



(12) 发明专利申请

(10) 申请公布号 CN 116064550 A

(43) 申请公布日 2023. 05. 05

(21) 申请号 202310128672.X

(22) 申请日 2019.12.13

(30) 优先权数据

62/782,461 2018.12.20 US

62/829,995 2019.04.05 US

(62) 分案原申请数据

201980091769.X 2019.12.13

(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 大辅·梶村

阿尔蒂·夏尔马-康宁

布列塔尼·杜波斯

古斯塔沃·德罗格特 家珍·萧

琼科·库诺 大卫·弗伦杜威

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 44202

专利代理师 郝传鑫

(51) Int. Cl.

C12N 15/113 (2010.01)

A01K 67/027 (2006.01)

C12N 15/85 (2006.01)

C12N 15/90 (2006.01)

C12N 15/12 (2006.01)

C12N 9/22 (2006.01)

C12Q 1/6883 (2018.01)

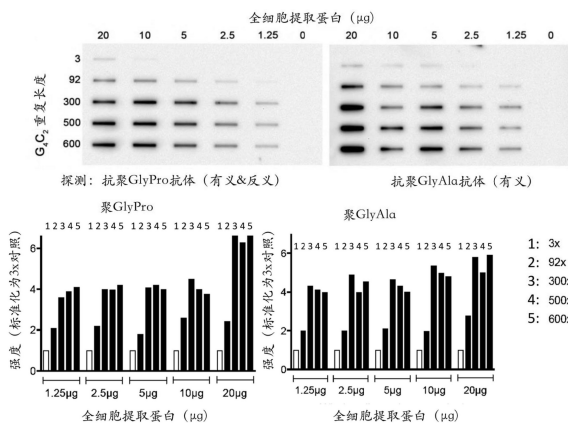
权利要求书5页 说明书71页
序列表(电子公布) 附图39页

(54) 发明名称

核酸酶介导的重复扩增

(57) 摘要

提供了用于扩增已经存在于基因组基因座上的重复的核酸酶介导的方法。包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物基因组、非人类动物细胞,和非人类动物,还提供了通过核酸酶介导的重复扩增制备此类非人类动物细胞和非人类动物的方法。还提供了使用非人类动物细胞或非人类动物识别可用于预防、延迟或治疗与在C9orf72基因座处的重复扩增相关的一种或多种神经退行性疾病的治疗候选物的方法。



1. 一种非人类动物基因组,其包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列。

2. 根据权利要求1所述的非人类动物基因组,其中所述异源六核苷酸重复扩增序列包含至少约300个重复的SEQ ID NO:1所示的六核苷酸序列。

3. 根据权利要求2所述的非人类动物基因组,其中所述异源六核苷酸重复扩增包含至少约500个重复的SEQ ID NO:1所示的六核苷酸序列。

4. 根据权利要求3所述的非人类动物基因组,其中所述异源六核苷酸重复扩增包含至少约600个重复的SEQ ID NO:1所示的六核苷酸序列。

5. 根据权利要求1-4中任一项所述的非人类动物基因组,其中所述重复在所述异源重复扩增序列中是连续的。

6. 根据权利要求1-4中任一项所述的非人类动物基因组,其中所述异源六核苷酸重复扩增序列位于所述内源C9orf72基因座的第一个非编码内源外显子和外显子2之间。

7. 根据权利要求1-4中任一项所述的非人类动物基因组,其中所述内源C9orf72基因座包含人C9ORF72核苷酸序列。

8. 根据权利要求7所述的非人类动物基因组,其中所述人C9ORF72核苷酸序列包含SEQ ID NO:46和/或SEQ ID NO:47。

9. 根据权利要求1-4和8中任一项所述的非人类动物基因组,其中所述非人类动物是啮齿动物。

10. 根据权利要求9所述的非人类动物基因组,其中所述啮齿动物是大鼠或小鼠。

11. 根据权利要求1-4、8和10中任一项所述的非人类动物基因组,其中所述非人类动物基因组对于所述异源六核苷酸重复扩增序列而言是杂合的。

12. 根据权利要求1-4、8和10中任一项所述的非人类动物基因组,其中包含所述非人类动物基因组的非人类动物或非人类动物细胞表现出:

(a) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,保留内含子序列的C9orf72转录物的增加的表达;和/或

(b) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,增加的RNA病灶数量;和/或

(c) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,增加的二肽重复蛋白水平。

13. 一种非人类动物细胞,所述非人类动物细胞在其基因组中包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列,且其中所述非人类动物细胞不是胚胎干(ES)细胞。

14. 根据权利要求13所述的非人类动物细胞,其中所述异源六核苷酸重复扩增序列包含至少约300个重复的SEQ ID NO:1所示的六核苷酸序列。

15. 根据权利要求14所述的非人类动物细胞,其中所述异源六核苷酸重复扩增包含至少约500个重复的SEQ ID NO:1所示的六核苷酸序列。

16. 根据权利要求15所述的非人类动物细胞,其中所述异源六核苷酸重复扩增包含至

少约600个重复的SEQ ID NO:1所示的六核苷酸序列。

17. 根据权利要求13-16中任一项所述的非人类动物细胞,其中所述重复在所述异源重复扩增序列中是连续的。

18. 根据权利要求13-16中任一项所述的非人类动物细胞,其中所述异源六核苷酸重复扩增序列位于所述内源C9orf72基因座的第一个非编码内源外显子和外显子2之间。

19. 根据权利要求13-16中任一项所述的非人类动物细胞,其中所述内源C9orf72基因座包含人C9ORF72核苷酸序列。

20. 根据权利要求19所述的非人类动物细胞,其中所述人C9ORF72核苷酸序列包含SEQ ID NO:46和/或SEQ ID NO:47。

21. 根据权利要求13-16和20中任一项所述的非人类动物细胞,其中所述非人类动物是啮齿动物。

22. 根据权利要求21所述的非人类动物细胞,其中所述非人类动物是大鼠或小鼠。

23. 根据权利要求13-16、20和22中任一项所述的非人类动物细胞,其中所述非人类动物细胞对于所述异源六核苷酸重复扩增序列而言是杂合的。

24. 根据权利要求13-16、20和22中任一项所述的非人类动物细胞,其中所述非人类动物细胞表现出:

(a) 与包含野生型C9orf72基因座的对照非人类动物细胞相比,保留内含子序列的C9orf72转录物的增加的表达;和/或

(b) 与包含野生型C9orf72基因座的对照非人类动物细胞相比,增加的RNA病灶数量;和/或

(c) 与包含野生型C9orf72基因座的对照非人类动物细胞相比,增加的二肽重复蛋白水平。

25. 根据权利要求13-16、20和22中任一项所述的非人类动物细胞,其中所述非人类动物细胞是胚胎干细胞衍生的运动神经元、脑细胞、皮层细胞、神经元细胞、肌肉细胞、心脏细胞,或生殖细胞。

26. 根据权利要求25所述的非人类动物细胞,其中所述非人类动物细胞是胚胎干细胞衍生的运动神经元。

27. 根据权利要求13-16、20、22和26中任一项所述的非人类动物细胞,其中所述非人类动物细胞是体外的。

28. 根据权利要求13-16、20、22和26中任一项所述的非人类动物细胞,其中所述非人类动物细胞是体内的。

29. 一种核酸,其包含非人类动物C9orf72基因,所述非人类动物C9orf72基因包含异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列。

30. 根据权利要求29所述的核酸,其中所述异源六核苷酸重复扩增序列包含至少约300个重复的SEQ ID NO:1所示的六核苷酸序列。

31. 根据权利要求30所述的核酸,其中所述异源六核苷酸重复扩增包含至少约500个重复的SEQ ID NO:1所示的六核苷酸序列。

32. 根据权利要求31所述的核酸,其中所述异源六核苷酸重复扩增包含至少约600个重

复的SEQ ID NO:1所示的六核苷酸序列。

33. 根据权利要求29-32中任一项所述的核酸,其中所述重复在所述异源重复扩增序列中是连续的。

34. 根据权利要求29-32中任一项所述的核酸,其中所述异源六核苷酸重复扩增序列位于所述内源C9orf72基因的第二个非编码内源外显子和外显子2之间。

35. 根据权利要求29-32中任一项所述的核酸,其中所述内源C9orf72基因包含人C9ORF72核苷酸序列。

36. 根据权利要求35所述的核酸,其中所述人C9ORF72核苷酸序列包含SEQ ID NO:46和/或SEQ ID NO:47。

37. 根据权利要求29-32和36中任一项所述的核酸,所述非人类动物是啮齿动物。

38. 根据权利要求37所述的核酸,其中所述啮齿动物是大鼠或小鼠。

39. 根据权利要求29-32、36和38中任一项所述的核酸,其中包含所述核酸的非人类动物表现出:

(a) 与包含野生型C9orf72基因座的对照非人类动物相比,保留内含子序列的C9orf72转录物的增加的表达;和/或

(b) 与包含野生型C9orf72基因座的对照非人类动物相比,增加的RNA病灶数量;和/或

(c) 与包含野生型C9orf72基因座的对照非人类动物相比,增加的二肽重复蛋白水平。

40. 评估用于治疗与C9orf72基因座处的六核苷酸重复扩增序列相关的疾病或病症的治疗候选物的方法,包括:

(a) 向在其基因组中包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物或非人类动物细胞施用候选药剂,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列;

(b) 进行一个或多个测定,以确定所述候选药剂是否对与所述疾病或病症相关的一个或多个体征或症状有影响;以及

(c) 识别对与所述疾病或病症相关的所述一个或多个体征或症状有影响的所述候选药剂作为治疗候选物。

41. 根据权利要求40所述的方法,其中将所述候选药剂体内施用给所述非人类动物,并且在施用所述候选药剂之后,在从所述非人类动物分离的细胞中体外进行所述一个或多个测定。

42. 根据权利要求40所述的方法,其中将所述候选药剂体外施用于非人胚胎干细胞衍生的运动神经元。

43. 根据权利要求40-42中任意一项所述的方法,其中所述一个或多个测定包括定量聚合酶链反应(qPCR)以检测含有内含子的C9orf72 RNA转录物。

44. 根据权利要求40-42中任意一项所述的方法,其中所述一个或多个测定包括测量包含C9orf72有义或反义RNA转录物的RNA病灶,可选地,其中通过荧光原位杂交测量所述RNA病灶。

45. 根据权利要求40-42中任意一项所述的方法,其中所述一个或多个测定包含测量二肽重复蛋白的积累,可选地,其中所述二肽重复蛋白是聚GA二肽重复蛋白或聚GP二肽重复蛋白,并且可选地,其中通过免疫组织化学测量所述二肽重复蛋白的积累。

46. 根据权利要求40-42中任意一项所述的方法,其中所述异源六核苷酸重复扩增序列包含至少约300个重复的SEQ ID NO:1所示的六核苷酸序列。

47. 根据权利要求46所述的方法,其中所述异源六核苷酸重复扩增包含至少约500个重复的SEQ ID NO:1所示的六核苷酸序列。

48. 根据权利要求47所述的方法,其中所述异源六核苷酸重复扩增包含至少约600个重复的SEQ ID NO:1所示的六核苷酸序列。

49. 根据权利要求40-42、47和48中任意一项所述的方法,其中所述重复在所述异源重复扩增序列中是连续的。

50. 根据权利要求40-42、47和48中任意一项所述的方法,其中所述异源六核苷酸重复扩增序列位于所述内源C9orf72基因座的第一个非编码内源外显子和外显子2之间。

51. 根据权利要求40-42、47和48中任意一项所述的方法,其中所述内源C9orf72基因座包含人C9ORF72核苷酸序列。

52. 根据权利要求51所述的方法,其中所述人C9ORF72核苷酸序列包含SEQ ID NO:46和/或SEQ ID NO:47。

53. 根据权利要求40-42、47、48和52中任意一项所述的方法,其中所述非人类动物是啮齿动物。

54. 根据权利要求53所述的方法,其中所述啮齿动物是大鼠或小鼠。

55. 根据权利要求40-42、47、48、52和54中任意一项所述的方法,其中所述非人类动物或所述非人类动物细胞对于所述异源六核苷酸重复扩增序列而言是杂合的。

56. 根据权利要求40-42、47、48、52和54中任意一项所述的方法,其中所述非人类动物或所述非人类动物细胞表现出:

(a) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,保留内含子序列的C9orf72转录物的增加的表达;和/或

(b) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,增加的RNA病灶数量;和/或

(c) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,增加的二肽重复蛋白水平。

57. 一种制备非人类动物的方法,所述非人类动物在其基因组中包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,所述方法包括:

(I) (a) 修饰非人类动物多能细胞的基因组以包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列;

(b) 识别或选择包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的遗传修饰的非人类动物多能细胞;

(c) 将该遗传修饰的非人类动物多能细胞导入非人类动物宿主胚胎;以及

(d) 在非人类动物代孕母体内孕育所述非人类动物宿主胚胎;或

(II) (a) 修饰非人类动物单细胞期胚胎的基因组以包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列;

(b) 选择包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的遗传修饰的胚胎;和

(c) 在非人类动物代孕母体内孕育所述非人类动物宿主胚胎。

58. 根据权利要求57所述的方法,其中所述异源六核苷酸重复扩增序列包含至少约300个重复的SEQ ID N0:1所示的六核苷酸序列。

59. 根据权利要求58所述的方法,其中所述异源六核苷酸重复扩增包含至少约500个重复的SEQ ID N0:1所示的六核苷酸序列。

60. 根据权利要求59所述的方法,其中所述异源六核苷酸重复扩增包含至少约600个重复的SEQ ID N0:1所示的六核苷酸序列。

61. 根据权利要求57-60中任意一项所述的方法,其中所述重复在所述异源重复扩增序列中是连续的。

62. 根据权利要求57-60中任意一项所述的方法,其中所述异源六核苷酸重复扩增序列位于所述内源C9orf72基因座的第一个非编码内源外显子和外显子2之间。

63. 根据权利要求57-60中任意一项所述的方法,其中所述内源C9orf72基因座包含人C9ORF72核苷酸序列。

64. 根据权利要求63所述的方法,其中所述人C9ORF72核苷酸序列包含SEQ ID N0:46和/或SEQ ID N0:47。

65. 根据权利要求57-60和64中任意一项所述的方法,其中所述非人类动物是啮齿动物。

66. 根据权利要求65所述的方法,其中所述啮齿动物是大鼠或小鼠。

67. 根据权利要求57-60、64和66中任意一项所述的方法,其中所述非人类动物对于所述异源六核苷酸重复扩增序列而言是杂合的。

68. 根据权利要求57-60、64和66中任意一项所述的方法,其中所述非人类动物表现出:

(a) 与包含野生型C9orf72基因座的对照非人类动物相比,保留内含子序列的C9orf72转录物的增加的表达;和/或

(b) 与包含野生型C9orf72基因座的对照非人类动物相比,增加的RNA病灶数量;和/或

(c) 与包含野生型C9orf72基因座的对照非人类动物相比,增加的二肽重复蛋白水平。

核酸酶介导的重复扩增

分案申请说明

[0001] 本申请是申请日为2019年12月13日,申请号为201980091769.X,发明名称为“核酸酶介导的重复扩增”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2018年12月20日提交的美国申请第62/782,461号和2019年4月5日提交的美国申请第62/839,995号的权益,这些申请各自以全文引用的方式并入本文中。

对通过EFS Web提交为XML文件的序列表的引用

[0003] 写在文件556983SEQLIST.xml中的序列表为147.0KB,创建于2023年2月13日(实际序列内容创建于2019年12月12日),通过引用并入。

背景技术

[0004] 由分散在整个人类基因组中的简单序列重复的扩增引起四十多种疾病,其中大部分主要影响神经系统。首先发现的是扩增的三核苷酸重复疾病,并且其仍然是最常见的。最近,四核苷酸、五核苷酸、六核苷酸甚至十二核苷酸重复扩增已被确定为人类疾病(包括一些最常见的神经遗传疾病)的病因。重复扩增疾病包括强直性肌营养不良(DM1和DM2)的病因、肌萎缩侧索硬化(ALS)/额颞叶痴呆(FTD)(C9ORF72)的最常见遗传原因、亨廷顿病和其它八种多聚谷氨酰胺疾病,包括最常见的显性遗传性共济失调、最常见的隐性共济失调(弗里德赖希共济失调),和最常见遗传性智力低下(脆性X综合征)。例如,在C9ORF72基因的非编码区内的GGGGCC(SEQ ID NO:1)的扩增六核苷酸重复序列已与ALS和FTD相关联。目前,这两种疾病都无法治愈。

[0005] 虽然各种实验动物模型广泛地用于大部分治疗剂的开发,但很少有模型能够以提供阐明所识别的遗传成分导致疾病的确切分子机制的方式来解决重复扩增疾病,例如神经退行性疾病和炎症性疾病。理想的动物模型含有相同的基因组分并且表现人类疾病的类似特征。鉴于物种之间的遗传差异,对于研发出近似地再现重复扩增疾病(例如人类神经退行性和/或炎症性疾病)的改进的动物模型存在高度未满足的需求。此类改进的动物模型为研发有效的治疗和/或预防药剂提供了重大价值。

发明内容

[0006] 提供了扩增重复扩增序列的方法,所述重复扩增序列在细胞中的靶基因组基因座处包含复数个拷贝的重复序列,并且还提供了非人类动物基因组、非人类动物细胞和非人类动物,它们的基因组中包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列。

[0007] 在一方面,提供了扩增重复扩增序列的方法,所述重复扩增序列在细胞中的靶基因组基因座处包含复数个拷贝的重复序列,或生产修饰的细胞的方法,所述修饰的细胞在靶基因组基因座处具有增加的重复序列拷贝数。一些此类方法包括将核酸酶试剂引入包含所述重复扩增序列的细胞群中,所述核酸酶试剂切割所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生具有扩增的重复扩增序列的修饰的细胞群。一些此类方法包

括将核酸酶试剂或编码所述核酸酶试剂的核酸引入包含重复扩增序列的细胞群中,其中所述核酸酶试剂切割所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生具有扩增的重复扩增序列的修饰的细胞群。一些此类方法包括将核酸酶试剂引入包含重复扩增序列的细胞群中,所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生具有扩增的重复扩增序列的修饰的细胞群。一些此类方法包括将核酸酶试剂或编码所述核酸酶试剂的核酸引入包含重复扩增序列的细胞群中,其中所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生具有扩增的重复扩增序列的修饰的细胞群。一些这样的方法包括:(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂引入所述细胞群,所述核酸酶试剂切割重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。一些这样的方法包括:(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂或编码所述核酸酶试剂的核酸引入所述细胞群,其中所述核酸酶试剂切割重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。一些这样的方法包括:(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂引入所述细胞群,其中所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。一些这样的方法包括:(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂或编码所述核酸酶试剂的核酸引入所述细胞群,其中所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。

[0008] 在一些这样的方法中,不将外源修复模板引入所述细胞群。

[0009] 在一些这样的方法中,所述核酸酶靶位点位于所述重复扩增序列的5'末端或3'末端的约1000、约500、约400、约300、约200、约100、约50、约40、约30、约20,或约10个核苷酸内,或与所述重复扩增序列的5'末端或3'末端重叠。在一些此类方法中,所述核酸酶靶位点位于所述重复扩增序列的5'末端或3'末端的约100、约50、约40、约30、约20,或约10个核苷酸内,或与所述重复扩增序列的5'末端或3'末端重叠。在一些此类方法中,所述核酸酶靶位点位于所述重复扩增序列的5'末端或3'末端的约50、约40、约30、约20,或约10个核苷酸内,或与所述重复扩增序列的5'末端或3'末端重叠。在一些此类方法中,所述核酸酶靶位点位于所述重复扩增序列的5'末端或3'末端的约25、约24、约23、约22、约21、约20、约19、约18、约17、约16、约15、约14、约13、约12、约11、约10、约9、约8、约7、约6、约5、约4、约3、约2或约1个核苷酸内,或与所述重复扩增序列的5'末端或3'末端重叠。在一些此类方法中,所述核酸酶靶位点与所述重复扩增序列的5'末端或3'末端重叠。

[0010] 在一些此类方法中,核酸酶试剂为锌指核酸酶(ZFN)、类转录激活因子效应物核酸

酶 (TALEN)、和成簇的规律间隔的短回文重复序列 (CRISPR) - 相关 (Cas) 蛋白和向导 RNA。可选地, 所述核酸酶试剂是 Cas 蛋白和向导 RNA。可选地, 所述 Cas 蛋白是 Cas9 蛋白。

[0011] 在一些此类方法中, 核酸酶靶位点靠近所述重复扩增序列的 5' 末端。在一些此类方法中, 所述核酸酶靶位点靠近所述重复扩增序列的 3' 末端。在一些此类方法中, 所述核酸酶靶位点位于所述重复扩增序列外。

[0012] 在一些此类方法中, 所述核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的 5' 末端或 3' 末端的约 1000、约 500、约 400、约 300、约 200、约 100、约 50、约 40、约 30、约 20 或约 10 个核苷酸内。在一些此类方法中, 所述核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的 5' 末端或 3' 末端的约 100、约 50、约 40、约 30、约 20, 或约 10 个核苷酸内, 或与所述重复扩增序列的 5' 末端或 3' 末端重叠。可选地, 所述核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的 5' 末端或 3' 末端的约 50、约 40、约 30、约 20 或约 10 个核苷酸内。在一些此类方法中, 所述核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的 5' 末端或 3' 末端的约 25、约 24、约 23、约 22、约 21、约 20、约 19、约 18、约 17、约 16、约 15、约 14、约 13、约 12、约 11、约 10、约 9、约 8、约 7、约 6、约 5、约 4、约 3、约 2, 或约 1 个核苷酸内, 或与所述重复扩增序列的 5' 末端或 3' 末端重叠。在一些此类方法中, 所述核酸酶试剂造成双链断裂或单链断裂的位点靠近重复扩增序列的 5' 末端。在一些此类方法中, 所述核酸酶试剂造成双链断裂或单链断裂的位点靠近重复扩增序列的 3' 末端。在一些此类方法中, 所述核酸酶试剂造成双链断裂或单链断裂的位点在所述重复扩增序列之外。在一些此类方法中, 在所述细胞修复所述双链断裂或单链断裂后, 所述核酸酶靶位点被保留。在一些此类方法中, 所述双链断裂或单链断裂的修复不导致所述重复扩增序列之外的插入或删除。

[0013] 在一些此类方法中, 所述核酸酶试剂在所述重复扩增序列的 5' 末端或 3' 末端附近造成双链断裂。在一些此类方法中, 所述核酸酶试剂为在所述重复扩增序列的 5' 末端或 3' 末端附近造成单链断裂的切口酶。

[0014] 在一些此类方法中, 将第一核酸酶试剂和第二核酸酶试剂引入所述细胞中, 其中所述第一核酸酶试剂切割所述重复扩增序列的 5' 末端附近的第一核酸酶靶位点, 所述第二核酸酶试剂切割所述重复扩增序列的 3' 末端附近的第二核酸酶靶位点。在一些此类方法中, 将第一核酸酶试剂或编码所述第一核酸酶试剂的核酸和第二核酸酶试剂或编码所述第二核酸酶试剂的核酸引入所述细胞中, 其中所述第一核酸酶试剂切割所述重复扩增序列的 5' 末端附近的第一核酸酶靶位点, 且所述第二核酸酶试剂切割所述重复扩增序列的 3' 末端附近的第二核酸酶靶位点。在一些此类方法中, 将第一核酸酶试剂和第二核酸酶试剂引入所述细胞中, 其中所述第一核酸酶试剂在所述重复扩增序列的 5' 末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂, 并且所述第二核酸酶试剂在所述重复扩增序列的 3' 末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂。在一些此类方法中, 将第一核酸酶试剂或编码所述第一核酸酶试剂的核酸和第二核酸酶试剂或所述编码第二核酸酶试剂的核酸引入所述细胞中, 其中所述第一核酸酶试剂在所述重复扩增序列的 5' 末端附近的第一个核酸酶靶位点处造成双链断裂或单链断裂, 并且所述第二核酸酶试剂在所述重复扩增序列的 3' 末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂。在一些此类方法中, 所述第一核酸酶试剂在所述重复扩增序列的 5' 末端附近的第一核酸酶靶位点处造成双链断

裂,且所述第二核酸酶试剂在所述重复扩增序列的3'末端附近的第二核酸酶靶位点处造成双链断裂。在一些此类方法中,所述第一核酸酶试剂在所述重复扩增序列的5'末端附近的第一核酸酶靶位点处造成单链断裂,且所述第二核酸酶试剂在所述重复扩增序列3'末端附近第二核酸酶靶位点处造成单链断裂。

[0015] 在一些此类方法中,所述重复扩增序列是异源重复扩增序列。

[0016] 在一些此类方法中,所述重复扩增序列包含所述重复序列的至少约10、至少约20、至少约30、至少约40、至少约50、至少约60、至少约70、至少约80或至少约90个拷贝。

[0017] 在一些此类方法中,重复序列的拷贝在重复扩增序列中是连续的。

[0018] 在一些此类方法中,其中重复序列是三核苷酸重复、四核苷酸重复、五核苷酸重复、六核苷酸重复,或十二核苷酸重复。

[0019] 在一些此类方法中,所述重复序列包含SEQ ID NO:1-12中的任一个。可选地,(i)所述重复序列包含SEQ ID NO:2(cag),并且所述靶基因组基因座为HTT、AR、ATN1、ATXN1、ATXN2、ATXN3、CACNA1A、ATXN7、PPP2R2B,或TBP基因座;(ii)所述重复序列包含SEQ ID NO:3(cgg),且所述靶基因组基因座为FMR1基因座;(iii)所述重复序列包含SEQ ID NO:4(ctg),且所述靶基因组基因座为DMPK、JPH3、ATXN8,或TCF4基因座;(iv)所述重复序列包含SEQ ID NO:5(gaa),且所述靶基因组基因座为FXN基因座;(v)所述重复序列包含SEQ ID NO:6(gcc),且所述靶基因组基因座为AFF2基因座;(vi)所述重复序列包含SEQ ID NO:7(gcg),且所述靶基因组基因座为PABPN1基因座;(vii)所述重复序列包含SEQ ID NO:8(cctg),且所述靶基因组基因座为CNBP基因座;(viii)所述重复序列包含SEQ ID NO:9(attct),所述靶基因组基因座为ATXN10基因座;(ix)所述重复序列包含SEQ ID NO:10(tggaa),且所述靶基因组基因座为TK2或BEAN1基因座;(x)所述重复序列包含SEQ ID NO:11(ggcctg),且所述靶基因组基因座为NOP56基因座;(xi)所述重复序列包含SEQ ID NO:1,且所述靶基因组基因座为C9ORF72基因座;或(xii)所述重复序列包含SEQ ID NO:12,且所述靶基因组基因座为CSTB基因座。可选地,所述重复序列包含SEQ ID NO:1,并且所述靶基因组基因座是C9ORF72基因座。可选地,所述核酸酶试剂为Cas9蛋白和向导RNA,且所述核酸酶靶位点包含SEQ ID NO:28或33。

[0020] 在一些此类方法中,所述细胞是非人类动物细胞。在一些此类方法中,所述细胞是非人类动物胚胎干细胞、胚胎干细胞衍生的运动神经元、脑细胞、皮质细胞、神经元细胞、肌肉细胞、心脏细胞,或生殖细胞。在一些此类方法,所述细胞是非人类动物单细胞期胚胎。在一些此类方法中,所述细胞是啮齿动物细胞。在一些此类方法中,所述细胞是小鼠细胞或大鼠细胞。可选地,所述细胞是小鼠细胞。可选地,所述细胞是小鼠胚胎干细胞或小鼠单细胞期胚胎。在一些此类方法中,所述细胞是人诱导性多能干细胞。在一些此类方法中,所述细胞在体外。在一些此类方法中,所述细胞在体内。

[0021] 在一些此类方法中,该方法包括多轮步骤(a)-(c),其中在第一轮之后的每一轮中,步骤(a)中的所述细胞群为从上一轮的步骤(c)中选择的所述修饰的细胞扩增而来的克隆细胞群。可选地,所述方法包括至少3轮或至少4轮。可选地,所述方法第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂切割所述重复扩增序列的5'末端附近的第一核酸酶靶位点,以产生第一修饰细胞,且所述方法第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂切割所述重复扩增序列的3'末端附近的第二核酸酶靶位

剂,所述第一核酸酶试剂在所述重复扩增序列的5'末端附近的第一核酸酶靶位点处造成单链断裂,以产生第一修饰细胞,且所述方法第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列的3'末端附近的第二核酸酶靶位点处造成单链断裂,以产生第二修饰细胞。在一些此类方法中,所述方法第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂在所述重复扩增序列的3'末端附近的第一核酸酶靶位点处造成单链断裂,以产生第一修饰细胞,且所述方法第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列的5'末端附近的第二核酸酶靶位点处造成单链断裂,以产生第二修饰细胞。在一些此类方法中,所述方法第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂在所述重复扩增序列的5'末端附近的第一核酸酶靶位点处造成单链断裂,以产生第一修饰细胞,且所述方法第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列的5'末端附近的第二核酸酶靶位点处造成单链断裂,以产生第二修饰细胞。在一些此类方法中,所述方法第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂在所述重复扩增序列的3'末端附近的第一核酸酶靶位点处造成单链断裂,以产生第一修饰细胞,且所述方法第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列的3'末端附近的第二核酸酶靶位点处造成单链断裂,以产生第二修饰细胞。在一些此类方法中,所述第一核酸酶靶位点与所述第二核酸酶靶位点相同。

[0023] 在一些此类方法中,所述核酸酶靶位点在重复扩增序列之外,所述核酸酶靶位点位于所述重复扩增序列的5'末端或3'末端的约50、约40、约30、约20,或约10个核苷酸内,所述核酸酶试剂是Cas9蛋白和向导RNA。在一些这样的方法中,所述核酸酶试剂是在所述重复扩增序列的5'末端或3'末端附近造成单链断裂的切口酶,在所述细胞修复所述单链断裂后,所述核酸酶靶位点被保留,且单链断裂的修复不会导致重复扩增序列之外的插入或删除。

[0024] 在另一方面,提供了一种非人类动物,所述非人类动物在其基因组中包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列。提供了一种非人类动物细胞,所述非人类动物细胞在其基因组中包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列。或提供了一种非人类动物基因组,其包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列。另一方面,提供了非人类动物C9orf72基因,所述非人类动物C9orf72基因包含异源六核苷酸重复扩增序列,其中所述异源六核苷酸重复扩增序列包含超过约100个重复的如SEQ ID NO:1所示的六核苷酸序列。

[0025] 在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述异源六核苷酸重复扩增序列包含至少约300个重复的或至少600个重复的如SEQ ID NO:1所示的六核苷酸序列。在一些此类非人类动物、非人类动物细胞或非人类动物基因组中,异源六核苷酸重复扩增序列包含至少约300个重复、至少约500个重复,或至少600个重复的SEQ ID NO:1所示的六核苷酸序列。在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述重复在所述异源重复扩增序列中是连续的。在一些此类非人类动物、非人类动物细胞,

或非人类动物基因组中,所述异源六核苷酸重复扩增序列位于所述内源C9orf72基因座的第一个非编码内源外显子和外显子2之间。

[0026] 在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述内源C9orf72基因座包含人C9ORF72核苷酸序列。可选地,所述人C9ORF72核苷酸序列包含SEQ ID NO:46和/或SEQ ID NO:47。

[0027] 在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述非人类动物是啮齿动物或所述非人类动物细胞是啮齿动物细胞。可选地,所述啮齿动物是大鼠或小鼠,或者其中所述非人类动物细胞是大鼠细胞或小鼠细胞。

[0028] 在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述非人类动物、非人类动物细胞,或非人类动物基因组对于所述异源六核苷酸重复扩增序列而言是纯合的。在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述非人类动物、非人类动物细胞,或非人类动物基因组对于所述异源六核苷酸重复扩增序列而言是杂合的。

[0029] 在一些此类非人类动物或细胞中,所述非人类动物或非人类动物细胞表现出:(a)与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,保留所述内含子序列的C9orf72转录物的增加的表达;和/或(b)与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,增加的RNA病灶数量;和/或(c)与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比,增加的二肽重复蛋白水平。

[0030] 在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述非人类动物细胞是胚胎干细胞、胚胎干细胞衍生的运动神经元、脑细胞、皮层细胞、神经元细胞、肌肉细胞、心脏细胞,或生殖细胞。在一些这样的非人类动物、非人类动物细胞,或非人类动物基因组中,所述非人类动物细胞是单细胞期胚胎。可选地,所述非人类动物细胞或基因组是体外的。可选地,所述非人类动物细胞是体内的。

[0031] 在一些此类非人类动物、非人类动物细胞,或非人类动物基因组中,所述非人类动物在其种系基因组中包含在所述内源C9orf72基因座处插入的所述异源六核苷酸重复扩增序列。

[0032] 在另一方面,提供了在生产修饰的细胞的方法中使用的核酸酶试剂,所述修饰细胞在靶基因组基因座处具有增加的拷贝数的重复序列,所述重复序列包含重复扩增序列,其包含复数个拷贝的所述靶基因组基因座处的所述重复序列,其中所述核酸酶试剂被设计成在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂。在另一方面,提供了编码核酸酶试剂的核酸,所述核酸酶试剂用于在靶基因组基因座处产生具有增加的拷贝数的重复序列的修饰的细胞的方法,所述重复序列包含重复扩增序列,其包含复数个拷贝的所述靶基因组基因座处的所述重复序列,其中所述核酸酶试剂被设计成在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂。

[0033] 在另一方面,提供了评估用于治疗与C9orf72基因座处的六核苷酸重复扩增序列相关的疾病或病症的治疗候选物的方法。一些这样的方法包括(a)向任何上述非人类动物或非人类动物细胞施用候选药剂;(b)进行一个或多个测定,以确定所述候选药剂是否对与所述疾病或病症相关的一个或多个体征或症状有影响;以及(c)识别对与所述疾病或病症

相关的所述一个或多个体征或症状有影响的所述候选药剂作为治疗候选物。

[0034] 在一些这样的方法中,将所述候选药剂体内施用给所述非人类动物,并且在施用所述候选药剂之后,在从所述非人类动物分离的细胞中体外进行所述一个或多个测定。在一些此类方法中,将所述候选药剂体外施用于非人胚胎干细胞衍生的运动神经元。

[0035] 在一些此类方法中,所述一个或多个测定包括定量聚合酶链反应(qPCR)以检测含有内含子的C9orf72 RNA转录物。在一些此类方法中,所述一个或多个测定包括测量包含C9orf72有义或反义RNA转录物的RNA病灶,可选地,其中通过荧光原位杂交测量所述RNA病灶。在一些此类方法中,所述一个或多个测定包含测量二肽重复蛋白的积累,可选地,其中所述二肽重复蛋白是聚GA二肽重复蛋白或聚GP二肽重复蛋白,并且可选地,其中通过免疫组织化学测量所述二肽重复蛋白的积累。

附图说明

[0036] 图1(未按比例)示出了野生型小鼠C9orf72基因座、野生型人类C9orf72基因座,和人类化小鼠C9orf72基因座的示意图,所述人类化小鼠C9orf72基因座包含SEQ ID NO:1所示的六核苷酸序列(GGGGCC)的92个重复(MAID8029a)。小鼠序列用虚线和虚线框表示,人类序列用实心黑色表示。外显子由方框表示。示出了六核苷酸重复序列的位置。

[0037] 图2(未按比例)示出了使用核酸酶试剂在重复序列附近引入双链断裂,以使用重组机制在靶基因组基因座处扩增所述重复序列(例如C9orf72基因座处的如SEQ ID NO:1所示的六核苷酸序列)的概念的示意图。

[0038] 图3(未按比例)示出了在C9orf72人源化、含有92x重复的ES细胞(MAID8029a)中的六核苷酸重复扩增序列末端附近的八个向导RNA靶序列的位置的示意图。

[0039] 图4示出了在无细胞系统中使用质粒(8028Stvec)评估的八个向导RNA的切割效率,该质粒包含与人源化、含有92x重复序列的等位基因(MAID8029a)相同的序列。

[0040] 图5(未按比例)示出了通过在30x重复扩增序列的5'末端附近或3'末端附近引入双链断裂(DSB),以在小鼠胚胎干(ES)细胞中扩增30x重复的方案示意图。人源化区域由标记和白框表示。示意了5' DSB和3' DSB的位置,以及用于PCR检测的引物的位置。示意了5' DSB和重复序列的5'边缘之间的距离以及3' DSB和重复序列的3'边缘之间的距离。

[0041] 图6A示出了小鼠ES细胞中C9orf72基因座的常规PCR结果,以评估在30x重复扩增序列的5'末端附近或3'末端附近切割后的重复区域的大小。使用箭头标记亲本克隆和进一步分析的扩增克隆。

[0042] 图6B(未按比例)示出了将亲本30x重复克隆与来自图6A的扩增克隆进行比较的示意图。

[0043] 图6C示出了亲本30x重复小鼠ES细胞克隆和来自图6A的扩增克隆的序列的比对,其从重复区的上游开始并一直持续到重复区的末端。

[0044] 图7A示出了C9orf72基因座的常规PCR的结果,以评估在小鼠ES细胞中30x重复扩增序列的5'末端附近或3'末端附近切割后的重复区域的大小。使用箭头标记亲本克隆和进一步分析的部分收缩的克隆(partially contracted clone)。

[0045] 图7B(未按比例)示出了将亲本30x重复克隆与来自图7A的部分收缩的克隆进行比较的示意图。

[0046] 图8A示出了C9orf72基因座的常规PCR结果,以评估在小鼠ES细胞中30x重复扩增序列的5'末端附近或3'末端附近切割后的重复区域的大小。使用箭头标记亲本克隆和进一步分析的保留30x重复序列的克隆。

[0047] 图8B(未按比例)示出了将亲本30x重复克隆与来自图8A的保留30x重复序列的克隆进行比较的示意图。

[0048] 图9A示出了C9orf72基因座的常规PCR的结果,以评估在小鼠ES细胞中30x重复扩增序列的5'末端附近或3'末端附近切割后的重复区域的大小。使用箭头标记亲本克隆和进一步分析的两个收缩的克隆。

[0049] 图9B(未按比例)示出了将亲本30x重复克隆与来自图9A的收缩克隆进行比较的示意图。

[0050] 图10(未按比例)示出了通过在小鼠ES细胞中的92x重复扩增序列的5'末端附近(方框2)、3'末端附近(方框3)或5'和3'末端附近(方框1)引入双链断裂,以来扩增92x重复的方案示意图。黑框表示内源小鼠序列;白框表示人源化区域。

[0051] 图11A(未按比例)示出了用于评估内源C9orf72 ES细胞克隆中如SEQ ID NO:1所示的六核苷酸序列的数量的常规双引物PCR的示意图。

[0052] 图11B(未按比例)示出了使用三个引物评估内源C9orf72 ES细胞克隆中如SEQ ID NO:1所示的六核苷酸序列的数量的prime PCR的示意图。

[0053] 图12A示出了C9orf72基因座的常规PCR结果,以评估在小鼠ES细胞中92x重复扩增序列的5'末端附近切割后的重复区域的大小。用星号标记扩增的重复。

[0054] 图12B示出了C9orf72基因座的常规PCR结果,以评估在小鼠ES细胞中92x重复扩增序列的3'末端附近切割后的重复区域的大小。用星号标记扩增的重复。

[0055] 图13显示了prime PCR的结果,以确认来自图12A的克隆之一和来自图12B的克隆之一中的重复的数目。亲本8029a(92x重复)克隆用作对照。该图示出了毛细管电泳的结果。信号强度在Y轴上,PCR产物大小在X轴上。读数是峰值的数量。

[0056] 图14示出了prime PCR的结果,以确认来自图12A的克隆之一和来自图12B的克隆之一中的重复的数目。

[0057] 图15示出了C9orf72基因座的常规PCR的结果,以评估在小鼠ES细胞中的92x重复扩增序列的5'末端附近切割后、在92x重复扩增序列的3'末端附近切割后,或92x重复扩增序列的5'和3'末端附近切割后的重复区域的大小。用箭头标记扩增的重复。

[0058] 图16A(未按比例)示出了通过首先在92x重复扩增序列的5'末端附近引入双链断裂,以产生第一个扩增克隆,然后在第一个扩增克隆中的重复扩增序列的3'末端附近引入双链断裂,以产生第二个扩增克隆,从而扩增小鼠ES细胞中的92x重复的方案示意图。人源化区域由标记和白框表示。示意了5' DSB和3' DSB的位置,以及用于PCR检测的引物的位置。

[0059] 图16B示出了C9orf72基因座的常规PCR的结果,以评估在小鼠ES中的250x重复扩增序列的3'末端附近切割250x重复扩增序列克隆后的第二次扩增后的重复区域的大小。用带箭头标记的指示扩增重复和亲本重复。

[0060] 图17A(未按比例)示出了通过在小鼠ES细胞中的92x重复扩增序列的3'末端附近引入单链断裂来扩增92x重复的方案示意图。人源化区域由标记和白框表示。示意了3'

DSB的位置,以及用于PCR检测或测序的引物的位置。

[0061] 图17B示出了C9orf72基因座的常规PCR的结果,以评估在小鼠ES细胞中的92x重复扩增序列的3'末端附近切割或切口(nicking)92x重复扩增序列克隆之后的扩增后的重复区域的大小。

[0062] 图17C示出了在92x重复扩增序列的3'末端附近切割或切口92x重复扩增序列克隆之后,与几个扩增的小鼠ES细胞克隆的测序结果的比对。

[0063] 图18A(未按比例)示出了通过在92x重复扩增序列的5'末端附近引入双链断裂(DSB),在小鼠单细胞期胚胎中扩增92x重复的方案示意图。人源化区域由标记和白框表示。示意了5' DSB的位置,以及用于PCR检测的引物的位置。

[0064] 图18B示出了C9orf72基因座的常规PCR的结果,以评估在小鼠单细胞期胚胎中,在92x重复扩增序列的5'末端附近切割92x重复扩增序列后,单细胞期胚胎的扩增和小鼠的产生之后的重复区域的大小。

[0065] 图19示出了靶基因基因座的常规PCR的结果,以评估在60x重复扩增序列的5'末端附近切割60x重复扩增序列后,在小鼠ES细胞中扩增后的三核苷酸重复区域的大小。

[0066] 图20A-20D示出柱状图,该柱状图示出来自C9orf72基因座的转录物(y轴)的表达水平(由如每幅图顶部的对C9orf72基因座的描述所示的TAQMAN®定量逆转录偶联PCR(RT-qPCR)测定确定),其中该转录物为胚胎干细胞衍生的运动神经元(ESMN)中的外显子1A-外显子2剪接转录物(图20A)、外显子1B-外显子2剪接转录物(图20B)、包含外显子1A附近的内含子序列(图20C),以及外显子1B附近的保留的内含子序列(图20D);相对于包含3个重复的SEQ ID NO:1所示的六核苷酸序列的ESMN,该胚胎干细胞衍生的运动神经元(ESMN)对于包含3、92、250或300个重复的SEQ ID NO:1所示的六核苷酸序列的修饰的C9orf72基因座而言是杂合的。

[0067] 图20E-20H示出柱状图,该柱状图示出来自C9orf72基因座的转录物(y轴)的表达水平(由如每幅图顶部的对C9orf72基因座的描述所示的TAQMAN®定量逆转录偶联PCR(RT-qPCR)测定确定),其中该转录物为胚胎干细胞衍生的运动神经元(ESMN)中的外显子1A-外显子2剪接转录物(图20E)、外显子1B-外显子2剪接转录物(图20F)、包含外显子1A附近的内含子序列(图20G),以及外显子1B附近的保留的内含子序列(图20H);相对于包含3个重复的SEQ ID NO:1所示的六核苷酸序列的ESMN,该胚胎干细胞衍生的运动神经元(ESMN)对于包含3、500或600个重复的SEQ ID NO:1所示的六核苷酸序列的修饰的C9orf72基因座而言是杂合的。

[0068] 图21(顶部)示出了来自胚胎干细胞衍生的运动神经元(ESMN)的裂解物的蛋白质狭缝印迹(western slot blot)图像,所述ESMN对于包含3、92、300、500或600个重复的SEQ ID NO:1所示的六核苷酸序列的修饰的C9orf72基因座而言是杂合的。用抗聚GlyPro抗体或抗聚GlyAla抗体印迹含有0 μ g、1.25 μ g、2.5 μ g、5 μ g、10 μ g,或20 μ g总蛋白的裂解物。图21(底部)示出了该图顶部的蛋白质狭缝印迹的量化。

[0069] 图22A-22B示出柱状图,该柱状图示出来自C9orf72基因座的转录物(y轴)的表达水平(由如每幅图顶部的对C9orf72基因座的描述所示的TAQMAN®定量逆转录偶联PCR(RT-qPCR)测定确定),其中该转录物包括胚胎干细胞衍生的运动神经元(ESMN)中的外显子

1A附近的内含子序列;相对于包含3个重复的SEQ ID NO:1所示的六核苷酸序列的ESMN,所述ESMN对于包含3、92、300、500或600个重复的SEQ ID NO:1所示的六核苷酸序列的修饰的C9orf72基因座而言是杂合的(图22A)。其中该转录物包括小鼠脑干和脊髓样本中的外显子1A附近的内含子序列;相对于包含3个重复的SEQ ID NO:1所示的六核苷酸序列的小鼠的脑干和脊髓样本,所述小鼠对于包含3或300个重复的SEQ ID NO:1所示的六核苷酸序列的修饰的(人源化的)C9orf72基因座而言是杂合的(图22B)。

[0070] 图23A-23D示出柱状图,该柱状图示出来自C9orf72基因座的转录物(y轴)的表达水平(由如每幅图顶部的对C9orf72基因座的描述所示的TAQMAN®定量逆转录偶联PCR(RT-qPCR)测定确定),其中该转录物为胚胎干细胞衍生的运动神经元(ESMN)中的外显子1A-外显子2剪接转录物(图23A)、外显子1B-外显子2剪接转录物(图23B)、包含外显子1A附近的内含子序列(图23C),以及外显子1B附近的保留的内含子序列(图23D);该胚胎干细胞衍生的运动神经元(ESMN)为近轴样运动神经元或肢体样运动神经元,且相对于包含3个重复的SEQ ID NO:1所示的六核苷酸序列的ESMN,该ESMN对于包含3、500或600个重复的SEQ ID NO:1所示的六核苷酸序列的修饰的C9orf72基因座而言是杂合的。

[0071] 图23E-23F示出柱状图,该柱状图示出来自C9orf72基因座的转录物(y轴)的表达水平(由如每幅图顶部的对C9orf72基因座的描述所示的TAQMAN®定量逆转录偶联PCR(RT-qPCR)测定确定),其中该转录物为胚胎干细胞衍生的运动神经元(ESMN)中的未剪接前体转录物(图23E)或剪接的C9orf72 RNA(图23F);该胚胎干细胞衍生的运动神经元(ESMN)为近轴样运动神经元或肢体样运动神经元,且相对于包含3个重复的SEQ ID NO:1所示的六核苷酸序列的ESMN,该ESMN对于包含3、500或600个重复的SEQ ID NO:1所示的六核苷酸序列的修饰的C9orf72基因座而言是杂合的。

具体实施方式

定义

[0072] 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。

[0073] 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)。

[0074] 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。

[0075] 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'。

[0076] 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。

[0077] 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。

[0078] 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。

[0079] 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,但不限于其它细胞蛋白质、核酸,或细胞的或细胞外的成分)。

[0080] 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。

[0081] 术语“内源序列”是指天然存在于细胞或非人类动物体内的核酸序列。例如,非人类动物的内源C9orf72序列是指天然存在于非人类动物中的C9orf72基因座处的天然C9orf72序列。

[0082] “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。

[0083] 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。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。

[0084] “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)与天然核酸序列相比具有更高使用频率的密码子替换。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)。

[0085] 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,“C9orf72基因座”可以指C9orf72基因、C9orf72 DNA序列、C9orf72编码序列,或C9orf72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。“C9orf72基因座”可包含C9orf72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。

[0086] 术语“基因”是指染色体中的DNA序列,其可能包含(如果天然存在)至少一个编码区和至少一个非编码区。染色体中编码产物(例如但不限于RNA产物和/或多肽产物)的DNA序列可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)。此外,其它非编码序列,包括调节序列(例如但不限于启动子、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),和基质附着区可能位于基因中。这些序列可以靠近基因的编码区(例如但不限于在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。

[0087] 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的,如果两个等位基因不同,则基因型被描述为杂合的。

[0088] “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(operably linked)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W0 2013/176772中找到,其全部内容通过引用并入本文。

[0089] “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组件上的功能。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)。

[0090] 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,和向导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和能力。这种生物学活性或功能可以包括,例如,Cas蛋白与向导RNA和靶DNA序列结合的能力。与原始分子相比,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),但是保留了该分子的基本生物学功能。

[0091] 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)。

[0092] 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(即去除蛋白质的中间部分)。

[0093] 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)。

[0094] “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(其不含添加或删除)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。

[0095] 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。

[0096] 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。表1总结了典型的氨基酸分类。

[0097] 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

[0098] “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。

[0099] 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。

[0100] 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。“报道蛋白”是指由报道基因编码的蛋白。

[0101] 本文所使用的术语“荧光报道蛋白”是指基于荧光可检测的报道蛋白, 其中荧光可以直接来自报道蛋白、报道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。

[0102] 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) 和非同源末端连接 (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。

[0103] 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; Mandaloset 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。

[0104] 非同源性末端接合 (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定整合。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介导的靶定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面的大范围的序列同一性的知识,这在尝试靶定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试剂产生的那些相容。参见,例如,US 2011/020722、WO 2014/033644、WO 2014/089290和Marescaet al. (2013) *Genome Res.* 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。

[0105] “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”。

[0106] “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。

[0107] 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。

[0108] 除非从上下文中另外显而易见,否则术语“约”涵盖所述值的 ± 5 的值。

[0109] 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。

[0110] 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。

[0111] 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。

[0112] 具有统计学意义的均值意指 $p \leq 0.05$ 。

详细说明

I. 概述

[0113] 本文公开了核酸酶介导的用于扩增已经存在于基因组基因座处的重复的方法。重复扩增疾病往往在重复长度和疾病严重程度之间显示出惊人的基因型-表型相关性。重复越长,疾病越严重,症状出现的时间也越早。参见,例如,Paulson (2018) *Handb. Clin. Neurol.*, 147:105-123, (2018), 出于所有目的通过引用整体并入本文。然而,与重复扩增疾病相关的重复序列中的高GC含量使得难以合成具有大量拷贝数的重复的DNA片段,并且难以在微生物中维持这样的重复。因此,准备材料以产生具有大量拷贝数的重复的靶向载体以及最终产生包含多个拷贝数的重复的转基因动物是具有挑战性的。这是为什

么很少存在针对重复扩增疾病的有用的动物模型的主要原因之一。本文公开的方法通过扩增已经插入基因组中正确位置的相对较短的重复(即较少数量的重复)而非从头进行(de novo)靶向载体生产和ES细胞靶向来克服这个问题。

[0114] 本文还公开了包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物基因组、非人类动物细胞和非人类动物,以及通过核酸酶介导的重复扩增制备此类非人类动物细胞和非人类动物的方法。还提供了使用非人类动物细胞或非人类动物识别可用于预防、延迟或治疗一种或多种与C9orf72基因座重复扩增相关的神经退行性疾病的治疗候选物的方法。

II. 核酸酶介导的重复扩增方法

[0115] 提供了扩增重复扩增序列的各种方法,所述重复扩增序列包括在细胞中的复数个拷贝数的重复序列,所述细胞在靶基因组基因座处包含所述重复扩增序列。提供了用于获得具有扩增的重复扩增序列的细胞的各种方法。还提供了通过此类方法产生的细胞。

[0116] 用于扩增重复扩增序列的此类方法可包括将核酸酶试剂引入包含重复扩增序列的细胞或细胞群中,所述核酸酶试剂切割所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生具有扩增的重复扩增序列的修饰的细胞或修饰的细胞群。用于扩增重复扩增序列的这些方法可包括将核酸酶试剂或编码所述核酸酶试剂的核酸引入包含重复扩增序列的细胞或细胞群中,其中述核酸酶试剂切割所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生具有扩增的重复扩增序列的修饰的细胞或修饰的细胞群。用于扩增重复扩增序列的这些方法可包括将核酸酶试剂引入包含重复扩增序列的细胞或细胞群中,所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生具有扩增的重复扩增序列的修饰的细胞或修饰的细胞群。用于扩增重复扩增序列的这些方法可包括将核酸酶试剂或编码所述核酸酶试剂的核酸引入包含重复扩增序列的细胞或细胞群中,其中所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生具有扩增的重复扩增序列的修饰的细胞或修饰的细胞群。重复扩增序列是特定基因座处重复序列的所有拷贝(例如连续重复)。重复扩增序列的5'末端是重复扩增序列中第一个重复的第一个核苷酸(即5'核苷酸)。重复扩增序列的3'末端是重复扩增序列中最后一个重复的最后一个核苷酸(即3'核苷酸)。重复扩增序列的5'末端附近的核酸酶靶位点可以位于重复扩增序列的5'末端的上游、重复扩增序列的5'末端的下游,或与重复扩增序列的5'末端重叠。重复扩增序列的3'末端附近的核酸酶靶位点可以位于重复扩增序列的3'末端的上游、重复扩增序列的3'末端的下游,或与重复扩增序列的3'末端重叠。在一些方法中,核酸酶靶位点在重复扩增序列之外。在一些方法中,核酸酶靶位点与重复扩增序列的5'末端或3'末端重叠。在一些方法中,核酸酶靶位点位于重复扩增序列内。

[0117] 此类方法可进一步包括量化修饰的细胞中重复序列的拷贝数或量化修饰的细胞群中重复序列的拷贝数并选择修饰的细胞,其中所述重复序列的拷贝数已增加。在一些方法中,重复扩增序列的平均扩增百分比(即添加到重复扩增序列的重复的序列的长度除以起始大小 $\times 100$)为至少约1%、至少约2%、至少约3%、至少约4%、至少约5%、至少约6%、至少约7%、至少约8%、至少约9%、至少约10%、至少约15%、至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%,或更高。

[0118] 一些这样的方法包括：(a) 提供包含重复扩增序列的细胞群，所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列；(b) 将核酸酶试剂引入所述细胞群，所述核酸酶试剂切割重复扩增序列的5'末端或3'末端附近的核酸酶靶位点，以产生修饰的细胞群；以及(c) 量化所述修饰的细胞群中的重复序列的拷贝数，并选择其中所述重复序列的拷贝数增加的修饰的细胞。一些这样的方法包括：(a) 提供包含重复扩增序列的细胞群，所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列；(b) 将核酸酶试剂或编码所述核酸酶试剂的核酸引入所述细胞群，其中所述核酸酶试剂切割重复扩增序列的5'末端或3'末端附近的核酸酶靶位点，以产生修饰的细胞群；以及(c) 量化所述修饰的细胞群中的重复序列的拷贝数，并选择其中所述重复序列的拷贝数增加的修饰的细胞。一些这样的方法包括：(a) 提供包含重复扩增序列的细胞群，所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列；(b) 将核酸酶试剂引入所述细胞群，其中所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂，以产生修饰的细胞群；以及(c) 量化所述修饰的细胞群中的重复序列的拷贝数，并选择其中所述重复序列的拷贝数增加的修饰的细胞。在筛选细胞群的一些方法中，扩增频率(具有重复扩增的细胞克隆占筛选克隆总数的百分比)为至少约1%、至少约2%、至少约3%、至少约4%、至少约5%、至少约6%、至少约7%、至少约8%、至少约9%、至少约10%，或更高。

[0119] 虽然实践中不需要对机制的理解，但据信在重复扩增序列附近切割靶基因组基因座，使得暴露的5'链仅包含重复序列，会导致Rad51丝进入错误位置，从而误导Rad51丝的同源性搜索，导致重复序列的扩增或收缩。参见图2。因此，在某些方法中，受损的染色单体利用未受损的姐妹染色单体作为其修复的模板。同样，在一些方法中，不将外源修复模板(外源供体序列)引入细胞或细胞群中。

[0120] 核酸酶试剂在重复扩增序列附近切割(例如造成双链断裂)。例如，核酸酶靶位点可位于所述重复扩增序列的5'末端或3'末端的约1000、约900、约800、约700、约600、约500、约400、约300、约200、约100、约90、约80、约70、约60、约50、约40、约30、约20或约10个核苷酸内。作为另一个例子，核酸酶靶位点可位于所述重复扩增序列的5'末端或3'末端的约100、约50、约40、约30、约20或约10个核苷酸内。例如，核酸酶靶位点可位于所述重复扩增序列的5'末端或3'末端的约100、约90、约80、约70、约60、约50、约40、约30、约20或约10个核苷酸内。作为另一个例子，核酸酶靶位点可位于所述重复扩增序列的5'末端或3'末端的约25、约24、约23、约22、约21、约20、约19、约18、约17、约16、约15、约14、约13、约12、约11、约10、约9、约8、约7、约6、约5、约4、约3、约2或约1个核苷酸内。替代地，核酸酶靶位点可以与重复扩增序列的5'末端或3'末端重叠。类似地，如果核酸酶是Cas蛋白，则原始间隔区相邻模体(PAM)序列可位于所述重复扩增序列的5'末端或3'末端的约1000、约900、约800、约700、约600、约500、约400、约300、约200、约100、约90、约80、约70、约60、约50、约40、约30、约20或约10个核苷酸内。作为另一个例子，PAM可位于所述重复扩增序列的5'末端或3'末端的约100、约50、

约40、约30、约20或约10个核苷酸内。例如，PAM可位于所述重复扩增序列的5'末端或3'末端的约100、约90、约80、约70、约60、约50、约40、约30、约20或约10个核苷酸内。作为另一个例子，PAM可位于所述重复扩增序列的5'末端或3'末端的约25、约24、约23、约22、约21、约20、约19、约18、约17、约16、约15、约14、约13、约12、约11、约10、约9、约8、约7、约6、约5、约4、约3、约2或约1个核苷酸内。

[0121] 核酸酶试剂造成双链断裂或单链断裂的位点靠近重复扩增序列。例如，核酸酶试剂造成双链断裂或单链断裂的位点可位于所述重复扩增序列的5'末端或3'末端的约1000、约900、约800、约700、约600、约500、约400、约300、约200、约100、约90、约80、约70、约60、约50、约40、约30、约20或约10个核苷酸内。作为另一个例子，核酸酶试剂造成双链断裂或单链断裂的位点可位于所述重复扩增序列的5'末端或3'末端的约100、约50、约40、约30、约20或约10个核苷酸内。作为一个例子，核酸酶试剂造成双链断裂或单链断裂的位点可位于所述重复扩增序列的5'末端或3'末端的约50、约40、约30、约20或约10个核苷酸内。作为另一个例子，核酸酶试剂造成双链断裂或单链断裂的位点可位于所述重复扩增序列的5'末端或3'末端的约25、约24、约23、约22、约21、约20、约19、约18、约17、约16、约15、约14、约13、约12、约11、约10、约9、约8、约7、约6、约5、约4、约3、约2或约1个核苷酸内(例如所述重复扩增序列的5'末端或3'末端的约20个核苷酸内、约16个核苷酸内、约15个核苷酸内、约11个核苷酸内、约10个核苷酸内、约7个核苷酸内、约5个核苷酸内或约2个核苷酸内)。在一个实施例中，重复扩增序列是六核苷酸重复扩增序列(例如在C90RF72基因座处)，并且核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的5'末端或3'末端的约10个核苷酸内、约7个核苷酸内、约5个核苷酸内或约2个核苷酸内(例如重复扩增序列的5'末端或3'末端的约7个核苷酸内、重复扩增序列的3'末端的约7个核苷酸内，或重复扩增序列的5'末端的约2个核苷酸内)。在另一个例子中，重复扩增序列是三核苷酸重复扩增序列，并且核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的5'末端或3'末端的约20个核苷酸内、约16个核苷酸内、约15个核苷酸内、约11个核苷酸内，或约10个核苷酸内(例如重复扩增序列的5'末端或3'末端的约16个核苷酸内、重复扩增序列的5'末端的约16个核苷酸内，或重复扩增序列的3'末端的约11个核苷酸内)。在一些方法中，核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的5'末端附近。在一些方法中，核酸酶试剂造成双链断裂或单链断裂的位点位于所述重复扩增序列的3'末端附近。在一些方法中，核酸酶试剂造成双链断裂或单链断裂的位点位于重复扩增序列内部。在一些方法中，核酸酶试剂造成双链断裂或单链断裂的位点位于重复扩增序列之外。在细胞修复了双链断裂或单链断裂后，可以破坏核酸酶试剂造成双链断裂或单链断裂的位点或保留核酸酶靶位点。然后可以重复使用核酸酶靶位点，进行后续轮的重复扩增。在一些方法中，双链断裂或单链断裂的修复不会导致核酸酶靶位点内的插入或删除。例如，在一些方法中，双链断裂或单链断裂的修复不会导致重复扩增序列外的插入或删除。

[0122] 可以使用任何合适的核酸酶试剂。例如，核酸酶试剂为锌指核酸酶(ZFN)、类转录激活因子效应物核酸酶(TALEN)、或成簇的规律间隔的短回文重复序列(CRISPR)-相关(Cas)蛋白和向导RNA(例如Cas9蛋白和向导RNA)。在本文其它地方更详细地公开了每个的示例和描述。

[0123] 在一些方法中，可以将两个核酸酶试剂引入细胞中。在一些方法中，可以将两个核

酸酶试剂或编码所述两个核酸酶试剂的核酸引入细胞。例如,第一核酸酶试剂可在重复扩增序列的5'末端附近切割第一核酸酶靶位点,并且第二核酸酶试剂可以在重复扩增序列3'末端附近切割第二核酸酶靶位点。在一个实施例中,第一核酸酶试剂在重复扩增序列的5'末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂,并且第二核酸酶试剂在重复扩增序列3'末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂。替代地,第一核酸酶试剂在重复扩增序列5'末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂,并且第二核酸酶试剂在重复扩增序列的5'末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂。替代地,第一核酸酶试剂在重复扩增序列的3'末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂,并且第二核酸酶试剂在重复扩增序列的3'末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂。在另一个实施例中,第一核酸酶试剂在重复扩增序列的5'末端附近的第一核酸酶靶位点处造成双链断裂,并且第二核酸酶试剂在重复扩增序列的3'末端附近的第二核酸酶靶位点处造成双链断裂。在另一个实施例中,第一核酸酶试剂在重复扩增序列的5'末端附近的第一核酸酶靶位点处造成单链断裂,并且第二核酸酶试剂在重复扩增序列的3'末端附近的第二核酸酶靶位点处造成单链断裂。在另一个实施例中,第一核酸酶试剂在重复扩增序列的5'末端附近的第一核酸酶靶位点处造成单链断裂,并且第二核酸酶试剂在重复扩增序列的3'末端附近的第二核酸酶靶位点处造成双链断裂。在另一个实施例中,第一核酸酶试剂在重复扩增序列的5'末端附近的第一核酸酶靶位点处造成双链断裂,并且第二核酸酶试剂在重复扩增序列的3'末端附近的第二核酸酶靶位点处造成单链断裂。第一核酸酶靶位点和第二核酸酶靶位点可以各自任意组合地位于重复扩增序列的5'末端或3'末端,且第一核酸酶试剂和第二核酸酶试剂可以各自任意组合地形成双链断裂或单链断裂。

[0124] 可以通过任何已知的方式将核酸酶试剂引入细胞。可以通过任何已知的方式将编码核酸酶试剂的核酸引入细胞。引入包括向细胞呈递分子(例如核酸或蛋白质),使其进入细胞内部。如果引入了多个组分,则它们可以同时或以任何组合顺序引入。另外,可以通过相同的递送方法或不同的递送方法将两种或更多种组分引入细胞中。类似地,如果细胞在体内(例如在非人类动物中),则可以通过相同的给药途径或不同的给药途径将两种或更多种成分引入非人类动物。

[0125] 引入细胞的分子可被提供于包含载体的组合物中,该载体增加了被引入的分子稳定性(例如延长降解产物的浓度在给定的储存条件下(例如-20°C、4°C,或环境温度)保持在阈值以下(例如起始核酸或蛋白质重量的0.5%以下)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(lipid cochleates),和脂质微管。

[0126] 本文提供了各种方法和组合物,以允许将核酸酶试剂引入细胞中。本文提供了各种方法和组合物,以允许将编码核酸酶试剂的核酸引入细胞中。将核酸引入各种细胞类型的方法是已知的,并包括,例如,稳定转染方法、瞬时转染方法,和病毒介导的方法。

[0127] 用于将核酸序列引入细胞的转染方案以及方案是多样的。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, Bacchettiet 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。

[0128] 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, 或通过核转染将核酸酶试剂引入细胞。核转染是一种改进的电穿孔技术, 其使核酸底物不仅可以递送到细胞质, 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)。在一个实施例中, 使用 LONZA[®] NUCLEOFECTOR[™] 系统进行核转染。

[0129] 也可以通过显微注射来将核酸酶试剂引入细胞 (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。mRNA 优选显微注射入细胞质 (例如将 mRNA 直接递送至翻译系统), 而蛋白或编码蛋白质或编码 RNA 的多核苷酸优选显微注射入细胞核/原核。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。如果将蛋白质注射到细胞质中并且需要靶向细胞核, 则它可以包含核定位信号以确保递送到细胞核/原核。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。

[0130] 将核酸酶试剂引入细胞的其它方法可包括, 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。将编码核酸酶试剂的核酸引入细胞的其它方法可包括, 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类动物。如果细胞在体内 (例如, 在非人类动物中), 递送至非人类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。

[0131] 可以通过体内的流体动力递送 (HDD) 来将核酸酶试剂引入细胞。可以通过体内的流体动力递送 (HDD) 来将编码核酸酶试剂的核酸引入细胞。流体动力递送已作为体内的细胞内 DNA 递送的方法而出现。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和合成载体相关的安全隐患。当注入血液时, DNA 能够到达血液可及的不同组织中的细胞。流体动力递送利用了将大量溶液快速注入循环里不可压缩的血液中所产生的力, 从而克服了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。

[0132] 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)来实现核酸酶试剂的引入。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)来实现编码核酸酶试剂的核酸的引入。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方面有缺陷)。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。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。

[0133] 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IVA)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。

[0134] 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。

[0135] 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合衣壳,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取的示例。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。

[0136] 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。然而,也可以使用单链AAV(ssAAV)载体。

[0137] 为了增加包装能力,可以在两个AAV转移质粒之间分配更长的转基因,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,转基因可以分开于两个转移质粒之间,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。

[0138] 核酸酶试剂的引入也可以通过脂质纳米颗粒(LNP)介导的递送来实现。编码核酸酶试剂的核酸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介导的递送可用于递送Cas mRNA和向导RNA的组合或Cas蛋白和向导RNA的组合。通过此类方法递送可导致瞬时Cas表达,并且可生物降解的脂质可提高清除率、提高耐受性并降低免疫原性。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),该隐形脂质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W02016/010840A1中找到,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)。

[0139] 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)用于稳定的中性脂质;(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):2227-2235和W0 2017/173054A1,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,货物(cargo)可包含向导RNA或编码向导RNA的核酸。在某些LNP中,货物可包含编码Cas核酸酶(例如Cas9)的mRNA以及向导RNA或编码向导RNA的核酸。

[0140] 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):2227-2235和W0 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。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)

基)丁酸酯(也称为Dlin-MC3-DMA(MC3))。

[0141] 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,至少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。

[0142] 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,脂质可以单独地具有约5.8至约6.2的pKa。

[0143] 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。例如,中性磷脂可以选自由二硬脂酰磷脂酰胆碱(DSPC)和二豆蔻酰基磷脂酰乙醇胺(DMPE)组成的组。

[0144] 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。

[0145] 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。

[0146] 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。

[0147] 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。

[0148] 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。

[0149] 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%。

[0150] 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。N/P比也可以是约4至约7或约4.5至约6。在具体实施例中,N/P之比可为4.5或可为6。

[0151] 在一些LNP中,货物可以包含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。替代地,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。替代地,LNP制剂可以包括约1:1至约1:2的Cas mRNA与gRNA核酸之比。Cas mRNA与gRNA核酸之比可为约1:1或约1:2。

[0152] 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,Finns et al. (2018) Cell Rep. 22(9):2227-2235,出于所有目的,通过引用将其全部内容整体并入本文。Cas9 mRNA与向导RNA的重量比可以为1:1。合适的LNP的另一个具体例子包含摩尔比为50:38.5:10;1.5的Dlin-MC3-DMA(MC3)、胆固醇、DSPC和PEG-DMG。

[0153] 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。Cas9mRNA与向导RNA的重量比可以为1:2。

[0154] 可以选择递送方式以降低免疫原性。例如,当递送多个组分时,可以通过不同的模式来递送这些组分(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mRNA或蛋白质)可以确保Cas/gRNA复合物仅在短时间内存在并起作用,并且可以降低由来自细菌衍生的Cas酶的肽引起的免疫原性被MHC分子展示在细胞表面。这种瞬时递送也可以减少脱靶修饰的可能性。

[0155] 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。

[0156] 重复扩增序列可以是异源重复扩增序列。当在核酸的上下文中使用时,术语“异源的”表示该核酸包含在同一分子中并非天然同时出现的至少两个片段。例如,当涉及核酸的片段或使用,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C90RF72序列,该C90RF72序列侧接在自然界中与该C90RF72序列不相关联的序列。替代地,核酸的重复扩增序列可包括侧接内源非人核酸序列的人核酸序列。例如,重复扩增序列可为侧接内源非人(例如小鼠)核酸序列的人核酸序列。在其它方法中,重复扩增序列可以是内源的。例如,细胞可以是人细胞(例如人诱导多能干细胞),并且重复扩增序列和侧接序列都可以包括人核酸序列。

[0157] 重复扩增序列包含多个重复。例如,待扩增的重复扩增序列可包含至少约10、至少约20、至少约30、至少约40、至少约50、至少约60、至少约70、至少约80或至少约100个拷贝的重复序列。在一些重复扩增序列中,重复可以是连续的(彼此相邻而没有插入序列)。

[0158] 由所述方法产生的扩增重复扩增序列可以具有任意数量的重复。例如,扩增的重复扩增序列可以包含超过约95个重复、超过约96个重复、超过约97个重复、超过约98个重复、超过约99个重复、超过约100个重复、超过约101个重复、超过约102个重复、超过约103个重复、超过约104个重复、超过约105个重复、超过约150个重复、超过约200个重复、超过约250个重复、超过约295个重复、超过约296个重复、超过约297个重复、超过约298个重复、超

超过约299个重复、超过约300个重复、超过约301个重复、超过约302个重复、超过约303个重复、超过约304个重复、超过约305个重复、超过约350个重复、超过约400个重复、超过约450个重复、超过约500个重复、超过约550个重复、超过约595个重复、超过约596个重复、超过约597个重复、超过约598个重复、超过约599个重复、超过约600个重复、超过约601个重复、超过约602个重复、超过约603个重复、超过约604个重复,或超过约605个重复。替代地,所述扩增的重复扩增序列可包含至少约95个重复、至少约96个重复、至少约97个重复、至少约98个重复、至少约99个重复、至少约100个重复、至少约101个重复、至少约102个重复、至少约103个重复、至少约104个重复、至少约105个重复、至少约150个重复、至少约200个重复、至少约250个重复、至少约295个重复、至少约296个重复、至少约297个重复、至少约298个重复、至少约299个重复、至少约300个重复、至少约301个重复、至少约302个重复、至少约303个重复、至少约304个重复、至少约305个重复、至少约350个重复、至少约400个重复、至少约450个重复、至少约500个重复、至少约550个重复、至少约595个重复、至少约596个重复、至少约597个重复、至少约598个重复、至少约599个重复、至少约600个重复、至少约601个重复、至少约602个重复、至少约603个重复、至少约604个重复,或至少约605个重复。在一个具体实施例中,扩增的重复扩增序列包含超过约100个重复、超过约300个重复、超过约600个重复、至少约100个重复、至少约300个重复,或至少约600个重复。

[0159] 重复序列(例如每个重复)可以包含任何数量的核苷酸。例如,重复序列可以是三核苷酸重复、四核苷酸重复、五核苷酸重复、六核苷酸重复,或十二核苷酸重复。重复序列可以是与疾病如神经系统疾病相关的重复扩增。作为一个实例,重复序列可以包含、基本上由,或由SEQ ID NO:1-12中的任一个组成。例如,重复序列可以是表2中列出的重复序列之一和/或可以位于包含表2中列出的基因之一的靶基因组基因座中和/或可以与表2中列出的疾病之一相关。

[0160] 超过40种疾病,其中大部分主要影响神经系统,是由分散在整个人类基因组中的简单序列重复的扩增引起的。扩增的三核苷酸重复疾病首先被发现,并且仍然是最常见的。最近,四核苷酸、五核苷酸、六核苷酸甚至十二核苷酸重复扩增已被确定为人类疾病(包括一些最常见的神经遗传疾病)的病因。重复扩增疾病包括强直性肌营养不良(DM1和DM2)的病因、肌萎缩侧索硬化/额颞叶痴呆(C9ORF72)的最常见遗传原因、亨廷顿病和其他八种多聚谷氨酰胺疾病,包括最常见的显性遗传性共济失调,最常见的隐性共济失调(弗里德赖希共济失调)和最常见的遗传性智力低下(脆性X综合征)。

[0161] 重复、它们出现的基因和它们相关的疾病的例子列于表2中。

[0162] 表2:导致神经系统疾病的重复扩增。

重复	SEQ ID NO	基因	相关的神经系统疾病
CAG	2	<i>HTT</i> <i>AR</i> <i>ATNI</i> <i>ATXN1</i> <i>ATXN2</i> <i>ATXN3</i> <i>CACNA1A</i> <i>ATXN7</i> <i>PPP2R2B</i> <i>TBP</i>	亨廷顿病 脊髓和延髓肌肉萎缩症 齿状核红核苍白球路易体萎缩症 脊髓小脑性共济失调1型 脊髓小脑性共济失调2型 脊髓小脑性共济失调3型 脊髓小脑性共济失调6型 脊髓小脑性共济失调7型 脊髓小脑性共济失调12型 脊髓小脑性共济失调17型
CGG	3	<i>FMRI (FRAXA)</i>	脆性X、脆性X震颤共济失调综合征
CTG	4	<i>DMPK</i> <i>JPH3</i> <i>ATXN8</i> <i>TCF4</i>	强直性肌营养不良1型 类亨廷顿病综合征2 (Huntington disease-like 2) 脊髓小脑性共济失调8型 富克斯角膜营养不良
GAA	5	<i>FXN (FRDA)</i>	弗里德赖希共济失调
GCC	6	<i>AFF2 (FRAXE)</i>	FRAXE智力低下
GCG	7	<i>PABPN1 (PABP2)</i>	眼咽型肌营养不良
CCTG	8	<i>CNBP</i>	强直性肌营养不良2型
ATTCT	9	<i>ATXN10</i>	脊髓小脑性共济失调10型
TGGAA	10	<i>TK2, BEANI (BEAN)</i>	脊髓小脑性共济失调31型
GGCCTG	11	<i>NOP56</i>	脊髓小脑性共济失调36型
GGGGCC	1	<i>C9ORF72</i>	额颞痴呆/肌萎缩侧索硬化
CCCCGCCCCGCG	12	<i>CSTB</i>	EPM1 (肌阵挛癫痫)

[0163] 重复扩增疾病的一般特征包括：(1) 重复扩增源于正常存在的多态性重复；(2) 扩增是不稳定的(动态的)，在传递给下一代时经常改变大小；(3) 更长的重复往往会导致更严重、更早发病的疾病；(4) 临床预判常见：发病早、后代的疾病更为严重；(5) 高度可变的表型，主要反映重复大小的差异。然而，与重复扩增疾病相关的重复序列中的高GC含量使得难以合成具有大量拷贝数的重复的DNA片段，并且难以在微生物中维持这样的重复。因此，准备材料以产生具有大量拷贝数的重复的靶向载体以及最终产生包含多个拷贝数的重复的转基因动物是具有挑战性的。这是为什么很少存在针对重复扩增疾病的有用的动物模型的主要原因之一。本文公开的方法通过扩增已经插入基因组中正确位置的相对较短的重复(即较少数量的重复)而非从头进行(de novo)靶向载体生产和ES细胞靶向来克服这个问题。

[0164] 在一个具体实例中，重复序列可包含GGGGCC(SEQ ID NO:1)、基本上由其组成或由其组成，并且靶基因组基因座可为C9ORF72基因座。在一个具体实例中，核酸酶靶位点可包含以下、基本由以下组成，或由以下组成：SEQ ID NO:28-35中的任一个，或者可包含以下、基本由以下组成，或由以下组成：SEQ ID NO:28或33。作为一个具体的例子，核酸酶试剂是CRISPR/Cas9(包含一个向导RNA)。向导RNA(例如，向导RNA的DNA靶向区段或向导序列)可包含以下、基本由以下组成，或由以下组成：SEQ ID NOS:84-91中的任一个，或者可包含以下、基本由以下组成，或由以下组成：SEQ ID NO:84或89。类似地，向导RNA(例如向导RNA的crRNA部分)可包含以下、基本由以下组成，或由以下组成：SEQ ID NO:56、57和76-81中的任一个，或者可包含以下、基本由以下组成，或由以下组成：SEQ ID NO:56或79。

[0165] 细胞可以是任何类型的细胞。可选地，细胞可以是人类细胞。例如，细胞可以是人诱导性多能干细胞。替代地，可以使用本文别处更详细公开的任何非人类动物细胞。作为一些具体实例，细胞可以是非人类动物单细胞期胚胎、非人类动物胚胎干细胞、源自胚胎干细

胞的运动神经元、脑细胞、皮质细胞、神经元细胞、肌肉细胞、心脏细胞、或生殖细胞。在一个例子中,细胞可以是啮齿动物细胞,例如小鼠细胞或大鼠细胞。例如,细胞可以是小鼠胚胎干细胞或小鼠单细胞期胚胎。细胞可以是体外的、离体的或体内的。

[0166] 重复序列的拷贝数的量化可以通过任何合适的方式来完成。例如,可以使用Southern印迹法或聚合酶链反应(PCR)基因分型方法。作为一个具体例子,可以使用常规的双引物PCR,或者可以使用利用三个引物的primePCR,如实施例中所述。Prime PCR是Warner等人(Warner et al. (1996) J. Med. Genet. 33 (12): 1022-1026, 出于所有目的通过引用整体并入本文)研发的三重重复引物PCR的改进版本。此方法使用可引发重复内聚合酶链反应的引物计算重复的数量,并与荧光标记的位点特异性引物一起使用。

[0167] 在一些方法中,该方法被重复多次(例如两次或三次)以实现多轮扩增。例如,该方法可以包括多轮的(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂引入所述细胞群,所述核酸酶试剂切割重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。例如,该方法可以包括多轮的(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂或编码所述核酸酶试剂的核酸引入所述细胞群,其中所述核酸酶试剂切割重复扩增序列的5'末端或3'末端附近的核酸酶靶位点,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。例如,该方法可以包括多轮的(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂或编码所述核酸酶试剂的核酸引入所述细胞群,其中所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。例如,该方法可以包括多轮的(a)提供包含重复扩增序列的细胞群,所述重复扩增序列在靶基因组基因座处包含多个拷贝的重复序列;(b)将核酸酶试剂或编码所述核酸酶试剂的核酸引入所述细胞群,其中所述核酸酶试剂在所述重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂,以产生修饰的细胞群;以及(c)量化所述修饰的细胞群中的重复序列的拷贝数,并选择其中所述重复序列的拷贝数增加的修饰的细胞。第一轮之后每一轮的步骤(a)中的细胞群可以从前一轮的步骤(c)中选择的修饰的细胞扩增而得的细胞克隆群。该方法可以重复任何次数(例如,至少约2次、至少约3次、至少约4次、至少约5次、至少约6次、至少约7次、至少约8次、至少约9次或至少约10次),只要在重复扩增序列的5'末端或3'末端附近保留有供核酸酶试剂靶向的核酸酶靶位点。

[0168] 例如,所述方法可在第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂切割所述重复扩增序列的5'末端附近的第一核酸酶靶位点,以产生第一修饰细胞,且所述方法可在第二次进行时使用第二核酸酶试剂,所述第二核酸酶试剂切割所述重复扩增序列的3'末端附近的第二核酸酶靶位点,以产生第二修饰细胞。替代地,所述方法可在第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂切割所述重复扩增序列的3'末端附近的第一核酸酶靶位点,以产生第一修饰细胞,且所述方法可在第二次进行时使用第二核酸酶试剂,

所述第二核酸酶试剂切割所述重复扩增序列的5'末端附近的第二核酸酶靶位点,以产生第二修饰细胞。

[0169] 作为另一个实施例,所述方法可在第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂在所述重复扩增序列5'末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂,以产生第一修饰细胞,且所述方法在第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列3'末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂,以产生第二修饰细胞。替代地,所述方法可在第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂在所述重复扩增序列3'末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂,以产生第一修饰细胞,且所述方法在第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列5'末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂,以产生第二修饰细胞。替代地,所述方法可在第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂在所述重复扩增序列5'末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂,以产生第一修饰细胞,且所述方法第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列5'末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂,以产生第二修饰细胞。在一些此类方法中,第一和第二核酸酶靶位点可以相同。在另外的方法中,它们可以不同。替代地,所述方法可在第一次进行时使用第一核酸酶试剂,所述第一核酸酶试剂在所述重复扩增序列3'末端附近的第一核酸酶靶位点处造成双链断裂或单链断裂,以产生第一修饰细胞,且所述方法第二次进行时,针对所述第一修饰细胞使用第二核酸酶试剂,所述第二核酸酶试剂在所述重复扩增序列3'末端附近的第二核酸酶靶位点处造成双链断裂或单链断裂,以产生第二修饰细胞。在一些此类方法中,第一和第二核酸酶靶位点可以相同。在另外的方法中,它们可以不同。在一些这样的方法中,第一核酸酶试剂和第二核酸酶试剂都造成双链断裂。在另外的方法中,第一核酸酶试剂和第二核酸酶试剂都造成单链断裂。在另外的方法中,第一核酸酶试剂造成双链断裂,且第二核酸酶试剂造成单链断裂,反之亦然。可以使用任何组合。

[0170] 在一种示例性方法中,所述核酸酶靶位点在重复扩增序列之外,所述核酸酶靶位点位于所述重复扩增序列的5'末端或3'末端的约50、约40、约30、约20,或约10个核苷酸内,所述核酸酶试剂是Cas9蛋白和向导RNA。在另一种示例性方法中,所述核酸酶靶位点在重复扩增序列之外,所述核酸酶靶位点位于所述重复扩增序列的5'末端或3'末端的约50、约40、约30、约20,或约10个核苷酸内,所述核酸酶试剂是Cas9蛋白和向导RNA,Cas9为在所述重复扩增序列的5'末端或3'末端附近造成单链断裂的切口酶,在所述细胞修复所述单链断裂后,所述核酸酶靶位点被保留,且单链断裂的修复不会导致重复扩增序列之外的插入或删除。

[0171] 一些方法从已经包含要在靶基因组基因座处扩增的重复扩增序列的细胞开始。预先存在的重复扩增序列可以是细胞中自然发生的重复扩增序列,也可以是先前使用外源修复模板插入的重复扩增序列。类似地,预先存在的重复扩增序列可以是先前在本文公开的核酸酶介导的方法的第一轮或多轮中扩增的重复扩增序列。其它方法可以包括生成包含要在靶基因组基因座处扩增的重复扩增序列的细胞的初始步骤。参见US 2018/0094267和WO 2018/064600,出于所有目的,将每篇通过引用整体并入本文。可以例如使用包含重复扩增

序列的外源修复模板(例如外源供体序列,如靶向载体),将要扩增的初始重复扩增序列插入靶基因座中。作为一个例子,靶向载体可以包含靶向内源靶基因座处的5'靶序列的5'同源臂和靶向内源靶基因座处的3'靶序列的3'同源臂。外源修复模板还可以包含核酸插入物,该核酸插入物包括要整合到靶基因座中的DNA片段(例如重复扩增序列)。核酸插入物在靶基因座中的整合可导致靶基因座中感兴趣的核酸序列(nucleic acid sequence of interest)的添加、靶基因座中感兴趣的核酸序列的删除,或靶基因座中感兴趣的核酸序列的替换(即删除和插入)。同源臂可以侧接包含重复扩增序列的插入核酸,以产生包含重复扩增序列的初始靶基因座(例如,在预先存在的重复扩增序列的核酸酶介导的扩增方法中使用)。

[0172] 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,并且它们可以是线性或环状的形式。例如,修复模板可以是单链寡脱氧核苷酸(ssODN)。

[0173] 外源修复模板也可以包含不存在于非靶定内源Slc30a8基因座处的异源序列。例如,外源修复模板可以包含选择盒,例如侧接重组酶识别位点的选择盒。

[0174] 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,则同源臂可以侧接核酸插入物。为了便于参考,本文将同源臂称为5'和3'(即上游和下游)同源臂。该术语涉及同源臂与外源修复模板内的核酸插入物的相对位置。5'和3'同源臂对应于靶基因座内的区域,其在本文分别称为“5'靶序列”和“3'靶序列”。

[0175] 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。给定靶序列与外源修复模板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,外源修复模板(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,靶基因座中的预期突变包括在被同源臂侧接的插入核酸中。

[0176] 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)在不存在核酸酶试剂(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。

[0177] 筛选步骤可以包括例如用于评估亲代染色体的等位基因修饰(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。

[0178] 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)。

[0179] 例如,修饰的细胞(例如修饰的多能细胞或胚胎干(ES)细胞,例如小鼠ES细胞或大鼠ES细胞)可通过例如重组产生,通过:(a)将一种或多种外源供体核酸(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,其中插入核酸包含重复扩增序列,以产生包含该重复扩增序列的靶基因座;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座处整合的插入核酸(即识别在靶基因座处包含重复扩增序列的至少一个细胞)。

[0180] 替代地,修饰的细胞可以通过以下产生:(a)向细胞中引入:(i)核酸酶试剂,其中所述核酸酶试剂在靶基因座内的靶位点处诱导切口或双链断裂;以及(ii)一个或多个外源供体核酸(例如靶向载体),其可选地包含插入核酸,所述插入核酸侧接,例如,对应于5'和3'靶位点的5'和3'同源臂,所述5'和3'靶位点位于与核酸酶靶位点足够接近之处,其中所述插入核酸包含重复扩增序列,以产生包含该重复扩增序列的靶基因座;以及(c)识别至少一个细胞,所述细胞在其基因组中包含在内源靶基因座处整合的插入核酸(即识别在靶基因座处包含重复扩增序列的至少一个细胞)。替代地,修饰的细胞可以通过以下产生:(a)向细胞中引入:(i)核酸酶试剂或编码所述核酸酶试剂的核酸,其中所述核酸酶试剂在靶基因座内的靶位点处诱导切口或双链断裂;以及(ii)一个或多个外源供体核酸(例如靶向载体),其可选地包含插入核酸,所述插入核酸侧接,例如,对应于5'和3'靶位点的5'和3'同源臂,所述5'和3'靶位点位于与核酸酶靶位点足够接近之处,其中所述插入核酸包含重复扩增序列,以产生包含该重复扩增序列的靶基因座;靶向载体,其包含侧面为对应于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突变的Slc30a8基因座或将在Slc30a8基因座处进行的突变;以及(c)识别至少一个细胞,所述细胞在其基因组中包含在内源靶基因座处整合的插入核酸(即识别在靶基因座处包含重复扩增序列的至少一个细胞)。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试剂。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。

A. 核酸酶试剂和核酸酶试剂的靶位点

[0181] 本文提供了用于生产修饰的细胞的方法中的核酸酶试剂,该修饰的细胞在靶基因组基因座处具有增加的拷贝数的重复序列,该修饰的细胞包含重复扩增序列,所述重复扩增序列在靶基因组基因座处包含多个拷贝数的重复序列。本文还提供了编码核酸酶试剂的核酸,其用于生产修饰的细胞的方法,该修饰的细胞在靶基因组基因座处具有增加的拷贝数的重复序列,该修饰的细胞包含重复扩增序列,所述重复扩增序列在靶基因组基因座处

包含多个拷贝数的重复序列。这种核酸酶试剂可设计为在重复扩增序列的5'末端或3'末端附近的核酸酶靶位点处造成双链断裂或单链断裂(例如切口酶)。

[0182] 术语“核酸酶试剂的靶位点”包括核酸酶试剂在其中诱导切口或双链断裂的DNA序列。核酸酶试剂的靶位点对于细胞可以是内源的(或天然的),或者所述靶位点对于细胞可以是外源的。对于细胞是外源的靶位点并非天然存在于细胞基因组中。靶位点对于人们希望将其放置于靶基因座的感兴趣的多核苷酸而言也可以是外源的。在某些情况下,靶位点在宿主细胞的基因组中仅出现一次。

[0183] 靶位点的长度可以是不同的,并且包括,例如,对于锌指核酸酶(ZFN)对而言约为30-36bp的靶位点(即对于每个ZFN而言约为15-18bp)、对于类转录激活因子效应物核酸酶(TALEN)而言约为36bp的靶位点、对于CRISPR/Cas9向导RNA而言为约20bp的靶位点。

[0184] 本文公开的方法和组合中可以使用任何诱导切口或双链断裂进入所需靶位点的核酸酶试剂。可以使用天然存在(naturally occurring)的或天然(native)的核酸酶试剂,只要该核酸酶试剂在所需靶位点诱导切口或双链断裂即可。替代地,可以使用修饰的或工程化的核酸酶试剂。“工程化核酸酶试剂”包括从其天然形式被工程化(修饰或衍生),以特异性靶向并诱导所需靶位点中的切口或双链断裂的核酸酶。因此,工程化的核酸酶试剂可以源自天然的、天然存在的核酸酶试剂,或者其可为人工产生的或合成的。工程化的核酸酶可以在例如靶位点中诱导切口或双链断裂,例如,其中靶位点并非天然的(非工程化的或未修饰的)核酸酶试剂所靶向的序列。核酸酶试剂的修饰可以少至蛋白质切割剂中的一个氨基酸,或核酸切割剂中的一个核苷酸。在靶位点或其它DNA中产生切口或双链断裂在本文中可称为“切下(cutting)”或“切割(cleaving)”靶位点或其它DNA。

[0185] 还提供了示例性靶位点的活性变体和片段。此类活性变体可包含与给定靶位点的至少65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列同一性,其中活性变体保留生物活性,因此能够以序列特异性方式被核酸酶试剂靶向和切割。利用核酸酶试剂测量靶位点的双链断裂的测定是已知的(例如,TaqMan® qPCR测定,Frendewey (2010) *Methods in Enzymology* 476:295-307,其全部内容以引用方式并入本文中)。

[0186] 核酸酶试剂的靶位点可位于靶基因座内或附近的任何位置。靶位点可以位于基因的编码区内,或位于影响基因表达的调控区内。核酸酶试剂的靶位点可以位于内含子、外显子、启动子、增强子、调节区,或任何非蛋白质编码区中。

[0187] 一种类型的核酸酶试剂是类转录激活因子效应物核酸酶(TALEN)。TAL效应物核酸酶是一类序列特异性核酸酶,可用于在原核或真核生物基因组中的特定靶序列处造成双链断裂。TAL效应物核酸酶是通过将天然或工程化类转录激活因子(TAL)效应物或其功能部分融合到核酸内切酶(例如FokI)的催化结构域而产生的。独特的、模块化的TAL效应物DNA结合域允许设计具有潜在任何给定的DNA靶向特异性的蛋白质。因此,TAL效应物核酸酶的DNA结合域可以被工程化为靶向特定的DNA靶位点,从而用于在所需的靶序列上造成双链断裂。参见WO 2010/079430;Morbiter et al. (2010) *PNAS* 10.1073/pnas.1013133107;Scholze & Boch (2010) *Virulence* 1:428-432;Christian et al. (2010) *Genetics* 186:757-761;Li et al. (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704;和Miller et al. (2011) *Nature Biotechnology* 29:143-148,出于所有目的,每篇文献均通过引用整体并入本文。

[0188] 合适的TAL核酸酶的例子和用于制备合适的TAL核酸酶的方法公开于,例如,US 2011/0239315 A1、US 2011/0269234 A1、US 2011/0145940 A1、US 2003/0232410 A1、US 2005/0208489 A1、US 2005/0026157 A1、US 2005/0064474 A1、US 2006/0188987 A1和US 2006/0063231 A1,出于所有目的,每篇都通过引用整体并入本文。

[0189] 在一些TALEN中,TALEN的每个单体包含33-35个TAL重复,其通过两个高变残基识别单个碱基对。在一些TALEN中,核酸酶试剂是嵌合蛋白,其包含可操作地连接至独立核酸酶(例如FokI核酸内切酶)的基于TAL重复的DNA结合域。例如,核酸酶试剂可包含第一基于TAL重复的DNA结合域和第二基于TAL重复的DNA结合域,其中第一和第二基于TAL重复的DNA结合域中的每一个可操作地连接至FokI核酸酶,其中第一和第二基于TAL重复的DNA结合域识别靶DNA序列的每条链中的两个连续靶DNA序列,其由不同长度(12-20bp)的间隔序列隔开,并且其中FokI核酸酶亚基二聚化以产生活性核酸酶,使靶序列上产生双链断裂。

[0190] 在本文公开的各种方法和组合物中使用的核酸酶试剂可以包括锌指核酸酶(ZFN)。在一些ZFN中,ZFN的每个单体包含3个或更多个基于锌指的DNA结合域,其中每个基于锌指的DNA结合域结合3bp亚位点。在其它ZFN中,ZFN是嵌合蛋白,包含基于锌指的DNA结合域,该域与独立的核酸酶(如FokI核酸内切酶)可操作地连接。例如,核酸酶试剂可包含第一ZFN和第二ZFN,其中第一ZFN和第二ZFN中的每一个与FokI核酸酶亚基可操作地连接,其中第一和第二ZFN识别靶DNA序列的每条链中的两个连续靶DNA序列,其由不同长度(5-7bp)的间隔序列隔开,并且其中FokI核酸酶亚基二聚化以产生活性核酸酶,使靶序列上产生双链断裂。参见,例如,US20060246567;US20080182332;US20020081614;US20030021776;WO/2002/057308A2;US20130123484;US20100291048;WO/2011/017293A2;和Gaj et al. (2013) Trends Biotechnol. 31(7):397-405,每篇出于所有目的通过引用整体并入本文。

[0191] 另一类核酸酶试剂是兆核酸酶。兆核酸酶根据保守序列模体分为四个家族,这些家族是LAGLIDADG、GIY-YIG、H-N-H和His-Cys盒家族。这些模体参与金属离子的配位和磷酸二酯键的水解。兆核酸酶以其长的靶位点以及在其DNA底物中耐受某些序列多态性而著称。兆核酸酶域、结构和功能是已知的,参见,例如,Guhan和Muniyappa (2003) Crit.Rev.Biochem.Mol.Biol. 38(3):199-248;Lucas et al. (2001) Nucleic Acids Res. 29(4):960-969;Jurica and Stoddard (1999) Cell.Mol.Life Sci. 55:1304-1326;Stoddard (2006) Q.Rev.Biophys. 38:49-95;和Moure et al. (2002) Nat.Struct.Biol. 9(10):764-770。在一些例子中,使用了天然存在的变体和/或工程化衍生的兆核酸酶。改变动力学、辅因子相互作用、表达、最佳条件和/或靶位点特异性以及筛选活性的方法是已知的。参见,例如,Epinat et al. (2003) Nucleic Acids Res. 31:2952-62;Chevalier et al. (2002) Mol.Cell 10:895-905;Gimble et al. (2003) Mol.Biol. 334:993-1008;Seligman et al. (2002) Nucleic Acids Res. 30:3870-9;Sussman et al. (2004) J.Mol.Biol. 342:31-41;Rosen et al. (2006) Nucleic Acids Res. 34:4791-800;Chames et al. (2005) Nucleic Acids Res 33:e178;Smith et al. (2006) Nucleic Acids Res. 34:e149;Gruen et al. (2002) Nucleic Acids Res. 30:e29;Chen和Zhao (2005) Nucleic Acids Res. 33:e154;W02005105989;W02003078619;W02006097854;W02006097853;W02006097784;和W02004031346,每篇出于所有目的通过引用整体并入本文。

[0192] 可以使用任何兆核酸酶,包括,例如,I-SceI、I-SceII、I-SceIII、I-SceIV、I-

SceV、I-SceVI、I-SceVII、I-CeuI、I-CeuAIIP、I-CreI、I-CrepsbIP、I-CrepsbIIP、I-CrepsbIIIP、I-CrepsbIVP、I-TliI、I-PpoI、PI-PspI、F-SceI、F-SceII、F-SuvI、F-TevI、F-TevII、I-AmaI、I-AniI、I-ChuI、I-CmoeI、I-CpaI、I-CpaII、I-CsmI、I-CvuI、I-CvuAIP、I-DdiI、I-DdiII、I-DirI、I-DmoI、I-HmuI、I-HmuII、I-HsNIP、I-LlaI、I-MsoI、I-NaaI、I-NanI、I-NcIIP、I-NgrIP、I-NitI、I-NjaI、I-Nsp236IP、I-PakI、I-PboIP、I-PcuIP、I-PcuAI、I-PcuVI、I-PgrIP、I-PobIP、I-PorI、I-PorIIP、I-PbpIP、I-SpBetaIP、I-ScaI、I-SexIP、I-SneIP、I-SpomI、I-SpomCP、I-SpomIP、I-SpomIIP、I-SquIP、I-Ssp6803I、I-SthPhiJP、I-SthPhiST3P、I-SthPhiSTe3bP、I-TdeIP、I-TevI、I-TevII、I-TevIII、I-UarAP、I-UarHGPAIP、I-UarHGPA13P、I-VinIP、I-ZbiIP、PI-MtuI、PI-MtuHIP、PI-MtuHIIP、PI-PfuI、PI-PfuII、PI-PkoI、PI-PkoII、PI-Rma43812IP、PI-SpBetaIP、PI-SceI、PI-TfuI、PI-TfuII、PI-ThyI、PI-TliI、PI-TliII,或其任何活性变体或片段。

[0193] 兆核酸酶可以靶向,例如,12至40个碱基对的双链DNA序列。在某些情况下,兆核酸酶靶向基因组中一个完美匹配的靶序列。

[0194] 一些兆核酸酶是归巢核酸酶。一种类型的归巢核酸酶(homing nuclease)是归巢核酸酶的LAGLIDADG家族,包括,例如,I-SceI、I-CreI和I-Dm1。

[0195] 核酸酶试剂可以进一步包括限制性内切核酸酶,其包括I型、II型、III型和IV型内切核酸酶。I型和III型限制性内切核酸酶靶向特定靶位点,但通常在离核酸酶结合位点的可变位置切割,该位点可能距离切割位点(靶位点)数百个碱基对。在II型系统中,限制活性与任何甲基化酶活性无关,切割通常发生在结合位点内或附近的特定位点。大多数II型酶切割回文序列,然而IIa型酶靶向非回文靶位点并在靶位点外切割,IIb型酶在靶位点外的两个位点切割序列两次,而IIs型酶靶向不对称靶位点并且在离靶位点约1-20个核苷酸的确定的距离处在一侧切割。IV型限制酶靶向甲基化DNA。在例如REBASE数据库(rebase.neb.com的网页;Roberts et al. (2003) *Nucleic Acids Res.* 31 (1):418-420; Roberts et al. (2003) *Nucleic Acids Res.* 31 (7):1805-1812;和Belfort et al. (2002) 于 *Mobile DNA II*, pp.761-783, Eds. Craigie et al. (ASM Press, Washington, DC) 中进一步描述和分类了限制酶,每篇文献均通过引用整体并入本文。

[0196] 还提供了核酸酶试剂(即工程化核酸酶试剂)的活性变体和片段。此类活性变体可包含与天然核酸酶试剂的至少65%、70%、75%、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目标靶位点处切割的能力,并因此保留切口或双链断裂诱导活性。例如,本文所述的任何核酸酶试剂都可以修饰自天然核酸内切酶序列,并设计为在未被天然核酸酶试剂靶向的靶位点处靶向和诱导切口或双链断裂。因此,一些工程化的核酸酶具有在不同于相应的天然核酸酶试剂靶位点的靶位点处诱导切口或双链断裂的特异性。对切口或双链断裂诱导活性的测定是已知的,并且通常测量内切核酸酶对含有靶位点的DNA底物的总体活性和特异性。

[0197] 可以通过本领域已知的任何方式将核酸酶试剂引入细胞中。编码核酸酶试剂的多肽可以直接导入细胞中。或者,可以将编码核酸酶试剂的多核苷酸引入细胞。当编码核酸酶试剂的多核苷酸被引入细胞时,核酸酶试剂可以在细胞内瞬时、条件性或组成性表达。因此,编码核酸酶试剂的多核苷酸可包含在表达盒中并且可操作地连接至条件启动子、诱导型启动子、组成型启动子或组织特异性启动子。此类启动子在本文别处进一步详细描述。或

者,核酸酶试剂可以作为编码核酸酶试剂的mRNA引入细胞中。

[0198] 编码核酸酶试剂的多核苷酸可以稳定地整合到细胞的基因组中并且可操作地连接到在细胞中有活性的启动子。或者,编码核酸酶试剂的多核苷酸可以在靶向载体中(例如包含插入多核苷酸的靶向载体,或在与包含插入多核苷酸的靶向载体分开的载体或质粒中)。

[0199] 当通过引入编码核酸酶试剂的多核苷酸向细胞提供核酸酶试剂时,这样的编码核酸酶试剂的多核苷酸可以被修饰以替代在感兴趣的细胞中与编码核酸酶试剂的天然存在的多核苷酸序列相比具有更高使用频率的密码子。例如,可以修饰编码核酸酶试剂的多核苷酸,以取代在给定的感兴趣的细胞中与编码核酸酶试剂的天然存在的多核苷酸序列相比具有更高使用频率的密码子,所述原核或真核细胞包括细菌细胞、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,或任何其它感兴趣的宿主细胞。

B. CRISPR/Cas系统

[0200] 本文公开的方法和组合物可以利用成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关(Cas)系统或此类系统的组件来修饰细胞内的基因组。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)。本文公开的方法和组合物可以通过利用CRISPR复合物(包括与Cas蛋白复合的向导RNA(gRNA))来使用CRISPR/Cas系统,以进行核酸的定点结合或切割。

[0201] 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,其已从其天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。

[0202] Cas蛋白。Cas蛋白通常包含至少一个可以与向导RNA相互作用的RNA识别或结合域。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蛋白通常会产生产物。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,从而在靶基因组基因座处造成双链断裂(例如,末端为平端的双链断裂),或者它可以是在靶基因组基因座上造成单链断裂的切口酶。

[0203] 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。

[0204] 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。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*)、伯克霍尔德里氏菌(*Burkholderiales bacterium*)、食萘极地单胞菌(*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 beccsii*)、金矿菌(*Candidatus Desulforudis*)、肉毒梭菌(*Clostridium botulinum*)、难辨梭菌(*Clostridium difficile*)、大芬戈尔德菌(*Fingoldia 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(*Petrogamobilis*)、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。来自化脓链球菌的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。来自脑膜炎奈瑟氏球菌的Cas9 (Nme2Cas9) 是另一示例性的Cas9蛋白。参见,例如,Edraki et al. (2019)Mol.Cell 73(4):714-726,出于所有目的通过引用整体并入本文。其它示例性的Cas9蛋白为来自嗜热链球菌的Cas9蛋白(例如由CRISPR1基因座编码的嗜热链球菌LMD-9Cas9 (St1Cas9) 或来自CRISPR3基因座的嗜热链球菌Cas9 (St3Cas9))。其它示例性的Cas9蛋白为来自新凶手弗朗西丝氏菌(Francisellanovicida)的Cas9 (FnCas9) 或识别替代性的PAM的RHA新凶手弗朗西丝氏菌Cas9变体(E1369R/E1449H/R1556A置换)。在例如Mamm.Genome 28(7):247-261中总结了这些Cas9蛋白和其它示例性Cas9蛋白,出于所有目的通过引用整体并入本文。

[0205] 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(Francisella1)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(Francisellatularensis 1)、新凶手土拉弗朗西斯氏菌(Francisellatularensis subsp.novicida)、苏格兰普雷沃氏菌(Prevotellaalbensis)、毛螺科菌MC2017 1(Lachnospiraceae bacterium MC2017 1)、解蛋白丁酸弧菌(Butyrivibrioproteoclasticus)、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)、史密斯氏互养菌属SCADC(Smithella sp.SCADC)、氨基酸球菌属BV3L6(Acidaminococcus sp.BV3L6)、毛螺科菌MA2020(Lachnospiraceae bacterium MA2020)、CandidatusMethanoplasmatermitum、挑剔真杆菌(Eubacteriumeligens)、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(Porphyromonascrevioricanis 3)、解糖脲普雷沃氏菌(Prevotelladisians)和猕猴卟啉单胞菌(Porphyromonasmacacae)。来自新凶手土拉弗朗西斯氏菌U112(FrancisellanovicidaU112)的Cpf1 (FnCpf1;分配的UniProt登录号为A0Q7Q2)是示例性Cpf1蛋白。

[0206] 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。

[0207] 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/R661A/Q695A/Q926A)。参见,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/K1003A/

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。这些和其它修饰的Cas蛋白在例如Cebrian-Serrano和Davies (2017) *Mamm. Genome* 28 (7): 247-261中进行了综述, 出于所有目的通过引用整体并入本文。另一个修饰的Cas9蛋白的例子是xCas9, 它是一种SpCas9变体, 可以识别更大范围的PAM序列。参见, 例如, Hu et al. (2018) *Nature* 556: 57-63, 出于所有目的通过引用整体并入本文。

[0208] 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)。

[0209] 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中造成双链断裂。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。

[0210] 一个或多个或所有核酸酶结构域可以被删除或突变, 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内造成单链断裂, 但不能造成双链断裂 (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863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, Saprunauskas 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。

[0211] xCas9的催化结构域中失活突变的实例与上文针对SpCas9所述的那些相同。金黄

色葡萄球菌Cas9蛋白的催化结构域中失活突变的例子也是已知的。例如,金黄色葡萄球菌Cas9酶(SaCas9)可以包含在N580位处的取代(例如,N580A取代)和在D10位处的取代(例如,D10A取代)以产生无核酸酶活性的Cas蛋白。参见,例如,WO2016/106236,出于所有目的通过引用将其整体并入本文。Nme2Cas9催化域中失活突变的例子也是已知的(例如,D16A和H588A的组合)。St1Cas9催化结构域中失活突变的例子也是已知的(例如,D9A、D598A、H599A和N622A的组合)。St3Cas9催化结构域中失活突变的例子也是已知的(例如D10A和N870A的组合)。CjCas9催化域中失活突变的例子也是已知的(例如D8A和H559A的组合)。FnCas9和RHA FnCas9催化结构域中失活突变的例子也是已知的(例如N995A)。

[0212] 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。

[0213] 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WO 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。

[0214] 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。

[0215] 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。

[0216] 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。

[0217] 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。

[0218] 可以任何形式提供Cas蛋白。例如,Cas蛋白可以以蛋白质的形式提供,所述蛋白质为,例如,与gRNA复合的Cas蛋白。或者,Cas蛋白可以以编码Cas蛋白的核酸形式提供,例如RNA(例如,信使RNA(mRNA))或DNA。可选地,可以对编码Cas蛋白的核酸进行密码子优化,以在特定细胞或生物体中有效翻译成蛋白质。例如,可以修饰编码Cas9蛋白的核酸,以替换在细菌细胞、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与天然核酸序列相比具有更高使用频率的密码子。当将编码Cas蛋白的核酸引入细胞时,Cas蛋白可以在细胞中瞬时、条件性或组成性表达。

[0219] 可以修饰以mRNA形式提供的Cas蛋白,以提高稳定性和/或免疫原性特性。可以对mRNA内的一个或多个核苷进行修饰。对mRNA核碱基进行化学修饰的例子包括假尿苷、1-甲基-假尿苷和5-甲基-胞苷。例如,可以使用包含N1-甲基假尿苷的加帽(capped)和聚腺苷酸化Cas mRNA。同样,Cas mRNA可以通过使用同义密码子去除尿苷来修饰。

[0220] 编码Cas蛋白的核酸可以稳定地整合到细胞的基因组中并且可操作地连接到在细胞中有活性的启动子。或者,编码Cas蛋白的核酸可以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指导基因或其他感兴趣的核酸序列(例如Cas基因)的表达,并且可以将此类感兴趣的核酸序列转移至靶细胞的任何核酸构建体。例如,编码Cas蛋白的核酸可以在包含编码gRNA的DNA的载体中。或者,它可以在与包含编码gRNA的DNA的载体分开的载体或质粒中。可用于表达构建体中的启动子包括在例如真核细胞、人类细胞、非人类细胞、哺乳动物细胞、非人类哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、多能细胞、胚胎干(ES)细胞、成体干细胞、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个或多个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。可选地,启动子可以是双向启动子,其在一个方向上驱动Cas蛋白的表达,而在另一方向上驱动向导RNA的表达。这种双向启动子可以由(1)完整的、常规的单向Pol 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Pol 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向导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。

[0221] 向导RNA。“向导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。向导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(targeter-RNA)”(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”、“单向导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,出于所有目的在此通过引用将每一篇的全部内容并入本文。例如,对于Cas9,单向导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“向导RNA”和“gRNA”包括双分子(即模块)gRNA和单分子gRNA。

[0222] 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:19)。位于DNA靶向段下游(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:GUUUUAGAGCUAUGCUGUUUUG(SEQ ID NO:83)。可将本文公开的任何靶向DNA的片段(即向导序列)(例如SEQ ID NO:84-91的任一)连接至SEQ ID NO:19或SEQ ID NO:83的5'末端以形成crRNA。

[0223] 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:

AGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUU (SEQ ID NO:20)。tracrRNA序列的另一个实施例包含以下,基本上由以下组成或由以下组成:

GUUGGAACCAUUCAAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU (SEQ ID NO:82)。tracrRNA序列的其它实施例包含以下,基本上由以下组成或由以下组成:

AAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (SEQ ID NO:116),或

GUUGGAACCAUUCAAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC (SEQ ID NO:117)。

[0224] 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。

[0225] 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。

[0226] 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。

[0227] 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(作为单向导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。单向导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。

[0228] 向导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。在一些向导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,错配在向导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)。

[0229] 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。

[0230] 单向导RNA可以包含DNA靶向段和支架序列(即向导RNA的蛋白质结合序列或Cas结合序列)。例如,此类向导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:

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA
GUCGGUGCU(版本1;SEQ ID NO:21);

GUUGGAACCAUUCAAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGG
CACCGAGUCGGUGC(版本2;SEQ ID NO:22);

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA
GUCGGUGC(版本3;SEQ ID NO:23);

GUUUUAGAGCUAUGCUGGAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA
GUGGCACCGAGUCGGUGC(版本4;SEQ ID NO:24);

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA
GUCGGUCUUUUUUU(版本5;SEQ ID NO:118);

GUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGA
GUCGGUCUUUUU(版本6;SEQ ID NO:119);或

GUUUAAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGCUUUUUU (版本7; SEQ ID NO: 120)。靶向本文公开的任何向导RNA靶序列的向导RNA可以包括,例如,与向导RNA的3'末端上的任何示例性向导RNA支架序列融合的向导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,以形成单向导RNA(嵌合向导RNA)。

[0231] 向导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、转录阻遏物、DNA甲基转移酶、DNA去甲基化酶、组蛋白乙酰转移酶、组蛋白去乙酰化酶等);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。

[0232] 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。向导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。其它可能的向导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。

[0233] 作为一个实例,向导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,向导RNA可以在向导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,向导RNA的5'和/或3'末端的核苷酸可以具有2'-O-甲基修饰。例如,向导RNA可以在向导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。本文的其它地方更详细地描述了其它可能的修改。在一个具体实施例中,向导RNA在前三个5'和3'末端RNA残基处包括2'-O-甲基类似物和3'硫代磷酸酯核苷酸间键。此类化学修饰可以,例如,为向导RNA提供更高的稳定性和针对外切核酸酶的保护,使它们在细胞内比未修饰的向导RNA持续更长时间。这种化学修饰还可以,例

如,防止先天细胞内免疫反应,这些免疫反应可以主动降解RNA或触发导致细胞死亡的免疫级联反应。

[0234] 向导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分子或提供为分别编码crRNA和tracrRNA的分离的DNA分子。

[0235] 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(例如编码Cas蛋白的核酸)的载体中,或者,其可以在与包含编码Cas蛋白的核酸的载体分开的载体或质粒中。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个或多个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。

[0236] 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)。向导RNA也可以是通过化学合成制备的、合成产生的分子。例如,可以化学合成向导RNA,以在前三个5'和3'末端RNA残基处包括2'-O-甲基类似物和3'硫代磷酸酯核苷酸间键。

[0237] 向导RNA(或编码向导RNA的核酸)可以在包含一个或多个向导RNA(例如1、2、3、4或更多个向导RNA)和增加向导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。

[0238] 向导RNA靶序列。向导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”。

[0239] 靶DNA既包括与向导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“向导RNA靶序列”具体是指非

互补链上的序列,该序列对应与向导RNA在互补链上杂交的序列(即其反向互补)。即,向导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')。向导RNA靶序列等同于向导RNA的DNA靶向段,但是胸腺嘧啶而非尿嘧啶。作为一个例子,SpCas9酶的向导RNA靶序列可以指非互补链上5'-NGG-3' PAM上游的序列。向导RNA被设计成与靶DNA的互补链具有互补性,其中向导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。如果在本文中向导RNA称为靶向向导RNA靶序列,则是指该向导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向导RNA靶序列的反向互补。

[0240] 靶DNA或向导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或向导RNA靶序列可以是细胞内源或外源的任何核酸序列。向导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。

[0241] Cas蛋白对靶DNA的位点特异性结合和切割可发生在由以下所决定的位置处:(i)向导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可以侧接向导RNA靶序列。可选地,向导RNA靶序列可以在3'末端上侧接PAM(例如对于Cas9)。替代地,向导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(例如在向导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的非互补链上紧邻向导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',向导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NNGRRRT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'。

[0242] 向导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,向导RNA靶序列加PAM的两个例子是GN₁₉NGG(SEQ ID NO:25)或N₂₀NGG(SEQ ID NO:26)。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。向导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:27),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。其它向导RNA靶序列加PAM可以具有SEQ ID NO:25-27的4-22个核苷酸的长度,包括5'G或GG和3'GG或NGG。其它向导RNA靶序列加PAM可以具有SEQ ID NO:25-27的14至20个核苷酸的长度。

[0243] 与靶DNA杂交的CRISPR复合物的形成可导致对应于向导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的非互补链上的向导RNA靶序列和与向导RNA杂交的互补链上的反向互补序列)。例如,切割位点可以在向导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蛋白在其上造成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位点处(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造成单链断裂,从而造成双链断裂。例如,第一切口酶可在双链DNA(dsDNA)的第一链上造成单链断裂,第二切口酶可在dsDNA的第二链上造成单链断裂,从而产生突出的序列。在一些情况下,第一链上的向导RNA靶序列或切口酶的切割位点与第二链上的向导RNA靶序列或切口酶的切割位点至少相隔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。

III、包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物、非人类动物细胞和非人类动物基因组

[0244] 本文公开的非人类动物基因组、非人类动物细胞和非人类动物可包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列。还公开了包含异源六核苷酸重复扩增序列的非人类动物C9orf72基因。本文还公开了包含在内源C9orf72基因座处插入的扩增的六核苷酸重复扩增序列的基因组、细胞(例如人类诱导的多能干细胞)和动物。还公开了包含扩增的六核苷酸重复扩增序列的动物C9orf72基因。例如,异源六核苷酸重复扩增序列可以包含超过100个重复的如SEQ ID NO:1所示的六核苷酸序列。

A. C9ORF72、六核苷酸重复扩增序列和相关疾病

[0245] 肌萎缩侧索硬化(ALS),也称为楼格瑞病,是最常见的成年发作型麻痹病症,其特征是上和/或下运动神经元损失。全美国发生ALS的个体多达20,000位,每年出现约5,000个新病例。额颞痴呆(FTD),也被称为皮克病、额颞叶变性,或FTLD,是脑额叶或颞叶出现进行性细胞退化而引起的一组病症。据报道,FTD在所有痴呆病例中占10-15%。介于人类C9ORF72基因的(和任选地跨越)外显子1A与1B(两种非编码外显子)之间的六核苷酸重复扩增序列已经与ALS和FTD均相关。据估计,GGGGCC(SEQ ID NO:1)六核苷酸重复扩增序列在家族性和许多非家族性ALS病例中占约50%。其存在于约25%的家族性FTD病例和约8%的散发性FTD病例中。

[0246] 已报告了与C9ORF72中的六核苷酸重复扩增序列相关的许多病理学方面,例如RNA病灶的形成与重复长度相关、特定RNA结合蛋白的隔离,以及二肽重复蛋白的聚积和聚集(例如由重复相关的非AUG(AUG)翻译产生的聚(甘氨酸-丙氨酸)、聚(甘氨酸-脯氨酸)、聚(甘氨酸-精氨酸)、聚(丙氨酸-脯氨酸),或聚(脯氨酸-精氨酸)二肽重复蛋白)。含有异源六核苷酸重复扩增序列(该序列包含六核苷酸序列(GGGGCC;SEQ ID NO:1)的66个重复)的小鼠已由腺相关病毒介导的体细胞脑转基因产生,并表现出由其神经元中重复相关的非AUG(AUG)翻译产生的RNA病灶和二肽蛋白聚集体。参见Chew et al. (2015) Science 348 (6239):1151-1154,出于所有目的通过引用整体并入本文。

[0247] 尽管已报道C9ORF72调节内体运输,但C9ORF72的大部分细胞功能仍然未知。确实,C9ORF72是一种编码功能未知的、未表征的蛋白质的基因。

[0248] 已报道了小鼠C9ORF72转录物变体。参见,例如,Koppers et al. (2015) Ann.Neurol.78:426-438和Atkinson et al. (2015) Acta Neuropathologica Communications 3:59,每一篇均出于所有目的通过引用整体并入本文。所报道的三种小鼠C9ORF72转录物变体的基因组信息还可以在Ensembl网站、在ENSMUST00000108127(V1)、

ENSMUST00000108126 (V2) 和ENSMUST00000084724 (V3) 名称下获得。示例性的非人(例如啮齿动物)C9orf72 mRNA和氨基酸序列在SEQ ID NO:40-43中列出。小鼠C9orf72的mRNA和氨基酸序列可以分别在GenBank登录号NM_001081343和NP_001074812中找到,并且出于所有目的通过引用整体并入本文。NM_001081343.1和NP_001074812.1的序列分别在SEQ ID NO:40和41中列出。大鼠C9orf72的mRNA和氨基酸序列可以分别在GenBank登录号NM_001007702和NP_001007703中找到,并且出于所有目的通过引用整体并入本文。NM_001007702.1和NP_001007703.1的序列分别在SEQ ID NOS:42和43中列出。

[0249] 人类C9ORF72转录物变体也是已知的。一个人类C9ORF72转录物变体的中心和3'编码区缺乏多个外显子,并且其3'末端外显子延伸超过变体3中所用的剪接位点(参见下文),从而产生新颖的3'非翻译区(UTR)(相较于变体3)。这种变体编码的多肽显著地更短并且其C末端氨基酸不同于其它两种变体所编码的氨基酸。这种变体的mRNA和氨基酸序列分别可以GenBank登录号NM_145005.6和NP_659442.2找到,并且以引用的方式出于所有目的并入本文中。NM_145005.6和NP_659442.2的序列分别列于SEQ ID NO:13和SEQ ID NO:14。第二人类C9ORF72转录物变体(2)的5'非翻译区(UTR)不同于变体3。这种变体的mRNA和氨基酸序列分别可以GenBank登录号NM_018325.4和NP_060795.1找到,并且以引用的方式出于所有目的并入本文中。NM_018325.4和NP_060795.1的序列分别列于SEQ ID NO:15和SEQ ID NO:16。第三人类C9ORF72转录物变体(3)含有所报道的三种变体中的最长序列并且编码更长的异型体。这种变体的mRNA和氨基酸序列分别可以GenBank登录号NM_001256054.2和NP_001242983.1找到,并且以引用的方式出于所有目的并入本文中。NM_001256054.2和NP_001242983.1的序列分别示于SEQ ID NO:17和SEQ ID NO:18。变体2和3编码相同的蛋白质。

B. 在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列

[0250] 本文描述的是包含插入内源C9orf72基因座(例如内源非人类动物C9orf72基因座)的六核苷酸重复扩增序列的基因组、细胞或动物(例如非人类动物基因组、细胞或动物)。细胞或动物可以是本文别处所描述的任何类型的细胞或动物。在一些基因组、细胞或动物中,六核苷酸重复扩增序列可以是异源序列。或者,它可以是内源序列,其随后使用本文公开的方法进一步扩增。术语“异源的”在核酸上下文中使用时表示该核酸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,当涉及核酸的片段使用时,术语“异源的”表示该核酸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,该核酸分子在自然界中不与另外的分子相关联。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。例如,核酸的异源区域可以包括侧接重复扩增序列的C9orf72序列,该重复扩增序列在自然界中未发现与C9orf72序列相关联。或者,核酸的异源区域可包括侧接内源非人核酸序列的人核酸序列。

[0251] C9ORF72六核苷酸重复扩增序列通常是包含至少两个(即两个重复)的如SEQ ID NO:1所示的六核苷酸序列GGGGCC的的核苷酸序列。为了插入内源非人C9orf72基因座中,异源六核苷酸重复扩增序列包含多个(即至少两个(重复))的如SEQ ID NO:1所示的六核苷酸序列。

[0252] 在一些异源六核苷酸重复扩增序列中,重复是连续的(彼此相邻而没有插入序

列)。

[0253] 异源六核苷酸重复扩增序列可以具有任意数量的重复。例如,重复扩增序列可包含多于约95个重复、多于约96个重复、多于约97个重复、多于约98个重复、多于约99个重复、多于约100个重复、多于约101个重复、多于约102个重复、多于约103个重复、多于约104个重复、多于约105个重复、多于约150个重复、多于约200个重复、多于约250个重复、多于约295个重复、多于约296个重复、多于约297个重复、多于约298个重复、多于约299个重复、多于约300个重复、多于约301个重复、多于约302个重复、多于约303个重复、多于约304个重复、多于约305个重复、多于约350个重复、多于约400个重复、多于约450个重复、多于约500个重复、多于约550个重复、多于约595个重复、多于约596个重复、多于约597个重复、多于约598个重复、多于约599个重复、多于约600个重复、多于约601个重复、多于约602个重复、多于约603个重复、多于约604个重复,或多于约605个重复。或者,重复扩增序列可包含至少约95个重复、至少约96个重复、至少约97个重复、至少约98个重复、至少约99个重复、至少约100个重复、至少约101个重复、至少约102个重复、至少约103个重复、至少约104个重复、至少约105个重复、至少约150个重复、至少约200个重复、至少约250个重复、至少约295个重复、至少约296个重复、至少约297个重复、至少约298个重复、至少约299个重复、至少约300个重复、至少约301个重复、至少约302个重复、至少约303个重复、至少约304个重复、至少约305个重复、至少约350个重复、至少约400个重复、至少约450个重复、至少约500个重复、至少约550个重复、至少约595个重复、至少约596个重复、至少约597个重复、至少约598个重复、至少约599个重复、至少约600个重复、至少约601个重复、至少约602个重复、至少约603个重复、至少约604个重复,或至少约605个重复。在一个具体实例中,异源六核苷酸重复扩增序列包含多于约100个重复、多于约300个重复、多于约600个重复、至少约100个重复、至少约300个重复,或至少约600个重复。

[0254] 重复扩增序列可以位于,例如,内源C9orf72基因座的第一个非编码内源外显子和外显子2之间。

[0255] 在一些例子中,修饰的内源非人类动物C9orf72基因座还可以包含额外的种间同源C9orf72序列。作为一个例子,非人类动物C9orf72基因座可以包含人C9ORF72核苷酸序列。例如,非人类动物C9orf72基因座可以包含用六核苷酸重复扩增序列和侧接的种间同源人C9ORF72序列(即,位于如SEQ ID NO:1所示的六核苷酸序列的重复的侧面的异源(例如人)序列)替换内源C9orf72基因座的5'非翻译和/或非编码内源非人序列。作为一个具体实施例,跨越内源外显子1(例如外显子1A和/或1B)和内源非人类C9orf72基因座的ATG起始密码子之间(并任选地包括其至少一部分)的非翻译和/或非编码序列或其部分可以被替换为异源六核苷酸重复扩增序列,并且可选地与侧接的种间同源人C9ORF72序列一起。例如,插入内源C9orf72基因座的序列可从5'至3'包含:第一异源六核苷酸侧接序列、多个重复的SEQ ID NO:1中所示出的六核苷酸,以及第二异源六核苷酸侧接序列。

[0256] 在一个实施例中,将跨越(以及可选地包括)人C9ORF72基因的全部或部分的外显子1A和/或外显子1B的异源人C9ORF72序列插入到内源C9orf72基因座中。例如,第一异源六核苷酸侧接序列可包含如SEQ IDNO:46所示的序列或其一部分,和/或第二异源六核苷酸侧接序列可包含如SEQ ID NO:47所示的序列或其一部分。因此,在一些实施例中,异源六核苷酸重复扩增序列所插入的内源C9orf72基因座可包含人序列,该人序列包含SEQ ID NO:46

和/或SEQ ID NO:47。

[0257] 在一个具体实施例中,内源C9orf72基因座包含用异源人六核苷酸重复扩增序列替换小鼠C9orf72基因座的部分非编码序列,所述异源人六核苷酸重复扩增序列与小鼠C9orf72启动子和/或人调控元件(例如,可能在人C9ORF72基因的外显子1A和/或1B中发现的那些)可操作地连接。参见US 2018/0094267和W0 2018/064600,出于所有目的,每篇都通过引用出于所有目的整体并入本文。为了插入内源非人C9orf72基因座,异源六核苷酸重复扩增序列可以包含多个(即至少两个(重复的))如SEQ ID NO:1所示的六核苷酸序列,并且可与跨越(以及可选地包括)人染色体9开放阅读框72(C9ORF72)的非编码外显子1A和外显子1B的基因组核酸序列(或其一部分)相同或基本相同。

[0258] 可选地,包含六核苷酸重复扩增序列的C9orf72基因座可以包含其它元件。此类元件的例子可包括选择盒、报告基因、重组酶识别位点,或其它元件。或者,基因座可以缺少其它元件(例如,可以缺少选择标记或选择盒)。合适的报告基因和报告蛋白的例子在本文别处公开。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)。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。

[0259] 其它元件,例如报告基因或选择盒,可以是侧接重组酶识别位点的自删除盒。参见,例如,US 8,697,851和US 2013/0312129,每篇都出于所有目的通过引用整体并入本文。例如,自删除盒可以包含与小鼠Prm1启动子可操作地连接的Crei基因(包含两个编码Cre重组酶的外显子,它们被内含子隔开)和与人泛素启动子可操作地连接的新霉素抗性基因。通过使用Prm1启动子,可以在F0动物的雄性生殖细胞中特异性删除自删除盒。编码选择标记的多核苷酸可以与在被靶向的细胞中具有活性的启动子可操作地连接。启动子的实例在本文别处描述。作为另一个具体实施例,自删除选择盒可以包含与一个或多个启动子(例如人泛素和EM7启动子)可操作地连接的潮霉素抗性基因编码序列,接着是多聚腺苷酸化信号,接着是可操作地连接到一个或多个启动子的Crei编码序列(例如mPrm1启动子),然后是另一个聚腺苷酸化信号,其中整个盒侧接loxP位点。

[0260] 基因座也可以是条件等位基因。例如,条件等位基因可以是多功能等位基因,如US 2011/0104799中所述,出于所有目的将其全文以引用方式并入本文。例如,条件等位基因可以包括:(a)相对于靶基因转录的有义方向的启动序列;(b)有义或反义方向的药物选择盒(DSC);(c)反义方向的感兴趣的核苷酸序列(NSI);以及(d)逆转条件模块(conditional by inversion module)(COIN,其利用外显子断裂内含子和可逆转的基因诱捕样模块)。参见,例如,US 2011/0104799。条件等位基因可进一步包含可重组单元,其在暴露于第一重组酶时重组以形成条件等位基因,该条件等位基因(i)缺乏启动序列和DSC;以及(ii)包含有义方向的NSI和反义方向的COIN。参见,例如,US 2011/0104799。

C. 包含插入于内源C9orf72基因座处的异源六核苷酸重复扩增序列的非人类动物

基因组、非人类动物细胞和非人类动物

[0261] 提供了动物基因组、动物细胞和动物，其包含如本文其它地方所述的，在内源C9orf72基因座处插入的六核苷酸重复扩增序列。在一些基因组、细胞或动物中，六核苷酸重复扩增序列可以是异源序列。或者，它可以是内源序列，其使用本文公开的方法进一步扩增。例如，提供了非人类动物基因组、非人类动物细胞和非人类动物，其包含如本文其它地方所述的，在内源C9orf72基因座处插入的六核苷酸重复扩增序列。或者，基因组或细胞可以是人类的（例如人诱导性多能干细胞）。基因组、细胞或非人类动物可以是雄性或雌性。对于在内源C9orf72基因座处插入的六核苷酸重复扩增序列，基因组、细胞或非人类动物可以是杂合的或纯合的。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。非人类动物可以包含插入到它们种系基因组中的内源C9orf72基因座处的异源六核苷酸重复扩增序列。

[0262] 本文提供的非人类动物或非人类动物细胞可表现出，例如，一个或多个或所有以下特征：(a) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比，保留内含子序列的C9orf72转录物的表达增加；和/或(b) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比，RNA病灶数量增加；和/或(c) 与包含野生型C9orf72基因座的对照非人类动物或对照非人类动物细胞相比，增加的二肽重复蛋白水平。这种二肽重复蛋白可以是，例如，聚(甘氨酸-丙氨酸)、聚(甘氨酸-脯氨酸)、聚(甘氨酸-精氨酸)、聚(丙氨酸-脯氨酸)，或聚(脯氨酸-精氨酸)二肽重复蛋白（例如，聚GA或聚GP二肽重复蛋白）。

[0263] 本文提供的非人类动物基因组或细胞可以是，例如，包含C9orf72基因座或与人类C9ORF72基因座同源或种间同源的基因组基因座的任何非人类动物基因组或细胞。基因组可以来自，或细胞可以是：真核细胞，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”包括动物界的任何成员，包括例如哺乳动物、鱼类，爬行动物、两栖动物、鸟类和蠕虫。哺乳动物细胞可以是，例如，非人哺乳动物细胞、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、猩猩、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。

[0264] 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。

[0265] 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,细胞可以是神经元细胞(例如ES细胞衍生的运动神经元)、脑细胞、皮质细胞,或心脏细胞。

[0266] 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。

[0267] 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。

[0268] 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。

[0269] 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。

[0270] 如本文所述的,包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物可以通过本文别处所述的方法制备。术语“动物”包括动物界的任何成员,包括例如哺乳动物、鱼类,爬行动物、两栖动物、鸟类和蠕虫。在具体实施例中,非人类动物是非人类哺乳动物。非人哺乳动物包括,例如,非人灵长类动物、猴子、猿、猩猩、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。还包括家养动物和农业动物。术语“非人”不包括人类。优选的非人类动物包括例如啮齿动物,例如小鼠和大鼠。

[0271] 非人类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)。

[0272] 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。

IV. 制备在其基因组中包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物或非人类动物细胞的方法

[0273] 提供了制备如本文其它地方所述的、包含在内源C9orf72基因座处插入的六核苷酸重复扩增序列的动物基因组、动物细胞或动物的各种方法。同样,提供了制备如本文其它地方所述的、包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物基因组、非人类动物细胞或非人类动物的各种方法。

[0274] 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类动物。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。例如,制备包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物可包括:(1) 修饰多能细胞的基因组以包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列;(2) 识别或选择包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列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动物宿主胚胎;以及(4) 将该宿主胚胎植入代孕母体内并孕育。例如,制备包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物可包括:(1) 修饰多能细胞的基因组以包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列;(2) 识别或选择包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列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动物宿主胚胎;以及(4) 在代孕母体中孕育该宿主胚胎。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,从而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。可选地,包含修饰的多能细胞(例如非人ES细胞)的宿主胚胎可在植入代孕母体内并在其中孕育以产生F0非人类动物之前培养至胚泡阶段。然后,该代孕母体可以产生包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物。

[0275] 该方法可以进一步包括识别具有修饰的靶基因组C9orf72基因座的细胞或动物。多种方法可用于识别具有靶向遗传修饰的细胞和动物。

[0276] 替代地,产生本文别处描述的非人类动物的方法可包括:(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,以包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列;(2) 识别遗传修饰的胚胎;以及(3) 将该遗传修饰的胚胎植入代孕母体中并在该代孕母体中孕育该遗传修饰的胚胎。替代地,产生本文别处描述的非人类动物的方法可包括:(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,以包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列;(2) 识别遗传修饰的胚胎;以及(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。

[0277] 核转移技术也可以用于产生非人哺乳动物。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法

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构细胞。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。重构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。

[0278] 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动物的细胞包含在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列。人们认识到,根据用于产生F0动物的方法,F0动物内具有在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的细胞数量会有所不同。通过例如VELOCIMOUSE®方法,将供体ES细胞引入来自相应的生物体的桑椹前阶段胚胎(例如8细胞期的小鼠胚胎),更大百分比的F0动物的细胞群能够包括具有感兴趣的核苷酸序列的细胞,该感兴趣的核苷酸序列包含靶向遗传修饰。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动物的细胞贡献可以包含具有靶向修饰的细胞群。

[0279] 遗传修饰的F0动物的细胞对于在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列可以是杂合的,或者在内源C9orf72基因座处插入的异源六核苷酸重复扩增序列可以是纯合的。

V. 使用在其基因组中包含于内源C9orf72基因座处插入的异源六核苷酸重复扩增序列的非人类动物或非人类动物细胞方法

[0280] 提供了多种方法,以利用如本文其它地方所公开的、包含在内源C9orf72基因座处插入的六核苷酸重复扩增序列的动物细胞或动物识别或评估用于治疗与C9orf72基因座处的六核苷酸重复扩增序列相关的疾病或病症的治疗候选物。同样,提供了多种方法,以利用如本文其它地方所公开的、包含在内源C9orf72基因座处插入的六核苷酸重复扩增序列的非人类动物细胞或非人类动物识别或评估用于治疗与异源C9orf72基因座处的六核苷酸重复扩增序列相关的疾病或病症的治疗候选物。

[0281] 此类方法可包括,例如,向非人类动物细胞或非人类动物施用候选药剂,进行一个或多个测定,以确定候选药剂是否对与疾病或病症相关的一个或多个相关体征或症状有影响,并且如果候选药剂对与疾病或病症相关的一个或多个体征或症状有影响,则将候选药剂识别为治疗候选物。

[0282] 疾病或病症可以是与C9orf72基因座处的重复扩增相关的神经变性疾病,例如肌萎缩侧索硬化(ALS)或额颞叶痴呆(FTD)。

[0283] 可以测试任何候选药剂。此类候选物可包括例如大分子(例如siRNA、抗体或CRISPR/Cas gRNA)或小分子。可以通过任何合适的途径以任何方式将候选药剂施用于非人类动物或非人类动物细胞。

[0284] 可以使用测量与疾病或病症相关的体征或症状的任何测定。作为第一个例子,体

征或症状可以是含有内含子的C9orf72 RNA转录本的表达。测量含有内含子的C9orf72 RNA转录物表达的测定的一个例子是定量聚合酶链反应(qPCR),其使用在C9orf72基因座的外显子1A和1B或外显子1B和2之间的的内含子区域内杂交的引物和探针。含内含子的C9orf72转录物的增加的表达与疾病病理有关(例如,在ALS患者中),并在本文所述的细胞中观察到,因此降低含内含子的C9orf72转录物表达的候选药剂可能是治疗候选物。

[0285] 作为第二个例子,体征或症状可以是包含C9orf72有义或反义RNA转录物的RNA病灶(例如核仁病灶)的存在。例如,可以通过荧光原位杂交来测量这样的RNA病灶。增加的包含有义或反义C9orf72 RNA的细胞质和/或核病灶(例如核仁)与疾病病理相关(例如在ALS患者中)并且在本文所述的细胞中观察到,因此减少RNA病灶的存在的候选药剂可以成为治疗候选物。

[0286] 作为另一个例子,体征或症状可以是二肽重复蛋白的存在或积累,例如聚(甘氨酸-丙氨酸)、聚(甘氨酸-脯氨酸)、聚(甘氨酸-精氨酸)、聚(丙氨酸-脯氨酸)或聚(脯氨酸-精氨酸)二肽重复蛋白。例如,这种二肽重复蛋白可以是聚GA二肽重复蛋白或聚GP二肽重复蛋白。二肽重复蛋白的存在或积累可以通过例如免疫组织化学、免疫荧光或通过蛋白质狭缝印迹来测量。RAN翻译产物(例如二肽重复蛋白)的积累与疾病病理学(例如,在ALS患者中)相关并且在本文所述的细胞中观察到,因此减少二肽重复的存在的候选药剂可以是治疗候选物。

[0287] 候选药剂可以在体内施用于非人类动物,并且可以在非人类动物中进行一个或多个测定。或者,可将候选药剂体内施用于非人类动物,且可在施用候选药剂后在从非人类动物分离的细胞中体外进行一个或多个测定。或者,候选药剂可以体外施用于细胞(例如非人胚胎干细胞衍生的运动神经元),并且可以在细胞中体外进行测定。此类细胞可以是例如胚胎干细胞、源自胚胎干细胞的运动神经b元、脑细胞、皮质细胞、神经元细胞、肌肉细胞或心脏细胞。

[0288] 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。尽管为了清楚和理解的目的已经通过说明和示例的方式对本发明进行了一些详细的描述,但是很明显可以在所附权利要求的范围内实施某些改变和修改。

序列简述

[0289] 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。

[0290] 表3:序列描述

SEQ ID NO	类型	描述
1	DNA	GGGGCC C9ORF72 六核苷酸重复
2	DNA	CAG 重复
3	DNA	CGG 重复
4	DNA	CTG 重复
5	DNA	GAA 重复
6	DNA	GCC 重复
7	DNA	GCG 重复
8	DNA	CCTG 重复
9	DNA	ATTCT 重复
10	DNA	TGGAA 重复
11	DNA	GGCCTG 重复
12	DNA	CCCCGCCCCGCG 重复
13	DNA	智人 (<i>Homo sapiens</i>) C9ORF72 mRNA (NM_145005.6)
14	蛋白质	智人 (<i>Homo sapiens</i>) C9ORF72 Amino Acid (NP_659442.2)
15	DNA	智人 (<i>Homo sapiens</i>) C9ORF72 mRNA (NM_018325.4)
16	蛋白质	智人 (<i>Homo sapiens</i>) C9ORF72 Amino Acid (NP_060795.1)
17	DNA	智人 (<i>Homo sapiens</i>) C9ORF72 mRNA (NM_001256054.2)
18	蛋白质	智人 (<i>Homo sapiens</i>) C9ORF72 Amino Acid (NP_001242983.1)
19	RNA	crRNA 尾
20	RNA	TracrRNA
21	RNA	gRNA 支架 v1
22	RNA	gRNA 支架 v2
23	RNA	gRNA 支架 v3
24	RNA	gRNA 支架 v4
25	DNA	通用向导 RNA 靶序列加 PAM v1
26	DNA	通用向导 RNA 靶序列加 PAM v2
27	DNA	通用向导 RNA 靶序列加 PAM v3
28	DNA	5' gRNA-1 靶序列
29	DNA	5' gRNA-2 靶序列
30	DNA	5' gRNA-3 靶序列
31	DNA	5' gRNA-4 靶序列
32	DNA	3' gRNA-1 靶序列
33	DNA	3' gRNA-2 靶序列
34	DNA	3' gRNA-3 靶序列
35	DNA	3' gRNA-4 靶序列
36	DNA	2-引物正向 (Fwd)
37	DNA	2-引物反向 (Rev)
38	DNA	锁核酸 (LNA) 有义 RNA 探针
39	DNA	锁核酸 (LNA) 反义 RNA 探针

SEQ ID NO	类型	描述
40	DNA	小家鼠 (<i>Mus musculus</i>) <i>C9orf72</i> mRNA (NM_001081343.1)
41	蛋白质	小家鼠 (<i>Mus musculus</i>) C9ORF72 氨基酸 (NP_001074812.1)
42	DNA	褐家鼠 (<i>Rattus norvegicus</i>) <i>C9orf72</i> mRNA (NM_001007702.1)
43	蛋白质	褐家鼠 (<i>Rattus norvegicus</i>) C9ORF72 氨基酸 (NP_001007703.1)
44	DNA	DNA 有义探针
45	DNA	DNA 反义探针
46	DNA	人类 5'侧翼序列
47	DNA	人类 3'侧翼序列
48	DNA	图 2 中的序列 1
49	DNA	图 2 中的序列 2
50	DNA	图 2 中的序列 3
51	DNA	图 2 中的序列 4, 即 tcagtcagt
52	DNA	来自图 3 的 gRNA_5side_repeat_2
53	DNA	来自图 3 的 gRNA_3side_repeat_1
54	DNA	来自图 3 的 gRNA_3side_repeat_3
55	DNA	来自图 3 的 gRNA_3side_repeat_4
56	RNA	5' gRNA-1 crRNA
57	RNA	5' gRNA-2 crRNA
58	DNA	图 3 中的 5'靶位点
59	DNA	图 3 中的 3'靶位点
60	DNA	图 11A 和 11B 中的序列
61	DNA	分析 A - 正向引物
62	DNA	分析 A - 反向引物
63	DNA	分析 A - 探针
64	DNA	分析 B - 正向引物
65	DNA	分析 B - 反向引物
66	DNA	分析 B - 探针
67	DNA	分析 D - 正向引物
68	DNA	分析 D - 反向引物
69	DNA	分析 D - 探针
70	DNA	分析 G - 正向引物
71	DNA	分析 G - 反向引物
72	DNA	分析 G - 探针
73	DNA	分析 H - 正向引物
74	DNA	分析 H - 反向引物
75	DNA	分析 H - 探针
76	RNA	5' gRNA-3 crRNA
77	RNA	5' gRNA-4 crRNA
78	RNA	3' gRNA-1 crRNA
79	RNA	3' gRNA-2 crRNA
80	RNA	3' gRNA-3 crRNA
81	RNA	3' gRNA-4 crRNA
82	RNA	TracrRNA v2
83	RNA	crRNA 尾 v2
84	RNA	5'-gRNA-1 指导序列
85	RNA	5'-gRNA-2 指导序列
86	RNA	5'-gRNA-3 指导序列

SEQ ID NO	类型	描述
87	RNA	5'-gRNA-4 指导序列
88	RNA	3'-gRNA-1 指导序列
89	RNA	3'-gRNA-2 指导序列
90	RNA	3'-gRNA-3 指导序列
91	RNA	3'-gRNA-4 指导序列
92	DNA	分析 C - 正向引物
93	DNA	分析 C - 反向引物
94	DNA	分析 C - 探针
95	DNA	分析 E - 正向引物
96	DNA	分析 E - 反向引物
97	DNA	分析 E - 探针
98	DNA	分析 F - 正向引物
99	DNA	分析 F - 反向引物
100	DNA	分析 F - 探针
101	DNA	分析 I - 正向引物
102	DNA	分析 I - 反向引物
103	DNA	分析 I - 探针
104	DNA	分析 J - 正向引物
105	DNA	分析 J - 反向引物
106	DNA	分析 J - 探针
107	DNA	图 6C 中的亲本克隆
108	DNA	图 6C 中的扩增克隆 (expanded clone)
109	DNA	图 17C 中的参考序列和 WT-Cas9 扩增克隆 1
110	DNA	图 17C 中的 WT-Cas9 扩增克隆 2
111	DNA	图 17C 中的 WT-Cas9 扩增克隆 3
112	DNA	图 17C 中的 WT-Cas9 扩增克隆 4
113	DNA	图 17C 中的 WT-Cas9 扩增克隆 5
114	DNA	图 17C 的 gRNA 靶序列加 PAM
115	DNA	图 17C 中的参考序列和切口酶扩增克隆
116	RNA	TracrRNA v3
117	RNA	TracrRNA v4
118	RNA	gRNA 支架 v5
119	RNA	gRNA 支架 v6
120	RNA	gRNA 支架 v7
121	DNA	测序引物

实施例

实施例1.C9orf72基因座上的六核苷酸重复扩增

[0291] 肌萎缩侧索硬化(ALS)是一种进行性神经退行性疾病,其导致运动神经元死亡,进而导致瘫痪。在美国,每年有五千人被诊断出患有ALS。10%的该疾病在家族中传播,在这一类别中,最常见的原因是C9ORF72基因座上的GGGGCC(SEQ ID NO:1)六核苷酸重复扩增。健康的个体通常具有小于30个重复,而受影响的患者通常具有超过1000个拷贝的该重复。由于缺乏有用的动物模型,这种长串的重复导致疾病的确切机制尚未阐明。六核苷酸重复序列中的高GC含量导致难以合成具有该重复的DNA片段,以及难以在微生物中保持该重复。因此,制备靶向载体等材料以产生转基因动物是非常具有挑战性的。这是迄今为止很少有用于C9ORF72重复扩增类型的ALS的有用动物模型的主要原因之一,并且该领域迫切需要更好的动物模型。

[0292] 为了绕过这些用于靶向载体构建的困难的中间步骤,我们采取的方法是扩增已经

插入基因组中正确位置的相对较短的六核苷酸重复,而非从头进行靶向载体生产和ES细胞靶向。作为重复扩增的起始材料,我们使用了杂合ES细胞克隆,其中小鼠C9orf72基因座的一部分被包含92x重复的GGGGCC (SEQ ID NO:1)六核苷酸的人类对应物替换。参见图1。我们之前开发了这个等位基因。参见US 2018/0094267和WO 2018/064600,出于所有目的,每篇都通过引用整体并入本文。

[0293] 当哺乳动物细胞染色体中发生DNA双链断裂(DSB)时,细胞识别DSB并通过同源重组(HR)或非同源重组途径(如非同源末端连接(NHEJ)或单链退火(SSA))修复损伤。在HR中,受损的染色单体利用未受损的染色单体作为其修复的模板。在这条通路中,一旦发生DSB,DSB的边缘会通过许多过程(包括3'链切除)进行修改。暴露的5'单链被多个Rad51分子包覆,然后Rad51细丝开始在基因组中寻找同源序列。Rad51细丝能够将自身插入双链DNA,其中暴露的5'单链具有同源性。侵入的5'链使用Rad51细丝所进入的序列启动恢复。Rad51细丝找到同源序列的准确性是保留正确DNA序列的关键。我们扩增重复的方法是基于这样的假设:通过在六核苷酸重复附近引入DSB以使暴露的5'链仅包含重复序列,Rad51细丝可能进入错误的位置,我们可能会误导Rad51细丝的同源性搜索,导致重复的扩增或收缩。见图2。

[0294] 具体而言,本文使用的方法是:通过使用CRISPR/Cas9核酸酶在六核苷酸重复的末端附近引入DSB来扩增人源化小鼠胚胎干细胞中C9orf72内含子1处的现有GGGGCC (SEQ ID NO:1)六核苷酸重复。几个向导RNA(gRNA)被设计在尽可能靠近六核苷酸重复序列的位置,以在切除后暴露5'单链上的重复序列。参见图3和表4。

[0295] 表4.C9ORF72 gRNA靶序列。

向导RNA	靶序列 (NGG PAM 的上游)	SEQ ID NO
5' gRNA-1	GCTATGCGATCGCCGTCTCG	28
5' gRNA-2	CCCCGGCCCCGGCCCCGAGA	29
5' gRNA-3	CGCCGTCTCGGGGCGGGGC	30
5' gRNA-4	CCGTCTCGGGGCGGGGCCG	31
3' gRNA-1	CGGCCGGCCCTCGAGGGTCT	32
3' gRNA-2	GGCCGGGGCCGAGACCCTCG	33
3' gRNA-3	GAGGGTCTCGGCCCGGCC	34
3' gRNA-4	CTCGGCCCGGCCCGGCC	35
向导RNA	靶序列	SEQ ID NO
5' gRNA-1	GCUAUGCGAUCGCCGUCUCG	84
5' gRNA-2	CCCCGGCCCCGGCCCCGAGA	85
5' gRNA-3	CGCCGUCUCGGGGCCGGGGC	86
5' gRNA-4	CCGUCUCGGGGCCGGGGCCG	87
3' gRNA-1	CGGCCGGCCUCGAGGGUCU	88
3' gRNA-2	GGCCGGGGCCGAGACCUCG	89
3' gRNA-3	GAGGGUCUCGGCCCCGGGCC	90
3' gRNA-4	CUCGGCCCCGGCCCCGGGCC	91
向导RNA	crRNA序列	SEQ ID NO
5' gRNA-1	GCUAUGCGAUCGCCGUCUCGGUUUUAGAGCUAUGCUGUUUUG	56
5' gRNA-2	CCCCGGCCCCGGCCCCGAGAGUUUUAGAGCUAUGCUGUUUUG	57
5' gRNA-3	CGCCGUCUCGGGGCCGGGGCGUUUUAGAGCUAUGCUGUUUUG	76
5' gRNA-4	CCGUCUCGGGGCCGGGGCCGUUUUAGAGCUAUGCUGUUUUG	77
3' gRNA-1	CGGCCGGCCUCGAGGGUCUUUUAGAGCUAUGCUGUUUUG	78
3' gRNA-2	GGCCGGGGCCGAGACCUCGGUUUUAGAGCUAUGCUGUUUUG	79
3' gRNA-3	GAGGGUCUCGGCCCCGGCCCCGUUUUAGAGCUAUGCUGUUUUG	80
3' gRNA-4	CUCGGCCCCGGCCCCGGCCCCGUUUUAGAGCUAUGCUGUUUUG	81
向导RNA	TracrRNA序列	SEQ ID NO
全部	GUUGGAACCAUUCAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUC CGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU	82

[0296] 目标是获得在尽可能接近重复的地方切割并且具有高切割效率的gRNA。使用质粒(8028Stvec)在体外测试DNA切割效率,该质粒包含与人源化的、含92x重复的等位基因相同的序列。从重复序列的5'侧和3'侧选择具有最高切割效率的gRNA。这些是5'gRNA-1和3'gRNA-2。见图4。

[0297] 在第一个实验中,我们通过电穿孔将SpCas9(化脓链球菌Cas9;Thermo Fisher)和gRNA作为RNA和蛋白质复合物(RNP)引入C9orf72人源化、含30x重复的ES细胞。由于引物-PCR不能提供精确的序列信息,而且多于~90x的重复太长而无法通过Sanger测序进行测序,因此使用30x重复等位基因进行了本实验,以通过扩增30x重复,表征重复区域周围的DNA序列和重复的精确数量。

[0298] 我们分别使用5'gRNA-1和3'gRNA-2测试了在该重复的5'侧引入单个DSB或在该重复的3'侧引入单个DSB。参见图5。其中的C9orf72被部分人源化并插入30x GGGGCC重复的小鼠ES细胞在VG ES培养基中生长。细胞通过电穿孔接受Cas9和gRNA RNPs。挑取ES细胞集落,集落在96孔培养板中生长,并纯化基因组DNA用于分析。然后对C9orf72基因座进行常规的双引物PCR,以评估在30x重复扩增序列的5'末端附近切割后或在30x重复扩增序列的3'末端附近切割后重复区域的大小。重复大小主要使用2-引物PCR,在琼脂糖凝胶上跑扩增子进行评估。我们挑选了扩增(>30x)、保留(~30x)、收缩(<30x)和折叠(~3x)的一些克隆,并对包含GGGGCC重复区域的PCR扩增子进行测序。我们确认重复数发生了变化,正如琼脂糖凝胶分析所预期的那样。我们还发现,在大多数克隆中,即使重复大小与琼脂糖凝胶上的原始30x相同,Cas9处理也会产生1-15bp的删除。结果示于图6A-9B。如图6A所示,使用5'gRNA-1的5'末端重复上游(5')的双链断裂触发了所选克隆中从30x到42x的重复扩增。参见图6B。扩增克隆与亲本30x克隆序列相比的序列如图6C所示。该实验表明,Cas9的双链断裂可以不间断地扩增简单的重复序列。这可能发生在双链断裂位点有或没有小删除的情况下,其可破坏向导RNA靶序列。在一些克隆中,观察到小的删除(例如,1-16bp)。

[0299] 如图7A和7B所示,重复序列5'末端上游的双链断裂也可以触发部分重复收缩而不中断重复。同样,这可能发生在双链断裂位点有或没有小删除的情况下。图7A和7B中示出的克隆从30个重复收缩到20个重复。

[0300] 如图8A和8B所示,一些克隆在重复序列5'末端上游切割后保留相同的重复数。同样,这可能发生在双链断裂位点有或没有小删除的情况下。

[0301] 当使用3'gRNA-2在重复序列的3'末端下游(3')引入双链断裂时,我们经常观察到大的删除。两个这样的克隆示于图9A和9B。在一个克隆中,在重复的下游观察到10-20bp的删除,并且该重复从30x收缩到1-3x。在另一个克隆中,重复从30x收缩到10-20x,并且观察到一个大于500bp的大删除,其包括部分该重复和下游序列(重复的3')。

[0302] 总的来说,我们观察到重复扩增效率取决于双链断裂的位点。重复序列的5'末端上游(5')的双链断裂比重复序列3'末端下游(3')的双链断裂更频繁地诱导重复扩增。表5示出了来自30x亲本克隆的数据总结。

[0303] 表5. 30x重复亲本系中重复扩张或收缩的评估。

重排	单个5' DSB (%)	单个3' DSB (%)
扩增	4.2	1.1
部分收缩	8.3	2.3

保留	50.0	21.6
完全收缩	0	61.4
没有检测到	37.5	13.6
总和	100	100

[0304] 我们接下来在含有92x重复序列的ES细胞中测试了相同的東西。为了修饰ES细胞中的DNA,通过电穿孔将SpCas9(化脓链球菌Cas9;Thermo Fisher)和gRNA作为RNA和蛋白质复合物(RNP)引入C9orf72人源化的、含有92x重复序列的ES细胞(MAID8029a)。挑取ES细胞集落,在96孔培养板中生长,并纯化基因组DNA用于分析。进行了三个实验:(1)在重复的5'侧引入单个DSB;(2)在重复的3'侧引入单个DSB;(3)在重复序列的5'和3'侧引入DSB。参见图10。使用传统的双引物PCR(图11A)和使用AmplideX PCR/CE C9orf72试剂盒的三个引物的prime PCR(图11B)进行分析。对于primePCR,来自ES细胞的基因组DNA用作模板。使用了三种引物:位于重复的5'外部的引物、位于重复的3'外部的引物,以及与重复内部的重复序列退火的第三个引物,如图11B所示。该PCR反应产生许多不同大小的PCR产物,如图11B所示。因为第三个引物可以在任何给定的重复单元中引发聚合酶反应,所以获得的PCR产物的数量基本上就是克隆具有的重复的数量。我们用毛细管电泳跑这些PCR产物,它检测荧光信号(其中一个引物是被荧光标记的)。

[0305] 根据制造商的说明使用AmplideX PCR/CE C9ORF72试剂盒(Asuragen),以确认内源C9orf72 ES细胞克隆中如SEQ ID NO:1所示的六核苷酸序列的数量。使用从3x重复克隆、92x重复克隆和30x重复克隆中纯化的mESC基因组总DNA作为对照。使用表6中的引物和来自AmplideX PCR/CE C9ORF72试剂盒的重复特异性引物在ABI 9700热循环仪(Thermo Fisher)上进行PCR。在ABI 3500xL GeneScan上使用POP-7聚合物(Thermo Fisher)和NuSieve琼脂糖凝胶(Lonza)通过毛细管电泳确定扩增子的大小。将2-log DNA阶梯(New England BioLabs)分子量标记加载到琼脂糖凝胶上进行比较,并使用SYBR Gold Nucleic Acid Stain(Thermo Fisher)观察条带。

[0306] 图6.PCR的引物。

引物名称	序列	SEQ ID NO:
2-引物正向	TGCGCCTCCGCCCGCGGGCGCAGGCACCGCAACCGCA	36
2-引物反向	CGCAGCCTGTAGCAAGCTCTGGAAGCTCAGGAGTCG	37

[0307] 如使用常规双引物PCR所证明的,在重复的任一側引入DSB导致重复区域的大小改变。参见图12A和12B。图12A和12B中示出了每种条件下88个菌落中的16个的结果。选择具有最大扩增的5'单个DSB克隆和3'单个DSB克隆进行进一步分析,使用prime PCR确认重复扩增。Prime PCR证实克隆9253D-B2(5'单个DSB)有145个重复,克隆9253A-B1(3'单个DSB)有130个重复和92个重复。参见图13和14。图13示出了毛细管电泳的结果。信号强度在Y轴上,PCR产物大小在X轴上。读数是峰值的数量。上面板来自92x亲本克隆,有92个峰。我们计算了中间面板中的150个峰值和底部面板中的130x个峰值。我们在底部面板中观察到两个高峰。这可能是因为我们挑选的ES细胞集落不是同质克隆(homogenous clone)。

[0308] 再次重复相同的实验,结果相似。参见图15。对于每种条件(单个5' DSB、单个3' DSB,或5'和3' DSB),测试了88个克隆。重复处的DSB导致重复不稳定。重复的任一側或两侧的DSB导致重复扩增(从92x到~300x)。表7中提供了对不同克隆中的扩增或收缩的完整分

析。通过TAQMAN测定或AsuragenAmplideX PCR/CE C90RF72试剂盒称为(called)完全收缩。总的来说,我们观察到重复扩增效率取决于亲本克隆中的起始重复长度。无论双链断裂是在重复序列的5'侧上游诱导还是在重复序列的3'侧下游诱导,在具有更大起始重复长度的亲本克隆中更频繁地观察到重复扩增。见表6。

[0309] 表7. 第一次重复扩增后对重复扩增或收缩的评估。

亲本系	92x 重复			30x 重复	
	5' & 3' DSB (%)	单个 5' DSB (%)	单个 3' DSB (%)	单个 5' DSB (%)	单个 3' DSB (%)
扩增 (>750bp = 92x)	4.5	18.2	5.7	4.2	1.1
部分收缩(<750bp, >200bp)	9.1	9.1	9.1	8.3	2.3
保留 (750bp)	19.3	51.1	5.7	50.0	21.6
完全收缩 (<200bp)	46.6	0	53.4	0	61.4
没有检测到	20.5	21.6	26.1	37.5	13.6
总和	100	100	100	100	100

[0310] 最大的重复是通过一起引入5'和3'DSB获得的。通过引入单个DSB获得的最大重复约为250X,其中3'gRNA位点仍然完整且重复可扩增。为了进一步扩增该重复,我们从250x重复克隆(MAID9253D-C5)开始,并如上所述通过引入Cas9/gRNA RNP在3'末端引入DSB。使用250x等位基因,我们能够将该重复进一步扩增到~600x的大小。参见图16A和16B以及表8。

[0311] 表8. 对使用单个3'DSB的第二次重复扩增后的重复扩增或收缩的评估。

亲本克隆	8029, 92x (%)	9253D-C5, 250x (%)
扩增(>亲本克隆)	5.6	13.6
部分收缩(<亲本克隆, >200bp)	9.1	43.2
保留(=亲本克隆)	5.6(28.4)*	30.7(43.2)*
完全收缩(<200bp)	55.8	4.5
没有检测到	27.3	2.3
总和	100	100

*观察到多个条带

[0312] 总而言之,在C90RF72六核苷酸重复序列的边缘诱导双链断裂触发了小鼠ES细胞中的重复扩增。当双链断裂在5'侧或3'侧引入时以不同的效率发生重复扩增。此外,当起始重复大小较大时,双链断裂诱导的重复扩增发生得更频繁。只要gRNA靶序列完整,重复至少可以扩增两次,我们能够以这种方式将C90RF72六核苷酸重复序列从92x扩增到大约600x。

[0313] 我们接下来评估单链断裂是否足以驱动重复扩增。为了修饰ES细胞中的DNA,将购自Integrated DNA Technologies的SpCas9 D10A和gRNA(3'gRNA-2)作为RNA和蛋白质复合物(RNP)通过电穿孔引入C9orf72人源化的、含有92x重复序列的ES细胞(MAID8029a)。挑取ES细胞集落,在96孔培养板中培养,提纯基因组DNA进行分析。使用如上所述的常规双引物PCR进行分析。实验配置如图17A所示,PCR结果如图17B所示。与双链断裂一样,重复区域3'末端下游(3')的单链断裂能够触发重复扩增。此外,与双链断裂不同,单链断裂不会引起任何大的删除(数据未显示)。然后我们使用单链断裂下游(3')的引物(SEQ ID NO:36中列出的引物或SEQ ID NO:121中列出的引物)进行测序PCR。如图17C所示,在使用Cas9核酸酶产生的5个为了造成双链断裂的重复扩增克隆中,4个在向导RNA靶序列和/或原始间隔区相邻模体(PAM)中存在删除。与此相对,在使用Cas9切口酶产生的8个为了造成单链断裂的重复

扩增克隆中,没有一个在向导RNA靶序列和/或PAM中存在删除。因此,向导RNA靶序列可以重新用于第二轮重复扩增。

[0314] 我们接下来评估了双链断裂是否可以驱动小鼠单细胞阶段胚胎中的重复扩增。SpCas9和gRNA (5' gRNA-1) 通过原核注射作为RNA和蛋白质复合物 (RNP) 引入C9orf72人源化、含有92x重复序列的小鼠单细胞阶段胚胎。从单细胞期胚胎产生小鼠,并收集尾基因组DNA。使用如上所述的常规双引物PCR进行分析。实验设置如图18A所示,PCR结果如图18B所示。与小鼠ES细胞一样,重复区域5'末端上游(5')的双链断裂能够触发重复扩增。在图18B所示的一个克隆中,重复区域从92个重复扩增到大约150个重复。

实施例2. 三核苷酸重复扩增

[0315] C9ORF72基因座上的六核苷酸重复扩增只是神经系统疾病中发生的一种重复扩增。已知重复扩增发生在其它神经系统疾病的其它基因中。这些包括,例如,三核苷酸重复、四核苷酸重复、五核苷酸重复、其它六核苷酸重复和十二核苷酸重复。

[0316] 我们接下来测试了双链断裂诱导的重复扩增是否可以发生在具有不同重复序列的不同基因座上。具体而言,我们测试了非C9ORF72的第二个靶基因中三核苷酸重复序列的扩增。亲本小鼠ES细胞克隆在靶基因中含60个重复的三核苷酸序列。我们设计了一个向导RNA来引导Cas9蛋白在重复序列的5'末端上游(5')16个碱基对的位置切割,我们设计了一个向导RNA来引导Cas9蛋白在重复序列的3'末端下游(3')11个碱基对的位置切割。通过电穿孔将SpCas9和5' gRNA作为RNA和蛋白质复合物 (RNP) 引入含有60x重复序列的ES细胞中。使用如上所述的常规双引物PCR进行分析。PCR结果如图19所示。与C9ORF72六核苷酸重复扩增一样,60x三核苷酸重复区域5'末端上游(5')的双链断裂能够触发重复扩增。在图19所示的一个克隆中,重复区域从60个重复扩大到85个重复。这表明双链断裂诱导的重复扩增可以发生在多个基因位点,并发生在多种不同类型的重复中。

[0317] 在不同位点、不同细胞中、用不同的起始重复数量、使用不同的Cas9蛋白(核酸酶对切口酶)、使用靶向不同链的不同向导RNA、使用不同Cas9切割位点位置,以及使用与重复之间的不同切割位点距离进行的多个实验的所有结果总结于表9。

[0318] 表9. 重复扩增频率。

基因	细胞	起始重复#	Cas9	gRNA	Cas9 切割位点位置	切割位点与重复的距离	扩增频率*	平均扩增长度**	最大的尺寸变化
C9orf72	ES 细胞	250x	WT	有义	3'	7 bp	13%	10%	~ 600x
C9orf72	ES 细胞	92x	WT	有义	3'	7 bp	5-10%	10%	~ 150x
C9orf72	ES 细胞	92x	D10A	有义	3'	7 bp	10%	5%	~ 150x
C9orf72	ES 细胞	92x	WT	有义	5'	2 bp	8-20%	10%	~ 250x
C9orf72	ES 细胞	92x	WT	有义	5'	100 bp	0%	N/A	N/A
C9orf72	ES 细胞	92x	WT	有义	5'	1000 bp	0%	N/A	N/A
C9orf72	ES 细胞	92x	D10A	有义	5'	2 bp	10%	5%	~ 150
C9orf72	ES 细胞	30x	WT	有义	3'	7 bp	0%	N/A	N/A
C9orf72	ES 细胞	30x	WT	有义	5'	2 bp	4-5%	30%	60x
C9orf72	ES 细胞	30x	D10A	有义	5'	2 bp	0%	N/A	N/A
C9orf72	胚胎	92x	WT	有义	5'	2 bp	1-2%***	50%	150x
C9orf72	胚胎	92x	WT	有义	3'	7 bp	少于 1%	10%	100x
三核苷酸重复靶基因	ES 细胞	60x	WT	有义	5'	16 bp	2-3%	33%	85x
三核苷酸重复靶基因	ES 细胞	60x	WT	反义	3'	11 bp****	0%	N/A	N/A

*观察到重复扩增的频率(#扩增的克隆/#筛选的克隆) x100

**扩增时的平均尺寸变化(平均扩增尺寸/起始尺寸) x100

***高度马赛克

***高度重复的区域

实施例3. 包含C9orf72基因座上六核苷酸重复扩增的运动神经元或脑组织的分析

[0319] 小鼠中六核苷酸重复大小的稳定性使用如上所述的AmplideX PCR/CE C9ORF72试剂盒(Asuragen)确认。

[0320] 如US 2018/0094267和WO 2018/064600中所述(每篇都出于所有目的通过引用整体并入本文),检查了小鼠胚胎干细胞衍生的运动神经元(ESMN)中的RNA转录物,其包含野生型C9orf72基因座(对照)或来自实施例1和2的遗传修饰的C9orf72基因座。在ESMN中评估RNA病灶和二肽重复蛋白水平,所述ESMN源自亲代胚胎干细胞,该亲代胚胎干细胞包含野生型C9orf72基因座(对照)或来自实施例1和2的遗传修饰的C9orf72基因座。材料和方法如下所述。

[0321] 此外,如US 2018/0094267和WO 2018/064600中所述(每篇都出于所有目的通过引用整体并入本文),检查了小鼠脑组织和亲本胚胎干细胞中的RNA转录物,其包含野生型C9orf72基因座(对照)或来自实施例1和2的遗传修饰的C9orf72基因座。在ESMN中评估RNA病灶和二肽重复蛋白水平,所述ESMN源自亲代胚胎干细胞,该亲代胚胎干细胞包含野生型C9orf72基因座(对照)或来自实施例1和2的遗传修饰的C9orf72基因座。材料和方法如下所述。

胚胎干细胞衍生的运动神经元

[0322] 来自实施例1的胚胎干细胞(ESC)在胚胎干细胞培养基(ESM;DMEM+15%胎牛血清+青霉素/链霉素+谷氨酰胺+非必需氨基酸+核苷+β-巯基乙醇+丙酮酸钠+LIF)中培养2天,在此期间每天更换培养基。在胰酶消化前1小时,将ES培养基更换为7mL ADFNK培养基(高级DMEM/F12+神经基础培养基+10%基因敲除血清+青霉素/链霉素+谷氨酰胺+β-巯基乙醇)。

吸出ADFNK培养基,用0.05%胰蛋白酶-EDTA对ESC进行胰酶消化。将沉淀的细胞重新悬浮在12mL的ADFNK中,并在悬浮液中生长两天。将细胞在补充有视黄酸(RA)、平滑激动剂和嘌呤吗啡胺的ADFNK中再培养4天,以获得肢体样运动神经元(ESMN)。分离的运动神经元在胚胎干细胞衍生的运动神经元培养基(ESMN;神经基础培养基+2%马血清+B27+谷氨酰胺+青霉素/链霉素+ β -巯基乙醇+10ng/mL GDNF、BDNF、CNTF)中铺板和成熟。

定量聚合酶链反应

[0323] 使用位于各个区域侧翼的引物和检测修饰的C9orf72基因座的那些区域的探针,提取来自每个样品的总RNA并进行逆转录。可检测区域包括跨越小鼠和人类序列连接处、仅人类序列或仅小鼠序列的区域。使用现成试剂盒的探针和引物对GAPDH、DROSHA或 β 2-微球蛋白进行qPCR。

[0324] 具体地,从包含野生型(WT)C9orf72基因座(对照)或遗传修饰的C9orf72基因座的胚胎干细胞衍生的运动神经元(ESMN)分离RNA。在其它实验中,还从包含野生型(WT)C9orf72基因座(对照)或遗传修饰的C9orf72基因座的胚胎干细胞的小鼠分离的亲本胚胎干(ES)细胞或总脑中分离RNA。

[0325] 根据制造商的方案(Zymo Research),使用Direct-zol RNA Miniprep plus试剂盒分离总RNA。根据制造商的方案(Invitrogen),使用Turbo DNA-free试剂盒用DNase处理总RNA,并稀释至20ng/ μ L。使用Quantitect Probe RT-PCR试剂盒(Qiagen),在一步反应中进行逆转录(RT)和PCR。qRT-PCR反应包含2 μ L RNA和8 μ L混合物,该混合物包含RT-PCR Master mix、ROX染料、RT-mix和20X基因特异性引物-探针混合物,最终体积为10 μ L。

[0326] 除非另有说明,最终引物和探针浓度分别为0.5 μ M和0.25 μ M。qRT-PCR在ViiATM7实时PCR检测系统(ThermoFisher)上进行。PCR反应一式四份进行,RT步骤在45 $^{\circ}$ C下10分钟,然后是95 $^{\circ}$ C下10分钟和两步循环95 $^{\circ}$ C下5秒,以及60 $^{\circ}$ C下30秒进行45个循环(在光学384孔板中)。表10中提供了用于每次分析(A、B、C、D、E、F、G、H、I和J)的引物和探针的序列以及SEQ ID NO。

[0327] 表10. 引物和探针。

分析A	
正向引物	CATCCCAATTGCCCTTTCC (SEQ ID NO: 61)
反向引物	CCCACACCTGCTCTTGCTAGA (SEQ ID NO: 62)
探针	TCTAGGTGAAAGTGGG (SEQ ID NO: 63)
分析B	
正向引物	GAGCAGGTGTGGGTTTAGGA (SEQ ID NO: 64)
反向引物	CCAGGTCTCACTGCATTCCA (SEQ ID NO: 65)
探针	ATTGCAAGCGTTCGGATAATGTGAGA (SEQ ID NO: 66)
分析C	
正向引物	GATAGTCGACATCCCTGCATC (SEQ ID NO: 92)
反向引物	GGTGGCGAGTGGCTATTG (SEQ ID NO: 93)
探针	AAGCGTTCGGATAATGTGAGACCTGG (SEQ ID NO: 94)
分析D	
正向引物	GCTGTACACGAAGGCTTCTC (SEQ ID NO: 67)
反向引物	GCACTGCTGCCAACTACAAC (SEQ ID NO: 68)
探针	TCAATGCCATCAGCTCACACCTGC (SEQ ID NO: 69)
分析E	
正向引物	TCTCACAGTACTCGCTGAGGGTGA (SEQ ID NO: 95)
反向引物	AAGAGCAGGTGTGGGTTTAG (SEQ ID NO: 96)
探针	CGGTTGTTTCCCTCCTTGT (SEQ ID NO: 97)
分析F	
正向引物	CCCACTACTTGCTCTCACAG (SEQ ID NO: 98)
反向引物	TACAGGCTGCGGTTGTTT (SEQ ID NO: 99)
探针	ACTCGCTGAGGGTGAACAAGAAA (SEQ ID NO: 100)
分析G	
正向引物	AAGAGGCGCGGGTAGAA (SEQ ID NO: 70)
反向引物	CAGCTTCGGTCAGAGAAATGAG (SEQ ID NO: 71)
探针	CTCTCCTCAGAGCTCGACGCATTT (SEQ ID NO: 72)
分析H	
正向引物	CTGCACAATTCAGCCCAAG (SEQ ID NO: 73)
反向引物	CAGGTCATGTCCCACAGAAT (SEQ ID NO: 74)
探针	CATATGAGGGCAGCAATGCAAGTC (SEQ ID NO: 75)
分析I	
正向引物	CGAGTGGGTGAGTGAGGA (SEQ ID NO: 101)
反向引物	TTCTACCCGCGCCTCTT (SEQ ID NO: 102)
探针	ATCCTGGCGGGTGGCTGTTT (SEQ ID NO: 103)
分析J	
正向引物	CGGATAATGTGAGACCTGGAAT (SEQ ID NO: 104)
反向引物	AAAGGTAGCCGCAACAA (SEQ ID NO: 105)
探针	ACCATCTCCTGCTGTTGCCAAGA (SEQ ID NO: 106)

蛋白质印迹分析

[0328] 收集分化的胚状体 (EB) 并在 SDS 样品缓冲液 (2% SDS、10% 甘油、5% β -巯基乙醇、60mM TrisHCl、pH 6.8、溴酚蓝) 中均质化。使用 RC DC 蛋白质测定法 (BioRad) 对蛋白质提取物进行定量。提取物 (10 μ g) 在 4-20% SDS-PAGE 凝胶 (ThermoFisher) 上电泳, 并使用 iBLOT 转移装置 (ThermoFisher) 将其转移到硝酸纤维素膜上。用针对 C90RF72 和 GAPDH (Millipore) 的初次抗体检测蛋白质印迹。通过与偶联辣根过氧化物酶 (Abcam) 的再次抗体培育, 然后使用 SuperSignal West Pico 化学发光底物 (Thermo Scientific) 进行化学发光来检测结合抗体。使用全速蓝色敏感医用 X 射线胶片 (Full Speed Blue sensitive medical X-Ray

film, Ewen Parker XRay Corporation), 通过放射自显影术检测信号。使用ImageJ计算相对蛋白质水平。数据未显示。

用于检测有义或反义RNA转录产物的荧光原位杂交 (FISH)

[0329] 荧光原位杂交 (FISH) 用于确定RNA的位置, 所述RNA在如上文所述产生的胚胎干细胞衍生的运动神经元 (ESMN) 中, 从如SEQ ID NO: 1所示的六核苷酸重复序列转录而来。简而言之, ESMN在四孔室载玻片 (Lab-Tek II室载玻片系统, ThermoFisher Scientific) 中生长, 并用在PBS中的4% PFA (Electron Microscopy Sciences) 固定。然后用焦碳酸二乙酯 (DEPC) PBS/0.2% Triton X-100 (Fisher Scientific, 目录#BP151) 对细胞进行透化, 用DEPC-PBS洗涤, 用LNA寡核苷酸封闭和染色, 以检测RNA转录产物, 如下所述。染色后, 随后将载玻片与合适的荧光染料一起孵育, 用Fluoromount G (Southern Biotech) 固定, 并使用共聚焦显微镜可视化。

[0330] 对于LNA探针, 用由50% 甲酰胺 (IBI Scientific, 目录#IB72020)、DEPC 2x SSC [300mM氯化钠、30mM柠檬酸钠 (pH 7.0)]、10% (w/v) 硫酸葡聚糖 (Sigma-Aldrich, 目录号#D8960) 和DEPC 50mM磷酸钠 (pH 7.0) 组成的缓冲液, 在66°C下预杂交载玻片30分钟。然后排干杂交缓冲液, 将杂交缓冲液中的400μL的40nM LNA探针混合物添加到每个载玻片中, 并在66°C下在黑暗中孵育3小时 (对于LNA探针)。将用LNA探针孵育的载玻片在室温下, 在DEPC 2x SSC/0.1% Tween 20 (Fisher Scientific, 目录号BP337) 中洗涤一次, 在65°C下, 在DEPC 0.1x SSC洗涤3次。随后将载玻片与1μg/mL的DAPI (Molecular Probes Inc.) 一起孵育。

[0331] 在另一个实验中, 用由50% 甲酰胺 (IBI Scientific, 目录#IB72020)、DEPC 2x SSC [300mM氯化钠、30mM柠檬酸钠 (pH 7.0)]、10% (w/v) 硫酸葡聚糖 (Sigma-Aldrich, 目录号#D8960) 和DEPC 50mM磷酸钠 (pH 7.0) 组成的缓冲液, 在66°C下 (对于LNA探针) 或55°C下 (对于DNA探针) 预杂交载玻片30分钟。然后排干杂交缓冲液, 将杂交缓冲液中的400μL的200ng/mL DNA探针混合物添加到每个载玻片中, 并在55°C下在黑暗中孵育3小时。将用DNA探针孵育的载玻片用2x SSC中的40% 甲酰胺洗涤3次, 并在PBS中简单洗涤一次。随后将载玻片与1μg/mL DAPI (Molecular Probes Inc.) 一起孵育。

[0332] 本实施例中使用的LNA和DNA寡核苷酸探针的序列和SEQ ID NO, 以及探针的杂交条件在下表11中提供。锁核酸 (LNA) 是一种核酸类似物, 其中核糖模体被一个额外的桥修饰, 该桥连接2' 氧和4' 碳。

[0333] 表11. LNA和DNA探针。

探针	序列 (SEQ ID NO)	杂交方法
LNA 有义 G ₄ C ₂ RNA	TYE563-CCCCGGCCCCGGCCCC (SEQ ID NO: 38)	66°C 杂交并在 0.1 X SSC 中洗涤
LNA 反义 G ₄ C ₂ RNA	TYE563-GGGGCCGGGCGGGGGGCC (SEQ ID NO: 39)	66°C 杂交并在 0.1 X SSC 中洗涤
DNA 有义 G ₄ C ₂ RNA	CCCCGGCCCCGGCCCCGG—Cy3 (SEQ ID NO: 44)	55°C 杂交并在 2X SSC 中洗涤
DNA 反义 G ₄ C ₂ RNA	GGGGCCGGGCGGGGC—Cy3 (SEQ ID NO: 45)	55°C 杂交并在 2X SSC 中洗涤

二肽重复蛋白产物的检测 (免疫荧光和蛋白质狭缝印迹)

[0334] 免疫荧光用于评估在如上所述产生的胚胎干细胞衍生的运动神经元 (ESMN) 中产生的二肽重复蛋白。简而言之, ESMN在四孔室载玻片 (Lab-Tek II室载玻片系统,

ThermoFisher Scientific) 中生长, 并用在PBS中的4%PFA (Electron Microscopy Sciences) 固定。然后用焦碳酸二乙酯 (DEPC) PBS/0.2% Triton X-100 (Fisher Scientific, 目录#BP151) 对细胞进行透化, 用DEPC-PBS洗涤, 用抗聚GA抗体封闭和染色以检测RAN翻译产物, 如下文所述。染色后, 随后将载玻片与合适的荧光染料一起孵育, 用Fluoromount G (Southern Biotech) 固定, 并使用共聚焦显微镜可视化。

[0335] 在透化之后, 用在含有0.2% Triton X100 (TBS-T) 的Tris缓冲盐水 (pH 7.4) 中稀释的5% 正常驴血清封闭载玻片。将载玻片与在TBS-T中用5% 正常驴稀释的抗聚GA (Millipore) 的初次抗体在4°C下孵育过夜。在用TBS-T洗涤3次后, 将载玻片与偶联Alexa 488或555 (TBS-T, ThermoFisher中1:1000) 和DAPI (1 μ g/mL) (Molecular Probes Inc.) 的物种特异性再次抗体在室温下孵育1小时。用TBS-T洗涤3次后, 用Fluoromount G (Southern Biotech) 固定载玻片并使用共聚焦显微镜可视化。

[0336] 对于狭缝印迹分析 (slot blot assay), 收集分化的胚状体 (EB) 并在SDS样品缓冲液 (2% SDS、10% 甘油、5% β -巯基乙醇、60mM TrisHCl、pH 6.8、溴酚蓝) 中均质化。使用RC DC蛋白质测定 (BioRad) 对蛋白质提取物进行定量。含有0 μ g、1.25 μ g、2.5 μ g、5 μ g、10 μ g或20 μ g的裂解物在真空下用Bio-Slot 48孔微滤系统 (Bio-Rad) 固定在硝酸纤维素膜上。将膜在TBS-T中洗涤并用抗聚 (GP) (1:5,000, Novus bios) 和聚GA (1:5000, Millipore) 的抗体印迹。膜与HRP偶联的再次抗体孵育后, 通过ECL加蛋白质印迹检测系统 (Pierce) 观察条带。

结果

[0337] 测试了包含人源化C9orf72等位基因的等位基因系列的胚胎干细胞衍生的运动神经元 (ESMN), 该人源化C9orf72等位基因具有G₄C₂六核苷酸序列的3个重复、92个重复、250个重复、300个重复、500个重复或600个重复。如图20C、图20D、图20G和图20H所示, 在C9orf72基因座处包含六核苷酸重复扩增序列的ESMN显示保留内含子1序列的C9orf72 mRNA转录物的增加的表达。此外, 更大的G₄C₂重复扩增增加了外显子1A的使用, 减少了外显子1B的使用。分别参见图20A和20B, 以及分别参见图20E和20F。包含G₄C₂重复扩增的ESMN还包含核和细胞质有义和反义C9orf72 RNA病灶 (数据未显示)。有义G₄C₂病灶更大并分布到核仁 (数据未显示)。在重复扩增大于300的ESMN中, RNA病灶与核仁的定位再现了ALS患者来源的细胞的病理学发现。此外, G₄C₂六核苷酸序列重复数的增加与从六核苷酸重复序列的转录本翻译而来 (通过RAN翻译, 一种非AUG机制) 的二肽重复蛋白 (聚GA和聚GP) 的增加的存在直接相关。参见, 例如, 图21。总之, 源自等位基因系列ES细胞的运动神经元再现ALS疾病的分子标志 (有义和反义重复RNA病灶、定位于核仁的重复RNA病灶、五种形式的二肽重复蛋白质中的至少两种, 以及含内含子转录物的增加的积累, 从而支持使用本文公开的非人类动物作为神经退行性疾病的疾病模型。

[0338] 在两个类型的ESMN中重复表10中的测定C、B、F、H、E和D的定量PCR反应, 所述的两个类型的ESMN为: 近轴样运动神经元 (hypaxial-like motor neurons) (MN) 和肢体样运动神经元 (limb-like motor neurons) (MN)。近轴样MN使模型中的近轴MN受神经支配。由近轴MN支配的肌肉的例子包括肋间肌、隔膜和腹壁肌肉。肢体样MN是由外侧运动柱产生并支配远端肢体肌肉的那些, 远端肢体肌肉例如前肢和后肢 (胫前肌、腓肠肌和臀肌)。主要的近轴MN是由添加视黄酸 (RA) 和声波刺猬激动剂 (sonic hedgehog agonist, SAG) 的方案生成的。可以通过在RA和SAG的基础上添加1 μ M purmorphamine, 生成肢体样MN, 我们可以生成

肢体样MN。

[0339] 测试了包含人源化C9orf72等位基因的等位基因系列的运动神经元,该人源化C9orf72等位基因具有G₄C₂六核苷酸序列的3个重复、92个重复、300个重复、500个重复或600个重复。如图23A和图23B所示,较大的G₄C₂重复扩增增加了外显子1A的使用,减少了外显子1B的使用。如图23C和图23D所示,在C9orf72基因座处包含六核苷酸重复扩增序列的ESMN显示保留内含子1序列的C9orf72 mRNA转录物的增加的表达。如图23E所示,较大的G₄C₂重复扩增增加了未剪接前体的表达。如图23F所示,剪接的C9orf72 mRNA的表达不会随重复大小而变化太大。

实施例4. 在C9orf72基因座处产生包含六核苷酸重复扩增的小鼠

[0340] F0小鼠是采用VELOCIMOUSE®方法,使用实施例1中描述的300x重复ES细胞克隆产生的。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,每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)细胞注射到桑椹胚前期胚胎(例如八细胞期胚胎)中,这有效地产生了完全由ES细胞衍生的F0代小鼠。

[0341] 和实施例3一样,测试了来自C9orf72 300x重复小鼠(即具有300个重复的G₄C₂六核苷酸序列的人源化C9orf72等位基因)和对照C9orf72 3x重复小鼠(即具有3个重复的人源化C9orf72等位基因)的脑干和脊髓样品的RNA转录物。和实施例3一样,在来自C9orf72 300x重复小鼠的脑干和脊髓样品中评估RNA病灶和二肽重复蛋白水平。如图22B所示,来自C9orf72 300x重复小鼠的脑干和脊髓样本显示保留内含子1序列的C9orf72 mRNA转录本的增加的表达。此外,对脊髓样本(即L4/L5腰(lumbar)脊髓运动神经元)进行了测试,结果显示其含有核和细胞质有义和反义C9orf72 RNA病灶(数据未显示)。类似地,这些样品具有从六核苷酸重复序列的转录本翻译而来(通过RAN翻译,一种非AUG机制)的二肽重复蛋白(聚GA)的增加的存在(数据未显示)。

[0342] 类似地,测试了包含C9orf72等位基因的等位基因系列的胚胎干细胞衍生的运动神经元(ESMN),该人源化C9orf72等位基因具有G₄C₂六核苷酸序列的3个重复、92个重复、300个重复、500个重复或600个重复。如图22A所示,在C9orf72基因座处包含六核苷酸重复扩增序列的ESMN显示保留内含子1序列的C9orf72 mRNA转录物的增加的表达。含有G₄C₂重复扩增的ESMN还含有核和细胞质有义和反义C9orf72 RNA病灶,以及从六核苷酸重复序列的转录本翻译而来(通过RAN翻译,一种非AUG机制)的二肽重复蛋白(聚GA)的增加的存在(数据未显示)。

[0343] 还使用已知方法产生了F0小鼠,该F0小鼠具有小鼠C9orf72基因座,该小鼠C9orf72基因座被包含约500个重复或约600个重复的GGGGCC(SEQ ID NO:1)六核苷酸的人类对应物替换。

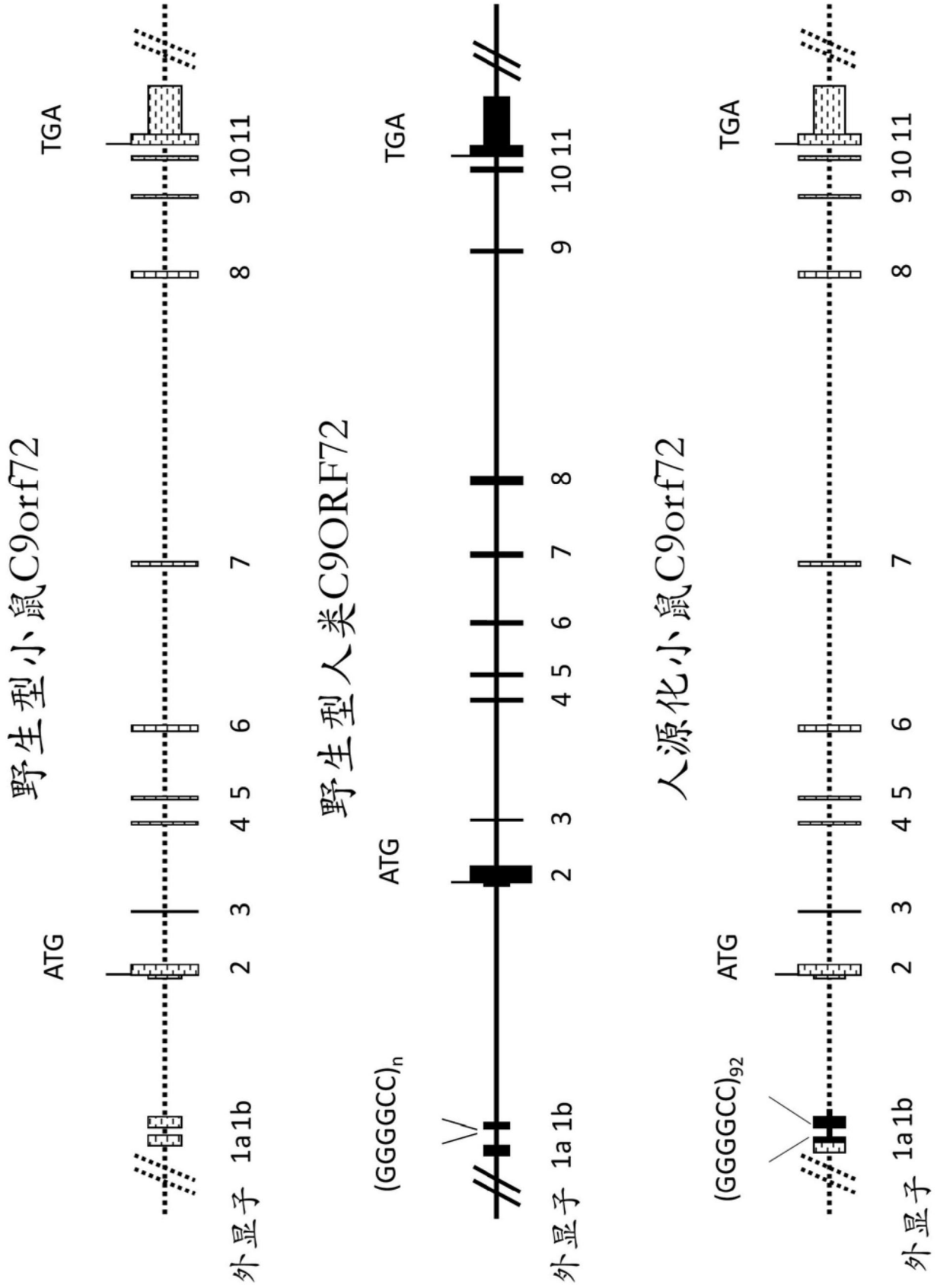


图1

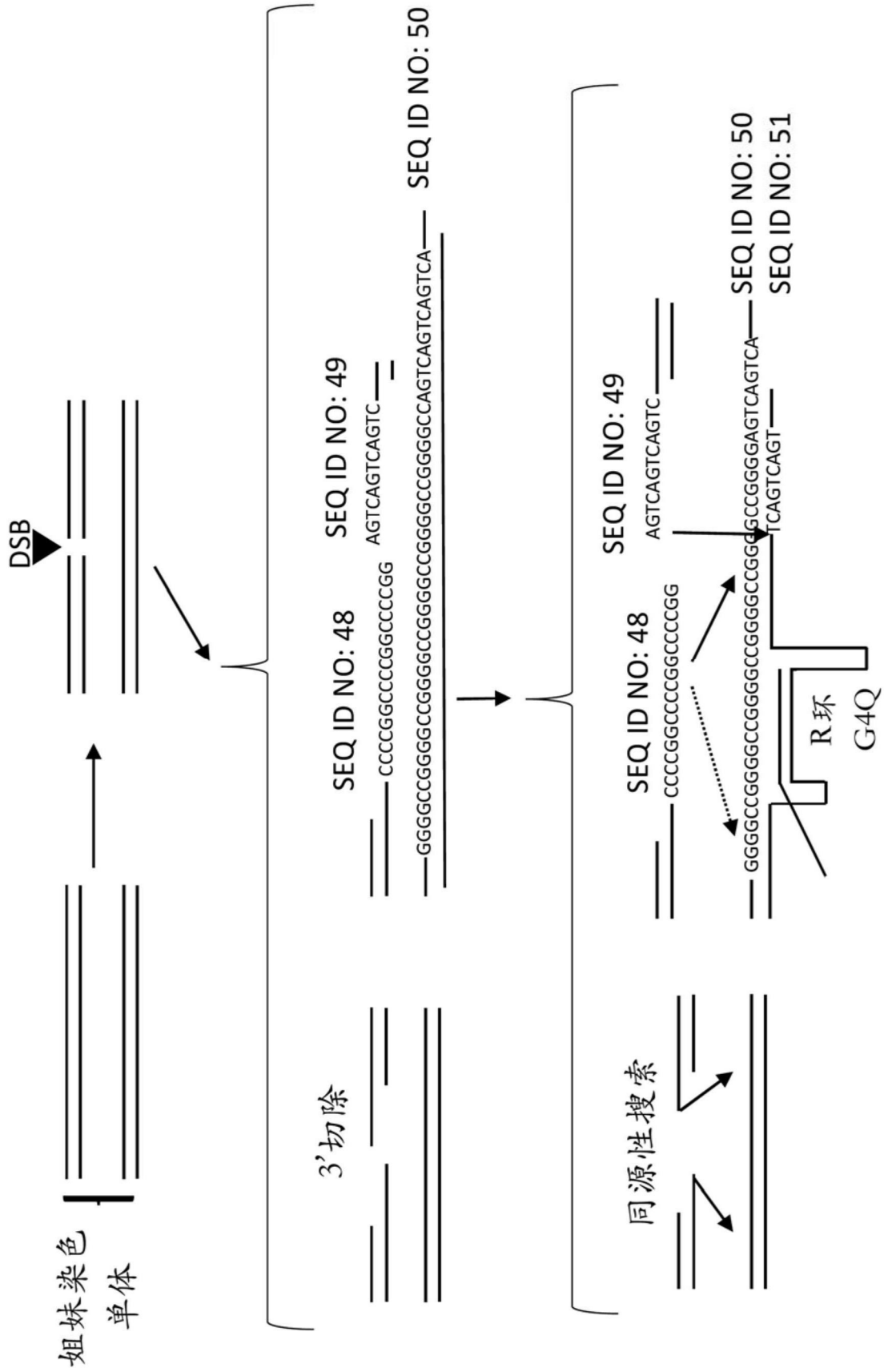


图2

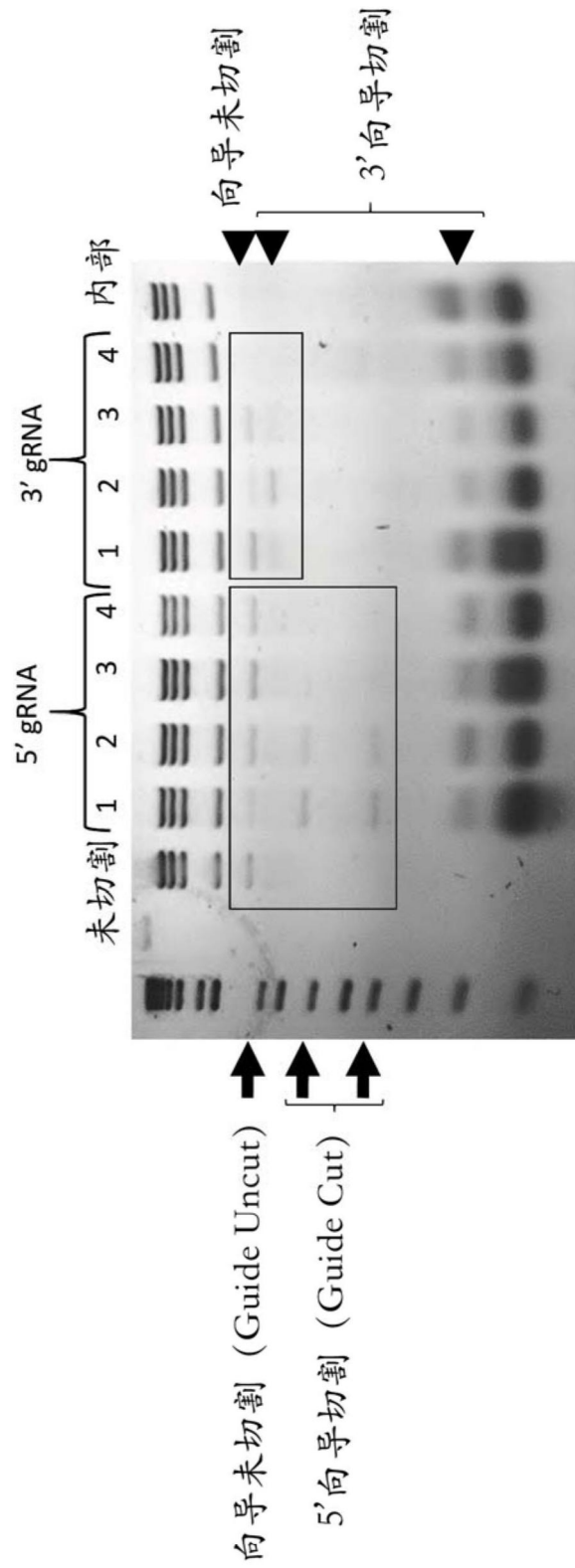


图4

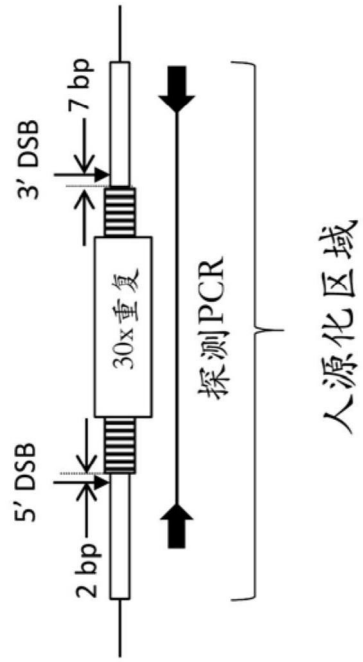


图5

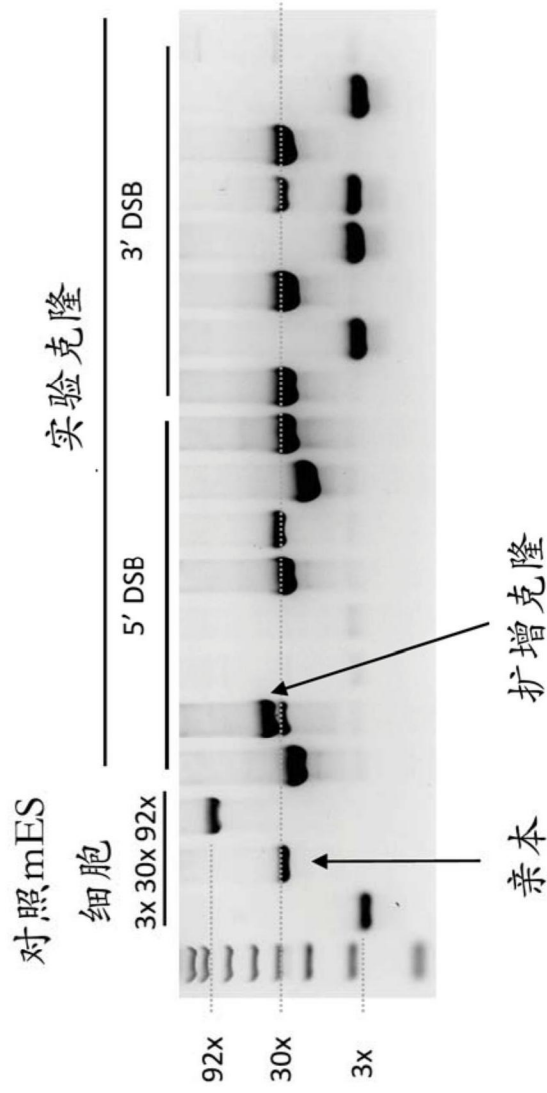


图6A

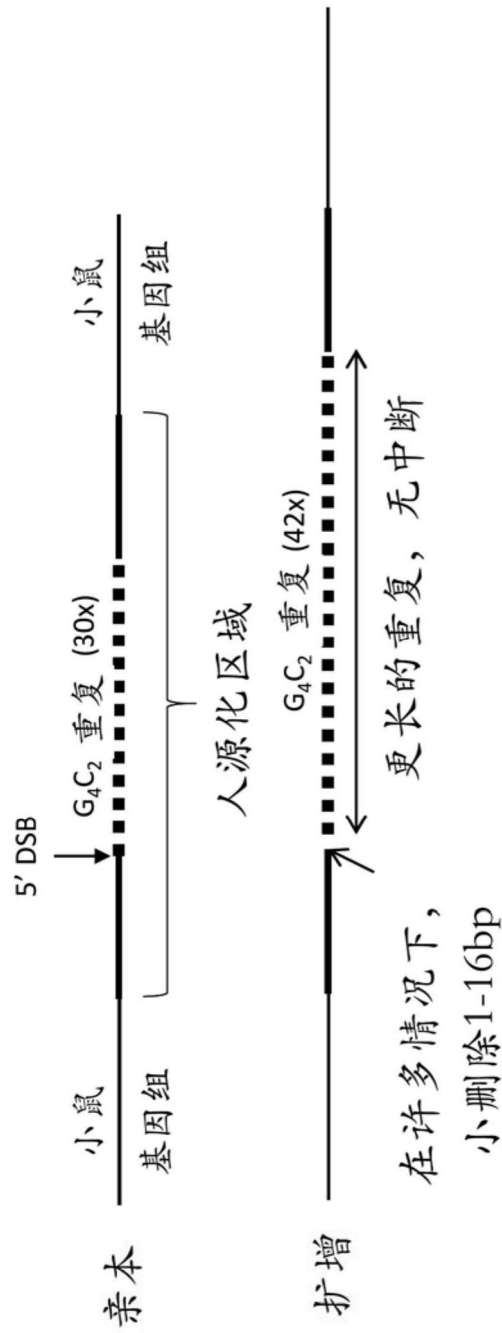


图6B

5' gRNA

亲本 (30x) CTCAGGAGTCGGCGGCTATGCGATCGCCGCTCGGGGCCGGGGCCGGGGCCGGGGCCGGG
 扩增克隆 (43x) CTCAGGAGTCGGCGGCTATGCGATCGCCGCTCGGGGCCGGGGCCGGGGCCGGGGCCGGG

亲本 (30x) GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGG
 扩增克隆 (43x) GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGG

亲本 (30x) GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGG
 扩增克隆 (43x) GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGG

亲本 (30x) GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGG
 扩增克隆 (43x) GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGG

亲本 (30x) (SEQ ID NO: 107)
 扩增克隆 (43x) GCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGGGCCGGG
 (SEQ ID NO: 108)

图6C

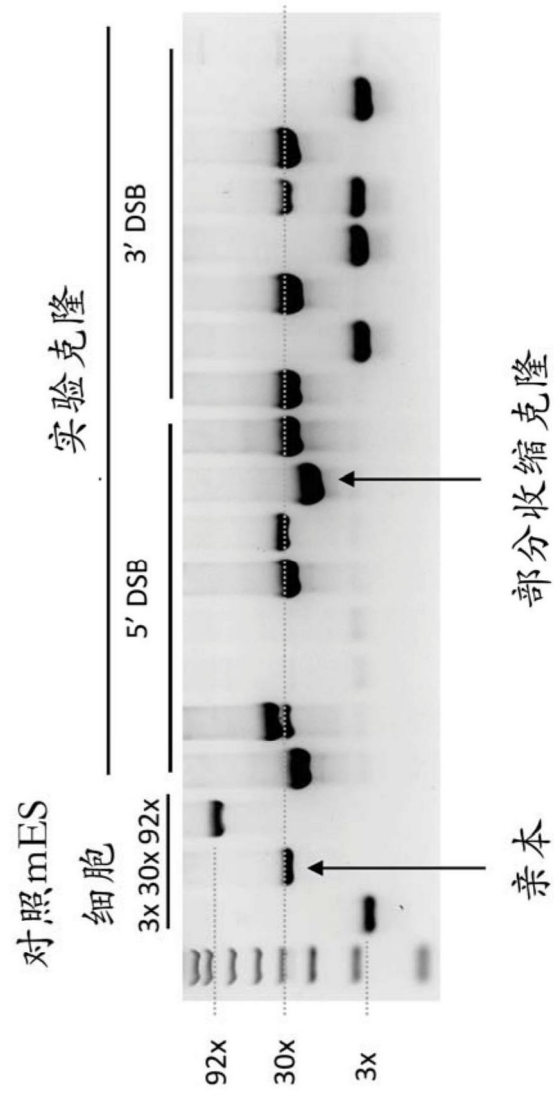


图7A

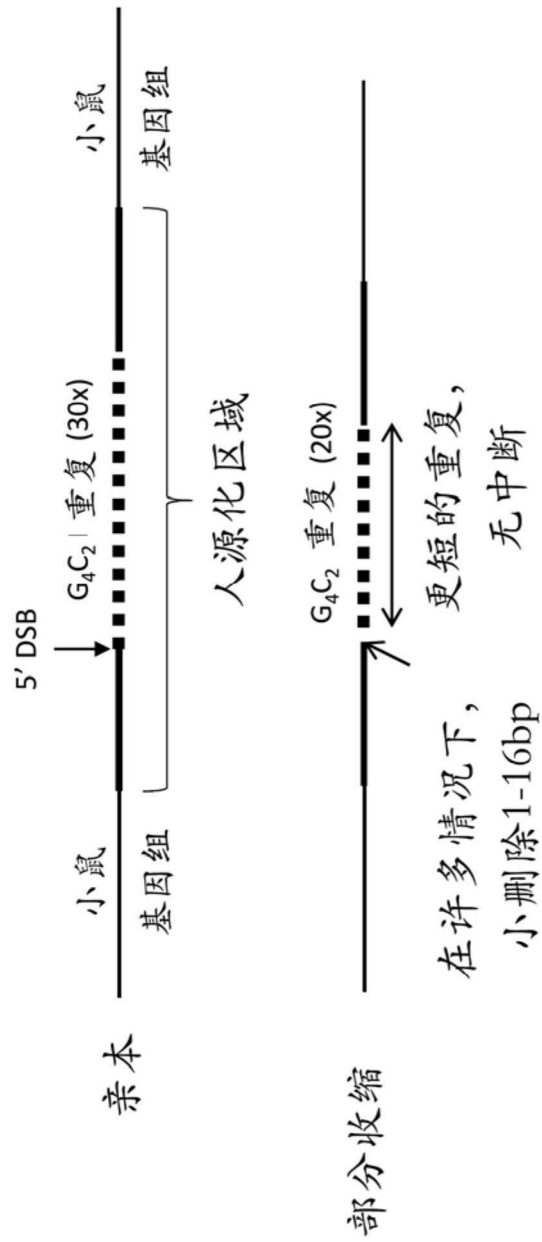


图7B

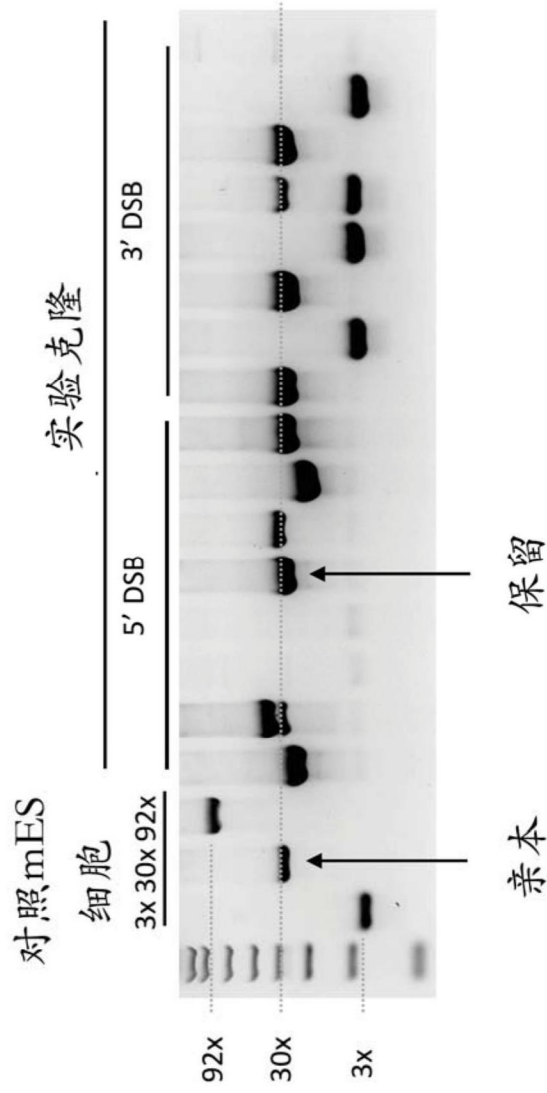


图8A

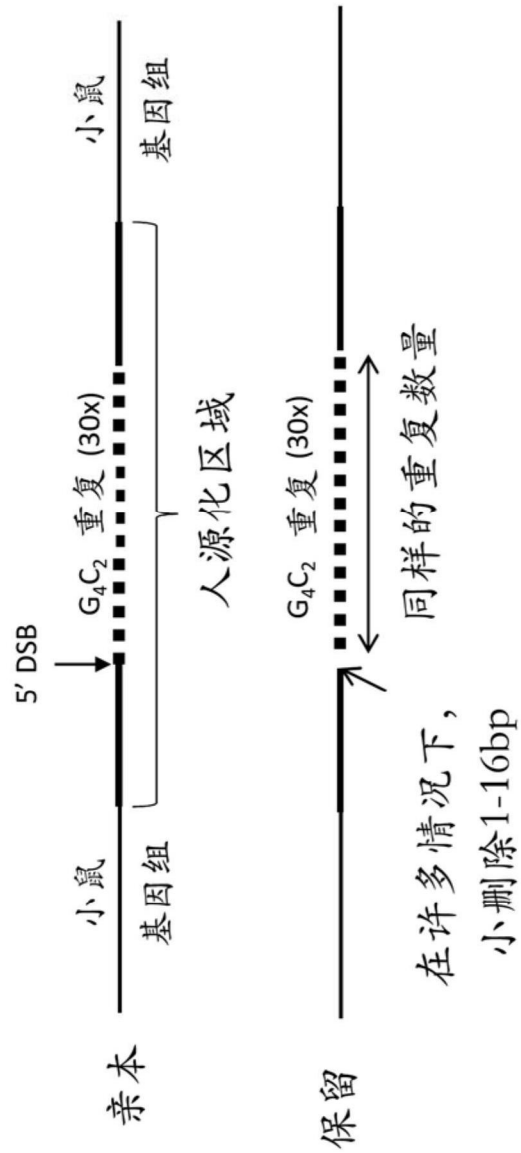


图8B

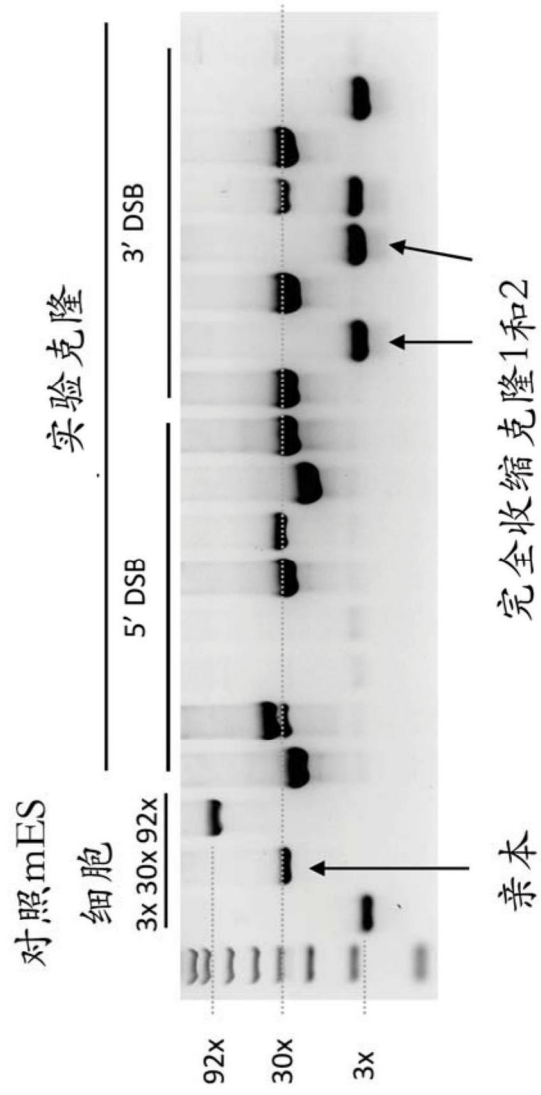


图9A

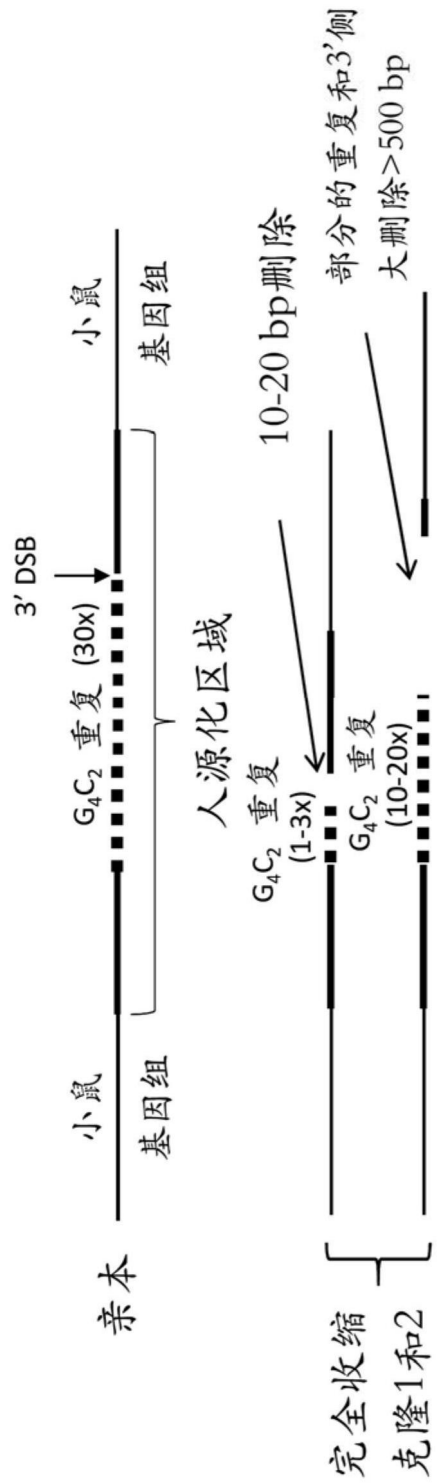


图9B

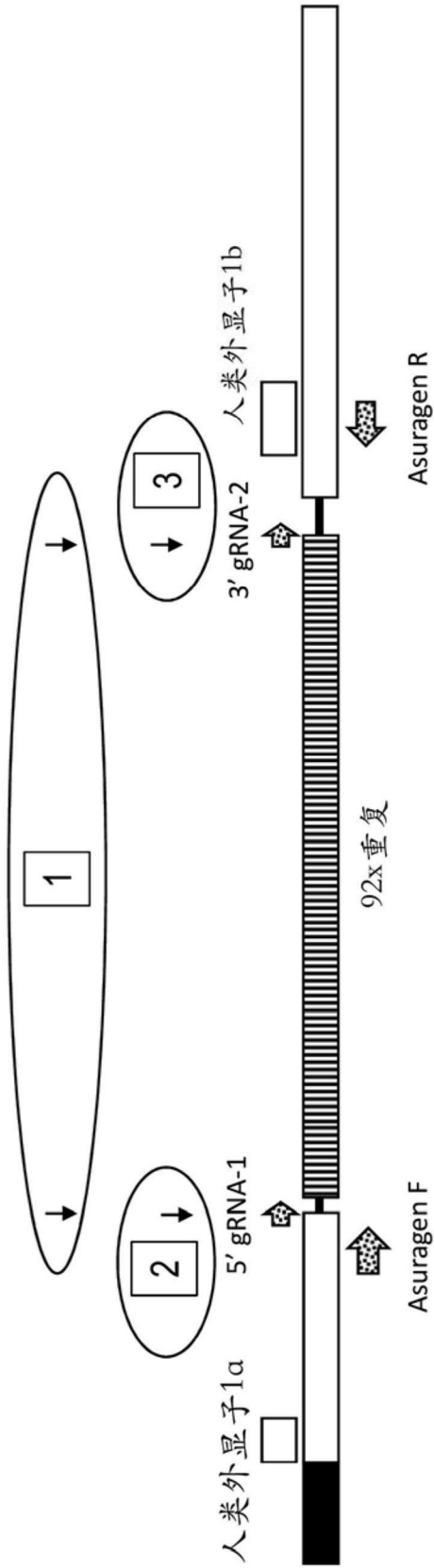


图10

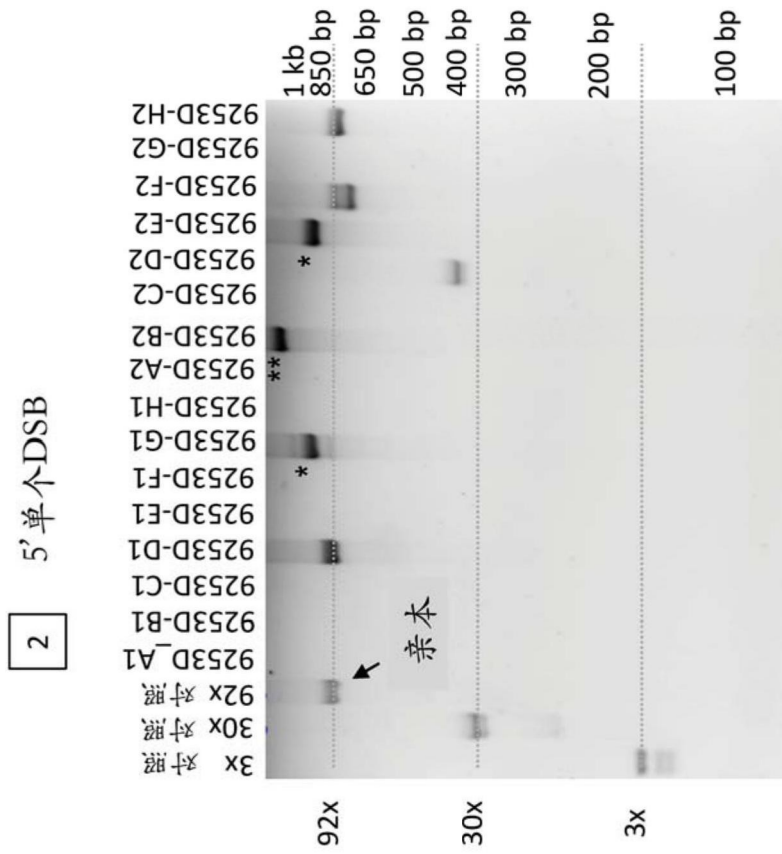


图12A

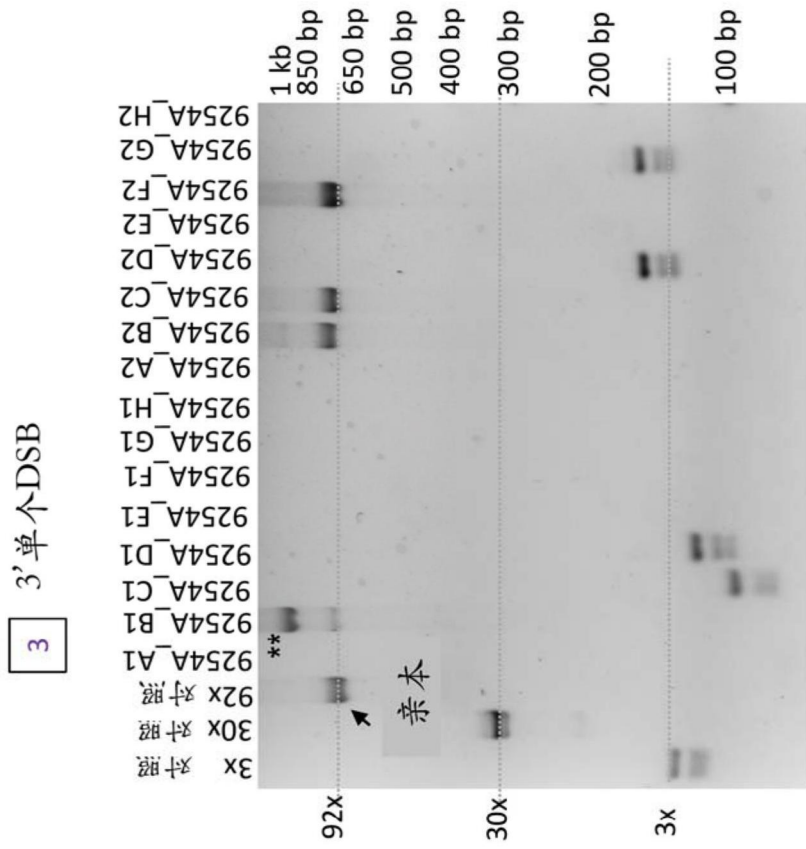


图12B

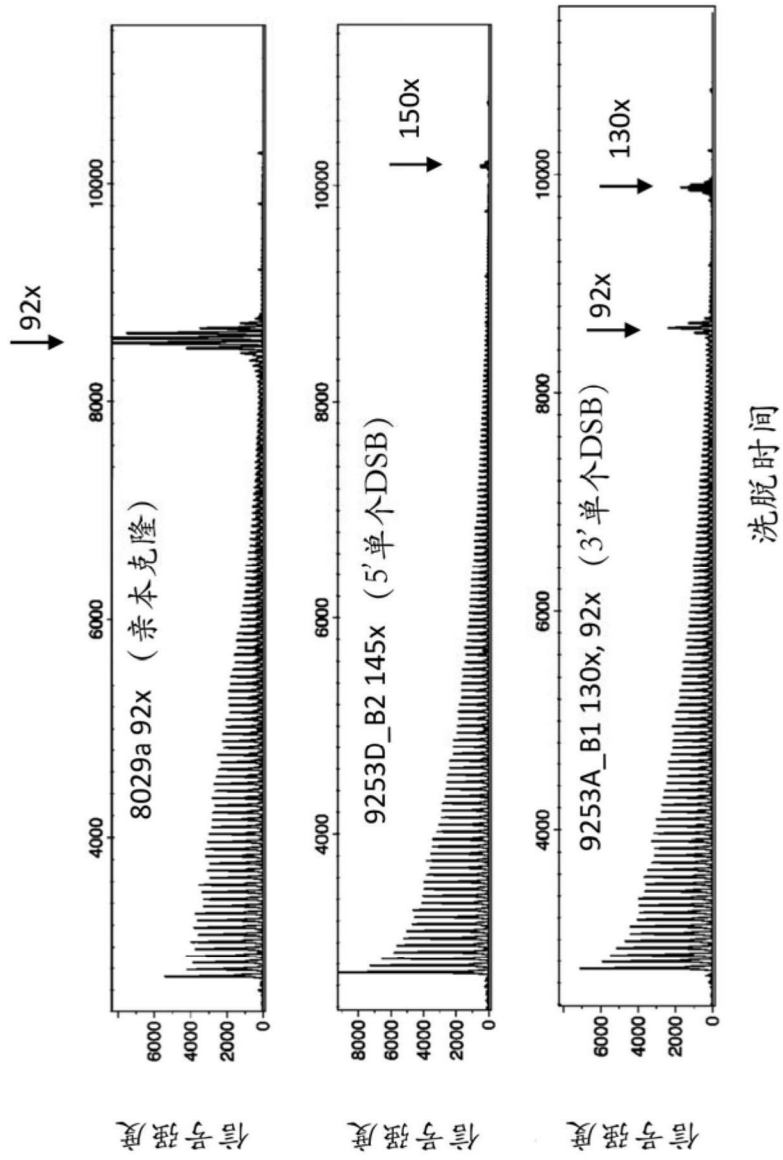


图13

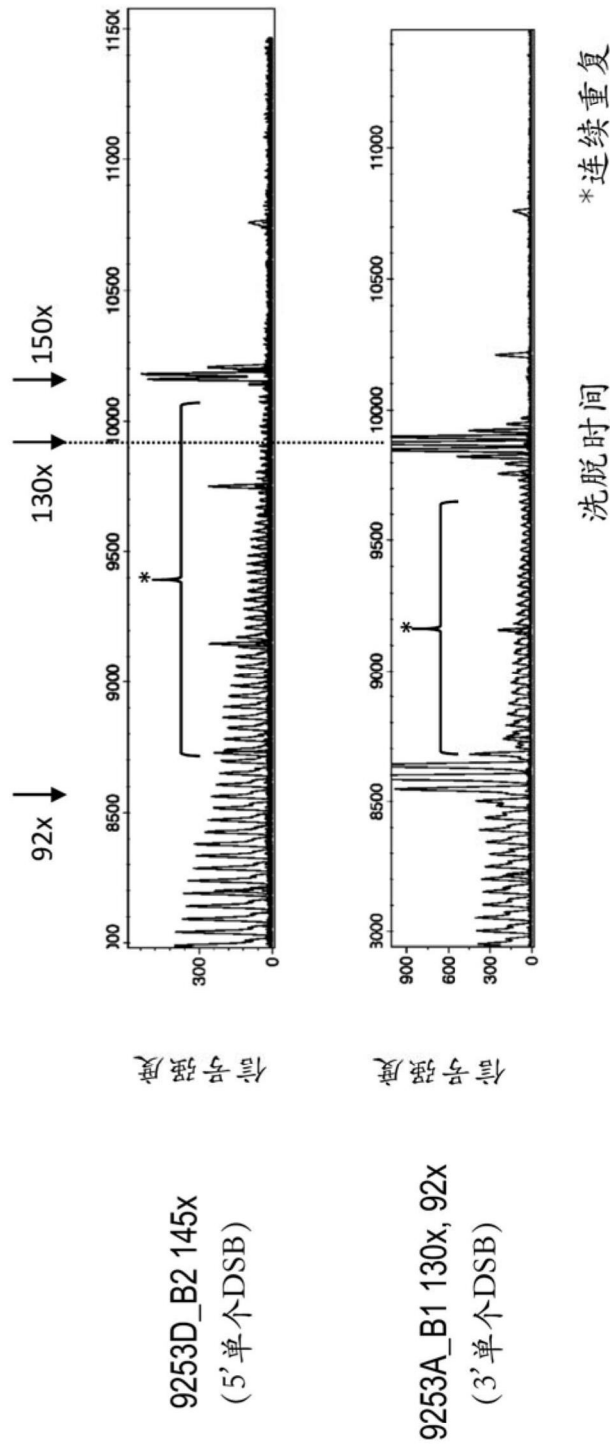


图14

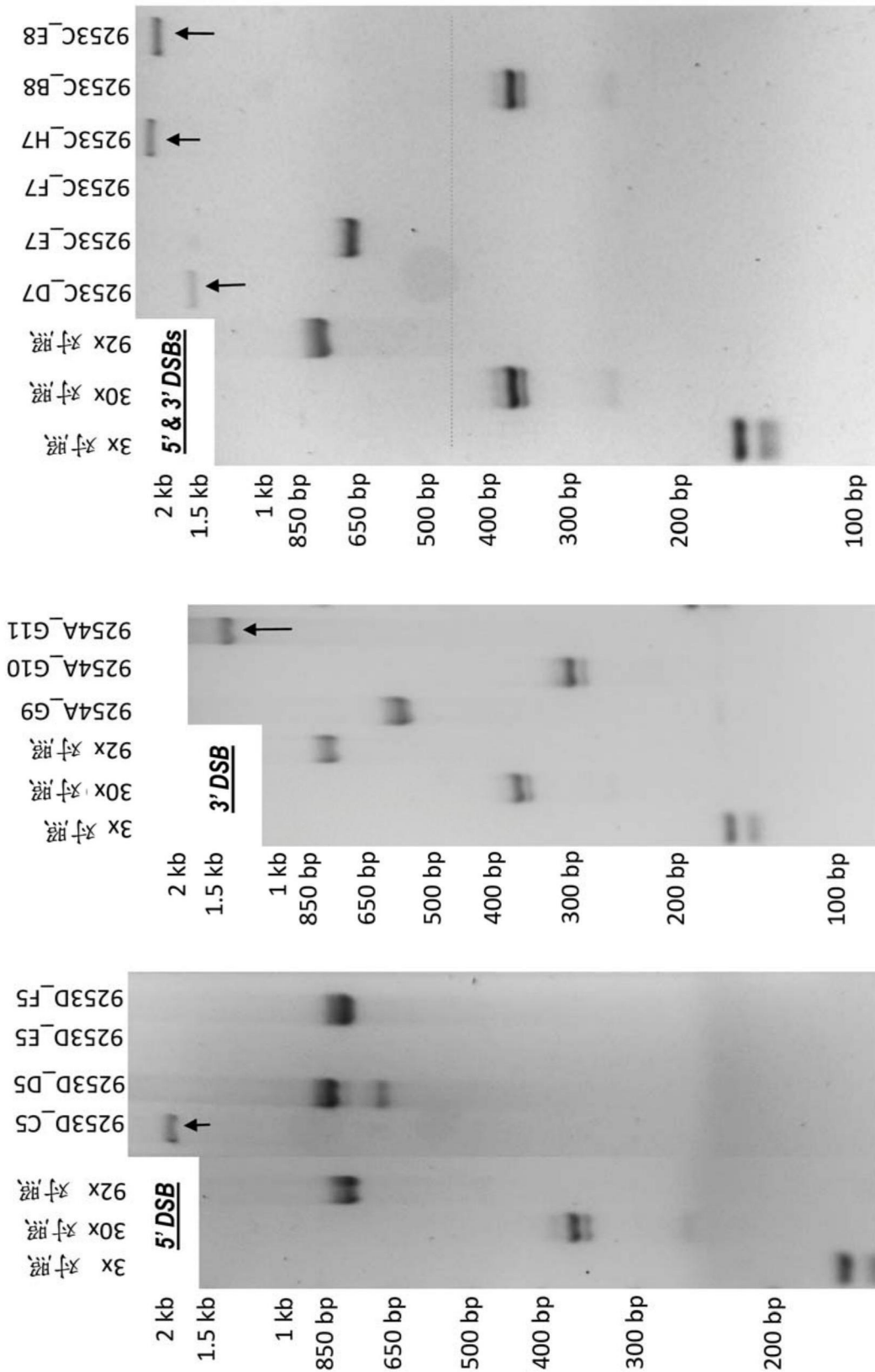


图15

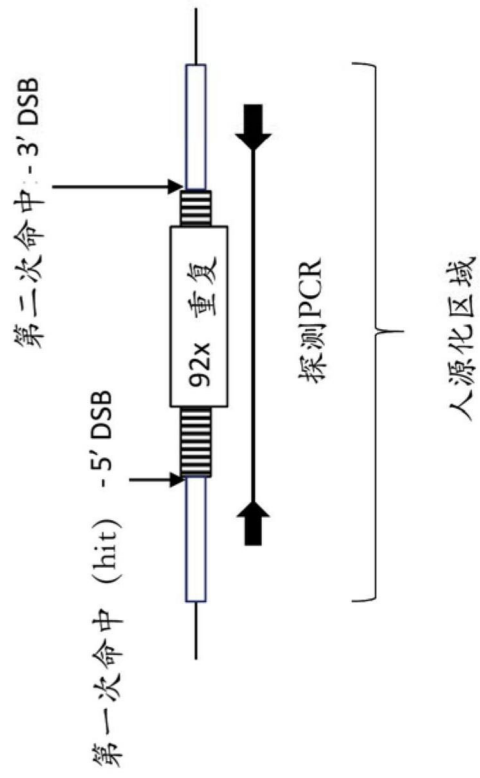


图16A

第二次扩增

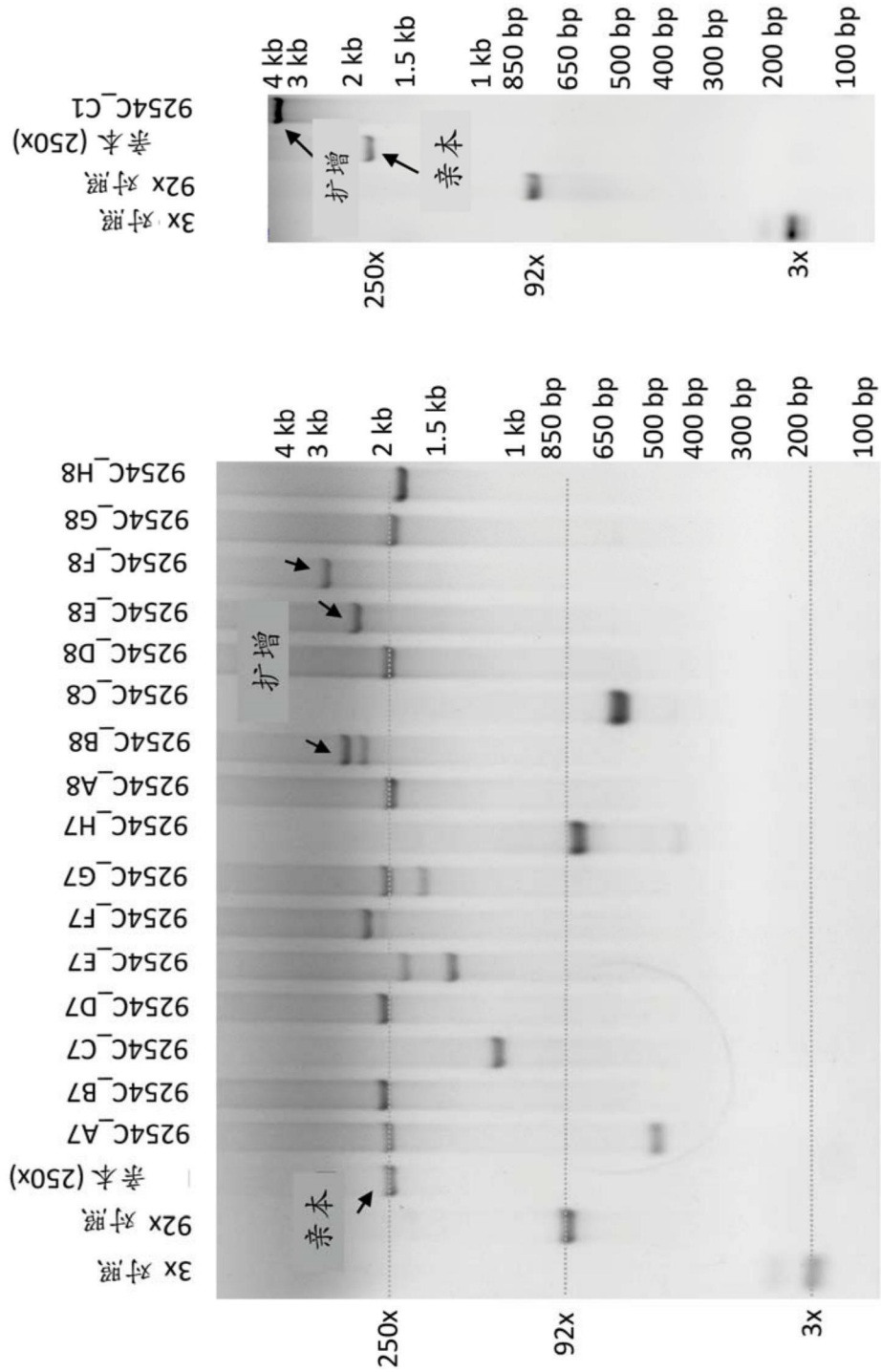


图16B

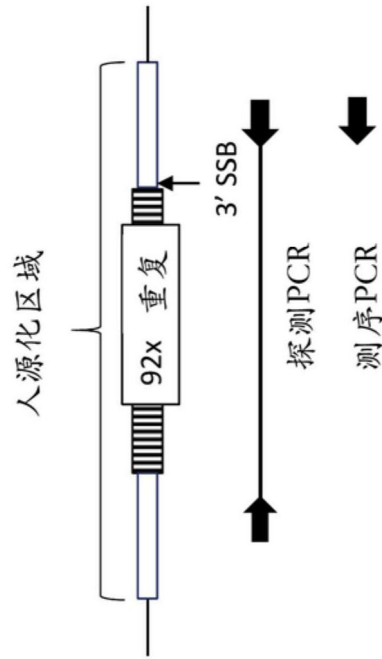


图17A

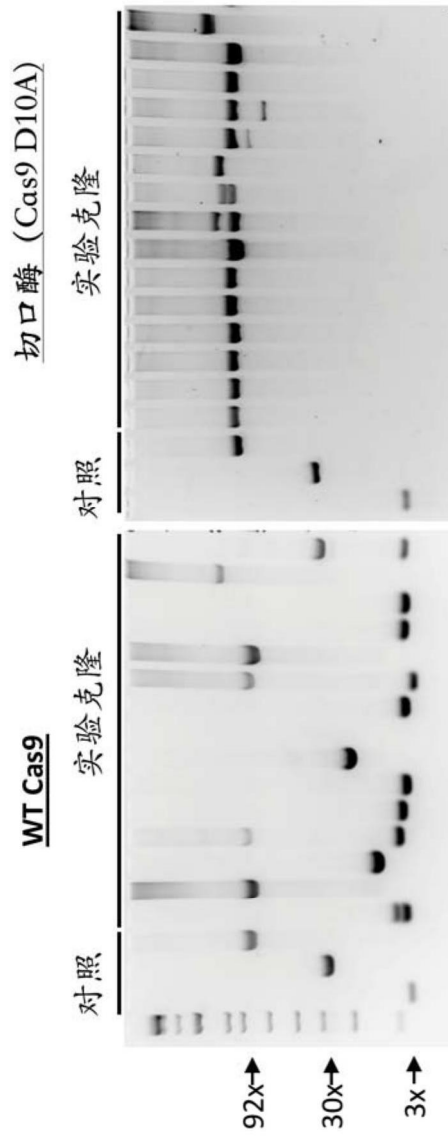


图17B

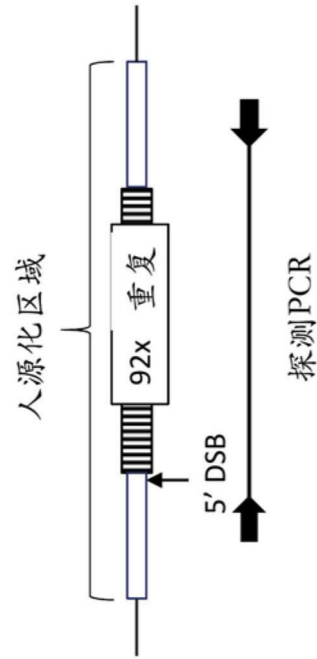


图18A

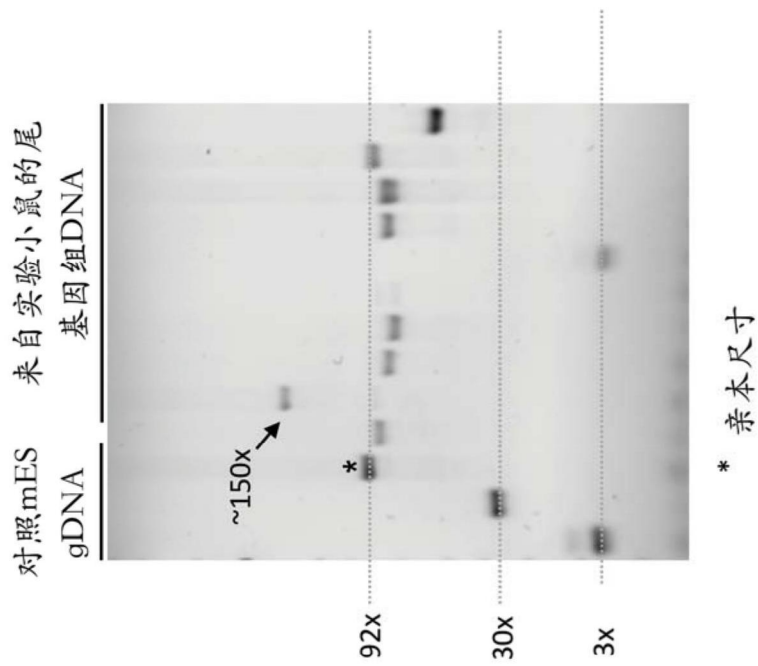


图18B

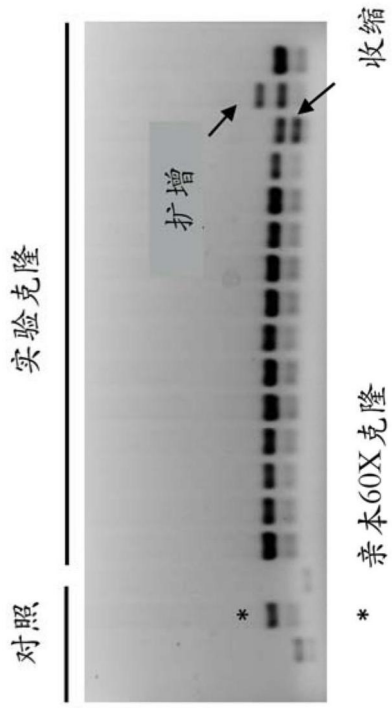


图19

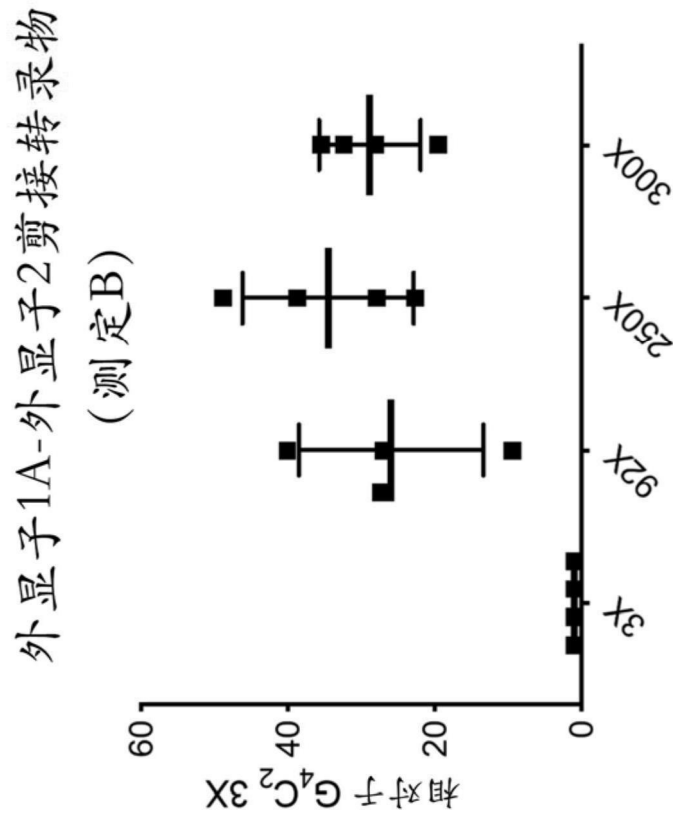


图20A

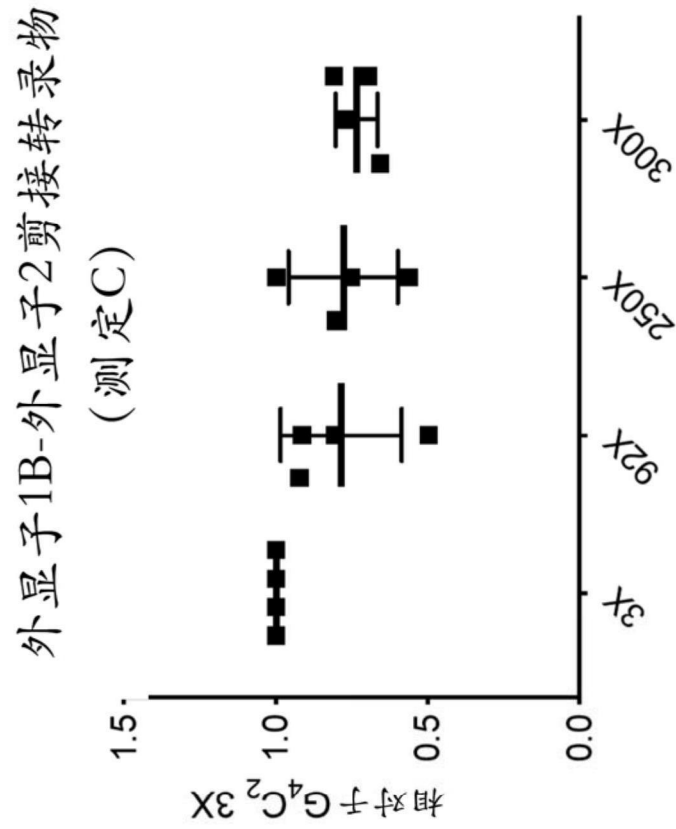


图20B

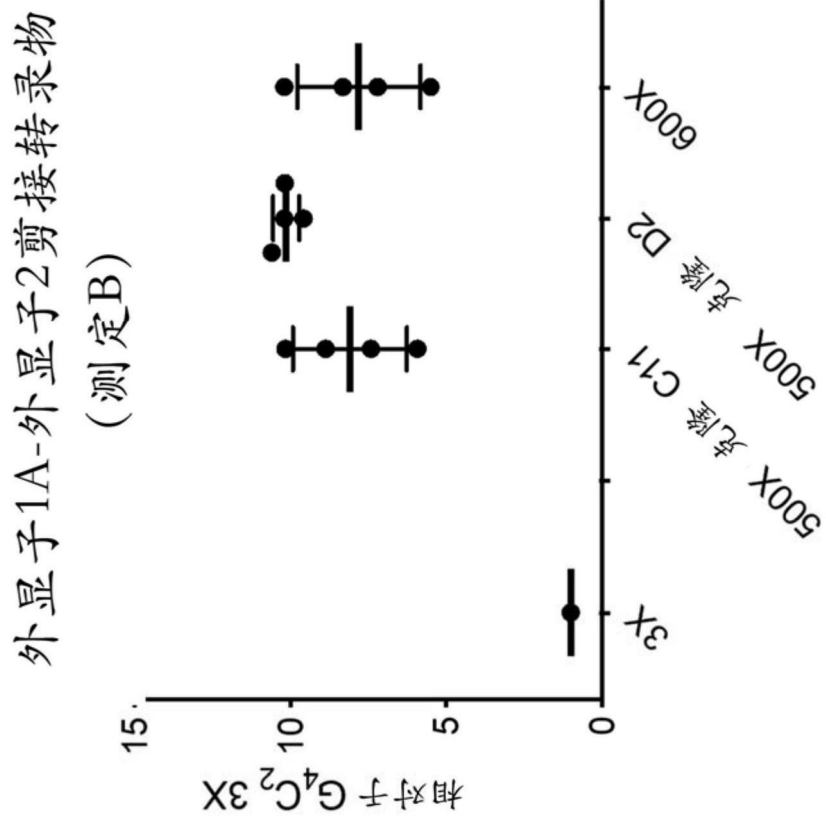


图20E

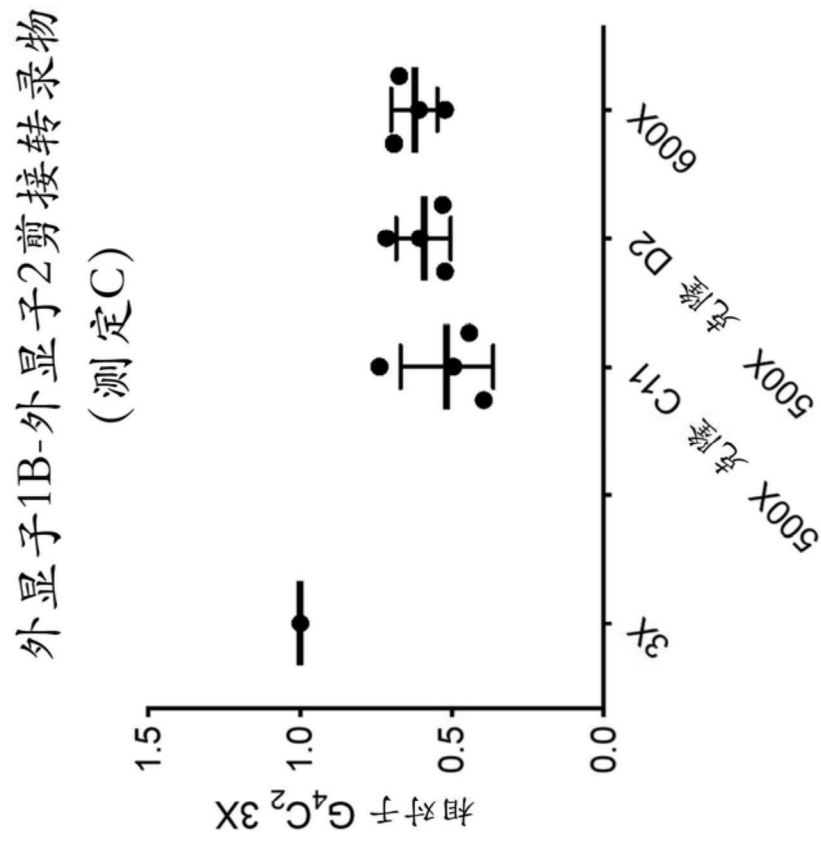


图20F

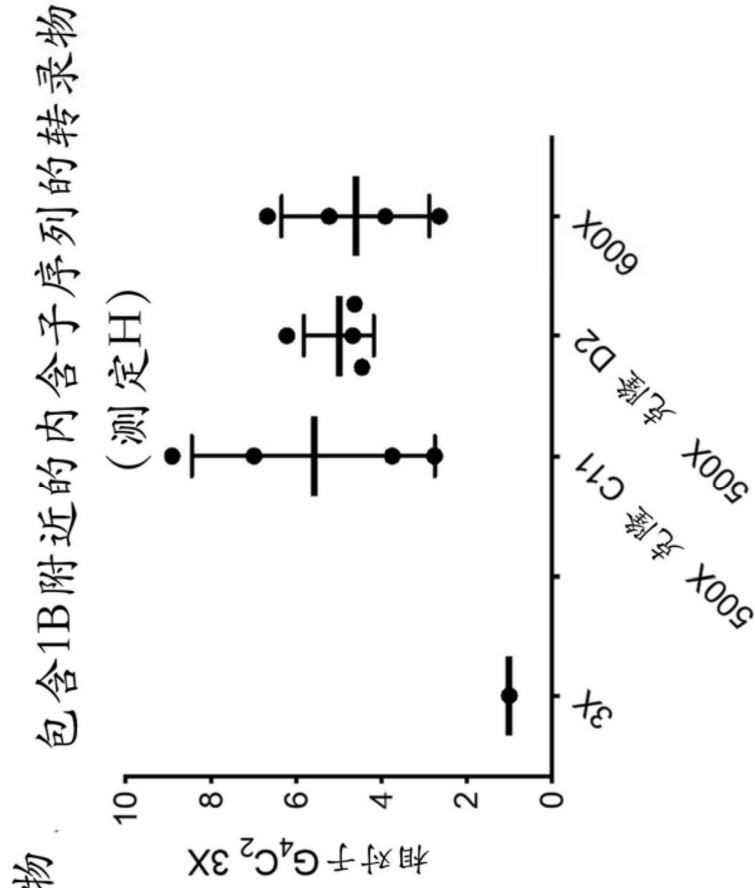


图20G

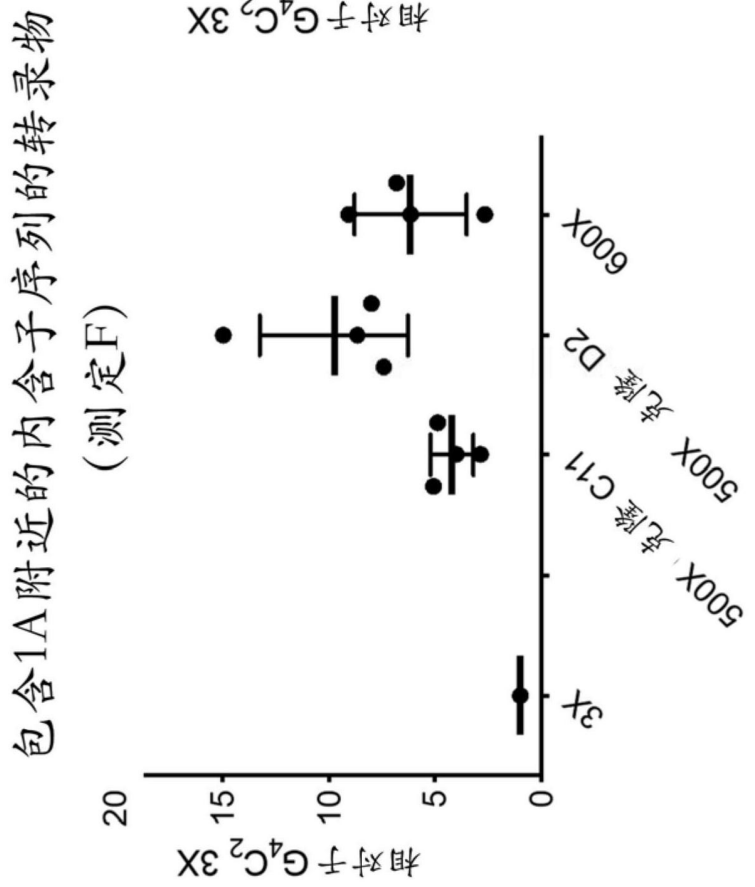


图20H

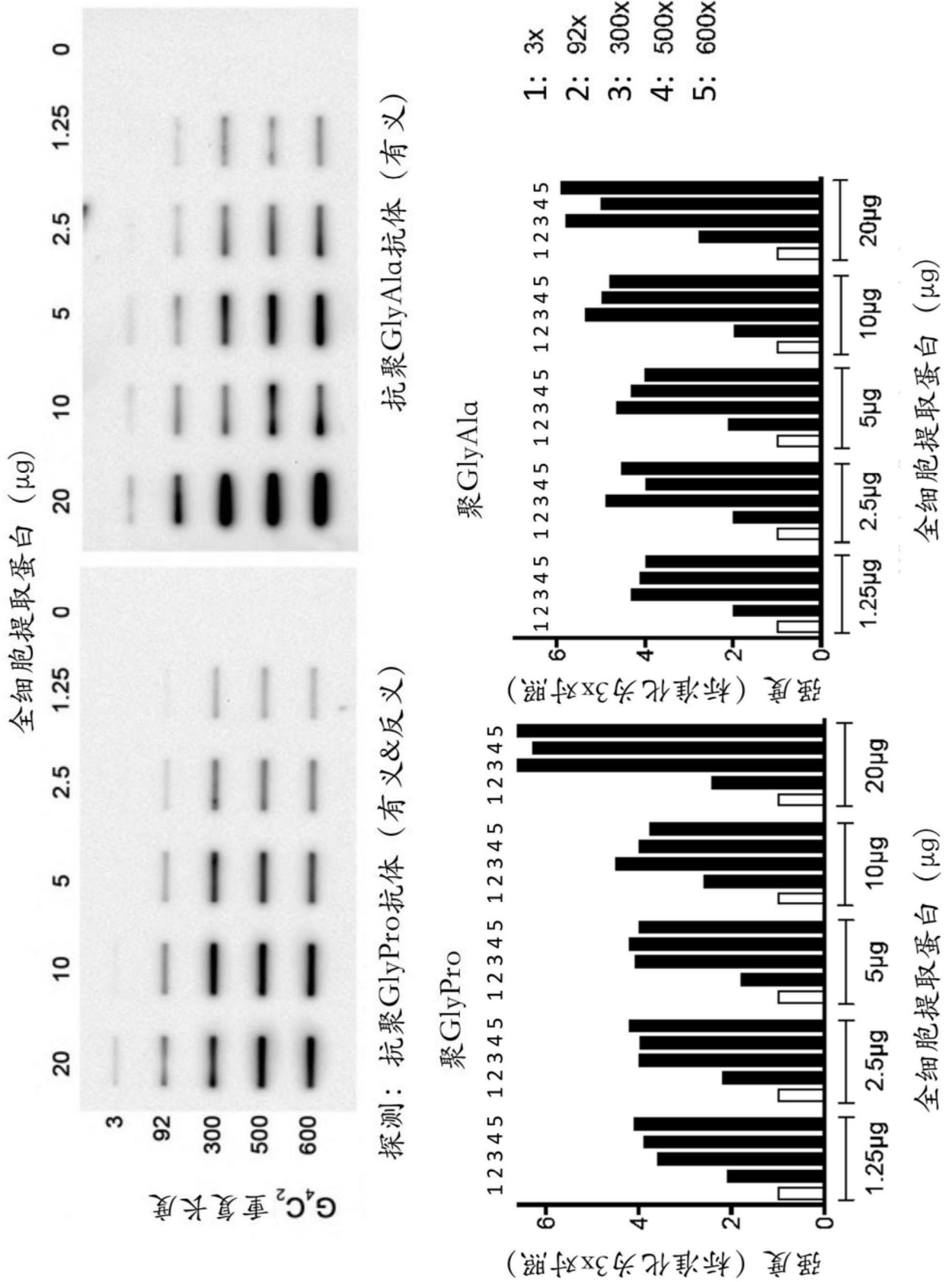


图21

包含1A附近的内含子序列的转录物

EC细胞衍生的运动神经元

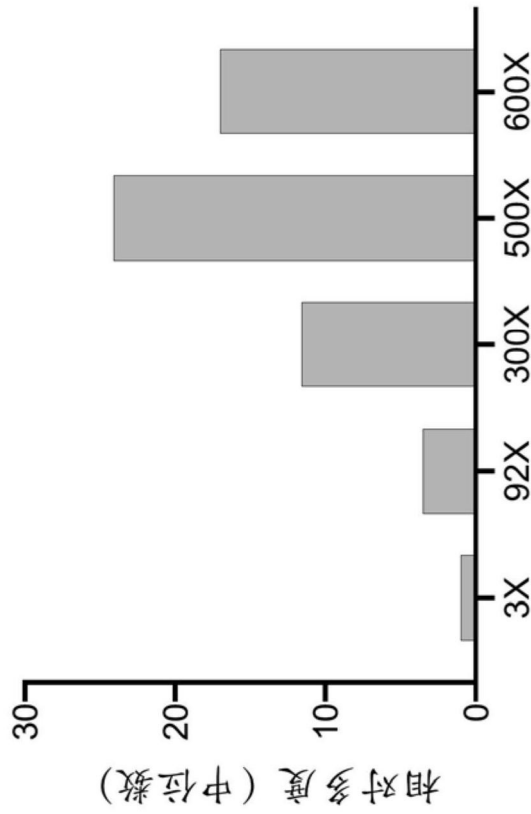


图22A

C9orf72 300x重复小鼠

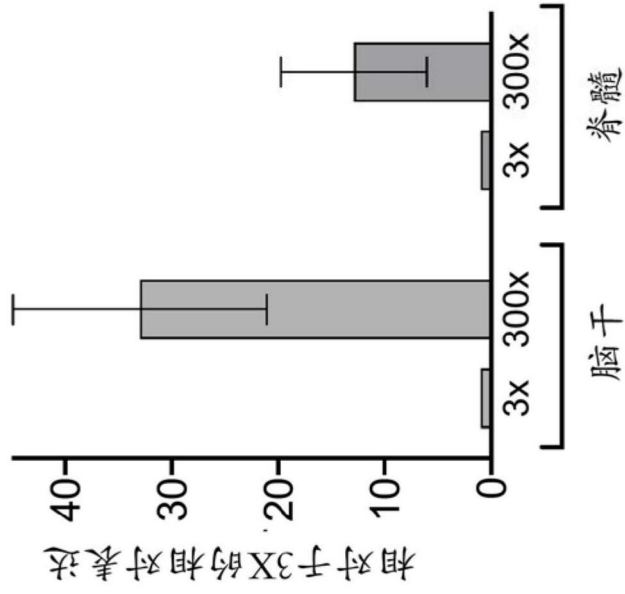


图22B

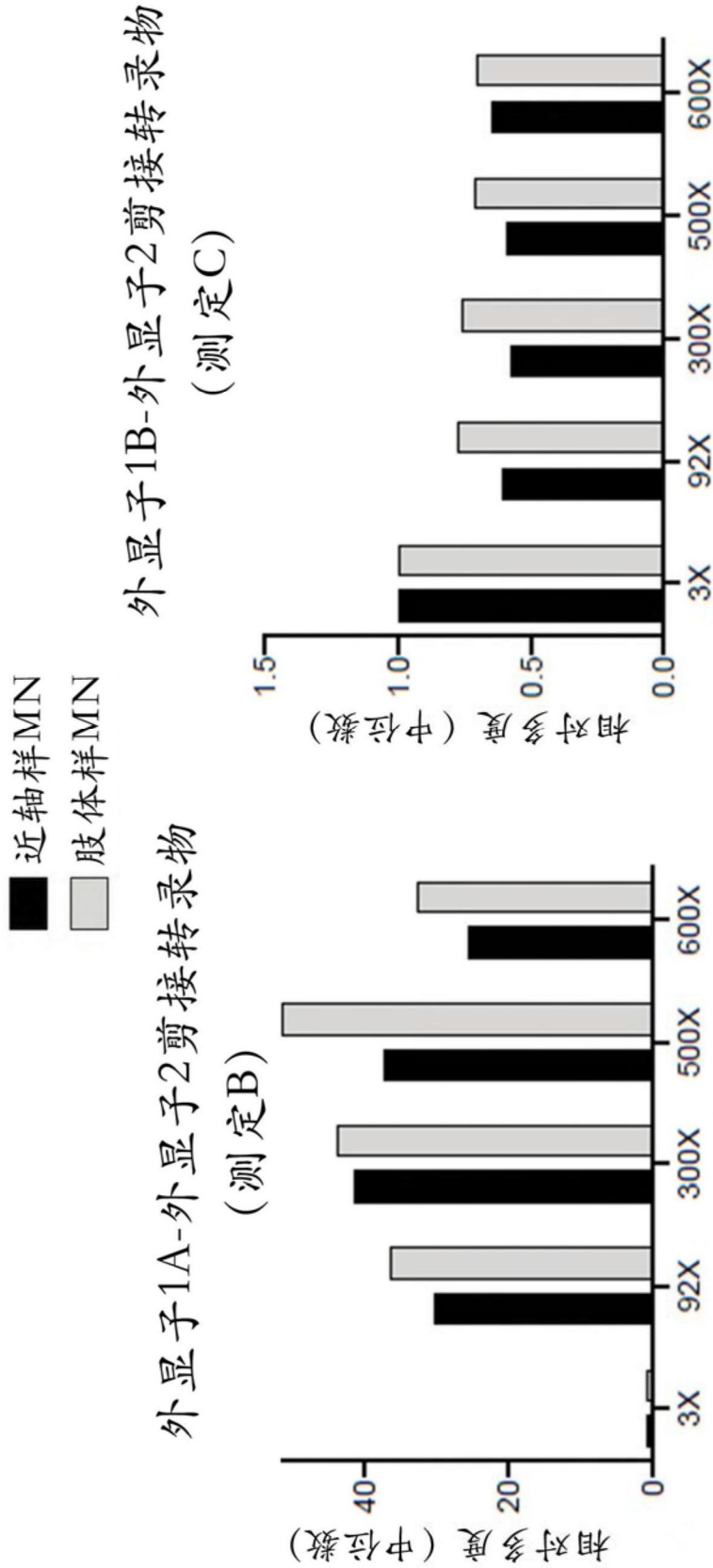


图23B

图23A

■ 近轴样MN
■ 肢体样MN

包含1B附近的内含子序列的转录物
(测定H)

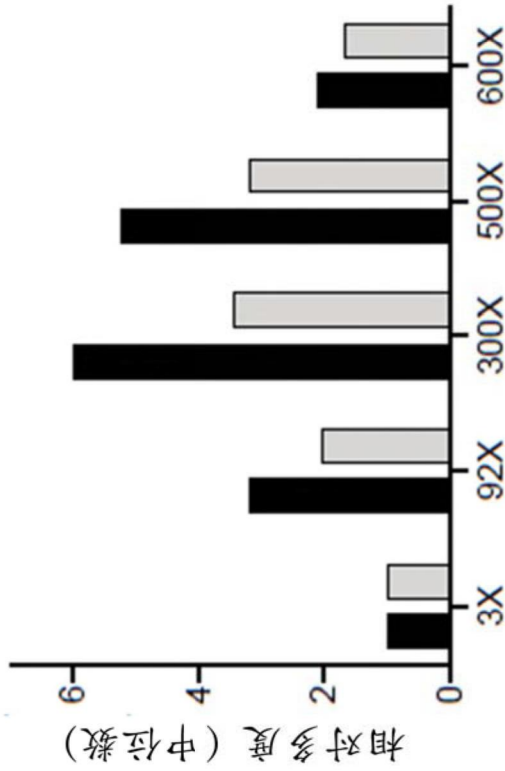


图23D

包含1A附近的内含子序列的转录物
(测定F)

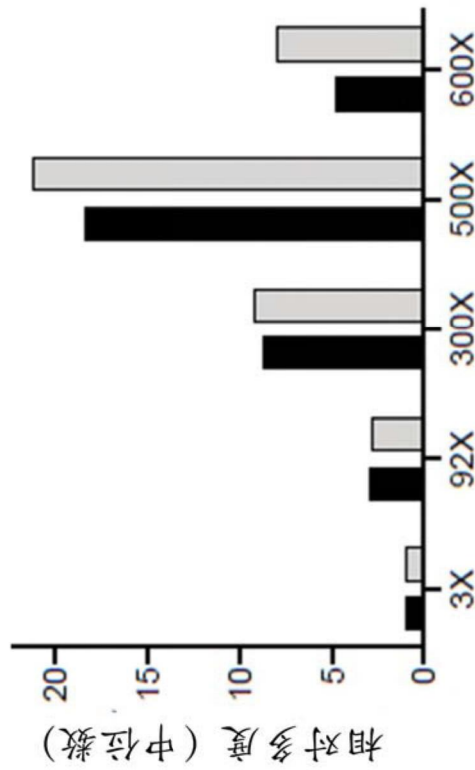


图23C

■ 近轴样MN
■ 肢体样MN

外显子1A-内含子 (未剪接前体)
(测定E)

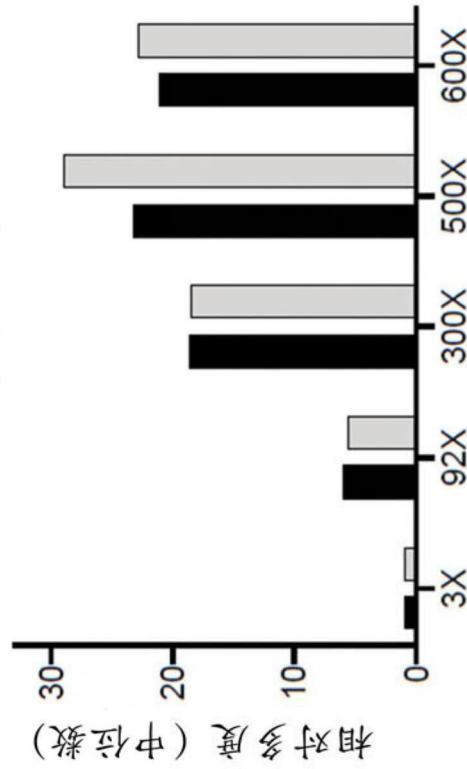


图23E

外显子5-6 (剪接mRNA)
(测定D)

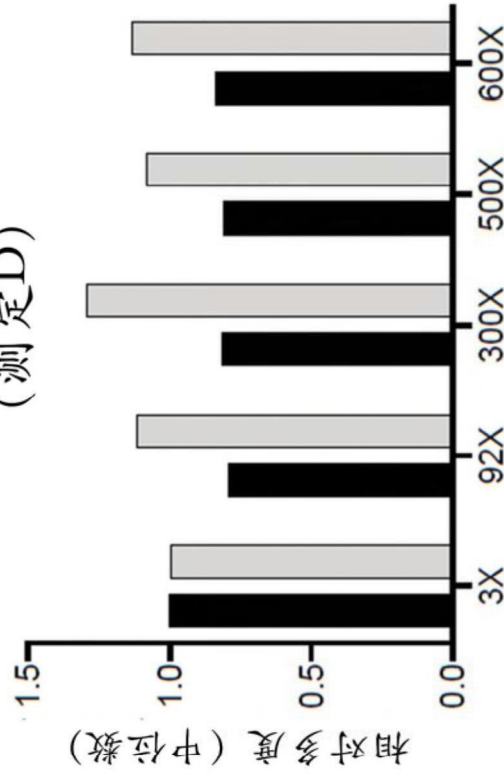


图23F