



(12)发明专利申请

(10)申请公布号 CN 107208113 A

(43)申请公布日 2017.09.26

(21)申请号 201580069507.5

布莱恩·扎姆布罗维兹

(22)申请日 2015.12.18

安德鲁·J·墨菲

(30)优先权数据

62/094,104 2014.12.19 US

62/167,408 2015.05.28 US

62/205,524 2015.08.14 US

(74)专利代理机构 广州三环专利商标代理有限公司 44202

代理人 郝传鑫

(85)PCT国际申请进入国家阶段日

2017.06.19

(51)Int.Cl.

C12N 15/90(2006.01)

C12N 9/22(2006.01)

A01K 67/027(2006.01)

(86)PCT国际申请的申请数据

PCT/US2015/066681 2015.12.18

(87)PCT国际申请的公布数据

W02016/100819 EN 2016.06.23

(71)申请人 瑞泽恩制药公司

地址 美国纽约州塔里敦

权利要求书5页 说明书71页

序列表19页 附图4页

(72)发明人 维拉·佛洛妮娜 林恩·麦克唐纳

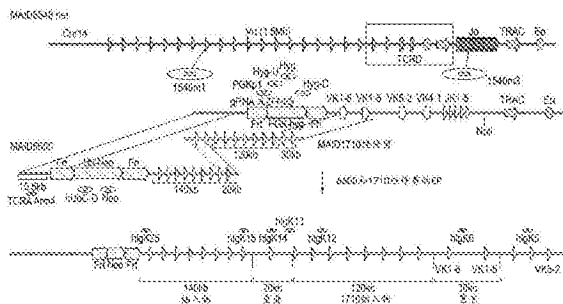
按照条约第19条修改的权利要求书5页

(54)发明名称

用于通过单步多重靶向进行靶向遗传修饰的方法和组合物

(57)摘要

本发明提供了方法和组合物,所述方法和组合物用于在细胞内的靶基因组基因座处进行一种或多种靶向遗传修饰,并用于产生包含所述经修饰的基因组基因座的非人动物。所述方法采用两个或更多个大靶向载体,所述两个或更多个大靶向载体能够在单个基因组靶向步骤中彼此重组且与所述靶基因组基因座重组。所述方法还可与核酸酶试剂结合使用。还提供了用于增强在细胞中靶基因组基因座处的同源重组的方法和组合物。所述方法采用包含一个或多个重叠序列的两个或更多个核酸。所述方法可与核酸酶试剂结合使用或不采用核酸酶试剂。



1. 一种用于修饰细胞中的靶基因组基因座的方法,包括:

(a) 向所述细胞中引入在所述靶基因组基因座内产生单链断裂或双链断裂的核酸酶试剂;

(b) 向所述细胞中引入第一大靶向载体 (LTVEC) 和第二LTVEC,所述第一大靶向载体为至少10kb长且包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物,并且所述第二LTVEC为至少10kb长且包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物,

其中所述第一LTVEC的所述第一3'同源臂具有与所述第二LTVEC的所述第二5'同源臂同源的第一重叠序列,并且所述第一LTVEC的所述第一5'同源臂和所述第二LTVEC的所述第二3'同源臂与所述靶基因组基因座内的对应基因组区段同源,

其中通过所述第一核酸插入物和所述第二核酸插入物在所述对应基因组区段之间的整合来修饰所述靶基因组基因座;以及

(c) 选择所靶向细胞,所述所靶向细胞包含整合到所述靶基因组基因座中的所述第一核酸插入物和所述第二核酸插入物。

2. 根据权利要求1所述的方法,其中所述第一核插入物和所述第一3'同源臂以及所述第二核酸插入物和第二5'同源臂是连续核酸的重叠片段,所述连续核酸通过整合所述第一核酸插入物和所述第二核酸插入物到所述靶基因组基因座内而重组。

3. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物或两者来自与所述细胞的物种不同的物种。

4. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物或两者都是人核酸。

5. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物和所述第二核酸插入物的组合大小为约50kb至约500kb,约50kb至约300kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,约275kb至约300kb,约300kb至约350kb,约350kb至约400kb,约400kb至约450kb,或约450kb至约500kb。

6. 根据权利要求5所述的方法,其中所述第一核酸插入物和所述第二核酸插入物的组合大小为约100kb至约500kb。

7. 根据权利要求6所述的方法,其中所述第一核酸插入物和所述第二核酸插入物的组合大小为约300kb。

8. 根据权利要求1-4中任一项所述的方法,其中所述所靶向细胞包含具有所述第一核酸插入物和所述第二核酸插入物两者的基因组DNA,所述第一核酸插入物和所述第二核酸插入物的组合大小在约5kb至约500kb的范围内。

9. 根据前述权利要求中任一项所述的方法,其中所述第一LTVEC的所述第一重叠序列与所述第二LTVEC的所述第一重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同源性。

10. 根据前述权利要求中任一项所述的方法,其中所述第一重叠序列的大小为约1kb至约70kb。

11. 根据前述权利要求中任一项所述的方法,其中所述第一重叠序列的大小为至少10kb或至少20kb。

12. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物或两者整合到所述靶基因组基因座中导致以下情况中的一种或多种:

(a) 在所述靶基因组基因座处加入外源序列;

(b) 在所述靶基因组基因座处缺失内源序列;以及

(c) 敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

13. 根据前述权利要求中任一项所述的方法,其中与使用单个LTVEC相比,所述第一LTVEC和所述第二LTVEC的组合使用导致靶向效率提升。

14. 根据权利要求13所述的方法,其中靶向效率的所述提升为至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍或20倍。

15. 根据权利要求1-14中任一项所述的方法,其中所述第一LTVEC或所述第二LTVEC的所述5'同源臂和所述3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。

16. 一种用于修饰细胞中的靶基因组基因座的方法,包括:

(a) 向所述细胞中引入在所述靶基因组基因座内产生单链断裂或双链断裂的核酸酶试剂;

(b) 向所述细胞中引入第一大靶向载体(LTVEC)、第二LTVEC和第三LTVEC,所述第一大靶向载体为至少10kb长且包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物,所述第二LTVEC为至少10kb长且包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物,并且所述第三LTVEC为至少10kb长且包含侧接有第三5'同源臂和第三3'同源臂的第三核酸插入物,

其中所述第一LTVEC的所述第一3'同源臂具有与所述第二LTVEC的所述第二5'同源臂同源的第一重叠序列,所述第二LTVEC的所述第二3'同源臂具有与所述第三LTVEC的所述第三5'同源臂同源的第二重叠序列,并且所述第一LTVEC的所述第一5'同源臂和所述第三LTVEC的所述第三3'同源臂与所述靶基因组基因座内的对应基因组区段同源,

其中通过所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物在所述对应基因组区段之间的整合来修饰所述靶基因组基因座;以及

(c) 选择所靶向细胞,所述所靶向细胞包含整合在所述靶基因组基因座中的所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物。

17. 根据权利要求16所述的方法,其中所述第一核插入物和所述第一3'同源臂以及所述第二核酸插入物和所述第二5'同源臂是连续核酸的重叠片段,并且所述第二核插入物和所述第二3'同源臂以及所述第三核酸插入物和所述第三5'同源臂是所述连续核酸的重叠片段,所述连续核酸通过整合所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物到所述靶基因组基因座内而重组。

18. 根据权利要求16或17所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物中的一者或多者来自与所述细胞的物种不同的物种。

19. 根据权利要求16-18中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物中的一者或多者是人核酸。

20. 根据权利要求16-19中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小为约50kb至约700kb,约50kb至约500kb,约50kb至约300kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,约275kb至约300kb,约300kb至约350kb,约350kb至约400kb,约400kb至约450kb,约450kb至约500kb,约500kb至约550kb,约550kb至约600kb,约600kb至约650kb,或约650kb至约700kb。

21. 根据权利要求20所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小为约100kb至约700kb。

22. 根据权利要求21所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小为约400kb。

23. 根据权利要求16-19中任一项所述的方法,其中所述所靶向细胞包含具有所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物三者的基因组DNA,所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小在约5kb至约700kb的范围内。

24. 根据权利要求16至23中任一项所述的方法,其中所述第一LTVEC的所述第一重叠序列与所述第二LTVEC的所述第一重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性,并且/或者所述第二LTVEC的所述第二重叠序列与所述第三LTVEC的所述第二重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。

25. 根据权利要求16-24中任一项所述的方法,其中所述第一重叠序列的大小为约1kb至约70kb,并且/或者所述第二重叠序列的大小为约1kb至约70kb。

26. 根据权利要求16-25中任一项所述的方法,其中所述第一重叠序列的大小为至少10kb或至少20kb,并且/或者所述第二重叠序列的大小为至少10kb或至少20kb。

27. 根据权利要求16-26中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物中的一者或多者整合到所述靶基因组基因座中导致以下情况中的一种或多种:

(a) 在所述靶基因组基因座处加入外源序列;

(b) 在所述靶基因组基因座处缺失内源序列;或

(c) 敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

28. 根据权利要求16-27中任一项所述的方法,其中所述第一LTVEC、所述第二LTVEC或所述第三LTVEC的所述5'同源臂和所述3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。

29. 根据权利要求12或27所述的方法,其中所述靶基因组基因座处的所述内源序列的所述缺失为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至

约800kb。

30. 一种用于增强细胞中的靶基因组基因座处的同源重组的方法, 包括向所述细胞内引入第一核酸和第二核酸, 其中所述第一核酸和所述第二核酸包含重叠核苷酸序列。

31. 根据权利要求30所述的方法, 其中在未使用核酸酶试剂的情况下, 所述靶基因组基因座处的同源重组增强。

32. 根据权利要求30所述的方法, 还包括向所述细胞内引入在所述靶基因组基因座处或附近产生单链断裂或双链断裂的核酸酶试剂。

33. 根据权利要求1-29和32中任一项所述的方法, 其中所述核酸酶试剂是锌指核酸酶(ZFN)、转录激活因子样效应物核酸酶(TALEN)或大范围核酸酶。

34. 根据权利要求1-29和32中任一项所述的方法, 其中所述核酸酶试剂包括成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白和向导RNA(gRNA)。

35. 根据权利要求34所述的方法, 其中所述Cas蛋白是Cas9。

36. 根据权利要求30-35中任一项所述的方法, 其中所述方法增强了在所述靶基因组基因座处所述第一核酸、所述第二核酸或两者的所述同源重组。

37. 根据权利要求36所述的方法, 其中所述同源重组的所述增强为至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍或20倍。

38. 根据权利要求30-37中任一项所述的方法, 其中所述第一核酸的所述重叠序列与所述第二核酸的所述重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。

39. 根据权利要求30-38中任一项所述的方法, 其中所述重叠序列为约1kb至约70kb。

40. 根据权利要求30-39中任一项所述的方法, 其中所述重叠序列为至少20kb。

41. 根据权利要求30-40中任一项所述的方法, 其中所述第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的靶向载体, 并且除了所述重叠序列以外, 所述第二核酸不包含与所述靶基因组基因座同源的核苷酸序列。

42. 根据权利要求30-40中任一项所述的方法, 其中所述第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一靶向载体, 并且所述第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二靶向载体。

43. 根据权利要求41或42所述的方法, 其中所述第一靶向载体是在约20kb至约200kb范围内的第一大靶向载体(LTVEC)。

44. 根据权利要求42所述的方法, 其中所述第二靶向载体是在约20kb至约200kb范围内的第二大靶向载体(LTVEC)。

45. 根据权利要求43或44所述的方法, 其中所述第一LTVEC的所述5'同源臂和所述3'同源臂的总和为10kb至约200kb。

46. 根据权利要求44或45所述的方法, 其中所述第二LTVEC的所述5'同源臂和所述3'同源臂的总和为10kb至约200kb。

47. 根据权利要求30-46中任一项所述的方法, 其中所述重叠序列位于所述第一核酸的3'端处和所述第二核酸序列的5'端处。

48. 根据权利要求30-47中任一项所述的方法, 其中所述重叠核苷酸序列有利于将重组

机制募集到所述靶基因组基因座处。

49. 根据前述权利要求中任一项所述的方法,其中所述细胞是人细胞。

50. 根据权利要求1-48中任一项所述的方法,其中所述细胞是非人细胞。

51. 根据前述权利要求中任一项所述的方法,其中所述细胞是多能细胞、造血干细胞、神经元干细胞或成纤维细胞。

52. 根据权利要求51所述的方法,其中所述多能细胞是胚胎干(ES)细胞或诱导多能干(iPS)细胞。

53. 根据权利要求50-52中任一项所述的方法,其中所述细胞是哺乳动物细胞。

54. 根据权利要求53所述的方法,其中所述哺乳动物细胞是啮齿动物细胞。

55. 根据权利要求54所述的方法,其中所述啮齿动物是小鼠或大鼠。

56. 一种用于产生F0代非人动物的方法,包括:

(a) 将非人ES细胞引入非人宿主胚胎,其中所述非人ES细胞是通过根据权利要求1-48和50-55中任一项所述的方法产生的;以及

(b) 在代孕母体中孕育所述非人宿主胚胎,

其中所述代孕母体产生包含所述修饰的所述F0代非人动物。

57. 根据权利要求56所述的方法,其中所述非人动物是小鼠或大鼠。

用于通过单步多重靶向进行靶向遗传修饰的方法和组合物

[0001] 相关申请的交叉引用

[0002] 本申请涉及2014年12月19日提交的美国申请号62/094,104、2015年5月28日提交的美国申请号62/167,408以及2015年8月14日提交的美国申请号62/205,524,这些申请中的每一个出于所有目的全文以引用方式并入本文。

[0003] 对通过EFS WEB作为文本文件提交的序列表的引用

[0004] 文件472224SEQLIST.txt中所写序列表大小为16.7kb,创建于2015年12月16日,该序列表据此全文以引用方式并入。

背景技术

[0005] 使用被设计成在基因组基因座处添加、缺失或取代特定核酸序列的靶向载体进行的同源重组是在非人动物中实现所需基因组修饰的常用方法。

[0006] 虽然通过同源重组的基因组修饰的技术在过去二十年内已取得显著进展,但在许多情况下,例如当啮齿动物基因组的很大一部分被大的人类基因组片段取代,或靶向某些细胞类型,例如成纤维细胞或其他体细胞时,使用非常大的靶向载体LTVEC实现可接受的靶向效率仍然存在困难。

发明内容

[0007] 提供了用于通过靶向系统修饰细胞内的靶基因组基因座的方法和组合物,该靶向系统利用能够彼此重组以形成单个连续核酸区段的两个或更多个靶向载体。任选地,靶向载体是大靶向载体(LTVEC)。任选地,LTVEC各自的大小为至少10kb。

[0008] 本发明提供了用于修饰细胞中的靶基因组基因座的方法,该方法包括:(a)向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂;(b)向细胞内引入包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一大靶向载体(LTVEC)以及包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二LTVEC,其中第一LTVEC的第一5'同源臂和第二LTVEC的第二3'同源臂与靶基因组基因座内的相应基因组区段同源,并且第一LTVEC的第一3'同源臂和第二LTVEC的5'同源臂彼此同源或分别与一个或多个另外的LTVEC的另外5'和3'同源臂同源,所述另外的LTVEC各自包含侧接有另外的5'同源臂和另外的3'同源臂的另外核酸插入物,其中通过在相应的基因组区段之间整合该第一核酸插入物、该一个或多个另外的LTVEC的另外核酸插入物(如果存在的话)以及该第二核酸插入物来修饰靶基因组基因座;以及(c)选择包含整合在靶基因组基因座中的该第一核酸插入物、该一个或多个另外核酸插入物(如果存在的话)以及该第二核酸插入物的所靶向细胞(targeted cell)。任选地,第一LTVEC、第二LTVEC和一个或多个另外的LTVEC的大小各自为至少10kb。在一些此类方法中,所述另外的LTVEC是当存在时被插入在第一LTVEC和第二LTVEC之间的一个或多个另外的LTVEC。

[0009] 本发明还提供了用于修饰细胞中的靶基因组基因座的双重靶向方法,该方法包括:(a)向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂;(b)向细胞

内引入第一大靶向载体 (LTVEC) 和第二大LTVEC, 该第一LTVEC的大小为至少10kb并且包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物, 该第二LTVEC的长度为至少10kb并且包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物, 其中第一LTVEC的第一3'同源臂具有与第二LTVEC的第二5'同源臂同源的第一重叠序列, 并且第一LTVEC的第一5'同源臂和第二LTVEC的第二3'同源臂与靶基因组基因座内的相应基因组区段同源, 其中通过在相应的基因组区段之间整合该第一核酸插入物和该第二核酸插入物来修饰靶基因组基因座; 以及 (c) 选择包含整合到靶基因组基因座内的该第一核酸插入物和该第二核酸插入物的所靶向细胞。

[0010] 任选地, 第一核酸插入物和第一3'同源臂以及第二核酸插入物和第二5'同源臂是连续核酸 (contiguous nucleic acid) 的重叠片段, 该连续核酸通过整合第一核酸插入物和第二核酸插入物到靶基因组基因座内而重组 (reform)。

[0011] 在一些此类方法中, 所述细胞为人细胞。在其他此类方法中, 所述细胞为非人细胞。在一些此类方法中, 所述细胞为多能细胞、造血干细胞、神经元干细胞或成纤维细胞。任选地, 多能细胞为胚胎干 (ES) 细胞或诱导多能干 (iPS) 细胞。在一些此类方法中, 所述细胞为哺乳动物细胞。任选地, 哺乳动物细胞为啮齿动物细胞。任选地, 啮齿动物细胞为小鼠细胞或大鼠细胞。

[0012] 在上述一些方法中, 核酸酶试剂是锌指核酸酶 (ZFN)、转录激活子样效应子核酸酶 (TALEN) 或大范围核酸酶。在上述一些方法中, 核酸酶试剂包括成簇规律间隔短回文重复序列 (CRISPR) 相关 (Cas) 蛋白和向导RNA (gRNA)。任选地, Cas蛋白为Cas9。

[0013] 在一些方法中, 第一核酸插入物、第二核酸插入物或两者来自与所述细胞的物种不同的物种。在一些方法中, 第一核酸插入物、第二核酸插入物或两者都是人核酸。

[0014] 在一些方法中, 第一核酸插入物和第二核酸插入物的组合大小为约50kb至约500kb, 约50kb至约300kb, 约50kb至约75kb, 约75kb至约100kb, 约100kb至125kb, 约125kb至约150kb, 约150kb至约175kb, 约175kb至约200kb, 约200kb至约225kb, 约225kb至约250kb, 约250kb至约275kb, 约275kb至约300kb, 约300kb至约350kb, 约350kb至约400kb, 约400kb至约450kb, 或约450kb至约500kb。任选地, 第一核酸插入物和第二核酸插入物的组合大小为约100kb至约500kb。任选地, 第一核酸插入物和第二核酸插入物的组合大小为约300kb。

[0015] 在一些方法中, 所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA, 第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。

[0016] 在一些方法中, 第一LTVEC的第一重叠序列与第二LTVEC的第一重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。在一些方法中, 第一重叠序列的大小为约1kb至约70kb。在一些方法中, 第一重叠序列的大小为至少10kb或至少20kb。

[0017] 在一些方法中, 将第一核酸插入物、第二核酸插入物或两者整合到靶基因组基因座中导致以下情况中的一种或多种: (a) 在靶基因组基因座处加入外源序列; (b) 在靶基因组基因座处缺失内源序列; 或者 (c) 敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换, 或它们的组合。任选地, 靶基因组基因座处的内源序列缺失为约5kb至约10kb, 约10kb至约20kb, 约20kb至约40kb, 约40kb至约60kb, 约60kb至约80kb, 约80kb至约100kb, 约100kb至约150kb, 约150kb至约200kb, 约200kb至约300kb, 约300kb至约

400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至约800kb。

[0018] 在一些方法中,与使用单个LTVEC相比,第一LTVEC和第二LTVEC的组合使用导致靶向效率(targeting efficiency)提升。任选地,靶向效率提升至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍,或20倍。

[0019] 在一些方法中,第一LTVEC或第二LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为约10kb至约150kb,并且第二LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在一些方法中,第一LTVEC或第二LTVEC的5'和3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb,并且第二LTVEC的5'和3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。

[0020] 本发明还提供了用于产生F0代非人动物的方法,包括:(a)将非人ES细胞引入非人宿主胚胎中,其中该非人ES细胞是通过上述方法中的任意一种产生的;以及(b)在代孕母体中孕育该非人宿主胚胎,其中代孕母体产生包含修饰的F0代非人动物。任选地,非人动物是小鼠或大鼠。

[0021] 本发明还提供了用于修饰细胞中的靶基因组基因座的三重靶向方法,该方法包括:(a)向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂;(b)向细胞内引入第一大靶向载体(LTVEC)、第二LTVEC和第三LTVEC,该第一LTVEC的长度为至少10kb并且包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物,该第二LTVEC的长度为至少10kb并且包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物,该第三LTVEC的长度为至少10kb并且包含侧接有第三5'同源臂和第三3'同源臂的第三核酸插入物,其中该第一LTVEC的第一3'同源臂具有与该第二LTVEC的第二5'同源臂同源的第一重叠序列,该第二LTVEC的第二3'同源臂具有与该第三LTVEC的第三5'同源臂同源的第三重叠序列,并且该第一LTVEC的第一5'同源臂和该第三LTVEC的第三3'同源臂与靶基因组基因座内的相应基因组区段同源,其中通过在相应基因组区段之间整合该第一核酸插入物、该第二核酸插入物和该第三核酸插入物来修饰靶基因组基因座;以及(c)选择包含整合到靶基因组基因座内的该第一核酸插入物、该第二核酸插入物和该第三核酸插入物的所靶向细胞。

[0022] 任选地,第一核酸插入物和第一3'同源臂以及第二核酸插入物和第二5'同源臂是连续核酸的重叠片段,并且第二核酸插入物和第二3'同源臂以及第三核酸插入物和第三5'同源臂是连续核酸的重叠片段,该连续核酸通过整合第一核酸插入物、第二核酸插入物和第三核酸插入物到靶基因组基因座内而重组。

[0023] 在一些此类方法中,所述细胞为人细胞。在其他此类方法中,所述细胞为非人细胞。在一些此类方法中,所述细胞为多能细胞、造血干细胞、神经元干细胞或成纤维细胞。任选地,多能细胞为胚胎干(ES)细胞或诱导多能干(iPS)细胞。在一些此类方法中,所述细胞为哺乳动物细胞。任选地,哺乳动物细胞为啮齿动物细胞。任选地,啮齿动物细胞为小鼠细胞或大鼠细胞。

[0024] 在一些此类方法中,核酸酶试剂是锌指核酸酶(ZFN)、转录激活子样效应子核酸酶(TALEN)或大范围核酸酶。在一些此类方法中,核酸酶试剂包括成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白和向导RNA(gRNA)。任选地,Cas蛋白为Cas9。

[0025] 在一些此类方法中,第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者来自与所述细胞的物种不同的物种。在一些此类方法中,第一核酸插入物、第二核酸插入物和第三核酸插入物都是人核酸。

[0026] 在一些此类方法中,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小为约50kb至约700kb,约50kb至约500kb,约50kb至约300kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,约275kb至约300kb,约300kb至约350kb,约350kb至约400kb,约400kb至约450kb,约450kb至约500kb,约500kb至约550kb,约550kb至约600kb,约600kb至约650kb,或约650kb至约700kb。任选地,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小为约100kb至约700kb。任选地,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小为约400kb。

[0027] 在一些此类方法中,所靶向细胞包含具有第一核酸插入物、第二核酸插入物和第三核酸插入物的基因组DNA,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小在约5kb至约700kb的范围内。

[0028] 在一些此类方法中,第一LTVEC的第一重叠序列与第二LTVEC的第一重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性,并且/或者第二LTVEC的第二重叠序列与第三LTVEC的第二重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。在一些此类方法中,第一重叠序列的大小为约1kb至约70kb,并且/或者第二重叠序列的大小为约1kb至约70kb。在一些此类方法中,第一重叠序列的大小为至少10kb或至少20kb,并且/或者第二重叠序列的大小为至少10kb或至少20kb。

[0029] 在一些此类方法中,将第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。任选地,靶基因组基因座处的内源序列缺失为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至约800kb。

[0030] 在一些方法中,第一LTVEC、第二LTVEC或第三LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为约10kb至约150kb,第二LTVEC的5'和3'同源臂的总和为约10kb至约150kb,并且第三LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在一些此类方法中,第一LTVEC、第二LTVEC或第三LTVEC的5'和3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约

80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb;第二LTVEC的5'和3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb;并且第三LTVEC的5'和3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。

[0031] 本发明还提供了用于产生F0代非人动物的方法,包括:(a)将非人ES细胞引入非人宿主胚胎中,其中该非人ES细胞是通过上述方法中的任何一种产生的;以及(b)在代孕母体中孕育该非人宿主胚胎;其中代孕母体产生包含修饰的F0代非人动物。任选地,非人动物是小鼠或大鼠。

[0032] 本发明还提供了用于增强细胞中的靶基因组基因座处的同源重组的方法,包括向细胞内引入第一核酸和第二核酸,其中该第一核酸和第二核酸包含重叠核苷酸序列。在一些此类方法中,与仅将单个核酸引入细胞中的方法相比,同源重组得到增强。

[0033] 在一些此类方法中,在未使用核酸酶试剂的情况下,靶基因组基因座处的同源重组增强。一些此类方法还包括向细胞内引入在靶基因组基因座处或其附近造成单链或双链断裂的核酸酶试剂。在一些此类方法中,核酸酶试剂是锌指核酸酶(ZFN)、转录激活因子样效应子核酸酶(TALEN)或大范围核酸酶。在一些此类方法中,核酸酶试剂包括成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白和向导RNA(gRNA)。任选地,Cas蛋白为Cas9。

[0034] 在一些此类方法中,该方法增强了在靶基因组基因座处第一核酸、第二核酸、或两者的同源重组。与其中引入第一核酸而不引入第二核酸的方法相比,一些此类方法增强了靶基因组基因座处第一核酸的同源重组。与其中引入第二核酸而不引入第一核酸的方法相比,一些此类方法增强了靶基因组基因座处第二核酸的同源重组。任选地,同源重组增强至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍,或20倍。

[0035] 在一些此类方法中,第一核酸的重叠序列与第二核酸的重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。在一些此类方法中,重叠序列为约1kb至约70kb。任选地,重叠序列为约1kb至约5kb,约5kb至约10kb,约10kb至约15kb,约15kb至约20kb,约20kb至约25kb,约25kb至约30kb,约30kb至约35kb,约35kb至约40kb,约40kb至约45kb,约45kb至约50kb,约50kb至约55kb,约55kb至约60kb,约60kb至约65kb,或约65kb至约70kb。在一些此类方法中,重叠序列为至少5kb、至少10kb、至少15kb、至少20kb、至少25kb、至少30kb、至少35kb、至少40kb、至少45kb、至少50kb、至少55kb、至少60kb、至少65kb,或至少70kb。任选地,重叠序列为至少20kb。

[0036] 在一些此类方法中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的靶向载体,并且除了所述重叠序列以外,第二核酸不包含与靶基因组基因座同源的核苷酸序列。任选地,第一靶向载体为约1kb至约2kb,约2kb至约5kb,或约5kb至约10kb。任选地,第一靶向载体为第一大靶向载体(LTVEC)。任选地,第一LTVEC的长度为至少10kb。任选地,第一靶向载体为在约20kb至约200kb的范围内的第一大靶向载体(LTVEC)。任选地,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb。

[0037] 在一些此类方法中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一靶向载体,并且第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的

第二核酸插入物的第二靶向载体。任选地,第一靶向载体为约1kb至约2kb,约2kb至约5kb,或约5kb至约10kb,并且/或者第二靶向载体为约1kb至约2kb,约2kb至约5kb,或约5kb至约10kb。任选地,第一靶向载体是第一大靶向载体(LTVEC),并且/或者第二靶向载体是第二大靶向载体(LTVEC)。任选地,第一LTVEC的长度为至少10kb,并且/或者第二LTVEC的长度为至少10kb。任选地,第一靶向载体是在约20kb至约200kb的范围内的第一大靶向载体(LTVEC),并且/或者第二靶向载体是在约20kb至约200kb的范围内的第二大靶向载体(LTVEC)。任选地,第一LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb,并且/或者第二LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。任选地,第一LTVEC或第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。任选地,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb,并且第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。

[0038] 在一些方法中,重叠序列位于第一核酸的3'端处和第二核酸序列的5'端处。在一些方法中,重叠核苷酸序列有利于将重组机制(recombination machinery)募集到靶基因组基因座。

[0039] 在一些此类方法中,所述细胞为人细胞。在其他此类方法中,所述细胞为非人细胞。在一些此类方法中,所述细胞为多能细胞、造血干细胞、神经元干细胞或成纤维细胞。任选地,多能细胞为胚胎干(ES)细胞或诱导多能干(iPS)细胞。在一些此类方法中,所述细胞为哺乳动物细胞。任选地,哺乳动物细胞为啮齿动物细胞。任选地,啮齿动物细胞为小鼠细胞或大鼠细胞。

[0040] 本发明还提供了用于产生F0代非人动物的方法,包括:(a)将非人ES细胞引入非人宿主胚胎中,其中该非人ES细胞是通过上述方法中任意一种产生的;以及(b)在代孕母体中孕育该非人宿主胚胎;其中代孕母体产生包含修饰的F0代非人动物。任选地,非人动物是小鼠或大鼠。

[0041] 提供了用于通过靶向系统修饰细胞内的靶基因组基因座的方法和组合物,该靶向系统利用能够彼此重组以形成单个连续核酸区段的两个或更多个靶向载体。在各种实施方案中,靶向载体是大靶向载体(LTVEC)。任选地,LTVEC各自的大小为至少10kb。

[0042] 在一个实施方案中,提供了一种用于修饰细胞中的靶基因组基因座的方法。此方法包括向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂;引入包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一大靶向载体(LTVEC)以及包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二LTVEC,其中第一LTVEC的第一5'同源臂和第二LTVEC的第二3'同源臂与靶基因组内的相应区段同源,并且第一LTVEC的第一3'同源臂和第二LTVEC的5'同源臂彼此同源或分别与一个或多个另外的LTVEC的另外5'和3'同源臂同源,所述另外的LTVEC各自包含侧接有另外的5'同源臂和另外的3'同源臂的另外插入物,其中通过在相应的基因组区段之间整合该第一插入物、该一个或多个另外的LTVEC的另外插入物(如果存在的话)以及该第二核酸插入物来修饰靶基因组基因座。任选地,第一LTVEC、第二LTVEC和一个或多个另外的LTVEC的大小各自为至少10kb。该方法还包括选择所靶向细胞,该所靶向细胞包含整合在靶基因组基因座中的第一核酸插入物、一个或多个另外的核酸插入物(如果存在的话)以及第二核酸插入物。在此类方法中,所述另

外的LTVEC是当存在时被插入在第一LTVEC和第二LTVEC之间的一个或多个另外的LTVEC。

[0043] 在另一个实施方案中,提供了一种用于修饰细胞中的靶基因组基因座的双重靶向方法。此类方法包括:向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂,引入第一大靶向载体(LTVEC)和第二LTVEC,该第一LTVEC包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物,并且该第二LTVEC包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物。任选地,第一LTVEC的大小为至少10kb,并且第二LTVEC的大小为至少10kb。在此类方法中,第一LTVEC的第一3'同源臂具有与第二LTVEC的第二5'同源臂同源的第一重叠序列,并且第一LTVEC的第一5'同源臂和第二LTVEC的第二3'同源臂与靶基因组内的对应区段同源,以便通过在相应基因组区段之间整合第一核酸插入物和第二核酸插入物来修饰靶基因组基因座。该方法还包括选择所靶向细胞,该所靶向细胞包含整合在靶基因组基因座中的第一核酸插入物和第二核酸插入物。

[0044] 在一些此类方法中,第一核酸插入物 and 第一3'同源臂以及第二核酸插入物 and 第二5'同源臂是连续核酸的重叠片段,该连续核酸通过整合第一核酸插入物 and 第二核酸插入物到靶基因组基因座内而重组。

[0045] 在另一个实施方案中,提供了一种用于修饰细胞中的靶基因组基因座的三重靶向方法。此类方法包括:向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂,引入第一大靶向载体(LTVEC)、第二LTVEC和第三LTVEC,该第一LTVEC包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物,该第二LTVEC包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物,并且该第三LTVEC包含侧接有第三5'同源臂和第三3'同源臂的第三核酸插入物。任选地,第一LTVEC的大小为至少10kb,第二LTVEC的大小为至少10kb,并且第三LTVEC的大小为至少10kb。在此类方法中,第一LTVEC的第一3'同源臂具有与第二LTVEC的第二5'同源臂同源的第一重叠序列,第二LTVEC的第二3'同源臂具有与第三LTVEC的第三5'同源臂同源的第二重叠序列,并且第一LTVEC的第一5'同源臂和第三LTVEC的第三3'同源臂与靶基因组内的对应区段同源,以便通过在相应基因组区段之间整合第一核酸插入物、第二核酸插入物和第三核酸插入物来修饰靶基因组基因座。该方法还包括选择所靶向细胞,该所靶向细胞包含整合在靶基因组基因座中的该第一核酸插入物、该第二核酸插入物 and 该第三核酸插入物。

[0046] 在一些此类方法中,第一核酸插入物 and 第一3'同源臂以及第二核酸插入物 and 第二5'同源臂是连续核酸的重叠片段,并且第二核酸插入物 and 第二3'同源臂以及第三核酸插入物 and 第三5'同源臂是连续核酸的重叠片段,该连续核酸是通过整合第一核酸插入物、第二核酸插入物 and 第三核酸插入物到靶基因组基因座内而重组。

[0047] 在一个实施方案中,所述细胞为多能细胞。在另一个实施方案中,多能细胞为胚胎干(ES)细胞。在一些实施方案中,多能细胞为造血干细胞或神经元干细胞。在另一个实施方案中,所述细胞为诱导性多能干(iPS)细胞。

[0048] 在一个实施方案中,靶基因组基因座位于细胞的基因组中。在另一个实施方案中,靶基因组基因座位于细胞内的染色体外DNA上。

[0049] 在一个实施方案中,所述细胞为成纤维细胞。

[0050] 在一些方法中,所述细胞为非人细胞。在其他方法中,所述细胞来自人。在一些实施方案中,该细胞是哺乳动物细胞。在另一个实施方案中,该哺乳动物细胞来自啮齿动物。

在一些情况下,啮齿动物为小鼠、大鼠或仓鼠。

[0051] 在上述一些方法中,核酸酶试剂由包含编码核酸酶的核酸序列的表达构建体表达,并且其中所述核酸有效连接至细胞中有活性的启动子。在其他方法中,核酸酶试剂由编码核酸酶的mRNA表达。在一些此类方法中,核酸酶是锌指核酸酶(ZFN)。在其他此类方法中,核酸酶为转录激活因子样效应物核酸酶(TALEN)。在另外的方法中,核酸酶为大范围核酸酶。

[0052] 在上述一些方法中,核酸酶试剂包括成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白和向导RNA(gRNA)。在一些此类方法中,Cas蛋白为Cas9。

[0053] 在上述一些方法中,第一核酸插入物、第二核酸插入物或两者来自与所述细胞的物种不同的物种。在一个实施方案中,第一核酸插入物、第二核酸插入物和/或第三核酸插入物来自不同的物种。在一些方法中,第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者来自与所述细胞的物种不同的物种。在一些方法中,第一核酸插入物、第二核酸插入物或两者都是人核酸。在另一个实施方案中,第一核酸插入物、第二核酸插入物和/或第三核酸插入物都是人核酸。在一些方法中,第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者是人核酸。

[0054] 在一个实施方案中,第一核酸插入物和第二核酸插入物的组合大小为约50kb至约500kb,约50kb至约300kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,约275kb至约300kb,约300kb至约350kb,约350kb至约400kb,约400kb至约450kb,或约450kb至约500kb。在另一个实施方案中,第一核酸插入物和第二核酸插入物的组合大小为约100kb至约500kb。在另一个实施方案中,第一核酸插入物和第二核酸插入物的组合大小为约300kb。

[0055] 在一些实施方案中,所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA,第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。

[0056] 在一个实施方案中,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小为约50kb至约700kb,约50kb至约500kb,约50kb至约300kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,约275kb至约300kb,约300kb至约350kb,约350kb至约400kb,约400kb至约450kb,约450kb至约500kb,约500kb至约550kb,约550kb至约600kb,约600kb至约650kb,或约650kb至约700kb。

[0057] 在一些实施方案中,所靶向细胞包含具有第一核酸插入物、第二核酸插入物和第三核酸插入物的基因组DNA,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小在约5kb至约700kb的范围内。任选地,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小为约100kb至约700kb。在一些实施方案中,第一核酸插入物、第二核酸插入物和第三核酸插入物的组合大小为约400kb。

[0058] 在上述一些方法中,第一LTVEC的第一重叠序列与第二LTVEC的第一重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。在上述一些方法中,第二LTVEC的第二重叠序列与第三LTVEC的第二重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。在上

述方法中的任一方法中,重叠序列为约1kb至约70kb。在一个具体实施方案中,重叠序列为至少10kb。在另一个具体实施方案中,重叠序列为至少20kb。在上述一些方法中,第一重叠序列和/或第二重叠序列为约1kb至约70kb。在一些方法中,第一重叠序列和/或第二重叠序列为至少10kb或至少20kb。

[0059] 在一些方法中,将第一核酸插入物、第二核酸插入物或两者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。在一些方法中,将第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

[0060] 在一些方法中,将第一核酸插入物、第二核酸插入物或两者整合到靶基因组基因座中导致靶基因组基因座处加入外源序列。在一个实施方案中,将第一核酸插入物、第二核酸插入物和/或第三核酸插入物整合到靶基因组基因座导致靶基因组基因座处加入外源序列。

[0061] 在一些方法中,将第一核酸插入物、第二核酸插入物或两者整合到靶基因组基因座中导致靶基因组基因座处缺失内源序列。在另一个实施方案中,将第一核酸插入物、第二核酸插入物和/或第三核酸插入物整合到靶基因组基因座导致靶基因组基因座处缺失内源序列。在一些此类方法中,靶基因组基因座处内源序列的缺失为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,或约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至约800kb。

[0062] 在一些方法中,将第一核酸插入物、第二核酸插入物或者这两种插入物整合到靶基因组基因座中导致敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。在另一个实施方案中,将第一核酸插入物、第二核酸插入物和/或第三核酸插入物整合到靶基因组基因座中导致敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

[0063] 在上述一些方法中,与使用单个LTVEC相比,第一LTVEC和第二LTVEC的组合使用导致靶向效率提升。任选地,靶向效率提升至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍,或20倍。

[0064] 在一些方法中,第一LTVEC或第二LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为约10kb至约150kb,并且第二LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在一些实施方案中,第一LTVEC、第二LTVEC或第三LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为约10kb至约150kb,第二LTVEC的5'和3'同源臂的总和为约10kb至约150kb,并且第三LTVEC的5'和3'同源臂的总和为约10kb至约150kb。在其他实施方案中,第一LTVEC、第二LTVEC或第三LTVEC的5'和3'同源臂的总和为约1kb至约5kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约

120kb,或约120kb至约150kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为约1kb至约5kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至150kb;第二LTVEC的5'和3'同源臂的总和为约1kb至约5kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至150kb;并且第三LTVEC的5'和3'同源臂的总和为约1kb至约5kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至150kb。

[0065] 还提供了一种产生F0代非人动物的方法。此类方法包括:将非人ES细胞引入非人宿主胚胎中,其中该非人ES细胞是通过上述方法中的任意一种产生的;以及在代孕母体中孕育该非人宿主胚胎,使得代孕母体产生包含修饰的F0代非人动物。任选地,非人动物是小鼠或大鼠。

[0066] 本发明还提供了用于增强细胞中的靶基因组基因座处的同源重组的方法,包括向细胞内引入第一核酸和第二核酸,其中该第一核酸和第二核酸包含重叠核苷酸序列。在一些此类方法中,与仅将单个核酸引入细胞中的方法相比,同源重组得到增强。在一些此类方法中,在未使用核酸酶试剂的情况下,靶基因组基因座处的同源重组增强。其他此类方法还包括向细胞内引入在靶基因组基因座处或其附近造成单链或双链断裂的核酸酶试剂。

[0067] 在一个方面,提供了一种在不使用核酸酶试剂的情况下增强细胞中基因组基因座处的同源重组的方法,其包括向细胞内引入第一核酸和第二核酸,其中该第一核酸和第二核酸包含重叠核苷酸序列。

[0068] 在一个实施方案中,该方法增强了在靶基因组基因座处第一核酸的同源重组。与其中引入第一核酸而不引入第二核酸的方法相比,一些此类方法增强了靶基因组基因座处第一核酸的同源重组。在一个实施方案中,该方法增强了在靶基因组基因座处第二核酸的同源重组。与其中引入第二核酸而不引入第一核酸的方法相比,一些此类方法增强了靶基因组基因座处第二核酸的同源重组。在一个实施方案中,该方法增强了在靶基因组基因座处第一核酸和第二核酸的同源重组。

[0069] 在一个实施方案中,同源重组增强至少0.5倍、1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍,或20倍。

[0070] 在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列同源。在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%,或99.9%的同一性。在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列具有100%的同一性。

[0071] 在一个实施方案中,重叠序列为约1kb至约70kb。在一些方法中,重叠序列为至少20kb。在一个实施方案中,重叠序列为约1kb至约5kb。在一个实施方案中,重叠序列为约5kb至约10kb。在一个实施方案中,重叠序列为约10kb至约15kb。在一个实施方案中,重叠序列为约15kb至约20kb。在一个实施方案中,重叠序列为约20kb至约25kb。在一个实施方案中,重叠序列为约25kb至约30kb。在一个实施方案中,重叠序列为约30kb至约35kb。在一个实施方案中,重叠序列为约35kb至约40kb。在一个实施方案中,重叠序列为约40kb至约45kb。在一个实施方案中,重叠序列为约45kb至约50kb。在一个实施方案中,重叠序列为约50kb至约55kb。在一个实施方案中,重叠序列为约55kb至约60kb。在一个实施方案中,重叠序列为约

60kb至约65kb。在一个实施方案中,重叠序列为约65kb至约70kb。

[0072] 在一个实施方案中,重叠序列为至少5kb。在一个实施方案中,重叠序列为至少10kb。在一个实施方案中,重叠序列为至少15kb。在一个实施方案中,重叠序列为至少20kb。在一个实施方案中,重叠序列为至少25kb。在一个实施方案中,重叠序列为至少30kb。在一个实施方案中,重叠序列为至少35kb。在一个实施方案中,重叠序列为至少40kb。在一个实施方案中,重叠序列为至少45kb。在一个实施方案中,重叠序列为至少50kb。在一个实施方案中,重叠序列为至少55kb。在一个实施方案中,重叠序列为至少60kb。在一个实施方案中,重叠序列为至少65kb。在一个实施方案中,重叠序列为至少70kb。

[0073] 在一个实施方案中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的靶向载体,并且除了所述重叠序列以外,第二核酸不包含与基因组基因座同源的核苷酸序列。

[0074] 在一个实施方案中,第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二靶向载体,并且除了所述重叠序列以外,第一核酸不包含与基因组基因座同源的核苷酸序列。

[0075] 在一个实施方案中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一靶向载体,并且第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二靶向载体。在一个实施方案中,第一核酸插入物和第二核酸插入物是连续核酸的重叠片段。

[0076] 在一个实施方案中,靶向载体为约1kb至约2kb。在一个实施方案中,靶向载体为约2kb至约5kb。在一个实施方案中,靶向载体为约5kb至约10kb。

[0077] 在一个实施方案中,靶向载体是大靶向载体(LTVEC)。在一些方法中,靶向载体是在约20kb至约200kb的范围内的LTVEC。在一些方法中,第一靶向载体是在约20kb至约200kb的范围内的第一LTVEC,并且/或者第二靶向载体是在约20kb至约200kb的范围内的第二LTVEC。在一个实施方案中,LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。在一些方法中,第一靶向载体是第一LTVEC,并且/或者第二靶向载体是第二LTVEC。在一些方法中,第一LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。在一些方法中,第二LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。在一些方法中,第一LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb,并且第二LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。

[0078] 在一个实施方案中,LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb,并且第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。

[0079] 在一个实施方案中,重叠序列位于第一核酸的3'端处和第二核酸序列的5'端处。在一个实施方案中,重叠序列位于第一核酸的5'端处和第二核酸序列的3'端处。

[0080] 在一个实施方案中,第一核酸插入物和/或第二核酸插入物来自不同的物种。在另一个实施方案中,第一核酸插入物和/或第二核酸插入物是人核酸。在一些方法中,第一核酸插入物、第二核酸插入物或两者来自与所述细胞的物种不同的物种。在一些方法中,第一核酸插入物、第二核酸插入物或两者都是人核酸。

[0081] 在一些方法中,将第一核酸插入物、第二核酸插入物或两者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。在一些方法中,将第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

[0082] 在一个实施方案中,将第一插入物和/或第二插入物整合到基因组基因座导致该基因组基因座处加入外源序列。

[0083] 在一些实施方案中,所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA,第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。在一些方法中,所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA,第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。

[0084] 在另一个实施方案中,将第一核酸插入物和/或第二核酸插入物整合到基因组基因座导致靶基因组基因座处缺失内源序列。在一个实施方案中,靶基因组基因座处的内源序列缺失为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,或约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至约800kb。

[0085] 在一些方法中,所述细胞为人细胞。在其他方法中,所述细胞为非人细胞。在一些方法中,所述细胞为多能细胞、造血干细胞、神经元干细胞或成纤维细胞。任选地,多能细胞为胚胎干(ES)细胞或诱导多能干(iPS)细胞。在一些方法中,所述细胞为哺乳动物细胞。任选地,哺乳动物细胞为啮齿动物细胞。任选地,啮齿动物细胞为小鼠细胞或大鼠细胞。

[0086] 在一个实施方案中,所述细胞为多能细胞。在另一个实施方案中,多能细胞为胚胎干(ES)细胞。在一些实施方案中,多能细胞为造血干细胞或神经元干细胞。在另一个实施方案中,所述细胞为诱导性多能干(iPS)细胞。

[0087] 在一个实施方案中,靶基因组基因座位于细胞的基因组中。在另一个实施方案中,靶基因组基因座位于细胞内的染色体外DNA上。

[0088] 在一个实施方案中,所述细胞为成纤维细胞。

[0089] 在一些方法中,所述细胞为非人细胞。在其他方法中,所述细胞来自人。在一些实施方案中,该细胞是哺乳动物细胞。在另一个实施方案中,该哺乳动物细胞来自啮齿动物。在一些情况下,啮齿动物为小鼠、大鼠或仓鼠。

[0090] 在上述一些方法中,重叠核苷酸序列有利于将重组机制募集到靶基因组基因座处。

[0091] 本发明还提供了用于产生F0代非人动物的方法,包括:(a)将非人ES细胞引入非人宿主胚胎中,其中该非人ES细胞是通过上述方法产生的;以及(b)在代孕母体中孕育该非人宿主胚胎;其中代孕母体产生包含修饰的F0代非人动物。任选地,非人动物是小鼠或大鼠。

[0092] 在另一方面,提供了一种用于在使用核酸酶试剂的情况下增强细胞中靶基因组基因座处的同源重组的方法,其包括将以下物质引入该细胞中:(i)第一核酸和第二核酸,其中第一核酸和第二核酸包含重叠核苷酸序列;以及(ii)在基因组基因座处或附近造成单链或双链断裂的核酸酶试剂。

[0093] 在一个实施方案中,该方法增强了在靶基因组基因座处第一核酸的同源重组。与其中引入第一核酸而不引入第二核酸的方法相比,一些此类方法增强了靶基因组基因座处第一核酸的同源重组。在一个实施方案中,该方法增强了在靶基因组基因座处第二核酸的同源重组。与其中引入第二核酸而不引入第一核酸的方法相比,一些此类方法增强了靶基因组基因座处第二核酸的同源重组。在一个实施方案中,该方法增强了在靶基因组基因座处第一核酸和第二核酸的同源重组。在一个实施方案中,同源重组增强至少0.5倍、1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍,或20倍。

[0094] 在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列同源。在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%,或99.9%的同一性。在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列具有100%的同一性。

[0095] 在一个实施方案中,重叠序列为约1kb至约70kb。在一个实施方案中,重叠序列为约1kb至约5kb。在一个实施方案中,重叠序列为约5kb至约10kb。在一个实施方案中,重叠序列为约10kb至约15kb。在一个实施方案中,重叠序列为约15kb至约20kb。在一个实施方案中,重叠序列为约20kb至约25kb。在一个实施方案中,重叠序列为约25kb至约30kb。在一个实施方案中,重叠序列为约30kb至约35kb。在一个实施方案中,重叠序列为约35kb至约40kb。在一个实施方案中,重叠序列为约40kb至约45kb。在一个实施方案中,重叠序列为约45kb至约50kb。在一个实施方案中,重叠序列为约50kb至约55kb。在一个实施方案中,重叠序列为约55kb至约60kb。在一个实施方案中,重叠序列为约60kb至约65kb。在一个实施方案中,重叠序列为约65kb至约70kb。

[0096] 在一个实施方案中,重叠序列为至少5kb。在一个实施方案中,重叠序列为至少10kb。在一个实施方案中,重叠序列为至少15kb。在一个实施方案中,重叠序列为至少20kb。在一个实施方案中,重叠序列为至少25kb。在一个实施方案中,重叠序列为至少30kb。在一个实施方案中,重叠序列为至少35kb。在一个实施方案中,重叠序列为至少40kb。在一个实施方案中,重叠序列为至少45kb。在一个实施方案中,重叠序列为至少50kb。在一个实施方案中,重叠序列为至少55kb。在一个实施方案中,重叠序列为至少60kb。在一个实施方案中,重叠序列为至少65kb。在一个实施方案中,重叠序列为至少70kb。

[0097] 在一个实施方案中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的靶向载体,并且除了所述重叠序列以外,第二核酸不包含与基因组基因座同源的核苷酸序列。

[0098] 在一个实施方案中,第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的第二

核酸插入物的第二靶向载体,并且除了所述重叠序列以外,第一核酸不包含与靶基因组基因座同源的核苷酸序列。

[0099] 在一个实施方案中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一靶向载体,并且第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二靶向载体。在一个实施方案中,第一核酸插入物和第二核酸插入物是连续核酸的重叠片段。

[0100] 在一个实施方案中,靶向载体为约1kb至约2kb。在一个实施方案中,靶向载体为约2kb至约5kb。在一个实施方案中,靶向载体为约5kb至约10kb。

[0101] 在一个实施方案中,靶向载体是大靶向载体(LTVEC)。在一些方法中,靶向载体是在约10kb至约200kb的范围内的一个大靶向载体。在一些方法中,第一靶向载体是第一LTVEC,并且/或者第二靶向载体是第二LTVEC。在一些方法中,第一靶向载体是在约10kb至约200kb的范围内的第一大靶向载体,并且/或者第二靶向载体是在约10kb至约200kb的范围内的第二大靶向载体。在一个实施方案中,LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。任选地,第一LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。任选地,第二LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。在一些方法中,第一LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb,并且第二LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。

[0102] 在一个实施方案中,LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb,并且第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。

[0103] 在一个实施方案中,重叠序列位于第一核酸的3'端处和第二核酸序列的5'端处。在一个实施方案中,重叠序列位于第一核酸的5'端处和第二核酸序列的3'端处。

[0104] 在一个实施方案中,核酸酶试剂由表达构建体表达,该表达构建体包含编码核酸酶的核酸序列,并且其中所述核酸有效连接至细胞中有活性的启动子。在一个实施方案中,核酸酶试剂由编码核酸酶的mRNA表达。在一个实施方案中,核酸酶是锌指核酸酶(ZFN)。在一个实施方案中,核酸酶为转录激活因子样效应物核酸酶(TALEN)。在一个实施方案中,核酸酶为大范围核酸酶。

[0105] 在一个实施方案中,核酸酶试剂包括成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白和向导RNA(gRNA)。在一个实施方案中,Cas蛋白为Cas9。

[0106] 在一个实施方案中,第一核酸插入物和/或第二核酸插入物来自不同的物种。在另一个实施方案中,第一核酸插入物和/或第二核酸插入物是人核酸。在一些方法中,第一核酸插入物、第二核酸插入物或两者来自与所述细胞的物种不同的物种。在一些方法中,第一核酸插入物、第二核酸插入物或两者都是人核酸。

[0107] 在一些方法中,将第一核酸插入物、第二核酸插入物或两者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。在一些方法中,将第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者整合到靶基因组基因座中,导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

[0108] 在一个实施方案中,将第一插入物和/或第二插入物整合到基因组基因座导致该基因组基因座处加入外源序列。

[0109] 在一些实施方案中,所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA,第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。在一些方法中,所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA,第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。

[0110] 在另一个实施方案中,将第一核酸插入物和/或第二核酸插入物整合到基因组基因座导致靶基因组基因座处缺失内源序列。在一个实施方案中,靶基因组基因座处的内源序列缺失为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,或约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至约800kb。

[0111] 在一些方法中,所述细胞为人细胞。在其他方法中,所述细胞为非人细胞。在一些方法中,所述细胞为多能细胞、造血干细胞、神经元干细胞或成纤维细胞。任选地,多能细胞为胚胎干(ES)细胞或诱导多能干(iPS)细胞。在一些方法中,所述细胞为哺乳动物细胞。任选地,哺乳动物细胞为啮齿动物细胞。任选地,啮齿动物细胞为小鼠细胞或大鼠细胞。

[0112] 在一个实施方案中,所述细胞为多能细胞。在另一个实施方案中,多能细胞为胚胎干(ES)细胞。在一些实施方案中,多能细胞为造血干细胞或神经元干细胞。在另一个实施方案中,所述细胞为诱导性多能干(iPS)细胞。

[0113] 在一个实施方案中,靶基因组基因座位于细胞的基因组中。在另一个实施方案中,靶基因组基因座位于细胞内的染色体外DNA上。

[0114] 在一个实施方案中,所述细胞为成纤维细胞。

[0115] 在一些方法中,所述细胞为非人细胞。在其他方法中,所述细胞来自人。在一些实施方案中,该细胞是哺乳动物细胞。在另一个实施方案中,该哺乳动物细胞来自啮齿动物。在一些情况下,啮齿动物为小鼠、大鼠或仓鼠。

[0116] 在上述一些方法中,重叠核苷酸序列有利于将重组机制募集到靶基因组基因座处。

[0117] 本发明还提供了用于产生F0代非人动物的方法,包括:(a)将非人ES细胞引入非人宿主胚胎中,其中该非人ES细胞是通过上述方法产生的;以及(b)在代孕母体中孕育该非人宿主胚胎;其中代孕母体产生包含修饰的F0代非人动物。任选地,非人动物是小鼠或大鼠。

[0118] 在另一个方面,提供了一种通过将重组机制加载到靶向载体上来增强细胞中靶基

基因组基因座处的同源重组的方法,其包括向细胞内引入第一核酸和第二核酸,其中第一核酸和第二核酸包含重叠核苷酸序列,并且其中重叠核苷酸序列有利于将重组机制募集到靶基因组基因座。

[0119] 在一个实施方案中,该方法增强了在靶基因组基因座处第一核酸的同源重组。与其中引入第一核酸而不引入第二核酸的方法相比,一些此类方法增强了靶基因组基因座处第一核酸的同源重组。在一个实施方案中,该方法增强了在靶基因组基因座处第二核酸的同源重组。与其中引入第二核酸而不引入第一核酸的方法相比,一些此类方法增强了靶基因组基因座处第二核酸的同源重组。一些此类方法增强了在靶基因组基因座处第一核酸和第二核酸的同源重组。在一个实施方案中,同源重组增强至少0.5倍、1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍,或20倍。

[0120] 在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列同源。在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%,或99.9%的同一性。在一个实施方案中,第一核酸的重叠序列与第二核酸的重叠序列具有100%的同一性。

[0121] 在一个实施方案中,重叠序列为约1kb至约70kb。在一个实施方案中,重叠序列为约1kb至约5kb。在一个实施方案中,重叠序列为约5kb至约10kb。在一个实施方案中,重叠序列为约10kb至约15kb。在一个实施方案中,重叠序列为约15kb至约20kb。在一个实施方案中,重叠序列为约20kb至约25kb。在一个实施方案中,重叠序列为约25kb至约30kb。在一个实施方案中,重叠序列为约30kb至约35kb。在一个实施方案中,重叠序列为约35kb至约40kb。在一个实施方案中,重叠序列为约40kb至约45kb。在一个实施方案中,重叠序列为约45kb至约50kb。在一个实施方案中,重叠序列为约50kb至约55kb。在一个实施方案中,重叠序列为约55kb至约60kb。在一个实施方案中,重叠序列为约60kb至约65kb。在一个实施方案中,重叠序列为约65kb至约70kb。

[0122] 在一个实施方案中,重叠序列为至少5kb。在一个实施方案中,重叠序列为至少10kb。在一个实施方案中,重叠序列为至少15kb。在一个实施方案中,重叠序列为至少20kb。在一个实施方案中,重叠序列为至少25kb。在一个实施方案中,重叠序列为至少30kb。在一个实施方案中,重叠序列为至少35kb。在一个实施方案中,重叠序列为至少40kb。在一个实施方案中,重叠序列为至少45kb。在一个实施方案中,重叠序列为至少50kb。在一个实施方案中,重叠序列为至少55kb。在一个实施方案中,重叠序列为至少60kb。在一个实施方案中,重叠序列为至少65kb。在一个实施方案中,重叠序列为至少70kb。

[0123] 在一个实施方案中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的靶向载体,并且除了所述重叠序列以外,第二核酸不包含与基因组基因座同源的核苷酸序列。

[0124] 在一个实施方案中,第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二靶向载体,并且除了所述重叠序列以外,第一核酸不包含与基因组基因座同源的核苷酸序列。

[0125] 在一个实施方案中,第一核酸是包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一靶向载体,并且第二核酸是包含侧接有第二5'同源臂和第二3'同源臂的

第二核酸插入物的第二靶向载体。在一个实施方案中,第一核酸插入物和第二核酸插入物是连续核酸的重叠片段。

[0126] 在一个实施方案中,靶向载体为约1kb至约2kb。在一个实施方案中,靶向载体为约2kb至约5kb。在一个实施方案中,靶向载体为约5kb至约10kb。

[0127] 在一个实施方案中,靶向载体是大靶向载体(LTVEC)。在一些方法中,靶向载体是在约10kb至约200kb的范围内的LTVEC。在一些方法中,第一靶向载体是第一LTVEC,并且/或者第二靶向载体是第二LTVEC。在一些方法中,第一靶向载体是在约10kb至约200kb的范围内的第一LTVEC,并且/或者第二靶向载体是在约10kb至约200kb的范围内的第二LTVEC。在一个实施方案中,LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。任选地,第一LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。任选地,第二LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。在一些方法中,第一LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb,并且第二LTVEC为约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,约120kb至约150kb,或约150kb至约200kb。

[0128] 在一个实施方案中,LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。在一些方法中,第一LTVEC的5'和3'同源臂的总和为10kb至约200kb,并且第二LTVEC的5'和3'同源臂的总和为10kb至约200kb。

[0129] 在一个实施方案中,重叠序列位于第一核酸的3'端处和第二核酸序列的5'端处。在一个实施方案中,重叠序列位于第一核酸的5'端处和第二核酸序列的3'端处。

[0130] 在一个实施方案中,第一核酸插入物和/或第二核酸插入物来自不同的物种。在另一个实施方案中,第一核酸插入物和/或第二核酸插入物是人核酸。在一些方法中,第一核酸插入物、第二核酸插入物或两者来自与所述细胞的物种不同的物种。在一些方法中,第一核酸插入物、第二核酸插入物或两者都是人核酸。

[0131] 在一些方法中,将第一核酸插入物、第二核酸插入物或两者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。在一些方法中,将第一核酸插入物、第二核酸插入物和第三核酸插入物中的一者或更多者整合到靶基因组基因座中导致以下情况中的一种或多种:(a)在靶基因组基因座处加入外源序列;(b)在靶基因组基因座处缺失内源序列;或者(c)敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

[0132] 在一个实施方案中,将第一插入物和/或第二插入物整合到基因组基因座中导致该基因组基因座处加入外源序列。

[0133] 在一些实施方案中,所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA,第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。

在一些方法中,所靶向细胞包含具有第一核酸插入物和第二核酸插入物两者的基因组DNA,第一核酸插入物和第二核酸插入物的组合大小在约5kb至约500kb的范围内。

[0134] 在另一个实施方案中,将第一核酸插入物和/或第二核酸插入物整合到基因组基因座中导致靶基因组基因座处缺失内源序列。在一个实施方案中,靶基因组基因座处的内源序列缺失为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,或约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至约800kb。

[0135] 在一些方法中,所述细胞为人细胞。在其他方法中,所述细胞为非人细胞。在一些方法中,所述细胞为多能细胞、造血干细胞、神经元干细胞或成纤维细胞。任选地,多能细胞为胚胎干(ES)细胞或诱导多能干(iPS)细胞。在一些方法中,所述细胞为哺乳动物细胞。任选地,哺乳动物细胞为啮齿动物细胞。任选地,啮齿动物细胞为小鼠细胞或大鼠细胞。

[0136] 在一个实施方案中,所述细胞为多能细胞。在另一个实施方案中,多能细胞为胚胎干(ES)细胞。在一些实施方案中,多能细胞为造血干细胞或神经元干细胞。在另一个实施方案中,所述细胞为诱导性多能干(iPS)细胞。

[0137] 在一个实施方案中,靶基因组基因座位于细胞的基因组中。在另一个实施方案中,靶基因组基因座位于细胞内的染色体外DNA上。

[0138] 在一个实施方案中,所述细胞为成纤维细胞。

[0139] 在一个实施方案中,所述细胞为非人细胞。在其他实施方案中,所述细胞来自人。在其他实施方案中,所述细胞为哺乳动物细胞。在另一个实施方案中,该哺乳动物细胞来自啮齿动物。在另一个实施方案中,啮齿动物为小鼠、大鼠或仓鼠。

[0140] 本发明还提供了用于产生F0代非人动物的方法,包括:(a)将非人ES细胞引入非人宿主胚胎中,其中该非人ES细胞是通过上述方法产生的;以及(b)在代孕母体中孕育该非人宿主胚胎;其中代孕母体产生包含修饰的F0代非人动物。任选地,非人动物是小鼠或大鼠。

附图说明

[0141] 图1提供了基因组双重靶向事件的示意图,在该事件中,靶向的是具有对包含潮霉素选择盒的小鼠染色体14上的TCR α 基因座进行杂合修饰的细胞。将潮霉素选择盒用锌指核酸酶(ZFN)或CRISPR/Cas复合物切割,并用包含新霉素选择盒和超过280kb的人免疫球蛋白 κ 链可变基因区段的两个大靶向载体进行靶向。所述大靶向载体各自包含大约20kb的重叠序列,该重叠序列允许在所述大靶向载体之间进行同源重组。该靶向事件在单个靶向步骤中精确地插入来自所述两个靶向载体的人免疫球蛋白 κ 链可变基因区段。采用被环形包围的矩形来示出用于确认靶向事件的各种探针的位置。图中向上的斜影线表示的是小鼠序列,无阴影线表示的是人类序列,并且向下的斜影虚线表示的是重组位点和选择盒。该示意图未按比例绘制,并且例如不反映可变基因片段的实际数目。

[0142] 图2提供了单一靶向事件的示意图,在该事件中使用包含新霉素选择盒和120kb人免疫球蛋白 κ 链可变基因区段的一个大靶向载体,来靶向具有对包含潮霉素选择盒的小鼠染色体14上的TCR α 基因座进行杂合修饰的细胞。采用被环形包围的矩形来示出用于确认靶向事件的各种探针的位置。图中向上的斜影线表示的是小鼠序列,无阴影线表示的是人类

序列,并且向下的斜影虚线表示的是重组位点和选择盒。该示意图未按比例绘制,并且例如不反映可变基因片段的实际数目。

[0143] 图3提供了使用CRISPR/Cas9系统来靶向并破坏潮霉素选择盒的示意图,并示出了靶向潮霉素基因中不同序列的各种向导RNA (gRNA) 在CRISPR识别位点的潮霉素基因内的位置。该示意图未按比例绘制。

[0144] 图4提供了基因组三重靶向事件的示意图,在该事件中,靶向的是具有对包含潮霉素选择盒的小鼠染色体14上的TCR α 基因座进行杂合修饰的细胞。将潮霉素选择盒用锌指核酸酶 (ZFN) 或CRISPR/Cas复合物切割,并用包含新霉素选择盒和约370kb的人免疫球蛋白 κ 链可变基因区段的三个大靶向载体进行靶向。所述大靶向载体各自包含约20kb至约60kb的重叠序列,该重叠序列允许在所述大靶向载体之间进行同源重组。该靶向事件在单个靶向步骤中精确地插入来自所述全部三个靶向载体的人免疫球蛋白 κ 链可变基因区段。采用被环形包围的矩形来示出用于确认靶向事件的各种探针的位置。图中向上的斜影线表示的是小鼠序列,无阴影线表示的是人类序列,并且向下的斜影虚线表示的是重组位点和选择盒。该示意图未按比例绘制,并且例如不反映可变基因片段的实际数目。

[0145] 定义

[0146] 在本文中可互换使用的术语“蛋白”、“多肽”和“肽”包括任何长度的氨基酸聚合形式,包括编码氨基酸和非编码氨基酸以及以化学方式或生化方式修饰或衍生的氨基酸。这些术语还包括经过修饰的聚合物,诸如具有经过修饰的肽骨架的多肽。

[0147] 在本文中可互换使用的术语“核酸”和“多核苷酸”包括任何长度的核苷酸聚合形式,包括核糖核苷酸、脱氧核糖核苷酸或它们的类似物或修饰形式。这些术语包括单链、双链和多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂交体、以及包含嘌呤碱基、嘧啶碱基、或其他天然的、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的聚合物。

[0148] 靶基因组基因座意指基因组的待使用靶向载体进行所靶向的修饰的区域。该区域可被定义为对应于靶向载体内同源臂的基因组DNA区段的外边界内的区域。靶基因组基因座可包括一个基因或一组基因、一个或多个内含子、一个或多个外显子、一个或多个调控序列等中的任一者或全部。

[0149] “密码子优化”一般包括通过以下方式修饰核酸序列以增强在特定宿主细胞中的表达的过程:将天然序列的至少一个密码子替换为在宿主细胞的基因中更频繁或最频繁使用的密码子,同时保持天然氨基酸序列。例如,可对编码Cas蛋白的核酸进行修饰,以替换成与天然存在的核酸序列相比在给定的原核细胞或真核细胞(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、仓鼠细胞或任何其他宿主细胞)中具有更高使用频率的密码子。密码子使用表是现成的,例如在“密码子使用数据库(Codon Usage Database)”处提供。这些表格可按多种方式进行改编。参见Nakamura et al. (2000) *Nucleic Acids Research* 28:292 (Nakamura等人,2000年,《核酸研究》,第28卷,第292页)。为实现在特定宿主中的表达而对特定序列进行密码子优化的计算机算法也是现成的(参见例如Gene Forge)。

[0150] “有效连接”或“有效连接的”包括两个或更多个组分(例如,启动子和另一个序列元件)的并置,使得这两个组分正常发挥功能并使这些组分中的至少一者有可能介导被施加在其他组分中的至少一者上的功能。例如,如果启动子响应于一个或多个转录调控因子

的存在或不存在而对编码序列的转录水平进行控制,则启动子可以是有效连接至编码序列。

[0151] 术语“多能细胞”或“多能干细胞”包括有能力发育成不止一种分化细胞类型的未分化细胞。此类多能细胞可以是例如哺乳动物胚胎干细胞(ES细胞)或哺乳动物诱导的多能干细胞(iPS细胞)。

[0152] 术语“胚胎干细胞”或“ES细胞”包括胚胎起源的全能或多能细胞,其能够在体外进行未分化增殖,并且能够在引入胚胎后促使发育中的胚胎成为任何组织。

[0153] 术语“诱导多能干细胞”或“iPS细胞”包括可以从经分化的成体细胞直接衍生的多能干细胞。可以通过将特定组的重编程因子引入非多能细胞来产生人iPS细胞,所述重编程因子可以包括例如Oct3/4、Sox家族转录因子(例如Sox1、Sox2、Sox3、Sox15)、Myc家族转录因子(例如,c-Myc、l-Myc、n-Myc)、Krüppel样家族(KLF)转录因子(例如KLF1、KLF2、KLF4、KLF5)和/或相关转录因子(诸如NANOG、LIN28和/或G1is1)。也可以例如通过使用miRNA、模拟转录因子的作用的小分子或谱系特异性分子来产生人iPS细胞。人iPS细胞的特征在于它们能够分化成三个脊椎动物胚层(例如内胚层、外胚层或中胚层)的任何细胞。人iPS细胞的特征还在于它们在合适的体外培养条件下无限繁殖的能力。参见例如,Takahashi和Yamanaka(Cell(2006)Vol.126(4),pp.663-676(《细胞》,2006年,第126卷第4期,第663-676页))。

[0154] 指涉免疫球蛋白核酸序列的术语“种系”包括可传给子代的核酸序列。

[0155] 核酸的“互补性”意指核酸的一条链中的核苷酸序列因其核碱基基团的取向而与相对核酸链上的另一个序列形成氢键。DNA中的互补碱基通常是A与T及C与G。在RNA中,它们通常是C与G及U与A。互补性可以是完全的或实质的/充分的。两个核酸之间的完全互补性意指这两个核酸可以形成双链体,其中双链体中的每个碱基按照沃森-克里克配对原则与互补碱基结合。“实质”或“充分”互补意指一条链中的序列不与相对链中的序列完全和/或完美互补,但在一组杂交条件(例如,盐浓度和温度)中这两条链上的碱基之间发生充分键合而形成稳定的杂交复合物。可通过以下方式预测此类条件:使用序列和标准数学计算来预测杂交链的 T_m (解链温度),或使用常规方法凭经验确定 T_m 。 T_m 包括两条核酸链之间形成的一组杂交复合物50%变性(即一组双链核酸分子一半解离为单链)时的温度。在低于 T_m 的温度下,有利于杂交复合物的形成,而在高于 T_m 的温度下,有利于杂交复合物中的两条链的解链或分离。可在1M NaCl水溶液中对具有已知G+C含量的核酸估计 T_m ,例如使用 $T_m = 81.5 + 0.41(\%G+C)$,而其他已知的 T_m 算法考虑了核酸结构特征。

[0156] “杂交条件”包括累积环境,其中一条核酸链通过互补链相互作用和氢键方式键合于第二核酸链,从而产生杂交复合物。此类条件包括含核酸的水溶液或有机溶液的化学组分及其浓度(例如,盐、螯合剂、甲酰胺)以及该混合物的温度。其他因素(例如,温育时间的长度或反应室尺寸)可对环境有影响。参见例如Sambrook et al.,Molecular Cloning,A Laboratory Manual,2.sup.nd ed.,pp.1.90-1.91,9.47-9.51,11.47-11.57(Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.,1989)(Sambrook等人,《分子克隆实验指南》,第2版,第1.90-1.91、9.47-9.51、11.47-11.57节,冷泉港实验室出版社,美国纽约州冷泉港,1989年)。

[0157] 杂交要求两个核酸包含互补序列,但允许碱基之间出现错配。适于两个核酸之间

的杂交的条件取决于核酸的长度和互补程度,这些变量是本领域众所周知的。两个核苷酸序列之间的互补程度越大,具有这些序列的核酸的杂交体的解链温度(T_m)值就越大。对于具有短序列段互补性(例如,在35个或更少、30个或更少、25个或更少、22个或更少、20个或更少、或18个或更少核苷酸内的互补性)的核酸之间的杂交,错配的位置变得重要(参见 Sambrook 等人,出处同上,11.7-11.8)。通常,可杂交核酸的长度为至少约10个核苷酸。可杂交核酸的示例性最小长度包括至少约15个核苷酸、至少约20个核苷酸、至少约22个核苷酸、至少约25个核苷酸以及至少约30个核苷酸。此外,可视需要根据诸如互补区域的长度和互补程度等因素来调节温度和洗涤溶液盐浓度。

[0158] 多核苷酸的序列不必与其靶核酸的序列100%互补,也能实现特异性杂交。此外,多核苷酸可在一个或多个区段内杂交,使得间插或相邻区段不参与杂交事件(例如,环结构或发夹结构)。多核苷酸(例如,gRNA)可与其靶向的靶核酸序列内的靶区域具有至少70%、至少80%、至少90%、至少95%、至少99%或100%序列互补性。例如,其中20个核苷酸中有18个与靶区域互补并因此特异性杂交的gRNA将具有90%互补性。在该示例中,剩余的非互补核苷酸可以成簇或散布在互补核苷酸内并且无需彼此邻接或与互补核苷酸邻接。

[0159] 通常可使用以下程序来确定核酸内的核酸序列的特定序列段之间的互补性百分比:使用本领域已知的BLAST程序(基本局部比对搜索工具)和PowerBLAST程序(Altschul et al. (1990) *J. Mol. Biol.* 215:403-410 (Altschul 等人,1990年,《分子生物学杂志》,第215卷,第403-410页);Zhang and Madden (1997) *Genome Res.* 7:649-656 (Zhang和Madden,1997年,《基因组研究》,第7卷,第649-656页))或使用Gap程序(威斯康星序列分析软件包,适用于Unix的版本8,遗传学计算机组,美国威斯康星州麦迪逊的大学研究园(Wisconsin Sequence Analysis Package,Version 8for Unix,Genetics Computer Group,University Research Park,Madison Wis.)),这些程序使用默认设置,这使用Smith和Waterman的算法(*Adv. Appl. Math.*, 1981, 2, 482-489 (《应用数学进展》,1981年,第2卷,第482-489页))。

[0160] 本文所提供的方法和组合物采用多种不同组分。在本说明书通篇中已经确认,一些组分可具有活性变体和片段。此类组分包括例如核酸酶试剂、Cas蛋白、CRISPR RNA、tracrRNA和向导RNA。这些组分中的每一者的生物活性在本文别处描述。

[0161] 在两个多核苷酸或多肽序列的语境中,“序列同一性”或“同一性”是指在指定比较窗内对齐以实现最大对应性时这两个序列中相同的残基。当使用序列同一性百分比指涉蛋白质时,应认识到,不相同的残基位置通常差别在于保守氨基酸置换,其中氨基酸残基被置换为具有类似化学特性(例如,电荷或疏水性)的其他氨基酸残基且因此不改变分子的功能特性。当序列差别在于保守置换时,可上调序列同一性百分比以校正置换的保守性质。差别在于此类保守置换的序列被称为具有“序列相似性”或“相似性”。用于进行该调节的方法是人们所熟知的。通常,这涉及将保守置换作为部分错配而非完全错配来评分,从而增加序列同一性百分比。因此,例如,若一个相同氨基酸被给定1的分数且一个非保守置换被给定0的分数,则一个保守置换被给定0至1之间的分数。保守置换的分数例如在程序PC/GENE(美国加利福尼亚州山景城的Intelligenetics公司(Intelligenetics,Mountain View,California))中所执行的那样来计算。

[0162] “序列同一性百分比”包括通过在比较窗内比较两个最佳比对的序列而确定的值,

其中与参考序列(其不包含添加或缺失)相比较,多核苷酸序列在比较窗中的部分可包含添加或缺失(即,空位),以便保证这两个序列的最佳比对。该百分比通过以下方式计算:确定其中相同的核酸碱基或氨基酸残基在两个序列中出现的位置的数目以产生匹配位置的数目,将匹配位置的数目除以在比较窗口中的位置总数,并且将结果乘以100以得到序列同一性百分比。

[0163] 除非另作说明,否则序列同一性/相似性值包括使用GAP版本10采用以下参数获得的值:核苷酸序列的同一性%和相似性%使用空位权重(GAP Weight)50和长度权重3及nwsgapdna.cmp评分矩阵;氨基酸序列的同一性%或相似性%使用空位权重8和长度权重2及BLOSUM62评分矩阵;或其任何等同程序。“等同程序”包括任何序列比较程序,其为所考虑的任何两个序列产生这样的比对,当与由GAP版本10产生的对应比对相比较时,该比对具有相同的核苷酸或氨基酸残基匹配和相同的序列同一性百分比。

[0164] “同源”序列包括与已知参考序列相同或基本上类似的核酸序列,使得该核酸序列与已知参考序列具有至少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%的同一性。“直系同源”序列包括来自一种物种的在功能上与另一物种中的已知参考序列等效的核酸序列。

[0165] 术语“体外”包括人工环境和发生在人工环境(例如试管)中的过程或反应。术语“体内”包括自然环境(例如细胞、生物体或身体)和发生在自然环境中的过程或反应。术语“离体”包括已从个体身体中移除的细胞和发生在此类细胞中的过程或反应。

[0166] “包含”或“包括”一个或多个所述及的要素的组合物或方法可包括未具体述及的其他要素。例如,“包含”或“包括”某种蛋白质的组合物可包含单独的该蛋白质或与其他成分组合的该蛋白质。

[0167] 值的范围的指定包括该范围内的或限定该范围的所有整数,以及由该范围内的整数所限定的所有子范围。

[0168] 除非从上下文明显看出,否则术语“约”涵盖落在规定值的标准测量误差容限(例如,SEM)内的那些值。

[0169] 除非上下文另外明确指出,否则单数形式的量词“一个”、“一种”和“该”包括复数指代物。例如,术语“Cas蛋白”或“至少一种Cas蛋白”可包括多种Cas蛋白,包括它们的混合物。

具体实施方式

[0170] I. 使用多个靶向载体修饰基因组基因座

[0171] 提供了用于修饰细胞中的靶基因组基因座的方法和组合物。此类方法采用能够与彼此重组以形成单个连续核酸区段的多个大靶向载体(LTVEC)。此类方法可利用单个靶向步骤中的1个、2个、3个、4个、5个、6个或更多个LTVEC。还提供了用于增强在细胞中靶基因组基因座处的同源重组的方法和组合物。此类方法采用包含一个或多个重叠序列的两个或更多个核酸。本文所公开的方法中的任一种可在体外、离体或体内发生。

[0172] A. 双重靶向

[0173] 提供了用于通过双重靶向方法修饰细胞内靶基因组基因座的方法和组合物。该方法和组合物采用能够彼此重组以形成单个连续核酸区段的两个大靶向载体(LTVEC)(即第

一LTVEC和第二LTVEC)。第一LTVEC包含第一核酸插入物,并且第二LTVEC包含第二核酸插入物。所述核酸插入物侧接5'同源臂和3'同源臂。第一核酸插入物及其3'同源臂和第二核酸插入物及其5'同源臂可以是同一连续核酸的重叠片段。第一LTVEC的3'同源臂和第二LTVEC的5'同源臂重叠(即,彼此互补),并且第一插入物和第二插入物侧接重叠的同源臂。此类方法包括可按任何顺序发生的三个重组事件:(1)第一LTVEC的3'同源臂和第二LTVEC的5'同源臂之间的重组;(2)第一LTVEC的5'同源臂和靶基因座中对应区段之间的重组;以及(3)第二LTVEC的3'同源臂和靶基因座中对应区段之间的重组。该三向重组使用位于第一核酸插入物和第二核酸插入物之间的同源臂的重叠序列重建靶基因座中的连续核酸。

[0174] LTVEC中的每一个还包含与靶基因组基因座中或附近的DNA区域同源的5'同源臂或3'同源臂,该区域允许单个连续核酸区段的重组和整合。因此,借助于三向重组事件可在单个靶向步骤中在靶基因座处进行大核酸修饰(即缺失、插入和/或替换)。

[0175] 三个重组事件可按任何顺序发生。在一个实施方案中,两个LTVEC的重叠序列之间的重组事件发生在与靶基因座的同源重组之前。在另一个实施方案中,与靶基因座的重组发生在两个LTVEC之间的重组之前。在又一个实施方案中,三个重组事件可同时发生。

[0176] 在一个实施方案中,提供了一种用于修饰细胞中的靶基因组基因座的方法。此方法包括以下步骤:引入包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一大靶向载体(LTVEC)和包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二LTVEC,其中第一LTVEC的第一3'同源臂具有与第二LTVEC的第二5'同源臂同源的重叠序列,并且第一LTVEC的第一5'同源臂和第二LTVEC的第二3'同源臂与靶基因组基因座中的对应基因组区段同源;其中通过将第一核酸插入物和第二核酸插入物整合在对应基因组区段之间来修饰靶基因组基因座。该方法还包括选择所靶向细胞,该所靶向细胞包含整合在靶基因组基因座中的第一核酸插入物和第二核酸插入物。

[0177] B. 三重靶向

[0178] 还提供用于通过三重靶向方法修饰细胞内靶基因组基因座的方法和组合物。该方法和组合物采用能够彼此重组以形成单个连续核酸区段的三个大靶向载体(LTVEC)(即第一LTVEC、第二LTVEC和第三LTVEC)。第一LTVEC包含第一核酸插入物,第二LTVEC包含第二核酸插入物,并且第三LTVEC包含第三核酸插入物。所述核酸插入物侧接5'同源臂和3'同源臂。第一核酸插入物及其3'同源臂和第二核酸插入物及其5'同源臂可以是同一连续核酸的重叠片段。第二核酸插入物及其3'同源臂和第三核酸插入物及其5'同源臂可以是同一连续核酸的重叠片段。第一LTVEC的3'同源臂和第二LTVEC的5'同源臂重叠(即,彼此互补),并且第一插入物和第二插入物侧接重叠的同源臂。第二LTVEC的3'同源臂和第三LTVEC的5'同源臂重叠(即彼此互补),并且第二插入物和第三插入物侧接重叠的同源臂。

[0179] 此类方法包括可按任何顺序发生的四个重组事件:(1)第一LTVEC的3'同源臂和第二LTVEC的5'同源臂之间的重组;(2)第二LTVEC的3'同源臂和第三LTVEC的5'同源臂之间的重组;(3)第一LTVEC的5'同源臂和靶基因座中对应区段之间的重组;以及(4)第三LTVEC的3'同源臂和靶基因座中对应区段之间的重组。该四向重组使用位于第一核酸插入物和第二核酸插入物之间以及第二核酸插入物和第三核酸插入物之间的同源臂的重叠序列重建靶基因座中的连续核酸。

[0180] 第一LTVEC和第三LTVEC还包含与靶基因组基因座中或附近的DNA区域同源的5'同

源臂或3'同源臂,这允许单个连续核酸区段的重组和整合。因此,借助于四向重组事件可在单个靶向步骤中在靶基因座处进行大核酸修饰(即缺失、插入和/或替换)。

[0181] 四个重组事件可按任何顺序发生。在一个实施方案中,三个LTVEC的重叠序列之间的重组事件发生在与靶基因座进行同源重组之前。在另一个实施方案中,与靶基因座的重组发生在三个LTVEC之间的重组之前。在又一个实施方案中,四个重组事件可同时发生。

[0182] 在一个实施方案中,提供了一种用于修饰细胞中的靶基因组基因座的方法。此方法包括以下步骤:引入包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一大靶向载体(LTVEC)、包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二LTVEC和包含侧接有第三5'同源臂和第三3'同源臂的第三核酸插入物的第三LTVEC,其中第一LTVEC的第一3'同源臂具有与第二LTVEC的第二5'同源臂同源的重叠序列,第二LTVEC的第二3'同源臂具有与第三LTVEC的第三5'同源臂同源的重叠序列,并且第一LTVEC的第一5'同源臂和第三LTVEC的第三3'同源臂与靶基因组基因座中的对应基因组区段同源;其中通过将第一核酸插入物、第二核酸插入物和第三核酸插入物整合到对应基因组区段之间来修饰靶基因组基因座。该方法还包括选择所靶向细胞,所靶向细胞包含整合在靶基因组基因座中的第一核酸插入物、第二核酸插入物和第三核酸插入物。

[0183] C. 使用多个LTVEC进行靶向

[0184] 本文提供的用于在单个靶向步骤中形成遗传修饰的靶向方法提供除了通过单个LTVEC靶向方法实现的那些之外的针对所靶向的基因修饰的新可能性和提高效率。使用能够彼此重组的两个、三个或更多个LTVEC进行靶向允许修饰DNA的较大区段。重组事件可按任何顺序发生。例如,LTVEC的重叠序列之间的重组事件可发生在与靶基因座进行同源重组之前。另选地,与靶基因座的重组可发生在LTVEC之间的重组之前,或者重组事件可同时发生。

[0185] 本文所述的靶向方法提供优于现有单个LTVEC靶向方法的若干优势,包括提高的靶向效率、可实现的遗传修饰的大小的增大、以及获得大基因组修饰所需的靶向步骤数量的减少,这一减少节约了时间并且维持了经修饰的胚胎干细胞的多能性。因为该方法允许在单个步骤中使用来自两个、三个或更多个LTVEC的核酸插入物的组合来修饰基因组基因座,所以这对于大基因组修饰尤其重要。因此,此类修饰可允许所靶向的基因组基因座中非常大(例如大于50kb)的缺失、替换和插入。

[0186] 例如,以顺序方式使用三个LTVEC来修饰靶基因组基因座以及筛选并确认所靶向的修饰所需的时间大约为9个月,然而同时用三个LTVEC进行同样的修饰和确认可在仅仅约4个月内完成。

[0187] 当多能细胞诸如胚胎干细胞被修饰时,顺序修饰还导致了多能性和种系传递潜能损失的较高风险。由于传代培养代数和电穿孔数量增加,染色体异常和核型异常积累并且可引起种系能力损失。参见例如Buehr et al. (2008) *Cell* 135:1287-1298 (Buehr等人,2008年,《细胞》,第135卷,第1287-1298页);Li et al. (2008) *Cell* 135(7):1299-1310 (Li等人,2008年,《细胞》,第135卷,第7期,第1299-1310页);以及Liu et al. (1997) *Dev. Dyn.* 209:85-91 (Liu等人,1997年,《发育动力学》,第209卷,第85-91页),所述文献中的每一篇出于所有目的全文以引用方式并入本文。不顺序使用多个LTVEC,而是同时使用多个LTVEC的靶向减少了传代数量和电穿孔数量,从而在保留多能细胞诸如胚胎干细胞的种系

能力的情况下增强了在多能细胞中执行遗传操作的能力。

[0188] 在特定实施方案中,遗传修饰包括对一个或多个内源性核酸的修饰、对一个或多个内源性核酸的置换、用异源核酸替换内源性核酸、敲除或敲入。在具体例子中,通过向细胞引入至少两个大靶向载体(LTVEC)来引入遗传修饰。在另一个例子中,通过向细胞引入至少三个大靶向载体(LTVEC)来引入遗传修饰。在此类例子中,LTVEC可包含待插入细胞的靶基因组基因座中的DNA。

[0189] 在一些实施方案中,用于修饰靶基因组基因座的方法包括向哺乳动物细胞引入遗传修饰。同样,本发明提供包含遗传修饰的哺乳动物细胞。

[0190] 可以使用用于在细胞中进行所靶向的遗传修饰的各种方法。例如,如上所述,所靶向的遗传修饰采用将通过同源重组事件产生所靶向的遗传修饰的体系。在其他情况下,可以使用在所靶向的基因组基因座处产生单链或双链断裂的核酸酶试剂来修饰细胞。然后通过非同源末端连接途径(NHEJ)修复单链或双链断裂。用于产生此类所靶向的遗传修饰的示例性方法在本文其他地方详细论述,包括例如使用大靶向载体。还可参见Wang et al. (2013) *Cell* 153:910-918 (Wang等人,2013年,《细胞》,第153卷,第910-918页),Mandalos et al. (2012) *PLOS ONE* 7:e45768:1-9 (Mandalos等人,2012年,《公共科学图书馆·综合》,第7卷第e45768期,第1-9页),以及Wang et al. (2013) *Nat Biotechnol.* 31:530-532 (Wang等人,2013年,《自然·生物技术》,第31卷,第530-532页),这些文献中的每一篇以引用方式并入本文。

[0191] 由靶向载体和靶基因组基因座之间的同源重组进行的所靶向的基因修饰可能是非常低效的,尤其是在非啮齿动物胚胎干细胞的细胞类型中。靶向载体与靶基因组基因座处核酸酶引导的双链DNA断裂结合使用可极大地提高修饰诸如缺失或插入的靶向效率。相似地,靶向载体与靶基因组基因座处核酸酶引导的单链DNA断裂结合使用可极大地提高修饰的靶向效率。

[0192] 在一些实施方案中,LTVEC可与在靶基因组基因座内造成单链或双链断裂的核酸酶试剂结合使用。此类方法还包括向细胞引入核酸酶试剂。在一个实施方案中,核酸酶试剂为锌指核酸酶(ZFN)。在另一个实施方案中,核酸酶试剂为成簇的规律间隔短回文重复序列(CRISPR)/CRISPR相关(Cas)体系。

[0193] 在一个实施方案中,提供了一种用于利用多个LTVEC来修饰细胞中的靶基因组基因座的方法。此类方法包括:(a)向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂;(b)引入包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一大靶向载体(LTVEC)以及包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二LTVEC,其中第一LTVEC的第一5'同源臂和第二LTVEC的第二3'同源臂与靶基因组基因座内的对应基因组区段同源,并且第一LTVEC的第一3'同源臂和第二LTVEC的5'同源臂彼此同源或分别与一个或多个另外的LTVEC的另外5'和3'同源臂同源,所述另外的LTVEC各自包含侧接有另外的5'同源臂和另外的3'同源臂的另外核酸插入物;其中通过在对应的基因组区段之间整合该第一核酸插入物、该一个或多个另外的LTVEC的另外核酸插入物(如果存在的话)以及该第二核酸插入物来修饰靶基因组基因座;以及(c)选择包含整合在靶基因组基因座中的该第一核酸插入物、该一个或多个另外核酸插入物(如果存在的话)以及该第二核酸插入物的靶细胞。在此类方法中,所述另外的LTVEC是当存在时被插入在第一LTVEC和第二LTVEC之间的一个或多个其他LTVEC。

[0194] 在一个实施方案中,提供了一种用于修饰细胞中的靶基因组基因座的双重靶向方法,该方法包括:(a)向细胞引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂;(b)引入包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一大靶向载体(LTVEC)和包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二LTVEC,其中第一LTVEC的第一3'同源臂具有与第二LTVEC的第二5'同源臂同源的重叠序列,并且第一LTVEC的第一5'同源臂和第二LTVEC的第二3'同源臂与靶基因组基因座内的对应基因组区段同源;其中通过在对应的基因组区段之间整合第一核酸插入物和第二核酸插入物来修饰靶基因组基因座;以及(c)选择包含整合到靶基因组基因座中的第一核酸插入物和第二核酸插入物的所靶向细胞。在此类方法中,第一核酸插入物和第一3'同源臂以及第二核酸插入物和第二5'同源臂是连续核酸的重叠片段,该连续核酸通过整合第一核酸插入物和第二核酸插入物到靶基因组基因座内而重组。

[0195] 在一个实施方案中,提供了一种用于修饰细胞中的靶基因组基因座的三重靶向方法,该方法包括:(a)向细胞内引入在靶基因组基因座内造成单链或双链断裂的核酸酶试剂;(b)引入包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的第一大靶向载体(LTVEC)、包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二LTVEC和包含侧接有第三5'同源臂和第三3'同源臂的第三核酸插入物的第三LTVEC,其中第一LTVEC的第一3'同源臂具有与第二LTVEC的第二5'同源臂同源的重叠序列,第二LTVEC的第二3'同源臂具有与第三LTVEC的第三5'同源臂同源的重叠序列,并且第一LTVEC的第一5'同源臂和第三LTVEC的第三3'同源臂与靶基因组基因座内的对应基因组区段同源;其中通过在对应基因组区段之间整合第一核酸插入物、第二核酸插入物和第三核酸插入物来修饰靶基因组基因座;以及(c)选择包含整合到靶基因组基因座内的第一核酸插入物、第二核酸插入物和第三核酸插入物的所靶向细胞。在此类三重靶向方法中,第一核酸插入物和第一3'同源臂以及第二核酸插入物和第二5'同源臂是连续核酸的重叠片段,并且第二核酸插入物和第二3'同源臂以及第三核酸插入物和第三5'同源臂是连续核酸的重叠片段,该连续核酸是通过整合第一核酸插入物、第二核酸插入物和第三核酸插入物到靶基因组基因座内而重组。

[0196] 在一些情况下,可同时引入两个、三个或更多个LTVEC。另选地,可顺序引入或可在不同时间引入两个、三个或更多个LTVEC。

[0197] 靶向体系的各种组成部分可包括例如靶向载体、核酸酶试剂、靶基因组基因座、核酸插入物、所关注的多核苷酸和/或其他组成部分,这些组成部分中的每一个在本文其他地方有详细描述。

[0198] D. 使用多个重叠核酸的靶向

[0199] 本文提供的用于在单个靶向步骤中形成遗传修饰的靶向方法提供除了通过单个核酸实现的那些之外的针对所靶向的基因修饰的新可能性和提高效率。使用能够彼此重组的两个、三个或更多个核酸的靶向允许修饰DNA的较大区段,并且即使在缺乏核酸酶试剂的情况下,也可提供相比于单独的单个核酸的提高了的靶向效率。因为使用核酸酶的方法所需的筛选更加复杂和耗时,其包括确认切割并检查脱靶效应的额外筛选步骤,所以未使用核酸酶试剂的此类方法可优于采用核酸酶试剂的那些方法。即使在缺乏靶向核酸酶的情况下,具有足够长度的重叠区域的核酸(例如LTVEC)也可增强在靶基因组基因座处的同源重组。例如,与使用单个核酸相比,使用具有足够长度的重叠区域的两个核酸可增强同源重

组。尽管理解机理并不是实践所需,但是据信,通过将重组机制(例如ExoI、Rad51、BRCA2等等)负载到核酸(例如LTVEC)上以在此类情况下增强同源重组,从而有利于将重组机制募集到靶基因座。

[0200] 本文提供了用于修饰靶基因组基因座或用于增强细胞中的靶基因组基因座处的同源重组的方法,该方法包括向细胞引入第一核酸和第二核酸,其中第一核酸和第二核酸包含重叠序列。第一核酸和第二核酸可例如为线性核酸。此类方法还可包括向细胞引入能够彼此重组的三个或更多个核酸。例如,第一核酸和第二核酸可具有第一重叠序列,并且第二核酸和第三核酸可具有第二重叠序列。在一些方法中,在没有核酸酶辅助的情况下,靶基因组基因座被修饰,或者靶基因组基因座处的同源重组被增强。在其他方法中,在核酸酶诸如锌指核酸酶、TALEN、大范围核酸酶或Cas9和向导RNA的辅助下,靶基因组基因座被修饰,或者靶基因组基因座处的同源重组被增强,该核酸酶在靶基因组基因座处或附近造成单链或双链断裂。

[0201] 该方法可增强靶基因组基因座处第一核酸的同源重组,可增强靶基因组基因座处第二核酸的同源重组,或者可增强靶基因组基因座处第一核酸和第二核酸两者的同源重组。例如,与其中引入第一核酸而不引入第二核酸的方法相比,可增强靶基因组基因座处第一核酸的同源重组。同样,与其中引入第二核酸而不引入第一核酸的方法相比,可增强靶基因组基因座处第二核酸的同源重组。同源重组可增强,例如,至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍或20倍。在一些方法中,不使用核酸酶情况下的增强可相当于使用核酸酶情况下的加强。例如,当与不使用核酸酶情况下的增强相比时,使用核酸酶情况下的增强的倍数变化可为0.5倍、0.6倍、0.7倍、0.8倍、0.9倍、1.0倍、1.1倍、1.2倍、1.3倍、1.4倍、1.5倍、1.6倍、1.7倍、1.8倍、1.9倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍或10倍。在一些情况下,不使用核酸酶情况下的增强可相同于或大于使用核酸酶情况下的增强。

[0202] 第一核酸的重叠序列可与第二核酸的重叠序列同源。例如,第一核酸的重叠序列可与第二核酸的重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。另选地,第一核酸的重叠序列可与第二核酸的重叠序列具有100%的同一性。

[0203] 重叠序列可为例如约1kb至约70kb或更多。例如,重叠序列可为约1kb至约5kb、约5kb至约10kb、约10kb至约15kb、约15kb至约20kb、约20kb至约25kb、约25kb至约30kb、约30kb至约35kb、约35kb至约40kb、约40kb至约45kb、约45kb至约50kb、约50kb至约55kb、约55kb至约60kb、约60kb至约65kb、约65kb至约70kb、约70kb至约80kb、约80kb至约90kb、约90kb至约100kb、约100kb至约120kb、约120kb至140kb、约140kb至约160kb、约160kb至约180kb、约180kb至约200kb、约200kb至约220kb、约220kb至约240kb、约240kb至约260kb、约260kb至约280kb,或约280kb至约300kb。例如,重叠序列可为约20kb至约60kb。另选地,重叠序列可为至少1kb、至少5kb、至少10kb、至少15kb、至少20kb、至少25kb、至少30kb、至少35kb、至少40kb、至少45kb、至少50kb、至少55kb、至少60kb、至少65kb、至少70kb、至少80kb、至少90kb、至少100kb、至少120kb、至少140kb、至少160kb、至少180kb、至少200kb、至少220kb、至少240kb、至少260kb、至少280kb,或至少300kb。

[0204] 重叠序列可位于第一核酸和第二核酸内的任何地方。例如,重叠序列可位于第一

核酸的3'端处和第二核酸的5'端处。另选地,重叠序列可位于第一核酸的5'端处和第二核酸的3'端处。

[0205] 在一些方法中,第一核酸为包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的靶向载体。第二核酸可为包含重叠序列的任一核酸,诸如质粒、靶向载体或大靶向载体。在一些方法中,除了重叠序列以外,第二核酸不包含与靶基因组基因座同源的核苷酸序列。例如,第二核酸可基本上由重叠序列组成或者由重叠序列组成。

[0206] 在一些方法中,第一核酸为包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物的靶向载体,并且第二核酸为包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物的第二靶向载体。

[0207] 第一靶向载体可具有任意大小。同样,第二靶向载体可具有任意大小。例如,第一靶向载体和/或第二靶向载体可为约1kb至约2kb、约2kb至约5kb、或约5kb至约10kb。第一靶向载体还可为大靶向载体(LTVEC)。同样,第二靶向载体可为LTVEC。在本文其他地方公开了LTVEC的示例性大小。例如,第一LTVEC和/或第二LTVEC可为约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、约120kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约450kb、约450kb至约500kb、约500kb至约550kb、约550kb至约600kb、约600kb至约650kb、约650kb至约700kb、约700kb至约750kb,或约750kb至约800kb。

[0208] 在一些方法中,第一核酸为LTVEC,并且第二核酸为包含重叠序列的较小核酸,诸如质粒或靶向载体。在一些方法中,除了重叠序列以外,第二核酸不包含与靶基因组基因座同源的核苷酸序列。例如,第二核酸可基本上由重叠序列组成或者由重叠序列组成。

[0209] 在一些方法中,第一核酸插入物和第二核酸插入物为连续核酸的重叠片段。在一些方法中,第一核酸插入物和/或第二核酸插入物可来自与所述细胞的物种不同的物种。例如,第一核酸插入物和/或第二核酸插入物可为人核酸。

[0210] 该方法可将第一核酸插入物和/或第二核酸插入物整合进靶基因组基因座中。该整合可导致靶基因组基因座处序列的添加、靶基因组基因座处序列的缺失或靶基因组基因座处序列的替换。例如,该整合可导致靶基因组基因座处外源序列的添加、靶基因组基因座处内源序列的缺失或靶基因组基因座处内源序列被外源序列替换。插入在靶基因组基因座处的第一核酸插入物、第二核酸插入物或第一核酸插入物和第二核酸插入物的组合可为例如约5kb至约500kb。在本文其他地方公开了其他示例性核酸插入物和插入大小。靶基因组基因座处的缺失可为例如约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约600kb、约600kb至约700kb,或约700kb至约800kb。在本文其他地方公开了其他示例性缺失大小。

[0211] 所靶向的细胞可为本文提供的细胞类型中的任一种,并且靶基因组基因座可为细胞内的任一DNA。例如,靶基因组基因座可在细胞的基因组中,或者它可在细胞内的染色体外DNA上。

[0212] II. 核酸插入物和靶向载体

[0213] A. 核酸插入物

[0214] 一个或多个核酸插入物可以用于本文公开的方法中,并且可以通过不同的靶向载

体或在相同的靶向载体上将它们引入细胞。核酸插入物包含待整合到基因组靶基因座的DNA区段。在靶基因座处整合核酸插入物可以导致向靶基因座添加所关注的核酸序列,靶基因座处所关注的核酸序列的缺失,和/或对靶基因座处所关注的核酸序列的替代。

[0215] 该方法提供使用核酸插入物对基因组基因座的修饰,所述核酸插入物的大小比可使用常规的单个靶向技术实现的核酸插入物(例如单个LTVEC)的大小大。在此类方法中,核酸插入物包含在两个、三个或更多个LTVEC上。LTVEC被设计成使得它们能够彼此重组以形成包含来自两个、三个或更多个LTVEC的组合核酸插入物的DNA的单个较大区段。

[0216] 在此类方法中,核酸插入物侧接5'同源臂和3'同源臂。侧接第一核酸插入物的3'同源臂和侧接第二核酸插入物的5'同源臂为同一连续核酸的重叠片段,该同一连续核酸随后通过同源臂的重叠片段之间的重组来重组。在此类方法中,两个LTVEC之间的重组产生了连续核酸插入物,其中同源臂的重叠序列位于第一核酸插入物和第二核酸插入物之间。三重靶向方法涉及第二LTVEC和第三LTVEC之间的额外重组,其中侧接第二核酸插入物的3'同源臂和侧接第三核酸插入物的5'同源臂为同一连续核酸的重叠片段,该同一连续核酸随后通过同源臂的重叠片段之间的重组来重组。在此类三重靶向方法中,三个LTVEC之间的重组产生了连续核酸插入物,其中同源臂的重叠序列位于第一核酸插入物、第二核酸插入物和第三核酸插入物之间。在一个实施方案中,同源臂的重叠序列包含核酸插入物的一部分。

[0217] 因此,这些方法允许在单个靶向步骤中使用来自两个、三个或更多个LTVEC的核酸插入物来修饰基因组基因座,从而在减少靶向步骤数量的同时,有效增大了核酸插入物的总大小。

[0218] 核酸插入物或被替代的靶基因座处的相应核酸可以是编码区、内含子、外显子、非翻译区、调控区、启动子、增强子,或它们的任何组合。此外,核酸插入物(即来自两个、三个或更多个LTVEC的组合核酸插入物)的大小或在靶基因座处被替换的对应核酸可具有任何期望的长度,包括例如介于10至100个核苷酸之间的长度,100至500个核苷酸的长度,500个核苷酸至1kb的长度,1kb至1.5kb的长度,1.5kb至2kb的长度,2kb至2.5kb的长度,2.5kb至3kb的长度,3kb至5kb的长度,5kb至8kb的长度,8kb至10kb的长度,或更大长度。在其他情况下,该长度可为约50kb至约700kb、约50kb至约500kb、约50kb至约300kb、约50kb至约75kb、约75kb至约100kb、约100kb至约125kb、约125kb至约150kb、约150kb至约175kb、约175kb至约200kb、约200kb至约225kb、约225kb至约250kb、约250kb至约275kb、约275kb至约300kb、约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约450kb、约450kb至约500kb、约500kb至约550kb、约550kb至约600kb、约600kb至约650kb、约650kb至约700kb、约700kb至约800kb、约800kb至1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约2.8Mb,或约2.8Mb至约3Mb。另选地,来自两个、三个或更多个LTVEC的组合核酸插入物或在靶基因座处被替换的对应核酸可为约3Mb至约4Mb、约4Mb至约5Mb、约5Mb至约6Mb、约6Mb至约7Mb、约7Mb至约8Mb、约8Mb至约9Mb,或约9Mb至约10Mb。在其他情况下,长度可以为至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或更大。例如,来自两个、三个或更多个LTVEC的组合核酸插入物或在靶基因座处被替换的对应核酸可为至少20kb、至少40kb、至

少60kb、至少80kb、至少100kb、至少150kb、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb、至少500kb、至少550kb、至少600kb、至少650kb、至少700kb、至少750kb、至少800kb、至少850kb、至少900kb、至少950kb、至少1Mb、至少1.5Mb、至少2Mb、至少2.5Mb、至少3Mb、至少4Mb、至少5Mb、至少6Mb、至少7Mb、至少8Mb、至少9Mb、或至少10Mb。在一个实施方案中,核酸插入物的大小为约5kb至约700kb。在一个实施方案中,核酸插入物的大小为约5kb至约500kb。在另一个实施方案中,核酸插入物的大小为约100kb至约700kb。在另一个实施方案中,核酸插入物的大小为约100kb至约500kb。在一个具体实施方案中,核酸插入物为约140kb。在另一个具体实施方案中,核酸插入物为约370kb。在另一个具体实施方案中,核酸插入物为约300kb。在另一个具体实施方案中,核酸插入物为约400kb。

[0219] 在一些单独靶向载体(即在与另一个靶向载体重组之前)中,核酸插入物可具有介于10至100个核苷酸之间的长度、100至500个核苷酸的长度、500个核苷酸至1kb的长度、1kb至1.5kb的长度、1.5kb至2kb的长度、2kb至2.5kb的长度、2.5kb至3kb的长度,或3kb至5kb的长度。在其他情况下,该长度可为约5kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约60kb至约70kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb,或约190kb至约200kb。另选地,核酸插入物可以为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,约150kb至约200kb,约200kb至约250kb,约250kb至约300kb,约300kb至约350kb,或者约350kb至约400kb。另选地,该核酸插入物可为约400kb至约450kb、约450kb至约500kb、约500kb至约550kb、约550kb至约600kb、约600kb至约650kb、约650kb至约700kb、约700kb至约750kb,或约750kb至约800kb。

[0220] 在一些情况下,替换靶基因座处的核酸导致核酸序列缺失在约1kb至约200kb、约2kb至约20kb或约0.5kb至约3Mb的范围内。在一些情况下,缺失的程度大于5'同源臂和3'同源臂的总长度。

[0221] 在一些情况下,核酸序列缺失程度在以下范围内:约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约70kb至约80kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约110kb至约120kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、约190kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约800kb、约800kb至1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约2.8Mb、约2.8Mb至约3Mb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb,或约2.5Mb至约3Mb。另选地,该缺失可为约3Mb至约4Mb、约4Mb至约5Mb、约5Mb至约10Mb、约10Mb至约20Mb、约20Mb至约30Mb、约30Mb至约40Mb、约40Mb至约50Mb、约50Mb至约60Mb、约60Mb至约70Mb、约70Mb至约80Mb、约80Mb至约90Mb,或约90Mb至约100Mb。

[0222] 在其他情况下,核酸插入物或在靶基因座处被替换的对应核酸可为至少10kb、至

少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少120kb、至少150kb、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb、至少500kb、至少550kb、至少600kb、至少650kb、至少700kb或更大。

[0223] 核酸插入物可以包含基因组DNA或任何其他类型的DNA。例如,核酸插入物可以来源于:原核生物、真核生物、酵母、禽类(例如,鸡)、非人哺乳动物、啮齿动物、人、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、猫、狗、白鼬、灵长类动物(例如,狨猴、恒河猴)、家养哺乳动物或农业哺乳动物,或任何其他所关注的生物体。

[0224] 核酸插入物和/或靶基因座处的核酸可包含编码序列或非编码序列,诸如调控元件(例如启动子、增强子,或转录阻遏物结合元件)。例如,核酸插入物可以包含内源性基因的至少一个外显子的敲入等位基因,或整个内源性基因的敲入等位基因(即“基因交换敲入”)。例如,核酸插入物可与靶基因组基因座处要被缺失的序列同源或直系同源。同源或直系同源的核酸插入物可替换在所关注的基因组基因座处要被缺失的序列。如果核酸插入物的插入产生了用同源或直系同源人核酸序列替换非人核酸序列(即该核酸插入物被插入替代在其内源基因组基因座处的对应非人DNA序列),那么这可导致基因座人源化。

[0225] 核酸插入物还可以包含条件型等位基因。条件型等位基因为如US2011/0104799中所述的多功能等位基因,该专利全文以引用方式并入本文。例如,条件型等位基因可以包含:(a)相对于靶基因的转录呈有义取向的致动序列;(b)呈有义或反义取向的药物选择盒(DSC);(c)呈反义取向的所关注核苷酸序列(NSI);以及(d)呈反义取向的倒转条件模块(COIN,其利用外显子断裂内含子和可倒转的基因诱捕样模块)。参见例如US 2011/0104799,该专利全文以引用方式并入本文。条件型等位基因还可以包含在暴露于第一重组酶后重组以形成条件型等位基因的可重组单元,所述条件型等位基因(i)缺乏致动序列和DSC;并且(ii)含有呈有义取向的NSI和呈反义取向的COIN。参见US 2011/0104799。

[0226] 一些核酸插入物包含编码选择标记物的多核苷酸。选择标记物可以包含在选择盒中。此类选择标记包括但不限于新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)或单纯疱疹病毒胸苷激酶(HSV-k),或者它们的组合。编码选择标记的多核苷酸可有效连接至在所靶向的细胞中有活性的启动子。启动子的例子在本文其他地方有所描述。

[0227] 在一些靶向载体中,核酸插入物包含报告基因。报告基因的例子为编码荧光素酶、 β -半乳糖苷酶、绿色荧光蛋白(GFP)、增强型绿色荧光蛋白(eGFP)、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)、增强型黄色荧光蛋白(eYFP)、蓝色荧光蛋白(BFP)、增强型蓝色荧光蛋白(eBFP)、DsRed、ZsGreen、MmGFP、mPlum、mCherry、tdTomato、mStrawberry、J-Red、mOrange、mKO、mCitrine、Venus、YPet、Emerald、CyPet、Cerulean、T-Sapphire、碱性磷酸酶,以及它们的组合的基因。此类报告基因可有效连接至在所靶向的细胞中有活性的启动子。启动子的例子在本文其他地方有所描述。

[0228] 在一些靶向载体中,核酸插入物包含一个或多个表达盒或缺失盒。给定表达盒可包含所关注的核苷酸序列、编码选择标记物的核酸,和/或报告基因,以及影响表达的各种调控组分。可以包括的选择性标记物和报告基因的示例在本文其他地方详细论述。

[0229] 在一些靶向载体中,核酸插入物包含侧接有位点特异性重组靶序列的核酸。虽然整个核酸插入物可侧接这种位点特异性重组靶序列,但该核酸插入物内的所关注的任何区

域或单独多核苷酸也可侧接此类位点。可侧接核酸插入物或核酸插入物中的任何所关注多核苷酸的位点特异性重组靶序列可包括例如loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attP、att、FRT、rox,以及它们的组合。在一个例子中,位点特异性重组位点侧接核酸插入物内所含的编码选择标记和/或报告基因的多核苷酸。在所靶向的基因座处整合核酸插入物之后,可除去位点特异性重组位点之间的序列。

[0230] B. 所关注的多核苷酸

[0231] 任何所关注的多核苷酸都可包含在各种核酸插入物中并由此整合在靶基因组基因座处。本文所公开的方法提供整合到所靶向的基因组基因座中的至少1、2、3、4、5、6个或更多个所关注的多核苷酸。

[0232] 当在靶基因组基因座处整合时,核酸插入物内的所关注多核苷酸可将一个或多个遗传修饰引入细胞中。所述基因修饰可包括缺失内源核酸序列和/或将外源或异源或直系同源多核苷酸加入到靶基因组基因座中。在一个实施方案中,所述基因修饰包括在靶基因组基因座处用所关注的外源多核苷酸替换内源核酸序列。因此,本文所提供的方法允许在靶基因组基因座中生成遗传修饰,所述遗传修饰包括敲除、缺失、插入、替换(“敲入”)、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。此类修饰可在第一、第二、第三、第四、第五、第六、第七或任何后续的核酸插入物整合到靶基因组基因座中后发生。

[0233] 在核酸插入物内的和/或在靶基因组基因座处整合的所关注多核苷酸可包括对于其被引入的细胞为天然的或同源的序列;所关注多核苷酸对于其被引入的细胞可为异源的;所关注多核苷酸对于其被引入的细胞可为外源的;所关注多核苷酸对于其被引入的细胞可为直系同源的;或者所关注多核苷酸可来自与其被引入的细胞不同的物种。指涉序列的“同源”包括对所述细胞天然的序列。指涉序列的“异源”包括来源于外来物种的序列,或者,如果序列来源于同一物种,则通过有意的人为干预在组成和/或基因组基因座方面从其天然形式进行了实质性修饰。指涉序列的“外源”包括源于外来物种的序列。“直系同源”包括来自一种物种的在功能上与另一物种中的已知参考序列等效的多核苷酸(即,物种变体)。所关注的多核苷酸可来自任何所关注的生物体,包括但不限于非人、啮齿动物、仓鼠、小鼠、大鼠、人、猴、禽、农业哺乳动物或非农业哺乳动物。目标多核苷酸还可包含编码区、非编码区、调控区或基因组DNA。因此,第1、第2、第3、第4、第5、第6、第7和/或后续核酸插入物中的任一者可包含此类序列。

[0234] 在一个实施方案中,在核酸插入物内和/或整合在靶基因组基因座处的所关注多核苷酸是与人核酸同源的。在更进一步的实施方案中,在靶基因座处整合的目标多核苷酸为基因组核酸的片段。在一个实施方案中,基因组核酸为小鼠基因组核酸、人基因组核酸、非人核酸、啮齿动物核酸、大鼠核酸、仓鼠核酸、猴核酸、农业哺乳动物核酸或非农业哺乳动物核酸,或者它们的组合。

[0235] 在一个实施方案中,如上所述,所关注多核苷酸可在约500个核苷酸至约200kb的范围内。所关注的多核苷酸可以是约500个核苷酸至约5kb,约5kb至约200kb,约5kb至约700kb,约5kb至约10kb,约10kb至约20kb,约20kb至约30kb,约30kb至约40kb,约40kb至约50kb,约60kb至约70kb,约80kb至约90kb,约90kb至约100kb,约100kb至约110kb,约120kb至约130kb,约130kb至约140kb,约140kb至约150kb,约150kb至约160kb,约160kb至约170kb,

约170kb至约180kb,约180kb至约190kb,约190kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,或约600kb至约700kb。

[0236] 在核酸插入物内和/或插入在靶基因组基因座处的所关注多核苷酸可编码多肽,可编码miRNA,可编码非编码的长RNA,或者其可包含任何所关注的调控区或非编码区,包括例如调控序列、启动子序列、增强子序列、转录阻遏物结合序列,或非蛋白编码序列的缺失,但不包含蛋白编码序列的缺失。另外,在核酸插入物内和/或插入在靶基因组基因座处的所关注多核苷酸可编码在神经系统、骨骼系统、消化系统、循环系统、肌肉系统、呼吸系统、心血管系统、淋巴系统、内分泌系统、泌尿系统、生殖系统或它们的组合中表达的蛋白。

[0237] 在核酸插入物内和/或整合在靶基因组基因座处的所关注多核苷酸可包含编码序列中的遗传修饰。此类遗传修饰包括但不限于编码序列的缺失突变或两个编码序列的融合。

[0238] 在核酸插入物内和/或整合在靶基因组基因座处的所关注多核苷酸可包含编码突变蛋白的多核苷酸。在一个实施方案中,所述突变体蛋白的特征在于改变的结合特性、改变的定位、改变的表达和/或改变的表达模式。在一个实施方案中,在核酸插入物内和/或整合在靶基因组基因座处的所关注多核苷酸包含至少一种疾病等位基因。在此类情况下,所述疾病等位基因可为显性等位基因,或者所述疾病等位基因为隐性等位基因。此外,所述疾病等位基因可包括单核苷酸多态性(SNP)等位基因。编码突变体蛋白的目标多核苷酸可来自任何生物体,包括但不限于编码突变体蛋白的哺乳动物、非人哺乳动物、啮齿动物、小鼠、大鼠、人、猴、农业哺乳动物或家养哺乳动物多核苷酸。

[0239] 在核酸插入物内和/或整合在靶基因组基因座处的所关注多核苷酸还可包含调控序列,包括例如启动子序列、增强子序列、转录阻遏物结合序列,或转录终止子序列。在具体实施方案中,在核酸插入物内和/或整合在靶基因组基因座处的所关注多核苷酸包括具有非蛋白编码序列缺失但不包含蛋白编码序列缺失的多核苷酸。在一个实施方案中,所述非蛋白编码序列的缺失包括调控序列的缺失。在另一个实施方案中,所述调控元件的缺失包括启动子序列的缺失。在一个实施方案中,所述调控元件的缺失包括增强子序列的缺失。此类目标多核苷酸可来自任何生物体,包括但不限于编码突变体蛋白的哺乳动物、非人哺乳动物、啮齿动物、小鼠、大鼠、人、猴、农业哺乳动物或家养哺乳动物多核苷酸。

[0240] 靶向基因修饰可以包括对所关注的多核苷酸的靶向改变。此类靶向修饰包括但不限于一个或多个核苷酸的添加、一个或多个核苷酸的缺失、一个或多个核苷酸的取代、所关注多核苷酸或其部分的敲除、所关注多核苷酸或其部分的敲入、用异源核酸序列替代内源性核酸序列,或它们的组合。在具体实施方案中,改变至少1个、2个、3个、4个、5个、7个、8个、9个、10个、100个、500个或更多个核苷酸,或者至少10kb至500kb或更多,以形成靶向基因组修饰。

[0241] C. 靶向载体

[0242] 可以采用靶向载体将核酸插入物引入基因组靶基因座,并且靶向载体包含核酸插入物和侧接于核酸插入物的同源臂。靶向载体可以是线形形式或环形形式,并且它们可以是单链或双链的。靶向载体可以是脱氧核糖核酸(DNA)或核糖核酸(RNA)。为了易于参考,同源臂在本文中称为5'和3'(即上游和下游)同源臂。该术语涉及靶向载体内的同源臂与核酸插入物的相对位置。5'和3'同源臂对应于所靶向的基因座内的区域或另一个靶向载体内的

区域,它们在本文中分别称为“5'靶序列”和“3'靶序列”。在一些情况下,同源臂也可充当5'或3'靶序列。

[0243] 本发明的方法采用能够彼此重组的两个、三个或更多个靶向载体。在各种实施方案中,靶向载体是如本文别处所述的大靶向载体(LTVEC)。在此类方法中,第一靶向载体、第二靶向载体和第三靶向载体各自包含5'和3'同源臂。第一靶向载体的3'同源臂包含与第二靶向载体的5'同源臂重叠的序列(即,重叠序列),从而允许第一LTVEC与第二LTVEC之间的同源重组。

[0244] 就双重靶向方法而言,第一靶向载体的5'同源臂和第二靶向载体的3'同源臂与靶基因组基因座内的对应区段(即,靶序列)同源,从而促进第一靶向载体和第二靶向载体与对应基因组区段的同源重组并修饰靶基因组基因座。

[0245] 就三重靶向方法而言,第二靶向载体的3'同源臂包含与第三靶向载体的5'同源臂重叠的序列(即,重叠序列),从而允许第二LTVEC与第三LTVEC之间的同源重组。第一靶向载体的5'同源臂和第三靶向载体的3'同源臂与靶基因组基因座内的对应区段(即,靶序列)同源,从而促进第一靶向载体和第三靶向载体与对应基因组区段的同源重组并修饰靶基因组基因座。

[0246] 当两个区域彼此共有足够水平的序列同一性时,同源臂和靶序列或两个同源臂彼此“对应”,从而充当同源重组反应的底物。术语“同源性”包括DNA序列与对应序列相同或共有序列同一性。给定靶序列与存在于靶向载体上的对应同源臂(即,重叠序列)之间或两个同源臂之间的序列同一性可为允许同源重组发生的任何程度的序列同一性。例如,靶向载体的同源臂(或其片段)与另一个靶向载体的靶序列或靶基因组基因座的靶序列(或其片段)共有的序列同一性的量可为至少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,以使得所述序列经历同源重组。

[0247] 此外,同源臂与对应靶序列之间的对应同源区可具有足以促进靶基因组基因座处同源重组的任何长度。例如,给定同源臂和/或对应靶序列可包含对应同源区,所述对应同源区为至少约5-10kb、5-15kb、5-20kb、5-25kb、5-30kb、5-35kb、5-40kb、5-45kb、5-50kb、5-55kb、5-60kb、5-65kb、5-70kb、5-75kb、5-80kb、5-85kb、5-90kb、5-95kb、5-100kb、100-200kb,或200-300kb长或更长(如在本文别处描述的LTVEC载体中所述),以使得同源臂与细胞靶基因组基因座内或另一个靶向载体内的对应靶序列具有足以经历同源重组的同源性。

[0248] 第一靶向载体的3'同源臂与第二靶向载体的5'同源臂的重叠序列或者第二靶向载体的3'同源臂与第三靶向载体的5'同源臂的重叠序列可具有足以促进靶向载体之间的同源重组的任何长度。例如,同源臂的给定重叠序列可包含对应重叠区域,所述对应重叠区域为至少约1-5kb、5-10kb、5-15kb、5-20kb、5-25kb、5-30kb、5-35kb、5-40kb、5-45kb、5-50kb、5-55kb、5-60kb、5-65kb、5-70kb、5-75kb、5-80kb、5-85kb、5-90kb、5-95kb、5-100kb、100-200kb,或200-300kb长或更长,使得同源臂的重叠序列与另一个靶向载体内的对应重叠序列具有足以经历同源重组的同源性。在一个实施方案中,重叠序列为1-5kb。在一个实施方案中,重叠序列为约1kb至约70kb。在一个实施方案中,重叠序列为约10kb至约70kb。在另一个实施方案中,重叠序列为约10kb至约50kb。在一个实施方案中,重叠序列为至少10kb。在另一个实施方案中,重叠序列为至少20kb。例如,重叠序列可为约1kb至约5kb,约

5kb至约10kb,约10kb至约15kb,约15kb至约20kb,约20kb至约25kb,约25kb至约30kb,约30kb至约35kb,约35kb至约40kb,约40kb至约45kb,约45kb至约50kb,约50kb至约60kb,约60kb至约70kb,约70kb至约80kb,约80kb至约90kb,约90kb至约100kb,约100kb至约120kb,约120kb至约140kb,约140kb至约160kb,约160kb至约180kb,约180kb至约200kb,约200kb至约220kb,约220kb至约240kb,约240kb至约260kb,约260kb至约280kb,或约280kb至约300kb。例如,重叠序列可为约20kb至约60kb。另选地,重叠序列可为至少1kb、至少5kb、至少10kb、至少15kb、至少20kb、至少25kb、至少30kb、至少35kb、至少40kb、至少45kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少120kb、至少140kb、至少160kb、至少180kb、至少200kb、至少220kb、至少240kb、至少260kb、至少280kb,或至少300kb。

[0249] 同源臂可与对细胞为天然的基因座(例如,靶基因座)相对应,或者另选地,它们可与整合到细胞基因组中的异源或外源DNA区段的区域相对应,所述区域包括例如转基因、表达盒或者异源或外源DNA区域。另选地,同源臂可与细胞中的靶向载体上的区域相对应。靶向载体的同源臂可与酵母人工染色体(YAC)、细菌人工染色体(BAC)、人类人工染色体的区域或在适当宿主细胞中包含的任何其他经改造的区域相对应。更进一步,靶向载体的同源臂可与BAC文库、粘粒文库或P1噬菌体文库的区域相对应,或者可来源于BAC文库、粘粒文库或P1噬菌体文库的区域。在某些情况下,靶向载体的同源臂与对于以下生物为天然、异源或外源的基因座相对应:原核生物、酵母、禽类(例如,鸡)、非人哺乳动物、啮齿动物、人、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、猫、狗、白鼬、灵长类动物(例如,狨猴、恒河猴)、家养哺乳动物、农业哺乳动物,或任何其他所关注的生物体。在一些情况下,同源臂与细胞中的基因座相对应,在不存在核酸酶试剂(例如,Cas蛋白)诱导的切口或双链断裂的情况下,该基因座无法使用常规方法靶向,或仅可不正确地或仅以显著较低效率靶向。在一些情况下,同源臂来源于合成DNA。

[0250] 在一些靶向载体中,5'或3'同源臂中的一者对应于所靶向的基因组基因座,而5'或3'同源臂中的另一者对应于另一个靶向载体上的区域。

[0251] 在一些靶向载体中,5'和3'同源臂对应于所靶向的基因组。另选地,同源臂可来自相关基因组。例如,所靶向的基因组为第一品系的小鼠基因组,且靶向臂来自第二品系的小鼠基因组,其中第一品系与第二品系不同。在某些情况下,同源臂来自相同动物的基因组或来自相同品系的基因组,例如所靶向的基因组为第一品系的小鼠基因组,且靶向臂来自相同小鼠或相同品系的小鼠基因组。

[0252] 靶向载体的同源臂可具有足以促进与对应靶序列的同源重组事件的任何长度,包括例如至少1-5kb、5-10kb、5-15kb、5-20kb、5-25kb、5-30kb、5-35kb、5-40kb、5-45kb、5-50kb、5-55kb、5-60kb、5-65kb、5-70kb、5-75kb、5-80kb、5-85kb、5-90kb、5-95kb、5-100kb、100-200kb,或200-300kb的长度或更长。如下文进一步详细描述,大靶向载体可采用更大长度的靶向臂。

[0253] 核酸酶试剂(例如,CRISPR/Cas体系)可与靶向载体组合使用,以帮助修饰靶基因座。此类核酸酶试剂可以促进靶向载体和靶基因座之间的同源重组。当核酸酶试剂与靶向载体组合使用时,靶向载体可以包含对应于定位为足够接近核酸酶切割位点的5'和3'靶序列的5'和3'同源臂,以促进在核酸酶切割位点处产生切口或双链断裂后靶序列与同源臂之间的同源重组事件。术语“核酸酶切割位点”包括在其处用核酸酶试剂产生切口或双链断裂

的DNA序列(例如,Cas9切割位点)。靶向基因座内对应于靶向载体的5'和3'同源臂的靶序列“位于足够接近”核酸酶切割位点处,前提是该距离能够促进在识别位点处产生切口或双链断裂后在5'和3'靶序列与同源臂之间发生同源重组事件。因此,在特定情况下,对应于靶向载体的5'和/或3'同源臂的靶序列在给定识别位点的至少1个核苷酸内,或者在给定识别位点的至少10个核苷酸至约14kb内。在一些情况下,核酸酶切割位点紧邻靶序列中的至少一者或两者。

[0254] 对应于靶向载体的同源臂的靶序列与核酸酶切割位点的空间关系可变化。例如,靶序列可位于核酸酶切割位点的5'端,靶序列可位于识别位点的3'端,或者靶序列可位于核酸酶切割位点两侧。

[0255] 与单独使用靶向载体相比,联合使用靶向载体(包括例如大靶向载体)与核酸酶试剂可以提高靶向效率。例如,当与单独使用靶向载体相比时,当靶向载体与核酸酶试剂联合使用时,靶向载体的靶向效率可提高至少两倍、至少三倍、至少4倍、至少10倍或在由这些整数形成的范围内,诸如2-10倍。

[0256] D. 大靶向载体

[0257] 一些靶向载体是“大靶向载体”或“LTVEC”,其包括包含对应于和来源于比由旨在细胞中进行同源重组的其他方法通常使用的那些核酸序列更大的核酸序列的同源臂的靶向载体。LTVEC可例如为至少10kb长,或5'同源臂和3'同源臂的总和可例如为至少10kb。LTVEC还包括包含具有比由旨在细胞中进行同源重组的其他方法通常使用的那些核酸序列更大的核酸序列的核酸插入物的靶向载体。例如,LTVEC使得能够对大基因座进行修饰,而传统的基于质粒的靶向载体由于有大小限制而无法实现这一点。例如,所靶向的基因座可以是(即,5'和3'同源臂可以对应于)在不存在核酸酶试剂(例如,Cas蛋白)诱导的切口或双链断裂的情况下,无法使用常规方法靶向,或仅可不正确地或仅以显著较低效率靶向的细胞基因座。

[0258] 如本文别处所述,本文所提供的方法采用能够在三向或四向重组事件中彼此重组且与靶基因组基因座重组的两个或三个LTVEC。这些方法使得对大基因座的修饰成为可能,而使用单一LTVEC无法实现这一点。

[0259] LTVEC的示例包括衍生自细菌人工染色体(BAC)、人类人工染色体或酵母人工染色体(YAC)的载体。LTVEC及其制备方法的示例描述于例如美国专利号6,586,251;美国专利号6,596,541;美国专利号7,105,348;和W02002/036789(PCT/US01/45375)中,这些专利每一者均以引用方式并入本文。LTVEC可以是线形形式或环形形式。

[0260] LTVEC可为任何长度,包括例如约20kb至约300kb,约20kb至约50kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,或约275kb至约300kb。另选地,LTVEC可以是至少10kb、至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少150kb、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb,或至少500kb或更多。LTVEC的尺寸可能太大,以致无法通过常规测定法如Southern印迹和长片段(例如,1kb-5kb)PCR来筛选靶向事件。

[0261] 在一些情况下,LTVEC包括在约5kb至约200kb,约5kb至约10kb,约10kb至约20kb,

约20kb至约30kb,约30kb至约40kb,约40kb至约50kb,约60kb至约70kb,约80kb至约90kb,约90kb至约100kb,约100kb至约110kb,约120kb至约130kb,约130kb至约140kb,约140kb至约150kb,约150kb至约160kb,约160kb至约170kb,约170kb至约180kb,约180kb至约190kb,或约190kb至约200kb范围内的核酸插入物。在其他情况下,核酸插入物可在约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,约150kb至约200kb,约200kb至约250kb,约250kb至约300kb,约300kb至约350kb,或约350kb至约400kb的范围内。在一些情况下,LTVEC包括在约400kb至约450kb,约450kb至约500kb,约500kb至约550kb,约550kb至约600kb,约600kb至约650kb,约650kb至约700kb,约700kb至约750kb,或约750kb至约800kb范围内的核酸插入物。

[0262] 在一些LTVEC中,5'同源臂和3'同源臂的总和为至少10kb。在其他LTVEC中,5'同源臂在约1kb至约100kb的范围内并且/或者3'同源臂在约1kb至约100kb的范围内。5'和3'同源臂的总和可为例如约1kb至约5kb,约5kb至约10kb,约10kb至约20kb,约20kb至约30kb,约30kb至约40kb,约40kb至约50kb,约50kb至约60kb,约60kb至约70kb,约70kb至约80kb,约80kb至约90kb,约90kb至约100kb,约100kb至约110kb,约110kb至约120kb,约120kb至约130kb,约130kb至约140kb,约140kb至约150kb,约150kb至约160kb,约160kb至约170kb,约170kb至约180kb,约180kb至约190kb,或约190kb至约200kb。另选地,每个同源臂可为至少5kb、至少10kb、至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少110kb、至少120kb、至少130kb、至少140kb、至少150kb、至少160kb、至少170kb、至少180kb、至少190kb,或至少200kb。同样,5'和3'同源臂的总和可为至少5kb、至少10kb、至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少110kb、至少120kb、至少130kb、至少140kb、至少150kb、至少160kb、至少170kb、至少180kb、至少190kb,或至少200kb。

[0263] 在一些情况下,LTVEC和核酸插入物被设计成允许在靶基因座处缺失约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,或约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,约700kb至约800kb,约500kb至约1Mb,约1Mb至约1.5Mb,约1.5Mb至约2Mb,约2Mb至约2.5Mb,或约2.5Mb至约3Mb的内源序列。另选地,该缺失可为约3Mb至约4Mb、约4Mb至约5Mb、约5Mb至约10Mb、约10Mb至约20Mb、约20Mb至约30Mb、约30Mb至约40Mb、约40Mb至约50Mb、约50Mb至约60Mb、约60Mb至约70Mb、约70Mb至约80Mb、约80Mb至约90Mb,或约90Mb至约100Mb。另选地,所述缺失可以是至少10kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少150kb、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb,或者至少500kb或更多。

[0264] 在其他情况下,LTVEC和核酸插入物经设计以允许向靶基因座插入在约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,约150kb至约200kb,约200kb至约250kb,约250kb至约300kb,约300kb至约350kb,或约350kb至约400kb的范围内的外源核酸序列。另选地,该插入可为约400kb至约450kb,约450kb至约500kb,约500kb至约550kb,约550kb至约600kb,约600kb至约650kb,约650kb至约700kb,约700kb至约750kb,或约750kb至约800kb。另选地,所述插入可

以是至少10kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少150kb、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb,或者至少500kb或更多。

[0265] 在其他情况下,核酸插入物和/或内源性基因座中所缺失的区域是至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或更大。

[0266] E. 基因组和靶基因组基因座

[0267] 通过本文所公开的方法修饰的基因组或靶基因组基因座可包含细胞内DNA的任何区段或区域。基因组或靶基因组基因座对细胞可以是天然的,可以是整合到细胞基因组中的DNA的异源或外源区段,或者可以是它们的组合。此类异源或外源DNA区段可包括转基因、表达盒、编码选择标记物的多核苷酸,或者基因组DNA的异源或外源区域。

[0268] 基因组或靶基因组基因座还可包含细胞中的染色体外DNA,诸如酵母人工染色体(YAC)、细菌人工染色体(BAC)、人类人工染色体,或包含在适当宿主细胞中的任何其他经改造的基因组区域。

[0269] III. 核酸酶试剂

[0270] 本文提供的用于修饰靶基因组基因座的方法和组合物可采用在所需识别位点中诱导切口或双链断裂的核酸酶试剂。

[0271] 术语“核酸酶试剂的识别位点”包括核酸酶试剂在其处诱导切口或双链断裂的DNA序列。核酸酶试剂的识别位点对于细胞可为内源的(或天然的),或识别位点对于细胞可为外源的。在具体实施方案中,识别位点对于细胞为外源的,从而在细胞基因组中不是天然存在的。在另一个实施方案中,识别位点对于细胞为外源的,并且对于希望被定位在靶基因座处的所关注多核苷酸为外源的。在进一步的实施方案中,外源或内源识别位点在宿主细胞的基因组中仅出现一次。在具体实施方案中,鉴定了在基因组内仅出现一次的内源或天然位点。然后可使用这种位点来设计将在内源识别位点处产生切口或双链断裂的核酸酶试剂。

[0272] 识别位点的长度可变,并且包括例如对于锌指核酸酶(ZFN)对为约30-36bp(即,对于每个ZFN为约15-18bp)、对于转录激活因子样效应物核酸酶(TALEN)为约36bp、或对于CRISPR/Cas9向导RNA为约20bp的识别位点。

[0273] 可在本文所公开的方法和组合物中使用任何会在所需识别位点中诱导切口或双链断裂的核酸酶试剂。可采用天然存在的或天然的核酸酶试剂,只要核酸酶试剂在所需识别位点中诱导切口或双链断裂即可。作为另一种选择,可采用经修饰或经改造的核酸酶试剂。“经改造的核酸酶试剂”包括由其天然形式改造(修饰或衍生)成在所需识别位点中特异性识别并诱导切口或双链断裂的核酸酶。因此,经改造的核酸酶试剂可衍生自原生的或天然存在的核酸酶试剂,或其可人工生成或合成。核酸酶试剂的修饰在蛋白切割剂中可少至一个氨基酸,或在核酸切割剂中可少至一个核苷酸。在一些实施方案中,经改造的核酸酶在识别位点中诱导切口或双链断裂,其中所述识别位点不是会被天然(未经改造的或未经修饰的)核酸酶试剂识别的序列。在识别位点或其他DNA中产生切口或双链断裂在本文中可称为“切开”或“切割”识别位点或其他DNA。

[0274] 本文还提供了示例性识别位点的活性变体和片段。此类活性变体可与给定识别位

点具有至少65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列同一性,其中所述活性变体保留生物活性,从而能够被核酸酶试剂以序列特异性方式识别并切割。测量核酸酶试剂对识别位点造成的双链断裂的测定法是本领域已知的(例如,TAQMAN[®] qPCR测定法,Frendewey D. et al., *Methods in Enzymology*, 2010, 476:295-307 (Frendewey D. 等人,《酶学方法》,2010年,第476卷,第295-307页),该文献全文以引用的方式并入本文)。

[0275] 核酸酶试剂的识别位点可以位于靶基因座中或附近的任何位置。识别位点可位于基因的编码区内,或位于影响基因表达的调控区内。核酸酶试剂的识别位点可以位于内含子、外显子、启动子、增强子、调控区,或任何非蛋白质编码区中。在具体实施方案中,识别位点被定位在编码选择标记的多核苷酸内。这种位置可位于选择标记的编码区内或者位于影响选择标记的表达的调控区内。因此,核酸酶试剂的识别位点可位于选择标记的内含子、编码选择标记的多核苷酸的启动子、增强子、调控区或任何非蛋白编码区中。在具体实施方案中,识别位点处的切口或双链断裂会破坏选择标记的活性。测定功能选择标记存在与否的方法是已知的。

[0276] 在一个实施方案中,核酸酶试剂为转录激活因子样效应物核酸酶(TALEN)。TAL效应物核酸酶是一类序列特异性核酸酶,其可用于在原核或真核生物基因组中的特定靶序列处产生双链断裂。可通过将天然的或经改造的转录激活因子样(TAL)效应物或其功能部分融合到内切核酸酶如FokI的催化结构域,来生成TAL效应物核酸酶。独特的模块化TAL效应物DNA结合结构域使得可以设计潜在地具有任何给定DNA识别特异性的蛋白质。因此,TAL效应物核酸酶的DNA结合结构域可被改造成识别特定DNA靶位点,故可用于在所需靶序列处产生双链断裂。参见W02010/079430;Morbitzer et al. (2010) PNAS 10.1073/pnas.1013133107 (Morbitzer 等人,2010年,《美国国家科学院院刊》,10.1073/pnas.1013133107);Scholze&Boch (2010) *Virulence* 1:428-432 (Scholze和Boch,2010年,《毒力》,第1卷,第428-432页);Christian et al. *Genetics* (2010) 186:757-761 (Christian 等人,《遗传学》,2010年,第186卷,第757-761页);Li et al. (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704 (Li 等人,2010年,《核酸研究》,2010年,doi:10.1093/nar/gkq704);以及Miller et al. (2011) *Nature Biotechnology* 29:143-148 (Miller 等人,2011年,《自然生物技术》,第29卷,第143-148页);所有这些文献均以引用的方式并入本文。

[0277] 合适TAL核酸酶以及用于制备合适TAL核酸酶的方法的示例公开于例如US 2011/0239315 A1、US 2011/0269234 A1、US 2011/0145940 A1、US2003/0232410 A1、US 2005/0208489 A1、US 2005/0026157 A1、US2005/0064474 A1、US 2006/0188987 A1和US 2006/0063231 A1中(每一份专利申请均据此以引用的方式并入本文)。在各种实施方案中,TAL效应物核酸酶被改造成在例如所关注的基因座或所关注的基因组基因座中的靶核酸序列中或附近进行切割,其中所述靶核酸序列位于靶向载体将要修饰的序列处或附近。适合与本文所提供的各种方法和组合物一起使用的TAL核酸酶,包括被专门设计成在本文所述的靶向载体将要修饰的靶核酸序列之处或附近进行结合的那些TAL核酸酶。

[0278] 在一个实施方案中,TALEN的每个单体包含经由两个高变残基识别单碱基对的33-35个TAL重复序列。在一个实施方案中,核酸酶试剂为嵌合蛋白,其包含有效连接至独立核酸酶的基于TAL重复序列的DNA结合结构域。在一个实施方案中,独立核酸酶为FokI内切核

酸酶。在一个实施方案中,核酸酶试剂包含第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域,其中所述第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域中的每一者均有效连接至FokI核酸酶,其中所述第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域识别每条DNA靶序列中被可变长度(12-20bp)的间隔序列隔开的两条邻接DNA靶序列,并且其中所述FokI核酸酶亚基发生二聚化,从而生成能在靶序列处产生双链断裂的活性核酸酶。

[0279] 在本文所公开的各种方法和组合物中采用的核酸酶试剂还可包括锌指核酸酶(ZFN)。在一个实施方案中,ZFN的每个单体包含3个或更多个基于锌指的DNA结合结构域,其中每个基于锌指的DNA结合结构域结合于3bp亚位点。在其他实施方案中,ZFN为包含有效连接至独立核酸酶的、基于锌指的DNA结合结构域的嵌合蛋白。在一个实施方案中,独立内切核酸酶为FokI内切核酸酶。在一个实施方案中,核酸酶试剂包含第一ZFN和第二ZFN,其中所述第一ZFN和第二ZFN中的每一者均有效连接至FokI核酸酶亚基,其中所述第一ZFN和第二ZFN识别DNA靶序列每条链中被约5-7bp间隔序列隔开的两条邻接DNA靶序列,并且其中所述FokI核酸酶亚基二聚化以生成能产生双链断裂的活性核酸酶。参见例如US20060246567; US20080182332; US20020081614; US20030021776; WO/2002/057308A2; US20130123484; US20100291048; WO/2011/017293A2; 以及Gaj et al. (2013) Trends in Biotechnology, 31 (7):397-405 (Gaj等人,2013年,《生物技术趋势》,第31卷,第7期,第397-405页); 这些文献中的每一篇均以引用的方式并入本文。

[0280] 在又一个实施方案中,核酸酶试剂为大范围核酸酶。已基于保守序列基序将大范围核酸酶分类为四个家族,这些家族为LAGLIDADG、GIY-YIG、H-N-H和His-Cys框家族。这些基序参与金属离子的配位和磷酸二酯键的水解。大范围核酸酶以其长识别位点以及耐受其DNA底物中的一些序列多态性而著称。大范围核酸酶结构域、结构和功能是已知的,参见例如,Guhan and Muniyappa (2003) Crit Rev Biochem Mol Biol 38:199-248 (Guhan和Muniyappa,2003年,《生物化学与分子生物学评论》,第38卷,第199-248页); Lucas et al., (2001) Nucleic Acids Res 29:960-9 (Lucas等人,2001年,《核酸研究》,第29卷,第960-969页); Jurica and Stoddard, (1999) Cell Mol Life Sci 55:1304-26 (Jurica和Stoddard,1999年,《细胞和分子生命科学》,第55卷,第1304-1326页); Stoddard, (2006) Q Rev Biophys 38:49-95 (Stoddard,2006年,《生物物理学季评》,第38卷,第49-95页); 以及Moure et al., (2002) Nat Struct Biol 9:764 (Moure等人,2002年,《自然结构生物学》,第9卷,第764页)。在一些示例中,使用天然存在的变体和/或经改造的衍生大范围核酸酶。用于调整动力学、辅因子相互作用、表达、最适条件和/或识别位点特异性及活性筛选的方法是已知的,参见例如,Epinat et al., (2003) Nucleic Acids Res 31:2952-62 (Epinat等人,2003年,《核酸研究》,第31卷,第2952-2962页); Chevalier et al., (2002) Mol Cell 10:895-905 (Chevalier等人,2002年,《分子细胞》,第10卷,第895-905页); Gimble et al., (2003) Mol Biol 334:993-1008 (Gimble等人,2003年,《分子生物学》,第334卷,第993-1008页); Seligman et al., (2002) Nucleic Acids Res 30:3870-9 (Seligman等人,2002年,《核酸研究》,第30卷,第3870-3879页); Sussman et al., (2004) J Mol Biol 342:31-41 (Sussman等人,2004年,《分子生物学杂志》,第342卷,第31-41页); Rosen et al., (2006) Nucleic Acids Res 34:4791-800 (Rosen等人,2006年,《核酸研究》,第34卷,第4791-4800页);

Chames et al., (2005) *Nucleic Acids Res* 33:e178 (Chames等人, 2005年, 《核酸研究》, 第33卷, 第e178页); Smith et al., (2006) *Nucleic Acids Res* 34:e149 (Smith等人, 2006年, 《核酸研究》, 第34卷, 第e149页); Gruen et al., (2002) *Nucleic Acids Res* 30:e29 (Gruen等人, 2002年, 《核酸研究》, 第30卷, 第e29页); Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154 (Chen和Zhao, 2005年, 《核酸研究》, 第33卷, 第e154页); W02005105989; W02003078619; W02006097854; W02006097853; W02006097784; 以及W02004031346。

[0281] 可在本发明中使用任何大范围核酸酶, 包括但不限于I-SceI、I-SceII、I-SceIII、I-SceIV、I-SceV、I-SceVI、I-SceVII、I-CeuI、I-CeuAIIIP、I-CreI、I-CrepsbIP、I-CrepsbIIP、I-CrepsbIIIP、I-CrepsbIIVP、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。

[0282] 在一个实施方案中, 所述大范围核酸酶识别12至40个碱基对的双链DNA序列。在一个实施方案中, 所述大范围核酸酶识别基因组中的一个完全匹配的靶序列。在一个实施方案中, 所述大范围核酸酶为归巢核酸酶。在一个实施方案中, 所述归巢核酸酶为归巢核酸酶的LAGLIDADG家族。在一个实施方案中, 归巢核酸酶的LAGLIDADG家族选自I-SceI、I-CreI和I-DmoI。

[0283] 核酸酶试剂还可包括限制性核酸内切酶, 它们包括I型、II型、III型和IV型核酸内切酶。I型和III型限制性内切核酸酶识别特定识别位点, 但通常在离核酸酶结合位点的可变位置处切割, 该核酸酶结合位点离切割位点(识别位点)可达数百个碱基对。在II型系统中, 酶切活性独立于任何甲基化酶活性, 并且通常在结合位点之内或附近的特定位点处发生切割。大多数II型酶切开回文序列, 但是IIa型酶识别非回文识别位点并在识别位点之外切割, IIb型酶在识别位点之外的两个位点处切开序列两次, 并且IIs型酶识别非对称识别位点并在一侧且离识别位点约1-20个核苷酸的限定距离处切割。IV型限制性内切酶靶向甲基化DNA。限制性内切酶进一步在例如REBASE数据库中进行说明和分类(地址为rebase.neb.com的网页; Roberts et al., (2003) *Nucleic Acids Res* 31:418-20 (Roberts等人, 2003年, 《核酸研究》, 第31卷, 第418-420页), Roberts et al., (2003) *Nucleic Acids Res* 31:1805-12 (Roberts等人, 2003年, 《核酸研究》, 第31卷, 第1805-1812页), 以及Belfort et al., (2002) in *Mobile DNA II*, pp.761-783, Eds. Craigie et al., (ASM Press, Washington, DC) (Belfort等人, 2002年, 载于《可移动的DNA II》, 第761-783页, Craigie等人编辑, 美国华盛顿特区ASM出版社)。

[0284] 在各种方法和组合中采用的核酸酶试剂还可以包含成簇的规律间隔短回文重复序列(CRISPR)/CRISPR相关(Cas)体系或此类体系的组分。CRISPR/Cas系统包括参与Cas

基因的表达或指导Cas基因的活性的转录物和其他元件。CRISPR/Cas系统可为I型、II型或III型系统。本文所公开的方法和组合物通过利用CRISPR复合物(包含与Cas蛋白复合的向导RNA(gRNA))来采用CRISPR/Cas系统对核酸进行定点切割。

[0285] 用于本文所公开的方法中的一些CRISPR/Cas系统为非天然存在的。“非天然存在的”系统包括任何表明受到人工干预的系统,诸如该系统的一个或多个组分从其天然存在的状态改变或突变,至少基本上不含其在自然界中与其天然关联的至少一个其他组分,或和与其天然关联的至少一个其他组分相关联。例如,一些CRISPR/Cas系统采用非天然存在的CRISPR复合物,这些复合物包含在天然情况下不会同时存在的gRNA和Cas蛋白。

[0286] Cas蛋白一般包含至少一个RNA识别或结合结构域。此类结构域可与向导RNA(gRNA,下文更详细地说明)相互作用。Cas蛋白还可包含核酸酶结构域(例如,DNA酶或RNA酶结构域)、DNA结合结构域、解旋酶结构域、蛋白-蛋白相互作用结构域、二聚化结构域以及其他结构域。核酸酶结构域具有用于核酸切割的催化活性。切割包括核酸分子共价键的断裂。切割可产生平头末端或交错末端,并且其可为单链或双链的。Cas蛋白可具有完全切割活性并在靶基因组基因座处形成双链断裂(例如,具有平头末端的双链断裂),或其可为在靶基因组基因座处形成单链断裂的切口酶。

[0287] Cas蛋白的示例包括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,以及它们的同源物或修饰形式。

[0288] 在一些情况下,Cas蛋白来源于II型CRISPR/Cas体系。例如,Cas蛋白可为Cas9蛋白或衍生自Cas9蛋白。这些Cas9蛋白通常共用具有保守架构的四个关键基序。基序1、2和4为RuvC样基序,并且基序3为HNH基序。Cas9蛋白可来自例如化脓性链球菌(*Streptococcus pyogenes*)、嗜热链球菌(*Streptococcus thermophilus*)、链球菌属物种(*Streptococcus sp.*)、达氏拟诺卡氏菌(*Nocardiopsis 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*)、酸热脂环酸芽孢杆菌(*Alicyclobacillus acidocaldarius*)、假蕈状芽孢杆菌(*Bacillus pseudomycoides*)、还原硒酸盐芽孢杆菌(*Bacillus selenitireducens*)、西伯利亚微小杆菌(*Exiguobacterium sibiricum*)、德氏乳杆菌(*Lactobacillus delbrueckii*)、唾液乳杆菌(*Lactobacillus salivarius*)、海洋微颤菌(*Microscilla marina*)、伯克氏菌(*Burkholderiales bacterium*)、萘降解极地单胞菌(*Polaromonas naphthalenivorans*)、极地单胞菌属物种(*Polaromonas sp.*)、瓦氏鳄球藻(*Crocospaera watsonii*)、蓝杆藻属物种(*Cyanothece sp.*)、铜绿微囊藻(*Microcystis aeruginosa*)、聚球藻属物种(*Synechococcus sp.*)、阿拉伯糖醋盐杆菌(*Acetohalobium arabaticum*)、制氮菌(*Ammonifex degensii*)、热解纤维素菌(*Caldicelulosiruptor beccii*)、*Candidatus Desulforudis*、肉毒梭菌(*Clostridium botulinum*)、艰难梭菌(*Clostridium difficile*)、大芬戈尔德菌(*Fingoldia magna*)、嗜热盐碱厌氧菌

(*Natranaerobius thermophilus*)、丙酸互营细菌(*Pelotomaculum thermopropionicum*)、喜温嗜酸硫杆菌(*Acidithiobacillus caldus*)、嗜酸氧化亚铁硫杆菌(*Acidithiobacillus ferrooxidans*)、紫色硫细菌(*Allochromatium vinosum*)、海杆菌属物种(*Marinobacter* sp.)、嗜盐亚硝化球菌(*Nitrosococcus halophilus*)、瓦氏亚硝化球菌(*Nitrosococcus watsoni*)、游海假交替单胞菌(*Pseudoalteromonashaloplanktis*)、纤线杆菌(*Ktedonobacter racemifer*)、甲烷盐菌(*Methanohalobium evestigatum*)、多变鱼腥藻(*Anabaena variabilis*)、泡沫节球藻(*Nodularia spumigena*)、念珠藻属物种(*Nostoc* sp.)、极大节螺藻(*Arthrospira maxima*)、钝顶节螺藻(*Arthrospira platensis*)、节螺藻属物种(*Arthrospira* sp.)、鞘丝藻属物种(*Lyngbya* sp.)、原型微鞘藻(*Microcoleus chthonoplastes*)、颤藻属物种(*Oscillatoria* sp.)、运动石袍菌(*Petrotoga mobilis*)、非洲栖热腔菌(*Thermosiphon africanus*)、或深海单细胞蓝细菌(*Acaryochloris marina*)。Cas9蛋白还可来自金黄色葡萄球菌(*Staphylococcus aureus*)。Cas9家族成员的附加示例包括在W0 2014/131833中描述的那些,该专利全文以引用的方式并入本文。在一个具体示例中,Cas9蛋白是来自化脓性链球菌(*S. pyogenes*)或其衍生的Cas9蛋白。来自化脓性链球菌的Cas9蛋白的氨基酸序列可以在例如SwissProt数据库中以登录号Q99ZW2找到。

[0289] Cas蛋白可为野生型蛋白(即,自然界存在的蛋白)、经修饰的Cas蛋白(即,Cas蛋白变体)、或者野生型或经修饰的Cas蛋白的片段。Cas蛋白也可以是野生型或经修饰的Cas蛋白的活性变体或片段。活性变体或片段可与野生型或经修饰的Cas蛋白或者其一部分具有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列同一性,其中所述活性变体保留了在所需切割位点处切开的能力,从而保留了切口诱导活性或双链断裂诱导活性。针对切口诱导活性或双链断裂诱导活性的测定法是已知的,并且一般测量Cas蛋白对包含切割位点的DNA底物的总体活性和特异性。

[0290] 可修饰Cas蛋白以提高或降低核酸结合亲和力、核酸结合特异性和/或酶活性。还可修饰Cas蛋白以改变蛋白的任何其他活性或特性,诸如稳定性。例如,Cas蛋白的一个或多个核酸酶结构域可以被修饰、缺失或失活,或者Cas蛋白可以被截短以去除对于蛋白质的功能并非必要的结构域,或优化(例如,增强或降低)Cas蛋白的活性。

[0291] 一些Cas蛋白包含至少两个核酸酶结构域,诸如DNA酶结构域。例如,Cas9蛋白可包含RuvC样核酸酶结构域和HNH样核酸酶结构域。RuvC结构域和HNH结构域各自可切开双链DNA的不同链,从而在DNA中产生双链断裂。参见例如Jinek et al. (2012) *Science* 337: 816-821 (Jinek等人,2012年,《科学》,第337卷,第816-821页),该文献全文据此以引用的方式并入。

[0292] 这些核酸酶结构域中的一者或两者可以被缺失或突变,使得它们不再具有功能或具有降低的核酸酶活性。如果核酸酶结构域之一被缺失或突变,则所得的Cas蛋白(例如,Cas9)可称为切口酶,并且可在双链DNA内的靶序列处生成单链断裂,但不会生成双链断裂(即,其可切割互补链或非互补链,但无法同时切割两者)。如果这两个核酸酶结构域都被缺失或突变,则所得的Cas蛋白(例如,Cas9)将具有降低的切割双链DNA两条链的能力(例如,无核酸酶的Cas蛋白)。将Cas9转变为切口酶的突变的示例是来自化脓性链球菌的Cas9的RuvC结构域中的D10A(Cas9的第10位处天冬氨酸至丙氨酸)突变。同样,来自化脓性链球菌的Cas9的HNH结构域中的H939A(氨基酸位置839处组氨酸至丙氨酸)或H840A(氨基酸位置

840处组氨酸至丙氨酸)可将Cas9转变为切口酶。将Cas9转变为切口酶的其他示例包括来自嗜热链球菌(*S. thermophilus*)的Cas9的对应突变。参见例如Saprunauskas et al. (2011) *Nucleic Acids Research* 39:9275-9282 (Saprunauskas等人, 2011年,《核酸研究》,第39卷,第9275-9282页)和W0 2013/141680,这些文献中的每一篇全文均以引用的方式并入本文。此类突变可使用诸如定点诱变、PCR介导的诱变或全基因合成的熟知方法来生成。其他形成切口酶的突变的示例可见于例如W0/2013/176772A1和W0/2013/142578A1中,这些专利中的每一篇均以引用的方式并入本文。

[0293] Cas蛋白也可为融合蛋白。例如,Cas蛋白可融合到切割结构域、表观遗传修饰结构域、转录激活结构域或转录阻遏物结构域。参见W02014/089290,该专利全文以引用的方式并入本文。Cas蛋白也可融合到异源多肽,从而提供增强或减弱的稳定性。融合的结构域或异源多肽可位于N端、C端或Cas蛋白的内部。

[0294] Cas融合蛋白的一个示例是与提供亚细胞定位的异源多肽融合的Cas蛋白。此类序列可以包括例如用于靶向细胞核的核定位信号(NLS)如SV40NLS、用于靶向线粒体的线粒体定位信号、ER滞留信号等。参见例如Lange et al. (2007) *J. Biol. Chem.* 282:5101-5105 (Lange等人, 2007年,《生物化学杂志》,第282卷,第5101-5105页)。Cas蛋白可包括例如一种或多种核定位信号(例如,两种核定位信号)。此类亚细胞定位信号可位于N端、C端或Cas蛋白内的任何位置处。NLS可包含一段碱性氨基酸,并且可为单分型(monopartite)序列或双分型(bipartite)序列。

[0295] Cas蛋白也可包含细胞穿透结构域。例如,细胞穿透结构域可衍生自HIV-1TAT蛋白、来自人乙肝病毒的TLM细胞穿透基序、MPG、Pep-1、VP22、来自单纯性疱疹病毒的细胞穿透肽、或多聚精氨酸肽序列。参见例如W0 2014/089290,该专利全文以引用的方式并入本文。细胞穿透结构域可位于N端、C端或Cas蛋白内的任何位置处。

[0296] Cas蛋白还可包含便于示踪或纯化的异源多肽,诸如荧光蛋白、纯化标签或表位标签。荧光蛋白的示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(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(mOrange、mKO、Kusabira-Orange、Monomeric Kusabira-Orange、mTangerine、tdTomato)以及任何其他合适的荧光蛋白。标签的示例包括谷胱甘肽-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)、多聚(NANP)、串联亲和纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、血凝素(HA)、nus、Softag 1、Softag 3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、组氨酸(His)、生物素羧基载体蛋白(BCCP)以及钙调蛋白。

[0297] Cas蛋白可以任何形式提供。例如,Cas蛋白可以蛋白的形式提供,诸如与gRNA复合的Cas蛋白。作为另一种选择,Cas蛋白可以编码Cas蛋白的核酸的形式提供,诸如RNA(例如,信使RNA(mRNA))或DNA。任选地,编码Cas蛋白的核酸可进行密码子优化,以在特定细胞或生物体中有效翻译成蛋白。当将编码Cas蛋白的核酸引入细胞中时,Cas蛋白可在细胞中瞬时

地、有条件地或组成性地表达。

[0298] 编码Cas蛋白的核酸可稳定整合在细胞的基因组中,并有效连接至细胞中有活性的启动子。作为另一种选择,编码Cas蛋白的核酸可有效连接至表达构建体中的启动子。表达构建体包括任何能够指导目标基因或其他核酸序列(例如,Cas基因)的表达并可将这种目标核酸序列转移到靶细胞中的核酸构建体。例如,编码Cas蛋白的核酸可位于包含编码gRNA的DNA的载体中。另选地,其可位于与包含编码gRNA的DNA的载体分离的载体或质粒中。可用于表达构建体的启动子包括例如在大鼠、真核生物、哺乳动物、非人哺乳动物、人、啮齿动物、小鼠或仓鼠多能细胞中有活性的启动子。其他启动子的示例在本文别处有描述。

[0299] “向导RNA”或“gRNA”包括结合到Cas蛋白并使Cas蛋白靶向靶DNA内特定位置的RNA分子。向导RNA可包含两个区段:“DNA靶向区段”和“蛋白结合区段”。“区段”包括分子的区段、部分或区域,诸如RNA中的一个邻接核苷酸段。一些gRNA包含两个单独的RNA分子:“激活因子-RNA”和“靶向因子-RNA”。其他gRNA为单个RNA分子(单条RNA多核苷酸),其也可称为“单分子gRNA”、“单向导RNA”或“sgRNA”。参见例如WO/2013/176772A1、WO/2014/065596A1、WO/2014/089290A1、WO/2014/093622A2、WO/2014/099750A2、WO/2013142578A1以及WO2014/131833A1,这些专利中的每一篇均以引用的方式并入本文。术语“向导RNA”和“gRNA”是包含性的,包括双分子gRNA和单分子gRNA。

[0300] 示例性双分子gRNA包含crRNA样(“CRISPR RNA”或“靶向因子-RNA”或“crRNA”或“crRNA重复序列”)分子以及对应的tracrRNA样(“反式作用CRISPR RNA”或“激活因子-RNA”或“tracrRNA”或“支架”)分子。crRNA包含gRNA的DNA靶向区段(单链)和一段核苷酸,该段核苷酸形成gRNA的蛋白结合区段的dsRNA双链体的一半。

[0301] 对应的tracrRNA(激活因子-RNA)包含一段核苷酸,该段核苷酸形成gRNA的蛋白结合区段的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,从而形成gRNA的蛋白结合结构域的dsRNA双链体。因此,每个crRNA可以说成具有对应的tracrRNA。

[0302] crRNA和对应的tracrRNA杂交以形成gRNA。crRNA另外提供了与靶序列杂交的单链DNA靶向区段。如果用于细胞内的修饰,则给定crRNA或tracrRNA分子的确切序列可被设计成对于将在其中使用这些RNA分子的物种具有特异性。参见例如Mali et al. (2013) *Science* 339:823-826 (Mali等人,2013年,《科学》,第339卷,第823-826页);Jinek et al. (2012) *Science*337:816-821 (Jinek等人,2012年,《科学》,第337卷,第816-821页);Hwang et al. (2013) *Nat. Biotechnol.* 31:227-229 (Hwang等人,2013年,《自然-生物技术》,第31卷,第227-229页);Jiang et al. (2013) *Nat. Biotechnol.* 31:233-239 (Jiang等人,2013年,《自然-生物技术》,第31卷,第233-239页);以及Cong et al. (2013) *Science* 339:819-823 (Cong等人,2013年,《科学》,第339卷,第819-823页),这些文献中的每一篇以引用方式并入本文。

[0303] 给定gRNA的DNA靶向区段(crRNA)包含与靶DNA中的序列互补的核苷酸序列。gRNA的DNA靶向区段通过杂交(即,碱基配对)以序列特异性方式与靶DNA相互作用。因此,DNA靶向区段的核苷酸序列可变化,并且决定将与gRNA和靶DNA相互作用的靶DNA内的位置。可修饰个体gRNA的DNA靶向区段,以与靶DNA内的任何所需序列杂交。天然存在的crRNA随Cas9系统和生物体不同而不同,但通常包含21至72个核苷酸长的靶向区段,该靶向区段被21至46

个核苷酸长的两个正向重复序列 (DR) 侧接 (参见例如WO2014/131833)。就化脓性链球菌而言,DR为36个核苷酸长,并且靶向区段为30个核苷酸长。位于3'的DR与对应的tracrRNA互补并杂交,继而结合于Cas9蛋白。

[0304] DNA靶向区段的长度可为约12个核苷酸至约100个核苷酸。例如,DNA靶向区段的长度可为约12个核苷酸(nt)至约80nt、约12nt至约50nt、约12nt至约40nt、约12nt至约30nt、约12nt至约25nt、约12nt至约20nt、或约12nt至约19nt。作为另一种选择,DNA靶向区段的长度可为约19nt至约20nt、约19nt至约25nt、约19nt至约30nt、约19nt至约35nt、约19nt至约40nt、约19nt至约45nt、约19nt至约50nt、约19nt至约60nt、约19nt至约70nt、约19nt至约80nt、约19nt至约90nt、约19nt至约100nt、约20nt至约25nt、约20nt至约30nt、约20nt至约35nt、约20nt至约40nt、约20nt至约45nt、约20nt至约50nt、约20nt至约60nt、约20nt至约70nt、约20nt至约80nt、约20nt至约90nt、或约20nt至约100nt。

[0305] 与靶DNA的核苷酸序列(靶序列)互补的DNA靶向区段的核苷酸序列的长度可为至少约12nt。例如,DNA靶向序列(即,与靶DNA内的靶序列互补的DNA靶向区段内的序列)可具有至少约12nt、至少约15nt、至少约18nt、至少约19nt、至少约20nt、至少约25nt、至少约30nt、至少约35nt或至少约40nt的长度。另选地,DNA靶向序列的长度可为约12个核苷酸(nt)至约80nt、约12nt至约50nt、约12nt至约45nt、约12nt至约40nt、约12nt至约35nt、约12nt至约30nt、约12nt至约25nt、约12nt至约20nt、约12nt至约19nt、约19nt至约20nt、约19nt至约25nt、约19nt至约30nt、约19nt至约35nt、约19nt至约40nt、约19nt至约45nt、约19nt至约50nt、约19nt至约60nt、约20nt至约25nt、约20nt至约30nt、约20nt至约35nt、约20nt至约40nt、约20nt至约45nt、约20nt至约50nt,或约20nt至约60nt。在一些情况下,DNA靶向序列可具有至少约20nt的长度。

[0306] TracrRNA可为任何形式(例如,全长tracrRNA或有活性的部分tracrRNA)并具有不同长度。它们可包括初级转录物或加工形式。例如,tracrRNA(作为单向导RNA的一部分或作为属于双分子gRNA的一部分的单独分子)可包含以下部分或由以下部分组成:野生型tracrRNA序列的全部或一部分(例如,野生型tracrRNA序列的约或大于约20、26、32、45、48、54、63、67、85个或更多个核苷酸)。来自化脓性链球菌的野生型tracrRNA序列的示例包括171个核苷酸、89个核苷酸、75个核苷酸以及65个核苷酸的形式。参见例如Deltcheva et al. (2011) Nature 471:602-607 (Deltcheva等人,2011年,《自然》,第471卷,第602-607页); WO 2014/093661,这些文献中的每一篇全文均以引用的方式并入本文。单向导RNA (sgRNA) 内的tracrRNA的示例包括存在于+48、+54、+67和+85形式的sgRNA内的tracrRNA区段,其中“+n”表示野生型tracrRNA的至多+n核苷酸包含在sgRNA中。参见US 8,697,359,该专利全文以引用的方式并入本文。

[0307] DNA靶向序列与靶DNA内的靶序列之间的互补性百分比可为至少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%)。在一些情况下,DNA靶向序列与靶DNA内的靶序列之间的互补性百分比在约20个邻接核苷酸内为至少60%。在一个示例中,在靶DNA的互补链内的靶序列的5'端的14个邻接核苷酸内,DNA靶向序列与靶DNA内的靶序列之间的互补性百分比为100%,并且在其余邻接核苷酸内低至0%。在这种情况下,DNA靶向序列可被视为14个核苷酸长。在另一个示例中,在靶DNA的互补链内的靶序列的5'端的七个邻接核苷酸内,DNA靶向序列与靶DNA内

的靶序列之间的互补性百分比为100%，并且在其余邻接核苷酸内低至0%。在这种情况下，DNA靶向序列可被视为7个核苷酸长。

[0308] gRNA的蛋白结合区段可包含彼此互补的两段核苷酸。蛋白结合区段的互补核苷酸杂交而形成双链RNA双链体(dsRNA)。对象gRNA的蛋白结合区段与Cas蛋白相互作用，并且gRNA经由DNA靶向区段指导结合的Cas蛋白到达靶DNA内的特异性核苷酸序列。

[0309] 向导RNA可包括提供额外所需特征(例如,经修饰或调控的稳定性;亚细胞靶向;用荧光标记物示踪;蛋白或蛋白复合物的结合位点;等等)的修饰或序列。此类修饰的例子包括例如5'帽(例如,7-甲基鸟苷酸帽(m7G));3'多聚腺苷酸化尾(即,3'多聚(A)尾);核糖开关序列(例如,以实现经调控的稳定性和/或经调控的蛋白和/或蛋白复合物可及性);稳定性控制序列;形成dsRNA双链体(即,发夹)的序列;使RNA靶向亚细胞位置(例如,细胞核、线粒体、叶绿体等)的修饰或序列;提供示踪的修饰或序列(例如,与荧光分子的直接缀合、与有利于荧光检测的部分的缀合、允许荧光检测的序列等);为蛋白质(例如,作用于DNA的蛋白质,包括转录激活因子、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)提供结合位点的修饰或序列;以及它们的组合。

[0310] 向导RNA可以任何形式提供。例如,gRNA可以RNA的形式(作为两分子(单独的crRNA和tracrRNA)或作为一分子(sgRNA))提供,并任选地以与Cas蛋白的复合物形式提供。gRNA也可以编码gRNA的DNA的形式提供。编码gRNA的DNA可编码单个RNA分子(sgRNA)或单独的RNA分子(例如,单独的crRNA和tracrRNA)。在后一情况下,编码gRNA的DNA可作为分别编码crRNA和tracrRNA的单独DNA分子提供。另选地,编码gRNA的DNA可作为一个DNA分子提供。

[0311] 当将编码gRNA的DNA引入细胞中时,gRNA可在细胞中瞬时地、有条件地或组成性地表达。编码gRNA的DNA可稳定整合在细胞的基因组中,并有效连接至在细胞中有活性的启动子。作为另一种选择,编码gRNA的DNA可有效连接至表达构建体中的启动子。例如,编码gRNA的DNA可位于包含编码Cas蛋白的核酸的载体中。另选地,其可位于与包含编码Cas蛋白的核酸的载体分离的载体或质粒中。可用于此类表达构建体中的启动子包括例如在大鼠、真核生物、哺乳动物、非人哺乳动物、人、啮齿动物、小鼠或仓鼠多能细胞中有活性的启动子。此类启动子可为例如条件启动子、诱导型启动子、组成型启动子或组织特异性启动子。在一些情况下,启动子是RNA聚合酶III启动子,例如人U6启动子。

[0312] 另选地,可通过各种其他方法制备gRNA。例如,可通过采用例如T7RNA聚合酶的体外转录来制备gRNA(参见例如WO 2014/089290和WO2014/065596,这些文献中的每一篇均出于所有目的全文以引用方式并入本文)。向导RNA也可为通过化学合成制备的合成产生的分子。

[0313] CRISPR/Cas体系的靶序列包括靶DNA中存在的这样的核酸序列,只要存在充分的结合条件,gRNA的DNA靶向区段就将与该核酸序列结合。例如,靶序列包括向导RNA被设计成与之具有互补性的序列,其中靶序列与DNA靶向序列之间的杂交促进CRISPR复合物的形成。不必要求完全互补性,只要存在足以引起杂交并促进CRISPR复合物形成的互补性即可。靶序列还包括下文更详细说明的Cas蛋白的切割位点。靶序列可包含任何多核苷酸,所述多核苷酸可位于例如细胞的细胞核或细胞质中,或位于细胞的细胞器如线粒体或叶绿体内。

[0314] 靶DNA内的靶序列可被Cas蛋白或gRNA所靶向(即,与之结合、或与之杂交、或与之互补)。合适的DNA/RNA结合条件包括通常存在于细胞中的生理条件。其他合适的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)(《分子克隆实验指南》,第3版,Sambrook等人,冷泉港实验室出版社,2001年))。与Cas蛋白或gRNA互补并杂交的靶DNA链可称为“互补链”,并且与“互补链”互补(并因此不与Cas蛋白或gRNA互补)的靶DNA链可称为“非互补链”或“模板链”。

[0315] Cas蛋白可在将与gRNA的DNA靶向区段结合的靶DNA中存在的核酸序列之内或之外的位点处切割核酸。“切割位点”包括Cas蛋白产生单链断裂或双链断裂的核酸位置。例如,CRISPR复合物(包含与靶序列杂交并与Cas蛋白复合的gRNA)的形成可导致将与gRNA的DNA靶向区段结合的靶DNA中存在的核酸序列之中或附近(例如,在相距1、2、3、4、5、6、7、8、9、10、20、50个或更多个碱基对内)的一条或两条链切割。如果切割位点位于将与gRNA的DNA靶向区段结合的靶DNA中存在的核酸序列之外,则切割位点仍被视为在“靶序列”内。切割位点可位于核酸的仅一条链上或两条链上。切割位点可位于核酸的两条链上的相同位置处(产生平头末端),或可位于每条链上的不同位点处(产生交错末端(即,悬垂序列))。可例如通过使用两种Cas蛋白来产生交错末端,这些Cas蛋白在不同链上的不同切割位点处产生单链断裂,从而产生双链断裂。例如,第一切口酶可在双链DNA(dsDNA)的第一链上形成单链断裂,而第二切口酶可在dsDNA的第二链上形成单链断裂,使得形成悬垂序列。在一些情况下,第一链上的切口酶的靶序列与第二链上的切口酶的靶序列相隔至少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。

[0316] Cas9对靶DNA的位点特异性切割可在由以下两者决定的位置处发生:(i) gRNA与靶DNA之间的碱基配对互补性,以及(ii)靶DNA中的短基序,称为前间区序列邻近基序(PAM)。PAM可侧接靶序列。任选地,靶序列可以在3'末端上侧接PAM。例如,Cas9的切割位点可为PAM序列上游或下游的约1至约10或者约2至约5个碱基对(例如,3个碱基对)。在一些情况下(例如,当使用来自化脓性链球菌的Cas9或密切相关的Cas9时),非互补链的PAM序列可为5'-N₁GG-3',其中N₁为任何DNA核苷酸并且紧邻靶DNA的非互补链的靶序列的3'。因此,互补链的PAM序列将为5'-CCN₂-3',其中N₂为任何DNA核苷酸并且紧邻靶DNA的互补链的靶序列的5'。在一些此类情况下,N₁和N₂可为互补的,并且N₁-N₂碱基对可为任何碱基对(例如,N₁=C且N₂=G;N₁=G且N₂=C;N₁=A且N₂=T;或N₁=T且N₂=A)。

[0317] 靶序列的示例包括与gRNA的DNA靶向区段互补的DNA序列、或除PAM序列之外的这种DNA序列。靶序列的一个示例包括核苷酸序列GNNNNNNNNNNNNNNNNNNNGG(GN₁₋₂₀GG;SEQ ID NO:1)。5'端的鸟嘌呤可有利于RNA聚合酶在细胞中进行转录。靶序列的其他例子可包括5'端处的两个鸟嘌呤核苷酸,以有利于T7聚合酶在体外进行有效转录。参见例如WO 2014/065596,该专利出于所有目的全文以引用方式并入本文。其他靶序列可具有4-22个核苷酸长的SEQ ID NO:1,包含5'G和3'GG。另外一些靶序列可具有14与20个核苷酸之间长度的SEQ ID NO:1。

[0318] 靶序列可为细胞内源或外源的任何核酸序列。靶序列可为编码基因产物(例如,蛋白质)的序列或非编码序列(例如,调控序列或垃圾DNA)或者可包括两者。

[0319] 本发明还提供了核酸酶试剂的活性变体和片段(即,经改造的核酸酶试剂)。此类活性变体可与天然核酸酶试剂具有至少65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更高的序列同一性,其中所述活性变体保留在所

需识别位点处切割的能力,从而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可由天然内切核酸酶序列修饰而成,并且可被设计成在不被天然核酸酶试剂识别的识别位点处识别并诱导切口或双链断裂。因此,在一些实施方案中,经改造的核酸酶具有在与对应天然核酸酶试剂识别位点不同的识别位点处诱导切口或双链断裂的特异性。针对切口或双链断裂诱导活性的测定法是已知的,并且一般测量内切核酸酶对包含识别位点的DNA底物的总体活性和特异性。

[0320] 可通过本领域已知的任何方式将核酸酶试剂导入多能细胞中。可将编码核酸酶试剂的多肽直接引入细胞中。另选地,可将编码核酸酶试剂的多核苷酸引入细胞中。当将编码核酸酶试剂的多核苷酸引入细胞中时,核酸酶试剂可在细胞内瞬时地、有条件地或组成性地表达。因此,编码核酸酶试剂的多核苷酸可包含在表达盒中,并有效连接至条件启动子、诱导型启动子、组成型启动子或组织特异性启动子。另选地,将核酸酶试剂作为编码核酸酶试剂的mRNA引入细胞中。

[0321] 在具体实施方案中,编码核酸酶试剂的多核苷酸稳定整合在细胞的基因组中,并有效连接至细胞中有活性的启动子。在其他实施方案中,编码核酸酶试剂的多核苷酸位于包含核酸插入物的相同靶向载体中,而在其他情况下,编码核酸酶试剂的多核苷酸位于与包含核酸插入物的靶向载体分离的载体或质粒中。

[0322] 当通过引入编码核酸酶试剂的多核苷酸来向细胞提供核酸酶试剂时,可修饰这种编码核酸酶试剂的多核苷酸,以置换与编码核酸酶试剂的天然存在的多核苷酸序列相比在目标细胞中具有更高使用频率的密码子。例如,可对编码核酸酶试剂的多核苷酸进行修饰,以替换成与天然存在的多核苷酸序列相比在原核细胞或真核细胞(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、仓鼠细胞或任何其他宿主细胞)中具有更高使用频率的密码子。

[0323] 可顺序地重复上述各种方法,以允许任何数量的核酸插入物靶向整合到染色体上给定的所靶向基因组基因座中。因此,所述各种方法可提供以用于向染色体上的靶基因组基因座中插入至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个核酸插入物。在特定实施方案中,此类顺序拼接方法允许将来自动物细胞或来自哺乳动物细胞(即,人、非人、啮齿动物、小鼠、猴、大鼠、仓鼠、家养哺乳动物或农业动物)的大基因组区域重建到染色体上所靶向的基因组基因座中。在此类情况下,包含编码区和非编码区两者的基因组区域的转移和重建允许通过至少部分地保留在天然基因组区域内发现的编码区、非编码区和拷贝数变异来保持给定区的复杂性。因此,所述各种方法提供了例如在细胞内生成“异源”或“外源”基因组区域的方法。

[0324] IV. 选择标记

[0325] 本文所提供的各种方法和组合物可将核酸酶试剂及其对应识别位点与选择标记结合使用。如本文所讨论,编码选择标记的多核苷酸中识别位点的位置可实现在靶基因座处鉴定整合事件的有效方法。此外,本文提供了各种方法,其中采用具有核酸酶识别位点的交替选择标记,以提高多个目标多核苷酸整合在给定所靶向基因座内的效率和功效。

[0326] 可在本文所公开的方法和组合物中使用各种选择标记。此类选择标记可例如赋予对抗生素如G418、潮霉素、杀稻瘟菌素、新霉素或嘌呤霉素的抗性。此类选择标记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-乙酰转移酶(puro^r)以及杀

稻瘟菌素S脱氢酶(bsr^r)。在另一些实施方案中,选择标记有效连接至诱导型启动子,并且选择标记的表达对细胞有毒性。此类选择标记的非限制性示例包括黄嘌呤/鸟嘌呤磷酸核糖转移酶(gpt)、次黄嘌呤-鸟嘌呤磷酸核糖转移酶(HGPRT)、或单纯疱疹病毒胸苷激酶(HSV-TK)。

[0327] 在一个实施方案中,核酸酶识别位点被定位在编码选择标记的基因内。在一个具体实施方案中,核酸酶识别位点被定位在潮霉素基因内。

[0328] 编码选择标记的多核苷酸有效连接至细胞中有活性的启动子。此类表达盒及其各种调控组分在本文别处更详细讨论。

[0329] V. 启动子

[0330] 本文所述的各种核酸序列可有效连接至启动子。此类启动子可例如在真核生物、哺乳动物、非人哺乳动物、人、啮齿动物、小鼠或仓鼠多能细胞中有活性。启动子可为例如组成性活性启动子、条件启动子、诱导型启动子、时间限制型启动子(例如,受发育调控的启动子)或空间限制型启动子(例如,细胞特异性或组织特异性启动子)。启动子的例子可见于例如WO 2013/176772,该专利全文以引用的方式并入本文。

[0331] 诱导型启动子的例子包括例如化学调控型启动子和物理调控型启动子。化学调控型启动子包括例如醇调控型启动子(例如,乙醇脱氢酶(alcA)基因启动子)、四环素调控型启动子(例如,四环素反应性启动子、四环素操纵子序列(tetO)、tet-On启动子或tet-Off启动子)、类固醇调控型启动子(例如,大鼠糖皮质激素受体、雌性激素受体的启动子或蜕化素受体的启动子)或金属调控型启动子(例如,金属蛋白启动子)。物理调控型启动子包括例如温度调控型启动子(例如,热休克启动子)和光调控型启动子(例如,光诱导型启动子或光阻抑型启动子)。

[0332] 组织特异性启动子可以是例如神经元特异性启动子、神经胶质特异性启动子、肌肉细胞特异性启动子、心脏细胞特异性启动子、肾脏细胞特异性启动子、骨细胞特异性启动子、内皮细胞特异性启动子或免疫细胞特异性启动子(例如,B细胞启动子或T细胞启动子)。

[0333] 受发育调控的启动子包括例如只在胚胎发育期或只在成人细胞中有活性的启动子。

[0334] 也可根据细胞类型选择启动子。例如,各种已知的启动子可用于真核细胞、哺乳动物细胞、非人细胞、非人哺乳动物细胞、多能细胞、非人多能细胞、人多能细胞、人ES细胞、人成体干细胞、发育受限的人祖细胞、人iPS细胞、人细胞、啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞。

[0335] VI. 表达盒

[0336] 本文提供了包含本文所提供的靶向系统的各种组成部分(即,核酸酶试剂、识别位点、核酸插入物、所关注的多核苷酸、靶向载体(即,LTVEC)、选择标记以及其他组成部分)的多核苷酸或核酸分子。

[0337] 还提供了包含靶向系统的各种组成部分的重组多核苷酸。术语“重组多核苷酸”和“重组DNA构建体”可在本文中互换使用。重组构建体包含核酸序列的人工或异源组合,例如自然界中不共存的调控序列和编码序列。在其他实施方案中,重组构建体可包含来自不同来源的调控序列和编码序列,或来自相同来源但以与自然界中的存在方式不同的方式排列的调控序列和编码序列。这种构建体可独自使用或可结合载体使用。如果使用载体,则载体

的选择取决于如本领域技术人员众所周知的用以转化宿主细胞的方法。例如,可使用质粒载体。本文提供了成功转化、选择和繁殖宿主细胞所需的且包含任一种分离核酸片段的遗传元件。筛选可尤其通过DNA的Southern分析、mRNA表达的Northern分析、蛋白表达的免疫印迹分析或表型分析来实现。

[0338] 在具体实施方案中,本文所述的靶向系统的组成部分中的一者或多者可提供于表达盒中,以便在原核细胞、真核细胞、细菌、酵母细胞、哺乳动物细胞或其他所关注的生物体或细胞类型中表达。所述盒可包括有效连接至本文所提供的多核苷酸的5' 调控序列和3' 调控序列。当用以提到两个蛋白编码区的连接时,有效连接意指编码区在同一阅读框中。在另一种情况下,编码蛋白的核酸序列可有效连接至调控序列(例如,启动子、增强子、沉默子序列等)以保持恰当的转录调控。

[0339] 所述盒可另外含有将共同引入生物体中的至少一个额外目标多核苷酸。或者,所述额外目标多核苷酸可提供于多个表达盒上。这种表达盒提供有多个限制位点和/或重组位点,以使重组多核苷酸的插入处于调控区的转录调控之下。所述表达盒可另外含有选择标记基因。

[0340] 所述表达盒在5' -3' 转录方向上可包含在哺乳动物细胞或所关注的宿主细胞中起作用的转录和翻译起始区(即,启动子)、本文所提供的重组多核苷酸、以及转录和翻译终止区(即,终止区)。所述调控区(即,启动子、转录调控区和翻译终止区)和/或本文所提供的多核苷酸对于宿主细胞或对于彼此可为天然/类似的。另选地,所述调控区和/或本文所提供的多核苷酸对于宿主细胞或对于彼此可为异源的。例如,有效连接至异源多核苷酸的启动子来自与得到该多核苷酸的物种不同的物种,或者如果来自相同/类似的物种,则对一者或两者由其原始形式和/或基因座进行了实质性修饰,或者启动子不是有效连接的多核苷酸的天然启动子。另选地,所述调控区和/或本文所提供的重组多核苷酸可以是完全合成的。

[0341] 所述终止区对于转录起始区而言可为天然的,对于有效连接的重组多核苷酸而言可为天然的,对于宿主细胞而言可为天然的,或者可来源于对于启动子、重组多核苷酸、宿主细胞或它们的任何组合而言为另一种的(即,外来的或异源的)来源。

[0342] 在制备表达盒时,可对各种DNA片段进行操纵,以便提供处于正确取向的DNA序列。为此目的,可采用衔接子或接头将DNA片段连接在一起,或者可涉及其他的操纵以提供便利的限制性位点、去除多余的DNA、去除限制性位点等。出于这个目的,可能涉及体外诱变、引物修复、限制性酶切、退火、再置换(例如转换和颠换)。

[0343] 多种启动子可用于本文所提供的表达盒中。可根据期望的结果来选择启动子。已经认识到,不同的应用可通过在表达盒中使用不同的启动子来增强,从而调整目标多核苷酸的表达的时机、位置和/或水平。如果需要,此类表达构建体还可含有启动子调控区(例如,赋予可诱导的、组成性的、环境或发育调控的、或细胞或组织特异性/选择性表达的启动子调控区)、转录起始位点、核糖体结合位点、RNA加工信号、转录终止位点和/或多聚腺苷酸化信号。

[0344] 含有本文所提供的多核苷酸的表达盒还可包含用于选择转化细胞的选择标记基因。利用选择标记基因来选择转化细胞或组织。

[0345] 在适当的情况下,可优化在所述方法和组合物(即,目标多核苷酸、核酸酶试剂等)中采用的序列,以便增加在细胞中的表达。也就是说,所述基因可使用给定目标细胞中偏好

的密码子来合成以便提高表达,所述密码子包括例如哺乳动物偏好密码子、人偏好密码子、啮齿动物偏好密码子、小鼠偏好密码子、大鼠偏好密码子等。

[0346] 在一个实施方案中,核酸酶试剂由表达构建体表达,所述表达构建体包含编码核酸酶的核酸序列,并且所述核酸有效连接至细胞中有活性的启动子。

[0347] VII. 制备经遗传修饰的非人动物的方法

[0348] 可采用本文所公开的各种方法产生经遗传修饰的非人动物。在一些情况下,产生经遗传修饰的非人动物的方法包括:(1)使用本文所述的方法修饰多能细胞的基因组;(2)选择经遗传修饰的多能细胞;(3)将经遗传修饰的多能细胞引入宿主胚胎中;以及(4)将包含经遗传修饰的多能细胞的宿主胚胎植入代孕母体中。产生来源于经遗传修饰的多能细胞的子代。可向任何时期诸如胚泡期或桑椹胚前期(即4-细胞期或8-细胞期)的宿主胚胎引入供体细胞。产生能够通过生殖系传递遗传修饰的子代。多能细胞可以是如本文别处讨论的ES细胞(例如小鼠ES细胞或大鼠ES细胞)。参见例如以引用方式并入本文的美国专利号7,294,754。

[0349] 也可使用核移植技术生成非人哺乳动物。简而言之,用于核移植的方法可包括以下步骤:(1)将卵母细胞去核,或者提供去核卵母细胞;(2)分离或提供供体细胞或核,以与去核卵母细胞混合;(3)将所述细胞或核插入到所述去核卵母细胞中,以形成重建细胞;(4)将所述重建细胞植入到动物的子宫中,以形成胚胎;以及(5)允许所述胚胎发育。在此类方法中,一般从处死的动物体内取出卵母细胞,但也可从活动物的输卵管和/或卵巢中分离卵母细胞。卵母细胞可在去核之前在本领域的技术人员已知的多种培养基中成熟。卵母细胞的去核可以本领域普通技术人员所熟知的多种方式进行。可在融合之前在透明带下显微注射供体细胞来将供体细胞或核插入到去核卵母细胞中以形成重建细胞。融合可通过跨接触/融合平面施加直流电脉冲(电融合)、通过将细胞暴露于促进融合的化学品如聚乙二醇或者借助灭活病毒如仙台病毒来诱导。重建细胞可在核供体和受体卵母细胞融合之前、期间和/或之后通过电和/或非电方式激活。激活方法包括电脉冲、化学诱导冲击、精子穿透、增加卵母细胞中二价阳离子水平以及减少卵母细胞中细胞蛋白磷酸化(如借助激酶抑制剂)。激活的重建细胞或胚胎可在本领域的技术人员所熟知的培养基中培养,然后移植到动物的子宫中。参见例如US20080092249、WO/1999/005266A2、US20040177390、WO/2008/017234A1以及美国专利号7,612,250,这些专利中的每一者均以引用的方式并入本文。

[0350] 所述方法还可包括产生F0代非人动物的方法,这些方法包括:(1)鉴定包含靶向修饰的非人ES细胞;(2)将包含靶向修饰的非人ES细胞引入非人宿主胚胎中;以及(3)在代孕母体中孕育非人宿主胚胎。然后代孕母体可产生包含靶向修饰的F0代非人动物。可孵育包含经遗传修饰的多能或全能细胞(例如,非人ES细胞)的宿主胚胎直至胚泡期,随后将其植入代孕母体中以产生F0动物。可经由如本文所述的等位基因修饰(MOA)测定法鉴定具有经遗传修饰的基因组基因座的动物。

[0351] 本文所提供的各种方法允许产生经遗传修饰的非人F0动物,其中经遗传修饰的F0动物的细胞包含靶向修饰。已经认识到,根据用于产生F0动物的方法的不同,F0动物体内具有靶向遗传修饰的细胞的数量将有所变化。经由例如**VELOCIMOUSE**[®]方法将供体ES细胞引入对应生物体的桑椹胚前期胚胎(例如,8细胞期小鼠胚胎)中,允许F0动物的更大百分比的细胞群包含具有靶向遗传修饰的细胞。在特定情况下,非人F0动物的细胞比例的至少

50%、60%、65%、70%、75%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%包含具有靶向修饰的细胞群。在其他情况下，F0动物的生殖细胞中的至少一个或多个具有靶向修饰。

[0352] 在一些情况下，经遗传修饰的F0动物的细胞对于靶向修饰而言为杂合的或复合杂合的。例如，经遗传修饰的F0动物的细胞对于靶向修饰而言可为半合子的。在其他情况下，经遗传修饰的F0动物的细胞对于靶向修饰而言为纯合的。

[0353] 在一些情况下，由本文所公开的方法和组合物产生的F0动物可以被培育成野生型动物，以产生对于靶向修饰而言为杂合的F1代。随后，来自F1代的动物可以相互繁殖，以产生对于靶向修饰而言为纯合的F2动物。可使用特异性引物和/或探针对F1子代进行基因分型，以确定靶向遗传修饰是否存在。

[0354] VIII. 将核酸和蛋白引入细胞中的方法

[0355] 本文提供各种方法和组合物以允许将核酸引入到细胞中。在一些情况下，用于引入核酸的系统允许在特定基因组基因座处的靶向整合。此类系统采用多种组成部分，并且为了易于参考，术语“靶向基因组整合系统”通常包括整合事件所需的所有组成部分（例如，核酸酶试剂、核酸酶切割位点、核酸插入物、靶向载体、靶基因组基因座和所关注的多核苷酸中的一者或多者）。

[0356] 本文所提供的方法可包括向细胞中引入包含靶向基因组整合系统的一种或多种组成部分的一个或多个多核苷酸或多肽构建体。“引入”包括以使得序列（多肽或多核苷酸）能够进入细胞内部的方式将序列呈递到细胞。本文所提供的方法并不取决于用于将核酸或蛋白引入细胞中的特定方法，只要使核酸或蛋白能够进入至少一个细胞内部即可。用于将核酸和蛋白引入到各种细胞类型中的方法是本领域已知的，并且包括，例如，稳定转染方法、瞬时转染方法和病毒介导方法。

[0357] 在一些情况下，在所述方法和组合物中采用的细胞具有稳定地并入到其基因组中的DNA构建体。“稳定地并入”或“稳定地引入”包括将多核苷酸引入细胞中，使得核苷酸序列整合到细胞的基因组中且能够遗传给其子代。可使用任何方案稳定地并入DNA构建体或靶向基因组整合系统的各种组成部分。

[0358] 转染方案以及将多肽或多核苷酸序列引入到细胞中的方案可有所差别。转染方法包括基于化学的转染方法，其使用脂质体；纳米粒子；磷酸钙（Graham et al. (1973) *Virology* 52 (2) :456-67 (Graham等人, 1973年,《病毒学》,第52卷,第2期,第456-467页), Bacchetti et al. (1977) *Proc Natl Acad Sci USA* 74 (4) :1590-4 (Bacchetti等人, 1977年,《美国国家科学院院刊》,第74卷,第4期,第1590-1594页), 以及Kriegler, M (1991) .*Transfer and Expression: A Laboratory Manual*. New York: W.H. Freeman and Company. pp. 96-97 (Kriegler, M, 1991年,《基因转染和表达实验手册》,纽约W.H. 弗里曼公司,第96-97页)); 树状体; 或阳离子聚合物, 诸如DEAE-葡聚糖或聚乙烯亚胺。非化学方法包括电穿孔、超声穿孔和光学转染。基于颗粒的转染包括使用基因枪或磁体辅助转染 (Bertram (2006) *Current Pharmaceutical Biotechnology* 7, 277-28 (Bertram, 2006年,《当今药物生物技术》,第7卷,第277-28页))。也可将病毒方法用于转染。

[0359] 在一些情况下,通过电穿孔、胞浆内注射、病毒感染、腺病毒、慢病毒、逆转录病毒、转染、脂质介导转染或Nucleofection™进行介导来向细胞内引入核酸或蛋白。

[0360] 在一段时间内可执行向细胞引入核酸或蛋白一次或多次。例如,在一段时间内可执行引入步骤至少两次、至少三次、至少四次、至少五次、至少六次、至少七次、至少八次、至少九次、至少十次、至少十一次、至少十二次、至少十三次、至少十四次、至少十五次、至少十六次、至少十七次、至少十八次、至少十九次,或至少二十次。

[0361] 当核酸酶试剂和靶向载体(例如,LTVEC)两者都引入细胞中时,它们可同时引入。另选地,核酸酶试剂可与靶向载体分别引入。例如,可在引入靶向载体之前引入核酸酶试剂,或者可在引入靶向载体之后引入核酸酶试剂。当两个或更多个LTVEC引入细胞中时,它们可同时引入,或另选地,它们可分别引入。

[0362] IX. 细胞和动物

[0363] 本文所提供的各种组合物和方法采用细胞,诸如来自动物的细胞。此类细胞可为非人细胞,其可来自非人动物。此类细胞可为真核细胞,包括例如真菌细胞(例如,酵母)、植物细胞、动物细胞、哺乳动物细胞和人细胞。哺乳动物细胞可为例如非人哺乳动物细胞、人细胞、啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞。真核细胞可为全能细胞、多能细胞诸如非人多能细胞(例如,小鼠胚胎干(ES)细胞或大鼠ES细胞)或人多能细胞,或者非多能细胞。全能细胞包括可产生任何细胞类型的未分化细胞,并且多能细胞包括具有发育成不止一种分化细胞类型的能力的未分化细胞。此类多能和/或全能细胞可为例如胚胎干(ES)细胞或ES样细胞,诸如诱导多能干(iPS)细胞。胚胎干细胞包括胚胎起源的全能或多能细胞,其能够在引入胚胎中后促使发育中的胚胎成为任何组织。ES细胞可来源于胚泡的内细胞团,并且能够分化成三个脊椎动物胚层(内胚层、外胚层和中胚层)中任一胚层的细胞。此类细胞还可为造血干细胞或神经元干细胞。

[0364] 真核细胞还可为不是原代体细胞的细胞。体细胞可包括不是配子、胚细胞、配子体或未分化干细胞的任何细胞。

[0365] 真核细胞还包括原代细胞。原代细胞包括直接从生物体、器官或组织分离出来的细胞或细胞培养物。原代细胞包括既非转化细胞又非永生细胞的细胞。它们包括获自生物体、器官或组织的任何细胞,所述细胞先前未以组织培养物传递,或者先前已经以组织培养物传递但不能无限地以组织培养物传递。此类细胞可通过常规技术分离,并且包括例如,体细胞、造血细胞、内皮细胞、上皮细胞、成纤维细胞、间充质干细胞、角质细胞、黑素细胞、单核白细胞、单核细胞、脂肪细胞、前脂肪细胞、神经细胞、神经胶质细胞、肝细胞、骨骼肌成肌细胞和平滑肌细胞。例如,原代细胞可来源于结缔组织、肌肉组织、神经系统组织或上皮组织。

[0366] 真核细胞还包括无限增殖化细胞。无限增殖化细胞包括来自多细胞生物体的细胞,该多细胞生物体通常不会无限增殖,但是由于突变或改变已逃避了正常细胞衰老并且反而可继续进行分裂。此类突变或改变可自然发生或被故意诱导。无限增殖化细胞的例子包括中国仓鼠卵巢(CHO)细胞、人胚肾细胞(如,HEK 293细胞)和小鼠胚胎成纤维细胞(如,3T3细胞)。无限增殖化细胞的很多类型在本领域中是熟知的。

[0367] 无限增殖化细胞或原代细胞包括典型地用于培养或表达重组基因或蛋白的细胞。

[0368] 关于细胞、多能和/或全能细胞、ES细胞、供体细胞和/或宿主胚胎,术语“动物”包括哺乳动物、鱼和鸟。哺乳动物包括,例如,人、非人灵长类动物、猴子、猩猩、猫、狗、马、公牛、鹿、野牛、绵羊、啮齿动物(如,小鼠、大鼠、仓鼠、豚鼠)、牲畜(如,牛物种诸如奶牛、阉牛

等;羊物种诸如绵羊、山羊等;和猪物种诸如小猪和公猪)。鸟包括例如鸡、火鸡、鸵鸟、鹅、鸭等。还包括驯养的动物和农用动物。术语“非人动物”不包括人。

[0369] 小鼠多能和/或全能细胞可来自129品系、C57BL/6品系、129和C57BL/6的混合、BALB/c品系或Swiss Webster品系。129品系的例子包括129P1、129P2、129P3、129X1、129S1(例如129S1/SV、129S1/SvIm)、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 (Festing等人,1999年,《哺乳动物基因组》,第10卷,第836页)。C57BL品系的例子包括C57BL/A、C57BL/An、C57BL/GrFa、C57BL/Ka1_wN、C57BL/6、C57BL/6J、C57BL/6ByJ、C57BL/6NJ、C57BL/10、C57BL/10ScSn、C57BL/10Cr和C57BL/01a。小鼠多能和/或全能细胞还可来自上述129品系和上述C57BL/6品系的混合(例如,50%129和50%C57BL/6)。同样地,小鼠多能和/或全能细胞可来自上述129品系的混合上述BL/6品系的混合(例如,129S6(129/SvEvTac)品系)。小鼠ES细胞的特定例子是VGF1小鼠ES细胞。参见例如Auerbach et al. (2000) *Biotechniques* 29,1024-1028,1030,1032 (Auerbach等人,2000年,《生物技术》,第29卷,第1024-1028、1030、1032页),该文献全文以引用的方式并入本文。

[0370] 大鼠多能和/或全能细胞可来自任何大鼠品系,包括例如ACI大鼠品系、黑刺(DA)大鼠品系、威斯塔鼠品系、LEA大鼠品系、Sprague Dawley (SD) 大鼠品系,或者Fischer大鼠品系诸如Fisher F344或Fisher F6。大鼠多能和/或全能细胞还可获自来源于以上列举的两种或更多种品系的混合的品系。例如,大鼠多能和/或全能细胞可来自DA品系或ACI品系。ACI大鼠品系的特点是具有黑刺、白色的腹和足,为RT1^{av1}单倍型。此类品系得自多种来源,包括Harlan实验室。来自ACI大鼠的大鼠ES细胞系的例子为ACI.G1大鼠ES细胞。黑刺(DA)大鼠品系的特点是具有花纹外皮并且为RT1^{av1}单倍型。此类大鼠得自多种来源,包括查尔斯河(Charles River)和Harlan实验室。来自DA大鼠的大鼠ES细胞系的例子为DA.2B大鼠ES细胞系和DA.2C大鼠ES细胞系。在一些情况下,大鼠多能细胞和/或全能细胞来自近交系大鼠品系。参见例如2014年2月20日提交的U.S.2014/0235933 A1和2014年4月16日提交的U.S.2014/0310828 A1,这两份文献全文以引用方式并入本文。

[0371] 人多能细胞的例子包括人ES细胞、人成体干细胞、发育受限人祖细胞,以及人诱导多能干(iPS)细胞诸如始发态人iPS细胞和原始态人iPS细胞。参见例如2014年10月15日提交且全文以引用方式并入本文的美国专利申请号14/515,503。诱导多能干细胞包括可以从经分化的成体细胞直接衍生的多能干细胞。可以通过将特定组的重编程因子引入细胞来产生人iPS细胞,所述重编程因子可以包括例如Oct3/4、Sox家族转录因子(例如Sox1、Sox2、Sox3、Sox15)、Myc家族转录因子(例如,c-Myc、l-Myc、n-Myc)、Krüppel样家族(KLF)转录因子(例如KLF1、KLF2、KLF4、KLF5)和/或相关转录因子(诸如NANOG、LIN28和/或G1is1)。也可以例如通过使用miRNA、模拟转录因子的作用的小分子或谱系特异性分子来产生人iPS细胞。人iPS细胞的特征在于它们能够分化成三个脊椎动物胚层(例如内胚层、外胚层或中胚层)的任何细胞。人iPS细胞的特征还在于它们在合适的体外培养条件下无限繁殖的能力。参见例如Takahashi and Yamanaka (2006) *Cell* 126:663-676 (Takahashi和Yamanaka,2006年,《细胞》,第126卷,第663-676页)。始发态人ES细胞和始发态人iPS细胞包括表达与植入后的外胚层细胞相似的特征,并且参与谱系特化和分化的细胞。原始态人ES细胞和原始态人iPS细胞包括表达与植入前胚胎的内细胞团的ES细胞相似的特征,并且不参与谱系特化

的细胞。参见例如Nichols and Smith(2009) Cell Stem Cell 4:487-492 (Nichols和Smith, 2009年,《细胞干细胞》,第4卷,第487-492页)。

[0372] 已植入到宿主胚胎中的细胞可称为“供体细胞”。经遗传修饰的多能和/或全能细胞可来自与宿主胚胎相同的品系或来自不同品系。同样地,代孕母体可来自与经遗传修饰的多能和/或全能细胞和/或宿主胚胎相同的品系,或者代孕母体可来自与经遗传修饰的多能和/或全能细胞和/或宿主胚胎不同的品系。

[0373] 可在本文所公开的方法和组合物中使用多种宿主胚胎。例如,具有靶向遗传修饰的多能和/或全能细胞可从对应生物体引入到桑椹胚前期胚胎(例如8-细胞期胚胎)中。参见例如US 7,576,259、US 7,659,442、US7,294,754和US 2008/0078000 A1,所有这些文献全文以引用方式并入本文。在其他情况下,供体ES细胞可植入桑椹胚前期例如2细胞期、4细胞期、8细胞期、16细胞期、32细胞期或64细胞期的宿主胚胎中。宿主胚胎还可以是胚泡或者可以是胚泡前期胚胎、桑椹胚前期胚胎、桑椹胚期胚胎、不密实的桑椹胚期胚胎或密实的桑椹胚期胚胎。当采用小鼠胚胎时,宿主胚胎期可以是Theiler期(Theiler Stage) 1(TS1)、TS2、TS3、TS4、TS5和TS6,参照Theiler(1989)“The House Mouse:Atlas of Mouse Development,”Springer-Verlag,New York(Theiler,1989年,“家鼠:小鼠发育图集”,纽约施普林格出版社)中描述的Theiler期。例如,Theiler期可选自TS1、TS2、TS3和TS4。在一些情况下,宿主胚胎包含透明带,并且供体细胞是经由透明带中的孔引入宿主胚胎中的ES细胞。在其他情况下,宿主胚胎是透明带较少的胚胎。在另一些情况下,桑椹胚期宿主胚胎聚集。

[0374] X. 鉴定具有经修饰的靶基因组基因座的细胞的方法

[0375] 上述方法中的一些方法还包括鉴定具有经修饰的靶基因组基因座(例如,经修饰的基因组)的细胞。可使用各种方法来鉴定具有靶向修饰诸如缺失或插入的细胞。此类方法可包括鉴定在靶基因座处具有靶向修饰的一个细胞。可完成筛选以鉴定具有经修饰基因组基因座的此类细胞。

[0376] 筛选步骤可包括用于评估亲本染色体的等位基因(MOA)修饰的定量测定。例如,可经由定量PCR诸如实时PCR(qPCR)进行定量测定。实时PCR可利用识别靶基因座的第一引物组和识别非靶向参考基因座的第二引物组。引物组可包含识别扩增序列的荧光探针。

[0377] 在其他情况下,使用包括例如Southern印迹分析、DNA测序、PCR分析或表型分析的方法来选择具有靶向遗传修饰的细胞。然后将此类细胞用于本文所述的各种方法和组合物中。

[0378] 合适的定量测定法的其他例子包括荧光介导原位杂交(FISH)、比较基因组杂交、等温DNA扩增、定量固定探针杂交、Invader Probes[®]、MMPassays[®]、TAQMAN[®]分子信标、或Eclipse[™]探针技术(参见例如US2005/0144655,该文献全文以引用方式并入本文)。

[0379] 筛选步骤一般还包括臂特异性测定,其是用于将核酸插入物向靶基因组基因座中的正确靶向插入与核酸插入物向靶基因组基因座以外的基因组位置中的随机转基因插入区分开,并且还用于检测两个或更多个重叠LTVEC向单个构建体中的正确组装。用于筛选靶向修饰的常规测定诸如长片段PCR或Southern印迹将插入的靶向载体连接至靶向基因座。然而,LTVEC由于其较大的同源臂大小,不允许通过此类常规测定进行筛选。为了筛选LTVEC靶向,可使用包括等位基因丢失(LOA)和等位基因获得(GOA)测定的等位基因修饰(MOA)测

定(参见例如US 2014/0178879和Frendewey et al. (2010) *Methods Enzymol.* 476:295-307 (Frendewey等人, 2010年,《酶学方法》,第476卷,第295-307页),所述文献出于所有目的全文以引用方式并入本文)。等位基因丢失(LOA)测定颠倒常规筛选逻辑,并且定量突变所定向到的天然基因座的拷贝数。在正确靶向的细胞克隆中,LOA测定检测两个天然等位基因中的一个(针对不在X或Y染色体上的基因),另一个等位基因被靶向修饰破坏。可将相同原理逆向应用为等位基因获得(GOA)测定,定量被插入的靶向载体的拷贝数。例如,GOA测定和LOA测定的组合使用将正确的靶向杂合克隆显示为已丢失天然靶基因的一个拷贝并且获得药物抗性基因或其他插入标记的一个拷贝。

[0380] 例如,定量聚合酶链反应(qPCR)可用作等位基因定量的方法,但是能够可靠区分靶基因的零个、一个和两个拷贝之间的差异或者核酸插入物的零个、一个和两个拷贝之间的差异的任何方法可用于发展MOA测定。例如,尤其是与参考基因(参见例如US 6,596,541,该专利出于所有目的全文以引用方式并入本文)相比,TAQMAN[®]可用于定量基因组DNA样品中DNA模板的拷贝数。在与一个或多个靶基因或一个或多个基因座相同的基因组DNA中定量参考基因。因此,执行两个TAQMAN[®]扩增(每个使用其相应的探针)。一个TAQMAN[®]探针确定参考基因的“Ct”(阈值循环),而其他探针确定一个或多个靶基因或一个或多个基因座的通过成功靶向(即LOA测定)置换的区域的Ct。Ct是反映针对TAQMAN[®]探针中每一个的最初DNA的数量的量,即较少丰度序列需要更多PCR循环以达到阈值循环。针对TAQMAN[®]反应减少一半模板序列拷贝数将会增加约一个Ct单元。当与来自非靶向细胞的DNA相比时,一个或多个靶基因或一个或多个基因座中的一个等位基因已通过同源重组置换的细胞中的TAQMAN[®]反应将会针对靶TAQMAN[®]反应增加一个Ct,而不会针对参考基因增加Ct。就GOA测定而言,可使用另一个TAQMAN[®]探针来确定核酸插入物的Ct,该核酸插入物通过成功靶向置换一个或多个靶基因或一个或多个基因座。

[0381] 可能有用的是增强标准LOA和GOA测定,以验证LTVEC进行正确靶向。例如,只进行LOA测定和GOA测定可能无法正确区分靶向的细胞克隆与其中靶基因组基因座的Cas诱导的缺失和基因组中其他地方的LTVEC随机整合同时发生的克隆。因为靶向细胞中的选择压力是基于选择盒的,所以基因组中其他地方的LTVEC随机转基因整合一般可包括选择盒和LTVEC的相邻区域,但不包括LTVEC的更远侧区域。例如,如果LTVEC的一部分被随机整合进基因组,并且LTVEC包含长度约5kb或更长、具有邻近3'同源臂的选择盒的核酸插入物,那么在一些情况下,3'同源臂而非5'同源臂将与选择盒进行转基因整合。另选地,如果选择盒邻近5'同源臂,那么在一些情况下,5'同源臂而非3'同源臂将与选择盒进行转基因整合。例如,如果LOA和GOA测定用于评估LTVEC的靶向整合,并且GOA测定利用针对选择盒或LTVEC的任何其他独特(非臂)区域的探针,那么与LTVEC随机转基因整合组合的靶基因组基因座处的杂合缺失将给出与在靶基因组基因座处LTVEC的杂合靶向整合相同的读数。为了验证通过LTVEC进行正确靶向,臂特异性测定可与LOA和/或GOA测定结合使用。

[0382] 臂特异性测定确定LTVEC同源臂中DNA模板的拷贝数。此类同源臂可包括不与另一个LTVEC重叠但与细胞中的靶序列相对应的LTVEC的同源臂(例如,与小鼠细胞中的基因组靶序列重叠的同源臂(mArm))。此类同源臂还可包括存在于两个重叠LTVEC中的重叠同源臂

(例如,第一LTVEC的3'同源臂和第二LTVEC的5'同源臂中的重叠人序列(hArm))。对于将多个重叠LTVEC引入细胞中的实验而言,筛选通常包括适用于所有独特插入的序列的LOA测定、GOA测定以及适用于所有同源区域(即,LTVEC与细胞中的靶序列之间,以及两个不同重叠LTVEC之间)的臂特异性测定。例如,在向小鼠细胞中引入三个重叠LTVEC而使野生型小鼠靶基因座人源化的情况下,杂合靶向插入的预期拷贝数将为如下:2个拷贝的5' mArm(与5'小鼠靶序列重叠的同源臂)、1个拷贝的hArm1(LTVEC 1与2之间的重叠序列)、1个拷贝的hArm2(LTVEC 2与3之间的重叠序列)以及2个拷贝的3' mArm(与3'小鼠靶序列重叠的同源臂)。在上述例子中,大于两个的mArm拷贝数通常将指示随机在靶基因组基因座之外而非在靶基因组基因座处的转基因LTVEC整合,这是不可取的。正确靶向的克隆将保持mArm拷贝数为两个。另外,在此类臂特异性测定中少于两个的mArm拷贝数通常将表明Cas介导的大段缺失延伸到靶向缺失的区域之外,这也是不可取的。同样,对于杂合靶向修饰而言,hArm1和hArm2的拷贝数都为1通常将指示所有三个LTVEC已组装到单个构建体中。

[0383] 上文或下文引用的所有专利申请、网站、其他出版物、登录号等等出于所有目的全文以引用方式并入,其程度如同每个单独项被具体且单独地指明以引用的方式这样并入。如果序列的不同版本与不同时间的登录号相关,则与本申请的有效提交日时的登录号相关联的版本是有意义的。有效提交日是指关于登录号的优先权申请(如果适用)的实际在先提交日或提交日。同样地,如果不同版本的出版物、网站等在不同时间被公布,则除非另有说明,否则最近在本申请的有效申请日公布的版本是有意义的。本发明的任何特征、步骤、元件、实施方案或方面可以与任何其他项组合使用,除非另有特别说明。尽管出于清楚理解的目的已通过说明和举例的方式对本发明进行了一些详细描述,但显而易见的是,在所附权利要求书的范围内还可作出某些变化和修改。

[0384] 实施例

[0385] 给出以下实施例是为了给本领域的普通技术人员提供如何实施并使用本发明的完整公开和描述,而非旨在限制本发明人视作其发明的范围,也不旨在表示下文的实验是所进行过的所有或仅有的实验。已尽量确保所使用的数字(例如量、温度等)的准确性,但应考虑到一些实验误差和偏差。除非另外指明,否则份数是重量份数,分子量是重均分子量,温度以摄氏度计,并且压力是大气压或接近大气压。

[0386] 实施例1:两个LTVEC与锌指核酸酶相结合对TCR α 基因座进行靶向

[0387] 双重靶向系统被设计成使得两个大靶向载体(LTVEC)在单个靶向步骤中修饰基因组基因座。如图1中所示,具有小鼠染色体14上的TCR α 基因座的杂合修饰(含潮霉素选择盒)的细胞通过双重靶向系统进行靶向,以产生包含附加Ig κ 可变基因区段的ES细胞。

[0388] 该双重靶向方法总结于图1中,其涉及两个不同大靶向载体(LTVEC)与编码核酸内切酶(例如,锌指核酸酶)的核苷酸序列一起双重靶向或共同电穿孔到ES细胞中,该核酸内切酶在靶基因座处或附近产生双链断裂。

[0389] 在该方法中,第一大靶向载体(标记为MAID 1710)包含具有人V κ 1-5和V κ 1-6基因区段的序列的3' 30kb同源臂、具有人V κ 3-7至V κ 3-15基因区段的120kb序列、以及具有人V κ 1-16基因区段的5' 20kb区域("重叠区")。第二大靶向载体(标记为MAID 6600)包含3' 20kb重叠区(与第一载体中一样,具有人V κ 1-16基因区段的区域)、具有人V κ 1-17至V κ 2-30基因区段的140kb序列、FRT-Ub-Neo-FRT选择盒以及15.5kb 3'小鼠TCR A同源臂。

[0390] 锌指核酸酶 (ZFN) 被设计成识别并切割潮霉素抗性基因内的靶序列,以便促进这两个LTVEC在靶TCR A基因座处的同源重组。图1中产生的ES细胞 (MAID 6548,对于所有人J κ 区段和四个功能性人V κ 基因区段为杂合的) 使用上述两个大靶向载体 (MAID6600和经修剪的MAID1700) 和表达各一半ZFN (1/2) 的两个质粒进行电穿孔,该ZFN结合于潮霉素抗性基因中的识别序列并催化靶位点 (TGCGATCGCTGCGGCCGAtccttagCCAGACGAGCGGGTTCGG (SEQ ID NO:2) 处的双链断裂;其中切割位点以小写字母表示) (参见表1)。两个附加ZFN被设计成靶向潮霉素:在核苷酸序列CGCTGCGGCCGATCTtagccaGACGAGCGGGTTCGG (SEQ ID NO:3) 处靶向潮霉素基因的ZFN (3/4);以及在核苷酸序列AGCGTGTCCGACCTGATGcagctcTCGGAGGGCGAAGAA (SEQ ID NO:4) 处靶向潮霉素基因的ZFN (5/6) (参见表1)。

[0391] 表1:潮霉素锌指核酸酶结合和切割位点(互补链未示出)

[0392]

锌指核酸酶	锌指半部	结合序列(5'-3')	切割位点 (5'-3')
Hyg-ZFN(1,2)	Hyg-ZF1	TGCGATCGCTGCGGCCGA (SEQ ID NO: 5)	TCTTAG
	Hyg-ZF2	CCGAACCCGCTCGTCTGG (SEQ ID NO: 6)	(SEQ ID NO: 11)

[0393]

Hyg-ZFN(3,4)	Hyg-ZF3	CGCTGCGGCCGATCT (SEQ ID NO: 7)	TAGCCA
	Hyg-ZF4	CCGAACCCGCTCGTC (SEQ ID NO: 8)	(SEQ ID NO: 12)
Hyg-ZFN(5,6)	Hyg-ZF5	AGCGTGTCCGACCTGATG (SEQ ID NO: 9)	CAGCTC
	Hyg-ZF6	TTCTTCGCCCTCCGA (SEQ ID NO: 10)	(SEQ ID NO: 13)

[0394] 通过同源重组将这两个大靶向载体插入DNA序列中,从而替换包含和围绕Hyg选择盒的区域。所得的ES细胞在内源TCR A基因座处包含具有人J κ 1至J κ 5和V κ 4-1至V κ 2-30基因区段的人免疫球蛋白可变结构域。使用上述TAQMAN[®]测定法确认这两个大靶向载体的成功并入 (Lie and Petropoulos,1998.Curr.Opin.Biotechnology 9:43-48 (Lie和Petropoulos,1998年,《生物技术当前述评》,第9卷,第43-48页),该文献以引用方式并入本文),该测定法使用图1中指示和下表2中列出的探针和引物(GOA=等位基因获得;LOA=等位基因丢失;拷贝数=检查序列的拷贝数以跟踪转基因整合与靶向整合;hArm1=第一大靶向载体 (MAID 1710) 的30kb 3'同源臂;hArm2=第一大靶向载体 (MAID 1710) 与第二大靶向载体 (MAID 6600) 的20kb重叠区,mArm=第二靶向载体 (MAID 6600) 的15.5kb 5'同源臂,野生型小鼠对照-小鼠TCR A基因座处存在的序列)。使用识别LTVEC的同源臂中的序列的实时PCR测定(称为臂特异性测定)来验证LTVEC向小鼠基因组中的正确靶向。确定这些臂特异性测定的拷贝数提供了进一步的说明,从而有助于将保留例如两个mArm拷贝数的正确靶向的ES克隆与这样的克隆区分开,其中Cas9诱导的靶小鼠基因座缺失恰逢LTVEC在基因座其他地方的随机整合,在这种情况下,mArm拷贝数将为三个(或更多个)。

[0395] 表2:TAQMAN引物和探针

[0396]

基因	测定法	正向引物	探针	反向引物
HYG	LOA	TGCGGCCGATCT	ACGAGCGGGTTC	TTGACCGATTCCT
		TAGCC (SEQ ID NO: 14)	GGCCCATTC (SEQ ID NO: 15)	TGCGG (SEQ ID NO: 16)

[0397]

HYG-U	LOA	CGACGTCTGTCG AGAAGTTTCTG (SEQ ID NO: 17)	AGTTCGACAGCG TGTCCGACCTGA (SEQ ID NO: 18)	CACGCCCTCCTA CATCGAA (SEQ ID NO: 19)
Hyg-D	LOA	TGTCGGGGCGTAC ACAAATCG (SEQ ID NO: 20)	CCGTCTGGACCG ATGGCTGTGT (SEQ ID NO: 21)	GGGCGTCGGTTT CCACTATC (SEQ ID NO: 22)
PGKp1 (Hyg 启动子)	LOA	CAAATGGAAGTA GCACGTCTCACT (SEQ ID NO: 23)	CTCGTGCAGATG GACAGCACCGC (SEQ ID NO: 24)	CCGCTGCCCCAA AGG (SEQ ID NO 25)
hlgK6	hArm1 的 拷贝数	GTCAAGCACTGC TGGCACAC (SEQ ID NO: 26)	AACCCTTGTGCT ATTGAATTGCTA TGCTGTCAG (SEQ ID NO: 27)	TGTTGTAGACCC TCCGCCAC (SEQ ID NO: 28)
hlgK12 (MAID 1710 插入物)	GOA	TTGCCTTTCTCAC ACCTGCAG (SEQ ID NO: 29)	CAGCCCATCCTG TCACTTCGCTGG A (SEQ ID NO: 30)	TGGCCCAACAGT ACAGCTCAG (SEQ ID NO :31)
hlgK13	hArm2 的 拷贝数	TCAGTCAATCAC CTTTCCCAGC (SEQ ID NO: 32)	TCCCCAGGTAGC CTCATGAACCAA TGTT (SEQ ID NO: 33)	CACATTACTGAG TCCCCACAGGG (SEQ ID NO: 34)
hlgK14	hArm2 的 拷贝数	CATTGTCAAAGA AGCACTGGAAAT G (SEQ ID NO: 35)	ACCATTGCAGTT TACCCACGGTTA GGATTTTT (SEQ ID NO: 36)	TCTTGCAATGGG ATCATCAGATG (SEQ ID NO: 37)
Neo	GOA	GGTGGAGAGGCT ATTCGGC (SEQ ID NO: 38)	TGGGCACAACAG ACAATCGGCTG (SEQ ID NO: 39)	GAACACGGCGGC ATCAG (SEQ ID NO: 40)

[0398]

hlgK15	GOA	CAGGTGCAAAGG TGACCACAG (SEQ ID NO: 41)	TGGGTCCTGCCC ATCCATGCA (SEQ ID NO: 42)	GGCAGCCTGAGT GTCAGAGC (SEQ ID NO: 43)
hlgK25	GOA	G TTCAGGCCCCA CAGACTCTC (SEQ ID NO: 44)	TCCTCTCTGGAG CAACCATGAAGT TCCCT (SEQ ID NO: 45)	CCTGAAGCCATG AGGGCAG (SEQ ID NO: 46)
hUbC-D (Neo 启动子)	GOA	AGGGTAGGCTCT CCTGAATCG (SEQ ID NO: 47)	ACAGGCGCCGGA CCTCTGGT (SEQ ID NO: 48)	CCAAAGAAACTG ACGCCTCAC (SEQ ID NO: 49)
TCRA Arm4	mArm 的 拷贝数	GCGCCACATGAA TTTGACCAG (SEQ ID NO: 50)	TGTACCCAATCTT CCAAAGAAAGAG CTG (SEQ ID NO: 51)	GGCATCCTGTCC TCCCTC (SEQ ID NO: 52)
亲本 1540m1	野生型小 鼠对照	CAGTAAGGGAAG AGACTACAACAG CAT (SEQ ID NO: 53)	TGCACACTGCTC ACCACTGCAAGC TAT (SEQ ID NO: 54)	TGCTGGTGGCCC CATCT (SEQ ID NO: 55)
亲本 1540m3	野生型小 鼠对照	GAACTCAGCTAT GATAGTGTGCGAA TGTA (SEQ ID NO: 56)	CAGCCCAGCAGC TGTGGGTCTC (SEQ ID NO: 57)	GCTCAGGGAGAA CACAGAACTTAG A (SEQ ID NO: 58)
hlgK5	MAID 6548 序 列	CCCCGTCCTCCTC CTTTTC (SEQ ID NO: 59)	TCATGTCCATTA ACCATTTACCTT TTGCCCA (SEQ ID NO: 60)	TGCAAGTGCTGC CAGCAAG (SEQ ID NO: 61)

[0399] ES细胞中所得的所靶向基因座具有下列接合序列,其中小鼠序列位于括号中,人序列采用正常字体,多克隆位点采用粗体,并且Frt序列采用斜体(表3)。

[0400] 表3:由双重ES细胞靶向得到的基因座的接合序列

[0401]

接合区	Seq ID No	序列
小鼠 Tcra/5' Frt	62	(GTCTTTTTTGTTCCTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTAAACTCTATGGAG TAGTCAGAACACACTCTTCA)GAAGGGACTC CTGATTTCAAAGGGTACCGAAGTTCCTATT CCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC
3' Frt/人 IgK	63	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG AAAGTATAGGAACTTCCTAGGGTTTCACCG GTGGCGCGCCTAACAGAGAGGAAAGTCAAA TTATAAAGAATATGAGATTCAGAATTCTGA TTAACTGTGG
人 IgK/小鼠 Tcra	64	GATAAATTATTTTGTTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC GCCACCGCGGTGGAGCTC(AGGTTTCCG GTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

[0402] 对分离的ES细胞集落进行等位遗传修饰 (MOA) 筛选,最终在所筛选的960个集落之中鉴定出27个正确靶向的克隆,靶向效率达2.81%。

[0403] 用于产生包含附加免疫球蛋白可变区基因区段的TCR A基因座的替代策略涉及使用相继的大靶向载体来连续靶向(参见例如图2)。因此,对于所有人J κ 基因区段和四个功能性人V κ 基因区段(MAID 6548)为杂合的ES细胞使用大靶向载体进行电穿孔,该大靶向载体从5'到3'包含:15.5kb 5'小鼠同源臂、Frt-Ub-Neo-Frt选择盒、具有V κ 3-7至V κ 3-15基因区段的120kb片段,以及具有V κ 1-5和V κ 1-6基因区段的30kb 3'人同源臂(也存在于MAID 6548序列中)。使用上述TAQMAN[®]测定法确认成功并入,该测定法使用上表2中列出和图2中指示的引物和探针:Hyg、hIgK5、hIgK6、hIgK12、Neo、亲本1540m3、亲本1540m1。具体地讲,TCRA Arm4和hIgK6探针用作臂特异性探针,以验证LTVEC的正确基因组靶向。还可使用附加组的引物和探针hIgK10,以确认成功并入:正向引物-CGATTATGACTGGTTAGGTAGAAAGGTG (SEQ ID NO:65);探针-GCCACTGGTTTCTCCAAATGTTTCAATCCAT (SEQ ID NO:66);反向引物-GGGAGTACTTGGAGATCCCTAAGC (SEQ ID NO:67)。

[0404] ES细胞中所得的所靶向基因座具有下列接合序列,其中小鼠序列位于括号中,人序列采用正常字体,多克隆位点采用粗体,并且Frt序列采用斜体(表4)。

[0405] 表4:由单ES细胞靶向得到的基因座的接合序列

[0406]

接合区	Seq ID No	序列
小鼠 Tcra/5' Frt	68	(TTGAGCTTCATCAAAGTCACATGGGTAA CTCTATGGAGTAGTCAGAACACACTCTCA) GAAGGGACTCCTGATTTCAAAGGGTACC GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAA AGTATAGGAACTTC
3' Frt/人 IgK	69	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG AAAGTATAGGAACTTCCTAGGGTTTCACCG GTGGCGCGCCAGGACCCAGGCTCTGACACT CAGGCTGCCAATACAATTGCCATGAAGACA GATGTTGATG

[0407]

人 IgK/小鼠 Tcra	64	GATAAATTATTTTGTGACACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC GCCACCGCGGTGGAGCTC(AGGTTTCCG GTACTIONAACACAGAGCACAGATTTAGT GGTGAGGGACTCT)
---------------	----	---

[0408] 对分离的集落进行MOA筛选,最终在所筛选的440个集落之中鉴定出5个正确靶向的克隆(单独的LTVEC),靶向效率达1.1%。表9中示出了对使用LTVEC+ZFN或LTVEC+CRISPR-Cas9靶向的分离的集落进行筛选的结果。

[0409] 在完成图2中所示的单靶向后,可使用大靶向载体相继靶向ES细胞,所述大靶向载体包含附加V_κ以便加起来达到功能性人免疫球蛋白V_κ基因区段的整个组库。

[0410] 在另外的其他替代策略中,可使用双重(两个大靶向载体)或单(一个大靶向载体)靶向方案完成相继附加人Ig V_κ基因区段的双重或单靶向,所述靶向方案涉及锌指核酸酶或CRISPR介导的一个或多个选择(例如,潮霉素)盒破坏。

[0411] 上述所靶向的ES细胞用作供体ES细胞,并且通过VELOCIMOUSE[®]方法引入桑椹胚前期胚胎(例如,8细胞期小鼠胚胎)中(参见例如US7,576,259、US 7,659,442、US 7,294,754和US 2008-0078000 A1)。将包含供体ES细胞的小鼠胚胎孵育直到胚泡期为止,然后植入代孕母体中,以产生完全来源于供体ES细胞的F0小鼠。使用检测独特基因序列的存在的等位遗传修饰测定进行基因分型,从而鉴定完全来源于供体ES细胞的独立地携带嵌合人IgK V-小鼠Tcra C基因的F0小鼠。

[0412] 实施例2:两个LTVEC与CRISPR/Cas系统相结合对潮霉素基因进行靶向

[0413] 还使用CRISPR/Cas9系统进行实施例1中所述利用锌指核酸酶的双重靶向方法。

[0414] 各种向导RNA (gRNA) 被设计成识别潮霉素抗性基因内的各种靶序列 (CRISPR识别序列)。潮霉素基因内的CRISPR识别序列如下:gRNA#1:ACGAGCGGGTTCGGCCCATTCGG (SEQ ID NO:70);gRNA#6:CTTAGCCAGACGAGCGGGTTCGG (SEQ ID NO:71);gRNA#10:GCCGATCTTAGCCAGACGAGCGG (SEQ ID NO:72);以及gRNA#16:CGACCTGATGCAGCTCTCGGAGG (SEQ ID NO:73)。图3中示出了潮霉素基因内的识别序列的位置,该图示出了靶向载体MAID 1545中潮霉素的CRISPR/Cas介导的破坏。对gRNA#1、gRNA#6、gRNA#10和gRNA#16进行筛选,并确认它们能特异性地靶向潮霉素基因(参见图3)。表5中提供了使用各种潮霉素特异性gRNA初步筛选得出的结果。

[0415] 表5:使用潮霉素特异性gRNA得出的初步筛选结果。

[0416]

gRNA/ZFN	初步候选物总数 (2个平板/电穿孔)	再次确认呈阳性的候选物/ 再次确认的总数
gRNA1	5	2/2
gRNA6	6	1/1
gRNA10	19	5/5
gRNA16	91	8/8
ZFN 1/2	10	4/4

[0417] ES细胞,例如图1中产生的ES细胞(MAID 6548,对于所有人J κ 区段和四个功能性人V κ 基因区段为杂合的)使用两个大靶向载体(实施例1中所述)与编码Cas9和gRNA(例如,gRNA#1、gRNA#6、gRNA#10或gRNA#16)的单个载体或多个载体一起进行电穿孔,所述Cas9和gRNA能识别并切割潮霉素抗性基因内的靶位点。

[0418] 通过同源重组将两个大靶向载体插入DNA序列中,从而替换包含和围绕Hyg选择盒的区域。使用TAQMAN[®]测定确认这两个大靶向载体的成功并入。

[0419] 上述所靶向的ES细胞将用作供体ES细胞,并且通过VELOCIMOUSE[®]方法引入桑椹胚前期胚胎(例如,8细胞期小鼠胚胎)中(参见例如US7,576,259、US 7,659,442、US 7,294,754和US 2008-0078000 A1)。将包含经遗传修饰的ES细胞的小鼠胚胎孵育直到胚泡期为止,然后植入代孕母体中,以产生完全来源于供体ES细胞的F0小鼠。使用检测独特基因序列的存在的等位遗传修饰测定进行基因分型,从而鉴定完全来源于供体ES细胞的F0小鼠。

[0420] 实施例3:三个LTVEC与锌指核酸酶相结合对TCR α 基因座进行靶向

[0421] 三重靶向系统被设计成使得三个大靶向载体(LTVEC)在单个靶向步骤中修饰基因组基因座。如图4中所示,具有小鼠染色体14上的TCR α 基因座的杂合修饰(含潮霉素选择盒)的细胞通过三重靶向系统进行靶向,以产生包含附加Ig κ 可变基因区段的ES细胞。

[0422] 该三重靶向方法总结于图4中,其涉及三个不同大靶向载体(LTVEC)(MAID 6647、MAID 6600和MAID 1710)与编码核酸内切酶(例如,锌指核酸酶或Cas9和gRNA)的核苷酸序列一起三重靶向或共同电穿孔到ES细胞中,该核酸内切酶在靶基因座处或附近产生双链断

裂。

[0423] 该方法中,第一大靶向载体(标记为MAID 1710)包含具有人V κ 1-5和V κ 1-6基因区段的序列的3' 30kb同源臂、具有人V κ 3-7至V κ 3-15基因区段的120kb序列、以及具有人V κ 1-16基因区段的5' 20kb区域(“重叠区”)。第二大靶向载体(标记为MAID 6600)包含3' 20kb重叠区(与第一载体中一样,具有人V κ 1-16基因区段的区域)、具有人V κ 1-17至V κ 2-24基因区段的140kb序列、以及具有人V κ 3-25至V κ 2-30的5' 60kb区域(“重叠区”)。第三大靶向载体(标记为MAID 6647)包含3' 60kb重叠区(与第二载体中一样,具有人V κ 3-25至V κ 2-30的区域)、具有人V κ 3-31至V κ 2-40的90kb序列、FRT-Ub-Neo-FRT选择盒以及15.5kb 5' 小鼠TCR A同源臂。

[0424] 锌指核酸酶(ZFN)被设计成识别并切割潮霉素抗性基因内的靶序列,以便促进三个LTVEC在靶TCR A基因座处的同源重组。图4中产生的ES细胞(MAID 6548,对于所有人J κ 区段和四个功能性人V κ 基因区段为杂合的)使用上述三个大靶向载体(经修剪的MAID6600、经修剪的MAID1700和MAID6647)和表达各一半ZFN(1/2)的两个质粒进行电穿孔,该ZFN结合于潮霉素抗性基因中的识别序列并催化靶位点(TGCGATCGCTGCGGCCGAtcttagCCAGACGAGCGGGTTCGG (SEQ ID NO:2)处的双链断裂;其中切割位点以小写字母表示)(参见表1)。

[0425] 通过同源重组将这三个大靶向载体插入DNA序列中,从而替换包含和围绕Hyg选择盒的区域。所得的ES细胞在内源TCR A基因座处包含具有人J κ 1至J κ 5和V κ 4-1至V κ 2-40基因区段的人免疫球蛋白可变结构域。使用上述TAQMAN[®]测定法确认这三个大靶向载体的成功并入(Lie and Petropoulos,1998.Curr.Opin.Biotechnology 9:43-48(Lie和Petropoulos,1998年,《生物技术当前述评》,第9卷,第43-48页),该文献以引用方式并入本文),该测定法使用图4中指示及上表2和下表6中列出的探针和引物(GOA=等位基因获得;LOA=等位基因丢失;拷贝数=检查序列的拷贝数以跟踪转基因整合与靶向整合;hArm1=第一大靶向载体(MAID 1710)的30kb 3' 同源臂;hArm2=第一大靶向载体(MAID 1710)与第二大靶向载体(MAID 6600)的20kb重叠区,hArm3=第二靶向载体(MAID 6600)与第三靶向载体(MAID6647)的60kb重叠区,mArm=第三靶向载体(MAID 6647)的15.5kb 5' 同源臂,野生型小鼠对照-小鼠TCR A基因座处存在的序列)。使用识别LTVEC的同源臂中的序列的实时PCR测定(称为臂特异性测定)来验证LTVEC向小鼠基因组中的正确靶向。确定这些臂特异性测定的拷贝数提供了进一步说明,从而有助于将正确靶向的ES克隆(对于小鼠探针(mArm)而言保留两个拷贝数,并且对于人探针(hArm1)而言保留一个拷贝数)与这样的克隆区分开,其中Cas9诱导的靶小鼠基因座缺失恰逢LTVEC在基因组其他地方的随机整合,在这种情况下,对于小鼠探针(mArm)而言将存在三个(或更多个)拷贝数,并且对于人探针(hArm1)而言将存在两个(或更多个)拷贝数。为了检测这三个LTVEC通过同源重组向所需基因座中的正确组装,我们利用了臂特异性TAQMAN[®]测定法。hArm2和hArm3的1个预期拷贝数指示所有三个LTVEC已组装到单个构建体中。

[0426] 表6:TAQMAN引物和探针

[0427]

基因	测定法	正向引物	探针	反向引物
HYG	LOA	TGCGGCCGATCTT AGCC (SEQ ID NO: 14)	ACGAGCGGGTTCGG CCCATTC (SEQ ID NO: 15)	TTGACCGATTC CTTGCGG (SEQ ID NO: 16)
HYG-U	LOA	CGACGTCTGTCGA GAAGTTTCTG (SEQ ID NO: 17)	AGTTCGACAGCGTGT CCGACCTGA (SEQ ID NO: 18)	CACGCCCTCCT ACATCGAA (SEQ ID NO: 19)
Hyg-D	LOA	TGTCGGGCGTACA CAAATCG (SEQ ID NO: 20)	CCGTCTGGACCGATG GCTGTGT (SEQ ID NO: 21)	GGGCGTCGGTT TCCACTATC (SEQ ID NO: 22)
hIgK6	hArm1 的 拷贝数	GTCAAGCACTGCT GGCACAC (SEQ ID NO: 26)	AACCCTTGTGCTATT GAATTGCTATGCTGT CAG (SEQ ID NO: 27)	TGTTGTAGACC CTCCGCCAC (SEQ ID NO: 28)
hIgK12 (MAID 1710 插入 物)	GOA	TTGCCTTTCTCAC ACCTGCAG (SEQ ID NO: 29)	CAGCCCATCCTGTCA CTTCGCTGGA (SEQ ID NO:30)	TGGCCCAACAG TACAGCTCAG (SEQ ID NO: 31)
hIgK13	hArm2 的 拷贝数	TCAGTCAATCACC TTTCCCAGC (SEQ ID NO: 32)	TCCCCAGGTAGCCTC ATGAACCAATGTT (SEQ ID NO: 33)	CACATTACTGA GTCCCCACAGG G (SEQ ID NO: 34)
hIgK14	hArm2 的 拷贝数	CATTGTCAAAGAA GCACTGGAAATG (SEQ ID NO: 35)	ACCATTGCAGTTTAC CCACGGTTAGGATT TT (SEQ ID NO: 36)	TCTTGCAATGG GATCATCAGAT G (SEQ ID NO: 37)
Neo	GOA	GGTGGAGAGGCT ATTCGGC (SEQ ID NO: 38)	TGGGCACAACAGAC AATCGGCTG (SEQ ID NO: 39)	GAACACGGCGG CATCAG (SEQ ID NO: 40)

[0428]

hIgK15	GOA	CAGGTGCAAAGG TGACCACAG (SEQ ID NO: 41)	TGGGTCCTGCCCATC CATGCA (SEQ ID NO: 42)	GGCAGCCTGAG TGTCAGAGC (SEQ ID NO: 43)
hIgK25	hArm3 的 拷贝数	G TTCAGGCCCCAC AGACTCTC (SEQ ID NO: 44)	TCCTCTCTGGAGCAA CCATGAAGTTCCT (SEQ ID NO: 45)	CCTGAAGCCAT GAGGGCAG (SEQ ID NO: 46)
TCRA Arm4	mArm 的拷 贝数	GCGCCACATGAAT TTGACCAG (SEQ ID NO: 50)	TGTACCCAATCTTCC AAAGAAAGAGCTG (SEQ ID NO: 51)	GGCATCCTGTC CTCCCTTC (SEQ ID NO: 52)
亲本 1540m1	野生型小 鼠对照	CAGTAAGGGAAG AGACTACAACAG CAT (SEQ ID NO: 53)	TGCACACTGCTCACC ACTGCAAGCTAT (SEQ ID NO: 54)	TGCTGGTGGCC CCATCT (SEQ ID NO: 55)
亲本 1540m3	野生型小 鼠对照	GAACTCAGCTATG ATAGTGTCGAATG TA (SEQ ID NO: 56)	CAGCCCAGCAGCTGT GGGTTCTC (SEQ ID NO: 57)	GCTCAGGGAGA ACACAGAACTT AGA (SEQ ID NO: 58)
hIgK5	MAID 6548 序列 (参见 表 2)	CCCCGTCCTCCTC CTTTTTC (SEQ ID NO: 59)	TCATGTCCATTAACC CATTTACCTTTTGCC CA (SEQ ID NO: 60)	TGCAAGTGCTG CCAGCAAG (SEQ ID NO: 61)
hIgK22	hArm3 的 拷贝数	TGGCTCCAAGAAC AGTTTGCC (SEQ ID NO: 74)	CCCTGACTTTGCTGC TCAACTCACAGCC (SEQ ID NO: 75)	GGTCCAGTGGA ATCTGCCATG (SEQ ID NO: 76)
hIgK21	GOA	CATTTGGCTACAT ATCAAAGCCG (SEQ ID NO: 77)	CCTGAGCCAGGGAA CAGCCCCTGATA (SEQ ID NO: 78)	ACATGGCTGAG GCAGACACC (SEQ ID NO: 79)

[0429]

hIgK26	GOA	TGGGCCGTTATGC TAGTACCA (SEQ ID NO: 80)	TGGCTTTACCCCTTT TGAAGGGCCC (SEQ ID NO: 81)	CACAGCTGAAG CAGGATGAGC (SEQ ID NO: 82)
hIgK30	GOA	TCTCTGAGCAGCC ATCCCC (SEQ ID NO: 83)	TTCTCCTTTGGTGTA GAGGGCACCAGC (SEQ ID NO: 84)	ACCAGGCATGG CAGAAAGG (SEQ ID NO: 85)

[0430] ES细胞中所得的所靶向基因座具有下列接合序列,其中小鼠序列位于括号中,人序列采用正常字体,多克隆位点采用粗体,并且Frt序列采用斜体(表7)。

[0431] 表7:由三重ES细胞靶向得到的基因座的接合序列

[0432]

接合区	Seq ID No	序列
小鼠 Tcra/5' Frt	62	(GTCTTTTTTGTTCCTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTTAAACTCTATGGAG TAGTCAGAACACACTCTTCA)GAAGGGACTC CTGATTTCAAAGGGTACCGAAGTTCCTATT CCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC
3' Frt/人 IgK	86	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG AAAGTATAGGAACTTCCTAGGGTTTCACCGGTG GCGCGCCTGAGTAGTGCTTTAGGTGTGTAATCA CCAAAGATTTAGTGAAGTCCCTGTGCAAGGAG
人 IgK/小鼠 Tcra	64	GATAAATTATTTTGTTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC GCCACCGCGGTGGAGCTC(AGGTTTCCG GTTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

[0433] 对分离的ES细胞集落进行等位遗传修饰(MOA)筛选,得到0.4%的靶向效率(参见表8)。

[0434] 表8:用3个LTVEC进行靶向的等位遗传修饰(MOA)筛选结果

[0435]

LTVEC数	核酸酶	缺失	插入物	效率
3个LTVEC	ZFN	hyg	370kb人	0.4%

3个LTVEC	gRNA#16/Cas9	hyg	370kb人	0.4%
3个LTVEC	无	hyg	370kb人	0%

[0436] 上述所靶向的ES细胞用作供体ES细胞,并且通过VELOCIMOUSE[®]方法引入桑椹胚前期胚胎(例如,8细胞期小鼠胚胎)中(参见例如US7,576,259、US 7,659,442、US 7,294,754和US 2008-0078000 A1)。将包含供体ES细胞的小鼠胚胎孵育直到胚泡期为止,然后植入代孕母体中,以产生完全来源于供体ES细胞的F0小鼠。使用检测独特基因序列的存在的等位遗传修饰测定进行基因分型,从而鉴定完全来源于供体ES细胞的独立地携带嵌合人IgK V-小鼠Tera C基因的F0小鼠。

[0437] 实施例4:三个LTVEC与CRISPR/Cas系统相结合对潮霉素基因进行靶向

[0438] 还使用CRISPR/Cas9系统进行实施例3中所述利用锌指核酸酶的三重靶向方法。

[0439] 各种向导RNA (gRNA) 被设计成识别潮霉素抗性基因内的各种靶序列(CRISPR识别序列)。霉素基因内的CRISPR识别序列如下:gRNA#1:ACGAGCGGGTTCGGCCATTCGG (SEQ ID NO:70);gRNA#6:CTTAGCCAGACGAGCGGGTTCGG (SEQ ID NO:71);gRNA#10:GCCGATCTTAGCCAGACGAGCGG (SEQ ID NO:72);以及gRNA#16:CGACCTGATGCAGCTCTCGGAGG (SEQ ID NO:73)。图3中示出了潮霉素基因内的识别序列的位置,对gRNA#1、gRNA#6、gRNA#10和gRNA#16进行筛选,并确认它们能特异性地靶向潮霉素基因(参见图3和表5)。

[0440] MAID 6548ES细胞(对于所有人J κ 区段和四个功能性人V κ 基因区段为杂合的)使用如实施例3中所述的三个大靶向载体与编码Cas9和gRNA#16的载体一起进行电穿孔,所述Cas9和gRNA#16能识别并切割潮霉素抗性基因内的靶位点。

[0441] 通过同源重组将三个大靶向载体插入DNA序列中,从而替换包含和围绕Hyg选择盒的区域。使用实施例3中所述的TAQMAN[®]测定法确认这三个大靶向载体的成功并入。

[0442] ES细胞中所得的所靶向基因座具有表7中所示的接合序列,其中小鼠序列位于括号中,人序列采用正常字体,多克隆位点采用粗体,并且Frt序列采用斜体。

[0443] 对分离的ES细胞集落进行等位遗传修饰(MOA)筛选,得到0.4%的靶向效率(参见表8)。

[0444] 上述所靶向的ES细胞将用作供体ES细胞,并且通过VELOCIMOUSE[®]方法引入桑椹胚前期胚胎(例如,8细胞期小鼠胚胎)中(参见例如US7,576,259、US 7,659,442、US 7,294,754和US 2008-0078000 A1)。将包含经遗传修饰的ES细胞的小鼠胚胎孵育直到胚泡期为止,然后植入代孕母体中,以产生完全来源于供体ES细胞的F0小鼠。使用检测独特基因序列的存在的等位遗传修饰测定进行基因分型,从而鉴定完全来源于供体ES细胞的F0小鼠。

[0445] 实施例5:经由两个LTVEC之间的重叠序列来增强LTVEC靶向

[0446] 采用实施例1中所述的双重靶向系统,使用两个大靶向载体(LTVEC)在单个靶向步骤中修饰基因组基因座。如图1中所示,具有小鼠染色体14上的TCR α 基因座的杂合修饰(含潮霉素选择盒)的细胞通过双重靶向系统进行靶向,以产生包含附加Ig κ 可变基因区段的ES细胞。将这两个不同LTVEC一起共同电穿孔到小鼠胚胎干(ES)细胞中。任选地,将编码核酸内切酶(锌指核酸酶(ZFN)或CRISPR-Cas9)的核酸共同电穿孔,以在靶基因座处或附近产生双链断裂。

[0447] 与实施例1中一样,LTVEC(标记为MAID 1710)包含具有人V κ 1-5和V κ 1-6基因区段

的序列的3' 30kb同源臂、具有人V κ 3-7至V κ 3-15基因区段的120kb序列、以及具有人V κ 1-16基因区段的5' 20kb区域(“重叠区”)。第二LTVEC(标记为MAID 6600)包含3' 20kb重叠区(与第一载体中一样,具有人V κ 1-16基因区段的区域)、具有人V κ 1-17至V κ 2-30基因区段的140kb序列、FRT-Ub-Neo-FRT选择盒以及15.5kb 3' 小鼠TCR A同源臂。

[0448] 成功靶向使得这两个LTVEC通过同源重组插入DNA序列中,从而替换包含和围绕Hyg选择盒的区域。所得的ES细胞在内源TCR A基因座处包含具有人J κ 1至J κ 5和V κ 4-1至V κ 2-30基因区段的人免疫球蛋白可变结构域。使用上述TAQMAN测定法确认这两个大靶向载体的成功并入(Lie and Petropoulos,1998.Curr.Opin.Biotechnology 9:43-48(Lie和Petropoulos,1998年,《生物技术当前述评》,第9卷,第43-48页),该文献以引用方式并入本文),该测定法使用图1和表2中指示的探针和引物。

[0449] 作为比较,也采用实施例1中所述的单LTVEC系统,使用单独或与ZFN或CRISPR-Cas9组合的单个LTVEC来修饰相同基因组基因座(参见图2)。通过上述TAQMAN测定法确认成功并入,该测定法使用上表2中列出和图2中指示的引物和探针。

[0450] 表9比较了各靶向实验中的靶向效率,这些靶向实验使用单个LTVEC(单独、与ZFN联用、或与Cas9联用),同时使用这两个LTVEC(单独、与ZFN联用、或与Cas9联用),或同时使用这两个LTVEC加上第三LTVEC(单独、与ZFN联用、或与Cas9联用)。表9中给出的靶向效率是通过使用表2中的TAQMAN引物和探针初始筛选、确认筛选和再确认筛选而确定为正确靶向的所筛选ESC克隆的百分比。用单个LTVEC单独进行靶向得到1.1%正确靶向的克隆。用ZFN切割将单个LTVEC的靶向效率提高至4.4%,并且用CRISPR-Cas9切割将单个LTVEC的靶向效率提高至5.5%。令人惊讶的是,用2个具有20kb重叠序列的LTVEC进行靶向,得到1.4%的靶向效率,即使未使用核酸酶时也是如此。当使用ZFN时,靶向效率提高至2.81%,并且当使用Cas9时,靶向效率提高至1.6%。

[0451] 表9:共同电穿孔的拼接LTVEC的靶向效率

[0452]

	MAID 6598: 第2插入			MAID 6600: 第2和第3插入			MAID 6647: 第2、第3和第4插入		
	仅 LTVEC	LTVEC + ZFN	LTVEC + Cas9	仅2个 LTVEC	2个LTVEC + ZFN	2个 LTVEC + Cas9	仅3个 LTVEC	3LTVEC + ZFN	3个 LTVEC + Cas9
靶向 效率	5 / 440 (1.1%)	17 / 384 (4.4%)	21 / 384 (5.5%)	13 / 960 (1.4%)	27 / 960 (2.8%)	15 / 960 (1.6%)	0 / 960 (0%)	4 / 960 (0.4%)	4 / 960 (0.4%)

序列表

<110> 维拉·佛洛妮娜

林恩·麦克唐纳

布莱恩·扎姆布罗维兹

安德鲁·J·墨菲

<120> 用于通过单步多重靶向
进行靶向遗传修饰的方法和组合物

<130> 57766-472224

<150> US 62/094,104

<151> 2014-12-19

<150> US 62/167,408

<151> 2015-05-28

<150> US 62/205,524

<151> 2015-08-14

<160> 86

<170> 适用于Windows的FastSEQ 4.0版

<210> 1

<211> 23

<212> DNA

<213> 人工序列

<220>

<223> 与向导RNA (gRNA)

连接的基因组靶序列

<220>

<221> misc_feature

<222> (2) ... (21)

<223> n = A、T、C或G

<400> 1

gnnnnnnnnn nnnnnnnnnn ngg 23

<210> 2

<211> 42

<212> DNA

<213> 人工序列

<220>

<223> ZFN1/2识别位点

<400> 2

tgcgatcget gcgccgatc ttagccagac gagcgggttc gg 42

<210> 3

<211> 36
<212> DNA
<213> 人工序列
<220>
<223> ZFN3/4识别位点
<400> 3
cgctgcggcc gatcttagcc agacgagcgg gttcgg 36
<210> 4
<211> 39
<212> DNA
<213> 人工序列
<220>
<223> ZFN5/6识别位点
<400> 4
agcgtgtccg acctgatgca gctctcggag ggccaagaa 39
<210> 5
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> ZFN1结合序列
<400> 5
tgcgatcgct gcggccga 18
<210> 6
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> ZFN2结合序列
<400> 6
ccgaaccgc tcgtctgg 18
<210> 7
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> ZFN3结合序列
<400> 7
cgctgcggcc gatct 15

<210> 8
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> ZFN4结合序列
<400> 8
ccgaaccgc tcgtc 15
<210> 9
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> ZFN5结合序列
<400> 9
agcgtgtccg acctgatg 18
<210> 10
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> ZFN6结合序列
<400> 10
tttctcgccc tccga 15
<210> 11
<211> 6
<212> DNA
<213> 人工序列
<220>
<223> ZFN1/2切割位点
<400> 11
tcttag 6
<210> 12
<211> 6
<212> DNA
<213> 人工序列
<220>
<223> ZFN3/4切割位点
<400> 12

tagcca 6
<210> 13
<211> 6
<212> DNA
<213> 人工序列
<220>
<223> ZFN5/6切割位点
<400> 13
cagctc 6
<210> 14
<211> 17
<212> DNA
<213> 人工序列
<220>
<223> HYG正向引物
<400> 14
tgcggccgat cttagcc 17
<210> 15
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> HYG探针
<400> 15
acgagcgggt tcggcccatt c 21
<210> 16
<211> 18
<212> DNA
<213> 人工序列
<220>
<223> HYG反向引物
<400> 16
ttgaccgatt ccttgccg 18
<210> 17
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> HYG-U正向引物

<400> 17
cgacgtctgt cgagaagttt ctg 23
<210> 18
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> HYG-U探针
<400> 18
agttcgacag cgtgtccgac ctga 24
<210> 19
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> HYG-U反向引物
<400> 19
cacgccctcc tacatcgaa 19
<210> 20
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> HYG-D正向引物
<400> 20
tgtcgggctg acacaaatcg 20
<210> 21
<211> 22
<212> DNA
<213> 人工序列
<220>
<223> HYG-D探针
<400> 21
ccgtctggac cgatggctgt gt 22
<210> 22
<211> 20
<212> DNA
<213> 人工序列
<220>

<223> HYG-D反向引物
<400> 22
gggcgtcggg ttccactatc 20
<210> 23
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> PGKp1 (Hyg启动子) 正向引物
<400> 23
caaatggaag tagcacgtct cact 24
<210> 24
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> PGKp1 (Hyg启动子) 探针
<400> 24
ctcgtgcaga tggacagcac cgc 23
<210> 25
<211> 15
<212> DNA
<213> 人工序列
<220>
<223> PGKp1 (Hyg启动子) 反向引物
<400> 25
ccgctgcccc aaagg 15
<210> 26
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> hIgK6正向引物
<400> 26
gtcaagcact gctggcacac 20
<210> 27
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> hIgK6探针
<400> 27
aaccccttgctg ctattgaatt gctatgctgt cag 33
<210> 28
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> hIgK6反向引物
<400> 28
tgttgtagac cctccgccac 20
<210> 29
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK12(MAID 1710插入物)正向引物
<400> 29
ttgcctttct cacacctgca g 21
<210> 30
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> hIgK12(MAID 1710插入物)探针
<400> 30
cagcccatcc tgcacttcg ctgga 25
<210> 31
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK12(MAID 1710插入物)反向引物
<400> 31
tggccaaca gtacagctca g 21
<210> 32
<211> 22
<212> DNA

<213> 人工序列
<220>
<223> hIgK13正向引物
<400> 32
tcagtcaatc acctttccca gc 22
<210> 33
<211> 28
<212> DNA
<213> 人工序列
<220>
<223> hIgK13探针
<400> 33
tccccaggta gcctcatgaa ccaatggt 28
<210> 34
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> hIgK13反向引物
<400> 34
cacattactg agtccccaca ggg 23
<210> 35
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> hIgK14正向引物
<400> 35
cattgtcaaa gaagcactgg aaatg 25
<210> 36
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> hIgK14探针
<400> 36
accattgcag ttaccacg gtaggattt tt 32
<210> 37
<211> 23

<212> DNA
<213> 人工序列
<220>
<223> hIgK14反向引物
<400> 37
tcttgcaatg ggatcatcag atg 23
<210> 38
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> Neo正向引物
<400> 38
ggtggagagg ctattcggc 19
<210> 39
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> Neo探针
<400> 39
tgggcacaac agacaatcgg ctg 23
<210> 40
<211> 17
<212> DNA
<213> 人工序列
<220>
<223> Neo反向引物
<400> 40
gaacacggcg gcatcag 17
<210> 41
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK15正向引物
<400> 41
caggtgcaaa ggtgaccaca g 21
<210> 42

<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK15探针
<400> 42
tgggtcctgc ccatccatgc a 21
<210> 43
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> hIgK15反向引物
<400> 43
ggcagcctga gtgtcagagc 20
<210> 44
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK25正向引物
<400> 44
gttcaggccc cacagactct c 21
<210> 45
<211> 29
<212> DNA
<213> 人工序列
<220>
<223> hIgK25探针
<400> 45
tcctctctgg agcaaccatg aagttcct 29
<210> 46
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> hIgK25反向引物
<400> 46
cctgaagcca tgagggcag 19

<210> 47
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hUbC-D (Neo启动子) 正向引物
<400> 47
agggtaggct ctcctgaatc g 21
<210> 48
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> hUbC-D (Neo启动子) 探针
<400> 48
acaggcgccg gacctctggt 20
<210> 49
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hUbC-D (Neo启动子) 反向引物
<400> 49
ccaaagaaac tgacgcctca c 21
<210> 50
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> TCRA Arm4正向引物
<400> 50
gcgccacatg aatttgacca g 21
<210> 51
<211> 28
<212> DNA
<213> 人工序列
<220>
<223> TCRA Arm4探针
<400> 51

tgtaccaat cttccaaaga aagagctg 28

<210> 52

<211> 19

<212> DNA

<213> 人工序列

<220>

<223> TCRA Arm4反向引物

<400> 52

ggcatcctgt cctcccttc 19

<210> 53

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 亲本1540m1正向引物

<400> 53

cagtaaggga agagactaca acagcat 27

<210> 54

<211> 27

<212> DNA

<213> 人工序列

<220>

<223> 亲本1540m1探针

<400> 54

tgcacactgc tcaccactgc aagctat 27

<210> 55

<211> 17

<212> DNA

<213> 人工序列

<220>

<223> 亲本1540m1反向引物

<400> 55

tgctggtggc cccatct 17

<210> 56

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> 亲本1540m3正向引物

<400> 56
gaactcagct atgatagtgt cgaatgta 28
<210> 57
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> 亲本1540m3探针
<400> 57
cagcccagca gctgtgggtt ctc 23
<210> 58
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> 亲本1540m3反向引物
<400> 58
gctcagggag aacacagaac ttaga 25
<210> 59
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> hIgK5正向引物
<400> 59
ccccgtcctc ctcctttttc 20
<210> 60
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> hIgK5探针
<400> 60
tcatgtccat taaccattt accttttgcc ca 32
<210> 61
<211> 19
<212> DNA
<213> 人工序列
<220>

<223> hIgK5反向引物

<400> 61

tgcaagtgct gccagcaag 19

<210> 62

<211> 156

<212> DNA

<213> 人工序列

<220>

<223> 小鼠Tera/5' Frt接合序列

<400> 62

gtcttttttg ttcttcacag ttgagettca tcaaagtcac atgggttaaa ctctatggag 60
tagtcagaac acaactettca gaagggactc ctgatttcaa agggtagcga agtttctatt 120
ccgaagtcc tattctctag aaagtatagg aacttc 156

<210> 63

<211> 132

<212> DNA

<213> 人工序列

<220>

<223> 3' Frt/人IgK接合序列

<400> 63

gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctc agggtttcac 60
cgggtggcgcg cctaacagag aggaaagtca aattataaag aatatgagat tcagaattct 120
gattaactgt gg 132

<210> 64

<211> 124

<212> DNA

<213> 人工序列

<220>

<223> 人IgK/小鼠Tera接合序列

<400> 64

gataaattat tttgtcagac aacaataaaa atcaatagea cgcctaaga gcgcccgcca 60
ccgcggtgga gctcaggttt ccggtactta acaacagagc acagatttag tggtgaggga 120
ctct 124

<210> 65

<211> 28

<212> DNA

<213> 人工序列

<220>

<223> hIgK10正向引物

- <400> 65
cgattatgac tggtaggta gaaagtg 28
- <210> 66
<211> 32
<212> DNA
<213> 人工序列
<220>
<223> hIgK10探针
- <400> 66
gccactgggtt tctccaaatg ttttcaatcc at 32
- <210> 67
<211> 24
<212> DNA
<213> 人工序列
<220>
<223> hIgK10反向引物
- <400> 67
gggagtactt ggagatccct aagc 24
- <210> 68
<211> 136
<212> DNA
<213> 人工序列
<220>
<223> 小鼠Tcra/5' Frt接合序列
- <400> 68
ttgagcttca tcaaagtcac atgggttaaa ctctatggag tagtcagaac acactcttca 60
gaagggactc ctgattcaa agggtagcga agttctatt ccgaagttcc tattctctag 120
aaagtatagg aacttc 136
- <210> 69
<211> 132
<212> DNA
<213> 人工序列
<220>
<223> 3' Frt/人IgK接合序列
- <400> 69
gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctc agggtttccac 60
cggtaggcgcg ccaggacceca ggetctgaca cttagctgc caatacaatt gccatgaaga 120
cagatgttga tg 132
- <210> 70

<211> 23
<212> DNA
<213> 人工序列
<220>
<223> gRNA#1 CRISPR识别序列
<400> 70
acgagcgggt tcggcccatt cgg 23
<210> 71
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> gRNA#6 CRISPR识别序列
<400> 71
cttagccaga cgagcgggtt cgg 23
<210> 72
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> gRNA#10 CRISPR识别序列
<400> 72
gccgatctta gccagacgag cgg 23
<210> 73
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> gRNA#16 CRISPR识别序列
<400> 73
cgacctgatg cagctctcgg agg 23
<210> 74
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK22正向引物
<400> 74
tggetccaag aacagtttgc c 21

<210> 75
<211> 28
<212> DNA
<213> 人工序列
<220>
<223> hIgK22探针
<400> 75
ccctgacttt gctgctcaac tcacagcc 28
<210> 76
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK22反向引物
<400> 76
ggtccagtgg aatctgcat g 21
<210> 77
<211> 23
<212> DNA
<213> 人工序列
<220>
<223> hIgK21正向引物
<400> 77
catttgcta catatcaaag ccg 23
<210> 78
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> hIgK21探针
<400> 78
cctgagccag ggaacagccc actgata 27
<210> 79
<211> 20
<212> DNA
<213> 人工序列
<220>
<223> hIgK21反向引物
<400> 79

acatggctga ggcagacacc 20
<210> 80
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK26正向引物
<400> 80
tgggccgtta tgctagtacc a 21
<210> 81
<211> 25
<212> DNA
<213> 人工序列
<220>
<223> hIgK26探针
<400> 81
tggettacc ccttttgaag ggccc 25
<210> 82
<211> 21
<212> DNA
<213> 人工序列
<220>
<223> hIgK26反向引物
<400> 82
cacagctgaa gcaggatgag c 21
<210> 83
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> hIgK30正向引物
<400> 83
tctctgagca gccatcccc 19
<210> 84
<211> 27
<212> DNA
<213> 人工序列
<220>
<223> hIgK30探针

<400> 84
ttctcctttg gtgtagaggg caccage 27
<210> 85
<211> 19
<212> DNA
<213> 人工序列
<220>
<223> hIgK30反向引物
<400> 85
accaggcatg gcagaaagg 19
<210> 86
<211> 130
<212> DNA
<213> 人工序列
<220>
<223> 3' Frt/人IgK接合
<400> 86
gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctc agggtttcac 60
cggtaggcgcg cctgagtagt gcttaggtg tgtaatcacc aaagatttag tgaagtcct 120
gtgcaaggag 130

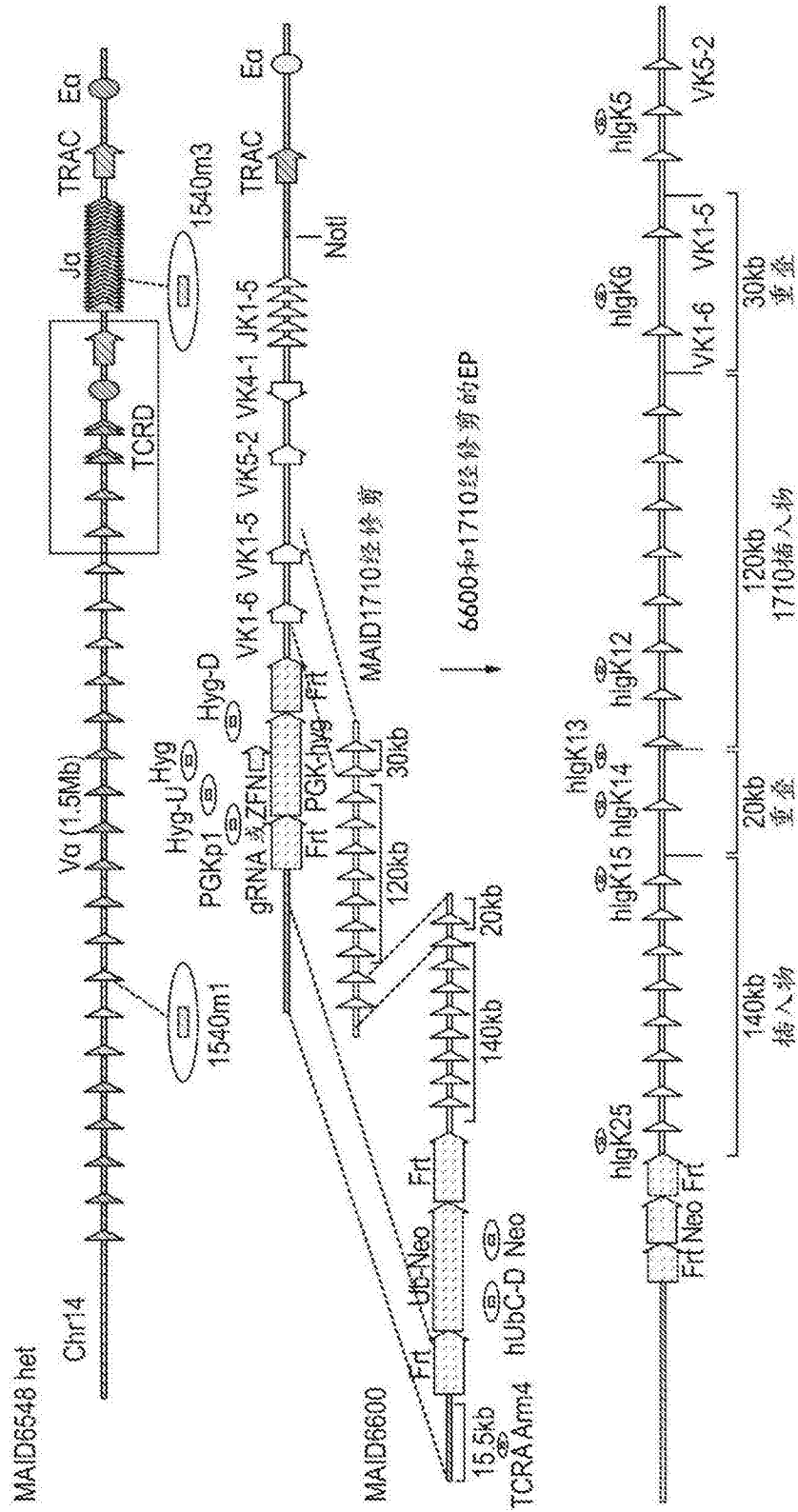


图1

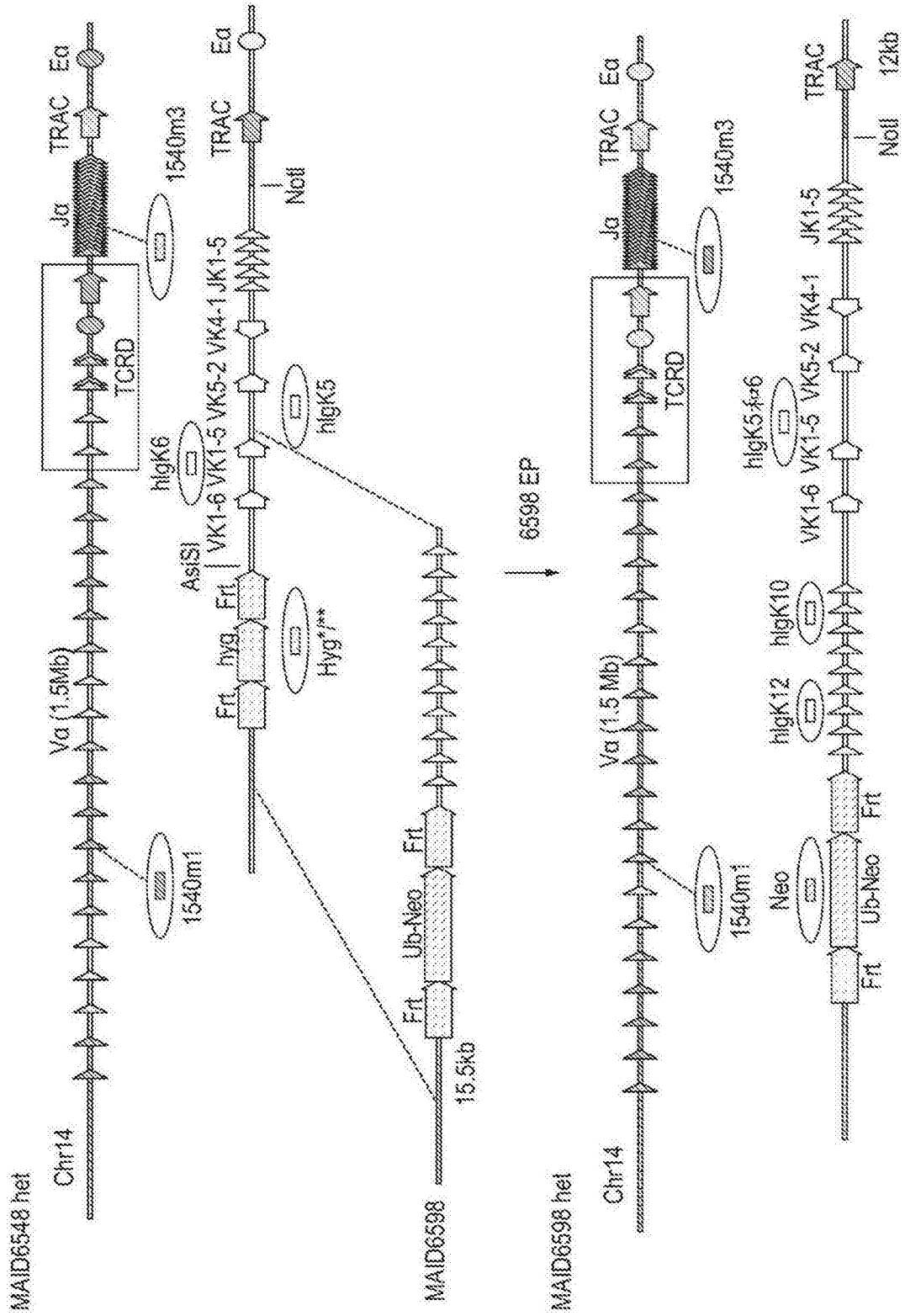


图2

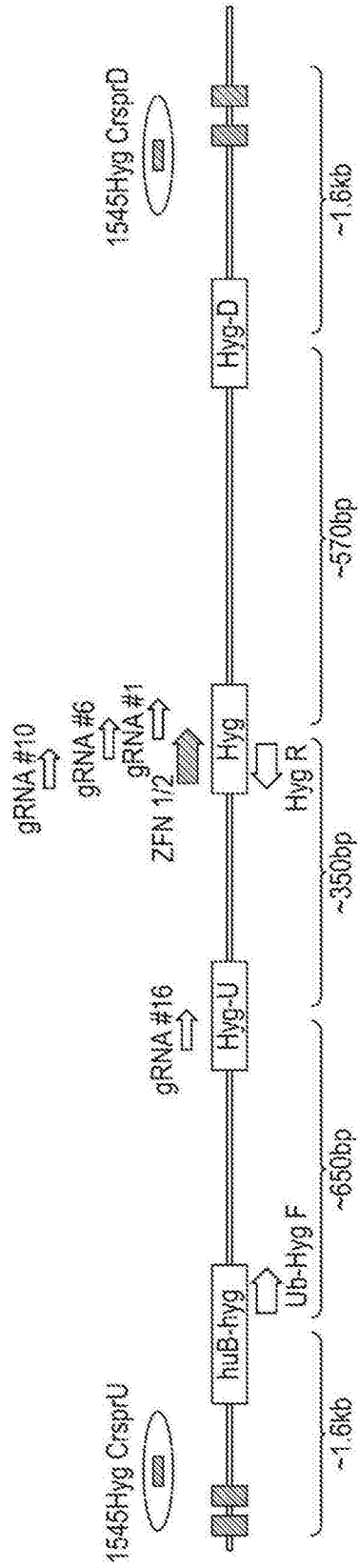


图3

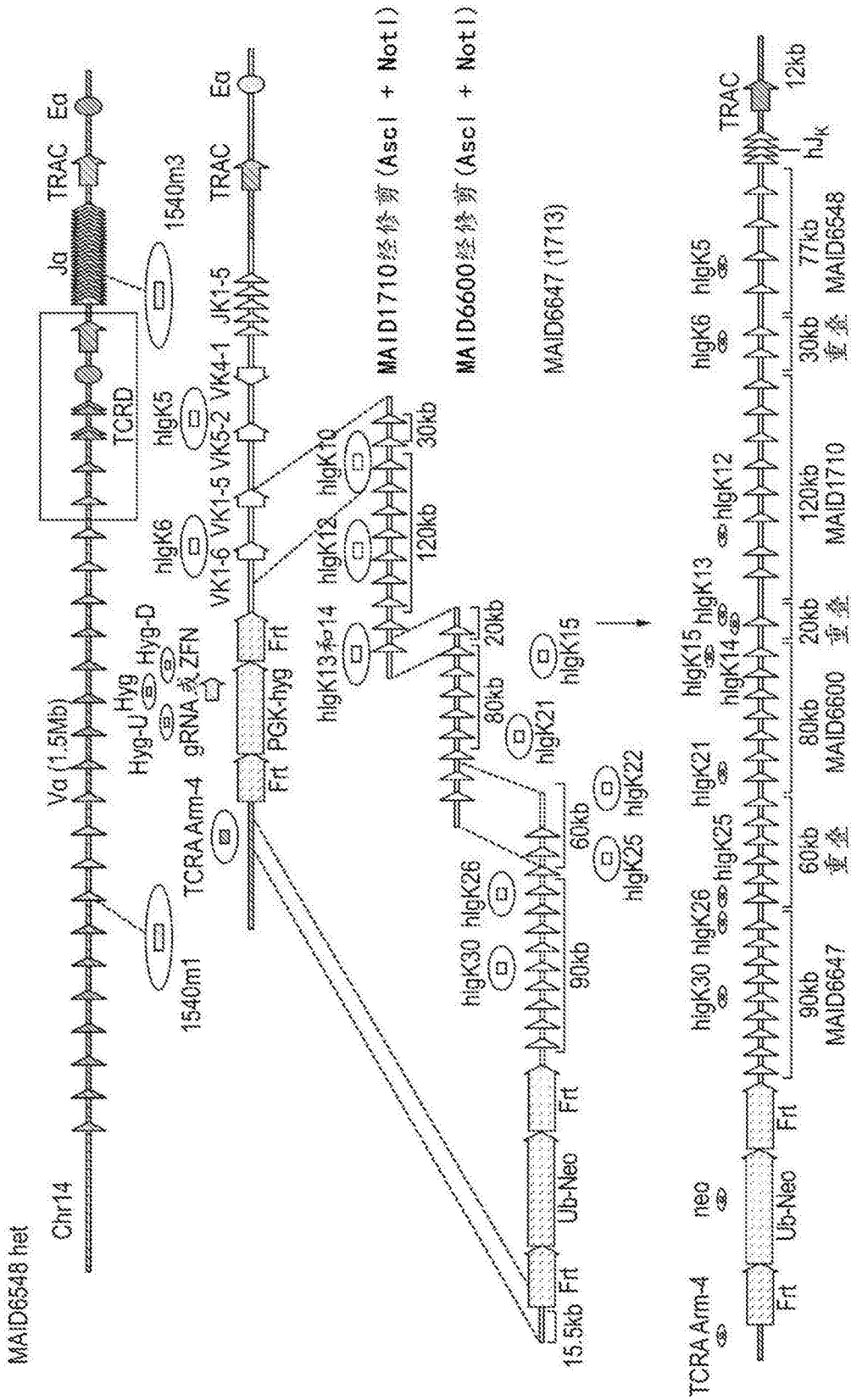


图4

1. 一种用于修饰细胞中的靶基因组基因座的方法,包括:

(a) 向所述细胞中引入在所述靶基因组基因座内产生单链断裂或双链断裂的核酸酶试剂;

(b) 向所述细胞中引入第一大靶向载体(LTVEC)和第二LTVEC,所述第一大靶向载体为至少10kb长且包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物,并且所述第二LTVEC为至少10kb长且包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物,

其中所述第一LTVEC的所述第一3'同源臂具有与所述第二LTVEC的所述第二5'同源臂同源的第一重叠序列,并且所述第一LTVEC的所述第一5'同源臂和所述第二LTVEC的所述第二3'同源臂与所述靶基因组基因座内的对应基因组区段同源,

其中通过所述第一核酸插入物和所述第二核酸插入物在所述对应基因组区段之间的整合来修饰所述靶基因组基因座;以及

(c) 选择所靶向细胞,所述所靶向细胞包含整合到所述靶基因组基因座中的所述第一核酸插入物和所述第二核酸插入物。

2. 根据权利要求1所述的方法,其中所述第一核酸插入物和所述第一3'同源臂以及所述第二核酸插入物和第二5'同源臂是连续核酸的重叠片段,所述连续核酸通过整合所述第一核酸插入物和所述第二核酸插入物到所述靶基因组基因座内而重组。

3. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物或两者来自与所述细胞的物种不同的物种。

4. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物或两者都是人核酸。

5. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物和所述第二核酸插入物的组合大小为约50kb至约500kb,约50kb至约300kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,约275kb至约300kb,约300kb至约350kb,约350kb至约400kb,约400kb至约450kb,或约450kb至约500kb。

6. 根据权利要求5所述的方法,其中所述第一核酸插入物和所述第二核酸插入物的组合大小为约100kb至约500kb。

7. 根据权利要求6所述的方法,其中所述第一核酸插入物和所述第二核酸插入物的组合大小为约300kb。

8. 根据权利要求1-4中任一项所述的方法,其中所述所靶向细胞包含具有所述第一核酸插入物和所述第二核酸插入物两者的基因组DNA,所述第一核酸插入物和所述第二核酸插入物的组合大小在约5kb至约500kb的范围内。

9. 根据前述权利要求中任一项所述的方法,其中所述第一LTVEC的所述第一重叠序列与所述第二LTVEC的所述第一重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同源性。

10. 根据前述权利要求中任一项所述的方法,其中所述第一重叠序列的大小为约1kb至约70kb。

11. 根据前述权利要求中任一项所述的方法,其中所述第一重叠序列的大小为至少10kb或至少20kb。

12. 根据前述权利要求中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物或两者整合到所述靶基因组基因座中导致以下情况中的一种或多种:

(a) 在所述靶基因组基因座处加入外源序列;

(b) 在所述靶基因组基因座处缺失内源序列;以及

(c) 敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

13. 根据前述权利要求中任一项所述的方法,其中与使用单个LTVEC相比,所述第一LTVEC和所述第二LTVEC的组合使用导致靶向效率提升。

14. 根据权利要求13所述的方法,其中靶向效率的所述提升为至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍或20倍。

15. 根据权利要求1-14中任一项所述的方法,其中所述第一LTVEC或所述第二LTVEC的所述5'同源臂和所述3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。

16. 一种用于修饰细胞中的靶基因组基因座的方法,包括:

(a) 向所述细胞中引入在所述靶基因组基因座内产生单链断裂或双链断裂的核酸酶试剂;

(b) 向所述细胞中引入第一大靶向载体(LTVEC)、第二LTVEC和第三LTVEC,所述第一大靶向载体为至少10kb长且包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物,所述第二LTVEC为至少10kb长且包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物,并且所述第三LTVEC为至少10kb长且包含侧接有第三5'同源臂和第三3'同源臂的第三核酸插入物,

其中所述第一LTVEC的所述第一3'同源臂具有与所述第二LTVEC的所述第二5'同源臂同源的第一重叠序列,所述第二LTVEC的所述第二3'同源臂具有与所述第三LTVEC的所述第三5'同源臂同源的第二重叠序列,并且所述第一LTVEC的所述第一5'同源臂和所述第三LTVEC的所述第三3'同源臂与所述靶基因组基因座内的对应基因组区段同源,

其中通过所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物在所述对应基因组区段之间的整合来修饰所述靶基因组基因座;以及

(c) 选择所靶向细胞,所述所靶向细胞包含整合在所述靶基因组基因座中的所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物。

17. 根据权利要求16所述的方法,其中所述第一核酸插入物和所述第一3'同源臂以及所述第二核酸插入物和所述第二5'同源臂是连续核酸的重叠片段,并且所述第二核酸插入物和所述第二3'同源臂以及所述第三核酸插入物和所述第三5'同源臂是所述连续核酸的重叠片段,所述连续核酸通过整合所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物到所述靶基因组基因座内而重组。

18. 根据权利要求16或17所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物中的一者或多者来自与所述细胞的物种不同的物种。

19. 根据权利要求16-18中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物中的一者或多者是人核酸。

20. 根据权利要求16-19中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小为约50kb至约700kb,约50kb至约500kb,约50kb至约300kb,约50kb至约75kb,约75kb至约100kb,约100kb至125kb,约125kb至约150kb,约150kb至约175kb,约175kb至约200kb,约200kb至约225kb,约225kb至约250kb,约250kb至约275kb,约275kb至约300kb,约300kb至约350kb,约350kb至约400kb,约400kb至约450kb,约450kb至约500kb,约500kb至约550kb,约550kb至约600kb,约600kb至约650kb,或约650kb至约700kb。

21. 根据权利要求20所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小为约100kb至约700kb。

22. 根据权利要求21所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小为约400kb。

23. 根据权利要求16-19中任一项所述的方法,其中所述所靶向细胞包含具有所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物三者的基因组DNA,所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物的组合大小在约5kb至约700kb的范围内。

24. 根据权利要求16至23中任一项所述的方法,其中所述第一LTVEC的所述第一重叠序列与所述第二LTVEC的所述第一重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性,并且/或者所述第二LTVEC的所述第二重叠序列与所述第三LTVEC的所述第二重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。

25. 根据权利要求16-24中任一项所述的方法,其中所述第一重叠序列的大小为约1kb至约70kb,并且/或者所述第二重叠序列的大小为约1kb至约70kb。

26. 根据权利要求16-25中任一项所述的方法,其中所述第一重叠序列的大小为至少10kb或至少20kb,并且/或者所述第二重叠序列的大小为至少10kb或至少20kb。

27. 根据权利要求16-26中任一项所述的方法,其中所述第一核酸插入物、所述第二核酸插入物和所述第三核酸插入物中的一者或多者整合到所述靶基因组基因座中导致以下情况中的一种或多种:

(a) 在所述靶基因组基因座处加入外源序列;

(b) 在所述靶基因组基因座处缺失内源序列;或

(c) 敲入、敲除、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换,或它们的组合。

28. 根据权利要求16-27中任一项所述的方法,其中所述第一LTVEC、所述第二LTVEC或所述第三LTVEC的所述5'同源臂和所述3'同源臂的总和为约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约120kb,或约120kb至约150kb。

29. 根据权利要求12或27所述的方法,其中所述靶基因组基因座处的所述内源序列的所述缺失为约5kb至约10kb,约10kb至约20kb,约20kb至约40kb,约40kb至约60kb,约60kb至约80kb,约80kb至约100kb,约100kb至约150kb,约150kb至约200kb,约200kb至约300kb,约300kb至约400kb,约400kb至约500kb,约500kb至约600kb,约600kb至约700kb,或约700kb至

约800kb。

30. 一种用于增强细胞中的靶基因组基因座处的同源重组的方法, 包括向所述细胞内引入第一靶向载体和第二靶向载体, 其中所述第一靶向载体包含侧接有第一5'同源臂和第一3'同源臂的第一核酸插入物, 并且所述第二靶向载体包含侧接有第二5'同源臂和第二3'同源臂的第二核酸插入物,

其中所述第一3'同源臂和所述第二5'同源臂包含重叠核苷酸序列, 并且其中所述第一5'同源臂和所述第二3'同源臂与所述靶基因组基因座内的对应区段同源。

31. 根据权利要求30所述的方法, 其中在未使用核酸酶试剂的情况下, 所述靶基因组基因座处的同源重组增强。

32. 根据权利要求30所述的方法, 还包括向所述细胞内引入在所述靶基因组基因座处或附近产生单链断裂或双链断裂的核酸酶试剂。

33. 根据权利要求1-29和32中任一项所述的方法, 其中所述核酸酶试剂是锌指核酸酶(ZFN)、转录激活因子样效应物核酸酶(TALEN)或大范围核酸酶。

34. 根据权利要求1-29和32中任一项所述的方法, 其中所述核酸酶试剂包括成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白和向导RNA(gRNA)。

35. 根据权利要求34所述的方法, 其中所述Cas蛋白是Cas9。

36. 根据权利要求30-35中任一项所述的方法, 其中所述方法增强了在所述靶基因组基因座处所述第一靶向载体、所述第二靶向载体或两者的所述同源重组。

37. 根据权利要求36所述的方法, 其中所述同源重组的所述增强为至少1.5倍、2倍、2.5倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、10倍、11倍、12倍、13倍、14倍、15倍、16倍、17倍、18倍、19倍或20倍。

38. 根据权利要求30-37中任一项所述的方法, 其中所述第一靶向载体的所述重叠序列与所述第二靶向载体的所述重叠序列具有至少70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、99.5%或99.9%的同一性。

39. 根据权利要求30-38中任一项所述的方法, 其中所述重叠序列为约1kb至约70kb。

40. 根据权利要求30-39中任一项所述的方法, 其中所述重叠序列为至少20kb。

41. 根据权利要求30-40中任一项所述的方法, 其中所述第一靶向载体是在约20kb至约200kb范围内的第一大靶向载体(LTVEC)。

42. 根据权利要求30-41中任一项所述的方法, 其中所述第二靶向载体是在约20kb至约200kb范围内的第二大靶向载体(LTVEC)。

43. 根据权利要求41或42所述的方法, 其中所述第一LTVEC的所述5'同源臂和所述3'同源臂的总和为10kb至约200kb。

44. 根据权利要求42或43所述的方法, 其中所述第二LTVEC的所述5'同源臂和所述3'同源臂的总和为10kb至约200kb。

45. 根据权利要求30-44中任一项所述的方法, 其中所述重叠核苷酸序列有利于将重组机制募集到所述靶基因组基因座处。

46. 根据前述权利要求中任一项所述的方法, 其中所述细胞是人细胞。

47. 根据权利要求1-45中任一项所述的方法, 其中所述细胞是非人细胞。

48. 根据前述权利要求中任一项所述的方法, 其中所述细胞是多能细胞、造血干细胞、

神经元干细胞或成纤维细胞。

49. 根据权利要求48所述的方法,其中所述多能细胞是胚胎干(ES)细胞或诱导多能干(iPS)细胞。

50. 根据权利要求47-49中任一项所述的方法,其中所述细胞是哺乳动物细胞。

51. 根据权利要求50所述的方法,其中所述哺乳动物细胞是啮齿动物细胞。

52. 根据权利要求51所述的方法,其中所述啮齿动物是小鼠或大鼠。

53. 一种用于产生F0代非人动物的方法,包括:

(a) 将非人ES细胞引入非人宿主胚胎,其中所述非人ES细胞是通过根据权利要求1-45和47-52中任一项所述的方法产生的;以及

(b) 在代孕母体中孕育所述非人宿主胚胎,

其中所述代孕母体产生包含所述修饰的所述F0代非人动物。

54. 根据权利要求53所述的方法,其中所述非人动物是小鼠或大鼠。