



(12)发明专利申请

(10)申请公布号 CN 109312340 A

(43)申请公布日 2019.02.05

(21)申请号 201780027036.0

(74)专利代理机构 北京市君合律师事务所  
11517

(22)申请日 2017.03.03

代理人 张怡 牡丹

(30)优先权数据

62/304,057 2016.03.04 US

(51)Int.Cl.

C12N 15/113(2006.01)

(85)PCT国际申请进入国家阶段日

2018.10.31

C12N 9/22(2006.01)

C12N 15/09(2006.01)

(86)PCT国际申请的申请数据

PCT/US2017/020598 2017.03.03

C12N 15/63(2006.01)

A61P 35/00(2006.01)

(87)PCT国际申请的公布数据

W02017/152015 EN 2017.09.08

(71)申请人 爱迪塔斯医药公司

地址 美国马萨诸塞州

(72)发明人 G·G·威尔斯戴德 H·嘉亚拉穆

王童瑶 J·A·左瑞斯

C·博尔赫斯

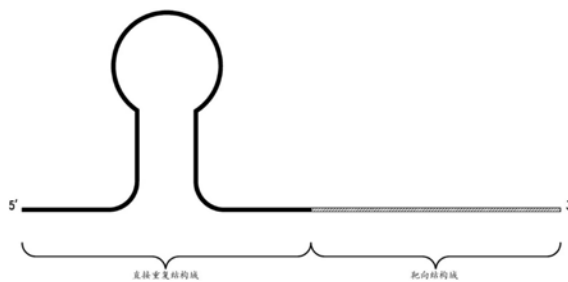
权利要求书7页 说明书62页 附图8页

(54)发明名称

用于癌症免疫疗法的CRISPR-CPF1相关方法、组合物和组分

(57)摘要

用于治疗癌症的CRISPR/Cpf1相关组合物和方法。



1. 一种抑制性着丝粒和启动子因子1 (Cpf1) gRNA分子,其包含与来自选自下组的一个T细胞表达的基因的靶结构域互补的靶向结构域,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

2. 如权利要求1所述的抑制性Cpf1 gRNA分子,其进一步包含直接重复结构域。

3. 如权利要求2所述的抑制性Cpf1 gRNA分子,其中所述直接重复结构域在长度上是15-20个核苷酸。

4. 如权利要求2或3所述的抑制性Cpf1 gRNA分子,其中所述直接重复结构域包含选自下组的核苷酸序列,该组由以下组成:SEQ ID NO:3708-3710。

5. 如权利要求1-4中任一项所述的抑制性Cpf1 gRNA分子,其中所述靶结构域被靶向用于敲除所述T细胞表达的基因的表达。

6. 如权利要求1-5中任一项所述的抑制性Cpf1 gRNA分子,其中所述靶结构域在所述T细胞表达的基因的编码区内。

7. 如权利要求6所述的抑制性Cpf1 gRNA分子,其中所述编码区是所述T细胞表达的基因的早期编码区。

8. 如权利要求1-4中任一项所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域被配置成在T细胞靶敲除位置的500、450、400、350、300、250、200、150、100、50、25或10个核苷酸内提供双链断裂。

9. 如权利要求1-8中任一项所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域包含与选自SEQ ID NO:1-3707的核苷酸序列相同或与所述核苷酸序列相差不超过1、不超过2、不超过3、不超过4或不超过5个核苷酸的核苷酸序列。

10. 如权利要求9所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域包含选自SEQ ID NO:1-3707的核苷酸序列。

11. 如权利要求1-10中任一项所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是15-25个核苷酸。

12. 如权利要求11所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是18个核苷酸。

13. 如权利要求11所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是19个核苷酸。

14. 如权利要求11所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是20个核苷酸。

15. 如权利要求11所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是21个核苷酸。

16. 如权利要求11所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是22个核苷酸。

17. 如权利要求11所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是23个核苷酸。

18. 如权利要求11所述的抑制性Cpf1 gRNA分子,其中所述靶向结构域在长度上是24个核苷酸。

19. 如权利要求1-18中任一项所述的抑制性Cpf1 gRNA分子,其中所述抑制性Cpf1

gRNA是单模块化抑制性Cpf1 gRNA分子。

20. 一种核酸组合物,其包含:(a) 编码第一抑制性Cpf1 gRNA分子的第一核苷酸序列,所述第一抑制性Cpf1 gRNA分子包含与来自选自下组的一个T细胞表达的基因的靶结构域互补的靶向结构域,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

21. 如权利要求20所述的核酸组合物,其中所述第一抑制性Cpf1 gRNA分子是如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子。

22. 如权利要求20或21所述的核酸组合物,其进一步包含直接重复结构域。

23. 如权利要求22所述的核酸组合物,其中所述直接重复结构域在长度上是15-20个核苷酸。

24. 如权利要求22或23所述的核酸组合物,其中所述直接重复结构域包含选自下组的核苷酸序列,该组由以下组成:SEQ ID NO:3708-3710。

25. 如权利要求20-24中任一项所述的核酸组合物,其中所述靶结构域被靶向用于敲除所述T细胞表达的基因的表达。

26. 如权利要求20-25中任一项所述的核酸组合物,其中所述靶结构域在所述T细胞表达的基因的编码区内。

27. 如权利要求26所述的核酸组合物,其中所述编码区是所述T细胞表达的基因的早期编码区。

28. 如权利要求20-27中任一项所述的核酸组合物,其中所述靶向结构域被配置成在T细胞靶敲除位置的500、450、400、350、300、250、200、150、100、50、25或10个核苷酸内提供双链断裂。

29. 如权利要求20-28中任一项所述的核酸组合物,其中所述靶向结构域包含与选自SEQ ID NO:1-3707的核苷酸序列相同或与所述核苷酸序列相差不超过1、不超过2、不超过3、不超过4或不超过5个核苷酸的核苷酸序列。

30. 如权利要求29所述的核酸组合物,其中所述靶向结构域包含选自SEQ ID NO:1-3707的核苷酸序列。

31. 如权利要求20-30中任一项所述的核酸组合物,其中所述靶向结构域在长度上是15-25个核苷酸。

32. 如权利要求31所述的核酸组合物,其中所述靶向结构域在长度上是18个核苷酸。

33. 如权利要求31所述的核酸组合物,其中所述靶向结构域在长度上是19个核苷酸。

34. 如权利要求31所述的核酸组合物,其中所述靶向结构域在长度上是20个核苷酸。

35. 如权利要求31所述的核酸组合物,其中所述靶向结构域在长度上是21个核苷酸。

36. 如权利要求31所述的核酸组合物,其中所述靶向结构域在长度上是22个核苷酸。

37. 如权利要求31所述的核酸组合物,其中所述靶向结构域在长度上是23个核苷酸。

38. 如权利要求31所述的核酸组合物,其中所述靶向结构域在长度上是24个核苷酸。

39. 如权利要求20-38中任一项所述的核酸组合物,其进一步包含:(b) 编码Cpf1分子的第二核苷酸序列。

40. 如权利要求39所述的核酸组合物,其中所述Cpf1分子在靶核酸中形成双链断裂。

41. 如权利要求39或40所述的核酸组合物,其中所述Cpf1分子选自下组,该组由以下组

成:氨基酸球菌属菌株BV3L6 Cpf1分子(AsCpf1)、毛螺菌科细菌ND2006 Cpf1分子(LbCpf1)和毛螺菌科细菌MA2020(Lb2Cpf1)。

42.如权利要求39-41中任一项所述的核酸组合物,其中所述第二核苷酸序列示于SEQ ID NO:3722、SEQ ID NO:3723或SEQ ID NO:3724中。

43.如权利要求39-42中任一项所述的核酸组合物,其中(a)和(b)存在于一个核酸分子上。

44.如权利要求43所述的核酸组合物,其中所述核酸分子是AAV载体。

45.如权利要求43或44中任一项所述的核酸组合物,其中:(a)存在于第一核酸分子上;并且(b)存在于第二核酸分子上。

46.如权利要求45所述的核酸组合物,其中所述第一核酸分子和第二核酸分子是AAV载体或LV载体。

47.如权利要求39-47中任一项所述的核酸组合物,其进一步包含:(c)编码第二抑制性Cpf1 gRNA分子的第三核苷酸序列,所述第二抑制性Cpf1 gRNA分子包含与选自下组的T细胞表达的基因的第二靶结构域互补的靶向结构域,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

48.如权利要求47所述的核酸组合物,其中所述第二抑制性Cpf1 gRNA分子是如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子。

49.如权利要求47或48所述的核酸组合物,其中所述第一抑制性Cpf1 gRNA分子和所述第二抑制性Cpf1 gRNA分子侧翼于T细胞靶敲除位置产生两组双链断裂。

50.如权利要求47-49中任一项所述的核酸组合物,其中所述第二抑制性Cpf1 gRNA分子的靶向结构域包含与选自SEQ ID NO:1-3707的核苷酸序列相同或与所述核苷酸序列相差不超过1、不超过2、不超过3、不超过4或不超过5个核苷酸的核苷酸序列。

51.如权利要求50所述的核酸组合物,其中所述第二抑制性Cpf1 gRNA的靶向结构域包含选自SEQ ID NO:1-3707的核苷酸序列。

52.如权利要求47-51中任一项所述的核酸组合物,其中(a)和(c)存在于一个核酸分子上。

53.如权利要求52所述的核酸组合物,其中所述核酸分子是AAV载体或LV载体。

54.如权利要求47-51中任一项所述的核酸组合物,其中:(a)存在于第一核酸分子上;并且(c)存在于第二核酸分子上。

55.如权利要求54所述的核酸组合物,其中所述第一核酸分子和第二核酸分子是AAV载体或LV载体。

56.如权利要求47-51中任一项所述的核酸组合物,其中(a)、(b)和(c)存在于一个核酸分子上。

57.如权利要求56所述的核酸组合物,其中所述核酸分子是AAV载体或LV载体。

58.如权利要求47-51中任一项所述的核酸组合物,其中:(a)、(b)和(c)之一在第一核酸分子上编码;并且(a)、(b)和(c)中的第二个和第三个在第二核酸分子上编码。

59.如权利要求58所述的核酸组合物,其中所述第一核酸分子和第二核酸分子是AAV载体或LV载体。

60.如权利要求58或59所述的核酸组合物,其中:(a)存在于第一核酸分子上;并且(b)

和(c)存在于第二核酸分子上。

61. 如权利要求58或59所述的核酸组合物,其中:(b)存在于第一核酸分子上;并且(a)和(c)存在于第二核酸分子上。

62. 如权利要求58或59所述的核酸组合物,其中:(c)存在于第一核酸分子上;并且(b)和(a)存在于第二核酸分子上。

63. 如权利要求20-62中任一项所述的核酸组合物,其中所述核酸组合物包含可操作地连接至(a)的启动子。

64. 如权利要求39-63中任一项所述的核酸组合物,其中所述核酸组合物包含可操作地连接至(b)的启动子。

65. 如权利要求47-64中任一项所述的核酸组合物,其中所述核酸组合物包含可操作地连接至(c)的第二启动子。

66. 一种组合物,其包含(a)如权利要求1-19中任一项所述的抑制性Cpf1gRNA分子。

67. 如权利要求66所述的组合物,其进一步包含(b)Cpf1分子。

68. 如权利要求67所述的组合物,其中所述Cpf1分子在靶核酸中形成双链断裂。

69. 如权利要求67或68所述的组合物,其中所述Cpf1分子选自下组,该组由以下组成:氨基酸球菌属菌株BV3L6 Cpf1分子(AsCpf1)、毛螺菌科细菌ND2006 Cpf1分子(LbCpf1)和毛螺菌科细菌MA2020(Lb2Cpf1)。

70. 如权利要求66-69中任一项所述的组合物,其进一步包含(c)第二抑制性Cpf1 gRNA分子。

71. 如权利要求66-70中任一项所述的组合物,其是包含Cpf1蛋白和编码所述抑制性Cpf1 gRNA分子的核糖核酸分子的核糖核蛋白组合物。

72. 一种改变细胞的方法,其包括使所述细胞与以下接触:

(a) 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA;和

(b) Cpf1分子。

73. 如权利要求72所述的方法,其中所述Cpf1分子在靶核酸中形成双链断裂。

74. 如权利要求72或73所述的方法,其中所述Cpf1分子选自下组,该组由以下组成:氨基酸球菌属菌株BV3L6 Cpf1分子(AsCpf1)、毛螺菌科细菌ND2006 Cpf1分子(LbCpf1)和毛螺菌科细菌MA2020(Lb2Cpf1)。

75. 如权利要求72-74中任一项所述的方法,其进一步包括使所述细胞与(c)第二抑制性Cpf1 gRNA分子接触,所述第二抑制性Cpf1 gRNA分子包含与选自下组的T细胞表达的基因的第二靶结构域互补的靶向结构域,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

76. 如权利要求72-75中任一项所述的方法,其中所述细胞来自患有癌症的受试者。

77. 如权利要求76所述的方法,其中所述癌症选自下组,该组由以下组成:淋巴瘤、慢性淋巴细胞白血病(CLL)、B细胞急性淋巴细胞白血病(B-ALL)、急性成淋巴细胞白血病、急性骨髓性白血病、非霍奇金淋巴瘤(NHL)、弥漫性大细胞淋巴瘤(DLCL)、多发性骨髓瘤、肾细胞癌(RCC)、成神经细胞瘤、结肠直肠癌、乳腺癌、卵巢癌、黑色素瘤、肉瘤、前列腺癌、肺癌、食道癌、肝细胞癌、胰腺癌、星形细胞瘤、间皮瘤、头颈癌以及成神经管细胞瘤。

78. 如权利要求72-77中任一项所述的方法,其中所述细胞来自将受益于在一个或多个

T细胞表达的基因中的一个或多个T细胞靶位置处的一个或多个改变的受试者。

79. 如权利要求72-78中任一项所述的方法,其中所述细胞是T细胞。

80. 如权利要求79所述的方法,其中所述T细胞是工程化T细胞。

81. 如权利要求80所述的方法,其中所述工程化T细胞是工程化的嵌合抗原受体(CAR) T细胞。

82. 如权利要求80所述的方法,其中所述工程化T细胞是工程化的TCR(T细胞受体) T细胞。

83. 如权利要求79-82中任一项所述的方法,其中所述T细胞被工程化为在所述T细胞表达的基因的T细胞靶敲除位置内引入改变之前表达TCR或CAR。

84. 如权利要求79-82中任一项所述的方法,其中所述T细胞被工程化为在所述T细胞表达的基因的T细胞靶敲除位置内引入改变之后表达TCR或CAR。

85. 如权利要求79-82中任一项所述的方法,其中所述T细胞被工程化为在所述T细胞表达的基因的T细胞靶敲除位置内引入改变的同时表达TCR或CAR。

86. 如权利要求72-85中任一项所述的方法,其中所述接触离体地进行。

87. 如权利要求72-86中任一项所述的方法,其中将所述接触的细胞返回所述受试者体内。

88. 如权利要求72-87中任一项所述的方法,其中所述接触包括使所述细胞与如权利要求20-65中任一项所述的核酸组合物接触。

89. 如权利要求72-87中任一项所述的方法,其中所述接触包括使所述细胞与如权利要求66-71中任一项所述的组合物接触。

90. 一种治疗受试者的方法,其包括使来自受试者的细胞与以下接触:

(a) 如权利要求1-35中任一项所述的抑制性Cpf1 gRNA;和

(b) Cpf1分子。

91. 如权利要求90所述的方法,其中所述Cpf1分子在靶核酸中形成双链断裂。

92. 如权利要求90或91所述的方法,其中所述Cpf1分子选自下组,该组由以下组成:氨基酸球菌属菌株BV3L6 Cpf1分子(AsCpf1)、毛螺菌科细菌ND2006 Cpf1分子(LbCpf1)和毛螺菌科细菌MA2020(Lb2Cpf1)。

93. 如权利要求90-92中任一项所述的方法,其进一步包括使所述细胞与(c) 第二抑制性Cpf1 gRNA分子接触,所述第二抑制性Cpf1 gRNA分子包含与选自下组的T细胞表达的基因的第二靶结构域互补的靶向结构域,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

94. 如权利要求90-93中任一项所述的方法,其中所述受试者患有癌症。

95. 如权利要求94所述的方法,其中所述癌症选自下组,该组由以下组成:淋巴瘤、慢性淋巴细胞白血病(CLL)、B细胞急性淋巴细胞白血病(B-ALL)、急性成淋巴细胞白血病、急性骨髓性白血病、非霍奇金淋巴瘤(NHL)、弥漫性大细胞淋巴瘤(DLCL)、多发性骨髓瘤、肾细胞癌(RCC)、成神经细胞瘤、结肠直肠癌、乳腺癌、卵巢癌、黑色素瘤、肉瘤、前列腺癌、肺癌、食道癌、肝细胞癌、胰腺癌、星形细胞瘤、间皮瘤、头颈癌以及成神经管细胞瘤。

96. 如权利要求90-95中任一项所述的方法,其中所述受试者将受益于在一个或多个T细胞表达的基因中的一个或多个T细胞靶位置处的一个或多个改变。

97. 如权利要求90-96中任一项所述的方法,其中所述细胞是T细胞。
98. 如权利要求97所述的方法,其中所述T细胞是工程化T细胞。
99. 如权利要求98所述的方法,其中所述工程化T细胞是工程化的嵌合抗原受体(CAR) T细胞。
100. 如权利要求98所述的方法,其中所述工程化T细胞是工程化的T细胞受体(TCR) T细胞。
101. 如权利要求97-100中任一项所述的方法,其中所述T细胞被工程化为在所述T细胞表达的基因的T细胞靶敲除位置内引入改变之前表达TCR或CAR。
102. 如权利要求97-100中任一项所述的方法,其中所述T细胞被工程化为在所述T细胞表达的基因的T细胞靶敲除位置内引入改变之后表达TCR或CAR。
103. 如权利要求97-100中任一项所述的方法,其中所述T细胞被工程化为在所述T细胞表达的基因的T细胞靶敲除位置内引入改变的同时表达TCR或CAR。
104. 如权利要求90-103中任一项所述的方法,其中所述接触离体地进行。
105. 如权利要求90-104中任一项所述的方法,其中将所述接触的细胞返回所述受试者体内。
106. 如权利要求90-105中任一项所述的方法,其中所述接触包括使所述细胞与如权利要求20-65中任一项所述的核酸组合物接触。
107. 如权利要求90-105中任一项所述的方法,其中所述接触包括使所述细胞与如权利要求66-71中任一项所述的组合物接触。
108. 一种反应混合物,其包含:
- (a) 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子、如权利要求20-65中任一项所述的核酸组合物或如权利要求66-71中任一项所述的组合物,以及
- (b) 来自将受益于在选自下组的一个或多个T细胞表达的基因中的一个或多个T细胞靶敲除位置处的一个或多个改变的受试者的细胞,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。
109. 一种试剂盒,其包含(a) 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子或编码所述抑制性Cpf1 gRNA的核酸组合物,以及以下中的一种或多种:
- (b) Cpf1分子;
- (c) 第二抑制性Cpf1 gRNA分子,其包含与选自下组的T细胞表达的基因的第二靶结构域互补的靶向结构域,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因;和
- (d) 编码(b)和(c)中的一种或多种的核酸组合物。
110. 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子,其中所述抑制性Cpf1 gRNA分子在其5'端或其附近包含修饰。
111. 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子,其中所述抑制性Cpf1 gRNA分子在其3'端或其附近包含修饰。
112. 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子,其中所述抑制性Cpf1 gRNA分子在其5'端或其附近包含修饰,并且在其3'端或其附近包含修饰。
113. 如权利要求110或112所述的抑制性Cpf1 gRNA分子,其中所述修饰在其5'端的1-

10、1-9、1-8、1-7、1-6、1-5、1-4、1-3或1-2个核苷酸内。

114. 如权利要求111或112所述的抑制性Cpf1 gRNA分子,其中所述修饰在其3'端的1-10、1-9、1-8、1-7、1-6、1-5、1-4、1-3或1-2个核苷酸内。

115. 如权利要求110-114中任一项所述的抑制性Cpf1 gRNA分子,其中所述修饰使得所述抑制性Cpf1 gRNA分子当被引入T细胞中时展现出增加的对于核酸酶的稳定性。

116. 如权利要求110-114中任一项所述的抑制性Cpf1 gRNA分子,其中所述修饰使得所述抑制性Cpf1 gRNA分子当被引入T细胞中时展现出减少的先天性免疫应答。

117. 如权利要求116所述的抑制性Cpf1 gRNA分子,其中所述先天性免疫应答涉及细胞因子表达的诱导。

118. 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子,用于治疗受试者的癌症。

119. 如权利要求118所述的抑制性Cpf1 gRNA分子,其中所述抑制性Cpf1 gRNA分子与(b) Cpf1分子组合使用。

120. 如权利要求1-19中任一项所述的抑制性Cpf1 gRNA分子在制备用于治疗受试者的癌症的药物中的用途。

121. 如权利要求120所述的用途,其中所述药物进一步包含(b) Cpf1分子。

122. 如权利要求20-65中任一项所述的核酸组合物,用于治疗受试者的癌症。

123. 如权利要求66-71中任一项所述的组合物,用于治疗受试者的癌症。

124. 如权利要求20-65中任一项所述的核酸组合物在制备用于治疗受试者的癌症的药物中的用途。

125. 如权利要求66-71中任一项所述的组合物在制备用于治疗受试者的癌症的药物中的用途。

## 用于癌症免疫疗法的CRISPR-CPF1相关方法、组合物和组分

[0001] 优先权要求

[0002] 本申请要求2016年3月4日提交的美国临时申请号62/304,057的优先权,将其内容通过引用以其整体特此并入本文。

[0003] 序列表

[0004] 本说明书参考了序列表(在2017年3月3日以名为“0841770140Sseqlist.txt”的.txt文件电子提交)。所述.txt文件生成于2017年3月2日并且大小为721,000个字节。将序列表的整个内容通过引用特此并入。

### 技术领域

[0005] 本发明涉及用于编辑靶核酸序列或调节靶核酸序列表达的CRISPR/Cpf1相关方法、组合物和组分,及其与包含工程化T细胞或T细胞前体的过继性转移的癌症免疫疗法的结合应用。

### 背景技术

[0006] 基因工程化T细胞的过继性转移已经作为癌症治疗模式进入临床测试。典型地,所述途径由以下步骤组成:1)通过单采术从受试者获得白细胞;2)选择/富集T细胞;3)通过细胞因子处理激活T细胞;4)通过逆转录病毒转导、慢病毒转导或电穿孔引入经克隆的T细胞受体(TCR)基因或嵌合抗原受体(CAR)基因;5)通过细胞因子处理扩增T细胞;6)调理受试者,通常通过淋巴细胞耗减(lymphodepletion);并且7)将工程化T细胞输注到受试者体内。

[0007] 经克隆的TCR基因(TRAC和TRBC)的来源包括从患有具体恶性肿瘤的个体中分离的稀有T细胞群以及从用特定肿瘤抗原或肿瘤细胞免疫的T细胞受体人源化小鼠中分离的T细胞克隆。过继性转移后,TCR工程化T细胞识别肿瘤细胞表面上的由主要组织相容性复合物(MHC)蛋白呈递的其同源抗原肽。抗原接合刺激信号转导通路,引起T细胞激活和增殖。经刺激的T细胞然后使细胞毒性抗肿瘤细胞应答增加,典型地涉及包含颗粒酶B、穿孔素和粒溶素的分泌复合物,从而诱导肿瘤细胞凋亡。

[0008] 嵌合抗原受体(CAR)基因编码包含经由铰链和跨膜结构域与胞质效应结构域融合的胞外肿瘤抗原结合结构域的人工T细胞受体,所述胞外肿瘤抗原结合结构域典型地衍生自单克隆抗体的单链抗体可变片段(scFv)结构域。所述效应结构域典型地衍生自T细胞辅助受体复合物的CD3- $\zeta$ 链,并且还可以包括衍生自CD28和/或CD137受体蛋白的结构域。所述CAR胞外结构域以MHC非依赖性方式结合肿瘤抗原,引起T细胞激活和增殖,以如针对TCR工程化T细胞所描述的细胞毒性抗肿瘤活性告终。

[0009] 迄今为止,已经在工程化T细胞的临床试验中靶向了至少15种不同的肿瘤抗原。在若干试验中,已经报告了抗肿瘤活性。已经在恶性血液病中实现了最大成功。例如,被工程化为靶向B细胞抗原CD19的CAR-T细胞的过继性转移在患有淋巴瘤、急性成淋巴细胞白血病、急性淋巴细胞白血病和B细胞急性淋巴细胞白血病的受试者体内导致了多种部分和完全应答。相比之下,靶向其他肿瘤类型尤其是实体瘤(包括肾细胞癌、成神经细胞瘤、结肠直

肠癌、乳腺癌、卵巢癌、黑色素瘤、肉瘤和前列腺癌)的试验一直不太成功。在许多这些试验中,非常少的患者经历了客观应答。因此,需要改进过继性转移的工程化T细胞的抗肿瘤功效。

## 发明内容

[0010] 本文所披露的方法和组合物提供了使用免疫疗法途径治疗癌症,所述免疫疗法途径包括向受试者给予基因工程化T细胞或T细胞前体。治疗患有癌症的受试者的途径是从所述受试者体内分离T细胞,将它们基因修饰成靶向由癌细胞表达的抗原,然后将它们重新引入所述受试者体内;一种称为过继性T细胞转移的过程。用于基因修饰T细胞的方法包括引入对应地编码跨膜TCR或CAR蛋白的T细胞受体(TCR)或嵌合抗原受体(CAR)基因,所述TCR或CAR蛋白特异性地识别特定癌症抗原。在某些实施例中,肿瘤表达的抗原与TCR或CAR蛋白的抗原结合结构域的接合启动信号传导级联,引起T细胞激活、增殖,并且最终经由细胞毒性免疫应答引起癌细胞的破坏(Kershaw等人,2013NatRevCancer[癌症自然评论]13,525-541)。

[0011] 利用基因修饰T细胞的过继性T细胞转移已经作为实体恶性肿瘤和恶性血液病的治疗剂进入临床测试。迄今为止结果不一。在恶性血液病(尤其是淋巴瘤、慢性淋巴细胞白血病(CLL)和急性淋巴细胞白血病(ALL))中,大多数患者在若干1期和2期试验中至少展现出部分应答,一些患者展现出完全应答(Kochenderfer, J.N.等人,2012Blood[血液]119,2709-2720)。然而,在大部分肿瘤类型(包括黑色素瘤、肾细胞癌和结肠直肠癌)中,观察到了较少的应答(Johnson, L.A.等人,2009Blood[血液]114,535-546; Lamers, C.H.等人,2013Mol. Ther. [分子疗法]21,904-912; Warren, R.S.等人,1998Cancer Gene Ther. [癌症基因疗法]5, S1-S2)。因此,需要改进经修饰T细胞的过继性转移在癌症治疗中的功效。

[0012] 限制基因修饰T细胞作为癌症治疗剂的功效的因素包括(1) T细胞增殖,例如过继性转移后T细胞增殖受限;(2) T细胞存活,例如肿瘤环境中的因子诱导T细胞凋亡;以及(3) T细胞功能,例如由宿主免疫细胞和癌细胞分泌的抑制因子对细胞毒性T细胞功能的抑制。本文所披露的方法和组合物通过修饰影响T细胞增殖、存活和/或功能的T细胞表达的基因的表达而解决了一个或多个这些限制。

[0013] 在某些实施例中,本文所披露的方法和组合物可以用于影响T细胞增殖(例如,通过使抑制T细胞增殖的基因失活)。在某些实施例中,本文所披露的方法和组合物可以用于影响T细胞存活(例如,通过使介导T细胞凋亡的基因失活)。在某些实施例中,本文所披露的方法和组合物可以用于影响T细胞功能(例如,通过使编码免疫抑制和抑制性(例如,无反应性诱导型)信号传导因子的基因失活)。在某些实施例中,本文所披露的方法和组合物可以用于改进T细胞持久性。在某些实施例中,本文所披露的方法和组合物可以单独地或组合地用于影响限制基因修饰T细胞作为癌症治疗剂的功效的因素中的一个或多个,例如T细胞增殖、T细胞存活、T细胞功能、T细胞持久性或其任何组合。

[0014] 本文所披露的方法和组合物可以用于通过改变一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的一个或多个)而影响T细胞增殖、存活、持久性和/或功能。在某些实施例中,本文所披露的方法和组合物可以用于通过改变一个或多个T细胞表达的基因(例如,CBLB和/或PTPN6基因)而影响T细胞增殖。在某

些实施例中,本文所披露的方法和组合物可以用于通过改变一个或多个T细胞表达的基因(例如,FAS和/或BID基因)而影响T细胞存活。在某些实施例中,本文所披露的方法和组合物可以用于通过改变一个或多个T细胞表达的基因(例如,CTLA4、PDCD1、TRAC和/或TRBC基因)而影响T细胞功能。在某些实施例中,本文所披露的方法和组合物可以用于通过改变B2M基因而改进T细胞持久性。

[0015] 在某些实施例中,一个或多个T细胞表达的基因(包括但不限于FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因)作为靶向敲除被独立地靶向,例如以影响T细胞增殖、存活、持久性和/或功能。在某些实施例中,本发明披露的方法包括敲除一个T细胞表达的基因(例如,选自下组的一个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因)。在某些实施例中,本发明披露的方法包括独立地敲除两个T细胞表达的基因(例如,选自下组的两个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因)。在某些实施例中,本发明披露的方法包括独立地敲除三个T细胞表达的基因,例如选自下组的三个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。在某些实施例中,本发明披露的方法包括独立地敲除四个T细胞表达的基因,例如选自下组的四个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。在某些实施例中,本发明披露的方法包括独立地敲除五个T细胞表达的基因,例如选自下组的五个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。在某些实施例中,本发明披露的方法包括独立地敲除六个T细胞表达的基因,例如选自下组的六个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。在某些实施例中,本发明披露的方法包括独立地敲除七个T细胞表达的基因,例如选自下组的七个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。在某些实施例中,本发明披露的方法包括独立地敲除八个T细胞表达的基因,例如FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的每一个。

[0016] 除上文所描述的基因之外,许多其他T细胞表达的基因可以被靶向以影响工程化T细胞的功效。这些基因包括但不限于TGFBR1、TGFBR2和TGFBR3(Kershaw等人2013NatRevCancer[癌症自然评论]13,525-541)。在某些实施例中,可以使用本文所披露的方法单独地或组合地改变TGFBR1、TGFBR2和TGFBR3基因中的一个或多个。在某些实施例中,可以使用本发明披露的方法单独地或与上文所描述的八个基因(即,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因)中的任何一个或多个组合地改变TGFBR1、TGFBR2和TGFBR3基因中的一个或多个。

[0017] 在某些实施例中,本文所披露的方法和组合物改变FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因,这是通过靶向所述一个或多个基因的位置(例如,敲除位置)(例如,非编码区(例如,启动子区)内的位置或编码区内的位置),或通过靶向所述一个或多个基因的转录序列(例如,内含子序列或外显子序列)。在某些实施例中,所述一个或多个基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的编码序列(例如编码区,例如早期编码区)被靶向用于改变和敲除表达。在某些实施例中,所述一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的非编码区(例如,启动子区)中的位置被靶向用于改变和敲除所述一个或多个T细胞表

达的基因的表达。

[0018] 在某些实施例中，本文所披露的方法和组合物改变FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因，这是通过靶向所述一个或多个基因的编码序列。在某些实施例中，所述编码序列是早期编码序列。在某些实施例中，所述一个或多个基因的编码序列被靶向用于敲除所述一个或多个T细胞表达的基因的表达。

[0019] 在某些实施例中，本文所披露的方法和组合物改变FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因，这是通过靶向所述一个或多个基因的非编码序列。在某些实施例中，所述非编码序列包括启动子区内的序列、增强子序列、内含子序列、3'UTR内的序列、多腺苷酸化信号序列或其组合。在某些实施例中，所述一个或多个基因的非编码序列被靶向用于敲除所述一个或多个基因的表达。

[0020] 在某些实施例中，本发明披露的方法包括敲除FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因的一个或两个等位基因，例如通过在所述一个或多个基因中诱导改变。在某些实施例中，所述改变包括插入、缺失、突变或其组合。

[0021] 在某些实施例中，所述靶向敲除途径是使用包含Cpf1酶的CRISPR/Cpf1系统由非同源末端连接(NHEJ)介导。

[0022] 如本文所使用的，“T细胞靶FAS敲除位置”是指FAS基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性FAS基因产物的表达减少或消除(例如，功能性FAS基因产物的表达敲除)的位置。在某些实施例中，所述位置在FAS基因的编码区(例如，早期编码区)中。

[0023] 如本文所使用的，“T细胞靶BID敲除位置”是指BID基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性BID基因产物的表达减少或消除(例如，功能性BID基因产物的表达敲除)的位置。在某些实施例中，所述位置在BID基因的编码区(例如，早期编码区)中。

[0024] 如本文所使用的，“T细胞靶CTLA4敲除位置”是指CTLA4基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性CTLA4基因产物的表达减少或消除(例如，功能性CTLA4基因产物的表达敲除)的位置。在某些实施例中，所述位置在CTLA4的编码区(例如，早期编码区)中。

[0025] 如本文所使用的，“T细胞靶PDCD1敲除位置”是指PDCD1基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性PDCD1基因产物的表达减少或消除(例如，功能性PDCD1基因产物的表达敲除)的位置。在某些实施例中，所述位置在PDCD1基因的编码区(例如，早期编码区)中。

[0026] 如本文所使用的，“T细胞靶CBLB敲除位置”是指CBLB基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性CBLB基因产物的表达减少或消除(例如，功能性CBLB基因产物的表达敲除)的位置。在某些实施例中，所述位置在CBLB基因的编码区(例如，早期编码区)中。

[0027] 如本文所使用的，“T细胞靶PTPN6敲除位置”是指PTPN6基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性PTPN6基因产物的表达减少或消除(例如，功能性PTPN6基因产物的表达敲除)的位置。在某些实施例中，所述位置在PTPN6基因的编码区(例如，早期编码区)中。

[0028] 如本文所使用的，“T细胞靶B2M敲除位置”是指B2M基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性B2M基因产物的表达减少或消除（例如，功能性B2M基因产物的表达敲除）的位置。在某些实施例中，所述位置在B2M基因的编码区（例如，早期编码区）中。如本文所使用的，“T细胞靶TRAC敲除位置”是指TRAC基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性TRAC基因产物的表达减少或消除（例如，功能性TRAC基因产物的表达敲除）的位置。在某些实施例中，所述位置在TRAC基因的编码区（例如，早期编码区）中。

[0029] 如本文所使用的，“T细胞靶TRBC敲除位置”是指TRBC基因中的如果例如通过NHEJ介导的改变被改变的话则导致功能性TRBC基因产物的表达减少或消除（例如，功能性TRBC基因产物的表达敲除）的位置。在某些实施例中，所述位置在TRBC基因的编码区（例如，早期编码区）中。

[0030] 如本文所使用的，“T细胞靶FAS位置”是指如本文所描述的T细胞靶FAS敲除位置中的任一个。

[0031] 如本文所使用的，“T细胞靶BID位置”是指如本文所描述的T细胞靶BID敲除位置中的任一个。

[0032] 如本文所使用的，“T细胞靶CTLA4位置”是指如本文所描述的T细胞靶CTLA4敲除位置中的任一个。

[0033] 如本文所使用的，“T细胞靶PDCD1位置”是指如本文所描述的T细胞靶PDCD1敲除位置中的任一个。

[0034] 如本文所使用的，“T细胞靶CBLB位置”是指如本文所描述的T细胞靶CBLB敲除位置中的任一个。

[0035] 如本文所使用的，“T细胞靶PTPN6位置”是指如本文所描述的T细胞靶PTPN6敲除位置中的任一个。

[0036] 如本文所使用的，“T细胞靶B2M位置”是指如本文所描述的T细胞靶B2M敲除位置中的任一个。

[0037] 如本文所使用的，“T细胞靶TRAC位置”是指如本文所描述的T细胞靶TRAC敲除位置中的任一个。

[0038] 如本文所使用的，“T细胞靶TRBC位置”是指如本文所描述的T细胞靶TRBC敲除位置中的任一个。

[0039] 如本文所使用的，“T细胞靶敲除位置”是指如本文所描述的T细胞靶FAS敲除位置、T细胞靶BID敲除位置、T细胞靶CTLA4敲除位置、T细胞靶PDCD1敲除位置、T细胞靶CBLB敲除位置、T细胞靶PTPN6敲除位置、T细胞靶B2M敲除位置、T细胞靶TRAC敲除位置或T细胞靶TRBC敲除位置中的任一个。

[0040] 如本文所使用的，“T细胞靶位置”是指如本文所描述的T细胞靶敲除位置中的任一个。

[0041] 在一个方面中，本文披露了gRNA分子（例如，分离的或非天然存在的gRNA分子），所述gRNA分子包含与来自选自下组的一个T细胞表达的基因的靶结构域互补的靶向结构域，该组由以下组成：FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0042] 在某些实施例中，所述gRNA分子的靶向结构域被配置成提供足够接近FAS、BID、

CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置(例如,T细胞靶敲除位置)的切割事件(例如,双链断裂),以允许FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置(例如,T细胞靶敲除位置)的改变(例如,与NHEJ相关的改变)。在某些实施例中,所述靶向结构域被配置成使得切割事件(例如,双链)被定位在T细胞靶位置(例如,T细胞靶敲除位置)的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。所述双链断裂可以被定位在FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞敲除靶位置(例如,T细胞靶敲除位置)的上游或下游。

[0043] 在某些实施例中,包含第二靶向结构域的第二gRNA分子被配置成提供足够接近FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置的切割事件(例如,双链断裂),以允许FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置的改变(例如,与NHEJ相关的改变),单独地或与由第一gRNA分子定位的断裂组合。在某些实施例中,所述第一gRNA分子和第二gRNA分子的靶向结构域被配置成使得双链断裂对于所述gRNA分子中的每者而言独立地被定位在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。在某些实施例中,所述两组双链断裂被定位在FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置(例如,T细胞靶敲除位置)的核苷酸的两侧。在某些实施例中,所述两组双链断裂被定位在FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置(例如,T细胞靶敲除位置)的一侧(例如,上游或下游)。在某些实施例中,双链断裂可以伴随有由第二gRNA分子定位的另外的双链断裂,如下文所讨论的。例如,第一gRNA分子的靶向结构域被配置成使得双链断裂被定位在FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置的上游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内;并且第二gRNA分子的靶向结构域被配置成使得双链断裂被定位在FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因中的T细胞靶位置的下游,例如在靶位置的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150、200、250、300、350、400、450或500个核苷酸内。

[0044] 在某些实施例中,当两种或更多种gRNA用于在靶核酸中定位两个或更多个切割事件(例如,两组双链断裂)时,所述两个或更多个切割事件可以由相同或不同Cpf1蛋白产生。在某些实施例中,当两种gRNA用于在靶核酸中定位两组双链断裂时,单Cpf1核酸酶用于产生两个双链断裂。在某些实施例中,当使用两种或更多种Cpf1蛋白时,所述Cpf1蛋白来自不同的物种。

[0045] 当细胞中的超过一个T细胞表达的基因被靶向用于改变时,所靶向的核酸可以被一种或多种Cpf1蛋白改变(例如,切割)。例如,如果两个基因被靶向用于改变,例如两个T细胞表达的基因被靶向用于敲除,则相同或不同的Cpf1蛋白可以用于靶向每个基因。在某些实施例中,两个T细胞表达的基因(或在细胞中所靶向的每个基因)被Cpf1核酸酶切割以产生双链断裂。在某些实施例中,两个T细胞表达的基因(或在细胞中所靶向的每个基因)被Cpf1分子切割以产生双链断裂。在某些实施例中,细胞中的一个或多个T细胞表达的基因可以通过用Cpf1核酸酶切割来改变。当两种或更多种Cpf1蛋白用于切割靶核酸(例如,细胞中

的不同基因)时,所述Cpf1蛋白可以来自不同的细菌物种。例如,细胞中的一个或多个T细胞表达的基因可以通过用来自一种细菌物种的Cpf1蛋白切割来改变,并且相同细胞中的一个或多个T细胞表达的基因可以通过用来自不同细菌物种的Cpf1蛋白切割来改变。在某些实施例中,当使用来自不同物种的两种或更多种Cpf1蛋白时,可以将它们同时递送或顺序地递送,以控制靶核酸中希望位置处的希望基因中的切割的特异性。

[0046] 在某些实施例中,所述第一gRNA分子的靶向结构域和所述第二gRNA分子的靶向结构域与靶核酸分子的相对链互补。在某些实施例中,所述gRNA分子和所述第二gRNA分子被配置成使得PAM朝外定向。

[0047] 在某些实施例中,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因的编码区(例如,早期编码区)中的位置被靶向,例如用于敲除。在某些实施例中,所述靶向结构域包含与选自SEQ ID NO:1-3707的核苷酸序列相同或相差不超过1个、不超过2个、不超过3个、不超过4个或不超过5个核苷酸的序列。在某些实施例中,所述靶向结构域包含选自SEQ ID NO:1-3707的核苷酸序列。

[0048] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的FAS编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:2326-3094。

[0049] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的BID编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:3284-3385。

[0050] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的CTLA4编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:64-370。

[0051] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的PDCD1编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:1-63。

[0052] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的CBLB编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:504-2325。

[0053] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的PTPN6编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:371-503。

[0054] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的B2M编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:3095-3283。

[0055] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的TRAC编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:3386-3588。

[0056] 在某些实施例中,当所述T细胞靶敲除位置是所述靶核酸序列中的TRBC编码区(例如,早期编码区),并且超过一种gRNA用于在所述靶核酸序列中定位断裂(例如,两个双链断

裂)例如以产生一个或多个indel时,每种指导RNA独立地选自SEQ ID NO:3589-3707。

[0057] 在某些实施例中,gRNA进一步包含直接重复结构域。在某些实施例中,所述直接重复结构域在长度上是15-20个核苷酸。在某些实施例中,所述直接重复结构域包含选自下组的核苷酸序列,该组由以下组成:SEQ ID NO:3708-3710。

[0058] 在某些实施例中,所述gRNA分子是单模块化(也称为“单分子”)gRNA。在某些实施例中,所述靶向结构域在长度上是15-25个核苷酸。在某些实施例中,所述靶向结构域在长度上是18个核苷酸。在某些实施例中,所述靶向结构域在长度上是19个核苷酸。在某些实施例中,所述靶向结构域在长度上是20个核苷酸。在某些实施例中,所述靶向结构域在长度上是21个核苷酸。在某些实施例中,所述靶向结构域在长度上是22个核苷酸。在某些实施例中,所述靶向结构域在长度上是23个核苷酸。

[0059] 切割事件(例如,双链断裂)通过Cpf1分子产生。在某些实施例中,所述Cpf1分子催化双链断裂。

[0060] 另外,本发明披露的主题提供了核酸组合物(例如,分离的或非天然存在的核酸组合物,例如DNA组合物),所述核酸组合物包含(a)编码如上文所描述的第一gRNA分子的第一核苷酸序列。在某些实施例中,所述核酸组合物进一步包含(b)编码Cpf1分子的第二核苷酸序列。所述Cpf1分子可以在靶核酸中形成双链断裂。在某些实施例中,所述Cpf1分子选自下组,该组由以下组成:氨基酸球菌属菌株BV3L6Cpf1分子(AsCpf1)、毛螺菌科细菌ND2006Cpf1分子(LbCpf1)和毛螺菌科细菌MA2020(Lb2Cpf1)。在某些实施例中,所述第二核苷酸序列示于SEQ ID NO:3722、SEQ ID NO:3723或SEQ ID NO:3724中。

[0061] 在某些实施例中,所述核酸组合物进一步包含(c)编码第二gRNA分子的第三核苷酸序列,所述第二gRNA分子包含与来自选自下组的一个T细胞表达的基因的靶结构域互补的靶向结构域,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。在某些实施例中,所述第二gRNA与所述第一gRNA分子靶向相同的T细胞靶位置。

[0062] 在某些实施例中,(a)和(b)存在于一个核酸分子(例如一个载体,例如一个病毒载体,例如AAV载体)上。可以在所描述的组合物和方法的任一种中使用的示例性AAV载体包括AAV1载体、经修饰的AAV1载体、AAV2载体、经修饰的AAV2载体、AAV3载体、AAV4载体、经修饰的AAV4载体、AAV5载体、经修饰的AAV5载体、经修饰的AAV3载体、AAV6载体、经修饰的AAV6载体、AAV7载体、经修饰的AAV7载体、AAV8载体、AAV9载体、AAV.rh10载体、经修饰的AAV.rh10载体、AAV.rh32/33载体、经修饰的AAV.rh32/33载体、AAV.rh43载体、经修饰的AAV.rh43载体、AAV.rh64R1载体以及经修饰的AAV.rh64R1载体。在某些实施例中,(a)存在于第一核酸分子(例如第一载体,例如第一病毒载体,例如第一AAV载体)上;并且(b)存在于第二核酸分子(例如第二载体,例如第二载体,例如第二AAV载体)上。所述第一核酸分子和第二核酸分子可以是AAV载体。

[0063] 在某些实施例中,(a)和(c)存在于一个核酸分子(例如一个载体,例如一个病毒载体,例如一个AAV载体)上。在某些实施例中,(a)和(c)在不同载体上。例如,(a)可以存在于第一核酸分子(例如第一载体,例如第一病毒载体,例如第一AAV载体)上;并且(c)可以存在于第二核酸分子(例如第二载体,例如第二载体,例如第二AAV载体)上。在某些实施例中,所述第一核酸分子和第二核酸分子是AAV载体。

[0064] 在某些实施例中,(a)、(b)和(c)存在于一个核酸分子(例如一个载体,例如一个病

毒载体,例如AAV载体)上。在某些实施例中,所述核酸分子是AAV载体。在某些实施例中,(a)、(b)和(c)之一存在于第一核酸分子(例如第一载体,例如第一病毒载体,例如第一AAV载体)上;并且(a)、(b)和(c)中的第二个和第三个在第二核酸分子(例如第二载体,例如第二AAV载体)上编码。所述第一核酸分子和第二核酸分子可以是AAV载体。

[0065] 在某些实施例中,(a)存在于第一核酸分子(例如第一载体,例如第一病毒载体,第一AAV载体)上;并且(b)和(c)存在于第二核酸分子(例如第二载体,例如第二载体,例如第二AAV载体)上。所述第一核酸分子和第二核酸分子可以是AAV载体。

[0066] 在某些实施例中,(b)存在于第一核酸分子(例如第一载体,例如第一病毒载体,例如第一AAV载体)上;并且(a)和(c)存在于第二核酸分子(例如第二载体,例如第二载体,例如第二AAV载体)上。所述第一核酸分子和第二核酸分子可以是AAV载体。

[0067] 在某些实施例中,(c)存在于第一核酸分子(例如第一载体,例如第一病毒载体,例如第一AAV载体)上;并且(b)和(a)存在于第二核酸分子(例如第二载体,例如第二载体,例如第二AAV载体)上。所述第一核酸分子和第二核酸分子可以是AAV载体。

[0068] 在某些实施例中,(a)、(b)和(c)各自存在于不同的核酸分子(例如不同的载体,例如不同的病毒载体,例如不同的AAV载体)上。例如,(a)可以在第一核酸分子上,(b)在第二核酸分子上,并且(c)在第三核酸分子上。所述第一核酸分子、第二核酸分子和第三核酸分子可以是AAV载体。

[0069] 本文所描述的核酸可以包含启动子,所述启动子可操作地连接至编码(a)的gRNA分子的序列,例如本文所描述的启动子。所述核酸可以进一步包含第二启动子,所述第二启动子可操作地连接至编码(c)的第二、第三和/或第四gRNA分子的序列,例如本文所描述的启动子。所述启动子和所述第二启动子彼此不同。在某些实施例中,所述启动子和第二启动子是相同的。本文所描述的核酸可以进一步包含启动子,所述启动子可操作地连接至编码(b)的Cpf1分子的序列,例如本文所描述的启动子。

[0070] 本发明披露的主题还提供了包含(a)如上文所描述的gRNA分子的组合。在某些实施例中,所述组合物进一步包含(b)Cpf1分子,例如如上文所描述的Cpf1分子。在某些实施例中,所述组合物进一步包含(c)如上文所描述的第二gRNA分子。在某些实施例中,如权利要求66-70中任一项所述的组合物,所述组合物是包含Cpf1蛋白和编码所述gRNA分子的核糖核酸分子的核糖核蛋白组合物。

[0071] 本发明披露的主题进一步提供了改变细胞(例如改变细胞的靶核酸的结构,例如改变序列)的方法,所述方法包括使所述细胞与以下接触:(a)如上文所描述的gRNA分子和(b)如上文所描述的Cpf1分子,以及任选地(c)如上文所描述的的第二gRNA分子。在另一个方面中,本文披露了治疗受试者(例如,患有癌症的受试者)(例如改变所述受试者的靶核酸的结构,例如序列)的方法,所述方法包括使所述受试者(或来自所述受试者的细胞)与以下接触:(a)如上文所描述的gRNA;和(b)如上文所描述的Cpf1分子,以及任选地(c)如上文所描述的的第二gRNA分子。

[0072] 在某些实施例中,改变细胞(例如改变细胞的靶核酸的结构,例如改变序列)的方法包括改变选自下组的两个或更多个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0073] 在某些实施例中,改变细胞的方法包括改变选自下组的两个或更多个T细胞表达

的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0074] 在某些实施例中,改变细胞的方法包括改变选自下组的三个或更多个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0075] 在某些实施例中,改变细胞的方法包括改变选自下组的四个或更多个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0076] 在某些实施例中,改变细胞的方法包括改变选自下组的五个或更多个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0077] 在某些实施例中,改变细胞的方法包括改变选自下组的六个或更多个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0078] 在某些实施例中,改变细胞的方法包括改变选自下组的七个或更多个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0079] 在某些实施例中,改变细胞的方法包括改变FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的每一个。

[0080] 在某些实施例中,所述方法包括接触来自患有癌症的受试者的细胞。在某些实施例中,所述癌症选自下组,该组由以下组成:淋巴瘤、慢性淋巴细胞白血病(CLL)、B细胞急性淋巴细胞白血病(B-ALL)、急性成淋巴细胞白血病、急性骨髓性白血病、非霍奇金淋巴瘤(NHL)、弥漫性大细胞淋巴瘤(DLCL)、多发性骨髓瘤、肾细胞癌(RCC)、成神经细胞瘤、结肠直肠癌、乳腺癌、卵巢癌、黑色素瘤、肉瘤、前列腺癌、肺癌、食道癌、肝细胞癌、胰腺癌、星形细胞瘤、间皮瘤、头颈癌以及成神经管细胞瘤。

[0081] 所述细胞可以来自将受益于在一个或多个T细胞表达的基因中的一个或多个T细胞靶位置处具有一个或多个改变的受试者。

[0082] 在某些实施例中,所述细胞是T细胞。接触可以离体地进行,并且接触步骤之后所接触的细胞可以被返回受试者体内。在某些实施例中,所述T细胞是工程化T细胞,例如工程化CAR(嵌合抗原受体)T细胞或工程化TCR(T细胞受体)T细胞。在某些实施例中,所述T细胞被工程化为在T细胞表达的基因的T细胞靶敲除位置内引入改变之前、之后或同时表达TCR或CAR。

[0083] 在某些实施例中,接触步骤包括使所述细胞与如上文所描述的核酸组合物接触。在某些实施例中,接触步骤包括使所述细胞与如上文所描述的组合物接触。在某些实施例中,所述组合物是核糖核蛋白组合物。

[0084] 在某些实施例中,接触包括使所述细胞与核酸分子(例如载体,例如AAV载体、AAV1载体、经修饰的AAV1载体、AAV2载体、经修饰的AAV2载体、AAV3载体、经修饰的AAV3载体、AAV4载体、经修饰的AAV4载体、AAV5载体、经修饰的AAV5载体、AAV6载体、经修饰的AAV6载体、AAV7载体、经修饰的AAV7载体、AAV8载体、AAV9载体、AAV.rh10载体、经修饰的AAV.rh10载体、AAV.rh32/33载体、经修饰的AAV.rh32/33载体、AAV.rh43载体、经修饰的AAV.rh43载体、AAV.rh64R1载体或经修饰的AAV.rh64R1载体)接触。

[0085] 在某些实施例中,接触包括向所述细胞递送作为蛋白质或mRNA的(b)的Cpf1分子以及编码(a)和任选地(c)的核酸。

[0086] 在某些实施例中,接触包括向所述细胞递送作为蛋白质或mRNA的(b)的Cpf1分子、作为RNA的(a)的gRNA以及任选地作为RNA的(c)的第二gRNA。

[0087] 在某些实施例中,接触包括向所述细胞递送作为RNA的(a)的gRNA、任选地作为RNA的(c)的第二gRNA以及编码(b)的Cpf1分子的核酸组合物。

[0088] 本发明披露的主题进一步提供了反应混合物,所述反应混合物包含如上文所描述的gRNA分子、如上文所描述的核酸组合物或如上文所描述的组合物以及细胞(例如,来自将受益于一个或多个T细胞表达的基因中的一个或多个T细胞靶位置处的一个或多个改变的受试者的细胞)。

[0089] 本发明披露的主题进一步提供了试剂盒,所述试剂盒包含(a)如上文所描述的gRNA分子或编码所述gRNA的核酸组合物,以及以下中的一种或多种:(b)如上文所描述的Cpf1分子;(c)如上文所描述的第二gRNA分子。

[0090] 本发明披露的主题进一步提供了经修饰的gRNA分子。在某些实施例中,所述经修饰的gRNA分子在其5'端或其附近包含修饰。在某些实施例中,所述gRNA分子在其3'端或其附近包含修饰。在某些实施例中,所述gRNA分子在其5'端或其附近包含修饰并且在其3'端或其附近包含修饰。在某些实施例中,所述修饰在其5'端的1-10、1-9、1-8、1-7、1-6、1-5、1-4、1-3或1-2个核苷酸内。在某些实施例中,所述修饰在其3'端的1-10、1-9、1-8、1-7、1-6、1-5、1-4、1-3或1-2个核苷酸内。在某些实施例中,所述修饰使得所述gRNA分子当被引入T细胞中时展现出增加的对于核酸酶的稳定性。在某些实施例中,所述修饰使得所述gRNA分子当被引入T细胞中时展现出降低的先天性免疫应答。在某些实施例中,所述先天性免疫应答涉及细胞因子表达的诱导。

[0091] 另外,本发明披露的主题提供了如上文所描述的gRNA分子,用于治疗受试者的癌症。在某些实施例中,所述gRNA分子与(b)Cpf1分子组合使用。

[0092] 本发明披露的主题进一步提供了如上文所描述的gRNA分子在制备用于治疗受试者的癌症的药物中的用途。在某些实施例中,所述药物进一步包含(b)Cpf1分子。

[0093] 本发明披露的主题进一步提供了如上文所描述的核酸组合物,用于治疗受试者的癌症。

[0094] 本发明披露的主题进一步提供了如上文所描述的组合物,用于治疗受试者的癌症。

[0095] 本发明披露的主题进一步提供了如上文所描述的核酸组合物在制备用于治疗受试者的癌症的药物中的用途。

[0096] 本发明披露的主题进一步提供了如上文所描述的组合物在制备用于治疗受试者的癌症的药物中的用途。

[0097] 除非另外定义,本文所使用的所有技术术语和科学术语具有与本发明所属领域的普通技术人员通常所理解的相同的含义。尽管可以在本发明的实践或测试中使用类似于或等效于本文描述的那些的方法和材料,但是下文描述了适合的方法和材料。本文提及的所有出版物、专利申请、专利和其他参考文献通过引用以其整体而并入。另外,所述材料、方法和实例仅是说明性的并不旨在是限制性的。

[0098] 标题(包括数字和字母标题和副标题)用于组织和呈现并不旨在是限制性的。

[0099] 本发明的其他特征和优势从具体实施方式、附图以及从权利要求书将变得显而易见。

## 附图说明

[0100] 图1描绘了根据本发明披露的主题的某些非限制性实施例的gRNA分子的结构。

[0101] 图2描绘了在生化切割测定中Cpf1RNP的评估。在体外切割测定中评估靶向TRAC基因座的若干RNP的活性。将对应于450bp的TRAC外显子1的PCR产物与所鉴定的TRAC RNP以1:1比率孵育。针对每种TRAC RNP列出了应产生成功切割的条带的大致预期的大小。虽然所有RNP似乎都有活性,但两种RNP看起来更具活性-GWED545和GWED546。

[0102] 图3A和3B描绘了用Cpf1TRAC特异性RNP处理的CD4<sup>+</sup>T细胞中TCR $\alpha$ / $\beta$ 表达的分析。(A) 将激活的人CD4<sup>+</sup>T细胞用被设计为靶向TRAC基因座的RNP电穿孔。在电穿孔后第4天,将细胞用TCR $\alpha$ / $\beta$ 抗体染色并通过FCM分析。针对测试的每种crRNA绘制TCR $\alpha$ / $\beta$ 阴性细胞的频率。用两种RNP (GWED545和GWED546) 处理产生显著频率的TCR $\alpha$ / $\beta$ 阴性细胞,表明这些RNP成功编辑TRAC基因座的能力导致表面蛋白表达的降低。(B) 针对与Cpf1apo对照和在此测定中未能缺失表面TCR $\alpha$ / $\beta$ 细胞的RNP相比产生显著的TCR $\alpha$ / $\beta$ 阴性细胞的两种RNP绘制的代表性FCM。

[0103] 图4描绘了Cpf1RNP的活力。用Cpf1RNP处理细胞不会导致活力损失。针对每种RNP通过前向散射对侧向散射的位置确定活淋巴细胞的频率。针对测试的每种RNP绘制活淋巴细胞的频率。

[0104] 图5描绘了T7E1对Cpf1编辑人T细胞的能力的分子分析。结果证实Cpf1可以编辑人T细胞。在电穿孔后第4天收获来自RNP处理的人T细胞的gDNA。使用特异性引物扩增TRAC基因座,并对PCR产物进行T7E1测定。简言之,使PCR产物变性,再退火,并且最后用T7E1酶处理。此酶切割在错配位点处错配的双链DNA。如通过琼脂糖凝胶上的切割产物的定量评估的T7E1酶的切割率与所靶向的基因座处的基因组编辑的百分比相关。绘制来自此测定的数据并且所述数据支持在图3中通过FACS观察到的数据。

[0105] 图6描绘了来自第二供体的用Cpf1TRAC特异性RNP处理的CD4<sup>+</sup>T细胞中TCR $\alpha$ / $\beta$ 表达的分析。将激活的人CD4<sup>+</sup>T细胞用对应于GWED545和GWED546的RNP电穿孔。在电穿孔后第4天,将细胞用TCR $\alpha$ / $\beta$ 抗体染色并通过FCM分析。绘制TCR $\alpha$ / $\beta$ 阴性细胞的频率。用Cpf1RNP GWED545和GWED546处理的细胞中TCR $\alpha$ / $\beta$ 的损失证明Cpf1可以在多个供体中可重复地编辑人T细胞。

[0106] 图7描绘了真核mRNA帽结构。

## 具体实施方式

### [0107] 定义

[0108] 如本文所使用的,术语“约”或“大约”意指在如本领域普通技术人员确定的特定值的可接受误差范围内,这将部分地取决于如何测量或确定所述值,即测量系统的限制。例如,根据本领域中的实践,“约”可以意指在3个或超过3个标准差内。可替代地,“约”可以意指给定值的至多20%、优选至多10%、更优选至多5%并且仍更优选至多1%的范围。可替代地,特别是对于生物系统或过程,所述术语可以意指在值的一个数量级内、优选地在5倍内并且更优选地在2倍内。

[0109] 如本文所使用的,“结构域”用于描述蛋白质或核酸的区段。除非另外指明,结构域不需要具有任何特定功能特性。

[0110] 如下进行两个序列之间的同源性或序列一致性(所述术语在本文可互换地使用)的计算。将所述序列进行比对用于最优比较的目的(例如,用于最优比对,可以在第一和第二氨基酸或核酸序列中的一个或两个中引入空位,并且出于比较的目的,可以不考虑非同源序列)。使用具有Blossum 62打分矩阵(其中空位罚分为12,空位延伸罚分为4,并且移码空位罚分为5)的GCG软件包中的GAP程序,将最优比对确定为最佳评分。然后比较相应的氨基酸位置或核苷酸位置处的氨基酸残基或核苷。当第一序列中的位置被与在第二序列中的相应位置相同的氨基酸残基或核苷酸占据时,则所述分子在那个位置是相同的。两个序列之间的百分比一致性是由所述序列共享的相同位置的数目的函数。

[0111] 如本文所使用的,“抑制性Cpf1gRNA分子”是指包含与核酸上的靶结构域互补的靶向结构域的gRNA分子,所述核酸包含编码被引入细胞或受试者中的CRISPR/Cpf1系统的组分的序列。抑制性Cpf1gRNA不靶向内源细胞或受试者序列。在某些实施例中,抑制性Cpf1gRNA分子包含与以下上的靶序列互补的靶向结构域:(a) 编码Cpf1分子的核酸分子;(b) 编码gRNA的核酸分子,所述gRNA包含靶向FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因的靶向结构域(靶基因gRNA);或编码CRISPR/Cpf1组分的超过一种核酸分子上,例如(a)和(b)两者。在某些实施例中,编码CRISPR/Cpf1组分(例如,编码Cpf1分子或靶基因gRNA)的核酸分子包含超过一个与抑制性Cpf1gRNA靶向结构域互补的靶结构域。在某些实施例中,抑制性Cpf1gRNA分子与Cpf1分子复合,并且例如通过切割或通过所述核酸分子结合而导致Cpf1介导的所靶向的核酸分子失活,并且导致CRISPR/Cpf1系统组分的产生停止或减少。在某些实施例中,所述Cpf1分子形成两种复合物:包含Cpf1分子与靶基因gRNA的复合物,所述复合物改变FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因;以及包含Cpf1分子与抑制性Cpf1gRNA分子的复合物,所述复合物发挥阻止CRISPR/Cpf1系统组分(例如,Cpf1分子或靶基因gRNA分子)的进一步产生的作用。在某些实施例中,抑制性Cpf1gRNA分子/Cpf1分子复合物与控制区序列(例如,启动子)结合或促进其切割,所述控制区序列可操作地连接至编码Cpf1分子的序列,即编码所述Cpf1分子的转录区、外显子或内含子的序列。在某些实施例中,抑制性Cpf1gRNA分子/Cpf1分子复合物与控制区序列(例如,启动子)结合或促进其切割,所述控制区序列可操作地连接至gRNA分子或编码所述gRNA分子的序列。在某些实施例中,所述抑制性Cpf1gRNA(例如,靶向Cpf1的抑制性Cpf1gRNA分子或靶向靶基因gRNA的抑制性Cpf1gRNA分子)限制所述Cpf1分子/靶基因gRNA分子复合物介导的基因靶向的作用。在某些实施例中,抑制性Cpf1gRNA对所述Cpf1分子/靶基因gRNA分子复合物的活性施加时间限制、表达水平限制或其他限制。在某些实施例中,抑制性Cpf1gRNA减少脱靶或其他不想要的活性。在某些实施例中,抑制性Cpf1gRNA分子抑制(例如,完全或基本上完全抑制)所述Cpf1系统的组分的产生并且由此限制其活性。

[0112] 如本文所使用的,“调节剂”是指可改变受试分子或遗传序列的活性(例如,酶活性、转录活性或翻译活性)、量、分布或结构的实体(例如,药物)。在某些实施例中,调节包括切割,例如,共价或非共价键的断裂或者共价或非共价键的形成,例如将部分附接至受试分子。在某些实施例中,调节剂改变受试分子的三维、二级、三级或四级结构。调节剂可以增加、降低、引发或消除受试活性。

[0113] 如本文所使用的,“大分子”是指具有至少2、3、5、10、20、30、40、50、60、70、80、90或100kD的分子量的分子。大分子包括蛋白质、多肽、核酸、生物制剂以及碳水化合物。

[0114] 如本文所使用的,“多肽”是指具有少于100个氨基酸残基的氨基酸的聚合物。在某些实施例中,它具有少于50、20或10个氨基酸残基。

[0115] 如本文所使用的,“非同源末端连接”或“NHEJ”是指连接介导的修复和/或非模板介导的修复,包括例如典型NHEJ(cNHEJ)、替代NHEJ(altnHEJ)、微同源性介导的末端连接(MMEJ)以及合成依赖性微同源性介导的末端连接(SD-MMEJ)。

[0116] 如本文所使用的,“参比分子”(例如,参比Cpf1分子或参比gRNA)是指与受试分子(例如,受试Cpf1分子或受试gRNA分子(例如,经修饰的或候选Cpf1分子))进行比较的分子。例如,可以将Cpf1分子表征为具有不超过参比Cpf1分子的核酸酶活性的10%。参比Cpf1分子的实例包括天然存在的未经修饰的Cpf1分子,例如天然存在的Cpf1分子,如氨基酸球菌属(例如,菌株BV3L6(“AsCpf1”)或毛螺菌科细菌(菌株ND2006(“LbCpf1”)或菌株MA2020(“Lb2Cpf1”))的Cpf1分子,如Zetsche等人,Cell[细胞](2015);163:759-771中所描述的。在某些实施例中,所述参比Cpf1分子是具有和与其进行比较的Cpf1分子最接近序列一致性或同源性的天然存在的Cpf1分子。在某些实施例中,所述参比Cpf1分子是如下序列(例如,天然存在的或已知的序列),所述序列是其上已经发生改变(例如,突变)的亲本形式。

[0117] 如本文关于分子的修饰所使用的“置换”或“置换的”不需要方法限制,但仅表明置换实体是存在的。

[0118] 如本文所使用的,“小分子”是指具有小于约2kD(例如,小于约2kD、小于约1.5kD、小于约1kD或小于约0.75kD)的分子量的化合物。

[0119] 如本文所使用的,“受试者”可以意指人或非人动物。所述术语包括但不限于哺乳动物(例如,人、其他灵长类动物、猪、啮齿动物(例如,小鼠和大鼠或仓鼠)、兔、豚鼠、奶牛、马、猫、狗、绵羊以及山羊)。在某些实施例中,所述受试者是人。在某些实施例中,所述受试者是家禽。

[0120] 如本文所使用的,“治疗(treat、treating和treatment)”意指,治疗哺乳动物(例如,人)的疾病,包括(a)抑制所述疾病,即抑制或防止其发展;(b)缓解所述疾病,即导致疾病状态的消退;以及(c)治愈所述疾病。

[0121] 如本文所使用的在氨基酸序列的语境下的“X”是指任何氨基酸(例如,二十种天然氨基酸中的任一种),除非另外说明。

[0122] 改进癌症免疫疗法

[0123] 在一个方面中,本文所披露的组合物和方法可以用于通过改变一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)来影响工程化T细胞的增殖。为了使工程化T细胞增加有效的抗肿瘤应答,它们需要:1)转移到所述受试者体内后充分地增殖,以提供足够数量的靶向肿瘤的特异性T细胞;2)在所述受试者体内存活足以维持所需抗肿瘤活性的时间长度;并且3)逃避由免疫细胞、肿瘤细胞和肿瘤环境中的其他细胞产生的抑制因子的影响,这样使得所述工程化T细胞维持功能性抗肿瘤表型。增殖和/或存活不足以及对抑制因子的敏感性可以导致工程化T细胞在患有癌症的受试者体内缺乏功效。本文所披露的方法和组合物解决了这些问题,以便改进工程化T细胞作为癌症治疗模式的功效。

[0124] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变CBLB基因而影响工程化T细胞的增殖。在某些实施例中,卡西塔斯(Casitas)B系淋巴瘤b蛋白(由CBLB编码)

的表达减少或不存在减少对外源白细胞介素信号传导的需要,以在递送到所述受试者体内后促进工程化T细胞的增殖 (Stromnes, I.M. 等人, 2010 *J. Clin. Invest.* [临床研究杂志] 120, 3722-3734)。

[0125] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变PTPN6基因而影响工程化T细胞的增殖。在某些实施例中,含Src同源区2结构域的磷酸酶-1蛋白(由PTPN6编码)的表达减少或不存在导致经转移的T细胞的短期积累增加以及随后的抗肿瘤活性改进 (Stromnes, I.M. 等人, 2012 *J. Immunol.* [免疫学杂志] 189, 1812-1825)。

[0126] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变FAS基因而影响工程化T细胞的增殖。在某些实施例中,Fas蛋白的表达减少或不存在将抑制Fas配体对T细胞凋亡的诱导;所述Fas配体是一种由许多癌症类型表达的因子 (Dotti, G. 等人, 2005 *Blood* [血液] 105, 4677-4684)。

[0127] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变BID基因而影响工程化T细胞的增殖。在某些实施例中,Bid蛋白的表达减少或不存在阻止Fas通路激活后T细胞凋亡的诱导 (Lei, X.Y. 等人, 2009 *Immunol. Lett.* [免疫学快报] 122, 30-36)。

[0128] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变CTLA4基因而降低免疫抑制因子对工程化T细胞的作用。在某些实施例中,细胞毒性T淋巴细胞相关抗原4(由CTLA4编码)的表达减少或不存在消除了结合肿瘤环境中的由抗原呈递细胞表达的CD80或CD86后无响应状态(“无反应性”)的诱导 (Shrikant, P. 等人, 1999 *Immunity* [免疫] 11, 483-493)。

[0129] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变PDCD1基因而降低免疫抑制因子对工程化T细胞的作用。在某些实施例中,编程性细胞死亡蛋白1(由PDCD1编码)的表达减少或不存在通过接合由肿瘤细胞或肿瘤环境中的细胞表达的PD1配体而阻止T细胞凋亡的诱导 (Topalian, S.L. 等人, 2012 *N. Engl. J. Med.* [新英格兰医学杂志] 366, 2443-2454)。

[0130] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变TRAC和/或TRBC基因而改进T细胞特异性和安全性。在某些实施例中,T细胞受体(由TRAC和TRBC编码)的表达减少或不存在通过消除宿主组织的T细胞受体识别和对宿主组织的应答而预防移植物抗宿主疾病。因此,这种途径可以用于产生“现成的”T细胞 (Torikai 等人, 2012 *Blood* [血液] 119, 5697-5705)。在某些实施例中,TRAC和/或TRBC基因的表达减少或不存在减少或消除内源T细胞受体与外源引入的工程化T细胞受体的错配,从而改进治疗功效 (Provasi 等人, 2012, *Nature Medicine* [自然医学] 18, 807-815)。

[0131] 在某些实施例中,本文所披露的组合物和方法可以用于通过改变B2M基因而改进T细胞持久性。在某些实施例中, $\beta$ -2-微球蛋白(由B2M编码)的表达减少或不存在导致T细胞上MHC I表面表达的丧失。由于衍生自所引入的CAR或工程化TCR的“外源”肽可以在所述工程化T细胞的表面上由I类MHC呈递,在某些实施例中,去除 $\beta$ -2-微球蛋白可以降低宿主排斥所输入的工程化T细胞的可能性,产生用于过继免疫疗法的低免疫原性细胞 (Mandal 等人, *Cell Stem Cell* [细胞与干细胞], 2014)。在某些实施例中, $\beta$ -2-微球蛋白的表达减少或不存在用于产生用于同种异体移植的“现成的”工程化T细胞 (Riolobos 等人 *Mol. Ther.* [分子治疗学], 2013)。

[0132] 在某些实施例中,本文所披露的组合物和方法可以用于降低选自下组的一个或多个基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因,以改进使用工程化T细胞的癌症免疫疗法的治疗。

[0133] 本文披露了使用本文所描述的组合物和方法经由免疫疗法治疗癌症的途径。

[0134] 在一种途径中,选自下组的一个或多个基因作为靶向敲除被靶向,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因,例如以影响T细胞增殖、存活、功能和/或持久性。在某些实施例中,所述途径包括敲除选自下组的一个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。在某些实施例中,所述途径包括敲除选自下组的两个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因。在某些实施例中,所述途径包括敲除选自下组的三个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因。在某些实施例中,所述途径包括敲除选自下组的四个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因。在某些实施例中,所述途径包括敲除选自下组的五个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因。在某些实施例中,所述途径包括敲除选自下组的六个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因。在某些实施例中,所述途径包括敲除选自下组的七个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因。在某些实施例中,所述途径包括敲除选自下组的八个T细胞表达的基因,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因。

[0135] 在某些实施例中,所述方法包括疾病发作之后启动对受试者的治疗。在某些实施例中,所述方法包括远远滞后于疾病发作启动对受试者的治疗,例如癌症发作之后1、2、3、4、5、6、7、8、9、10、12、24或36个月。

[0136] 在某些实施例中,所述方法包括在疾病的晚期启动对受试者的治疗。

[0137] 总体上,在疾病的所有阶段启动对受试者的治疗预期对受试者有益。

[0138] 可以使用本文所披露的组合物和方法治疗的癌症包括血液癌症和实体瘤。例如,可以使用本文所披露的组合物和方法治疗的癌症包括但不限于淋巴瘤、慢性淋巴细胞白血病(CLL)、B细胞急性淋巴细胞白血病(B-ALL)、急性成淋巴细胞白血病、急性骨髓性白血病、非霍奇金淋巴瘤(NHL)、弥漫性大细胞淋巴瘤(DEL)、多发性骨髓瘤、肾细胞癌(RCC)、成神经细胞瘤、结肠直肠癌、乳腺癌、卵巢癌、黑色素瘤、肉瘤、前列腺癌、肺癌、食道癌、肝细胞癌、胰腺癌、星形细胞瘤、间皮瘤、头颈癌以及成神经管细胞瘤。

[0139] 改变一个或多个T细胞表达的基因的方法

[0140] 如本文所披露的,可以例如使用如本文所描述的CRISPR-Cpf1介导的方法通过基因编辑靶向(例如,改变)一个或多个T细胞表达的基因(例如,选自下组的一个或多个,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因)。

[0141] 本文所披露的方法和组合物提供了靶向(例如,改变)一个或多个T细胞表达的基因(例如,选自下组的一个或多个,该组由以下组成:FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因)中的T细胞靶位置。可以例如使用CRISPR-Cpf1介导的方法通过基因编辑靶向(例如,改变)T细胞靶位置,以靶向(例如改变)一个或多个T细胞表达的基因(例

如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)。

[0142] 本文披露了用于靶向(例如,改变)一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)中的T细胞靶位置的方法。

[0143] 例如通过以下方式实现靶向(例如,改变)T细胞靶位置:敲除一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因):

[0144] (a) 非常接近一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因的编码区(例如,早期编码区)或在所述编码区(例如,早期编码区)内的一个或多个核苷酸的插入或缺失(例如,NHEJ介导的插入或缺失),或者

[0145] (b) 包括一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的至少一部分的基因组序列的缺失(例如,NHEJ介导的缺失)。所有途径都造成靶向(例如,改变)一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)。

[0146] 在某些实施例中,本文所描述的方法在一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的至少一个等位基因中的编码区附近引入一个或多个断裂。在某些实施例中,本文所描述的方法侧翼于一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的至少一部分引入两个或更多个断裂。所述两个或更多个断裂去除(例如,缺失)包括一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的至少一部分的基因组序列。本文所描述的所有方法都导致靶向(例如,改变)一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)。

[0147] 一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的靶向(例如,改变)可以由任何机制介导。可以与一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的改变相关的示范性机制包括但不限于非同源末端连接(例如,经典的或替代的)、微同源性介导的末端连接(MMEJ)、同源定向修复(例如,内源供体模板介导的)以及SDSA(合成依赖性链退火)。

[0148] 通过引入插入或缺失敲除一个或多个T细胞表达的基因

[0149] 在某些实施例中,所述方法包括非常接近一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的T细胞靶敲除位置(例如,早期编码区)引入一个或多个核苷酸的插入或缺失。如本文所描述的,在某些实施例中,所述方法包括向一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的T细胞靶敲除位置例如编码区(例如早期编码区,例如在离起始密码子500bp或剩余的编码序列内,例如在离起始密码子的前500bp的下游)引入一个或多个断裂(例如,双链断裂)。在某些实施例中,NHEJ介导的所述一个或多个断裂的修复允许NHEJ介导的在T细胞靶敲除位置内非常接近地引入indel。

[0150] 在某些实施例中,在一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)中的T细胞靶敲除位置处或非常接近所述位置引入双链断裂(例如,由一种gRNA分子定位)。在某些实施例中,单gRNA分子(例如,与Cpf1核酸酶)用于在T细胞靶敲除位置例如编码区(例如早期编码区,例如在离起始密码子500bp或剩余的编码序列内,例如在离起始密码子的前500bp的下游)处或非常接近所述位置产生双链

断裂。在某些实施例中,所述断裂被定位成避免不想要的靶染色体元件(如重复元件,例如Alu重复)。

[0151] 在某些实施例中,在一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)中的T细胞靶敲除位置处或非常接近所述位置引入两组断裂(例如,两个双链断裂)(例如,由两种gRNA分子定位)。在某些实施例中,两种gRNA分子(例如,与一种或两种Cpf1核酸酶)用于侧翼于T细胞靶敲除位置例如编码区(例如早期编码区,例如在离起始密码子500bp或剩余的编码序列内,例如在离起始密码子的前500bp的下游)产生两个双链断裂。在某些实施例中,所述gRNA分子被配置成使得两组断裂均被定位在T细胞靶敲除位置的上游或下游。在某些实施例中,所述gRNA分子被配置成使得一组断裂被定位在T细胞靶敲除位置的上游并且第二组断裂被定位在T细胞靶敲除位置的下游。在某些实施例中,所述断裂被定位成避免不想要的靶染色体元件(如重复元件,例如Alu重复)。

[0152] 在某些实施例中,两种或更多种(例如,三种或四种)gRNA分子与一种Cpf1分子一起使用。在某些实施例中,当两种或更多种(例如,三种或四种)gRNA与两种或更多种Cpf1分子一起使用时,至少一种Cpf1分子来自与其他一种或多种Cpf1分子不同的物种。例如,当两种gRNA分子与两种Cpf1分子一起使用时,一种Cpf1分子可以来自一种物种并且另一种Cpf1分子可以来自不同的物种。根据需要,两种Cpf1种类均用于产生双链断裂。

[0153] 当细胞中的超过一个基因被靶向用于改变时,所靶向的核酸可以被一种或多种Cpf1蛋白(例如,Cpf1核酸酶)改变(例如,切割)。例如,如果两个基因被靶向用于改变,例如两个基因被靶向用于敲除,则相同或不同的Cpf1蛋白可以用于靶向每个基因。在某些实施例中,两个基因(或在细胞中所靶向的每个基因)被Cpf1核酸酶切割以产生双链断裂。在某些实施例中,两个基因(或在细胞中所靶向的每个基因)被Cpf1核酸酶切割以产生双链断裂。在某些实施例中,细胞中的一个或多个基因可以通过用Cpf1核酸酶切割来改变。当两种或更多种Cpf1蛋白用于切割靶核酸(例如,细胞中的不同基因)时,所述Cpf1蛋白可以来自不同的细菌物种。例如,细胞中的一个或多个基因可以通过用来自一种细菌物种的Cpf1蛋白切割来改变,并且相同细胞中的一个或多个基因可以通过用来自不同细菌物种的Cpf1蛋白切割来改变。在某些实施例中,当使用来自不同物种的两种或更多种Cpf1蛋白时,可以将它们同时递送或顺序地递送,以控制靶核酸中希望位置处的希望基因中的切割的特异性。

[0154] 在某些实施例中,所述第一gRNA分子的靶向结构域和所述第二gRNA分子的靶向结构域与靶核酸分子的相对链互补。在某些实施例中,所述gRNA分子和所述第二gRNA分子被配置成使得PAM朝外定向。

[0155] 通过缺失(例如,NHEJ介导的缺失)包括一个或多个T细胞表达的基因的至少一部分的基因组序列敲除一个或多个T细胞表达的基因

[0156] 在某些实施例中,所述方法包括引入包含一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的至少一部分的基因组序列的缺失。如本文所描述的,在某些实施例中,所述方法包括引入两个双链断裂—一个在(即,侧翼于)T细胞靶敲除位置的5' 并且另一个在(即,侧翼于)T细胞靶敲除位置的3'。在某些实施例中,两种gRNA(例如,单分子的(或嵌合的)或模块化的gRNA分子)被配置成将两组断裂(例如,两个双链断裂)定位在一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、

CBLB、PTPN6、B2M、TRAC和/或TRBC基因)中的T细胞靶敲除位置的相对侧上。

[0157] 在某些实施例中,所述方法包括缺失(例如,NHEJ介导的缺失)包括一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的至少一部分的基因组序列。如本文所描述的,在某些实施例中,所述方法包括侧翼于一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)中的区域(例如,编码区(例如,早期编码区)或非编码区(例如,启动子区、增强子区、内含子、3' UTR和/或多腺苷酸化信号序列))引入两组断裂(例如,一对双链断裂)。在某些实施例中,NHEJ介导的所述一个或多个断裂的修复允许改变一个或多个如本文所描述的T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因),这减少或消除所述基因的表达,例如以敲除一个或多个T细胞表达的基因的一个或两个等位基因。

[0158] 在某些实施例中,在一个或多个T细胞表达的基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)中的T细胞靶敲除位置处或非常接近所述位置引入两组断裂(例如,两个双链断裂)(例如,由两种gRNA分子定位)。在某些实施例中,两种gRNA分子(例如,与一种或两种Cpf1核酸酶)用于侧翼于T细胞靶敲除位置产生两组断裂,例如所述gRNA分子被配置成使得一组断裂被定位在T细胞靶敲除位置的上游并且第二组断裂被定位在T细胞靶敲除位置的下游。在某些实施例中,所述断裂被定位成避免不想要的靶染色体元件(如重复元件,例如Alu重复)。

[0159] 在某些实施例中,两种或更多种(例如,三种或四种)gRNA分子与一种Cpf1分子一起使用。在某些实施例中,当两种或更多种(例如,三种或四种)gRNA与两种或更多种Cpf1分子一起使用时,至少一种Cpf1分子来自与其他一种或多种Cpf1分子不同的物种。例如,当两种gRNA分子与两种Cpf1分子一起使用时,一种Cpf1分子可以来自一种物种并且另一种Cpf1分子可以来自不同的物种。根据需要,两种Cpf1种类均用于产生双链断裂。

[0160] 当细胞中的超过一个基因被靶向用于改变时,所靶向的核酸可以被一种或多种Cpf1蛋白(例如,Cpf1核酸酶)改变(例如,切割)。例如,如果两个基因被靶向用于改变,例如两个基因被靶向用于敲除,则相同或不同的Cpf1蛋白可以用于靶向每个基因。在某些实施例中,两个基因(或在细胞中所靶向的每个基因)被Cpf1核酸酶切割以产生双链断裂。在某些实施例中,两个基因(或在细胞中所靶向的每个基因)被Cpf1核酸酶切割以产生双链断裂。在某些实施例中,细胞中的一个或多个基因可以通过用Cpf1核酸酶切割来改变。当两种或更多种Cpf1蛋白用于切割靶核酸(例如,细胞中的不同基因)时,所述Cpf1蛋白可以来自不同的细菌物种。例如,细胞中的一个或多个基因可以通过用来自一种细菌物种的Cpf1蛋白切割来改变,并且相同细胞中的一个或多个基因可以通过用来自不同细菌物种的Cpf1蛋白切割来改变。在某些实施例中,当使用来自不同物种的两种或更多种Cpf1蛋白时,可以将它们同时递送或顺序地递送,以控制靶核酸中希望位置处的希望基因中的切割的特异性。

[0161] 在某些实施例中,所述第一gRNA分子的靶向结构域和所述第二gRNA分子的靶向结构域与靶核酸分子的相对链互补。在某些实施例中,所述gRNA分子和所述第二gRNA分子被配置成使得PAM朝外定向。

[0162] 在某些实施例中,基因工程化T细胞的过继性转移可以提供癌症的潜在治疗。将编码细胞表面受体的基因插入所述T细胞中。所述基因工程化T细胞能够检测肿瘤相关抗原,所述肿瘤相关抗原可以用于区别肿瘤细胞与大多数正常组织。

[0163] 可以在疾病发作之后,但优选在病程的早期进行靶基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因)的一个或两个等位基因的敲除。

[0164] I. gRNA分子

[0165] 如该术语在本文使用的,gRNA分子是指促进gRNA分子/Cpf1分子复合物向靶核酸特异性靶向或归巢的核酸。gRNA分子可以是单模块化的(具有单RNA分子,也称为“单分子的”,例如嵌合gRNA)或模块化的(包含超过一种并且典型地两种分开的RNA分子)。在某些实施例中,所述gRNA分子是单模块化gRNA。

[0166] 在某些实施例中,gRNA分子从5'到3'包含:直接重复结构域和靶向结构域,如图1所示。在某些实施例中,所述gRNA分子在长度上是15至100(例如,15至30、30至50、50至70、70至80或80至100)个核苷酸。在某些实施例中,所述gRNA分子在长度上是30至50(例如,30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49或50)个核苷酸。在某些实施例中,所述gRNA分子在长度上是38个核苷酸。在某些实施例中,所述gRNA分子在长度上是39个核苷酸。在某些实施例中,所述gRNA分子在长度上是40个核苷酸。在某些实施例中,所述gRNA分子在长度上是41个核苷酸。在某些实施例中,所述gRNA分子在长度上是42个核苷酸。在某些实施例中,所述gRNA分子在长度上是43个核苷酸。在某些实施例中,所述gRNA分子在长度上是44个核苷酸。在某些实施例中,所述gRNA分子在长度上是45个核苷酸。

[0167] 直接重复结构域

[0168] 在某些实施例中,所述直接重复结构域在长度上是10至30(例如,10至15、15至20、20至25或20至30)个核苷酸。在某些实施例中,所述直接重复结构域在长度上是15至25(例如,15至20或20至25)个核苷酸。在某些实施例中,所述直接重复结构域在长度上是15至20(例如,15、16、17、18、19或20)个核苷酸。在某些实施例中,所述直接重复结构域在长度上是20至25(例如,20、21、22、23、24或25)个核苷酸。在某些实施例中,所述直接重复结构域在长度上是20个核苷酸。在某些实施例中,所述直接重复结构域在长度上是21个核苷酸。

[0169] 在某些实施例中,所述直接重复结构域包含示于下文提供的SEQ ID NO:3708中的核苷酸序列。

[0170] UAAUUUCUACUCUUGUAGAU[SEQ ID NO:3708]

[0171] 在某些实施例中,所述直接重复结构域包含示于下文提供的SEQ ID NO:3709中的核苷酸序列。

[0172] UAAUUUCUACUAAGUGUAGAU[SEQ ID NO:3709]

[0173] 在某些实施例中,所述直接重复结构域包含示于下文提供的SEQ ID NO:3710中的核苷酸序列。

[0174] GAAUUUCUACUAUUGUAGAU[SEQ ID NO:3710]

[0175] gRNA分子AsCpf1包含含有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域。

[0176] gRNA分子LbCpf1包含含有示于SEQ ID NO:3709中的核苷酸序列的直接重复结构域。

[0177] gRNA分子Lb2Cpf1包含含有示于SEQ ID NO:3710中的核苷酸序列的直接重复结构域。

[0178] 在某些实施例中,所述直接重复结构域包含一个单茎环。

[0179] 在某些实施例中,所述直接重复结构域包含与示于SEQ ID NO:3708、SEQ ID NO:3709或SEQ ID NO:3710中的序列至少约80%、至少约81%、至少约82%、至少约83%、至少约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%或约99%同源的核苷酸序列。

[0180] 在某些实施例中,所述直接重复结构域包含含有示于SEQ ID NO:3708、SEQ ID NO:3709或SEQ ID NO:3710中的序列的至少一个修饰的核苷酸序列。所述修饰基本上或不干扰或影响所述gRNA分子的切割活性。在某些实施例中,所述至少一个修饰选自插入、缺失、突变及其组合。在某些实施例中,所述至少一个修饰在保留RNA双链体的茎环中。在某些实施例中,所述至少一个修饰包含至少一个突变。在某些实施例中,所述至少一个修饰不破坏茎环双链体结构。在某些实施例中,所述至少一个修饰不在示于SEQ ID NO:3708、SEQ ID NO:3709或SEQ ID NO:3710中的序列的3'端的最后三个核苷酸内。在某些实施例中,所述至少一个修饰不在示于SEQ ID NO:3708、SEQ ID NO:3709或SEQ ID NO:3710中的序列的3'端的最后两个核苷酸内。在某些实施例中,所述至少一个修饰不是针对示于SEQ ID NO:3708、SEQ ID NO:3709或SEQ ID NO:3710中的序列的3'端的最后一个核苷酸。在某些实施例中,所述直接重复结构域包含或具有不超过5个修饰、不超过4个修饰、不超过3个修饰、不超过2个修饰或1个修饰。在某些实施例中,所述直接重复结构域包含或具有一个、两个、三个、四个或五个修饰。

#### [0181] 靶向结构域

[0182] 本发明披露的gRNA的靶向结构域与所述靶核酸上的靶结构域互补。在某些实施例中,所述靶向结构域与FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因中的靶核酸互补,例如具有选自下组的核苷酸序列的靶向结构域,该组由以下组成:SEQ ID NO:1-3707。关于靶向结构域选择的指导可以在例如Zetsche等人,Ce11[细胞](2015);163:759-771中找到。

[0183] 在某些实施例中,所述靶向结构域包含与所述靶核酸上的靶序列互补(例如,至少约80%、约85%、约90%、约95%、约98%或约99%互补,例如完全互补)的核苷酸序列。所述靶向结构域是RNA分子的一部分并且因此将包含碱基尿嘧啶(U),同时编码所述gRNA分子的任何DNA均将包含碱基胸腺嘧啶(T)。在某些实施例中,所述靶向结构域与所述靶序列的互补性有利于所述gRNA分子/Cpf1分子复合物与靶核酸的相互作用的特异性。在某些实施例中,在靶向结构域和靶序列对中,所述靶向结构域中的尿嘧啶碱基与所述靶序列中的腺嘌呤碱基配对。在某些实施例中,所述靶向结构域在长度上是5至50(例如,5至10、10至20、20至30、30至40或40至50)个核苷酸。在某些实施例中,所述靶向结构域在长度上是15至30(例如,15至25或25至30)个核苷酸。在某些实施例中,所述靶向结构域在长度上是15至25(例如,15、16、17、18、19、20、21、22、23、24或25)个核苷酸。

[0184] 在某些实施例中,所述靶向结构域在长度上是18个核苷酸。在某些实施例中,所述靶向结构域在长度上是19个核苷酸。在某些实施例中,所述靶向结构域在长度上是20个核苷酸。在某些实施例中,所述靶向结构域在长度上是21个核苷酸。在某些实施例中,所述靶向结构域在长度上是22个核苷酸。在某些实施例中,所述靶向结构域在长度上是23个核苷酸。在某些实施例中,所述靶向结构域在长度上是24个核苷酸。

[0185] 典型地,所述靶向结构域与所述靶序列具有完全互补性。在某些实施例中,所述靶

结构域具有或包含1、2、3、4、5、6、7或8个与所述靶向结构域的相应核苷酸不互补的核苷酸。

[0186] 在某些实施例中,所述靶结构域在其5'端的5个核苷酸内包含1、2、3、4或5个与所述靶向结构域的相应核苷酸互补的核苷酸。在某些实施例中,所述靶结构域在其3'端的5个核苷酸内包含1、2、3、4或5个与所述靶向结构域的相应核苷酸互补的核苷酸。

[0187] 在某些实施例中,所述靶结构域在其5'端的5个核苷酸内包含1、2、3或4个与所述靶向结构域的相应核苷酸不互补的核苷酸。在某些实施例中,所述靶结构域在其3'端的5个核苷酸内包含1、2、3或4个与所述靶向结构域的相应核苷酸不互补的核苷酸。

[0188] 在某些实施例中,互补程度连同gRNA的其他特性足以允许Cpf1分子靶向所述靶核酸。

[0189] 在某些实施例中,所述靶向结构域包含两个与所述靶结构域不互补的连续核苷酸(“非互补核苷酸”),例如所述靶向结构域的5'端的5个核苷酸内、所述靶向结构域的3'端的5个核苷酸内或距所述靶向结构域的一端或两端超过5个核苷酸的两个连续非互补核苷酸。

[0190] 在某些实施例中,在所述靶向结构域的5'端的5个核苷酸内、所述靶向结构域的3'端的5个核苷酸内或在距所述靶向结构域的一端或两端超过5个核苷酸的区域内没有两个连续核苷酸与所述靶向结构域不互补。

[0191] 在某些实施例中,在所述靶向结构域的5'端的5个核苷酸内、所述靶向结构域的3'端的5个核苷酸内或在距所述靶向结构域的一端或两端超过5个核苷酸的区域内没有非互补核苷酸。

[0192] 在某些实施例中,所述靶向结构域包含选自下组的核苷酸序列,该组由以下组成: SEQ ID NO:1-3707。

[0193] 示于SEQ ID NO:1-63中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6(“AsCpf1”)或毛螺菌科细菌(菌株ND2006(“LbCpf1”)或菌株MA2020(“Lb2Cpf1”)) Cpf1一起使用以敲除PDCD1基因。

[0194] 示于SEQ ID NO:64-370中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6(“AsCpf1”)或毛螺菌科细菌(菌株ND2006(“LbCpf1”)或菌株MA2020(“Lb2Cpf1”)) Cpf1一起使用以敲除CTLA4基因。

[0195] 示于SEQ ID NO:371-503中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6(“AsCpf1”)或毛螺菌科细菌(菌株ND2006(“LbCpf1”)或菌株MA2020(“Lb2Cpf1”)) Cpf1一起使用以敲除PTPN6基因。

[0196] 示于SEQ ID NO:504-2325中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6(“AsCpf1”)或毛螺菌科细菌(菌株ND2006(“LbCpf1”)或菌株MA2020(“Lb2Cpf1”)) Cpf1一起使用以敲除CBLB基因。

[0197] 示于SEQ ID NO:2326-3094中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6(“AsCpf1”)或毛螺菌科细菌(菌株ND2006(“LbCpf1”)或菌株MA2020(“Lb2Cpf1”)) Cpf1一起使用以敲除FAS基因。

[0198] 示于SEQ ID NO:3095-3283中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6(“AsCpf1”)或毛螺菌科细菌(菌株ND2006(“LbCpf1”)或菌株MA2020(“Lb2Cpf1”)) Cpf1一起使用以敲除B2M基因。

[0199] 示于SEQ ID NO:3284-3385中的任何靶向结构域均可以与氨基酸球菌属(例如,菌

株BV3L6 (“AsCpf1”)或毛螺菌科细菌(菌株ND2006 (“LbCpf1”)或菌株MA2020 (“Lb2Cpf1”)) Cpf1一起使用以敲除BID基因。

[0200] 示于SEQ ID NO:3386-3588中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6 (“AsCpf1”)或毛螺菌科细菌(菌株ND2006 (“LbCpf1”)或菌株MA2020 (“Lb2Cpf1”)) Cpf1一起使用以敲除TRAC基因。

[0201] 示于SEQ ID NO:3589-3707中的任何靶向结构域均可以与氨基酸球菌属(例如,菌株BV3L6 (“AsCpf1”)或毛螺菌科细菌(菌株ND2006 (“LbCpf1”)或菌株MA2020 (“Lb2Cpf1”)) Cpf1一起使用以敲除TRBC基因。

[0202] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3433中的核苷酸序列。

[0203] AGAAUCAAAAUCGGUGAAUAGGC (SEQ ID NO:3433)

[0204] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3587中的核苷酸序列。

[0205] UUUGAGAAUCAAAAUCGGUGAAU (SEQ ID NO:3587)

[0206] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3538中的核苷酸序列。

[0207] GUCUGUGAUUACACAUCAGAAU (SEQ ID NO:3538)

[0208] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3461中的核苷酸序列。

[0209] CACAUGCAAAGUCAGAUUUGUUG (SEQ ID NO:3461)

[0210] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3475中的核苷酸序列。

[0211] CAUGUGCAAACGCCUUCAACAAC (SEQ ID NO:3475)

[0212] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3524中的核苷酸序列。

[0213] GAUUCUCAAAACAAAUGUGUCACA (SEQ ID NO:3524)

[0214] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3566中的核苷酸序列。

[0215] UCUGUGAUUACACAUCAGAAUC (SEQ ID NO:3566)

[0216] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3517中的核苷酸序列。

[0217] GAGUCUCUCAGCUGGUACACGGC (SEQ ID NO:3517)

[0218] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3573中的核苷酸序列。

[0219] UGACACAUUUGUUUGAGAAUCA (SEQ ID NO:3573)

[0220] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3580中的核苷酸序列。

[0221] UUGCUCAGGCCACAGCACUGUU (SEQ ID NO:3580)

[0222] 在某些实施例中,所述靶向结构域包含示于下文提供的SEQ ID NO:3454中的核苷酸序列。

酸序列。

[0223] AUUCUCAAAACAAAUGUGUCACAA (SEQ ID NO:3454)

[0224] 在某些实施例中,所述靶向结构域包含与选自下组的一个序列至少约80%、至少约81%、至少约82%、至少约83%、至少约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%或约99%同源的核苷酸序列,该组由以下组成:SEQ ID NO:1-3707。

[0225] 在某些实施例中,所述靶向结构域包含含有选自下组的一个序列的至少一个修饰的核苷酸序列,该组由以下组成:SEQ ID NO:1-3707。在某些实施例中,所述一个或多个修饰是第VIII部分中披露的一个或多个修饰。在某些实施例中,至少一个修饰使得所述靶向结构域较不易受降解影响或更具生物相容性,例如较低免疫原性。作为举例,所述靶向结构域的骨架可以用硫代磷酸酯或来自第VIII部分的其他一个或多个修饰来修饰。在某些实施例中,所述靶向结构域的核苷酸可以包含2' 修饰,例如2-乙酰化,例如2' 甲基化,或来自第VIII部分的其他一个或多个修饰。

[0226] 在某些实施例中,所述至少一个修饰选自插入、缺失、突变及其组合。在某些实施例中,所述靶向结构域包含1、2、3、4、5、6、7或8个或更多个修饰。在某些实施例中,所述靶向结构域在其5' 端的5个核苷酸内包含1、2、3或4个修饰。在某些实施例中,所述靶向结构域在其3' 端的5个核苷酸内包含多达1、2、3或4个修饰。

[0227] 在某些实施例中,所述靶向结构域包含在两个连续核苷酸处的修饰,例如所述靶向结构域的5' 端的5个核苷酸内、所述靶向结构域的3' 端的5个核苷酸内或距所述靶向结构域的一端或两端超过5个核苷酸的两个连续核苷酸。

[0228] 在某些实施例中,在所述靶向结构域的5' 端的5个核苷酸内、所述靶向结构域的3' 端的5个核苷酸内或在距所述靶向结构域的一端或两端超过5个核苷酸的区域内没有两个连续核苷酸被修饰。某些实施例中,在所述靶向结构域的5' 端的5个核苷酸内、所述靶向结构域的3' 端的5个核苷酸内或在距所述靶向结构域的一端或两端超过5个核苷酸的区域内没有核苷酸被修饰。

[0229] 可以选择所述靶向结构域中的修饰以便不干扰靶向功效,这可以通过测试描述于第IV部分中的系统中的候选修饰来进行评估。具有候选靶向结构域的gRNA可以在第IV部分中的系统中进行评估,所述候选靶向结构域具有选定的长度、序列、互补程度或修饰程度。所述候选靶向结构域可以被单独地或与一种或多种其他候选变化放置在已知与选定的靶具有功能性的gRNA分子/Cpf1分子系统中并且进行评估。

[0230] 在某些实施例中,全部的经修饰的核苷酸互补于并且能够杂交到所述靶结构域中存在的相应核苷酸上。在某些实施例中,1、2、3、4、5、6、7或8个或更多个经修饰的核苷酸不互补于或不能够杂交到所述靶结构域中存在的相应核苷酸上。

[0231] II. 用于设计gRNA的方法

[0232] 提供了用于选择、设计和验证用于本文所描述的gRNA的靶向结构域的方法。本文还提供了用于掺入gRNA的示例性靶向结构域。

[0233] 用于选择和验证靶序列连同脱靶分析的方法描述于例如Mali等人,2013SCIENCE [科学]339(6121):823-826;Hsu等人NAT BIOTECHNOL[自然生物技术],31(9):827-32;Fu等人,2014NAT BIOTECHNOL[自然生物技术],doi:10.1038/nbt.2808.PubMed PMID:

24463574;Heigwer等人,2014NAT METHODS[自然方法]11(2):122-3.doi:10.1038/nmeth.2812.PubMed PMID:24481216;Bae等人,2014BIOINFORMATICS[生物信息学]PubMed PMID:24463181;Xiao A等人,2014BIOINFORMATICS[生物信息学]PubMed PMID:24389662;以及Zetsche等人,Cell[细胞](2015);163:759-771中。

[0234] 在某些实施例中,软件工具可以用来优化在使用者的靶序列内的gRNA的靶向结构域的选择,例如以跨基因组最小化总脱靶活性。脱靶活性可以不同于切割。例如,对于有待与Cpf1分子(例如,AsCpf1、LbCpf1或Lb2Cpf1)一起使用的gRNA分子的每个可能的靶向结构域,软件工具可以鉴定跨基因组所有潜在的脱靶序列,所述脱靶序列含有高达一定数量(例如,1、2、3、4、5、6、7、8、9或10)的错配碱基对。在每个脱靶序列处的切割效率是可以预测的,例如使用实验衍生的加权方案。其他功能(例如,用于gRNA载体构建的自动化试剂设计、用于中靶Surveyor测定的引物设计和用于高通量检测以及经由下一代测序对脱靶切割进行定量的引物设计)也可以被包括在所述工具中。候选gRNA分子可以通过本领域已知的方法或如第IV部分中所描述的进行评估。

[0235] 在某些实施例中,使用DNA序列检索算法(例如,使用基于公共工具cas-offinder的定制gRNA设计软件)鉴定用于与Cpf1分子一起使用的gRNA(Bae等人Bioinformatics[生物信息学].2014;30(10):1473-1475)。所述定制gRNA设计软件在计算指导物的全基因组脱靶倾向之后为指导物打分。典型地,针对长度范围从17至24的指导物考虑范围从完美匹配至7个错配的匹配。一旦经计算确定了脱靶位点,便计算每种指导物的总分并且使用web界面以表格输出总结。除鉴定与PAM序列(例如,(T)<sub>x</sub>N PAM,例如TTN PAM)邻近的潜在gRNA位点之外,所述软件还鉴定与选定的gRNA位点相差1、2、3个或更多个核苷酸的所有PAM邻近序列。从UCSC基因组浏览器获得每个基因的基因组DNA序列,并且使用可公开获得的RepeatMasker程序针对重复元件对序列进行筛选。RepeatMasker针对重复元件和具有低复杂性的区域对输入DNA序列进行检索。输出是存在于给定查询序列中的重复的详细注释。

[0236] 鉴定后,基于它们与靶位点的距离(基于在人基因组中含有相关PAM的靠近匹配的鉴定)将gRNA排成等级。在某些实施例中,对于AsCpf1、LbCpf1或Lb2Cpf1,所述PAM是TTN PAM。第一等级gRNA分子的靶向结构域靶向FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因中起始密码子下游的编码序列的前500bp内。第二等级gRNA分子的靶向结构域靶向FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因的编码序列的剩余部分。在某些情况下,当基因的编码序列短于500bp时,gRNA分子的靶向结构域都被包括在第一等级中。需注意等级是非包容性的(对于所述策略,每种gRNA仅被列出一次)。

[0237] 表1提供了根据第一设计和分级策略的示例性靶向结构域。作为一个实例,设计了18-mer、19-mer、20-mer、21-mer、22-mer、23-mer或24-mer靶向结构域。被设计为与AsCpf1、LbCpf1或Lb2Cpf1分子一起使用的示例性gRNA(由SEQ ID NO提及)提供于表1中,所述示例性gRNA是使用此基于分级的途径相对于敲除FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因的表达而鉴定。在某些实施例中,所述靶向结构域通过互补碱基配对与所述靶结构域杂交。表1的示于SEQ ID NO:1-3707中的任何靶向结构域均可以与AsCpf1、LbCpf1或Lb2Cpf1分子一起使用,以减少、降低或抑制FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因的表达。

[0238] 表1.示例性gRNA的SEQ ID NO(AsCpf1、LbCpf1或Lb2Cpf1)

[0239]

等级	<i>PDCD1</i>	<i>CTLA4</i>	<i>PTPN6</i>	<i>CBLB</i>	<i>FAS</i>	<i>B2M</i>	<i>BID</i>	<i>TRAC</i>	<i>TRBC</i>
1	SEQ ID NO: 1-15	SEQ ID NO: 64-167	SEQ ID NO: 371-405	SEQ ID NO: 504-788	SEQ ID NO: 2326-2664	SEQ ID NO: 3095-3220	SEQ ID NO: 3284-3385	SEQ ID NO: 3386-3588	SEQ ID NO: 3589-3707
2	SEQ ID NO: 15-63	SEQ ID NO: 168-370	SEQ ID NO: 406-503	SEQ ID NO: 789-2325	SEQ ID NO: 2665-3094	SEQ ID NO: 3221-3283	无	无	无

[0240] III.Cpf1分子

[0241] 多个物种的Cpf1分子可以用于本文所披露的方法和组合物中。在某些实施例中，所述Cpf1分子选自氨基酸球菌属（例如，菌株BV3L6）分子（“AsCpf1”）、毛螺菌科细菌（例如，菌株ND2006）分子（“LbCpf1”）和毛螺菌科细菌（例如，菌株MA2020）分子（“Lb2Cpf1”）。

[0242] 如该术语在本文使用的，Cpf1分子或Cpf1多肽是指可以与gRNA分子相互作用并且与所述gRNA分子一起归巢或定位至包含靶结构域和PAM序列的位点的分子或多肽。如那些术语在本文使用的，Cpf1分子和Cpf1多肽包括天然存在的Cpf1分子以及例如与参比序列（例如，最相似的天然存在的Cpf1分子）相差至少一个氨基酸残基的工程化的、改变的或经修饰的Cpf1分子或Cpf1多肽。

[0243] 已经确定了新凶手弗朗西斯菌 (*Francisella novicida*) U112Cpf1 (“FnCpf1”) 分子的结构 (Zetsche等人, *Cell* [细胞] (2015); 163:759-771)。Cpf1是缺乏反式激活crRNA (tracrRNA) 的单RNA指导的内切核酸酶。参见Zetsche (2015), 将其通过引用以其整体而并入。Cpf1利用短的富含T的原型间隔子邻近基序 (“PAM”) 来切割靶DNA。参见同上。Cpf1经由具有4或5-nt 5'突出端的交错的DNA双链断裂来切割DNA。参见同上。5'突出端可以经由非同源末端连接 (NHEJ) 机制促进基因插入。已经显示AsCpf1和LbCpf1在人细胞中展现出核酸酶活性。参见同上。天然存在的Cpf1分子包含N末端混合的 $\alpha/\beta$ 结构域和C末端RuvC样内切核酸酶结构域 (“RuvC结构域”)。所述RuvC结构域包含三个分割 (split) RuvC基序: RuvC I、RuvCII和RuvCIII。与Cas9不同, Cpf1缺乏HNH内切核酸酶结构域。另外, 天然存在的Cpf1分子包含RuvC I和RuvC II之间的螺旋区以及RuvC II和RuvC III之间的锌指样结构域。天然存在的Cpf1分子的结构描述于Zetsche (2015) 中。

[0244] 所述RuvC结构域保留所有催化残基, 例如对于FnCpf1为D<sup>917</sup>、E<sup>1006</sup>和D<sup>1255</sup>。所述RuvC结构域切割靶DNA (例如, FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC) 的两条链, 并且产生双链断裂。

[0245] 本发明披露的Cpf1分子或Cpf1多肽与指导RNA (gRNA) 分子相互作用并且与所述gRNA分子一起定位至包含靶结构域和PAM序列的位点。在某些实施例中, Cpf1分子或Cpf1多肽与靶核酸相互作用并且切割靶核酸的能力是PAM序列依赖性的。PAM序列是在所述靶核酸中的序列。在某些实施例中, 所述靶核酸的切割发生在PAM序列的上游。在某些实施例中, 所

述靶核酸的切割发生在PAM序列的下游。来自不同细菌物种的Cpf1分子可以识别不同序列基序(例如,PAM序列)。在某些实施例中,所述PAM是富含T的PAM。在某些实施例中,所述PAM具有核苷酸序列(T)<sub>x</sub>N,其中X是1-10,并且N是A、G、C或T。在某些实施例中,X是2,并且因此所述PAM是TTN。在某些实施例中,X是3,并且因此所述PAM是TTTN。在某些实施例中,Cpf1分子(例如,AsCpf1、LbCpf1或Lb2Cpf1)识别序列基序TTTN并引导在该序列下游1-24(例如,1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23或24)bp的靶核酸序列的切割。Cpf1分子识别PAM序列的能力可以通过Pattanayak等人,NATURE BIOTECHNOLOGY [自然生物技术] (2013);31(9):839-843中描述的体外选择测定来确定。

[0246] 在某些实施例中,Cpf1分子或Cpf1多肽包含如下氨基酸序列:

[0247] 所述氨基酸序列与本文所披露的任何Cpf1分子序列或天然存在的Cpf1分子序列(例如,来自本文所列物种的或Zetsche (2015)中描述的Cpf1分子)至少约60%、约65%、约70%、约75%、约80%、约81%、约82%、约83%、约84%、约85%、约86%、约87%、约88%、约89%、约90%、约91%、约92%、约93%、约94%、约95%、约96%、约97%、约98%或约99%同源;

[0248] 所述氨基酸序列当与本文所披露的任何Cpf1分子序列或天然存在的Cpf1分子序列(例如,来自本文所列物种的或Zetsche (2015)中描述的Cpf1分子)相比时相差不超过约2%、约5%、约10%、约15%、约20%、约30%或40%的氨基酸残基;

[0249] 所述氨基酸序列与本文所披露的任何Cpf1分子序列或天然存在的Cpf1分子序列(例如,来自本文所列物种的或Zetsche (2015)中描述的Cpf1分子)相差至少1、2、5、10或20个氨基酸但相差不超过100、80、70、60、50、40或30个氨基酸;或者

[0250] 与本文所披露的任何Cpf1分子序列或天然存在的Cpf1分子序列(例如,来自本文所列物种的或Zetsche (2015)中描述的Cpf1分子)相同。

[0251] 在某些实施例中,所述Cpf1分子或Cpf1多肽是不同于参比Cpf1分子或参比Cpf1多肽的工程化的Cpf1分子或工程化的Cpf1多肽。在某些实施例中,所述参比Cpf1分子或参比Cpf1多肽是天然存在的Cpf1分子或Cpf1多肽。在某些实施例中,工程化的Cpf1分子或工程化的Cpf1多肽保留或基本上保留所述参比Cpf1分子或参比Cpf1多肽的核酸酶(例如,内切核酸酶)活性。在某些实施例中,所述工程化Cpf1分子或工程化Cpf1多肽保留所述参比Cpf1分子或参比Cpf1多肽的至少约70%、约80%、约90%、约95%或约99%的核酸酶活性。

[0252] 可以相对于参比Cpf1分子(例如,天然存在的Cpf1分子)引入一个或多个突变或改变。这样的—个或多个突变或者一个或多个改变可以包括取代(例如,保守取代或非必需氨基酸的取代);插入;和/或缺失。如本文所使用的,术语“非必需”氨基酸残基是指可以从Cpf1分子(例如,天然存在的Cpf1分子)的野生型序列改变而不消除或基本上不改变Cpf1活性(例如,核酸酶/切割活性)的残基。在某些实施例中,Cpf1分子或Cpf1多肽多肽相对于参比Cpf1分子(例如,天然存在的Cpf1分子)包含一个或多个突变或改变,例如至少1、2、3、4、5、10、15、20、30、40或50个突变或改变但少于200、100或80个突变或改变。所述一个或多个突变和一个或多个改变可以存在于混合的 $\alpha/\beta$ 结构域、RuvC结构域或混合的 $\alpha/\beta$ 结构域和RuvC结构域中。

[0253] 在某些实施例中,工程化的Cpf1分子或工程化的Cpf1多肽包含不同于天然存在的Cpf1分子(例如,不同于具有最接近同源性的天然存在的Cpf1分子)的切割特性。例如,Cpf1

分子或Cpf1多肽与天然存在的Cpf1分子可以具有如下不同：例如与天然存在的Cpf1分子（例如，AsCpf1、LbCpf-1或Lb2Cpf1）相比，其调节（例如，降低或增加）双链核酸的切割的能力（内切核酸酶活性）；或者切割核酸分子（例如，双链核酸分子）的能力可以被消除。

[0254] 在某些实施例中，所述工程化Cpf1分子或Cpf1多肽是例如两种或更多种不同Cpf1分子或Cpf1多肽（例如，不同物种的两种或更多种天然存在的Cpf1分子）的融合体。例如，可以将一个物种的天然存在的Cpf1分子的片段融合到第二物种的Cpf1分子的片段上。

[0255] 天然存在的Cpf1分子可以识别特定的PAM序列，例如上文针对AsCpf1、LbCpf-1和Lb2Cpf1所描述的5'-TTN PAM序列。在某些实施例中，与所述参比Cpf1分子相比，工程化的Cpf1分子或工程化的Cpf1多肽具有改变的PAM特异性（例如，影响PAM识别）。例如，可以改变天然存在的Cpf1分子，例如以改变PAM识别，例如以改变Cpf1分子或Cpf1多肽识别以减少脱靶位点和/或改进特异性或消除PAM识别需要的PAM序列。在某些实施例中，可以改变Cpf1分子，例如以增加PAM识别序列的长度和/或改进对高水平一致性的Cpf1特异性，例如以减少脱靶位点并增加特异性。在某些实施例中，所述PAM识别序列的长度在长度上是至少4、5、6、7、8、9、10或15个氨基酸。

[0256] 可以使用定向进化产生识别不同PAM序列和/或具有降低的脱靶活性的Cpf1分子或Cpf1多肽。可以用于Cpf1分子定向进化的示例性方法和系统描述于例如Esvelt等人Nature[自然]2011, 472(7344): 499-503中。可以例如通过描述于第IV部分中的方法对候选Cpf1分子进行评估。

[0257] 在某些实施例中，所述工程化Cpf1分子和工程化Cpf1多肽包含一个或多个缺失，其减小所述分子的尺寸，同时保留或基本上保留（例如，基本上不影响或降低）所需的Cpf1特性，例如基本上天然的构象、Cpf1核酸酶活性（例如，内切核酸酶活性；即切割双链核酸的两条链并产生双链断裂的能力）和/或核酸分子（例如，靶核酸或gRNA）的识别活性。所述工程化Cpf1分子的较小尺寸允许提高递送方法的灵活性，并且由此增加基因组编辑的实用性。在某些实施例中，所述工程化Cpf1分子和工程化Cpf1多肽进一步包含一个或多个接头，其中接头布置于所述缺失侧翼的氨基酸残基之间。

[0258] 在某些实施例中，编码Cpf1分子或Cpf1多肽的核酸组合物包含合成核酸序列。例如，所述合成核酸序列可以进行化学修饰。在某些实施例中，所述序列mRNA具有一种或多种（例如，所有）以下特性：它被加帽，多腺苷酸化，被5-甲基胞苷和/或假尿苷取代。

[0259] 另外地或可替代地，可以对所述合成核酸序列进行密码子优化，例如至少一个非常见密码子或低不常见密码子已经被常见密码子替换。例如，所述合成核酸可以引导优化的信使mRNA的合成（例如，对于在哺乳动物表达系统（例如，本文所描述的）中的表达进行优化）。

[0260] 另外地或可替代地，编码序列分子或序列多肽的核酸可以包含核定位序列（NLS）。核定位序列在本领域是已知的。

[0261] 编码AsCpf1的示例性人密码子优化的核酸序列示于下文提供的SEQ ID NO:3722中。

[0262] ATGACCCAGTTCGAGGGCTTCACCAACCTGTACCAGGTGAGCAAGACCCTGCGGTTTCGAGCTGATCCC  
CCAGGGCAAGACCCTGAAGCACATCCAGGAGCAGGGCTTCATCGAGGAGGACAAGGCCGGAACGACCACTACAAG  
GAGCTGAAGCCCATCATCGACCGGATCTACAAGACCTACGCCGACCAGTGCCTGCAGCTGGTGCAGCTGGACTGGG

AGAACCTGAGCGCCGCCATCGACAGCTACCGGAAGGAGAAGACCGAGGAGACCCGGAACGCCCTGATCGAGGAGCA  
GGCCACCTACCGGAACGCCATCCACGACTACTTCATCGGCCGGACCGACAACCTGACCGACGCCATCAACAAGCGG  
CACGCCGAGATCTACAAGGGCCTGTTCAAGGCCGAGCTGTTCAACGGCAAGGTGCTGAAGCAGCTGGGCACCGTGA  
CCACCACCGAGCACGAGAACGCCCTGCTGCGGAGCTTCGACAAGTTACCACCTACTTCAGCGGCTTCTACGAGAA  
CCGGAAGAACGTGTTACAGCGCCGAGGACATCAGCACCGCCATCCCCACCGGATCGTGCAGGACAACCTCCCCAAG  
TTCAAGGAGAACTGCCACATCTTCACCCGGCTGATCACCGCCGTGCCAGCCTGCGGGAGCACTTCGAGAACGTGA  
AGAAGGCCATCGGCATCTTCGTGAGCACCAGCATCGAGGAGGTGTTACGTTCCCTTCTACAACCAGCTGCTGAC  
CCAGACCCAGATCGACCTGTACAACCAGCTGCTGGGCGGCATCAGCCGGGAGGCCGGCACCGAGAAGATCAAGGGC  
CTGAACGAGGTGCTGAACCTGGCCATCCAGAAGAACGACGAGACCGCCACATCATCGCCAGCCTGCCCCACCGGT  
TCATCCCCCTGTTCAAGCAGATCCTGAGCGACCGGAACACCCTGAGCTTCATCCTGGAGGAGTTCAAGAGCGACGA  
GGAGGTGATCCAGAGCTTCTGCAAGTACAAGACCCTGCTGCGGAACGAGAACGTGCTGGAGACCGCCGAGGCCCTG  
TTCAACGAGCTGAACAGCATCGACCTGACCCACATCTTCATCAGCCACAAGAAGCTGGAGACCATCAGCAGCGCCC  
TGTGCGACCACTGGGACACCCTGCGGAACGCCCTGTACGAGCGGGCGATCAGCGAGCTGACCGCAAGATCACCAA  
GAGCGCCAAGGAGAAGGTGCAGCGGAGCCTGAAGCACGAGGACATCAACCTGCAGGAGATCATCAGCGCCGCCGGC  
AAGGAGCTGAGCGAGGCCTTCAAGCAGAAGACCAGCGAGATCCTGAGCCACGCCACGCCGCCCTGGACCAGCCCC  
TGCCCACCACCCTGAAGAAGCAGGAGGAGAAGGAGATCCTGAAAAGCCAGCTGGACAGCCTGCTGGGCCTGTACCA  
CCTGCTGGACTGGTTCGCCGTGGACGAGAGCAACGAGGTGGACCCCCGAGTTCAGCGCCCGGCTGACCGGCATCAAG  
CTGGAGATGGAGCCCAGCCTGAGCTTCTACAACAAGGCCCGGAACCTACGCCACCAAGAAGCCCTACAGCGTGGAGA  
AGTTCAAGCTGAACTTCAGATGCCACCCTGGCCAGCGGCTGGGACGTGAACAAGGAGAAGAACAACGGCGCCAT  
CCTGTTTCGTGAAGAACGGCCTGTACTACCTGGGCATCATGCCAAGCAGAAGGGCCGGTACAAGGCCCTGAGCTTC  
GAGCCCACCAGAAAGACCAGCGAGGGCTTCGACAAGATGTACTACGACTACTTCCCCGACGCCGCAAGATGATCC  
CCAAGTGCAGCACCCAGCTGAAGGCCGTGACCGCCACTTCCAGACCCACACCACCCCATCCTGCTGAGCAACAA  
CTTCATCGAGCCCCTGGAGATACCAAGGAGATCTACGACCTGAACAACCCCGAGAAGGAGCCCAAGAAGTTCCAG  
ACCGCCTACGCCAAGAAGACCGGCGACCAGAAGGGCTACCGGGAGGCCCTGTGCAAGTGGATCGACTTACCCGGG  
ACTTCCTGAGCAAGTACACCAAGACCACCAGCATCGACCTGAGCAGCCTGCGGCCAGCAGCCAGTACAAGGACCT  
GGGCGAGTACTACGCCGAGCTGAACCCCTGCTGTACCACATCAGCTTCCAGCGGATCGCCGAGAAGGAGATCATG  
GACGCCGTGGAGACCGGCAAGCTGTACCTGTTCCAGATCTACAACAAGGACTTCGCCAAGGGCCACCACGGCAAGC  
CCAACCTGCACACCCTGTACTGGACCGGCCTGTTAGCCCCGAGAACCTGGCCAAGACCAGCATCAAGCTGAACGG  
CCAGGCCGAGCTGTTCTACCGGCCAAGAGCCGGATGAAGCGGATGGCCCACCGGCTGGGCGAGAAGATGCTGAAC  
AAGAAGCTGAAGGACCAGAAGACCCCATCCCCGACACCCTGTACCAGGAGCTGTACGACTACGTGAACCACCGGC  
TGAGCCACGACCTGAGCGACGAGGCCCGGGCCCTGCTGCCAACGTGATCACCAAGGAGGTGAGCCACGAGATCAT  
CAAGGACCGGCGTTACCAGCGACAAGTTCTTCTTCCACGTGCCATCACCTGAACTACCAGGCCGCAACAGC  
CCCAGCAAGTTCAACCAGCGGTGAACGCCTACCTGAAGGAGCACCCCGAGACCCCATCATCGGCATCGACCGGG  
GCGAGCGGAACCTGATCTACATCACCGTATCGACAGCACCGGAAGATCCTGGAGCAGCGGAGCCTGAACACCAT  
CCAGCAGTTCGACTACCAGAAGAAGCTGGACAACCGGGAGAAGGAGCGGGTGGCCGCCCGGAGGCCTGGAGCGTG  
GTGGGCACCATCAAGGACCTGAAGCAGGGCTACCTGAGCCAGGTGATCCACGAGATCGTGGACCTGATGATCCACT  
ACCAGGCCGTGGTGGTGCTGGAGAACCTGAACTTCGGCTTCAAGAGCAAGCGGACCGGCATCGCCGAGAAGGCCGT  
GTACCAGCAGTTCGAGAAGATGCTGATCGACAAGCTGAACTGCCTGGTGTGAAGGACTACCCCGCCGAGAAGGTG  
GGCGGCGTGCTGAACCCCTACCAGCTGACCGACCAGTTACCAGCTTCGCCAAGATGGGCACCCAGAGCGGCTTCC

TGTTCTACGTGCCCGCCCCCTACACCAGCAAGATCGACCCCTGACCGGCTTCGTGGACCCCTTCGTGTGGAAGAC  
CATCAAGAACCACGAGAGCCGGAAGCACTTCTGGAGGGCTTCGACTTCTGCACTACGACGTGAAGACCGGCGAC  
TTCATCCTGCACTTCAAGATGAACCGGAACCTGAGCTTCCAGCGGGCCTGCCCGGCTTCATGCCCGCTGGGACA  
TCGTGTTGAGAAGAACGAGACCCAGTTCGACGCCAAGGGCACCCCTTCATCGCCGGAAGCGGATCGTGCCCGT  
GATCGAGAACCACCGGTTACCGGCCGGTACCGGGACCTGTACCCCGCCAACGAGCTGATCGCCCTGCTGGAGGAG  
AAGGGCATCGTGTCCGGGACGGCAGCAACATCCTGCCAAGCTGCTGGAGAACGACGACAGCCACGCCATCGACA  
CCATGGTGGCCCTGATCCGGAGCGTGTGCAGATGCGGAACAGCAACGCCGCCACCGGCGAGGACTACATCAACAG  
CCCCGTGCGGGACCTGAACGGCGTGTGCTTCGACAGCCGGTTCAGAACCCCGAGTGGCCCATGGACGCCGACGCC  
AACGGCGCTACCACATCGCCCTGAAGGGCCAGCTGCTGCTGAACCACCTGAAGGAGAGCAAGGACCTGAAGCTGC  
AGAACGGCATCAGCAACCAGGACTGGCTGGCTACATCCAGGAGCTGCGGAACAAGCGGCCCGCCACCAAGAA  
GGCCGGCCAGGCCAAGAAGAAGGGCAGCTACCCCTACGACGTGCCCGACTACGCCTACCCCTACGACGTGCC  
GACTACGCCTACCCCTACGACGTGCCCGACTACGCCTGA (SEQ ID NO:3722)

[0263] 由示于SEQ ID NO:3722中的核苷酸序列编码的相应氨基酸序列示于下文提供的SEQ ID NO:3725中。

[0264] MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQQFIEEDKARNDDHYKELKPI IDRIYKTYADQCLQ  
LVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDFYIGRTDNLDAINKRHAIEYKGLFKAELFNGKVL  
KQLGTVTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEIDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLR  
EHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHII  
ASLPHRFIPLFKQILSDRNTLSFILLEFKSDEEVIQSFCKYKTLRNENVLETAELFNELNSIDLTHIFISHKKL  
ETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEI ISAAGKELSEAFKQKTSEILSHA  
HAALDQPLPTTLKQEEKEILKSQLDLGLYHLLDWFVAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYAT  
KKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIIMPQKQKRYKALSFEPTTEKTSEGFDKMYDYF  
PDAAKMIPKSTQLKAVTAHFQTHHTPILLSNFIETKEIYDLNPEKEPKKFQAYAKKTGDQKGYREALC  
KWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDF  
AKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAEFYRPKSRMKRMAHRLGKMLNKKLKDQKTPIPDTLYQEL  
YDYVNHRLSHDLSDEARALLPNVITKEVSHEI IKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPET  
PIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKLDNREKERVAAARQAWSVVGTIKDLKQGYLSQVIHE  
IVDLMIHYQAVVVLENLNFQKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLPYQLTDQFTSFAK  
MGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWTKIKNHESRKHFLGFDLHYDVKTGDFILHFKMNRNLSFQRGLP  
GFMPAWDIVFEKNETQFDAQGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGI VFRDGSNILPKLLEN  
DDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPWPMDADANGAYHIALKGQLLLNHLK  
ESKDLKLQNGISNQDWLAYIQELRN (SEQ ID NO:3725)

[0265] 编码LbCpf1的示例性人密码子优化的核酸序列示于下文提供的SEQ ID NO:3723中。

[0266] ATGAGCAAGCTGGAGAAGTTCACCAACTGCTACAGCCTGAGCAAGACCCTGCGGTTCAAGGCCATCCC  
CGTGGGCAAGACCCAGGAGAACATCGACAACAAGCGGCTGCTGGTGGAGGACGAGAAGCGGGCCGAGGACTACAAG  
GGCGTGAAGAAGCTGCTGGACCGGTAACCTGAGCTTCATCAACGACGTGCTGCACAGCATCAAGCTGAAGAACC  
TGAACAACACTACATCAGCCTGTCCGGAAGAAGACCCGACCGAGAAGGAGAACAAGGAGCTGGAGAACCTGGAGAT  
CAACCTGCGGAAGGAGATCGCCAAGGCCTTCAAGGGCAACGAGGGCTACAAGAGCCTGTTCAAGAAGGACATCATC

GAGACCATCCTGCCCGAGTTCCTGGACGACAAGGACGAGATCGCCCTGGTGAACAGCTTCAACGGCTTACCACCG  
CCTTACCAGCTTCTTCGACAACCGGGAGAACATGTTTCAGCGAGGAGGCCAAGAGCACCAGCATCGCCTTCCGGTG  
CATCAACGAGAACCTGACCCGGTACATCAGCAACATGGACATCTTCGAGAAGGTGGACGCCATCTTCGACAAGCAC  
GAGGTGCAGGAGATCAAGGAGAAGATCCTGAACAGCGACTACGACGTGGAGGACTTCTTCGAGGGCGAGTTCTTCA  
ACTTCGTGCTGACCCAGGAGGGCATCGACGTGTACAACGCCATCATCGGGCGCTTCGTGACCGAGAGCGGCGAGAA  
GATCAAGGGCTGAACGAGTACATCAACCTGTACAACCAGAAGACCAAGCAGAAGCTGCCCAAGTTCAAGCCCCTG  
TACAAGCAGGTGCTGAGCGACCGGGAGAGCCTGAGCTTCTACGGCGAGGGCTACACCAGCGACGAGGAGGTGCTGG  
AGGTGTTCCGGAACACCCTGAACAAGAACAGCGAGATCTTCAGCAGCATCAAGAAGCTGGAGAAGCTGTTCAAGAA  
CTTCGACGAGTACAGCAGCGCCGGCATCTTCGTGAAGAACGGCCCCGCCATCAGCACCATCAGCAAGGACATCTTC  
GGCGAGTGGAACGTGATCCGGGACAAGTGAACGCCGAGTACGACGACATCCACCTGAAGAAGAAGGCCGTGGTGA  
CCGAGAAGTACGAGGACGACCGGGGAAAAGCTTCAAGAAGATCGGCAGCTTCAGCCTGGAGCAGCTGCAGGAGTA  
CGCCGACGCCGACCTGAGCGTGGTGGAGAAGCTGAAGGAGATCATCATCCAGAAGGTGGACGAGATCTACAAGGTG  
TACGGCAGCAGCGAGAAGCTGTTTCGACGCCGACTTCGTGCTGGAGAAAAGCCTGAAGAAGAACGACCGCGTGGTGG  
CCATCATGAAGGACCTGCTGGACAGCGTAAAAGCTTCGAGAACTACATCAAGGCCTTCTTCGGCGAGGGCAAGGA  
GACCAACCGGGACGAGAGCTTCTACGGCGACTTCGTGCTGGCCTACGACATCCTGCTGAAGGTGGACCACATCTAC  
GACGCCATCCGGAACCTACGTGACCCAGAAGCCCTACAGCAAGGACAAGTTCAAGCTGTACTTCCAGAACCCCCAGT  
TCATGGGCGGCTGGGACAAGGACAAGGAGACCGACTACCGGGCCACCATCCTGCGGTACGGCAGCAAGTACTACCT  
GGCCATCATGGACAAGAAGTACGCCAAGTGCCTGCAGAAGATCGACAAGGACGACGTGAACGGCAACTACGAGAAG  
ATCAACTACAAGCTGCTGCCCGCCCCAACAAGATGCTGCCCAAGGTGTTCTTCAGCAAGAAGTGGATGGCCTACT  
ACAACCCAGCGAGGACATCCAGAAGATCTACAAGAACGGCACCTTCAAGAAGGGCGACATGTTCAACCTGAACGA  
CTGCCACAAGCTGATCGACTTCTTCAAGGACAGCATCAGCCGGTACCCCAAGTGGAGCAACGCCTACGACTTCAAC  
TTCAGCGAGACCGAGAAGTACAAGGACATCGCCGGCTTCTACCGGGAGGTGGAGGAGCAGGGCTACAAGGTGAGCT  
TCGAGAGCGCCAGCAAGAAGGAGGTGGACAAGCTGGTGGAGGAGGGCAAGCTGTACATGTTCCAGATCTACAACAA  
GGACTTCAGCGACAAGAGCCACGGCACCCCCAACCTGCACACCATGTACTTCAAGCTGCTGTTTCGACGAGAACAAC  
CACGGCCAGATCCGGCTGAGCGGGCGCGCCGAGCTGTTTCATGCGGGCGGCCAGCCTGAAGAAGGAGGAGCTGGTGG  
TGCACCCCGCCAACAGCCCCATCGCCAACAAGAACCCCGACAACCCCAAGAAGACCACCACCCTGAGCTACGACGT  
GTACAAGGACAAGCGGTTTCAGCGAGGACCAGTACGAGCTGCACATCCCCATCGCCATCAACAAGTGCCCCAAGAAC  
ATCTTCAAGATCAACACCGAGGTGCGGGTGTGCTGCTGAAGCACGACGACAACCCCTACGTGATCGGCATCGACCGGG  
GCGAGCGGAACCTGCTGTACATCGTGGTGGTGGACGGCAAGGGCAACATCGTGGAGCAGTACAGCCTGAACGAGAT  
CATCAACAACCTTCAACGGCATCCGGATCAAGACCGACTACCACAGCCTGCTGGACAAGAAGGAGAAGGAGCGGTTTC  
GAGGCCCGGCAGAACTGGACCAGCATCGAGAACATCAAGGAGCTGAAGGCCGGCTACATCAGCCAGGTGGTGCACA  
AGATCTGCGAGCTGGTGGAGAAGTACGACCGCGTATCGCCCTGGAGGACCTGAACAGCGGCTTCAAGAACAGCCG  
GGTGAAGGTGGAGAAGCAGGTGTACCAGAAGTTCGAGAAGATGCTGATCGACAAGCTGAACTACATGGTGGACAAG  
AAAAGCAACCCCTGCGCCACCGGGCGGCCCTGAAGGGCTACCAGATCACCAACAAGTTCGAGAGCTTCAAGAGCA  
TGAGCACCCAGAACGGCTTCATCTTCTACATCCCCGCTGGCTGACCAGCAAGATCGACCCAGCACCAGCTTCGT  
GAACCTGCTGAAGACCAAGTACACCAGCATCGCCGACAGCAAGAAGTTCATCAGCAGCTTCGACCGGATCATGTAC  
GTGCCCAGGAGGACCTGTTTCGAGTTCGCCCTGGACTACAAGAACTTCAGCCGGACCGACCGGACTACATCAAGA  
AGTGAAGCTGTACAGCTACGGCAACCGGATCCGGATCTTCGGAACCCCAAGAAGAACAACGTGTTTCGACTGGGA  
GGAGGTGTGCTGACCAGCGCCTACAAGGAGCTGTTCAACAAGTACGGCATCAACTACCAGCAGGGCGACATCCGG

GCCCTGCTGTGCGAGCAGAGCGACAAGGCCTTCTACAGCAGCTTCATGGCCCTGATGAGCCTGATGCTGCAGATGC  
GGAACAGCATCACCGGCCGACCGACGTGGACTTCTGATCAGCCCCGTGAAGAACAGCGACGGCATCTTCTACGA  
CAGCCGGAACTACGAGGCCAGGAGAACGCCATCCTGCCAAGAACGCCGACGCCAACGGCGCCTACAACATCGCC  
CGGAAGGTGCTGTGGGCCATCGGCCAGTTCAAGAAGGCCGAGGACGAGAAGCTGGACAAGGTGAAGATCGCCATCA  
GCAACAAGGAGTGGCTGGAGTACGCCAGACCAGCGTGAAGCACAAGCGGCCCGCCGCCACCAAGAAGGCCGCCA  
GGCCAAGAAGAAGAAGGGCAGCTACCCCTACGACGTGCCCGACTACGCCTACCCCTACGACGTGCCCGACTACGCC  
TACCCCTACGACGTGCCCGACTACGCCTGA (SEQ ID NO:3723)

[0267] 由示于SEQ ID NO:3723中的核苷酸序列编码的相应氨基酸序列示于下文提供的  
SEQ ID NO:3726中。

[0268] MSKLEKFTNCYLSKTLRFKAIPVGKTQENIDNKRLLEVEDEKRAEDYKGVKKLLDRYLSFINDVLHS  
IKLKNLNNYISLFRKKTRTEKENKELENLEINLRKEIAKAFKGNENYKSLFKKDI IETILPEFLDDKDEIALVNSF  
NGFTTAFTGFFDNRENMFSEEAKSTSI AFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFF  
EGEFFNFVLTQEGIDVYNAI IGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVLSDRESLSFYGEGYTS  
DEEVLEVFRNTLNKNSIFSSIKKLEKLFKNFDEYSSAGIFVKNGPAISTISKDIFGEWNVIRDKWNAEYDDIHLK  
KKAVVTEKYEDDRRKSFKKIGSFSLEQLQEYADADLSVVEKLKEII IQKVDEIYKVGSSSEKLFDAFDVLEKSLKK  
NDAVVAIMKDLLDSVKSFENYKAFKGGEGKTNRDESFYGFVFLAYDILLKVDHIYDAIRNYVTQKPYSKDKFKLY  
FQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYAKCLQKIDKDDVNGNYEKINYKLLPGPNKMLPKVFFSK  
KWMAYNPSEDIQKIYKNGTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVEEQ  
GYKVSFESASKKEVDKLEEGKLYMFQIYNKDFSDKSHGTPNLHTMYFKLLFDENNHGQIRLSGGAELFMRRASLK  
KEELVVHPANSP IANKNPDNPKTTTLSYDVYKDKRFSEDQYELHIPAIANKCPKNIFKINTEVRVLLKHDDNPYV  
IGIDRGERNLLYIVVVDGKGNIVEQYSLNEI INNFGIRIKTDYHSLDDKKEKERFEARQNWTSIENIKELKAGYI  
SQVVHKICELV EKYDAVIALEDLNSGFKNRDKVVEKQVYQKFEKMLIDKLNVMVDKSNPCATGGALKGYQITNKF  
ESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNLLKTKYTSIADSKKFISSFDRIMYVPEEDLFEFALDYKNFSRTD  
ADYIKKWKLYSYGNRIRIFRNPKNVDFWEEVCLTSAYKELFNKYGINYQQGDIRALLCEQSDKAFYSSFMALMS  
LMLQMRNSITGRDQVDFLISPVKNSDGFYDSRNYEAQENAILPKNADANGAYNIARKVLWAIGQFKKAEDEKLDK  
VKIAISNKEWLEYAQTQSVKH (SEQ ID NO:3726)

[0269] 编码Lb2Cpf1的示例性人密码子优化的核酸序列示于下文提供的SEQ ID NO:3724  
中。

[0270] ATGTACTACGAGAGCCTGACCAAGCAGTACCCCGTGAGCAAGACCATCCGGAACGAGCTGATCCCCAT  
CGGCAAGACCCTGGACAACATCCGGCAGAACAACATCCTGGAGAGCGACGTGAAGCGGAAGCAGAACTACGAGCAC  
GTGAAGGGCATCCTGGACGAGTACCACAAGCAGCTGATCAACGAGGCCCTGGACAACCTGCACCCTGCCAGCCTGA  
AGATCGCCGCCGAGATCTACCTGAAGAACCAGAAGGAGGTGAGCGACCGGGAGGACTTCAACAAGACCCAGGACCT  
GCTGCGGAAGGAGGTGGTGGAGAAGCTGAAGGCCACGAGAACTTACCAAGATCGGCAAGAAGGACATCCTGGAC  
CTGCTGGAGAAGCTGCCAGCATCAGCGAGGACGACTACAACGCCCTGGAGAGCTTCCGGAACCTTACACCTACT  
TCACCAGCTACAACAAGGTGCGGGAGAACCTGTACAGCGACAAGGAGAAAAGCAGCACCGTGGCCTACCGGCTGAT  
CAACGAGAACTTCCCAAGTTCCTGGACAACGTGAAAAGCTACCGGTTCTGTAAGACCGCCGGCATCCTGGCCGAC  
GGCCTGGGCGAGGAGGAGCAGGACAGCCTGTTTCATCGTGGAGACCTTCAACAAGACCCTGACCCAGGACGGCATCG  
ACACCTACAACAGCCAGGTGGGCAAGATCAACAGCAGCATCAACCTGTACAACCAGAAGAACCAGAAGGCCAACGG  
CTTCCGGAAGATCCCAAGATGAAGATGCTGTACAAGCAGATCCTGAGCGACCGGGAGGAGAGCTTCATCGACGAG



[0271] 由示于SEQ ID NO:3724中的核苷酸序列编码的相应氨基酸序列示于下文提供的SEQ ID NO:3727中。

[0272] MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKKRQNYEHVKGILDEYHKQLINEALDNC  
TLPSLKIAAAEIYLNQKEVSDREDFNKTQDLLRKEVVEKLKAHENFTKIGKKDILDLEKLPSISEDDYNALESFR  
NFYTYFTSYNKVRENLYSDKEKSSTVAYRLINENFPKFLDNVKSYPVKTAGILADGLGEEEQDSLFIIVETFNKTL  
TQDGDITYNSQVGKINSSINLYNQKNQKANGFRKIPKMKMLYKQILSDREESFIDEFQSDEVLIDNVESYGSVLIE  
SLKSSKVS AFFDALRESKGNVYKNDLAKTAMSNI VFENWRTFDDLLNQEYDLANENKKKDDKYFEKRQKELKKN  
KSYSLEHL CNLSEDCNLIENYIHQISDDIENI I INNETFLRIVINEHDRSRKLAKNRKAVKAIKDFLDSIKVLER  
ELKLINSSGQELEKDLIVYSAHEELLVELKQVDSLYNMTRNYLTKKPFSTEKVKLNFNRSTLLNGWDRNKETDNLG  
VLLLKDGKY YLGIMNTSANKAFVNPVAKTEKVFKKVDYKLLPVPNQMLPKVFFAKSNIDFYNPSSEIYSNYKKG  
HKKGNMFSLEDCHNLIDFFKESISKHEDWSKFGFKFSDTASYNDISEFYREVEKQGYKLYTDTIDETYINDLIERN  
ELYLFQIYNKDFSMYSGKLNHLTYFMMLFDQRNIDDVVYKLNGEAEVYRPA SISEDELI IHKAGEEIKNKNPN  
RARTKETSTFSYDIVKDKRYSKDKFTLHIPITMNFVDEVKRFNDVNSAIRIDENNVIGIDRGERNLLYVVVID  
SKGNILEQISLNSI INKEYDIETDYHALLDEREGGRDKARKDWNTVENIRDLKAGYLSQVVNVAKLVLYNAI IC  
LEDLNFQFKRGRQKVEKQVYQKFEKMLIDKLNLYVIDKSREQTSPKELGGALNALQLTSKFKSFKELGKQSGVIYY  
VPAYLTSKIDPTTG FANLFYMKCENVEKSKRFFDGFDFIRFNALENVFEFGFDYRSFTQRACGINSKWTVCTNGER  
IIKYRNPDKNNMFDEKVVVVVTDEMKNLFEQYKIPYEDGRNVKDMI ISNEEA E FYRRLYRLLQQTLQMRNSTSDGTR  
DYI I SPVKNKREAYFNSELSDGSPKADANGAYNIARKGLWVLEQIRQKSEGEKINLAMTNAEWLEYA QTHLL  
(SEQ ID NO:3727)

[0273] 在空间定位上远离在Cpf1活性中所涉及的区域(例如,与靶核酸分子和/或gRNA相互作用)的较不保守的或不保守的区域代表作为用于缺失而基本上不影响或降低Cpf1活性的候选物的区域或结构域。

[0274] 对于序列比较,典型地一个序列作为参比序列,测试序列与其进行比较。当使用序列比较算法时,将测试序列和参比序列输入计算机,指定子序列坐标(如果需要的话),并且指定序列算法程序参数。可以使用默认程序参数,或者指定备选参数。序列比较算法然后基于程序参数计算测试序列相对于参比序列的百分比序列一致性。用于比较的序列比对方法在本领域是熟知的。可以例如通过Smith和Waterman, (1970) Adv. Appl. Math. [应用数学进展]2:482c的局部同源性算法、通过Needleman和Wunsch, (1970) J. Mol. Biol. [分子生物学杂志]48:443的同源比对算法、通过Pearson和Lipman, (1988) Proc. Nat' l. Acad. Sci. USA [美国国家科学院院刊]85:2444的检索相似性方法、通过这些算法的计算机化实现(GAP、BESTFIT、FASTA和TFASTA,在威斯康星遗传软件包(Wisconsin Genetics Software Package)中,遗传学计算机组(Genetics Computer Group), 575 Science Dr., 麦迪逊,威斯康星州)或通过人工比对和目视检查(参见例如,Brent等人, (2003) Current Protocols in Molecular Biology [分子生物学实验指南])进行用于比较的序列的最佳比对。

[0275] 适于确定百分比序列一致性和序列相似性的算法的两个实例是BLAST和BLAST 2.0算法,其分别描述于Altschul等人, (1977) Nuc. Acids Res. [核酸研究]25:3389-3402;和Altschul等人, (1990) J. Mol. Biol. [分子生物学杂志]215:403-410中。用于执行BLAST分析的软件可通过美国国家生物技术信息中心(National Center for Biotechnology Information)公开地获得。

[0276] 也可以使用已经被整合到ALIGN程序(2.0版)中的E.Meyers和W.Miller, (1988) Comput. Appl. Biosci. [计算机应用生物科学] 4:11-17的算法,使用PAM120权重残基表、12的空位长度罚分和4的空位罚分确定两个氨基酸序列之间的百分比一致性。此外,可以使用已经被整合到GCG软件包中的GAP程序(可在www.gcg.com获得)中的Needleman和Wunsch (1970) J. Mol. Biol. [分子生物学杂志] 48:444-453算法,使用Blossom 62矩阵或PAM250矩阵,以及16、14、12、10、8、6或4的空位权重和1、2、3、4、5或6的长度权重确定两个氨基酸序列之间的百分比一致性。

[0277] IV. 候选分子的功能分析

[0278] 可以通过本领域已知的方法或如本文所描述的来评估候选Cpf1分子、候选gRNA分子、候选Cpf1分子/gRNA分子复合物。例如,用于评估Cpf1分子的内切核酸酶活性的示例性方法描述于例如Zetsche等人, Cell [细胞] (2015); 163:759-771中。

[0279] 结合和切割测定:测试Cpf1分子的内切核酸酶活性

[0280] 可以在质粒切割测定中对Cpf1分子/gRNA分子复合物结合至并且切割靶核酸的能力进行评估。在此测定中,在反应之前通过加热至95°C并且缓慢冷却至室温,将合成或体外转录的gRNA分子预退火。在37°C下,将天然或限制酶切消化-线性化的质粒DNA (300ng (约8nM)) 用纯化的Cpf1蛋白分子 (50-500nM) 和gRNA (50-500nM, 1:1) 在具有或不具有10mM MgCl<sub>2</sub>的Cpf1质粒切割缓冲液 (20mM HEPES pH 7.5, 150mM KCl, 0.5mM DTT, 0.1mM EDTA) 中孵育60min。用5X DNA加样缓冲液 (30%甘油, 1.2% SDS, 250mM EDTA) 终止反应,通过0.8%或1%琼脂糖凝胶电泳进行解析并且通过溴化乙锭染色进行可视化。所得切割产物指示所述Cpf1分子是切割两条DNA链还是仅切割两条链中的一条。例如,线状DNA产物指示切割两条DNA链。带切口的开环产物指示仅切割两条链中的一条。

[0281] 可替代地,可以在寡核苷酸DNA切割测定中对Cpf1分子/gRNA分子复合物结合至并且切割靶核酸的能力进行评估。在此测定中,在37°C下,在50μL反应中,通过用在1X T4多核苷酸激酶反应缓冲液中的5单位T4多核苷酸激酶以及约3-6pmol (约20-40mCi) [γ-<sup>32</sup>P]-ATP孵育30min,对DNA寡核苷酸 (10pmol) 进行放射性标记。在热灭活后 (65°C持续20min),通过柱对反应进行纯化以去除未结合的标记。通过在95°C下用等摩尔量的未标记的互补寡核苷酸退火标记的寡核苷酸持续3min,随后缓慢冷却至室温来产生双链体底物 (100nM)。对于切割测定,通过加热至95°C持续30s,随后缓慢冷却至室温对gRNA分子进行退火。在9μl的总体积中,将Cpf1 (500nM终浓度) 与退火的gRNA分子 (500nM) 在切割测定缓冲液 (20mM HEPES pH 7.5, 100mM KCl, 5mM MgCl<sub>2</sub>, 1mM DTT, 5%甘油) 中进行预孵育。通过添加1μl靶DNA (10nM) 开始反应并且在37°C下孵育1h。通过添加20μl的加样染料 (5mM EDTA, 0.025% SDS, 5%甘油, 在甲酰胺中) 淬灭反应并且加热至95°C持续5min。将切割产物在含有7M脲的12%变性聚丙烯酰胺凝胶上进行解析并且通过磷成像进行可视化。所得切割产物指示互补链、非互补链还是两者被切割。

[0282] 这些测定中的一个或两个可以用于评估候选gRNA分子或候选Cpf1分子的适合性。

[0283] 结合测定:测试Cpf1分子与靶DNA的结合

[0284] 用于评估Cpf1分子与靶DNA的结合的示例性方法描述于例如Zetsche等人, Cell [细胞] (2015); 163:759-771中。

[0285] 例如,在电泳迁移率变动测定中,通过在去离子水中混合每条链 (10nmol)、加热至

95°C持续3min并且缓慢冷却至室温而形成靶DNA双链体。将所有DNA在含有1X TBE的8%非变性凝胶上进行纯化。将DNA条带通过UV遮蔽进行可视化,切除,并且通过将凝胶片浸泡在DEPC处理的H<sub>2</sub>O中进行洗脱。将洗脱的DNA进行乙醇沉淀并且溶解在DEPC处理的H<sub>2</sub>O中。在37°C下,使用T4多核苷酸激酶将DNA样品用[ $\gamma$ -<sup>32</sup>P]-ATP进行5'端标记持续30min。将多核苷酸激酶在65°C下热变性20min,并且使用柱去除未结合的放射性标记。在10 $\mu$ l的总体积中,在含有20mM HEPES pH 7.5、100mM KCl、5mM MgCl<sub>2</sub>、1mM DTT和10%甘油的缓冲液中进行结合测定。用等摩尔量的预退火的gRNA分子对Cpf1蛋白分子进行程序化,并且从100pM滴定至1 $\mu$ M。将放射性标记的DNA添加至20pM的终浓度。将样品在37°C下孵育1h并且在4°C下在含有1X TBE和5mM MgCl<sub>2</sub>的8%天然聚丙烯酰胺凝胶上进行解析。将凝胶干燥并且通过磷成像进行DNA可视化。

#### [0286] 差示扫描荧光测定法 (DSF)

[0287] 可以经由DSF测量Cpf1-gRNA核糖核蛋白 (RNP) 复合物的热稳定性。这种技术测量蛋白质的热稳定性,其可以在有利条件下(如添加结合RNA分子,例如gRNA)增加。

[0288] 使用两种不同的方案进行测定,一种方案用于测试gRNA:Cpf1蛋白的最佳化学计量比,并且另一种方案用于确定RNP形成的最佳溶液条件。

[0289] 为了确定形成RNP复合物的最佳溶液,将水+10x SYPRO Orange® (生命技术公司 (Life Technologies) 目录#S-6650) 中的Cpf1的2 $\mu$ M溶液分配到384孔板中。然后添加稀释于溶液中的具有不同pH和盐的等摩尔量的gRNA。在室温下孵育10'并短暂离心以去除任何气泡之后,使用带有Bio-Rad CFX Manager软件的Bio-Rad CFX384™ Real-Time System C1000Touch™热循环仪运行从20°C至90°C的梯度,其中温度每10秒增加1°。

[0290] 第二次测定由在来自上述测定1的最适缓冲液中混合不同浓度的gRNA与2 $\mu$ M Cpf1并在384孔板中于RT下孵育10'组成。添加等体积的最适缓冲液+10x SYPRO Orange® (生命技术公司目录#S-6650) 并且将板用Microseal®B粘合剂 (MSB-1001) 密封。短暂离心以去除任何气泡后,使用带有Bio-Rad CFX Manager软件的Bio-Rad CFX384™ Real-Time System C1000Touch™热循环仪运行从20°C至90°C的梯度,其中温度每10秒增加1°。

#### [0291] V. 基因组编辑途径

[0292] 通常,应当理解的是根据本文所描述的方法对任何基因的改变可以由任何机制介导并且任何方法都不限于具体机制。可以与基因的改变相关的示例性机制包括但不限于非同源末端连接(例如,经典的或替代的)、微同源性介导的末端连接(MMEJ)、同源定向修复(例如,内源供体模板介导的)以及SDSA(合成依赖性链退火)。本文描述了用于使用NHEJ靶向敲除FAS、BID、CTLA4、PDCD1、CBLB或PTPN6、B2M、TRAC和/或TRBC基因的一个或两个等位基因的示例性方法(参见第V.1部分)。在某些实施例中,所披露的方法可以靶向FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的两个或更多个用于敲除。

#### [0293] V.1用于基因靶向的NHEJ途径

[0294] 如本文所描述的,核酸酶诱导的非同源末端连接(NHEJ)可以用于靶向基因特异性敲除。核酸酶诱导的NHEJ还可以用于去除(例如,缺失)感兴趣基因的序列插入。

[0295] 在某些实施例中,与本文所描述的方法相关的基因组改变依赖于核酸酶诱导的NHEJ和NHEJ修复途径的易错性质。NHEJ通过将两端连接在一起修复DNA中的双链断裂;然而,通常,只有两个相容端(完全如它们通过双链断裂所形成的)是完全连接的,原始序列才

被恢复。在末端重新连接之前，双链断裂的DNA端常常是酶加工的受试者，在一条或两条链处产生核苷酸的添加或去除。这使得NHEJ修复位点处的DNA序列中存在插入和/或缺失(indel)突变。典型地，这些突变中的三分之二改变阅读框并且因此产生非功能蛋白。另外，维持阅读框但插入或缺失大量的序列的突变可以破坏蛋白质的功能性。这是基因座依赖性的，因为关键功能结构域中的突变可能比蛋白质的非关键区域中的突变耐受性低。

[0296] 由NHEJ产生的indel突变在性质上是不可预测的；然而，在给定的断裂位点处，某些indel序列是有利的并且在群体中过度表达，这可能归因于微同源性的小区域。缺失的长度可以广泛变化；最常见地在1-50bp范围内，但是它们可以容易地达到大于100-200bp。插入往往是较短的并且常常包括紧密围绕断裂位点的序列的短的重叠。然而，有可能获得大的插入，并且在这些情况下，插入序列通常已经被追溯至基因组的其他区域或至存在于细胞中的质粒DNA。

[0297] 因为NHEJ是诱变的过程，它还可以用于缺失小序列基序，只要不需要产生特异性最终序列。如果将双链断裂靶向短靶序列附近，由NHEJ修复引起的缺失突变通常跨越，并且因此去除不想要的核苷酸。对于较大的DNA区段的缺失，引入两个双链断裂(序列的每侧上一个双链断裂)可以在末端之间产生NHEJ，其中去除了整个间插序列。这两种途径都可以用于缺失特异性DNA序列；然而，NHEJ的易出错性质仍可能在修复位点处产生indel突变。

[0298] 被靶向感兴趣基因(例如基因的编码区，例如早期编码区)的NHEJ介导的indel可以用于敲除感兴趣的基因(即，消除其表达)。例如，感兴趣基因的早期编码区包括紧跟着转录起始位点、在编码序列的第一外显子内或在转录起始位点的500bp(例如，小于500、450、400、350、300、250、200、150、100或50bp)内的序列。

[0299] 在某些实施例中，将NHEJ介导的indel引入一个或多个T细胞表达的基因中。在Cpf1双链核酸酶用于将突变引入两个T细胞表达的基因(例如，FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的任何两个)的某些实施例中，提供了靶向两个基因的单体的gRNA或gRNA对连同所述Cpf1双链核酸酶。在Cpf1双链核酸酶用于将突变引入三个T细胞表达的基因(例如，FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的任何三个)的某些实施例中，提供了靶向所有三个基因的单体的gRNA或gRNA对连同所述Cpf1双链核酸酶。在Cpf1双链核酸酶用于将突变引入四个T细胞表达的基因(例如，FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的任何四个)的某些实施例中，提供了靶向所有四个基因的单体的gRNA或gRNA对连同所述Cpf1双链核酸酶。在Cpf1双链核酸酶用于将突变引入五个T细胞表达的基因(例如，FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的任何五个)的某些实施例中，提供了靶向所有五个基因的单体的gRNA或gRNA对连同所述Cpf1双链核酸酶。在Cpf1双链核酸酶用于将突变引入六个T细胞表达的基因(例如，FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的任何六个)的某些实施例中，提供了靶向所有六个基因的单体的gRNA或gRNA对连同所述Cpf1双链核酸酶。在Cpf1双链核酸酶用于将突变引入七个T细胞表达的基因(例如，FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的任何七个)的某些实施例中，提供了靶向所有七个基因的单体的gRNA或gRNA对连同所述Cpf1双链核酸酶。在Cpf1双链核酸酶用于将突变引入八个T细胞表达的基因(例如，FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和TRBC基因中的每一个)的某些实施例中，提供了靶向所有八个基因的单体的gRNA或gRNA对连同所述Cpf1双链

核酸酶。

[0300] 双链断裂相对于靶位置的布置

[0301] 在gRNA和Cpf1核酸酶产生双链断裂用于诱导NHEJ介导的indel的目的的某些实施例中,gRNA(例如,单模块化gRNA分子)被配置成将一个双链断裂定位在非常接近靶位置的核苷酸之处。在某些实施例中,所述切割位点在距靶位置0-30bp之间(例如,距靶位置小于30、小于25、小于20、小于15、小于10、小于9、小于8、小于7、小于6、小于5、小于4、小于3、小于2或1bp)。在某些实施例中,所述切割位点在距靶位置1-24bp之间。

[0302] 所述双链切割性Cpf1分子可以用于本文所描述的方法和组合物中以在靶位置的两侧均产生断裂。可以在靶位置的两侧产生双链断裂,以去除两个切口之间的核酸序列(例如,两个断裂之间的区域被缺失)。在某些实施例中,两种gRNA被配置成将双链断裂定位在靶位置的两侧。

[0303] V.2基因组编辑方法中的gRNA和Cpf1分子

[0304] 本文所披露的gRNA分子(例如第I部分中披露的那些)可以与本文所披露的Cpf1分子(例如第III部分中披露的那些)一起使用,其产生双链断裂以改变靶核酸的序列,例如靶位置或靶基因标签。在某些实施例中,例如当靶向产生双链断裂的Cpf1分子时,所述gRNA将双链断裂定位(i)在靶位置的50、100、150、200、250、300、350、400、450或500个核苷酸内,或(ii)足够接近使得所述靶位置在末端切除的区域内。

[0305] VI.靶细胞

[0306] 在多种细胞中,可以将Cpf1分子和gRNA分子(例如,Cpf1分子/gRNA分子复合物)用于操纵细胞,例如以编辑靶核酸。

[0307] 在某些实施例中,通过编辑一个或多个(例如,如本文所描述的)靶基因(例如,在其中诱导突变)来操纵细胞。在某些实施例中,调节一个或多个靶基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的表达。在某些实施例中,通过编辑一个或多个靶基因(例如,在其中诱导突变)和/或调节一个或多个靶基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)的表达来离体地操纵细胞,并向受试者给予。用于离体操纵的靶细胞的来源可以包括例如受试者的血液、受试者的脐血或受试者的骨髓。用于离体操纵的靶细胞的来源还可以包括例如异源供体血液、脐血或骨髓。

[0308] 可以将本文所描述的Cpf1和gRNA分子递送至靶细胞。在某些实施例中,所述靶细胞是T细胞(例如,CD8<sup>+</sup>T细胞(例如,CD8<sup>+</sup>天然T细胞、中枢记忆T细胞或效应记忆T细胞)、CD4<sup>+</sup>T细胞、自然杀伤T细胞(NKT细胞)、调节性T细胞(Treg)、干细胞记忆T细胞)、淋巴祖细胞、造血干细胞、自然杀伤细胞(NK细胞)或树突细胞。在某些实施例中,所述靶细胞是诱导多能干(iPS)细胞或衍生自iPS细胞的细胞(例如产生自受试者的iPS细胞),所述细胞被操纵以改变一个或多个靶基因(例如,FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因)(例如,在其中诱导突变)或操纵所述基因的表达,并分化成例如T细胞(例如,CD8<sup>+</sup>T细胞(例如,CD8<sup>+</sup>天然T细胞、中枢记忆T细胞或效应记忆T细胞)、CD4<sup>+</sup>T细胞、干细胞记忆T细胞)、淋巴祖细胞或造血干细胞。

[0309] 在某些实施例中,所述靶细胞已经被改变成含有特定T细胞受体(TCR)基因(例如,TRAC和TRBC基因)。在某些实施例中,TCR对肿瘤相关抗原(例如,癌胚抗原(CEA)、GP100、由T细胞1识别的黑色素瘤抗原(MART1)、黑色素瘤抗原A3(MAGEA3)、NYESO1或p53)具有结合特

异性。

[0310] 在某些实施例中,所述靶细胞已经被改变成含有特定嵌合抗原受体(CAR)。在某些实施例中,CAR对肿瘤相关抗原(例如,CD19、CD20、碳酸酐酶IX(CAIX)、CD171、CEA、ERBB2、GD2、 $\alpha$ -叶酸受体、Lewis Y抗原、前列腺特异性膜抗原(PSMA)或肿瘤相关糖蛋白72(TAG72))具有结合特异性。

[0311] 在某些实施例中,所述靶细胞已经被改变成例如通过TCR或CAR结合以下肿瘤抗原中的一种或多种。肿瘤抗原可以包括但不限于AD034、AKT1、BRAP、CAGE、CDX2、CLP、CT-7、CT8/HOM-TE5-85、cTAGE-1、Fibulin-1、HAGE、HCA587/MAGE-C2、hCAP-G、HCE661、HER2/neu、HLA-Cw、HOM-HD-21/半乳糖凝集素9、HOM-MEEL-40/SSX2、HOM-RCC-3.1.3/CAXII、HOXA7、HOXB6、Hu、HUB1、KM-HN-3、KM-KN-1、KOC1、KOC2、KOC3、KOC3、LAGE-1、MAGE-1、MAGE-4a、MPP11、MSLN、NNP-1、NY-BR-1、NY-BR-62、NY-BR-85、NY-CO-37、NY-CO-38、NY-ES0-1、NY-ES0-5、NY-LU-12、NY-REN-10、NY-REN-19/LKB/STK11、NY-REN-21、NY-REN-26/BCR、NY-REN-3/NY-CO-38、NY-REN-33/SNC6、NY-REN-43、NY-REN-65、NY-REN-9、NY-SAR-35、OGFr、PLU-1、Rab38、RBPJ $\kappa$ 、RHAMM、SCP1、SCP-1、SSX3、SSX4、SSX5、TOP2A、TOP2B或酪氨酸酶。

[0312] VII. 递送、配制品和给药途径

[0313] 可以使用多种递送方法和配制品将多种形式的组分(例如,Cpf1分子和gRNA分子)引入靶细胞中,参见例如表2和3。当Cpf1或gRNA组分被编码为DNA用于递送时,所述DNA可以典型地但不一定包括控制区(例如,包含启动子)以影响表达。Cpf1分子序列的有用启动子包括例如CMV启动子、EF-1a启动子、EFS启动子、MSCV启动子、PGK启动子、CAG启动子、骨骼 $\alpha$ 肌动蛋白启动子、肌肉肌酸激酶启动子、抗肌萎缩蛋白启动子、 $\alpha$ 肌球蛋白重链启动子以及平滑肌肌动蛋白启动子。gRNA的有用启动子包括例如H1、7SJ、EF-1a、tRNA或U6启动子。可以选择具有类似或不同强度的启动子来调谐组分的表达。编码Cpf1分子的序列可以包含核定位信号(NLS),例如SV40NLS。在某些实施例中,编码Cpf1分子的序列包含至少两个核定位信号。在某些实施例中,Cpf1分子或gRNA分子的启动子可以独立地是诱导型的、组织特异性的或细胞特异性的。

[0314] 表2-可以将组分递送至靶细胞的形式实例

[0315]

要素		注解
一种或多种 Cpf1 分子	一种或多种 gRNA 分子	
DNA	DNA	在某些实施例中，Cpf1 分子和 gRNA 转录自 DNA。在某些实施例中，它们在单独的分子上进行编码。
DNA		在某些实施例中，Cpf1 分子和 gRNA 转录自 DNA，这里转录自单分子。
DNA	RNA	在某些实施例中，Cpf1 分子转录自 DNA，并且 gRNA 被提供为体外转录或合成的 RNA。
mRNA	RNA	在某些实施例中，Cpf1 分子翻译自体外转录的 mRNA，并且 gRNA 被提供为体外转录或合成的 RNA。
mRNA	DNA	在某些实施例中，Cpf1 分子翻译自体外转录的 mRNA，并且 gRNA 转录自 DNA。
蛋白质	DNA	在某些实施例中，Cpf1 分子被提供为蛋白质，并且 gRNA 转录自 DNA。
蛋白质	RNA	在某些实施例中，Cpf1 分子被提供为蛋白质，并且 gRNA 被提供为转录或合成的 RNA。这种递送方法被称为“RNP 递送”。

[0316] 表3概括了CRISPR/Cpf1系统的组分(例如,如本文所描述的Cpf1分子组分和gRNA分子组分)的各种递送方法。

[0317] 表3

[0318]

递送载体/方式	递送到非分裂细胞中	表达的持续时间	基因组整合	所递送分子的类型

[0319]

递送载体/方式		递送到非分裂细胞中	表达的持续时间	基因组整合	所递送分子的类型
物理的 (例如, 电穿孔、基因枪、磷酸钙转染、细胞压缩或挤压)		是	瞬时的	否	核酸和蛋白质
病毒的	逆转录病毒	否	稳定的	是	RNA
	慢病毒	是	稳定的	是/否具有修饰	RNA
	腺病毒	是	瞬时的	否	DNA
	腺相关病毒 (AAV)	是	稳定的	否	DNA
	痘苗病毒	是	非常瞬时的	否	DNA
	单纯疱疹病毒	是	稳定的	否	DNA
非病毒的	阳离子脂质体	是	瞬时的	取决于递送什么	核酸和蛋白质
	聚合物纳米粒子	是	瞬时的	取决于递送什么	核酸和蛋白质
生物非病毒递送载体	减毒细菌	是	瞬时的	否	核酸
	工程化的噬菌体	是	瞬时的	否	核酸
	哺乳动物病毒样粒子	是	瞬时的	否	核酸
	生物脂质体: 红细胞血影和外泌体	是	瞬时的	否	核酸

[0320] 基于DNA的Cpf1分子和/或一种或多种gRNA分子的递送

[0321] 可以通过本领域已知的方法或如本文所描述的将编码Cpf1分子和/或gRNA分子的核酸组合物给予受试者或递送到细胞中。例如,编码Cpf1和/或编码gRNA的DNA可以通过例如通过载体(例如,病毒或非病毒载体)、非基于载体的方法(例如,使用裸DNA或DNA复合物)或其组合进行递送。

[0322] 在某些实施例中,编码Cpf1和/或gRNA的DNA是通过载体(例如,病毒载体/病毒或质粒)进行递送的。

[0323] 在某些实施例中,所述载体包含编码Cpf1分子和/或gRNA分子的序列。在某些实施例中,所述载体进一步包含编码融合到例如Cpf1分子序列上的信号肽(例如,用于核定位、核仁定位、线粒体定位)的序列。例如,载体可以包含融合到编码Cpf1分子的序列上的核定位序列(例如,来自SV40)。

[0324] 一个或多个调节/控制元件(例如,启动子、增强子、内含子、多腺苷酸化信号、Kozak共有序列、内部核糖体进入位点(IRES)、2A序列和剪接受体或供体)可以被包括在所述载体中。在某些实施例中,所述启动子由RNA聚合酶II识别(例如,CMV启动子)。在某些实施例中,所述启动子由RNA聚合酶III识别(例如,U6启动子)。在某些实施例中,所述启动子是受调节的启动子(例如,诱导型启动子)。在某些实施例中,所述启动子是组成型启动子。在某些实施例中,所述启动子是组织特异性启动子。在某些实施例中,所述启动子是病毒启动子。在某些实施例中,所述启动子是非病毒启动子。

[0325] 在某些实施例中,所述载体或递送运载体是病毒载体(例如,用于产生重组病毒)。在某些实施例中,所述病毒是DNA病毒(例如,dsDNA或ssDNA病毒)。在某些实施例中,所述病毒是RNA病毒(例如,ssRNA病毒)。示例性病毒载体/病毒包括例如逆转录病毒、慢病毒、腺病毒、腺相关病毒(AAV)、痘苗病毒、痘病毒以及单纯疱疹病毒。

[0326] 在某些实施例中,所述病毒感染分裂细胞。在某些实施例中,所述病毒感染非分裂细胞。在某些实施例中,所述病毒感染分裂细胞和非分裂细胞两者。在某些实施例中,所述病毒可以整合到宿主基因组中。在某些实施例中,所述病毒被工程化为具有降低的免疫性(例如,在人中)。在某些实施例中,所述病毒是有复制能力的。在某些实施例中,所述病毒是复制缺陷型的(例如,另外多轮的病毒粒子复制和/或包装所需的基因的一个或多个编码区被其他基因替换或缺失)。在某些实施例中,所述病毒引起Cpf1分子和/或gRNA分子的瞬时表达。在某些实施例中,所述病毒引起Cpf1分子和/或gRNA分子的持久(例如,至少1周、2周、1个月、2个月、3个月、6个月、9个月、1年、2年或永久)表达。所述病毒的包装能力可以在例如至少约4kb到至少约30kb(例如,至少约5kb、10kb、15kb、20kb、25kb、30kb、35kb、40kb、45kb或50kb)之间变化。

[0327] 在某些实施例中,所述病毒载体识别特定的细胞类型或组织。例如,所述病毒载体可以用不同/替代的病毒包膜糖蛋白进行假型包装;用细胞类型特异性受体进行工程化(例如,对一种或多种病毒包膜糖蛋白进行一个或多个基因修饰以结合靶向配体(如肽配体、单链抗体或生长因子));和/或被工程化为具有双重特异性的分子桥,其中一端识别病毒糖蛋白而另一端识别靶细胞表面的部分(例如,配体-受体、单克隆抗体、亲和素-生物素和化学缀合)。

[0328] 在某些实施例中,编码Cpf1和/或gRNA的DNA是通过重组逆转录病毒进行递送的。在某些实施例中,所述逆转录病毒(例如,莫洛尼(Moloney)鼠白血病病毒)包括(例如,允许

整合进宿主基因组中的) 逆转录酶。在某些实施例中, 所述逆转录病毒是有复制能力的。在某些实施例中, 所述逆转录病毒是复制缺陷型的 (例如, 另外多轮的病毒粒子复制和包装所需的基因的一个或多个编码区被其他基因替换或缺失)。

[0329] 在某些实施例中, 编码Cpf1和/或gRNA的DNA是通过重组慢病毒进行递送的。例如, 所述慢病毒是复制缺陷型的 (例如, 不包含病毒复制所需的一个或多个基因)。

[0330] 在某些实施例中, 编码Cpf1和/或gRNA的DNA是通过重组腺病毒进行递送的。在某些实施例中, 所述腺病毒被工程化为在人中具有降低的免疫性。在某些实施例中, 编码Cpf1和/或gRNA的DNA是通过重组AAV进行递送的。在某些实施例中, 所述AAV不将其基因组整合到宿主细胞 (例如, 本文所描述的靶细胞) 的基因组中。在某些实施例中, 所述AAV可以将其基因组整合到宿主细胞 (例如, 如本文所描述的靶细胞) 的基因组中。在某些实施例中, 所述AAV是自我互补腺相关病毒 (scAAV) (例如, 对一起退火以形成双链DNA的两条链进行包装的scAAV)。可以用于所披露的方法中的AAV血清型包括AAV1、AAV2、经修饰的AAV2 (例如, 在Y444F、Y500F、Y730F和/或S662V处修饰)、AAV3、经修饰的AAV3 (例如, 在Y705F、Y731F和/或T492V处修饰)、AAV4、AAV5、AAV6、经修饰的AAV6 (例如, 在S663V和/或T492V处修饰)、AAV8、AAV 8.2、AAV9、AAV rh 10, 并且假型AAV (如AAV2/8、AAV2/5和AAV2/6) 也可以在所披露的方法中使用。在某些实施例中, 可以用于本文所描述的方法中的AAV衣壳是来自血清型AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV.rh8、AAV.rh10、AAV.rh32/33、AAV.rh43、AAV.rh64R1或AAV7m8的衣壳序列。

[0331] 在某些实施例中, 编码Cpf1和/或gRNA的DNA以再工程化AAV衣壳递送, 其例如与来自血清型AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV.rh8、AAV.rh10、AAV.rh32/33、AAV.rh43或AAV.rh64R1的衣壳序列具有50%或更大 (例如, 60%或更大、70%或更大、80%或更大、90%或更大或者95%或更大) 的序列同源性。

[0332] 在某些实施例中, 编码Cpf1和/或gRNA的DNA是通过嵌合AAV衣壳进行递送的。示例性的嵌合AAV衣壳包括但不限于AAV9i1、AAV2i8、AAV-DJ、AAV2G9、AAV2i8G9或AAV8G9。

[0333] 在某些实施例中, 编码Cpf1和/或gRNA的DNA是通过杂合病毒 (例如, 本文所描述的一种或多种病毒的杂合体) 进行递送的。在某些实施例中, 所述杂合病毒是AAV (例如, 任何AAV血清型) 与博卡病毒、B19病毒、猪AAV、鹅AAV、猫AAV、犬AAV或MVM的杂合体。

[0334] 使用包装细胞形成能够感染靶细胞的病毒粒子。这样的细胞包括可以包装腺病毒的293细胞和可以包装逆转录病毒的 $\psi$ 2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生产者细胞系产生。所述载体典型地含有包装以及随后整合进宿主或靶细胞 (如果适用的话) 所需的最低量病毒序列, 而其他病毒序列由编码有待表达的蛋白质 (例如Cpf1) 的表达盒替换。例如, 在基因疗法中使用的AAV载体典型地仅具有来自AAV基因组的反向末端重复 (ITR) 序列, 所述序列为包装并在宿主或靶细胞中基因表达所需。失去的病毒功能由包装细胞系反式地提供。此后, 将病毒DNA包装进如下细胞系中, 所述细胞系含有编码辅助质粒的其他AAV基因, 即rep和cap, 但缺少ITR序列。所述细胞系还被作为辅助者的腺病毒感染。所述辅助病毒促进AAV载体的复制和从辅助质粒表达AAV基因。由于缺少ITR序列, 未以显著的量包装所述辅助质粒。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。

[0335] 在某些实施例中, 所述病毒载体具有识别细胞类型的能力。例如, 所述病毒载体可

以用不同/替代的病毒包膜糖蛋白进行假型包装;用细胞类型特异性受体进行工程化(例如,对病毒包膜糖蛋白进行基因修饰以结合靶向配体(如肽配体、单链抗体、生长因子));和/或被工程化为具有双重特异性的分子桥,其中一端识别病毒糖蛋白而另一端识别靶细胞表面的部分(例如,配体-受体、单克隆抗体、亲和素-生物素和化学缀合)。

[0336] 在某些实施例中,所述病毒载体实现细胞类型特异性表达。例如,可以构建组织特异性启动子以仅在特定靶细胞中限制转基因(Cpf1和gRNA)的表达。所述载体的特异性也可以由转基因表达的微小RNA依赖性控制所介导。在某些实施例中,所述病毒载体具有增加的病毒载体和靶细胞膜的融合效率。例如,可以结合融合蛋白(如融合感受态血球凝集素(HA))以增加病毒摄取进入细胞中。在某些实施例中,所述病毒载体具有核定位的能力。例如,可以将需要分解核膜(在细胞分裂期间)并且因此将不感染非分裂细胞的病毒改变成结合病毒的基质蛋白中的核定位肽,由此能够实现非增殖细胞的转导。

[0337] 在某些实施例中,编码Cpf1和/或gRNA的DNA是通过非基于载体的方法(例如,使用裸DNA或DNA复合物)进行递送的。例如,DNA可以例如通过有机改性的二氧化硅或硅酸盐(Ormosil)、电穿孔、瞬时细胞压缩或挤压(例如,如在Lee等人[2012]Nano Lett[纳米快报]12:6322-27中所描述的)、基因枪、声致穿孔、磁转染、脂质介导的转染、树枝状聚合物、无机纳米粒子、磷酸钙或其组合进行递送。

[0338] 在某些实施例中,经由电穿孔递送包括将细胞与编码Cpf1和/或gRNA的DNA在盒、室或比色皿中混合并且施加一个或多个限定持续时间和幅度的电脉冲。在某些实施例中,使用如下系统进行经由电穿孔的递送,在所述系统中将细胞与编码Cpf1和/或gRNA的DNA在连接至装置(例如,泵)的容器中混合,所述装置向盒、室或比色皿中给料混合物,在所述盒、室或比色皿中施加一个或多个限定持续时间和幅度的电脉冲,之后将细胞递送至第二容器。

[0339] 在某些实施例中,编码Cpf1和/或gRNA的DNA是通过载体和非基于载体的方法的组合进行递送的。例如,病毒体包含与灭活病毒(例如,HIV或流感病毒)组合的脂质体,这可以导致比单独的病毒或脂质体方法更有效的基因转移。

[0340] 在某些实施例中,所述递送运载体是非病毒载体。在某些实施例中,所述非病毒载体是无机纳米粒子。示例性的无机纳米粒子包括例如磁性纳米粒子(例如, $\text{Fe}_3\text{MnO}_2$ )和二氧化硅。可以将纳米粒子的外表面与带正电荷的聚合物(例如,聚乙烯亚胺、聚赖氨酸、聚丝氨酸)缀合,这允许有效载荷的附接(例如,缀合或截留)。在某些实施例中,所述非病毒载体是有机纳米粒子。示例性的有机纳米粒子包括例如含有阳离子脂质连同中性辅助脂质的SNALP脂质体(其涂覆有聚乙二醇(PEG))以及涂覆有脂质涂层的鱼精蛋白-核酸复合物。用于基因转移的示例性脂质示于下表4中。

[0341] 表4-用于基因转移的脂质

[0342]

脂质	缩写	特征
1,2-二油酰基-sn-甘油基-3-磷脂酰胆碱	DOPC	辅助性的
1,2-二油酰基-sn-甘油基-3-磷脂酰乙醇胺	DOPE	辅助性的
胆固醇		辅助性的
<i>N</i> -[1-(2,3-二油烯基氧基)丙基] <i>N,N,N</i> -三甲基氯化铵	DOTMA	阳离子的

[0343]

脂质	缩写	特征
1,2-二油烯基氧基-3-三甲基铵-丙烷	DOTAP	阳离子的
双十八烷基氨基甘氨酸精胺	DOGS	阳离子的
<i>N</i> -(3-氨基丙基)- <i>N,N</i> -二甲基-2,3-双(十二烷氧基)-1-溴化丙铵	GAP-DLRIE	阳离子的
溴化十六烷基三甲基铵	CTAB	阳离子的
6-月桂氧己基鸟氨酸酯	LHON	阳离子的
1-(2,3-二油酰基氧基丙基)-2,4,6-三甲基吡啶鎓	2Oc	阳离子的
2,3-二油烯基氧基- <i>N</i> -[2(精胺甲酰胺基-乙基)- <i>N,N</i> -二甲基-1-三氟乙酸丙铵	DOSPA	阳离子的
1,2-二油烯基-3-三甲基铵-丙烷	DOPA	阳离子的
<i>N</i> -(2-羟基乙基)- <i>N,N</i> -二甲基-2,3-双(十四烷氧基)-1-溴化丙铵	MDRIE	阳离子的
二肉豆蔻酰氧基丙基二甲基羟基乙基溴化铵	DMRI	阳离子的
3β-[ <i>N,N'</i> -二甲基氨基乙烷]-氮甲酰基]胆固醇	DC-Chol	阳离子的
双-胍鎓-三氨基乙胺-胆固醇	BGTC	阳离子的
1,3-双脱氧基-2-(6-羧基-精胺基)-丙酰胺	DOSPER	阳离子的
二甲基十八烷基溴化铵	DDAB	阳离子的
双十八烷基氨基甘氨酸基亚精胺	DSL	阳离子的
外消旋-[(2,3-双十八烷氧基丙基)(2-羟基乙基)]-二甲基氯化铵	CLIP-1	阳离子的
外消旋-[2(2,3-双十六烷氧基丙基-氧基甲氧基)乙基]三甲基溴化铵	CLIP-6	阳离子的
乙基二肉豆蔻酰磷脂酰胆碱	EDMPC	阳离子的
1,2-二硬脂氧基- <i>N,N</i> -二甲基-3-氨基丙烷	DSDMA	阳离子的
1,2-二肉豆蔻酰基-三甲基铵丙烷	DMTAP	阳离子的
<i>O,O'</i> -二肉豆蔻基- <i>N</i> -赖氨酸基天冬氨酸酯	DMKE	阳离子的
1,2-二硬脂酰基- <i>sn</i> -甘油基-3-乙基磷酸胆碱	DSEPC	阳离子的
<i>N</i> -棕榈酰 D-赤-鞘氨苷甲酰基-精胺	CCS	阳离子的
<i>N</i> -叔丁基- <i>N</i> O-十四烷基-3-十四烷基氨基丙脒	二 C14-脒	阳离子的

[0344]

脂质	缩写	特征
十八碳烯醇基氧基[乙基-2-十七烯基-3 羟基乙基]氯化咪唑啉鎓	DOTIM	阳离子的
N1-胆固醇氧基羰基-3,7-二氮杂壬烷-1,9-二胺	CDAN	阳离子的
2-(3-[双(3-氨基-丙基)-氨基]丙胺基)-N-双十四烷基氨基甲酰基甲-乙基-乙酰胺	RPR209120	阳离子的
1,2-二亚油基氧基-3-二甲基氨基丙烷	DLinDMA	阳离子的
2,2-二亚油基-4-二甲基氨基乙基-[1,3]-二氧戊环	DLin-KC2-DMA	阳离子的
二亚油基-甲基-4-二甲基氨基丁酸酯	DLin-MC3-DMA	阳离子的

[0345] 用于基因转移的示例性聚合物示于下表5中。

[0346] 表5. 用于基因转移的聚合物

[0347]

聚合物	缩写
聚乙二醇	PEG
聚乙烯亚胺	PEI
二硫代双(琥珀酰亚胺基丙酸酯)	DSP
二甲基-3,3'-二硫代双丙亚氨酸酯	DTBP
聚(乙烯亚胺)双氨基甲酸酯	PEIC
聚(L-赖氨酸)	PLL
组氨酸修饰的 PLL	
聚( <i>N</i> -乙烯基吡咯烷酮)	PVP
聚(丙烯亚胺)	PPI
聚(酰胺胺)	PAMAM
聚(酰胺基乙烯亚胺)	SS-PAEI
三亚乙基四胺	TETA
聚( $\beta$ -氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚( $\alpha$ -[4-氨基丁基]-L-乙醇酸)	PAGA

[0348]

聚合物	缩写
聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(N-乙基-4-乙烯基溴化吡啶鎓)	
聚(磷腈)	PPZ
聚(磷酸酯)	PPE
聚(氨基磷酸酯)	PPA
聚(N-2-羟基丙基甲基丙烯酰胺)	pHPMA
聚(2-(二甲基氨基)甲基丙烯酸乙酯)	pDMAEMA
聚(2-氨基乙基丙烯磷酸酯)	PPE-EA
壳聚糖	
半乳糖基化壳聚糖	
N-十二烷基化的壳聚糖	
组蛋白	
胶原	
葡聚糖-精胺	D-SPM

[0349] 在某些实施例中,所述运载体具有靶向修饰以增加靶细胞摄入纳米粒子和脂质体(例如,细胞特异性抗原、单克隆抗体、单链抗体、适配体、聚合物、糖和细胞穿透肽)。在某些实施例中,所述运载体使用融合肽和内体去稳定肽/聚合物。在某些实施例中,所述运载体经历酸触发的构象变化(例如,以加速负荷物的内体逃逸)。在某些实施例中,使用刺激可切割的聚合物,例如用于在细胞区室中释放。例如,可以使用在还原性细胞环境中被切割的基于二硫化物的阳离子型聚合物。

[0350] 在某些实施例中,所述递送运载体是生物非病毒递送运载体。在某些实施例中,所述运载体是减毒细菌(例如,天然或人工工程化成侵入性的,但减毒以防止发病原和表达转基因(例如,单核细胞增生李斯特菌、某些沙门氏菌属菌株、长双歧杆菌和经修饰的大肠杆菌)、具有营养和组织特异性向性以靶向特定细胞的细菌、具有经修饰的表面蛋白以改变靶细胞特异性的细菌)。在某些实施例中,所述运载体是基因修饰噬菌体(例如,具有大包装能力、较少免疫原性、含有哺乳动物质粒维持序列并且具有结合的靶向配体的工程化噬菌体)。在某些实施例中,所述运载体是哺乳动物病毒样粒子。例如,可以产生经修饰的病毒粒子(例如,通过纯化“空心”粒子,随后用所需负荷物离体组装病毒)。所述运载体也可以被工程化为结合靶向配体从而改变靶组织特异性。在某些实施例中,所述运载体是生物脂质体。例如,所述生物脂质体是衍生自人细胞的基于磷脂的粒子(例如,红细胞血影,其是红细胞分解成衍生自受试者的球状结构(例如,可以通过附接不同组织或细胞特异性配体来实

现组织靶向)、或分泌外泌体-受试者衍生的内吞起源的膜结合纳米运载体(30-100nm)(例如,可以产生自不同细胞类型并且因此可以被细胞吸收,而不需要靶向配体)。

[0351] 在某些实施例中,递送除了CRISPR/Cpf1系统的组分(例如,本文所描述的Cpf1分子组分和/或gRNA分子组分)外的一种或多种核酸组合物(例如,DNA分子)。在某些实施例中,所述核酸组合物是在递送CRISPR/Cpf1系统的一种或多种组分的同时进行递送的。在某些实施例中,所述核酸组合物是在递送Cpf1系统的一种或多种组分之前或之后(例如,小于约30分钟、1小时、2小时、3小时、6小时、9小时、12小时、1天、2天、3天、1周、2周或4周)进行递送的。在某些实施例中,所述核酸组合物是通过不同于递送CRISPR/Cpf1系统的一种或多种组分(例如,Cpf1分子组分和/或gRNA分子组分)的方式进行递送的。所述核酸组合物可以通过本文所描述的任何递送方法进行递送。例如,所述核酸分子可以通过病毒载体(例如,逆转录病毒或慢病毒)进行递送,并且所述Cpf1分子组分和/或所述gRNA分子组分可以通过电穿孔进行递送。在某些实施例中,所述核酸组合物编码FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC或TRBC基因。

#### [0352] 编码Cpf1分子的RNA的递送

[0353] 可以通过本领域已知的方法或如本文所描述的将编码Cpf1分子和/或gRNA分子的RNA递送到细胞(例如,本文所描述的靶细胞)中。例如,编码Cpf1的和/或编码gRNA的RNA可以例如通过显微注射、电穿孔、瞬时细胞压缩或挤压(例如,如在Lee等人[2012]Nano Lett[纳米快报]12:6322-27中所描述的)、脂质介导的转染、肽介导的递送或其组合进行递送。编码Cpf1的和/或编码gRNA的RNA可以与分子缀合,从而促进被靶细胞(例如,本文所描述的靶细胞)摄取。

[0354] 在某些实施例中,经由电穿孔递送包括将细胞与编码Cpf1分子和/或gRNA分子的RNA在盒、室或比色皿中混合并且施加一个或多个限定持续时间和幅度的电脉冲。在某些实施例中,使用如下系统进行经由电穿孔的递送,在所述系统中将细胞与编码Cpf1分子和/或gRNA分子的RNA在连接至装置(例如,泵)的容器中混合,所述装置向盒、室或比色皿中给料混合物,在所述盒、室或比色皿中施加一个或多个限定持续时间和幅度的电脉冲,之后将细胞递送至第二容器。编码Cpf1的和/或编码gRNA的RNA可以与分子缀合,从而促进被靶细胞(例如,本文所描述的靶细胞)摄取。

#### [0355] Cpf1分子蛋白的递送

[0356] Cpf1分子可以通过本领域已知的方法或如本文所描述的递送到细胞中。例如,Cpf1蛋白分子可以例如通过显微注射、电穿孔、瞬时细胞压缩或挤压(例如,如在Lee等人[2012]Nano Lett[纳米快报]12:6322-27中所描述的)、脂质介导的转染、肽介导的递送或其组合进行递送。递送可以与编码gRNA的DNA或与gRNA相伴。

[0357] 在某些实施例中,经由电穿孔递送包括将细胞与Cpf1分子、与或不与gRNA分子在盒、室或比色皿中混合并且施加一个或多个限定持续时间和幅度的电脉冲。在某些实施例中,使用如下系统进行经由电穿孔的递送,在所述系统中将细胞与Cpf1分子、与或不与gRNA分子在连接至装置(例如,泵)的容器中混合,所述装置向盒、室或比色皿中给料混合物,在所述盒、室或比色皿中施加一个或多个限定持续时间和幅度的电脉冲,之后将细胞递送至第二容器。

#### [0358] Cpf1分子蛋白和gRNA的RNP递送

[0359] 在某些实施例中,所述Cpf1分子和gRNA经由核糖核蛋白(RNP)递送而递送至靶细胞。在某些实施例中,所述Cpf1分子被提供为蛋白质,并且所述gRNA分子被提供为转录或合成的RNA。所述gRNA分子可以通过化学合成产生。在某些实施例中,所述gRNA分子在递送至靶细胞之前在合适的条件下与所述Cpf1分子形成RNP复合物。在某些实施例中,合适的条件包括将所述Cpf1分子蛋白和所述gRNA在室温下孵育至少约10分钟。所述RNP复合物可以通过本领域已知的任何合适的方法递送至靶细胞,例如通过电穿孔、脂质介导的转染、基于蛋白质或DNA的穿梭、机械力或液压力。在某些实施例中,将所述RNP复合物通过电穿孔递送至靶细胞。

[0360] VIII. 经修饰的核苷、核苷酸和核酸

[0361] 经修饰的核苷和经修饰的核苷酸可以存在于核酸(例如特别是gRNA,但是还有编码Cpf1分子的核酸)中。除了或代替上述任何特定的gRNA分子修饰,可以进行本节中披露的修饰。如本文所描述的,“核苷”被定义为含有五碳糖分子(戊糖或核糖)或其衍生物以及有机碱(嘌呤或嘧啶)或其衍生物的化合物。如本文所描述的,“核苷酸”被定义为进一步包含磷酸酯基团的核苷。

[0362] 经修饰的核苷和核苷酸可以包括以下中的一项或多项:

[0363] (i) 磷酸二酯骨架键联中的一个或两个非连接磷酸氧和/或一个或多个连接磷酸氧的改变,例如置换;

[0364] (ii) 核糖的成分(例如,核糖上的2' 羟基)的改变,例如置换;

[0365] (iii) “脱磷酸”接头对磷酸部分的完全置换;

[0366] (iv) 天然存在的核碱基的修饰或置换;

[0367] (v) 核糖-磷酸骨架的置换或修饰;

[0368] (vi) 寡核苷酸的3' 端或5' 端的修饰,例如末端磷酸酯基团的去除、修饰或置换或部分的缀合;以及

[0369] (vii) 糖的修饰。

[0370] 以上列出的修饰可以组合,以提供可以具有两种、三种、四种或更多种修饰的经修饰的核苷和核苷酸。例如,经修饰的核苷或核苷酸可以具有经修饰的糖和经修饰的核碱基。在某些实施例中,修饰gRNA的每个碱基,例如所有碱基都具有经修饰的磷酸酯基团,例如所有经修饰的磷酸酯基团都是硫代磷酸酯基团。在某些实施例中,单分子的或模块化的gRNA分子的所有或基本上所有磷酸酯基团被硫代磷酸酯基团置换。

[0371] 在某些实施例中,可以将经修饰的核苷酸(例如,具有如本文所描述的修饰的核苷酸)掺入核酸(例如“经修饰的核酸”)中。在某些实施例中,经修饰的核酸包含一个、两个、三个或更多个经修饰的核苷酸。在某些实施例中,经修饰的核酸中的至少5% (例如,至少约5%、至少约10%、至少约15%、至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%或约100%) 的位置是经修饰的核苷酸。

[0372] 未经修饰的核酸可以易于被例如细胞核酸酶降解。例如,核酸酶可以水解核酸磷酸二酯键。因此,在一个方面中,本文所描述的经修饰的核酸可以含有一个或多个经修饰的核苷或核苷酸,例如以引入对核酸酶的稳定性。

[0373] 在某些实施例中,本文所描述的经修饰的核苷、经修饰的核苷酸和经修饰的核酸当被引入细胞群中时可以展现出降低的先天性免疫应答。术语“先天性免疫应答”包括对外源核酸的细胞应答,所述外源核酸包括通常是病毒或细菌来源的单链核酸,所述细胞应答涉及细胞因子(特别是干扰素)表达与释放以及细胞死亡的诱导。在某些实施例中,本文所描述的经修饰的核苷、经修饰的核苷酸和经修饰的核酸可以破坏大沟相互作用配偶体与核酸的结合。在某些实施例中,本文所描述的经修饰的核苷、经修饰的核苷酸和经修饰的核酸当被引入细胞群中时可以展现出降低的先天性免疫应答,并且还破坏大沟相互作用配偶体与核酸的结合。

#### [0374] 化学基团的定义

[0375] 如本文所使用的,“烷基”意在指直链的或支链的饱和烃基。示例性烷基基团包括甲基(Me)、乙基(Et)、丙基(例如,正丙基和异丙基)、丁基(例如,正丁基、异丁基、叔丁基)、戊基(例如,正戊基、异戊基、新戊基)等。烷基基团可以含有从1至约20、从2至约20、从1至约12、从1至约8、从1至约6、从1至约4或从1至约3个碳原子。

[0376] 如本文所使用的,“芳基”是指单环或多环(例如,具有2、3或4个稠环)的芳香烃,例如像苯基、萘基、蒽基、菲基、茛满基、茛基等。在某些实施例中,芳基基团具有从6至约20个碳原子。

[0377] 如本文所使用的,“烯基”是指含有至少一个双键的脂肪族基团。

[0378] 如本文所使用的,“炔基”是指含有2-12个碳原子并且特征在于具有一个或多个三键的直链的或支链的炔链。炔基基团的实例包括但不限于乙炔基、炔丙基和3-己炔基。

[0379] 如本文所使用的,“芳基烷基”或“芳烷基”是指烷基氢原子被芳基基团置换的烷基部分。芳烷基包括超过一个氢原子已经被芳基基团置换的基团。“芳基烷基”或“芳烷基”的实例包括苄基、2-苯基乙基、3-苯基丙基、9-苄基、二苯甲基以及三苯甲基基团。

[0380] 如本文所使用的,“环烷基”是指具有3至12个碳的环状的、二环的、三环的或多环的非芳香烃基团。环烷基部分的实例包括但不限于环丙基、环戊基和环己基。

[0381] 如本文所使用的,“杂环基”是指杂环系统的单价基。代表性杂环基包括但不限于四氢呋喃基、四氢噻吩基、吡咯烷基、吡咯烷酮基、哌啶基、吡咯啉基、哌嗪基、二噁烷基、二氧戊环基、二氮杂卓基、氧氮杂卓基、硫氮杂卓基以及吗啉基。

[0382] 如本文所使用的,“杂芳基”是指杂芳香环系统的单价基。杂芳基部分的实例包括但不限于咪唑基、噁唑基、噻唑基、三唑基、吡咯基、呋喃基、吡啶基、苯硫基、吡啶基、吡啶基、吡嗪基、哒嗪基、嘧啶基、吡嗪基、嘌呤基、嘧啶基、喹啉基以及蝶啶基。

#### [0383] 磷酸骨架修饰

##### [0384] 磷酸酯基团

[0385] 在某些实施例中,可以通过用不同取代基置换一个或多个氧来修饰经修饰的核苷酸的磷酸酯基团。此外,经修饰的核苷酸(例如,存在于经修饰的核酸中的经修饰的核苷酸)可以包括如本文所描述的经修饰的磷酸酯对未经修饰的磷酸酯部分的完全置换。在某些实施例中,磷酸骨架的修饰可以包括产生不带电接头或具有不对称电荷分布的带电接头的改变。

[0386] 经修饰的磷酸酯基团的实例包括硫代磷酸酯、硒代磷酸酯(phosphoroselenate)、硼磷酸酯(borano phosphate)、硼磷酸酯(borano phosphate ester)、氢磷酯、磷酰胺酯

(phosphoroamidate)、烷基或芳基磷酸酯和磷酸三酯。在某些实施例中,磷酸骨架部分中的非桥连磷酸氧原子之一可以被以下基团中的任一项置换:硫(S)、硒(Se)、BR<sub>3</sub>(其中R可以是例如氢、烷基或芳基)、C(例如,烷基基团、芳基基团等)、H、NR<sub>2</sub>(其中R可以是例如氢、烷基或芳基)或OR(其中R可以是例如烷基或芳基)。未经修饰的磷酸酯基团中的磷原子是非手性的。然而,以上原子或原子的基团之一对非桥连磷酸氧之一的置换可以使得磷原子是手性的;也就是说以这种方式修饰的磷酸酯基团中的磷原子是立构中心。立构磷原子可以具有“R”构型(本文是R<sub>p</sub>)或“S”构型(本文是S<sub>p</sub>)。

[0387] 二硫代磷酸酯具有两个被硫置换的非桥连氧。二硫代磷酸酯中的磷中心是非手性的,这阻止寡核糖核苷酸非对映异构体的形成。在某些实施例中,对一个或两个非桥连氧的修饰还可以包括用以下基团置换非桥连氧,所述基团独立地选自S、Se、B、C、H、N以及OR(R可以是例如烷基或芳基)。

[0388] 还可以通过用氮(桥连的磷酰胺酯)、硫(桥连的硫代磷酸酯)和碳(桥连的亚甲基磷酸酯)置换桥连氧(即,将磷酸连接至核苷的氧)来修饰磷酸酯接头。置换可以发生在连接氧或发生在两个连接氧处。

[0389] 磷酸酯基团的置换

[0390] 磷酸酯基团可以被不含磷连接物置换。在某些实施例中,带电磷酸酯基团可以被中性部分置换。

[0391] 可以置换磷酸酯基团的部分的实例可以包括但不限于例如甲基磷酸酯、羟氨基、硅氧烷、碳酸酯、羧甲基、氨基甲酸酯、酰胺、硫醚、环氧乙烷接头、磺酸酯、磺酰胺、硫代甲缩醛(thioformacetal)、甲缩醛(formacetal)、胍、亚甲亚氨基、亚甲甲基亚氨基、亚甲胍基、亚甲二甲基胍基以及亚甲氧基甲基亚氨基。

[0392] 核糖磷酸骨架的置换

[0393] 还可以构建可以模拟核酸的支架,其中磷酸酯接头和核糖被核酸酶抗性核苷或核苷酸替代物置换。在某些实施例中,可以通过替代骨架拴住核碱基。实例可以包括但不限于吗啉代、环丁基、吡咯烷和肽核酸(PNA)核苷替代物。

[0394] 糖修饰

[0395] 经修饰的核苷和经修饰的核苷酸可以包括对糖基的一种或多种修饰。例如,2' 羟基基团(OH)可以被多种不同的“氧基”或“脱氧”取代基修饰或置换。在某些实施例中,对2' 羟基基团的修饰可以增强核酸的稳定性,因为羟基不再可以被去质子化以形成2' -醇盐离子。2' -醇盐可以通过接头磷原子上的分子内亲核攻击而催化降解。

[0396] “氧基”-2' 羟基基团修饰的实例可以包括烷氧基或芳氧基(OR,其中“R”可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖);聚乙二醇(PEG),O(CH<sub>2</sub>CH<sub>2</sub>O)<sub>n</sub>CH<sub>2</sub>CH<sub>2</sub>OR,其中R可以是例如H或任选取代的烷基,并且n可以是0至20(例如,从0至4、从0至8、从0至10、从0至16、从1至4、从1至8、从1至10、从1至16、从1至20、从2至4、从2至8、从2至10、从2至16、从2至20、从4至8、从4至10、从4至16以及从4至20)的整数。在某些实施例中,“氧基”-2' 羟基基团修饰可以包括“锁”核酸(LNA),其中2' 羟基可以例如通过C<sub>1-6</sub>亚烷基或C<sub>1-6</sub>杂亚烷基桥连接至同一核糖的4' 碳,其中示例性桥可以包括亚甲基、亚丙基、醚或氨基桥;O-氨基(其中氨基可以是例如NH<sub>2</sub>;烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基)和氨基烷氧基O(CH<sub>2</sub>)<sub>n</sub>-氨基(其中氨基可以是例如NH<sub>2</sub>;烷氨基、二烷氨基、杂

环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基)。在某些实施例中，“氧基”-2' 羟基基团修饰可以包括甲氧基乙基基团(MOE) ( $\text{OCH}_2\text{CH}_2\text{OCH}_3$ , 例如PEG衍生物)。

[0397] “脱氧”修饰可以包括氢(即脱氧核糖, 例如在部分ds RNA的突出端部分); 卤素(例如, 溴、氯、氟或碘); 氨基(其中氨基可以是例如 $\text{NH}_2$ ; 烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基、二杂芳氨基或氨基酸);  $\text{NH}(\text{CH}_2\text{CH}_2\text{NH})_n\text{CH}_2\text{CH}_2$ -氨基(其中氨基可以是例如如本文所描述的),  $-\text{NHC}(\text{O})\text{R}$ (其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖), 氰基; 巯基; 烷基-硫代-烷基; 硫代烷氧基; 以及烷基、环烷基、芳基、烯基和炔基, 其可以任选地被例如如本文所描述的氨基取代。

[0398] 糖基还可以含有一个或多个具有与核糖中的对应碳相反的立体化学构型的碳。因此, 经修饰的核酸可以包括含有例如阿拉伯糖作为糖的核苷酸。核苷酸“单体”可以在糖的1' 位处具有 $\alpha$ 键联, 例如 $\alpha$ -核苷。经修饰的核酸还可以包括“无碱基”糖, 其在C-1' 处缺乏核碱基。这些无碱基糖还可以在一个或多个构成性糖原子处被进一步修饰。经修饰的核酸还可以包括一种或多种处于L型的糖, 例如L-核苷。

[0399] 通常, RNA包括糖基核糖, 它是具有氧的5元环。示例性的经修饰的核苷和经修饰的核苷酸可以包括但不限于核糖中氧的置换(例如, 用硫(S)、硒(Se)或亚烷基, 例如像亚甲基或亚乙基); 双键的添加(例如, 以用环戊烯基或环己烯基置换核糖); 核糖的缩环(例如, 以形成环丁烷或氧杂环丁烷的4元环); 核糖的扩环(例如, 以形成具有另外的碳或杂原子的6元或7元环, 例如像脱水己糖醇、阿卓糖醇、甘露醇、环己烷基、环己烯基以及吗啉代, 其也具有氨基磷酸酯骨架)。在某些实施例中, 经修饰的核苷酸可以包括多环形式(例如, 三环; 和“解锁”形式, 如二醇核酸(GNA)(例如, R-GNA或S-GNA, 其中核糖被附接至磷酸二酯键的二醇单元置换), 苏糖核酸(TNA, 其中核糖被 $\alpha$ -L-苏呋喃糖基(threofuranosyl)-(3'  $\rightarrow$  2') 置换)。

#### [0400] 核碱基上的修饰

[0401] 能够被掺入经修饰的核酸中的本文所描述的经修饰的核苷和经修饰的核苷酸可以包括经修饰的核碱基。核碱基的实例包括但不限于腺嘌呤(A)、鸟嘌呤(G)、胞嘧啶(C)以及尿嘧啶(U)。这些核碱基可以被修饰或全部置换, 以提供可以被掺入经修饰的核酸中的经修饰的核苷和经修饰的核苷酸。核苷酸的核碱基可以独立地选自嘌呤、嘧啶、嘌呤或嘧啶类似物。在某些实施例中, 核碱基可以包括例如天然存在的碱基及其合成衍生物。

#### [0402] 尿嘧啶

[0403] 在某些实施例中, 经修饰的核碱基是经修饰的尿嘧啶。具有经修饰的尿嘧啶的示例性核碱基和核苷包括但不限于假尿苷( $\psi$ )、吡啶-4-酮核糖核苷、5-氮杂-尿苷、6-氮杂-尿苷、2-硫代-5-氮杂-尿苷、2-硫代-尿苷( $s2U$ )、4-硫代-尿苷( $s4U$ )、4-硫代-假尿苷、2-硫代-假尿苷、5-羟基-尿苷( $ho^5U$ )、5-氨基烯丙基-尿苷、5-卤代-尿苷(例如, 5-碘代-尿苷或5-溴代-尿苷)、3-甲基-尿苷( $m^3U$ )、5-甲氧基-尿苷( $mo^5U$ )、尿苷5-氧基乙酸( $cmo^5U$ )、尿苷5-氧基乙酸甲酯( $mcmo^5U$ )、5-羧甲基-尿苷( $cm^5U$ )、1-羧甲基-假尿苷、5-羧基羟甲基-尿苷( $chm^5U$ )、5-羧基羟甲基-尿苷甲酯( $mchm^5U$ )、5-甲氧羰基甲基-尿苷( $mcm^5U$ )、5-甲氧羰基甲基-2-硫代-尿苷( $mcm^5s2U$ )、5-氨基甲基-2-硫代-尿苷( $nm^5s2U$ )、5-甲基氨基-尿苷( $mnm^5U$ )、5-甲基氨基-2-硫代-尿苷( $mnm^5s2U$ )、5-甲基氨基-2-硒代-尿苷( $mnm^5se^2U$ )、5-氨基酰基甲基-尿苷( $ncm^5U$ )、5-羧甲基氨基甲基-尿苷( $cmnm^5U$ )、5-羧甲基氨基甲基-2-硫代-尿苷( $cmnm$

<sup>5</sup>s2U)、5-丙炔基-尿苷、1-丙炔基-假尿苷、5-牛磺酸甲基-尿苷( $\tau\text{cm}^5\text{U}$ )、1-牛磺酸甲基-假尿苷、5-牛磺酸甲基-2-硫代-尿苷( $\tau\text{m}^5\text{s}2\text{U}$ )、1-牛磺酸甲基-4-硫代-假尿苷、5-甲基-尿苷( $\text{m}^5\text{U}$ ,即具有核碱基脱氧胸腺嘧啶)、1-甲基-假尿苷( $\text{m}^1\psi$ )、5-甲基-2-硫代-尿苷( $\text{m}^5\text{s}2\text{U}$ )、1-甲基-4-硫代-假尿苷( $\text{m}^1\text{s}^4\psi$ )、4-硫代-1-甲基-假尿苷、3-甲基-假尿苷( $\text{m}^3\psi$ )、2-硫代-1-甲基-假尿苷、1-甲基-1-去氮杂-假尿苷、2-硫代-1-甲基-1-去氮杂-假尿苷、二氢尿苷(D)、二氢假尿苷、5,6-二氢尿苷、5-甲基-二氢尿苷( $\text{m}^5\text{D}$ )、2-硫代-二氢尿苷、2-硫代-二氢假尿苷、2-甲氧基-尿苷、2-甲氧基-4-硫代-尿苷、4-甲氧基-假尿苷、4-甲氧基-2-硫代-假尿苷、N1-甲基-假尿苷、3-(3-氨基-3-羧丙基)尿苷( $\text{acp}^3\text{U}$ )、1-甲基-3-(3-氨基-3-羧丙基)假尿苷( $\text{acp}^3\psi$ )、5-(异戊烯基氨基)尿苷( $\text{inm}^5\text{U}$ )、5-(异戊烯基氨基)-2-硫代-尿苷( $\text{inm}^5\text{s}2\text{U}$ )、 $\alpha$ -硫代-尿苷、2'-O-甲基-尿苷(Um)、5,2'-O-二甲基-尿苷( $\text{m}^5\text{Um}$ )、2'-O-甲基-假尿苷( $\psi\text{m}$ )、2-硫代-2'-O-甲基-尿苷( $\text{s}2\text{Um}$ )、5-甲氧基羧甲基-2'-O-甲基-尿苷( $\text{mcm}^5\text{Um}$ )、5-氨基甲酰基甲基-2'-O-甲基-尿苷( $\text{ncm}^5\text{Um}$ )、5-羧甲基氨基甲基-2'-O-甲基-尿苷( $\text{cmm}^5\text{Um}$ )、3,2'-O-二甲基-尿苷( $\text{m}^3\text{Um}$ )、5-(异戊烯基氨基)-2'-O-甲基-尿苷( $\text{inm}^5\text{Um}$ )、1-硫代-尿苷、脱氧胸苷、2'-F-阿糖(ara)-尿苷、2'-F-尿苷、2'-OH-阿糖-尿苷、5-(2-甲氧甲酰基乙烯基)尿苷、5-[3-(1-E-丙烯基氨基)尿苷、吡啶并[3,4-d]嘧啶、黄嘌呤以及次黄嘌呤。

#### [0404] 胞嘧啶

[0405] 在某些实施例中,经修饰的核碱基是经修饰的胞嘧啶。具有经修饰的胞嘧啶的示例性核碱基和核苷包括但不限于5-氮杂-胞苷、6-氮杂-胞苷、假异胞苷、3-甲基-胞苷( $\text{m}^3\text{C}$ )、N4-乙酰基-胞苷(act)、5-甲酰基-胞苷( $\text{f}^5\text{C}$ )、N4-甲基-胞苷( $\text{m}^4\text{C}$ )、5-甲基-胞苷( $\text{m}^5\text{C}$ )、5-卤代-胞苷(例如,5-碘代-胞苷)、5-羟甲基-胞苷( $\text{hm}^5\text{C}$ )、1-甲基-假异胞苷、吡咯并-胞苷、吡咯并-假异胞苷、2-硫代-胞苷( $\text{s}2\text{C}$ )、2-硫代-5-甲基-胞苷、4-硫代-假异胞苷、4-硫代-1-甲基-假异胞苷、4-硫代-1-甲基-1-去氮杂-假异胞苷、1-甲基-1-去氮杂-假异胞苷、泽布拉林(zebularine)、5-氮杂-泽布拉林、5-甲基-泽布拉林、5-氮杂-2-硫代-泽布拉林、2-硫代-泽布拉林、2-甲氧基-胞苷、2-甲氧基-5-甲基-胞苷、4-甲氧基-假异胞苷、4-甲氧基-1-甲基-假异胞苷、赖西丁( $\text{k}^2\text{C}$ )、 $\alpha$ -硫代-胞苷、2'-O-甲基-胞苷(Cm)、5,2'-O-二甲基-胞苷( $\text{m}^5\text{Cm}$ )、N4-乙酰基-2'-O-甲基-胞苷( $\text{ac}^4\text{Cm}$ )、N4,2'-O-二甲基-胞苷( $\text{m}^4\text{Cm}$ )、5-甲酰基-2'-O-甲基-胞苷( $\text{f}^5\text{Cm}$ )、N4,N4,2'-O-三甲基-胞苷( $\text{m}^4_2\text{Cm}$ )、1-硫代-胞苷、2'-F-阿糖-胞苷、2'-F-胞苷以及2'-OH-阿糖-胞苷。

#### [0406] 腺嘌呤

[0407] 在某些实施例中,经修饰的核碱基是经修饰的腺嘌呤。具有经修饰的腺嘌呤的示例性核碱基和核苷包括但不限于2-氨基-嘌呤、2,6-二氨基嘌呤、2-氨基-6-卤代-嘌呤(例如,2-氨基-6-氯代-嘌呤)、6-卤代-嘌呤(例如,6-氯代-嘌呤)、2-氨基-6-甲基-嘌呤、8-叠氮基-腺苷、7-去氮杂-腺苷、7-去氮杂-8-氮杂-腺苷、7-去氮杂-2-氨基-嘌呤、7-去氮杂-8-氮杂-2-氨基-嘌呤、7-去氮杂-2,6-二氨基嘌呤、7-去氮杂-8-氮杂-2,6-二氨基嘌呤、1-甲基-腺苷( $\text{m}^1\text{A}$ )、2-甲基-腺苷( $\text{m}^2\text{A}$ )、N6-甲基-腺苷( $\text{m}^6\text{A}$ )、2-甲硫基-N6-甲基-腺苷( $\text{ms}2\text{m}^6\text{A}$ )、N6-异戊烯基-腺苷( $\text{i}^6\text{A}$ )、2-甲硫基-N6-异戊烯基-腺苷( $\text{ms}^2\text{i}^6\text{A}$ )、N6-(顺羟基异戊烯基)腺苷( $\text{io}^6\text{A}$ )、2-甲硫基-N6-(顺羟基异戊烯基)腺苷( $\text{ms}2\text{io}^6\text{A}$ )、N6-缩水甘油基氨基甲酰基-腺苷( $\text{g}^6\text{A}$ )、N6-苏氨酰基氨基甲酰基-腺苷( $\text{t}^6\text{A}$ )、N6-甲基-N6-苏氨酰基氨基甲酰基-腺苷( $\text{m}^6\text{t}^6\text{A}$ )、2-甲硫基-N6-苏氨酰基氨基甲酰基-腺苷( $\text{ms}^2\text{g}^6\text{A}$ )、N6,N6-二甲基-腺苷( $\text{m}^6_2\text{A}$ )、N6-羟基正缬氨

酰基氨甲酰基-腺苷(hn<sup>6</sup>A)、2-甲硫基-N6-羟基正缬氨酰基氨甲酰基-腺苷(ms2hn<sup>6</sup>A)、N6-乙酰基-腺苷(ac<sup>6</sup>A)、7-甲基-腺苷、2-甲硫基-腺苷、2-甲氧基-腺苷、 $\alpha$ -硫代-腺苷、2'-O-甲基-腺苷(Am)、N<sup>6</sup>,2'-O-二甲基-腺苷(m<sup>6</sup>Am)、N<sup>6</sup>-甲基-2'-脱氧腺苷、N6,N6,2'-O-三甲基-腺苷(m<sup>6</sup><sub>2</sub>Am)、1,2'-O-二甲基-腺苷(m<sup>1</sup>Am)、2'-O-核糖基腺苷(磷酸盐)(Ar(p))、2-氨基-N6-甲基-嘌呤、1-硫代-腺苷、8-叠氮基-腺苷、2'-F-阿糖-腺苷、2'-F-腺苷、2'-Oh-阿糖-腺苷以及N6-(19-氨基-五氧杂十九烷基)-腺苷。

#### [0408] 鸟嘌呤

[0409] 在某些实施例中,经修饰的核碱基是经修饰的鸟嘌呤。具有经修饰的鸟嘌呤的示例性核碱基和核苷包括但不限于肌苷(I)、1-甲基-肌苷(m<sup>1</sup>I)、怀俄苷(imG)、甲基怀俄苷(mimG)、4-去甲基-怀俄苷(imG-14)、异怀俄苷(imG2)、怀丁苷(yW)、过氧怀丁苷(o<sub>2</sub>yW)、羟基怀丁苷(OHyW)、修饰不足的羟基怀丁苷(OHyW\*)、7-去氮杂-鸟苷、癸苷(Q)、环氧癸苷(oQ)、半乳糖基-癸苷(galQ)、甘露糖基-癸苷(manQ)、7-氰基-7-去氮杂-鸟苷(preQ<sub>0</sub>)、7-氨基甲基-7-去氮杂-鸟苷(preQ<sub>1</sub>)、古嘌呤(G<sup>+</sup>)、7-去氮杂-8-氮杂-鸟苷、6-硫代-鸟苷、6-硫代-7-去氮杂-鸟苷、6-硫代-7-去氮杂-8-氮杂-鸟苷、7-甲基-鸟苷(m<sup>7</sup>G)、6-硫代-7-甲基-鸟苷、7-甲基-肌苷、6-甲氧基-鸟苷、1-甲基-鸟苷(m<sup>1</sup>G)、N2-甲基-鸟苷(m<sup>2</sup>G)、N2,N2-二甲基-鸟苷(m<sup>2</sup><sub>2</sub>G)、N2,7-二甲基-鸟苷(m<sup>2</sup>,7G)、N2,N2,7-二甲基-鸟苷(m<sup>2</sup>,2,7G)、8-氧代-鸟苷、7-甲基-8-氧代-鸟苷、1-甲基-6-硫代-鸟苷、N2-甲基-6-硫代-鸟苷、N2,N2-二甲基-6-硫代-鸟苷、 $\alpha$ -硫代-鸟苷、2'-O-甲基-鸟苷(Gm)、N2-甲基-2'-O-甲基-鸟苷(m<sup>2</sup>Gm)、N2,N2-二甲基-2'-O-甲基-鸟苷(m<sup>2</sup><sub>2</sub>Gm)、1-甲基-2'-O-甲基-鸟苷(m<sup>1</sup>Gm)、N2,7-二甲基-2'-O-甲基-鸟苷(m<sup>2</sup>,7Gm)、2'-O-甲基-肌苷(Im)、1,2'-O-二甲基-肌苷(m<sup>1</sup>Im)、0<sup>6</sup>-苯基-2'-脱氧肌苷、2'-O-核糖基鸟苷(磷酸盐)(Gr(p))、1-硫代-鸟苷、0<sup>6</sup>-甲基-鸟苷、0<sup>6</sup>-甲基-2'-脱氧鸟苷、2'-F-阿糖-鸟苷以及2'-F-鸟苷。

#### [0410] 示例性的经修饰的gRNA

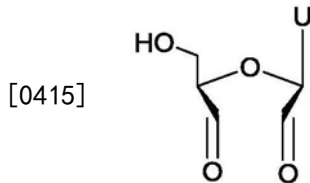
[0411] 在某些实施例中,经修饰的核酸可以是经修饰的gRNA。应当理解的是本文所描述的任何gRNA均可以根据本部分进行修饰,包括包含含有选自下组的核苷酸序列的靶向结构域的任何gRNA,该组由以下组成:SEQ ID NO:1-3707。如本文所讨论的,瞬时表达的或递送的核酸可以易于被例如细胞核酸酶降解。因此,在一个方面中,本文所描述的经修饰的gRNA可以含有一个或多个经修饰的核苷或核苷酸,其引入对核酸酶的稳定性。在某些实施例中,本文所描述的某些经修饰的gRNA可以引起来自某些细胞(特别是本发明的细胞(例如,T细胞))的减少的先天性免疫应答。如上所述的,术语“先天性免疫应答”包括对外源核酸的细胞应答,所述外源核酸包括通常是病毒或细菌来源的单链核酸,所述细胞应答涉及细胞因子(特别是干扰素)表达与释放以及细胞死亡的诱导。

[0412] 例如,如本文所讨论的,当gRNA的5'端通过包含真核mRNA帽结构或帽类似物进行修饰时,我们已经在某些细胞类型(例如,T细胞)中的基因离体编辑方面看到改进。本发明涵盖这样的认识,即通过5'加帽的gRNA观察到的改进可以扩展到已经以其他方式(例如,通过包含经修饰的核苷或核苷酸、通过包含3'聚A束和/或当修饰体外转录的gRNA时通过用磷酸酶如牛小肠碱性磷酸酶处理以去除5'三磷酸酯基团)修饰以实现相同类型的结构或功能结果的gRNA上。

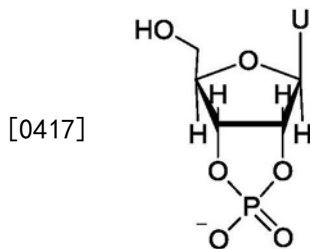
[0413] 因此,在某些实施例中,本文所讨论的方法和组合物提供了其中的gRNA已经在其

5' 端处或附近(例如,在其5' 端的1-10、1-9、1-8、1-7、1-6、1-5、1-4、1-3或1-2个核苷酸内)被修饰的方法和组合物。在某些实施例中,gRNA的5' 端通过包含真核mRNA帽结构或帽类似物(例如,G(5') ppp(5') G帽类似物、m7G(5') ppp(5') G帽类似物或3' -O-Me-m7G(5') ppp(5') G抗反向帽类似物(ARCA))进行修饰,如在图7中所描绘的。所述帽或帽类似物可以在gRNA的化学合成或体外转录期间被包括。在某些实施例中,体外转录的gRNA通过用磷酸酶(例如,牛小肠碱性磷酸酶)处理进行修饰,以去除5' 三磷酸酯基团。

[0414] 在某些实施例中,gRNA在其3' 端处或附近(例如,在其3' 端的1-10、1-9、1-8、1-7、1-6、1-5、1-4、1-3或1-2个核苷酸内)包含修饰。例如,在某些实施例中,gRNA的3' 端通过添加一个或多个(例如,25-200个)腺嘌呤(A)残基进行修饰。聚A束可以被包含在编码gRNA的核酸(例如,质粒、PCR产物、病毒基因组)中,或者可以在化学合成期间,或者在使用聚腺苷聚合酶(例如,大肠杆菌聚(A)聚合酶)体外转录后被添加到gRNA上。在某些实施例中,gRNA可以在3' 末端U核糖处被修饰。例如,U核糖的两个末端羟基基团可以被氧化为醛基基团和核糖环的伴随开口,以提供如下所示的经修饰的核苷:



[0416] 其中“U”可以是未经修饰的或经修饰的尿苷。在某些实施例中,可以用如下所示的2' 3' 环状磷酸酯修饰3' 末端U:



[0418] 其中“U”可以是未经修饰的或经修饰的尿苷。在某些实施例中,所述gRNA分子可以含有3' 核苷酸,其可以例如通过掺入本文所描述的一个或多个经修饰的核苷酸而相对于降解进行稳定化。在某些实施例中,例如,尿苷可以被经修饰的尿苷(例如,5-(2-氨基)丙基尿苷和5-溴代尿苷)或被本文所描述的任何经修饰的尿苷置换;腺苷和鸟苷可以被经修饰的腺苷和鸟苷(例如,在8位处具有修饰,例如8-溴代鸟苷)或被本文所描述的任何经修饰的腺苷和鸟苷置换。

[0419] 在某些实施例中,gRNA既在其5' 端处或附近包含修饰又在其3' 端处或附近包含修饰。在某些实施例中,体外转录的gRNA既含有5' 帽结构或帽类似物又含有3' 聚A束。在某些实施例中,体外转录的gRNA通过用磷酸酶(例如,牛小肠碱性磷酸酶)处理进行修饰以去除5' 三磷酸酯基团,并且包含3' 聚A束。

[0420] 虽然前述内容聚焦于末端修饰,应该理解的是本文所讨论的方法和组合物可以使用在gRNA序列内的一个或多个非末端位置和/或一个或多个末端位置处包括一个或多个经修饰的核苷或核苷酸的gRNA。

[0421] 在某些实施例中,可以向gRNA中掺入糖-修饰的核糖核苷酸,例如其中2' OH-基团

被选自以下的基团置换：H、-OR、-R（其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖）、卤素、-SH、-SR（其中R可以是例如烷基、环烷基、芳基、芳烷基、杂芳基或糖）、氨基（其中氨基可以是例如NH<sub>2</sub>；烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基、二杂芳氨基或氨基酸）；或氰基（-CN）。在某些实施例中，可以例如用硫代磷酸酯基团如本文所描述的修饰磷酸骨架。在某些实施例中，所述gRNA的一个或多个核苷酸可以各自独立地是经修饰的或未经修饰的核苷酸，包括但不限于2'-糖修饰的如2'-O-甲基、2'-O-甲氧基乙基，或2'-氟修饰的，包括例如，2'-F或2'-O-甲基腺苷（A）、2'-F或2'-O-甲基胞苷（C）、2'-F或2'-O-甲基尿苷（U）、2'-F或2'-O-甲基胸苷（T）、2'-F或2'-O-甲基鸟苷（G）、2'-O-甲氧基乙基-5-甲基尿苷（Teo）、2'-O-甲氧基乙基腺苷（Aeo）、2'-O-甲氧基乙基-5-甲基胞苷（m5Ceo）及其任何组合。

[0422] 在某些实施例中，gRNA可以包括“锁”核酸（LNA），其中2' OH-基团可以例如通过C1-6亚烷基或C1-6杂亚烷基桥连接至同一核糖的4' 碳，其中示例性桥可以包括亚甲基、亚丙基、醚或氨基桥；O-氨基（其中氨基可以是例如NH<sub>2</sub>；烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基）和氨基烷氧基或O（CH<sub>2</sub>）<sub>n</sub>-氨基（其中氨基可以是例如NH<sub>2</sub>；烷氨基、二烷氨基、杂环基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基）。

[0423] 在某些实施例中，gRNA可以包括经修饰的核苷酸，其是多环的（例如，三环；和“解锁”形式，如二醇核酸（GNA）（例如，R-GNA或S-GNA，其中核糖被附接至磷酸二酯键的二醇单元置换），或苏糖核酸（TNA，其中核糖被 $\alpha$ -L-苏呋喃糖基-（3' → 2'）置换）。

[0424] 通常，gRNA分子包括糖基核糖，它是具有氧的5元环。示例性的经修饰的gRNA可以包括但不限于核糖中氧的置换（例如，用硫（S）、硒（Se）或亚烷基，例如像亚甲基或亚乙基）；双键的添加（例如，以用环戊烯基或环己烯基置换核糖）；核糖的缩环（例如，以形成环丁烷或氧杂环丁烷的4元环）；核糖的扩环（例如，以形成具有另外的碳或杂原子的6元或7元环，例如像脱水己糖醇、阿卓糖醇、甘露醇、环己烷基、环己烯基以及吗啉代，其也具有氨基磷酸酯骨架）。尽管大多数的糖类类似物改变被定位至2' 位，其他位点也适于修饰，包括4' 位。在某些实施例中，gRNA包含4'-S、4'-Se或4'-C-氨基甲基-2'-O-Me修饰。

[0425] 在某些实施例中，可以将去氮杂核苷酸（例如，7-去氮杂-腺苷）掺入gRNA中。在某些实施例中，可以将O-烷基化的和N-烷基化的核苷酸（例如，N6-甲基腺苷）掺入gRNA中。在某些实施例中，gRNA分子中的一个或多个或所有核苷酸是脱氧核苷酸。

[0426] XI. 抑制性Cpf1gRNA分子及其用于限制Cpf1系统活性的用途

[0427] 使用或包括编码Cpf1分子或gRNA分子的核酸（例如，DNA）的方法和组合物可以另外使用或包括“抑制性Cpf1gRNA分子”。所述抑制性Cpf1gRNA可以限制被引入细胞或受试者中的其他CRISPR/Cfp1组分的活性。在某些实施例中，gRNA分子包含与核酸上的靶结构域互补的靶向结构域，所述核酸包含编码被引入细胞或受试者中的CRISPR/Cpf1系统的组分的序列。在某些实施例中，抑制性Cpf1gRNA分子包含与以下上的靶序列互补的靶向结构域：(a) 编码Cpf1分子的核酸；(b) 编码gRNA分子的核酸，所述gRNA分子包含靶向FAS、BID、CTLA4、PDCD1、CBLB、PTPN6、B2M、TRAC和/或TRBC基因的靶向结构域（靶基因gRNA）；或编码CRISPR/Cpf1组分的超过一种核酸上，例如（a）和（b）两者。所述抑制性Cpf1gRNA分子可以与所述Cpf1分子复合，以使所述系统组分失活。在某些实施例中，Cpf1分子/抑制性Cpf1gRNA

分子复合物使包含编码Cpf1分子的序列的核酸失活。在某些实施例中，Cpf1分子/抑制性Cpf1gRNA分子复合物使包含编码靶基因gRNA分子的序列的核酸失活。在某些实施例中，Cpf1分子/抑制性Cpf1gRNA分子复合物对所述Cpf1分子/靶基因gRNA分子复合物的活性施加时间限制、表达水平限制或其他限制。在某些实施例中，Cpf1分子/抑制性Cpf1gRNA分子复合物减少脱靶或其他不想要的活性。在某些实施例中，抑制性Cpf1gRNA分子靶向有待负向调节的CRISPR/Cpf1系统组分的编码序列或控制区（例如，启动子）。例如，抑制性Cpf1gRNA可以靶向Cpf1分子的编码序列、或调节Cpf1分子分子编码序列的表达的控制区（例如，启动子）、或布置在两者之间的序列。在某些实施例中，抑制性Cpf1gRNA分子靶向靶基因gRNA的编码序列或控制区（例如，启动子）。在某些实施例中，抑制性Cpf1gRNA（例如，靶向Cpf1或靶向靶基因gRNA的抑制性Cpf1gRNA分子）或编码它的核酸被单独地引入，例如晚于Cpf1分子或编码它的核酸。例如，第一载体（例如病毒载体，例如AAV载体）可以引入编码Cpf1分子和一种或多种靶基因gRNA分子的核酸，并且第二载体（例如病毒载体，例如AAV载体）可以引入编码抑制性Cpf1gRNA分子（例如，靶向Cpf1或靶向靶基因gRNA的gRNA分子）的核酸。在某些实施例中，第二载体可以在第一载体之后引入。在某些实施例中，抑制性Cpf1gRNA分子（例如，靶向Cpf1或靶向靶基因gRNA的抑制性Cpf1gRNA分子）或编码它的核酸可以与Cpf1分子或编码它的核酸一起引入，例如同时或在同一载体中，但是例如在转录控制元件（例如，启动子或增强子）之下，所述转录控制元件稍后被激活，例如这样使得在一段时间之后，Cpf1的转录被减少。在某些实施例中，转录控制元件被内在地激活。在某些实施例中，转录元件经由引入外触发被激活。

[0428] 典型地，编码抑制性Cpf1gRNA分子（例如，靶向Cpf1的gRNA分子）的核酸序列处于与它负向调节的组分（例如，编码Cpf1分子的核酸）不同的控制区（例如，启动子）的控制之下。在某些实施例中，“不同的控制区”仅仅是指未处于在功能上与两个受控序列偶联的一个控制区（例如，启动子）的控制之下。在某些实施例中，不同是指在控制区的种类或类型上“不同的控制区”。例如，编码抑制性Cpf1gRNA分子（例如，靶向Cpf1的gRNA分子）的序列处于具有较低水平的表达的控制区（例如，启动子）的控制之下，或者其表达晚于编码它负向调节的组分的序列（例如，编码Cpf1分子的核酸）。

[0429] 通过举例，编码抑制性Cpf1gRNA分子（例如，靶向Cpf1的抑制性Cpf1gRNA分子）的序列可以处于本文所描述的控制区（例如，人U6小核启动子或人H1启动子）的控制之下。在某些实施例中，编码它负向调节的组分的序列（例如，编码Cpf1分子的核酸）可以处于本文所描述的的控制区（例如，启动子）（例如，CMV、EF-1a、MSCV、PGK、CAG控制启动子）的控制之下。

[0430] 实例

[0431] 以下实例仅仅是说明性的，并不旨在以任何方式限制本发明的范围或内容。

[0432] 实例1-将Cpf1/crRNA RNP递送至T细胞

[0433] 为了证明在原代CD4<sup>+</sup>T细胞中Cpf1介导的切割，将经纯化的氨基酸球菌属BV3L6Cpf1（“AsCpf1”）与针对TCR $\alpha$ 链设计的11种不同的gRNA（也称为“crRNA”）（表6）复合。GWED539包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3433中的核苷酸序列的靶向结构域。GWED540包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3587中的核苷酸序列的靶向结构域。

GWED541包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3538中的核苷酸序列的靶向结构域。GWED542包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3461中的核苷酸序列的靶向结构域。GWED543包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3475中的核苷酸序列的靶向结构域。GWED544包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3524中的核苷酸序列的靶向结构域。GWED545包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3566中的核苷酸序列的靶向结构域。GWED546包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3517中的核苷酸序列的靶向结构域。GWED547包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3573中的核苷酸序列的靶向结构域。GWED548包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3580中的核苷酸序列的靶向结构域。GWED549包含具有示于SEQ ID NO:3708中的核苷酸序列的直接重复结构域和具有示于SEQ ID NO:3454中的核苷酸序列的靶向结构域。

[0434] 表6

<u>crRNA ID</u>	<u>crRNA序列</u>	<u>SEQ ID NO</u>
GWED539	UAAUUUCUACUCUUGUAGAUAGAAUCAAAAUCGGUGAAUAGGC	3711
GWED540	UAAUUUCUACUCUUGUAGAUUUUGAGAAUCAAAAUCGGUGAAU	3712
GWED541	UAAUUUCUACUCUUGUAGAUGUCUGUGAUUACACAUCAGAAU	3713
GWED542	UAAUUUCUACUCUUGUAGAUCAUGCAAAGUCAGAUUUGUUG	3714
GWED543	UAAUUUCUACUCUUGUAGAUCAUGUGCAAACGCCUUAACAAC	3715
GWED544	UAAUUUCUACUCUUGUAGAUUUCUCAAAACAAAUGUGUCACA	3716
GWED545	UAAUUUCUACUCUUGUAGAUUCUGUGAUUACACAUCAGAAUC	3717
GWED546	UAAUUUCUACUCUUGUAGAUAGUCUCUCAGCUGGUACACGGC	3718
GWED547	UAAUUUCUACUCUUGUAGAUUGACACAUUUGUUUGAGAAUCA	3719
GWED548	UAAUUUCUACUCUUGUAGAUUUGCUCCAGGCCACAGCACUGUU	3720
GWED549	UAAUUUCUACUCUUGUAGAUUUCUCAAAACAAAUGUGUCACAA	3721

[0435] [0436] 通过电穿孔递送每种RNP复合物,并且通过流式细胞术(FCM)和T7E1测定评估各个RNP靶向和切割TRAC基因座的能力。

[0437] AsCpf1酶的纯化

[0438] 通过基因合成与附加的SapI Electra克隆臂(DNA 2.0克隆系统)和C末端核质蛋白NLS获得对应于AsCpf1的1307个残基的3.9Kb基因区段。将克隆到pUC57载体中的合成构建体通过SapI消化进行切割并Electra克隆(DNA 2.0,Electra克隆系统)到pD441-NH、pD441-CH(高拷贝)和pD421-NH(低拷贝)大肠杆菌表达载体中。

[0439] 将含有AsCpf1基因(N末端His(高拷贝)、C末端His(高拷贝)或N末端His(低拷贝))的质粒转化或电穿孔到细菌表达菌株中之后,使用以下方法表达和纯化蛋白质。所有AsCpf1构建体在N末端或C末端上还含有C末端核质蛋白NLS序列,但也可以含有SV40NLS。在

将质粒转化或电穿孔到蛋白质表达细菌细胞(例如,Rosetta 2)中之后,将若干所得菌落加入0.5mL脑心肺(BHL)培养基或另一种不含抗生素的富培养基中。30分钟至1小时之后,或当细胞悬浮液明显混浊时,向培养物中加入0.5mL BHL培养基或其他富培养基和抗生素(氯霉素和卡那霉素)。一旦培养物变得明显混浊( $OD=0.6$ )便将培养体积加倍,直至体积达到8mL。然后将完整的培养物转移至1L Terrific肉汤(Teknova公司)培养基+抗生素+1mL 1000x金属溶液(Teknova公司)+200 $\mu$ L 1M硫酸镁溶液中。使培养物在37 $^{\circ}$ C下生长。在1-2小时之后或当培养烧瓶变得轻微混浊时测量培养物OD。此时,每1小时或在适当时测量OD直至OD达到1.0-1.5。一旦OD达到1.0-1.5,便将烧瓶转移至较低温度(18 $^{\circ}$ C至25 $^{\circ}$ C)并在30分钟至1小时后再次检查OD。当培养物达到约2.0的OD时,通过添加IPTG诱导蛋白质表达。使细胞在18 $^{\circ}$ C下生长12-16小时(这个时间可以根据需要变化直至3天)。然后收获培养物并使用大型离心机沉淀并立即裂解或在-80 $^{\circ}$ C下冷冻至裂解当天。

[0440] 使用微流化器裂解表达AsCpf1的细胞沉淀,其中将1-10g干细胞沉淀用70mL裂解缓冲液(50mM Tris pH 8.0,1mM TCEP[三(2-羧乙基)膦]或DTT(二硫苏糖醇),10%-20%甘油和300-1000mM NaCl或KCl)重悬浮。可替代地,将细胞裂解并使用BPER(赛默飞世尔公司(ThermoFisher))化学裂解试剂盒、BugBuster(EMD密理博公司(EMD Millipore))化学裂解试剂盒或含有相同缓冲液但在1%Triton X-100存在下的内部化学裂解试剂分解细胞膜来提取折叠的AsCpf1。最后,使用微流化器方法中相同的裂解缓冲液但不使用1%Triton X-100或任何其他温和的两性离子洗涤剂,可以使用超声波仪来裂解细胞。

[0441] 将细胞裂解物在离心机中离心并弃去细胞碎片沉淀。将上清液通过0.2 $\mu$ m或0.45 $\mu$ m过滤器过滤并加样到HisTrap Ni-NTA柱(通用医疗集团(GE Healthcare))或使用HISpur Ni-NTA浆液(赛默飞世尔公司)的重力柱上。在两种情况下,将HisTrap或重力柱中的浆液用裂解缓冲液平衡,然后暴露于用还含有30mM咪唑的裂解缓冲液洗涤的若干5x-20x柱体积洗涤剂。最后,使用具有250-500mM咪唑的裂解缓冲液,来自HisTrap或重力柱的Ni-NTA树脂上洗脱His标记的AsCpf1。

[0442] 然后使用截留分子量为100kDa或更低的过滤器将AsCpf1蛋白浓缩至大约5mL并加样到配有尺寸排阻柱的AKTA Pure(通用医疗集团)FPLC仪器上。可替代地,由于6xHis-AsCpf1-NLS构建体的净正电荷(+8),AsCpf1也可以加样到阳离子交换柱上并使用100mM至1000mM NaCl梯度在40分钟内纯化。NaCl也可以用KCl代替。可替代地,具有3X FLAG标签(-7净电荷)且无NLS(+5净电荷)的AsCpf1(+3净电荷)将带负电荷并且能够结合阴离子交换柱且能够用增加的盐梯度纯化。用于尺寸排阻纯化方法或阳离子交换纯化方法的缓冲液是50mM HEPES(pH 7.5)、1mM TCEP或DTT、10%-20%甘油以及250mM NaCl恒定离子强度(尺寸排阻)或100mM至1000mM NaCl离子强度(阳离子交换)。HEPES可以用缓冲能力在pH 6.5-8.5范围内的任何缓冲液代替。将AsCpf1蛋白洗脱成2mL级分,产生一个主要峰,如在FPLC仪器上通过UV吸光度检测的。合并来自此峰的级分,并且确定存在AsCpf1并且纯度高,如通过SDS-PAGE分析的,SDS-PAGE显示在预期的150kDa分子量标记处的清晰条带而没有污染条带。此外,AsCpf1吸收光谱的检查显示干净的蛋白质而没有可测量的核酸污染,如通过260/280UV吸光度比测量的。

[0443] 基于如在NCBI网站上的UniProt Prot-param工具和所测量的280吸光度预测的AsCpf1消光系数(143,940M<sup>-1</sup>cm<sup>-1</sup>)将AsCpf1的合并级分浓缩至50 $\mu$ M。将这些等分试样储存

在4℃下立即使用或在液氮中快速冷冻并储存在-80℃下长期储存。

[0444] 通过化学合成产生crRNA,并且将其在递送至细胞之前通过将蛋白质和crRNA在室温下孵育至少15分钟与上述经纯化的蛋白质复合。将经复合的RNP用于两个单独的测定中。将RNP用于体外切割测定,其中将RNP与对应于TRAC基因座的外显子1的PCR产物一起孵育。将RNP与PCR产物以1:1摩尔比孵育并允许在37℃下孵育15分钟。孵育后,将反应用蛋白酶K在42℃下处理20分钟。在PAGE凝胶上对切割进行可视化。虽然观察到大多数RNP的切割,但两种RNP显示出显著的切割(图2)。还将一部分每种RNP复合物通过电穿孔以1 $\mu$ g/100,000个细胞的比率引入激活的原代人CD4<sup>+</sup>T细胞(在补充有IL-2、IL-7、IL-15的完全培养基中培养)中。使用对TCR $\alpha$ / $\beta$ 有特异性的Brilliant Violet 421(生物传奇公司(BioLegend))缀合的抗体通过FCM在电穿孔后4天监测细胞上的TCR $\alpha$ / $\beta$ 表达。与不影响TCR $\alpha$ / $\beta$ 表达的许多RNP相比,两种RNP(GWED545和GWED546)显著降低了转染细胞上TCR $\alpha$ / $\beta$ 的表达(图3)。另外,用Cpf1RNP处理不会对细胞活力产生负面影响(图4)。为了确认TCR $\alpha$ / $\beta$ 阴性细胞的产生是TRAC基因座处基因组编辑的结果,收获gDNA并进行T7E1测定。简言之,T7E1测定涉及450bp PCR产物的扩增、纯化和尺寸验证,PCR产物的变性以及通过加热至95℃并且然后缓慢冷却重新杂交。然后用识别并切割非完全匹配DNA的T7内切核酸酶I(或其他错配敏感酶)消化杂交的PCR产物。如果indel存在于原始模板DNA中,当使扩增子变性并重退火时,这导致具有不同indel的DNA链杂交并且因此产生不完全匹配的双链DNA。可以通过凝胶电泳或通过毛细管电泳使消化产物可视化。被切割DNA的分数(切割产物的密度除以切割和未切割的密度)用于使用以下等式估计百分比NHEJ: %NHEJ = (1 - (1 - 切割的分数)<sup>1/2</sup>)。T7E1测定对低至约2%-5%的NHEJ是敏感的。实际上,数据证实对于降低TCR $\alpha$ / $\beta$ 表达的RNP在TRAC基因座处存在DNA修饰(图5)。

[0445] 为了确定Cpf1是否可以切割来自超过一个供体的人T细胞,将包含GWED545和GWED546的RNP电穿孔到来自第二供体的激活的人T细胞中。使用对TCR $\alpha$ / $\beta$ 有特异性的Brilliant Violet 421(生物传奇公司)缀合的抗体通过FCM在电穿孔后第4天通过FACS分析评估编辑效率。通过TCR $\alpha$ / $\beta$ 表达的丧失证实了这两种RNP跨两个供体编辑人T细胞的能力(图6)。

[0446] 通过引用并入

[0447] 本文提及的所有出版物、专利、序列表和专利申请都通过引用以其整体而特此并入,如同每一单独的出版物、专利、序列表或专利申请具体且单独地指明通过引用而并入一样。在有冲突的情况下,以本申请(包括本文的任何定义)为准。

[0448] 等效物

[0449] 本领域的普通技术人员仅使用常规实验就应认识到或能够确定本文所描述的本发明的具体实施例的许多等效物。此类等效物旨在由以下权利要求书涵盖。某些实施例也在以下权利要求内。

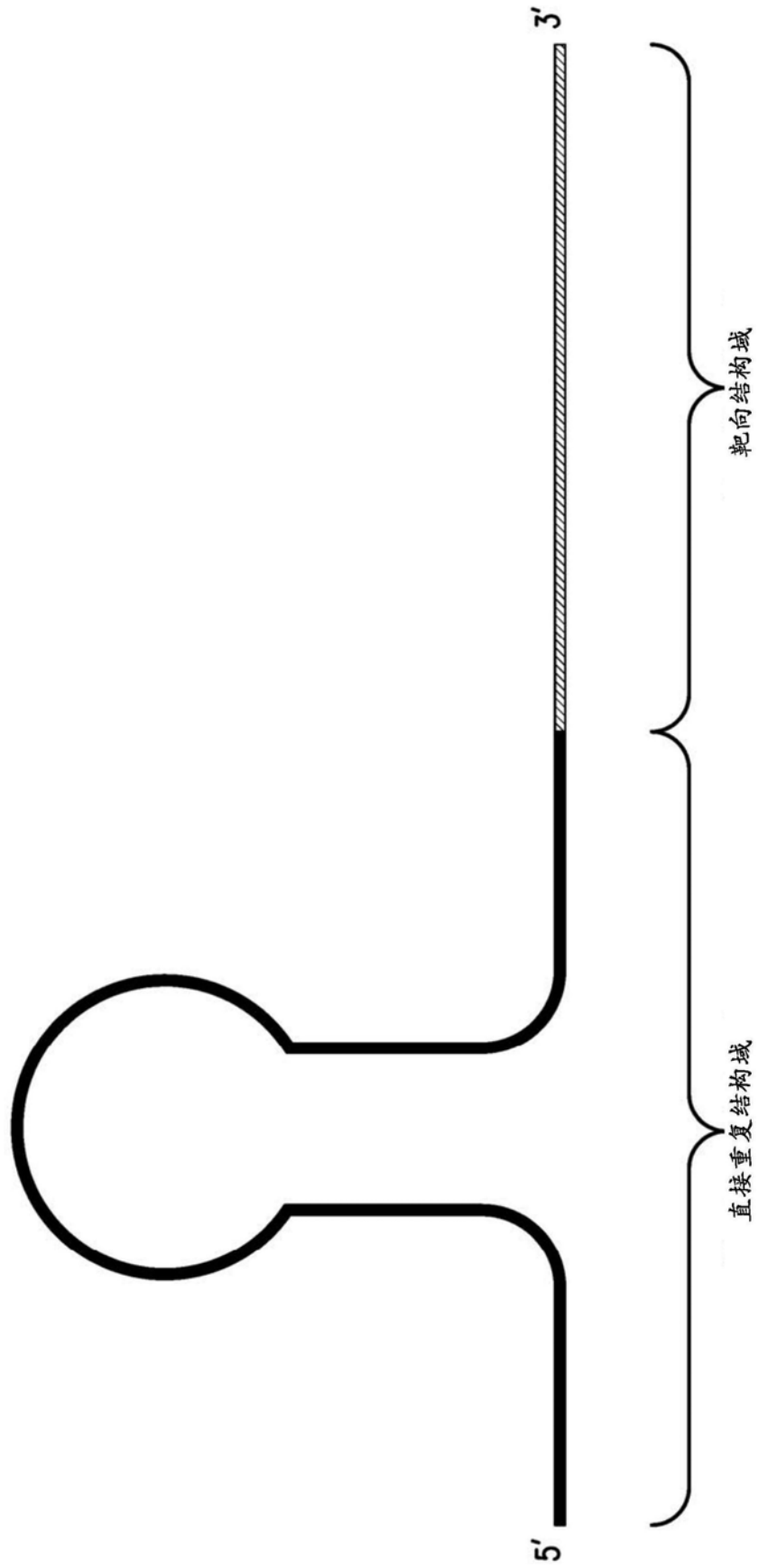
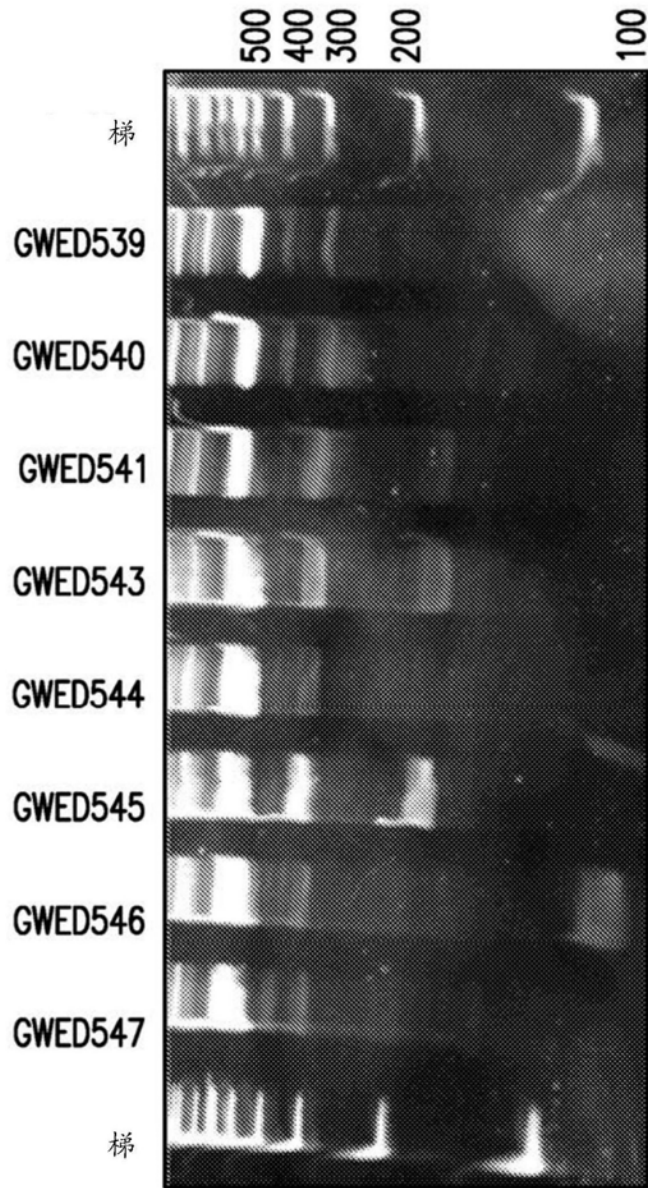


图1



crRNA ID:	预期切割尺寸:
GWED539	142, 331
GWED540	146, 327
GWED541	190, 283
GWED543	184, 289
GWED544	159, 314
GWED545	189, 284
GWED546	96, 377
GWED547	157, 136

全长PCR产物 = 450bp

图2

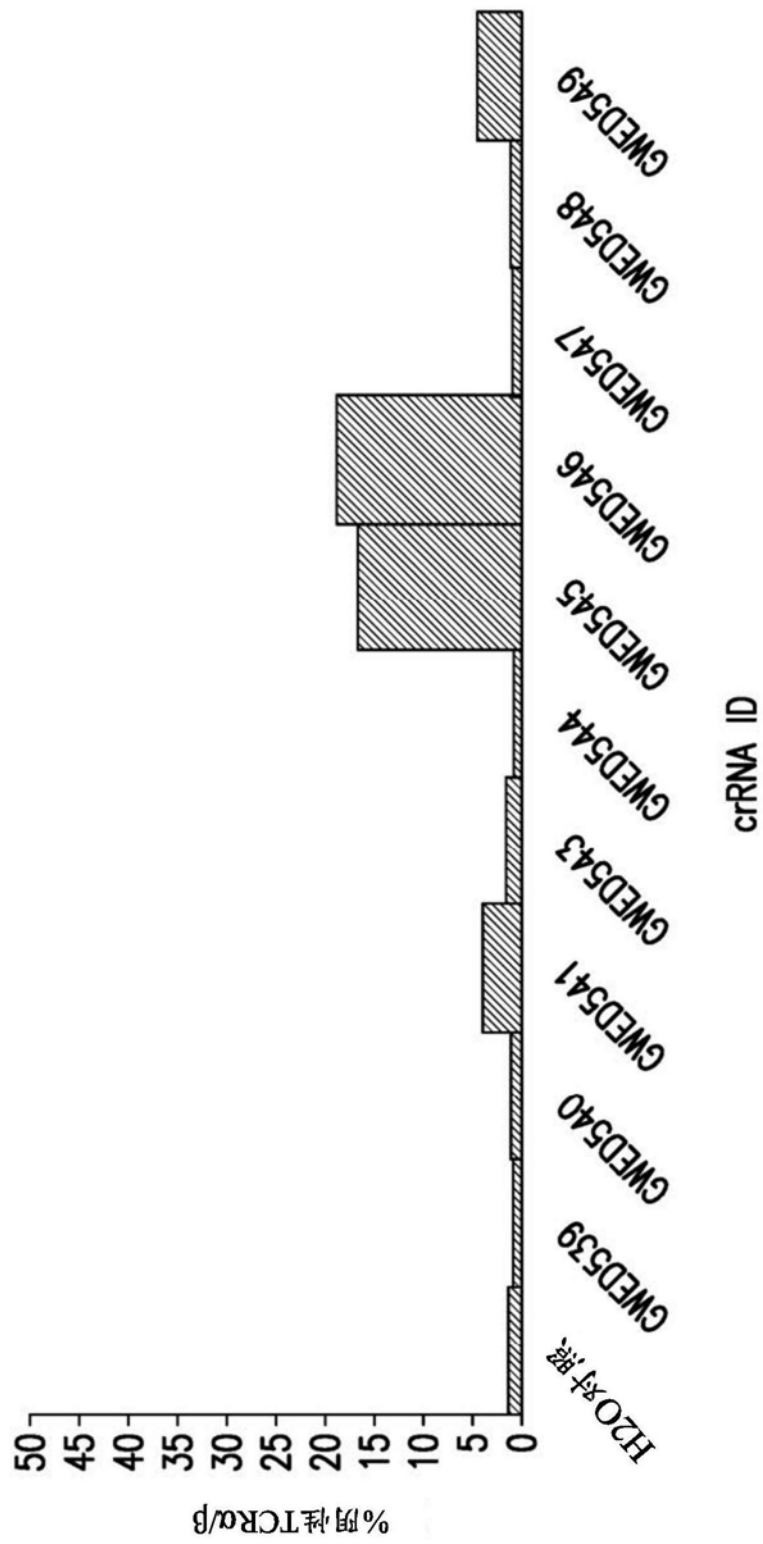


图3A

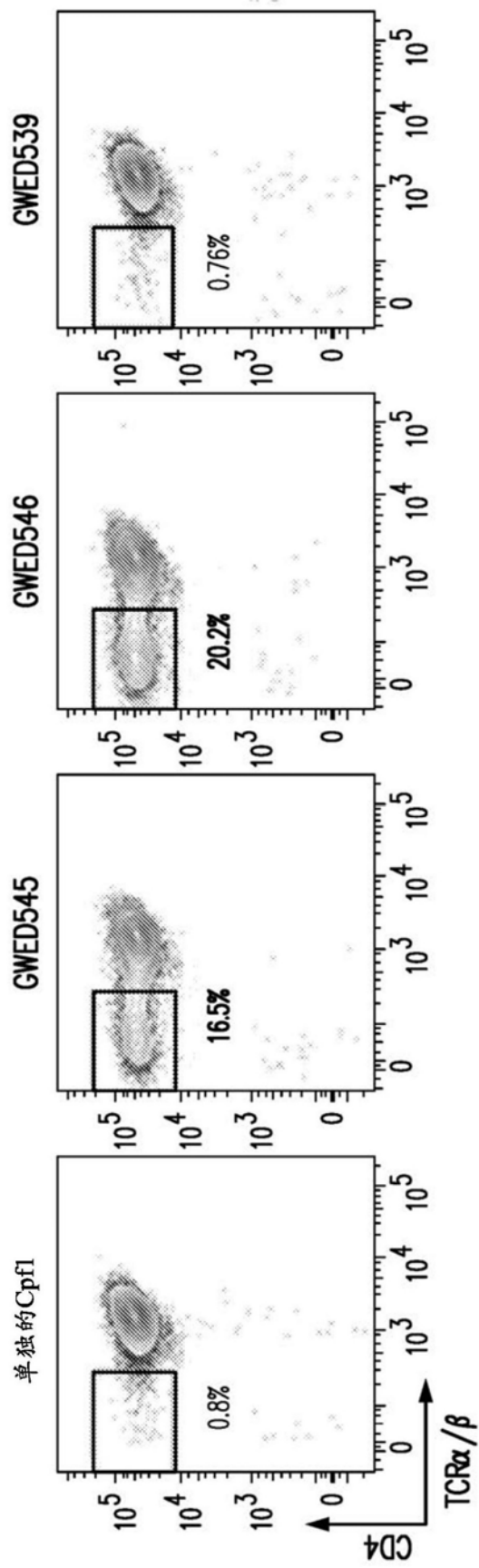


图3B

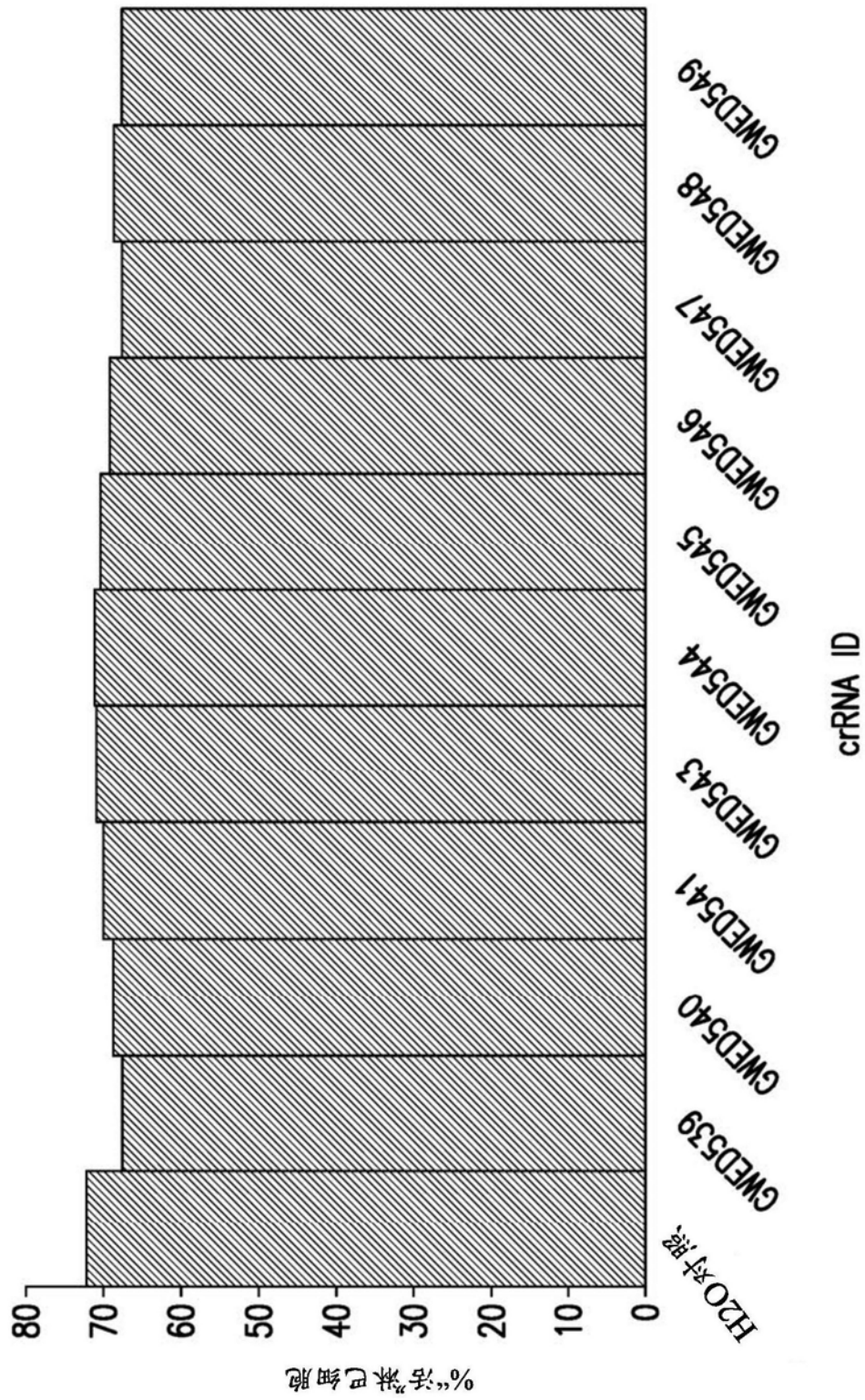


图4

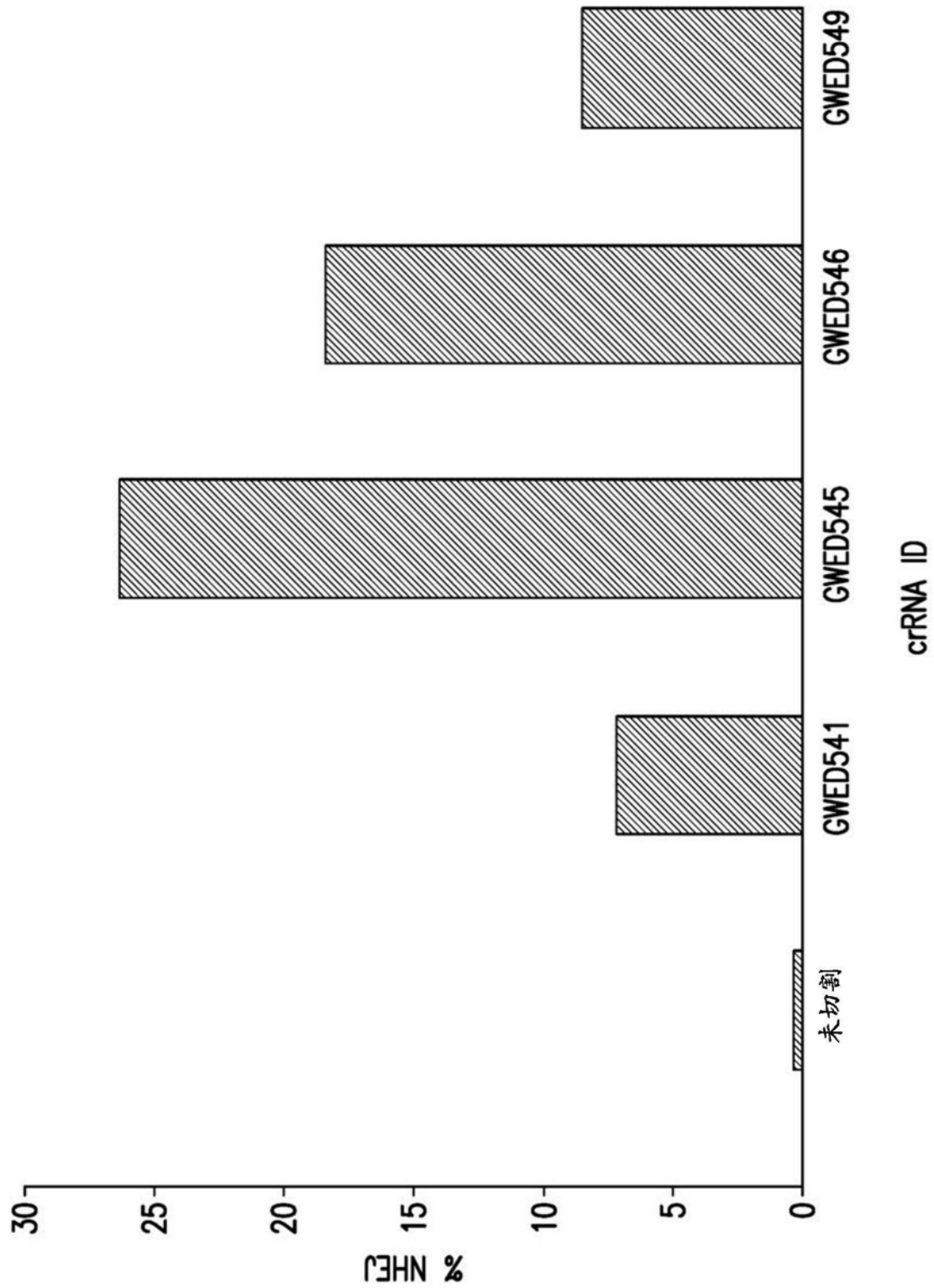


图5

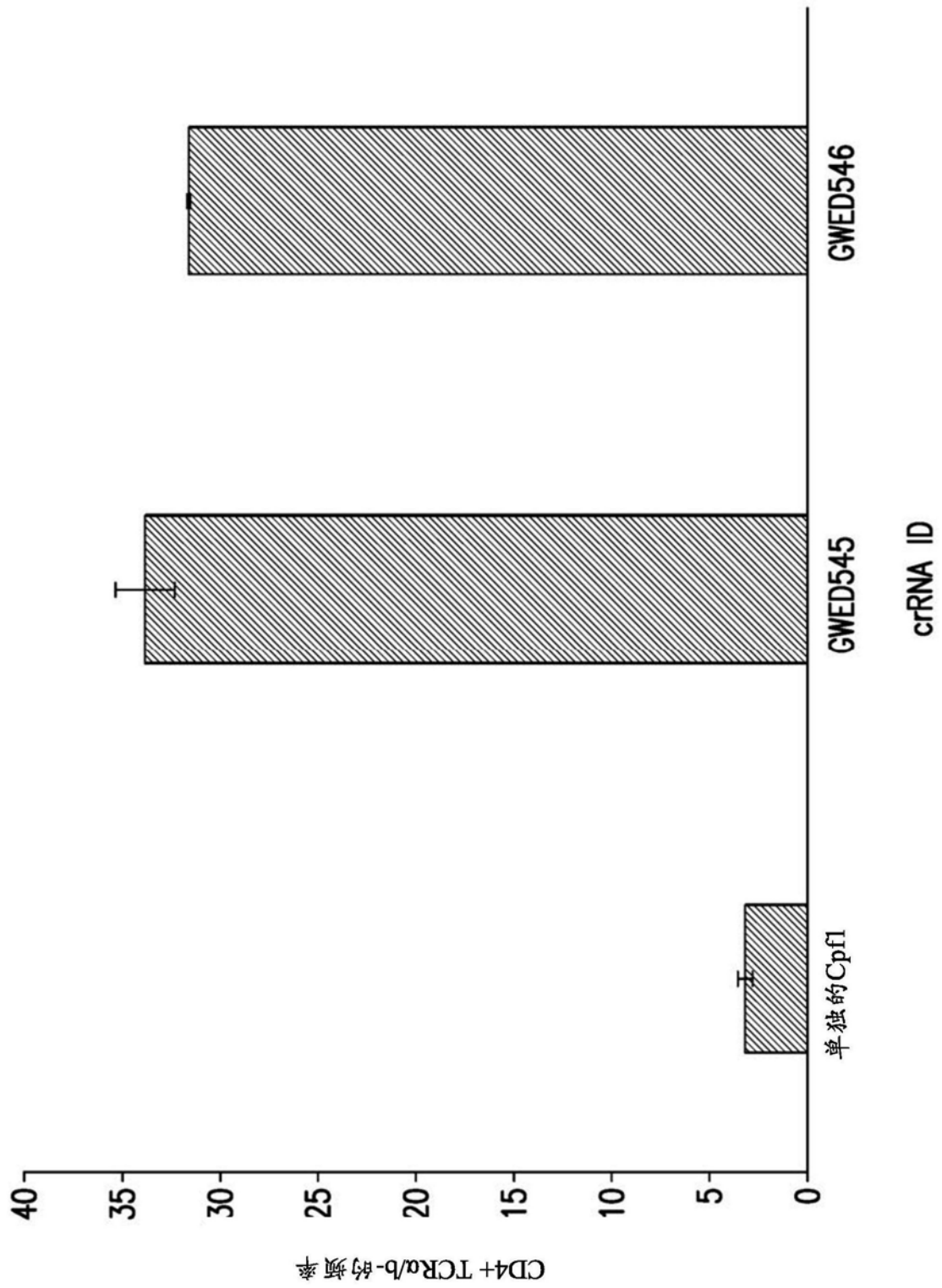


图6

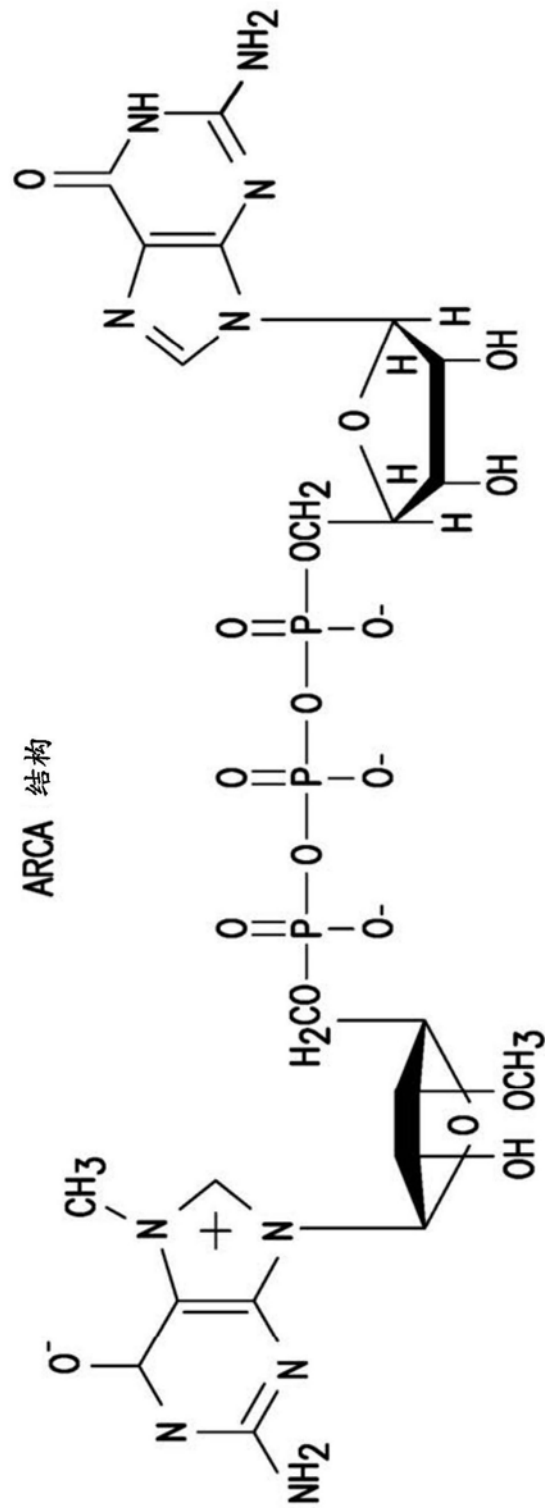


图7