



(12) 发明专利

(10) 授权公告号 CN 109475109 B

(45) 授权公告日 2021. 10. 29

(21) 申请号 201780044071.3  
(22) 申请日 2017.05.19  
(65) 同一申请的已公布的文献号  
申请公布号 CN 109475109 A  
(43) 申请公布日 2019.03.15  
(30) 优先权数据  
62/339,472 2016.05.20 US  
62/368,604 2016.07.29 US  
(85) PCT国际申请进入国家阶段日  
2019.01.16  
(86) PCT国际申请的申请数据  
PCT/US2017/033648 2017.05.19  
(87) PCT国际申请的公布数据  
W02017/201476 EN 2017.11.23  
(73) 专利权人 瑞泽恩制药公司  
地址 美国纽约州塔里敦  
(72) 发明人 维拉·沃罗宁那 林恩·麦克唐纳

马林·普里塞特  
卡曼·维纳斯·莱  
阿肖克·巴迪斯  
安德鲁·J·莫菲  
古斯塔沃·德罗格特  
大卫·佛伦杜威  
布莱恩·扎姆布罗维兹

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

代理人 郝传鑫

(51) Int. Cl.  
C07K 16/18 (2006.01)  
C12N 15/90 (2006.01)  
C12N 15/85 (2006.01)  
C07K 16/00 (2006.01)  
A01K 67/027 (2006.01)

审查员 陈龙

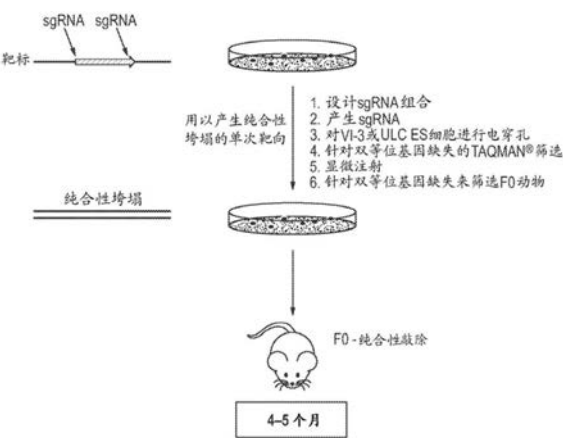
权利要求书9页 说明书134页  
序列表33页 附图41页

(54) 发明名称

用于使用多个引导RNA来破坏免疫耐受性的方法

(57) 摘要

提供用于制备对目标外来抗原的耐受性降低的非人动物以及使用所述动物来制备针对所述目标外来抗原的抗原结合蛋白的方法和组合物。所述方法和组合物采用使用多个引导RNA的CRISPR/Cas9系统来使与目标外来抗原同源或共有目标表位的自身抗原的表达降低或消除,或使所述自身抗原上的与目标外来抗原共有的表位的表达降低或消除。



1. 一种产生针对目标外来抗原的抗原结合蛋白的方法,其包括:

(a) 制备对目标外来抗原的耐受性降低的经遗传修饰的、为大鼠或小鼠的非人动物,其包括:

(i) 向一群大鼠或小鼠单细胞期胚胎或一群大鼠或小鼠胚胎干 (ES) 细胞中引入:

(I) Cas9蛋白或编码所述Cas9蛋白的核酸;

(II) 杂交至靶基因组基因座内的第一引导RNA识别序列的第一引导RNA或编码所述第一引导RNA的DNA,其中所述靶基因组基因座包含编码与所述目标外来抗原同源或共有目标表位的自身抗原的基因的全部或部分;和

(III) 杂交至所述靶基因组基因座内的第二引导RNA识别序列的第二引导RNA或编码所述第二引导RNA的DNA;

(ii) 从所述一群大鼠或小鼠单细胞期胚胎或所述一群大鼠或小鼠ES细胞中筛选经修饰的大鼠或小鼠单细胞期胚胎或经修饰的大鼠或小鼠ES细胞,其中在一对相应第一染色体和第二染色体中所述靶基因组基因座被修饰以产生具有双等位基因修饰的所述经修饰的大鼠或小鼠单细胞期胚胎或所述经修饰的大鼠或小鼠ES细胞,其中所述自身抗原的表达被消除;和

(iii) 从所述经修饰的大鼠或小鼠单细胞期胚胎或所述经修饰的大鼠或小鼠ES细胞产生经遗传修饰的F0代大鼠或小鼠,其中在所述经遗传修饰的F0代大鼠或小鼠中的所述一对相应第一染色体和第二染色体中所述靶基因组基因座被修饰以使所述自身抗原的表达被消除;

(b) 用所述目标外来抗原使在步骤(a)中产生的所述经遗传修饰的F0代大鼠或小鼠免疫;和

(c) 在足以引发对所述目标外来抗原的免疫应答的条件下维持所述经遗传修饰的F0代大鼠或小鼠,其中所述经遗传修饰的F0代大鼠或小鼠产生针对所述目标外来抗原的抗原结合蛋白。

2. 如权利要求1所述的方法,其中所述Cas9蛋白具有双链断裂诱导活性,并且其中所述Cas9蛋白和所述引导RNA在所述靶基因组基因座内的不同位点处产生成对双链断裂。

3. 如权利要求1或2所述的方法,其中步骤(a) (i) 中的所述细胞或胚胎是所述大鼠或小鼠ES细胞,并且在步骤(a) (iii) 中产生所述经遗传修饰的F0代大鼠或小鼠包括:

(I) 将所述经修饰的大鼠或小鼠ES细胞引入宿主胚胎中;和

(II) 将所述宿主胚胎植入代孕母体中以产生所述经遗传修饰的F0代大鼠或小鼠,其中所述一对相应第一染色体和第二染色体中的所述靶基因组基因座被修饰以使所述自身抗原的表达被消除。

4. 如权利要求1或2所述的方法,其中步骤(a) (i) 中的所述细胞或胚胎是所述大鼠或小鼠单细胞期胚胎,并且在步骤(a) (iii) 中产生所述经遗传修饰的F0代大鼠或小鼠包括将所述经修饰的大鼠或小鼠单细胞期胚胎植入代孕母体中以产生所述经遗传修饰的F0代大鼠或小鼠,其中所述一对相应第一染色体和第二染色体中的所述靶基因组基因座被修饰以使所述自身抗原的表达被消除。

5. 如权利要求1或2所述的方法,其进一步包括:

(d) (i) 从分离自经免疫的经遗传修饰F0代大鼠或小鼠的B细胞制备杂交瘤;或

(ii) 从所述经免疫的经遗传修饰F0代大鼠或小鼠获得编码针对所述目标外来抗原的所述抗原结合蛋白的免疫球蛋白重链可变结构域的第一核酸序列和/或编码针对所述目标外来抗原的所述抗原结合蛋白的免疫球蛋白轻链可变结构域的第二核酸序列。

6. 如权利要求5所述的方法, 其中所述第一核酸序列和/或所述第二核酸序列从所述经遗传修饰的F0代大鼠或小鼠的淋巴细胞或从由所述淋巴细胞产生的杂交瘤获得。

7. 如权利要求5所述的方法, 其中所述经遗传修饰的F0代小鼠或大鼠包含人源化免疫球蛋白基因座, 并且其中所述第一核酸序列编码人免疫球蛋白重链可变结构域, 并且所述第二核酸序列编码人免疫球蛋白轻链可变结构域。

8. 如权利要求1、2、6和7中任一项所述的方法, 其中相比于在用所述目标外来抗原使对照大鼠或小鼠免疫之后由在所述靶基因组基因座处是野生型的所述对照大鼠或小鼠产生的抗原结合蛋白, 由所述经遗传修饰的F0代大鼠或小鼠产生的针对所述目标外来抗原的所述抗原结合蛋白具有更高效价。

9. 如权利要求1、2、6和7中任一项所述的方法, 其中相较于在用所述目标外来抗原使对照大鼠或小鼠免疫之后由在所述靶基因组基因座处是野生型的所述对照大鼠或小鼠产生的抗原结合蛋白, 在用所述目标外来抗原使所述经遗传修饰的F0代大鼠或小鼠免疫之后由所述经遗传修饰的F0代大鼠或小鼠产生更具多样性的针对所述目标外来抗原的抗原结合蛋白谱系。

10. 如权利要求1、2、6和7中任一项所述的方法, 其中相较于在用所述目标外来抗原使对照大鼠或小鼠免疫之后由在所述靶基因组基因座处是野生型的所述对照大鼠或小鼠产生的抗原结合蛋白, 由所述经遗传修饰的F0代大鼠或小鼠产生的针对所述目标外来抗原的所述抗原结合蛋白使用更大多样性的重链V基因区段和/或轻链V基因区段。

11. 如权利要求1、2、6和7中任一项所述的方法, 其中由所述经遗传修饰的F0代大鼠或小鼠产生的一些针对所述目标外来抗原的所述抗原结合蛋白与所述自身抗原交叉反应。

12. 如权利要求1、2、6和7中任一项所述的方法, 其中在所述靶基因组基因座中, 所述第一引导RNA识别序列在所述第二引导RNA识别序列的5', 并且

其中步骤(a) (ii) 进一步包括进行保留测定以确定在所述第一引导RNA识别序列的5'以及1kb内的区域和/或在所述第二引导RNA识别序列的3' 以及1kb内的区域的拷贝数是2。

13. 如权利要求1、2、6和7中任一项所述的方法, 其中所述目标外来抗原是所述自身抗原的直系同源物。

14. 如权利要求1、2、6和7中任一项所述的方法, 其中所述目标外来抗原包含人蛋白质的全部或一部分。

15. 如权利要求1、2、6和7中任一项所述的方法, 其中所述靶基因组基因座被修饰以包含一个或多个核苷酸的插入、一个或多个核苷酸的缺失或一个或多个核苷酸的替换。

16. 如权利要求15所述的方法, 其中所述靶基因组基因座被修饰以包含所述一个或多个核苷酸的缺失。

17. 如权利要求16所述的方法, 其中所述缺失是无随机插入和缺失(插入缺失)的精确缺失。

18. 如权利要求1、2、6、7、16和17中任一项所述的方法, 其中所述第一引导RNA识别序列包含编码所述自身抗原的所述基因的起始密码子, 或在所述起始密码子的10、20、30、40、

50、100、200、300、400、500或1,000个核苷酸内,并且所述第二引导RNA识别序列包含编码所述自身抗原的所述基因的终止密码子,或在所述终止密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。

19.如权利要求1、2、6、7、16和17中任一项所述的方法,其中所述第一引导RNA识别序列和所述第二引导RNA识别序列不同,并且所述第一引导RNA识别序列和所述第二引导RNA识别序列中的每一个包含编码所述自身抗原的所述基因的起始密码子,或在所述起始密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。

20.如权利要求1、2、6、7、16和17中任一项所述的方法,其中所述靶基因组基因座被修饰以包含在0.1kb至200kb之间的双等位基因缺失。

21.如权利要求1、2、6、7、16和17中任一项所述的方法,其中所述修饰包括编码所述自身抗原的所述基因的全部或一部分的双等位基因缺失。

22.如权利要求1、2、6、7、16和17中任一项所述的方法,其中所述修饰包括编码所述自身抗原的所述基因的起始密码子的双等位基因破坏。

23.如权利要求1、2、6、7、16和17中任一项所述的方法,其中引入步骤(a)(i)进一步包括向所述大鼠或小鼠ES细胞或所述大鼠或小鼠单细胞期胚胎中引入:

(iv)杂交至所述靶基因组基因座内的第三引导RNA识别序列的第三引导RNA或编码所述第三引导RNA的DNA;和/或

(v)杂交至所述靶基因组基因座内的第四引导RNA识别序列的第四引导RNA或编码所述第四引导RNA的DNA。

24.如权利要求1、2、6、7、16和17中任一项所述的方法,其中步骤(a)(i)中的所述细胞或胚胎是所述大鼠或小鼠ES细胞,并且将所述Cas9蛋白、所述第一引导RNA和所述第二引导RNA各自以DNA形式引入所述大鼠或小鼠ES细胞中。

25.如权利要求1、2、6、7、16和17中任一项所述的方法,其中步骤(a)(i)中的所述细胞或胚胎是所述大鼠或小鼠ES细胞,并且将所述Cas9蛋白、所述第一引导RNA和所述第二引导RNA各自通过电穿孔或核转染来引入所述大鼠或小鼠ES细胞中。

26.如权利要求1、2、6、7、16和17中任一项所述的方法,其中步骤(a)(i)中的所述细胞或胚胎是所述大鼠或小鼠单细胞期胚胎,并且将所述Cas9蛋白、所述第一引导RNA和所述第二引导RNA各自以RNA形式引入所述大鼠或小鼠单细胞期胚胎中。

27.如权利要求1、2、6、7、16和17中任一项所述的方法,其中步骤(a)(i)中的所述细胞或胚胎是所述大鼠或小鼠单细胞期胚胎,并且将所述Cas9蛋白、所述第一引导RNA和所述第二引导RNA通过原核注射或细胞质注射来引入所述大鼠或小鼠单细胞期胚胎中。

28.如权利要求1、2、6、7、16和17中任一项所述的方法,其中不在步骤(a)(i)中引入外源性修复模板。

29.如权利要求1、2、6、7、16和17中任一项所述的方法,其中所述引入步骤(a)(i)进一步包括向所述大鼠或小鼠ES细胞或所述大鼠或小鼠单细胞期胚胎中引入外源性修复模板,所述外源性修复模板包含在所述靶基因组基因座处杂交至5'靶标序列的5'同源臂和在所述靶基因组基因座处杂交至3'靶标序列的3'同源臂,前提是如果步骤(a)(i)中的所述细胞或胚胎是所述大鼠或小鼠单细胞期胚胎,那么所述外源性修复模板的长度是不超过5kb。

30.如权利要求29所述的方法,其中所述外源性修复模板进一步包含由所述5'同源臂

和所述3'同源臂侧接的核酸插入物。

31.如权利要求30所述的方法,其中所述核酸插入物与所述靶基因组基因座同源。

32.如权利要求30所述的方法,其中所述核酸插入物与所述靶基因组基因座直系同源。

33.如权利要求29所述的方法,其中所述外源性修复模板的长度在50个核苷酸至1kb之间。

34.如权利要求33所述的方法,其中所述外源性修复模板的长度在80个核苷酸至200个核苷酸之间。

35.如权利要求29所述的方法,其中所述外源性修复模板是单链寡脱氧核苷酸。

36.如权利要求29所述的方法,其中步骤(a)(i)中的所述细胞或胚胎是所述大鼠或小鼠ES细胞,并且其中:

(a)所述外源性修复模板是长度是至少10kb的大型靶向载体(LTVEC);或

(b)所述外源性修复模板是LTVEC,其中所述LTVEC的5'同源臂和3'同源臂的总和的长度是至少10kb。

37.如权利要求29所述的方法,其中所述靶基因组基因座被修饰以包含一个或多个核苷酸的缺失,并且其中经缺失核酸序列由在所述5'靶标序列与所述3'靶标序列之间的核酸序列组成。

38.如权利要求29所述的方法,其中所述外源性修复模板包含由所述5'同源臂和所述3'同源臂侧接的核酸插入物,

其中所述核酸插入物与经缺失核酸序列同源,

其中所述靶基因组基因座被修饰以包含一个或多个核苷酸的缺失,并且

其中所述核酸插入物替换所述经缺失核酸序列。

39.如权利要求29所述的方法,其中所述外源性修复模板包含由所述5'同源臂和所述3'同源臂侧接的核酸插入物,

其中所述核酸插入物与经缺失核酸序列直系同源,

其中所述靶基因组基因座被修饰以包含一个或多个核苷酸的缺失,并且

其中所述核酸插入物替换所述经缺失核酸序列。

40.如权利要求1、2、6、7、16、17和30-39中任一项所述的方法,其中所述大鼠或小鼠包含人源化免疫球蛋白基因座。

41.如权利要求1、2、6、7、16、17和30-39中任一项所述的方法,其中所述非人动物是所述小鼠。

42.如权利要求41所述的方法,其中小鼠品系包括BALB/c品系。

43.如权利要求42所述的方法,其中所述小鼠品系包括BALB/c、C57BL/6和129品系的混合。

44.如权利要求43所述的方法,其中所述小鼠品系是50%

BALB/c、25% C57BL/6和25% 129。

45.如权利要求41所述的方法,其中所述小鼠的MHC单倍型是MHC<sup>b/d</sup>。

46.如权利要求41所述的方法,其中所述小鼠在它的种系中包含插入在内源性小鼠免疫球蛋白可变区基因座处的人未重排可变区基因区段。

47.如权利要求46所述的方法,其中所述人未重排可变区基因区段是重链基因区段,并

且所述小鼠免疫球蛋白基因座是重链基因座,和/或其中所述人未重排可变区基因区段是 $\kappa$ 链或 $\lambda$ 轻链区段,并且所述小鼠免疫球蛋白基因座是轻链基因座。

48.如权利要求46或47所述的方法,其中所述小鼠在它的种系中包含可操作地连接至小鼠恒定区基因的人未重排可变区基因区段,其中所述小鼠缺乏人恒定区基因,并且其中所述小鼠恒定区基因在内源性小鼠免疫球蛋白基因座处。

49.如权利要求46或47所述的方法,其中所述小鼠包含:

(a) 杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和

(b) 杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;

其中(a)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(b)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中所述小鼠不能形成包含人可变区和人恒定区的抗体。

50.如权利要求41所述的方法,其中所述小鼠在它的种系中包含可操作地连接至小鼠轻链恒定区的人源化免疫球蛋白轻链可变基因座,其包含至多一个或至多两个重排人轻链V/J序列,并且其中所述小鼠进一步包含可操作地连接至小鼠重链恒定区基因的人源化免疫球蛋白重链可变基因座,其包含至少一个未重排人V、至少一个未重排人D和至少一个未重排人J区段。

51.如权利要求50所述的方法,其中所述小鼠包含人源化重链免疫球蛋白可变基因座和人源化轻链免疫球蛋白可变基因座,其中所述小鼠表达单一轻链。

52.如权利要求51所述的方法,其中所述小鼠包含:

(a) 编码免疫球蛋白轻链的人 $V_L$ 结构域的单一重排人免疫球蛋白轻链可变区 $V_L/J_L$ ,其中所述单一重排人 $V_L/J_L$ 区选自人 $V_{\kappa}1-39/J_{\kappa}5$ 基因区段或人 $V_{\kappa}3-20/J_{\kappa}1$ 基因区段;和

(b) 内源性重链可变 $V_H$ 基因区段被一个或多个人 $V_H$ 基因区段的替换,其中所述人 $V_H$ 基因区段可操作地连接至内源性重链恒定 $C_H$ 区基因,并且所述人 $V_H$ 基因区段能够重排并形成人/小鼠嵌合重链基因。

53.如权利要求51所述的方法,其中所述小鼠表达抗体的群体,并且所述小鼠的种系仅包括是重排人种系 $\kappa$ 轻链可变区基因的单一免疫球蛋白 $\kappa$ 轻链可变区基因,

其中所述小鼠就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是杂合的,因为它仅含有一个拷贝,或就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是纯合的,因为它含有两个拷贝,所述小鼠的特征在于活性亲和力成熟,以使:

(i) 所述群体的各免疫球蛋白 $\kappa$ 轻链都包含由所述重排人种系 $\kappa$ 轻链可变区基因或由其体细胞突变变体编码的轻链可变结构域;

(ii) 所述群体包括包含其轻链可变结构域由所述重排人种系 $\kappa$ 轻链可变区基因编码的免疫球蛋白 $\kappa$ 轻链的抗体和包含其轻链可变结构域由所述其体细胞突变变体编码的免疫球蛋白 $\kappa$ 轻链的抗体;以及

(iii) 所述小鼠产生与所述免疫球蛋白 $\kappa$ 轻链成功配对以形成所述群体所述抗体的体细胞突变的高亲和力重链的多样性集合。

54. 如权利要求51所述的方法,其中所述小鼠在它的种系中就以下来说是杂合的或纯合的:

(a) 重排V<sub>κ</sub>/J<sub>κ</sub>序列在内源性小鼠κ免疫球蛋白轻链可变区基因座处的插入,所述重排V<sub>κ</sub>/J<sub>κ</sub>序列包含:

(i) 单一人种系V<sub>κ</sub>序列,所述单一人种系V<sub>κ</sub>序列存在于SEQ ID NO:148或SEQ ID NO:149中;和

(ii) 单一人种系J<sub>κ</sub>序列,其中所述重排V<sub>κ</sub>/J<sub>κ</sub>序列可操作地连接至内源性小鼠κ恒定区;和

(b) 多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

55. 如权利要求46、47和50-54中任一项所述的方法,其中所述小鼠包含对免疫球蛋白重链基因座的修饰,其中所述修饰使内源性ADAM6功能降低或消除,

其中所述小鼠包含编码小鼠ADAM6蛋白、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其同源物或其片段在雄性小鼠中具有功能性,并且

其中编码所述小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的所述异位核酸序列存在于人重链可变区基因座处。

56. 如权利要求1、2、6、7、16、17、30-39、42-47和50-54中任一项所述的方法,其中所述非人动物是所述小鼠,其中所述小鼠至少部分地源于BALB/c品系,并且所述小鼠包含人源化免疫球蛋白基因座,

其中所述目标外来抗原是与所述自身抗原直系同源的人蛋白质的全部或一部分,

其中所述第一引导RNA识别序列包含编码所述自身抗原的所述基因的起始密码子,或在所述起始密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且所述第二引导RNA识别序列包含编码所述自身抗原的所述基因的终止密码子,或在所述终止密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且

其中所述修饰包括编码所述自身抗原的所述基因的全部或一部分的双等位基因缺失,由此所述自身抗原的表达被消除。

57. 如权利要求1、2、6、7、16、17、30-39、42-47和50-54中任一项所述的方法,其中所述非人动物是所述小鼠,其中所述小鼠至少部分地源于BALB/c品系,并且所述小鼠包含人源化免疫球蛋白基因座,

其中所述目标外来抗原是与所述自身抗原直系同源的人蛋白质的全部或一部分,

其中所述第一引导RNA识别序列包含编码所述自身抗原的所述基因的起始密码子或在所述起始密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且所述第二引导RNA识别序列包含编码所述自身抗原的所述基因的所述起始密码子,或在所述起始密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且

其中所述修饰包括对编码所述自身抗原的所述基因的起始密码子的双等位基因破坏,由此所述自身抗原的表达被消除。

58. 如权利要求56所述的方法,其中所述小鼠包含:

(a) 编码小鼠ADAM6蛋白、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其同源物或其片段在雄性小鼠中具有功能性;

(b) 杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和

(c) 杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;

其中(b)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(c)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中所述小鼠不能形成包含人可变区和人恒定区的抗体。

59. 如权利要求56所述的方法,其中所述小鼠在它的种系中就以下来说是杂合的或纯合的:

(a) 编码小鼠ADAM6蛋白、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其同源物或其片段在雄性小鼠中具有功能性;

(b) 重排V<sub>κ</sub>/J<sub>κ</sub>序列在内源性小鼠κ免疫球蛋白轻链可变区基因座处的插入,所述重排V<sub>κ</sub>/J<sub>κ</sub>序列包含:

(i) 单一人种系V<sub>κ</sub>序列,所述单一人种系V<sub>κ</sub>序列存在于SEQ ID NO:148或SEQ ID NO:149中;和

(ii) 单一人种系J<sub>κ</sub>序列,其中所述重排V<sub>κ</sub>/J<sub>κ</sub>序列可操作地连接至所述内源性小鼠κ恒定区;和

(c) 多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

60. 如权利要求1、2、6、7、16、17、30-39、42-47、50-54、58和59中任一项所述的方法,其中所述大鼠或小鼠ES细胞是杂交细胞,或所述大鼠或小鼠单细胞期胚胎是杂交单细胞期胚胎,并且其中所述方法进一步包括:

(a') 比较所述一对相应第一染色体和第二染色体在所述靶基因组基因座内的序列,以及在接触步骤(a)之前基于以下来选择所述靶基因组基因座内的靶区域:在所述一对相应第一染色体与第二染色体之间,相对于所述靶基因组基因座的其余部分的全部或一部分,所述靶区域具有更高序列同一性百分比,其中所述靶区域包含:

所述第一引导RNA识别序列和在所述第一引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb或10kb的侧接序列,和/或

所述第二引导RNA识别序列和在所述第二引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900 bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb或10kb的侧接序列。

61. 如权利要求60所述的方法,其中在所述一对相应第一染色体与第二之间,相对于所述靶基因组基因座的其余部分,所述靶区域具有更高序列同一性百分比。

62. 如权利要求61所述的方法,其中在所述一对相应第一染色体与第二染色体之间,所述靶区域具有至少99.9%序列同一性,并且在所述一对相应第一染色体与第二染色体之间,所述靶基因组基因座的其余部分具有不超过99.8%序列同一性。

63. 一种制备对目标外来抗原的耐受性降低的经遗传修饰的、为大鼠或小鼠的非人动物的方法,其包括:

(a) 向一群大鼠或小鼠单细胞期胚胎或不是单细胞期胚胎的一群大鼠或小鼠胚胎干(ES)细胞中引入:

(i) Cas9蛋白或编码所述Cas9蛋白的核酸;

(ii) 杂交至靶基因组基因座内的第一引导RNA识别序列的第一引导RNA或编码所述第一引导RNA的DNA,其中所述靶基因组基因座包含编码与所述目标外来抗原同源或共有目标表位的自身抗原的基因的全部或一部分;和

(iii) 杂交至所述靶基因组基因座内的第二引导RNA识别序列的第二引导RNA或编码所述第二引导RNA的DNA;

(b) 从所述一群大鼠或小鼠单细胞期胚胎或所述大鼠或小鼠ES细胞中筛选经修饰的大鼠或小鼠单细胞期胚胎或经修饰的大鼠或小鼠ES细胞,其中一对相应第一染色体和第二染色体中的所述靶基因组基因座被修饰以产生具有双等位基因修饰的所述经修饰的大鼠或小鼠单细胞期胚胎或所述经修饰的大鼠或小鼠ES细胞,其中所述自身抗原的表达被消除;和

(c) 从所述经修饰的大鼠或小鼠单细胞期胚胎或所述经修饰的大鼠或小鼠ES细胞产生经遗传修饰的F0代大鼠或小鼠,其中在所述经遗传修饰的F0代大鼠或小鼠中所述一对相应第一染色体和第二染色体中所述靶基因组基因座被修饰以使所述自身抗原的表达被消除。

64. 如权利要求63所述的方法,其中步骤(a)中的所述细胞或胚胎是所述大鼠或小鼠ES细胞,并且在步骤(c)中产生所述经遗传修饰的F0代大鼠或小鼠包括:

(I) 将所述经修饰大鼠或小鼠ES细胞引入宿主胚胎中;和

(II) 将所述宿主胚胎植入代孕母体中以产生所述经遗传修饰的F0代大鼠或小鼠,其中在所述一对相应第一染色体和第二染色体中所述靶基因组基因座被修饰以使所述自身抗原的表达被消除。

65. 如权利要求63所述的方法,其中步骤(a)中的所述细胞或胚胎是所述大鼠或小鼠单细胞期胚胎,并且在步骤(c)中产生所述经遗传修饰的F0代大鼠或小鼠包括将所述经修饰的大鼠或小鼠单细胞期胚胎植入代孕母体中以产生所述经遗传修饰的F0代大鼠或小鼠,其中在所述一对相应第一染色体和第二染色体中所述靶基因组基因座被修饰以使所述自身抗原的表达被消除。

66. 如权利要求63-65中任一项所述的方法,其中所述目标外来抗原是所述自身抗原的直系同源物。

67. 如权利要求63-65中任一项所述的方法,其中所述第一引导RNA识别序列包含编码所述自身抗原的所述基因的起始密码子,或在所述起始密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且所述第二引导RNA识别序列包含编码所述自身抗原的所述基因的终止密码子,或在所述终止密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。

68. 如权利要求63-65中任一项所述的方法, 其中所述第一引导RNA识别序列和所述第二引导RNA识别序列不同, 并且所述第一引导RNA识别序列和所述第二引导RNA识别序列中的每一个包含编码所述自身抗原的所述基因的起始密码子, 或在所述起始密码子的10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。

69. 如权利要求63-65中任一项所述的方法, 其中在所述靶基因组基因座中, 所述第一引导RNA识别序列在所述第二引导RNA识别序列的5', 并且

其中步骤(b)进一步包括进行保留测定以确定在所述第一引导RNA识别序列的5'以及1kb内的区域和/或在所述第二引导RNA识别序列的3'以及1kb内的区域的拷贝数是2。

70. 如权利要求63-65中任一项所述的方法, 其中所述修饰包括编码所述自身抗原的所述基因的全部或一部分的双等位基因缺失。

71. 如权利要求63-65中任一项所述的方法, 其中所述修饰包括编码所述自身抗原的所述基因的起始密码子的双等位基因破坏。

72. 如权利要求63-65中任一项所述的方法, 其中所述非人动物是所述小鼠。

## 用于使用多个引导RNA来破坏免疫耐受性的方法

[0001] 通过EFS网以文本文件形式提交的序列表的引用

[0002] 写入文件497023SEQLIST.txt中的序列表是38.3千字节,于2017年5月18日创建,并且据此以引用的方式并入。

### 背景技术

[0003] 用“非自身”蛋白质使非人动物(例如啮齿动物,诸如小鼠或大鼠)免疫是一种通常使用的用以获得特异性抗原结合蛋白诸如单克隆抗体的方法。然而,这个方法依赖于非人动物中的天然蛋白质与经免疫以使得非人动物的免疫系统能够将免疫原识别为非自身(即外来)的蛋白质之间的序列趋异性。产生针对与自身抗原具有高度同源性的抗原的抗体可由于免疫耐受性而是一项困难任务。因为蛋白质的在功能上重要的区域在物种之间倾向于保守,所以对自身抗原的免疫耐受性常常对产生针对这些关键表位的抗体造成挑战。

[0004] 尽管已在靶向各种基因组基因座方面取得进展,但仍然有许多不能以常规靶向策略高效靶向的基因组基因座或不能以常规靶向策略高效实现的基因组修饰。CRISPR/Cas系统已提供一种用于基因组编辑的新工具,但仍然存在困难。举例来说,当试图特别是在真核细胞和生物体中产生大型靶向基因组缺失或其他大型靶向遗传修饰时,困难可仍然在一些情形下出现。

[0005] 此外,可难以在无后续交配步骤下高效产生就靶向遗传修饰来说是纯合的细胞或动物,并且一些基因座相比于其他基因座可更加难以进行靶向以产生纯合性靶向修饰。举例来说,尽管就某一大型靶向基因组缺失来说是杂合的F0代小鼠可有时通过常规靶向策略来获得,但产生就所述缺失来说是纯合的F1代小鼠需要这些杂合性小鼠的后续交配。这些额外交配步骤是昂贵的和耗时的。

### 发明内容

[0006] 提供用于制备对目标外来抗原的耐受性降低的非人动物以及使用所述动物来产生结合所述目标外来抗原的抗原结合蛋白的方法和组合物。在一个方面,本发明提供一种制备对目标外来抗原的耐受性降低的非人动物的方法,其包括:(a)使不是单细胞期胚胎的非人动物多能细胞的基因组与以下各物接触:(i) Cas9蛋白;(ii) 杂交至第一靶基因组基因座内的第一引导RNA识别序列的第一引导RNA,其中所述第一靶基因组基因座影响与所述目标外来抗原同源或共有目标表位的第一自身抗原的表达;和(iii) 杂交至所述第一靶基因组基因座内的第二引导RNA识别序列的第二引导RNA;其中一对第一染色体和第二染色体中的所述第一靶基因组基因座被修饰以产生具有双等位基因修饰的经修饰非人动物多能细胞,其中所述第一自身抗原的表达降低;(b) 将所述经修饰非人动物多能细胞引入宿主胚胎中;和(c) 将所述宿主胚胎植入代孕母体中以产生经遗传修饰的F0代非人动物,其中所述一对第一染色体和第二染色体中的所述第一靶基因组基因座被修饰以使所述第一自身抗原的表达降低。任选地,多能细胞是胚胎干(ES)细胞。任选地,接触包括通过核转染来将Cas9蛋白、第一引导RNA和第二引导RNA引入非人动物多能细胞中。任选地,将Cas9蛋白以编码

Cas9蛋白的DNA形式引入非人动物多能细胞中,将第一引导RNA以编码第一引导RNA的DNA形式引入非人动物多能细胞中,并且将第二引导RNA以编码第二引导RNA的DNA形式引入非人动物多能细胞中。

[0007] 在一些所述方法中,接触步骤(a)进一步包括使基因组与:(iv)杂交至第一靶基因组基因座内的第三引导RNA识别序列的第三引导RNA;和/或(v)杂交至第一靶基因组基因座内的第四引导RNA识别序列的第四引导RNA接触。在一些所述方法中,接触步骤(a)进一步包括使基因组与以下各物接触:(iv)杂交至第二靶基因组基因座内的第三引导RNA识别序列的第三引导RNA,其中所述第二靶基因组基因座影响第一自身抗原或与目标外来抗原同源或共有目标表位的第二自身抗原的表达;和/或(v)杂交至所述第二靶基因组基因座内的第四引导RNA识别序列的第四引导RNA。

[0008] 在一些所述方法中,接触步骤(a)进一步包括使基因组与包含会在靶基因组基因座处杂交于5'靶标序列的5'同源臂和会在靶基因组基因座处杂交于3'靶标序列的3'同源臂的外源性修复模板接触。任选地,外源性修复模板进一步包含由5'同源臂和3'同源臂侧接的核酸插入物。在一些所述方法中,核酸插入物与第一靶基因组基因座同源或直系同源。在一些所述方法中,外源性修复模板的长度在约50个核苷酸至约1kb之间。在一些所述方法中,外源性修复模板的长度在约80个核苷酸至约200个核苷酸之间。在一些所述方法中,外源性修复模板是单链寡脱氧核苷酸。在一些所述方法中,外源性修复模板是长度是至少10kb的大型靶向载体(LTVEC),和/或外源性修复模板是LTVEC,其中LTVEC的5'同源臂和3'同源臂的总和的长度是至少10kb。

[0009] 一些所述方法进一步包括:(d)用目标外来抗原使在步骤(c)中产生的经遗传修饰的F0代非人动物免疫;(e)在足以引发对目标外来抗原的免疫应答的条件下维持经遗传修饰的F0代非人动物;和(f)从经遗传修饰的F0代非人动物获得编码人免疫球蛋白重链可变结构域的第一核酸序列和/或编码人免疫球蛋白轻链可变结构域的第二核酸序列。

[0010] 在一些所述方法中,相比于在使在第一靶基因组基因座处是野生型的对照非人动物免疫之后获得的抗原结合蛋白,在用目标外来抗原使经遗传修饰的F0代非人动物免疫之后获得的针对目标外来抗原的抗原结合蛋白具有更高效价。在一些所述方法中,相较于在使在第一靶基因组基因座处是野生型的对照非人动物免疫之后获得的抗原结合蛋白,在用目标外来抗原使经遗传修饰的F0代非人动物免疫之后获得更具多样性的针对目标外来抗原的抗原结合蛋白谱系。

[0011] 在一些所述方法中,第一自身抗原的表达被消除。

[0012] 在一些所述方法中,目标外来抗原是第一自身抗原的直系同源物。在一些所述方法中,目标外来抗原包含以下、基本上由以下组成或由以下组成:人蛋白质的全部或一部分。

[0013] 在一些所述方法中,第一靶基因组基因座被修饰以包含一个或多个核苷酸的插入、一个或多个核苷酸的缺失或一个或多个核苷酸的替换。在一些所述方法中,第一靶基因组基因座被修饰以包含一个或多个核苷酸的缺失。在一些所述方法中,接触步骤(a)包括使基因组与包含会在靶基因组基因座处杂交于5'靶标序列的5'同源臂和会在靶基因组基因座处杂交于3'靶标序列的3'同源臂的外源性修复模板接触,前提是如果基因组在单细胞期胚胎中,那么所述外源性修复模板的长度是不超过5kb,其中所述外源性修复模板包含由所

述5'同源臂和所述3'同源臂侧接的核酸插入物,其中所述核酸插入物与经缺失核酸序列同源或直系同源,并且其中所述核酸插入物替换经缺失核酸序列。在一些所述方法中,缺失是无随机插入和缺失(插入缺失)的精确缺失。在一些所述方法中,接触步骤(a)包括使基因组与包含会在靶基因组基因座处杂交于5'靶标序列的5'同源臂和会在靶基因组基因座处杂交于3'靶标序列的3'同源臂的外源性修复模板接触,前提是如果基因组在单细胞期胚胎中,那么所述外源性修复模板的长度是不超过5kb,其中经缺失核酸序列由在5'靶标序列与3'靶标序列之间的核酸序列组成。

[0014] 在一些所述方法中,第一靶基因组基因座包含以下、基本上由以下组成或由以下组成:编码第一自身抗原的基因的全部或一部分。在一些所述方法中,修饰包括以下、基本上由以下组成或由以下组成:编码第一自身抗原的基因的全部或一部分的纯合性缺失。在一些所述方法中,修饰包括以下、基本上由以下组成或由以下组成:对编码第一自身抗原的基因的起始密码子的纯合性破坏。

[0015] 在一些所述方法中,第一引导RNA识别序列包含编码第一自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且第二引导RNA识别序列包含编码第一自身抗原的基因的终止密码子,或在终止密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。任选地,第一引导RNA识别序列包含起始密码子,并且第二引导RNA识别序列包含终止密码子。在一些所述方法中,第一引导RNA识别序列包含第一Cas9裂解位点,并且第二引导RNA识别序列包含第二Cas9裂解位点,其中第一靶基因组基因座被修饰以包含所述第一Cas9裂解位点与所述第二Cas9裂解位点之间的缺失。任选地,缺失是精确缺失,其中经缺失核酸序列由第一Cas9裂解位点与第二Cas9裂解位点之间的核酸序列组成。

[0016] 在一些所述方法中,第一引导RNA识别序列和第二引导RNA识别序列不同,并且第一引导RNA识别序列和第二引导RNA识别序列中的每一个包含编码第一自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。任选地,第一引导RNA识别序列和第二引导RNA识别序列中的每一个包含起始密码子。

[0017] 在一些所述方法中,第一核酸序列和/或第二核酸序列从经遗传修饰的非人动物的淋巴细胞或从由所述淋巴细胞产生的杂交瘤获得。

[0018] 在一些所述方法中,非人动物包含人源化免疫球蛋白基因座。在一些所述方法中,非人动物是啮齿动物。在一些所述方法中,啮齿动物是小鼠。任选地,小鼠品系包括BALB/c品系。任选地,小鼠品系包括BALB/c、C57BL/6和129品系。任选地,小鼠品系是50%BALB/c、25%C57BL/6和25%129。任选地,小鼠的MHC单倍型是MHC<sup>b/d</sup>。

[0019] 在一些所述方法中,小鼠在它的种系中包含插入在内源性小鼠免疫球蛋白基因座处的人未重排可变区基因区段。任选地,人未重排可变区基因区段是重链基因区段,并且小鼠免疫球蛋白基因座是重链基因座。任选地,人未重排可变区基因区段是轻链区段,并且小鼠免疫球蛋白基因座是轻链基因座。任选地,轻链基因区段是人κ或λ轻链基因区段。在一些所述方法中,小鼠在它的种系中包含可操作地连接至小鼠恒定区基因的人未重排可变区基因区段,其中小鼠缺乏人恒定区基因,并且其中所述小鼠恒定区基因在内源性小鼠免疫球蛋白基因座处。在一些所述方法中,小鼠包含:(a)杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小

鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(b)杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(a)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(b)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。在一些所述方法中,小鼠包含对免疫球蛋白重链基因座的修饰,其中所述修饰使内源性ADAM6功能降低或消除,并且其中小鼠包含编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于人重链可变区基因座处。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于除人重链可变区基因座以外的位置处。

[0020] 在一些所述方法中,小鼠在它的种系中包含可操作地连接至轻链恒定区的人源化免疫球蛋白轻链可变基因座,其包含至多一个或至多两个重排人轻链V/J序列。任选地,轻链恒定区基因是小鼠基因。在一些所述方法中,小鼠进一步包含可操作地连接至重链恒定区基因的人源化免疫球蛋白重链可变基因座,其包含至少一个未重排人V、至少一个未重排人D和至少一个未重排人J区段。任选地,重链恒定区基因是小鼠基因。在一些所述方法中,小鼠包含人源化重链免疫球蛋白可变基因座和人源化轻链免疫球蛋白可变基因座,其中小鼠表达单一轻链。在一些所述方法中,小鼠包含:(a)编码免疫球蛋白轻链的人 $V_L$ 结构域的单重排人免疫球蛋白轻链可变区( $V_L/J_L$ ),其中所述单重排人 $V_L/J_L$ 选自人 $V_{\kappa}1-39/J$ 基因区段或人 $V_{\kappa}3-20/J$ 基因区段;和(b)内源性重链可变( $V_H$ )基因区段被一个或多个人 $V_H$ 基因区段的替换,其中所述人 $V_H$ 基因区段可操作地连接至内源性重链恒定( $C_H$ )区基因,并且所述人 $V_H$ 基因区段能够重排并形成人/小鼠嵌合重链基因。在一些所述方法中,小鼠表达抗体的群体,并且小鼠的种系仅包括是重排人种系 $\kappa$ 轻链可变区基因的单一免疫球蛋白 $\kappa$ 轻链可变区基因,其中小鼠就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是杂合的,因为它仅含有一个拷贝,或就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是纯合的,因为它含有两个拷贝,小鼠的特征在于活性亲和力成熟,以使:(i)所述群体的各免疫球蛋白 $\kappa$ 轻链都包含由所述重排人种系 $\kappa$ 轻链可变区基因或由其体细胞突变变体编码的轻链可变结构域;(ii)所述群体包括包含轻链可变结构域由所述重排人种系 $\kappa$ 轻链可变区基因编码的免疫球蛋白 $\kappa$ 轻链的抗体和包含轻链可变结构域由所述其体细胞突变变体编码的免疫球蛋白 $\kappa$ 轻链的抗体;以及(iii)小鼠产生与所述免疫球蛋白 $\kappa$ 轻链成功配对以形成所述群体的所述抗体的体细胞突变的高亲和力重链的多样性集合。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a)重排 $V_{\kappa}/J_{\kappa}$ 序列在内源性小鼠 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排 $V_{\kappa}/J_{\kappa}$ 序列包含:(i)单一人种系 $V_{\kappa}$ 序列,所述单一人种系 $V_{\kappa}$ 序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii)单一人种系 $J_{\kappa}$ 序列,其中所述重排 $V_{\kappa}/J_{\kappa}$ 序列可操作地连接至内源性小鼠 $\kappa$ 恒定区;和(b)多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。在一些所述方法中,小鼠包含对免疫球蛋

白重链基因座的修饰,其中所述修饰使内源性ADAM6功能降低或消除,并且其中小鼠包含编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于人重链可变区基因座处。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于除人重链可变区基因座以外的位置处。

[0021] 在一些所述方法中,小鼠具有包含对免疫球蛋白重链基因座的修饰的基因组,其中所述修饰使内源性ADAM6功能降低或消除,并且小鼠进一步包含编码非人动物ADAM6蛋白或其直系同源物或同源物或相应ADAM6蛋白的功能性片段的核酸序列。任选地,小鼠的基因组包含:(a) ADAM6基因的异位放置;和(b)人免疫球蛋白重链可变区基因座,其包含一个或多个人 $V_H$ 基因区段、一个或多个人 $D_H$ 基因区段和一个或多个人 $J_H$ 基因区段向内源性非人动物重链基因座中的插入,其中所述人 $V_H$ 、 $D_H$ 和 $J_H$ 基因区段可操作地连接至重链恒定区基因;以使小鼠的特征在于:(i)它是能生育的;以及(ii)当它用抗原免疫时,它产生包含可操作地连接至由所述重链恒定区基因编码的重链恒定结构域的由所述一个或多个人 $V_H$ 、一个或多个人 $D_H$ 和一个或多个人 $J_H$ 基因区段编码的重链可变结构域的抗体,其中所述抗体显示与所述抗原的特异性结合。

[0022] 在一些所述方法中,非人动物是至少部分地源于BALB/c品系的小鼠,其中所述小鼠包含人源化免疫球蛋白基因座,其中目标外来抗原是与第一自身抗原直系同源的人蛋白质的全部或一部分,并且第一靶基因组基因座包含编码所述第一自身抗原的基因的全部或一部分,其中第一引导RNA识别位点包含编码所述第一自身抗原的所述基因的起始密码子,并且第二引导RNA识别位点包含编码所述第一自身抗原的所述基因的终止密码子,并且其中修饰包括编码所述第一自身抗原的所述基因的全部或一部分的纯合性缺失,由此所述第一自身抗原的表达被消除。任选地,小鼠包含:(a)编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b)杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(c)杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(b)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(c)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a)编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b)重排 $V_K/J_K$ 序列在内源性小鼠 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排 $V_K/J_K$ 序列包含:(i)单一人种系 $V_K$ 序列,所述单一人种系 $V_K$ 序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii)单一人种系 $J_K$ 序列,其中所述重排 $V_K/J_K$ 序列可操作地连接至内源性小鼠 $\kappa$ 恒定区;和(c)多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球

蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

[0023] 在一些所述方法中,其中非人动物是至少部分地源于BALB/c品系的小鼠,其中所述小鼠包含人源化免疫球蛋白基因座,其中目标外来抗原是与第一自身抗原直系同源的人蛋白质的全部或一部分,并且第一靶基因组基因座包含编码所述第一自身抗原的基因的全部或一部分,其中第一引导RNA识别位点包含编码所述第一自身抗原的所述基因的起始密码子,并且第二引导RNA识别位点包含编码所述第一自身抗原的所述基因的终止密码子,并且其中修饰包括对编码所述第一自身抗原的所述基因的所述起始密码子的纯合性破坏,由此所述第一自身抗原的表达被消除。任选地,小鼠包含:(a) 编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b) 杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(c) 杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(b) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(c) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a) 编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b) 重排V<sub>κ</sub>/J<sub>κ</sub>序列在内源性小鼠κ免疫球蛋白轻链可变区基因座处的插入,所述重排V<sub>κ</sub>/J<sub>κ</sub>序列包含:(i) 单一人种系V<sub>κ</sub>序列,所述单一人种系V<sub>κ</sub>序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii) 单一人种系J<sub>κ</sub>序列,其中所述重排V<sub>κ</sub>/J<sub>κ</sub>序列可操作地连接至内源性小鼠κ恒定区;和(c) 多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

[0024] 在一些方法中,非人动物多能细胞是杂交细胞,并且方法进一步包括:(a') 比较一对同源染色体中相应第一染色体和第二染色体在第一靶基因组基因座内的序列,以及在接触步骤(a) 之前基于以下来选择第一靶基因组基因座内的靶区域:在所述一对同源染色体中所述相应第一染色体与第二染色体之间,相对于第一靶基因组基因座的其余部分的全部或一部分,所述靶区域具有更高序列同一性百分比。任选地,在一对同源染色体中相应第一染色体与第二染色体之间,相对于第一靶基因组基因座的其余部分,靶区域具有更高序列同一性百分比。任选地,在相应第一染色体与第二染色体之间,靶区域具有至少99.9%序列同一性,并且在相应第一染色体与第二染色体之间,第一靶基因组基因座的其余部分具有不超过99.8%序列同一性。任选地,在一对同源染色体中相应第一染色体和第二染色体中,靶区域相同。任选地,靶区域在第一靶基因组基因座内具有连续等位基因序列同一性的最长可能链段内。

[0025] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:第一引导RNA识别序列和在第一引导RNA识别序列的5' 侧、3' 侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、

2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,以及第二引导RNA识别序列和在第二引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:不存在于基因组中其他地方的不同引导RNA识别序列和在所述不同引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,以及将相对于其他区段具有最高序列同一性百分比的两个区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中而非存在于基因组中其他地方的各不同引导RNA识别序列相对应的区段。

[0026] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的区域。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的区域,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他区段具有最高序列同一性百分比的区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中各对不同引导RNA识别序列相对应的区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0027] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的区域和在第一引导RNA识别序列与第二引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的区域和在所述一对不同引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他区段具有最高序列同一性百分比的区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中各对不同引导RNA识别序列相对应的区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0028] 在一些所述方法中,其中靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至

少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个非连续区段,其中各非连续区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他非连续区段具有最高序列同一性百分比的非连续区段选作靶区域。任选地,一个或多个非连续区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中各对不同引导RNA识别序列相对应的非连续区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0029] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的基因组区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个非连续区段,其中各非连续区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的基因组区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他非连续区段具有最高序列同一性百分比的非连续区段选作靶区域。任选地,一个或多个非连续区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中各对不同引导RNA识别序列相对应的非连续区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0030] 在一些所述方法中,步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成:由5'靶标序列和3'靶标序列侧接的区域。在一些所述方法中,步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成:由5'靶标序列和3'靶标序列侧接且包括所述5'靶标序列和所述3'靶标序列的区域。在一些所述方法中,步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成:5'靶标序列和/或3'靶标序列。任选地,步骤(a')中的靶基因组基因座包含以下、基本上由以下组成或由以下组成:5'靶标序列和3'靶标序列。在一些所述方法中,步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成:在5'靶标序列与3'靶标序列之间的区域和在5'靶标序列与3'靶标序列之间的区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。在一些所述方法中,步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成:在5'靶标

序列与3'靶标序列之间的区域和在5'靶标序列与3'靶标序列之间的区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。在一些所述方法中,步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成:在5'靶标序列与3'靶标序列之间的区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。在一些所述方法中,步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成:在5'靶标序列与3'靶标序列之间的区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。

[0031] 在另一方面,本发明提供一种制备对目标外来抗原的耐受性降低的非人动物的方法,其包括:(a)使非人动物单细胞期胚胎的基因组与以下各物接触:(i) Cas9蛋白;(ii) 杂交至第一靶基因组基因座内的第一引导RNA识别序列的第一引导RNA,其中所述第一靶基因组基因座影响与所述目标外来抗原同源或共有目标表位的第一自身抗原的表达;和(iii) 杂交至所述第一靶基因组基因座内的第二引导RNA识别序列的第二引导RNA;其中一对第一染色体和第二染色体中的所述第一靶基因组基因座被修饰以产生双等位基因修饰,其中经修饰非人动物单细胞期胚胎中所述第一自身抗原的表达降低;和(b)将所述经修饰非人动物单细胞期胚胎植入代孕母体中以产生经遗传修饰的F0代非人动物,其中所述一对第一染色体和第二染色体中的所述第一靶基因组基因座被修饰以使所述第一自身抗原的表达降低。任选地,接触包括通过核转染来将Cas9蛋白、第一引导RNA和第二引导RNA引入非人动物单细胞期胚胎中。任选地,将Cas9蛋白以编码Cas9蛋白的DNA形式引入非人动物单细胞期胚胎中,将第一引导RNA以编码第一引导RNA的DNA形式引入非人动物单细胞期胚胎中,并且将第二引导RNA以编码第二引导RNA的DNA形式引入非人动物单细胞期胚胎中。

[0032] 在一些所述方法中,接触步骤(a)进一步包括使基因组与:(iv) 杂交至第一靶基因组基因座内的第三引导RNA识别序列的第三引导RNA;和/或(v) 杂交至第一靶基因组基因座内的第四引导RNA识别序列的第四引导RNA接触。在一些所述方法中,接触步骤(a)进一步包括使基因组与以下各物接触:(iv) 杂交至第二靶基因组基因座内的第三引导RNA识别序列的第三引导RNA,其中所述第二靶基因组基因座影响第一自身抗原或与目标外来抗原同源或共有目标表位的第二自身抗原的表达;和/或(v) 杂交至所述第二靶基因组基因座内的第四引导RNA识别序列的第四引导RNA。

[0033] 在一些所述方法中,接触步骤(a)进一步包括使基因组与包含会在靶基因组基因座处杂交于5'靶标序列的5'同源臂和会在靶基因组基因座处杂交于3'靶标序列的3'同源臂的外源性修复模板接触,其中所述外源性修复模板的长度在约50个核苷酸至约5kb之间。任选地,外源性修复模板进一步包含由5'同源臂和3'同源臂侧接的核酸插入物。在一些所述方法中,核酸插入物与第一靶基因组基因座同源或直系同源。在一些所述方法中,外源性修复模板的长度在约50个核苷酸至约1kb之间。在一些所述方法中,外源性修复模板的长度

在约80个核苷酸至约200个核苷酸之间。在一些所述方法中,外源性修复模板是单链寡脱氧核苷酸。

[0034] 一些所述方法进一步包括:(c)用目标外来抗原使在步骤(b)中产生的经遗传修饰的F0代非人动物免疫;(d)在足以引发对目标外来抗原的免疫应答的条件下维持经遗传修饰的F0代非人动物;和(e)从经遗传修饰的F0代非人动物获得编码人免疫球蛋白重链可变结构域的第一核酸序列和/或编码人免疫球蛋白轻链可变结构域的第二核酸序列。

[0035] 在一些所述方法中,相比于在使在第一靶基因组基因座处是野生型的对照非人动物免疫之后获得的抗原结合蛋白,在用目标外来抗原使经遗传修饰的F0代非人动物免疫之后获得的针对目标外来抗原的抗原结合蛋白具有更高效价。在一些所述方法中,相较于在使在第一靶基因组基因座处是野生型的对照非人动物免疫之后获得的抗原结合蛋白,在用目标外来抗原使经遗传修饰的F0代非人动物免疫之后获得更具多样性的针对目标外来抗原的抗原结合蛋白谱系。

[0036] 在一些所述方法中,第一自身抗原的表达被消除。

[0037] 在一些所述方法中,目标外来抗原是第一自身抗原的直系同源物。在一些所述方法中,目标外来抗原包含以下、基本上由以下组成或由以下组成:人蛋白质的全部或一部分。

[0038] 在一些所述方法中,第一靶基因组基因座被修饰以包含一个或多个核苷酸的插入、一个或多个核苷酸的缺失或一个或多个核苷酸的替换。在一些所述方法中,第一靶基因组基因座被修饰以包含一个或多个核苷酸的缺失。在一些所述方法中,接触步骤(a)包括使基因组与包含会在靶基因组基因座处杂交于5'靶标序列的5'同源臂和会在靶基因组基因座处杂交于3'靶标序列的3'同源臂的外源性修复模板接触,前提是如果基因组在单细胞期胚胎中,那么所述外源性修复模板的长度是不超过5kb,其中所述外源性修复模板包含由所述5'同源臂和所述3'同源臂侧接的核酸插入物,其中所述核酸插入物与经缺失核酸序列同源或直系同源,并且其中所述核酸插入物替换经缺失核酸序列。在一些所述方法中,缺失是无随机插入和缺失(插入缺失)的精确缺失。在一些所述方法中,接触步骤(a)包括使基因组与包含会在靶基因组基因座处杂交于5'靶标序列的5'同源臂和会在靶基因组基因座处杂交于3'靶标序列的3'同源臂的外源性修复模板接触,前提是如果基因组在单细胞期胚胎中,那么所述外源性修复模板的长度是不超过5kb,其中经缺失核酸序列由在5'靶标序列与3'靶标序列之间的核酸序列组成。

[0039] 在一些所述方法中,第一靶基因组基因座包含以下、基本上由以下组成或由以下组成:编码第一自身抗原的基因的全部或一部分。在一些所述方法中,修饰包括以下、基本上由以下组成或由以下组成:编码第一自身抗原的基因的全部或一部分的纯合性缺失。在一些所述方法中,修饰包括以下、基本上由以下组成或由以下组成:对编码第一自身抗原的基因的起始密码子的纯合性破坏。

[0040] 在一些所述方法中,第一引导RNA识别序列包含编码第一自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且第二引导RNA识别序列包含编码第一自身抗原的基因的终止密码子,或在终止密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。任选地,第一引导RNA识别序列包含起始密码子,并且第二引导RNA识别序列包含终止密码子。在一些所述方法中,

第一引导RNA识别序列包含第一Cas9裂解位点,并且第二引导RNA识别序列包含第二Cas9裂解位点,其中第一靶基因组基因座被修饰以包含所述第一Cas9裂解位点与所述第二Cas9裂解位点之间的缺失。任选地,缺失是精确缺失,其中经缺失核酸序列由第一Cas9裂解位点与第二Cas9裂解位点之间的核酸序列组成。

[0041] 在一些所述方法中,第一引导RNA识别序列和第二引导RNA识别序列不同,并且第一引导RNA识别序列和第二引导RNA识别序列中的每一个包含编码第一自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。任选地,第一引导RNA识别序列和第二引导RNA识别序列中的每一个包含起始密码子。

[0042] 在一些所述方法中,第一核酸序列和/或第二核酸序列从经遗传修饰的非人动物的淋巴细胞或从由所述淋巴细胞产生的杂交瘤获得。

[0043] 在一些所述方法中,非人动物包含人源化免疫球蛋白基因座。在一些所述方法中,非人动物是啮齿动物。在一些所述方法中,啮齿动物是小鼠。任选地,小鼠品系包括BALB/c品系。任选地,小鼠品系包括BALB/c、C57BL/6和129品系。任选地,小鼠品系是50%BALB/c、25%C57BL/6和25%129。任选地,小鼠的MHC单倍型是MHC<sup>b/d</sup>。

[0044] 在一些所述方法中,小鼠在它的种系中包含插入在内源性小鼠免疫球蛋白基因座处的人未重排可变区基因区段。任选地,人未重排可变区基因区段是重链基因区段,并且小鼠免疫球蛋白基因座是重链基因座。任选地,人未重排可变区基因区段是轻链区段,并且小鼠免疫球蛋白基因座是轻链基因座。任选地,轻链基因区段是人 $\kappa$ 或 $\lambda$ 轻链基因区段。在一些所述方法中,小鼠在它的种系中包含可操作地连接至小鼠恒定区基因的人未重排可变区基因区段,其中小鼠缺乏人恒定区基因,并且其中所述小鼠恒定区基因在内源性小鼠免疫球蛋白基因座处。在一些所述方法中,小鼠包含:(a) 杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(b) 杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(a) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(b) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。在一些所述方法中,小鼠包含对免疫球蛋白重链基因座的修饰,其中所述修饰使内源性ADAM6功能降低或消除,并且其中小鼠包含编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于人重链可变区基因座处。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于除人重链可变区基因座以外的位置处。

[0045] 在一些所述方法中,小鼠在它的种系中包含可操作地连接至轻链恒定区的人源化免疫球蛋白轻链可变基因座,其包含至多一个或至多两个重排人轻链V/J序列。任选地,轻链恒定区基因是小鼠基因。在一些所述方法中,小鼠进一步包含可操作地连接至重链恒定区基因的人源化免疫球蛋白重链可变基因座,其包含至少一个未重排人V、至少一个未重排人D和至少一个未重排人J区段。任选地,重链恒定区基因是小鼠基因。在一些所述方法中,

小鼠包含人源化重链免疫球蛋白可变基因座和人源化轻链免疫球蛋白可变基因座,其中小鼠表达单一轻链。在一些所述方法中,小鼠包含:(a) 编码免疫球蛋白轻链的人 $V_L$ 结构域的单一重排人免疫球蛋白轻链可变区( $V_L/J_L$ ),其中所述单一重排人 $V_L/J_L$ 区选自人 $V_{\kappa}1-39/J$ 基因区段或人 $V_{\kappa}3-20/J$ 基因区段;和(b) 内源性重链可变( $V_H$ )基因区段被一个或多个个人 $V_H$ 基因区段的替换,其中所述人 $V_H$ 基因区段可操作地连接至内源性重链恒定( $C_H$ )区基因,并且所述人 $V_H$ 基因区段能够重排并形成人/小鼠嵌合重链基因。在一些所述方法中,小鼠表达抗体的群体,并且小鼠的种系仅包括是重排人种系 $\kappa$ 轻链可变区基因的单一免疫球蛋白 $\kappa$ 轻链可变区基因,其中小鼠就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是杂合的,因为它仅含有一个拷贝,或就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是纯合的,因为它含有两个拷贝,小鼠的特征在于活性亲和力成熟,以使:(i) 所述群体的各免疫球蛋白 $\kappa$ 轻链都包含由所述重排人种系 $\kappa$ 轻链可变区基因或由其体细胞突变变体编码的轻链可变结构域;(ii) 所述群体包括包含轻链可变结构域由所述重排人种系 $\kappa$ 轻链可变区基因编码的免疫球蛋白 $\kappa$ 轻链的抗体和包含轻链可变结构域由所述其体细胞突变变体编码的免疫球蛋白 $\kappa$ 轻链的抗体;以及(iii) 小鼠产生与所述免疫球蛋白 $\kappa$ 轻链成功配对以形成所述群体的所述抗体的体细胞突变的高亲和力重链的多样性集合。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a) 重排 $V_{\kappa}/J_{\kappa}$ 序列在内源性小鼠 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排 $V_{\kappa}/J_{\kappa}$ 序列包含:(i) 单一人种系 $V_{\kappa}$ 序列,所述单一人种系 $V_{\kappa}$ 序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii) 单一人种系 $J_{\kappa}$ 序列,其中所述重排 $V_{\kappa}/J_{\kappa}$ 序列可操作地连接至内源性小鼠 $\kappa$ 恒定区;和(b) 多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。在一些所述方法中,小鼠包含对免疫球蛋白重链基因座的修饰,其中所述修饰使内源性ADAM6功能降低或消除,并且其中小鼠包含编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于人重链可变区基因座处。任选地,编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列存在于除人重链可变区基因座以外的位置处。

[0046] 在一些所述方法中,小鼠具有包含对免疫球蛋白重链基因座的修饰的基因组,其中所述修饰使内源性ADAM6功能降低或消除,并且小鼠进一步包含编码非人动物ADAM6蛋白或其直系同源物或同源物或相应ADAM6蛋白的功能性片段的核酸序列。任选地,小鼠的基因组包含:(a) ADAM6基因的异位放置;和(b) 人免疫球蛋白重链可变区基因座,其包含一个或多个个人 $V_H$ 基因区段、一个或多个个人 $D_H$ 基因区段和一个或多个个人 $J_H$ 基因区段向内源性非人动物重链基因座中的插入,其中所述人 $V_H$ 、 $D_H$ 和 $J_H$ 基因区段可操作地连接至重链恒定区基因;以使小鼠的特征在于:(i) 它是能生育的;以及(ii) 当它用抗原免疫时,它产生包含可操作地连接至由所述重链恒定区基因编码的重链恒定结构域的由所述一个或多个个人 $V_H$ 、一个或多个个人 $D_H$ 和一个或多个个人 $J_H$ 基因区段编码的重链可变结构域的抗体,其中所述抗体显示与所述抗原的特异性结合。

[0047] 在一些所述方法中,非人动物是至少部分地源于BALB/c品系的小鼠,其中所述小

鼠包含人源化免疫球蛋白基因座,其中目标外来抗原是与第一自身抗原直系同源的人蛋白质的全部或一部分,并且第一靶基因组基因座包含编码所述第一自身抗原的基因的全部或一部分,其中第一引导RNA识别位点包含编码所述第一自身抗原的所述基因的起始密码子,并且第二引导RNA识别位点包含编码所述第一自身抗原的所述基因的终止密码子,并且其中修饰包括编码所述第一自身抗原的所述基因的全部或一部分的纯合性缺失,由此所述第一自身抗原的表达被消除。任选地,小鼠包含:(a) 编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b) 杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(c) 杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(b) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(c) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a) 编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b) 重排V<sub>k</sub>/J<sub>k</sub>序列在内源性小鼠 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排V<sub>k</sub>/J<sub>k</sub>序列包含:(i) 单一人种系V<sub>k</sub>序列,所述单一人种系V<sub>k</sub>序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii) 单一人种系J<sub>k</sub>序列,其中所述重排V<sub>k</sub>/J<sub>k</sub>序列可操作地连接至内源性小鼠 $\kappa$ 恒定区;和(c) 多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

[0048] 在一些所述方法中,其中非人动物是至少部分地源于BALB/c品系的小鼠,其中所述小鼠包含人源化免疫球蛋白基因座,其中目标外来抗原是与第一自身抗原直系同源的人蛋白质的全部或一部分,并且第一靶基因组基因座包含编码所述第一自身抗原的基因的全部或一部分,其中第一引导RNA识别位点包含编码所述第一自身抗原的所述基因的起始密码子,并且第二引导RNA识别位点包含编码所述第一自身抗原的所述基因的终止密码子,并且其中修饰包括对编码所述第一自身抗原的所述基因的所述起始密码子的纯合性破坏,由此所述第一自身抗原的表达被消除。任选地,小鼠包含:(a) 编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b) 杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(c) 杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(b) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(c) 重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a) 编码小鼠ADAM6蛋

白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b)重排V<sub>κ</sub>/J<sub>κ</sub>序列在内源性小鼠 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排V<sub>κ</sub>/J<sub>κ</sub>序列包含:(i)单一人种系V<sub>κ</sub>序列,所述单一人种系V<sub>κ</sub>序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii)单一人种系J<sub>κ</sub>序列,其中所述重排V<sub>κ</sub>/J<sub>κ</sub>序列可操作地连接至内源性小鼠 $\kappa$ 恒定区;和(c)多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

[0049] 在一些方法中,非人动物单细胞期胚胎是杂交单细胞期胚胎,并且方法进一步包括:(a')比较一对同源染色体中相应第一染色体和第二染色体在第一靶基因组基因座内的序列,以及在接触步骤(a)之前基于以下来选择第一靶基因组基因座内的靶区域:在所述一对同源染色体中所述相应第一染色体与第二染色体之间,相对于第一靶基因组基因座的其余部分的全部或一部分,所述靶区域具有更高序列同一性百分比。任选地,在一对同源染色体中相应第一染色体与第二染色体之间,相对于第一靶基因组基因座的其余部分,靶区域具有更高序列同一性百分比。任选地,在相应第一染色体与第二染色体之间,靶区域具有至少99.9%序列同一性,并且在相应第一染色体与第二染色体之间,第一靶基因组基因座的其余部分具有不超过99.8%序列同一性。任选地,在一对同源染色体中相应第一染色体和第二染色体中,靶区域相同。任选地,靶区域在第一靶基因组基因座内具有连续等位基因序列同一性的最长可能链段内。

[0050] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:第一引导RNA识别序列和在第一引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,以及第二引导RNA识别序列和在第二引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:不存在于基因组中其他地方的不同引导RNA识别序列和在所述不同引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,以及将相对于其他区段具有最高序列同一性百分比的两个区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中而非存在于基因组中其他地方的各不同引导RNA识别序列相对应的区段。

[0051] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的区域。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:

在一对不同引导RNA识别序列之间的区域,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他区段具有最高序列同一性百分比的区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中各对不同引导RNA识别序列相对应的区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0052] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的区域和在第一引导RNA识别序列与第二引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的区域和在所述一对不同引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他区段具有最高序列同一性百分比的区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中各对不同引导RNA识别序列相对应的区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0053] 在一些所述方法中,其中靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地,步骤(a')包括比较第一靶基因组基因座的两个或更多个非连续区段,其中各非连续区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的基因组区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他非连续区段具有最高序列同一性百分比的非连续区段选作靶区域。任选地,一个或多个非连续区段包含以下、基本上由以下组成或由以下组成:与第一靶基因组基因座中各对不同引导RNA识别序列相对应的非连续区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0054] 在一些所述方法中,靶区域包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列之间的基因组区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、

90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。任选地，步骤(a')包括比较第一靶基因组基因座的两个或更多个非连续区段，其中各非连续区段包含以下、基本上由以下组成或由以下组成：在一对不同引导RNA识别序列之间的基因组区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列，其中所述引导RNA识别序列不存在于基因组中其他地方，以及将相对于其他非连续区段具有最高序列同一性百分比的非连续区段选作靶区域。任选地，一个或多个非连续区段包含以下、基本上由以下组成或由以下组成：与第一靶基因组基因座中各对不同引导RNA识别序列相对应的非连续区段，其中所述引导RNA识别序列不存在于基因组中其他地方。

[0055] 在一些所述方法中，步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成：由5'靶标序列和3'靶标序列侧接的区域。在一些所述方法中，步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成：由5'靶标序列和3'靶标序列侧接且包括所述5'靶标序列和所述3'靶标序列的区域。在一些所述方法中，步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成：5'靶标序列和/或3'靶标序列。任选地，步骤(a')中的靶基因组基因座包含以下、基本上由以下组成或由以下组成：5'靶标序列和3'靶标序列。在一些所述方法中，步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成：在5'靶标序列与3'靶标序列之间的区域和在5'靶标序列与3'靶标序列之间的区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。在一些所述方法中，步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成：在5'靶标序列与3'靶标序列之间的区域和在5'靶标序列与3'靶标序列之间的区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。在一些所述方法中，步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成：在5'靶标序列与3'靶标序列之间的区域的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。在一些所述方法中，步骤(a')中的靶区域包含以下、基本上由以下组成或由以下组成：在5'靶标序列与3'靶标序列之间的区域的各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。

[0056] 在另一方面，提供一种产生针对目标外来抗原的抗原结合蛋白的方法，其包括：(a) 制备对目标外来抗原的耐受性降低的经遗传修饰的非人动物，其包括：(i) 向非人动物单细胞期胚胎或不是单细胞期胚胎的非人动物多能细胞中引入：(I) Cas9蛋白；(II) 杂交至靶基因组基因座内的第一引导RNA识别序列的第一引导RNA，其中所述靶基因组基因座包含

编码与所述目标外来抗原同源或共有目标表位的自身抗原的基因的全部或一部分；和 (III) 杂交至所述靶基因组基因座内的第二引导RNA识别序列的第二引导RNA；其中一对相应第一染色体和第二染色体中的所述靶基因组基因座被修饰以产生具有双等位基因修饰的经修饰非人动物单细胞期胚胎或经修饰非人动物多能细胞，其中所述自身抗原的表达被消除；和 (ii) 从所述经修饰非人动物单细胞期胚胎或所述经修饰非人动物多能细胞产生经遗传修饰的F0代非人动物，其中所述经遗传修饰的F0代非人动物中所述一对相应第一染色体和第二染色体中的所述靶基因组基因座被修饰以使所述自身抗原的表达被消除；(b) 用所述目标外来抗原使在步骤 (a) 中产生的所述经遗传修饰的F0代非人动物免疫；和 (c) 在足以引发对所述目标外来抗原的免疫应答的条件下维持所述经遗传修饰的F0代非人动物，其中所述经遗传修饰的F0代非人动物产生针对所述目标外来抗原的抗原结合蛋白。

[0057] 在一些方法中，步骤 (a) (i) 中的细胞是非人动物多能干细胞，并且在步骤 (a) (ii) 中产生经遗传修饰的F0代非人动物包括：(I) 将经修饰非人动物多能细胞引入宿主胚胎中；和 (II) 将所述宿主胚胎植入代孕母体中以产生经遗传修饰的F0代非人动物，其中一对相应第一染色体和第二染色体中的靶基因组基因座被修饰以使自身抗原的表达被消除。任选地，多能细胞是胚胎干 (ES) 细胞。在一些方法中，步骤 (a) (i) 中的细胞是非人动物单细胞期胚胎，并且在步骤 (a) (ii) 中产生经遗传修饰的F0代非人动物包括将经修饰非人动物单细胞期胚胎植入代孕母体中以产生经遗传修饰的F0代非人动物，其中一对相应第一染色体和第二染色体中的靶基因组基因座被修饰以使自身抗原的表达被消除。

[0058] 一些所述方法进一步包括从分离自经免疫的经遗传修饰F0代非人动物的B细胞制备杂交瘤。一些所述方法进一步包括从经免疫的经遗传修饰F0代非人动物获得编码一种针对目标外来抗原的抗原结合蛋白的免疫球蛋白重链可变结构域的第一核苷酸序列和/或编码一种针对目标外来抗原的抗原结合蛋白的免疫球蛋白轻链可变结构域的第二核苷酸序列。任选地，第一核苷酸序列和/或第二核苷酸序列从经遗传修饰的F0代非人动物的淋巴细胞 (例如B细胞) 或从由所述淋巴细胞产生的杂交瘤获得。任选地，经遗传修饰的F0代非人动物包含人源化免疫球蛋白基因座，并且其中第一核苷酸序列编码人免疫球蛋白重链可变结构域，并且第二核苷酸序列编码人免疫球蛋白轻链可变结构域。

[0059] 在一些所述方法中，相比于在用目标外来抗原使对照非人动物免疫之后由在靶基因组基因座处是野生型的所述对照非人动物产生的抗原结合蛋白，由经遗传修饰的F0代非人动物产生的针对目标外来抗原的抗原结合蛋白具有更高效价。在一些所述方法中，相较于在用目标外来抗原使对照非人动物免疫之后由在靶基因组基因座处是野生型的所述对照非人动物产生的抗原结合蛋白，在用目标外来抗原使经遗传修饰的F0代非人动物免疫之后由经遗传修饰的F0代非人动物产生更具多样性的针对目标外来抗原的抗原结合蛋白谱系。在一些所述方法中，相较于在用目标外来抗原使对照非人动物免疫之后由在靶基因组基因座处是野生型的所述对照非人动物产生的抗原结合蛋白，由经遗传修饰的F0代非人动物产生的针对目标外来抗原的抗原结合蛋白使用更大多样性的重链V基因区段和/或轻链V基因区段。在一些所述方法中，由经遗传修饰的F0代非人动物产生的一些针对目标外来抗原的抗原结合蛋白与自身抗原交叉反应。

[0060] 在一些所述方法中，在靶基因组基因座中，第一引导RNA识别序列在第二引导RNA识别序列的5'，并且步骤 (a) (i) 进一步包括进行保留测定以确定在第一引导RNA识别序列

的5' 以及约1kb内的区域和/或在第二引导RNA识别序列的3' 以及约1kb内的区域的拷贝数是2。

[0061] 在一些所述方法中,目标外来抗原是自身抗原的直系同源物。在一些所述方法中,目标外来抗原包含人蛋白质的全部或一部分。

[0062] 在一些所述方法中,靶基因组基因座被修饰以包含一个或多个核苷酸的插入、一个或多个核苷酸的缺失或一个或多个核苷酸的替换。任选地,缺失是无随机插入和缺失(插入缺失)的精确缺失。

[0063] 在一些所述方法中,第一引导RNA识别序列包含编码自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且第二引导RNA识别序列包含编码自身抗原的基因的终止密码子,或在终止密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。在一些所述方法中,第一引导RNA识别序列和第二引导RNA识别序列不同,并且第一引导RNA识别序列和第二引导RNA识别序列中的每一个包含编码自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。

[0064] 在一些所述方法中,靶基因组基因座被修饰以包含在约0.1kb至约200kb之间的双等位基因缺失。在一些所述方法中,修饰包括编码自身抗原的基因的全部或一部分的双等位基因缺失。在一些所述方法中,修饰包括对编码自身抗原的基因的起始密码子的双等位基因破坏。

[0065] 在一些所述方法中,引入步骤(a)(i)进一步包括向非人动物多能细胞或非人动物单细胞期胚胎中引入:(iv)杂交至靶基因组基因座内的第三引导RNA识别序列的第三引导RNA;和/或(v)杂交至靶基因组基因座内的第四引导RNA识别序列的第四引导RNA。

[0066] 在一些所述方法中,步骤(a)(i)中的细胞是非人动物多能干细胞,并且将Cas9蛋白、第一引导RNA和第二引导RNA各自以DNA形式引入非人动物多能干细胞中。在一些所述方法中,步骤(a)(i)中的细胞是非人动物多能干细胞,并且将Cas9蛋白、第一引导RNA和第二引导RNA各自通过电穿孔或核转染来引入非人动物多能干细胞中。在一些所述方法中,步骤(a)(i)中的细胞是非人动物单细胞期胚胎,并且将Cas9蛋白、第一引导RNA和第二引导RNA各自以RNA形式引入非人动物单细胞期胚胎中。在一些所述方法中,步骤(a)(i)中的细胞是非人动物单细胞期胚胎,并且将Cas9蛋白、第一引导RNA和第二引导RNA通过原核注射或细胞质注射来引入非人动物单细胞期胚胎中。

[0067] 在一些所述方法中,不在步骤(a)(i)中引入外源性修复模板。在一些所述方法中,引入步骤(a)(i)进一步包括向非人动物多能细胞或非人动物单细胞期胚胎中引入外源性修复模板,其包含会在靶基因组基因座处杂交于5' 靶标序列的5' 同源臂和会在靶基因组基因座处杂交于3' 靶标序列的3' 同源臂,前提是如果步骤(a)(i)中的细胞是非人动物单细胞期胚胎,那么所述外源性修复模板的长度是不超过约5kb。任选地,外源性修复模板进一步包含由5' 同源臂和3' 同源臂侧接的核酸插入物。任选地,核酸插入物与靶基因组基因座同源或直系同源。任选地,外源性修复模板的长度在约50个核苷酸至约1kb之间。任选地,外源性修复模板的长度在约80个核苷酸至约200个核苷酸之间。任选地,外源性修复模板是单链寡脱氧核苷酸。任选地,步骤(a)(i)中的细胞是非人动物多能细胞,并且(a)外源性修复模板是长度是至少10kb的大型靶向载体(LTVEC);或(b)外源性修复模板是LTVEC,其中LTVEC

的5'同源臂和3'同源臂的总和的长度是至少10kb。任选地,靶基因组基因座被修饰以包含一个或多个核苷酸的缺失,并且经缺失核酸序列由在5'靶标序列与3'靶标序列之间的核酸序列组成。任选地,外源性修复模板包含由5'同源臂和3'同源臂侧接的核酸插入物,所述核酸插入物与经缺失核酸序列同源或直系同源,靶基因组基因座被修饰以包含一个或多个核苷酸的缺失,并且所述核酸插入物替换经缺失核酸序列。

[0068] 在一些所述方法中,非人动物包含人源化免疫球蛋白基因座。在一些所述方法中,非人动物是啮齿动物。任选地,啮齿动物是小鼠。任选地,小鼠品系包括BALB/c品系。任选地,小鼠品系包括BALB/c、C57BL/6和129品系。任选地,小鼠品系是50%BALB/c、25%C57BL/6和25%129。任选地,小鼠的MHC单倍型是MHC<sup>b/d</sup>。

[0069] 在一些所述方法中,小鼠在它的种系中包含插入在内源性小鼠免疫球蛋白基因座处的人未重排可变区基因区段。任选地,人未重排可变区基因区段是重链基因区段,并且小鼠免疫球蛋白基因座是重链基因座,和/或其中人未重排可变区基因区段是 $\kappa$ 或 $\lambda$ 轻链区段,并且小鼠免疫球蛋白基因座是轻链基因座。任选地,小鼠在它的种系中包含可操作地连接至小鼠恒定区基因的人未重排可变区基因区段,其中小鼠缺乏人恒定区基因,并且其中所述小鼠恒定区基因在内源性小鼠免疫球蛋白基因座处。任选地,小鼠包含:(a)杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(b)杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(a)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(b)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。

[0070] 在一些所述方法中,小鼠在它的种系中包含可操作地连接至小鼠轻链恒定区的人源化免疫球蛋白轻链可变基因座,其包含至多一个或至多两个重排人轻链V/J序列,并且其中小鼠进一步包含可操作地连接至小鼠重链恒定区基因的人源化免疫球蛋白重链可变基因座,其包含至少一个未重排人V、至少一个未重排人D和至少一个未重排人J区段。任选地,小鼠包含人源化重链免疫球蛋白可变基因座和人源化轻链免疫球蛋白可变基因座,其中小鼠表达单一轻链。任选地,小鼠包含:(a)编码免疫球蛋白轻链的人V<sub>L</sub>结构域的唯一重排人免疫球蛋白轻链可变区(V<sub>L</sub>/J<sub>L</sub>),其中所述单一重排人V<sub>L</sub>/J<sub>L</sub>区选自人V $\kappa$ 1-39/J $\kappa$ 5基因区段或人V $\kappa$ 3-20/J $\kappa$ 1基因区段;和(b)内源性重链可变(V<sub>H</sub>)基因区段被一个或多个人VH基因区段的替换,其中所述人V<sub>H</sub>基因区段可操作地连接至内源性重链恒定(C<sub>H</sub>)区基因,并且所述人V<sub>H</sub>基因区段能够重排并形成人/小鼠嵌合重链基因。任选地,小鼠表达抗体的群体,并且小鼠的种系仅包括是重排人种系 $\kappa$ 轻链可变区基因的唯一免疫球蛋白 $\kappa$ 轻链可变区基因,其中小鼠就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是杂合的,因为它仅含有一个拷贝,或就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是纯合的,因为它含有两个拷贝,小鼠的特征在于活性亲和力成熟,以使:(i)所述群体的各免疫球蛋白 $\kappa$ 轻链都包含由所述重排人种系 $\kappa$ 轻链可变区基因或由其体细胞突变变体编码的轻链可变结构域;(ii)所述群体包括包含轻链可变结构域由所述重排人种系 $\kappa$ 轻链可变区基因编码的免疫球蛋白 $\kappa$ 轻链的抗体和包含轻链可变结构域由所述其体细胞突变变体编码的免疫球蛋白 $\kappa$ 轻链的抗体;以及(iii)小鼠产

生与所述免疫球蛋白 $\kappa$ 轻链成功配对以形成所述群体的所述抗体的体细胞突变的高亲和力和重链的多样性集合。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a)重排V $\kappa$ /J $\kappa$ 序列在内源性小鼠 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排V $\kappa$ /J $\kappa$ 序列包含:(i)单一人种系V $\kappa$ 序列,所述单一人种系V $\kappa$ 序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii)单一人种系J $\kappa$ 序列,其中所述重排V $\kappa$ /J $\kappa$ 序列可操作地连接至内源性小鼠 $\kappa$ 恒定区;和(b)多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

[0071] 在一些所述方法中,小鼠包含对免疫球蛋白重链基因座的修饰,其中所述修饰使内源性ADAM6功能降低或消除,其中小鼠包含编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性,并且其中编码所述小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的所述异位核酸序列存在于人重链可变区基因座处。

[0072] 在一些所述方法中,非人动物是至少部分地源于BALB/c品系的小鼠,并且所述小鼠包含人源化免疫球蛋白基因座,其中目标外来抗原是与自身抗原直系同源的人蛋白质的全部或一部分,其中第一引导RNA识别序列包含编码自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且第二引导RNA识别序列包含编码自身抗原的基因的终止密码子,或在终止密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且其中修饰包括编码自身抗原的基因的全部或一部分的双等位基因缺失,由此自身抗原的表达被消除。在一些所述方法中,非人动物是至少部分地源于BALB/c品系的小鼠,并且所述小鼠包含人源化免疫球蛋白基因座,其中目标外来抗原是与自身抗原直系同源的人蛋白质的全部或一部分,其中第一引导RNA识别序列包含编码自身抗原的基因的起始密码子,并且第二引导RNA识别序列包含编码自身抗原的基因的终止密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且其中修饰包括对编码自身抗原的基因的起始密码子的双等位基因破坏,由此自身抗原的表达被消除。任选地,小鼠包含:(a)编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b)杂合重链基因座,其包含人免疫球蛋白重链V、D和J基因区段的插入,其中所述人重链免疫球蛋白V、D和J基因区段可操作地连接至小鼠免疫球蛋白重链基因,其中所述小鼠免疫球蛋白重链基因在内源性小鼠免疫球蛋白基因座处;和(c)杂合轻链基因座,其包含人免疫球蛋白轻链V和J基因区段的插入,其中所述人V和J基因区段可操作地连接至小鼠免疫球蛋白轻链恒定区基因序列;其中(b)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合重链序列,以及(c)重排以形成包含可操作地连接至小鼠恒定区的人可变区的杂合轻链序列,并且其中小鼠不能形成包含人可变区和人恒定区的抗体。任选地,小鼠在它的种系中就以下来说是杂合的或纯合的:(a)编码小鼠ADAM6蛋白、其直系同源物、其同源物或其片段的异位核酸序列,其中所述ADAM6蛋白、其直系同源物、其同源物或其片段在雄性小鼠中具有功能性;(b)重排V $\kappa$ /J $\kappa$ 序列在内源性小鼠 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排V $\kappa$ /J $\kappa$ 序列包含:(i)单一人种系V $\kappa$ 序列,所述单

一人种系V<sub>κ</sub>序列存在于SEQ ID NO:148或SEQ ID NO:149中;和(ii)单一人种系J<sub>κ</sub>序列,其中所述重排V<sub>κ</sub>/J<sub>κ</sub>序列可操作地连接至内源性小鼠κ恒定区;和(c)多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性小鼠免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/小鼠嵌合免疫球蛋白重链基因。

[0073] 在一些所述方法中,非人动物多能细胞是杂交细胞,或非人哺乳动物单细胞期胚胎是杂交单细胞期胚胎,并且其中方法进一步包括:(a')比较一对相应第一染色体和第二染色体在靶基因组基因座内的序列,以及在接触步骤(a)之前基于以下来选择靶基因组基因座内的靶区域:在所述一对相应第一染色体与第二染色体之间,相对于靶基因组基因座的其余部分的全部或一部分,所述靶区域具有更高序列同一性百分比,其中所述靶区域包含:第一引导RNA识别序列和在第一引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb或10kb的侧接序列,和/或第二引导RNA识别序列和在第二引导RNA识别序列的5'侧、3'侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb或10kb的侧接序列。任选地,在一对相应第一染色体与第二染色体之间,相对于靶基因组基因座的其余部分,靶区域具有更高序列同一性百分比。任选地,在一对相应第一染色体与第二染色体之间,靶区域具有至少99.9%序列同一性,并且在一对相应第一染色体与第二染色体之间,靶基因组基因座的其余部分具有不超过99.8%序列同一性。

[0074] 在另一方面,提供制备对目标外来抗原的耐受性降低的经遗传修饰的非人动物的方法,其包括:(a)向非人动物单细胞期胚胎或不是单细胞期胚胎的非人动物多能细胞中引入:(i)Cas9蛋白;(ii)杂交至靶基因组基因座内的第一引导RNA识别序列的第一引导RNA,其中所述靶基因组基因座包含编码与所述目标外来抗原同源或共有目标表位的自身抗原的基因的全部或一部分;和(iii)杂交至所述靶基因组基因座内的第二引导RNA识别序列的第二引导RNA;其中一对相应第一染色体和第二染色体中的所述靶基因组基因座被修饰以产生具有双等位基因修饰的经修饰非人动物单细胞期胚胎或经修饰非人动物多能细胞,其中所述自身抗原的表达被消除;和(b)从所述经修饰非人动物单细胞期胚胎或所述经修饰非人动物多能细胞产生经遗传修饰的F0代非人动物,其中所述经遗传修饰的F0代非人动物中所述一对相应第一染色体和第二染色体中的所述靶基因组基因座被修饰以使所述自身抗原的表达被消除。

[0075] 所述方法可包括例如产生针对目标外来抗原的抗原结合蛋白的方法的任何以上公开的变化。举例来说,在一些所述方法中,步骤(a)中的细胞是非人动物多能干细胞,并且在步骤(b)中产生经遗传修饰的F0代非人动物包括:(I)将经修饰非人动物多能细胞引入宿主胚胎中;和(II)将所述宿主胚胎植入代孕母体中以产生经遗传修饰的F0代非人动物,其中一对相应第一染色体和第二染色体中的靶基因组基因座被修饰以使自身抗原的表达被消除。任选地,多能细胞是胚胎干(ES)细胞。在一些所述方法中,步骤(a)中的细胞是非人动物单细胞期胚胎,并且在步骤(b)中产生经遗传修饰的F0代非人动物包括将经修饰非人动物单细胞期胚胎植入代孕母体中以产生经遗传修饰的F0代非人动物,其中一对相应第一染色体和第二染色体中的靶基因组基因座被修饰以使自身抗原的表达被消除。在一些所述方

法中,目标外来抗原是自身抗原的直系同源物。在一些所述方法中,第一引导RNA识别序列包含编码自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且第二引导RNA识别序列包含编码自身抗原的基因的终止密码子,或在终止密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。在一些所述方法中,第一引导RNA识别序列和第二引导RNA识别序列不同,并且第一引导RNA识别序列和第二引导RNA识别序列中的每一个包含编码自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。在一些所述方法中,在靶基因组基因座中,第一引导RNA识别序列在第二引导RNA识别序列的5',并且步骤(a)(i)进一步包括进行保留测定以确定在第一引导RNA识别序列的5'以及约1kb内的区域和/或在第二引导RNA识别序列的3'以及约1kb内的区域的拷贝数是2。在一些所述方法中,修饰包括编码自身抗原的基因的全部或一部分的双等位基因缺失。在一些所述方法中,修饰包括对编码自身抗原的基因的起始密码子的双等位基因破坏。在一些所述方法中,非人动物是小鼠。

## 附图说明

[0076] 图1显示用以破坏VELOCIMMUNE<sup>®</sup>小鼠(VI-3;在IgH与Igκ两者处均是纯合性人源化的)中的免疫耐受性的传统方法。在传统方法中,在F1H4胚胎干(ES)细胞中产生编码与目标外来靶抗原同源的自身抗原的基因的杂合性敲除(无效)等位基因。从设计靶向载体至产生就敲除来说是杂合的F0小鼠的时间是约5个月。接着使VI-3小鼠与在编码与目标外来靶抗原同源的自身抗原的内源性基因处携带杂合性敲除突变的F0小鼠交配。为产生适于免疫的三重纯合性小鼠(目标靶标的纯合性无效和在IgH与Igκ两者处的纯合性人源化),需要再进行两代交配。从设计靶向载体至产生三重纯合性小鼠的整个过程花费约15至16个月。

[0077] 图2显示用于破坏VELOCIMMUNE<sup>®</sup>(VI-3)小鼠中或通用轻链(ULC或共同轻链)小鼠中的免疫耐受性的加速方法。在这个方法中,靶向源于VI-3或ULC小鼠的ES细胞以产生编码与目标外来靶抗原同源的自身抗原的内源性基因的杂合性无效等位基因。依序靶向步骤为获得纯合性无效VI-3或ULC ES细胞克隆所需。

[0078] 图3显示用于破坏VELOCIMMUNE<sup>®</sup>(VI-3)小鼠中或通用轻链(或共同轻链)(ULC)小鼠中的耐受性的另一加速方法。在这个方法中,用CRISPR/Cas9和成对引导RNA靶向VI-3或ULC ES细胞以在单一步骤中产生编码与目标外来靶抗原同源的自身抗原的内源性基因的纯合性垮塌。TAQMAN<sup>®</sup>筛选可包括例如等位基因丧失测定与保留测定两者。

[0079] 图4显示使用大型靶向载体(LTVEC)和成对上游和下游引导RNA(gU和gD)来同时使编码与目标外来靶抗原同源的自身抗原的小鼠基因缺失和用新霉素选择标记进行替换的一般性示意图。由两个引导RNA引导的Cas9裂解位点的位置由箭号指示在小鼠基因序列以下。TAQMAN<sup>®</sup>测定探针由水平线指示,包括保留测定探针以及上游、中间和下游等位基因丧失(LOA)测定探针。图的底部部分指示预期的靶向等位基因类型。

[0080] 图5显示使用大型靶向载体(LTVEC)和各自靶向小鼠ATG起始密码子的三个重叠引导RNA来同时使编码与目标外来靶抗原同源的自身抗原的小鼠基因缺失和用floxed新霉素

选择标记和lacZ进行替换的一般性示意图。引导RNA由水平箭号指示,并且TAQMAN<sup>®</sup>测定探针由经包围水平线指示。图的底部部分指示预期的靶向等位基因类型。

[0081] 图6显示野生型通用轻链 (ULC 1-39) 小鼠中以及就编码与靶标8直系同源的自身抗原(自身抗原8)的内源性基因来说是纯合性无效的ULC 1-39小鼠中针对人靶抗原(靶标8)的抗体效价数据。

[0082] 图7显示被进行来产生杂交VGF1 (F1H4) ES细胞 (C57BL6 (X<sup>B6</sup>) /129S6 (Y<sup>129</sup>)) 的交配。

[0083] 图8显示使用LTVEC以及一个或两个5' 区域gRNA、中间区域gRNA和3' 区域gRNA来同时使小鼠基因或小鼠基因的一部分缺失和用相应人形式进行替换的示意图。LTVEC显示于图的顶部部分中,并且小鼠基因座显示于图的底部部分中。由八个引导RNA引导的Cas9裂解位点的位置由竖直箭号指示在小鼠基因序列以下。

[0084] 图9A显示使用LTVEC和两个引导RNA(引导RNA A和B)来同时使小鼠基因缺失和用相应人形式进行替换的一般性示意图。LTVEC显示于图9A的顶部部分中,并且小鼠基因座显示于图9A的底部部分中。由两个引导RNA引导的Cas9裂解位点的位置由箭号指示在小鼠基因序列以下。

[0085] 图9B-图9E显示当使用两个引导RNA时在更大频率下发生的独特双等位基因修饰(等位基因类型)。具有斜影线的粗线指示小鼠基因,虚线指示小鼠基因中的缺失,并且粗黑线指示人基因的插入。图9B显示纯合性垮塌等位基因(大型CRISPR诱导缺失)。图9C显示纯合性靶向等位基因。图9D显示半合性靶向等位基因。图9E显示复合杂合性等位基因。

[0086] 图10A和图10B显示确认所选克隆的基因型的PCR测定。图10A显示来自使用引物m-lr-f和m-5'-r针对所选ES细胞克隆进行的长程PCR测定的结果,所述测定确定人插入物与在与5' 同源臂同源的那些序列的外部的序列之间的关联,由此证明正确靶向。图10B显示来自5' Del J、5' Ins J、Del A+F和Del A+E2PCR测定的结果。5' Del J描绘使用m-5'-f和m-5-r引物获得的PCR产物,所述测定扩增围绕gRNA A裂解位点的野生型序列以确定这个序列的保留或丧失。5' Ins J描绘使用m-5'-f和h-5'-r引物获得的PCR产物,所述测定确定人插入物与小鼠基因组之间的关联。测定将在靶向克隆与随机整合克隆两者的情况下给出阳性结果。Del A+F描绘关于克隆B0-F10和AW-A8中由双重gRNA A和F裂解介导的大型缺失的预期扩增子大小(359bp)和实际条带。Del A+E2描绘用于克隆BA-A7的相同理念。NT指示无模板,+/+指示亲本VGF1杂交ES细胞野生型对照,H/+指示杂合性人源化基因型,H/ $\Delta$ 指示半合性人源化基因型,H/H指示纯合性人源化基因型,并且 $\Delta/\Delta$ 指示纯合性缺失基因型。

[0087] 图11A-图11C显示对用Lrp5人源化LTVEC与Cas9和两个gRNA组合靶向的小鼠ES细胞克隆AW-D9(图11A)和BA-D5(图11C)以及用单独LTVEC靶向的克隆BS-C4(图11B)的荧光原位杂交(FISH)分析。箭号指示染色体19的条带B上的杂交信号的位置。红色信号指示仅与小鼠探针杂交(虚线箭号,图11B)。黄色混合颜色信号指示与红色小鼠探针和绿色人探针两者杂交。对于BS-C4克隆,一个染色体19条带B具有红色信号(虚线箭号)以及另一染色体19条带B具有黄色信号(实线箭号)确认靶向正确基因座以及杂合性基因型(图11B)。对于AW-D9和BS-C4克隆,两个染色体19的B条带均具有黄色信号(实线箭号,图11A和图11C)确认靶向正确基因座以及纯合性基因型。

[0088] 图12显示在VGF1杂交ES细胞的情况下,染色体19连同被设计以通过分析杂合性丧失(LOH)来考查由两个引导RNA介导的基因转换或有丝分裂重组事件的测定一起的示意图。

**TAQMAN<sup>®</sup>** qPCR染色体拷贝数 (CCN) 探针的近似位置由箭号显示。结构变体 (SV) 多态性 PCR探针的近似位置由V形符连同上方给出的它们离Lrp5基因座的距离 (以Mb计) 一起加以显示。单核苷酸变体 (SNV) **TAQMAN<sup>®</sup>** 等位基因辨别探针的近似位置由箭头连同下方给出的它们离Lrp5基因座的距离 (以Mb计) 一起加以显示。F、E2、D、B2和A的gRNA识别序列的位置由在Lrp5基因图示的上方的斜箭号显示。

[0089] 图13A和图13B显示对用Hc人源化LTVEC与Cas9和两个gRNA组合靶向的小鼠ES细胞克隆Q-E9 (图13A) 和O-E3 (图13B) 的荧光原位杂交 (FISH) 分析。箭号指示染色体2的条带B上的杂交信号的位置。红色信号指示仅与小鼠探针杂交 (虚线箭号, 图13A)。黄色混合颜色信号指示与红色小鼠探针和绿色人探针两者杂交 (实线箭号)。对于Q-E9克隆, 一个染色体2条带B具有红色信号 (虚线箭号) 以及另一染色体2条带B具有黄色信号 (实线箭号) 确认靶向正确基因座以及杂合性基因型 (图13A)。对于O-E3克隆, 两个染色体2的B条带均具有黄色信号 (实线箭号, 图13B) 确认靶向正确基因座图以及纯合性基因型。

[0090] 图14显示在VGF1杂交ES细胞的情况下, 含有小鼠C5基因的染色体连同被设计以通过分析杂合性丧失 (LOH) 来考查由两个引导RNA介导的基因转换或有丝分裂重组事件的测定一起的示意图。结构变体 (SV) 多态性PCR探针的近似位置由水平箭号连同上方给出的它们离C5基因座的距离 (以Mb计) 一起加以显示。E2和A的gRNA识别序列的位置由在C5基因座图示的上方的斜箭号显示。

[0091] 图15A-图15E显示克隆BR-B4、BP-G7、B0-G11、B0-F10、B0-A8和BC-H9的结构变化 (SV) 测定的结果, 其中VGF1 (F1H4)、129和B6DNA用作对照。在Lrp5基因座的端粒方向在以下距离下进行测定: 13.7Mb (图15A)、20.0Mb (图15B)、36.9Mb (图15C)、48.3Mb (图15D) 和 56.7Mb (图15E)。B6和129等位基因的PCR产物的位置由箭号显示。

[0092] 图16A-图16C显示在Lrp5的着丝粒方向0.32Mb (图16A)、在Lrp5的端粒方向1.2Mb (图16B) 以及在Lrp5的端粒方向57.2Mb (图16C) 的等位基因辨别图。各轴上的值代表相对荧光强度。各图描绘各样品的四个重复测定, 其显示为实心点 (B6等位基因)、空心点 (129等位基因) 和具有斜线的点 (B6等位基因/129等位基因两者)。

[0093] 图17A-图17C是显示在细胞周期的G2期期间可产生纯合性事件的有丝分裂重组和通过杂合性丧失检测的广泛分布的基因转换的可能机理的示意图。图17A显示就129同源物上的靶向人源化来说是杂合的杂交129/B6ES细胞中的显示两个染色单体的经复制同源染色体。双头箭号指示由双重gRNA引导的Cas9裂解产生的潜在双链断裂, 所述裂解促进通过同源染色体上的染色单体之间的同源性重组达成的相互交换, 在所靶向等位基因的着丝粒侧上显示为交叉, 从而产生图17B中所示的杂交染色单体。图17C显示在有丝分裂和细胞分裂之后, 有可能有四种类型的染色体分离进入子细胞中。保留杂合性的两种, 即亲本型杂合子 (Hum/+, 左上) 和通过均等交换获得的杂合子 (Hum/+, 右上) 不能通过LOH测定来区分。其他两种显示杂合性丧失, 即丧失端粒方向B6等位基因的人源化纯合子 (Hum/Hum, 例如克隆 B0-A8, 左下) 和丧失端粒方向129等位基因的野生型纯合子 (+/+, 右下)。这个后述类型将丧失, 因为它未保留人源化等位基因的药物抗性盒。

[0094] 图18A-图18F显示解释在F1杂交小鼠ES细胞的情况下在CRISPR/Cas9辅助的人源化实验中观察到的包括杂合性丧失 (LOH) 的结果的可能机理, 所述细胞具有一个源于

129S6/SvEvTac小鼠品系的单倍体染色体组和一个源于C57BL/6NTac (B6) 小鼠品系的单倍体染色体组。图18A显示由有丝分裂交叉达成的相互染色单体交换,其中杂合性修饰在基因组复制之前或在基因组复制之后继之以姐妹染色单体之间的基因转换而发生在129染色体上。图18B显示由有丝分裂交叉达成的相互染色单体交换,其中单一129染色单体在基因组复制之后被修饰。图18C显示由有丝分裂交叉达成的相互染色单体交换,其中尚未发生LTVEC靶向,但Cas9裂解已发生在129或B6染色体上(显示的是B6裂解)。图18D显示由断裂诱导的复制达成的染色单体拷贝,其中杂合性修饰在基因组复制之前或在基因组复制之后继之以姐妹染色单体之间的基因转换而发生在129染色体上。图18E显示由断裂诱导的复制达成的染色单体拷贝,其中单一129染色单体在基因组复制之后被修饰。图18F显示由断裂诱导的复制达成的染色单体拷贝,其中尚未发生LTVEC靶向,但Cas9裂解已发生在129或B6染色体上(显示的是B6裂解)。

[0095] 图19显示在VGF1杂交ES细胞中,使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人LRP5基因座进行替换的小鼠Lrp5基因座的示意图。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0096] 图20显示在VGF1杂交ES细胞中,使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人形式进行替换的小鼠Hc基因座的示意图。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0097] 图21显示在VGF1杂交ES细胞中,使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人形式进行替换的小鼠Trpa1基因座的示意图。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0098] 图22显示在VGF1杂交ES细胞中,使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人形式进行替换的小鼠Adamts5基因座的示意图。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0099] 图23显示在VGF1杂交ES细胞中,使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人形式进行替换的小鼠Folh1基因座的示意图。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0100] 图24显示在VGF1杂交ES细胞中,使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人形式进行替换的小鼠Dpp4基因座的示意图。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0101] 图25显示在VGF1杂交ES细胞中,使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人形式进行替换的小鼠Ror1基因座的示意图。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0102] 图26显示包括编码跨膜蛋白的基因的小鼠基因座的示意图;在VGF1杂交ES细胞中,所述小鼠基因座使用LTVEC和一个或多个gRNA加以靶向以达成缺失和用相应人形式进行替换。矩形代表靶标基因组区域内的不同基因。在竖直虚线内部的区域是靶向区域(在LTVEC的5' 靶标序列和3' 靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系MP变体、来自Taconic Biosciences的C57BL/6N品系RGC变体、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。MP变体和RGC变体是来自相同品系的不同小鼠。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0103] 图27A-图27C是显示在细胞周期的G2期期间可产生纯合性事件的有丝分裂重组和通过局部杂合性丧失检测的基因转换的可能机理的示意图。图27A显示就129同源物上的靶向人源化来说是杂合的杂交129/B6ES细胞中的显示两个染色单体的经复制同源染色体。129同源物上的杂合性修饰在基因组复制之前发生,或单一129染色单体在基因组复制之后被修饰,继之以染色单体间基因转换。双头箭号指示由双重gRNA引导的Cas9裂解产生的潜在双链断裂,所述裂解促进双重链侵入和合成引导的修复,由斜虚线箭号加以显示,从而产生通过拷贝一个经修饰染色单体的一小部分的基因转换事件产生的杂交染色单体,如图27B中所示。图27C显示在有丝分裂和细胞分裂之后,有可能有两种类型的染色体分离进入子细胞中:一种保留杂合性(不具有杂合性丧失的亲本型杂合子(Hum/+,上部)),以及一种

具有围绕靶向修饰的局部杂合性丧失 (Hum/Hum, 底部, 保留129等位基因)。

[0104] 图28显示单独或与大型靶向载体组合使用靶向编码自身抗原的基因的起始密码子和终止密码子区域的成对引导RNA, 在VI-3和ULC 1-39胚胎干 (ES) 细胞中对于具有不同大小的不同自身抗原靶标进行CRISPR/Cas9介导的缺失的效率。

[0105] 图29显示在使用靶向编码自身抗原的基因的起始密码子和终止密码子区域的成对引导RNA, 用靶向用于缺失的具有不同大小的不同自身抗原靶标的CRISPR/Cas9靶向VI-3和ULC 1-39单细胞期胚胎之后产生的具有垮塌等位基因的小鼠幼仔的百分比。

[0106] 图30A和图30B显示在用靶标9全长DNA免疫之后, 在亲本VI-3T3细胞和被工程化来表达靶标9的VI-3T3细胞情况下, 野生型VI-3-Adam6小鼠中 (图30B) 以及就编码与靶标9直系同源的自身抗原 (自身抗原9) 的内源性基因来说是纯合性无效的VI3-Adam6小鼠中 (图30A) 针对人靶抗原 (靶标9) 的抗体效价数据。

[0107] 图31A和图31B显示针对人靶抗原 (靶标4) 以及针对相应直系同源小鼠自身抗原 (自身抗原4) 的抗体效价数据。图31A显示就编码自身抗原4的内源性基因来说是纯合性无效的VI3-Adam6小鼠中针对人靶标4和小鼠自身抗原4的抗体效价数据。图31B显示就编码自身抗原4的内源性基因来说是纯合性无效的ULC 1-39小鼠中针对人靶标4和小鼠自身抗原4的组的抗体效价数据。

[0108] 图32显示VI3-Adam6小鼠和ULC 1-39小鼠中的免疫球蛋白重链基因座 (顶部) 和免疫球蛋白轻链基因座 (底部) 的示意图, 所述小鼠各自具有50% BALB/cTac、25% C57BL/6NTac和25% 129S6/SvEvTac的遗传背景。在VI3-Adam6小鼠中, 内源性小鼠免疫球蛋白重链和轻链可变区用相应人DNA以及重新插入的小鼠Adam6基因 (由梯形表示的Adam6b和Adam6a) 替换。在通用轻链 (ULC 1-39) 小鼠中, 内源性小鼠免疫球蛋白重链可变区用相应人DNA以及重新插入的小鼠Adam6基因替换, 并且免疫球蛋白轻链可变区包含可操作地连接至hVκ3-15启动子的单一重排人免疫球蛋白轻链核苷酸序列 (Vκ1-39/Jκ5)。人区段以黑色描绘, 并且小鼠区段由斜线指示。

[0109] 定义

[0110] 在本文中可互换使用的术语“蛋白质”、“多肽”和“肽”包括任何长度的聚合氨基酸形式, 包括编码和非编码氨基酸以及化学或生物化学修饰或衍生的氨基酸。所述术语也包括已被修饰的聚合物, 诸如具有经修饰肽骨架的多肽。

[0111] 蛋白质被称为具有“N末端”和“C末端”。术语“N末端”涉及蛋白质或多肽的起始部分, 由具有游离胺基团 (-NH<sub>2</sub>) 的氨基酸封端。术语“C末端”涉及氨基酸链 (蛋白质或多肽) 的终止部分, 由游离羧基 (-COOH) 封端。

[0112] 在本文中可互换使用的术语“核酸”和“多核苷酸”包括任何长度的聚合核苷酸形式, 包括核糖核苷酸、脱氧核糖核苷酸或其类似物或经修饰形式。它们包括单链、双链和多链DNA或RNA、基因组DNA、cDNA、DNA-RNA杂合物、以及包含嘌呤碱基、嘧啶碱基或其他天然核苷酸碱基、化学修饰的核苷酸碱基、生物化学修饰的核苷酸碱基、非天然核苷酸碱基或衍生的核苷酸碱基的聚合物。

[0113] 核酸被称为具有“5'末端”和“3'末端”, 因为单核苷酸以使得一个单核苷酸戊糖环的5'磷酸在一个方向上通过磷酸二酯键连接至它的相邻单核苷酸戊糖环的3'氧的方式反应以产生寡核苷酸。如果寡核苷酸的末端的5'磷酸未连接至单核苷酸戊糖环的3'氧, 那么

所述末端被称为“5'末端”。如果寡核苷酸的末端的3'氧未连接至另一单核苷酸戊糖环的5'磷酸,那么所述末端被称为“3'末端”。即使在较大寡核苷酸的内部,核酸序列也可被称为具有5'末端和3'末端。在线性或环状DNA分子中,离散元件被称为在“下游”或3'元件的“上游”或5'。

[0114] 术语“野生型”包括具有如在正常(与突变、患病、改变等相对比)状态或情形下所见的结构和/或活性的实体。野生型基因和多肽常常以多种不同形式(例如等位基因)存在。

[0115] 关于蛋白质和核酸的术语“经分离”包括蛋白质和核酸关于可通常原位存在的其他细菌、病毒或细胞组分是相对纯化的,直至并包括蛋白质和多核苷酸的大致上纯净制剂。术语“经分离”也包括蛋白质和核酸不具有天然存在的对应物,已被化学合成,因此大致上未由其他蛋白质或核酸污染,或已与它们天然地与其相伴的大多数其他细胞组分(例如其他细胞蛋白质、多核苷酸或细胞组分)分离或从所述大多数其他细胞组分纯化。

[0116] “外源性”分子或序列包括通常不以那个形式存在于细胞中的分子或序列。通常存在包括关于细胞的特定发育阶段和环境条件而存在。外源性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,或可包括对应于细胞内的内源性序列,但呈不同形式(即不在染色体内)的序列。相比之下,内源性分子或序列包括通常以那个形式在特定发育阶段在特定环境条件下存在于特定细胞中的分子或序列。

[0117] “密码子优化”通常包括通过用更频繁或最频繁用于宿主细胞的基因中的密码子替换天然序列的至少一个密码子,同时维持天然氨基酸序列来修饰核酸序列以达成在特定宿主细胞中的表达增强的过程。举例来说,编码Cas9蛋白的多核苷酸可被修饰以用相较于天然存在的核酸序列,在给定原核或真核细胞(包括细菌细胞、酵母细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、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进行替代。密码子使用表例如可易于在“密码子使用数据库(Codon Usage Database)”处获得。这些表可以许多方式改适。参见Nakamura等(2000) *Nucleic Acids Research* 28:292,其出于所有目的以引用的方式整体并入本文。用于针对在特定宿主中的表达对特定序列进行密码子优化的计算机算法也是可用的(参见例如Gene Forge)。

[0118] 术语“基因座”是指基因(或有意义序列)、DNA序列、多肽编码序列或位置在生物体的基因组的染色体上的特定位置。举例来说,“Lrp5基因座”可指Lrp5基因、Lrp5DNA序列、LRP5编码序列或Lrp5位置在生物体的基因组的已被鉴定为这种序列驻留所处的染色体上的特定位置。“Lrp5基因座”可包含Lrp5基因的调控元件,包括例如增强子、启动子、5' UTR和/或3' UTR或其组合。

[0119] 术语“基因”是指染色体中的DNA序列,其编码产物(例如RNA产物和/或多肽产物),并且包括被非编码内含子中断的编码区和在5'末端与3'末端两者上邻近于编码区定位的序列,以致基因对应于全长mRNA(包括5'非翻译序列和3'非翻译序列)。术语“基因”也包括其他非编码序列,包括调控序列(例如启动子、增强子和转录因子结合位点)、多腺苷酸化信号、内部核糖体进入位点、沉默子、绝缘序列和基质附着区。这些序列可接近于基因的编码区(例如在10kb内),或在远端位点处,并且它们影响基因的转录和翻译的水平或速率。

[0120] 术语“等位基因”是指基因的变异形式。一些基因具有多种不同形式,其位于染色体上的相同位置或遗传基因座处。二倍体生物体具有两个在各遗传基因座处的等位基因。

各对等位基因代表特定遗传基因座的基因型。如果在特定基因座处存在两个相同等位基因,那么基因型被描述为纯合,而如果两个等位基因不同,那么基因型被描述为杂合。

[0121] “启动子”是通常包含TATA框的DNA调控区,所述TATA框能够指导RNA聚合酶II以使RNA合成在特定多核苷酸序列的适当转录起始位点处起始。启动子可另外包含影响转录起始速率的其他区域。本文公开的启动子序列调节可操作地连接的多核苷酸的转录。启动子可在一个或多个本文公开的细胞类型(例如真核细胞、非人哺乳动物细胞、人细胞、啮齿动物细胞、多能细胞、单细胞期胚胎、分化细胞或其组合)中具有活性。启动子可为例如组成型活性启动子、条件型启动子、诱导型启动子、时间限制启动子(例如发育调控的启动子)、或空间限制启动子(例如细胞特异性或组织特异性启动子)。启动子的实例可例如见于以引用的方式整体并入本文的W0 2013/176772中。

[0122] 诱导型启动子的实例包括例如化学调控的启动子和物理调控的启动子。化学调控的启动子包括例如醇调控的启动子(例如醇脱氢酶(alcA)基因启动子)、四环素调控的启动子(例如四环素应答型启动子、四环素操纵子序列(tetO)、tet-On启动子或tet-Off启动子)、类固醇调控的启动子(例如大鼠糖皮质激素受体的启动子、雌激素受体的启动子、或蜕皮激素(ecdysone)受体的启动子)、或金属调控的启动子(例如金属蛋白(metalloprotein)启动子)。物理调控的启动子包括例如温度调控的启动子(例如热激启动子)和光调控的启动子(例如光诱导型启动子或光阻遏型启动子)。

[0123] 组织特异性启动子可为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)。

[0124] 发育调控的启动子包括例如仅在胚胎发育阶段期间或仅在成人细胞中具有活性的启动子。

[0125] “可操作连接”或处于“可操作地连接”包括两个或更多个组分(例如启动子和另一序列元件)毗邻以使两个组分均正常起作用,并且使得有可能至少一个组分可介导对至少一个其他组分施加的作用。举例来说,如果启动子应答于存在或不存在一个或多个转录调控因子来控制编码序列的转录水平,那么所述启动子可为可操作地连接至所述编码序列。可操作连接可包括所述序列彼此邻接,或以反式起作用(例如调控序列可在某一距离下起控制编码序列的转录的作用)。作为另一实例,可使免疫球蛋白可变区(或V(D)J区段)的核酸序列可操作地连接至免疫球蛋白恒定区的核酸序列以便允许在序列之间适当重组以得到免疫球蛋白重链或轻链序列。

[0126] 核酸的“互补性”意指一个核酸链中的核苷酸序列由于它的核苷碱基基团的定向而与相对核酸链上的另一序列形成氢键。DNA中的互补碱基通常是A与T以及C与G。在RNA中,它们通常是C与G以及U与A。互补性可为完全的或大致上的/充分的。两个核酸之间的完全互补性意指两个核酸可形成双链体,其中所述双链体中的每个碱基都通过沃森-克里克(Watson-Crick)配对来键合于互补碱基。“大致上”或“充分”互补性意指一个链中的序列不全然和/或完全互补于相对链中的序列,但在一组杂交条件(例如盐浓度和温度)下在两个链上的碱基之间发生充分键合以形成稳定杂交复合物。所述条件可通过使用序列和用以预测杂交链的 $T_m$ (熔解温度)的标准数学计算,或通过使用常规方法对 $T_m$ 进行经验确定来预测。 $T_m$ 包括在两个核酸链之间形成的一群杂交复合物被50%变性(即一群双链核酸分子变

得有半数解离成单链)所处的温度。在低于 $T_m$ 的温度下,有利于形成杂交复合物,而在高于 $T_m$ 的温度下,有利于杂交复合物中各链的熔解或分离。可通过使用例如 $T_m=81.5+0.41(\%G+C)$ 来估计具有已知G+C含量的核酸在1M NaCl水溶液中的 $T_m$ ,但其他已知 $T_m$ 计算虑及核酸结构特征。

[0127] “杂交条件”包括一个核酸链通过互补链相互作用和氢键合来键合于第二核酸链以产生杂交复合物所处的累积环境。所述条件包括含有核酸的水溶液或有机溶液的化学组分和它们的浓度(例如盐、螯合剂、甲酰胺)以及混合物的温度。其他因素诸如孵育时长或反应室尺寸可对环境有影响。参见例如Sambrook等, *Molecular Cloning, A Laboratory Manual*, 第2版, 第1.90-1.91、9.47-9.51、11.47-11.57页(Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), 其出于所有目的以引用的方式整体并入本文。

[0128] 杂交要求两个核酸含有互补性序列,但碱基之间的不匹配是可能的。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、本领域中熟知的变量。两个核苷酸序列之间的互补程度越大,具有那些序列的核酸的杂交物的熔解温度( $T_m$ )的值越大。对于互补性链段较短(例如历经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个或更少核苷酸具有互补性)的核酸之间的杂交,不匹配的位置变得重要(参见Sambrook等, 上文, 11.7-11.8)。通常,可杂交核酸的长度是至少约10个核苷酸。可杂交核酸的说明性最小长度包括至少约15个核苷酸、至少约20个核苷酸、至少约22个核苷酸、至少约25个核苷酸和至少约30个核苷酸。此外,必要时可根据诸如互补区域的长度和互补程度的因素来调整温度和洗涤溶液盐浓度。

[0129] 多核苷酸的序列无需100%互补于它的将可特异性杂交的靶标核酸的序列。此外,多核苷酸可历经一个或多个区段进行杂交以使间插或邻近区段不涉及于杂交事件中(例如环结构或发夹结构)。多核苷酸(例如gRNA)可包含与靶标核酸序列内它们所靶向的靶标的至少70%、至少80%、至少90%、至少95%、至少99%、或100%序列互补性。举例来说,其中18/20核苷酸互补于靶标,并且将因此进行特异性杂交的gRNA将代表90%互补性。在这个实例中,其余非互补性核苷酸可群集或与互补性核苷酸交替分散,并且无需邻接于彼此或互补性核苷酸。

[0130] 核酸内特定核酸序列链段之间的互补性百分比可通过以下方式常规确定:使用本领域中已知的BLAST程序(基本局部比对搜索工具)和PowerBLAST程序(Altschul等(1990) *J. Mol. Biol.* 215:403-410; Zhang和Madden(1997) *Genome Res.* 7:649-656),或利用缺省设置,使用Gap程序(威斯康星序列分析包(Wisconsin Sequence Analysis Package), 第8版,用于Unix, Genetics Computer Group, University Research Park, Madison Wis.), 其使用Smith和Waterman的算法(*Adv. Appl. Math.*, 1981, 2, 482-489)。

[0131] 本文提供的方法和组合物采用多种不同组分。在整篇描述中都应认识到一些组分可具有活性变体和片段。所述组分包括例如Cas9蛋白、CRISPR RNA、tracrRNA和引导RNA。这些组分各自的生物活性在本文中其他地方描述。

[0132] 在两个多核苷酸或多肽序列的情形下的“序列同一性”或“同一性”涉及当历经指定比较窗进行比对以达成最大对应性时两个序列中相同的残基。当关于蛋白质使用序列同一性百分比时,应认识到不同一的残基位置常常差异在于保守性氨基酸取代,其中氨基酸

残基被取代成具有类似化学性质(例如电荷或疏水性)的其他氨基酸残基,因此不改变分子的功能性质。当序列差异在于保守性取代时,序列同一性百分比可被调高以关于取代的保守性质加以修正。差异在于所述保守性取代的序列被称为具有“序列类似性”或“类似性”。用于进行这个调整的手段为本领域技术人员所熟知。通常,这涉及将保守性取代作为部分而非完全不匹配进行评分,由此使序列同一性百分比增加。因此,举例来说,当同一氨基酸被给与评分1,并且非保守性取代被给与评分0时,保守性取代被给与在0与1之间的评分。例如如程序PC/GENE(Intelligenetics,Mountain View,California)中所执行来计算保守性取代评分。

[0133] “序列同一性百分比”包括通过历经比较窗比较两个最优对准序列来确定的值,其中相较于参照序列(其不包含添加或缺失),比较窗中的多核苷酸序列的一部分可包含添加或缺失(即空位)以达成两个序列的最优比对。通过以下方式来计算百分比:确定同一核苷酸碱基或氨基酸残基存在于两个序列中所处的位置的数目以产生匹配位置的数目,用匹配位置的数目除以比较窗中的位置总数,以及用100乘以结果以产生序列同一性百分比。

[0134] 除非另外陈述,否则序列同一性/类似性值包括使用第10版GAP利用以下参数获得的值:核苷酸序列的同一性%和类似性%,使用GAP权重50和长度权重3以及nwsgapdna.cmp评分矩阵;氨基酸序列的同一性%和类似性%,使用GAP权重8和长度权重2以及BLOSUM62评分矩阵;或其任何等效程序。“等效程序”包括对于任何两个所论述的序列,当相较于由第10版GAP产生的相应比对时,产生具有相同核苷酸或氨基酸残基匹配和相同序列同一性百分比的比对的任何序列比较程序。

[0135] 如本文用于涉及共有表位的术语“大致上同一性”包括在相应位置中含有同一残基的序列。举例来说,如果历经相关残基链段,两个序列的相应残基中的至少70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更多是同一的,那么它们可被视为大致上同一。相关链段可为例如完整序列或可为至少5、10、15个或更多残基。

[0136] 术语“保守性氨基酸取代”是指用具有类似尺寸、电荷或极性的不同氨基酸取代通常存在于序列中的氨基酸。保守性取代的实例包括将非极性(疏水性)残基诸如异亮氨酸、缬氨酸或亮氨酸取代成另一非极性残基。同样,保守性取代的实例包括将一个极性(亲水性)残基取代成另一残基,诸如在精氨酸与赖氨酸之间,在谷氨酰胺与天冬酰胺之间,或在甘氨酸与丝氨酸之间。另外,将碱性残基诸如赖氨酸、精氨酸或组氨酸取代成另一残基,或将一个酸性残基诸如天冬氨酸或谷氨酸取代成另一酸性残基是保守性取代的额外实例。非保守性取代的实例包括将非极性(疏水性)氨基酸残基诸如异亮氨酸、缬氨酸、亮氨酸、丙氨酸或甲硫氨酸取代成极性(亲水性)残基诸如半胱氨酸、谷氨酰胺、谷氨酸或赖氨酸,和/或将极性残基取代成非极性残基。以下概述典型氨基酸分类。

|        |      |     |   |     |    |      |
|--------|------|-----|---|-----|----|------|
| [0137] | 丙氨酸  | Ala | A | 非极性 | 中性 | 1.8  |
|        | 精氨酸  | Arg | R | 极性  | 正性 | -4.5 |
|        | 天冬酰胺 | Asn | N | 极性  | 中性 | -3.5 |
|        | 天冬氨酸 | Asp | D | 极性  | 负性 | -3.5 |
|        | 半胱氨酸 | Cys | C | 非极性 | 中性 | 2.5  |
|        | 谷氨酸  | Glu | E | 极性  | 负性 | -3.5 |
|        | 谷氨酰胺 | Gln | Q | 极性  | 中性 | -3.5 |
|        | 甘氨酸  | Gly | G | 非极性 | 中性 | -0.4 |
|        | 组氨酸  | His | H | 极性  | 正性 | -3.2 |
|        | 异亮氨酸 | Ile | I | 非极性 | 中性 | 4.5  |
|        | 亮氨酸  | Leu | L | 非极性 | 中性 | 3.8  |
|        | 赖氨酸  | Lys | K | 极性  | 正性 | -3.9 |
|        | 甲硫氨酸 | Met | M | 非极性 | 中性 | 1.9  |
|        | 苯丙氨酸 | Phe | F | 非极性 | 中性 | 2.8  |
|        | 脯氨酸  | Pro | P | 非极性 | 中性 | -1.6 |
|        | 丝氨酸  | Ser | S | 极性  | 中性 | -0.8 |
|        | 苏氨酸  | Thr | T | 极性  | 中性 | -0.7 |
|        | 色氨酸  | Trp | W | 非极性 | 中性 | -0.9 |
|        | 酪氨酸  | Tyr | Y | 极性  | 中性 | -1.3 |
|        | 缬氨酸  | Val | V | 非极性 | 中性 | 4.2  |

[0138] 关于免疫球蛋白核酸序列的术语“种系”包括可传至子代的核酸序列。

[0139] 术语“抗原结合蛋白”包括结合抗原的任何蛋白质。抗原结合蛋白的实例包括抗体、抗体的抗原结合片段、多特异性抗体(例如双特异性抗体)、scFV、双scFV、微型双功能抗体、微型三功能抗体、微型四功能抗体、V-NAR、VHH、VL、F(ab)、F(ab)<sub>2</sub>、DVD(双可变结构域抗原结合蛋白)、SVD(单可变结构域抗原结合蛋白)、双特异性T细胞衔接物(BiTE)或Davisbody(美国专利号8,586,713,其出于所有目的以引用的方式整体并入本文)。

[0140] 术语“抗原”是指是整个分子或分子内的结构域的物质,其能够引发产生与其具有结合特异性的抗体。术语抗原也包括在野生型宿主生物体中将借助于自身识别而不引发抗体产生,但可在进行适当遗传工程化以破坏免疫耐受性的宿主动物中引发这种应答的物质。

[0141] 术语“表位”是指抗原上由抗原结合蛋白(例如抗体)所结合的位点。表位可由连续氨基酸或通过一个或多个蛋白质的三级折叠而毗邻的非连续氨基酸形成。由连续氨基酸形成的表位(也称为线性表位)在暴露于变性溶剂时通常得以保留,而通过三级折叠形成的表位(也称为构象性表位)在用变性溶剂处理时通常丧失。表位通常包括至少3个,并且更通常至少5个或8-10个呈独特空间构象的氨基酸。确定表位的空间构象的方法包括例如x射线晶体照相术和2维核磁共振。参见例如Epitope Mapping Protocols, Methods in Molecular Biology, 第66卷, Glenn E. Morris编(1996), 其出于所有目的以引用的方式整体并入本文。

[0142] 术语“自身”在与抗原或表位联合使用时描述抗原或表位将借助于包括在通常由宿主物种生物合成或宿主物种通常所暴露于的物质之中而不由宿主物种的野生型成员的B

细胞受体识别或仅由其不良识别。所述物质诱导宿主免疫系统的耐受性。术语“外来”在与抗原或表位联合使用时描述抗原或表位不是自身抗原或自身表位。外来抗原是通常不由宿主物种产生的任何抗原。

[0143] 术语“抗体”包括免疫球蛋白分子,其包含四个多肽链,即通过二硫键相互连接的两个重(H)链和两个轻(L)链。各重链包含重链可变结构域和重链恒定区( $C_H$ )。重链恒定区包含三个结构域: $C_H1$ 、 $C_H2$ 和 $C_H3$ 。各轻链包含轻链可变结构域和轻链恒定区( $C_L$ )。重链和轻链可变结构域可进一步再分成称为互补决定区(CDR)的高变区,其与称为框架区(FR)的更保守区域交替分散。各重链和轻链可变结构域包含三个CDR和四个FR,从氨基末端至羧基末端按以下顺序安置:FR1、CDR1、FR2、CDR2、FR3、CDR3、FR4(重链CDR可缩写为HCDR1、HCDR2和HCDR3;轻链CDR可缩写为LCDR1、LCDR2和LCDR3)。术语“高亲和力”抗体是指关于它的靶标表位具有约 $10^{-9}$ M或更低(例如约 $1 \times 10^{-9}$ M、 $1 \times 10^{-10}$ M、 $1 \times 10^{-11}$ M或约 $1 \times 10^{-12}$ M)的 $K_D$ 的抗体。在一个实施方案中, $K_D$ 通过表面等离子体共振例如BIACORE™来测量;在另一实施方案中, $K_D$ 通过ELISA来测量。

[0144] 术语“重链”或“免疫球蛋白重链”包括来自任何生物体的免疫球蛋白重链序列,包括免疫球蛋白重链恒定区序列。除非另外规定,否则重链可变结构域包括三个重链CDR和四个FR区。重链的片段包括CDR、CDR和FR、及其组合。在可变结构域之后(从N末端至C末端),典型重链具有 $C_H1$ 结构域、铰链、 $C_H2$ 结构域和 $C_H3$ 结构域。重链的功能性片段包括能够特异性识别表位(例如以在微体积摩尔浓度、纳体积摩尔浓度或皮体积摩尔浓度的范围内的 $K_D$ 识别表位),能够由细胞表达并分泌,以及包含至少一个CDR的片段。重链可变结构域由可变区核苷酸序列编码,所述序列通常包含源于种系中存在的 $V_H$ 、 $D_H$ 和 $J_H$ 区段的谱系的 $V_H$ 、 $D_H$ 和 $J_H$ 区段。各种生物体的V、D和J重链区段的序列、位置和命名可见于IMGT数据库中,所述数据库可通过因特网在万维网(www)上以URL“imgt.org”访问。

[0145] 术语“轻链”包括来自任何生物体的免疫球蛋白轻链序列,并且除非另外规定,否则包括人 $\kappa$ 和 $\lambda$ 轻链和VpreB以及替代轻链。除非另外规定,否则轻链可变结构域通常包括三个轻链CDR和四个框架(FR)区。通常,全长轻链从氨基末端至羧基末端包括可变结构域(其包括FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4)和轻链恒定区氨基酸序列。轻链可变结构域由轻链可变区核苷酸序列编码,所述序列通常包含源于种系中存在的轻链V和J基因区段的谱系的轻链 $V_L$ 和轻链 $J_L$ 基因区段。各种生物体的轻链V和J基因区段的序列、位置和命名可见于IMGT数据库中,所述数据库可通过因特网在万维网(www)上以URL“imgt.org”访问。轻链包括例如不选择性结合由它们出现于其中的表位结合蛋白选择性结合的第一或第二表位的那些。轻链也包括结合以及识别一个或多个由它们出现于其中的表位结合蛋白选择性结合的表位,或辅助重链结合以及识别所述表位的那些。

[0146] 如本文所用的术语“互补决定区”或“CDR”包括由生物体的免疫球蛋白基因的通常(即在野生型动物中)出现在免疫球蛋白分子(例如抗体或T细胞受体)的轻链或重链的可变区中两个框架区之间的核酸序列编码的氨基酸序列。CDR可由例如种系序列或重排序列以及例如由原初或成熟B细胞或T细胞编码。CDR可被以体细胞方式突变(例如从在动物的种系中编码的序列变化),人源化,和/或用氨基酸取代、添加或缺失来修饰。在一些情况下(例如对于CDR3),CDR可由两个或更多个不连续(例如在未重排核酸序列中),但例如由于对序列进行剪接或连接(例如V-D-J重组以形成重链CDR3)而在B细胞核酸序列中连续的序列(例如

种系序列) 编码。

[0147] 术语“未重排”包括免疫球蛋白基因座的以下状态:其中V基因区段和J基因区段(对于重链,也包括D基因区段)被分开维持,但能够接合以形成包含V(D)J谱系的单一V、(D)、J的重排V(D)J基因。

[0148] 术语重链可变区基因座包括例如小鼠染色体的染色体上见到野生型重链可变( $V_H$ )、重链多样性( $D_H$ )和重链接合( $J_H$ )区域DNA序列所处的位置。

[0149] 术语 $\kappa$ 轻链可变区基因座包括例如小鼠染色体的染色体上见到野生型 $\kappa$ 可变( $V_K$ )和 $\kappa$ 接合( $J_K$ )区域DNA序列所处的位置。

[0150] 术语 $\lambda$ 轻链可变区基因座包括例如小鼠染色体的染色体上见到野生型 $\lambda$ 可变( $V_\lambda$ )和 $\lambda$ 接合( $J_\lambda$ )区域DNA序列所处的位置。

[0151] “同源性”序列(例如核酸序列)包括以下序列:其与已知参照序列同一或大致上类似,以致它与所述已知参照序列例如至少50%、至少55%、至少60%、至少65%、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%、或100%同一。同源序列可包括例如直系同源序列和旁系同源序列。同源性基因例如通常通过物种形成事件(直系同源基因)或遗传重复事件(旁系同源基因)而起源于共同祖先DNA序列。“直系同源”基因包括不同物种中通过物种形成而从共同祖先基因进化的基因。直系同源物通常在进化的过程中保留相同功能。“旁系同源”基因包括由于基因组内的重复而相关的基因。旁系同源物可在进化的过程中进化出新功能。

[0152] 术语“体外”包括人工环境以及发生在人工环境(例如试管)内的过程或反应。术语“体内”包括天然环境(例如细胞或生物体或身体)以及发生在天然环境内的过程或反应。术语“离体”包括已从个体的身体移除的细胞以及发生在所述细胞内的过程或反应。

[0153] 术语“杂交物”包括在一对同源染色体中第一染色体与第二染色体之间,在一个或多个靶基因组基因座处具有一个或多个序列变化(例如具有等位基因变化)的细胞或品系。举例来说,杂交细胞可源于两个遗传相异亲本之间交配(即在一个或多个基因方面不同的亲本之间杂交)的子代。作为一实例,杂交物可通过使两个不同近交系(即被交配以获得遗传同质性的品系)杂交来产生。所有人都被视为杂交物。

[0154] “包含”或“包括”一个或多个叙述的要素的组合物或方法可包括未明确叙述的其他要素。举例来说,“包含”或“包括”蛋白质的组合物可含有单独或与其他成分组合的所述蛋白质。

[0155] 对值的范围的指定包括所述范围内或界定所述范围的所有整数,以及由所述范围内的整数界定的所有子范围。

[0156] 除非另外根据上下文显而易见,否则术语“约”涵盖在陈述值的标准测量误差界限(例如SEM)内的值。

[0157] 除非上下文另外明确规定,否则单数形式的冠词“一个(种)(a/an)”和“这个(种)(the)”包括复数个(种)指示物。举例来说,术语“一个(种)Cas9蛋白”或“至少一个(种)Cas9蛋白”可包括复数个(种)Cas9蛋白,包括其混合物。

[0158] 统计显著意指 $p \leq 0.05$ 。

## 具体实施方式

### [0159] I. 概述

[0160] 本文提供用于产生结合与自身抗原共有表位或与所述自身抗原同源的目标外来靶抗原(例如目标人靶抗原)上的所述表位的抗原结合蛋白(例如抗体)的组合物和改进方法。所述方法包括通过采用两个或更多个引导RNA(gRNA)以在单一靶基因组基因座内的不同位点处产生成对双链断裂来使非人动物诸如啮齿动物(例如小鼠或大鼠)(任选在它们的种系中包含人源化免疫球蛋白重链和/或轻链基因座)中对外来抗原的耐受性降低。任选地,包含靶基因组基因座的细胞是杂交细胞,并且方法进一步包括选择靶基因组基因座内用以经受靶向遗传修饰的靶区域,以使在一对同源染色体中相应第一染色体与第二染色体之间,相对于所述靶基因组基因座的其余部分的全部或一部分,所述靶区域具有更高序列同一性程度。所述成对双链断裂影响自身抗原的表达以使自身抗原的表达降低或消除,或使来自自身抗原的与外来抗原共有的表位的表达降低或消除。包含人源化免疫球蛋白重链和轻链基因座以及也在靶基因组基因座中具有这种突变的所述经遗传修饰的非人动物可接着用外来抗原免疫,可在足以使非人动物产生对外来抗原的免疫应答的条件下维持非人动物,并且可从非人动物或来自非人动物的细胞获得结合外来抗原的抗原结合蛋白。

[0161] 用于产生针对人抗原的抗体的小鼠,诸如在它们的种系中包含人源化免疫球蛋白重链和/或轻链基因座的小鼠,通常源于品系的组合,所述组合由于相较于其他小鼠品系,BALB/c品系用于产生多样性抗体谱系的能力增加而包括BALB/c。然而,相较于通常用于在小鼠中产生靶向遗传修饰的胚胎干(ES)细胞(例如本文所述的F1H4(VGF1)细胞),源于所述抗体产生性小鼠品系的ES细胞在培养中被靶向的能力和/或用于产生具有靶向遗传修饰以及通过种系传递靶向修饰的F0代小鼠的能力通常降低。因此,用以产生靶标敲除小鼠以克服耐受性的常规方法涉及多轮交配和/或连续靶向,其中用于分娩就在目标靶标处的无效等位基因来说是纯合的以及准备用于免疫的小鼠的整个过程花费约15-16个月。

[0162] 本文所述的方法有利地使这个时间降低至约4至5个月(并且就在目标靶标处的无效等位基因来说是纯合的小鼠幼仔可在约3个月内分娩)。除时间范围较短之外,本文所述的方法也使为产生纯合性修饰所需的电穿孔的轮数降低,使所需传代数和培养时间降低,使所需细胞数降低,并且由于不需要靶向载体以及筛选相应地得以简化而使过程精简。相较于其中自身抗原的表达未被消除的小鼠,由于重链和轻链V基因区段的使用增加,所以本文所述的方法有利地导致在用目标外来抗原免疫之后,抗体的多样性增加。此外,由于产生与相应自身抗原交叉反应的抗体(即结合在自身抗原与目标外来抗原之间重叠的表位的抗体),本文所述的方法产生在用目标外来抗原免疫之后针对较多样性的表位产生的抗体,由此使得能够产生较大的针对目标外来抗原的抗体汇集物。

### [0163] II. 修饰靶基因组基因座以破坏耐受性的方法

[0164] 用“非自身”蛋白质使在它们的种系中包含人源化免疫球蛋白重链和/或轻链基因座的非人动物(例如啮齿动物,诸如小鼠或大鼠)免疫是一种通常使用的用以获得特异性抗原结合蛋白诸如单克隆抗体的方法。免疫方法具有吸引力,因为它有可能提供已在体内成熟的高亲和力抗原结合蛋白,并且可具有成本效益与时间效益两者。然而,这个方法依赖于非人动物中的天然蛋白质与经免疫以使得非人动物的免疫系统能够将免疫原识别为非自身(即外来)的蛋白质之间的序列趋异性。

[0165] B细胞受体通过一系列重组事件来由基因区段(例如V、D和J)的有序排列加以装配,并且已知这个基因区段装配是不精确的并产生对包括自身抗原的各种抗原具有亲和力的受体。尽管具有产生结合自身分子的B细胞受体的这个能力,但免疫系统具备若干自身耐受性机理以避免所述自体反应性B细胞受体的产生和扩增,并且辨别自身与非自身,由此防止自体免疫性。参见例如Shlomchik (2008) *Immunity* 28:18-28以及Kumar和Mohan (2008) 40 (3):208-23,其各自出于所有目的以引用的方式整体并入本文。因此,在具有人源化免疫球蛋白基因座的非人动物中产生针对与非人动物的自身抗原具有高度同源性(例如结构同源性或序列同源性)的人抗原的人抗体可由于免疫耐受性而是一项困难任务。因为蛋白质的在功能上重要的区域在物种之间倾向于保守,所以对自身抗原的免疫耐受性常常对产生针对这些关键表位的抗体造成挑战。用高度类似或“同源”的外来(例如人)抗原使非人动物(例如啮齿动物,诸如小鼠或大鼠)免疫产生微弱或不存在抗体应答,因此,使得获得针对所述人抗原进行结合的抗原结合蛋白(例如抗体)成问题。作为一实例,由内源性蛋白质(自身抗原)和外来靶抗原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,以致免疫系统不将靶抗原识别为外来。举例来说,外来抗原与非人动物中的自身抗原之间共有的表位可使得在所述非人动物中发动针对所述外来抗原的有效免疫应答成问题,因为免疫耐受性使表达针对所述外来抗原的中和抗体的B细胞消减和/或缺失。为克服这个耐受性以及非人动物中获得结合自身抗原或其同源物(例如人同源物)的单克隆抗体,可产生特定经遗传修饰或敲除非人动物以移除与它的编码用于免疫的抗原的人对应物基因共有显著同源性和/或高度保守的编码非人动物蛋白质的基因(或目标共有表位)。参见例如美国专利号7,119,248,其出于所有目的以引用的方式整体并入本文。然而,产生所述非人动物可为昂贵的和耗时的。

[0166] 用以产生靶标敲除小鼠以克服耐受性的常规方法涉及多轮交配和/或连续靶向。用于产生针对人抗原的抗体的小鼠,诸如在它们的种系中包含人源化免疫球蛋白重链和/或轻链基因座的小鼠(例如VELOCIMMUNE<sup>®</sup>小鼠,其在IgH基因座与Igκ基因座两者处均是纯合性人源化的),通常源于品系的组合,所述组合由于相较于其他小鼠品系,BALB/c品系用于产生多样性抗体谱系的能力增加而包括BALB/c。然而,相较于通常用于在小鼠中产生靶向遗传修饰的胚胎干(ES)细胞(例如本文所述的F1H4(VGF1)细胞,其包含50%129SvS6品系和50%C57BL/6N品系),源于所述抗体产生性小鼠品系的ES细胞在培养中被靶向的能力和/或用于产生具有靶向遗传修饰以及通过种系传递靶向修饰的F0代小鼠的能力通常降低。因此,用以破坏抗体产生性小鼠诸如VELOCIMMUNE<sup>®</sup>小鼠中的免疫耐受性的传统方法涉及首先靶向更易于接受靶向以及通过种系传递靶向修饰的ES细胞系(例如F1H4)中的编码自身抗原的基因。在这种方法中,设计大型靶向载体(LTVEC),在F1H4ES细胞中产生敲除(无效)等位基因,并且产生在目标靶标处携带杂合性敲除突变的F0小鼠(典型时间范围是5个月)。接着使VELOCIMMUNE<sup>®</sup>小鼠与在目标靶标处携带杂合性敲除突变的F0小鼠交配。为产生适于免疫的三重纯合性小鼠(目标靶标的纯合性无效和在IgH与Igκ两者处的纯合性人源化),需要再进行两代交配。整个过程花费约15至16个月(参见例如图1),并且比下述连续靶向方法(参见例如图2)更有效。

[0167] 或者,可设计和构建大型靶向载体(LTVEC),接着电穿孔至源于抗体产生性小鼠(例如 VELOCIMMUNE<sup>®</sup>小鼠或包含功能性异位小鼠 Adam 6 基因的 VELOCIMMUNE<sup>®</sup>小鼠(“VI-3小鼠”))的胚胎干(ES)细胞中以在编码与靶抗原同源或共有目标表位的自身抗原的内源性基因中产生杂合性修饰。接着进行第二轮靶向以产生纯合性修饰。尽管耗时少于上述交配方法,但这个方法可仍然是耗时的,花费约9至10个月来产生准备用于用靶抗原进行的免疫的F0小鼠(参见例如图2)。此外,所述方法需要多轮电穿孔和伴有更多次传代的较长培养时间,其全都导致多能性降低和产生用于产生抗原结合蛋白的F0小鼠的能力降低。参见例如Buehr等(2008)Cell 135:1287-1298;Li等(2008)Cell 135(7):1299-1310;以及Liu等(1997)Dev.Dyn.209:85-91,其各自出于所有目的以引用的方式整体并入本文。

[0168] 本文所述的方法有利地使这个时间降低至约4至5个月(参见例如图3;就在目标靶标处的无效等位基因来说是纯合的小鼠幼仔可在约3个月内分娩,但接着在免疫之前老化4-5周)。除时间范围较短之外,本文所述的方法也使为产生纯合性修饰所需的电穿孔的轮数降低,使所需传代数和培养时间降低,并且使所需细胞数降低。筛选更加简单和精简,因为例如不需要等位基因获得探针,并且不需要拷贝数校正。相较于其中自身抗原的表达未被消除的小鼠,由于重链和轻链V基因区段的使用增加,所以本文所述的方法也导致在用目标外来抗原免疫之后,抗体的多样性增加。此外,由于产生与相应自身抗原交叉反应的抗体(即结合在自身抗原与目标外来抗原之间重叠的表位的抗体),本文所述的方法可产生在用目标外来抗原免疫之后针对较大多样性的表位产生的抗体,由此使得能够产生较大的针对目标外来抗原的抗体汇集物。

[0169] 本文提供用于修饰靶基因组基因座以破坏耐受性的各种方法。方法可离体或在体内进行,并且它们可利用两个或更多个靶向影响与目标外来抗原同源或共有目标表位的自身抗原的表达的单一靶基因组基因座内的不同区域,并且与Cas蛋白形成两个或更多个复合物并裂解靶标核酸的引导RNA(例如两个gRNA、三个引导RNA或四个引导RNA)。两个或更多个引导RNA可单独或与外源性修复模板组合使用,前提是如果细胞是例如单细胞期胚胎,那么外源性修复模板的长度可小于5kb。所述方法促进在靶基因座处产生双等位基因遗传修饰,并且可包括基因组垮塌或其他靶向修饰,诸如同时缺失基因组内的核酸序列和用外源性核酸序列进行替换。相较于在低频率下产生双等位基因修饰的用一个gRNA进行的靶向,用两个或更多个gRNA进行的靶向导致在显著增加的比率下产生双等位基因修饰(例如以纯合方式靶向的细胞、以纯合方式缺失的细胞和包括以半合方式靶向的细胞的复合的以杂合方式靶向的细胞)。

[0170] 应答于双链断裂(DSB)的修复主要通过以下两个保守DNA修复路径来发生:非同源性末端接合(NHEJ)和同源性重组(HR)。参见Kasperek和Humphrey(2011)Seminars in Cell&Dev.Biol.22:886-897,其出于所有目的以引用的方式整体并入本文。NHEJ包括通过使断裂末端直接连接至彼此或外源性序列而无需同源性模板来修复核酸中的双链断裂。由NHEJ达成的对非连续序列的连接可常常在双链断裂的位点附近导致缺失、插入或易位。

[0171] 由外源性修复模板介导的对靶标核酸的修复可包括在两个多核苷酸之间交换遗传信息的任何过程。举例来说,NHEJ也可通过使断裂末端与外源性修复模板的末端直接连

接(即基于NHEJ的捕集)来导致外源性修复模板的靶向整合。当同源性引导的修复(HDR)路径不易于可用时(例如在非分裂细胞、原代细胞和不良进行基于同源性的DNA修复的细胞中),所述NHEJ介导的靶向整合可优选用于插入外源性修复模板。此外,与同源性引导的修复相比,不需要了解侧接于裂解位点的具有序列同一性的大型区域(超出由Cas介导的裂解产生的突出部分),此在试图向具有对基因组序列的了解有限的基因组的生物体中进行靶向插入时可有益的。整合可通过外源性修复模板与经裂解基因组序列之间的钝性末端连接,或通过使用由可与经裂解基因组序列中的由Cas蛋白产生的那些突出部分相容的突出部分侧接的外源性修复模板进行的粘性末端(即具有5'或3'突出部分)连接来进行。参见例如US 2011/020722、WO 2014/033644、WO 2014/089290以及Maresca等(2013) *Genome Res.* 23 (3):539-546,其各自出于所有目的以引用的方式整体并入本文。如果连接钝性末端,那么可需要进行靶标和/或供体切除以产生具有为片段接合所需的微同源性的区域,此可在靶标序列中产生非所要改变。

[0172] 修复也可通过同源性引导的修复(HDR)或同源性重组(HR)来发生。HDR或HR包括可需要核苷酸序列同源性,使用“供体”分子作为“靶标”分子(即经受双链断裂的分子)的修复模板,并且导致遗传信息从供体向靶标转移的核酸修复形式。在不希望受任何特定理论束缚下,所述转移可涉及对在断裂的靶标与供体之间形成的异源双链体DNA的不匹配修正,和/或合成依赖性链退火,其中供体用于再合成将成为靶标的一部分的遗传信息,和/或相关过程。在一些情况下,供体多核苷酸、供体多核苷酸的一部分、供体多核苷酸的拷贝、或供体多核苷酸的拷贝的一部分整合至靶标DNA中。参见Wang等(2013) *Cell* 153:910-918; Mandalos等(2012) *PLOS ONE* 7:e45768:1-9;以及Wang等(2013) *Nat Biotechnol.* 31:530-532,其各自出于所有目的以引用的方式整体并入本文。

[0173] 为制备对目标外来靶抗原的耐受性降低的非人动物,可靶向一个或多个影响与所述目标外来抗原同源或共有表位的自身抗原的表达的靶基因组基因座以使所述自身抗原的表达降低。优选地,自身抗原的表达被消除。如果自身抗原不再表达(例如如果自身抗原是蛋白质,那么所述蛋白质不再表达,或如果自身抗原是蛋白质上的特定表位,那么包含那个表位的蛋白质不再表达),那么自身抗原的表达被视为得以消除。

[0174] 在一个实例中,可使不是单细胞期胚胎的非人动物多能细胞(例如胚胎干(ES)细胞)的基因组与Cas蛋白、杂交至靶基因组基因座内的第一引导RNA识别序列的第一引导RNA和杂交至靶基因组基因座内的第二引导RNA识别序列的第二引导RNA接触。在另一实例中,可使非人动物单细胞期胚胎的基因组与Cas蛋白、杂交至靶基因组基因座内的第一引导RNA识别序列的第一引导RNA和杂交至靶基因组基因座内的第二引导RNA识别序列的第二引导RNA接触。

[0175] 在一些本文提供的方法中,所靶向的细胞是如在本文中其他地方定义的杂交细胞。所述方法也可包括选择靶基因组基因座内的靶区域,如在本文中其他地方所述。可选择靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,相对于靶基因组基因座的其他区段或靶基因组基因座的其余部分,它具有高序列同一性百分比。作为一实例,选择靶区域可包括比较一对同源染色体中相应第一染色体和第二染色体在靶基因组基因座内的序列,以及选择在所述一对同源染色体中所述相应第一染色体与第二染色体之间,相对于所述靶基因组基因座的其余部分的全部或一部分具有更高序列同一性百分比的靶区

域。选择靶区域的方法如在本文中其他地方更详细所述。

[0176] 任选地,可进一步使基因组与杂交至靶基因组基因座内(或影响自身抗原的表达或影响与目标外来抗原同源或共有目标表位的第二自身抗原的表达的第二靶基因组基因座内)的引导RNA识别序列的额外引导RNA接触,诸如杂交至靶基因组基因座内的第三引导RNA识别序列的第三引导RNA,或所述第三引导RNA和杂交至靶基因组基因座内的第四引导RNA识别序列的第四引导RNA。接触可包括将Cas蛋白和引导RNA以任何形式以及通过任何手段引入细胞中,如在本文中其他地方更详细所述。引导RNA与Cas蛋白形成复合物,并且将它引导至靶基因组基因座处的引导RNA识别序列,在所述引导RNA识别序列处,Cas蛋白在引导RNA识别序列内的Cas蛋白裂解位点处裂解靶基因组基因座。由Cas蛋白达成的裂解可产生双链断裂或单链断裂(例如如果Cas蛋白是切口酶)。可用于方法中的Cas蛋白和引导RNA的实例和变化形式在本文中其他地方描述。由Cas蛋白在靶基因组基因座处达成的裂解可修饰一对第一染色体和第二染色体中的靶基因组基因座以产生使自身抗原的表达降低的双等位基因修饰。

[0177] 目标外来抗原可为需要其抗原结合蛋白的任何外来抗原。举例来说,目标外来抗原可包含以下、基本上由以下组成或由以下组成:病毒蛋白质、细菌蛋白质、哺乳动物蛋白质、猿猴蛋白质、犬蛋白质、猫蛋白质、马蛋白质、牛蛋白质、啮齿动物蛋白质(例如大鼠或小鼠)或人蛋白质的全部或一部分。举例来说,目标外来抗原可包含以下、基本上由以下组成或由以下组成:具有一个或多个突变或变化的人蛋白质。目标外来抗原和自身抗原可为同源的。举例来说,目标外来抗原和自身抗原可为直系同源的或旁系同源的。或者或此外,目标外来抗原和自身抗原可包含以下、基本上由以下组成或由以下组成:共有表位。共有表位可存在于同源性蛋白质之间,或可存在于不同源的相异蛋白质之间。可共有表位的线性氨基酸序列和/或构象契合(例如甚至在不存在一级序列同源性下的类似抗原性表面)。举例来说,共有表位包括大致上相同的表位。如果两个抗原之间共有表位,那么针对第一抗原上的表位的抗体将通常也结合第二抗原上的表位。

[0178] 接触可在不存在外源性修复模板下或在存在与靶基因组基因座重组的外源性修复模板下发生以产生靶向遗传修饰。举例来说,细胞可为单细胞期胚胎,并且外源性修复模板的长度可小于5kb。外源性修复模板的实例在本文中其他地方描述。

[0179] 在一些所述方法中,通过外源性修复模板修复靶标核酸通过同源性引导的修复(HDR)来发生。当Cas蛋白在靶基因组基因座处裂解两个DNA链以产生双链断裂时,当Cas蛋白是在靶基因组基因座处裂解一个DNA链以产生单链断裂的切口酶时,或当成对Cas切口酶用于产生由两个偏移切口形成的双链断裂时,可发生同源性引导的修复。在所述方法中,外源性修复模板包含对应于靶基因组基因座处的5'靶标序列和3'靶标序列的5'同源臂和3'同源臂。引导RNA识别序列或裂解位点可邻近于5'靶标序列,邻近于3'靶标序列,邻近于5'靶标序列与3'靶标序列两者,或既不邻近于5'靶标序列,也不邻近于3'靶标序列。彼此邻近的序列包括在彼此的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内的序列。任选地,外源性修复模板进一步包含由5'同源臂和3'同源臂侧接的核酸插入物,并且所述核酸插入物插入在5'靶标序列与3'靶标序列之间。如果不存在核酸插入物,那么外源性修复模板可起使5'靶标序列与3'靶向序列之间的基因组序列缺失的作用。

[0180] 或者,通过外源性修复模板修复靶标核酸可通过非同源性末端接合(NHEJ)介导的

连接来发生。在所述方法中,外源性修复模板的至少一个末端包含互补于至少一个由Cas介导的在靶基因组基因座处的裂解产生的突出部分的短单链区域。外源性修复模板中的互补性末端可侧接于核酸插入物。举例来说,外源性修复模板的各末端都可包含互补于由Cas介导的在靶基因组基因座处的裂解产生的突出部分的短单链区域,并且外源性修复模板中的这些互补性区域可侧接于核酸插入物。突出部分(即交错末端)可通过切除由Cas介导的裂解产生的双链断裂的钝性末端来产生。所述切除可产生具有为片段接合所需的微同源性的区域,但这可在靶标核酸中产生非所要或不可控制的改变。或者,所述突出部分可通过使用成对Cas切口酶来产生。举例来说,如果Cas蛋白是切口酶,那么可使靶基因组基因座与靶向相对DNA链的第一引导RNA和第二引导RNA接触,由此基因组通过进行双重切口来修饰。这可通过使靶基因组基因座与杂交至靶基因组基因座内的不同引导RNA识别序列的两个引导RNA接触来实现。两个引导RNA与Cas切口酶形成两个复合物,并且Cas切口酶在一个引导RNA识别序列内对具有靶基因组基因座的第一链进行切口,并且在另一引导RNA识别序列内对具有靶基因组基因座的第二链进行切口。外源性修复模板接着与靶基因组基因座重组以产生靶向遗传修饰。

[0181] 在一些方法中,核酸插入物包含与自身抗原编码的基因的全部或一部分同源或直系同源的序列。这可例如在敲除自身抗原可导致胚胎致死时适用。核酸插入物可在呈本文所述的任何形式(例如靶向载体、LTVEC、ssODN等)的外源性修复模板中,并且核酸插入物可进一步包含选择盒(例如自我缺失性选择盒)或可缺乏选择盒。在所述方法中,举例来说,编码自身抗原的基因的全部或一部分可被缺失以及用相应同源或直系同源序列替换。举例来说,编码自身抗原的基因的全部可被缺失以及用相应同源或直系同源序列替换,或所述基因的编码自身抗原的特定基序或区域的一部分可被缺失以及用相应同源或直系同源序列替换。任选地,相应同源或直系同源序列可来自另一物种。举例来说,如果自身抗原是小鼠抗原,那么相应同源或直系同源序列可为例如同源或直系同源大鼠、仓鼠、猫、狗、龟、狐猴或人序列。或者或另外,相较于所替换的序列,同源或直系同源序列可包含一个或多个点突变(例如1、2、3、4、5个或更多个)。所述点突变可例如起使自身抗原中的一个或多个表位的表达消除的作用。所述表位可为与目标外来抗原共有的表位。任选地,所述点突变可导致所编码多肽中的保守性氨基酸取代(例如天冬氨酸[Asp,D]被谷氨酸[Glu,E]取代)。所述氨基酸取代可导致表达保留野生型自身抗原的功能,但缺乏存在于目标外来抗原上以及与野生型自身抗原共有的表位的自身抗原。同样,缺失编码自身抗原的基因的全部或一部分以及用缺乏在目标外来抗原与自身抗原之间共有的表位的相应同源或直系同源序列替换可导致表达自身抗原的同源物或直系同源物,其保留野生型自身抗原的功能,但缺乏存在于目标外来抗原上以及与野生型自身抗原共有的表位。可接着产生针对那些表位的抗原结合蛋白。

[0182] 经修饰非人动物多能细胞可接着用于使用在本文中其他地方所述的方法来产生经遗传修饰的非人动物。举例来说,可将经修饰非人动物多能细胞引入宿主胚胎中,并且可将所述宿主胚胎植入代孕母体中以产生经遗传修饰的F0代非人动物,其中一对第一染色体和第二染色体中的靶基因组基因座被修饰以具有双等位基因修饰,以使自身抗原的表达被降低或消除。在单细胞期胚胎的情况下,可选择经遗传修饰的胚胎,接着植入代孕母体中以产生经遗传修饰的F0代非人动物,其中一对第一染色体和第二染色体中的靶基因组基因座

被修饰以具有双等位基因修饰,以使自身抗原的表达被降低或消除。F0代非人动物可接着用于使用在本文中其他地方所述的方法来产生针对目标外来抗原的抗原结合蛋白。

[0183] A. 选择靶区域

[0184] 通过外源性修复模板(例如靶向载体)与靶基因组基因座之间的同源性重组达成的靶向基因修饰可极其低效,尤其是在除啮齿动物胚胎干细胞以外的细胞类型中。由CRISPR/Cas9引导的裂解对一个或多个双链DNA断裂的诱导可促进通过外源性修复模板(例如靶向载体)与靶基因组基因座之间的同源性重组(HR)进行纯合性基因靶向。CRISPR/Cas9也可促进通过非同源性末端接合(NHEJ)修复机理进行纯合性插入或缺失突变(即相同的双等位基因改变)。对于涉及极大型人源化的基因修饰,使靶向载体与由靶向单一靶基因组基因座的两个引导RNA引导的CRISPR/Cas9核酸酶系统组合可进一步使靶向效率增强超过用一个引导RNA实现的靶向效率。相较于在低频率下产生双等位基因修饰或完全不产生双等位基因修饰的用一个引导RNA进行的靶向,用两个引导RNA进行的靶向导致在显著增加的比率下产生以纯合方式靶向的细胞、以纯合方式缺失的细胞和复合的以杂合方式靶向的细胞(包括以半合方式靶向的细胞)。然而,在一些基因组基因座处,获得以纯合方式靶向的细胞或以纯合方式缺失的细胞可仍然是困难的。

[0185] 不同于在通常用于实验室环境下的在它们的实际上所有基因组基因座处都是纯合的近交小鼠和大鼠品系中,在杂交细胞(例如在所有人中)中的靶基因组基因座处的两个等位基因的序列将通常不是100%同一的。然而,如在本文提供的实施例中所证明,无论初始CRISPR/Cas9诱导的修饰通过HR产生还是通过NHEJ产生,纯合性基因组改变的频率都取决于靶基因组基因座的两个等位基因之间的序列类似性的程度。这个观察结果暗示CRISPR/Cas9诱导的纯合性基因修饰是一种同源性依赖性现象。支持这个的是,CRISPR/Cas9诱导的纯合性修饰常常伴有在同一染色体上与靶基因组基因座相关联的等位基因序列和结构变体(单核苷酸变体SNV或结构变体SV)的杂合性丧失(LOH),如本文实施例中所证明。LOH可涉及针对在靶基因组基因座的任一侧上的变体的局部基因转换机理或涉及在靶基因组基因座的端粒侧上的所有变体的长程基因转换(极性基因转换)。所述基因转换事件必须是同源性驱动的有丝分裂重组机理的结果。

[0186] 这个认识为设计CRISPR/Cas9辅助的纯合性靶向实验提供指导。选择其中两个等位基因共有高度序列同一性的靶区域会赋予最高成功机会。CRISPR/Cas9辅助的在两个等位基因之间具有高度序列变化的靶区域处进行的纯合性靶向取得成功的可能性较小。即使在具有高密度的SNV和SV的基因座处,成功率也可通过使用识别在靶基因组基因座内具有连续等位基因序列同一性的最长可能链段内或在靶基因组基因座的其中等位基因序列同一性最大的链段内的序列的引导RNA或核酸酶试剂来改进。

[0187] 本文所述的方法可涉及选择靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,靶区域的全部或一部分的序列同一性可最大化。在杂交细胞中,一对同源染色体的一个拷贝上的序列在相较于染色体对的另一拷贝时将通常具有一些差异(例如单核苷酸变化)。因此,所述方法可包括比较一对同源染色体(例如人细胞具有23对同源染色体)中相应第一染色体和第二染色体在靶基因组基因座中的序列,接着选择所述靶基因组基因座内的靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,所述靶区域的全部或一部分的序列同一性最大化。如果无序列可用,那么所述方法可进一步包括在

比较序列之前对一对同源染色体内的各单一染色体上的靶基因组基因座测序。

[0188] 靶区域可包含以下、基本上由以下组成或由以下组成：例如在本文公开的方法中由两个或更多个引导RNA中的一者或一个或多个外源性修复模板靶向的任何区段或区域，或侧接于在本文公开的方法中由两个或更多个引导RNA中的一者或一个或多个外源性修复模板靶向的区段或区域的任何区段或区域。靶区域可为连续基因组序列或非连续基因组序列。举例来说，靶区域可包含以下、基本上由以下组成或由以下组成：通过本文公开的方法被靶向以达成缺失的基因组区段或区域、被靶向以达成替换的基因组区段或区域、或被靶向以达成插入的基因组区段或区域，和/或可包含以下、基本上由以下组成或由以下组成：侧接于通过本文公开的方法被靶向以达成缺失、替换或插入的基因组区段或基因组区域的5' 和/或3' 序列。优选地，靶区域包含以下、基本上由以下组成或由以下组成：紧靠通过本文公开的方法被靶向以达成缺失、替换或插入的区域的上游的序列和/或紧靠所述区域的下流的序列（例如在两个引导RNA识别序列或裂解位点之间的区域的上游和/或下游的序列，或在外源性修复模板的5' 靶标序列与3' 靶标序列之间的区域的上游和/或下游的序列）。作为一实例，如果使用两个引导RNA，那么靶区域可包含以下、基本上由以下组成或由以下组成：侧接于引导RNA识别序列或Cas裂解位点之间的区域的5'（即上游）和3'（即下游）序列。侧接序列的长度的实例在本文中其他地方公开。

[0189] 在一些方法中，举例来说，可首先设计外源性修复模板，并且可接着设计由外源性修复模板的5' 靶标序列和3' 靶标序列侧接的区域内的引导RNA以使引导RNA识别序列内的区域和/或侧接于（5' 侧、3' 侧或各侧）引导RNA识别序列（例如如果使用两个或更多个引导RNA，那么侧接于相隔最远的两个引导RNA识别序列之间的区域）的区域中的序列同一性最大化。或者，在一些方法中，举例来说，可首先设计两个或更多个引导RNA，并且可接着设计外源性修复模板以使5' 靶标序列和3' 靶标序列侧接于两个或更多个引导RNA识别序列，并且以使5' 靶标序列和3' 靶标序列内的区域和/或侧接于（5' 侧、3' 侧或各侧）5' 靶标序列和3' 靶标序列（例如侧接于5' 靶标序列与3' 靶标序列之间的区域）的区域中的序列同一性最大化。

[0190] 作为一实例，靶区域可包含以下、基本上由以下组成或由以下组成：两个或更多个引导RNA中的一者的引导RNA识别序列。或者或此外，靶区域可包含以下、基本上由以下组成或由以下组成：侧接于引导RNA识别序列的5' 和/或3' 序列。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样，3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0191] 作为另一实例，靶区域可包含以下、基本上由以下组成或由以下组成：两个或更多个引导RNA识别序列。或者或此外，靶区域可包含以下、基本上由以下组成或由以下组成：侧接于引导RNA识别序列的5' 和/或3' 序列。在其中使用两个引导RNA的方法中，举例来说，靶区域可包含以下、基本上由以下组成或由以下组成：由两个引导RNA识别序列或裂解位点侧接的基因组区域或由两个引导RNA识别序列或裂解位点侧接且包括所述两个引导RNA识别序列或裂解位点的基因组区域。或者或此外，靶区域可包含以下、基本上由以下组成或由以

下组成：侧接于两个引导RNA识别序列或裂解位点之间的区域的5' 和/或3' 序列，或侧接于两个引导RNA识别序列或裂解位点之间的区域且包括所述两个引导RNA识别序列或裂解位点的5' 和/或3' 序列。可在其中使用超过两个引导RNA的方法中选择类似靶区域，例外之处是替代如上由两个引导RNA识别序列或裂解位点侧接的基因组区域的将是由相隔最远的引导RNA识别序列或裂解位点侧接的基因组区域。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样，3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0192] 在其中使用外源性修复模板的方法中，举例来说，靶区域可包含以下、基本上由以下组成或由以下组成：由5' 靶标序列和3' 靶标序列侧接的区域或由5' 靶标序列和3' 靶标序列侧接且包括所述5' 靶标序列和所述3' 靶标序列的区域。或者或另外，靶区域可包含以下、基本上由以下组成或由以下组成：侧接于5' 靶标序列与3' 靶标序列之间的基因组区域的5' 和/或3' 序列或侧接于5' 靶标序列与3' 靶标序列之间的基因组区域的5' 和/或3' 序列。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样，3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0193] 可使靶区域的全部或靶区域的一部分的等位基因序列同一性最大化。作为一实例，可使与至少一个或各个引导RNA识别序列相对应的基因组区域或包含至少一个或各个引导RNA识别序列的区域的等位基因序列同一性最大化。举例来说，可使至少一个或各个引导RNA识别序列的等位基因序列同一性最大化。或者，可使至少一个或各个引导RNA识别序列和侧接于所述至少一个或各个引导RNA识别序列的5' 和/或3' 序列的等位基因序列同一性最大化。或者，可使侧接于至少一个或各个引导RNA识别序列的5' 和/或3' 序列的等位基因序列同一性最大化。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样，3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0194] 或者或另外，可使与外源性修复模板的5' 靶标序列和/或3' 靶标序列相对应的基因组区域或包含5' 靶标序列和3' 靶标序列中的至少一者或各者的区域的等位基因序列同一性最大化。举例来说，可使5' 靶标序列和3' 靶标序列中的至少一者或各者的等位基因序列同一性最大化。或者，可使5' 靶标序列和3' 靶标序列中的至少一者或各者和侧接于所述5' 靶标序列和所述3' 靶标序列中的所述至少一者或各者的5' 和/或3' 序列的等位基因序列同一性最大化。或者，可使侧接于5' 靶标序列和3' 靶标序列中的至少一者或各者的5' 和/或3' 序列的等位基因序列同一性最大化。5' 侧接序列可为例如至少10、20、30、40、50、100、

200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样,3'侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0195] 或者或另外,可使侧接于被靶向以达成缺失、替换或插入的区域的序列的等位基因序列同一性最大化。举例来说,在使用两个引导RNA的方法中,可使侧接于两个裂解位点或两个引导RNA识别序列之间的区域的5'和/或3'序列的等位基因序列同一性最大化。在使用三个或更多个引导RNA的方法中,可使侧接于相隔最远的两个裂解位点或两个引导RNA识别序列之间的区域的5'和/或3'序列的等位基因序列同一性最大化。作为另一实例,在使用外源性修复模板的方法中,可使侧接于外源性修复模板的5'靶标序列与3'靶标序列之间的区域(即由外源性修复模板靶向以达成缺失的基因组区域)的5'和/或3'序列的等位基因序列同一性最大化。5'侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。同样,3'侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列。

[0196] 选择靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,所述靶区域的全部或一部分的序列同一性最大化未必意指着眼于在一对同源染色体中第一染色体和第二染色体上的靶基因组基因座,以及挑选相对于所述靶基因组基因座的其余部分具有最高等位基因序列同一性的区域,而是代之以可考虑其他因素。举例来说,如果靶区域包含以下、基本上由以下组成或由以下组成:一个或多个引导RNA识别序列和/或侧接于所述一个或多个引导RNA识别序列的序列,那么可考虑的其他因素包括例如什么推定引导RNA识别序列位于区域中,推定引导RNA识别序列是否独特,推定引导RNA识别序列位于区域内何处,预测区域中的推定引导RNA识别序列多么有成效或特有,区域内的推定引导RNA识别序列与适于外源性修复模板的5'靶标序列和3'靶标序列的邻近性,区域内的推定引导RNA识别序列与其他推定引导RNA识别序列的邻近性,区域内的推定引导RNA识别序列与被靶向以达成修正的突变的邻近性等。举例来说,优选地,引导RNA识别序列是不存在于基因组中其他地方的独特靶标位点。参见例如US 2014/0186843,其出于所有目的以引用的方式整体并入本文。同样,引导RNA特异性可涉及GC含量和靶向性序列长度并可通过改变GC含量和靶向性序列长度来优化,并且算法可用于设计或评估使引导RNA的脱靶结合或相互作用最小化的引导RNA靶向性序列。参见例如WO 2016/094872,其出于所有目的以引用的方式整体并入本文。在一些方法中,可考虑或使用来自不同物种的Cas9蛋白(例如化脓性链球菌(*S. pyogenes*)Cas9和金黄色葡萄球菌(*S. aureus*)Cas9)以使潜在引导RNA识别序列的数目由于可用PAM序列的数目增加而增加。

[0197] 在一个实例中,可选择靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,靶区域的全部或部分具有高序列同一性百分比。举例来说,可选择靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,靶区域的全部或部分具有诸如

至少95%、95.5%、96%、96.5%、97%、97.5%、98%、98.5%、99%、99.1%、99.2%、99.3%、99.4%、99.5%、99.55%、99.6%、99.65%、99.7%、99.75%、99.8%、99.85%、99.9%、99.95%或100%序列同一性的最小序列同一性百分比。

[0198] 在另一实例中,可选择靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,靶区域的全部或部分具有低数目或低密度的单核苷酸变化。举例来说,可选择靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,靶区域的全部或部分具有诸如每kb序列不超过5、4.9、4.8、4.7、4.6、4.5、4.4、4.3、4.2、4.1、4、3.9、3.8、3.7、3.6、3.5、3.4、3.3、3.2、3.1、3、2.9、2.8、2.7、2.6、2.5、2.4、2.3、2.2、2.1、2、1.9、1.8、1.7、1.6、1.5、1.4、1.3、1.2、1.1、1、0.9、0.8、0.7、0.6、0.5、0.4、0.3、0.2、0.1或0个单核苷酸变化的最大单核苷酸变化密度。

[0199] 任选地,在一对同源染色体中相应第一染色体和第二染色体中,靶区域可相同。任选地,靶区域可在靶基因组基因座内具有连续序列同一性的最长可能链段内。

[0200] 或者或另外,可选择靶基因组基因座内的靶区域以使在一对同源染色体中相应第一染色体与第二染色体之间,相对于所述靶基因组基因座内的其他区域,靶区域的全部或部分具有高序列同一性百分比或低数目或低密度的单核苷酸变化。

[0201] 举例来说,相对于靶基因组基因座的其余部分的全部或部分,靶区域可具有更高序列同一性百分比或更低密度的单核苷酸变化。举例来说,在相应第一同源染色体与第二同源染色体之间,靶区域可具有至少99.9%序列同一性,并且在相应第一染色体与第二染色体之间,靶基因组基因座的其余部分具有不超过99.8%序列同一性。

[0202] 举例来说,靶区域可包含以下、基本上由以下组成或由以下组成:一个或多个与一个或多个引导RNA识别序列相对应的靶标基因组区域,并且相对于靶基因组基因座的其他区段,诸如与靶基因组基因座内的一个或多个其他潜在引导RNA识别序列相对应的基因组区域,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。作为一个实例,靶区域可包含以下、基本上由以下组成或由以下组成:一个或多个引导RNA识别序列中的至少一者或各者,并且相对于靶基因组基因座的其他区段,诸如靶基因组基因座内的一个或多个其他潜在引导RNA识别序列,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。作为另一实例,靶区域可包含以下、基本上由以下组成或由以下组成:一个或多个引导RNA识别序列中的至少一者或各者和侧接于所述一个或多个引导RNA识别序列中的所述至少一者或各者的5' 和/或3' 序列,并且相对于靶基因组基因座的其他区段,诸如靶基因组基因座内的一个或多个其他潜在引导RNA识别序列和它们的5' 和/或3' 侧接序列,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。作为另一实例,靶区域可包含以下、基本上由以下组成或由以下组成:侧接于一个或多个引导RNA识别序列中的至少一者或各者的5' 和/或3' 序列,并且相对于靶基因组基因座的其他区段,诸如靶基因组基因座内的一个或多个其他潜在引导RNA识别序列的5' 和/或3' 侧接序列,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样,3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序

列。

[0203] 在其中使用两个引导RNA的方法中,靶区域可包含以下、基本上由以下组成或由以下组成:与第一引导RNA识别序列相对应的第一靶标基因组区域和/或与第二引导RNA识别序列相对应的第二靶标基因组区域,并且相对于靶基因组基因座的其他区段,诸如与靶基因组基因座内的一个或多个其他潜在引导RNA识别序列相对应的基因组区域,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。举例来说,靶区域可包含以下、基本上由以下组成或由以下组成:第一引导RNA识别序列和/或第二引导RNA识别序列,并且相对于靶基因组基因座的其他区段,诸如靶基因组基因座内的一个或多个其他潜在引导RNA识别序列,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。作为另一实例,靶区域可包含以下、基本上由以下组成或由以下组成:高百分比的第一引导RNA识别序列和侧接于所述第一引导RNA识别序列的5' 和/或3' 序列和/或第二引导RNA识别序列和侧接于所述第二引导RNA识别序列的5' 和/或3' 序列,并且相对于靶基因组基因座的其他区段,诸如与靶基因组基因座内的一个或多个其他潜在引导RNA识别序列和它们的5' 和/或3' 侧接序列相对应的基因组区域,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。作为另一实例,靶区域可包含以下、基本上由以下组成或由以下组成:侧接于第一引导RNA识别序列的5' 和/或3' 序列和/或侧接于第二引导RNA识别序列的5' 和/或3' 序列,并且相对于靶基因组基因座的其他区段,诸如侧接于靶基因组基因座内的一个或多个其他潜在引导RNA识别序列的5' 和/或3' 序列,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样,3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0204] 因此,在其中在选择靶区域时考虑一个引导RNA的方法中,举例来说,选择靶区域可包括比较靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:不存在于基因组中其他地方的不同引导RNA识别序列和在所述不同引导RNA识别序列的5' 侧、3' 侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,以及将相对于其他区段具有最高序列同一性百分比的区段选作靶区域。如果使用两个或更多个引导RNA,方法可包括将相对于其他区段具有最高序列同一性百分比的两个或更多个区段选作靶区域。任选地,一个或多个区段可包含以下、基本上由以下组成或由以下组成:与靶基因组基因座中而非存在于基因组中其他地方的各引导RNA识别序列相对应的区段。

[0205] 或者或另外,在其中使用两个引导RNA的方法中,靶区域可包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列或第一裂解位点与第二裂解位点之间的区域,并且相对于靶基因组基因座的其他区段,诸如在靶基因组基因座内的一对或多对其他潜在引导RNA识别序列或裂解位点之间的区域,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。如果使用三个或更多个引导RNA,相关区域将是

在相隔最远的两个引导RNA识别序列或两个裂解位点之间的区域。

[0206] 因此,在其中使用两个引导RNA的方法中,举例来说,选择靶区域可包括比较靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的区域,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他区段具有最高序列同一性百分比的区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与靶基因组基因座中各对不同引导RNA识别序列相对应的区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0207] 或者或另外,在其中使用两个引导RNA的方法中,靶区域可包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列或第一裂解位点与第二裂解位点之间的区域和侧接于在所述第一引导RNA识别序列与所述第二引导RNA识别序列或所述第一裂解位点与所述第二裂解位点之间的基因组区域的5' 和/或3' 序列,并且相对于靶基因组基因座的其他区段,诸如在靶基因组基因座内的一对或多对其他潜在引导RNA识别序列或裂解位点之间的区域和侧接于在一对或多对其他潜在引导RNA识别序列或裂解位点之间的基因组区域的5' 和/或3' 序列,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。优选地,靶区域可包含以下、基本上由以下组成或由以下组成:在第一引导RNA识别序列与第二引导RNA识别序列或第一裂解位点与第二裂解位点之间的基因组区域和侧接于在所述第一引导RNA识别序列与所述第二引导RNA识别序列或所述第一裂解位点与所述第二裂解位点之间的所述基因组区域的5' 和3' 序列,并且相对于靶基因组基因座的其他区段,诸如在靶基因组基因座内的一对或多对其他潜在引导RNA识别序列或裂解位点之间的区域和侧接于在一对或多对其他潜在引导RNA识别序列或裂解位点之间的基因组区域的5' 和3' 序列,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。如果使用三个或更多个引导RNA,那么相关区域将是侧接于在相隔最远的两个引导RNA识别序列或两个裂解位点之间的基因组区域的5' 和/或3' 序列。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样,3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0208] 因此,在其中使用两个引导RNA的方法中,举例来说,选择靶区域可包括比较靶基因组基因座的两个或更多个区段,其中各区段包含以下、基本上由以下组成或由以下组成:在一对不同引导RNA识别序列之间的区域和在所述一对不同引导RNA识别序列之间的基因组区域的5' 侧、3' 侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列,其中所述引导RNA识别序列不存在于基因组中其他地方,以及将相对于其他区段具有最高序列同一性百分比的区段选作靶区域。任选地,一个或多个区段包含以下、基本上由以下组成或由以下组成:与靶基因组基因座中各对不同引导RNA识别序列相对应的区段,其中所述引导RNA识别序列不存在于基因组中其他地方。

[0209] 或者或另外,在其中使用两个引导RNA的方法中,靶区域可包含以下、基本上由以

下组成或由以下组成：侧接于在第一引导RNA识别序列与第二引导RNA识别序列或第一裂解位点与第二裂解位点之间的基因组区域的5' 和/或3' 序列，并且相对于靶基因组基因座的其他区段，诸如侧接于在靶基因组基因座内的一对或多对其他潜在引导RNA识别序列或裂解位点之间的基因组区域的5' 和/或3' 序列，靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。优选地，靶区域可包含以下、基本上由以下组成或由以下组成：侧接于在第一引导RNA识别序列与第二引导RNA识别序列或第一裂解位点与第二裂解位点之间的基因组区域的5' 和3' 序列，并且相对于靶基因组基因座的其他区段，诸如侧接于在靶基因组基因座内的一对或多对其他潜在引导RNA识别序列或裂解位点之间的基因组区域的5' 和3' 序列，靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。如果使用三个或更多个引导RNA，那么相关区域将是侧接于在相隔最远的两个引导RNA识别序列或两个裂解位点之间的基因组区域的5' 和/或3' 序列。5' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样，3' 侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0210] 因此，在其中使用两个引导RNA的方法中，举例来说，选择靶区域可包括比较靶基因组基因座的两个或更多个非连续区段，其中各非连续区段包含以下、基本上由以下组成或由以下组成：在一对不同引导RNA识别序列之间的基因组区域的5' 侧、3' 侧或各侧上的至少10bp、20bp、30bp、40bp、50bp、100bp、200bp、300bp、400bp、500bp、600bp、700bp、800bp、900bp、1,000bp、1kb、2kb、3kb、4kb、5kb、6、kb、7kb、8kb、9kb、10kb、20kb、30kb、40kb、50kb、60kb、70kb、80kb、90kb、100kb、110kb、120kb、130kb、140kb或150kb的侧接序列，其中所述引导RNA识别序列不存在于基因组中其他地方，以及将相对于其他非连续区段具有最高序列同一性百分比的非连续区段选作靶区域。任选地，一个或多个非连续区段包含以下、基本上由以下组成或由以下组成：与靶基因组基因座中各对不同引导RNA识别序列相对应的非连续区段，其中所述引导RNA识别序列不存在于基因组中其他地方。

[0211] 在其中使用外源性修复模板的方法中，靶区域可包含以下、基本上由以下组成或由以下组成：在5' 靶标序列与3' 靶标序列之间的区域，并且相对于靶基因组基因座的其他区段，靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。或者或另外，靶区域可包含以下、基本上由以下组成或由以下组成：5' 靶标序列和/或3' 靶标序列，并且相对于靶基因组基因座的其他区段，靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。优选地，靶区域可包含以下、基本上由以下组成或由以下组成：5' 靶标序列和3' 靶标序列，并且相对于靶基因组基因座的其他区段，靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。举例来说，靶区域可包含以下、基本上由以下组成或由以下组成：由5' 靶标序列和3' 靶标序列侧接且包括所述5' 靶标序列和所述3' 靶标序列的区域，并且相对于靶基因组基因座的其他区段，靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。

[0212] 同样，在其中使用外源性修复模板的方法中，靶区域可包含以下、基本上由以下组成或由以下组成：侧接于在所述外源性修复模板的5' 靶标序列与3' 靶标序列之间的基因组区域的5' 和/或3' 序列或侧接于在所述外源性修复模板的5' 靶标序列与3' 靶标序列之间的

基因组区域且包括所述5'靶标序列和所述3'靶标序列的5'和/或3'序列,并且相对于靶基因组基因座的其他区段,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。优选地,靶区域可包含以下、基本上由以下组成或由以下组成:侧接于在外源性修复模板的5'靶标序列与3'靶标序列之间的基因组区域的5'和3'序列或侧接于在外源性修复模板的5'靶标序列与3'靶标序列之间的基因组区域且包括所述5'靶标序列和所述3'靶标序列的5'和3'序列,并且相对于靶基因组基因座的其他区段,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。或者,靶区域可包含以下、基本上由以下组成或由以下组成:在外源性修复模板的5'靶标序列与3'靶标序列之间的区域和侧接于在所述5'靶标序列与所述3'靶标序列之间的基因组区域的5'和/或3'序列,并且相对于靶基因组基因座的其他区段,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。优选地,靶区域可包含以下、基本上由以下组成或由以下组成:在外源性修复模板的5'靶标序列与3'靶标序列之间的区域和侧接于在所述5'靶标序列与所述3'靶标序列之间的基因组区域的5'和3'序列,并且相对于靶基因组基因座的其他区段,靶区域可具有高序列同一性百分比或低密度的单核苷酸变化。5'侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。同样,3'侧接序列可为例如至少10、20、30、40、50、100、200、300、400、500、600、700、800、900或1,000bp的侧接序列或至少1、2、3、4、5、6、7、8、9、10、20、30、40、50、60、70、80、90、100、110、120、130、140或150kb的侧接序列。

[0213] 通过本文公开的方法修饰的靶区域可包括细胞内的任何DNA区段或区域(连续或非连续)。靶区域可为细胞所天然具有,可为整合至细胞的基因组中的异源性或外源性DNA区段,或可为其组合。所述异源性或外源性DNA区段可包括转基因、表达盒、编码选择标记的多核苷酸、或基因组DNA的异源性或外源性区域。

#### [0214] B. CRISPR/Cas系统

[0215] 本文公开的方法利用成簇规律散布短回文重复序列(CRISPR)/CRISPR相关蛋白(Cas)系统或所述系统的组分来修饰细胞内的基因组。CRISPR/Cas系统包括Cas基因的表达中涉及或指导Cas基因的活性的转录物和其他要素。CRISPR/Cas系统可为I型、II型或III型系统。或者,CRISPR/Cas系统可为例如V型系统(例如V-A亚型或V-B亚型)。本文公开的方法和组合物通过利用CRISPR复合物(包含引导RNA(gRNA)与Cas蛋白复合)对核酸进行定点裂解来采用CRISPR/Cas系统。

[0216] 用于本文公开的方法中的CRISPR/Cas系统是非天然存在的。“非天然存在”系统包括指示涉及人工的任何事物,诸如系统的一个或多个组分被从它们的天然存在状态改变或突变,至少大致上不含至少一种它们在自然界中与其天然相伴的其他组分,或与至少一种它们不天然与其相伴的其他组分相伴。举例来说,一些CRISPR/Cas系统采用非天然存在的CRISPR复合物,其包含不天然存在在一起的gRNA和Cas蛋白。其他CRISPR/Cas系统采用不天然存在的Cas蛋白,并且其他CRISPR/Cas系统采用不天然存在的gRNA。

#### [0217] (1) Cas蛋白

[0218] Cas蛋白通常包含至少一个可与引导RNA(gRNA,以下更详细描述)相互作用的RNA识别或结合结构域。Cas蛋白也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、蛋白质-蛋白质相互作用结构域、二聚化结构域和其他结构域。核酸

酶结构域具有用于核酸裂解的催化活性,其包括使核酸分子的共价键断裂。裂解可产生钝性末端或交错末端,并且这可为单链或双链。举例来说,野生型Cas9蛋白将通常产生钝性裂解产物。或者,野生型Cpf1蛋白(例如FnCpf1)可产生具有5核苷酸5'突出部分的裂解产物,其中裂解在非靶向链上发生在从PAM序列开始第18碱基对之后,而在靶向链上发生在第23碱基之后。Cas蛋白可具有完全裂解活性以在靶标核酸中产生双链断裂(例如具有钝性末端的双链断裂),或它可为在靶标核酸中产生单链断裂的切口酶。

[0219] Cas蛋白的实例包括Cas1、Cas1B、Cas2、Cas3、Cas4、Cas5、Cas5e (CasD)、Cas6、Cas6e、Cas6f、Cas7、Cas8a1、Cas8a2、Cas8b、Cas8c、Cas9 (Csn1或Csx12)、Cas10、Cas10d、CasF、CasG、CasH、Csy1、Csy2、Csy3、Cse1 (CasA)、Cse2 (CasB)、Cse3 (CasE)、Cse4 (CasC)、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csx1、Csx15、Csf1、Csf2、Csf3、Csf4和Cu1966及其同源物或经修饰形式。

[0220] 一示例性Cas蛋白是来自II型CRISPR/Cas系统的Cas9蛋白或源于Cas9蛋白的蛋白质。Cas9蛋白来自II型CRISPR/Cas系统,并且通常共有四个具有保守构造的关键基序。基序1、2和4是RuvC样基序,并且基序3是HNH基序。示例性Cas9蛋白来自化脓性链球菌(*Streptococcus pyogenes*)、嗜热链球菌(*Streptococcus thermophilus*)、链球菌属某种(*Streptococcus sp.*)、金黄色葡萄球菌(*Staphylococcus aureus*)、达松维尔拟诺卡氏菌(*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、玫瑰链孢囊菌(*Streptosporangium roseum*)、玫瑰链孢囊菌、酸热脂环芽孢杆菌(*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*)、戴耳布吕克氏乳酸杆菌(*Lactobacillus delbrueckii*)、唾液乳酸杆菌(*Lactobacillus salivarius*)、海洋微颤菌(*Microscilla marina*)、伯克氏菌目细菌(*Burkholderiales bacterium*)、萘降解极地单胞菌(*Polaromonas naphthalenivorans*)、极地单胞菌某种(*Polaromonas sp.*)、瓦氏鳄球藻(*Crocospaera watsonii*)、蓝丝菌属某种(*Cyanothece sp.*)、铜绿微囊藻(*Microcystis aeruginosa*)、聚球藻属某种(*Synechococcus sp.*)、阿拉伯糖醋盐杆菌(*Acetohalobium arabaticum*)、丹氏制氮菌(*Ammonifex degensii*)、毕氏热解纤维素菌(*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*)、热丙酸盐暗色厌氧香肠状菌(*Pelotomaculum thermopropionicum*)、喜温嗜酸硫杆菌(*Acidithiobacillus caldus*)、氧化亚铁嗜酸硫杆菌(*Acidithiobacillus ferrooxidans*)、酒色别样着色菌(*Allochromatium vinosum*)、海杆菌属某种(*Marinobacter sp.*)、嗜盐亚硝化球菌(*Nitrosococcus halophilus*)、沃特森亚硝化球菌(*Nitrosococcus watsoni*)、游海假交替单胞菌(*Pseudoalteromonas haloplanktis*)、成簇细枝菌(*Ktedonobacter racemifer*)、调查甲烷盐菌(*Methanohalobium evestigatum*)、多变鱼腥藻(*Anabaena variabilis*)、泡沫节球藻(*Nodularia spumigena*)、念珠藻属某种(*Nostoc sp.*)、极大节旋藻(*Arthrospira maxima*)、钝顶节旋藻(*Arthrospira platensis*)、节旋藻属某种(*Arthrospira sp.*)、鞘丝

藻属某种 (*Lyngbya* sp.)、喜泥微鞘藻 (*Microcoleus chthonoplastes*)、颤藻属某种 (*Oscillatoria* sp.)、运动石袍菌 (*Petrotoga mobilis*)、非洲栖热腔菌 (*Thermosipho africanus*)、*Acaryochloris marina*、脑膜炎奈瑟氏菌 (*Neisseria meningitidis*) 或空肠弯曲杆菌 (*Campylobacter jejuni*)。Cas9 家族成员的额外实例描述于出于所有目的以引用的方式整体并入本文的 W0 2014/131833 中。来自化脓性链球菌的 Cas9 (SpCas9) (指定 SwissProt 登录号 Q99ZW2) 是一示例性 Cas9 蛋白。来自金黄色葡萄球菌的 Cas9 (Sa Cas9) (指定 UniProt 登录号 J7RUA5) 是另一示例性 Cas9 蛋白。来自空肠弯曲杆菌的 Cas9 (CjCas9) (指定 UniProt 登录号 Q0P897) 是另一示例性 Cas9 蛋白。参见例如 Kim 等 (2017) *Nat. Comm.* 8: 14500, 其出于所有目的以引用的方式整体并入本文。SaCas9 小于 SpCas9, 并且 CjCas9 小于 SaCas9 与 SpCas9 两者。

[0221] Cas 蛋白的另一实例是 Cpf1 (来自普雷沃菌属 (*Prevotella*) 和弗朗西斯氏菌属 (*Francisella*) 的 CRISPR 1) 蛋白。Cpf1 是一种大型蛋白质 (约 1300 个氨基酸), 其含有与 Cas9 的相应结构域同源的 RuvC 样核酸酶结构域以及 Cas9 的特征性精氨酸富集簇的对应物。然而, Cpf1 缺乏存在于 Cas9 蛋白中的 HNH 核酸酶结构域, 并且 RuvC 样结构域在 Cpf1 序列中是连续的, 这不同于 Cas9, 在 Cas9 中, 它含有包括 HNH 结构域的长插入物。参见例如 Zetsche 等 (2015) *Cell* 163 (3): 759-771, 其出于所有目的以引用的方式整体并入本文。示例性 Cpf1 蛋白来自野兔热弗朗西斯氏菌 1 (*Francisella tularensis* 1)、野兔热弗朗西斯氏菌新凶手亚种 (*Francisella tularensis* subsp. *novicida*)、苏格兰普雷沃菌 (*Prevotella albensis*)、毛螺菌科细菌 MC2017 1 (*Lachnospiraceae* bacterium MC2017 1)、瘤胃溶纤维丁酸弧菌 (*Butyrivibrio proteoclasticus*)、异域菌门细菌 GW2011\_GWA2\_33\_10 (*Peregrinibacteria* bacterium GW2011\_GWA2\_33\_10)、俭菌总门细菌 GW2011\_GWC2\_44\_17 (*Parcubacteria* bacterium GW2011\_GWC2\_44\_17)、史密斯氏菌属某种 SCADC (*Smithella* sp. SCADC)、氨基酸球菌属某种 BV3L6 (*Acidaminococcus* sp. BV3L6)、毛螺菌科细菌 MA2020 (*Lachnospiraceae* bacterium MA2020)、*Candidatus Methanoplasma termitum*、挑剔真杆菌 (*Eubacterium eligens*)、牛眼莫拉菌 237 (*Moraxella bovoculi* 237)、稻田氏钩端螺旋体 (*Leptospira inadai*)、毛螺菌科细菌 ND2006、狗口腔卟啉单胞菌 3 (*Porphyromonas crevioricanis* 3)、解糖豚普雷沃菌 (*Prevotella disiens*) 和猕猴卟啉单胞菌 (*Porphyromonas macacae*)。来自新凶手弗朗西斯氏菌 U112 (*Francisella novicida* U112) 的 Cpf1 (FnCpf1; 指定 UniProt 登录号 A0Q7Q2) 是一示例性 Cpf1 蛋白。

[0222] Cas 蛋白可为野生型蛋白质 (即存在于自然界中的那些)、经修饰 Cas 蛋白 (即 Cas 蛋白变体)、或野生型或经修饰 Cas 蛋白的片段。Cas 蛋白也可为关于野生型或经修饰 Cas 蛋白的催化活性的活性变体或片段。关于催化活性的活性变体或片段可包含与野生型或经修饰 Cas 蛋白或其一部分的至少 80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99% 或更大的序列同一性, 其中活性变体保留在所需裂解位点处切割的能力, 因此保留切口诱导或双链断裂诱导活性。针对切口诱导或双链断裂诱导活性的测定是已知的, 并且通常测量 Cas 蛋白对含有裂解位点的 DNA 底物的总体活性和特异性。

[0223] 经修饰 Cas 蛋白的一个实例是经修饰 SpCas9-HF1 蛋白, 其是化脓性链球菌 Cas9 的具有被设计来使非特异性 DNA 接触降低的改变 (N497A/R661A/Q695A/Q926A) 的高保真变体。参见例如 Kleinstiver 等 (2016) *Nature* 529 (7587): 490-495, 其出于所有目的以引用的方

式整体并入本文。经修饰Cas蛋白的另一实例是被设计来使脱靶效应降低的经修饰eSpCas9变体(K848A/K1003A/R1060A)。参见例如Slaymaker等(2016) *Science* 351(6268):84-88,其出于所有目的以引用的方式整体并入本文。其他SpCas9变体包括K855A和K810A/K1003A/R1060A。

[0224] Cas蛋白可被修饰以使核酸结合亲和力、核酸结合特异性和酶活性中的一者或者增加或降低。Cas蛋白也可被修饰以改变蛋白质的任何其他活性或性质,诸如稳定性。举例来说,Cas蛋白的一个或多个核酸酶结构域可被修饰、缺失或失活,或Cas蛋白可被截短以移除不为蛋白质的功能所必需的结构域或使Cas蛋白的活性最优化(例如增强或降低)。

[0225] Cas蛋白可包含至少一个核酸酶结构域,诸如DNA酶结构域。举例来说,野生型Cpf1蛋白通常包含会裂解靶标DNA的两个链,或许呈二聚构型的RuvC样结构域。Cas蛋白也可包含至少两个核酸酶结构域,诸如DNA酶结构域。举例来说,野生型Cas9蛋白通常包含RuvC样核酸酶结构域和HNH样核酸酶结构域。RuvC结构域和HNH结构域可各自切割双链DNA的不同链以在DNA中产生双链断裂。参见例如Jinek等(2012) *Science* 337:816-821,其出于所有目的以引用的方式整体并入本文。

[0226] 一个或两个核酸酶结构域可被缺失或突变以使它们不再具有功能性或具有降低的核酸酶活性。如果一个核酸酶结构域被缺失或突变,那么所得Cas蛋白(例如Cas9)可被称为切口酶,并且可在双链DNA内在引导RNA识别序列处产生单链断裂而非双链断裂(即它可裂解互补链或非互补链,而非两者)。如果两个核酸酶结构域均被缺失或突变,那么所得Cas蛋白(例如Cas9)裂解双链DNA的两个链的能力将降低(例如核酸酶无效Cas蛋白)。使Cas9转变成切口酶的突变的一实例是来自化脓性链球菌的Cas9的RuvC结构域中的D10A(在Cas9的位置10处,天冬氨酸变为丙氨酸)突变。同样,来自化脓性链球菌的Cas9的HNH结构域中的H939A(在氨基酸位置839处,组氨酸变为丙氨酸)或H840A(在氨基酸位置840处,组氨酸变为丙氨酸)或N863A(在氨基酸位置N863处,天冬酰胺变为丙氨酸)可使Cas9转变成切口酶。使Cas9转变成切口酶的其他实例包括来自嗜热链球菌(*S. thermophilus*)的Cas9的相应突变。参见例如Saprunauskas等(2011) *Nucleic Acids Research* 39:9275-9282和WO 2013/141680,其各自出于所有目的以引用的方式整体并入本文。所述突变可使用诸如定点诱变、PCR介导的诱变或全基因合成的方法来产生。产生切口酶的其他突变的实例可例如见于WO 2013/176772和WO 2013/142578中,所述专利各自出于所有目的以引用的方式整体并入本文。如果Cas蛋白中的所有核酸酶结构域都被缺失或突变(例如Cas9蛋白中的两个核酸酶结构域均被缺失或突变),那么所得Cas蛋白(例如Cas9)裂解双链DNA的两个链的能力将降低(例如核酸酶无效或核酸酶非活性Cas蛋白)。一个特定实例是D10A/H840A化脓性链球菌Cas9双重突变体,或在来自另一物种的Cas9的情况下在与化脓性链球菌Cas9最优比对时的相应双重突变体。另一特定实例是D10A/N863A化脓性链球菌Cas9双重突变体,或在来自另一物种的Cas9的情况下在与化脓性链球菌Cas9最优比对时的相应双重突变体。

[0227] 金黄色葡萄球菌Cas9蛋白的催化结构域中的失活性突变的实例也是已知的。举例来说,金黄色葡萄球菌Cas9酶(SaCas9)可包含在位置N580处的取代(例如N580A取代)和在位置D10处的取代(例如D10A取代)以产生核酸酶非活性Cas蛋白。参见例如WO 2016/106236,其出于所有目的以引用的方式整体并入本文。

[0228] Cpf1蛋白的催化结构域中的失活性突变的实例也是已知的。关于来自新凶手弗朗

西斯氏菌U112 (FnCpf1)、氨基酸球菌属某种BV3L6 (AsCpf1)、毛螺菌科细菌ND2006 (LbCpf1) 和牛眼莫拉菌237 (MbCpf1 Cpf1) 的Cpf1蛋白,所述突变可包括在AsCpf1的位置908、993或1263或Cpf1直系同源物中的相应位置、或LbCpf1的位置832、925、947或1180或Cpf1直系同源物中的相应位置处的突变。所述突变可包括例如AsCpf1的突变D908A、E993A和D1263A或Cpf1直系同源物中的相应突变、或LbCpf1的突变D832A、E925A、D947A和D1180A或Cpf1直系同源物中的相应突变中的一者或多者。参见例如US 2016/0208243,其出于所有目的以引用的方式整体并入本文。

[0229] Cas蛋白也可被可操作地连接至异源性多肽,呈融合蛋白形式。举例来说,可使Cas蛋白融合于裂解结构域、表观遗传修饰结构域、转录活化结构域或转录阻遏物结构域。参见WO 2014/089290,其出于所有目的以引用的方式整体并入本文。也可使Cas蛋白融合于异源性多肽,从而提供增加或降低的稳定性。融合结构域或异源性多肽可位于N末端、C末端或Cas蛋白内部。

[0230] Cas融合蛋白的一实例是融合于提供亚细胞定位的异源性多肽的Cas蛋白。所述异源性多肽可包括例如用于靶向核的一个或多个核定位信号(NLS) 诸如SV40 NLS、用于靶向线粒体的线粒体定位信号、ER保留信号等。参见例如Lange等(2007) J. Biol. Chem. 282: 5101-5105,其出于所有目的以引用的方式整体并入本文。其他适合NLS包括 $\alpha$ -输入蛋白NLS。所述亚细胞定位信号可位于N末端、C末端或Cas蛋白内的任何地方。NLS可包含碱性氨基酸链段,并且可为单分型序列或双分型序列。任选地,Cas蛋白包含两个或更多个NLS,包括在N末端的NLS (例如 $\alpha$ -输入蛋白NLS) 和/或在C末端的NLS (例如SV40 NLS)。

[0231] 也可使Cas蛋白可操作地连接至细胞穿透结构域。举例来说,细胞穿透结构域可源于HIV-1 TAT蛋白、来自人乙型肝炎病毒的TLM细胞穿透基序、MPG、Pep-1、VP22、来自单纯疱疹病毒的细胞穿透肽或聚精氨酸肽序列。参见例如WO 2014/089290,其出于所有目的以引用的方式整体并入本文。细胞穿透结构域可位于N末端、C末端或Cas蛋白内的任何地方。

[0232] 也可使Cas蛋白可操作地连接至异源性多肽以便追踪或纯化,诸如荧光蛋白、纯化标签或表位标签。荧光蛋白的实例包括绿色荧光蛋白 (例如GFP、GFP-2、tagGFP、turboGFP、eGFP、Emerald、Azami Green、Monomeric Azami Green、CopGFP、AceGFP、ZsGreen1)、黄色荧光蛋白 (例如YFP、eYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1)、蓝色荧光蛋白 (例如eBFP、eBFP2、Azurite、mKalamal、GFPuv、Sapphire、T-sapphire)、青色荧光蛋白 (例如eCFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan)、红色荧光蛋白 (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) 和任何其他适合荧光蛋白。标签的实例包括谷胱甘肽-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)、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) 和钙调蛋白 (calmodulin)。

[0233] 也可使Cas9蛋白栓系于外源性修复模板或经标记核酸。所述栓系 (即物理连接) 可通过共价相互作用或非共价相互作用来实现,并且栓系可为直接的 (例如通过直接融合或化学缀合,此可通过对蛋白质上半胱氨酸或赖氨酸残基的修饰或内含肽修饰来实现),或可

通过一个或多个间插接头或衔接物分子诸如链霉亲和素或适体来实现。参见例如Pierce等(2005) *Mini Rev. Med. Chem.* 5 (1) : 41-55; Duckworth等(2007) *Angew. Chem. Int. Ed. Engl.* 46 (46) : 8819-8822; Schaeffer和Dixon(2009) *Australian J. Chem.* 62 (10) : 1328-1332; Goodman等(2009) *Chembiochem.* 10 (9) : 1551-1557; 以及Khatwani等(2012) *Bioorg. Med. Chem.* 20 (14) : 4532-4539, 其各自出于所有目的以引用的方式整体并入本文。用于合成蛋白质-核酸缀合物的非共价策略包括生物素-链霉亲和素和镍-组氨酸方法。共价蛋白质-核酸缀合物可通过使用广泛多种化学过程使适当官能化的核酸和蛋白质连接来合成。这些化学过程中的一些涉及使寡核苷酸直接连接至蛋白质表面上的氨基酸残基(例如赖氨酸胺或半胱氨酸硫醇), 而其他更复杂流程需要蛋白质的翻译后修饰或涉及催化性或反应性蛋白质结构域。用于使蛋白质共价连接至核酸的方法可包括例如使寡核苷酸化学交联于蛋白质赖氨酸或半胱氨酸残基、表达蛋白连接、化学酶法和使用光适体。可使外源性修复模板或经标记核酸栓系于C末端、N末端或Cas9蛋白内的内部区域。优选地, 使外源性修复模板或经标记核酸栓系于Cas9蛋白的C末端或N末端。同样, 可使Cas9蛋白栓系于5'末端、3'末端或外源性修复模板或经标记核酸内的内部区域。也就是说, 外源性修复模板或经标记核酸可以任何定向和极性加以栓系。优选地, 使Cas9蛋白栓系于外源性修复模板或经标记核酸的5'末端或3'末端。

[0234] Cas蛋白可以任何形式提供。举例来说, Cas蛋白可以蛋白质形式诸如与gRNA复合的Cas蛋白提供。或者, Cas蛋白可以编码所述Cas蛋白的核酸形式诸如RNA(例如信使RNA(mRNA))或DNA提供。任选地, 编码Cas蛋白的核酸可经密码子优化以达成在特定细胞或生物体中高效翻译成蛋白质。举例来说, 编码Cas蛋白的核酸可被修饰以用相较于天然存在的多核苷酸序列, 在细菌细胞、酵母细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其他目标宿主细胞中具有更高使用频率的密码子进行替代。当将编码Cas蛋白的核酸引入细胞中时, Cas蛋白可在细胞中短暂、条件性或组成性表达。

[0235] 编码Cas蛋白的核酸可被稳定整合在细胞基因组中, 并且可操作地连接至在细胞中具有活性的启动子。或者, 编码Cas蛋白的核酸可被可操作地连接至表达构建体中的启动子。表达构建体包括能够指导基因或其他目标核酸序列(例如Cas基因)的表达, 以及可将这种目标核酸序列转移至靶标细胞中的任何核酸构建体。举例来说, 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, 它可在不同于包含核酸插入物的靶向载体和/或不同于包含编码gRNA的DNA的载体的载体或质粒中。可用于表达构建体中的启动子包括例如在真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞、兔细胞、多能细胞、胚胎干(ES)细胞或接合子中的一者或多者中具有活性的启动子。所述启动子可为例如条件型启动子、诱导型启动子、组成型启动子或组织特异性启动子。任选地, 启动子可为在一个方向上驱动Cas蛋白的表达, 并且在另一方向上驱动引导RNA的表达的双向启动子。所述双向启动子可由以下组成: (1) 完整常规单向Pol III启动子, 其含有3个外部控制元件: 远端序列元件(DSE)、近端序列元件(PSE)和TATA框; 和(2) 包括PSE和TATA框的第二基本Pol III启动子, 其以相反定向融合于DSE的5'末端。举例来说, 在H1启动子中, DSE邻近于PSE和TATA框, 并且可通过产生杂合启动子来致使启动子具有双向性, 其中通过附接源于U6启动子的PSE和TATA框来控制反向上的转录。参见例如US 2016/0074535, 其出于所有目的以引用的方式整体并入

本文。使用双向启动子来同时表达编码Cas蛋白的基因和编码引导RNA的基因允许产生紧凑表达盒以有助于进行递送。

[0236] (2) 引导RNA

[0237] “引导RNA”或“gRNA”是结合Cas蛋白(例如Cas9蛋白),并且使所述Cas蛋白靶向靶标DNA内的特定位置的RNA分子。引导RNA可包含两个区段:“DNA靶向区段”和“蛋白质结合区段”。“区段”包括分子的节段或区域,诸如RNA中的连续核苷酸链段。一些gRNA,诸如针对Cas9的那些,可包含两个单独RNA分子:“活化RNA”(例如tracrRNA)和“靶向RNA”(例如CRISPR RNA或crRNA)。其他gRNA是单一RNA分子(单一RNA多核苷酸),其也可称为“单分子gRNA”、“单引导RNA”或“sgRNA”。参见例如WO 2013/176772、WO 2014/065596、WO 2014/089290、WO 2014/093622、WO 2014/099750、WO 2013/142578和WO 2014/131833,其各自出于所有目的以引用的方式整体并入本文。对于Cas9,举例来说,单引导RNA可包含crRNA融合于tracrRNA(例如通过接头)。对于Cpf1,举例来说,仅crRNA为实现与靶标序列的结合或裂解所需。术语“引导RNA”和“gRNA”包括双分子gRNA(即模块gRNA)与单分子gRNA两者。

[0238] 一示例性双分子gRNA包含crRNA样(“CRISPR RNA”或“靶向RNA”或“crRNA”或“crRNA重复”)分子和相应tracrRNA样(“反式作用性CRISPR RNA”或“活化RNA”或“tracrRNA”)分子。crRNA包含gRNA的DNA靶向区段(单链)与形成gRNA的蛋白质结合区段的dsRNA双链体的一半的核苷酸链段两者。

[0239] 相应tracrRNA(活化RNA)包含形成gRNA的蛋白质结合区段的dsRNA双链体的另一半的核苷酸链段。crRNA的核苷酸链段与tracrRNA的核苷酸链段互补和杂交以形成gRNA的蛋白质结合结构域的dsRNA双链体。因此,各crRNA可被称为具有相应tracrRNA。

[0240] 在其中需要crRNA与tracrRNA两者的系统中,crRNA和相应tracrRNA杂交以形成gRNA。在其中仅需要crRNA的系统中,crRNA可为gRNA。crRNA另外提供杂交至引导RNA识别序列的单链DNA靶向区段。如果用于在细胞内的修饰,那么给定crRNA或tracrRNA分子的确切序列可被设计来对其中将使用RNA分子的物种来说具有特异性。参见例如Mali等(2013) Science 339:823-826;Jinek等(2012) Science 337:816-821;Hwang等(2013) Nat.Biotechnol.31:227-229;Jiang等(2013) Nat.Biotechnol.31:233-239;以及Cong等(2013) Science 339:819-823,其各自出于所有目的以引用的方式整体并入本文。

[0241] 给定gRNA的DNA靶向区段(crRNA)包含互补于靶标DNA中的序列(即引导RNA识别序列)的核苷酸序列。gRNA的DNA靶向区段与靶标DNA以序列特异性方式通过杂交(即碱基配对)来相互作用。因此,DNA靶向区段的核苷酸序列可变化,并且决定靶标DNA内gRNA和靶标DNA将进行相互作用的位置。主题gRNA的DNA靶向区段可被修饰以杂交于靶标DNA内的任何所需序列。天然存在的crRNA视CRISPR/Cas系统和生物体而不同,但常常含有长度在21至72个核苷酸之间的靶向区段,由长度在21至46个核苷酸之间的两个同向重复序列(DR)侧接(参见例如WO 2014/131833,其出于所有目的以引用的方式整体并入本文)。在化脓性链球菌的情况下,DR的长度是36个核苷酸,并且靶向区段的长度是30个核苷酸。位于3'的DR与相应tracrRNA互补和杂交,所述tracrRNA转而结合Cas蛋白。

[0242] DNA靶向区段可具有至少约12个核苷酸、至少约15个核苷酸、至少约17个核苷酸、至少约18个核苷酸、至少约19个核苷酸、至少约20个核苷酸、至少约25个核苷酸、至少约30个核苷酸、至少约35个核苷酸或至少约40个核苷酸的长度。所述DNA靶向区段可具有约12个

核苷酸至约100个核苷酸、约12个核苷酸至约80个核苷酸、约12个核苷酸至约50个核苷酸、约12个核苷酸至约40个核苷酸、约12个核苷酸至约30个核苷酸、约12个核苷酸至约25个核苷酸、或约12个核苷酸至约20个核苷酸的长度。举例来说，DNA靶向区段可为约15个核苷酸至约25个核苷酸（例如约17个核苷酸至约20个核苷酸、或约17个核苷酸、约18个核苷酸、约19个核苷酸或约20个核苷酸）。参见例如US 2016/0024523，其出于所有目的以引用的方式整体并入本文。对于来自化脓性链球菌的Cas9，典型DNA靶向区段的长度在16与20个核苷酸之间，或长度在17与20个核苷酸之间。对于来自金黄色葡萄球菌的Cas9，典型DNA靶向区段的长度在21与23个核苷酸之间。对于Cpf1，典型DNA靶向区段的长度是至少16个核苷酸，或长度是至少18个核苷酸。

[0243] TracrRNA可呈任何形式（例如全长tracrRNA或活性部分tracrRNA），并且具有不同长度。它们可包括初级转录物或经加工形式。举例来说，tracrRNA（作为单引导RNA的一部分，或呈作为双分子gRNA的一部分的单独分子形式）可包含以下或由以下组成：野生型tracrRNA序列的全部或一部分（例如野生型tracrRNA序列的约或超过约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序列的实例包括171核苷酸、89核苷酸、75核苷酸和65核苷酸形式。参见例如Deltcheva等（2011）Nature 471:602-607；WO 2014/093661，其各自出于所有目的以引用的方式整体并入本文。单引导RNA（sgRNA）内的tracrRNA的实例包括见于sgRNA的+48、+54、+67和+85形式内的tracrRNA区段，其中“+n”指示直至野生型tracrRNA的+n核苷酸被包括在sgRNA中。参见US 8,697,359，其出于所有目的以引用的方式整体并入本文。

[0244] DNA靶向序列与靶标DNA内的引导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%）。历经约20个连续核苷酸，DNA靶向序列与靶标DNA内的引导RNA识别序列之间的互补性百分比可为至少60%。作为一实例，在靶标DNA的互补链内的引导RNA识别序列的5'末端，历经14个连续核苷酸，DNA靶向序列与靶标DNA内的引导RNA识别序列之间的互补性百分比是100%，并且历经其余部分，互补性百分比低至0%。在这种情况下，DNA靶向序列的长度可被视为是14个核苷酸。作为另一实例，在靶标DNA的互补链内的引导RNA识别序列的5'末端，历经7个连续核苷酸，DNA靶向序列与靶标DNA内的引导RNA识别序列之间的互补性百分比是100%，并且历经其余部分，互补性百分比低至0%。在这种情况下，DNA靶向序列的长度可被视为是7个核苷酸。在一些引导RNA中，DNA靶向序列内的至少17个核苷酸互补于靶标DNA。举例来说，DNA靶向序列的长度可为20个核苷酸，并且可包含1、2或3个与靶标DNA（引导RNA识别序列）的不匹配。优选地，不匹配不邻近于原间隔体邻近基序（PAM）序列（例如不匹配在DNA靶向序列的5'末端，或不匹配离PAM序列至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18或19个碱基对）。

[0245] gRNA的蛋白质结合区段可包含彼此互补的两个核苷酸链段。蛋白质结合区段的互补性核苷酸杂交以形成双链RNA双链体（dsRNA）。主题gRNA的蛋白质结合区段与Cas蛋白相互作用，并且gRNA通过DNA靶向区段来将结合的Cas蛋白引导至靶标DNA内的特定核苷酸序列。

[0246] 单引导RNA具有DNA靶向区段和骨架序列（即引导RNA的蛋白质结合或Cas结合序列）。示例性骨架序列包括：GTTGGAACCATTCAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGTTATC

AACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO:150); GTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO:151); 和 GTTTAAGAGCTAT GCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO:152)。

[0247] 引导RNA可包括提供额外合乎需要特征(例如经改进或经调控稳定性;亚细胞靶向;用荧光标记进行追踪;蛋白质或蛋白质复合物的结合位点等)的修饰或序列。所述修饰的实例包括例如5'帽(例如7-甲基鸟苷酸帽(m7G));3'聚腺苷酸化尾部(即3'聚腺苷酸(poly(A))尾部);核糖开关序列(例如以允许达成经调控稳定性和/或经调控蛋白质和/或蛋白质复合物可及性);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(例如核、线粒体、叶绿体等)的修饰或序列;提供追踪的修饰或序列(例如直接缀合于荧光分子,缀合于有助于荧光检测的部分,允许达成荧光检测的序列,等);提供蛋白质(例如作用于DNA的蛋白质,包括转录活化子、转录阻遏物、DNA甲基转移酶、DNA脱甲基酶、组蛋白乙酰基转移酶、组蛋白脱乙酰基酶等)的结合位点的修饰或序列;及其组合。修饰的其他实例包括工程化的茎环双链体结构、工程化的突起区域、在茎环双链体结构的3'的工程化发夹或其任何组合。参见例如US 2015/0376586,其出于所有目的以引用的方式整体并入本文。突起可为由crRNA样区域和最小tracrRNA样区域组成的双链体内的未配对核苷酸区域。突起可在双链体的一侧包含未配对5'-XXX-3',其中X是任何嘌呤,并且Y可为可与相对链上的核苷酸形成摇摆配对的核苷酸,并且突起可在双链体的另一侧包含未配对核苷酸区域。

[0248] 引导RNA可以任何形式提供。举例来说,gRNA可以呈两个分子形式(单独crRNA和tracrRNA)或呈一个分子形式(sgRNA)的RNA形式提供,并且任选地与Cas蛋白的复合物形式提供。举例来说,gRNA可通过使用例如T7RNA聚合酶进行体外转录来制备(参见例如WO 2014/089290和WO 2014/065596,其各自出于所有目的以引用的方式整体并入本文)。引导RNA也可通过化学合成制备。

[0249] gRNA也可以编码gRNA的DNA形式提供。编码gRNA的DNA可编码单一RNA分子(sgRNA)或单独RNA分子(例如单独crRNA和tracrRNA)。在后述情况下,编码gRNA的DNA可以一个DNA分子形式或以分别编码crRNA和tracrRNA的单独DNA分子形式提供。

[0250] 当gRNA以DNA形式提供时,所述gRNA可在细胞中短暂、条件性或组成性表达。编码gRNA的DNA可被稳定整合至细胞基因组中,并且可操作地连接至在细胞中具有活性的启动子。或者,编码gRNA的DNA可被可操作地连接至表达构建体中的启动子。举例来说,编码gRNA的DNA可在包含外源性修复模板的载体和/或包含编码Cas蛋白的核酸的载体中。或者,它可在不同于包含外源性修复模板的载体和/或包含编码Cas蛋白的核酸的载体的载体或质粒中。可用于所述表达构建体中的启动子包括例如在真核细胞、人细胞、非人细胞、哺乳动物细胞、非人哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞、仓鼠细胞、兔细胞、多能细胞、胚胎干(ES)细胞或接合子中的一者或多者中具有活性的启动子。所述启动子可为例如条件型启动子、诱导型启动子、组成型启动子或组织特异性启动子。所述启动子也可例如双向启动子。适合启动子的特定实例包括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。

[0251] (3) 引导RNA识别序列

[0252] 术语“引导RNA识别序列”包括存在于靶标DNA中的由gRNA的DNA靶向区段将与其结合的核酸序列,前提是存在足以达成结合的条件。举例来说,引导RNA识别序列包括引导RNA被设计来与其具有互补性的序列,其中引导RNA识别序列与DNA靶向序列之间的杂交促进形成CRISPR复合物。未必需要完全互补性,前提是存在足以导致杂交和促进形成CRISPR复合物的互补性。引导RNA识别序列也包括Cas蛋白的裂解位点,以下更详细描述。引导RNA识别序列可包含任何多核苷酸,其可位于例如细胞的核或细胞质中或细胞的细胞器诸如线粒体或叶绿体内。

[0253] 靶标DNA内的引导RNA识别序列可由Cas蛋白或gRNA靶向(即由其结合,与其杂交,或与其互补)。适合DNA/RNA结合条件包括通常存在于细胞中的生理条件。其他适合DNA/RNA结合条件(例如在无细胞系统的情况下的条件)在本领域中是已知的(参见例如Molecular Cloning:A Laboratory Manual,第3版(Sambrook等,Harbor Laboratory Press 2001),其出于所有目的以引用的方式整体并入本文)。靶标DNA的与Cas蛋白或gRNA互补和杂交的链可称为“互补链”,并且靶标DNA的互补于“互补链”(并且因此不互补于Cas蛋白或gRNA)的链可称为“非互补链”或“模板链”。

[0254] Cas蛋白可在存在于靶标DNA中的由gRNA的DNA靶向区段将与其结合的核酸序列的内部或外部的位点处裂解核酸。“裂解位点”包括核酸的由Cas蛋白产生单链断裂或双链断裂所处的位置。举例来说,形成CRISPR复合物(包含gRNA杂交于引导RNA识别序列以及与Cas蛋白复合)可导致在存在于靶标DNA中的由gRNA的DNA靶向区段将与其结合的核酸序列中或附近(例如在离所述核酸序列1、2、3、4、5、6、7、8、9、10、20、50个或更多个碱基对内)对一个或两个链的裂解。如果裂解位点在gRNA的DNA靶向区段将与其结合的核酸序列的外部,那么裂解位点仍然被视为在“引导RNA识别序列”内。裂解位点可在核酸的仅一个链上或在两个链上。裂解位点可在核酸的两个链上的相同位置处(产生钝性末端),或可在各链上的不同位点处(产生交错末端(即突出部分))。交错末端可例如通过使用两个Cas蛋白来产生,所述Cas蛋白各自在不同链上的不同裂解位点处产生单链断裂,由此产生双链断裂。举例来说,第一切口酶可在双链DNA(dsDNA)的第一链上产生单链断裂,并且第二切口酶可在dsDNA的第二链上产生单链断裂,以致产生突出序列。在一些情况下,切口酶在第一链上的引导RNA识别序列与切口酶在第二链上的引导RNA识别序列分隔至少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。

[0255] 由Cas蛋白对靶标DNA的位点特异性结合和裂解可发生在由(i) gRNA与靶标DNA之间的碱基配对互补性与(ii) 靶标DNA中的称为原间隔体邻近基序(PAM)的短基序两者决定的位置处。PAM可侧接于引导RNA识别序列。任选地,引导RNA识别序列可由PAM侧接在3'末端上。或者,引导RNA识别序列可由PAM侧接在5'末端上。举例来说,Cas蛋白的裂解位点可为在PAM序列的上游或下游的约1至约10或约2至约5个碱基对(例如3个碱基对)。在一些情况下(例如当使用来自化脓性链球菌的Cas9或密切相关Cas9时),非互补链的PAM序列可为5'-N<sub>1</sub>GG-3',其中N<sub>1</sub>是任何DNA核苷酸,并且紧靠靶标DNA的非互补链的引导RNA识别序列的3'。因此,互补链的PAM序列将是5'-CCN<sub>2</sub>-3',其中N<sub>2</sub>是任何DNA核苷酸,并且紧靠靶标DNA的互补链的引导RNA识别序列的5'。在一些所述情况下,N<sub>1</sub>和N<sub>2</sub>可互补,并且N<sub>1</sub>-N<sub>2</sub>碱基对可为任何碱基对(例如N<sub>1</sub>=C且N<sub>2</sub>=G;N<sub>1</sub>=G且N<sub>2</sub>=C;N<sub>1</sub>=A且N<sub>2</sub>=T;或N<sub>1</sub>=T且N<sub>2</sub>=A)。在来自金黄色葡萄球菌的Cas9的情况下,PAM可为NNGRRT(SEQ ID NO:146)或NNGRR(SEQ ID NO:147),其中N

可为A、G、C或T,并且R可为G或A。在来自空肠弯曲杆菌的Cas9的情况下,PAM可为例如NNNNACAC或NNNNRYAC,其中N可为A、G、C或T,并且R可为G或A。在一些情况下(例如对于FnCpf1),PAM序列可在5'末端的上游,并且具有序列5'-TTN-3'。

[0256] 引导RNA识别序列的实例包括互补于gRNA的DNA靶向区段的DNA序列,或这种DNA序列加上PAM序列。举例来说,靶标基序可为紧靠在由Cas9蛋白识别的NGG基序之前的20核苷酸DNA序列,诸如GN<sub>19</sub>NGG(SEQ ID NO:1)或N<sub>20</sub>NGG(SEQ ID NO:2)(参见例如WO 2014/165825,其出于所有目的以引用的方式整体并入本文)。在5'末端的鸟嘌呤可有助于在细胞中通过RNA聚合酶达成的转录。引导RNA识别序列的其他实例可在5'末端包括两个鸟嘌呤核苷酸(例如GGN<sub>20</sub>NGG;SEQ ID NO:3)以有助于在体外通过T7聚合酶达成的高效转录。参见例如WO 2014/065596,其出于所有目的以引用的方式整体并入本文。其他引导RNA识别序列可具有SEQ ID NO:1-3的长度在4-22个之间的核苷酸,包括5' G或GG和3' GG或NGG。其他引导RNA识别序列可具有SEQ ID NO:1-3的长度在14与20个之间的核苷酸。

[0257] 引导RNA识别序列可为任何细胞内源性或外源性核酸序列。引导RNA识别序列可为编码基因产物(例如蛋白质)的序列或非编码序列(例如调控序列),或可包括两者。

#### [0258] C. 外源性修复模板

[0259] 本文公开的方法和组合物可利用外源性修复模板来在用Cas蛋白裂解靶基因组基因座之后修饰所述靶基因组基因座。举例来说,细胞可为单细胞期胚胎,并且外源性修复模板的长度可小于5kb。在除单细胞期胚胎以外的细胞类型中,外源性修复模板(例如靶向载体)可更长。举例来说,在除单细胞期胚胎以外的细胞类型中,外源性修复模板可为如在本文中其他地方所述的大型靶向载体(LTVEC)(例如具有至少10kb的长度或具有总和是至少10kb的5'同源臂和3'同源臂的靶向载体)。使用外源性修复模板与Cas蛋白组合可通过促进同源性引导的修复来在靶基因组基因座处产生更精确修饰。

[0260] 在所述方法中,Cas蛋白裂解靶基因组基因座以产生单链断裂(切口)或双链断裂,并且外源性修复模板通过非同源性末端接合(NHEJ)介导的连接或通过同源性引导的修复事件来使靶标核酸重组。任选地,用外源性修复模板进行的修复会移除或破坏引导RNA识别序列或Cas裂解位点以使已被靶向的等位基因不能再由Cas蛋白靶向。

[0261] 外源性修复模板可包含脱氧核糖核酸(DNA)或核糖核酸(RNA),它们可为单链或双链,并且它们可呈线性或环状形式。举例来说,外源性修复模板可为单链寡脱氧核苷酸(ssODN)。参见例如Yoshimi等(2016)Nat. Commun. 7:10431,其出于所有目的以引用的方式整体并入本文。一示例性外源性修复模板的长度在约50个核苷酸至约5kb之间,在约50个核苷酸至约3kb之间,或在约50至约1,000个核苷酸之间。其他示例性外源性修复模板的长度在约40至约200个核苷酸之间。举例来说,外源性修复模板的长度可在约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、或约190至约200个核苷酸之间。或者,外源性修复模板的长度可在约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,外源性修复模板的长度可在约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,外源性修复模板

的长度可为例如不超过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。在除单细胞期胚胎以外的细胞类型中，外源性修复模板（例如靶向载体）可更长。举例来说，在除单细胞期胚胎以外的细胞类型中，外源性修复模板可为如在本文中其他地方所述的大型靶向载体（LTVEC）。

[0262] 在一个实例中，外源性修复模板是长度在约80个核苷酸与约200个核苷酸之间的ssODN。在另一实例中，外源性修复模板是长度在约80个核苷酸与约3kb之间的ssODN。这种ssODN可具有例如各自长度在约40个核苷酸与约60个核苷酸之间的同源臂。这种ssODN也可具有例如各自长度在约30个核苷酸与100个核苷酸之间的同源臂。同源臂可为对称的（例如各自长度是40个核苷酸或各自长度是60个核苷酸），或它们可为不对称的（例如一个同源臂的长度是36个核苷酸，并且一个同源臂的长度是91个核苷酸）。

[0263] 外源性修复模板可包括提供额外合乎需要特征（例如经改进或经调控稳定性；用荧光标记进行追踪或检测；蛋白质或蛋白质复合物的结合位点等）的修饰或序列。外源性修复模板可包含一个或多个荧光标记、纯化标签、表位标签或其组合。举例来说，外源性修复模板可包含一个或多个荧光标记（例如荧光蛋白或其他荧光团或染料），诸如至少1、至少2、至少3、至少4或至少5个荧光标记。示例性荧光标记包括荧光团诸如荧光素（例如6-羧基荧光素（6-FAM））、Texas Red、HEX、Cy3、Cy5、Cy5.5、Pacific Blue、5-（和-6）-羧基四甲基若丹明（TAMRA）和Cy7。广泛范围的用于标记寡核苷酸的荧光染料可商购获得（例如从Integrated DNA Technologies）。所述荧光标记（例如内部荧光标记）可例如用于检测已被直接整合至具有可与外源性修复模板的末端相容的突出末端的经裂解靶标核酸中的所述外源性修复模板。标记或标签可在5'末端、3'末端或外源性修复模板内部。举例来说，可使外源性修复模板在5'末端与来自Integrated DNA Technologies的IR700荧光团（5'IRDYE<sup>®</sup>700）缀合。

[0264] 外源性修复模板也可包含包括待整合在靶基因组基因座处的DNA区段的核酸插入物。在靶基因组基因座处整合核酸插入物可导致向所述靶基因组基因座中添加目标核酸序列，使在所述靶基因组基因座处的目标核酸序列缺失，或替换在所述靶基因组基因座处的目标核酸序列（即缺失和插入）。一些外源性修复模板被设计来在靶基因组基因座处插入核酸插入物而在所述靶基因组基因座处无任何相应缺失。其他外源性修复模板被设计来使在靶基因组基因座处的目标核酸序列缺失而无核酸插入物的任何相应插入。其他外源性修复模板被设计来使在靶基因组基因座处的目标核酸序列缺失，并且用核酸插入物替换它。

[0265] 核酸插入物或在靶基因组基因座处被缺失和/或替换的相应核酸可具有各种长度。一示例性核酸插入物或在靶基因组基因座处被缺失和/或替换的相应核酸的长度在约1个核苷酸至约5kb之间，或长度在约1个核苷酸至约1,000个核苷酸之间。举例来说，核酸插入物或在靶基因组基因座处被缺失和/或替换的相应核酸的长度可在约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、或约190至约200个核苷酸之间。同样，核酸插入物或在靶基因组基因座处被缺失和/或替换的相应核酸的长度可在约1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。同样,核酸插入物或在靶基因组基因座处被缺失和/或替换的相应核酸的长度可在约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。从靶基因组基因座缺失的核酸也可在约1kb至约5kb、约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约70kb至约80kb、约80kb至约90kb、约90kb至约100kb、约100kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约600kb、约600kb至约700kb、约700kb至约800kb、约800kb至约900kb、约900kb至约1Mb之间或更长。或者,从靶基因组基因座缺失的核酸可在约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。

[0266] 核酸插入物可包含基因组DNA或任何其他类型的DNA。举例来说,核酸插入物可来自原核生物、真核生物、酵母、鸟(例如鸡)、非人哺乳动物、啮齿动物、人、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、猫、狗、雪貂、灵长类动物(例如狨猴、恒河猴)、驯化哺乳动物、农业哺乳动物、龟或任何其他目标生物体。

[0267] 核酸插入物可包含与编码自身抗原的基因的全部或一部分(例如所述基因的编码自身抗原的特定基序或区域的一部分)同源或直系同源的序列。同源序列可来自不同物种或相同物种。举例来说,核酸插入物可包含相较于在靶基因组基因座处被靶向以进行替换的序列,包含一个或多个点突变(例如1、2、3、4、5个或更多个)的序列。任选地,所述点突变可导致所编码多肽中的保守性氨基酸取代(例如天冬氨酸[Asp,D]被谷氨酸[Glu,E]取代)。

[0268] 核酸插入物或在靶基因组基因座处被缺失和/或替换的相应核酸可为编码区诸如外显子;非编码区诸如内含子、非翻译区或调控区(例如启动子、增强子或转录阻遏物结合元件);或其任何组合。

[0269] 核酸插入物也可包含条件性等位基因。条件性等位基因可为如US2011/0104799中所述的多功能性等位基因,所述专利出于所有目的以引用的方式整体并入本文。举例来说,条件性等位基因可包含:(a)关于靶标基因的转录的处于有义定向的触动序列;(b)处于有义或反义定向的药物选择盒(DSC);(c)处于反义定向的目标核苷酸序列(NSI);和(d)处于相反定向的倒位条件性模块(COIN,其利用外显子隔裂性内含子和可倒位基因捕集样模块)。参见例如US 2011/0104799。条件性等位基因可进一步包含可重组单元,其在暴露于第一重组酶后重组以形成(i)缺乏触动序列和DSC;和(ii)含有处于有义定向的NSI和处于反义定向的COIN的条件性等位基因。参见例如US 2011/0104799。

[0270] 核酸插入物也可包含编码选择标记的多核苷酸。或者,核酸插入物可缺乏编码选择标记的多核苷酸。选择标记可含于选择盒中。任选地,选择盒可为自我缺失性盒。参见例如US 8,697,851和US 2013/0312129,其各自出于所有目的以引用的方式整体并入本文。作为一实例,自我缺失性盒可包含可操作地连接至小鼠Prm1启动子的Crei基因(包含编码Cre重组酶的两个外显子,由内含子分隔)和可操作地连接至人泛素启动子的新霉素抗性基因。通过采用Prm1启动子,自我缺失性盒可在F0动物的雄性生殖细胞中被特异性缺失。示例性选择标记包括新霉素磷酸转移酶(neo<sup>r</sup>)、潮霉素B磷酸转移酶(hyg<sup>r</sup>)、嘌呤霉素-N-乙酰基转

移酶(puro<sup>r</sup>)、杀稻瘟菌素S脱氨酶(bsr<sup>r</sup>)、黄嘌呤/鸟嘌呤磷酸核糖基转移酶(gpt)、或单纯疱疹病毒胸苷激酶(HSV-k)或其组合。编码选择标记的多核苷酸可被可操作地连接至在所靶向的细胞中具有活性的启动子。启动子的实例在本文中其他地方描述。

[0271] 核酸插入物也可包含报道体基因。示例性报道体基因包括编码以下的那些:荧光素酶、 $\beta$ -半乳糖苷酶、绿色荧光蛋白(GFP)、增强绿色荧光蛋白(eGFP)、青色荧光蛋白(CFP)、黄色荧光蛋白(YFP)、增强黄色荧光蛋白(eYFP)、蓝色荧光蛋白(BFP)、增强蓝色荧光蛋白(eBFP)、DsRed、ZsGreen、MmGFP、mPlum、mCherry、tdTomato、mStrawberry、J-Red、mOrange、mKO、mCitrine、Venus、YPet、Emerald、CyPet、Cerulean、T-Sapphire和碱性磷酸酶。所述报道体基因可被可操作地连接至在所靶向的细胞中具有活性的启动子。启动子的实例在本文中其他地方描述。

[0272] 核酸插入物也可包含一个或多个表达盒或缺失盒。给定盒可包含目标核苷酸序列、编码选择标记的多核苷酸和报道体基因中的一者或多者以及影响表达的各种调控组分。可包括的可选择标记和报道体基因的实例在本文中其他地方详细讨论。

[0273] 核酸插入物可包含侧接有位点特异性重组靶标序列的核酸。或者,核酸插入物可包含一个或多个位点特异性重组靶标序列。尽管整个核酸插入物可由所述位点特异性重组靶标序列侧接,但核酸插入物内的任何目标区域或个别目标多核苷酸也可由所述位点侧接。可侧接于核酸插入物或核酸插入物中的任何目标多核苷酸的位点特异性重组靶标序列可包括例如loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attP、att、FRT、rox或其组合。在一个实例中,位点特异性重组位点侧接于核酸插入物内含有的编码选择标记的多核苷酸和/或报道体基因。在核酸插入物整合在所靶向基因座处之后,可移除位点特异性重组位点之间的序列。任选地,可使用两个外源性修复模板,各自具有包含位点特异性重组位点的核酸插入物。可使外源性修复模板靶向侧接于目标核酸的5'和3'区域。在两个核酸插入物整合至靶基因组基因座中之后,可移除两个经插入位点特异性重组位点之间的目标核酸。

[0274] 核酸插入物也可包含限制核酸内切酶(即限制酶)的一个或多个限制位点,所述限制核酸内切酶包括I型、II型、III型和IV型核酸内切酶。I型和III型限制核酸内切酶识别特定识别位点,但通常在离开核酸酶结合位点的可变位置处进行裂解,所述结合位点可远离裂解位点(识别位点)数百个碱基对。在II型系统中,限制活性不依赖于任何甲基酶活性,并且裂解通常发生在结合位点内或附近的特定位点处。大多数II型酶切割回文序列,然而,IIa型酶识别非回文识别位点并在识别位点的外部进行裂解,IIb型酶切割序列两次,其中两个位点均在识别位点的外部,并且IIc型酶识别不对称识别位点并在一侧以及在离识别位点约1-20个核苷酸的确定距离下进行裂解。IV型限制酶靶向甲基化DNA。例如在REBASE数据库(网页在rebase.neb.com处;Roberts等,(2003)Nucleic Acids Res.31:418-420;Roberts等,(2003)Nucleic Acids Res.31:1805-1812;以及Belfort等(2002),Mobile DNA II,第761-783页,Craigie等人编,(ASM Press,Washington,DC))中进一步描述和分类限制酶。

[0275] (1) 用于非同源性末端接合介导的插入的修复模板

[0276] 一些外源性修复模板在5'末端和/或3'末端具有互补于一个或多个由Cas蛋白介导的在靶基因组基因座处的裂解产生的突出部分的短单链区域。这些突出部分也可被称为

5' 同源臂和3' 同源臂。举例来说,一些外源性修复模板在5' 末端和/或3' 末端具有互补于一个或多个由Cas蛋白介导的在靶基因组基因座处的5' 靶标序列和/或3' 靶标序列处的裂解产生的突出部分的短单链区域。一些所述外源性修复模板仅在5' 末端或仅在3' 末端具有互补性区域。举例来说,一些所述外源性修复模板仅在5' 末端具有互补于在靶基因组基因座处的5' 靶标序列处产生的突出部分的互补性区域,或仅在3' 末端具有互补于在靶基因组基因座处的3' 靶标序列处产生的突出部分的互补性区域。其他所述外源性修复模板在5' 末端与3' 末端两者处均具有互补性区域。举例来说,其他所述外源性修复模板在5' 末端与3' 末端两者处均具有互补性区域,其例如分别互补于由Cas介导的在靶基因组基因座处的裂解产生的第一突出部分和第二突出部分。举例来说,如果外源性修复模板是双链,那么单链互补性区域可从修复模板的顶链的5' 末端和修复模板的底链的5' 末端延伸,从而在各末端上产生5' 突出部分。或者,单链互补性区域可从修复模板的顶链的3' 末端和从模板的底链的3' 末端延伸,从而产生3' 突出部分。

[0277] 互补性区域可具有足以促进外源性修复模板与靶标核酸之间的连接的任何长度。示例性互补性区域的长度在约1至约5个核苷酸之间,在约1至约25个核苷酸之间,或在约5至约150个核苷酸之间。举例来说,互补性区域的长度可为至少约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸。或者,互补性区域的长度可为约5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个核苷酸,或更长。

[0278] 所述互补性区域可互补于由两对切口酶产生的突出部分。两个具有交错末端的双链断裂可通过使用裂解相对DNA链以产生第一双链断裂的第一切口酶和第二切口酶、以及裂解相对DNA链以产生第二双链断裂的第三切口酶和第四切口酶来产生。举例来说,Cas蛋白可用于对与第一、第二、第三和第四引导RNA相对应的第一、第二、第三和第四引导RNA识别序列进行切口。第一引导RNA识别序列和第二引导RNA识别序列可被定位来产生第一裂解位点以使由第一切口酶和第二切口酶在第一DNA链和第二DNA链上产生的切口产生双链断裂(即所述第一裂解位点包含第一引导RNA识别序列和第二引导RNA识别序列内的切口)。同样,第三引导RNA识别序列和第四引导RNA识别序列可被定位来产生第二裂解位点以使由第三切口酶和第四切口酶在第一DNA链和第二DNA链上产生的切口产生双链断裂(即所述第二裂解位点包含第三引导RNA识别序列和第四引导RNA识别序列内的切口)。优选地,第一引导RNA识别序列和第二引导RNA识别序列和/或第三引导RNA识别序列和第四引导RNA识别序列内的切口可为产生突出部分的偏移切口。偏移窗可为例如至少约5bp、10bp、20bp、30bp、40bp、50bp、60bp、70bp、80bp、90bp、100bp或更多。参见Ran等(2013)Cell 154:1380-1389; Mali等(2013)Nat. Biotech. 31:833-838;以及Shen等(2014)Nat. Methods 11:399-404,其各自出于所有目的以引用的方式整体并入本文。在所述情况下,双链外源性修复模板可被设计成具有互补于由第一引导RNA识别序列和第二引导RNA识别序列内的切口以及由第三引导RNA识别序列和第四引导RNA识别序列内的切口产生的突出部分的单链互补性区域。这种外源性修复模板可接着通过非同源性末端接合介导的连接来插入。

[0279] (2) 用于通过同源性引导的修复进行的插入的修复模板

[0280] 一些外源性修复模板包含同源臂。如果外源性修复模板也包含核酸插入物,那么

同源臂可侧接于所述核酸插入物。为便于引用,同源臂在本文中被称为5'和3' (即上游和下游) 同源臂。这个术语涉及同源臂相对于外源性修复模板内的核酸插入物的相对位置。5'同源臂和3'同源臂分别对应于靶基因组基因座内的在本文中被称为“5'靶标序列”和“3'靶标序列”的区域。

[0281] 当两个区域彼此共有足以充当同源性重组反应的底物的序列同一性水平时,同源臂和靶标序列“对应”或彼此“对应”。术语“同源性”包括DNA序列与相应序列同一或共有与相应序列的序列同一性。给定靶标序列与见于外源性修复模板中的相应同源臂之间的序列同一性可为允许发生同源性重组的任何程度的序列同一性。举例来说,由外源性修复模板的同源臂(或其片段)和靶标序列(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,以使序列经受同源性重组。此外,同源臂与相应靶标序列之间具有同源性的相应区域可具有足以促进同源性重组的任何长度。示例性同源臂的长度在约25个核苷酸至约2.5kb之间,在约25个核苷酸至约1.5kb之间,或在约25至约500个核苷酸之间。举例来说,给定同源臂(或各同源臂)和/或相应靶标序列可包含长度在约25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个核苷酸之间的相应同源性区域,以使同源臂具有足以经受与靶标核酸内的相应靶标序列的同源性重组的同源性。或者,给定同源臂(或各同源臂)和/或相应靶标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之间的相应同源性区域。举例来说,同源臂的长度可各自是约750个核苷酸。同源臂可为对称的(各自长度大小大约相同),或它们可为不对称的(一个长于另一个)。

[0282] 同源臂可对应于为细胞所天然具有的基因座(例如所靶向基因座)。或者,举例来说,它们可对应于整合至细胞的基因组中的异源性或外源性DNA区段的区域,所述区段包括例如转基因、表达盒或异源性或外源性DNA区域。或者,靶向载体的同源臂可对应于酵母人工染色体(YAC)、细菌人工染色体(BAC)、人人工染色体的区域或适当宿主细胞中含有的任何其他工程化区域。更进一步,靶向载体的同源臂可对应于或源于BAC文库、粘粒文库或P1噬菌体文库的区域,或可源于合成DNA。

[0283] 当CRISPR/Cas系统与外源性修复模板组合使用时,5'靶标序列和3'靶标序列优选足够邻近于Cas裂解位点加以定位(例如在引导RNA识别序列的足够邻近内),以便在Cas裂解位点处的单链断裂(切口)或双链断裂后,促进在靶标序列与同源臂之间发生同源性重组事件。术语“Cas裂解位点”包括由Cas酶(例如与引导RNA复合的Cas9蛋白)产生切口或双链断裂所处的DNA序列。如果距离诸如将在Cas裂解位点处的单链断裂或双链断裂后促进在5'靶标序列和3'靶标序列与同源臂之间发生同源性重组事件,那么所靶向基因座内的对应于外源性修复模板的5'同源臂和3'同源臂的靶标序列“足够邻近于所述Cas裂解位点加以定位”。因此,对应于外源性修复模板的5'同源臂和/或3'同源臂的靶标序列可例如在给定Cas裂解位点的至少1个核苷酸内或在给定Cas裂解位点的至少10个核苷酸至约1,000个核苷酸内。作为一实例,Cas裂解位点可紧邻靶标序列中的至少一者或两者。

[0284] 或者,给定裂解位点离5'靶标序列、3'靶标序列或两个靶标序列具有不同长度。举

例来说,如果使用两个引导RNA,那么第一引导RNA识别序列和/或第二引导RNA识别序列或第一裂解位点和/或第二裂解位点可位于5'靶标序列与3'靶标序列之间,或可邻近于或接近于5'靶标序列和/或3'靶标序列,诸如在5'靶标序列和/或3'靶标序列的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。或者,第一引导RNA识别序列和/或第二引导RNA识别序列或第一裂解位点和/或第二裂解位点可离5'靶标序列和/或3'靶标序列至少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加以定位。举例来说,第一引导RNA识别序列和/或第二引导RNA识别序列或第一裂解位点和/或第二裂解位点可离5'靶标序列和/或3'靶标序列在约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之间加以定位。或者,第一引导RNA识别序列和/或第二引导RNA识别序列或第一裂解位点和/或第二裂解位点可离5'靶标序列和/或3'靶标序列超过50bp、超过100bp、超过200bp、超过300bp、超过400bp、超过500bp、超过600bp、超过700bp、超过800bp、超过900bp、超过1kb、超过2kb、超过3kb、超过4kb、超过5kb、超过6kb、超过7kb、超过8kb、超过9kb、超过10kb、超过20kb、超过30kb、超过40kb、超过50kb、超过60kb、超过70kb、超过80kb、超过90kb或超过100kb加以定位。举例来说,第一引导RNA识别序列或第一裂解位点可离5'靶标序列或离5'靶标序列与3'靶标序列两者超过50bp、超过100bp、超过200bp、超过300bp、超过400bp、超过500bp、超过600bp、超过700bp、超过800bp、超过900bp、超过1kb、超过2kb、超过3kb、超过4kb、超过5kb、超过6kb、超过7kb、超过8kb、超过9kb、超过10kb、超过20kb、超过30kb、超过40kb、超过50kb、超过60kb、超过70kb、超过80kb、超过90kb或超过100kb加以定位。同样,第二引导RNA识别序列或第二裂解位点可离3'靶标序列或离5'靶标序列与3'靶标序列两者超过50bp、超过100bp、超过200bp、超过300bp、超过400bp、超过500bp、超过600bp、超过700bp、超过800bp、超过900bp、超过1kb、超过2kb、超过3kb、超过4kb、超过5kb、超过6kb、超过7kb、超过8kb、超过9kb、超过10kb、超过20kb、超过30kb、超过40kb、超过50kb、超过60kb、超过70kb、超过80kb、超过90kb或超过100kb加以定位。

[0285] 对应于外源性修复模板的同源臂的靶标序列和Cas裂解位点的空间关系可不同。举例来说,靶标序列可位于Cas裂解位点的5',靶标序列可位于Cas裂解位点的3',或靶标序列可侧接于Cas裂解位点。

[0286] 在除单细胞期胚胎以外的细胞中,外源性修复模板可为“大型靶向载体”或“LTVEC”,其包括包含同源臂的靶向载体,所述同源臂对应于和源于大于通常由意图在细胞

中进行同源性重组的其他方法使用的那些核酸序列的核酸序列。LTVEC也包括包含核酸插入物的靶向载体,所述核酸插入物具有大于通常由意图在细胞中进行同源性重组的其他方法使用的那些核酸序列的核酸序列。举例来说,LTVEC使得有可能修饰由于它们的大小限制而不能被传统基于质粒的靶向载体容纳的大型基因座。举例来说,所靶向基因座可为(即5'同源臂和3'同源臂可对应于)细胞的不可使用常规方法靶向或在不存在由核酸酶试剂(例如Cas蛋白)诱导的切口或双链断裂下仅可不正确靶向或仅可以显著较低效率靶向的基因座。

[0287] LTVEC的实例包括源于细菌人工染色体(BAC)、人人工染色体或酵母人工染色体(YAC)的载体。LTVEC和用于制备它们的方法的非限制性实例例如描述于美国专利号6,586,251;6,596,541;和7,105,348中;以及WO 2002036789中,所述专利各自出于所有目的以引用的方式整体并入本文。LTVEC可呈线性形式或呈环状形式。

[0288] LTVEC可具有任何长度,并且长度通常是至少10kb。举例来说,LTVEC可为约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。LTVEC也可为约50kb至约500kb、约100kb至约125kb、约300kb至约325kb、约325kb至约350kb、约350kb至约375kb、约375kb至约400kb、约400kb至约425kb、约425kb至约450kb、约450kb至约475kb、或约475kb至约500kb。或者,LTVEC可为至少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或更大。LTVEC的大小可太大以致不能通过常规测定例如southern印迹和长程(例如1kb至5kb)PCR来筛选靶向事件。

[0289] LTVEC中的5'同源臂和3'同源臂的总和通常是至少10kb。作为一实例,5'同源臂可在约5kb至约100kb的范围内和/或3'同源臂可在约5kb至约100kb的范围内。作为另一实例,5'同源臂可在约5kb至约150kb的范围内和/或3'同源臂可在约5kb至约150kb的范围内。各同源臂可为例如约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'同源臂和3'同源臂的总和可为例如约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'同源臂和3'同源臂的总和也可为例如约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb。或者,各同源臂可为至少5kb、至少10kb、至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少110kb、至少120kb、至少130kb、至少140kb、至少150kb、至少160kb、至少170kb、至少180kb、至少190kb或至少200kb。同样,5'同源臂和3'同源臂的总和可为至少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。各同源臂也可为至少250kb、至少300kb、至少350kb或至少400kb。

[0290] LTVEC可包含核酸插入物,其具有大于通常由意图在细胞中进行同源性重组的其他方法使用的那些核酸序列的核酸序列。举例来说,LTVEC可包含在约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、约350kb至约400kb的范围内或更大的核酸插入物。LTVEC也可包含例如在约1kb至约5kb、约400kb至约450kb、约450kb至约500kb的范围内或更大的核酸插入物。或者,核酸插入物可为至少1kb、至少5kb、至少10kb、至少20kb、至少30kb、至少40kb、至少60kb、至少80kb、至少100kb、至少150kb、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。

[0291] D. 接触细胞的基因组以及将核酸或蛋白质引入细胞中

[0292] 接触细胞的基因组可包括将一个或多个Cas蛋白或编码Cas蛋白的核酸、一个或多个引导RNA或编码引导RNA的核酸(即一个或多个CRISPR RNA和一个或多个tracrRNA)和一个或多个外源性修复模板引入细胞中,前提是如果细胞是例如单细胞期胚胎,那么外源性修复模板的长度可小于5kb。接触细胞的基因组(例如接触细胞)可包括将仅一个以上组分、一个或多个组分或全部组分引入细胞中。“引入”包括以序列进入细胞的内部的方式向细胞递呈核酸或蛋白质。引入可通过任何手段来实现,并且一个或多个组分(例如两个组分或全部组分)可以任何组合同时或依序引入细胞中。举例来说,外源性修复模板可在引入Cas蛋白和引导RNA之前引入,或它可在引入所述Cas蛋白和所述引导RNA之后引入(例如所述外源性修复模板可在引入所述Cas蛋白和所述引导RNA之前或之后约1、2、3、4、8、12、24、36、48或72小时施用)。参见例如US 2015/0240263和US 2015/0110762,其各自出于所有目的以引用的方式整体并入本文。

[0293] Cas蛋白可以蛋白质形式诸如与gRNA复合的Cas蛋白,或以编码所述Cas蛋白的核酸形式诸如RNA(例如信使RNA(mRNA))或DNA引入细胞中。当以DNA形式引入时,编码引导RNA的DNA可被可操作地连接至在细胞中具有活性的启动子。所述DNA可在一个或多个表达构建体中。

[0294] 引导RNA可以RNA形式或以编码所述引导RNA的DNA形式引入细胞中。当以DNA形式引入时,编码引导RNA的DNA可被可操作地连接至在细胞中具有活性的启动子。所述DNA可在一个或多个表达构建体中。举例来说,所述表达构建体可为单一核酸分子的组分。或者,它们可以任何组合分开在两个或更多个核酸分子之中(即编码一个或多个CRISPR RNA的DNA和编码一个或多个tracrRNA的DNA可为单独核酸分子的组分)。

[0295] 在一些方法中,编码核酸酶试剂(例如Cas蛋白和引导RNA)的DNA和/或编码外源性修复模板的DNA可通过DNA小环来引入细胞中。参见例如WO 2014/182700,其出于所有目的以引用的方式整体并入本文。DNA小环是可用于非病毒性基因转移的超卷曲DNA分子,其不具有复制起点,也不具有抗生素选择标记。因此,DNA小环在大小方面通常小于质粒载体。这些DNA缺乏细菌DNA,因此缺乏见于细菌DNA中的未甲基化CpG基序。

[0296] 本文提供的方法不取决于用于将核酸或蛋白质引入细胞中的特定方法,只是为了

核酸或蛋白质进入至少一个细胞的内部。用于将核酸和蛋白质引入各种细胞类型中的方法在本领域中是已知的,并且包括例如稳定转染方法、短暂转染方法和病毒介导的方法。

[0297] 转染方案以及用于将核酸或蛋白质引入细胞中的方案可变化。非限制性转染方法包括基于化学的转染方法,使用脂质体;纳米粒子;磷酸钙(Graham等(1973) *Virology* 52(2):456-67, Bacchetti等(1977) *Proc. Natl. Acad. Sci. USA* 74(4):1590-4以及Kriegler, M(1991). *Transfer and Expression: A Laboratory Manual*. New York: W. H. Freeman and Company. 第96-97页);树枝状聚合物;或阳离子聚合物诸如DEAE-右旋糖酐或聚乙烯亚胺。非化学方法包括电穿孔、声致穿孔和光学转染。基于粒子的转染包括使用基因枪或磁体辅助转染(Bertram(2006) *Current Pharmaceutical Biotechnology* 7, 277-28)。病毒方法也可用于转染。

[0298] 将核酸或蛋白质引入细胞中也可由电穿孔、细胞质内注射、病毒性感染、腺病毒、慢病毒、逆转录病毒、转染、脂质介导的转染或核转染介导。将核酸或蛋白质引入细胞中也可由腺相关病毒介导。核转染是一种改进电穿孔技术,其使得核酸基质能够不仅被递送至细胞质中,而且也穿过核膜以及进入核中。此外,在本文公开的方法中使用核转染通常需要比常规电穿孔少得多的细胞(例如相较于由常规电穿孔所需的700万,仅需要约200万)。在一个实例中,核转染使用**LONZA**<sup>®</sup> NUCLEOFECTOR<sup>™</sup>系统来进行。

[0299] 将核酸或蛋白质引入细胞(例如单细胞期胚胎)中也可通过显微注射来实现。在单细胞期胚胎中,显微注射可向母亲和/或父亲前核中或向细胞质中进行。如果显微注射仅向一个前核中进行,那么父亲前核由于它的尺寸较大而是优选的。mRNA的显微注射优选向细胞质中进行(例如以将mRNA直接递送至翻译机构中),而Cas蛋白或编码Cas蛋白或编码RNA的核酸的显微注射优选向核/前核中进行。或者,显微注射可通过向核/前核与细胞质两者中注射来进行:可首先将针引入核/前核中,并且可注射第一量,以及当从单细胞期胚胎移除针时,可将第二量注射至细胞质中。如果将Cas蛋白注射至细胞质中,那么所述Cas蛋白优选包含核定位信号以确保递送至核/前核中。用于进行显微注射的方法是熟知的。参见例如Nagy等(Nagy A, Gertszenstein M, Vintersten K, Behringer R., 2003, *Manipulating the Mouse Embryo*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Meyer等(2010) *Proc. Natl. Acad. Sci. USA* 107:15022-15026以及Meyer等(2012) *Proc. Natl. Acad. Sci. USA* 109:9354-9359。向单细胞期胚胎中的引入也可通过电穿孔来实现。

[0300] 用于将核酸或蛋白质引入细胞中的其他方法可包括例如载体递送、粒子介导的递送、外体介导的递送、脂质纳米粒子介导的递送、细胞穿透肽介导的递送或可植入装置介导的递送。

[0301] 将核酸或蛋白质引入细胞中可一次或历经一段时期多次进行。举例来说,引入可历经一段时期至少两次,历经一段时期至少三次,历经一段时期至少四次,历经一段时期至少五次,历经一段时期至少六次,历经一段时期至少七次,历经一段时期至少八次,历经一段时期至少九次,历经一段时期至少十次,至少十一次,历经一段时期至少十二次,历经一段时期至少十三次,历经一段时期至少十四次,历经一段时期至少十五次,历经一段时期至少十六次,历经一段时期至少十七次,历经一段时期至少十八次,历经一段时期至少十九次,或历经一段时期至少二十次进行。

[0302] 在一些情况下,方法和组合物中采用的细胞具有稳定并入它们的基因组中的DNA构建体。在所述情况下,接触可包括提供具有已稳定并入它的基因组中的构建体的细胞。举例来说,本文公开的方法中采用的细胞可具有稳定并入它的基因组中的预先存在的Cas编码基因(即Cas就绪细胞)。“稳定并入”或“稳定引入”或“稳定整合”包括将多核苷酸引入细胞中以使核苷酸序列整合至细胞基因组中,并且能够通过其子代遗传。任何方案都可用于稳定并入DNA构建体或靶向基因组整合系统的各种组分。

[0303] E. 靶基因组基因座以及引导RNA识别序列的位置

[0304] 靶基因组基因座可为影响与目标外来靶抗原同源或共有目标表位的自身抗原的表达的任何基因组基因座。优选地,靶基因组基因座包含以下、基本上由以下组成或由以下组成:编码自身抗原的基因的全部或一部分。作为一实例,靶基因组基因座可包含以下、基本上由以下组成或由以下组成:包含编码自身抗原的基因的起始密码子的区域,或可包含以下、基本上由以下组成或由以下组成:所述基因的整个编码区。或者,靶基因组基因座可包含以下、基本上由以下组成或由以下组成:影响编码自身抗原的基因的表达的另一基因组基因座。这种基因组基因座的一实例是编码为编码自身抗原的基因的表达所需的转录调控子的基因的全部或一部分。在一些方法中,可靶向多个靶基因组基因座。作为一实例,如果存在编码与目标外来抗原同源或共有目标表位的多个自身抗原的多个基因,那么多个基因各自可被依序或同时靶向。

[0305] 第一引导RNA识别序列和第二引导RNA识别序列可在靶基因组基因座内的任何地方。举例来说,第一引导RNA识别序列和第二引导RNA识别序列可侧接于编码与目标外来靶抗原同源或共有目标表位的自身抗原的基因的全部或一部分。在一个实例中,第一引导RNA识别序列包含编码自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且第二引导RNA识别序列包含编码自身抗原的基因的终止密码子,或在终止密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。举例来说,第一引导RNA识别序列可包含起始密码子,并且第二引导RNA识别序列可包含终止密码子。如果也使用第三引导RNA和第四引导RNA,那么第三引导RNA识别序列和第四引导RNA识别序列也可在靶基因组基因座内的任何地方。举例来说,两个引导RNA识别序列(例如第一和第三,其中第一引导RNA识别序列和第三引导RNA识别序列不同且任选重叠)可包含编码自身抗原的基因的起始密码子,或可在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内,并且另外两个引导RNA识别序列(例如第二和第四,其中第二引导RNA识别序列和第四引导RNA识别序列不同且任选重叠)可包含编码自身抗原的基因的终止密码子或可在终止密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。靶向起始密码子与终止密码子两者可导致编码自身抗原的基因的编码序列的缺失,并且由此消除自身抗原的表达。

[0306] 在另一实例中,第一引导RNA识别序列和第二引导RNA识别序列不同,并且中的每一个包含编码自身抗原的基因的起始密码子,或在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。举例来说,第一引导RNA识别序列和第二引导RNA识别序列可重叠,并且可各自包含起始密码子。如果也使用第三引导RNA和/或第四引导RNA,那么第三引导RNA识别序列和第四引导RNA识别序列可在靶基因组基因座内的任何地方。举例来说,第三引导RNA识别序列和第四引导RNA识别序列可不同于彼此以及不同于第一引导

RNA识别序列和第二引导RNA识别序列,并且第三引导RNA识别序列和第四引导RNA识别序列各自也可包含编码自身抗原的基因的起始密码子,或可在起始密码子的约10、20、30、40、50、100、200、300、400、500或1,000个核苷酸内。靶向起始密码子可破坏起始密码子,并且由此消除编码自身抗原的基因的表达。

[0307] 如果使用第三引导RNA和第四引导RNA(或额外引导RNA),那么也可靶向影响第一自身抗原的表达或影响与目标外来抗原同源或共有目标表位的其他自身抗原(例如第二自身抗原)的表达的额外靶基因组基因座以使第一自身抗原和/或其他自身抗原的表达降低。作为一实例,在一些方法中,可靶向编码与目标外来抗原同源或共有目标表位的第一自身抗原的基因,并且可靶向编码与目标外来抗原同源或共有目标表位的第二自身抗原的第二基因。

[0308] F. 重组机理以及用于改变非同源性末端接合、基因转换或同源性重组的普遍性的方法

[0309] 重组包括在两个多核苷酸之间交换遗传信息的任何过程,并且可通过任何机理来发生。应答于双链断裂(DSB)的重组主要通过以下两个保守DNA修复路径来发生:非同源性末端接合(NHEJ)和同源性重组(HR)。参见Kasperek和Humphrey(2011) *Seminars in Cell& Dev. Biol.* 22:886-897,其出于所有目的以引用的方式整体并入本文。同样,由外源性修复模板介导的对靶标核酸的修复可包括在两个多核苷酸之间交换遗传信息的任何过程。

[0310] NHEJ包括通过使断裂末端直接连接至彼此或外源性序列而无需同源性模板来修复核酸中的双链断裂。由NHEJ达成的对非连续序列的连接可常常在双链断裂的位点附近导致缺失、插入或易位。举例来说,NHEJ也可通过使断裂末端与外源性修复模板的末端直接连接(即基于NHEJ的捕集)来导致外源性修复模板的靶向整合。当同源性引导的修复(HDR)路径不易于可用时(例如在非分裂细胞、原代细胞和不良进行基于同源性的DNA修复的细胞中),所述NHEJ介导的靶向整合可优选用于插入外源性修复模板。此外,与同源性引导的修复相比,不需要了解侧接于裂解位点的具有序列同一性的大型区域(超出由Cas介导的裂解产生的突出部分),此在试图向具有对基因组序列的了解有限的基因组的生物体中进行靶向插入时可有益。整合可通过外源性修复模板与经裂解基因组序列之间的钝性末端连接,或通过使用由可与经裂解基因组序列中的由Cas蛋白产生的那些突出部分相容的突出部分侧接的外源性修复模板进行的粘性末端(即具有5'或3'突出部分)连接来进行。参见例如US 2011/020722、WO 2014/033644、WO 2014/089290以及Maresca等(2013) *Genome Res.* 23(3):539-546,其各自出于所有目的以引用的方式整体并入本文。如果连接钝性末端,那么可需要进行靶标和/或供体切除以产生具有为片段接合所需的微同源性的区域,此可在靶标序列中产生非所要改变。

[0311] 重组也可通过同源性引导的修复(HDR)或同源性重组(HR)来发生。HDR或HR包括可需要核苷酸序列同源性,使用“供体”分子作为“靶标”分子(即经受双链断裂的分子)的修复模板,并且导致遗传信息从供体向靶标转移的核酸修复形式。在不希望受任何特定理论束缚下,所述转移可涉及对在断裂的靶标与供体之间形成的异源双链体DNA的不匹配修正,和/或合成依赖性链退火,其中供体用于再合成将成为靶标的一部分的遗传信息,和/或相关过程。在一些情况下,供体多核苷酸、供体多核苷酸的一部分、供体多核苷酸的拷贝、或供体多核苷酸的拷贝的一部分整合至靶标DNA中。参见Wang等(2013) *Cell* 153:910-918;

Mandalos等(2012) PLOS ONE 7:e45768:1-9;以及Wang等(2013) Nat Biotechnol.31:530-532,其各自出于所有目的以引用的方式整体并入本文。

[0312] 重组可在一对同源染色体中第一染色体与第二染色体之间。所述手段可包括例如杂合性丧失 (LOH)、基因转换或通过任何已知重组机理而发生的交叉事件。在不希望受理论束缚下,LOH可例如通过伴有或不伴有基因转换的有丝分裂重组,或通过染色体损失和重复来发生。参见例如Lefebvre等(2001) Nat. Genet. 27:257-258,其出于所有目的以引用的方式整体并入本文。在这个情形下的基因转换可包括遗传物质从供体序列向高度同源性接受体的单向转移(即遗传信息从一个分子向它的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理来拷贝等位基因的任何手段。举例来说,基因转换可涉及遗传信息从完整序列向含有双链断裂的同源性区域的非相互转移,并且它可发生在姐妹染色单体、同源染色体、或同一染色单体上或不同染色体上的同源序列之间。参见例如Chen等(2007) Nat. Rev. Genet. 8:762-775,其出于所有目的以引用的方式整体并入本文。在特定情况下,基因转换直接由作为从同源染色体拷贝遗传信息的结果的同源性重组所致。当同源序列不同一时,这可导致局部化杂合性丧失 (LOH)。

[0313] 作为一实例,LOH可通过由有丝分裂交叉达成的相互染色单体交换,或通过利用断裂诱导的复制进行的染色单体拷贝来发生。在任一情况下,都可发生杂合性修饰,其中一个染色体在基因组复制之前被靶向。或者,单一染色单体可在基因组复制之后被靶向,继之以染色单体间基因转换。

[0314] 在任何本文公开的方法中,细胞可为已被修饰来使NHEJ活性增加或降低的细胞。同样,细胞可为已被修饰来使基因转换或HDR活性增加的细胞。所述修饰可包括在调控NHEJ、基因转换和/或HDR中涉及的基因的表达或活性方面的改变。举例来说,使NHEJ的活性降低和/或使HDR的活性增加可促进对应于两个核酸酶试剂(例如Cas蛋白和两个引导RNA)的核酸酶识别序列(例如引导RNA识别序列)之间的基因组区域的双等位基因垮塌。在不希望受任何特定理论束缚下,双等位基因基因组垮塌可发生所采用的一种机理通过在第一等位基因内NHEJ介导的修复或HDR介导的修复以及通过HDR机理诸如基因转换来产生相同第二等位基因来进行(参见实施例1)。因此,促进HDR介导的路径(例如通过使NHEJ活性降低或通过使HDR活性增加)也可促进基因组区域的双等位基因垮塌。类似地,在不希望受任何特定理论束缚下,如果NHEJ活性降低,并且HDR活性(例如基因转换活性)相应增加,那么通过使用靶向单一基因座的成对核酸酶试剂(例如Cas蛋白和成对引导RNA)进行的杂合性细胞向纯合性细胞的转换可得到促进。

[0315] 抑制剂可用于使NHEJ活性增加或降低或使HDR活性增加或降低。所述抑制剂可为例如对基因转录物具有特异性的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)) 或反义寡核苷酸。抑制剂可针对NHEJ或HDR中涉及的酶或它们的由通过例如磷酸化、泛素化和苏素化进行的翻译后修饰达成的上游调控。

[0316] 在哺乳动物细胞中,NHEJ是主要DSB修复机理,并且在整个细胞周期中都具有活性。在脊椎动物中,“典型”或“经典”NHEJ路径(C-NHEJ) 需要包括DNA-PK、Ku70-80、Artemis、连接酶IV (Lig4)、XRCC4、CLF和Pol $\mu$ 的若干核心因子来修复DSB。参见Kasperek和Humphrey (2011) Seminars in Cell&Dev. Biol. 22:886-897,其出于所有目的以引用的方式整体并入

本文。在NHEJ期间,DNA末端由高度丰富的末端保护性Ku蛋白结合,所述Ku蛋白充当用于装载其他NHEJ组分的对接站。

[0317] 因此,在一些本文公开的方法中,细胞已被修饰来使C-NHEJ中涉及的因子的表达或活性降低或消除或使所述表达或活性增加。举例来说,在一些方法中,细胞已被修饰来使DNA-PK、Ku70-80、Artemis、连接酶IV (Lig4)、XRCC4、CLF和/或Polμ表达或活性降低或消除。在特定方法中,细胞已被修饰来使DNA-PK表达或活性降低或消除,或使DNA-PK表达或活性(例如DNA-PKcs的表达或活性;示例性UniProt序列指定为P97313)增加。DNA-PKcs抑制剂的实例包括例如NU7026和NU7441。参见例如美国专利号6,974,867,其出于所有目的以引用的方式整体并入本文。在特定方法中,细胞已被修饰来使连接酶IV表达或活性降低或消除,或使连接酶IV表达或活性增加。连接酶IV抑制剂的一实例是SCR7。

[0318] 靶向细胞周期检查点蛋白如ATM(例如KU55933)、CHK1/CHK2(例如KLD1162或CHIR-124)和ATR(例如VE 821)的抑制剂也可用于协同增强特定DNA修复抑制剂的作用,或防止非意图副作用如细胞周期阻滞和/或凋亡(参见Ciccia等(2010) Mol Cell 40:179,其出于所有目的以引用的方式整体并入本文)。

[0319] 破坏C-NHEJ可使由“替代性”NHEJ(A-NHEJ)路径介导的异常接合的水平增加,并且也可使HR修复增加。A-NHEJ路径显示朝向微同源性介导的接合的偏倚,并且相比于C-NHEJ,遵循较慢动力学。已提出包括MRN复合物(MRE11、RAD50、NBS1)、CtIP、XRCC1、PARP、Lig1和Lig3的若干因子进行参与。参见Kasperek和Humphrey(2011) Seminars in Cell& Dev. Biol. 22:886-897以及Claybon等(2010) Nucleic Acids Res. 38(21):7538-7545,其各自出于所有目的以引用的方式整体并入本文。

[0320] 因此,在一些本文公开的方法中,细胞已被修饰来使A-NHEJ中涉及的因子的表达或活性降低或消除或使所述表达或活性增加。举例来说,在一些方法中,细胞已被修饰来使MRE11、RAD50、NBS1、CtIP、XRCC1、PARP(例如PARP1)、Lig1和/或Lig3表达或活性降低或消除。在其他方法中,细胞已被修饰来使MRE11、RAD50、NBS1、CtIP、XRCC1、PARP(例如PARP1)、Lig1和/或Lig3表达或活性增加。在特定方法中,细胞已被修饰来使PARP1表达或活性降低或消除,或使PARP1表达或活性增加(示例性UniProt序列指定为P11103)。PARP抑制剂(例如NU1025、依尼帕尼(Iniparib)、奥拉帕尼(Olaparib))的实例包括烟酰胺;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;啡啶和菲啶酮;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(caffeine)、茶碱(theophylline)和胸苷及其类似物和衍生物。参见例如美国专利号8,071,579,其出于所有目的以引用的方式整体并入本文。

[0321] C-NHEJ也与HR展现竞争性关系,以致破坏C-NHEJ也可导致HR修复增加。可利用NHEJ与HR之间的所述竞争,因为破坏NHEJ可由于随机整合降低以及有可能由于通过同源性重组进行的靶标整合增加而导致基因靶向增强。

[0322] 存在若干形式的同源性重组修复,包括单链退火、基因转换、交叉以及断裂诱导的复制。单链退火是次要形式的HR修复,其中在经切除DSB的任一侧上的同源性单链序列进行退火,从而导致染色体重构。视分隔具有序列同源性的两个区域的距离而定,单链退火产生不同大小的缺失。基因转换包括遗传信息从一个分子向它的同源物的非相互交换,直接由

作为从同源染色体拷贝遗传信息的结果的HR所致。当同源序列不同时,这可导致局部化LOH。通常,基因转换的程度限于几百个碱基对。然而,已报道在包括RAD51C缺乏的一些遗传背景下的长地带基因转换。参见Nagaraju等(2006) *Mol. Cell. Biol.* 26:8075-8086,其出于所有目的以引用的方式整体并入本文。交叉可例如发生在同源染色体之间,并且如果发生在G1期,那么有可能导致相互易位,或如果发生在G2期,那么有可能导致非相互易位和从断裂位点延伸至远端端粒的LOH。断裂诱导的复制是HR的一种变化形式,其中在链侵入之后,DNA复制继续进行直至染色体的末端。因此,存在HR可促进LOH所采用的许多机理。

[0323] 因此,在一些本文公开的方法中,细胞已被修饰来使HR中涉及的因子的表达或活性降低或消除或使所述表达或活性增加。举例来说,在一些方法中,细胞已被修饰来使RAD51、RAD52、RAD54、RAD55、RAD51C、BRCA1和/或BRCA2表达或活性增加。在其他方法中,细胞已被修饰来使RAD51、RAD52、RAD54、RAD55、RAD51C、BRCA1和/或BRCA2表达或活性降低或消除。

[0324] 在一些方法中,可改变调控NHEJ和/或HR中涉及的其他蛋白质的表达或活性。举例来说,在一些方法中,细胞已被修饰来使Chk2表达或活性降低或消除,使Clspn表达或活性降低或消除,使Setd2表达或活性降低或消除,使Kat2a表达或活性增加,和/或使Rad51表达或活性增加。在其他方法中,细胞已被修饰来使Chk2表达或活性增加,使Clspn表达或活性增加,使Setd2表达或活性增加,使Kat2a表达或活性降低或消除,和/或使Rad51表达或活性降低或消除。

[0325] Chk2(也称为Chek2和Rad53;裂殖酵母(*S. pombe*)同源物是Cds1)是一种为检查点介导的细胞周期阻滞、DNA修复的活化、以及应答于存在DNA双链断裂的凋亡所需的丝氨酸/苏氨酸蛋白质激酶。参见Blaikley等(2014) *Nucleic Acids Research* 42:5644-5656,其出于所有目的以引用的方式整体并入本文。Clspn(也称为Claspin;裂殖酵母同源物是Mrc1)是一种为应答于DNA损害的由检查点介导的细胞周期阻滞所需的蛋白质。已报道裂殖酵母中的Chk2或Clspn同源物的缺失会导致高重组表型,其相较于野生型展现显著升高水平的断裂诱导的基因转换。具体来说,基因转换的水平据报道得以显著增加,而非同源性末端接合(NHEJ)、姐妹染色单体转换(SCC)和杂合性丧失(LOH)的水平据报道得以降低。参见Blaikley等(2014) *Nucleic Acids Research* 42:5644-5656。

[0326] Kat2a(也称为Gcn5和Gcn5l2)是一种促进转录活化,并且已被报道与双链断裂修复相关联的遍在性组蛋白乙酰基转移酶。Kat2a依赖性组蛋白H3赖氨酸36(H3K36)乙酰化使染色质可及性增加,使切除增加,并且促进同源性重组,同时抑制非同源性末端接合。参见Pai等(2014) *Nat. Commun.* 5:4091,其出于所有目的以引用的方式整体并入本文。Setd2(也称为Kiaa1732、Kmt3a和Set2)是一种组蛋白甲基转移酶,其使用脱甲基化赖氨酸36(H3K36me2)作为底物,使组蛋白H3的赖氨酸36特异性三甲基化(H3K36me3)。Setd2依赖性H3K36甲基化使染色质可及性降低,使切除降低,并且促进NHEJ。参见Pai等(2014) *Nat. Commun.* 5:4091。

[0327] Rad 51(也称为Reca、Rad51A和DNA修复蛋白Rad51同源物1)是一种与Rad52和其他蛋白质一起在同源性重组期间起实现链交换,从而形成异源双链体DNA的作用的蛋白质,所述异源双链体DNA通过不匹配修复来解决以产生基因转换地带。在哺乳动物细胞中,已报道Rad51和Rad52过度表达会使同源性重组和基因转换的频率增加。参见Yanez和Porter

(1999) *Gene Ther.* 6:1282-1290以及Lambert和Lopez (2000) *EMBO J.* 19:3090-3099,其出于所有目的以引用的方式整体并入本文。

[0328] 调控NHEJ、基因转换和/或同源性引导的修复中涉及的基因的表达或活性的改变可具有空间特异性或时间特异性,并且也可为可诱导的或暂时的和可逆的。举例来说,可构建各种盒形式以允许在特定细胞或组织类型中,在特定发育阶段或在诱导后达成缺失。所述盒可采用重组酶系统,其中盒由重组酶识别位点侧接在两侧上,并且可使用在所需细胞类型中表达,在所需发育阶段表达,或在诱导后得以表达或活化的重组酶移除。所述盒可进一步构建来包括一系列各对不同重组酶识别位点,其被放置以致可产生无效等位基因、条件性等位基因、或组合条件性/无效等位基因,如出于所有目的以引用的方式整体并入本文的US 2011/0104799中所述。对重组酶基因的调控可以各种方式控制,诸如通过使重组酶基因可操作地连接至细胞特异性、组织特异性或发育调控启动子(或其他调控元件),或通过使重组酶基因可操作地连接至包含仅在特定细胞类型、组织类型或发育阶段中具有活性的miRNA的识别位点的3' -UTR。重组酶也可例如通过采用融合蛋白,将重组酶置于效应物或代谢物的控制下(例如CreER<sup>T2</sup>,其活性由他莫昔芬(tamoxifen)正性控制),或通过将重组酶基因置于诱导型启动子的控制下(例如活性由多西环素(doxycycline)和TetR或TetR变体控制的重组酶)来调控。各种盒形式和调控重组酶基因的手段的实例例如提供在US 8,518,392;US 8,354,389;和US 8,697,851中,所述专利各自以引用的方式整体并入本文。

[0329] 在其他本文公开的方法中,细胞已被修饰来通过阻断处于细胞周期的某一时期诸如细胞周期的M期或S期的细胞来使NHEJ活性增加或降低,或使基因转换或HDR活性增加。参见例如WO 2016/036754,其出于所有目的以引用的方式整体并入本文。这可用细胞周期阻断组合物实现。所述组合物的实例包括诺考达唑(nocodazole)、羟基脲(hydroxyurea);秋水仙碱(colchicine);地美可辛(demecolcine)(秋水仙胺(colcemid));洛伐他汀(lovastatin);含羞草氨酸(mimosine);胸苷;阿非迪霉素(aphidicolin);拉春库林A(latrunculin A);和拉春库林B。所述修饰可包括在调控NHEJ、基因转换和/或HDR中涉及的基因的表达或活性方面的改变。

#### [0330] G. 靶向遗传修饰的类型

[0331] 各种类型的靶向遗传修饰可使用本文所述的方法加以引入。所述靶向遗传修饰可包括使与目标外来靶抗原同源或共有目标表位的自身抗原的表达降低或消除的任何修饰。优选地,所述修饰破坏靶基因组基因座。破坏的实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,框移突变,截短突变,无效突变,或插入或缺失少数核苷酸(例如导致框移突变)。破坏可导致等位基因的失活(即功能丧失)或丧失。所述靶向遗传修饰可包括例如插入一个或多个核苷酸,缺失一个或多个核苷酸,或取代(替换)一个或多个核苷酸。所述插入、缺失或替换可例如导致点突变,敲除目标核酸序列或其一部分,敲入目标核酸序列或其一部分,用异源性或外源性核酸序列替换内源性核酸序列,改变调控元件(例如启动子或增强子),错义突变,无义突变,框移突变,截短突变,无效突变或其组合。举例来说,可改变(例如缺失、插入或取代)至少1、2、3、4、5、7、8、9、10个或更多个核苷酸以形成靶向遗传修饰。缺失、插入或替换可具有任何大小,如在本文中其他地方所公开。参见例如Wang等(2013) *Cell* 153:910-918;Mandalos等(2012) *PLoS One* 7:e45768;以及Wang等(2013) *Nat Biotechnol.* 31:530-532,其各自以引用的方式整体并入本文。所述突变可导致自身抗原的

表达(例如mRNA和/或蛋白质表达)降低或表达消除(例如等位基因的缺失)。

[0332] 靶向遗传修饰(例如插入、缺失或取代)可发生在靶基因组基因座中的一个或多个位置处。举例来说,如果使用两个外源性修复模板,那么靶向遗传修饰可包括在靶基因组基因座内的两个位置处的两个单独修饰。

[0333] 在其中使用外源性修复模板的方法中,举例来说,缺失可在5'靶标序列与3'靶标序列之间。在其中使用两个或更多个引导RNA的方法中,缺失可在第一引导RNA识别序列与第二引导RNA识别序列或第一Cas裂解位点与第二Cas裂解位点之间。所述缺失可为任何长度。经缺失核酸可为例如约1bp至约5bp、约5bp至约10bp、约10bp至约50bp、约50bp至约100bp、约100bp至约200bp、约200bp至约300bp、约300bp至约400bp、约400bp至约500bp、约500bp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、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约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。

[0334] 或者,经缺失核酸可为例如至少1bp、至少5bp、至少10bp、至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少1kb、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或更大。在一些情况下,经缺失核酸可为至少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(例如染色体的大部分)。

[0335] 在一特定实例中,缺失大小可在约0.1kb与约1Mb之间,在约0.1kb与约900kb之间,在约0.1kb与约400kb之间,在约0.1kb与约200kb之间,在约0.1kb与约100kb之间,或多达约1Mb,多达约900kb,多达约400kb,多达约200kb或多达约100kb。在一特定实例中,缺失大小可在约0.1-200、0.1-190、0.1-180、0.1-170、0.1-160、0.1-150、0.1-140、0.1-130、0.1-120、0.1-110、0.1-100、0.1-90、0.1-80、0.1-70、0.1-60、0.1-50、0.1-40、0.1-30、0.1-20、0.1-10、0.1-9、0.1-8、0.1-7、0.1-6、0.1-5、0.1-4、0.1-3、0.1-2、或0.1-1kb之间。所靶向细胞克隆诸如所靶向胚胎干细胞克隆中的双等位基因缺失(垮塌)效率(即具有双等位基因缺失的筛选克隆的百分比)可在约1-100%、1-90%、1-80%、1-70%、1-60%、1-50%、1-40%、1-30%、或1-27%之间,或可为至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%或25%。举例来说,在一个实施方案中,缺失大小是约50kb或更小,并且双等位基因缺失效率在约1-30%或1-27%之间,或缺失大小是约50kb或更大(例如在约50kb至约200kb之间),并且双等位基因缺失效率是约1-5%或1-3%。在其中靶向单细胞期胚胎的实验中,在单细胞期胚胎中进行CRISPR/Cas注射之后出生的活体幼仔中的双等位基因缺失(垮塌)效率(即具有双等位基因缺失的活体幼仔的百分比)可在约1-100%、1-90%、或1-85%之间,或可为至少约1%、2%、3%、4%、5%、6%、7%、8%、9%、10%、11%、12%、13%、14%、15%、16%、17%、18%、19%、20%、25%、30%、40%、50%、60%、70%、80%、或85%。举例来说,在一个实施方案中,缺失大小是约

50kb或更小,并且双等位基因缺失效率在约1-85%或20-85%之间,或缺失大小是约50kb或更大(例如在约50kb至约100kb之间),并且双等位基因缺失效率是约1-20%或1-15%。

[0336] 在其中使用外源性修复模板的方法中,举例来说,插入可在5' 靶标序列与3' 靶标序列之间。所述插入可具有任何长度。举例来说,插入核酸可为例如约1bp至约5bp、约5bp至约10bp、约10bp至约50bp、约50bp至约100bp、约100bp至约200bp、约200bp至约300bp、约300bp至约400bp、约400bp至约500bp、约500bp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250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约450kb、约450kb至约500kb或更大。或者,插入可为至少1bp、至少5bp、至少10bp、至少50bp、至少100bp、至少200bp、至少300bp、至少400bp、至少500bp、至少1kb、至少5kb、至少10kb、至少20kb、至少30kb、至少40kb、至少60kb、至少80kb、至少100kb、至少150kb、至少200kb、至少250kb、至少300kb、至少350kb、至少400kb、至少450kb或至少500kb。

[0337] 靶向遗传修饰可为精确修饰或不精确修饰。举例来说,在使用外源性修复模板的方法中,缺失可为精确缺失,其中经缺失核酸仅由5' 同源臂与3' 同源臂之间的核酸序列组成,以致在经修饰靶基因组基因座处不存在额外插入或缺失(插入缺失)。类似地,如果使用侧接于编码自身抗原的基因的整个编码区的成对gRNA,那么第一Cas蛋白裂解位点与第二Cas蛋白裂解位点之间的缺失可为精确缺失,其中经缺失核酸仅由第一Cas蛋白裂解位点与第二Cas蛋白裂解位点之间的核酸序列组成,以致在经修饰靶基因组基因座处不存在额外插入或缺失(插入缺失)。在其中使用外源性修复模板与侧接于目标区域的成对gRNA两者的方法中,缺失可为以上提及的精确缺失中的任一者。或者,第一Cas蛋白裂解位点与第二Cas蛋白裂解位点之间的缺失可为延伸超出第一Cas蛋白裂解位点和第二Cas蛋白裂解位点的不精确缺失,这符合通过非同源性末端接合(NHEJ)进行的不精确修复,从而在经修饰基因组基因座处产生额外缺失和/或插入。举例来说,缺失可延伸超出第一Cas蛋白裂解位点和第二Cas蛋白裂解位点约1、2、3、4、5、10、20、30、40、50、100、200、300、400或500bp或更多。同样,经修饰基因组基因座可包含符合通过NHEJ进行的不精确修复的额外插入,诸如插入约1、2、3、4、5、10、20、30、40、50、100、200、300、400、或500bp或更多。使用外源性修复模板(例如单链寡脱氧核苷酸(ssODN))以及CRISPR/Cas9可通过促进同源性引导的修复而非NHEJ来使精确修饰的机会增加。

[0338] 靶向修饰可包括用相应同源或直系同源序列替换在靶基因组基因座处的序列(例如编码自身抗原的基因的全部或一部分,诸如所述基因的编码自身抗原的特定区域或基序的一部分)。缺失编码自身抗原的基因的全部或一部分以及用缺乏在目标外来抗原与自身抗原之间共有的表位的相应同源或直系同源序列替换可导致表达自身抗原的同源物或直系同源物,其保留野生型自身抗原的功能,但缺乏存在于目标外来抗原上以及与野生型自身抗原共有的表位。或者或另外,靶向修饰可包括在靶基因组基因座(例如编码自身抗原的基因的全部或一部分)处的一个或多个点突变(例如1、2、3、4、5个或更多个)。所述点突变可例如起使自身抗原中与目标外来抗原共有的一个或多个表位的表达消除的作用。任选地,所述点突变可导致所编码多肽中的保守性氨基酸取代(例如天冬氨酸[Asp,D]被谷氨酸[Glu,E]取代)。所述氨基酸取代可导致表达保留野生型自身抗原的功能,但缺乏存在于目

标外来抗原上以及与野生型自身抗原共有的表位的自身抗原。

[0339] 本文所述的方法促进和增加双等位基因修饰并且特别是纯合性修饰的频率。特定来说,相较于使细胞与任一单独引导RNA接触,通过使细胞与靶向靶基因组基因座内的第一引导RNA识别序列和第二引导RNA识别序列的第一引导RNA和第二引导RNA接触,产生双等位基因修饰的效率可得以增加。通过使细胞与靶向靶基因组基因座内的引导RNA识别序列的第一引导RNA、第二引导RNA和第三引导RNA接触,或与靶向靶基因组基因座内的引导RNA识别序列的第一引导RNA、第二引导RNA、第三引导RNA和第四引导RNA接触,产生双等位基因修饰的效率也可得以增加。此外或或者,通过选择靶基因组基因座以使在一对同源染色体中相应第一染色体与第二染色体之间,在靶基因组基因座的全部或一部分中的序列同一性最大化,产生双等位基因修饰并且特别是纯合性修饰的效率可得以增加。用于选择所述靶基因组基因座的方法在本文中其他地方进一步详细描述。

[0340] 优选地,靶向遗传修饰是双等位基因修饰。双等位基因修饰包括其中对相应同源染色体(例如在二倍体细胞中)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的事件。同源染色体(即一对同源染色体)包括在相同基因座处具有相同基因,但可能具有不同等位基因的染色体(例如在减数分裂期间成对的染色体)。术语等位基因包括遗传序列的一种或多种替代性形式中的任一者。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。

[0341] 双等位基因修饰可产生就靶向遗传修饰来说的纯合性。纯合性包括靶基因组基因座的两个等位基因(即两个同源染色体上的相应等位基因)均具有靶向遗传修饰所处的情况。举例来说,双等位基因修饰可包括以下、基本上由以下组成或由以下组成:编码自身抗原的基因的全部或一部分的纯合性缺失,或双等位基因修饰可包括以下、基本上由以下组成或由以下组成:对编码自身抗原的基因的起始密码子的纯合性破坏,以致起始密码子不再具有功能性。

[0342] 或者,双等位基因修饰可产生就靶向修饰来说的复合杂合性(例如半合性)。复合杂合性包括靶基因组基因座的两个等位基因(即两个同源染色体上的等位基因)均已被修饰,但它们已以不同方式被修饰(例如一个等位基因中具有靶向修饰,而另一等位基因被失活或破坏)所处的情况。举例来说,在不具有靶向修饰的等位基因中,由Cas蛋白产生的双链断裂可通过非同源性末端接合(NHEJ)介导的DNA修复来修复,此产生包含核酸序列的插入或缺失的突变等位基因,并且由此导致对那个基因组基因座的破坏。举例来说,如果细胞具有一个具有靶向修饰的等位基因和另一不能够被表达的等位基因,那么双等位基因修饰可产生复合杂合性。复合杂合性包括半合性。半合性包括仅存在靶基因组基因座的一个等位基因(即两个同源染色体中的一者上的等位基因)所处的情况。举例来说,如果靶向修饰发生在一个等位基因中,伴有另一等位基因的相应丧失或缺失,那么双等位基因修饰可产生就所述靶向修饰来说的半合性。

[0343] 在一特定实例中,双等位基因修饰可包括一对第一同源染色体和第二同源染色体中第一引导RNA识别序列或Cas裂解位点与第二引导RNA识别序列或Cas裂解位点之间的纯合性缺失。或者,双等位基因修饰可包括一对第一同源染色体和第二同源染色体中第一引导RNA识别序列或Cas裂解位点与第二引导RNA识别序列或Cas裂解位点之间的双等位基因缺失(即两个染色体中的缺失,但未必在各自中具有相同缺失)。缺失可同时发生,或缺失可

初始在第一同源染色体中发生,纯合性接着由细胞使用第一同源染色体作为供体序列以通过同源性重组诸如通过基因转换修复第二同源染色体中的一个或多个双链断裂来实现。

[0344] 在另一特定实例中,双等位基因修饰可包括对一对第一同源染色体和第二同源染色体中的靶标基因的起始密码子区域的纯合性破坏。或者,对一对第一同源染色体和第二同源染色体中的靶标基因的起始密码子区域进行双等位基因破坏(即两个染色体中的破坏,但未必在各自中具有相同修饰)。修饰可同时发生,或修饰可初始在第一同源染色体中发生,纯合性接着由细胞使用第一同源染色体作为供体序列以通过同源性重组诸如通过基因转换修复第二同源染色体中的一个或多个双链断裂来实现。

[0345] 如果使用供体序列(例如外源性修复模板),那么双等位基因修饰可包括在一对第一同源染色体和第二同源染色体中第一引导RNA识别序列或Cas裂解位点与第二引导RNA识别序列或Cas裂解位点之间的缺失以及5'靶标序列与3'靶标序列之间的核酸插入物插入,由此产生纯合性经修饰基因组。或者,双等位基因修饰可包括在一对第一同源染色体和第二同源染色体中5'靶标序列与3'靶标序列之间的缺失以及5'靶标序列与3'靶标序列之间的核酸插入物插入,由此产生纯合性经修饰基因组。缺失和插入可同时发生在两个染色体中,或缺失和插入可初始发生在第一同源染色体中,纯合性接着由细胞使用第一同源染色体作为供体序列以通过同源性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来实现。举例来说,在不希望受任何特定理论束缚下,核酸插入物的插入可发生在第一同源染色体中(伴有或不伴有由Cas蛋白达成的裂解),并且第二同源染色体可接着通过由Cas蛋白对第二同源染色体的裂解刺激的基因转换事件来修饰。

[0346] 或者,如果外源性修复模板包含不具有核酸插入物的5'同源臂和3'同源臂,那么双等位基因修饰可包括一对第一同源染色体和第二同源染色体中5'靶标序列与3'靶标序列之间的缺失,由此产生纯合性经修饰基因组。缺失可同时发生在两个染色体中,或缺失可初始发生在第一同源染色体中,纯合性接着由细胞使用第一同源染色体作为供体序列以通过同源性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来实现。举例来说,在不希望受任何特定理论束缚下,缺失可发生在第一同源染色体中(伴有或不伴有由Cas蛋白达成的裂解),并且第二同源染色体可接着通过由Cas蛋白对第二同源染色体的裂解刺激的基因转换事件来修饰。

[0347] 第一引导RNA识别序列与第二引导RNA识别序列之间的缺失或5'靶标序列与3'靶标序列之间的缺失可为精确缺失,其中经缺失核酸仅由第一核酸酶裂解位点与第二核酸酶裂解位点之间的核酸序列组成,或仅由5'靶标序列与3'靶标序列之间的核酸序列组成,以致在经修饰基因组靶基因座处不存在额外缺失或插入。第一引导RNA识别序列与第二引导RNA识别序列之间的缺失也可为延伸超出第一核酸酶裂解位点和第二核酸酶裂解位点的不精确缺失,这符合通过非同源性末端接合(NHEJ)进行的不精确修复,从而在经修饰基因组基因座处产生额外缺失和/或插入。举例来说,缺失可延伸超出第一Cas蛋白裂解位点和第二Cas蛋白裂解位点约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或更多。同样,经修饰基因组基因座可包含符合通过NHEJ进行的不精确修复的额外插入,诸如插入约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或更多。

[0348] 通过使用外源性修复模板产生的靶向插入可具有任何大小。外源性修复模板中的核酸插入物的实例和核酸插入物的大小的实例在本文中其他地方描述。

[0349] 纯合性靶向遗传修饰是有利的,因为用于制备具有这些修饰的经遗传修饰的动物的方法(以下更详细描述)可更高效和耗时较少。在许多情况下,诸如移除或破坏基因以研究不存在它时的影响,仅有就靶向遗传修饰来说的杂合性(即在一个等位基因中具有修饰而对另一等位基因无变化)是不足够的。在常规靶向策略的情况下,就大型靶向基因组缺失来说是杂合的F0代动物可能是可获得的,但产生就缺失来说是纯合的F1代动物需要这些杂合性动物的后续异种交配。这些额外交配步骤是昂贵的和耗时的。能够产生就靶向遗传修饰来说是纯合的F0代经遗传修饰的动物会导致显著效率增加和时间节约,因为需要的交配步骤较少。

[0350] H. 鉴定具有靶向遗传修饰的细胞

[0351] 本文公开的方法可进一步包括鉴定具有经修饰靶标核酸(例如经修饰基因组)的细胞。各种方法可用于鉴定具有靶向遗传修饰诸如缺失或插入的细胞。所述方法可包括鉴定一个在靶基因组基因座处具有靶向遗传修饰的细胞。可进行筛选以鉴定具有经修饰基因组基因座的所述细胞。

[0352] 筛选步骤可包括用于评估亲本染色体的等位基因修饰(MOA)的定量测定(例如等位基因丧失(LOA)和/或等位基因获得(GOA)测定)。举例来说,定量测定可通过定量PCR诸如实时PCR(qPCR)来进行。实时PCR可利用识别靶基因组基因座的第一引物组和识别非靶向参照基因座的第二引物组。引物组可包含识别扩增的序列的荧光探针。

[0353] 为鉴定纯合性垮塌ES细胞克隆,相较于传统方法,TAQMAN<sup>®</sup>探针qPCR策略可以更大效率和准确性加以使用。由于包括“中间”LOA测定(参见例如图4中的mTM探针)以及不存在GOA测定,纯合性垮塌等位基因可用一个qPCR板鉴定。因为用于筛选ES细胞克隆的每个测定都是LOA测定,所以可在不使用任何非小鼠DNA校正物下准确计算每个测试区域的拷贝数。

[0354] 筛选步骤也可包括保留测定,其是一种用于区分核酸插入物向靶基因组基因座中的正确靶向插入与所述核酸插入物向所述靶基因组基因座的外部的基因组位置中的随机转基因插入的测定。保留测定也可用于区分正确缺失与延伸超出靶向缺失区域的缺失。用于筛选靶向修饰的常规测定诸如长程PCR或Southern印迹使所插入靶向载体与所靶向基因座相关联。然而,由于它们的同源臂大小较大,LTVEC不容许通过所述常规测定来筛选。为筛选LTVEC靶向,可使用等位基因修饰(MOA)测定,包括等位基因丧失(LOA)测定和等位基因获得(GOA)测定(参见例如US 2014/0178879以及Frendewey等(2010)Methods Enzymol.476:295-307,其各自出于所有目的以引用的方式整体并入本文)。等位基因丧失(LOA)测定使常规筛选逻辑倒转,并且对突变所针对的天然基因座的拷贝数定量。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者(针对不在X或Y染色体上的基因),另一等位基因由靶向修饰破坏。相同原理可相反地用作等位基因获得(GOA)测定以对所插入靶向载体的拷贝数定量。举例来说,组合使用GOA测定和LOA测定将会把正确靶向的杂合性克隆揭示为已丧失天然靶标基因的一个拷贝,并且获得药物抗性基因或其他插入标记的一个拷贝。

[0355] 作为一实例,定量聚合酶链式反应(qPCR)可用作等位基因定量的方法,但可可靠地区分零个、一个和两个靶标基因拷贝之间的差异或零个、一个和两个核酸插入物拷贝之

间的差异的任何方法都可用于开发MOA测定。举例来说,TAQMAN<sup>®</sup>可用于定量基因组DNA样品中的DNA模板的拷贝数,尤其是通过与参照基因进行比较(参见例如US 6,596,541,其出于所有目的以引用的方式整体并入本文)。对与一个或多个靶标基因或基因座在相同基因组DNA中的参照基因定量。因此,进行两个TAQMAN<sup>®</sup>扩增(各自具有它的相应探针)。一个TAQMAN<sup>®</sup>探针测定参照基因的“Ct”(阈值循环),而另一探针测定一个或多个通过成功靶向而被替换的所靶向基因或基因座的区域的Ct(即LOA测定)。Ct是反映各TAQMAN<sup>®</sup>探针的起始DNA的数量的量,即丰度较小的序列需要更多个PCR循环来达到阈值循环。使TAQMAN<sup>®</sup>反应的模板序列的拷贝数降低一半将导致增加约一个Ct单位。当相较于来自非靶向细胞的DNA时,在其中一个或多个靶标基因或基因座的一个等位基因已通过同源性重组而被替换的细胞的情况下的TAQMAN<sup>®</sup>反应将导致靶标TAQMAN<sup>®</sup>反应增加一个Ct,而参照基因的Ct不增加。对于GOA测定,另一TAQMAN<sup>®</sup>探针可用于测定通过成功靶向而替换一个或多个所靶向基因或基因座的核酸插入物的Ct。

[0356] 因为成对gRNA可在靶基因组基因座处产生大型Cas介导的缺失,所以可适用的是使标准LOA和GOA测定加强以验证由LTVEC达成的正确靶向(即在除单细胞期胚胎以外的细胞中)。举例来说,单独LOA和GOA测定可能不区分正确靶向的细胞克隆与其中靶基因组基因座的大型Cas诱导缺失符合LTVEC随机整合在基因组中其他地方的克隆,特别是如果GOA测定采用针对LTVEC插入物内的选择盒的探针。因为所靶向细胞中的选择压力基于选择盒,所以LTVEC在基因组中其他地方进行的随机转基因整合将通常包括LTVEC的选择盒和邻近区域,但将排除LTVEC的更远端区域。举例来说,如果LTVEC的一部分随机整合至基因组中,并且LTVEC包含长度约5kb或更多,具有邻近于3'同源臂的选择盒的核酸插入物,那么通常3'同源臂而非5'同源臂将与所述选择盒一起以转基因方式整合。或者,如果选择盒邻近于5'同源臂,那么通常5'同源臂而非3'同源臂将与选择盒一起以转基因方式整合。作为一实例,如果LOA测定和GOA测定用于评估LTVEC的靶向整合,并且GOA测定利用针对选择盒的探针,那么在靶基因组基因座处的杂合性缺失与LTVEC的随机转基因整合组合将与LTVEC在靶基因组基因座处的杂合性靶向整合给出相同读出。为验证由LTVEC达成的正确靶向,可单独或与LOA和/或GOA测定联合使用保留测定。

[0357] 保留测定对5'靶标序列(对应于LTVEC的5'同源臂)和/或3'靶标序列(对应于LTVEC的3'同源臂)中的DNA模板的拷贝数进行测定。特定来说,测定对应于与选择盒邻近的同源臂的靶标序列中的DNA模板的拷贝数是适用的。在二倍体细胞中,大于2的拷贝数通常指示LTVEC随机转基因整合在靶基因组基因座的外部而非在靶基因组基因座处,此是不合需要的。正确靶向的克隆将保留拷贝数2。此外,在所述保留测定中拷贝数小于2通常指示存在延伸超出靶向缺失区域的大型Cas介导的缺失,此也是不合需要的。

[0358] 在用于鉴定核酸插入物在二倍体细胞中的靶基因组基因座处的靶向插入的一示例性保留测定中,首先从具有已与大型靶向载体(LTVEC)接触的基因组的细胞获得DNA,所述大型靶向载体包含由杂交至第一靶标序列的第一同源臂和杂交至第二靶标序列的第二同源臂侧接的核酸插入物,其中所述核酸插入物包含邻近于所述第一同源臂的选择盒。任

选地,选择盒可包含药物抗性基因。然后使DNA暴露于在第一靶标序列内结合的探针、在核酸插入物内结合的探针和在具有已知拷贝数的参照基因内结合的探针,其中各探针在结合后产生可检测信号。接着检测来自各探针的结合的信号。将来自参照基因探针的信号与来自第一靶标序列探针的信号进行比较以测定第一靶标序列的拷贝数,并且将来自参照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测定核酸插入物的拷贝数。核酸插入物拷贝数1或2以及第一靶标序列拷贝数2通常指示核酸插入物靶向插入在靶基因组基因座处,而核酸插入物拷贝数1或更多以及第一靶标序列拷贝数3或更多通常指示核酸插入物随机插入在除靶基因组基因座以外的基因组基因座处。

[0359] 来自第一靶标序列探针的结合的信号可用于测定第一靶标序列的阈值循环(Ct)值,来自参照基因探针的结合的信号可用于测定参照基因的阈值循环(Ct)值,并且第一靶标序列的拷贝数可通过将第一靶标序列Ct值与参照基因Ct值进行比较来测定。同样,来自核酸插入物探针的结合的信号可用于测定核酸插入物的阈值循环(Ct)值,并且核酸插入物的拷贝数可通过将第一靶标序列Ct值与参照基因Ct值进行比较来测定。

[0360] LTVEC中的核酸插入物可为例如至少5、10、20、30、40、50、60、70、80、90、100、150、200、250、300、350、400、450或500kb。探针在第一靶标序列和选择盒中所结合的序列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。

[0361] 所述方法可进一步包括额外保留测定以对第二靶标序列的拷贝数进行测定。举例来说,所述方法可进一步包括使细胞的DNA暴露于结合第二靶标序列的探针,检测来自第二靶标序列探针的结合的信号,以及将来自参照基因探针的信号与来自第二靶标序列探针的信号进行比较以测定第二靶标序列的拷贝数。

[0362] 同样,所述方法可进一步包括额外GOA测定以对核酸插入物内的一个或多个额外序列的拷贝数进行测定。举例来说,所述方法可进一步包括使细胞的DNA暴露于一个或多个结合核酸插入物的额外探针,检测来自一个或多个额外探针的结合的信号,以及将来自参照基因探针的信号与来自一个或多个额外核酸插入物探针的信号进行比较以测定核酸插入物内的一个或多个额外序列的拷贝数。

[0363] 同样,当LTVEC被设计来使内源性序列从靶基因组基因座缺失时,或当使用成对gRNA(例如以在单一基因组靶基因座内的不同位点处产生成对双链断裂以及使间插内源性序列缺失)时,所述方法可进一步包括LOA测定以对在靶基因组基因座处的内源性序列的拷贝数进行测定。举例来说,所述方法可进一步包括使细胞的DNA暴露于结合在靶基因组基因座处的内源性序列的探针,检测来自内源性序列探针的结合的信号,以及将来自参照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测定内源性序列的拷贝数。

[0364] 保留测定也可用于其中使用成对gRNA,但未必使用外源性修复模板的实验中。因为成对gRNA可在靶基因组基因座处产生大型Cas介导的缺失,所以可适用的是使标准LOA测定加强以与由于在NHEJ修复之后的插入缺失而延伸超出靶向缺失区域的缺失相对比来验证由成对gRNA达成的正确靶向性缺失。

[0365] 保留测定对包含第一引导RNA识别序列(即5'引导RNA识别序列)和/或在第一引导RNA识别序列的上游的区域和/或包含第二引导RNA识别序列(即3'引导RNA识别序列)和/或

在第二引导RNA识别序列的下游以及邻近于第二引导RNA识别序列的区域中的DNA模板的拷贝数进行测定。在二倍体细胞中,拷贝数小于1将指示存在延伸超出靶向缺失区域的大型NHEJ介导的缺失,此是不合需要的。正确靶向的克隆将保留拷贝数2。用以测定拷贝数的探针可例如在引导RNA识别序列的约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内。

[0366] 适合定量测定的其他实例包括荧光介导的原位杂交(FISH)、比较基因组杂交、等温DNA扩增、与固定化探针定量杂交、INVADER<sup>®</sup>探针、TAQMAN<sup>®</sup>分子信标探针或ECLIPSE<sup>™</sup>探针技术(参见例如US 2005/0144655,其出于所有目的以引用的方式整体并入本文)。也可使用用于筛选靶向修饰的常规测定,诸如长程PCR、Southern印迹或Sanger测序。所述测定通常用于获得所插入靶向载体与所靶向基因组基因座之间存在关联的证据。举例来说,对于长程PCR测定,一个引物可识别所插入DNA内的序列,而另一引物识别超出靶向载体的同源臂的末端的靶基因组基因座序列。

[0367] 下一代测序(NGS)也可用于筛选,特别是在已被修饰的单细胞期胚胎的情况下。下一代测序也可被称为“NGS”或“大规模平行测序”或“高通量测序”。除MOA测定和保留测定之外,所述NGS也可用作筛选工具以确定靶向遗传修饰的确切性质以及检测镶嵌现象。镶嵌现象是指在一个已从单一受精卵(即接合子)发育的个体中存在两个或更多个具有不同基因型的细胞群体。在本文公开的方法中,并非必须使用选择标记筛选所靶向克隆。举例来说,可在不使用选择盒下依赖于本文所述的MOA测定和NGS测定。

[0368] 也可针对与目标外来抗原同源或共有目标表位的自身抗原的表达降低或消除来筛选所靶向细胞。举例来说,如果自身抗原是蛋白质,那么可通过用于测定蛋白质表达的任何已知技术来评估表达,所述技术包括例如Western印迹分析或蛋白质免疫染色。

[0369] III. 制备经遗传修饰的非人动物的方法

[0370] 可采用各种本文公开的方法产生经遗传修饰的非人动物。用于产生经遗传修饰的生物体的任何适宜方法或方案,包括本文所述的方法,都适于产生这种经遗传修饰的非人动物。以对多能细胞诸如胚胎干(ES)细胞进行遗传修饰起始的所述方法通常包括:(1)使用本文所述的方法修饰不是单细胞期胚胎的多能细胞的基因组;(2)鉴定或选择经遗传修饰的多能细胞;(3)将所述经遗传修饰的多能细胞引入宿主胚胎中;和(4)在代孕母体中植入和孕育包含所述经遗传修饰的多能细胞的所述宿主胚胎。代孕母体可接着产生包含靶向遗传修饰以及能够通过种系传递靶向遗传修饰的F0代非人动物。携带经遗传修饰的基因组基因座的动物可通过如本文所述的等位基因修饰(MOA)测定来鉴定。可将供体细胞引入处于任何时期诸如胚泡期或前桑椹胚期(即4细胞期或8细胞期)的宿主胚胎中。产生能够通过种系传递遗传修饰的子代。多能细胞可为例如如在本文中其他地方讨论的ES细胞(例如啮齿动物ES细胞、小鼠ES细胞或大鼠ES细胞)。参见例如美国专利号7,294,754,其出于所有目的以引用的方式整体并入本文。

[0371] 或者,以对单细胞期胚胎进行遗传修饰起始的所述方法通常包括:(1)使用本文所述的方法修饰单细胞期胚胎的基因组;(2)鉴定或选择经遗传修饰的胚胎;和(3)在代孕母体中植入和孕育所述经遗传修饰的胚胎。代孕母体可接着产生包含靶向遗传修饰以及能够通过种系传递靶向遗传修饰的F0代非人动物。携带经遗传修饰的基因组基因座的动物可通

过如本文所述的等位基因修饰(MOA)测定来鉴定。

[0372] 核转移技术也可用于产生非人哺乳动物。简要说,用于核转移的方法可包括以下步骤:(1)使卵母细胞去核或提供去核卵母细胞;(2)分离或提供待与所述去核卵母细胞组合的供体细胞或核;(3)将所述细胞或核插入所述去核卵母细胞中以形成重构细胞;(4)将所述重构细胞植入非人动物的子宫中以形成胚胎;和(5)使所述胚胎发育。在所述方法中,卵母细胞通常从死亡动物检索,但它们也可从活体动物的输卵管和/或卵巢分离。可在去核之前使卵母细胞在为本领域普通技术人员所知的多种培养基中成熟。可以为本领域普通技术人员所熟知的许多方式进行卵母细胞去核。可通过在融合之前将供体细胞显微注射在透明带下来将供体细胞或核插入去核卵母细胞中以形成重构细胞。融合可通过跨越接触/融合平面施加DC电脉冲(电融合),通过使细胞暴露于融合促进化学物质诸如聚乙二醇,或通过失活病毒诸如仙台病毒(Sendai virus)的方式来诱导。重构细胞可在使核供体和接受卵母细胞融合之前、期间和/或之后通过电学和/或非电学手段来活化。活化方法包括电脉冲、化学诱导冲击、通过精子来渗透、增加卵母细胞中的二价阳离子的水平、以及降低卵母细胞中的细胞蛋白质的磷酸化(如通过激酶抑制剂的方式)。活化的重构细胞或胚胎可在为本领域普通技术人员所熟知的培养基中培养,接着转移至动物的子宫中。参见例如US 2008/0092249、WO 1999/005266、US 2004/0177390、WO 2008/017234和美国专利号7,612,250,其各自出于所有目的以引用的方式整体并入本文。

[0373] 各种本文提供的方法允许产生经遗传修饰的非人F0动物,其中所述经遗传修饰的F0动物的细胞包含靶向遗传修饰。应认识到视用于产生F0动物的方法而定,F0动物内具有靶向遗传修饰的细胞的数目将变化。通过例如**VELOCIMOUSE<sup>®</sup>**方法来将供体ES细胞引入来自相应生物体的前桑椹胚期胚胎(例如8细胞期小鼠胚胎)中使得F0动物的细胞群体中有较大百分比包含具有靶向遗传修饰的细胞。举例来说,非人F0动物的至少50%、60%、65%、70%、75%、85%、86%、87%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的细胞贡献可由具有靶向遗传修饰的细胞群体组成。此外,F0动物的至少一个或多个生殖细胞可具有靶向遗传修饰。

[0374] A. 非人动物和细胞的类型

[0375] 本文提供的方法采用非人动物和细胞以及来自非人动物的胚胎。所述非人动物优选是哺乳动物,诸如啮齿动物(例如大鼠、小鼠和仓鼠)。其他非人哺乳动物包括例如人、非人灵长类动物、猴、猿、猫、狗、兔、马、公牛、鹿、野牛、家畜(例如牛物种诸如母牛、阉牛等;羊物种诸如绵羊、山羊等;以及猪物种诸如猪和公猪)。术语“非人”排除人。在一些本文提供的方法中,非人动物和细胞以及来自非人动物的胚胎是杂交物。

[0376] 本文提供的方法中采用的非人动物细胞可为例如全能细胞或多能细胞(例如胚胎干(ES)细胞诸如啮齿动物ES细胞、小鼠ES细胞或大鼠ES细胞)。全能细胞包括可产生任何细胞类型的未分化细胞,并且多能细胞包括具有发育成超过一种分化细胞类型的能力的未分化细胞。所述多能和/或全能细胞可为例如ES细胞或ES样细胞,诸如诱导多能干(iPS)细胞。ES细胞包括在引入胚胎中后能够促成发育胚胎的任何组织的胚胎源性全能或多能细胞。ES细胞可源于胚泡的内细胞群,并且能够分化成三个脊椎动物胚层(内胚层、外胚层和中胚层)中的任一者的细胞。

[0377] 本文提供的方法中采用的非人动物细胞也可包括单细胞期胚胎(即受精卵母细胞

或接合子)。单细胞期胚胎是通过两个配子之间的受精事件来形成的真核细胞。所述单细胞期胚胎可来自任何遗传背景(例如BALB/c、C57BL/6、129或其组合),可为新鲜的或冷冻的,并且可源于自然交配或体外受精。

[0378] 本文提供的方法中采用的小鼠和小鼠细胞可例如来自129品系、C57BL/6品系、BALB/c品系、Swiss Webster品系、129品系和C57BL/6品系的混合物、BALB/c品系和C57BL/6品系的混合物、129品系和BALB/c品系的混合物、以及BALB/c品系、C57BL/6品系和129品系的混合物。举例来说,本文提供的方法中采用的小鼠或小鼠细胞可至少部分地来自BALB/c品系(例如至少约25%、至少约50%、至少约75%源于BALB/c品系,或约25%、约50%、约75%或约100%源于BALB/c品系)。在一个实例中,小鼠或小鼠细胞可具有包含50%BALB/c、25%C57BL/6和25%129的品系。或者,小鼠或小鼠细胞可包括排除BALB/c的品系或品系组合。在所述小鼠中,BALB/c背景不为产生针对目标外来抗原的足够抗原结合蛋白谱系所需。

[0379] 129品系的实例包括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等(1999) *Mammalian Genome* 10(8):836,其出于所有目的以引用的方式整体并入本文。C57BL品系的实例包括C57BL/A、C57BL/An、C57BL/GrFa、C57BL/Ka1\_wN、C57BL/6、C57BL/6J、C57BL/6ByJ、C57BL/6NJ、C57BL/10、C57BL/10ScSn、C57BL/10Cr和C57BL/01a。本文提供的方法中采用的小鼠和小鼠细胞也可来自以上提及的129品系和以上提及的C57BL/6品系的混合物(例如50%129和50%C57BL/6)。同样,本文提供的方法中采用的小鼠和小鼠细胞可来自以上提及的129品系的混合物或以上提及的BL/6品系的混合物(例如129S6(129/SvEvTac)品系)。小鼠ES细胞的一特定实例是VGF1小鼠ES细胞。VGF1小鼠ES细胞(也称为F1H4)源于通过使雌性C57BL/6NTac小鼠与雄性129S6/SvEvTac小鼠杂交产生的杂交胚胎。参见例如Auerbach等(2000) *Biotechniques* 29,1024-1028,其出于所有目的以引用的方式整体并入本文。

[0380] 本文提供的方法中采用的小鼠和小鼠细胞也可具有MHC单倍型的任何组合。MHC分子的功能在于结合外来肽片段,并且将它们展示在细胞表面上以由适当T细胞识别。举例来说,小鼠和小鼠细胞可包含MHC<sup>b</sup>单倍型(例如C57BL/6)、MHC<sup>d</sup>单倍型(例如BALB/c),或可包含MHC<sup>b</sup>与MHC<sup>d</sup>两者(例如C57BL/6和BALB/c的组合)。所述MHC组合可导致抗体效价增加。

[0381] 本文提供的方法中采用的大鼠或大鼠细胞可来自任何大鼠品系,包括例如ACI大鼠品系、Dark Agouti (DA) 大鼠品系、Wistar大鼠品系、LEA大鼠品系、Sprague Dawley (SD) 大鼠品系或Fischer大鼠品系诸如Fisher F344或Fisher F6。大鼠或大鼠细胞也可从源于两个或更多个以上叙述的品系的混合物的品系获得。举例来说,大鼠或大鼠细胞可来自DA品系或ACI品系。ACI大鼠品系被表征为具有黑色环纹,具有白色腹部和足部以及RT1<sup>av1</sup>单倍型。所述品系可从包括Harlan Laboratories的多种来源获得。来自ACI大鼠的大鼠ES细胞系的一实例是ACI.G1大鼠ES细胞。Dark Agouti (DA) 大鼠品系被表征为具有环纹皮毛和RT1<sup>av1</sup>单倍型。所述大鼠可从包括Charles River和Harlan Laboratories的多种来源获得。来自DA大鼠的大鼠ES细胞系的实例是DA.2B大鼠ES细胞系和DA.2C大鼠ES细胞系。在一些情况下,大鼠或大鼠细胞来自近交大鼠品系。参见例如US 2014/0235933 A1,其出于所有目的以引用的方式整体并入本文。在其他情况下,大鼠或大鼠细胞来自杂交大鼠品系。

[0382] 已被植入宿主胚胎中的细胞可被称为“供体细胞”。供体细胞可与宿主胚胎来自相

同品系,或与宿主胚胎来自不同品系。同样,代孕母体可与供体细胞和/或宿主胚胎来自相同品系,或代孕母体可与供体细胞和/或宿主胚胎来自不同品系。

[0383] 多种宿主胚胎可用于本文公开的方法和组合物中。举例来说,可将供体细胞(例如供体ES细胞)引入来自相应生物体的前桑椹胚期胚胎(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,其各自出于所有目的以引用的方式整体并入本文。在其他方法中,可将供体细胞植入处于2细胞期、4细胞期、8细胞期、16细胞期、32细胞期或64细胞期的宿主胚胎中。宿主胚胎也可为胚泡,或可为前胚泡胚胎、前桑椹胚期胚胎、桑椹胚期胚胎、未致密桑椹胚期胚胎或致密桑椹胚期胚胎。当采用小鼠胚胎时,宿主胚胎时期可为蒂勒期1(Theiler Stage 1,TS1)、TS2、TS3、TS4、TS5和TS6,参照Theiler(1989)“The House Mouse:Atlas of Mouse Development,”Springer-Verlag, New York中所述的蒂勒期,所述文献出于所有目的以引用的方式整体并入本文。举例来说,蒂勒期可选自TS1、TS2、TS3和TS4。在一些方法中,宿主胚胎包含透明带,并且供体细胞是穿过透明带中的孔引入宿主胚胎中的ES细胞。在其他方法中,宿主胚胎是无带胚胎。在其他方法中,桑椹胚期宿主胚胎是聚集的。

[0384] B. 用于产生抗原结合蛋白的非人动物

[0385] 用于本文提供的方法中的非人动物可为能够产生抗原结合蛋白的任何非人动物,诸如哺乳动物、啮齿动物、大鼠或小鼠。举例来说,可使用被遗传修饰来使抗体产生最优化的非人动物(例如啮齿动物,诸如大鼠或小鼠)。所述非人动物可为被工程化来有助于大规模生产可用作人治疗剂的抗体的非人动物,包括包含人源化免疫球蛋白基因座的非人动物。举例来说,非人动物(例如啮齿动物,诸如大鼠或小鼠)可在它的种系中包含一个或多个以下修饰:非人动物(例如啮齿动物,诸如大鼠或小鼠)重链可变区基因座完全或部分被人重链可变基因座替换;非人动物(例如啮齿动物,诸如大鼠或小鼠) $\kappa$ 轻链可变区基因座完全或部分被人 $\kappa$ 轻链可变区基因座替换;非人动物(例如啮齿动物,诸如大鼠或小鼠) $\lambda$ 轻链可变区基因座完全或部分被人 $\lambda$ 轻链可变区基因座替换;并且重链和轻链可变区基因座完全被它们的人同源物或直系同源物替换。非人动物(例如啮齿动物,诸如大鼠或小鼠)也可在它的种系中包含一个或多个以下修饰:完全人重链和轻链可变区基因座可操作地连接至非人动物(例如啮齿动物,诸如大鼠或小鼠)恒定区核酸序列以使所述非人动物(例如啮齿动物,诸如大鼠或小鼠)产生包含人可变结构域融合于非人动物(例如啮齿动物,诸如大鼠或小鼠)恒定结构域的B细胞或抗体;或人重链和/或轻链可变区可操作地连接至非人动物(例如啮齿动物,诸如大鼠或小鼠)恒定区核酸序列以使所述非人动物(例如啮齿动物,诸如大鼠或小鼠)产生包含人可变结构域融合于非人动物(例如啮齿动物,诸如大鼠或小鼠)恒定区的B细胞或抗体。作为一实例,可使用VELOCIMMUNE<sup>®</sup>小鼠。参见例如US 6,596,541、US 8,791,323、US 8,895,802、US 8,895,801、US 7,105,348、US 2002/0106629、US 2007/0061900、US 2011/0258710、US 2011/0283376、US 2013/0210137、US 2014/0017781、US 2014/0020124、US 2014/0020125、US 2014/0017782、US 2014/0018522、US 2014/0033337、US 2014/0033336、US 2014/0041068、US 2014/0073010、US 2014/0023637、US 2014/0017238、US 2014/0013457、US 2014/0017229、US 2002/0183275、US 8,502,018、US 2012/0322108、US 2013/0254911、US 2014/0213773、US 2015/0201589、US 2015/0210776、US 2014/0017228、US 8,642,835、US 8,697,940以及Murphy等(2014)

Proc.Natl.Acad.Sci.U.S.A.111 (14):5153-5158,其各自出于所有目的以引用的方式整体并入本文。**VELOCIMMUNE**<sup>®</sup>小鼠含有在内源性基因座处用相应人免疫球蛋白可变区对编码小鼠免疫球蛋白重链(IgH)和免疫球蛋白轻链(例如κ轻链,Igκ)的种系可变区的精确大规模替换。这个精确替换产生具有制备具有人可变区和小鼠恒定区的重链和轻链的杂合免疫球蛋白基因座的小鼠。对小鼠V<sub>H</sub>-D<sub>H</sub>-J<sub>H</sub>和Vκ-Jκ区段的精确替换保持侧接小鼠序列在杂合免疫球蛋白基因座处完整并具有功能性。所述小鼠的体液免疫系统如同野生型小鼠的体液免疫系统起作用。B细胞发育在任何重要方面都未受阻碍,并且在抗原激发后在小鼠中产生丰富多样性的人可变区。

[0386] 上述非人动物(例如啮齿动物,诸如大鼠或小鼠)(例如**VELOCIMMUNE**<sup>®</sup>小鼠)也可在它们的种系中包含编码非人动物(例如啮齿动物,诸如大鼠或小鼠)ADAM6基因或其同源物或直系同源物或功能性片段的功能性异位核酸序列。举例来说,这种非人动物(例如啮齿动物,诸如大鼠或小鼠)可缺乏功能性内源性ADAM6基因,并且包含功能性异位核酸序列以弥补非人动物(例如啮齿动物,诸如大鼠或小鼠)ADAM6功能的丧失。举例来说,功能性异位序列可包含一个或多个Adam6基因,诸如小鼠Adam6a基因、小鼠Adam6b基因、或Adam6a基因与Adam6b基因两者。异位核酸序列可存在于人重链可变区基因座处或其他地方。参见例如US 2012/0322108;US 2013/0254911;US 2014/0213773;US 2015/0201589;US 2015/0210776;US 2014/0017228;和US 2013/0198879,其各自出于所有目的以引用的方式整体并入本文。

[0387] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括被遗传修饰来从有限人轻链可变区基因区段谱系表达有限人轻链可变结构域谱系或单一人轻链可变结构域的非人动物(例如啮齿动物,诸如大鼠或小鼠)。所述非人动物产生“通用轻链”或“共同轻链”,并且可适用于制备双特异性抗体。参见例如US 2011/0195454;US 2012/0021409;US 2012/0192300;US 2015/0059009;US 2013/0045492;US 2013/0198880;US 2013/0185821;US 2013/0302836;US 2013/0247234;US 2014/0329711;和US 2013/0198879,其各自出于所有目的以引用的方式整体并入本文。举例来说,非人动物(例如啮齿动物,诸如大鼠或小鼠)可被遗传工程化来包括单一未重排人轻链可变区基因区段(或两个人轻链可变区基因区段),其重排以形成表达单一轻链(或表达两个轻链中的任一者或两者)的重排人轻链可变区基因(或两个重排轻链可变区基因)。重排人轻链可变结构域能够与由非人动物(例如啮齿动物,诸如大鼠或小鼠)选择的多个亲和力成熟的人重链配对,其中重链可变区特异性结合不同表位。

[0388] 为实现有限轻链选项谱系,非人动物(例如啮齿动物,诸如大鼠或小鼠)可被工程化来致使它的制备或重排天然非人动物(例如啮齿动物,诸如大鼠或小鼠)轻链可变结构域的能力具有非功能性或大致上非功能性。这可例如通过使非人动物(例如啮齿动物,诸如大鼠或小鼠)的轻链可变区基因区段缺失来实现。内源性非人动物(例如啮齿动物,诸如大鼠或小鼠)基因座可接着通过所选外源性适合人轻链可变区基因区段来修饰,所述人轻链可变区基因区段以使得所述外源性人可变区基因区段可重排以及与内源性非人动物(例如啮齿动物,诸如大鼠或小鼠)轻链恒定区基因重组并形成重排反向嵌合轻链基因(人可变,非人动物(例如啮齿动物,诸如大鼠或小鼠)恒定)的方式可操作地连接至非人动物(例如啮齿

动物,诸如大鼠或小鼠)轻链恒定区。

[0389] 上述非人动物(例如啮齿动物,诸如大鼠或小鼠)(例如“通用轻链”或“共同轻链”)也可在它们的种系中包含编码非人动物(例如啮齿动物,诸如大鼠或小鼠)ADAM6基因或其同源物或直系同源物或功能性片段的功能性异位核酸序列。类似地,任何本文所述的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)都也可在它们的种系中包含编码非人动物(例如啮齿动物,诸如大鼠或小鼠)ADAM6基因或其同源物或直系同源物或功能性片段的功能性异位核酸序列。举例来说,这种非人动物(例如啮齿动物,诸如大鼠或小鼠)可缺乏功能性内源性ADAM6基因,并且包含功能性异位核酸序列以弥补非人动物(例如啮齿动物,诸如大鼠或小鼠)ADAM6功能的丧失。异位核酸序列可存在于人重链可变区基因座处或其他地方。参见例如US 2012/0322108;US 2013/0254911;US 2014/0213773;US 2015/0201589;US 2015/0210776;US 2014/0017228;和US 2013/0198879,其各自出于所有目的以引用的方式整体并入本文。

[0390] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系中包含可操作地连接至重链恒定区核酸序列的未重排轻链V区段和未重排J区段的非人动物。参见例如US 2012/0096572、US 2014/0130194和US 2014/0130193,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一个实例是其种系基因组包含经修饰内源性免疫球蛋白重链基因座的非人动物,所述经修饰内源性免疫球蛋白重链基因座包含用包含多个未重排人免疫球蛋白轻链可变(V<sub>κ</sub>)基因区段和多个未重排人免疫球蛋白轻链接合(J<sub>κ</sub>)基因区段,并且可操作地连接至内源性非人动物免疫球蛋白重链恒定(C<sub>H</sub>)区的核苷酸序列对在非人动物免疫球蛋白重链基因座处的所有功能性内源性非人动物免疫球蛋白重链可变(V<sub>H</sub>)基因区段、所有功能性内源性非人动物免疫球蛋白重链多样性(D<sub>H</sub>)基因区段和所有功能性内源性非人动物免疫球蛋白重链接合(J<sub>H</sub>)基因区段的替换,其中所述多个未重排人免疫球蛋白轻链V基因区段和所述多个未重排人免疫球蛋白轻链J基因区段在B细胞发育期间参与B细胞中的重排以形成重排人免疫球蛋白轻链V<sub>κ</sub>/J<sub>κ</sub>基因序列,其可操作地连接至在所述经修饰内源性重链基因座处的内源性非人动物免疫球蛋白重链C<sub>H</sub>区。这种非动物的另一实例是在它的种系中包含可操作地连接至在内源性非人动物重链基因座处的内源性非人动物重链恒定区的第一未重排人κ轻链可变(V<sub>κ</sub>)基因区段和未重排人κ轻链接合(J<sub>κ</sub>)基因区段的非人动物,其中所述第一未重排人V<sub>κ</sub>基因区段和所述未重排人J<sub>κ</sub>基因区段替换所有功能性内源性非人动物重链可变(V<sub>H</sub>)基因区段、所有功能性内源性非人动物多样性(D<sub>H</sub>)基因区段和所有功能性内源性非人动物重链接合(J<sub>H</sub>)基因区段,其中所述第一未重排人V<sub>κ</sub>基因区段和所述未重排人J<sub>κ</sub>基因区段在所述非人动物中参与重排以形成可操作地连接至内源性非人动物重链恒定区的重排V<sub>κ</sub>/J<sub>κ</sub>序列,并且其中所述非动物在它的种系中进一步包含可操作地连接至非人动物轻链恒定基因的第二人轻链可变(V<sub>L</sub>)基因区段和人轻链接合(J<sub>L</sub>)基因区段。这种非动物的另一实例是其基因组包含以下的非人动物:(a)内源性免疫球蛋白重链基因座,其被修饰来包含用第一多个未重排人轻链可变(V<sub>κ</sub>)基因区段和第一多个未重排人轻链接合(J<sub>κ</sub>)基因区段对在非人动物免疫球蛋白重链基因座处的所有功能性内源性非人动物免疫球蛋白重链可变(V<sub>H</sub>)基因区段、所有功能性内源性非人动物免疫球蛋白重链多样性(D<sub>H</sub>)基因区段和所有功能性内源性非人动物免疫球蛋白重链接合(J<sub>H</sub>)基因区段的替换,其中所述第一多个未重排人免疫球蛋白轻链V<sub>κ</sub>和J<sub>κ</sub>基因区段可

操作地连接至在内源性免疫球蛋白重链基因座处的内源性重链恒定 ( $C_H$ ) 区,并且在B细胞发育期间参与B细胞中的重排以形成可操作地连接至内源性非人动物 $C_H$ 区核酸序列的第一重排人轻链 $V_K/J_K$ 基因序列;和(b)经修饰免疫球蛋白轻链基因座,其包含可操作地连接至在内源性非人动物轻链基因座处的内源性非人动物轻链恒定 ( $C_K$ ) 区核酸序列的第二多个未重排人轻链可变 ( $V_K$ ) 基因区段和第二多个未重排人轻链接合 ( $J_K$ ) 基因区段,其中所述第二多个未重排人免疫球蛋白轻链 $V_K$ 和 $J_K$ 基因区段替换在内源性链基因座处的所有功能性内源性非人动物轻链可变 ( $V_K$ ) 基因区段和所有功能性内源性非人动物轻链接合 ( $J_K$ ) 基因区段,并且在B细胞发育期间参与B细胞中的重排以形成可操作地连接至内源性非人动物 $C_K$ 区核酸序列的第二重排人免疫球蛋白轻链 $V_K/J_K$ 区基因序列。

[0391] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系基因组中包含免疫球蛋白重链基因座的非人动物,所述免疫球蛋白重链基因座包含可操作地连接至内源性非人动物免疫球蛋白恒定区基因序列的重排人免疫球蛋白重链可变区核苷酸序列。参见例如US 2015/0020224、US 2014/0245468、US 2016/0100561、US 9,204,624和US 14/961,642,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一个实例是在它的种系基因组中在内源性免疫球蛋白重链基因座处包含可操作地连接至内源性重链恒定区基因序列的重排人免疫球蛋白重链可变区核苷酸序列的非人动物,其中所述重排重链可变区核苷酸序列编码序列 $V_H3-23/X_1X_2/J_H$ ,其中 $X_1$ 是任何氨基酸,并且 $X_2$ 是任何氨基酸。这种非人动物的另一实例是在它的种系基因组中包含经遗传修饰的内源性免疫球蛋白重链基因座的非人动物,所述经遗传修饰的内源性免疫球蛋白重链基因座包含可操作地连接至内源性非人免疫球蛋白恒定区基因序列的重排人免疫球蛋白重链可变区核苷酸序列,其中所述非人动物关于B细胞群体展现大致上类似于野生型非人动物的体液免疫系统。这种非人动物的另一实例是包含经遗传修饰的内源性免疫球蛋白重链基因座的非人动物,所述经遗传修饰的内源性免疫球蛋白重链基因座包含重排人免疫球蛋白重链可变区核苷酸序列,其包含通过间隔体可操作地连接至重链J区段( $J_H$ )序列的重链V区段( $V_H$ )序列,其中所述间隔体包含编码至少两个氨基酸残基,其中所述重排人免疫球蛋白重链可变区核苷酸序列可操作地连接至内源性非人动物免疫球蛋白恒定区基因序列。在一个实例中, $V_H$ 区段是 $V_H3-23$ 。

[0392] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括其种系基因组包含以下的非人动物:限制免疫球蛋白重链基因座,其特征存在于存在可操作地连接至非人免疫球蛋白重链恒定区核酸序列的单一未重排 $V_H$ 基因区段、一个或多个人未重排 $D_H$ 基因区段和一个或多个人未重排 $J_H$ 基因区段,其中所述非人动物进一步包含含有源于所述限制免疫球蛋白重链基因座的重排人重链可变区基因序列的B细胞。参见例如US 2013/0323791和US 2013/0096287,其各自出于所有目的以引用的方式整体并入本文。在一些所述非人动物中,单一未重排人 $V_H$ 基因区段是 $V_H1-69$ 。在一些所述非人动物中,单一未重排人 $V_H$ 基因区段是 $V_H1-2$ 。可使用的其他非人动物包括其内源性免疫球蛋白重链基因座由于它包含单一 $V_H$ 基因区段、一个或多个人 $D_H$ 基因区段和一个或多个人 $J_H$ 基因区段而受限制,并且不包含功能性内源性免疫球蛋白重链可变区基因座的非人动物;所述非人动物进一步包含一个或多个可操作地连接至一个或多个人 $J_L$ 基因区段的人免疫球蛋白 $V_L$ 基因区段,其中所述单一 $V_H$ 基因区段、一个或多个人 $D_H$ 基因区段和一个或多个 $J_H$ 基因区段可操作地连接至非人免疫球蛋

白重链恒定区基因,其中所述单一人 $V_H$ 基因区段是 $V_H1-69$ 或其多态性变体。

[0393] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系基因组中包含经遗传修饰的免疫球蛋白重链基因座的非人动物,所述经遗传修饰的免疫球蛋白重链基因座包含未重排人免疫球蛋白重链可变区核苷酸序列,其中所述未重排重链可变区核苷酸序列包含至少一个组氨酸密码子的添加或至少一个非组氨酸密码子被组氨酸密码子的取代,其中所述组氨酸密码子不由相应人种系重链可变区基因区段编码;并且其中添加或进行取代的组氨酸密码子存在于互补决定区3(CDR3)编码序列中。参见例如US 2013/0247235、US 9,301,510和US 14/046,501,其各自出于所有目的以引用的方式整体并入本文。

[0394] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括包含种系遗传修饰的非人动物,所述种系遗传修饰包括编码内源性IgG恒定区基因的 $C_H1$ 结构域的核苷酸序列的至少一部分的缺失;其中所述非人动物表达包含功能性 $C_H1$ 结构域的IgM恒定区基因,并且所述非人动物在它的血清中表达完全或部分缺乏 $C_H1$ 结构域并且缺乏同源轻链的IgG抗体。参见例如US 2011/0145937、US 2014/0289876、US 2015/0197553、US 2015/0197554、US 2015/0197555、US 2015/0196015、US 2015/0197556、US 2015/0197557和US 8,754,287,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是包含种系修饰的非人动物,所述修饰包括:(a) 编码内源性IgG恒定区基因的 $C_H1$ 结构域的核苷酸序列的缺失;和(b) 一个或多个人重链可变区基因区段的纳入,其中所述一个或多个人重链可变区基因区段可操作地连接至(a)的内源性IgG恒定区;其中所述非人动物包含完整IgM恒定区基因,并且所述非人动物表达包含人可变结构域,完全或部分缺乏CH1结构域,并且缺乏同源轻链的IgG重链抗体,并且将所述IgG重链抗体分泌至它的血清中。参见例如US 2011/0145937。这种非人动物的另一实例是包含种系修饰的非人动物,所述修饰包括:(a) 编码内源性IgG恒定区基因的 $C_H1$ 结构域和铰链区的核酸序列的缺失;和(b) 一个或多个人重链可变区基因区段的纳入,其中所述一个或多个人重链可变区基因区段可操作地连接至(a)的内源性IgG恒定区;其中所述非人动物包含完整IgM恒定区基因。参见例如US 2015/0197553。这种非人动物的另一实例是包含种系修饰的非人动物,所述修饰包含:(a) 编码内源性IgG恒定区基因的 $C_H1$ 结构域的核酸序列的缺失;(b) 内源性IgG2a恒定区基因的缺失;(c) 内源性IgG2b恒定区基因的缺失;和(d) 一个或多个人重链可变区基因区段的纳入,其中所述一个或多个人重链可变区基因区段可操作地连接至(a)的内源性IgG恒定区;其中所述非人动物包含完整IgM恒定区基因。参见例如US 2015/0197554。这种非人动物的另一实例是包含种系修饰的非人动物,所述修饰包含:(a) 编码内源性IgG恒定区基因的 $C_H1$ 结构域和铰链区的核酸序列的缺失;(b) 内源性IgG2a恒定区基因的缺失;(c) 内源性IgG2b恒定区基因的缺失;和(d) 一个或多个人重链可变区基因区段的纳入,其中所述一个或多个人重链可变区基因区段可操作地连接至(a)的内源性IgG恒定区;其中所述非人动物包含完整IgM恒定区基因。参见例如US2015/0197555。这种非人动物的另一实例是包含种系修饰的非人动物,所述修饰包含:(a) 编码内源性IgG1恒定区基因的 $C_H1$ 结构域的核酸序列的缺失;(b) 内源性IgD恒定区基因的缺失;(c) 内源性IgG3恒定区基因的缺失;(d) 内源性IgG2a恒定区基因的缺失;(e) 内源性IgG2b恒定区基因的缺失;(f) 内源性IgE恒定区基因的缺失;(g) 内源性IgA恒定区基因的缺失;和(h) 一个或多个人重链可变区基因区段的纳入,其中所述一个

或多个人重链可变区基因区段可操作地连接至(a)的内源性IgG1恒定区;其中所述非人动物包含完整IgM恒定区基因。参见例如US 2015/0196015。这种非人动物的另一实例是包含种系修饰的非人动物,所述修饰包含:(a)编码内源性IgG1恒定区基因的C<sub>H</sub>1结构域的核酸序列的缺失;(b)编码内源性IgG2a恒定区基因的C<sub>H</sub>1结构域的核酸序列的缺失;(c)内源性IgD恒定区基因的缺失;(d)内源性IgG3恒定区基因的缺失;(e)内源性IgG2b恒定区基因的缺失;(f)内源性IgE恒定区基因的缺失;(g)内源性IgA恒定区基因的缺失;和(h)一个或多个人重链可变区基因区段的纳入,其中所述一个或多个人重链可变区基因区段可操作地连接至(a)的内源性IgG1恒定区;其中所述非人动物包含完整IgM恒定区基因。参见例如US 2015/0197556。这种非人动物的另一实例是包含种系修饰的非人动物,所述修饰包含:(a)编码内源性IgG1恒定区基因的C<sub>H</sub>1结构域和铰链区的核酸序列的缺失;(b)内源性IgD恒定区基因的缺失;(c)内源性IgG3恒定区基因的缺失;(d)内源性IgG2a恒定区基因的缺失;(e)内源性IgG2b恒定区基因的缺失;(f)内源性IgE恒定区基因的缺失;(g)内源性IgA恒定区基因的缺失;和(h)一个或多个人重链可变区基因区段的纳入,其中所述一个或多个人重链可变区基因区段可操作地连接至(a)的内源性IgG1恒定区;其中所述非人动物包含完整IgM恒定区基因。参见例如US 2015/0197557。

[0395] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括包含与非人动物 $\kappa$ 轻链恒定区序列邻接的 $\lambda$ 轻链可变区序列(V $\lambda$ )和至少一个J序列(J)的非人动物。参见例如US 2012/0073004、US 2014/0137275、US 2015/0246976、US 2015/0246977、US 2015/0351371、US 9,035,128、US 9,066,502、US 9,163,092和US 9,150,662,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一个实例是包含以下的非人动物:(a)在内源性非人动物轻链基因座处的至少12至至少40个未重排人 $\lambda$ 轻链可变区基因区段和至少一个人J $\lambda$ 基因区段;(b)位于所述至少12至至少40个人轻链可变区基因区段与所述至少一个人J $\lambda$ 序列之间的人V $\kappa$ -J $\kappa$ 基因间序列;其中所述非人动物表达包含含有人V $\lambda$ 结构域和非人动物C $\kappa$ 结构域的轻链的抗体。这种非人动物的另一实例是在它的种系中在内源性 $\kappa$ 轻链基因座处包含以下的非人动物:(a)未重排轻链可变区,其包含多个连续未重排功能性人 $\lambda$ 轻链V(hV $\lambda$ )基因区段和多个连续未重排功能性人 $\lambda$ 轻链J(hJ $\lambda$ )基因区段,其中所述多个hV $\lambda$ 基因区段和所述多个hJ $\lambda$ 基因区段是所述未重排轻链可变区中仅有的功能性可变区基因区段;和(b)非人动物 $\kappa$ 轻链恒定区基因,其中所述多个连续未重排人 $\lambda$ 轻链V(hV $\lambda$ )基因区段和所述多个连续未重排人 $\lambda$ 轻链J(hJ $\lambda$ )基因区段可操作地连接至所述非人动物 $\kappa$ 轻链恒定区基因以使所述未重排轻链可变区能够重排以形成重排人 $\lambda$ 轻链可变区,并且所述非人动物表达包含含有由所述重排人 $\lambda$ 轻链可变区编码的可变区和由所述非人动物 $\kappa$ 轻链恒定区基因编码的恒定区的轻链的抗体。这种非人动物的另一实例是在它的种系中在内源性 $\kappa$ 轻链基因座处包含以下的非人动物:(a)未重排轻链可变区,其包含:(i)至少12个连续未重排功能性人 $\lambda$ 轻链可变区(hV $\lambda$ )基因区段和多个连续未重排功能性人 $\lambda$ 轻链J(hJ $\lambda$ )基因区段,其中所述至少12个功能性hV $\lambda$ 基因区段和所述多个功能性hJ $\lambda$ 基因区段是所述未重排轻链可变区中仅有的功能性可变区基因区段;和(ii)位于所述连续hV $\lambda$ 基因区段与所述多个连续hJ $\lambda$ 基因区段之间的人V $\kappa$ -J $\kappa$ 基因间序列;和(b)非人动物 $\kappa$ 轻链恒定区基因;其中所述至少12个连续未重排功能性人 $\lambda$ 轻链V(hV $\lambda$ )基因区段和所述多个连续未重排功能性人 $\lambda$ 轻链J(hJ $\lambda$ )基因区段可操作地连接至所述非人动物 $\kappa$ 轻链恒定区基因以使所述未重排轻链可变

区能够重排以形成重排人 $\lambda$ 轻链可变区,并且所述非人动物表达包含含有由所述重排人 $\lambda$ 轻链可变区编码的可变区和由所述非人动物 $\kappa$ 轻链恒定区基因编码的恒定区的轻链的抗体。这种非人动物的另一实例是在它的种系中包含以下的非人动物:(a) 未重排轻链可变区,其包含多个连续未重排功能性人 $\lambda$ 轻链V(hV $\lambda$ )基因区段和多个连续未重排功能性人 $\lambda$ 轻链J(hJ $\lambda$ )基因区段,其中所述多个hV $\lambda$ 基因区段和所述多个hJ $\lambda$ 基因区段是所述未重排轻链可变区中仅有的功能性可变区基因区段;和(b) 非人动物 $\kappa$ 轻链恒定区基因,其中所述多个连续未重排功能性hV $\lambda$ 基因区段和所述多个连续未重排功能性hJ $\lambda$ 基因区段可操作地连接至所述非人动物 $\kappa$ 轻链恒定区基因以使所述未重排轻链可变区能够重排以形成重排人 $\lambda$ 轻链可变区,并且所述非人动物表达包含含有由所述重排人 $\lambda$ 轻链可变区编码的可变结构域和由所述非人动物 $\kappa$ 轻链恒定区基因编码的恒定结构域的轻链的抗体。这种非人动物的另一实例是在它的种系中包含以下的非人动物:(a) 未重排轻链可变区,其包含:(i) 至少12个连续未重排功能性人 $\lambda$ 轻链V(hV $\lambda$ )基因区段和多个连续未重排功能性人 $\lambda$ 轻链J(hJ $\lambda$ )基因区段,其中所述至少12个功能性hV $\lambda$ 基因区段和所述多个功能性hJ $\lambda$ 基因区段是所述未重排轻链可变区中仅有的功能性可变区基因区段;和(ii) 位于所述连续hV $\lambda$ 基因区段与所述多个连续hJ $\lambda$ 基因区段之间的人V $\kappa$ -J $\kappa$ 基因间序列;和(b) 非人动物 $\kappa$ 轻链恒定区基因;其中所述至少12个连续未重排功能性hV $\lambda$ 基因区段和所述多个连续未重排功能性hJ $\lambda$ 基因区段可操作地连接至所述非人动物 $\kappa$ 轻链恒定区基因以使所述未重排轻链可变区能够重排以形成重排人 $\lambda$ 轻链可变区,并且所述非人动物表达包含含有由所述重排人 $\lambda$ 轻链可变区编码的可变结构域和由所述非人动物 $\kappa$ 轻链恒定区基因编码的恒定结构域的轻链的抗体。这种非人动物的另一实例是其基因组包含免疫球蛋白基因座的非人动物,所述免疫球蛋白基因座包含可操作地连接至非人动物C $\kappa$ 基因的人V $\lambda$ 和J $\lambda$ 基因区段,以致所述非人动物表达包含人 $\lambda$ 可变结构域序列与非人动物 $\kappa$ 恒定结构域融合的免疫球蛋白轻链。参见例如US 9,226,484。

[0396] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系中在内源性非人动物轻链基因座处包含人 $\lambda$ 轻链可变区序列的非人动物,其中所述人 $\lambda$ 可变区序列表达在包含非人动物免疫球蛋白恒定区基因序列的轻链中。参见例如US 2013/0323790、US 2013/0326647、US 2015/0089680、US 2015/0173331、US 2015/0176002、US 2015/0173332、US 2012/0070861、US 2015/0320023、US 2016/0060359、US 2016/0057979、US 9,029,628、US 9,006,511、US 9,012,717、US 9,206,261、US 9,206,262、US 9,206,263和US 9,226,484,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是表达包含人 $\lambda$ 可变序列与非人动物恒定区融合的免疫球蛋白轻链的非人动物,其中所述非人动物展现约1:1的 $\kappa$ 使用与 $\lambda$ 使用比率。参见例如US 9,029,628。这种非人动物的另一实例是其基因组包含内源性未重排 $\kappa$ 轻链免疫球蛋白基因座的非人动物,所述内源性未重排 $\kappa$ 轻链免疫球蛋白基因座包含用人V $\lambda$ 和J $\lambda$ 基因区段对内源性V $\kappa$ 和J $\kappa$ 基因区段的替换,并且其中所述人V $\lambda$ 和J $\lambda$ 基因区段可操作地连接至非人动物C $\kappa$ 基因以使所述非人动物表达包含人 $\lambda$ 可变序列与非人动物 $\kappa$ 恒定区融合的免疫球蛋白轻链。参见例如US 9,006,511。这种非人动物的另一实例是其基因组包含内源性 $\lambda$ 轻链免疫球蛋白基因座的非人动物,所述内源性 $\lambda$ 轻链免疫球蛋白基因座包含:(i) 第一内源性V $\lambda$ -J $\lambda$ -C $\lambda$ 基因簇的缺失;和(ii) 用人V $\lambda$ 和J $\lambda$ 基因区段对第二内源性V $\lambda$ -J $\lambda$ -C $\lambda$ 基因簇中的内源性V $\lambda$ 和J $\lambda$ 基因区段的片段的替换,其中所述人V $\lambda$ 和J $\lambda$ 基因区段包含至少一个人V $\lambda$ 基因区段和至少一个人J $\lambda$ 基因区段,并且其中所述人V $\lambda$ 和J $\lambda$ 基因区

段可操作地连接至非人动物C $\lambda$ 基因。参见例如US 9,012,717。

[0397] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括具有包含对免疫球蛋白重链基因座的修饰的基因组的非人动物,其中所述修饰使内源性ADAM6功能降低或消除,并且所述非人动物进一步包含编码非人动物ADAM6蛋白或其直系同源物或同源物或相应ADAM6蛋白的功能性片段的核酸序列。参见例如US 2012/0322108、US 2013/0254911、US 2014/0213773、US 2015/0201589、US 2015/0210776、US 2014/0017228、US 8,642,835和US 8,697,940,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是其基因组包含以下的非人动物:(a) ADAM6基因的异位放置;和(b) 人免疫球蛋白重链可变区基因座,其包含一个或多个人V<sub>H</sub>基因区段、一个或多个人D<sub>H</sub>基因区段和一个或多个人J<sub>H</sub>基因区段向内源性非人动物重链基因座中的插入,其中所述人V<sub>H</sub>、D<sub>H</sub>和J<sub>H</sub>基因区段可操作地连接至重链恒定区基因;以使所述非人动物的特征在于:(i) 它是能生育的;以及(ii) 当它用抗原免疫时,它产生包含可操作地连接至由所述重链恒定区基因编码的重链恒定结构域的由所述一个或多个人V<sub>H</sub>、一个或多个人D<sub>H</sub>和一个或多个人J<sub>H</sub>基因区段编码的重链可变结构域的抗体,其中所述抗体显示与所述抗原的特异性结合。参见例如US 8,642,835。

[0398] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括包含以下的非人动物:(a) 一个或多个人V $\lambda$ 和J $\lambda$ 基因区段在非人免疫球蛋白轻链恒定区的上游的插入,(b) 一个或多个人V<sub>H</sub>、一个或多个人D<sub>H</sub>和一个或多个人J<sub>H</sub>基因区段在非人免疫球蛋白重链恒定区的上游的插入,和(c) 编码ADAM6蛋白或其功能性片段的核苷酸序列,其中所述ADAM6蛋白从异位ADAM6核酸序列表达。参见例如US 2013/0160153和US 2014/0017228,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是其基因组包含以下的非人动物:(a) 一个或多个人V $\lambda$ 基因区段和一个或多个人J $\lambda$ 基因区段在非人动物免疫球蛋白轻链恒定区基因的上游的插入,(b) 一个或多个人V<sub>H</sub>基因区段、一个或多个人D<sub>H</sub>基因区段和一个或多个人J<sub>H</sub>基因区段在非人动物免疫球蛋白重链恒定区基因的上游的插入,和(c) 编码非人动物ADAM6蛋白的异位核苷酸序列,其中所述非人动物ADAM6蛋白从所述异位核苷酸序列表达。参见例如US 2013/0160153。

[0399] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系中包含免疫球蛋白基因座的非人动物,所述免疫球蛋白基因座包含未重排免疫球蛋白可变基因序列,其在CDR3编码序列中包含至少一个非组氨酸密码子被组氨酸密码子的取代或至少一个组氨酸密码子的插入,其中所述非人动物进一步在体内包含多样性抗体谱系,所述抗体各自对目标抗原具有特异性,并且在可变结构域的CDR3中包含至少一个由所述未重排免疫球蛋白可变基因序列中的至少一个组氨酸密码子取代或插入编码的组氨酸氨基酸。参见例如US 2013/0247236和US 2014/0082760,其各自出于所有目的以引用的方式整体并入本文。在一个实例中,第一免疫球蛋白可变区基因座包含未重排免疫球蛋白重链可变区序列的包含未重排V<sub>H</sub>、D<sub>H</sub>和J<sub>H</sub>基因区段的功能性部分,并且其中所述未重排V<sub>H</sub>、D<sub>H</sub>和J<sub>H</sub>基因区段中的一者或多者包含不由相应野生型种系基因区段编码的插入或进行取代的组氨酸密码子。在另一实例中,未重排V<sub>H</sub>、D<sub>H</sub>和J<sub>H</sub>基因区段是未重排人V<sub>H</sub>、未重排人D<sub>H</sub>和未重排人J<sub>H</sub>基因区段。在另一实施方案中,在它的种系中包含含有免疫球蛋白轻链可变区序列的第二免疫球蛋白可变区基因座,所述免疫球蛋白轻链可变区序列包含至少一个组氨酸密码子的插入或至少一个非组氨酸密码子被组氨酸密码子的取代,其中插入或进行取代的组氨酸密码子不由相应

野生型种系免疫球蛋白可变区序列编码,其中非人动物表达包含由非人动物的种系中的组氨酸取代或插入获得的组氨酸的免疫球蛋白轻链可变结构域。参见例如US 2013/0247236。

[0400] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括包含以下的非人动物:(a)一个或多个人 $V_L$ 和一个或多个人 $J_L$ 基因区段在非人免疫球蛋白轻链恒定区的上游的插入;(b)一个或多个人 $V_L$ 和一个或多个人 $J_L$ 基因区段在非人免疫球蛋白重链恒定区的上游的插入;和(c)编码ADAM6蛋白或其功能性片段的核苷酸序列,其中所述ADAM6蛋白从异位ADAM6核酸序列表达。参见例如US 2013/0212719,其出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是其基因组包含以下的非人动物:(a)一个或多个人 $V_L$ 基因区段和一个或多个人 $J_L$ 基因区段在非人免疫球蛋白轻链恒定区基因的上游的插入,其中所述一个或多个人 $V_L$ 基因区段和一个或多个人 $J_L$ 基因区段可操作地连接至所述非人免疫球蛋白轻链恒定区基因;(b)一个或多个人 $V_L$ 基因区段和一个或多个人 $J_L$ 基因区段在非人免疫球蛋白重链恒定区基因的上游的插入,其中所述一个或多个人 $V_L$ 基因区段和一个或多个人 $J_L$ 基因区段可操作地连接至所述非人免疫球蛋白重链恒定区基因;和(c)编码非人动物(例如啮齿动物,诸如大鼠或小鼠)ADAM6蛋白的插入核酸序列,其中所述非人动物(例如啮齿动物,诸如大鼠或小鼠)ADAM6蛋白从所述插入核酸序列表达,以使所述非人动物的B细胞表达各自包括两个免疫球蛋白轻链与两个免疫球蛋白重链配对的抗体,其中各轻链包含人轻链可变结构域和非人轻链恒定结构域,并且各重链包含人轻链可变结构域和非人重链恒定结构域。参见例如US 2013/0212719。

[0401] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系中具有以下的非人动物:(a)包含单一人 $V_H$ 基因区段、一个或多个 $D_H$ 基因区段和一个或多个 $J_H$ 基因区段的人基因组序列;和(b)编码在雄性非人动物中具有功能性的ADAM6蛋白的序列,其中编码所述ADAM6的所述序列位于不同于野生型非人动物的ADAM6基因座的位置处。参见例如US2013/0333057,其出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是在它的种系中具有以下的非人动物:(a)包含单一人 $V_H$ 基因区段、一个或多个人 $D_H$ 基因区段和一个或多个人 $J_H$ 基因区段的未重排人基因组序列,其中所述单一人 $V_H$ 基因区段是 $V_H1-2$ 、 $V_H1-69$ 、 $V_H2-26$ 、 $V_H2-70$ 或其多态性变体;和(b)编码在雄性非人动物中具有功能性的ADAM6蛋白的序列,其中编码所述ADAM6蛋白的所述序列位于不同于野生型非人动物的ADAM6基因座的位置处。参见例如US 2013/0333057。

[0402] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括包含以下的非人动物:(a)编码免疫球蛋白轻链的人 $V_L$ 结构域的单一重排人免疫球蛋白轻链可变区( $V_L/J_L$ ),其中所述单一重排人 $V_L/J_L$ 区选自人 $V_{\kappa1-39}/J$ 基因区段或人 $V_{\kappa3-20}/J$ 基因区段(例如 $V_{\kappa1-39}/J_{\kappa5}$ 基因区段或人 $V_{\kappa3-20}/J_{\kappa1}$ 基因区段);和(b)用一个或多个人 $V_H$ 基因区段对内源性重链可变( $V_H$ )基因区段的替换,其中所述人 $V_H$ 基因区段可操作地连接至内源性重链恒定( $C_H$ )区基因,并且所述人 $V_H$ 基因区段能够重排并形成人/非人动物嵌合重链基因。所述非人动物可被称为“通用轻链”(ULC)或“共同轻链”非人动物。参见例如US 2011/0195454、US 2012/0021409、US 2012/0192300、US 2015/0059009、US 2013/0045492、US 2013/0198880、US 2013/0185821、US 2013/0302836、US 2015/0313193和US 15/056,713,其各自出于所有目的以引用的方式整体并入本文。同样,可使用的另一非人动物(例如啮齿动物,诸如大鼠或小鼠)包括表达抗体群体的非人动物,其中所述非人动物的种系仅包括是重排人种系 $\kappa$ 轻链

可变区基因的单一免疫球蛋白 $\kappa$ 轻链可变区基因,所述非人动物就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是杂合的,因为它仅含有一个拷贝,或就所述单一免疫球蛋白 $\kappa$ 轻链可变区基因来说是纯合的,因为它含有两个拷贝;所述非人动物的特征在于活性亲和力成熟,以使:(i)所述群体的各免疫球蛋白 $\kappa$ 轻链都包含由所述重排人种系 $\kappa$ 轻链可变区基因或由其体细胞突变体编码的轻链可变结构域;(ii)所述群体包括包含轻链可变结构域由所述重排人种系 $\kappa$ 轻链可变区基因编码的免疫球蛋白 $\kappa$ 轻链的抗体和包含轻链可变结构域由所述其体细胞突变体编码的免疫球蛋白 $\kappa$ 轻链的抗体;以及(iii)所述非人动物产生与所述免疫球蛋白 $\kappa$ 轻链成功配对以形成所述群体的所述抗体的体细胞突变的高亲和力重链的多样性集合。这种非人动物的一实例是在它的种系中就以下来说是杂合的或纯合的非人动物:(a)重排 $V_{\kappa}/J_{\kappa}$ 序列在内源性非人动物 $\kappa$ 免疫球蛋白轻链可变区基因座处的插入,所述重排 $V_{\kappa}/J_{\kappa}$ 序列包含:单一人种系 $V_{\kappa}$ 序列,所述单一人种系 $V_{\kappa}$ 序列存在于SEQ ID NO:148或SEQ ID NO:149中;和单一人种系 $J_{\kappa}$ 序列,其中所述重排 $V_{\kappa}/J_{\kappa}$ 序列可操作地连接至内源性非人动物 $\kappa$ 恒定区;和(b)多个人免疫球蛋白重链可变区基因区段在内源性非人动物免疫球蛋白重链可变区基因座处的插入,其中所述人免疫球蛋白重链可变区基因区段可操作地连接至内源性非人动物免疫球蛋白重链恒定区,并且所述人免疫球蛋白重链可变区基因区段能够重排并形成重排人/非人动物嵌合免疫球蛋白重链基因。SEQ ID NO:148是工程化的人 $V_{\kappa}1-39J_{\kappa}5$ 基因座的序列,并且SEQ ID NO:149是工程化的人 $V_{\kappa}3-20J_{\kappa}1$ 基因座的序列。参见例如US 2011/0195454,其出于所有目的以引用的方式整体并入本文。

[0403] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括适用于产生人 $V_L/C_H \times ULC$ 结构域的非人动物,其在它的种系基因组中包含:(i)编码免疫球蛋白杂合链的杂合免疫球蛋白基因座,其中所述杂合免疫球蛋白基因座包含可操作地连接至包含一个或多个重链恒定区基因的免疫球蛋白重链恒定区核酸序列的未重排人免疫球蛋白轻链可变区基因区段( $V_L$ 和 $J_L$ ),所述重链恒定区基因各自至少编码功能性 $C_H1$ 结构域,其中所述 $V_L$ 和 $J_L$ 基因区段能够重排以形成包含可操作地连接至所述免疫球蛋白重链恒定区核酸序列的重排人 $V_L/J_L$ 基因序列的杂合序列;(ii)编码人通用轻链,并且包含可操作地连接至免疫球蛋白轻链恒定区核酸序列的人通用重排轻链可变区核苷酸序列的轻链基因座;其中所述非人动物能够产生包含源于所述杂合基因座的人免疫球蛋白杂合链和源于所述轻链基因座的同源人通用轻链的抗原结合蛋白,其中所述人免疫球蛋白杂合链包含人免疫球蛋白轻链可变( $hV_L/C_H \times ULC$ )结构域融合于包含功能性 $C_H1$ 结构域的重链恒定IgD、IgG、IgE或IgA区,并且其中所述人通用轻链包含人免疫球蛋白轻链融合于轻链恒定结构域。参见例如PCT/US2016/023289,其出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是适用于产生人 $V_L/C_H \times ULC$ 结构域的非人动物,其在它的种系基因组中包含:(i)经修饰内源性免疫球蛋白重链基因座,其包含用可操作地连接至包含一个或多个重链恒定区基因的内源性非人动物免疫球蛋白重链恒定区核酸的多个未重排人免疫球蛋白轻链可变 $V_{\kappa}$ 基因区段和多个未重排人免疫球蛋白轻链接合 $J_{\kappa}$ 基因区段对所有功能性内源性非人动物免疫球蛋白重链可变 $V_H$ 基因区段、所有功能性内源性非人动物免疫球蛋白重链多样性 $D_H$ 基因区段和所有功能性内源性非人动物免疫球蛋白重链接合 $J_H$ 基因区段的替换,所述重链恒定区基因各自至少编码功能性 $C_H1$ 结构域,其中所述多个未重排人免疫球蛋白轻链 $V_{\kappa}$ 基因区段和所述多个未重排人免疫球蛋白轻链 $J_{\kappa}$ 基因区段在B细胞发育期间参与B细胞中的重排以形成可

操作地连接至在内源性非人动物免疫球蛋白重链基因座处的所述内源性非人动物免疫球蛋白重链恒定区核酸序列的第一重排人免疫球蛋白轻链可变区V<sub>κ</sub>/J<sub>κ</sub>核苷酸序列;和(ii)经修饰内源性轻链基因座,其包含源于重排V<sub>κ</sub>1-39/J<sub>κ</sub>5或V<sub>κ</sub>3-20/J<sub>κ</sub>1基因序列的单一重排人免疫球蛋白轻链可变区基因序列,其中所述单一重排人免疫球蛋白轻链可变区基因序列可操作地连接至内源性非人动物免疫球蛋白轻链恒定区κ基因序列;其中所述非人动物能够产生包含源于所述经修饰内源性免疫球蛋白重链基因座的人免疫球蛋白杂合链和源于所述经修饰内源性轻链基因座的同源人通用轻链的抗原结合蛋白,其中所述人免疫球蛋白杂合链包含人免疫球蛋白轻链可变(hV<sub>L</sub>/C<sub>H</sub> x ULC)结构域融合于包含功能性C<sub>H</sub>1结构域的重链恒定IgD、IgG、IgE或IgA区,并且其中所述人通用轻链包含人免疫球蛋白轻链融合于轻链恒定结构域。

[0404] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系基因组中例如在内源性非人轻链基因座处包含轻链免疫球蛋白基因座的非人动物,所述轻链免疫球蛋白基因座包含可操作地连接至免疫球蛋白轻链恒定区核酸序列的重排人免疫球蛋白轻链可变区核苷酸序列,其中可操作地连接至免疫球蛋白轻链恒定区核酸序列的所述重排人免疫球蛋白轻链可变区核苷酸序列编码通用轻链,并且其中所述非人动物能够产生或会产生表达包含免疫球蛋白杂合链和所述通用轻链的抗原结合蛋白的细胞例如淋巴细胞例如B细胞。参见例如US 2013/0247234、US 2014/0329711、US 2014/0013456、US 2015/0119556、US 2015/0250151、US 9,334,334和US 9,332,742,其各自出于所有目的以引用的方式整体并入本文。一些所述非人动物就重排人免疫球蛋白轻链可变区核苷酸序列来说是纯合的。一些所述非人动物就重排人免疫球蛋白轻链可变区核苷酸序列来说是杂合的。在一些所述非人动物中,轻链恒定区核酸序列是κ序列。在一些所述非人动物中,轻链恒定区核酸序列是λ序列。在一些所述非人动物中,第二免疫球蛋白基因座是轻链κ基因座。在一些实施方案中,第二免疫球蛋白基因座是轻链λ基因座。这种非人动物的一实例是在它的种系中包含免疫球蛋白轻链基因座的非人动物,所述免疫球蛋白轻链基因座包含单一重排人免疫球蛋白轻链可变区基因序列,其包含人V<sub>κ</sub>和J<sub>κ</sub>区段序列,其中所述V<sub>κ</sub>区段序列源于人V<sub>κ</sub>1-39或V<sub>κ</sub>3-20基因区段,并且其中所述单一重排人免疫球蛋白轻链可变区基因序列包含所述V<sub>κ</sub>区段序列的至少一个非组氨酸密码子被在选自105、106、107、108、109、111及其组合(根据IMGT编号)组成的组的位置处的组氨酸密码子的取代。

[0405] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括其基因组包含以下的非人动物:(a)可操作地连接至重链恒定区基因的人源化免疫球蛋白重链可变基因座,其包含至少一个未重排人V<sub>H</sub>、至少一个未重排人D<sub>H</sub>和至少一个未重排人J<sub>H</sub>区段;(b)可操作地连接至轻链恒定区基因的人源化免疫球蛋白轻链可变基因座,其包含不超过一个或不超过两个重排人轻链V/J序列;和(c)表达功能性非人动物ADAM6蛋白或其功能性直系同源物或功能性同源物或功能性片段的异位核酸序列。参见例如US 2013/0198879,其出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是在它的种系中包含以下的非人动物:(a)人源化免疫球蛋白重链可变基因座,其包含至少一个未重排人V<sub>H</sub>基因区段、至少一个未重排人D<sub>H</sub>基因区段和至少一个未重排人J<sub>H</sub>基因区段,其中所述人源化免疫球蛋白重链可变基因座可操作地连接至免疫球蛋白重链恒定区基因;(b)人源化免疫球蛋白轻链可变基因座,其包含(i)单一重排人轻链V/J序列,其中所述单一重排人轻链V/J序列是重排人V<sub>κ</sub>

1-39/J $\kappa$ 序列或重排人V $\kappa$ 3-20/J $\kappa$ 序列,或(ii)不超过一个人轻链V基因区段和不超过一个人轻链J基因区段,其中所述不超过一个人轻链V基因区段是V $\kappa$ 1-39或V $\kappa$ 3-20,其中所述人源化免疫球蛋白轻链可变基因座可操作地连接至免疫球蛋白轻链恒定区基因;和(c)表达在雄性非人动物中具有功能性的非人动物ADAM6蛋白或其直系同源物或同源物或功能性片段的异位核酸序列。

[0406] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系中包含以下的非人动物:(a)编码在内源性免疫球蛋白重链基因座处的至少一个内源性免疫球蛋白重链恒定区基因的C $\mu$ 1结构域的核苷酸序列中的缺失或失活性突变,其中所述至少一个内源性免疫球蛋白重链恒定区基因是IgG、IgA、IgE、IgD或其组合;和(b)以下中的任一者或两者:(i)包含至少一个未重排免疫球蛋白轻链可变区(V $_L$ )基因区段和至少一个未重排免疫球蛋白轻链接合(J $_L$ )基因区段的核酸序列,其中所述未重排V $_L$ 和J $_L$ 基因区段能够重组以形成可操作地连接至包含编码所述C $\mu$ 1结构域的所述核苷酸序列中的所述缺失或失活性突变的免疫球蛋白重链恒定区基因的重排免疫球蛋白轻链可变区(V $_L$ /J $_L$ )核苷酸序列;和/或(ii)包含含有V $_L$ 和J $_L$ 基因区段序列的单一重排免疫球蛋白轻链可变区V $_L$ /J $_L$ 基因序列的免疫球蛋白轻链基因座,其中所述单一重排免疫球蛋白轻链可变区基因序列可操作地连接至免疫球蛋白轻链恒定区基因序列。参见例如US 2015/0289489,其出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是包含以下的非人动物:(a)用以下在非人动物重链基因座处对所有或大致上所有内源性免疫球蛋白重链V、D和J基因区段的替换:(i)一个或多个未重排人免疫球蛋白重链V $_H$ 基因区段、一个或多个未重排人免疫球蛋白重链D $_H$ 基因区段和一个或多个未重排人免疫球蛋白重链J $_H$ 基因区段,其中所述一个或多个人未重排免疫球蛋白重链V $_H$ 、D $_H$ 和J $_H$ 基因区段可操作地连接至非人动物重链恒定区基因序列;或(ii)一个或多个未重排人轻链V $_L$ 基因区段和一个或多个人未重排轻链J $_L$ 基因区段,其中所述一个或多个未重排人轻链V $_L$ 和J $_L$ 基因区段可操作地连接至非人动物重链恒定区基因序列,其中所述非人动物重链恒定区基因序列包含全长IgM基因和编码选自由IgG1、IgG2a、IgG2b、IgG2c、IgG3及其组合组成的组的IgG基因中的C $\mu$ 1结构域的核苷酸序列中的缺失或失活性突变;和(b)用单一重排人可变V $\kappa$ /J $\kappa$ 基因序列对所有或大致上所有内源性免疫球蛋白轻链V和J基因区段的替换,并且其中所述非人动物表达包含IgM重链与同源轻链缔合的B细胞受体。

[0407] 可使用的其他非人动物(例如啮齿动物,诸如大鼠或小鼠)包括在它的种系中包含免疫球蛋白轻链基因座的非人动物,所述免疫球蛋白轻链基因座包含可操作地连接至免疫球蛋白轻链恒定区序列的不超过两个人V $_L$ 基因区段和一个或多个人J $_L$ 基因区段,其中所述不超过两个人V $_L$ 基因区段中的每一个包含至少一个不由相应人种系V $_L$ 基因区段编码的组氨酸密码子,并且其中所述人V $_L$ 基因区段和J $_L$ 基因区段能够重排以及编码抗体的人轻链可变结构域。参见例如US 2014/0013456、US 2015/0119556、US 2015/0250151、US 2013/0247234和US 9,332,742,其各自出于所有目的以引用的方式整体并入本文。这种非人动物的一实例是包含不超过两个人V $_L$ 基因区段的非人动物,所述人V $_L$ 基因区段各自能够与人J $_L$ 基因区段(选自一个或多个J $_L$ 区段)重排以及编码免疫球蛋白轻链的人可变结构域,其中所述不超过两个V $_L$ 基因区段各自和/或所述J $_L$ 基因区段包含至少一个非组氨酸残基被组氨酸残基的取代。参见例如US 2014/0013456。这种非人动物的另一实例是在它的种系中包含免

疫球蛋白轻链基因座的非人动物,所述免疫球蛋白轻链基因座包含可操作地连接至免疫球蛋白轻链恒定区序列的两个未重排人 $V_k$ 基因区段和一个或多个未重排人 $J_k$ 基因区段,其中所述两个未重排人 $V_k$ 基因区段是人 $V_k1-39$ 和 $V_k3-20$ 基因区段,中的每一个包含非组氨酸密码子被组氨酸密码子的一个或多个取代,并且其中所述人 $V_k$ 和 $J_k$ 基因区段能够重排,并且所述人 $V_k$ 和 $J_k$ 基因区段编码在选自105、106、107、108、109、111(根据IMGT编号)及其组合组成的组的位置处包含一个或多个组氨酸的人轻链可变结构域,其中所述一个或多个组氨酸由所述一个或多个取代获得。参见例如US 2015/0250151。

#### [0408] IV. 产生抗原结合蛋白的方法

[0409] 通过本文公开的方法产生的经遗传修饰的F0代非人动物可用于制备针对目标外来靶抗原的抗原结合蛋白。已描述用于产生抗原结合蛋白(例如抗体)的若干技术。抗原结合蛋白可直接从经免疫小鼠的B细胞分离(参见例如US 2007/0280945,其出于所有目的以引用的方式整体并入本文),和/或经免疫小鼠的B细胞可用于制备杂交瘤(参见例如Kohler和Milstein(1975)Nature 256:495-497,其出于所有目的以引用的方式整体并入本文)。编码来自如本文所述的非人动物的抗原结合蛋白(重链和/或轻链)的DNA可易于使用常规技术来分离和测序。源于如本文所述的非人动物的杂交瘤和/或B细胞充当所述DNA的优选来源。一旦分离,即可将DNA放置至表达载体中,接着将所述表达载体转染至不另外产生免疫球蛋白的宿主细胞中,以在重组宿主细胞中获得对单克隆抗体的合成。

[0410] 举例来说,可使通过本文公开的方法产生的经遗传修饰的F0代非人动物暴露于靶抗原,并且在足以引发对目标外来靶抗原的免疫应答的条件下加以维持。可接着从经遗传修饰的F0代非人动物获得编码人免疫球蛋白重链可变结构域的第一核酸序列和/或编码人免疫球蛋白轻链可变结构域的第二核酸序列。或者,可接着从经遗传修饰的F0代非人动物分离抗原结合蛋白。作为一实例,可鉴定表达特异性结合目标外来抗原的抗体的克隆选择的淋巴细胞。

[0411] 在一个实例中,抗原结合蛋白可通过以下方式产生:用目标外来靶抗原使经遗传修饰的F0代非人动物免疫,使非人动物发动免疫应答,从经免疫动物收集淋巴细胞(例如B细胞),使所述淋巴细胞与骨髓瘤细胞融合以形成杂交瘤细胞,从所述杂交瘤细胞获得编码特异性结合靶抗原的 $V_H$ 结构域的核酸序列和/或编码特异性结合靶抗原的 $V_L$ 结构域的核酸序列,将所述核酸序列与编码免疫球蛋白恒定区或其功能性片段序列的核酸序列同框克隆(即可操作连接)以产生免疫球蛋白重链和/或免疫球蛋白轻链,以及在能够表达抗原结合蛋白的细胞(例如CHO细胞)中表达所述重链和所述轻链。

[0412] 在另一实例中,抗原结合蛋白可通过以下方式产生:用目标外来靶抗原使经遗传修饰的F0代非人动物免疫,使非人动物发动免疫应答,从经免疫动物收集淋巴细胞(例如B细胞),从所述淋巴细胞获得编码特异性结合靶抗原的 $V_H$ 结构域的核酸序列和/或编码特异性结合靶抗原的 $V_L$ 结构域的核酸序列,将所述核酸序列与编码免疫球蛋白恒定区或其功能性片段序列的核酸序列同框克隆(即可操作连接)以产生免疫球蛋白重链和/或免疫球蛋白轻链,以及在能够表达抗原结合蛋白的细胞(例如CHO细胞)中表达所述重链和所述轻链。

[0413] 用目标外来抗原免疫可用蛋白质、DNA、DNA和蛋白质的组合、或表达目标外来抗原的细胞进行。获得的淋巴细胞可来自经免疫动物的任何来源,包括例如脾、淋巴结或骨髓。

[0414] 在一些所述方法中, $V_H$ 结构域和/或 $V_L$ 结构域是人结构域(例如当经遗传修饰的F0

代非人动物在IgH与Igκ两者处均是纯合性人源化的时),将V<sub>H</sub>结构域和/或V<sub>L</sub>结构域与编码人恒定区的核酸序列同框克隆,并且产生的抗原结合蛋白是完全人抗体。

[0415] 当相较于对照非人动物(即在第一靶基因组基因座处是野生型)时,在本文所述的经遗传修饰的F0代非人动物(即在第一靶基因组基因座处经遗传修饰)中产生的针对目标外来抗原的抗原结合蛋白的产量通常增加。也就是说,相比于在使在第一靶基因组基因座处是野生型的对照非人动物免疫之后获得的抗原结合蛋白,在本文所述的经遗传修饰的F0代非人动物(即在第一靶基因组基因座处经遗传修饰)中产生的针对目标外来抗原的抗原结合蛋白通常具有更高效价。举例来说,效价可为至少1.5倍、2倍、3倍、4倍、5倍、6倍、7倍、8倍、9倍、或10倍高。术语抗体效价包括存在于血清中的特异性抗体的浓度的测量结果。举例来说,抗体效价可为生物体已产生多少识别特定表位的抗体的测量结果,表示为仍然产生阳性结果的最大稀释度的倒数。同样,相较于在使在第一靶基因组基因座处是野生型的对照非人动物免疫之后获得的抗原结合蛋白,在用目标外来抗原使经遗传修饰的F0代非人动物免疫之后,通常获得更具多样性的针对目标外来抗原的抗原结合蛋白谱系。对照非人动物是指在第一靶基因组基因座处是野生型的非人动物。优选地,经遗传修饰的F0代非人动物与对照动物之间的唯一实质性差异是第一靶基因组基因座的状况。举例来说,优选地,对照动物不具有其他实质性遗传修饰,并且是与非人动物相同的物种,是与非人动物相同的品系,具有相同遗传背景(除第一靶基因组基因座以外),并且与经遗传修饰的F0代非人动物具有相同龄期。

[0416] 以上或以下引用的所有专利申请、网站、其他出版物、登录号等都出于所有目的以引用的方式整体并入本文,所述引用的程度就好像明确地和个别地指示将各个别条目都如此以引用的方式并入本文一样。如果在不同时间有序列的不同版本与登录号相关联,那么意指的是在本申请的有效申请日期时与所述登录号相关联的版本。有效申请日期意指实际申请日期或涉及所述登录号的优先权申请的申请日期(如果可适用)的早先日期。同样,如果在不同时间有出版物、网站等的不同版本被公布,那么除非另外指示,否则意指的是在申请的有效申请日期时最新近公布的版本。除非另外明确指示,否则本发明的任何特征、步骤、要素、实施方案或方面都可与任何其他特征、步骤、要素、实施方案或方面组合使用。尽管本发明已出于明晰和理解的目的通过说明和举例的方式相当详细地加以描述,但将显而易见的是某些变化和修改可在随附权利要求的范围内加以实施。

[0417] 序列简述

[0418] 使用标准核苷酸碱基字母缩写和氨基酸三字母代码显示随附序列中所列的核苷酸和氨基酸序列。核苷酸序列遵循在序列的5'末端开始,并且前进(即在各行中从左至右)直至3'末端的标准惯例。仅显示各核苷酸序列的一个链,但互补链应理解为通过对所显示链的任何提及来包括。氨基酸序列遵循在序列的氨基末端开始,并且前进(即在各行中从左至右)直至羧基末端的标准惯例。

[0419] 表1. 序列描述。

[0420]

| SEQ ID NO | 类型  | 描述                                                       |
|-----------|-----|----------------------------------------------------------|
| 1         | DNA | 引导 RNA 识别序列 v.1                                          |
| 2         | DNA | 引导 RNA 识别序列 v.2                                          |
| 3         | DNA | 引导 RNA 识别序列 v.3                                          |
| 4         | DNA | <i>C5 (Hc)</i> gRNA A DNA 靶向区段(离靶基因座端点 100 bp)           |
| 5         | DNA | <i>C5 (Hc)</i> gRNA B DNA 靶向区段(离靶基因座端点 500 bp)           |
| 6         | DNA | <i>C5 (Hc)</i> gRNA C DNA 靶向区段(离靶基因座端点 38200 和 37500 bp) |
| 7         | DNA | <i>C5 (Hc)</i> gRNA D DNA 靶向区段(离靶基因座端点 43500 和 32200 bp) |
| 8         | DNA | <i>C5 (Hc)</i> gRNA E DNA 靶向区段(离靶基因座端点 500 bp)           |
| 9         | DNA | <i>C5 (Hc)</i> gRNA E2 DNA 靶向区段(离靶基因座端点 100 bp)          |
| 10        | DNA | <i>Lrp5</i> gRNA A DNA 靶向区段(离靶基因座端点 50 bp)               |
| 11        | DNA | <i>Lrp5</i> gRNA B DNA 靶向区段(离靶基因座端点 500 bp)              |
| 12        | DNA | <i>Lrp5</i> gRNA B2 DNA 靶向区段(离靶基因座端点 1000 bp)            |
| 13        | DNA | <i>Lrp5</i> gRNA C DNA 靶向区段(离靶基因座端点 29900 和 38430 bp)    |
| 14        | DNA | <i>Lrp5</i> gRNA D DNA 靶向区段(离靶基因座端点 29950 和 38380 bp)    |
| 15        | DNA | <i>Lrp5</i> gRNA E2 DNA 靶向区段(离靶基因座端点 1000 bp)            |
| 16        | DNA | <i>Lrp5</i> gRNA E DNA 靶向区段(离靶基因座端点 500 bp)              |
| 17        | DNA | <i>Lrp5</i> gRNA F DNA 靶向区段(离靶基因座端点 50 bp)               |
| 18        | DNA | <i>Ror1</i> gRNA A DNA 靶向区段(离靶基因座端点 200 bp)              |
| 19        | DNA | <i>Ror1</i> gRNA B DNA 靶向区段(离靶基因座端点 1000 bp)             |
| 20        | DNA | <i>Ror1</i> gRNA D DNA 靶向区段(离靶基因座端点 54300 和 55500 bp)    |
| 21        | DNA | <i>Ror1</i> gRNA C DNA 靶向区段(离靶基因座端点 54500 和 55300 bp)    |
| 22        | DNA | <i>Ror1</i> gRNA E DNA 靶向区段(离靶基因座端点 1000 bp)             |
| 23        | DNA | <i>Ror1</i> gRNA F DNA 靶向区段(离靶基因座端点 200 bp)              |
| 24        | DNA | <i>Trpa1</i> gRNA A DNA 靶向区段(离靶基因座端点 100 bp)             |
| 25        | DNA | <i>Trpa1</i> gRNA A2 DNA 靶向区段(离靶基因座端点 500 bp)            |
| 26        | DNA | <i>Trpa1</i> gRNA B DNA 靶向区段(离靶基因座端点 1000 bp)            |
| 27        | DNA | <i>Trpa1</i> gRNA C DNA 靶向区段(离靶基因座端点 25600 和 19740 bp)   |
| 28        | DNA | <i>Trpa1</i> gRNA D DNA 靶向区段(离靶基因座端点 26970 和 18370 bp)   |
| 29        | DNA | <i>Trpa1</i> gRNA E2 DNA 靶向区段(离靶基因座端点 1000 bp)           |
| 30        | DNA | <i>Trpa1</i> gRNA E DNA 靶向区段(离靶基因座端点 500 bp)             |
| 31        | DNA | <i>Trpa1</i> gRNA F DNA 靶向区段(离靶基因座端点 100 bp)             |
| 32        | DNA | 190045 正向引物                                              |
| 33        | DNA | 190061 正向引物                                              |
| 34        | DNA | 190068 正向引物                                              |
| 35        | DNA | 190030 正向引物                                              |
| 36        | DNA | 190033 正向引物(与图 6 中的 SV 48.3 的正向引物相同)                     |

[0421]

| SEQ ID NO | 类型  | 描述                                   |
|-----------|-----|--------------------------------------|
| 37        | DNA | 190013 正向引物                          |
| 38        | DNA | 190045 反向引物                          |
| 39        | DNA | 190061 反向引物                          |
| 40        | DNA | 190068 反向引物                          |
| 41        | DNA | 190030 反向引物                          |
| 42        | DNA | 190033 反向引物(与图 6 中的 SV 48.3 的反向引物相同) |
| 43        | DNA | 190013 反向引物                          |
| 44        | DNA | C2 探针(B6)-图 6 中的 SNV 0.32            |
| 45        | DNA | T3 探针(B6)-图 6 中的 SNV 1.2             |
| 46        | DNA | T6 探针(B6)-图 6 中的 SNV 11.1            |
| 47        | DNA | T7 探针(B6)-图 6 中的 SNV 13.2            |
| 48        | DNA | T8 探针(B6)-图 6 中的 SNV 17.5            |
| 49        | DNA | T9 探针(B6)-图 6 中的 SNV 25.8            |
| 50        | DNA | T10 探针(B6)-图 6 中的 SNV 33.0           |
| 51        | DNA | T11 探针(B6)-图 6 中的 SNV 38.3           |
| 52        | DNA | T13 探针(B6)-图 6 中的 SNV 49.6           |
| 53        | DNA | T14 探针(B6)-图 6 中的 SNV 57.2           |
| 54        | DNA | C2 探针(129)-图 6 中的 SNV 0.32           |
| 55        | DNA | T3 探针(129)-图 6 中的 SNV 1.2            |
| 56        | DNA | T6 探针(129)-图 6 中的 SNV 11.1           |
| 57        | DNA | T7 探针(129)-图 6 中的 SNV 13.2           |
| 58        | DNA | T8 探针(129)-图 6 中的 SNV 17.5           |
| 59        | DNA | T9 探针(129)-图 6 中的 SNV 25.8           |
| 60        | DNA | T10 探针(129)-图 6 中的 SNV 33.0          |
| 61        | DNA | T11 探针(129)-图 6 中的 SNV 38.3          |
| 62        | DNA | T13 探针(129)-图 6 中的 SNV 49.6          |
| 63        | DNA | T14 探针(129)-图 6 中的 SNV 57.2          |
| 64        | DNA | C2 正向引物-图 6 中的 SNV 0.32              |
| 65        | DNA | T3 正向引物-图 6 中的 SNV 1.2               |
| 66        | DNA | T6 正向引物-图 6 中的 SNV 11.1              |
| 67        | DNA | T7 正向引物-图 6 中的 SNV 13.2              |
| 68        | DNA | T8 正向引物-图 6 中的 SNV 17.5              |
| 69        | DNA | T9 正向引物-图 6 中的 SNV 25.8              |
| 70        | DNA | T10 正向引物-图 6 中的 SNV 33.0             |
| 71        | DNA | T11 正向引物-图 6 中的 SNV 38.3             |
| 72        | DNA | T13 正向引物-图 6 中的 SNV 49.6             |
| 73        | DNA | T14 正向引物-图 6 中的 SNV 57.2             |
| 74        | DNA | C2 反向引物-图 6 中的 SNV 0.32              |
| 75        | DNA | T3 反向引物-图 6 中的 SNV 1.2               |
| 76        | DNA | T6 反向引物-图 6 中的 SNV 11.1              |
| 77        | DNA | T7 反向引物-图 6 中的 SNV 13.2              |
| 78        | DNA | T8 反向引物-图 6 中的 SNV 17.5              |
| 79        | DNA | T9 反向引物-图 6 中的 SNV 25.8              |
| 80        | DNA | T10 反向引物-图 6 中的 SNV 33.0             |
| 81        | DNA | T11 反向引物-图 6 中的 SNV 38.3             |
| 82        | DNA | T13 反向引物-图 6 中的 SNV 49.6             |

[0422]

| SEQ ID NO | 类型  | 描述                              |
|-----------|-----|---------------------------------|
| 83        | DNA | T14 反向引物-图 6 中的 SNV 57.2        |
| 84        | DNA | 图 6 中的 SV 13.7 的正向引物            |
| 85        | DNA | 图 6 中的 SV 13.7 的反向引物            |
| 86        | DNA | 图 6 中的 SV 20.0 的正向引物            |
| 87        | DNA | 图 6 中的 SV 20.0 的反向引物            |
| 88        | DNA | 图 6 中的 SV 36.9 的正向引物            |
| 89        | DNA | 图 6 中的 SV 36.9 的反向引物            |
| 90        | DNA | 图 6 中的 SV 56.7 的正向引物            |
| 91        | DNA | 图 6 中的 SV 56.7 的反向引物            |
| 92        | DNA | <i>Lrp5</i> 基因座的 m-lr-f 引物      |
| 93        | DNA | <i>Lrp5</i> 基因座的 m-5'-f 引物      |
| 94        | DNA | <i>Lrp5</i> 基因座的 m-A 引物         |
| 95        | DNA | <i>Lrp5</i> 基因座的 h-lr-r 引物      |
| 96        | DNA | <i>Lrp5</i> 基因座的 m-5'-r 引物      |
| 97        | DNA | <i>Lrp5</i> 基因座的 h-5'-r 引物      |
| 98        | DNA | <i>Lrp5</i> 基因座的 m-F 引物         |
| 99        | DNA | <i>Lrp5</i> 基因座的 m-E2 引物        |
| 100       | DNA | 7064retU 正向引物                   |
| 101       | DNA | 7064retU 反向引物                   |
| 102       | DNA | 7064retU TAQMAN <sup>®</sup> 探针 |
| 103       | DNA | 7064retD 正向引物                   |
| 104       | DNA | 7064retD 反向引物                   |
| 105       | DNA | 7064retD TAQMAN <sup>®</sup> 探针 |
| 106       | DNA | 7140retU 正向引物                   |
| 107       | DNA | 7140retU 反向引物                   |
| 108       | DNA | 7140retU TAQMAN <sup>®</sup> 探针 |
| 109       | DNA | 7140retD 正向引物                   |
| 110       | DNA | 7140retD 反向引物                   |
| 111       | DNA | 7140retD TAQMAN <sup>®</sup> 探针 |
| 112       | DNA | <i>Folh1</i> gRNA A DNA 靶向区段    |
| 113       | DNA | <i>Folh1</i> gRNA A2 DNA 靶向区段   |
| 114       | DNA | <i>Folh1</i> gRNA B DNA 靶向区段    |
| 115       | DNA | <i>Folh1</i> gRNA C DNA 靶向区段    |
| 116       | DNA | <i>Folh1</i> gRNA D DNA 靶向区段    |
| 117       | DNA | <i>Folh1</i> gRNA E DNA 靶向区段    |
| 118       | DNA | <i>Folh1</i> gRNA E2 DNA 靶向区段   |
| 119       | DNA | <i>Folh1</i> gRNA F DNA 靶向区段    |
| 120       | DNA | <i>Adamts5</i> gRNA A DNA 靶向区段  |
| 121       | DNA | <i>Adamts5</i> gRNA A2 DNA 靶向区段 |
| 122       | DNA | <i>Adamts5</i> gRNA B DNA 靶向区段  |
| 123       | DNA | <i>Adamts5</i> gRNA C DNA 靶向区段  |
| 124       | DNA | <i>Adamts5</i> gRNA D DNA 靶向区段  |
| 125       | DNA | <i>Adamts5</i> gRNA E2 DNA 靶向区段 |
| 126       | DNA | <i>Adamts5</i> gRNA E DNA 靶向区段  |
| 127       | DNA | <i>Adamts5</i> gRNA F DNA 靶向区段  |
| 128       | DNA | <i>Dpp4</i> gRNA A DNA 靶向区段     |

[0423]

| SEQ ID NO | 类型  | 描述                                    |
|-----------|-----|---------------------------------------|
| 129       | DNA | <i>Dpp4</i> gRNA B DNA 靶向区段           |
| 130       | DNA | <i>Dpp4</i> gRNA B2 DNA 靶向区段          |
| 131       | DNA | <i>Dpp4</i> gRNA C DNA 靶向区段           |
| 132       | DNA | <i>Dpp4</i> gRNA D DNA 靶向区段           |
| 133       | DNA | <i>Dpp4</i> gRNA E2 DNA 靶向区段          |
| 134       | DNA | <i>Dpp4</i> gRNA E DNA 靶向区段           |
| 135       | DNA | <i>Dpp4</i> gRNA F DNA 靶向区段           |
| 136       | DNA | 图 8 中的 SV 6.1 的正向引物                   |
| 137       | DNA | 图 8 中的 SV 6.1 的反向引物                   |
| 138       | DNA | 图 8 中的 SV 6.3 的正向引物                   |
| 139       | DNA | 图 8 中的 SV 6.3 的反向引物                   |
| 140       | DNA | 图 8 中的 SV 7.8 的正向引物                   |
| 141       | DNA | 图 8 中的 SV 7.8 的反向引物                   |
| 142       | DNA | 图 8 中的 SV 16 的正向引物                    |
| 143       | DNA | 图 8 中的 SV 16 的反向引物                    |
| 144       | DNA | 图 8 中的 SV 25.5 的正向引物                  |
| 145       | DNA | 图 8 中的 SV 25.5 的反向引物                  |
| 146       | DNA | 金黄色葡萄球菌 Cas9 PAM 序列                   |
| 147       | DNA | 金黄色葡萄球菌 Cas9 PAM 序列                   |
| 148       | DNA | 工程化的人 V $\kappa$ 1-39J $\kappa$ 5 基因座 |
| 149       | DNA | 工程化的人 V $\kappa$ 3-20J $\kappa$ 1 基因座 |
| 150       | DNA | 引导 RNA 骨架 v1                          |
| 151       | DNA | 引导 RNA 骨架 v2                          |
| 152       | DNA | 引导 RNA 骨架 v3                          |

[0424] 实施例

[0425] 实施例1. 使用靶向起始密码子和终止密码子的成对引导RNA产生用于抗体产生的K0胚胎干(ES)细胞、单细胞期胚胎和小鼠。

[0426] **VELOCIGENE<sup>®</sup>**和**VELOCIMOUSE<sup>®</sup>**技术已允许产生使得能够产生完全人抗体的**VELOCIMMUNE<sup>®</sup>**小鼠。**VELOCIMMUNE<sup>®</sup>**小鼠表达免疫球蛋白 $\kappa$  (Ig $\kappa$ ) 和重 (IgH) 链, 其中完全人源化可变区连结至小鼠恒定区。因为蛋白质的在功能上重要的区域在物种之间倾向于保守, 所以对自身抗原的免疫耐受性常常对产生针对这些关键表位的抗体造成挑战。在传统上, 使**VELOCIMMUNE<sup>®</sup>**小鼠与在目标自身抗原靶标处携带杂合性敲除突变的F0小鼠交配以克服免疫耐受性。为产生适于免疫的三重纯合性小鼠 (目标靶标的纯合性无效和在IgH与Ig $\kappa$ 两者处的纯合性人源化), 需要再进行两代交配以及15至16个月的总时间。为使这个过程加速, 获得**VELOCIMMUNE<sup>®</sup>**胚胎干(ES)细胞, 其可被靶向来在目标靶标处产生无效等位基因。然而, 不幸的是, 需要依序靶向步骤来获得纯合性无效**VELOCIMMUNE<sup>®</sup>**ES细胞克隆, 此是耗时的。更重要的是, 不仅**VELOCIMMUNE<sup>®</sup>**ES细胞克隆在传统上在用于免疫计划的K0情况下展现用以产生完全ES细胞源性F0**VELOCIMICE<sup>®</sup>** (即由将ES细胞注射至8细胞期胚胎中获得的完全ES细胞源性F0代小鼠) 的能力低下 (参见例如表2), 而且依序靶向的**VELOCIMMUNE<sup>®</sup>**ES细胞克隆展现用以产

生完全ES细胞源性F0 **VELOCIMICE**<sup>®</sup> (即由将ES细胞注射至8细胞期胚胎中获得的完全ES细胞源性F0代小鼠)的能力甚至进一步降低。参见例如表3 (比较使用用于产生靶向遗传修饰和**VELOCIMICE**<sup>®</sup>的典型ES细胞系 (F1H4ES细胞系) 以及两种通用轻链 (ULC) ES细胞系和**VELOCIMMUNE**<sup>®</sup> ES细胞系 (VI-3Adam6) 达成的 **VELOCIMOUSE**<sup>®</sup>产生效率)。

[0427] 表2.ES细胞系在用于免疫计划的K0情况下的**VELOCIMOUSE**<sup>®</sup>产生效率。

| ESC 株系     | 总体经基因分型的<br><b>VELOCIMICE</b> <sup>®</sup> | 总体经注射的<br>胚胎 | <b>VELOCIMOUSE</b> <sup>®</sup><br>生产率 |
|------------|--------------------------------------------|--------------|----------------------------------------|
| ULC1-39 F2 | 36                                         | 6788         | 0.50%                                  |
| ULC1-39 A4 | 2                                          | 150          | 1.30%                                  |
| VI3Adam-B3 | 163                                        | 2112         | 7.72%                                  |

[0429] 表3.ES细胞系的总体**VELOCIMOUSE**<sup>®</sup>产生效率。

| ESC 株系       | <b>VELOCIMICE</b> <sup>®</sup> /显微注射胚胎% |       |
|--------------|-----------------------------------------|-------|
|              | 第一电穿孔                                   | 依序电穿孔 |
| F1H4         | 18.5%                                   | 16.6% |
| ULC1-39 A4   | 2.0%                                    | 2.6%  |
| ULC 1-39 F2  | 2.1%                                    | 0.9%  |
| VI-3Adam6 B3 | 11.1%                                   | 5.2%  |

[0431] 为产生对目标外来靶抗原的耐受性降低的小鼠,我们已开发一种用以在单一修饰步骤中快速产生包含功能性异位小鼠Adam6基因的就在目标靶标处的无效等位基因来说是纯合的**VELOCIMMUNE**<sup>®</sup> ES细胞的方法。我们已使使用一对引导RNA来在包含功能性异位小鼠Adam6基因的**VELOCIMMUNE**<sup>®</sup> ES细胞中的目标靶标的两个等位基因上高效产生大型缺失的程序最优化,由此避免需要设计和产生大型靶向载体 (LTVEC)。使用这个方法,就在目标靶标处的无效等位基因来说是纯合的以及准备用于免疫的F0 **VELOCIMICE**<sup>®</sup>可在4至5个月而非15至16个月内分娩 (就在目标靶标处的无效等位基因来说是纯合的小鼠幼仔可在约3个月内分娩,但接着老化4-5周以进行免疫)。在这个实验中,设计和克隆成对引导RNA以靶向用于纯合性缺失的与那些目标外来靶抗原直系同源的自身抗原。引导RNA被设计来靶向编码自身抗原的内源性基因的起始密码子和终止密码子区域。对于一些靶标,设计两对gRNA (v1和v2)。引导RNA设计过程描述于以下材料和方法中。将引导RNA连同Cas9一起电穿孔或核转染至源于**VELOCIMMUNE**<sup>®</sup>小鼠的ES细胞中,所述小鼠包括功能性异位小鼠Adam6基因 (VI-3Adam6) 小鼠 (用相应人DNA以及重新插入的小鼠Adam6基因替换内源性小鼠免疫球蛋白重链和轻链可变区) 或通用轻链 (ULC 1-39) 小鼠 (具有是人Vκ1-39/J基因区段的单一重排人免疫球蛋白轻链可变区的小鼠)。参见图32。电穿孔和核转染方案描述于以下材料和方法中。在一些实验中,将Cas9和成对引导RNA连同靶向用于缺失的编码自身抗原的内源性基因的大型靶向载体 (LTVEC) 一起电穿孔 (参见例如图4)。在与或不与LTVEC一起使用CRISPR/Cas9 (CC9) 的情况下观察到相当的缺失效率 (参见表4)。

[0432] 表4. 双等位基因缺失效率。

[0433]

| 自身抗原            | 亲本 ESC       | EP 类型         | 筛选的克隆 | 具有双等位基因缺失的克隆 | 效率 (%) |
|-----------------|--------------|---------------|-------|--------------|--------|
| 自身抗原 1<br>(细胞质) | ULC1-39 F2   | LTVEC + CC9v1 | 384   | 19           | 4.9    |
|                 |              | CC9v1         | 384   | 20           | 5.2    |
|                 |              | CC9v2         | 176   | 14           | 7.9    |
|                 | VI-3Adam6 B3 | LTVEC + CC9v1 | 384   | 9            | 2.3    |
|                 |              | CC9v1         | 384   | 15           | 3.9    |
|                 |              | CC9v2         | 352   | 19           | 5.4    |
| 自身抗原 2<br>(跨膜)  | ULC1-39 F2   | LTVEC + CC9v1 | 384   | 14           | 3.6    |
|                 |              | CC9v1         | 384   | 21           | 5.5    |
|                 |              | CC9v2         | 352   | 42           | 11.9   |
|                 | VI-3Adam6 B3 | LTVEC + CC9v2 | 384   | 11           | 2.9    |
|                 |              | CC9v2         | 384   | 11           | 2.9    |
|                 |              | CC9v2         | 176   | 20           | 11.3   |

[0434]

| 自身抗原           | 亲本 ESC       | EP 类型       | 筛选的克隆 | 具有双等位基因缺失的克隆 | 效率 (%) |
|----------------|--------------|-------------|-------|--------------|--------|
| 自身抗原 3<br>(跨膜) | ULC1-39 F2   | LTVEC + CC9 | 384   | 12           | 3.1    |
|                |              | CC9         | 384   | 11           | 2.9    |
|                | VI-3Adam6 B3 | LTVEC + CC9 | 384   | 11           | 2.9    |
| 自身抗原 4<br>(跨膜) | ULC1-39 F2   | LTVEC + CC9 | 176   | 11           | 6.3    |
|                |              | CC9         | 176   | 3            | 1.7    |
|                | VI-3Adam6 B3 | CC9         | 352   | 76           | 21.6   |
| 自身抗原 5<br>(跨膜) | ULC1-39 F2   | LTVEC + CC9 | 192   | 8            | 4.2    |
|                |              | CC9         | 384   | 8            | 2.1    |
|                | VI-3Adam6 B3 | CC9         | 352   | 15           | 4.3    |
| 自身抗原 6<br>(跨膜) | ULC1-39 F2   | CC9         | 176   | 10           | 5.6    |
| 自身抗原 7<br>(跨膜) | ULC1-39 F2   | LTVEC + CC9 | 352   | 10           | 2.8    |
|                |              | CC9         | 352   | 5            | 1.4    |
|                | VI-3Adam6 B3 | CC9         | 352   | 7            | 2      |

[0435] 从实验开始 (gRNA设计) 至结束 (对于编码自身抗原的内源性基因, 具有纯合性无效等位基因的经基因分型的F0小鼠) 的时间安排是约3个月。作为一实例, 用于产生就对应于靶标1的自身抗原 (自身抗原1) 来说是纯合性无效的F0小鼠的时间安排显示于表5中。

[0436] 表5. 用以在VI-3 Adam6小鼠中递送自身抗原1的纯合性无效等位基因的时间安排。

[0437]

| 过程                            | 日期       |
|-------------------------------|----------|
| gRNA 和 TAQMAN <sup>®</sup> 设计 | 9/21/15  |
| gRNA 制备                       | 10/6/15  |
| 电穿孔                           | 10/16/15 |
| 初级 ES 细胞筛选                    | 11/2/15  |
| 再确认筛选                         | 11/18/15 |
| 显微注射 ES 细胞克隆                  | 11/23/15 |
| F0 小鼠出生日期                     | 12/11/15 |
| F0 小鼠基因分型                     | 12/21/15 |

[0438] 单独或连同靶向用于缺失的自身抗原的大型靶向载体 (LTVEC) 一起使用靶向各自自身抗原的起始密码子和终止密码子区域的成对引导RNA, 进行若干实验以靶向来自VI-3-Adam6小鼠和ULC 1-39小鼠的胚胎干 (ES) 细胞中的用于缺失的各种自身抗原。将Cas9和引导RNA以DNA形式引入ES细胞中。如表6和图28中所示, 对于所有测试自身抗原都实现缺失 (即垮塌), 其中缺失大小在0.1kb与165kb之间的范围内, 并且在缺失 (即垮塌) 的大小与产生缺失 (即垮塌) 的效率之间存在负性相关。也可实现大得多的大小的双等位基因垮塌。举例来说, 我们已实现约400kb的缺失大小的双等位基因垮塌。同样, 通过使用两个5' gRNA和两个3' gRNA以及修复载体, 以约1.2%的效率实现在小鼠IgH基因座处的约900 kb-1 Mb双等位基因垮塌 (数据未显示)。

[0439] 表6. 缺失 (垮塌) 大小对缺失 (垮塌) 效率的影响。

| 小鼠靶标    | 垮塌大小(kb) | 筛选的克隆 | NHEJ 双等位基因插入缺失 |         |
|---------|----------|-------|----------------|---------|
|         |          |       | 垮塌             | 垮塌效率(%) |
| 自身抗原 4  | 0.1      | 528   | 76             | 14.3    |
| 自身抗原 10 | 0.1      | 352   | 48             | 13.6    |
| 自身抗原 11 | 1.3      | 352   | 4              | 1.1     |
| 自身抗原 12 | 1.6      | 176   | 29             | 16.5    |
| 自身抗原 5  | 2.2      | 736   | 7              | 0.9     |
| 自身抗原 13 | 2.6      | 352   | 52             | 14.8    |
| 自身抗原 14 | 2.8      | 352   | 18             | 5.1     |
| 自身抗原 15 | 4.3      | 352   | 39             | 11.1    |
| 自身抗原 16 | 4.5      | 352   | 30             | 8.5     |
| 自身抗原 17 | 4.6      | 352   | 36             | 10.2    |
| 自身抗原 18 | 6        | 176   | 24             | 13.6    |
| 自身抗原 2  | 15.1     | 528   | 62             | 11.7    |
| 自身抗原 9  | 18.4     | 440   | 119            | 27      |
| 自身抗原 19 | 24.5     | 176   | 25             | 14.2    |
| 自身抗原 7  | 25.3     | 704   | 12             | 1.7     |
| 自身抗原 20 | 25.7     | 352   | 45             | 12.8    |
| 自身抗原 21 | 26.2     | 352   | 30             | 8.5     |
| 自身抗原 6  | 28.9     | 176   | 10             | 5.7     |
| 自身抗原 3  | 39       | 280   | 11             | 3.9     |
| 自身抗原 1  | 45.7     | 528   | 33             | 6.3     |
| 自身抗原 22 | 58       | 176   | 2              | 1.1     |
| 自身抗原 23 | 84.4     | 352   | 9              | 2.6     |
| 自身抗原 24 | 95.4     | 792   | 8              | 1       |
| 自身抗原 25 | 165      | 704   | 13             | 2.4     |

[0441] 类似于使用ES细胞的程序, 为产生对目标外来靶抗原的耐受性降低的小鼠, 我们也已开发一种用以在单一修饰步骤中快速产生就在目标靶标处的无效等位基因来说是纯合的单细胞期胚胎的方法。我们已使用一对引导RNA来在单细胞期胚胎中的目标靶标的两个等位基因上高效产生大型缺失的程序最优化, 由此避免需要设计和产生大型靶向载体 (LTVEC)。此外, 相较于使用ES细胞系 (例如ULC 1-39 ES细胞系), 使用单细胞期胚胎可改

进所靶向小鼠的产生效率。使用这个方法,准备用于免疫的就在目标靶标处的无效等位基因来说是纯合的F0小鼠可在4至5个月(就在目标靶标处的无效等位基因来说是纯合的F0小鼠幼仔可在约3个月内分娩)而非15至16个月内分娩。在这个实验中,设计和克隆成对引导RNA以靶向用于纯合性缺失的与那些目标外来靶抗原直系同源的自身抗原。引导RNA被设计来靶向编码自身抗原的内源性基因的起始密码子和终止密码子区域。引导RNA设计过程描述于以下材料和方法中。简要说,使超排卵雌性与配种雄性交配以产生胚胎。如果仅有少许雄性可用,那么使用体外受精。雌性龄期范围是3-16周,每个供体的卵母细胞在15-46的范围内(中值=32个卵母细胞),并且每个供体的接合子在5-32的范围内(中值=15个接合子)。将引导RNA连同Cas9 mRNA一起显微注射(细胞质注射)至来自 **VELOCIMMUNE<sup>®</sup>** 小鼠的单细胞期胚胎中,所述小鼠包括功能性异位小鼠Adam6基因(VI-3Adam6)小鼠(用相应人DNA以及重新插入的小鼠Adam6基因替换内源性小鼠免疫球蛋白重链和轻链可变区)或通用轻链(ULC 1-39)小鼠(具有是人V $\kappa$ 1-39/J $\kappa$ 5基因区段的单一重排人免疫球蛋白轻链可变区的小鼠)。注射的胚胎的数目在99-784的范围内(中值=334),存活的胚胎的百分比在56%-73%的范围内(中值=63%),转移的胚胎的数目在59-442的范围内(中值=226),各计划的幼仔的数目在10-46的范围内(中值=32),并且出生率在2%-59%的范围内(中值=13%)。如表7和图29中所示,对于所有测试自身抗原都产生携带靶向缺失(即垮塌)的活体幼仔,其中缺失大小在0.1kb与94kb之间的范围内,并且在缺失(即垮塌)的大小与产生携带缺失(即垮塌)的小鼠幼仔的效率之间存在负性相关。

[0442] 表7. 通过在胚胎中进行Cas9注射达成的敲除。

| 自身抗原编号 | 大小(kb) | 龄期(周) | 卵供体 | 卵/供体 | 接合子/供体 | 注射数 | ET数 | 幼仔数 | 出生% | 外显子1 NHEJ效率 | 活体幼仔的垮塌效率(无效幼仔数) |
|--------|--------|-------|-----|------|--------|-----|-----|-----|-----|-------------|------------------|
| 3      | 39     | 7     | 5   | 33   | 20     | 99  | 59  | 35  | 59% | 66%         | 27%<br>(10个幼仔)   |
| 4      | 0.1    | 3~7   | 30  | 25   | 9*     | 275 | 169 | 28  | 17% | 96%         | 25%<br>(7个幼仔)    |
| 26     | 15     | 6~7   | 20  | 16   | 9*     | 267 | 189 | 19  | 10% | 100%        | 58%<br>(11个幼仔)   |
| 14     | 2.7    | 7~10  | 21  | 35   | 16     | 334 | 226 | 43  | 19% | 98%         | 85%<br>(36个幼仔)   |
| 10     | 3.9    | 7~15  | 49  | 30   | 13     | 784 | 442 | 10  | 2%  | N/A         | 66%<br>(6个幼仔)    |
| 17     | 4.5    | 10    | 6   | 38   | 16     | 290 | 212 | 39  | 18% | 99%         | 57%<br>(22个幼仔)   |
| 6      | 29     | 10~11 | 21  | 39   | 20     | 556 | 366 | 46  | 13% | 91%         | 31%<br>(14个幼仔)   |
| 1      | 46     | 10~16 | 25  | 32   | 19     | 470 | 275 | 32  | 12% | 84%         | 22%<br>(7个幼仔)    |
| 27     | 19.8   | 11~16 | 24  | 32   | 20     | 491 | 307 | 24  | 8%  | 80%         | 45%<br>(10个幼仔)   |
| 13     | 2.6    | 14~15 | 17  | 31   | 19     | 326 | 223 | 47  | 21% | 100%        | 86%<br>(40个幼仔)   |
| 16     | 4.5    | 14~15 | 17  | 34   | 17     | 286 | 173 | 43  | 25% | 97%         | 83%<br>(34个幼仔)   |
| 23     | 84     | 11~13 | 20  | 34   | 19     | 253 | 130 | 11  | 8%  | 76%         | 7%<br>(3个幼仔)     |
| 24     | 94     | 12~15 | 36  | 25   | 13     | 463 | 317 | 56  | 18% | 83%         | 13%<br>(1个幼仔)    |
| 15     | 4.2    | 12~15 | 23  | 24   | 12     | 286 | 174 | 44  | 25% | 88%         | 50%<br>(10个幼仔)   |

[0444] \*IVF或三联体(而非成对自然交配)

[0445] 材料和方法

[0446] 引导RNA和**TAQMAN<sup>®</sup>**测定设计:基于各基因座的共有编码序列(CCDS)以格式

5' NNNNNNNNNNNNNNNnnNNNGG 3' (SEQ ID NO:2) 设计具有23个碱基对的长度的引导RNA (gRNA), 其中N是任何核苷酸。最后三个核苷酸 (NGG) 是原间隔体邻近基序 (PAM), 并且由 Cas9 酶达成的双链钝性末端DNA裂解发生在NGG的5' 的3个核苷酸处 (在以上小写残基之间)。基于从包括 [crispr.mit.edu](http://crispr.mit.edu), [crispr.med.harvard.edu/sgRNAScorer/](http://crispr.med.harvard.edu/sgRNAScorer/) 和 [broadinstitute.org/rnai/public/analysis-tools/sgrna-design](http://broadinstitute.org/rnai/public/analysis-tools/sgrna-design) 的各种gRNA搜索引擎获得的评分选择gRNA。简要说, 在两个DNA链上分别测定紧邻起始ATG的5' 和3' 的100-150bp序列和紧邻终止密码子的5' 和3' 的100-150bp以获得gRNA。进一步探询根据使用的所有搜索引擎都具有高分的接近ATG的两个gRNA (彼此重叠不超过25%) 和接近终止密码子的两个gRNA在小鼠基因组中的独特性, 以及在通用轻链 (ULC或共同轻链)、包含功能性异位小鼠Adam6基因的 **VELOCIMMUNE**<sup>®</sup> 小鼠 (VI-3-Adam6) 和VGB6 **VELOCIGENE**<sup>®</sup> 小鼠胚胎干细胞 (ESC) 系中无单核苷酸变化 (SNV)。如果使用以上搜索规程未发现高评分引导物, 那么搜索在ATG和终止密码子周围的额外序列直至发现两个高品质引导物。

[0447] 采用 **APPLIED BIOSYSTEMS**<sup>®</sup> 定制 **TAQMAN**<sup>®</sup> MGB 探针, 使用 **PRIMER EXPRESS**<sup>®</sup> 设计 **TAQMAN**<sup>®</sup> 测定以使探针序列始终重叠于各引导物的 cas9 切割位点。

一些 **TAQMAN**<sup>®</sup> 测定也使用 Biosearch Technologies 双重标记的 **BHQ**<sup>®</sup> 探针 ([biosearchtech.com/ProbeITy/design/inputsequences.aspx](http://biosearchtech.com/ProbeITy/design/inputsequences.aspx)) 获得。如果 Cas9 切割由引导物结合的序列, 那么这些测定充当等位基因丧失测定。所有测定都针对 SNV 加以筛选。如下对引导物命名: mGU 和 mGU2 (针对小鼠基因组上游); 以及 mGD 和 mGD2 (针对小鼠基因组下游)。如下对 **TAQMAN**<sup>®</sup> 测定命名: mTGU 和 mTGU2 (针对分别涵盖 mGU 和 mGU2 的 **TAQMAN**<sup>®</sup> 测定), 以及 mTGD 和 mTGD2 (针对分别涵盖 mGD 和 mGD2 的 **TAQMAN**<sup>®</sup> 测定)。设计大致与引导物 mGU/mGU2 和 mGD/mGD2 等距离, 在待垮塌的基因座的中间进行的额外 **TAQMAN**<sup>®</sup> 测定, 被称为 mTM (针对小鼠 **TAQMAN**<sup>®</sup> Middle)。这个等位基因丧失测定确定是否发生由引导物侧接的区域的缺失 (垮塌)。

[0448] 设计在 mGU/mGU2 (无论哪个在最5') 的上游以及在 mGD/mGD2 (无论哪个在最3') 的下游 200-800bp 进行的其他 **TAQMAN**<sup>®</sup> 测定。这些测定被分别称为 retU (针对上游保留) 和 retD (针对下游保留)。这些测定描绘最大可接受的缺失大小, 并且如上针对 SNV 加以筛选。

[0449] 引导RNA克隆: 设计和合成引导RNA双链体。因为U6启动子优选以鸟嘌呤起始, 所以如果序列未以鸟嘌呤起始, 那么将鸟嘌呤添加至5'。将冻干gRNA双链体用无菌水再混悬至100μM, 并且在0.5mL微量离心管中建立以下连接反应: 14.5μL PCR合格水; 2μL 10X T4 DNA连接酶缓冲液 (NEB)、1μL pMB\_sgRNA\_BsmBI载体 (约60ng)、1μL gRNA双链体 (100μM) 和 1.5μL T4 DNA连接酶 (40U/μL; NEB)。接着在室温下孵育连接反应物1小时, 并且随后用于在 TOP10 细胞的情况下进行的转化反应中。接着挑选菌落, 并且通过PCR和测序来检查。

[0450] **BTX**<sup>®</sup> 电穿孔方案: 如下制备引导混合物: 10μg 各sgRNA质粒和5μg Cas9野生型质

粒。在电穿孔当天,在电穿孔过程之前半小时至一小时,将细胞用ES培养基喂养。接着将细胞用PBS洗涤两次,并且添加0.25%胰蛋白酶-EDTA并在37℃下孵育细胞15分钟。在孵育之后将一个或多个板轻叩,添加ES培养基以中和胰蛋白酶,将细胞温和吸移4次以拆散细胞凝块并转移至一个或多个凝胶板中,并且在37℃下孵育细胞20分钟。将一个或多个板振荡并用培养基温和洗涤一次,并且接着将所有细胞都转移至15-mL管中,接着在1200rpm下使所述管旋转5分钟。在10mL PBS中合并所有沉淀,并且对细胞计数并在必要时进行稀释。将20μl体积的细胞混悬液添加至CELLOMETER<sup>®</sup>载片,并且使用Nexcelom CELLOMETER AUTO T4<sup>™</sup>细胞活力计数器进行计数。接着在1200rpm下使管离心5分钟。将沉淀再混悬于电穿孔缓冲液中,各次电穿孔使用 $7.5 \times 10^6$ 个细胞。将细胞添加至微型离心管中的引导混合物中,其中各管中的体积是120μl。将管混合2-3次并使用宽孔尖端转移至96孔电穿孔比色皿(2mm间隙)中,并且将比色皿密封。使用BTX<sup>®</sup> ECM<sup>®</sup> 630电穿孔仪,在700V、400 Ω、25uF下递送电脉冲。接着在冰上孵育比色皿10分钟。接着将经电穿孔细胞转移至深孔板中(在比色皿在冰上时添加0.8mL/孔)。将细胞涂铺于2x15cm凝胶板上/用25mL培养基在各板中进行扩散。短暂选择以1μg/mL嘌呤霉素起始,持续3天,并且直至电穿孔后10天将培养基变换为非选择培养基,此时挑选集落。

[0451] **NUCLEOFECTOR<sup>®</sup>**电穿孔方案:在电穿孔当天,在电穿孔过程之前半小时至一小时,将细胞用ES培养基喂养。接着将细胞用10mL PBS洗涤两次,并且添加2mL 0.25%胰蛋白酶-EDTA并在37℃下孵育细胞15分钟。在孵育之后将一个或多个板轻叩,添加8mL ES培养基以中和胰蛋白酶,将细胞温和吸移4次以拆散细胞凝块并转移至一个或多个凝胶板中,并且在37℃下孵育细胞20分钟。将一个或多个板振荡并用培养基温和洗涤一次,并且接着将所有细胞都转移至15-mL管中,接着在90X g下使所述管旋转3分钟。将沉淀再混悬于10mL PBS中,并且对细胞计数并在必要时进行稀释。将20μl体积的细胞混悬液添加至CELLOMETER<sup>®</sup>载片中,并且使用Nexcelom Vision CBA系统进行计数。将总计 $2 \times 10^6$ 个细胞等分,并且在EPPENDORF<sup>®</sup>管中在90X g下离心3分钟。接着以100μL的总体积将沉淀再混悬于与5μg Cas9野生型质粒和2.5μg各sgRNA质粒混合的LONZA<sup>®</sup> P4缓冲液中。接着将细胞转移至大的LONZA<sup>®</sup>比色皿中。使用LONZA<sup>®</sup> 4D-NUCLEOFECTOR<sup>™</sup>和程序CP-105递送电脉冲。添加400μL体积的新鲜ES培养基,并且将细胞转移至新的EPPENDORF<sup>®</sup>管中以进行混合。接着将细胞涂铺于具有10mL ES培养基的2x 10cm凝胶板上。短暂选择在EP后2天用嘌呤霉素(1.5μg/mL)起始,持续2天。在选择之后,直至电穿孔后10天使用非选择培养基,此时挑选集落。

[0452] 筛选:使用TAQMAN<sup>®</sup>测定mTGU、mTGU2、mTGD和mTGD2评估采用引导物mGU、mGU2、mGD和mGD2由Cas9达成的切割。当拷贝数从2(亲本未修饰对照DNA)降低至1时,确定在一个等位基因而非另一等位基因处进行切割。当测定产生拷贝数0时,确定由Cas9达成纯合性裂解。因为Cas9在ATG和终止密码子附近的切割不保证移除间插序列,所以当mTM测定数值从2(亲本)变为1或0时分别评估杂合性垮塌和纯合性垮塌。最后,使用retU测定和retD测定设置缺失大小的外部限值。在拷贝数是2的情况下,保留测定将保持完整(得以保留),如同亲

本一样。

[0453] 在用mGU、mGU2、mGD和mGD2或其某一组合进行电穿孔之后获得的ESC克隆首先使用测定mTGU、mTGU2、mTM、mTGD和mTGD2或其某一组合针对Cas9裂解和/或垮塌加以筛选。对于所有测定都具有0拷贝数的集落接着使用retU和retD加以进一步筛选,并且仅有retU和retD拷贝数是2的集落得以通过以进行进一步分析。

[0454] 对小鼠胚胎干细胞的初级和再确认筛选:通过TAQMAN<sup>®</sup> LOA (等位基因丧失)多路(4路) qPCR来针对靶基因座的纯合性缺失筛选经修饰mESC集落。对于第一轮筛选(初级),在两个96孔板的第1-11列中分离176个独特克隆的DNA。第12列用先前从相同mESC亲本品系分离并用作拷贝数校正物的野生型ES细胞DNA填充;因此待筛选的各DNA板含有88个经修饰克隆和8个校正克隆。将各克隆的DNA一式四份分配至384孔板中,并且利用用于同时在靶基因的相对上游、中间和下游区域中测定拷贝数的采用FAM、VIC、ABY和Quasar的TAQMAN<sup>®</sup>探针,在单一反应混合物中跨越靶基因座的三个区域测定纯合性LOA,其中Quasar扩增Wnt-2b以校正DNA浓度。在测定拷贝数之后,选择跨越靶基因座的所有三个测定的多达8个具有0拷贝的“最佳”品质克隆以进行后续生长扩增,再涂铺,以及经受扩增谱系的拷贝数测定(再确认)。将各扩增克隆涂铺,并且以6个重复分离DNA,从而占据96孔板的1行(A-H)的前6列,由此为使用的额外测定提供额外遗传物质和数据重复。重复用于初级筛选中的测定以确认初级基因型,并且保留测定用于确定缺失程度。保留测定定位在就在靶向缺失区域的上游和下游,并且通常等于2个拷贝。额外测定用于确认在小鼠免疫球蛋白重(IgH)和κ(Igκ)基因座处的亲本ESC基因型(对于IgH和Igκ小鼠是LOA,并且对于人源化是GOA)。

[0455] 用以鉴定Cas9介导的等位基因的下一代测序(NGS):使用标准盐沉淀方法提取来自Cas9修饰的F0小鼠的小型尾部活检组织的基因组DNA。对于各靶基因座,在虑及以下考虑事项下设计PCR引物:(1)扩增子大小的长度在280-380bp之间;(2)gRNA裂解位点在PCR产物内的中心,其中引物离gRNA裂解位点至少35bp以容纳较大插入/缺失(插入缺失),(3)引物的长度是22-25bp,具有在62-65℃之间的熔解温度(T<sub>m</sub>),在3'末端上具有2bp CG钳;和(4)针对BALB/c、C57BL/6或129品系的基因组序列单核苷酸变化检查引物。接着将由ILLUMINA<sup>®</sup>提供的特定通用衔接子序列添加至基因座特异性序列。所得扩增子在琼脂糖凝胶上可视化,并且使用THERMO FISHER SCIENTIFIC<sup>®</sup> SEQUALPREP<sup>™</sup>标准化板试剂盒加以纯化/标准化。产物通过QUBIT<sup>®</sup>来定量,并且1ng各产物用于通过用NEXTERA<sup>®</sup>引物和NEXTERA<sup>®</sup>PCR主混合物进行的额外PCR确定条型码的模板。在热循环仪中采用以下条件进行PCR:在72℃下持续3分钟,在95℃下持续30秒,12个循环的{95℃持续10秒,55℃持续30秒,72℃持续30秒},72℃持续5分钟,以及10℃保持。所得条型码化PCR产物通过AMPURE<sup>®</sup> XP珠粒来纯化,使用NEXTERA<sup>®</sup> XT试剂盒中的ILLUMINA<sup>®</sup>标准化珠粒加以标准化,汇合,并且装载至MISEQ<sup>™</sup>中以进行测序和原始数据收集。

[0456] 对8细胞期小鼠胚胎的显微注射:将约2mL标准ES细胞培养基(-LIF)添加至无菌35mm培养皿盖中,并且用经过滤矿物油覆盖。使用口吸移器将ES细胞涂铺于培养皿的下半

部上。将低温保存的8细胞期SW宿主胚胎朝向培养皿的上部放置。为帮助使在注射期间的胚胎损害最小化,新的注射吸移器的尖端通过相对于固持吸移器进行温和敲击来钝化。基于形态和亮度选择ES细胞,并且采集至注射吸移器中。使胚胎位于固持吸移器上以使分裂球之间的间隙存在于3点钟位置。通过在3点钟位置小心穿过透明带进行穿刺以及将细胞放置在那个点处来将ES细胞引入胚胎的卵黄周隙中。每个胚胎引入总计7-9个ES细胞。将经注射胚胎放置至含有用经过滤矿物油覆盖的一滴KSOM胚胎培养基的35mm培养皿中,并且将胚胎在37.0℃下在7.5%CO<sub>2</sub>下培养过夜。次日早晨,将胚胎通过手术转移至假妊娠雌性中。

[0457] 实施例2.使用靶向起始密码子的区域的多个引导RNA产生用于抗体产生的KO ES细胞和小鼠。

[0458] 在用以产生对目标外来靶抗原的耐受性降低的小鼠的另一实验中,设计和克隆三个引导RNA以靶向用于纯合性缺失的与那些外来靶抗原直系同源的自身抗原。三个重叠引导RNA被设计来靶向涵盖编码自身抗原的内源性基因的起始密码子的重叠区域(参见图5)。将引导RNA连同Cas9一起电穿孔或核转染至源于通用轻链(ULC 1-39)小鼠(在它们的种系中包含以下的小鼠:(i)重排V<sub>κ</sub>/J<sub>κ</sub>序列在内源性小鼠κ免疫球蛋白轻链可变区基因座处的插入,所述重排V<sub>κ</sub>/J<sub>κ</sub>序列包含:单一人种系V<sub>κ</sub>序列;和单一人种系J<sub>κ</sub>序列,其中所述重排V<sub>κ</sub>/J<sub>κ</sub>序列可操作地连接至内源性小鼠κ恒定区;和(ii)可操作地连接至内源性小鼠免疫球蛋白重链恒定区的多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入)的ES细胞中。在一些实验中,将Cas9和三个引导RNA连同靶向用于缺失的编码自身抗原的内源性基因的大型靶向载体(LTVEC)一起电穿孔(参见例如图5)。与CRISPR/Cas9(CC9)组合使用LTVEC使在靶基因座处得到双等位基因突变的机会显著增加(参见表8),但用LTVEC和CRISPR/Cas9进行靶向需要多得多的筛选来排除假阳性。

[0459] 表8.双等位基因缺失效率。

| [0460] | 自身抗原           | 亲本 ESC     | EP 类型       | 筛选的<br>集落 | 具有双等位基<br>因修饰的克隆 | 效率<br>(%) |
|--------|----------------|------------|-------------|-----------|------------------|-----------|
|        | 自身抗原 8<br>(跨膜) | ULC1-39 F2 | LTVEC + CC9 | 384       | 111              | 28.9      |
|        |                |            | CC9         | 192       | 28               | 14.6      |

[0461] 实施例3.使小鼠免疫以及分析对免疫原的血清抗体应答。

[0462] 免疫

[0463] 使用多种免疫原诸如蛋白质,用众多跨膜靶标使包含功能性异位小鼠Adam6基因的VELOCIMMUNE<sup>®</sup>小鼠(VI-3)、通用轻链(ULC 1-39)小鼠(在它们的种系中包含以下的小鼠:(i)重排V<sub>κ</sub>/J<sub>κ</sub>序列在内源性小鼠κ免疫球蛋白轻链可变区基因座处的插入,所述重排V<sub>κ</sub>/J<sub>κ</sub>序列包含:单一人种系V<sub>κ</sub>序列;和单一人种系J<sub>κ</sub>序列,其中所述重排V<sub>κ</sub>/J<sub>κ</sub>序列可操作地连接至内源性小鼠κ恒定区;和(ii)可操作地连接至内源性小鼠免疫球蛋白重链恒定区的多个人免疫球蛋白重链可变区基因区段在内源性小鼠免疫球蛋白重链可变区基因座处的插入)、KO(敲除)/VI-3小鼠(其中与外来靶抗原直系同源的自身抗原被敲除的VI-3小鼠)和KO/ULC 1-39小鼠(其中与外来靶抗原直系同源的自身抗原被敲除的ULC 1-39小鼠)免疫。在启动免疫之前从小鼠收集免疫前血清。使用标准佐剂,通过不同途径在不同时间间隔下使小鼠加强免疫,总计进行3-6次加强免疫。将小鼠定期放血,并且测定关于相应抗原的抗血清效价。在靶标8的实例中,通过足垫途径用靶标8的具有小鼠Fc标签的重组细胞外

结构域使小鼠免疫。效价来自第2次放血(对于ULC 1-39,在致敏+6次加强免疫之后)或第3次放血(对于自身抗原-8-KO/ULC 1-39,在致敏+3次加强免疫之后)。

#### [0464] 抗血清效价测定

[0465] 使用ELISA测定血清中的针对相应免疫原的抗体效价。将96孔微量滴定板(THERMO SCIENTIFIC®)用于磷酸盐缓冲盐水(PBS, IRVINE SCIENTIFIC®)中的相应靶抗原在2μg/mL下涂布过夜。将板用含有0.05%吐温20(Tween 20)的磷酸盐缓冲盐水(PBS-T, SIGMA-ALDRICH®)洗涤,并且用250μl含0.5%牛血清白蛋白(BSA, SIGMA-ALDRICH®)的PBS在室温下阻断1小时。将板用PBS-T洗涤。将免疫前抗血清和免疫抗血清于0.5%BSA-PBS中连续稀释3倍,并且在室温下添加至板中,持续1小时。将板洗涤,并且将山羊抗小鼠IgG-Fc-辣根过氧化物酶(HRP)缀合的二级抗体(Jackson ImmunoResearch)添加至板中并在室温下孵育1小时。将板洗涤,并且使用TMB/H<sub>2</sub>O<sub>2</sub>作为底物,通过孵育20分钟来显色。反应用酸终止,并且在分光光度计(VICTOR®, PERKINELMER®)上在450nm下对板进行读取。使用GRAPHPAD PRISM®软件计算抗体效价。在靶标8的实例中,使用的滴定抗原是人靶标8的具有Myc-Myc-His标签的重组细胞外结构域。

#### [0466] 结果

[0467] 通过用不同跨膜靶标进行免疫来探究在VI-3、ULC1-39、KO/VI-3和KO/ULC 1-39小鼠中的体液免疫应答。对于免疫的所有靶标,在KO/VI-3和KO/ULC 1-39品系中都引发高抗体效价。在VI-3和ULC 1-39小鼠品系中,效价也较高。然而,一般来说,KO品系似乎具有更大的效价应答。引发的免疫应答在表9中表示为抗体效价,其定义为抗原结合吸光度是本底的两倍高所处的最高血清稀释度的倒数。因此,数值越高,对免疫原的体液免疫应答越大。总之,超过16种靶标已在KO品系中得以成功免疫。针对靶标1和9的单克隆抗体已通过BST和杂交瘤平台来分离,并且针对靶标4和5的单克隆抗体已通过BST来分离,并且正在进行对这些抗体的进一步表征。ULC 1-39和自身抗原-8-KO/ULC 1-39小鼠中针对一种目标人靶抗原(靶标8;与小鼠自身抗原8直系同源,上文)的抗体产生的数据提供于表9中和图6中。相比于野生型ULC 1-39小鼠,F0KO小鼠引发的对蛋白质激发的应答是约5倍高,如由针对靶标的中值抗体效价所指示。表9中也提供特异性结合至抗原的抗体的数目(在两倍于本底吸光度的吸光度下)。在自身抗原-9-KO/VI-3小鼠相较于VI-3小鼠的情况下显示类似结果。参见图30A和图30B。在这个实验中,通过真皮内途径来用编码野生型靶标9的任一DNA使野生型VI-3-Adam6小鼠和自身抗原-9-KO/VI-3-Adam6小鼠免疫。使用被工程化来表达靶标9的细胞或亲本VI-3T3细胞测定效价。尽管来自VI-3-Adam6小鼠的抗体效价与对照差不多,但自身抗原-9-KO/VI-3-Adam6小鼠中的抗体效价得以极大增加。这显示KO/VI-3品系与KO/ULC品系两者均引发稳健免疫应答。

[0468] 表9.比较ULC 1-39品系和KO/ULC 1-39品系中的免疫应答。

| 靶标   | 针对靶标的中值抗体效价          |                    | 抗原阳性单克隆抗体            |                   |
|------|----------------------|--------------------|----------------------|-------------------|
|      | KO/ULC 1-39<br>(n=5) | ULC 1-39<br>(n=10) | KO/ULC 1-39<br>(n=2) | ULC 1-39<br>(n=2) |
| 靶标 8 | 986,890              | 200,387            | 76                   | 61                |

[0470] 实施例4.使小鼠免疫以及分析抗体多样性和V基因区段的使用

[0471] 用靶标3使包含功能性异位小鼠Adam6基因的VELOCIMMUNE<sup>®</sup>小鼠(VI-3)和自身抗原-3-KO(敲除)/VI-3小鼠免疫。在启动免疫之前从小鼠收集免疫前血清。使用标准佐剂,通过不同途径在不同时间间隔下使小鼠加强免疫,总计进行3-6次加强免疫。将小鼠定期放血,并且测定关于相应抗原的抗血清效价。

[0472] 从野生型VI-3小鼠和自身抗原-3-KO VI-3小鼠的脾分离B细胞,并且对抗体测序以确定V基因使用。直接从单一抗原阳性B细胞分离编码VH结构域和VL结构域的DNA,并且测序。参见例如US 7,582,298,其出于所有目的以引用的方式整体并入本文。野生型VI-3小鼠的V基因使用数据呈现于表10中,并且自身抗原-3-KO VI-3小鼠的V基因使用数据呈现于表11中。如表10和表11中所示,相较于野生型VI-3小鼠,在自身抗原-3-KO VI-3小鼠的情况下观察到在重链V基因区段与轻链V基因区段两者的使用方面的更多样性。举例来说,在野生型VI-3小鼠中,仅4个重链V基因区段和6个轻链V基因区段用于抗体,并且79%的抗体使用IgH V4-59和IgκV1-12V基因区段。相比之下,在自身抗原-3-KO VI-3小鼠中,6个重链V基因区段和10个轻链V基因区段用于抗体,其中最普遍的使用组合(IgH V3-23和IgκV4-1)占抗体的仅42%。

[0473] 表10.野生型VI-3小鼠中针对靶标3的抗体的V基因使用。

[0474]

| WT<br>VI3 小<br>鼠 | Igκ<br>V1-5 | Igκ<br>V1-9 | Igκ<br>V1-12 | Igκ<br>V1-16 | Igκ<br>V1-17 | Igκ<br>V1-33 | Igκ<br>V1-39 | Igκ<br>V3-11 | Igκ<br>V3-15 | Igκ<br>V3-20 | Igκ<br>V4-1 | 无 Vκ<br>序列 | 总计  |
|------------------|-------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|------------|-----|
| IgH<br>V1-18     |             |             |              |              |              |              |              |              |              |              |             |            | 0   |
| IgH<br>V3-11     |             |             |              |              |              |              |              | 1            |              |              |             | 2          | 3   |
| IgH<br>V3-23     |             |             | 29           |              |              |              |              |              |              |              |             |            | 29  |
| IgH<br>V3-33     |             |             |              |              |              |              |              |              |              |              |             |            | 0   |
| IgH<br>V3-7      |             |             |              |              |              |              |              |              |              |              |             |            | 0   |
| IgH<br>V3-9      |             |             |              |              |              |              |              |              |              | 1            |             |            | 1   |
| IgH<br>V4-59     | 3           | 1           | 150          |              |              | 2            |              |              |              |              |             |            | 156 |
| 无 VH<br>序列       |             |             | 1            |              |              |              |              |              |              |              |             |            | 1   |
| 总计               | 3           | 1           | 180          | 0            | 0            | 2            | 0            | 1            | 0            | 1            | 0           | 2          | 190 |

[0475] 表11.自身抗原-3-KO VI-3小鼠中针对靶标3的抗体的V基因使用。

[0476]

| KO<br>VI3<br>小鼠 | Igκ<br>V1-5 | Igκ<br>V1-9 | Igκ<br>V1-12 | Igκ<br>V1-16 | Igκ<br>V1-17 | Igκ<br>V1-33 | Igκ<br>V1-39 | Igκ<br>V3-11 | Igκ<br>V3-15 | Igκ<br>V3-20 | Igκ<br>V4-1 | 无 Vκ<br>序列 | 总计  |
|-----------------|-------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-------------|------------|-----|
| IgH<br>V1-18    |             |             |              |              |              | 2            |              |              |              |              |             |            | 2   |
| IgH<br>V3-11    | 1           |             | 1            |              |              |              |              |              |              |              |             |            | 2   |
| IgH<br>V3-23    |             | 33          |              |              | 2            | 1            |              |              |              |              | 56          |            | 92  |
| IgH<br>V3-33    |             |             | 1            |              |              |              | 2            |              |              |              |             |            | 3   |
| IgH<br>V3-7     |             |             |              |              |              |              |              |              |              |              | 1           |            | 1   |
| IgH<br>V3-9     |             |             |              |              |              |              |              |              |              |              |             |            | 0   |
| IgH<br>V4-59    | 4           | 1           | 3            | 3            | 1            | 8            | 6            |              | 1            | 4            |             |            | 31  |
| 无 VH<br>序列      |             |             | 1            |              |              |              |              |              |              |              |             |            | 1   |
| 总计              | 5           | 34          | 6            | 3            | 3            | 11           | 8            | 0            | 1            | 4            | 57          | 0          | 132 |

[0477] 此外,在自身抗原-3-KO VI-3小鼠中产生与小鼠自身抗原3具有交叉反应性的抗体(即结合人靶标3与小鼠自身抗原3两者的抗体)(参见例如表12)。类似结果用自身抗原4和人靶标4在VI-3小鼠与ULC 1-39小鼠两者中观察到。参见图31A和图31B。在这个实验中,使用足垫途径,用His标签化人靶标4蛋白和/或His标签化小鼠自身抗原4蛋白(His标签化)使自身抗原-4-KO/VI-3-Adam6小鼠和自身抗原-4-KO/ULC 1-39小鼠免疫。使用His标签化人靶标4、His标签化小鼠自身抗原4或His标签化Fe1 d 1(对照)作为涂布抗原来测定效价。

[0478] 产生针对在小鼠自身抗原3与靶标3之间共有(或在小鼠自身抗原4与人靶标4之间共有)的表位的抗体的能力是有利的,因为这使抗体汇集物扩增:在野生型VI-3小鼠中未产生与小鼠自身抗原3具有交叉反应性的抗体。此外,由于交叉反应性抗体与野生型小鼠中的内源性自身抗原具有交叉反应性,所以可更易于在体内测试它们的药物动力学性质。因此,可不需要产生被遗传工程化来表达所述交叉反应性抗体的靶抗原(例如人靶抗原)的小鼠。

[0479] 表12.自身抗原-3-KO VI-3小鼠中产生的与自身抗原3具有交叉反应性的抗体。

[0480]

| VH        | Vκ       | 抗体的数目 |
|-----------|----------|-------|
| IgH V3-23 | IgκV1-17 | 2     |
| IgH V3-23 | IgκV1-9  | 1     |
| IgH V3-23 | IgκV4-1  | 55    |
| IgH V4-59 | IgκV3-20 | 1     |

[0481] 实施例5.使用一个引导RNA或两个引导RNA达成的CRISPR/Cas9介导的靶向。

[0482] 材料和方法

[0483] ES细胞培养、筛选和电穿孔

[0484] 本文所述的实验用VGF1(即我们的C57BL6NTac/129S6SvEvF1杂交XY ES细胞系)进行(Poueymirou等(2007)Nat.Biotechnol.25:91-99;Valenzuela等(2003)Nat.Biotechnol.21:652-659)。ES细胞如先前所述加以培养(Matise等(2000),Joyner,A.L.编Gene Targeting:a practical approach,第100-132页,Oxford University Press,New York)。VGF1细胞通过使雌性C57BL/6NTac小鼠与雄性129S6/SvEvTac小鼠杂交以产生C57BL6( $X^{B6}$ )/129S6( $Y^{129}$ )小鼠来产生。参见图7。

[0485] 用750万个细胞在2mm间隙比色皿中以0.12ml的最终体积进行电穿孔(EP)。用于EP的电学条件是700V、400欧姆电阻和25 $\mu$ F电容,使用BTX ECM 630电穿孔系统(Harvard Apparatus, Holliston, MA)。每次EP的LTVEC量是0.0015mg, Cas9表达性质粒是0.005mg, 并且sgRNA表达性质粒是0.010mg。一些EP在添加100ng赋予嘌呤霉素抗性的质粒下进行以允许在不关于由LTVEC表达的新霉素抗性进行选择下选择克隆。在EP之后,将细胞涂铺于两个15cm凝胶化培养皿上,并且每日更换培养基。含有100 $\mu$ g/ml G-418硫酸盐或0.0015mg/ml嘌呤霉素的选择培养基在EP之后48小时开始,并且持续直至EP后10天。将集落挑选于PBS中,并且添加至含有0.05%胰蛋白酶的96孔培养皿中并使其解离15分钟,用培养基中和,并且用于分离DNA以进行筛选。

[0486] 等位基因修饰方法(Frendewey等(2010) *Methods Enzymol.* 476:295-307)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。

[0487] 引导序列的设计

[0488] 将在上游与下游两者围绕Lrp5或其他所靶向基因的经缺失部分内部50bp、100bp、500bp或1kb位置的约200bp的DNA输入CRISPR设计工具([crispr.mit.edu](http://crispr.mit.edu))中以检索可能的gRNA序列。接着过滤潜在gRNA序列以确保它们将仅允许切割内源性DNA而非LTVEC中的人源化插入物。

[0489] 单引导RNA克隆

[0490] 将sgRNA以双链体寡聚物(IDT)形式在融合于77bp骨架的BsmBI位点处克隆至pMB<sub>sgRNA</sub>(U6启动子)中以达成无缝RNA表达,或以经验证表达质粒形式从GeneCopoeia购买(LRP5引导物A、B、B2、E2、E和F)。通过PCR和Sanger测序来确认内部产生的质粒。

[0491] 用于基因型确认的DNA模板

[0492] 从ES细胞、源于已用靶向载体以及表达Cas9的质粒和表达若干引导RNA(gRNA)中的一者的质粒或表达不同gRNA组合的两个质粒电穿孔的ES细胞的克隆纯化DNA。选择由等位基因修饰(即等位基因丧失或等位基因获得)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诱导的缺失的克隆进行追踪常规PCR测定。

[0493] 寡核苷酸设计

[0494] 设计用于gRNA的每个组合的两个PCR测定。第一PCR是用以检测不同gRNA组合的引导RNA识别序列之间的垮塌的缺失测定。作为5'测定的第二PCR测定包括两个PCR测定。第一是针对人源化等位基因的5'人测定,并且跨越小鼠-人接合部加以设计。第二是针对内源性小鼠等位基因的5'小鼠测定,并且跨越5'靶向缺失接合部加以设计。

[0495] PCR反应和TOPO克隆

[0496] TaKaRa LA Taq DNA聚合酶(目录号RR002M)用于扩增ES细胞DNA模板。各PCR测定反应混合物与水阴性对照一起运行。测定混合物含有以下各物:0.005mL ES细胞DNA模板;1X LA PCR缓冲液II(Mg<sup>2+</sup>);0.01mM dNTP混合物;0.0075mM正向寡聚物(各自);0.0075mM反向寡聚物(各自);5000单位/mL LA Taq聚合酶;和ddH<sub>2</sub>O,补足至0.025mL。

[0497] PCR热循环程序由以下组成:94 $^{\circ}$ C,持续1分钟;继之以35个循环的94 $^{\circ}$ C,持续30秒,60 $^{\circ}$ C退火梯度,持续30秒,以及68 $^{\circ}$ C,持续1分钟(每扩增1kb);随后在72 $^{\circ}$ C下聚合10分钟。

[0498] PCR产物通过在2%琼脂糖凝胶上进行电泳来分级分离,采用Invitrogen 1kb plus DNA梯状条带(目录号10787-018)和/或Invitrogen 50bp DNA梯状条带(目录号

10416-014)。遵循来自Invitrogen的用于测序的TOP0 TA克隆试剂盒(目录号K4575-02)的说明书,将剩余PCR产物克隆至pCR4-TOP0载体中。将克隆反应物化学转化至One Shot Top10细胞中,并且涂铺在0.06mg/mL X-gal和0.025mg/mL卡那霉素琼脂板上。

#### [0499] 测序

[0500] 将白色菌落接种至含有0.025mg/mL卡那霉素的LB中,并且在37℃下在振荡下孵育过夜。各菌落代表来自测定产物群体的一个扩增子。使用QIAGEN质粒小型制备试剂盒(目录号12123)从各细菌培养物提取DNA。在包括0.002mL TOP0克隆PCR产物、1x PCR增强溶液(10x储备物)(目录号X11495-017)、0.0075mM寡聚物(M13F或M13R)和补足至0.015mL的ddH<sub>2</sub>O的测序反应混合物中测定插入物的DNA序列。

#### [0501] 测序分析

[0502] 从测序结果去除不确定序列和pCR4-TOP0载体序列,从而分离PCR插入物序列。接着将经测序片段与参照进行比对并分析变化。

#### [0503] 对垮塌克隆测序

[0504] 遵循制造商说明书(Invitrogen目录号K4575-02),将来自垮塌阳性克隆的PCR产物克隆至pCR4-TOP0载体中,接着化学转化至One Shot Top10细胞中,并且涂铺在0.060mg/mL X-gal和0.025mg/mL卡那霉素琼脂板上。使用QIAGEN质粒小型制备试剂盒(目录号12123)从细菌培养物提取DNA。接着将插入物测序结果与预测垮塌参照进行比对,并且分析插入缺失变化。预测Cas9会在离PAM 3个碱基对处进行裂解以获得由gRNA识别的序列。将经预测裂解内的序列从参照缺失,并且剩余序列用于与结果进行比对。

#### [0505] 针对单核苷酸变体(SNV)的TAQMAN<sup>®</sup>等位基因辨别测定

[0506] TAQMAN<sup>®</sup>等位基因辨别反应是0.008mL,含有基因组DNA、针对各多态性的特异性探针/引物以及TAQMAN<sup>®</sup>基因表达PCR主混合物。探针从Life Technologies(Thermo)订购,并且引物来自IDT。针对等位基因129的探针用VIC染料标记;针对等位基因B6的探针用FAM染料标记。各TAQMAN<sup>®</sup>等位基因测定一式四份在384孔板上进行,并且在Applied BioSystems ViiA 7平台上运行。SNV PCR循环程序如下:95℃,持续10分钟,继之以40个以下循环:95℃,持续15秒,60℃,持续60秒,以及60℃,持续30秒。对运行的分析以及对结果的评估使用ViiA 7软件v1.1进行。

#### [0507] FISH分析

[0508] 由Cell Line Genetics(Madison,Wisconsin)或Van Andel Institute(Grand Rapids,Michigan)通过它们的标准程序来使用荧光原位杂交(FISH)对所选ES细胞克隆进行分析。我们提供小鼠BAC和人BAC作为用于2色分析的探针。

#### [0509] 靶基因座的增强基因组垮塌和/或人源化

[0510] 为实现啮齿动物基因的全部或一部分的精确单步缺失以及任选同时用它的人同源物的全部或一部分进行的替换,我们通过电穿孔来将以下核酸分子引入啮齿动物ES细胞中:(1)LTVEC;(2)编码Cas9核酸内切酶的质粒或mRNA;和(3)一种或多种编码一个或多个CRISPR单引导RNA(gRNA)的质粒或gRNA自身。在各实验中,LTVEC是线性化的。在一些实验中,LTVEC包含编码基因产物(蛋白质或RNA)的人基因的全部或一部分,其由被设计来引导使啮齿动物基因缺失以及插入人基因的同源性重组事件的啮齿动物DNA同源臂侧接。在其

他实验中,LTVEC被设计来靶向单独基因座,诸如Ch25h基因座。在任一情况下,LTVEC都携带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)的表达的药物选择盒。

[0511] 摄取LTVEC并将它掺入它们的基因组中的ES细胞能够在组织培养皿上在含有抗生素药物的生长培养基中生长并形成集落。因为我们引入的CRISPR/Cas9编码核酸分子和gRNA编码核酸分子相比于LTVEC分子是500至1,000倍多,所以大多数含LTVEC的药物抗性集落也至少短暂含有CRISPR/Cas9组分。我们挑选药物抗性集落,并且通过等位基因修饰方法(Valenzuela等(2003) *Nat. Biotech.* 21:652-660; Frendewey等(2010) *Methods Enzymol.* 476:295-307;以引用的方式整体并入本文)来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,被称为保留测定的识别LTVEC的同源臂中的序列的实时PCR测定用于验证LTVEC向小鼠基因组中的正确靶向。这些保留测定对拷贝数的测定提供用以帮助区分保留拷贝数2的正确靶向的ES克隆与其中靶标小鼠基因座的大型Cas9诱导缺失符合LTVEC随机整合在基因组中其他地方(在所述情况下,保留测定具有拷贝数3(或更多))的克隆的进一步阐明。成对gRNA能够在靶标小鼠基因座处产生大型Cas9介导的缺失意味着如先前所述的标准LOA和GOA测定可由保留测定加强以提供进一步阐明以及验证正确靶向。因此,设计保留测定,并且与LOA和GOA测定联合使用。

[0512] 在各实验中,使用一个或两个gRNA。单一使用的gRNA引导Cas9在靶基因座的5'末端、靶基因座的中间、或靶基因座的3'末端附近的裂解(即靶向小鼠基因缺失)。当使用两个gRNA时,一个gRNA引导Cas9在靶基因座的5'末端附近的裂解,并且另一gRNA引导Cas9在靶基因座的中间或在靶基因座的3'末端附近的裂解。

[0513] Lrp5基因座

[0514] 在一组实验中,LTVEC被设计来产生小鼠Lrp5(低密度脂蛋白受体相关蛋白5)基因的编码外结构域的部分的68kb缺失,以及同时用来自人LRP5基因的同源序列的91kb片段进行的替换(参见图8)。LTVEC包含人LRP5基因的91kb片段,其由含有源于小鼠Lrp5基因座的侧接于小鼠Lrp5基因的意图缺失的68kb序列的部分的7kb和33kb的基因组DNA的同源臂侧接。在单独实验中,使Lrp5人源化LTVEC与编码Cas9的质粒和编码被设计来在小鼠Lrp5基因的靶向缺失区域内产生双链断裂的八个gRNA(A、B、B2、C、D、E2、E、F)中的一者的第二质粒组合。gRNA被设计来避免识别人LRP5基因的插入部分中的任何序列。在其他实验中,我们使LTVEC和Cas9编码质粒与编码靶向小鼠Lrp5基因的靶向缺失区域内的不同位点的两个不同gRNA的质粒组合。

[0515] 通过针对缺失内的序列以及针对药物选择盒和人基因插入物内的序列的等位基因修饰测定(Valenzuela等(2003) *Nat. Biotechnol.* 21:652-659; Frendewey等(2010) *Methods Enzymol.* 476:295-307)来针对靶向人源化筛选药物抗性ES细胞克隆。如果克隆已丧失两个内源性小鼠基因序列中的一者并获得人插入物的一个拷贝,并且也保留两个拷贝的保留序列(位于LTVEC的同源臂中),那么将它们评分为被正确靶向。用于这个筛选的两个保留测定是使用以下引物和探针进行的TAQMAN<sup>®</sup>测定:7064retU正向引物 CCTCCTGAGCTTTCCTTTGCAG (SEQ ID NO:100);7064retU反向引物 CCTAGACAACACAGACACTGTATCA (SEQ ID NO:101);7064retU TAQMAN<sup>®</sup>探针

TTCTGCCCTTGAAAAGGAGAGGC (SEQ ID NO:102); 7064retD正向引物CCTCTGAGGCCACCTGAA (SEQ ID NO:103); 7064retD反向引物CCCTGACAAGTTCTGCCTTCTAC (SEQ ID NO:104); 7064retD **TAQMAN**<sup>®</sup>探针TGCCCAAGCCTCTGCAGCTTT (SEQ ID NO:105)。

[0516] CRISPR/Cas9辅助的Lrp5基因人源化的结果概述于表13中。当将单独LTVEC引入ES细胞中时, 1.9%的筛选药物抗性克隆携带正确靶向的杂合性人源化等位基因(参见表13中的杂合性靶向列, 其包括其中非靶向等位基因完全未突变或具有小型CRISPR诱导突变诸如由NHEJ引起的小型缺失的克隆)。相比之下, 使LTVEC与由八个测试gRNA (A、B、B2、C、D、E2、E和F; 参见表1) 中的七者引导的Cas9核酸内切酶组合在2.1至7.8%的范围内的效率下产生正确靶向的单等位基因杂合性突变。对于由B2和D达成的Cas9引导的裂解, 除单等位基因靶向之外, 也在1.0-2.1%的频率下检测到双等位基因纯合性人源化。我们从未用独自LTVEC观察到双等位基因靶向, 即使对于小型简单缺失等位基因。纯合性Lrp5人源化ES细胞可通过**VELOCIMOUSE**<sup>®</sup>方法 (Poueymirou等 (2007) Nat. Biotech. 25:91-99, 以引用的方式整体并入本文) 来直接转化至准备用于表型和药物功效研究的完全ES细胞源性小鼠中。

[0517] 被设计来检测在预测裂解位点处或附近的gRNA/Cas9诱导的NHEJ突变的MOA测定证明所有测试gRNA都具有突变活性(数据未显示)。在所有测定克隆之中检测的单等位基因gRNA诱导突变或双等位基因gRNA诱导突变的比例根据基因座和位置而变化。在gRNA突变活性与LTVEC靶向性之间不存在强烈关联, 但最低靶向效率常常与具有最低突变频率的gRNA相关联。

[0518] 将识别Lrp5基因的靶向缺失区域的不同末端的两个gRNA组合使总体人源化靶向效率增加, 主要是通过使五个测试组合中的三者的纯合性靶向事件的频率增加(表13)。因为gRNA的组合有可能在Cas9裂解位点之间产生由gRNA加以程序化的大型缺失, 所以我们也观察到携带在一个Lrp5等位基因上的靶向人源化和在另一等位基因上的大型CRISPR诱导缺失的半合性ES细胞克隆(gRNA组合A+F, 表13)。此外, 对于两个gRNA组合(A+F和A+E2), 我们鉴定了具有以下独特基因型的ES细胞克隆: 在两个Lrp5等位基因上的大型CRISPR介导的缺失。

[0519] 表13. 使用个别gRNA和组合gRNA对Lrp5外结构域进行的CRISPR/Cas9辅助人源化的筛选结果。

[0520]

| gRNA    | gRNA 位点离靶向缺失的 5'/3'末端的距离(bp)   | 根据等位基因类型计的靶向效率 |            |            |           |            |
|---------|--------------------------------|----------------|------------|------------|-----------|------------|
|         |                                | 杂合性靶向(效率%)     | 半合性靶向(效率%) | 纯合性靶向(效率%) | 总体靶向(效率%) | 纯合性缺失(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')/<br>38430 (3')      | 4.1            |            |            | 4.1       |            |
| D       | 29950 (5')/<br>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')<br>F: 50 (3')       | 6.6            | 2.9        | 2.2        | 11.7      | 2.9        |
| B + E   | B: 500 (5')<br>E: 500 (3')     | 2.5            |            |            | 2.5       |            |
| B2 + E2 | B2: 1000 (5')<br>E2: 1000 (3') | 4.2            |            | 2.1        | 6.3       |            |
| A + E   | A: 50 (5')<br>E: 500 (3')      | 4.6            |            | 6.2        | 10.8      |            |
| A + E2  | A: 50 (5')<br>E2: 1000 (3')    | 2.0            |            | 4.0        | 6.0       | 4.0        |
| 无       | N/A                            | 1.9            |            |            | 1.9       |            |

[0521] 如表13中所证明,当使用靶向单一基因座的两个gRNA而非一个gRNA时(参见图9A),观察到具有双等位基因靶向的克隆的百分比显著增加,从而指示使用gRNA组合会促进双等位基因修饰。图9A显示使用LTVEC和两个引导RNA(A和B)来同时使小鼠基因缺失和用相应人形式进行替换的一般性示意图。当使用两个gRNA时在高得多的频率下观察到的独特突变等位基因类型包括纯合性垮塌的等位基因(图9B;  $\Delta/\Delta$ )、以纯合方式靶向的等位基因(图9C; Hum/Hum)、以半合方式靶向的等位基因(图9D; (Hum/ $\Delta$ ))以及其他复合的以杂合方式靶向的等位基因(例如一个等位基因具有LTVEC靶向的人源化,并且另一等位基因具有CRISPR诱导突变诸如小型缺失)(图9E)。

[0522] 进行若干PCR测定以支持和确认基于MOA测定的基因型。引物可见于表1中。Lrp5 LTVEC具有足够短(6.9kb)以通过PCR来证明靶向性的5'同源臂,所述PCR测定人插入物与邻近小鼠基因组序列之间的实体连接。我们用来自评分为杂合性、半合性或纯合性的克隆的DNA,而非用来自亲本ES细胞系或来自评分为具有双等位基因大型缺失的克隆的DNA观察到预期7.5kb PCR产物(图10A),因此确认通过MOA(即LOA和GOA)筛选得到的靶向识别,并且支持推断的双等位基因大型缺失。考查在缺失和插入接合部处的序列的5'-De1-J PCR测定(图10B)在来自亲本ES细胞系以及来自大多数杂合性人源化克隆的DNA的情况下产生330bp产物(数据未显示)。对于杂合性克隆AW-C3,5'-De1-J测定产生小于预期的产物(图10B),从而表明gRNA A/Cas9裂解在非靶向等位基因上诱导小型缺失突变,其也通过针对gRNA A裂

解的MOA测定检测到(数据未显示)。如所预期,5'-De1-J测定对于具有半合性、纯合性和双等位基因缺失等位基因的克隆是阴性的。考查在人DNA插入物的5'末端与邻近小鼠侧接序列之间的接合部处的序列的5'-Ins-J PCR(图10B)在杂合性、半合性和纯合性克隆的情况下产生478bp产物,因为这些克隆具有至少一个所靶向人源化等位基因。对于具有双等位基因大型缺失的克隆,5'-Ins-J PCR测定不产生产物(图10B)。为确认半合性和双等位基因缺失克隆中的大型缺失,我们用识别双重gRNA靶标位点的外部的序列的引物进行PCR。测定A gRNA位点与F gRNA位点之间的缺失的De1 (A+F) PCR在来自克隆AW-A8和BO-F10的DNA的情况下产生约360bp的单一产物(图10B),从而确认至少一个Lrp5等位基因具有大型缺失。同样,测定A gRNA位点与E2gRNA位点之间的大型缺失的De1 (A+E2) PCR在来自克隆BA-A7的DNA的情况下产生约250bp的单一产物。缺失PCR连同接合部测定、LOA测定和GOA测定一起支持双等位基因大型缺失基因型。图10A和图10B中所示的测定结果是我们除荧光原位杂交(FISH;图11A-图11C)之外也进行的用以确认表13中概述的双等位基因基因型的类似测定的代表性实例。

[0523] 荧光原位杂交(FISH)用于确认Lrp5基因的纯合性靶向人源化。将来自其中使Lrp5人源化LTVEC与Cas9和两个gRNA(A加F或A加E2)组合的靶向实验的由定量和常规PCR测定评分为被纯合性靶向的ES细胞克隆送给商业细胞学服务机构以进行FISH和核型分析。将携带小鼠Lrp5基因的细菌人工染色体(BAC)用红色荧光标记加以标记并用作用以鉴定内源性Lrp5基因座的探针,并且将携带人LRP5基因的BAC用绿色荧光标记加以标记并用作用以鉴定用人插入物靶向的染色单体的探针。使经标记BAC探针与来自所靶向克隆的中期涂片杂交,并且通过荧光显微术来可视化。涂片上的染色体通过用DAPI(4',6-二脒基-2-苯基吡啶)染色来可视化,并且各克隆的单独核型通过吉姆萨染色(Giemsa staining)来测定。克隆AW-D9的典型结果显示于图11A中,发现所述克隆具有正常40XY核型(未显示)。图11A中的复合照片显示红色小鼠BAC探针信号与绿色人BAC探针信号两者共同定位于作为Lrp5基因的已知位置的小鼠染色体19的两个拷贝上的细胞学条带B。图11C中的复合照片显示对于另一克隆(BA-D5)的相同纯合性靶向。这些结果确认在克隆AW-D9和BA-D5中,人源化LTVEC中的人LRP5基因的91kb片段被正确插入在两个染色体19同源物上的预定小鼠Lrp5基因座处。相比之下,图11B中的复合照片显示红色小鼠BAC探针信号与绿色人BAC探针信号两者共同定位于小鼠染色体19的单一拷贝上的细胞学条带B(实线箭头),而仅有红色小鼠BAC探针信号定位于小鼠染色体19的另一拷贝上的细胞学条带B。这些结果确认人源化LTVEC中的人LRP5基因的91kb片段被正确插入在染色体19的仅一个拷贝上的预定小鼠Lrp5基因座处(杂合性靶向)。它们也指示(连同未显示的其他对照一起)人BAC探针不与小鼠Lrp5基因座交叉杂交,而是仅识别人LRP5插入物。

[0524] 在某些克隆中存在通过明显非同源性末端接合修复来在两个等位基因处形成的相同CRISPR诱导插入缺失突变表明在F1H4杂交细胞(其包含50%129SvS6品系和50%C57BL/6N品系)中发生基因转换事件。为获得对在使用两个gRNA时增强双等位基因靶向的背后机理的见地,在用LTVEC以及A加F或A加E2 gRNA组合进行靶向之后筛选了具有靶向纯合性人源化或纯合性CRISPR诱导大型缺失的七个克隆。

[0525] 图12显示被设计来考查由两个引导RNA介导的基因转换事件的测定的实例。具体来说,通过分析F1H4杂交ES细胞(其包含50%129 SvS6品系和50%C57BL/6N品系)中的杂合

性丧失 (LOH) 来考查基因转换的可能性。基因转换可通过129SvS6 (129) 与C57BL/6N (B6) 之间的已知多态性的杂合性丧失来证明,因此PCR测定被设计来区分这两种等位基因类型。结构变体 (SV) 多态性通过被设计来检测129等位基因与B6等位基因之间的差异的常规PCR来测定。尽管仅一个以下使用的SV测定显示于图12中,但该概念对于各个SV测定都是相同的。基于B6小鼠品系与129小鼠品系之间的结构变化 (SV) 设计引物,并且显示于表1中。约束引物设计条件以鉴定约25bp SV以及产生约300bp PCR产物;选择这些条件以使任何变化都可通过凝胶电泳见到。

[0526] 在对克隆运行PCR之前,相对于来自B6品系、129品系以及来自F1H4ES细胞系的野生型ES细胞DNA对测定进行验证和优化。选择产生为B6或129等位基因所特有的可区分PCR条带并且在使用F1H4 DNA的情况下在产生这两个相同可区分条带方面一致的引物组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、190061、190068、190030、190033、190013—一对由等位基因修饰 (MOA) 测定和常规PCR将基因分型为“纯合性靶向”或“纯合性垮塌”的Lrp5人源化克隆进行使用。SV PCR测定沿染色体19,从Lrp5基因座至染色体的端粒端加以隔开,离Lrp5基因座在约13.7至约56.2Mb的范围内。SV测定在染色体19上离Lrp5基因座的近似距离 (以Mb计) 如下:对于测定190045是13.7,对于测定190061是19.0,对于测定190068是35.0,对于测定190030是37.4,对于测定190033是48.3,并且对于测定190013是56.2。仅有测定190033显示于图12中 (显示为SV 48.3),但用于测定190045、190061、190068、190030、190033和190013的引物显示于表1中。

[0527] 对来自这些克隆的DNA以及对F1H4对照DNA、129对照DNA和B6对照DNA运行PCR。PCR产物通过在6%聚丙烯酰胺凝胶上进行电泳来分级分离,所述凝胶随后用GelRed染色。产生两个条带的克隆符合F1H4对照,此根据先前优化显示顶部条带为129等位基因所特有,并且底部条带为B6等位基因所特有。仅产生一个条带的克隆显示只是B6条带或只是129条带。克隆AW-A7、AW-F10、BA-D5、BA-F2、BC-H9和BR-B4对于所有六个测定都仅显示B6条带,而克隆B0-A8对于所有六个测定都仅显示129条带。如先前所提及,这些克隆由MOA和/或PCR将基因分型为纯合性靶向或纯合性垮塌,并且涉及各种gRNA组合 (A加F、A加E2、B2和D)。只是存在单一等位基因条带表明基因转换事件正在发生—如果不存在转换,那么两个条带均将仍然如同在F1H4对照的情况下而存在。

[0528] 此外,129等位基因与B6等位基因之间的单核苷酸变体 (SNV) 通过TAQMAN<sup>®</sup>等位基因辨别测定来测定。SNV测定在图12中的染色体19图谱上的近似位置由箭头连同下方给出的它们离Lrp5基因座的距离 (以Mb计) 一起加以显示。离Lrp5基因座的距离 (以Mb计) 如下:在Lrp5的着丝粒方向0.32 (C2),在Lrp5的端粒方向1.2 (T3),在Lrp5的端粒方向11.1 (T6),在Lrp5的端粒方向13.2 (T7),在Lrp5的端粒方向17.5 (T8),在Lrp5的端粒方向25.8 (T9),在Lrp5的端粒方向33.0 (T10),在Lrp5的端粒方向38.3 (T11),在Lrp5的端粒方向49.6 (T13),以及在Lrp5的端粒方向57.2 (T14)。129特异性探针和B6特异性探针以及引物对显示于表1中。

[0529] 表14显示根据SV等位基因与SNV等位基因的LOH,在染色体19的长臂上,在Lrp5靶基因座的端粒方向上展现明显基因转换事件的ES细胞克隆的七个实例。ES细胞克隆源于如所指示的使Lrp5人源化LTVEC与一个或两个gRNA组合的独立靶向实验。gRNA识别序列的位置显示在图12中Lrp5基因图示的上方 (指向左侧的粗箭号)。基因分型测定指示七个克隆中

的六者具有Lrp5基因的以纯合方式靶向的人源化,而一者具有纯合性垮塌(在gRNA位点之间的大型缺失)。在七个克隆中的六者中,129等位基因丧失,从而仅保留B6等位基因。在另一克隆中,B6等位基因丧失,从而仅保留129等位基因。所有克隆就在Lrp5基因座的着丝粒侧上测定的等位基因来说都保持杂合(即在C2 SNV测定的情况下,所有克隆都是杂合性B6/129)。在七个克隆中观察到的LOH指示当使LTVEC与一个gRNA或更经常与两个gRNA组合时获得纯合性经遗传修饰的等位基因所采用的一种机理是在一个等位基因上进行第一靶向遗传修饰,继之以同源性引导的重组基因转换事件,其从一个染色体将所述靶向遗传修饰拷贝至它的同源物。

[0530] 表14. 杂合性丧失测定结果。

| 克隆     | gRNA | Lrp5等位基因类型 | 杂合性丧失测定 (SV和SNV) |
|--------|------|------------|------------------|
| AW-A7  | A+F  | 纯合性靶向      | 仅检测到B6等位基因       |
| AW-F10 | A+F  | 纯合性垮塌      | 仅检测到B6等位基因       |
| BO-A8  | A+F  | 纯合性靶向      | 仅检测到129等位基因      |
| BA-D5  | A+E2 | 纯合性靶向      | 仅检测到B6等位基因       |
| BA-F2  | A+E2 | 纯合性靶向      | 仅检测到B6等位基因       |
| BC-H9  | B2   | 纯合性靶向      | 仅检测到B6等位基因       |
| BR-B4  | D    | 纯合性靶向      | 仅检测到B6等位基因       |

[0532] C5 (Hc) 基因座

[0533] 在另一组实验中,LTVEC被设计来产生小鼠补体组分5 (C5或Hc (溶血补体)) 基因的76kb缺失,以及同时用同源性人C5基因的97kb片段进行的替换。靶基因座包含C5 (Hc) 基因的外显子2至终止密码子。LTVEC包含人C5基因的97kb片段,其由含有源于小鼠C5 (Hc) 基因座的侧接于小鼠C5 (Hc) 基因的意图缺失的76kb序列的部分的35kb和31kb的基因组DNA的同源臂侧接。在单独实验中,使C5 (Hc) 人源化LTVEC与编码Cas9的质粒和编码被设计来在小鼠C5 (Hc) 基因的靶向缺失区域内产生双链断裂的六个gRNA (A、B、C、D、E和E2;参见表1) 中的一者的第二质粒组合。gRNA被设计来避免识别人C5基因的插入部分中的任何序列。在其他实验中,我们使LTVEC和Cas9编码质粒与编码靶向小鼠C5 (Hc) 基因的靶向缺失区域内的不同位点的两个不同gRNA的质粒组合。在一些实验中,靶向Ch25h基因座的对照LTVEC替代C5 (Hc) 人源化LTVEC加以使用。被设计来使Ch25h的整个编码序列(约1kb) 缺失以及将嘌呤霉素和新霉素选择盒插入Ch25h基因座中的对照LTVEC用作用以选择未被靶向以达成在C5 (Hc) 基因座处的同源性重组的药物抗性克隆的手段。

[0534] CRISPR/Cas9辅助的C5 (Hc) 基因人源化的结果显示于表15中,并且类似于对于CRISPR/Cas9辅助的Lrp5基因人源化获得的结果。相比于对于Lrp5,用单独LTVEC对于C5 (Hc) 人源化达成的靶向效率更高(6.1%),但添加Cas9和gRNA使六个测试gRNA中的四者的靶向效率增强。如同Lrp5一样,将gRNA组合(即使用两个gRNA)用于C5 (Hc) 人源化使总体靶向效率进一步增加,主要是通过使半合性和纯合性靶向事件的频率增加。我们也发现在两个等位基因上均具有大型CRISPR诱导缺失的ES细胞克隆(在1.8%至3.6%的频率下观察到)。此外,当靶向Ch25h基因座的LTVEC与两个C5 (Hc) gRNA组合使用时,在1.2%至6%的频率下观察到具有在两个引导RNA识别序列之间垮塌的纯合性等位基因的克隆,从而指示在靶基因座处,垮塌事件独立于同源性重组事件而发生。如同Lrp5一样,保留测定用于确认正

确靶向的克隆。用于这个筛选的两个保留测定是使用以下引物和探针进行的TAQMAN<sup>®</sup>测定:7140retU正向引物CCCAGCATCTGACGACACC (SEQ ID NO:106);7140retU反向引物GACCACTGTGGGCATCTGTAG (SEQ ID NO:107);7140retU TAQMAN<sup>®</sup>探针CCGAGTCTGCTGTTACTGTTAGCATCA (SEQ ID NO:108);7140retD正向引物CCCAGACCTTCTGAGCATG (SEQ ID NO:109);7140retD反向引物TGCAGGCTGAGTCAGGATTTG (SEQ ID NO:110);7140retD TAQMAN<sup>®</sup>探针TAGTCACGTTTTGTGACACCCAGA (SEQ ID NO:111)。

[0535] 表15.使用个别gRNA和组合gRNA对C5 (Hc) 基因进行的CRISPR/Cas9辅助人源化的筛选结果。

| gRNA   | gRNA 位点离<br>靶向缺失的<br>5'/3'末端的距<br>离(bp) | LTVEC | 根据等位基因类型计的靶向效率     |                    |                    |                   |                    |
|--------|-----------------------------------------|-------|--------------------|--------------------|--------------------|-------------------|--------------------|
|        |                                         |       | 杂合性<br>靶向(效<br>率%) | 半合性<br>靶向(效<br>率%) | 纯合性<br>靶向(效<br>率%) | 总体靶<br>向(效<br>率%) | 纯合性<br>缺失(效<br>率%) |
| A      | 100 (5')                                | C5    | 16.6               |                    |                    | 16.6              |                    |
| B      | 500 (5')                                | C5    | 14.5               |                    |                    | 14.5              |                    |
| C      | 38200 (5')/<br>37500 (3')               | C5    | 11.4               |                    |                    | 11.4              |                    |
| D      | 43500 (5')/<br>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:<br>37500 (3')            | C5    | 19.6               | 7.1                | 0.6                | 27.3              | 0.6                |
| A + C  | A: 100 (5') C:<br>37500 (3')            | Ch25h | N/A                | N/A                | N/A                | N/A               | 6.0                |
| A + E2 | A: 100 (5')<br>E2: 100 (3')             | C5    | 19.0               | 3.6                | 1.2                | 23.8              | 3.0                |
| A + E2 | A: 100 (5')<br>E2: 100 (3')             | Ch25h | N/A                | N/A                | N/A                | N/A               | 1.2                |
| 无      | N/A                                     | C5    | 6.1                |                    |                    | 6.1               |                    |

[0537] 荧光原位杂交 (FISH) 用于确认C5 (Hc) 基因的纯合性靶向人源化。将来自其中使C5 (Hc) 人源化LTVEC与Cas9和两个gRNA组合的靶向实验的由定量和常规PCR测定评分为被纯合性靶向的ES细胞克隆送给商业细胞学服务机构以进行FISH和核型分析。将携带小鼠C5 (Hc) 基因的细菌人工染色体 (BAC) 用红色荧光标记加以标记并用作用以鉴定内源性基因座的探针,并且将携带人C5基因的BAC用绿色荧光标记加以标记并用作用以鉴定用人插入物靶向的染色单体的探针。使经标记BAC探针与来自所靶向克隆的中期涂片杂交,并且通过荧光显微术来可视化。涂片上的染色体通过用DAPI (4',6-二脒基-2-苯基吡啶) 染色来可视化,并且各克隆的单独核型通过吉姆萨染色来测定。克隆0-E的典型结果显示于图13B中。图13B中的复合照片显示红色小鼠BAC探针信号与绿色人BAC探针信号两者共同定位于作为C5 (Hc) 基因的已知位置的小鼠染色体2的两个拷贝上的C5 (Hc) 基因座。这些结果确认在克隆0-E3中,人源化LTVEC中的人C5基因的97kb片段被正确插入在两个染色体2同源物上的预定小鼠C5 (Hc) 基因座处。相比之下,图13A中的复合照片显示红色小鼠BAC探针信号与绿色人

BAC探针信号两者共同定位在小鼠染色体2的单一拷贝上(实线箭头),而仅有红色小鼠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。这些结果确认在克隆Q-E9中,人源化LTVEC中的人C5基因的97kb片段被正确插入在染色体2的仅一个拷贝上的预定小鼠C5(Hc)基因座处(杂合性靶向)。

[0538] 接着测定克隆以考查由两个引导RNA介导的基因转换事件。具体来说,通过分析F1H4杂交ES细胞(其包含50%129 SvS6品系和50%C57BL/6N品系)中的杂合性丧失(LOH)来考查基因转换的可能性。基因转换可通过129SvS6(129)与C57BL/6N(B6)之间的已知多态性的杂合性丧失来证明,因此PCR测定被设计来区分这两种等位基因类型。结构变体(SV)多态性通过被设计来检测129等位基因与B6等位基因之间的差异的常规PCR来测定。基于B6小鼠品系与129小鼠品系之间的结构变化(SV)设计引物,并且显示于表1中。约束引物设计条件以鉴定约25bp SV以及产生约300bp PCR产物;选择这些条件以使任何变化都将可通过凝胶电泳可视化。

[0539] 在对克隆运行PCR之前,相对于来自B6品系、129品系以及来自F1H4 ES细胞系的野生型ES细胞DNA对测定进行验证和优化。选择产生为B6或129等位基因所特有的可区分PCR条带并且在使用F1H4 DNA的情况下在产生这两个相同可区分条带方面一致的引物组用于对克隆进行测试。选择五个引物组-ID SV 6.1、SV 6.3、SV 7.8、SV 16和SV 25.5一对来自靶向实验的克隆进行使用。四个SV PCR测定沿染色体,从C5基因座至染色体的端粒端加以隔开,离C5基因座在约6.3至约25.5Mb的范围内。最终SV PCR测定在C5基因座的着丝粒方向约6.1Mb处。SV测定离C5基因座的近似距离(以Mb计)如下:对于测定SV 6.1是6.1(在着丝粒方向),对于测定SV 6.3是6.3(在端粒方向),对于测定SV 7.8是7.8(在端粒方向),对于测定SV 16.0是16.0,并且对于测定SV25.5是25.5(参见图14)。

[0540] 所有21个克隆就在C4基因座的着丝粒侧上测定的等位基因来说都保持杂合(即所有克隆都是杂合性B6/129)。根据LOH,21个测试克隆中的两者在C5靶基因座的端粒方向上展现明显基因转换事件(参见表16)。基因分型测定指示克隆之一具有C5基因的以纯合方式靶向的人源化,并且另一克隆具有纯合性垮塌。在两个克隆中观察到的LOH指示当使LTVEC与一个gRNA或更经常与两个gRNA组合时获得纯合性经遗传修饰的等位基因所采用的一种机理是在一个等位基因上进行第一靶向遗传修饰,继之以同源性引导的重组基因转换事件,其从一个染色体将所述靶向遗传修饰拷贝至它的同源物。

[0541] 表16. 杂合性丧失测定结果。

[0542]

| 克隆   | gRNA | C5等位基因类型 | 基因转换测定      |
|------|------|----------|-------------|
| R-E2 | A+E2 | 纯合性靶向    | 仅检测到129等位基因 |
| R-E8 | A+E2 | 纯合性垮塌    | 仅检测到129等位基因 |

[0543] Ror1基因座

[0544] 在另一组实验中,LTVEC被设计来产生小鼠Ror1(酪氨酸蛋白质激酶跨膜受体ROR1)基因的110kb缺失,以及同时用同源性人ROR1基因的134kb片段进行的替换。LTVEC包含人ROR1基因的134kb片段,其由含有源于小鼠Ror1基因座的侧接于小鼠Ror1基因的意图缺失的110kb序列的部分的41.8kb和96.4kb的基因组DNA的同源臂侧接。在单独实验中,使Ror1人源化LTVEC与编码Cas9的质粒和编码被设计来在小鼠Ror1基因的靶向缺失区域内产生双链断裂的六个gRNA(A、B、C、D、E和F;参见表1)中的一者的第二质粒组合。gRNA被设计来

避免识别人ROR1基因的插入部分中的任何序列。在其他实验中,我们使LTVEC和Cas9编码质粒与编码靶向Ror1基因内的被靶向以达成缺失的不同位点的两个不同gRNA的质粒组合。

[0545] CRISPR/Cas9辅助的Ror1基因人源化的结果显示于表17中,并且类似于对于CRISPR/Cas9辅助的Lrp5和C5 (Hc) 基因人源化获得的结果。用单独LTVEC达成的靶向效率是0.3%,并且添加Cas9和gRNA使六个测试gRNA中的两者的靶向效率略微增加。将A gRNA和F gRNA组合通过使杂合性靶向事件与半合性靶向事件两者的频率增加来使总体Ror1靶向效率增加至6.3%。我们也发现在两个等位基因上均具有大型CRISPR诱导缺失的ES细胞克隆(在1.6%的频率下观察到)。未观察到纯合性靶向克隆。在额外实验中,也使gRNA A和D组合,但仍然未观察到纯合性靶向。

[0546] 表17. 使用个别gRNA和组合gRNA对Ror1基因进行的CRISPR/Cas9辅助人源化的筛选结果。

| gRNA  | gRNA 位点离靶向缺失的5'/3'末端的距离(bp)  | 根据等位基因类型计的靶向效率 |            |            |           |            |
|-------|------------------------------|----------------|------------|------------|-----------|------------|
|       |                              | 杂合性靶向(效率%)     | 半合性靶向(效率%) | 纯合性靶向(效率%) | 总体靶向(效率%) | 纯合性缺失(效率%) |
| A     | 200 (5')                     | 0.7            |            |            | 0.7       |            |
| B     | 1000 (5')                    | 0.0            |            |            | 0.0       |            |
| D     | 54300 (5')/<br>55500 (3')    | 0.7            |            |            | 0.7       |            |
| C     | 54500 (5')/<br>55300 (3')    | 0.0            |            |            | 0.0       |            |
| E     | 1000 (3')                    | 0.0            |            |            | 0.0       |            |
| F     | 200 (3')                     | 0.3            |            |            | 0.3       |            |
| A + F | A: 200 (5')<br>F: 200 (3')   | 4.2            | 2.1        |            | 6.3       | 1.6        |
| A + D | A: 200 (5')<br>D: 55500 (3') | 1.0            |            |            | 1.0       |            |
| 无     | N/A                          | 0.3            |            |            | 0.3       |            |

[0548] Trpa1基因座

[0549] 在另一组实验中,LTVEC被设计来产生小鼠Trpa1(瞬时受体电位阳离子通道,亚家族A,成员1)基因的45.3kb缺失,以及同时用同源性人TRPA1基因的54.5kb片段进行的替换。LTVEC包含人TRPA1基因的54.5kb片段,其由含有源于小鼠Trpa1基因座的侧接于小鼠Trpa1基因的意图缺失的45.3kb序列的部分的41.0kb和58.0kb的基因组DNA的同源臂侧接。在单独实验中,使Trpa1人源化LTVEC与编码Cas9的质粒和编码被设计来在小鼠Trpa1基因的靶向缺失区域内产生双链断裂的八个gRNA(A、A2、B、C、D、E2、E和F;参见表1)中的一者的第二质粒组合。gRNA被设计来避免识别人TRPA1基因的插入部分中的任何序列。在其他实验中,我们使LTVEC和Cas9编码质粒与编码靶向被靶向以达成缺失的Trpa1基因内的不同位点的两个不同gRNA的质粒组合。

[0550] CRISPR/Cas9辅助的Trpa1基因人源化的结果显示于表18中,并且类似于对于CRISPR/Cas9辅助的Lrp5和C5 (Hc) 基因人源化获得的结果。用单独LTVEC达成的靶向效率是0.3%,并且添加Cas9和gRNA使八个测试gRNA中的六者的靶向效率增加。将B gRNA和F gRNA组合通过使杂合性、半合性和纯合性靶向事件的频率增加来使总体Trpa1靶向效率增加至

3.4%。我们也发现在两个等位基因上均具有大型CRISPR诱导缺失的ES细胞克隆(在0.3%的频率下观察到)。

[0551] 表18. 使用个别gRNA和组合gRNA对Trpa1基因进行的CRISPR/Cas9辅助人源化的筛选结果。

| gRNA | gRNA 位点离靶向缺失的 5'/3'末端的距离(bp) | 根据等位基因类型计的靶向效率 |            |            |           |            |
|------|------------------------------|----------------|------------|------------|-----------|------------|
|      |                              | 杂合性靶向(效率%)     | 半合性靶向(效率%) | 纯合性靶向(效率%) | 总体靶向(效率%) | 纯合性缺失(效率%) |
| A    | 100 (5')                     | 1.0            |            |            | 1.0       |            |
| A2   | 500 (5')                     | 2.1            |            |            | 2.1       |            |
| B    | 1000 (5')                    | 1.4            |            |            | 1.4       |            |
| C    | 25600 (5')/<br>19740 (3')    | 1.0            |            |            | 1.0       |            |
| D    | 26970 (5')/<br>18370 (3')    | 2.1            |            |            | 2.1       |            |
| E2   | 1000 (3')                    | 0.0            |            |            | 0.0       |            |

| gRNA  | gRNA 位点离靶向缺失的 5'/3'末端的距离(bp) | 根据等位基因类型计的靶向效率 |            |            |           |            |
|-------|------------------------------|----------------|------------|------------|-----------|------------|
|       |                              | 杂合性靶向(效率%)     | 半合性靶向(效率%) | 纯合性靶向(效率%) | 总体靶向(效率%) | 纯合性缺失(效率%) |
| E     | 500 (3')                     | 0.0            |            |            | 0.0       |            |
| F     | 100 (3')                     | 0.7            |            |            | 0.7       |            |
| B + F | B: 1000 (5')<br>F: 100 (3')  | 2.8            | 0.3        | 0.3        | 3.4       | 0.3        |
| 无     | N/A                          | 0.3            |            |            | 0.3       |            |

[0554] 如这些实例所说明,相较于单一gRNA,在广泛分隔的位点处使用双重引导RNA使对杂合性人源化的增强作用改进。此外,相较于单一gRNA,使用双重引导RNA促进双等位基因事件。与用一个gRNA进行的靶向相比,用两个gRNA进行的靶向导致产生以纯合方式靶向的细胞(Hum/Hum),其中两个等位基因均具有靶向人源化;以纯合方式缺失的细胞( $\Delta/\Delta$ ),其中两个等位基因都不被人源化LTVEC靶向,但两者均具有大型缺失;以及以半合方式靶向的细胞(Hum/ $\Delta$ ),其中一个等位基因具有靶向人源化,并且另一等位基因具有大型双重gRNA/Cas9诱导缺失。首先,我们发现在两个靶标等位基因处均具有精确和相同的极其大型人源化的正确靶向的克隆(例如就靶向基因修饰来说是纯合的细胞)。尽管当我们使用一个gRNA来实现Lrp5人源化时也观察到以纯合方式靶向的克隆,但相比于当我们采用两个gRNA时,它们在低得多的频率下出现(参见表13)。同样,当使用一个gRNA来实现C5(Hc)人源化或Trpa1人源化时,我们未观察到纯合性靶向,但当与靶向载体一起使用两个gRNA时,我们确实观察到纯合性靶向(参见表15和表18)。类似地,对于Lrp5靶向、C5(Hc)靶向、Ror1靶向和Trpa1靶向,我们发现就基因修饰来说是半合的正确靶向的克隆(即它们具有在一个等位基因上的精确靶向的人源化,以及在另一等位基因上的极其大型(有时是基因消除性)缺失)。当使用一个gRNA来实现Lrp5、C5(Hc)、Ror1或Trpa1人源化时,所述修饰完全不发生(分别参见表13、表15、表17和表18)。

[0555] 其次,我们发现在两个所靶向等位基因上均具有由受两个gRNA引导的Cas9裂解事

件诱导的相同极大大型缺失(>45kb)的克隆(即细胞就在靶基因座处的大型(有时是基因消除性)缺失来说是纯合的)。这些类型的突变不需要针对同一基因的靶向载体。举例来说,如表15中所示,我们已通过使Cas9和两个gRNA与针对与由gRNA靶向的基因无关的不同基因的靶向载体组合来获得具有纯合性CRISPR诱导缺失的ES细胞。因此,由两个gRNA引导的Cas9核酸酶可在不添加靶向载体下在细胞中诱导大型缺失。在所述情况下,由表达药物抗性基因的载体提供的短暂或稳定药物选择可有助于通过使已摄取DNA的ES细胞富集来分离稀有纯合性缺失克隆。

[0556] 实施例6. 分析由组合gRNA诱导的大型缺失。

[0557] 由组合gRNA诱导的大型缺失的等位基因结构

[0558] 对包含由受两个gRNA引导的Cas9裂解事件诱导的大型缺失的克隆进行额外序列分析(参见表19)。这些大型缺失似乎独立于在同一基因座处的LTVEC引导的同源性重组事件,因为当我们使gRNA与Lrp5 LTVEC或靶向相距接近30Mb的Ch25h基因的LTVEC组合时,我们在Lrp5基因座处在近似相同频率下获得了大型缺失(数据未显示)。为表征大型缺失,我们对来自6个人源化操作的40个克隆(15个是半合性的,并且25个具有双等位基因大型缺失)进行缺失跨越性PCR,并且对PCR产物的个别克隆测序。序列确认在38kb至109kb的范围内的的大型缺失。三个ES细胞克隆(Lrp5克隆AW-A8和BP-D3以及Adamts5克隆X-B11)在预测的Cas9裂解位点之间具有完全修复的精确缺失(68.2kb),而一个克隆(Hc克隆P-B12)除38.1kb缺失之外也具有单一碱基对插入。27个ES细胞克隆具有延伸超出Cas9裂解位点的缺失,这符合通过非同源性末端接合(NHEJ)进行的不精确修复。其余9个ES细胞克隆具有组合了明显NHEJ诱导的缺失和插入的突变(例如Lrp5克隆BP-F6和Hc克隆O-E4),其中五者具有我们可作图至它们的来源基因组基因座的大于200bp的插入(数据未显示)。Lrp5克隆B0-E9中的210bp插入相对于在着丝粒方向的位于gRNA F靶标位点的外部约2,600bp的相同序列(染色体19+, 3589138-3589347)处于反向定向。这个序列存在于Lrp5LTVEC的长3'同源臂中。Lrp5克隆BP-F6和BP-G7源于其中我们使Lrp5gRNA A和F与Cas9和靶向在端粒方向离开Lrp5 30Mb的Ch25h基因的LTVEC组合的实验。克隆BP-F6具有似乎源于Ch25h LTVEC的一端的266bp插入,因为它由与载体骨架的一部分相同的103bp片段连接于与在Ch25h附近的序列相同,并且也存在于LTVEC的长臂中的163bp片段(染色体19+, 34478136-34478298)组成;这个片段相对于内源性染色体序列以反向定向插入在缺失处。Hc克隆O-E4具有254bp插入,其相对于见于经缺失序列内的离开gRNA A识别序列约3.1kb的相同序列处于反向。Hc克隆S-D5中的1,304bp插入由以下两个片段组成:1,238bp片段,其与见于经缺失序列内的离开预测的gRNA E2引导的Cas9裂解位点约1.4kb的相同序列处于相同定向;和第二66bp片段,其是处于在gRNA E2切割位点的外部25bp的相同序列的反向定向的重复序列。

[0559] 表19. 由组合gRNA诱导的大型缺失的等位基因结构。

[0560]

| 基因             | ES 细胞克隆 | 基因型 <sup>1</sup> | gRNA   | 靶向缺失内的位置(bp)     | 缺失大小(kb) | 额外经缺失序列(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          |        | 1      |
|                | N-E1    | Hum/Δ            |        |                  |          | 10          |        | 13     |
|                | N-E9    | Hum/Δ            |        |                  |          | 20          |        | 16     |
|                | O-C5    | Hum/Δ            |        |                  |          | 31          |        | 21     |
|                | 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     |
|                | S-A11   | Δ/Δ              |        |                  |          | 122         |        | 11     |
|                |         |                  |        |                  |          | 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      |
| <i>Dpp4</i>    | S-F1    | Δ/Δ              |        | 5′-50/38,100-3′  | 40.7     | 18          | 877    | 20     |
|                | S-G6    | Δ/Δ              |        |                  |          | 35          | 3      | 17     |
| <i>Adamts5</i> | X-B11   | Δ/Δ              |        | 5′-1000/100-3′   | 37.4     |             |        | 11     |

[0561] <sup>1</sup>Hum/+, 两个天然等位基因中的一者具有靶向人源化, 从而产生杂合性基因型; Hum/ $\Delta$ , 双等位基因修饰, 其中一个等位基因具有靶向人源化, 并且另一等位基因具有大型 Cas9-gRNA 诱导缺失, 从而产生半合性基因型; Hum/Hum, 双等位基因修饰, 其中两个等位基因均具有靶向人源化, 从而产生纯合性基因型;  $\Delta/\Delta$ , 双等位基因修饰, 其中两个等位基因

均具有大型Cas9-gRNA诱导缺失。

[0562] 在纯合性等位基因处的基因转换的证据

[0563] 具有双等位基因大型缺失的25个ES细胞克隆中的24者仅具有单一独特序列(表19),从而指示它们是纯合性等位基因。对于Hc克隆S-A11,我们发现在12个PCR克隆中的11者中具有相同序列。具有不同序列的单一克隆可能表明两个不同缺失等位基因,但我们 also 发现两个Hc半合性克隆N-D11和O-F12具有相同结果。多个克隆中的不同纯合性缺失等位基因表明它们可能已通过基因转换机理而出现,其中一个染色体上的缺失充当用于对同源染色体上的Cas9裂解进行同源性重组修复的模板。我们利用VGF1ES细胞系的129S6SvEvTac(129)和C57BL/6NTac(B6)F1杂交组合物(Poueymirou等(2007)Nat.Biotechnol.25:91-99;Valenzuela等(2003)Nat.Biotechnol.21:652-659)来将基因转换测定为品系之间在染色体19上的Lrp5基因座(对于以下使用的五个SV测定和十个SNV测定,参见图12)和染色体2上的Hc基因座(未显示)周围的结构变体(SV)和单核苷酸变体(SNV)的杂合性丧失(Lefebvre等(2001)Nat.Genet.27:257-258)。为确认任何杂合性丧失都不是整个染色体丧失的结果,我们在129品系与B6品系之间相同的位点处进行染色体拷贝数(CCN)测定。对于Lrp5人源化或经缺失等位基因,我们测定在端粒方向离开Lrp5 1.2Mb加以定位直至染色体19的长臂的末端的多个SV和SNV(图12)。由于Lrp5的位置靠近着丝粒,所以我们未发现SV,并且仅在基因的着丝粒侧上发现一个SNV。对于Hc,我们能够在染色体2上的基因的任一侧上测定多个SV和SNV(未显示)。六个Lrp5克隆的结果显示于图15A-图15E和图16A-图16C中。

[0564] 图15A-图15E显示五个SV测定的结果,所述测定的位置在离开Lrp513.7Mb至56.7Mb(在长臂的端粒端附近)的范围内。五个SV测定对于129、B6和VGF1对照中的129(较大)和B6(较小)等位基因产生两个不同大小的产物。SV测定在染色体19图谱上的近似位置显示于图12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)。测定编号代表在Lrp5的端粒方向的Mb数目。用于这些测定的引物显示于表1中,并且结果显示于图15A-图15E中。两个克隆BC-H9(Lrp5<sup>Hum/Hum</sup>,gRNA B2)和BR-B4(Lrp5<sup>Hum/Hum</sup>,gRNA D)显示保留所有B6SV等位基因的杂合性丧失,而第三克隆B0-A8(Lrp5<sup>Hum/Hum</sup>,gRNA A+F)保留所有129等位基因。其他三个克隆B0-F10(Lrp5<sup>Hum/Hum</sup>,gRNA A+F)、B0-G11(Lrp5<sup>Hum/Hum</sup>,gRNA A+F)和BP-G7(Lrp5<sup>Δ/Δ</sup>,gRNA A+F)保持杂合。

[0565] 此外,129等位基因与B6等位基因之间的单核苷酸变体(SNV)通过TAQMAN<sup>®</sup>等位基因辨别测定来测定。SNV测定在图12中的染色体19图谱上的近似位置由箭头连同在下部的测定编号一起加以显示,并且下方给出它们离Lrp5基因座的距离(以Mb计)。离Lrp5基因座的距离(以Mb计)如下:在Lrp5的着丝粒方向0.32(C2),在Lrp5的端粒方向1.2(T3),在Lrp5的端粒方向11.1(T6),在Lrp5的端粒方向13.2(T7),在Lrp5的端粒方向17.5(T8),在Lrp5的端粒方向25.8(T9),在Lrp5的端粒方向33.0(T10),在Lrp5的端粒方向38.3(T11),在Lrp5的端粒方向49.6(T13),以及在Lrp5的端粒方向57.2(T14)。129特异性探针和B6特异性探针以及引物对显示于表1中。根据SV测定,显示端粒方向杂合性丧失(L0H)的三个克隆(BC-H9、B0-A8和BR-B4)的结果显示于图16A-图16C中。SNV测定(图16A-图16C并且数据未显示)确认在Lrp5的端粒侧上在染色体19的长臂上的基因转换事件(SNV 1.2和SNV 57.2;分别参见图16B和图16C),但SNV 0.32测定(参见图16A)显示所有克隆就在着丝粒侧上离开Lrp5 320kb的等位基因来说都保持杂合。在24个Lrp5<sup>Hum/Hum</sup>或Lrp5<sup>Δ/Δ</sup>测定克隆中,我们发现

六个在Lrp5的端粒侧上在染色体19的整个长臂上具有杂合性丧失的迹象。五个克隆(四个Lrp5<sup>Hum/Hum</sup>和一个Lrp5<sup>Δ/Δ</sup>)从杂合性B6转换成纯合性B6,而第六克隆(Lrp5<sup>Hum/Hum</sup>)转换成纯合性129。CCN测定证明染色体19的两个拷贝得以保留。用于21个Hc纯合性克隆的类似杂合性丧失测定揭示对于在Hc基因的端粒侧上的所有SV和SNV,两个克隆R-E2(Hc<sup>Hum/Hum</sup>,gRNA A+F)和R-E8(Hc<sup>Δ/Δ</sup>,gRNA A+F)显示杂合性丧失而成为纯合性129,而对于在着丝粒侧上的所有等位基因,保留杂合性。CCN测定指示染色体2未丧失。

[0566] 我们的结果首次证明CRISPR/Cas9可增强用于达成超过100kb的大型单步人源化的同源性引导的修复,此使大规模基因组工程化的可能性扩大。使LTVEC和gRNA/Cas9组合的最显著和出乎意料的益处是它们能够促进纯合性靶向人源化。尽管双等位基因突变和纯合性靶向事件已在其他CRISPR/Cas9实验中加以报道,但这些基因修饰和插入中的大多数具有的数量级小于我们的人源化等位基因。在使用CRISPR/Cas9之前,我们从未发现由LTVEC达成的纯合性靶向,我们也尚未在我们使靶向单独基因的多个LTVEC组合时观察到同时靶向超过一个基因。鉴于这个经验,gRNA/Cas9诱导的纯合性靶向表明并非两个LTVEC分别靶向两个等位基因,在一个等位基因上的初始靶向事件可能充当用于由一个或多个Cas9切割促进的使另一等位基因同源性转换的模板。对双重gRNA/Cas9诱导的大型双等位基因缺失也具有纯合性的揭示(表19)提供对基因转换机理的进一步支持。

[0567] 杂合性丧失测定(图12)证明多个等位基因的涵盖在靶标基因的端粒侧上的大型染色体片段的大规模基因转换导致一些纯合性人源化和大型缺失。这个类型的长程有方向性基因转换与在细胞周期的G2期,同源染色体的经复制染色单体之间的有丝分裂重组一致(Lefebvre等(2001)Nat.Genet.27:257-258)(图17A-图17C)。尽管这仅解释少数纯合性事件,但这个机理可提供gRNA/Cas9裂解可用于促进在染色体的大部分上的多个等位基因从杂合向纯合进行大规模转换所采用的手段。然而,大多数纯合性事件似乎已是其机理值得进一步探究的局部基因转换的结果。

[0568] 长程有方向性基因转换的其他证据通过分析在用编码Lrp5gRNA A和F的质粒、编码Cas9的质粒和靶向在端粒方向离开Lrp5 30Mb的Ch25h基因的LTVEC对F1H4杂交ES细胞(其包含50%129SvS6品系和50%C57BL/6N品系)进行电穿孔之后获得的三个克隆来提供。在使用TAQMAN<sup>®</sup>测定在2个gRNA之间的预测缺失内部(离5'末端500bp以及离3'末端2kb)进行初级筛选之后,三个克隆初始被评分为野生型,但测定129等位基因与B6等位基因之间的单核苷酸变体(SNV)的后续TAQMAN<sup>®</sup>等位基因辨别测定惊人地揭示杂合性丧失。所用SNV测定是一个着丝粒方向测定(SNV 0.32)和两个端粒方向测定(SNV 1.2和SNV 57.2)(参见图12)。如表20中所示,着丝粒方向SNV测定(0.32Mb)确认在所有三个克隆中都保留杂合性。然而,两个端粒方向SNV测定均显示BP-E7和BP-H4就129等位基因来说是纯合的,并且两个端粒方向SNV测定均显示BP-E6就B6等位基因来说是纯合的。所有三个克隆都显示染色体19的两个拷贝得以保留,并且所有三个克隆就LTVEC靶向来说都是转基因的(即Ch25h基因座被靶向)。这些结果打开了使用靶向CRISPR/Cas9裂解达成强制纯合性的可能性。

[0569] 表20.SNV等位基因辨别测定的筛选结果。

|        |    |          |         |          |
|--------|----|----------|---------|----------|
| [0570] | 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   |

[0571] 若干可能机理可解释在 小鼠 F1H4 杂交 ES 细胞 (其包含 50% 129SvS6 品系和 50% C57BL/6N 品系) 的情况下在 CRISPR/Cas9 辅助的 LTVEC 人源化实验中观察到的结果 (参见图 18A-图 18F)。所述机理可通过由有丝分裂交叉达成的相互染色单体交换 (参见图 18A-图 18C), 或通过利用断裂诱导的复制进行的染色单体拷贝 (参见图 18D-图 18E) 来发生。在任一情况下, 都可发生杂合性修饰, 其中在基因组复制之前 129 染色体或 B6 染色体由 LTVEC 靶向 (参见图 18A 和图 18D)。或者, 单一 129 染色单体或单一 B6 染色单体可在基因组复制之后由 LTVEC 靶向, 继之以染色单体间基因转换 (参见图 18B 和图 18E)。或者, 在靶基因组基因座处可缺乏 LTVEC 靶向, 但 Cas9 裂解可在 129 或 B6 染色体上发生 (参见图 18C 和图 18F)。这个后述可能性可解释用 BP-E7、BP-H4 和 BP-E6 克隆观察到的结果。潜在结果显示于图 18A-图 18F 中。对于图 18F, 如果 Cas9 裂解 129 染色单体, 那么也有可能观察到保留 B6 等位基因的杂合性丧失 (LOH)。在上述实验中, 已观察到导致两个等位基因均被靶向 (Hum/Hum) 或两个等位基因均是野生型等位基因 (+/+) 的杂合性丧失事件。

[0572] 实施例 7. 对在 B6 等位基因与 129 等位基因之间具有最小变化的基因的纯合性靶向。

[0573] 也测试若干其他基因座的纯合性靶向。在另一实验中, LTVEC 被设计来产生 小鼠 Adamts5 (具有血栓反应素基序的解整合素和金属蛋白酶 5) 基因的 38kb 缺失, 以及同时用人 ADAMTS5 基因的 43kb 片段进行的替换。LTVEC 包含人 ADAMTS5 基因的 43kb 片段, 其由含有源于 小鼠 Adamts5 基因座的侧接于 小鼠 Adamts5 基因的意图缺失的 38kb 序列的部分的 22kb 和 46kb 的基因组 DNA 的同源臂侧接。在单独实验中, 我们使 Adamts5 人源化 LTVEC 与编码 Cas9 的质粒和编码被设计来在 小鼠 Adamts5 基因的靶向缺失区域内产生双链断裂的八个 sgRNA (gA、gA2、gB、gC、gD、gE、gE2、和 gF) 中的一者或两者的第二质粒或多个质粒组合。sgRNA 被设计来避免识别人 ADAMTS5 的插入部分中的任何序列。

[0574] CRISPR/Cas9 辅助的 Adamts5 基因人源化的结果显示于表 21 中。当将单独 LTVEC 引入 ES 细胞中时, 我们发现 96 个筛选的药物抗性克隆都不携带正确靶向的单等位基因杂合性人源化等位基因。相比之下, 使 LTVEC 与由八个测试 sgRNA 中的两者 (B 和 F; 参见表 1) 引导的 Cas9 核酸内切酶组合在 1.0% 的效率下产生正确靶向的单等位基因杂合性突变或双等位基因复合杂合性突变。未观察到纯合性靶向修饰。在额外实验中, 也使 gRNA A2 和 E2 组合, 但仍然未观察到纯合性靶向。

[0575] 表 21. CRISPR/Cas9 辅助的 Adamts5 基因人源化的筛选结果。

[0576]

| sgRNA 位置 | 离缺失端点的近似距离(bp) | gRNA    | CRISPR 活性(%) | 筛选的克隆 | 杂合性 靶向 | 复合杂合性 | 纯合性 靶向 |
|----------|----------------|---------|--------------|-------|--------|-------|--------|
| 5'       | 100            | gRNA A  | 85.7         | 96    | 0      | 0     | 0      |
| 5'       | 500            | gRNA A2 | 54.8         | 96    | 0      | 0     | 0      |
| 5'       | 1000           | gRNA B  | 66.7         | 96    | 1      | 0     | 0      |
| 中间       | 18700/18950    | gRNA C  | 9.5          | 96    | 0      | 0     | 0      |
| 中间       | 18800/18850    | gRNA D  | 4.8          | 96    | 0      | 0     | 0      |
| 3'       | 1000           | gRNA F  | 36.9         | 96    | 0      | 1     | 0      |
| 3'       | 500            | gRNA E  | 54.8         | 96    | 0      | 0     | 0      |
| 3'       | 100            | gRNA E2 | 54.8         | 96    | 0      | 0     | 0      |
| 5'和 3'   | 500/100        | A2 + E2 | 无测定          | 96    | 0      | 0     | 0      |
| N/A      | N/A            | 无       | N/A          | 96    | 0      | 0     | 0      |

[0577] 在另一实验中,LTVEC被设计来产生小鼠Dpp4(二肽基肽酶4)基因的79kb缺失,以及同时用同源性人DPP4基因的82kb片段进行的替换。LTVEC包含由5'同源臂和3'同源臂侧接的人DPP4基因的82kb片段,所述同源臂各自含有源于小鼠Dpp4基因座的侧接于小鼠Dpp4基因的意图缺失的79kb序列的部分的46kb的基因组DNA。在单独实验中,我们使Dpp4人源化LTVEC与编码Cas9的质粒和编码被设计来在小鼠Dpp4基因的靶向缺失区域内产生双链断裂的八个sgRNA(gA、gB、gB2、gC、gD、gE、gE2和gF)中的一者或两者的第二质粒或多个质粒组合。sgRNA被设计来避免识别人DPP4基因的插入部分中的任何序列。

[0578] CRISPR/Cas9辅助的Dpp4基因人源化的结果显示于表22中。当将单独LTVEC引入ES细胞中时,我们发现2.1%的筛选药物抗性克隆携带正确靶向的单等位基因杂合性人源化等位基因。相比之下,使LTVEC与由八个测试sgRNA(A、B、B2、C、D、E、E2和F;参见表1)中的任一者引导的Cas9核酸内切酶组合在2.1至7.3%的范围内的效率下产生正确靶向的单等位基因杂合性突变。未观察到纯合性靶向修饰。在额外实验中,使gRNA A和F或gRNA A和D组合,但仍然未观察到纯合性靶向。

[0579] 表22. CRISPR/Cas9辅助的Dpp4基因人源化的筛选结果。

[0580]

| sgRNA 位置 | 离缺失端点的近似距离(bp) | gRNA    | CRISPR 活性(%) | 筛选的克隆 | 杂合性 靶向 | 复合杂合性 | 纯合性 靶向 |
|----------|----------------|---------|--------------|-------|--------|-------|--------|
| 5'       | 50             | gRNA A  | 无测定          | 96    | 7      | 0     | 0      |
| 5'       | 400            | gRNA B  | 无测定          | 96    | 2      | 0     | 0      |
| 5'       | 900            | gRNA B2 | 无测定          | 96    | 5      | 0     | 0      |
| 中间       | 38800/40200    | gRNA C  | 无测定          | 96    | 3      | 0     | 0      |
| 中间       | 40800/38100    | gRNA D  | 无测定          | 96    | 3      | 0     | 0      |
| 3'       | 900            | gRNA E2 | 无测定          | 96    | 2      | 0     | 0      |
| 3'       | 500            | gRNA E  | 无测定          | 96    | 6      | 0     | 0      |
| 3'       | 200            | gRNA F  | 无测定          | 96    | 5      | 0     | 0      |
| 5'和 3'   | 50/38100       | A + D   | 无测定          | 384   | 4      | 0     | 0      |
| 5'和 3'   | 50/200         | A + F   | 无测定          | 384   | 9      | 0     | 0      |
| N/A      | N/A            | 无       | N/A          | 96    | 2      | 0     | 0      |

[0581] 在另一实验中,LTVEC被设计来产生小鼠Folh1(谷氨酸羧肽酶2)基因的55kb缺失,

以及同时用同源性人FOLH1基因的61kb片段进行的替换。LTVEC包含人FOLH1基因的61kb片段,其由含有源于小鼠Folh1基因座的侧接于小鼠Folh1基因的意图缺失的55kb序列的部分的22kb和46kb的基因组DNA的同源臂侧接。在单独实验中,我们使Folh1人源化LTVEC与编码Cas9的质粒和编码被设计来在小鼠Folh1基因的靶向缺失区域内产生双链断裂的八个sgRNA (gA、gA2、gB、gC、gD、gF、gE和gE2) 中的一者或两者的第二质粒或多个质粒组合.sgRNA被设计来避免识别人FOLH1基因的插入部分中的任何序列。

[0582] CRISPR/Cas9辅助的Folh1基因人源化的结果显示于表23中。当将单独LTVEC引入ES细胞中时,我们发现96个筛选的药物抗性克隆都不携带正确靶向的单等位基因杂合性人源化等位基因。相比之下,使LTVEC与由六个测试sgRNA中的三者 (A、D和E2;参见表1) 引导的Cas9核酸内切酶组合在1.0至3.1%的范围内的效率下产生正确靶向的单等位基因杂合性突变。未观察到纯合性靶向修饰。在额外实验中,使gRNA A和E2或gRNA A和D组合,但仍然未观察到纯合性靶向。

[0583] 表23. CRISPR/Cas9辅助的Folh1基因人源化的筛选结果。

| sgRNA<br>位置 | 离缺失端点的<br>近似距离(bp) | gRNA    | CRISPR<br>活性(%) | 筛选的<br>克隆 | 杂合性<br>靶向 | 复合杂<br>合性 | 纯合性<br>靶向 |
|-------------|--------------------|---------|-----------------|-----------|-----------|-----------|-----------|
| 5'          | 100                | gRNA A  | 45.2            | 96        | 2         | 0         | 0         |
| 5'          | 500                | gRNA A2 | 61.9            | 96        | 0         | 0         | 0         |
| 5'          | 1000               | gRNA B  | N/A             | N/A       | N/A       | N/A       | N/A       |
| 中间          | 30300/24800        | gRNA C  | 7.1             | 96        | 0         | 0         | 0         |
| 中间          | 31290/23810        | gRNA D  | 39.2            | 96        | 1         | 0         | 0         |
| 3'          | 1000               | gRNA F  | N/A             | N/A       | N/A       | N/A       | N/A       |
| 3'          | 500                | gRNA E2 | 无测定             | 96        | 1         | 0         | 0         |
| 3'          | 100                | gRNA E  | 1.2             | 96        | 0         | 0         | 0         |
| 5'和 3'      | 100/23810          | A + D   | 无测定             | 96        | 3         | 0         | 0         |
| 5'和 3'      | 100/500            | A + E2  | 无测定             | 96        | 0         | 0         | 0         |
| N/A         | N/A                | 无       | N/A             | 96        | 0         | 0         | 0         |

[0585] 在靶向不同基因座时观察到的纯合性靶向克隆的概述提供于表24中。

[0586] 表24. 在不同基因座的情况下的纯合性靶向克隆的数目。

| 基因              | <i>Adamts5</i> | <i>Trpa1</i> | <i>Folh1</i>     | <i>Lrp5</i>                                      | <i>C5</i>        | <i>Dpp4</i>     | <i>Ror1</i>     |
|-----------------|----------------|--------------|------------------|--------------------------------------------------|------------------|-----------------|-----------------|
| MAID #          | 7028           | 7002         | 7044             | 7064                                             | 7140             | 7326            | 7292            |
| Del/Ins (kb)    | 38/43          | 45/55        | 55/61            | 68/91                                            | 76/97            | 79/82           | 110/134         |
| gRNA 组合         | A2 + E2        | B + F        | A + D,<br>A + E2 | A + F,<br>B + E,<br>B2 + E2,<br>A + E,<br>A + E2 | A + C,<br>A + E2 | A + F,<br>A + D | A + D,<br>A + F |
| 纯合性靶向克隆         | 0              | 1            | 0                | 12                                               | 4                | 0               | 0               |
| 用于设计同源臂<br>的基因组 | 129 Bac        | 129 Bac      | B6 Bac           | 129 Bac                                          | 129 Bac          | B6 Bac          | B6 Bac          |

[0588] 在这些实验中,对于在B6等位基因与129等位基因之间具有最小序列变化的基因,纯合性靶向性最高。这在图19-图25中加以说明。在各图中,在竖直虚线内部的区域指示靶向区域(在LTVEC的5'靶标序列和3'靶标序列内部的区域)。举例来说,对于Lrp5(参见图19),基于129基因组来设计LTVEC的同源臂。用于确定单核苷酸变化的参照序列是来自

Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系、来自Taconic Biosciences的C57BL/6N品系、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系进行比较。竖直线代表相较于参照序列的单核苷酸变化。图20-图25分别提供对Hc、Trpa1、Adamts5、Folh1、Dpp4、Ror1和CD3的类似分析。

[0589] 如表24中以及图19-图25中所示,在Lrp5基因座(12个纯合性克隆)和Hc/C5基因座(4个纯合性克隆)的情况下产生最高数目的纯合性靶向克隆。这些靶基因组基因座各自具有极少数单核苷酸变化,特别是在gRNA识别序列处或附近或侧接于意图进行缺失和替换的区域(参见图19(Lrp5)和图20(C5))。

[0590] 相比之下,对于在B6等位基因与129等位基因之间,特别是在gRNA识别序列处或附近或侧接于意图进行缺失和替换的区域具有高密度的等位基因序列变化的基因,纯合性靶向低下或不存在。举例来说,当靶向Adamts5、Folh1、Dpp4或Ror1基因座时,未产生纯合性克隆(分别是图22-图25)。然而,当靶向在意图进行缺失和替换的区域的3'具有高密度的等位基因序列变化,但在意图进行缺失和替换的区域的5' (即在5' gRNA识别序列处或附近)具有低密度的等位基因序列变化的Trpa1基因座时,产生纯合性克隆(图21)。

[0591] 实施例8.使用相对于同源染色体对中的各染色体设计的靶向载体不使纯合性靶向增加。

[0592] 为进一步测试纯合性靶向修饰是通过在同源染色体对中的各染色体上发生独立靶向事件,还是通过在同源染色体对中的一个染色体上发生靶向事件,接着还在所述同源染色体对之间进行基因转换或杂合性丧失来产生,靶向在同源染色体对之间、在gRNA识别序列处或附近或侧接于意图进行缺失和替换的区域具有大量等位基因序列变化的另一基因组基因座。参见例如图26。这允许我们考查等位基因变化对纯合性垮塌或纯合性靶向的影响。在竖直虚线内部的区域是靶向区域(在LTVEC的5'靶标序列和3'靶标序列内部的区域)。用于确定单核苷酸变化的参照序列是来自Jackson Laboratory的C57BL/6J小鼠品系的基因组序列。将这个参照序列与来自Taconic Biosciences的129S6/SvEv品系MP变体、来自Taconic Biosciences的C57BL/6N品系RGC变体、以及由129S6/SvEv品系和C57BL/6N品系产生的VGF1杂交细胞系(表示在图的底部部分中的三行中)进行比较。三行中的各者中的竖直线代表相较于参照序列的单核苷酸变化。

[0593] 在这个实验中,两个LTVEC被设计来产生小鼠基因座的33kb缺失,以及同时用直系同源人基因的包括三个区段(6.8kb、0.1kb和1.7kb)的34.5kb片段进行的替换,其中在人区段之间具有小鼠基因座的间插区段。实验用VGF1(F1H4)(即我们的C57BL6NTac/129S6SvEvF1杂交XY ES细胞系)进行(Poueymirou等(2007)Nat.Biotechnol.25:91-99; Valenzuela等(2003)Nat.Biotechnol.21:652-659)。ES细胞如先前所述加以培养(Matise等(2000),Joyner,A.L.编Gene Targeting:a practical approach,第100-132页,Oxford University Press,New York)。VGF1细胞通过使雌性C57BL/6NTac小鼠与雄性129S6/SvEvTac小鼠杂交以产生C57BL6( $X^{B6}$ )/129S6( $Y^{129}$ )小鼠来产生。参见图7。

[0594] 一个LTVEC具有相对于VGF1细胞中的129染色体设计的同源臂,并且包括Neo选择盒(MAID#7170),并且另一LTVEC具有相对于C57BL6染色体设计的同源臂,并且包括Hyg选择盒(MAID#7314)。两个LTVEC在其他方面相同。

[0595] 在单独实验中,我们使两个人源化LTVEC与编码Cas9的质粒和编码被设计来在小鼠基因的靶向缺失区域内产生双链断裂的四个sgRNA (mGU、mGU2、mGD、mGD2) 的第二质粒或多个质粒组合。sgRNA被设计来避免识别人基因的插入部分中的任何序列。

[0596] 筛选总计192个Neo<sup>+</sup>克隆,128个Hyg<sup>+</sup>克隆,以及16个Neo<sup>+</sup>/Hyg<sup>+</sup>克隆。使LTVEC与由四个sgRNA引导的Cas9核酸内切酶组合产生一些杂合性靶向克隆、半合性靶向克隆、双等位基因垮塌克隆、具有NHEJ缺失的杂合性靶向克隆、具有双等位基因NHEJ缺失的克隆、以及具有NHEJ缺失的杂合性垮塌克隆。然而,未观察到纯合性靶向克隆。这表明并非是在同源染色体对内的各染色体上发生的单独靶向事件,而是局部基因转换事件导致在其他实验中观察到的纯合性靶向克隆。如果同源染色体对内的各染色体上发生的独立靶向事件导致在其他实验中观察到的纯合性靶向克隆,那么使用针对同源染色体对内的两个染色体中的各者特别加以调适的两个靶向载体将被预期会产生纯合性靶向克隆,尽管在5' 同源臂和3' 同源臂的5' 靶标序列和3' 靶标序列内具有高百分比的等位基因序列变化,因为针对各染色体加以调适的靶向载体将那个等位基因序列变化定址在5' 靶标序列和3' 靶标序列内。然而,使用两个LTVEC未产生任何纯合性靶向克隆。这进一步支持纯合性靶向修饰通过局部基因转换事件来产生的理念,如图27中所描绘。相比于极性基因转换,我们已在更高比率下观察到在靶向缺失和插入的两侧上的局部杂合性丧失。

- [0001] 序列表
- [0002] <110> 瑞泽恩制药公司
- [0003] <120> 用于使用多个引导RNA来破坏免疫耐受性的方法
- [0004] <130> 057766/497023
- [0005] <150> 62/339,472
- [0006] <151> 2016-05-20
- [0007] <150> 62/368,604
- [0008] <151> 2016-07-29
- [0009] <160> 152
- [0010] <170> 用于Windows的4.0版FastSEQ
- [0011] <210> 1
- [0012] <211> 23
- [0013] <212> DNA
- [0014] <213> 人工序列
- [0015] <220>
- [0016] <223> 引导RNA识别序列
- [0017] <220>
- [0018] <221> 尚未归类的特征
- [0019] <222> (2) ... (21)
- [0020] <223> n = A、T、C或G
- [0021] <400> 1
- [0022] gnnnnnnnnnn nnnnnnnnnn ngg 23
- [0023] <210> 2
- [0024] <211> 23
- [0025] <212> DNA
- [0026] <213> 人工序列
- [0027] <220>
- [0028] <223> 引导RNA识别序列
- [0029] <220>
- [0030] <221> 尚未归类的特征
- [0031] <222> (1) ... (21)
- [0032] <223> n = A、T、C或G
- [0033] <400> 2
- [0034] nnnnnnnnnnn nnnnnnnnnn ngg 23
- [0035] <210> 3
- [0036] <211> 25
- [0037] <212> DNA
- [0038] <213> 人工序列
- [0039] <220>
- [0040] <223> 引导RNA识别序列
- [0041] <220>

- [0042] <221> 尚未归类的特征
- [0043] <222> (3) ... (23)
- [0044] <223> n = A、T、C或G
- [0045] <400> 3
- [0046] ggnnnnnnnnn nnnnnnnnnn nnnngg 25
- [0047] <210> 4
- [0048] <211> 20
- [0049] <212> DNA
- [0050] <213> 人工序列
- [0051] <220>
- [0052] <223> C5 (Hc) gRNA A DNA靶向区段(离靶基因座端点100 bp)
- [0053] <400> 4
- [0054] atcacaacc agttaaccgg 20
- [0055] <210> 5
- [0056] <211> 20
- [0057] <212> DNA
- [0058] <213> 人工序列
- [0059] <220>
- [0060] <223> C5 (Hc) gRNA B DNA靶向区段(离靶基因座端点500 bp)
- [0061] <400> 5
- [0062] tttcagacga gccgaccgg 20
- [0063] <210> 6
- [0064] <211> 20
- [0065] <212> DNA
- [0066] <213> 人工序列
- [0067] <220>
- [0068] <223> C5 (Hc) gRNA C DNA靶向区段(离靶基因座端点38200和37500 bp)
- [0069] <400> 6
- [0070] tgtgtgtcat agcgatgtcg 20
- [0071] <210> 7
- [0072] <211> 20
- [0073] <212> DNA
- [0074] <213> 人工序列
- [0075] <220>
- [0076] <223> C5 (Hc) gRNA D DNA靶向区段(离靶基因座端点43500和32200 bp)
- [0077] <400> 7
- [0078] aacaggtacc ctatcctcac 20
- [0079] <210> 8
- [0080] <211> 20
- [0081] <212> DNA
- [0082] <213> 人工序列
- [0083] <220>

- [0084] <223> C5 (Hc) gRNA E DNA靶向区段(离靶基因座端点500 bp)  
[0085] <400> 8  
[0086] ggcccggacc tagtctctct 20  
[0087] <210> 9  
[0088] <211> 20  
[0089] <212> DNA  
[0090] <213> 人工序列  
[0091] <220>  
[0092] <223> C5 (Hc) gRNA E2 DNA靶向区段(离靶基因座端点100 bp)  
[0093] <400> 9  
[0094] tcgtggttgc atgcgcactg 20  
[0095] <210> 10  
[0096] <211> 20  
[0097] <212> DNA  
[0098] <213> 人工序列  
[0099] <220>  
[0100] <223> Lrp5 gRNA A DNA靶向区段(离靶基因座端点50 bp)  
[0101] <400> 10  
[0102] gggaacccac agcatactcc 20  
[0103] <210> 11  
[0104] <211> 20  
[0105] <212> DNA  
[0106] <213> 人工序列  
[0107] <220>  
[0108] <223> Lrp5 gRNA B DNA靶向区段(离靶基因座端点500 bp)  
[0109] <400> 11  
[0110] gaatcatgca cggctacccc 20  
[0111] <210> 12  
[0112] <211> 20  
[0113] <212> DNA  
[0114] <213> 人工序列  
[0115] <220>  
[0116] <223> Lrp5 gRNA B2 DNA靶向区段(离靶基因座端点1000 bp)  
[0117] <400> 12  
[0118] tgctcctatg gggaggcgcg 20  
[0119] <210> 13  
[0120] <211> 20  
[0121] <212> DNA  
[0122] <213> 人工序列  
[0123] <220>  
[0124] <223> Lrp5 gRNA C DNA靶向区段(离靶基因座端点29900和38430 bp)  
[0125] <400> 13

|        |                                                  |
|--------|--------------------------------------------------|
| [0126] | actgagatca atgaccccga 20                         |
| [0127] | <210> 14                                         |
| [0128] | <211> 20                                         |
| [0129] | <212> DNA                                        |
| [0130] | <213> 人工序列                                       |
| [0131] | <220>                                            |
| [0132] | <223> Lrp5 gRNA D DNA靶向区段(离靶基因座端点29950和38380 bp) |
| [0133] | <400> 14                                         |
| [0134] | gggtcgcccg gaacctctac 20                         |
| [0135] | <210> 15                                         |
| [0136] | <211> 20                                         |
| [0137] | <212> DNA                                        |
| [0138] | <213> 人工序列                                       |
| [0139] | <220>                                            |
| [0140] | <223> Lrp5 gRNA E2 DNA靶向区段(离靶基因座端点1000 bp)       |
| [0141] | <400> 15                                         |
| [0142] | cttggataac attgataccc 20                         |
| [0143] | <210> 16                                         |
| [0144] | <211> 20                                         |
| [0145] | <212> DNA                                        |
| [0146] | <213> 人工序列                                       |
| [0147] | <220>                                            |
| [0148] | <223> Lrp5 gRNA E DNA靶向区段(离靶基因座端点500 bp)         |
| [0149] | <400> 16                                         |
| [0150] | ggggcagagc ccttatatca 20                         |
| [0151] | <210> 17                                         |
| [0152] | <211> 20                                         |
| [0153] | <212> DNA                                        |
| [0154] | <213> 人工序列                                       |
| [0155] | <220>                                            |
| [0156] | <223> Lrp5 gRNA F DNA靶向区段(离靶基因座端点50 bp)          |
| [0157] | <400> 17                                         |
| [0158] | tcgctcacat taatccctag 20                         |
| [0159] | <210> 18                                         |
| [0160] | <211> 20                                         |
| [0161] | <212> DNA                                        |
| [0162] | <213> 人工序列                                       |
| [0163] | <220>                                            |
| [0164] | <223> Ror1 gRNA A DNA靶向区段(离靶基因座端点200 bp)         |
| [0165] | <400> 18                                         |
| [0166] | tgtgggcctt tgctgatcac 20                         |
| [0167] | <210> 19                                         |

|        |                                                  |
|--------|--------------------------------------------------|
| [0168] | <211> 20                                         |
| [0169] | <212> DNA                                        |
| [0170] | <213> 人工序列                                       |
| [0171] | <220>                                            |
| [0172] | <223> Ror1 gRNA B DNA靶向区段(离靶基因座端点1000 bp)        |
| [0173] | <400> 19                                         |
| [0174] | aatctatgat cctatggcct 20                         |
| [0175] | <210> 20                                         |
| [0176] | <211> 20                                         |
| [0177] | <212> DNA                                        |
| [0178] | <213> 人工序列                                       |
| [0179] | <220>                                            |
| [0180] | <223> Ror1 gRNA D DNA靶向区段(离靶基因座端点54300和55500 bp) |
| [0181] | <400> 20                                         |
| [0182] | tgccaatagc agtgacttga 20                         |
| [0183] | <210> 21                                         |
| [0184] | <211> 20                                         |
| [0185] | <212> DNA                                        |
| [0186] | <213> 人工序列                                       |
| [0187] | <220>                                            |
| [0188] | <223> Ror1 gRNA C DNA靶向区段(离靶基因座端点54500和55300 bp) |
| [0189] | <400> 21                                         |
| [0190] | gggaagaatg ggctattgtc 20                         |
| [0191] | <210> 22                                         |
| [0192] | <211> 20                                         |
| [0193] | <212> DNA                                        |
| [0194] | <213> 人工序列                                       |
| [0195] | <220>                                            |
| [0196] | <223> Ror1 gRNA E DNA靶向区段(离靶基因座端点1000 bp)        |
| [0197] | <400> 22                                         |
| [0198] | ggttgtttgt gctgatgacg 20                         |
| [0199] | <210> 23                                         |
| [0200] | <211> 20                                         |
| [0201] | <212> DNA                                        |
| [0202] | <213> 人工序列                                       |
| [0203] | <220>                                            |
| [0204] | <223> Ror1 gRNA F DNA靶向区段(离靶基因座端点200 bp)         |
| [0205] | <400> 23                                         |
| [0206] | ccgtcctagg ccttctacgt 20                         |
| [0207] | <210> 24                                         |
| [0208] | <211> 20                                         |
| [0209] | <212> DNA                                        |

- [0210] <213> 人工序列  
[0211] <220>  
[0212] <223> Trpa1 gRNA A DNA靶向区段 (离靶基因座端点100 bp)  
[0213] <400> 24  
[0214] gtactgggga atcggtggtc 20  
[0215] <210> 25  
[0216] <211> 20  
[0217] <212> DNA  
[0218] <213> 人工序列  
[0219] <220>  
[0220] <223> Trpa1 gRNA A2 DNA靶向区段 (离靶基因座端点500 bp)  
[0221] <400> 25  
[0222] cacgcactcc aaatttatcc 20  
[0223] <210> 26  
[0224] <211> 20  
[0225] <212> DNA  
[0226] <213> 人工序列  
[0227] <220>  
[0228] <223> Trpa1 gRNA B DNA靶向区段 (离靶基因座端点1000 bp)  
[0229] <400> 26  
[0230] ctaagtgtgt atcagtacat 20  
[0231] <210> 27  
[0232] <211> 20  
[0233] <212> DNA  
[0234] <213> 人工序列  
[0235] <220>  
[0236] <223> Trpa1 gRNA C DNA靶向区段 (离靶基因座端点25600和19740 bp)  
[0237] <400> 27  
[0238] tgccctgcac aataagcgca 20  
[0239] <210> 28  
[0240] <211> 20  
[0241] <212> DNA  
[0242] <213> 人工序列  
[0243] <220>  
[0244] <223> Trpa1 gRNA D DNA靶向区段 (离靶基因座端点26970和18370 bp)  
[0245] <400> 28  
[0246] actcattgaa acgttatggc 20  
[0247] <210> 29  
[0248] <211> 20  
[0249] <212> DNA  
[0250] <213> 人工序列  
[0251] <220>

[0252] <223> Trpa1 gRNA E2 DNA靶向区段(离靶基因座端点1000 bp)  
[0253] <400> 29  
[0254] agtaagggtg gattaaattc 20  
[0255] <210> 30  
[0256] <211> 20  
[0257] <212> DNA  
[0258] <213> 人工序列  
[0259] <220>  
[0260] <223> Trpa1 gRNA E DNA靶向区段(离靶基因座端点500 bp)  
[0261] <400> 30  
[0262] gccatctaga ttcattgtaac 20  
[0263] <210> 31  
[0264] <211> 20  
[0265] <212> DNA  
[0266] <213> 人工序列  
[0267] <220>  
[0268] <223> Trpa1 gRNA F DNA靶向区段(离靶基因座端点100 bp)  
[0269] <400> 31  
[0270] gactagaaat gttctgcacc 20  
[0271] <210> 32  
[0272] <211> 21  
[0273] <212> DNA  
[0274] <213> 人工序列  
[0275] <220>  
[0276] <223> 190045正向引物  
[0277] <400> 32  
[0278] gagctcatag ccaacagctt g 21  
[0279] <210> 33  
[0280] <211> 20  
[0281] <212> DNA  
[0282] <213> 人工序列  
[0283] <220>  
[0284] <223> 190061正向引物  
[0285] <400> 33  
[0286] atgcatcaga tcacgctcag 20  
[0287] <210> 34  
[0288] <211> 20  
[0289] <212> DNA  
[0290] <213> 人工序列  
[0291] <220>  
[0292] <223> 190068正向引物  
[0293] <400> 34

---

|        |                                       |
|--------|---------------------------------------|
| [0294] | gtccttgtgg catttccaac 20              |
| [0295] | <210> 35                              |
| [0296] | <211> 24                              |
| [0297] | <212> DNA                             |
| [0298] | <213> 人工序列                            |
| [0299] | <220>                                 |
| [0300] | <223> 190030正向引物                      |
| [0301] | <400> 35                              |
| [0302] | ccagtatggg gtcagttaat agcg 24         |
| [0303] | <210> 36                              |
| [0304] | <211> 19                              |
| [0305] | <212> DNA                             |
| [0306] | <213> 人工序列                            |
| [0307] | <220>                                 |
| [0308] | <223> 190033正向引物(与图6中的SV 48.3的正向引物相同) |
| [0309] | <400> 36                              |
| [0310] | ctgtgcagaa agcagcctc 19               |
| [0311] | <210> 37                              |
| [0312] | <211> 20                              |
| [0313] | <212> DNA                             |
| [0314] | <213> 人工序列                            |
| [0315] | <220>                                 |
| [0316] | <223> 190013正向引物                      |
| [0317] | <400> 37                              |
| [0318] | cctctccctc taggcacctg 20              |
| [0319] | <210> 38                              |
| [0320] | <211> 20                              |
| [0321] | <212> DNA                             |
| [0322] | <213> 人工序列                            |
| [0323] | <220>                                 |
| [0324] | <223> 190045反向引物                      |
| [0325] | <400> 38                              |
| [0326] | tctttaaggg ctccgttgtc 20              |
| [0327] | <210> 39                              |
| [0328] | <211> 20                              |
| [0329] | <212> DNA                             |
| [0330] | <213> 人工序列                            |
| [0331] | <220>                                 |
| [0332] | <223> 190061反向引物                      |
| [0333] | <400> 39                              |
| [0334] | aagaccaacc attcaccag 20               |
| [0335] | <210> 40                              |

|        |                                        |
|--------|----------------------------------------|
| [0336] | <211> 20                               |
| [0337] | <212> DNA                              |
| [0338] | <213> 人工序列                             |
| [0339] | <220>                                  |
| [0340] | <223> 190068反向引物                       |
| [0341] | <400> 40                               |
| [0342] | ttcccagtc aagtcaaagg 20                |
| [0343] | <210> 41                               |
| [0344] | <211> 20                               |
| [0345] | <212> DNA                              |
| [0346] | <213> 人工序列                             |
| [0347] | <220>                                  |
| [0348] | <223> 190030反向引物                       |
| [0349] | <400> 41                               |
| [0350] | ctgttatctg caaggcaccc 20               |
| [0351] | <210> 42                               |
| [0352] | <211> 20                               |
| [0353] | <212> DNA                              |
| [0354] | <213> 人工序列                             |
| [0355] | <220>                                  |
| [0356] | <223> 190033反向引物 (与图6中的SV 48.3的反向引物相同) |
| [0357] | <400> 42                               |
| [0358] | acaactggat cctgattcgc 20               |
| [0359] | <210> 43                               |
| [0360] | <211> 20                               |
| [0361] | <212> DNA                              |
| [0362] | <213> 人工序列                             |
| [0363] | <220>                                  |
| [0364] | <223> 190013反向引物                       |
| [0365] | <400> 43                               |
| [0366] | taagagggca tgggtgagac 20               |
| [0367] | <210> 44                               |
| [0368] | <211> 20                               |
| [0369] | <212> DNA                              |
| [0370] | <213> 人工序列                             |
| [0371] | <220>                                  |
| [0372] | <223> C2探针 (B6) -图6中的SNV 0.32          |
| [0373] | <400> 44                               |
| [0374] | aattcagaag acctatcgta 20               |
| [0375] | <210> 45                               |
| [0376] | <211> 21                               |
| [0377] | <212> DNA                              |

[0378] <213> 人工序列  
[0379] <220>  
[0380] <223> T3探针 (B6) -图6中的SNV 1.2  
[0381] <400> 45  
[0382] tatgtgtata ggtgtttgga t 21  
[0383] <210> 46  
[0384] <211> 19  
[0385] <212> DNA  
[0386] <213> 人工序列  
[0387] <220>  
[0388] <223> T6探针 (B6) -图6中的SNV 11.1  
[0389] <400> 46  
[0390] tacattgcta aatgaaacc 19  
[0391] <210> 47  
[0392] <211> 16  
[0393] <212> DNA  
[0394] <213> 人工序列  
[0395] <220>  
[0396] <223> T7探针 (B6) -图6中的SNV 13.2  
[0397] <400> 47  
[0398] cgcagtcacatg cacata 16  
[0399] <210> 48  
[0400] <211> 20  
[0401] <212> DNA  
[0402] <213> 人工序列  
[0403] <220>  
[0404] <223> T8探针 (B6) -图6中的SNV 17.5  
[0405] <400> 48  
[0406] ttataaaagcc cagtatgtac 20  
[0407] <210> 49  
[0408] <211> 14  
[0409] <212> DNA  
[0410] <213> 人工序列  
[0411] <220>  
[0412] <223> T9探针 (B6) -图6中的SNV 25.8  
[0413] <400> 49  
[0414] tgctgcataa tcag 14  
[0415] <210> 50  
[0416] <211> 18  
[0417] <212> DNA  
[0418] <213> 人工序列  
[0419] <220>

[0420] <223> T10探针 (B6) -图6中的SNV 33.0  
[0421] <400> 50  
[0422] tcaggagtga attggata 18  
[0423] <210> 51  
[0424] <211> 16  
[0425] <212> DNA  
[0426] <213> 人工序列  
[0427] <220>  
[0428] <223> T11探针 (B6) -图6中的SNV 38.3  
[0429] <400> 51  
[0430] ctgctactta cctttg 16  
[0431] <210> 52  
[0432] <211> 13  
[0433] <212> DNA  
[0434] <213> 人工序列  
[0435] <220>  
[0436] <223> T13探针 (B6) -图6中的SNV 49.6  
[0437] <400> 52  
[0438] aggaggaaaa cgc 13  
[0439] <210> 53  
[0440] <211> 17  
[0441] <212> DNA  
[0442] <213> 人工序列  
[0443] <220>  
[0444] <223> T14探针 (B6) -图6中的SNV 57.2  
[0445] <400> 53  
[0446] cctttgttcc tcataag 17  
[0447] <210> 54  
[0448] <211> 20  
[0449] <212> DNA  
[0450] <213> 人工序列  
[0451] <220>  
[0452] <223> C2探针 (129) -图6中的SNV 0.32  
[0453] <400> 54  
[0454] aattcagaag acctattgta 20  
[0455] <210> 55  
[0456] <211> 21  
[0457] <212> DNA  
[0458] <213> 人工序列  
[0459] <220>  
[0460] <223> T3探针 (129) -图6中的SNV 1.2  
[0461] <400> 55

[0462] tatgtgtata ggtgtttgca t 21  
[0463] <210> 56  
[0464] <211> 16  
[0465] <212> DNA  
[0466] <213> 人工序列  
[0467] <220>  
[0468] <223> T6探针 (129) -图6中的SNV 11.1  
[0469] <400> 56  
[0470] cattgctaca tgaaac 16  
[0471] <210> 57  
[0472] <211> 16  
[0473] <212> DNA  
[0474] <213> 人工序列  
[0475] <220>  
[0476] <223> T7探针 (129) -图6中的SNV 13.2  
[0477] <400> 57  
[0478] cgcagtcacg cacgta 16  
[0479] <210> 58  
[0480] <211> 23  
[0481] <212> DNA  
[0482] <213> 人工序列  
[0483] <220>  
[0484] <223> T8探针 (129) -图6中的SNV 17.5  
[0485] <400> 58  
[0486] tgagaatttta taaagcccaa tat 23  
[0487] <210> 59  
[0488] <211> 14  
[0489] <212> DNA  
[0490] <213> 人工序列  
[0491] <220>  
[0492] <223> T9探针 (129) -图6中的SNV 25.8  
[0493] <400> 59  
[0494] tgctgcatga tcag 14  
[0495] <210> 60  
[0496] <211> 15  
[0497] <212> DNA  
[0498] <213> 人工序列  
[0499] <220>  
[0500] <223> T10探针 (129) -图6中的SNV 33.0  
[0501] <400> 60  
[0502] tcaggagtga atcgg 15  
[0503] <210> 61

[0504] <211> 16  
[0505] <212> DNA  
[0506] <213> 人工序列  
[0507] <220>  
[0508] <223> T11探针(129)-图6中的SNV 38.3  
[0509] <400> 61  
[0510] ctgctagtta cctttg 16  
[0511] <210> 62  
[0512] <211> 15  
[0513] <212> DNA  
[0514] <213> 人工序列  
[0515] <220>  
[0516] <223> T13探针(129)-图6中的SNV 49.6  
[0517] <400> 62  
[0518] aggaggaaga cgcag 15  
[0519] <210> 63  
[0520] <211> 17  
[0521] <212> DNA  
[0522] <213> 人工序列  
[0523] <220>  
[0524] <223> T14探针(129)-图6中的SNV 57.2  
[0525] <400> 63  
[0526] ctttgttctt cataagc 17  
[0527] <210> 64  
[0528] <211> 25  
[0529] <212> DNA  
[0530] <213> 人工序列  
[0531] <220>  
[0532] <223> C2正向引物-图6中的SNV 0.32  
[0533] <400> 64  
[0534] atgagggatt tccttaatca gacaa 25  
[0535] <210> 65  
[0536] <211> 29  
[0537] <212> DNA  
[0538] <213> 人工序列  
[0539] <220>  
[0540] <223> T3正向引物-图6中的SNV 1.2  
[0541] <400> 65  
[0542] tggtatgttt attcttactc aaggttttg 29  
[0543] <210> 66  
[0544] <211> 22  
[0545] <212> DNA

[0546] <213> 人工序列  
[0547] <220>  
[0548] <223> T6正向引物-图6中的SNV 11.1  
[0549] <400> 66  
[0550] gggcaactga tggaaagaac tc 22  
[0551] <210> 67  
[0552] <211> 23  
[0553] <212> DNA  
[0554] <213> 人工序列  
[0555] <220>  
[0556] <223> T7正向引物-图6中的SNV 13.2  
[0557] <400> 67  
[0558] gactgacgca caaactgtgc ctt 23  
[0559] <210> 68  
[0560] <211> 26  
[0561] <212> DNA  
[0562] <213> 人工序列  
[0563] <220>  
[0564] <223> T8正向引物-图6中的SNV 17.5  
[0565] <400> 68  
[0566] cccaaagcat ataacaagaa caaatg 26  
[0567] <210> 69  
[0568] <211> 18  
[0569] <212> DNA  
[0570] <213> 人工序列  
[0571] <220>  
[0572] <223> T9正向引物-图6中的SNV 25.8  
[0573] <400> 69  
[0574] gcaggacgca ggcgttta 18  
[0575] <210> 70  
[0576] <211> 23  
[0577] <212> DNA  
[0578] <213> 人工序列  
[0579] <220>  
[0580] <223> T10正向引物-图6中的SNV 33.0  
[0581] <400> 70  
[0582] gcatactcat ggcagtctac atc 23  
[0583] <210> 71  
[0584] <211> 20  
[0585] <212> DNA  
[0586] <213> 人工序列  
[0587] <220>

[0588] <223> T11正向引物-图6中的SNV 38.3  
[0589] <400> 71  
[0590] cctgccccctt gatgagtgtt 20  
[0591] <210> 72  
[0592] <211> 23  
[0593] <212> DNA  
[0594] <213> 人工序列  
[0595] <220>  
[0596] <223> T13正向引物-图6中的SNV 49.6  
[0597] <400> 72  
[0598] ccctctttga tatgctcgtg tgt 23  
[0599] <210> 73  
[0600] <211> 22  
[0601] <212> DNA  
[0602] <213> 人工序列  
[0603] <220>  
[0604] <223> T14正向引物-图6中的SNV 57.2  
[0605] <400> 73  
[0606] tcccacaggt ccatgtcttt aa 22  
[0607] <210> 74  
[0608] <211> 28  
[0609] <212> DNA  
[0610] <213> 人工序列  
[0611] <220>  
[0612] <223> C2反向引物-图6中的SNV 0.32  
[0613] <400> 74  
[0614] agactacaat gagctaccat cataaggt 28  
[0615] <210> 75  
[0616] <211> 24  
[0617] <212> DNA  
[0618] <213> 人工序列  
[0619] <220>  
[0620] <223> T3反向引物-图6中的SNV 1.2  
[0621] <400> 75  
[0622] caaccatcta aaactccagt tcca 24  
[0623] <210> 76  
[0624] <211> 28  
[0625] <212> DNA  
[0626] <213> 人工序列  
[0627] <220>  
[0628] <223> T6反向引物-图6中的SNV 11.1  
[0629] <400> 76

[0630] tgtgtaacag gacagttgaa tgtagaga 28  
[0631] <210> 77  
[0632] <211> 19  
[0633] <212> DNA  
[0634] <213> 人工序列  
[0635] <220>  
[0636] <223> T7反向引物-图6中的SNV 13.2  
[0637] <400> 77  
[0638] cttaaaaccc gccctgcat 19  
[0639] <210> 78  
[0640] <211> 26  
[0641] <212> DNA  
[0642] <213> 人工序列  
[0643] <220>  
[0644] <223> T8反向引物-图6中的SNV 17.5  
[0645] <400> 78  
[0646] ctacaggaga tgtggctggt ctatgt 26  
[0647] <210> 79  
[0648] <211> 22  
[0649] <212> DNA  
[0650] <213> 人工序列  
[0651] <220>  
[0652] <223> T9反向引物-图6中的SNV 25.8  
[0653] <400> 79  
[0654] tcagcgtgat tcgcttgtag tc 22  
[0655] <210> 80  
[0656] <211> 26  
[0657] <212> DNA  
[0658] <213> 人工序列  
[0659] <220>  
[0660] <223> T10反向引物-图6中的SNV 33.0  
[0661] <400> 80  
[0662] tgcatagctg tttgaataat gacaag 26  
[0663] <210> 81  
[0664] <211> 21  
[0665] <212> DNA  
[0666] <213> 人工序列  
[0667] <220>  
[0668] <223> T11反向引物-图6中的SNV 38.3  
[0669] <400> 81  
[0670] tgcagcatct ctgtcaagca a 21  
[0671] <210> 82

[0672] <211> 24  
[0673] <212> DNA  
[0674] <213> 人工序列  
[0675] <220>  
[0676] <223> T13反向引物-图6中的SNV 49.6  
[0677] <400> 82  
[0678] gcaacaacat aaccacagc ataa 24  
[0679] <210> 83  
[0680] <211> 23  
[0681] <212> DNA  
[0682] <213> 人工序列  
[0683] <220>  
[0684] <223> T14反向引物-图6中的SNV 57.2  
[0685] <400> 83  
[0686] gctaagcggtt tggaagaaat tcc 23  
[0687] <210> 84  
[0688] <211> 20  
[0689] <212> DNA  
[0690] <213> 人工序列  
[0691] <220>  
[0692] <223> 图6中的SV 13.7的正向引物  
[0693] <400> 84  
[0694] taggctctaa ggatgctggc 20  
[0695] <210> 85  
[0696] <211> 20  
[0697] <212> DNA  
[0698] <213> 人工序列  
[0699] <220>  
[0700] <223> 图6中的SV 13.7的反向引物  
[0701] <400> 85  
[0702] aagcagcttc aaaccctctg 20  
[0703] <210> 86  
[0704] <211> 20  
[0705] <212> DNA  
[0706] <213> 人工序列  
[0707] <220>  
[0708] <223> 图6中的SV 20.0的正向引物  
[0709] <400> 86  
[0710] ttacttggcc ttggaactgc 20  
[0711] <210> 87  
[0712] <211> 21  
[0713] <212> DNA

[0714] <213> 人工序列  
[0715] <220>  
[0716] <223> 图6中的SV 20.0的反向引物  
[0717] <400> 87  
[0718] tgattcgtaa tcgtcactgc c 21  
[0719] <210> 88  
[0720] <211> 20  
[0721] <212> DNA  
[0722] <213> 人工序列  
[0723] <220>  
[0724] <223> 图6中的SV 36.9的正向引物  
[0725] <400> 88  
[0726] tcctgtcccg agaaactgtc 20  
[0727] <210> 89  
[0728] <211> 20  
[0729] <212> DNA  
[0730] <213> 人工序列  
[0731] <220>  
[0732] <223> 图6中的SV 36.9的反向引物  
[0733] <400> 89  
[0734] agctggcttt cagagagctg 20  
[0735] <210> 90  
[0736] <211> 20  
[0737] <212> DNA  
[0738] <213> 人工序列  
[0739] <220>  
[0740] <223> 图6中的SV 56.7的正向引物  
[0741] <400> 90  
[0742] ttagaaagtg ccaaccaggc 20  
[0743] <210> 91  
[0744] <211> 20  
[0745] <212> DNA  
[0746] <213> 人工序列  
[0747] <220>  
[0748] <223> 图6中的SV 56.7的反向引物  
[0749] <400> 91  
[0750] ctctggctag gaacaatggc 20  
[0751] <210> 92  
[0752] <211> 25  
[0753] <212> DNA  
[0754] <213> 人工序列  
[0755] <220>

[0756] <223> Lrp5基因座的m-lr-f引物  
[0757] <400> 92  
[0758] gtttaggtgca gggctctactc agctg 25  
[0759] <210> 93  
[0760] <211> 20  
[0761] <212> DNA  
[0762] <213> 人工序列  
[0763] <220>  
[0764] <223> Lrp5基因座的m-5' -f引物  
[0765] <400> 93  
[0766] ggaggagagg agaagcagcc 20  
[0767] <210> 94  
[0768] <211> 20  
[0769] <212> DNA  
[0770] <213> 人工序列  
[0771] <220>  
[0772] <223> Lrp5基因座的m-A引物  
[0773] <400> 94  
[0774] ggaggagagg agaagcagcc 20  
[0775] <210> 95  
[0776] <211> 26  
[0777] <212> DNA  
[0778] <213> 人工序列  
[0779] <220>  
[0780] <223> Lrp5基因座的h-lr-r引物  
[0781] <400> 95  
[0782] gcaaacagcc ttcttccac attcgg 26  
[0783] <210> 96  
[0784] <211> 24  
[0785] <212> DNA  
[0786] <213> 人工序列  
[0787] <220>  
[0788] <223> Lrp5基因座的m-5' -r引物  
[0789] <400> 96  
[0790] ttgctttcag tagttcaggt gtgc 24  
[0791] <210> 97  
[0792] <211> 20  
[0793] <212> DNA  
[0794] <213> 人工序列  
[0795] <220>  
[0796] <223> Lrp5基因座的h-5' -r引物  
[0797] <400> 97

|        |                                |
|--------|--------------------------------|
| [0798] | ggcgttgtca ggaagttgcc 20       |
| [0799] | <210> 98                       |
| [0800] | <211> 22                       |
| [0801] | <212> DNA                      |
| [0802] | <213> 人工序列                     |
| [0803] | <220>                          |
| [0804] | <223> Lrp5基因座的m-F引物            |
| [0805] | <400> 98                       |
| [0806] | tgaagttgag aggcacatga gg 22    |
| [0807] | <210> 99                       |
| [0808] | <211> 24                       |
| [0809] | <212> DNA                      |
| [0810] | <213> 人工序列                     |
| [0811] | <220>                          |
| [0812] | <223> Lrp5基因座的m-E2引物           |
| [0813] | <400> 99                       |
| [0814] | tagagtagcc acaggcagca aagc 24  |
| [0815] | <210> 100                      |
| [0816] | <211> 22                       |
| [0817] | <212> DNA                      |
| [0818] | <213> 人工序列                     |
| [0819] | <220>                          |
| [0820] | <223> 7064retU正向引物             |
| [0821] | <400> 100                      |
| [0822] | cctcctgagc tttcctttgc ag 22    |
| [0823] | <210> 101                      |
| [0824] | <211> 25                       |
| [0825] | <212> DNA                      |
| [0826] | <213> 人工序列                     |
| [0827] | <220>                          |
| [0828] | <223> 7064retU反向引物             |
| [0829] | <400> 101                      |
| [0830] | cctagacaac acagacactg tatca 25 |
| [0831] | <210> 102                      |
| [0832] | <211> 23                       |
| [0833] | <212> DNA                      |
| [0834] | <213> 人工序列                     |
| [0835] | <220>                          |
| [0836] | <223> 7064retU TAQMAN探针        |
| [0837] | <400> 102                      |
| [0838] | ttctgccctt gaaaaggaga ggc 23   |
| [0839] | <210> 103                      |

---

|        |                              |
|--------|------------------------------|
| [0840] | <211> 18                     |
| [0841] | <212> DNA                    |
| [0842] | <213> 人工序列                   |
| [0843] | <220>                        |
| [0844] | <223> 7064retD正向引物           |
| [0845] | <400> 103                    |
| [0846] | cctctgaggc cacctgaa 18       |
| [0847] | <210> 104                    |
| [0848] | <211> 23                     |
| [0849] | <212> DNA                    |
| [0850] | <213> 人工序列                   |
| [0851] | <220>                        |
| [0852] | <223> 7064retD反向引物           |
| [0853] | <400> 104                    |
| [0854] | ccctgacaag ttctgccttc tac 23 |
| [0855] | <210> 105                    |
| [0856] | <211> 21                     |
| [0857] | <212> DNA                    |
| [0858] | <213> 人工序列                   |
| [0859] | <220>                        |
| [0860] | <223> 7064retD TAQMAN探针      |
| [0861] | <400> 105                    |
| [0862] | tgcccaagcc tctgcagctt t 21   |
| [0863] | <210> 106                    |
| [0864] | <211> 19                     |
| [0865] | <212> DNA                    |
| [0866] | <213> 人工序列                   |
| [0867] | <220>                        |
| [0868] | <223> 7140retU正向引物           |
| [0869] | <400> 106                    |
| [0870] | cccagcatct gacgacacc 19      |
| [0871] | <210> 107                    |
| [0872] | <211> 21                     |
| [0873] | <212> DNA                    |
| [0874] | <213> 人工序列                   |
| [0875] | <220>                        |
| [0876] | <223> 7140retU反向引物           |
| [0877] | <400> 107                    |
| [0878] | gaccactgtg ggcactctgta g 21  |
| [0879] | <210> 108                    |
| [0880] | <211> 27                     |
| [0881] | <212> DNA                    |

[0882] <213> 人工序列  
[0883] <220>  
[0884] <223> 7140retU TAQMAN探针  
[0885] <400> 108  
[0886] ccgagtcctgc tgttactgtt agcatca 27  
[0887] <210> 109  
[0888] <211> 20  
[0889] <212> DNA  
[0890] <213> 人工序列  
[0891] <220>  
[0892] <223> 7140retD正向引物  
[0893] <400> 109  
[0894] cccgacacct tctgagcatg 20  
[0895] <210> 110  
[0896] <211> 21  
[0897] <212> DNA  
[0898] <213> 人工序列  
[0899] <220>  
[0900] <223> 7140retD反向引物  
[0901] <400> 110  
[0902] tgcaggctga gtcaggattt g 21  
[0903] <210> 111  
[0904] <211> 25  
[0905] <212> DNA  
[0906] <213> 人工序列  
[0907] <220>  
[0908] <223> 7140retD TAQMAN探针  
[0909] <400> 111  
[0910] tagtcacgtt ttgtgacacc ccaga 25  
[0911] <210> 112  
[0912] <211> 20  
[0913] <212> DNA  
[0914] <213> 人工序列  
[0915] <220>  
[0916] <223> Folh1 gRNA A DNA靶向区段  
[0917] <400> 112  
[0918] tgaaccaatt gtgtagcctt 20  
[0919] <210> 113  
[0920] <211> 20  
[0921] <212> DNA  
[0922] <213> 人工序列  
[0923] <220>

[0924] <223> Folh1 gRNA A2 DNA靶向区段  
[0925] <400> 113  
[0926] aatagtggta aagcaccatg 20  
[0927] <210> 114  
[0928] <211> 20  
[0929] <212> DNA  
[0930] <213> 人工序列  
[0931] <220>  
[0932] <223> Folh1 gRNA B DNA靶向区段  
[0933] <400> 114  
[0934] gtgtgctaag gatcgaagtc 20  
[0935] <210> 115  
[0936] <211> 20  
[0937] <212> DNA  
[0938] <213> 人工序列  
[0939] <220>  
[0940] <223> Folh1 gRNA C DNA靶向区段  
[0941] <400> 115  
[0942] caccgagatg cttgggtatt 20  
[0943] <210> 116  
[0944] <211> 20  
[0945] <212> DNA  
[0946] <213> 人工序列  
[0947] <220>  
[0948] <223> Folh1 gRNA D DNA靶向区段  
[0949] <400> 116  
[0950] tgtaaccgcc ctgaatgacc 20  
[0951] <210> 117  
[0952] <211> 20  
[0953] <212> DNA  
[0954] <213> 人工序列  
[0955] <220>  
[0956] <223> Folh1 gRNA E DNA靶向区段  
[0957] <400> 117  
[0958] aaaaggcat cataaatccc 20  
[0959] <210> 118  
[0960] <211> 20  
[0961] <212> DNA  
[0962] <213> 人工序列  
[0963] <220>  
[0964] <223> Folh1 gRNA E2 DNA靶向区段  
[0965] <400> 118

[0966] tcaaaaatag tcatacacct 20  
[0967] <210> 119  
[0968] <211> 20  
[0969] <212> DNA  
[0970] <213> 人工序列  
[0971] <220>  
[0972] <223> Folh1 gRNA F DNA靶向区段  
[0973] <400> 119  
[0974] ggtctctagt acattgtaga 20  
[0975] <210> 120  
[0976] <211> 20  
[0977] <212> DNA  
[0978] <213> 人工序列  
[0979] <220>  
[0980] <223> Adamts5 gRNA A DNA靶向区段  
[0981] <400> 120  
[0982] ggtggtggtg ctgacggaca 20  
[0983] <210> 121  
[0984] <211> 20  
[0985] <212> DNA  
[0986] <213> 人工序列  
[0987] <220>  
[0988] <223> Adamts5 gRNA A2 DNA靶向区段  
[0989] <400> 121  
[0990] tatgagatca acactcgcta 20  
[0991] <210> 122  
[0992] <211> 20  
[0993] <212> DNA  
[0994] <213> 人工序列  
[0995] <220>  
[0996] <223> Adamts5 gRNA B DNA靶向区段  
[0997] <400> 122  
[0998] ccaaggactt cccacgtta 20  
[0999] <210> 123  
[1000] <211> 20  
[1001] <212> DNA  
[1002] <213> 人工序列  
[1003] <220>  
[1004] <223> Adamts5 gRNA C DNA靶向区段  
[1005] <400> 123  
[1006] tgcttccctt atgcaagatt 20  
[1007] <210> 124

[1008] <211> 20  
[1009] <212> DNA  
[1010] <213> 人工序列  
[1011] <220>  
[1012] <223> Adamts5 gRNA D DNA靶向区段  
[1013] <400> 124  
[1014] ttaggtaccc tatttgaata 20  
[1015] <210> 125  
[1016] <211> 20  
[1017] <212> DNA  
[1018] <213> 人工序列  
[1019] <220>  
[1020] <223> Adamts5 gRNA E2 DNA靶向区段  
[1021] <400> 125  
[1022] tgcagtgggt gacaggtcca 20  
[1023] <210> 126  
[1024] <211> 20  
[1025] <212> DNA  
[1026] <213> 人工序列  
[1027] <220>  
[1028] <223> Adamts5 gRNA E DNA靶向区段  
[1029] <400> 126  
[1030] agggttatac tgacgttg 20  
[1031] <210> 127  
[1032] <211> 20  
[1033] <212> DNA  
[1034] <213> 人工序列  
[1035] <220>  
[1036] <223> Adamts5 gRNA F DNA靶向区段  
[1037] <400> 127  
[1038] tgtctttcaa ggagggtac 20  
[1039] <210> 128  
[1040] <211> 20  
[1041] <212> DNA  
[1042] <213> 人工序列  
[1043] <220>  
[1044] <223> Dpp4 gRNA A DNA靶向区段  
[1045] <400> 128  
[1046] actagtagac ctgagggtt 20  
[1047] <210> 129  
[1048] <211> 20  
[1049] <212> DNA

[1050] <213> 人工序列  
[1051] <220>  
[1052] <223> Dpp4 gRNA B DNA靶向区段  
[1053] <400> 129  
[1054] gctccagtgt ttaggccttg 20  
[1055] <210> 130  
[1056] <211> 20  
[1057] <212> DNA  
[1058] <213> 人工序列  
[1059] <220>  
[1060] <223> Dpp4 gRNA B2 DNA靶向区段  
[1061] <400> 130  
[1062] ggcaagctga aaacgcatgc 20  
[1063] <210> 131  
[1064] <211> 20  
[1065] <212> DNA  
[1066] <213> 人工序列  
[1067] <220>  
[1068] <223> Dpp4 gRNA C DNA靶向区段  
[1069] <400> 131  
[1070] gtagatcgct ttccactacc 20  
[1071] <210> 132  
[1072] <211> 20  
[1073] <212> DNA  
[1074] <213> 人工序列  
[1075] <220>  
[1076] <223> Dpp4 gRNA D DNA靶向区段  
[1077] <400> 132  
[1078] gaactccact gctcgtgagc 20  
[1079] <210> 133  
[1080] <211> 20  
[1081] <212> DNA  
[1082] <213> 人工序列  
[1083] <220>  
[1084] <223> Dpp4 gRNA E2 DNA靶向区段  
[1085] <400> 133  
[1086] ataggtgggc actattgaag 20  
[1087] <210> 134  
[1088] <211> 20  
[1089] <212> DNA  
[1090] <213> 人工序列  
[1091] <220>

|        |                           |
|--------|---------------------------|
| [1092] | <223> Dpp4 gRNA E DNA靶向区段 |
| [1093] | <400> 134                 |
| [1094] | atgggaaggt ttataccagc 20  |
| [1095] | <210> 135                 |
| [1096] | <211> 20                  |
| [1097] | <212> DNA                 |
| [1098] | <213> 人工序列                |
| [1099] | <220>                     |
| [1100] | <223> Dpp4 gRNA F DNA靶向区段 |
| [1101] | <400> 135                 |
| [1102] | cggtgtaaaa acaacgggaa 20  |
| [1103] | <210> 136                 |
| [1104] | <211> 20                  |
| [1105] | <212> DNA                 |
| [1106] | <213> 人工序列                |
| [1107] | <220>                     |
| [1108] | <223> 图8中的SV 6.1的正向引物     |
| [1109] | <400> 136                 |
| [1110] | ggaatgccaa ggctactgtc 20  |
| [1111] | <210> 137                 |
| [1112] | <211> 20                  |
| [1113] | <212> DNA                 |
| [1114] | <213> 人工序列                |
| [1115] | <220>                     |
| [1116] | <223> 图8中的SV 6.1的反向引物     |
| [1117] | <400> 137                 |
| [1118] | aaaagtctgc tttgggtggt 20  |
| [1119] | <210> 138                 |
| [1120] | <211> 21                  |
| [1121] | <212> DNA                 |
| [1122] | <213> 人工序列                |
| [1123] | <220>                     |
| [1124] | <223> 图8中的SV 6.3的正向引物     |
| [1125] | <400> 138                 |
| [1126] | cttcatgaac ctactcagg a 21 |
| [1127] | <210> 139                 |
| [1128] | <211> 21                  |
| [1129] | <212> DNA                 |
| [1130] | <213> 人工序列                |
| [1131] | <220>                     |
| [1132] | <223> 图8中的SV 6.3的反向引物     |
| [1133] | <400> 139                 |

[1134] tctcggagtc aggatttacc t 21  
[1135] <210> 140  
[1136] <211> 20  
[1137] <212> DNA  
[1138] <213> 人工序列  
[1139] <220>  
[1140] <223> 图8中的SV 7.8的正向引物  
[1141] <400> 140  
[1142] tgtctctttg cctgttgctg 20  
[1143] <210> 141  
[1144] <211> 22  
[1145] <212> DNA  
[1146] <213> 人工序列  
[1147] <220>  
[1148] <223> 图8中的SV 7.8的反向引物  
[1149] <400> 141  
[1150] tctgctctac aaggcttacg tg 22  
[1151] <210> 142  
[1152] <211> 20  
[1153] <212> DNA  
[1154] <213> 人工序列  
[1155] <220>  
[1156] <223> 图8中的SV 16的正向引物  
[1157] <400> 142  
[1158] caaccaggca gacttacagc 20  
[1159] <210> 143  
[1160] <211> 20  
[1161] <212> DNA  
[1162] <213> 人工序列  
[1163] <220>  
[1164] <223> 图8中的SV 16的反向引物  
[1165] <400> 143  
[1166] ggcctaggaa ccagtcaaaa 20  
[1167] <210> 144  
[1168] <211> 23  
[1169] <212> DNA  
[1170] <213> 人工序列  
[1171] <220>  
[1172] <223> 图8中的SV 25.5的正向引物  
[1173] <400> 144  
[1174] gcttactgga aagctacata ggg 23  
[1175] <210> 145

|        |                            |
|--------|----------------------------|
| [1176] | <211> 22                   |
| [1177] | <212> DNA                  |
| [1178] | <213> 人工序列                 |
| [1179] | <220>                      |
| [1180] | <223> 图8中的SV 25.5的反向引物     |
| [1181] | <400> 145                  |
| [1182] | caacaacata gaaaccctg tc 22 |
| [1183] | <210> 146                  |
| [1184] | <211> 6                    |
| [1185] | <212> DNA                  |
| [1186] | <213> 人工序列                 |
| [1187] | <220>                      |
| [1188] | <223> 金黄色葡萄球菌Cas9 PAM序列    |
| [1189] | <220>                      |
| [1190] | <221> 尚未归类的特征              |
| [1191] | <222> (1) ... (2)          |
| [1192] | <223> n = A、T、C或G          |
| [1193] | <220>                      |
| [1194] | <221> 尚未归类的特征              |
| [1195] | <222> (4) ... (5)          |
| [1196] | <223> r = A或G              |
| [1197] | <400> 146                  |
| [1198] | nngrrt 6                   |
| [1199] | <210> 147                  |
| [1200] | <211> 5                    |
| [1201] | <212> DNA                  |
| [1202] | <213> 人工序列                 |
| [1203] | <220>                      |
| [1204] | <223> 金黄色葡萄球菌Cas9 PAM序列    |
| [1205] | <220>                      |
| [1206] | <221> 尚未归类的特征              |
| [1207] | <222> (1) ... (2)          |
| [1208] | <223> n = A、T、C或G          |
| [1209] | <220>                      |
| [1210] | <221> 尚未归类的特征              |
| [1211] | <222> (4) ... (5)          |
| [1212] | <223> r = A或G              |
| [1213] | <400> 147                  |
| [1214] | nngrr 5                    |
| [1215] | <210> 148                  |
| [1216] | <211> 3155                 |
| [1217] | <212> DNA                  |

|        |                                                                    |      |
|--------|--------------------------------------------------------------------|------|
| [1218] | <213> 人工序列                                                         |      |
| [1219] | <220>                                                              |      |
| [1220] | <223> 工程改造的人Vk1-39Jk5基因座                                           |      |
| [1221] | <400> 148                                                          |      |
| [1222] | ggcgcgccgt agctttgaat tttaaacatc tatttgacaa gaaatgcata gttcctttctc | 60   |
| [1223] | tttaaaataa tgtaatgttt ctttcaagaa taagcttggt ttgatgcctc tctccccaac  | 120  |
| [1224] | atgatagaag tgtagcataa atctatgaaa aattccattt cctgtgcct acaacaacta   | 180  |
| [1225] | cctgggattg aaaactttct ccttgctct agtcctttct tctacaccta ctccacatc    | 240  |
| [1226] | atctgtgact caaaacaata ctgtcagga aagatcccgg aaagagcaaa aaagacttcc   | 300  |
| [1227] | ttagagggtg cagagattcc tatgccacta tctgtcatct ctagaagggg ttgtgagtat  | 360  |
| [1228] | gaggaagagc agagcttgta aattttctac ttgctttgac ttccactgta tttcctaaca  | 420  |
| [1229] | acaacaacca cagcaacacc cataacatca caggacaaac ttctagtact tccaaggctt  | 480  |
| [1230] | tagtctcagt aaatcttctc tacctccatc acagcagcta gaaggtttga tactcataca  | 540  |
| [1231] | aatagtactg tagctttctg ttcataattg gaaaaataga caagacccaa tgtaatacag  | 600  |
| [1232] | gctttccttc agccagttag cgttcagttt ttgatcacc attgcacaca tataccagc    | 660  |
| [1233] | atatgtctaa tatatatgta gaaatccgtg aagcaagagt tataatagct tgtgttttct  | 720  |
| [1234] | attgtattgt attttctct tatatcatct tcttcttctg tcattaaaaa aaaaccgttc   | 780  |
| [1235] | aagtaggtct aaattaatta ttgatcata agtagataaa atattttatt tcataacaca   | 840  |
| [1236] | ttgacccgat gaatatgttt ctttgccaga catagtctc atttccaagg taacaagcct   | 900  |
| [1237] | gaaaaaatta tactggagca agtcaacagg taatgatggg agcttttctt tattgtctg   | 960  |
| [1238] | gggcaagaat aagacaaaag ataacagggt agaataaaga ttgtgtaaga aagaaggaca  | 1020 |
| [1239] | gcaacaggac atgggaacct ttatagagt aacattttga taatggatga tgagaattaa   | 1080 |
| [1240] | tgagttagac agggatgggt gggaatgatt gaaggtgtga gtactttagc acagattaag  | 1140 |
| [1241] | accaaactcat taggatttaa agagttgtgt agagttagt aaggaaaagc cttagaatta  | 1200 |
| [1242] | aatttggtg cggtataaac attcttggt tagactgaag actcttttct gtgctaagta    | 1260 |
| [1243] | agtatattta tgataatgat gatgactgta gtgctgaata tttaataaat aaaaacaaaa  | 1320 |
| [1244] | ttaattgccg catacataat gtcctgaata ctattgtaa tgttttatct tatttccttt   | 1380 |
| [1245] | aaactgtcta cagcactata aggtaggtag cagtattgtc acagttacac agatatggaa  | 1440 |
| [1246] | accgagacac agggaagtta agttacttga tcaatttcaa gcaatcggca agccatggag  | 1500 |
| [1247] | catctatgtc agggctgcca ggacatgtga ctgtaaacag aagtttttca ctttttaact  | 1560 |
| [1248] | caaagagggt atgtggctgg gttaatggaa agcttcagga ccctcagaaa acattactaa  | 1620 |
| [1249] | caagcaaatg aaaggtgtat ctggaagatt aagttttaac agactcttca tttccatcga  | 1680 |
| [1250] | tccaataatg cacttaggga gatgactggg catattgagg ataggaagag agaagtgaag  | 1740 |
| [1251] | acacagcttt ttatatgtt cttacaggc ttgtgccaaa catcttctgg gtggatttag    | 1800 |
| [1252] | gtgattgagg agaagaaaga cacaggagcg aaattctctg agcacaaggg aggagttcta  | 1860 |
| [1253] | cactcagact gagccaacag acttttctgg cctgacaacc agggcggcgc aggatgtc    | 1920 |
| [1254] | gtgcagagag gaagaagcag gtggtctttg cagctgaaag ctcagctgat ttgcatatgg  | 1980 |
| [1255] | agtcattata caacatccca gaattcttta agggcagctg ccaggaagct aagaagcatc  | 2040 |
| [1256] | ctctcttcta gctctcagag atggagacag acacactcct gctatgggtg ctgctgctct  | 2100 |
| [1257] | gggttccagg tgagggtaca gataagtgt atgagcaacc tctgtggcca ttatgatgct   | 2160 |
| [1258] | ccatgcctct ctgttcttga tctactaat tagggcattt gtcactggtt ttaagtttcc   | 2220 |
| [1259] | ccagtcacct gaattttcca ttttctcaga gtgatgtcca aaattattct taaaaattta  | 2280 |

|        |                                                                     |                    |
|--------|---------------------------------------------------------------------|--------------------|
| [1260] | aatgaaaagg tcctctgctg tgaaggcttt taaagatata taaaaataat ctttgtgttt   | 2340               |
| [1261] | atcattccag gtgccagatg tgacatccag atgaccagct ctccatcctc cctgtctgca   | 2400               |
| [1262] | tctgtaggag acagagtcac catcacttgc cgggcaagtc agagcattag cagctattta   | 2460               |
| [1263] | aattggtatc agcagaaacc agggaaagcc cctaagctcc tgatctatgc tgcattccagt  | 2520               |
| [1264] | ttgcaaagtg ggggtcccatc aaggttcagt ggcagtgatg ctgggacaga tttcactctc  | 2580               |
| [1265] | accatcagca gtctgcaacc tgaagatttt gcaacttact actgtcaaca gagttacagt   | 2640               |
| [1266] | accctccga tcacctcgg ccaagggaca cgactggaga ttaaactgaa gtaatttttc     | 2700               |
| [1267] | actattgtct tctgaaattt ggggtctgatg gccagtattg acttttagag gcttaaatag  | 2760               |
| [1268] | gagtttggtg aagattggta aatgagggca tttaagattt gccatgggtt gcaaaagtta   | 2820               |
| [1269] | aactcagctt caaaaatgga tttggagaaa aaaagattaa attgctctaa actgaatgac   | 2880               |
| [1270] | acaaagtaaa aaaaaaagt gtaactaaaa aggaaccctt gtatttctaa ggagcaaaag    | 2940               |
| [1271] | taaatttatt tttgttcaact cttgccaaat attgtattgg ttgttgctga ttatgcatga  | 3000               |
| [1272] | tacagaaaag tggaaaaata catttttttag tctttctccc ttttgtttga taaattattt  | 3060               |
| [1273] | tgtcagacaa caataaaaat caatagcacg ccctaagatc tagatgcatg ctcgagtgcc   | 3120               |
| [1274] | atttcattac ctctttctcc gcacccgaca tagat                              | 3155               |
| [1275] | <210>                                                               | 149                |
| [1276] | <211>                                                               | 3166               |
| [1277] | <212>                                                               | DNA                |
| [1278] | <213>                                                               | 人工序列               |
| [1279] | <220>                                                               |                    |
| [1280] | <223>                                                               | 工程改造的人Vk3-20Jk1基因座 |
| [1281] | <400>                                                               | 149                |
| [1282] | ggcgcgccgt agctttgaat tttaaaccatc tatttgacaa gaaatgcata gttccttctc  | 60                 |
| [1283] | tttaaaataa tgtaatgttt ctttcaagaa taagcttggg ttgatgcctc tctcccaac    | 120                |
| [1284] | atgatagaag tgtagcataa atctatgaaa aattccattt ccctgtgcct acaacaacta   | 180                |
| [1285] | cctgggattg aaaacttctt cccttgctct agtcctttct tctacaccta cttccacatc   | 240                |
| [1286] | atctgtgact caaaacaata cttgtcagga aagatcccgg aaagagcaaa aaagacttcc   | 300                |
| [1287] | ttagaggtgt cagagattcc tatgccacta tctgtcatct ctagaagggg ttgtgagtat   | 360                |
| [1288] | gaggaagagc agagcttgta aattttctac ttgctttgac ttccactgta tttcctaaca   | 420                |
| [1289] | acaacaacca cagcaacacc cataacatca caggacaaac ttctagtact tccaaggctt   | 480                |
| [1290] | tagtctcagt aaatcttctc tacctccatc acagcagcta gaaggtttga tactcataca   | 540                |
| [1291] | aatagtactg tagctttctg ttcataattg gaaaaataga caagacccaa tgtaatacag   | 600                |
| [1292] | gctttccttc agccagttag cgttcagttt ttggatcacc attgcacaca tataccagc    | 660                |
| [1293] | atatgtctaa tatatatgta gaaatccgtg aagcaagagt tataatagct tgtgttttct   | 720                |
| [1294] | attgtattgt attttctct tatatcatct tcttcttcgt tcattaaaaa aaaaccgttc    | 780                |
| [1295] | aagtaggtct aaattaatta ttggatcata agtagataaa atattttatt tcataacaca   | 840                |
| [1296] | ttgaccgat gaatatgttt ctttgccaga catagtctc atttccaagg taacaagcct     | 900                |
| [1297] | gaaaaaatta tactggagca agtcaacagg taatgatggg agcttttcct tattgtctg    | 960                |
| [1298] | gggcaagaat aagacaaaag ataacagggt agaataaaga ttgtgtaaga aagaaggaca   | 1020               |
| [1299] | gcaacaggac atgggaacct tttatagagt aacattttga taatggatga tgagaattaa   | 1080               |
| [1300] | tgagttagac agggatgggt gggaatgatt gaaggtgtga gtactttagc acagattaag   | 1140               |
| [1301] | accaaactcat taggattttaa agagttgtgt agagttagtg aaggaaaagc cttagaatta | 1200               |

|        |                                                                    |           |
|--------|--------------------------------------------------------------------|-----------|
| [1302] | aatttgctg cggataaaac attcttggat tagactgaag actcttttct gtgctaagta   | 1260      |
| [1303] | agtatatatta tgataatgat gatgactgta gtgctgaata tttataaat aaaaacaaaa  | 1320      |
| [1304] | ttaattgccg catacataat gtctgaata ctattgtaaa tgttttatct tatttccttt   | 1380      |
| [1305] | aaactgtcta cagcactata aggtaggtac cagtattgtc acagttacac agatatggaa  | 1440      |
| [1306] | accgagacac agggaaagta agttacttga tcaatttcaa gcaatcggca agccatggag  | 1500      |
| [1307] | catctatgtc agggctgccca ggacatgtga ctgtaaacag aagtttttca ctttttaact | 1560      |
| [1308] | caaagagggt atgtggctgg gttaatggaa agcttcagga ccttcagaaa acattactaa  | 1620      |
| [1309] | caagcaaatg aaaggtgtat ctggaagatt aagttttaac agactcttca tttccatcga  | 1680      |
| [1310] | tccaataatg cacttaggga gatgactggg catattgagg ataggaagag agaagtgaaa  | 1740      |
| [1311] | acacagcttt ttatatgtt cttaacaggc ttgtgccaaa catcttctgg gtggatttag   | 1800      |
| [1312] | gtgattgagg agaagaaaga cacaggagcg aaattctctg agcacaaggg aggagtctta  | 1860      |
| [1313] | cactcagact gagccaacag acttttctgg cctgacaacc agggcggcgc aggatgctca  | 1920      |
| [1314] | gtgcagagag gaagaagcag gtggtctttg cagctgaaag ctcagctgat ttgcatatgg  | 1980      |
| [1315] | agtcattata caacatccca gaattcttta agggcagctg ccaggaagct aagaagcatc  | 2040      |
| [1316] | ctctcttcta gctctcagag atggagacag acacactcct gctatgggtg ctgctgctct  | 2100      |
| [1317] | gggttccagg tgagggtaca gataagtgtt atgagcaacc tctgtggcca ttatgatgct  | 2160      |
| [1318] | ccatgcctct ctgttcttga tcactataat tagggcattt gtcactggtt ttaagtttcc  | 2220      |
| [1319] | ccagtcccct gaattttcca ttttctcaga gtgatgtcca aaattattct taaaaattta  | 2280      |
| [1320] | aatgaaaagg tcctctgctg tgaaggcttt taaagatata taaaaataat ctttgtgtt   | 2340      |
| [1321] | atcattccag gtgccagatg tataccaccg gagaaattgt gttgacgcag tctccaggca  | 2400      |
| [1322] | ccctgtcttt gtctccaggg gaaagagcca ccctctcctg cagggccagt cagagtgtta  | 2460      |
| [1323] | gcagcagcta cttagcctgg taccagcaga aacctggcca ggctcccagg ctctcatct   | 2520      |
| [1324] | atggtgcac cagcagggcc actggcatcc cagacaggtt cagtggcagt gggctctggga  | 2580      |
| [1325] | cagacttcac tctcaccatc agcagactgg agcctgaaga ttttgcagtg tattactgtc  | 2640      |
| [1326] | agcagtatgg tagctcacct tggacgttcg gccaaaggac caaggtggaa atcaaacgta  | 2700      |
| [1327] | agtaattttt cactattgtc ttctgaaatt tgggtctgat ggccagtatt gactttttaga | 2760      |
| [1328] | ggcttaaata ggagtttggg aaagattggg aaatgagggc atttaagatt tgccatgggt  | 2820      |
| [1329] | tgcaaaagt aaactcagct tcaaaaatgg atttggagaa aaaaagatta aattgctcta   | 2880      |
| [1330] | aactgaatga cacaaagtaa aaaaaaaaag tgtaactaaa aaggaaccct tgtatttcta  | 2940      |
| [1331] | aggagcaaaa gtaaatattt tttgttcac tcttgccaaa tattgtattg gttgttgctg   | 3000      |
| [1332] | attatgcatg atacagaaaa gtggaaaaat acatttttta gtctttctcc cttttgtttg  | 3060      |
| [1333] | ataaattatt ttgtcagaca acaataaaaa tcaatagcac gccctaagat ctatgcat    | 3120      |
| [1334] | gctcgagtgc catttcatta cctctttctc cgcacccgac atagat                 | 3166      |
| [1335] | <210>                                                              | 150       |
| [1336] | <211>                                                              | 82        |
| [1337] | <212>                                                              | DNA       |
| [1338] | <213>                                                              | 人工序列      |
| [1339] | <220>                                                              |           |
| [1340] | <223>                                                              | 引导RNA骨架v1 |
| [1341] | <400>                                                              | 150       |
| [1342] | gttgaacca ttcaaacag catagcaagt taaaataagg ctagtccgtt atcaacttga    | 60        |
| [1343] | aaaagtggca ccgagtcggt gc                                           | 82        |

---

[1344] <210> 151  
[1345] <211> 76  
[1346] <212> DNA  
[1347] <213> 人工序列  
[1348] <220>  
[1349] <223> 引导RNA骨架v2  
[1350] <400> 151  
[1351] gtttttagagc tagaaatagc aagttaaaat aaggctagtc cgttatcaac ttgaaaaagt 60  
[1352] ggcaccgagt cgggtgc 76  
[1353] <210> 152  
[1354] <211> 86  
[1355] <212> DNA  
[1356] <213> 人工序列  
[1357] <220>  
[1358] <223> 引导RNA骨架v3  
[1359] <400> 152  
[1360] gtttaagagc tatgctggaa acagcatagc aagtttaaat aaggctagtc cgttatcaac 60  
[1361] ttgaaaaagt ggcaccgagt cgggtgc 86

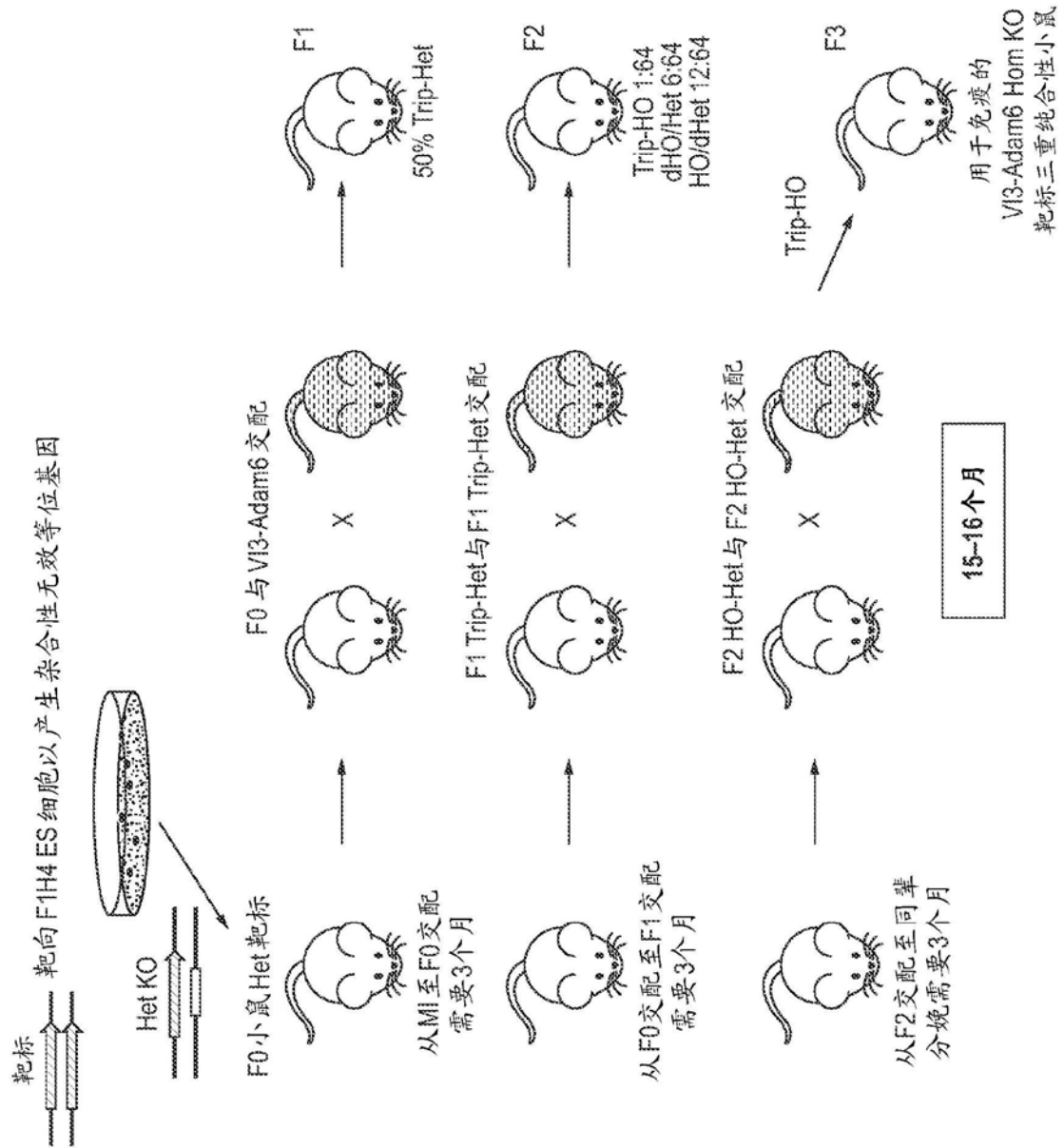


图1

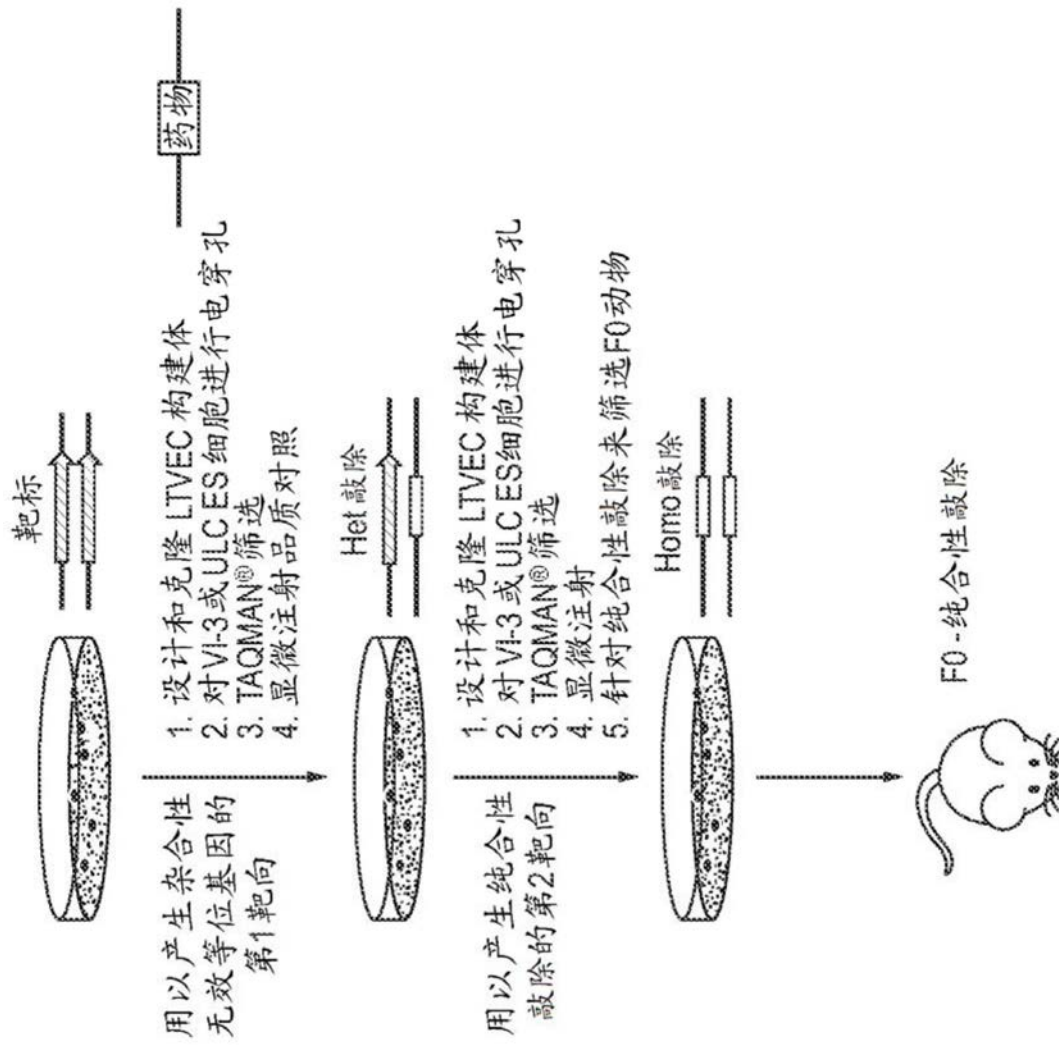


图2

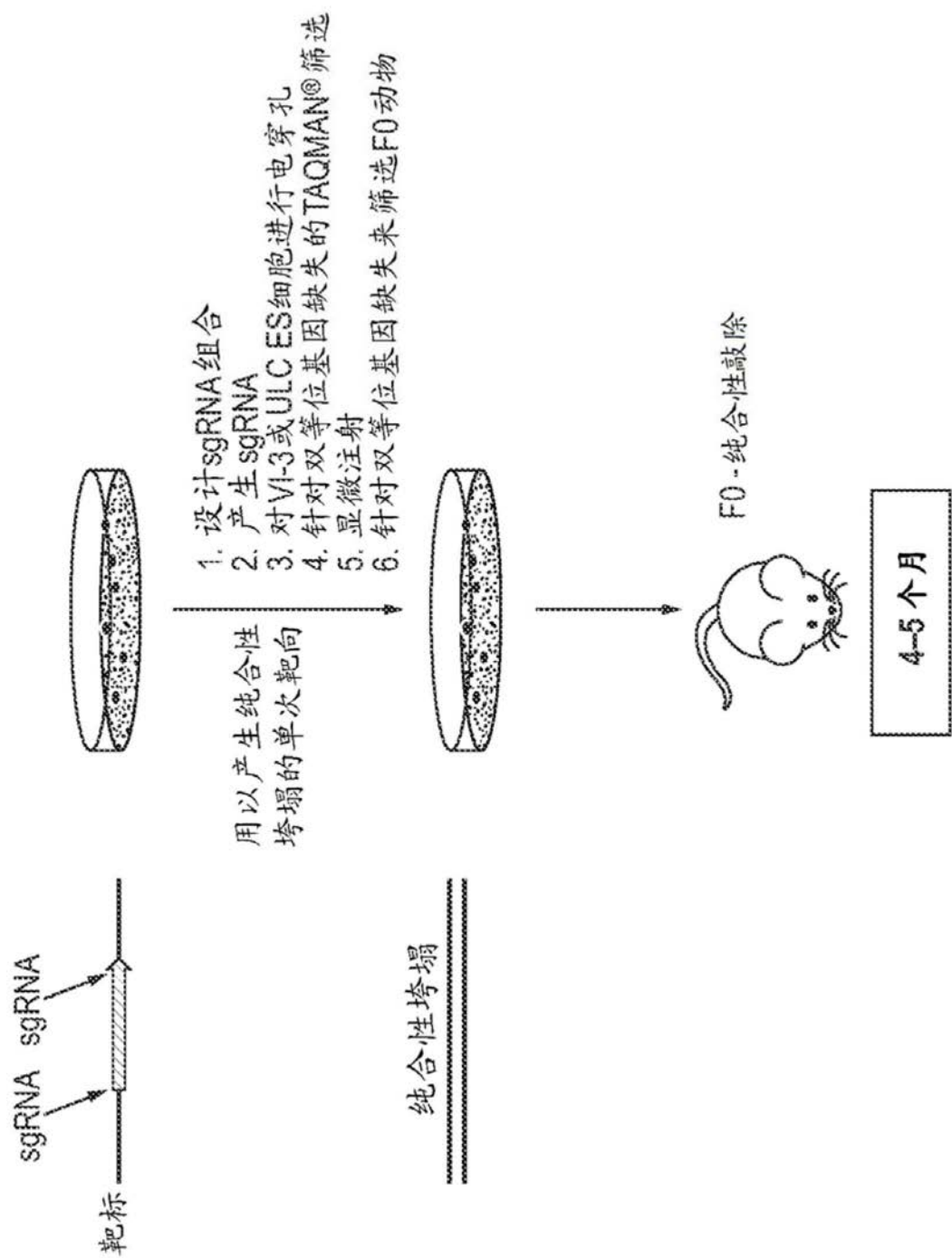


图3



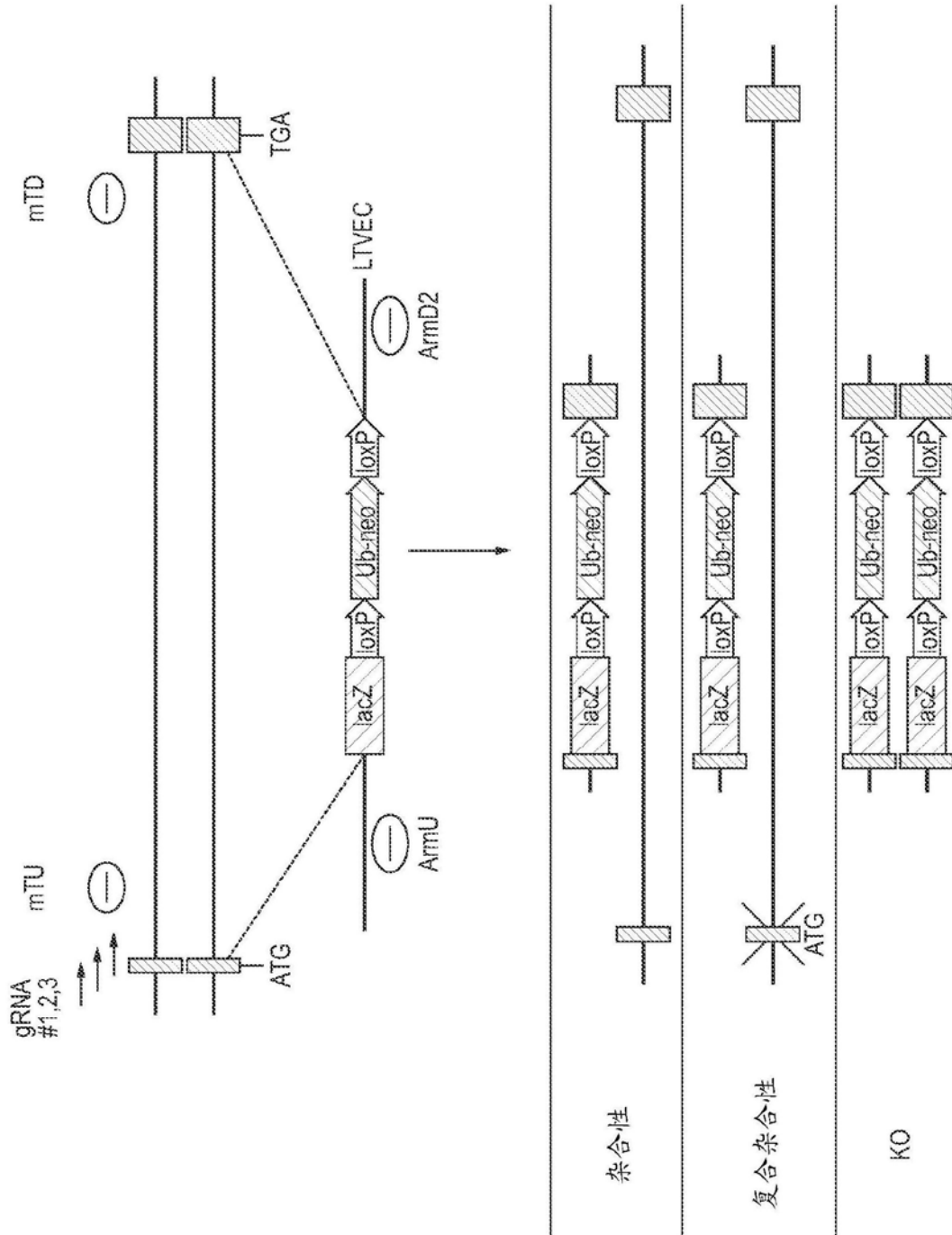


图5

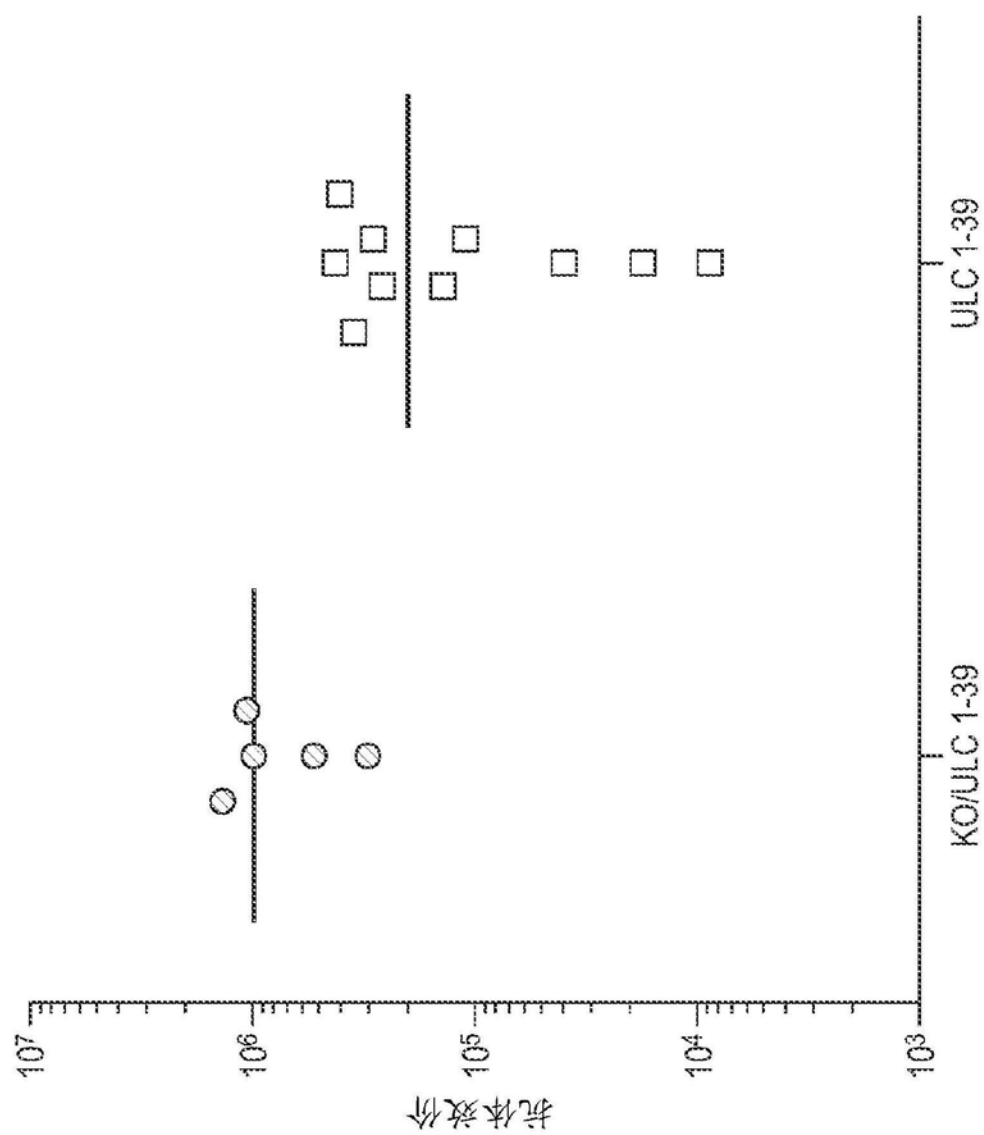


图6

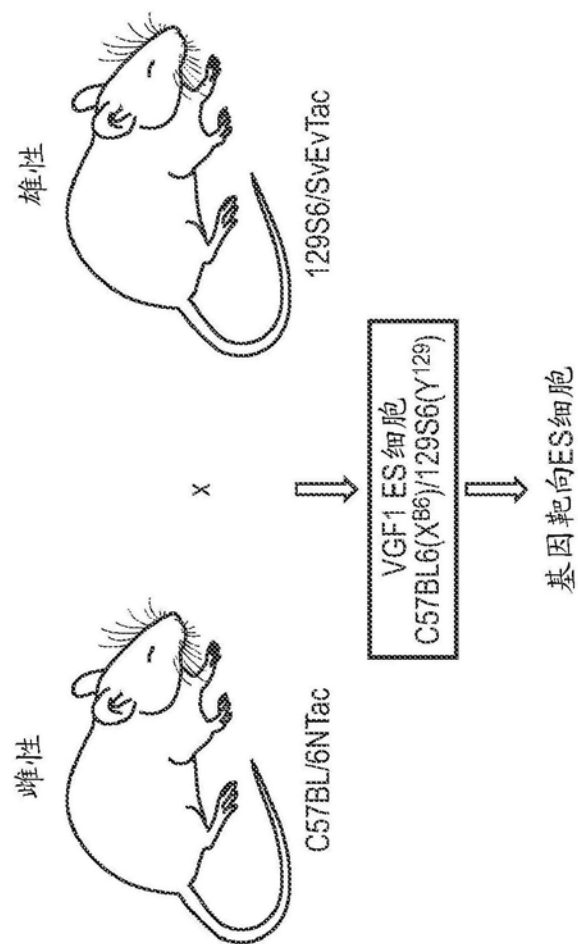


图7

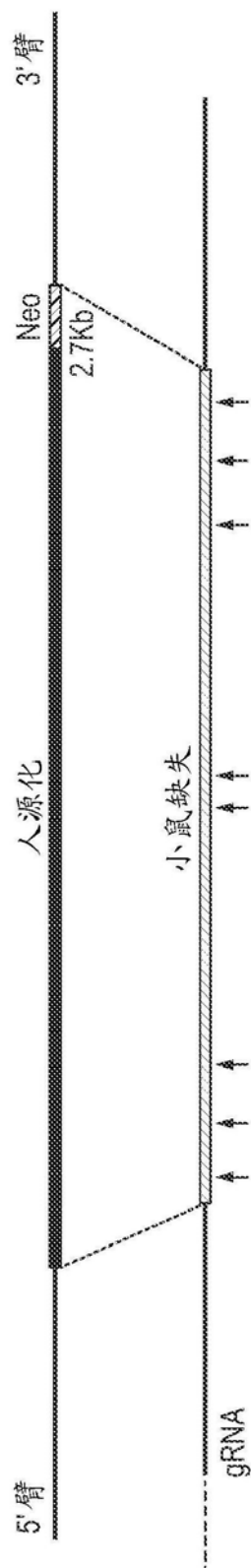


图8

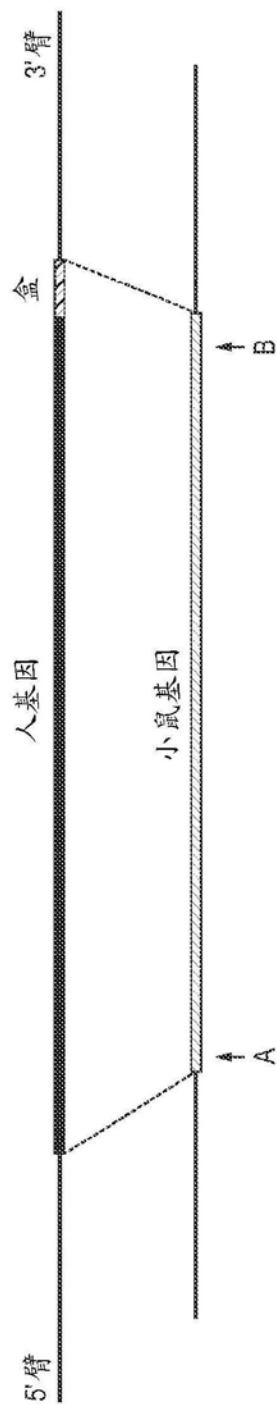


图9A



图9B



图9C



图9D



图9E

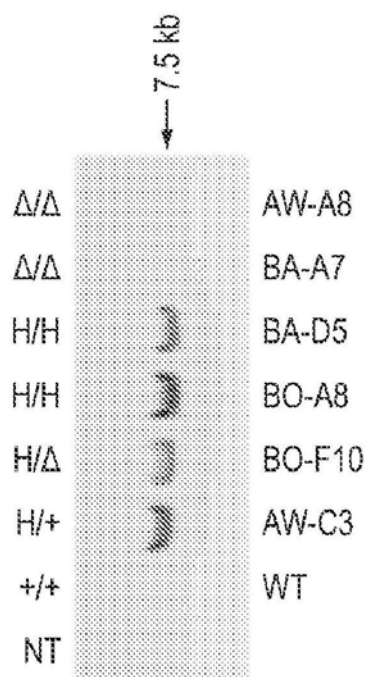


图10A

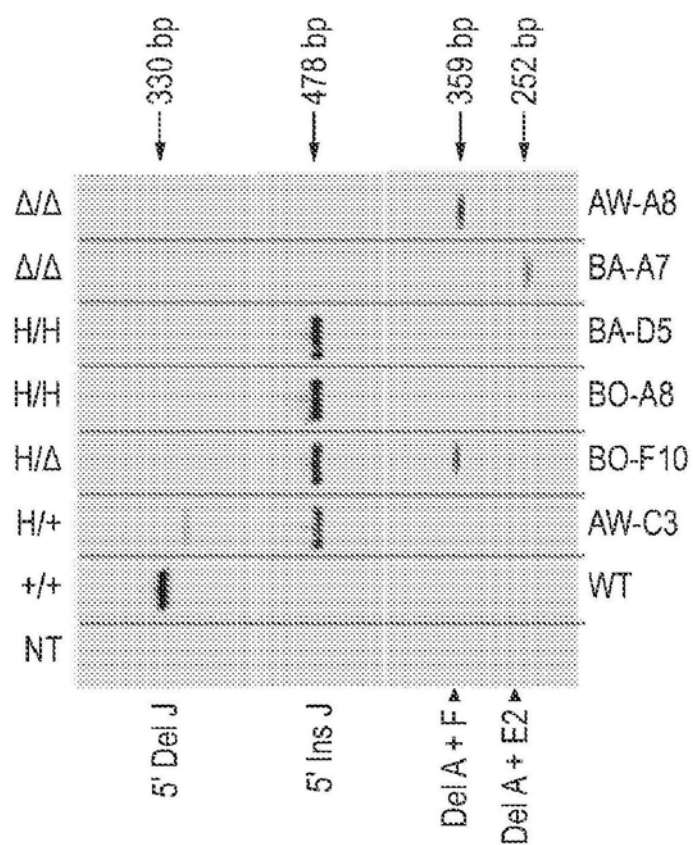


图10B

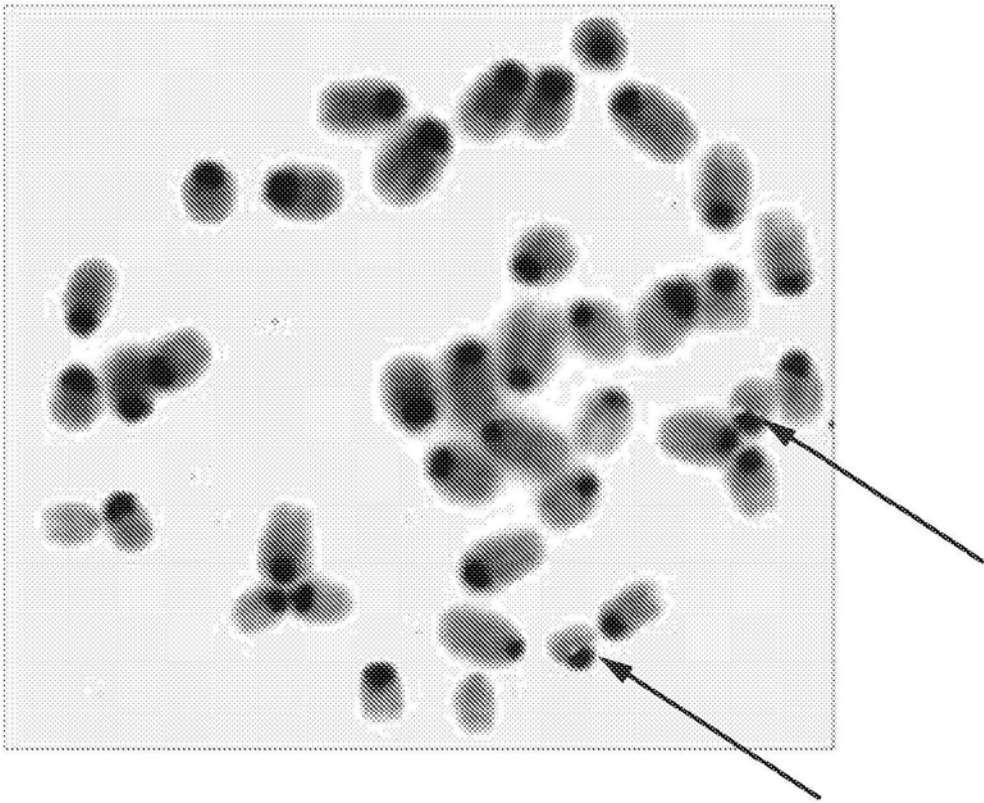


图11A

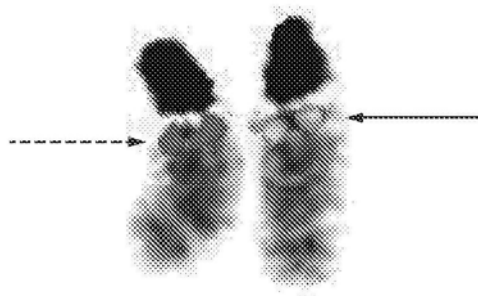


图11B



图11C

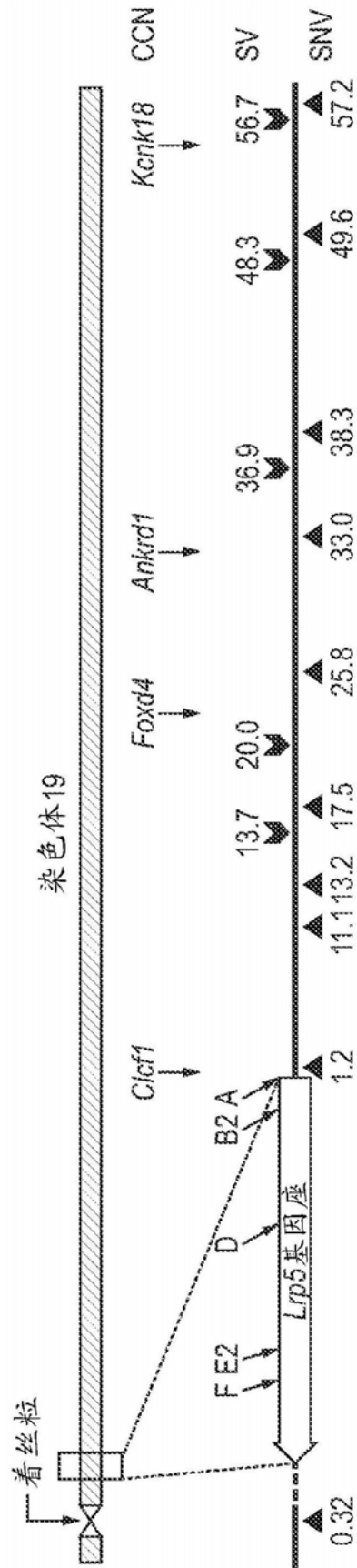


图12

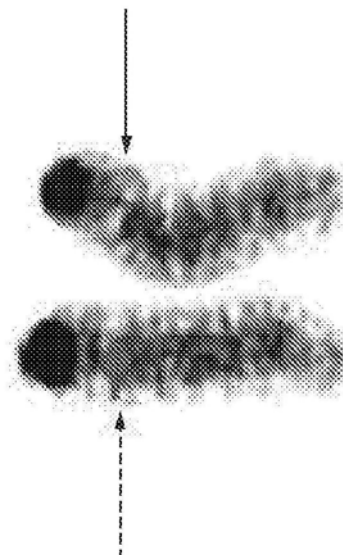


图13A

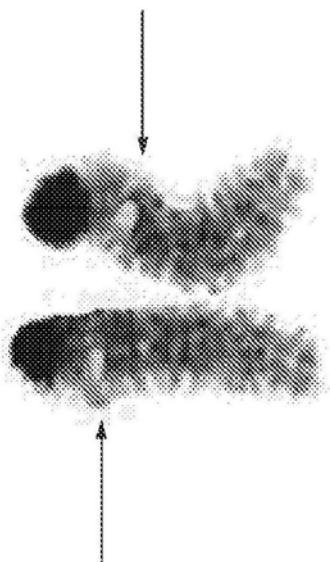


图13B

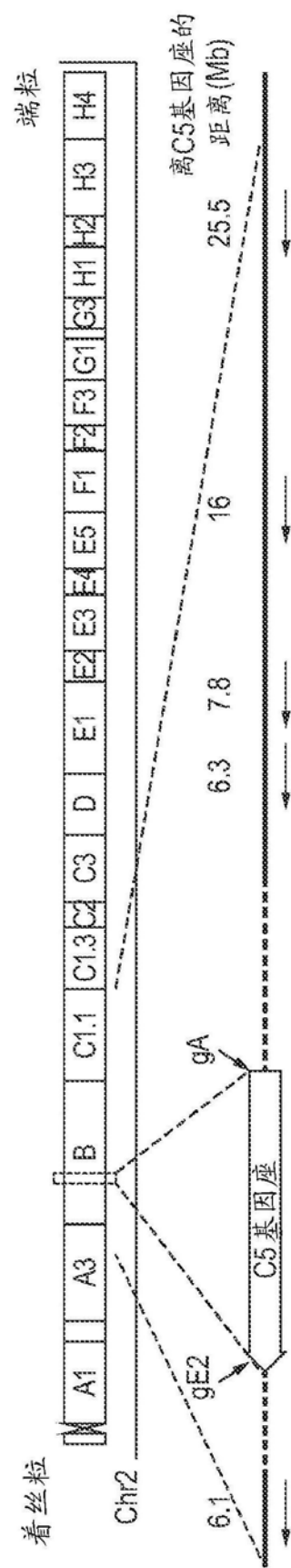
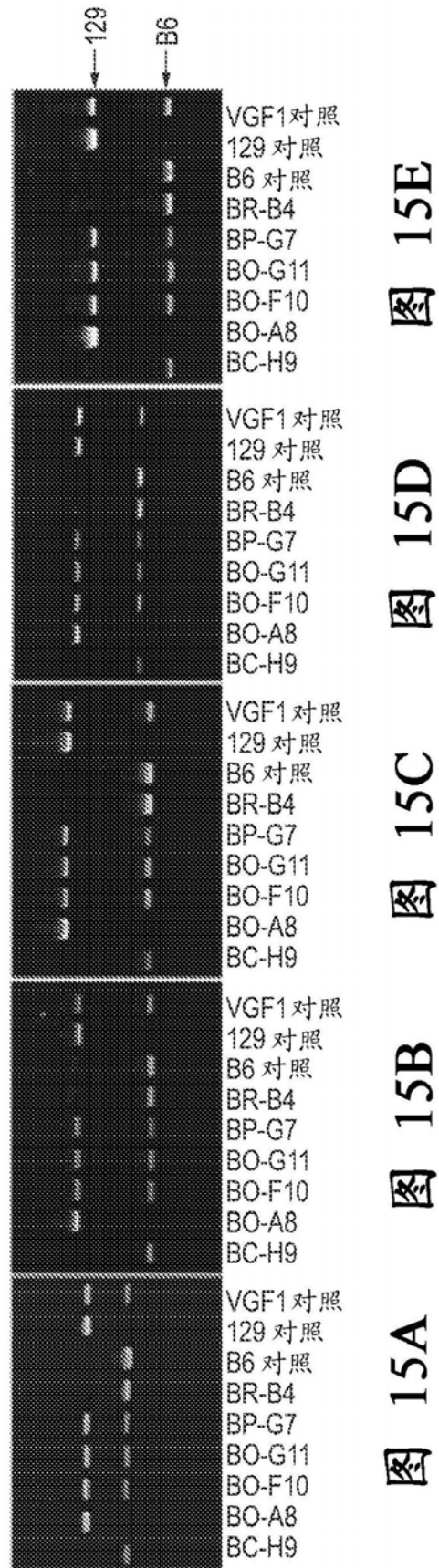


图14



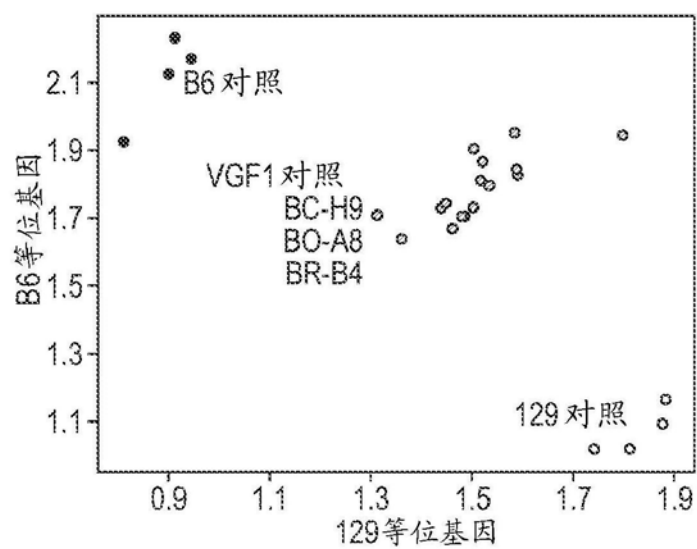


图16A

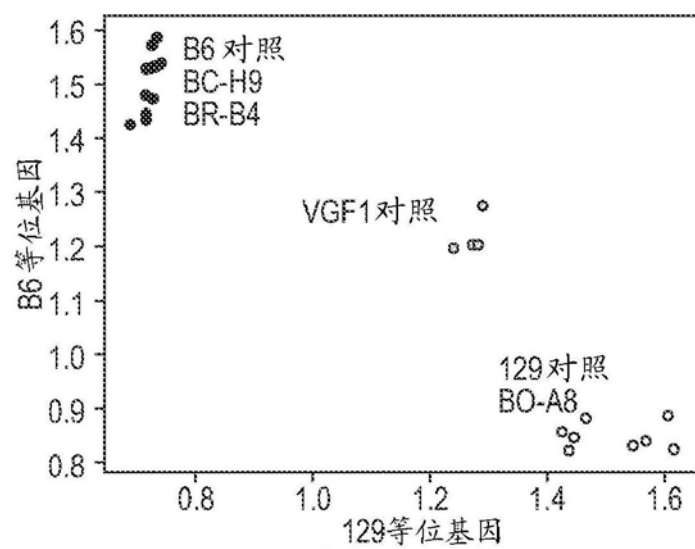


图16B

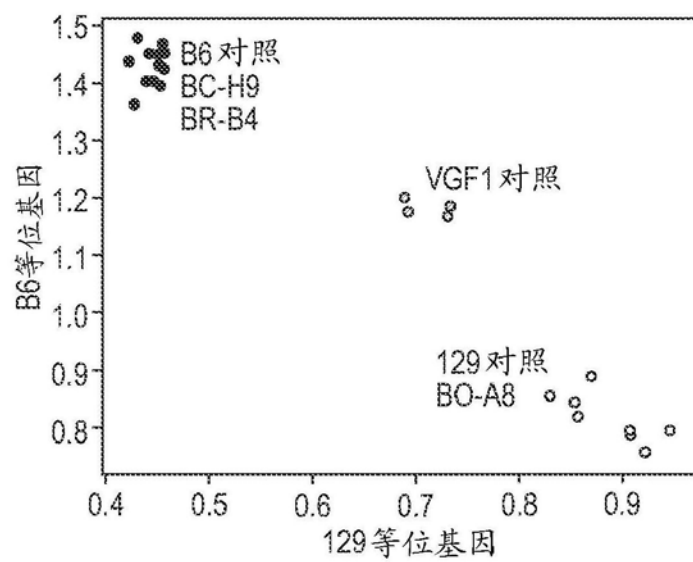


图16C

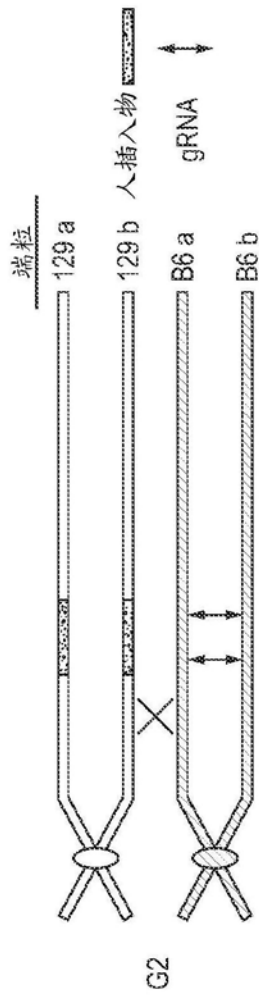


图 17A

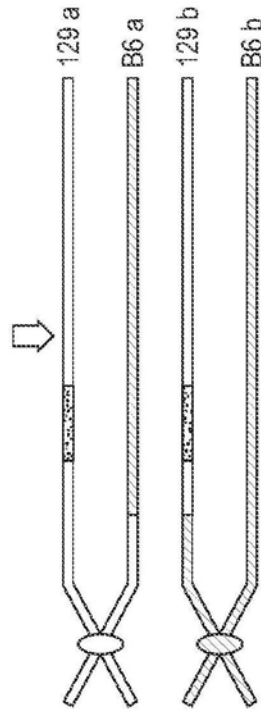


图 17B

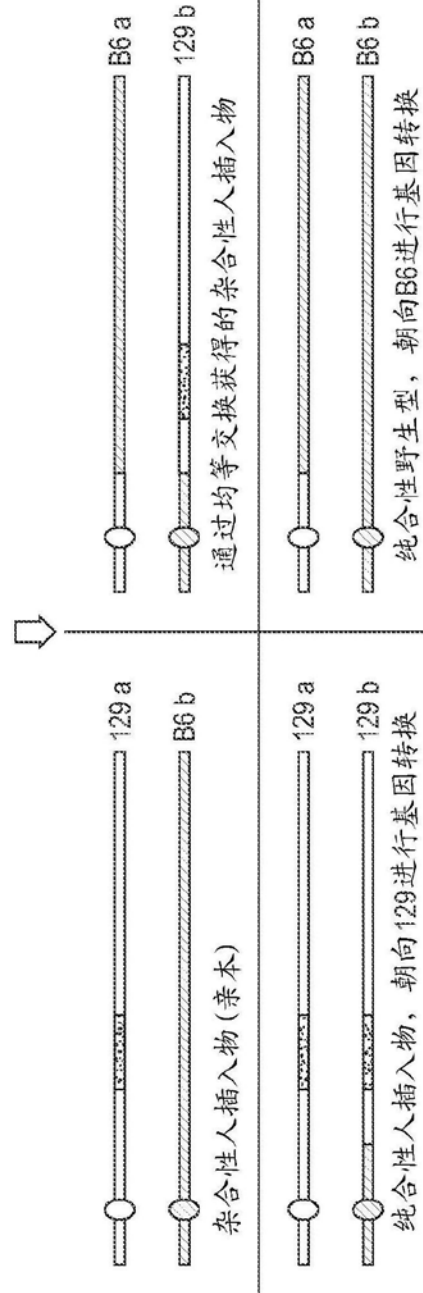


图 17C

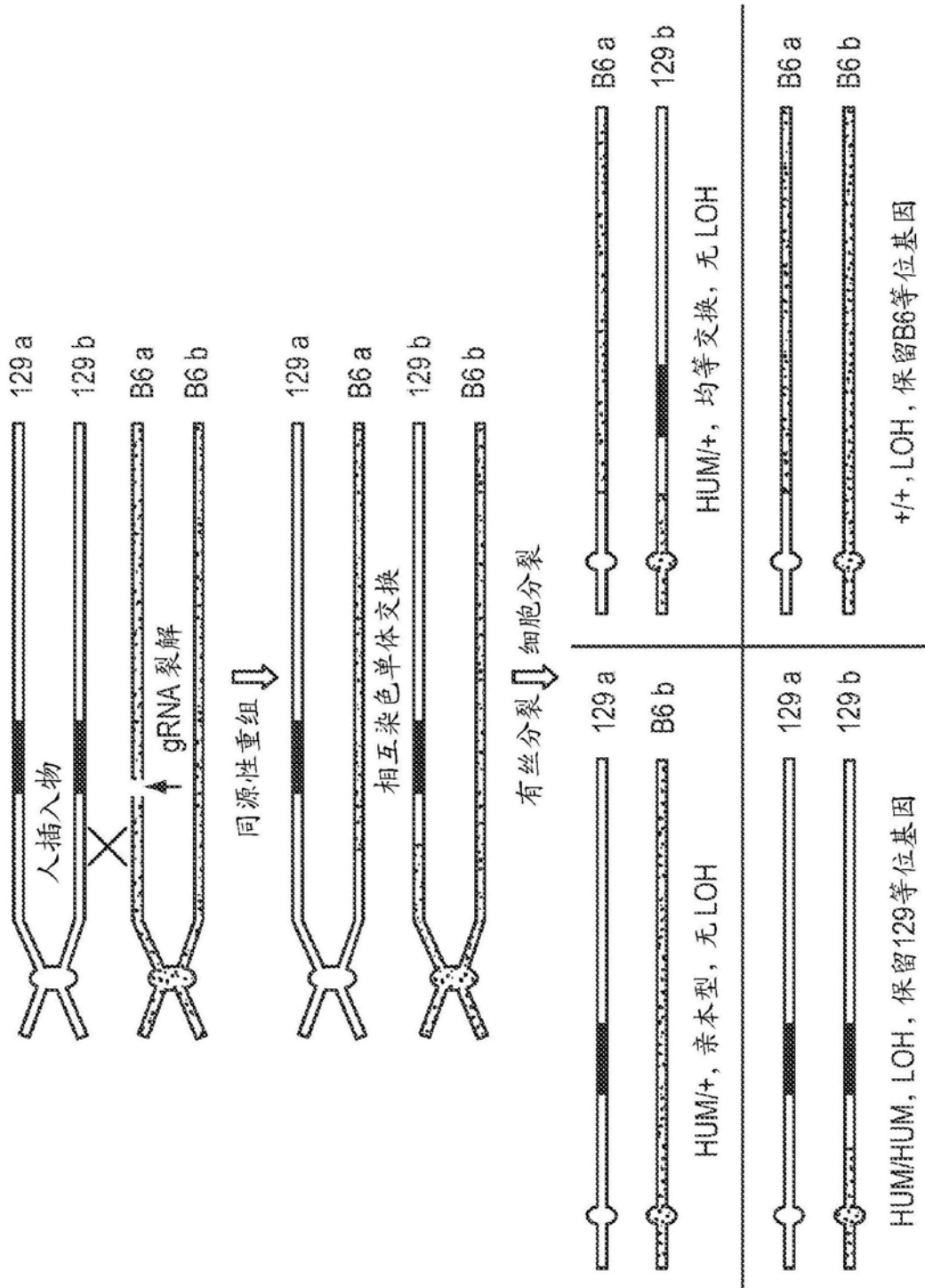


图18A

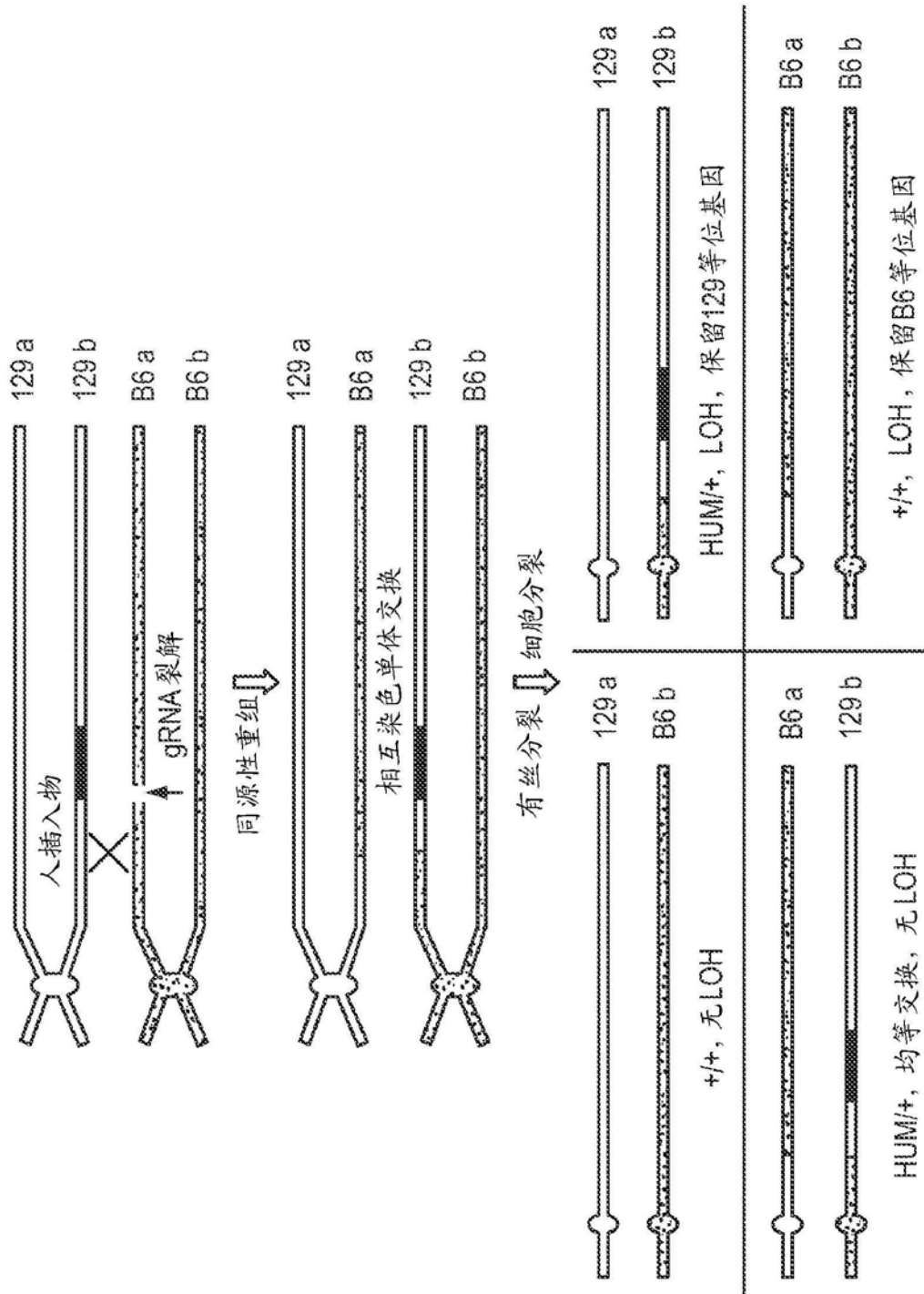


图18B

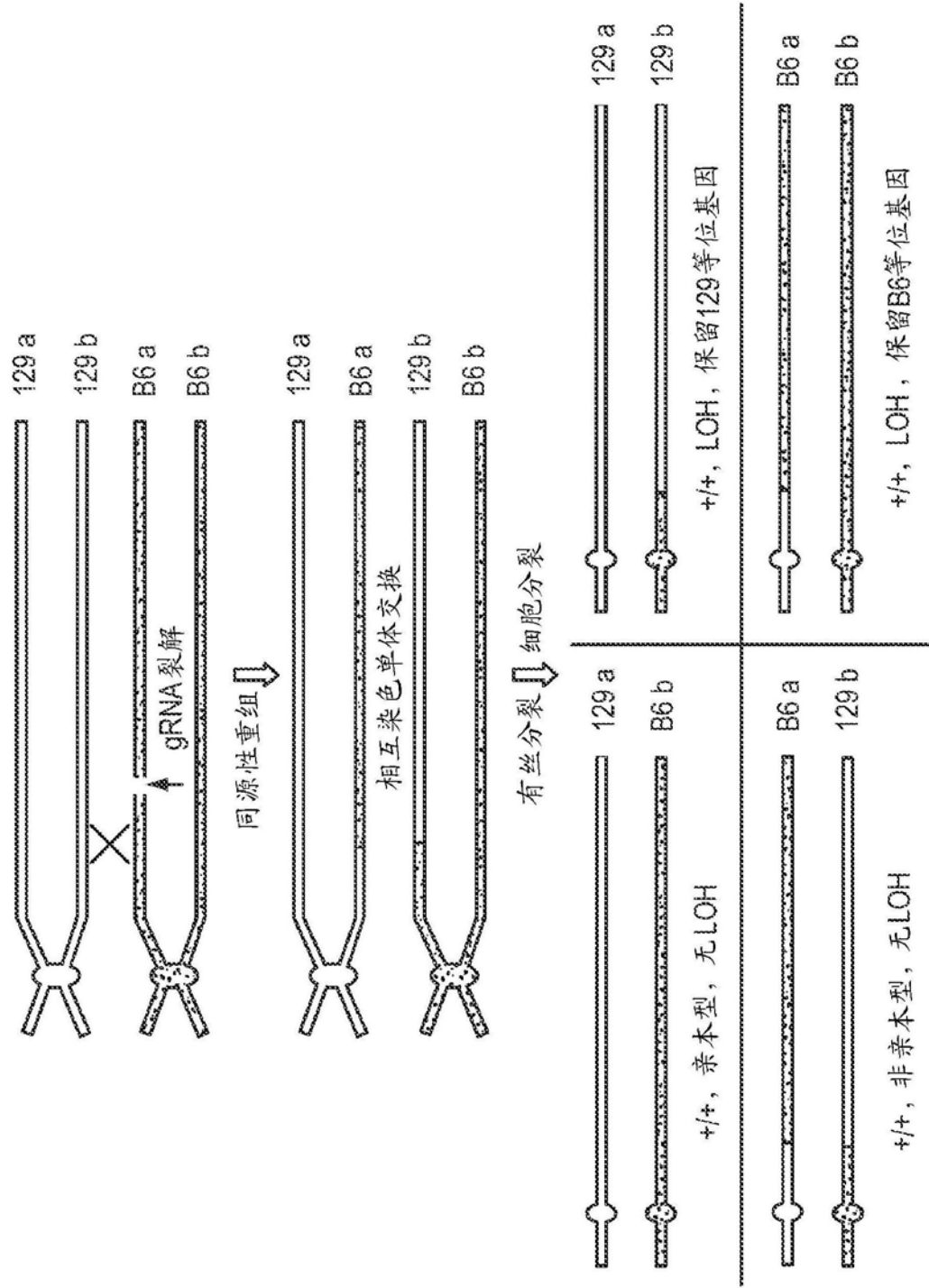


图18C

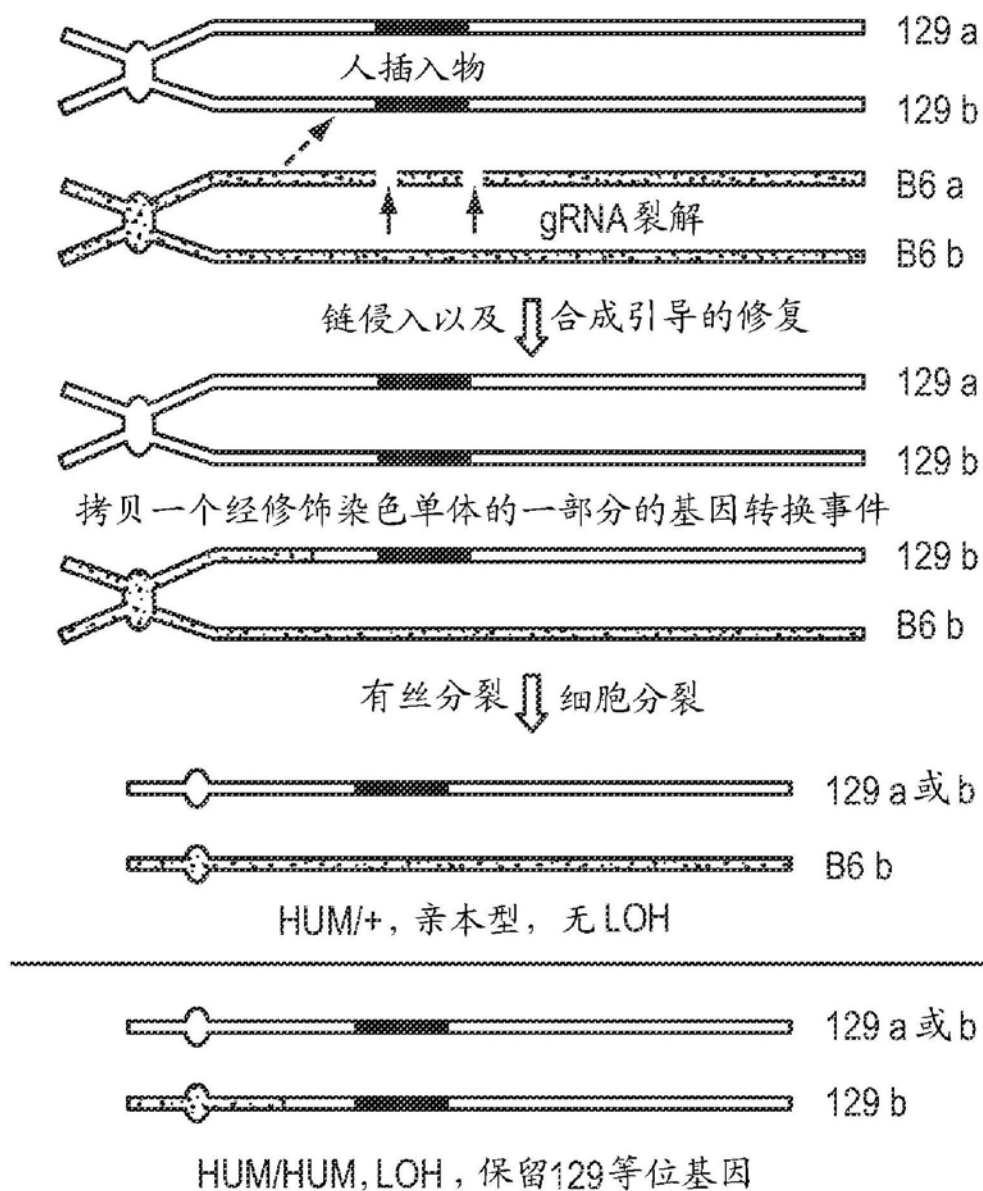


图18D

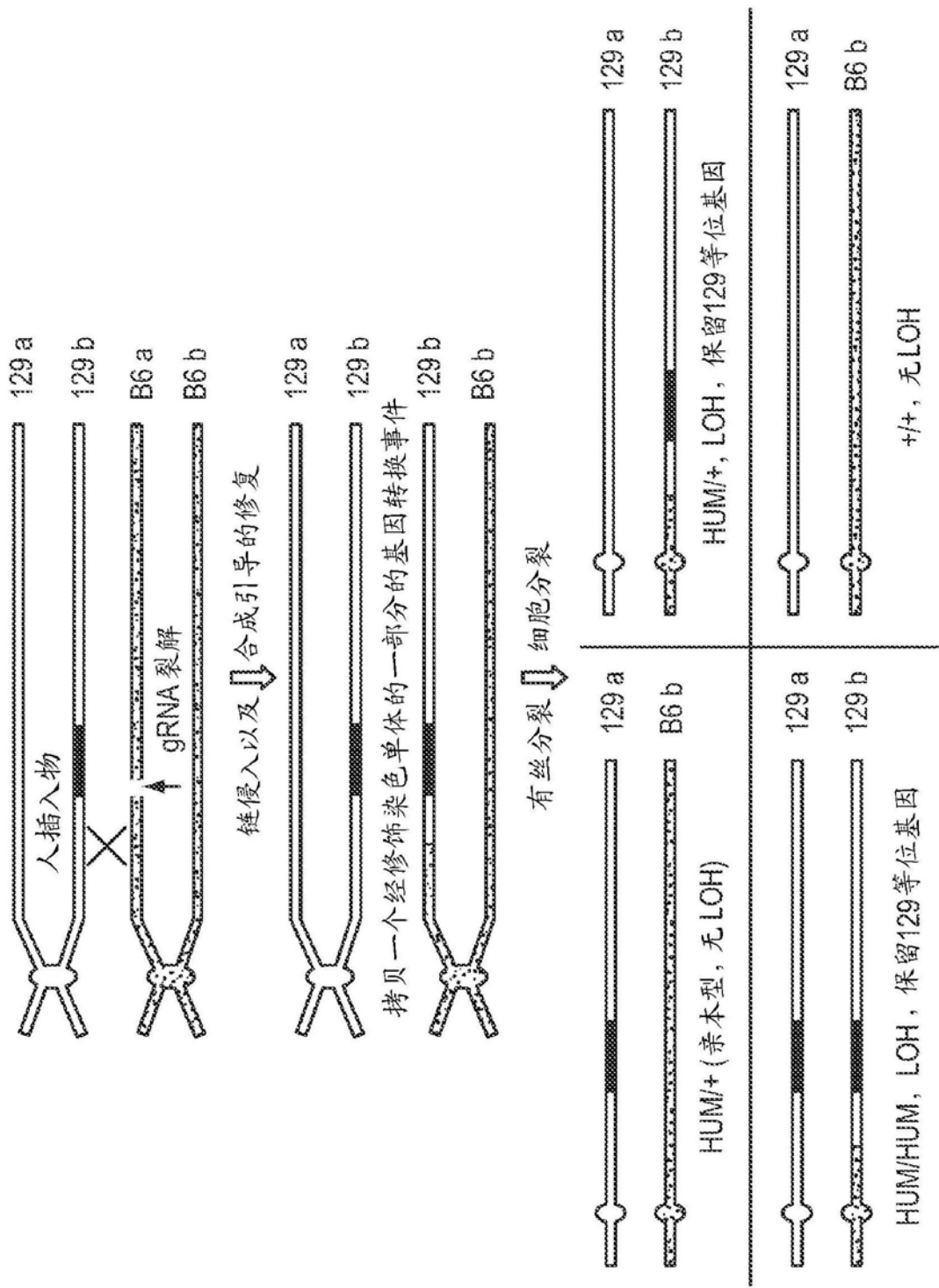


图18E

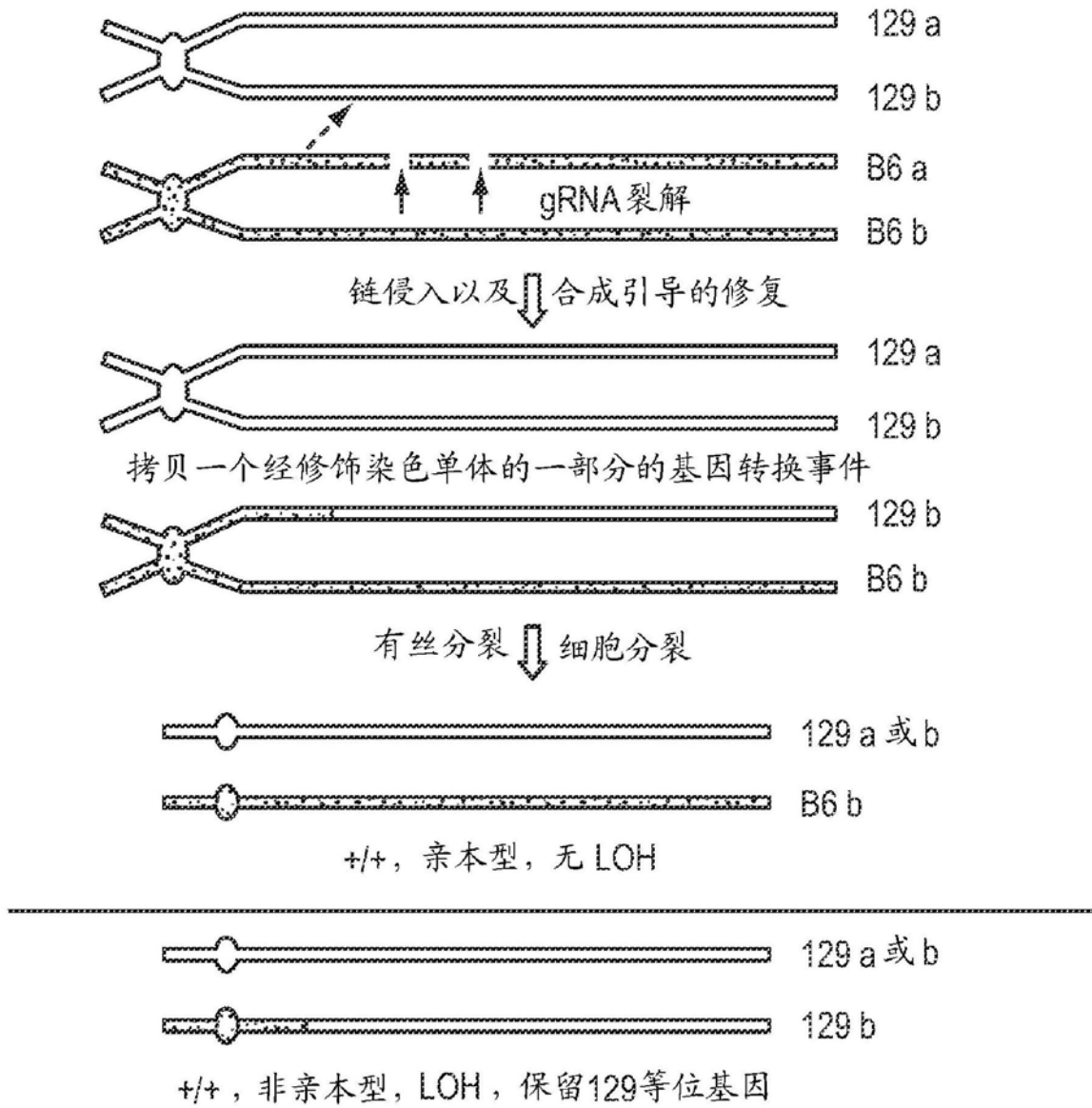


图18F

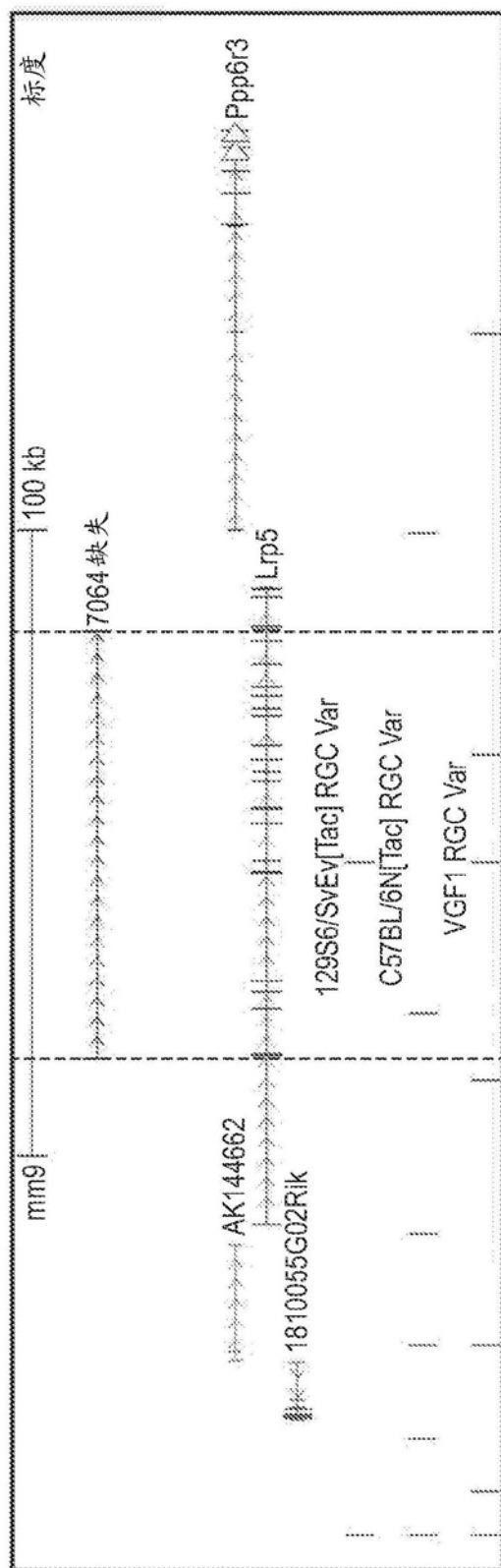


图19

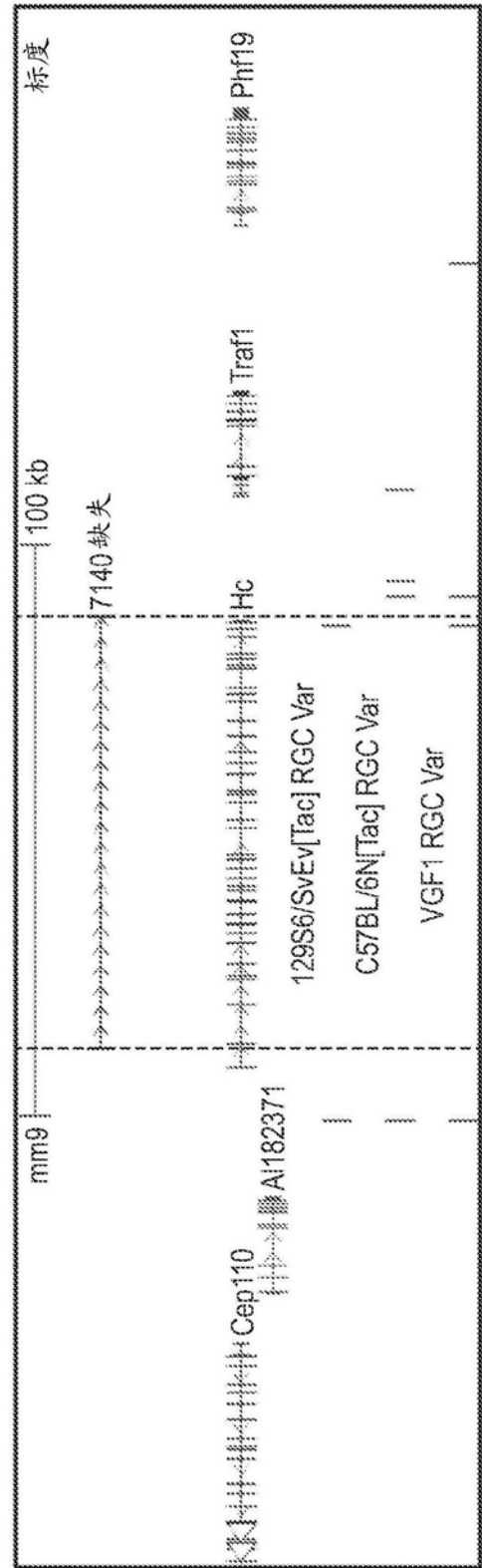


图20

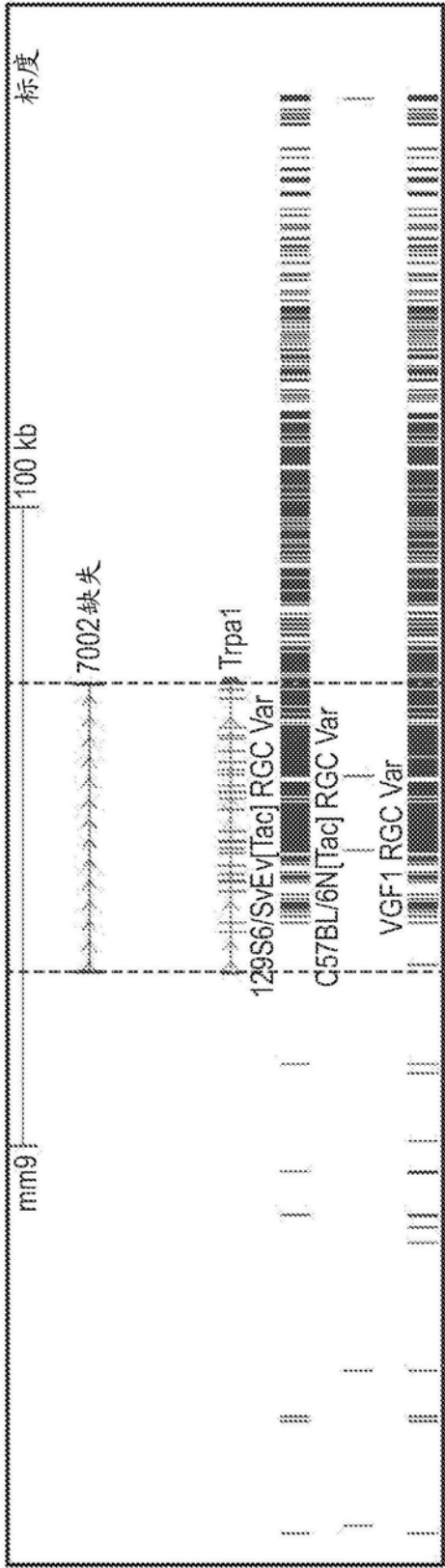


图21

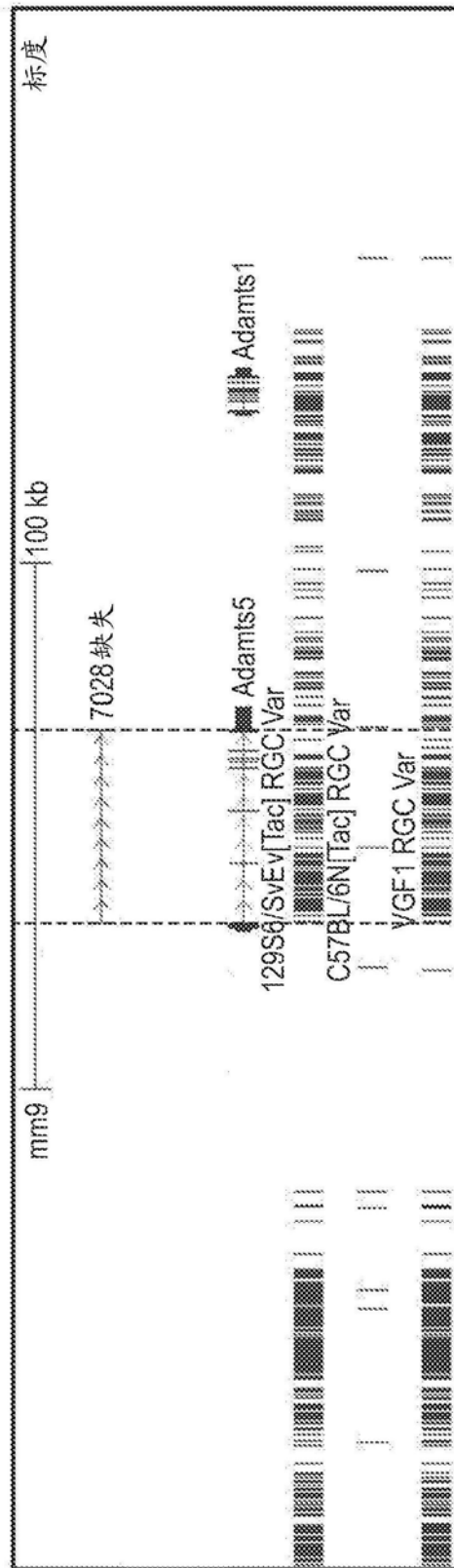


图22

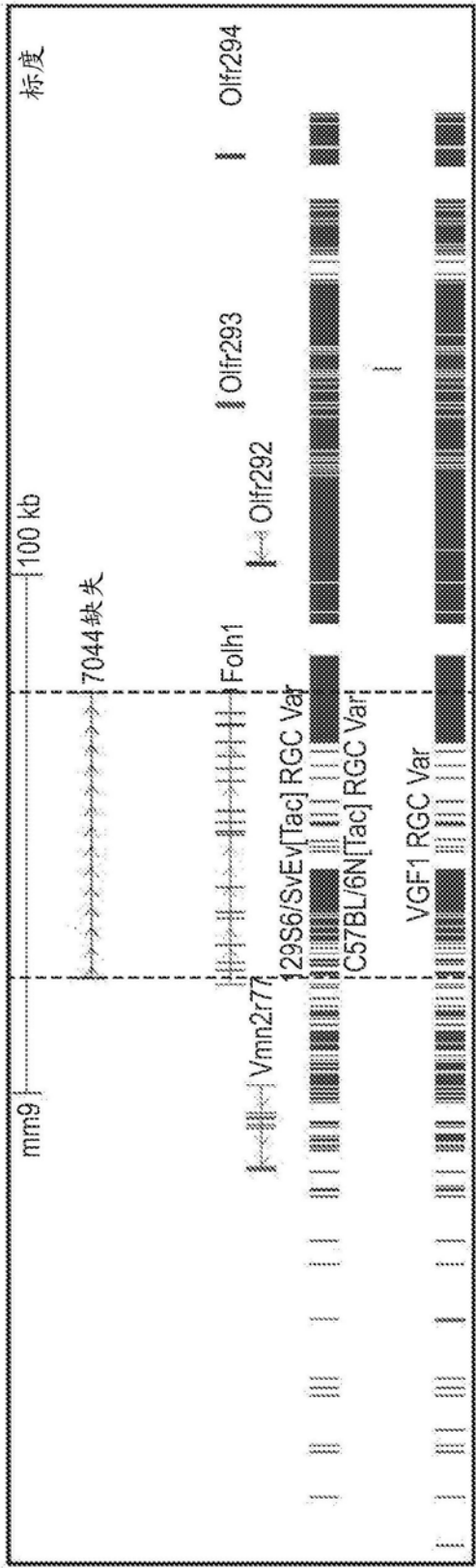


图23

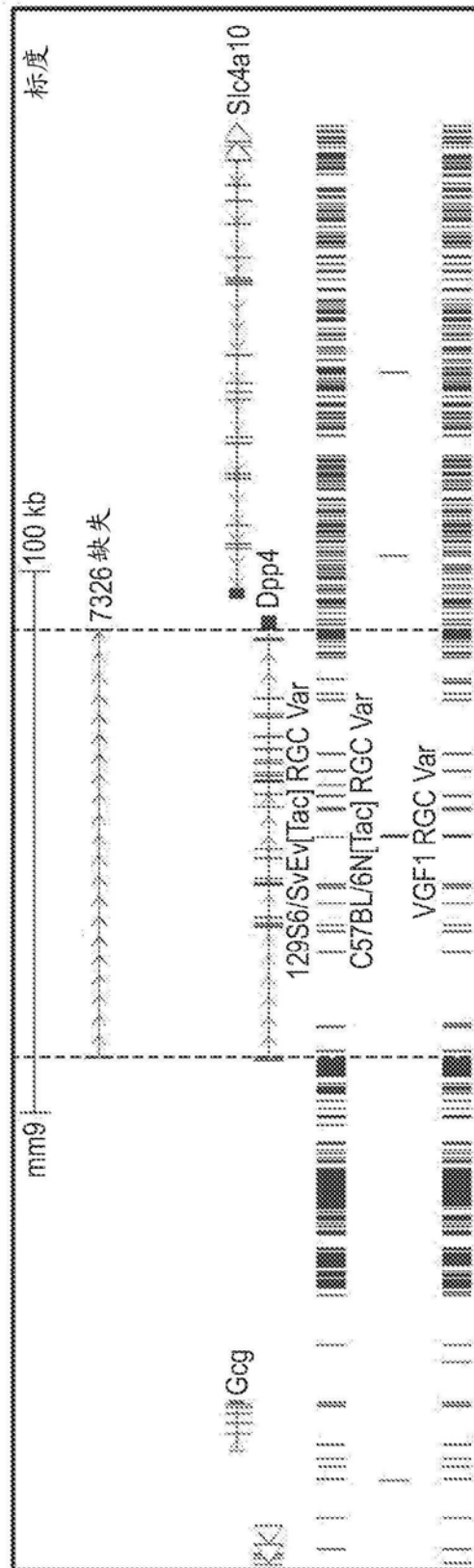


图24

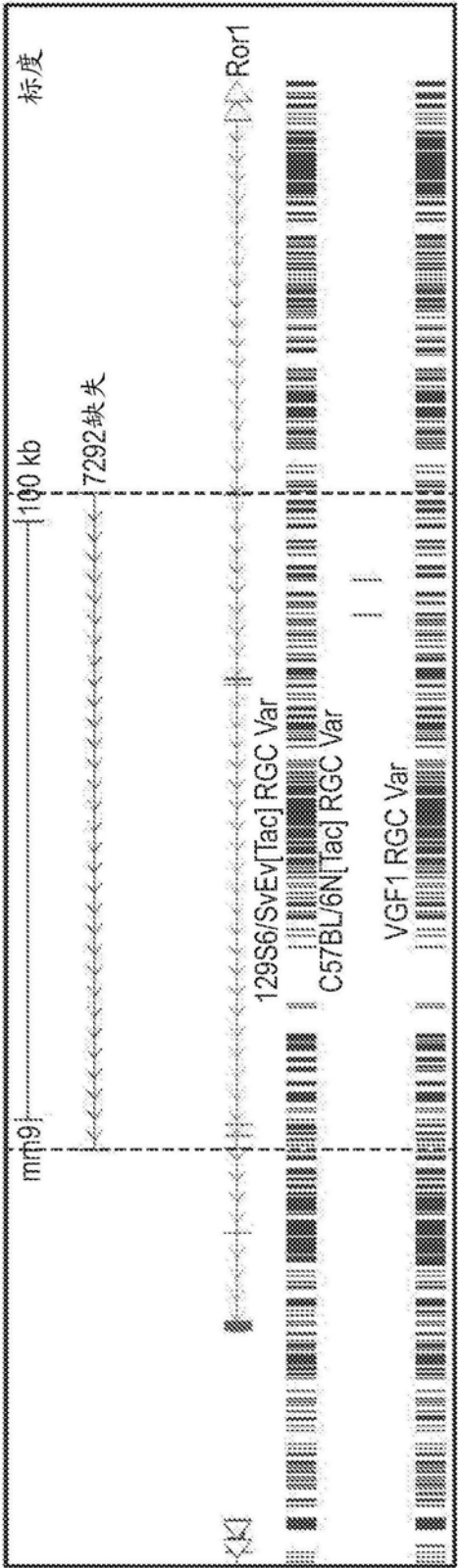


图25

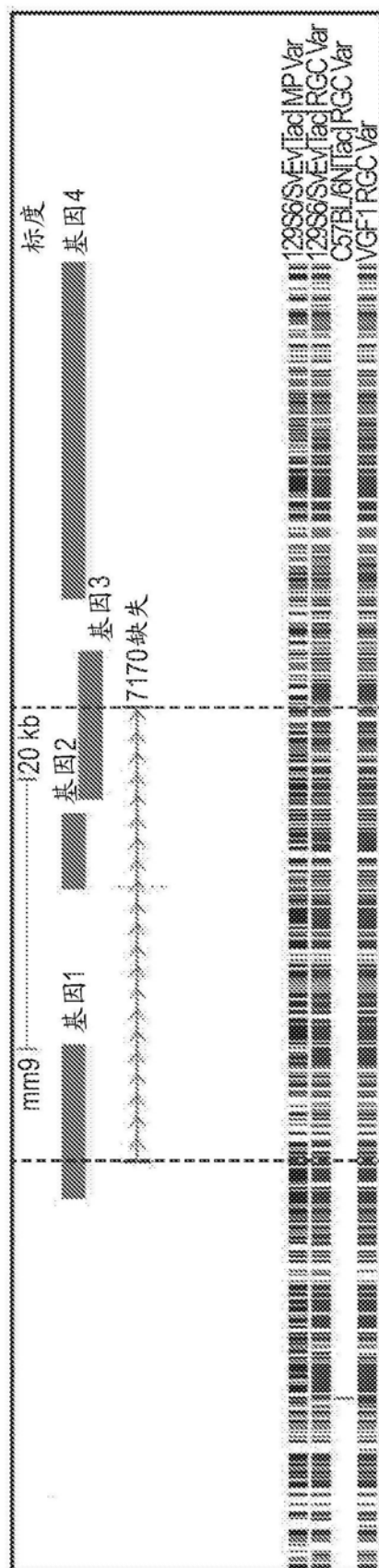


图26

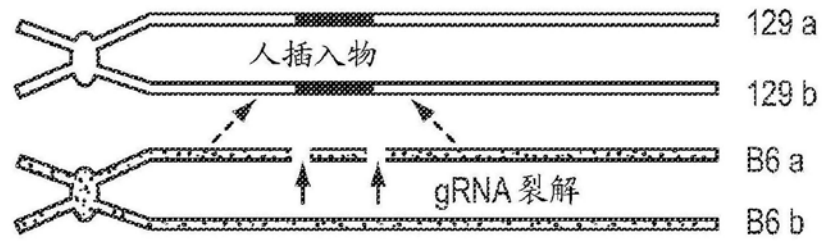


图27A

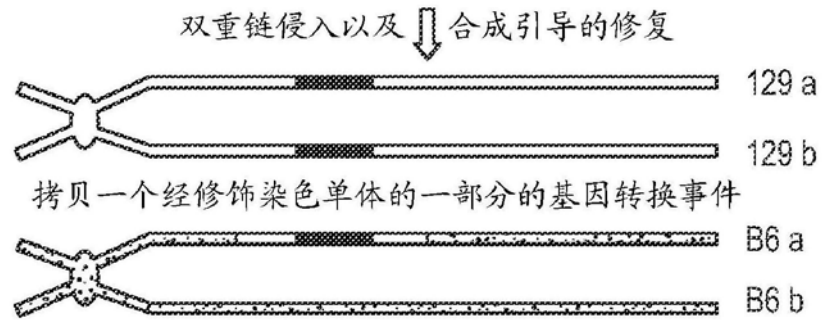


图27B

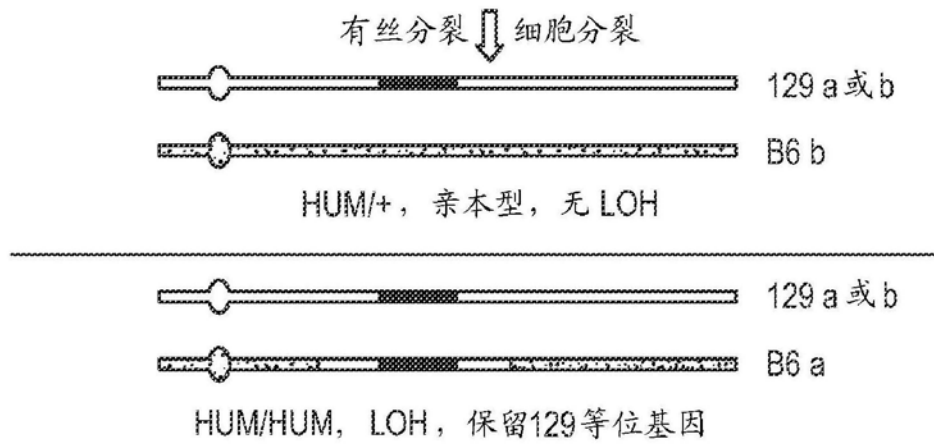


图27C

不同大小的CRISPR-Cas9缺失的效率

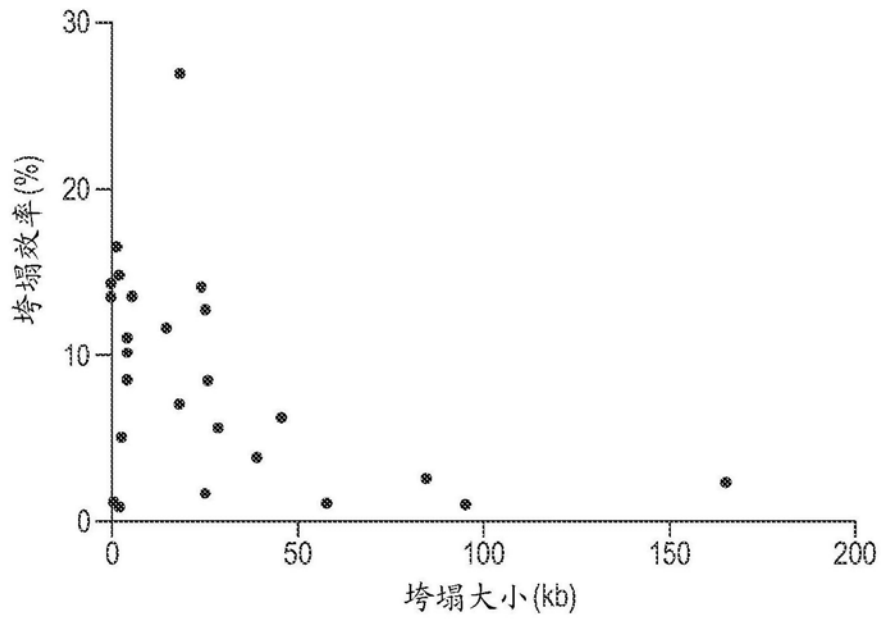


图28

垮塌大小相对于效率

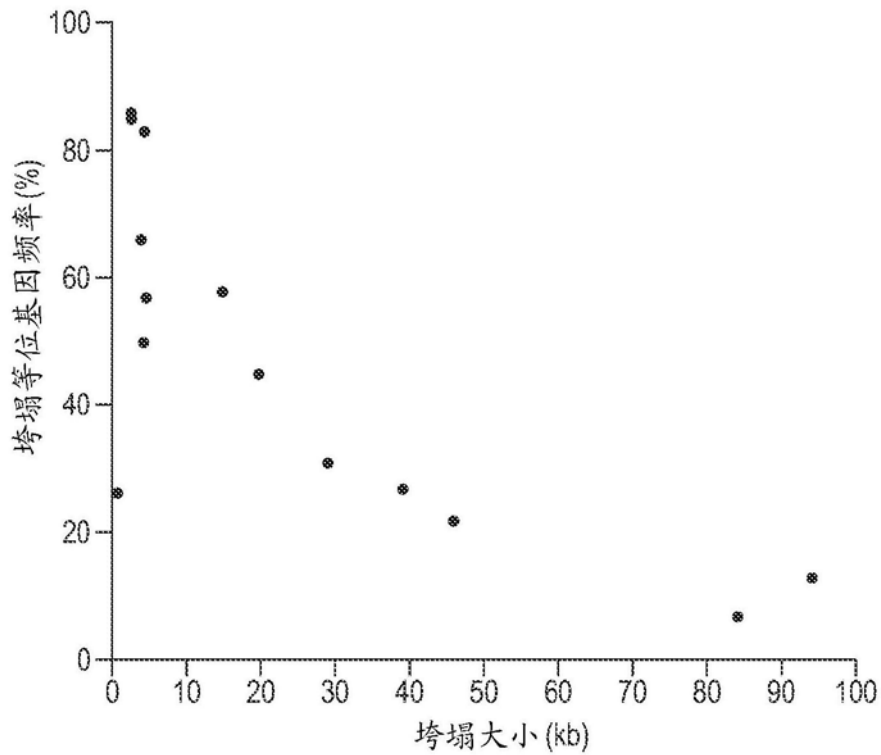


图29

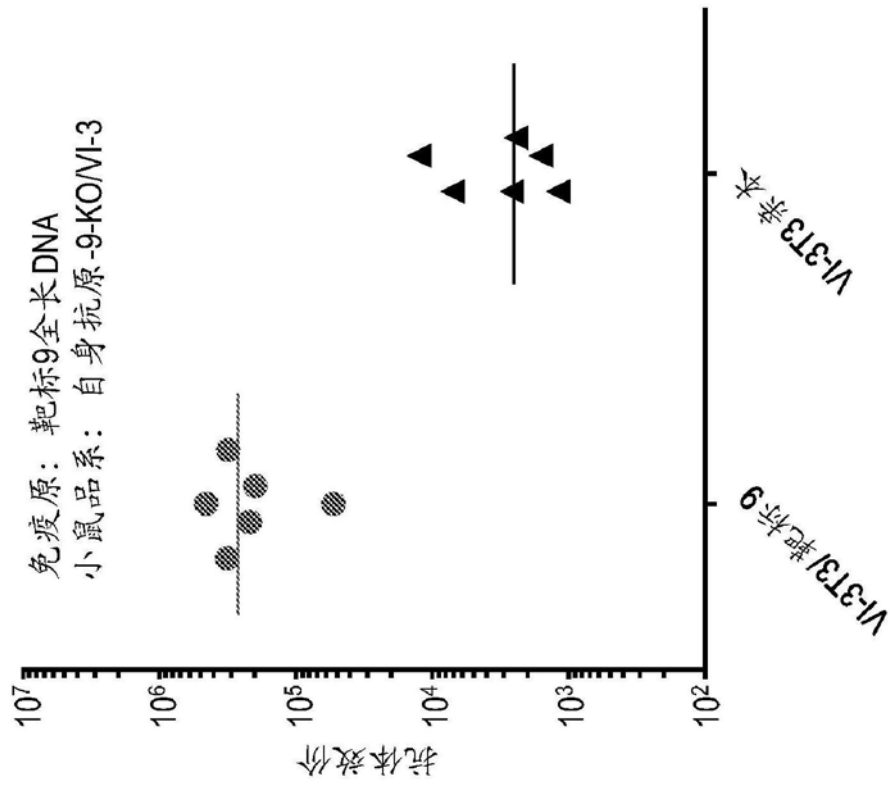


图30A

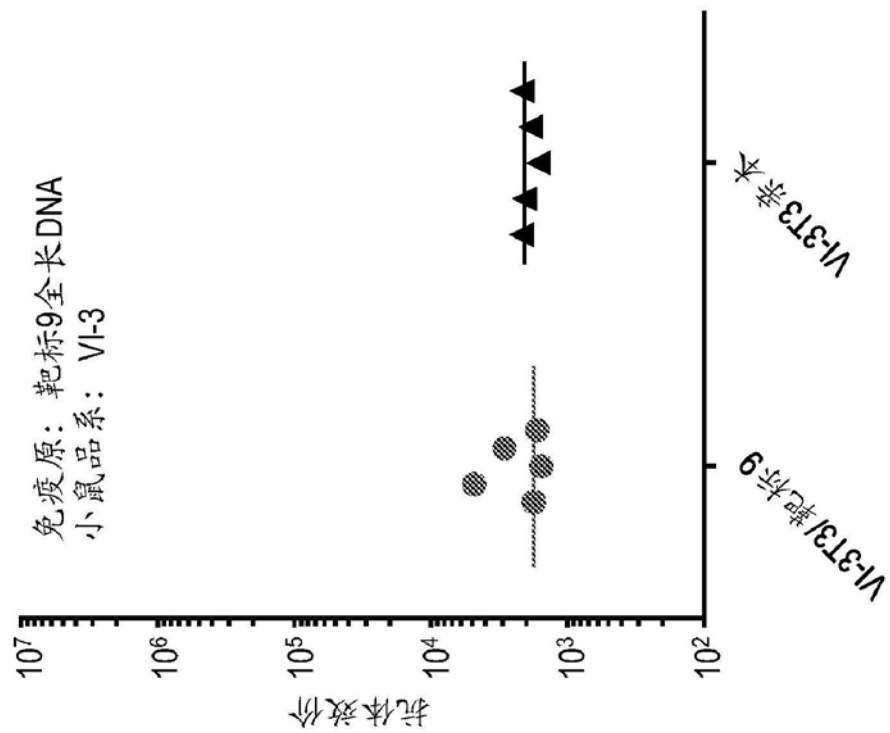


图30B

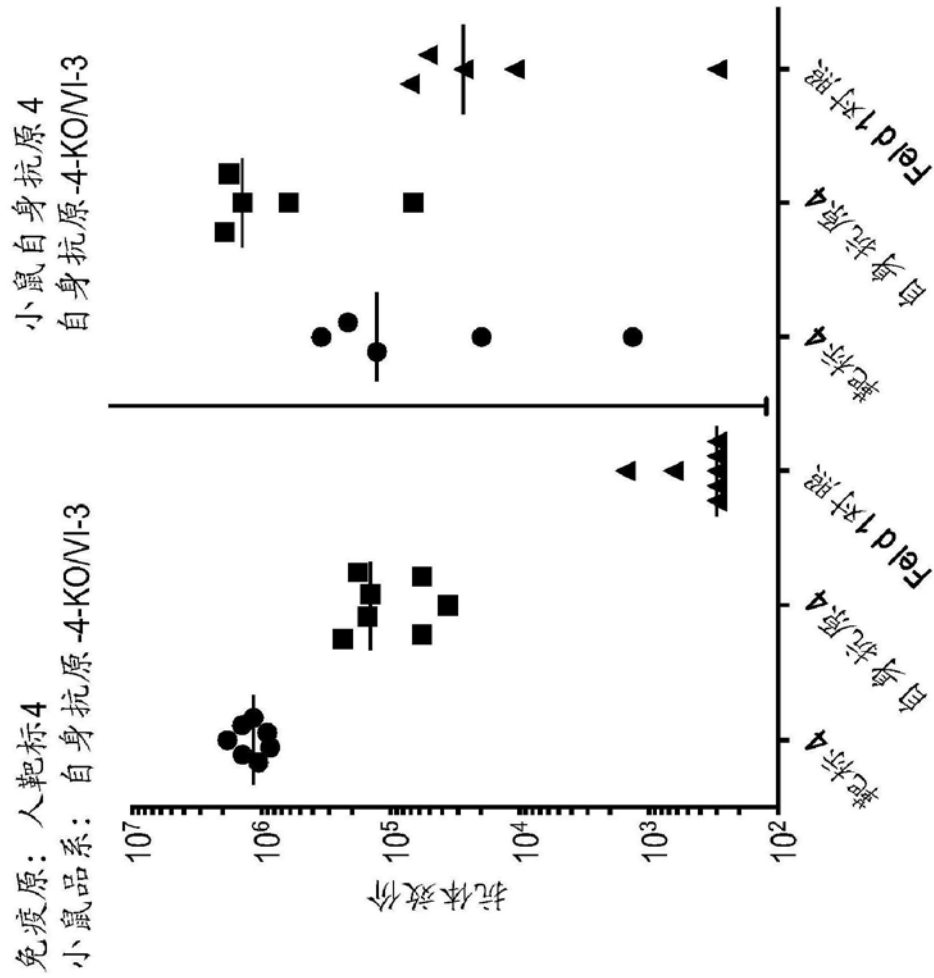


图31A

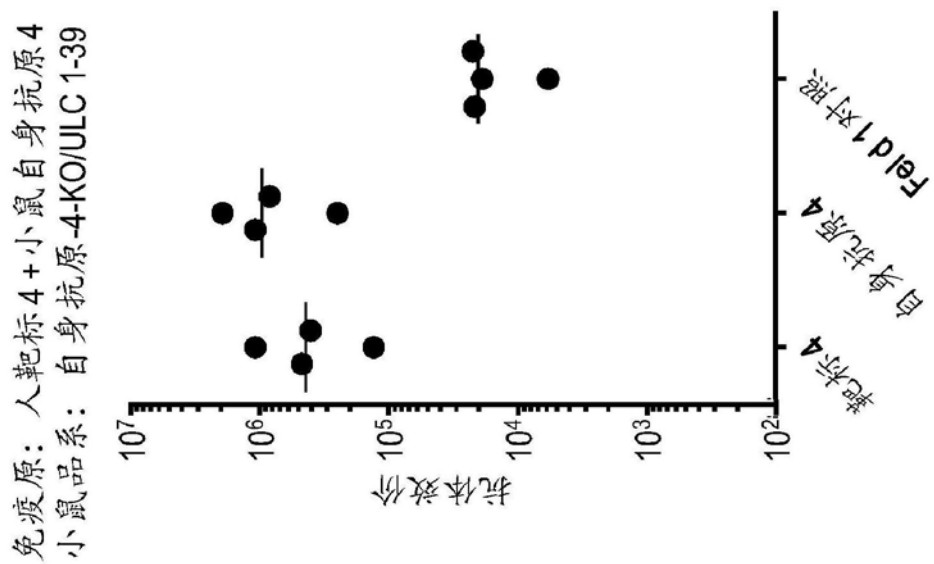


图31B

50% BALB/cTac, 25% C57BL/6NTac, 25% 129S6/SvEvTac

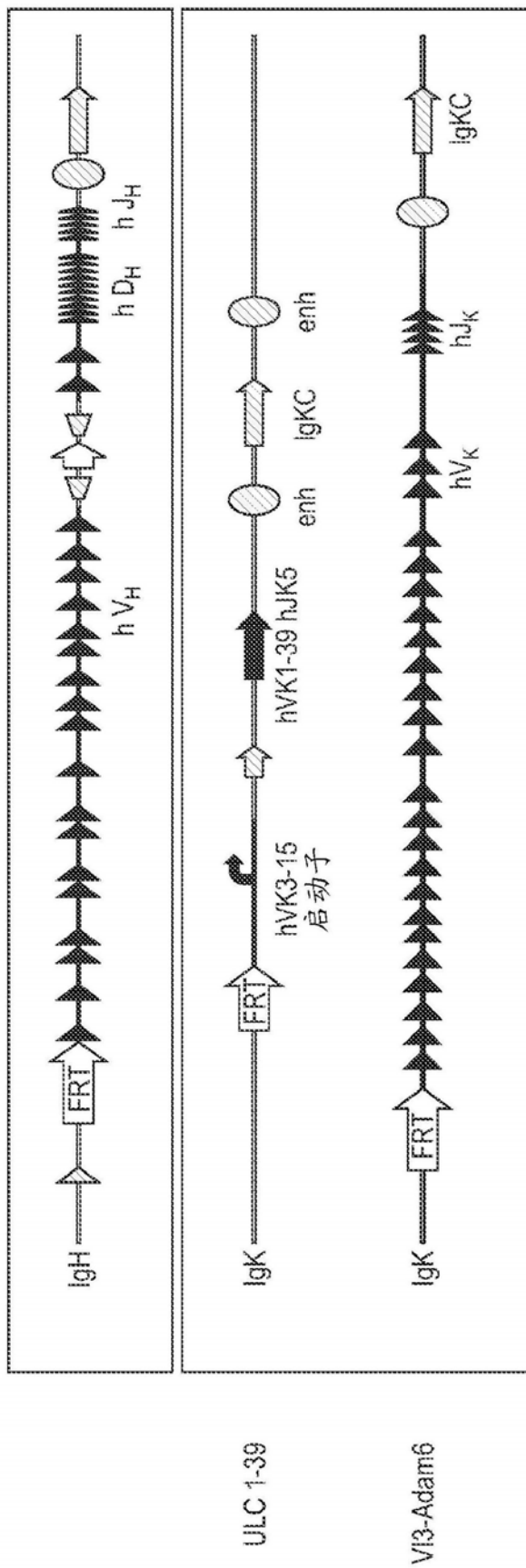


图32