



(12) 发明专利申请

(10) 申请公布号 CN 115491373 A

(43) 申请公布日 2022.12.20

(21) 申请号 202211088995.2

(51) Int.Cl.

(22) 申请日 2016.10.28

C12N 15/10 (2006.01)

(30) 优先权数据

C12N 15/113 (2010.01)

62/249,071 2015.10.30 US

C12N 9/22 (2006.01)

62/249,159 2015.10.30 US

(62) 分案原申请数据

201680074667.3 2016.10.28

(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 A·E·弗里德兰 P·奥唐纳尔

D·A·邦姆克罗特

(74) 专利代理机构 北京市君合律师事务所

11517

专利代理人 张怡 张璐

权利要求书10页 说明书137页 附图30页

(54) 发明名称

治疗单纯疱疹病毒的CRISPR/CAS相关方法  
及组合物

(57) 摘要

本发明描述了用于编辑人类细胞中RS1、RL2和/或LAT基因的CRISPR/CAS相关系统、组合物和方法,本发明还描述了细胞和组合物,包括根据其编辑的细胞。

1. 一种基因组编辑系统,所述系统包括:

gRNA分子,所述gRNA分子包括靶向结构域,所述靶向结构域与单纯疱疹病毒(HSV)病毒基因的靶序列互补,所述HSV病毒基因的靶序列选自由RS1基因、RL2基因和LAT基因组成的群组;以及

Cas9分子。

2. 根据权利要求1所述的基因组编辑系统,其中所述靶向结构域被构造成在约500bp、约450bp、约400bp、约350bp、约300bp、约250bp、约200bp、约150bp、约100bp、约50bp、约25bp或约10bp的HSV靶位置中形成双链断裂或单链断裂,从而改变所述HSV病毒基因。

3. 根据权利要求2所述的基因组编辑系统,其中所述改变HSV病毒基因包括敲除所述HSV病毒基因、敲低所述HSV病毒基因或同时敲除和敲低所述HSV病毒基因。

4. 根据权利要求1-3中任一项所述的基因组编辑系统,其中所述靶向结构域被构造成靶向所述HSV病毒基因的编码区或非编码区,其中所述非编码区包括所述HSV病毒基因的启动子区、增强子区、内含子、3' UTR、5' UTR或多聚腺苷酸化信号区;且所述编码区包括所述HSV病毒基因的早期编码区。

5. 根据权利要求1-4中任一项所述的基因组编辑系统,其中所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:208至58749的核苷酸序列相同,或具有不超过3个核苷酸的不同。

6. 根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲除所述HSV-1RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:208至2509组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

7. 根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:7098至9292组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

8. 根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲除所述HSV-1RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:21324至22744组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

9. 根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:26613至28037组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

10. 根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲除所述HSV-1LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:31730至32746组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

11. 根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2LAT基因,且所述靶向结构域包括核苷酸

序列,所述核苷酸序列与选自由SEQ ID NO:35617至36926组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

12.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-1RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:2510至7073组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

13.根据权利要求12所述的基因组编辑系统,其中所述靶向结构域包括选自由SEQ ID NO:243、2515、3362和3363组成的群组中的核苷酸序列。

14.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:9293至13614组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

15.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-1RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:22745至26601组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

16.根据权利要求15所述的基因组编辑系统,其中所述靶向结构域包括选自由SEQ ID NO:23519、23527、23535、23571和23583组成的群组中的核苷酸序列。

17.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:28038至31720组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

18.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-1LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:32747至35600组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

19.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:36927至40871组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

20.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:13637至14794组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

21.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:17753至18784组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

22.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球

菌Cas9分子,所述基因组编辑系统敲低所述HSV-1RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:40886至42078组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

23.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:49498至50652组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

24.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:45340至46479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

25.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:53858至55056组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

26.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:14795至17741组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

27.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:18785至21311组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

28.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:42079至45315组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

29.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:50653至53824组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

30.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:46480至49479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

31.根据权利要求1-5中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:55057至58731组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

32. 根据权利要求6-11和20-25中任一项所述的基因组编辑系统,其中所述化脓链球菌Cas9分子识别前间区序列邻近基序(Protospacer Adjacent Motif,PAM)NGG。

33. 根据权利要求12-19和26-31中任一项所述的基因组编辑系统,其中所述金黄色葡萄球菌Cas9分子识别NNGRRT (SEQ ID NO:204) 或NNGRRV (SEQ ID NO:205) 中任一个PAM。

34. 根据权利要求1-33中任一项所述基因组编辑系统,其中所述Cas9分子选自由酶活化Cas9 (eaCas9) 分子、酶失活Cas9 (eiCas9) 分子和eiCas9融合蛋白组成的群组。

35. 根据权利要求34所述的基因组编辑系统,其中所述eaCas9分子包括HNH样结构域切割活性,但不具有或不显著具有N-末端RuvC样结构域切割活性。

36. 根据权利要求34所述的基因组编辑系统,其中所述eaCas9分子是HNH样结构域切口酶。

37. 根据权利要求34所述的基因组编辑系统,其中所述eaCas9分子包括N-末端RuvC样结构域切割活性,但不具有或不显著具有HNH样结构域切割活性。

38. 根据权利要求34所述的基因组编辑系统,其中所述eaCas9分子是N-末端RuvC样结构域切口酶。

39. 根据权利要求1-38中任一项所述的基因组编辑系统,其中所述Cas9分子包括野生型Cas9分子、突变型Cas9分子或二者的组合。

40. 根据权利要求39所述的基因组编辑系统,其中所述突变型Cas9分子包括选自由D10、E762、D986、H840、N854、N863和N580组成的群组中的突变。

41. 根据权利要求1-40中任一项所述的基因组编辑系统,其中所述Cas9分子是金黄色葡萄球菌Cas9分子或化脓链球菌Cas9分子。

42. 根据权利要求1-41中任一项所述的基因组编辑系统,其中所述gRNA是模块化gRNA分子或嵌合gRNA分子。

43. 根据权利要求1-42中任一项所述的基因组编辑系统,其中所述靶向结构域具有16、17、18、19、20、21、22、23、24、25或26个核苷酸的长度。

44. 根据权利要求1-43中任一项所述的基因组编辑系统,其中所述gRNA分子包括,从5'到3' :

靶向结构域;

第一互补结构域;

连接结构域;

第二互补结构域;

近端结构域;以及

尾部结构域。

45. 根据权利要求44所述的基因组编辑系统,其中所述连接结构域的长度不超过25个核苷酸。

46. 根据权利要求44或45所述的基因组编辑系统,其中所述近端结构域和所述尾部结构域一起长度为至少20、至少25、至少30或至少40个核苷酸。

47. 根据权利要求1-46中任一项所述的基因组编辑系统,所述系统包括两个、三个或四个gRNA分子。

48. 根据权利要求1-47中任一项所述的基因组编辑系统,其用于改变细胞中所述HSV病

毒基因。

49. 根据权利要求48所述的基因组编辑系统,其中所述细胞感染了HSV。

50. 一种组合物,所述组合物包括gRNA分子,所述gRNA分子包括与HSV病毒基因的靶序列互补的靶向结构域,所述HSV病毒基因选自由RS1基因、RL2基因和LAT基因组成的群组。

51. 根据权利要求50所述的组合物,其中所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:208至58749的核苷酸序列相同,或具有不超过3个核苷酸不同。

52. 根据权利要求50或51所述的组合物,所述组合物包括一个、两个、三个或四个gRNA分子。

53. 根据权利要求50-52中任一项所述的组合物,所述组合物还包括至少一个Cas9分子。

54. 根据权利要求53所述的组合物,其中所述至少一个Cas9分子是化脓链球菌Cas9分子或金黄色葡萄球菌Cas9分子。

55. 根据权利要求53或54所述的组合物,其中所述至少一个Cas9分子包括野生型Cas9分子、突变型Cas9分子或二者的组合。

56. 根据权利要求55所述的组合物,其中所述突变型Cas9分子包括选自由D10、E762、D986、H840、N854、N863和N580组成的群组中的突变。

57. 根据权利要求53所述的组合物,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:208至2509和13637至14794组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

58. 根据权利要求53所述的组合物,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:7098至9292和17753至18784组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

59. 根据权利要求53所述的组合物,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:21324至22744和40886至42078组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

60. 根据权利要求53所述的组合物,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2RL2基因的靶序列互补的靶向结构域且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:26613至28037和49498至50652组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

61. 根据权利要求53所述的组合物,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:31730至32746和45340至46479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

62. 根据权利要求53所述的组合物,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:35617至36926和53858至55056组成的群组中的

核苷酸序列相同,或具有不超过3个核苷酸的不同。

63. 根据权利要求53所述的组合物,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:2510至7073和14795至17741组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

64. 根据权利要求63所述的组合物,其中所述靶向结构域包括选自由SEQ ID NOS:243、2515、3362和3363组成的群组中的核苷酸序列。

65. 根据权利要求53所述的组合物,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:9293至13614和18785至21311组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

66. 根据权利要求53所述的组合物,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:22745至26601和42079至45315组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

67. 根据权利要求66所述的组合物,其中所述靶向结构域包括选自由SEQ ID NOS:23519、23527、23535、23571和23583组成的群组中的核苷酸序列。

68. 根据权利要求53所述的组合物,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:28038至31720和50653至53824组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

69. 根据权利要求53所述的组合物,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1LAT基因的靶序列互补靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:32747至35600和46480至49479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

70. 根据权利要求53所述的组合物,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:36927至40871和55057至58731组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

71. 根据权利要求57-62中任一项所述的组合物,其中所述化脓链球菌Cas9分子识别前间区序列邻近基序(Protospacer Adjacent Motif, PAM) NGG。

72. 根据权利要求63-70中任一项所述的基因组编辑系统,其中所述金黄色葡萄球菌Cas9分子识别NNGRRT (SEQ ID NO:204) 或NNGRRV (SEQ ID NO:205) 中任一个PAM。

73. 根据权利要求50-72中任一项所述的组合物用于改变细胞中的所述HSV病毒基因。

74. 根据权利要求73所述的组合物,其中所述细胞感染了HSV。

75. 一种载体,所述载体包括编码gRNA分子的多核苷酸,所述gRNA分子包括与HSV病毒基因的靶序列互补的靶向结构域,所述HSV病毒基因选自由RS1基因、RL2基因和LAT基因组成的群组。

76. 根据权利要求75所述的载体,其中所述靶向结构域包括核苷酸序列,所述核苷酸序

列与选自SEQ ID NO:208至58749的核苷酸序列相同,或具有不超过3个核苷酸的不同。

77.根据权利要求75或76中任一项所述的载体,所述载体还包括编码Cas9分子的多核苷酸。

78.根据权利要求77所述的载体,其中所述至少一个Cas9分子是化脓链球菌Cas9分子或金黄色葡萄球菌Cas9分子。

79.根据权利要求77或78所述的载体,其中所述至少一个Cas9分子包括野生型Cas9分子、突变型Cas9分子或二者的组合。

80.根据权利要求80所述的载体,其中所述突变型Cas9分子包括选自由D10、E762、D986、H840、N854、N863和N580组成的群组中的突变。

81.根据权利要求77所述的载体,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:208至2509和13637至14794组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

82.根据权利要求77所述的载体,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:7098至9292和17753至18784组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

83.根据权利要求77所述的载体,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:21324至22744和40886至42078组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

84.根据权利要求77所述的载体,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:26613至28037和49498至50652组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

85.根据权利要求77所述的载体,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:31730至32746和45340至46479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

86.根据权利要求77所述的载体,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:35617至36926和53858至55056组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

87.根据权利要求77所述的载体,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:2510至7073和14795至17741组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

88.根据权利要求87所述的载体,其中所述靶向结构域包括选自由SEQ ID NO:243、2515、3362和3363组成的群组中的核苷酸序列。

89. 根据权利要求77所述的载体,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:9293至13614和18785至21311组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

90. 根据权利要求77所述的载体,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:22745至26601和42079至45315组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

91. 根据权利要求90所述的载体,其中所述靶向结构域包括选自由SEQ ID NO:23519、23527、23535、23571和23583组成的群组中的核苷酸序列。

92. 根据权利要求77所述的载体,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:28038至31720和50653至53824组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

93. 根据权利要求77所述的载体,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1LAT基因的靶序列互补靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:32747至35600和46480至49479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

94. 根据权利要求77所述的载体,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2LAT基因的靶序列互补靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:36927至40871和55057至58731组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

95. 根据权利要求81-86中任一项所述的载体,其中所述化脓链球菌Cas9分子识别前间区序列邻近基序(Protospacer Adjacent Motif, PAM) NGG。

96. 根据权利要求87-94中任一项所述的基因组编辑系统,其中所述金黄色葡萄球菌Cas9分子识别NNGRRT (SEQ ID NO:204) 或NNGRRV (SEQ ID NO:205) 中任一个PAM。

97. 根据权利要求75-96中任一项所述的载体,其中所述载体是病毒载体。

98. 根据权利要求75-97中任一项所述的载体用于改变细胞中的所述HSV病毒基因。

99. 根据权利要求98所述的载体,其中所述细胞感染了HSV。

100. 一种改变HSV病毒基因的方法,所述HSV病毒基因选自由细胞中的RS1基因、RL2基因和LAT基因组成的群组,所述方法包括给予细胞下述内容中的一个:

(i) 基因编辑系统,所述基因编辑系统包括gRNA分子和至少一个Cas9分子,所述gRNA分子包括与所述HSV病毒基因的靶向序列互补的靶向结构域;

(ii) 载体,所述载体包括编码gRNA分子的多核苷酸和编码Cas9分子的多核苷酸,所述gRNA分子包括与所述HSV病毒基因的靶向序列互补的靶向结构域;或

(iii) 组合物,所述组合物包括gRNA分子和至少一个Cas9分子,所述gRNA分子包括与所述HSV病毒基因的靶向序列互补的靶向结构域。

101. 根据权利要求100所述的方法,其中所述改变包括敲除所述HSV病毒基因、敲低所述HSV病毒基因或同时敲除和敲低所述HSV病毒基因。

102. 根据权利要求100或101所述的方法,其中所述细胞感染了HSV。

103. 根据权利要求100-102中任一项所述的方法,其中所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:208至58749的核苷酸序列相同,或具有不超过3个核苷酸的不同。

104. 根据权利要求100-103中任一项所述的方法,其中所述Cas9分子是金黄色葡萄球菌Cas9分子或化脓链球菌Cas9分子。

105. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:208至2509和13637至14794组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

106. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:7098至9292和17753至18784组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

107. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:21324至22744和40886至42078组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

108. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:26613至28037和49498至50652组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

109. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-1LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:31730至32746和45340至46479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

110. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是化脓链球菌Cas9分子,所述gRNA分子包括与HSV-2LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:35617至36926和53858至55056组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

111. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:2510至7073和14795至17741组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

112. 根据权利要求111所述的方法,其中所述靶向结构域包括选自由SEQ ID NO:243、2515、3362和3363组成的群组中的核苷酸序列。

113. 根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2RS1基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:9293至13614和18785至

21311组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

114.根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:22745至26601和42079至45315组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

115.根据权利要求114所述的方法,其中所述靶向结构域包括选自由SEQ ID NO:23519、23527、23535、23571和23583组成的群组中的核苷酸序列。

116.根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2RL2基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:28038至31720和50653至53824组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

117.根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-1LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:32747至35600和46480至49479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

118.根据权利要求100-104中任一项所述的方法,其中所述Cas9分子是金黄色葡萄球菌Cas9分子,所述gRNA分子包括与HSV-2LAT基因的靶序列互补的靶向结构域,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NO:36927至40871和55057至58731组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

119.一种gRNA分子,所述gRNA分子包括与HSV病毒基因的靶序列互补的靶向结构域,所述HSV病毒基因选自由细胞中的RS1基因、RL2基因和LAT基因组成的群组。

120.根据权利要求119所述的gRNA分子,其中所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:208至58749的核苷酸序列相同,或具有不超过3个核苷酸的不同。

121.一种细胞,所述细胞包括根据权利要求1-47中任一项所述的基因组编辑系统、权利要求50-72中任一项所述的组合物或权利要求75-97中任一项所述的载体。

122.根据权利要求121所述的细胞,其中所述细胞选自由上皮细胞、神经细胞和视神经细胞组成的群组。

## 治疗单纯疱疹病毒的CRISPR/CAS相关方法及组合物

[0001] 本申请是申请号为201680074667.3、申请日为2016年10月28日、发明名称为“治疗单纯疱疹病毒的CRISPR/CAS相关方法及组合物”的中国发明专利申请的分案申请，原申请为国际申请号为PCT/US2016/059502的国际申请的国家阶段申请，该国际申请要求2015年10月30日提交的美国临时专利申请第62/249,071号和2015年10月30日提交的美国临时专利申请第62/249,159号的优先权。

[0002] 优先权要求

[0003] 本申请要求2015年10月30日提交的美国临时申请No.62/249,071以及2015年10月30日提交的美国临时申请No.62/249,159的优先权，其内容通过引用整体并入本文。

[0004] 政府利益声明

[0005] 本发明是在国立卫生研究院授予的基金号1R43A1120302-01的资助下完成的。政府对本发明享有一定的权利。

[0006] 序列表

[0007] 本说明书提及序列表(以电子方式于2016年10月28日提交名为“2016-10-28\_084177\_0133\_ST25.txt”的.txt文件)。“2016-10-28\_084177\_0133\_ST25.txt”文件于2016年10月28日生成，文件大小为11,189,396字节。序列表的全部内容在此引入作为参考。

### 技术领域

[0008] 本公开涉及用于编辑靶核酸序列(例如编辑RS1、RL2和/或LAT基因)的CRISPR/CAS相关方法、组合物和基因组编辑系统及其与单纯疱疹病毒(HSV)的应用。

### 背景技术

[0009] 单纯疱疹病毒(HSV)，例如1型单纯疱疹病毒(HSV-1)和2型单纯疱疹病毒(HSV-2)，是普遍存在且具备高度传染性的病原体。HSV-1通常引起口腔和粘膜的间歇性疼痛起泡。HSV-2通常在生殖器部位引起间歇性、疼痛起泡。单纯疱疹病毒可引起终身反复发作的病毒反应。

[0010] 大多数人群在童年期间发生HSV-1感染。成年后，美国高达80%的人口感染HSV-1。新的HSV-1感染发生率为年平均每100人1.6例(Langenberg等，1999年；新英格兰医学杂志341:1432-1438)。HSV-1感染的最严重表现包括例如角膜炎、脑炎和脑膜炎。

[0011] 全球有超过5亿人感染HSV-2。每年首次感染人数高达2300万人。在美国，大约1/5的成人对于HSV-2感染是血清反应阳性的(Xu等人，美国传染病学会第42届年会摘要739；2004年9月30日)。HSV-2的发病率正在增加：自1976年以来，美国成年人中HSV-2血清阳性的存在增加了30% (Fleming等，新英格兰医学杂志1997年；337:1105-11)。感染HSV-2会增加感染艾滋病毒的风险，尤其是在患有活动性损伤的患者中。

[0012] 感染HSV-1和/或HSV-2是永久性的。在最初感染HSV-1或HSV-2后，病毒建立持续宿主一生的潜伏感染。最初感染HSV-1通常会引起口腔粘膜(包括嘴唇、嘴和鼻子)的疼痛起泡。HSV-1初次感染较少影响肛门生殖器区域，而引起生殖器和肛门区域粘膜的疼痛起泡。

最初感染HSV-2通常引起肛门生殖器部位粘膜的疼痛起泡。HSV-2初始感染较少影响口腔,而引起嘴唇、嘴和鼻子粘膜的疼痛起泡。

[0013] 初次感染后,HSV-1和HSV-2在所有对象中建立潜伏感染。在建立潜伏感染后,HSV-1或HSV-2的再激活可以在对象的一生中的任何时间点发作。HSV-1或HSV-2的再激活更可能发生在老年人和免疫功能低下的个体中,包括那些患有癌症的人、那些患有HIV/AIDs的人以及那些接受过实体器官或造血干细胞移植的人身上。

[0014] HSV-1和HSV-2均引起眼疱疹。从历史角度来看,HSV-1一直是主要的眼部疱疹感染的病原体。然而,近年来全球范围内HSV-2相关眼部感染的发病率在不断上升。

[0015] HSV-1或HSV-2的眼部感染可影响眼睛上皮,引起角膜炎。角膜炎是眼疱疹最常见的形式。HSV相关性角膜炎是发达国家感染性失明最常见的原因(Dawson等人,眼科综述,1976年;21 (2) :121-135)。在全世界范围内,每年约有150万例HSV相关眼病和4万例HSV相关失明或严重单眼视力损伤(Krawczyk等人,公共科学图书馆一,2015年;10 (1) :e0116800, Farooq和Shukla,2012年;眼科综述57 (5) :448-462)。美国眼部HSV感染的发病率似乎在上升(Liesegang等人,1989年;眼科档案,107:1155-1159. Baratz等人,2009年,眼科视觉科学研究50e-abstract5044)。每10万人每年有15.6例上皮性角膜炎,美国每年约有5万例(Farooq和Shukla,2012年;眼科综述57 (5) :448-462)。

[0016] 眼疱疹也可能影响视网膜,引起视网膜炎。与HSV相关的角膜炎相比,HSV相关性视网膜炎的发生率较低,但是具有更高的永久性视力损害风险。HSV相关性视网膜炎最常影响成人,并可导致急性视网膜坏死(ARN)。ARN在超过50%的对象中引起永久性视力损伤(Roy等人,眼部免疫学和炎症,2014年;22 (3) :170-174)。

[0017] 新生儿是患有严重HSV-1和HSV-2感染风险的人群。这种疾病在分娩过程中由母亲传染给胎儿。母亲在怀孕期间发生原发性HSV-1或HSV-2感染的情况下,母胎传播的机会最高。新生儿疱疹的发病率约为每100,000名新生儿4-30人(Brown ZA等人,2003年;美国医学协会杂志;289 (2) :203-209,Dinh T-H等人,2008年;性传播疾病;35 (1) :19-21)。新生儿可出现严重的HSV相关性角膜炎、视网膜炎、脑炎和/或脑膜炎。新生儿眼部疱疹可能导致即刻、永久性视力丧失。眼部HSV使新生儿面临后期发展ARN的风险。HSV-1或HSV-2没有治愈或预防性治疗措施。治疗主要在急性感染期间给予。原发性HSV-1或HSV-2感染可采用抗病毒治疗措施,包括阿昔洛韦、伐昔洛韦和泛昔洛韦。这些治疗措施可以减少病毒脱落、减少疼痛并改善损伤的愈合时间。再激活的潜伏感染可能在没有治疗的情况下解决(可能是自限性的)或可能用抗病毒治疗措施来治疗。抗病毒治疗措施可在某些情况下预防性给予,包括在患有最近感染或再激活HSV-1或HSV-2的母亲的分娩期间。

[0018] 正在开发用于预防HSV-1和HSV-2感染的疫苗。然而,在对照临床试验中,疫苗接种效力受到限制。最近针对HSV-1和HSV-2感染的疫苗在预防HSV-1感染方面仅35%有效(Belshe等人,2012年;新英格兰医学杂志366 (1) :34-43)。

[0019] 尽管抗逆转录病毒疗法有所进展,但仍需要治疗、预防和/或减少HSV-1和HSV-2感染,特别是治疗、预防和/或减少HSV-1和HSV-2相关的眼部感染,包括角膜炎和视网膜炎。治愈、预防或治疗HSV-1和HSV-2眼部感染的疗法将优于目前的护理标准。

## 发明内容

[0020] 本发明所述的方法、基因组编辑系统和组合物提供了对单纯疱疹病毒(HSV)感染例如眼部感染的治疗、预防和/或减轻措施。

[0021] 本发明所述的方法、基因组编辑系统和组合物可用于提供治疗、预防和/或减轻单纯疱疹病毒眼部感染(包括由1型单纯疱疹病毒(HSV-1)和/或2型单纯疱疹病毒(HSV-2)或其症状引起的眼部感染),或治疗、预防和/或减轻其症状,例如,可通过改变(例如,敲除和/或敲低)HSV-1或HSV-2病毒基因中的一个或多个来提供,例如,通过敲除和/或敲低RS1、RL2、和/或LAT基因中的一个、两个或三个来进行。RL2包括开放染色质区,所述开放染色质区与潜伏期、再激活和裂解感染期间的LAT基因和HSV-1病毒基因表达的调节相关(J.Gen.Viro.,2008年1月;89(Pt 1):68-77)。

[0022] 本发明所述的方法、基因组编辑系统和组合物通过敲除RS1、RL2、和/或LAT基因来提供对单纯疱疹病毒眼部感染或其症状的治疗、预防和/或减轻,所述单纯疱疹病毒眼部感染包括由HSV-1和/或HSV-2引起的眼部感染。本发明所述的方法、基因组编辑系统和组合物通过敲低RS1、RL2、和/或LAT基因来提供对单纯疱疹病毒眼部感染或其症状的治疗、预防和/或减轻,所述单纯疱疹病毒眼部感染包括由HSV-1和/或HSV-2引起的眼部感染。本发明所述的方法、基因组编辑系统和组合物通过同时敲除和敲低RS1、RL2、和/或LAT基因来提供对单纯疱疹病毒眼部感染或其症状的治疗、预防和/或减轻,所述单纯疱疹病毒眼部感染包括由HSV-1和/或HSV-2引起的眼部感染。

[0023] 本发明所述的方法、基因组编辑系统和组合物通过改变RS1、RL2、或LAT基因内的一个或多个位置而导致其从受感染的细胞中被破坏和/或被消除来提供对单纯疱疹病毒(HSV)眼部感染的治疗、预防和/或减轻,所述单纯疱疹病毒眼部感染包括由HSV-1和/或HSV-2或其症状引起的眼部感染。

[0024] 在一个方面中,本发明所述的方法、基因组编辑系统和组合物可用于通过靶向基因(例如,靶向所述基因的非编码或编码区)来改变(例如,敲除或敲低)RS1、RL2、和/或LAT基因中的一种、两种或三种的表达以治疗、预防和/或减轻HSV-1或HSV-2感染。

[0025] 在某些实施方式中,RS1、RL2和/或LAT基因中的一个、两个或三个的编码序列(例如,编码区(本发明中也称作编码序列))被靶向用于改变以及敲除和/或敲低表达。在某些实施方式中,所述编码区是早期编码区,例如,RS1、RL2和/或LAT基因的早期编码区。例如,而不是作为限制,本发明所述的方法、基因组编辑系统和组合物用于通过靶向RS1、RL2和/或LAT基因中的一个、两个或三个的编码序列(例如,内含子或外显子序列)来改变RS1、RL2和/或LAT基因中的一个、两个或三个以治疗、预防和/或减轻HSV-1或HSV-2感染。在某些实施方式中,所述基因(例如,RS1、RL2和/或LAT基因中的一个、两个或三个的编码序列)被靶向以敲除和/或敲低RS1、RL2和/或LAT基因中的一个、两个或三个,例如,以消除RS1、RL2和/或LAT基因中的一个、两个或三个的表达;和/或以敲除RS1、RL2和/或LAT基因中的一个、两个或三个的一个或多个拷贝,例如,通过引入一种改变,所述改变包括RS1、RL2和/或LAT基因中的一个、两个或三个中的突变(例如,插入或缺失)。在某些实施方式中,所述方法、基因组编辑系统和组合物提供一种改变,所述改变包括RS1、RL2和/或LAT基因中的一个、两个或三个中的插入或缺失。

[0026] 在某些实施方式中,RS1、RL2和/或LAT基因中的一个、两个或三个的早期编码序列

被靶向以敲除或敲低RS1、RL2和/或LAT基因中的一个、两个或三个。在某些实施方式中，靶向影响所述RS1、RL2和/或LAT基因的一个或多个拷贝。在某些实施方式中，靶向敲除或靶向敲低方法减少或消除了一种、两种或多种RS1、RL2和/或LAT基因产物的表达。在某些实施方式中，所述方法、基因组编辑系统和组合物提供一种改变，所述改变包括RS1、RL2和/或LAT基因中的一个、两个或三个中的插入或缺失。

[0027] 在另一个方面中，所述方法、基因组编辑系统和组合物包括所述RS1、RL2和/或LAT基因，例如，启动子、增强子、内含子、5' UTR、3' UTR、聚腺苷酸化信号和/或开放染色质区。在某些实施方式中，所述基因 (RS1、RL2和/或LAT的非编码序列) 被靶向以敲除所述基因，例如，以消除所述基因的表达，例如，以敲除所述RS1、RL2和/或LAT基因的一个或多个拷贝，例如，通过引入改变，所述改变包括所述RS1、RL2和/或LAT基因中的突变(例如，插入或缺失)。在某些实施方式中，所述方法、基因组编辑系统和组合物提供改变，所述改变包括RS1、RL2和/或LAT基因中的插入或缺失。

[0028] 在某些实施方式中，改变(例如，敲除或敲低)所述RS1基因指：(1)减少或消除RS1基因表达，(2)干扰转录调节因子ICP4蛋白质的活性和/或功能，其由所述RS1基因编码或(3)减少或消除转录调节因子ICP4蛋白的细胞内、血清和/或脑实质内水平。

[0029] 在某些实施方式中，改变(例如，敲除或敲低)所述RL2基因指：(1)减少或消除RL2基因表达，(2)干扰ICP0蛋白质的活性和/或功能，所述ICP0蛋白质由所述RL2基因编码和/或(3)减少或消除ICP0蛋白的细胞内、血清和/或脑实质内水平。

[0030] 在某些实施方式中，改变(例如，敲除或敲低)所述LAT基因指：(1)减少或消除LAT基因表达，(2)干扰由所述LAT基因编码的蛋白质的活性和/或功能，和/或(3)减少或消除由所述LAT基因编码的蛋白质的细胞内、血清和/或脑实质内水平。

[0031] 在某些实施方式中，本披露所述的方法、基因组编辑系统和组合物提供一种改变，所述改变包括通过插入或缺失一个或多个核苷酸来破坏所述RS1、RL2和/或LAT基因，所述插入或缺失一个或多个核苷酸由如下所述的Cas9分子(例如，酶活性Cas9 (eaCas9)，例如，Cas9核酸酶或Cas9切口酶)或Cas9融合蛋白介导。该类型的改变也称作“敲除”所述RS1、RL2和/或LAT基因。在某些实施方式中，敲除所述RS1、RL2和/或LAT基因包括敲除所述RS1、RL2和/或LAT基因的一个或多个拷贝，例如，通过引入改变，所述改变包括所述RS1、RL2和/或LAT基因中的突变(例如，插入或缺失)。在某些实施方式中，所述改变包括所述RS1、RL2和/或LAT基因中的插入或缺失。在某些实施方式中，靶向敲除方法由非同源末端连接(NHEJ)介导，例如，使用CRISPR/Cas系统，所述系统包括Cas9分子(例如，eaCas9分子)或Cas9-融合蛋白。在某些实施方式中，Cas9分子或Cas9-融合蛋白是Cas9变体，例如，化脓链球菌Cas9变体或金黄色葡萄球菌Cas9变体。在某些实施方式中，所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中，所述化脓链球菌Cas9变体是VRER变体。在某些实施方式中，靶向敲除方法减少或消除了功能性RL2基因产物的表达。在某些实施方式中，靶向敲除方法减少或消除了功能性LAT基因产物的表达。在某些实施方式中，靶向敲除方法减少或消除了功能性RS1基因产物的表达。

[0032] 在某些实施方式中，本披露所述的方法、基因组编辑系统和组合物提供对所述RS1、RL2和/或LAT基因的表达的改变，所述改变不包括所述RS1、RL2和/或LAT基因中的核苷酸插入或缺失。该类型的改变也称作“敲低”所述RS1、RL2和/或LAT基因的表达。在某些实施

方式中,该方法引起所述RS1、RL2和/或LAT基因的表达的减少、降低、抑制或消除。在某些实施方式中,靶向敲低方法由CRISPR/Cas系统介导,以改变(例如,阻断、减少或降低转录)RS1、RL2和/或LAT基因的转录,所述CRISPR/Cas系统包括Cas9分子(例如,无酶促活性Cas9(eiCas9)分子)或Cas9-融合蛋白(例如,eiCas9融合蛋白(例如,与转录抑制子结构域或染色质修饰蛋白融合的eiCas9))。在某些实施方式中,所述RS1、RL2和/或LAT基因的非编码区(例如,增强子区、启动子区、5' UTR、3' UTR、聚腺苷酸化信号和/或开放染色质区)被靶向以改变所述RS1、RL2和/或LAT基因的表达。在某些实施方式中,所述RL2基因的开放染色质区被靶向以改变所述RL2基因的表达。在某些实施方式中,转录调节区(例如,启动子区(例如,控制所述RS1、RL2和/或LAT基因的转录的启动子区))被靶向以改变(例如,敲低)所述RS1、RL2和/或LAT基因的表达。在某些实施方式中,一个或多个gRNA分子包括靶向结构域,所述靶向结构域被构造成靶向足够接近所述转录调节区(例如,启动子区(例如,控制所述RS1、RL2和/或LAT基因的转录的启动子区))的eiCas9分子或eiCas9融合蛋白,以减少、降低或抑制所述RS1、RL2和/或LAT基因的表达。在某些实施方式中,所述RL2基因的编码区被靶向以改变(例如,敲低)所述RL2基因的表达。在某些实施方式中,所述RS1基因的编码区被靶向以改变(例如,敲低)所述RS1基因的表达。在某些实施方式中,所述LAT基因的编码区被靶向以改变(例如,敲低)所述LAT基因的表达。在某些实施方式中,所述eiCas9分子是Cas9变体,例如,化脓链球菌Cas9变体或金黄色葡萄球菌Cas9变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。在某些实施方式中,靶向敲低方法减少或消除了功能性RL2基因产物的表达。在某些实施方式中,靶向敲低方法减少或消除了功能性LAT基因产物的表达。在某些实施方式中,靶向敲低方法减少或消除了功能性RS1基因产物的表达。

[0033] 在某些实施方式中,敲低所述RS1、RL2和/或LAT基因可治愈HSV感染。在某些实施方式中,敲低所述RS1、RL2和/或LAT基因提供HSV感染的功能性治愈。在某些实施方式中,敲低所述RS1、RL2和/或LAT基因导致对HSV感染的持续性病毒学反应。

[0034] 在某些实施方式中,所述RS1、RL2和/或LAT基因中已知被整合到所述受试者基因组中的区域被靶向用于敲低。在某些实施方式中,所述RS1、RL2和/或LAT基因中已知未被整合到所述受试者基因组中的区域被靶向用于敲除。在某些实施方式中,所述方法包括敲除所述RS1、RL2和/或LAT基因中未被整合进所述受试者基因组中的区域。

[0035] 敲除、敲低以及同时敲除和敲低所述RS1、RL2和/或LAT基因可降低HSV感染、复制和包装,并因此可治疗、预防和/或减少HSV感染。敲除、敲低以及同时敲除和敲低所述RS1、RL2和/或LAT基因的表达可单独或组合地引起以下情况中的任何一种:减少HSV DNA的产生、降低病毒的感染力、减少病毒颗粒的包装、减少病毒的脱落和/或减少由所述RS1、RL2和/或LAT基因编码的病毒蛋白(例如,ICP0和/或ICP4蛋白)的产生。在某些实施方式中,所述方法包括同时1)敲除和2)敲低所述RS1、RL2和/或LAT基因的两个不同区域,例如,1)敲低所述RS1、RL2和/或LAT基因中被整合进所述受试者基因组中的区域以及2)敲除所述RS1、RL2和/或LAT基因中未被整合进所述受试者基因组中的不同区域。

[0036] 本发明披露的主题提供一种基因组编辑系统,所述基因组编辑系统包括:gRNA分子和Cas9分子,所述gRNA分子包括靶向结构域,所述靶向结构域与单纯疱疹病毒(HSV)病毒基因的靶序列互补,所述HSV病毒基因的靶序列选自由RS1基因、RL2基因、和LAT基因组成的

群组。在某些实施方式中，所述靶向结构域被构造成在约500bp、450bp、400bp、350bp、300bp、250bp、200bp、150bp、100bp、50bp、25bp或10bp的HSV靶位置中形成双链断裂或单链断裂，从而改变所述HSV病毒基因。在某些实施方式中，改变所述HSV病毒基因包括敲除所述HSV病毒基因、敲低所述HSV病毒基因或同时敲除和敲低所述HSV病毒基因。

[0037] 在某些实施方式中，所述靶向结构域被构造成靶向所述HSV病毒基因的编码区或非编码区，其中所述非编码区包括所述HSV病毒基因的启动子区、增强子区、内含子、3' UTR、5' UTR或多聚腺苷酸化信号区；且所述编码区包括所述HSV病毒基因的早期编码区。

[0038] 在某些实施方式中，所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:208至58749的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0039] 在某些实施方式中，所述Cas9分子是化脓链球菌Cas9分子，所述基因组编辑系统敲除所述HSV-1 RS1基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:208至2509组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0040] 在某些实施方式中，所述Cas9分子是化脓链球菌Cas9分子，所述基因组编辑系统敲除所述HSV-2 RS1基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:7098至9292组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0041] 在某些实施方式中，所述Cas9分子是化脓链球菌Cas9分子，所述基因组编辑系统敲除所述HSV-1 RL2基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:21324至22744组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0042] 在某些实施方式中，所述Cas9分子是化脓链球菌Cas9分子，所述基因组编辑系统敲除所述HSV-2 RL2基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:26613至28037组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0043] 在某些实施方式中，所述Cas9分子是化脓链球菌Cas9分子，所述基因组编辑系统敲除所述HSV-1LAT基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:31730至32746组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0044] 在某些实施方式中，所述Cas9分子是化脓链球菌Cas9分子，所述基因组编辑系统敲除所述HSV-2LAT基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:35617至36926组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0045] 在某些实施方式中，所述Cas9分子是金黄色葡萄球菌Cas9分子，所述基因组编辑系统敲除所述HSV-1 RS1基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:2510至7073组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0046] 在某些实施方式中，所述Cas9分子是金黄色葡萄球菌Cas9分子，所述基因组编辑系统敲除所述HSV-2 RS1基因，且所述靶向结构域包括核苷酸序列，所述核苷酸序列与选自由SEQ ID NOS:9293至13614组成的群组中的核苷酸序列相同，或具有不超过3个核苷酸的不同。

[0047] 在某些实施方式中，所述Cas9分子是金黄色葡萄球菌Cas9分子，所述基因组编辑

系统敲除所述HSV-1 RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:22745至26601组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0048] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2 RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:28038至31720组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0049] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-1 LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:32747至35600组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0050] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲除所述HSV-2 LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:36927至40871组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0051] 在某些实施方式中,所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1 RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:13637至14794组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0052] 在某些实施方式中,所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2 RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:17753至18784组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0053] 在某些实施方式中,所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1 RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:40886至42078组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0054] 在某些实施方式中,所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2 RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:49498至50652组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0055] 在某些实施方式中,所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1 LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:45340至46479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0056] 在某些实施方式中,所述Cas9分子是化脓链球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2 LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:53858至55056组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0057] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1 RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自

由SEQ ID NOS:14795至17741组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0058] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2 RS1基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:18785至21311组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0059] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1 RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:42079至45315组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0060] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2 RL2基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:50653至53824组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0061] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-1LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:46480至49479组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0062] 在某些实施方式中,所述Cas9分子是金黄色葡萄球菌Cas9分子,所述基因组编辑系统敲低所述HSV-2LAT基因,且所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自由SEQ ID NOS:55057至58731组成的群组中的核苷酸序列相同,或具有不超过3个核苷酸的不同。

[0063] 在某些实施方式中,所述化脓链球菌Cas9分子识别前间区序列邻近基序(Protospacer Adjacent Motif, PAM) NGG。

[0064] 在某些实施方式中,所述金黄色葡萄球菌Cas9分子识别NNGRRT (SEQ ID NO:204) 和NNGRRV (SEQ ID NO:205) 中任一个PAM。

[0065] 本发明披露的主题提供一种gRNA分子,例如,分离的或非天然存在的gRNA分子,所述gRNA分子包括靶向结构域,所述靶向结构域与RS1、RL2或LAT基因的靶向结构域(也称作“靶序列”)互补。本发明披露的主题提供一种包括该gRNA分子的组合物。此外,本发明披露的主题提供一种包括该gRNA分子的载体。另外,本发明披露的主题提供包括本发明披露的基因组编辑系统、载体或组合物的细胞。在某些实施方式中,所述细胞选自由上皮细胞、神经细胞和视觉细胞组成的群组。

[0066] 在某些实施方式中,所述gRNA分子的靶向结构域被构造成提供足够接近HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的切割事件(例如,双链断裂或单链断裂),以允许HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变(例如,与NHEJ相关的改变)。在某些实施方式中,所述靶向结构域被构造成使切割事件(例如,双链断裂或单链断裂)位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。所述断裂(例如,双链或单链断裂)可位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游或下游。在某些实施方式中,所述gRNA分子的靶向结构域被构造

成在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的500个(500、400、300、250、200、150、100、80、60、40、20或10个内)个核苷酸内提供选自双链断裂和单链断裂的切割事件。

[0067] 在某些实施方式中,第二gRNA分子包括第二靶向结构域,所述第二靶向结构域被构造成提供足够接近HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的切割事件(例如,双链断裂或单链断裂),以单独地或与由所述第一gRNA分子定位的断裂组合地允许HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变(例如,与NHEJ相关的改变)。在某些实施方式中,所述第一和第二gRNA分子的靶向结构域被构造成使得切割事件(例如,双链断裂或单链断裂)对于每个所述gRNA分子独立地位于所述靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。在某些实施方式中,所述断裂(例如,双链或单链断裂)位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的两侧。在某些实施方式中,所述断裂(例如,双链或单链断裂)位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的一侧,例如,上游或下游。在某些实施方式中,所述第一和/或第二gRNA分子的靶向结构域被构造成在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的500个(500、400、300、250、200、150、100、80、60、40、20或10个内)个核苷酸内提供选自双链断裂和单链断裂的切割事件。

[0068] 在某些实施方式中,单链断裂伴随着另一个单链断裂,所述另一个单链断裂由第二gRNA分子定位,如下所述。例如,所述靶向结构域被构造成使切割事件(例如,两个单链断裂)位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。在某些实施方式中,所述第一和第二gRNA分子被构造成彼此足够接近,在引导Cas9分子或Cas9-融合蛋白(例如,Cas9切口酶)时,单链断裂伴随另一个单链断裂(由第二gRNA分子定位),以产生对HSV RS1、HSV RL2靶标或HSV LAT靶标的改变。在某些实施方式中,所述第一和第二gRNA分子被构造成使得例如在所述Cas9分子或Cas9-融合蛋白是切口酶时,由所述第二gRNA定位的单链断裂在由所述第一gRNA分子定位的断裂的10、20、30、40或50个核苷酸内。在某些实施方式中,所述两个gRNA分子被构造成在不同链上的相同位置或彼此的几个核苷酸内进行切割,例如,基本模拟双链断裂。

[0069] 在某些实施方式中,双链断裂可伴随着另一个双链断裂,所述另一个双链断裂由第二gRNA分子定位,如下所述。例如,第一gRNA分子的靶向结构域被构造成使双链断裂位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游,例如,位于所述靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内,而第二gRNA分子的靶向结构域被构造成使双链断裂位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的下游,例如,位于所述靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。

[0070] 在某些实施方式中,双链断裂可伴随着两个其他单链断裂,所述两个其他单链断裂由第二和第三gRNA分子定位。例如,第一gRNA分子的靶向结构域被构造成使双链断裂位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游,例如,位于所述靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内,而第二和第三gRNA分子的靶向结构域被构造成使两个单链断裂位于HSV RS1靶标或HSV RL2靶标或HSV LAT靶标的下游,例如,位于所述靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。在某些实施方式中,所述第一、第二和第三gRNA分子的靶向结构

域被构造成使得切割事件(例如,双链断裂或单链断裂)对于每个所述gRNA分子是独立的。

[0071] 在某些实施方式中,第一和第二单链断裂可伴随着两个其他单链断裂,所述两个其他单链断裂由第三gRNA分子和第四gRNA分子定位。例如,第一和第二gRNA分子的靶向结构域被构造成使两个单链断裂位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游,例如,位于所述靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内,而第三和第四gRNA分子的靶向结构域被构造成使两个单链断裂位于HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的下游,例如,位于所述靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。在某些实施方式中,所述第一、第二、第三和/或第四gRNA分子的靶向结构域被构造成在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的500个(500、400、300、250、200、150、100、80、60、40、20或10个内)个核苷酸内提供选自双链断裂和单链断裂的切割事件。

[0072] 在某些实施方式中,当使用多个gRNA产生(1)紧邻的两个单链断裂,(2)两个双链断裂,例如,双侧接HSV RS1靶标、HSV RL2靶标或HSV LAT靶标(例如,以移除一段DNA、以创建缺失突变),或在所述基因中创建不止一个indel(例如,在编码区、早期编码区中),(3)一个双链断裂和两个配对缺口,双侧接HSV RS1靶标、HSV RL2靶标或HSV LAT靶标(例如,以移除一段DNA、以插入缺失)或(4)四个单链断裂(两个位于位置的每一侧),它们靶向相同的HSV RS1靶标、HSV RL2靶标或HSV LAT靶标。在某些实施方式中,多个gRNA可用于靶向不止一个HSV RS1靶标、HSV RL2靶标或HSV LAT靶标。

[0073] 在某些实施方式中,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所述靶核酸分子的相对链互补。在某些实施方式中,所述gRNA分子和所述第二gRNA分子被构造成使所述PAM向外朝向。

[0074] 在某些实施方式中,gRNA分子的靶向结构域被构造成避免不需要的靶染色体元素,所述不需要的靶染色体元素包括但不限于所述靶向结构域中的重复元素,例如Alu重复。如本发明所述,gRNA分子可以是第一、第二、第三和/或第四gRNA分子。

[0075] 在某些实施方式中,gRNA分子的靶向结构域被构造成将切割事件定位在足够远离预选核苷酸(例如,编码区的核苷酸)的位置,从而可不改变所述核苷酸。在某些实施方式中,gRNA分子的靶向结构域被构造成将内含子切割事件定位在离内含子/外显子边界或天然存在的剪接信号足够远的位置,以避免改变外显子序列或不需要的剪接事件。如本发明所述,gRNA分子可以是第一、第二、第三和/或第四gRNA分子。

[0076] 在某些实施方式中,靶向HSV1 RS1靶向敲除位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:208至2509和2510至7073中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0077] 在某些实施方式中,靶向HSV2 RS1靶向敲除位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:7098至9292和9293至13614中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0078] 在某些实施方式中,靶向HSV1 RL2靶向敲除位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:21324至22744和22745至26601中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0079] 在某些实施方式中,靶向HSV2 RL2靶向敲除位置的gRNA分子中的靶向结构域包括

核苷酸序列,所述核苷酸序列与选自SEQ ID NO:26613至28037和28038至31720中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0080] 在某些实施方式中,靶向HSV1 LAT靶向敲除位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:31730至32746和32747至35600中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0081] 在某些实施方式中,靶向HSV2 LAT靶向敲除位置的gRNA分子中的所述靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:35617至36926和36927至40871中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0082] 在某些实施方式中,靶向HSV1 RS1靶向敲低位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:13637至14794和14795至17741中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0083] 在某些实施方式中,靶向HSV2 RS1靶向敲低位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:17753至18784和18785至21311中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0084] 在某些实施方式中,靶向HSV1 RL2靶向敲低位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:40886至42078和42079至45315中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0085] 在某些实施方式中,靶向HSV2 RL2靶向敲低位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:49498至50652和50653至53824中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0086] 在某些实施方式中,靶向HSV1 LAT靶向敲低位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:45340至46479和46480至49479中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0087] 在某些实施方式中,靶向HSV2 LAT靶向敲低位置的gRNA分子中的靶向结构域包括核苷酸序列,所述核苷酸序列与选自SEQ ID NO:53858至55056和55057至58731中的核苷酸序列相同,或存在不超过1个、2个、3个、4个或5个核苷酸的不同。

[0088] 在某些实施方式中,所述gRNA分子是单分子或嵌合gRNA分子。

[0089] 在某些实施方式中,本发明披露的gRNA分子的靶向结构域长度为16、17、18、19、20、21、22、23、24、25或26个核苷酸。

[0090] 在某些实施方式中,所述gRNA分子包括从5'到3'的:靶向结构域(包括“核心结构域”,和可选地包括“次级结构域”);第一互补结构域;连接结构域;第二互补结构域;和近端结构域。在某些实施方式中,所述gRNA分子还包括尾部结构域。在某些实施方式中,所述近端结构域和尾部结构域一起作为单个结构域。

[0091] 在某些实施方式中,gRNA分子包括长度不超过25个核苷酸的连接结构域、一起长度至少为20、30、35或40个核苷酸的近端和尾部结构域以及长度等于或大于16、17、18、19、20、21、22、23、24、25或26个核苷酸的靶向结构域。

[0092] 切割事件(例如,双链或单链断裂)可通过Cas9分子或Cas9-融合蛋白来产生。所述Cas9分子或Cas9-融合蛋白可以是酶活性的Cas9(eaCas9)分子,例如,在靶核酸中形成双链断裂的eaCas9分子或在靶核酸中形成单链断裂的eaCas9分子(例如,切口酶分子)。在某些

实施方式中,所述eaCas9分子可以是Cas9变体。例如,但不作为限制,所述Cas9变体可以是化脓链球菌Cas9变体或金黄色葡萄球菌Cas变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。除了Cas9分子或Cas9-融合蛋白外,还可使用本发明披露的其他核酸酶来产生切割事件。

[0093] 在某些实施方式中,所述eaCas9分子或eaCas9-融合蛋白催化双链断裂。

[0094] 在某些实施方式中,所述eaCas9分子或eaCas9-融合蛋白包括HNH样结构域切割活性,但不具有或不显著具有N-末端RuvC样结构域切割活性。在这种情况下,所述eaCas9分子或eaCas9-融合蛋白是HNH样结构域切口酶,例如,eaCas9分子或eaCas9-融合蛋白在D10处包括突变,例如,D10A。在某些实施方式中,所述eaCas9分子或eaCas9-融合蛋白包括N-末端RuvC样结构域切割活性,但不具有或不显著具有HNH样结构域切割活性。在某些实施方式中,所述eaCas9分子或eaCas9-融合蛋白是N-末端RuvC样结构域切口酶,例如,所述eaCas9分子在H840处包括突变,例如,H840A。在某些实施方式中,所述eaCas9分子或eaCas9-融合蛋白是N-末端RuvC样结构域切口酶,例如,所述eaCas9分子或eaCas9-融合蛋白在H863处包括突变,例如,H863A。

[0095] 在某些实施方式中,单链断裂形成在所述靶核酸中与所述gRNA分子的靶向结构域互补的链中。在某些实施方式中,单链断裂形成在所述靶核酸中不是与所述gRNA分子的靶向结构域互补的链中。

[0096] 此外,本发明披露的主题提供一种核酸组合物,例如,分离的或非天然存在的核酸组合物,例如,DNA,所述DNA包括(a)编码本发明披露的gRNA分子的第一核苷酸序列,例如,包括靶向结构域的gRNA分子,所述靶向结构域与RS1、RL2或LAT基因的靶序列互补,例如,HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处的靶序列。

[0097] 在某些实施方式中,所述核酸组合物还包括(b)编码本发明所述的Cas9分子或Cas9-融合蛋白的第二核苷酸序列。在某些实施方式中,所述Cas9分子是eiCas9分子。在某些实施方式中,所述Cas9分子是eaCas9分子。在某些实施方式中,所述Cas9分子(例如,eiCas9分子或eaCas9分子)可以是Cas9变体,例如,化脓链球菌Cas9变体或金黄色葡萄球菌Cas9变体。

[0098] 所述Cas9分子或Cas9-融合蛋白可以是切口酶分子、酶活化Cas9(eaCas9)分子或eaCas9-融合蛋白,例如,在靶核酸中形成双链断裂的eaCas9分子或eaCas9-融合蛋白和/或在靶核酸中形成单链断裂的eaCas9分子或eaCas9-融合蛋白。在某些实施方式中,单链断裂形成在所述靶核酸中与所述gRNA分子的靶向结构域互补的链中。在某些实施方式中,单链断裂形成在所述靶核酸中不是与所述gRNA分子的靶向结构域互补的链中。

[0099] 在某些实施方式中,所述核酸组合物还包括(c)(i)编码本发明所述的第二gRNA分子的第三核苷酸序列,所述第三核苷酸序列包括靶向结构域,所述靶向结构域与所述RS1、RL2或LAT基因的第二靶序列互补以及可选地(c)(ii)编码本发明所述的第三gRNA分子的序列,所述序列具有靶向结构域,所述靶向结构域与所述RS1、RL2或LAT基因的第三靶序列互补;以及可选地(c)(iii)编码本发明所述的第四gRNA分子的序列,所述序列具有靶向结构域,所述靶向结构域与所述RS1、RL2或LAT基因的第四靶序列互补。

[0100] 在某些实施方式中,(a)和(b)存在于相同的核酸分子上,例如,相同的载体上,例如,相同的病毒载体上,例如,相同的腺相关病毒(AAV)载体或慢病毒(LV)载体上。在某些实

施方式中,所述核酸分子是LV载体。在某些实施方式中,所述核酸分子是AAV载体。可用在所述组合物和方法中任一个中的示例性AAV载体包括AAV2载体、修饰的AAV2载体、AAV3载体、修饰的AAV3载体、AAV6载体、修饰的AAV6载体、AAV8载体和AAV9载体。在某些实施方式中,所述Cas9分子是eiCas9分子。在某些实施方式中,所述Cas9分子是eaCas9分子。在某些实施方式中,所述Cas9分子(例如,eiCas9分子或eaCas9分子)可以是Cas9变体,例如,化脓链球菌Cas9变体或金黄色葡萄球菌Cas9变体。

[0101] 在某些实施方式中,(a)存在于第一核酸分子上,例如,第一载体上,例如,第一病毒载体上,例如,第一AAV载体或第一LV载体上;以及(b)存在于第二核酸分子上,例如,第二载体上,例如,第二载体上,例如,第二AAV载体或第二LV载体上。所述第一和第二核酸分子可以是AAV载体。在某些实施方式中,所述第一和第二核酸分子可以是LV载体。

[0102] 在某些实施方式中,(a)和(c)(i)存在于一个核酸分子上,例如,一个载体上,例如,一个病毒载体上,例如,一个AAV载体或LV载体上。在某些实施方式中,所述核酸分子是AAV载体。在某些实施方式中,所述核酸分子是LV载体。在某些实施方式中,(a)和(c)(i)在不同的载体上。在某些实施方式中,(a)存在于第一核酸分子上,例如,第一载体上,例如,第一病毒载体上,例如,第一AAV载体或LV载体上;以及(c)(i)存在于第二核酸分子上,例如,第二载体上,例如,第二载体上,例如,第二AAV载体或第二LV载体上。在某些实施方式中,所述第一和第二核酸分子是AAV载体。在某些实施方式中,所述第一和第二核酸分子是LV载体。

[0103] 在某些实施方式中,(a)、(b)和(c)(i)中的每个存在于一个核酸分子上,例如,一个载体上,例如,一个病毒载体上,例如,AAV载体或LV载体上。在某些实施方式中,所述核酸分子是AAV载体。在某些实施方式中,所述核酸分子是LV载体。在某些实施方式中,(a)、(b)和(c)(i)中的一个在第一核酸分子上编码,例如,第一载体上,例如,第一病毒载体上,例如,第一AAV载体或LV载体上;以及(a)、(b)和(c)(i)中的第二和第三个在第二核酸分子上编码,例如,第二载体上,例如,第二载体上,例如,第二AAV载体或LV载体上。所述第一和第二核酸分子可以是AAV载体或LV载体。

[0104] 在某些实施方式中,(a)存在于第一核酸分子上,例如,第一载体上,例如,第一病毒载体上,例如,第一AAV载体或LV载体上;以及(b)和(c)(i)存在于第二核酸分子上,例如,第二载体上,例如,第二载体上,例如,第二AAV载体或LV载体上。所述第一和第二核酸分子可以是AAV载体或LV载体。

[0105] 在某些实施方式中,(b)存在于第一核酸分子上,例如,第一载体上,例如,第一病毒载体上,例如,第一AAV载体或LV载体上;以及(a)和(c)(i)存在于第二核酸分子上,例如,第二载体上,例如,第二载体上,例如,第二AAV载体或LV载体上。所述第一和第二核酸分子可以是AAV载体或LV载体。

[0106] 在某些实施方式中,(c)(i)存在于第一核酸分子上,例如,第一载体上,例如,第一病毒载体上,例如,第一AAV载体或LV载体上;以及(b)和(a)存在于第二核酸分子上,例如,第二载体上,例如,第二载体上,例如,第二AAV载体或LV载体上。所述第一和第二核酸分子可以是AAV载体。在某些实施方式中,所述第一和第二核酸分子可以是LV载体。

[0107] 在某些实施方式中,(a)、(b)和(c)(i)中的每个存在于不同的核酸分子上,例如,不同的载体上,例如,不同的病毒载体上,例如,不同的AAV载体或LV载体上。例如,(a)可在

第一核酸分子上, (b) 在第二核酸分子上, 以及 (c) (i) 在第三核酸分子上。所述第一、第二和第三核酸分子可以是AAV载体。在某些实施方式中, 所述第一、第二和第三核酸分子可以是LV载体。

[0108] 在某些实施方式中, 当存在第三和/或第四gRNA分子时, (a)、(b)、(c) (i)、(c) (ii) 和 (c) (iii) 中的每个可存在于一个核酸分子上, 例如, 一个载体上, 例如, 一个病毒载体上, 例如, AAV载体或LV载体上。在某些实施方式中, 所述核酸分子是AAV载体。在某些实施方式中, 所述核酸分子是LV载体。在某些实施方式中, (a)、(b)、(c) (i)、(c) (ii) 和 (c) (iii) 中的每个可存在于不同的核酸分子上, 例如, 不同的载体上, 例如, 不同的病毒载体上, 例如, 不同的AAV载体或不同的LV载体上。在其他实施方式中, (a)、(b)、(c) (i)、(c) (ii) 和 (c) (iii) 中的每个可存在于多于一个但少于五个核酸分子上, 例如, AAV载体或LV载体上。

[0109] 在某些实施方式中, 所述第二gRNA分子被构造成提供足够接近HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的切割事件(例如, 双链断裂或单链断裂), 以单独地或与由所述第一gRNA分子定位的断裂组合地导致HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变(例如, 与NHEJ相关的改变)。

[0110] 在某些实施方式中, 所述第三gRNA分子被构造成提供足够接近HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的切割事件(例如, 双链断裂或单链断裂), 以单独地或与由所述第一和/或第二gRNA分子定位的断裂组合地允许HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变(例如, 与NHEJ相关的改变)。

[0111] 在某些实施方式中, 所述第四gRNA分子被构造成提供足够接近HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的切割事件(例如, 双链断裂或单链断裂), 以单独地或与由所述第一gRNA分子、所述第二gRNA分子和/或所述第三gRNA分子定位的断裂组合地允许HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变(例如, 与NHEJ相关的改变)。

[0112] 在某些实施方式中, 选择所述第二gRNA与第一gRNA分子同样靶向相同的HSV RS1靶标、HSV RL2靶标或HSV LAT靶标。在某些实施方式中, 选择所述第三gRNA和所述第四gRNA分子与第一和第二gRNA分子同样靶向相同的HSV RS1靶标、HSV RL2靶标或HSV LAT靶标。

[0113] 所述第二、第三和第四gRNA分子的靶向结构域可单独选自SEQ ID NO:208至58749所示的核苷酸序列。所述第二、第三和第四gRNA分子可以是模块化gRNA分子或嵌合gRNA分子。

[0114] 本发明所述核酸组合物中存在的一个或多个核酸和/或核酸组合物可包括可操作连接至所述核苷酸序列的启动子, 所述核苷酸序列编码 (a) 的所述gRNA分子, 例如, 本发明所述的启动子。所述核酸和/或核酸组合物还可包括可操作连接至所述核苷酸序列的第二启动子, 所述核苷酸序列编码 (c) 的所述第二、第三和/或第四gRNA, 例如, 本发明所述的启动子。所述启动子和第二启动子彼此不同。在某些实施方式中, 所述启动子和第二启动子相同。

[0115] 本发明所述的核酸组合物还可包括可操作连接至所述核苷酸序列的启动子, 所述核苷酸序列编码 (b) 的所述Cas9分子或Cas9-融合蛋白, 例如, 本发明所述的启动子。

[0116] 本发明披露的主题还提供一种组合物, 所述组合物包括 (a) 本发明公开的gRNA分子, 例如, 包括靶向结构域的gRNA分子, 所述靶向结构域与RS1、RL2或LAT基因的靶序列互补。在某些实施方式中, 所述组合物还包括 (b) Cas9分子(例如, eaCas9分子或eiCas9分子)

或Cas9-融合蛋白(如本发明所述)。在某些实施方式中,所述Cas9分子(例如,eaCas9分子或eiCas9分子)可以是Cas9变体。例如,但不作为限制,所述Cas9变体可以是化脓链球菌Cas9变体或金黄色葡萄球菌Cas变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。在某些实施方式中,所述组合物还包括(c)第二、第三和/或第四gRNA分子,例如,本发明所述的第二、第三和/或第四gRNA分子。在某些实施方式中,所述组合物是药物组合物。例如根据本发明公开的方法,本发明所述的组合物(例如,本发明所述的药物组合物)可用于治疗、预防和/或减轻受试者中的HSV-1或HSV-2感染。

[0117] 本发明披露的主题还提供一种改变HSV病毒基因的方法,所述HSV病毒基因选自由细胞中的RS1基因、RL2基因和LAT基因组成的群组,所述方法包括给予细胞下组中的一个:

[0118] (i)基因编辑系统,所述基因编辑系统包括gRNA分子和至少一个Cas9分子,所述gRNA分子包括靶向结构域,所述靶向结构域与所述HSV病毒基因的靶序列互补;

[0119] (ii)载体,所述载体包括编码gRNA分子的多核苷酸和编码Cas9分子的多核苷酸,所述gRNA分子包括靶向结构域,所述靶向结构域与所述HSV病毒基因的靶序列互补;或

[0120] (iii)组合物,所述组合物包括gRNA分子和至少一个Cas9分子,所述gRNA分子包括靶向结构域,所述靶向结构域与所述HSV病毒基因的靶序列互补。

[0121] 在另一个方面,本发明披露一种改变细胞中的RS1、RL2或LAT基因的方法,例如,改变细胞的靶核酸的结构,例如,改变细胞的靶核酸的序列,所述方法包括将所述细胞与以下接触:(a)本发明公开的gRNA分子,以及(b)Cas9分子(例如eaCas9分子)或Cas9-融合蛋白,例如,本发明所述的Cas9分子;以及可选地,(c)靶向RS1、RL2和LAT基因的第二、第三和/或第四gRNA分子,例如,本发明所述的第二、第三和/或第四gRNA分子。在某些实施方式中,所述Cas9分子可以是Cas9变体。

[0122] 在某些实施方式中,所述方法包括接触来自患有或可能发展成HSV-1和/或HSV-2的受试者的细胞。所述细胞可来自受益于在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处具有突变的受试者。

[0123] 在某些实施方式中,所述接触步骤可在体内进行。

[0124] 在某些实施方式中,本发明所述改变细胞的方法包括在所述接触步骤之前,获知所述细胞中HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的序列。获知所述细胞中HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的序列可通过对所述RS1、RL2或LAT基因中的一个或多个或所述RS1、RL2或LAT基因的一部分进行测序。

[0125] 在某些实施方式中,所述方法的接触步骤包括将所述细胞与核酸组合物(例如,表达(a)、(b)和(c)中的至少一个的载体,例如,AAV载体或LV载体)接触。在某些实施方式中,所述方法的接触步骤包括将所述细胞与核酸组合物(例如,表达(a)、(b)和(c)中的每个的载体,例如,AAV载体或LV载体)接触。在某些实施方式中,所述方法的接触步骤包括向所述细胞递送Cas9分子或(b)的Cas9-融合蛋白,以及核酸组合物,所述核酸组合物编码(a)的gRNA分子,和可选地,编码(c)(i)的第二gRNA分子,以及再可选地,编码(c)(ii)的第三gRNA分子和/或编码(c)(iii)的第四gRNA分子。

[0126] 在某些实施方式中,所述接触步骤包括将所述细胞与核酸组合物(例如,AAV载体或LV载体)接触,其表达(a)、(b)、(c)和(d)中的至少一个。在某些实施方式中,所述方法的

接触步骤包括将所述细胞与核酸组合物(例如,载体或AAV载体)接触,其表达(a)、(b)和(c)中的每个。在某些实施方式中,所述方法的接触步骤包括向所述细胞递送(b)的Cas9分子或Cas9-融合蛋白以及核酸组合物,所述核酸组合物编码(a)的gRNA分子和(d)的模板核酸,和可选地,编码(c)(i)的第二gRNA分子,以及再可选地,编码(c)(iv)的第三gRNA分子和/或(c)(iii)的第四gRNA分子。

[0127] 在某些实施方式中,所述接触步骤包括将所述细胞与核酸组合物接触,所述核算组合物是本发明所述的,例如,载体,例如,AAV载体,例如,AAV2载体、修饰的AAV2载体、AAV3载体、修饰的AAV3载体、AAV6载体、修饰的AAV6载体、AAV8载体或AAV9载体。在某些实施方式中,所述载体是LV载体。

[0128] 在某些实施方式中,接触包括向所述细胞递送作为蛋白质或mRNA的(b)的Cas9分子或Cas9-融合蛋白,以及核酸组合物,所述核酸组合物编码(a)的gRNA分子以及可选地(c)的第二、第三和/或第四gRNA分子。

[0129] 在某些实施方式中,所述接触步骤包括向所述细胞递送作为蛋白质或mRNA的(b)的Cas9分子或Cas9-融合蛋白,作为RNA的(a)的所述gRNA分子以及可选地作为RNA的(c)的第二、第三和/或第四gRNA分子。

[0130] 在某些实施方式中,接触包括向所述细胞递送作为RNA的(a)的gRNA分子,可选地作为RNA的(c)的第二、第三和/或第四gRNA分子,以及编码所述(b)的Cas9分子或Cas9-融合蛋白的核酸组合物。

[0131] 本发明披露的主题还提供一种治疗、预防和/或减轻患有或可能发展HSV-1和/或HSV-2的受试者的方法,例如通过改变所述受试者的靶核酸的结构(例如序列),所述方法包括将所述受试者(或来自所述受试者的细胞)与以下接触:

[0132] (a) 本发明披露的gRNA分子,例如,靶向RS1、RL2或LAT基因的gRNA分子;

[0133] (b) Cas9分子(例如,eaCas9分子或eiCas9分子)或Cas9-融合蛋白,例如,本发明所述的Cas9分子;以及

[0134] 可选地,(c)(i)靶向RS1、RL2或LAT基因的第二gRNA分子,例如,本发明披露的第二gRNA分子,以及

[0135] 再可选地,(c)(ii)第三gRNA分子,并且仍然可选地,(c)(iii)靶向RL2或LAT或RS1基因的第四gRNA,例如,本发明披露的第三和第四gRNA分子。

[0136] 在某些实施方式中,所述方法包括在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处引入突变,例如,通过NHEJ。

[0137] 在某些实施方式中,在体内将所述受试者的细胞与(例如,通过静脉内递送)、(a)、(b)和可选地(c)(i)、进一步可选地与(c)(ii)、和仍进一步可选地(c)(iii)接触。

[0138] 在某些实施方式中,所述接触步骤包括将所述受试者与核酸组合物(例如,本发明所述的载体、AAV载体或LV载体)接触,例如,编码(a)、(b)和可选地(c)(i)、进一步可选地(c)(ii)以及仍进一步可选地(c)(iii)中的至少一个的核酸组合物。

[0139] 在某些实施方式中,所述接触步骤包括向所述受试者递送作为蛋白质或mRNA的(b)的Cas9分子或Cas9-融合蛋白,以及核酸组合物,所述核酸组合物编码(a)和可选地(c)(i)、进一步可选地(c)(ii)、以及仍进一步可选地(c)(iii)。

[0140] 在某些实施方式中,所述接触步骤包括向所述受试者递送作为蛋白质或mRNA的

(b) 的Cas9分子或Cas9-融合蛋白、作为RNA的 (a) 的gRNA分子、可选地作为gRNA的 (c) (i) 的第二gRNA、进一步可选地 (c) (ii) 以及仍进一步可选地 (c) (iii) 的gRNA分子。

[0141] 在某些实施方式中,所述接触步骤包括向所述受试者递送作为RNA的 (a) 的gRNA分子、可选地作为RNA的 (c) (i) 的第二gRNA、进一步可选地 (c) (ii) 以及仍进一步可选地 (c) (iii),和编码所述 (b) 的Cas9分子或Cas9-融合蛋白的核酸组合物。

[0142] 当所述方法包括: (1) 通过NHEJ在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处引入突变或 (2) 敲低所述RS1、RL2和/或LAT基因中的一个或多个的表达(例如,通过靶向所述启动子区)时, (b) 的Cas9分子或Cas9-融合蛋白和至少一个gRNA分子(例如, (a) 的gRNA分子)包含在所述接触步骤中。

[0143] 本发明披露的主题提供了反应混合物,所述反应混合物包括gRNA分子、核酸组合物或本发明所述的组合物以及细胞(例如,来自具有或可能发展HSV-1和/或HSV-2的受试者的细胞,或来自将从HSV RL2靶标或HSV LAT靶标或HSV RS1靶标处的突变受益的受试者的细胞)。

[0144] 本发明披露的主题提供试剂盒,所述试剂盒包括: (a) 本发明所述的gRNA分子或编码所述gRNA分子的核酸组合物,以及以下中的一种或多种:

[0145] (b) Cas9分子(例如,eaCas9分子或eiCas9分子)或Cas9-融合蛋白,例如,本发明所述的Cas9分子或编码所述Cas9的核酸或mRNA;

[0146] (c) (i) 第二gRNA分子,例如,本发明所述的第二gRNA分子或编码 (c) (i) 的核酸;

[0147] (c) (ii) 第三gRNA分子,例如,本发明所述的第三gRNA分子或编码 (c) (ii) 的核酸;或

[0148] (c) (iii) 第四gRNA分子,例如,本发明所述的第四gRNA分子或编码 (c) (iii) 的核酸。在某些实施方式中,所述Cas9分子可以是Cas9变体。例如,但不作为限制,所述Cas9变体可以是化脓链球菌Cas9变体或金黄色葡萄球菌Cas变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。

[0149] 在某些实施方式中,所述试剂盒包括编码 (a) 、(b) 、(c) (i) 、(c) (ii) 和 (c) (iii) 中一个或多个的核酸,例如,AAV载体或LV载体。

[0150] 本发明披露的主题提供一种用于治疗,预防、减轻或延迟受试者中HSV-1和/或HSV-2感染的发作或进展的gRNA分子(例如本发明所述的gRNA分子),例如,根据本文所述的治疗、预防、减轻或延迟本发明所述HSV-1和/或HSV-2感染的发作或进展的方法。

[0151] 在某些实施方式中,所述gRNA分子与Cas9分子(例如,eaCas9分子或eiCas9分子)或Cas9-融合蛋白(例如,本发明所述的Cas9分子)组合使用。例如,但不是作为限制,所述Cas9分子或Cas9-融合蛋白是Cas9变体。例如,但不作为限制,所述Cas9变体可以是化脓链球菌Cas9变体或金黄色葡萄球菌Cas变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。另外或替代地,在某些实施方式中,所述gRNA分子与第二、第三和/或第四gRNA分子组合使用,例如,本发明所述的第二、第三和/或第四gRNA分子。

[0152] 本发明披露的主题提供在制备用于治疗,预防、减轻或延迟受试者中HSV-1和/或HSV-2的发作或进展的gRNA分子(例如本发明所述的gRNA分子)的药物中的用途,例如,根据本文所述的治疗、预防、减轻或延迟本发明所述HSV-1和/或HSV-2的发作或进展的方法。

[0153] 在某些实施方式中,所述药物包括Cas9分子(例如,eaCas9分子或eiCas9分子)或Cas9-融合蛋白(例如,本发明所述的Cas9分子)。另外或替代地,在某些实施方式中,所述药物包括第二、第三和/或第四gRNA分子,例如,本发明所述的第二、第三和/或第四gRNA分子。在某些实施方式中,所述Cas9分子可以是Cas9变体。例如,但不作为限制,所述Cas9变体可以是化脓链球菌Cas9变体或金黄色葡萄球菌Cas变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。

[0154] 本发明披露的gRNA分子、基因组编辑系统、方法、组合物、反应混合物和试剂盒还可包括控制性gRNA分子,例如,本发明披露的控制性gRNA分子。

[0155] 除非另有明确定义,本发明使用的所有技术性和科技性术语与本发明所述领域的本领域技术人员所理解的意义相同。尽管与本发明所述的方法和材料相似或等同的方法和材料可以用于本发明的实践和测试,但是以下描述了合适的方法和材料。本发明提及的所有出版物、专利申请、专利和其他引用文献均通过整体并入。此外,材料、方法和实施例仅仅为示例性的,并非意图进行限制。

[0156] 标题,包括数字和字母标题以及副标题,用于组织和演示,而不是限制性的。

[0157] 本发明的其他特征和优点从发明详述、附图和权利要求中将是显而易见的。

[0158] 附图简述

[0159] 图1A-1I是若干示例性gRNA的表示。图1A描绘了部分来源于(或部分地在序列上建模)化脓链球菌呈双链体结构的模块化gRNA分子(按照出现次序分别是SEQ ID NO:39和40);图1B描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:41);图1C描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:42);图1D描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:43);图1E描绘了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:44);图1F描绘了部分来源于嗜热链球菌呈双链体结构的模块化gRNA分子(按照出现次序分别是SEQ ID NO:45和46);图1G描绘了化脓链球菌和嗜热链球菌的模块化gRNA分子(按照出现次序分别是SEQ ID NO:39、45、47和46)的比对。图1H-1I描绘了单分子gRNA分子的另外示例性结构。图1H示出了部分来源于化脓链球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:42)的示例性结构。图1I示出了部分来源于金黄色葡萄球菌呈双链体结构的单分子gRNA分子(SEQ ID NO:38)的示例性结构。

[0160] 图2A-2G描绘了Cas9序列的比对(Chylinski 2013)。N-末端RuvC样结构域加框并且以“Y”指示。其他两个RuvC样结构域加框并且以“B”指示。HNH样结构域加框并且以“G”指示。Sm:变形链球菌(SEQ ID NO:1);Sp:化脓链球菌(SEQ ID NO:2);St:嗜热链球菌(SEQ ID NO:4);以及Li:无害利斯特菌(SEQ ID NO:5)。“基序”(SEQ ID NO:14)是基于四个序列的共有序列。所有四个序列中的保守残基由单字母氨基酸缩写表示;“\*”指示在这四个序列的任一者的相应位置中发现的任何氨基酸;并且“-”指示不存在。

[0161] 图3A-3B示出了来自披露于Chylinski 2013中的Cas9分子(SEQ ID NO:52-95、120-123)的N-末端RuvC样结构域的比对。图3B的最后一行鉴定了4个高度保守的残基。

[0162] 图4A-4B示出了来自披露于Chylinski 2013中的除去序列异常值的Cas9分子(SEQ ID NO:52-123)的N-末端RuvC样结构域的比对。图4B的最后一行鉴定了3个高度保守的残基。

[0163] 图5A-5C示出了来自披露于Chylinski 2013中的Cas9分子 (SEQ ID NO:124-198) 的HNH样结构域的比对。图5C的最后一行鉴定了保守残基。

[0164] 图6A-6B示出了来自披露于Chylinski 2013中的除去序列异常值的Cas9分子 (SEQ ID NO:124-141、148、149、151-153、162、163、166-174、177-187、194-198) 的HNH样结构域的比对。图6B的最后一行鉴定了3个高度保守的残基。

[0165] 图7示出了使用示例性gRNA序列 (SEQ ID NO:42) 的gRNA结构域命名法。

[0166] 图8A和8B提供了化脓链球菌Cas9的结构域组织的示意性表示。图8A参照Cas9的两种叶片 (识别 (REC) 叶片和核酸酶 (NUC) 叶片) 示出了Cas9结构域的组织, 包括氨基酸位置。图8B示出了83个Cas9直向同源物中每个结构域的百分比同源性。

[0167] 图9是pAF025质粒图谱的示意图。

[0168] 图10A-10B示出了Cas9介导的质粒pAF025中HSV-1靶序列的切割。(A) 示出了通过表18 (参见实施例3) 中列出的gRNA来进行HSV1的RL2/LAT的靶向敲低。(B) 示出了通过表19 (参见实施例3) 中列出的的HSV1的RS1靶向敲低。

[0169] 发明详述

[0170] 为了披露清晰的目的而非用于限制, 将发明详述分为以下小节:

[0171] 1. 定义

[0172] 2. 单纯疱疹病毒

[0173] 3. 治疗、预防和/或减轻HSV相关的眼部感染的方法

[0174] 4. 改变RS1、RL2和/或LAT基因的方法

[0175] 5. 向导RNA (gRNA) 分子

[0176] 6. 用于设计gRNA的方法

[0177] 7. Cas9分子

[0178] 8. 候选分子的功能分析

[0179] 9. 基因组编辑方法

[0180] 10. 靶细胞

[0181] 11. 递送、制剂和给药途径

[0182] 12. 修饰的核苷、核苷酸和核酸

[0183] 1. 定义

[0184] 如本发明所使用的, 术语“约”或“近似”是指在由本领域普通技术人员确定的特定值的可接受误差范围内, 其可部分取决于如何测量或确定该值, 即, 测量系统的限制。例如, 根据本领域的实践, “约”可表示在3个或3个以上的标准偏差内。或者, “约”可表示给定值的至多20%, 优选至多10%, 更优选至多5%, 还更优选至多1%的范围。或者, 特别是关于生物系统或过程, 该术语可意指数值的一个数量级以内, 优选在5倍以内, 更优选在2倍以内。

[0185] 如本发明所用, “基因组编辑系统”是指具有RNA引导的DNA编辑活性的任何系统。本发明披露的基因编辑系统包括至少两种从天然存在的CRISPR系统调试来的组成部分: 向导RNA (gRNA) 和RNA引导的核酸酶。这两个组成部分形成复合物, 该复合物能够与细胞中的特定核酸序列结合并编辑该核酸序列中或周围的DNA, 例如, 通过制造一个或多个单链断裂 (SSB或缺口) 、双链断裂 (DSB) 和/或点突变来完成。

[0186] 在各个实施方式中, 基因编辑系统可包括 (a) 一个或多个Cas9/gRNA复合物, 和 (b)

分开的Cas9分子和能够在细胞中缔合以形成一个或多个Cas9/gRNA复合物的gRNA。根据本发明披露的一种基因组编辑系统可由一个或多个核苷酸(例如, RNA、DNA)编码, 所述一个或多个核苷酸包括用于Cas9和/或gRNA的编码序列, 所述Cas9和/或gRNA可缔合以形成Cas9/gRNA复合物, 所述一个或多个编码所述基因组编辑系统的核苷酸可由本发明所述的载体来运载。

[0187] 在某些实施方式中, 所述基因组编辑系统靶向HSV病毒基因, 所述HSV病毒基因选自由RS1基因、RL2基因和LAT基因组成的群组。本发明披露的基因组编辑系统可用于改变(敲除或敲低)一个或多个HSV病毒基因, 例如, RS1基因、RL2基因和LAT基因。

[0188] 在某些实施方式中, 所述基因组编辑系统靶向RS1基因。在某些实施方式中, 所述RS1基因是人RS1基因。在某些实施方式中, 所述基因组编辑系统靶向RL2基因。在某些实施方式中, 所述RL2基因是人RL2基因。在某些实施方式中, 所述基因组编辑系统靶向LAT基因。在某些实施方式中, 所述LAT基因是人LAT基因。在某些实施方式中, 所述基因组编辑系统靶向RS1、RL2和LAT基因中的两个或三个。

[0189] 在某些实施方式中, 靶向RS1基因的所述基因编辑系统包括第一gRNA分子或编码其的多核苷酸和至少一个Cas9分子或编码其的多核苷酸, 所述第一gRNA分子包括与RS1基因中的靶标结构域(也称为“靶序列”)互补的靶向结构域。在某些实施方式中, 靶向RS1基因的所述基因组编辑系统还包括第二gRNA分子, 所述第二gRNA分子包括与所述RS1基因中的第二靶标结构域互补的靶向结构域或编码其的多核苷酸。靶向RS1基因的所述基因组编辑系统还可包括靶向RS1基因的第三和第四gRNA分子。

[0190] 在某些实施方式中, 靶向RL2基因的所述基因编辑系统包括第一gRNA分子或编码其的多核苷酸和至少一个Cas9分子或编码其的多核苷酸, 所述第一gRNA分子包括与RL2基因中的靶标结构域互补的靶向结构域。在某些实施方式中, 靶向RL2基因的所述基因组编辑系统还包括第二gRNA分子, 所述第二gRNA分子包括与所述RL2基因中的第二靶标结构域互补的靶向结构域或编码其的多核苷酸。靶向RL2基因的所述基因组编辑系统还可包括靶向RL2基因的第三和第四gRNA分子。

[0191] 在某些实施方式中, 靶向LAT基因的所述基因编辑系统包括第一gRNA分子或编码其的多核苷酸和至少一个Cas9分子或编码其的多核苷酸, 所述第一gRNA分子包括与LAT基因中的靶标结构域互补的靶向结构域。在某些实施方式中, 靶向LAT基因的所述基因组编辑系统还包括第二gRNA分子, 所述第二gRNA分子包括与所述LAT基因中的第二靶标结构域互补的靶向结构域或编码其的多核苷酸。靶向LAT基因的所述基因组编辑系统还可包括靶向LAT基因的第三和第四gRNA分子。

[0192] 在某些实施方式中, 所述基因组编辑系统在细胞中或在体外或体内接触中实施。在某些实施方式中, 所述基因组编辑系统在药物中使用, 例如, 用于修饰一种或多种靶基因(例如, RS1、RL2和/或LAT基因)的药物或用于治疗、预防和/或减轻HSV感染(HSV-1或HSV-2感染)的药物。在某些实施方式中, 所述基因组编辑系统用于治疗。

[0193] 如本发明所用的“靶基因”是指编码已知或推定的基因产物的任何核苷酸序列。在某些实施方式中, 所述靶基因是HSV病毒基因。本发明使用的“HSV病毒基因”指(HSV-1或HSV-2)RS1基因、(HSV-1或HSV-2)RL2基因或(HSV-1或HSV-2)LAT基因。

[0194] 本发明使用的“HSV1 RS1靶敲除位置”指HSV1的RS1基因中的位置, 若该位置被

NHEJ介导的改变而改变,则使功能性RS1基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述RS1基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述RS1基因的非编码区中。

[0195] 本发明使用的“HSV2 RS1靶敲除位置”指HSV2的RS1基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性RS1基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述RS1基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述RS1基因的非编码区中。

[0196] 本发明使用的“HSV RS1靶敲除位置”指HSV(例如,HSV-1或HSV-2)的RS1基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性RS1基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述RS1基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述RS1基因的非编码区中。

[0197] 本发明使用的“HSV1 RS1靶敲低位置”指HSV1的RS1基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性RS1基因产物的表达减少或消除。在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于RS1基因的启动子区中(例如,RS1基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0198] 本发明使用的“HSV2 RS1靶敲低位置”指HSV2的RS1基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性RS1基因产物的表达减少或消除。在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于RS1基因的启动子区中(例如,RS1基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0199] 本发明使用的“HSV RS1靶敲低位置”指HSV(例如,HSV-1或HSV-2)的RS1基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性RS1基因产物的表达减少或消除。在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于RS1基因的启动子区中(例如,RS1基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0200] 本发明使用的“HSV RS1靶标”包括HSV RS1靶敲低位置和/或HSV RS1靶敲除位置。

[0201] 本发明使用的“HSV1 RL2靶敲除位置”指HSV1的RL2基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性RL2基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述RL2基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述RL2基因的非编码区中。

[0202] 本发明使用的“HSV2 RL2靶敲除位置”指HSV2的RL2基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性RL2基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述RL2基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述RL2基因的非编码区中。

[0203] 本发明使用的“HSV RL2靶敲除位置”指HSV(例如,HSV-1或HSV-2)的RL2基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性RL2基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述RL2基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述RL2基因的非编码区中。

[0204] 本发明使用的“HSV1 RL2靶敲低位置”指HSV1的RL2基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性RL2基因产物的表达减少或消除。

在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于RL2基因的启动子区中(例如,RL2基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0205] 本发明使用的“HSV2 RL2靶敲低位置”指HSV2的RL2基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性RL2基因产物的表达减少或消除。在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于RL2基因的启动子区中(例如,RL2基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0206] 本发明使用的“HSV RL2靶敲低位置”指HSV(例如,HSV-1或HSV-2)的RL2基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性RL2基因产物的表达减少或消除。在某些实施方式中,减少或消除转录。在某些实施方式中,所述位置位于RL2基因的启动子区中(例如,RL2基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0207] 本发明使用的“HSV RL2靶标位置”包括HSV RL2靶敲低位置和/或HSV RL2靶敲除位置。

[0208] 本发明使用的“HSV1 LAT靶敲除位置”指HSV1的LAT基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性功能LAT基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述LAT基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述LAT基因的非编码区中。

[0209] 本发明使用的“HSV2 LAT靶敲除位置”指HSV2的LAT基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性功能LAT基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述LAT基因的编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述LAT基因的非编码区中。

[0210] 本发明使用的“HSV LAT靶敲除位置”指HSV(例如,HSV-1或HSV-2)的LAT基因中的位置,若该位置被NHEJ介导的改变而改变,则使功能性LAT基因产物的表达减少或消除。在某些实施方式中,所述位置位于所述LAT基因编码区中(例如,早期编码区)。在某些实施方式中,所述位置位于所述LAT基因的非编码区中。

[0211] 本发明使用的“HSV1 LAT靶敲低位置”指HSV1的LAT基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性LAT基因产物的表达减少或消除。在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于LAT基因的启动子区中(例如,LAT基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0212] 本发明使用的“HSV2 LAT靶敲低位置”指HSV2的LAT基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性LAT基因产物的表达减少或消除。在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于LAT基因的启动子区中(例如,LAT基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0213] 本发明使用的“HSV LAT靶敲低位置”指HSV(例如,HSV-1或HSV-2)的LAT基因中的位置,若该位置被本发明所述的eiCas9或eiCas9-融合蛋白靶向,则使功能性LAT基因产物的表达减少或消除。在某些实施方式中,转录被减少或消除。在某些实施方式中,所述位置位于LAT基因的启动子区中(例如,LAT基因中由eiCas9或eiCas9-融合蛋白靶向的启动子区中的位置)。

[0214] 本发明使用的“HSV LAT靶标位置”包括HSV LAT靶敲低位置和/或HSV LAT靶敲除

位置。

[0215] 如本文所用的“结构域”是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。

[0216] 如下进行两个序列之间的同源性或序列一致性(这些术语在本文可互换地使用)的计算。将这些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酸。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。

[0217] 本发明使用的“支配性gRNA分子”指包括靶向结构域的gRNA分子,所述靶向结构域与核酸上的靶标结构域互补,所述核酸包括编码引入细胞或受试者中的CRISPR/Cas系统的组成部分的序列。支配性gRNA不靶向内源细胞或受试者序列。在某些实施方式中,支配性gRNA分子包括与靶序列互补的靶向结构域,所述靶序列位于:(a)编码Cas9分子的核酸上;(b)编码gRNA的核酸上,所述gRNA包括靶向RS1、RL2或LAT基因(靶基因gRNA)的靶向结构域;或位于编码CRISPR/Cas组成部分((a)和(b)二者)的多于一个的核酸上。在某些实施方式中,编码CRISPR/Cas组成部分(例如,编码Cas9分子或靶基因gRNA)的核酸分子包括多于一个的与支配性gRNA靶向结构域互补的靶标结构域。在某些实施方式中,支配性gRNA分子与Cas9分子复合并导致被靶向的核酸的Cas9介导的失活(例如,通过切割或通过与核酸结合)以及导致CRISPR/Cas系统组成部分产生的停止或减少。在某些实施方式中,所述Cas9分子形成两种复合物:一种复合物包括Cas9分子与靶基因gRNA,该复合物可改变RS1、RL2或LAT基因;而另一种复合物包括Cas9分子与支配性gRNA分子,该复合物可用于防止CRISPR/Cas系统组成部分(例如,Cas9分子或靶基因gRNA分子)的进一步产生。在某些实施方式中,支配性gRNA分子/Cas9分子复合物结合至控制区序列(例如启动子)或促进控制区序列的切割,所述控制区序列与编码Cas9分子的序列、编码转录区域、外显子或内含子的序列可操作连接。在某些实施方式中,支配性gRNA分子/Cas9分子复合物结合至控制区序列(例如,启动子)或促进控制区序列的切割,所述控制区序列与gRNA分子或编码所述gRNA分子的序列可操作连接。在某些实施方式中,支配性gRNA(例如,Cas9-靶向支配性gRNA分子或靶基因gRNA靶向控制gRNA分子)限制了Cas9分子/靶基因gRNA分子复合物介导的基因靶向的作用。在某些实施方式中,支配性gRNA对Cas9分子/靶基因gRNA分子复合物的活性施加时间上、表达水平上或其他限制。在某些实施方式中,支配性gRNA减少脱靶或其他不需要的活性。在某些实施方式中,支配性gRNA分子抑制(例如,完全或基本上完全抑制)Cas9系统的组分的产生并由此限制或支配其活性。

[0218] 本发明使用的“调节剂”指可改变对象分子或基因序列的活性(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。在某些实施方式中,调节包括切割,例如,共价或非共价键的断裂,或共价或非共价键的形成,例如,部分(moiety)连接至对象分子。在某些实施方式中,调节剂改变了对象分子的三维、二级、三级或四级结构。调节剂可增加、减少、启动或消除对象活性。

[0219] 本发明中使用的“大分子”指分子量至少为2、3、5、10、20、30、40、50、60、70、80、90或100kD的分子。大分子包括蛋白质、多肽、核酸、生物制剂和碳水化合物。

[0220] 如本发明所用的“多肽”是指具有少于100个氨基酸残基的氨基酸的聚合物。在某些实施方式中,它具有少于50、20、或10个氨基酸残基。

[0221] 如本文所用的“Cas9分子”或“Cas9多肽”分别是指可以与gRNA分子相互作用,并且与gRNA分子一起定位至包括靶标结构域(并且在某些实施方式中,是PAM序列)(也称作“靶序列”)的位点的分子或多肽。Cas9分子和Cas9多肽包括天然存在的Cas9分子和Cas9多肽,以及工程化的、改变的或经修饰的Cas9分子或Cas9多肽,它们与参考序列(例如最相似的天然存在的Cas9分子)相差例如至少一个氨基酸残基。

[0222] 在某些实施方式中,Cas9分子是识别NGG PAM序列的野生型化脓链球菌Cas9。在某些实施方式中,Cas9分子是化脓链球菌Cas9 EQR变体,其识别NGAG PAM序列、NGCG PAM序列、NGGG PAM序列、NGTG PAM序列、NGAA PAM序列、NGAT PAM序列或NGAC PAM序列。在某些实施方式中,Cas9分子是化脓链球菌Cas9 VRER变体,其识别NGCG PAM序列、NGCA PAM序列、NGCT PAM序列、NGCC PAM序列。在某些实施方式中,Cas9分子是识别NNGRRT PAM序列或NNGRRV PAM序列的野生型金黄色葡萄球菌Cas9。

[0223] 如本发明所用的“参考分子”是指经修饰或候选分子与其作比较的分子。例如,参考Cas9分子是指经修饰或候选Cas9分子与其作比较的Cas9分子。同样地,参考gRNA是指经修饰或候选gRNA分子与其作比较的gRNA分子。经修饰或候选分子可以基于序列(例如,经修饰或候选分子可以与参考分子具有X%序列一致性或同源性)或活性(例如,经修饰或候选分子可以具有参考分子的X%的活性)与参考分子进行比较。例如,在参考分子是Cas9分子的情况下,经修饰或候选分子可以表征为具有不多于参考Cas9分子的核酸酶活性的10%。参考Cas9分子的示例包括天然存在的未经修饰的Cas9分子,例如来自化脓链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏菌的天然存在的Cas9分子。在某些实施方式中,参考Cas9分子是具有和与其进行比较的经修饰或候选Cas9分子最接近序列一致性或同源性的天然存在的Cas9分子。在某些实施方式中,参考Cas9分子是具有天然存在或已知序列的亲本分子,在其上进行突变以到达修饰或候选Cas9分子。

[0224] 如本发明关于分子的修饰所用的“替换”或“替换的”不需要方法限制,而仅表明替换实体是存在的。

[0225] 本发明使用的“小分子”指分子量小于约2kD的化合物,例如小于约2kD、小于约1.5kD、小于约1kD或小于约0.75kD。

[0226] 如本发明所用的“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,人类、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)。在某些实施方式中,所述受试者是人。在其他实施方式中,所述受试者是家禽。

[0227] 本发明使用的“治疗”(treat、treating、treatment)指治疗哺乳动物例如人的疾病,包括(a)抑制疾病,即阻止或预防其发展或进展;(b)缓解疾病,即导致疾病状态消退;(c)缓解疾病的一种或多种症状;和(d)治愈疾病。

[0228] 本发明使用的“预防”(prevent、preventing、prevention)指预防哺乳动物例如人类中的疾病,包括(a)避免或排除疾病;(b)影响疾病倾向;(c)预防或延迟疾病的至少一种

症状的发作。

[0229] 如本文所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何一种),除非另外说明。

[0230] 2.单纯疱疹病毒

[0231] 单纯疱疹病毒(HSV)分为至少两种类型:单纯疱疹病毒1型(HSV-1)和单纯疱疹病毒2型(HSV-2)。HSV-1和HSV-2也分别称为人类疱疹病毒1(HHV-1)和人类疱疹病毒(HHV-2)。

[0232] 疱疹病毒的结构包括包裹在二十面体蛋白笼(衣壳)内的相对较大的双链线性DNA基因组,该二十面体蛋白笼被包裹在称为(包膜)的脂双层中。包膜通过被膜与衣壳接合。该完整的颗粒称作病毒体(Mettenleiter等人,(2006) *Curr. Opin. Microbiol.* 9 (4) :423-429)。HSV-1和HSV-2各自在其基因组中包含至少74个基因(或开放阅读框、ORF)或甚至多达84个独特的蛋白质编码基因,该编码基因由94个推定的ORF组成(McGeoch等人(2006) *Virus Res.* 117 (1) :90-104;Rajcáni等人(2004) *Virus Genes* 28 (3) :293-310)。这些基因编码参与形成病毒的衣壳、皮膜和包膜的各种蛋白质,以及控制病毒的复制和感染性。

[0233] HSV-1和HSV-2的基因组是复杂的并包括两个独特区域,即长独特区域(UL)和短独特区域(US),每个区域含有多个病毒基因。即刻早期基因编码例如调节早期和晚期病毒基因表达的蛋白质。早期基因编码例如涉及DNA复制和某些包膜糖蛋白产生的酶。晚期基因编码例如形成病毒颗粒的蛋白质。HSV基因的转录由受感染宿主的RNA聚合酶II催化(McGeoch等人(2006) *Virus Res.* 117 (1) :90-104)。

[0234] HSV进入宿主细胞涉及几种在包膜的病毒表面上的糖蛋白(例如,糖蛋白B(gB)、糖蛋白C(gC)、糖蛋白D(gD)、糖蛋白H(gH)和糖蛋白L(gL))与在宿主细胞的表面上的受体(例如疱疹病毒进入介体(HVEM)、Nectin-1或3-O硫酸化硫酸乙酰肝素)的相互作用。当与细胞表面上的特定受体结合时,包膜将与宿主细胞膜融合并产生孔,病毒通过该孔进入宿主细胞。在与受体结合后,病毒也可被内吞,且融合可发生在内体。病毒衣壳进入细胞质后,被运送到细胞核。一旦在核入口孔附着于细胞核,衣壳通过衣壳入口将DNA内容物排出。在感染细胞后,产生级联的疱疹病毒蛋白,例如即刻早期、早期和晚期蛋白。

[0235] 单纯疱疹病毒可能会以静止但持久的形式持续存在,称为潜伏感染。在潜伏感染细胞期间,HSV表达潜伏相关转录物(LAT)RNA。LAT可调节宿主细胞基因组并干扰天然细胞死亡机制。通过维持宿主细胞,LAT的表达保留了病毒储库,这导致随后的、通常有症状的周期性复发或非潜伏期的“暴发”特征。无论复发是否有症状,都将发生病毒脱落以产生进一步感染。疱疹病毒DNA含有编码ICP4的基因,该基因是与溶解性感染相关的基因的反式激活因子(Pinnoji等人(2007) *Virol. J.* 4:56)。人类神经元蛋白神经元限制性沉默因子(NRSF)或人类阻遏元件沉默转录因子(REST)可与ICP4基因周围的元件结合并导致组蛋白脱乙酰化,从而阻止从该基因转录起始,进而阻止其他参与裂解周期的病毒基因的转录(Pinnoji等人(2007) *Virol. J.* 4:56;Bedadala等人(2007) *Cell Res.* 17 (6) :546-555)。ICP4蛋白质合成的抑制作用可通过病毒蛋白质ICP0来逆转,这种病毒蛋白质使得NRSF与ICP4基因分离,从而防止了病毒DNA的沉默(Roizman等人(2005) *Cell Cycle* 4 (8) :1019-21)。

[0236] 2.1HSV-感染

[0237] 单纯疱疹病毒通过感染皮肤和粘膜内的上皮细胞进入宿主。最常见的是,HSV-1通过口咽上皮细胞(包括口腔、嘴唇和鼻子的上皮细胞)的感染进入宿主。最常见的是,HSV-2

通过感染肛门生殖器区域的上皮细胞(包括生殖器和肛门的上皮)进入宿主。然而,HSV-1可主要感染肛门生殖器区域,而HSV-2可主要感染口咽。

[0238] HSV-1引起口腔和粘膜的间歇性疼痛。它是一种无处不在、高度传染的病原体。最初感染HSV-1通常引起嘴唇和口腔黏膜的疼痛起泡。

[0239] HSV-2是一种性传播病毒。它通常被称为生殖器疱疹。初次感染HSV-2通常引起生殖器部位的疼痛起泡。该疾病可引起终身反复发作的病毒反应。它是高传染性的并会增加感染HIV的风险,尤其是在患有活动性病变的患者中。

[0240] HSV-1和HSV-2感染在宿主的一生中持续存在。在初次感染期间,病毒通常感染口咽和肛门生殖器区域的细胞,在受影响的区域引起疼痛的囊泡。HSV感染的再激活通常发生在口咽或肛门生殖器区域。然而,再次激活眼睛和中枢神经系统的感染是最严重和有害的HSV表现,因为它们可分别导致失明和永久性神经障碍。原发性和再激活感染可导致永久性神经系统后遗症和失明。HSV-2还增加了受试者发展HIV的风险。对于治疗、预防和/或减少HSV-1和/或HSV-2感染的方法存在相当大的需求。

[0241] 单纯疱疹病毒在上皮细胞内产生即刻早期基因,其编码病毒合成所需的酶和结合蛋白。原发感染后,病毒通过逆行运输从感觉神经轴突运传输到感觉神经背根神经节(DRG)。HSV-1主要传输到达三叉DRG,但其可根据原发感染部位传播到其他感觉神经节。HSV-2主要传输到位于骶骨内的感觉DRG,但可根据原发感染部位到达其他感觉神经节。在DRG中,病毒建立了潜伏感染。潜伏感染在宿主的一生中持续存在。在DRG细胞内,病毒脱掉衣壳,病毒DNA被运输到细胞核中,与潜伏期相关的关键病毒RNA被转录(包括LAT RNA)。

[0242] 在原发性感染期间,受试者通常在持续4-15天的口腔或肛门生殖器区域中经历疼痛起泡。最常见的疼痛包括HSV-1原发感染中的嘴唇、牙龈和鼻粘膜。较少见的是,HSV-1原发感染可能涉及肛门生殖器区域。HSV-2原发感染最常见于肛门生殖器区域,包括大腿周围的阴道、阴唇、子宫颈、阴茎、阴囊、肛门和皮肤。较少见的是,HSV-2原发感染涉及口咽部。极少见的是,HSV-1和HSV-2原发感染可能涉及眼睛、中枢神经系统、手指和指甲病床(疱疹白癜风)。HSV-1感染主要通过唾液和/或性行为传播。HSV-2感染主要通过性行为传播,但也可能通过唾液传播。HSV感染的水泡可能破裂,释放出高度传染性的清澈液体。原发感染常伴有流感样疾病,包括发烧、寒战和肌肉酸痛。

[0243] 宿主免疫防御对抗HSV感染非常重要。CD4+T细胞和CD8+细胞负责识别和清除病原体。T细胞反应受损的受试者,包括HIV患者、器官移植后接受免疫抑制剂的患者以及免疫系统发育中的新生儿,都易发生HSV-1和HSV-2感染的最严重表现。

[0244] 潜伏感染的重新激活通常不那么严重,并且持续时间可能较短。重新激活HSV-1感染最通常影响口腔区域,但也可能影响其他区域,包括肛门生殖器区域、眼睛、中枢神经系统(CNS)、指甲和咽部。重新激活HSV-2感染通常影响肛门生殖器区域,但也可能影响其他区域,包括口腔区域、眼睛、中枢神经系统(CNS)、指甲和咽部。HSV-1或HSV-2感染的重新激活可导致眼科疾病,包括角膜炎(上皮性角膜炎,基质角膜炎和盘状角膜炎)。一般来说,HSV-1和HSV-2的眼科表现包括疼痛、撕裂、眼睛发红和对光敏感。大多数与HSV相关的眼部感染的消退没有永久性视力损害。然而,眼疱疹感染很少会导致疤痕,细菌病原体继发感染,且很少失明。HSV-1或HSV-2感染的重新激活也可引起视网膜炎。HSV相关的视网膜炎很罕见但较严重,并且具有永久失明的高风险。

[0245] 新生儿是患有严重HSV-1和HSV-2感染风险的人群。这种疾病在分娩过程中由母亲传染给胎儿。母亲在怀孕期间发生原发性HSV感染的情况下,母胎传播的机会最高。新生儿疱疹的发病率约为每100,000名新生儿4-30人。新生儿可出现严重的HSV-1或HSV-2脑炎和/或脑膜炎。尽管接受即时的抗病毒治疗,感染HSV-1或HSV-2的新生儿的永久性神经系统后遗症的发生率还是较为显著。在一项用高剂量抗病毒疗法治疗的HSV脑炎或脑膜炎婴儿的研究中,发现死亡率为4%,幸存者中69%有永久性神经系统后遗症(Kimberlin等人, *Pediatrics*, 2001; 108:230-238)。

[0246] 原发性HSV-1和HSV-2感染可采用抗病毒治疗措施,包括阿昔洛韦、伐昔洛韦和泛昔洛韦。这些治疗措施已表现为病毒脱落减少、疼痛减少和损伤的愈合时间的改善。潜伏感染的重新激活可能在没有治疗的情况下消退(可能是自限性的)或可能用抗病毒治疗措施来治疗。治疗主要在急性感染期间给予。没有治愈或预防性治疗措施。治疗措施可在某些情况下预防性给予,包括在患有最近HSV-1或HSV-2感染或重新激活的母亲的分娩期间。

[0247] 没有有效的治疗方法可预防HSV-1或HSV-2感染。在活动性感染期间使用抗病毒治疗以及使用安全套可将传播速率降低约50%。

[0248] 人类免疫缺陷病毒-1(HIV-1)感染率在HSV-2血清阳性的受试者中显著增加。感染HIV-1的风险是HSV-2患者的3倍。抗病毒药物对降低感染HIV的风险没有影响。

[0249] 2.2HSV相关的眼部疾病

[0250] HSV感染,例如眼睛的HSV-1和/或HSV-2感染(原发感染或重新激活感染),被称为HSV相关眼病。HSV相关眼病最常引起眼前房感染,称为角膜炎、基质角膜炎和/或盘状角膜炎。HSV相关的眼部疾病可能更罕见地引起眼后房感染,称为视网膜炎。HSV-1角膜炎是剧烈疼痛和不愉快的。其极少会导致疤痕、细菌病原体继发感染,且很少会失明。HSV相关性视网膜炎是HSV相关眼病的罕见表现,但是具有更高的永久性视力损伤风险。

[0251] 通过将病毒从三叉神经节沿着三叉神经的眼科分支(第五颅神经)顺行运送到眼中并进入眼睛,在眼中发生重新激活感染。病毒的重新激活也可能发生在角膜内。通过两种机制之一建立三叉神经节内的潜伏期。首先,HSV-1或HSV-2可通过沿三叉神经从眼睛(眼感染后)逆行转运到三叉神经节。或者,它可以在感染口腔粘膜、生殖器部位或其他眼外部位后通过血源性播散扩散至三叉神经节。在建立三叉神经节的潜伏感染后,在任何时候,特别是在免疫受损的宿主中,病毒可通过沿三叉神经顺行行进并重新建立感染而进入眼睛。

[0252] 当眼部疱疹影响眼后房时,会引起视网膜炎。在成人中,HSV-1是造成HSV-视网膜炎的大多数病例的原因(Pepose等人, *眼科感染和免疫*, 1996; Mosby 1155-1168)。在新生儿和儿童中,HSV-2是造成HSV-视网膜炎的大多数病例的原因(Pepose等人, *眼科感染和免疫*, 1996; Mosby 1155-1168)。与HSV有关的视网膜炎可导致急性视网膜坏死(ARN),这将在2周内没有治疗的情况下破坏视网膜(Banerjee和Rouse, *人类疱疹病毒*, 2007; 剑桥大学出版社,第35章)。即使接受治疗,ARN后永久性视力损害的风险也高于50% (Roy等人, *眼科免疫学和炎症* 2014; 22 (3): 170-174)。

[0253] 角膜炎是眼疱疹最常见的形式。HSV角膜炎可表现为树突状角膜炎、基质角膜炎、睑缘炎和结膜炎。HSV-1是造成大部分HSV相关性角膜炎的原因,占病例的58% (Dawson等人, *眼科综述*, 1976; 21 (2): 121-135)。HSV-2占HSV相关性角膜炎病例的其余部分,约占病例的42%。在美国,每年约有48,000例复发性或原发性HSV相关性角膜炎感染病例(Liesegang

等人,1989;107 (8) :1155-1159)。在HSV相关性角膜炎的所有病例中,约1.5-3%的受试者经历严重的永久性视力障碍(Wilhelmus等人,眼科档案,1981;99 (9) :1578-82)。HSV相关眼病导致的永久性视力损伤的风险随着眼相关HSV重新激活的数量的增加而增加。

[0254] 总的来说,基质性角膜炎约占角膜炎的15%,与最高的角膜炎永久性视力损伤的风险有关。间质性角膜炎导致疤痕和不规则散光。先前的眼部HSV感染增加了发展为基质感染的风险,这意味着先前有过眼部HSV感染的受试者在重新激活后有增加的永久性视觉损伤的风险。在儿童中,基质性角膜炎占所有角膜炎病例的60%。因此,儿童特别容易受到HSV相关性角膜炎的永久性视力损害。美国1950-1982年的一项回顾性研究发现,每100,000人年约有2.6例新发或复发性基质性角膜炎病例,或每年约8,000例基质性角膜炎(Liesegang等人,1989;107 (8) :1155-1159)。2002年在法国进行的一项最新研究估计新发或复发性基质性角膜炎病例的发病率为9.6/10万(Labetoulle等人,眼科,2005;112 (5) :888-895)。在发达国家,HSV相关性角膜炎的发病率可能会增加(Farooq和Shukla,2012;眼科综述,57 (5) :448-462)。

[0255] 本发明所述的基因组编辑系统、组合物和方法可用于治疗、预防和/或减少HSV-1和/或HSV-2的眼部感染,包括但不限于HSV-1基质角膜炎、HSV-1树突状角膜炎、HSV-1睑缘炎、HSV-1结膜炎、HSV-1视网膜炎、HSV-2基质角膜炎、HSV-2树突状角膜炎、HSV-2睑缘炎、HSV-2结膜炎和HSV-2视网膜炎。

[0256] 3. 治疗、预防和/或减轻HSV相关的眼部感染的方法

[0257] 本发明披露了使用本发明所述方法、基因组编辑系统和组合物来治疗、预防和/或减少HSV相关眼部感染的途径。HSV相关的眼部感染可能由HSV-1和/或HSV-2感染引起。例如,而不是作为限制,本发明披露的方法、基因组编辑系统和组合物可用来治疗、预防和/或减少HSV-1感染、HSV-2感染或HSV-1和HSV-2二种感染。

[0258] HSV-1和HSV-2中的RS1、RL2和LAT基因与与病毒感染、增殖和组装以及潜伏期维持和病毒重新激活的维持有关。敲除或敲低任何这些单独的或组合的基因可减少HSV-1和/或HSV-2感染。由于HSV-1或HSV-2病毒在身体内分散的局部区域建立起潜伏,因此非常适合于在潜伏区域提供失效性治疗的局部递送。在分散区域的靶向敲除(例如三叉神经背根神经节、角膜、颈背根神经节或骶背根神经节)可通过使HSV-1和/或HSV-2病毒失效来减少或消除潜伏感染。

[0259] 本发明描述了通过敲除或敲低病毒基因来治疗、预防和/或减少HSV-1和/或HSV-2感染的方法。本发明描述的方法包括敲除或敲低以下HSV-1和/或HSV-2编码的基因:RL2、LAT和RS1,或其任意组合(例如,任何单个基因,例如,RL2,例如LAT,例如RS1,或任何两个基因,例如RL2和LAT,例如,RL2和RS1,例如,RS1和LAT,或三个基因)。当存在两个改变事件(例如,敲低或敲除RS1、RL2和/或LAT基因的表达)时,两个改变事件可按顺序地或同时发生。在某些实施方式中,RS1、RL2和/或LAT基因的敲除发生在RS1、RL2和/或LAT基因的敲低之前。在某些实施方式中,RS1、RL2和/或LAT基因的敲除与RS1、RL2和/或LAT基因的敲低同时发生。在某些实施方式中,RS1、RL2和/或LAT基因的敲除在RS1、RL2和/或LAT基因的敲低之后发生。在某些实施方案中,改变的效果是协同的。

[0260] RL2编码基因ICP0,即作为基因表达的反式激活因子的775个氨基酸的蛋白质。RL2基因是由疱疹病毒表达的五种即刻早期基因之一。ICP0参与延迟的早期和晚期基因的表达

的激活 (Lees-Miller等人, 1996, 病毒学杂志, 70 (11) : 7471-7477)。ICP0被认为涉及神经毒力。在细胞培养中,已经发现ICP0是从潜伏期重新激活所需的 (Leib等人, 1989, 病毒学杂志, 63: 759-768)。已经显示不表达RL2的缺失突变体不能在体外复制 (Sacks和Schaffer, 1987, 病毒学杂志, 61 (3) : 829-839)。在某些实施方式中,敲除RL2可使HSV-1和/或HSV-2失去从潜伏重新激活的能力。在某些实施方式中,敲除或敲低RL2可使HSV-1和/或HSV-2失去复制的能力。在某些实施方式中,敲除或敲低RL2可使HSV-1和/或HSV-2失去感染和/或建立神经组织中的潜伏感染的能力。

[0261] LAT编码在潜伏期内由疱疹病毒表达的唯一基因。潜伏期是病毒在宿主组织中建立静止感染的时间,通常在神经组织中,包括三叉神经节或骶神经节。LAT被认为参与疱疹病毒感染的重新激活,使病毒再次感染上皮和其他组织。在某些实施方式中,敲除或敲低LAT可使HSV-1和/或HSV-2基因潜伏期和/或再激活失效,破坏HSV-1和/或HSV-2在潜伏感染后维持潜伏感染和/或再激活的能力。在某些实施方式中,敲除或敲低LAT表达消除了HSV-1和/或HSV-2的潜伏感染。在某些实施方式中,敲除或敲低LAT表达缩短了HSV-1和/或HSV-2感染的持续时间、治疗和/或治愈HSV-1和/或HSV-2感染。

[0262] RS1在由HSV-1和HSV-2表达即刻早期基因中起重要作用。RS1是由疱疹病毒表达的五种即刻早期基因之一,并是主要的转录调节因子。RS1编码病毒蛋白质ICP4。ICP4对于控制由HSV-1和HSV-2产生的早期和晚期基因的整体表达是重要的。RS1基因在HSV-1和HSV-2中类似。

[0263] 在某些实施方式中,敲除或敲低RS1、RL2和/或LAT基因使HSV-1和/或HSV-2基因表达失活,或减少病毒复制、装配、成熟、包装或感染中的一种或多种。在某些实施方式中,敲除RS1、RL2和/或LAT基因表达缩短了HSV-1和/或HSV-2感染的持续时间。在某些实施方式中,敲除或敲低RS1、RL2和/或LAT基因表达治疗或治愈了HSV-1和/或HSV-2感染。

[0264] 在某些实施方式中,减少与眼相关的HSV再激活的持续时间、数量和/或频率可降低感染HSV-1和/或HSV-2的受试者的永久性视觉损伤的风险。

[0265] 在某些实施方式中,单独或组合敲除和/或敲低RS1、RL2和/或LAT基因可使HSV-1和/或HSV-2更易于抗病毒治疗。重要基因的突变可使HSV-1、HSV-2和其他病毒更容易受到抗病毒药物的治疗 (Zhou等人, 病毒学杂志, 2014; 88 (19) : 11121-11129)。单独或组合敲除或敲低RL2和LAT和/或RS1基因可与抗病毒疗法组合以治疗、预防和/或减少HSV-1和/或HSV-2感染。本发明描述的组合物和方法可与另一种抗病毒治疗方法 (例如, 另一种抗HSV-1治疗或抗HSV-2治疗方法) 组合使用来治疗、预防和/或减少HSV-1或HSV-2感染。

[0266] 在一种方法中,一个、两个或三个RS1、RL2和LAT基因被靶向作为靶向敲除或敲低,例如,以抑制一种或多种病毒功能,包括例如病毒基因调控、病毒基因转录、病毒基因组复制、病毒潜伏基因的表达和病毒衣壳形成。在某些实施方式中,所述方法包括敲除一个HSV-1和HSV-2基因 (例如,RS1、RL2或LAT)。在某些实施方式中,所述方法包括敲低一个HSV-1和HSV-2基因 (例如,RS1、RL2或LAT)。在某些实施方式中,所述方法包括敲除两个HSV-1和/或HSV-2基因,例如,RL2和LAT基因,例如,RL2和RS1基因,例如,RS1和LAT基因。在某些实施方式中,所述方法包括敲低两个HSV-1和/或HSV-2基因,例如,RL2和LAT基因,例如,RL2和RS1基因,例如,RS1和LAT基因。在某些实施方式中,所述方法包括敲除三个HSV-1和/或HSV-2基因,例如,所有三个RL2、LAT和RS1基因。在某些实施方式中,所述方法包括敲低三个HSV-1

和/或HSV-2基因,例如,所有三个RS1、RL2和LAT基因。

[0267] 在某些实施方式中,抑制一种或多种病毒功能(例如病毒基因调节、病毒基因转录、病毒基因组复制和病毒衣壳形成)减少了原发性或复发性感染的持续时间和/或减少病毒颗粒的脱落。受试者也可能经历较短的疾病持续时间、降低向性伴侣传播的风险、在怀孕的情况下降低向胎儿传播风险和/或有可能完全清除HSV-1和/或HSV-2(治愈)。

[0268] 敲除或敲低一个或多个靶基因(例如,RS1、RL2或LAT基因)的一个或多个拷贝(例如,1、2、3、4、5、6、7、8、9、10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000或更多拷贝)可在疾病发作或疾病发作之后进行,优选在疾病过程中的早期进行。

[0269] 在某些实施方式中,所述方法包括在疾病发作之前开始治疗受试者。

[0270] 在某些实施方式中,所述方法包括在疾病发作之后开始治疗受试者。

[0271] 在某些实施方式中,所述方法包括在疾病发作后开始治疗受试者,例如,HSV-1和/或HSV-2感染发病后的1、2、3、4、5、6、7、8、9、10、12、16、24、36、48或更多个月。在某些实施方式中,所述方法包括在疾病发作后开始治疗受试者,例如,HSV-1和/或HSV-2感染发病后的1、2、3、4、5、10、15、20、25、40、50或60年。

[0272] 在某些实施方式中,所述方法包括在疾病的晚期阶段对受试者开始治疗,例如在急性或潜伏期期间进行治疗。在某些实施方式中,所述方法包括在影响中枢神经系统、眼睛、口咽、生殖器区域和/或其他区域的疾病严重急性阶段开始治疗受试者。

[0273] 总体而言,预期疾病各阶段受试者开始治疗可改善愈合,减少疾病持续时间并有益于受试者。

[0274] 在某些实施方式中,所述方法包括在疾病进展之前开始治疗受试者。在某些实施方式中,所述方法包括在疾病的早期阶段开始治疗受试者,例如,当受试者已经暴露于HSV-1和/或HSV-2或被认为已经暴露于HSV-1和/或HSV-2时。

[0275] 在某些实施方式中,所述方法包括在疾病进展之前开始治疗受试者。在某些实施方式中,所述方法包括在疾病的早期阶段开始治疗受试者,例如,当受试者对HSV-1和/或HSV-2感染呈阳性但没有体征或症状时。

[0276] 在某些实施方式中,所述方法包括在出现与HSV-1和/或HSV-2感染一致或相关的下一个或多个发现时开始治疗:发热、头痛、身体疼痛、肛门生殖器起疱、口腔溃疡、脑炎或角膜炎。

[0277] 在某些实施方式中,所述方法包括在口腔内或口腔周围出现疼痛性起泡(例如口腔或口咽,例如婴儿、儿童、成人或年轻成人)时开始治疗受试者。

[0278] 在某些实施方式中,所述方法包括在例如婴儿、儿童、成人或年轻成人出现肛门生殖器区域中的疼痛性起疱、生殖器溃疡和/或流感样症状时开始治疗受试者。

[0279] 在某些实施方式中,所述方法包括开始治疗疑似患有HSV-1和/或HSV-2脑膜炎和/或HSV-1和/或HSV-2脑炎的受试者。

[0280] 在某些实施方式中,所述方法包括在出现与HSV-1和/或HSV-2脑膜炎和/或脑炎一致或相关的一种或多种以下症状时开始治疗:发热、头痛、呕吐、畏光、癫痫发作、意识水平下降昏睡或嗜睡。

[0281] 在某些实施方式中,所述方法包括在出现与HSV-1和/或HSV脑膜炎和/或脑炎一致或相关的任何以下迹象时开始治疗:针对HSV-1和/或HSV-2的阳性CSF培养物、CSF中升高的

WBC、颈部僵硬/阳性布鲁津斯基征的迹象。在某些实施方式中,所述方法包括对具有与HSV-1和/或HSV-2脑炎和/或脑膜炎相一致的符合EEG、CSF检查、MRI、CSF标本的PCR和/或脑活检标本的PCR的患者开始治疗。

[0282] 在某些实施方式中,所述方法包括在出现与视觉HSV-1和/或HSV-2一致或相关的任何以下症状时开始治疗:疼痛、畏光、视力模糊、撕裂、发红/注射、视力丧失、飞蚊症或闪光。

[0283] 在某些实施方式中,所述方法包括在出现与眼HSV-1和/或HSV-2(也称为HSV-1和/或HSV-2角膜炎)一致或相关的眼科检查的任何以下发现时开始治疗:在角膜上皮出现的小的突出的清亮囊泡;角膜表面不规则,点状上皮糜烂;致密的基质浸润;溃疡;坏死;局灶性、多灶性或弥漫性细胞浸润;免疫环;新生血管形成;或在角膜出现的任何级别的影子血管。

[0284] 在某些实施方式中,所述方法包括在出现与HSV-1和/或HSV-2视网膜炎或急性视网膜坏死一致或相关的眼科检查的任何以下发现时开始治疗:视敏度降低;葡萄膜炎;玻璃体炎;巩膜注射;前房和/或玻璃体腔的炎症;玻璃浑浊;视神经水肿;周边视网膜增白;视网膜撕裂;视网膜脱落;视网膜坏死;闭塞性血管病变伴有动脉受累的证据,包括小动脉鞘和小动脉衰减。

[0285] 在某些实施方式中,所述方法包括在出现与眼睛、口咽、肛门生殖器区域或中枢神经系统的HSV-1或HSV-2感染一致或相关的症状和/或体征时开始治疗。在某些实施方式中,在疾病过程早期对可疑的HSV-1或HSV-2感染的情况开始治疗HSV-1和/或HSV-2感染是有益的。

[0286] 在某些实施方式中,所述方法包括开始宫内治疗。在某些实施方式中,受试者处于母胎传播的高风险中。

[0287] 在某些实施方式中,所述方法包括在母亲患有HSV-1和/或HSV-2活性感染或具有最近的原发性HSV-1和/或HSV-2感染的情况下开始治疗。

[0288] 在某些实施方式中,所述方法包括在器官移植之前或器官移植之后立即开始治疗。

[0289] 在某些实施方式中,所述方法包括在疑似暴露于HSV-1和/或HSV-2的情况下开始治疗。

[0290] 在某些实施方式中,所述方法包括在疑似HSV脑炎或脑膜炎的情况下预防性地开始治疗。

[0291] 在某些实施方式中,认为HIV阳性受试者和移植后受试者可能由于免疫缺陷而经历严重的HSV-1和/或HSV-2激活或再激活,包括HSV-脑炎和脑膜炎。新生儿在分娩过程中也由于母胎传播而处于严重HSV脑炎的风险中。抑制一种或多种病毒功能(例如,病毒基因调节、病毒基因转录、病毒基因组复制和病毒衣壳形成)可为处于严重HSV-1和/或HSV-2感染风险的所述群体提供优良的保护。受试者可能经历低发病率的HSV-1和/或HSV-2脑炎和/或低发病率的HSV-1和/或HSV-2脑炎后严重神经后遗症,这将极大地改善生活质量。

[0292] 在某些实施方式中,所述方法包括开始治疗患有或有风险发展HSV-1和/或HSV-2感染的严重表现的受试者,例如新生儿、患有HIV的受试者、正在进行免疫抑制剂疗法的受试者(例如,在器官移植之后),患有癌症的受试者、正在接受化疗的受试者、将接受化疗的受试者、正在接受放射疗法的受试者、将接受放射疗法的受试者。

[0293] 在某些实施方式中, HIV阳性受试者和移植后受试者可能由于免疫缺陷而经历严重的HSV-1和/或HSV-2激活或再激活, 包括HSV-脑炎和脑膜炎。新生儿在分娩过程中也由于母胎传播而处于严重HSV脑炎的风险中。抑制主要的病毒功能(例如, 病毒基因调节、病毒基因转录、病毒潜伏基因的表达、病毒基因组复制和病毒衣壳形成)可为处于严重HSV-1和/或HSV-2感染风险的所述群体提供优良的保护。受试者可能经历低发病率的HSV-1和/或HSV-2脑炎和/或低发病率的HSV-1和/或HSV-2脑炎后严重神经后遗症, 这将极大地改善生活质量。

[0294] 在某些实施方式中, 所述方法包括开始对HSV-1和/或HSV-2测试呈阳性的受试者进行治疗。

[0295] 在某些实施方式中, 所述方法包括开始对HSV-1和/或HSV-2感染呈阳性的受试者进行治疗。例如, 可使用病毒培养、直接荧光抗体研究、皮肤活检、PCR、血液血清学测试、CSF血清学测试、CSF PCR或脑活检来测试HSV-1和/或HSV-2感染。在某些实施方式中, 所述方法包括通过诊断性玻璃体切割术、视网膜内活检或水性流体PCR、对玻璃体样品进行PCR来对经测试HSV-2感染呈阳性的受试者开始治疗。

[0296] 在某些实施方式中, 所述方法包括在暴露于HSV-1和/或HSV-2并在处于高风险HSV感染的严重后遗症的受试者中开始治疗。

[0297] 在某些实施方式中, 通过编辑(例如, 引入突变)一个或多个靶基因, 例如RS1、RL2或LAT基因来操纵细胞。在某些实施方式中, 例如, 在体内调节一种或多种靶基因(例如本发明所述的一种或多种RS1、RL2或LAT基因)的表达。

[0298] 在某些实施方式中, 所述方法包括通过腺伴随病毒(AAV)递送gRNA分子。在某些实施方式中, 所述方法包括通过慢病毒(LV)递送gRNA分子。在某些实施方式中, 所述方法包括通过纳米颗粒递送gRNA分子。

[0299] 在某些实施方式中, 所述方法还包括向受试者施用第二种抗病毒疗法或治疗剂, 例如本发明所述的抗HSV-1或抗HSV-2疗法或治疗剂。组合物和其他治疗或治疗剂可按任何顺序施用。例如, 本发明描述的组合物可与一种或多种另外的治疗剂或治疗剂同时、在其之前或之后施用。在某些实施方式中, 两种或更多种疗法或治疗剂的效果是协同的。示例性的抗HSV-1和抗HSV-2疗法和治疗剂包括但不限于阿昔洛韦、伐昔洛韦、泛昔洛韦、喷昔洛韦或疫苗。

[0300] 4. 改变RS1、RL2和/或LAT基因的方法

[0301] 本发明披露的RS1、RL2和/或LAT基因可通过本发明所述的基因组编辑系统、组合物和方法来改变。

[0302] 本发明提供所述方法、基因组编辑系统和组合物改变(例如, 敲除或敲低)HSV RS1靶标、HSV RL2靶标或HSV LAT靶标。

[0303] 本发明披露的HSV RS1靶标、HSV RL2靶标或HSV LAT靶标可通过基因编辑单独或组合改变, 例如, 使用本发明所述CRISPR-Cas9介导的方法、基因组编辑系统和组合物。可例如, 通过以下方式实现(例如, 敲除或敲低)HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变:

[0304] (1) 敲除所述RS1、RL2或LAT基因:

[0305] (a) 插入或缺失紧邻或在所述RS1、RL2或LAT基因的早期编码区内一个或多个核

昔酸(例如,NHEJ介导的插入或缺失);或

[0306] (b) 缺失包括所述RS1、RL2或LAT基因的至少一部分的基因组序列或多个基因组序列(例如,NHEJ介导的缺失);

[0307] (2) 通过靶向所述RS1、RL2或LAT基因的非编码区(例如,启动子区)来敲低由eCas9分子或eCas9-融合蛋白介导的RS1、RL2或LAT基因。

[0308] 所有方法引起所述RS1、RL2和/或LAT基因的改变(例如,敲除或敲低)。可与所述RS1、RL2和/或LAT基因之一或两者的改变相关联的示例性机制包括,但不限于非同源末端连接(例如,经典或替代)、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)、SDSA(合成依赖性链退火)、单链退火或单链侵入。

[0309] 在某些实施例中,本发明所述方法、基因组编辑系统和组合物在所述RS1、RL2和/或LAT基因的早期编码区附近引入一个或多个断裂。在某些实施例中,本发明所述方法、基因组编辑系统和组合物在所述RS1、RL2和/或LAT基因的至少一部分侧面引入两个或多个断裂。所述两个或多个断裂移除(例如,缺失)包括所述RS1、RL2和/或LAT基因的至少一部分的基因组序列。在某些实施例中,本发明所述方法包括通过靶向HSV RL2和/或HSV LAT和/或RS1靶标敲低位置的启动子区域来敲低由eCas9分子或eCas9-融合蛋白介导的RS1、RL2和/或LAT基因。所有方法引起所述RS1、RL2和/或LAT基因的改变(例如,敲除或敲低)。

[0310] 4.1通过在RS1、RL2或LAT基因中引入indel或缺失来敲除RS1、RL2或LAT基因

[0311] 在某些实施方式中,所述方法包括紧邻RS1、RL2和/或LAT基因的HSV RS1靶标敲除位置、HSV RL2靶标敲除位置或HSV LAT靶标敲除位置(例如,早期编码区)引入一个或多个核昔酸的插入或缺失。如本发明所述,在某些实施方式中,所述方法包括在足够接近HSV RL2靶标敲除位置或HSV LAT靶标敲除位置的早期编码区(例如,5'或3'端)来引入一个或多个断裂(例如,单链断裂或双链断裂),从而可合理预期断裂引起的indel跨越HSV RL2靶标敲除位置或HSV LAT靶标敲除位置(例如,早期编码区)。NHEJ介导的断裂修复允许在接近或在HSV RL2靶标敲除位置或HSV LAT靶标敲除位置的早期编码区域内引入NHEJ介导的indel。

[0312] 在某些实施方式中,所述方法包括引入包括RS1、RL2和/或LAT基因的至少一部分的基因组序列的缺失。如本发明所述,在一个实施方式中,所述方法包括在RL2、LAT或RS1靶标位置的一个5'或另一个3'(即侧翼)处引入双链断裂。在一个实施方式中,两个gRNA(例如单分子(或嵌合)或模块化gRNA分子)被构造成将所述两个双链断裂定位在RS1、RL2和/或LAT基因中RL2、LAT或RS1靶标敲除位置的相对侧上。

[0313] 在某些实施方式中,在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处引入单链断裂(例如,由一个gRNA分子定位)。在某些实施方式中,单个gRNA分子(例如,具有Cas9切口酶)用来在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处产生单链断裂,例如,所述gRNA分子被构造成使所述单链断裂定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游(例如,上游200bp内)或下游(例如,下游200bp内)处。在某些实施方式中,定位所述断裂以避免出现不期望的靶染色体元件,例如重复元件,例如,Alu重复。

[0314] 在某些实施方式中,在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处引入双链断裂(例如,由一个gRNA分子定位)。在某些实施方式中,单个gRNA分子(例如,具有Cas9核酸酶而不是Cas9切口酶)用来在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密

靠近处产生双链断裂,例如,所述gRNA分子被构造成使所述双链断裂定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游(例如,上游200bp内)或下游(例如,下游200bp内)处。在某些实施方式中,定位所述断裂以避免出现不期望的靶染色体元件,例如重复元件,例如,Alu重复。

[0315] 在某些实施方式中,在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处引入两个单链断裂(例如,由两个gRNA分子定位)。在某些实施方式中,两个gRNA分子(例如,具有一个或两个Cas9切口酶)用来在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处产生两个单链断裂,例如,所述gRNA分子被构造成使所述单链断裂均定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游(例如,上游200bp内)或下游(例如,下游200bp内)处。在某些实施方式中,两个gRNA分子(例如,具有两个Cas9切口酶)用来在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处产生两个单链断裂,例如,所述gRNA分子被构造成使一个单链断裂定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游处(例如,上游200bp内),且使另一个定位在其下游(例如,下游200bp内)处。在某些实施方式中,定位所述断裂以避免出现不期望的靶染色体元件,例如重复元件,例如,Alu重复。

[0316] 在某些实施方式中,在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处引入两个双链断裂(例如,由两个gRNA分子定位)。在某些实施方式中,两个gRNA分子(例如,具有一个或两个Cas9核酸酶而不是Cas9切口酶)用来在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标两侧产生两个双链断裂,例如,所述gRNA分子被构造成使一个双链断裂定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游处(例如,上游200bp内),且使第二个双链断裂定位在下游(例如,下游200bp内)处。在某些实施方式中,定位所述断裂以避免出现不期望的靶染色体元件,例如重复元件,例如,Alu重复。

[0317] 在某些实施方式中,在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处引入一个双链断裂和两个单链断裂(例如,由三个gRNA分子定位)。在某些实施方式中,三个gRNA分子(例如,具有Cas9核酸酶而不是Cas9切口酶和一个和两个Cas9切口酶)用来在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的侧翼处产生一个双链断裂和两个单链断裂,例如,所述gRNA分子被构造成使所述双链断裂定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游或下游处(例如,上游或下游200bp内),且使所述两个单链断裂定位在相反位置,例如,HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的下游或上游处(例如,下游或上游200bp内)。在某些实施方式中,定位所述断裂以避免出现不期望的靶染色体元件,例如重复元件,例如,Alu重复。

[0318] 在某些实施方式中,在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处或紧密靠近处引入四个单链断裂(例如,由四个gRNA分子定位)。在某些实施方式中,四个gRNA分子(例如,具有一个或多个Cas9切口酶)用来在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标侧翼处产生四个单链断裂,例如,所述gRNA分子被构造成使第一个和第二个单链断裂定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的上游处(例如,上游200bp内),且使第三个和第四个定位在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的下游(例如,下游200bp内)处。在某些实施方式中,定位所述断裂以避免出现不期望的靶染色体元件,例如重复元件,例如,Alu重复。

[0319] 在某些实施方式中,两个或多个(例如,三个或四个)gRNA分子与一个Cas9分子或Cas9-融合蛋白一起使用。在某些实施方式中,当两个或多个(例如,三个或四个)gRNA与两

个或多个Cas9分子一起使用时,至少一个Cas9分子来自与其他Cas9分子不同的物种。例如,当两个gRNA分子与两个Cas9分子一起使用时,一个Cas9分子可来自一个物种,而另一个Cas9分子可来自不同的物种。需要的话,Cas9物种均用来产生单链或双链断裂。

[0320] 4.2. 通过使包括所述RS1、RL2和/或LAT基因的至少一部分的基因组序列或多个基因组序列缺失(例如,NHEJ介导的缺失)来敲除RS1、RL2和/或LAT基因中的一个或多个

[0321] 在某些实施方式中,所述方法包括使包括所述RS1、RL2和/或LAT基因的至少一部分的基因组序列或使包括所述RS1、RL2和/或LAT的至少一部分的多个基因组序列缺失(例如,NHEJ介导的缺失)。在某些实施方式中,所述方法包括在HSV RS1靶标敲除位置、HSV RL2靶标敲除位置或HSV LAT靶标敲除位置的一个5'或另一个3'(即侧翼)处引入两个双链断裂。在某些实施方式中,两个gRNA(例如,单分子(或嵌合)或模块化gRNA分子)被构造成将两个双链断裂定位在所述RL2基因中HSV RL2靶标敲除位置的相反位置上。在某些实施方式中,两个gRNA(例如,单分子(或嵌合)或模块化gRNA分子)被构造成将所述两个双链断裂定位在所述LAT基因中HSV LAT靶标敲除位置的相反位置上。在某些实施方式中,两个gRNA(例如,单分子(或嵌合)或模块化gRNA分子)被构造成将所述两个双链断裂定位在所述RS1基因中HSV RS1靶标敲除位置的相反位置上。

[0322] 4.3. 敲低由无酶促活性的Cas9 (eiCas9) 分子或eiCas9-融合蛋白介导的所述RS1、RL2和/或LAT基因中的一个或多个

[0323] 靶向敲低方法减少或消除功能性RS1、RL2和/或LAT基因产物的表达。如本发明所述,在某些实施方式中,靶向敲低通过将无酶促活性的Cas9 (eiCas9) 分子或eiCas9-融合蛋白(例如与转录抑制子结构域或染色质修饰蛋白融合的eiCas9) 靶向RS1、RL2和/或LAT基因中的一个、两个或三个来介导。

[0324] 本发明所述方法和组合物可用来通过靶向转录调节区(例如,启动子区(例如,控制所述RS1、RL2和/或LAT基因中的一个或多个的转录的启动子区))来改变所述RS1、RL2和/或LAT基因的表达,以治疗或预防HSV-1或HSV-2感染。在某些实施方式中,所述启动子区被靶向以敲低RS1、RL2和/或LAT基因中的一个或多个的表达。靶向敲低方法减少或消除功能性RS1、RL2和/或LAT基因产物的表达。

[0325] 在某些实施方式中,一种或多种eiCas9分子可用于阻断一种或多种内源转录因子的结合。在某些实施方式中,eiCas9可与染色质修饰蛋白融合。改变染色质状态可以导致靶基因的表达降低。与一种或多种染色质修饰蛋白融合的一种或多种eiCas9可以用于改变染色质状态。

[0326] 在某些实施方式中,eiCas9介导的对所述RS1、RL2和/或LAT基因中的一个或多个的表达的减少导致所述RS1、RL2和/或LAT RNA转录的减少和/或停止。在某些实施方式中,eiCas9介导的对所述RS1、RL2和/或LAT基因中的一个或多个的表达的减少导致由所述RS1、RL2和/或LAT基因编码的HSV-1或HSV-2蛋白(例如,ICP0蛋白和/或LAT蛋白和/或转录调节因子ICP4蛋白)的翻译的减少和/或停止。

[0327] 在某些实施方式中,eiCas9介导的对所述RS1、RL2和/或LAT基因中的一个或多个的表达的减少单独或组合产生以下中的任何现象:HSV DNA产生减少、HSV脱落减少、HSV复制减少、病毒感染力降低、病毒颗粒包装减少、病毒蛋白例如ICP0蛋白例如转录调节因子ICP4蛋白产生的减少。

[0328] 在某些实施方式中,敲低所述RS1、RL2和/或LAT基因中的一个或多个治愈HSV-1或HSV-2感染。在某些实施方式中,敲低所述RS1、RL2和/或LAT基因中的一个或多个导致HSV-1或HSV-2感染的功能性治愈。在某些实施方式中,敲低所述RS1、RL2和/或LAT基因中的一个或多个导致HSV-1或HSV-2感染的持续病毒学应答。急性发作期间的靶向敲低方法可减少病毒脱落、复制,这导致炎症减少,这可减少对眼睛的伤害。在某些实施方式中,如本发明所述,所述eiCas9分子可以是Cas9变体。例如,但不作为限制,所述Cas9变体可以是化脓链球菌Cas9变体或金黄色葡萄球菌Cas变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。

[0329] 5.向导RNA(gRNA)分子

[0330] 如该术语在本文使用的,gRNA分子是指促进gRNA分子/Cas9分子复合物向靶核酸特异性靶向或归巢的核酸。gRNA分子可以是单分子的(具有单个RNA分子)(例如,嵌合)、或模块化的(包含多于一种并且典型地两种分开的RNA分子)。本发明提供的gRNA分子包括靶向结构域,所述靶向结构域包括与靶标结构域(也称作“靶序列”)完全或部分互补的核苷酸序列、由其组成或基本由其组成。在某些实施方式中,所述gRNA分子进一步包含一个或多个另外的结构域,包括例如第一互补结构域、连接结构域、第二互补结构域、近端结构域、尾部结构域和5’延伸结构域。下面详细讨论了这些结构域中的每一个。在某些实施方式中,所述gRNA分子中的一个或多个结构域包括与例如来自化脓链球菌、金黄色葡萄球菌或嗜热链球菌的天然存在的序列相同或与其共享序列同源性的核苷酸序列。在某些实施方式中,所述gRNA分子中的一个或多个结构域包括与例如来自化脓链球菌或金黄色葡萄球菌的天然存在的序列相同或与其共享序列同源性的核苷酸序列。

[0331] 图1A-1I中提供了若干示例性gRNA结构。关于gRNA的三维形式、或活化形式的链内或链间相互作用,高度互补的区域在图1A-1I和本文提供的其他描绘中有时显示为双链体。图7说明了使用SEQ ID NO:42的gRNA序列的gRNA结构域命名法,所述gRNA序列在tracrRNA衍生区域中含有一个发夹环。在某些实施方式中,gRNA可以在该区域中含有多个(例如,两个、三个或更多个)发夹环(参见例如,图1H-1I)。

[0332] 在某些实施方式中,单分子的或嵌合的gRNA包含,优选地从5’到3’:

[0333] 靶向结构域,所述靶向结构域与RL2、LAT或RS1基因中的靶标结构域互补,例如,包括选自SEQ ID NO:208至58749中的核苷酸序列的靶向结构域;

[0334] 第一互补结构域;

[0335] 连接结构域;

[0336] 第二互补结构域(其与所述第一互补结构域互补);

[0337] 近端结构域;以及

[0338] 任选地,尾部结构域。

[0339] 在某些实施方式中,模块化gRNA包含:

[0340] 第一链,其包含,优选地从5’到3’:

[0341] 靶向结构域,所述靶向结构域与RL2、LAT或RS1基因中的靶标结构域互补,例如,包括选自SEQ ID NO:208至58749中的核苷酸序列的靶向结构域;

[0342] 第一互补结构域;以及

[0343] 第二链,其包含,优选地从5’到3’:

- [0344] 任选地,5'延伸结构域;
- [0345] 第二互补结构域;
- [0346] 近端结构域;和
- [0347] 任选地,尾部结构域。

#### [0348] 5.1靶向结构域

[0349] 靶向结构域(有时可替代地称为指导序列)包括与RL2、LAT或RS1基因中的靶核酸序列互补或部分互补的核酸序列,由其组成或基本上由其组成。全部或部分靶向结构域与其互补或部分互补的RL2、LAT或RS1基因中的核酸序列在本文中称为靶标结构域。

[0350] 用于选择靶向结构域的方法本领域是已知的(参见例如,Fu 2014; Sternberg 2014)。适用于本发明所述的方法、组合物和试剂盒的靶向结构域的实例包括如SEQ ID NO:208 to 58749中所述的核苷酸序列。

[0351] 包含靶标结构域的靶核酸的链在本文中称为互补链,因为其与靶向结构域序列互补。由于靶向结构域是gRNA分子的一部分,所以它包含碱基尿嘧啶(U)而非胸腺嘧啶(T);相反,编码gRNA分子的任何DNA分子可包括胸腺嘧啶而非尿嘧啶。在靶向结构域/靶标结构域对中,靶向结构域中的尿嘧啶碱基将与靶标结构域中的腺嘌呤碱基配对。在某些实施方式中,靶向结构域与靶结构域之间的互补程度足以允许将Cas9分子靶向靶核酸。

[0352] 在某些实施方式中,靶向结构域包括核心结构域和任选的第二结构域。在这些实施方式的某些中,核心结构域位于第二结构域的3'处,并且在这些实施方式的某些中,核心结构域位于靶向结构域的3'端或其附近。在这些实施方式的某些中,核心结构域由靶向结构域的3'端处的约8至约13个核苷酸组成或基本上由其组成。在某些实施方式中,只有核心结构域与靶标结构域的相应部分互补或部分互补,并且在这些实施方式的某些中,核心结构域与靶标结构域的相应部分完全互补。在某些实施方式中,第二结构域也与靶标结构域的一部分互补或部分互补。在某些实施方式中,核心结构域与靶标结构域中的核心结构域靶标互补或部分互补,而第二结构域与靶标结构域中的第二靶标结构域互补或部分互补。在某些实施方式中,核心结构域和第二结构域与靶标结构的它们各自对应的部分具有相同的互补程度。在某些实施方式中,核心结构域与其靶标之间的互补程度和第二结构域与其靶标之间的互补程度可以不同。在这些实施方式的某些中,核心结构域可以对其靶标具有比第二结构域更高的互补程度,而在其他实施方式中,第二结构域可以具有比核心结构域更高的互补程度。

[0353] 在某些实施方式中,靶向结构域和/或靶向结构域内的核心结构域的长度为3至100、5至100、10至100或20至100个核苷酸,并且在这些实施方式的某些中,靶向结构域或核心结构域的长度为3至15、3至20、5至20、10至20、15至20、5至50、10至50或20至50个核苷酸。在某些实施方式中,靶向结构域和/或靶向结构域内的核心结构域的长度为6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个核苷酸。在某些实施方式中,靶向结构域和/或靶向结构域内的核心结构域的长度为6+/-2、7+/-2、8+/-2、9+/-2、10+/-2、10+/-4、10+/-5、11+/-2、12+/-2、13+/-2、14+/-2、15+/-2、或16+/-2、20+/-5、30+/-5、40+/-5、50+/-5、60+/-5、70+/-5、80+/-5、90+/-5、或100+/-5个核苷酸。

[0354] 在靶向结构域包括核心结构域的某些实施方式中,核心结构域的长度为3至20个核苷酸,并且在这些实施方式的某些中,核心结构域的长度为5至15或8至13个核苷酸。在靶

向结构域包括第二结构域的某些实施方式中,第二结构域的长度为0、1、2、3、4、5、6、7、8、9、10、11、12、13、14或15个核苷酸。在靶向结构域包含长度为8至13个核苷酸的核心结构域的某些实施方式中,各自地,靶向结构域的长度为26、25、24、23、22、21、20、19、18、17、或16个核苷酸,并且第二结构域的长度为13至18、12至17、11至16、10至15、9至14、8至13、7至12、6至11、5至10、4至9、或3至8个核苷酸。

[0355] 在某些实施方式中,靶向结构域与靶结构域完全互补。同样地,在靶向结构域包含核心结构域和/或第二结构域的情况下,在某些实施方式中,核心结构域和第二结构域中的一个或两个与靶标结构域的相应部分完全互补。在某些实施方式中,靶向结构域与靶标结构域部分互补,并且在靶向结构域包含核心结构域和/或第二结构域的这些实施方式的某些中,核心结构域和第二结构域中的一个或两个与靶标结构域的相应部分部分互补。在这些实施方式的某些中,靶向结构域或靶向结构域内的核心结构域与靶标结构域或靶标结构域的相应部分至少约80%、约85%、约90%、或约95%互补。在某些实施方式中,靶向结构域和/或靶向结构域内的核心或第二结构域包括与靶标结构域或其部分不互补的一个或多个核苷酸,并且在这些实施方式的某些中,靶向结构域和/或靶向结构域内的核心或第二结构域包括与靶标结构域不互补的1、2、3、4、5、6、7或8个核苷酸。在某些实施方式中,核心结构域包括与靶标结构域的相应部分不互补的1、2、3、4或5个核苷酸。在靶向结构域包括与靶标结构域不互补的一个或多个核苷酸的某些实施方式中,所述非互补核苷酸中的一个或多个位于靶向结构域的5'或3'端的五个核苷酸内。在这些实施方式的某些中,靶向结构域包括在其5'端、3'端或其5'和3'端的五个核苷酸内的与靶标结构域不互补的1、2、3、4或5个核苷酸。在靶向结构域包括与靶标结构域不互补的两个或更多个核苷酸的某些实施方式中,所述非互补核苷酸中的两个或更多个彼此相邻,并且在这些实施方式的某些中,所述两个或更多个连续的非互补核苷酸位于靶向结构域的5'或3'端的五个核苷酸内。在某些实施方式中,所述两个或更多个连续的非互补核苷酸都位于离靶向结构域的5'和3'端超过五个核苷酸处。

[0356] 在某些实施方式中,靶向结构域、核心结构域和/或第二结构域不包含任何修饰。在某些实施方式中,靶向结构域、核心结构域和/或第二结构域或其中的一个或多个核苷酸具有修饰,包括但不限于以下阐述的修饰。在某些实施方式中,靶向结构域、核心结构域和/或第二结构域的一个或多个核苷酸可以包含2'修饰(例如,在核糖上2'位置处的修饰),例如2-乙酰化,例如2'甲基化。在某些实施方式中,可以用硫代磷酸酯修饰靶向结构域的骨架。在某些实施方式中,对靶向结构域、核心结构域和/或第二结构域的一个或多个核苷酸的修饰使得靶向结构域和/或包含靶向结构域的gRNA不易降解或更加生物相容,例如具有更低的免疫原性。在某些实施方式中,靶向结构域和/或核心或第二结构域包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施方式的某些中,靶向结构域和/或核心或第二结构域包括它们各自5'端的五个核苷酸内的1、2、3或4个修饰,和/或它们各自3'端的五个核苷酸内的1、2、3或4个修饰。在某些实施方式中,靶向结构域和/或核心或第二结构域包括在两个或更多个连续核苷酸处的修饰。

[0357] 在靶向结构域包括核心和第二结构域的某些实施方式中,核心和第二结构域含有相同数量的修饰。在这些实施方式的某些中,这两个结构域都不含修饰。在其他实施方式中,核心结构域包括比第二结构域更多的修饰,或反之亦然。在某些实施方式中,选择对靶

向结构域(包括核心或第二结构域)中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过使用如下阐述的系统测试候选修饰来评价。具有候选靶向结构域的gRNA可以使用如下阐述的系统进行评价,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/Cas9分子系统中并且进行评价。

[0358] 在某些实施方式中,全部的修饰核苷酸互补于并且能够杂交到靶标结构域中存在的相应核苷酸上。在某些实施方式中,1、2、3、4、5、6、7或8个或更多个修饰核苷酸不互补于或不能够杂交到靶标结构域中存在的相应核苷酸上。

[0359] 5.2第一和第二互补结构域

[0360] 第一和第二互补(有时可替代地分别称为crRNA衍生的发夹序列和tracrRNA衍生的发夹序列)结构域互相完全或部分互补。在某些实施方式中,互补程度足以使所述两个结构域在至少一些生理条件下形成双链体区域。在某些实施方式中,第一与第二互补结构域之间的互补程度与gRNA的其他特性一起足以允许Cas9分子靶向靶核酸。第一和第二互补结构域的示例在图1A-1G中进行了阐述。

[0361] 在某些实施方式中(参见例如,图1A-1B),第一和/或第二互补结构域包括与相应互补结构域缺乏互补性的一个或多个核苷酸。在某些实施方式中,第一和/或第二互补结构域包括不与相应互补结构域互补的1、2、3、4、5或6个核苷酸。例如,第二互补结构域可以含有与第一互补结构域中的相应核苷酸不配对的1、2、3、4、5或6个核苷酸。在某些实施方式中,第一或第二互补结构域上不与相应互补结构域互补的核苷酸从在第一与第二互补结构域之间形成的双链体环出。在这些实施方式的某些中,未配对环出位于第二互补结构域上,并且在这些实施方式的某些中,未配对区域从离第二互补结构域的5'端1、2、3、4、5或6个核苷酸处开始。

[0362] 在某些实施方式中,第一互补结构域的长度为5至30、5至25、7至25、5至24、5至23、7至22、5至22、5至21、5至20、7至18、7至15、9至16、或10至14个核苷酸,并且在这些实施方式的某些中,第一互补结构域的长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、或25个核苷酸。在某些实施方式中,第二互补结构域的长度为5至27、7至27、7至25、5至24、5至23、5至22、5至21、7至20、5至20、7至18、7至17、9至16、或10至14个核苷酸,并且在这些实施方式的某些中,第二互补结构域的长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个核苷酸。在某些实施方式中,第一和第二互补结构域的长度各自独立地为6+/-2、7+/-2、8+/-2、9+/-2、10+/-2、11+/-2、12+/-2、13+/-2、14+/-2、15+/-2、16+/-2、17+/-2、18+/-2、19+/-2、或20+/-2、21+/-2、22+/-2、23+/-2、或24+/-2个核苷酸。在某些实施方式中,第二互补结构域长于第一互补结构域(例如,长出2、3、4、5、或6个核苷酸)。

[0363] 在某些实施方式中,第一和/或第二互补结构域各自独立地包含三个亚结构域,其按5'到3'方向是:5'亚结构域、中央亚结构域、和3'亚结构域。在某些实施方式中,第一互补结构域的5'亚结构域和3'亚结构域分别与第二互补结构域的3'亚结构域和5'亚结构域完全或部分互补。

[0364] 在某些实施方式中,第一互补结构域的5'亚结构域的长度为4至9个核苷酸,并且在这些实施方式的某些中,5'结构域的长度为4、5、6、7、8或9个核苷酸。在某些实施方式中,

第二互补结构域的5'亚结构域的长度为3至25、4至22、4至18、或4至10个核苷酸，并且在这些实施方式的某些中，5'结构域的长度为3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、或25个核苷酸。在某些实施方式中，第一互补结构域的中央亚结构域的长度为1、2或3个核苷酸。在某些实施方式中，第二互补结构域的中央亚结构域的长度为1、2、3、4或5个核苷酸。在某些实施方式中，第一互补结构域的3'亚结构域的长度为3至25、4至22、4至18、或4至10个核苷酸，并且在这些实施方式的某些中，3'亚结构域的长度为3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸。在某些实施方式中，第二互补结构域的3'亚结构域的长度为4至9个(例如4、5、6、7、8或9个)核苷酸。

[0365] 第一和/或第二互补结构域可以与天然存在的或参考的第一和/或第二互补结构域共享同源性，或者从其衍生。在这些实施方式的某些中，第一和/或第二互补结构域与天然存在的或参考的第一和/或第二互补结构域具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性或与其相差不多于1、2、3、4、5或6个核苷酸。在这些实施方式的某些中，第一和/或第二互补结构域可以与来自化脓链球菌或金黄色葡萄球菌的第一和/或第二互补结构域具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性。

[0366] 在某些实施方式中，第一和/或第二互补结构域不包含任何修饰。在其他实施方式中，第一和/或第二互补结构域或其中的一个或多个核苷酸具有修饰，包括但不限于以下阐述的修饰。在某些实施方式中，第一和/或第二互补结构域的一个或多个核苷酸可以包含2'修饰(例如在核糖上的2'位置处的修饰)，例如2-乙酰化，例如2'甲基化。在某些实施方式中，可以用硫代磷酸酯修饰靶向结构域的骨架。在某些实施方式中，对第一和/或第二互补结构域的一个或多个核苷酸的修饰使得第一和/或第二互补结构域和/或包含第一和/或第二互补结构域的gRNA不易降解或更具有生物相容，例如具有更低的免疫原性。在某些实施方式中，第一和/或第二互补结构域各自独立地包括1、2、3、4、5、6、7或8个或更多个修饰，并且在这些实施方式的某些中，第一和/或第二互补结构域各自独立地包括它们各自的5'端、3'端或它们的5'和3'端的五个核苷酸内的1、2、3或4个修饰。在某些实施方式中，第一和/或第二互补结构域各自独立地在它们各自的5'端、3'端或它们的5'和3'端的五个核苷酸内不含修饰。在某些实施方式中，第一和第二互补结构域中的一个或两个包含在两个或更多个连续核苷酸处的修饰。

[0367] 在某些实施方式中，选择对第一和/或第二互补结构域中的一个或多个核苷酸的修饰以不干扰靶向功效，这可以通过在如下阐述的系统中测试候选修饰来评价。具有候选第一和第二互补结构域的gRNA可以在如下阐述的系统中进行评价，所述第一和第二互补结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选互补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/Cas9分子系统中并且进行评价。

[0368] 在某些实施方式中，由第一和第二互补结构域形成的双链体区域的长度为例如6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21或22bp，排除任何环出的或未配对的核苷酸。

[0369] 在某些实施方式中，当成双链体时，第一和第二互补结构域包含11个配对的核苷酸(参见例如，SEQ ID NO:48的gRNA)。在某些实施方式中，当成双链体时，第一和第二互补

结构域包含15个配对的核苷酸(参见例如,SEQ ID NO:50的gRNA)。在某些实施方式中,当成双链体时,第一和第二互补结构域包含16个配对的核苷酸(参见例如,SEQ ID NO:51的gRNA)。在某些实施方式中,当成双链体时,第一和第二互补结构域包含21个配对的核苷酸(参见例如,SEQ ID NO:29的gRNA)。在某些实施方式中,在第一与第二互补结构域之间交换一个或多个核苷酸以去除聚-U束。例如,可以交换SEQ ID NO:48的gRNA的核苷酸23和48或核苷酸26和45,以分别产生SEQ ID NO:49或31的gRNA。类似地,SEQ ID NO:29的gRNA的核苷酸23和39可以与核苷酸50和68交换以产生SEQ ID NO:30的gRNA。

[0370] 5.3连接结构域

[0371] 连接结构域被布置在单分子gRNA或嵌合gRNA中的第一和第二互补结构域之间并用于连接第一和第二互补结构域。图1B-1E提供了连接结构域的实例。在某些实施方式中,连接结构域的一部分来自crRNA衍生区域,并且另一部分来自tracrRNA衍生区域。

[0372] 在某些实施方式中,连接结构域共价连接第一和第二互补结构域。在这些实施方式的某些中,连接结构域由共价键组成或包含共价键。在其他实施方式中,连接结构域非共价连接第一和第二互补结构域。在某些实施方式中,连接结构域的长度为十个或更少的核苷酸,例如1、2、3、4、5、6、7、8、9或10个核苷酸。在其他实施方式中,连接结构域的长度为大于10个核苷酸,例如11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个或更多个核苷酸。在某些实施方式中,连接结构域的长度为2至50、2至40、2至30、2至20、2至10、2至5、10至100、10至90、10至80、10至70、10至60、10至50、10至40、10至30、10至20、10至15、20至100、20至90、20至80、20至70、20至60、20至50、20至40、20至30或20至25个核苷酸。在某些实施方式中,连接结构域的长度为10+/-5、20+/-5、20+/-10、30+/-5、30+/-10、40+/-5、40+/-10、50+/-5、50+/-10、60+/-5、60+/-10、70+/-5、70+/-10、80+/-5、80+/-10、90+/-5、90+/-10、100+/-5或100+/-10个核苷酸。

[0373] 在某些实施方式中,连接结构域与天然存在的序列(例如,对所述第二互补结构域是5'的tracrRNA的序列)共享同源性,或从其衍生。在某些实施方式中,连接结构域与本文所披露的连接结构域(例如,图1B-1E的连接结构域)具有至少约50%、约60%、约70%、约80%、约90%或约95%的同源性或与其相差不多于1、2、3、4、5或6个核苷酸。

[0374] 在某些实施例中,连接结构域不包含任何修饰。在其他实施例中,连接结构域或其中的一个或多个核苷酸具有修饰,包括但不限于以下阐述的修饰。在某些实施例中,连接结构域的一个或多个核苷酸可以包含2'修饰(例如在核糖上的2'位置处的修饰),例如2-乙酰化,例如2'甲基化。在某些实施例中,可以用硫代磷酸酯修饰连接结构域的骨架。在某些实施例中,对连接结构域的一个或多个核苷酸的修饰使得连接结构域和/或包含连接结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,连接结构域包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,连接结构域包括其5'和/或3'端的五个核苷酸内的1、2、3或4个修饰。在某些实施例中,连接结构域包含在两个或更多个连续核苷酸处的修饰。

[0375] 在某些实施方式中,选择对连接结构域中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过在如下阐述的系统中测试候选修饰来评价。具有候选连接结构域的gRNA可以在如下阐述的系统中进行评价,所述候选连接结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选连接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/Cas9分子系统中并且进行评价。

[0376] 在某些实施例中,连接结构域包含典型地邻近于第一互补结构域的3'端和/或第二互补结构域的5'端或在其1、2或3个核苷酸内的双链体区域。在这些实施例的某些中,连接区域的双链体区域的长度为10+/-5、15+/-5、20+/-5、20+/-10或30+/-5bp。在某些实施例中,连接结构域的双链体区域的长度为1、2、3、4、5、6、7、8、9、10、11、12、13、14或15bp。在某些实施例中,形成连接结构域的双链体区域的序列是完全互补的。在其他实施例中,形成双链体区域的一个或两个序列含有与其他双链体序列不互补的一个或多个核苷酸(例如,1、2、3、4、5、6、7或8个核苷酸)。

[0377] 5.4 5'延伸结构域

[0378] 在某些实施方式中,如本文所披露的模块化gRNA包含5'延伸结构域,即第二互补结构域的5'的一个或多个另外的核苷酸(参见例如,图1A)。在某些实施方式中,5'延伸结构域的长度为2至10或更多、2至9、2至8、2至7、2至6、2至5或2至4个核苷酸,并且在这些实施方式的某些中,5'延伸结构域的长度为2、3、4、5、6、7、8、9或10个或更多个核苷酸。

[0379] 在某些实施方式中,5'延伸结构域核苷酸不包含修饰,例如以下提供的类型的修饰。然而,在某些实施方式中,5'延伸结构域包含一个或多个修饰,例如,使其较不易降解或更生物相容(例如,更低的免疫原性)的修饰。作为举例,5'延伸结构域的骨架可以用硫代磷酸酯、或如下阐述的其他一个或多个修饰来修饰。在某些实施方式中,5'延伸结构域的核苷酸可以包含2'修饰(例如,在核糖上的2'位置处的修饰),例如2-乙酰化,例如2'甲基化,或如下阐述的其他一个或多个修饰。

[0380] 在某些实施方式中,5'延伸结构域可以包含多达1、2、3、4、5、6、7或8个修饰。在某些实施方式中,5'延伸结构域包含其5'端的5个核苷酸内的多达1、2、3或4个修饰,例如在模块化gRNA分子中。在某些实施方式中,5'延伸结构域包含其3'端的5个核苷酸内的多达1、2、3或4个修饰,例如在模块化gRNA分子中。

[0381] 在某些实施方式中,5'延伸结构域包含在两个连续核苷酸处的修饰,例如5'延伸结构域的5'端的5个核苷酸内、5'延伸结构域的3'端的5个核苷酸内、或远离5'延伸结构域的一端或两端超过5个核苷酸的两个连续核苷酸。在某些实施方式中,在5'延伸结构域的5'端的5个核苷酸内、5'延伸结构域的3'端的5个核苷酸内、或在远离5'延伸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。在某些实施方式中,在5'延伸结构域的5'端的5个核苷酸内、5'延伸结构域的3'端的5个核苷酸内、或在远离5'延伸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。

[0382] 可以选择5'延伸结构域中的修饰以便不干扰gRNA分子功效,这可以通过在如下阐述的系统中测试候选修饰来评价。具有候选5'延伸结构域的gRNA可以在如下阐述的系统中进行评价,所述候选5'延伸结构域具有选定的长度、序列、互补程度、或修饰程度。候选5'延伸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/Cas9分子系统中并且进行评价。

[0383] 在某些实施方式中,5'延伸结构域与参考5'延伸结构域(例如天然存在的(例如,化脓链球菌、金黄色葡萄球菌或嗜热链球菌)5'延伸结构域)、或本文所述的5'延伸结构域(例如,来自图1A-1G))具有至少约60%、约70%、约80%、约85%、约90%或约95%的同源性,或与其相差不多于1、2、3、4、5、或6个核苷酸。

[0384] 5.5近端结构域

[0385] 图1A-1G提供了近端结构域的实例。

[0386] 在某些实施例中,近端结构域的长度为5至20个或更多个核苷酸,例如长度为5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25或26个核苷酸。在这些实施例的某些中,近端结构域的长度为6+/-2、7+/-2、8+/-2、9+/-2、10+/-2、11+/-2、12+/-2、13+/-2、14+/-2、14+/-2、16+/-2、17+/-2、18+/-2、19+/-2、或20+/-2个核苷酸。在某些实施例中,近端结构域的长度为5至20、7至18、9至16、或10至14个核苷酸。

[0387] 在某些实施方式中,近端结构域可以与天然存在的近端结构域共享同源性,或从其衍生。在这些实施方式的某些中,近端结构域与本文所披露的近端结构域(例如化脓链球菌、金黄色葡萄球菌或嗜热链球菌近端结构域,包括图1A-1G所阐述的那些)具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性,或与其相差不多于1、2、3、4、5或6个核苷酸。

[0388] 在某些实施例中,近端结构域不包含任何修饰。在其他实施例中,近端结构域或其中的一个或多个核苷酸具有修饰,包括但不限于本文所阐述的修饰。在某些实施例中,近端结构域的一个或多个核苷酸可以包含2'修饰(例如在核糖上的2'位置处的修饰),例如2-乙酰化,例如2'甲基化。在某些实施例中,可以用硫代磷酸酯修饰近端结构域的骨架。在某些实施例中,对近端结构域的一个或多个核苷酸的修饰使得近端结构域和/或包含近端结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,近端结构域包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,近端结构域包括其5'和/或3'端的五个核苷酸内的1、2、3或4个修饰。在某些实施例中,近端结构域包含在两个或更多个连续核苷酸处的修饰。

[0389] 在某些实施例中,选择对近端结构域中的一个或多个核苷酸的修饰以不干扰靶向功效,这可以通过在如下阐述的系统中测试候选修饰来评价。具有候选近端结构域的gRNA可以在如下阐述的系统中进行评价,所述候选近端结构域具有选定的长度、序列、互补程度、或修饰程度。所述候选近端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/Cas9分子系统中并且进行评价。

[0390] 5.6尾部结构域

[0391] 广谱的尾部结构域适于在本文所披露的gRNA分子中使用。图1A和1C-1G提供了这种尾部结构域的实例。

[0392] 在某些实施例中,不存在尾部结构域。在其他实施例中,尾部结构域的长度为1至100个或更多个核苷酸,例如长度为1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90或100个核苷酸。在某些实施例中,尾部结构域的长度为1至5、1至10、1至15、1至20、1至50、10至100、20至100、10至90、20至90、10至80、20至80、10至70、20至70、10至60、20至60、10至50、20至50、10至40、20至40、10至30、20至30、20至25、10至20或10至15个核苷酸。在某些实施例中,尾部结构域的长度为5+/-5、10+/-5、20+/-10、20+/-5、25+/-10、30+/-10、30+/-5、40+/-10、40+/-5、50+/-10、50+/-5、60+/-10、60+/-5、70+/-10、70+/-5、80+/-10、80+/-5、90+/-10、90+/-5、100+/-10或100+/-5个核苷酸。

[0393] 在某些实施方式中,尾部结构域可以与天然存在的尾部结构域或天然存在的尾部结构域的5'端共享同源性,或从其衍生。在这些实施方式的某些中,近端结构域与本文所披露

露的天然形成的尾部结构域(例如化脓链球菌、金黄色葡萄球菌或嗜热链球菌尾部结构域,包括图1A和图1C-1G所阐述的那些)具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性,或与其相差不多于1、2、3、4、5或6个核苷酸。

[0394] 在某些实施例中,尾部结构域包括彼此互补,并且在至少一些生理条件下形成双链体区域的序列。在这些实施例的某些中,尾部结构域包含尾部双链体结构域,其可以形成尾部双链体区域。在某些实施例中,尾部双链体区域的长度为3、4、5、6、7、8、9、10、11或12bp。在某些实施例中,尾部结构域包含不形成双链体的尾部双链体结构域的3'的单链结构域。在这些实施例的某些中,单链结构域的长度为3至10个核苷酸(例如3、4、5、6、7、8、9、10个)或长度为4至6个核苷酸。

[0395] 在某些实施例中,尾部结构域不包含任何修饰。在其他实施例中,尾部结构域或其中的一个或多个核苷酸具有修饰,包括但不限于本文所阐述的修饰。在某些实施例中,尾部结构域的一个或多个核苷酸可以包含2'修饰(例如在核糖上的2'位置处的修饰),例如2-乙酰化,例如2'甲基化。在某些实施例中,可以用硫代磷酸酯修饰尾部结构域的骨架。在某些实施例中,对尾部结构域的一个或多个核苷酸的修饰使得尾部结构域和/或包含尾部结构域的gRNA不易降解或更生物相容,例如更低的免疫原性。在某些实施例中,尾部结构域包括1、2、3、4、5、6、7或8个或更多个修饰,并且在这些实施例的某些中,尾部结构域包括其5'和/或3'端的五个核苷酸内的1、2、3或4个修饰。在某些实施例中,尾部结构域包含在两个或更多个连续核苷酸处的修饰。

[0396] 在某些实施方式中,尾部结构域包括在3'端与体外或体内转录方法相关的核苷酸。当将T7启动子用于gRNA的体外转录时,这些核苷酸可以是DNA模板的3'端前存在的任何核苷酸。在某些实施方式中,gRNA分子包括通过从DNA模板体外转录而制备的3'聚腺苷酸尾部。在某些实施方式中,gRNA分子的靶向结构域的5'核苷酸是鸟嘌呤核苷酸,DNA模板包含位于与靶向结构域对应的序列的紧上游的T7启动子序列,且T7启动子序列的3'核苷酸不是鸟嘌呤核苷酸。在某些实施方式中,gRNA分子的靶向结构域的5'核苷酸不是鸟嘌呤核苷酸,DNA模板包含位于与靶向结构域对应的序列的紧上游的T7启动子序列,且T7启动子序列的3'核苷酸是位于鸟嘌呤核苷酸以外的核苷酸下游的鸟嘌呤核苷酸。

[0397] 在某些实施例中,尾部结构域包括在3'端与体外或体内转录方法相关的核苷酸。当将T7启动子用于gRNA的体外转录时,这些核苷酸可以是DNA模板的3'端前存在的任何核苷酸。当将U6启动子用于体内转录时,这些核苷酸可以是序列UUUUUU。当将H1启动子用于转录时,这些核苷酸可以是序列UUUU。当使用替代的pol-III启动子时,这些核苷酸可以是各种数量的尿嘧啶碱基,这取决于例如pol-III启动子的终止信号,或者它们可以包括替代碱基。

[0398] 在某些实施例中,所述近端结构域和尾部结构域一起包含SEQ ID N0:32、33、34、35、36或37所示的序列,由其组成或基本上由其组成。

[0399] 5.7示例性单分子/嵌合gRNA

[0400] 在某些实施方式中,本发明所述gRNA具有以下结构:5' [靶向结构域]-[第一互补结构域]-[连接结构域]-[第二互补结构域]-[近端结构域]-[尾部结构域]-3',其中:

[0401] 靶向结构域包含核心结构域和任选地第二结构域,并且长度为10至50个核苷酸;

[0402] 第一互补结构域的长度为5至25个核苷酸,并且在某些实施方式中,与本文所披露

的参考第一互补结构域具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性；

[0403] 连接结构域的长度为1至5个核苷酸；

[0404] 第二互补结构域的长度为5至27个核苷酸，并且在某些实施方式中，与本文所披露的参考第二互补结构域具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性；

[0405] 近端结构域的长度为5至20个核苷酸，并且在某些实施方式中，与本文所披露的参考近端结构域具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性；以及

[0406] 尾部结构域的长度为1至50个核苷酸，并且在某些实施方式中，与本文所披露的参考尾部结构域具有至少约50%、约60%、约70%、约80%、约85%、约90%或约95%的同源性。

[0407] 在某些实施方式中，来自(a)、(b)、和/或(c)的序列与天然存在的gRNA的相应序列或与本文所述的gRNA具有至少约50%、约60%、约70%、约75%、约60%、约70%、约80%、约85%、约90%、约95%或约99%的同源性。

[0408] 在某些实施方式中，如本文所披露的单分子gRNA包含，优选地从5'到3'：

[0409] 靶向结构域，其例如包含10-50个核苷酸；

[0410] 第一互补结构域，其例如包含15、16、17、18、19、20、21、22、23、

[0411] 24、25、或26个核苷酸；

[0412] 连接结构域；

[0413] 第二互补结构域；

[0414] 近端结构域；以及

[0415] 尾部结构域，

[0416] 其中，

[0417] (a)当一起考虑时，近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸；

[0418] (b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、

[0419] 30、31、35、40、45、49、50、或53个核苷酸；或

[0420] (c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、

[0421] 31、32、36、41、46、50、51、或54个核苷酸，所述核苷酸与第一互补结构域的相应核苷酸互补。

[0422] 在某些实施方式中，来自(a)、(b)、和/或(c)的序列与天然存在的gRNA的相应序列或与本文所述的gRNA具有至少50%、60%、70%、75%、80%、85%、90%、95%或99%的同源性。

[0423] 在某些实施方式中，当一起考虑时，近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸。

[0424] 在某些实施方式中，第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0425] 在某些实施方式中，第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、

26、31、32、36、41、46、50、51或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0426] 在某些实施方式中,靶向结构域由16、17、18、19、20、21、22、23、24、25或26个与靶结构域或其部分互补或部分互补的核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)组成,基本上由其组成或包含其,例如,靶向结构域的长度为16、17、18、19、20、21、22、23、24、25或26个核苷酸。在这些实施方式的某些中,靶向结构域在靶向结构域的整个长度、靶结构域的整个长度或两者上与靶结构域互补。

[0427] 在某些实施方式实施方式中,本文所披露的单分子的或嵌合的gRNA分子(包含靶向结构域、第一互补结构域、连接结构域、第二互补结构域、近端结构域以及任选地尾部结构域)包含SEQ ID NO:42所示的氨基酸序列,其中靶向结构域被列为20个N(残基1-20)但长度范围可以从16至26个核苷酸,并且其中最后的六个残基(残基97-102)表示U6启动子的终止信号,但可以不存在或数目更少。在某些实施方式实施方式中,所述单分子的或嵌合的gRNA分子是化脓链球菌gRNA分子。

[0428] 在某些实施方式中,本文所披露的单分子的或嵌合的gRNA分子(包含靶向结构域、第一互补结构域、连接结构域、第二互补结构域、近端结构域以及任选地尾部结构域)包含SEQ ID NO:38所示的氨基酸序列,其中靶向结构域被列为20个N(残基1-20)但长度范围可以从16至26个核苷酸,并且其中最后的六个残基(残基97-102)表示U6启动子的终止信号,但可以不存在或数目更少。在某些实施方式中,所述单分子的或嵌合的gRNA分子是金黄色葡萄球菌gRNA分子。

[0429] 示例性嵌合gRNA的序列和结构也示于图1H-1I中。

[0430] 5.8示例性模块化gRNA

[0431] 在某些实施方式中,本文所披露的模块化gRNA包含:

[0432] 第一链,其包含,优选地从5'到3':

[0433] 靶向结构域,其例如包含15、16、17、18、19、20、21、22、23、

[0434] 24、25、或26个核苷酸;

[0435] 第一互补结构域;以及

[0436] 第二链,其包含,优选地从5'到3':

[0437] 任选地,5'延伸结构域;

[0438] 第二互补结构域;

[0439] 近端结构域;和

[0440] 尾部结构域,

[0441] 其中:

[0442] (a)当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、

[0443] 30、31、35、40、45、49、50、或53个核苷酸;

[0444] (b)第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、

[0445] 30、31、35、40、45、49、50、或53个核苷酸;或

[0446] (c)第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、

[0447] 31、32、36、41、46、50、51、或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0448] 在某些实施方式中,来自(a)、(b)或(c)的序列与天然存在的gRNA的相应序列或与本文所述的gRNA具有至少约50%、约60%、约70%、约80%、约85%、约90%、约95%或约99%的同源性。

[0449] 在某些实施方式中,当一起考虑时,近端结构域和尾部结构域包含至少15、18、20、25、30、31、35、40、45、49、50、或53个核苷酸。

[0450] 在某些实施方式中,第二互补结构域的最后一个核苷酸的3'存在至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。在某些实施方式中,第二互补结构域的最后一个核苷酸的3'存在至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,所述核苷酸与第一互补结构域的相应核苷酸互补。

[0451] 在某些实施方式中,靶向结构域包含、具有、或由16、17、18、19、20、21、22、23、24、25或26个与靶标结构域具有互补性的核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)组成,例如,靶向结构域的长度为16、17、18、19、20、21、22、23、24、25或26个核苷酸。

[0452] 在某些实施方式中,靶向结构域由16、17、18、19、20、21、22、23、24、25或26个与靶标结构域或其部分互补的核苷酸(例如,16、17、18、19、20、21、22、23、24、25或26个连续核苷酸)组成,基本上由其组成或包含其。在这些实施方式的某些中,靶向结构域在靶向结构域的整个长度、靶标结构域的整个长度或两者上与靶标结构域互补。

[0453] 在某些实施方式中,靶向结构域包括、具有或由16个具有与靶标结构域的互补性的核苷酸(例如,16个连续核苷酸),例如,靶向结构域为16个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0454] 在某些实施方式中,靶向结构域包括、具有或由16个具有与靶标结构域的互补性的核苷酸(例如,16个连续核苷酸),例如,靶向结构域为16个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0455] 在某些实施方式中,靶向结构域包括、具有或由16个具有与靶标结构域的互补性的核苷酸(例如,16个连续核苷酸),例如,靶向结构域为16个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0456] 在某些实施方式中,靶向结构域具有或由17个具有与靶标结构域的互补性的核苷酸(例如,17个连续核苷酸),例如,靶向结构域为17个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0457] 在某些实施方式中,靶向结构域具有或由17个具有与靶标结构域的互补性的核苷酸(例如,17个连续核苷酸),例如,靶向结构域为17个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0458] 在某些实施方式中,靶向结构域具有或由17个具有与靶标结构域的互补性的核苷酸(例如,17个连续核苷酸),例如,靶向结构域为17个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0459] 在某些实施方式中,靶向结构域具有或由18个具有与靶标结构域的互补性的核苷酸(例如,18个连续核苷酸),例如,靶向结构域为18个核苷酸长度;当近端结构域和尾部结

构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0460] 在某些实施方式中,靶向结构域具有或由18个具有与靶标结构域的互补性的核苷酸(例如,18个连续核苷酸),例如,靶向结构域为18个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0461] 在某些实施方式中,靶向结构域具有或由18个具有与靶标结构域的互补性的核苷酸(例如,18个连续核苷酸),例如,靶向结构域为18个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0462] 在某些实施方式中,靶向结构域包括、具有或由19个具有与靶标结构域的互补性的核苷酸(例如,19个连续核苷酸),例如,靶向结构域为19个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0463] 在某些实施方式中,靶向结构域包括、具有或由19个具有与靶标结构域的互补性的核苷酸(例如,19个连续核苷酸),例如,靶向结构域为19个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0464] 在某些实施方式中,靶向结构域包括、具有或由19个具有与靶标结构域的互补性的核苷酸(例如,19个连续核苷酸),例如,靶向结构域为19个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0465] 在某些实施方式中,靶向结构域包括、具有或由20个具有与靶标结构域的互补性的核苷酸(例如,20个连续核苷酸),例如,靶向结构域为20个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0466] 在某些实施方式中,靶向结构域包括、具有或由20个具有与靶标结构域的互补性的核苷酸(例如,20个连续核苷酸),例如,靶向结构域为20个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0467] 在某些实施方式中,靶向结构域包括、具有或由20个具有与靶标结构域的互补性的核苷酸(例如,20个连续核苷酸),例如,靶向结构域为20个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0468] 在某些实施方式中,靶向结构域包括、具有或由21个具有与靶标结构域的互补性的核苷酸(例如,21个连续核苷酸),例如,靶向结构域为21个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0469] 在某些实施方式中,靶向结构域包括、具有或由21个具有与靶标结构域的互补性的核苷酸(例如,21个连续核苷酸),例如,靶向结构域为21个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0470] 在某些实施方式中,靶向结构域包括、具有或由21个具有与靶标结构域的互补性的核苷酸(例如,21个连续核苷酸),例如,靶向结构域为21个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0471] 在某些实施方式中,靶向结构域包括、具有或由22个具有与靶标结构域的互补性

的核苷酸(例如,22个连续核苷酸),例如,靶向结构域为22个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0472] 在某些实施方式中,靶向结构域包括、具有或由22个具有与靶标结构域的互补性的核苷酸(例如,22个连续核苷酸),例如,靶向结构域为22个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0473] 在某些实施方式中,靶向结构域包括、具有或由22个具有与靶标结构域的互补性的核苷酸(例如,22个连续核苷酸),例如,靶向结构域为22个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0474] 在某些实施方式中,靶向结构域包括、具有或由23个具有与靶标结构域的互补性的核苷酸(例如,23个连续核苷酸),例如,靶向结构域为23个核苷酸长度;当近端结构域和尾部结构域一起考虑,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0475] 在某些实施方式中,靶向结构域包括、具有或由23个具有与靶标结构域的互补性的核苷酸(例如,23个连续核苷酸),例如,靶向结构域为23个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0476] 在某些实施方式中,靶向结构域包括、具有或由23个具有与靶标结构域的互补性的核苷酸(例如,23个连续核苷酸),例如,靶向结构域为23个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0477] 在某些实施方式中,靶向结构域包括、具有或由24个具有与靶标结构域的互补性的核苷酸(例如,24个连续核苷酸),例如,靶向结构域为24个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0478] 在某些实施方式中,靶向结构域包括、具有或由24个具有与靶标结构域的互补性的核苷酸(例如,24个连续核苷酸),例如,靶向结构域为24个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0479] 在某些实施方式中,靶向结构域包括、具有或由24个具有与靶标结构域的互补性的核苷酸(例如,24个连续核苷酸),例如,靶向结构域为24个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0480] 在某些实施方式中,靶向结构域包括、具有或由25个具有与靶标结构域的互补性的核苷酸(例如,25个连续核苷酸),例如,靶向结构域为25个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0481] 在某些实施方式中,靶向结构域包括、具有或由25个具有与靶标结构域的互补性的核苷酸(例如,25个连续核苷酸),例如,靶向结构域为25个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0482] 在某些实施方式中,靶向结构域包括、具有或由25个具有与靶标结构域的互补性的核苷酸(例如,25个连续核苷酸),例如,靶向结构域为25个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0483] 在某些实施方式中,靶向结构域包括、具有或由26个具有与靶标结构域的互补性的核苷酸(例如,26个连续核苷酸),例如,靶向结构域为26个核苷酸长度;当近端结构域和尾部结构域一起考虑时,包括15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0484] 在某些实施方式中,靶向结构域包括、具有或由26个具有与靶标结构域的互补性的核苷酸(例如,26个连续核苷酸),例如,靶向结构域为26个核苷酸长度;且在第二互补结构域的最后一个核苷酸的3'至少有15、18、20、25、30、31、35、40、45、49、50或53个核苷酸。

[0485] 在某些实施方式中,靶向结构域包括、具有或由26个具有与靶标结构域的互补性的核苷酸(例如,26个连续核苷酸),例如,靶向结构域为26个核苷酸长度;且在与第一互补结构域中对应核苷酸互补的第二互补结构域中的最后一个核苷酸的3'至少有16、19、21、26、31、32、36、41、46、50、51或54个核苷酸。

[0486] 5.9gRNA递送

[0487] 在本发明提供的方法的某些实施方式中,所述方法包括如本发明所述的一个或多个(例如,两个、三个或四个)gRNA分子的递送。在这些实施方式中的某些中,gRNA分子通过静脉注射、肌内注射、皮下注射或吸入进行递送。在某些实施方式中,gRNA分子在基因组编辑系统中与Cas9分子一起递送。

[0488] 6.用于设计gRNA的方法

[0489] 提供用于选择、设计和验证本发明所述gRNA中所用的靶向结构域的方法。本发明还提供了用于掺入gRNA的示例性靶向结构域。

[0490] 先前已经描述了用于靶序列的选择和验证以及脱靶分析的方法(参见例如,Mali 2013;Hsu 2013;Fu 2014;Heigwer 2014;Bae 2014;Xiao 2014)。例如,软件工具可以用来优化与使用者的靶序列对应的潜在靶向结构域的选择,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。对于使用化脓链球菌Cas9的每个可能的靶向结构域选择,所述工具可以鉴别跨基因组所有脱靶序列(以上的NAG或NGG PAM),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。然后将每个可能的靶向结构域进行排序,根据其总的预测的脱靶切割;最高排名的靶向结构域表示可能具有最大中靶切割和最少脱靶切割的那些。其他功能(例如,用于CRISPR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。可以使用本领域已知和/或本文所阐述的方法对候选靶向结构域和包含那些靶向结构域的gRNA进行功能评价。

[0491] 作为非限制性实例,使用DNA序列检索算法鉴定用于与化脓链球菌、金黄色葡萄球菌和脑膜炎奈瑟球菌Cas9一起使用的gRNA中使用的靶向结构域。为化脓性链球菌和脑膜炎奈瑟氏球菌靶标设计了17聚体和20聚体靶向结构域,而18聚体、19聚体、20聚体、21聚体、22聚体、23聚体和24聚体靶向结构域设计用于金黄色葡萄球菌靶标。使用基于公共工具cas-offinder (Bae 2014)的定制gRNA设计软件进行gRNA设计。该软件在计算向导物的全基因组脱靶倾向之后为向导物打分。通常针对长度范围从17至24的向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,便计算每种向导物的总分并且使用web界面以表格输出总结。除鉴定与PAM序列邻近的潜在靶位点之外,所述软件还鉴定与所选靶位点相差1、2、3个或更多于3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，并使用公开可用的RepeatMasker程序针对重复元件对序列进行筛选。RepeatMasker检索输入DNA序列的重复元件和低复杂度区域。输出是存在于给定查询序列中的重复的详细注释。

[0492] 鉴定后，靶向结构域根据它们与靶位点的距离、它们的正交性和存在的5' G (基于鉴定包括相关PAM的人类基因组中的紧密匹配，例如，野生型化脓链球菌Cas9分子的NGG PAM) 而被分级；针对野生型金黄色葡萄球菌Cas9分子的NNGRRT (SEQ ID NO:204) 或NNGRRV (SEQ ID NO:205) PAM，或针对野生型脑膜炎奈瑟氏菌Cas9分子的NNNNGATT或NNNNGCTT PAM；针对化脓性链球菌Cas9 EQR变体，选自由NGAG、NGCG、NGGG、NGTG、NGAA、NGAT和NGAC组成的组的PAM；或选自由NGCG、NGCA、NGCT和NGCC组成的组的用于化脓链球菌Cas9 VRER变体的PAM)。正交性是指在人类基因组中含有与靶序列最小数目的错配的序列的数目。“高水平的正交性”或“良好的正交性”可以例如是指20聚体靶向结构域，其除预期靶标之外在人类基因组中既没有一致序列，又没有含有靶序列中的一个或两个错配的任何序列。具有良好正交性的靶向结构域被选择用于最小化脱靶DNA切割。

[0493] 针对单-gRNA核酸酶切割并且针对双-gRNA成对的“切口酶”策略两者对靶向结构域进行鉴定。用于选择靶向结构域并且确定哪些靶向结构域可以用于双-gRNA成对的“切口酶”策略的标准是基于两个考虑：

[0494] (1) 靶向结构域对应该在DNA上定向成使得PAM朝外并且用D10A Cas9切口酶切割将产生5'突出端；和

[0495] (2) 假设用双切口酶对切割将以合理频率导致整个插入序列的缺失。然而，用双切口酶对切割还可以仅在所述gRNA之一的位点处导致indel突变。

[0496] 可以针对它们如何有效地去除整个序列对比引起一个靶向结构域的靶位点处的indel突变对候选对成员进行测试。

[0497] 6.1用于敲除HSV-1 RS1基因的靶向结构域

[0498] 与本发明披露方法结合在gRNA中用于敲除HSV-1 RS1基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌，以及5级用于脑膜炎奈瑟氏球菌。

[0499] 对于化脓性链球菌和脑膜炎奈瑟氏球菌，1级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离，例如在靶位点(例如起始密码子)的500bp(例如，下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离，例如在靶位点(例如起始密码子)的500bp(例如，下游)内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离，例如在靶位点(例如起始密码子)的500bp(例如，下游)内和(2)5' G的存在来选择。4级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离，例如在靶位点(例如起始密码子)的500bp(例如，下游)内来选择。5级gRNA分子的靶向结构域基于与靶位点(例如起始密码子)的距离来选择，例如，在剩下的编码序列的范围内，例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)。

[0500] 对于金黄色葡萄球菌，1级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离，例如在靶位点(例如起始密码子)的500bp(例如，下游)内，(2)高水平的正交性、(3)5' G的存在以及(4)其中PAM是NNGRRT (SEQ ID NO:204) 来选择。2级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离，例如在靶位点(例如起始密码子)的

500bp (例如, 下游) 内, (2) 高水平的正交性以及 (3) 其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于 (1) 与靶位点 (例如起始密码子) 的距离, 例如在靶位点 (例如起始密码子) 的500bp (例如, 下游) 内以及 (2) 其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于 (1) 与靶位点 (例如起始密码子) 的距离, 例如在靶位点 (例如起始密码子) 的500bp (例如, 下游) 内以及 (2) 其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于 (1) 与靶位点 (例如起始密码子) 的距离来选择, 例如, 在剩下的编码序列的范围内, 例如在编码序列的第一个500bp的下游 (例如+500 (相对于起始密码子) 至终止密码子的任何地方) (2) 5' G的存在以及 (3) 其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于 (1) 与靶位点 (例如起始密码子) 的距离来选择, 例如, 在剩下的编码序列的范围内, 例如在编码序列的第一个500bp的下游 (例如+500 (相对于起始密码子) 至终止密码子的任何地方) 以及 (2) 其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于 (1) 与靶位点 (例如起始密码子) 的距离来选择, 例如, 在剩下的编码序列的范围内, 例如在编码序列的第一个500bp的下游 (例如+500 (相对于起始密码子) 至终止密码子的任何地方) 以及 (2) 其中PAM是NNGRRV来选择。请注意, 层级是非包容性的 (对该策略每个gRNA仅列出一次)。在某些示例中, 没有基于特定层级的标准来鉴定gRNA。注意, 层级是非包容性的 (对该策略每个靶向结构域仅列出一次)。在某些示例中, 没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表1。

[0501] 表1用于敲除HSV-1 RS1基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
<b>1 级</b>	SEQ ID NO:208 至 254	SEQ ID NO:2510 至 2531	SEQ ID NO:7074
<b>2 级</b>	SEQ ID NO:255 至 333	SEQ ID NO:2532 至 2563	SEQ ID NO:7075
<b>3 级</b>	SEQ ID NO:334 至 425	SEQ ID NO:2564 至 2586	不适用
<b>4 级</b>	SEQ ID NO:426 至 516	SEQ ID NO:2587 至 3244	不适用
<b>5 级</b>	SEQ ID NO:517 至 2509	SEQ ID NO:3245 至 3368	SEQ ID NO:7076 至 7097
<b>6 级</b>	不适用	SEQ ID NO:3369 至 3545	不适用
<b>7 级</b>	不适用	SEQ ID NO:3546 至 7073	不适用

[0504] 6.2用于敲除HSV-2 RS1基因的靶向结构域

[0505] 与本发明披露方法结合在gRNA中用于敲除HSV-2 RS1基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌, 以及5级用于脑膜炎奈瑟氏球菌。

[0506] 对于化脓性链球菌和脑膜炎奈瑟氏球菌, 1级gRNA分子的靶向结构域基于 (1) 与靶位点 (例如起始密码子) 的距离, 例如在靶位点 (例如起始密码子) 的500bp (例如, 下游) 内、

(2) 高水平的正交性和(3) 5' G的存在来选择。2级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内和(2) 高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内和(2) 5' G的存在来选择。4级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内来选择。5级gRNA分子的靶向结构域基于与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)。

[0507] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内,(2) 高水平的正交性、(3) 5' G的存在以及(4) 其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内,(2) 高水平的正交性以及(3) 其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内以及(2) 其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内以及(2) PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)(2) 5' G的存在以及(3) 其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子)以及(2) 其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1) 与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子)以及(2) 其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0508] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表2。

[0509] 表2用于敲除HSV-2 RS1基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0510]	1 级 SEQ ID NO:7098 至 7148	SEQ ID NO:9293 至 9314	SEQ ID NO:13615
	2 级 SEQ ID NO:7149 至 7236	SEQ ID NO:9315 至 9342	SEQ ID NO:13616 至 13618
	3 级 SEQ ID NO:7237 至 7286	SEQ ID NO:9343 至 9355	不适用
	4 级 SEQ ID NO:7287 至 7341	SEQ ID NO:9356 至 9911	不适用
	5 级 SEQ ID NO:7342 至 9292	SEQ ID NO:9912 至 10034	SEQ ID NO:13619 至 13636
	6 级 不适用	SEQ ID NO:10035 至 10191	不适用
	7 级 不适用	SEQ ID NO:10192 至 13614	不适用

[0511] 6.3 用于敲除HSV-1 RL2基因的靶向结构域

[0512] 与本发明披露方法结合在gRNA中用于敲除HSV-1 RL2基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0513] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内和(2)5' G的存在来选择。4级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内来选择。5级gRNA分子的靶向结构域基于与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)。

[0514] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内,(2)高水平的正交性、(3)5' G的存在以及(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内,(2)高水平的正交性以及(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内以及(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内以及(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)(2)5' G的存在以及(3)其中PAM是

NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)以及(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)以及(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0515] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次策略)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表3。

[0516] 表3用于敲除HSV-1 RL2基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0517]	<b>1 级</b> SEQ ID NO:21324 至 21368	SEQ ID NO:22745 至 22755	SEQ ID NO:26602
	<b>2 级</b> SEQ ID NO:21369 至 21441	SEQ ID NO:22756 至 22769	SEQ ID NO:26603
	<b>3 级</b> SEQ ID NO:21442 至 21505	SEQ ID NO:22770 至 22800	不适用
	<b>4 级</b> SEQ ID NO:21506 至 21567	SEQ ID NO:22801 至 23486	SEQ ID NO:26604
	<b>5 级</b> SEQ ID NO:21568 至 22744	SEQ ID NO:23487 至 23587	SEQ ID NO:26605 至 26612
	<b>6 级</b> 不适用	SEQ ID NO:23588 至 23745	不适用
	<b>7 级</b> 不适用	SEQ ID NO:23746 至 26601	不适用

[0518] 6.4用于敲除HSV-2 RL2基因的靶向结构域

[0519] 与本发明披露方法结合在gRNA中用于敲除HSV-2 RL2基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0520] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内和(2)高水平的正交性来选择。基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内和(2)5' G的存在来选择1级gRNA分子的靶向结构域。4级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内来选择。5级gRNA分子的靶向结构域基于与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)。

[0521] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内,(2)高水平的正交

性、(3) 5' G的存在以及(4) 其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内,(2)高水平的正交性以及(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内以及(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离,例如在靶位点(例如起始密码子)的500bp(例如,下游)内以及(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)(2) 5' G的存在以及(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)以及(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)与靶位点(例如起始密码子)的距离来选择,例如,在剩下的编码序列的范围内,例如在编码序列的第一个500bp的下游(例如+500(相对于起始密码子)至终止密码子的任何地方)以及(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次策略)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0522] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表4。

[0523] 表4用于敲除HSV-2 RL2基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
<b>1 级</b>	SEQ ID NO:26613 至 26638	SEQ ID NO:28038 至 28056	SEQ ID NO:31721 至 31722
<b>2 级</b>	SEQ ID NO:26639 至 26670	SEQ ID NO:28057 至 28107	SEQ ID NO:31723 至 31726
<b>3 级</b>	SEQ ID NO:26671 至 26743	SEQ ID NO:28108 至 28138	不适用
<b>4 级</b>	SEQ ID NO:26744 至 26863	SEQ ID NO:28139 至 28740	SEQ ID NO:31727
<b>5 级</b>	SEQ ID NO:26864 至 28037	SEQ ID NO:28741 至 28846	SEQ ID NO:31728 至 31729
<b>6 级</b>	不适用	SEQ ID NO:28847 至 28999	不适用
<b>7 级</b>	不适用	SEQ ID NO:29000 至 31720	不适用

[0525] 6.5用于敲除HSV-1LAT内含子的靶向结构域

[0526] 与本发明披露方法结合在gRNA中用于敲除HSV-1LAT内含子的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0527] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的前500bp内,(2)高水平的正交性和(3)5' G的存在。2级gRNA的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的第一个500bp内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的前500bp内和(2)5' G的存在。4级gRNA的靶向结构域基于到靶位点的距离,例如在LAT内含子的第一个500bp内,来选择。5级gRNA分子的靶向结构域基于到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游,来选择。

[0528] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内,(2)高水平的正交性,(3)5' G的存在和(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内,(2)高水平的正交性和(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内和(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游、(2)5' G的存在以及(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游以及(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游以及(2)其中PAM是NNGRRV来选择。请注意,层级不包括在内(每个gRNA仅列出一次策略)。在某些示例中,没有根据特定级别的标准确定gRNA。

[0529] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表5。

[0530] 表5用于敲除HSV-1LAT内含子的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0531]	<b>1 级</b> SEQ ID NO:31730 至 31762	SEQ ID NO:32747 至 32782	SEQ ID NO:35601
	<b>2 级</b> SEQ ID NO:31763 至 31809	SEQ ID NO:32783 至 32841	SEQ ID NO:35602 至 35603
	<b>3 级</b> SEQ ID NO:31810 至 31897	SEQ ID NO:32842 至 32893	不适用

[0532]	<b>4 级</b>	SEQ ID NO:31898 至 32025	SEQ ID NO:32894 至 33621	SEQ ID NO:35604
	<b>5 级</b>	SEQ ID NO:32026 至 32746	SEQ ID NO:33622 至 33716	SEQ ID NO:35605 至 35616
	<b>6 级</b>	不适用	SEQ ID NO:33717 至 33932	不适用
	<b>7 级</b>	不适用	SEQ ID NO:33933 至 35600	不适用

[0533] 6.6 用于敲除HSV-2LAT内含子的靶向结构域

[0534] 与本发明披露方法结合在gRNA中用于敲除HSV-2LAT内含子基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0535] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的前500bp内,(2)高水平的正交性和(3)5' G的存在。2级gRNA的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的第一个500bp内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的前500bp内和(2)5' G的存在。4级gRNA的靶向结构域基于到靶位点的距离,例如在LAT内含子的第一个500bp内,来选择。5级gRNA分子的靶向结构域基于到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游,来选择。

[0536] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内,(2)高水平的正交性,(3)5' G的存在和(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内,(2)高水平的正交性和(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内和(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)到目标位点的距离,例如在LAT内含子的第一个500bp内和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游、(2)5' G的存在以及(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游以及(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在LAT内含子的剩下的序列内,例如在LAT内含子的第一个500bp的下游以及(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0537] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表6。

[0538] 表6用于敲除HSV-2LAT内含子的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0539]	<b>1 级</b> SEQ ID NO:35617 至 35640	SEQ ID NO:36927 至 36941	SEQ ID NO:40872 至 40873
	<b>2 级</b> SEQ ID NO:35641 至 35704	SEQ ID NO:36942 至 36980	不适用
	<b>3 级</b> SEQ ID NO:35705 至 35795	SEQ ID NO:36981 至 37038	不适用
	<b>4 级</b> SEQ ID NO:35796 至 35911	SEQ ID NO:37039 至 37860	不适用
	<b>5 级</b> SEQ ID NO:35912 至 36926	SEQ ID NO:37861 至 37960	SEQ ID NO:40874 至 40885
	<b>6 级</b> 不适用	SEQ ID NO:37961 至 38098	不适用
	<b>7 级</b> 不适用	SEQ ID NO:38099 至 40871	不适用

[0540] 6.7 用于敲低HSV-1 RS1基因的靶向结构域

[0541] 与本发明披露方法结合在gRNA中用于敲低HSV-1 RS1基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0542] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)5' G的存在来选择。4级gRNA的靶向结构域基于到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内来选择。5级gRNA分子的靶向结构域基于到靶位点的距离来选择,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb。

[0543] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性,(3)5' G的存在和(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性和(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb、(2)5' G的存在和(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb和(2)PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点的距

离,例如在转录起始位点 (TSS) 的另外500bp (例如上游或下游) 内,例如延伸至TSS上游和下游的1kb和 (2) PAM是NNGRRV来选择。请注意,层级是非包容性的 (对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0544] 请注意,层级是非包容性的 (对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定级次的标准确定靶向结构域。鉴定的靶向结构域总结如下表7。

[0545] 表7用于敲低HSV-1 RS1基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0546]	<b>1 级</b> SEQ ID NO:13637 至 13743	SEQ ID NO:14795 至 14841	SEQ ID NO:17742 至 17743
	<b>2 级</b> SEQ ID NO:13744 至 13957	SEQ ID NO:14842 至 14938	SEQ ID NO:17744 至 17748
	<b>3 级</b> SEQ ID NO:13958 至 14099	SEQ ID NO:14939 至 14965	不适用
	<b>4 级</b> SEQ ID NO:14100 至 14274	SEQ ID NO:14966 至 16394	不适用
	<b>5 级</b> SEQ ID NO:14275 至 14794	SEQ ID NO:16395 至 16458	SEQ ID NO:17749 至 17752
	<b>6 级</b> 不适用	SEQ ID NO:16459 至 16548	不适用
	<b>7 级</b> 不适用	SEQ ID NO:16549 至 17741	不适用

[0547] 6.8用于敲低HSV-2 RS1基因的靶向结构域

[0548] 与本发明披露方法结合在gRNA中用于敲低HSV-2 RS1基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0549] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于 (1) 与靶位点的距离,例如在起始密码子的500bp (例如上游或下游) 内、(2) 高水平的正交性和 (3) 5' G的存在来选择。2级gRNA的靶向结构域基于 (1) 到靶位点的距离,例如在起始密码子的500bp (例如上游或下游) 内和 (2) 高水平的正交性来选择。3级gRNA分子的靶向结构域基于 (1) 与靶位点的距离,例如在起始密码子的500bp (例如上游或下游) 内和 (2) 5' G的存在来选择。4级gRNA的靶向结构域基于到靶位点的距离,例如在起始密码子的500bp (例如上游或下游) 内来选择。5级gRNA分子的靶向结构域基于到靶位点的距离来选择,例如在起始密码子的另外500bp (例如上游或下游) 内,例如延伸至起始密码子上游和下游的1kb。

[0550] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于 (1) 与靶位点的距离,例如在起始密码子的500bp内 (例如上游或下游) 、(2) 高水平的正交性, (3) 5' G的存在和 (4) 其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于 (1) 与靶位点的距离,例如在起始密码子的500bp内 (例如上游或下游) 、(2) 高水平的正交性和 (3) 其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于 (1) 与靶位点的距离,例如在起始密码子的500bp内 (例如上游或下游) 和 (2) 其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于 (1) 与靶位点的距

离,例如在起始密码子的500bp内(例如上游或下游)和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在起始密码子的另外500bp(例如上游或下游)内,例如延伸至起始密码子上游和下游的1kb、(2)5' G的存在和(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在起始密码子的另外500bp(例如上游或下游)内,例如延伸至起始密码子上游和下游的1kb和(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在起始密码子的另外500bp(例如上游或下游)内,例如延伸至起始密码子上游和下游的1kb和(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0551] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表8。

[0552] 表8用于敲低HSV-2 RS1基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0553]	<b>1 级</b> SEQ ID NO:17753 至 17850	SEQ ID NO:18785 至 18816	SEQ ID NO:21312
	<b>2 级</b> SEQ ID NO:17851 至 18016	SEQ ID NO:18817 至 18868	SEQ ID NO:21313 至 21319
	<b>3 级</b> SEQ ID NO:18017 至 18102	SEQ ID NO:18869 至 18889	不适用
	<b>4 级</b> SEQ ID NO:18103 至 18257	SEQ ID NO:18890 至 20108	不适用
	<b>5 级</b> SEQ ID NO:18258 至 18784	SEQ ID NO:20109 至 20151	SEQ ID NO:21320 至 21323
[0554]	<b>6 级</b> 不适用	SEQ ID NO:20152 至 20213	不适用
	<b>7 级</b> 不适用	SEQ ID NO:20214 至 21311	不适用

[0555] 6.9用于敲低HSV-1 RL2基因的靶向结构域

[0556] 与本发明披露方法结合在gRNA中用于敲低HSV-1 RL2基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0557] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)5' G的存在来选择。4级gRNA的靶向结构域基于到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内来选择。5级gRNA分子的靶向结构域基于到靶位点的距

离来选择,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb。

[0558] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性,(3)5' G的存在和(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性和(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb、(2)5' G的存在和(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb和(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb和(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0559] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表9。

[0560] 表9用于敲低HSV-1 RL2基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0561]	1 级	SEQ ID NO:40886 至 40976	SEQ ID NO:42079 至 42149
	2 级	SEQ ID NO:40977 至 41180	SEQ ID NO:42150 至 42299
	3 级	SEQ ID NO:41181 至 41305	SEQ ID NO:42300 至 42323
	4 级	SEQ ID NO:41306 至 41493	SEQ ID NO:42324 至 43715
	5 级	SEQ ID NO:41494 至 42078	SEQ ID NO:43716 至 43784
	6 级	不适用	SEQ ID NO:43785 至 43921
	7 级	不适用	SEQ ID NO:43922 至 45315

[0562] 6.10用于敲低HSV-2 RL2基因的靶向结构域

[0563] 与本发明披露方法结合在gRNA中用于敲低HSV-2 RL2基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在起始位点(TSS)的500bp(例如上游或下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA的靶向结构域基于(1)到靶位点的距离,例如在起始位点(TSS)的500bp(例如上游或下游)内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在起始密码子的500bp(例如上游或下游)内和(2)5' G的存在来选择。4级gRNA的靶向结构域基于到靶位点的距离,例如在起始密码子的500bp(例如上游或下游)内来选择。5级gRNA分子的靶向结构域基于到靶位点的距离来选择,例如在起始密码子的另外500bp(例如上游或下游)内,例如延伸至起始密码子上游和下游的1kb。

[0564] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在起始密码子的500bp内(例如上游或下游)、(2)高水平的正交性,(3)5' G的存在和(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在起始密码子的500bp内(例如上游或下游)、(2)高水平的正交性和(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在起始密码子的500bp内(例如上游或下游)和(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在起始密码子的500bp内(例如上游或下游)和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在起始密码子的另外500bp(例如上游或下游)内,例如延伸至起始密码子上游和下游的1kb、(2)5' G的存在和(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在起始密码子的另外500bp(例如上游或下游)内,例如延伸至起始密码子上游和下游的1kb和(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在起始密码子的另外500bp(例如上游或下游)内,例如延伸至起始密码子上游和下游的1kb和(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0565] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表10。

[0566] 表10用于敲低HSV-2 RL2基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
[0567]	<b>1 级</b> SEQ ID NO:49498 至 49587	SEQ ID NO:50653 至 50725	SEQ ID NO:53825 至 53834
	<b>2 级</b> SEQ ID NO:49588 至 49738	SEQ ID NO:50726 至 50857	SEQ ID NO:53835 至 53843
	<b>3 级</b> SEQ ID NO:49739 至 49899	SEQ ID NO:50858 至 50911	不适用
	<b>4 级</b> SEQ ID NO:49900 至 50151	SEQ ID NO:50912 至 52535	不适用
	<b>5 级</b> SEQ ID NO:50152 至 50652	SEQ ID NO:52536 至 52587	SEQ ID NO:53844 至 53857
	<b>6 级</b> 不适用	SEQ ID NO:52588 至 52696	不适用
	<b>7 级</b> 不适用	SEQ ID NO:52697 至 53824	不适用

[0568] 6.11 用于敲低HSV-1LAT基因的靶向结构域

[0569] 与本发明披露方法结合在gRNA中用于敲低HSV-1LAT基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0570] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)5' G的存在来选择。4级gRNA的靶向结构域基于到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内来选择。5级gRNA分子的靶向结构域基于到靶位点的距离来选择,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb。

[0571] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性,(3)5' G的存在和(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性和(3)其中PAM是NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb、(2)5' G的存在和(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb和(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点

的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb和(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0572] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表11。

[0573] 表11用于敲低HSV-1LAT基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
<b>1 级</b>	SEQ ID NO:45340 至 45427	SEQ ID NO:46480 至 46533	不适用
<b>2 级</b>	SEQ ID NO:45428 至 45589	SEQ ID NO:46534 至 46619	SEQ ID NO:49480 至 49488
<b>3 级</b>	SEQ ID NO:45590 至 45707	SEQ ID NO:46620 至 46664	不适用
<b>4 级</b>	SEQ ID NO:45708 至 45897	SEQ ID NO:46665 至 47789	SEQ ID NO:49489
<b>5 级</b>	SEQ ID NO:45898 至 46479	SEQ ID NO:47790 至 47887	SEQ ID NO:49490 至 49497
<b>6 级</b>	不适用	SEQ ID NO:47888 至 48041	不适用
<b>7 级</b>	不适用	SEQ ID NO:48042 至 49479	不适用

[0576] 6.12用于敲低HSV-2LAT基因的靶向结构域

[0577] 与本发明披露方法结合在gRNA中用于敲低HSV-2LAT基因的靶向结构域被识别和分级为5级用于化脓性链球菌、7级用于金黄色葡萄球菌,以及5级用于脑膜炎奈瑟氏球菌。

[0578] 对于化脓性链球菌和脑膜炎奈瑟氏球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内、(2)高水平的正交性和(3)5' G的存在来选择。2级gRNA的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)高水平的正交性来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内和(2)5' G的存在来选择。4级gRNA的靶向结构域基于到靶位点的距离,例如在转录起始位点(TSS)的500bp(例如上游或下游)内来选择。5级gRNA分子的靶向结构域基于到靶位点的距离来选择,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb。

[0579] 对于金黄色葡萄球菌,1级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性,(3)5' G的存在和(4)其中PAM是NNGRRT来选择。2级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)、(2)高水平的正交性和(3)其中PAM是

NNGRRT来选择。3级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRT来选择。4级gRNA分子的靶向结构域基于(1)与靶位点的距离,例如在转录起始位点(TSS)的500bp内(例如上游或下游)和(2)其中PAM是NNGRRV来选择。5级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb、(2)5' G的存在和(3)其中PAM是NNGRRT来选择。6级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb和(2)其中PAM是NNGRRT来选择。7级gRNA分子的靶向结构域基于(1)到靶位点的距离,例如在转录起始位点(TSS)的另外500bp(例如上游或下游)内,例如延伸至TSS上游和下游的1kb和(2)其中PAM是NNGRRV来选择。请注意,层级是非包容性的(对该策略每个gRNA仅列出一次)。在某些示例中,没有根据特定层级的标准确定gRNA。

[0580] 请注意,层级是非包容性的(对该策略每个靶向结构域仅列出一次)。在某些示例中,没有根据特定层级的标准确定靶向结构域。鉴定的靶向结构域总结如下表12。

[0581] 表12用于敲低HSV-2LAT基因的化脓性链球菌、金黄色葡萄球菌和脑膜炎奈瑟氏球菌靶向结构域的核苷酸序列

	化脓性链球菌	金黄色葡萄球菌	脑膜炎奈瑟氏球菌
<b>1 级</b>	SEQ ID NO:53858 至 53916	SEQ ID NO:55057 至 55080	SEQ ID NO:58732 至 58736
<b>2 级</b>	SEQ ID NO:53917 至 54030	SEQ ID NO:55081 至 55119	SEQ ID NO:58737 至 58743
<b>3 级</b>	SEQ ID NO:54031 至 54240	SEQ ID NO:55120 至 55140	不适用
<b>4 级</b>	SEQ ID NO:54241 至 54468	SEQ ID NO:55141 至 56792	不适用
<b>5 级</b>	SEQ ID NO:54469 至 55056	SEQ ID NO:56793 至 56865	SEQ ID NO:58744 至 58749
<b>6 级</b>	不适用	SEQ ID NO:56866 至 56970	不适用
<b>7 级</b>	不适用	SEQ ID NO:56971 至 58731	不适用

[0582] [0583] 本发明所述的一种或多种gRNA分子(例如包括表1-12中所述的靶向域的gRNA分子)可与至少一种Cas9分子(例如化脓性链球菌Cas9分子和/或金黄色葡萄球菌Cas9分子)一起使用以形成单链或双链切割,例如用Cas9切口酶分子产生单链断裂或用Cas9核酸酶分子产生双链断裂。

[0584] 在某些实施方式中,当单个gRNA分子用于靶向Cas9切口酶以在RS1、RL2或LAT靶标处附近产生单链断裂时,例如,该gRNA用于靶向RS1、RL2或LAT靶标的上游(例如,上游500bp内)或下游(例如,下游500bp内)。

[0585] 在某些实施方式中,当单个gRNA分子用于靶向Cas9核酸酶以在RL2、LAT或RS1靶标处附近产生双链断裂时,例如,该gRNA用于靶向RS1、RL2或LAT靶标的上游(例如,上游500bp

内)或下游(例如,下游500bp内)。

[0586] 在某些实施方式中,两个或多个(例如,三个或四个)gRNA分子与一个Cas9分子或Cas9-融合蛋白一起使用。在某些实施方式中,当两种或多种(例如,三种或四种)gRNA与两种或多种Cas9分子或Cas9-融合蛋白一起使用时,至少一种Cas9分子来自与其他Cas9分子不同的物种。当设计用于靶向两个Cas9分子的两个gRNA时,一种Cas9可以是一个物种,第二种Cas9可来自不同的物种。需要的话,Cas9物种均用来产生单链或双链断裂。

[0587] 表1-12中描述的任何上游gRNA可与表1-12中描述的任何下游gRNA配对。当设计用于一种Cas9的上游gRNA与设计用于不同物种Cas9的下游gRNA配对时,根据需要,两种Cas9种类都用于产生单链或双链断裂。

[0588] 7. Cas9分子

[0589] 多个物种的Cas9分子可以用于本发明所述的方法和组合物中。尽管化脓链球菌、金黄色葡萄球菌和脑膜炎奈瑟菌Cas9分子是本发明的披露的大部分的主题,也可以使用本文所列出的其他物种的Cas9蛋白的、从中衍生的、或基于其的Cas9分子。这些包括例如来自以下项的Cas9分子:燕麦食酸菌(*Acidovorax avenae*)、胸膜肺炎放线杆菌(*Actinobacillus pleuropneumoniae*)、琥珀酸放线杆菌(*Actinobacillus succinogenes*)、猪放线杆菌(*Actinobacillus suis*)、放线菌属(*Actinomyces* sp.)、*cycliphilus denitrificans*、*Aminomonas paucivorans*、蜡样芽孢杆菌(*Bacillus cereus*)、史氏芽孢杆菌(*Bacillus smithii*)、苏云金芽孢杆菌(*Bacillus thuringiensis*)、拟杆菌属(*Bacteroides* sp.)、*Blastopirellula marina*、慢生根瘤菌属(*Bradyrhizobium* sp.)、侧孢短芽孢杆菌(*Brevibacillus laterosporus*)、结肠弯曲菌(*Campylobacter coli*)、空肠弯曲菌(*Campylobacter jejuni*)、红嘴鸥弯曲杆菌(*Campylobacter lari*)、*Candidatus puniceispirillum*、解纤维梭菌(*Clostridium cellulolyticum*)、产气荚膜梭菌(*Clostridium perfringens*)、拥挤棒杆菌(*Corynebacterium accolens*)、白喉棒状杆菌(*Corynebacterium diphtheriae*)、*Corynebacterium matruchotii*、恒雄芝氏沟鞭藻玫瑰杆菌(*Dinoroseobacter shibae*)、细长真杆菌(*Eubacterium dolichum*)、 $\gamma$ -变形菌纲(*gamma proteobacterium*)、重氮营养葡萄糖酸醋杆菌(*Gluconacetobacter diazotrophicus*)、副流感嗜血杆菌(*Haemophilus parainfluenzae*)、嗜血弯曲杆菌(*Haemophilus sputorum*)、*Helicobacter canadensis*、同性恋螺杆菌(*Helicobacter cinaedi*)、鼬鼠螺杆菌(*Helicobacter mustelae*)、*Ilyobacter polytropus*、金氏金氏菌(*Kingella kingae*)、卷曲乳酸杆菌(*Lactobacillus crispatus*)、伊氏李斯特菌(*Listeria ivanovii*)、单核细胞增生李斯特菌(*Listeria monocytogenes*)、李斯特氏菌科菌(*Listeriaceae bacterium*)、甲基孢囊菌属(*Methylocystis* sp.)、甲烷氧化菌(*Methylosinus trichosporium*)、羞怯动弯杆菌(*Mobiluncus mulieris*)、奈瑟氏杆菌(*Neisseria bacilliformis*)、灰色奈瑟球菌(*Neisseria cinerea*)、浅黄色奈瑟氏菌(*Neisseria flavescens*)、乳糖奈瑟氏菌(*Neisseria lactamica*)、奈瑟氏菌属(*Neisseria* sp.)、瓦茨瓦尔西奈瑟氏菌(*Neisseria wadsworthii*)、亚硝化单胞菌属(*Nitrosomonas* sp.)、食清洁剂细小棒菌(*Parvibaculum lavamentivorans*)、出血败血性巴士杆菌(*Pasteurella multocida*)、*Phascolarctobacterium succinatutens*、*Ralstonia syzygii*、沼泽红假单胞菌(*Rhodopseudomonas palustris*)、小红卵菌属(*Rhodovulum*

sp.)、米氏西蒙斯氏菌 (Simonsiella muelleri)、鞘氨醇单胞菌属 (Sphingomonas sp.)、Sporolactobacillus vineae、路邓葡萄球菌 (Staphylococcus lugdunensis)、链球菌属 (Streptococcus sp.)、Subdoligranulum sp.、运动替斯崔纳菌 (Tistrella mobilis)、密螺旋体属 (Treponema sp.) 或 Verminephrobacter eiseniae。

[0590] 7.1 Cas9结构域

[0591] 已经确定了两种不同的天然存在的细菌Cas9分子 (Jinek 2014) 和具有指导RNA (例如, crRNA和tracrRNA的合成融合体) 的化脓链球菌Cas9 (Nishimasu 2014; Anders 2014) 的晶体结构。

[0592] 天然存在的Cas9分子包含两种叶片:识别 (REC) 叶片和核酸酶 (NUC) 叶片;其各自进一步包含本文所述的结构域。图8A-8B提供了重要的Cas9结构域的一级结构的组织的示意图。贯穿本披露使用的由每个结构域所涵盖的结构域命名和氨基酸残基编号是如先前 (Nishimasu 2014) 所描述的。氨基酸残基的编号是参照来自化脓链球菌的Cas9。

[0593] REC叶片包含富精氨酸的桥螺旋 (BH)、REC1结构域、和REC2结构域。REC叶片与其他已知蛋白不享有结构相似性,指示它是Cas9特异性功能结构域。BH结构域是长的螺旋且富精氨酸的区域并且包含化脓链球菌Cas9序列的氨基酸60-93。REC1结构域对于例如gRNA或tracrRNA的重复:抗重复双链体的识别而言是重要的,并且因此对于识别靶序列的Cas9活性而言是关键的。REC1结构域在化脓链球菌Cas9序列的氨基酸94至179和308至717处包含两个REC1基序。尽管在线性一级结构中被REC2结构域分开,这两个REC1结构域在三级结构中组装以形成REC1结构域。REC2结构域、或其部分在重复:抗重复双链体的识别中也可以发挥作用。REC2结构域包括化脓链球菌Cas9序列的氨基酸180至307。

[0594] NUC叶片包含RuvC结构域、HNH结构域和PAM相互作用 (PI) 结构域。RuvC结构域与逆转录病毒整合酶超家族成员享有结构相似性,并且切割靶核酸分子的单链(例如,非互补链)。RuvC结构域由化脓链球菌Cas9序列的氨基酸1-59、718-769和909-1098处的三个分割RuvC基序 (RuvCI、RuvCII和RuvCIII,在本领域其通常分别称为RuvCI结构域或N-末端RuvC结构域、RuvCII结构域和RuvCIII结构域) 组装而来。与REC1结构域类似,三个RuvC基序被一级结构中的其他结构域线性分开。然而,在三级结构中,这三个RuvC基序组装并形成RuvC结构域。HNH结构域与HNH内切核酸酶享有结构相似性,并且切割靶核酸分子的单链(例如,非互补链)。HNH结构域位于RuvC II-III基序之间并包含化脓链球菌Cas9序列的氨基酸775-908。PI结构域与靶核酸分子的PAM相互作用,并包含化脓链球菌Cas9序列的氨基酸1099-1368。

[0595] 7.1.1 RuvC样结构域和HNH样结构域

[0596] 在某些实施方式中,Cas9分子或Cas9多肽包含HNH样结构域和RuvC样结构域,并且在这些实施方式的某些中,切割活性取决于RuvC样结构域和HNH样结构域。Cas9分子或Cas9多肽可以包含RuvC样结构域和HNH样结构域中的一者或更多者。在某些实施方式中,Cas9分子或Cas9多肽包含RuvC样结构域(例如,如下所述的RuvC样结构域)和/或HNH样结构域(例如,如下所述的HNH样结构域)。

[0597] RuvC样结构域

[0598] 在某些实施方式中,RuvC样结构域切割靶核酸分子的单链(例如,非互补链)。Cas9分子或Cas9多肽可以包括多于一个RuvC样结构域(例如,一个、两个、三个或更多个RuvC样

结构域)。在某些实施方式中,RuvC样结构域的长度为至少5、6、7、8个氨基酸但长度不多于20、19、18、17、16或15个氨基酸。在某些实施方式中,Cas9分子或Cas9多肽包含长度为约10至20个氨基酸(例如,约15个氨基酸)的N-末端RuvC样结构域。

[0599] 7.1.2N-末端RuvC样结构域

[0600] 一些天然存在的Cas9分子包含多于一个RuvC样结构域,其中切割取决于N-末端RuvC样结构域。因此,Cas9分子或Cas9多肽可以包含N-末端RuvC样结构域。示例性N-末端RuvC样结构域如下所述。

[0601] 在某些实施方式中,Cas9分子或Cas9多肽包含N-末端RuvC样结构域,所述N-末端RuvC样结构域包含具有化学式I的氨基酸序列:

[0602] D-X<sub>1</sub>-G-X<sub>2</sub>-X<sub>3</sub>-X<sub>4</sub>-X<sub>5</sub>-G-X<sub>6</sub>-X<sub>7</sub>-X<sub>8</sub>-X<sub>9</sub> (SEQ ID NO:20),

[0603] 其中

[0604] X<sub>1</sub>选自I、V、M、L和T(例如,选自I、V和L);

[0605] X<sub>2</sub>选自T、I、V、S、N、Y、E和L(例如,选自T、V和I);

[0606] X<sub>3</sub>选自N、S、G、A、D、T、R、M和F(例如,A或N);

[0607] X<sub>4</sub>选自S、Y、N和F(例如,S);

[0608] X<sub>5</sub>选自V、I、L、C、T和F(例如,选自V、I和L);

[0609] X<sub>6</sub>选自W、F、V、Y、S和L(例如,W);

[0610] X<sub>7</sub>选自A、S、C、V和G(例如,选自A和S);

[0611] X<sub>8</sub>选自V、I、L、A、M和H(例如,选自V、I、M和L);并且

[0612] X<sub>9</sub>选自任何氨基酸或是不存在的(例如,选自T、V、I、L、Δ、F、S、A、Y、M和R,或例如,选自T、V、I、L和Δ)。

[0613] 在某些实施方式中,N-末端RuvC样结构域与SEQ ID NO:20的序列相差多达1个但不多于2、3、4或5个残基。

[0614] 在某些实施方式中,N-末端RuvC样结构域是有切割能力的。在其他实施方式中,N-末端RuvC样结构域是无切割能力的。

[0615] 在某些实施方式中,Cas9分子或Cas9多肽包含N-末端RuvC样结构域,所述N-末端RuvC样结构域包含具有化学式II的氨基酸序列:

[0616] D-X<sub>1</sub>-G-X<sub>2</sub>-X<sub>3</sub>-S-X<sub>5</sub>-G-X<sub>6</sub>-X<sub>7</sub>-X<sub>8</sub>-X<sub>9</sub> (SEQ ID NO:21),

[0617] 其中

[0618] X<sub>1</sub>选自I、V、M、L和T(例如,选自I、V和L);

[0619] X<sub>2</sub>选自T、I、V、S、N、Y、E和L(例如,选自T、V和I);

[0620] X<sub>3</sub>选自N、S、G、A、D、T、R、M和F(例如,A或N);

[0621] X<sub>5</sub>选自V、I、L、C、T和F(例如,选自V、I和L);

[0622] X<sub>6</sub>选自W、F、V、Y、S和L(例如,W);

[0623] X<sub>7</sub>选自A、S、C、V和G(例如,选自A和S);

[0624] X<sub>8</sub>选自V、I、L、A、M和H(例如,选自V、I、M和L);并且

[0625] X<sub>9</sub>选自任何氨基酸或是不存在的(例如,选自T、V、I、L、Δ、F、S、A、Y、M和R,或选自例如,T、V、I、L和Δ)。

[0626] 在某些实施方式中,N-末端RuvC样结构域与SEQ ID NO:21的序列相差多达1个但

不多于2、3、4或5个残基。

[0627] 在某些实施方式中,N-末端RuvC样结构域包含具有化学式III的氨基酸序列:

[0628] D-I-G-X<sub>2</sub>-X<sub>3</sub>-S-V-G-W-A-X<sub>8</sub>-X<sub>9</sub> (SEQ ID NO:22) ,

[0629] 其中

[0630] X<sub>2</sub>选自T、I、V、S、N、Y、E和L (例如,选自T、V和I) ;

[0631] X<sub>3</sub>选自N、S、G、A、D、T、R、M和F (例如,A或N) ;

[0632] X<sub>8</sub>选自V、I、L、A、M和H (例如,选自V、I、M和L) ;并且

[0633] X<sub>9</sub>选自任何氨基酸或是不存在的 (例如,选自T、V、I、L、Δ、F、S、A、Y、M和R,或选自例如,T、V、I、L和Δ) 。

[0634] 在某些实施方式中,N-末端RuvC样结构域与SEQ ID NO:22的序列相差多达1个但不多于2、3、4或5个残基。

[0635] 在某些实施方式中,N-末端RuvC样结构域包含具有化学式IV的氨基酸序列:

[0636] D-I-G-T-N-S-V-G-W-A-V-X (SEQ ID NO:23) ,

[0637] 其中

[0638] X是非极性烷基氨基酸或羟基氨基酸,例如,X选自V、I、L和T (例如,Cas9分子可以包含示于图2A-2G中的N-末端RuvC样结构域(描绘为Y))。

[0639] 在某些实施方式中,N-末端RuvC样结构域与SEQ ID NO:23的序列相差多达1个但不多于2、3、4或5个残基。在某些实施方式中,N-末端RuvC样结构域与本发明(例如,在图3A-3B中)所披露的N-末端RuvC样结构域的序列相差多达1个但不多于2、3、4或5个残基。在一个实施方式中,在图3A-3B中鉴定出的高度保守残基中的1个、2个、3个或全部是存在的。

[0640] 在某些实施方式中,N-末端RuvC样结构域与本发明(例如,在图4A-4B中)所披露的N-末端RuvC样结构域的序列相差多达1个但不多于2、3、4或5个残基。在某些实施方式中,在图4A-4B中鉴定出的高度保守残基中的1个、2个或全部是存在的。

[0641] 7.1.3另外的RuvC样结构域

[0642] 除了N-末端RuvC样结构域之外,Cas9分子或Cas9多肽可以包含一个或多个另外的RuvC样结构域。在某些实施方式中,Cas9分子或Cas9多肽包含两个另外的RuvC样结构域。在某些实施方式中,另外的RuvC样结构域的长度为至少5个氨基酸,并且例如长度为小于15个氨基酸,例如长度为5至10个氨基酸,例如长度为8个氨基酸。

[0643] 另外的RuvC样结构域可以包含具有化学式V的氨基酸序列:

[0644] I-X<sub>1</sub>-X<sub>2</sub>-E-X<sub>3</sub>-A-R-E (SEQ ID NO:15)

[0645] 其中,

[0646] X<sub>1</sub>是V或H;

[0647] X<sub>2</sub>是I、L或V (例如,I或V) ;并且

[0648] X<sub>3</sub>是M或T。

[0649] 在某些实施方式中,另外的RuvC样结构域包含具有化学式VI的氨基酸序列:

[0650] I-V-X<sub>2</sub>-E-M-A-R-E (SEQ ID NO:16) ,

[0651] 其中

[0652] X<sub>2</sub>是I、L或V (例如,I或V) (例如,Cas9分子或Cas9多肽可以包含示于图2A-2G中的另外的RuvC样结构域(描绘为B))。

[0653] 另外的RuvC样结构域可以包含具有化学式VII的氨基酸序列：

[0654] H-H-A-X<sub>1</sub>-D-A-X<sub>2</sub>-X<sub>3</sub> (SEQ ID NO:17) ,

[0655] 其中

[0656] X<sub>1</sub>是H或L;

[0657] X<sub>2</sub>是R或V;并且

[0658] X<sub>3</sub>是E或V。

[0659] 在某些实施方式中,另外的RuvC样结构域包含如下氨基酸序列:H-H-A-H-D-A-Y-L (SEQ ID NO:18)。

[0660] 在某些实施方式中,另外的RuvC样结构域与SEQ ID NO:15-18的序列相差多达1个但不多于2、3、4或5个残基。

[0661] 在某些实施方式中,N-末端RuvC样结构域侧翼的序列具有化学式VIII的氨基酸序列:

[0662] K-X<sub>1</sub>'-Y-X<sub>2</sub>'-X<sub>3</sub>'-X<sub>4</sub>'-Z-T-D-X<sub>9</sub>'-Y (SEQ ID NO:19) ,

[0663] 其中

[0664] X<sub>1</sub>' 选自K和P;

[0665] X<sub>2</sub>' 选自V、L、I和F (例如,V、I和L) ;

[0666] X<sub>3</sub>' 选自G、A和S (例如,G) ;

[0667] X<sub>4</sub>' 选自L、I、V和F (例如,L) ;

[0668] X<sub>9</sub>' 选自D、E、N和Q;并且

[0669] Z是N-末端RuvC样结构域,例如,如以上所描述的,例如,具有5到20个氨基酸。

[0670] 7.1.4 HNH样结构域

[0671] 在某些实施方式中,HNH样结构域切割双链核酸分子的单链互补结构域(例如,互补链)。在某些实施方式中,HNH样结构域的长度为至少15、20或25个氨基酸但长度为不多于40、35或30个氨基酸,例如长度为20至35个氨基酸,例如长度为25至30个氨基酸。示例性的 HNH样结构域如下所述。

[0672] 在某些实施方式中,Cas9分子或Cas9多肽包含HNH样结构域,所述HNH样结构域具有化学式IX的氨基酸序列:

[0673] X<sub>1</sub>-X<sub>2</sub>-X<sub>3</sub>-H-X<sub>4</sub>-X<sub>5</sub>-P-X<sub>6</sub>-X<sub>7</sub>-X<sub>8</sub>-X<sup>9</sup>-X<sup>10</sup>-X<sup>11</sup>-X<sup>12</sup>-X<sup>13</sup>-X<sup>14</sup>-X<sup>15</sup>-N-X<sup>16</sup>-X<sup>17</sup>-X<sup>18</sup>-X<sup>19</sup>-X<sub>20</sub>-X<sub>21</sub>-X<sub>22</sub>-X<sub>23</sub>-N (SEQ ID NO:25) ,其中

[0674] X<sub>1</sub>选自D、E、Q和N (例如,D和E) ;

[0675] X<sub>2</sub>选自L、I、R、Q、V、M和K;

[0676] X<sub>3</sub>选自D和E;

[0677] X<sub>4</sub>选自I、V、T、A和L (例如,A、I和V) ;

[0678] X<sub>5</sub>选自V、Y、I、L、F和W (例如,V、I和L) ;

[0679] X<sub>6</sub>选自Q、H、R、K、Y、I、L、F和W;

[0680] X<sub>7</sub>选自S、A、D、T和K (例如,S和A) ;

[0681] X<sub>8</sub>选自F、L、V、K、Y、M、I、R、A、E、D和Q (例如,F) ;

[0682] X<sub>9</sub>选自L、R、T、I、V、S、C、Y、K、F和G;

[0683] X<sub>10</sub>选自K、Q、Y、T、F、L、W、M、A、E、G和S;

- [0684]  $X_{11}$  选自D、S、N、R、L和T(例如,D)；
- [0685]  $X_{12}$  选自D、N和S；
- [0686]  $X_{13}$  选自S、A、T、G和R(例如,S)；
- [0687]  $X_{14}$  选自I、L、F、S、R、Y、Q、W、D、K和H(例如,I、L和F)；
- [0688]  $X_{15}$  选自D、S、I、N、E、A、H、F、L、Q、M、G、Y和V；
- [0689]  $X_{16}$  选自K、L、R、M、T和F(例如,L、R和K)；
- [0690]  $X_{17}$  选自V、L、I、A和T；
- [0691]  $X_{18}$  选自L、I、V和A(例如,L和I)；
- [0692]  $X_{19}$  选自T、V、C、E、S和A(例如,T和V)；
- [0693]  $X_{20}$  选自R、F、T、W、E、L、N、C、K、V、S、Q、I、Y、H和A；
- [0694]  $X_{21}$  选自S、P、R、K、N、A、H、Q、G和L；
- [0695]  $X_{22}$  选自D、G、T、N、S、K、A、I、E、L、Q、R和Y；并且
- [0696]  $X_{23}$  选自K、V、A、E、Y、I、C、L、S、T、G、K、M、D和F。
- [0697] 在某些实施方式中, HNH样结构域与SEQ ID NO:25的序列相差至少一个但不多于2、3、4、或5个残基。
- [0698] 在某些实施方式中, HNH样结构域是有切割能力的。在某些实施方式中, HNH样结构域是无切割能力的。
- [0699] 在某些实施方式中, Cas9分子或Cas9多肽包含 HNH样结构域, 所述 HNH样结构域包括具有化学式X的氨基酸序列:
- [0700]  $X_1-X_2-X_3-H-X_4-X_5-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-X_{19}-X_{20}-X_{21}-X_{22}-X_{23}-N$   
(SEQ ID NO:26) ,
- [0701] 其中
- [0702]  $X_1$  选自D和E；
- [0703]  $X_2$  选自L、I、R、Q、V、M和K；
- [0704]  $X_3$  选自D和E；
- [0705]  $X_4$  选自I、V、T、A和L(例如,A、I和V)；
- [0706]  $X_5$  选自V、Y、I、L、F和W(例如,V、I和L)；
- [0707]  $X_6$  选自Q、H、R、K、Y、I、L、F和W；
- [0708]  $X_8$  选自F、L、V、K、Y、M、I、R、A、E、D和Q(例如,F)；
- [0709]  $X_9$  选自L、R、T、I、V、S、C、Y、K、F和G；
- [0710]  $X_{10}$  选自K、Q、Y、T、F、L、W、M、A、E、G和S；
- [0711]  $X_{14}$  选自I、L、F、S、R、Y、Q、W、D、K和H(例如,I、L和F)；
- [0712]  $X_{15}$  选自D、S、I、N、E、A、H、F、L、Q、M、G、Y和V；
- [0713]  $X_{19}$  选自T、V、C、E、S和A(例如,T和V)；
- [0714]  $X_{20}$  选自R、F、T、W、E、L、N、C、K、V、S、Q、I、Y、H和A；
- [0715]  $X_{21}$  选自S、P、R、K、N、A、H、Q、G和L；
- [0716]  $X_{22}$  选自D、G、T、N、S、K、A、I、E、L、Q、R和Y；并且
- [0717]  $X_{23}$  选自K、V、A、E、Y、I、C、L、S、T、G、K、M、D和F。
- [0718] 在某些实施方式中, HNH样结构域与SEQ ID NO:26的序列相差1、2、3、4或5个残基。

[0719] 在某些实施方式中,Cas9分子或Cas9多肽包含HNH样结构域,所述HNH样结构域包括具有化学式XI的氨基酸序列:

[0720]  $X_1-V-X_3-H-I-V-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-T-X_{20}-X_{21}-X_{22}-X_{23}-N$  (SEQ ID NO:27) ,

[0721] 其中

[0722]  $X_1$ 选自D和E;

[0723]  $X_3$ 选自D和E;

[0724]  $X_6$ 选自Q、H、R、K、Y、I、L和W;

[0725]  $X_8$ 选自F、L、V、K、Y、M、I、R、A、E、D和Q(例如,F) ;

[0726]  $X_9$ 选自L、R、T、I、V、S、C、Y、K、F和G;

[0727]  $X_{10}$ 选自K、Q、Y、T、F、L、W、M、A、E、G和S;

[0728]  $X_{14}$ 选自I、L、F、S、R、Y、Q、W、D、K和H(例如,I、L和F) ;

[0729]  $X_{15}$ 选自D、S、I、N、E、A、H、F、L、Q、M、G、Y和V;

[0730]  $X_{20}$ 选自R、F、T、W、E、L、N、C、K、V、S、Q、I、Y、H和A;

[0731]  $X_{21}$ 选自S、P、R、K、N、A、H、Q、G和L;

[0732]  $X_{22}$ 选自D、G、T、N、S、K、A、I、E、L、Q、R和Y;并且

[0733]  $X_{23}$ 选自K、V、A、E、Y、I、C、L、S、T、G、K、M、D和F。

[0734] 在某些实施方式中,HNH样结构域与SEQ ID NO:27的序列相差1、2、3、4或5个残基。

[0735] 在某些实施方式中,Cas9分子或Cas9多肽包含HNH样结构域,所述HNH样结构域具有化学式XII的氨基酸序列:

[0736]  $D-X_2-D-H-I-X_5-P-Q-X_7-F-X_9-X_{10}-D-X_{12}-S-I-D-N-X_{16}-V-L-X_{19}-X_{20}-S-X_{22}-X_{23}-N$  (SEQ ID NO:28) ,

[0737] 其中

[0738]  $X_2$ 选自I和V;

[0739]  $X_5$ 选自I和V;

[0740]  $X_7$ 选自A和S;

[0741]  $X_9$ 选自I和L;

[0742]  $X_{10}$ 选自K和T;

[0743]  $X_{12}$ 选自D和N;

[0744]  $X_{16}$ 选自R、K和L;

[0745]  $X_{19}$ 选自T和V;

[0746]  $X_{20}$ 选自S和R;

[0747]  $X_{22}$ 选自K、D和A;并且

[0748]  $X_{23}$ 选自E、K、G和N(例如,Cas9分子或Cas9多肽可以包含如本文所描述的HNH样结构域)。

[0749] 在某些实施方式中,HNH样结构域与SEQ ID NO:28的序列相差多达1个但不多于2、3、4或5个残基。

[0750] 在某些实施方式中,Cas9分子或Cas9多肽包含具有化学式XIII的氨基酸序列:

[0751]  $L-Y-Y-L-Q-N-G-X_1'-D-M-Y-X_2'-X_3'-X_4'-X_5'-L-D-I-X_6'-X_7'-L-S-X_8'-Y-Z-N-R-$

X<sub>9</sub>’-K-X<sub>10</sub>’-D-X<sub>11</sub>’-V-P (SEQ ID NO:24) ,

[0752] 其中

[0753] X<sub>1</sub>’选自K和R;

[0754] X<sub>2</sub>’选自V和T;

[0755] X<sub>3</sub>’选自G和D;

[0756] X<sub>4</sub>’选自E、Q和D;

[0757] X<sub>5</sub>’选自E和D;

[0758] X<sub>6</sub>’选自D、N和H;

[0759] X<sub>7</sub>’选自Y、R和N;

[0760] X<sub>8</sub>’选自Q、D和N;

[0761] X<sub>9</sub>’选自G和E;

[0762] X<sub>10</sub>’选自S和G;

[0763] X<sub>11</sub>’选自D和N;并且

[0764] Z是HNH样结构域,例如,如以上所描述的。

[0765] 在某些实施方式中,Cas9分子或Cas9多肽包含与SEQ ID NO:24的序列相差多达1个但不多于2、3、4或5个残基的氨基酸序列。

[0766] 在某些实施方式中,HNH样结构域与本文(例如,在图5A-5C中)所披露的HNH样结构域的序列相差多达1个但不多于2、3、4或5个残基。在某些实施方式中,在图5A-5C中鉴定出的高度保守残基中的1个或2个是存在的。

[0767] 在某些实施方式中,HNH样结构域与本发明(例如,在图6A-6B中)所披露的HNH样结构域的序列相差多达1个但不多于2、3、4或5个残基。在某些实施方式中,在图6A-6B中鉴定出的高度保守残基中的1个、2个、3个或全部是存在的。

[0768] 7.2Cas9活性

[0769] 在某些实施方式中,Cas9分子或Cas9多肽能够切割靶核酸分子。典型地,野生型Cas9分子切割靶核酸分子的两条链。Cas9分子和Cas9多肽可以被工程化以改变核酸酶切割(或其他特性),例如以提供作为切口酶、或缺乏切割靶核酸能力的Cas9分子或Cas9多肽。能够切割靶核酸分子的Cas9分子或Cas9多肽在本文中称为eaCas9(酶促活性Cas9)分子或eaCas9多肽。

[0770] 在某些实施方式中,eaCas9分子或eaCas9多肽包含以下一种或多种酶活性:

[0771] 切口酶活性,即切割核酸分子的单链(例如非互补链或互补链)的能力;

[0772] 双链核酸酶活性,即切割双链核酸的两条链并且产生双链断裂的能力,其在某些实施方式中是在两种切口酶活性的存在下;

[0773] 内切核酸酶活性;

[0774] 外切核酸酶活性;以及

[0775] 解旋酶活性,即解旋双链核酸的螺旋结构的能力。

[0776] 在某些实施方式中,酶促活性的Cas9(“eaCas9”)分子或eaCas9多肽切割两条DNA链,并且产生双链断裂。在某些实施方式中,eaCas9分子或eaCas9多肽仅切割一条链,例如,gRNA杂交到的链、或互补于与gRNA杂交的链的链。在某些实施方式中,eaCas9分子或eaCas9多肽包含与HNH结构域相关的切割活性。在某些实施方式中,eaCas9分子或eaCas9多肽包含

与RuvC结构域相关的切割活性。在某些实施方式中,eaCas9分子或eaCas9多肽包含与HNH结构域相关的切割活性以及与RuvC结构域相关的切割活性。在某些实施方式中,eaCas9分子或eaCas9多肽包含有活性、或有切割能力的HNH结构域以及无活性、或无切割能力的RuvC结构域。在某些实施方式中,eaCas9分子或eaCas9多肽包含无活性、或无切割能力的HNH结构域和有活性、或有切割能力的RuvC结构域。

[0777] 在某些实施方式中,Cas9分子或Cas9多肽具有与gRNA分子相互作用,并且结合所述gRNA分子以定位至核心靶结构域的能力,但不能切割靶核酸、或不能以有效速率进行切割。不具有或实质上不具有切割活性的Cas9分子在本发明中称为无酶促活性Cas9(“eiCas9”)分子或eiCas9多肽。例如,eiCas9分子或eiCas9多肽可以缺乏切割活性或具有实质上小于参考Cas9分子或eiCas9多肽,例如小于20%、10%、5%、1%或0.1%的切割活性,如通过本文所述的测定所测量的。

[0778] 7.3靶向和PAM

[0779] Cas9分子或Cas9多肽可以与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(并且在某些实施方式中,是PAM序列)的位点。在某些实施方式中,可使用W02015/089465中披露的gRNA来靶向本发明披露的Cas9分子或Cas9多肽(例如,eaCas9或eiCas9),该文献通过引用整体并入本发明。在某些实施方式中,使用W02015/089465中披露的gRNA靶向的Cas9分子或Cas9多肽是化脓性链球菌Cas9。在某些实施方式中,使用W02015/089465中披露的gRNA靶向的Cas9分子或Cas9多肽是金黄色葡萄球菌Cas9。

[0780] 在某些实施方式中,eaCas9分子或eaCas9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序列是在靶核酸中的序列。在某些实施方式中,靶核酸的切割发生在PAM序列的上游。来自不同细菌物种的eaCas9分子可以识别不同序列基序(例如,PAM序列)。在某些实施方式中,化脓链球菌的eaCas9分子识别序列基序NGG并且指导切割靶核酸序列的在该序列的上游的1至10(例如,3至5)bp(参见例如,Mali 2013)。

[0781] 在某些实施方式中,Cas9分子是化脓链球菌Cas9 EQR变体或化脓链球菌Cas9 VRER变体。

[0782] 在某些实施方式中,化脓性链球菌Cas9 EQR变体识别NGAG、NGCG、NGGG、NGTG、NGAA、NGAT或NGAC的PAM序列,并指导在该序列上游1至10个(例如,3至5个)碱基对切割靶核酸序列。在某些实施方式中,化脓性链球菌Cas9 EQR变体识别NGAG的PAM序列,并指导在该序列上游1至10个(例如,3至5个)碱基对切割靶核酸序列。参见Kleinstiver等人,NATURE 2015;523 (7561) :481-5。

[0783] 在某些实施方式中,化脓性链球菌Cas9 VRER变体识别NGCG、NGCA、NGCT或NGCC的PAM序列,并指导在该序列上游1至10个(例如,3至5个)碱基对切割靶核酸序列。在某些实施方式中,化脓性链球菌Cas9 VRER变体识别NGCG的序列基序,并指导在该序列上游1至10个(例如,3至5个)碱基对切割靶核酸序列。参见Kleinstiver等人,NATURE 2015;523 (7561) :481-5。

[0784] 在某些实施方式中,嗜热链球菌Cas9分子识别NGGNG (SEQ ID NO:199) 和/或NNAGAAW (W=A或T) (SEQ ID NO:200) 的PAM序列,并指导在这些序列上游的1至10(例如,3至5)bp切割靶核酸序列(参见例如,Horvath 2010;Deveau 2008)。

[0785] 在某些实施方式中,变形链球菌Cas9分子识别NGG和/或NAAR (R=A或G) (SEQ ID

NO:201)的PAM序列,并指导在这些序列上游的1至10(例如,3至5)bp切割靶核酸序列(参见例如,Deveau 2008)。

[0786] 在某些实施方式中,金黄色葡萄球菌Cas9分子识别NNGRR (R=A或G) (SEQ ID NO: 202)的PAM序列,并指导在该序列的上游的1至10(例如,3至5)bp切割靶核酸序列。

[0787] 在某些实施方式中,金黄色葡萄球菌识别NNGRRN (R=A或G) (SEQ ID NO:203)的PAM序列,并指导在该序列的上游的1至10(例如,3至5)bp切割靶核酸序列。在某些实施方式中,金黄色葡萄球菌Cas9分子识别NNGRRT (R=A或G) (SEQ ID NO:204)的PAM序列,并指导在该序列的上游的1至10(例如,3至5)bp切割靶核酸序列。在某些实施方式中,金黄色葡萄球菌Cas9分子识别NNGRRV (R=A或G) (SEQ ID NO:205)的PAM序列,并指导在该序列的上游的1至10(例如,3至5)bp切割靶核酸序列。

[0788] 在某些实施方式中,脑膜炎奈瑟菌Cas9分子识别NNNNGATT或NNNGCTT的PAM序列,并指导在该序列的上游的1至10(例如,3至5)碱基对切割靶核酸序列。参见,例如,Hou等人,PNAS Early Edition 2013,1-6。

[0789] Cas9分子识别PAM序列的能力可以例如使用如先前所述的转化测定(Jinek2012)来确定。在上述实施方式中,N可以是任何核苷酸残基,例如A、G、C或T中的任一项。

[0790] 先前已经描述了示例性的天然存在的Cas9分子(参见例如,Chylinski 2013)。此类Cas9分子包括以下项的Cas9分子:簇1细菌家族、簇2细菌家族、簇3细菌家族、簇4细菌家族、簇5细菌家族、簇6细菌家族、簇7细菌家族、簇8细菌家族、簇9细菌家族、簇10细菌家族、簇11细菌家族、簇12细菌家族、簇13细菌家族、簇14细菌家族、簇15细菌家族、簇16细菌家族、簇17细菌家族、簇18细菌家族、簇19细菌家族、簇20细菌家族、簇21细菌家族、簇22细菌家族、簇23细菌家族、簇24细菌家族、簇25细菌家族、簇26细菌家族、簇27细菌家族、簇28细菌家族、簇29细菌家族、簇30细菌家族、簇31细菌家族、簇32细菌家族、簇33细菌家族、簇34细菌家族、簇35细菌家族、簇36细菌家族、簇37细菌家族、簇38细菌家族、簇39细菌家族、簇40细菌家族、簇41细菌家族、簇42细菌家族、簇43细菌家族、簇44细菌家族、簇45细菌家族、簇46细菌家族、簇47细菌家族、簇48细菌家族、簇49细菌家族、簇50细菌家族、簇51细菌家族、簇52细菌家族、簇53细菌家族、簇54细菌家族、簇55细菌家族、簇56细菌家族、簇57细菌家族、簇58细菌家族、簇59细菌家族、簇60细菌家族、簇61细菌家族、簇62细菌家族、簇63细菌家族、簇64细菌家族、簇65细菌家族、簇66细菌家族、簇67细菌家族、簇68细菌家族、簇69细菌家族、簇70细菌家族、簇71细菌家族、簇72细菌家族、簇73细菌家族、簇74细菌家族、簇75细菌家族、簇76细菌家族、簇77细菌家族、或簇78细菌家族。

[0791] 示例性的天然存在的Cas9分子包括簇1细菌家族的Cas9分子。实例包括以下项的Cas9分子:金黄色葡萄球菌、化脓链球菌(例如,菌株SF370、MGAS10270、MGAS10750、MGAS2096、MGAS315、MGAS5005、MGAS6180、MGAS9429、NZ131和SSI-1)、嗜热链球菌(例如,菌株LMD-9)、假豕链球菌(例如,菌株SPIN 20026)、变形链球菌(例如,菌株NCTC11558)、*S.gallolyticus*(例如,菌株UCN34、ATCC BAA-2069)、*S.equines*(例如,菌株TCC 9812、MGCS 124)、*S.dysdalactiae*(例如,菌株GGS 124)、*S.bovis*(例如,菌株ATCC 700338)、*S.anginosus*(例如,菌株F0211)、*S.agalactiae*(例如,菌株NEM316、A909)、李斯特菌(例如,菌株F6854)、无害斯特菌(无害斯特菌,例如,菌株Clip11262)、意大利肠道球菌(例如,菌株DSM 15952)或屎肠球菌(例如,菌株1,231,408)。

[0792] 另外的示例性Cas9分子是脑膜炎奈瑟球菌的Cas9分子(Hou等人,PNAS Early Edition 2013,1-6)和金黄色葡萄球菌cas9分子。

[0793] 在某些实施方式中,Cas9分子或Cas9多肽包含以下氨基酸序列:

[0794] 与本发明所述的任何Cas9分子序列或与天然存在的Cas9分子序列(例如来自本发明所列物种的Cas9分子(例如,SEQ ID N0:1,2,4-6或12)或在Chylinski2013中描述的)相比,具有约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%、约96%、约97%、约98%或约99%的同源性;当与其相比时,与其相差不多于约2%、约5%、约10%、约15%、约20%、约30%或约40%的氨基酸残基;或

[0795] 与其相差至少1、2、5、10或20个氨基酸但相差不多于100、80、70、60、50、40或30个氨基酸;或与其相同。在某些实施方式中,Cas9分子或Cas9多肽包含以下一种或多种活性:切口酶活性;双链切割活性(例如,内切核酸酶和/或外切核酸酶活性);解旋酶活性;或连同gRNA分子定位至靶核酸的能力。

[0796] 在某些实施方式中,Cas9分子或Cas9多肽包含图2A-2G中共有序列的任何氨基酸序列,其中“\*”表示在化脓性链球菌、嗜热链球菌、变异链球菌或无害利斯特氏菌的Cas9分子的氨基酸序列的相应位置中发现的任何氨基酸,“-”表示不存在。在某些实施方式中,Cas9分子或Cas9多肽与图2A-2G中所披露的共有序列的序列相差多达1个但不多于2、3、4、5、6、7、8、9或10个氨基酸残基。在某些实施方式中,Cas9分子或Cas9多肽包含SEQ ID N0:2的氨基酸序列。在其他实施方式中,Cas9分子或Cas9多肽与SEQ ID N0:2的序列相差至少1个但不多于2、3、4、5、6、7、8、9或10个氨基酸残基。

[0797] 多个Cas9分子的序列的比较表明某些区域是保守的。这些鉴定如下:

[0798] 区域1(残基1至180,或在区域1'的情况下,残基120至180);

[0799] 区域2(残基360至480);

[0800] 区域3(残基660至720);

[0801] 区域4(残基817至900);以及

[0802] 区域5(残基900至960)。

[0803] 在某些实施方式中,Cas9分子或Cas9多肽包含区域1-5,连同足够的另外的Cas9分子序列以提供生物活性分子(例如,具有至少一种本文所述的活性的Cas9分子)。在某些实施方式中,区域1-5各自独立地与本发明所述的Cas9分子或Cas9多肽(例如,来自图2A-2G的序列的相应残基具有约50%、约60%、约70%、约80%、约85%、约90%、约95%、约96%、约97%、约98%或约99%的同源性。

[0804] 在某些实施方式中,Cas9分子或Cas9多肽包含以下称为区域1的氨基酸序列:

[0805] 与化脓链球菌的Cas9的氨基酸序列的氨基酸1-180具有约50%、约60%、约70%、约80%、约85%、约90%、约95%、约96%、约97%、约98%或约99%的同源性(编号是根据图2中的基序序列;图2A-2G中的四个Cas9序列中的52%的残基是保守的);

[0806] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸1-180相差至少1、2、5、10或20个氨基酸但相差不多于90、80、70、60、50、40或30个氨基酸;或

[0807] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸1-180相同。

- [0808] 在某些实施方式中,Cas9分子或Cas9多肽包含以下称为区域1'的氨基酸序列:
- [0809] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸120-180具有约55%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%、约96%、约97%、约98%或约99%的同源性(图2中的四个Cas9序列中的55%的残基是保守的);
- [0810] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸120-180相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或
- [0811] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸120-180相同。
- [0812] 在某些实施方式中,Cas9分子或Cas9多肽包含以下称为区域2的氨基酸序列:
- [0813] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸360-480具有约50%、约55%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%、约96%、约97%、约98%或约99%的同源性(图2中的四个Cas9序列中的52%的残基是保守的);
- [0814] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸360-480相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或
- [0815] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸360-480相同。
- [0816] 在某些实施方式中,Cas9分子或Cas9多肽包含以下称为区域3的氨基酸序列:
- [0817] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸660-720具有约55%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%、约96%、约97%、约98%或约99%的同源性(图2中的四个Cas9序列中的56%的残基是保守的);
- [0818] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸660-720相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或
- [0819] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸660-720相同。
- [0820] 在某些实施方式中,Cas9分子或Cas9多肽包含以下称为区域4的氨基酸序列:
- [0821] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸817-900具有约50%、约55%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%、约96%、约97%、约98%或约99%的同源性(图2A-2G中的四个Cas9序列中的55%的残基是保守的);
- [0822] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸817-900相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸;或
- [0823] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸817-900相同。
- [0824] 在某些实施方式中,Cas9分子或Cas9多肽包含以下称为区域5的氨基酸序列:
- [0825] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸900-960具有约50%、约55%、约60%、约65%、约70%、约75%、约80%、约85%、约

90%、约95%、约96%、约97%、约98%或约99%的同源性(图2A-2G中的四个Cas9序列中的60%的残基是保守的)；

[0826] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸900-960相差至少1、2或5个氨基酸但相差不多于35、30、25、20或10个氨基酸；或

[0827] 与化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌的Cas9的氨基酸序列的氨基酸900-960相同。

[0828] 7.4工程化的或改变的Cas9

[0829] 本发明所述的Cas9分子和Cas9多肽(可以具有多种特性中的任一种，包括：核酸酶活性(例如，内切核酸酶和/或外切核酸酶活性)；解旋酶活性；在功能上与gRNA分子相关联的能力；以及靶向(或定位至)核酸上的位点的能力(例如，PAM识别和特异性)。在某些实施方式中，Cas9分子或Cas9多肽可以包括这些特性的全部或子集。在某些实施方式中，Cas9分子或Cas9多肽具有与gRNA分子相互作用，并且与所述gRNA分子一起定位至核酸中的位点的能力。其他活性(例如，PAM特异性、切割活性、或解旋酶活性)在Cas9分子和Cas9多肽中可以更广泛地变化。

[0830] Cas9分子包括工程化的Cas9分子和工程化的Cas9多肽(如在此背景下使用的，工程化的仅仅意指所述Cas9分子或Cas9多肽不同于参考序列，并且没有暗示过程或来源限制)。工程化的Cas9分子或Cas9多肽可以包含改变的酶特性，例如改变的核酸酶活性(与天然存在的或其他参考Cas9分子相比)或改变的解旋酶活性。如本文所讨论的，工程化的Cas9分子或Cas9多肽可以具有切口酶活性(与双链核酸酶活性相反)。在某些实施方式中，工程化的Cas9分子或Cas9多肽可以具有改变其尺寸的改变，例如减小其尺寸的氨基酸序列缺失，例如对一种或多种或任何Cas9活性没有显著影响。在某些实施方式中，工程化的Cas9分子或Cas9多肽可包括影响PAM识别的改变。在某些实施方式中，工程化的Cas9分子被改变以识别除内源性野生型PI结构域识别的PAM序列以外的PAM序列。在某些实施方式中，Cas9分子或Cas9多肽在序列上可以不同于天然存在的Cas9分子，但是在一种或多种Cas9活性上没有显著改变。

[0831] 具有所希望特性的Cas9分子或Cas9多肽可以通过多种方式制成，例如，通过改变亲本(例如天然存在的)Cas9分子或Cas9多肽，以提供具有所希望特性的改变的Cas9分子或Cas9多肽。例如，可以相对于亲本Cas9分子(例如，天然存在的或工程化的Cas9分子)引入一个或多个突变或差异。此类突变和差异包括：取代(例如，保守取代或非必需氨基酸的取代)；插入；或缺失。在某些实施方式中，Cas9分子或Cas9多肽相对于参考(例如，亲本)Cas9分子可以包含一个或多个突变或差异，例如，至少1、2、3、4、5、10、15、20、30、40或50个突变但少于200、100或80个突变。

[0832] 在某些实施方式中，一个突变或多个突变对Cas9活性(例如，本发明所述的Cas9活性)不具有实质影响。在某些实施方式中，一个突变或多个突变对Cas9活性(例如，本发明所述的Cas9活性)具有实质影响。

[0833] 7.5经修饰的切割Cas9

[0834] 在某些实施方式中，Cas9分子或Cas9多肽包含不同于天然存在的Cas9分子(例如，不同于具有最接近同源性的天然存在的Cas9分子)的切割特性。例如，Cas9分子或Cas9多肽可不同于天然存在的Cas9分子，例如化脓性链球菌的Cas9分子，如下：例如，与天然存在的

Cas9分子(例如,化脓性链球菌的Cas9分子)相比,其调节(例如,减少或增加)的双链核酸切割(内切核酸酶和/或核酸外切酶活性)的能力;与天然存在的Cas9分子(例如,化脓性链球菌的Cas9分子)相比,其调节(例如,降低或增加)单链核酸(例如,核酸分子的非互补链或核酸分子的互补链)的切割能力(切口酶活性);或可消除切割核酸分子(例如,双链或单链核酸分子)的能力。

[0835] 在某些实施方式中,eaCas9分子或eaCas9多肽包含以下一种或多种活性:与N-末端RuvC样结构域相关的切割活性;与HNH样结构域相关的切割活性;与HNH样结构域相关的切割活性和与N-末端RuvC样结构域相关的切割活性。

[0836] 在某些实施方式中,eaCas9分子或eaCas9多肽包含有活性、或有切割能力的HNH样结构域(例如,本文所述的HNH样结构域,例如,SEQ ID N0:24-28)和无活性、或无切割能力的N-末端RuvC样结构域。示例性的无活性、或无切割能力的N-末端RuvC样结构域可以在N-末端RuvC样结构域中具有天冬氨酸的突变(例如,披露于图2A-2G中的共有序列的位置9处的天冬氨酸或SEQ ID N0:2的位置10处的天冬氨酸例如可以被丙氨酸取代)。在某些实施方式中,eaCas9分子或eaCas9多肽与野生型的区别在于N-末端RuvC样结构域并且不切割靶核酸、或以显著低于参考Cas9分子的切割活性(例如,低于约20%、约10%、约5%、约1%或约0.1%)的效率进行切割,如通过本发明所述的测定所测量的。参考Cas9分子可以是天然存在的未经修饰的Cas9分子,例如天然存在的Cas9分子,如化脓链球菌、金黄色葡萄球菌或嗜热链球菌的Cas9分子。在某些实施方式中,参考Cas9分子是具有最接近序列一致性或同源性的天然存在的Cas9分子。

[0837] 在某些实施方式中,eaCas9分子或eaCas9多肽包含无活性、或无切割能力的HNH结构域和有活性、或有切割能力的N-末端RuvC样结构域(例如,本发明所述的N-末端RuvC样结构域,例如SEQ ID N0:15-23)。示例性的无活性或无切割能力的HNH样结构域可具有在以下一处或多处的突变:HNH样结构域中的组氨酸(例如,在披露于图2A-2G中的共有序列的位置856处所示的组氨酸,例如,可被丙氨酸取代);以及HNH样结构域中的一个或多个天冬酰胺(例如,在披露于图2A-2G中的共有序列的位置870处和/或披露于图2A-2G中的共有序列的位置879处所示的天冬酰胺,例如,可被丙氨酸取代)。在某些实施方式中,eaCas9与野生型的区别在于HNH样结构域并且不切割靶核酸、或以显著低于参考Cas9分子的切割活性(例如,低于约20%、约10%、约5%、约1%或约0.1%)的效率进行切割,如通过本发明所述的测定所测量的。参考Cas9分子可以是天然存在的未经修饰的Cas9分子,例如天然存在的Cas9分子,如化脓链球菌、金黄色葡萄球菌或嗜热链球菌的Cas9分子。在某些实施方式中,参考Cas9分子是具有最接近序列一致性或同源性的天然存在的Cas9分子。

[0838] 在某些实施方式中,示例性Cas9活性包括PAM特异性、切割活性、和解旋酶活性中的一项或多项。一个或多个突变可以存在于,例如:一个或多个RuvC样结构域(例如,N-末端RuvC样结构域);HNH结构域;RuvC结构域和HNH结构域之外的区域中。在某些实施方式中,一个或多个突变存在于RuvC结构域中。在某些实施方式中,一个或多个突变存在于HNH结构域中。在某些实施方式中,突变存在于RuvC结构域和HNH结构域二者中。

[0839] 可参照化脓链球菌Cas9序列在RuvC结构域中进行的示例性突变包括:D10A、E762A和/或D986A。可参照化脓链球菌Cas9序列在HNH结构域中进行的示例性突变包括:H840A、N854A和/或N863A。可参照金黄色葡萄球菌Cas9序列在RuvC结构域中进行的示例性突变包

括:D10A(参见,例如,SEQ ID N0:10)。可参照金黄色葡萄球菌Cas9序列在HNH结构域中进行的示例性突变包括:N580A(参见,例如,SEQ ID N0:11)。

[0840] 无论具体序列(例如,取代)是否可以影响一种或多种活性(如靶向活性、切割活性等),例如可以通过评价所述突变是否是保守的来评价或预测。在某些实施方式中,“非必需”氨基酸残基,如在Cas9分子的背景下所使用的,是可以改变自Cas9分子的野生型序列(例如,天然存在的Cas9分子(例如,eaCas9分子))的残基,不会消除或更优选地不会实质上改变Cas9活性(例如,裂解活性),而改变“必需”氨基酸残基导致活性(例如,裂解活性)的实质性损失。

[0841] 在某些实施方式中,Cas9分子或Cas9多肽包含不同于天然存在的Cas9分子(例如,不同于具有最接近同源性的天然存在的Cas9分子)的切割特性。例如,Cas9分子可不同于天然存在的Cas9分子,例如,金黄色葡萄球菌或化脓性链球菌的Cas9分子,如下:例如,与天然存在的Cas9分子(例如,金黄色葡萄球菌或化脓性链球菌的Cas9分子)相比,其调节(例如,减少或增加)的双链断裂(内切核酸酶和/或核酸外切酶活性)的能力;与天然存在的Cas9分子(例如,金黄色葡萄球菌或化脓性链球菌的Cas9分子)相比,其调节(例如,降低或增加)单链核酸(例如,核酸分子的非互补链或核酸分子的互补链)的切割能力(切口酶活性);或可消除切割核酸分子(例如,双链或单链核酸分子)的能力。在某些实施方式中,切口酶是包括序列SEQ ID N0:10 (D10A) 或SEQ ID N0:11 (N580A) (Friedland 2015) 的金黄色葡萄球菌Cas9衍生的切口酶。

[0842] 在某些实施方式中,改变的Cas9分子是包含以下一种或多种活性的eaCas9分子:与RuvC结构域相关的切割活性;与HNH结构域相关的切割活性;与HNH结构域相关的切割活性和与RuvC结构域相关的切割活性。

[0843] 在某些实施方式中,改变的Cas9分子或Cas9多肽包含如下序列,其中:

[0844] 对应于披露于图2A-2G中的共有序列的固定序列的序列与披露于图2A-2G中的共有序列中的固定残基的区别不多于约1%、约2%、约3%、约4%、约5%、约10%、约15%或约20%;并且

[0845] 对应于披露于图2A-2G中的共有序列中由“\*”鉴定的残基的序列与来自天然存在的Cas9分子(例如,化脓链球菌、嗜热链球菌、变形链球菌或无害利斯特菌Cas9分子)的对应序列的“\*”残基的区别不多于约1%、约2%、约3%、约4%、约5%、约10%、约15%、约20%、约25%、约30%、约35%或约40%。

[0846] 在某些实施方式中,改变的Cas9分子或Cas9多肽是包括披露于图2A-2G中的化脓链球菌Cas9的氨基酸序列的eaCas9分子或eaCas9多肽,其中在由披露于图2A-2G中的共有序列中的“\*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100或200个氨基酸残基)处具有不同于化脓链球菌的序列的一个或多个氨基酸(例如,取代)。

[0847] 在某些实施方式中,改变的Cas9分子或Cas9多肽是包括披露于图2A-2G中的嗜热链球菌Cas9的氨基酸序列的eaCas9分子或eaCas9多肽,其中在由披露于图2A-2G中的共有序列中的“\*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100或200个氨基酸残基)处具有不同于嗜热链球菌的序列的一个或多个氨基酸(例如,取代)。

[0848] 在某些实施方式中,改变的Cas9分子或Cas9多肽是包括披露于图2A-2G中的变形链球菌Cas9的氨基酸序列的eaCas9分子或eaCas9多肽,其中在由披露于图2A-2G中的共有序列中的“\*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100或200个氨基酸残基)处具有不同于变形链球菌的序列的一个或多个氨基酸(例如,取代)。

序列中的“\*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100或200个氨基酸残基)处具有不同于变形链球菌的序列的一个或多个氨基酸(例如,取代)。

[0849] 在某些实施方式中,改变的Cas9分子或Cas9多肽是包括披露于图2A-2G中的无害利斯特菌Cas9的氨基酸序列的eaCas9分子或eaCas9多肽,其中在由披露于图2A-2G中的共有序列中的“\*”表示的一个或多个残基(例如,2、3、5、10、15、20、30、50、70、80、90、100或200个氨基酸残基)处具有不同于无害利斯特菌的序列的一个或多个氨基酸(例如,取代)。

[0850] 在某些实施方式中,改变的Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)可以是例如多种不同Cas9分子(例如,不同物种的两种或更多种天然存在的Cas9分子)中的两种的融合体。例如,可以将一个物种的天然存在的Cas9分子的片段融合到第二物种的Cas9分子的片段上。作为实例,可以将包含N-末端RuvC样结构域的化脓链球菌的Cas9分子的片段融合到包含HNH样结构域的不同于化脓链球菌的物种(例如,嗜热链球菌)的Cas9分子的片段上。

[0851] 7.6具有改变的PAM识别或无PAM识别的Cas9

[0852] 天然存在的Cas9分子可以识别特异性PAM序列,例如如上针对例如化脓链球菌、嗜热链球菌、变形链球菌和金黄色葡萄球菌描述的PAM识别序列。

[0853] 在某些实施方式中,Cas9分子或Cas9多肽具有与天然存在的Cas9分子相同的PAM特异性。在某些实施方式中,Cas9分子或Cas9多肽具有与天然存在的Cas9分子不相关的PAM特异性、或与它与之具有最接近序列同源性的天然存在的Cas9分子不相关的PAM特异性。例如,可以改变天然存在的Cas9分子,例如以改变PAM识别,例如以改变Cas9分子或Cas9多肽识别的PAM序列以减少脱靶位点和/或改进特异性;或消除PAM识别需要。在某些实施方式中,可以改变Cas9分子或Cas9多肽,例如以增加PAM识别序列的长度和/或提高Cas9对高水平一致性(例如,gRNA与PAM序列之间的约98%、约99%或约100%匹配)的特异性,例如以减少脱靶位点和/或增加特异性。在某些实施方式中,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。在某些实施方式中,Cas9特异性需要gRNA与PAM序列之间的至少约90%、约95%、约96%、约97%、约98%、或约99%的同源性。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Cas9分子或Cas9多肽。描述了可以用于Cas9分子定向进化的示例性方法和系统(参见例如,Esvelt 2011)。例如,可通过以下所述的方法对候选Cas9分子进行评价。

[0854] 7.7尺寸优化的Cas9

[0855] 本文所描述的工程化的Cas9分子和工程化的Cas9多肽包括包含减小分子的尺寸但仍保留所希望Cas9特性(例如,基本上天然的构象、Cas9核酸酶活性、和/或靶核酸分子识别)的缺失的Cas9分子或Cas9多肽。本文提供了包含一个或多个缺失和任选地一个或多个接头的Cas9分子或Cas9多肽,其中接头被布置于在所述缺失的侧翼的氨基酸残基之间。用于鉴定参考Cas9分子中的适合缺失的方法、用于产生具有缺失和接头的Cas9分子的方法、以及使用此类Cas9分子的方法在审查本文献后对于本领域的普通技术人员应是清楚的。

[0856] 具有缺失的Cas9分子(例如,金黄色葡萄球菌或化脓链球菌Cas9分子)比对应的天然存在的Cas9分子小,例如具有减少数目的氨基酸。Cas9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。Cas9分子可以包含一个或多个不会实质上影响或降低本文所描述的所得Cas9分子的活性的缺失。包含如本文所描述的缺失的Cas9

分子中所保留的活性包括以下项中的一种或多种：

[0857] 切口酶活性,即切割核酸分子的单链(例如,非互补链或互补链)的能力;双链核酸酶活性,即切割双链核酸的两条链并且产生双链断裂的能力,其在某些实施方式中是在两种切口酶活性的存在下;

[0858] 内切核酸酶活性;

[0859] 外切核酸酶活性;

[0860] 解旋酶活性,即解旋双链核酸的螺旋结构的能力;

[0861] 以及核酸分子(例如,靶核酸或gRNA)的识别活性。

[0862] 可以使用本文所述的或本领域的活性测定来评估本文所述的Cas9分子的活性。

[0863] 7.8 鉴定适于缺失的区域

[0864] 可通过多种方法鉴定Cas9分子的适于缺失的区域。可在化脓链球菌Cas9的晶体结构上建模来自不同细菌物种的天然存在的直向同源Cas9分子(Nishimasu2014),以便相对于所述蛋白的三维构象跨所选的Cas9直向同源物检查保守水平。在空间定位上远离在Cas9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不实质上影响或降低Cas9活性的候选物的区域或结构域。

[0865] 7.9 编码Cas9分子的核酸

[0866] 本文提供了编码Cas9分子或Cas9多肽(例如,eaCas9分子或eaCas9多肽)的核酸。先前已经描述了编码Cas9分子或Cas9多肽的示例性核酸(参见例如,Cong 2013;Wang 2013;Mali 2013;Jinek 2012)。

[0867] 在某些实施方式中,编码Cas9分子或Cas9多肽的核酸可以是合成核酸序列。例如,合成核酸分子可以进行化学修饰,例如如本文所述。在某些实施方式中,Cas9 mRNA具有以下一种或多种(例如,所有)特性:它被5-甲基胞苷和/或假尿苷加帽、聚腺苷酸化、取代。

[0868] 另外或替代地,可对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取代。例如,合成的核酸可以指导优化的信使mRNA的合成(例如,针对在哺乳动物表达系统(例如,本文描述的)中的表达进行优化)。

[0869] 另外或替代地,编码Cas9分子或Cas9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。

[0870] 编码化脓链球菌的Cas9分子的示例性密码子优化的核酸序列示于SEQ ID N0:3中。化脓链球菌Cas9分子的对应氨基酸序列示于SEQ ID N0:2中。在某些实施方式中,化脓链球菌Cas9分子是化脓链球菌Cas9变体。在某些实施方式中,化脓性链球菌Cas9变体是具有SEQ ID N0:208所示序列的EQR变体。在某些实施方式中,化脓性链球菌Cas9变体是具有SEQ ID N0:209所示序列的VRER变体。

[0871] 编码金黄色葡萄球菌Cas9分子的示例性密码子优化的核酸序列示于SEQ ID N0:7-9,206和207中。在某些实施方式中,Cas9分子是包括D10A突变的突变金黄色葡萄球菌Cas9分子。在某些实施方式中,包括D10A突变的突变金黄色葡萄球菌Cas9分子具有SEQ ID N0:10所示的序列。在某些实施方式中,Cas9分子是包括N580突变的突变金黄色葡萄球菌Cas9分子。在某些实施方式中,包括N580突变的突变金黄色葡萄球菌Cas9分子具有SEQ ID N0:11所示的序列。金黄色葡萄球菌Cas9分子的氨基酸序列示于SEQ ID N0:6中。

[0872] 如果任何上述Cas9序列与肽或多肽在C-末端处融合,则应理解的是终止密码子将

被去除。

[0873] 7.10 其他Cas分子和Cas多肽

[0874] 不同类型的Cas分子或Cas多肽可以用来实践本文所披露的发明。在某些实施方式中, 使用II型Cas系统的Cas分子。在某些实施方式中, 使用其他Cas系统的Cas分子。例如, 可以使用I型或III型Cas分子。先前已经描述了示例性Cas分子(和Cas系统)(参见例如, Haft 2005和Makarova 2011)。示例性Cas分子(和Cas系统)也示于表13中。

[0875]

表 13: Cas 系统					
基 因 名称 <sup>‡</sup>	系 统 类型 或 亚 型	来 自 2005 <sup>§</sup> 的 名称	Haft 所 编 码 蛋 白 的 结 构 ( PDB 登录 ) <sup>¶</sup>	所 编 码 蛋 白 的 家 族 ( 和 超 家 族 ) <sup>##</sup>	代 表
<i>cas1</i>	• I型 • II型 • III型	<i>cas1</i>	3GOD、3LFX 和 2YZS	COG1518	SERP2463、SPy1047 和 <i>ygbT</i>

表 13: Cas 系统

[0876]	<i>cas2</i>	• I型 • II型 • III型	<i>cas2</i>	2IVY、2I8E 和 3EXC	COG1343 和 COG3512	SERP2462、SPy1048、SPy1723 (N-末端结构域) 和 <i>ygbF</i>
	<i>cas3</i> ′	• I型 <sup>‡‡</sup>	<i>cas3</i>	NA	COG1203	APE1232 和 <i>ygcB</i>
	<i>cas3</i> ′	• 亚型 I-A • 亚型 I-B	NA	NA	COG2254	APE1231 和 BH0336
	<i>cas4</i>	• 亚型 I-A • 亚型 I-B • 亚型 I-C • 亚型 I-D • 亚型 II-B	<i>cas4</i> 和 <i>csa1</i>	NA	COG1468	APE1239 和 BH0340
	<i>cas5</i>	• 亚型 I-A • 亚型 I-B • 亚型 I-C • 亚型 I-E	<i>Cas5a</i> 、 <i>cas5d</i> 、 <i>cas5e</i> 、 <i>cas5h</i> 、 <i>cas5p</i> 、 <i>cas5t</i> 、和 <i>cmx5</i>	3KG4	COG1688 (RAMP)	APE1234、 BH0337、 <i>devS</i> 和 <i>ygcI</i>
	<i>cas6</i>	• 亚型 I-A • 亚型 I-B • 亚型 I-D • 亚型 III-A • 亚型 III-B	<i>cas6</i> 和 <i>cmx6</i>	3I4H	COG1583 和 COG5551 (RAMP)	PF1131 和 slr7014
	<i>cas6e</i>	• 亚型 I-E	<i>cse3</i>	1WJ9	(RAMP)	<i>ygcH</i>
	<i>cas6f</i>	• 亚型 I-F	<i>csy4</i>	2XLJ	(RAMP)	<i>yI727</i>
	<i>cas7</i>	• 亚型 I-A • 亚型 I-B • 亚型 I-C • 亚型 I-E	<i>csa2</i> 、 <i>csd2</i> 、 <i>cse4</i> 、 <i>csh2</i> 、 <i>csp1</i> 、和 <i>cst2</i>	NA	COG1857 和 COG3649 (RAMP)	<i>devR</i> 和 <i>ygcJ</i>
	<i>cas8a1</i>	• 亚型 I-A <sup>‡‡</sup>	<i>cmx1</i> 、 <i>cst1</i> 、 <i>csx8</i> 、 <i>csx13</i> 、和 CXXC-CXXC	NA	BH0338 样	LA3191 <sup>§§</sup> 和 PG2018 <sup>§§</sup>

表 13: Cas 系统

[0877]	<i>cas8a2</i>	• 亚型 I-A <sup>‡‡</sup>	<i>csa4</i> 和 <i>csx9</i>	NA	PH0918	AF0070、 AF1873、 MJ0385、 PF0637、 PH0918 和 SSO1401
	<i>cas8b</i>	• 亚型 I-B <sup>‡‡</sup>	<i>csh1</i> 和 TM1802	NA	BH0338 样	MTH1090 和 TM1802
	<i>cas8c</i>	• 亚型 I-C <sup>‡‡</sup>	<i>csd1</i> 和 <i>csp2</i>	NA	BH0338 样	BH0338
	<i>cas9</i>	• II 型 <sup>‡‡</sup>	<i>csn1</i> 和 <i>csx12</i>	NA	COG3513	FTN_0757 和 SPy1046
	<i>cas10</i>	• III 型 <sup>‡‡</sup>	<i>cmr2</i> 、 <i>csm1</i> 和 <i>csx11</i>	NA	COG1353	MTH326、 Rv2823c <sup>§§</sup> 和 TM1794 <sup>§§</sup>
	<i>cas10d</i>	• 亚型 I-D <sup>‡‡</sup>	<i>csc3</i>	NA	COG1353	slr7011
	<i>csy1</i>	• 亚型 I-F <sup>‡‡</sup>	<i>csy1</i>	NA	y1724 样	y1724
	<i>csy2</i>	• 亚型 I-F	<i>csy2</i>	NA	(RAMP)	y1725
	<i>csy3</i>	• 亚型 I-F	<i>csy3</i>	NA	(RAMP)	y1726
	<i>cse1</i>	• 亚型 I-E <sup>‡‡</sup>	<i>cse1</i>	NA	YgcL 样	ygcL
	<i>cse2</i>	• 亚型 I-E	<i>cse2</i>	2ZCA	YgcK 样	ygcK
	<i>csc1</i>	• 亚型 I-D	<i>csc1</i>	NA	alr1563 样 (RAMP)	alr1563
	<i>csc2</i>	• 亚型 I-D	<i>csc1</i> 和 <i>csc2</i>	NA	COG1337 (RAMP)	slr7012
	<i>csa5</i>	• 亚型 I-A	<i>csa5</i>	NA	AF1870	AF1870、 MJ0380、 PF0643 和 SSO1398
	<i>csn2</i>	• 亚型 II-A	<i>csn2</i>	NA	SPy1049 样	SPy1049
	<i>csm2</i>	• 亚型 III-A <sup>‡‡</sup>	<i>csm2</i>	NA	COG1421	MTH1081 和 SERP2460
	<i>csm3</i>	• 亚型 III-A	<i>csc2</i> 和 <i>csm3</i>	NA	COG1337 (RAMP)	MTH1080 和 SERP2459
	<i>csm4</i>	• 亚型 III-A	<i>csm4</i>	NA	COG1567 (RAMP)	MTH1079 和 SERP2458
	<i>csm5</i>	• 亚型 III-A	<i>csm5</i>	NA	COG1332 (RAMP)	MTH1078 和 SERP2457

表 13: Cas 系统

<i>csm6</i>	• 亚型 III-A	APE2256 和 <i>csm6</i>	2WTE	COG1517	APE2256 和 SSO1445
<i>cmr1</i>	• 亚型 III-B	<i>cmr1</i>	NA	COG1367 (RAMP)	PF1130
<i>cmr3</i>	• 亚型 III-B	<i>cmr3</i>	NA	COG1769 (RAMP)	PF1128
<i>cmr4</i>	• 亚型 III-B	<i>cmr4</i>	NA	COG1336 (RAMP)	PF1126
<i>cmr5</i>	• 亚型 III-B <sup>††</sup>	<i>cmr5</i>	2ZOP 和 2OEB	COG3337	MTH324 和 PF1125
<i>cmr6</i>	• 亚型 III-B	<i>cmr6</i>	NA	COG1604 (RAMP)	PF1124
<i>csb1</i>	• 亚型 I-U	GSU0053	NA	(RAMP)	Balac_1306 和 GSU0053
<i>csb2</i>	• 亚型 I-U <sup>§§</sup>	NA	NA	(RAMP)	Balac_1305 和 GSU0054
<i>csb3</i>	• 亚型 I-U	NA	NA	(RAMP)	Balac_1303 <sup>§§</sup>
<i>csx17</i>	• 亚型 I-U	NA	NA	NA	Btus_2683
<i>csx14</i>	• 亚型 I-U	NA	NA	NA	GSU0052
<i>csx10</i>	• 亚型 I-U	<i>csx10</i>	NA	(RAMP)	Caur_2274
<i>csx16</i>	• 亚型 III-U	VVA1548	NA	NA	VVA1548
<i>csaX</i>	• 亚型 III-U	<i>csaX</i>	NA	NA	SSO1438
<i>csx3</i>	• 亚型 III-U	<i>csx3</i>	NA	NA	AF1864
<i>csx1</i>	• 亚型 III-U	<i>csa3</i> 、 <i>csx1</i> 、 <i>csx2</i> 、 DXTHG、 NE0113 和 TIGR0271 0	1XMX 和 2I71	COG1517 和 COG4006	MJ1666、 NE0113、 PF1127 和 TM1812
<i>csx15</i>	• 未知	NA	NA	TTE2665	TTE2665
<i>csf1</i>	• U型	<i>csf1</i>	NA	NA	AFE_1038
<i>csf2</i>	• U型	<i>csf2</i>	NA	(RAMP)	AFE_1039
<i>csf3</i>	• U型	<i>csf3</i>	NA	(RAMP)	AFE_1040
<i>csf4</i>	• U型	<i>csf4</i>	NA	NA	AFE_1037

[0878]

8. 候选分子的功能分析

[0880] 可通过领域已知的方法或如本发明所述的方法来评价候选Cas9分子、候选gRNA分

子、候选Cas9分子/gRNA分子复合物。例如,先前已经描述了用于评价Cas9分子的内切核酸酶活性的示例性方法(Jinek 2012)。

[0881] 8.1融合和切割测定:测试Cas9核酸内切酶活性

[0882] 可在质粒切割测定中对Cas9分子/gRNA分子复合物结合至并且切割靶核酸的能力进行评价。在这个测定中,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Cas9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Cas9质粒切割缓冲液(20mM HEPES pH 7.5、150mM KC1、0.5mM DTT、0.1mM EDTA)中孵育60分钟。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Cas9分子是否切割两条DNA链、或仅切割两条链中的一条。例如,线性DNA产物指示两条DNA链的切割。有缺口的开放圆形产物指示两条链中只有一条被切割。

[0883] 可替代地,可以在寡核苷酸DNA切割测定中对Cas9分子/gRNA分子复合物结合至并且切割靶核酸的能力进行评价。在这个测定中,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)[ $\gamma$ -32P]-ATP孵育30分钟,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3分钟,随后缓慢冷却至室温而生成双链体底物(100nM)。对于切割测定,通过加热至95°C持续30秒,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Cas9(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 KC1、5mM MgCl<sub>2</sub>、1mM DTT、5%甘油)中进行预孵育。通过添加1μL靶DNA(10nM)开始反应并在37°C下孵育1小时。将反应通过添加20μL的加样染料(5mM EDTA、0.025%SDS、5%甘油,在甲酰胺中)淬灭并加热至95°C持续5分钟。将切割产物在含有7M尿素的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、或两者是否被切割。

[0884] 这些测定中的一个或两个可以用于评价候选gRNA分子或候选Cas9分子的适用性。

[0885] 8.2结合测定:测试Cas9分子与靶DNA的结合

[0886] 先前已经描述了用于评价Cas9分子与靶DNA的结合的示例性方法,例如,Jinek等人,Science 2012;337(6096):816-821。

[0887] 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3分钟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,裁切,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[ $\gamma$ -32P]-ATP进行5'端标记持续30分钟。将多核苷酸激酶在65°C下热变性持续20分钟,并且使用柱去除未结合的放射性标记。在10μL的总体积中,在含有20mM HEPES pH 7.5、100mM KC1、5mM MgCl<sub>2</sub>、1mM DTT以及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Cas9蛋白分子进行程序化,并且从100pM滴定至1μM。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1小时并且在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

并且通过感光成像进行DNA可视化。

[0888] 8.3差示扫描荧光测定法 (DSF)

[0889] 可以通过差示扫描荧光测定法 (DSF) 来检测Cas9-gRNA核糖核蛋白 (RNP) 复合物的热稳定性。该技术测量蛋白质的热稳定性,蛋白质的热稳定性可以在有利条件下增加,例如添加结合RNA分子,例如gRNA。

[0890] 使用两种不同的方案进行测定,一种测试gRNA:Cas9蛋白的最佳化学计量比,另一种测定RNP形成的最佳溶液条件。

[0891] 为了确定形成RNP复合物的最佳溶液,将水中的Cas9的2μM溶液与10x SYPRO Orange® (生命技术公司 (Life Technologies) 目录#S-6650) 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分钟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 C1000 Touch™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°C。

[0892] 第二个测定由在来自上述测定1的最优缓冲液中混合不同浓度的gRNA与2μM Cas9 并在384孔板中于RT下孵育10分钟组成。添加等体积的最适缓冲液与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。在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 C1000 Touch™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°。

[0893] 9.基因组编辑方法

[0894] 本发明所述是用于(例如,使用本发明所述方法或途径中的一种或多种,例如,使用NHEJ)靶向改变(例如,敲除)RS1、RL2和/或LAT基因(例如,RS1、RL2和/或LAT基因中的一个或两个等位基因)的组合物、基因组编辑系统和方法。本发明所述也是用于被靶向敲低的RS1、RL2和/或LAT基因的方法。

[0895] 9.1用于基因靶向的NHEJ方法

[0896] 在本发明提供的方法中的某些实施方式中,NHEJ介导的改变被用来改变HSV RS1靶标、HSV RL2靶标或HSV LAT靶标。如本发明所述,核酸酶诱导的非同源末端连接 (NHEJ) 可被用来靶向基因特异性敲除。核酸酶诱导的NHEJ还可被用来去除(例如,删除)感兴趣基因中的序列插入。

[0897] 在某些实施方式中,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以及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,只有两个相容末端(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末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。这是实施方式依赖性的,因为关键功能结构域中的突变可能比蛋白质的非关键区中的突变耐受性低。

[0898] 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;它们最常见地在1-50bp范围内,但是可以达到大于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复。然而,有可能获得大插入,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。

[0899] 因为NHEJ是诱变的过程,所以它还可以用于缺失小序列基序(例如,长度上小于或等于50个核苷酸的基序),只要不需要产生特异性最终序列。如果双链断裂被靶向靶序列附近,则由NHEJ修复导致的缺失突变常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。以这种方式,可以缺失大至几百千碱基的DNA区段。这两种方法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产生indel突变。

[0900] 双链切割性eaCas9分子和单链、或切口酶,eaCas9分子均可以用于本文所描述的方法和组合物以产生NHEJ介导的indel。靶向感兴趣基因的早期编码区的NHEJ介导的indel可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之内、或在起始位点的500bp之内(例如,小于500、450、400、350、300、250、200、150、100或50bp)的序列。

[0901] 9.1.1相对于靶标放置双链或单链断裂

[0902] 在某些实施方式中,其中gRNA和Cas9核酸酶为了诱导NHEJ介导的插入缺失(indel)而产生双链断裂,将gRNA(例如单分子(或嵌合)或模块化gRNA分子)配置为在靠近目标位置的核苷酸处定位一个双链断裂。在某些实施方式中,该切割位点距靶标0-30bp之间(例如,距靶标小于30、25、20、15、10、9、8、7、6、5、4、3、2或1bp位置)。

[0903] 在某些实施方式中,其中两个与Cas9切口酶复合的gRNA为了诱导NHEJ介导的indel的目的而诱导两条单链断裂,将两个gRNA例如独立地单分子(或嵌合)或模块化gRNA构造成将两个单链断裂定位以提供NHEJ修复靶位点的核苷酸。在某些实施方式中,gRNA被构造成在不同链上的相同位置或彼此的几个核苷酸内进行基本模拟双链断裂的切割。在某些实施方式中,更接近的切口位于离靶标0-30bp之间(例如,距靶标小于30、25、20、15、10、9、8、7、6、5、4、3、2或1bp位置),且两个切口在彼此的25-55bp内(例如,25至50、25至45、25至40、25至35、25至30、50至55、45至55、40至55、35至55、30至55、30至50,35至50、40至50、45至50、35至45或40至45bp)并且彼此不超过100bp(例如,不超过90、80、70、60、50、40、30、20或10bp)。在某些实施方式中,gRNA被构造成在靶标的核苷酸的任一侧上放置单链断裂。

[0904] 双链切割性eaCas9分子和单链、或切口酶,eaCas9分子均可以用于本文所描述的方法和组合物以在靶标的两侧处产生断裂。可在靶标的两侧产生双链或成对单链断裂以去除两个切割之间的核酸序列(例如,两个断裂之间的区域被删除)。在某些实施方式中,两个gRNA(例如独立的单分子(或嵌合)或模块化gRNA)被构造成在靶标的两侧定位双链断裂。在一个替代实施方式中,例如独立地、单分子(或嵌合)或模块化gRNA的三种gRNA被构造成在靶标的任一侧定位双链断裂(即,一种与cas9核酸酶的gRNA复合物)和两个单链断裂或成对单链断裂(即,两个gRNA与Cas9切口酶复合)。在某些实施方式中,四个gRNA(例如独立的单分子(或嵌合)或模块化gRNA)被构造成在靶标的任一侧产生两对单链断裂(即两对与Cas9切口酶复合的两个gRNA)。理想情况下,双链断裂或两个单链缺口较近的一对可位于靶标的0-500bp内(例如,距目标位置不超过450、400、350、300、250、200、150、100、50或25bp)。当使用切口酶时,一对中的两个切口在彼此的25-55bp内(例如,25至50、25至45、25至40、25至

35、25至30、50至55、45至55、40至55、35至55、30至55、30至50,35至50、40至50、45至50、35至45或40至45bp)并且彼此不超过100bp(例如,不超过90、80、70、60、50、40、30、20或10bp)。

[0905] 9.2HDR修复、HDR介导的敲入和模板核酸

[0906] 在本发明提供的方法的某些实施方式中,通过使用外源提供的模板核酸(本发明也称为供体构建体),使用HDR介导的序列改变来改变RS1、RL2或LAT基因中一个或多个核苷酸的序列。在某些实施方式中,通过HDR与外源提供的供体模板或模板核酸发生HDR介导的HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变。例如,供体构建体或模板核酸提供HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变。在某些实施方式中,将质粒供体用作用于同源重组的模板。在某些实施方式中,使用单链供体模板作为通过靶序列与供体模板之间的替代HDR方法(例如,单链退火)改变HSV RS1靶标、HSV RL2靶标或HSV LAT靶标位点的模板。HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的供体模板影响的改变取决于Cas9分子的切割。经由Cas9切割可以包括双链断裂或两个单链断裂。

[0907] 在某些实施方式中,使用HDR介导的序列改变来改变RS1、RL2或LAT基因中一个或多个核苷酸的序列而不使用外源提供的模板核酸。在某些实施方式中,HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变通过具有内源基因组供体序列的HDR发生。例如,内源基因组供体序列提供RL2、LAT或RS1靶位置的改变。在某些实施方式中,内源基因组供体序列位于与靶序列相同的染色体上。在某些实施方式中,内源基因组供体序列位于与靶序列不同的染色体上。通过内源基因组供体序列改变HSV RS1靶标、HSV RL2靶标或HSV LAT靶标取决于Cas9分子的切割。经由Cas9切割可以包括双链断裂或两个单链断裂。

[0908] 在本发明提供的方法的某些实施方式中,HDR介导的改变用于改变RL2、LAT或RS1基因中的单个核苷酸。这些实施方式可利用一个双链断裂或两个单链断裂。在某些实施方式中,单核苷酸改变的并入是使用:(1)一个双链断裂,(2)两个单链断裂,(3)两个双链断裂,其中断裂发生在靶位置的每一侧,(4)一个双链断裂和两个单链断裂,其中双链断裂和两个单链断裂发生在靶位置的每一侧,(5)四个单链断裂,其中一对单链断裂发生在靶位置的每一侧,或(6)一个单链断裂。

[0909] 在使用单链模板核酸(例如,供体模板)的某些实施方式中,靶位置可以通过替代的HDR来改变。

[0910] HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的供体模板影响的改变取决于Cas9分子的切割。通过Cas9的切割可包括缺口、双链断裂或两个单链断裂,例如,靶核酸每条链上的一个断裂。在靶核酸上引入断裂之后,在断裂端发生切除,产生单链的突出DNA区域。

[0911] 在典型HDR中,引入包含与靶核酸同源的序列的双链供体模板,其可直接掺入靶核酸中或用作模板以改变靶核酸序列。在断裂处切除后,修复可以通过不同的途径进行,例如通过双霍利迪连接模型(或双链断裂修复(DSBR)途径)或合成依赖性链退火(SDSA)途径。在双霍利迪连接模型中,发生由靶核酸的两个单链突出端进行链入侵到供体模板中的同源序列,导致形成具有两个霍利迪连结的中间体。连接点随着入侵链的末端合成新的DNA以填补由切除产生的空位而迁移。新合成的DNA的末端连接至切除末端,且连接处被解离,导致靶核酸的改变。与供体模板的交叉可能会在连接点分解时发生。在SDSA途径中,只有一个单链突出端侵入到供体模板,并从入侵链的末端合成新的DNA以填补由切除产生的空位。然后新合成的DNA与剩余的单链突出端退火,合成新的DNA以填补空位,并将链连接以产生改变的

DNA双链体。

[0912] 在可替代的HDR中,引入单链供体模板,例如模板核酸。用于改变所希望的靶标的靶核酸处的切口、单链断裂或双链断裂由例如本文所述的Cas9分子介导,并且在断裂处发生切除以显示单链突出端。如上所述,掺入模板核酸序列以改变HSV RS1靶标、HSV RL2靶标或HSV LAT靶标通常通过SDSA途径发生。

[0913] 在国际申请PCT/US2014/057905中的标题为“模板核酸”的第IV节中提供了关于模板核酸的另外的细节。

[0914] 在某些实施方式中,通过具有与HNH-样结构域相关的切割活性和与RuvC样结构域(例如,N端RuvC样结构域,例如野生型Cas9)相关的切割活性的Cas9分子实现双链切割。这些实施方式仅需要单个gRNA。

[0915] 在某些实施方式中,一个单链断裂或切口由具有切口酶活性的Cas9分子(例如,本发明所述的Cas9切口酶(比如,D10A Cas9切口酶))实现。带切口的靶核酸可以是alt-HDR的底物。

[0916] 在某些实施方式中,两个单链断裂或缺口由具有切口酶活性(例如,与HNH-样结构域相关的切割活性或与N端RuvC样结构域相关的切割活性)的Cas9分子实现。这些实施方式通常需要两个gRNA,一个用于每个单链断裂的放置。在某些实施方式中,具有切口酶活性的Cas9分子切割与gRNA杂交的链,但不切割与gRNA杂交的链互补的链。在某些实施方式中,具有切口酶活性的Cas9分子不切割与gRNA杂交的链,但相反切割与gRNA杂交的链互补的链。

[0917] 在某些实施方式中,切口酶具有HNH活性,例如,具有失活的RuvC活性的Cas9分子,例如,在D10处具有突变的Cas9分子,例如,D10A突变(参见,例如SEQ ID NO:10)。D10A使RuvC失活;因此,Cas9切口酶具有(仅)HNH活性并可切割gRNA杂交的链(例如,互补链,其上不具有NGG PAM)。在某些实施方式中,具有H840(例如H840A)突变的Cas9分子可用作切口酶。H840A使HNH失活;因此,Cas9切口酶具有(仅)RuvC活性并切割非互补链(例如,具有NGG PAM且其序列与gRNA相同的链)。在某些实施方式中,具有N863(例如,N863A突变)突变的Cas9分子可用作切口酶。N863A使HNH失活;因此,Cas9切口酶具有(仅)RuvC活性并切割非互补链(具有NGG PAM且其序列与gRNA相同的链)。在某些实施方式中,具有N580(例如,N580A突变)突变的Cas9分子可用作切口酶。N580A使HNH失活;因此,Cas9切口酶具有(仅)RuvC活性并切割非互补链(具有NGG PAM且其序列与gRNA相同的链)。

[0918] 在某些实施方式中,其中使用切口酶和两种gRNA来定位两个单链切口,一个切口位于+链上,一个切口位于靶核酸的-链上。PAM可面向外部。可选择gRNA使得gRNA分开约0-50、0-100或0-200个核苷酸。在某些实施方式中,在与两种gRNA的靶向结构域互补的靶序列之间没有重叠。在某些实施方式中,gRNA不重叠并分开多达50、100或200个核苷酸。在某些实施方式中,使用两种gRNA可增加特异性,例如,通过减少脱靶融合(Ran 2013)。

[0919] 在某些实施方式中,可使用单个切口来诱导HDR,例如,alt-HDR。在某些实施方式中,可使用单个切口来增加给定切割位点处HR与NHEJ的比率。在某些实施方式中,单链断裂形成在所述靶核酸中与所述gRNA的靶向结构域互补的链中。在某些实施方式中,单链断裂形成在所述靶核酸中不是与所述gRNA的靶向结构域互补的链中。

[0920] 9.2.1相对于靶标放置双链或单链断裂

[0921] 其中一条链中的双链断裂或单链断裂应该足够接近HSV RS1靶标、HSV RL2靶标或

HSV LAT靶标,以便在所需区域产生改变。在某些实施方式中,该距离不超过50、100、200、300、350或400个核苷酸。在某些实施方式中,断裂应该足够接近靶标,使得靶标在末端切除期间位于受到核酸外切酶介导的去除的区域内。如果HSV RS1靶标、HSV RL2靶标或HSV LAT靶标与断裂之间的距离太大,则期望改变的序列可能不包括在末端切除中,因此可能不会被改变来作为供体序列。外源提供的供体序列或内源性基因组供体序列,在某些实施方式中,仅用于改变末端切除区域内的序列。

[0922] 在某些实施方式中,本发明所述的方法在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标附近引入一个或多个断裂。在这些中的某些实施方式中,在HSV RS1靶标、HSV RL2靶标或HSV LAT靶标两侧引入两个或多个断裂。该两个或多个断裂去除了(例如,缺失)包括HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的基因组序列。所有本发明所述方法都在RS1、RL2或LAT基因中产生对HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的改变。

[0923] 在某些实施方式中,gRNA靶向结构域被构造成使切割事件(例如,双链断裂或单链断裂)位于期望改变的区域(例如,突变)的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。断裂(例如,双链或单链断裂)可位于期望改变的区域的上游或下游,例如,突变。在某些实施方式中,断裂位于期望改变的区域内,例如,在由至少两个突变核苷酸限定的区域内。在某些实施方式中,断裂位于紧邻希望改变的区域,例如,紧邻突变的上游或下游。

[0924] 在某些实施方式中,单链断裂伴随着另一个单链断裂,所述另一个单链断裂由第二gRNA分子标位,如下所述。例如,所述靶向结构域结合被构造成使切割事件(例如,两个单链断裂)位于靶标的1、2、3、4、5、10、15、20、25、30、35、40、45、50、60、70、80、90、100、150或200个核苷酸内。在某些实施方式中,第一和第二gRNA分子被构造成使得当引导Cas9切口酶时,单链断裂可伴随另外的由第二gRNA定位的单链断裂,其足够接近彼此以导致期望区域的改变。在某些实施方式中,所述第一和第二gRNA分子被构造成使得例如在所述Cas9是切口酶时,由所述第二gRNA标位的单链断裂在由所述第一gRNA分子标位的断裂的10、20、30、40或50个核苷酸内。在某些实施方式中,所述两个gRNA分子被构造成在不同链上的相同位置或彼此的几个核苷酸内进行切割,例如,基本模拟双链断裂。

[0925] 在某些实施方式中,其中为了诱导HDR介导的序列改变的目的,gRNA(单分子(或嵌合)或模块化gRNA)和Cas9核酸酶诱导双链断裂,切割位点在远离靶标的0-200bp之间(例如,0至175、0至150、0至125、0至100、0至75、0至50、0至25、25至200、25至175、25至150、25至125、25至100、25至75、25至50、50至200、50至175、50至150、50至125、50至100、50至75、75至200、75至175、75至150、75至125、75至100个碱基对)。在某些实施方式中,该切割位点距靶标0-100bp之间(例如,0至75、0至50、0至25、25至100、25至75、25至50、50至100、50至75、75至100bp)。

[0926] 在某些实施方式中,可通过使用切口酶来产生突出的断裂来促进HDR。虽然不希望受理论束缚,但突出端的单链性质可增强细胞通过HDR修复断裂的可能性,与例如NHEJ相反。具体而言,在某些实施方式中,通过选择将第一切口酶靶向第一靶序列的第一gRNA和将第二切口酶靶向第二靶序列的第二gRNA促进HDR,所述第二靶序列位于与第一靶标相对的DNA链上并与第一个切口错开。

[0927] 在某些实施方式中,gRNA分子的所述靶向结构域被构造成将切割事件标位成足够

远离预选核苷酸从而可不改变所述核苷酸。在某些实施方式中, gRNA分子的所述靶向结构域被构造成将内含子切割事件定位在足够远离内含子/外显子边界或天然存在的剪接信号, 以避免改变外显子序列或不需要的剪接事件。如本发明所述, 所述gRNA分子可以是第一、第二、第三和/或第四gRNA分子。

[0928] 9.2.2 相对彼此的第一断裂和第二断裂的放置

[0929] 在某些实施方式中, 双链断裂可伴随着另一个双链断裂, 所述另一个单链断裂由第二gRNA分子标位, 如下所述。

[0930] 在某些实施方式中, 双链断裂可伴随着两个其他单链断裂, 所述两个其他单链断裂由第二和第三gRNA分子标位。

[0931] 在某些实施方式中, 第一和第二单链断裂可伴随着两个其他单链断裂, 所述两个其他单链断裂由第三gRNA分子和第四gRNA分子标位。

[0932] 当两个或更多个gRNA用于在靶核酸中定位两个或更多个切割事件(例如, 双链或单链断裂)时, 两个或更多个切割事件可由相同或不同的Cas9蛋白完成。例如, 当使用两个gRNA来定位两个双链断裂时, 单个Cas9核酸酶可用于产生两个双链断裂。当使用两个或更多个gRNA来定位两个或更多个单链断裂(切口)时, 可使用单个Cas9切口酶来产生两个或更多个切口。当使用两个或更多个gRNA来定位至少一个双链断裂和至少一个单链断裂时, 可使用两个Cas9蛋白, 例如, 一种Cas9核酸酶和一种Cas9切口酶。在某些实施方式中, 使用两种或更多种Cas9蛋白, 并可按顺序递送两种或更多种Cas9蛋白以控制靶核酸中期望位置的双链对单链断裂的特异性。

[0933] 在某些实施方式中, 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所述靶核酸分子的相对链互补。在某些实施方式中, 所述gRNA分子和所述第二gRNA分子被构造成使所述PAM向外朝向。

[0934] 在某些实施方式中, 选择两个gRNA以在相互预选距离的两个位置处指导Cas9介导的切割。在某些实施方式中, 两个切割点位于靶核酸的相反链上。在某些实施方式中, 两个切割点形成钝端断裂, 并在其他实施方式中, 它们被错开以使DNA末端包含一个或两个突出端(例如一个或多个5'突出端和/或一个或多个3'突出端)。在某些实施方式中, 每次切割事件是一个切口。在某些实施方式中, 切口足够接近以致它们形成被双链断裂机制识别的断裂(而不是被例如SSBr机制所识别)。在某些实施方式中, 切口足够远以至于它们产生作为HDR底物的突出端, 即断裂的放置模仿经历了一些切除的DNA底物。例如, 在某些实施方式中, 切口间隔以产生作为进行性切除的第五的突出物。在某些实施方式中, 两个断裂间隔在25-65个核苷酸之间。例如, 两个断裂彼此间隔约25、30、35、40、45、50、55、60或65个核苷酸。例如, 两个断裂彼此间隔至少约25、30、35、40、45、50、55、60或65个核苷酸。例如, 两个断裂彼此间隔至多约30、35、40、45、50、55、60或65个核苷酸。在某些实施方式中, 两个断裂彼此间隔约25-30、30-35、35-40、40-45、45-50、50-55、55-60、或60-65个核苷酸。

[0935] 在某些实施方式中, 模仿切除的断裂的断裂包括3'突出端(例如, 由DSB和切口产生, 其中切口留下3'突出端)、5'突出(例如, 由DSB和切口产生, 其中切口留下5'突出端)、3'和5'突出端(例如, 由三个切口产生)、两个3'突出端(例如, 由彼此错开的两个切口产生)或两个5'突出端(例如由彼此错开的两个切口产生)。

[0936] 在其中与Cas9切口酶复合的两个gRNA(独立地、单分子(或嵌合)或模块化gRNA)诱

导两个单链断裂以诱导HDR介导的改变的某些实施方式中,更接近的切口在远离靶标的0-200bp之间(例如,0至175、0至150、0至125、0至100、0至75、0至50、0至25、25至200、25至175、25至150、25至125、25至100、25至75、25至50、50至200、50至175、50至150、50至125、50至100、50至75、75至200、75至175、75至150、75至125或75至100bp),且两个切口理想地可在彼此的25-65bp内(例如,25至50、25至45、25至40、25至35、25至30、30至55、30至50、30至45、30至40、30至35、35至55、35至50、35至45、35至40、40至55、40至50、40至45bp、45至50bp、50至55bp、55至60bp或60至65bp)并彼此不超过100bp(例如,彼此相距不超过90、80、70、60、50、40、30、20、10或5bp)。在某些实施方式中,该切割位点距靶标0-100bp之间(例如,0至75、0至50、0至25、25至100、25至75、25至50、50至100、50至75、75至100bp)。

[0937] 在某些实施方式中,两个gRNA(例如独立的单分子(或嵌合)或模块化gRNA)被构造成在靶标的两侧定位双链断裂。在某些实施方式中,例如独立地、单分子(或嵌合)或模块化gRNA的三种gRNA被构造成在靶标的任一侧定位双链断裂(即,一种与cas9核酸酶的gRNA复合物)和两个单链断裂或成对单链断裂(即,两个gRNA与Cas9切口酶复合)。在某些实施方式中,四个gRNA(例如独立的单分子(或嵌合)或模块化gRNA)被构造成在靶标的任一侧产生两对单链断裂(即两对与Cas9切口酶复合的两个gRNA)。理想情况下,双链断裂或一对中两个单链缺口中较近的可位于靶标的0-500bp内(例如,距目标位置不超过450、400、350、300、250、200、150、100、50或25bp)。当使用切口酶时,在某些实施方式中,一对中的两个切口在彼此的25-65bp内(例如,25至55、25至50、25至45、25至40、25至35、25至30、50至55、45至55、40至55、35至55、30至55、30至50、35至50、40至50、45至50、35至45、40至45bp、45至50bp、50至55bp、55至60bp或60至65bp)并且彼此不超过100bp(例如,不超过90、80、70、60、50、60、30或20或10bp)。

[0938] 当使用两种gRNA靶向Cas9分子进行断裂时,可设想不同的Cas9分子组合。在某些实施方式中,使用第一gRNA将第一Cas9分子靶向第一靶标,使用第二gRNA靶向第二Cas9分子靶向第二靶标。在某些实施方式中,第一个Cas9分子在靶核酸的第一条链上产生切口,而第二个Cas9分子在相对的链上产生切口,导致双链断裂(例如,钝端切口或突出端切口)。

[0939] 可选择切口酶的不同组合来将一个单链断裂靶向一条链,并将第二个单链断裂靶向相反链。当选择一种组合时,可考虑有切口酶具有一个活性RuvC样结构域,且切口酶具有一个活性HNH结构域。在某些实施方式中,RuvC样结构域切割靶核酸分子的非互补链。在某些实施方式中,HNH样结构域切割双链核酸分子的单链互补结构域(例如,互补链)。一般而言,如果两个Cas9分子具有相同的活性结构域(例如,两者都具有活性RuvC结构域或两者都具有活性HNH结构域),则可选择结合靶的相对链的两种gRNA。更详细来说,在某些实施方式中,第一gRNA与靶核酸的第一链互补并结合具有活性RuvC样结构域的切口酶且引起该切口酶切割与该第一gRNA不互补的链,即靶核酸的第二链;且第二gRNA与靶核酸的第二链互补并结合具有活性RuvC样结构域的切口酶且引起该切口酶切割与该第二gRNA不互补的链,即靶核酸的第一链。相反,在某些实施方式中,第一gRNA与靶核酸的第一链互补并结合具有活性HNH结构域的切口酶且引起该切口酶切割与该第一gRNA互补的链,即靶核酸的第一链;且第二gRNA与靶核酸的第二链互补并结合具有活性HNH结构域的切口酶且引起该切口酶切割与该第二gRNA互补的链,即靶核酸的第二链。在另一种布置中,如果一个Cas9分子具有活性RuvC样结构域且另一个Cas9分子具有活性HNH结构域,则两个Cas9分子的gRNA可与靶核酸

的相同链互补,以便具有活性RuvC样结构域的Cas9分子可切割非互补链且具有HNH结构域的Cas9分子可以切割互补链,产生双链断裂。

[0940] 9.2.3 供体模板的同源臂

[0941] 同源臂应至少延伸至可能发生末端切除的区域,例如,以便允许切除的单链突出端在供体模板内找到互补区域。总长度可能受限于诸如质粒大小或病毒包装限制等参数。在某些实施方式中,同源臂不延伸成重复的元件,例如,Alu重复序列或LINE重复序列。

[0942] 示例性的同源臂长度包括至少50、100、250、500、750、1000、2000、3000、4000或5000个核苷酸。在某些实施方式中,同源臂长度为50-100、100-250、250-500、500-750、750-1000、1000-2000、2000-3000、3000-4000或4000-5000个核苷酸。

[0943] 如本发明使用的术语,模板核酸是指可与Cas9分子和gRNA分子组合使用以改变HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的结构的核酸序列。在某些实施方式中,HSV RS1靶标、HSV RL2靶标或HSV LAT靶标可以是添加了一个或多个核苷酸的目标核酸上的两个核苷酸之间的位点,例如,相邻的核苷酸。或者,HSV RS1靶标、HSV RL2靶标或HSV LAT靶标可包含一个或多个被模板核酸改变的核苷酸。

[0944] 在某些实施方式中,靶核酸被修饰以具有模板核酸的一些或全部序列,通常在切割位点处或切割位点附近。在某些实施方式中,模板核酸是单链的。在某些实施方式中,模板核酸是双链的。在某些实施方式中,模板核酸是DNA,例如,双链DNA。在某些实施方式中,模板核酸是单链DNA。在某些实施方式中,模板核酸在相同的载体骨架上编码,例如,AAV基因组、质粒DNA,如Cas9和gRNA。在某些实施方式中,模板核酸在体内从载体骨架切除,例如,其侧翼为gRNA识别序列。在某些实施方式中,模板核酸包括内源基因组序列。

[0945] 在某些实施方式中,模板核酸通过参与HDR事件来改变靶标的结构。在某些实施方式中,模板核酸改变靶标的序列。在某些实施方式中,模板核酸导致将修饰的或非天然存在的碱基掺入到靶核酸中。

[0946] 通常,模板序列经历由断裂介导的或催化的与靶序列的重组。在某些实施方式中,模板核酸包括对应于通过eaCas9介导的切割事件切割的靶序列上的位点的序列。在某些实施方式中,模板核酸包括对应于在第一Cas9介导的事件中被切割的靶序列上的第一位点和在第二Cas9介导的事件中被切割的靶序列上的第二位点的序列。

[0947] 模板核酸通常包含以下组分:

[0948] [5' 同源臂]-[替换序列]-[3' 同源臂]。

[0949] 同源臂用于重组到染色体中,从而用替换序列取代不需要的元件,例如突变或标记(signature)。在某些实施方式中,同源臂位于最远切割位点的两侧。

[0950] 在某些实施方式中,5' 同源臂的3' 末端是替换序列5' 末端旁边的位置。在某些实施方式中,5' 同源臂从替换序列的5' 末端的5' 端可延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸。

[0951] 在某些实施方式中,3' 同源臂的5' 末端是替换序列3' 末端旁边的位置。在某些实施方式中,3' 同源臂从替换序列的3' 末端的3' 端可延伸至少10、20、30、40、50、100、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸。

[0952] 在某些实施方式中,为了改变HSV RS1靶标、HSV RL2靶标或HSV LAT靶标处的一个或多个核苷酸,同源臂(例如5' 和3' 同源臂)各自可包括最远端的gRNA两侧的约1000bp的序

列(例如,HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的任一侧上的序列的1000bp)。

[0953] 在某些实施方式中,可缩短一个或两个同源臂以避免包括某些序列重复元件,例如Alu重复序列或LINE元件。例如,可缩短5'同源臂以避免序列重复元件。在某些实施方式中,可缩短3'同源臂以避免序列重复元件。在某些实施方式中,可缩短5'和3'同源臂以避免包含某些序列重复元件。

[0954] 在某些实施方式中,可设计用于改变HSV RS1靶标、HSV RL2靶标或HSV LAT靶标的序列的模板核酸作为单链寡核苷酸,例如单链寡聚脱氧核苷酸(ssODN)。当使用ssODN时,5'和3'同源臂的长度可高达约200bp,例如至少25、50、75、100、125、150、175或200bp长。随着寡核苷酸合成的不断改进,更长的同源臂也可用于ssODN。在某些实施方式中,通过化学合成以外的方法制备较长的同源臂,例如,通过使长双链核酸变性和纯化其中一条链,例如,通过对锚定于固体底物的链特异性序列的亲和力来纯化。

[0955] 在某些实施方式中,当模板核酸与切口的5'端(即切口链的5'方向)具有延伸的同源性时,alt-HDR更有效地进行。因此,在一些实施方式中,模板核酸具有较长的同源臂和较短的同源臂,其中较长的同源臂可在切口的5'端退火。在某些实施方式中,与切口5'端退火的同源臂在切口或在替换序列的5'或3'末端至少50、75、100、125、150、175、200、300、400、500、600、700、800、900、1000、1500、2000、3000、4000或5000个核苷酸处。在某些实施方式中,可与切口5'端退火的同源臂比可在切口3'端退火的同源臂长至少约10%、约20%、约30%、约40%或约50%。在某些实施方式中,可与切口5'端退火的同源臂比可在切口3'端退火的同源臂长至少2倍、3倍、4倍或5倍。取决于ssDNA模板是否可与完整链或切口链退火,与切口5'端退火的同源臂可分别位于ssDNA模板的5'末端或ssDNA模板的3'末端。

[0956] 同样,在某些实施方式中,模板核酸具有5'同源臂、替换序列和3'同源臂,使得模板核酸与切口5'端具有延伸的同源性。例如,5'同源臂和3'同源臂可具有基本相同的长度,但是替换序列从切口的5'端延伸的比从切口3'端延伸得更远。在某些实施方式中,替换序列在切口的5'端的延伸比其在切口的3'端延伸长至少约10%、约20%、约30%、约40%、约50%、2倍、3倍、4倍或5倍。

[0957] 在某些实施方式中,当模板核酸以切口为中心时,alt-HDR更有效地进行。因此,在某些实施方式中,模板核酸具有两个基本上相同大小的同源臂。例如,模板核酸的第一同源臂的长度约为模板核酸的第二同源臂长度的约10%、约9%、约8%、约7%、约6%、约5%、约4%、约3%、约2%或约1%。

[0958] 同样,在某些实施方式中,模板核酸具有5'同源臂、替换序列和3'同源臂,使得模板核酸在切口的任一侧上延伸基本上相同距离。例如,同源臂可具有不同的长度,但是可选择替换序列以补偿这一点。例如,替换序列可从切口的5'端延伸的比从切口3'端延伸得更远,但是切口的5'端同源臂比切口的3'端同源臂更短以进行补偿。相反也是可能的,例如,替换序列可从切口的3'端延伸的比从切口5'端延伸得更远,但是切口的3'端同源臂比切口的5'端同源臂更短以进行补偿。

[0959] 9.2.4模板核酸

[0960] 在某些实施方式中,模板核酸是双链的。在某些实施方式中,模板核酸是单链的。在某些实施方式中,模板核酸包括单链部分和双链部分。在某些实施方式中,模板核酸在切口和/或替换序列的任一侧上包含约50至100bp(例如,55至95、60至90、65至85或70至80bp)

的同源臂。在某些实施方式中，模板核酸包括切口或替换序列5' 端、切口或切口或替换序列的3' 端、或切口的5' 端以及3' 端或替换序列的约50、55、60、65、70、75、80、85、90、95或100bp 的同源臂。

[0961] 在某些实施方式中，模板核酸在切口和/或替换序列的3' 端上包括约150至200bp (例如,155至195、160至190、165至185或170至180bp) 的同源臂。在某些实施方式中，模板核酸在切口或替换序列的3' 端上包括约150、155、160、165、170、175、180、185、190、195或200bp 的同源臂。在某些实施方式中，模板核酸在切口或替换序列的5' 端上包括少于约100、90、80、70、60、50、40、30、20、15或10bp 的同源臂。

[0962] 在某些实施方式中，模板核酸在切口和/或替换序列的5' 端上包括约150至200bp (例如,155至195、160至190、165至185或170至180bp) 的同源臂。在某些实施方式中，模板核酸在切口或替换序列的5' 端上包括约150、155、160、165、170、175、180、185、190、195或200bp 的同源臂。在某些实施方式中，模板核酸在切口或替换序列的3' 端上包括少于约100、90、80、70、60、50、40、30、20、15或10bp 的同源臂。

[0963] 在某些实施方式中，模板核酸包括例如一个或多个核苷酸的核苷酸序列，所述核苷酸序列可被加入到或可引起靶核酸中的变化。在其他实施方式中，模板核酸包括可用来修饰靶标的核苷酸序列。

[0964] 模板核酸可包括替换序列。在某些实施方式中，模板核酸包括5' 同源臂。在某些实施方式中，模板核酸包括3' 同源臂。

[0965] 在某些实施方式中，模板核酸是线性双链DNA。例如，长度可约为150-200bp，例如，约150、160、170、180、190或200bp。例如，长度可至少为150、160、170、180、190或200bp。在某些实施方式中，长度不超过150、160、170、180、190或200bp。在某些实施方式中，双链模板核酸具有约160bp的长度，例如，约155-165、约150-170、约140-180、约130-190、约120-200、约110-210、约100-220、约90-230或约80-240bp。

[0966] 模板核酸可为线性单链DNA。在某些实施方式中，模板核酸是 (i) 可与靶核酸的切割链退火的线性单链DNA, (ii) 可与靶核酸的完整链退火的线性单链DNA, (iii) 可与靶核酸的正链退火的线性单链DNA, (iv) 可与目标核酸的负链退火的线性单链DNA, 或前述多于一种的线性单链DNA。例如，长度可约为150-200个核苷酸，例如，约150、160、170、180、190或200个核苷酸。例如，长度可至少为150、160、170、180、190或200个核苷酸。在某些实施方式中，长度不超过150、160、170、180、190或200个核苷酸。在某些实施方式中，单链模板核酸具有约160个核苷酸的长度，例如，约155-165、150-170、140-180、130-190、120-200、110-210、100-220、90-230或80-240个核苷酸。

[0967] 在某些实施方式中，模板核酸是环状双链DNA，例如质粒。在某些实施方式中，模板核酸在替换序列和/或切口的任一侧上包括约500至1000bp的同源臂。在某些实施方式中，模板核酸包括切口或替换序列的5' 端、切口或替换序列的3' 端或切口或替换序列的5' 以及3' 端的约300、400、500、600、700、800、900、1000、1500或2000bp的同源臂。在某些实施方式中，模板核酸包括切口或替换序列的5' 端、切口或替换序列的3' 端或切口或替换序列的5' 以及3' 端的至少300、400、500、600、700、800、900、1000、1500或2000bp的同源臂。在某些实施方式中，模板核酸包括切口或替换序列的5' 端、切口或替换序列的3' 端或切口或替换序列的5' 以及3' 端的不超过300、400、500、600、700、800、900、1000、1500或2000bp的同源臂。

[0968] 在某些实施方式中,可缩短一个或两个同源臂以避免包括某些序列重复元件,例如Alu重复序列、LINE元件。例如,可缩短5'同源臂以避免序列重复元件,可缩短3'同源臂以避免序列重复元件。在某些实施方式中,可缩短5'和3'同源臂以避免包含某些序列重复元件。

[0969] 在某些实施方式中,模板核酸是腺病毒载体,例如AAV载体,例如具有允许其包装在AAV衣壳中的长度和序列的ssDNA分子。载体可以是例如小于5kb并可含有促进包装到衣壳中的ITR序列。载体可能是整合缺陷的。在某些实施方式中,模板核酸在替换序列和/或切口的任一侧上包括约150至1000个核苷酸的同源臂。在某些实施方式中,模板核酸包括切口或替换序列的5'端、切口或替换序列的3'端或切口或替换序列的5'以及3'端的约100、150、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸的同源臂。在某些实施方式中,模板核酸包括切口或替换序列的5'端、切口或替换序列的3'端或切口或替换序列的5'以及3'端的至少100、150、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸的同源臂。在某些实施方式中,模板核酸包括切口或替换序列的5'端、切口或替换序列的3'端或切口或替换序列的5'以及3'端的至多100、150、200、300、400、500、600、700、800、900、1000、1500或2000个核苷酸的同源臂。

[0970] 在某些实施方式中,模板核酸是慢病毒载体,例如IDLV(整合缺陷型慢病毒)。在某些实施方式中,模板核酸在替换序列和/或切口的任一侧上包含约500至1000bp的同源臂。在某些实施方式中,模板核酸包括切口或替换序列的5'端、切口或替换序列的3'端或切口或替换序列的5'以及3'端的约300、400、500、600、700、800、900、1000、1500或2000bp的同源臂。在某些实施方式中,模板核酸包括切口或替换序列的5'端、切口或替换序列的3'端或切口或替换序列的5'以及3'端的至少300、400、500、600、700、800、900、1000、1500或2000bp的同源臂。在某些实施方式中,模板核酸包括切口或替换序列的5'端、切口或替换序列的3'端或切口或替换序列的5'以及3'端的不超过300、400、500、600、700、800、900、1000、1500或2000bp的同源臂。

[0971] 在某些实施方式中,模板核酸包括阻止Cas9识别和切割模板核酸的一个或多个突变,例如沉默突变。相对于待改变的细胞基因组中的相应序列,模板核酸可包括例如至少1、2、3、4、5、10、20或30个沉默突变。在某些实施方式中,相对于待改变的细胞基因组中的相应序列,模板核酸可包括至多2、3、4、5、10、20、30或50个沉默突变。在某些实施方式中,cDNA包括阻止Cas9识别和切割模板核酸的一个或多个突变,例如沉默突变。相对于待改变的细胞基因组中的相应序列,模板核酸可包括例如至少1、2、3、4、5、10、20或30个沉默突变。在某些实施方式中,相对于待改变的细胞基因组中的相应序列,模板核酸可包括至多2、3、4、5、10、20、30或50个沉默突变。

[0972] 在某些实施方式中,5'和3'同源臂各自包含位于对应于替代序列的核苷酸两侧的一段序列。在某些实施方式中,模板核酸包括侧接5'同源臂和3'同源臂的替换序列,每个同源臂各自独立地包含10个或多个、20个或多个、50个或多个、100个或多个、150个或多个、200个或多个、250个或多个、300个或多个、350个或多个、400个或多个、450个或多个、500个或多个、550个或多个、600个或多个、650个或多个、700个或多个、750个或多个、800个或多个、850个或多个、900个或多个、1000个或多个、1100个或多个、1200个或多个、1300个或多个、1400个或多个、1500个或多个、1600个或多个、1700个或多个、1800个或多个、1900个或

多个或2000个或多个核苷酸。在某些实施方式中，模板核酸包含侧接5'同源臂和3'同源臂的替换序列，其各自独立地包括至少50、100或150个核苷酸，但不足以包括重复元件。在某些实施方式中，模板核酸包含侧接5'同源臂和3'同源臂的替换序列，其各自独立地包括5至100个、10至150个或20至150个核苷酸。在某些实施方式中，替换序列可选择性地包括启动子和/或polyA信号。

[0973] 9.3单链退火

[0974] 单链退火 (SSA) 是另一种修复存在于靶核酸中的两个重复序列之间的双链断裂的DNA修复过程。由SSA途径所利用的重复序列的长度通常大于30个核苷酸。在断裂端处发生切除，以显示靶核酸的两条链上的重复序列。切除之后，将含有重复序列的单链突出端用RPA蛋白涂覆，以防止重复序列不适当退火，例如退火至自身。RAD52结合至突出端上的重复序列中的每者上并且将所述序列对齐以使得互补的重复序列能够退火。退火之后，切割突出端的单链翼。新的DNA合成填充任何空位，并且连接恢复DNA双链体。作为所述处理的结果，两个重复之间的DNA序列被缺失。缺失的长度可以取决于很多因素，包括所利用的两个重复的位置、以及切除的途径或持续进行能力。

[0975] 与HDR途径相反，SSA不需要模板核酸来改变靶核酸序列。而是利用互补的重复序列。

[0976] 9.4其他DNA修复途径

[0977] 9.4.1SSBR(单链断裂修复)

[0978] 基因组中的单链断裂 (SSB) 由SSBR途径来修复，所述SSBR途径是不同于上文讨论的DSB修复机制的机制。SSBR途径具有四个主要阶段：SSB检测、DNA末端处理、DNA空位填充、以及DNA连接。更详细的解释给出于Caldecott 2008中，并且这里给出概述。

[0979] 在第一阶段中，当形成SSB时，PARP1和/或PARP2识别断裂并募集修复机器。DNA断裂处的PARP1结合和活性是瞬时的，并且它似乎通过促进损伤处的SSBr蛋白复合物的病灶积累或稳定性而加速SSBr。可论证地，这些SSBr蛋白中最重要的是XRCC1，它作为分子支架起作用，所述分子支架与SSBr过程的多种酶组分（包括负责清除DNA 3'和5'端的蛋白质）相互作用，使所述酶组分稳定化，并且刺激所述酶组分。例如，XRCC1与促进末端处理的若干蛋白质（DNA聚合酶β、PNK、和三种核酸酶APE1、APTX和APLF）相互作用。APE1具有内切核酸酶活性。APLF展示出内切核酸酶和3'到5'外切核酸酶活性。APTX具有内切核酸酶和3'到5'外切核酸酶活性。

[0980] 这种末端处理是SSBR的重要阶段，因为大部分（若非全部）SSB的3'-和/或5'-末端是‘被损伤的’。末端处理通常涉及将被损伤的3'-端恢复到羟基化状态和/或将被损伤的5'端恢复成磷酸酯部分，这样使得所述末端变得有连接能力。可以处理被损伤的3'末端的酶包括PNKP、APE1、和TDP1。可以处理被损伤的5'末端的酶包括PNKP、DNA聚合酶β、和APTX。LIG3（DNA连接酶III）也可以参与末端处理。一旦将末端清除，便发生空位填充。

[0981] 在DNA空位填充阶段，典型存在的蛋白质是PARP1、DNA聚合酶β、XRCC1、FEN1（翼内切核酸酶1）、DNA聚合酶δ/ε、PCNA、以及LIG1。存在两种空位填充方式，短补丁修复（short patch repair）和长补丁修复（long patch repair）。短补丁修复涉及插入丢失的单核苷酸。在一些SSB处，“空位填充”可能继续取代两个或更多个核苷酸（已经报道了多达12个碱基的取代）。FEN1是去除被取代的5'-残基的内切核酸酶。多种DNA聚合酶（包括Polβ）涉及在

SSB的修复中,其中DNA聚合酶的选择受SSB的来源和类型的影响。

[0982] 在第四阶段中,DNA连接酶如LIG1(连接酶I)或LIG3(连接酶III)催化末端连接。短补丁修复使用连接酶III,并且长补丁修复使用连接酶I。

[0983] 有时,SSBR是与复制偶联的。这条途径可以涉及CtIP、MRN、ERCC1、和FEN1中的一者或多者。可促进SSBR的另外的因子包括:aPARP、PARP1、PARP2、PARG、XRCC1、DNA聚合酶b、DNA聚合酶d、DNA聚合酶e、PCNA、LIG1、PNK、PNKP、APE1、APTX、APLF、TDP1、LIG3、FEN1、CtIP、MRN、以及ERCC1。

[0984] 9.4.2MMR(错配修复)

[0985] 细胞包含三种切除修复途径:MMR、BER和NER。所述切除修复途径具有的共同特点在于它们典型地识别DNA一条链上的损伤,然后外切/内切核酸酶去除所述损伤并且留下随后被DNA聚合酶填充的1-30个核苷酸的空位并且最终用连接酶密封。更完整的图片给出于Li,Cell Research (2008) 18:85-98中,并且这里提供了概述。

[0986] 错配修复(MMR)在错配的DNA碱基上运行。

[0987] MSH2/6或MSH2/3两种复合物都具有在错配识别和修复启动中发挥重要作用的ATP酶活性。MSH2/6优先识别碱基-碱基错配并且鉴定1或2个核苷酸的错配,而MSH2/3优先识别较大的ID错配。

[0988] hMLH1与hPMS2杂二聚化,以形成hMutL $\alpha$ ,其具有ATP酶活性并且对于MMR的多个步骤而言是重要的。它具有PCNA/复制因子C(RFC)依赖性内切核酸酶活性,所述活性在涉及EX01的3'切口指导的MMR中发挥重要作用。(EX01是HR和MMR两者中的参与者。)它规定了失配激发切除的终止。它调节错配引起的切除的终止。可促进MMR的其他因素包括:EX01、MSH2、MSH3、MSH6、MLH1、PMS2、MLH3、DNA Pol d、RPA、HMGB1、RFC、以及DNA连接酶I。

[0989] 9.4.3碱基切除修复(BER)

[0990] 碱基切除修复(BER)途径贯穿细胞周期是激活的;它主要负责从基因组中去除小的、非螺旋扭曲碱基损伤。相比之下,相关的核苷酸切除修复途径(在下一部分中讨论)修复庞大的螺旋扭曲损伤。更详细的解释给出于Caldecott,Nature Reviews Genetics 9,619-631(2008年8月)中,并且这里给出概述。

[0991] DNA碱基损伤后,碱基切除修复(BER)启动并且所述过程可以被简化为五个主要步骤:(a)去除被损伤的DNA碱基;(b)切开后续碱基位点;(c)清理DNA末端;(d)将所需核苷酸插入修复空位中;以及(e)连接DNA骨架中的剩余切口。这些最后的步骤类似于SSBR。

[0992] 在第一步中,损伤特异性DNA糖基化酶通过切割将碱基连接至糖磷酸骨架上的N-糖苷键而切除被损伤的碱基。然后具有相关裂解酶活性的AP内切核酸酶-1(APE1)或双功能DNA糖基化酶切开磷酸二酯骨架以产生DNA单链断裂(SSB)。BER的第三步涉及清理DNA末端。BER中的第四步由Pol $\beta$ 进行,它将新的互补核苷酸添加到修复空位中,并且在最终步骤中,XRCC1/连接酶III密封DNA骨架中的剩余切口。这完成了短补丁BER途径,其中大多数(约80%)的被损伤的DNA碱基得到修复。然而,如果在步骤3中在通过Pol $\beta$ 插入一个核苷酸后5'-端对末端处理活性有抗性,则将聚合酶换为复制型DNA聚合酶Pol $\delta/\epsilon$ ,所述复制型DNA聚合酶然后再将约2-8个核苷酸添加到DNA修复空位中。这产生了5'翼结构,其被与持续合成能力因子增殖细胞核抗原(PCNA)相关的翼内切核酸酶-1(FEN-1)识别并切除。DNA连接酶I然后密封DNA骨架中的剩余切口并完成长补丁BER。可促进BER途径的其他因素包括:DNA糖

基化酶、APE1、Polb、Pold、Pole、XRCC1、连接酶III、FEN-1、PCNA、RECQL4、WRN、MYH、PNKP、以及APTX。

[0993] 9.4.4核苷酸切除修复 (NER)

[0994] 核苷酸切除修复 (NER) 是从DNA中去除庞大的螺旋扭曲损伤的重要切除机制。在 Marteijn等人, Nature Reviews Molecular Cell Biology 15, 465-481 (2014) 中给出了关于NER的其他细节, 并且在此给出了概述。NER是涵盖两条更小途径的宽途径: 全基因组NER (GG-NER) 和转录偶联修复NER (TC-NER)。GG-NER和TC-NER使用不同的因子来识别DNA损伤。然而, 它们利用相同的机器进行损伤切开、修复、和连接。

[0995] 一旦识别出损伤, 细胞去除含有所述损伤的短的单链DNA区段。内切核酸酶XPF/ERCC1和XPG (由ERCC5编码) 通过切割损伤任一侧上的被损伤的链而去除损伤, 产生22-30个核苷酸的单链空位。接着, 细胞进行DNA空位填充合成和连接。在这个过程中涉及的是: PCNA、RFC、DNA Pol $\delta$ 、DNA Pol $\epsilon$ 或DNA Pol $\kappa$ 、以及DNA连接酶I或XRCC1/连接酶III。复制型细胞倾向于使用DNA pol $\epsilon$ 和DNA连接酶I进行连接步骤, 而非复制型细胞倾向于使用DNA Pol $\delta$ 、DNA Pol $\kappa$ 、和XRCC1/连接酶III复合物进行连接步骤。

[0996] NER可以涉及以下因子: XPA-G、POLH、XPF、ERCC1、XPA-G、以及LIG1。转录偶联NER (TC-NER) 可以涉及以下因子: CSA、CSB、XPB、XPD、XPG、ERCC1、以及TTDA。可以促进NER修复途径的另外的因子包括XPA-G、POLH、XPF、ERCC1、XPA-G、LIG1、CSA、CSB、XPA、XPB、XPC、XPD、XPF、XPG、TTDA、UVSSA、USP7、CETN2、RAD23B、UV-DDB、CAK子复合物、RPA、以及PCNA。

[0997] 9.4.5链间交联 (ICL)

[0998] 称为ICL修复途径的专用途径修复链间交联。可以在复制或转录期间在不同DNA链中的碱基之间发生链间交联、或共价交联。ICL修复涉及多个修复过程的协作, 具体地溶核活性、跨损伤合成 (TLS) 、和HDR。核酸酶被募集以切除被交联的碱基的任一侧上的ICL, 同时TLS和HDR协作以修复被切割的链。ICL修复可以涉及以下因子: 内切核酸酶 (例如, XPF和RAD51C) 、内切核酸酶 (如RAD51) 、跨损伤聚合酶 (例如, DNA聚合酶 $\zeta$ 和Rev1) 、以及范科尼贫血 (FA) 蛋白 (例如, FancJ) 。

[0999] 9.4.6其他途径

[1000] 在哺乳动物体内存在若干其他DNA修复途径。

[1001] 跨损伤合成 (TLS) 是用于修复有缺陷复制事件之后留下的单链断裂的途径, 并且涉及跨损伤聚合酶 (例如, DNA pol $\beta$ 和Rev1) 。

[1002] 无误复制后修复 (PRR) 是另一条用于修复有缺陷复制事件之后留下的单链断裂的途径。

[1003] 9.5靶向敲低

[1004] 与CRISPR/Cas介导的基因敲除不同 (其通过在DNA水平使基因 (例如, RS1、RL2或LAT基因) 突变而永久消除表达), CRISPR/Cas敲低通过使用人工转录因子允许基因表达暂时减少。使Cas9蛋白的两个DNA切割结构域中的关键残基突变 (例如, D10A和H840A突变) 导致产生无催化活性Cas9 (eiCas9, 也称为失活的Cas9或dCas9) 分子。无催化活性Cas9与gRNA复合并且定位至由gRNA的靶向结构域指定的DNA序列, 然而, 其不切割靶DNA。将dCas9融合至效应子结构域 (例如, 转录阻遏结构域) 能够将效应子募集到由gRNA指定的任何DNA位点。尽管当被募集到编码序列中的早期区时无酶促活性Cas9 (eiCas9) 本身可阻断转录, 但可以

通过将转录阻遏结构域(例如KRAB、SID或ERD)融合至Cas9并且将其募集到靶敲低位置(例如在起始密码子的3'序列的1000bp内或在基因(例如,RS1、RL2或LAT基因)的起始密码子的5'启动子区的500bp内)来实现更稳健的阻遏。很可能靶向启动子的DNaseI超敏位点(DHSs)可产生更有效的基因阻遏或激活,因为这些区域更有可能接触到Cas9蛋白并且也更有可能为内源转录因子携带位点。尤其针对基因阻遏,阻断内源转录因子的结合位点将有助于下调基因表达。在某些实施方式中,一种或多种eICas9分子可以用于阻断一种或多种内源转录因子的结合。在某些实施方式中,eICas9分子可以与染色质修饰蛋白融合。改变染色质状态可以导致靶基因的表达降低。与一种或多种染色质修饰蛋白融合的一种或多种eICas9分子可以用于改变染色质状态。

[1005] 在某些实施方式中,可以将gRNA分子靶向已知的转录应答元件(例如,启动子、增强子等)、已知的上游激活序列(UAS)、和/或具有怀疑能够控制靶DNA表达的未知或已知功能的序列。

[1006] CRISPR/Cas介导的基因敲低可以用于减少不想要的等位基因或转录物的表达。在某些实施方式中,永久破坏基因是不理想的情形。在这些实施方式中,位点特异性阻遏可以用于暂时降低或消除表达。在某些实施方式中,Cas阻遏物的脱靶效应可能比Cas核酸酶的脱靶效应更不严重,因为核酸酶可以切割任何DNA序列并且造成突变,而Cas阻遏物仅在其靶向转录活跃基因的启动区时才可具有效应。然而,尽管核酸酶介导的敲除是永久的,只有Cas阻遏物存在于细胞中,阻遏才可以持续。一旦阻遏物不再存在,很可能内源转录因子和基因调节元件将使表达恢复至它的天然状态。

[1007] 9.6基因组编辑方法中的gRNA的示例

[1008] 如本发明所述的gRNA分子可与产生双链断裂或单链断裂的Cas9分子一起使用以改变靶核酸的序列,例如,靶标或靶遗传标记。下面描述了用于这些方法的gRNA分子。

[1009] 在某些实施方式中,gRNA(例如嵌合gRNA)被构造以使其包括一种或多种以下性质;

[1010] (a) 例如,当靶向产生双链断裂的Cas9分子时,其可定位靶标的(i)50、100、150、200、250、300、350、400、450或500个核苷酸内的双链断裂或(ii)足够接近以使目标位置位于末端切除区域内的双链断裂;

[1011] (b) 其具有至少16个核苷酸的靶向结构域,例如(i)16、(ii)17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25或(xi)26个核苷酸的靶向结构域;以及

[1012] (c) (i)近端和尾端结构域一起时包括至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部和近端结构域的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1013] (c) (ii)第二互补结构域的最后一个核苷酸的3'端有至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌的对应序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1014] (c) (iii)与第一互补结构域中对应的核苷酸互补的第二互补结构域的最后一个核苷酸的3'端有至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,例如,来自天然

存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌的对应序列的至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1015] (c) (iv) 尾部结构域长度至少为10、15、20、25、30、35或40个核苷酸,例如,其包括来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部结构域的至少10、15、20、25、30、35或40个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;或

[1016] (c) (v) 尾部结构域包括15、20、25、30、35、40个核苷酸或天然形成的尾部结构域的所有对应部分,例如,天然形成的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部结构域。

[1017] 在某些实施方式中,gRNA被构造成使其包含以下特性:a和b (i) ;a和b (ii) ;a和b (iii) ;a和b (iv) ;a和b (v) ;a和b (vi) ;a和b (vii) ;a和b (viii) ;a和b (ix) ;a和b (x) ;a和b (xi) ;a和c;a、b和c;a (i) 、b (i) 和c (i) ;a (i) 、b (i) 和c (ii) ;a (i) 、b (ii) 和c (i) ;a (i) 、b (ii) 和c (ii) ;a (i) 、b (iii) 和c (i) ;a (i) 、b (iii) 和c (ii) ;a (i) 、b (iv) 和c (i) ;a (i) 、b (iv) 和c (ii) ;a (i) 、b (v) 和c (i) ;a (i) 、b (v) 和c (ii) ;a (i) 、b (vi) 和c (i) ;a (i) 、b (vi) 和c (ii) ;a (i) 、b (vii) 和c (i) ;a (i) 、b (vii) 和c (ii) ;a (i) 、b (viii) 和c (i) ;a (i) 、b (viii) 和c (ii) ;a (i) 、b (ix) 和c (i) ;a (i) 、b (ix) 和c (ii) ;a (i) 、b (x) 和c (i) ;a (i) 、b (x) 和c (ii) ;a (i) 、b (xi) 或c (i) ;a (i) 、b (xi) 和c (ii) 。

[1018] 在某些实施方式中,gRNA(例如嵌合gRNA)被构造以使其包括一种或多种以下性质;

[1019] (a) 例如,当靶向产生单链断裂的Cas9分子时,一个或两个gRNA可将单链断裂定位在(i)靶标的50、100、150、200、250、300、350、400、450或500个核苷酸内,或(ii)足够接近以使得目标位置在末端切除区域内;

[1020] (b) 一个或两个具有至少16个核苷酸的靶向结构域,例如(i)16、(ii)

[1021] 17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25或(xi)26个核苷酸的靶向结构域;以及

[1022] (c) (i) 近端和尾端结构域一起时包括至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部和近端结构域的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1023] (c) (ii) 第二互补结构域的最后一个核苷酸的3'端有至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌的对应序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1024] (c) (iii) 与第一互补结构域中对应的核苷酸互补的第二互补结构域的最后一个核苷酸的3'端有至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,例如,来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌的对应序列的至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1025] (c) (iv) 尾部结构域长度至少为10、15、20、25、30、35或40个核苷酸,例如,其包括来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部结构域的至少10、15、20、25、30、35或40个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;或

[1026] (c) (v) 尾部结构域包括15、20、25、30、35、40个核苷酸或天然形成的尾部结构域的所有对应部分,例如,天然形成的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部结构域。

[1027] 在某些实施方式中,gRNA被构造成使其包含以下特性:a和b (i) ;a和b (ii) ;a和b (iii) ;a和b (iv) ;a和b (v) ;a和b (vi) ;a和b (vii) ;a和b (viii) ;a和b (ix) ;a和b (x) ;a和b (xi) ;a和c;a、b和c;a (i) 、b (i) 和c (i) ;a (i) 、b (i) 和c (ii) ;a (i) 、b (ii) 和c (i) ;a (i) 、b (ii) 和c (ii) ;a (i) 、b (iii) 和c (i) ;a (i) 、b (iii) 和c (ii) ;a (i) 、b (iv) 和c (i) ;a (i) 、b (iv) 和c (ii) ;a (i) 、b (v) 和c (i) ;a (i) 、b (v) 和c (ii) ;a (i) 、b (vi) 和c (i) ;a (i) 、b (vi) 和c (ii) ;a (i) 、b (vii) 和c (i) ;a (i) 、b (vii) 和c (ii) ;a (i) 、b (viii) 和c (i) ;a (i) 、b (viii) 和c (ii) ;a (i) 、b (ix) 和c (i) ;a (i) 、b (ix) 和c (ii) ;a (i) 、b (x) 和c (i) ;a (i) 、b (x) 和c (ii) ;a (i) 、b (xi) 或c (i) ;a (i) 、b (xi) 和c (ii) 。

[1028] 在某些实施方式中,gRNA与具有HNH活性的Cas9切口酶分子一起使用,例如,具有失活的RuvC活性的Cas9分子,例如,在D10处具有突变的Cas9分子,例如,D10A突变。在某些实施方式中,gRNA与具有RuvC活性的Cas9切口酶分子一起使用,例如,具有失活的HNH活性的Cas9分子,例如,在840处具有突变的Cas9分子,例如,H840A突变。在某些实施方式中,gRNA与具有RuvC活性的Cas9切口酶分子一起使用,例如,具有失活的HNH活性的Cas9分子,例如,在N863处具有突变的Cas9分子,例如,N863A突变。在某些实施方式中,gRNA与具有RuvC活性的Cas9切口酶分子一起使用,例如,具有失活的HNH活性的Cas9分子,例如,在N580处具有突变的Cas9分子,例如,N580A突变。

[1029] 在某些实施方式中,包括第一和第二gRNA的gRNA对(例如,嵌合gRNA对)被构造成使其包括一种或多种以下性质:

[1030] (a) 例如,当靶向产生单链断裂的Cas9分子时,一个或两个gRNA可将单链断裂定位在(i)靶标的50、100、150、200、250、300、350、400、450或500个核苷酸内,或(ii)足够接近以使得目标位置在末端切除区域内;

[1031] (b) 一个或两个具有至少16个核苷酸的靶向结构域,例如(i)16、(ii)17、(iii)18、(iv)19、(v)20、(vi)21、(vii)22、(viii)23、(ix)24、(x)25或(xi)26个核苷酸的靶向结构域;

[1032] (c) (i) 近端和尾端结构域一起时包括至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部和近端结构域的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1033] (c) (ii) 第二互补结构域的最后一个核苷酸的3'端有至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,例如,来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌的对应序列的至少15、18、20、25、30、31、35、40、45、49、50或53个核苷酸,或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1034] (c) (iii) 与第一互补结构域中对应的核苷酸互补的第二互补结构域的最后一个核苷酸的3' 端有至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸, 例如, 来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌的对应序列的至少16、19、21、26、31、32、36、41、46、50、51或54个核苷酸, 或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列;

[1035] (c) (iv) 尾部结构域长度至少为10、15、20、25、30、35或40个核苷酸, 例如, 其包括来自天然存在的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部结构域的至少10、15、20、25、30、35或40个核苷酸, 或相差不超过1、2、3、4、5、6、7、8、9或10个核苷酸的序列; 或

[1036] (c) (v) 尾部结构域包括15、20、25、30、35、40个核苷酸或天然形成的尾部结构域的所有对应部分, 例如, 天然形成的化脓性链球菌、金黄色葡萄球菌或脑膜炎奈瑟氏球菌尾部结构域。

[1037] (d) gRNA被构造成使得当与靶核酸杂交时, 其被0-50、0-100、0-200、至少10、至少20、至少30或至少50个核苷酸分开;

[1038] (e) 第一个gRNA和第二个gRNA产生的断裂位于不同的链上; 以及

[1039] (f) PAM向外朝向。

[1040] 在某些实施方式中, 一个或两个gRNA被构造成使其包括以下特性:a和b (i) ;a和b (ii) ;a和b (iii) ;a和b (iv) ;a和b (v) ;a和b (vi) ;a和b (vii) ;a和b (viii) ;a和b (ix) ;a和b (x) ;a和b (xi) ;a和c;a、b和c;a (i) 、b (i) 和c (i) ;a (i) 、b (i) 和c (ii) ;a (i) 、b (i) 、c和d;a (i) 、b (i) 、c和e;a (i) 、b (i) 、c、d和e;a (i) 、b (ii) 和c (i) ;a (i) 、b (ii) 和c (ii) ;a (i) 、b (ii) 、c和d;a (i) 、b (ii) 、c和e;a (i) 、b (ii) 、c、d和e;a (i) 、b (iii) 和c (i) ;a (i) 、b (iii) 和c (ii) ;a (i) 、b (iii) 、c和d;a (i) 、b (iii) 、c和e;a (i) 、b (iii) 、c、d和e;a (i) 、b (iv) 和c (i) ;a (i) 、b (iv) 、c和d;a (i) 、b (iv) 、c和e;a (i) 、b (iv) 、c、d和e;a (i) 、b (v) 和c (i) ;a (i) 、b (v) 和c (ii) ;a (i) 、b (v) 、c和d;a (i) 、b (v) 、c和e;a (i) 、b (v) 、c、d和e;a (i) 、b (vi) 和c (i) ;a (i) 、b (vi) 和c (ii) ;a (i) 、b (vi) 、c和d;a (i) 、b (vi) 、c和e;a (i) 、b (vi) 、c、d和e;a (i) 、b (vii) 和c (i) ;a (i) 、b (vii) 和c (ii) ;a (i) 、b (vii) 、c和d;a (i) 、b (vii) 、c和e;a (i) 、b (vii) 、c、d和e;a (i) 、b (viii) 和c (i) ;a (i) 、b (viii) 和c (ii) ;a (i) 、b (viii) 、c和d;a (i) 、b (viii) 、c和e;a (i) 、b (viii) 、c、d和e;a (i) 、b (ix) 和c (i) ;a (i) 、b (ix) 、c和d;a (i) 、b (ix) 、c和e;a (i) 、b (ix) 、c、d和e;a (i) 、b (x) 和c (i) ;a (i) 、b (x) 和c (ii) ;a (i) 、b (x) 、c和d;a (i) 、b (x) 、c和e;a (i) 、b (x) 、c、d和e;a (i) 、b (xi) 和c (i) ;a (i) 、b (xi) 、c和d;a (i) 、b (xi) 、c和e;a (i) 、b (xi) 、c、d和e。

[1041] 在某些实施方式中, gRNA与具有HNH活性的Cas9切口酶分子一起使用, 例如, 具有失活的RuvC活性的Cas9分子, 例如, 在D10处具有突变的Cas9分子, 例如, D10A突变。在某些实施方式中, gRNA与具有RuvC活性的Cas9切口酶分子一起使用, 例如, 具有失活的HNH活性的Cas9分子, 例如, 在H840处具有突变的Cas9分子, 例如, N840A突变。在某些实施方式中, gRNA与具有RuvC活性的Cas9切口酶分子一起使用, 例如, 具有失活的HNH活性的Cas9分子, 例如, 在N863处具有突变的Cas9分子, 例如, N863A突变。在某些实施方式中, gRNA与具有RuvC活性的Cas9切口酶分子一起使用, 例如, 具有失活的HNH活性的Cas9分子, 例如, 在N580处具有突变的Cas9分子, 例如, N580A突变。

[1042] 10. 靶细胞

[1043] 可使用Cas9分子(例如,eaCas9分子或eiCas9分子)或Cas9融合蛋白和gRNA分子(例如Cas9分子/gRNA分子复合物)来操作细胞,例如,编辑种类繁多的细胞中的靶核酸。

[1044] 在某些实施方式中,通过编辑(例如,在其中引入突变)一种或多种靶基因,例如HSV-1或HSV-2基因,例如RS1、RL2和/或LAT基因中的一种、两种或三种来操纵细胞。在某些实施方式中,使用HSV-1和/或HSV-2来感染细胞。在某些实施方式中,细胞来自具有HSV-1和/或HSV-2感染的受试者。在某些实施方式中,细胞或受试者具有潜伏的HSV-1和/或HSV-2感染。在某些实施方式中,例如,在体内调节一种或多种靶基因(例如,本发明所述的一种或多种靶基因,例如RS1、RL2和/或LAT基因中的一种、两种或三种)的表达。

[1045] 可将Cas9和gRNA分子递送至靶细胞,例如,本发明所述的细胞。在某些实施方式中,靶细胞是上皮细胞,例如,口咽上皮细胞(包括,例如鼻子、牙龈、嘴唇、舌或咽的上皮细胞)、手指或指甲床的上皮细胞或肛门-生殖器上皮细胞(包括,例如阴茎、阴囊、外阴、阴道、宫颈、肛门或大腿的上皮细胞)。在某些实施方式中,靶细胞是神经元细胞,例如,颅神经节神经元(例如三叉神经节神经元,例如,动眼神经神经节神经元,例如,外展神经神经节神经元,例如滑车神经神经节神经元)、颈神经节神经元(例如骶神经节神经元、感觉神经节神经元、皮层神经元、小脑神经元或海马神经元)。在某些实施方式中,靶细胞是一个视觉细胞,例如,眼睛的上皮细胞,例如,眼睑的上皮细胞,例如结膜细胞,例如结膜上皮细胞,例如角膜基质(corneal keratocyte)细胞,例如角膜缘细胞,例如角膜上皮细胞,例如角膜基质(corneal stromal)细胞,例如睫状体,体细胞,例如巩膜细胞,例如晶状体细胞,例如脉络膜细胞,例如视网膜细胞,例如视杆状光感受器细胞,例如视锥光感受器细胞,例如视网膜色素上皮细胞,例如水平细胞,例如无长突细胞,例如神经节细胞。

[1046] 11. 递送、配制品和施用途径

[1047] 组成部分(例如Cas9分子、一个或多个gRNA分子(例如,Cas9分子/gRNA分子复合物))和供体模板核酸或全部三种能以各种形式被递送、制剂或施用,参见例如表14和表15。在某些实施方式中,Cas9分子、一种或多种gRNA分子(例如,两种gRNA分子)一起存在于基因组编辑系统中。在某些实施方式中,编码Cas9分子的序列和编码两种或多种(例如,2、3、4种或多种)不同gRNA分子的序列存在于相同的核酸分子,例如AAV载体上。在某些实施方式中,编码Cas9分子的两种序列和编码两种或多种(例如,2、3、4种或更多种)不同gRNA分子的序列存在于相同的核酸分子,例如AAV载体上。当Cas9或gRNA组分编码为用于递送的DNA时,DNA通常可包括控制区以实现表达,例如,包含启动子。用于Cas9分子序列的有用启动子包括CMV、EFS、EF-1a、MSCV、PGK和CAG、骨骼肌α肌动蛋白启动子、肌肉肌酸激酶启动子、肌营养不良蛋白启动子、α肌球蛋白重链启动子和平滑肌肌动蛋白启动子。在某些实施方式中,该启动子是组成型启动子。在某些实施方式中,该启动子是组织特异启动子。gRNA的有用启动子包括T7.H1、EF-1a、7SK、U6、U1和tRNA启动子。可选择具有相似或不同强度的启动子来调整组分的表达。编码Cas9分子的序列可包括核定位信号(NLS),例如,SV40NLS。在某些实施方式中,编码Cas9分子的序列包括至少两个核定位信号。在某些实施方式中,Cas9分子或gRNA分子的启动子可以是独立的、可诱导的、组织特异性的或细胞特异性的。表14提供了可如何制剂、递送、或施用所述组分的实例。

[1048] 表14

元件			
Cas9 分子	gRNA 分子	供体模板核酸	注解
DNA	DNA	DNA	在某些实施方式中，Cas9 分子（例如，eaCas9 分子或 eiCas9 分子）和 gRNA 转录自 DNA。在某些实施方式中，它们在单独的分子上进行编码。在某些实施方式中，提供所述供体模板作为单独的 DNA 分子。
DNA	DNA		在某些实施方式中，Cas9 分子（例如，eaCas9 分子或 eiCas9 分子）和 gRNA 转录自 DNA。在某些实施方式中，它们在单独的分子上进行编码。在某些实施方式中，在编码所述 gRNA 的相同 DNA 分子上提供所述供体模板。
DNA	DNA		在某些实施方式中，Cas9 分子（例如，eaCas9 分子或 eiCas9 分子）和 gRNA 转录自 DNA，此处自单个分子。在某些实施方式中，提供所述供体模板作为单独的 DNA 分子。
DNA	<u>DNA</u>		在某些实施方式中，Cas9 分子（例如，eaCas9 分子或 eiCas9 分子）和 gRNA 转录自 DNA。在某些实施方式中，它们在单独的分子上进行编码。在某些实施方式中，在编码所述 Cas9 的相同 DNA 分子上提供所述供体模板。
DNA	RNA	DNA	在某些实施方式中，Cas9 分子（例如，eaCas9 或 eiCas9 分子）转录自 DNA，并且 gRNA 被提供为体外转

[1049]

			录或合成的 RNA。在某些实施方式中，提供所述供体模板作为单独的 DNA 分子。
DNA	RNA		在某些实施方式中，Cas9 分子（例如，eaCas9 或 eiCas9 分子）转录自 DNA，并且 gRNA 被提供为体外转录或合成的 RNA。在某些实施方式中，在编码所述 Cas9 的相同 DNA 分子上提供所述供体模板。
mRNA	RNA	DNA	在某些实施方式中，Cas9 分子（例如，eaCas9 或 eiCas9 分子）在体外翻译自 mRNA，并且 gRNA 被提供为体外转录或合成的 RNA。在某些实施方式中，提供所述供体模板作为 DNA 分子。
mRNA	DNA	DNA	在某些实施方式中，Cas9 分子（例如，eaCas9 或 eiCas9 分子）在体外翻译自 mRNA，并且 gRNA 转录自 DNA。在某些实施方式中，提供所述供体模板作为单独的 DNA 分子。
mRNA	DNA		在某些实施方式中，Cas9 分子（例如，eaCas9 或 eiCas9 分子）在体外翻译自 mRNA，并且 gRNA 转录自 DNA。在某些实施方式中，在编码所述 gRNA 的相同 DNA 分子上提供所述供体模板。
蛋白质	DNA	DNA	在某些实施方式中，提供 Cas9 分子（例如，eaCas9 分子或 eiCas9 分子）作为蛋白质，且 gRNA 转录自 DNA。在某些实施方式中，提供所述供体模板作为单独的 DNA 分子。
蛋白质	DNA		在某些实施方式中，提供 Cas9 分子（例如，eaCas9 分子或 eiCas9 分子）作为蛋白质，且 gRNA 转录自 DNA。在某些实施方式中，在编码

[1050]

			所述 gRNA 的相同 DNA 分子上提供所述供体模板。	
[1051]	蛋白质	RNA	DNA	在某些实施方式中，提供（例如，eaCas9 分子或 eiCas9 分子）作为蛋白质，并且提供 gRNA 作为转录或合成的 RNA。该递送方法称作“RNP 递送”。在某些实施方式中，提供所述供体模板作为 DNA 分子。

[1052] 表15概括了Cas系统的组分(例如,如本发明所述的Cas9分子组分和gRNA分子组分)的各种递送方法。

[1053] 表15

载体递送/模式		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
[1054]	病毒	逆转录病毒	否	稳定的	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	阳离子脂质体	是	瞬时的	取决于递送什么	核酸和蛋白
	聚合物纳米粒子	是	瞬时的	取决于递送什么	核酸和蛋白

[1055]	生物非病毒递送运载体	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	

[1056] 11.1Cas9分子和/或一个或多个gRNA分子的基于DNA的递送

[1057] 编码Cas9分子(例如,eaCas9分子或eiCas9分子)、gRNA分子、供体模板核酸或其任何组合(例如两种或全部)的核酸组合物可通过领域已知的方法或如本发明所述的给予受试者或递送至细胞。例如,Cas9编码和/或gRNA编码的DNA以及供体模板可通过例如载体(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。

[1058] 编码Cas9分子(例如eaCas9分子或eiCas9分子)和/或gRNA分子的核酸组合物可通过靶细胞(例如本发明所述的靶细胞)与促进摄取的分子(例如N-乙酰半乳糖胺)进行缀合。供体模板分子可同样通过靶细胞(例如本文所述的靶细胞)与促进摄取的分子(例如N-乙酰半乳糖胺)进行缀合。

[1059] 在某些实施方式中,通过载体(例如病毒载体/病毒或质粒)递送Cas9编码和/或gRNA编码的DNA。

[1060] 载体可包括编码Cas9分子和/或gRNA分子的序列和/或与被靶向的区域(例如,靶序列)具有高度同源性的供体模板。在某些实施方式中,供体模板包含靶序列的全部或部分。示例性供体模板是修复模板,例如基因校正模板或基因突变模板,例如点突变(例如单核苷酸(nt)取代)模板。载体还可包括编码融合到例如Cas9分子序列上的信号肽(例如,用于核定位、核仁定位或线粒体定位)的序列。例如,载体可包括与编码Cas9分子的序列融合的核定位序列(例如,来自SV40)。

[1061] 载体可包括一个或多个调节/控制元件,例如,启动子、增强子、内含子、多聚腺苷酸化信号、Kozak共有序列、内部核糖体进入位点(IRES)、2A序列、和/或剪接受体或供体。在某些实施方式中,启动子被RNA聚合酶II识别(例如,CMV启动子)。在其他实施方式中,启动子被RNA聚合酶III识别(例如,U6启动子)。在某些实施方式中,该启动子是调控的启动子(例如,诱导型启动子)。在某些实施方式中,该启动子是组成型启动子。在某些实施方式中,该启动子是组织特异启动子。在某些实施方式中,该启动子是病毒启动子。在某些实施方式中,该启动子是非病毒启动子。

[1062] 在某些实施方式中,载体或递送载体是病毒载体(例如,用于产生重组病毒)。在某些实施方式中,所述病毒是DNA病毒(例如,dsDNA或ssDNA病毒)。在某些实施方式中,所述病毒是RNA病毒(例如,ssRNA病毒)。在某些实施方式中,病毒感染分裂的细胞。在某些实施方式中,病毒感染非分裂细胞。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。

[1063] 在某些实施方式中,病毒感染分裂的细胞。在其他实施方式中,病毒感染非分裂细胞。在某些实施方式中,病毒感染分裂和非分裂细胞。在某些实施方式中,病毒可整合到宿主基因组中。在某些实施方式中,该病毒被设计成具有降低的免疫力,例如在人体内。在某些实施方式中,病毒具有复制能力。在其他实施方式中,该病毒是复制缺陷型的,例如,具有一个或多个用于其他轮次病毒粒复制和/或包装的必要基因编码区被其他基因替代或缺失。在某些实施方式中,病毒引起Cas9分子或分子和/或gRNA分子或分子的瞬时表达。在其他实施方式中,该病毒引起Cas9分子或分子和/或gRNA分子或分子的长期持续的表达,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。病毒的包装能力可变化,例如,从至少约4kb至至少约30kb,例如,至少约5kb、10kb、15kb、20kb、25kb、30kb、35kb、40kb、45kb或50kb。

[1064] 在某些实施方式中,病毒载体识别特定的细胞类型或组织。例如,病毒载体可用不同的/可选的病毒包膜糖蛋白假型包装;用细胞类型特异性受体(例如,一种或多种病毒包膜糖蛋白的基因修饰以掺入靶配体,比如肽配体、单链抗体或生长因子)进行工程化;和/或工程改造成具有双特异性的分子桥,一端识别病毒糖蛋白,另一端识别靶细胞表面的一部分(例如,配体-受体、单克隆抗体、亲和素-生物素和化学缀合物)。

[1065] 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。

[1066] 在某些实施方式中,通过重组逆转录病毒传递Cas9和/或gRNA编码序列。在某些实施方式中,逆转录病毒(例如莫洛尼鼠白血病病毒)包括,例如允许整合到宿主基因组中的逆转录酶。在某些实施方式中,逆转录病毒具有复制能力。在其他实施方式中,逆转录病毒是复制缺陷型的,例如,具有一个或多个用于其他轮次病毒粒复制和包装所需的基因编码区被其他基因替代,或缺失。

[1067] 在某些实施方式中,通过重组慢病毒传递Cas9和/或gRNA编码核酸序列。在某些实施方式中,通过重组逆转录病毒递送供体模板核酸。例如,慢病毒是复制缺陷的,例如,不包含病毒复制所需的一个或多个基因。

[1068] 在某些实施方式中,通过重组腺病毒传递Cas9和/或gRNA编码核酸序列。在某些实施方式中,通过重组腺病毒递送供体模板核酸。在某些实施方式中,腺病毒被设计成在人体内具有降低的免疫力。

[1069] 在某些实施方式中,通过重组AAV传递Cas9和/或gRNA编码核酸序列。在某些实施方式中,通过重组AAV递送供体模板核酸。在某些实施方式中,AAV不将其基因组整合到宿主细胞的基因组中,例如本发明所述的靶细胞。在某些实施方式中,AAV可将其基因组的至少一部分整合到宿主细胞,例如本发明所述的靶细胞的基因组中。在某些实施方式中,AAV是自身互补的腺伴随病毒(scAAV),例如scAAV,其包装退火在一起形成双链DNA的两条链。可在所公开的方法中使用的AAV血清型包括AAV1、AAV2、修饰的AAV2(例如在Y444F、Y500F、Y730F和/或S662V处的修饰)、AAV3、修饰的AAV3(例如在Y705F、Y731F和/或T492V)、AAV4、AAV5、AAV6、修饰的AAV6(例如在S663V和/或T492V的修饰)、AAV8、AAV8.2、AAV9、AAVrh10和假型AAV如AAV2/8、AAV2/5和AAV2/6也用于所公开的方法中。在某些实施方式中,可用于本发明所述方法的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。

[1070] 在某些实施方式中,编码Cas9和/或gRNA的核酸序列在重构工程AAV衣壳中递送,例如具有约50%或更多、例如约60%或更多、约70%或更多、约80%或更多、约90%或更多、或约95%或更多的与来自血清型AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV.rh8、AAV.rh10、AAV.rh32/33、AAV.rh43或AAV.rh64R1的衣壳序列的序列同源性。

[1071] 在某些实施方式中,通过嵌合AAV衣壳递送Cas9和/或gRNA编码核酸序列。在某些实施方式中,通过嵌合AAV衣壳递送供体模板核酸。示例性嵌合AAV衣壳包括,但不限于AAV9i1、AAV2i8、AAV-DJ、AAV2G9、AAV2i8G9或AAV8G9。

[1072] 在某些实施方式中,AAV是自身互补的腺伴随病毒(scAAV),例如scAAV,其包装退火在一起形成双链DNA的两条链。

[1073] 在某些实施方式中,编码Cas9和/或gRNA的DNA由混合病毒递送,例如,本发明所述的一种或多种病毒的杂合。在某些实施方式中,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。

[1074] 包装细胞用于形成能够感染靶细胞的病毒颗粒。示例性包装细胞包括可包装腺病毒的293细胞和可包装逆转录病毒的 $\psi$ 2或PA317细胞。用于基因治疗的病毒载体通常由将核酸载体包装入病毒颗粒的生产细胞系产生。该载体通常包含包装和随后整合到宿主或靶细胞(如果适用)所需的最小病毒序列,同时其他病毒序列被编码待表达蛋白质的表达盒代替,例如Cas9分子的组分,例如,两个Cas9组分。例如,用于基因治疗的AAV载体通常仅具有来自AAV基因组的反向末端重复序列(ITS),其是宿主或靶细胞中包装和基因表达所需的。如“三重转染方案”所述,缺失的病毒功能可通过包装细胞系和/或含有腺病毒的E2A、E4和VA基因以及编码来自AAV的Rep和Cap基因的质粒来反式提供。此后,病毒DNA被包装在细胞系中,该细胞系含有编码其他AAV基因(即rep和cap)但缺乏ITS序列的辅助质粒。在某些实施方式中,将病毒DNA包装在含有来自腺病毒的E1A和/或E1B基因的生产细胞系中。细胞系也作为辅助物被腺病毒感染。辅助病毒(例如,腺病毒或HSV)或辅助质粒促进AAV载体的复制和用ITS从辅助质粒表达AAV基因。由于缺乏ITS序列,辅助质粒未被显著包装。通过例如腺病毒比AAV更敏感的热处理,可降低腺病毒的污染。

[1075] 在某些实施方式中,病毒载体能够进行细胞类型和/或组织类型识别。例如,病毒载体可用不同的/可选的病毒包膜糖蛋白假型包装;用细胞类型特异性受体(例如,病毒包膜糖蛋白的基因修饰以掺入靶配体,比如肽配体、单链抗体、生长因子)进行工程化;和/或工程改造成具有双特异性的分子桥,一端识别病毒糖蛋白,另一端识别靶细胞表面的一部分(例如,配体-受体、单克隆抗体、亲和素-生物素和化学缀合物)。

[1076] 在某些实施方式中,病毒载体实现细胞类型特异性表达。例如,可构建组织特异性启动子以限制转基因(Cas9和gRNA)仅在靶细胞中表达。载体的特异性也可通过转基因表达的微RNA依赖性控制来介导。在某些实施方式中,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,可掺入融合蛋白如融合感受态血凝素(HA)来增加病毒被摄入细胞中。在某些实施方式中,病毒载体具有核定位能力。例如,需要破坏核膜(在细胞分裂过程中)并因此不能感染非分裂细胞的病毒可被改变以将核定位肽掺入病毒的基质蛋白中,从而使非增殖细胞的转导成为可能。

[1077] 在某些实施方式中,通过基于非载体的方法(例如,使用裸DNA或DNA复合物)递送编码Cas9和/或gRNA的核酸序列。例如,可通过例如有机修饰的二氧化硅或硅酸盐

(Ormosil)、电穿孔、瞬时细胞压缩或挤压(例如Lee等人,2012,Nano Lett 12:6322-27)、基因枪、声孔、磁转染、脂质介导的转染、树枝状聚合物、无机纳米颗粒、磷酸钙或其组合来递送DNA。

[1078] 在某些实施方式中,通过电穿孔的递送包括将细胞与编码Cas9和/或gRNA的DNA在盒、腔室或比色杯中混合并施加具有限定的持续时间和幅度的一个或多个电脉冲。在某些实施方式中,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Cas9分子和/或gRNA编码的DNA在连接至装置(例如,泵)的容器中混合,所述装置向盒、室或比色皿中供给混合物,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。

[1079] 在某些实施方式中,编码Cas9和/或gRNA的核酸序列通过基于载体和非基于载体的方法的组合来递送。在某些实施方式中,通过载体和非基于载体的方法的组合递送供体模板核酸。例如,病毒体结合与灭活病毒(例如,HIV或流感病毒)组合的脂质体,其可导致比单独的病毒或脂质体方法更有效的基因转移,例如,在呼吸道上皮细胞中。

[1080] 在某些实施方式中,递送载体是非病毒递送载体。在某些实施方式中,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,Fe<sub>3</sub>MnO<sub>2</sub>)和二氧化硅。纳米粒子的外表面可与允许附着(例如缀合或包埋)有效负载的带正电荷的聚合物(例如,聚乙烯亚胺,聚赖氨酸,聚丝氨酸)缀合。在某些实施方式中,非病毒载体是有机纳米颗粒(例如,纳米颗粒内的有效载荷的截留)。示例性的有机纳米粒子包括,例如含有阳离子脂质以及用聚乙二醇(PEG)和鱼精蛋白包被的中性辅助脂质以及用脂质包衣包被的核酸复合物的SNALP脂质体。

[1081] 用于基因转移的示例性脂质示于下表16中。

[1082] 表16:用于基因转移的脂质

[1083]

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
N-[1-(2,3-二油烯基氧基)丙基]N,N,N-三甲基氯化铵	DOTMA	阳离子的
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酰精胺	DOGS	阳离子的
N-(3-氨基丙基)-N,N-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	LHON	阳离子的
1-(2,3-二油酰基氧基丙基)-2,4,6-三甲基吡啶	2Oc	阳离子的
2,3-二油烯基氧基-N-[2(精胺甲酰胺基-乙基)-N,N-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
N-(2-羟基乙基)-N,N-二甲基-2,3-双(十四烷氧基)-1-溴化丙铵	MDRIE	阳离子的
二肉豆蔻酰氧丙基 二甲基 羟乙基 溴化铵	DMRI	阳离子的
3 $\beta$ -[N-(N',N' -二甲基氨基乙烷)-氨甲酰基]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-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酰基亚精胺	DSL	阳离子的
外消旋-[(2,3-双十八烷氧基丙基)(2-羟乙基)]-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-氧甲基氧甲基)乙基]三甲基溴化铵	CLIP-6	阳离子的

[1084]	乙基二肉豆蔻酰磷脂酰胆碱	EDMPC	阳离子的
	1,2-二硬脂酰基-N,N-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
	O,O'-二肉豆蔻基-N-赖氨酸-天冬氨酸酯	DMKE	阳离子的
	1,2-二硬脂酰-sn-甘油基-3-乙基胆碱磷酸	DSEPC	阳离子的
	N-棕榈酰 D-赤-鞘氨昔氨基甲酰基-精胺	CCS	阳离子的
	N-叔丁基-N0-十四烷基-3-十四烷基氨基丙脒	二 C14-脒	阳离子的
	十八碳烯醇基氨基[乙基-2-十七烯基-3 羟乙基]氯化咪唑啉鎓	DOTIM	阳离子的
	N1-胆固醇氨基-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
[1085]	1,2-二亚油基氨基-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

[1085] 用于基因转移的示例性聚合物示于下表17中。

[1086] 表17: 用于基因转移的聚合物

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(N-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(β-氨基酯)	
聚(4-羟基-L-脯氨酸酯)	PHP
聚(烯丙基胺)	
聚(α-[4-氨基丁基]-L-乙醇酸)	PAGA

[1088]

聚(D,L-乳酸-共-乙醇酸)	PLGA
聚(N-乙基-4-乙烯基吡啶鎓溴化物)	
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-(二甲氨基)甲基丙烯酸乙酯)	pDMAEMA
聚(2-氨基丙烯磷酸酯)	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

[1089] 在某些实施方式中,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(例如,N-乙酰半乳糖胺(Ga1NAc))和细胞穿透肽)。在某些实施方式中,所述运载体使用融合肽和内体去稳定肽/聚合物。在某些实施方式中,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)。在某些实施方式中,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,可使用在还原性细胞环境中裂解的基于二硫化物的阳离子聚合物。

[1090] 在某些实施方式中,递送运载体是生物非病毒递送运载体。在某些实施方式中,运载体是减毒的细菌(例如,天然或人工改造为侵入性的但减毒以防止发病机理并表达转基因(例如,单核细胞增生利斯特氏菌,某些沙门氏菌菌株、长双歧杆菌和修饰的大肠杆菌、针对靶标特异性组织具有营养和组织特异性的细菌、具有修饰表面蛋白以改变靶组织特异性的细菌)。在某些实施方式中,运载体是遗传修饰的噬菌体(例如,具有大包装能力、免疫原性较低、含有哺乳动物质粒维持序列并具有并入的靶向配体的工程噬菌体)。在某些实施方式中,运载体是哺乳动物病毒样颗粒。例如,可产生修饰的病毒颗粒(例如,通过纯化“空”颗粒,随后用期望的货物离体组装病毒)。运载体也可设计成掺入靶向配体以改变靶组织特异性。在某些实施方式中,运载体是生物脂质体。例如,生物脂质体是来源于人类细胞的磷脂类粒子(例如红细胞,其是分解成来自受试者的球形结构的红细胞(例如,组织靶向可通过附着各种组织或细胞特异性配体)或分泌型外泌体-来自内吞来源的主体(即,患者)膜结合纳米囊泡(30-100nm)(例如,可由各种细胞类型产生并因此可被细胞吸收而不需要靶向配体)。

[1091] 在某些实施方式中,递送不同于Cas系统组分的一种或多种核酸分子(例如DNA分子),例如,本发明所述的Cas9分子组分或组分和/或gRNA分子组分或组分。在某些实施方式中,这些核酸分子与Cas系统的一种或多种组分同时进行递送。在某些实施方式中,核酸分子是在递送Cas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。在某些实施方式中,通过不同于Cas系统的一个或多个组分的方式递送核酸分子,例如,递送Cas9分子组分和/或

gRNA分子组分。核酸分子可通过本文所述的任何递送方法进行递送。例如,核酸分子可通过病毒载体例如整合缺陷型慢病毒递送,并通过电穿孔递送Cas9分子组分或组分和/或gRNA分子组分或组分,例如,从而可减少核酸(例如,DNA)引起的毒性。在某些实施方式中,核酸分子编码治疗性蛋白质,例如,本发明所述的蛋白质。在某些实施方式中,核酸分子编码RNA分子,例如,本发明所述的RNA分子。

[1092] 11.2 递送编码Cas9分子的RNA

[1093] 编码Cas9分子(例如eaCas9分子或eiCas9分子)和/或gRNA分子的RNA可通过领域已知的方法或如本发明所述的方法递送到细胞(例如,本发明所述的靶细胞)中。例如,可通过例如显微注射、电穿孔、瞬时细胞压缩或挤压(例如,如Lee等,2012,Nano Lett 12:6322-27)、脂质介导的转染、肽介导的递送或其组合来进行递送编码Cas9和/或编码gRNA的RNA。编码Cas9和/或编码gRNA的RNA可与促进靶细胞(例如本发明所述的靶细胞)摄取的分子进行缀合。

[1094] 在某些实施方式中,经由电穿孔的递送包括在盒、腔室或比色皿中将细胞与编码Cas9分子的RNA(例如,eaCas9分子、eiCas9分子或eiCas9融合蛋白)和/或gRNA分子以及存在或不存在供体模板核酸分子混合,并且施加一个或多个定义的持续时间和幅度的电脉冲。在某些实施方式中,经由电穿孔的递送使用其中细胞与编码Cas9分子的RNA(例如,eaCas9分子、eiCas9分子或eiCas9融合蛋白)和/或具有或不具有供体模板核酸分子的gRNA分子在与装置(例如泵)连接的容器中混合的系统,所述装置将混合物输送到盒、腔室或比色皿中,其中施加了具有限定持续时间和幅度的一个或多个电脉冲,之后将细胞递送到第二容器。Cas9编码和/或gRNA编码的RNA可与促进靶细胞(例如本发明所述的靶细胞)摄取的分子进行缀合。

[1095] 11.3 Cas9分子蛋白质的递送

[1096] Cas9分子(例如,eaCas9分子或eiCas9分子)可通过领域已知的方法或如本文所述的方法递送到细胞中。例如,可通过例如显微注射、电穿孔、瞬时细胞压缩或挤压(例如,如Lee等,2012,Nano Lett 12:6322-27)、脂质介导的转染、肽介导的递送或其组合来进行递送Cas9蛋白质分子。递送可以与编码gRNA的DNA或与gRNA相伴。Cas9蛋白可以与促进靶细胞(例如本文所述的靶细胞)摄取的分子进行缀合。

[1097] 在某些实施方式中,经由电穿孔的递送包括在盒、腔室或比色皿中将细胞与Cas9分子(例如,eaCas9分子、eiCas9分子或eiCas9融合蛋白)和/或将gRNA分子与或不与供体核酸混合,并且施加一个或多个定义的持续时间和幅度的电脉冲。在某些实施方式中,经由电穿孔的递送使用其中细胞与Cas9分子(例如,eaCas9分子、eiCas9分子或eiCas9融合蛋白)和/或gRNA分子在与装置(例如泵)连接的容器中混合的系统,所述装置将混合物输送到盒、腔室或比色皿中,其中施加了具有限定持续时间和幅度的一个或多个电脉冲,之后将细胞递送到第二容器。编码Cas9和/或编码gRNA的RNA可与促进靶细胞(例如本发明所述的靶细胞)摄取的分子进行缀合。

[1098] 11.4 Cas9分子蛋白质和gRNA的RNP递送

[1099] 在某些实施方式中,通过核糖核蛋白(RNP)递送将Cas9分子和gRNA分子递送至靶细胞。在某些实施方式中,提供Cas9分子作为蛋白质,并且提供转录或合成的RNA作为gRNA。gRNA分子可通过化学合成来产生。在某些实施方式中,在递送至靶细胞之前,gRNA分子在合

适的条件下与Cas9分子蛋白形成RNP复合物。可通过本领域已知的任何合适方法,例如通过电穿孔、脂质介导的转染、基于蛋白质或DNA的穿梭、机械力或液压力将RNP复合物递送至靶细胞。在某些实施方式中,通过电穿孔将RNP复合物递送至靶细胞。

[1100] 11.5施用路径

[1101] 全身施用模式包括口服和肠胃外途径。肠胃外途径包括,举例来说,静脉内、动脉内、骨内、肌内、皮内、皮下、鼻内以及腹膜内途径。全身给予的组分可以被修饰或制剂成将组分靶向上皮或神经元细胞。

[1102] 例如,局部施用模式包括鞘内、脊柱内、脑室内和脑实质内(例如,进入脑或脊髓的实质)。

[1103] 在某些实施方式中,局部施用模式包括在三叉神经的水平由实质进入背根神经节。在某些实施方式中,局部施用模式包括在骶神经节的水平由实质进入背根神经节。在某些实施方式中,局部施用模式包括在腰椎神经节的水平由实质进入背根神经节。在某些实施方式中,局部施用模式包括在胸神经节的水平由实质进入背根神经节。在某些实施方式中,局部施用模式包括在颈椎神经节,例如颈上神经节,例如颈中神经节,例如颈下神经节,的水平由实质进入背根神经节。在某些实施方式中,局部施用模式包括脑实质内进入颅神经节水平由实质进入背根神经节,例如,颅神经节I-XII。

[1104] 在某些实施方式中,相比于当全身给予(例如,静脉内)时,当局部给予时,显著更少量的组分(相比于全身方法)可以发挥效果。局部给予模式可以降低或消除潜在毒副作用的发生率,当全身给予治疗有效量的组分时可能发生毒副作用。

[1105] 在某些实施方式中,将本发明所述组分递送至上皮细胞,例如,口咽上皮细胞(包括,例如鼻子、牙龈、嘴唇、舌或咽的上皮细胞)、手指或指甲床的上皮细胞或肛门-生殖器上皮细胞(包括,例如阴茎、外阴、阴道或肛门的上皮细胞)。在某些实施方式中,将本发明所述组分递送至眼睛(包括例如角膜上皮,例如角膜基质,例如上眼睑和下眼睑的上皮,例如晶状体)。

[1106] 可以以周期性推注或从内部储库或外部储库(例如,从静脉袋)持续输注的方式提供给药。组分可以局部施用,例如通过从缓释药物递送装置连续释放。

[1107] 可以从内部储库(例如,从布置在眼内或眼外位置的植入物(参见美国专利号5,443,505和5,766,242))或从外部储库(例如,来自静脉袋)。组分可局部施用,例如,通过从固定在眼内壁上的持续释放药物递送装置中连续释放或通过靶向的透巩膜的控制释放进入脉络膜(参见,例如,PCT/US00/00207,PCT/US02/14279,Ambati等人,(2000)Invest.Ophthalmol.Vis.Sci.41:1181-1185;Ambati等人(2000)Invest.Ophthalmol.Vis.Sci.41:1186-1191)。本领域已知各种适用于将组分局部施用到眼内的装置。参见,例如,美国专利6,251,090、6,299,895、6,416,777、6,413,540和PCT/US00/28187。

[1108] 在某些实施方式中,可将组分制剂成允许在延长时段内释放。释放系统可包括生物可降解材料的基质或通过扩散释放掺入的组分的材料。这些组分可均匀或非均匀地分布在释放系统内。各种释放系统可能是有用的,但是,选择适当系统将取决于特定应用所需的释放速率。可使用不可降解和可降解的释放系统。合适的释放系统包括聚合物和聚合物基质、非聚合物基质或无机和有机赋形剂和稀释剂,例如但不限于碳酸钙和糖(例如,海藻

糖)。释放系统可以是天然的或合成的。然而,合成释放系统是优选的,因为它们通常更可靠、更可重复且产生更多限定的释放曲线。可选择释放系统材料,使得具有不同分子量的组分通过材料的扩散或降解而释放。

[1109] 代表性的合成的生物可降解聚合物包括,例如:聚酰胺,比如,聚(氨基酸)和聚(肽);诸如聚(乳酸)、聚(乙醇酸)、聚(乳酸-乙醇酸共聚物)和聚(己内酯)之类的聚酯;聚(酐);多原酸酯;聚碳酸酯;及其化学衍生物(取代、添加化学基团,例如烷基,亚烷基,羟基化,氧化和本领域技术人员常规制备的其它修饰)、其共聚物和混合物。代表性的合成不可降解聚合物包括,聚醚,例如:聚(环氧乙烷)、聚(乙二醇)和聚(四氢呋喃)等;乙烯基聚合物-聚丙烯酸酯和聚甲基丙烯酸酯如甲基、乙基、其他烷基、甲基丙烯酸羟乙酯、丙烯酸和甲基丙烯酸,以及其他如聚乙烯醇、聚乙烯吡咯烷酮和聚醋酸乙烯酯;聚(氨基甲酸酯);纤维素及其衍生物如烷基、羟烷基、醚、酯、硝基纤维素和各种醋酸纤维素;聚硅氧烷;及其任何化学衍生物(取代、添加化学基团,例如烷基,亚烷基,羟基化,氧化和本领域技术人员常规制备的其它修饰)、其共聚物和混合物。

[1110] 聚(丙交酯-共-乙交酯)微球也可以用于注射。典型地,微球体由乳酸和乙醇酸的聚合物组成,其被构造成形成空心球体。球体的直径可以大约为15-30微米,并且可装载本发明所述的组分。

[1111] 11.6组分的双模式或差异递送

[1112] 分开递送Cas系统的组分,例如,Cas9分子或Cas9-融合蛋白组分和gRNA分子组分,并且更具体地通过不同模式递送组分可增强性能,例如,通过改善组织特异性和安全性。在某些实施方式中,所述Cas9分子是Cas9变体。例如,但不作为限制,所述Cas9变体可以是化脓链球菌Cas9变体或金黄色葡萄球菌Cas变体。在某些实施方式中,所述化脓链球菌Cas9变体是EQR变体。在某些实施方式中,所述化脓链球菌Cas9变体是VRER变体。

[1113] 在某些实施方式中,Cas9分子(例如,eaCas9或eiCas9分子)或Cas9融合蛋白和gRNA分子通过不同的模式递送,或者在本发明中有时称为差异模式。如本发明所用的不同或差异模式是指赋予对象组分分子(例如,Cas9分子或gRNA分子)不同的药效学或药代动力学性质的递送模式。例如,递送模式可导致不同的组织分布、不同的半衰期或不同的时间分布,例如,在选定的腔室、组织或器官中。

[1114] 一些递送模式,例如,通过持续存在于细胞中的核酸载体递送,或通过例如通过自主复制或插入细胞核酸的细胞子代递送,导致组分更持久的表达和存在。示例包括病毒的递送,例如,腺伴随病毒或慢病毒递送。

[1115] 例如,例如Cas9分子和gRNA分子等组分可通过在所产生的身体部位或特定腔室、组织或器官中所递送的组分的半衰期或持久性方面不同的模式递送。在某些实施方式中,gRNA分子可通过这些模式进行递送。Cas9分子组分可通过导致持久性较低或较少暴露于身体或特定腔室或组织或器官的模式递送。

[1116] 更一般地,在某些实施方式中,第一种递送模式用于递送第一组分,第二种递送模式用于递送第二组分。第一种递送模式赋予第一药效学或药代动力学性质。第一药效学性质可以是例如组分或编码组分的核酸在体内、腔室、组织或器官中的分布、持久性或暴露。第二种递送模式赋予第二药效学或药代动力学性质。第二药效学性质可以是例如组分或编码组分的核酸在体内、腔室、组织或器官中的分布、持久性或暴露。

[1117] 在某些实施方式中,第一药效学或药代动力学性质,例如分布、持久性或暴露,比第二药效学或药代动力学性质更受限制。

[1118] 在某些实施方式中,选择第一递送模式以优化例如最小化药效学或药代动力学性质,例如分布、持久性或暴露。

[1119] 在某些实施方式中,选择第二递送模式以优化例如最大化药效学或药代动力学性质,例如分布、持久性或暴露。

[1120] 在某些实施方式中,第一种递送模式包括使用相对持久的元件,例如核酸,例如质粒或病毒载体,例如AAV或慢病毒。由于这种载体相对来说是持久的,因此从它们转录来的产品会相对持久。

[1121] 在某些实施方式中,第二种递送模式包括相对短暂的元件,例如, RNA或蛋白质。

[1122] 在某些实施方式中,第一组分包括gRNA,且递送模式相对持久,例如,gRNA从质粒或病毒载体(例如,AAV或慢病毒)转录。这些基因的转录几乎没有生理学后果,因为基因不编码蛋白质产物,且gRNA不能独立行事。以瞬时方式递送第二组分,即Cas9分子(例如eaCas9分子或eiCas9分子),例如,作为mRNA或作为蛋白质,确保完整的Cas9分子/gRNA分子复合物仅存在和保持活性很短的时间。

[1123] 此外,组分可以不同的分子形式递送,或者以不同的递送载体相互补充以增强安全性和组织特异性。

[1124] 使用差异递送模式可提高性能、安全性和有效性。例如,可减少最终的脱吧修饰的可能性。由于来自细菌来源的Cas酶的肽通过MHC分子在细胞表面展示,所以免疫原性组分如Cas9分子(例如,eaCas9或eiCas9分子)或Cas9-融合蛋白以较不持久的模式递送可降低免疫原性。两部分递送系统可减轻这些缺点。

[1125] 差异递送模式可用于将组分递送给不同但重叠的靶标区域。活性复合体的形成在靶标区域的重叠之外被最小化。因此,在某些实施方式中,通过第一递送模式递送第一组分,例如,gRNA分子,其导致第一空间分布,例如,组织分布。通过第二递送模式递送第二组分,例如Cas9分子(例如,eaCas9分子或eiCas9分子),其导致第二空间分布,例如,组织分布。在某些实施方式中,第一种模式包括选自脂质体、纳米颗粒(例如,聚合物纳米颗粒)和核酸(例如,病毒载体)的第一元件。第二模式包括从该组中选择的第二元件。在某些实施方式中,第一种递送模式包括第一靶向元件,例如,细胞特异性受体或抗体,第二种递送模式不包括该元件。在实施方式中,第二种递送模式包括第二靶向元件,例如,第二细胞特异性受体或第二抗体。

[1126] 当Cas9分子(例如,eaCas9分子或eiCas9分子)在病毒递送载体、脂质体或聚合物纳米颗粒中递送时,当可能仅需要靶向单一的组织时,有可能递送至多个组织并在其中有治疗活性。两部分传递系统可解决这一挑战并提高组织特异性。如果gRNA分子和Cas9分子被包装在具有不同但重叠的组织向性的分离的递送载体中,则全功能复合物仅在被两种载体靶向的组织中形成。

[1127] 11.7离体递送

[1128] 在某些实施方式中,将表14中描述的基因组编辑系统的每个组分引入细胞中,然后将所述细胞引入受试者体内,例如,将细胞从受试者体内移除、离体操作并引入受试者。引入组分的方法可以包括例如表15中所述的任何递送方法。

[1129] 12. 修饰的核苷、核苷酸和核酸

[1130] 修饰的核苷和修饰的核苷酸可以存在于核酸中,例如特别是gRNA,但是还有其他形式的RNA,例如mRNA、RNAi或siRNA。如本文所描述的,“核苷”被定义为包含五碳糖分子(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基团的核苷。

[1131] 修饰的核苷和核苷酸可以包括以下项中的一项或多项:

[1132] (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;

[1133] (ii) 核糖的组分(例如,核糖上的2'羟基)的改变,例如置换;

[1134] (iii) “脱磷酸”接头对磷酸部分的完全置换;

[1135] (iv) 天然存在的核碱基的修饰或置换;

[1136] (v) 核糖-磷酸骨架的置换或修饰;

[1137] (vi) 寡核苷酸的3'端或5'端的修饰,例如,末端磷酸基团的去除、修饰或置换或部分的结合;以及

[1138] (vii) 糖的修饰。

[1139] 以上列出的修饰可以组合,以提供可以具有两个、三个、四个或更多个修饰的修饰的核苷和核苷酸。例如,修饰的核苷或核苷酸可以具有修饰的糖和修饰的核碱基。在某些实施方式中,修饰gRNA的每个碱基,例如所有碱基都具有修饰的磷酸酯基团,例如所有修饰的磷酸酯基团都是硫代磷酸酯基团。在某些实施方式中,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替换。

[1140] 在某些实施方式中,修饰的核苷酸,例如,具有如本发明所述的修饰的核苷酸,可掺入到核酸中,例如,“修饰的核酸”中。在某些实施方式中,修饰的核酸包括一个、两个、三个或更多个修饰的核苷酸。在某些实施方式中,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%) 的位置是修饰的核苷酸。

[1141] 未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,在某些实施方式中,本发明描述的修饰的核酸可以含有一个或多个修饰的核苷或核苷酸,例如,以引入对核酸酶的稳定性。

[1142] 在某些实施方式中,本文描述的修饰的核苷、修饰的核苷酸和修饰的核酸当被在体内和离体地引入细胞群体中时都可以展现出减少的先天性免疫应答。术语“先天性免疫应答”包括对外源核酸的细胞应答,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。在某些实施方式中,本文描述的修饰的核苷、修饰的核苷酸和修饰的核酸可以破坏大沟相互作用配偶体与核酸的结合。在某些实施方式中,本文描述的修饰的核苷、修饰的核苷酸和修饰的核酸当被在体内和离体地引入细胞群体中时都可以展现出减少的先天性免疫应答,并且还破坏大沟相互作用配偶体与核酸的结合。

[1143] 12.1 化学基团的定义

[1144] 如本文所用的,“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。烷基基团可以包含从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。

[1145] 如本发明所用的,“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、茚满基、茚基等。在某些实施方式中,芳基基团具有从6至约20个碳原子。

[1146] 如本文所用的,“烯基”是指包含至少一个双键的脂肪族基团。

[1147] 如本文所用的,“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烃链。炔基基团的实例包括但不限于乙炔基、炔丙基和3-己炔基。

[1148] 如本文所用的,“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。芳烷基包括一个以上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-芴基、二苯甲基以及三苯甲基基团。

[1149] 如本文所用的,“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。

[1150] 如本文所用的,“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。

[1151] 如本文所用的,“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、吲哚基、苯硫基、吡唑基、吡啶基、吡嗪基、哒嗪基、嘧啶基、吲嗪基、嘌呤基、萘啶基、喹啉基以及蝶啶基。

[1152] 12.2磷酸骨架修饰

[1153] 12.2.1磷酸基团

[1154] 在某些实施方式中,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,修饰的核苷酸(例如,存在于修饰的核酸中的修饰的核苷酸)可以包括如本文描述的修饰的磷酸酯对未修饰的磷酸酯部分的完全置换。在某些实施方式中,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。

[1155] 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)、氢膦酸酯、磷酰胺酯(phosphoroamidate)、烷基或芳基膦酸酯和磷酸三酯。在某些实施方式中,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)、 $BR_3$ (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、 $NR_2$ (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Rp)或“S”构型(本文是Sp)。

[1156] 二硫代磷酸酯的两个非桥连氧被硫置换。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。在某些实施方式中,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)。

[1157] 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,将磷酸连接至核昔的氧)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。

[1158] 12.2.2磷酸基团的置换

[1159] 磷酸酯基团可以被不含磷连接物置换。在某些实施方式中,带电磷酸酯基团可以被中性部分置换。

[1160] 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)、肟、亚甲亚氨基、亚甲甲基亚氨基、亚甲肼基、亚甲二甲基肼基以及亚甲氧基甲基亚氨基。

[1161] 12.2.3核糖磷酸骨架的置换

[1162] 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昔或核昔酸替代物置换。在某些实施方式中,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)核昔替代物。

[1163] 12.3糖修饰

[1164] 修饰的核昔和修饰的核昔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”取代基修饰或替换。在某些实施方式中,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。

[1165] “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), $O(CH_2CH_2O)_nCH_2CH_2OR$ ,其中R可以是例如H或任选取代的烷基,并且n可以是从0至20的整数(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)。在某些实施方式中,“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)。在某些实施方式中,“氧基”-2'羟基基团修饰可以包括甲氧基乙基基团(MOE)(OCH<sub>2</sub>CH<sub>2</sub>OCH<sub>3</sub>,例如PEG衍生物)。

[1166] “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);NH(CH<sub>2</sub>CH<sub>2</sub>NH)nCH<sub>2</sub>CH<sub>2</sub>-氨基(其中氨基可以是例如如本文描述的),-NHC(O)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描述的氨基取代。

[1167] 糖基还可以包含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,修饰的核酸可以包括含有例如阿拉伯糖作为糖的核昔酸。核昔酸“单体”可以在糖的1'位置处具有 $\alpha$ 键联,例如 $\alpha$ -核昔。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。修饰的核酸还可以包括一种或多种处于L型的糖,例如L-核昔。

[1168] 通常,RNA包括糖基核糖,它是具有氧的5元环。示例性修饰的核昔和修饰的核昔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)。在某些实施方式中,修饰的核昔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)。

#### [1169] 12.4核碱基上的修饰

[1170] 能够被掺入修饰的核酸中的本文描述的修饰的核昔和修饰的核昔酸可以包括修饰的核碱基。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替换,以提供可以被掺入修饰的核酸中的修饰的核昔和修饰的核昔酸。核昔酸的核碱基可以独立地选自嘌呤、嘧啶、嘌呤或嘧啶类似物。在某些实施方式中,核碱基可以包括例如天然存在的碱基及其合成的衍生物。

#### [1171] 12.4.1尿嘧啶

[1172] 在某些实施方式中,修饰的核碱基是修饰的尿嘧啶。具有修饰的尿嘧啶的示例性核碱基和核昔包括但不限于假尿昔( $\Psi$ )、毗啶-4-酮核糖核昔、5-氮杂-尿昔、6-氮杂-尿昔、2-硫代-5-氮杂-尿昔、2-硫代-尿昔(s2U)、4-硫代-尿昔(s4U)、4-硫代-假尿昔、2-硫代-假尿昔、5-羟基-尿昔(ho5U)、5-氨基烯丙基-尿昔、5-卤代-尿昔(例如,5-碘代-尿昔或5-溴代-尿昔)、3-甲基-尿昔(m3U)、5-甲氧基-尿昔(mo5U)、尿昔5-氨基乙酸(cmo5U)、尿昔5-氨基乙酸甲酯(mcmo5U)、5-羧甲基-尿昔(cm5U)、1-羧甲基-假尿昔、5-羧基羟甲基-尿昔(chm5U)、5-羧基羟甲基-尿昔甲酯(mchm5U)、5-甲氧羰基甲基-尿昔(mcm5U)、5-甲氧羰基甲基-2-硫代-尿昔(mcm5s2U)、5-氨基甲基-2-硫代-尿昔(nm5s2U)、5-氨基甲基-尿昔(mnm5U)、5-氨基甲基-2-硫代-尿昔(mnm5s2U)、5-氨基甲基-2-硒代-尿昔(mnm5se2U)、5-氨基甲酰基甲基-尿昔(ncm5U)、5-羧甲基氨基甲基-尿昔(cmmnm5U)、5-羧甲基氨基甲基-2-硫代-尿昔(cmnm5s2U)、5-丙炔基-尿昔、1-丙炔基-假尿昔、5-牛磺酸甲基-尿昔( $\tau$ cm5U)、1-牛磺酸甲基-假尿昔、5-牛磺酸甲基-2-硫代-尿昔( $\tau$ m5s2U)、1-牛磺酸甲基-4-硫代-假尿昔、5-甲基-尿昔(m5U,即具有核碱基脱氧胸腺嘧啶)、1-甲基-假尿昔(m1 $\Psi$ )、5-甲基-2-硫代-尿昔(m5s2U)、1-甲基-4-硫代-假尿昔(m1s4 $\Psi$ )、4-硫代-1-甲基-假尿昔、3-甲基-假尿昔(m3 $\Psi$ )、2-硫代-1-甲基-假尿昔、1-甲基-1-去氮杂-假尿昔、2-硫代-1-甲基-1-去氮杂-假尿昔、二氢尿昔(D)、二氢假尿昔、5,6-二氢尿昔、5-甲基-二氢尿昔(m5D)、2-硫代-二氢尿昔、2-硫代-二氢假尿昔、2-甲氧基-尿昔、2-甲氧基-4-硫代-尿昔、4-甲氧基-假尿昔、4-甲氧基-2-硫代-假尿昔、N1-甲基-假尿昔、3-(3-氨基-3-羧丙基)尿昔(acp3U)、1-甲基-3-(3-氨基-3-羧丙基)假尿昔(acp3 $\Psi$ )、5-(异戊烯基氨基甲基)尿昔(inm5U)、5-(异戊烯基氨基甲基)-2-硫代-尿昔(inm5s2U)、 $\alpha$ -硫代-尿昔、2'-0-甲基-尿昔(Um)、5,2'-0-二甲基-尿昔(m5Um)、2'-0-甲基-假尿昔( $\Psi$ m)、2-硫代-2'-0-甲基-尿昔(s2Um)、5-甲氧基羰甲基-2'-0-甲基-尿昔(mcm5Um)、5-氨基甲酰基甲基-2'-0-甲基-尿昔(ncm5Um)、5-羧甲基氨基甲基-2'-0-甲基-尿昔。

昔 (cmnm 5Um)、3,2'-0-二甲基-尿昔 (m3Um)、5-(异戊烯基氨基甲基)-2'-0-甲基-尿昔 (inm 5Um)、1-硫代-尿昔、脱氧胸昔、2'-F-阿糖 (ara)-尿昔、2'-F-尿昔、2'-OH-阿糖-尿昔、5-(2-甲氧甲酰基乙烯基) 尿昔、5-[3-(1-E-丙烯基氨基) 尿昔、吡唑并[3,4-d]嘧啶、黄嘌呤以及次黄嘌呤。

[1173] 12.4.2 胞嘧啶

[1174] 在某些实施方式中, 修饰的核碱基是修饰的胞嘧啶。具有修饰的胞嘧啶的示例性核碱基和核昔包括但不限于5-氮杂-胞昔、6-氮杂-胞昔、假异胞昔、3-甲基-胞昔 (m3C)、N4-乙酰基-胞昔 (act)、5-甲酰基-胞昔 (f5C)、N4-甲基-胞昔 (m4C)、5-甲基-胞昔 (m5C)、5-卤代-胞昔 (例如, 5-碘代-胞昔)、5-羟甲基-胞昔 (hm5C)、1-甲基-假异胞昔、吡咯并-胞昔、吡咯并-假异胞昔、2-硫代-胞昔 (s2C)、2-硫代-5-甲基-胞昔、4-硫代-假异胞昔、4-硫代-1-甲基-假异胞昔、4-硫代-1-甲基-1-去氮杂-假异胞昔、1-甲基-1-去氮杂-假异胞昔、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-胞昔、2-甲氧基-5-甲基-胞昔、4-甲氧基-假异胞昔、4-甲氧基-1-甲基-假异胞昔、赖西丁 (k2C)、 $\alpha$ -硫代-胞昔、2'-0-甲基-胞昔 (Cm)、5,2'-0-二甲基-胞昔 (m5Cm)、N4-乙酰基-2'-0-甲基-胞昔 (ac4Cm)、N4,2'-0-二甲基-胞昔 (m4Cm)、5-甲酰基-2'-0-甲基-胞昔 (f 5Cm)、N4,N4,2'-0-三甲基-胞昔 (m42Cm)、1-硫代-胞昔、2'-F-阿糖-胞昔、2'-F-胞昔以及2'-OH-阿糖-胞昔。

[1175] 12.4.3 腺嘌呤

[1176] 在某些实施方式中, 修饰的核碱基是修饰的腺嘌呤。具有修饰的腺嘌呤的示例性核碱基和核昔包括但不限于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-腺昔、7-去氮杂-腺昔、7-去氮杂-8-氮杂-腺昔、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-腺昔 (m1A)、2-甲基-腺昔 (m2A)、N6-甲基-腺昔 (m6A)、2-甲硫基-N6-甲基-腺昔 (ms2m6A)、N6-异戊烯基-腺昔 (i6A)、2-甲硫基-N6-异戊烯基-腺昔 (ms2i6A)、N6-(顺羟基异戊烯基) 腺昔 (io6A)、2-甲硫基-N6-(顺羟基异戊烯基) 腺昔 (ms2io6A)、N6-缩水甘油基氨基甲酰基-腺昔 (g6A)、N6-苏氨酰基氨基甲酰基-腺昔 (t6A)、N6-甲基-N6-苏氨酰基氨基甲酰基-腺昔 (m6t6A)、2-甲硫基-N6-苏氨酰基氨基甲酰基-腺昔 (ms2g6A)、N6,N6-二甲基-腺昔 (m62A)、N6-羟基正缬氨酰基氨基甲酰基-腺昔 (hn6A)、2-甲硫基-N6-羟基正缬氨酰基氨基甲酰基-腺昔 (ms2hn6A)、N6-乙酰基-腺昔 (ac6A)、7-甲基-腺昔、2-甲硫基-腺昔、2-甲氧基-腺昔、 $\alpha$ -硫代-腺昔、2'-0-甲基-腺昔 (Am)、N6,2'-0-二甲基-腺昔 (m6Am)、N6-甲基-2'-脱氧腺昔、N6,N6,2'-0-三甲基-腺昔 (m62Am)、1,2'-0-二甲基-腺昔 (m1Am)、2'-0-核糖基腺昔 (磷酸盐) (Ar (p))、2-氨基-N6-甲基-嘌呤、1-硫代-腺昔、8-叠氮基-腺昔、2'-F-阿糖-腺昔、2'-F-腺昔、2'-OH-阿糖-腺昔以及N6-(19-氨基-五氧杂十九烷基)-腺昔。

[1177] 12.4.4 鸟嘌呤

[1178] 在某些实施方式中, 修饰的核碱基是修饰的鸟嘌呤。具有修饰的鸟嘌呤的示例性核碱基和核昔包括但不限于肌昔 (I)、1-甲基-肌昔 (m1I)、怀俄昔 (imG)、甲基怀俄昔 (mimG)、4-去甲基-怀俄昔 (imG-14)、异怀俄昔 (imG2)、怀丁昔 (yW)、过氧怀丁昔 (o2yW)、羟基怀丁昔 (OHyW)、修饰不足的羟基怀丁昔 (OHyW\*)、7-去氮杂-鸟昔、辨昔 (Q)、环氧辨昔

(oQ)、半乳糖基-辨昔 (galQ)、甘露糖基-辨昔 (manQ)、7-氰基-7-去氮杂-鸟昔 (preQ0)、7-氨基甲基-7-去氮杂-鸟昔 (preQ1)、古嘌昔 (G+)、7-去氮杂-8-氮杂-鸟昔、6-硫代-鸟昔、6-硫代-7-去氮杂-鸟昔、6-硫代-7-去氮杂-8-氮杂-鸟昔、7-甲基-鸟昔 (m7G)、6-硫代-7-甲基-鸟昔、7-甲基-肌昔、6-甲氨基-鸟昔、1-甲基-鸟昔 (m'G)、N2-甲基-鸟昔 (m2G)、N2,N2-二甲基-鸟昔 (m2 2G)、N2,7-二甲基-鸟昔 (m2,7G)、N2,N2,7-二甲基-鸟昔 (m2,2,7G)、8-氧代-鸟昔、7-甲基-8-氧代-鸟昔、1-甲基-6-硫代-鸟昔、N2-甲基-6-硫代-鸟昔、N2,N2-二甲基-6-硫代-鸟昔、 $\alpha$ -硫代-鸟昔、2'-0-甲基-鸟昔 (Gm)、N2-甲基-2'-0-甲基-鸟昔 (m2Gm)、N2,N2-二甲基-2'-0-甲基-鸟昔 (m22Gm)、1-甲基-2'-0-甲基-鸟昔 (m'Gm)、N2,7-二甲基-2'-0-甲基-鸟昔 (m2,7Gm)、2'-0-甲基-肌昔 (Im)、1,2'-0-二甲基-肌昔 (m'Im)、06-苯基-2'-脱氧肌昔、2'-0-核糖基鸟昔 (磷酸盐) (Gr (p))、1-硫代-鸟昔、06-甲基-鸟昔、06-甲基-2'-脱氧鸟昔、2'-F-阿糖-鸟昔以及2'-F-鸟昔。

[1179] 12.5示例性修饰的gRNA

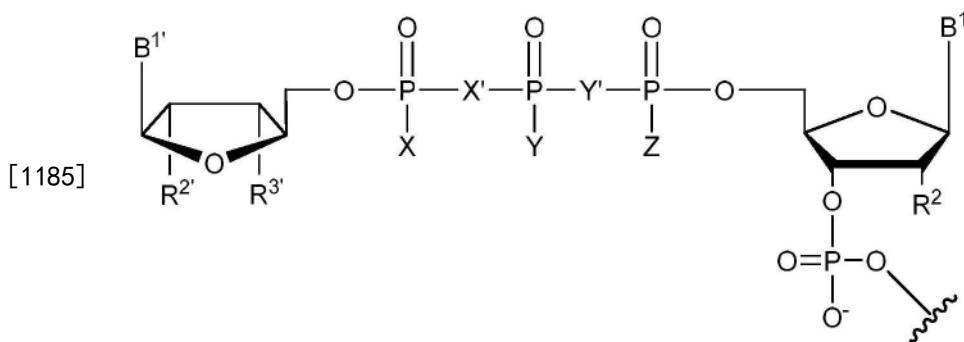
[1180] 在某些实施方式中,修饰的核酸可以是修饰的gRNA。应当理解,本发明所述的任何gRNA可根据本节进行修饰,包括任何包含靶向结构域的gRNA,所述靶向结构域包括选自SEQ ID NOS:208to 58749中所述的核苷酸序列。

[1181] 如上所述,瞬时表达或递送的核酸可易于被例如细胞核酸酶降解。因此,在一个方面中,本发明描述的修饰的gRNA分子可含有一个或多个修饰的核昔或核昔酸,以引入对核酸酶的稳定性。在某些实施方式中,当将本发明所述的某些修饰的gRNA分子引入细胞群,特别是本发明公开的细胞时,可表现出降低的先天性免疫应答。如上所述,术语“先天性免疫应答”包括对外源核酸的细胞应答,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)表达与释放的诱导以及细胞死亡的诱导。

[1182] 虽然本节中讨论的一些示例性修饰可包括在gRNA序列内的任何位置,但在某些实施方式中,gRNA分子在其5'端或其附近包含修饰(例如,在其5'末端的1-10、1-5或1-2个核苷酸内)。在某些实施方式中,gRNA在其3'末端或其附近(例如,在其3'末端的1-10、1-5或1-2个核苷酸内)包括修饰。在某些实施方式中,gRNA分子包括在其5'端或其附近的修饰以及在其3'端或其附近的修饰。

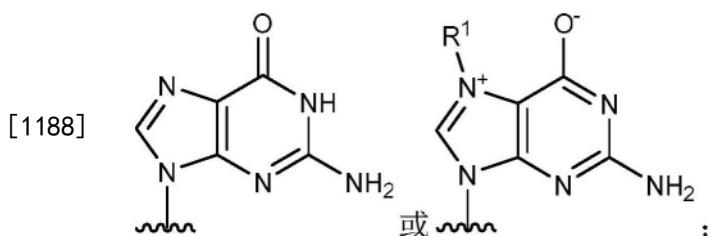
[1183] 在某些实施方式中,gRNA分子的5'末端缺少5'三磷酸基团。在某些实施方式中,靶向结构域的5'末端缺少5'三磷酸基团。在某些实施方式中,gRNA分子的5'末端包括5'盖。在某些实施方式中,靶向结构域的5'末端包括5'盖。在某些实施方式中,gRNA分子缺少5'三磷酸基团。在某些实施方式中,gRNA分子包括靶向结构域,且靶向结构域的5'末端缺少5'三磷酸基团。在某些实施方式中,gRNA分子包括5'盖。在某些实施方式中,gRNA分子包括靶向结构域,且靶向结构域的5'末端包括5'盖。

[1184] 在某些实施方式中,gRNA的5'末端通过包含真核mRNA帽结构或帽类似物(例如,无限制地,G(5')ppp(5')G帽类似物、m7G(5')ppp(5')G帽类似物、或3'-0-Me-m7G(5')ppp(5')G抗反向帽类似物(ARCA))进行修饰。在某些实施方式中,5'帽包含修饰的鸟嘌呤核苷酸,该鸟嘌呤核苷酸通过5'-5'三磷酸键与gRNA分子的其余部分连接。在某些实施方式中,5'帽类似物包含两个通过5'-5'三磷酸键连接的任选修饰的鸟嘌呤核苷酸。在某些实施方式中,gRNA分子的5'末端具有以下化学式:



[1186] 其中：

[1187]  $B^1$  和  $B^{1'}$  各自单独为



[1189] 每个  $R^1$  单独为  $C_{1-4}$  烷基, 可选地由苯基或6-元杂芳基取代;

[1190]  $R^2$ 、 $R^{2'}$  和  $R^{3'}$  中的每个单独为 H、F、OH 或  $O-C_{1-4}$  烷基;

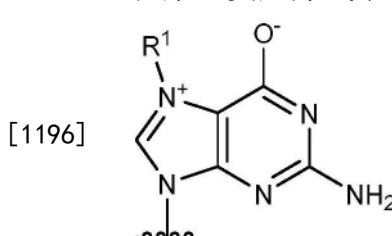
[1191] X、Y 和 Z 中的每个单独为 O 或 S; 以及

[1192]  $X'$  和  $Y'$  中的每个单独为 O 或  $CH_2$ 。

[1193] 在某些实施方式中, 每个  $R^1$  单独为  $-CH_3$ 、 $-CH_2CH_3$  或  $-CH_2C_6H_5$ 。

[1194] 在某些实施方式中,  $R^1$  为  $-CH_3$ 。

[1195] 在某些实施方式中,  $B^1$  为

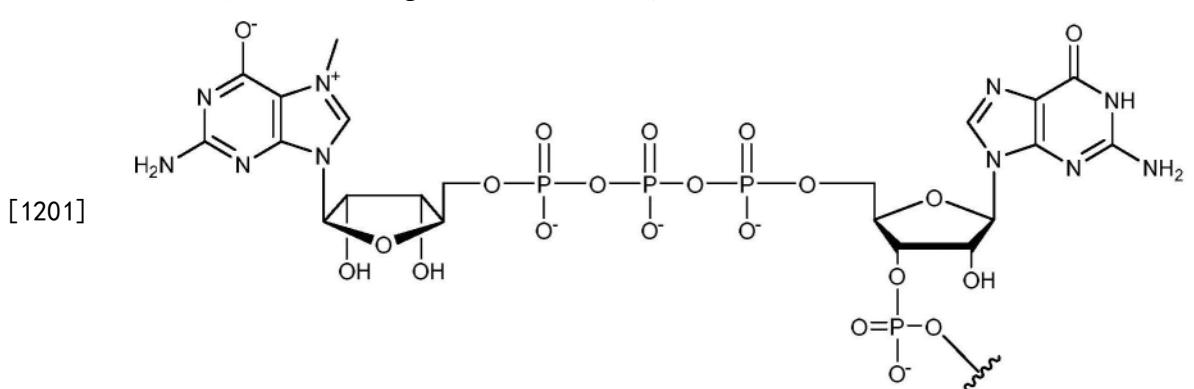


[1197] 在某些实施方式中,  $R^2$ 、 $R^{2'}$  和  $R^{3'}$  中的每个单独为 H、OH 或  $O-CH_3$ 。

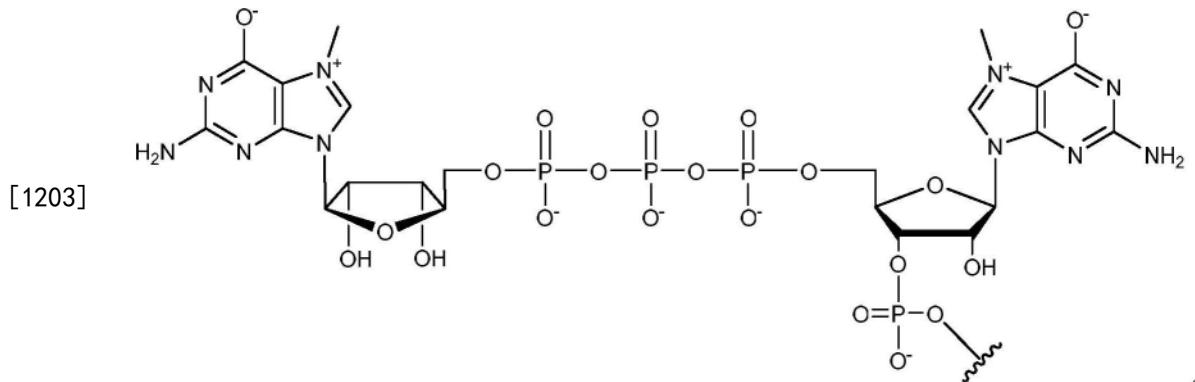
[1198] 在某些实施方式中, X、Y 和 Z 中的每个为 0。

[1199] 在某些实施方式中,  $X'$  和  $Y'$  中的每个为 0。

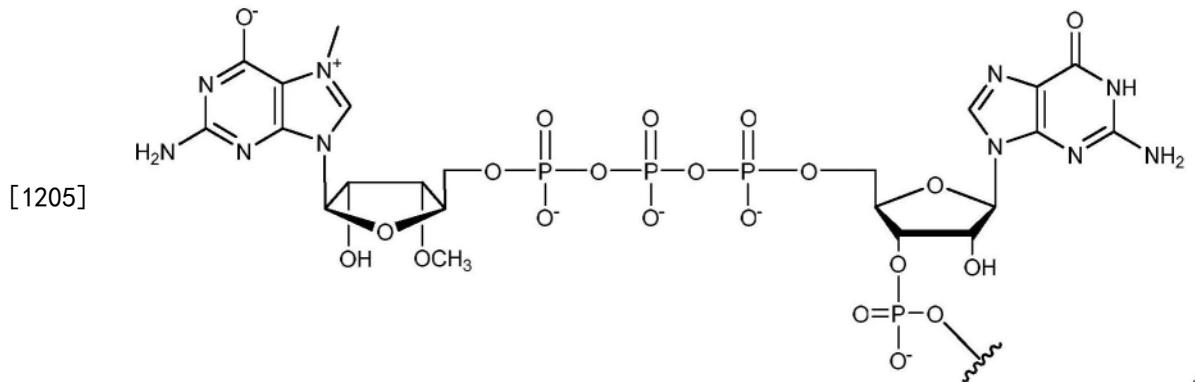
[1200] 在某些实施方式中, gRNA 分子的 5' 末端具有以下化学式:



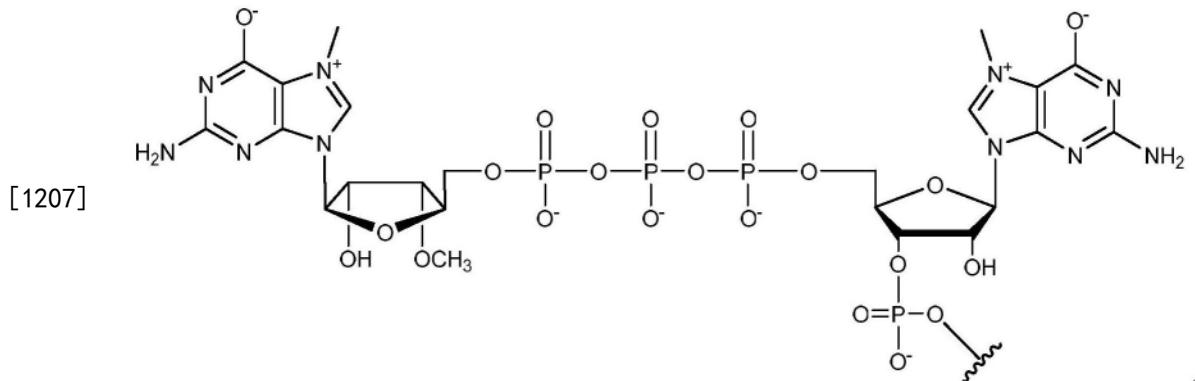
[1202] 在某些实施方式中, gRNA分子的5' 末端具有以下化学式:



[1204] 在某些实施方式中, gRNA分子的5' 末端具有以下化学式:



[1206] 在某些实施方式中, gRNA分子的5' 末端具有以下化学式:



[1208] 在某些实施方式中, X为S, Y和Z为0。

[1209] 在某些实施方式中, Y为S, X和Z为0。

[1210] 在某些实施方式中, Z为S, X和Y为0。

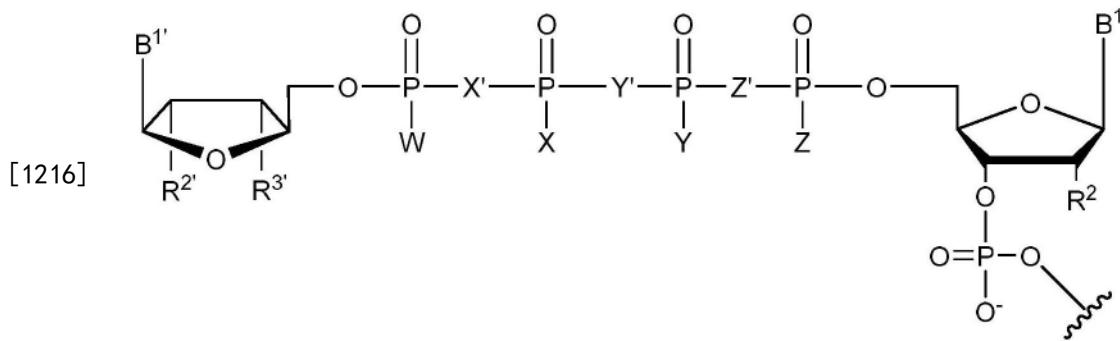
[1211] 在某些实施方式中, 硫代磷酸酯是Sp非对映体。

[1212] 在某些实施方式中, X' 为CH<sub>2</sub>, Y' 为0。

[1213] 在某些实施方式中, X' 为0, Y' 为CH<sub>2</sub>。

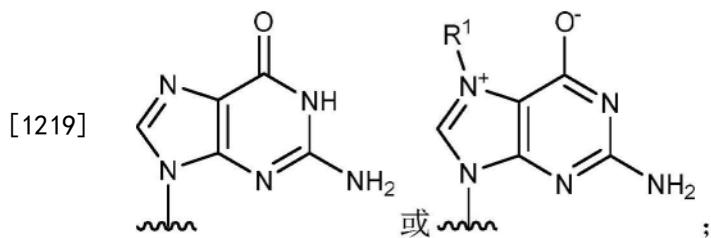
[1214] 在某些实施方式中, 5' 帽包含两个通过可选修饰的5' - 5' 四磷酸键连接的任选修饰的鸟嘌呤核苷酸。

[1215] 在某些实施方式中, gRNA分子的5' 末端具有以下化学式:



[1217] 其中：

[1218]  $B^1$  和  $B^{1'}$  各自单独为



[1220] 每个  $R^1$  单独为  $C_{1-4}$  烷基, 可选地由苯基或6-元杂芳基取代;

[1221]  $R^2$ 、 $R^2'$  和  $R^3$  中的每个单独为 H、F、OH 或  $O-C_{1-4}$  烷基;

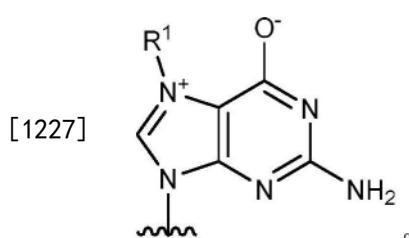
[1222]  $W$ 、 $X$ 、 $Y$  和  $Z$  中的每个单独为 0 或 S; 以及

[1223]  $X'$ 、 $Y'$  和  $Z'$  中的每个单独为 0 或  $CH_2$ 。

[1224] 在某些实施方式中, 每个  $R^1$  单独为  $-CH_3$ 、 $-CH_2CH_3$  或  $-CH_2C_6H_5$ 。

[1225] 在某些实施方式中,  $R^1$  为  $-CH_3$ 。

[1226] 在某些实施方式中,  $B^{1'}$  为



[1228] 在某些实施方式中,  $R^2$ 、 $R^2'$  和  $R^3$  中的每个单独为 H、OH 或  $O-CH_3$ 。

[1229] 在某些实施方式中,  $W$ 、 $X$ 、 $Y$  和  $Z$  中的每个为 0。

[1230] 在某些实施方式中,  $X'$ 、 $Y'$  和  $Z'$  中的每个为 0。

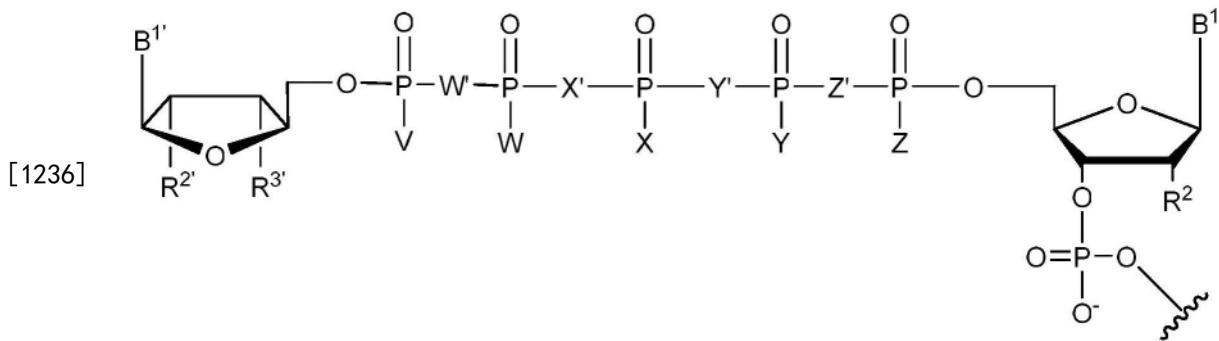
[1231] 在某些实施方式中,  $X'$  为  $CH_2$ ,  $Y'$  和  $Z'$  为 0。

[1232] 在某些实施方式中,  $Y'$  为  $CH_2$ ,  $X'$  和  $Z'$  为 0。

[1233] 在某些实施方式中,  $Z'$  为  $CH_2$ ,  $X'$  和  $Y'$  为 0。

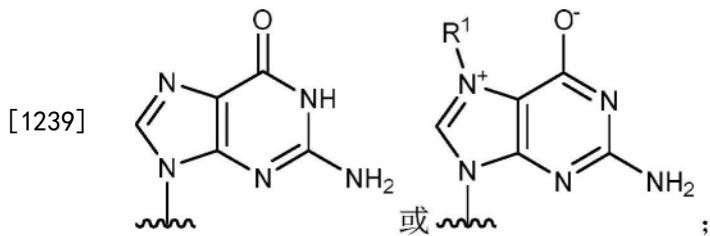
[1234] 在某些实施方式中, 5' 帽包含两个通过可选修饰的 5' - 5' 五磷酸键连接的任选修饰的鸟嘌呤核苷酸。

[1235] 在某些实施方式中, gRNA 分子的 5' 末端具有以下化学式:



[1237] 其中：

[1238]  $B^1$  和  $B^{1'}$  各自单独为



[1240] 每个  $R^1$  单独为  $C_{1-4}$  烷基, 可选地由苯基或6-元杂芳基取代;

[1241]  $R^2$ 、 $R^2'$  和  $R^3'$  中的每个单独为 H、F、OH 或  $O-C_{1-4}$  烷基;

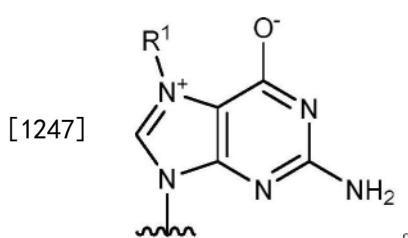
[1242] V、W、X、Y 和 Z 中的每个单独为 O 或 S; 以及

[1243]  $W'$ 、 $X'$ 、 $Y'$  和  $Z'$  中的每个单独为 O 或  $CH_2$ 。

[1244] 在某些实施方式中, 每个  $R^1$  单独为  $-CH_3$ 、 $-CH_2CH_3$  或  $-CH_2C_6H_5$ 。

[1245] 在某些实施方式中,  $R^1$  为  $-CH_3$ 。

[1246] 在某些实施方式中,  $B^1$  为



[1248] 在某些实施方式中,  $R^2$ 、 $R^2'$  和  $R^3'$  中的每个单独为 H、OH 或  $O-CH_3$ 。

[1249] 在某些实施方式中, V、W、X、Y 和 Z 中的每个为 O。

[1250] 在某些实施方式中,  $W'$ 、 $X'$ 、 $Y'$  和  $Z'$  中的每个为 O。

[1251] 如本发明使用的“5' 帽”涵盖传统的 mRNA 5' 帽结构, 但也包括这些的类似物。例如, 除了被上述化学结构所涵盖的5' 帽结构之外, 还可使用例如具有亚甲基-双(磷酸酯)部分的四磷酸类似物(例如参见Rydzik, A. M等人, (2009) Org Biomol Chem 7 (22) :4763-76)、具有非桥联氧硫取代的类似物(例如, 参见, Grudzien-Nogalska, E. 等人, (2007) RNA 13 (10) :1745-1755)、N7-苯化二核苷四磷酸类似物(例如, 参见Grudzien, E. 等人(2004) RNA 10 (9) :1479-1487)或防逆转帽类似物(例如, 参见美国专利7,074,596和Jemielity, J. 等人(2003) RNA 9 (9) :1108-1122和Stepinski, J. 等人, (2001) RNA 7 (10) :1486-1495)。本申请还包括使用具有卤素基团的帽类似物代替 OH 或 OMe(例如, 参见美国专利8,304,529); 与至少一个硫代磷酸酯(PS)连接的帽类似物(例如, 参见美国专利8,153,773和Kowalska, J. 等人(2012) RNA 18 (10) :2675-2685)。

人, (2008) RNA 14 (6) :1119-1131); 和具有至少一个硼烷磷酸酯或磷酸硒酯键的帽类似物(例如, 参见美国专利8,519,110); 以及炔基衍生的5' 帽类似物(例如, 参见美国专利8,969,545)。

[1252] 通常, 在化学合成或gRNA的体外转录过程中可包括5' 帽。在某些实施方式中, 不使用5' 帽, 而是通过用磷酸酶(例如, 小牛肠碱性磷酸酶) 处理除去5' 三磷酸基团而改变gRNA(例如体外转录的gRNA)。

[1253] 本发明公开的主题还提供了通过使用包含3' 聚腺苷酸尾巴(本发明中也称为聚腺苷酸酶(polyA)区域)的gRNA进行基因编辑的方法、基因组编辑系统和组合物。例如, 可通过在gRNA分子前体的体外转录后使用聚腺苷聚合酶将polyA尾添加至gRNA分子前体来制备此类gRNA。例如, 在某些实施方式中, 可使用聚合酶例如, 大肠杆菌polyA聚合酶(E-PAP)来酶促加入polyA尾。包含polyA尾的gRNA也可通过从DNA模板体外转录来制备。在某些实施方式中, 将限定长度的polyA尾编码在DNA模板上, 并通过RNA聚合酶(比如, T7 RNA聚合酶)与gRNA一起转录。体外转录后, 也可通过使用RNA连接酶或DNA连接酶在具有或不具有与gRNA分子前体和polyA寡核苷酸互补的夹板DNA寡核苷酸将polyA寡核苷酸连接至gRNA分子前体来制备具有polyA尾的gRNA。例如, 在一些实施方式中, 限定长度的polyA尾被合成为合成的寡核苷酸并在具有或不具有与向导RNA和polyA寡核苷酸互补的夹板DNA寡核苷酸的情况下, 用RNA连接酶或DNA连接酶连接到gRNA的3' 末端。包括polyA尾的gRNA也可以合成方式使一个或多个片段通过RNA连接酶或具有或不具有一个或多个夹板DNA寡核苷酸的DNA连接酶连接在一起制备。

[1254] 在某些实施方式中, polyA尾包含少于50个腺嘌呤核苷酸, 例如, 少于45个腺嘌呤核苷酸、少于40个腺嘌呤核苷酸、少于35个腺嘌呤核苷酸、少于30个腺嘌呤核苷酸、少于25个腺嘌呤核苷酸或少于20个腺嘌呤核苷酸。在某些实施方式中, polyA尾包含5至50个腺嘌呤核苷酸, 例如, 5至40个腺嘌呤核苷酸、5至30个腺嘌呤核苷酸、10至50个腺嘌呤核苷酸或15至25个腺嘌呤核苷酸。在某些实施方式中, polyA尾包含约20个腺嘌呤核苷酸。

[1255] 本发明还通过使用包含本发明所述的一个或多个修饰的核苷或核苷酸的gRNA来提供用于基因编辑(例如, 离体基因编辑)的方法、基因组编辑系统和组合物。

[1256] 虽然本节中讨论的一些示例性修饰可包括在gRNA序列内的任何位置, 但在某些实施方式中, gRNA在其5' 端或其附近包含修饰(例如, 在其5' 末端的1-10、1-5或1-2个核苷酸)。在某些实施方式中, gRNA在其3' 末端或其附近(例如, 在其3' 末端的1-10、1-5或1-2个核苷酸内)包括修饰。在某些实施方式中, gRNA包括在其5' 端或其附近的修饰以及在其3' 端或其附近的修饰。

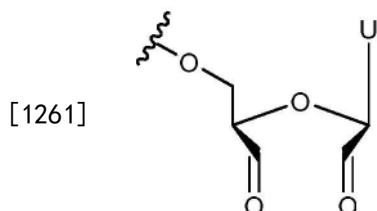
[1257] 在某些实施方式中, gRNA分子(例如, 体外转录的gRNA)包含与来自真核细胞中表达的基因的靶标结构域互补的靶向结构域, 其中gRNA分子在其5' 末端被修饰且包括3' polyA尾。例如, 该gRNA分子可缺少5' 三磷酸基团(例如, 靶向结构域的5' 末端缺少5' 三磷酸基团)。在某些实施方式中, 通过用磷酸酶(例如, 小牛肠碱性磷酸酶)处理以除去5' 三磷酸基团并包含如本发明所述的3' 聚腺苷酸尾来修饰gRNA(例如, 体外转录的gRNA)。可选地, gRNA分子可包括5' 帽(例如, 靶向结构域的5' 末端包含5' 帽)。在某些实施方式中, gRNA(例如, 体外转录的gRNA)含有本发明所述的5' 帽结构或帽类似物和3' polyA尾。在某些实施方式中, 5' 帽包含修饰的鸟嘌呤核苷酸, 该鸟嘌呤核苷酸通过5' -5' 三磷酸键与gRNA分子的其

余部分连接。在某些实施方式中,5' 帽包含两个通过可选修饰的5' - 5' 三磷酸键连接的任选修饰的鸟嘌呤核苷酸(例如,如上所述)。在某些实施方式中,polyA尾包含5至50个腺嘌呤核苷酸,例如,5至40个腺嘌呤核苷酸、5至30个腺嘌呤核苷酸、10至50个腺嘌呤核苷酸、15至25个腺嘌呤核苷酸、少于30个腺嘌呤核苷酸、少于25个腺嘌呤核苷酸或约20个腺嘌呤核苷酸。

[1258] 在某些实施方式中,本发明公开的主题提供一种gRNA分子,所述gRNA分子包括与来自真核细胞中表达的基因的靶标结构域互补的靶向结构域,其中,所述gRNA分子包括3 'polyA尾,所述polyA包含少于30个腺嘌呤核苷酸(例如,少于25个腺嘌呤核苷酸、15至25个腺嘌呤核苷酸或约20个腺嘌呤核苷酸)。在某些实施方式中,这些gRNA分子在它们的5' 末端被进一步修饰(例如,通过用磷酸酶处理以除去5' 三磷酸基团修饰gRNA分子或修饰gRNA分子以包括如本发明所述的5' 帽)。

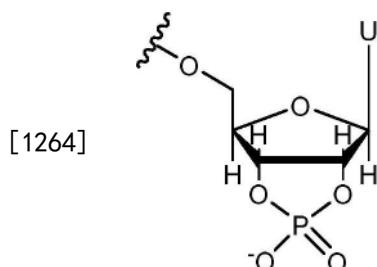
[1259] 在某些实施方式中,gRNA可在3' 末端U核糖处被修饰。在某些实施方式中,gRNA的5' 末端和3' 末端U核糖被修饰(例如,通过用磷酸酶处理以除去5' 三磷酸基团修饰gRNA分子或修饰gRNA分子以包括如本发明所述的5' 帽)。

[1260] 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,以提供如下所示的修饰的核苷:



[1262] 其中“U”可以是未修饰的或修饰的尿苷。

[1263] 在某些实施方式中,可以用如下所示的2' 3' 环状磷酸酯修饰3' 末端U:



[1265] 其中“U”可以是未修饰的或修饰的尿苷。

[1266] 在某些实施方式中,gRNA分子可以含有3' 核苷酸,其可以例如通过掺入本文描述的一个或多个修饰的核苷酸而相对于降解进行稳定化。在这个实施方式中,例如,尿苷可以被修饰的尿苷(例如,5-(2-氨基)丙基尿苷和5-溴代尿苷)或被本文描述的任何修饰的尿苷置换;腺苷、胞苷和鸟苷可以被修饰的腺苷、胞苷和鸟苷(例如,在8-位具有修饰,例如8-溴代鸟苷)或被本文描述的任何修饰的腺苷、胞苷和鸟苷置换。

[1267] 在某些实施方式中,可以向gRNA中掺入糖-修饰的核糖核苷酸,例如其中2' OH-基团被选自以下项的基团置换: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、杂环基氨基、芳氨基、二芳氨基、杂芳氨基、二杂芳氨基或氨基酸);或氰基(-CN)。在某些实施方式中,可以例如用硫代磷酸酯基团如本文

描述的修饰磷酸骨架。在某些实施方式中, gRNA的一个或多个核苷酸可以各自独立地是修饰的或未修饰的核苷酸, 包括但不限于2' -糖修饰的如2' -0-甲基、2' -0-甲氧基乙基, 或2' -氟修饰的, 包括例如, 2' -F或2' -0-甲基腺苷(A)、2' -F或2' -0-甲基胞苷(C)、2' -F或2' -0-甲基尿苷(U)、2' -F或2' -0-甲基胸苷(T)、2' -F或2' -0-甲基鸟苷(G)、2' -0-甲氧基乙基-5-甲基尿苷(Teo)、2' -0-甲氧基乙基腺苷(Aeo)、2' -0-甲氧基乙基-5-甲基胞苷(m5Ceo)、及其任何组合。

[1268] 在某些实施方式中, 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; 0-氨基(其中氨基可以是例如NH<sub>2</sub>; 烷氨基、二烷氨基、杂环基氨基、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0(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、杂环基氨基、芳氨基、二芳氨基、杂芳氨基或二杂芳氨基、乙二胺或聚氨基)。

[1269] 在某些实施方式中, gRNA可以包括修饰的核苷酸, 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'  $\rightarrow$  2')置换)。

[1270] 通常, gRNA分子包括糖基核糖, 它是具有氧的5元环。示例性修饰的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。在某些实施方式中, gRNA包含4' -S、4' -Se或4' -C-氨基甲基-2' -0-Me修饰。

[1271] 在某些实施方式中, 可以将去氮杂核苷酸(例如, 7-去氮杂-腺苷)掺入gRNA中。在某些实施方式中, 可以将0-和N-烷基化的核苷酸(例如, N6-甲基腺苷)掺入gRNA中。在某些实施方式中, gRNA分子中的一个或多个或所有核苷酸是脱氧核苷酸。

[1272] 12.6miRNA结合位点

[1273] 微小RNA(或miRNA)是天然存在的19-25个核苷酸长的细胞非编码RNA。它们结合至例如在mRNA的3' UTR中具有适当miRNA结合位点的核酸分子, 并且下调基因表达。在某些实施方式中, 这种下调通过减少核酸分子稳定性或抑制转录而发生。本发明所公开的RNA种类(例如, 编码Cas9的mRNA)可以例如在其3' UTR中包含miRNA结合位点。miRNA结合位点可以被选择为促进所选细胞类型中的表达下调。例如, 掺入miR-122的结合位点(一种丰富的肝脏中的微RNA)可抑制肝脏中目的基因的表达。

## 实施例

[1274] 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。

[1275] 实施例1: 候选指导RNA(gRNA)的评估

[1276] 候选gRNA的适用性可如本实施例所述进行评估。虽然描述了嵌合gRNA, 但该方法也可用于评估模块化gRNA。

[1277] 将gRNA克隆进载体中

[1278] 针对每种gRNA,设计并获得一对重叠寡核苷酸。使寡核苷酸退火并将其连接进含有上游U6启动子和长嵌合gRNA的剩余序列的经消化的载体骨架中。对质粒进行序列验证并制备以产生足够量的转染质量DNA。备选的启动子可用于驱动体内转录(例如,H1启动子)或用于体外转录(例如,T7启动子)。

[1279] 将gRNA克隆进线性dsDNA分子中(STITCHR)

[1280] 针对每种gRNA,设计并获得单个重叠寡核苷酸。将U6启动子和gRNA支架(例如除靶向结构域包括一切,例如包括衍生自crRNA和tracrRNA的序列,例如包括第一互补结构域;连接结构域;第二互补结构域;近端结构域;和尾部结构域)分开地PCR扩增并纯化为dsDNA分子。在PCR反应中使用gRNA特异性寡核苷酸,以便将通过寡核苷酸中指定的靶向结构域连接的U6和gRNA支架缝在一起。将所得dsDNA分子(STITCHR产物)纯化用于转染。备选的启动子可用于驱动体内转录(例如,H1启动子)或用于体外转录(例如,T7启动子)。任何gRNA支架都可以用于创造与来自任何细菌物种的Cas9相容的gRNA。

[1281] 将待测试的每种gRNA连同表达Cas9的质粒和少量的表达GFP的质粒转染进人类细胞中。在预实验中,这些细胞可以是永生人类细胞系,如293T、K562或U2OS。可替代地,可以使用原代人类细胞。在这种情况下,细胞可能与最终的治疗性细胞靶相关。使用类似于潜在治疗靶细胞群的原代细胞可以提供关于在内源染色质和基因表达的背景下的基因靶向率的重要信息。

[1282] 可使用脂质转染(如Lipofectamine或Fugene)或通过电穿孔(如Lonza Nucleofection)进行转染。转染后,可以通过荧光显微法或通过流式细胞术确定GFP表达,以确认一致且高水平的转染。这些预转染可包括不同的gRNA和不同的靶向途径(17-mer、20-mer、核酸酶、双切口酶等),以确定哪些gRNA/gRNA组合给出最大活性。

[1283] 可通过T7E1型分析或测序、通过测量NHEJ诱导的靶基因座上的indel形成来评估每种gRNA的切割效率。或者,也可使用其他错配敏感性酶,例如,CeII/Surveyor核酸酶。

[1284] 对于T7E1测定,PCR扩增子是大约500-700bp,其中预期的切割位点不对称地放置。对PCR产物进行扩增、纯化和尺寸验证后,通过加热至95°C并且然后缓慢冷却使DNA变性并重新杂交。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。如果在原始模板DNA中存在indel,当扩增子变性并重新退火时,这会导致带有不同插入缺失的DNA链的杂交,并因此导致不完全匹配的双链DNA。可通过凝胶电泳或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)可以用于使用以下等式估计百分比NHEJ: %NHEJ =  $(1 - (1 - \text{切割的分数})^{1/2})$ 。T7E1测定对低至约2%-5%的NHEJ是敏感的。

[1285] 可使用测序代替T7E1测定法,或除T7E1测定法外,可使用测序。对于Sanger测序,将经纯化的PCR扩增子克隆进质粒骨架中,转化,小量制备并用单个引物测序。Sanger测序可以用于在通过T7E1确定NHEJ率之后确定indel的确切性质。

[1286] 测序还可使用下一代测序技术进行。当使用下一代测序时,扩增子可以为300-500bp,其中预期的切割位点不对称地放置。PCR后,可以将下一代测序衔接子和条形码(例如Illumina多元衔接子和索引)添加到扩增子的末端,例如用于在高通量测序(例如在Illumina MiSeq上)中使用。这种方法允许检测非常低的NHEJ率。

[1287] 实施例2:通过NHEJ评估基因靶向

[1288] 可以选择在初始测试中诱导最大水平的NHEJ的gRNA用于基因靶向效率的进一步评价。在这种情况下,细胞来源于疾病受试者,并且因此具有相关突变。

[1289] 转染后(通常是转染后2-3天),可以从大量的转染细胞中分离基因组DNA,并且PCR可以用于扩增靶区域。PCR后,可以通过测序确定用于产生所希望突变(敲除靶基因或去除靶序列基序)的基因靶向效率。对于Sanger测序,PCR扩增子可以是500-700bp长。对于下一代测序,PCR扩增子可以是300-500bp长。如果目的是敲除基因功能,则测序可以用于评估多少百分比的病毒拷贝已经经历导致将预期破坏基因功能的移码或大的缺失或插入的NHEJ诱导的indel。如果目的是去除特定序列基序,则测序可以用于评估多少百分比的病毒拷贝已经经历跨越这个序列的NHEJ诱导的缺失。

[1290] 实施例3:靶向合成HSV-1构建体的单个gRNA的活性的评估

[1291] 构建含有HSV-1序列的质粒作为报道基因以测量Cas9介导的靶DNA切割。该报道质粒pAF025编码由CMV启动子驱动的绿色荧光蛋白(GFP)。将靶HSV-1序列插入GFP的框架N-末端中,在它们之间具有P2A自切割肽序列。

[1292] 使用基于公共工具cas-offinder的定制向导RNA设计软件鉴定RNA分子(Bae等人, *Bioinformatics*. 2014; 30 (10) :1473-1475)。在本实施例中测试并列于表18和19中的每种gRNA分子作为STITCHR产物产生,并与表达金黄色葡萄球菌Cas9(pAF003)的质粒共转染到HEK293FT细胞中。pAF003质粒编码金黄色葡萄球菌Cas9,同时带有由CMV启动子驱动的N端和C端核定位信号(NLS)和C端三标记标签。通过Mirus TransIT-293转染试剂将编码gRNA和Cas9的DNA与靶质粒pAF025一起导入细胞中。转染两天后,通过胰蛋白酶消化将细胞从其生长板中移出,在PBS缓冲液中洗涤,并用BD Accuri流式细胞仪分析。

[1293] 图9示出了pAF025的质粒图谱。图10A-10B示出了由质粒pAF025中由于Cas9介导的HSV-1靶序列切割引起的转染细胞群的平均荧光(或相对荧光单位,RFU)测量的GFP表达降低。表18列出了图10A中示出的gRNA,表19列出了图10B中示出的gRNA。

[1294] 表18(图10A中示出的gRNA)

gRNA名称	gRNA的靶向结构域的核苷酸序列号SEQ ID NO.
HSV1-RL2-1094	GAGGCCGCCGAGGACGUCAG (SEQ ID NO:23519)
HSV1-RL2-1413	GGGGGGGUUGGGGUUGGGGU (SEQ ID NO:23587)
HSV1-RL2-2202	GCCCCUCCGGGGGUUGGGGU (SEQ ID NO:23583)
HSV1-RL2-2199	GUCUGGCCCUCCGGGGGU (SEQ ID NO:23580)
HSV1-RL2-2191	GGGGGGCGUCUGGCCCUCCGG (SEQ ID NO:23571)
HSV1-RL2-1248	GGGGCGUCUGGCCCUCCGG (SEQ ID NO:23569)
HSV1-RL2-2150	GCCCCCCCCGGCCCUAGUCGGAGG (SEQ ID NO:23527)
HSV1-RL2-2154	GCCUGUGGGAGAGGCCGGGG (SEQ ID NO:23531)
HSV1-RL2-899	GGGGGAGUCGCUGAACACUA (SEQ ID NO:23489)
HSV1-RL2-2158	GUCUCUGUUGUUUGCAAGGGGG (SEQ ID NO:23535)

[1296] 表19(图10B中示出的gRNA)

gRNA名称	gRNA的靶向结构域的核苷酸序列号SEQ ID NO.
HSV1-RS1-3106	GCGUCAUCGACCUCGUCGGACU (SEQ ID NO:3363)
HSV1-RS1-3105	GUCAUCGACCUCGUCGGACU (SEQ ID NO:3362)

HSV1-RS1-2314	GCGACAGGCGGUCCGUUGGGU (SEQ ID NO:2522)
HSV1-RS1-2307	GGGCGCGCGACAGGCGGUCCG (SEQ ID NO:2515)
HSV1-RS1-36	GCGCGCGACAGGCGGUCCG (SEQ ID NO:243)

[1298] 通过引用结合

[1299] 本文提及的所有出版物、专利和专利申请都通过引用以其全文而特此结合,如同每一单独的出版物、专利或专利申请具体且单独地指明通过引用而结合一样。在有冲突的情况下,以本申请(包括本文的任何定义)为准。

[1300] 等效物

[1301] 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。

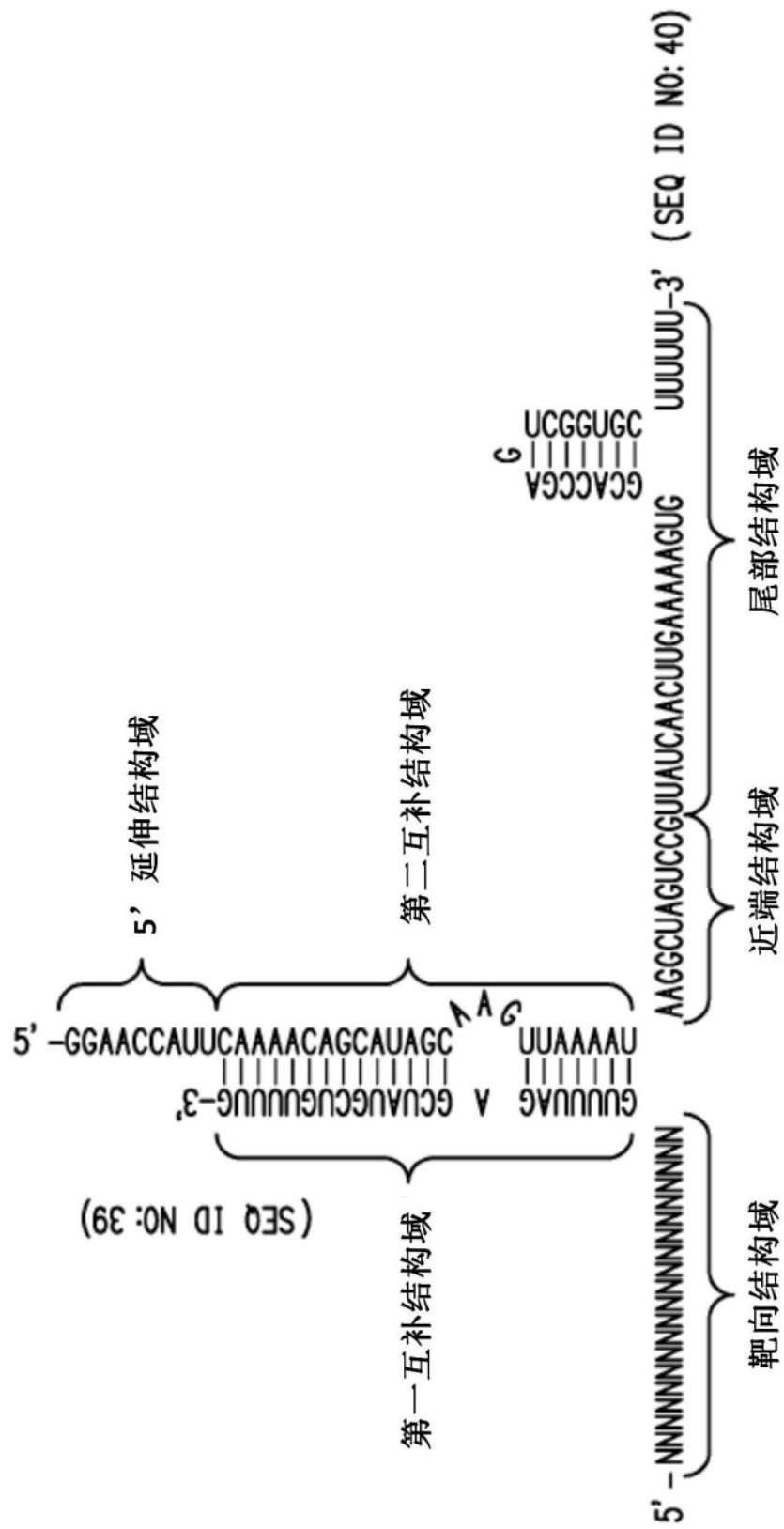


图1A

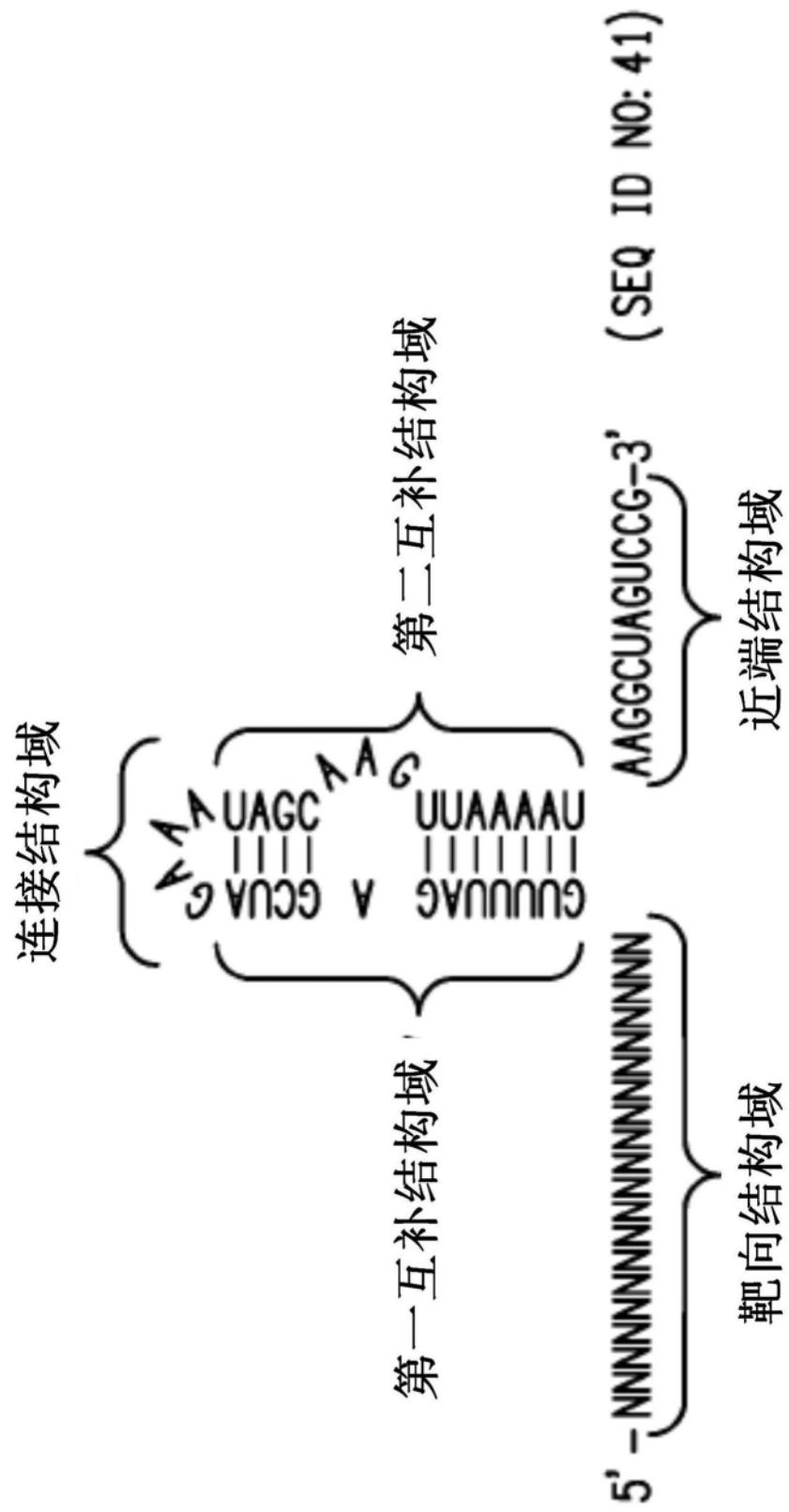


图1B

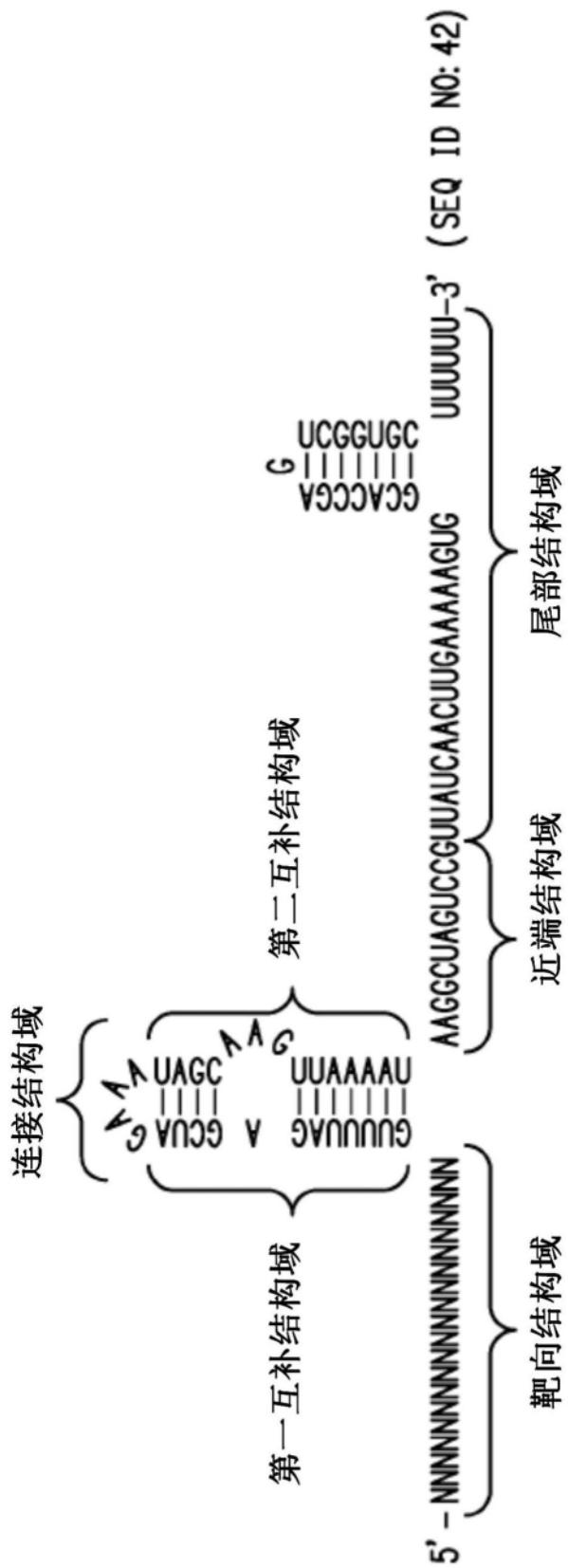


图1C

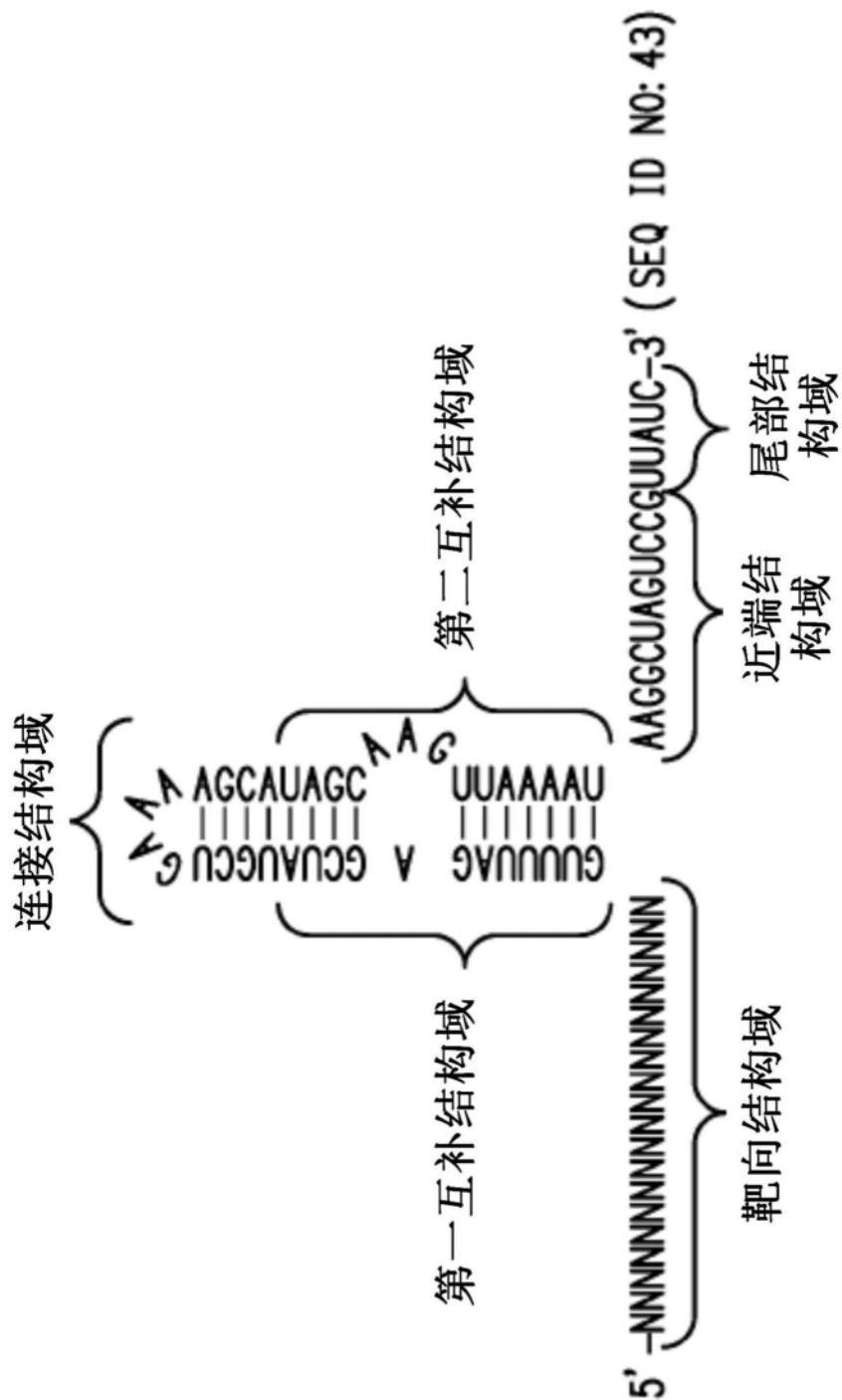


图1D

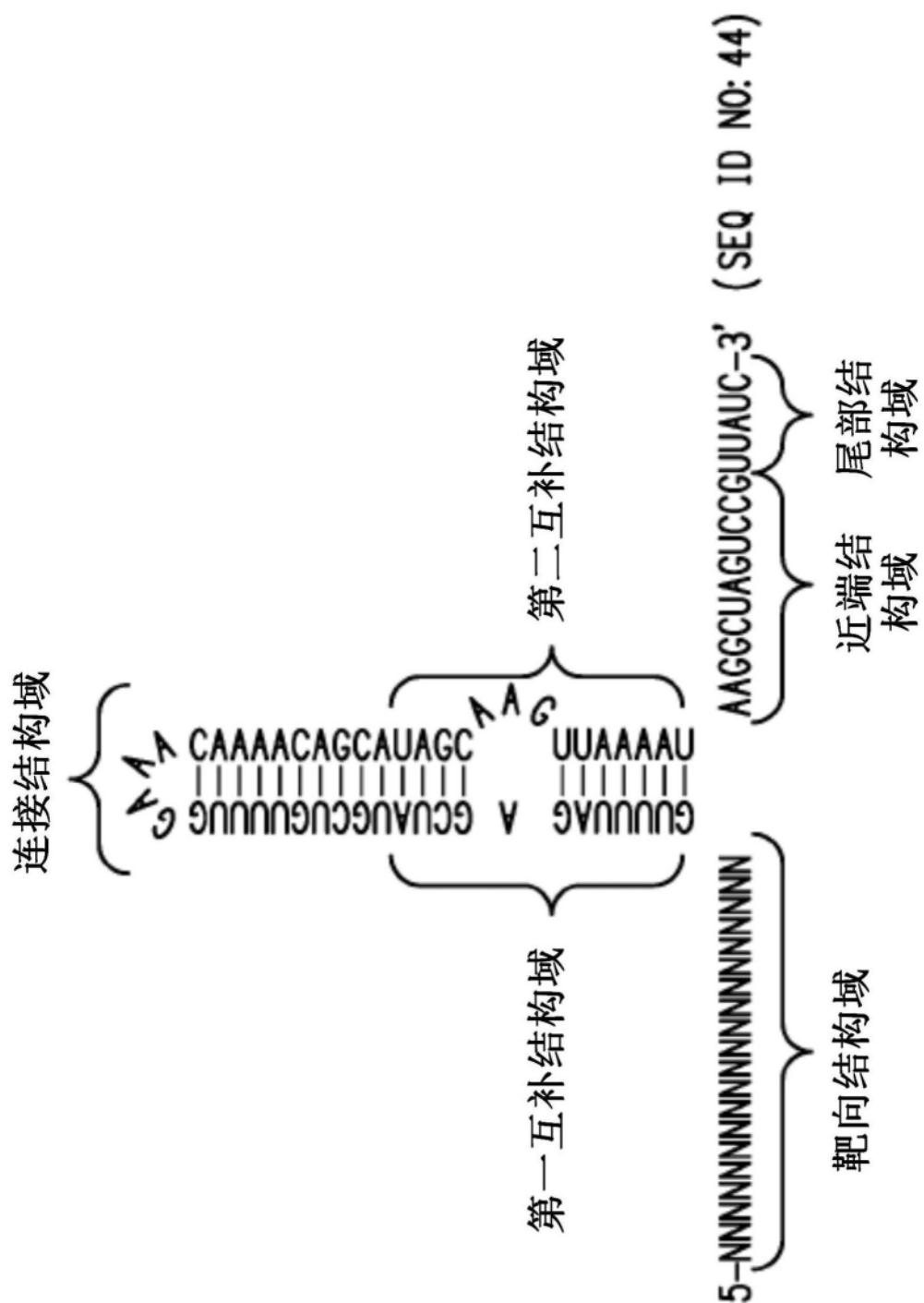


图1E

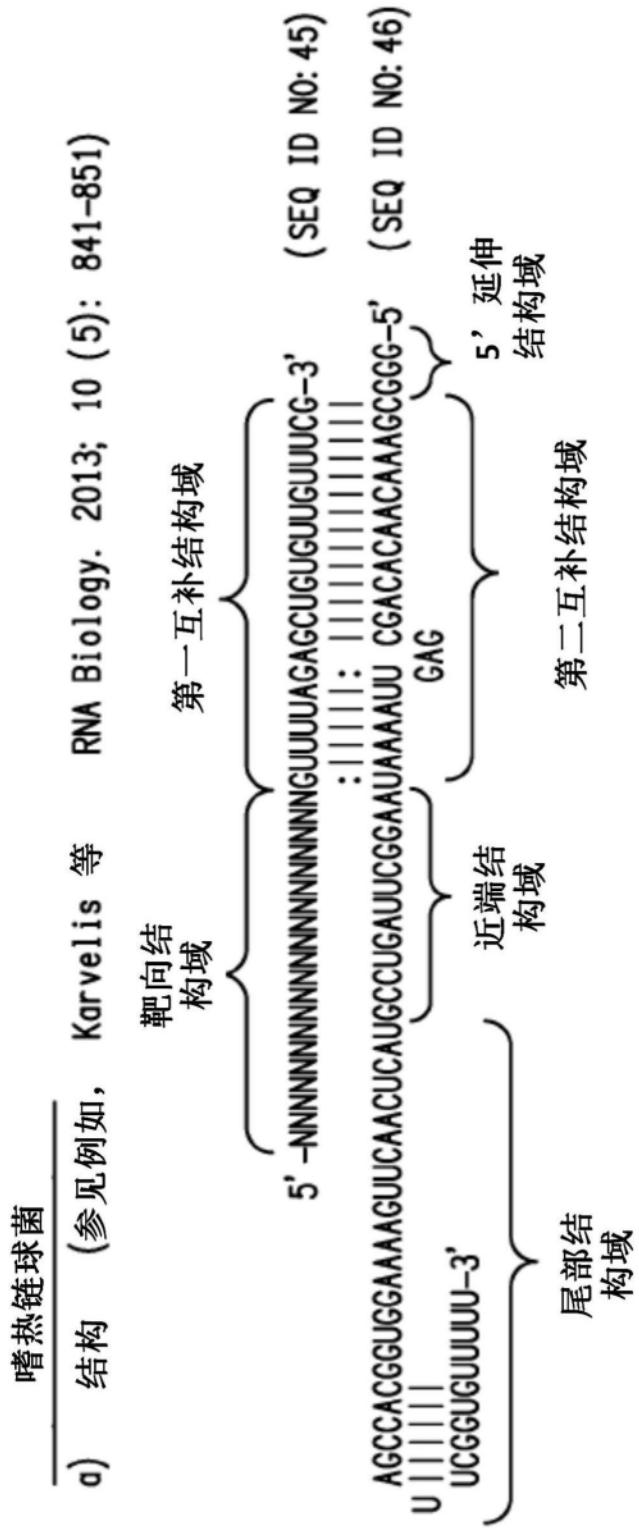


图1F

图1G

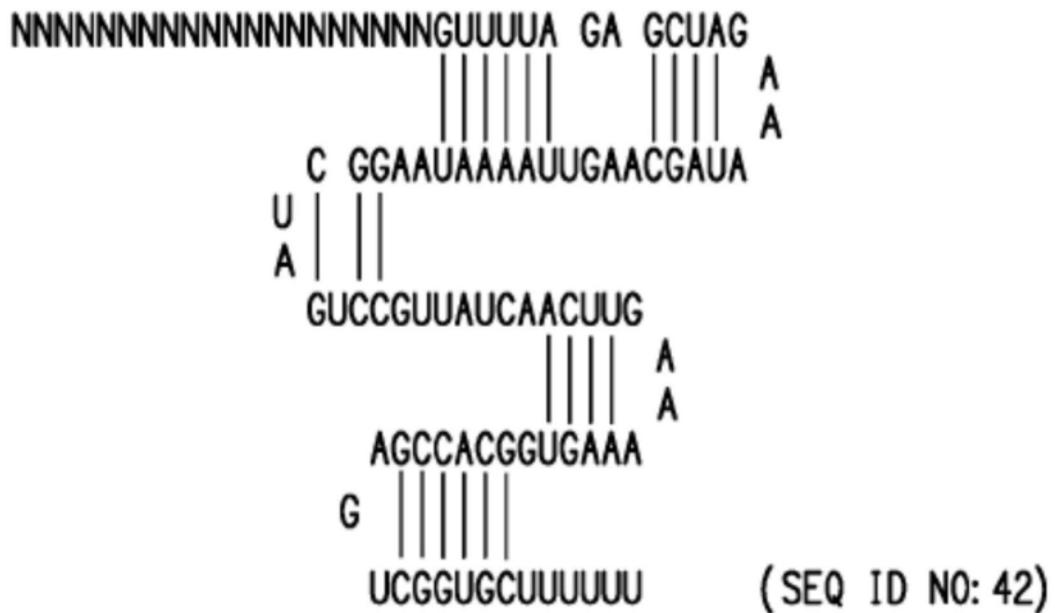


图 1H



图 11

通过MAFFT的CLUSTAL形式比对(v7.058b)

SM	MKKPYSIGDGTNSVGVAVYTDYKVPAKKMVKVLGNTDKSHIEKVLILLGALLFDSGNTAED	RLRKRTARRYTRRRNRILYQEIIFSEMEGKVDDDSFFHRLEDSFLVTEDKHERHPIFGN	LEEVKYHENFPTIYHLRQYLADNPEKYVDIRLVLVYALAHIIKFRGHHFLIEGKFDTRNNDV	SM	LEEVEYTHKNYPTIYHLREELVNSSEKADRLVYALAHIIKRYGNFLIEGALDTQNTSV
SP	MDKKYSIGDGTNSVGVAVYTDYKVPSKKFVKVLGNTDRHSIKKVLILLGALLFDSGCTREA	TRLRKRTARRYTRRRNRILYQEIIFSEMEGKVDDDSFFHRLEESFLVTEDKHERHPIFGN	IVDEVAYHEKPTIYHLRKKLVVDSTKADRLIYIYALAHMIKFRGHHFLIEGDLNPDNSDV	SP	*****;
ST	MTKPYSIGDGTNSVGVAVYTDYKVPSKKMVKVLGNTSKYIKKVLILLGVLLFDSGCTAEG	RRRKRTARRYTRRRNRILYQEIIFSEMEGKVDDDSFLVPOODKRSKXPIFGN	LVEEKAYHDEFPTIYHLRKYLADSTKADRLVYALAHMIKYRGHHFLIEGEFNSKNDI	ST	*****;
LI	MKKPYTIGDGTNSVGVAVYTDYQDLVKRMKIAGDSEKKQIKKNNFMGVRLFDEGQTAAD	REMRARTARRRIERRRNRIISYLGQIFAEMMSKTDANFFCRLSDSFYVUNNEKRNSRHPFFAT	*****;	LI	*****;
基序	M-K-Y*IGDIGTNSVGVAV-TD*Y-*--*K*K*-G*-*-I*KW*-G--LFD-G-TA--	*****;	*****;	基序	*****;

图2A

基序	SM	QALFQEFIAVINTPFENS-----LOEQINVQVEEILTDKISKSAKKDRVLKLFNEKSIN
	SP	DKLFQIOLVOTYNQOLFEENP-----INASGVDAKAILSARLSKSRRLENLIAQLPGEKKIN
	ST	QRNFQDFPLDTYNAIFESDL-----SENSKQLEEIVEDKISKLEKKDRILKLFPGEKNS
	LI	DGIIYKQFIQTYNQVFASGIEDGSLMKKLEDWAKILIVEKTRKEKLERILKLYPGEKSAA
		*****Y*-----F-----*
基序	SM	GRFAEFLKLIVGNOADFKKHFELEEKAPLOFSKDTYEELEVLAQIGDNYAELFLSAKK
	SP	GLFGNLLIALSGLGLTPNFKSNFDLAEDAKLQLSKOTYDDDLQNLIAQIGDQYADLFLAAKN
	ST	GIFSEFPLKLIVGNOADFPRKCFNLDEKAISLHFSKESYDDELETLLGIGDYYSDVFLKAKK
	LI	GMFAQFISLIVGSKGNFQKPFDLIERSDIECKADSYEEDLESLLALIGDEYASFLVAKRN
		*****-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*
基序	SM	LYDSILLSGILTVDVGTKAPLSASMIQRYNEHQMDLAQLKQFIHQQLSDKYNEVFSQDV
	SP	LSDAILLSDILRVRNTETIKAPLSASMIKRYDEHHQDLTLKALVRQQLPEKYKEIFFDQS
	ST	LYDAILLSSGFLTVTVDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLXTYNEVFKDDT
	LI	AYSAVVLLSIITVAETETNAKLSASMIERFDTHEEDLGELKAIFIKLHLPKHYEEIFSNT
		*****-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*
基序	SM	KDGYAGYIDGKTNQEAFFYKYLGLINKIEGSGYIFLDKIEREDFLRKQRTFDNGSIPHQIH
	SP	KNGYAGYIDGKTNQEEFYKFKPILEKNDGTTEELLVKLNREDLLRKQRTFDNGSIPHQIH
	ST	KNGYAGYIDGKTNQDFYVYILKKLIAEFGADYFLEKIDREDFLRKQRTFDNGSIPHQIH
	LI	KHGYAGYIDGKTKQADFYKYMONTLIEGADYFLAKIEKENFLRKQRTFDNGAIPHQIH
		*****-----*-----*-----*-----*-----*-----*-----*-----*-----*-----*
基序	SM	K-GYAGYIDG-----Q-----FY-----K-----L-----DL-----H-----E-----F-----*
	SP	
	ST	
	LI	

图2B

SM	LNEMALIRROAEFFPLADNQDRIEKLTFRIPIYYVPLARGKSDFAWLSRKSADKITP	SM	WNEFDEIVDKESSAEEAFINRNTNDYLPNQKVLPKHSLLYEKFTVYNELTKVYKITE-Q3	SM	KTAFFDANMKQEIFFDGVPFKVYRKVTDKLMDFLEKEFDEFRIVDLTGLDKENKVNASY	SM	TYHDLCKIL-DKDFLUNSKNEKILEDIVVLTLEDFREMIRKRLENYSSDLLTKEQVKLLE
SP	LGELHAILERQEDFYFLXDNREKIEKLTFRIPIYYVPLARGNSRFAMWTRKSEETITTP	SP	WNEFEDVVDKGASAQSFIERNNTSDLYLPEEKVLEKHSLLYETFNVYNEELTKVYRFLAESMR	SP	KPAFLSGEOKKAIIVDLLFPTKTRKUTVKQLKEDYFKKIECFDSVEISGVEDR-...-FNASLG	SP	TYHDLKLTIKDKDFLNEENEDILEDIVVLTLEDFREMIEERLKTYAHLFDKWMQQLK
ST	LOEMRAILDKQAKRFYFLAKNEKERIEKLTFRIPIYYVPLARGNSDFAWSIRKNEKITP	ST	WNEEERKUDFGKSAUDFIEKNTNQDYLPEKENVLUKHSLLCYQKYLVUNELTKVYRIND-Q5	ST	DYQFLDSKQKRDIVRLYFDDKRVTDKDIIEYL-HAIYGYDGIELKGIEKQ-...-FNSSLS	ST	TYHDLNLIINDKEFLDSSNEAILEEIIHTLIEDREMIKQRLSEKFENIEFDKSVLKKLS
LI	LEELEAILHQQAKYYPEFLKENYDKISLVTFRIFYFVGPLANGOSEFAWLTRKADGEIRP	LI	WNEEERKUDFGKSAUDFIEKNTNQDYLPEKENVLUKHSLLCYQKYLVUNELTKVYRIND-Q5	LI	KTSYFSGQQERKEQFLNDLFQOKRKVKICKDLEFLF-RNMSHVESPTIEGLEDS-...-FNSSYS	LI	TYHDLKVGIGKQELLINPVNTMLENIVKILTVEDKEMIKEQLOQQFSQSDVLDGVVLUKMLE
基 序 :	L-E*-AI*-Q--*YPFYL--N--*I--*T--*TFRIPIY*VGPLA-G*S-FAN--RK--*Y-P	基 序 :	WNEEERKUDFGKSAUDFIEKNTNQDYLPEKENVLUKHSLLCYQKYLVUNELTKVYRIND-Q5	基 序 :	WNEEERKUDFGKSAUDFIEKNTNQDYLPEKENVLUKHSLLCYQKYLVUNELTKVYRIND-Q5	基 序 :	WNEEERKUDFGKSAUDFIEKNTNQDYLPEKENVLUKHSLLCYQKYLVUNELTKVYRIND-Q5

图 2C

图2D

SM KILLSAKLITQRKFDNLTKAERGGLTDDKAGFIKRQLVETRQITKHVARILDERFNTETD  
 SP QLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDLSRMNTKYD  
 ST QLLKSKLISQRKFDNLTKAERGGLSPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKED  
 LI KLYQGNLMSRKFDYLTKAERGGLTEADKARFIHRQLVETRQITKNVANILHQRFNYEKD  
 基序 : :\* ..: \* : : \* \* \* \* : \*\*\* \* : \* \* \* \* \* : \* ..: \* : \*  
 \*L---\*L\*\*\*RKFD-LTKAERGGL---DKA-FI\*RQLVETRQITK\*VA-\*L---N---D

SM EINNKKIRQVKIVTLKSNLVSNFRKEFELYKREINDY\*HANHDAYLN\*NAVIGKALLGVYPQL  
 SP ENDKLIREVVKITLKSCLVSDFRKDFQFYKVRREINNY\*HANHDAYLN\*AVVGTALIKKYPKL  
 ST EINNRAVRTVKIIITLKSTLVSQFRKDFELYKVRREINDF\*HANHDAYLN\*AVVASALLKKYPKL  
 LI DHGNTMKQVRIVTLKSALVSQFRKQFQLYKVRDVNDY\*HANHDAYLN\*GVVANTLLKVYPQL  
 基序 : : .. : \* : : \* \* \* \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \*  
 \*-----V\*\*\*TLKS-LVS\*FRK\*F\*\*LYKV\*\*N\*\*HANHDAYLN-V\*---L\*---YPL

SM EPEFVYGDYPFHGHKE-----NK-ATAKKFFYSNIMMFKKDDVRTD-----  
 SP ESEFVYGDYKVDVRKMIAKSEQEIGK-ATAKYFFYSNIMMFKKTEITLANGEIRKRPLI  
 ST EPEFVYGDYPKYNFSRE-----RKSATEKVVYFYSNIMMFKKSIISLADGRVIERPLI  
 LI EPEFVYGDYHQFDWPKA-----NK-ATAKQFYTNIMLFQAQKDRIID-----  
 基序 : \*.\* \* \* : .. : \* \* \* \* : \* .. : \* : \* : \* : \* : \* : \* : \* : \*  
 E-EFVYGDY-----K-AT-K--FY\*NIM-\*F-----\*

SM -----KNGEI IWKKDEHISNIKKVLSYPQVNIVKKVEEQTGGFSKE-----SILPK  
 SP ETNGETGEIIVWDKGDRDFATVRKVLSPQVNIVKKTEVQTGGFSKE-----SILPK  
 ST EVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRGKPGLFNANLSSPK  
 LI -----ENGEI LWDK-KYLDTVKKVMSYRQMNIVKKTEI0KGEFSKA-----TIKPK  
 基序 : : .. : \* : .. : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \* : \*  
 -----GE-\*W-K-----V\*M--Q\*N\*VKK-E-Q-----\*-----PK

图2E

图 2F

SM	KR-YTSTTEILNATLHQSIITGLYETRIDLNKLGGD	(SEQ ID NO: 1)
SP	KR-YTSTKEVLDATLHQSIITGLYETRIDLSQLGGD	(SEQ ID NO: 2)
ST	YRDYTPSSLLKDATLHQSIITGLYETRIDLAKLGEGLGG	(SEQ ID NO: 4)
LI	KR-YNNLKELLNSTLIIYQSIITGLYESRKRLD----D	(SEQ ID NO: 5)
	*****:*****:*****:*****:*****:*****:	
	-R-Y-----*-----T*I*QS*TGLYE*R--L-----	(SEQ ID NO: 14)

图2G

在Chybinski等中公开的N末端RuvC样结构域的比对  
(排除序列超出部分)  
(通过MAFFT的CLUSTAL形式比对 (v7.058b))

1, 12	DIGTNSVGMWAVT	(SEQ ID NO: 120)
3, 20	DVGITNSVGMWAVT	(SEQ ID NO: 121)
15	DVGITSSVGMWAVT	(SEQ ID NO: 122)
4	DIGTASVGMWAVT	(SEQ ID NO: 123)
7	DVGITGSVGMWAVT	(SEQ ID NO: 52)
6	DIGTNSVGMWAVV	(SEQ ID NO: 53)
9	DIGTNSVGMWAVV	(SEQ ID NO: 54)
10	DIGTNSVGMWAVV	(SEQ ID NO: 55)
11	DIGTNSVGMWAVV	(SEQ ID NO: 56)
42	DLGITNSIGGMWAVV	(SEQ ID NO: 57)
48	DLGITNSIGGMWAVI-	(SEQ ID NO: 58)
43	DLGITNSIGGMWALV	(SEQ ID NO: 59)
2	DIGTNSVGMWCVT	(SEQ ID NO: 60)
14	DIGTNSVGMWCYT	(SEQ ID NO: 61)
5	DMGTCSSLGMWAVT	(SEQ ID NO: 62)
16	DIGTSSVGMWAI	(SEQ ID NO: 63)
8	DLGTCGSVGMWAVV	(SEQ ID NO: 64)
22	DLGVGSVGMWAVV	(SEQ ID NO: 65)
23	DLGIASIGGMWAI	(SEQ ID NO: 66)
24	DLGIASVGMWAVV	(SEQ ID NO: 67)
25	DLGVASVGMWIV	(SEQ ID NO: 68)
26	DIGIASVGMWAIL	(SEQ ID NO: 69)
28	DLGISSSVGMWSVI	(SEQ ID NO: 70)
32	DIGIASVGMWSVI	(SEQ ID NO: 71)
33	DVGIGSIGGMWAVI	(SEQ ID NO: 72)
39	DLGVGSIGGMWAVI	(SEQ ID NO: 73)
34	DIGYASIGGMWAVI	(SEQ ID NO: 74)
47	DTGITNSLGMWAVV	(SEQ ID NO: 75)
50	DLGITNSIGGMWLL	(SEQ ID NO: 76)
49	DIGTDSLGMWAVF	(SEQ ID NO: 77)
18	DIGGNSIGGMFAVV	(SEQ ID NO: 78)
41	DLGVGSIGGMWAVV	(SEQ ID NO: 79)
45	DLGIASCGWWGIV	(SEQ ID NO: 80)

图3A

21	DIGTASVGMCILT	(SBO ID NO: 81)
27	DIGICGSVGCGIL	(SBO ID NO: 82)
29	DIGITSVGYELI	(SBO ID NO: 83)
30	DIGITSVGFGLI	(SBO ID NO: 84)
31	DVGITSTGYAVL	(SBO ID NO: 85)
40	DLGITSPGYAIL	(SBO ID NO: 86)
17	DIGNASVGWSAF	(SBO ID NO: 87)
19	DVGTNSCCGVVAM	(SBO ID NO: 88)
35	DVGERSIGLAAV	(SBO ID NO: 89)
36	DVGGLNSVGILAAV	(SBO ID NO: 90)
37	DVGGLMSVGILAAI	(SBO ID NO: 91)
38	DVGTPSVGLAAI	(SBO ID NO: 92)
13	DIGTGSVGYACM	(SBO ID NO: 93)
44	DLGTTSIGFAHI	(SBO ID NO: 94)
46	DLGTTSIGSSVR	(SBO ID NO: 95)
	* * *	

图3B

在Chyliński等中公开的N末端RuvC样结构域的比对  
(通过MAFFT的CLUSTAL形式比对 (v7.058b) )

1, 12	D----IGTNSVGWAVT	(SEQ ID NO:120)
3, 20	D----VGTNSVGWAVT	(SEQ ID NO:121)
15	D----MGTNSVGWAVT	(SEQ ID NO:122)
4	D----VGTSSVGWAVT	(SEQ ID NO:123)
7	D----IGTASVGWAVT	(SEQ ID NO:52)
6	D----VGTGSVGWAVT	(SEQ ID NO:53)
9	D----IGTNSVGWAVV	(SEQ ID NO:54)
10	D----IGTNSVGWAVI	(SEQ ID NO:55)
52	D----IGTNSIGGWAVI	(SEQ ID NO:96)
11	D----IGTNSVGWAVL	(SEQ ID NO:56)
42	D----LGTNSIGGWAVV	(SEQ ID NO:57)
48	D----LGTNSIGGWAVI	(SEQ ID NO:58)
43	D----LGTNSIGWALV	(SEQ ID NO:59)
2, 14	D----IGTNSVGWCVT	(SEQ ID NO:60)
5	D----IGTNSVGYAVT	(SEQ ID NO:61)
16	D----MGTGSLGVWAVT	(SEQ ID NO:62)
8	D----IGTSSVGWAAI	(SEQ ID NO:63)
22	D----LGTGSVGWAVV	(SEQ ID NO:64)
23	D----LGVCSSVGWAVV	(SEQ ID NO:65)
24	D----LGIAASIGWAI	(SEQ ID NO:66)
68	D----LGIAASVGWAVV	(SEQ ID NO:67)
25	D----LGIAASVGWSIV	(SEQ ID NO:68)
26	D----LGIAASVGWAIL	(SEQ ID NO:69)
66	D----LGIAASVGWAVL	(SEQ ID NO:98)
59	D----IGIAASIGWAVI	(SEQ ID NO:99)
61	D----IGIAASVGWAI	(SEQ ID NO:100)
64	D----VGIASVGWAVI	(SEQ ID NO:101)
62	D----IGIAASVGWAL-	(SEQ ID NO:102)
67	D----IGIAASVGWAVV	(SEQ ID NO:103)
32	D----IGIAASVGWSVI	(SEQ ID NO:71)
28	D----LGISSVGWSVI	(SEQ ID NO:70)
63	D----IGITSVGWAVI	(SEQ ID NO:104)

图4A

33	D----VGIGSIGWAVI
57	D----LGISSLGWAIIV
39	D----LGVGSIGFAIV
34	D----IGYASIGWAVI
50	D----LGTNSIGWCLL
54	D----LGTNSIGWGLL
47	D----TGTNSLGWAIIV
49	D----IGTDSLGWAVF
51	D----LGSTSLGWAIF
58	D----IGISSIGWAFS
21	D----LGIASVGWCLT
45	D----LGIASCWGVV
18	D----IGSNSIGFAVV
65	D----IGTTSIGPSVI
29	D----IGITSVGYGLI
30	D----IGITSVGFGLI
44	D----LGTTSIGFAHI
27	D----IGIGSVGVGIL
41	D----LGVGSIGVAVA
31	D----VGITSTGYAVL
40	D----LGITSFGYAIL
53	D----IGTSSIGWLY
55	D----LGSNSLGFVFT
56	D----LGANSLGWPVV
17	D----IGNASVGWSAF
19	D----VGTNSCGWVAM
35	D----VGERSIGLAAV
36	D----VGLNSVGLAAV
37	D----VGLMSVGLAAI
38	D----VGTFSVGLAAI
13	D----IGTGSVGYACM
46	D----LGTNSIGSSVR
60	DIGLRIGITSCGWSI-
69	D----MGAKYTGVFYA
73	D----LGGKNTGFFSF
74	D----LGVKNTGVPSA
70	D----LGAKFTGVALY
71	D----LGGKFTGVCLS
72	D----LGGTYTGTFIT

图4B

在Chybinski等中公开的HNH样结构域的比对  
(通过MAFFT的CLUSTAL形式比对 (v7.058b) )

1	YDIDHIYPKS-LTKD-----	DSP-DNLVLCERTAN
2	-DIDHIIYPKS-KIKD-----	-DSP-DNLVLVULKNN
3	-DRDHIIYPQS-KIKD-----	-DSI-DNLVLVNKTIN
4	-DIDHIIYPKS-KIKD-----	-DSI-TNRLVLEKDDIN
5	-DIDHIIYPQS-KIKD-----	-DSI-SNRLVLVSSCN
6	-DIDHIIYPQS-KTMD-----	-DSL-NNRVLVKQYNN
7	-DQDHIIYPKS-KIYD-----	-DSL-ENRVLVKQNLN
8	-QIDHIIYPQS-LVKD-----	-DSP-DNRVLVVPSEN
9	-DIDHIIIPQA-FIKD-----	-NSI-DNRFVLTTSSKEN
10	-DIDHIIIPQA-FLKD-----	-NSI-DNKVLVSSASN
11	-DIDHIIIPQA-YTKD-----	-NSL-DNRFVLVSNTIN
12	-DIDHIIYPQS-FITD-----	-NSI-DNLVLTTSSAGN
13	-DVDHIIYPQS-FLKD-----	-DSI-DNKVLTRSDDN
14	-NIDHIIYPQS-MVKD-----	-DSL-DNKVLVQSEBIN
15	-DIDHILPQS-LIKD-----	-DSL-DNRFVLNATIN
16	-DIDHILPQS-FIKD-----	-DSL-ENRVLVKKAVN
17	-EVDHIIIPRS-YIKD-----	-DSI-DNKVLVIKKMN
18	-EVDHIIIPRS-YIKD-----	-DSP-ENKVLVYREEN
19	-DIDHIIIPQA-VTON-----	-DSI-DNRVLVARAEN
20	-EIDHIIIPYS-ISPD-----	-DSS-SNKVLVLAESN
21	-EIDHIIIPYS-LCPD-----	-DSS-ANKVLVHKQSN
22	-DIDHIIIPYS-RSMD-----	-DSY-SNKVLVLSGEN
23	-DIDHIIIPYS-KSMD-----	-DSF-NNKVLCLAEEN
24	-EIDHIIIPYS-RSPD-----	-DSY-MNKVLVFTQIN
25	-QIDHIIYPQS-RSMD-----	-DSY-MNKVLVLTDEN
26	-EIDHIIIPYS-RSPD-----	-DSL-SNKVLVLGSEN
27	-EIDHALPPS-RTWD-----	-DSF-NNKVLVLGSEN
28	-EIDHALPPS-RTWD-----	-DSF-NNKVLVLSEN
29	-EIDHIIIPIS-ISLD-----	-DSI-NNKVLVLSKAN
30	-EVDHIIIPIS-ISLD-----	-DSI-TNKLVLVTHEN
31	-QVDHALPPS-RSYD-----	-DSK-NNKVLVLTHEN
32	-EVDHILPLS-ITFD-----	-DSL-ANKVLVYATAN
33	-EVDHILPLS-ITFD-----	-DAR-SNKVLVYRSN

图5A

29  
 67  
 58  
 51  
 55  
 57  
 56  
 54  
 52  
 31  
 45  
 53  
 60  
 21  
 23  
 25  
 49  
 33  
 42  
 43  
 44  
 20  
 66  
 61  
 46  
 47  
 48  
 50  
 39  
 41  
 40  
 35  
 36  
 37  
 38  
 70  
 71  
 73

-EVDHIIIPRS-VSFD-----NSY-HNKVULVKQSEN  
 -DIDHILPPS-ITFD-----DSP-RNKVLVTSQEN  
 -EIDHILPPS-PSAD-----DSP-ANKVULCLARAN  
 -EIEHILPPS-LTLD-----DSM-ANKTVCFRQAN  
 -DIDHILPPS-VSLD-----DSA-ANKVUCLREAN  
 -DIDHILPPS-1SWD-----DSA-ANKVUCLMRYAN  
 -DIDHILPPS-MTLD-----DSP-ANK11CMRYAN  
 -DVDHILPPS-RTLD-----DSP-PNRTLCLREAN  
 -EIEHILPPS-RTLD-----DSL-NMNTVAMRAN  
 -EVDHIIIPYS-ISMD-----DSY-TNKVULTSAKCN  
 -QVDHILPPS-RFGD-----DSY-LANKTLCTARSN  
 -QVDHILPPS-RTLD-----DSP-ANKVLAQHDAN  
 -QIDHAPPIS-RSLD-----DSQ-SNKVULCLTSN  
 -DIDHIVPRS-ISFD-----DSP-SNLV1VNKLDN  
 -EIEHIIIPYS-MSYD-----NSQ-ANKILTEKAEN  
 -EIDHIVPRS-NSAD-----DSW-FNKLLVKKSTN  
 -EMDHILPPS-RSLD-----NCW-HNRVLVHGKDN  
 -EVDHIVPRS-LIIL-----NTI-HNKALVIAEN  
 -EIEHIVPOS-LYFD-----DSF-SNKV1CEAEVN  
 -DIEHIVPQA-RLFD-----DSP-SNKTLCEARSVN  
 -EIEHIVPKA-RYFD-----DSP-SNKTLTFHRIN  
 -DKDHIIPQS-MKKD-----DSI-IMNLV1VNKNAN  
 -EVEHIIWPRS-RSFD-----NSP-RNKTLCRKDVN  
 -IVNHIIPYN-RSFD-----DTY-HNRVLTLETK  
 -DMEHTIPKS-ISFD-----NSD-ONLTLCESYYN  
 -DIEHTIPRS-AGCD-----STK-MNLTLCSSRFN  
 -DIEHTIPRS-1SQD-----NSQ-MNKTLCSLKFN  
 -DIDHIVPLA-RGGR-----DSL-DMWVLQSDAN  
 -DIEHLPPIA-ESED-----NGR-NNLVISHACN  
 -DVDHIFPRD-DTAD-----NSY-GNKVVAHRQCN  
 -DIEHIVPQS-LGGL-----STD-YNTIVTLKSVN  
 -ELDHIVPRT-DGGS-----NRH-ENLAITCGACN  
 -EMDHIVPRKGVGST-----NTR-TNFAAVCACN  
 -EMDHIVPRKGVGST-----NTR-VNLAAACACN  
 -EMDHIVPRAGQGST-----NTR-ENLVAVCHRCN  
 -EIDHILPPS-LIKDARGIVNAE-PNLIYASSRGN  
 -EIDHIIIPRS-LTGRIKNTVENSE-ANLIYCSSKGN  
 -EIDHIIIPRS-LTLKKSESIYNSE-VNLIFUSAQGN  
 -EIDHIIIPRS-LTLKKSESIYNSE-VNLIFUSAQGN

图5B

(SEQ ID NO: 191)  
(SEQ ID NO: 192)  
(SEQ ID NO: 193)  
(SEQ ID NO: 194)

-EIDHTYPRS-LSKIKHPCV1FNSE-VNLIYCSSQGN  
-EIDHILPRS-HTLKIYGTVNPE-GNLIYVHQKCN  
-ELDHIIPRS-HKYY---GTLNDE-ANLICVTRGDN  
-BLEBHIYPHS-PRQS---NAL-SSLVLTWPGVN  
: \* . . .

图5C

在Chybinski等中公开的HNH样结构域的比对（排除序列超出部分）  
(通过MAFFT的CLUSTAL形式比对 (v7.058b))

1	YDIDHIIYPRS-LTKDDS-FDNLVLCERTAN
2	-DRDHIYPRSKV1KDDS-FDNLVLV1KNNEN
3	-DIDHIIYPRS-KIKDDS-IDNLVLV1KNTIN
4	-DIDHIIYPRS-KIKDDS-ITNRLVLVERDIN
5	-DIDHIIYPOS-KIKDDS-1SNRVLVYCSSCN
6	-DIDHIIYPOS-KTMDDDS-LNNRVLV1KNNN
7	-DQDHIIYPKS-KIYDDDS-LENRVLV1KNNL
8	-QIDHIIYPOS-LVYDDDS-FDNRVLVYVPSEN
9	-DIDHIIYPOA-PIKDDNS-IDNRLVLTTSKEN
10	-DIDHIIYPOA-PLKDDNS-IDNKVLVSSASN
11	-DIDHIIYPOA-YTKDDNS-LDNRLVLSNITN
12	-DIDHIIYPOS-PITDDNS-IDNLVLVLTSSAGN
13	-DVDHIIYPOS-PLKDDDS-IDNKVLVLTSDQN
14	-NIDHIIYPOS-MVKDDDS-LDNKVLVQSEIN
15	-DIDHIIYPOS-LIKDDDS-LDNRVLV1NATIN
16	-DIDHIIYPOS-PIKDDDS-LENRVLV1KAVN
17	-EVDHIIYPRS-PIKDDDS-IDNKVLV1KMMN
18	-EVDHIIYPRS-YIKDDDS-PENKVLVYREEN
19	-DIDHIIYPOA-VTQDDNS-IDNRLV1VARAEN
20	-DIDHIIYPRS-ISFDDDS-PSNLVLV1VNLIN
21	-EIDHIIYPPS-ISFDDDS-PSNLVLV1VNLIN
22	-EIDHIIYPPS-ISFDDDS-SANKVLV1LAESN
23	-EIDHIIYPPS-LCFDDDS-SANKVLV1QSN
24	-EIDHIIYPPS-ISLDDDS-1NNKVLV1LSKAN
25	-EVDHIIYPPS-ISLDDDS-1NNKVLV1YTHREN
26	-EVDHIIYPLS-ITFDDDS-LANKVLVYATAN
27	-EIDHIIYPRS-ISFDDA-RSNKVLVYRSEN
28	-EVDHIIYPRS-VSFDDNS-YHNNKVLV1QSEN
29	-EVDHIIYPPS-ISWDDDS-YTNKVLTSAKCN
30	-DIDHIIYPPS-RSMDDDS-YSNKVLV1LGEN
31	-EIEHIIYPPS-MSYDdns-QANKKILTEKAEN
32	-EVDHIIYPPS-LILDNT-1NNKVLV1YAEEN
33	-EIDHIIYPPS-KSADDSS-WFNKVLV1KSTN
34	-EMDHIIYPLS-RSLDNG-WHNRVLVHGKDN
35	-EIEHVIYPOS-LYFDDDS-PSNKV1CEAEVN
36	-DIEHIIYPOA-RLFDDDS-PSNK1LEARSVN

图6A

44	-EIEHTIVPKA- RVFDDDS- PSNKTLTFPHRIN	(SEQ ID NO: 173)
20	-DKDHTIIPQS- MKKDDSIINNLVLVNNQNN	(SEQ ID NO: 174)
45	-QVDHTILPWS- RFGDDDS- YLNKTLCTARSN	(SEQ ID NO: 163)
50	-DIDHVIPLA- RGGRDS- LDNMVLQQSDAN	(SEQ ID NO: 180)
46	-DMEHTIPKS- ISFDNS- DQNLTLCESYYN	(SEQ ID NO: 177)
47	-DTEHTIPRS- AGGDST- KMNLTLCSSRFN	(SEQ ID NO: 178)
48	-DTEHTIPRS- ISQDNS- QMNKTLCILKFN	(SEQ ID NO: 179)
39	-DIEHLFPIA- ESEDNG- RNNLVISHSACN	(SEQ ID NO: 181)
41	-DVDHTIFPRD- DTADNS- YGNKVVVAHRQCN	(SEQ ID NO: 182)
40	-DIEHTIVPQS- LGGLST- DYNTIVTLKSVN	(SEQ ID NO: 183)
35	-ELDHTIVPRT- DGGSNR- HENLAITCGACN	(SEQ ID NO: 184)
36	-EMDHTIVPRKGVGSTNT- RTNFAAVCAECN	(SEQ ID NO: 185)
37	-EMDHTIVPRKGVGSTNT- RVNLAACAAACN	(SEQ ID NO: 186)
38	-EMDHTIVPRAQGQGSTNT- RENLVAVCHRCN	(SEQ ID NO: 187)
34	-ELEHTIVPHS- FRQSNA- LSSLVLTWPGVN	(SEQ ID NO: 194)

图6B

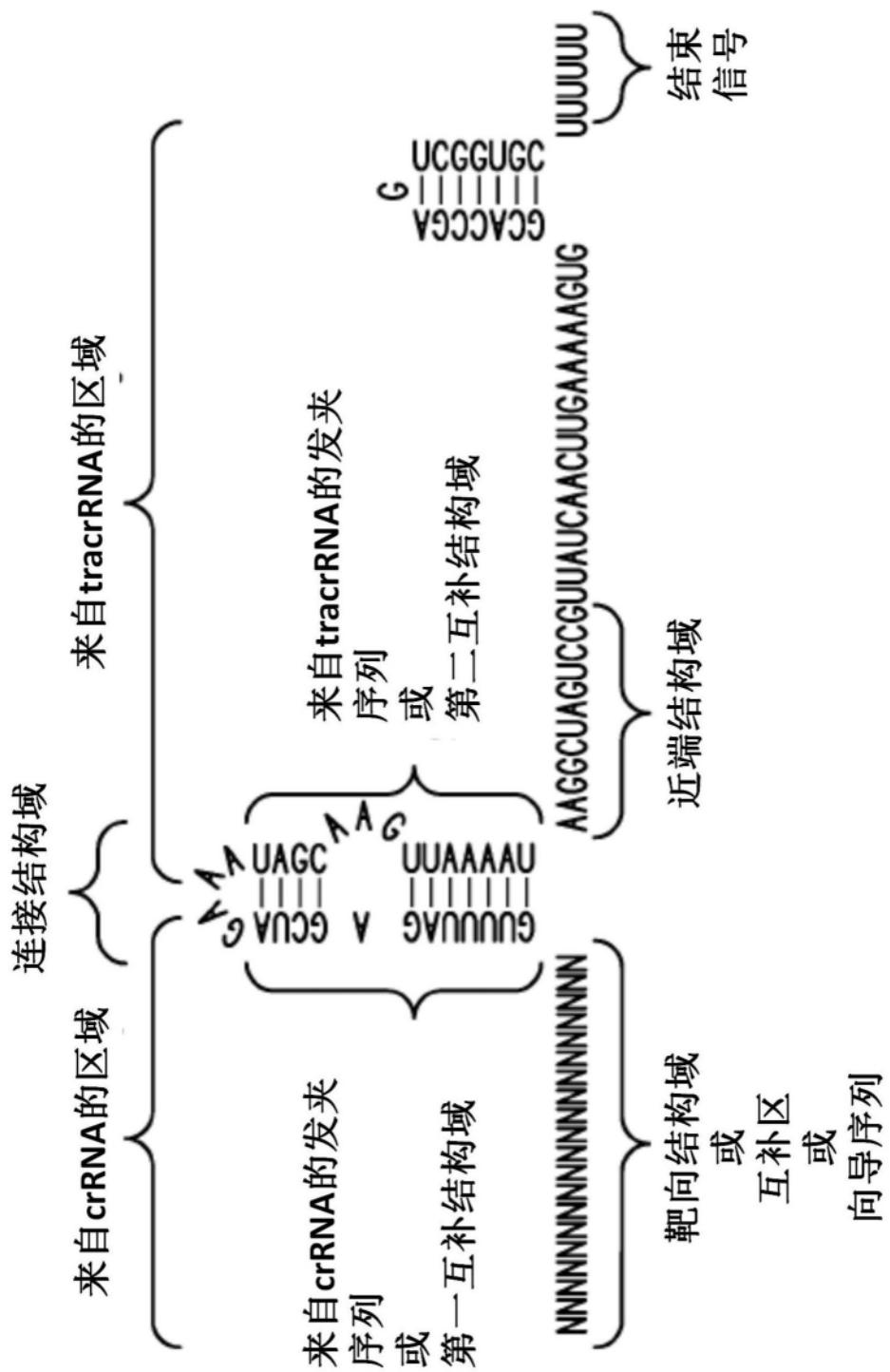


图7

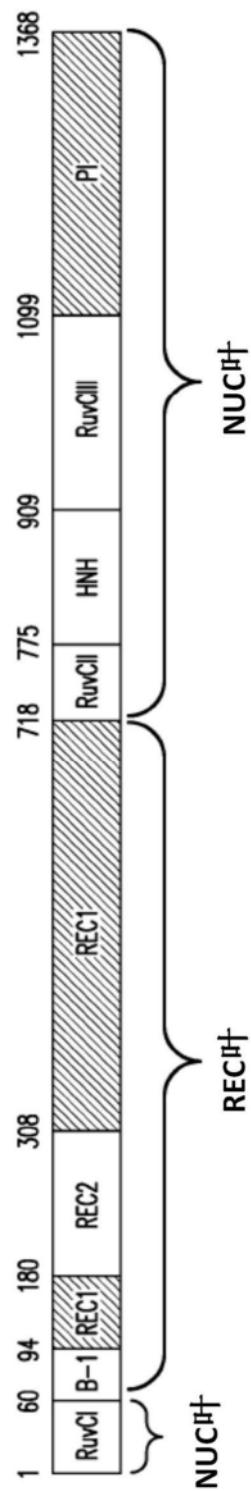


图8A

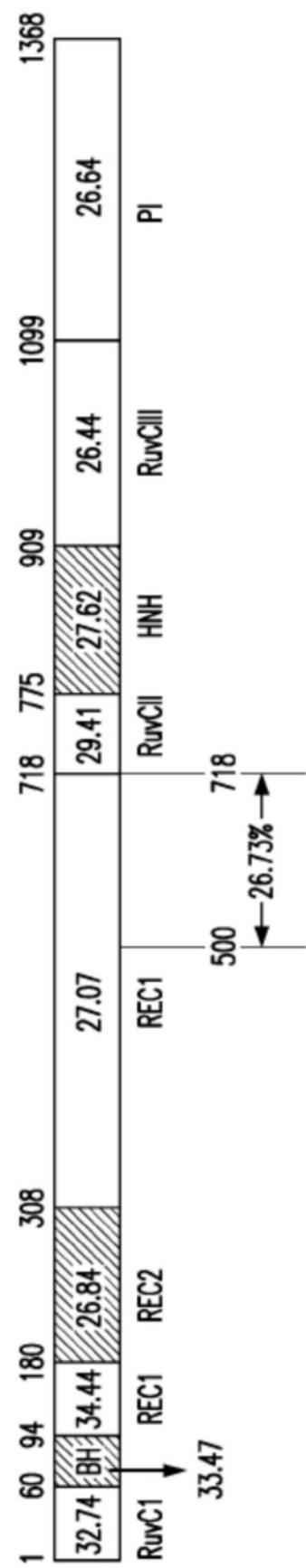
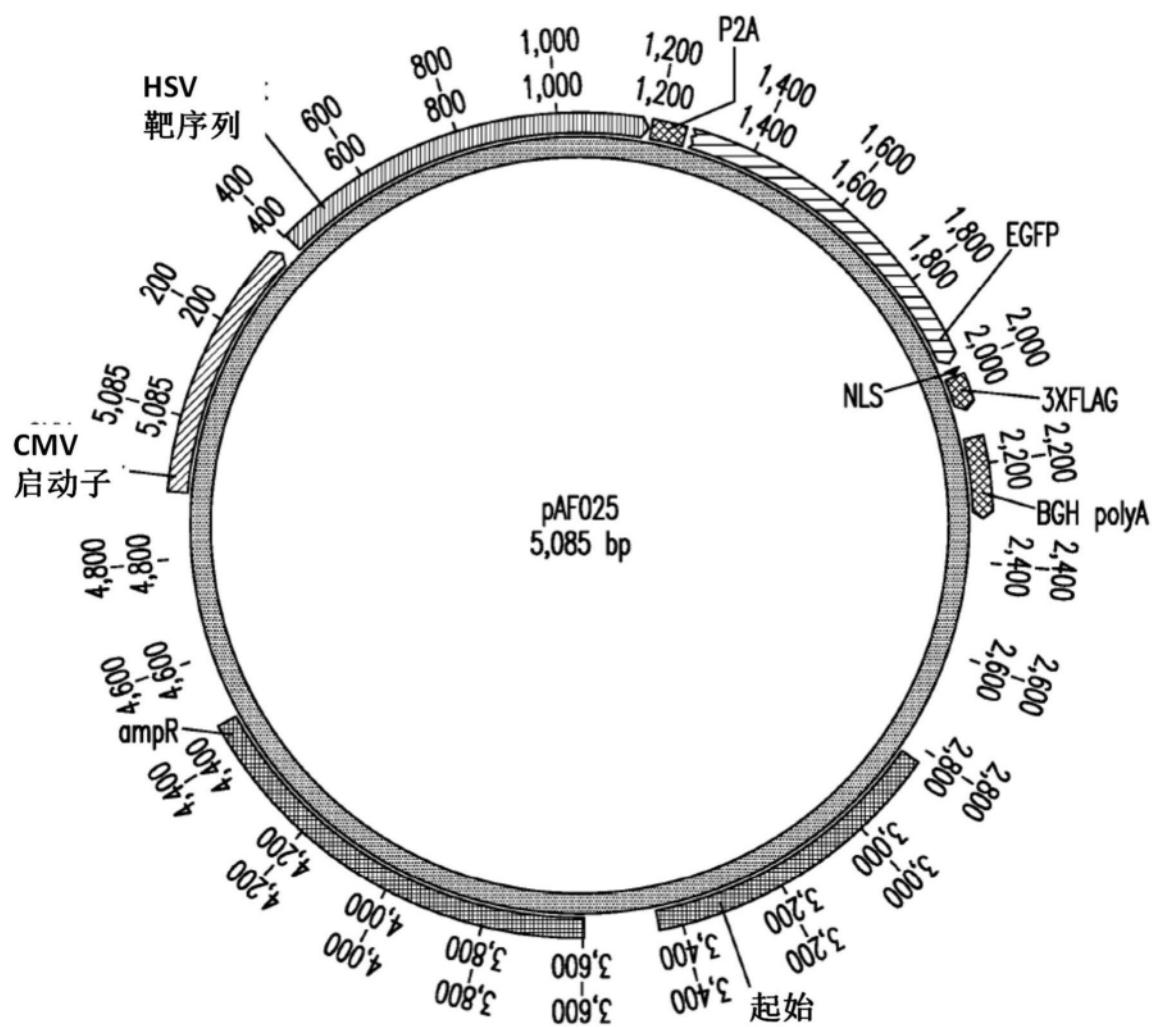


图8B



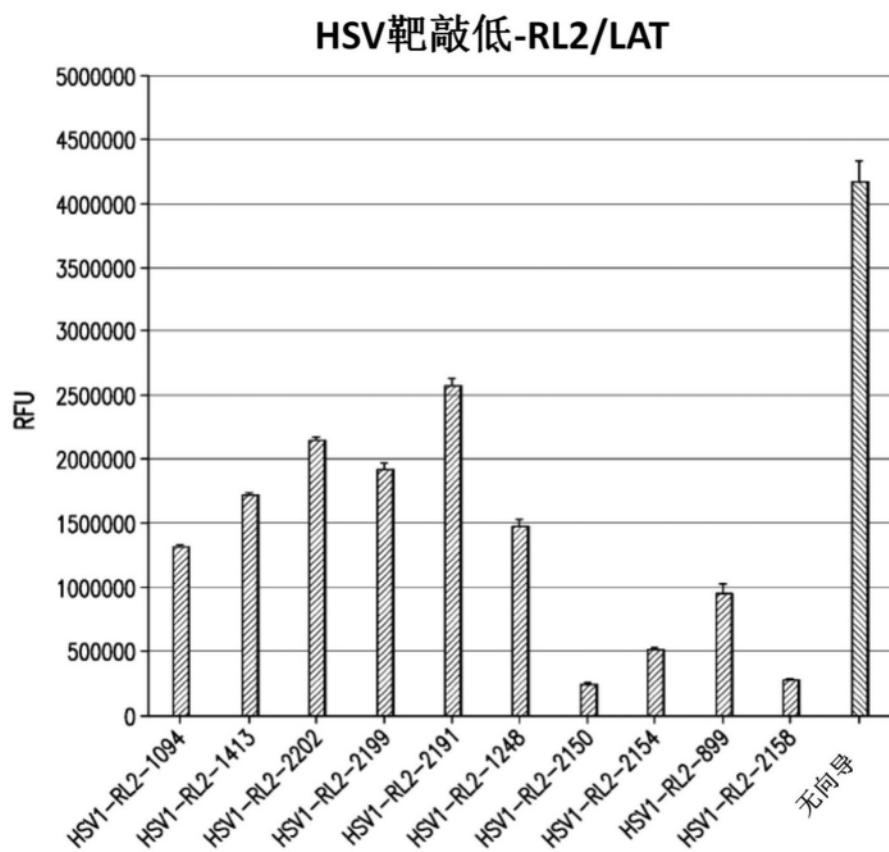


图10A

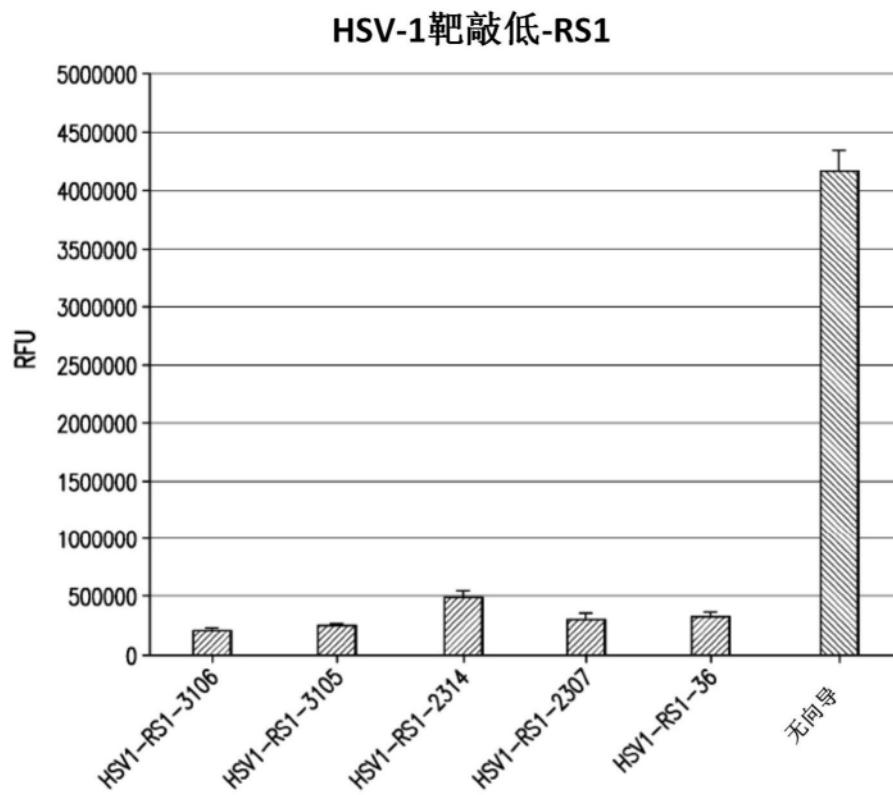


图10B