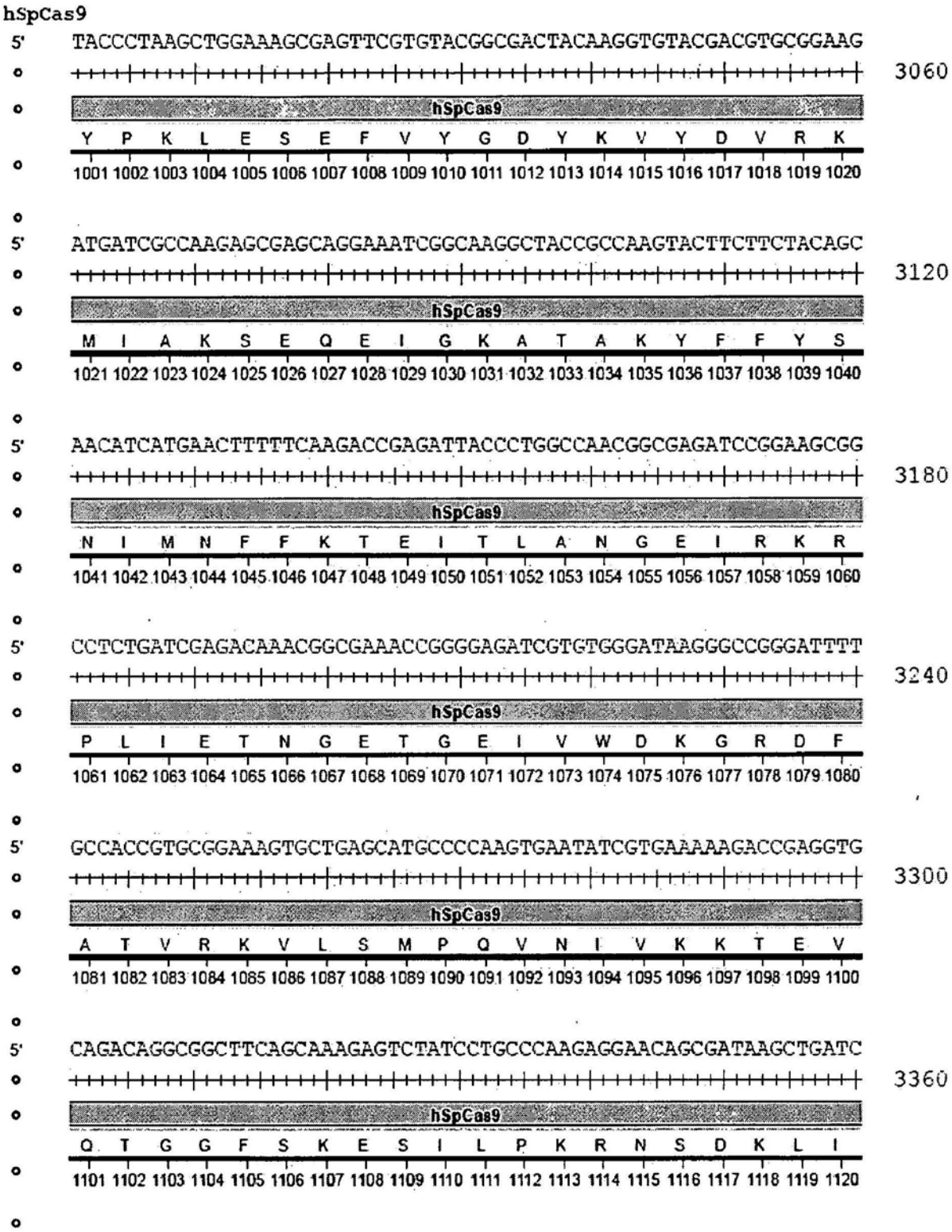

hSpCas9
5' TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAG
3060
hSpCas9
Y P K L E S E F V Y G D Y K V Y D V R K
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
5' ATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
3120
hSpCas9
M I A K S E Q E I G K A T A K Y F F Y S
1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
5' AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
3180
hSpCas9
N I M N F F K T E I T L A N G E I R K R
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
5' CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT
3240
hSpCas9
P L I E T N G E T G E I V W D K G R D F
1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
5' GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG
3300
hSpCas9
A T V R K V L S M P Q V N I V K K T E V
1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
5' CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATC
3360
hSpCas9
Q T G G F S K E S I L P K R N S D K L I
1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120

图24J

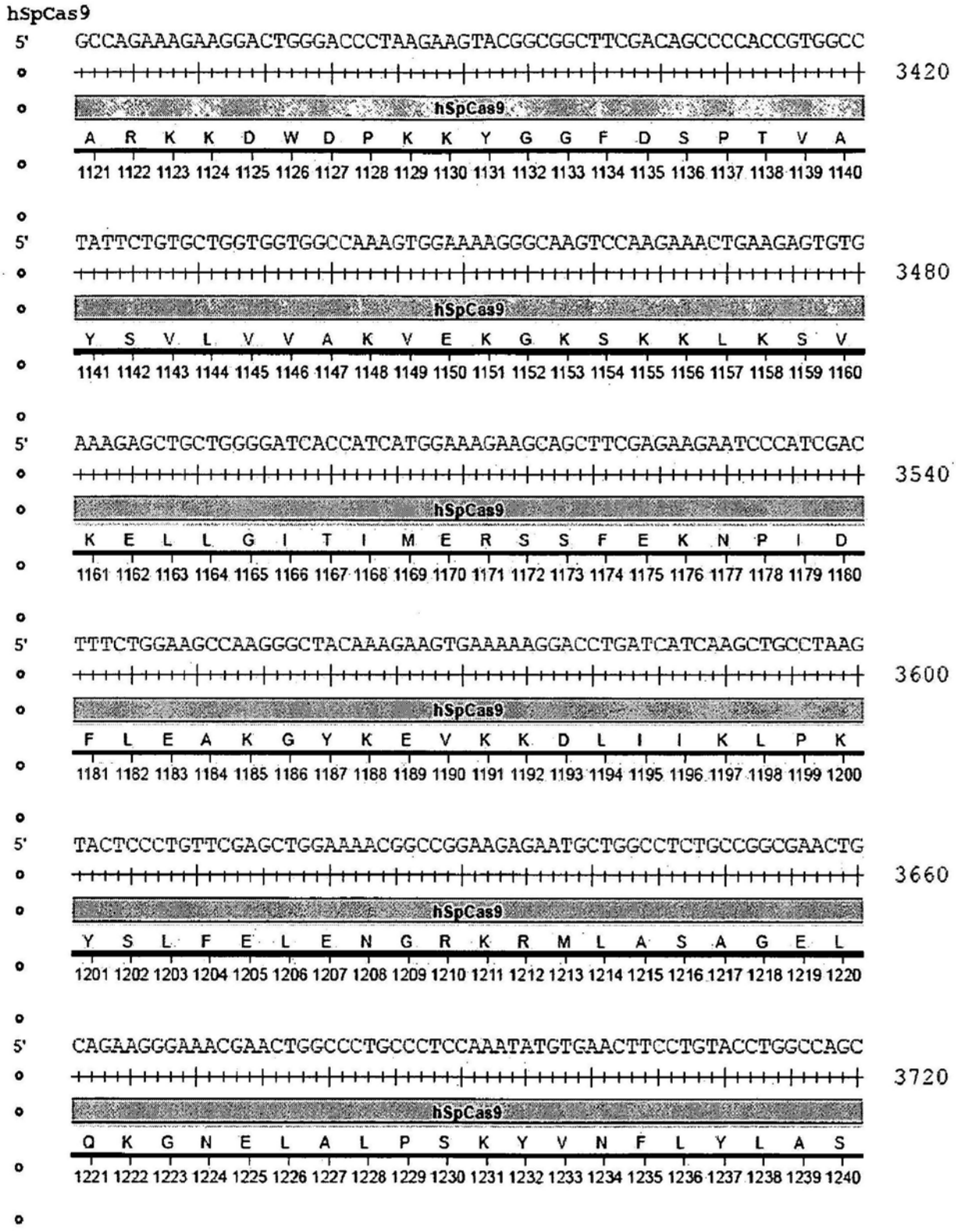

hSpCas9
5' GCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCC
3420
hSpCas9
A R K K D W D P K K Y G G F D S P T V A
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
5' TATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTG
3480
hSpCas9
Y S V L V V A K V E K G K S K K L K S V
1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
5' AAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGAC
3540
hSpCas9
K E L L G I T I M E R S S F E K N P I D
1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
5' TTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG
3600
hSpCas9
F L E A K G Y K E V K K D L I I K L P K
1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
5' TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTG
3660
hSpCas9
Y S L F E L E N G R K R M L A S A G E L
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
5' CAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGC
3720
hSpCas9
Q K G N E L A L P S K Y V N F L Y L A S
1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240

图24K

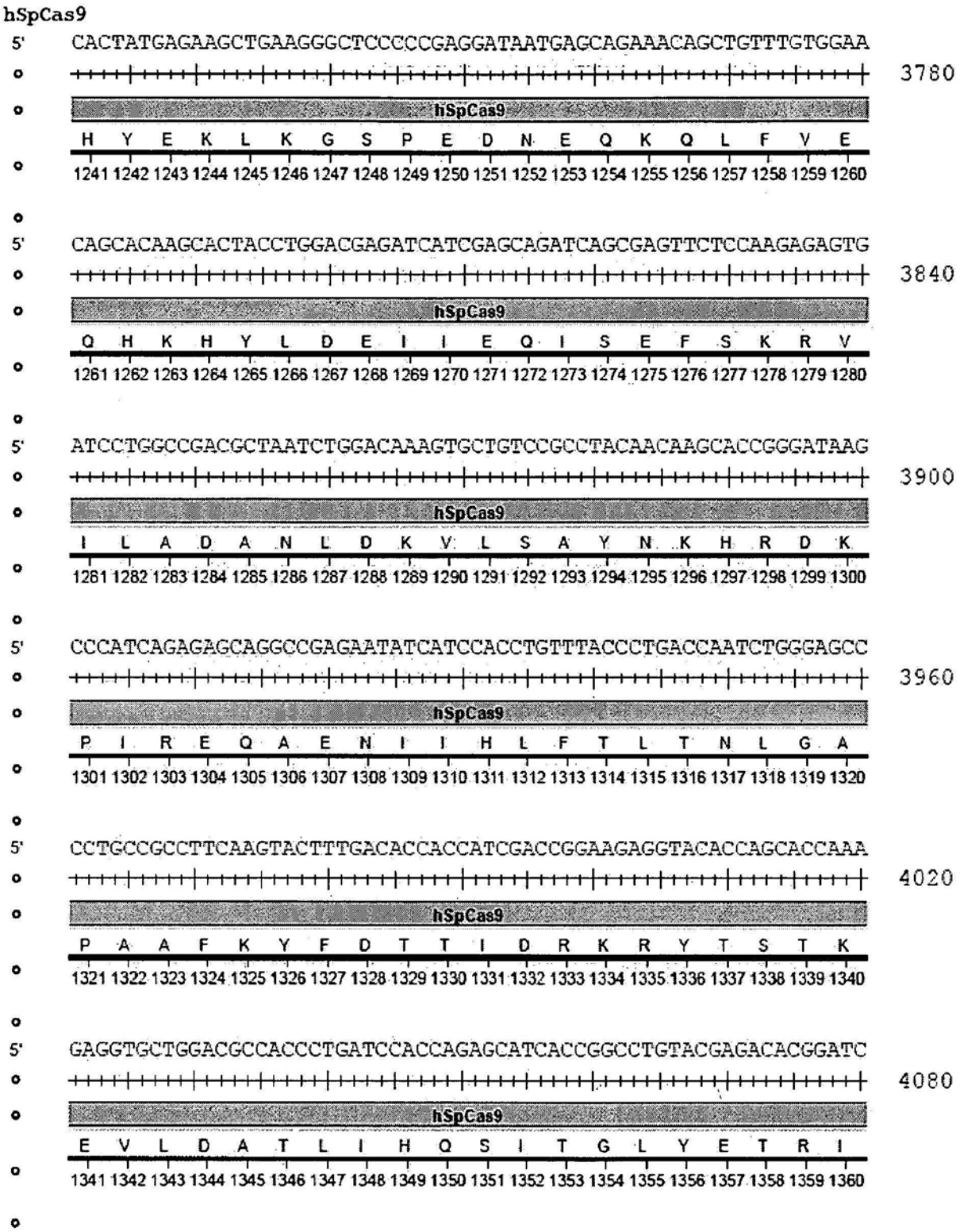

hSpCas9
5' CACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA 3780
hSpCas9
H Y E K L K G S P E D N E Q K Q L F V E
1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
5' CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG 3840
hSpCas9
Q H K H Y L D E I I E Q I S E F S K R V
1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
5' ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG 3900
hSpCas9
I L A D A N L D K V L S A Y N K H R D K
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
5' CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCC 3960
hSpCas9
P I R E Q A E N I I H L F T L T N L G A
1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
5' CCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA 4020
hSpCas9
P A A F K Y F D T T I D R K R Y T S T K
1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
5' GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATC 4080
hSpCas9
E V L D A T L I H Q S I T G L Y E T R I
1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360

图24L

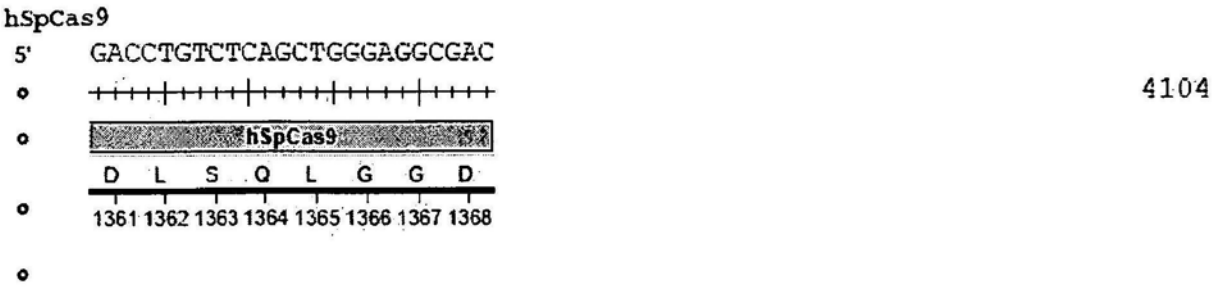

hSpCas9
5'
GACCTGTCTCAGCTGGGAGGCGAC
4104
hSpCas9
D L S Q L G G D
1361 1362 1363 1364 1365 1366 1367 1368

图24M

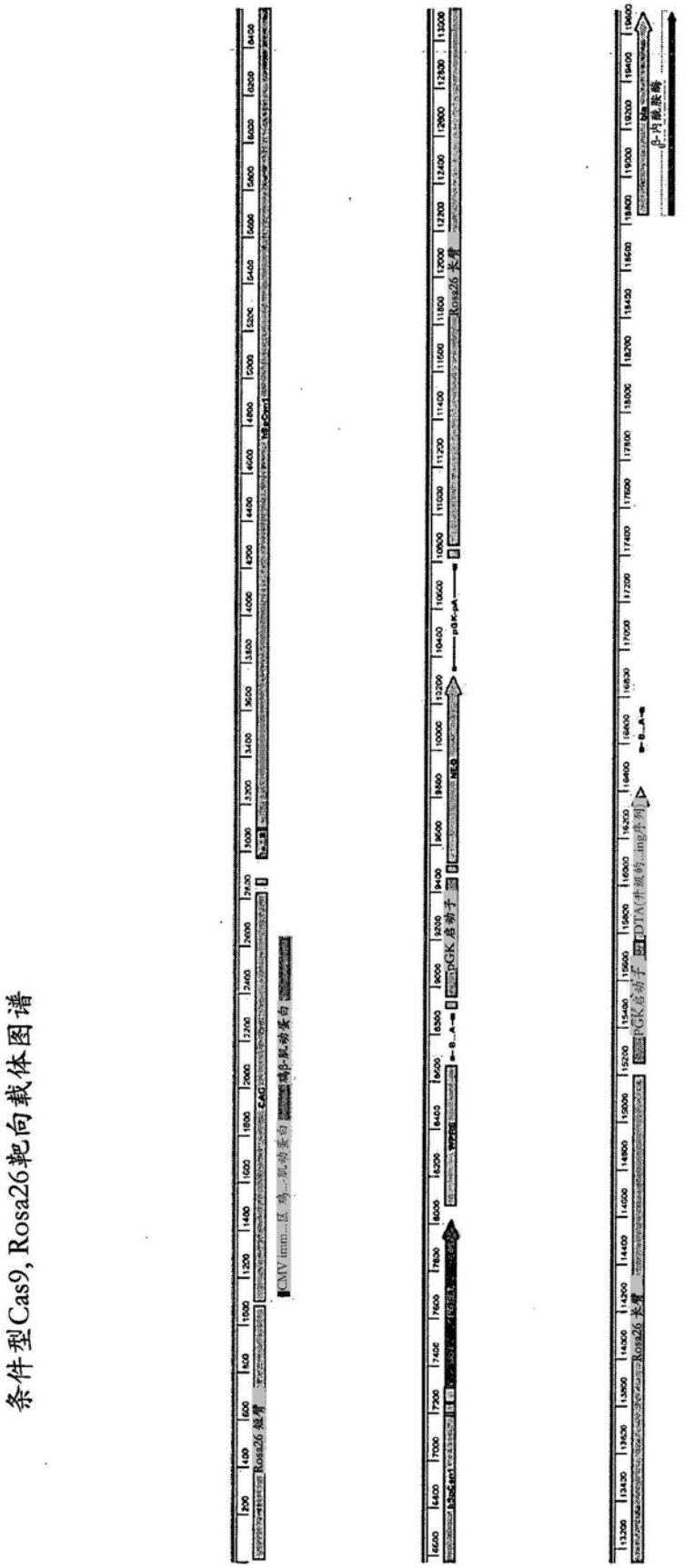

条件型Cas9, Rosa26靶向载体图谱
Rosa26 短臂
CMV imm...区 鸡...肌动蛋白
PGK 启动子
Rosa26 长臂
pGK-pA
PGK启动子
DTA(作用的...mp序列)
β-内酰胺酶

图25A

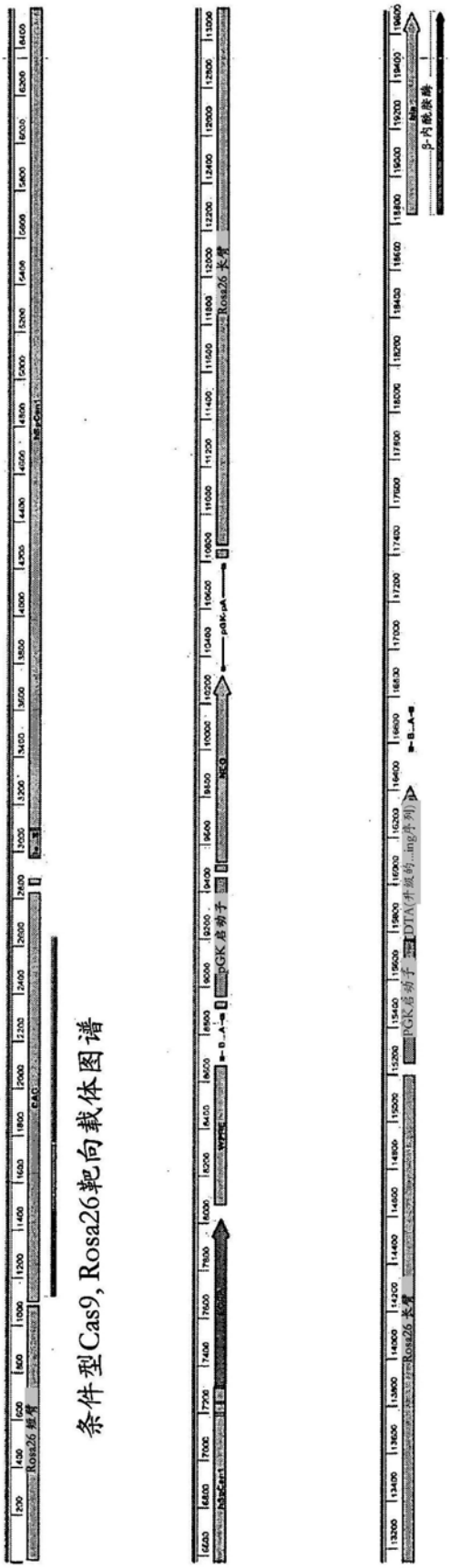


图25B

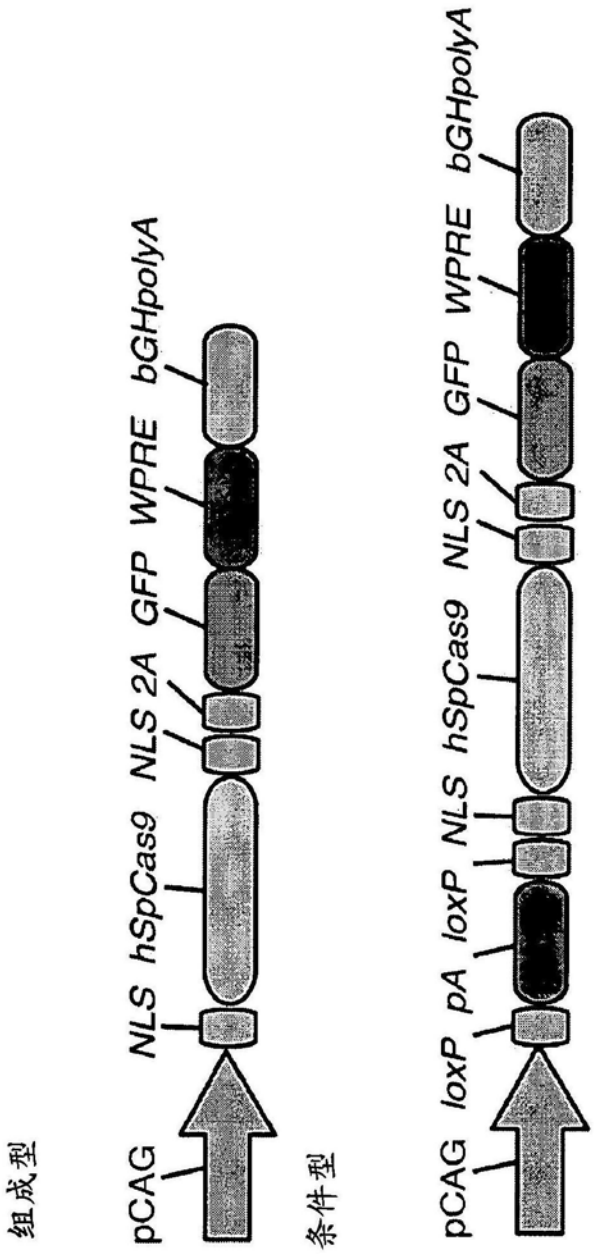
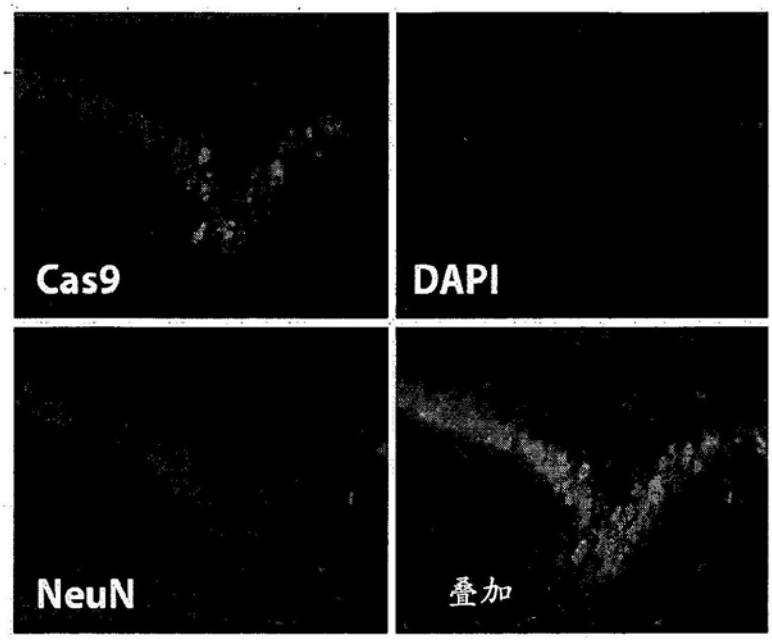

组成型
pCAG
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA
条件型
pCAG
loxP
pA
loxP
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA

图26

小鼠海马中的Cas9表达（AAV）



小鼠皮层中的Cas9表达（AAV）

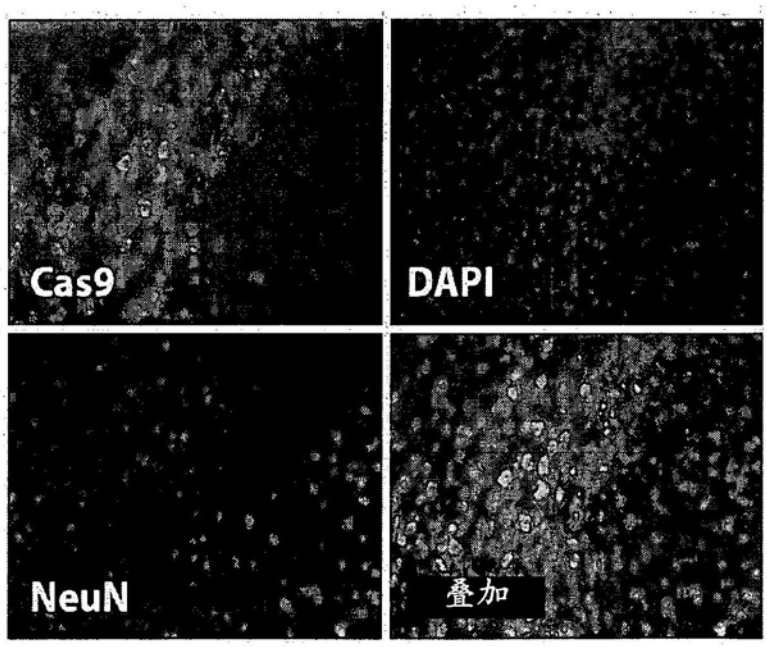


图27

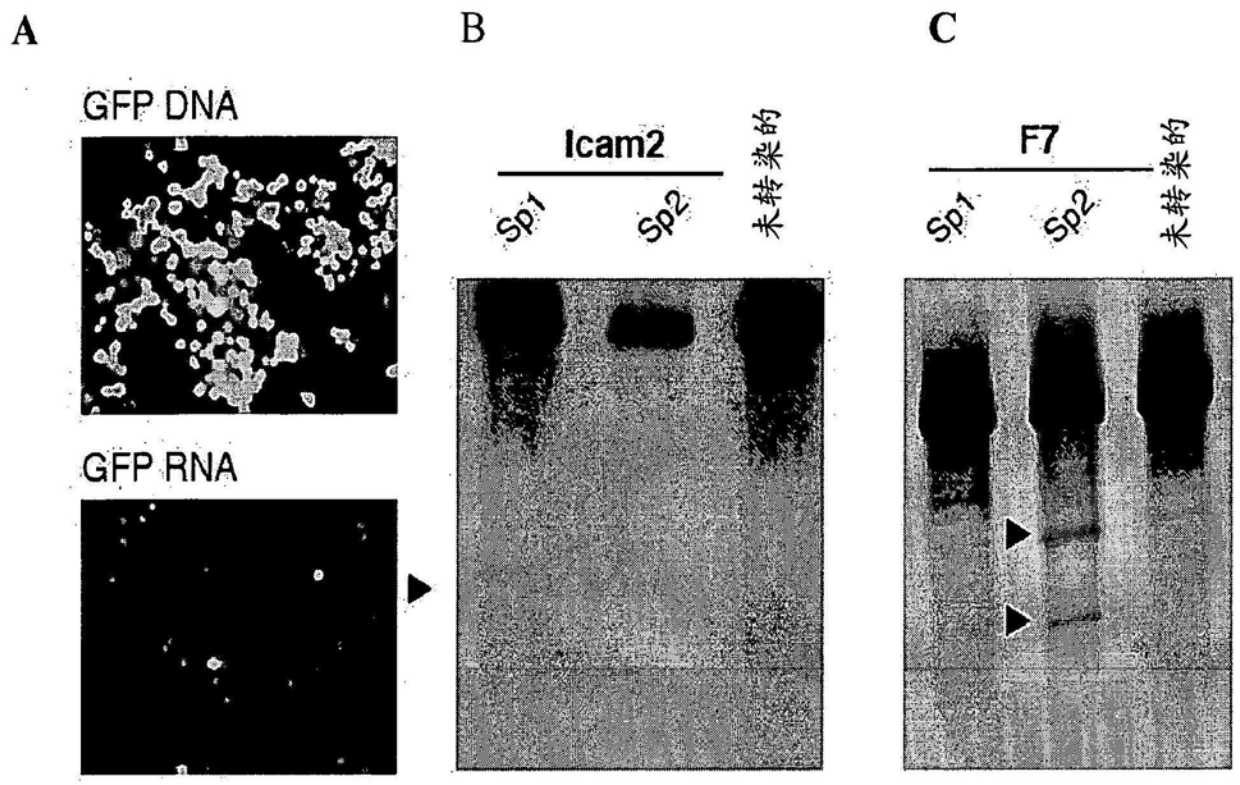


图28

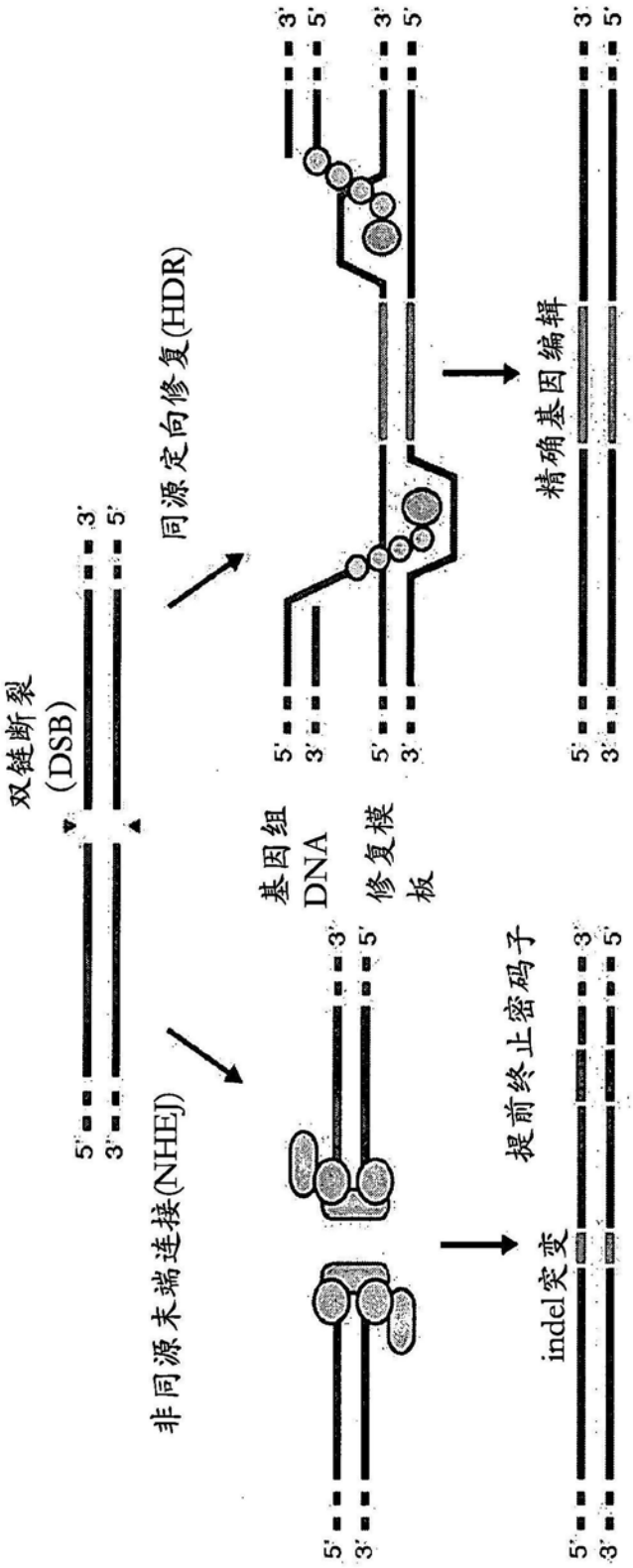

双链断裂
(DSB)
同源定向修复(HDR)
精确基因编辑
基因组
DNA
修复模板
非同源末端连接(NHEJ)
提前终止密码子
indel突变
5'
3'

图29

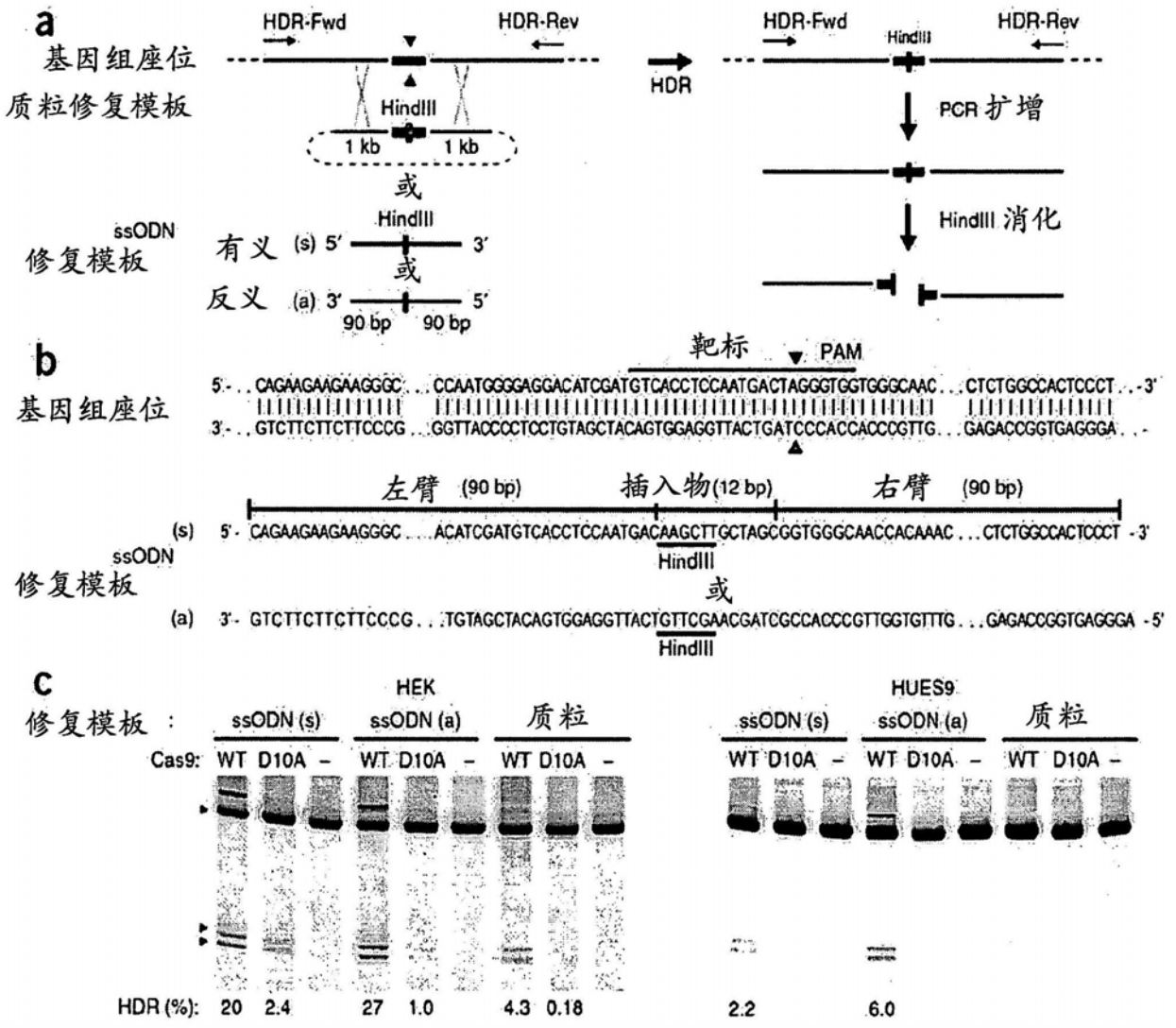

a
基因组座位
质粒修复模板
HDR-Fwd
HDR-Rev
HindIII
1 kb
1 kb
或
ssODN
修复模板
有义 (s) 5′
3′
反义 (a) 3′
5′
90 bp
90 bp
HDR
PCR 扩增
HindIII 消化
b
基因组座位
靶标
PAM
5′- . . CAGAAGAAGAAGGGC . . . CCAATGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGTGGGCAAC . . . CTCTGGCCACTCCCT . . -3′
3′- . . GTCTTCTTCTTCCCG . . GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCACCCGTTG . . . GAGACCGGTGAGGGA . . -
左臂 (90 bp)
插入物(12 bp)
右臂 (90 bp)
(s) 5′- CAGAAGAAGAAGGGC . . . ACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAAC . . . CTCTGGCCACTCCCT -3′
ssODN
修复模板
HindIII
或
(a) 3′- GTCTTCTTCTTCCCG . . . TGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGCCACCCGTTGGTGTTTG . . . GAGACCGGTGAGGGA -5′
HindIII
c
修复模板 :
HEK
HUES9
ssODN (s)
ssODN (a)
质粒
Cas9: WT D10A –
HDR (%): 20 2.4 27 1.0 4.3 0.18 2.2 6.0

图30A-C

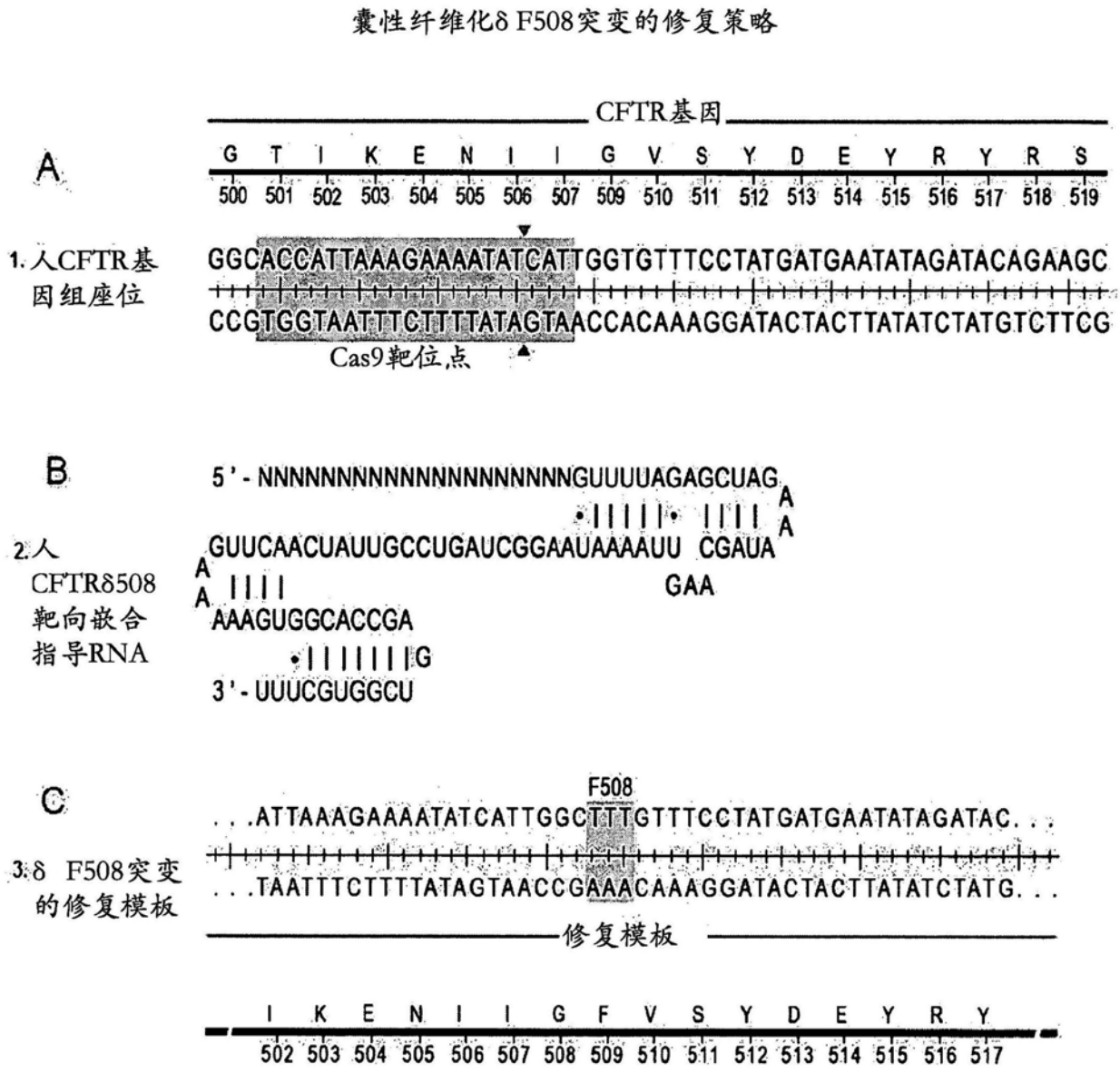

囊性纤维化δ F508突变的修复策略
CFTR基因
A
G T I K E N I I G V S Y D E Y R Y R S
500 501 502 503 504 505 506 507 509 510 511 512 513 514 515 516 517 518 519
1. 人CFTR基因组座位
GGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGATACAGAAGC
CCGTGGTAATTTCTTTTATAGTAACCACAAAGGATACTACTTATATCTATGTCTTCG
Cas9靶位点
B
5'- NNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
AAAGUGGCACCGA
3'- UUUCGUGGCU
2.人CFTRδ508靶向嵌合指导RNA
C
F508
...ATTAAAGAAAATATCATTGGCTTTGTTTCCTATGATGAATATAGATAC...
...TAATTTCTTTTATAGTAACCGAAACAAAGGATACTACTTATATCTATG...
修复模板
I K E N I I G F V S Y D E Y R Y
502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517
3. δ F508突变的修复模板

图31

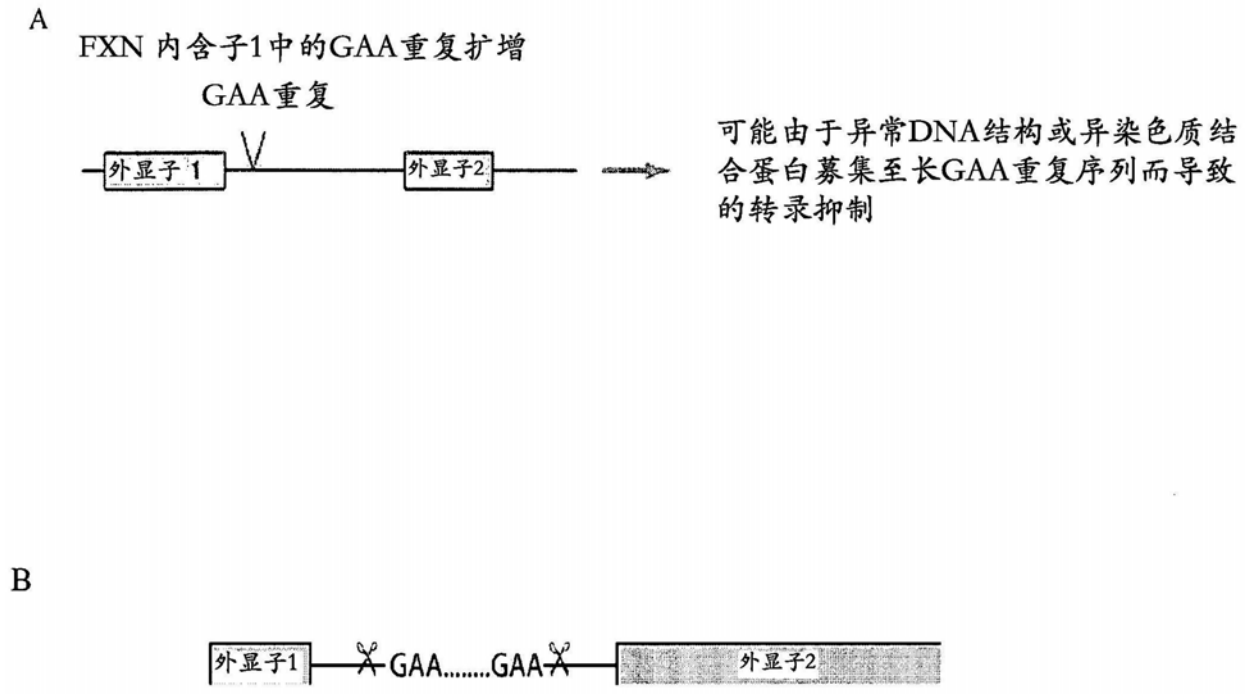

A
FXN 内含子1中的GAA重复扩增
GAA重复
外显子 1
外显子2
可能由于异常DNA结构或异染色质结合蛋白募集至长GAA重复序列而导致的转录抑制
B
外显子1
GAA.....GAA
外显子2

图32

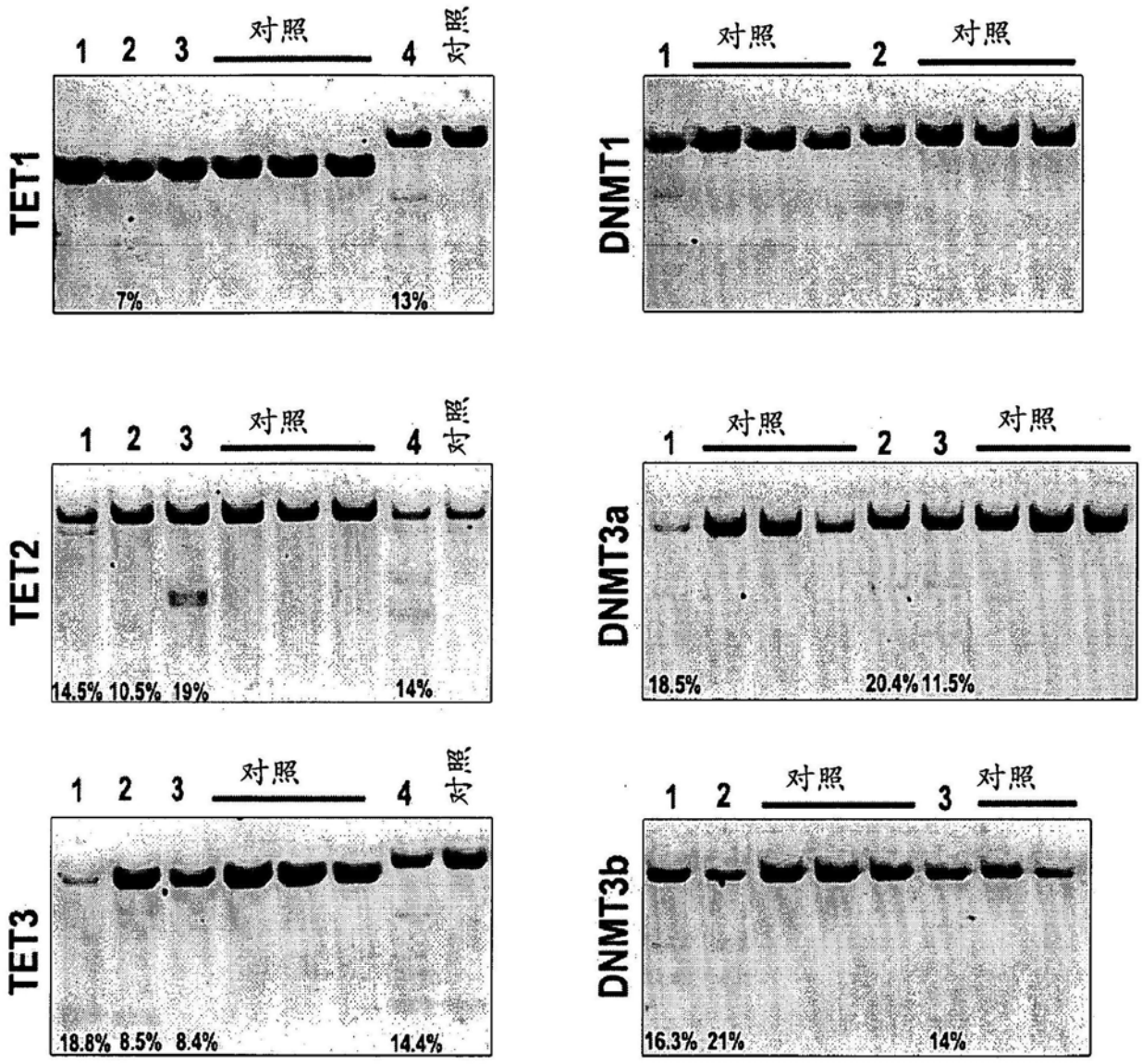

TET1
1 2 3 对照 4 对照
7%
13%
DNMT1
1 对照 2 对照
TET2
1 2 3 对照 4 对照
14.5% 10.5% 19%
14%
DNMT3a
1 对照 2 3 对照
18.5%
20.4% 11.5%
TET3
1 2 3 对照 4 对照
18.8% 8.5% 8.4%
14.4%
DNMT3b
1 2 对照 3 对照
16.3% 21%
14%

图33

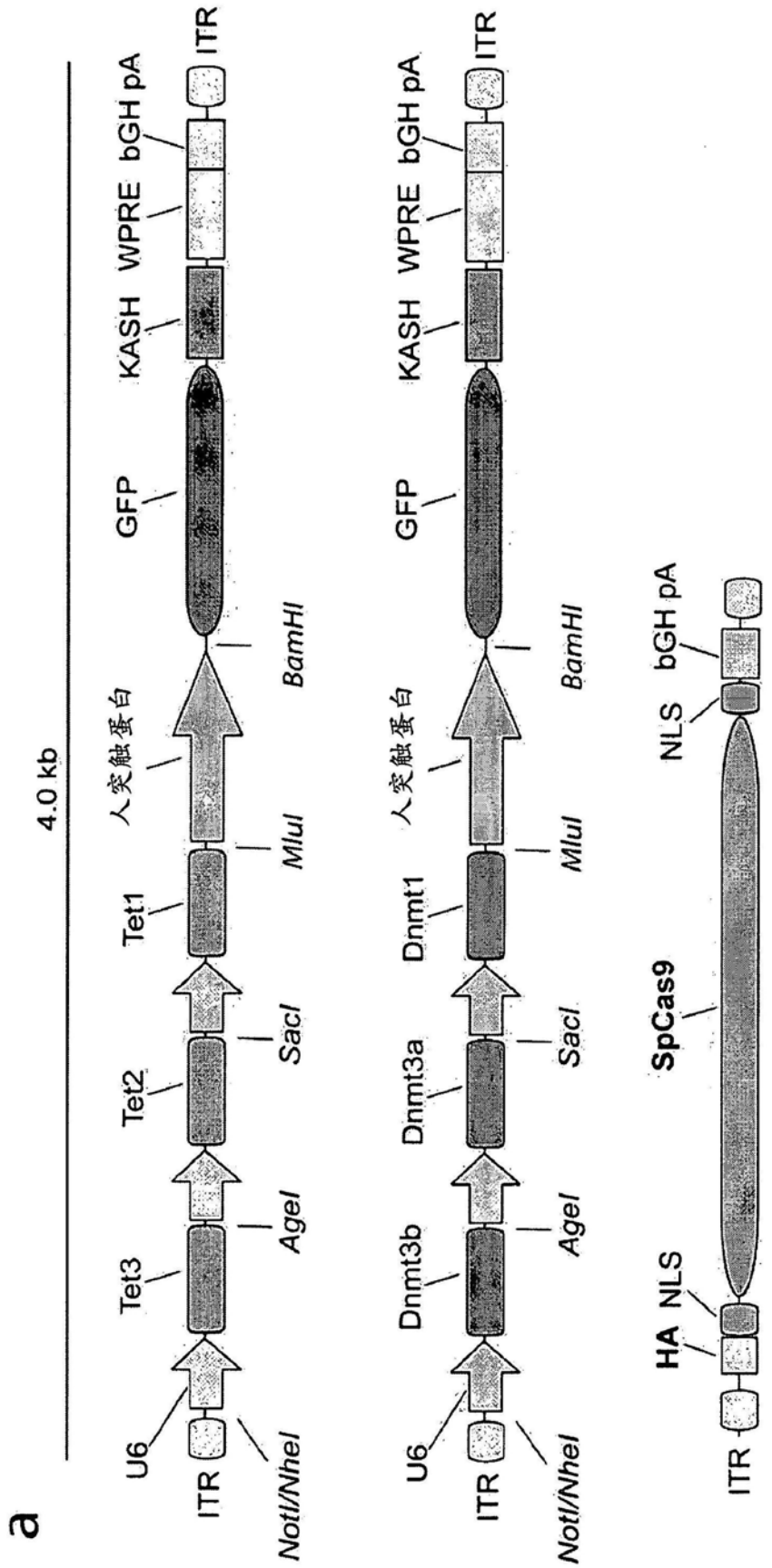

a
4.0 kb
U6
ITR
NotI/NheI
Tet3
AgeI
Tet2
SacI
Tet1
MluI
人突触蛋白
BamHI
GFP
KASH
WPRE
bGH pA
ITR
U6
ITR
NotI/NheI
Dnmt3b
AgeI
Dnmt3a
SacI
Dnmt1
MluI
人突触蛋白
BamHI
GFP
KASH
WPRE
bGH pA
ITR
HA
NLS
SpCas9
NLS
bGH pA
ITR

图34

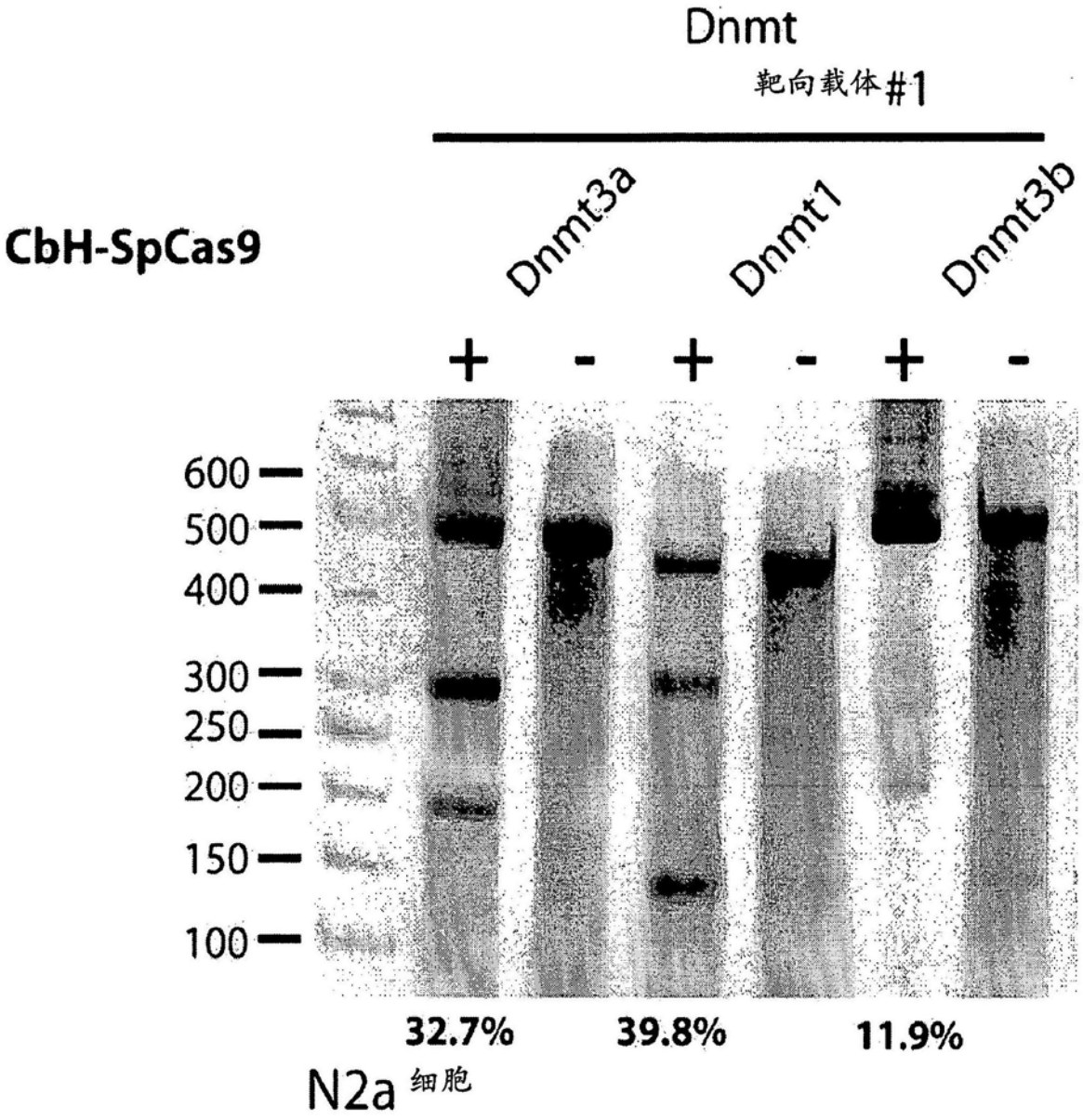

Dnmt
靶向载体#1
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+ - + - + -
600
500
400
300
250
200
150
100
32.7%
39.8%
11.9%
N2a 细胞

图35

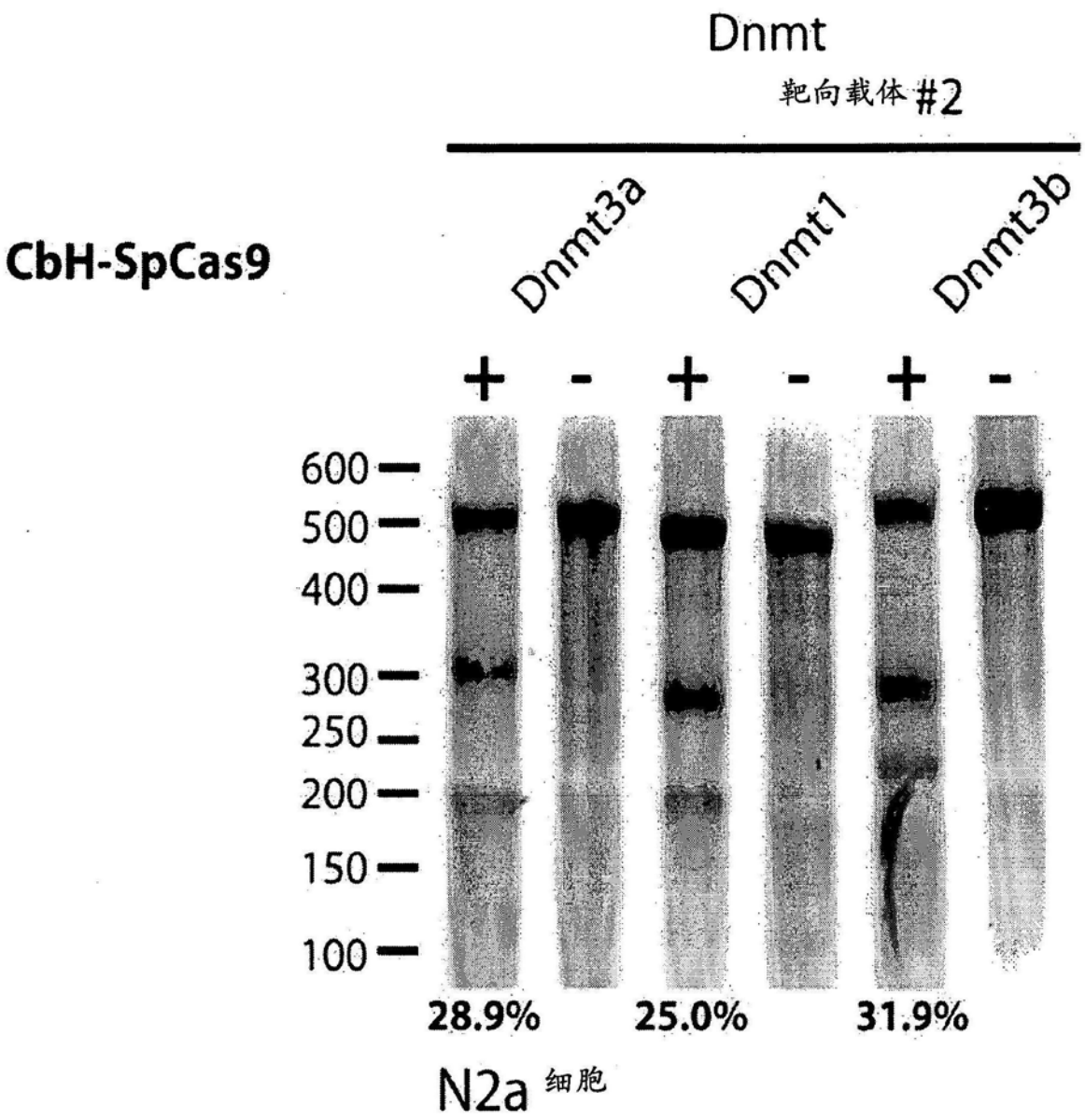

Dnmt
靶向载体 #2
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+ - + - + -
600
500
400
300
250
200
150
100
28.9%
25.0%
31.9%
N2a 细胞

图36

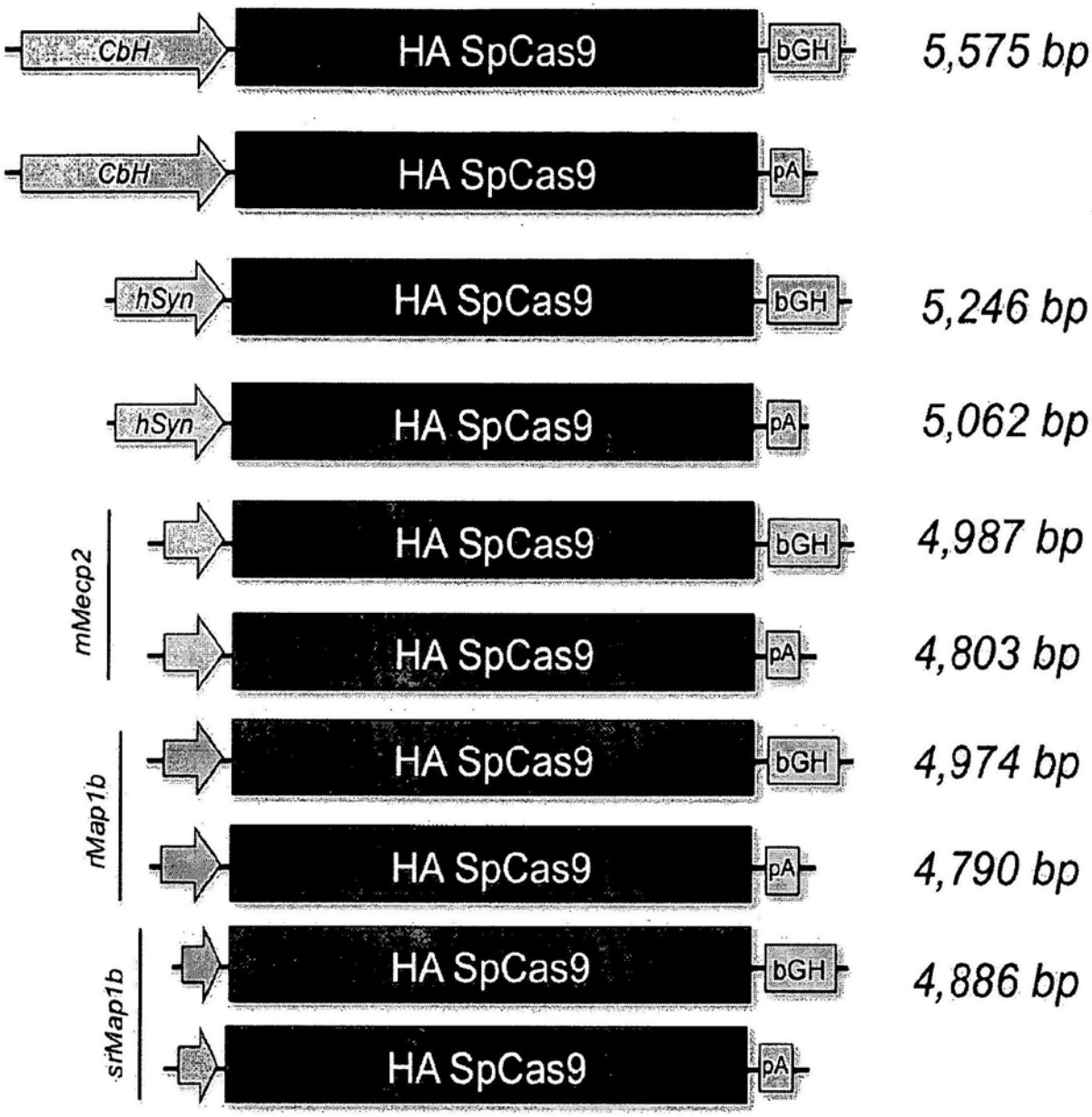

CbH
HA SpCas9
bGH
5,575 bp
CbH
HA SpCas9
pA
hSyn
HA SpCas9
bGH
5,246 bp
hSyn
HA SpCas9
pA
5,062 bp
mMecp2
HA SpCas9
bGH
4,987 bp
HA SpCas9
pA
4,803 bp
rMap1b
HA SpCas9
bGH
4,974 bp
HA SpCas9
pA
4,790 bp
srMap1b
HA SpCas9
bGH
4,886 bp
HA SpCas9
pA

图37

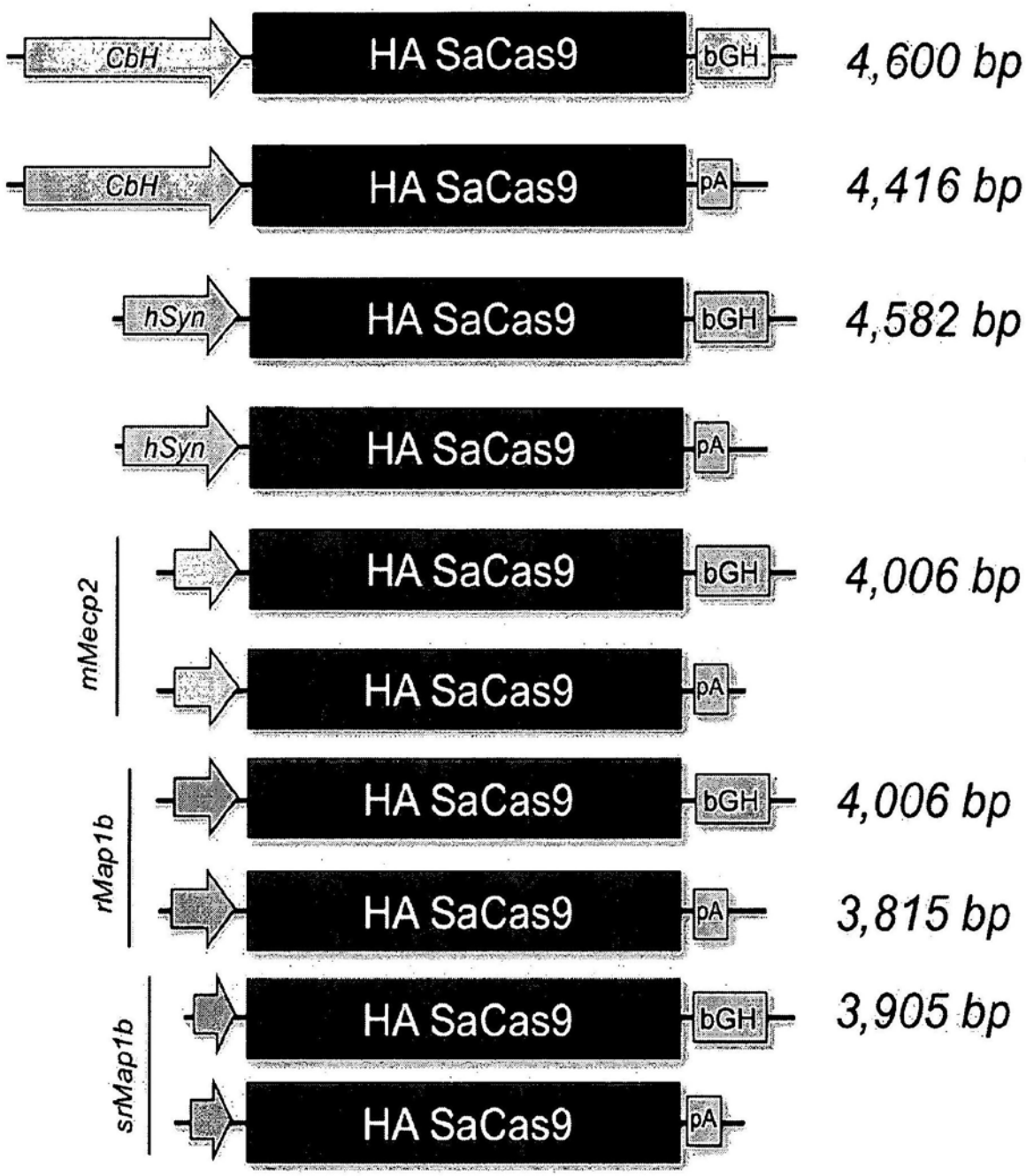

CbH
HA SaCas9
bGH
4,600 bp
CbH
HA SaCas9
pA
4,416 bp
hSyn
HA SaCas9
bGH
4,582 bp
hSyn
HA SaCas9
pA
mMecp2
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
rMap1b
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
3,815 bp
srMap1b
HA SaCas9
bGH
3,905 bp
HA SaCas9
pA

图38

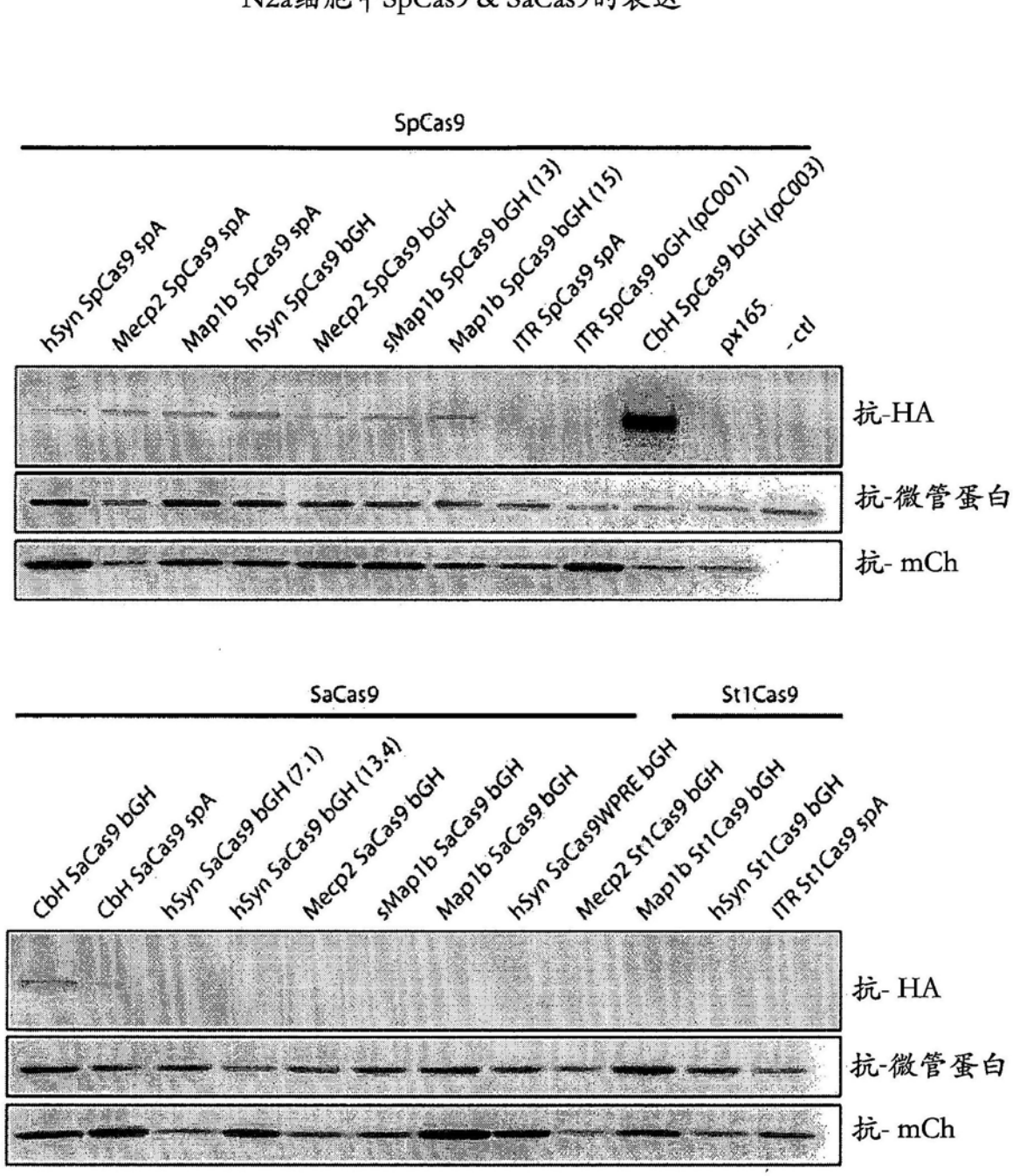

N2a细胞中SpCas9 & SaCas9的表达
SpCas9
hSyn SpCas9 spA
Mecp2 SpCas9 spA
Map1b SpCas9 spA
hSyn SpCas9 bGH
Mecp2 SpCas9 bGH
sMap1b SpCas9 bGH (13)
Map1b SpCas9 bGH (15)
ITR SpCas9 spA
ITR SpCas9 bGH (pC001)
CbH SpCas9 bGH (pC003)
px165
ctl
抗-HA
抗-微管蛋白
抗- mCh
SaCas9
St1Cas9
CbH SaCas9 bGH
CbH SaCas9 spA
hSyn SaCas9 bGH (7.1)
hSyn SaCas9 bGH (13.4)
Mecp2 SaCas9 bGH
sMap1b SaCas9 bGH
Map1b SaCas9 bGH
hSyn SaCas9WPRE bGH
Mecp2 St1Cas9 bGH
Map1b St1Cas9 bGH
hSyn St1Cas9 bGH
ITR St1Cas9 spA
抗- HA
抗-微管蛋白
抗- mCh

图39

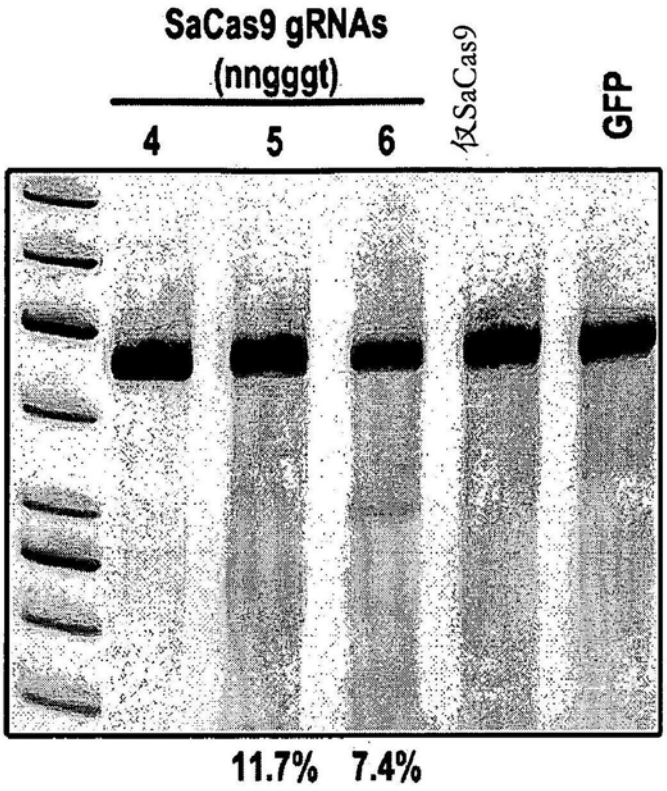


图40

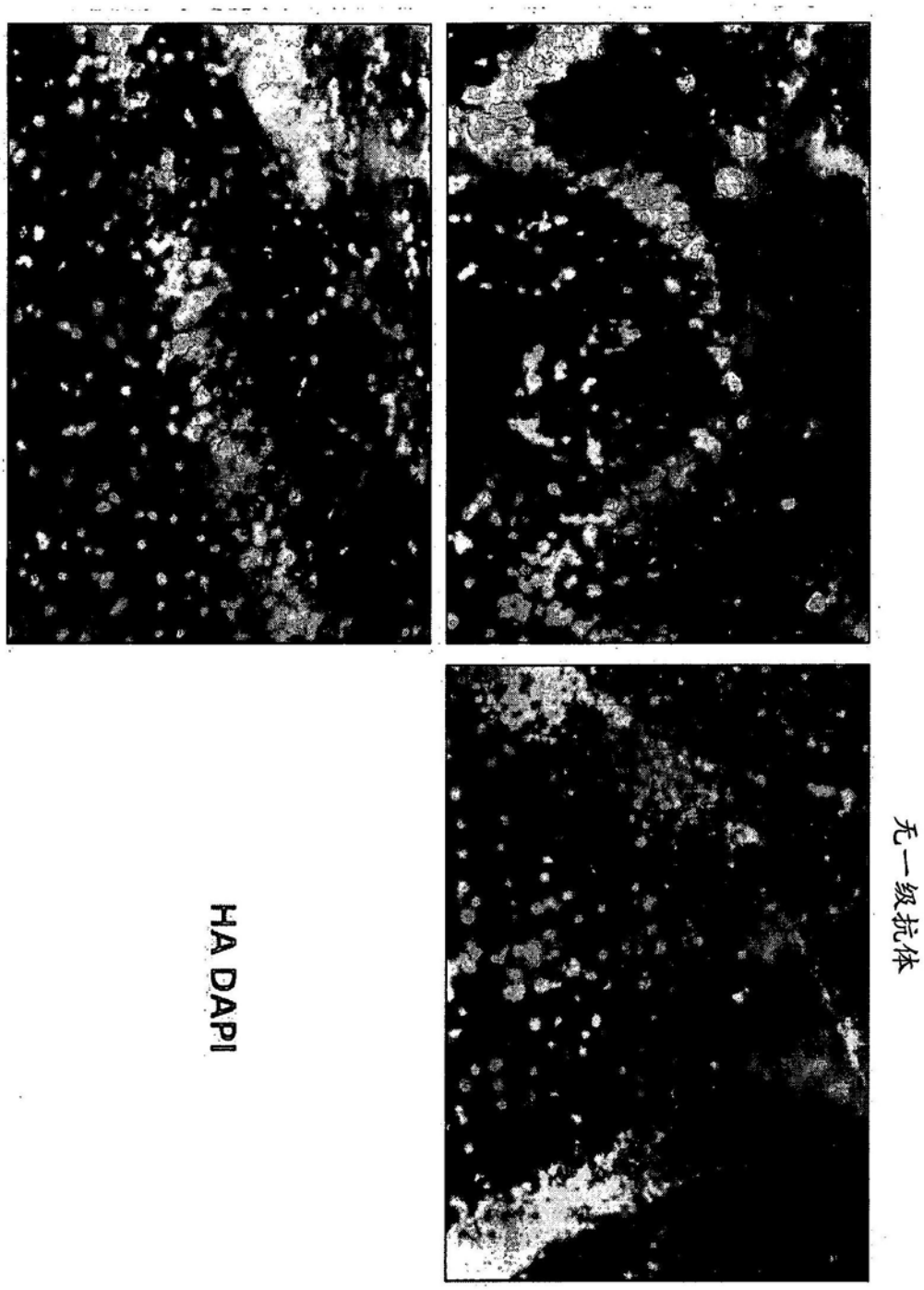


图41

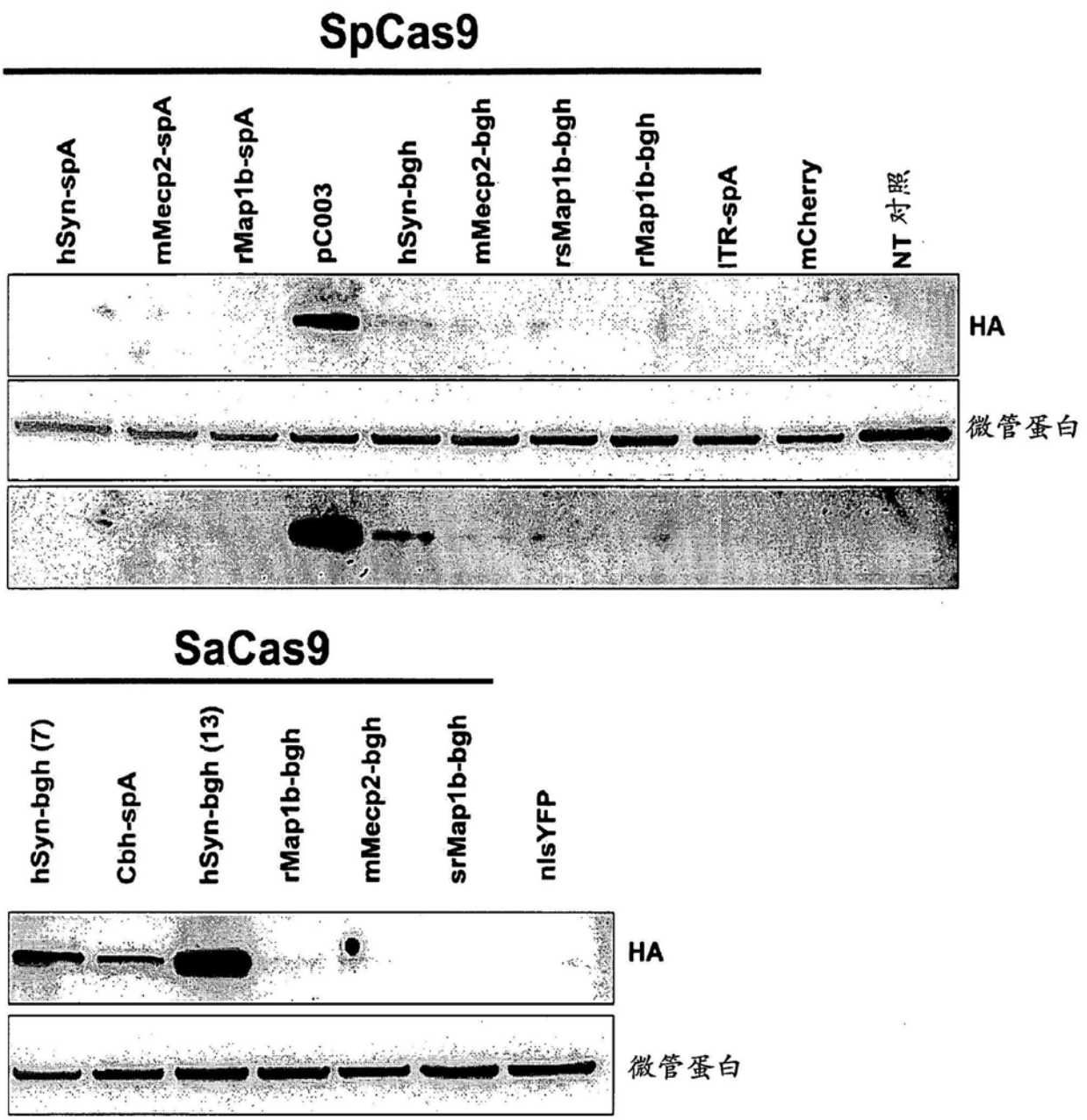

SpCas9
hSyn-spA
mMecp2-spA
rMap1b-spA
pC003
hSyn-bgh
mMecp2-bgh
rsMap1b-bgh
rMap1b-bgh
ITR-spA
mCherry
NT 对照
HA
微管蛋白
SaCas9
hSyn-bgh (7)
Cbh-spA
hSyn-bgh (13)
rMap1b-bgh
mMecp2-bgh
srMap1b-bgh
nlsYFP
HA
微管蛋白

图42

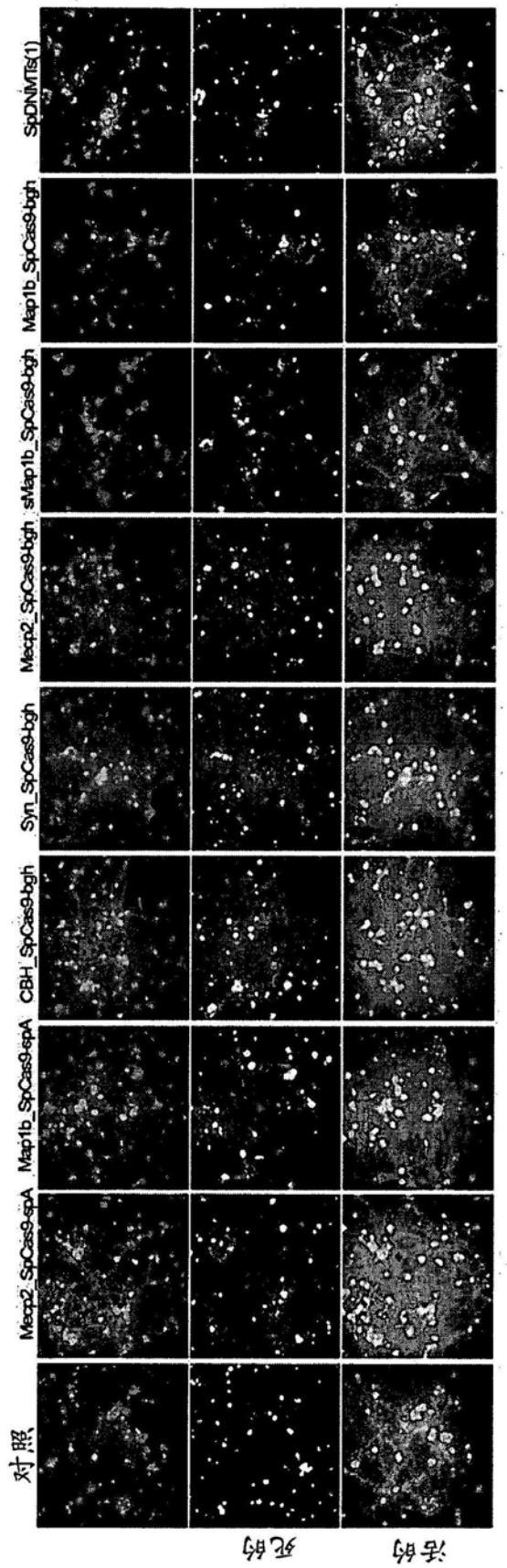

对照
Mecp2_SpCas9-spA
Map1b_SpCas9-spA
CBH_SpCas9-bgh
Syn_SpCas9-bgh
Mecp2_SpCas9-bgh
sMap1b_SpCas9-bgh
Map1b_SpCas9-bgh
SpDNMT3(1)
死的
活的

图43

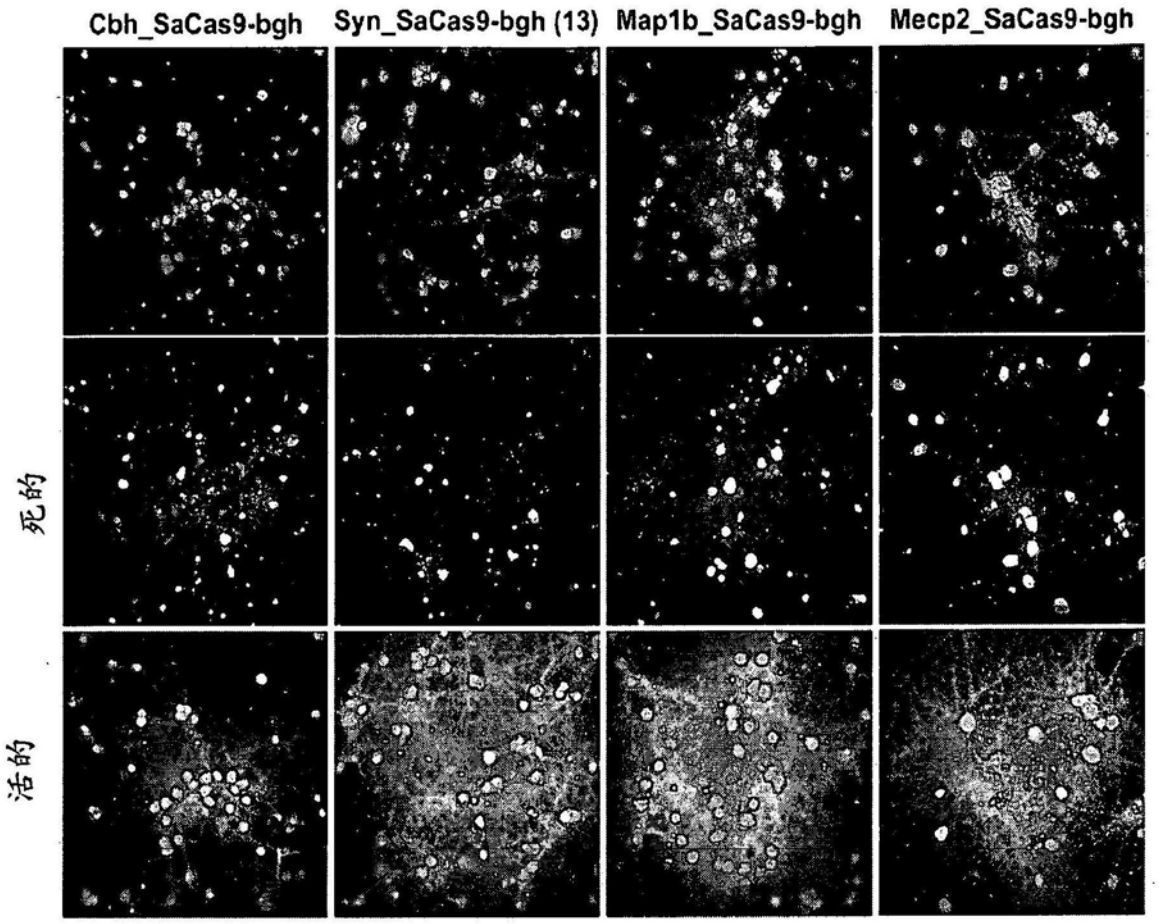

Cbh_SaCas9-bgh
Syn_SaCas9-bgh (13)
Map1b_SaCas9-bgh
Mecp2_SaCas9-bgh
死的
活的

图44

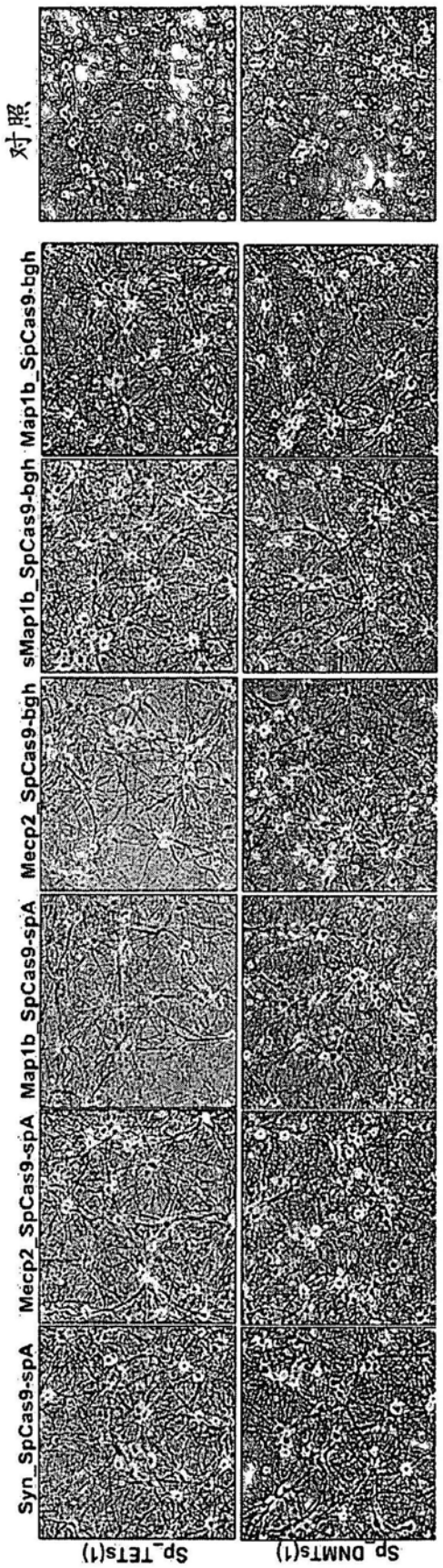

对照
Map1b_SpCas9-bgh
sMap1b_SpCas9-bgh
Mecp2_SpCas9-bgh
Map1b_SpCas9-spA
Mecp2_SpCas9-spA
Syn_SpCas9-spA
Sp_TETs(1)
Sp_DNMTs(1)

图45

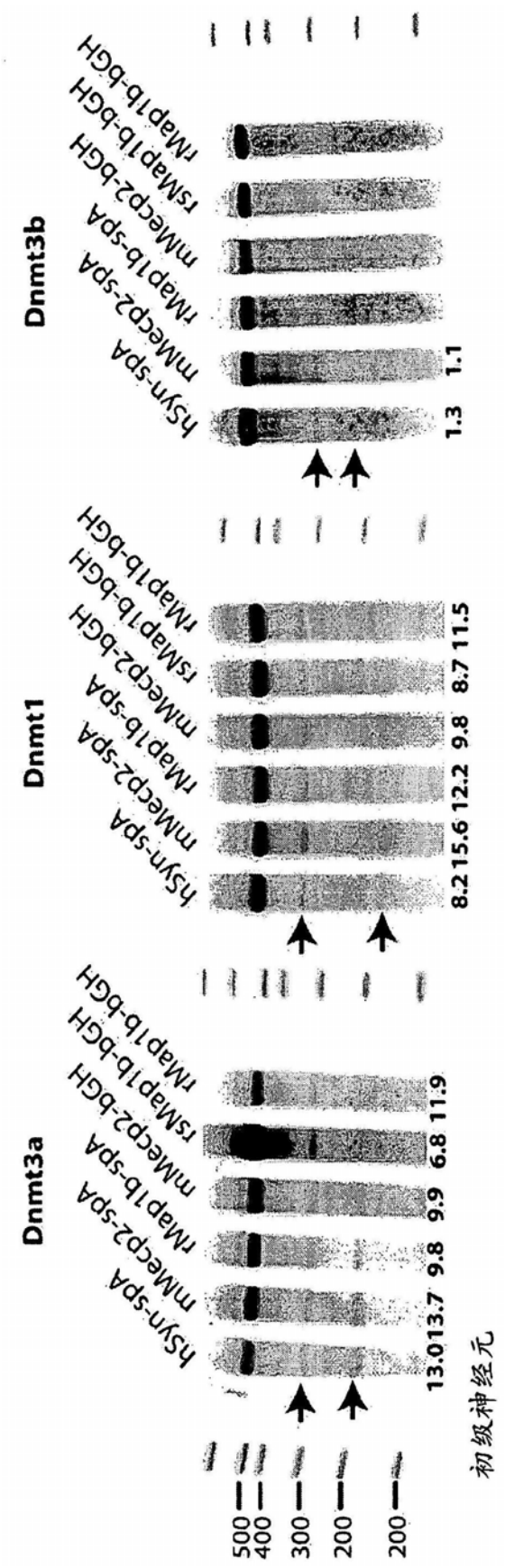

Dnmt3b
rMap1b-bGH
rSMap1b-bGH
rSMecp2-bGH
mMap1b-spA
rMecp2-spA
mMecp2-spA
hSyn-spA
1.1
1.3
Dnmt1
rMap1b-bGH
rSMap1b-bGH
rSMecp2-bGH
mMap1b-spA
rMecp2-spA
mMecp2-spA
hSyn-spA
11.5
8.7
9.8
12.2
15.6
8.2
Dnmt3a
rMap1b-bGH
rSMap1b-bGH
rSMecp2-bGH
mMap1b-spA
rMecp2-spA
mMecp2-spA
hSyn-spA
11.9
6.8
9.9
9.8
13.7
13.0
500
400
300
200
200
初级神经元

图46

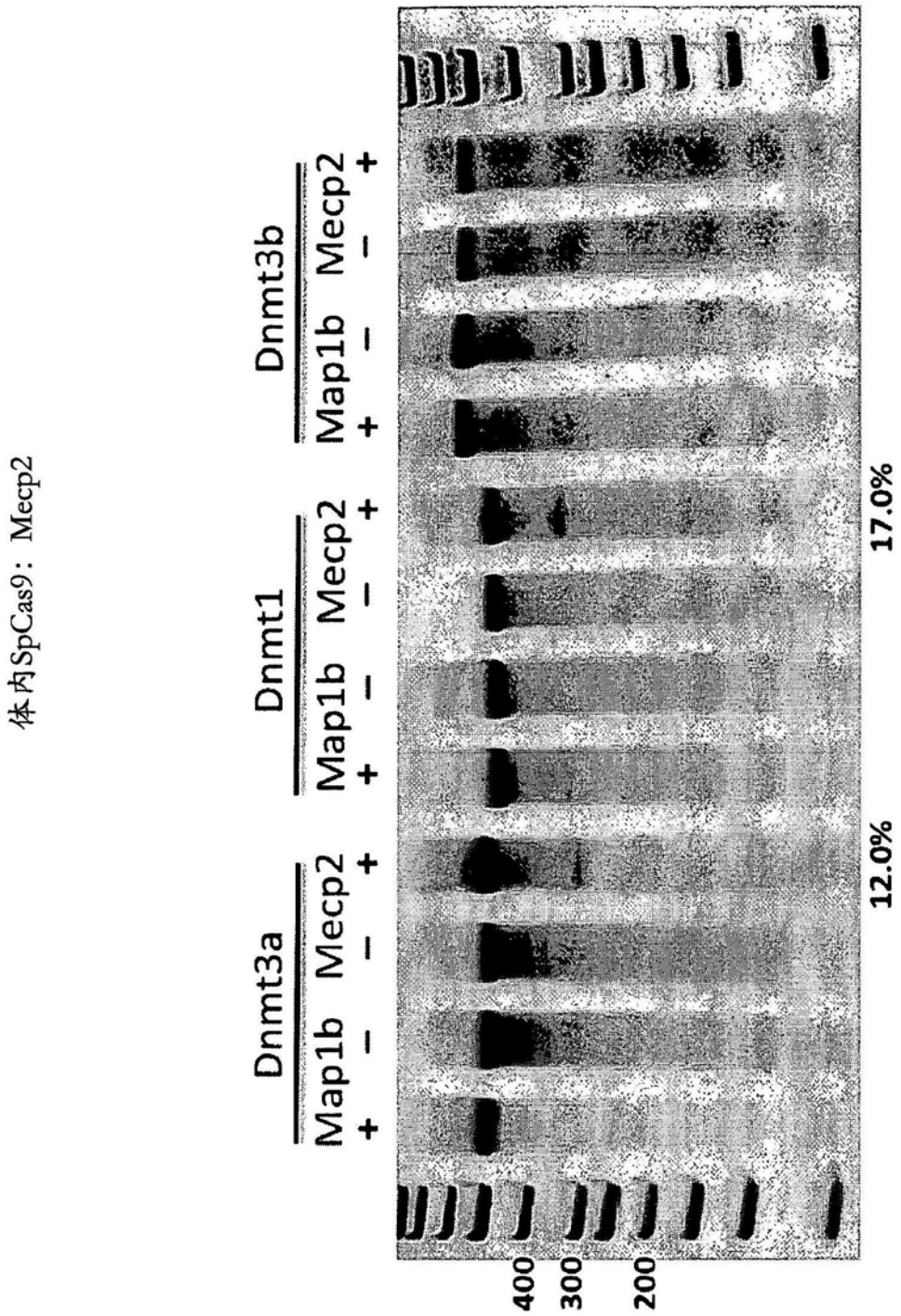

体内SpCas9：Mecp2
Dnmt3a
Map1b Mecp2
+ − +
Dnmt1
Map1b Mecp2
+ − +
Dnmt3b
Map1b Mecp2
+ − +
400
300
200
12.0%
17.0%

图47

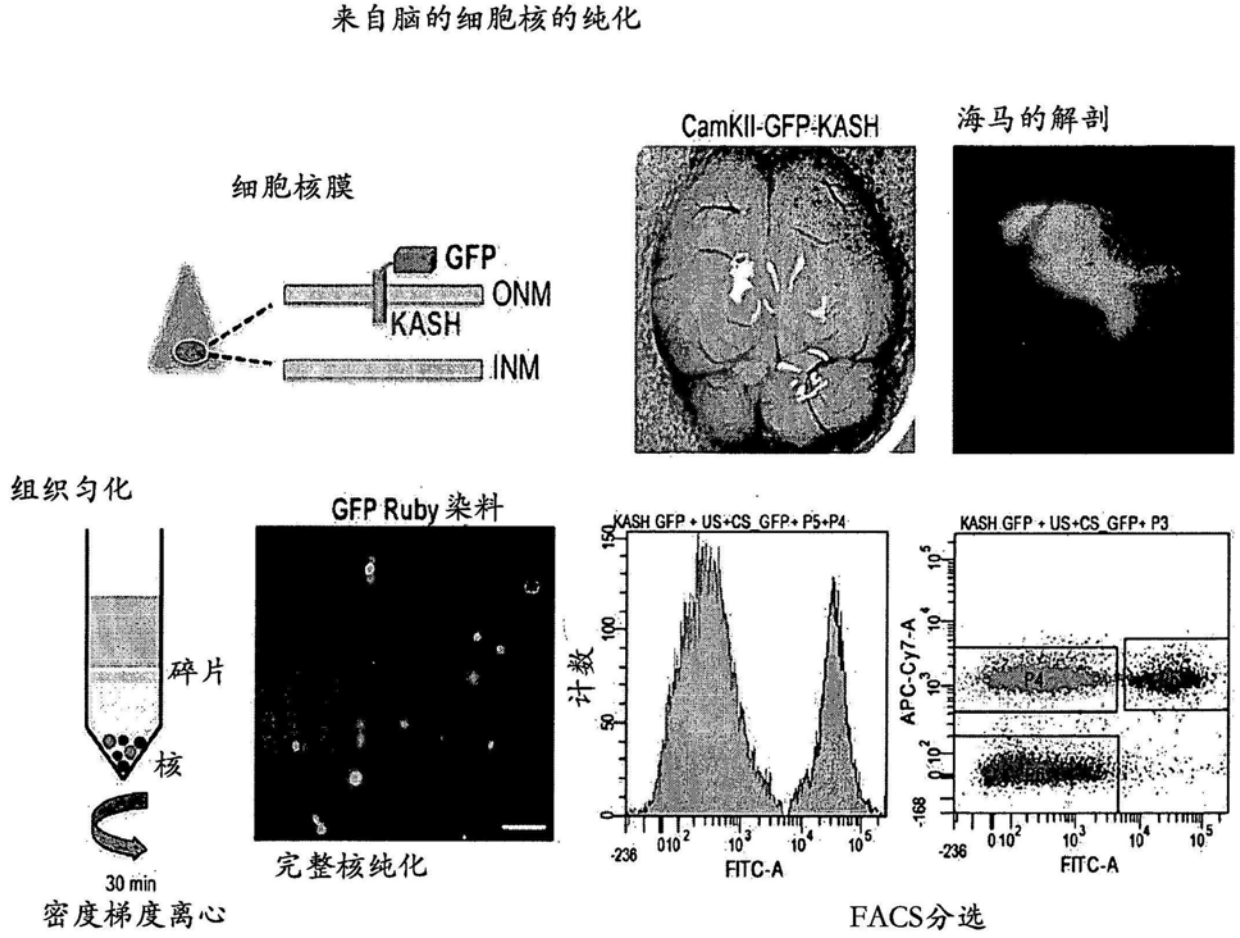

来自脑的细胞核的纯化
细胞核膜
GFP
ONM
KASH
INM
CamKII-GFP-KASH
海马的解剖
组织匀化
碎片
核
30 min
密度梯度离心
GFP Ruby 染料
完整核纯化
KASH GFP + US+CS_GFP+ P5+P4
计数
FITC-A
KASH GFP + US+CS_GFP+ P3
APC-Cy7-A
P4
FITC-A
FACS分选

图48

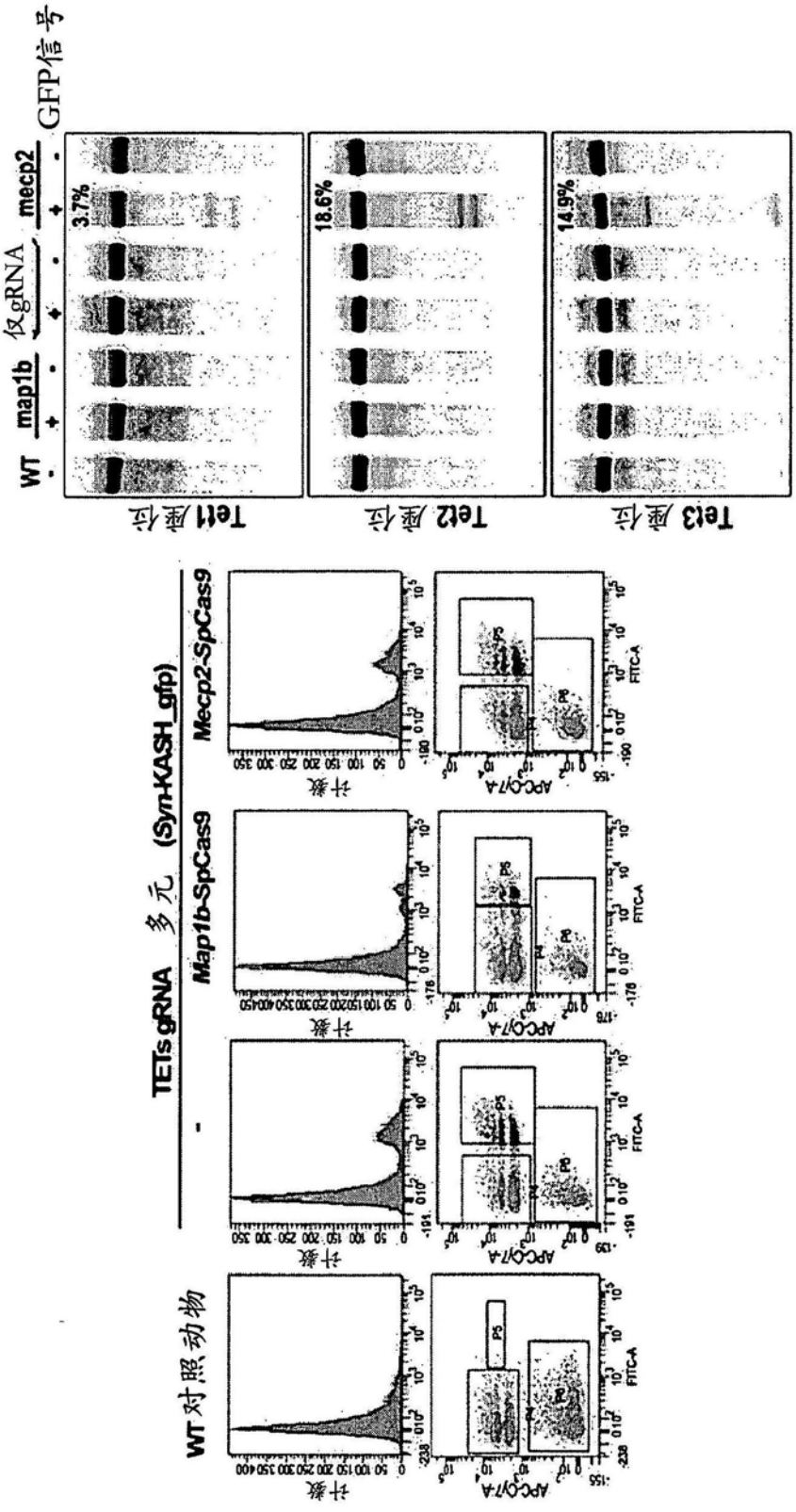

GFP信号
WT
map1b
仅gRNA
mecp2
3.7%
18.6%
14.9%
Tet1 座位
Tet2 座位
Tet3 座位
TETs gRNA 多元 (SynKASH_gfp)
-
Map1b-SpCas9
Mecp2-SpCas9
WT 对照动物
计数
APC-Cy7-A
FITC-A
P5
P6

图49

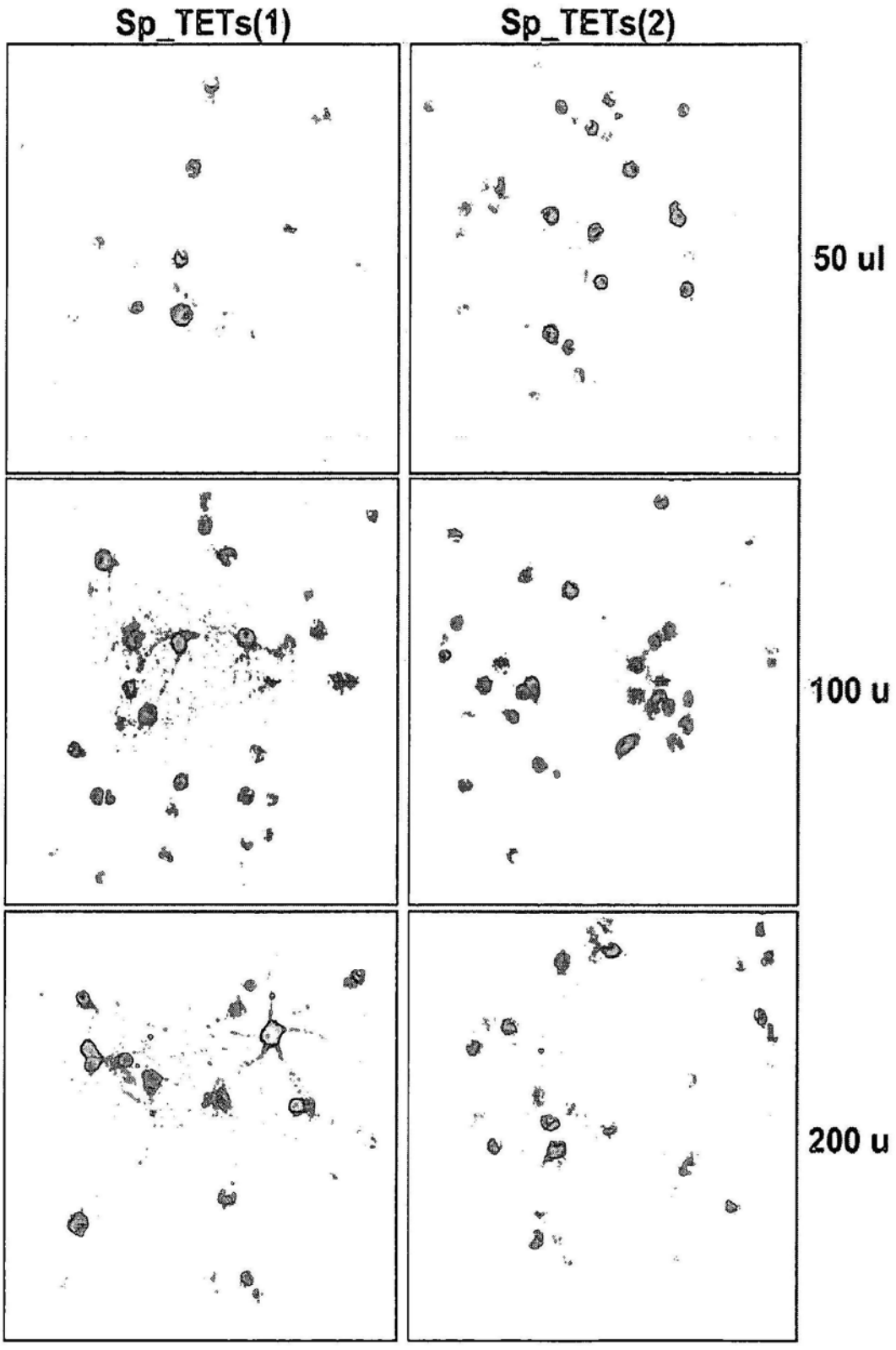


图50

	模式	突出端
1	CCnNNNNNNNNNNNNNNNnnnnnnnnnnnnNNNNNNNNNNNNNNNNnGG	38
2	CCnNNNNNNNNNNNNNNNNnnnnnnnnNNNNNNNNNNNNNNNNnGG	34
3	CCnNNNNNNNNNNNNNNNNnnnnNNNNNNNNNNNNNNNNnGG	30
4	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	26
5	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	22
6	CCnNNNNNNNNNNNNNNNNNNNNNNNNnGG	18
7	CCnNNNNNNNNNNNNNNNNNNNNnGG	14
8	CCnNNNNNNNNNNNNNNNNnGG	10
9	CCnNNNNNNNNNNNNnGG	6
10	CCnNNNNNNNNNNnGG	4
11	CCnNNNNNNNNNnGG	3
12	CCnNNNNNNNNnGG	2
13	CCnNNNNNNNnGG	1
14	CCnNNNNNNnGG	0
15	CCnNNNNNnGG	1
16	CCnNNNNnGG	2
17	CCnNNNnGG	3
18	CCnnGG	4
19	CCGG	8
20	GGCC	10
21	GGNNCC	12
22	GGNNNCC	14
23	GGNNNNCC	16
24	GGNNNNNNNCC	20

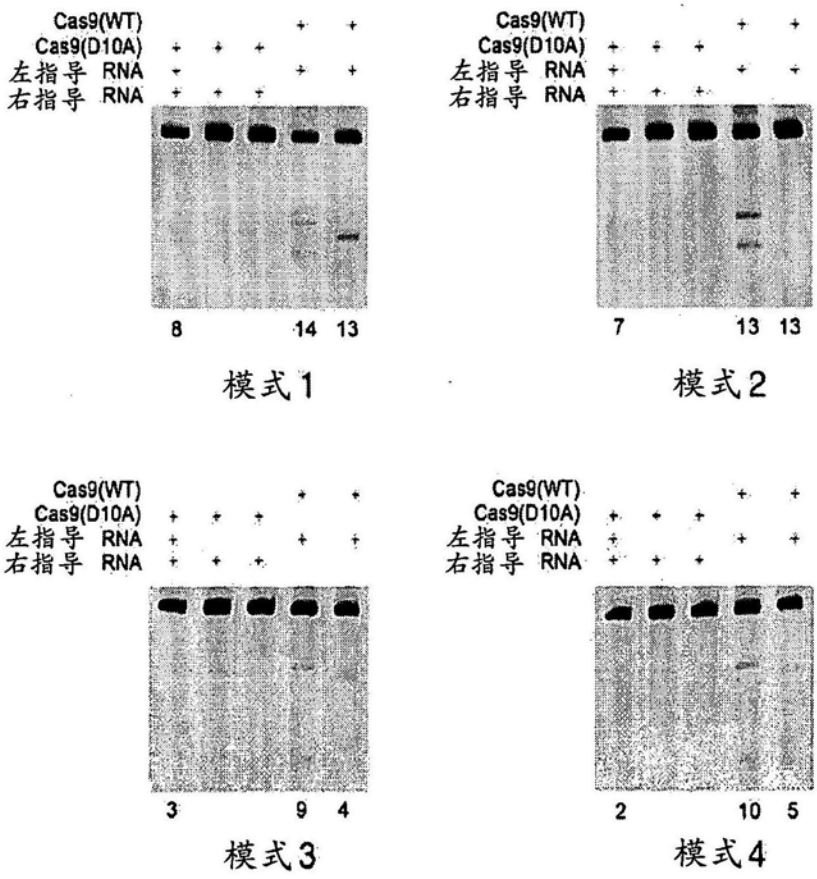


图51

用于产生U6-指导RNA表达盒的反向PCR引物（与U6正向引物配对）

Z0260	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACATCAACCGGTGGCGCAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 1
Z0261	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACATCAACCGGTGGCGCAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 2
Z0262	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 3
Z0263	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 4
Z0264	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTGGCGCATTGCCACGAAGCGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 5
Z0265	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCGAGGGCAGAGTGCTGCTTGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 6
Z0266	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGTCCGAGCAGAAGAAGAAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 7
Z0267	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 8
Z0268	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGCAGAAGAAGAAGGGCTCCGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 9
Z0269	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACATCAACCGGTGGCGCAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 10
Z0270	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCTGGCCCAGGTGAAGGTGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 11
Z0271	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCCCTCCCTGGCCCAGGTGAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 12
Z0272	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAACCGGAGGACAAAGTACAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 13
Z0273	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTGAAGGTGTGGTTCCAGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 14
Z0274	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGTGAAGGTGTGGTTCCAGAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 15
Z0275	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAACCGGAGGACAAAGTACAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 16
Z0276	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCTGGCCCAGGTGAAGGTGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 17
Z0277	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTGAAGGTGTGGTTCCAGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 18
Z0278	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 19
Z0279	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGGAGGGAGGGGGCACAGATGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 20
Z0280	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCACCTTCACCTGGGCCAGGGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 21
Z0281	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACACCCTAGTCATTGGAGGTGAGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 22
Z0282	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCAGAGCAGCCACTGGGGCCTGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 23
Z0283	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCACCTTCACCTGGGCCAGGGGGTGTTTCGTCCTTTCCACAAG	左指导 RNA 24
Z0284	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCCATTGGCCTGCTTCGTGGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 1
Z0285	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACATTGGCCTGCTTCGTGGCAAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 2
Z0286	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCCTCCTCCAGCTTCTGCCGGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 3
Z0287	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCTCCAGCTTCTGCCGTTTGGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 4
Z0288	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACATTGGCCTGCTTCGTGGCAAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 5
Z0289	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGCAGCAAGCAGCACTCTGCCGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 6
Z0290	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTTCTTCTTCTGCTCGGACTCGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 7
Z0291	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACACCGGAGGACAAAGTACAAAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 8
Z0292	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCTTCTTCTGCTCGGACTCAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 9
Z0293	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTTGATGTGATGGGAGCCCTGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 10
Z0294	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGGCCAGGGAGGGAGGGGCAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 11
Z0295	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGGAGGGAGGGGCACAGATGGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 12
Z0296	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCGGTTCTGGAACCACACCTGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 13
Z0297	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACCTGGGCCAGGGAGGGAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 14
Z0298	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACTCACCTGGGCCAGGGAGGGAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 15
Z0299	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTTCTGGAACCACACCTTCAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 16
Z0300	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGGAGGGAGGGGCACAGATGGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 17
Z0301	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGGGCCAGGGAGGGAGGGGCAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 18
Z0302	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGTTCTGGAACCACACCTTCAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 19
Z0303	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACAGGTGAAGGTGTGGTTCCAGGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 20
Z0304	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAACCGGAGGACAAAGTACAGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 21
Z0305	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCAAACCCACGAGGGCAGAGTGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 22
Z0306	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGTTTCTCATCTGTGCCCCGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 23
Z0307	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACGAGGACAAAGTACAAACGGCGGTGTTTCGTCCTTTCCACAAG	右指导 RNA 24

图52

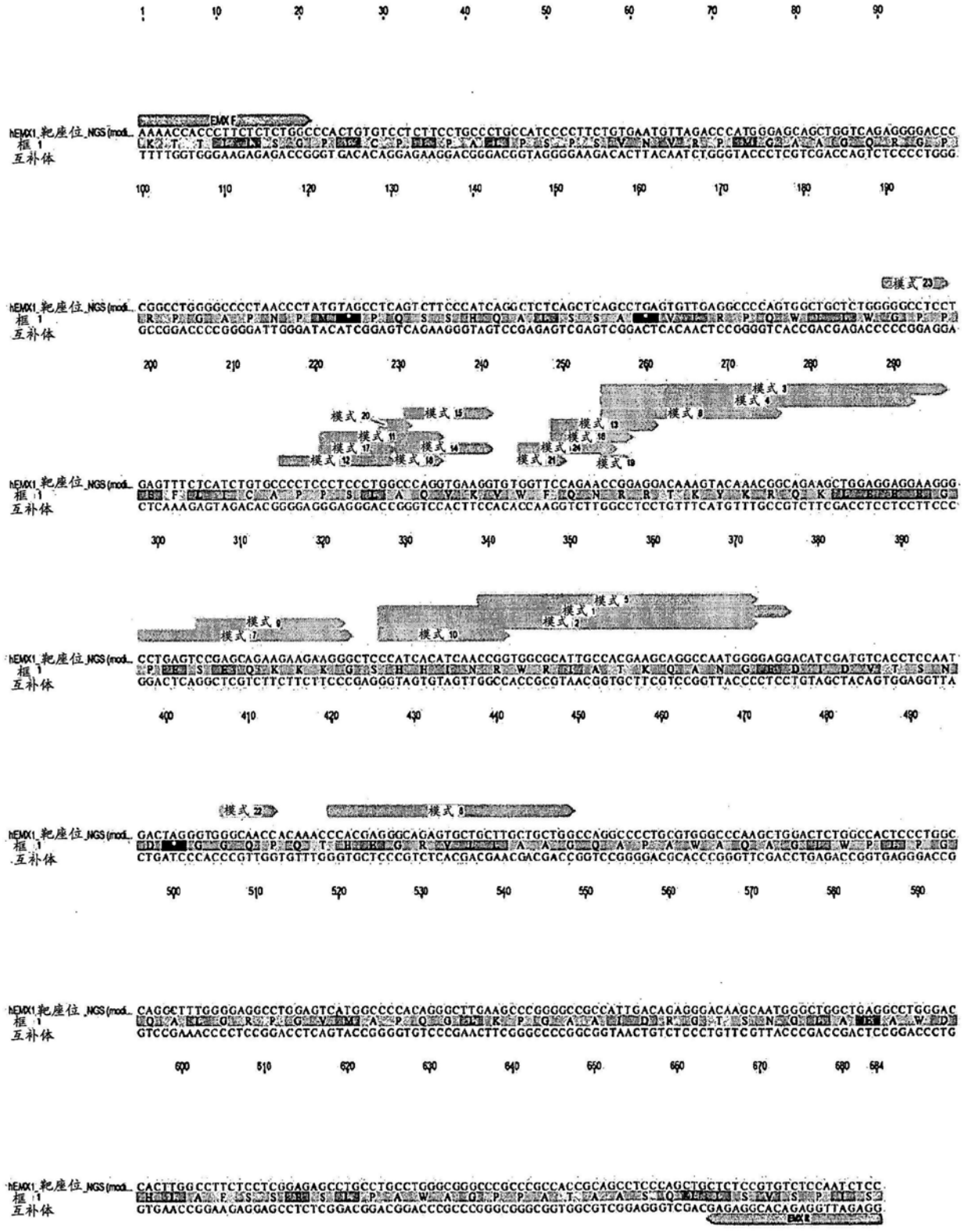
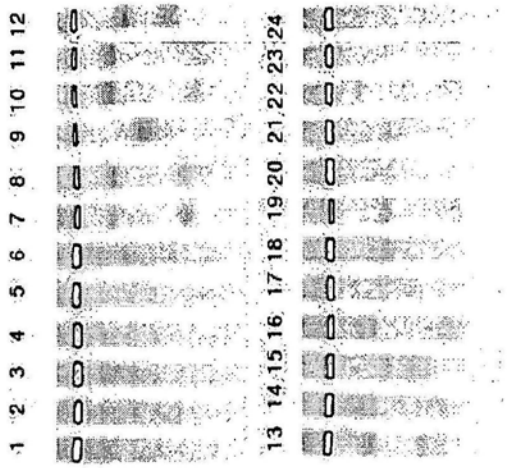

hEMX1_靶座位_NGS (modi...
框 1
互补体
EMX F
模式 23
模式 3
模式 4
模式 8
模式 20
模式 15
模式 11
模式 17
模式 14
模式 12
模式 16
模式 13
模式 18
模式 24
模式 21
模式 19
模式 9
模式 7
模式 5
模式 1
模式 2
模式 10
模式 22
模式 6
EMX R

图53



泳道 #	使用的Cas9	左 gRNA	右 gRNA	突出端 (D10A) - 从5'突出的碱基数目
1	D10A	左23	左12	-36
2	D10A	右 4	左 9	-25
3	D10A	左23	_右 23	-16
4	D10A	右7	左 10	-15
5	D10A	右 16	左 3	-8
6	D10A	右 22	右 6	26
7	D10A	左 12	_右 16	31
8	D10A	左 12	右 13	34
9	D10A	左 10	_右 1	38
10	D10A	_右 23	右 16	51
11	D10A	右 23	_右 13	54
12	D10A	_左3	_右 7	57
13	D10A	左12	右 4	65
14	D10A	左12	右 3	69
15	D10A	左 3	右 10	76
16	D10A	_右 23	右 4	85
17	D10A	左 12	右 9	95
18	D10A	左 12	_右 10	115
19	D10A	_右 23	右 10	135
20	D10A	左12	右 2	145
21	D10A	左 12	左22	181
22	D10A	右 23	左22	201
23	D10A	左 12	_右 6	222
24	D10A	右23	右6	242

图54

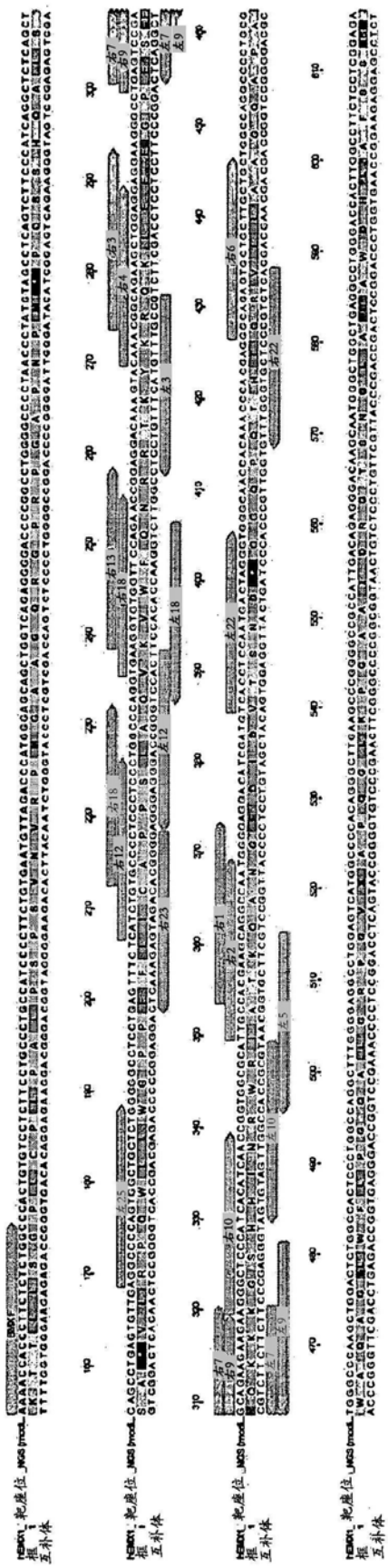


图55

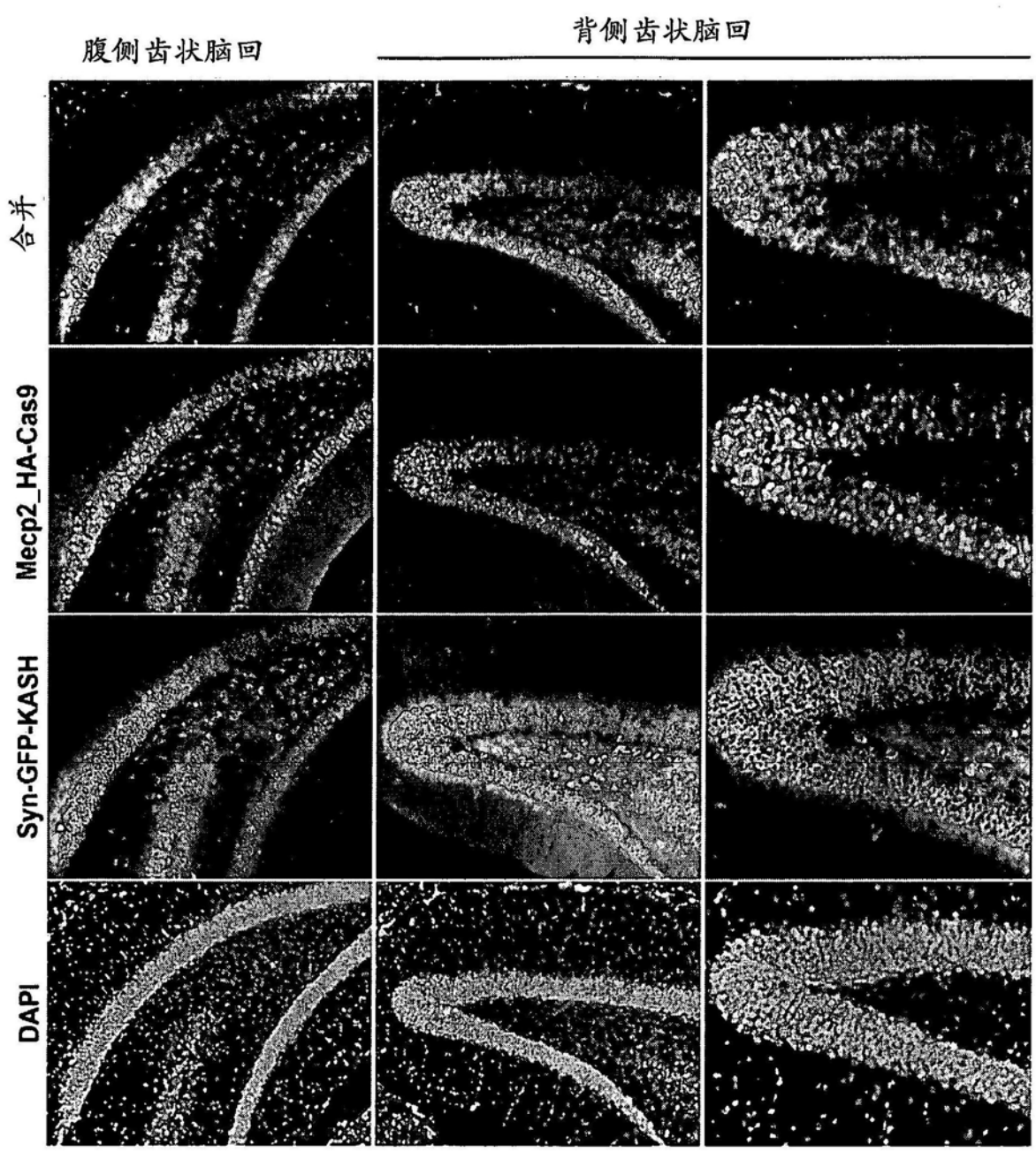
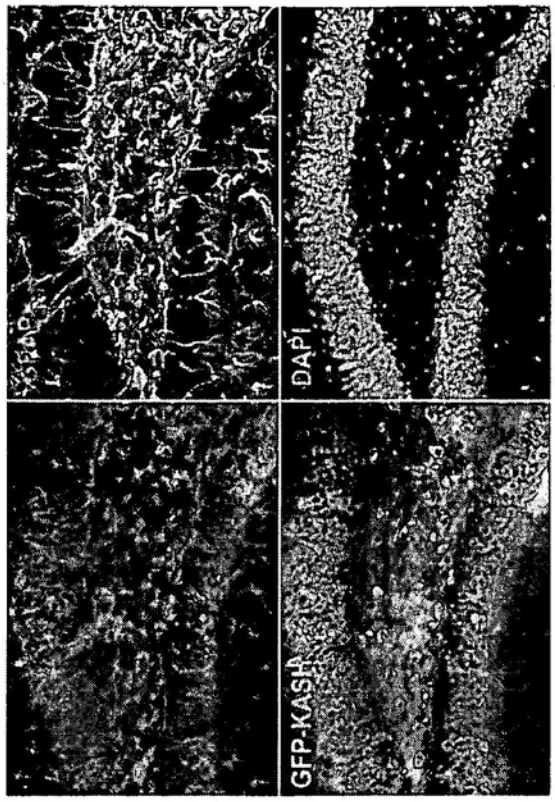


图56



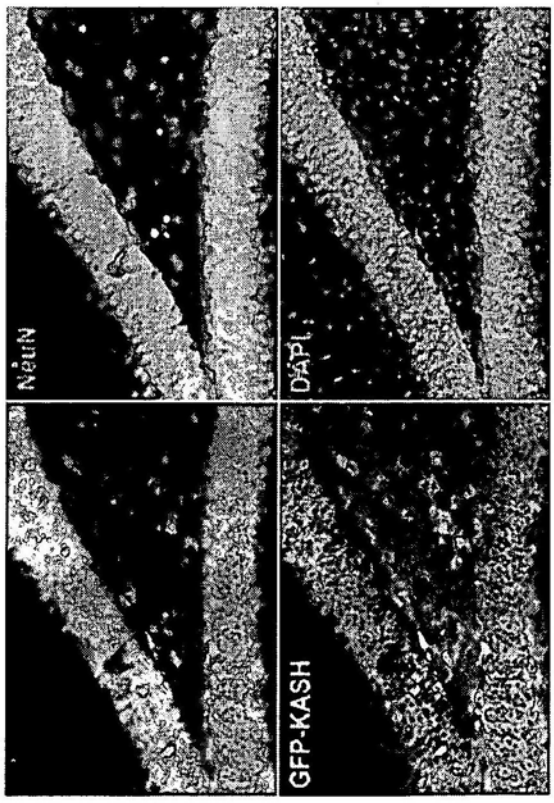


图57

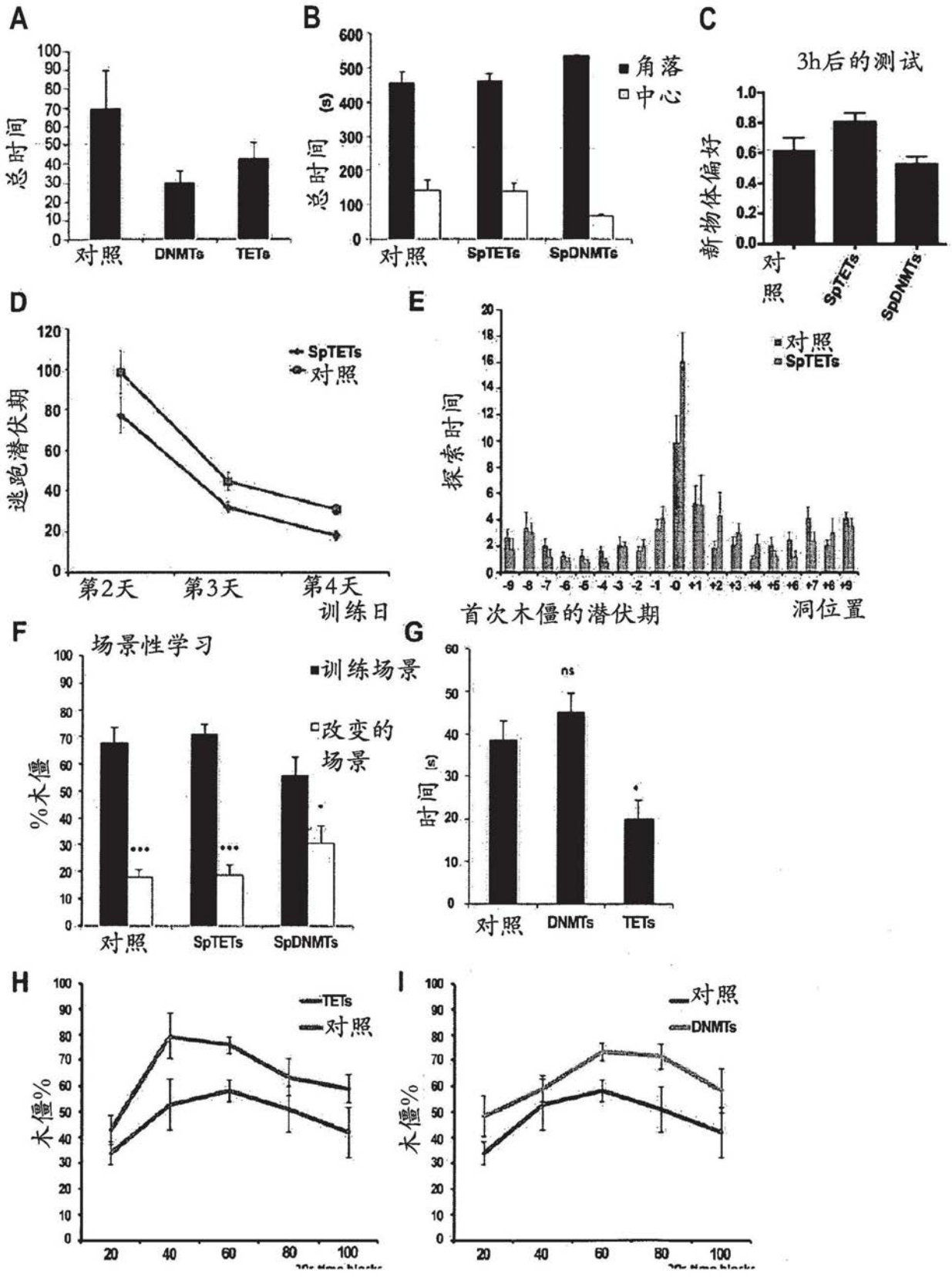

A
总时间
对照
DNMTs
TETs
B
总时间 (s)
角落
中心
对照
SpTETs
SpDNMTs
C
3h后的测试
新物体偏好
对照
SpTETs
SpDNMTs
D
逃跑潜伏期
SpTETs
对照
第2天
第3天
第4天
训练日
E
探索时间
对照
SpTETs
首次木僵的潜伏期
洞位置
F
场景性学习
训练场景
改变的场景
%木僵
对照
SpTETs
SpDNMTs
G
时间 (s)
ns
对照
DNMTs
TETs
H
木僵%
TETs
对照
I
木僵%
对照
DNMTs

图58

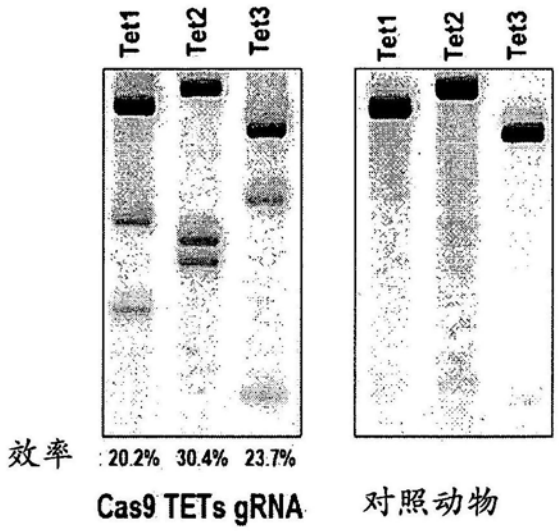

Tet1
Tet2
Tet3
Tet1
Tet2
Tet3
效率
20.2%
30.4%
23.7%
Cas9 TETs gRNA
对照动物

图59

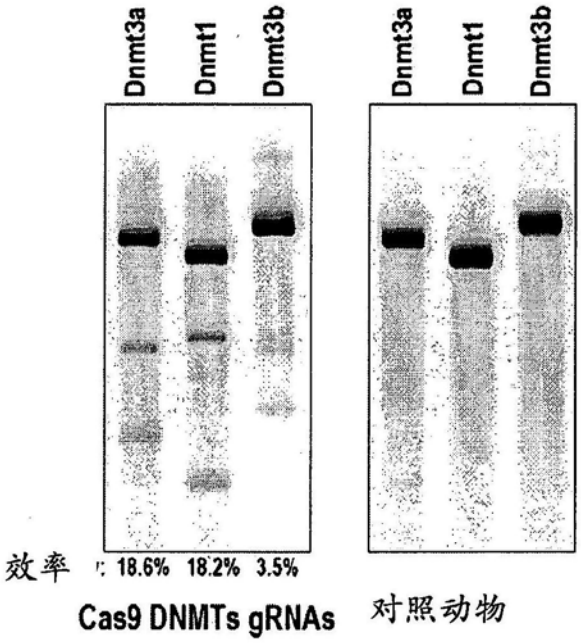

Dnmt3a
Dnmt1
Dnmt3b
Dnmt3a
Dnmt1
Dnmt3b
效率
18.6%
18.2%
3.5%
Cas9 DNMTs gRNAs
对照动物

图60

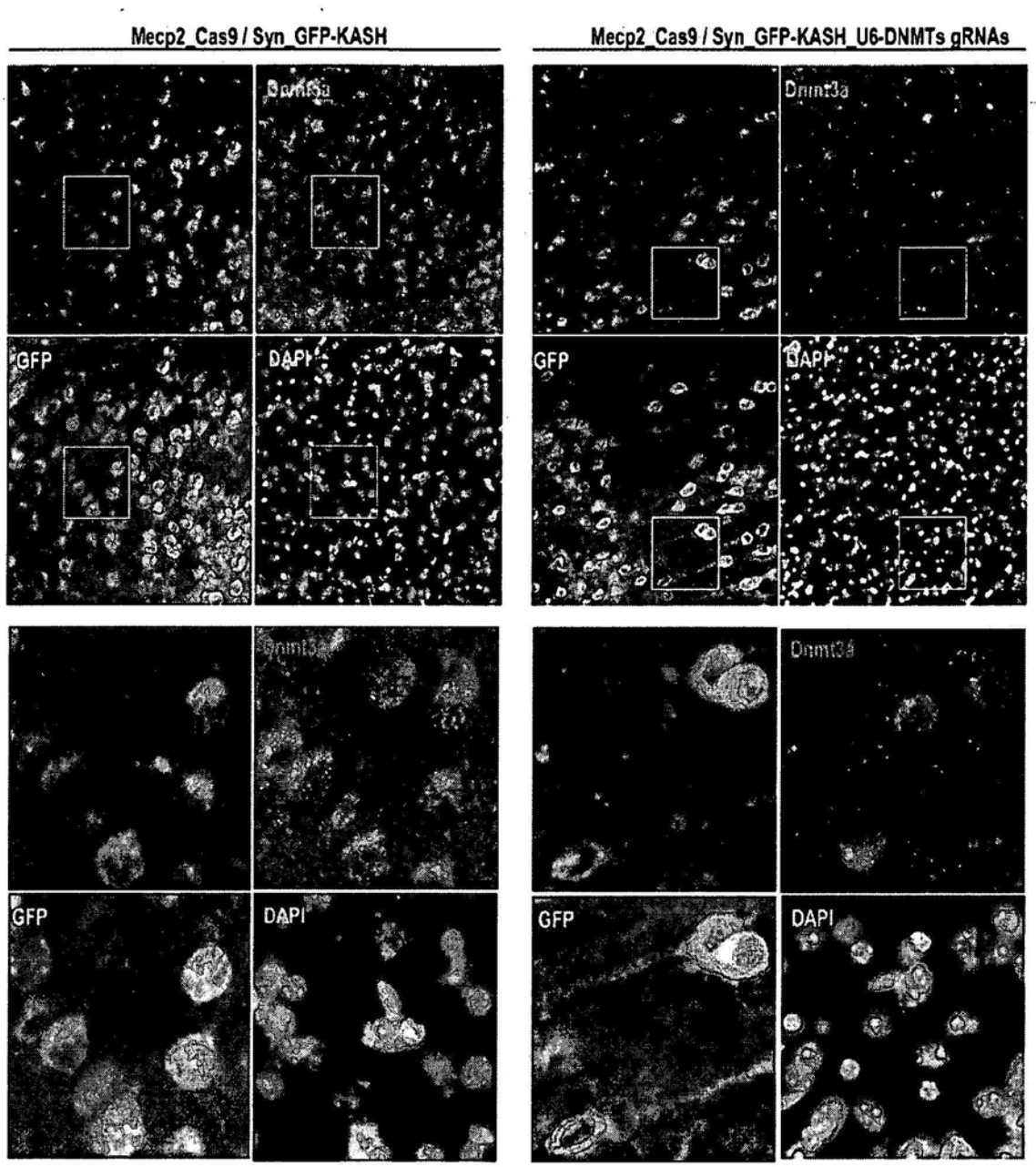

Mecp2_Cas9 / Syn_GFP-KASH
Mecp2_Cas9 / Syn_GFP-KASH_U6-DNMTs gRNAs
Dnmt3a
GFP
DAPI

图61

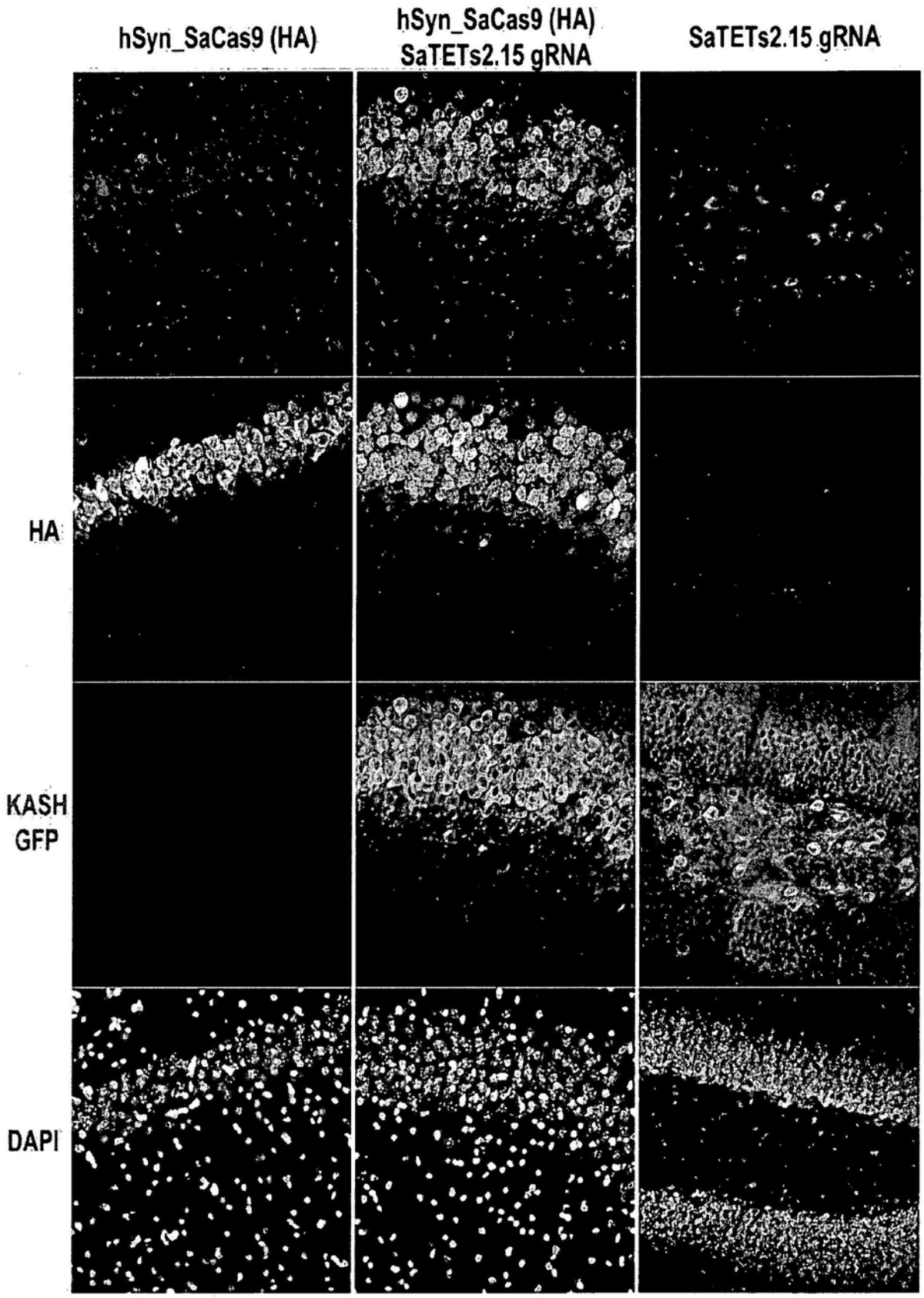

hSyn_SaCas9 (HA)
hSyn_SaCas9 (HA)
SaTETs2.15 gRNA
SaTETs2.15 gRNA
HA
KASH
GFP
DAPI

图62

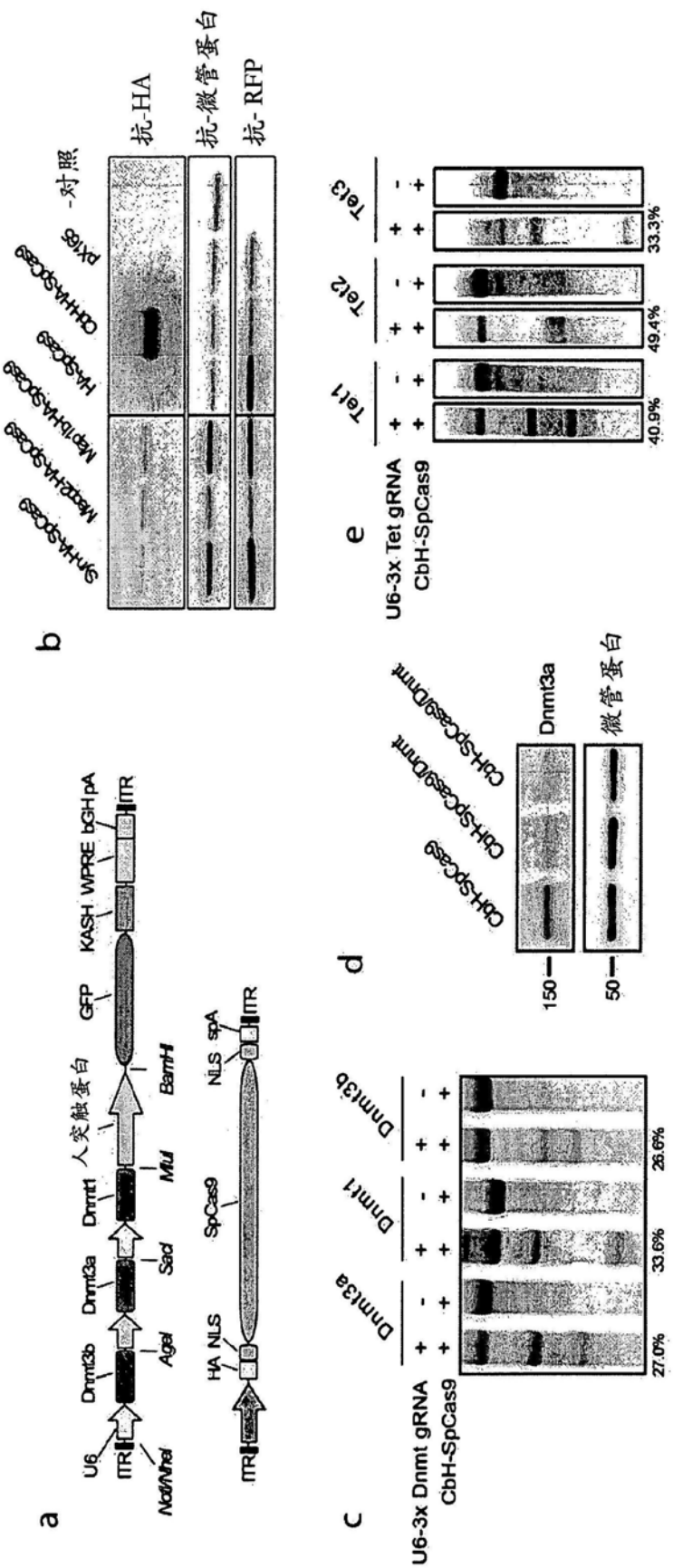

a
U6
ITR
NotI/NheI
Dnmt3b
AgeI
Dnmt3a
SacI
Dnmt1
MluI
人突触蛋白
BamHI
GFP
KASH
WPRE
bGH pA
ITR
HA NLS
SpCas9
NLS
SpA
ITR
b
-对照
pX165
CbH-HA-SpCas9
Hp-SpCas9
Mep1b-HA-SpCas9
Mea2-HA-SpCas9
Sph-HA-SpCas9
抗-HA
抗-微管蛋白
抗- RFP
c
U6-3x Dnmt gRNA
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
27.0%
33.6%
26.6%
d
CbH-SpCas9
CbH-SpCas9/Dnmt
CbH-SpCas9/Dnmt
Dnmt3a
微管蛋白
150
50
e
U6-3x Tet gRNA
CbH-SpCas9
Tet1
Tet2
Tet3
40.9%
49.4%
33.3%

图63

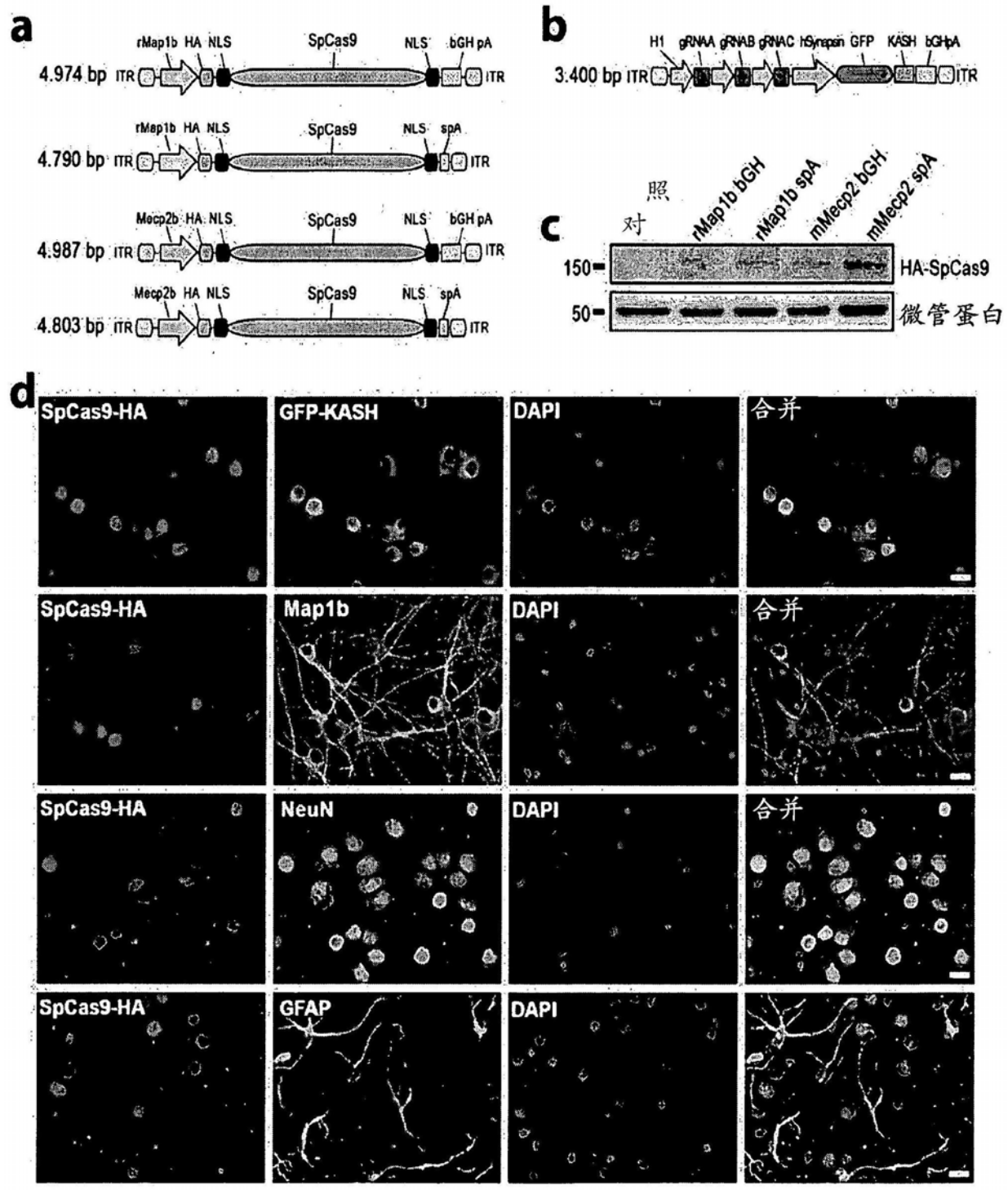

a
rMap1b HA NLS SpCas9 NLS bGH pA
4.974 bp ITR ITR
rMap1b HA NLS SpCas9 NLS spA
4.790 bp ITR ITR
Mecp2b HA NLS SpCas9 NLS bGH pA
4.987 bp ITR ITR
Mecp2b HA NLS SpCas9 NLS spA
4.803 bp ITR ITR
b
H1 gRNAA gRNAB gRNAC hSynapsn GFP KASH bGHpA
3.400 bp ITR ITR
c
对照
rMap1b bGH
rMap1b spA
mMecp2 bGH
mMecp2 spA
150
HA-SpCas9
50
微管蛋白
d
SpCas9-HA
GFP-KASH
DAPI
合并
SpCas9-HA
Map1b
DAPI
合并
SpCas9-HA
NeuN
DAPI
合并
SpCas9-HA
GFAP
DAPI

图64

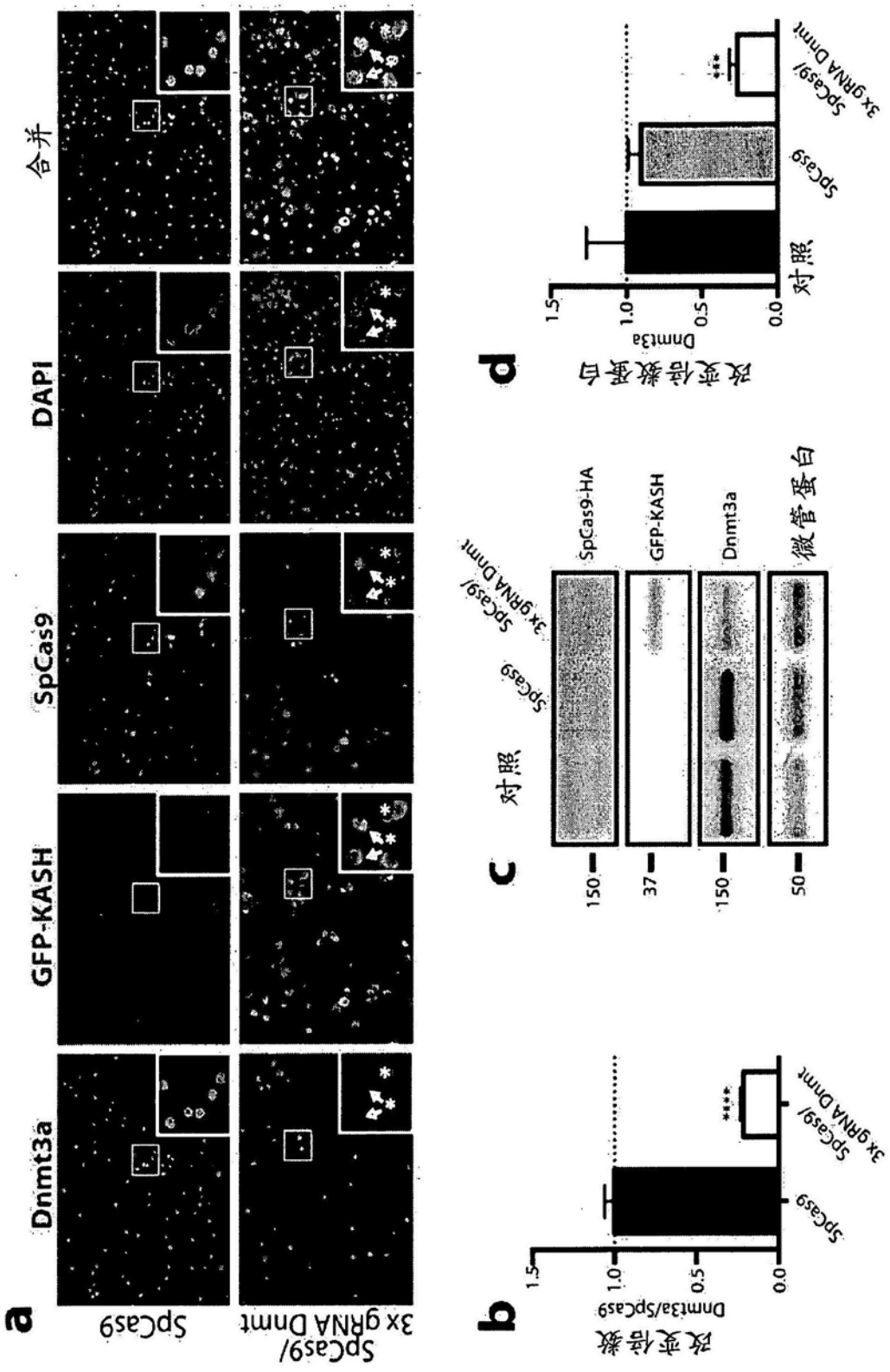

a
Dnmt3a
GFP-KASH
SpCas9
DAPI
合并
SpCas9
SpCas9/ 3x gRNA Dnmt
b
改变倍数
Dnmt3a/SpCas9
1.5
1.0
0.5
0.0
****
SpCas9
SpCas9/ 3x gRNA Dnmt
c
对照
SpCas9
SpCas9/ 3x gRNA Dnmt
SpCas9-HA
GFP-KASH
Dnmt3a
微管蛋白
150
37
150
50
d
改变倍数蛋白
Dnmt3a
1.5
1.0
0.5
0.0
对照
SpCas9
SpCas9/ 3x gRNA Dnmt
**

图65

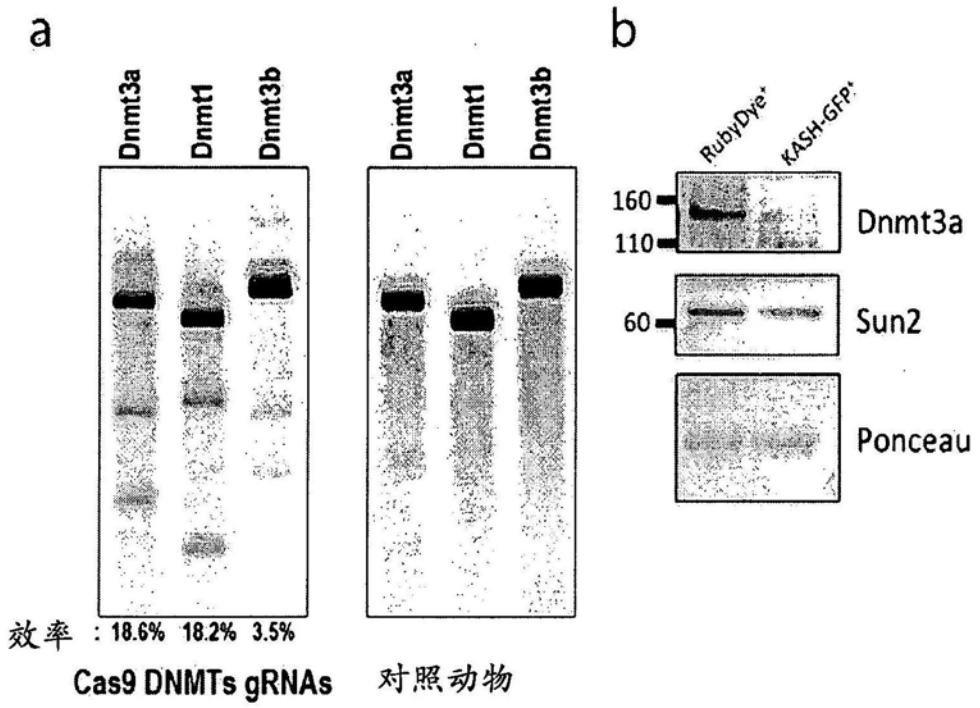

a
Dnmt3a
Dnmt1
Dnmt3b
Dnmt3a
Dnmt1
Dnmt3b
效率 ：18.6% 18.2% 3.5%
Cas9 DNMTs gRNAs
对照动物
b
RubyDye+
KASH-GFP+
160
110
Dnmt3a
60
Sun2
Ponceau

图66

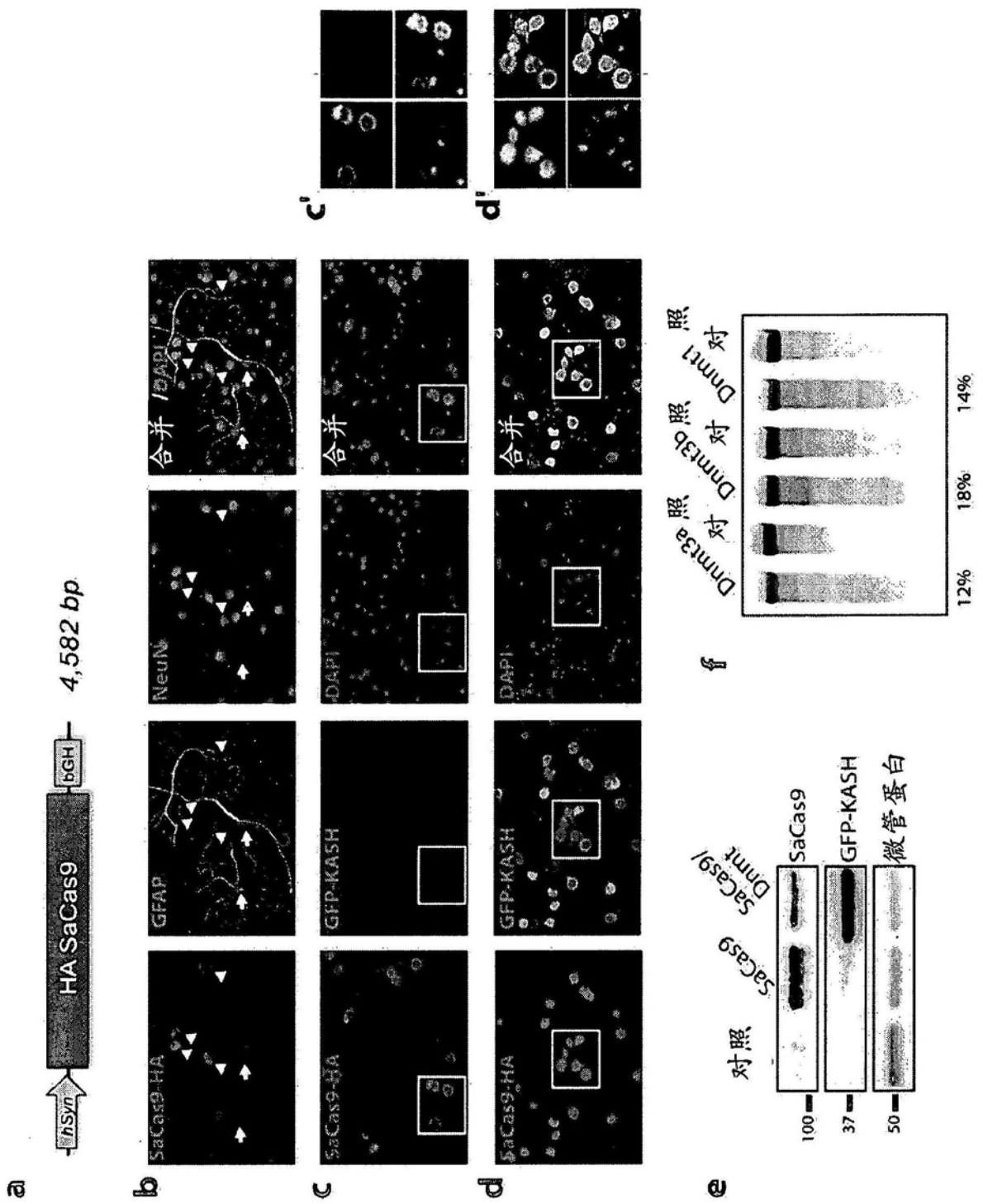

a
hSyn
HA SaCas9
bGH
4,582 bp
b
SaCas9-HA
GFAP
NeuN
合并 /DAPI
c
SaCas9-HA
GFP-KASH
DAPI
合并
d
SaCas9-HA
GFP-KASH
DAPI
合并
c'
d'
e
对照
SaCas9
SaCas9/Dnmt
SaCas9
GFP-KASH
微管蛋白
100
37
50
f
Dnmt3a 对照
Dnmt3a
Dnmt3b 对照
Dnmt3b
Dnmt1 对照
Dnmt1
12%
18%
14%

图67

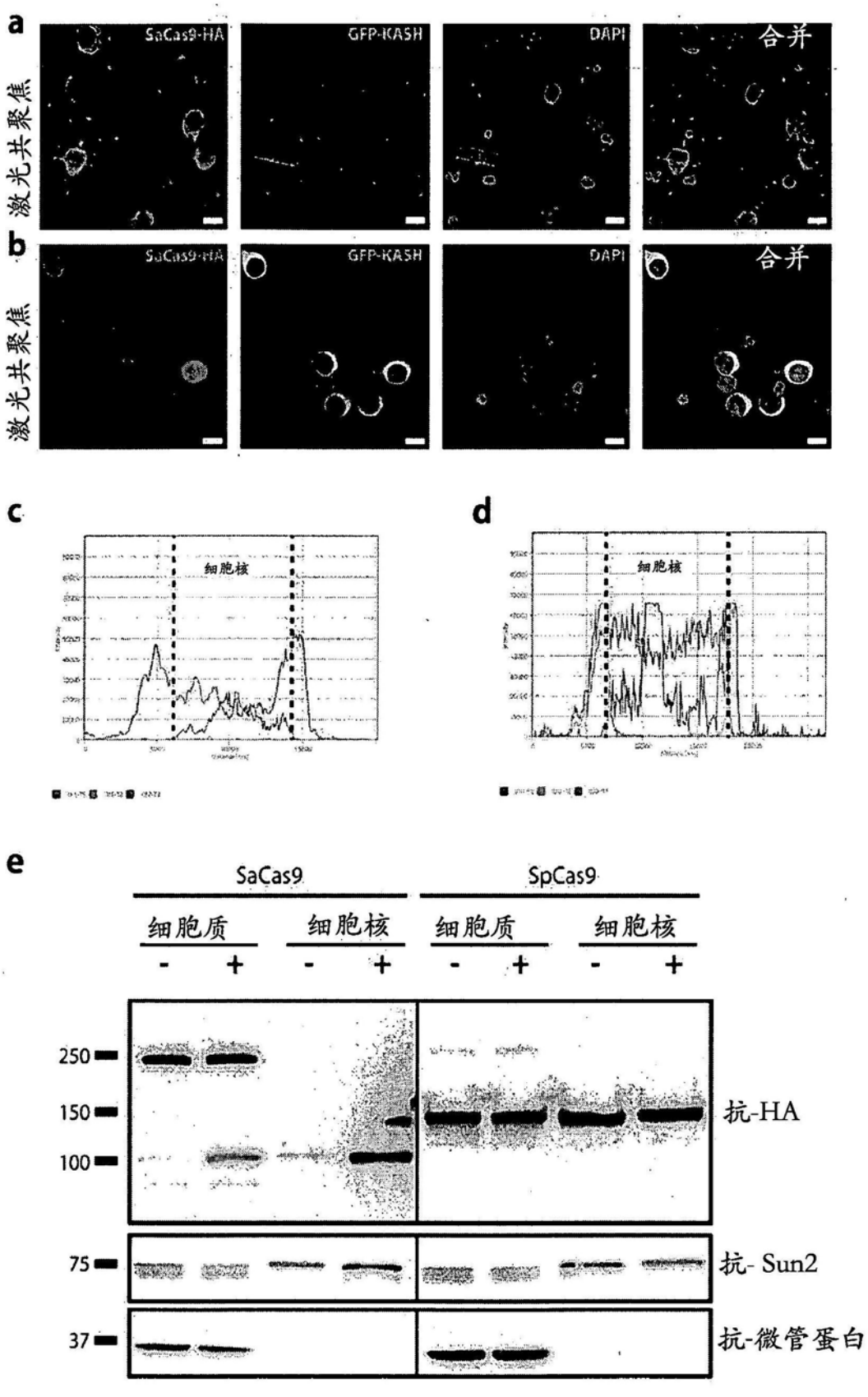
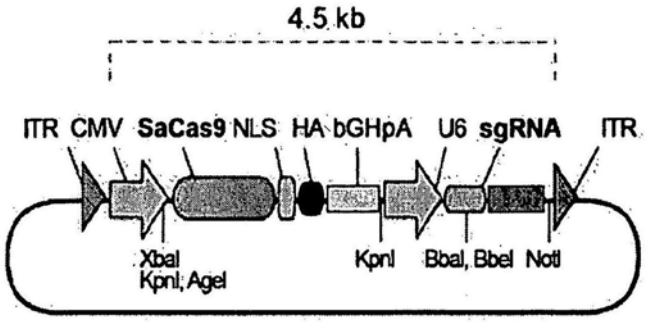
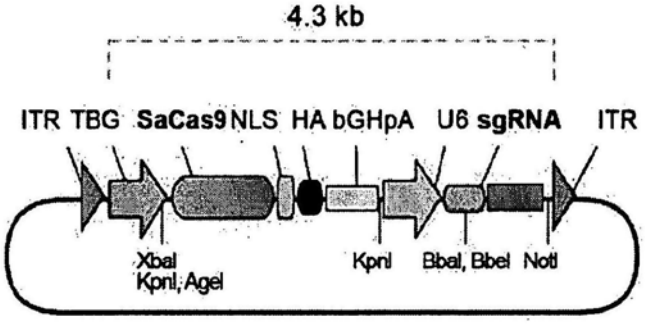

a
激光共聚焦
SaCas9-HA
GFP-KASH
DAPI
合并
b
激光共聚焦
SaCas9-HA
GFP-KASH
DAPI
合并
c
细胞核
d
细胞核
e
SaCas9
SpCas9
细胞质
细胞核
细胞质
细胞核
- + - + - + - +
250
150
100
抗-HA
75
抗- Sun2
37
抗-微管蛋白

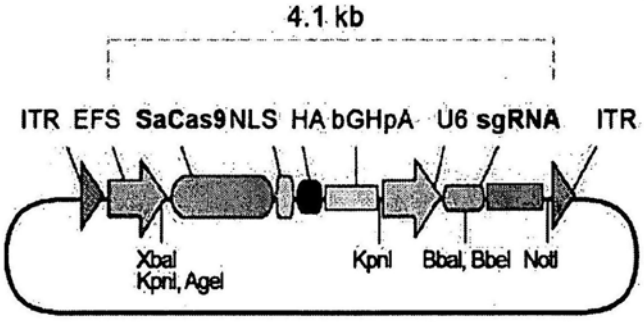
图68



AAV-Sa-Cas9 载体 :
CMV-SaCas9-NLS-HA-U6-sgRNA



肝特异性AAV-Sa-Cas9载体:
TBG-SaCas9-NLS-HA-U6-sgRNA



可替代的AAV-Sa-Cas9载体:
EFS-SaCas9-NLS-HA-U6-sgRNA

图69

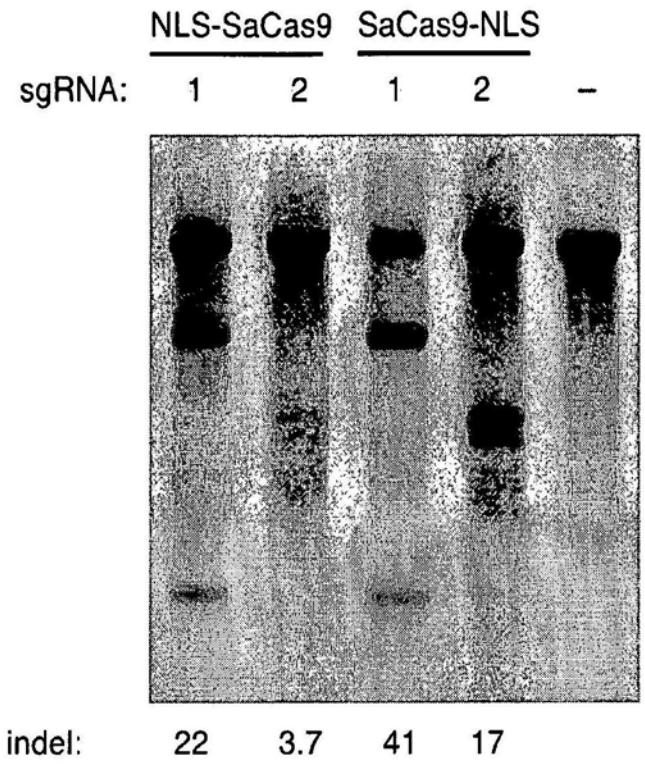


图70

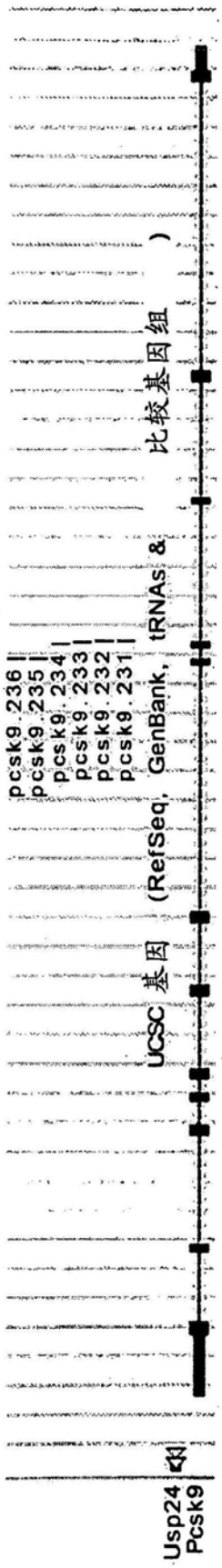

Usp24
Pcsk9
UCSC 基因 (RefSeq, GenBank, tRNAs & 比较基因组)
pcsk9_236
pcsk9_235
pcsk9_234
pcsk9_233
pcsk9_232
pcsk9_231

图71

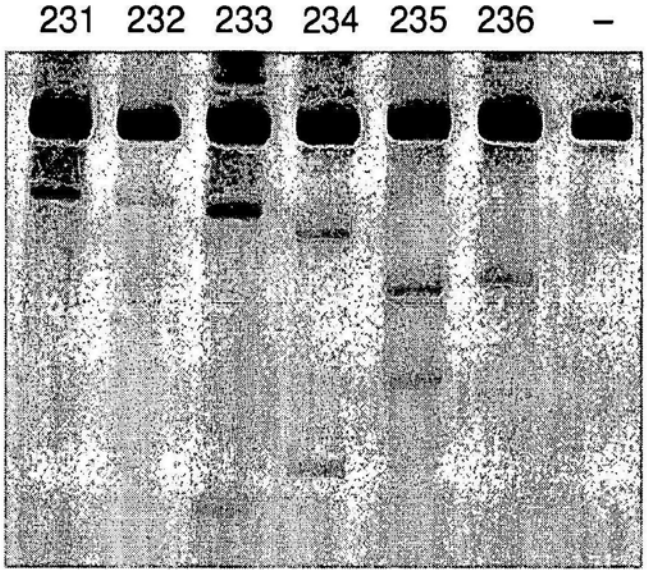

231
232
233
234
235
236
–

图72

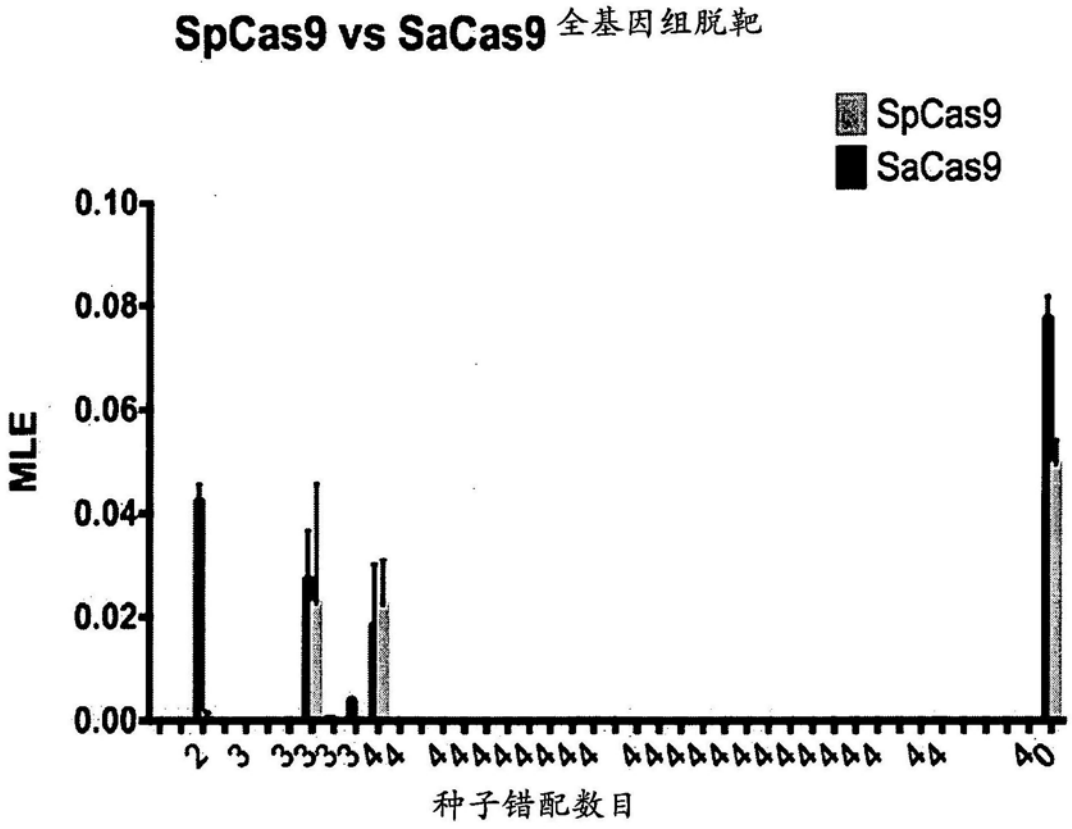

SpCas9 vs SaCas9 全基因组脱靶
SpCas9
SaCas9
0.10
0.08
0.06
0.04
0.02
0.00
MLE
种子错配数目

图73