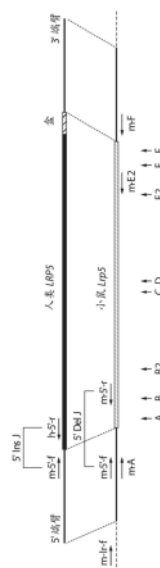




(45) 授权公告日 2025.02.11

序列表38页 附图30页

本文提供了用于形成和促进对细胞内基因组的双等位基因靶向修饰以及用于产生具有经修饰的基因组的非人类动物的组合物和方法。本文还提供了用于将对于等位基因而言为杂合的细胞内的基因组修饰变成对于该等位基因而言为纯合的组合物和方法。所述方法利用Cas蛋白和两种或更多种靶向同一基因组靶基因座内不同位置的向导RNA。本文还提供了鉴定具有经修饰的基因组的细胞的方法。



[接上页]

(56) 对比文件

Yoshimi et al..Allele-specific genome editing and correction of disease-associated phenotypes in rats using the CRISPR-Cas platform.Nature communications.2014,2附图8和第5页左栏第1

段.

Zhou et al..Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting.FEBS Journal.2014,摘要和图1-2.

Ran et al..Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity.Cell.2013,摘要.

1. 一种用于非治疗目的的对哺乳动物细胞内的基因组进行修饰的方法,所述基因组对第一等位基因而言为杂合,包括:

(I) 使所述基因组与下列物质接触:

(a) 第一Cas蛋白;以及

(b) 第一向导RNA,其中所述第一向导RNA包含第一tracrRNA和第一CRISPR RNA,所述第一CRISPR RNA与第一非等位基因特异性CRISPR RNA识别序列杂交;其中所述第一等位基因在第一同源染色体上并且所述CRISPR RNA识别序列位于第二同源染色体上与所述第一等位基因相对应的基因座的近着丝粒侧;以及

(c) 第二向导RNA,其中所述第二向导RNA包含第二tracrRNA和第二CRISPR RNA,所述第二CRISPR RNA与第二非等位基因特异性CRISPR RNA识别序列杂交,所述第二非等位基因特异性CRISPR RNA识别序列位于所述第二同源染色体上与所述第一等位基因相对应的基因座的近着丝粒侧,

其中所述第一等位基因距离所述第一CRISPR RNA识别序列100kb至60Mb,

其中所述Cas蛋白和所述第一向导RNA天然情况下不会同时存在,

其中所述Cas蛋白对双链DNA的两条链都具有核酸酶活性;且

其中所述Cas蛋白切割所述第一CRISPR RNA识别序列,以产生双链断裂,并且所述细胞经修饰变成对于所述第一等位基因而言是纯合的;且

(II) 鉴定对于所述第一等位基因而言是纯合的经修饰的细胞。

2. 根据权利要求1所述的方法,其中步骤(I) 包括向所述细胞中引入: (a) 所述Cas蛋白或编码所述Cas蛋白的核酸;以及 (b) 第一向导RNA或编码所述第一向导RNA的DNA。

3. 根据权利要求2所述的方法,其中步骤(I) 包括将编码所述Cas蛋白的核酸引入所述细胞并将所述第一向导RNA引入所述细胞,其中编码所述Cas蛋白的核酸是RNA。

4. 根据权利要求2所述的方法,其中步骤(I) 包括将编码所述Cas蛋白的核酸和编码所述第一向导RNA的DNA引入所述细胞,其中编码所述Cas蛋白的核酸是DNA。

5. 根据权利要求2所述的方法,其中将所述Cas蛋白和所述第一向导RNA作为第一蛋白-RNA复合物引入所述细胞中。

6. 根据权利要求4所述的方法,其中所述第一CRISPR RNA和所述第一tracrRNA是单独的RNA分子,并且其中:

(1) 编码所述Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子;和/或

(2) 编码所述第一CRISPR RNA的DNA有效连接至第二表达构建体中的第二启动子;和/或

(3) 编码所述第一tracrRNA的DNA有效连接至第三表达构建体中的第三启动子;

其中所述第一启动子、所述第二启动子和所述第三启动子在所述细胞中有活性。

7. 根据权利要求6所述的方法,其中所述第一表达构建体、所述第二表达构建体和所述第三表达构建体是单个核酸分子的组分。

8. 根据权利要求4所述的方法,其中:

(1) 编码所述Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子;和/或

(2) 编码所述第一向导RNA的DNA有效连接至第二表达构建体中的第二启动子,

其中所述第一启动子和所述第二启动子在所述细胞中有活性。

9. 根据权利要求8所述的方法, 其中所述第一表达构建体和所述第二表达构建体是单个核酸分子的组分。

10. 根据前述权利要求中任意一项所述的方法, 其中在所述双链断裂的端粒侧处出现杂合性丢失。

11. 根据权利要求1-9中任意一项所述的方法, 其中步骤(I)还包括将所述第二向导RNA或编码所述第二向导RNA的DNA引入所述细胞中。

12. 根据权利要求1-9中任意一项所述的方法, 其中所述Cas蛋白切割所述第一CRISPR RNA识别序列和所述第二CRISPR RNA识别序列。

13. 根据权利要求1-9中任意一项所述的方法, 其中所述第一CRISPR RNA识别序列和所述第二CRISPR RNA识别序列位于所述第二同源染色体上而非所述第一同源染色体上。

14. 根据权利要求1-9中任意一项所述的方法, 其中所述第一CRISPR RNA识别序列距离所述着丝粒100bp至1kb、1kb至10kb、10kb至100kb、100kb至1Mb、1Mb至10Mb、10Mb至20Mb、20Mb至30Mb、30Mb至40Mb、40Mb至50Mb、50Mb至60Mb、60Mb至70Mb、70Mb至80Mb、80Mb至90Mb, 或90Mb至100Mb。

15. 根据权利要求1-9中任意一项所述的方法, 其中所述第一等位基因距离所述第一CRISPR RNA识别序列1Mb至10Mb、10Mb至20Mb、20Mb至30Mb、30Mb至40Mb、40Mb至50Mb、或50Mb至60Mb。

16. 根据权利要求1-9中任意一项所述的方法, 其中所述第一等位基因距离所述第一CRISPR RNA识别序列至少1Mb、至少10Mb、至少20Mb、至少30Mb、至少40Mb、或至少50Mb。

17. 根据权利要求1-9中任意一项所述的方法, 其中所述第二同源染色体中被杂合性丢失置换的区域为100bp至1kb、1kb至10kb、10kb至100kb、100kb至1Mb、1Mb至10Mb、10Mb至20Mb、20Mb至30Mb、30Mb至40Mb、40Mb至50Mb、或50Mb至60Mb。

18. 根据权利要求1-9中任意一项所述的方法, 其中所述第二同源染色体中被杂合性丢失置换的区域为至少100bp、至少1kb、至少10kb、至少100kb、至少1Mb、至少10Mb、至少20Mb、至少30Mb、至少40Mb、或至少50Mb。

19. 根据权利要求1-9中任意一项所述的方法, 其中

(1) 所述第一等位基因包含突变; 或者

(2) 所述第一等位基因为野生型等位基因, 且在所述第二同源染色体上的相应位点包含突变。

20. 根据权利要求19所述的方法, 其中所述第一等位基因包含突变, 其中所述突变是靶向修饰。

21. 根据权利要求1-9中任意一项所述的方法, 其中所述哺乳动物细胞是人类细胞或啮齿动物细胞。

22. 根据权利要求21所述的方法, 其中所述哺乳动物细胞是所述人类iPS细胞。

23. 根据权利要求21所述的方法, 其中所述哺乳动物细胞是小鼠胚胎干(ES)细胞或大鼠ES细胞。

24. 根据权利要求23所述的方法, 其中所述哺乳动物细胞是所述小鼠ES细胞。

25. 根据权利要求1-9中任意一项所述的方法, 其中所述Cas蛋白是Cas9。

26. 根据权利要求1-9中任意一项所述的方法, 其中所述第一向导RNA中的所述第一

CRISPR RNA和所述第一tracrRNA融合在一起。

27. 根据权利要求1-9中任意一项所述的方法, 其中所述第一CRISPR RNA和所述第一tracrRNA是单独的RNA分子。

28. 根据权利要求1-9中任意一项所述的方法, 其中已经对所述细胞进行了修饰, 以减少非同源末端连接 (NHEJ) 和/或增加基因转换或同源介导修复 (HDR)。

29. 根据权利要求28所述的方法, 其中已经对所述细胞进行了修饰, 以降低下列物质中一者或多者的表达或活性: DNA-PK、PARP1和连接酶IV。

30. 根据权利要求29所述的方法, 其中表达或活性的降低是可诱导的、可逆的、时间特异的和/或空间特异的。

31. 根据权利要求1-9中任意一项所述的方法, 其中步骤 (II) 中的所述鉴定包括等位基因修饰测定。

32. 根据权利要求1-9中任意一项所述的方法, 其中步骤 (II) 中的所述鉴定包括筛选杂合性丢失。

33. 根据权利要求32所述的方法, 其中所述筛选包括进行所述第一CRISPR RNA识别序列的端粒侧的结构变体多态性PCR测定和/或单核苷酸变异等位基因分型测定。

34. 根据权利要求33所述的方法, 其中所述筛选包括在所述第一CRISPR RNA识别序列的端粒侧的不同位置处进行多重结构变体多态性PCR测定和/或多重单核苷酸变异等位基因分型测定。

35. 根据权利要求1-9中任意一项所述的方法, 其中所述细胞对于一个或多个额外等位基因而言是杂合的, 所述第一CRISPR RNA识别序列位于在与在所述第二同源染色体上与所述一个或多个额外等位基因相对应的基因座的近着丝粒侧, 并且同源重组对所述细胞进行修饰, 使其变得对所述一个或多个额外等位基因而言为纯合, 并且步骤 (II) 包括鉴定对于所述一个或多个额外等位基因而言为纯合的经修饰的细胞。

使用成对向导RNA进行靶向遗传修饰的方法和组合物

[0001] 分案申请说明

[0002] 本申请是申请日为2015年11月20日,申请号为201580063428.3,发明名称为“使用成对向导RNA进行靶向遗传修饰的方法和组合物”的发明专利申请的分案申请。

[0003] 相关申请的交叉引用

[0004] 本申请要求2014年11月21日提交的美国专利申请No.62/083,005、2015年6月19日提交的美国专利申请No.62/182,314以及2015年8月28日提交的美国专利申请No.62/211,421的权益,这些美国专利申请中的每一篇均出于所有目的全文以引用方式并入本文。

[0005] 提交的序列表的引用

[0006] 作为通过EFS WEB提交的文本文件

[0007] 写入文件472225SEQLIST.txt中的序列表为32.7kb,创建日期为2015年11月20日,并据此以引用方式并入本文。

背景技术

[0008] 虽然在靶向各种基因组基因座方面已经取得了一定进展,但仍然存在许多类型的无法有效地靶向的基因组基因座或者使用常规靶向策略不能正确地或有效地实现的多种基因组修饰。例如,当试图尤其是在真核细胞和生物体中形成大靶向基因组缺失或其他大靶向遗传修饰时,就会出现问题。

[0009] 具体地讲,当使用常规靶向策略时,很难有效地产生对大靶向基因组缺失或其他基因组修饰而言为纯合的或复合杂合的(如半合子的)细胞或动物。例如,虽然可通过常规靶向策略获得对大靶向基因组缺失而言为杂合的F0代小鼠,但是要产生对该缺失而言为纯合的F2代小鼠,需要使这些杂合的小鼠进一步交配。这些额外的交配步骤不仅成本高昂,而且耗时。

发明内容

[0010] 本发明提供了用于修饰细胞内的基因组的方法和组合物。在一个方面,本发明提供了用于对细胞内的基因组进行修饰的方法,包括使基因组与下列物质接触:(a) 第一Cas蛋白;(b) 与基因组靶基因座内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA;(c) 与基因组靶基因座内的第二CRISPR RNA识别序列杂交的第二CRISPR RNA;(d) tracrRNA;以及(e) 包含侧接5'同源臂和3'同源臂的核酸插入物的靶向载体,其中5'同源臂与5'靶序列杂交且3'同源臂与3'靶序列杂交,前提条件是如果细胞为1细胞期胚胎,则靶向载体的长度不超过5kb;其中该基因组包括含基因组靶基因座的一对第一同源染色体和第二同源染色体;并且其中第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者,以在第一同源染色体和第二同源染色体的至少一者中产生至少一处双链断裂。在一个方面,本发明提供了用于对细胞内的基因组进行双等位基因修饰的方法,包括使基因组与下列物质接触:(a) 第一Cas蛋白;(b) 与基因组靶基因座内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA;(c) 与基因组靶基因座内的第二CRISPR RNA识别序列杂交的第二

CRISPR RNA; (d) tracrRNA; 以及 (e) 包含侧接5' 同源臂和3' 同源臂的核酸插入物的靶向载体, 其中5' 同源臂与5' 靶序列杂交且3' 同源臂与3' 靶序列杂交, 前提条件是如果细胞为1细胞期胚胎, 则靶向载体的长度不超过5kb; 其中该基因组包括含基因组靶基因座的一对第一同源染色体和第二同源染色体; 并且其中第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者, 以在第一同源染色体和第二同源染色体的至少一者中产生至少一处双链断裂。

[0011] 这些方法还可包括鉴定包含经修饰的基因组的细胞。在一些方法中, 核酸插入物包含邻近与第一靶序列杂交的第一同源臂的选择盒, 其中第一同源臂为5' 同源臂且第一靶序列为5' 靶序列, 或者其中第一同源臂为3' 同源臂且第一靶序列为3' 靶序列, 其中鉴定过程包括: (a) 获得细胞中的DNA; (b) 将细胞的DNA暴露于在第一靶序列内结合的探针、在核酸插入物内结合的探针以及在具有已知拷贝数的参考基因内结合的探针, 其中每个探针在结合时产生可检测信号; (c) 检测来自每个探针结合的信号; 以及 (d) 将来自参考基因探针的信号与来自第一靶序列探针的信号进行比较以确定第一靶序列的拷贝数, 并将来自参考基因探针的信号与来自核酸插入物探针的信号进行比较以确定核酸插入物的拷贝数, 其中核酸插入物的拷贝数为一个或两个并且第一靶序列的拷贝数为两个表明核酸插入物在基因组靶基因座处的靶向插入, 并且其中核酸插入物的拷贝数为一个或多个并且第一靶序列的拷贝数为三个或更多个表明核酸插入物在除靶基因组基因座以外的基因组基因座处的随机插入。

[0012] 在一些方法中, 第一Cas蛋白在第一同源染色体和第二同源染色体的每一者中切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者, 以在第一同源染色体和第二同源染色体的每一者中产生至少一处双链断裂。在一些方法中, 第一Cas蛋白在第一同源染色体和第二同源染色体中的至少一者中切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列, 以在第一同源染色体和第二同源染色体中的至少一者中产生至少两处双链断裂。

[0013] 一些方法还包括使基因组与下列物质接触: 与基因组靶基因座内的第三CRISPR RNA识别序列杂交的第三CRISPR RNA; 以及与基因组靶基因座内的第四CRISPR RNA识别序列杂交的第四CRISPR RNA。任选地, 第一CRISPR RNA识别序列与第三CRISPR RNA识别序列相隔约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb至约5kb、约5kb至约6kb、约6kb至约7kb、约7kb至约8kb、约8kb至约9kb、约9kb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或约90kb至约100kb。任选地, 第二CRISPR RNA识别序列与第四CRISPR RNA识别序列相隔约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb至约5kb、约5kb至约6kb、约6kb至约7kb、约7kb至约

8kb、约8kb至约9kb、约9kb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或约90kb至约100kb。任选地,第一CRISPR RNA识别序列和第三CRISPR RNA识别序列为第一对CRISPR RNA识别序列,第二CRISPR RNA识别序列和第四CRISPR RNA识别序列为第二对CRISPR RNA识别序列,其中第一对与第二对相隔约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约5kb、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1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约3Mb、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。

[0014] 在一些方法中,第一Cas蛋白切割第一CRISPR RNA识别序列、第二CRISPR RNA识别序列、第三CRISPR RNA识别序列和第四CRISPR RNA识别序列中的至少两者,以在第一同源染色体和第二同源染色体中的至少一者中产生至少两处双链断裂。在一些方法中,第一Cas蛋白切割第一CRISPR RNA识别序列、第二CRISPR RNA识别序列、第三CRISPR RNA识别序列和第四CRISPR RNA识别序列中的至少两者,以在第一同源染色体和第二同源染色体中均产生至少两处双链断裂。

[0015] 在一些方法中,核酸插入物在5'和3'靶序列之间插入。任选地,5'和3'靶序列在基因组靶基因座内。任选地,细胞不是1细胞期胚胎,并且靶向载体是至少10kb的大靶向载体(LTVEC)。

[0016] 在一些方法中,与基因组单独接触第一CRISPR RNA或第二CRISPR RNA相比,基因组同时接触第一CRISPR RNA和第二CRISPR RNA导致双等位基因修饰效率提高。在一些方法中,细胞是二倍体,并且双等位基因修饰导致基因组靶基因座处的纯合性或复合杂合性。任选地,复合杂合性为半合性。在一些方法中,双等位基因修饰包括在第一同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失。在一些方法中,双等位基因修饰包括在第一同源染色体和第二同源染色体两者中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失。在一些方法中,双等位基因修饰还包括核酸插入物在第一同源染色体和第二同源染色体两者中的5'和3'靶序列之间的插入。在一些方法中,双等位基因修饰包括:(1)在第一同源染色体和第二同源染色体两者中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失;以及(2)核酸插入物在第一同源染色体而非在第二同源染色体中的5'和3'靶序列之间的插入。在一些方法中,双等位基因修饰包括:(1)在第一同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失;以及(2)在第二同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的基因座破坏。在一些方法中,双等位基因修饰包括:(1)在第一同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失;(2)核酸插入物在第一同源染色体中的5'和3'靶序列之间的插入;以及(3)在第二同源染色体中的5'和3'靶序列之间

的基因座破坏。在一些方法中,双等位基因修饰包括:(1)在第一同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失;(2)核酸插入物在第一同源染色体中的5'和3'靶序列之间的插入,其中核酸插入物序列与所缺失序列是同源或直系同源的。

[0017] 在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相隔约1kb至约5kb、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。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相隔至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相隔约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp或约900bp至约1kb。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相隔小于25bp、小于50bp、小于100bp、小于150bp、小于200bp、小于250bp、小于300bp、小于350bp、小于400bp、小于450bp、小于500bp、小于600bp、小于700bp、小于800bp、小于900bp、小于1kb、小于2kb、小于3kb、小于4kb、小于5kb或小于10kb。

[0018] 在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列各自位于距5'和3'靶序列两者至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp、至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、至少6kb、至少7kb、至少8kb、至少9kb、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或至少100kb处。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列各自位于距5'和3'靶序列两者约50bp至约100bp、约200bp至约300bp、约300bp至约400bp、约400bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb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或约50kb至约100kb处。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列各自位于距5'和3'靶序列两者大于50bp、大于100bp、大于200bp、大于300bp、大于400bp、大于500bp、大于600bp、大于700bp、大于800bp、大于900bp、大于1kb、大于2kb、大于3kb、大于4kb、大于5kb、大于6kb、大于7kb、大于8kb、大于9kb、大于10kb、大于20kb、大于30kb、大于40kb、大于50kb、大于60kb、大于70kb、大于80kb、大于90kb或大于100kb处。

[0019] 在一些方法中,所缺失核酸为约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。在一些方法中,所缺失核酸为至少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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。任选地,所缺失核酸为至少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或至少2Mb。

[0020] 在一些方法中,靶向载体是线形形式。任选地,靶向载体是单链或双链的。在一些方法中,细胞不是1细胞期胚胎,并且靶向载体是至少10kb的大靶向载体(LTVEC)。在一些方法中,细胞不是1细胞期胚胎,并且靶向载体是大靶向载体(LTVEC),其中LTVEC的5'和3'同源臂的总和为至少10kb。任选地,LTVEC为约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或约275kb至约300kb。任选地,LTVEC的5'和3'同源臂的总和为约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。

[0021] 在一些方法中,细胞为非人类细胞、非多能的人类细胞,或直接源自分化的成年细胞的人类诱导性多能干细胞。任选地,细胞为哺乳动物细胞、非多能的人类细胞、非人类细胞、啮齿动物细胞、小鼠细胞和大鼠细胞。任选地,细胞为非多能细胞、非人类多能细胞、啮齿动物多能细胞、小鼠多能细胞、大鼠多能细胞、小鼠胚胎干(ES)细胞、大鼠ES细胞或直接源自分化的成年细胞的人类诱导性多能干(iPS)细胞。任选地,细胞为非人类哺乳动物1细胞期胚胎。任选地,细胞为非人类哺乳动物1细胞期胚胎,并且靶向载体的长度为约50个核苷酸至约5kb。任选地,细胞为非人类哺乳动物1细胞期胚胎,并且靶向载体为单链DNA,其长度为约60至约200个核苷酸。

[0022] 在一些方法中,第一Cas蛋白为Cas9。在一些方法中,第一Cas蛋白在双链DNA的两条链上均具有核酸酶活性。

[0023] 在一些方法中,第一Cas蛋白为切口酶。一些方法还包括使基因组与下列物质接触:(f)作为切口酶的第二Cas蛋白;(g)与第三CRISPR RNA识别序列杂交的第三CRISPR RNA;以及(h)与第四CRISPR RNA识别序列杂交的第四CRISPR RNA;其中第一Cas蛋白在第一CRISPR RNA识别序列内和在第二CRISPR RNA识别序列内切割基因组DNA的第一链,并且第二Cas蛋白在第三CRISPR RNA识别序列内和在第四CRISPR RNA识别序列内切割基因组DNA的第二链。

[0024] 在一些方法中,第一CRISPR RNA和tracrRNA融合在一起作为第一向导RNA(gRNA),并且/或者第二CRISPR RNA和tracrRNA融合在一起作为第二gRNA。在一些方法中,第一CRISPR RNA和tracrRNA是单独的RNA分子,并且/或者第二CRISPR RNA和tracrRNA是单独的RNA分子。

[0025] 在一些方法中,接触过程包括将第一Cas蛋白、第一CRISPR RNA和第二CRISPR RNA

以及tracrRNA引入细胞中。在一些方法中, (a) 第一Cas蛋白以蛋白质、编码第一Cas蛋白的信使RNA (mRNA) 或编码第一Cas蛋白的DNA的形式引入细胞中; (b) 第一CRISPR RNA以RNA的形式或以编码第一CRISPR RNA的DNA的形式引入细胞中; (c) 第二CRISPR RNA以RNA的形式或以编码第二CRISPR RNA的DNA的形式引入细胞中; 并且/或者 (d) tracrRNA以RNA的形式或以编码tracrRNA的DNA的形式引入细胞中。在一些方法中, 第一Cas蛋白、第一CRISPR RNA和tracrRNA作为第一蛋白-RNA复合物引入细胞中, 并且/或者第一Cas蛋白、第二CRISPR RNA和tracrRNA作为第二蛋白-RNA复合物引入细胞中。在一些方法中, (a) 编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子; (b) 编码第一CRISPR RNA的DNA有效连接至第二表达构建体中的第二启动子; (c) 编码第二CRISPR RNA的DNA有效连接至第三表达构建体中的第三启动子; 并且/或者 (d) 编码tracrRNA的DNA有效连接至第四表达构建体中的第四启动子; 其中第一启动子、第二启动子、第三启动子和第四启动子在细胞中有活性。任选地, 第一表达构建体、第二表达构建体、第三表达构建体和/或第四表达构建体是单一核酸分子的组分。在一些方法中, (a) 编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子; (b) 编码第一CRISPR RNA和tracrRNA的DNA在编码第一向导RNA (gRNA) 的DNA中融合在一起, 并有效连接至第二表达构建体中的第二启动子; 并且/或者 (c) 编码第二CRISPR RNA和tracrRNA的DNA在编码第二gRNA的DNA中融合在一起, 并有效连接至第三表达构建体中的第三启动子; 其中第一启动子、第二启动子和第三启动子在细胞中有活性。任选地, 第一表达构建体、第二表达构建体和/或第三表达构建体是单一核酸分子的组分。

[0026] 在一些方法中, 已经对细胞进行了修饰, 以减少非同源末端连接 (NHEJ) 和/或增加基因转换或同源介导修复 (HDR)。任选地, 已经对细胞进行了修饰, 以降低DNA-PK的表达或活性和/或降低PARP1的表达或活性。任选地, 已经对细胞进行了修饰, 以降低连接酶IV的表达或活性。任选地, 表达或活性的降低是可诱导的、可逆的、时序特异的和/或空间特异的。

[0027] 在一些方法中, (1) 细胞不是1细胞期胚胎, 并且靶向载体是大靶向载体, 其中5' 和3' 同源臂的总和为至少10kb; (2) 第一CRISPR RNA识别序列和第二CRISPR RNA识别序列各自位于距5' 和3' 靶序列两者大于200bp、大于300bp、大于400bp、大于500bp、大于600bp、大于700bp、大于800bp、大于900bp、大于1kb、大于2kb、大于3kb、大于4kb、大于5kb、大于6kb、大于7kb、大于8kb、大于9kb、大于10kb、大于20kb、大于30kb、大于40kb、大于50kb、大于60kb、大于70kb、大于80kb、大于90kb或大于100kb处; (3) 第一Cas蛋白在第一同源染色体和第二同源染色体中的至少一者中切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列, 以在第一同源染色体和第二同源染色体中的至少一者中产生至少两处双链断裂; 以及 (4) 双等位基因修饰包括在第一同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失以及核酸插入物在第一同源染色体中的5' 和3' 靶序列之间的插入, 其中核酸插入物序列与所缺失序列是同源或直系同源的。

[0028] 本发明还提供了用于产生F0代非人类动物的方法, 包括: (a) 将非人类ES细胞引入非人类宿主胚胎中, 其中非人类ES细胞是通过任意上述方法产生的; 以及 (b) 在代孕母体中孕育非人类宿主胚胎; 其中代孕母体产生包含双等位基因修饰的F0代非人类动物。一些方法包括: (a) 使非人类ES细胞中的基因组与下列物质接触: (i) 第一Cas蛋白; (ii) 与基因组靶基因座内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA; (iii) 与基因组靶基因座内的第二CRISPR RNA识别序列杂交的第二CRISPR RNA; (iv) tracrRNA; 以及 (v) 包含侧接5'

同源臂和3'同源臂的核酸插入物的靶向载体,其中该基因组包括含基因组靶基因座的一对第一同源染色体和第二同源染色体;并且其中第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者,以在第一同源染色体和第二同源染色体中的至少一者中产生至少一处双链断裂;(b) 鉴定包含双等位基因修饰的非人类ES细胞;(c) 将包含双等位基因修饰的非人类ES细胞引入非人类宿主胚胎中;以及(d) 在代孕母体中孕育非人类宿主胚胎;其中代孕母体产生包含双等位基因修饰的F0代非人类动物。

[0029] 在一些方法中,第一Cas蛋白在第一同源染色体和第二同源染色体的每一者中切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者,以在第一同源染色体和第二同源染色体的每一者中产生至少一处双链断裂。在一些方法中,第一Cas蛋白在第一同源染色体和第二同源染色体中的至少一者中切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列,以在第一同源染色体和第二同源染色体中的至少一者中产生至少两处双链断裂。

[0030] 在一些方法中,非人类动物为小鼠,非人类ES细胞为小鼠ES细胞,并且非人类宿主胚胎为小鼠宿主胚胎。在一些方法中,非人类动物为大鼠,非人类ES细胞为大鼠ES细胞,并且非人类宿主胚胎为大鼠宿主胚胎。

[0031] 在一些方法中,双等位基因修饰导致基因组靶基因座处的纯合性或复合杂合性。任选地,复合杂合性为半合性。

[0032] 在一些方法中,第一CRISPR RNA和tracrRNA融合在一起作为第一向导RNA(gRNA),并且/或者第二CRISPR RNA和tracrRNA融合在一起作为第二gRNA。在一些方法中,第一CRISPR RNA和tracrRNA是单独的RNA分子,并且/或者第二CRISPR RNA和tracrRNA是单独的RNA分子。

[0033] 在一些方法中,接触过程包括将第一Cas蛋白、第一CRISPR RNA和第二CRISPR RNA以及tracrRNA引入细胞中。在一些方法中,(a) 第一Cas蛋白以蛋白质、编码第一Cas蛋白的信使RNA(mRNA)或编码第一Cas蛋白的DNA的形式引入细胞中;(b) 第一CRISPR RNA以RNA的形式或以编码第一CRISPR RNA的DNA的形式引入细胞中;(c) 第二CRISPR RNA以RNA的形式或以编码第二CRISPR RNA的DNA的形式引入细胞中;并且/或者(d) tracrRNA以RNA的形式或以编码tracrRNA的DNA的形式引入细胞中。在一些方法中,第一Cas蛋白、第一CRISPR RNA和tracrRNA作为第一蛋白-RNA复合物引入细胞中,并且/或者第一Cas蛋白、第二CRISPR RNA和tracrRNA作为第二蛋白-RNA复合物引入细胞中。在一些方法中,(a) 编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子;(b) 编码第一CRISPR RNA的DNA有效连接至第二表达构建体中的第二启动子;(c) 编码第二CRISPR RNA的DNA有效连接至第三表达构建体中的第三启动子;并且/或者(d) 编码tracrRNA的DNA有效连接至第四表达构建体中的第四启动子;其中第一启动子、第二启动子、第三启动子和第四启动子在细胞中有活性。任选地,第一表达构建体、第二表达构建体、第三表达构建体和/或第四表达构建体是单一核酸分子的组分。在一些方法中,(a) 编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子;(b) 编码第一CRISPR RNA和tracrRNA的DNA在编码第一向导RNA(gRNA)的DNA中融合在一起,并有效连接至第二表达构建体中的第二启动子;并且/或者(c) 编码第二CRISPR RNA和tracrRNA的DNA在编码第二gRNA的DNA中融合在一起,并有效连接至第三表达构建体中的第三启动子;其中第一启动子、第二启动子和第三启动子在细胞中有活性。任选

地,第一表达构建体、第二表达构建体和/或第三表达构建体是单一核酸分子的组分。

[0034] 在一些方法中,第一Cas蛋白为Cas9。在一些方法中,第一Cas蛋白在双链DNA的两条链上均具有核酸酶活性。

[0035] 在一些方法中,第一Cas蛋白为切口酶。一些方法还包括使基因组与下列物质接触:(f)作为切口酶的第二Cas蛋白;(g)与第三CRISPR RNA识别序列杂交的第三CRISPR RNA;以及(h)与第四CRISPR RNA识别序列杂交的第四CRISPR RNA;其中第一Cas蛋白在第一CRISPR RNA识别序列内和在第二CRISPR RNA识别序列内切割基因组DNA的第一链,并且第二Cas蛋白在第三CRISPR RNA识别序列内和在第四CRISPR RNA识别序列内切割基因组DNA的第二链。

[0036] 在一些方法中,已经对细胞进行了修饰,以减少非同源末端连接(NHEJ)和/或增加基因转换或同源介导修复(HDR)。任选地,已经对细胞进行了修饰,以降低DNA-PK的表达或活性和/或降低PARP1的表达或活性。任选地,已经对细胞进行了修饰,以降低连接酶IV的表达或活性。任选地,表达或活性的降低是可诱导的、可逆的、时序特异的和/或空间特异的。

[0037] 本发明还提供了用于产生F0代非人类动物的方法,包括将通过任意上述方法产生的经遗传修饰的1细胞期胚胎植入代孕母体中;其中代孕母体产生包含双等位基因修饰的F0代非人类动物。

[0038] 本发明还提供了用于修饰对第一等位基因而言为杂合的细胞内的基因组的方法,包括使基因组与下列物质接触:(a)第一Cas蛋白;(b)tracrRNA;(c)与第二等位基因内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA,其中第一等位基因在第一同源染色体上,并且第二等位基因在第二同源染色体上的对应基因座处;以及(d)与第二等位基因内的第二CRISPR RNA识别序列杂交的第二CRISPR RNA;其中第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者,以产生至少一处双链断裂和发生重组的末端序列,其中该重组发生在第一等位基因和第二等位基因之间,以形成对第一等位基因而言为纯合的经修饰的基因组。一些方法还包括鉴定对第一等位基因而言为纯合的细胞。

[0039] 在一些方法中,第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列。在一些方法中,第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列,以产生至少两处双链断裂和发生重组的末端序列。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列位于第二等位基因而非第一等位基因内。在一些方法中,Cas蛋白和第一CRISPR RNA在天然情况下不会同时存在。

[0040] 在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相隔约1kb至约5kb、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。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相隔至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相

隔约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp或约900bp至约1kb。在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列相隔小于25bp、小于50bp、小于100bp、小于150bp、小于200bp、小于250bp、小于300bp、小于350bp、小于400bp、小于450bp、小于500bp、小于600bp、小于700bp、小于800bp、小于900bp、小于1kb、小于2kb、小于3kb、小于4kb、小于5kb或小于10kb。

[0041] 在一些方法中,第一等位基因和第二等位基因之间的序列差异跨越约100bp至约200bp、约200bp至约400bp、约400bp至约600bp、约600bp至约800bp、约800bp至约1kb、约1kb至约2kb、约2kb至约3kb、约4kb至约5kb、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。在一些方法中,第一等位基因和第二等位基因之间的序列差异跨越至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少800bp、至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、至少6kb、至少7kb、至少8kb、至少9kb、至少10kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。

[0042] 在一些方法中,第一等位基因包含靶向修饰,并且第二等位基因是野生型等位基因。在一些方法中,第一等位基因是野生型等位基因,并且第二等位基因包含致病突变。

[0043] 在一些方法中,重组包括基因转换。在一些方法中,重组包括杂合性丢失(LOH)。

[0044] 在一些方法中,细胞为非人类细胞、非多能的人类细胞,或直接源自分化的成年细胞的人类诱导性多能干细胞细胞。任选地,细胞为哺乳动物细胞、非多能的人类细胞、非人类细胞、啮齿动物细胞、小鼠细胞或大鼠细胞。任选地,细胞为非多能细胞、非人类多能细胞、啮齿动物多能细胞、小鼠多能细胞、大鼠多能细胞、小鼠胚胎干(ES)细胞、大鼠ES细胞或直接源自分化的成年细胞的人类诱导性多能干(iPS)细胞。

[0045] 在一些方法中,第一Cas蛋白为Cas9。在一些方法中,第一Cas蛋白在双链DNA的两条链上均具有核酸酶活性。

[0046] 在一些方法中,第一Cas蛋白为切口酶。一些方法还包括使基因组与下列物质接触:(f)作为切口酶的第二Cas蛋白;(g)与第三CRISPR RNA识别序列杂交的第三CRISPR RNA;以及(h)与第四CRISPR RNA识别序列杂交的第四CRISPR RNA;其中第一Cas蛋白在第一CRISPR RNA识别序列内和在第二CRISPR RNA识别序列内切割基因组DNA的第一链,并且第二Cas蛋白在第三CRISPR RNA识别序列内和在第四CRISPR RNA识别序列内切割基因组DNA的第二链。

[0047] 在一些方法中,第一CRISPR RNA和tracrRNA融合在一起作为第一向导RNA(gRNA),并且/或者第二CRISPR RNA和tracrRNA融合在一起作为第二gRNA。在一些方法中,第一CRISPR RNA和tracrRNA是单独的RNA分子,并且/或者第二CRISPR RNA和tracrRNA是单独的RNA分子。

[0048] 在一些方法中,接触过程包括将第一Cas蛋白、第一CRISPR RNA和第二CRISPR RNA以及tracrRNA引入细胞中。在一些方法中,(a) 第一Cas蛋白以蛋白质、编码第一Cas蛋白的信使RNA(mRNA)或编码第一Cas蛋白的DNA的形式引入细胞中;(b) 第一CRISPR RNA以RNA的形式或以编码第一CRISPR RNA的DNA的形式引入细胞中;(c) 第二CRISPR RNA以RNA的形式或以编码第二CRISPR RNA的DNA的形式引入细胞中;并且/或者(d) tracrRNA以RNA的形式或以编码tracrRNA的DNA的形式引入细胞中。在一些方法中,第一Cas蛋白、第一CRISPR RNA和tracrRNA作为第一蛋白-RNA复合物引入细胞中,并且/或者第一Cas蛋白、第二CRISPR RNA和tracrRNA作为第二蛋白-RNA复合物引入细胞中。在一些方法中,(a) 编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子;(b) 编码第一CRISPR RNA的DNA有效连接至第二表达构建体中的第二启动子;(c) 编码第二CRISPR RNA的DNA有效连接至第三表达构建体中的第三启动子;并且/或者(d) 编码tracrRNA的DNA有效连接至第四表达构建体中的第四启动子;其中第一启动子、第二启动子、第三启动子和第四启动子在细胞中有活性。任选地,第一表达构建体、第二表达构建体、第三表达构建体和/或第四表达构建体是单一核酸分子的组分。在一些方法中,(a) 编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子;(b) 编码第一CRISPR RNA和tracrRNA的DNA在编码第一向导RNA(gRNA)的DNA中融合在一起,并有效连接至第二表达构建体中的第二启动子;并且/或者(c) 编码第二CRISPR RNA和tracrRNA的DNA在编码第二gRNA的DNA中融合在一起,并有效连接至第三表达构建体中的第三启动子;其中第一启动子、第二启动子和第三启动子在细胞中有活性。任选地,第一表达构建体、第二表达构建体和/或第三表达构建体是单一核酸分子的组分。

[0049] 在一些方法中,已经对细胞进行了修饰,以减少非同源末端连接(NHEJ)和/或增加基因转换或同源介导修复(HDR)。任选地,已经对细胞进行了修饰,以降低DNA-PK的表达或活性和/或降低PARP1的表达或活性。任选地,已经对细胞进行了修饰,以降低连接酶IV的表达或活性。任选地,表达或活性的降低是可诱导的、可逆的、时序特异的和/或空间特异的。

[0050] 本发明还提供了用于修饰对第一等位基因而言为杂合的细胞内的基因组的方法,包括使基因组与下列物质接触:(a) 第一Cas蛋白;(b) tracrRNA;以及(c) 与第一非等位基因特异性CRISPR RNA识别序列杂交的第一CRISPR RNA,其中第一等位基因在第一同源染色体上,并且CRISPR RNA识别序列位于与第二同源染色体上的第一等位基因相对应的基因座的近着丝粒侧;并且其中第一Cas蛋白切割第一CRISPR RNA识别序列,以产生双链断裂,并且细胞被修饰以变得对第一等位基因而言为纯合。一些方法还包括鉴定对第一等位基因而言为纯合的细胞。在一些方法中,Cas蛋白和第一CRISPR RNA在天然情况下不会同时存在。

[0051] 这类方法还可包括使基因组和与第二非等位基因特异性CRISPR RNA识别序列杂交的第二CRISPR RNA接触,该CRISPR RNA识别序列位于与第二同源染色体上的第一等位基因相对应的基因座的近着丝粒侧,其中第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者,以产生至少一处双链断裂。在一些方法中,第一Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列。

[0052] 在一些方法中,在双链断裂处的端粒侧出现杂合性丢失。

[0053] 在一些方法中,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列位于第二同源染色体而非第一同源染色体上。在一些方法中,第一CRISPR RNA识别位点距着丝粒约100bp至约1kb、约1kb至约10kb、约10kb至约100kb、约100kb至约1Mb、约1Mb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。在一些方法中,第一等位基因距第一CRISPR RNA识别位点约100bp至约1kb、约1kb至约10kb、约10kb至约100kb、约100kb至约1Mb、约1Mb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。在一些方法中,被杂合性丢失置换的第二同源染色体区域为约100bp至约1kb、约1kb至约10kb、约10kb至约100kb、约100kb至约1Mb、约1Mb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。

[0054] 在一些方法中,细胞为非人类细胞、非多能的人类细胞,或直接源自分化的成年细胞的人类诱导性多能干细胞细胞。任选地,细胞为哺乳动物细胞、非多能的人类细胞、非人类细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、非多能细胞、非人类多能细胞、啮齿动物多能细胞、小鼠多能细胞、大鼠多能细胞、小鼠胚胎干(ES)细胞、大鼠ES细胞或直接源自分化的成年细胞的人类诱导性多能干(iPS)细胞或非人类哺乳动物1细胞期胚胎。

[0055] 在一些方法中,第一Cas蛋白为Cas9。在一些方法中,第一Cas蛋白在双链DNA的两条链上均具有核酸酶活性。在一些方法中,第一Cas蛋白为切口酶。任选地,第一Cas蛋白为切口酶,并且其中该方法还包括使基因组与作为切口酶的第二Cas蛋白、与第三CRISPR RNA识别序列杂交的第三CRISPR RNA以及与第四CRISPR RNA识别序列杂交的第四CRISPR RNA接触,其中第一Cas蛋白在第一CRISPR RNA识别序列内和在第二CRISPR RNA识别序列内切割基因组DNA的第一链,并且第二Cas蛋白在第三CRISPR RNA识别序列内和在第四CRISPR RNA识别序列内切割基因组DNA的第二链。

[0056] 在一些方法中,第一CRISPR RNA和tracrRNA融合在一起作为第一向导RNA(gRNA),并且/或者第二CRISPR RNA和tracrRNA融合在一起作为第二gRNA。在一些方法中,第一CRISPR RNA和tracrRNA是单独的RNA分子,并且/或者第二CRISPR RNA和tracrRNA是单独的RNA分子。

[0057] 在一些方法中,接触过程包括将第一Cas蛋白、第一CRISPR RNA和第二CRISPR RNA以及tracrRNA引入细胞中。在一些方法中,第一Cas蛋白以蛋白质、编码第一Cas蛋白的信使RNA(mRNA)或编码第一Cas蛋白的DNA的形式引入细胞中。任选地,编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子,其中第一启动子在细胞中有活性。在一些方法中,第一CRISPR RNA以RNA的形式或以编码第一CRISPR RNA的DNA的形式引入细胞中。任选地,编码第一CRISPR RNA的DNA有效连接至第二表达构建体中的第二启动子,其中第二启动子在细胞中有活性。在一些方法中,第二CRISPR RNA以RNA的形式或以编码第二CRISPR RNA的DNA的形式引入细胞中。任选地,编码第二CRISPR RNA的DNA有效连接至第三表达构建体中的第三启动子,其中第三启动子在细胞中有活性。在一些方法中,tracrRNA以RNA的形式或以编码tracrRNA的DNA的形式引入细胞中。任选地,编码tracrRNA的DNA有效连接至第四表达构建体中的第四启动子,其中第四启动子在细胞中有活性。任选地,第一表达构建体、第二表达构建体、第三表达构建体和/或第四表达构建体是单一核酸分子的组分。

[0058] 任选地,编码第一Cas蛋白的DNA有效连接至第一表达构建体中的第一启动子;编码第一CRISPR RNA和tracrRNA的DNA在编码第一向导RNA(gRNA)的DNA中融合在一起,并有

效连接至第二表达构建体中的第二启动子；并且/或者编码第二CRISPR RNA和tracrRNA的DNA在编码第二gRNA的DNA中融合在一起，并有效连接至第三表达构建体中的第三启动子；其中第一启动子、第二启动子和第三启动子在细胞中有活性。任选地，第一表达构建体、第二表达构建体和/或第三表达构建体是单一核酸分子的组分。

[0059] 任选地，第一Cas蛋白、第一CRISPR RNA和tracrRNA作为第一蛋白-RNA复合物引入细胞中，并且/或者第一Cas蛋白、第二CRISPR RNA和tracrRNA作为第二蛋白-RNA复合物引入细胞中。

[0060] 在一些方法中，已经对细胞进行了修饰，以减少非同源末端连接 (NHEJ) 和/或增加基因转换或同源介导修复 (HDR)。任选地，已经对细胞进行了修饰，以降低下列物质中一者或多者的表达或活性：DNA-PK、PARP1和连接酶IV。任选地，表达或活性的降低是可诱导的、可逆的、时序特异的和/或空间特异的。

[0061] 在一些方法中，第一等位基因包含突变。任选地，突变是靶向修饰。在一些方法中，第一等位基因是野生型等位基因，并且第二同源染色体上的对应基因座包含突变。

[0062] 本发明还提供了用于鉴定核酸插入物在不是1细胞期胚胎的二倍体细胞中的靶基因组基因座处的靶向插入的方法，包括：(a) 获得细胞中的DNA，其中已经使细胞与大靶向载体 (LTVEC) 接触，该LTVEC包含侧接与第一靶序列杂交的第一同源臂和与第二靶序列杂交的第二同源臂的核酸插入物，其中核酸插入物包含邻近第一同源臂的选择盒；(b) 将细胞的DNA暴露于在第一靶序列内结合的探针、在核酸插入物内结合的探针以及在具有已知拷贝数的参考基因内结合的探针，其中每个探针在结合时产生可检测信号；(c) 检测来自每个探针结合的信号；以及(d) 将来自参考基因探针的信号与来自第一靶序列探针的信号进行比较以确定第一靶序列的拷贝数，并将来自参考基因探针的信号与来自核酸插入物探针的信号进行比较以确定核酸插入物的拷贝数，其中核酸插入物的拷贝数为一个或两个并且第一靶序列的拷贝数为两个表明核酸插入物在靶基因组基因座处的靶向插入，并且其中核酸插入物的拷贝数为一个或多个并且第一靶序列的拷贝数为三个或更多个表明核酸插入物在除靶基因组基因座以外的基因组基因座处的随机插入。

[0063] 在一些方法中，来自第一靶序列探针结合的信号用于确定第一靶序列的阈值循环 (Ct) 值，来自参考基因探针结合的信号用于确定参考基因的阈值循环 (Ct) 值，并且第一靶序列的拷贝数是通过比较第一靶序列Ct值和参考基因Ct值确定的。在一些方法中，来自核酸插入物探针结合的信号用于确定核酸插入物的阈值循环 (Ct) 值，并且核酸插入物的拷贝数是通过比较第一靶序列Ct值和参考基因Ct值确定的。

[0064] 在一些方法中，选择盒包含药物抗性基因。

[0065] 在一些方法中，核酸插入物为至少5kb、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或至少500kb。在一些方法中，第一靶序列中探针所结合的序列和选择盒之间的距离不超过100个核苷酸、200个核苷酸、300个核苷酸、400个核苷酸、500个核苷酸、600个核苷酸、700个核苷酸、800个核苷酸、900个核苷酸、1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb或5kb。

[0066] 一些方法还包括确定第二靶序列的拷贝数。任选地，步骤 (b) 还包括将细胞的DNA暴露于结合第二靶序列的探针，步骤 (c) 还包括检测来自第二靶序列探针结合的信号，并且

步骤(d)还包括将来自参考基因探针的信号与来自第二靶序列探针的信号进行比较以确定第二靶序列的拷贝数。

[0067] 一些方法还包括确定核酸插入物内的一种或多种附加序列的拷贝数。任选地,步骤(b)还包括将细胞的DNA暴露于结合核酸插入物的一种或多种额外探针,步骤(c)还包括检测来自一种或多种额外探针结合的信号,并且步骤(d)还包括将来自参考基因探针的信号与来自一种或多种额外核酸插入物探针的信号进行比较以确定核酸插入物内的一种或多种附加序列的拷贝数。任选地,核酸插入物内的一种或多种附加序列包括邻近第二靶序列的序列。

[0068] 在一些方法中,LTVEC被设计成从靶基因组基因座中缺失内源性序列,或者还已经使细胞与Cas蛋白、与靶基因组基因座内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA、与靶基因组基因座内的第二CRISPR RNA识别序列杂交的第二CRISPR RNA以及tracrRNA接触。任选地,这类方法还包括确定靶基因组基因座处的内源性序列的拷贝数。任选地,步骤(b)还包括将细胞的DNA暴露于在靶基因组基因座处结合内源性序列的探针,步骤(c)还包括检测来自内源性序列探针结合的信号,并且步骤(d)还包括将来自参考基因探针的信号与来自内源性序列探针的信号进行比较以确定内源性序列的拷贝数。

附图说明

[0069] 图1示出了使用LTVEC以及一个或两个5'区域(A,B,B2)、中间区域(C,D)和3'区域(E2,E,F)gRNA来使小鼠Lrp5胞外域缺失,并同时用对应的人类LRP5形式置换该胞外域的示意图。LTVEC示于图的顶部,小鼠Lrp5基因座示于图的底部。由八个向导RNA导向的Cas9切割位点的位置由小鼠基因序列下方的竖直箭头示出。水平箭头表示小鼠和人类序列的PCR引物。

[0070] 图2A示出了使用LTVEC和两个向导RNA(向导RNA A和B)使小鼠基因缺失,并同时用对应的人类形式置换该基因的一般示意图。LTVEC示于图2A的顶部,小鼠基因座示于图2A的底部。由两个向导RNA导向的Cas9切割位点的位置由小鼠基因序列下方的箭头示出。

[0071] 图2B至图2E示出了当使用两个向导RNA时,以较高频率出现的独特双等位基因修饰(等位基因类型)。具有斜影线的粗线表示小鼠基因,虚线表示小鼠基因中的缺失,并且粗黑线表示人类基因的插入。图2B示出了纯合的剪切(collapsed)等位基因(CRISPR诱导的大缺失)。图2C示出了纯合的靶向等位基因。图2D示出了半合子的靶向等位基因。图2E示出了复合杂合的等位基因。

[0072] 图3A和3B示出了确认所选克隆的基因型的PCR测定结果。图3A示出了使用引物m-lr-f和m-5'-r对所选ES细胞克隆进行长片段PCR测定的结果,所述引物在人类插入物和与5'同源臂同源的序列之外的序列之间形成连接,从而提供正确的靶向。图3B示出了5' De1 J、5' Ins J、De1 A+F和De1 A+E2的PCR测定结果。5' De1 J示出了使用m-5'-f和m-5-r引物得到的PCR产物,所述引物对gRNA A切割位点周围的野生型序列进行扩增,以确定这种序列的保留或丢失。5' Ins J示出了使用m-5'-f和h-5'-r引物得到的PCR产物,所述引物在人类插入物和小鼠基因组之间形成连接。这一测定在靶向和随机整合克隆中均会得到阳性结果。De1 A+F示出了在克隆B0-F10和AW-A8中由双重gRNA A和F切割介导的大缺失的预期扩增子大小(359bp)和实际条带。De1 A+E2示出了克隆BA-A7的相同概念。NT表示无模板,+/+表示

亲本VGF1杂交ES细胞野生型对照,H/+表示杂合的人源化基因型,H/ Δ 表示半合子的人源化基因型,H/H表示纯合的人源化基因型,并且 Δ/Δ 表示纯合的缺失基因型。

[0073] 图4A至图4C示出了用Lrp5人源化LTVEC结合Cas9和两个gRNA靶向的小鼠ES细胞克隆AW-D9(图4A)和BA-D5(图4C)以及单独用LTVEC靶向的克隆BS-C4(图4B)的荧光原位杂交(FISH)分析。箭头表示杂交信号在19号染色体的条带B上的位置。红色信号表示只与小鼠探针杂交(虚线箭头,图4B)。混合黄色信号表示同时与红色小鼠探针和绿色人类探针杂交。一条19号染色体条带B具有红色信号(虚线箭头)且另一条19号染色体条带B具有黄色信号(实线箭头),确认了靶向BS-C4克隆正确的基因座和杂合的基因型(图4B)。两条19号染色体条带B均具有黄色信号(实线箭头,图4A和图4C),确认了靶向AW-D9和BS-C4克隆正确的基因座和纯合的基因型。

[0074] 图5示出了19号染色体的测定结果示意图,所述测定被设计成通过分析VGF1杂交ES细胞中的杂合性丢失(LOH)来检查由两个向导RNA介导的基因转换或有丝分裂重组事件。**TaqMan**[®]qPCR染色体拷贝数(CCN)测定的大概位置由箭头示出。结构变体(SV)多态性PCR测定的大概位置由山形示出,其上给出了其距离Lrp5基因座的距离(以Mb计)。单核苷酸变体(SNV) **TaqMan**[®]等位基因分型测定的大概位置由箭头状物示出,其上给出了其距离Lrp5基因座的距离(以Mb计)。F、E2、D、B2和A的gRNA识别位点的位置由Lrp5基因图示上方的斜箭头示出。

[0075] 图6示出了使用LTVEC和一个或两个5'区域(A,B)、中间区域(C,D)和3'区域(E,E2)gRNA来使小鼠C5(Hc)基因中从外显子2到终止密码子的区域缺失,同时用对应的人类C5形式置换该区域的示意图。LTVEC示于图的顶部,小鼠C5(Hc)基因座示于图的底部。由六个向导RNA导向的Cas9切割位点的位置由小鼠基因序列下方的箭头示出。

[0076] 图7A和图7B示出了用Hc人源化LTVEC结合Cas9和两个gRNA靶向的小鼠ES细胞克隆Q-E9(图7A)和O-E3(图7B)的荧光原位杂交(FISH)分析。箭头表示杂交信号在2号染色体的条带B上的位置。红色信号表示只与小鼠探针杂交(虚线箭头,图7A)。混合黄色信号表示同时与红色小鼠探针和绿色人类探针杂交(实线箭头)。一条2号染色体条带B具有红色信号(虚线箭头)且另一条2号染色体条带B具有黄色信号(实线箭头),确认了靶向Q-E9克隆正确的基因座和杂合的基因型(图7A)。两条2号染色体条带B均具有黄色信号(实线箭头,图7B),确认了靶向O-E3克隆正确的基因座和纯合的基因型。

[0077] 图8示出了使用LTVEC以及一个或两个5'区域(A,B)、中间区域(D,C)和3'区域(E,F)gRNA来使小鼠Ror1基因缺失,并同时用对应的人类ROR1形式置换该基因的示意图。LTVEC示于图的顶部,小鼠Ror1基因座示于图的底部。由六个向导RNA导向的Cas9切割位点的位置由小鼠基因序列下方的箭头示出。

[0078] 图9示出了使用LTVEC以及一个或两个5'区域(A,A2,B)、中间区域(C,D)和3'区域(E2,E,F)gRNA来使小鼠Trpa1基因缺失,并同时用对应的人类TRPA1形式置换该基因的示意图。LTVEC示于图的顶部,小鼠Trpa1基因座示于图的底部。由八个向导RNA导向的Cas9切割位点的位置由小鼠基因序列下方的箭头示出。

[0079] 图10A至图10E示出了克隆BR-B4、BP-G7、B0-G11、B0-F10、B0-A8和BC-H9的结构变异(SV)测定的结果,其中VGF1(F1H4)、129和B6 DNA用作对照。这些测定是在端粒到Lrp5基

因座的距离如下的情况下进行的:13.7Mb (图10A)、20.0Mb (图10B)、36.9Mb (图10C)、48.3Mb (图10D)和56.7Mb (图10E)。B6和129等位基因的PCR产物的位置由箭头示出。

[0080] 图11A至图11C示出了在Lrp5的着丝粒侧0.32Mb (图11A)、在Lrp5的端粒侧1.2Mb (图11B)以及在Lrp5的端粒侧57.2Mb (图11C)的等位基因分型图。每条轴线上的值表示相对荧光强度。这些图表示每个样品的四次重复,结果以实心点 (B6等位基因)、空心点 (129等位基因)和斜线点 (B6/129等位基因两者)示出。

[0081] 图12A至图12C为示出在细胞周期的G2期有丝分裂重组的可能机制的示意图,所述重组可产生由杂合性丢失检测到的纯合事件和广泛基因转换。图12A示出了复制的同源染色体,其示出了对129同源物上的靶向人源化而言为杂合的杂交体129/B6 ES细胞中的两条染色单体。双箭头表示通过双重gRNA导向的Cas9切割产生的可能的双链断裂,所述切割通过同源染色体上的染色单体之间的同源重组促进相互交换,所述交换被示为靶向等位基因的着丝粒侧上的交叉,最终得到图12B中所示的杂交染色单体。图12C示出了在有丝分裂和细胞分裂之后,四种类型的染色体可能分离到子细胞中。保留杂合性的染色体即亲本型杂合子 (Hum/+,左上方)和经等量交换的杂合子 (Hum/+,右上方)这两种类型不能由LOH测定辨别。其他两种类型表现出杂合性丢失,其中人源化纯合子 (Hum/Hum,例如克隆B0-A8,左下方)表现出端粒侧B6等位基因丢失,野生型纯合子 (+/+,右下方)表现出端粒侧129等位基因丢失。后者即野生型纯合子将丢失,因为其未保留人源化等位基因的药物抗性盒。

[0082] 图13示出了使用同源臂大小为35kb和31kb的靶向载体 (LTVEC)或每个同源臂大小为5kb的靶向载体 (sTVEC)以及一个或两个5'区域 (A,B)、中间区域 (C,D)和3'区域 (E,E2) gRNA来使小鼠C5 (Hc)基因中从外显子2到终止密码子的区域缺失,同时用对应的人类C5形式置换该区域的示意图。这两种靶向载体示于图的顶部,小鼠C5 (Hc)基因座示于图的底部。由六个向导RNA导向的Cas9切割位点的位置由小鼠基因序列下方的竖直箭头示出,并且用于筛选的引物由水平箭头示出。确定插入物拷贝数的等位基因获得 (GOA)测定和确定靶向缺失的小鼠序列的等位基因丢失 (LOA)测定的位置由三角形示出。

[0083] 图14示出了使用LTVEC和两个5'区域 (A,B) gRNA来使小鼠Cmah基因的前五个外显子缺失,同时用lacZ报告基因和潮霉素抗性选择盒置换该部分的示意图。LTVEC示于图的顶部,小鼠Cmah基因座示于图的底部。由两个向导RNA导向的Cas9切割位点的位置由小鼠基因序列下方的竖直箭头示出,并且确定插入物拷贝数的GOA测定和确定靶向缺失的小鼠序列的LOA测定的位置由三角形示出。

[0084] 图15示出了当用两个5'区域gRNA (A和B;分别为SEQ ID NOS:107和SEQ ID NOS:108)靶向小鼠Cmah基因座 (SEQ ID NO:109)时,切割事件和所产生的切除产物 (SEQ ID NO:112)的示意图。与Cmah基因座杂交的gRNA序列以粗体示出,Cas9蛋白由带斑点的椭圆表示,Cas9切割位点由竖直箭头示出,并且前间区序列邻近基序 (PAM)用方框括出。**TaqMan[®]** LOA测定正向引物、探针和反向引物的大概位置由在图上方的水平条和箭头示出。进行切割和切除之后所产生的5'和3'片段分别为SEQ ID NOS:110和SEQ ID NOS:111。

[0085] 图16A至图16E示出了阐释在F1杂交小鼠ES细胞中进行的CRISPR/Cas9辅助的人源化实验中所观察到的结果 (包括杂合性丢失 (LOH))的可能机制,所述F1杂交小鼠ES细胞具有一个来源于129S6/SvEvTac小鼠细胞株的单倍体染色体组和一个来源于C57BL/6NTac (B6)小鼠细胞株的单倍体染色体组。图16A示出了通过有丝分裂交换进行的相互染色单体

交换,其中杂合修饰在基因组复制之前或在基因组复制之后发生在129号染色体上,然后在姊妹染色单体之间进行基因转换。图16B示出了通过有丝分裂交换进行的相互染色单体交换,其中单条129号染色单体在基因组复制之后被修饰。图16C示出了通过有丝分裂交换进行的相互染色单体交换,其中尚未发生LTVEC靶向,但Cas9切割已经在129号或B6染色体(示出B6切割)上发生。图16D示出了通过断裂诱导的复制进行的染色单体拷贝,其中杂合修饰在基因组复制之前或在基因组复制之后在129号染色体上发生,然后在姊妹染色单体之间进行基因转换。图16E示出了通过断裂诱导的复制进行的染色单体拷贝,其中单条129号染色单体在基因组复制之后被修饰。图16F示出了通过断裂诱导的复制进行的染色单体拷贝,其中尚未发生LTVEC靶向,但Cas9切割已经在129号或B6染色体(示出B6切割)上发生。

[0086] 图17A至图17C示出了靶向修饰的筛选策略。图17A示出了标准等位基因修饰(MOA)筛选策略以通过大靶向载体(LTVEC)检测杂合靶向,其中小鼠染色体中的内源性序列缺失并被Neo-SDC插入物置换。该策略使用TaqMan[®]探针mTU和mTD针对靶向缺失的内源性序列的上游和下游区域。图17B示出了使用TaqMan[®]保留测定(retU和retD探针)结合等位基因修饰(MOA)测定(mTGU、mTM和mTGD探针用于等位基因丢失(LOA)测定,hTU和hTD探针用于等位基因获得(GOA)测定)来筛选CRISPR/Cas9辅助的人源化。图17C示出了使用TaqMan[®]保留测定(retU和retD探针)结合等位基因丢失(LOA)测定(mTGU、mTM和mTGD探针)来通过成对向导RNA(gU和gD)筛选CRISPR/Cas9辅助的缺失。

[0087] 图18示出了小鼠免疫球蛋白重链基因座(其中可变区基因区段被人类对等物(三角形)置换)以及具有侧接loxP位点的Pgk-Neo插入物(有效连接至新霉素磷酸转移酶基因的磷酸甘油酸激酶I启动子)的靶向载体的大约900kb区域的示意图(未按比例绘制)。两个gRNA用于在5'端处切割小鼠免疫球蛋白重链基因座,两个gRNA用于在3'端处切割基因座,靶向载体使小鼠免疫球蛋白重链基因座缺失并用Pgk-Neo插件置换该基因座。由四个向导RNA导向的Cas9切割位点的位置由靶基因座下方的竖直箭头示出。圈内水平线表示用于等位基因修饰(MOA)测定(hIgH31、hIgH1、mIgHA1、mIgHA7和hIgH9)和保留测定(5' IgH臂1、5' IgH臂2、mIgM-398和mIgM-1045)的TaqMan[®]探针。

[0088] 定义

[0089] 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。

[0090] 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。

[0091] “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),该文献出于所有目的全文以引用方式并入本文。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)。

[0092] “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。

[0093] 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彻底和/或完全互补,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 包括在两条核酸链之间形成的一群杂交复合物发生50%变性时的温度。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。

[0094] “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,附录第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),该文献出于所有目的全文以引用方式并入本文。

[0095] 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。

[0096] 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。

[0097] 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 8 for 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))。

[0098] 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。

[0099] 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”。进行这种调节的方式是本领域技术人员众所周知的。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。

[0100] “序列同一性百分比”包括通过比较窗内比较两个最佳对齐的序列而确定的值,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对齐。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。

[0101] 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。

[0102] 术语“体外”包括人工环境(例如试管)以及发生在人工环境内的过程或反应。术语“体内”包括自然环境(例如细胞或生物体或机体)以及发生在自然环境内的过程或反应。术语“离体”包括已经从个体的机体取下的细胞以及发生在这类细胞内的过程或反应。

[0103] “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。

[0104] 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。

[0105] 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。

[0106] 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

[0107] I.概述

[0108] 本发明提供了用于修饰细胞内的基因组的方法和组合物。这些方法和组合物采用CRISPR/Cas系统,使用两个向导RNA(gRNA)靶向单个基因组靶基因座内不同的位点。例如,这些方法和组合物可采用CRISPR/Cas系统,使用两个向导RNA(gRNA)在单个基因组靶基因座内不同的位点处产生成对的双链断裂。或者,这些方法和组合物可采用CRISPR/Cas系统,使用两个向导RNA(gRNA)在单个基因组靶基因座内不同的位点处产生成对的单链断裂。在一些方法中,可使用两个或更多个(例如三个或四个)向导RNA例如在单个基因组靶基因座内不同的位点处产生两个或更多个单链断裂或双链断裂。

[0109] 一些方法促进双等位基因遗传修饰并包括基因组剪切,由此使大核酸序列从两个切割位点之间的染色体中缺失。其他方法促进双等位基因遗传修饰并包括细胞内核酸序列的缺失以及同时被外源核酸序列置换。如下文进一步详细概述,使用两个gRNA的这些方法通过促进在单个靶向步骤中产生细胞和动物,提高了用双等位基因靶向遗传修饰产生这类细胞或动物的效率。因此,用双等位基因靶向遗传修饰产生动物所必需的动物和交配的数量有所减少。

[0110] 其他方法包括基因转换或杂合性丢失,由此通过在由在对应同源染色体上的对应等位基因中的两个gRNA确定的位点处进行切割,来修饰对等位基因而言为杂合的基因组,使其变得对等位基因而言为纯合。如下文进一步详细概述,在这些方法中使用两个gRNA提高了基因转换频率并实现了在大段染色体DNA上进行基因转换。

[0111] II.CRISPR/Cas系统

[0112] 本文所公开的方法和组合物可利用成簇的规律间隔的短回文重复序列(CRISPR)/CRISPR相关(Cas)系统或此类系统的组分来修饰细胞内的基因组。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。

[0113] 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。

[0114] A.Cas RNA引导的核酸内切酶

[0115] 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。

[0116] 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。

[0117] 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.*)、金黄色葡萄球菌(*Staphylococcus aureus*)、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(*Pseudoalteromonas haloplanktis*)、纤线杆菌(*Ktedonobacter racemifer*)、甲烷盐菌(*Methanohalobium vestigatum*)、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(*Petrogloia mobilis*)、非洲栖热腔菌(*Thermosiphon africanus*)、或深海单细胞蓝细菌(*Acaryochloris marina*)。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*)或从其衍生的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。

[0118] 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。

[0119] 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。

[0120] 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),该文献出于所有目的全文以引用方式并入本文。

[0121] 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CRISPR RNA识别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。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)突变。

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Sapranauskas et al. (2011) *Nucleic Acids Research* 39:9275-9282 (Sapranauskas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均出于所有目的全文以引用方式并入本文。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出于所有目的全文以引用方式并入本文。

[0122] Cas蛋白也可融合蛋白。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W0 2014/089290,该专利出于所有目的全文以引用方式并入本文。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。

[0123] 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SV40 NLS、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),该文献出于所有目的全文以引用方式并入本文。例如,Cas蛋白可融合到一种或多种核定位信号(例如两种或三种核定位信号)。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。

[0124] 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W02014/089290,该专利出于所有目的全文以引用方式并入本文。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。

[0125] 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、mK0、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。

[0126] 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。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

多核苷酸序列相比在细菌细胞、酵母细胞、非多能的人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其他目标宿主细胞中具有更高使用频率的密码子。当将编码Cas蛋白的核酸引入细胞中时,Cas蛋白可在细胞中瞬时地、条件性地或组成型地表达。

[0127] 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蛋白的核酸可在包含核酸插入物的靶向载体和/或包含编码gRNA的DNA的载体中。或者,其可在与包含核酸插入物的靶向载体分离的载体或质粒中,和/或与包含编码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。还可使用在1细胞期胚胎中有活性的启动子。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。其他启动子的示例在本文别处有描述。

[0128] B. 向导RNA (gRNA)

[0129] “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WO 2014/131833A1,这些专利中的每一篇均出于所有目的全文以引用方式并入本文。术语“向导RNA”和“gRNA”是包含性的,包括双分子gRNA和单分子gRNA。

[0130] 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。

[0131] 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。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。tracrRNA的形式可包括初级转录物或加工形式。例如,在化脓性链球菌(*S. pyogenes*)中,不同形式的tracrRNA包括171-核苷酸、89-核苷酸、75-核苷酸和65-核苷酸形式。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)和WO 2014/093661,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。

[0132] crRNA和对应的tracrRNA杂交以形成gRNA。crRNA另外提供了与CRISPR RNA识别序列杂交的单链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均出于所有目的全文以引用方式并入本文。

[0133] 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W02014/131833,该专利出于所有目的全文以引用方式并入本文)。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。

[0134] 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。

[0135] 与靶DNA的核苷酸序列(CRISPR RNA识别序列)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CRISPR RNA识别序列互补的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。

[0136] 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)和WO 2014/093661,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。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,该专利出于所有目的全文以引用方式并入本文。

[0137] DNA靶向序列与靶DNA内的CRISPR RNA识别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靶向序列和靶DNA内的CRISPR RNA识别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%。例如,在靶DNA的互补链内的CRISPR RNA识别序列的5'端的14个邻接核苷酸内,DNA靶向序列与靶DNA内的CRISPR RNA识别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。又如,在靶DNA的互补链内的CRISPR RNA识别序列的5'端的七个邻接核苷酸内,DNA靶向序列与靶DNA内的CRISPR RNA识别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。

[0138] 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。

[0139] 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。此类修饰的示例包括例如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。

[0140] gRNA可包含编码crRNA和tracrRNA的核酸序列。例如,gRNA可包含:(a)具有核酸序列5'-GUUUUAGAGCUAGAAAUAGCAAGUUAUAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'(SEQ ID NO:1)的嵌合RNA;或(b)具有核酸序列5'-GUUUUAGAGCUAGAAAUAGCAAGUUAUAAAUAAGGCUAGUCCG-3'(SEQ ID NO:2)的嵌合RNA。

[0141] 在一些情况下,crRNA包含5'-GUUUUAGAGCUAGAAAUAGCAAGUUAUAAAU-3'(SEQ ID NO:3);5'-GUUUUAGAGCUAGAAAUAGCAAGUUAUAAAUAAG(SEQ ID NO:4);或5'-GAGUCCGAGCAGAAGAAGAAGUUUA-3'(SEQ ID NO:5)。

[0142] 在一些情况下,tracrRNA包含5'-AAGGCUAGUCCG-3'(SEQ ID NO:6)或5'-AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUG CUUUU-3'(SEQ ID NO:7)。

[0143] 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RNA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。

[0144] 当将编码gRNA的DNA引入细胞中时,gRNA可在细胞中瞬时地、条件性地或组成型地表达。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可以在包含核酸插入物的靶向载体和/或包含编码Cas蛋白的核酸的载体中。或者,其可以在与包含核酸插入物的靶向载体分离的载体或质粒中,和/或与包含编码Cas蛋白的核酸的载体分离的载体或质粒中。可用于这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、非人类哺乳动物、人类、啮齿动物、小鼠或仓鼠多能细胞中有活性的启动子。还可使用在1细胞期胚胎中有活性的启动子。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,所述启动子为RNA聚合酶III启动子,诸如人U6启动子、大鼠U6聚合酶III启动子、或小鼠U6聚合酶III启动子。其他启动子的示例在本文别处有描述。

[0145] 作为另一种选择,可通过各种其他方法制备gRNA。例如,可通过采用例如T7 RNA聚合酶的体外转录来制备gRNA(参见例如WO 2014/089290和WO 2014/065596,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也可通过化学合成制备的合成产生的分子。

[0146] C. CRISPR RNA识别序列

[0147] 术语“CRISPR RNA识别序列”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,CRISPR RNA识别序列包括向导RNA被设计成与之具有互补性的序列,其中CRISPR RNA识别序列与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。CRISPR RNA识别序列还包括下文更详细说明的Cas蛋白的切割位点。CRISPR RNA识别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。

[0148] 靶DNA内的CRISPR RNA识别序列可被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) (Sambrook等人,《分子克隆实验指南》,第3版,冷泉港实验室出版社,2001年),该文献出于所有目的全文以引用方式并入本文。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”。

[0149] 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CRISPR RNA识别序列杂交并与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靶向区段结合的核酸序列之外,则切割位点仍被视为在“CRISPR RNA识别序列”

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)。可例如通过使用两种Cas蛋白来产生交错末端,每种Cas蛋白在每条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,并且第二切口酶可在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CRISPR RNA识别序列与第二链上的切口酶的CRISPR RNA识别序列相隔至少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。

[0150] 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CRISPR RNA识别序列。任选地,CRISPR RNA识别序列可在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CRISPR RNA识别序列的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CRISPR RNA识别序列的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)。

[0151] CRISPR RNA识别序列的示例包括与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。例如,靶基序可为紧接在Cas蛋白所识别的NGG基序前面的20个核苷酸的DNA序列,诸如GN₁₉NGG(SEQ ID NO:8)或N₂₀NGG(SEQ ID NO:9)(参见例如WO 2014/165825,该专利出于所有目的全文以引用方式并入本文)。5'端的鸟嘌呤核苷酸可有利于RNA聚合酶在细胞中进行转录。CRISPR RNA识别序列的其他示例可包括5'端的两个鸟嘌呤核苷酸(例如,GGN₂₀NGG;SEQ ID NO:10),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CRISPR RNA识别序列可具有4-22个核苷酸长的SEQ ID NO:8-10,包括5' G或GG以及3' GG或NGG。另外一些CRISPR RNA识别序列可具有14至20个核苷酸长的SEQ ID NO:8-10。CRISPR RNA识别序列的具体示例包括与包含SEQ ID NOS:11-38中任一者的核酸互补的DNA序列。

[0152] CRISPR RNA识别序列可为细胞内源或外源的任何核酸序列。CRISPR RNA识别序列可为编码基因产物(例如,蛋白)的序列或非编码序列(例如,调控序列)或者可包括两者。在一些情况下,CRISPR RNA识别序列可以在与疾病相关的基因或核酸内和/或在与信号通路相关的基因或核酸内。与疾病相关的基因或核酸包括与非疾病对照的组织或细胞相比,在来源于受疾病感染的组织的细胞中产生异常水平或异常形式的转录或翻译产物的任何基因或核酸。例如,与疾病相关的基因可具有一种或多种突变或遗传变异,这些突变或遗传变异与疾病的病因学直接相关或者和与疾病的病因学相关的一个或多个基因存在连锁不平衡。转录或翻译产物可能是已知或未知的,并可能处于正常或异常水平。疾病相关的基因和核酸的示例可得自马里兰州巴尔的摩的约翰霍普金斯大学McKusick-Nathans遗传医学研究所(McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD))以及马里兰州贝塞斯达的美国国家医学图书馆美国国家生物技术信息中心(National Center for Biotechnology Information, National Library of Medicine

(Bethesda, MD)), 具体信息可见于万维网。对于疾病相关的基因和核酸的另外示例, 参见美国专利No. 8, 697, 359, 该专利出于所有目的全文以引用方式并入本文。

[0153] 致病基因中的突变可以是隐性突变或显性突变。二倍体生物体(即每一染色体具有两个拷贝的生物体)通常携带每个基因的两个拷贝。如果个体中的两个拷贝是相同的, 则该个体对该基因而言为纯合的。如果这两个拷贝是不同的等位基因, 则该个体对该基因而言为杂合的。术语“基因型”包括个体在单个基因(或多个基因)中是否携带突变, 并且术语“表型”包括该基因型的物理或功能结果。隐性突变包括这样的突变: 其中两个等位基因必须都突变才能观察到突变体表型(即, 生物体对于该突变体等位基因而言必须是纯合的, 才能表现突变体表型)。隐性突变可例如使受感染基因失活并导致功能丧失。例如, 隐性突变可移除染色体中的所有或部分基因、破坏基因的表达或改变所编码蛋白的结构, 从而改变其功能。相比之下, 显性突变包括这样的突变: 其中在对突变而言为杂合的生物体中也能观察到突变体表型(即, 生物体携带一个突变体等位基因和一个野生型等位基因)。显性突变可例如导致功能获得。例如, 显性突变可提高给定基因产物的活性、赋予基因产物新活性, 或导致其不恰当的时序或空间表达。显性突变也可与功能丧失相关。在一些情况下, 如果需要两个拷贝的基因来发挥正常功能, 那么移除单个拷贝可能引起突变体表型。这类基因是单倍剂量不足的。在其他情况下, 一个等位基因的突变可能导致蛋白质结构变化, 这种变化会干扰由其他等位基因编码的野生型蛋白的功能。这类突变为显性负突变。一些等位基因可能同时与隐性和显性表型相关。

[0154] 一些CRISPR RNA识别序列在包含突变的基因或核酸内。这种突变可例如为显性突变或隐性突变。在某些情况下, 显性突变在对显性突变而言为杂合的细胞内(也就是说, 细胞包含野生型等位基因和具有显性突变的突变体等位基因)。在一些这类情况下, CRISPR RNA识别序列可能在突变体等位基因而非野生型等位基因内。或者, CRISPR RNA识别序列可能在野生型等位基因而非突变体等位基因内。

[0155] III. 靶向载体和核酸插入物

[0156] 本文所公开的方法和组合物也可利用包含核酸插入物和同源臂的靶向载体来修饰细胞内的基因组。在此类方法中, 核酸插入物通过同源物重组事件被整合到由同源臂确定的基因组靶基因座中。本文所提供的方法可利用核酸酶试剂(例如Cas蛋白)与同源重组事件结合。此类方法采用由核酸酶试剂在核酸酶切割位点处形成的切口或双链断裂并结合同源重组, 以便于将核酸插入物靶向整合到基因组靶基因座中。

[0157] A. 用于除1细胞期胚胎之外的细胞的靶向载体和核酸插入物

[0158] (1) 核酸插入物

[0159] 一个或多个分开的核酸插入物可以用于本文公开的方法中, 并且可以通过分开的靶向载体或在相同的靶向载体上将它们导入细胞。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(即缺失和插入)。

[0160] 核酸插入物或在靶基因座处被置换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, 核酸插入物或在靶基因座处被置换的相应核酸可以是任何期望的长度, 包括例如在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核苷酸的长度、1.5kb至2kb核苷酸的长度、2kb至2.5kb核苷酸的长度、2.5kb至3kb核苷酸的长度、3kb至5kb核苷酸的长度、5kb至8kb核苷酸的长度、8kb至10kb核苷酸或更长的长度。在其他情况下,长度可以为约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、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。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。一些核酸插入物可能甚至更小。例如,可以插入长度为约4个核苷酸至约12个核苷酸的插入物,以形成限制性内切酶位点。

[0161] 在一些靶向载体中,核酸插入物可以为约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、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或约350kb至约400kb。

[0162] 在一些情况下,置换靶基因座处的核酸导致约1kb至约200kb、约2kb至约20kb或约0.5kb至约3Mb范围内的核酸序列缺失。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。

[0163] 在一些情况下,核酸序列的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。

[0164] 在其他情况下,核酸插入物或在靶基因座处被置换的相应核酸可以是至少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。

[0165] 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)、非人类哺乳动物、啮齿动物、人类、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。

[0166] 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”)。

[0167] 例如,核酸插入物可以与在基因组靶基因座处靶向缺失的序列同源或直系同源。同源或直系同源核酸插入物可置换在基因组靶基因座处靶向缺失的序列。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,使得其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。如果核酸插入物的插入致使非人类核酸序列被同源或直系同源人类核酸序列置换(即,核酸插入物代替对应的非人类DNA序列被插入在其内源性基因组基因座处),则直系同源序列可导致基因座的人源化。

[0168] 核酸插入物还可以包含条件型等位基因。条件型等位基因可为如US2011/0104799中所述的多功能等位基因,该专利出于所有目的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US2011/0104799,该专利出于所有目的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US2011/0104799,该专利出于所有目的全文以引用方式并入本文。

[0169] 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。编码选择标记物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。启动子的示例在本文别处有所描述。

[0170] 在一些靶向载体中,核酸插入物包含报告基因。报告基因的示例为编码以下物质的基因: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,或它们的组合。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。启动子的示例在本文别处有所描述。

[0171] 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。

[0172] 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。可位于核酸插入物或核酸插入物中任何所关注的多核苷酸两侧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。在一个示例中,位点特异性重组位点位于核酸插入物内所包含的编码选择标记物和/或报告基因的多核苷酸两侧。在靶向基因座处整合核酸插入物之后,可移除位点特异性重组位点之间的序列。

[0173] (2) 靶向载体

[0174] 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位于核酸插入物两侧的同源臂。靶向载体可以是线形形式或环形形式,并且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'同源臂对应于靶向基因座内的区域,该区域在本文中分别被称为“5'靶序列”和“3'靶序列”。一些靶向载体包含没有核酸插入物的5'和3'同源臂。此类靶向载体可起到在不插入核酸插入物的情况下使5'和3'靶序列之间的序列缺失的作用。

[0175] 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%序列同一性,以使得所述序列经历同源重组。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在切割的识别位点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(如在本文其他地方描述的LTVEC载体中所述),以使得同源臂与细胞基因组内的对应靶序列具有足以经历同源重组的同源性。

[0176] 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,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而言为天然、异源或外源的基因座相对应:原核生物、酵母、禽类(例如,鸡)、非人类哺乳动物、啮齿动物、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,或任何其他目标生物体。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。

[0177] 在一些靶向载体中,5'和3'同源臂对应于靶向基因组。或者,同源臂可来自相关基因组。例如,靶向基因组为第一细胞株的小鼠基因组,靶向臂来自第二细胞株的小鼠基因组,其中第一细胞株与第二细胞株不同。在某些情况下,同源臂来自相同动物的基因组或来自相同细胞株的基因组,例如靶向基因组为第一细胞株的小鼠基因组,靶向臂来自相同小鼠的小鼠基因组或来自相同细胞株的小鼠基因组。

[0178] 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。

[0179] 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。

[0180] 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,靶序列可位于核酸酶切割位点的3'端,或者靶序列可位于核酸酶切割位点两侧。

[0181] 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以提高至少两倍、至少三倍、至少4倍或至少10倍。

[0182] (3) 大靶向载体

[0183] 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。

[0184] 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及其制备方法的非限制性示例描述于例如美国专利No.6,586,251;美国专利No.6,596,541;美国专利No.7,105,348;和WO 2002/036789(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。

[0185] LTVEC可以为任何长度,包括例如约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,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。

[0186] 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,核酸插入物可以在约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。

[0187] 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5kb至约100kb的范围内,并且/或者3'同源臂在约5kb至约100kb的范围内。每个同源臂可以为例如约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。5'和3'同源臂的总和可以为例如约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。或者,每个同源臂可以为至少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'同源臂的总和可以为至少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。

[0188] 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、约150kb至约200kb、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。或者,缺失可以为至少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。

[0189] 在其他情况下,LTVEC和核酸插入物被设计成允许向靶基因座插入约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范围内的外源核酸序列。另选地,插入可以为至少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。

[0190] 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。

[0191] B. 用于1细胞期胚胎的靶向载体和核酸插入物

[0192] 用于1细胞期胚胎中的靶向载体的长度不超过5kb,并且可以是脱氧核糖核酸(DNA)或核糖核酸(RNA),它们可以是单链或双链的,并且它们可以是环形形式或线形形式。用于1细胞期胚胎中的示例性靶向载体的长度在约50个核苷酸至约5kb之间。例如,用于1细胞期胚胎中的靶向载体的长度可以在约50至约100、约100至约200、约200至约300、约300至约400、约400至约500、约500至约600、约600至约700、约700至约800、约800至约900或约900至约1,000个核苷酸之间。或者,用于1细胞期胚胎中的靶向载体的长度可以在约1kb至约1.5kb、约1.5kb至约2kb、约2kb至约2.5kb、约2.5kb至约3kb、约3kb至约3.5kb、约3.5kb至约4kb、约4kb至约4.5kb或约4.5kb至约5kb之间。或者,用于1细胞期胚胎中的靶向载体的长度可以为例如不超过5kb、4.5kb、4kb、3.5kb、3kb、2.5kb、2kb、1.5kb、1kb、900个核苷酸、800个核苷酸、700个核苷酸、600个核苷酸、500个核苷酸、400个核苷酸、300个核苷酸、200个核苷酸、100个核苷酸或50个核苷酸。就单链DNA供体而言,示例性靶向载体可以在约60个核苷酸和约200个核苷酸(例如,约60个核苷酸至约80个核苷酸、约80个核苷酸至约100个核苷酸、约100个核苷酸至约120个核苷酸、约120个核苷酸至约140个核苷酸、约140个核苷酸至约160个核苷酸、约160个核苷酸至约180个核苷酸或约180个核苷酸至约200个核苷酸)之间。

[0193] 这类靶向载体包含对应于靶向基因座内的区域(分别为5'靶序列和3'靶序列)的5'和3'同源臂。任选地,靶向载体包含侧接5'和3'同源臂的核酸插入物(例如,在基因组靶基因座处整合的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,或对靶基因座处所关注的核酸序列的替代(即缺失和插入)。

[0194] 同源臂与对应靶序列之间的对应同源区可具有足以促进同源重组的任何长度。用于1细胞期胚胎中的示例性同源臂的长度在约20个核苷酸至约2.5kb之间(例如长度为约30个核苷酸至约100个核苷酸)。例如,给定同源臂和/或对应靶序列可包含长度在约20至约30、约30至约40、约40至约50、约50至约60、约60至约70、约70至约80、约80至约90、约90至约100、约100至约150、约150至约200、约200至约250、约250至约300、约300至约350、约350至约400、约400至约450或约450至约500个核苷酸之间的对应同源区,以使得同源臂与细胞基因组内的对应靶序列具有足以发生同源重组的同源性。或者,给定同源臂和/或对应靶序列可包含长度在约0.5kb至约1kb、约1kb至约1.5kb、约1.5kb至约2kb或约2kb至约2.5kb之间

的对应同源区。就单链DNA供体而言,示例性同源臂可以在约30个核苷酸和约60个核苷酸之间(例如,约30个核苷酸至约40个核苷酸、约40个核苷酸至约50个核苷酸或约50个核苷酸至约60个核苷酸)。

[0195] 如上文所述,同源臂可与对细胞而言为天然的基因座(例如,靶基因座)相对应,或者,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。如上文所述,5'和3'靶序列优选地位于足够接近Cas切割位点,以促进在Cas切割位点处产生单链断裂(切口)或双链断裂后在靶序列与同源臂之间发生同源重组事件。

[0196] 核酸插入物或在靶基因座处缺失和/或被置换的相应核酸可具有各种长度。示例性核酸插入物或在靶基因座处缺失和/或被置换的相应核酸的长度在约10个核苷酸至约5kb之间。例如,核酸插入物或在靶基因座处缺失和/或被置换的相应核酸的长度可以在约1至约10、约10至约20、约20至约30、约30至约40、约40至约50、约50至约60、约60至约70、约70至约80、约80至约90、约90至约100、约100至约110、约110至约120、约120至约130、约130至约140、约140至约150、约150至约160、约160至约170、约170至约180、约180至约190、约190至约200、约200至约300、约300至约400、约400至约500、约500至约600、约600至约700、约700至约800、约800至约900或约900至约1,000个核苷酸之间。例如,可以插入长度为约4个核苷酸至约12个核苷酸的插入物,以形成限制性内切酶位点。同样,核酸插入物或在靶基因座处缺失和/或被置换的相应核酸的长度可以在约1kb至约1.5kb、约1.5kb至约2kb、约2kb至约2.5kb、约2.5kb至约3kb、约3kb至约3.5kb、约3.5kb至约4kb、约4kb至约4.5kb或约4.5kb至约5kb之间。从基因组靶基因座缺失的核酸的长度还可以在约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200kb、约kb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800kb至约900kb、约900kb至约1Mb之间或更长。或者,从基因组靶基因座缺失的核酸的长度可以在约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约3Mb、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之间。

[0197] 如上文所述,核酸插入物可包含基因组DNA或任何其他类型的DNA,核酸插入物或在靶基因座处缺失和/或被置换的相应核酸可以是编码区或非编码区,并且核酸插入物可与在靶基因组基因座处靶向缺失的序列同源或直系同源。核酸插入物还可包含条件型等位基因、编码选择标记物的多核苷酸、报告基因、一个或多个表达盒、一个或多个缺失盒或包含上述侧接有位点特异性重组靶序列的核酸的核酸插入物。

[0198] C. 启动子

[0199] 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大鼠、真核生物、哺乳动物、非人类哺乳动物、人类、啮齿动物、小鼠或仓鼠多能细胞中有活性。还可使用在1细胞期胚胎中有活性的启动子。启动子可以是例如组成型活性启动子、条件型启动子、诱导型启动子、时序限制型启动子(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)。启动子的示例可见于例如WO 2013/176772,该专利全文以引用方式并入本文。

[0200] 诱导型启动子的示例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)。

[0201] 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)。

[0202] 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。

[0203] 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非人类哺乳动物细胞、非人类多能细胞、直接源自分化的成年细胞的人类诱导性多能干细胞、非多能的人类细胞、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。

[0204] IV. 修饰基因组和产生经遗传修饰的非人类动物的方法

[0205] A. 修饰基因组的方法

[0206] 提供了通过使用两个向导RNA来靶向单个基因组靶基因座内的不同区域,从而修饰细胞内基因组的各种方法。还提供了使用两个或更多个向导RNA(例如,三个向导RNA或四个向导RNA)来靶向单个基因组靶基因座内的不同区域的方法。这些方法可在体外、离体或体内进行。这类方法促进双等位基因遗传修饰的形成并可包括基因组剪切或其他靶向修饰,诸如基因组内核酸序列的缺失以及同时被外源核酸序列置换。

[0207] 通过在靶向载体与靶基因座之间发生同源重组而产生的靶向基因修饰可能效率非常低,特别是在除啮齿动物胚胎干细胞之外的细胞类型中。使用靶向载体并结合靶基因座处核酸酶导向的双链DNA断裂可大大提高简单修饰(诸如小缺失或插入)的杂合靶向效率。

[0208] 将靶向载体与由一个向导RNA(gRNA)导向的CRISPR/Cas9核酸酶结合还可提高极大且极低效率基因修饰(诸如小鼠基因的缺失以及同时被人类对等物(人源化)置换)的杂合靶向效率。这类修饰可涉及极大的(例如>50kb)缺失和插入(参见实例1中的Lrp5、C5(Hc)、Ror1和Trpa1靶向)。

[0209] 在通过核酸酶(诸如Cas9)在靶基因组基因座处产生的一处或多处双链断裂的同源介导修复期间,首先通过切除5'端对该一处或多处断裂进行加工,以形成3'单链悬垂序列。然后Rad51在单链DNA上聚合以寻找同源序列,链侵入到未损坏的同源模板双螺旋DNA(例如,靶向载体)中,并形成中间D环结构以有利于使用未损坏的同源DNA(例如,靶向载体)作为模板修复该一处或多处双链断裂。然后通过涉及两侧同源区的双交换事件,用来自靶向载体的核酸插入物置换染色体序列。这一过程是否能够顺利进行受到若干种因素的影响,诸如核酸插入物的大小、与靶向载体的同源臂同源的区域长度,以及与靶向载体的同源臂同源的区域位置(例如,相对于一处或多处双链断裂的位置)。

[0210] 随着核酸插入物或在靶基因组基因座处缺失的序列的大小增加,切除过程变得更

加不可预测,中间D环结构的稳定性降低并变得更加不可预测,并且重组过程的成功率大体上有所降低并变得更加不可预测。例如,随着靶向修饰的大小增加,内部重组的风险也增大,尤其是当被置换的序列与所插入的序列具有一定相似度时。当发生这种内部重组时,同源重组交换发生在预期靶区域内部,并且完整核酸插入物不会掺入基因组靶基因座中。此外,一般认为,随着在双链断裂与突变或插入位点之间的距离增加(例如,超过100bp或200bp),HR介导的插入效率有所降低。参见Beumer et al. (2013) *Genes\Genomes\Genetics* 3:657-664 (Beumer等人,2013年,《基因\基因组学\遗传学》,第3卷,第657-664页);Elliott et al. (1998) *Mol.Cell.Biol.* 18:93-101 (Elliott等人,1998年,《分子细胞生物学》,第18卷,第93-101页);以及Byrne et al. (2015) *Nucleic Acids Research* 43(3):e21 (Byrne等人,2015年,《核酸研究》,第43卷,第3期,第e21页),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。

[0211] 为了使得靶向基因修饰能够在靶基因组基因座处形成大缺失并同时插入一大段外来DNA,需要形成双 ω 结构作为重组中间体。修饰程度越大,结构稳定性越低。在除1细胞期胚胎之外的细胞类型中,可使用总和为10kb或总同源性更大的LTVEC。同源臂的总和为10kb或总同源性更大的LTVEC提高了双 ω 重组中间体的稳定性,从而有利于核酸酶介导的大缺失以及同时被大核酸插入物置换,并且进一步不仅使得邻近同源区的双链断裂发生以增强靶向效率,而且还使得远离同源区的双链断裂发生以增强靶向效率。

[0212] 对于涉及极大程度人源化的基因修饰,将靶向载体与由两个gRNA导向的CRISPR/Cas9核酸酶系统结合可进一步增强靶向效率,使其超过用一个gRNA所达到的效率(参见实例1中的Lrp5、C5(Hc)、Ror1和Trpa1人源化)。就这一点而言,使用两个gRNA可产生意想不到的结果。与用一个gRNA靶向(其产生双等位基因修饰的频率低或根本不产生)相比,用两个gRNA靶向导致产生纯合靶向的细胞、纯合缺失的细胞以及复合杂合靶向的细胞(包括半合靶向的细胞)的速率显著提高。

[0213] 用于在单次靶向实验中产生三种等位基因类型(即,纯合靶向、纯合缺失和复合杂合靶向(尤其是半合靶向))的方法为靶向基因修饰提供了新的可能性和增强的效率。对于简单的基因修饰,诸如小鼠ES细胞中基因的靶向缺失以及其被编码报告基因表达的蛋白(例如, β -半乳糖苷酶或荧光蛋白)的序列置换,将靶向载体与由两个gRNA导向的CRISPR/Cas9系统结合促进了杂合靶向的ES细胞的产生,然后可通过 **VelociMouse[®]** 方法将这些细胞用于产生完全ES细胞衍生的F0代小鼠。参见Poueymirou et al. (2007) *Nat.Biotech.* 25: 91-99 (Poueymirou等人,2007年,《自然生物技术》,第25卷,第91-99页),该文献出于所有目的全文以引用方式并入本文。这些小鼠可用于利用报告基因敲入等位基因来研究组织特异性基因表达。可将在同一实验中产生的纯合靶向的ES细胞克隆转化成具有纯合靶向的基因缺失的VelociMice,然后可研究其基因敲除的表型结果以及受体中的基因表达。由具有靶基因的纯合CRISPR诱导性缺失的ES细胞产生VelociMice使得能够验证纯合靶向的小鼠中所观察到的敲除表型,并且可揭示纯粹的基因缺失与伴有受体插入和药物选择盒的缺失之间的表型差异。同时携带靶向缺失-插入等位基因和CRISPR诱导性缺失的复合杂合(尤其是半合子)ES细胞克隆使得产生VelociMice用于研究的机会与产生来源于纯合靶向的克隆和纯合缺失的克隆的机会相同。此外,可使这些小鼠交配,以由单个ES细胞克隆同时形成靶向缺失和简单缺失突变体小鼠品系。

[0214] 当扩展到人源化的情况时,这些优势更具有价值。人源化小鼠基因的一种重要用途是形成用于测试人特异性治疗剂的动物模型。为了使入源化成为有效的模型,必须除去小鼠基因或使之失活以避免小鼠与人类基因产物之间发生相互作用,这些相互作用可能损害生物功能或与药物的合适相互作用。同时,人类基因必须能够取代其小鼠对等物的生物功能。这些要求可通过将由两个gRNA导向的Cas9核酸酶与靶向载体结合来进行测试,其中该靶向载体被设计成使小鼠基因缺失并同时被人类基因置换。可将来源于具有纯合靶向人源化的ES细胞的VelociMice与来源于具有小鼠基因的纯合CRISPR诱导性缺失的ES细胞的VelociMice进行比较。如果敲除缺失引起可观察到的突变体表型且人源化小鼠不表达这种表型而是正常的,则人类基因能够取代小鼠基因的生物功能。纯合人源化小鼠或携带人源化等位基因和CRISPR诱导性缺失等位基因的复合杂合(例如,半合子)组合的人源化小鼠可用作动物模型来研究人特异性治疗剂的作用机制和功效。复合杂合(例如,半合子)VelociMice也可用于通过常规交配产生小鼠的人源化和缺失敲除品系。因此,从将两个gRNA CRISPR系统与靶向载体结合的单次基因靶向实验中,产生了经遗传修饰的小鼠,从而得到用于治疗剂的临床前测试的重要小鼠模型以及用于研究人药物靶标的小鼠同源物的生物功能的敲除品系。

[0215] (1)产生、提高或增加双等位基因修饰频率的方法

[0216] 本文提供了对细胞内基因组进行双等位基因修饰的方法,或者提高或增加对细胞内基因组进行双等位基因修饰的频率的方法。这类方法可导致例如基因组剪切以移除随后重组的基因组DNA的两条序列之间的大部分基因组DNA。这类方法还可导致核酸插入物的插入或大部分基因组DNA的缺失以及其被核酸插入物置换。

[0217] 本文所提供的用于修饰细胞内基因组的方法包括使基因组与第一Cas蛋白、与基因组靶基因座内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA、与基因组靶基因座内的第二CRISPR RNA识别序列杂交的第二CRISPR RNA以及tracrRNA接触。任选地,可使基因组进一步和与基因组靶基因座内的CRISPR RNA识别序列杂交的额外CRISPR RNA接触,所述额外CRISPR RNA为诸如与基因组靶基因座内的第三CRISPR RNA识别序列杂交的第三CRISPR RNA和/或与基因组靶基因座内的第四CRISPR RNA识别序列杂交的第四CRISPR RNA。双等位基因修饰可通过使基因组与第一Cas蛋白、与基因组靶基因座内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA、与基因组靶基因座内的第二CRISPR RNA识别序列杂交的第二CRISPR RNA以及tracrRNA接触来产生。如下文进一步详细描述,可以任何形式并通过任何方式将Cas蛋白、CRISPR RNA和tracrRNA引入细胞中。同样,可同时或以任何组合按顺序引入Cas蛋白、CRISPR RNA和tracrRNA中的全部或一些。基因组的接触可直接(即,组分直接接触基因组本身)或间接(即,组分与直接接触基因组的其他组分相互作用)进行。

[0218] 基因组可包含一对第一同源染色体和第二同源染色体,这对染色体包含基因组靶基因座。第一Cas蛋白可在第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的一者或两者内(即,在第一CRISPR RNA识别序列内的第一切割位点处和/或在第二CRISPR RNA识别序列内的第二切割位点处)切割这些染色体中的一者或两者。如果还使用第三CRISPR RNA和/或第四CRISPR RNA,则第一Cas蛋白可在第三CRISPR RNA识别序列和/或第四CRISPR RNA识别序列中的一者或两者内(即,在第三CRISPR RNA识别序列内的第三切割位点处和/或在第四CRISPR RNA识别序列内的第四切割位点处)切割这些染色体中的一者或两者。然

后切割事件可在这些染色体的一者或两者中产生至少一处双链断裂。切割事件还可在这些染色体的一者或两者中产生至少两处双链断裂。如果使用Cas切口酶,则切割事件可在这些染色体的一者或两者中产生至少一处单链断裂,或在这些染色体的一者或两者中产生至少两处单链断裂。如果使用第三CRISPR RNA和/或第四CRISPR RNA,则切割事件可在这些染色体的一者或两者中产生至少三处或四处单链或双链断裂。通过双链断裂产生的末端序列随后可发生重组,或者通过单链断裂产生的末端序列随后可发生重组。然后可对具有包含双等位基因修饰的经修饰基因组的细胞进行鉴定。

[0219] 例如,第一Cas蛋白可在第一同源染色体和第二同源染色体中的第一CRISPR RNA识别序列内的第一切割位点处并在至少第一同源染色体中的第二CRISPR RNA识别序列内的第二切割位点处切割基因组,从而在第一同源染色体和第二同源染色体中产生末端序列。末端序列随后可发生重组,以形成具有包含靶向修饰的双等位基因修饰的基因组。靶向修饰可包括至少第一染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失。

[0220] 第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可在基因组靶基因座内任何位置。第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可在任何所关注的基因组区域两侧。例如,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可在基因诸如Lrp5基因座、C5 (Hc) 基因座、Ror1基因座或Trpa1基因座的所有或部分编码序列两侧。第一CRISPR RNA识别序列和第二CRISPR RNA识别序列还可在Cmah基因的所有或部分编码序列两侧。或者,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可在非编码序列(诸如,调控元件(例如启动子),或者编码序列和非编码序列两者)两侧。第三CRISPR RNA识别序列和第四CRISPR RNA识别序列可例如在侧接第一CRISPR RNA识别序列和第二CRISPR RNA识别序列的所关注的基因组区域内的任何位置。

[0221] 例如,第三CRISPR RNA识别序列可邻近第一CRISPR RNA识别序列,并且第四CRISPR RNA识别序列可邻近第二CRISPR RNA识别序列。因此,第一CRISPR RNA识别序列和第三CRISPR RNA识别序列可为第一对CRISPR RNA识别序列,并且第二CRISPR RNA识别序列和第四CRISPR RNA识别序列可为第二对CRISPR RNA识别序列。例如,第一CRISPR RNA识别序列和第三CRISPR RNA识别序列(和/或第二CRISPR RNA识别序列和第四CRISPR RNA识别序列)可相隔约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约1.5kb、约1.5kb至约2kb、约2kb至约2.5kb、约2.5kb至约3kb、约3kb至约3.5kb、约3.5kb至约4kb、约4kb至约4.5kb、约4.5kb至约5kb、约5kb至约6kb、约6kb至约7kb、约7kb至约8kb、约8kb至约9kb或约9kb至约10kb。例如,第一CRISPR RNA识别序列和第三CRISPR RNA识别序列(和/或第二CRISPR RNA识别序列和第四CRISPR RNA识别序列)可相隔约100bp至约1kb。或者,第一CRISPR RNA识别序列和第三CRISPR RNA识别序列(和/或第二CRISPR RNA识别序列和第四CRISPR RNA识别序列)可重叠。

[0222] 第一对CRISPR RNA识别序列可位于靠近基因组靶基因座的5'端的位置,并且第二对CRISPR RNA识别序列可位于靠近基因组靶基因座的3'端的位置。或者,第一对和第二对

均可位于靠近基因组靶基因座的5'端的位置或均可位于靠近靶基因座的3'端的位置。或者,一对或两对可位于基因组靶基因座内部。例如,第一CRISPR RNA识别序列或第一对CRISPR RNA识别序列距基因组靶基因座的5'端可小于25bp、小于50bp、小于100bp、小于150bp、小于200bp、小于250bp、小于300bp、小于350bp、小于400bp、小于450bp、小于500bp、小于600bp、小于700bp、小于800bp、小于900bp、小于1kb、小于2kb、小于3kb、小于4kb、小于5kb或小于10kb。同样,第二CRISPR RNA识别序列或第一对CRISPR RNA识别序列距基因组靶基因座的3'端可小于25bp、小于50bp、小于100bp、小于150bp、小于200bp、小于250bp、小于300bp、小于350bp、小于400bp、小于450bp、小于500bp、小于600bp、小于700bp、小于800bp、小于900bp、小于1kb、小于2kb、小于3kb、小于4kb、小于5kb或小于10kb。

[0223] 或者,第一CRISPR RNA识别序列或第一对CRISPR RNA识别序列距基因组靶基因座的5'端可为例如至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、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。同样,第二CRISPR RNA识别序列或第一对CRISPR RNA识别序列距基因组靶基因座的3'端可为例如至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、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。

[0224] 或者,第一CRISPR RNA识别序列或第一对CRISPR RNA识别序列距基因组靶基因座的5'端可为例如约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb至约5kb、约5kb至约6kb、约6kb至约7kb、约7kb至约8kb、约8kb至约9kb、约9kb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或约90kb至约100kb。同样,第二CRISPR RNA识别序列或第一对CRISPR RNA识别序列距基因组靶基因座的3'端可为例如约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb至约5kb、约5kb至约6kb、约6kb至约7kb、约7kb至约8kb、约8kb至约9kb、约9kb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或约90kb至约100kb。

[0225] 第一切割位点和第二切割位点或第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可相隔例如约1kb至约5kb、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。第一切割位点和第二切割位点或第一CRISPR RNA识别序列和第二CRISPR RNA识别序列也可相隔例如约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。第一切割位点和第二切割位点或第一CRISPR RNA识别序列和第二CRISPR RNA识别序列也可相隔例如约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp或约900bp至约1kb。同样,第一对CRISPR RNA识别序列与第二对CRISPR RNA识别序列可相隔例如约25bp至约50bp、约50bp至约100bp、约100bp至约150bp、约150bp至约200bp、约200bp至约250bp、约250bp至约300bp、约300bp至约350bp、约350bp至约400bp、约400bp至约450bp、约450bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约5kb、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1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约3Mb、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。

[0226] 或者,第一切割位点和第二切割位点或第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可相隔例如至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb或更远。同样,第一对CRISPR RNA识别序列与第二对CRISPR RNA识别序列可相隔例如至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb或更远。

[0227] 或者,第一切割位点和第二切割位点或第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可相隔小于25bp、小于50bp、小于100bp、小于150bp、小于200bp、小于250bp、小于300bp、小于350bp、小于400bp、小于450bp、小于500bp、小于600bp、小于700bp、小于800bp、小于900bp、小于1kb、小于2kb、小于3kb、小于4kb、小于5kb或小于10kb。同样,第一对CRISPR RNA识别序列与第二对CRISPR RNA识别序列可相隔例如小于25bp、小于50bp、小于100bp、小于150bp、小于200bp、小于250bp、小于300bp、小于350bp、小于400bp、小于450bp、小于500bp、小于600bp、小于700bp、小于800bp、小于900bp、小于1kb、小于2kb、小于3kb、小于4kb、小于5kb或小于10kb。

[0228] 通过在第一切割位点和/或第二切割位点处切割基因组而产生的末端序列可为平头末端或交错末端,并且第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失可包括第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的所有或部分核酸序列(包括第一CRISPR RNA识别序列和第二CRISPR RNA识别序列)。同样,通过在第三切割位点和/或第四切割位点处切割基因组而产生的末端序列可为平头末端或交错末端。例如,缺失

可包括第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的仅一部分核酸序列和/或仅一部分第一CRISPR RNA识别序列和/或第二CRISPR RNA识别序列。或者,第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失可包括在第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的所有核酸序列。同样,缺失可包括第一CRISPR RNA识别序列和/或第二CRISPR RNA识别序列或其部分。在一些情况下,缺失还包括位于第一CRISPR RNA识别序列和第二CRISPR RNA识别序列外部的序列(即,不包括且不在第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的序列)。

[0229] 在第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失可为任意长度。例如,所缺失核酸可为约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。

[0230] 或者,所缺失核酸可为例如至少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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb或更大。在一些情况下,所缺失核酸可为至少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10Mb、至少20Mb、至少30Mb、至少40Mb、至少50Mb、至少60Mb、至少70Mb、至少80Mb、至少90Mb或至少100Mb(例如大部分染色体)。

[0231] 第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失可为精确缺失,其中所缺失核酸由仅在第一Cas蛋白切割位点和第二Cas蛋白切割位点之间的核酸序列组成,使得在经修饰的基因组靶基因座处不存在另外的缺失或插入。第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失也可为延伸到第一Cas蛋白切割位点和第二Cas蛋白切割位点之外的不精确缺失(符合通过非同源末端连接(NHEJ)进行的不精确修复),从而导致在经修饰的基因组基因座处产生另外的缺失和/或插入。例如,该缺失可延伸超过第一Cas蛋白切割位点和第二Cas蛋白切割位点之外约1bp、约2bp、约3bp、约4bp、约5bp、约10bp、约20bp、约30bp、约40bp、约50bp、约100bp、约200bp、约300bp、约400bp、约500bp或更远。同样,经修饰的基因组基因座可包含符合通过NHEJ进行的不精确修复的另外插入,诸如约1bp、约2bp、约3bp、约4bp、约5bp、约10bp、约20bp、约30bp、约40bp、约50bp、约100bp、约200bp、约300bp、约400bp、约500bp或更大的插入。

[0232] 接触过程可在不存在外源供体序列或存在外源供体序列的情况下进行,前提条件是如果细胞为1细胞期胚胎,则外源供体序列的长度不超过5kb。外源分子或序列包括非正常存在于细胞中的分子或序列。正常存在包括相对于细胞的特定发育阶段和环境条件而存在。外源分子或序列例如可包括细胞内的相应内源性序列的突变形式,诸如内源性序列的人源化形式。相比之下,内源性分子或序列包括在特定环境条件下正常存在于特定发育阶段的特定细胞中的分子或序列。

[0233] 外源供体序列可在靶向载体内并可包含侧接与基因组内的5'和3'靶序列相对应的5'和3'同源臂的核酸插入物,前提条件是如果细胞为1细胞期胚胎,则靶向载体的长度不

超过5kb。在除1细胞期胚胎之外的细胞类型中,靶向载体可能更长。在除1细胞期胚胎之外的细胞类型中,靶向载体可为例如本文所述的大靶向载体(LTVEC),并且可为至少10kb。因此,在一些方法中,使基因组进一步与靶向载体接触,并且核酸插入物在5'和3'靶序列之间插入。

[0234] 或者,外源供体序列可包括不含核酸插入物的5'和3'同源臂。这类不含核酸插入物的靶向载体可有利于基因组内的5'和3'靶序列之间的精确缺失。这类精确缺失可为例如至少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、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或至少2Mb或更长。

[0235] 在一些这类方法中,5'和3'同源臂对应于在包含第一CRISPR RNA的第一CRISPR RNA识别序列和/或第二CRISPR RNA的第二CRISPR RNA识别序列的基因组靶基因座处的5'和3'靶序列。第一CRISPR RNA识别序列和第二CRISPR RNA识别序列或者第一切割位点和第二切割位点可邻近5'靶序列,邻近3'靶序列,或者既不邻近5'靶序列也不邻近3'靶序列。或者,第一CRISPR RNA识别序列或第一切割位点可邻近5'靶序列,并且第二CRISPR RNA识别序列或第二切割位点可邻近3'靶序列。或者,第一CRISPR RNA识别序列或第一切割位点可邻近5'靶序列或3'靶序列,并且第二CRISPR RNA识别序列或第二切割位点可既不邻近5'靶序列也不邻近3'靶序列。

[0236] 例如,第一CRISPR RNA识别序列和/或第二CRISPR RNA识别序列可位于5'和3'靶序列之间,或者可邻近或靠近5'靶序列和/或3'靶序列,诸如在5'和/或3'靶序列的1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb、150kb、160kb、170kb、180kb、190kb、200kb、250kb、300kb、350kb、400kb、450kb或500kb内。同样,第一切割位点和/或第二切割位点可位于5'和3'靶序列之间或可邻近或靠近5'靶序列和/或3'靶序列,诸如在5'和/或3'靶序列的1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb、150kb、160kb、170kb、180kb、190kb、200kb、250kb、300kb、350kb、400kb、450kb或500kb内。例如,第一CRISPR RNA识别序列或第一切割位点可在5'靶序列或5'和3'靶序列两者的50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb、150kb、160kb、170kb、180kb、190kb、200kb、250kb、300kb、350kb、400kb、450kb或500kb内。同样,第二CRISPR RNA识别序列或第二切割位点可在3'靶序列或5'和3'靶序列两者的50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb、150kb、160kb、170kb、180kb、190kb、200kb、250kb、300kb、350kb、400kb、450kb或500kb内。

[0237] 或者,第一CRISPR RNA识别序列和/或第二CRISPR RNA识别序列可位于距5'和/或3'靶序列至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp、至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、至少

6kb、至少7kb、至少8kb、至少9kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb处。同样,第一切割位点和/或第二切割位点可位于距5' 和/或3' 靶序列至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp、至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、至少6kb、至少7kb、至少8kb、至少9kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb处。例如,第一CRISPR RNA识别序列或第一切割位点可位于距5' 靶序列或距5' 和3' 靶序列两者至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp、至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、至少6kb、至少7kb、至少8kb、至少9kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb处。同样,第二切割位点的第二CRISPR RNA识别序列可位于距3' 靶序列或距5' 和3' 靶序列两者至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少900bp、至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、至少6kb、至少7kb、至少8kb、至少9kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb处。

[0238] 例如,第一CRISPR RNA识别序列和/或第二CRISPR RNA识别序列可位于距5' 和/或3' 靶序列约50bp至约100bp、约200bp至约300bp、约300bp至约400bp、约400bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb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100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约300kb、约300kb至约400kb或约400kb至约500kb处。同样,第一切割位点和/或第二切割位点可位于距5' 和/或3' 靶序列约50bp至约100bp、约200bp至约300bp、约300bp至约400bp、约400bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb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100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约300kb、约300kb至约400kb或约400kb至约500kb处。例如,第一CRISPR RNA识别序列或第一切割位点可位于距5' 靶序列或距5' 和3' 靶序列两者约50bp至约100bp、约200bp至约300bp、约300bp至约400bp、约400bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约

2kb、约2kb至约3kb、约3kb至约4kb、约4kb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100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约300kb、约300kb至约400kb或约400kb至约500kb处。同样，第二CRISPR RNA识别序列或第二切割位点可位于距3'靶序列或距5'和3'靶序列两者约50bp至约100bp、约200bp至约300bp、约300bp至约400bp、约400bp至约500bp、约500bp至约600bp、约600bp至约700bp、约700bp至约800bp、约800bp至约900bp、约900bp至约1kb、约1kb至约2kb、约2kb至约3kb、约3kb至约4kb、约4kb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100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约300kb、约300kb至约400kb或约400kb至约500kb处。

[0239] 或者，第一CRISPR RNA识别序列和/或第二CRISPR RNA识别序列或者第一切割位点和/或第二切割位点可位于距5'和/或3'靶序列大于50bp、大于100bp、大于200bp、大于300bp、大于400bp、大于500bp、大于600bp、大于700bp、大于800bp、大于900bp、大于1kb、大于2kb、大于3kb、大于4kb、大于5kb、大于6kb、大于7kb、大于8kb、大于9kb、大于10kb、大于20kb、大于30kb、大于40kb、大于50kb、大于60kb、大于70kb、大于80kb、大于90kb或大于100kb处。例如，第一CRISPR RNA识别序列或第一切割位点可位于距5'靶序列或距5'和3'靶序列两者大于50bp、大于100bp、大于200bp、大于300bp、大于400bp、大于500bp、大于600bp、大于700bp、大于800bp、大于900bp、大于1kb、大于2kb、大于3kb、大于4kb、大于5kb、大于6kb、大于7kb、大于8kb、大于9kb、大于10kb、大于20kb、大于30kb、大于40kb、大于50kb、大于60kb、大于70kb、大于80kb、大于90kb或大于100kb处。同样，第二CRISPR RNA识别序列或第二切割位点可位于距3'靶序列或距5'和3'靶序列两者大于50bp、大于100bp、大于200bp、大于300bp、大于400bp、大于500bp、大于600bp、大于700bp、大于800bp、大于900bp、大于1kb、大于2kb、大于3kb、大于4kb、大于5kb、大于6kb、大于7kb、大于8kb、大于9kb、大于10kb、大于20kb、大于30kb、大于40kb、大于50kb、大于60kb、大于70kb、大于80kb、大于90kb或大于100kb处。

[0240] 本文所述的方法提高并增加了双等位基因修饰的频率。具体地讲，与使基因组单独与第一CRISPR RNA或第二CRISPR RNA接触相比，通过使基因组同时与第一CRISPR RNA和第二CRISPR RNA接触，可增加产生双等位基因修饰的效率。通过使基因组与第一CRISPR RNA、第二CRISPR RNA和第三CRISPR RNA或者第一CRISPR RNA、第二CRISPR RNA、第三CRISPR RNA和第四CRISPR RNA接触也可增加产生双等位基因修饰的效率。双等位基因修饰包括这样的事件：其中对相应同源染色体（例如，二倍体细胞中的）上的相同基因座进行相同修饰，或其中对相应同源染色体上的相同基因座进行不同修饰。同源染色体包括在相同基因座处的基因相同但等位基因可能不同的染色体（例如，在减数分裂期间配对的染色体）。术语“等位基因”包括一种或多种替代形式中任一种形式的基因序列。在二倍体细胞或生物体中，给定序列的两个等位基因通常占据一对同源染色体上的相应基因座。

[0241] 双等位基因修饰可导致靶向修饰的纯合性或靶向修饰的复合杂合性（例如，半合性）。对一组细胞进行单次靶向实验可产生对靶向修饰（例如，基因座的人源化）而言为纯合的细胞、对该靶向修饰而言为复合杂合的细胞（包括对该靶向修饰而言为半合子的细胞），以及在第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间纯合剪切（即，在两个CRISPR RNA识别序列之间缺失大核酸序列）的细胞。纯合性包括这样的情况：其中靶基因座

的两个等位基因(即,两条同源染色体上的对应等位基因)均具有靶向修饰。复合杂合性包括这样的情况:其中靶基因座的两个等位基因(即,两条同源染色体上的等位基因)都已经被修饰,但它们以不同的方式被修饰(例如,对一个等位基因进行靶向修饰,而使另一个等位基因失活或受到破坏)。内源性核酸序列的破坏可例如在用非同源末端连接(NHEJ)介导的DNA修复来修复由Cas蛋白产生的双链断裂时,导致该修复产生包含核酸序列的插入或缺失的突变体等位基因,从而导致对该基因组基因座的破坏。破坏的示例包括改变调控元件(例如启动子或增强子)、错义突变、无义突变、移码突变、截短突变、无效突变,或少量核苷酸的插入或缺失(例如,引起移码突变)。破坏可导致失活(即,功能丧失)或等位基因的丧失。

[0242] 例如,如果细胞具有经靶向修饰的一个等位基因并具有不能被表达或不能以其他方式起作用的另一个等位基因,则双等位基因修饰可导致复合杂合性。复合杂合性包括半合性。半合性包括这样的情况:仅存在靶基因座的一个等位基因(即,两条同源染色体之一上的等位基因)。例如,如果靶向修饰发生在一个等位基因中,而另一个等位基因具有相应的丢失或缺失,则双等位基因修饰可导致靶向修饰的半合性。

[0243] 在特定示例中,双等位基因修饰可包括在第一同源染色体和第二同源染色体对中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失。这种缺失可同时发生,或者这种缺失可最初在第一同源染色体中发生,其中通过细胞使用第一同源染色体作为供体序列来经由同源重组(诸如通过基因转换)修复第二同源染色体中的一处或多处双链断裂而实现纯合性。第一同源染色体和第二同源染色体中的所缺失核酸序列可相同、部分重叠或不同。或者,双等位基因修饰可包括第一同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失以及第二同源染色体中的相应等位基因或基因座的丢失。或者,双等位基因修饰可包括第一同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失以及第二同源染色体中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的相应等位基因或基因座的失活或破坏。

[0244] 如果使用供体序列,则双等位基因修饰可包括第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失以及核酸插入物在第一同源染色体和第二同源染色体对中的5'和3'靶序列之间的插入,从而导致纯合的经修饰基因组。所述缺失和插入可在两条染色体中同时发生,或者所述缺失和插入可最初在第一同源染色体中发生,其中通过细胞使用第一同源染色体作为供体序列来经由同源重组(诸如通过基因转换)修复第二同源染色体中的双链断裂而实现纯合性。例如,不希望受任何特定理论的束缚,核酸插入物的插入可发生在第一同源染色体中(存在或不存在由Cas蛋白进行切割的情况),并且第二同源染色体可随后通过基因转换事件被修饰,该事件通过由第二同源染色体上的Cas蛋白进行的切割刺激。

[0245] 或者,双等位基因修饰可导致复合杂合的经修饰基因组。例如,靶向修饰可包括第一同源染色体和第二同源染色体两者中的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失以及核酸插入物在第一同源染色体中而非在第二同源染色体中的插入。或者,靶向修饰可包括第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失,以及核酸插入物在第一同源染色体中的插入和第二同源染色体中的相应等位基因或基因座的失活或破坏。或者,双等位基因修饰可导致半合子的经修饰基因组,其中靶向修饰可包

括第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失以及核酸插入物在第一同源染色体中的插入和第二同源染色体中的相应等位基因或基因座的丢失或缺失。

[0246] 纯合和复合杂合的(尤其是半合子的)靶向基因修饰是有利的,因为利用这些修饰产生经遗传修饰的动物的过程(在下文中更详细地描述)可能更加高效且耗时更少。在许多情况下(诸如移除基因以研究其缺失的影响),靶向遗传修饰的纯粹杂合性(即,对一个等位基因进行修饰但不改变另一个等位基因)是不够的。采用常规靶向策略,可能会获得对大靶向基因组缺失而言为杂合的F0代动物,但需要这些杂合的动物进行后续相互交配才能产生对该缺失而言为纯合的F1代动物。这些额外的交配步骤不仅成本高昂,而且耗时。能够形成对靶向基因修饰而言为纯合的或复合杂合的(尤其是半合子的)F0代经遗传修饰的动物,显著提高了效率并节省了时间,因为需要的交配步骤更少。

[0247] (2) 基因转换或杂合性丢失

[0248] 在一些方法中,待修饰的基因组在对第一等位基因而言为杂合的细胞内,并且对该基因进行修饰以使其变得对第一等位基因而言为纯合。术语“杂合的”包括这样的情况:其中基因组在一个或多个相应染色体基因座处包含不同的等位基因(例如,在同源染色体上相应基因座处包含不同的等位基因)。术语“纯合的”包括这样的情况:其中基因组在相应染色体基因座处(例如,在相应同源染色体上)包含相同的等位基因。在一些这类方法中,可通过细胞使用第一等位基因作为供体序列来经由同源重组(诸如基因转换)修复相应第二等位基因中的双链断裂而实现纯合性。通常,基因转换的程度限于数百个碱基对。参见例如,Kasperek&Humphrey (2011) *Seminars in Cell&Dev. Biol.* 22:886-897 (Kasperek&Humphrey等人,2011年,《细胞与发育生物学研讨文辑》,第22卷,第886-897页),该文献出于所有目的全文以引用方式并入本文。但是,使用在单个基因座内不同切割位点处引导切割的成对向导RNA可提高并增强更长段上的基因转换能力。

[0249] 这类方法可用于多种情况。第一等位基因可包含突变。在一些方法中,例如第一等位基因包含期望的靶向基因修饰。如果例如目标是形成对靶向遗传修饰而言为纯合的非人类动物,则实现该修饰的纯合性可节省大量的时间和成本。在其他方法中,第一等位基因是基因的野生型等位基因,其对应于该基因的第二致病等位基因。或者,第二等位基因可包含任何突变。然后可使用这些方法,实现利用野生型等位基因在其天然染色体基因座处置换致病等位基因这一最终基因治疗目标。

[0250] 在用于修饰对第一等位基因而言为杂合的基因组以使其变得对第一等位基因而言为纯合的一些此类方法中,使基因组与Cas蛋白、tracrRNA、与第二等位基因内的第一CRISPR RNA识别序列杂交的第一CRISPR RNA以及与第二等位基因内的第二CRISPR RNA识别序列杂交的第二CRISPR RNA接触,其中第一等位基因在第一同源染色体上,第二等位基因在第二同源染色体上的相应基因座处(即,第一等位基因和第二等位基因可为第一同源染色体和第二同源染色体对中的相应等位基因)。任选地,可使基因组和与第二等位基因内的CRISPR RNA识别序列杂交的额外CRISPR RNA(例如,第三CRISPR RNA,或者第三CRISPR RNA和第四CRISPR RNA)接触。Cas蛋白可切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的一者或两者(即,在第一CRISPR RNA识别序列内的第一切割位点处和/或在第二CRISPR RNA识别序列内的第二切割位点处进行切割)。基因组在第一切割位点和/或第二切割位点处进行切割可在基因组DNA中形成平头末端或可形成交错末端。切割位点随后可

通过第一等位基因和第二等位基因之间的重组而被修复,从而导致对第一等位基因而言为纯合的经修饰基因组。然后可对具有经修饰基因组的细胞进行鉴定。

[0251] 在一些方法中,第一CRISPR RNA识别序列和/或第二CRISPR RNA识别序列位于第二等位基因而非第一等位基因内。第一等位基因和/或第二等位基因可为野生型等位基因或可包含靶向修饰或其他与野生型等位基因的偏差。例如,第一等位基因可包含期望的靶向修饰,并且第二等位基因可为野生型等位基因。或者,第一等位基因可为野生型等位基因,并且第二等位基因可包含不期望的修饰,诸如致病突变。在一些此类方法中,发生靶向基因修复或靶向基因校正,使得第二等位基因中的致病突变通过使用第一等位基因作为供体序进行重组而被校正。

[0252] 第一切割位点和第二切割位点或第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可相隔例如约1kb至约5kb、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。

[0253] 或者,第一切割位点和第二切割位点或第一CRISPR RNA识别序列和第二CRISPR RNA识别序列可相隔例如至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。

[0254] 在一些方法中,第一等位基因和第二等位基因之间的序列差异跨越约100bp至约200bp、约200bp至约400bp、约400bp至约600bp、约600bp至约800bp、约800bp至约1kb、约1kb至约2kb、约2kb至约3kb、约4kb至约5kb、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。

[0255] 在其他方法中,第一等位基因和第二等位基因之间的差异跨越至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少600bp、至少700bp、至少800bp、至少800bp、至少1kb、至少2kb、至少3kb、至少4kb、至少5kb、至少6kb、至少7kb、至少8kb、至少9kb、至少10kb、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、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。

[0256] 在其他用于修饰对第一等位基因而言为杂合的细胞中的基因组以使其变得对第一等位基因而言为纯合的这类方法中,使基因组与Cas蛋白、tracrRNA以及与第一CRISPR RNA识别序列杂交的第一非等位基因特异性CRISPR RNA接触。第一等位基因在第一同源染色体上,并且CRISPR RNA识别序列位于与第二同源染色体上的第一等位基因相对应的基因座的近着丝粒侧(即,更靠近着丝粒)。Cas蛋白可切割第一CRISPR RNA识别序列以产生双链断裂。然后可发生重组以修饰细胞,使其变得对第一等位基因而言为纯合。

[0257] 任选地,细胞对一个或多个额外等位基因而言为杂合的,第一CRISPR RNA识别序列位于与第二同源染色体上的一个或多个额外等位基因相对应的基因座的近着丝粒侧,并且重组对细胞进行修饰,使其变得对一个或多个额外等位基因而言为纯合。

[0258] 任选地,该方法还可包括使基因组和与第二CRISPR RNA识别序列杂交的第二非等位基因特异性CRISPR RNA接触,该第二CRISPR RNA识别序列位于与第二同源染色体中的第一等位基因相对应的基因座的近着丝粒侧,其中Cas蛋白切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的至少一者,以产生至少一处双链断裂。任选地,该方法还可包括使基因组和与CRISPR RNA识别序列杂交的额外非等位基因特异性CRISPR RNA(例如,第三CRISPR RNA,或第三CRISPR RNA和第四CRISPR RNA)接触,该CRISPR RNA识别序列位于与第二同源染色体上的第一等位基因相对应的基因座的近着丝粒侧。然后可对具有经修饰基因组的细胞进行鉴定。

[0259] 在一些方法中,第一CRISPR RNA识别序列(或第二CRISPR RNA识别序列、第三CRISPR RNA识别序列或第四CRISPR RNA识别序列)位于第二同源染色体而非第一同源染色体上。第一CRISPR RNA识别位点(或第二CRISPR RNA识别位点、第三CRISPR RNA识别位点或第四CRISPR RNA识别位点)可距着丝粒约100bp至约1kb、约1kb至约10kb、约10kb至约100kb、约100kb至约1Mb、约1Mb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。

[0260] 第一等位基因和/或一个或多个额外等位基因可包含突变,诸如靶向修饰。或者,第一等位基因和/或一个或多个额外等位基因可为野生型等位基因,并且第二同源染色体上的相应基因座可包含突变,诸如致病突变。第一等位基因可距第一CRISPR RNA识别位点约100bp至约1kb、约1kb至约10kb、约10kb至约100kb、约100kb至约1Mb、约1Mb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。或者,第一等位基因可距第一CRISPR RNA识别位点至少100bp、至少1kb、至少10kb、至少100kb、至少1Mb、至少10Mb、至少20Mb、至少30Mb、至少40Mb、至少50Mb、至少60Mb、至少70Mb、至少80Mb、至少90Mb或至少100Mb或更远。

[0261] Cas蛋白可以是Cas9。其在双链DNA的两条链上可具有核酸酶活性,或其可以是切口酶。在一些方法中,Cas蛋白和第一CRISPR RNA在天然情况下不会同时存在。

[0262] 所述重组可包括双链断裂的端粒侧(即,朝向端粒)杂合性丢失(例如,极性或定向基因转换或杂合性丢失)。被杂合性丢失置换的第二同源染色体区域可为约100bp至约1kb、约1kb至约10kb、约10kb至约100kb、约100kb至约1Mb、约1Mb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。或者,被置换的第二同源染色体区域可为至少100bp、至少1kb、至少10kb、至少100kb、至少1Mb、至少10Mb、至少20Mb、至少30Mb、至少40Mb、至少50Mb、至少60Mb、至少70Mb、至少80Mb、至少90Mb或至少100Mb或更大。例如,大部分染色体可被置换。

[0263] B. 制备经遗传修饰的非人动物的方法

[0264] 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) 使用上述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非人哺乳动物ES细胞(例如小鼠ES细胞或大鼠ES细胞)。参见例如,美国专利No.7,294,754,该专利出于所有目的全文以引用方式并入本文。

[0265] 另选地,产生经遗传修饰的非人动物的方法可包括：(1) 使用上述方法修饰1-细胞期胚胎的基因组；(2) 选择经遗传修饰的胚胎；以及(3) 将经遗传修饰的胚胎植入到代孕母体中。产生能够通过生殖系传递遗传修饰的子代。

[0266] 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普通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普通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No.7,612,250,这些专利中的每一者均出于所有目的全文以引用的方式并入本文。

[0267] 制备经遗传修饰的非人动物的一些方法包括产生F0代非人动物的方法。此类方法可包括使非人ES细胞中的基因组与Cas蛋白、与第一CRISPR RNA识别序列杂交的第一CRISPR RNA、与第二CRISPR RNA识别序列杂交的第二CRISPR RNA和tracrRNA接触。Cas蛋白可切割第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的基因组以产生末端序列。末端序列可经历重组以形成带有靶向修饰的基因组,并且靶向修饰可包括第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间的缺失。

[0268] 该方法还可包括：(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。

[0269] 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,F0动物体中具有有关

注的核苷酸序列和不含重组酶盒和/或选择盒的细胞数量将会改变。经由例如 **VELOCIMOUSE**[®] 方法将供体ES细胞从对应生物体(例如8-细胞期小鼠胚胎)引入到桑椹胚前期胚胎中允许更大百分比的F0动物的细胞群包含具有所关注的核苷酸序列的细胞,所关注的核苷酸序列包含靶向遗传修饰。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%、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。

[0270] 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。

[0271] 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代的动物可以彼此交配,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。

[0272] C. 基因组和靶基因组基因座

[0273] 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。

[0274] 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。

[0275] D. Cas9和向导RNA的形式

[0276] 在一些方法中,基因组接触包括将一个或多个Cas蛋白、一个或多个CRISPR RNA和一个或多个tracrRNA引入到细胞中。可通过任何手段来完成引入,并且组分中的一种或多种(例如组分中的两种,或组分中的所有)可被同时或以任何组合依次引入到细胞中。

[0277] CRISPR RNA和tracrRNA可融合在一起作为待引入到细胞中的向导RNA(gRNA)。另选地,CRISPR RNA和tracrRNA可以是不同的RNA分子。CRISPR RNA可以RNA的形式或以编码CRISPR RNA的DNA的形式引入到细胞中。同样地,tracrRNA可以RNA的形式或以编码tracrRNA的DNA的形式引入到细胞中,并且gRNA可以RNA的形式或以编码gRNA的DNA的形式引入到细胞中。

[0278] Cas蛋白可以蛋白、编码Cas蛋白的信使RNA(mRNA)或编码Cas蛋白的DNA的形式引入到细胞中。在一些方法中,Cas蛋白、CRISPR RNA和tracrRNA可作为蛋白-RNA复合物引入到细胞中。同样地,Cas蛋白和gRNA可作为蛋白-RNA复合物引入到细胞中。Cas蛋白可以是细胞可透性Cas蛋白(例如,具有细胞穿透结构域的Cas蛋白)。

[0279] 编码Cas蛋白、CRISPR RNA或tracrRNA的DNA可有效连接至细胞中有活性的启动子。此类DNA可以为一个或多个表达构建体。在一些方法中,此类表达构建体中的一个或多个可以是单一核酸分子的组分。例如,编码一个或多个Cas蛋白的DNA、编码一个或多个CRISPR RNA的DNA和编码一个或多个tracrRNA的DNA都可以是单一核酸分子的组分。另选

地,在两个、三个、四个或更多个核酸分子之间,它们可以任何组合的方式分离。

[0280] 相似地,编码Cas蛋白的DNA或编码gRNA的DNA可有效连接至细胞中有活性的启动子。此类DNA也可以为一个或多个表达构建体。在一些方法中,此类表达构建体中的一个或多个可以是单一核酸分子的组分。例如,编码一个或多个Cas蛋白的DNA和编码一个或多个gRNA的DNA都可以是单一核酸分子的组分。另选地,在两个、三个、四个或更多个核酸分子之间,它们可以任何组合的方式分离。

[0281] 在一些方法中,Cas蛋白和CRISPR RNA和/或tracrRNA在天然情况下不会同时存在。在一些方法中,例如,Cas蛋白和第一CRISPR RNA在天然情况下不会同时存在,Cas蛋白和第二CRISPR RNA在天然情况下不会同时存在,并且/或者Cas蛋白和tracrRNA在天然情况下不会同时存在。

[0282] 在一些方法中,Cas蛋白为Cas9蛋白。Cas蛋白可融合到异源多肽,诸如核定位信号(NLS)。Cas蛋白可具有全切割活性并且可在基因组DNA中形成双链断裂(例如,具有平头末端的双链断裂),或者它可以是能够只切割基因组DNA的链的切口酶。

[0283] 在一些方法中,采用了成对切口酶。例如,基因组可与切割DNA相对链的第一切口酶和第二切口酶接触,由此,通过双切口修饰基因组。第一切口酶可切割基因组DNA的第一链(即,互补链),并且第二切口酶可切割基因组DNA的第二链(即,非互补链)。另选地,这两个切口酶都可切割相同链。可例如通过使第一切口酶的RuvC结构域中的催化残基突变(例如,本文别处所述的D10A突变)以及使第二切口酶的HNH结构域中的催化残基突变(例如,本文别处所述的H840A突变)来形成第一切口酶和第二切口酶。另选地,第一切口酶可用于形成两个切口。

[0284] 在一些此类方法中,可采用双切口以形成具有交错末端的一个或多个双链断裂。例如,采用双切口以在第一切割位点和第二切割位点处形成交错末端。第一切口酶可切割与第一CRISPR RNA和第二CRISPR RNA杂交的第一CRISPR RNA识别序列和第二CRISPR RNA识别序列中的DNA的第一链,并且第二切口酶可切割与第三CRISPR RNA和第四CRISPR RNA杂交的第三CRISPR RNA靶识别序列和第四CRISPR RNA靶识别序列中的DNA的第二链。另选地,第一切口酶可用于使第一CRISPR RNA识别序列、第二CRISPR RNA识别序列、第三CRISPR RNA识别序列和第四CRISPR RNA识别序列产生切口。可定位第一CRISPR RNA靶识别序列和第三CRISPR RNA靶识别序列以形成第一切割位点,使得通过第一切口酶和第二切口酶在DNA的第一链和第二链上形成的切口形成双链断裂(即,第一切割位点包含第一CRISPR RNA识别序列和第三CRISPR RNA识别序列中的切口)。同样地,可定位第二CRISPR RNA识别序列和第四CRISPR RNA识别序列以形成第二切割位点,使得通过第一切口酶和第二切口酶在DNA的第一链和第二链上形成的切口形成双链断裂(即,第二切割位点包含第二CRISPR RNA识别序列和第四CRISPR RNA识别序列中的切口)。在一些情况下,第一CRISPR RNA识别序列和第三CRISPR RNA识别序列和/或第二CRISPR RNA识别序列和第四CRISPR RNA识别序列中的切口可以是偏置切口。偏置窗口可以为,例如,至少约5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp或更多。参见Ran et al. (2013) Cell 154:1380-1389(Ran等人,2013年,《细胞》,第154卷,第1380-1389页);Mali et al. (2013) Nat. Biotech. 31:833-838(Mali等人,2013年,《自然生物技术》,第31卷,第833-838页);以及Shen et al. (2014) Nat. Methods 11:399-404(Shen等人,2014年,《自然方法》,第11卷,第399-404页),上述文献

中的每一篇出于所有目的全文以引用方式并入本文。

[0285] E. 将核酸和蛋白引入到细胞中的方法

[0286] 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分(例如,核酸酶试剂、核酸酶切割位点、插入DNA多核苷酸、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)。

[0287] 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。

[0288] 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掺入到其基因组中的DNA构建体。例如,本文所公开的方法中采用的细胞可具有稳定地掺入到其基因组中的原生Cas-编码基因(即Cas-就绪细胞)。“稳定地掺入”或“稳定地引入”包括将多核苷酸引入到细胞中,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掺入DNA构建体或靶向基因组整合系统的各种组分。

[0289] 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。非限制性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));树枝状体;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, J. (2006) *Current Pharmaceutical Biotechnology* 7, 277-28 (Bertram, J., 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。

[0290] 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。

[0291] 也可通过显微注射完成向细胞(例如1-细胞期胚胎)内引入核酸或蛋白。在1-细胞期胚胎中,显微注射可进入到雌性原核和/或雄性原核中或进入到细胞质中。如果显微注射只进入到一个原核中,那么由于其具有较大尺寸,雄性原核是优选的。mRNA显微注射优选地进入到细胞质中(例如,以直接将mRNA递送到翻译机器中),而Cas蛋白或编码Cas蛋白或编码RNA的核酸分子的显微注射则优选地进入到核/原核中。另选地,可通过向核/原核和细胞质中注射来进行显微注射:可首先向核/原核中引入针,并且可注射第一量,然后当将针从1-细胞期胚胎中移除时,可向细胞质中注射第二量。如果Cas蛋白被注射到细胞质中,那么Cas蛋白优选地包含核定位信号以确保递送到核/原核。用于进行显微注射的方法是众所周

知的。参见例如Nagy等人(Nagy A,Gertsenstein M,Vintersten K,Behringer R.,2003, Manipulating the Mouse Embryo.Cold Spring Harbor,New York:Cold Spring Harbor Laboratory Press(Nagy A,Gertsenstein M,Vintersten K,Behringer R.,2003年,《小鼠胚胎实验操作指南》,美国纽约州冷泉港,冷泉港实验室出版社),Meyer et al.(2010) Pros Natl Acad Sci USA 107:15022-15026(Meyer等人,2010年,《美国国家科学院院刊》,第107卷,第15022-15026页)和Meyer et al.(2012) Proc Natl Acad Sci USA 109:9354-9359(Meyer等人,2012年,《美国国家科学院院刊》,第109卷,第9354-9359页),上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。

[0292] 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。

[0293] 当核酸酶试剂和靶向载体(例如用于细胞而非1-细胞期胚胎的LTVEC)都被引入到细胞中时,它们可被同时引入。另选地,核酸酶试剂可与靶向载体分开引入。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。

[0294] F. 重组机制及用于改变非同源末端连接、基因转换或同源重组的发生率的方法

[0295] 重组包括两个多核苷酸之间遗传信息交换的任何过程并且可通过任何机制发生。响应于双链断裂(DSB)的重组主要通过两条保守的DNA修复途径:非同源末端连接(NHEJ)和同源重组(HR)发生。参见Kasperek&Humphrey(2011) Seminars in Cell&Dev.Biol.22:886-897(Kasperek和Humphrey,2011年,《细胞与发育生物学研讨文辑》,第22卷,第886-897页),该文献出于所有目的全文以引用方式并入本文。NHEJ包括通过在不需要同源模板的情况下将断裂端彼此直接连接来修复核酸中的双链断裂。通过NHEJ连接不连续序列通常可导致双链断裂位点附近的缺失、插入或易位。还可通过同源介导修复(HDR)或同源重组(HR)发生重组。HDR或HR包括可能需要核苷酸序列同源性,使用“供体”分子作为模板来修复“靶”分子(即,经历了双链断裂的分子),并导致将遗传信息从供体转移到靶的核酸修复形式。不希望受任何特定理论的束缚,此类转移可能涉及对在断裂的靶和供体之间形成的异源双链DNA进行错配校正,和/或其中供体用于重新合成将变成靶的一部分的基因信息的合成依赖性链退火,和/或相关过程。在一些情况下,供体多核苷酸,供体多核苷酸的一部分,供体多核苷酸的拷贝,或供体多核苷酸拷贝的一部分整合到靶DNA中。

[0296] 在将对于等位基因而言为杂合的细胞基因组修饰变成对于等位基因而言为纯合的语境中,重组可包括从杂合细胞中衍生出纯合细胞的任何手段。此类手段可包括,例如,杂合性丢失(LOH)、基因转换或通过任何已知重组机制发生的交换事件。不希望受理论的束缚,LOH可例如,在进行或不进行基因转换的情况下经由有丝分裂重组,或者经由染色体丢失和复制发生。参见例如Lefebvre et al.(2001) Nat.Genet.27:257-258(Lefebvre等人,2001年,《自然-遗传学》,第27卷,第257-258页),该文献出于所有目的全文以引用方式并入本文。在此语境中的基因转换可包括遗传物质从供体序列向高度同源受体的单向转移(即遗传信息从一个分子向其同源物的非相互交换)。基因转换包括用于通过任何已知重组机制拷贝等位基因的任何手段。例如,基因转换可涉及遗传信息从一个完整序列向含有双链断裂的同源区域的非相互交换,并且它可在或者相同染色单体或者不同染色体上的姊妹染

色单体之间、同源染色体之间或同源序列之间发生。参见例如Chen et al. (2007) Nat. Rev. Genet. 8:762-775 (Chen等人, 2007年,《自然评论-遗传学》,第8卷,第762-775页), 该文献出于所有目的全文以引用方式并入本文。在特定情况下,由于从同源染色体拷贝遗传信息,所以直接通过同源重组产生基因转换。当同源序列不同时,这可导致局部的杂合性丢失 (LOH)。

[0297] 例如,LOH可通过有丝分裂交换进行的相互染色单体交换,或通过断裂诱导的复制进行的染色单体拷贝发生。在任一种情况下,可发生杂合修饰,其中在基因组复制之前靶向一个染色体。另选地,在基因组复制之后可靶向单个染色单体,随后进行染色单体间基因转换。

[0298] 在本文所公开的方法的任一个中,所述细胞可以是已被修饰为提高或降低NHEJ活性的细胞。同样地,所述细胞可以是已被修饰为提高基因转换或NHEJ活性的细胞。此类修饰可包括参与调控NHEJ、基因转换和/或HDR的基因的表达或活性的修饰。例如,降低NHEJ活性和/或提高HDR活性可促进对应于两个gRNA的CRISPR RNA识别序列之间的基因组区域的双等位基因剪切。不受任何特定理论的束缚,通过其可发生双等位基因基因组剪切的一个机制为在第一等位基因中经受NHEJ介导修复或HDR介导修复,并经由HDR机制 (诸如基因转换) 形成相同的第二等位基因 (参见实例1)。因此,促进HDR介导途径 (例如,通过降低NHEJ活性或通过提高HDR活性) 也可促进基因组区域的双等位基因剪切。相似地,不受任何特定理论的束缚,如果NHEJ活性降低并且HDR活性 (例如基因转换活性) 相应地提高,那么可促进通过使用靶向单个基因座的成对向导RNA来进行的从杂合细胞向纯合细胞的转换。

[0299] 抑制因子可用于提高或降低NHEJ活性,或者提高或降低HDR活性。此类抑制因子可以是,例如,小分子或抑制性核酸诸如短干扰核酸 (例如,短干扰RNA (siRNA)、双链RNA (dsRNA)、微小RNA (miRNA) 和短发夹RNA (shRNA)) 或针对基因转录物具有特异性的反义寡核苷酸。抑制因子可被定向于酶处,这些酶参与NHEJ或HDR,或经由例如磷酸化、泛素化和苏素化的翻译后修饰进行的其上游调控。

[0300] 在哺乳动物细胞中,NHEJ为主要的DSB修复机制,并且活跃在整个细胞周期内。在脊椎动物中,“标准”或“典型”NHEJ途径 (C-NHEJ) 需要若干核心因子,包括DNA-PK、Ku70-80、Artemis、连接酶IV (Lig4)、XRCC4、CLF和Pol μ 以修复DSB。参见Kasperek&Humphrey (2011) Seminars in Cell&Dev. Biol. 22:886-897 (Kasperek和Humphrey, 2011年,《细胞与发育生物学研讨文辑》,第22卷,第886-897页), 该文献出于所有目的全文以引用方式并入本文。在NHEJ期间,通过用作用于负载其他NHEJ组分的系泊部位的高丰度末端保护Ku蛋白结合DNA末端。

[0301] 因此,在本文所公开的方法的一些中,所述细胞已被修饰用于减少或消除或用于增加参与C-NHEJ的因子的表达或活性。例如,在一些方法中,所述细胞已被修饰用于减少或消除DNA-PK、Ku70-80、Artemis、连接酶IV (Lig4)、XRCC4、CLF和/或Pol μ 的表达或活性。在特定方法中,所述细胞已被修饰用于减少或消除DNA-PK的表达或活性,或用于增加DNA-PK的表达或活性 (例如DNA-PKcs的表达或活性;指定为P97313的示例性UniProt序列)。DNA-PKcs抑制因子的例子包括,例如,NU7026和NU7441。参见,例如,美国专利No. 6,974,867,该专利出于所有目的全文以引用方式并入本文。在特定方法中,所述细胞已被修饰用于减少或消除连接酶IV的表达或活性或用于增加连接酶IV的表达或活性。连接酶IV抑制因子的一个例

子是SCR7。

[0302] 靶向细胞周期关卡蛋白的抑制因子如ATM (例如KU55933)、CHK1/CHK2 (例如KLD1162或CHIR-124) 和ATR (例如VE 821) 还可用于协同增强特定DNA修复抑制因子的效应或预防非预期的不良效应如细胞周期停滞和/或细胞凋亡(参见Ciccica et al. (2010) *Mol Cell* 40:179(Ciccica等人, 2010年,《分子细胞》,第40卷,第179页),该文献出于所有目的全文以引用方式并入本文。

[0303] C-NHEJ的破坏可增大通过“另选”NHEJ (A-NHEJ) 途径介导的异常连接水平,并且还可增加HR修复。A-NHEJ途径显示出朝向微同源介导的连接的偏倚,并且遵循比C-NHEJ慢的动力学。已提出若干因子,包括MRN复合物(MRE11、RAD50、NBS1)、CtIP、XRCC1、PARP、Lig1和Lig3参与。参见Kasperek&Humphrey (2011) *Seminars in Cell&Dev. Biol.* 22:886-897 (Kasperek和Humphrey, 2011年,《细胞与发育生物学研讨编辑》,第22卷,第886-897页) 和Claybon et al. (2010) *Nucleic Acids Res.* 38(21):7538-7545 (Claybon等人, 2010年,《核酸研究》,第38卷第21期,第7538-7545页),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。

[0304] 因此,在本文所公开的方法的一些中,所述细胞已被修饰用于减少或消除或用于增加参与A-NHEJ的因子的表达或活性。例如,在一些方法中,所述细胞已被修饰用于减少或消除MRE11、RAD50、NBS1、CtIP、XRCC1、PARP (例如PARP1)、Lig1和/或Lig3的表达或活性。在其他方法中,所述细胞已被修饰用于增加MRE11、RAD50、NBS1、CtIP、XRCC1、PARP (例如PARP1)、Lig1和/或Lig3的表达或活性。在特定方法中,所述细胞已被修饰用于减少或消除PARP1的表达或活性或用于增加PARP1的表达或活性(指定为P11103的示例性UniProt序列)。PARP抑制因子(例如NU1025、Iniparib、Olaparib)的例子包括烟酰胺;异喹啉酮和二氢异喹啉酮;苯并咪唑和吡啶;酞嗪-1(2H)-酮和喹啉酮;异吡啶酮及其类似物和衍生物;菲啶和菲啶酮;苯并吡喃酮及其类似物和衍生物;不饱和羧酸衍生物及其类似物和衍生物;哒嗪,包括稠合哒嗪及其类似物和衍生物;和/或其他化合物诸如咖啡因、茶碱和胸苷及其类似物和衍生物。参见,例如,美国专利No. 8,071,579,该专利出于所有目的全文以引用方式并入本文。

[0305] C-NHEJ还表现出与HR的竞争关系,使得破坏C-NHEJ还可导致增加的HR修复。可利用NHEJ和HR之间的这种竞争,这是因为破坏NHEJ可通过减少随机整合来增强基因靶向,并且可能通过同源重组来增加靶整合。

[0306] 存在若干同源重组修复形式,包括单链退火、基因转换、交换和断裂诱导复制。单链退火是HR修复的次要形式,其中同源单链序列位于经切除的DSB退火链的任一侧上,从而导致染色体重建。根据序列同源性的两个区域的分隔距离,单链退火产生变化尺寸的缺失。基因转换包括遗传信息从一个分子到其同源物的非相互交换,这直接源于从同源染色体拷贝遗传信息造成的HR。当同源序列不相同,这可导致局部的LOH。通常,基因转换程度受限数百碱基对。然而,已报告了在一些遗传背景下的长段基因转换,包括RAD51C缺失。参见Nagaraju et al. (2006) *Mol. Cell. Biol.* 26:8075-8086 (Nagaraju等人, 2006年,《分子细胞生物学》,第26卷,第8075-8086页),该文献出于所有目的全文以引用方式并入本文。交换可在例如同源染色体之间发生,并且如果发生在G1中,那么交换具有发生相互易位的可能性,或者如果发生在G2中,那么交换具有发生非相互易位以及自断裂位点延伸至远侧端粒的

LOH的可能性。断裂诱导复制是HR的一种变体,其中伴随着链侵入,DNA复制持续进行至染色体的末端。因此,存在很多机制,通过这些机制HR可促进LOH。

[0307] 因此,在本文所公开的方法的一些中,所述细胞已被修饰用于减少或消除或用于增加参与HR的因子的表达或活性。例如,在一些方法中,所述细胞已被修饰用于增加RAD51、RAD52、RAD54、RAD55、RAD51C、BRCA1和/或BRCA2的表达或活性。在其他方法中,所述细胞已被修饰用于减少或消除RAD51、RAD52、RAD54、RAD55、RAD51C、BRCA1和/或BRCA2的表达或活性。

[0308] 在一些方法中,可改变参与调控NHEJ和/或HR的其他蛋白的表达或活性。例如,在一些方法中,所述细胞已被修饰用于减少或消除Chk2的表达或活性,用于减少或消除Clspn的表达或活性,用于减少或消除Setd2的表达或活性,用于增加Kat2a的表达或活性,并且/或者用于增加Rad51的表达或活性。在其他方法中,所述细胞已被修饰用于增加Chk2的表达或活性,用于增加Clspn的表达或活性,用于增加Setd2的表达或活性,用于减少或消除Kat2a的表达或活性,并且/或者用于减少或消除Rad51的表达或活性。

[0309] Chk2(也称为Chek2和Rad53;S.pombe同源物为Cds1)为响应于DNA双链断裂存在的关卡介导细胞周期停滞、DNA修复激活和细胞凋亡需要的丝氨酸/苏氨酸蛋白激酶。参见Blaikley et al. (2014) *Nucleic Acids Research* 42:5644-5656 (Blaikley等人,2014年,《核酸研究》,第42卷,第5644-5656页),该文献出于所有目的全文以引用方式并入本文。Clspn(也称为Claspin;S.pombe同源物为Mrc1)是响应于DNA损伤的关卡介导细胞周期停滞所需的蛋白。S.pombe中的Chk2或Clspn的同源物缺失已被报告会产生超重重组表型,与野生型相比,所述超重重组表型表现出断裂诱导基因转换水平显著升高。具体地讲,基因转换水平已被报告显著增大,反之非同源末端连接(NHEJ)、姐妹染色单体转换(SCC)和杂合性丢失(LOH)的水平已被报告有所降低。参见Blaikley et al. (2014) *Nucleic Acids Research* 42:5644-5656 (Blaikley等人,2014年,《核酸研究》,第42卷,第5644-5656页)。

[0310] Kat2a(也称为Gcn5和Gcn5l2)为普遍存在的组蛋白乙酰转移酶,所述组蛋白乙酰转移酶促进转录激活并且已被报告与双链断裂修复相关联。当抑制非同源末端连接时,Kat2a依赖的组蛋白H3赖氨酸36(H3K36)乙酰化增大染色质可接近性,提高切除率,并且促进同源重组。参见Pai et al. (2014) *Nat. Commun.* 5:4091 (Pai等人,2014年,《自然通讯》,第5卷,第4091页),该文献出于所有目的全文以引用方式并入本文。Setd2(也称为Kiaa1732、Kmt3a和Set2)为组蛋白甲基转移酶,所述组蛋白甲基转移酶使用脱甲基化赖氨酸36(H3K36me2)作为底物来特异性地三甲基化组蛋白H3的赖氨酸36(H3K36me3)。Setd2依赖的H3K36甲基化减少染色质可接近性,提高切除率,并且促进NHEJ。参见Pai et al. (2014) *Nat. Commun.* 5:4091 (Pai等人,2014年,《自然通讯》,第5卷,第4091页)。

[0311] Rad 51(也称为Reca、Rad51A和DNA修复蛋白Rad51同源物1)为一种蛋白,与Rad52和其他蛋白一起作用以实现同源重组期间的链交换,从而形成由失配修复解决的异源双链DNA以产生基因转换段。在哺乳动物细胞中,Rad51和Rad52过表达已被报告会增加同源重组和基因转换的频率。参见Yanez&Porter (1999) *Gene Ther.* 6:1282-1290 (Yanez和Porter,1999年,《基因治疗》,第6卷,第1282-1290页)和Lambert&Lopez (2000) *EMBO J.* 19:3090-3099 (Lambert和Lopez,2000年,《EMBO杂志》,第19卷,第3090-3099页),该文献出于所有目的全文以引用方式并入本文。

[0312] 参与调控NHEJ、基因转换和/或同源介导修复的基因的表达或活性中的修饰可为空间特异的或时序特异的,并且还可以是可诱导的或暂时的和可逆的。例如,可构建各种形式的盒,以便可以在特定细胞或组织类型中、在特定发育阶段或在诱导时缺失。此类盒可采用重组酶系统,其中所述盒在两侧上侧接重组酶识别位点,并且可使用在所需细胞类型中表达的、在所需发育阶段表达的、或在诱导时表达或激活的重组酶来去除。可进一步构建此类盒,以包括一系列成对的不同重组酶识别位点,这些位点被布置成使得可生成无效等位基因、条件等位基因或条件/无效组合等位基因,如US 2011/0104799中所述,该专利出于所有目的全文以引用方式并入本文。可按各种方式控制重组酶基因的调控,诸如将重组酶基因有效连接至细胞特异性、组织特异性或受发育调控的启动子(或其他调控元件),或将重组酶基因有效连接至3' -UTR,该3' -UTR包含仅在特定细胞类型、组织类型或发育阶段中活跃的miRNA的识别位点。重组酶还可例如通过以下方式调控:采用使重组酶处于效应物或代谢物的控制之下的融合蛋白(例如,其活性受到它莫西芬控制的CreER^{T2}),或使重组酶基因处于诱导型启动子(例如,其活性受到强力霉素和TetR或TetR变体控制的诱导型启动子)的控制之下。例如在US 8,518,392、US 8,354,389和US 8,697,851中提供了各种形式的盒及调控重组酶基因的方式的示例,这些专利中的每一篇全文均以引用的方式并入。

[0313] G. 细胞和动物

[0314] 本文提供的各种组合物和方法采用细胞,诸如来自动物的细胞。此类细胞可来自非人动物。此类细胞可以是真核细胞,包括例如真菌细胞(例如酵母)、植物细胞、动物细胞、哺乳动物细胞和非多能的人细胞。哺乳动物细胞可以是例如非人哺乳动物细胞、非多能的人细胞、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。真核细胞可以是非人全能细胞、非人多能细胞(例如小鼠胚胎干(ES)细胞或大鼠ES细胞)或非多能细胞。非人全能细胞包括未分化细胞,该未分化细胞可产生任何细胞类型,并且多能细胞包括未分化细胞,该未分化细胞具有发育成多于一种分化细胞类型的能力。此类多能和/或全能细胞可以是例如非人胚胎干(ES)细胞或类ES细胞诸如非人诱导多能干(iPS)细胞。非人胚胎干细胞包括胚胎起源的非人全能或多能细胞。非人ES细胞可来源于胚胎的内细胞团,并且能够分化成为三个脊椎动物胚层(内胚层、外胚层和中胚层)中任一层的细胞。

[0315] 真核细胞还可以是非原代体细胞的细胞。体细胞可包括不是配子、生殖细胞、配子母细胞或未分化干细胞的任何细胞。

[0316] 真核细胞还包括非人原代细胞。非人原代细胞包括直接从非人生物体、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。它们包括获自非人生物体、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。

[0317] 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非人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相反地可继续进行分裂。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。

[0318] 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。

[0319] 真核细胞还可包括非人类哺乳动物的 1-细胞期胚胎 (即, 受精卵母细胞或受精卵)。此类 1-细胞期胚胎可来自任何遗传背景 (例如 BALB/c、C57BL/6、129 或它们的组合), 可以是新鲜的或冷冻的, 并且可以来源于体外受精。

[0320] 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, 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。

[0321] 小鼠多能和/或全能细胞可来自 129 细胞株、C57BL/6 细胞株、129 和 C57BL/6 的混合株、BALB/c 细胞株或 Swiss Webster 细胞株。129 细胞株的例子包括 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), 该文献出于所有目的全文以引用方式并入本文。C57BL 细胞株的例子包括 C57BL/A、C57BL/An、C57BL/GrFa、C57BL/Kal_wN、C57BL/6、C57BL/6J、C57BL/6ByJ、C57BL/6NJ、C57BL/10、C57BL/10ScSn、C57BL/10Cr 和 C57BL/01a。小鼠多能和/或全能细胞还可来自上述 129 细胞株和上述 C57BL/6 细胞株的混合株 (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 细胞株的混合株或上述 BL/6 细胞株的混合株 (例如, 129S6 (129/SvEvTac) 细胞株)。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), 该文献出于所有目的全文以引用方式并入本文。

[0322] 大鼠多能和/或全能细胞可来自任何大鼠细胞株, 包括例如 ACI 大鼠细胞株、黑刺 (DA) 大鼠细胞株、威斯塔鼠细胞株、LEA 大鼠细胞株、Sprague Dawley (SD) 大鼠细胞株, 或者 Fischer 大鼠细胞株诸如 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细胞株的混合株的细胞株。例如, 大鼠多能和/或全能细胞可来自 DA 细胞株或 ACI 细胞株。ACI 大鼠细胞株的特点是具有黑刺、白色的腹和足, 为 RT1^{av1} 单倍型。此类细胞株得自多种来源, 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) 大鼠细胞株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 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细胞株。参见例如 2014 年 2 月 20 日提交的 U.S. 2014/0235933A1, 该文献出于所有目的全文以引用方式并入本文。

[0323] 人多能细胞的例子包括直接源自分化的成年细胞的人类诱导性多能干细胞 (人 iPS 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),该文献出于所有目的全文以引用方式并入本文。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),该文献出于所有目的全文以引用方式并入本文。

[0324] 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细胞株或来自不同细胞株。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/或宿主胚胎相同的细胞株,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/或宿主胚胎不同的细胞株。

[0325] 可在本文所公开的方法和组合物中使用多种非人宿主胚胎。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、US 7,294,754和US 2008/0078000 A1,这些专利中的每一篇均出于所有目的全文以引用方式并入本文。在其他情况下,非人供体ES细胞可在2-细胞期、4-细胞期、8-细胞期、16-细胞期、32-细胞期或64-细胞期被植入到非人宿主胚胎中。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:小鼠发育图谱”,纽约施普林格出版社)中所述的Theiler期,该文献出于所有目的全文以引用方式并入本文。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。

[0326] H. 鉴定具有经修饰基因组的细胞的方法

[0327] 上述方法中的一些还包括鉴定具有经修饰基因组的细胞的步骤。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(例如在第一CRISPR RNA识别序列和第二CRISPR RNA识别序列之间)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。

[0328] 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。

[0329] 筛选步骤还可包括保留测定,该保留测定是用于区分将核酸插入物正确靶向插入到靶基因组基因座中和将核酸插入物随机转基因插入到靶基因组基因座外的基因组位置中的测定法。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由于其较大同源臂尺寸,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。

[0330] 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。

[0331] 因为成对gRNA可在靶基因组基因座处形成较大Cas介导的缺失,所以它可有助于增强标准LOA测定和GOA测定来证实通过LTVEC进行的正确靶向 (即在细胞中而非1-细胞期胚胎中)。例如,具体地讲,如果GOA测定采用针对LTVEC插入物中的选择盒的探针,那么只是进行LOA测定和GOA测定可能无法区分正确靶向的细胞克隆与其中靶基因组基因座的较大Cas诱导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会包括选择盒和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,那么一般来讲,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,那么一般来讲,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测定和GOA测定用于评估LTVEC靶向整合,并且GOA测定利用针对选择盒的探针,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进行的正确靶向,可单独使用保留测定,或者将保留测定与LOA测定和/或GOA测定结合使用。

[0332] 保留测定确定5'靶序列(对应于LTVEC的5'同源臂)和/或3'靶序列(对应于LTVEC的3'同源臂)中的DNA模板的拷贝数。具体地讲,确定与邻近选择盒的同源臂对应的靶序列中DNA模板的拷贝数是有用的。在二倍体细胞中,大于2的拷贝数通常表明在靶基因组基因座外部而非在靶基因组基因座处的LTVEC随机转基因整合,这是不可取的。正确靶向的克隆将会保持拷贝数为2。此外,在此类保留测定中少于2的拷贝数通常表明延伸超出靶向缺失的区域的较大Cas介导缺失,这也是不可取的。

[0333] 在用于鉴定二倍体细胞中靶基因组基因座处核酸插入物的靶向插入的示例性保留测定中,DNA最初获自己与大靶向载体(LTVEC)接触的细胞,该大靶向载体包含侧接与第一靶序列杂交的第一同源臂和与第二靶序列杂交的第二同源臂的核酸插入物,其中该核酸插入物包含邻近第一同源臂的选择盒。任选地,选择盒可包含药物抗性基因。随后将所述DNA暴露于结合在第一靶序列内的探针、结合在核酸插入物内的探针以及结合在具有已知拷贝数的参考基因内的探针,其中每个探针在结合后产生可检测信号。随后,检测来自所述探针中每个探针结合的信号。将来自参考基因探针的信号与来自第一靶序列探针的信号进行比较以确定第一靶序列的拷贝数,并将来自参考基因探针的信号与来自核酸插入物探针的信号进行比较以确定核酸插入物的拷贝数。核酸插入物拷贝数为一个或两个并且第一靶序列拷贝数为两个通常表明核酸插入物在靶基因组基因座处的靶向插入,而核酸插入物拷贝数为一个或多个并且第一靶序列拷贝数为三个或更多个通常表明核酸插入物在基因组基因座处而非靶基因组基因座处的随机插入。

[0334] 来自第一靶序列探针结合的信号可用于确定第一靶序列的阈值循环(Ct)值,来自参考基因探针结合的信号可用于确定参考基因的阈值循环(Ct)值,并且第一靶序列的拷贝数可通过比较第一靶序列Ct值和参考基因Ct值确定。同样地,来自核酸插入物探针结合的信号可用于测定用于核酸插入物的阈值循环(Ct)值,并且可通过对比第一靶序列Ct值和参考基因Ct值来测定核酸插入物的拷贝数。

[0335] LTVEC中的核酸插入物可为例如至少5kb、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或至少500kb。第一靶序列中探针所结合的序列和选择盒之间的距离可为,例如,不超过100个核苷酸、200个核苷酸、300个核苷酸、400个核苷酸、500个核苷酸、600个核苷酸、700个核苷酸、800个核苷酸、900个核苷酸、1kb、1.5kb、2kb、2.5kb、3kb、3.5kb、4kb、4.5kb或5kb。

[0336] 此类方法还可包括附加的保留测定来确定第二靶序列的拷贝数。例如,此类方法还可包括:将细胞的DNA暴露于结合第二靶序列的探针,检测来自第二靶序列探针结合的信号,并且将来自参考基因探针的信号与来自第二靶序列探针的信号进行比较以确定第二靶序列的拷贝数。

[0337] 同样地,此类方法还可包括附加的GOA测定来确定核酸插入物内的一种或多种附加序列的拷贝数。例如,此类方法还可包括:将细胞的DNA暴露于结合核酸插入物的一种或多种附加探针,检测来自一种或多种附加探针结合的信号,以及将来自参考基因探针的信号与来自一种或多种附加核酸插入物探针的信号进行比较以确定核酸插入物内一种或多种附加序列的拷贝数。

[0338] 同样地,当LTVEC被设计用于从靶基因组基因座中缺失内源性序列时,或者当使用

成对gRNA(例如,以在单个靶基因组基因座中不同位点处形成成对双链断裂并且缺失间插内源性序列)时,此类方法还可包括LOA测定以确定在靶基因组基因座处的内源性序列的拷贝数。例如,此类方法还可包括:将细胞的DNA暴露于结合靶基因组基因座处内源性序列的探针,检测来自内源性序列探针结合的信号,以及将来自参考基因探针的信号与来自内源性序列探针的信号进行比较以确定内源性序列的拷贝数。

[0339] 合适的定量测定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[®]、MMP assays[®]、TaqMan[®]分子信标或Eclipse[™]探针技术(参见例如US2005/0144655,该专利出于所有目的全文以引用方式并入本文)。

[0340] 针对未使用LTVEC产生的靶向遗传修饰,可使用筛选靶向修饰的常规测定诸如长片段PCR、Southern印迹或Sanger测序。此类测定通常用于获得被插入的靶向载体和靶向基因组基因座之间连接的证据。例如,对于长片段PCR测定,一个引物可识别插入的DNA中的序列,而另一个引物识别超出靶向载体同源臂末端的靶基因组序列。

[0341] 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、实施例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详细描述,但显而易见的是,在所附权利要求的范围内还可作出某些变化和修改。

[0342] 表1.序列说明。

[0343]

SEQ ID NO	类型	说明
1	RNA	gRNA
2	RNA	gRNA
3	RNA	crRNA
4	RNA	crRNA
5	RNA	crRNA
6	RNA	tracrRNA
7	RNA	tracrRNA
8	DNA	CRISPR RNA 识别序列
9	DNA	CRISPR RNA 识别序列
10	DNA	CRISPR RNA 识别序列

[0344]

11	DNA	<i>C5 (Hc)</i> gRNA A DNA-靶向区段 (距离靶基因座端点 100bp)
12	DNA	<i>C5 (Hc)</i> gRNA B DNA-靶向区段 (距离靶基因座端点 500bp)
13	DNA	<i>C5 (Hc)</i> gRNA C DNA-靶向区段 (距离靶基因座端点 38200bp 和 37500bp)
14	DNA	<i>C5 (Hc)</i> gRNA D DNA-靶向区段 (距离靶基因座端点 43500bp 和 32200bp)
15	DNA	<i>C5 (Hc)</i> gRNA E DNA-靶向区段 (距离靶基因座端点 500bp)
16	DNA	<i>C5 (Hc)</i> gRNA E2 DNA-靶向区段 (距离靶基因座端点 100bp)
17	DNA	<i>Lrp5</i> gRNA A DNA-靶向区段 (距离靶基因座端点 50bp)
18	DNA	<i>Lrp5</i> gRNA B DNA-靶向区段 (距离靶基因座端点 500bp)
19	DNA	<i>Lrp5</i> gRNA B2 DNA-靶向区段 (距离靶基因座端点 1000bp)
20	DNA	<i>Lrp5</i> gRNA C DNA-靶向区段 (距离靶基因座端点 29900bp 和 38430bp)
21	DNA	<i>Lrp5</i> gRNA D DNA-靶向区段 (距离靶基因座端点 29950bp 和 38380bp)
22	DNA	<i>Lrp5</i> gRNA E2 DNA-靶向区段 (距离靶基因座端点 1000bp)
23	DNA	<i>Lrp5</i> gRNA E DNA-靶向区段 (距离靶基因座端点 500bp)
24	DNA	<i>Lrp5</i> gRNA F DNA-靶向区段 (距离靶基因座端点 50bp)
25	DNA	<i>Ror1</i> gRNA A DNA-靶向区段 (距离靶基因座端点 200bp)
26	DNA	<i>Ror1</i> gRNA B DNA-靶向区段 (距离靶基因座端点 1000bp)
27	DNA	<i>Ror1</i> gRNA D DNA-靶向区段 (距离靶基因座端点 54300bp 和 55500bp)
28	DNA	<i>Ror1</i> gRNA C DNA-靶向区段 (距离靶基因座端点 54500bp 和 55300bp)
29	DNA	<i>Ror1</i> gRNA E DNA-靶向区段 (距离靶基因座端点 1000bp)
30	DNA	<i>Ror1</i> gRNA F DNA-靶向区段 (距离靶基因座端点 200bp)
31	DNA	<i>Trpa1</i> gRNA A DNA-靶向区段 (距离靶基因座端点 100bp)
32	DNA	<i>Trpa1</i> gRNA A2 DNA-靶向区段 (距离靶基因座端点 500bp)
33	DNA	<i>Trpa1</i> gRNA B DNA-靶向区段 (距离靶基因座端点 1000bp)
34	DNA	<i>Trpa1</i> gRNA C DNA-靶向区段 (距离靶基因座端点 25600bp 和 19740bp)
35	DNA	<i>Trpa1</i> gRNA D DNA-靶向区段 (距离靶基因座端点 26970bp 和 18370bp)
36	DNA	<i>Trpa1</i> gRNA E2 DNA-靶向区段 (距离靶基因座端点 1000bp)
37	DNA	<i>Trpa1</i> gRNA E DNA-靶向区段 (距离靶基因座端点 500bp)
38	DNA	<i>Trpa1</i> gRNA F DNA-靶向区段 (距离靶基因座端点 100bp)
39	DNA	190045 正向引物

[0345]

40	DNA	190061 正向引物
41	DNA	190068 正向引物
42	DNA	190030 正向引物
43	DNA	190033 正向引物 (与图 5 中针对 SV 48.3 的正向引物相同)
44	DNA	190013 正向引物
45	DNA	190045 反向引物
46	DNA	190061 反向引物
47	DNA	190068 反向引物
48	DNA	190030 反向引物
49	DNA	190033 反向引物 (与图 5 中针对 SV 48.3 的反向引物相同)
50	DNA	190013 反向引物
51	DNA	C2 探针(B6) -图 5 中 SNV 0.32
52	DNA	T3 探针(B6) -图 5 中 SNV 1.2
53	DNA	T6 探针(B6) -图 5 中 SNV 11.1
54	DNA	T7 探针(B6) -图 5 中 SNV 13.2
55	DNA	T8 探针(B6) -图 5 中 SNV 17.5
56	DNA	T9 探针(B6) -图 5 中 SNV 25.8
57	DNA	T10 探针(B6) -图 5 中 SNV 33.0
58	DNA	T11 探针(B6) -图 5 中 SNV 38.3
59	DNA	T13 探针(B6) -图 5 中 SNV 49.6
60	DNA	T14 探针(B6) -图 5 中 SNV 57.2
61	DNA	C2 探针(129) -图 5 中 SNV 0.32
62	DNA	T3 探针(129) -图 5 中 SNV 1.2
63	DNA	T6 探针(129) -图 5 中 SNV 11.1
64	DNA	T7 探针(129) -图 5 中 SNV 13.2
65	DNA	T8 探针(129) -图 5 中 SNV 17.5
66	DNA	T9 探针(129) -图 5 中 SNV 25.8
67	DNA	T10 探针(129) -图 5 中 SNV 33.0
68	DNA	T11 探针(129) -图 5 中 SNV 38.3

[0346]

69	DNA	T13 探针(129)-图 5 中 SNV 49.6
70	DNA	T14 探针(129)-图 5 中 SNV 57.2
71	DNA	C2 正向引物-图 5 中 SNV 0.32
72	DNA	T3 正向引物-图 5 中 SNV 1.2
73	DNA	T6 正向引物-图 5 中 SNV 11.1
74	DNA	T7 正向引物-图 5 中 SNV 13.2
75	DNA	T8 正向引物-图 5 中 SNV 17.5
76	DNA	T9 正向引物-图 5 中 SNV 25.8
77	DNA	T10 正向引物-图 5 中 SNV 33.0
78	DNA	T11 正向引物-图 5 中 SNV 38.3
79	DNA	T13 正向引物-图 5 中 SNV 49.6
80	DNA	T14 正向引物-图 5 中 SNV 57.2
81	DNA	C2 反向引物-图 5 中 SNV 0.32
82	DNA	T3 反向引物-图 5 中 SNV 1.2
83	DNA	T6 反向引物-图 5 中 SNV 11.1
84	DNA	T7 反向引物-图 5 中 SNV 13.2
85	DNA	T8 反向引物-图 5 中 SNV 17.5
86	DNA	T9 反向引物-图 5 中 SNV 25.8
87	DNA	T10 反向引物-图 5 中 SNV 33.0
88	DNA	T11 反向引物-图 5 中 SNV 38.3
89	DNA	T13 反向引物-图 5 中 SNV 49.6
90	DNA	T14 反向引物-图 5 中 SNV 57.2
91	DNA	正向引物-图 5 中 SV 13.7
92	DNA	反向引物-图 5 中 SV 13.7
93	DNA	正向引物-图 5 中 SV 20.0
94	DNA	反向引物-图 5 中 SV 20.0
95	DNA	正向引物-图 5 中 SV 36.9
96	DNA	反向引物-图 5 中 SV 36.9
97	DNA	正向引物-图 5 中 SV 56.7

[0347]

98	DNA	反向引物-图 5 中 SV 56.7
99	DNA	图 1 中 m-lr-f 引物
100	DNA	图 1 中 m-5'-f 引物
101	DNA	图 1 中 m-A 引物
102	DNA	图 1 中 h-lr-r 引物
103	DNA	图 1 中 m-5'-r 引物
104	DNA	图 1 中 h-5'-r 引物
105	DNA	图 1 中 m-F 引物
106	DNA	图 1 中 m-E2 引物
107	RNA	<i>Cmah</i> gRNA A DNA-靶向区段
108	RNA	<i>Cmah</i> gRNA B DNA-靶向区段
109	DNA	<i>Cmah</i> 基因座
110	DNA	gRNA A 切口上游的 <i>Cmah</i> 基因座
111	DNA	gRNA B 切口下游的 <i>Cmah</i> 基因座
112	RNA	由 gRNA A 和 gRNA B 切除的 <i>Cmah</i> 基因座序列
113	DNA	图 13 中 C5 引物 m-5'-F
114	DNA	图 13 中 C5 引物 m-5'-R
115	DNA	图 13 中 C5 引物 h-5'-R
116	DNA	图 15 中 <i>Cmah</i> TaqMan [®] 正向引物
117	DNA	图 15 中 <i>Cmah</i> TaqMan [®] 探针
118	DNA	图 15 中 <i>Cmah</i> TaqMan [®] 反向引物
119	DNA	7064retU 正向引物
120	DNA	7064retU 反向引物
121	DNA	7064retU TaqMan [®] 探针
122	DNA	7064retD 正向引物
123	DNA	7064retD 反向引物
124	DNA	7064retD TaqMan [®] 探针
125	DNA	7140retU 正向引物
126	DNA	7140retU 反向引物

[0348]

127	DNA	7140retU TaqMan [®] 探针
128	DNA	7140retD 正向引物
129	DNA	7140retD 反向引物
130	DNA	7140retD TaqMan [®] 探针
131	DNA	mADAM6-2 LOA 正向引物
132	DNA	mADAM6-2 LOA 反向引物
133	DNA	mADAM6-2 LOA 探针
134	DNA	hIgH31 LOA 正向引物
135	DNA	hIgH31 LOA 反向引物
136	DNA	hIgH31 LOA 探针
137	DNA	hIgH9 LOA 正向引物
138	DNA	hIgH9 LOA 反向引物
139	DNA	hIgH9 LOA 探针
140	DNA	hIgH1 LOA 正向引物
141	DNA	hIgH1 LOA 反向引物
142	DNA	hIgH1 LOA 探针
143	DNA	Neo GOA 正向引物
144	DNA	Neo GOA 反向引物
145	DNA	Neo GOA 探针
146	DNA	5'IgH Arm1 保留测定正向引物
147	DNA	5'IgH Arm1 保留测定反向引物
148	DNA	5'IgH Arm1 保留测定探针
149	DNA	mIgM398 保留测定正向引物
150	DNA	mIgM398 保留测定反向引物
151	DNA	mIgM398 保留测定探针
152	DNA	mIgM1045 保留测定正向引物
153	DNA	mIgM1045 保留测定反向引物
154	DNA	mIgM1045 保留测定探针
155	DNA	3' IgH Arm2 保留测定正向引物

[0349]

156	DNA	3' IgH Arm2 保留测定反向引物
157	DNA	3' IgH Arm2 保留测定探针
158	DNA	mIgHp2 亲本正向引物
159	DNA	mIgHp2 亲本反向引物
160	DNA	mIgHp2 亲本探针
161	DNA	mIgKd2 亲本正向引物
162	DNA	mIgKd2 亲本反向引物
163	DNA	mIgKd2 亲本探针
164	DNA	hIgK5 亲本正向引物
165	DNA	hIgK5 亲本反向引物
166	DNA	hIgK5 亲本探针
167	DNA	3'gRNA_I DNA 靶向序列
168	DNA	3'gRNA_II DNA 靶向序列
169	DNA	5' gRNA_I DNA 靶向序列
170	DNA	5' gRNA_II DNA 靶向序列
171	DNA	5' IgH Arm2 保留测定正向引物
172	DNA	5' IgH Arm2 保留测定反向引物
173	DNA	5' IgH Arm2 保留测定探针
174	DNA	3'IgH Arm1 保留测定正向引物
175	DNA	3'IgH Arm1 保留测定反向引物
176	DNA	3'IgH Arm1 保留测定探针

[0350] 实例

[0351] 实例1:使用一个向导RNA或两个向导RNA的CRISPR/Cas9介导的靶向。

[0352] 材料和方法

[0353] 非人ES细胞培养、筛选和电穿孔

[0354] 使用VGF1,即我们的C57BL6NTac/129S6SvEvF1混合XY ES细胞系(Poueymirou et al. (2007) Nat. Biotechnol. 25:91-99 (Poueymirou等人, 2007年,《自然生物技术》,第25卷,第91-99页); Valenzuela et al. (2003) Nat. Biotechnol. 21:652-659 (Valenzuela等人, 2003年,《自然生物技术》,第21卷,第652-659页)进行本文所述的实验。如前所述培养ES细胞(Matise et al. (2000) in Joyner, A.L. ed. Gene Targeting: a practical approach, pp.100-132, Oxford University Press, New York (等人, 2000年, 载于Joyner, A.L. 编辑的《经典基因敲除指南》,第100-132页, 牛津大学出版社, 纽约))。

[0355] 使用2mm间隙比色皿中最终体积为0.12ml的750万细胞进行电穿孔(EP)。使用BTX ECM 630电穿孔系统(美国马萨诸塞州霍利斯顿的哈佛仪器公司(Harvard Apparatus,

Holliston, MA)) 进行EP的电气条件是700V、400欧姆电阻和25微法电容。每EP的LTVEC的量为0.0015mg, Cas9表达质粒为0.005mg, 并且sgRNA表达质粒为0.010mg。通过添加100ng赋予嘌呤霉素抗性的质粒以允许在不选择LTVEC表达的新霉素抗性的情况下选择克隆, 来进行一些EP。EP之后, 将细胞接种到两个15cm胶凝化的皿上, 并且每日更换培养基。在EP后48小时开始选择含有100 μ g/ml G-418硫酸盐或0.0015mg/ml嘌呤霉素的培养基, 并且持续进行直至EP后10天。在PBS中挑取集落, 并且添加到96孔皿中, 该皿含有0.05%胰蛋白酶并被允许解离15分钟, 用培养基中和并用于分离DNA以供筛选。

[0356] 等位基因修饰方法 (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)) 用于鉴定正确靶向的ES细胞克隆以及确定小鼠等位基因基因型。

[0357] 向导序列设计

[0358] 将DNA的围绕Lrp5或其他靶向基因的缺失部分内部、上游和下游50bp、100bp、500bp或1kb位置的大约200bp输入到CRISPR设计工具 (crispr.mit.edu), 以检索可能的gRNA序列。随后, 过滤可能的gRNA序列以确保它们只允许切割内源性DNA, 而不会切割LTVEC中的人源化插入物。

[0359] 单个向导RNA克隆

[0360] sgRNA要么作为双链寡核苷酸 (IDT) 被克隆进pMB_sgRNA (U6启动子) 中的BsmbI位点处, 与用于无缝RNA表达的77bp支架融合, 要么作为经验证的表达质粒购自GeneCopoeia (LRP5向导A、B、B2、E2、E和F)。通过PCR和Sanger测序来确认内部生产的质粒。

[0361] 用于基因型确认的DNA模板

[0362] 从非人ES细胞、来源于非人ES细胞的克隆纯化出DNA, 所述非人ES细胞已用靶向载体、表达Cas9的质粒、表达若干向导RNA (gRNA) 中的一个的质粒或者表达不同gRNA组合的两个质粒进行电穿孔。选择用于后续常规PCR测定的克隆, 这些克隆由等位基因修饰 (即等位基因丢失或等位基因获得) 定量PCR测定鉴定为具有小鼠靶基因座的靶向缺失和靶向载体的插入或具有Cas9/gRNA诱导缺失。

[0363] 寡核苷酸设计

[0364] 针对gRNA的每个组合设计两种PCR测定。第一PCR为用于检测不同gRNA组合的CRISPR RNA识别序列之间的剪切的缺失测定。为5' 测定的第二PCR测定包括两种PCR测定。第一种是针对人源化等位基因的5' 人测定并且被设计为跨小鼠-人接合区。第二种是针对内源性小鼠等位基因的5' 小鼠测定并且被设计为跨5' 靶向缺失接合区。

[0365] PCR反应和TOPO克隆

[0366] TaKaRa LA Taq DNA聚合酶 (目录号RR002M) 用于扩增非人ES细胞DNA模板。以水为阴性对照进行每个PCR测定反应混合。测定混合物含有以下组分: 0.005mL ES细胞DNA模板; 1X LA PCR缓冲液II (加入了Mg²⁺); 0.01mM dNTP混合物; 0.0075mM正向寡核苷酸 (各自); 0.0075mM反向寡核苷酸 (各自); 5000单位/mL LA Taq聚合酶; 以及0.025mL ddH₂O。

[0367] PCR热循环程序包括94℃下进行1分钟; 然后在94℃下进行35个循环30秒, 60℃梯度退火30秒, 并且每扩增1kb在68℃下进行1分钟; 然后在72℃下进行聚合10分钟。

[0368] 用Invitrogen的1kb plus DNA ladder (目录号10787-018) 和/或Invitrogen的50bp DNA Ladder (目录号10416-014) 在2%琼脂糖凝胶上通过电泳分级获得PCR产物。遵循

Invitrogen用于测序的TOPO TA克隆试剂盒(目录号K4575-02)的说明书将剩余的PCR产物克隆进入pCR4-TOPO载体中。将克隆反应物化学转化到One Shot Top10细胞中并接种在0.06mg/mL X-gal和0.025mg/mL卡那霉素琼脂平板上。

[0369] 测序

[0370] 将白色集落接种到含有0.025mg/mL卡那霉素的LB中,并在37℃下振荡孵育过夜。每个集落代表来自一组经测定产物的一个扩增子。使用QIAGEN质粒小量提取试剂盒(目录号12123)从每个细菌培养物中提取DNA。在测序反应混合物中确定插入物的DNA序列,该测序反应混合物包括0.002mL TOPO克隆的PCR、1x PCRx增强子溶液(10x母液)(目录号X11495-017)、0.0075mM寡核苷酸(M13F或M13R)和0.015mL ddH₂O。

[0371] 测序分析

[0372] 从测序结果剔除不确定序列和pCR4-TOPO载体序列,从而分离出PCR插入物序列。随后,将经测序的片段与参考序列比对,然后分析变异。

[0373] 对剪切的克隆测序

[0374] 按照制造商的说明书(Invitrogen目录号K4575-02)将来自经剪切的阳性克隆的PCR产物克隆到pCR4-TOPO载体中,随后将该PCR产物化学转化到One Shot Top10细胞中并接种在0.060mg/mL X-gal和0.025mg/mL卡那霉素琼脂平板上。使用QIAGEN质粒小量提取试剂盒(目录号12123)从细菌培养物中提取DNA。随后,将插入物测序结果与预测的剪切参考序列比对,并分析indel变异。经预测,Cas9将来自PAM的3个碱基对切割为可由gRNA识别的序列。从参考序列中缺失预测的切割位点中的序列,并且将剩余的序列用于与结果比对。

[0375] 用于单核苷酸变体(SNV)的TaqMan[®]等位基因分型测定

[0376] TaqMan[®]等位基因分型反应物为0.008mL,含有基因组DNA、针对每种多态性的特异性探针/引物和TaqMan[®]基因表达PCR主混合物。从赛默飞世尔的生命技术分公司(Life Technologies (Thermo))订购探针,从IDT订购引物。用于等位基因129的探针用VIC染料标记;用于等位基因B6的探针用FAM染料标记。每个TaqMan[®]等位基因测定在384孔板上一式四份地进行,并且在Applied BioSystems ViiA 7平台上运作。SNV PCR循环程序如下所述:在95℃下进行10分钟,然后将以下步骤进行40个循环:在95℃下进行15秒,在60℃下进行60秒,并且在60℃下进行30秒。使用ViiA 7软件v1.1完成运行分析和结果评估。

[0377] FISH分析

[0378] 使用荧光原位杂交(FISH)按照其标准流程由威斯康星州麦迪逊的Cell Line Genetics公司(Cell Line Genetics (Madison, Wisconsin))或者密歇根州大急流城的Van Andel研究所(Van Andel Institute (Grand Rapids, Michigan))来分析选定的ES细胞克隆。我们提供小鼠和人BAC作为用于双色分析的探针。

[0379] 靶基因座的增强基因组剪切和/或人源化

[0380] 为了在单一步骤中精确地使啮齿动物基因的全部或部分缺失,同时任选地用其人类同源物的全部或部分进行置换,我们通过电穿孔向啮齿动物ES细胞中导入以下核酸分子:(1)LTVEC;(2)编码Cas9核酸内切酶的质粒或mRNA;以及(3)编码一种或多种CRISPR单向导RNA(gRNA)的一种或多种质粒或者gRNA本身。在各个实验中,LTVEC是线形的。在一些实验

中,LTVEC包括人类基因的全部或部分,其编码侧接啮齿动物DNA同源臂的基因产物(蛋白质或RNA),所述同源臂被设计成引导使啮齿动物基因缺失并插入人类基因的同源重组事件。在其他实验中,LTVEC被设计成靶向单独的基因座,诸如Ch25h基因座。在任一种情况下,LTVEC还携带药物选择盒,所述药物选择盒指导赋予对抗生素药物(例如,G418)的抗性的酶(例如,新霉素磷酸转移酶)的表达。

[0381] 接受LTVEC并将其并入基因组的ES细胞能够在组织培养皿上于含有抗生素药物的生长培养基中生长并形成集落。因为我们引入了为LTVEC分子的500至1,000倍的CRISPR/Cas9编码和gRNA编码核酸分子,所以大多数含有LTVEC的耐药集落也至少瞬时地含有CRISPR/Cas9组分。我们挑取耐药性集落并通过等位基因修饰法(Valenzuela et al. (2003) Nat. Biotech. 21:652-660 (Valenzuela等人,2003年,《自然-生物技术》,第21卷,第652-660页);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);这些文献全文以引用方式并入本文)筛选它们,以鉴定具有正确靶向的人源化等位基因的克隆。此外,使用称为保留测定(retention assay)的识别LTVEC同源臂中序列的实时PCR测定,来验证LTVEC到小鼠基因组内的正确靶向。确定这些保留测定结果的拷贝数提供了进一步的说明,来帮助区分保留拷贝数为2的正确靶向ES克隆。在克隆中靶小鼠基因座的Cas9诱导的大缺失与基因组中其他地方的LTVEC随机整合同时发生的情况下,保留测定结果的拷贝数为3(或更多)。成对的gRNA在靶小鼠基因座处产生Cas9介导的大缺失的能力意味着,先前所述的标准LOA测定法和GOA测定法可以通过保留测定来增强,以提供进一步的说明并验证正确的靶向。因此,保留测定被设计成结合LOA测定法和GOA测定法使用。

[0382] 在各个实验中,使用一种或两种gRNA。所使用的gRNA各自在靶基因座的5'末端附近(即靶向小鼠基因缺失)、靶基因座的中间或靶基因座的3'末端附近指导Cas9切割。当使用两种gRNA时,一种gRNA在靶基因座的5'末端附近指导Cas9切割,另一种gRNA在靶基因座的中间或靶基因座的3'末端附近指导Cas9切割。

[0383] Lrp5基因座

[0384] 在一组实验中,LTVEC被设计成产生小鼠Lrp5(低密度脂蛋白受体相关蛋白5)基因中编码胞外域的部分的68kb缺失,同时用来自人类LRP5基因的同源序列的91kb片段置换(图1)。LTVEC包含人类LRP5基因中侧接同源臂的91kb片段,该同源臂含有7kb和33kb的来源于小鼠Lrp5基因座的部分的基因组DNA,所述部分侧接意欲缺失的小鼠Lrp5基因的68kb序列。在单独的实验中,将Lrp5人源化LTVEC与编码Cas9的质粒和编码设计用于在靶向缺失的小鼠Lrp5基因的区域产生双链断裂的八种gRNA(A、B、B2、C、D、E2、E、F)之一的第二质粒组合。所述gRNA被设计成避免识别人类LRP5基因的插入部分中的任何序列。在其他实验中,我们将LTVEC和编码Cas9的质粒以及编码两种靶向要缺失的小鼠Lrp5基因的区域不同位点的不同gRNA的质粒组合。

[0385] 通过对缺失内的序列以及药物选择盒和人类基因插入物内的序列进行等位基因修饰测定(Valenzuela et al. (2003) Nat. Biotechnol. 21:652-659 (Valenzuela等人,2003年,《自然生物技术》,第21卷,第652-659页);Frendewey et al. (2010) Methods Enzymol. 476:295-307 (Frendewey等人,《酶学方法》,2010年,第476卷,第295-307页))来筛选耐药ES细胞克隆,以用于靶向人源化。如果克隆丢失了两个内源性小鼠基因序列中的一

个并获得了人类插入物的一个拷贝,并且还保留了保留序列的两个拷贝(位于LTVEC的同源臂中),则将克隆评分为正确靶向。用于此筛选的两种保留测定是使用以下引物和探针的 **TaqMan**[®] 测定:7064retU正向引物CCTCCTGAGCTTTCCTTTGCAG (SEQ ID NO:119);7064retU反向引物CCTAGACAACACAGACACTGTATCA (SEQ ID NO:120);7064retU **TaqMan**[®] 探针TTCTGCCCTTGAAAAGGAGAGGC (SEQ ID NO:121);7064retD正向引物CCTCTGAGGCCACCTGAA (SEQ ID NO:122);7064retD反向引物CCCTGACAAGTTCTGCCTTCTAC (SEQ ID NO:123);7064retD **TaqMan**[®] 探针TGCCCAAGCCTCTGCAGCTTT (SEQ ID NO:124)。

[0386] Lrp5基因的CRISPR/Cas9辅助人源化的结果汇总在表2中。将单独的LTVEC引入ES细胞后,1.9%的所筛选耐药性克隆携带正确靶向的杂合人源化等位基因(参见表2中的杂合靶向列,其包括其中非靶向等位基因全部未突变的克隆,或具有CRISPR诱导的小突变(诸如由NHEJ引起的小缺失)的克隆)。相比之下,将LTVEC与八种经测试的gRNA(A、B、B2、C、D、E2、E和F;参见表1)中的七种导向的Cas9核酸内切酶组合产生了正确靶向的单等位基因杂合突变,突变效率在2.1%至7.8%的范围内。对于B2和D进行的Cas9导向的切割,除了单等位基因靶向外,还检测到了频率为1.0-2.1%的双等位基因纯合人源化。我们尚未观察到使用LTVEC自身的双等位基因靶向,即使对于简单的小缺失等位基因也是如此。纯合的Lrp5人源化ES细胞可以通过 **VELOCIMOUSE**[®] 方法(Poueymirou et al.(2007) Nat.Biotech.25:91-99(Poueymirou等人,2007年,《自然生物技术》,第25卷,第91-99页),该文献全文以引用的方式并入本文)直接转化到完全ES细胞来源的小鼠内以用于表型和药物功效研究。

[0387] 设计用于检测在预测的切割位点处或附近的gRNA/Cas9诱导的NHEJ突变的MOA测定法展示了所有测试的gRNA的突变活性(数据未显示)。在所有测定的克隆中检测到的单等位基因或双等位基因gRNA诱导的突变的比例随基因座和位置而变化。gRNA突变活性与LTVEC靶向之间没有强相关性,但是最低的靶向效率通常与具有最低突变频率的gRNA相关。

[0388] 将识别靶向缺失的Lrp5基因的区域的不同末端的两种gRNA组合主要是通过增加测试的五种组合中的三种中纯合靶向事件的频率来提高总人源化靶向效率(表2)。因为gRNA的组合具有在由gRNA编程的Cas9切割位点之间产生大缺失的潜力,我们还观察到在一个Lrp5等位基因上携带靶向人源化的半合子ES细胞克隆和在另一个等位基因上的大CRISPR诱导的缺失(gRNA组合A+F,表2)。此外,对于两种gRNA组合(A+F和A+E2),我们鉴定了具有独特基因型的ES细胞克隆:在Lrp5等位基因上都有CRISPR介导的大缺失。

[0389] 表2. 使用单独gRNA和组合gRNA进行Lrp5胞外域的CRISPR/Cas9辅助人源化的筛选结果。

[0390]

		等位基因型靶向效率				
gRNA	gRNA 位点距靶向缺失的 5'/3'末端的距离 (bp)	杂合靶向 (效率%)	半合子靶向 (效率%)	纯合靶向 (效率%)	总靶向 (效率%)	纯合缺失 (效率%)
A	50 (5')	7.8			7.8	
B	500 (5')	4.2			4.2	
B2	1000 (5')	6.2		1.0	7.2	
C	29900 (5')/ 38430 (3')	4.1			4.1	
D	29950 (5')/ 38380 (3')	5.2		2.1	7.3	
E2	1000 (3')	2.1			2.1	
E	500 (3')	0.0			0.0	
F	50 (3')	4.2			4.2	
A + F	A: 50 (5') F: 50 (3')	6.6	2.9	2.2	11.7	2.9
B + E	B: 500 (5') E: 500 (3')	2.5			2.5	
B2 + E2	B2: 1000 (5') E2: 1000 (3')	4.2		2.1	6.3	
A + E	A: 50 (5') E: 500 (3')	4.6		6.2	10.8	
A + E2	A: 50 (5') E2: 1000 (3')	2.0		4.0	6.0	4.0
无	不适用	1.9			1.9	

[0391] 如表2所显示,当使用靶向单个基因座的两种gRNA而不是一种gRNA时,观察到具有双等位基因靶向的克隆的百分比显著增加(参见图2A),这表明使用gRNA组合促进了双等位基因修饰。图2A示出了使用LTVEC和两个向导RNA(A和B)使小鼠基因缺失,并同时用对应的人类形式置换该基因的一般示意图。在使用两种gRNA时以高得多的频率观察到的独特的突变等位基因类型包括纯合剪切等位基因(图2B; Δ / Δ)、纯合靶向等位基因(图2C; HUM/HUM)、半合子靶向等位基因(图2D; (HUM/ Δ))和其他复合的杂合靶向等位基因(例如,一个等位基因具有LTVEC靶向人源化,而另一个等位基因具有CRISPR诱导的突变,诸如小缺失)(图2E)。

[0392] 进行若干PCR测定以支持并确认基于MOA测定法的基因型。引物示出于图1中,并且可见于表1。Lrp5LTVEC具有足够短的(6.9kb)5'同源臂,以通过测试人类插入物和相邻小鼠基因序列之间的物理连接的PCR来证明靶向(图1)。我们观察到具有来自评分为杂合、半合子或纯合的克隆的DNA,而不具有来自亲本ES细胞系或来自被评分为具有双等位基因大缺失的克隆的DNA的预期7.5kb PCR产物(图3A),因此证实了通过MOA(即,LOA和GOA)筛选产生

的靶向细胞并且支持推断的双等位基因大缺失。检查缺失和插入连接处的序列的5'-De1-J PCR测定(图3B)产生了具有来自亲本ES细胞系和大多数杂合人源化克隆的DNA的330bp产物(数据未显示)。对于杂合克隆AW-C3,5'-De1-J测定产生小于预期的产物(图3B),表明gRNA A/Cas9切割诱导非靶向等位基因上的小缺失突变,此突变也通过对gRNA A切割的MOA测定检测出(数据未显示)。可以预知,5'-De1-J测定对具有半合子、纯合和双等位基因缺失的等位基因的克隆是阴性的。检查人类DNA插入物的5'末端与相邻的小鼠旁侧序列之间的连接处的序列的5'-Ins-J PCR(图3B)在杂合、半合子和纯合克隆中产生了478bp的产物,因为它们具有至少一个靶向的人源化等位基因。5'-Ins-J PCR测定未产生具有双等位基因大缺失的克隆产物(图3B)。为了证实半合子和双等位基因缺失克隆中的大缺失,我们用识别双重gRNA靶位点以外的序列的引物进行PCR。测定A gRNA位点和F gRNA位点之间的缺失的De1(A+F)PCR(图1)产生了具有来自克隆AW-A8和B0-F10的约360bp的单一产物(图3B),证实Lrp5等位基因中的至少一个具有大缺失。同样,测定A gRNA位点和E2 gRNA位点之间大缺失的De1(A+E2)CR产生了具有来自克隆BA-A7的DNA的约250bp的单一产物。缺失PCR,与连接测定、LOA测定和GOA测定一起支持双等位基因大缺失的基因型。图3A和图3B中所示的测定结果是除了荧光原位杂交(FISH;图4A至图4C)之外还进行的类似测定的代表性实例,以证实表2中汇总的双等位基因的基因型。

[0393] 荧光原位杂交(FISH)用于证实Lrp5基因的纯合靶向人源化。将如靶向实验中所纯合靶向的、通过定量和常规PCR测定评分的ES细胞克隆送到商业细胞学服务机构进行FISH和核型分析,在所述靶向实验中Lrp5人源化LTVEC(图1)与Cas9和两种gRNA(A加F或A加E2)组合。携带小鼠Lrp5基因的细菌人工染色体(BAC)用红色荧光标记物标记并用作探针来鉴定内源性Lrp5基因座,携带人类LRP5基因的BAC用绿色荧光标记物标记并用作探针来鉴定用人类插入物靶向的染色单体。标记的BAC探针与来自靶向克隆的中期染色体涂片杂交,并通过荧光显微镜进行肉眼观察。通过用DAPI(4',6-二脒基-2-苯基吡啶)染色来观察涂片上的染色体,并通过吉姆萨染色测定每个克隆的单独核型。图4A中示出了克隆AW-D9的典型结果,发现克隆AW-D9具有正常的40XY核型(未示出)。图4A中的复合照片显示红色小鼠BAC探针信号和绿色人类BAC探针信号在小鼠染色体19(Lrp5基因的已知位置)的两个拷贝上共同定位于细胞学带B。图4C中的复合照片显示了另一个克隆(BA-D5)的相同的纯合靶向。这些结果证实,人源化LTVEC中的人类LRP5基因(图1)的91kb片段在克隆AW-D9和BA-D5的两个染色体19同源物上的预期小鼠Lrp5基因座处正确插入。相比之下,图4B中的复合照片显示,红色小鼠BAC探针信号和绿色人类BAC探针信号在小鼠染色体19的单一拷贝(实线箭头)上共同定位于细胞学带B,而只有红色小鼠BAC探针信号定位于小鼠染色体19的另一拷贝上的细胞学带B。这些结果证实,人源化LTVEC中的人类LRP5基因(图1)的91kb片段在染色体19的仅一个拷贝上的预期小鼠Lrp5基因座处正确插入(杂合靶向)。他们还(与其他未示出的对照一起)指示人类BAC探针不与小鼠Lrp5基因座交叉杂交,而是仅识别人类LRP5插入物。

[0394] 通过明显的非同源末端连接修复在两个等位基因上形成相同的CRISPR诱导的插入缺失突变的某些克隆的存在表明,F1H4杂交细胞(其由50%的129SvS6细胞株和50%的C57BL/6N细胞株构成)中发生了基因转换事件。为了深入了解使用两种gRNA时增强双等位基因靶向的机制,筛选了七种克隆,其具有靶向的纯合人源化或纯合CRISPR诱导的大缺失,之后用LTVEC以及A加F gRNA组合或A加E2 gRNA组合进行靶向。

[0395] 图5示出了被设计成检查由两种向导RNA介导的基因转换事件的测定实例。具体来说,通过分析F1H4杂交ES细胞(由50%的129SvS6细胞株和50%的C57BL/6N细胞株构成)的杂合性丢失(LOH)来检验基因转换的可能性。基因转换可通过129SvS6(129)和C57BL/6N(B6)之间的已知多态性的杂合性丢失来证明,并且由此设计PCR测定来区分这两种等位基因类型。通过被设计用于检测129和B6等位基因之间的差异的常规PCR来测定结构变体(SV)多态性。尽管图5中只示出了下面使用的SV测定中的一种,但各个SV测定的概念是相同的。引物是基于B6和129小鼠细胞株之间的结构变异(SV)设计的,并且示于表1中。引物设计条件限制为鉴定约25bp的SV,并产生约300bp的PCR产物;选择这些条件以使得通过凝胶电泳可看出任何变化。

[0396] 在对克隆进行PCR之前,对来自B6细胞株、129细胞株和F1H4 ES细胞系的野生型ES细胞的DNA进行了测定验证和优化。产生特异于B6或129等位基因的可区别PCR条带并且在使用F1H4 DNA产生这些相同的两个可区别条带方面一致的引物组被用于对克隆进行测试。对于染色体19(Lrp5基因的位置),选择六个引物组(ID 190045、ID 190061、ID 190068、ID 190030、ID 190033和ID 190013)用于通过等位基因修饰(MOA)测定法和常规PCR基因分型为“纯合靶向”或“纯合剪切”的Lrp5人源化克隆。SV PCR测定结果沿着染色体19从Lrp5基因座到染色体端粒末端隔开,范围为Lrp5基因座的约13.7Mb到约56.2Mb。来自Lrp5基因座的染色体19上的SV测定的近似距离(以Mb计)如下:测定190045为13.7,测定190061为19.0,测定190068为35.0,测定190030为37.4,测定190033为48.3,测定190013为56.2。图5中仅示出了测定190033(显示为SV 48.3),但是用于测定190045、190061、190068、190030、190033以及190013的引物示于表1中。

[0397] 对来自这些克隆的DNA以及F1H4对照DNA、129对照DNA和B6对照DNA进行PCR。通过在6%聚丙烯酰胺凝胶上电泳分级PCR产物,随后用GelRed进行染色。将产生两个条带的克隆与F1H4对照匹配,从先前的优化表明,顶带特异于129等位基因,底带特异于B6等位基因。仅产生一个带的克隆只显示B6带或仅显示129带。克隆AW-A7、AW-F10、BA-D5、BA-F2、BC-H9和BR-B4仅显示所有六种测定中的B6带,而克隆B0-A8仅显示所有六种测定中的129带。如前所述,这些克隆通过MOA和/或PCR基因分型为纯合靶向或纯合剪切,并涉及各种gRNA组合(A+F、A+E2以及B2+D)。只存在一个等位基因带表明在发生基因转换事件,如果没有转换,则会像在F1H4对照中那样仍存在两个带。

[0398] 此外,通过TaqMan[®]等位基因分型测定测定129等位基因和B6等位基因之间的单核苷酸变体(SNV)。图5中染色体19图上的SNV测定的近似位置由箭头显示,其距Lrp5基因座的距离(以Mb计)在下面给出。距Lrp5基因座的距离(以Mb计)如下:在Lrp5的着丝粒侧0.32(C2),在Lrp5的端粒侧1.2(T3),在Lrp5的端粒侧11.1(T6),在Lrp5的端粒侧13.2(T7),在Lrp5的端粒侧17.5(T8),在Lrp5的端粒侧25.8(T9),在Lrp5的端粒侧33.0(T10),在Lrp5的端粒侧38.3(T11),在Lrp5的端粒侧49.6(T13),以及在Lrp5的端粒侧57.2(T14)。129特异性探针和B6特异性探针以及引物对示于表1中。

[0399] 表3示出了非人ES细胞克隆的七个实例,这些克隆通过SV和SNV等位基因的LOH在染色体19的长臂上从Lrp5靶基因座起在端粒方向上具有明显的基因转换事件。ES细胞克隆得自将Lrp5人源化LTVEC(图1)与一种或两种gRNA组合的独立靶向实验,如所指出的那样。gRNA识别位点的位置显示在图5中的Lrp5基因图示之上(向左指向的粗箭头)。基因分型测

定表明,七个克隆中的六个具有Lrp5基因的纯合靶向人源化,一个具有纯合剪切(gRNA位点之间的大缺失)。在七个克隆中的六个中,129等位基因丢失,仅留下B6等位基因。在另一个克隆中,B6等位基因丢失,仅留下129等位基因。所有克隆对于在Lrp5基因座的着丝粒侧测定的等位基因而言保持杂合(即,C2 SNV测定的所有克隆均为杂合的B6/129)。在七个克隆中观察到的LOH表明,当LTVEC与一种或(更频繁地)两种gRNA组合时获得纯合遗传修饰的等位基因的一种机制是一个等位基因上的第一靶向遗传修饰,随后是同源定向重组基因转换事件,所述事件将靶向遗传修饰从一个染色体复制到其同系物。

[0400] 表3. 杂合性丢失测定结果。

[0401]

克隆	gRNA	Lrp5等位基因型	杂合性丢失测定(SV和SNV)
AW-A7	A+F	纯合靶向	仅检测到B6等位基因
AW-F10	A+F	纯合剪切	仅检测到B6等位基因
BO-A8	A+F	纯合靶向	仅检测到129等位基因
BA-D5	A+E2	纯合靶向	仅检测到B6等位基因
BA-F2	A+E2	纯合靶向	仅检测到B6等位基因
BC-H9	B2	纯合靶向	仅检测到B6等位基因
BR-B4	D	纯合靶向	仅检测到B6等位基因

[0402] C5(Hc) 基因座

[0403] 在另一组实验中,LTVEC被设计成形成补体组分5(C5或Hc(溶血性补体))的小鼠基因的76kb缺失,并且同时用同源人类C5基因的97kb片段置换(图6)。靶基因座包含C5(Hc)基因的止密码子的外显子2。LTVEC包含人类C5基因中侧接同源臂的97kb片段,该同源臂含有35kb和31kb的来源于小鼠C5(Hc)基因座的部分的基因组DNA,所述部分侧接意欲缺失的小鼠C5(Hc)基因的76kb序列。在单独的实验中,将C5(Hc)人源化LTVEC与编码Cas9的质粒和编码设计用于在靶向缺失的小鼠C5(Hc)基因的区域产生双链断裂的六种gRNA(A、B、C、D、E以及E2,见表1)之一的第二质粒组合。所述gRNA被设计成避免识别人类C5基因的插入部分中的任何序列。在其他实验中,我们将LTVEC和编码Cas9的质粒以及编码两种靶向要缺失的小鼠C5(Hc)基因的区域不同位点的不同gRNA的质粒组合。在一些实验中,使用靶向Ch25h基因座的对照LTVEC代替C5(Hc)人源化LTVEC。对照LTVEC(其被设计成缺失Ch25h(约1kb)的整个编码序列并将嘌呤霉素和新霉素选择盒插入Ch25h基因座)被用作在C5(Hc)基因座上选择不是靶向同源重组的耐药克隆的手段。

[0404] C5(Hc)基因的CRISPR/Cas9辅助人源化的结果示于表4中,并且与Lrp5基因的CRISPR/Cas9辅助人源化获得的结果相似。单独使用LTVEC的靶向效率在用于C5(Hc)人源化时比用于Lrp5时高(高出6.1%),但是添加Cas9和gRNA提高了所测试的六种gRNA中四种的靶向效率。与Lrp5一样,将gRNA(例如,两种gRNA)结合到C5(Hc)人源化进一步提高了总靶向效率,主要是通过增加半合子和纯合靶向事件频率实现的。我们还发现在两种等位基因上具有大CRISPR诱导缺失的ES细胞克隆(以1.8%至3.6%的频率观察到)。另外,当LTVEC靶向Ch25h基因座与两个C5(Hc)RNA组合使用时,以1.2%至6%的频率观察到在两种gRNA CRISPR RNA识别序列之间被剪切的具有纯合等位基因的克隆,表明剪切事件独立于靶基因座上的同源重组事件发生。与Lrp5一样,使用保留测定来确认正确靶向的克隆。用于此筛选的两种保留测定是使用以下引物和探针的TaqMan®测定:7140retU正向引物

CCCAGCATCTGACGACACC (SEQ ID NO:125); 7140retU反向引物GACCACTGTGGGCATCTGTAG (SEQ ID NO:126); 7140retU **TaqMan**[®]探针CCGAGTCTGCTGTTACTGTAGCATCA (SEQ ID NO:127); 7140retD正向引物CCCGACACCTTCTGAGCATG (SEQ ID NO:128); 7140retD反向引物TGCAGGCTGAGTCAGGATTTG (SEQ ID NO:129); 7140retD **TaqMan**[®]探针TAGTCACGTTTTGTGACACCCAG (SEQ ID NO:130)。

[0405] 表4. 使用单独gRNA和组合gRNA进行C5 (Hc) 基因的CRISPR/Cas9辅助人源化的筛选结果。

gRNA	gRNA 位点距靶向缺失的 5'/3'末端的距离(bp)	LTVEC	等位基因型靶向效率				
			杂合靶向 (效率%)	半合子靶向 (效率%)	纯合靶向 (效率%)	总靶向 (效率%)	纯合缺失 (效率%)
A	100 (5')	C5	16.6			16.6	
B	500 (5')	C5	14.5			14.5	
C	38200 (5')/ 37500 (3')	C5	11.4			11.4	
D	43500 (5')/ 32200 (3')	C5	7.3			7.3	
E	500 (3')	C5	4.2			4.2	
E2	100 (3')	C5	6.2			6.2	
A + C	A: 100 (5') C:37500 (3')	C5	19.6	7.1	0.6	27.3	0.6
A + C	A: 100 (5') C:37500 (3')	Ch25h	不适用	不适用	不适用	不适用	6.0
A + E2	A: 100 (5') E2: 100 (3')	C5	19.0	3.6	1.2	23.8	3.0
A + E2	A: 100 (5') E2: 100 (3')	Ch25h	不适用	不适用	不适用	不适用	1.2
无	不适用	C5	6.1			6.1	

[0408] 荧光原位杂交 (FISH) 用于证实C5 (Hc) 基因的纯合靶向人源化。将如靶向实验中所纯合靶向的、通过定量和常规PCR测定评分的ES细胞克隆送到商业细胞学服务机构以进行FISH和核型分析, 在所述靶向实验中C5 (Hc) 人源化LTVEC (图6) 与Cas9和两种gRNA组合。携带小鼠C5 (Hc) 基因的细菌人工染色体 (BAC) 用红色荧光标记物标记并用作探针来鉴定内源性基因座, 携带人类C5基因的BAC用绿色荧光标记物标记并用作探针来鉴定用人类插入物

靶向的染色单体。标记的BAC探针与来自靶向克隆的中期染色体涂片杂交,并通过荧光显微镜进行肉眼观察。通过用DAPI (4',6-二脒基-2-苯基吡啶)染色来观察涂片上的染色体,并通过吉姆萨染色测定每个克隆的单独核型。图7B中示出了克隆0-E的典型结果。图7B中的复合照片显示红色小鼠BAC探针信号和绿色人类BAC探针信号在小鼠染色体2 (C5 (Hc) 基因的已知位置) 的两个拷贝上共同定位于C5 (Hc) 基因座。这些结果证实,人源化LTVEC中的人类C5基因 (图6) 的97kb片段在克隆0-E3的两个染色体2同源物上的预期小鼠C5 (Hc) 基因座处正确插入。相比之下,图7A中的复合照片显示,红色小鼠BAC探针信号和绿色人类BAC探针信号共同定位于小鼠染色体2的单一拷贝 (实线箭头) 上,而只有红色小鼠BAC探针信号定位于小鼠染色体2的另一拷贝上的C5 (Hc) 基因座。这些结果证实,在人源化LTVEC (图6) 中的人类C5基因的97kb片段在克隆Q-E9的染色体2的仅一个拷贝上的预期小鼠C5 (Hc) 基因座处正确插入 (杂合靶向)。

[0409] Ror1基因座

[0410] 在另一组实验中,LTVEC被设计成形成小鼠Ror1 (酪氨酸-蛋白激酶跨膜受体ROR1) 基因的110kb缺失,并且同时用同源人类ROR1基因的134kb片段置换 (图8)。LTVEC包含人类ROR1基因中侧接同源臂的134kb片段,该同源臂含有41.8kb和96.4kb的来源于小鼠Ror1基因座的部分的基因组DNA,所述部分侧接意欲缺失的小鼠Ror1基因的110kb序列。在单独的实验中,将Ror1人源化LTVEC与编码Cas9的质粒和编码设计用于在靶向缺失的小鼠Ror1基因的区域产生双链断裂的六种gRNA (A、B、C、D、E以及F,见表1) 之一的第二质粒组合。所述gRNA被设计成避免识别人类ROR1基因的插入部分中的任何序列。在其他实验中,我们将LTVEC和编码Cas9的质粒以及编码两种靶向要缺失的Ror1基因内不同位点的不同gRNA的质粒组合。

[0411] Ror1基因的CRISPR/Cas9辅助人源化的结果显示于表5中,并且与Lrp5基因和C5 (Hc) 的CRISPR/Cas9辅助人源化获得的结果相似。单独使用LTVEC的靶向效率为0.3%,而添加Cas9和gRNA略微提高了所测试的六种gRNA中两种的靶向效率。结合A gRNA和F gRNA通过增加杂合和半合子靶向事件的频率将总Ror1靶向效率提高到6.3%。我们还发现在两种等位基因上具有大CRISPR诱导缺失的ES细胞克隆 (以1.6%的频率观察到)。

[0412] 表5. 使用单独gRNA和组合gRNA进行Ror1基因的CRISPR/Cas9辅助人源化的筛选结果。

gRNA	gRNA 位点距靶向缺失的 5'/3'末端的距离(bp)	等位基因型靶向效率				
		杂合靶向 (效率%)	半合子靶向 (效率%)	纯合靶向 (效率%)	总靶向 (效率%)	纯合缺失 (效率%)
[0413]	A	200 (5')	0.7		0.7	
	B	1000 (5')	0.0		0.0	
	D	54300 (5')/ 55500 (3')	0.7		0.7	
	C	54500 (5')/ 55300 (3')	0.0		0.0	
	E	1000 (3')	0.0		0.0	
	F	200 (3')	0.3		0.3	
[0414]	A + F	A: 200 (5') F: 200 (3')	4.2	2.1	6.3	1.6
	无	不适用	0.3		0.3	

[0415] Trpa1基因座

[0416] 在另一组实验中,LTVEC被设计成形成小鼠Trpa1 (瞬时受体电位阳离子通道,亚族A,成员1) 基因的45.3kb缺失,并且同时用同源人类TRPA1基因的54.5kb片段置换(图9)。LTVEC包含人类TRPA1基因中侧接同源臂的54.5kb片段,该同源臂含有41.0kb和58.0kb的来源于小鼠Trpa1基因座的部分的基因组DNA,所述部分侧接意欲缺失的小鼠Trpa1基因的45.3kb序列。在单独的实验中,将Trpa1人源化LTVEC与编码Cas9的质粒和编码设计用于在靶向缺失的小鼠Trpa1基因的区域产生双链断裂的八种gRNA(A、A2、B、C、D、E、E2以及F,见表1)之一的第二质粒组合。所述gRNA被设计成避免识别人类TRPA1基因的插入部分中的任何序列。在其他实验中,我们将LTVEC和编码Cas9的质粒以及编码两种靶向要缺失的Trpa1基因内不同位点的不同gRNA的质粒组合。

[0417] Trpa1基因的CRISPR/Cas9辅助人源化的结果显示于表6中,并且与Lrp5基因和C5(Hc)的CRISPR/Cas9辅助人源化获得的结果相似。单独使用LTVEC的靶向效率为0.3%,而添加Cas9和gRNA提高了所测试的八种gRNA中六种的靶向效率。结合B gRNA和F gRNA通过增加杂合、半合子和纯和靶向事件的频率将总Trpa1靶向效率提高到3.4%。我们还发现在两种等位基因上具有大CRISPR诱导缺失的ES细胞克隆(以0.3%的频率观察到)。

[0418] 表6. 使用单独gRNA和组合gRNA进行Trpa1基因的CRISPR/Cas9辅助人源化的筛选结果。

[0419]

		等位基因型靶向效率				
gRNA	gRNA 位点距靶向缺失的 5'/3'末端的距离(bp)	杂合靶向 (效率%)	半合子靶向 (效率%)	纯合靶向 (效率%)	总靶向 (效率%)	纯合缺失 (效率%)
A	100 (5')	1.0			1.0	
A2	500 (5')	2.1			2.1	

[0420]

B	1000 (5')	1.4			1.4	
C	25600 (5')/ 19740 (3')	1.0			1.0	
D	26970 (5')/ 18370 (3')	2.1			2.1	
E2	1000 (3')	0.0			0.0	
E	500 (3')	0.0			0.0	
F	100 (3')	0.7			0.7	
B + F	B: 1000 (5') F: 100 (3')	2.8	0.3	0.3	3.4	0.3
无	不适用	0.3			0.3	

[0421] 如这些实例所示出,与单种gRNA相比,在广泛分离的位点处使用双重向导RNA改善了杂合人源化的增强。此外,与单种gRNA相比,双重向导RNA的使用促进双等位基因事件。与使用一种gRNA的靶向相反,用两种gRNA靶向导致产生纯合靶向细胞(Hum/Hum)、纯合缺失细胞(Δ/Δ)以及半合子靶向细胞(Hum/ Δ),其中:纯合靶向细胞中的两个等位基因都具有靶向人源化;纯合缺失细胞中的两个等位基因均未被靶向人源化LTVEC,但都具有大缺失;半合子靶向细胞的其中一个等位基因具有靶向人源化,而另一个具有双重gRNA/Cas9诱导的大缺失。首先,我们发现在靶等位基因上具有精确且相同的非常大的人源化的正确靶向克隆(例如,对于靶向基因修饰而言为纯合的细胞)。虽然当我们使用一种gRNA来实现Lrp5人源化时,也观察到纯合靶向克隆,但它们的发生频率比使用两种gRNA时的频率低得多(参见表2)。同样,当使用一种gRNA实现C5(Hc)人源化或Trpa1人源化时,我们没有观察到纯合靶向,但是当使用两种具有靶向载体的gRNA时,我们确实观察到纯合靶向(参见表4和表6)。类似地,我们发现对于Lrp5靶向、C5(Hc)靶向、Ror1靶向和Trpa1靶向的基因修饰(即,它们在一个等位基因上具有精确靶向的人源化,而在另一个等位基因上具有非常大的、有时为基因消融的缺失)为半合子的正确靶向的克隆。当使用一种gRNA来实现Lrp5、C5(Hc)、Ror1或Trpa1人源化时,完全不会发生这种修饰(分别参见表2、表4、表5和表6)。

[0422] 第二,我们发现在两个靶向等位基因上由两种gRNA导向的Cas9切割事件诱导的具有相同的非常大的缺失(>45kb)的克隆(即,细胞对于靶基因座处的大的、有时为基因消融的缺失而言是纯合的)。这些类型的突变不需要针对相同基因的靶向载体。例如,如表4所示,我们已经通过将Cas9以及两种gRNA与针对不同基因(与gRNA靶向的一个基因无关)导向

的靶向载体进行结合而获得具有纯合CRISPR诱导的缺失的ES细胞。因此,由两种gRNA导向的Cas9核酸酶可以在不添加靶向载体的情况下引起细胞中的大缺失。在这种情况下,由表达药物抗性基因的载体提供的瞬时或稳定的药物选择可通过富集已经接受DNA的ES细胞来促进罕见的纯合缺失克隆的分离。

[0423] 实例2.组合gRNA引起的大缺失的分析。

[0424] 组合gRNA引起的大缺失的等位基因结构

[0425] 对包含由两种gRNA导向的Cas9切割事件诱导的大缺失的克隆进行附加序列分析(参见表7)。这些大缺失似乎与相同基因座上的LTVEC导向的同源重组事件无关,因为当我们将gRNA与Lrp5LTVEC或靶向近30Mb远的Ch25h基因的一种LTVEC相结合时,我们以大致相同的频率在Lrp5基因座处获得了大缺失(数据未显示)。为了表征大缺失,我们从四组人源化中的37个克隆、15个半合子和22个具有双等位基因大缺失的克隆进行跨缺失PCR,并对PCR产物的各个克隆进行测序。这些序列确认了大缺失,其范围从38kb到109kb。ES细胞克隆中的两种(Lrp5克隆AW-A8和BP-D3)在预测的Cas9切割位点之间已经完全修复了精确的缺失(68.2kb),而一种克隆(Hc克隆P-B12)除了38.1kb的缺失外还具有单碱基对插入。ES细胞克隆中的二十七种具有延伸超出Cas9切割位点的缺失,与通过非同源末端连接(NHEJ)的不精确修复一致。ES细胞克隆中的其余七种具有将明显的NHEJ诱导的缺失和插入(例如,Lrp5克隆BP-F6和Hc克隆O-E4)组合的突变,其中四种具有大于200bp的插入,我们可以确定这些插入映射到其源基因组基因座(数据未示出)。在Lrp5克隆B0-E9中的210bp插入相对于位于着丝粒方向(染色体19+,3589138-3589347)的gRNA F靶位点外约2600bp的相同序列处于反向取向。此序列存在于Lrp5 LTVEC的长3'同源臂中。Lrp5克隆BP-F6和BP-G7源自我们将Lrp5 gRNA A和F与Cas9和LTVEC组合的实验,其中LTVEC靶向在端粒方向上与Lrp5相距30Mb的Ch25h基因。克隆BP-F6具有似乎衍生自Ch25h LTVEC的一端的266bp插入,其由与连接到163bp片段的载体主链的部分相同的103bp片段组成,其中163bp片段与在Ch25h附近并且存在于LTVEC的长臂(染色体19+,34478136-34478298)上的序列相同;将该片段以相对于内源染色体序列反向的取向插入缺失。Hc克隆O-E4具有254bp插入,其相对于在距离gRNA A识别位点近3.1kb的缺失序列中发现的相同序列是反向的。Hc克隆S-D5中的1304bp插入由两个片段组成:1238bp片段,其与距离预测的gRNA E2导向的Cas9切割位点约1.4kb的缺失序列中发现的相同序列具有相同的取向;以及第二个66bp片段,其是在gRNA E2切割位点外25bp的相同序列的反向取向的复制。

[0426] 表7.组合gRNA引起的大缺失的等位基因结构。

[0427]

基因	ES 细胞 克隆	基因型 ¹	gRNA	靶向缺失内的 位置(bp)	缺失的 尺寸(kb)	缺失的附 加序列(bp)	插入(bp)	PCR 克隆
<i>Lrp5</i>	AW-A8	Δ/Δ	A + F	5'-50/50-3'	68.2	-	-	40
	BO-E9	Δ/Δ				12	210	17
	BP-D3	Δ/Δ				-	-	11
	BP-F6	Δ/Δ				30	266	6
	BP-G7	Δ/Δ				77		9
	BA-A7	Δ/Δ	A + E2	5'-50/1,000-3'	67.3	7		19
	BA-C7	Δ/Δ				84		32
<i>Hc</i>	N-A11	Δ/Δ	A + C	5'-100/38,200-3'	38.1	14		12
	N-D4	Δ/Δ				10		15
	N-D11	Hum/ Δ				20		10
							10	
	N-E1	Hum/ Δ				10		13
	N-E9	Hum/ Δ				20		16
	O-C5	Hum/ Δ				31		21

[0428]		O-D2	Hum/ Δ				5		12
		O-E4	Hum/ Δ				19	254	18
		O-E5	Hum/ Δ				35	2	16
		O-E6	Hum/ Δ				6		17
		O-F11	Hum/ Δ				12	7	18
		O-F12	Hum/ Δ				41		6
							35		1
		P-B12	Δ/Δ					1	7
		P-C12	Δ/Δ				20		15
		P-D1	Δ/Δ				33		10
		P-G8	Δ/Δ				5		2
		Q-F5	Hum/ Δ	A + E2	5'-100/100-3'	75.6	3	3	15
		Q-F10	Δ/Δ				46		13
		R-A5	Δ/Δ				18		14
		R-A7	Δ/Δ				37		15
		R-A9	Hum/ Δ				261		8
		R-C8	Hum/ Δ				180		11
		R-D12	Hum/ Δ				182		10
		R-F11	Hum/ Δ				19		11
							122		11
		S-A11	Δ/Δ				46		1
		S-D5	Δ/Δ				216	1304	8
	<i>Ror1</i>	Y-B5	Δ/Δ	A + F	5'-200/200-3'	109	18		6
		Y-C7	Δ/Δ				23		7
		Y-E1	Δ/Δ				12		3
	<i>Trpa1</i>	AD-C7	Δ/Δ	B + F	5'-1,000/100-3'	44.6	30		8

[0429] 1 Hum/+, 导致杂合基因型的两个天然等位基因之一的靶向人源化; Hum/ Δ , 导致半合子基因型的双等位基因修饰, 其中一个等位基因具有靶向人源化, 另一个等位基因具有Cas9-gRNA诱导的大缺失; Hum/Hum, 导致纯合基因型的双等位基因修饰, 其中两个等位基因均具有靶向人源化; Δ/Δ , 双等位基因修饰, 其中两个等位基因均具有Cas9-gRNA诱导的大缺失。

[0430] 纯合等位基因发生基因转换的证据

[0431] 具有双等位基因大缺失的22个ES细胞克隆中有21个仅具有单一的独特序列(表

7),表明它们是纯合等位基因。对于Hc克隆S-A11,我们在12个PCR克隆中的11个中发现了相同的序列。具有不同序列的单个克隆可能表明两个不同的缺失等位基因,但是我们还发现两个Hc半合子克隆N-D11和O-F12的结果相同。多个克隆中不同纯合缺失等位基因表明它们可能由基因转换机制引起,其中一条染色体上的缺失作为同源染色体上Cas9切割的同源重组修复的模板。我们利用VGF1 ES细胞系的129S6SvEvTac (129) 和C57BL/6NTac (B6) F1混合物(Poueymirou et al. (2007) Nat. Biotechnol. 25:91-99 (Poueymirou等人,2007年,《自然生物技术》,第25卷,第91-99页);Valenzuela et al. (2003) Nat. Biotechnol. 21:652-659 (Valenzuela等人,2003年,《自然生物技术》,第21卷,第652-659页))来测定基因转换,所述基因转换为各品系之间在染色体19上的Lrp5基因座(参见图5中的五次SV测定和下文所用的十次SNV测定)和染色体2上的Hc基因座(未示出)附近的结构变异(SV)和单核苷酸变异(SNV)的杂合性丢失(Lefebvre et al. (2001) Nat. Genet. 27:257-258 (Lefebvre等人,2001年,《自然遗传学》,第27卷,第257-258页))。为了证实任何杂合性丢失不是整个染色体丢失的结果,我们在129细胞株和B6细胞株之间相同的位点进行了染色体拷贝数(CCN)测定。对于Lrp5人源化或缺失等位基因,我们测定了位于自在端粒酶方向上距离Lrp5 1.2Mb处至染色体19的长臂末端的多个SV和SNV(图5)。由于Lrp5位于接近着丝粒的位置,我们发现基因的着丝粒侧没有SV,只有一个SNV。对于Hc,我们能够测定染色体2上基因任一侧上的多个SV和SNV(未示出)。六个Lrp5克隆的结果示出于图10A至图10E和图11A至图11C中。

[0432] 图10A至图10E示出了5个SV测定的结果,其位置范围为距离Lrp5 13.7Mb处到距离长臂端粒末端附近56.7Mb处。对于129、B6和VGF1对照中的129(较大)和B6(较小)等位基因,五个SV测定产生两种不同大小的产物。染色体19图上的SV测定的近似位置示出于图5中(参见测定SV 13.7、测定SV 20.0、测定SV 36.9、测定SV 48.3和测定SV 56.7)。测定数表示端粒到Lrp5的Mb数。用于这些测定的引物示于表1中,并且结果示于图10A至图10E中。BC-H9 (Lrp5^{Hum/Hum}, gRNA B2) 和BR-B4 (Lrp5^{Hum/Hum}, gRNA D) 的两个克隆显示了保留所有B6 SV等位基因的杂合性丢失,而第三个克隆B0-A8 (Lrp5^{Hum/Hum}, gRNA A+F) 保留了所有129等位基因。另外三个克隆B0-F10 (Lrp5^{Hum/Hum}, gRNA A+F)、B0-G11 (Lrp5^{Hum/Hum}, gRNA A+F) 和BP-G7 (Lrp5^{Δ/Δ}, gRNA A+F) 保持杂合。

[0433] 此外,通过TaqMan[®]等位基因分型测定测定129等位基因和B6等位基因之间的单核苷酸变体(SNV)。图5中染色体19图上的SNV测定的近似位置由下方具有测定数的箭头显示,并且其距离Lrp5基因座的距离(以Mb计)在下面给出。距Lrp5基因座的距离(以Mb计)如下:在Lrp5的着丝粒侧0.32 (C2),在Lrp5的端粒侧1.2 (T3),在Lrp5的端粒侧11.1 (T6),在Lrp5的端粒侧13.2 (T7),在Lrp5的端粒侧17.5 (T8),在Lrp5的端粒侧25.8 (T9),在Lrp5的端粒侧33.0 (T10),在Lrp5的端粒侧38.3 (T11),在Lrp5的端粒侧49.6 (T13),以及在Lrp5的端粒侧57.2 (T14)。129特异性探针和B6特异性探针以及引物对示于表1中。通过SV测定显示端粒杂合性丢失(LOH)的三个克隆(BC-H9、B0-A8和BR-B4)的结果显示在图11A至图11C中。SNV测定(图11A至图11C和数据未示出)证实了Lrp5的端粒侧(SNV 1.2和SNV 57.2;分别参见图11B和图11C)上跨染色体19长臂的基因转换事件,但是SNV 0.32测定(参见图11A)显示,所有克隆对于在距离着丝粒侧的Lrp5 320Kb处的等位基因而言保持杂合。在测定的24个Lrp5^{Hum/Hum}或Lrp5^{Δ/Δ}克隆中,我们发现六个具有Lrp5的端粒侧上跨染色体19整个长臂的杂合性丢失的证据。五个克隆(四个Lrp5^{Hum/Hum}和一个Lrp5^{Δ/Δ})从杂合转变为纯合B6,而第六个

克隆 (Lrp5^{Hum/Hum}) 转变为纯合129。CCN测定证明了染色体19的两个拷贝的保留。21个Hc纯合克隆的相似杂合性丢失测定表明,其中两个R-E2 (Hc^{Hum/Hum}, gRNA A+F) 和R-E8 (Hc^{Δ/Δ}, gRNAs A+F) 在Hc基因的端粒侧的所有SV和SNV显示出纯合129的杂合性丢失,同时保持着丝粒侧上所有等位基因的杂合性。CCN测定表明染色体2没有丢失。

[0434] 我们的结果首次证明,CRISPR/Cas9可以增强超过100kb的大单步人源化的同源介导修复,其增加了大规模基因组改造的可能性。结合LTVEC和gRNA/Cas9的最显著和最意想不到的益处是它们促进纯合靶向人源化的能力。尽管在其他CRISPR/Cas9实验中已经报道了双等位基因突变和纯合靶向事件,但是这些基因修饰和插入中的大多数的数量级已经比我们的人源化等位基因小。在使用CRISPR/Cas9之前,我们从未发现LTVEC纯合靶向;当我们组合靶向单独基因的多个LTVEC时,我们也没有看到同时靶向多个基因。鉴于这个实验, gRNA/Cas9诱导的纯合靶向表明,一个等位基因上的初始靶向事件(而不是分别靶向两个等位基因的两个LTVEC) 可以用作由一个或多个Cas9切割促进的另一等位基因的同源转换的模板。双重gRNA/Cas9诱导的大双等位基因缺失也是纯合的(表7),这一发现为基因转换机制提供了进一步的支持。

[0435] 杂合性丢失的测定(图5)表明,覆盖靶向基因的端粒侧上的染色体大片段的多重等位基因的大规模基因转换是一些纯合人源化和大量缺失的原因。这种类型的远程定向基因转换与细胞周期的G2期中同源染色体的复制染色单体之间的有丝分裂重组一致(Lefebvre et al. (2001) Nat. Genet. 27:257-258 (Lefebvre等人, 2001年,《自然遗传学》,第27卷,第257-258页)) (图12)。虽然它仅解释了少数纯合事件,但是这种机制可以提供一种手段,通过其可以使用gRNA/Cas9切割来促进在染色体的大部分上多个等位基因从杂合到纯合的大规模转换。然而,大多数纯合事件看起来是局部基因转换的结果,其机制值得进一步研究。

[0436] 对三个克隆的分析提供了长片段定向基因转换的进一步证据,其中所述三个克隆是使用编码Lrp5gRNA A和F的质粒、编码Cas9的质粒以及靶向在端粒方向上距离Lrp5 30Mb的Ch25h基因的LTVEC对F1H4杂交ES细胞(其由50%的129SvS6细胞株和50%的C57BL/6N细胞株构成)进行电穿孔后获得的。在两种gRNA(距5'末端500bp以及3'末端2kb)之间的预测缺失内使用TaqMan[®]测定进行初步筛选后,筛选出初始评分为野生型的三个克隆,但是后续进行的TaqMan[®]等位基因分型测定法测定的129和B6等位基因之间的单核苷酸变体(SNV)令人惊讶地表明了杂合性丢失。所用的SNV测定结果是一个着丝粒测定结果(SNV 0.32)和两个端粒测定结果(SNV 1.2和SNV 57.2)(参见图5)。如表8所示,着丝粒SNV测定结果(0.32Mb)证实了所有三个克隆中的杂合性保留。然而,两种端粒SNV测定结果都表明,BP-E7和BP-H4对于129等位基因而言是纯合的,并且两种端粒SNV测定结果都表明,BP-E6对于B6等位基因而言是纯合的。所有三个克隆都显示了染色体19的两个拷贝的保留,并且所有三个克隆都是LTVEC靶向的转基因(即,靶向Ch25h基因座)。这些结果提供了使用靶向CRISPR/Cas9切割强制纯合的可能性。

[0437] 表8. SNV等位基因分型测定的筛选结果。

[0438]

克隆	SNV 0.32	SNV 1.2	SNV 57.2
BP-E7	129/B6	129/129	129/129

BP-H4	129/B6	129/129	129/129
BP-E6	129/B6	B6/B6	B6/B6

[0439] 几种可能的机制可以解释在小鼠F1H4杂交ES细胞(其由50%的129SvS6细胞株和50%的C57BL/6N细胞株构成)中的CRISPR/Cas9辅助LTVEC人源化实验中观察到的结果(参见图16A至图16F)。这种机制可以通过有丝分裂交换进行的相互染色单体交换(参见图16A至图16C)或通过断裂诱导的复制进行的染色单体拷贝(参见图16D至图16E)发生。在任一种情况下,可能发生杂合修饰,其中129染色体或B6染色体在基因组复制前由LTVEC靶向(参见图16A和图16D)。或者,单个129染色单体或单个B6染色单体可以在基因组复制后由LTVEC靶向,随后进行染色单体间的基因转换(参见图16B和图16E)。或者,靶基因组基因座缺乏LTVEC靶向,但是Cas9切割可发生在129或B6染色体上(参见图16C和图16F)。后一种可能性可解释通过BP-E7、BP-H4和BP-E6克隆看到的结果。潜在的结果如图16A至图16F所示。对于图16F,如果Cas9切割了129染色单体,也可观察到保留B6等位基因的杂合性丢失(LOH)。在上述实验中,已经观察到杂合性丢失事件,其导致两个等位基因被靶向(Hum/Hum)或两个等位基因都是野生型等位基因(+/+)。

[0440] 实例3.LTVEC同源臂尺寸对靶向效率的影响

[0441] 为了确定同源臂尺寸对靶向效率的影响,对被设计形成补体组分5(C5或Hc(溶血性补体))的小鼠基因的76kb缺失并且同时用同源人类C5基因的97kb片段置换的两种LTVEC进行比较(图13)。靶基因座包含C5(Hc)基因的止密码子的外显子2。第一LTVEC包含人类C5基因中侧接同源臂的97kb片段,该同源臂含有35kb和31kb的来源于小鼠C5(Hc)基因座的部分的基因组DNA,所述部分侧接意欲缺失的小鼠C5(Hc)基因的76kb序列(参见图13的标记了LTVEC的靶向载体)。第二LTVEC包含人类C5基因中侧接同源臂的97kb片段,该同源臂含有各为5kb的来源于小鼠C5(Hc)基因座的部分的基因组DNA,所述部分侧接意欲缺失的小鼠C5(Hc)基因的76kb序列(参见图13的标记了sTVEC的靶向载体)。

[0442] 在单独的实验中,将C5(Hc)人源化LTVEC与编码Cas9的质粒和编码设计用于在靶向缺失的小鼠C5(Hc)基因的区域产生双链断裂的六种gRNA(A、B、C、D、E以及E2,见表1)中一种或两种的第二质粒组合。所述gRNA被设计成避免识别人类C5基因的插入部分中的任何序列。

[0443] C5(Hc)基因的CRISPR/Cas9辅助人源化的结果示出在表9中。单独的第一LTVEC(35kb和31kb的同源臂)的靶向效率高于单独的第二LTVEC(5kb和5kb的同源臂)的靶向效率。然而,当与gRNA A和E2组合时,各个LTVEC的总靶向效率几乎相同(参见表9),表明当将CRISPR/Cas9与LTVEC靶向组合使用来靶向C5(Hc)基因座时,5kb的同源臂尺寸(即总和为10kb)足以促使观察到的靶向效率增加。

[0444] 表9.使用具有不同同源臂尺寸的LTVEC进行C5(Hc)基因的CRISPR/Cas9辅助人源化的筛选结果。

	gRNA	LTVEC	LOA 筛选的靶向效率(%)			
			杂合	半合子	纯合	总计
[0445]	A + E2	1	12.8	6.4	3.8	23.0
	无	1	8.3	0	0	8.3
	A + E2	2	7.2	13.1	4.3	24.6
	无	2	5.4	0	0	5.4

[0446] 实例4.CRISPR RNA识别序列之间的较短距离对靶向效率的影响

[0447] 为了确定CRISPR RNA识别序列和切割位点之间的较短距离对靶向效率的影响,将LTVEC设计成形成用于胞苷一磷酸-N-乙酰神经氨酸羟化酶(Cmah)小鼠基因的18.2kb缺失,并且同时用插入片段替换,所述插入片段包含lacZ报告基因和潮霉素抗性选择盒。使用两种gRNA与LTVEC靶向间隔紧密的序列(图14)。靶基因座包含Cmah基因的前5个外显子。LTVEC包含侧接同源臂的8.8kb lacZ-hyg^r插入物,该同源臂含有120kb和157kb的来源于小鼠Cmah基因座的部分的基因组DNA,所述部分侧接意欲缺失的小鼠Cmah基因的18.2kb序列。将LTVEC与编码Cas9的质粒和被设计成在靶向缺失的小鼠Cmah基因的区域5'末端附近形成双链断裂的两种gRNA(A和B)的质粒组合。两种gRNA靶向意欲缺失的序列的5'末端的ATG附近的间隔紧密的序列,其中靶向切割位点间隔为27bp(参见图15)。用两种gRNA导向的Cas9切割产生27bp的平头末端切除序列。单独使用LTVEC作为对照。

[0448] 通过对缺失内的序列以及药物选择盒和人类基因插入物内的序列进行等位基因修饰测定(Valenzuela et al. (2003) Nat. Biotechnol. 21:652-659 (Valenzuela等人, 2003年,《自然生物技术》,第21卷,第652-659页);Frendewey et al. (2010) Methods Enzymol. 476:295-307 (Frendewey等人,《酶学方法》,2010年,第476卷,第295-307页))来筛选耐药ES细胞克隆,以用于靶向人源化。如果它们丢失了两个内源性小鼠基因序列中的一个,并且获得了lacZ-hyg^r插入片段的一个拷贝,则将克隆评分为正确靶向。此外,使用称为保留测定(retention assay)的识别LTVEC同源臂中序列的实时PCR测定,来验证LTVEC到小鼠基因组内的正确靶向。确定这些保留测定结果的拷贝数提供了进一步的说明,来帮助区分保留拷贝数为2的正确靶向ES克隆。在克隆中靶小鼠基因座的Cas9诱导的大缺失与基因组中其他地方的LTVEC随机整合同时发生的情况下,保留测定结果的拷贝数为3(或更多)。成对的gRNA在靶小鼠基因座处产生Cas9介导的大缺失的能力意味着,先前所述的标准LOA测定法和GOA测定法可以通过保留测定来增强,以提供进一步的说明并验证正确的靶向。因此,保留测定被设计成结合LOA测定法和GOA测定法使用。

[0449] Cmah靶向实验的结果汇总在表10中。在单独使用LTVEC的对照靶向实验中,5.4%(3/56)的筛选的克隆具有杂合的(Het)缺失-置换突变;95%的克隆在Cmah基因座处保持为野生型(WT)。在CRISPR靶向实验中,除了几种WT克隆之外,我们观察到五种不同的突变等位基因类型。我们观察到三种类型的LTVEC靶向等位基因:(1) Het; (2) Hom(纯合缺失-置换);以及(3) Hemi(一个等位基因上的缺失-置换和另一个等位基因上的gRNA/Cas9诱导的突变)。这三种类型占有所有筛选的克隆的43.5%(106/244)。与单独的LTVEC相比,我们观察到Cmah基因靶向(其中至少一个等位基因被靶向)增强了8倍。我们还观察到两种仅携带gRNA/

Cas9-插入缺失突变的等位基因：(1) Het, 其中我们检测到两个WT等位基因中一者上的插入缺失；以及(2) 双等位基因标记突变, 其可以是纯合的 (Hom) 或半合子的 (Hemi)。仅3.7%的筛选的克隆在Cmah基因座处保持WT, 没有可检测到的突变。总的来说, 当使用gRNA A和B的组合时, 超过94%的克隆具有Cas9诱导的突变。

[0450] 表10.Cmah靶向的筛选结果。

对照靶向 (单独的 LTVEC)				CRISPR 靶向: LTVEC + Cas9 + gRNAs A + B			
突变	等位基因型	基因型	所筛选克隆%	突变	等位基因型	基因型	所筛选克隆%
缺失-置换	Het	<i>lacZ</i> /+	5.4	缺失-置换	Het	<i>lacZ</i> /+	2.5
缺失-置换	Hom	<i>lacZ</i> / <i>lacZ</i>	0	缺失-置换	Hom	<i>lacZ</i> / <i>lacZ</i>	15
缺失-置换/ 插入缺失	Hemi	<i>lacZ</i> /插入缺失	0	缺失-置换/ 插入缺失	Hemi	<i>lacZ</i> /插入缺失	26
插入缺失	Het	插入缺失/+	0	插入缺失	Het	插入缺失/+	11
插入缺失	Hom 或 Hemi	插入缺失/插入缺失	0	插入缺失	Hom 或 Hemi	插入缺失/插入缺失	42
无	WT	+/+	95	无	WT	+/+	3.7

[0452] 实例5. 在1细胞期胚胎中使用成对的gRNA的大剪切

[0453] 为了在1细胞期胚胎中实现大靶向缺失, 设计实验以形成小鼠Lrp5 (低密度脂蛋白受体相关蛋白5) 基因中编码胞外域的部分的68kb缺失, 并且任选地同时通过使用具有侧接两个60-核苷酸同源臂的4核苷酸插入物的单链DNA供体序列 (长度为124核苷酸) 来用4-核苷酸插入替换。当插入靶基因座时, 4-核苷酸插入物产生限制性内切酶位点。在单独的实验中, 通过细胞质注射 (CI) 递送蛋白质形式的Cas9蛋白质, 或通过原核注射 (PNI) 或电穿孔 (EP) 以mRNA形式递送Cas9。将Cas9与两种被设计成在靶向缺失的小鼠Lrp5基因区域内形成双链断裂的gRNA (A+F) 组合, 并且任选地具有同源重组供体。以RNA形式注射gRNA。然后评估所得到的单等位基因和双等位突变的频率。

[0454] 结果汇总在表11中, 其包括在两个向导RNA的靶位点之间的NHEJ介导的缺失或通过与ssDNA供体的同源介导修复辅助的缺失。当成对的向导RNA和Cas9通过细胞质注射与ssDNA供体一起引入时, 观察到双等位基因突变。在每个观察到的双等位基因突变中, 一个染色体通过NHEJ介导的缺失修饰, 一个染色体通过HDR辅助的缺失修饰。这些结果表明细胞质mRNA压电注射导致与供体的纯合重组的潜力的一致同源介导修复。

[0455] 表11.Cas9递送方法比较。

靶	HR 供体	递送	Cas9	浓度 (mg)Cas9/sgRNA/ 供体	单等位基因突变		双等位基因突变	
					NHEJ 剪切	HDR 剪切	NHEJ 剪切	HDR 剪切
[0456]	Lrp5-68kb 剪切(A+F sgRNA)	-	PNI 蛋白	50/60/0	7%	不适用	0%	不适用
			CI mRNA	100/100/0	15%	不适用	不适用	不适用
			EP mRNA	250/150/0	3%	不适用	0%	不适用
	Lrp5-68kb 剪切(A+F sgRNA)	+	PNI 不适用	不适用	不适用	不适用	不适用	不适用
			CI mRNA	100/50/100	16.6%	19.4%	5.5%	5.5%
			EP mRNA	250/200/250	1.4%	2.1%	0%	0%

[0457] 实例6.用于区分靶向插入和转基因插入之间的保留测定以及区分靶向缺失和延伸超出靶向区域的缺失之间的保留测定

[0458] 标准等位基因修饰 (MOA) 筛选策略 (参见例如图17A) 通过将每个样品的四个生物学重复Ct值的平均值与所有样品的Ct中值进行比较来确定 **TaqMan[®]** 拷贝数。对于等位基因丢失, 针对正在靶向缺失的靶基因组基因座区域的上游 (mTU) 和下游 (mTD) 区域使用 **TaqMan[®]** 探针。对于等位基因获得, 针对新霉素抗性盒使用 **TaqMan[®]** 探针。然而, 可以针对核酸插入片段的任何区域设计这样的探针。对于二倍体杂合靶向克隆, mTU、mTD和Neo探针各自的 **TaqMan[®]** 拷贝数应为1。对于二倍体纯合靶向克隆, mTU和mTD各自的 **TaqMan[®]** 拷贝数应为0, Neo的 **TaqMan[®]** 拷贝数应为2。同样, 对于二倍体未靶向克隆, mTU和mTD各自的 **TaqMan[®]** 拷贝数应为2, Neo的 **TaqMan[®]** 拷贝数应为0。对于二倍体杂合剪切克隆, mTU和mTD的 **TaqMan[®]** 拷贝数应为1, Neo的拷贝数应为0。对于二倍体纯合剪切克隆, mTU、mTD和Neo各自的 **TaqMan[®]** 拷贝数应为0。

[0459] 然而, 因为成对gRNA可在靶基因组基因座处形成较大Cas介导的缺失, 它可有助于增强标准LOA测定和GOA测定来证实通过LTVEC进行的正确靶向。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诱导的大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会包括选择盒和LTVEC的相邻区域, 但不包括LTVEC的更远侧区域。例如, 如果LOA测定和GOA测定用于评估LTVEC靶向整合, 并且GOA测定利用针对选择盒的探针, 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进行的正确靶向, 可单独使用保留测定, 或者将保留测定与LOA测定和/或GOA测定结合使用。

[0460] 当进行 **TaqMan[®]** 保留测定时, 使用与5' 同源臂的5' 靶序列相对应的上游探针 (retU探针) 以及3' 同源臂的3' 靶序列相对应的下游探针 (retD探针) (参见图17B, 其示出将

TaqMan[®] 保留测定与GOA和LOA测定组合使用以使用新霉素选择筛选CRISPR/Cas9辅助的人源化)。图17B还示出了核酸插入物中不同的探针如何用于GOA测定(参见上游hTU探针和下游hTD探针)。不同类型的靶向修饰和转基因插入的GOA测定、LOA测定和保留测定结果示于表12中。

[0461] 表12. 不同修饰类型的GOA测定、LOA测定和保留测定的预测拷贝数读数。

[0462]

修饰类型	retU	mTGU	mTM	mTGD	retD	Neo
纯合靶向	2	0	0	0	2	2
杂合靶向	2	1	1	1	2	1
杂合靶向+杂合剪切	2	0	0	0	2	1
杂合剪切	2	1	1	1	2	0
具有转基因插入的杂合剪切	2	1	1	1	3*	1*
纯合剪切	2	0	0	0	2	0
具有转基因插入的纯合剪切	2	0	0	0	3*	1*

[0463] TaqMan保留测定也可与LOA测定联合使用,以使用配对的gRNA筛选CRISPR/Cas9辅助缺失(参见图17C)。在这种测定中,retU和retD的拷贝数在所有情况下都应保持为2。拷贝数小于2表明Cas9介导的大缺失超出了正在靶向缺失的区域。LOA测定和不同类型的剪切相关修饰的保留测定结果示于表13中。

[0464] 表13. 不同剪切修饰的LOA测定和保留测定的预测拷贝数读数。

[0465]

修饰类型	retU	mTGU	mTM	mTGD	retD
纯合剪切	2	0	0	0	2
杂合剪切	2	1	1	1	2
具有附加下游缺失的杂合剪切	2	1	1	1	1*

[0466] 实例7. 使用四种向导RNA进行CRISPR/Cas9介导的靶向。

[0467] 为了实现修饰的约900kb区域的小鼠免疫球蛋白重链基因座的精确单步缺失,并且用侧接1oxP位点的P_{gk}-Neo插入物(磷酸甘油酸激酶I启动子可操作地连接到新霉素磷酸转移酶基因)同时置换,我们通过电穿孔向小鼠ES细胞中引入以下核酸分子:(1)LTVEC;(2)编码Cas9核酸内切酶的质粒;以及(3)编码四种CRISPR单向导RNA(gRNA)的一种或多种质粒。在各个实验中,LTVEC是线形的。靶向修饰的基因座是用人类对应物置换具有可变区基因片段(V_H、D_H、J_H)的约900kb的小鼠免疫球蛋白重链基因座区域(参见图18)。LTVEC包含P_{gk}-Neo插入片段,该插入片段的长度约为2kb,其侧接19kb的5'同源臂和13kb的3'同源臂,所述同源臂被设计成引导同源重组事件,所述同源重组事件缺失约900kb的靶基因座区域并插入药物选择盒,该药物选择盒引导赋予G418抗性的新霉素磷酸转移酶表达。

[0468] 在使用的四种gRNA中,两种引导靶基因座5'末端附近的Cas9切割(图18中的5' gRNA_I和5' gRNA_II),两种引导靶基因座3'末端附近的Cas9切割(图18中的3' gRNA_I和3' gRNA_II)。5' gRNA_I和5' gRNA_II靶序列彼此间隔约150bp,3' gRNA_I和3' gRNA_II靶序列重叠,其中3' gRNA_II靶位点相对于3' gRNA_I靶位点偏移1bp。

[0469] 接受LTVEC并将其并入基因组的ES细胞能够在组织培养皿上于含有抗生素药物的生长培养基中生长并形成集落。我们挑取耐药性集落并通过等位基因修饰法(Valenzuela

et al. (2003) Nat. Biotech. 21:652-660 (Valenzuela等人, 2003年,《自然-生物技术》,第21卷,第652-660页); 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); 这些文献全文以引用方式并入本文) 筛选它们,以鉴定具有正确靶向的人源化等位基因的克隆(参见下表14)。此外,使用称为保留测定的识别LTVEC同源臂中序列的实时PCR测定,来验证LTVEC到小鼠基因组内的正确靶向(参见下表14)。

[0470] 表14. 用于证实使用LTVEC和4种gRNA靶向的探针。

探针	测定法	SEQ ID NO	距离	Kb
5' IgH 臂 1	保留	148	5' gRNA	1.5
5' IgH 臂 2	保留	173	5' gRNA	0.315
[0471] hIgH31	LOA	136	5' gRNA	154
hIgH1	LOA	142	5' gRNA	747
hIgH1	LOA	142	3' gRNA	116
hIgH9	LOA	139	3' gRNA	3.2
3' IgH 臂 1	保留	176	3' gRNA	0.484
3' IgH 臂 2	保留	157	3' gRNA	1.2
[0472] mIgM-398	保留	151	3' gRNA	6.4
mIgM-1045	保留	154	3' gRNA	7

[0473] 在所得到的靶向ES细胞中,缺失了约900kb的区域并用两种等位基因中的Pgk-Neo插入物替换(参见图18)。以出乎意料地高的效率实现了这种大缺失和替换(双等位基因缺失的效率为约1.2%)。

- [0001] 序列表
- [0002] <110> 瑞泽恩制药公司
- [0003] 安德鲁·J·墨菲
- [0004] 大卫·弗伦杜威
- [0005] 卡曼·维纳斯·莱
- [0006] 沃基特克·奥尔巴赫
- [0007] 古斯塔沃·德罗格特
- [0008] 安东尼·加戈利亚地
- [0009] 大卫·M·巴伦苏埃拉
- [0010] 维拉·佛洛妮娜
- [0011] 林恩·麦克唐纳
- [0012] 乔治·D·扬科波洛斯
- [0013] <120> 使用成对向导RNA进行靶向遗传修饰的方法和组合物
- [0014] <130> 57766-472225
- [0015] <150> US 62/083,005
- [0016] <151> 2014-11-21
- [0017] <150> US 62/182,314
- [0018] <151> 2015-06-19
- [0019] <150> US 62/211,421
- [0020] <151> 2015-08-28
- [0021] <160> 176
- [0022] <170> 适用于Windows的FastSEQ 4.0版
- [0023] <210> 1
- [0024] <211> 80
- [0025] <212> RNA
- [0026] <213> 人工序列
- [0027] <220>
- [0028] <223> gRNA
- [0029] <400> 1
- [0030] guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguaaucaac uugaaaaagu 60
- [0031] ggcaccgagu cggugcuuuu 80
- [0032] <210> 2
- [0033] <211> 42
- [0034] <212> RNA
- [0035] <213> 人工序列
- [0036] <220>
- [0037] <223> gRNA
- [0038] <400> 2

[0039] guuuuagagc uagaaauagc aaguuaaaaau aaggcuaguc cg 42
[0040] <210> 3
[0041] <211> 30
[0042] <212> RNA
[0043] <213> 人工序列
[0044] <220>
[0045] <223> crRNA
[0046] <400> 3
[0047] guuuuagagc uagaaauagc aaguuaaaaau 30
[0048] <210> 4
[0049] <211> 33
[0050] <212> RNA
[0051] <213> 人工序列
[0052] <220>
[0053] <223> crRNA
[0054] <400> 4
[0055] guuuuagagc uagaaauagc aaguuaaaaau aag 33
[0056] <210> 5
[0057] <211> 26
[0058] <212> RNA
[0059] <213> 人工序列
[0060] <220>
[0061] <223> crRNA
[0062] <400> 5
[0063] gaguccgagc agaagaagaa guuuua 26
[0064] <210> 6
[0065] <211> 12
[0066] <212> RNA
[0067] <213> 人工序列
[0068] <220>
[0069] <223> tracrRNA
[0070] <400> 6
[0071] aaggcuaguc cg 12
[0072] <210> 7
[0073] <211> 50
[0074] <212> RNA
[0075] <213> 人工序列
[0076] <220>
[0077] <223> tracrRNA

[0078] <400> 7
[0079] aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 50
[0080] <210> 8
[0081] <211> 23
[0082] <212> DNA
[0083] <213> 人工序列
[0084] <220>
[0085] <223> CRISPR RNA识别序列
[0086] <220>
[0087] <221> misc_feature
[0088] <222> (2) ... (21)
[0089] <223> n=a、t、c或g
[0090] <400> 8
[0091] gnnnnnnnnnn nnnnnnnnnnn ngg 23
[0092] <210> 9
[0093] <211> 23
[0094] <212> DNA
[0095] <213> 人工序列
[0096] <220>
[0097] <223> CRISPR RNA识别序列
[0098] <220>
[0099] <221> misc_feature
[0100] <222> (1) ... (21)
[0101] <223> n=a、t、c或g
[0102] <400> 9
[0103] nnnnnnnnnnn nnnnnnnnnnn ngg 23
[0104] <210> 10
[0105] <211> 25
[0106] <212> DNA
[0107] <213> 人工序列
[0108] <220>
[0109] <223> CRISPR RNA识别序列
[0110] <220>
[0111] <221> misc_feature
[0112] <222> (3) ... (23)
[0113] <223> n=a、t、c或g
[0114] <400> 10
[0115] ggnnnnnnnnnn nnnnnnnnnnn nnngg 25
[0116] <210> 11

[0117] <211> 20
[0118] <212> DNA
[0119] <213> 人工序列
[0120] <220>
[0121] <223> C5 (Hc) gRNA A DNA-靶向区段
[0122] <400> 11
[0123] atcacaaacc agttaaccgg 20
[0124] <210> 12
[0125] <211> 20
[0126] <212> DNA
[0127] <213> 人工序列
[0128] <220>
[0129] <223> C5 (Hc) gRNA B DNA-靶向区段
[0130] <400> 12
[0131] tttcagacga gccgaccgg 20
[0132] <210> 13
[0133] <211> 20
[0134] <212> DNA
[0135] <213> 人工序列
[0136] <220>
[0137] <223> C5 (Hc) gRNA C DNA-靶向区段
[0138] <400> 13
[0139] tgtgtgtcat agcgatgtcg 20
[0140] <210> 14
[0141] <211> 20
[0142] <212> DNA
[0143] <213> 人工序列
[0144] <220>
[0145] <223> C5 (Hc) gRNA D DNA-靶向区段
[0146] <400> 14
[0147] aacaggtacc ctatcctcac 20
[0148] <210> 15
[0149] <211> 20
[0150] <212> DNA
[0151] <213> 人工序列
[0152] <220>
[0153] <223> C5 (Hc) gRNA E DNA-靶向区段
[0154] <400> 15
[0155] ggcccggacc tagtctctct 20

[0156] <210> 16
[0157] <211> 20
[0158] <212> DNA
[0159] <213> 人工序列
[0160] <220>
[0161] <223> C5 (Hc) gRNA E2 DNA-靶向区段
[0162] <400> 16
[0163] tcgtggttgc atgcgcactg 20
[0164] <210> 17
[0165] <211> 20
[0166] <212> DNA
[0167] <213> 人工序列
[0168] <220>
[0169] <223> Lrp5 gRNA A DNA-靶向区段
[0170] <400> 17
[0171] gggaacccac agcatactcc 20
[0172] <210> 18
[0173] <211> 20
[0174] <212> DNA
[0175] <213> 人工序列
[0176] <220>
[0177] <223> Lrp5 gRNA B DNA-靶向区段
[0178] <400> 18
[0179] gaatcatgca cggctacccc 20
[0180] <210> 19
[0181] <211> 20
[0182] <212> DNA
[0183] <213> 人工序列
[0184] <220>
[0185] <223> Lrp5 gRNA B2 DNA-靶向区段
[0186] <400> 19
[0187] tgctcctatg gggaggcgcg 20
[0188] <210> 20
[0189] <211> 20
[0190] <212> DNA
[0191] <213> 人工序列
[0192] <220>
[0193] <223> Lrp5 gRNA C DNA-靶向区段
[0194] <400> 20

[0195] actgagatca atgaccccga 20
[0196] <210> 21
[0197] <211> 20
[0198] <212> DNA
[0199] <213> 人工序列
[0200] <220>
[0201] <223> Lrp5 gRNA D DNA-靶向区段
[0202] <400> 21
[0203] gggtcgccccg gaacctctac 20
[0204] <210> 22
[0205] <211> 20
[0206] <212> DNA
[0207] <213> 人工序列
[0208] <220>
[0209] <223> Lrp5 gRNA E2 DNA-靶向区段
[0210] <400> 22
[0211] cttggataac attgataccc 20
[0212] <210> 23
[0213] <211> 20
[0214] <212> DNA
[0215] <213> 人工序列
[0216] <220>
[0217] <223> Lrp5 gRNA E DNA-靶向区段
[0218] <400> 23
[0219] ggggcagagc ccttatatca 20
[0220] <210> 24
[0221] <211> 20
[0222] <212> DNA
[0223] <213> 人工序列
[0224] <220>
[0225] <223> Lrp5 gRNA F DNA-靶向区段
[0226] <400> 24
[0227] tcgctcacat taatccctag 20
[0228] <210> 25
[0229] <211> 20
[0230] <212> DNA
[0231] <213> 人工序列
[0232] <220>
[0233] <223> Ror1 gRNA A DNA-靶向区段

[0234] <400> 25
[0235] tgtgggcctt tgctgacac 20
[0236] <210> 26
[0237] <211> 20
[0238] <212> DNA
[0239] <213> 人工序列
[0240] <220>
[0241] <223> Ror1 gRNA B DNA-靶向区段
[0242] <400> 26
[0243] aatctatgat cctatggcct 20
[0244] <210> 27
[0245] <211> 20
[0246] <212> DNA
[0247] <213> 人工序列
[0248] <220>
[0249] <223> Ror1 gRNA D DNA-靶向区段
[0250] <400> 27
[0251] tgccaatagc agtgacttga 20
[0252] <210> 28
[0253] <211> 20
[0254] <212> DNA
[0255] <213> 人工序列
[0256] <220>
[0257] <223> Ror1 gRNA C DNA-靶向区段
[0258] <400> 28
[0259] gggaagaatg ggctattgtc 20
[0260] <210> 29
[0261] <211> 20
[0262] <212> DNA
[0263] <213> 人工序列
[0264] <220>
[0265] <223> Ror1 gRNA E DNA-靶向区段
[0266] <400> 29
[0267] ggttggttgt gctgatgacg 20
[0268] <210> 30
[0269] <211> 20
[0270] <212> DNA
[0271] <213> 人工序列
[0272] <220>

[0273] <223> Ror1 gRNA F DNA-靶向区段
[0274] <400> 30
[0275] ccgtcctagg ccttctacgt 20
[0276] <210> 31
[0277] <211> 20
[0278] <212> DNA
[0279] <213> 人工序列
[0280] <220>
[0281] <223> Trpa1 gRNA A DNA-靶向区段
[0282] <400> 31
[0283] gtactgggga atcggtggtc 20
[0284] <210> 32
[0285] <211> 20
[0286] <212> DNA
[0287] <213> 人工序列
[0288] <220>
[0289] <223> Trpa1 gRNA A2 DNA-靶向区段
[0290] <400> 32
[0291] cacgcactcc aaatttatcc 20
[0292] <210> 33
[0293] <211> 20
[0294] <212> DNA
[0295] <213> 人工序列
[0296] <220>
[0297] <223> Trpa1 gRNA B DNA-靶向区段
[0298] <400> 33
[0299] ctaagtgtgt atcagtacat 20
[0300] <210> 34
[0301] <211> 20
[0302] <212> DNA
[0303] <213> 人工序列
[0304] <220>
[0305] <223> Trpa1 gRNA C DNA-靶向区段
[0306] <400> 34
[0307] tgccctgcac aataagegca 20
[0308] <210> 35
[0309] <211> 20
[0310] <212> DNA
[0311] <213> 人工序列

- [0312] <220>
[0313] <223> Trpa1 gRNA D DNA-靶向区段
[0314] <400> 35
[0315] actcattgaa acgttatggc 20
[0316] <210> 36
[0317] <211> 20
[0318] <212> DNA
[0319] <213> 人工序列
[0320] <220>
[0321] <223> Trpa1 gRNA E2 DNA-靶向区段
[0322] <400> 36
[0323] agtaagggtg gattaaattc 20
[0324] <210> 37
[0325] <211> 20
[0326] <212> DNA
[0327] <213> 人工序列
[0328] <220>
[0329] <223> Trpa1 gRNA E DNA-靶向区段
[0330] <400> 37
[0331] gccatctaga ttcatgtaac 20
[0332] <210> 38
[0333] <211> 20
[0334] <212> DNA
[0335] <213> 人工序列
[0336] <220>
[0337] <223> Trpa1 gRNA F DNA-靶向区段
[0338] <400> 38
[0339] gactagaaat gttctgcacc 20
[0340] <210> 39
[0341] <211> 21
[0342] <212> DNA
[0343] <213> 人工序列
[0344] <220>
[0345] <223> 190045正向引物
[0346] <400> 39
[0347] gagctcatag ccaacagctt g 21
[0348] <210> 40
[0349] <211> 20
[0350] <212> DNA

[0351] <213> 人工序列
[0352] <220>
[0353] <223> 190061正向引物
[0354] <400> 40
[0355] atgcatcaga tcacgctcag 20
[0356] <210> 41
[0357] <211> 20
[0358] <212> DNA
[0359] <213> 人工序列
[0360] <220>
[0361] <223> 190068正向引物
[0362] <400> 41
[0363] gtccttgtgg catttccaac 20
[0364] <210> 42
[0365] <211> 24
[0366] <212> DNA
[0367] <213> 人工序列
[0368] <220>
[0369] <223> 190030正向引物
[0370] <400> 42
[0371] ccagtatggt gtcagttaat agcg 24
[0372] <210> 43
[0373] <211> 19
[0374] <212> DNA
[0375] <213> 人工序列
[0376] <220>
[0377] <223> 190033正向引物
[0378] <400> 43
[0379] ctgtgcagaa agcagcctc 19
[0380] <210> 44
[0381] <211> 20
[0382] <212> DNA
[0383] <213> 人工序列
[0384] <220>
[0385] <223> 190013正向引物
[0386] <400> 44
[0387] cctctccctc taggcacctg 20
[0388] <210> 45
[0389] <211> 20

[0390] <212> DNA
[0391] <213> 人工序列
[0392] <220>
[0393] <223> 190045反向引物
[0394] <400> 45
[0395] tctttaaggg ctccgttgtc 20
[0396] <210> 46
[0397] <211> 20
[0398] <212> DNA
[0399] <213> 人工序列
[0400] <220>
[0401] <223> 190061反向引物
[0402] <400> 46
[0403] aagaccaacc attcaccag 20
[0404] <210> 47
[0405] <211> 20
[0406] <212> DNA
[0407] <213> 人工序列
[0408] <220>
[0409] <223> 190068反向引物
[0410] <400> 47
[0411] ttcccagtc aagtcaaagg 20
[0412] <210> 48
[0413] <211> 20
[0414] <212> DNA
[0415] <213> 人工序列
[0416] <220>
[0417] <223> 190030反向引物
[0418] <400> 48
[0419] ctgttatctg caaggcaccc 20
[0420] <210> 49
[0421] <211> 20
[0422] <212> DNA
[0423] <213> 人工序列
[0424] <220>
[0425] <223> 190033反向引物
[0426] <400> 49
[0427] acaactggat cctgattcgc 20
[0428] <210> 50

[0429] <211> 20
[0430] <212> DNA
[0431] <213> 人工序列
[0432] <220>
[0433] <223> 190013反向引物
[0434] <400> 50
[0435] taagaggga tgggtgagac 20
[0436] <210> 51
[0437] <211> 20
[0438] <212> DNA
[0439] <213> 人工序列
[0440] <220>
[0441] <223> C2探针 (B6)
[0442] <400> 51
[0443] aattcagaag acctatcgta 20
[0444] <210> 52
[0445] <211> 21
[0446] <212> DNA
[0447] <213> 人工序列
[0448] <220>
[0449] <223> T3探针 (B6)
[0450] <400> 52
[0451] tatgtgtata ggtgtttgga t 21
[0452] <210> 53
[0453] <211> 19
[0454] <212> DNA
[0455] <213> 人工序列
[0456] <220>
[0457] <223> T6探针 (B6)
[0458] <400> 53
[0459] tacattgcta aatgaaacc 19
[0460] <210> 54
[0461] <211> 16
[0462] <212> DNA
[0463] <213> 人工序列
[0464] <220>
[0465] <223> T7探针 (B6)
[0466] <400> 54
[0467] cgcagtcattg cacata 16

[0468] <210> 55
[0469] <211> 20
[0470] <212> DNA
[0471] <213> 人工序列
[0472] <220>
[0473] <223> T8探针 (B6)
[0474] <400> 55
[0475] ttataaagcc cagtatgtac 20
[0476] <210> 56
[0477] <211> 14
[0478] <212> DNA
[0479] <213> 人工序列
[0480] <220>
[0481] <223> T9探针 (B6)
[0482] <400> 56
[0483] tgctgcataa tcag 14
[0484] <210> 57
[0485] <211> 18
[0486] <212> DNA
[0487] <213> 人工序列
[0488] <220>
[0489] <223> T10探针 (B6)
[0490] <400> 57
[0491] tcaggagtga attggata 18
[0492] <210> 58
[0493] <211> 16
[0494] <212> DNA
[0495] <213> 人工序列
[0496] <220>
[0497] <223> T11探针 (B6)
[0498] <400> 58
[0499] ctgctactta cctttg 16
[0500] <210> 59
[0501] <211> 13
[0502] <212> DNA
[0503] <213> 人工序列
[0504] <220>
[0505] <223> T13探针 (B6)
[0506] <400> 59

[0507] aggaggaaaa cgc 13
[0508] <210> 60
[0509] <211> 17
[0510] <212> DNA
[0511] <213> 人工序列
[0512] <220>
[0513] <223> T14探针 (B6)
[0514] <400> 60
[0515] cctttgttcc tcataag 17
[0516] <210> 61
[0517] <211> 20
[0518] <212> DNA
[0519] <213> 人工序列
[0520] <220>
[0521] <223> C2探针 (129)
[0522] <400> 61
[0523] aattcagaag acctattgta 20
[0524] <210> 62
[0525] <211> 21
[0526] <212> DNA
[0527] <213> 人工序列
[0528] <220>
[0529] <223> T3探针 (129)
[0530] <400> 62
[0531] tatgtgtata ggtgtttgca t 21
[0532] <210> 63
[0533] <211> 16
[0534] <212> DNA
[0535] <213> 人工序列
[0536] <220>
[0537] <223> T6探针 (129)
[0538] <400> 63
[0539] cattgctaca tgaaac 16
[0540] <210> 64
[0541] <211> 16
[0542] <212> DNA
[0543] <213> 人工序列
[0544] <220>
[0545] <223> T7探针 (129)

[0546] <400> 64
[0547] cgcagtcacgta 16
[0548] <210> 65
[0549] <211> 23
[0550] <212> DNA
[0551] <213> 人工序列
[0552] <220>
[0553] <223> T8探针 (129)
[0554] <400> 65
[0555] tgagaattta taaagcccaa tat 23
[0556] <210> 66
[0557] <211> 14
[0558] <212> DNA
[0559] <213> 人工序列
[0560] <220>
[0561] <223> T9探针 (129)
[0562] <400> 66
[0563] tgctgcatga tcag 14
[0564] <210> 67
[0565] <211> 15
[0566] <212> DNA
[0567] <213> 人工序列
[0568] <220>
[0569] <223> T10探针 (129)
[0570] <400> 67
[0571] tcaggagtga atcgg 15
[0572] <210> 68
[0573] <211> 16
[0574] <212> DNA
[0575] <213> 人工序列
[0576] <220>
[0577] <223> T11探针 (129)
[0578] <400> 68
[0579] ctgctagtta cctttg 16
[0580] <210> 69
[0581] <211> 15
[0582] <212> DNA
[0583] <213> 人工序列
[0584] <220>

[0585]	<223> T13探针 (129)
[0586]	<400> 69
[0587]	aggaggaaga cgcag 15
[0588]	<210> 70
[0589]	<211> 17
[0590]	<212> DNA
[0591]	<213> 人工序列
[0592]	<220>
[0593]	<223> T14探针 (129)
[0594]	<400> 70
[0595]	ctttgttctt cataagc 17
[0596]	<210> 71
[0597]	<211> 25
[0598]	<212> DNA
[0599]	<213> 人工序列
[0600]	<220>
[0601]	<223> C2正向引物
[0602]	<400> 71
[0603]	atgagggatt tccttaatca gacaa 25
[0604]	<210> 72
[0605]	<211> 29
[0606]	<212> DNA
[0607]	<213> 人工序列
[0608]	<220>
[0609]	<223> T3正向引物
[0610]	<400> 72
[0611]	tggtatgttt attcttactc aaggttttg 29
[0612]	<210> 73
[0613]	<211> 22
[0614]	<212> DNA
[0615]	<213> 人工序列
[0616]	<220>
[0617]	<223> T6正向引物
[0618]	<400> 73
[0619]	gggcaactga tggaaagaac tc 22
[0620]	<210> 74
[0621]	<211> 23
[0622]	<212> DNA
[0623]	<213> 人工序列

[0624] <220>
[0625] <223> T7正向引物
[0626] <400> 74
[0627] gactgacgca caaacttgtc ctt 23
[0628] <210> 75
[0629] <211> 26
[0630] <212> DNA
[0631] <213> 人工序列
[0632] <220>
[0633] <223> T8正向引物
[0634] <400> 75
[0635] cccaaagcat ataacaagaa caaatg 26
[0636] <210> 76
[0637] <211> 18
[0638] <212> DNA
[0639] <213> 人工序列
[0640] <220>
[0641] <223> T9正向引物
[0642] <400> 76
[0643] gcaggacgca ggcgttta 18
[0644] <210> 77
[0645] <211> 23
[0646] <212> DNA
[0647] <213> 人工序列
[0648] <220>
[0649] <223> T10正向引物
[0650] <400> 77
[0651] gcatcctcat ggcagtctac atc 23
[0652] <210> 78
[0653] <211> 20
[0654] <212> DNA
[0655] <213> 人工序列
[0656] <220>
[0657] <223> T11正向引物
[0658] <400> 78
[0659] cctgcccctt gatgagtgtt 20
[0660] <210> 79
[0661] <211> 23
[0662] <212> DNA

[0663] <213> 人工序列
[0664] <220>
[0665] <223> T13正向引物
[0666] <400> 79
[0667] ccctctttga tatgctcgtg tgt 23
[0668] <210> 80
[0669] <211> 22
[0670] <212> DNA
[0671] <213> 人工序列
[0672] <220>
[0673] <223> T14正向引物
[0674] <400> 80
[0675] tcccacaggt ccatgtcttt aa 22
[0676] <210> 81
[0677] <211> 28
[0678] <212> DNA
[0679] <213> 人工序列
[0680] <220>
[0681] <223> C2反向引物
[0682] <400> 81
[0683] agactacaat gagctaccat cataaggt 28
[0684] <210> 82
[0685] <211> 24
[0686] <212> DNA
[0687] <213> 人工序列
[0688] <220>
[0689] <223> T3反向引物
[0690] <400> 82
[0691] caaccatcta aaactccagt tcca 24
[0692] <210> 83
[0693] <211> 28
[0694] <212> DNA
[0695] <213> 人工序列
[0696] <220>
[0697] <223> T6反向引物
[0698] <400> 83
[0699] tgtgtaacag gacagttgaa tgtagaga 28
[0700] <210> 84
[0701] <211> 19

[0702] <212> DNA
[0703] <213> 人工序列
[0704] <220>
[0705] <223> T7反向引物
[0706] <400> 84
[0707] cttaaaaccc gccctgcat 19
[0708] <210> 85
[0709] <211> 26
[0710] <212> DNA
[0711] <213> 人工序列
[0712] <220>
[0713] <223> T8反向引物
[0714] <400> 85
[0715] ctacaggaga tgtggctggt ctatgt 26
[0716] <210> 86
[0717] <211> 22
[0718] <212> DNA
[0719] <213> 人工序列
[0720] <220>
[0721] <223> T9反向引物
[0722] <400> 86
[0723] tcagcgtgat tcgcttgtag tc 22
[0724] <210> 87
[0725] <211> 26
[0726] <212> DNA
[0727] <213> 人工序列
[0728] <220>
[0729] <223> T10反向引物
[0730] <400> 87
[0731] tgcatagctg tttgaataat gacaag 26
[0732] <210> 88
[0733] <211> 21
[0734] <212> DNA
[0735] <213> 人工序列
[0736] <220>
[0737] <223> T11反向引物
[0738] <400> 88
[0739] tgcagcatct ctgtcaagca a 21
[0740] <210> 89

[0741] <211> 24
[0742] <212> DNA
[0743] <213> 人工序列
[0744] <220>
[0745] <223> T13反向引物
[0746] <400> 89
[0747] gcaacaacat aacccacagc ataa 24
[0748] <210> 90
[0749] <211> 23
[0750] <212> DNA
[0751] <213> 人工序列
[0752] <220>
[0753] <223> T14反向引物
[0754] <400> 90
[0755] gctaagcggtt tggaagaaat tcc 23
[0756] <210> 91
[0757] <211> 20
[0758] <212> DNA
[0759] <213> 人工序列
[0760] <220>
[0761] <223> SV 13.7正向引物
[0762] <400> 91
[0763] taggctctaa ggatgctggc 20
[0764] <210> 92
[0765] <211> 20
[0766] <212> DNA
[0767] <213> 人工序列
[0768] <220>
[0769] <223> SV 13.7 reverse primer
[0770] <400> 92
[0771] aagcagcttc aaaccctctg 20
[0772] <210> 93
[0773] <211> 20
[0774] <212> DNA
[0775] <213> 人工序列
[0776] <220>
[0777] <223> SV 20.0正向引物
[0778] <400> 93
[0779] ttacttggcc ttggaactgc 20

[0780]	<210> 94
[0781]	<211> 21
[0782]	<212> DNA
[0783]	<213> 人工序列
[0784]	<220>
[0785]	<223> SV 20.0 reverse primer
[0786]	<400> 94
[0787]	tgattcgtaa tcgtcactgc c 21
[0788]	<210> 95
[0789]	<211> 20
[0790]	<212> DNA
[0791]	<213> 人工序列
[0792]	<220>
[0793]	<223> SV 36.9正向引物
[0794]	<400> 95
[0795]	tcctgtcccg agaaactgtc 20
[0796]	<210> 96
[0797]	<211> 20
[0798]	<212> DNA
[0799]	<213> 人工序列
[0800]	<220>
[0801]	<223> SV 36.9 reverse primer
[0802]	<400> 96
[0803]	agctggcttt cagagagctg 20
[0804]	<210> 97
[0805]	<211> 20
[0806]	<212> DNA
[0807]	<213> 人工序列
[0808]	<220>
[0809]	<223> SV 56.7正向引物
[0810]	<400> 97
[0811]	ttagaaagtg ccaaccaggc 20
[0812]	<210> 98
[0813]	<211> 20
[0814]	<212> DNA
[0815]	<213> 人工序列
[0816]	<220>
[0817]	<223> SV 56.7 reverse primer
[0818]	<400> 98

[0819] ctctggctag gaacaatggc 20
[0820] <210> 99
[0821] <211> 25
[0822] <212> DNA
[0823] <213> 人工序列
[0824] <220>
[0825] <223> m-lr-f引物
[0826] <400> 99
[0827] gttagggtgca gggctctactc agctg 25
[0828] <210> 100
[0829] <211> 20
[0830] <212> DNA
[0831] <213> 人工序列
[0832] <220>
[0833] <223> m-5'-f引物
[0834] <400> 100
[0835] ggaggagagg agaagcagcc 20
[0836] <210> 101
[0837] <211> 20
[0838] <212> DNA
[0839] <213> 人工序列
[0840] <220>
[0841] <223> m-A引物
[0842] <400> 101
[0843] ggaggagagg agaagcagcc 20
[0844] <210> 102
[0845] <211> 26
[0846] <212> DNA
[0847] <213> 人工序列
[0848] <220>
[0849] <223> h-lr-r引物
[0850] <400> 102
[0851] gcaaacagcc ttcttccac attcgg 26
[0852] <210> 103
[0853] <211> 24
[0854] <212> DNA
[0855] <213> 人工序列
[0856] <220>
[0857] <223> m-5'-r引物

[0858] <400> 103
[0859] ttgctttcag tagttcaggt gtgc 24
[0860] <210> 104
[0861] <211> 20
[0862] <212> DNA
[0863] <213> 人工序列
[0864] <220>
[0865] <223> h-5'-r引物
[0866] <400> 104
[0867] ggcgttgtca ggaagttgcc 20
[0868] <210> 105
[0869] <211> 22
[0870] <212> DNA
[0871] <213> 人工序列
[0872] <220>
[0873] <223> m-F引物
[0874] <400> 105
[0875] tgaagttgag aggcatga gg 22
[0876] <210> 106
[0877] <211> 24
[0878] <212> DNA
[0879] <213> 人工序列
[0880] <220>
[0881] <223> m-E2引物
[0882] <400> 106
[0883] tagagtagcc acaggcagca aagc 24
[0884] <210> 107
[0885] <211> 21
[0886] <212> RNA
[0887] <213> 人工序列
[0888] <220>
[0889] <223> Cmah gRNA A DNA-靶向区段
[0890] <400> 107
[0891] gugacaggaa ggcucucac c 21
[0892] <210> 108
[0893] <211> 21
[0894] <212> RNA
[0895] <213> 人工序列
[0896] <220>

- [0897] <223> Cmah gRNA B DNA-靶向区段
- [0898] <400> 108
- [0899] gcuuacaagc aaauugcuga c 21
- [0900] <210> 109
- [0901] <211> 108
- [0902] <212> DNA
- [0903] <213> 人工序列
- [0904] <220>
- [0905] <223> Cmah基因座
- [0906] <400> 109
- [0907] ttgccagtga ccctgtttgc agtttagagtt gacaggaagg cttctcaccc gggacatddd 60
- [0908] aaatgaccca gtcagcaaatt tgcttgtaag ttttggtggt ctttcatt 108
- [0909] <210> 110
- [0910] <211> 46
- [0911] <212> DNA
- [0912] <213> 人工序列
- [0913] <220>
- [0914] <223> gRNA A切口上游的Cmah基因座
- [0915] <400> 110
- [0916] ttgccagtga ccctgtttgc agtttagagtt gacaggaagg cttctc 46
- [0917] <210> 111
- [0918] <211> 35
- [0919] <212> DNA
- [0920] <213> 人工序列
- [0921] <220>
- [0922] <223> gRNA B切口下游的Cmah基因座
- [0923] <400> 111
- [0924] agcaaattgc ttgtaagttt tggtgttctt tcatt 35
- [0925] <210> 112
- [0926] <211> 27
- [0927] <212> DNA
- [0928] <213> 人工序列
- [0929] <220>
- [0930] <223> 由gRNA A和gRNA B切除的Cmah基因座序列
- [0931] <400> 112
- [0932] acccgggaca ttttaaattga cccagtc 27
- [0933] <210> 113
- [0934] <211> 21
- [0935] <212> DNA

[0936] <213> 人工序列
[0937] <220>
[0938] <223> C5引物m-5'-F
[0939] <400> 113
[0940] accccagcat ctgacgacac c 21
[0941] <210> 114
[0942] <211> 20
[0943] <212> DNA
[0944] <213> 人工序列
[0945] <220>
[0946] <223> C5引物m-5'-R
[0947] <400> 114
[0948] agaaagaccg cagtggaacc 20
[0949] <210> 115
[0950] <211> 21
[0951] <212> DNA
[0952] <213> 人工序列
[0953] <220>
[0954] <223> C5引物h-5'-R
[0955] <400> 115
[0956] tccccacatg cctagtagga g 21
[0957] <210> 116
[0958] <211> 22
[0959] <212> DNA
[0960] <213> 人工序列
[0961] <220>
[0962] <223> Cmah TaqMan正向引物
[0963] <400> 116
[0964] gtgaccctgt ttgcagttag ag 22
[0965] <210> 117
[0966] <211> 23
[0967] <212> DNA
[0968] <213> 人工序列
[0969] <220>
[0970] <223> Cmah TaqMan探针
[0971] <400> 117
[0972] acaggaaggc ttctcaccgc gga 23
[0973] <210> 118
[0974] <211> 24

[0975] <212> DNA
[0976] <213> 人工序列
[0977] <220>
[0978] <223> Cmah TaqMan反向引物
[0979] <400> 118
[0980] tgaagtcagg aaactgttcc aatg 24
[0981] <210> 119
[0982] <211> 22
[0983] <212> DNA
[0984] <213> 人工序列
[0985] <220>
[0986] <223> 7064retU正向引物
[0987] <400> 119
[0988] cctcctgagc tttcctttgc ag 22
[0989] <210> 120
[0990] <211> 25
[0991] <212> DNA
[0992] <213> 人工序列
[0993] <220>
[0994] <223> 7064retU反向引物
[0995] <400> 120
[0996] cctagacaac acagacactg tatca 25
[0997] <210> 121
[0998] <211> 23
[0999] <212> DNA
[1000] <213> 人工序列
[1001] <220>
[1002] <223> 7064retU TaqMan探针
[1003] <400> 121
[1004] ttctgccctt gaaaaggaga ggc 23
[1005] <210> 122
[1006] <211> 18
[1007] <212> DNA
[1008] <213> 人工序列
[1009] <220>
[1010] <223> 7064retD正向引物
[1011] <400> 122
[1012] cctctgaggc cacctgaa 18
[1013] <210> 123

[1014]	<211> 23
[1015]	<212> DNA
[1016]	<213> 人工序列
[1017]	<220>
[1018]	<223> 7064retD反向引物
[1019]	<400> 123
[1020]	ccctgacaag ttctgccttc tac 23
[1021]	<210> 124
[1022]	<211> 21
[1023]	<212> DNA
[1024]	<213> 人工序列
[1025]	<220>
[1026]	<223> 7064retD TaqMan探针
[1027]	<400> 124
[1028]	tgcccaagcc tctgcagctt t 21
[1029]	<210> 125
[1030]	<211> 19
[1031]	<212> DNA
[1032]	<213> 人工序列
[1033]	<220>
[1034]	<223> 7140retU正向引物
[1035]	<400> 125
[1036]	cccagcatct gacgacacc 19
[1037]	<210> 126
[1038]	<211> 21
[1039]	<212> DNA
[1040]	<213> 人工序列
[1041]	<220>
[1042]	<223> 7140retU反向引物
[1043]	<400> 126
[1044]	gaccactgtg ggcattctgta g 21
[1045]	<210> 127
[1046]	<211> 27
[1047]	<212> DNA
[1048]	<213> 人工序列
[1049]	<220>
[1050]	<223> 7140retU TaqMan探针
[1051]	<400> 127
[1052]	ccgagtctgc tggtactgtt agcatca 27

[1053]	<210> 128
[1054]	<211> 20
[1055]	<212> DNA
[1056]	<213> 人工序列
[1057]	<220>
[1058]	<223> 7140retD正向引物
[1059]	<400> 128
[1060]	cccgacacct tctgagcatg 20
[1061]	<210> 129
[1062]	<211> 21
[1063]	<212> DNA
[1064]	<213> 人工序列
[1065]	<220>
[1066]	<223> 7140retD反向引物
[1067]	<400> 129
[1068]	tgcaggctga gtcaggattt g 21
[1069]	<210> 130
[1070]	<211> 25
[1071]	<212> DNA
[1072]	<213> 人工序列
[1073]	<220>
[1074]	<223> 7140retD TaqMan探针
[1075]	<400> 130
[1076]	tagtcacgtt ttgtgacacc ccaga 25
[1077]	<210> 131
[1078]	<211> 22
[1079]	<212> DNA
[1080]	<213> 人工序列
[1081]	<220>
[1082]	<223> mADAM6-2 LOA正向引物
[1083]	<400> 131
[1084]	agggctgagg gagaacatat ac 22
[1085]	<210> 132
[1086]	<211> 20
[1087]	<212> DNA
[1088]	<213> 人工序列
[1089]	<220>
[1090]	<223> mADAM6-2 LOA反向引物
[1091]	<400> 132

[1092] aggcctgatg caggagctat 20
[1093] <210> 133
[1094] <211> 26
[1095] <212> DNA
[1096] <213> 人工序列
[1097] <220>
[1098] <223> mADAM6-2 LOA探针
[1099] <400> 133
[1100] tcctctcagc tggattaaca gcatca 26
[1101] <210> 134
[1102] <211> 20
[1103] <212> DNA
[1104] <213> 人工序列
[1105] <220>
[1106] <223> hIgH31 LOA正向引物
[1107] <400> 134
[1108] atcacactca tcccatcccc 20
[1109] <210> 135
[1110] <211> 20
[1111] <212> DNA
[1112] <213> 人工序列
[1113] <220>
[1114] <223> hIgH31 LOA反向引物
[1115] <400> 135
[1116] cacagggaag caggaactgc 20
[1117] <210> 136
[1118] <211> 29
[1119] <212> DNA
[1120] <213> 人工序列
[1121] <220>
[1122] <223> hIgH31 LOA探针
[1123] <400> 136
[1124] cccttccta agtaccacag agtgggctc 29
[1125] <210> 137
[1126] <211> 19
[1127] <212> DNA
[1128] <213> 人工序列
[1129] <220>
[1130] <223> hIgH9 LOA正向引物

[1131]	<400> 137
[1132]	tcctccaacg acaggtccc 19
[1133]	<210> 138
[1134]	<211> 20
[1135]	<212> DNA
[1136]	<213> 人工序列
[1137]	<220>
[1138]	<223> hIgH9 LOA反向引物
[1139]	<400> 138
[1140]	gatgaactga cgggcacagg 20
[1141]	<210> 139
[1142]	<211> 24
[1143]	<212> DNA
[1144]	<213> 人工序列
[1145]	<220>
[1146]	<223> hIgH9 LOA探针
[1147]	<400> 139
[1148]	tccctggaac tctgccccga caca 24
[1149]	<210> 140
[1150]	<211> 19
[1151]	<212> DNA
[1152]	<213> 人工序列
[1153]	<220>
[1154]	<223> hIgH1 LOA正向引物
[1155]	<400> 140
[1156]	cagtcccgtt gatccagcc 19
[1157]	<210> 141
[1158]	<211> 21
[1159]	<212> DNA
[1160]	<213> 人工序列
[1161]	<220>
[1162]	<223> hIgH1 LOA反向引物
[1163]	<400> 141
[1164]	ggatatgcag cactgtgcca c 21
[1165]	<210> 142
[1166]	<211> 30
[1167]	<212> DNA
[1168]	<213> 人工序列
[1169]	<220>

- [1170] <223> hIgH1 LOA探针
[1171] <400> 142
[1172] cccatcaggg attttgtatc tctgtggacg 30
[1173] <210> 143
[1174] <211> 19
[1175] <212> DNA
[1176] <213> 人工序列
[1177] <220>
[1178] <223> Neo GOA正向引物
[1179] <400> 143
[1180] ggtggagagg ctattcggc 19
[1181] <210> 144
[1182] <211> 17
[1183] <212> DNA
[1184] <213> 人工序列
[1185] <220>
[1186] <223> Neo GOA反向引物
[1187] <400> 144
[1188] gaacacggcg gcatcag 17
[1189] <210> 145
[1190] <211> 23
[1191] <212> DNA
[1192] <213> 人工序列
[1193] <220>
[1194] <223> Neo GOA探针
[1195] <400> 145
[1196] tgggcacaac agacaatcgg ctg 23
[1197] <210> 146
[1198] <211> 19
[1199] <212> DNA
[1200] <213> 人工序列
[1201] <220>
[1202] <223> 5' IgH Arm1保留测定正向引物
[1203] <400> 146
[1204] aggatgctgg gaaacagac 19
[1205] <210> 147
[1206] <211> 20
[1207] <212> DNA
[1208] <213> 人工序列

[1209]	<220>
[1210]	<223> 5' IgH Arm1保留测定反向引物
[1211]	<400> 147
[1212]	gaccactctc aggactctca 20
[1213]	<210> 148
[1214]	<211> 24
[1215]	<212> DNA
[1216]	<213> 人工序列
[1217]	<220>
[1218]	<223> 5' IgH Arm1保留测定探针
[1219]	<400> 148
[1220]	tggaaggtcc caaaggaaac caca 24
[1221]	<210> 149
[1222]	<211> 22
[1223]	<212> DNA
[1224]	<213> 人工序列
[1225]	<220>
[1226]	<223> mIgM398保留测定正向引物
[1227]	<400> 149
[1228]	gagctcacac cttgaccttt ca 22
[1229]	<210> 150
[1230]	<211> 22
[1231]	<212> DNA
[1232]	<213> 人工序列
[1233]	<220>
[1234]	<223> mIgM398保留测定反向引物
[1235]	<400> 150
[1236]	tggtgggacg aacacattta ca 22
[1237]	<210> 151
[1238]	<211> 24
[1239]	<212> DNA
[1240]	<213> 人工序列
[1241]	<220>
[1242]	<223> mIgM398保留测定探针
[1243]	<400> 151
[1244]	ccagctgtcg cagagatgaa cccc 24
[1245]	<210> 152
[1246]	<211> 22
[1247]	<212> DNA

- [1248] <213> 人工序列
[1249] <220>
[1250] <223> mIgM1045保留测定正向引物
[1251] <400> 152
[1252] tccctccaca gacatcctaa cc 22
[1253] <210> 153
[1254] <211> 21
[1255] <212> DNA
[1256] <213> 人工序列
[1257] <220>
[1258] <223> mIgM1045保留测定反向引物
[1259] <400> 153
[1260] gttagcggac ttgctgagga a 21
[1261] <210> 154
[1262] <211> 22
[1263] <212> DNA
[1264] <213> 人工序列
[1265] <220>
[1266] <223> mIgM1045保留测定探针
[1267] <400> 154
[1268] tcaccatccc cccctccttt gc 22
[1269] <210> 155
[1270] <211> 22
[1271] <212> DNA
[1272] <213> 人工序列
[1273] <220>
[1274] <223> 3' IgH Arm2保留测定正向引物
[1275] <400> 155
[1276] ggtcatgtgg caaggctatt tg 22
[1277] <210> 156
[1278] <211> 21
[1279] <212> DNA
[1280] <213> 人工序列
[1281] <220>
[1282] <223> 3' IgH Arm2保留测定正向引物
[1283] <400> 156
[1284] agcctggact ttcggtttgg t 21
[1285] <210> 157
[1286] <211> 30

- [1287] <212> DNA
[1288] <213> 人工序列
[1289] <220>
[1290] <223> 3' IgH Arm2保留测定探针
[1291] <400> 157
[1292] ccactaggta aacttgttagc tgtggtttga 30
[1293] <210> 158
[1294] <211> 17
[1295] <212> DNA
[1296] <213> 人工序列
[1297] <220>
[1298] <223> mIgHp2亲本正向引物
[1299] <400> 158
[1300] gccatgcaag gccaagc 17
[1301] <210> 159
[1302] <211> 24
[1303] <212> DNA
[1304] <213> 人工序列
[1305] <220>
[1306] <223> mIgHp2亲本反向引物
[1307] <400> 159
[1308] agttcttgag ccttagggtg ctag 24
[1309] <210> 160
[1310] <211> 24
[1311] <212> DNA
[1312] <213> 人工序列
[1313] <220>
[1314] <223> mIgHp2亲本探针
[1315] <400> 160
[1316] ccaggaaaat gctgccagag cctg 24
[1317] <210> 161
[1318] <211> 20
[1319] <212> DNA
[1320] <213> 人工序列
[1321] <220>
[1322] <223> mIgKd2亲本正向引物
[1323] <400> 161
[1324] gcaaacaaaa accactggcc 20
[1325] <210> 162

- [1326] <211> 19
[1327] <212> DNA
[1328] <213> 人工序列
[1329] <220>
[1330] <223> mIgKd2亲本反向引物
[1331] <400> 162
[1332] ggccacattc catgggttc 19
[1333] <210> 163
[1334] <211> 37
[1335] <212> DNA
[1336] <213> 人工序列
[1337] <220>
[1338] <223> mIgKd2亲本探针
[1339] <400> 163
[1340] ctgttcctct aaaactggac tccacagtaa atggaaa 37
[1341] <210> 164
[1342] <211> 20
[1343] <212> DNA
[1344] <213> 人工序列
[1345] <220>
[1346] <223> hIgK5亲本正向引物
[1347] <400> 164
[1348] ccccgctctc ctcctttttc 20
[1349] <210> 165
[1350] <211> 19
[1351] <212> DNA
[1352] <213> 人工序列
[1353] <220>
[1354] <223> hIgK5亲本反向引物
[1355] <400> 165
[1356] tgcaagtgct gccagcaag 19
[1357] <210> 166
[1358] <211> 32
[1359] <212> DNA
[1360] <213> 人工序列
[1361] <220>
[1362] <223> hIgK5亲本探针
[1363] <400> 166
[1364] tcatgtccat taaccattt accttttgcc ca 32

[1365]	<210> 167
[1366]	<211> 21
[1367]	<212> DNA
[1368]	<213> 人工序列
[1369]	<220>
[1370]	<223> 3' gRNA_I DNA-靶向序列
[1371]	<400> 167
[1372]	gactactacg gtatggacgt c 21
[1373]	<210> 168
[1374]	<211> 21
[1375]	<212> DNA
[1376]	<213> 人工序列
[1377]	<220>
[1378]	<223> 3' gRNA_II DNA-靶向序列
[1379]	<400> 168
[1380]	gctactacgg tatggacgtc t 21
[1381]	<210> 169
[1382]	<211> 20
[1383]	<212> DNA
[1384]	<213> 人工序列
[1385]	<220>
[1386]	<223> 5' gRNA_I DNA-靶向序列
[1387]	<400> 169
[1388]	gaagctgact agtttacgca 20
[1389]	<210> 170
[1390]	<211> 21
[1391]	<212> DNA
[1392]	<213> 人工序列
[1393]	<220>
[1394]	<223> 5' gRNA_II DNA-靶向序列
[1395]	<400> 170
[1396]	gtagcattct tacacctagc a 21
[1397]	<210> 171
[1398]	<211> 23
[1399]	<212> DNA
[1400]	<213> 人工序列
[1401]	<220>
[1402]	<223> 5' IgH Arm2保留测定正向引物
[1403]	<400> 171

[1404] ctgagcatatc tgctgcctaa cac 23
[1405] <210> 172
[1406] <211> 25
[1407] <212> DNA
[1408] <213> 人工序列
[1409] <220>
[1410] <223> 5' IgH Arm2保留测定反向引物
[1411] <400> 172
[1412] gagcagtgc tttcttagtt aagga 25
[1413] <210> 173
[1414] <211> 26
[1415] <212> DNA
[1416] <213> 人工序列
[1417] <220>
[1418] <223> 5' IgH Arm2保留测定探针
[1419] <400> 173
[1420] tgaaatggca gttcttctcc agctgg 26
[1421] <210> 174
[1422] <211> 20
[1423] <212> DNA
[1424] <213> 人工序列
[1425] <220>
[1426] <223> 3' IgH Arm1保留测定正向引物
[1427] <400> 174
[1428] ggtggagtcc ctggatgatg 20
[1429] <210> 175
[1430] <211> 21
[1431] <212> DNA
[1432] <213> 人工序列
[1433] <220>
[1434] <223> 3' IgH Arm1保留测定反向引物
[1435] <400> 175
[1436] atccctccag ccataggatt g 21
[1437] <210> 176
[1438] <211> 28
[1439] <212> DNA
[1440] <213> 人工序列
[1441] <220>
[1442] <223> 3' IgH Arm1保留测定探针

[1443] <400> 176

[1444] ctttggaggc tcatttgagg gagatgct 28

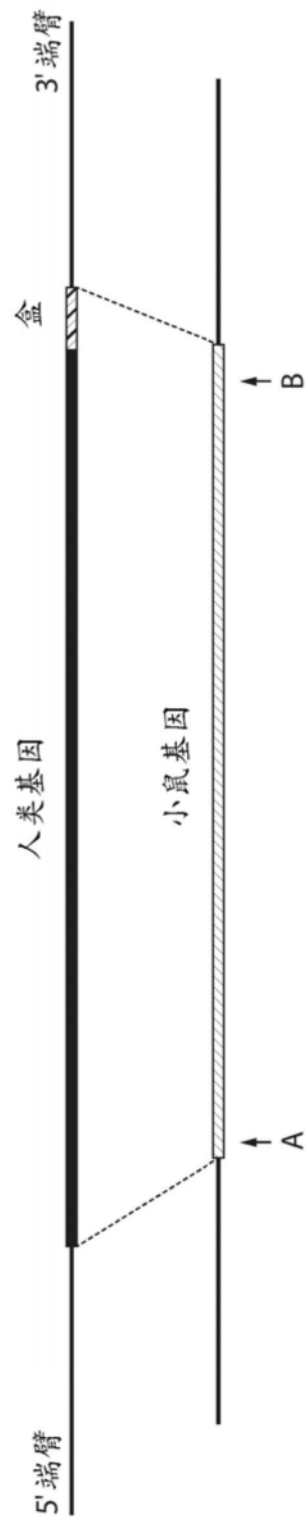


图2A



图2B



图2C



图2D



图2E

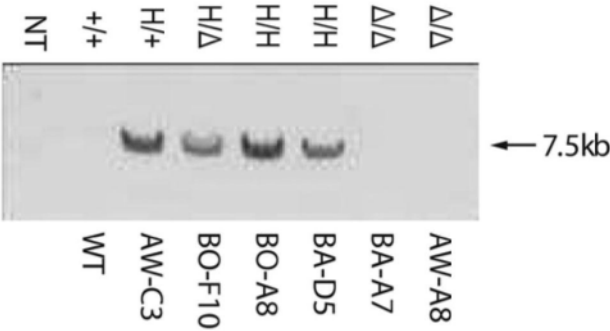


图3A

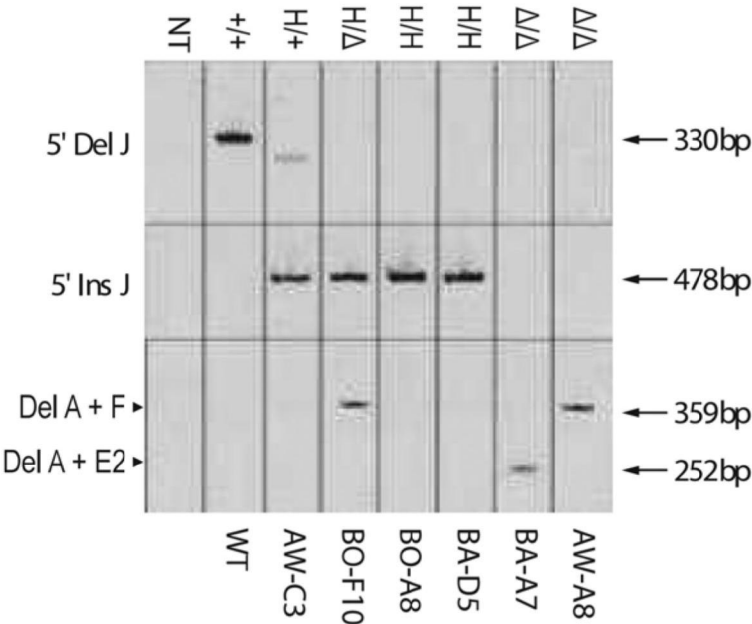


图3B

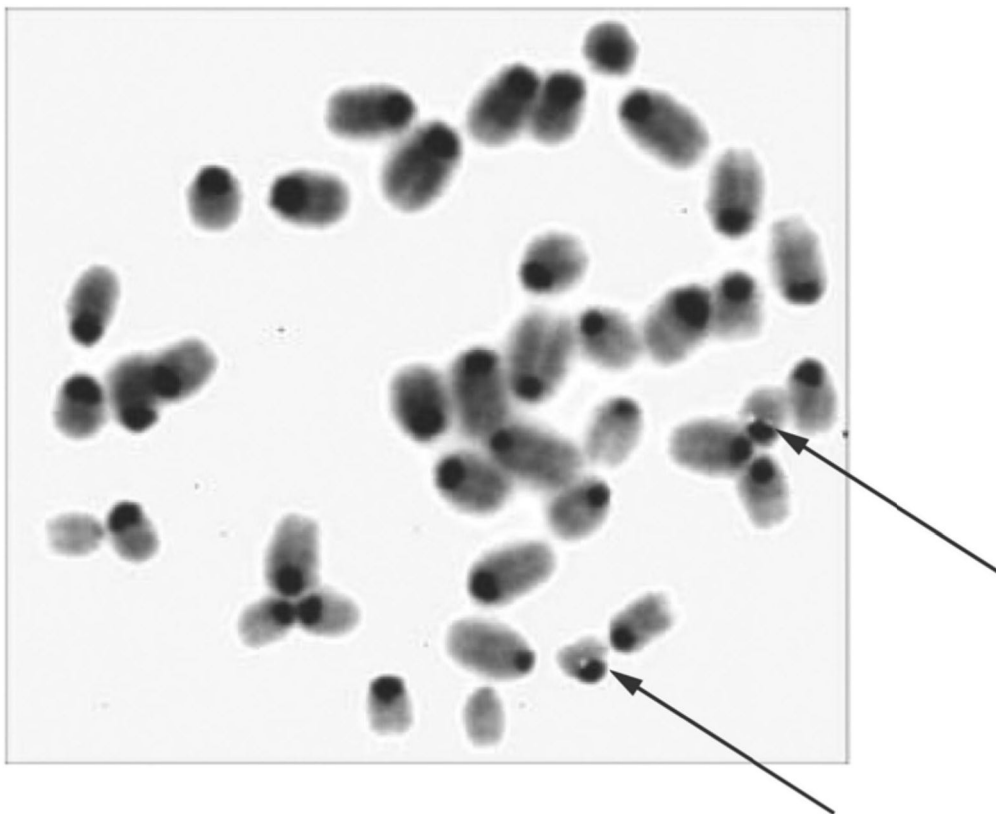


图4A



图4B



图4C

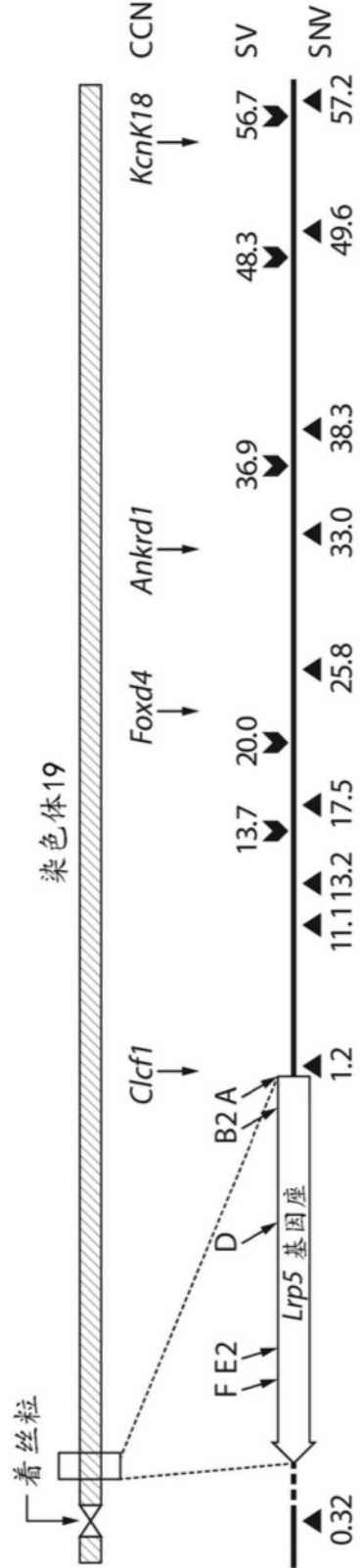


图5

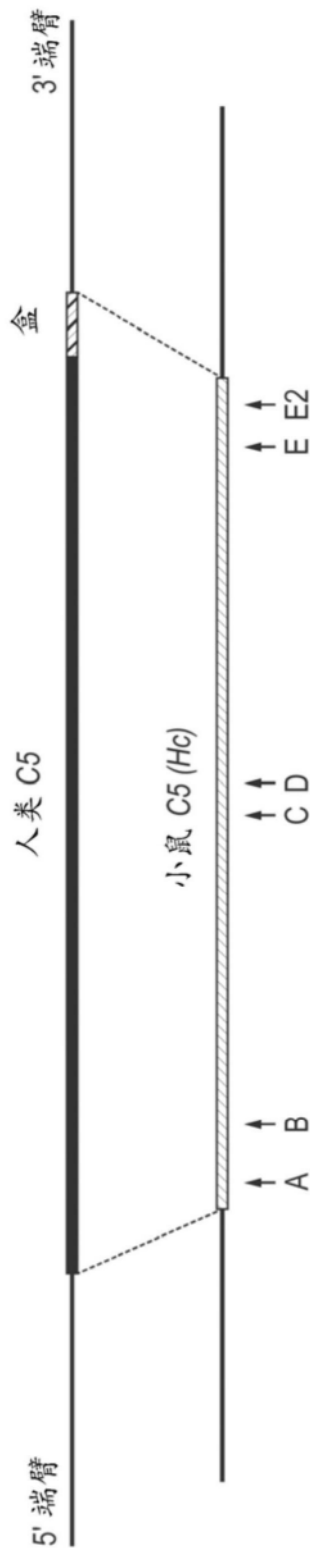


图6

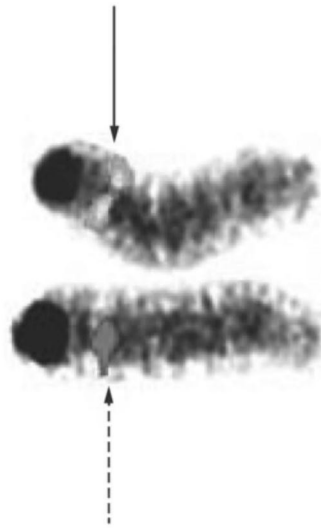


图7A

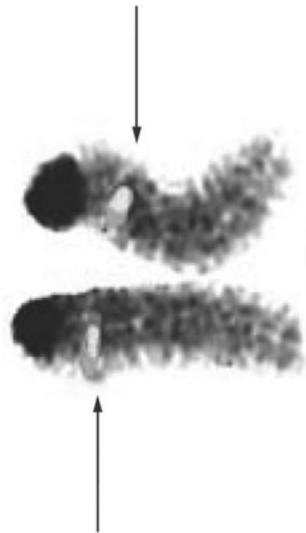


图7B

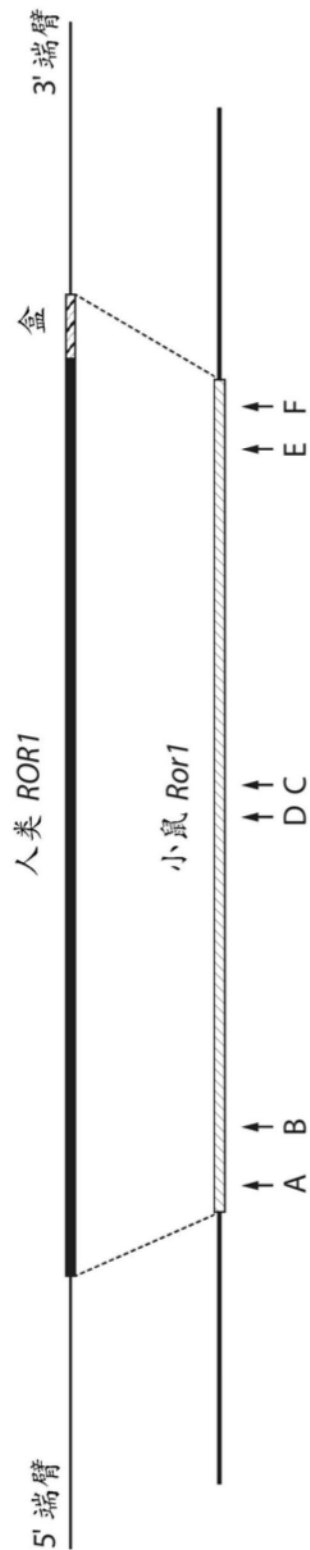


图8

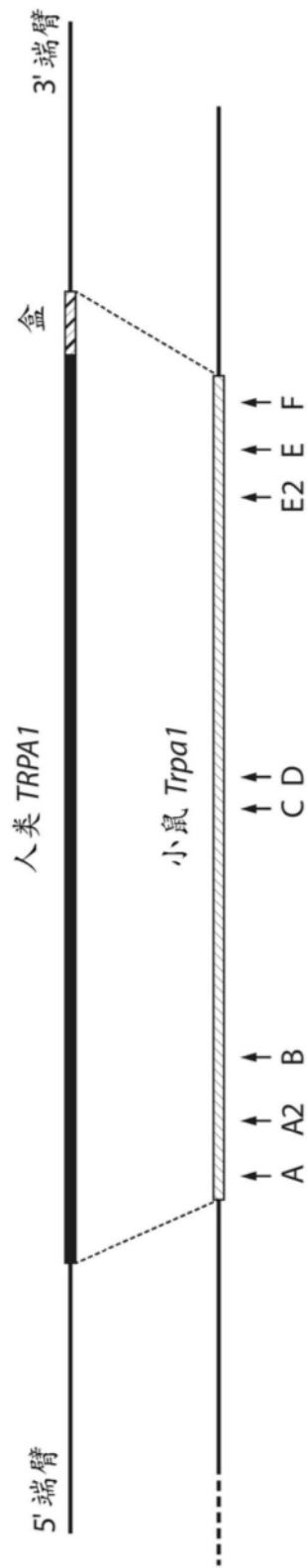
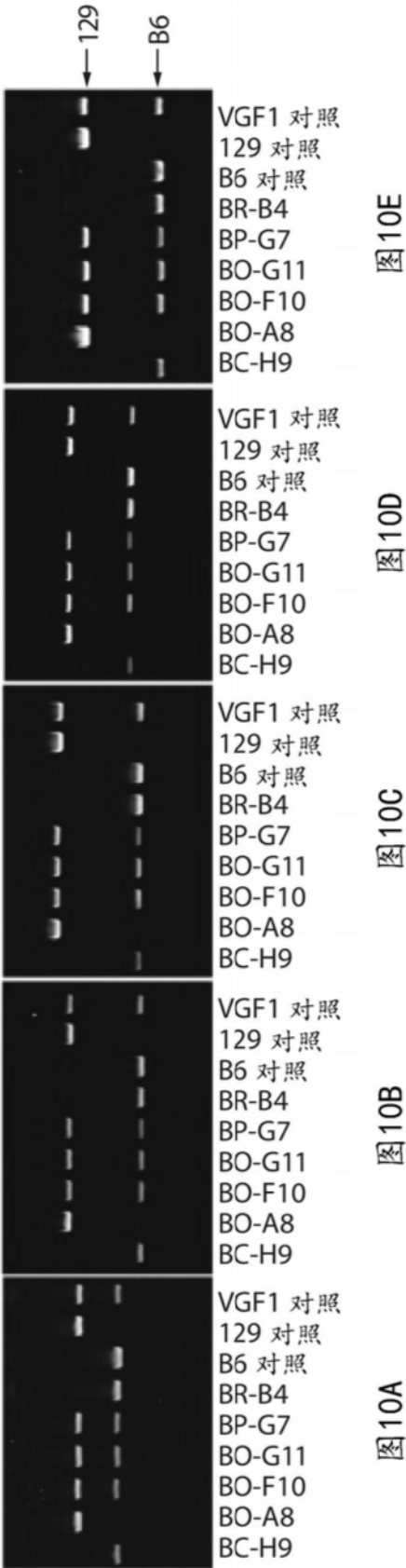


图9



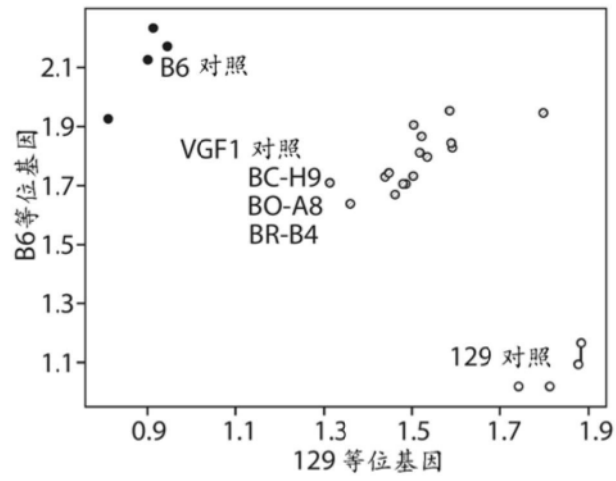


图11A

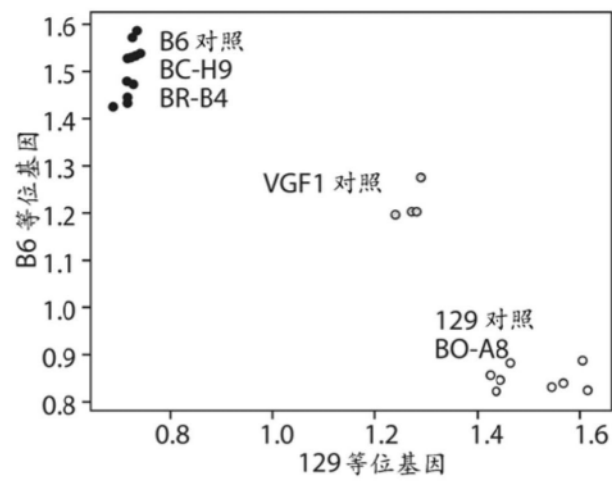


图11B

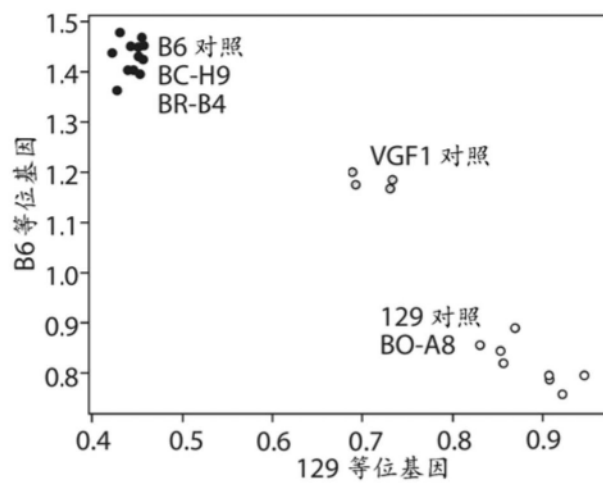


图11C

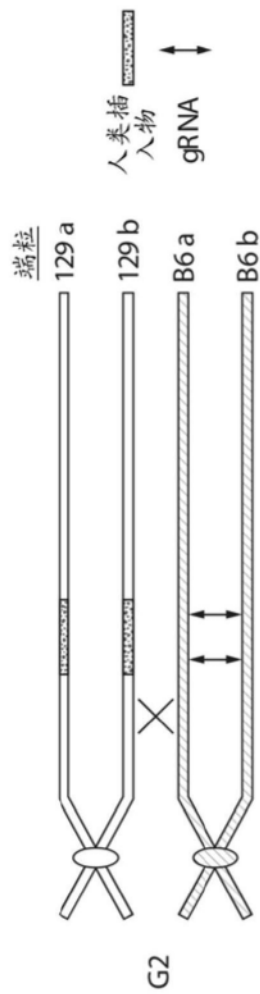


图12A

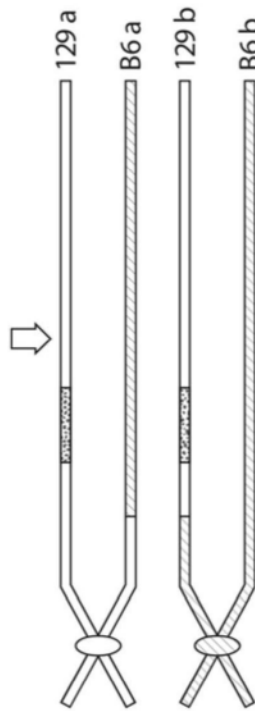


图12B

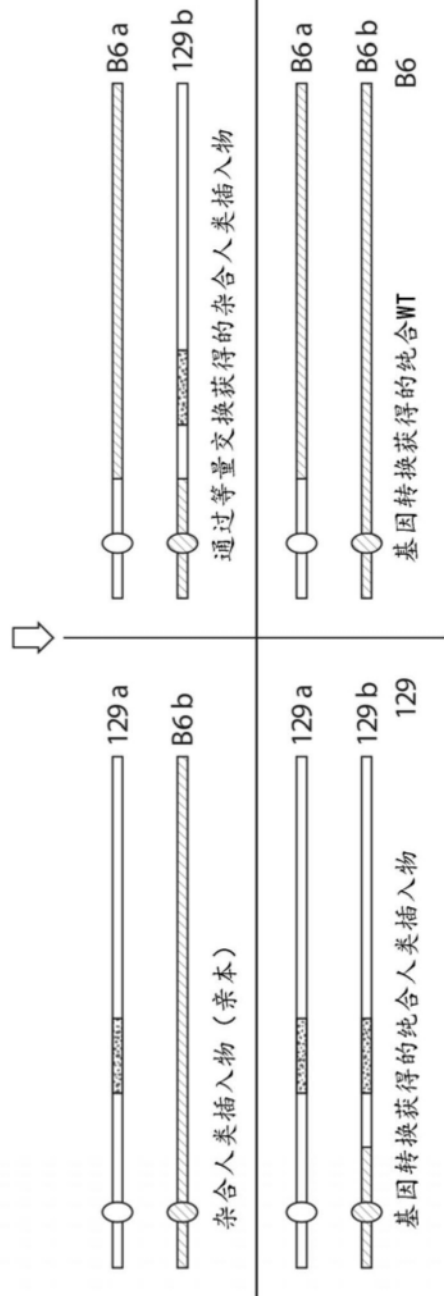


图12C

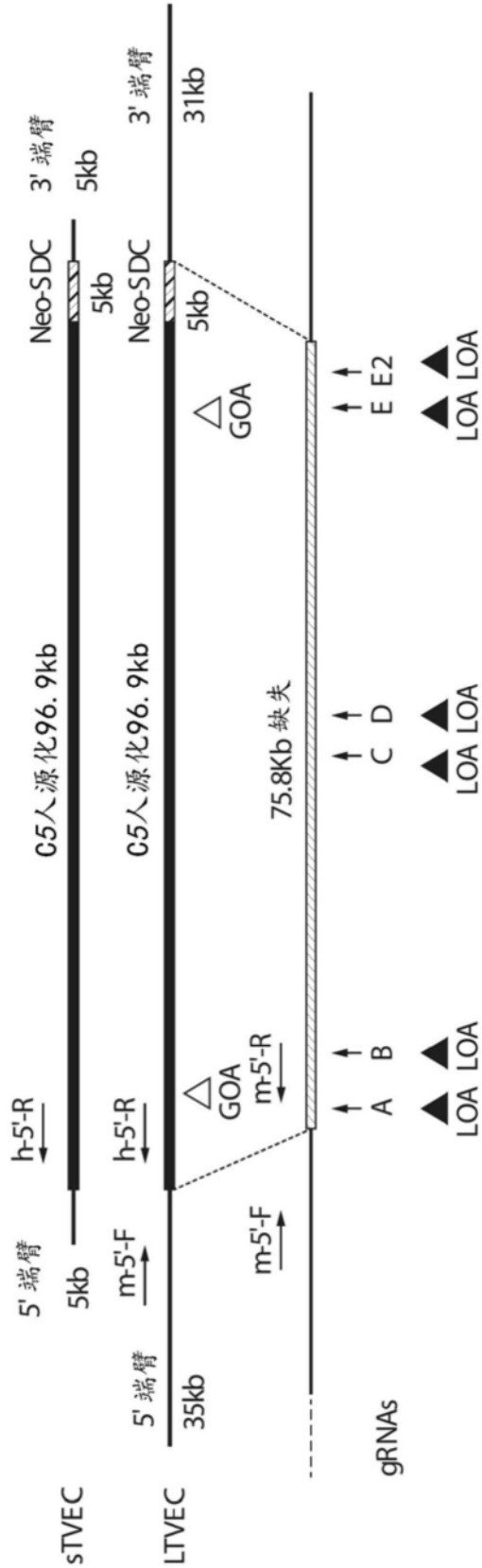


图13

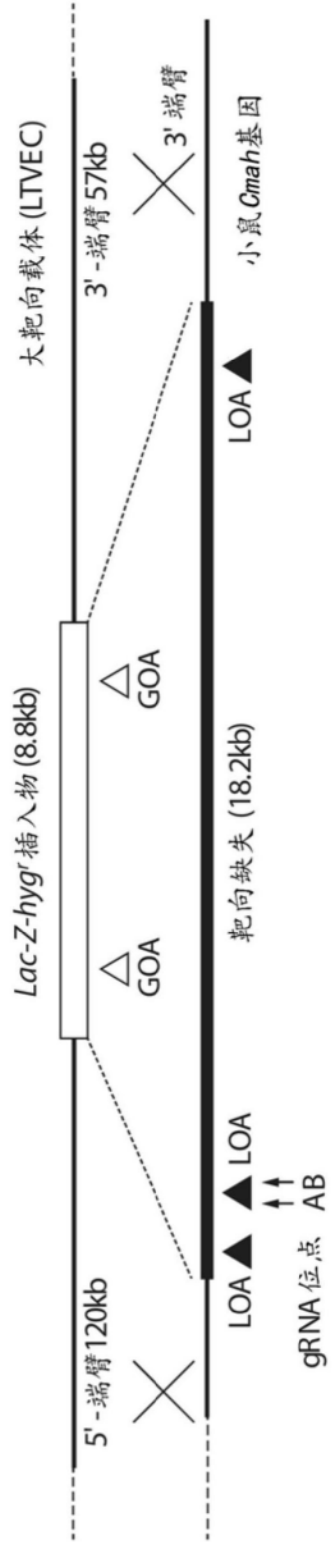


图14

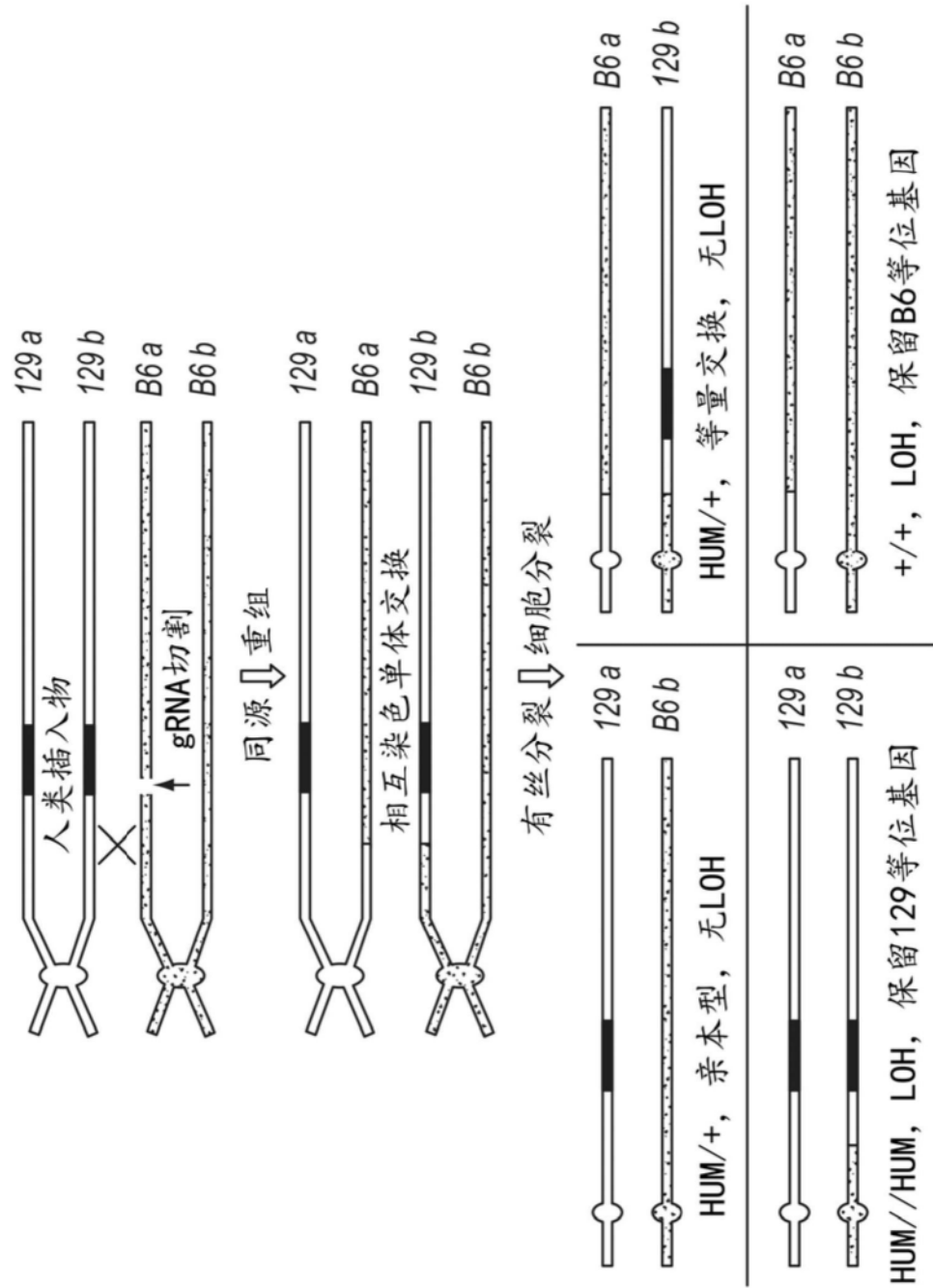


图16A

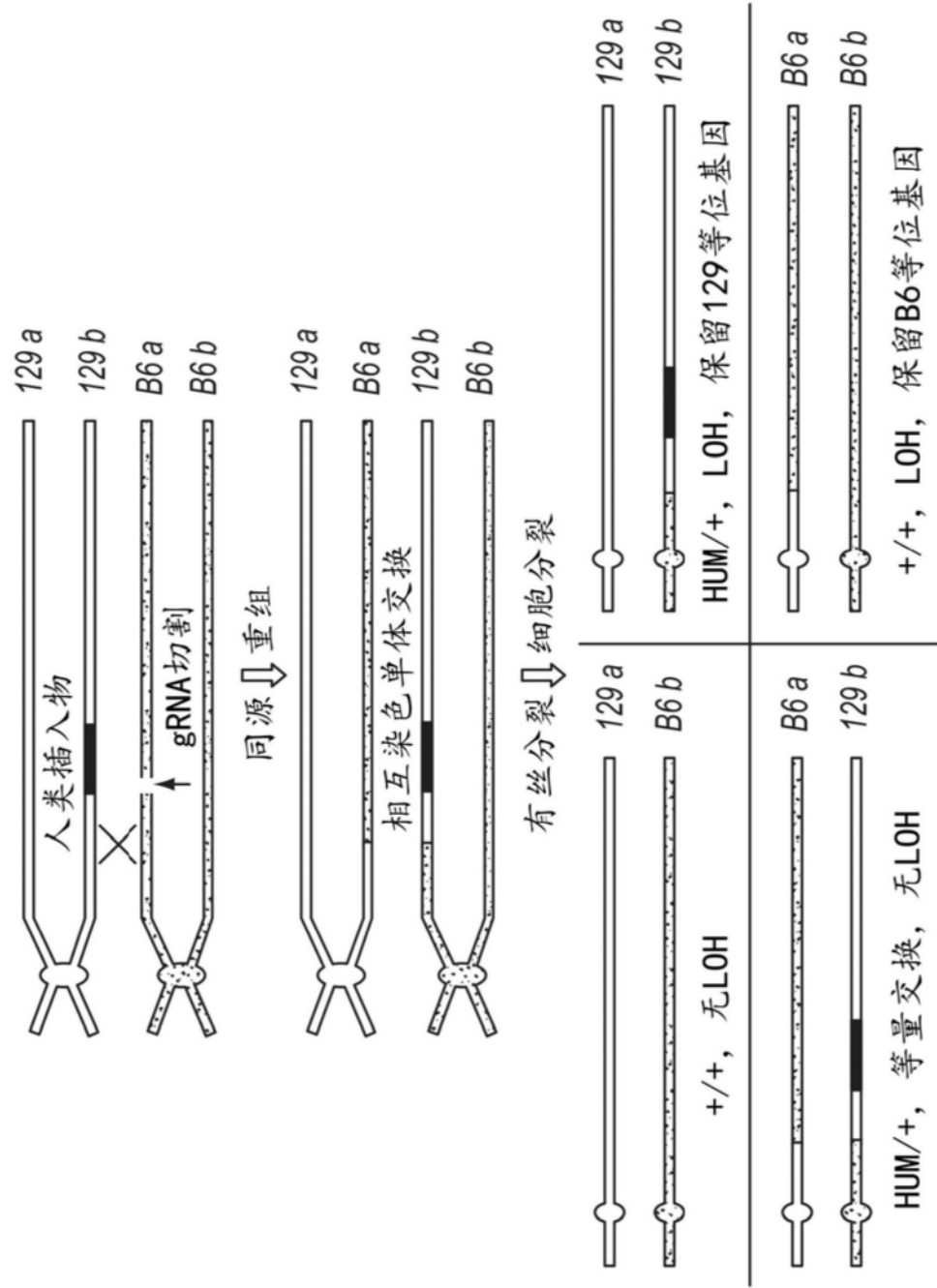


图16B

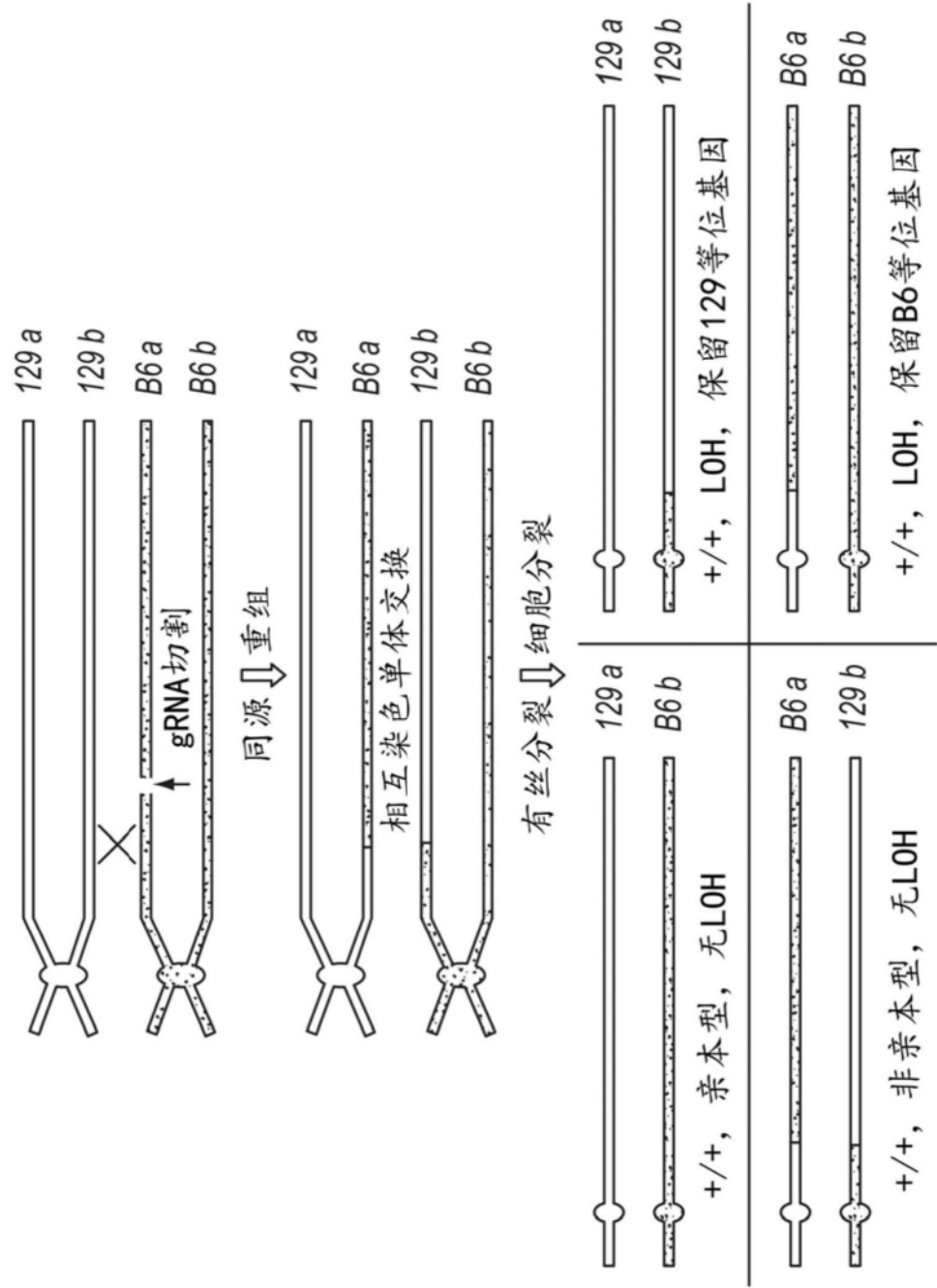


图16C

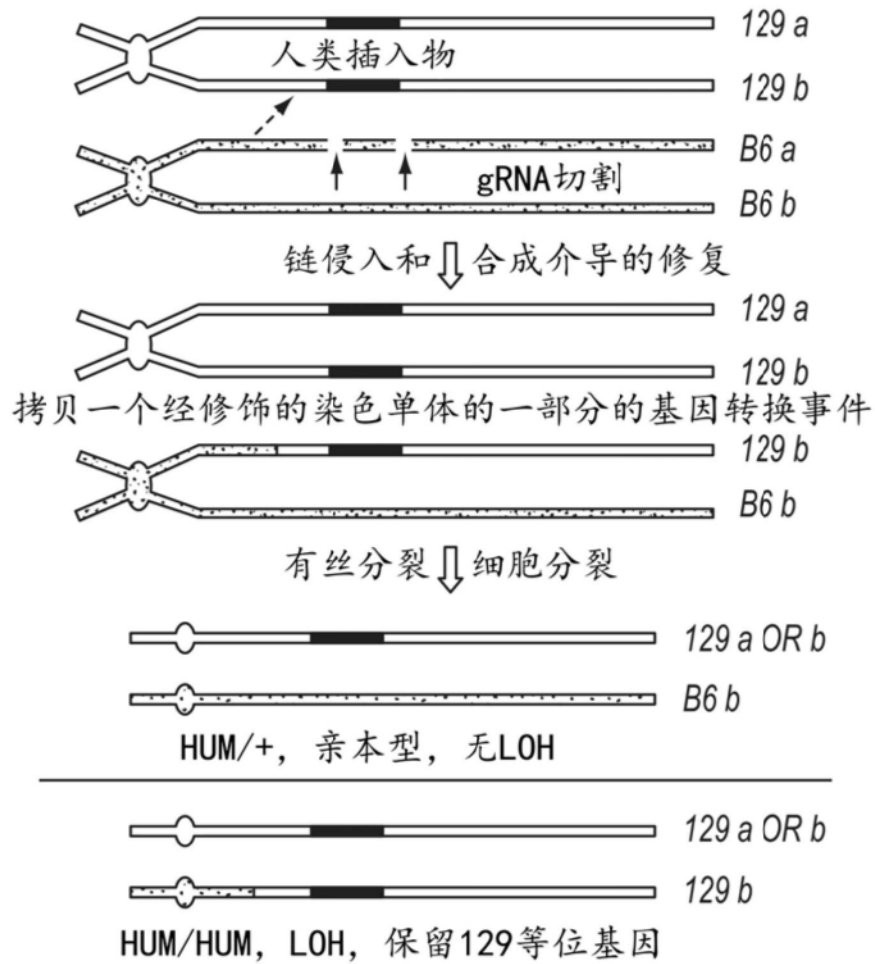


图16D

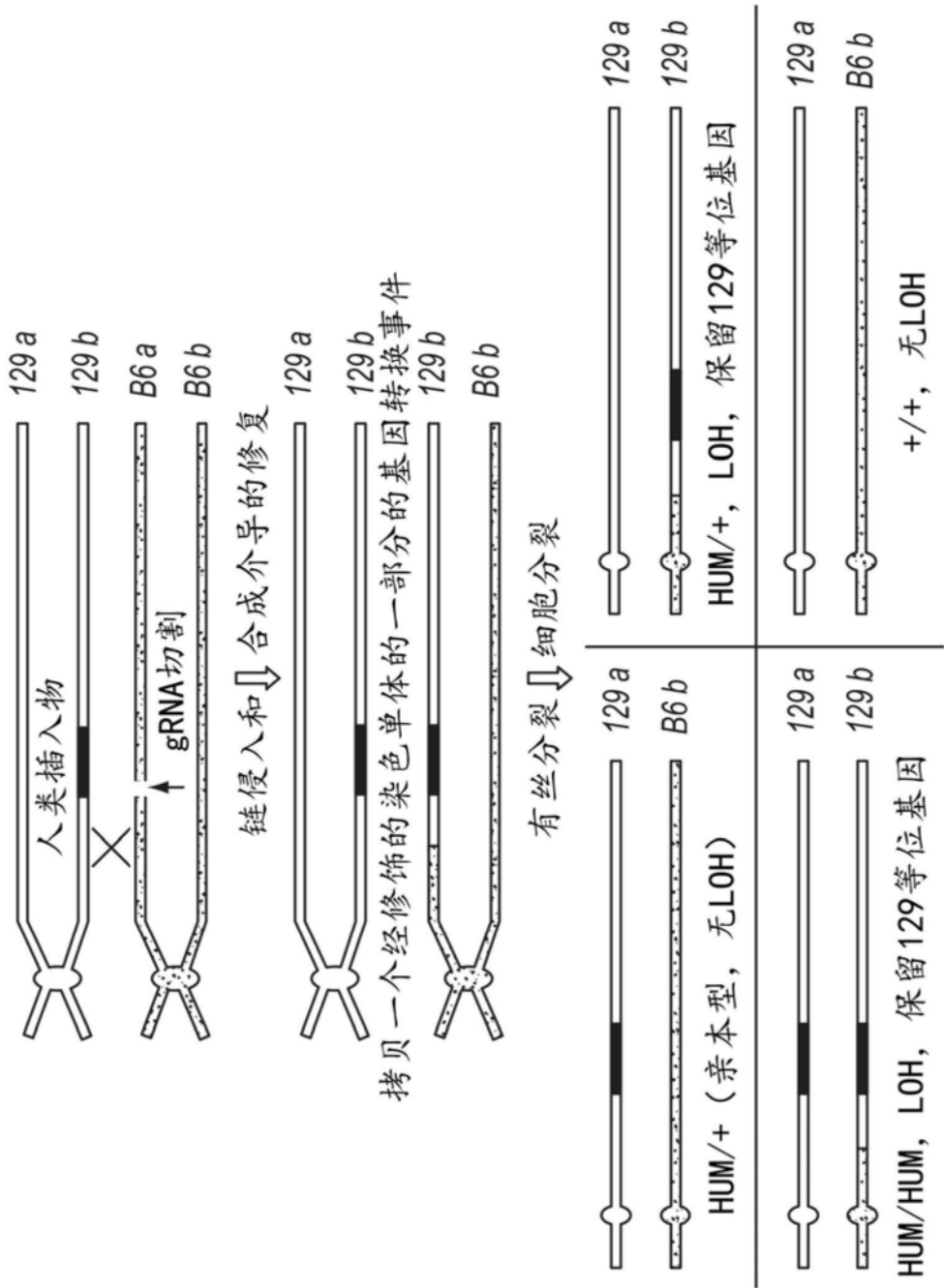


图16E

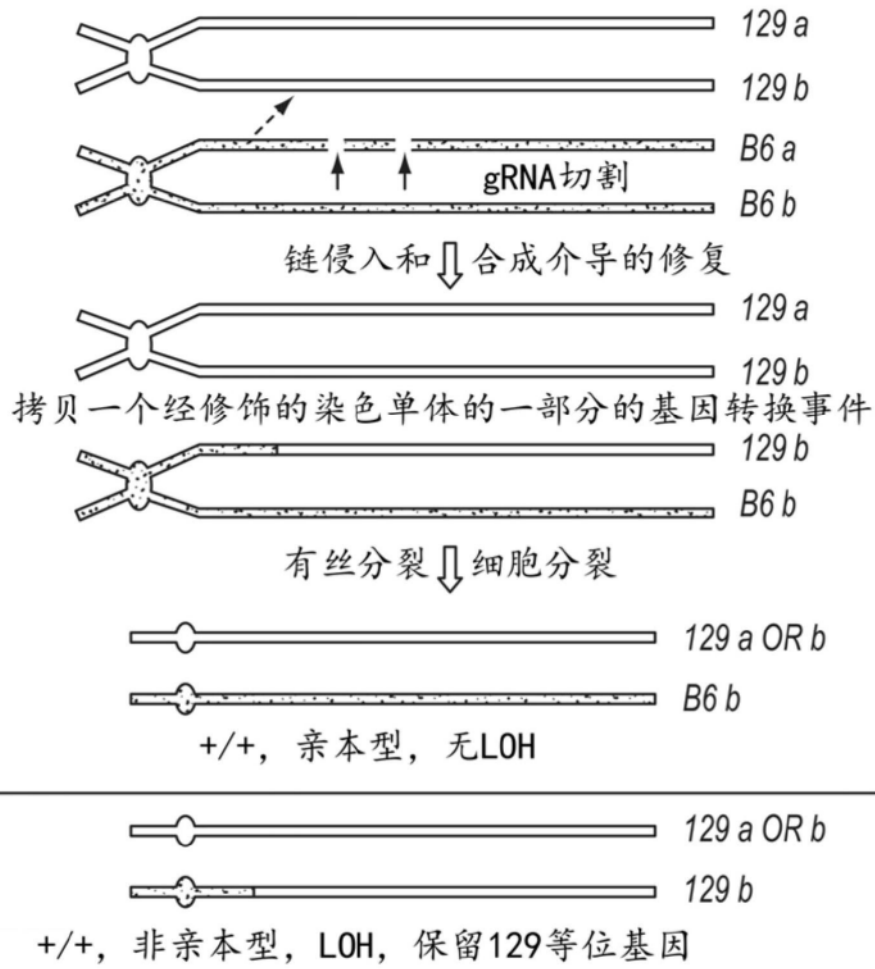


图16F

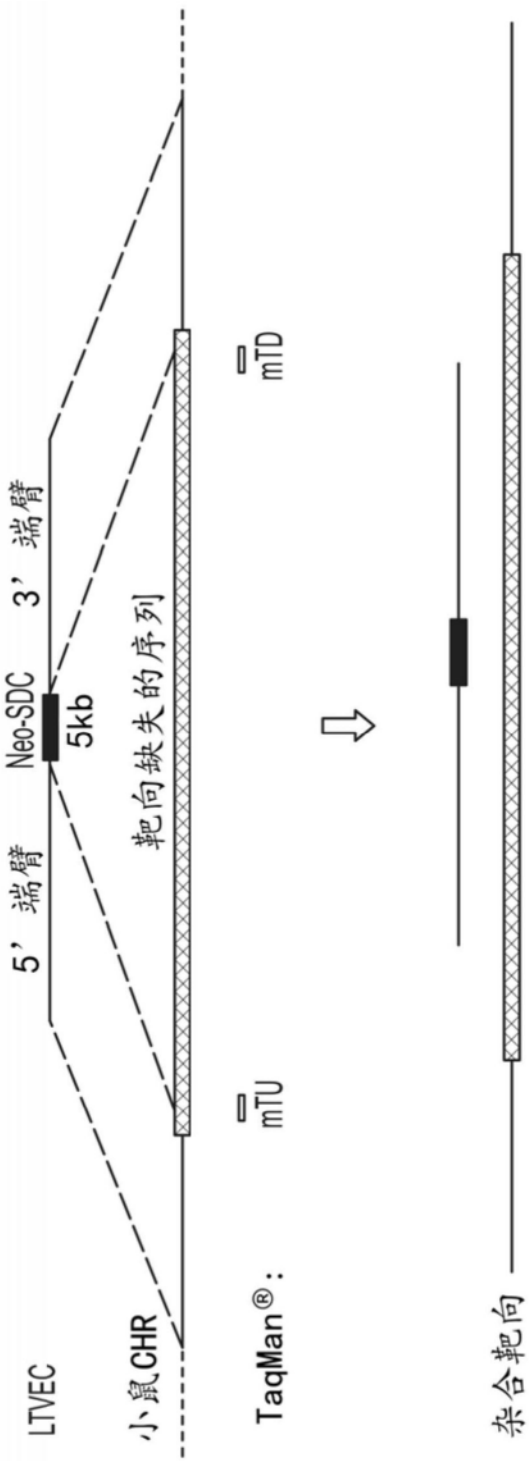


图17A

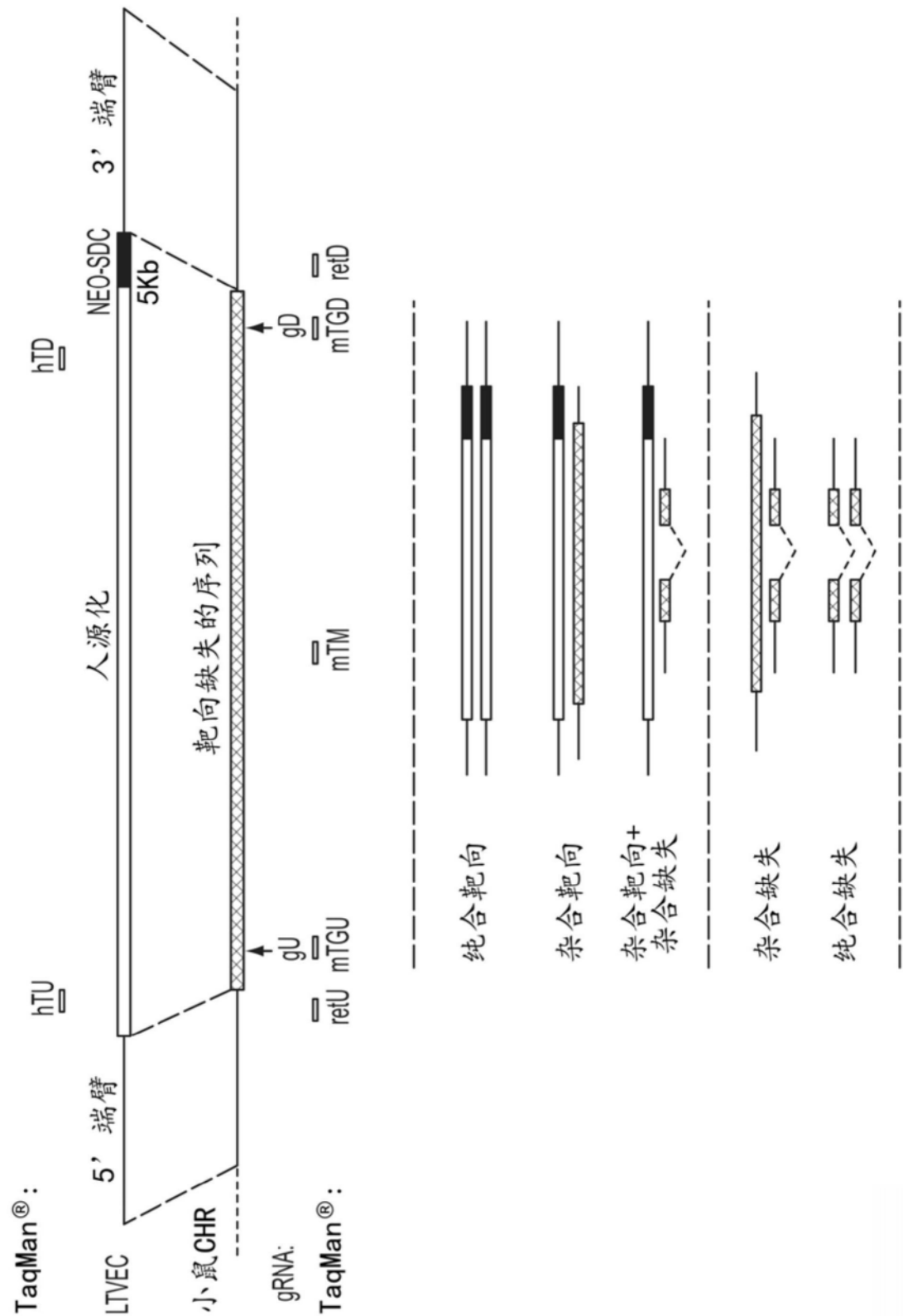


图17B

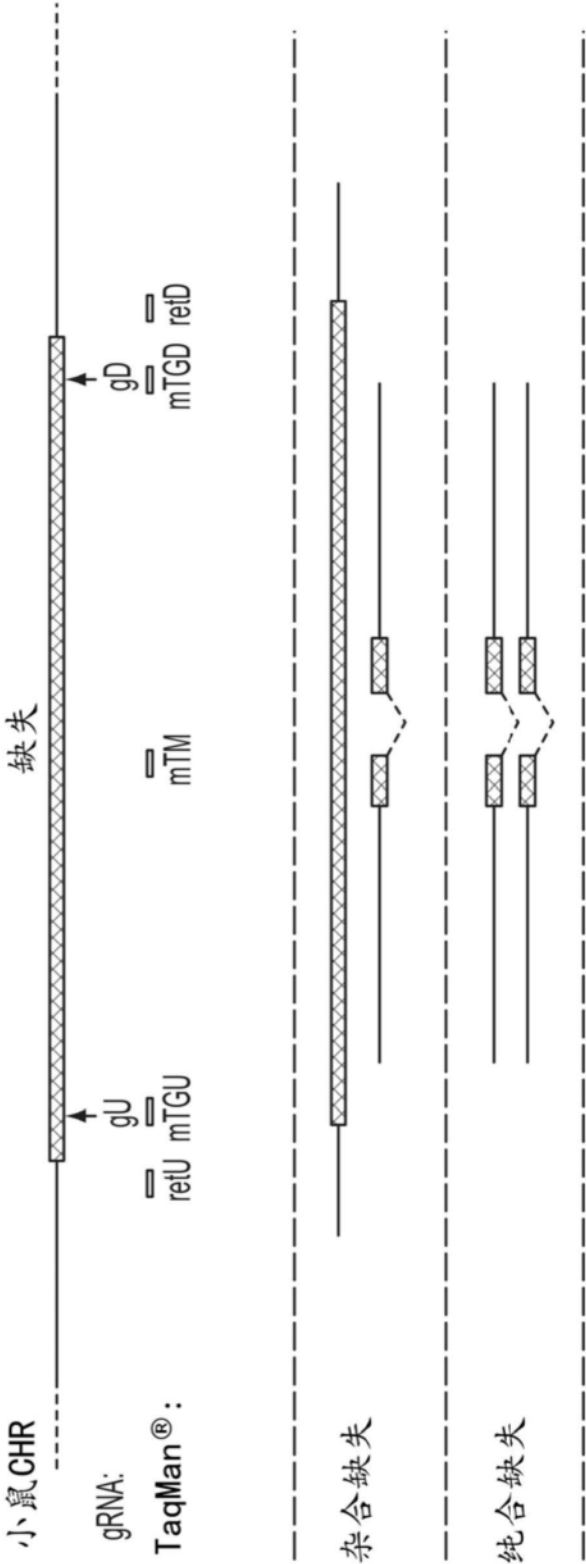


图17C

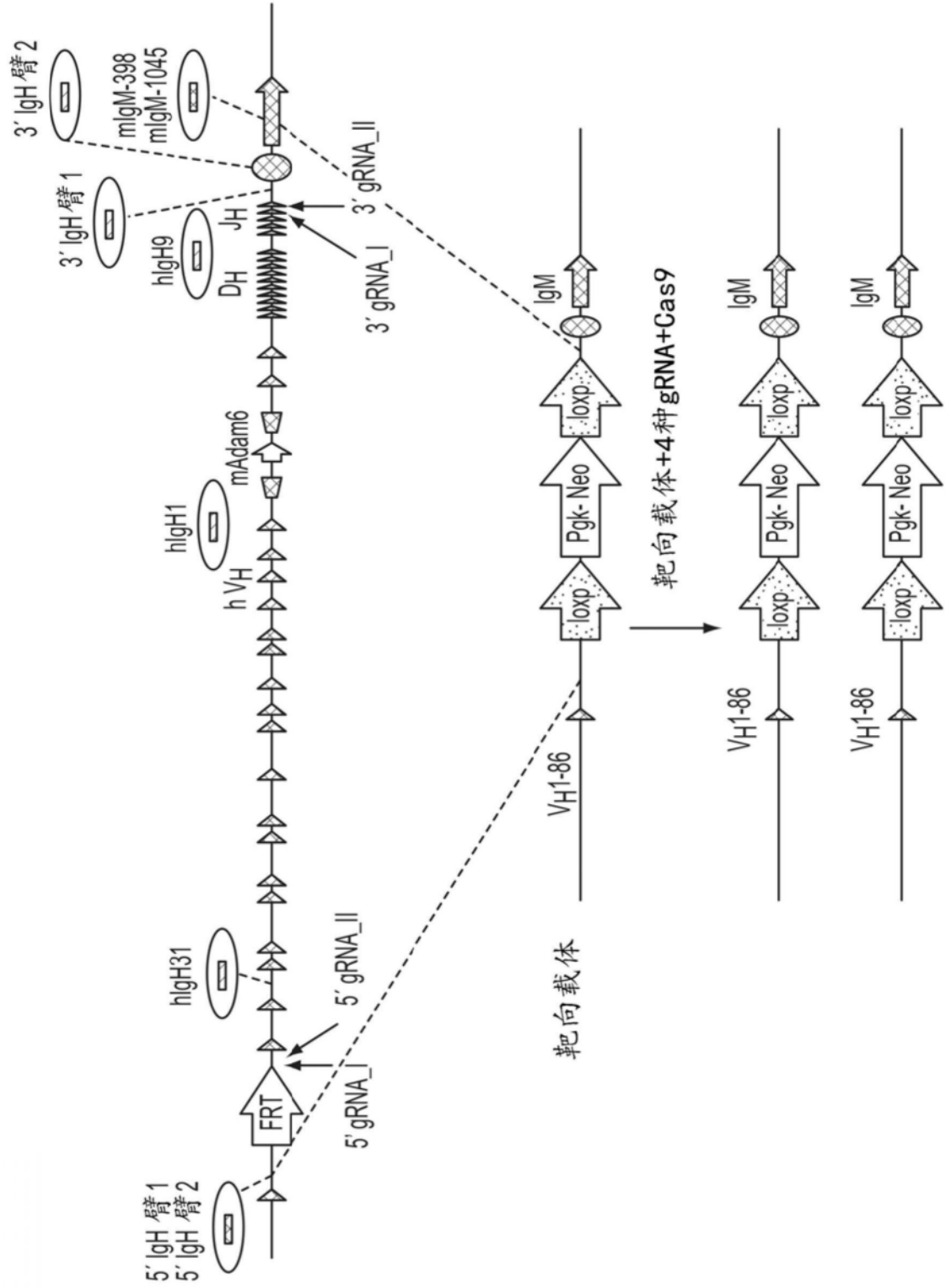


图18