



(12) 发明专利申请

(10) 申请公布号 CN 116349651 A

(43) 申请公布日 2023. 06. 30

(21) 申请号 202310370809.2

C12N 5/10 (2006.01)

(22) 申请日 2019.03.19

C12N 15/864 (2006.01)

(30) 优先权数据

C12N 15/85 (2006.01)

62/644,961 2018.03.19 US

C12N 9/22 (2006.01)

C12N 15/113 (2010.01)

(62) 分案原申请数据

201980020470.5 2019.03.19

(71) 申请人 瑞泽恩制药公司

地址 美国纽约州塔里敦

(72) 发明人 查琳·亨特 苏珊娜·哈特福德

国春·龚 布莱恩·扎姆布罗维兹

(74) 专利代理机构 广州三环专利商标代理有限公司

公司 44202

专利代理师 郝传鑫

(51) Int. Cl.

A01K 67/027 (2006.01)

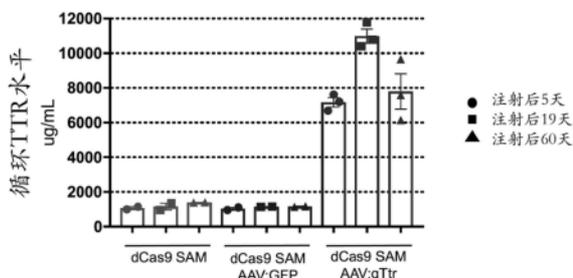
权利要求书8页 说明书72页
序列表(电子公布) 附图39页

(54) 发明名称

使用CRISPR/CAS系统对动物进行转录调制

(57) 摘要

提供了包含CRISPR/Cas协同激活介质系统组分的非人动物细胞和非人动物,以及制备和使用这种非人动物细胞和非人动物的方法。提供了使用此类非人动物增加体内靶基因表达并评估CRISPR/Cas协同激活介质系统增加体内靶基因表达能力的方法。



1. 一种非人动物,其包含第一基因组整合表达盒,其中所述第一表达盒包含:

(a) 编码嵌合成簇的规律间隔的短回文重复序列(CRISPR)相关(Cas)蛋白的核酸,所述蛋白包含与一个或多个转录激活域融合的核酸酶失活的Cas蛋白;以及

(b) 编码嵌合衔接体蛋白的核酸,所述嵌合衔接体蛋白包含与一个或多个转录激活域融合的衔接体。

2. 根据权利要求1所述的非人动物,还包含一个或多个指导RNA或编码所述一个或多个指导RNA的表达盒,每个所述指导RNA包含一个或多个衔接体结合元件,所述嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件,且

其中所述一个或多个指导RNA中的每一个能够与所述Cas蛋白形成复合物并将所述Cas蛋白引导至靶基因内的靶序列。

3. 根据权利要求1所述的非人动物,还包含编码一个或多个指导RNA的第二基因组整合表达盒,每个所述指导RNA包含一个或多个衔接体结合元件,所述嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件,且

其中所述一个或多个指导RNA中的每一个能够与所述Cas蛋白形成复合物并将所述Cas蛋白引导至靶基因内的靶序列。

4. 根据权利要求2或3所述的非人动物,其中所述靶序列包含所述靶基因内的调控序列。

5. 根据权利要求4所述的非人动物,其中所述调控序列包含启动子或增强子。

6. 根据权利要求2-5中任意一项所述的非人动物,其中所述靶序列在所述靶基因的转录起始位点的200个碱基对内。

7. 根据权利要求6所述的非人动物,其中所述靶序列在所述转录起始位点上游的200个碱基对和所述转录起始位点下游的1个碱基对的区域内。

8. 根据权利要求2-7中任意一项所述的非人动物,其中编码所述一个或多个指导RNA中的每一个的序列可操作地连接至不同的U6启动子。

9. 根据权利要求2-8中任意一项所述的非人动物,其中所述一个或多个指导RNA中的每一个包含两个衔接体结合元件,所述嵌合衔接体蛋白可以特异性地与所述两个衔接体结合元件结合。

10. 根据权利要求9所述的非人动物,其中第一衔接体结合元件在所述一个或多个指导RNA的每一个的第一环内,且第二衔接体结合元件在所述一个或多个指导RNA的每一个的第二环内。

11. 根据权利要求10所述的非人动物,其中所述一个或多个指导RNA中的每一个是单指导RNA,所述单指导RNA包含融合至反式激活CRISPR RNA(tracrRNA)部分的CRISPR RNA(crRNA)部分,且

其中所述第一环是对应于SEQ ID NO:12的残基13-16的四环,且所述第二环是对应于SEQ ID NO:12的残基53-56的茎环2。

12. 根据权利要求2-11中任意一项所述的非人动物,其中所述衔接体结合元件包含SEQ ID NO:16所示的序列。

13. 根据权利要求12所述的非人动物,其中所述一个或多个指导RNA中的每一个包含SEQ ID NO:40或63所示的序列。

14. 根据权利要求2-13中任意一项所述的非人动物,其中所述一个或多个指导RNA中的至少一个靶向Ttr基因,可选地,其中Ttr靶向的指导RNA靶向包含SEQ ID NO:34-36中任一所示的序列的序列,或可选地,其中所述Ttr靶向的指导RNA包含SEQ ID NO:37-39中任一所示的序列。

15. 根据权利要求2-13中任意一项所述的非人动物,其中所述一个或多个指导RNA中的至少一个靶向Pcsk9基因,可选地,其中Pcsk9靶向的指导RNA靶向包含SEQ ID NO:89-91中任一所示的序列的序列,或可选地,其中所述Pcsk9靶向的指导RNA包含SEQ ID NO:92-94中任一所示的序列。

16. 根据权利要求2-13中任意一项所述的非人动物,其中所述一个或多个指导RNA中的至少一个靶向Ldlr基因,可选地,其中Ldlr靶向的指导RNA靶向包含SEQ ID NO:75-77中任一所示的序列的序列,或可选地,其中所述Ldlr靶向的指导RNA包含SEQ ID NO:78-80中任一所示的序列。

17. 根据权利要求2-16中任意一项所述的非人动物,其中所述一个或多个指导RNA靶向两个或多个靶基因。

18. 根据权利要求2-17中任意一项所述的非人动物,其中所述一个或多个指导RNA包含靶向单个靶基因的多个指导RNA。

19. 根据权利要求2-18中任意一项所述的非人动物,其中所述一个或多个指导RNA包含靶向单个靶基因的至少三个指导RNA。

20. 根据权利要求19所述的非人动物,其中所述至少三个指导RNA靶向小鼠Ttr基因座,其中第一指导RNA靶向包含SEQ ID NO:34的序列或包含SEQ ID NO:37所示的序列,第二指导RNA靶向包含SEQ ID NO:35的序列或包含SEQ ID NO:38所示的序列,且第三指导RNA靶向包含SEQ ID NO:36的序列或包含SEQ ID NO:39所示的序列。

21. 根据权利要求19所述的非人动物,其中所述至少三个指导RNA靶向小鼠Pcsk9基因座,其中第一指导RNA靶向包含SEQ ID NO:89的序列或包含SEQ ID NO:92所示的序列,第二指导RNA靶向包含SEQ ID NO:90的序列或包含SEQ ID NO:93所示的序列,且第三指导RNA靶向包含SEQ ID NO:91的序列或包含SEQ ID NO:94所示的序列。

22. 根据权利要求19所述的非人动物,其中所述至少三个指导RNA靶向小鼠Ldlr基因座,且其中第一指导RNA靶向包含SEQ ID NO:75的序列或包含SEQ ID NO:78所示的序列,第二指导RNA靶向包含SEQ ID NO:76的序列或包含SEQ ID NO:79所示的序列,且第三指导RNA靶向包含SEQ ID NO:77的序列或包含SEQ ID NO:80中所示的序列。

23. 根据前述权利要求中任意一项所述的非人动物,其中所述Cas蛋白是Cas9蛋白。

24. 根据权利要求23所述的非人动物,其中所述Cas9蛋白是化脓链球菌Cas9蛋白。

25. 根据权利要求23或24所述的非人动物,其中当与化脓链球菌Cas9蛋白最佳比对时,所述Cas9蛋白包含对应于D10A和N863A的突变。

26. 根据前述权利要求中任意一项所述的非人动物,其中对编码所述Cas蛋白的序列进行密码子优化以在所述非人动物中表达。

27. 根据前述权利要求中任意一项所述的非人动物,其中嵌合Cas蛋白中的一个或多个转录激活因子结构域选自:VP16、VP64、p65、MyoD1、HSF1、RTA、SET7/9及其组合。

28. 根据权利要求27所述的非人动物,其中所述嵌合Cas蛋白中的一个或多个转录激活

因子结构域包含VP64。

29. 根据权利要求28所述的非人动物,其中所述嵌合Cas蛋白从N末端到C末端包括:无催化活性的Cas蛋白质、核定位信号,和VP64转录激活因子结构域。

30. 根据权利要求29所述的非人动物,其中所述嵌合Cas蛋白包含与SEQ ID NO:1所示的序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。

31. 根据权利要求30所述的非人动物,其中编码所述嵌合Cas蛋白的第一表达盒的片段包含与SEQ ID NO:25所示的序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。

32. 根据前述权利要求中任意一项所述的非人动物,其中所述第一表达盒进一步包含在编码所述嵌合Cas蛋白的片段的上游的多腺苷酸化信号或转录终止子,

其中所述多腺苷酸化信号或所述转录终止子被重组酶识别位点侧接,且

其中所述多腺苷酸化信号或所述转录终止子已以组织特异性方式切除。

33. 根据权利要求32所述的非人动物,其中已经在肝脏中切除了所述多腺苷酸化信号或所述转录终止子。

34. 根据权利要求32或33所述的非人动物,其中所述重组酶是Cre重组酶。

35. 根据权利要求32-34中任意一项所述的非人动物,其中所述非人动物还包含基因组整合的重组酶表达盒,所述基因组整合的重组酶表达盒包含可操作地连接至组织特异性启动子的重组酶编码序列。

36. 根据权利要求32-35中任意一项所述的非人动物,其中重组酶基因可操作地连接于表2所示的启动子之一。

37. 根据前述权利要求中任意一项所述的非人动物,其中衔接体位于所述嵌合衔接体蛋白的N末端,且一个或多个转录激活域位于所述嵌合衔接体蛋白的C末端。

38. 根据前述权利要求中任意一项所述的非人动物,其中衔接体包含MS2外壳蛋白或其功能片段或变体。

39. 根据前述权利要求中任意一项所述的非人动物,其中所述嵌合衔接体蛋白中的一个或多个转录激活域选自:VP16、VP64、p65、MyoD1、HSF1、RTA、SET7/9及其组合。

40. 根据权利要求39所述的非人动物,其中所述嵌合衔接体蛋白中的一个或多个转录激活域包含p65和HSF1。

41. 根据权利要求40所述的非人动物,其中所述嵌合衔接体蛋白从N末端至C末端包含:MS2外壳蛋白、核定位信号、p65转录激活域,和HSF1转录激活域。

42. 根据权利要求41所述的非人动物,其中所述嵌合衔接体蛋白包含与SEQ ID NO:6所示的序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。

43. 根据权利要求42所述的非人动物,其中编码所述嵌合衔接体蛋白的第一表达盒的片段包含与SEQ ID NO:27所示的序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。

44. 根据前述权利要求中任意一项所述的非人动物,其中所述第一表达盒是多顺反子。

45. 根据权利要求44所述的非人动物,其中通过内部核糖体进入位点(IRES)将编码所述嵌合Cas蛋白的所述第一表达盒的片段与编码所述嵌合衔接体蛋白的所述第一表达盒的片段分开。

46. 根据权利要求44所述的非人动物,其中通过编码2A肽的核酸将编码所述嵌合Cas蛋白的所述第一表达盒的片段与编码所述嵌合衔接体蛋白的所述第一表达盒的片段分开。

47. 根据权利要求46所述的非人动物,其中所述2A肽是T2A肽。

48. 根据前述权利要求中任意一项所述的非人动物,其中所述第一表达盒被整合到安全港基因座中。

49. 根据权利要求2-48中任意一项所述的非人动物,其中所述第一表达盒和/或所述第二表达盒被整合到安全港基因座中。

50. 根据权利要求49所述的非人动物,其中所述非人动物对于所述第一表达盒是杂合的,且对于所述第二表达盒是杂合的,且

所述第一表达盒基因组整合于所述安全港基因座的第一等位基因中,且所述第二表达盒基因组整合于所述安全港基因座的第二等位基因中。

51. 根据权利要求48-50中任意一项所述的非人动物,其中所述安全港基因座是Rosa26基因座。

52. 根据权利要求48-51中任意一项所述的非人动物,其中所述第一表达盒与所述安全港基因座中的内源启动子可操作地连接。

53. 根据前述权利要求中任意一项所述的非人动物,其中所述非人动物是哺乳动物。

54. 根据权利要求53所述的非人动物,其中所述哺乳动物是啮齿动物。

55. 根据权利要求54所述的非人动物,其中所述啮齿动物是大鼠或小鼠。

56. 根据权利要求55所述的非人动物,其中所述啮齿动物是小鼠。

57. 包含基因组整合的表达盒的非人动物细胞,其中所述表达盒包含:

(a) 编码嵌合成簇的规律间隔的短回文重复序列(CRISPR)相关(Cas)蛋白的核酸,所述蛋白包含与一个或多个转录激活域融合的核酸酶失活的Cas蛋白;以及

(b) 编码嵌合衔接体蛋白的核酸,所述嵌合衔接体蛋白包含与一个或多个转录激活域融合的衔接体。

58. 包含基因组整合的表达盒的非人动物基因组,其中所述表达盒包含:

(a) 编码嵌合成簇的规律间隔的短回文重复序列(CRISPR)相关(Cas)蛋白的核酸,所述蛋白包含与一个或多个转录激活域融合的核酸酶失活的Cas蛋白;以及

(b) 编码嵌合衔接体蛋白的核酸,所述嵌合衔接体蛋白包含与一个或多个转录激活域融合的衔接体。

59. 靶向载体,其包含被在靶基因组基因座处靶向5'靶序列的5'同源臂和在所述靶基因组基因座处靶向3'靶序列的3'同源臂侧接的插入核酸,其中所述插入核酸包含表达盒,其中所述表达盒包含:

(a) 编码嵌合成簇的规律间隔的短回文重复序列(CRISPR)相关(Cas)蛋白的核酸,所述蛋白包含与一个或多个转录激活域融合的核酸酶失活的Cas蛋白;以及

(b) 编码嵌合衔接体蛋白的核酸,所述嵌合衔接体蛋白包含与一个或多个转录激活域融合的衔接体。

60. 制造根据权利要求1-56中任意一项所述的非人动物的方法,包括:

(a) 向非人动物胚胎干(ES)细胞引入:

(i) 靶向靶基因组基因座中的靶序列的核酸酶剂;和

(ii) 靶向载体,其包含核酸插入物,所述核酸插入物包含第一表达盒,所述第一表达盒被在靶基因组基因座中对应5'靶序列的5'同源臂和在靶基因组基因座中对应3'靶序列的3'同源臂侧接,

其中所述靶向载体与所述靶基因组基因座重组以产生遗传修饰的非人ES细胞,所述遗传修饰的非人ES细胞在其基因组中包含所述靶基因组基因座处的所述第一表达盒;

(b) 将所述遗传修饰的非人ES细胞引入非人动物宿主胚胎中;以及

(c) 在代孕母体中孕育所述非人动物宿主胚胎,其中所述代孕母体产生F0代遗传修饰的非人动物,所述非人动物的基因组中包含所述靶基因组基因座处的所述第一表达盒。

61. 根据权利要求60所述的方法,其中所述靶向载体为长度至少为10kb的大靶向载体,或其中所述5'同源臂和所述3'同源臂的长度总和至少为10kb。

62. 制造根据权利要求1-56中任意一项所述的非人动物的方法,包括:

(a) 向非人动物单细胞期胚胎引入:

(i) 靶向靶基因组基因座中的靶序列的核酸酶剂;和

(ii) 靶向载体,其包含核酸插入物,所述核酸插入物包含第一表达盒,所述第一表达盒被在所述靶基因组基因座中对应5'靶序列的5'同源臂和在所述靶基因组基因座中对应3'靶序列的3'同源臂侧接,

其中所述靶向载体与所述靶基因组基因座重组以产生遗传修饰的非人ES细胞,所述遗传修饰的非人ES细胞在其基因组中包含所述靶基因组基因座处的所述第一表达盒;

(b) 在代孕母体中孕育遗传修饰的非人动物单细胞期胚胎,以产生遗传修饰的F0代非人动物,所述遗传修饰的F0代非人动物在其基因组中包含在所述靶基因组基因座处的所述第一表达盒。

63. 根据权利要求60-62中任意一项所述的方法,其中所述核酸酶剂包括Cas蛋白和指导RNA。

64. 根据权利要求63所述的方法,其中所述Cas蛋白是Cas9蛋白。

65. 根据权利要求63或64所述的方法,其中步骤(a)进一步包括引入在所述靶基因组基因座内靶向第二靶序列的第二指导RNA。

66. 根据权利要求60-65中任意一项所述的方法,其中所述非人动物是小鼠或大鼠。

67. 根据权利要求66所述的方法,其中所述非人动物是小鼠。

68. 一种增加靶基因在权利要求1-56中任意一项所述的非人动物体内的表达的方法,包括将一个或多个指导RNA引入所述非人动物,每个所述指导RNA包含一个或多个衔接体结合元件,所述嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件,

其中所述一个或多个指导RNA与所述嵌合Cas蛋白和嵌合衔接体蛋白形成复合物并将它们引导至所述靶基因内的靶序列,从而增加所述靶基因的表达。

69. 根据权利要求68所述的方法,其中通过腺相关病毒(AAV)介导的递送引入所述一个或多个指导RNA。

70. 根据权利要求69所述的方法,其中所述AAV是AAV8。

71. 根据权利要求68所述的方法,其中通过脂质纳米颗粒介导的递送或流体动力递送来引入所述一个或多个指导RNA。

72. 根据权利要求68-71中任意一项所述的方法,其中所述靶基因是在肝脏中表达的基

因。

73. 根据权利要求68-72中任意一项所述的方法, 其中将所述一个或多个指导RNA给予所述非人动物的途径是静脉内注射、脑内注射、腹膜内注射、滴鼻法, 或玻璃体内注射。

74. 根据权利要求68-72中任意一项所述的方法, 其中所述靶序列包含所述靶基因内的调控序列。

75. 根据权利要求74所述的方法, 其中所述调控序列包含启动子或增强子。

76. 根据权利要求68-75中任意一项所述的方法, 其中所述靶序列在所述靶基因的所述转录起始位点的200个碱基对内。

77. 根据权利要求76所述的方法, 其中所述靶序列在所述转录起始位点上游的200个碱基对和所述转录起始位点下游的1个碱基对的区域内。

78. 根据权利要求68-77中任意一项所述的方法, 其中所述一个或多个指导RNA以RNA的形式被引入。

79. 根据权利要求68-77中任意一项所述的方法, 其中所述一个或多个指导RNA以DNA的形式被引入。

80. 根据权利要求79所述的方法, 其中所述一个或多个指导RNA中的每一个可操作地连接至不同U6启动子。

81. 根据权利要求68-80中任意一项所述的方法, 其中所述一个或多个指导RNA中的每一个均包含两个衔接体结合元件, 所述嵌合衔接体蛋白可以特异性地与所述两个衔接体结合元件结合。

82. 根据权利要求81所述的方法, 其中第一衔接体结合元件在所述一个或多个指导RNA的每一个的第一环内, 且第二衔接体结合元件在所述一个或多个指导RNA的每一个的第二环内。

83. 根据权利要求82所述的方法, 其中所述一个或多个指导RNA中的每一个是单指导RNA, 所述单指导RNA包含融合至反式激活CRISPR RNA (tracrRNA) 部分的CRISPR RNA (crRNA) 部分, 且

其中所述第一环是对应于SEQ ID NO:12的残基13-16的四环, 所述第二环是对应于SEQ ID NO:12的残基53-56的茎环2。

84. 根据权利要求68-83中任意一项所述的方法, 其中所述衔接体结合元件包含SEQ ID NO:16所示的序列。

85. 根据权利要求84所述的方法, 其中所述一个或多个指导RNA中的每一个包含SEQ ID NO:40或63所示的序列。

86. 根据权利要求68-85中任意一项所述的方法, 其中所述一个或多个指导RNA中的至少一个靶向疾病相关基因。

87. 根据权利要求86所述的方法, 其中所述疾病相关基因是Ttr基因, 可选地, 其中Ttr靶向的指导RNA靶向包含SEQ ID NO:34-36中任一个所示的序列的序列, 或可选地, 其中所述Ttr靶向的指导RNA包含SEQ ID NO:37-39中任意一项所示的序列。

88. 根据权利要求68-85中任意一项所述的方法, 其中所述一个或多个指导RNA中的至少一个靶向Pcsk9基因, 可选地, 其中Pcsk9靶向的指导RNA靶向包含SEQ ID NO:89-91中任一个所示的序列的序列, 或可选地, 其中所述Pcsk9靶向的指导RNA包含SEQ ID NO:92-94中

任一个所示的序列。

89. 根据权利要求88所述的方法,其中所述方法在所述非人动物中引起高胆固醇血症。

90. 根据权利要求68-85中任意一项所述的方法,其中所述一个或多个指导RNA中的至少一个靶向Ldlr基因,可选地,其中Ldlr靶向的指导RNA靶向包含SEQ ID NO:75-77中任一一个所示的序列的序列,或可选地,其中所述Ldlr靶向的指导RNA包含SEQ ID NO:78-80中任一一个所示的序列。

91. 根据权利要求68-90中任意一项所述的方法,其中所述一个或多个指导RNA靶向两个或多个靶基因。

92. 根据权利要求68-91中任意一项所述的方法,其中所述一个或多个指导RNA包含靶向单个靶基因的多个指导RNA。

93. 根据权利要求68-92中任意一项所述的方法,其中所述一个或多个指导RNA包含靶向单个靶基因的至少三个指导RNA。

94. 根据权利要求93所述的方法,其中所述至少三个指导RNA靶向所述小鼠Ttr基因座,其中第一指导RNA靶向包含SEQ ID NO:34的序列或包含SEQ ID NO:37所示的序列,第二指导RNA靶向包含SEQ ID NO:35的序列或包含SEQ ID NO:38所示的序列,且第三指导RNA靶向包含SEQ ID NO:36的序列或包含SEQ ID NO:39所示的序列。

95. 根据权利要求93所述的方法,其中所述至少三个指导RNA靶向所述小鼠Pcsk9基因座,且其中第一指导RNA靶向包含SEQ ID NO:89的序列或包含SEQ ID NO:92所示的序列,第二指导RNA靶向包含SEQ ID NO:90的序列或包含SEQ ID NO:93所示的序列,且第三指导RNA靶向包含SEQ ID NO:91的序列或包含SEQ ID NO:94所示的序列。

96. 根据权利要求93所述的方法,其中所述至少三个指导RNA靶向所述小鼠Ldlr基因座,其中第一指导RNA靶向包含SEQ ID NO:75的序列或包含SEQ ID NO:78所示的序列,第二指导RNA靶向包含SEQ ID NO:76的序列或包含SEQ ID NO:79所述的序列,且第三指导RNA靶向包含SEQ ID NO:77的序列或包含SEQ ID NO:80中所述的序列。

97. 根据权利要求68-96中任意一项所述的方法,其中所述靶基因的表达相对于对照非人动物增加至少0.5倍、1倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍,或20倍。

98. 根据权利要求68-97中任意一项所述的方法,其中所述第一表达盒在编码所述嵌合Cas蛋白的片段的上游还包含多腺苷酸化信号或转录终止子,

其中所述多腺苷酸化信号或所述转录终止子被重组酶识别位点侧接,所述重组酶识别位点被位点特异性重组酶识别,且

其中所述方法进一步包括将所述重组酶引入所述非人动物中。

99. 根据权利要求98所述的方法,其中所述重组酶是Cre重组酶。

100. 根据权利要求98或99所述的方法,其中所述重组酶通过腺相关病毒(AAV)介导的递送而引入。

101. 根据权利要求100所述的方法,其中所述AAV是AAV8。

102. 根据权利要求98或99所述的方法,其中所述重组酶通过脂质纳米颗粒介导的递送或流体动力递送引入。

103. 根据权利要求98-102中任意一项所述的方法,其中以组织特异性的方式引入或表达所述重组酶。

104. 根据权利要求98-102中任意一项所述的方法,其中以蛋白形式引入所述重组酶。

105. 根据权利要求98-102中任意一项所述的方法,其中以DNA或RNA的形式引入所述重组酶。

106. 根据权利要求105所述的方法,其中以与表2中列出的启动子之一可操作地连接的DNA的形式引入所述重组酶。

107. 根据权利要求98-106中任意一项所述的方法,其中向所述非人动物施用所述重组酶的途径是静脉内注射、脑内注射、腹膜内注射、滴鼻法,或玻璃体内注射。

108. 根据权利要求68-107中任意一项所述的方法,其中所述一个或多个指导RNA通过腺相关病毒(AAV)介导的递送被引入,其中所述一个或多个指导RNA中的每一个可操作地连接至不同的U6启动子,并且其中所述一个或多个指导RNA包括靶向单个靶基因的多个指导RNA。

109. 一种在非人动物体内模拟高胆固醇血症的方法,包括将一个或多个靶向Pcsk9的指导RNA引入权利要求1-56中任意一项所述的非人动物,其中所述一个或多个指导RNA中的每一个包含一个或多个嵌合体结合元件,所述嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件,

其中所述一个或多个指导RNA与所述嵌合Cas蛋白和嵌合衔接体蛋白形成复合物,并将它们引导至Pcsk9内的靶序列,从而增加Pcsk9的表达并引起高胆固醇血症。

使用CRISPR/CAS系统对动物进行转录调制

分案申请说明

[0001] 本申请是申请日为2019年03月19日,申请号为201980020470.5,发明名称为“使用CRISPR/CAS系统对动物进行转录调制”的发明专利申请的分案申请。

相关申请的交叉引用

[0002] 本申请要求于2018年3月19日提交的美国申请号62/644,961的权益,出于所有目的将通过引用整体并入本文。

对通过EFS Web提交为XML文件的序列表的引用

[0003] 写在文件057766-549767.xml中的序列表为234KB,创建于2023年3月30日(实际序列内容创建于2019年03月19日),通过引用并入。

背景技术

[0004] 在许多生物过程(如发育和疾病)中,基因表达受到严格控制。转录因子通过与靶基因的增强子和启动子区域的特定DNA序列结合来调节基因表达,并通过其效应子域调制(modulate)转录。基于相同的原理,通过将各种功能域融合到经工程化以与目标基因结合的DNA结合域中以产生人工转录因子(ATF),从而调节该目标基因的表达。然而,这些ATF的结合特异性通常是简并(degenerate)的,可能难以预测,并且复杂及费时的设计和产生限制了其应用。

[0005] CRISPR/Cas技术是一种有前途的新治疗方式,其不仅可以用于进行靶向的基因组修饰,而且可以用于调节靶基因的转录。然而,需要一种更好的方法来评估体内引入的CRISPR/Cas试剂的效率。在体内测试该系统的局限性之一是需要同时将所有组分引入活生物体。引入这些组分的典型方法是将DNA构建体瞬时转染到将生成适当的RNA和蛋白质的细胞中。尽管有效,但是这种方法具有固有的缺点,因为细胞必须依靠质粒DNA构建体先进行转录,然后翻译,其后Cas蛋白才可用于与sgRNA组分相互作用。需要更好的方法和工具来更有效地评估引入的CRISPR/Cas试剂的活性,并评估靶向体内特定组织或细胞类型的不同递送方法和参数。

[0006] 另外,诸如CRISPR/Cas试剂之类的生物活性剂向对象的递送通常由于组分到达靶细胞或组织的困难而受到阻碍。这些限制可能导致,例如,需要使用比达到结果所需的浓度高得多的试剂浓度,这增加了毒性作用和副作用的风险。需要改进的体内递送方法和评估此类递送方法的方法。

发明内容

[0007] 提供了包含CRISPR/Cas协同激活介质(SAM)系统的非人动物,以及使用此类非人动物(例如SAM就绪(SAM-ready)的非人动物)评估CRISPR/Cas SAM试剂在体内激活靶基因转录的能力或评估在体内激活靶基因的转录或增加靶基因的表达的影响。还提供了包含CRISPR/Cas协同激活介质(SAM)系统的非人动物基因组或细胞。

[0008] 在一个方面,提供了非人动物基因组、非人动物细胞或非人动物,其包含一个或多个

个基因组整合的协同激活介质表达盒。此类非人动物基因组、非人动物细胞或非人动物可包含,例如,第一基因组整合表达盒,其中所述第一表达盒包含:(a) 编码嵌合成簇的规律间隔的短回文重复序列(CRISPR)相关(Cas)蛋白的核酸,其包含与一个或多个转录激活域融合的核酸酶失活(nuclease-inactive)的Cas蛋白;以及(b) 编码嵌合衔接体蛋白的核酸,所述嵌合衔接体蛋白包含与一个或多个转录激活域融合的衔接体。

[0009] 此类非人动物基因组、非人动物细胞,或非人动物可进一步包含一个或多个指导RNA或编码所述一个或多个指导RNA的表达盒,每个指导RNA包含一个或多个衔接体结合元件,所述嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件,其中所述一个或多个指导RNA中的每一个能够与所述Cas蛋白形成复合物并将其引导至靶基因内的靶序列。可选地,编码所述一个或多个指导RNA的表达盒在腺相关病毒(AAV),例如AAV8中。可选地,编码所述一个或多个指导RNA的表达盒在AAV中,所述一个或多个指导RNA中的每一个可操作地连接至不同的U6启动子,并且所述一个或多个指导RNA包含靶向单个基因的多个指导RNA。

[0010] 此类非人动物基因组、非人动物细胞,或非人动物可进一步包含编码一个或多个指导RNA的第二基因组整合表达盒,每个指导RNA包含一个或衔接体结合元件,所述嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件,其中所述一个或多个指导RNA中的每一个能够与所述Cas蛋白形成复合物并将其引导至靶基因内的靶序列。

[0011] 在一些非人动物基因组、非人动物细胞,或非人动物中,靶序列包含靶基因内的调控序列。可选地,所述调控序列包含启动子或增强子。在一些非人动物基因组、非人动物细胞,或非人动物中,所述靶序列在靶基因的转录起始位点的200个碱基对内。可选地,所述靶序列在转录起始位点上游的200个碱基对和转录起始位点下游的1个碱基对的区域内。

[0012] 在一些非人动物基因组、非人动物细胞,或非人动物中,编码所述一个或多个指导RNA中的每一个的序列可操作地连接至不同的启动子,例如U6启动子。在一些非人动物基因组、非人动物细胞,或非人动物中,所述一个或多个指导RNA中的每一个均包含两个衔接体结合元件,所述嵌合衔接体蛋白可以特异性地与所述两个衔接体结合元件结合。可选地,第一衔接体结合元件在所述一个或多个指导RNA的每一个的第一环内,且第二衔接体结合元件在所述一个或多个指导RNA的每一个的第二环内。可选地,所述一个或多个指导RNA中的每一个是单指导RNA,所述单指导RNA包含融合至反式激活CRISPR RNA(tracrRNA)部分的CRISPR RNA(crRNA)部分,其中所述第一环是对应于SEQ ID NO:12的残基13-16的四环(tetraloop),所述第二环是对应于SEQ ID NO:12的残基53-56的茎环2。在一些非人动物基因组、非人动物细胞,或非人动物中,衔接体结合元件包含SEQ ID NO:16所示的序列。可选地,所述一个或多个指导RNA中的每一个包含SEQ ID NO:40或63所示的序列。

[0013] 在一些非人动物基因组、非人动物细胞,或非人动物中,所述一个或多个指导RNA中的至少一个靶向疾病相关基因。可选地,所述一个或多个指导RNA中的至少一个靶向Ttr基因,可选地,其中Ttr靶向的指导RNA靶向包含SEQ ID NO:34-36中任一个所示的序列的序列,或可选地,其中所述Ttr靶向的指导RNA包含SEQ ID NO:37-39中任一个所示的序列。可选地,所述一个或多个指导RNA中的至少一个靶向Pcsk9基因,可选地,其中Pcsk9靶向的指导RNA靶向包含SEQ ID NO:89-91中任一个所示的序列的序列,或可选地,其中所述Pcsk9靶向的指导RNA包含SEQ ID NO:92-94中任一个所示的序列。可选地,所述一个或多个指导RNA中的至少一个靶向Ldlr基因,可选地,其中Ldlr靶向的指导RNA靶向包含SEQ ID NO:75-77

中任一个所示的序列的序列,或可选地,其中所述Ldlr靶向的指导RNA包含SEQ ID NO:78-80中任一个所示的序列。

[0014] 在一些非人动物基因组、非人动物细胞,或非人动物中,所述一个或多个指导RNA靶向两个或多个靶基因。在一些非人动物基因组、非人动物细胞,或非人动物中,所述一个或多个指导RNA包含靶向单个靶基因的多个指导RNA。在一些非人动物基因组、非人动物细胞,或非人动物中,所述一个或多个指导RNA包含靶向单个靶基因的至少三个指导RNA。可选地,所述至少三个指导RNA靶向小鼠Ttr基因座,其中第一指导RNA靶向包含SEQ ID NO:34的序列或包含SEQ ID NO:37所示的序列,第二指导RNA靶向包含SEQ ID NO:35的序列或包含SEQ ID NO:38所示的序列,且第三指导RNA靶向包含SEQ ID NO:36的序列或包含SEQ ID NO:39所示的序列。可选地,所述至少三个指导RNA靶向小鼠Pcsk9基因座,其中第一指导RNA靶向包含SEQ ID NO:89的序列或包含SEQ ID NO:92所示的序列,第二指导RNA靶向包含SEQ ID NO:90的序列或包含SEQ ID NO:93所示的序列,且第三指导RNA靶向包含SEQ ID NO:91的序列或包含SEQ ID NO:94所示的序列。可选地,所述至少三个指导RNA靶向小鼠Ldlr基因座,其中第一指导RNA靶向包含SEQ ID NO:75的序列或包含SEQ ID NO:78所示的序列,第二指导RNA靶向包含SEQ ID NO:76的序列或包含SEQ ID NO:79所示的序列,且第三指导RNA靶向包含SEQ ID NO:77的序列或包含SEQ ID NO:80中所示的序列。

[0015] 在一些非人动物基因组、非人动物细胞,或非人动物中,Cas蛋白是Cas9蛋白。可选地,Cas9蛋白是化脓链球菌(*Streptococcus pyogenes*)Cas9蛋白。任选地,当与化脓链球菌(*Streptococcus pyogenes*)Cas9蛋白最佳比对时,Cas9蛋白包含对应于D10A和N863A的突变。可选地,对编码Cas蛋白的序列进行密码子优化以在非人动物中表达。

[0016] 在一些非人动物基因组、非人动物细胞,或非人动物中,嵌合Cas蛋白中的一个或多个转录激活因子结构域(transcriptional activator domain)选自:VP16、VP64、p65、MyoD1、HSF1、RTA、SET7/9及其组合。可选地,嵌合Cas蛋白中的一个或多个转录激活因子结构域包含VP64。

[0017] 在一些非人动物基因组、非人动物细胞,或非人动物中,嵌合Cas蛋白包括从N末端到C末端的:无催化活性的Cas蛋白质;和VP64转录激活因子结构域。在一些非人动物基因组、非人动物细胞,或非人动物中,嵌合Cas蛋白包括从N末端到C末端的:无催化活性的Cas蛋白质、核定位信号,和VP64转录激活因子结构域。可选地,嵌合Cas蛋白包含与SEQ ID NO:1所示序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。可选地,编码嵌合Cas蛋白的第一表达盒的片段包含与SEQ ID NO:25所示的序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。

[0018] 在一些非人动物基因组、非人动物细胞,或非人动物中,所述第一表达盒进一步包含在编码所述嵌合Cas蛋白的片段的上游的多腺苷酸化信号或转录终止子,其中所述多腺苷酸化信号或转录终止子被重组酶识别位点侧接,其中所述多腺苷酸化信号或转录终止子已以组织特异性方式切除。可选地,已经在肝脏中切除了多腺苷酸化信号或转录终止子。可选地,重组酶是Cre重组酶。可选地,非人动物基因组、非人动物细胞,或非人动物还包含基因组整合的重组酶表达盒,所述基因组整合的重组酶表达盒包含可操作地连接至组织特异性启动子的重组酶编码序列。可选地,重组酶基因可操作地连接于表2所示的启动子之一。

[0019] 在一些非人动物基因组、非人动物细胞,或非人动物中,衔接体位于嵌合衔接体蛋

白的N末端,一个或多个转录激活域位于嵌合衔接体蛋白的C末端。可选地,衔接体包含MS2外壳蛋白或其功能片段或变体。可选地,嵌合衔接体蛋白中的一个或多个转录激活域选自:VP16、VP64、p65、MyoD1、HSF1、RTA、SET7/9及其组合。可选地,嵌合衔接体蛋白中的一个或多个转录激活域包含p65和HSF1。

[0020] 在一些非人动物基因组、非人动物细胞,或非人动物中,嵌合衔接体蛋白包含从N末端至C末端:MS2外壳蛋白、核定位信号、p65转录激活域,和HSF1转录激活域。任选地,嵌合衔接体蛋白包含与SEQ ID NO:6所示的序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。可选地,编码嵌合衔接体蛋白的第一表达盒的片段包含与SEQ ID NO:27所示的序列至少90%、95%、96%、97%、98%、99%,或100%相同的序列。

[0021] 在一些非人动物基因组、非人动物细胞,或非人动物中,第一表达盒是多顺反子(multicistronic)。可选地,通过内部核糖体进入位点(IRES)将编码嵌合Cas蛋白的第一表达盒的片段与编码嵌合衔接体蛋白的第一表达盒的片段分开。可选地,通过编码2A肽的核酸将编码嵌合Cas蛋白的第一表达盒的片段与编码嵌合衔接体蛋白的第一表达盒的片段分开。可选地,2A肽是T2A肽。

[0022] 在一些非人动物基因组、非人动物细胞,或非人动物中,第一表达盒被整合到安全港基因座(safe harbor locus)中。在一些非人动物基因组、非人动物细胞,或非人动物中,第一表达盒和/或第二表达盒被整合到安全港基因座中。任选地,非人动物基因组、非人动物细胞,或非人动物对于第一表达盒是杂合的,对于第二表达盒是杂合的,且第一表达盒基因组整合于所述安全港基因座的第一等位基因中,且第二表达盒基因组整合于所述安全港基因座的第二等位基因中。可选地,所述安全港基因座是Rosa26基因座。可选地,所述第一表达盒与所述安全港基因座中的内源启动子可操作地连接。

[0023] 一些这样的非人动物是哺乳动物。可选地,哺乳动物是啮齿动物。可选地,所述啮齿动物是大鼠或小鼠。可选地,所述啮齿动物是小鼠。

[0024] 在另一方面,提供了用于制备以上公开的任何非人动物基因组、非人动物细胞,和非人动物的靶向载体。此类靶向载体可包含被在靶基因组基因座处靶向5'靶序列的5'同源臂和在所述靶基因组基因座处靶向3'靶序列的3'同源臂侧接的插入核酸,其中插入核酸包含表达盒,其中所述表达盒包含(a)编码嵌合成簇的规律间隔的短回文重复序列(CRISPR)相关(Cas)蛋白的核酸,其包含与一个或多个转录激活域融合的核酸酶失活的Cas蛋白;以及(b)编码嵌合衔接体蛋白的核酸,所述嵌合衔接体蛋白包含与一个或多个转录激活域融合的衔接体。

[0025] 在另一方面,提供了制造以上公开的任何非人动物的方法。一些这样的方法包括:(a)向非人动物胚胎干(ES)细胞引入:(i)靶向靶基因组基因座中的靶序列的核酸酶剂;(ii)靶向载体,其包含核酸插入物,所述核酸插入物包含第一表达盒,所述第一表达盒被在靶基因组基因座中对应5'靶序列的5'同源臂和在靶基因组基因座中对应3'靶序列的3'同源臂侧接,其中所述靶向载体与所述靶基因组基因座重组以产生遗传修饰的非人ES细胞,所述遗传修饰的非人ES细胞在其基因组中包含所述靶基因组基因座处的第一表达盒;(b)将经遗传修饰的非人ES细胞引入非人动物宿主胚胎中;以及(c)在代孕母体中孕育非人动物宿主胚胎,其中所述代孕母体产生F0代遗传修饰的非人动物,该非人动物的基因组中包含靶基因组基因座处的第一表达盒。可选地,靶向载体为长度至少为10kb的大靶向载体,或

其中5'和3'同源臂的长度总和至少为10kb。

[0026] 一些这样的方法包括：(a) 向非人动物单细胞期胚胎中引入：(i) 靶向靶基因组基因座中的靶序列的核酸酶剂；以及(ii) 靶向载体，所述靶向载体包含核酸插入物，所述核酸插入物包含第一表达盒，所述第一表达盒被靶基因组基因座中对应5'靶序列的5'同源臂和在靶基因组基因座中对应3'靶序列的3'同源臂侧接，其中所述靶向载体与所述靶基因组基因座重组以产生遗传修饰的非人ES细胞，所述遗传修饰的非人ES细胞在其基因组中包含所述靶基因组基因座处的第一表达盒；(b) 在代孕母体中孕育遗传修饰的非人动物单细胞期胚胎，以产生遗传修饰的F0代非人动物，所述遗传修饰的F0代非人动物在其基因组中包含在靶基因组基因座处的第一表达盒。

[0027] 在一些这样的方法中，所述核酸酶剂包括Cas蛋白和指导RNA。可选地，所述Cas蛋白是Cas9蛋白。可选地，这样的方法可以包括引入在靶基因组基因座内靶向第二靶序列的第二指导RNA。

[0028] 在一些这样的方法中，所述非人动物是小鼠或大鼠。可选地，所述非人动物是小鼠。

[0029] 在另一方面，提供了增加靶基因在任何所述非人动物体内的表达的方法。这样的方法可以包括，例如，将一个或多个指导RNA引入非人动物，每个所述指导RNA包含一个或多个衔接体结合元件，所述嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件，其中所述一个或多个指导RNA与所述嵌合Cas蛋白和嵌合衔接体蛋白形成复合物并将它们引导至靶基因内的靶序列，从而增加靶基因的表达。可选地，靶基因是在肝脏中表达的基因。

[0030] 在一些此类方法中，通过腺相关病毒(AAV)介导的递送引入所述一个或多个指导RNA。可选地，AAV是AAV8。在一些此类方法中，通过脂质纳米颗粒介导的递送或流体动力递送来引入所述一个或多个指导RNA。在一些这样的方法中，将所述一个或多个指导RNA给予非人动物的途径是静脉内注射、脑内(intraparenchymal)注射、腹膜内注射、滴鼻法(nasal installation)，或玻璃体内注射。

[0031] 在一些这样的方法中，所述靶序列包含靶基因内的调控序列。可选地，所述调控序列包含启动子或增强子。在一些这样的方法中，所述靶序列在靶基因的转录起始位点的200个碱基对内。可选地，所述靶序列在转录起始位点上游的200个碱基对和转录起始位点下游的1个碱基对的区域内。

[0032] 在一些这样的方法中，所述一个或多个指导RNA以RNA的形式被引入。在一些这样的方法中，所述一个或多个指导RNA以DNA的形式被引入。可选地，所述一个或多个指导RNA中的每一个可操作地连接至不同的启动子，例如U6启动子。

[0033] 在一些这样的方法中，所述一个或多个指导RNA中的每一个均包含两个衔接体结合元件，所述嵌合衔接体蛋白可以特异性地与所述两个衔接体结合元件结合。可选地，第一衔接体结合元件在所述一个或多个指导RNA的每一个的第一环内，且第二衔接体结合元件在所述一个或多个指导RNA的每一个的第二环内。可选地，所述一个或多个指导RNA中的每一个是单指导RNA，所述单指导RNA包含融合至反式激活CRISPR RNA(tracrRNA)部分的CRISPR RNA(crRNA)部分，其中所述第一环是对应于SEQ ID NO:12的残基13-16的四环，所述第二环是对应于SEQ ID NO:12的残基53-56的茎环2。

[0034] 在一些这样的方法中，所述衔接体结合元件包含SEQ ID NO:16所示的序列。可选

地,所述一个或多个指导RNA中的每一个包含SEQ ID NO:40或63所示的序列。

[0035] 在一些这样的方法中,所述一个或多个指导RNA中的至少一个靶向疾病相关基因。可选地,所述疾病相关基因是Ttr基因,可选地,其中所述Ttr靶向的指导RNA靶向包含SEQ ID NO:34-36中任一个所示的序列的序列,或可选地,其中所述Ttr靶向的指导RNA包含SEQ ID NO:37-39中任一个所示的序列。在一些这样的方法中,所述一个或多个指导RNA中的至少一个靶向Pcsk9基因,可选地,其中Pcsk9靶向的指导RNA靶向包含SEQ ID NO:89-91中任一个所示的序列的序列,或可选地,其中所述Pcsk9靶向的指导RNA包含SEQ ID NO:92-94中任一个所示的序列。可选地,所述方法在非人动物中引起高胆固醇血症。可选地,所述一个或多个指导RNA中的至少一个靶向Ldlr基因,可选地,其中Ldlr靶向的指导RNA靶向包含SEQ ID NO:75-77中任意一项所示的序列的序列,或可选地,其中所述Ldlr靶向的指导RNA包含SEQ ID NO:78-80中任一个所示的序列。

[0036] 在一些这样的方法中,所述一个或多个指导RNA靶向两个或多个靶基因。在一些这样的方法中,所述一个或多个指导RNA包含靶向单个靶基因的多个指导RNA。在一些这样的方法中,所述一个或多个指导RNA包含靶向单个靶基因的至少三个指导RNA。可选地,所述至少三个指导RNA靶向小鼠Ttr基因座,其中第一指导RNA靶向包含SEQ ID NO:34的序列或包含SEQ ID NO:37所示的序列,第二指导RNA靶向包含SEQ ID NO:35的序列或包含SEQ ID NO:38所示的序列,且第三指导RNA靶向包含SEQ ID NO:36的序列或包含SEQ ID NO:39所示的序列。可选地,所述至少三个指导RNA靶向小鼠Pcsk9基因座,其中第一指导RNA靶向包含SEQ ID NO:89的序列或包含SEQ ID NO:92所示的序列,第二指导RNA靶向包含SEQ ID NO:90的序列或包含SEQ ID NO:93所示的序列,且第三指导RNA靶向包含SEQ ID NO:91的序列或包含SEQ ID NO:94所示的序列。可选地,所述至少三个指导RNA靶向小鼠Ldlr基因座,其中第一指导RNA靶向包含SEQ ID NO:75的序列或包含SEQ ID NO:78所示的序列,第二指导RNA靶向包含SEQ ID NO:76的序列或包含SEQ ID NO:79所述的序列,且第三指导RNA靶向包含SEQ ID NO:77的序列或包含SEQ ID NO:80中所述的序列。

[0037] 在一些这样的方法中,靶基因的表达相对于对照非人动物增加至少0.5倍、1倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍,或20倍。

[0038] 在一些这样的方法中,第一表达盒在编码嵌合Cas蛋白的片段的上游还包含多腺苷酸化信号或转录终止子,其中所述多腺苷酸化信号或转录终止子被重组酶识别位点侧接,所述重组酶识别位点被位点特异性重组酶识别,且其中该方法进一步包括将重组酶引入非人动物中。可选地,重组酶是Cre重组酶。可选地,重组酶通过腺相关病毒(AAV)介导的递送而引入。可选地,AAV是AAV8。可选地,重组酶通过脂质纳米颗粒介导的递送或流体动力递送引入。可选地,以组织特异性的方式引入或表达重组酶。可选地,以蛋白形式引入重组酶。可选地,以DNA或RNA的形式引入重组酶。可选地,以与表2中列出的启动子之一可操作地连接的DNA的形式引入重组酶。可选地,向非人动物施用重组酶的途径是静脉内注射、脑内注射、腹膜内注射、滴鼻法,或玻璃体内注射。

[0039] 在一些这样的方法中,一个或多个指导RNA通过腺相关病毒(AAV)介导的递送被引入,所述一个或多个指导RNA中的每一个可操作地连接至不同的U6启动子,并且所述一个或多个指导RNA包括靶向单个基因的多个指导RNA。

[0040] 在另一方面,提供了一种在上述任何非人动物中模拟高胆固醇血症的方法。这样

的方法可包括将一个或多个靶向Pcsk9的指导RNA引入所述非人动物,其中所述一个或多个指导RNA中的每一个包含一个或多个嵌合体结合元件,嵌合衔接体蛋白可以特异性结合至所述衔接体结合元件,其中一个或多个指导RNA与嵌合Cas蛋白和嵌合衔接体蛋白形成复合物,并将它们引导至Pcsk9内的靶序列,从而增加Pcsk9的表达并引起高胆固醇血症。

附图说明

[0041] 图1A(未按比例)示出了lox-stop-lox (LSL) dCas9协同激活介质 (SAM) 等位基因,由5'到3'包含:一个3'剪接序列;第一loxP位点;新霉素抗性基因;多腺苷酸化信号;第二loxP位点;dCas9-NLS-VP64编码序列;T2A肽编码序列;MCP-NLS-p65-HSF1编码序列;以及土拨鼠肝炎病毒转录后调控元件(WPRE)。

[0042] 图1B(未按比例)示出了来自图1A的等位基因,其中floxed新霉素抗性基因和多腺苷酸化信号被去除。

[0043] 图2示出了将图1A中的dCas9 SAM等位基因靶向到Rosa26基因座的第一内含子中的一般示意图。

[0044] 图3A示出了F1H4野生型 (WT) 小鼠胚胎干细胞 (mESCs)、Cas9WT mESCs、lox-stop-lox (LSL) dCas9 SAM mESCs (如图1A所示,具有处于floxed多腺苷酸化信号下游的dCas9 SAM等位基因的mESCs) 和dCas9SAM mESCs (如图1B所示,其中floxed多腺苷酸化信号被Cre重组酶切除的dCas9 SAM等位基因的mESCs) 中的Cas9 mRNA表达水平。

[0045] 图3B示出了F1H4野生型 (WT) 小鼠胚胎干细胞 (mESCs)、Cas9WT mESCs、LSL dCas9 SAM mESCs (见图1A), 和dCas9 SAM mESCs (见图1B) 中的p65 mRNA表达水平。

[0046] 图4示出了F1H4野生型 (WT) 小鼠胚胎干细胞 (mESCs)、Cas9 WT mESCs、LSL dCas9 SAM mESCs (见图1A), 和dCas9 SAM mESCs (见图1B) 中的Cas9蛋白表达水平。

[0047] 图5(未按比例)示出了将指导RNA阵列 (array) 等位基因引入dCas9SAM小鼠胚胎干细胞的示意图。指导RNA阵列等位基因从5'到3'包括:3'剪接序列;第一rox位点;嘌呤霉素抗性基因;多腺苷酸化信号;第二rox位点;第一U6启动子;第一指导RNA编码序列;第二U6启动子;第二指导RNA编码序列;第三U6启动子;第三指导RNA编码序列。

[0048] 图6(未按比例)示出了设计靶向Ttr的转录起始位点上游的三个指导RNA的示意图。

[0049] 图7示出了通用单指导RNA (SEQ ID NO:63) 的示意图,其中四环和茎环2已被MS2结合的适配体替代,以促进融合至转录激活域的嵌合MS2外壳蛋白 (MCP) 的募集。

[0050] 图8A至8C分别示出了被Ttr指导RNA阵列靶定的杂合dCas9 SAM小鼠胚胎干细胞 (mESC) 克隆中的Ttr、Dsg2, 和B4gal t6 mRNA表达水平。通过RT-qPCR确定表达水平。y轴表示循环阈 (ct) 值。使用F1H4野生型mESCs、LSL dCas9 SAM (见图1A), 和dCas9 SAM (见图1B) mESC克隆作为对照。

[0051] 图9A-9L示出了从野生型小鼠、杂合dCas9 SAM小鼠, 和对Ttr指导RNA阵列也是杂合的杂合dCas9 SAM小鼠分离的各种组织中的TTR蛋白表达。

[0052] 图10A和10B分别示出了从野生型小鼠、杂合dCas9 SAM小鼠, 和对Ttr指导RNA阵列也是杂合的杂合dCas9 SAM小鼠分离的肺和脾中的Ttr mRNA表达水平。通过RT-qPCR确定表达水平。y轴表示循环阈 (ct) 值。

[0053] 图10C和10D分别示出了从野生型小鼠、杂合dCas9 SAM小鼠,和对Ttr指导RNA阵列也是杂合的杂合dCas9 SAM小鼠分离的肺和脾中的Dsg2 mRNA表达水平。通过RT-qPCR确定表达水平。y轴表示循环阈(ct)值。

[0054] 图10E和10F分别示出了从野生型小鼠、杂合dCas9 SAM小鼠,和对Ttr指导RNA阵列也是杂合的杂合dCas9 SAM小鼠分离的肺和脾中Bgal^{t6} mRNA的表达水平。通过RT-qPCR确定表达水平。y轴表示循环阈(ct)值。

[0055] 图11示出了通过ELISA测定的野生型小鼠、杂合dCas9 SAM小鼠,和对Ttr指导RNA阵列也是杂合的杂合dCas9 SAM小鼠中TTR的血清水平。

[0056] 图12示出了通过ELISA测定的未经处理的杂合dCas9 SAM小鼠、经AAV8-GFP处理的杂合dCas9 SAM小鼠,和经包含Ttr指导RNA阵列的AAV8处理的杂合dCas9 SAM小鼠中的TTR血清水平,示出了注射后5天、19天,和60天的结果。

[0057]]图13示出了通过ELISA测定的野生型小鼠、杂合dCas9 SAM小鼠,和对Ttr指导RNA阵列也是杂合的杂合dCas9 SAM小鼠中的TTR循环血清水平,示出了注射后3-13个月的结果。

[0058] 图14示出了通过ELISA测定的未经处理的杂合dCas9 SAM小鼠、经AAV8-GFP处理的杂合dCas9 SAM小鼠,和经包含Ttr指导RNA阵列的AAV8处理的杂合dCas9 SAM小鼠中的TTR循环血清水平,显示了注射后5天、19天、2个月、3个月、4个月、5个月、6个月、7个月,和8个月的结果。

[0059] 图15示出了通过ELISA测定的未经处理的纯合dCas9 SAM小鼠、经AAV8-GFP处理的纯合dCas9 SAM小鼠,和经包含Ttr指导RNA阵列或单指导RNA τ 1、2或3的AAV8处理的纯合dCas9 SAM小鼠中的TTR循环血清水平,显示了注射后1星期、2星期和3星期的结果。

[0060] 图16A和16B分别示出了在未经处理的纯合dCas9 SAM小鼠(PCSK9前、LDLR前、WT前、2周WT,或5周WT)、用包含Pcsk9指导RNA阵列的AAV8处理的纯合dCas9 SAM小鼠,或用包含Ldlr指导RNA阵列的AAV8处理的纯合dCas9 SAM小鼠中的胆固醇和LDL水平。

[0061] 图17A和17B示出了从未经处理的纯合dCas9 SAM小鼠、用包含Pcsk9指导RNA阵列的AAV8处理的纯合dCas9 SAM小鼠,和用包含Ldlr指导RNA阵列的AAV8处理的纯合dCas9 SAM小鼠分离的肝脏中的相对Ldlr和Pcsk9 mRNA表达水平。

[0062] 图18A和18B分别示出了在未经处理的纯合dCas9 SAM小鼠(UNT)或用包含Ldlr指导RNA阵列的AAV8处理的纯合dCas9 SAM小鼠(LDLR(HFD))中的胆固醇和LDL水平。

[0063] 图19示出了从未经处理的小鼠、用包含靶基因1指导RNA#1的AAV8处理的纯合dCas9 SAM小鼠、用包含靶基因1指导RNA#2的AAV8处理的纯合dCas9 SAM小鼠,或用包含靶基因1指导RNA#1&2的AAV8处理的纯合dCas9 SAM小鼠分离的肝脏中的靶基因1的相对mRNA表达水平。通过RT-qPCR确定表达水平。y轴表示相对于未处理样品的表达。*表示与未处理相比 $p < 0.0001$ 。**表示与指导RNA#1或指导RNA#2相比, $p < 0.001$ 。

具体实施方式

定义

[0064] 本文可互换使用的术语“蛋白质”、“多肽”,和“肽”包括任何长度的氨基酸的聚合形式,包括编码和非编码氨基酸以及化学或生物化学修饰或衍生的氨基酸。该术语还包括

已被修饰的聚合物,例如具有修饰的肽主链的多肽。术语“结构域”是指具有特定功能或结构的蛋白质或多肽的任何部分。

[0065] 蛋白质被称为具有“N末端”和“C末端”。术语“N末端”是指蛋白质或多肽的起始,末端为具有游离胺基(-NH₂)的氨基酸。术语“C末端”是指氨基酸链(蛋白质或多肽)的末端,其末端为是游离羧基(-COOH)。

[0066] 本文可互换使用的术语“核酸”和“多核苷酸”包括任何长度的核苷酸的聚合形式,包括核糖核苷酸、脱氧核糖核苷酸或其类似物或修饰形式。它们包括单链、双链,和多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂交,以及包含嘌呤碱基、嘧啶碱基,或其它天然的、化学修饰的、生物化学修饰的、非天然的,或衍生的核苷酸碱基。

[0067] 核酸被称为具有“5'末端”和“3'末端”,因为单核苷酸通过使一个单核苷酸戊糖环的5'磷酸通过磷酸二酯键在一个方向上与其邻居的3'氧连接的方式反应生成寡核苷酸。如果寡核苷酸的5'磷酸未与单核苷酸戊糖环的3'氧连接,则将其称为“5'末端”。如果寡核苷酸的3'氧未与另一个单核苷酸戊糖环的5'磷酸连接,则将其末端称为“3'末端”。即使是较大的寡核苷酸内部的核酸序列也可被称为具有5'和3'末端。在线性或环状DNA分子中,离散元素被称为“下游”或3'元素的“上游”或5'。

[0068] 术语“基因组整合”是指已被引入细胞以使核苷酸序列整合到细胞基因组中的核酸。可以使用任何方案将核酸稳定地掺入细胞的基因组中。

[0069] 术语“表达载体”或“表达构建体”或“表达盒”是指重组核酸,其包含可操作地连接至所需编码序列,该所需编码序列可操作地连接至在特定宿主细胞或生物中表达可操作连接的编码序列所必需的合适核酸序列。在原核生物中的表达所需的核酸序列通常包括启动子、操纵子(可选)和核糖体结合位点,以及其它序列。真核细胞通常利用启动子、增强子,以及终止和多腺苷酸化信号,尽管可以在不牺牲必要表达的情况下删除一些元素并添加其它元素。

[0070] 术语“靶向载体”是指可以通过同源重组、非同源末端连接介导的连接,或以任何其它重组方式引入细胞基因组中的靶向位置的重组核酸。

[0071] 术语“病毒载体”是指重组核酸,其包含至少一种病毒来源的元件,并且包括足以或允许包装到病毒载体颗粒中的元件。可以将载体和/或颗粒用于将DNA、RNA,或其它核酸离体或体内转移到细胞中的目的。多种形式的病毒载体是已知的。

[0072] 对于蛋白质、核酸和细胞而言的术语“分离的”包括相对于通常可能原位存在的其它细胞或生物成分而言相对纯化的蛋白质、核酸,和细胞,直至并包括蛋白质、核酸,或细胞的基本纯净的制剂。术语“分离的”还包括不具有天然存在的对应物的蛋白质和核酸,或者化学合成并因此基本上未被其它蛋白质或核酸污染的蛋白质或核酸。术语“分离的”还包括已经与蛋白质、核酸,或细胞所天然伴随的大多数其它细胞成分或生物成分分离或纯化的蛋白质、核酸或细胞(例如,其它细胞蛋白质、核酸,或细胞或细胞外成分)。

[0073] 术语“野生型”包括具有在正常(相对于突变、患病,改变等)的状态或情景中发现的结构和/或活性的实体。野生型基因和多肽通常以多种不同形式(例如等位基因)存在。

[0074] 术语“内源序列”是指天然存在于细胞或非人动物体内的核酸序列。例如,非人动物的内源Rosa26序列是指天然存在于非人动物的Rosa26基因座处的天然Rosa26序列。

[0075] “外源”分子或序列包括通常不以该形式存在于细胞中的分子或序列。通常的存在

包括在细胞的特定发育阶段和环境条件的存在。例如,外源分子或序列可以包括细胞内相应内源序列的突变形式(例如内源序列的人源化版本),或者可以包括与细胞内的内源序列相对应但形式不同(即不在染色体内)的序列。与此相对,内源分子或序列包括在特定环境条件下,在特定发育阶段的特定细胞中通常以该形式存在的分子或序列。

[0076] 当在核酸或蛋白质的上下文中使用时,术语“异源的”表示该核酸或蛋白质包含在同一分子中并非天然同时出现的至少两个片段(segment)。例如,当涉及核酸的片段或蛋白质的片段使用时,术语“异源的”表示该核酸或蛋白质包含两个或更多个在自然界中彼此之间没有相同关系(例如结合在一起)的子序列。作为一个实例,核酸载体的“异源”区域是在另一核酸分子内或附着于另一核酸分子的核酸片段,其在自然界中不与另一分子相关联。例如,核酸载体的异源区域可以包括编码序列,该编码序列侧接在自然界中与该编码序列不相关联的序列。同样地,蛋白质的“异源”区域是在另一肽分子内或附着于另一肽分子的氨基酸片段,其在自然界中不与另一肽分子相关联(例如融合蛋白或带有标签的蛋白)。类似地,核酸或蛋白质可包含异源标记或异源分泌或定位序列。

[0077] “密码子优化”利用密码子的简并性,如可指定氨基酸的三碱基对密码子组合的多样性所展示的,并且通常包括修饰核酸序列的过程,以在保持天然氨基酸序列的前提下,通过用宿主细胞基因中更常用或最常用的密码子替换天然序列的至少一个密码子,从而增强在特定宿主中的表达。例如,可以修饰编码Cas9蛋白的核酸,以用在给定的原核或真核细胞中(包括细菌细胞、酵母细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞,或任何其它宿主细胞)替换天然核酸序列相比具有更高使用频率的密码子。密码子使用情况表很容易例如在“密码子使用情况数据库(Codon Usage Database)”中获得。这些表可以以多种方式适用。参见Nakamura et al. (2000) *Nucleic Acids Research* 28:292,出于所有目的通过引用整体并入本文。也可获得用于在特定宿主中表达的特定序列的密码子优化的计算机算法(参见,例如Gene Forge)。

[0078] 术语“基因座”是指基因(或重要序列)、DNA序列、多肽编码序列的特定位置,或生物体基因组的染色体上的位置。例如,“Ttr基因座”可以指Ttr基因、TtrDNA序列、TTR编码序列,或Ttr在生物体基因组染色体上的特定位置,已识别出这类序列存在于该生物体基因组染色体的何处。“Ttr基因座”可包含Ttr基因的调节元件,包括例如增强子、启动子、5'和/或3'非翻译区(UTR)或其组合。

[0079] 术语“基因”是指染色体中编码产物(例如RNA产物和/或多肽产物)的DNA序列,并包括被非编码内含子打断的编码区和与5'和3'末端上的与该编码区相邻的序列,使得该基因对应于全长mRNA(包括5'和3'非翻译序列)。术语“基因”还包括其它非编码序列,包括调节序列(例如启动子、增强子,和转录因子结合位点)、多腺苷酸化信号、内部核糖体进入位点、沉默子、隔绝序列(insulating sequence),和基质附着区。这些序列可以靠近基因的编码区(例如在10kb之内)或在远处位点,并且它们影响基因转录和翻译的水平或速率。

[0080] 术语“等位基因”是指基因的变体形式。一些基因具有多种不同形式,它们位于染色体上的相同位置或遗传位点上。二倍体生物在每个遗传位点具有两个等位基因。每对等位基因代表特定遗传基因座的基因型。如果在特定位点有两个相同的等位基因,则基因型被描述为纯合,如果两个等位基因不同,则基因型被描述为杂合。

[0081] “启动子”是DNA的调节区,其通常包含能够引导RNA聚合酶II在特定多核苷酸序列

的适当转录起始位点起始RNA合成的TATA盒。启动子可以另外包含影响转录起始速率的其它区域。本文公开的启动子序列调节可操作连接的多核苷酸的转录。启动子可以在本文公开的一种或多种细胞类型(例如真核细胞、非人类哺乳动物细胞、人类细胞、啮齿动物细胞、多能细胞、单细胞期胚胎、分化的细胞,或其组合)中具有活性。启动子可以是例如组成型活性启动子、条件启动子(conditional promoter)、诱导型启动子、时间受限的启动子(temporally restricted promoter)(例如发育调控的启动子(developmentally regulated promoter)),或空间受限的启动子(spatially restricted promoter)(例如细胞特异性或组织特异性启动子)。启动子的例子可以在例如WO 2013/176772中找到,其全部内容通过引用并入本文。

[0082] 组成型启动子是在所有发育阶段的所有组织或特定组织中都有活性的启动子。组成型启动子的实例包括人即早期(immediate early)巨细胞病毒(hCMV)启动子、小鼠即早期巨细胞病毒(mCMV)启动子、人延伸因子1 α (hEF1 α)启动子、小鼠延伸因子1 α (mEF1 α)启动子、小鼠磷酸甘油酸酯激酶(PGK)启动子、鸡 β 肌动蛋白杂种(CAG或CBh)启动子、SV40早期启动子,和 β 2微管蛋白启动子。

[0083] 诱导型启动子的实例包括,例如,化学调节的启动子和物理调节的启动子。化学调节的启动子包括,例如,醇调节的启动子(例如,醇脱氢酶(alcA)基因启动子)、四环素调节的启动子(例如四环素响应性启动子、四环素操纵子序列(tetO)、tet-On启动子,或tet-Off启动子)、类固醇调节的启动子(例如大鼠糖皮质激素受体、雌激素受体的启动子,或蜕皮素受体的启动子),或金属调节的启动子(例如金属蛋白启动子)。物理调节的启动子包括,例如,温度调节的启动子(例如热激启动子)和光调节的启动子(例如光诱导型启动子或光抑制型启动子(light-repressible promoter))。

[0084] 组织特异性启动子可以是,例如,神经元特异性启动子、胶质细胞特异性启动子、肌肉细胞特异性启动子、心脏细胞特异性启动子、肾细胞特异性启动子、骨细胞特异性启动子、内皮细胞特异性启动子,或免疫细胞特异性启动子(例如,B细胞启动子或T细胞启动子)。

[0085] 发育调节的启动子包括,例如,仅在胚胎发育阶段或仅在成年细胞中才有活性的启动子。

[0086] 可操作的连接”或“可操作地连接”是指两个或更多个成分(例如启动子和另一个序列元件)的并置,使得两个成分均正常起作用,并允许至少一个成分可以介导施加在至少一个其它成分上的功能。例如,如果启动子响应一种或多种转录调节因子的存在或不存在而控制编码序列的转录水平,则该启动子可以可操作地连接至该编码序列。可操作的连接可以包括这样的序列,所述序列彼此连续或反式作用(例如,调节序列可以在一定距离处起作用以控制编码序列的转录)。

[0087] 核酸的“互补性”是指核酸的一条链中的核苷酸序列由于其核碱基的取向而与相对的核酸链上的另一序列形成氢键。DNA中的互补碱基通常是A和T,C和G。在RNA中,它们通常是C和G,U和A。互补性可以是完美的或大量的/充分的。两个核酸之间的完美互补性意味着两个核酸可以形成双链体,该双链体中的每个碱基都通过沃森-克里克配对与互补碱基结合。“大量的”(substantial)或“充分的”互补意指一条链中的序列与相对链中的序列不完全和/或不完美互补,但是两条链上的碱基之间在一系列杂交条件(例如盐浓度和温度)

下发生足够的键合以形成稳定的杂合复合体。可以通过使用序列和标准数学计算来预测杂交链的 T_m (解链温度),或通过使用常规方法凭实验确定 T_m 来预测此类条件。 T_m 包括两条核酸链之间形成的杂交复合物群体50%变性(即双链核酸分子群体一半解离为单链)的温度。低于 T_m 的温度有利于杂交复合物的形成,而高于 T_m 的温度有利于杂交复合物中链的解链或分离。可以通过使用例如 $T_m=81.5+0.41(\%G+C)$ 估算1M NaCl水溶液中具有已知G+C含量的核酸的 T_m ,尽管其它已知的 T_m 计算考虑了核酸结构特征。

[0088] “杂交条件”包括累积环境(cumulative environment),在该累积环境中一个核酸链通过互补链相互作用和氢键结合至第二核酸链以产生杂交复合物。此类条件包括含有核酸的水溶液或有机溶液的化学成分及其浓度(例如盐、螯合剂、甲酰胺),以及混合物的温度。其它因素(例如孵育时间的长度或反应室的尺寸)可能会对环境造成影响。参见,例如, Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2. sup. nd ed., pp. 1.90-1.91, 9.47-9.51, 11.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 纽约, 1989), 出于所有目的通过引用将其全部内容合并于此。

[0089] 杂交需要两个核酸包含互补序列,尽管碱基之间可能错配。适用于两种核酸之间杂交的条件取决于核酸的长度和互补程度,这是众所周知的变量。两个核苷酸序列之间的互补程度越大,具有这些序列的核酸的杂交体的解链温度(T_m)的值越大。对于具有短互补序列片段(例如35个或更少、30个或更少、25个或更少、22个或更少、20个或更少,或18个或更少的核苷酸的互补性)的核酸之间的杂交,错配的位置变得很重要(参见Sambrook等,同上,11.7-11.8)。通常,可杂交核酸的长度为至少约10个核苷酸。可杂交核酸的说明性最小长度包括至少约15个核苷酸、至少约20个核苷酸、至少约22个核苷酸、至少约25个核苷酸,和至少约30个核苷酸。此外,可以根据诸如互补区域的长度和互补程度等因素,根据需要调节温度和洗涤液盐浓度。

[0090] 多核苷酸的序列不必与其靶核酸的序列100%互补以特异性杂交。此外,多核苷酸可在一个或多个片段上杂交,使得在杂交事件中不涉及间插(intervening)或相邻片段(例如环结构或发夹结构)。多核苷酸(例如gRNA)可包含与其靶向的靶核酸序列内的靶区域至少70%、至少80%、至少90%、至少95%、至少99%,或100%的序列互补性。例如,其中20个核苷酸中的18个与靶区域互补并且因此将特异性杂交的gRNA将代表90%的互补性。在该实例中,剩余的非互补核苷酸可以成簇(clustered)或散在(interspersed)于互补核苷酸,并且不需要彼此或与互补核苷酸邻接(contiguous)。

[0091] 可以使用BLAST程序(基本局部比对搜索工具)和PowerBLAST程序(Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; Zhang和Madden (1997) *Genome Res.* 7:649-656),或使用Gap程序(Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.)的默认设置(其使用Smith和Waterman的算法(*Adv. Appl. Math.*, 1981, 2, 482-489)),常规地确定核酸中特定核酸序列段之间的互补性百分比。

[0092] 本文提供的方法和组合物采用多种不同的组分。说明书全文中的某些组分可能具有有活性的变体和片段。这样的组分包括,例如,Cas蛋白、CRISPR RNA、tracrRNA,和指导RNA。这些组分的每一种的生物活性在本文其它地方描述。术语“功能性”是指蛋白质或核酸(或其片段或变体)表现出生物学活性或功能的固有能力。这种生物学活性或功能可以包

括,例如,Cas蛋白与指导RNA和靶DNA序列结合的能力。与原始片段相比,功能片段或变体的生物学功能可以相同或实际上可以改变(例如就其特异性或选择性或功效而言),但是保留了基本生物学功能。

[0093] 术语“变体”是指不同于群体中最普遍的序列的核苷酸序列(例如一个核苷酸不同)或不同于群体中最普遍的序列的蛋白序列(例如一个氨基酸不同)。

[0094] 当涉及蛋白质时,术语“片段”是指比全长蛋白质短或具有更少氨基酸的蛋白质。当涉及核酸时,术语“片段”是指比全长核酸短或具有更少核苷酸的核酸。片段可以是,例如,N末端片段(即去除蛋白质的C末端的一部分)、C末端片段(即去除蛋白质的N末端的一部分),或内部片段。

[0095] 在两个多核苷酸或多肽序列的上下文中,“序列同一性”或“同一性”是指两个序列中的残基,当在指定的比较窗口上比对以获得最大对应性时它们是相同的。当使用序列同一性百分比来表示蛋白质时,不相同的残基位置的不同之处通常为保守的氨基酸取代,其中氨基酸残基被具有相似化学性质(例如电荷或疏水性)的其它氨基酸残基取代,并因此不改变分子的功能特性。当序列的不同之处为保守取代时,可以向上调整序列同一性百分比以对取代的保守性质进行校正。因这种保守取代而不同的序列被称为具有“序列相似性”或“相似性”。进行这种调整的方法是众所周知的。通常,这涉及将保守取代计为部分错配而不是完全错配,从而增加序列同一性百分比。因此,例如,在相同氨基酸的评分为1,非保守取代的评分为0的情况下,保守取代的评分为0与1之间的值。保守取代的得分的计算通过例如PC/GENE程序实现(加利福尼亚州,山景城,Intelligenetics)。

[0096] “序列同一性百分比”包括通过比较窗口中比较两个最佳比对的序列(完美匹配残基的最大数目)而确定的值,其中为了两个序列的最佳比对,比较窗口中的多核苷酸序列的部分可以包括与参考序列(其不含添加或缺失)相比的添加或删除(即空隙)。通过确定两个序列中出现相同核酸碱基或氨基酸残基的位置数,以产生匹配位置数,将匹配位置数除以比较窗口中的位置总数,然后将结果乘以100,以得到序列同一性百分比。除非另有说明(例如较短的序列包括连接的异源序列),否则比较窗口是被比较的两个序列中较短的序列的全长。

[0097] 除非另有说明,否则序列同一性/相似性值包括使用GAP版本10,利用以下参数获得的值:使用GAP权重50和长度权重3以及nws gapdna.cmp评分矩阵获得核苷酸序列的同一性%和相似性%;使用GAP权重8和长度权重2,以及BLOSUM62评分矩阵获得氨基酸序列同一性%和相似性%;或其任何等效程序。“等效程序”包括任何序列比较程序,当与由GAP版本10生成的相应比对进行比较时,对于所讨论的任何两个序列,所述序列比较程序针对所讨论的任何两个序列生成具有相同核苷酸或氨基酸残基匹配和相同百分比的序列同一性的比对。

[0098] 术语“保守氨基酸取代”是指序列中通常存在的氨基酸被具有相似大小,电荷或极性的不同氨基酸取代。保守取代的实例包括用非极性(疏水)残基如异亮氨酸、缬氨酸,或亮氨酸取代另一个非极性残基。同样地,保守取代的例子包括一个极性(亲水)残基被另一个所取代,例如精氨酸和赖氨酸之间的取代、谷氨酰胺和天冬酰胺之间的取代,或甘氨酸和丝氨酸之间的取代。另外,保守取代的其它实例是用碱性残基例如赖氨酸、精氨酸,或组氨酸取代另一种碱性残基,或用一个酸性残基例如天冬氨酸或谷氨酸替代另一种酸性残基。非

保守取代的实例包括将非极性(疏水)氨基酸残基,例如异亮氨酸、缬氨酸、亮氨酸、丙氨酸,或蛋氨酸取代为极性(亲水)残基,如半胱氨酸、谷氨酰胺、谷氨酸或赖氨酸和/或将极性残基取代为非极性残基。下文的表1总结了典型的氨基酸分类。

[0099] 表1:氨基酸分类

丙氨酸	Ala	A	非极性	中性	1.8
精氨酸	Arg	R	极性	阳性	-4.5
天冬酰胺	Asn	N	极性	中性	-3.5
天冬氨酸	Asp	D	极性	阴性	-3.5
半胱氨酸	Cys	C	非极性	中性	2.5
谷氨酸	Glu	E	极性	阴性	-3.5
谷氨酰胺	Gln	Q	极性	中性	-3.5
甘氨酸	Gly	G	非极性	中性	-0.4
组氨酸	His	H	极性	阳性	-3.2
异亮氨酸	Ile	I	非极性	中性	4.5
亮氨酸	Leu	L	非极性	中性	3.8
赖氨酸	Lys	K	极性	阳性	-3.9
蛋氨酸	Met	M	非极性	中性	1.9
苯丙氨酸	Phe	F	非极性	中性	2.8
脯氨酸	Pro	P	非极性	中性	-1.6
丝氨酸	Ser	S	极性	中性	-0.8
苏氨酸	Thr	T	极性	中性	-0.7
色氨酸	Trp	W	非极性	中性	-0.9
酪氨酸	Tyr	Y	极性	中性	-1.3
缬氨酸	Val	V	非极性	中性	4.2

[0100] “同源”序列(例如核酸序列)是指与已知参考序列相同或基本相似的序列,使得其例如与已知参考序列至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%,或100%相同。同源序列可以包括,例如,种间同源序列和种内同源序列。同源基因例如通常通过物种形成事件(speciation event)(种间同源基因)或遗传复制事件(genetic duplication event)(种内同源基因)源自共同的祖先DNA序列。“种间同源”基因包括通过物种形成从共同祖先基因进化而来的不同物种中的基因。种间同源物(orthologs)通常在进化过程中保持相同的功能。“种内同源”基因包括通过基因组内的复制而相关的基因。种内同源物(paralogs)可以在进化过程中进化出新功能。

[0101] 术语“体外”包括人工环境以及在人工环境(例如试管)内发生的过程或反应。术语

“体内”包括自然环境(例如细胞或生物体或身体)以及在自然环境中发生的过程或反应。术语“离体”包括已经从个体体内移出的细胞以及在这种细胞内发生的过程或反应。

[0102] 术语“报告基因”是指具有编码基因产物(通常是酶)的序列的核酸,当包含与内源或异源启动子和/或增强子元件可操作地连接的报告基因序列的构建体被引入含有(或可以使其含有)该启动子和/或增强子元件的激活所必需的因子的细胞中时,该序列可被容易地和定量地测定。报告基因的例子包括但不限于编码 β -半乳糖苷酶(lacZ)的基因、细菌氯霉素乙酰基转移酶(cat)基因、萤火虫荧光素酶基因、编码 β -葡糖醛酸糖苷酶(GUS)的基因,和编码荧光蛋白的基因。“报告蛋白”是指由报告基因编码的蛋白。

[0103] 本文所使用的术语“荧光报告蛋白”是指基于荧光可检测的报告蛋白,其中荧光可以直接来自报告蛋白、报告蛋白在荧光底物上的活性,或具有结合荧光标记的化合物的亲和力的蛋白。荧光蛋白的例子包括绿色荧光蛋白(例如GFP、GFP-2、tagGFP、turboGFP、eGFP、Emerald、Azami Green、单体Azami Green、CopGFP、AceGFP,和ZsGreen1)、黄色荧光蛋白(例如YFP、eYFP、Citrine、Venus、YPet、PhiYFP,和ZsYellow1)、蓝色荧光蛋白(例如BFP、eBFP、eBFP2、Azurite、mKalamal、GFPuv, Sapphire, 和T-sapphire)、青色荧光蛋白(例如CFP、eCFP、Cerulean、CyPet、AmCyan1,和Midoriishi-Cyan)、红色荧光蛋白(例如RFP、mKate、mKate2、mPlum、DsRed单体、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-Monomer、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry,和Jred)、橙色荧光蛋白(例如mOrange、mKO、Kusabira-Orange、Monomeric Kusabira-Orange、mTangerine,和tdTomato),以及可以通过流式细胞方法检测到其在细胞中的存在的任何其它合适的荧光蛋白。

[0104] 响应双链断裂(DSB)的修复主要通过两个保守的DNA修复途径发生:同源重组(HR)和非同源性末端接合(NHEJ)。参见Kasperek&Humphrey(2011) *Seminars in Cell& Dev. Biol.* 22:886-897,出于所有目的通过引用整体并入本文。同样,由外源供体核酸介导的靶核酸的修复可以包括两个多核苷酸之间交换遗传信息的任何过程。

[0105] 术语“重组”包括两个多核苷酸之间的遗传信息交换的任何过程,并且可以通过任何机制发生。重组可通过同源定向修复(HDR)或同源重组(HR)发生。HDR或HR包括一种可能需要核苷酸序列同源性的核酸修复形式,其使用“供体”分子作为模板来修复“靶标”分子(即经历双链断裂的分子),并导致遗传信息从供体转移到靶标。不希望受到任何特定理论的束缚,这种转移可能涉及在断裂的靶标和供体之间形成的异源双链DNA的错配校正,和/或合成依赖性链退火,其中供体用于重新合成遗传信息,该遗传信息将成为靶标的一部分,和/或相关过程。在一些情况下,供体多核苷酸、供体多核苷酸的一部分、供体多核苷酸的拷贝,或供体多核苷酸的拷贝的一部分整合到靶DNA中。参见Wang et al. (2013) *Cell* 153: 910-918; Mandalos et al. (2012) *PLoS ONE* 7:e45768:1-9;和Wang et al. (2013) *Nat Biotechnol.* 31:530-532,出于所有目的将其每一篇均通过引用整体并入本文。

[0106] NHEJ包括通过将断裂末端彼此直接连接或与外源序列直接连接而不需要同源模板来修复核酸中的双链断裂。通过NHEJ对非连续序列进行连接通常会导致双链断裂位点附近的删除、插入,或易位。例如,NHEJ还可以通过将断裂末端与外源供体核酸的末端直接连接(即基于NHEJ的捕获)来导致外源供体核酸的靶向整合。当同源定向修复(HDR)途径不能立即使用时(例如,在非分裂细胞、原代细胞,以及较差地进行基于同源性的DNA修复的细胞

中),此类NHEJ介导的靶向整合可优选用于插入外源供体核酸。另外,与同源定向修复相反,不需要关于切割位点侧接的大范围的序列同一性的知识,这在尝试靶向插入到具有对基因组序列的认识有限的基因组的生物中时可能是有益的。整合可以通过在外源供体核酸与切割的基因组序列之间平端的连接或通过使用侧面为单链突出端的外源供体核酸连接黏性末端(即具有5'或3'单链突出端)来进行,所述单链突出端与在切割的基因组序列中由核酸酶剂产生的那些相容。参见,例如,US 2011/020722、WO 2014/033644、WO 2014/089290和Maresca et al. (2013) GenomeRes. 23(3):539-546,出于所有目的将每一篇均通过引用整体并入本文。如果连接平端,可能需要进行靶标和/或供体切除,以产生片段连接所需的微同源性区域(regions of microhomology),这可能会在靶序列中产生不想要的改变。

[0107] “包含”或“包括”一个或多个所列举的要素的组合物或方法可以包括未具体列举的其它要素。例如,“包含”或“包括”蛋白质的组合物可以单独包含蛋白质或包含蛋白质以及其它成分。过渡用语“基本上由……组成”是指权利要求的范围应解释为涵盖权利要求中所述的特定要素以及不实质上影响所要求保护的发明的基本和新颖特征的要素。因此,当在本发明的权利要求书中使用时,术语“基本上由……组成”并不意图被解释为等同于“包括”。

[0108] “可选的”或“可选地”是指随后描述的事件或情况可能发生或可能不会发生,并且说明书包括事件或情况发生的情况以及事件或情况没有发生的情况。

[0109] 值范围的指定包括该范围内或定义该范围的所有整数,以及该范围内的整数定义的所有子范围。

[0110] 除非从上下文中另外显而易见,否则术语“约”涵盖所述值的标准测量误差范围(例如SEM)内的值。

[0111] 术语“和/或”是指并涵盖一个或多个相关列出的项目的任何和所有可能的组合,以及当以备选方式(“或”)解释时组合的不存在。

[0112] 术语“或”是指特定列表的任何一个成员,并且还包含该列表的成员的任意组合。

[0113] 除非上下文另外明确指出,否则冠词“一”,“一个”和“所述”的单数形式包括复数形式。例如,术语“一种蛋白质”或“至少一种蛋白质”可包括多种蛋白质,包括其混合物。

[0114] 具有统计学意义意指 $p \leq 0.05$ 。

详细说明

I. 概述

[0115] 成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关(Cas) (CRISPR/Cas9) 系统是用于基因组工程和调节靶基因表达的强大工具。该系统在体内的一个局限性是需要将所有组分同时引入到活生物体中。通常,这些成分是通过将DNA构建体转染到会产生适当RNA和蛋白质的细胞中而瞬时引入的。尽管有效,但是这种方法具有固有的缺点,因为细胞必须依靠质粒DNA构建体先进行转录,然后翻译,然后才能使用Cas蛋白与sgRNA组分相互作用。需要更好的方法和工具来更有效地评估CRISPR/Cas试剂的活性,并评估靶向体内特定组织或细胞类型的不同递送方法和参数。

[0116] 在示例性的CRISPR/Cas协同激活介质(SAM)系统中,几个激活域相互作用以引起比单独由任何一个因子诱导的转录激活更大的转录激活。要使用SAM系统,通常需要引入三种病毒。第一种病毒包含直接与VP64域融合的催化失活的Cas蛋白、由四个串联重复的单纯

疱疹病毒蛋白16组成的转录激活因子。当VP64与在转录起始位点附近结合的蛋白融合时，它作用为强转录激活因子。第二种病毒携带入MS2外壳蛋白(MCP)，MS2外壳蛋白与另外两个激活的转录因子融合：热休克因子1(HSF1)；和转录因子65(p65)。MCP自然结合到MS2茎环。在示例性的SAM系统中，MCP与被工程化进CRISPR相关的sgRNA的MS2茎环相互作用，从而将结合的转录因子穿梭到适当的基因组位置。第三种病毒引入了包含MS2环的sgRNA。

[0117] 本文提供了用于体内和离体激活靶基因的转录以及体内和离体评估CRISPR/Cas介导的转录激活活性的方法和组合物。所述方法和组合物使用细胞和非人动物，其包含嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒，或协同激活介质(SAM)表达盒(例如嵌合Cas蛋白编码序列和嵌合衔接体蛋白序列)，使得这些组分可以组成性地获得(constitutively available)，或者例如组织特定的或时间特定的方式获得。盒可以是基因组整合的。这样的细胞和非人动物还可以包含如本文其它地方公开的指导RNA表达盒和/或重组酶表达盒。替代地，可通过其它方式将一种或多种组分(例如指导RNA和/或重组酶)引入细胞和非人动物中，以诱导靶基因的转录激活。

[0118] 包含SAM表达盒的非人动物简化了在体内测试CRISPR/Cas组分的递送和活性的过程，因为仅需要将指导RNA引入非人动物以激活靶基因的转录。如果非人动物还包含指导RNA表达盒，则可以在不引入任何其它组分的情况下研究靶基因激活的作用。另外，SAM表达盒或指导RNA表达盒可以可选地是条件表达盒，其可以在特定组织或发育阶段中选择性表达，这可以，例如，降低体内Cas介导的毒性的风险。替代地，此类表达盒可以被组成性地表达，以使得能够测试任何和所有类型的细胞、组织，和器官中的活性。

[0119] 还提供了用于制造和使用这些非人动物的方法和组合物，以测试和测量基于Cas的SAM系统在体内激活靶基因的转录的能力，或评估体内增加靶基因转录的影响。

II. 包含协同激活介质(SAM)表达盒的非人动物

[0120] 本文公开的非人动物基因组、非人动物细胞，和非人动物包含基于成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关(Cas)(CRISPR/Cas9)的协同激活介质(SAM)表达盒，以用于体内或离体激活靶基因转录的方法，并评估SAM系统或此类系统的组分(例如引入非人动物或细胞中的指导RNA)体内或离体激活靶基因组基因座转录的能力。本文公开的方法和组合物利用包含成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关(Cas)(CRISPR/Cas9)的协同激活介质(SAM)表达盒的非人动物或细胞，以用于体内或离体激活靶基因转录的方法，并评估SAM系统或此类系统的组分(例如引入非人动物或细胞中的指导RNA)体内或离体激活靶基因组基因座转录的能力。本文所述的SAM系统包含嵌合Cas蛋白和嵌合衔接体蛋白，并且可以与本文其它地方所述的指导RNA一起使用，以激活靶基因的转录。指导RNA可以通过基因组整合的表达盒编码，或者可以通过AAV或任何其它合适的方式提供。嵌合Cas蛋白和嵌合衔接体蛋白(例如包含与指导RNA内的衔接体结合元件特异性结合的衔接体；以及一个或多个异源转录激活域)在本文其它地方进一步详细描述。

[0121] CRISPR/Cas系统包括与Cas基因的表达或引导Cas基因的活性有关的转录物和其它元件。CRISPR/Cas系统可以是例如I型、II型、III型系统或V型系统(例如，V-A亚型或V-B亚型)。本文公开的组合物和方法中使用的CRISPR/Cas系统可以是非天然存在的。“非天然存在”的系统包括指示人手所参与到的任何事物，例如系统的一个或多个组分，其已从其自然存在的状态更改或变异、其至少基本上没有至少一种与它们在自然界中天然相关联的其

它组分的存在,或者其与至少一个与它们非天然相关联的其它组件相关联。例如,某些CRISPR/Cas系统采用非天然存在的CRISPR复合物,该复合物包含不天然一起存在的gRNA和Cas蛋白、采用不天然存在的Cas蛋白,或采用不天然存在的gRNA。

[0122] 本文公开的方法和组合物通过使用或测试CRISPR复合物(包含与嵌合Cas蛋白和嵌合衔接体蛋白复合的指导RNA(gRNA))在体内诱导靶基因组位点转录激活的能力来采用CRISPR/Cas系统。

[0123] 本文公开的基因组、细胞,和非人动物包含嵌合Cas蛋白表达盒和/或嵌合衔接体蛋白表达盒。例如,本文公开的基因组、细胞,和非人动物可包含协同激活介质(SAM)表达盒,其包含嵌合Cas蛋白编码序列和嵌合衔接体蛋白编码序列。

[0124] 此类包含SAM表达盒的基因组、细胞,或非人动物的优势在于:仅需要递送指导RNA,即可诱导靶基因组基因座的转录激活。一些这样的基因组、细胞,或非人动物也包含指导RNA表达盒,从而使得靶基因转录激活所需的所有组分已存在。SAM系统可用于此类细胞中,以任何所需方式提供靶基因的增加表达。例如,一种或多种靶基因的表达可以以组成性方式或调节性方式(例如可诱导的、组织特异性的、时间调节性的等)增加。

A. 嵌合Cas蛋白

[0125] 提供了可以与本文其它地方公开的指导RNA结合以激活靶基因转录的嵌合Cas蛋白。此类嵌合Cas蛋白可包含:(a) DNA结合结构域,其为成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关(Cas)蛋白或其功能片段或变体,其能够与指导RNA形成复合物,并与靶序列结合;以及(b) 一个或多个转录激活域或其功能片段或变体。例如,此类融合蛋白可包含1、2、3、4、5,或更多个转录激活域(例如两个或更多个异源转录激活域,或三个或更多个异源转录激活域)。在一个实施例中,嵌合Cas蛋白可包含催化失活的Cas蛋白(例如dCas9)和VP64转录激活域或其功能片段或变体。例如,这种嵌合Cas蛋白可包含,基本上由,或由以下组成:与SEQ ID NO:1所示的dCas9-VP64嵌合Cas蛋白序列至少具有85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性的氨基酸序列。然而,还提供了嵌合Cas蛋白,其中转录激活域包含其它转录激活域或功能片段或其变体,和/或其中还提供了该包含其它Cas蛋白(例如催化失活的Cas蛋白)的Cas蛋白。其它合适的转录激活域的实例在本文其它地方提供。

[0126] 转录激活域可以位于N末端、C末端,或Cas蛋白内的任何位置。例如,转录激活域可以附着于化脓链球菌(*Streptococcus pyogenes*) Cas9蛋白的Rec1结构域、Rec2结构域、HNH结构域,或PI结构域,或当与化脓链球菌Cas9蛋白质最佳比对时的种间同源Cas9蛋白或同源或种间同源Cas蛋白的任何相应区域。例如,转录激活域可以在553位置附着于Rec1结构域,在575位置附着于Rec1结构域,在175-306位置中的任何位置附着于Rec2结构域或替换175-306位置的部分或整个区域,在715-901位置内的任何位置附着于HNH结构域或替换715-901位置的部分或整个区域,或在化脓链球菌Cas9蛋白的1153位置附着于PI结构域。参见例如W0 2016/049258,出于所有目的通过引用将其全文并入本文。转录激活域可以如本文其它地方所述,在其一侧或两侧上侧接一个或多个接头。

[0127] 嵌合Cas蛋白也可以可操作地连接或融合至其它异源多肽。融合或连接的异源多肽可以位于嵌合Cas蛋白的N末端、C末端,或其内部的任何地方。例如,嵌合Cas蛋白可以进一步包含核定位信号。合适的核定位信号和对Cas蛋白的其它修饰的实例在本文其它地方

进一步详细描述。

(1) Cas蛋白

[0128] Cas蛋白通常包含至少一个可与指导RNA相互作用的RNA识别或结合结构域。Cas蛋白的功能片段或功能变体是保留与指导RNA形成复合物并结合靶基因中的靶序列的能力(并且例如激活靶基因的转录)的功能片段或功能变体。

[0129] 除了本文其它地方所述的转录激活域之外, Cas蛋白还可包含核酸酶结构域(例如, DNA酶结构域或RNA酶结构域)、DNA结合结构域、解旋酶结构域、蛋白质-蛋白质相互作用结构域、二聚化结构域, 和其它域。一些此类结构域(例如DNA酶结构域)可以来自天然Cas蛋白。可以添加其它此类结构域以制备修饰的Cas蛋白。核酸酶结构域具有用于核酸切割的催化活性, 核酸切割包括核酸分子的共价键的断裂。切割可产生平端(blunt ends)或交错端(staggered ends), 并且可以是单链或双链的。例如, 野生型Cas9蛋白通常会平端(blunt)切割产物。替代地, 野生型Cpf1蛋白(例如FnCpf1)可产生具有5-核苷酸5'单链突出端的切割产物, 该切割发生于非靶向链上的PAM序列的第18个碱基对之后和靶向链的第23个碱基之后。Cas蛋白可以具有完全的切割活性, 从而在靶基因组位点处产生双链断裂(例如, 末端为平端的双链断裂), 或者它可以是在靶基因组位点上产生单链断裂的切口酶。在一个实例中, 本文公开的嵌合Cas蛋白的Cas蛋白部分已被修饰为具有降低的核酸酶活性(例如核酸酶活性与野生型Cas蛋白相比降低了至少约70%、75%、80%、85%、90%、95%、96%、97%、98%、99%, 或100%)或基本没有任何核酸酶活性(即核酸酶活性与野生型Cas蛋白相比降低了至少90%、95%、97%、98%、99%, 或100%, 或具有不超过野生型Cas蛋白的核酸酶活性的约0%、1%、2%、3%、5%, 或10%的核酸酶活性)。核酸酶失活的Cas蛋白是一种具有突变的Cas蛋白, 已知该突变是在其催化(即核酸酶)结构域中的失活突变(例如: Cpf1蛋白中的RuvC样核酸内切酶结构域中的失活突变, 或Cas9中的HNH核酸内切酶结构域和RuvC样核酸内切酶结构域中的失活突变), 或该核酸酶失活的Cas蛋白是与野生型Cas蛋白相比, 核酸酶活性减少了至少约97%、98%、99%, 或100%的Cas蛋白。下文公开了减少或基本消除核酸酶活性的不同的Cas蛋白突变的实例。

[0130] Cas蛋白的实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5e (CasD)、Cas6、Cas6e、Cas6f、Cas7、Cas8a1、Cas8a2、Cas8b、Cas8c、Cas9 (Csn1或Csx12)、Cas10、Cas10d、CasF、CasG、CasH、Csy1、Csy2、Csy3、Cse1 (CasA)、Cse2 (CasB)、Cse3 (CasE)、Cse4 (CasC)、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4, 和Cu1966, 以及其同源物或修饰版本。

[0131] 示例性的Cas蛋白是Cas9蛋白或衍生自Cas9蛋白的蛋白。Cas9蛋白来自II型CRISPR/Cas系统, 通常与保守结构具有相同的四个关键模体(motifs)。模体1、2和4是RuvC样模体, 模体3是HNH模体。示例性Cas9蛋白质来自化脓链球菌(*Streptococcus pyogenes*)、嗜热链球菌(*Streptococcus thermophilus*)、链球菌属(*Streptococcus sp.*)、金黄色葡萄球菌(*Staphylococcus aureus*)、达森维尔诺卡氏菌(*Nocardiosis dassonvillei*)、始旋链霉菌(*Streptomyces pristinaespiralis*)、绿色产色链霉菌(*Streptomyces viridochromogenes*)、绿色产色链霉菌(*Streptomyces viridochromogenes*)、粉红链孢囊菌(*Streptosporangium roseum*)、粉红链孢囊菌(*Streptosporangium roseum*)、酸性脂环

酸杆菌 (*Alicyclobacillus acidocaldarius*)、假蕈状芽孢杆菌 (*Bacillus pseudomycooides*)、还原硒酸盐芽孢杆菌 (*Bacillus selenitireducens*)、西伯利亚微小杆菌 (*Exiguobacterium sibiricum*)、保加利亚乳杆菌 (*Lactobacillus delbrueckii*)、唾液乳杆菌 (*Lactobacillus salivarius*)、海洋微颤菌 (*Microscilla marina*)、伯克霍尔德氏菌 (*Burkholderiales bacterium*)、食茶极地单胞菌 (*Polaromonas naphthalenivorans*)、极地单胞菌属 (*Polaromonas sp.*)、海洋固氮蓝藻 (*Crocospaera watsonii*)、蓝丝细菌属 (*Cyanothece sp.*)、铜绿微囊藻 (*Microcystis aeruginosa*)、聚球藻属 (*Synechococcus sp.*)、阿拉伯醋酸喜盐菌 (*Acetohalobium arabaticum*)、丹氏制氮菌 (*Ammonifex degensii*)、热解纤维素菌 (*Caldicelulosiruptor becsii*)、金矿菌 (*Candidatus Desulforudis*)、肉毒梭菌 (*Clostridium botulinum*)、难辨梭菌 (*Clostridium difficile*)、大芬戈尔德菌 (*Finegoldia magna*)、嗜热盐碱厌氧菌 (*Natranaerobius thermophilus*)、嗜热丙酸降解发酵菌 (*Pelotomaculum thermopropionicum*)、喜温嗜酸硫杆菌 (*Acidithiobacillus caldus*)、氧化亚铁硫杆菌 (*Acidithiobacillus ferrooxidans*)、酒色别样着色菌 (*Allochromatium vinosum*)、海杆菌属 (*Marinobacter sp.*)、嗜盐亚硝化球菌 (*Nitrosococcus halophilus*)、亚硝化球菌 (*Nitrosococcus watsoni*)、游海假交替单胞菌 (*Pseudoalteromonas haloplanktis*)、成簇细枝菌 (*Ktedonobacter racemifer*)、调查甲烷盐菌 (*Methanohalobium vestigatum*)、鱼腥藻 (*Anabaena variabilis*)、泡沫节球藻 (*Nodularia spumigena*)、念珠藻属 (*Nostoc sp.*)、极大节旋藻 (*Arthrospira maxima*)、钝顶节旋藻 (*Arthrospira platensis*)、节旋藻属 (*Arthrospira sp.*)、鞘丝藻 (*Lyngbya sp.*)、原型微鞘藻 (*Microcoleus chthonoplastes*)、颤藻属 (*Oscillatoria sp.*)、运动石孢菌 (*Petrotoga mobilis*)、非洲栖热腔菌 (*Thermosiphon africanus*)、深海单细胞蓝细菌 (*Acaryochloris marina*)、脑膜炎奈瑟氏球菌 (*Neisseria meningitidis*)，或空肠弯曲杆菌 (*Campylobacter jejuni*)。Cas9 家族成员的另外的例子在 WO 2014/131833 中描述，出于所有目的通过引用将其全文并入本文。化脓链球菌的 Cas9 (SpCas9) (分配的 SwissProt 登录号为 Q99ZW2) 是示例性的 Cas9 蛋白。来自金黄色葡萄球菌的 Cas9 (SaCas9) (分配的 UniProt 登录号为 J7RUA5) 是另一种示例性 Cas9 蛋白。来自空肠弯曲杆菌的 Cas9 (CjCas9) (分配的 UniProt 登录号为 Q0P897) 是另一种示例性 Cas9 蛋白。参见，例如，Kim et al. (2017) Nat. Comm. 8:14500，出于所有目的通过引用整体并入本文。SaCas9 小于 SpCas9，而 CjCas9 小于 SaCas9 和 SpCas9。

[0132] Cas 蛋白的另一个例子是 Cpf1 (来自普雷沃氏菌 (*Prevotella*) 和弗朗西丝氏菌 1 (*Francisella 1*) 的 CRISPR) 蛋白。Cpf1 是一种大蛋白 (约 1300 个氨基酸)，其中包含与 Cas9 的相应结构域同源的 RuvC 样核酸酶结构域以及与 Cas9 的特征性精氨酸富集簇相对应的部分。但是，Cpf1 缺少 Cas9 蛋白中存在的 HNH 核酸酶结构域，而 RuvC 样结构域在 Cpf1 序列中是连续的；这与包含 HNH 域的长插入片段的 Cas9 不同。参见，例如，Zetsche et al. (2015) Cell 163 (3):759-771，出于所有目的通过引用整体并入本文。示例性的 Cpf1 蛋白来自土拉热弗朗西斯氏菌 1 (*Francisella tularensis 1*)、新凶手土拉弗朗西斯氏菌 (*Francisella tularensis subsp. novicida*)、苏格兰普雷沃氏菌 (*Prevotella albensis*)、毛螺科菌 MC2017 1 (*Lachnospiraceae bacterium MC2017 1*)、解蛋白丁酸弧菌 (*Butyrivibrio proteoclasticus*)、佩莱格里尼菌科细菌 GW2011_GWA2_33_10 (*Peregrinibacteria*

bacterium GW2011_GWA2_33_10)、俭菌细菌GW2011_GWC2_44_17 (Parcubacteria bacterium GW2011_GWC2_44_17)、斯密斯氏互养菌属SCADC(*Smithella* sp.SCADC)、氨基酸球菌属BV3L6 (*Acidaminococcus* sp.BV3L6)、毛螺科菌MA2020 (*Lachnospiraceae* bacterium MA2020)、*Candidatus Methanoplasma termitum*、挑剔真杆菌 (*Eubacterium eligens*)、牛眼莫拉氏菌237 (*Moraxella bovoculi* 237)、稻田钩端螺旋体 (*Leptospira inadai*)、毛螺科菌ND2006 (*Lachnospiraceae* bacterium ND2006)、狗口腔卟啉单胞菌3 (*Porphyromonas crevioricanis* 3)、解糖脲普雷沃氏菌 (*Prevotella disiens*) 和猕猴卟啉单胞菌 (*Porphyromonas macacae*)。来自新凶手土拉弗朗西斯菌U112 (*Francisella novicida* U112)的Cpf1 (FnCpf1;分配的UniProt登录号为A0Q7Q2)是示例性Cpf1蛋白。

[0133] Cas蛋白可以是野生型蛋白(即自然界中存在的那些)、修饰的Cas蛋白(即Cas蛋白变体),或野生型Cas蛋白或修饰的Cas蛋白的片段。Cas蛋白也可以是相对于野生型Cas蛋白或修饰的Cas蛋白的催化活性的活性变体或片段。针对催化活性的活性变体或片段可包含与野生型或修饰的Cas蛋白或其部分的至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或更高的序列同一性,其中活性变体保留在所需切割位点切割的能力,因此保留了缺口诱导或双链断裂诱导活性。缺口诱导或双链断裂诱导活性的测定是已知的,并且通常测量Cas蛋白在包含切割位点的DNA底物上的总体活性和特异性。

[0134] 修饰的Cas蛋白的一个例子是修饰的SpCas9-HF1蛋白,其为化脓链球菌Cas9的一种高保真变体,具有设计为减少非特异性DNA接触的改变(N497A τ /R661A τ /Q695A/Q926A)。例如,Kleinstiver et al. (2016) *Nature* 529(7587):490-495,出于所有目的通过引用整体并入本文。修饰的Cas蛋白的另一个实例是修饰的eSpCas9变体(K848A τ /K1003A τ /R1060A),其设计为减少脱靶效应。参见,例如,Slaymaker et al. (2016) *Science* 351(6268):84-88,出于所有目的通过引用全文并入本文。其它SpCas9变体包括K855A和K810A/K1003A/R1060A。

[0135] Cas蛋白可以被修饰以增加或降低核酸结合亲和力、核酸结合特异性,和酶活性中的一种或多种。Cas蛋白也可以被修饰以改变蛋白的任何其它活性或性质,例如稳定性。例如,可以修饰、删除,或灭活Cas蛋白的一个或多个核酸酶结构域,或者可以截断Cas蛋白以去除对于蛋白功能非必需的结构域或对Cas蛋白的活性或特性进行优化(例如增强或降低)。

[0136] Cas蛋白可以包含至少一个核酸酶结构域,例如DNA酶结构域。例如,野生型Cpf1蛋白通常包含RuvC样结构域,其切割靶DNA的两条链,可能是二聚体构型。Cas蛋白也可以包含至少两个核酸酶结构域,例如DNA酶结构域。例如,野生型Cas9蛋白通常包含RuvC样核酸酶结构域和HNH样核酸酶结构域。RuvC和HNH结构域可以各自切割双链DNA的一条不同的链,以在DNA中产生双链断裂。参见,例如,Jinek et al. (2012) *Science* 337:816-821,出于所有目的通过引用整体并入本文。

[0137] 一个或多个或所有核酸酶结构域可以被缺失或突变,使得它们不再起作用或具有降低的核酸酶活性。例如,如果Cas9蛋白中的一个核酸酶结构域被删除或突变,则所得的Cas9蛋白可以称为切口酶,并且可以在双链的靶DNA内产生单链断裂,但不能产生双链断裂(即它可以切割互补链或非互补链,但不能同时切割两者)。如果两个核酸酶结构域都被删除或突变,则所得的Cas蛋白(例如Cas9)切割双链DNA的两条链的能力会降低(例如,无核酸

酶的 (nuclease-null) 或核酸酶失活的Cas蛋白, 或催化死亡的Cas蛋白 (dCas)。将Cas9转化为切口酶的突变的一个例子是来自化脓链球菌的Cas9的RuvC结构域中的D10A(在Cas9的位置10的天冬氨酸变为丙氨酸) 突变。同样地, 来自化脓链球菌的Cas9的HNH结构域中的H939A(在氨基酸位置839的组氨酸变为丙氨酸)、H840A(在氨基酸位置840处的组氨酸变为丙氨酸), 或N863A(在氨基酸位置N863处的天冬酰胺变为丙氨酸) 可以将Cas9转化为切口酶。将Cas9转化为切口酶的突变的其它实例包括来自嗜热链球菌的Cas9的相应突变。参见, 例如, Sapranaukas et al. (2011) *Nucleic Acids Research* 39:9275-9282和W0 2013/141680, 出于所有目的将其每一篇通过引用整体并入本文。可以使用诸如定点突变、PCR介导的突变, 或全基因合成等方法产生此类突变。可以在例如W0 2013/176772和W0 2013/142578中找到产生切口酶的其他突变的实例, 出于所有目的将其每一篇的全部内容通过引用并入本文。如果在Cas蛋白中所有核酸酶结构域均删除或突变(例如在Cas9蛋白中的两个核酸酶结构域均删除或突变), 则所得的Cas蛋白(例如Cas9) 切割双链DNA的两条链的能力将降低(例如无核酸酶的或核酸酶失活的Cas蛋白)。一个具体实例是D10A/H840A化脓链球菌Cas9双突变体或当与化脓链球菌Cas9最佳比对时来自另一物种的Cas9中相应的双突变体。另一个具体实例是D10A/N863A化脓链球菌Cas9双突变体或当与化脓链球菌Cas9最佳比对时来自另一个物种的Cas9中相应的双突变体。催化失活的Cas9蛋白 (dCas9) 的一个示例包含以下、基本上由以下组成, 或由以下组成: 与SEQ ID NO:2中列出的dCas9蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%, 或100%的同一性的氨基酸序列。

[0138] 金黄色葡萄球菌Cas9蛋白的催化结构域中的失活突变的实例也是已知的。例如, 金黄色葡萄球菌Cas9酶 (SaCas9) 可包含在位置N580处的取代(例如N580A取代) 和在位置D10处的取代(例如D10A取代), 以产生核酸酶失活的Cas蛋白。参见, 例如, W0 2016/106236, 出于所有目的通过引用将其全文并入本文。

[0139] Cpf1蛋白的催化结构域中的失活突变的实例也是已知的。对于来自新凶手土拉弗朗西斯菌U112的Cpf1蛋白 (FnCpf1)、来自氨基酸球菌属BV3L6的Cpf1蛋白 (AsCpf1)、来自毛螺科菌ND2006的Cpf1蛋白 (LbCpf1), 和来自牛眼莫拉氏菌237的Cpf1蛋白 (MbCpf1 Cpf1), 此类突变可包括AsCpf1的位置908、993, 或1263或Cpf1种间同源物的相应位置处的突变, 或LbCpf1的位置832、925、947, 或1180或Cpf1种间同源物的相应位置处的突变。此类突变可包括, 例如, AsCpf1的突变D908A、E993A, 和D1263A或在Cpf1种间同源物的相应突变, 或LbCpf1的D832A、E925A、D947A和D1180A或在Cpf1种间同源物的相应突变中的一种或多种。参见, 例如, US 2016/0208243, 出于所有目的通过引用将其全部内容并入本文。

[0140] Cas蛋白还可以作为融合蛋白可操作地连接至异源多肽。例如, 除了转录激活域之外, Cas蛋白还可与切割结构域或表观遗传修饰结构域融合。参见W0 2014/089290, 出于所有目的通过引用将其全部内容合并于此。Cas蛋白也可以与提供增加或减少的稳定性的异源多肽融合。融合结构域或异源多肽可以位于Cas蛋白的N端、C端或内部。

[0141] 作为一个实例, 可以将Cas蛋白与提供亚细胞定位的一种或多种异源多肽融合。此类异源多肽可包括, 例如, 用于靶向细胞核的一个或多个核定位信号 (NLS), 例如单分型 (monopartite) SV40 NLS和/或二分型 (bipartite) α -输入蛋白NLS, 用于靶向线粒体的线粒体定位信号, ER保留信号等。参见, 例如, Lange et al. (2007) *J. Biol. Chem.* 282:5101-

5105,其全部内容出于所有目的通过引用整体并入本文。此类亚细胞定位信号可以位于N末端、C末端,或Cas蛋白内的任何位置。NLS可以包含一段碱性氨基酸,并且可以是单分型序列或二分型序列。可选地,Cas蛋白可包含两个或更多个NLS,包括位于N末端的NLS(例如 α -输入蛋白NLS或单分型NLS)和在C末端的NLS(例如SV40 NLS或二分型NLS)。Cas蛋白还可在N末端包含两个或多个NLS和/或在C末端包含两个或多个NLS。

[0142] Cas蛋白也可以可操作地连接至细胞穿透结构域或蛋白转导结构域。例如,细胞穿透域可以源自HIV-1TAT蛋白、来自人类乙型肝炎病毒的TLM细胞穿透模体、MPG、Pep-1、VP22、来自单纯疱疹病毒的细胞穿透肽,或聚精氨酸肽序列。参见,例如,WO 2014/089290和WO 2013/176772,出于所有目的,将其每一篇均通过引用整体并入本文。细胞穿透结构域可以位于N末端、C末端或Cas蛋白内的任何位置。

[0143] 为了易于追踪或纯化,Cas蛋白也可以可操作地连接至异源多肽,例如荧光蛋白、纯化标签,或表位标签。荧光蛋白的实例包括绿色荧光蛋白(例如GFP、GFP-2、tagGFP、turboGFP、eGFP、Emerald、Azami Green、Monomeric Azami Green、CopGFP、AceGFP、ZsGreen1),黄色荧光蛋白(例如YFP、eYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1),蓝色荧光蛋白(例如eBFP、eBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-sapphire),青色荧光蛋白(例如eCFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan),红色荧光蛋白(例如mKate、mKate2、mPlum、DsRed monomer、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-Monomer、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred),橙色荧光蛋白(例如mOrange、mK0、Kusabira-Orange、Monomeric Kusabira-Orange、mTangerine、tdTomato),和任何其它合适的荧光蛋白。标签的例子包括谷胱甘肽-S-转移酶(GST)、几丁质结合蛋白(CBP)、麦芽糖结合蛋白、硫氧还蛋白(TRX)、聚(NANP)、串联亲和纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、血凝素蛋白(HA)、nus、Softag 1、Softag3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、组氨酸(His)、生物素羧基载体蛋白(BCCP),和钙调蛋白。

[0144] Cas蛋白也可以被束缚(tethered)到标记的核酸。可以通过共价相互作用或非共价相互作用实现这种束缚(即物理连接),并且束缚可以是直接的(例如通过直接融合或化学缀合,这可以通过修饰蛋白质上的半胱氨酸或赖氨酸残基或修饰内含肽基来实现),或可以通过一种或多种间插连接子(intervening linkers)或衔接体分子(例如链霉亲和素或适配体)来实现。参见,例如,Pierce et al. (2005) *Mini Rev. Med. Chem.* 5 (1):41-55; Duckworth et al. (2007) *Angew. Chem. Int. Ed. Engl.* 46 (46):8819-8822; Schaeffer和Dixon (2009) *Australian J. Chem.* 62 (10):1328-1332; Goodman et al. (2009) *Chembiochem.* 10 (9):1551-1557;和Khatwani et al. (2012) *Bioorg. Med. Chem.* 20 (14):4532-4539,出于所有目的,将其每一篇通过引用整体并入本文。合成蛋白质-核酸缀合物的非共价策略包括生物素-链霉亲和素和镍-组氨酸方法。可以通过使用多种化学方法连接适当官能化的核酸和蛋白质来合成共价蛋白质-核酸缀合物。这些化学方法中的一些涉及将寡核苷酸直接附着到蛋白质表面的氨基酸残基(例如赖氨酸胺或半胱氨酸硫醇),而其它更复杂的方案则需要蛋白质的翻译后修饰,或催化性或反应性蛋白结构域的参与。使蛋白质与核酸共价连接的方法可以包括,例如,寡核苷酸与蛋白质赖氨酸残基或半胱氨酸残基的化学交联、表达的蛋白质连接(expressed protein-ligation)、化学酶法和光适配体

(photoaptamers)的使用。可以将标记的核酸束缚在C末端、N末端,或Cas蛋白内的内部区域。在一实例中,标记的核酸束缚在Cas蛋白的C末端或N末端。同样,Cas蛋白可以束缚在5'末端、3'末端,或标记的核酸中的内部区域。即标记的核酸可以任何方向和极性束缚。例如,可以将Cas蛋白束缚在标记核酸的5'末端或3'末端。

(2) 转录激活域

[0145] 本文公开的嵌合Cas蛋白可包含一个或多个转录激活域。转录激活域包括天然存在的转录因子的区域,其可与DNA结合域(例如,与指导RNA复合的催化失活的Cas蛋白)一起通过直接接触转录机制或通过例如共激活因子的蛋白质接触转录机制来激活启动子的转录。转录激活域还包括转录因子的此类区域的功能片段或变体,以及工程化的转录激活域,该工程化的转录激活域衍生自天然的、天然存在的转录激活域,或是人工创建或合成的,以激活靶基因的转录。功能片段是当可操作地连接至合适的DNA结合域时能够激活靶基因转录的片段。功能变体是当可操作地连接至合适的DNA结合域时能够激活靶基因转录的变体

[0146] 用于本文公开的嵌合Cas蛋白中的特异转录激活域包含VP64转录激活域或其功能片段或变体。VP64是来自单纯疱疹VP16激活域的最小激活域的四聚体重复。例如,转录激活域可包含以下、基本上由以下组成,或由以下组成:与SEQ ID NO:3中列出的VP64转录激活域蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性的氨基酸序列。

[0147] 转录激活域的其他实例包括单纯疱疹病毒VP16反式激活域、VP64(单纯疱疹病毒VP16的四重串联重复)、NF- κ B p65(NF- κ B反式激活亚基p65)激活域、MyoD1反式激活结构域、HSF1反式激活结构域(来自人类热休克因子1的反式激活结构域)、RTA(EB病毒R反式激活因子激活结构域)、SET7/9反式激活结构域、p53激活结构域1、p53激活结构域2,CREB(cAMP反应元件结合蛋白)激活域、E2A激活域、NFAT(活化T细胞核因子)激活域,及其功能片段和变体。参见,例如,US2016/0298125、US 2016/0281072,和WO 2016/049258,出于所有目的将其全部内容通过引用整体并入本文。转录激活域的其他实例包括Gcn4、MLL、Rtg3、Gln3、Oaf1、Pip2、Pdr1、Pdr3、Pho4、Leu3及其功能片段和变体。参见,例如,US 2016/0298125,出于所有目的通过引用将其全部内容合并于此。转录激活域的其他实例包括Sp1、Vax、GATA4及其功能片段和变体。参见,例如,WO 2016/149484,出于所有目的通过引用将其全部内容合并于此。其他例子包括来自Oct1、Oct-2A、AP-2、CTF1、P300、CBP、PCAF、SRC1、PvALF、ERF-2、OsGAI、HALF-1、C1、AP1、ARF-5、ARF-6、ARF-7、ARF-8、CPRF1、CPRF4、MYC-RP/GP和TRAB1PC4的激活结构域及其功能片段和变体。参见,例如,US 2016/0237456、EP3045537,和WO 2011/146121,出于所有目的将其每一篇均通过引用整体并入。还已知其他合适的转录激活域。参见,例如,WO 2011/146121,出于所有目的通过引用将其全部内容合并于此。

B. 嵌合衔接体蛋白

[0148] 还提供了可以与本文其他地方公开的指导RNA结合的嵌合衔接体蛋白。本文公开的嵌合衔接体蛋白可用于dCas协同激活介质(SAM)样系统中,以增加针对靶基因内的靶序列的转录激活域的数量和多样性,以激活靶基因的转录。可以将编码嵌合衔接体蛋白的核酸基因组整合到本文其他地方公开的细胞或非人动物(例如包含基因组整合的嵌合Cas蛋白表达盒的细胞或非人动物)中,或嵌合衔接体蛋白或核酸可以使用本文其他地方公开的方法(例如LNP介导的递送或AAV介导的递送)引入此类细胞和非人动物中。

[0149] 此类嵌合衔接体蛋白包含：(a) 与指导RNA中的衔接体结合元件特异性结合的衔接体（即衔接体结构域或衔接体蛋白）；以及(b) 一个或多个异源转录激活域。例如，此类融合蛋白可包含1、2、3、4、5，或更多个转录激活域（例如，两个或更多个异源转录激活域或三个或更多个异源转录激活域）。在一个例子中，此类嵌合衔接体蛋白可包含：(a) 与指导RNA中的衔接体结合元件特异性结合的衔接体（即衔接体结构域或衔接体蛋白）；以及(b) 两个或多个转录激活域。例如，嵌合衔接体蛋白可以包含：(a) 与指导RNA中的一个或多个MS2适配体特异性结合的MS2外壳蛋白衔接体（例如指导RNA中不同位置的两个MS2适配体）；以及(b) 一个或多个（例如两个或多个）转录激活域。例如，两个转录激活域可以是p65和HSF1转录激活域或其功能片段或变体。然而，还提供了嵌合衔接体蛋白，其中转录激活域包含其它转录激活域或其功能片段或变体。

[0150] 一个或多个转录激活域可直接融合至衔接体。替代地，一个或多个转录激活域可通过接头或接头的组合或通过一个或多个另外的域连接至衔接体。同样，如果存在两个或多个转录激活域，则它们可以彼此直接融合，或者可以通过接头或接头组合或一个或多个其它结构域彼此连接。可以在这些融合蛋白中使用的接头可以包括不干扰融合蛋白功能的任何序列。示例性的接头是短的（例如2-20个氨基酸），并且通常是柔性的（例如包含具有高自由度的氨基酸，例如甘氨酸、丙氨酸，和丝氨酸）。接头的一些具体实例包括一个或多个由GGGS (SEQ ID NO:4) 或GGGGS (SEQ ID NO:5) 组成的单元，例如GGGS (SEQ ID NO:4) 的两个，三个，四个或更多个重复或GGGGS (SEQ ID NO:5) 的任何组合。也可以使用其它接头序列。

[0151] 一个或多个转录激活域和衔接体可以在嵌合衔接体蛋白内以任何顺序排列。作为一种选择，一个或多个转录激活域可以在衔接体的C末端，而该衔接体可以在一个或多个转录激活域的N末端。例如，一个或多个转录激活域可以在嵌合衔接体蛋白的C末端，且该衔接体可以在嵌合衔接体蛋白的N末端。然而，一个或多个转录激活域可以在衔接体的C末端而不在嵌合衔接体蛋白的C末端（例如如果核定位信号在嵌合衔接体蛋白的C末端）。同样地，衔接体可以在一个或多个转录激活域的N末端而不在嵌合衔接体蛋白的N末端（例如如果核定位信号在嵌合衔接体蛋白的N末端）。作为另一选择，一个或多个转录激活域可以在衔接体的N末端，而衔接体可以在一个或多个转录激活域的C末端。例如，一个或多个转录激活域可以在嵌合衔接体蛋白的N末端，而该衔接体可以在嵌合衔接体蛋白的C末端。作为另一选择，如果嵌合衔接体蛋白包含两个或更多个转录激活域，则两个或更多个转录激活域可以侧接衔接体。

[0152] 嵌合衔接体蛋白也可以可操作地连接或融合至其它异源多肽。融合或连接的异源多肽可位于嵌合衔接体蛋白内的N末端、C末端或其内部的任何位置。例如，嵌合衔接体蛋白可以进一步包含核定位信号。这种蛋白质的一个具体实例包括MS2外壳蛋白（适配体）和HSF1转录激活域，所述MS2外壳蛋白（适配体）直接或通过NLS连接p65转录激活域，该p65转录激活域位于MS2外壳蛋白（MSP）的C末端；所述HSF1转录激活域位于p65转录激活域的C末端。这样的蛋白质可以从N末端到C末端包含：MCP；核定位信号；p65转录激活域；和HSF1转录激活域。例如，嵌合衔接体蛋白可包含以下，基本上由以下组成，或由以下组成：与SEQ ID NO:6所示的MCP-p65-HSF1嵌合衔接体蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的氨基酸序列。

[0153] 嵌合衔接体蛋白也可以与提供亚细胞定位的一种或多种异源多肽融合或连接。此

类异源多肽可包括,例如,一个或多个核定位信号(NLS),例如用于靶向核的SV40 NLS和/或 α -输入蛋白NLS,用于靶向线粒体的线粒体定位信号,ER保留信号等。参见,例如,Lange et al. (2007) *J. Biol. Chem.* 282:5101-5105,出于所有目的通过引用整体并入本文。NLS可以包含,例如,一段碱性氨基酸,并且可以是单分型序列或二分型序列。可选地,嵌合衔接体蛋白包含两个或更多个NLS,包括在N末端的NLS(例如, α -输入蛋白NLS)和/或在C末端的NLS(例如SV40 NLS)。

[0154] 嵌合衔接体蛋白也可以可操作地连接至细胞穿透结构域或蛋白转导结构域。例如,细胞穿透域可以源自HIV-1TAT蛋白、来自人类乙型肝炎病毒的TLM细胞穿透模体、MPG、Pep-1、VP22、来自单纯疱疹病毒的细胞穿透肽,或聚精氨酸肽序列。参见,例如,WO 2014/089290和WO 2013/176772,出于所有目的,将其每一篇均通过引用整体并入本文。作为另一个例子,可以将嵌合衔接体蛋白融合或连接至提供增加或降低的稳定性的异源多肽。

[0155] 为了易于追踪或纯化,嵌合衔接体蛋白也可以可操作地连接至异源多肽,例如荧光蛋白、纯化标签,或表位标签。荧光蛋白的实例包括绿色荧光蛋白(例如GFP、GFP-2、tagGFP、turboGFP、eGFP、Emerald、Azami Green、Monomeric Azami Green、CopGFP、AceGFP、ZsGreen1),黄色荧光蛋白(例如YFP、eYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1),蓝色荧光蛋白(例如eBFP、eBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-sapphire),青色荧光蛋白(例如eCFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan),红色荧光蛋白(例如mKate、mKate2、mPlum、DsRed monomer、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-Monomer、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred),橙色荧光蛋白(例如mOrange、mKO、Kusabira-Orange、Monomeric Kusabira-Orange、mTangerine、tdTomato),和任何其它合适的荧光蛋白。标签的例子包括谷胱甘肽-S-转移酶(GST)、几丁质结合蛋白(CBP)、麦芽糖结合蛋白、硫氧还蛋白(TRX)、聚(NANP)、串联亲和纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、血凝素蛋白(HA)、nus、Softag 1、Softag 3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、组氨酸(His)、生物素羧基载体蛋白(BCCP),和钙调蛋白。

[0156] 嵌合衔接体蛋白也可以被束缚到标记的核酸。可以通过共价相互作用或非共价相互作用实现这种束缚(即物理连接),并且束缚可以是直接的(例如通过直接融合或化学缀合,这可以通过修饰蛋白质上的半胱氨酸或赖氨酸残基或修饰内含肽基来实现),或可以通过一种或多种间插连接子或衔接体分子(例如链霉亲和素或适配体)来实现。参见,例如,Pierce et al. (2005) *Mini Rev. Med. Chem.* 5 (1):41-55; Duckworth et al. (2007) *Angew. Chem. Int. Ed. Engl.* 46 (46):8819-8822; Schaeffer和Dixon (2009) *Australian J. Chem.* 62 (10):1328-1332; Goodman et al. (2009) *Chembiochem.* 10 (9):1551-1557; 和 Khatwani et al. (2012) *Bioorg. Med. Chem.* 20 (14):4532-4539,出于所有目的,将其每一篇通过引用整体并入本文。合成蛋白质-核酸缀合物的非共价策略包括生物素-链霉亲和素和镍-组氨酸方法。可以通过使用多种化学方法连接适当官能化的核酸和蛋白质来合成共价蛋白质-核酸缀合物。这些化学方法中的一些涉及将寡核苷酸直接附着到蛋白质表面的氨基酸残基(例如赖氨酸胺或半胱氨酸硫醇),而其它更复杂的方案则需要蛋白质的翻译后修饰,或催化性或反应性蛋白结构域的参与。使蛋白质与核酸共价连接的方法可以包括,例如,寡核苷酸与蛋白质赖氨酸残基或半胱氨酸残基的化学交联、表达的蛋白质连接、化学酶

法和光适配体的使用。可以将标记的核酸束缚在C末端、N末端,或Cas蛋白内的内部区域。同样,Cas蛋白可以束缚在5'末端、3'末端,或标记的核酸中的内部区域。即标记的核酸可以任何方向和极性束缚。

(1) 适配体

[0157] 衔接体(即衔接体结构域或衔接体蛋白)是特异性识别并结合不同序列(例如以序列特异性方式结合不同的DNA和/或RNA序列,例如适配体)的核酸结合结构域(例如DNA结合结构域和/或RNA结合结构域)。适配体包括通过其采取特定三维构象的能力可以以高亲和力和特异性结合至靶分子的核酸。这样的衔接体可以与例如特定的RNA序列和二级结构结合。可以将这些序列(即衔接体结合元件)工程化为指导RNA。例如,可以将MS2适配体工程化成指导RNA,以特异性结合MS2外壳蛋白(MCP)。例如,衔接体可包含以下、基本上由以下组成,或由以下组成:与SEQ ID NO:7中列出的MCP序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性的氨基酸序列。

[0158] 衔接体和靶标的一些具体实例包括存在于噬菌体外壳蛋白多样性内的RNA结合蛋白/适配体组合。例如,可以使用以下衔接体蛋白或其功能片段或变体:MS2外壳蛋白(MCP)、PP7、Qβ、F2、GA、fr、JP501、M12、R17、BZ13、JP34、JP500、KU1、M11、MX1、TW18、VK、SP、FI、ID2、NL95、TW19、AP205、ΦCb5、ΦCb8r、ΦCb12r、ΦCb23r、7s和PRR1。参见,例如,WO 2016/049258,出于所有目的通过引用将其全文并入本文。衔接体蛋白的功能片段或功能变体保留结合特定的衔接体结合元件的能力(例如,以序列特异性方式与特定的衔接体结合序列结合的能力)。例如,可以使用PP7假单胞菌(*Pseudomonas*)噬菌体外壳蛋白变体,其中氨基酸68-69突变为SG,而氨基酸70-75从野生型蛋白中删除。参见,例如,Wu et al. (2012) *Biophys J* 102(12):2936-2944和Chao et al. (2007) *Nature Structural&Molecular Biology* 15(1):103-105,出于所有目的,将每一篇均通过引用整体并入本文。同样,可以使用MCP变体,例如N55K突变体。参见,例如,Spingola和Peabody (1994) *J Biol Chem* 269(12):9006-9010,出于所有目的通过引用整体并入本文。

[0159] 可以使用的衔接体蛋白的其它例子包括核糖核酸内切酶Csy4或λN蛋白的全部或部分(例如DNA结合形式)。参见,例如,US 2016/0312198,出于所有目的通过引用将其全部内容合并于此。

(2) 转录激活域

[0160] 本文公开的嵌合衔接体蛋白包含一个或多个转录激活域。这样的转录激活域可以是天然存在的转录激活域,可以是天然存在的转录激活域的功能片段或功能变体,或者可以是工程的或合成的转录激活域。可以使用的转录激活域包括本文其它地方描述为用于嵌合Cas蛋白的那些。

[0161] 用于本文公开的嵌合衔接体蛋白的具体的转录激活域包含p65和/或HSF1转录激活域或其功能片段或变体。HSF1转录激活域可以是人热休克因子1(HSF1)的转录激活域。p65转录激活域可以是转录因子p65的转录激活域(也称为由RELA基因编码的核因子NF-κ-Bp65亚基)。作为一个实例,转录激活域可包含以下、基本上由以下组成,或由以下组成:与SEQ ID NO:8中列出的p65转录激活域蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性的氨基酸序列。作为另一个例子,转录激活域可以包含以下、基本上由以下组成,或由以下组成:与SEQ ID NO:9中列出的HSF1

转录激活域蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%，或100%的同一性的氨基酸序列。

C. 指导RNA和指导RNA阵列

[0162] 还提供了可与本文其它各处公开的嵌合Cas蛋白和嵌合衔接体蛋白结合以激活靶基因转录的指导RNA或指导RNA阵列。如本文其它地方所公开的，可以将编码指导RNA的核酸基因组整合到细胞或非人动物（例如SAM就绪细胞（SAM-ready cell）或非人动物）中，或者可以使用本文其它地方公开的方法（例如LNP介导的递送或AAV介导的递送）将指导RNA或核酸引入此类细胞和非人动物中。如本文其它地方所公开的，可以选择递送方法以提供重组酶的组织特异性递送。

[0163] 编码指导RNA或指导RNA阵列的核酸可以编码一种或多种指导RNA（或如果将指导RNA引入细胞或非人动物中，则可以引入一种或多种指导RNA）。例如，可以编码或引入2个或更多个、3个或更多个、4个或更多个，或5个或更多个指导RNA。每个指导RNA编码序列可以可操作地连接至相同的启动子（例如U6启动子）或不同的启动子（例如，每个指导RNA编码序列可操作地连接至其自身的U6启动子）。两个或更多个指导RNA可以在单个靶基因中靶向不同的靶序列。例如，2个或更多、3个或更多、4个或更多，或5个或更多的指导RNA可以在单个靶基因中分别靶向不同的靶序列。类似地，指导RNA可以靶向多个靶基因（例如2个或更多个、3个或更多个、4个或更多个，或5个或更多个靶基因）。指导RNA靶序列的实例在本文其它地方公开。

(1) 指导RNA

[0164] “指导RNA”或“gRNA”是与Cas蛋白（例如Cas9蛋白）结合并且将Cas蛋白靶向靶DNA内的特定位置的RNA分子。指导RNA可以包含两个段（segment）：“DNA靶向段”和“蛋白质结合段”。“段”包括分子的一部分或区域，例如RNA中核苷酸的连续链。某些gRNA（例如Cas9的gRNA）可以包含两个单独的RNA分子：“激活剂RNA”（例如tracrRNA）和“靶标RNA（targeterRNA）”（例如CRISPR RNA或crRNA）。其它gRNA是单RNA分子（单RNA多核苷酸），也可以称为“单分子gRNA”、“单指导RNA”，或“sgRNA”。参见，例如，WO 2013/176772、WO 2014/065596、WO 2014/089290、WO 2014/093622、WO 2014/099750、WO 2013/142578，和WO 2014/131833，在出于所有目的此通过引用将每一篇的全部内容并入本文。例如，对于Cas9，单指导RNA可以包含与tracrRNA融合的crRNA（例如通过接头）。例如，对于Cpf1，仅需要crRNA即可实现与靶序列的结合。术语“指导RNA”和“gRNA”包括双分子（即模块）gRNA和单分子gRNA。

[0165] 示例性的两分子gRNA包含crRNA样（“CRISPR RNA”或“靶标RNA”或“crRNA”或“crRNA重复”）分子和相应的tracrRNA样（“反式作用CRISPR基因”或“激活剂RNA”或“tracrRNA”）分子。crRNA既包含gRNA的DNA靶向段（单链），又包含形成gRNA的蛋白质结合片段的dsRNA双链体一半的核苷酸链。位于DNA靶向片段下游（3'）的crRNA尾巴的例子包含以下，基本上由以下组成，或由以下组成：GUUUUAGAGCUAUGCU（SEQ ID NO:10）。可将本文公开的任何靶向DNA的片段连接至SEQ ID NO:10的5'末端以形成crRNA。

[0166] 相应的tracrRNA（活化剂RNA）包含一段核苷酸，其形成gRNA的蛋白质结合段的dsRNA双链体的另一半。crRNA的一段核苷酸与tracrRNA的一段核苷酸互补并杂交，以形成gRNA蛋白质结合域的dsRNA双链体。这样，每个crRNA可以称为具有对应的tracrRNA。tracrRNA序列的实例包含以下，基本上由以下组成或由以下组成：AGCAUAGCAAGUAAAAUA

GGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCUUU (SEQ ID NO:11)。

[0167] 在需要crRNA和tracrRNA的系统中,crRNA和相应的tracrRNA杂交以形成gRNA。在仅需要crRNA的系统中,crRNA可以是gRNA。crRNA还提供了与靶DNA互补链杂交的单链DNA靶向片段。如果用于细胞内的修饰,则可以将给定的crRNA或tracrRNA分子的确切序列设计为对要使用RNA分子的物种具有特异性。参见,例如,Mali et al. (2013) Science 339:823-826;Jinek et al. (2012) Science 337:816-821;Hwanget al. (2013) Nat. Biotechnol. 31: 227-229;Jiang et al. (2013) Nat. Biotechnol. 31:233-239;和Cong et al. (2013) Science 339:819-823,出于所有目的,将每一篇均通过引用整体并入本文。

[0168] 如下文所更详细描述,给定的gRNA的DNA靶向段(crRNA)包含与靶DNA的互补链上的序列互补的核苷酸序列。gRNA的DNA靶向片段通过杂交(即碱基配对),以序列特异性方式与靶DNA相互作用。这样,DNA靶向段的核苷酸序列可以变化,并确定gRNA和靶DNA将与之相互作用的靶DNA内的位置。可以修饰对象(subject)gRNA的DNA靶向段,以与靶DNA内的任何所需序列杂交。取决于CRISPR/Cas系统和生物体,天然存在的crRNA是不同的,但通常包含长度在21至72个核苷酸之间的靶向片段,其侧接长度在21至46个核苷酸之间的两个直接重复序列(DR)(参见,例如,WO 2014/131833,出于所有目的通过引用整体并入本文)。对于化脓链球菌,DR的长度为36个核苷酸,而靶向段的长度为30个核苷酸。位于3'的DR与相应的tracrRNA互补并与之杂交,后者又与Cas蛋白结合。

[0169] DNA靶向段可以具有例如至少约12、15、17、18、19、20、25、30、35或40个核苷酸的长度。此类DNA靶向段的长度可为,例如,约12至约100、约12至约80、约12至约50、约12至约40、约12至约30、约12至约25,或约12至约20个核苷酸。例如,DNA靶向段可以为约15至约25个核苷酸(例如约17至约20个核苷酸,或约17、18、19或20个核苷酸)。参见,例如,US 2016/0024523,出于所有目的通过引用将其全部内容合并于此。对于来自化脓链球菌的Cas9,典型的DNA靶向段的长度为16至20个核苷酸之间或17至20个核苷酸之间。对于来自金黄色葡萄球菌的Cas9,典型的DNA靶向段的长度在21至23个核苷酸之间。对于Cpf1,典型的DNA靶向段的长度为至少16个核苷酸或至少18个核苷酸。

[0170] TracrRNA可以是任何形式(例如全长tracrRNA或活性部分tracrRNA(active partial tracrRNAs)),并且具有不同的长度。它们可以包括初级转录物或加工后的形式。例如,tracrRNA(作为单指导RNA的一部分,或作为两分子gRNA的一部分的另外的分子)可以包含以下,基本上由以下组成,或由以下组成:野生型tracrRNA序列的全部或一部分(例如野生型tracrRNA序列的约20个、26个、32个、45个、48个、54个、63个、67个、85个或更多个核苷酸)。来自化脓链球菌的野生型tracrRNA序列的实例包括171个核苷酸、89个核苷酸、75个核苷酸,和65个核苷酸的版本。参见,例如,Deltcheva et al. (2011) Nature 471:602-607; WO 2014/093661,出于所有目的,其每一篇均通过引用整体并入本文。单指导RNA(sgRNA)中的tracrRNA的例子包括在+48、+54、+67,和+85版本的sgRNA中发现的tracrRNA片段,其中“+n”表示在sgRNA中包含野生型tracrRNA的最多+n核苷酸。参见US 8,697,359,出于所有目的将其全部内容并入本文作为参考。

[0171] 指导RNA的DNA靶向段与靶DNA的互补链之间的互补性百分比可以为至少60%(例如,至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少98%、至少99%,或100%)。在约20个连续核苷酸上,DNA靶向段与靶DNA的互补链之间的

互补性百分比可以为至少60%。例如,DNA靶向段与靶DNA互补链之间的互补性百分比可以是靶DNA互补链5'端的14个连续核苷酸的100%,而剩下的部分则低至0%。在这种情况下,DNA靶向段可以被认为是14个核苷酸的长度。作为另一个例子,DNA靶向段与靶DNA的互补链之间的互补性百分比可以是靶DNA互补链的5'端的7个连续核苷酸的100%,而剩下的部分则低至0%。在这种情况下,DNA靶向段可被认为是7个核苷酸的长度。在一些指导RNA中,DNA靶向段内的至少17个核苷酸与靶DNA的互补链互补。例如,DNA靶向段的长度可以是20个核苷酸,并且可以与靶DNA的互补链包含1、2或3个错配。在一个例子中,错配不与互补链的对应于原间隔子相邻基序(PAM)序列的区域(即,PAM序列的反向互补序列)相邻(例如,错配在指导RNA的DNA靶向段的5'末端,或错配距离对应于PAM序列的互补链的区域至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18或19个碱基对)。

[0172] gRNA的蛋白质结合段可包含两个互补的核苷酸片段。蛋白结合段的互补核苷酸杂交形成双链RNA双链体(dsRNA)。对象gRNA的蛋白质结合段与Cas蛋白相互作用,并且gRNA通过DNA靶向段将结合的Cas蛋白导向靶DNA内的特定核苷酸序列。

[0173] 单指导RNA可以包含DNA靶向段和支架序列(即指导RNA的蛋白质结合序列或Cas结合序列)。例如,此类指导RNA可以具有连接3'支架序列的5'DNA靶向段。示例性的支架序列包含以下,基本上由以下组成或由以下组成:

GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCG
AGUCGGUGCU(版本1;SEQ ID NO:12);

GUUGGAACCAUUCAAAACAGCAUAGCAAGUAAAAUAAGGCUAGU CCGUUAUCAACUUGAAAAAGUG
GCACCGAGUCGGUGC(版本2;SEQ ID NO:13);

G U U U U A G A G C U A G A A A U A G C A A G U U A A A A U A A G G C U A G U C C G U U A
UCAACUUGAAAAAGUGGCACCGAGUCGGUGC(版本3;SEQ ID NO:14);和

GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUAAAAUAAGGC UAGUCCGUUAUCAACUUGAAAA
AGUGGCACCGAGUCGGUGC(版本4;SEQ ID NO:15)。靶向本文公开的任何指导RNA靶序列的指导RNA可以包括,例如,与指导RNA的3'末端上的任何示例性指导RNA支架序列融合的指导RNA的5'末端上的DNA靶向段。也就是说,本文公开的任何DNA靶向段都可以连接至上述任何一个支架序列的5'末端,以形成单指导RNA(嵌合指导RNA)。

[0174] 指导RNA可以包括提供其它所需特征的修饰或序列(例如修饰的或调节的稳定性;亚细胞靶向;荧光标记的跟踪;蛋白质或蛋白质复合物的结合位点等)。此类修饰的示例包括,例如,5'帽(例如7-甲基鸟苷帽(m7G));3'聚腺苷酸尾(即3'聚(A)尾);核糖开关序列(例如允许调节的稳定性和/或调节的蛋白质和/或蛋白质复合物的可及性);稳定性控制序列;形成dsRNA双链体的序列(即发夹);将RNA靶向亚细胞位置的修饰或序列(例如核、线粒体、叶绿体等);提供跟踪的修饰或序列(例如与荧光分子直接缀合、与有助于荧光检测的模体缀合、允许荧光检测的序列等);提供蛋白质结合位点的修饰或序列(例如,作用于DNA的蛋白质,例如转录激活剂);及其组合。修饰的其它例子包括工程化的茎环双链体结构、工程化的凸起(bulge)区域、茎环双链体结构的3'末端的工程化的发夹,或其任何组合。参见,例如,US 2015/0376586,出于所有目的通过引用将其全部内容并入本文。凸起可以由crRNA样区域和最小tracrRNA样区域组成的双链体中的核苷酸的不成对区域。凸起可在双链体的一侧包含未配对的5'-XXX-3',其中X是任何嘌呤,Y可以是可与相反链上的核苷酸形成摆

动对的核苷酸,以及位于双链体的另一侧的未配对的核苷酸区域。

[0175] 未修饰的核酸可能易于降解。外源核酸也可以诱导先天免疫应答。修饰可以帮助引入稳定性并降低免疫原性。指导RNA可以包含修饰的核苷和修饰的核苷酸,包括例如以下的一种或多种:(1)改变或替换一种或两种非连接的磷酸氧和/或一种或多种在磷酸二酯主链连接中的连接的磷酸氧;(2)改变或替换核糖的成分,例如改变或替换核糖上的2'羟基;(3)用去磷接头取代磷酸模体;(4)修饰或替换天然存在的核碱基;(5)取代或修饰核糖磷酸主链;(6)修饰寡核苷酸的3'末端或5'末端(例如末端磷酸基的去除、修饰或替换,或模体的缀合);以及(7)糖的修饰。其它可能的指导RNA修饰包括尿嘧啶或聚尿嘧啶束(poly-uracil tracts)的修饰或替换。参见,例如,WO 2015/048577和US 2016/0237455,出于所有目的,将每一篇通过引用整体并入本文。可以对Cas编码核酸如Cas mRNA进行类似的修饰。例如,可以通过使用同义密码子去除尿苷来修饰Cas mRNA。

[0176] 作为一个实例,指导RNA的5'或3'末端的核苷酸可以包括硫代磷酸酯(phosphorothioate)键(例如,碱基可以具有为硫代磷酸酯基团的修饰的磷酸基)。例如,指导RNA可以在指导RNA的5'或3'末端的2,3,或4个末端核苷酸之间包含硫代磷酸酯键。作为另一个例子,指导RNA的5'和/或3'末端的核苷酸可以具有2'-O-甲基修饰。例如,指导RNA可以在指导RNA的5'和/或3'末端(例如5'末端)的2,3或4个末端核苷酸处包含2'-O-甲基修饰。参见,例如,WO 2017/173054 A1和Finn et al. (2018) Cell Reports 22:1-9,出于所有目的,通过引用将其全文并入本文。

[0177] 在一些指导RNA(例如,单指导RNA)中,通过插入与一个或多个衔接体(即衔接蛋白或结构域)结合的不同RNA序列来修饰指导RNA的至少一个环(例如两个环)。此类衔接体蛋白可用于进一步募集一个或多个异源功能结构域,例如转录激活域。包含此类衔接体蛋白(即嵌合衔接体蛋白)的融合蛋白的实例在本文其它地方公开。例如,MS2结合环ggccAACAU GAGGAUCACCCAUGUCUGCAGggcc (SEQ ID NO:16)可以取代SEQ ID NO:12或SEQ ID NO:14所示的sgRNA支架(主链)的核苷酸+13至+16和核苷酸+53至+56,或WO 2016/049258和Koner mann et al. (2015) Nature 517(7536):583-588所描述的化脓链球菌CRISPR/Cas9系统的sgRNA骨架,出于所有目的,将每一篇均通过引用整体并入本文。参见,例如,图7。本文所使用的指导RNA编号是指在指导RNA支架序列(即指导RNA的DNA靶向段的下游的序列)中的核苷酸编号。例如,指导RNA支架的第一个核苷酸为+1,支架的第二个核苷酸为+2,依此类推。对应于SEQ ID NO:12或SEQ ID NO:14中核苷酸+13至+16的残基是跨越SEQ ID NO:12或SEQ ID NO:14中核苷酸+9至+21的区域中的环序列;在本文中将该区域称为四环。对应于SEQ ID NO:12或SEQ ID NO:14中核苷酸+53至+56的残基是跨越SEQ ID NO:12或SEQ ID NO:14中核苷酸+48至+61的区域中的环序列;在本文中将该区域称为茎环2。SEQ ID NO:12或SEQ ID NO:14中的其它茎环序列包括茎环1(核苷酸+33至+41)和茎环3(核苷酸+63至+75)。所得结构是sgRNA支架,其中四环和茎环2序列中的每一个均已被MS2结合环取代。四环和茎环2从Cas9蛋白突出,使得添加MS2结合环不会干扰任何Cas9残基。此外,四环和茎环2位点与DNA的接近性表明,定位于这些位置可能会导致DNA与任何募集的蛋白质(例如转录激活因子)之间的高度的相互作用。因此,在一些sgRNA中,对应于SEQ ID NO:12或SEQ ID NO:14所示的指导RNA支架的+13至+16的核苷酸和/或对应于+53至+56的核苷酸或与这些支架/主链中的任意一个最佳比对时的相应的残基被能够结合一个或多个衔接体蛋白或结构域的独特RNA序

列取代。替代地或另外地,可以将衔接体结合序列添加到指导RNA的5'末端或3'末端。在四环和茎环2区域中包含MS2结合环的示例性指导RNA支架可以包含以下,基本上由以下组成,或由以下组成:SEQ ID NO:40所示的序列。示例性的在四环和茎环2区域中包含MS2结合环的示例性通用指导RNA可以包含以下,基本上由以下组成,或由以下组成:SEQ ID NO:63所示的序列。

[0178] 指导RNA可以以任何形式提供。例如,可以以两个分子(分离的crRNA和tracrRNA)或一个分子(sgRNA)的RNA的形式提供gRNA,并且可选地以与Cas蛋白的复合物的形式提供。gRNA也可以以编码gRNA的DNA的形式提供。编码gRNA的DNA可以编码单个RNA分子(sgRNA)或分离的RNA分子(例如分离的crRNA和tracrRNA)。在后一种情况下,编码gRNA的DNA可以提供为一个DNA分子或作为分别编码crRNA和tracrRNA的分离的DNA分子。

[0179] 当以DNA形式提供gRNA时,该gRNA可以在细胞中瞬时地、条件地,或组成地表达。可以将编码gRNA的DNA稳定整合到细胞基因组中,并可操作地连接在细胞中活跃的启动子。替代地,可将编码gRNA的DNA可操作地连接至表达构建体中的启动子。例如,编码gRNA的DNA可以在包含异源核酸的载体中。可用于此类表达构建体的启动子包括在例如真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、多能细胞、胚胎干(ES)细胞、成年干细胞、发育受限的祖细胞、诱导性多能干(iPS)细胞,或单细胞期胚胎的一种或多种中具有活性的启动子。这样的启动子可以是,例如,条件启动子、诱导型启动子、组成型启动子,或组织特异性启动子。这样的启动子也可以是,例如,双向启动子。合适的启动子的具体实例包括RNA聚合酶III启动子,例如人U6启动子、大鼠U6聚合酶III启动子,或小鼠U6聚合酶III启动子。

[0180] 替代地,可以通过各种其它方法制备gRNA。例如,可以使用例如T7 RNA聚合酶,通过体外转录来制备gRNA(参见,例如,WO

2014/089290和WO 2014/065596,出于所有目的将其每一篇的全部内容通过引用并入本文)。指导RNA也可以是通过化学合成制备的、合成产生的分子。

[0181] 指导RNA(或编码指导RNA的核酸)可以在包含一个或多个指导RNA(例如1、2、3、4或更多个指导RNA)和增加指导RNA稳定性(例如,在给定的储存条件下(例如-20°C、4°C,或环境温度)延长降解产物的含量低于阈值(例如低于起始核酸或蛋白质的重量的0.5%)的时间;或增加体内稳定性)的载体的组合物中。此类载体的非限制性实例包括聚(乳酸)(PLA)微球、聚(D,L-乳酸-甘醇酸)(poly(D,L-lactic-coglycolic-acid), PLGA)微球、脂质体、胶束、反胶束、脂质卷(lipid cochleates),和脂质微管。这样的组合物可以进一步包含Cas蛋白,例如Cas9蛋白,或编码Cas蛋白的核酸。

(2) 指导RNA靶序列

[0182] 指导RNA的靶DNA包括存在于DNA中的核酸序列,在存在足够的结合条件的前提下,gRNA的DNA靶向段将与该核酸序列结合。合适的DNA/RNA结合条件包括细胞中通常存在的生理条件。其它合适的DNA/RNA结合条件(例如,无细胞系统中的条件)是本领域已知的(参见,例如,Molecular Cloning:A Laboratory Manual,3rd Ed. (Sambrook et al., Harbor Laboratory Press 2001),出于所有目的通过引用整体并入本文)。与gRNA互补并杂交的靶DNA链可以称为“互补链”,而与“互补链”互补的靶DNA链(因此与Cas蛋白或gRNA不互补)称为“非互补链”或“模板链”。

[0183] 靶DNA既包括与指导RNA杂交的互补链上的序列,又包括非互补链上的相应序列(例如,与原间隔子相邻基序(PAM)相邻)。本文所使用的术语“指导RNA靶序列”具体是指非互补链上的序列,该序列对应与指导RNA在互补链上杂交的序列(即其反向互补)。即,指导RNA靶序列是指与PAM相邻的非互补链上的序列(例如对于Cas9,是PAM的上游或5')。指导RNA靶序列等同于指导RNA的DNA靶向段,但是胸腺嘧啶而非尿嘧啶。作为一个例子,SpCas9酶的指导RNA靶序列可以指非互补链上5'-NGG-3' PAM上游的序列。指导RNA被设计成与靶DNA的互补链具有互补性,其中指导RNA的DNA靶向段与靶DNA的互补链之间的杂交促进了CRISPR复合物的形成。如果存在足够的互补性以引起杂交并促进CRISPR复合物的形成,则不一定需要完全互补。如果在本文中指导RNA称为靶向指导RNA靶序列,则是指该指导RNA与靶DNA的互补链序列杂交,该互补链序列是非互补序列上的指导RNA靶序列的反向互补。

[0184] 靶DNA或指导RNA靶序列可以包含任何多核苷酸,并且可以位于例如细胞(例如线粒体或叶绿体)的细胞核或细胞质中或细胞的细胞器内。靶DNA或指导RNA靶序列可以是细胞内源或外源的任何核酸序列。指导RNA靶序列可以是编码基因产物(例如蛋白质)的序列或非编码序列(例如调节序列),或者可以包括两者。

[0185] 优选地,靶序列与基因的转录起始位点相邻。例如,靶序列可以在转录起始位点的1000、900、800、700、600、500、400、300、200、190、180、170、160、150、140、130、120、110、100、90、80、70、60、50、40、30、20、10、5或1个碱基对之内,在转录起始位点上游的1000、900、800、700、600、500、400、300、200、190、180、170、160、150、140、130、120、110、100、90、80、70、60、50、40、30、20、10、5或1个碱基对之内,或在转录起始位点下游的1000、900、800、700、600、500、400、300、200、190、180、170、160、150、140、130、120、110、100、90、80、70、60、50、40、30、20、10、5或1个碱基对之内。可选地,靶序列在转录起始位点上游的200个碱基对至转录起始位点下游的1个碱基对的区域内(-200至+1)。

[0186] 靶序列可以在希望被靶向转录激活的任何基因内。在某些情况下,靶基因可以是非表达的基因或表达较弱的基因(例如,仅在背景以上最低表达的基因,例如1.1倍、1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍,或2倍)。靶基因也可以是与对照基因相比以低水平表达的基因。靶基因也可以是表观遗传沉默的。术语“表观遗传沉默的(epigenetically silenced)”是指未转录的基因或相对于对样品(例如相应的对照细胞,例如正常细胞)中的基因转录水平而言,由于除遗传变化(例如突变)以外的其它机制而在降低的水平上转录的基因。基因沉默的表观遗传机制是众所周知的,并且包括,例如,基因5'调控区的CpG岛中CpG二核苷酸的超甲基化,以及由于例如组蛋白乙酰化导致的染色质结构变化,从而减少或抑制基因转录。

[0187] 靶基因可以包括在特定器官或组织(例如肝脏)中表达的基因。靶基因可以包括疾病相关基因。疾病相关基因是指与非疾病对照的组织或细胞相比,在源自受疾病影响的组织的细胞中以异常水平或异常形式产生转录或翻译产物的任何基因。它可能是一个以异常高水平表达的基因,其中表达的改变与疾病的发生和/或进展相关。疾病相关基因还指具有引起疾病病因的突变或遗传变异的基因。转录或翻译的产物可能是已知的或未知的,并且可能处于正常或异常水平。例如,靶基因可以是与蛋白质聚集疾病和失调(例如阿尔茨海默氏病、帕金森氏病、亨廷顿氏病、肌萎缩性侧索硬化、朊病毒疾病,和淀粉样变(amyloidoses),例如甲状腺素视黄质运载蛋白淀粉样变)相关的基因(例如Ttr)。靶基因也

可以是涉及与疾病或病症(例如高胆固醇血症或动脉粥样硬化)相关的途径的基因,或者当过表达时可以模拟此类疾病或病症的基因。靶基因也可以是在一种或多种类型的癌症中表达或过表达的基因。参见,例如,Santarius et al.(2010)Nat.Rev.Cancer 10(1):59-64,出于所有目的通过引用整体并入本文。

[0188] 这种靶基因的一个具体实例是Ttr基因。可选地,Ttr基因可以包含病原性突变(例如,引起淀粉样变的突变)。此类突变的例子提供于例如WO 2018/007871,出于所有目的通过引用将其全文并入本文。示例性的人TTR蛋白和示例性的人TTR基因分别由UniProt ID P02766和Entrez基因ID 7276识别。示例性的小鼠TTR蛋白和示例性的小鼠Ttr基因分别由UniProt ID P07309和Entrez基因ID 22139识别。运甲状腺素蛋白(TTR)是一种在血清和脑脊髓液中发现的蛋白质,其携带甲状腺激素和视黄醇结合蛋白至视黄醇。肝脏将TTR分泌到血液中,而脉络丛将TTR分泌到脑脊液中。TTR也在视网膜色素上皮中产生并分泌到玻璃体中。错折叠和聚集的TTR聚集在淀粉样变疾病,包括系统性老年性淀粉样变(SSA)、家族性淀粉样变多发性神经病(FAP),和家族性淀粉样心肌病(FAC),的多个组织和器官中。运甲状腺素蛋白(TTR)是一种127个氨基酸,55kDa的血清和脑脊液转运蛋白,主要由肝脏合成,但也由脉络丛产生。它也被称为前白蛋白、甲状腺素结合前白蛋白、ATTR、TBPA、CTS、CTS1、HEL111、HsT2651,和PALB。在其天然状态下,TTR以四聚体形式存在。在纯合子中,纯四聚体包含相同的、127个氨基酸的、富含 β -折叠的亚基。在杂合子中,TTR四聚体可以由变异体和/或野生型亚基组成,通常以统计学方式组合。TTR负责在血清和脑脊髓液中携带甲状腺素(T4)和视黄醇结合的RBP(视黄醇结合蛋白)。小鼠Ttr基因中的指导RNA靶序列(不包括PAM)的实例分别在SEQ ID NO:34、35和36中列出。SEQ ID NO:34位于Ttr转录起始位点的-63(基因组坐标:build mm10,chr18,+链,20665187-20665209),SEQ ID NO:35位于Ttr转录起始位点的-134(基因组坐标:build mm10,chr18,+链,20665116-20665138),而SEQ ID NO:36位于Ttr转录起始位点的-112(基因组坐标:build mm10,chr18,+链,20665138-

20665160)。与分别在SEQ ID NO:34、35和36中列出的指导RNA靶序列相对应的指导RNA的DNA靶向段分别在SEQ ID NO:41、42和43中列出。包含这些DNA靶向段的单指导RNA的实例分别在SEQ ID NO:37、38和39中列出。

[0189] 靶基因的其它实例是前蛋白转化酶枯草杆菌蛋白酶/kexin 9型(PCSK9)和低密度脂蛋白(LDL)受体(LDLR)。示例性人PCSK9蛋白和示例性人PCSK9基因分别通过UniProt ID Q8NBP7和Entrez基因ID 255738识别。示例性的小鼠PCSK9蛋白和示例性的小鼠Pcsk9基因分别由UniProt ID Q80W65和Entrez基因ID 100102鉴定。示例性的人LDLR蛋白和示例性的人LDLR基因分别由UniProt ID P01130和Entrez基因ID 3949识别。示例性的小鼠LDLR蛋白和示例性的小鼠Ldlr基因分别由UniProt ID P35951和Entrez基因ID 16835识别。

[0190] LDLR介导富含胆固醇的LDL的内吞作用,从而维持血浆LDL水平。这发生在所有带核细胞中,但主要发生在肝脏中;肝脏从循环中清除了约70%的LDL。LDL受体结合并开始将LDL颗粒从细胞外液摄入细胞,从而降低了LDL颗粒的浓度。当LDL与LDLR结合时,它会诱导内体中LDLR-LDL复合物的内在化。内体环境的酸性诱导LDLR采用发夹结构。构象变化导致LDLR释放其LDL配体,并且受体被循环回到质膜。在人类中,由于LDL-胆固醇在血液中的积累,LDL直接参与动脉粥样硬化的发展,这是导致大多数心血管疾病的过程。

[0191] 当PCSK9结合至LDLR时,PCSK9阻止受体-配体复合物的构象变化。这种抑制作用将

LDLR重定向至溶酶体。PCSK9在胆固醇稳态中起主要调节作用,主要是通过降低质膜上的LDLR水平。降低的LDLR水平会导致LDL颗粒的代谢下降,从而导致高胆固醇血症。如果PCSK9被阻断,则更多的LDLR被回收并存在于细胞表面,以从细胞外液中去除LDL颗粒。因此,阻断PCSK9可以降低血液LDL颗粒浓度,而增加PCSK9的表达可以增加血液LDL颗粒浓度。因此,如本文其它地方所述的激活Pcsk9的表达可用于模拟高胆固醇血症(血液中胆固醇水平高),其可导致动脉粥样硬化(动脉硬化)。

[0192] 小鼠Ldlr基因中的指导RNA靶序列(不包括PAM)的例子分别在SEQ ID NO:75、76和77中列出。分别在SEQ ID NO:81、82和83中列出分别对应于SEQ ID NO:75、76和77中列出的指导RNA靶序列的指导RNA的DNA靶向段。包含这些DNA靶向段的单指导RNA的实例分别在SEQ ID NO:78、79和80中列出。

[0193] 小鼠Pcsk9基因中的指导RNA靶序列(不包括PAM)的实例分别在SEQ ID NO:89、90和91中列出。分别在SEQ ID NOS:95、96和97中列出与分别在SEQ ID NO:89、90和91中列出的指导RNA靶序列相对应的指导RNA的DNA靶向段。包含这些DNA靶向段的单指导RNA的实例分别在SEQ ID NO:92、93和94中列出。

[0194] Cas蛋白对靶DNA的位点特异性结合和切割可发生在由(i)指导RNA和靶DNA的互补链之间的碱基配对互补性和(ii)位于靶DNA的非互补链中,被称为原间隔子相邻基序(PAM)的短基序。PAM可以侧接指导RNA靶序列。可选地,指导RNA靶序列可以在3'末端上侧接PAM(例如对于Cas9)。替代地,指导RNA靶序列可以在5'末端上侧接PAM(例如对于Cpf1)。例如,Cas蛋白的切割位点可以是(例如在指导RNA靶序列内)PAM序列上游或下游的约1至约10或约2至约5个碱基对(例如3个碱基对)。在SpCas9中,PAM序列(即非互补链上)可以是5'-N₁GG-3',其中N₁是任何DNA核苷酸,并且PAM在靶DNA的非互补链上紧邻指导RNA靶序列的3'。这样,在互补链上与PAM相对应的序列(即反向互补序列)将是5'-CCN₂-3',其中N₂是任何DNA核苷酸,并且紧邻序列的5',指导RNA的DNA靶向段与该序列在靶DNA的互补链上杂交。在一些这样的情况下,N₁和N₂可以是互补的,并且N₁-N₂碱基对可以是任何碱基对(例如N₁=C和N₂=G;N₁=G和N₂=C;N₁=A和N₂=T;或N₁=T,和N₂=A)。对于来自金黄色葡萄球菌的Cas9,PAM可以是NNGRRT或NNGRR,其中N可以是A、G、C,或T,而R可以是G或A。对于来自空肠弯曲杆菌的Cas9,PAM可以是,例如,NNNNACAC或NNNNRYAC,其中N可以是A、G、C,或T,R可以是G或A。在某些情况下(例如对于FnCpf1),PAM序列可以位于5'末端的上游并具有序列5'-TTN-3'。

[0195] 指导RNA靶序列的例子是紧接在由SpCas9蛋白识别的NGG基序之前的20个核苷酸的DNA序列。例如,指导RNA靶序列加PAM的两个例子是GN₁₉NGG(SEQ ID NO:17,其中N=A、T、C、或G)或N₂₀NGG(SEQ ID NO:18,其中N=A、T、C、或G)。参见,例如,WO 2014/165825,出于所有目的通过引用将其全部内容并入本文。5'末端的鸟嘌呤可促进细胞中RNA聚合酶的转录。指导RNA靶序列加PAM的其它例子可包括在5'末端的两个鸟嘌呤核苷酸(例如GGN₂₀NGG;SEQ ID NO:19,其中N=A、T、C、或G),以促进T7聚合酶在体外的有效转录。参见,例如,WO 2014/065596,出于所有目的通过引用将其全部内容合并于此。其它指导RNA靶序列加PAM可以具有SEQ ID NO:17-19的4-22个核苷酸的长度,包括5'G或GG和3'GG或NGG。其它指导RNA靶序列加PAM可以具有SEQ ID NO:17-19的14至20个核苷酸的长度。

[0196] 与靶DNA杂交的CRISPR复合物的形成可导致对应于指导RNA靶序列的区域内或附近的靶DNA的一条或两条链的切割(即靶DNA的非互补链上的指导RNA靶序列和与指导RNA杂

交的互补链上的反向互补序列)。例如,切割位点可以在指导RNA靶序列内(例如,相对于PAM序列的限定(defined)位置)。“切割位点”包括Cas蛋白在其上产生单链断裂或双链断裂的靶DNA的位置。切割位点可以仅在双链DNA的一条链上(例如当使用切口酶时)或在两条链上。切割位点可以在两条链上的相同位置上(产生平端;例如Cas9)或可以在每条链上的不同部位(产生交错端(即突出端);例如Cpf1)。可以例如通过使用两个Cas蛋白产生交错端,每个Cas蛋白在不同链的不同切割位点产生单链断裂,从而产生双链断裂。例如,第一切口酶可在双链DNA(dsDNA)的第一链上产生单链断裂,第二切口酶可在dsDNA的第二链上产生单链断裂,从而产生突出的序列。在一些情况下,第一链上的切口酶的指导RNA靶序列或切割位点与第二链上的切口酶的指导RNA靶序列或切割位点至少相隔2、3、4、5、6、7、8、9、10、15、20、25、30、40、50、75、100、250、500或1,000个碱基对。

D. 重组酶和重组酶删除子(Deleter)非人动物

[0197] 本文所公开的细胞或非人动物包含嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒、SAM表达盒、指导RNA表达盒,或重组酶表达盒,其中所述盒在多腺苷酸化信号的下游,或在侧接被由位点特异性重组酶识别的重组酶识别位点的转录终止子的下游;该细胞或非人动物可进一步包含驱动位点特异性重组酶表达的重组酶表达盒。编码重组酶的核酸可以被基因组整合,或者可以使用本文其它地方公开的方法(例如LNP介导的递送或AAV介导的递送)将重组酶或核酸引入此类细胞和非人动物中。可以选择递送方法以提供重组酶的组织特异性递送,如本文其它地方所公开的。

[0198] 位点特异性重组酶包括可促进重组酶识别位点之间的重组的酶,其中两个重组位点在单个核酸内物理上分开或在分开的核酸上。重组酶的实例包括Cre、Flp,和Dre重组酶。Cre重组酶基因的一个例子是Crei,其中两个编码Cre重组酶的外显子被内含子隔开,以防止其在原核细胞中表达。此类重组酶可进一步包含核定位信号,以促进定位至核(例如NLS-Crei)。重组酶识别位点包括被位点特异性重组酶识别的核苷酸序列,并且可以用作重组事件的底物。重组酶识别位点的例子包括FRT、FRT11、FRT71、attp、att、rox,和lox位点,例如loxP、lox511、lox2272、lox66、lox71、loxM2和lox5171。

[0199] 重组酶表达盒可以整合在与本文公开的其它表达盒不同的靶基因组基因座上,或者可以基因组整合在相同的靶基因组基因座上(例如Rosa26基因座,例如整合到Rosa26基因座的第一内含子中)。例如,细胞或非人动物可以对于SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)和重组酶表达盒中的每一个是杂合的,靶基因组基因座的一个等位基因包含SAM表达盒,靶基因组基因座的第二等位基因包含重组酶表达盒。同样,细胞或非人动物可以对于指导RNA表达盒(例如指导RNA阵列表达盒)和重组酶表达盒中的每一个是杂合的,靶基因组位点的一个等位基因包含指导RNA表达盒,靶基因组基因座的第二等位基因包含重组酶表达盒。

[0200] 重组酶表达盒中的重组酶基因可以可操作地连接至任何合适的启动子。启动子的实例在本文其它地方公开。例如,启动子可以是组织特异性启动子或发育阶段特异性启动子。这样的启动子是有利的,因为它们可以在期望的组织中或仅在期望的发育阶段选择性地激活靶基因的转录。例如,就Cas蛋白而言,这可以减少体内的Cas介导的毒性的可能性。表2提供了小鼠重组酶删除品系(strain)的示例性启动子的非限制性列表。

[0201] 表2:小鼠重组酶删除子品系中使用的示例性启动子。

启动子(物种)	表达位点
<i>ACTA1</i> (人类)	躯体和心脏的成年横纹肌纤维和胚胎横纹肌细胞
<i>Adipoq</i> , 脂连蛋白, 包括 C1Q 和胶原蛋白域(小鼠)	白色脂肪组织 (WAT) 和棕色脂肪组织 (BAT)
<i>Agrp</i> (小鼠)	下丘脑中的 ArGP 神经元
<i>Alb</i> , 白蛋白(大鼠)	肝脏
<i>Alb1</i> , 白蛋白(小鼠)	肝脏
<i>Amh</i> (小鼠)	睾丸支持细胞
<i>Aqp2</i> (小鼠)	肾细胞 (收集管, 左) 和睾丸 (精子, 右)。
<i>Calb2</i> , 钙结合蛋白 2	大脑和皮质中的钙视网膜蛋白中间神经元 (calretinin interneurons)
<i>Camk2a</i> , 钙/钙调蛋白依赖性蛋白激酶 II α (小鼠)	前脑, 特别是海马中的 CA1 锥体细胞层
<i>Cck</i> , 胆囊收缩素(小鼠)	皮质及成年脊髓和胚胎第 15.5 天的脊髓和心脏的胆囊收缩素阳性神经元 (中间神经元)
CD2, CD2 分子 (人类)	T 细胞和 B 细胞 (所有定向的 (comitted) B 细胞和 T 细胞祖细胞)
<i>Cd19</i>	B 细胞
<i>Cdh5</i> , 钙黏着蛋白 5	发育中的血管和静止血管的内皮细胞, 以及造血细胞的子集
<i>Chd16</i> (小鼠)	肾小管, 特别是收集管、髓祥, 和远端小管
<i>Chat</i> , 胆碱乙酰转移酶(小鼠)	胆碱能神经元
<i>Ckmm</i> (小鼠)	骨骼肌和心肌。
<i>Cort</i> , 皮质抑素(cortistatin)	表达 Cort 的细胞 (CST 阳性神经元)
<i>Crh</i> , 促肾上腺皮质激素释放激素	CRH 阳性神经元

启动子(物种)	表达位点
Cspg4 (小鼠)	中枢神经系统中表达 NG2 的神经胶质细胞 (多突胶质细胞、少突胶质细胞的祖细胞) 和其它器官中表达 NG2 的细胞; 胼胝体; 中枢神经系统和其它组织, 例如睾丸和血管
<i>Cyp39a1</i> , 细胞色素 P450, 家族 39, 亚家族 a, 多肽 1 (小鼠)	脑皮质、海马、纹状体, 嗅球, 和小脑
<i>dlx6a</i> , 无远端 (distal-less) 同源盒基因 6a	γ -氨基丁酸能前脑神经元
Ella, 腺病毒(腺病毒)	各种各样的组织, 包括将遗传改变传递给子代的生殖细胞
<i>Emx1</i> , 空通气孔同源物 1 (果蝇)	新皮质和海马的神经元以及皮质区 (pallium) 神经胶质细胞中的神经元
<i>En1</i> , engrailed 1	脊髓 V1 中间神经元, E9 产生的胚胎中脑和菱脑节 1, 以及在四肢腹侧外胚层中, 体细胞的一个子集中, 和一些中胚层衍生的组织
<i>Fabp4</i> , 脂肪酸结合蛋白 4	棕色和白色脂肪组织。
<i>Foxd1</i> (小鼠)	在注定会变成肾脏的基质细胞的细胞中的后肾间充质 (metanephric mesenchyme) 的肾脏发育, 以及遍布全身的多个器官
<i>Foxp3</i> (小鼠)	<i>Cd4+Cd25<高>Cd127<低></i> 来自淋巴结, 脾脏和胸腺的 T 细胞; 卵巢
<i>Gad2</i> , 谷氨酸脱羧酶 2	Gad2 阳性神经元
GFAP, 胶质纤维酸性蛋白 (人类)	中枢神经系统, 包括星形胶质细胞、少突胶质细胞, 室管膜和一些神经元; 还有肝的门静脉周围 (periportal) 细胞
<i>Gfap</i> (小鼠)	脑和脊髓中的星形胶质细胞, 以及产后和成年表达 GFAP 的神经干细胞及其在大脑中的子代; 在 e15.5d 软骨原基; 胸腺、心肌、眼晶状体, 嵌入成体的膀胱和肠道肌肉的周围神经
<i>Gfap</i> (小鼠)	整个健康的大脑和脊髓中的大多数星形胶质细胞, 以及中枢神经系统 (CNS) 损伤后的基本上所有星形胶质细胞; 脑室下区成体干细胞 (adult stems) 的亚群
<i>Grik4</i> , 谷氨酸受体, 离子型 (ionotropic), kainate 4 (小鼠)	14 天大时的海马 CA3 区, 以及在 8 周龄时, CA3 区近 100% 的锥体细胞中观察到重组; 其它脑区域
<i>Hspa2</i> , 热激蛋白 2 (小鼠)	s 细线期 / 偶线期 精母细胞 (leptotene/zygotene spermatocyte)
<i>Ins2</i> , 胰岛素 2 (大鼠)	胰腺 β 细胞以及下丘脑

启动子(物种)	表达位点
<i>Itgax</i> , 整合素 α X (小鼠)	CD8-, CD8 + 树突状细胞, 来自淋巴结、肺和表皮的组织来源的树突状细胞以及浆细胞样树突状细胞
<i>Kap</i> (小鼠)	雄性小鼠肾皮质近端小管细胞; 子宫和肝脏
<i>KRT14</i> , 角蛋白 14 (人类)	皮肤, 口腔外层, 包括在 11.75 d.p.c. 时的牙板和在 14.5 d.p.c. 前的牙齿上皮。
<i>Lck</i> , 淋巴细胞蛋白酪氨酸激酶(小鼠)	胸腺细胞
<i>Lck</i> (小鼠)	胸腺
<i>Lepr</i> (小鼠)	下丘脑 (弓状、背内侧、外侧和腹内侧核)、边缘和皮质脑区 (额叶底杏仁核 (basolateral amygdaloid nucleus), 梨状皮层和外侧内嗅皮层) 以及脾后皮层 (retrosplenial cortex)
<i>Lyve1</i> (小鼠)	淋巴内皮
<i>Lyz2</i> , Lysozyme 2 (小鼠)	骨髓细胞, 包括单核细胞, 成熟的巨噬细胞和粒细胞
MMTV	乳腺、唾液腺、精囊、皮肤、红细胞、B 细胞和 T 细胞; 在肺、肾、肝和脑组织中较低
<i>Mnx1</i> , 运动神经元和胰腺同源盒 1 (小鼠)	运动神经元
<i>Myf5</i> , 生肌因子 5	骨骼肌和真皮以及一些异位位置
<i>Myh6</i> (小鼠)	心脏组织
<i>Nes</i> , 神经上皮干细胞蛋白 (大鼠)	中枢和周围神经系统; 一些孤立的肾脏和心脏细胞
<i>Neurog3</i> , 神经元素 3 (neurogenin 3), (大鼠)	成年胰岛, 小肠肠内分泌细胞, 胃的内分泌部分, 所有胰腺内分泌细胞以及一些非内分泌肠细胞
<i>Nkx2-1</i>	Cre 重组酶活性通过 Nkx2.1 启动子/增强子区导向脑中间神经元祖细胞、发育中的肺、甲状腺和垂体
<i>NPHS2</i> (人类)	肾小球发育的毛细血管袢末期阶段 (late capillary loop stage) 的足细胞和成熟肾小球的足细胞
<i>Nr5a1</i> , 核受体亚家族 5 A 组成员 1 (小鼠)	下丘脑腹内侧核 (Ventromedial Hypothalamus)、皮质、肾上腺、垂体和性腺,
<i>Omp</i> , 嗅觉标记蛋白(小鼠)	成熟的嗅觉感觉神经元
<i>Pax3</i> , 配对盒基因 3	E9 至 11.5 胚胎的背神经管和体节以及 E11.5 胚胎的心脏神经嵴细胞和结肠上皮细胞
<i>Pf4</i> , 血小板因子 4(小鼠)	巨核细胞
<i>Pomc1</i> (小鼠)	下丘脑弓状核中的 POMC 神经元分散在海马齿状回中的 POMC 神经元
<i>Prdm1</i> (小鼠)	原始生殖细胞
<i>Prm</i> (小鼠)	雄性种系

启动子(物种)	表达位点
<i>Pvalb</i> , 小清蛋白	表达小清蛋白的神经元, 例如大脑中的中间神经元和背根神经节的本体感受传入感觉神经元 (proprioceptive afferent sensory neurons)
<i>Scnn1a</i> (小鼠)	皮质、丘脑、中脑, 和小脑
<i>Shh</i> , 音猬因子	内源性 <i>Shh</i> 表达模式
<i>Sim1</i> , 单意同源物 1 (果蝇)(小鼠)	下丘脑室旁核 (paraventricular hypothalamu) 和大脑其它部位
<i>Slc6a3</i> , 溶质载体家族 6 (神经递质转运蛋白、多巴胺), 成员 3	多巴胺能细胞群 (黑质 (SN) 和腹侧被盖区 (VTA), 以及在红核后区 (retrosubthalamic field))
<i>Slc17a6</i> (小鼠)	兴奋性谷氨酸能神经元细胞体 (excitatory glutamatergic neuron cell bodies)
<i>Sst</i> , 生长抑素	生长抑素阳性神经元 (包括树突状抑制中间神经元, 如马提诺帝细胞 (Martinotti cells) 和方位腔隙分子细胞 (Oriens-Lacunosum-Moleculare cells))
<i>Stra8</i> (小鼠)	产后, 减数分裂前, 男性生殖细胞
<i>Syn1</i> (大鼠)	早至 E12.5 的神经细胞, 包括脑、脊髓和 DRG, 以及成体中的神经元
<i>Tagln</i> , 胶转蛋白 (小鼠)	平滑肌
<i>Tagln</i> (小鼠)	成体平滑肌细胞 (如动脉、静脉和内脏器官) 和心肌细胞
<i>Tek</i> (小鼠)	胚胎发生和成年期的内皮细胞
<i>Thy1</i> (小鼠)	皮质和海马神经元
<i>Twist2</i> , 扭转碱性螺旋-环-螺旋转录因子 2	早至胚胎第 9.5 天时的中胚层, 在中胚层组织中, 例如鳃弓和体节, 以及在浓缩的间充质来源的软骨细胞和成骨细胞中
<i>Vav1</i> (小鼠)	花斑 (variegated) 生殖细胞 (睾丸和卵巢) 以及心脏和肠道
<i>Vill1</i> , 绒毛蛋白 1 (小鼠)	小肠和大肠的绒毛和隐窝
<i>Vip</i> , 血管活性肠多肽	一些 γ -氨基丁酸能中间神经元
<i>Wnt1</i> , 无翅型 MMTV 整合位点 1 (小鼠)	胚胎神经管、中脑、中脑和尾侧间脑 (caudal diencephalon) 的背和腹中线、中后脑交界部和背脊髓
<i>Wnt1</i> (小鼠)	发育中的神经嵴和中脑
<i>Krt17</i> , 角蛋白 17 (小鼠)	内源性角蛋白 17 表达模式
<i>Osr2</i> , 奇略过相关 2 (果蝇), 小鼠, 实验室	发育中的腮和泌尿生殖道
<i>Trp63</i> , 转化相关蛋白 63 (小鼠)	内源性 <i>Trp63</i> 表达模式

启动子(物种)	表达位点
<i>Prrx1</i> , 配对相关的同源异形框 1(大鼠)	早期肢芽间充质和颅面间充质的子集中, 以及雌性种系的有限表达
<i>Tbx22</i> , T-box 转录因子 22(小鼠)	内源性 <i>Tbx22</i> 表达模式
<i>Tgfb3</i> , 转化生长因子, $\beta 3$ (小鼠)	胚胎和胎儿发育过程中的心脏、咽弓、听泡、中脑、肢芽, 中线腭上皮 (midline palatal epithelium) 和晶须滤泡 (whisker follicles)
<i>Wnt1</i> , 无翅型 MMTV 整合位点家族, 成员 1(小鼠)	胚胎神经管、中脑、尾侧间脑 (caudal diencephalon)、中-后脑交界, 背脊髓 (dorsal spinal cord)、神经嵴细胞
<i>ACTB</i> , 肌动蛋白, β (鸡)	大多数组织类型
<i>Col2a1</i> , 胶原, II 型, $\alpha 1$ (小鼠)	胚胎发生和出生后的软骨细胞系 (软骨)。
<i>Dlx5</i> , 末梢更少同源框 5	皮层
<i>KRT14</i> , 角蛋白 14(人类)	角质形成细胞
<i>Lgr5</i> , 富含亮氨酸重复 G 蛋白偶联受体 5	结肠和小肠中的隐窝基底柱状细胞 (小肠的干细胞)
<i>Myh6</i> , 肌球蛋白, 重肽(小鼠)	发育中和成体心脏
<i>Plp1</i> , 蛋白脂质蛋白(髓磷脂) 1(小鼠)	少突胶质细胞和雪旺氏细胞
<i>UBC</i> , 泛素 C(人类)	所有组织类型
<i>Wfs1</i> , 钨综合征 1 同源物(人类)	皮质、海马、纹状体、丘脑和小脑
<i>Gt(ROSA)26Sor</i> (小鼠)	从植入前开始的大多数组织类型, 包括发育中的种系
鸡 β -肌动蛋白启动子和 hCMV 立即早期增强子	泛素

E. 编码嵌合Cas蛋白、嵌合衔接体蛋白、指导RNA, 协同激活介质或重组酶的核酸

[0202] 还提供了编码嵌合Cas蛋白、嵌合衔接体蛋白、指导RNA、重组酶, 或其任何组合的核酸。嵌合Cas蛋白、嵌合衔接体蛋白、指导RNA和重组酶在本文其它地方更详细地描述。例如, 核酸可以是嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒、协同激活介质 (SAM) 表达盒 (其包含编码嵌合Cas蛋白和嵌合衔接体蛋白两者的核酸)、指导RNA或指导RNA阵列表达盒、重组酶表达盒, 或其任何组合。这样的核酸可以是RNA (例如信使RNA (mRNA)) 或DNA, 可以是单链或双链, 并且可以是线性或环状的。DNA可以是载体的一部分, 例如表达载体或靶向载体。载体也可以是病毒载体, 例如腺病毒载体、腺相关病毒载体、慢病毒载体和逆转录病毒载体。当将本文公开的任何核酸引入细胞中时, 编码的嵌合DNA-靶向蛋白, 嵌合衔接体蛋白或指导RNA可以在细胞中瞬时、条件或组成性表达。

[0203] 可选地, 可以对核酸进行密码子优化以在特定细胞或生物中有效翻译成蛋白质。例如, 可以对核酸进行修饰以替代与天然多核苷酸序列相比在细菌细胞、酵母细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其它感兴趣的宿主细胞中具有较高使用频率的密码子。

[0204] 核酸或表达盒可以稳定地整合到细胞或非人动物的基因组 (即染色体) 中, 或者其可以位于染色体外 (例如染色体外复制DNA)。稳定整合的表达盒或核酸可以随机整合到非

人动物的基因组中(即转基因),或者它们可以整合到非人动物的基因组中的预定区域中(即敲入)。在一个实例中,如本文其它地方所述,将核酸或表达盒稳定整合到安全港基因座中。稳定整合有核酸或表达盒的靶基因组位点可以是对核酸或表达盒杂合的,或对核酸或表达盒纯合的。例如,靶基因组基因座或细胞或非人动物对SAM表达盒可以是杂合的,对于指导RNA表达盒可以是杂合的,可选地,每个在不同等位基因上处于相同的靶基因组基因座。

[0205] 本文所述的核酸或表达盒可以可操作地连接至任何合适的启动子,以在非人动物体内表达或离体在细胞内表达。非人动物可以是本文其它各处所述的任何合适的非人动物。作为一个实例,核酸或表达盒(例如嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒,或包含编码嵌合Cas蛋白和嵌合衔接体蛋白的核酸的SAM盒)可以可操作地连接在靶基因组基因座上的内源启动子,例如Rosa26启动子。替代地,盒核酸或表达盒可以可操作地连接至外源启动子,例如组成型活性启动子(例如CAG启动子或U6启动子)、条件启动子(conditional promoter)、诱导型启动子、时间受限的启动子(temporally restricted promoter)(例如发育调控的启动子(developmentally regulated promoter)),或空间受限的启动子(spatially restricted promoter)(例如细胞特异性或组织特异性启动子)。这样的启动子是众所周知的,并在本文其它地方讨论。可用于表达构建体的启动子包括在例如真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞、兔细胞、多能细胞、胚胎干(ES)细胞,或合子中具有活性的启动子。这样的启动子可以是例如条件启动子、诱导型启动子、组成型启动子或组织特异性启动子。

[0206] 例如,可以将编码指导RNA的核酸可操作地连接至U6启动子,例如人U6启动子或小鼠U6启动子。合适的启动子(例如用于表达指导RNA的启动子)的具体实例包括RNA聚合酶III启动子,例如人U6启动子、大鼠U6聚合酶III启动子,或小鼠U6聚合酶III启动子。

[0207] 可选地,启动子可以是双向启动子,其驱动一个基因(例如编码嵌合DNA靶向蛋白的基因)和另一个方向的第二个基因(例如编码指导RNA或嵌合衔接体蛋白的基因)的表达。这种双向启动子可以由以下组成:(1)完整的、常规的单向Po1 III启动子,该启动子包含3个外部控制元件:远端序列元件(DSE)、近端序列元件(PSE)和TATA盒;和(2)第二种基本的Po1 III启动子,其包括PSE和以反向方向融合到DSE的5'末端的TATA盒。例如,在H1启动子中,DSE与PSE和TATA盒相邻,并且通过创建杂交启动子,可以使启动子双向表达;在该杂交启动子中,通过附加衍生自U6启动子的PSE和TATA盒来控制反向转录。参见,例如,US 2016/0074535,出于所有目的通过引用将其全部内容合并于此。使用双向启动子来同时表达两个基因可以产生致密(compact)表达盒以促进递送。

[0208] 一种或多种核酸可以在多顺反子表达构建体中在一起。例如,编码嵌合Cas蛋白的核酸和编码嵌合衔接体蛋白的核酸可以一起在双顺反子表达构建体中。参见例如图1A和1B。多顺反子表达载体同时表达来自同一mRNA的两个或多个单独的蛋白质(即由同一启动子产生的转录本)。蛋白质多顺反子表达的合适策略包括,例如,使用2A肽和使用内部核糖体进入位点(IRES)。例如,这样的构建体可以包含:(1)编码一种或多种嵌合Cas蛋白和一种或多种嵌合衔接体蛋白的核酸;(2)编码两个或多个嵌合衔接体蛋白的核酸;(3)编码两个或更多个嵌合Cas蛋白的核酸;(4)编码两个或更多个指导RNA或两个或更多个指导RNA阵列的核酸;(5)编码一个或多个嵌合Cas蛋白和一个或多个指导RNA或指导RNA阵列的核酸;(6)

编码一个或多个嵌合衔接体蛋白和一个或多个指导RNA或指导RNA阵列的核酸；(7) 编码一个或多个嵌合Cas蛋白、一个或多个嵌合衔接体蛋白，以及一个或多个指导RNA或指导RNA阵列的核酸。作为一个实例，这样的多顺反子载体可以使用一个或多个内部核糖体进入位点(IRES)以允许从mRNA的内部区域开始翻译。作为另一个例子，这样的多顺反子载体可以使用一个或多个2A肽。这些肽是小的“自切割”肽，通常具有18-22个氨基酸的长度，并从同一mRNA中产生等摩尔水平的多个基因。核糖体跳过了2A肽C端的甘氨酸-脯氨酸肽键的合成，导致2A肽与其直接下游肽之间的“切割”。参见，例如，Kim et al. (2011) PLoS One 6(4) : e18556, 出于所有目的通过引用整体并入本文。“切割”发生在C末端的甘氨酸和脯氨酸残基之间，这意味着上游顺反子将在其末端添加一些额外的残基，而下游顺反子将从脯氨酸开始。结果，“切割(cleaved-off)”的下游肽在其N端具有脯氨酸。2A介导的切割是所有真核细胞中的普遍现象。已经从小RNA病毒、昆虫病毒和C型轮状病毒中识别出2A肽。参见，例如，Szymczak et al. (2005) Expert Opin Biol Ther 5:627-638, 出于所有目的通过引用整体并入本文。可以使用的2A肽的例子包括四体(Thataasigna)病毒2A(T2A)；猪捷申病毒-1 2A(P2A)；马甲型鼻炎病毒(ERAV) 2A(E2A)；和FMDV 2A(F2A)。示例性的T2A、P2A、E2A和F2A序列包括以下：T2A(EGRGSLTTCGDVEENPGP；SEQ ID NO:20)；P2A(ATNFSLLKQAGDVEENPGP；SEQ ID NO:21)；E2A(QCTNYALLKLAGDVESNPGP；SEQ ID NO:22)；和F2A(VKQTLNFDLLKLAGDVESNPGP；SEQ ID NO:23)。可以将GSG残基添加到任何这些肽的5'末端，以提高切割效率。

[0209] 任何核酸或表达盒也可在编码序列上游包含多腺苷酸化信号或转录终止子。例如，嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒、SAM表达盒、指导RNA表达盒，或重组酶表达盒可在表达盒中的编码序列上游包含多腺苷酸化信号或转录终止子。多腺苷酸化信号或转录终止子可以侧接位点特异性重组酶识别的重组酶识别位点。可选地，重组酶识别位点也可侧接选择盒，该选择盒包括，例如，抗药性蛋白的编码序列。可选地，重组酶识别位点不侧接选择盒。多腺苷酸化信号或转录终止子阻止了由编码序列编码的蛋白质或RNA(例如嵌合Cas蛋白、嵌合衔接体蛋白、指导RNA，或重组酶)的转录和表达。但是，暴露于位点特异性重组酶后，将切除多腺苷酸化信号或转录终止子，并且可以表达蛋白质或RNA。

[0210] 如果多腺苷酸化信号或转录终止子以组织特异性或发育阶段特异性方式被切除，用于表达盒(例如嵌合Cas蛋白表达盒或SAM表达盒)的这种构型能够在包含表达盒的非人动物中进行组织特异性表达或发育阶段特异性表达。例如，在嵌合Cas蛋白的情况下，由于嵌合Cas蛋白在细胞或非人动物中的长时间表达或嵌合Cas蛋白在不希望的发育阶段或在非人动物中不希望的细胞或组织类型中的表达，这可能降低毒性。参见，例如，Parikh et al. (2015) PLoS One 10(1) : e0116484, 出于所有目的，通过引用将其全文并入本文。如果包含表达盒的非人动物还包含可操作地连接至组织特异性或发育阶段特异性启动子的位点特异性重组酶的编码序列，则可以以组织特异性或发育阶段特异性方式实现多腺苷酸化信号或转录终止子的切除。然后仅在那些组织或那些发育阶段中切除多腺苷酸化信号或转录终止子，从而实现组织特异性表达或发育阶段特异性表达。在一实例中，嵌合Cas蛋白、嵌合衔接体蛋白、嵌合Cas蛋白和嵌合衔接体蛋白，或指导RNA可以肝特异性方式表达。本文中其它地方公开了已用于研发非人动物的此类“重组酶删除子”品系的此类启动子的实例。

[0211] 可以使用任何转录终止子或多腺苷酸化信号。本文所使用的“转录终止子”是指引起转录终止的DNA序列。在真核生物中，转录终止子由蛋白质因子识别，终止后是多腺苷酸

化,其为在poly(A)聚合酶在场的情况下将poly(A)尾巴添加到mRNA转录物中的过程。哺乳动物的poly(A)信号通常由一个约45个核苷酸长的核心序列组成,该核心序列可侧接各种辅助序列,这些辅助序列用于增强切割和多腺苷酸化效率。核心序列由mRNA中的高度保守的上游元件(称为AATAAA或AAUAAA)(称为poly A识别模体或poly A识别序列,被切割和多腺苷酸化特异性因子(CPSF)识别)和定义不明确的下游区域(富含U,或G和U)组成;该下游区域受裂解刺激因子(CstF)约束。可以使用的转录终止子的例子包括,例如,人生长激素(HGH)多腺苷酸化信号、猿猴病毒40(SV40)晚期多腺苷酸化信号、兔 β -球蛋白多腺苷酸化信号、牛生长激素(BGH)多腺苷酸化信号、磷酸甘油酸激酶(PGK)多腺苷酸化信号、AOX1转录终止序列、CYC1转录终止序列,或已知适合调节真核细胞中基因表达的任何转录终止序列。

[0212] 位点特异性重组酶包括可促进重组酶识别位点之间的重组的酶,其中两个重组位点在单个核酸内物理上分开或在分开的核酸上。重组酶的实例包括Cre、Flp,和Dre重组酶。Cre重组酶基因的一个例子是Crei,其中两个编码Cre重组酶的外显子被内含子隔开,以防止其在原核细胞中表达。此类重组酶可进一步包含核定位信号,以促进定位至核(例如NLS-Crei)。重组酶识别位点包括被位点特异性重组酶识别的核苷酸序列,并且可以用作重组事件的底物。重组酶识别位点的例子包括FRT、FRT11、FRT71、attp、att、rox,和lox位点,例如loxP、lox511、lox2272、lox66、lox71、loxM2和lox5171。

[0213] 本文公开的表达盒也可以包含其它组分。此类表达盒(例如嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒、SAM表达盒、指导RNA表达盒,或重组酶表达盒)可进一步在表达盒的5'末端包含3'剪接序列和/或在编码序列(例如编码嵌合Cas蛋白、嵌合衔接体蛋白、指导RNA,或重组酶)之后的第二多腺苷酸化信号。术语“3'剪接序列”是指在3'内含子/外显子边界的核酸序列,其可以被剪接机制识别并结合。表达盒可进一步包括选择盒,其包含例如抗药性蛋白的编码序列。合适的选择标记的例子包括新霉素磷酸转移酶(neo_r)、潮霉素B磷酸转移酶(hyg_r)、嘌呤霉素-N-乙酰基转移酶(puro_r)、杀稻瘟菌素S脱氨酶(bsr_r)、黄嘌呤/鸟嘌呤磷酸核糖基转移酶(gpt),和单纯疱疹病毒胸苷激酶(HSV-k)。可选地,选择盒可以侧接位点特异性重组酶的重组酶识别位点。如果表达盒还包含如上所述的侧接在编码序列上游的聚、多腺苷酸化信号的重组酶识别位点,则选择盒可以侧接相同的重组酶识别位点或者可以侧接为被不同重组酶识别的不同组的重组酶识别位点。

[0214] 表达盒还可包含编码一种或多种报道蛋白,例如荧光蛋白(例如绿色荧光蛋白),的核酸。可以使用任何合适的报道蛋白。例如,可以使用如本文其它地方所定义的荧光报告蛋白,或者可以使用非荧光报告蛋白。荧光报告蛋白的实例在本文其它地方提供。非荧光报告蛋白包括,例如,可用于组织化学或生物发光测定的报告蛋白,例如 β -半乳糖苷酶、荧光素酶(例如,海肾荧光素酶、萤火虫荧光素酶和NanoLuc荧光素酶)和 β -葡萄糖醛酸苷酶。表达盒可以包括可以在流式细胞术测定中检测到的报道蛋白(例如,荧光报道蛋白,例如绿色荧光蛋白)和/或可以在组织化学测定中检测到的报道蛋白(例如 β -半乳糖苷酶蛋白)。这种组织化学测定的一个例子是通过水解X-Gal(5-溴-4-氯-3-吲哚基-b-D-吡喃半乳糖苷)产生的蓝色沉淀,或使用荧光底物(例如 β -甲基伞形半乳糖苷(MUG)和荧光素二半乳糖苷(FDG)),组织化学地可视化 β -半乳糖苷酶的原位表达。

[0215] 本文所述的表达盒可以是任何形式。例如,表达盒可以在载体或质粒,例如病毒载体中。该表达盒可以与能够引导蛋白质或RNA表达的表达构建体中的启动子可操作地连接

(例如在去除上游多腺苷酸化信号时)。替代地,表达盒可以在靶向载体中。例如,靶向载体可以包含侧接表达盒的同源臂,其中该同源臂适合于引导与期望的靶基因组基因座的重组,以促进基因组整合和/或内源序列的替换。

[0216] 本文所述的表达盒可以是体外的,它们可以离体在细胞(例如胚胎干细胞)内(例如基因组整合的或染色体外的),或者它们可以在生物体(例如非人动物)体内(例如,基因组整合或染色体外)。如果是离体的,则表达盒可以在来自任何生物的任何类型的细胞中,例如全能细胞,例如胚胎干细胞(例如,小鼠或大鼠胚胎干细胞)或诱导的多能干细胞(例如人诱导的多能干细胞)。如果在体内,则表达盒可以在任何类型的生物体中(例如,如本文其它地方进一步描述的非人动物)。

[0217] 编码催化失活的Cas蛋白的核酸的具体实例可以包含以下,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:2所示的dCas9蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性。可选地,该核酸可以包含,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:24中所示的序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性(可选地,其中该序列编码蛋白,该蛋白与SEQ ID NO:2所示的dCas9蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性)。

[0218] 编码嵌合Cas蛋白的核酸的具体实例可以包含以下,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:1所示的嵌合Cas蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性。可选地,该核酸可以包含,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:25中所示的序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性(可选地,其中该序列编码蛋白,该蛋白与SEQ ID NO:1所示的嵌合Cas蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性)。

[0219] 编码衔接体的具体实例可以包含以下,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:7所示的MCP序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性。可选地,该核酸可以包含,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:26中所示的序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性(可选地,其中该序列编码蛋白,该蛋白与SEQ ID NO:7所示的MCP序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性)。

[0220] 编码嵌合衔接体蛋白的具体实例可以包含以下,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:6所示的嵌合衔接体蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性。可选地,该核酸可以包含,基本上由以下组成,或由以下组成:编码氨基酸序列的核酸,该氨基酸序列与SEQ ID NO:27中所示的序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的同一性(可选地,其中该序列编码蛋白,该蛋白与SEQ ID

NO:6所示的嵌合衔接体蛋白序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%，或100%的同一性)。

[0221] 编码转录激活域的具体实例可以包含以下，基本上由以下组成，或由以下组成：编码氨基酸序列的核酸，该氨基酸序列与SEQ ID NO:3、8，或9所示的VP64、p65，或HSF1序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%，或100%的同一性。可选地，该核酸可以包含，基本上由以下组成，或由以下组成：编码氨基酸序列的核酸，该氨基酸序列与SEQ ID NO:28、29，或30中所示的序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%，或100%的同一性(可选地，其中该序列编码蛋白，该蛋白与SEQ ID NO:3、8，或9所示的VP64、p65，或HSF1序列具有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%，或100%的同一性)

[0222] 一种示例性的协同激活介质(SAM)表达盒从5'至3'包含：(a) 3'剪接序列；(b) 第一重组酶识别位点(例如loxP位点)；(c) 药物抗性基因的编码序列(例如，新霉素磷酸转移酶(neo^r)编码序列)；(d) 多腺苷酸化信号；(e) 第二重组酶识别位点(例如loxP位点)；(f) 嵌合Cas蛋白编码序列(例如dCas9-NLS-VP64融合蛋白)；(g) 2A蛋白编码序列(例如T2A编码序列)；和(e) 嵌合衔接体蛋白编码序列(例如，MCP-NLS-p65-HSF1)。参见，例如，图1A和SEQ ID NO:31(SEQ ID NO:64所示的编码序列和SEQ ID NO:44所示的编码蛋白)。

[0223] 一种示例性的通用指导RNA阵列表达盒从5'至3'包含：(a) 3'剪接序列；(b) 第一重组酶识别位点(例如rox位点)；(c) 药物抗性基因的编码序列(例如，嘌呤霉素-N-乙酰基转移酶(puro^r)编码序列)；(d) 多腺苷酸化信号；(e) 第二重组酶识别位点(例如rox位点)；(f) 指导RNA，其包含一个或多个指导RNA基因(例如，第一U6启动子，然后是第一指导RNA编码序列；第二U6启动子，然后是第二指导RNA编码序列；第三U6启动子，然后是第三指导RNA编码序列)。参见，例如，图5和SEQ ID NO:32。SEQ ID NO:32的包含启动子和指导RNA编码序列的区域在SEQ ID NO:65中列出。指导RNA阵列表达盒中的重组酶识别位点可以与SAM表达盒中的重组酶识别位点相同或不同(例如，可以被相同的重组酶或不同的重组酶识别)。编码靶向小鼠Ttr的指导RNA的这种示例性指导RNA阵列表达盒在SEQ ID NO:33中列出。SEQ ID NO:33的包含启动子和指导RNA编码序列的区域在SEQ ID NO:66中列出。

[0224] 另一个示例性通用指导RNA阵列表达盒包含一个或多个指导RNA基因(例如，第一U6启动子，然后是第一指导RNA编码序列；第二U6启动子，然后是第二指导RNA编码序列；第三U6启动子，然后是第三指导RNA编码序列)。这种示例性的通用指导RNA阵列表达盒在SEQ ID NO:66中列出。用于特定基因的这种指导RNA阵列表达盒的例子例如在SEQ ID NO:33、66、67、71、84、85，和98中列出。

F. 用于整合的基因组位点

[0225] 本文所述的核酸和表达盒可以基因组整合在细胞或非动物的靶基因组基因座处。可以使用能够表达基因的任何靶基因组基因座。

[0226] 可以稳定地整合本文所述的核酸或盒的靶基因组基因座的实例是非人动物基因组中的安全港基因座。整合的外源DNA与宿主基因组之间的相互作用会限制整合的可靠性和安全性，并可能导致明显的表型效应；这不是由于靶向遗传修饰，而是由于整合对周围内源基因的意外影响。例如，随机插入的转基因可能会受到位置效应和沉默的影响，从而使其表达不可靠且不可预测。同样，将外源DNA整合到染色体基因座中会影响周围的内源基因和

染色质,从而改变细胞行为和表型。安全港基因座包括染色体基因座,其中转基因或其它外源核酸插入物可以在所有目的组织中稳定、可靠地表达,而不会明显改变细胞行为或表型(即,对宿主细胞没有任何有害作用)。参见,例如,Sadelain et al. (2012) Nat. Rev. Cancer 12:51-58,出于所有目的通过引用整体并入本文。例如,安全港基因座可以是其中插入的基因序列的表达不受邻近基因的任何通读表达干扰的基因。例如,安全港基因座可以包括染色体基因座,其中外源DNA可以以可预测的方式整合和发挥功能,而不会不利地影响内源基因的结构或表达。安全港基因座可包括基因外区域或基因内区域,例如,非必需、可省去,或可被破坏而无明显表型后果的基因内的基因座。

[0227] 例如,Rosa26基因座及其在人类中的等同物在所有组织中提供开放的染色质构型,并在胚胎发育期间和成体中普遍表达。参见,例如,Zambrowicz et al. (1997) Proc. Natl. Acad. Sci. USA 94:3789-3794,出于所有目的通过引用整体并入本文。此外,可以高效地靶向Rosa26基因座,并且Rosa26基因的破坏不会产生明显的表型。安全港基因座的其它示例包括CCR5、HPRT、AAVS1,和白蛋白。参见,例如,美国专利号7,888,121;7,972,854;7,914,796;7,951,925;8,110,379;8,409,861;8,586,526;和美国专利公开号2003/0232410;2005/0208489;2005/0026157;2006/0063231;2008/0159996;2010/00218264;2012/0017290;2011/0265198;2013/0137104;2013/0122591;2013/0177983;2013/0177960;和2013/0122591,出于所有目的,通过引用将其每一篇全部内容整体并入本文。安全港基因座(例如Rosa26基因座)的双等位基因靶向没有负面影响,因此可以将不同的基因或报告基因靶向两个Rosa26等位基因。在一个实例中,表达盒被整合到Rosa26基因座的内含子,例如Rosa26基因座的第一内含子。参见,例如,图2。

[0228] 整合到靶基因组基因座中的表达盒可以与靶基因组基因座处的内源启动子可操作地连接,或者可以和与靶基因组基因座异源的外源启动子可操作地连接。在一个实例中,嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒,或协同激活介质(SAM)表达盒被整合到靶基因组基因座(例如Rosa26基因座)中,并在靶基因组基因座处与内源启动子(例如Rosa26启动子)可操作地连接。在另一个实例中,将指导RNA表达盒整合到靶基因组基因座(例如Rosa26基因座)中,并且可操作地连接至一个或多个异源启动子(例如U6启动子,例如位于每个指导RNA编码序列上游的不同U6启动子)。

G. 非人动物基因组、非人动物细胞,和非人动物

[0229] 还提供了包含本文所述的核酸或表达盒的非人动物基因组、非人动物细胞,和非人动物。基因组、细胞,或非人动物可以是雄性或雌性。核酸或表达盒可以稳定地整合到细胞或非人动物的基因组中(即染色体中),或者可以位于染色体之外(例如染色体外复制DNA)。可以将核酸或表达盒随机整合到非人动物的基因组中(即转基因),或者可以将其整合到非人动物的基因组的预定区域(例如安全港基因座)中(即敲入)。稳定整合有核酸或表达盒的靶基因组基因座可以对于核酸或表达盒杂合,或对于核酸或表达盒纯合。二倍体生物在每个遗传位点具有两个等位基因。每对等位基因代表特定遗传基因座的基因型。如果在特定位点有两个相同的等位基因,则将基因型描述为纯合;如果两个等位基因不同,则将基因型描述为杂合。包含本文所述的稳定整合的核酸或表达盒的非人动物可在其种系中包含所述核酸或表达盒。

[0230] 例如,非人动物基因组、非人动物细胞,或非人动物可包含嵌合Cas蛋白表达盒、嵌

合衔接体蛋白表达盒,或协同激活介质(SAM)表达盒(包含嵌合Cas蛋白编码序列和嵌合衔接体蛋白序列两者)。在一实例中,基因组、细胞或非人动物包含SAM表达盒,该SAM表达盒包含嵌合Cas蛋白编码序列和嵌合衔接体蛋白编码序列两者。在一实例中,SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)被稳定地整合到基因组中。稳定整合的SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)可以随机整合到非人动物的基因组中(即转基因),也可以整合到非人动物的基因组的预定区域中(即敲入)。在一个实例中,SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)被稳定整合到基因组的预定区域,例如安全港基因座(例如Rosa26)。稳定整合SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)的靶基因组基因座可以对SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)杂合或纯合。

[0231] 可选地,上述基因组、细胞,或非人动物可进一步包含指导RNA表达盒(例如指导RNA阵列表达盒)。指导RNA表达盒可以稳定地整合到细胞或非人动物的基因组中(即染色体中),或者可以位于染色体之外(例如染色体外复制DNA或通过AAV、LNP,或本文公开的任何其它方式引入细胞或非人类细胞中)。指导RNA表达盒可以随机整合到非人动物的基因组中(即转基因),或者可以将其整合到非人动物的基因组的预定区域(例如安全港基因座)中(即敲入)。稳定整合有指导RNA表达盒的靶基因组基因座可以对于指导RNA表达盒杂合,或对于指导RNA表达盒纯合。在一个实例中,基因组、细胞或非人动物包括SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)和指导RNA表达盒。在一实例中,两个盒均基因组整合。指导RNA表达盒可以与SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)整合在不同的靶基因组基因座上或者可以基因组整合在相同的靶基因座上(例如,Rosa26基因座,例如整合在Rosa26基因座的第一个内含子中)。例如,基因组、细胞,或非人动物可以对于每个SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒)和指导RNA表达盒是杂合的,其中一个靶基因组基因座(例如Rosa26)的等位基因包含SAM表达盒(或嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒),而靶基因组基因座的第二等位基因包含指导RNA表达盒表达盒。

[0232] 可选地,上述基因组、细胞,或非人动物中的任一个可以进一步包含重组酶表达盒。重组酶表达盒可以稳定地整合到细胞或非人动物的基因组(即染色体)中,或者可以位于染色体之外(例如染色体外复制DNA,或通过AAV、LNP、HDD,或此处公开的任何其它方式引入细胞或非人动物中)。重组酶表达盒可以随机整合到非人动物的基因组中(即转基因),或者其可以整合到非人动物的基因组的预定区域(例如安全港基因座)中(即敲入)。重组酶表达盒被稳定整合的靶基因组基因座可以对重组酶表达盒为杂合或纯合。重组酶表达盒可以与本文公开的任何其它表达盒整合在不同的靶基因组基因座上,或者可以基因组整合在相同的靶基因座上(例如Rosa26基因座,例如整合到Rosa26基因座的第一内含子中)。

[0233] 本文提供的基因组或细胞可以是,例如,真核生物基因组或细胞,其包括,例如,真菌细胞(例如酵母)、植物细胞、动物细胞、哺乳动物细胞、非人哺乳动物细胞,和人类细胞。术语“动物”包括哺乳动物、鱼类,和鸟类。哺乳动物基因组或细胞可以是,例如,非人哺乳动物细胞、人类细胞、啮齿动物细胞、大鼠细胞、小鼠细胞,或仓鼠细胞。其它非人哺乳动物包括,例如,非人灵长类动物、猴子、猿、猫、狗、兔、马、公牛、鹿、野牛、家畜(例如牛科动物如奶牛和阉牛等;绵羊科动物如绵羊和山羊等;以及猪科动物如猪和野猪)。鸟类包括,例如,鸡、

火鸡、鸵鸟、鹅、鸭等。还包括家养动物和农业动物。术语“非人”不包括人类。

[0234] 细胞也可以是任何类型的未分化或分化状态。例如，细胞可以是全能细胞、多能细胞（例如人多能细胞或非人多能细胞，例如小鼠胚胎干（ES）细胞或大鼠ES细胞）或非多能细胞。全能细胞包括可以产生任何细胞类型的未分化细胞，多能细胞包括具有发展为一种以上分化细胞类型的能力的未分化细胞。这样的多能和/或全能细胞可以是，例如，ES细胞或ES样细胞，例如诱导性多能干（iPS）细胞。ES细胞包括胚胎来源的全能或多能细胞，其在引入胚胎后能够对发育中的胚胎的任何组织作出贡献。ES细胞可以源自胚泡的内部细胞团，并且能够分化为三个脊椎动物胚层（内胚层、外胚层，和中胚层）中任何一个的细胞。

[0235] 人多能细胞的例子包括人ES细胞、人成体干细胞、发育受限的人祖细胞，和人诱导性多能干（iPS）细胞，例如primed的人iPS细胞和naïve的人iPS细胞。诱导的多能干细胞包括可以直接源自分化的成体细胞的多能干细胞。可以通过将特定的重编程因子集引入细胞，以生成iPS细胞，其可以包括例如Oct3/4、Sox家族转录因子（例如Sox1、Sox2、Sox3、Sox15）、Myc家族转录因子（例如c-Myc、l-Myc、n-Myc），Krüppel样家族（KLF）转录因子（例如KLF1、KLF2、KLF4、KLF5）和/或相关的转录因子，例如NANOG、LIN28和/或Glis1。还可以通过例如使用miRNA、模仿转录因子作用的小分子，或谱系特异性分子（lineage specifiers）来生成iPS细胞。人iPS细胞的特征在于它们能够分化为三个脊椎动物胚层（例如内胚层、外胚层，或中胚层）的任何细胞。人iPS细胞的特征还在于其在合适的体外培养条件下无限增殖的能力。参见例如Takahashi和Yamanaka（2006）Cell 126:663-676，出于所有目的通过引用将其全部内容合并于此。Primed的人ES细胞和primed的人iPS细胞包括表达与植入后的上胚层细胞相似的特征并致力于谱系特异性和分化的细胞。Naïve的人类ES细胞和naïve的iPS细胞包括表达与植入前胚胎内部细胞团的ES细胞相似的特征，并不致力于谱系特异性的细胞。参见，例如，Nichols和Smith（2009）Cell Stem Cell 14:487-492，出于所有目的通过引用整体并入本文。

[0236] 本文提供的细胞也可以是生殖细胞（例如精子或卵母细胞）。该细胞可以是有丝分裂潜能细胞（mitotically competent cells）或有丝分裂惰性细胞、减数分裂潜能细胞或减数分裂惰性细胞。类似地，细胞也可以是初级体细胞或并非初级体细胞的细胞。体细胞包括不是配子、生殖细胞、配子母细胞，或未分化干细胞的任何细胞。例如，细胞可以是肝细胞、肾细胞、造血细胞、内皮细胞、上皮细胞、成纤维细胞、间充质细胞、角质形成细胞、血细胞、黑素细胞、单核细胞、单个核细胞、单核细胞前体、B细胞、红细胞-巨核细胞、嗜酸性粒细胞、巨噬细胞、T细胞、胰岛β细胞、外分泌细胞、胰腺祖细胞、内分泌祖细胞、脂肪细胞、前脂肪细胞、神经元、神经胶质细胞、神经干细胞、神经元、肝母细胞（hepatoblasts）、肝细胞、心肌细胞、骨骼成肌细胞、平滑肌细胞、导管细胞、泡细胞、α细胞、β细胞、δ细胞、PP细胞、胆囊上皮细胞（cholangiocytes）、白色或棕色脂肪细胞，或眼细胞（例如，小梁网细胞、视网膜色素上皮细胞、视网膜微血管内皮细胞、视网膜周细胞、结膜上皮细胞、结膜成纤维细胞、虹膜色素上皮细胞、角膜细胞、晶状体上皮细胞、非色素性睫状上皮细胞、眼球脉络膜成纤维细胞、感光细胞、神经节细胞、双极细胞、水平细胞，或无长突细胞）。

[0237] 本文提供的合适的细胞还包括初级细胞。初级细胞包括直接从生物体、器官，或组织分离的细胞或细胞培养物。初级细胞包括既未转化也非永生的细胞。它们包括获自生物体、器官，或组织的任何细胞，所述细胞先前没有在组织培养中进行传代或先前已经在组织

培养中进行了传代但是无法在组织培养中进行无限传代。这样的细胞可以通过常规技术分离,并且包括,例如,体细胞、造血细胞、内皮细胞、上皮细胞、成纤维细胞、间充质细胞、角质形成细胞、黑素细胞、单核细胞(monocytes)、单个核细胞(mononuclear cells)、脂肪细胞、前脂肪细胞、神经元、神经胶质细胞、肝细胞、骨骼肌成肌细胞,和平滑肌细胞。例如,原代细胞可以衍生自结缔组织、肌肉组织、神经系统组织,或上皮组织。

[0238] 本文提供的其它合适的细胞包括永生细胞。永生细胞包括来自多细胞生物的细胞,这些细胞通常不会无限期增殖,但是由于突变或改变而逃避了正常的细胞衰老,可以继续分裂。这样的突变或改变可以自然发生或有意诱导。永生细胞的实例包括中国仓鼠卵巢(CHO)细胞、人胚肾细胞(例如HEK 293细胞或293T细胞),和小鼠胚成纤维细胞(例如3T3细胞)。永生细胞的许多类型是众所周知的。永生细胞或原代细胞包括通常用于培养或表达重组基因或蛋白质的细胞。

[0239] 本文提供的细胞还包括单细胞期胚胎(即受精卵母细胞或受精卵)。这样的单细胞期胚胎可以来自任何遗传背景(例如,BALB/c、C57BL/6、129,或它们的组合),可以是新鲜的或冷冻的,并且可以源自自然育种或体外受精。

[0240] 本文提供的细胞可以是正常,健康的细胞,或者可以是患病或携带突变的细胞。

[0241] 包含本文所述核酸或表达盒的非人动物可以通过本文其它地方描述的方法制备。术语“动物”包括哺乳动物、鱼类,和蠕虫。哺乳动物包括,例如,人类、非人灵长类动物、猴子、猿、猫、狗、马、公牛、鹿、野牛、绵羊、兔子、啮齿动物(如小鼠、大鼠、仓鼠,以及豚鼠),以及家畜(例如牛科动物如奶牛和阉牛;绵羊科动物如绵羊和山羊;以及猪科动物如猪和野猪)。鸟类包括,例如,鸡、火鸡、鸵鸟、鹅、鸭。还包括家养动物和农业动物。术语“非人动物”不包括人类。优选的非人动物包括例如啮齿动物,例如小鼠和大鼠。

[0242] 非人动物可以来自任何遗传背景。例如,合适的小鼠可以来自129种系、C57BL/6种系、129种系和C57BL/6种系的混合、BALB/c种系,或Swiss Webster种系。129种系的实例包括129P1、129P2、129P3、129X1、129S1(例如129S1/SV、129S1/Sv1m)、129S2、129S4、129S5、129S9/SvEvH、129S6(129/SvEvTac)、129S7、129S8、129T1,和129T2。参见,例如,Festing et al.(1999)Mammalian Genome 10:836,出于所有目的通过引用整体并入本文。C57BL种系的实例包括C57BL/A、C57BL/An、C57BL/GrFa、C57BL/Kal_wN、C57BL/6、C57BL/6J、C57BL/6ByJ、C57BL/6NJ、C57BL/10、C57BL/10ScSn、C57BL/10Cr,和C57BL/01a。合适的小鼠还可来自于前述129种系和前述C57BL/6种系的混合(例如50%129和50%C57BL/6)。类似地,合适的小鼠可来自于前述129种系的混合或前述BL/6种系的混合(例如129S6(129/SvEvTac)种系)。

[0243] 类似地,大鼠可来自于任何大鼠种系,包括例如ACI大鼠种系、Dark Agouti(DA)大鼠种系、Wistar大鼠种系、LEA大鼠种系、Sprague Dawley(SD)大鼠种系,或Fischer大鼠种系(例如Fisher F344或Fisher F6)。大鼠还可获自衍生自上述两种或多种种系的混合的种系。例如,合适的大鼠可获自DA种系或ACI种系。ACI大鼠种系被表征为具有黑色刺豚鼠,其具有白色腹部和足部以及RT1^{av1}单体型。这些种系可获自包括Harlan实验室在内的多种不同的来源。Dark Agouti(DA)大鼠种系被表征为具有刺鼠皮(agouti coat)和RT1^{av1}单体型。这些大鼠获自包括Charles River和Harlan实验室在内的多种不同的来源。一些合适的大鼠可来自同系交配大鼠种系。参见,例如,US2014/0235933,其全部内容出于所有目的通过引用并入本文。

III. 增加靶基因的转录/表达和评估CRISPR/Cas体内活性的方法

[0244] 提供了多种方法,以使用协同激活介质系统以及本文所述的细胞和非人动物在体内激活一个或多个靶基因的转录或评估CRISPR/Cas的递送以及评估CRISPR/Cas在活体动物的组织和器官中的活性。此类方法利用了包含本文其它地方所述的表达盒的非人动物。

A. 增加靶基因的表达或测试CRISPR/Cas在体内或体外激活靶基因转录的能力的方法

[0245] 提供了多种方法来增加/激活靶基因的表达/转录或评估本文所述的CRISPR/Cas协同激活介质(SAM)系统在体内使用本文所述的非人动物增加/激活靶基因的表达/转录的能力。这样的非人动物例如可以包含SAM表达盒(包含嵌合Cas蛋白编码序列和嵌合衔接体蛋白编码序列),或者可以包含嵌合Cas蛋白表达盒或嵌合衔接体蛋白表达盒。这样的方法可以包括将一个或多个指导RNA引入非人动物,所述指导RNA各自包含一个或多个本文公开的嵌合衔接体蛋白可以特异性结合的衔接体结合元件。一个或多个指导RNA可以与嵌合Cas蛋白和嵌合衔接体蛋白形成复合物,并引导它们靶向一种或多种靶基因内的靶序列,从而增加一个或多个靶基因的表达。这样的方法可以进一步包括评估一种或多种靶基因的表达或转录。

[0246] 可选地,可以引入两个或更多个指导RNA,每个设计成靶向靶基因内的不同指导RNA靶序列。例如,可以将2个或更多个、3个或更多个、4个或更多个,或5个或更多个指导RNA设计为靶向单个靶基因。替代地或另外地,可以引入两个或更多个指导RNA,每个被设计为在两个或更多个不同的靶基因中靶向不同的指导RNA靶序列(即复用(multiplexing))。

[0247] 可选地,在其中的嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒,或协同激活介质表达盒(包括嵌合Cas蛋白编码序列和嵌合衔接体蛋白编码序列)包含位于编码序列上游的多腺苷酸化信号或转录终止子,且多腺苷酸化信号或转录终止子的侧接由位点特异性重组酶识别的重组酶识别位点的方法中,该方法可进一步包括将重组酶引入非人动物中。重组酶可以切除多腺苷酸化信号或转录终止子,从而允许下游编码序列的表达。

[0248] 在其中的非人动物已经包含如本文其它地方所述的指导RNA表达盒的一些方法中,该方法可以简单地包括将重组酶引入非人动物中,其中该重组酶切除上游多腺苷酸化信号或转录终止子,从而表达嵌合Cas蛋白和/或嵌合衔接体蛋白,从而增加/激活靶基因的表达/转录。

[0249] 可选地,在其中的非人动物包括嵌合Cas蛋白表达盒但不包括嵌合衔接体蛋白表达盒的方法中,可以将嵌合衔接体蛋白引入非人动物中。同样地,在其中的非人动物包括嵌合衔接体蛋白表达盒但不包括嵌合Cas蛋白表达盒的方法中,可以将嵌合Cas蛋白表达盒引入非人动物中。

[0250] 上文提供的用于评估体内CRISPR/Cas活性的各种方法也可以用于使用如本文其它地方所述的包含Cas表达盒的细胞离体评估CRISPR/Cas活性。

[0251] 指导RNA和可选的重组酶可以如本文其它各处所公开的,通过任何递送方法(例如AAV、LNP或HDD)和任何给药途径,以任何形式(用于指导RNA的DNA或RNA;用于重组酶的DNA、RNA或蛋白质)引入细胞或非人动物。在一些方法中,指导RNA或重组酶可以以组织特异性方式引入。在特定方法中,递送是通过AAV介导的递送。例如,如果靶向肝,则可以使用AAV8。同样,如果非人动物或细胞包含嵌合Cas蛋白表达盒但不包含嵌合衔接体蛋白表达盒,则嵌合

衔接体蛋白可以以任何形式 (DNA、RNA, 或蛋白质) 通过任何递送方法 (例如AAV、LNP, 或HDD) 和本文其它各处公开的任何给药途径引入细胞或非人动物。或者, 如果非人动物或细胞包含嵌合衔接体蛋白表达盒但不包含嵌合Cas蛋白表达盒, 则嵌合Cas蛋白可以以任何形式 (DNA、RNA或DNA), 通过任何递送方法 (例如AAV、LNP或HDD) 和任何给药途径 (如本文其它地方所公开的) 引入细胞或非人动物中。

[0252] 用于评估靶基因组基因座的增加的转录或表达的方法在本文其它地方提供, 并且是众所周知的。评估可以是本文其它各处所公开的任何细胞类型、任何组织类型或任何器官类型。在一些方法中, 例如通过评估肝细胞中靶基因组基因座表达的分泌蛋白的血清水平来评估靶基因在肝细胞中的表达。如果靶基因编码具有特定酶活性的蛋白质, 则评估可以包括测量靶基因的表达和/或靶基因编码的蛋白质的活性。替代地或另外地, 评估可以包括评估在从非人动物分离的一种或多种细胞中的表达。评估可包括从非人动物分离靶器官或组织, 并评估靶基因在靶器官或组织中的表达。评估还可包括评估靶器官或组织内两种或更多种不同细胞类型中靶基因的表达。类似地, 评估可包括从非人动物分离非靶器官或组织 (例如两个或更多个非靶器官或组织), 并评估靶基因在非靶器官或组织中的表达。

[0253] 在一些方法中, 靶基因可以是如本文其它地方所述的疾病相关基因。例如, 靶基因可以是与蛋白质聚集疾病或病症相关的基因。作为特定实例, 靶基因可以是与蛋白质聚集疾病或病症相关的基因 (例如Ttr), 并且该方法可以包括增加该靶基因的表达以模拟蛋白质聚集疾病或病症。在某些特定方法中, 靶基因可以是Ttr。可选地, Ttr基因可以包含病理性突变 (例如引起淀粉样变性的突变) 或病理性突变的组合。此类突变的例子提供于例如WO 2018/007871, 出于所有目的通过引用将其全文并入本文。

[0254] 在其它方法中, 靶基因可以是参与与疾病或病症 (例如高胆固醇血症或动脉粥样硬化) 相关的途径的一种。在某些特定方法中, 靶基因可以是Pcsk9或Ldlr。在其它方法中, 靶基因可以是当过表达时可以模拟此类疾病或病症的基因。例如, 靶基因可以是Pcsk9, 并且该方法可以包括增加Pcsk9的表达以模拟高胆固醇血症。

B. 优化CRISPR/Cas体内或体外增加靶基因表达能力的方法

[0255] 提供了多种方法来优化CRISPR/Cas向细胞或非人动物的递送或优化CRISPR/Cas的体内转录激活活性。这样的方法可以包括, 例如: (a) 在第一非人动物或第一细胞中第一次进行如上所述的测试CRISPR/Cas修饰靶基因组基因座的能力的方法; (b) 更改变量, 并在具有更改后的变量的第二非人动物 (即同一物种) 或第二细胞中第二次进行该方法; 以及 (c) 比较步骤 (a) 中靶基因的表达/转录与步骤 (b) 中靶基因的表达/转录, 并选择导致靶基因最高表达/转录的方法。

[0256] 替代地或附加地, 可以选择导致最高功效、最高一致性, 或最高特异性的方法。较高的功效是指靶基因的表达/转录水平更高 (例如在特定靶细胞类型内、特定靶组织内, 或特定靶器官内靶向较高百分比的细胞)。较高的一致性是指如果靶向多种类型的细胞、组织, 或器官, 则靶基因在不同类型的靶细胞、组织, 或器官之间的表达/转录更加一致地增加 (例如靶器官中更多的细胞类型的表达/转录增加)。如果靶向特定器官, 则较高的一致性也可以指器官内所有位置的表达/转录更加一致的增加。更高的特异性可以指相对于靶基因或被靶向的基因的更高特异性、相对于靶定的细胞类型的更高特异性、相对于靶定的组织类型的更高特异性, 或相对于靶定的器官的更高特异性。例如, 增加的靶特异性是指其它基

因上的脱靶作用较小(例如,代替或除了修饰靶基因组基因座外,在非预期的脱靶基因组基因座(例如邻近基因组基因座)上转录增加的靶细胞的百分比更低)。同样,增加的细胞类型、组织,或器官类型特异性是指如果特定细胞类型、组织类型,或器官类型被靶向,则脱靶细胞类型、组织类型,或器官类型中的作用(即增加的表达/转录)较少。(例如,当特定器官被靶向(例如肝脏)时,在并非预期靶标的器官或组织中的细胞中的作用(即增加的表达/转录)较少)。

[0257] 改变的变量可以是任何参数。作为一个例子,改变的变量可以是包装或递送方法,通过该包装或递送方法将指导RNA(或可选地重组酶或其它组分)引入细胞或非人动物中。递送方法(例如LNP、HDD,和AAV)的例子在本文其它地方公开。例如,改变的变量可以是AAV血清型。作为另一个例子,改变的变量可以是用于将指导RNA(或可选地重组酶或其它成分)引入细胞或非人动物的给药途径。给药途径的实例(例如静脉内,玻璃体内、实质内和鼻内滴注)在本文其它地方公开。

[0258] 作为另一个例子,改变的变量可以是引入的指导RNA(或可选地,重组酶或其它组分)的浓度或量。作为另一个例子,改变的变量可以是引入指导RNA(或可选地,重组酶或其它组分)的次数或频率。作为另一个例子,改变的变量可以是指导RNA(或可选地,重组酶或其它组分)被引入的形式。例如,指导RNA可以以DNA的形式或以RNA的形式被引入。类似地,指导RNA(或可选地,重组酶或其它组分)可包含修饰的各种组合,以实现稳定性,减少脱靶效应,促进递送等。作为另一个例子,改变的变量可以是被引入的指导RNA的序列(例如,引入以不同的序列的不同的指导RNA)。

C. 将指导RNA和其它成分引入细胞和非人动物

[0259] 本文公开的方法包括将一个或多个如本文其它地方所述的指导RNA、指导RNA阵列、重组酶,或其它组分引入细胞或非人动物中。“引入”包括以这样的方式向细胞或非人动物呈现核酸或蛋白质,使得该核酸或蛋白质能够进入细胞内部或非人动物内的细胞内部。引入可以通过任何方式完成,并且可以将两个或更多个组分(例如组分中的两个或全部的组分)同时或以任何组合顺序引入细胞或非人动物中。例如,可以在引入第二指导RNA之前将第一指导RNA引入细胞或非人动物中。另外,可以通过相同的递送方法或不同的递送方法将两个或更多组分引入细胞或非人动物中。类似地,可以通过相同的给药途径或不同的给药途径将两个或更多组分引入非人动物。

[0260] 指导RNA可以以RNA的形式(例如体外转录的RNA)或以编码该指导RNA的DNA的形式被引入细胞。同样地,诸如重组酶的蛋白质组分可以以DNA、RNA,或蛋白质的形式引入细胞。当以DNA的形式引入时,编码指导RNA的DNA可以可操作地连接至在细胞中有活性的启动子。例如,指导RNA可以经由AAV递送并在U6启动子下在体内表达。这样的DNA可以在一个或多个表达构建体中。例如,这样的表达构建体可以是单个核酸分子的组分。替代地,它们可以以任何组合在两个或多个核酸分子之间分离(即,编码一个或多个CRISPR RNA的DNA和编码一个或多个tracrRNA的DNA可以是分离的核酸分子的组分)。

[0261] 编码指导RNA或重组酶(或其它组分)的核酸可与表达构建体中的启动子可操作地连接。表达构建体包括能够引导基因或其它目的核酸序列的表达并且可以将这种目的核酸序列转移至靶细胞的任何核酸构建体。可以在表达构建体中使用的合适的启动子包括在例如真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细

胞、大鼠细胞、仓鼠细胞、兔细胞、多能细胞、胚胎干(ES)细胞、成年干细胞、发育受限的祖细胞、诱导性多能干(iPS)细胞,或单细胞期胚胎中的一个或多个中有活性的启动子。这样的启动子可以是例如条件启动子、诱导型启动子、组成型启动子,或组织特异性启动子。可选地,启动子可以是双向启动子,其在一个方向上驱动指导RNA的表达,而在另一方向上驱动另一组分的表达。这种双向启动子可以由(1)完整的、常规的单向Po1 III启动子,该启动子包含3个外部控制元件:远端序列元件(DSE)、近端序列元件(PSE)和TATA盒;和(2)第二基本Po1 III启动子,其包括PSE和反向融合到DSE的5'末端的TATA盒组成。例如,在H1启动子中,DSE与PSE和TATA盒相邻,并且可以通过创建杂交启动子,使启动子为双向;在该杂交启动子中,通过附加衍生自U6启动子的PSE和TATA盒来控制反向的转录。参见,例如,US 2016/0074535,出于所有目的通过引用将其全部内容合并于此。使用双向启动子来同时表达编码指导RNA的基因和另一种组分允许紧凑表达盒的产生,以促进递送。

[0262] 指导RNA或编码指导RNA(或其它组分)的核酸可以在包含载体的组合物中提供,所述载体增加了指导RNA的稳定性(例如,在给定的储存条件下(例如-20°C、4°C,或环境温度)延长降解产物的含量低于阈值(例如低于起始核酸或蛋白质的重量的0.5%)的时间;或增加体内稳定性)。此类载体的非限制性实例包括聚(乳酸)(PLA)微球、聚(D,L-乳酸-甘醇酸)(poly(D,L-lactic-coglycolic-acid), PLGA)微球、脂质体、胶束、反胶束、脂质卷),和脂质微管。

[0263] 本文提供了各种方法和组合物,以允许将核酸或蛋白质引入细胞或非人动物中。将核酸引入各种细胞类型的方法是本领域已知的,包括例如稳定的转染方法、瞬时转染方法和病毒介导的方法。

[0264] 转染方案以及用于将核酸序列引入细胞的方案是多样的。非限制性转染方法包括基于化学品的转染方法,其使用脂质体;纳米粒子;磷酸钙(Graham et al.(1973)Virology 52(2):456-67, Bacchetti et al.(1977)Proc.Natl.Acad.Sci.USA 74(4):1590-4,和 Kriegler,M(1991).Transfer and Expression:A Laboratory Manual.New York:W.H.Freeman and Company.pp.96-97);树状聚合物;或阳离子聚合物(例如DEAE-葡聚糖或聚乙烯亚胺)。非化学方法包括电穿孔、声纳穿孔,和光学转染。基于粒子的转染包括使用基因枪或磁辅助转染(Bertram(2006)Current Pharmaceutical Biotechnology 7,277-28)。病毒方法也可以用于转染。

[0265] 还可以通过电穿孔、通过卵细胞质内注射(intracytoplasmic injection)、通过病毒感染、通过腺病毒、通过腺相关病毒、通过慢病毒、通过逆转录病毒、通过转染、通过脂质介导的转染,或通过核转染将人TRKB靶向试剂引入细胞。核转染是一种改进的电穿孔技术,其使核酸底物不仅可以传递到细胞质,还可以通过核膜递送到细胞核。另外,在本文公开的方法中使用核转染通常比常规电穿孔需要更少的细胞(例如仅需约200万,而常规电穿孔则需700万)。在一实施例中,使用LONZA[®] NUCLEOFECTOR[™]系统进行核转染。

[0266] 也可以通过显微注射来将核酸或蛋白引入细胞(例如合子)。在合子(即单细胞期胚胎)中,显微注射可以注入母系和/或父系原核或注入细胞质中。如果显微注射仅进入一个原核,则父系原核因其尺寸较大是优选的。替代地,显微注射可以通过注射到细胞核/原核和细胞质中执行:可以先将针头引入细胞核/原核,然后可以注射第一量,然后当从单细胞期胚胎中取出针头时,可将第二量注入细胞质中。进行显微注射的方法是众所周知的。参

见,例如,Nagy et al. (Nagy A, Gertsenstein M, Vintersten K, Behringer R., 2003, *Manipulating the Mouse Embryo*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); 还参见 Meyer et al. (2010) *Proc. Natl. Acad. Sci. USA* 107:15022-15026 和 Meyer et al. (2012) *Proc. Natl. Acad. Sci. USA* 109:9354-9359。

[0267] 将核酸或蛋白质引入细胞或非人动物的其它方法可包括,例如,载体递送、颗粒介导的递送、外体介导的递送、脂质纳米颗粒介导的递送、细胞穿膜肽介导的递送,或可植入装置介导的递送。作为具体实例,可以将核酸或蛋白质在诸如聚(乳酸)(PLA)微球、聚(D,L-乳酸-甘醇酸)(PLGA)微球、脂质体、胶束、反胶束、脂质卷,或脂质微管的载体中引入细胞或非人动物。递送至非人动物的一些具体实例包括流体动力递送、病毒介导的递送(例如,腺相关病毒(AAV)介导的递送),和脂质纳米颗粒介导的递送。

[0268] 核酸和蛋白质向细胞或非人动物中的引入可以通过流体动力递送(HDD)来完成。流体动力递送已经作为体内的细胞内DNA递送的方法出现。为了将基因递送至实质细胞,仅需要通过选定的血管注射必需的DNA序列,从而消除了与当前的病毒载体和合成载体有关的安全隐患。当注射入血液时,DNA能够到达血液可及的不同组织中的细胞。流体动力递送利用通过将大量溶液快速注射到循环中不可压缩的血液中所产生的力来克服内皮和细胞膜的物理屏障,该物理屏障阻止了大的和膜不可渗透的化合物进入实质细胞。除了DNA的递送外,此方法还可用于体内的RNA、蛋白质,和其它小化合物的有效细胞内递送。参见,例如, Bonamassa et al. (2011) *Pharm. Res.* 28(4):694-701,出于所有目的通过引用整体并入本文。

[0269] 还可以通过病毒介导的递送(例如AAV介导的递送或慢病毒介导的递送)来实现核酸的引入。其它示例性病毒/病毒载体包括逆转录病毒、腺病毒、痘苗病毒、痘病毒,和单纯疱疹病毒。病毒可以感染分裂细胞、非分裂细胞,或分裂和非分裂细胞两者。病毒可以整合到宿主基因组中,也可以不整合到宿主基因组中。还可以对此类病毒进行工程化,使其免疫力降低。病毒可以是可复制型的,也可以是复制缺陷型的(例如,在另一轮病毒体复制和/或包装中必需的一个或多个基因中有缺陷)。病毒可引起瞬时表达、长期表达(例如至少1周、2周、1个月、2个月,或3个月),或永久表达(例如Cas9和/或gRNA的表达)。示例性病毒滴度(例如AAV滴度)包括 10^{12} 、 10^{13} 、 10^{14} 、 10^{15} ,和 10^{16} 载体基因组/mL。

[0270] ssDNA AAV基因组由两个开放阅读框Rep和Cap组成,其侧接允许合成互补DNA链的两个反向末端重复序列。当构建AAV转移质粒时,将转基因置于两个ITR之间,并且Rep和Cap可以反式提供。除了Rep和Cap之外,AAV还可能需要一个包含来自腺病毒的基因的辅助质粒。这些基因(E4、E2a,和VA)介导AAV复制。例如,可以将转移质粒Rep/Cap和辅助质粒转染到含有腺病毒基因E1+的HEK293细胞中,以产生感染性AAV颗粒。替代地,可以将Rep、Cap,和腺病毒辅助基因合并为一个质粒。类似的包装细胞和方法可用于其它病毒,例如逆转录病毒。

[0271] 已经识别出多种血清型的AAV。这些血清型在它们感染的细胞类型(即它们的向性(tropism))上有所不同,从而允许特定细胞类型的优先转导。CNS组织的血清型包括AAV1、AAV2、AAV4、AAV5、AAV8,和AAV9。心脏组织的血清型包括AAV1、AAV8,和AAV9。肾组织的血清型包括AAV2。肺组织的血清型包括AAV4、AAV5、AAV6,和AAV9。胰腺组织的血清型包括AAV8。感光细胞的血清型包括AAV2、AAV5,和AAV8。视网膜色素上皮组织的血清型包括AAV1、AAV2、

AAV4、AAV5，和AAV8。骨骼肌组织的血清型包括AAV1、AAV6、AAV7、AAV8，和AAV9。肝组织的血清型包括AAV7、AAV8，和AAV9，尤其是AAV8。

[0272] 可以通过假型(pseudotyping)进一步完善向性；假型是来自不同病毒血清型的衣壳和基因组的混合。例如，AAV2/5表示一种病毒，其包含包装在血清型5的衣壳中的血清型2的基因组。使用假型病毒可以提高转导效率，并改变向性。衍生自不同血清型的混合衣壳也可用于改变病毒向性。例如，AAV-DJ包含来自八种血清型的杂交衣壳，并且在体内多种细胞类型中显示出高感染性。AAV-DJ8是另一个显示AAV-DJ的属性，但具有增强的大脑摄入的示例。AAV血清型也可以通过突变进行修饰。AAV2的突变修饰的例子包括Y444F、Y500F、Y730F，和S662V。AAV3的突变修饰的例子包括Y705F、Y731F，和T492V。AAV6的突变修饰的例子包括S663V和T492V。其它假型/修饰的AAV变体包括AAV2/1、AAV2/6、AAV2/7、AAV2/8、AAV2/9、AAV2.5、AAV8.2，和AAV/SASTG。

[0273] 为了加速转基因表达，可以使用自身互补的AAV(scAAV)变体。由于AAV依赖于细胞的DNA复制机制来合成AAV单链DNA基因组的互补链，因此转基因表达可能会延迟。为了解决此延迟，可以使用包含能够在感染后自发退火的互补序列的scAAV，从而消除了宿主细胞DNA合成的需求。

[0274] 为了增加包装能力，可以在两个AAV转移质粒之间分裂更长的转基因，第一个具有3'剪接供体，第二个具有5'剪接受体。细胞共同感染(co-infection)后，这些病毒形成多联体，被剪接在一起，全长转基因可以得到表达。尽管这允许更长的转基因表达，但表达效率较低。用于增加能力的类似方法利用同源重组。例如，转基因可以在两个转移质粒之间分开，但是具有实质的序列重叠，使得共同表达诱导全长转基因的同源重组和表达。

[0275] 核酸和蛋白质的引入也可以通过脂质纳米颗粒(LNP)介导的递送来实现。例如，LNP介导的递送可以用于递送RNA形式的指导RNA。通过这种方法的递送导致了指导RNA的瞬时存在，并且可生物降解的脂质提高了清除率，提高了耐受性并降低了免疫原性。脂质制剂可以保护生物分子免于降解，同时改善其细胞摄取。脂质纳米颗粒是包含通过分子间力彼此物理结合的多个脂质分子的颗粒。这些包括微球(包括单层和多层囊泡，例如脂质体)、乳液中的分散相、胶束或悬浮液中的内相。此类脂质纳米颗粒可用于封装一种或多种核酸或蛋白质以进行递送。包含阳离子脂质的制剂在用于递送聚阴离子(如核酸)时是有用的。可以包括的其它脂质是中性脂质(即不带电荷的或两性离子脂质)、阴离子脂质、增强转染的辅助脂质(helper lipid)，以及隐形脂质(stealth lipid)，其增加了纳米粒子可以在体内存在的时间长度。合适的阳离子脂质、中性脂质、阴离子脂质、辅助脂质，和隐形脂质的实例可以在W0 2016/010840 A1和W0 2017/173054A1中找到，其每一篇出于所有目的通过引用整体并入本文。示例性的脂质纳米颗粒可包含阳离子脂质和一种或多种其它组分。在一个实例中，其它组分可包含辅助脂质，例如胆固醇。在另一个实例中，其它组分可以包含辅助脂质，例如胆固醇和诸如DSPC的中性脂质。在另一个实例中，其它组分可以包含辅助脂质(例如胆固醇)、可选的中性脂质(例如DSPC)，以及隐形脂质(例如S010、S024、S027、S031，或S033)。

[0276] LNP可包含以下一种或多种或全部：(i) 用于包封和用于内体逸出的脂质；(ii) 稳定的中性脂质；(iii) 用于稳定的辅助脂质；(iv) 隐形脂质。参见，例如，Finn et al. (2018) Cell Reports 22:1-9和W0 2017/173054 A1，出于所有目的，通过引用将其全部内容整体

并入本文。在某些LNP中,运货(cargo)可包含指导RNA或编码指导RNA的核酸。在某些LNP中,运货可包含编码Cas核酸酶(例如Cas9)的mRNA以及指导RNA或编码指导RNA的核酸。

[0277] 用于包封和内体逸出的脂质可以是阳离子脂质。脂质也可以是可生物降解的脂质,例如可生物降解的可电离脂质。合适脂质的一个实例是脂质A或LP01,其为(9Z,12Z)-3-((4,4-双(辛氧基氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基十八烷基-9,12-二烯酸酯,也称为3-((4,4-双(辛基氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基(9Z,12Z)-十八烷基-9,12-二烯酸酯。参见,例如,Finn et al. (2018) Cell Reports 22:1-9和WO 2017/173054 A1,出于所有目的,通过引用将其每一篇的全部内容整体并入本文。合适的脂质的另一个实例是脂质B,其为((5-((二甲基氨基)甲基)-1,3-亚苯基)双(氧基))双(辛烷-8,1-二基)双(癸酸酯),也称为((5-((二甲基氨基)甲基)-1,3-亚苯基)双(氧基))双(辛烷-8,1-二基)双(癸酸酯)。合适的脂质的另一个例子是脂质C,其为2-((4-((3-(二甲基氨基)丙氧基)羰基)氧基)-十六烷酰基)氧基)丙烷-1,3-二基(9Z,9'Z,12Z,12'Z)-双(十八烷基-9,12-二烯酸酯)。合适的脂质的另一个实例是脂质D,其为3-(((3-(二甲基氨基)丙氧基)羰基)氧基)-13-(辛酰氧基)十三烷基-3-辛基十一酸酯。

[0278] 适用于本文所述的LNP的一些此类脂质在体内是可生物降解的。例如,包含这种脂质的LNP包括其中至少75%的脂质在8、10、12、24,或48小时或3、4、5、6、7,或10天内从血浆中清除的那些。作为另一个例子,50%的LNP在8、10、12、24,或48小时,或3、4、5、6、7,或10天之内从血浆中清除。

[0279] 此类脂质可以根据它们所在的介质的pH值而为可电离的。例如,在弱酸性介质中,脂质可以被质子化并因此带有正电荷。相反,在轻微弱碱性的介质中,例如pH值约为7.35的血液中,脂质可能不会质子化,因此不带电荷。在一些实施方案中,脂质可以在至少约9、9.5,或10的pH下质子化。这种脂质带有电荷的能力与其固有的pKa有关。例如,脂质可以独立地具有约5.8至约6.2的pKa。

[0280] 中性脂质起到稳定和改善LNP加工的作用。合适的中性脂质的实例包括各种中性、不带电或两性离子脂质。适用于本公开的中性磷脂的实例包括,但不限于,5-七癸基苯-1,3-二醇(间苯二酚)、二棕榈酰磷脂酰胆碱(DPPC)、二硬脂酰磷脂酰胆碱(DSPC)、磷酸胆碱(DOPC)、二肉豆蔻酰磷脂酰胆碱(DMPC)、磷脂酰胆碱(PLPC)、1,2-双硬脂酰-sn-甘油-3-磷酸胆碱(DAPC)、磷脂酰乙醇胺(PE)、卵磷脂酰胆碱(EPC)、二月桂酰磷脂酰胆碱(DLPC)、二肉豆蔻酰磷脂酰胆碱(DMPC)、1-肉豆蔻酰基-2-棕榈酰基磷脂酰胆碱(MPPC)、1-棕榈酰基-2-肉豆蔻酰基磷脂酰胆碱(PMPC)、1-棕榈酰基-2-硬脂酰基磷脂酰胆碱(PSPC)、1,2-二硬脂酰基-sn-甘油-3-磷酸胆碱(DBPC)、1-硬脂酰基-2-棕榈酰基磷脂酰胆碱(SPPC)、1,2-双二十碳烯酰基(dieicosenoyl)-sn-甘油基-3-磷酸胆碱(DEPC)、棕榈酰基油酰基磷脂酰胆碱(POPC)、溶血磷脂酰胆碱、二油酰基磷脂酰乙醇胺(DOPE)、二亚油酰基磷脂酰胆碱二硬脂酰基磷脂酰乙醇胺(DSPE)、二豆蔻酰基磷脂酰乙醇胺(DMPE)、二棕榈酰基磷脂酰乙醇胺(DPPE)、棕榈酰基油酰基磷脂酰乙醇胺(POPE)、溶血磷脂酰乙醇胺,以及它们的组合。

[0281] 辅助脂质包括增强转染的脂质。辅助脂质提高转染的机制可能包括例如提高粒子的稳定性。在一些情况下,辅助脂质可以改善膜的融合性(fusogenicity)。辅助脂质包括类固醇、固醇,和烷基邻苯二酚。合适的辅助脂质的实例包括胆固醇、5-十七烷基间苯二酚,和

胆固醇半琥珀酸酯。在一实例中,辅助脂质可以是胆固醇或胆固醇半琥珀酸酯。

[0282] 隐形脂质包括改变纳米颗粒可以在体内存在的时间长度的脂质。隐形脂质可以通过例如减少颗粒聚集和控制颗粒大小来辅助配制过程。隐形脂质可能会调节LNP的药代动力学特性。合适的隐形脂质包括具有连接至脂质部分的亲水头部基团的脂质。

[0283] 隐形脂质的亲水性头部基团可包括,例如,选自基于PEG的聚合物的聚合物部分(有时称为聚(环氧乙烷))、聚(恶唑啉)、聚(乙烯醇)、聚(甘油)、聚(N-乙基吡咯烷酮)、聚氨基酸,和聚N-(2-羟丙基)甲基丙烯酰胺。术语“PEG”是指任何聚乙二醇或其它聚亚烷基醚聚合物。在某些LNP制剂中,PEG是PEG-2K,也称为PEG 2000,其平均分子量为约2,000道尔顿。参见,例如,WO 2017/173054 A1,出于所有目的通过引用将其全部内容合并于此。

[0284] 隐形脂质的脂质部分可以衍生自,例如,二酰基甘油或二酰基甘油酰胺,包括那些包含烷基链的二烷基甘油或二烷基甘油酰胺基团的分子,其中该烷基链的长度独立地包含约C4至约C40饱和或不饱和碳原子,其中该链可包含一个或多个官能团,例如酰胺或酯。二烷基甘油或二烷基甘油酰胺基可进一步包含一个或多个取代的烷基。

[0285] 作为一个实例,隐形脂质可以选自PEG-二月桂酰基甘油、PEG-二肉豆蔻酰基甘油(PEG-DMG)、PEG-二棕榈酰基甘油、PEG-二硬脂酰基甘油(PEG-DSPE)、PEG-二月桂基甘油酰胺、PEG-二肉豆蔻基甘油酰胺、PEG-二棕榈酰基甘油酰胺,和PEG-二硬脂酰基甘油酰胺、PEG-胆固醇(1-[8'-(胆甾-5-烯-3 β -氧基)羧酰胺-3',6'-二氧辛基]氨基甲酰基- ω -甲基-聚(乙二醇)、PEG-DMB(3,4-二十四烷氧基苄基- ω -甲基-聚(乙二醇)醚)、1,2-二肉豆蔻酰基-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000](PEG2k-DMG)、1,2-二硬脂酰基-sn-甘油-3-磷酸乙醇胺-N-[甲氧基(聚乙二醇)-2000]

(PEG2k-DSPE)、1,2-二硬脂酰基-sn-甘油、甲氧基聚乙二醇(PEG2k-DSG)、聚(乙二醇)-2000-二甲基丙烯酸酯(PEG2k-DMA),和1,2-二硬脂基氧基丙基-3-胺-N-[甲氧基(聚乙二醇)-2000](PEG2k-DSA)。在一个特定实例中,隐形脂质可以是PEG2k-DMG。

[0286] LNP可以在制剂中包含组分脂质的各自不同的摩尔比。CCD脂质的摩尔%可以是例如约30摩尔%至约60摩尔%、约35摩尔%至约55摩尔%、约40摩尔%至约50摩尔%、约42摩尔%至约47摩尔%,或约45%。辅助脂质的摩尔%可以是例如约30摩尔%至约60摩尔%、约35摩尔%至约55摩尔%、约40摩尔%至约50摩尔%、约41摩尔%至约46摩尔%,或约44摩尔%。中性脂质的摩尔%可以是例如约1摩尔%至约20摩尔%、约5摩尔%至约15摩尔%、约7摩尔%至约12摩尔%,或约9摩尔%。隐形脂质的摩尔%可以是例如约1摩尔%至约10摩尔%、约1摩尔%至约5摩尔%、约1摩尔%至约3摩尔%、约2摩尔%,或约1摩尔%。

[0287] LNP在要被包封的核酸的可生物降解脂质的带正电荷的胺基(N)和带负电荷的磷酸基(P)之间可以具有不同的比率。这可在数学上由等式N/P表示。例如,N/P之比可为约0.5至约100、约1至约50、约1至约25、约1至约10、约1至约7、约3至5、约4至约5、约4、约4.5,或约5。

[0288] 在一些LNP中,运货可以包含Cas mRNA和gRNA。Cas mRNA和gRNA的比例可以不同。例如,LNP制剂可包括Cas mRNA与gRNA核酸之比为约25:1至约1:25、约10:1至约1:10、约5:1至约1:5,或约1:1。替代地,LNP制剂可以包括约1:1至约1:5,或约10:1的Cas mRNA与gRNA核酸之比。备选地,LNP制剂可以包括约1:10、25:1、10:1、5:1、3:1、1:1、1:3、1:5、1:10或1:25的Cas mRNA与gRNA核酸之比。

[0289] 合适的LNP的具体实例的氮磷(N/P)比为4.5,并且以45:44:9:2的摩尔比包含可生物降解的阳离子脂质、胆固醇、DSPC,和PEG2k-DMG。可生物降解的阳离子脂质可以是(9Z,12Z)-3-((4,4-双(辛氧基氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基十八烷基-9,12-二烯酸酯,也称为3-((4,4-双(辛基氧基)丁酰基)氧基)-2-(((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙基(9Z,12Z)-十八烷基-9,12-二烯酸酯。参见,例如,Finn et al.(2018)Cell Reports 22:1-9,出于所有目的,通过引用将其全部内容整体并入本文。合适的脂质的另一个实例是脂质B,其为((5-((二甲基氨基)甲基)-1,3-亚苯基)双(氧基))双(辛烷-8,1-二基)双(癸酸酯),也称为((5-((二甲基氨基)甲基)-1,3-亚苯基)双(氧基))双(辛烷-8,1-二基)双(癸酸酯)。合适的脂质的另一个例子是脂质C,其为2-((4-(((3-(二甲基氨基)丙氧基)羰基)氧基)-十六烷酰基)氧基)丙烷-1,3-二基(9Z,9'Z,12Z,12'Z)-双(十八烷基-9,12-二烯酸酯)。合适的脂质的另一个实例是脂质D,其为3-(((3-(二甲基氨基)丙氧基)羰基)氧基)-13-(辛酰氧基)十三烷基-3-辛基十一酸酯。其它合适的脂质包括庚二酸-6,9,28,31-四烯-19-基-4-(二甲基氨基)丁酸酯(也称为Dlin-MC3-DMA(MC3))。

[0290] 合适的LNP的另一个具体实例的氮磷(N/P)比为6,并且以50:38:9:3的摩尔比包含可生物降解的阳离子脂质、胆固醇、DSPC,和PEG2k-DMG。可生物降解的阳离子脂质可以是(9Z,12Z)-3-((4,4-双(辛氧基氧基)丁酰基)氧基)-2-(((3-(二乙氨基)丙氧基)羰基)氧基)甲基)丙基十八烷基-9,12-二烯酸酯,也称为3-((4,4-双(辛基氧基)丁酰基)氧基)-2-(((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙基(9Z,12Z)-十八烷基-9,12-二烯酸酯。

[0291] 可以选择递送方式以降低免疫原性。例如,可以通过不同的模式来递送不同的组分(例如双模式的递送)。这些不同的模式可以赋予对象递送的分子不同的药效学或药代动力学性质。例如,不同的模式可以导致不同的组织分布、不同的半衰期,或不同的时间分布。某些递送模式(例如通过自主复制或基因组整合而保留在细胞中的核酸载体的递送)导致分子更持久的表达和存在,而其它递送方式是瞬时的且持久性较低(例如递送RNA或蛋白质)。以更瞬时的方式递送组分(例如RNA)可以确保Cas/gRNA复合物仅在短时间内存在并起作用,并且可以降低免疫原性。这种瞬时传递也可以减少脱靶修饰的可能性。

[0292] 体内给药可以通过任何合适的途径进行,包括,例如,肠胃外、静脉内、口服、皮下、动脉内、颅内、鞘内、腹膜内、局部、鼻内,或肌肉内。全身性给药方式包括,例如,口服和肠外途径。肠外途径的实例包括静脉内、动脉内、骨内(intraosseous),肌肉内、皮内、皮下、鼻内,和腹膜内途径。一个具体的例子是静脉输液。局部给药方式包括,例如,鞘内、脑室内、实质内(例如局部实质内递送至纹状体(例如进入尾状核(caudate)或壳核)、大脑皮层、中央前回、海马(例如进入齿状回或CA3区)、颞叶皮层(temporal cortex)、杏仁核、额叶皮层(frontal cortex)、丘脑、小脑、延髓、下丘脑、顶盖、被盖,或黑质)、眼内、眶内、结膜下、玻璃体内、视网膜下,和巩膜途径。与全身给药(例如静脉内)相比,当局部给药(例如实质内或玻璃体内)时,显著更少量的组分(与全身途径相比)即可以发挥作用。局部给药方式还可以减少或消除全身给药治疗有效量的组分时可能发生的潜在毒性副作用的发生。

[0293] 体内给药可以通过任何合适的途径进行,包括,例如,肠胃外、静脉内、口服、皮下、动脉内、颅内、鞘内、腹膜内、局部、鼻内,或肌肉内。一个具体的例子是静脉输液。鼻滴注和玻璃体内注射是其它的具体实例。可以使用一种或多种生理和药学上可接受的载体、稀释

剂、赋形剂,或助剂配制包含指导RNA(或编码指导RNA的核酸)的组合物。该制剂可以取决于选择的给药途径。术语“药学上可接受的”是指载体、稀释剂、赋形剂,或助剂与制剂的其它成分相容并且对其接受者基本上无害。

[0294] 给药频率和剂量数目可以取决于外源供体核酸或指导RNA(或编码该指导RNA的核酸)的半衰期和给药途径以及其它因素。可以在一段时间内进行一次或多次的将核酸或蛋白质引入细胞或非人动物中。例如,引入可以在一段时间内至少执行两次、在一段时间内至少执行三次、在一段时间内至少执行四次、在一段时间内至少执行五次、在一段时间内至少执行六次、在一段时间内至少执行七次、在一段时间内至少执行八次、在一段时间内至少执行九次、在一段时间内至少执行十次、至少执行十一次、在一段时间内至少执行十二次、在一段时间内至少执行十三次、在一段时间内至少执行十四次、在一段时间内至少执行十五次、在一段时间内至少执行十六次、在一段时间内至少执行十七次、在一段时间内至少执行十八次、在一段时间内至少执行十九次,或在一段时间内至少执行二十次。E. 体内测量CRISPR/Cas的活性并评估靶基因的表达

[0295] 本文公开的方法可以进一步包括评估靶基因的表达。测量表达或活性的方法将取决于被修饰的靶基因。

[0296] 例如,如果靶基因包括编码RNA或蛋白质的基因,则评估表达的方法可包括测量编码的RNA和/或蛋白质的表达或活性。例如,如果编码的蛋白质是释放到血清中的蛋白质,则可以测量编码的蛋白质的血清水平。用于测量RNA和蛋白质的水平和活性的测定是众所周知的。

[0297] 可以在来自任何组织或器官的任何细胞类型中评估靶基因在非人动物中的表达。例如,可以在来自相同组织或器官的多种细胞类型或在来自组织或器官内的多个位置的细胞中评估靶基因的表达。这可以提供关于靶组织或器官内的哪些细胞类型被靶向或CRISPR/Cas达到并修改了组织或器官的哪些部分的信息。作为另一个例子,可以在多种类型的组织或多个器官中评估靶基因的表达。在将特定组织或器官被靶定的方法中,这可以提供有关该组织或器官被靶向的效率以及在其它组织或器官中是否存在脱靶效应的信息。

IV. 制作包含Cas表达盒和/或重组酶表达盒的非人动物的方法

[0298] 提供了各种方法来制备非人动物,其包含如本文别处所公开的协同激活介质(SAM)表达盒(包含嵌合Cas蛋白编码序列和嵌合衔接体蛋白表达编码序列)、指导RNA表达盒,和重组酶表达中的一个或多个或全部。同样,提供了各种方法来制备非人动物,其包含如本文别处所公开的嵌合Cas蛋白表达盒、嵌合衔接体蛋白表达盒、指导RNA表达盒,和重组酶表达中的一个或多个或全部。用于生产遗传修饰生物的任何方便的方法或方案都适用于生产这种遗传修饰的非人动物。参见,例如,Cho et al. (2009) *Current Protocols in Cell Biology* 42:19.11:19.11.1-19.11.22和Gama Sosa et al. (2010) *Brain Struct. Funct.* 214(2-3):91-109,出于所有目的,将每一篇均通过引用整体并入本文。这样的遗传修饰的非人动物可以例如通过在靶基因座(例如安全港基因座,如Rosa26)处的基因敲入或通过使用随机整合的转基因来产生。参见,例如,WO 2014/093622和WO 2013/176772,出于所有目的,将它们的每一篇通过引用整体并入本文。在例如US 2012/0017290、US 2011/0265198,和US 2013/0236946中描述了将构建体靶向Rosa26基因座的方法,出于所有目的将其每一篇的全部内容通过引用并入本文。

[0299] 例如,产生包含本文中其它地方公开的一个或多个或所有表达盒的非人动物的方法可包括:(1)修饰多能细胞的基因组以包含一个或多个或全部的表达盒;(2)识别或选择包含一个或多个或所有所述表达盒的遗传修饰的多能细胞;(3)将该遗传修饰的多能细胞导入非人类物宿主胚胎中;以及(4)将宿主胚胎植入代孕母体中并在代孕母体中孕育宿主胚胎。例如,产生包含本文中其它地方公开的一个或多个或所有表达盒的非人动物的方法可包括:(1)修饰多能细胞的基因组以包含一个或多个或所有表达盒;(2)识别或选择包含一个或多个或所有表达盒的遗传修饰的多能细胞;(3)将经遗传修饰的多能细胞导入非人动物宿主胚胎中;(4)在代孕母体中孕育宿主胚胎。可选地,可以将包含修饰的多能细胞(例如非人ES细胞)的宿主胚胎培育至胚泡阶段,然后将其植入代孕母体并在代孕母体中孕育,以产生F0非人动物。代孕母体然后可以产生包含一个或多个或全部表达盒的F0代非人动物。

[0300] 所述方法可以进一步包括识别具有修饰的靶基因组基因座的细胞或动物。可以使用多种方法来识别具有靶向遗传修饰的细胞和动物。

[0301] 修饰基因组的步骤可以例如利用外源供体核酸(例如靶向载体)来修饰靶基因组基因座,以包含本文别处所公开的一个或多个或所有表达盒。作为一个实例,靶向载体可以包括靶向靶基因组基因座处的5'靶序列的5'同源臂和和靶向靶基因组基因座处的3'靶序列的3'同源臂。外源修复模板也可以包含核酸插入物,该核酸插入物包括要整合在靶基因组基因座中的DNA片段。靶基因组基因座中的核酸插入物的整合可导致在靶基因组基因座中添加目标核酸序列、在靶基因组基因座中删除目标核酸序列,或替换靶基因组基因座中的目标核酸序列(即删除和插入)。同源臂可以侧接包含本文别处所公开的一个或多个或全部表达盒的插入核酸,以产生靶基因组基因座。

[0302] 外源修复模板可以用于非同源末端连接介导的插入或同源重组。外源修复模板可以包含脱氧核糖核酸(DNA)或核糖核酸(RNA),它们可以是单链或双链的,并且可以是线性或环状形式。例如,修复模板可以是单链寡脱氧核苷酸(ssODN)。

[0303] 外源修复模板还可以包含在未靶向的内源靶基因组基因座处不存在的异源序列。例如,外源修复模板可包含选择盒,例如侧接重组酶识别位点的选择盒。

[0304] 一些外源修复模板包含同源臂。如果外源修复模板还包含核酸插入物,则同源臂可以侧接该核酸插入物。为了便于参考,在本文中同源臂称为5'和3'(即上游和下游)同源臂。该术语涉及外源修复模板内的同源臂与核酸插入物的相对位置。5'和3'同源臂对应于靶基因组基因座内的区域,其在本文分别称为“5'靶序列”和“3'靶序列”。

[0305] 当两个区域彼此共享足够水平的序列同一性以充当同源重组反应的底物时,同源臂和靶序列“对应”或彼此“对应”。术语“同源”包括与相应序列相同或共享序列同一性的DNA序列。给定靶序列与外源供体核酸中发现的相应同源臂之间的序列同一性可以是允许同源重组发生的任何程度的序列同一性。例如,外源供体核酸(或其片段)的同源臂和靶序列(或其片段)共有的序列同一性的量可以为至少50%、55%、60%、65%、70%、75%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的序列同一性,以使这些序列进行同源重组。此外,同源臂和对应靶序列之间的对应同源区域可以为足以促进同源重组的任何长度。在一些靶向载体中,靶基因组基因座中的预期突变包括在侧接同源臂的插入核酸中。

[0306] 在单细胞期胚胎以外的细胞中,外源修复模板可以是“大靶向载体”或“LTVEC”,其包括靶向载体,所述靶向载体包含对应于并衍生自核酸序列的同源臂,该核酸序列大于用于在细胞中进行同源重组的其它方法所通常使用的核酸序列。LTVEC还包括靶向载体,该靶向载体包含具有核酸序列的核酸插入物,该核酸序列大于用于在细胞中进行同源重组的其它方法所通常使用的核酸序列。例如,LTVEC使大基因座的修饰成为可能;由于其尺寸限制,传统的基于质粒的靶向载体无法容纳该大基因座。例如,靶定的基因座可以是(即5'和3'同源臂可以对应于)在不存在核酸酶剂(例如Cas蛋白)诱导的切口或双链断裂时使用常规方法不可靶定的或只能被错误地靶定的或仅以非常低的效率被靶定的细胞的基因座。LTVEC可以是任何长度,通常长度至少为10kb。LTVEC中5'同源臂和3'同源臂的总和通常至少为10kb。

[0307] 筛选步骤可以包括例如用于评估亲代染色体的等位遗传修饰(MOA)的定量测定。例如,可以通过定量PCR,例如实时PCR(qPCR),来进行定量测定。实时PCR可以利用识别靶基因座的第一引物组和识别非靶定参考基因座的第二引物组。引物组可以包含识别扩增的序列的荧光探针。

[0308] 合适的定量测定的其它实例包括荧光介导原位杂交(FISH)、比较基因组杂交、等温DNA扩增、定量杂交至固定探针、INVADER[®]探针、TAQMAN[®]分子信标探针,或ECLIPSE[™]探针技术(参见例如US2005/0144655,其出于所有目的以全文引用方式并入本文)。

[0309] 合适的多能细胞的例子是胚胎干(ES)细胞(例如小鼠ES细胞或大鼠ES细胞)。修饰的多能细胞可通过例如以下产生:通过(a)将一个或多个外源供体核酸(例如靶向载体)引入细胞中,该外源供体核酸包含侧接例如对应于5'和3'靶位点的5'和3'同源臂的插入核酸,其中插入核酸包含本文所公开的一个或多个或全部表达盒;以及(b)识别至少一个细胞,所述细胞在其基因组中包含在靶基因组基因座处整合的插入核酸。替代地,修饰的多能细胞可以例如通过以下产生:通过(a)向细胞中引入:(i)核酸酶剂,其中所述核酸酶剂在靶基因组基因座内的识别位点处诱导切口或双链断裂;以及(ii)一个或多个靶向载体,其包含侧接对应于5'和3'靶位点的5'和3'同源臂的插入核酸,所述5'和3'靶位点位于与识别位点足够接近之处,其中所述插入核酸包含一个或多个或全部的表达盒;以及(c)识别至少一个在靶基因组基因座处包含修饰(例如插入核酸的整合)的细胞。可以使用诱导切口或双链断裂进入所需的识别位点的任何核酸酶剂。合适的核酸酶的实例包括类转录激活因子效应物核酸酶(TALEN)、锌指核酸酶(ZFN)、兆核酸酶,和成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关(Cas)系统(例如CRISPR/Cas9系统)或此类系统的组分(例如CRISPR/Cas9)。参见,例如,US 2013/0309670和US 2015/0159175,出于所有目的将每一篇的全部内容通过引用整体并入本文。

[0310] 供体细胞可以在任何阶段(例如胚泡期或桑椹胚前期(pre-morula stage,即4细胞期或8细胞期))引入宿主胚胎。产生能够通过种系传递遗传修饰的后代。参见,例如,美国专利号7,294,754,其出于所有目的通过引用整体并入本文。

[0311] 替代地,产生本文其它地方所述的非人动物的方法可包括:(1)使用上述修饰多能细胞的方法修饰单细胞期胚胎的基因组,以包含一个或多个或全部的表达盒;(2)选择遗传修饰的胚胎;和(3)将该遗传修饰的胚胎植入代孕母体到并在代孕母体中孕育。替代

地,产生本文其它地方所述的非人动物的方法可包括:(1)使用上述修饰多能细胞的方法修饰单细胞期胚胎的基因组,以包含一个或多个或全部的表达盒;(2)选择遗传修饰的胚胎;和(3)在代孕母体中孕育该遗传修饰的胚胎。产生了能够通过种系传递遗传修饰的后代。

[0312] 核转移技术也可以用于产生非人动物。简而言之,核转移的方法可以包括以下步骤:(1)将卵母细胞去核或提供去核卵母细胞;(2)分离或提供将与该去核卵母细胞结合的供体细胞或细胞核;(3)将细胞或细胞核插入去核卵母细胞中以形成重组细胞;(4)将重组细胞植入动物子宫内形成胚胎;以及(5)允许该胚胎发育。在这种方法中,卵母细胞通常是从死去的动物身上获得的,尽管它们也可以从活体动物的输卵管和/或卵巢中分离。去核之前,卵母细胞可在多种众所周知的培养基中成熟。卵母细胞的去核可以多种公知的方式进行。可以通过融合前在透明带下显微注射供体细胞,将供体细胞或细胞核插入去核卵母细胞以形成重组细胞。可以通过在接触/融合平面上施加直流电脉冲(电融合),通过将细胞暴露于融合促进化学物质(例如聚乙二醇),或通过灭活病毒(例如仙台病毒)来诱导融合。重建的细胞可以在核供体和受体卵母细胞融合之前、之中,和/或之后通过电和/或非电方式激活。激活方法包括电脉冲、化学诱导的电击、精子穿透、卵母细胞中二价阳离子水平的增加,以及卵母细胞中细胞蛋白的磷酸化的降低(通过激酶抑制剂的方式)。活化的重组细胞或胚胎可以在众所周知的培养基中培养,然后转移到动物的子宫中。参见,例如,US

2008/0092249、WO 1999/005266、US 2004/0177390、WO 2008/017234,和美国专利号7,612,250,出于所有目的将其每一篇的全部内容通过引用整体并入本文。

[0313] 本文提供的各种方法允许产生遗传修饰的非人F0动物,其中遗传修饰的F0动物的细胞包含一个或多个或全部的表达盒。已经认识到,取决于用于产生F0动物的方法,F0动物中具有一个或多个或全部的表达盒的细胞数量将有所不同。通过例如VELOCI小鼠[®]方法将供体ES细胞从相应的生物体(例如8细胞期小鼠胚胎)导入到桑椹胚前期胚胎中,从而使F0动物的细胞群体的更高百分比包含具有目的核苷酸序列的细胞,所述目的核苷酸序列包含靶定的遗传修饰。例如,至少50%、60%、65%、70%、75%、85%、86%、87%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%,或100%的非人F0动物的细胞贡献可包括具有靶定修饰的细胞群体。

[0314] 遗传修饰的F0动物的细胞对此处公开的一个或多个或全部的表达盒可以是杂合的,或者对此处公开的一个或多个或全部的表达盒可以是纯合的。

[0315] 出于所有目的,上文或下文所引用的所有专利文件、网站、其它出版物、登录号等均以引用的方式整体并入本文,其程度与将每个单独的项目具体并单独地指示为通过引用并入本文的程度相同。如果序列的不同版本在不同时间与登录号相关联,则意指着在本申请的有效申请日与登录号相关联的版本。有效申请日是指实际的申请日或提及登录号的优先权申请(如有的话)的申请日中较早的日期。同样,如果在不同时间发布了出版物,网站等的不同版本,则除非另有说明,是指在本申请的有效申请日最近发布的版本。除非另外明确指出,否则本发明的任何特征、步骤、元素、实施例,或方面可以与任何其它相结合使用。尽管出于清楚和理解的目的已经通过图示和示例的方式对本发明进行了详细描述,但是显而易见的是,可以在所附权利要求的范围内进行某些改变和修改。

序列简述

[0316] 使用核苷酸碱基的标准字母缩写和氨基酸的三字母代码示出所附序列表中列出

的核苷酸和氨基酸序列。核苷酸序列遵循从序列5'末端开始并向前(即每行从左到右)到3'末端的标准约定。每个核苷酸序列仅显示一条链,但任何对所示出的链的提及均应理解为包括互补链。当提供编码氨基酸序列的核苷酸序列时,应当理解还提供了其编码相同氨基酸序列的密码子简并变体。氨基酸序列遵循从序列的氨基末端开始并向前(即在每行中从左至右)到达羧基末端的标准约定。

[0317] 表3:序列描述。

SEQ ID NO	类型	描述
1	蛋白质	dCas9-VP64 嵌合 Cas 蛋白
2	蛋白质	dCas9 蛋白
3	蛋白质	VP64 转录激活域
4	蛋白质	接头 v1
5	蛋白质	接头 v2
6	蛋白质	MCP-p65-HSF1 嵌合衔接体蛋白
7	蛋白质	MS2 外壳蛋白质(MCP)
8	蛋白质	p65 转录激活域
9	蛋白质	HSF1 转录激活域
10	RNA	crRNA 尾
11	RNA	tracrRNA
12	RNA	gRNA 支架 v1
13	RNA	gRNA 支架 v2
14	RNA	gRNA 支架 v3
15	RNA	gRNA 支架 v4
16	RNA	MS2 结合环
17	DNA	指导 RNA 靶序列加 PAM v1
18	DNA	指导 RNA 靶序列加 PAM v2
19	DNA	指导 RNA 靶序列加 PAM v3
20	蛋白质	T2A
21	蛋白质	P2A
22	蛋白质	E2A
23	蛋白质	F2A
24	DNA	编码 dCas9 蛋白的核酸
25	DNA	编码 dCas9-VP64 嵌合 Cas 蛋白的核酸

SEQ ID NO	类型	描述
26	DNA	编码 MCP 的核酸
27	DNA	编码 MCP-p65-HSF1 嵌合衔接体蛋白的核酸
28	DNA	编码 VP64 转录激活域的核酸
29	DNA	编码 p65 转录激活域的核酸
30	DNA	编码 HSF1 转录激活域的核酸
31	DNA	协同激活介质 (SAM) 双顺反子表达盒(dCas9-VP64-T2A-MCP-p65-HSF1)
32	DNA	通用指导 RNA 阵列表达盒
33	DNA	<i>Ttr</i> 指导 RNA 阵列表达盒
34	DNA	小鼠 <i>Ttr</i> 指导 RNA 靶序列 v1
35	DNA	小鼠 <i>Ttr</i> 指导 RNA 靶序列 v2
36	DNA	小鼠 <i>Ttr</i> 指导 RNA 靶序列 v3
37	RNA	小鼠 <i>Ttr</i> 单指导 RNA v1
38	RNA	小鼠 <i>Ttr</i> 单指导 RNA v2
39	RNA	小鼠 <i>Ttr</i> 单指导 RNA v3
40	RNA	具有 MS2 结合环的 gRNA 支架
41	RNA	小鼠 <i>Ttr</i> 指导 RNA DNA 靶向片段 v1
42	RNA	小鼠 <i>Ttr</i> 指导 RNA DNA 靶向片段 v2
43	RNA	小鼠 <i>Ttr</i> 指导 RNA DNA 靶向片段 v3
44	蛋白质	协同激活介质 (SAM) (dCas9-VP64-T2A-MCP-p65-HSF1)
45	DNA	XBA-TDP fw
46	DNA	XBA-TDP 探针
47	DNA	XBA-TDP rev
48	DNA	Neo fw
49	DNA	Neo 探针
50	DNA	Neo rev
51	DNA	SAM TD fw
52	DNA	SAM TD 探针
53	DNA	SAM TD rev
54	DNA	MS2_T fw
55	DNA	MS2_T 探针
56	DNA	MS2_T rev
57	DNA	P65_T fw
58	DNA	P65_T 探针
59	DNA	P65_T rev
60	DNA	WPRE_TP fw
61	DNA	WPRE_TP 探针
62	DNA	WPRE_TP rev
63	RNA	具有 MS2 结合环的通用单 gRNA
64	DNA	协同激活介质 (SAM) 编码序列(dCas9-VP64-T2A-MCP-p65-HSF1)
65	DNA	通用指导 RNA 阵列启动子和指导 RNA 编码序列
66	DNA	<i>Ttr</i> 指导 RNA 阵列启动子和指导 RNA 编码序列
67	DNA	pscAAV Ttr 阵列
68	DNA	pAAV Ttr g1
69	DNA	pAAV Ttr g2
70	DNA	pAAV Ttr g3
71	DNA	pscAAV Ldlr 阵列
72	DNA	pAAV Ldlr g1
73	DNA	pAAV Ldlr g2

SEQ ID NO	类型	描述
74	DNA	pAAV Ldlr g3
75	DNA	小鼠 Ldlr 指导 RNA 靶序列 v1
76	DNA	小鼠 Ldlr 指导 RNA 靶序列 v2
77	DNA	小鼠 Ldlr 指导 RNA 靶序列 v3
78	RNA	小鼠 Ldlr 单指导 RNA v1
79	RNA	小鼠 Ldlr 单指导 RNA v2
80	RNA	小鼠 Ldlr 单指导 RNA v3
81	RNA	小鼠 Ldlr 指导 RNA DNA 靶向片段 v1
82	RNA	小鼠 Ldlr 指导 RNA DNA 靶向片段 v2
83	RNA	小鼠 Ldlr 指导 RNA DNA 靶向片段 v3
84	DNA	Ldlr 指导 RNA 阵列启动子和指导 RNA 编码序列
85	DNA	pcsAAV Pcsk9 阵列
86	DNA	pAAV Pcsk9 g1
87	DNA	pAAV Pcsk9 g2
88	DNA	pAAV Pcsk9 g3
89	DNA	小鼠 Pcsk9 指导 RNA 靶序列 v1
90	DNA	小鼠 Pcsk9 指导 RNA 靶序列 v2
91	DNA	小鼠 Pcsk9 指导 RNA 靶序列 v3
92	RNA	小鼠 Pcsk9 单指导 RNA v1
93	RNA	小鼠 Pcsk9 单指导 RNA v2
94	RNA	小鼠 Pcsk9 单指导 RNA v3
95	RNA	小鼠 Pcsk9 指导 RNA DNA 靶向片段 v1
96	RNA	小鼠 Pcsk9 指导 RNA DNA 靶向片段 v2
97	RNA	小鼠 Pcsk9 指导 RNA DNA 靶向片段 v3
98	DNA	Pcsk9 指导 RNA 阵列启动子和指导 RNA 编码序列

实施例

实施例1: SAM就绪的小鼠的产生

[0318] 要使用dCas9协同激活介质(SAM)系统,通常需要引入三个组分:(1)直接融合到VP64域的dCas9;(2)与两个其它激活转录因子(热休克因子1(HSF1)和转录因子65(p65))融合的MS2外壳蛋白(MCP);和(3)含MS2环的sgRNA。通常需要将每个组分引入单独的慢病毒中。尽管所描述的三组分系统在细胞培养中具有一定的灵活性,但在动物模型中这种设置不太理想。相反,我们选择引入dCas9 SAM组件(dCas9-VP64和MCP-p65-HSF1)作为由内源Rosa26启动子驱动的一个转录本。最初,dCas9SAM系统的表达被floxed的新霉素停止表达盒的存在而阻断。引入Cre重组酶后,删除终止盒,并打开dCas9 SAM表达。然后通过将指导RNA或指导RNA阵列(例如表达自U6启动子)整合入另一个Rosa26等位基因或通过AAV引入来引入指导RNA或指导RNA阵列。通过将dCas9 SAM等位基因与各种Cre递送方法配对,可以控制基因调节的时机和组织特异性。如下所示,该系统可用于在体内诱导基因表达,并可用于疾病模拟等应用。

[0319] 产生了包含靶向小鼠的Rosa26基因座的同源臂的大靶向载体(LTVEC),以将dCas9 SAM表达盒引入Rosa26基因座的第一内含子。在例如US 6,586,251和Valenzuela et al. (2003)Nat. Biotechnol. 21(6):652-659中描述了使用**VELOCIGENE**[®]基因工程技术,通过细菌同源重组(BHR)反应从细菌人工染色体(BAC)DNA衍生的LTVEC的产生和使用;出于所有目的,通过引用将其每一篇的全部内容合并于此。在例如US 2015/0376628和WO 2015/200334中描述了通过体外组装方法产生LTVEC的方法,出于所有目的将其全部内容通过引

用并入本文。

[0320] 对表达盒中的化脓链球菌dCas9编码序列(CDS)进行密码子优化以在小鼠中表达。编码的dCas9包含以下使Cas9核酸酶失活的突变:D10A和N863A。dCas9-NLS-VP64-T2A-MCP-NLS-p65-HSF1-WPRE表达盒如图1A和SEQ ID NO:31所示。协同激活介质(SAM)编码序列(dCas9-VP64-T2A-MCP-p65-HSF1)如SEQ ID NO:64所示,并编码SEQ ID NO:44所示的蛋白质。该表达盒靶向Rosa26基因座的第一个内含子(见图2),以利用Rosa26基因座的强普遍表达以及靶向Rosa26基因座的简单程度。表达盒之前是带有合适剪接信号和强多腺苷酸化(polyA)信号的floxed新霉素抗性盒(neo盒)。dCas9 SAM表达盒从5'到3'的组分如下表4所示。

[0321] 表4:dCas9 SAM表达盒组分。

组分	SEQ ID NO: 31 中的核苷酸区域
第一 loxP 位点	1 - 34
编码对新霉素家族抗生素(例如 G418)具有抗性的新霉素磷酸转移酶的序列	125 - 928
多腺苷酸化信号	937 - 2190
第二 loxP 位点	2218 - 2251
密码子优化的 dCas9 编码序列	2306 - 6457
NLS	2309 - 2356
NLS	6512 - 6532
VP64	6533 - 6719
带有 5' GSG 的 T2A	6719 - 6781
MCP	6782 - 7171
NLS	7226 - 7246
p65	7262 - 7804
HSF1	7829 - 8200
土拨鼠肝炎病毒转录后调控元件(WPRE)	8224 - 8820

[0322] 为了产生靶向的Rosa26等位基因,将LTVEC引入F1H4小鼠胚胎干细胞中。在抗生素选择后,通过TAQMAN[®]以挑选、扩增,和筛选集落(colonies)。在例如US 2014/0178879;US 2016/0145646;WO 2016/081923;和Frendewey et al. (2010) Methods Enzymol. 476:295-307中描述了包括等位基因丢失(LOA)和等位基因获得(GOA)测定的等位遗传修饰(MOA)测定,出于所有目的,通过引用将其每一篇的全部内容合并于此。等位基因丢失(LOA)测定颠倒了传统的筛选逻辑,并量化了突变所针对的天然基因座的基因组DNA样本中的拷贝数。在正确靶向的杂合细胞克隆中,LOA测定检测到两个天然等位基因之一(对于不在X或Y染色体上的基因),另一个等位基因被靶向修饰破坏。可以将相同的原理反向应用于等位基因获得(GOA)测定,以量化基因组DNA样品中插入的靶向载体的拷贝数。表5提供了用于筛选的引物和探针。

[0323] 表5:引物和探针。

引物/探针	序列	SEQ ID NO
-------	----	-----------

XBA-TDP fw	CGTGATCTGCAACTCCAGTCTT	45
XBA-TDP探针	AGATGGGCGGGAGTCTTCTGGGC	46
XBA-TDP rev	CACACCAGGTTAGCCTTTAAGCC	47
Neo fw	GGTGGAGAGGCTATTCGGC	48
Neo探针	TGGGCACAACAGACAATCGGCTG	49
Neo rev	GAACACGGCGGCATCAG	50
SAM TD fw	ACCGGCTGTCCGACTACGAT	51
SAM TD探针	TGGACCACATCGTGCCTCAGA	52
SAM TD rev	CGGGCCTTGTCGCTTCTG	53
MS2_T fw	GGCTCCTTCTAATTTTCGCTAATGG	54
MS2_T探针	TGGCAGAGTGGATCAGCTCCA	55
MS2_T rev	CTGACGCTGCATGTCACCTT	56
P65_T fw	AGGGCGTGTCCATGTCTCATAG	57
P65_T探针	ACAGCCGAACCAATGCTGATGGA	58
P65_T rev	CCAGCCGGGTAATGGCTTC	59
WPRE_TP fw	TGTGTTGCCACCTGGATTCTG	60
WPRE_TP探针	CGCGGGACGTCCTTCTGCTAC	61
WPRE_TP rev	GGAAGTCCGCTGGATTGAG	62

[0324] 使用**VELOCIMOUSE**[®]方法产生了F0小鼠。参见，例如，US 7,576,259；US 7,659,442；US 7,294,754；US 2008/0078000；和Poueymirou et al. (2007) Nat. Biotechnol. 25 (1) :91-99，出于所有目的，将其每一篇通过引用整体并入本文。在**VELOCIMOUSE**[®]方法中，通过激光辅助注射将靶向的小鼠胚胎干(ES)细胞注入前桑葚胚期(pre-morula stage)的胚胎，例如八细胞期的胚胎，从而有效地产生了完全为ES细胞衍生的F0代小鼠。

[0325] 在通过Cre重组酶的作用除去floxed的新霉素抗性盒(neo盒)之前，转录和翻译新霉素抗性基因；然而，由于存在有效阻断了贯穿转录的强poly(A)区域，因此没有表达dCas9-NLS-VP64 CDS和MCP-NLS-p65-HSF1CDS。参见图1A。但是，在通过Cre重组酶的作用去除neo盒后，dCas9和MCP融合蛋白的杂合mRNA由Rosa26启动子组成性表达。参见图1B。通过从已去除floxed新霉素抗性盒(neo盒)的靶向mESC中提取总RNA来验证dCas9和MCP的表达，然后进行反转录以生成cDNA和**TAQMAN**[®]qPCR，以检测反转录的cDNA(RT-qPCR)。测量Cas9和p65 mRNA水平。分别参见图3A和3B。dCas9表达也通过蛋白质印迹证实。参见图4。综上所述，所创建的系统能够在mESC和衍生自它们的小鼠中连续或有条件地(需要除去新霉素抗性盒)表达水平一致的dCas9融合蛋白和MCP融合蛋白。

实施例2. 用Ttr指导RNA验证SAM就绪小鼠

[0326] 为了在体内验证该系统，用Ttr指导RNA阵列靶向载体靶向杂合的dCas9 SAM mESC，所述Ttr指导RNA阵列靶向载体靶向小鼠Rosa26基因座的第一内含子。Ttr指导RNA阵列在图5和SEQ ID NO:33中描述。包括启动子和指导RNA编码序列的区域在SEQ ID NO:66中列出。阵列中的指导RNA所靶向的小鼠Ttr基因中的指导RNA靶序列(不包括PAM)分别在SEQ

ID NO:34 (ACGGTTGCCCTCTTTTCCCAA)、SEQ ID NO:35 (ACTGTCAGACTCAAAGGTGC), 和SEQ ID NO:36

(GACAATAAGTAGTCTTACTC)中列出。SEQ ID NO:34位于Ttr转录起始位点的-63, SEQ ID NO:35位于Ttr转录起始位点的-134, 而SEQ ID NO:36位于Ttr转录起始位点的-112。靶向这些指导RNA靶序列的单指导RNA分别在SEQ ID NOS:37、38和39中列出。同源臂侧接Ttr指导RNA阵列, 该阵列包含三个靶向Ttr基因座的、含MS2茎环的指导RNA。将Ttr指导RNA阵列与roxed的嘌呤霉素终止盒整合在Rosa26基因座上。此盒可防止Rosa26贯穿转录本, 干扰U6启动子活性。终止盒后, U6启动子串联表达包含MS2茎环的三个sgRNA序列, 并用扩展的PolIII终止序列将它们分开。设计指导(guides)的目的是将dCas9 SAM组分引导至Ttr转录起始位点(TSS)上游100-200bp的区域。参见图6。Ttr指导RNA阵列表达盒从5'到3'的组分如下表6所示。每个指导RNA的结构总体示意图, 包括MS2茎环, 示于图7。

[0327] 表6. Ttr指导RNA阵列表达盒组分。

组分	SEQ ID NO: 33 中的核苷酸区域
第一 rox 位点	1 – 32
编码嘌呤霉素-N-乙酰基转移酶以对嘌呤霉素家族抗生素具有抗性的序列	111 – 710
多腺苷酸化信号	797 – 2338
第二 rox 位点	2363 – 2394
第一 U6 启动子	2401 – 2640
第一 Ttr 指导 RNA 编码序列	2642 – 2798
第二 U6 启动子	2884 – 3123
第二 Ttr 指导 RNA 编码序列	3125 – 3281
第三 U6 启动子	3366 – 3605
第三 Ttr 指导 RNA 编码序列	3606 – 3762

[0328] 在确认靶向的mESC克隆对于dCas9 SAM表达盒为杂合且对于指导RNA阵列表达盒为杂合后, 我们使用RT-qPCR来确定相对基因表达。对于Ttr, 在我们的WT mESC中、包含由终止盒封闭的dCas9 SAM组分的mESC中, 以及带有主动表达的dCas9 SAM等位基因(移除终止盒)的mESC中, RT-qPCR的ct值达到35。但是, 在将U6 SAM Ttr指导阵列靶向到每条细胞系后, 只有包含活性dCas9 SAM系统加指导表达的细胞系才显示出ct值减小为20。请参见图8A。ct值下降15意味着相对基因表达增加了2500倍。随着Ttr表达的显著增加, 我们希望确保dCas9 SAM激活成分的紧密接近不会影响邻近的基因。为此, 通过RT-qPCR评估Dsg2和B4gal1t6 (Ttr每一侧的基因), 并确定在上述任何细胞系中的表达均没有显著增加。分别参见图8B和8C。

[0329] 为了验证该基因上调是稳定的并且可以转化为小鼠模型, 将靶向克隆微注射到8细胞小鼠胚胎中以衍生出小鼠品系。具体而言, 在透明带上创建一个小孔, 以促进靶向mESC的注射。将这些注射的8细胞胚胎转移给代孕母体, 以产生携带转基因的幼仔。在代孕母体中妊娠后, 注射的胚胎产生了F0小鼠, 其没有携带可检测的宿主胚胎。完全由ES细胞衍生的

小鼠是正常、健康和可育的(通过种系传播)。

[0330] 在从具有基因组整合的Ttr指导RNA阵列的野生型小鼠、dCas9SAM小鼠,和dCas9SAM小鼠收获的各种组织中观察到通过RT-qPCR测定的Ttr mRNA表达。这些组织中的每一个都提取了RNA。基因组DNA被降解,因此不会计入qPCR反应。逆转录RNA,然后使用对Ttr特异的测定法检测Ttr转录本。在实验中,通过RT-qPCR测定了来自每个组织的等量RNA。数据显示,Ttr表达水平在所有组织中均升高,包括一些正常情况下预计不会出现TTR的器官。参见表7。在所有检查的组织(包括肝、脾、心脏、肺、骨骼肌、睾丸、胸腺、眼、胰腺、淋巴结、肾,和脑)中,TTR蛋白的表达也升高。分别参见图9A-9L。然而,表达的相对性受组织的影响。总体而言,来自对照小鼠的组织中较低的Ttr表达与SAM系统较高的上调相关。由肺和脾脏中的循环阈值确定的Ttr mRNA表达水平分别示于图10A和10B中。在RT-qPCR中,通过荧光信号的积累检测到阳性反应。循环阈值(ct)定义为荧光信号超过背景水平所需的循环数——ct值越低表示表达越高。此外,与在mESC克隆中进行筛选一样,通过RT-qPCR评估Dsg2和B4gal1t6 (Ttr两侧的基因),并确定其表达没有明显增加。参见例如图10C、10D、10E,和10F。

[0331] 表7. 相对于对照,TTR表达的增加。

组织	WT TTR (平均 Ct)	R26SAM TTR (平均 CT)	R26TTR:R26SAM TTR (平均 CT)	相对表达
肝	14.11	14.45	13.01	2.9
脑	17.89	18.04	16.29	3.19
眼	18.84	19.38	16.78	6.89
肾	19.49	20.25	13.05	110.13
胰腺	22.72	23.04	14.29	238.68
胸腺	27.28	27.93	20.29	247.99
睾丸	27.47	27.60	17.64	1935.15
心脏	28.36	29.56	17.71	1015.33
脾	29.70	28.63	18.08	3323.59
肺	30.43	31.90	15.17	79415.55
骨骼肌	32.32	29.89	19.86	5918.21

[0332] 与野生型小鼠或只表达dCas9 SAM组分的小鼠相比,对dCas9SAM组分和靶向Ttr的SAM指导RNA阵列杂合的F0小鼠显示从血清中通过ELISA检测的1000 μ g/mL的循环TTR增加至4000 μ g/mL。参见图11。

[0333] 我们接下来评估了在表达靶向来自Rosa26基因座的Ttr的指导RNA的小鼠中TTR水平的增加是否稳定。使用三组小鼠:(1) F1H4 (WT); (2) 杂合的Rosa26-dCas9-SAM; 和(3) Rosa26-dCas9-SAM:Rosa26-U6-TTR指导阵列(3个靶向Ttr的指导)。如上所述,从mESC产生这些小鼠,并且将F0代老化至一年。每月通过ELISA测量TTR的血清量,并观察动物的任何病理变化。尽管这些动物在一年内未观察到病理变化,但它们的循环TTR保持了2倍至2.5倍的增长。参见图13。这些数据表明,在表达靶向来自Rosa26基因座的Ttr的指导的小鼠中,Ttr表达和循环TTR水平稳定至少一年。

[0334] 其中的dCas9 SAM组分和指导RNA阵列的表达构建体均基因组整合的系统是在小鼠的整个寿命中产生非常高的基因表达的良好系统。然而,我们还希望能够模仿表达的急剧增加。为此,我们将具有相同Ttr指导阵列的AAV引入了成年小鼠中,其对Rosa26表达的dCas9 SAM组分杂合。通过尾静脉注射将AAV(血清型8)引入靶肝细胞。我们在注射后5、19和60天通过ELISA分析了循环TTR,以确定注射是否成功。令人惊讶地,到第5天,在小鼠中循环的TTR水平从未经处理的小鼠或用表达GFP的AAV处理的对照小鼠中的1000 $\mu\text{g}/\text{mL}$ 跃升至在AAV处理的小鼠中的7000 $\mu\text{g}/\text{mL}$ 。到第19天,血清水平继续增加到11,000 $\mu\text{g}/\text{mL}$ 。到第60天,血清水平仍约为8000 $\mu\text{g}/\text{mL}$ 。参见图12。

[0335] 我们接下来在注射后8个月进行该实验。如上所述,评估了三组小鼠:(1)纯合的Rosa26-dCas9-SAM(未处理);(2)纯合的Rosa26-dCas9-SAM(AAV8-GFP);(3)纯合的Rosa26-dCas9-SAM(AAV8-gTTR阵列(3个靶向TTR的指导))。这些小鼠在8周龄时被注射AAV8-GFP或AAV8-gTTR阵列,并被追踪到注射后8个月。通过ELISA在各个早期时间点测定TTR的血清量,然后每月一次测定TTR的血清量,并观察这些动物的任何病理变化。尽管这些动物在注射后8个月未观察到病理变化,但在第19天时它们的循环TTR具有11倍的最初增加,而在注射后5个月,该水平处于TTR升高了约4倍的稳定状态。参见图14。

[0336] 为了追踪在体内的上调是否需要多个指导RNA,或者单指导RNA是否足够,我们从指导RNA阵列中取出每个指导RNA,并将其分别包装到AAV8中。在该实验中评估了六组小鼠:(1)纯合的Rosa26-dCas9-SAM(未处理);(2)纯合的Rosa26-dCas9-SAM(AAV8-GFP);(3)纯合的Rosa26-dCas9-SAM(AAV8-gTTR阵列(3个靶向TTR的指导));(4)纯合的Rosa26-dCas9-SAM(AAV8-gTTR#1);(5)纯合的Rosa26-dCas9-SAM(AAV8-gTTR#2);和(6)纯合的Rosa26-dCas9-SAM(AAV8-gTTR#3)。组(3)-(6)中的指导RNA表达盒的序列分别在SEQ ID NO:67-70中列出。这些小鼠在8周龄时被注射了含有AAV8的指导RNA或GFP,并被追踪到注射后8个月。通过ELISA在不同的早期时间点至3周测量TTR的血清量。结果如图15所示。注射后1周,gTTR指导阵列的循环TTR较对照组增加了6.5倍,而每个单指导RNA的血清中的循环TTR则增加了3倍。在注射后2周,gTTR指导阵列的循环TTR降低至对照组的5.5倍,而单指导中的两个的血清中的循环TTR保持为3.5倍,且gRNA#3的血清中的循环TTR跃升至几乎5倍。在注射后3周,与WT对照相比,所有gRNA的血清中的循环TTR具有增加约3.5倍的水平。这些结果表明,指导RNA阵列可以提供最初的高蛋白爆发,但随着时间的流逝,单个gRNA可以在基因上调方面同样表现出色,从而导致循环TTR蛋白。

[0337] 为了继续评估整合到AAV中的单指导或多指导在Rosa26-dCas9-SAM小鼠中的基因上调中是否更成功,我们使用一种或两种指导RNA评估了肝中靶基因1的表达。注射后三周通过TAQMAN评估RNA表达。我们观察到与未经处理的纯合Rosa26-dCas9-SAM相比,所有三个组((1)纯合的Rosa26-dCas9-SAM(AAV8-靶基因1指导RNA#1)、(2)纯合Rosa26-dCas9-SAM(AAV8-靶基因1指导RNA#2)和(3)纯合Rosa26-dCas9-SAM(AAV8-靶基因1指导RNA#1&2))中的靶基因1的显著上调。与注射后3周的一个指导RNA组相比,两个指导RNA组的RNA表达显著增加。将AAV8与两个指导RNA结合使用可导致肝脏表达比未处理的多200倍,而将AAV8与一个指导RNA结合使用可以导致肝脏表达比未经处理的多100倍。参见图19。

[0338] 尽管上述实验主要集中于小鼠Ttr基因的上调,但是当靶向其它基因时也观察到类似的表达增加(数据未显示)。此外,通过使用不同的血清型或通过组织特异性Cre处理来

控制dCas9 SAM表达,我们可以控制基因上调的时机和组织特异性以生成强大、可靠的疾病模型。

实施例3.用Pcsk9和Ldlr指导RNA验证SAM就绪小鼠

[0339] 作为该系统在体内的进一步验证,选择参与胆固醇途径的两个基因(Pcsk9和Ldlr)作为上调的靶标,并且在五周的时间过程中观察对胆固醇水平的生理作用。评估三组小鼠:(1)纯合Rosa26-dCas9-SAM(AAV8-Pcsk9指导阵列);(2)纯合Rosa26-dCas9-SAM(AAV8-Ldlr指导阵列);(3)纯合Rosa26-dCas9-SAM(未处理)。

[0340] Pcsk9指导RNA阵列的序列在SEQ ID NO:85中列出。该指导RNA阵列编码三个指导RNA。包括启动子和指导RNA编码序列的区域在SEQ ID NO:98中列出。由阵列中的指导RNA靶向的小鼠Pcsk9基因中的指导RNA靶序列(不包括PAM)在SEQ ID NO:89-91中列出。靶向这些指导RNA靶序列的单指导RNA分别在SEQ ID NO:92-94中列出。

[0341] Ldlr指导RNA阵列的序列在SEQ ID NO:71中列出。指导RNA阵列编码三个指导RNA。包括启动子和指导RNA编码序列的区域在SEQ ID NO:84中列出。由阵列中的指导RNA靶向的小鼠Ldlr基因中的指导RNA靶序列(不包括PAM)在SEQ ID NO:75-77中列出。靶向这些指导RNA靶序列的单指导RNA分别在SEQ ID NO:78-80中列出。

[0342] 结果在图16A中示出。注射后两周,Rosa26-dCas9-SAM(AAV8-Pcsk9指导阵列)的胆固醇水平比注射前的胆固醇水平增加3.5倍。与此相对,Rosa26-dCas9-SAM(AAV8-Ldlr指导阵列)显示总胆固醇水平比注射前水平降低了75%。未经处理的动物保持其胆固醇。注射后5周,Rosa26-dCas9-SAM(AAV8-Pcsk9指导阵列)的胆固醇水平比注射前的胆固醇水平增加3倍。与此相对,Rosa26-dCas9-SAM(AAV8-Ldlr指导阵列)显示总胆固醇水平比注射前水平降低了50%。未经处理的动物保持其胆固醇。在LDL水平处观察到类似的效果。参见图16B。

[0343] 接下来,评估了Ldlr和Pcsk9的表达。在图17A和17B中分别示出了在上述实验中,在注射后5周,小鼠肝脏中Ldlr和Pcsk9的TAQMAN表达水平。

[0344] 作为进一步的验证,示出通过dCAS9-SAM系统而增加的LDLR可以带来长期利益,我们在喂养高脂饮食的纯合dCas9-SAM小鼠中评估了AAV8-Ldlr指导RNA阵列,并在注射AAV8-Ldlr指导RNA阵列后对其进行了20周的追踪。预先给小鼠放血以测定其初始胆固醇水平,然后向其饲喂高脂肪饮食(HFD)8周(每4周放血以测试胆固醇水平)。胆固醇和LDL水平的结果分别示于图18A和18B。8周后,向小鼠注射AAV8-Ldlr指导阵列或不进行处理。每月给小鼠放血,并评估其总胆固醇和LDL水平。在此时间框架内,与未经处理的、饲喂HFD的小鼠相比,经AAV8-Ldlr指导阵列处理的小鼠具有较低的总胆固醇和较低的LDL水平。

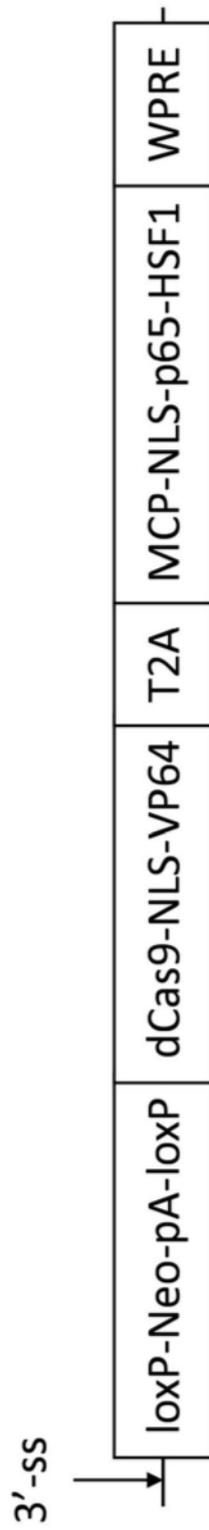


图1A



图1B

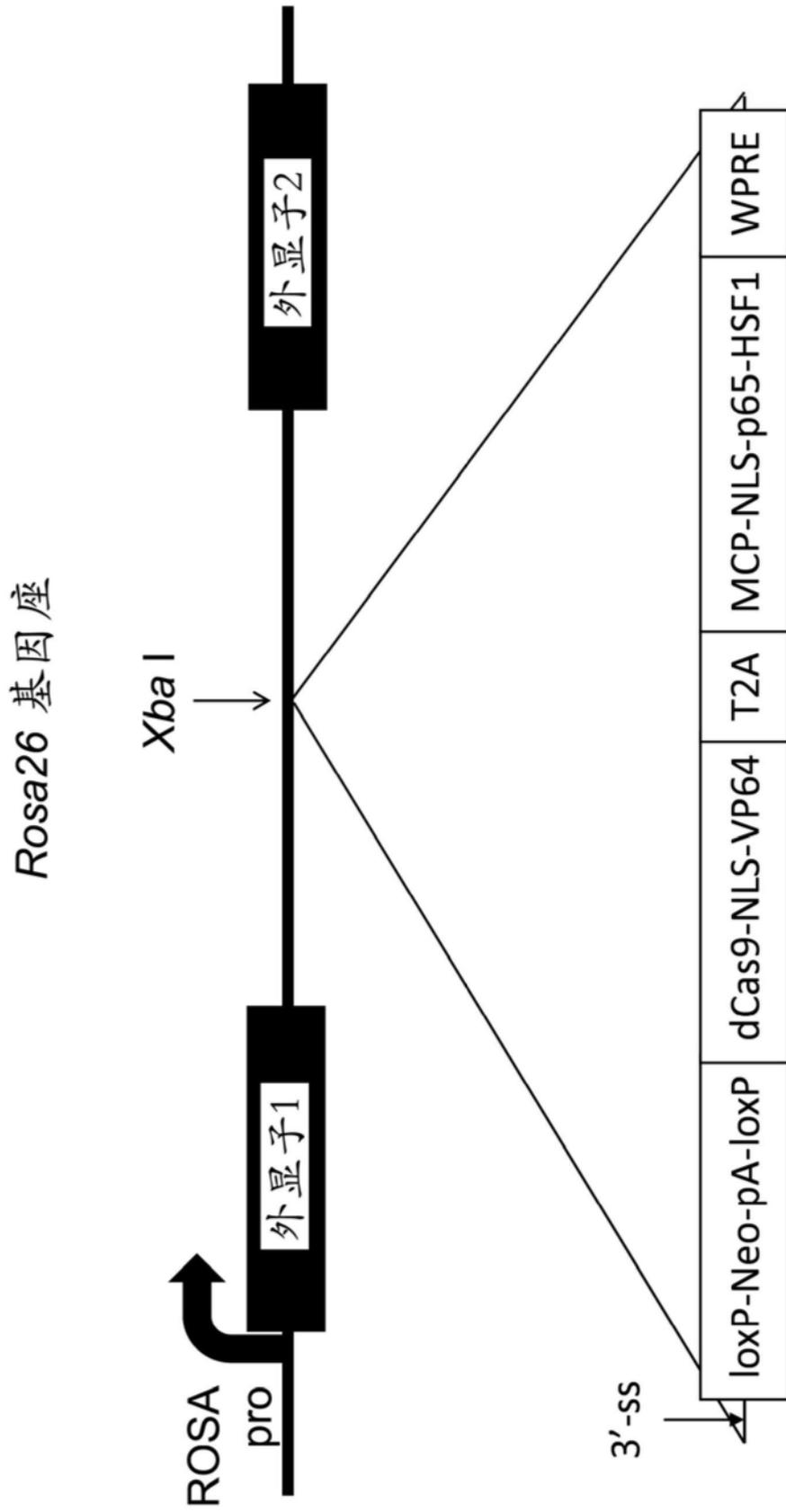


图2

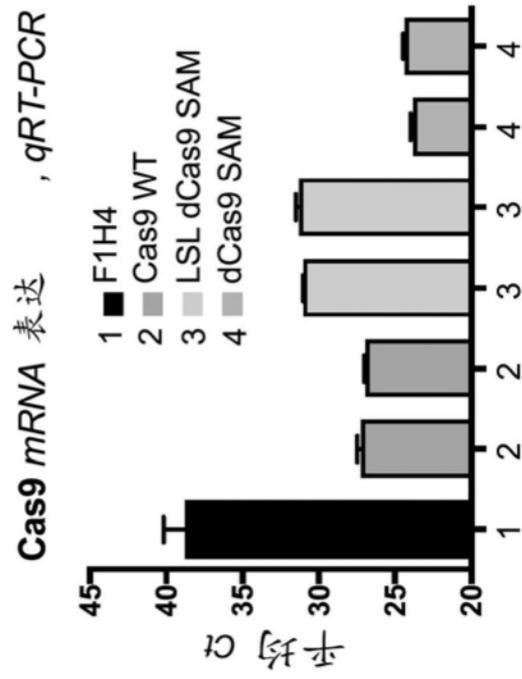


图3A

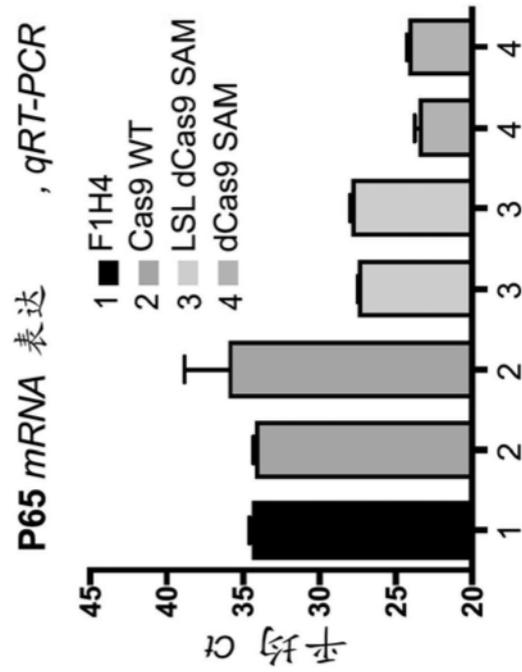


图3B

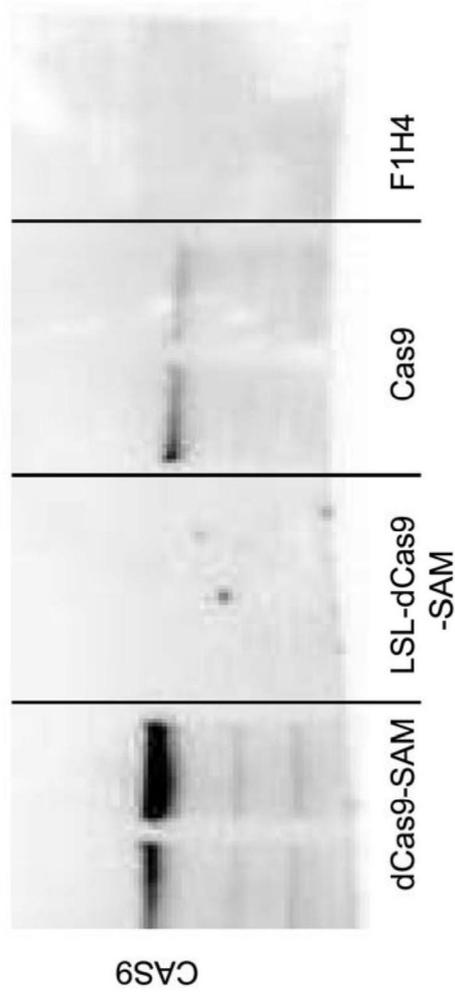


图4

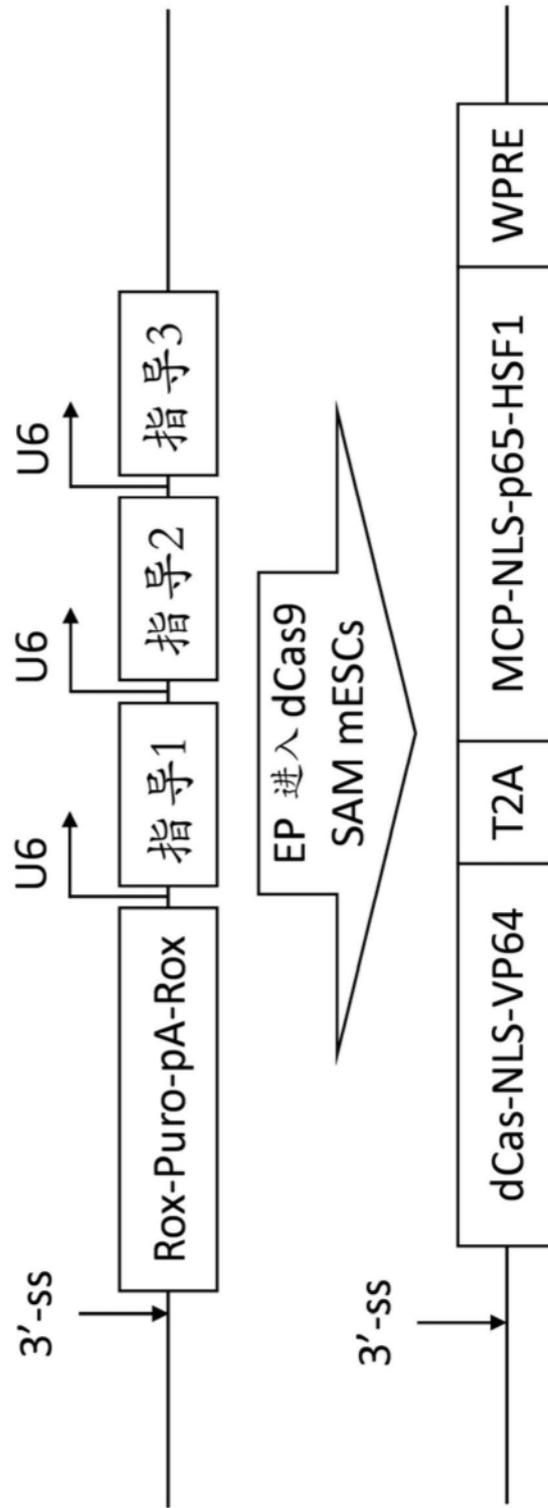


图5

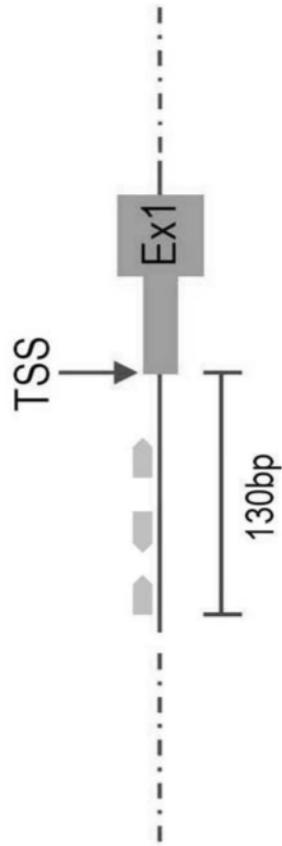


图6

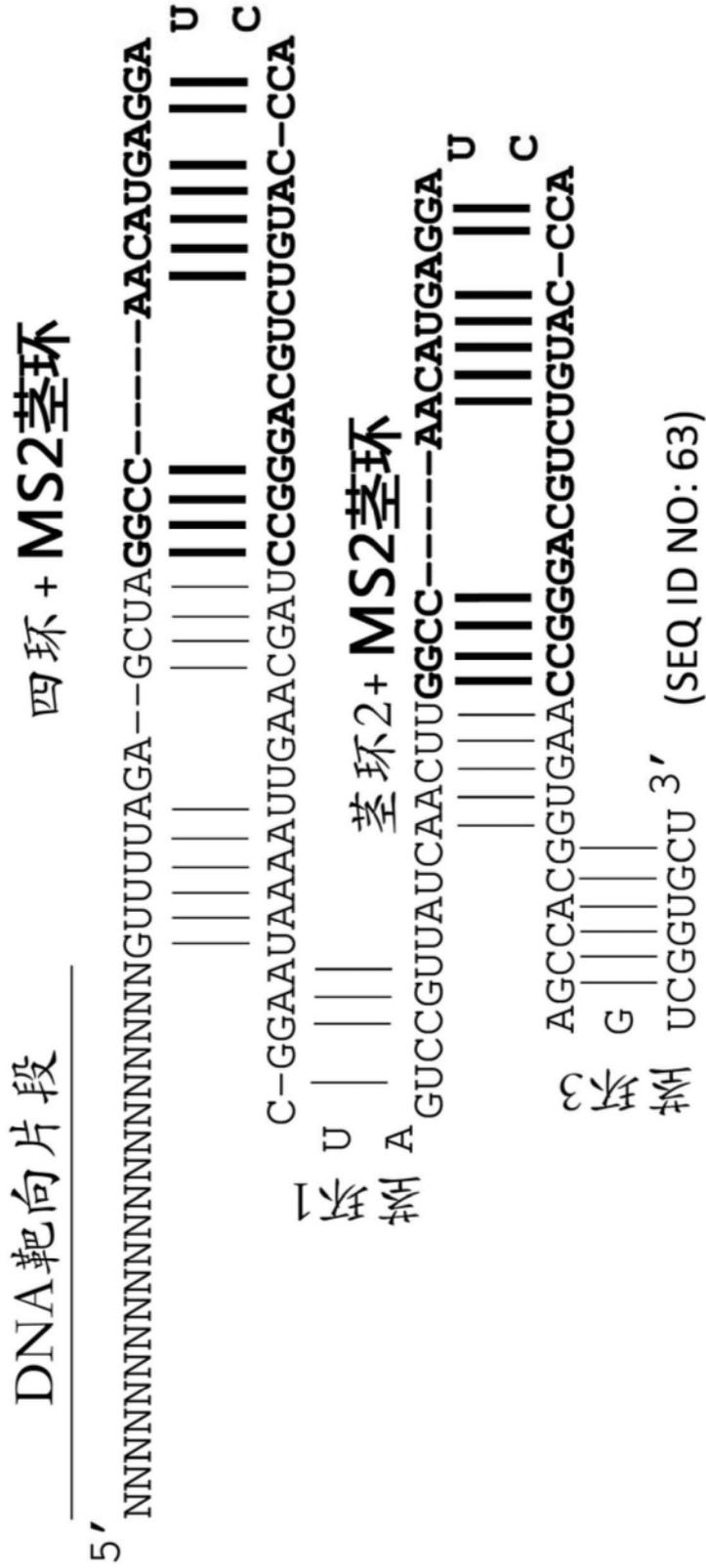


图7

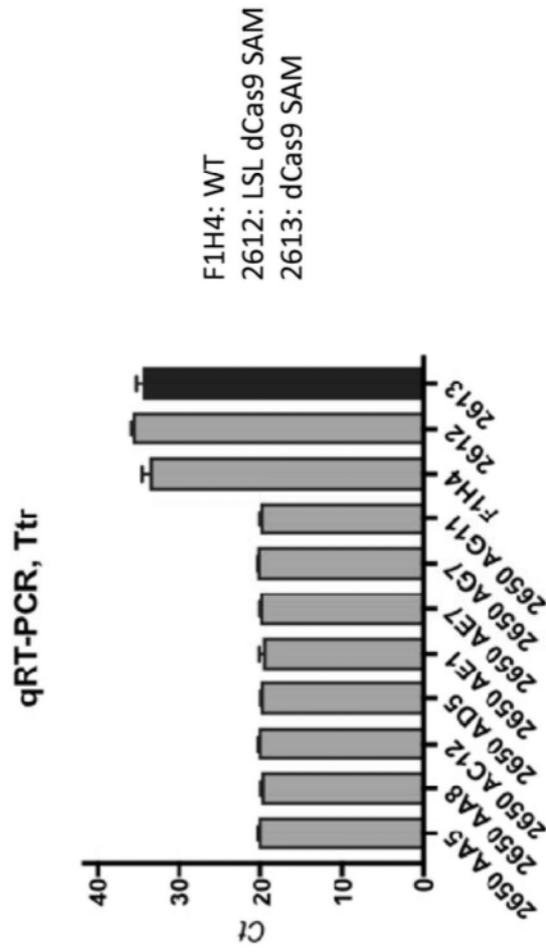


图8A

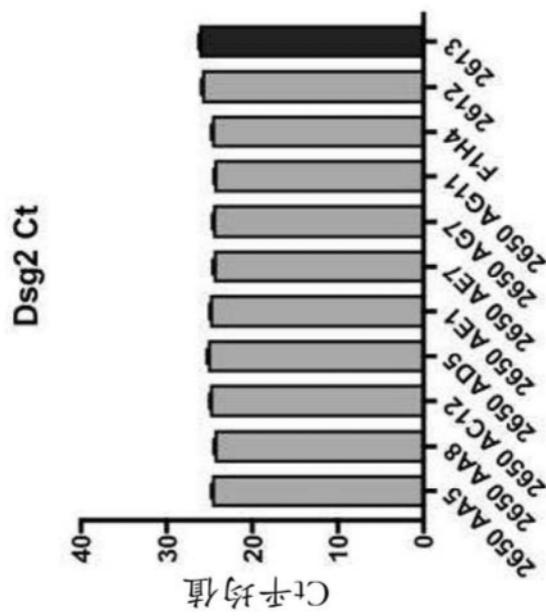


图8B

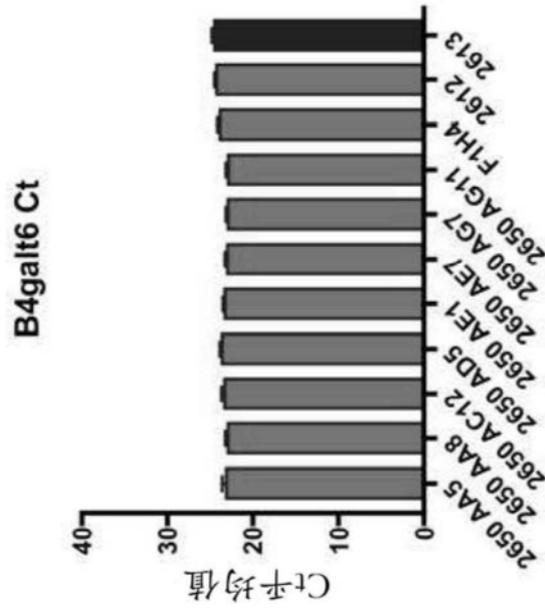


图8C

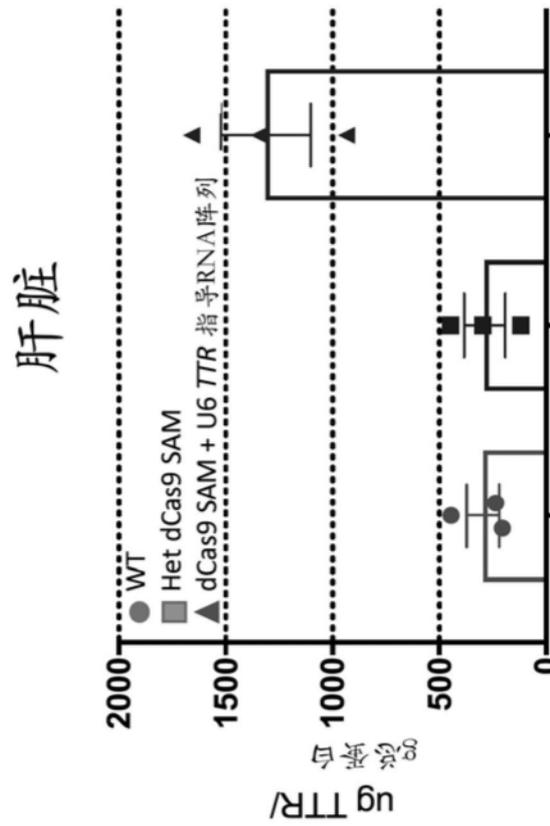


图9A

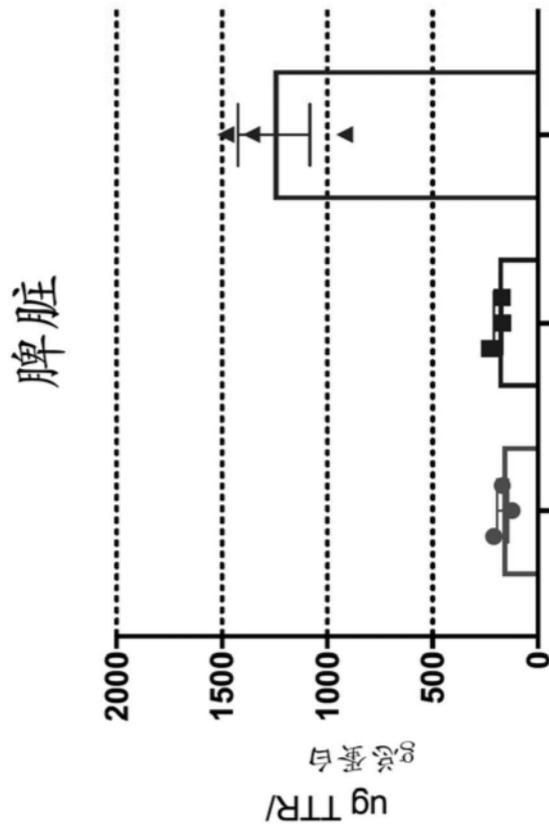


图9B

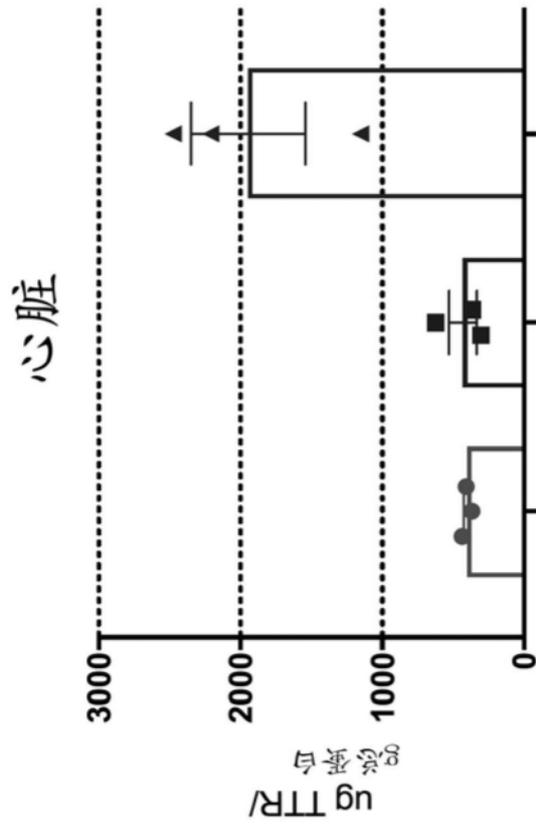


图9C

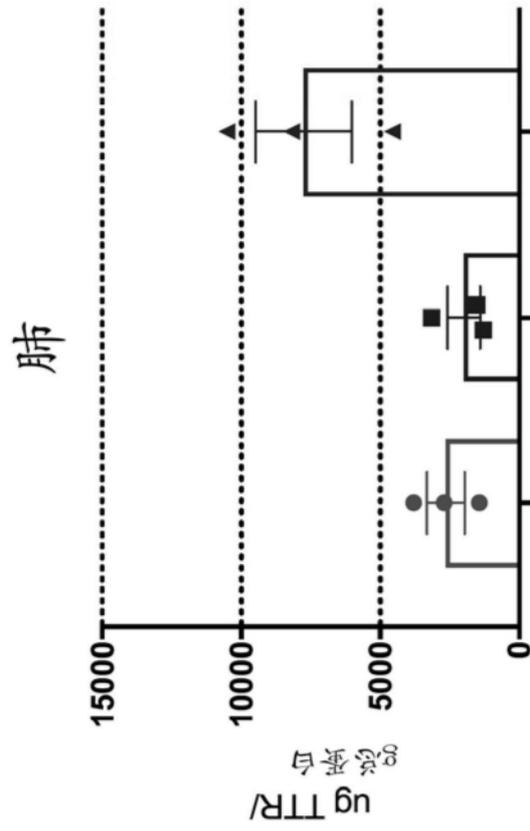


图9D

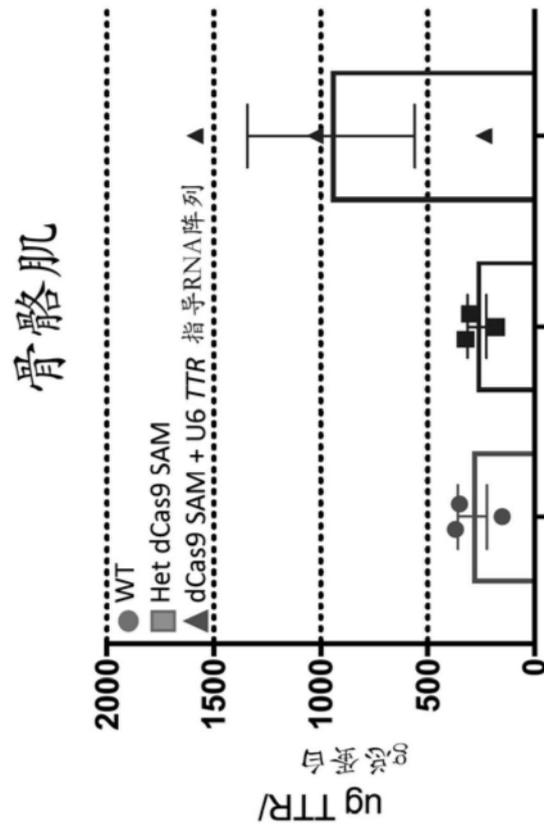


图9E

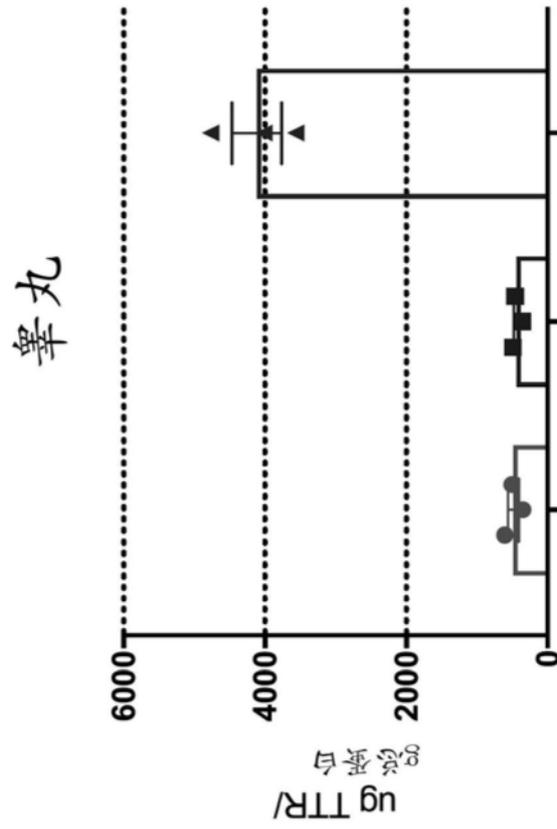


图9F

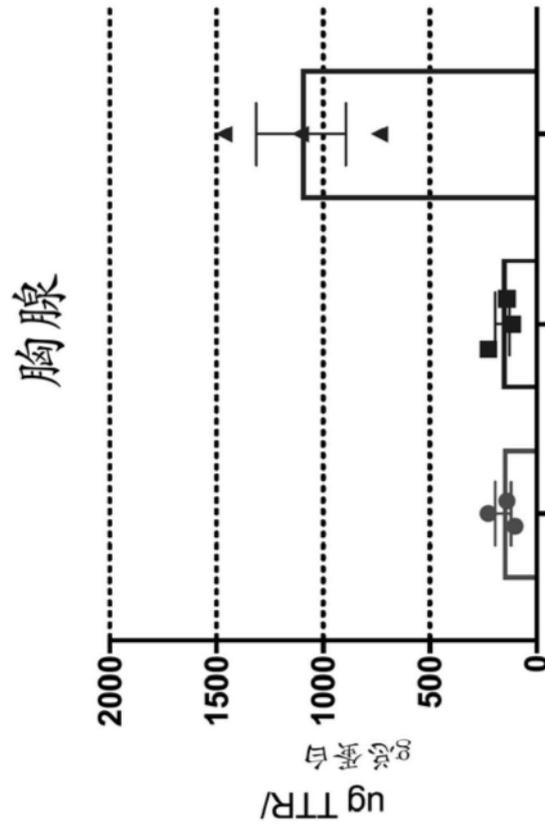


图9G

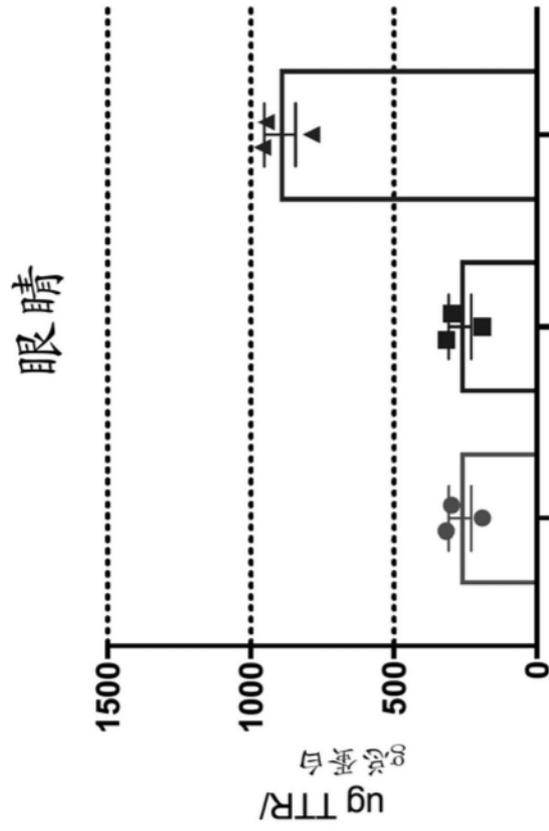


图9H

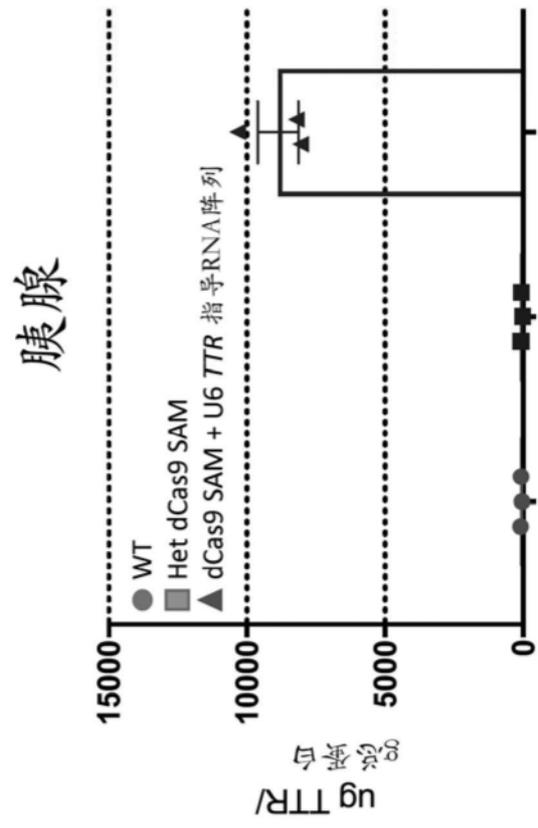


图9I

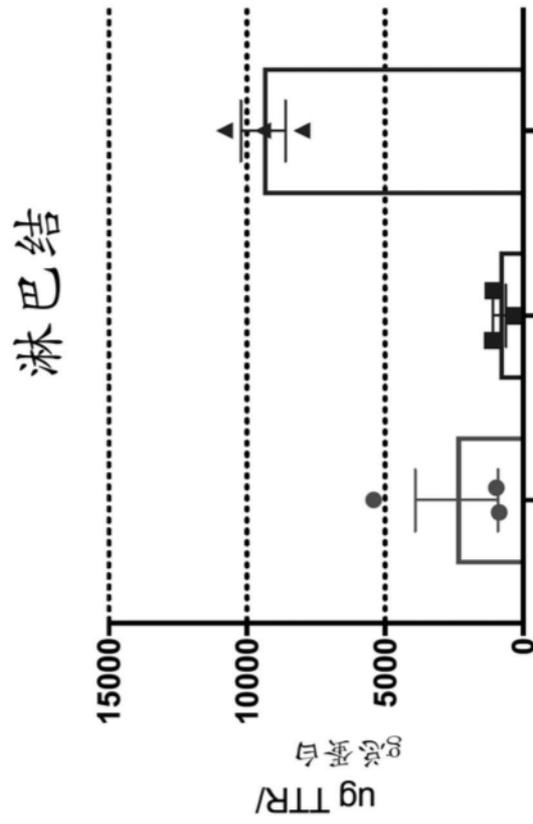


图9J

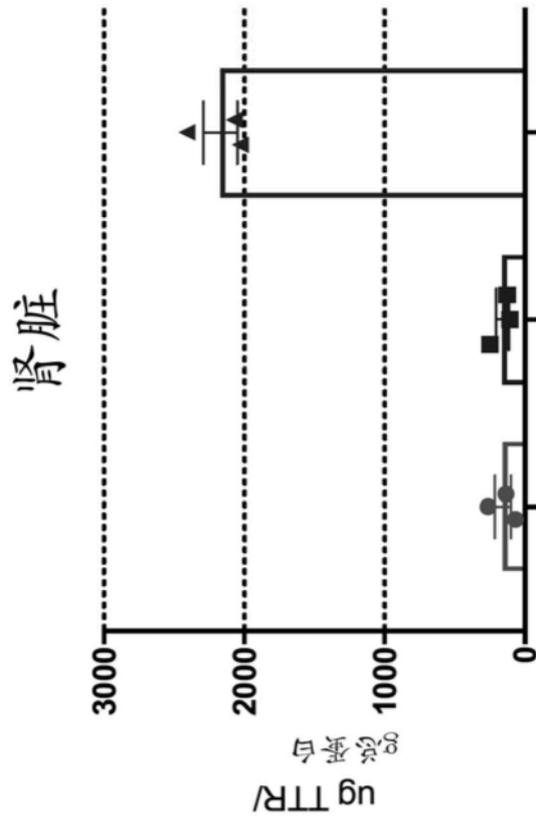


图9K

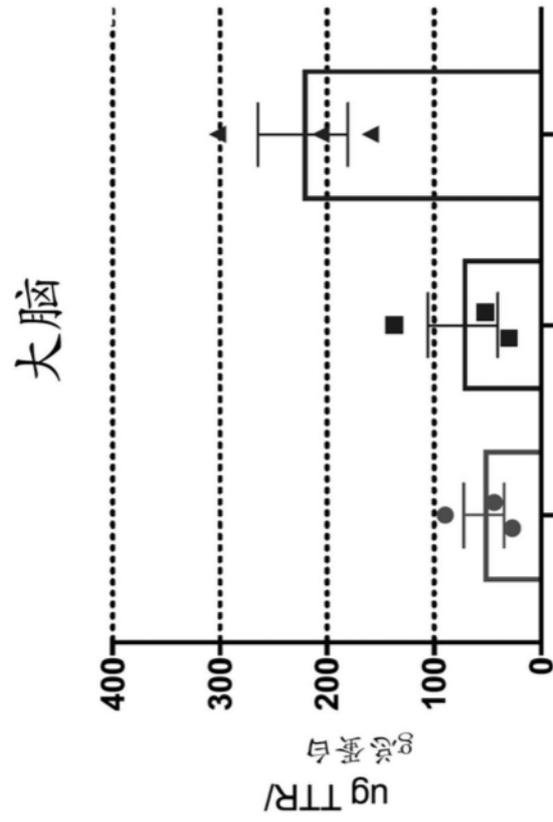


图9L

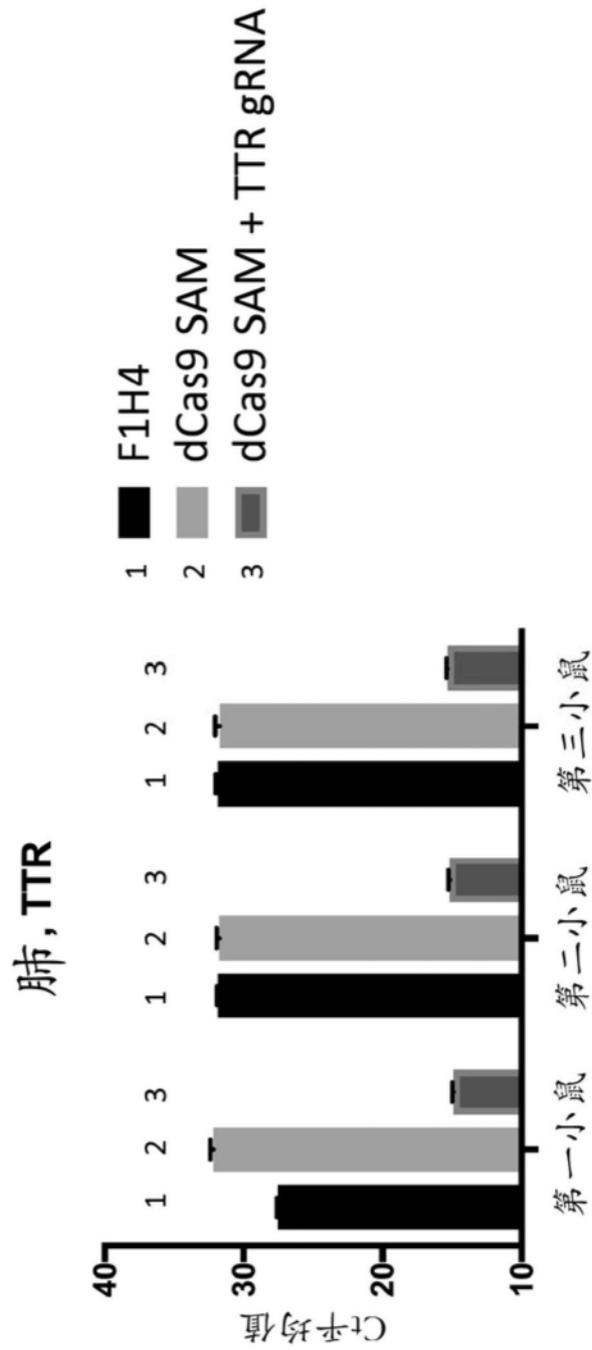


图10A

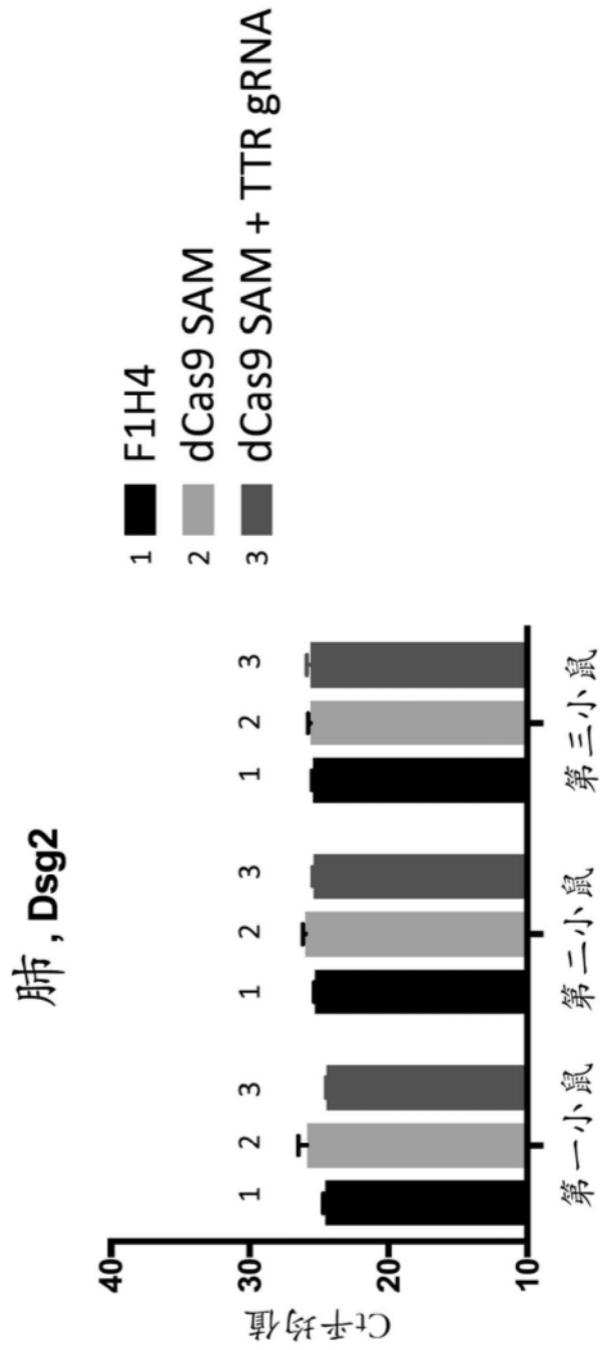


图10C

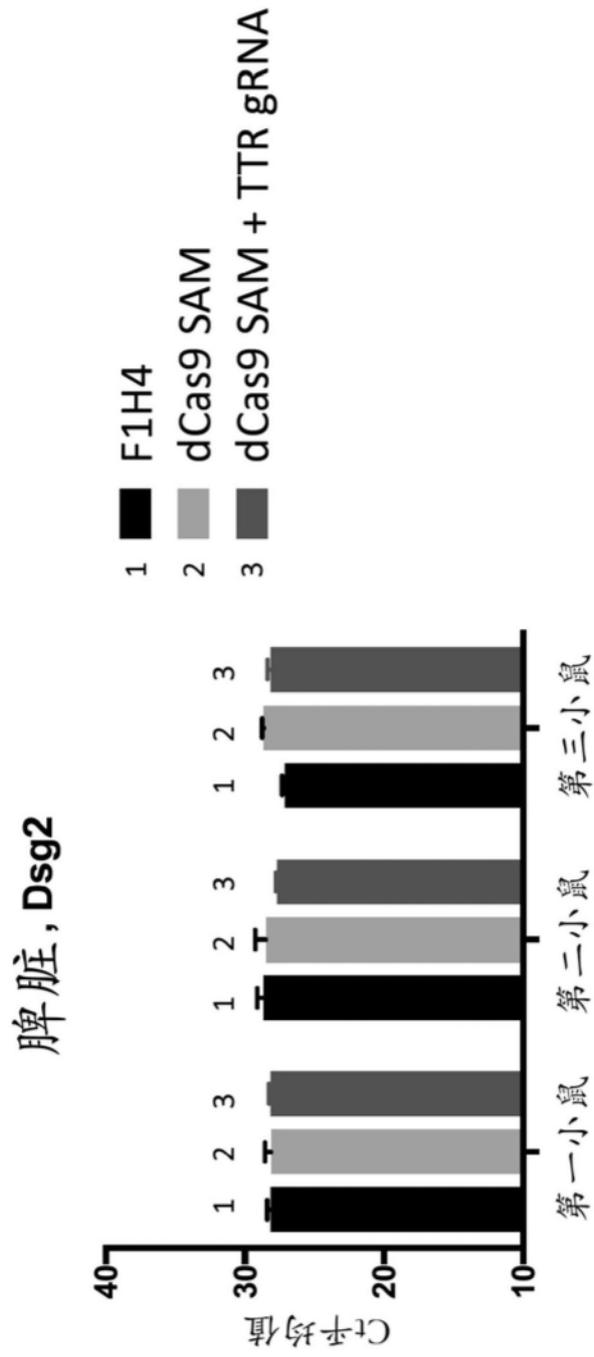


图10D

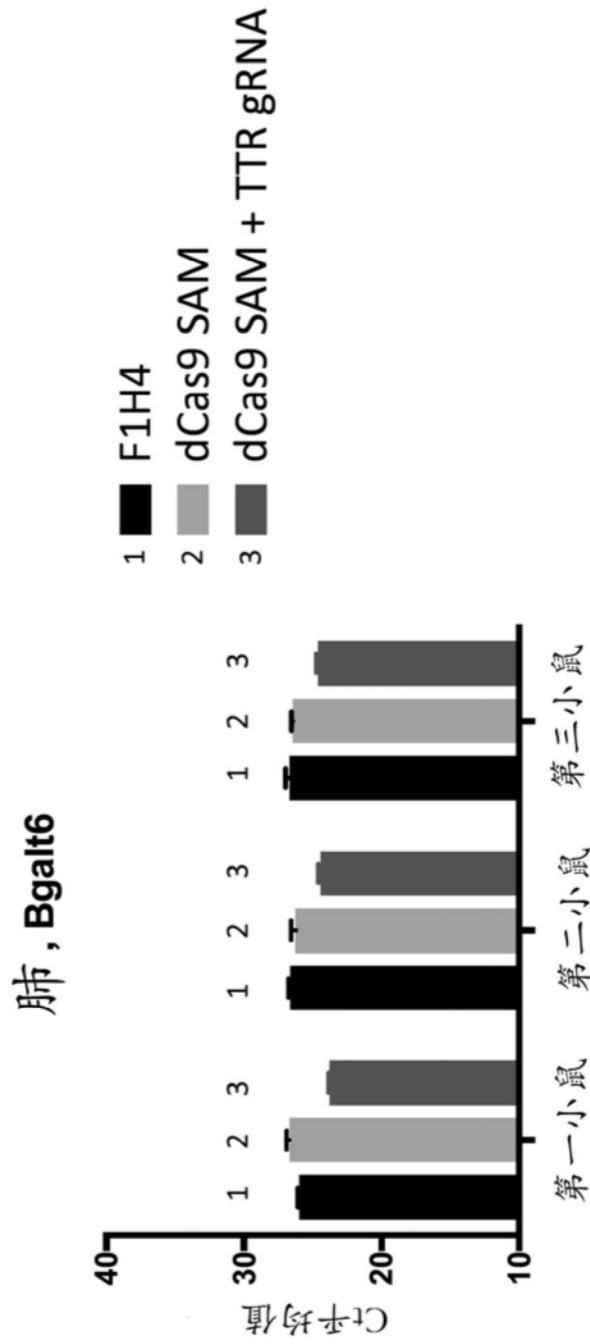


图10E

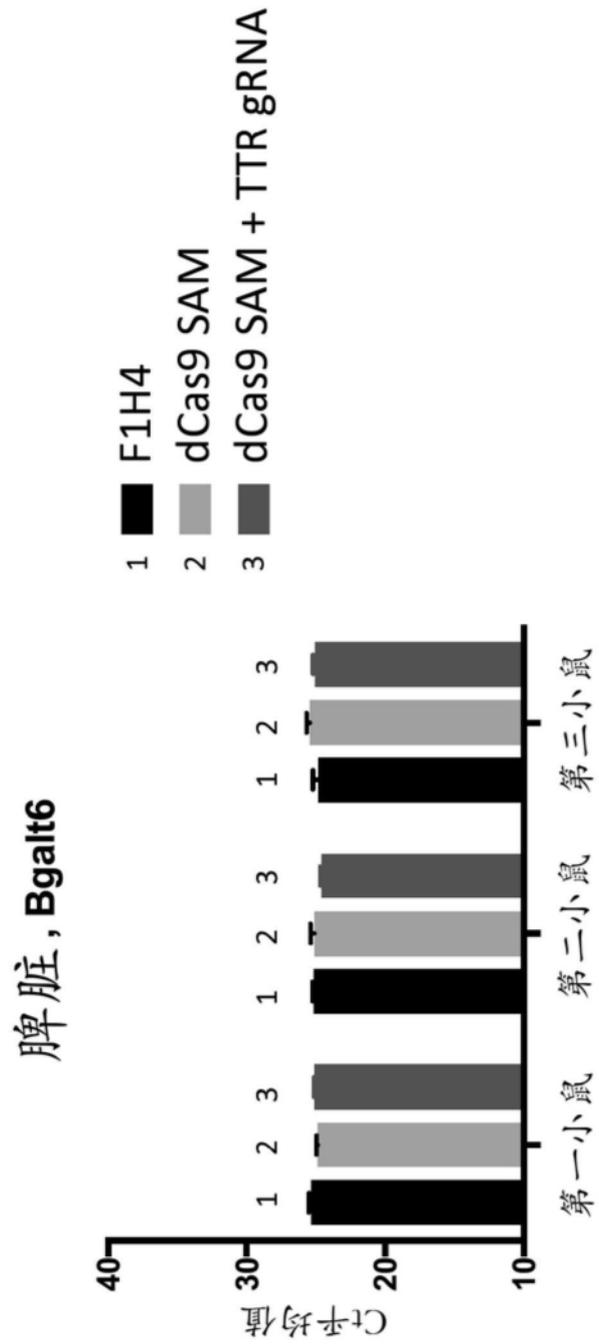


图10F

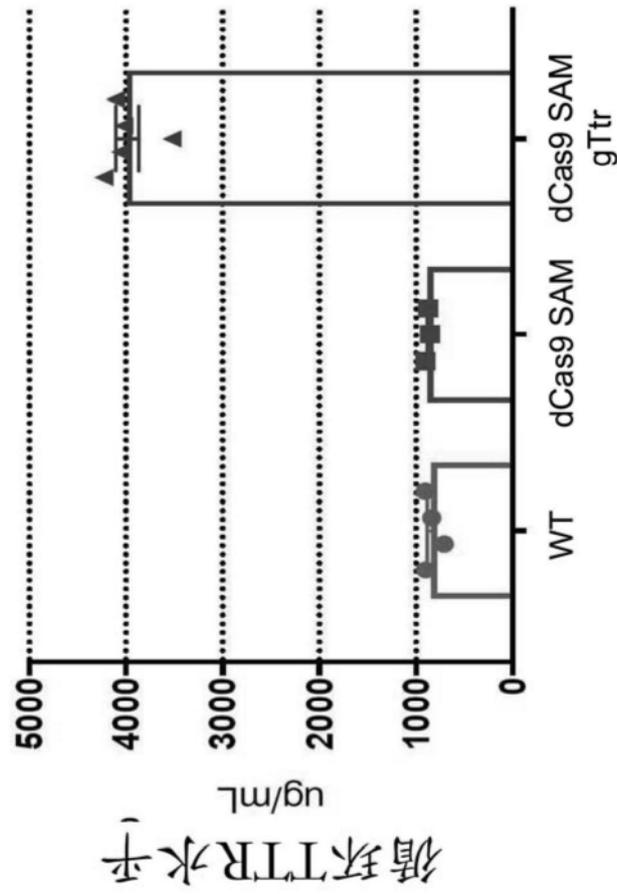


图11

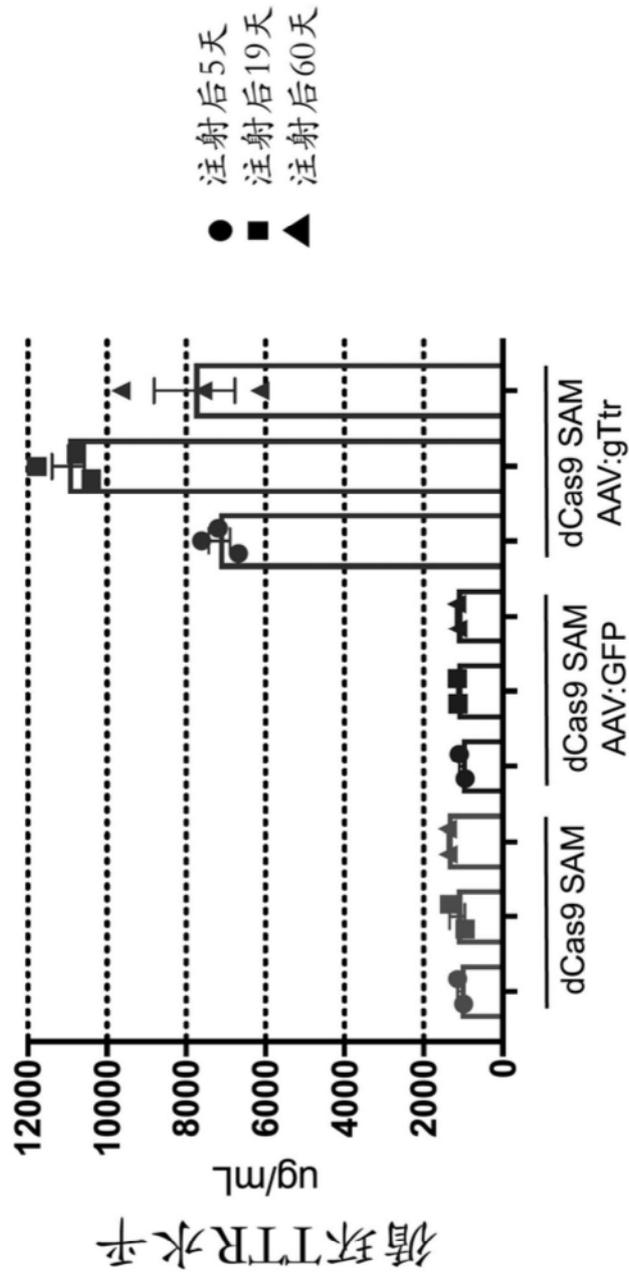


图12

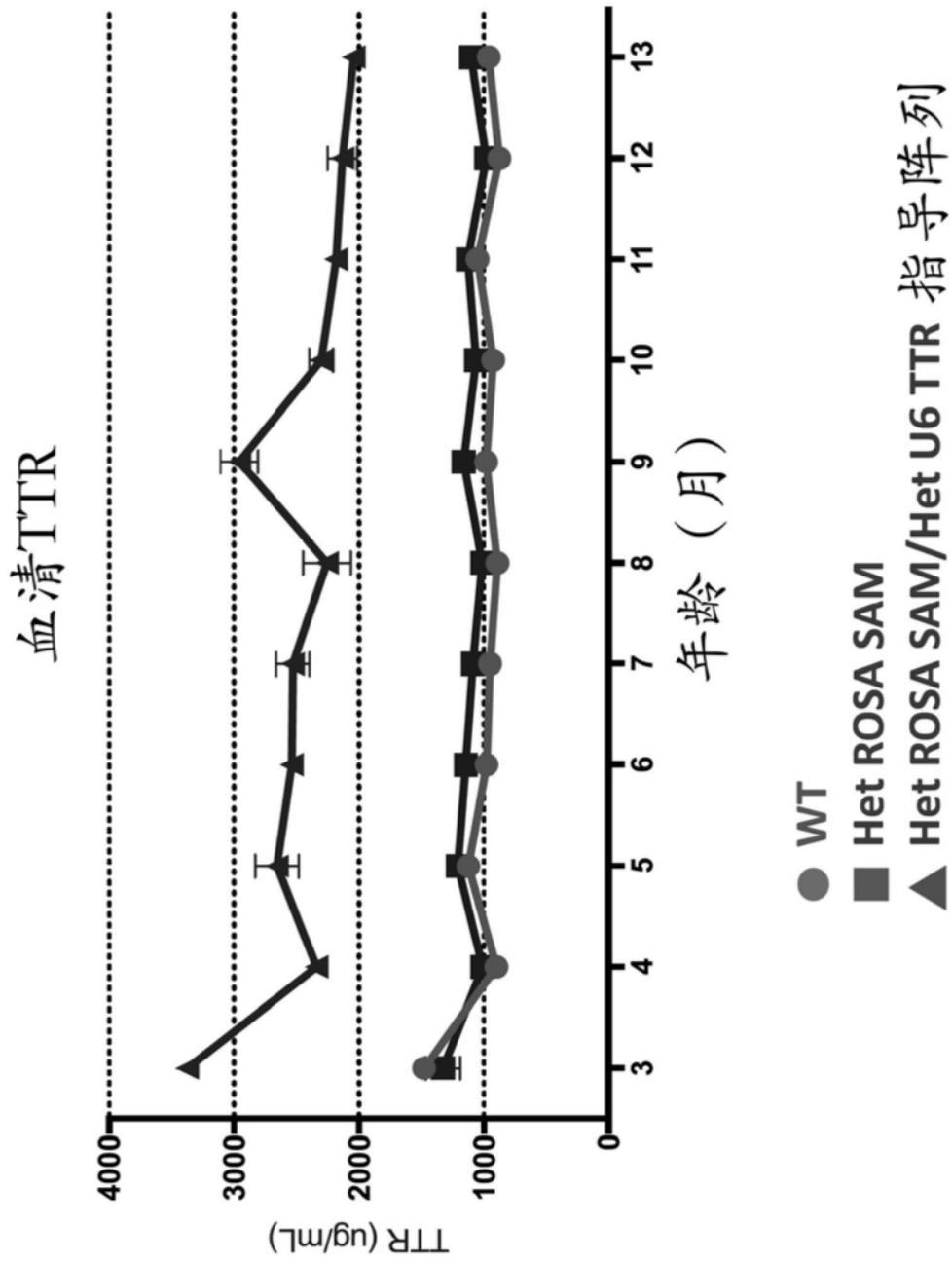


图13

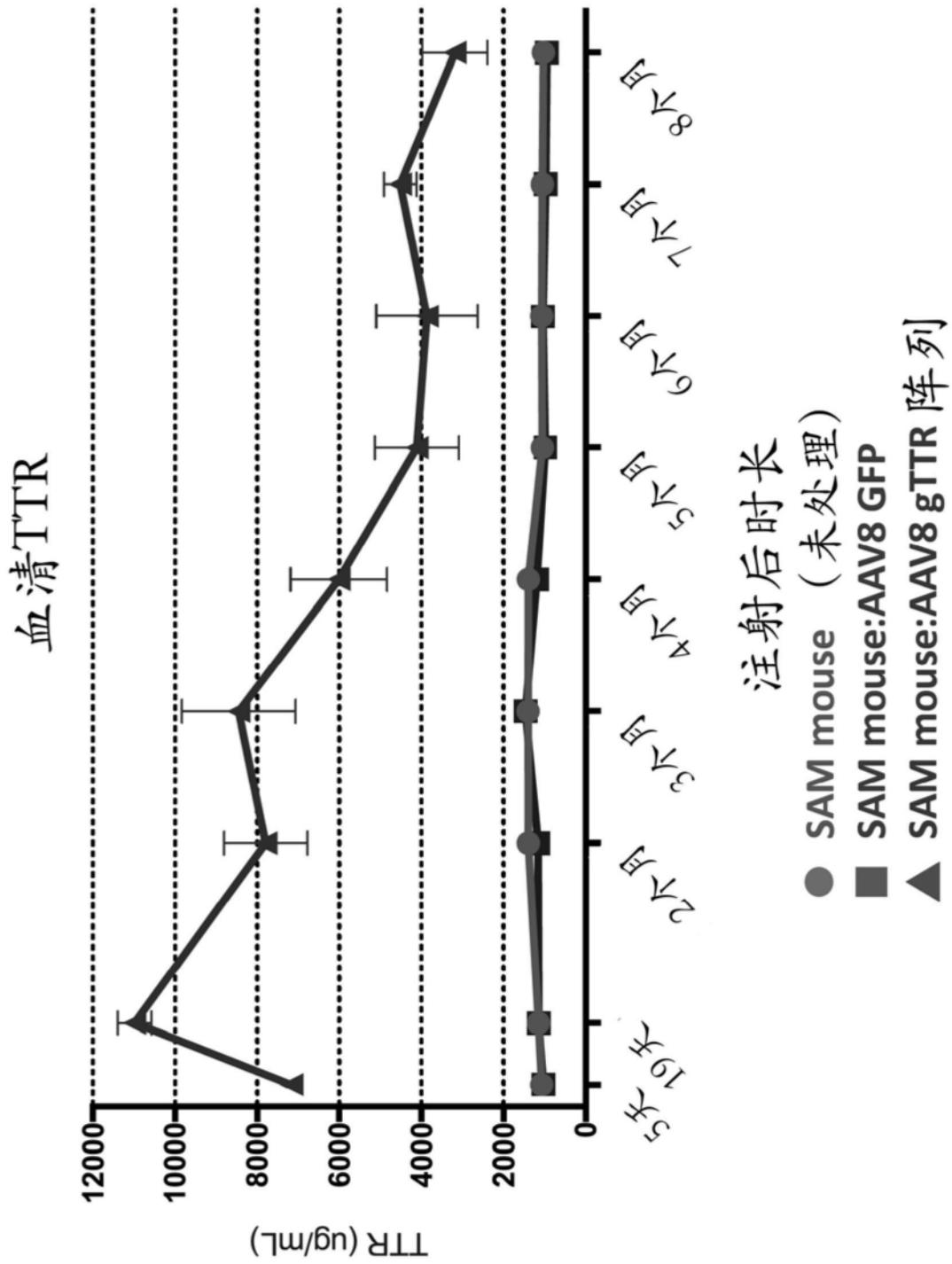


图14

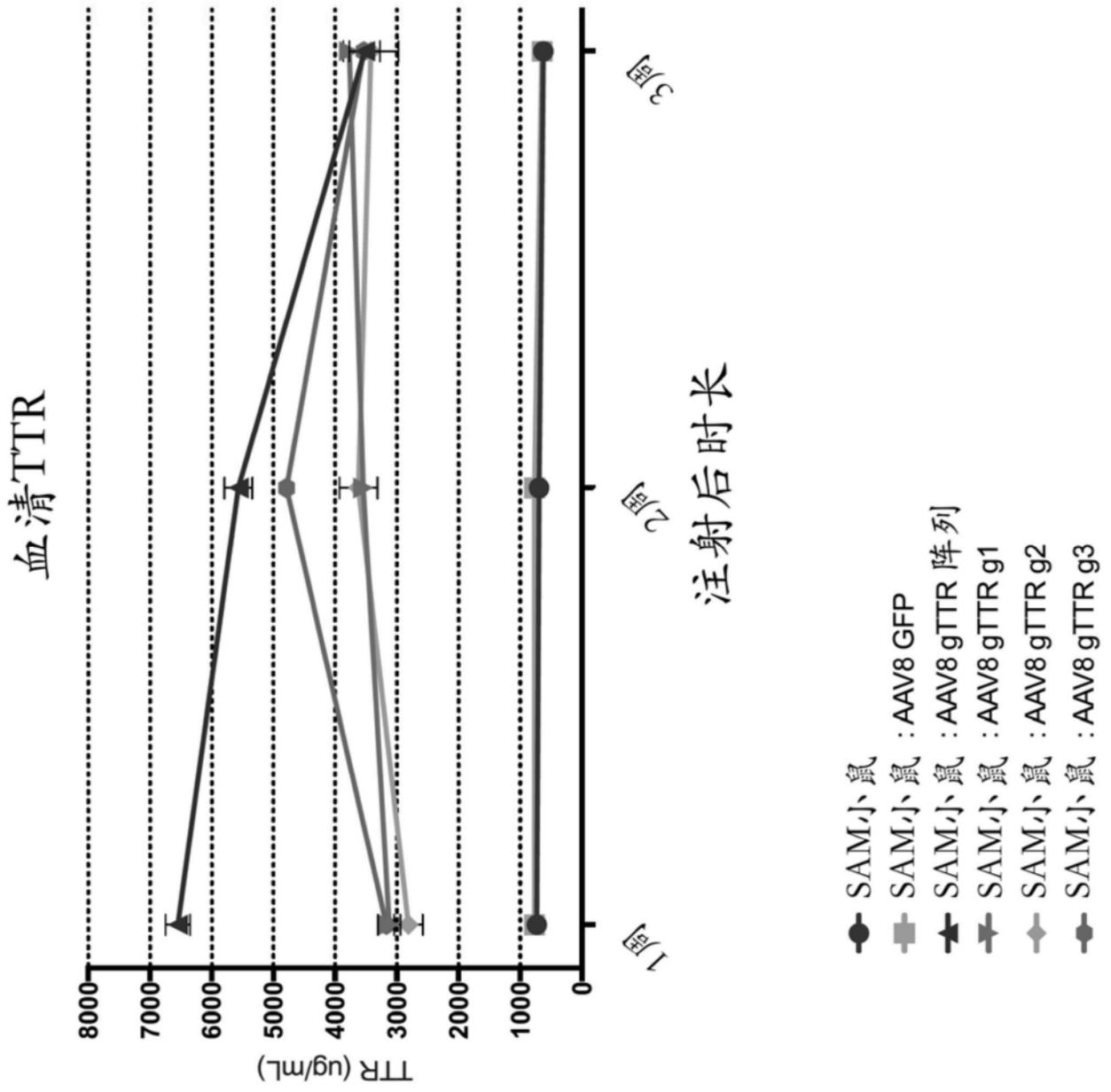


图15

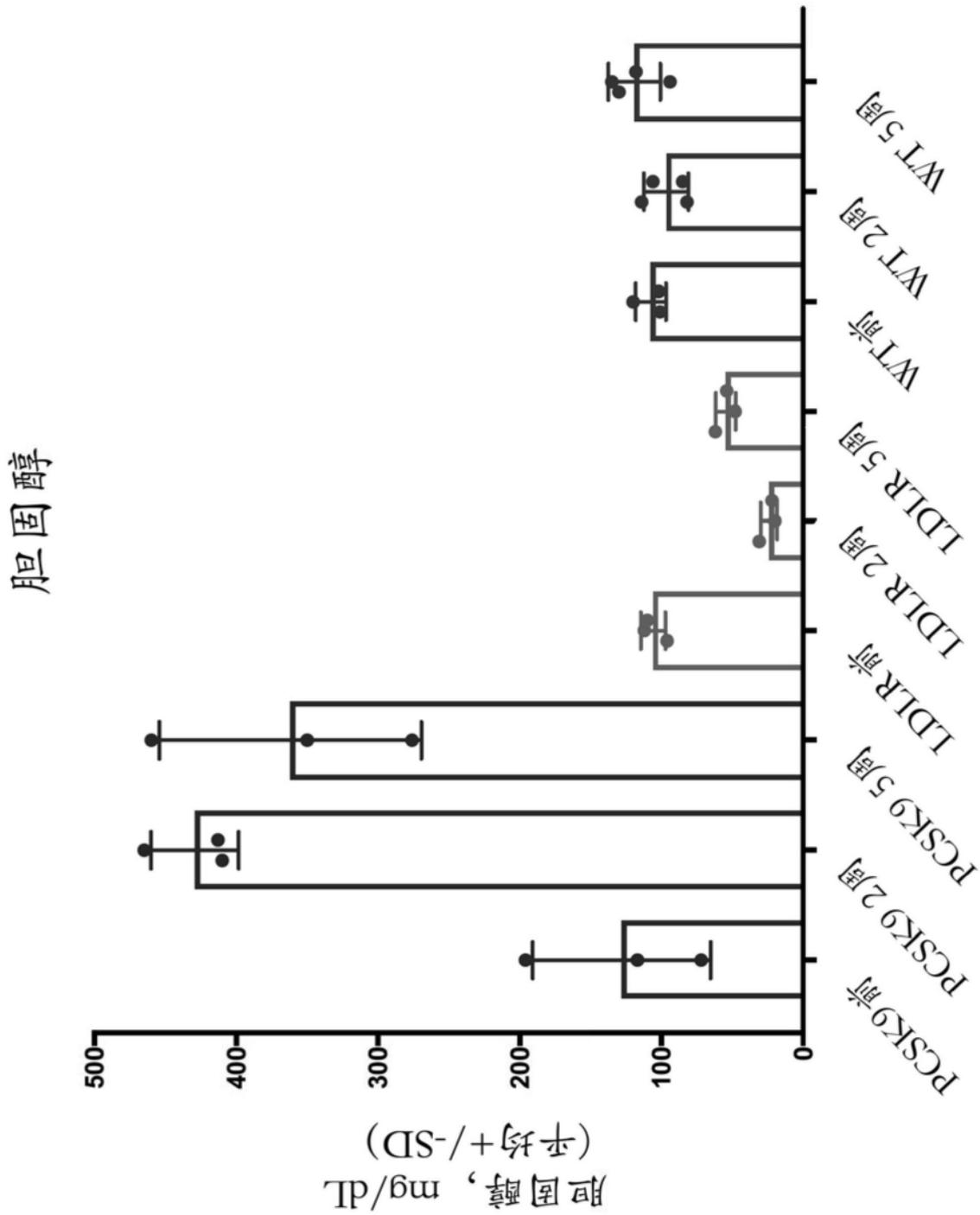


图16A

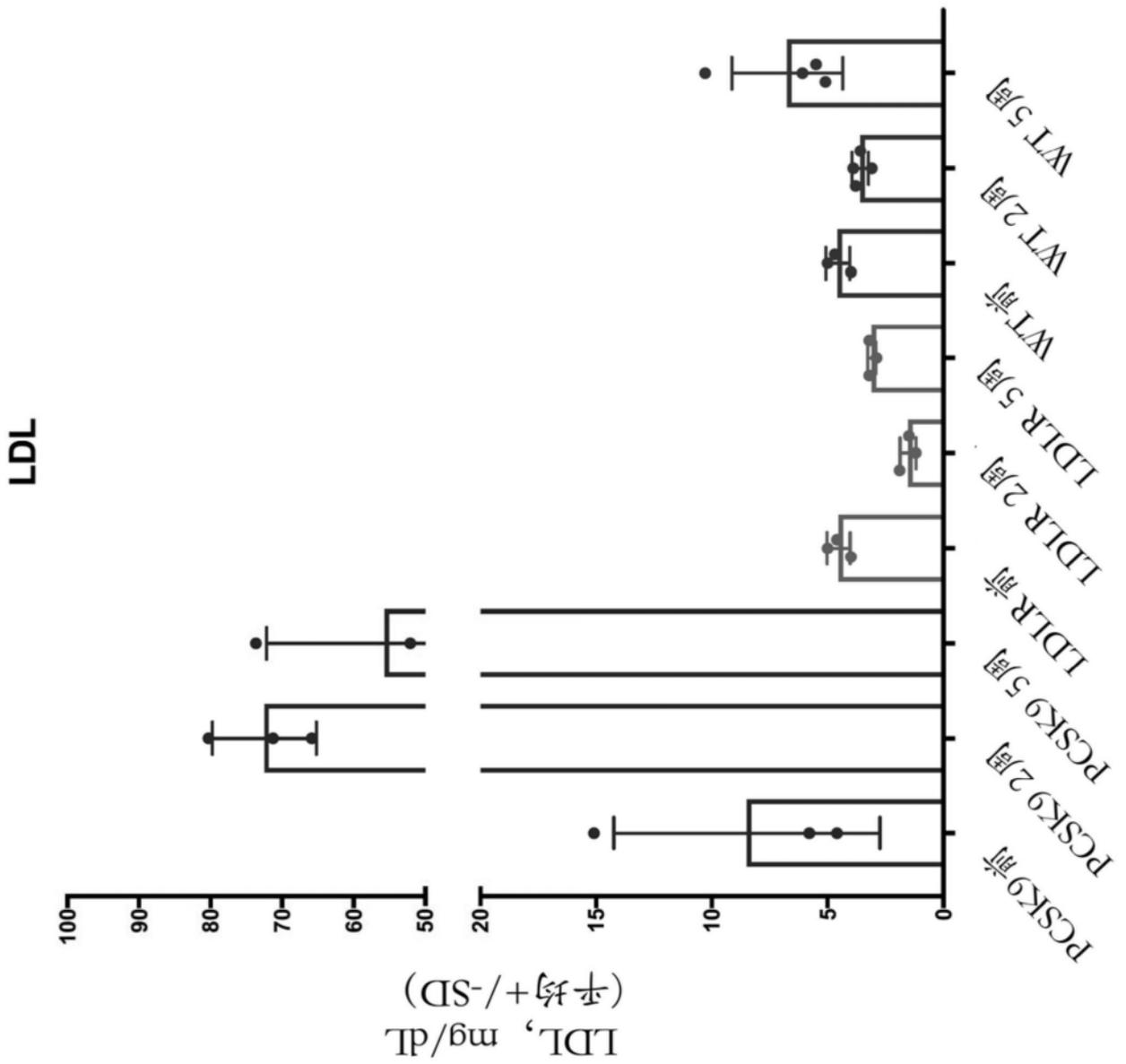


图16B

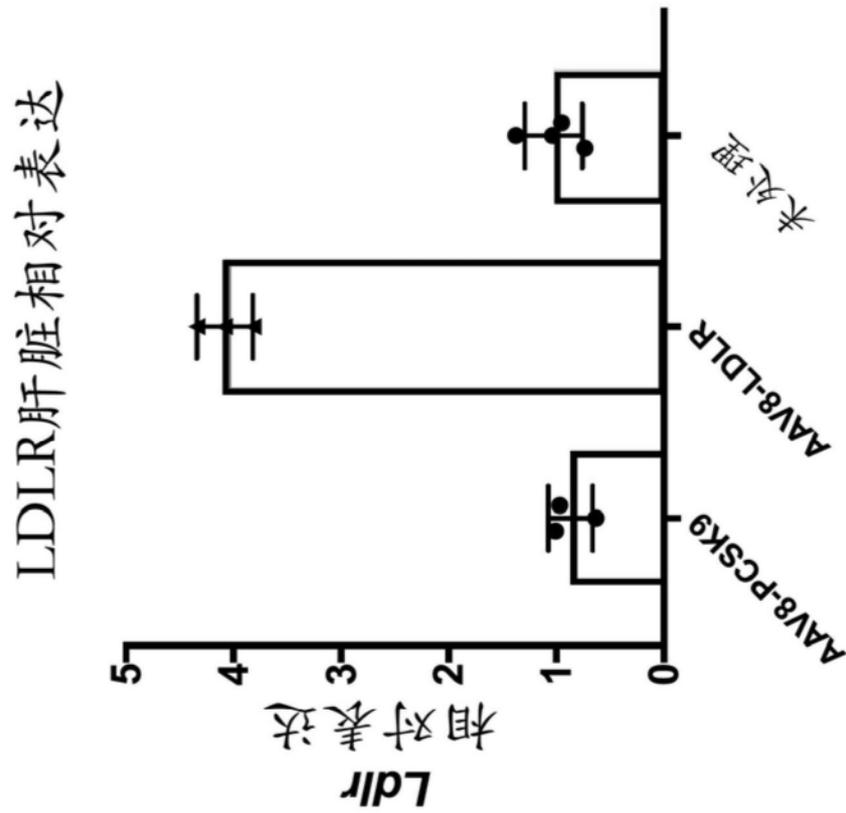


图17A

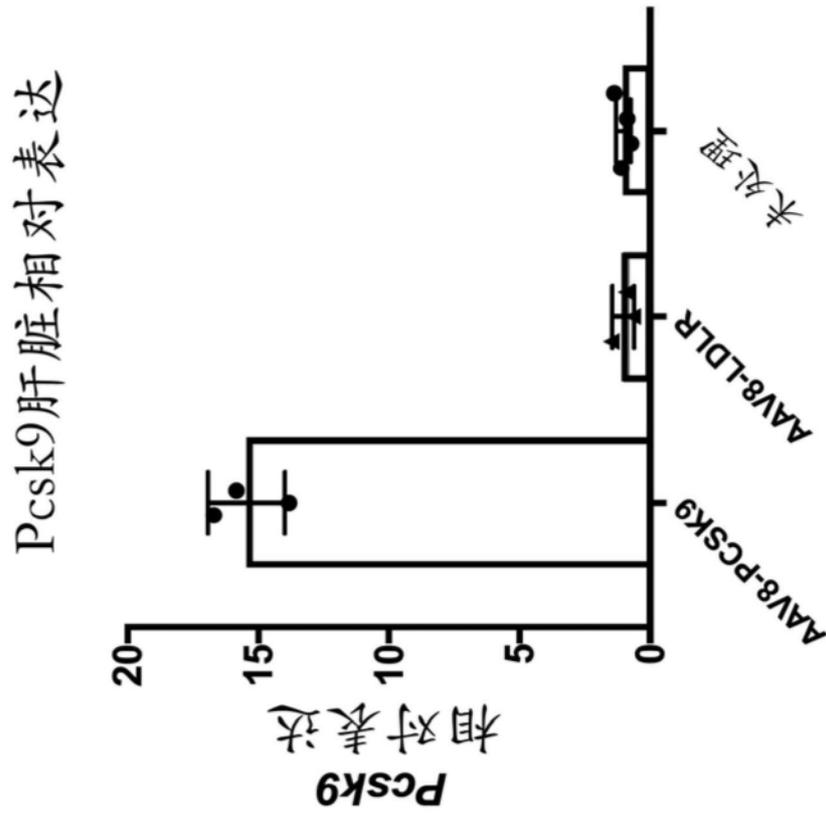


图17B

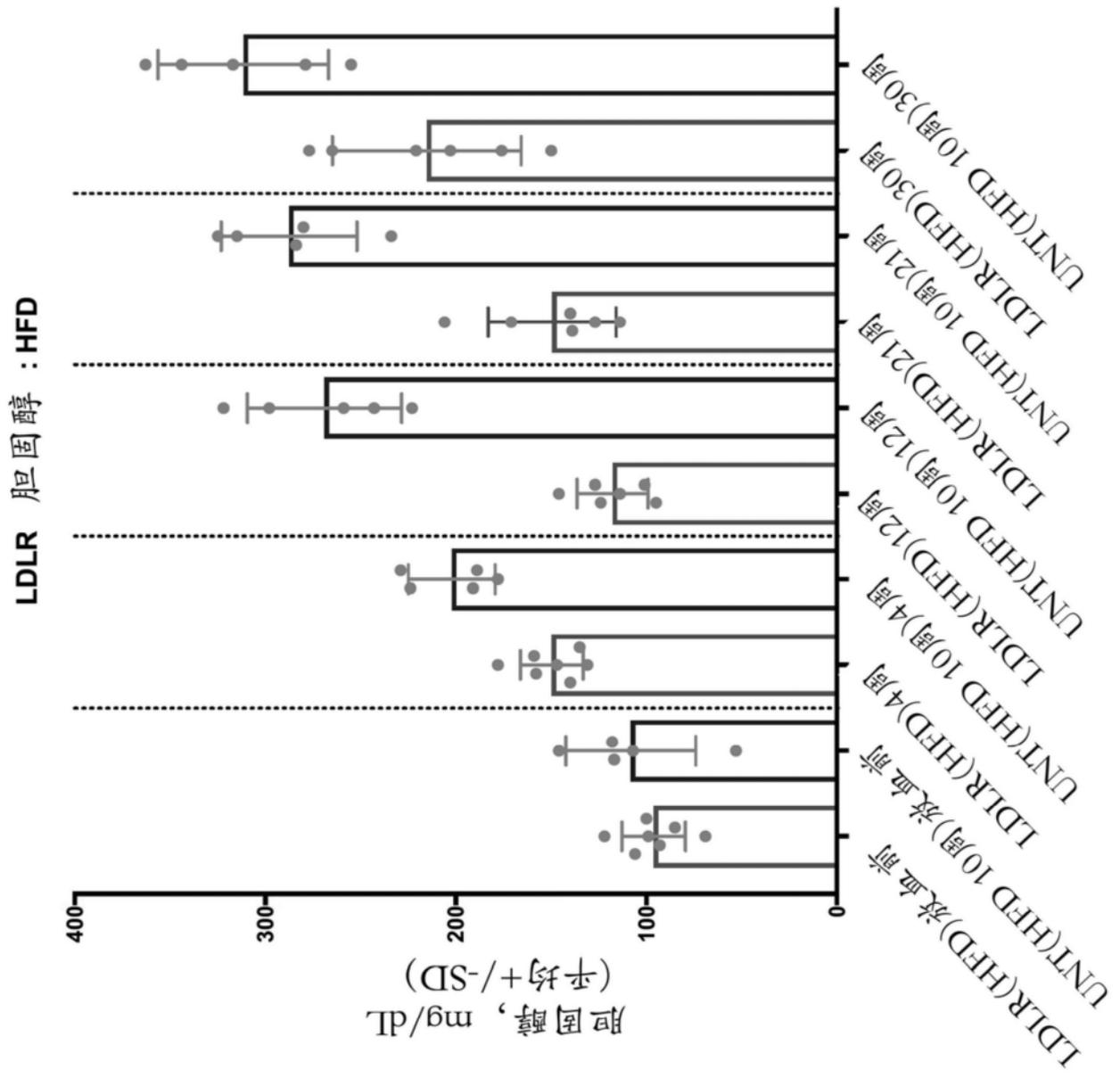


图18A

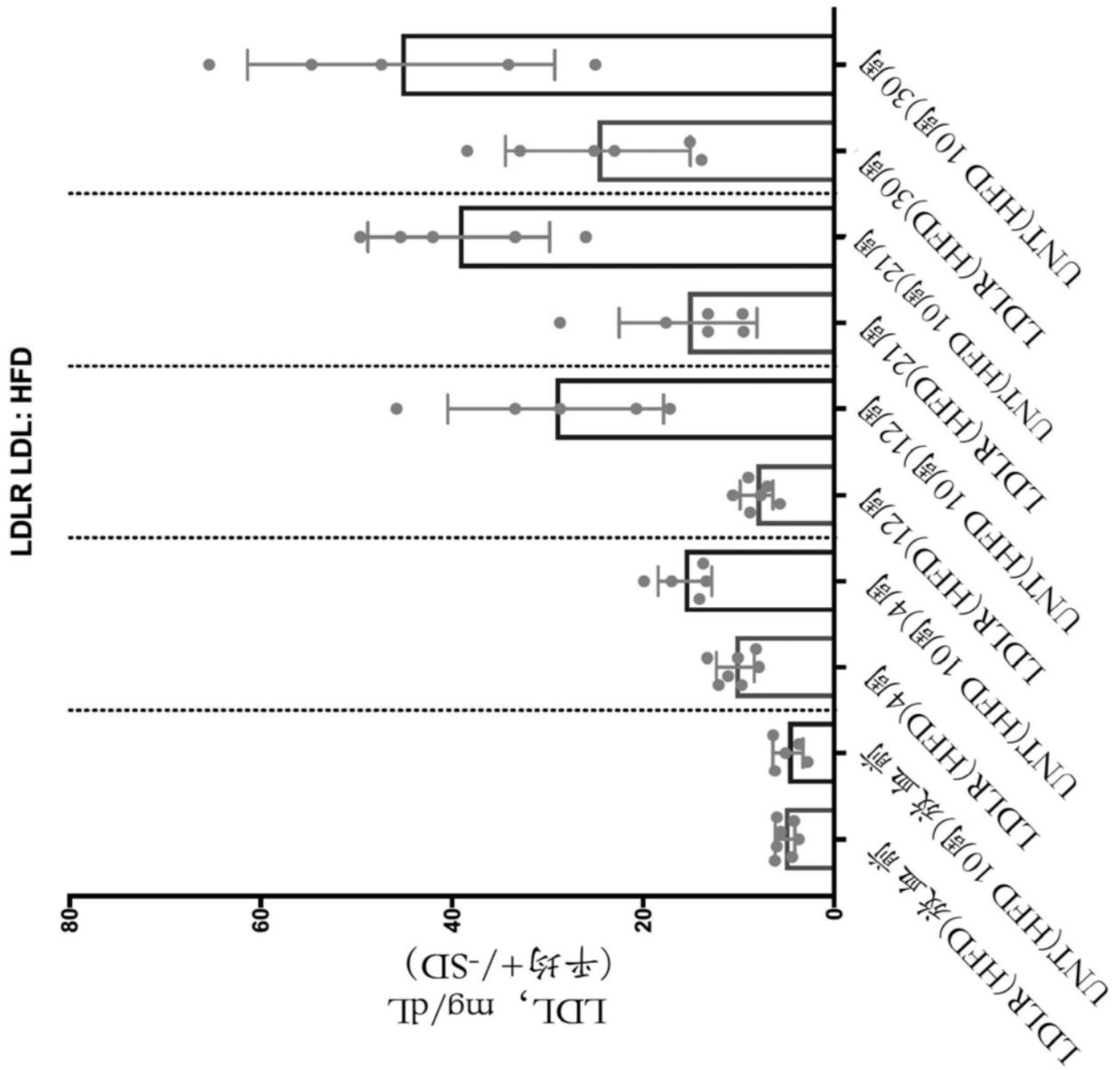


图18B

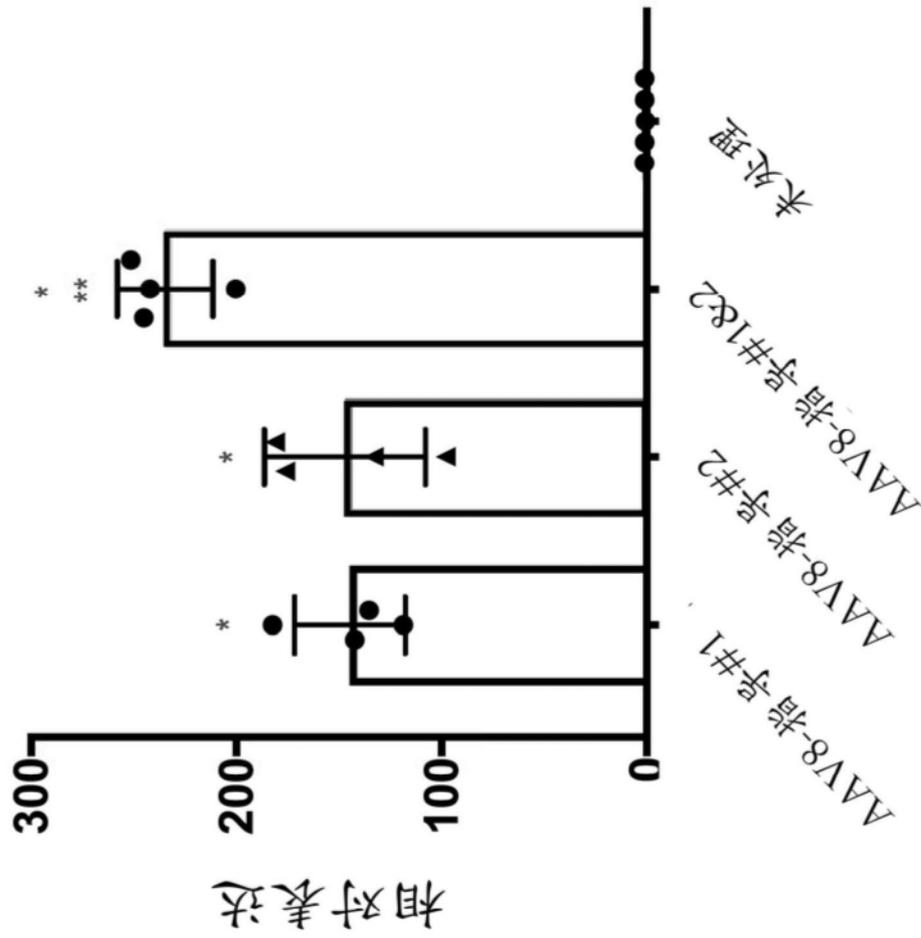


图19