



(12) 发明专利

(10) 授权公告号 CN 110951779 B

(45) 授权公告日 2024. 04. 16

(21) 申请号 201911060314.X

(22) 申请日 2014.10.15

(65) 同一申请的已公布的文献号  
申请公布号 CN 110951779 A

(43) 申请公布日 2020.04.03

(30) 优先权数据  
61/914,768 2013.12.11 US  
62/017,416 2014.06.26 US  
62/029,261 2014.07.25 US  
62/052,906 2014.09.19 US  
62/059,527 2014.10.03 US

(62) 分案原申请数据  
201480074803.X 2014.10.15

(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 D·弗伦德维 W·奥尔巴克

K·V·莱 久野淳子

D·M·瓦伦泽拉

G·D·扬科普洛斯

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

专利代理师 郝传鑫

(51) Int. Cl.  
C12N 15/85 (2006.01)  
C12N 15/90 (2006.01)  
A01K 67/0278 (2024.01)  
C07K 14/715 (2006.01)  
C07K 14/775 (2006.01)

(56) 对比文件  
CN 105308184 A, 2016.02.03

审查员 张谨

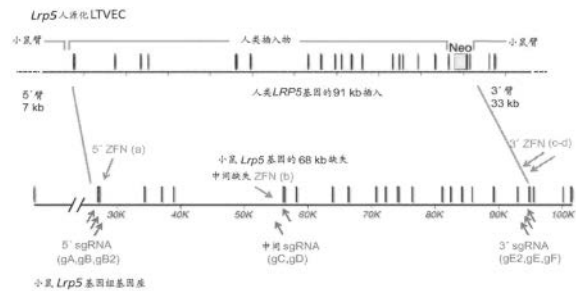
权利要求书6页 说明书142页  
序列表25页 附图77页

(54) 发明名称

用于靶向修饰基因组的方法和组合物

(57) 摘要

本发明提供使用如本文所述的包含各种内源或外源核酸序列的大靶向载体 (LTVEC) 修饰在真核细胞、哺乳动物细胞、人类细胞或非人类哺乳动物细胞中的目标基因组基因座的组合物和方法。另外的方法组合使用所述LTVEC和CRISPR/Cas系统。本发明还提供用于产生在其种系中包含一种或多种靶向基因修饰的基因修饰的非人类动物的组合物和方法。



1. 一种在人类诱导的多潜能干细胞中修饰目标基因组基因座处的基因组的体外方法, 包括:

向所述人类诱导的多潜能干细胞引入Cas9蛋白或编码所述Cas9蛋白的核酸、在所述目标基因组基因座处与CRISPR靶序列杂交的CRISPR RNA或编码所述CRISPR RNA的DNA、tracrRNA或编码所述tracrRNA的DNA, 以及大靶向载体, 所述大靶向载体大小至少为10kb, 其包括插入核酸, 所述插入核酸侧接:

(i) 与在所述目标基因组基因座处的5'靶序列同源的5'同源臂; 和

(ii) 与在所述目标基因组基因座处的3'靶序列同源的3'同源臂, 以及

其中在将所述Cas9蛋白或编码所述Cas9蛋白的核酸、所述CRISPR RNA或编码所述CRISPR RNA的所述DNA、所述tracrRNA或编码所述tracrRNA的所述DNA, 以及所述大靶向载体引入所述人类诱导的多潜能干细胞后, 所述人类诱导的多潜能干细胞的基因组被修饰以包含靶向基因修饰, 所述靶向基因修饰包括 (i) 所述目标基因组基因座的区的缺失, 所述的缺失至少为30kb和/或 (ii) 在所述目标基因组基因座处插入所述插入核酸, 其中所述插入至少为30kb。

2. 如权利要求1所述的方法, 其中将所述CRISPR RNA和所述tracrRNA作为包含所述CRISPR RNA和所述tracrRNA的单一核酸分子引入。

3. 如权利要求2所述的方法, 其中所述单一核酸分子包含以单向导RNA形式融合在一起的所述CRISPR RNA和所述tracrRNA。

4. 如权利要求1所述的方法, 其中所述CRISPR RNA和所述tracrRNA单独地引入。

5. 如权利要求1-4中任一项所述的方法, 其中:

(a) 所述Cas9蛋白以蛋白质、编码所述Cas9蛋白的信使RNA (mRNA) 或编码所述Cas9蛋白的DNA的形式引入;

(b) 所述CRISPR RNA以编码所述CRISPR RNA的RNA或DNA的形式引入; 且

(c) 所述tracrRNA以编码所述tracrRNA的RNA或DNA的形式引入。

6. 如权利要求5所述的方法, 其中将所述Cas9蛋白、所述CRISPR RNA和所述tracrRNA作为蛋白质-RNA复合体引入。

7. 如权利要求5所述的方法, 其中:

(a) 编码所述Cas9蛋白的所述DNA呈包含操作性连接编码所述Cas9蛋白的第一核酸的第一启动子的第一表达构建体的形式;

(b) 编码所述CRISPR RNA的所述DNA呈包含操作性连接编码所述CRISPR RNA的第二核酸的第二启动子的第二表达构建体的形式; 且

(c) 编码所述tracrRNA的所述DNA呈包含操作性连接编码所述tracrRNA的第三核酸的第三启动子的第三表达构建体的形式;

其中所述第一启动子、所述第二启动子和所述第三启动子在所述人类诱导的多潜能干细胞中具有活性, 且

其中所述第一表达构建体、所述第二表达构建体和所述第三表达构建体在单一核酸分子上或在多个核酸分子上。

8. 如权利要求5所述的方法, 其中:

(a) 编码所述Cas9蛋白的所述DNA呈包含操作性连接编码所述Cas9蛋白的第一核酸的

第一启动子的第一表达构建体的形式;且

(b) 编码所述CRISPR RNA的所述DNA和编码所述tracrRNA的所述DNA呈包含操作性连接编码包含所述CRISPR RNA和所述tracrRNA的gRNA的第二核酸的第二启动子的第二表达构建体的形式;

其中所述第一启动子和所述第二启动子在所述人类诱导的多潜能干细胞中具有活性,且

其中所述第一表达构建体和所述第二表达构建体在单一核酸分子上或在单独的核酸分子上。

9. 如权利要求1-4和6-8中任一项所述的方法,其中所述靶向基因修饰同时包括在所述目标基因组基因座处缺失内源核酸序列和在所述目标基因组基因座处插入所述插入核酸。

10. 如权利要求9所述的方法,其中所述缺失的内源核酸序列为30kb至110kb,且所述插入核酸为40kb至140kb。

11. 如权利要求1-4、6-8和10中任一项所述的方法,其中所述靶向基因修饰为双等位基因基因修饰。

12. 如权利要求11所述的方法,其中所述双等位基因基因修饰包括在两种同源染色体中在所述目标基因组基因座处内源核酸序列的缺失和所述插入核酸的插入。

13. 如权利要求11所述的方法,其中所述修饰的人类诱导的多潜能干细胞在所述目标基因组基因座处为复合杂合的或半合的。

14. 如权利要求13所述的方法,其中在一种染色体中在所述目标基因组基因座处的所述靶向基因修饰包括内源核酸序列的缺失和所述插入核酸的插入。

15. 如权利要求14所述的方法,其中所述靶向基因修饰包括:(1)在第一同源染色体和第二同源染色体中在所述目标基因组基因座处缺失内源核酸序列;和(2)所述插入核酸插入在所述第一同源染色体中的所述目标基因组基因座中并破坏在所述第二同源染色体中的所述目标基因组基因座。

16. 如权利要求1-4、6-8、10和12-15中任一项所述的方法,其中所述大靶向载体为至少40kb;或

其中所述靶向基因修饰包括所述目标基因组基因座的区的缺失,其中所述缺失至少为30kb,而所述大靶向载体为至少15kb。

17. 如权利要求1-4、6-8、10和12-15中任一项所述的方法,其中靶向基因修饰包括所述插入核酸的插入,其中所述插入核酸为至少40kb;或

其中所述靶向基因修饰包括所述目标基因组基因座的区的缺失和所述插入核酸的插入,其中所述缺失至少为30kb,其中所述插入核酸至少为10kb。

18. 如权利要求1-4、6-8、10和12-15中任一项所述的方法,其中所述插入核酸为40kb至140kb。

19. 如权利要求1-4、6-8、10和12-15中任一项所述的方法,其中所述CRISPR靶序列由原间隔区邻近基序序列直接侧接。

20. 如权利要求1-4、6-8、10和12-15中任一项所述的方法,其中所述大靶向载体的所述5'同源臂和所述3'同源臂的总和为10kb至150kb。

21. 如权利要求20所述的方法,其中所述大靶向载体的所述5'同源臂和所述3'同源臂

的总和为30kb至150kb。

22. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述大靶向载体的长度为100kb至300kb。

23. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述靶向基因修饰包括:

(a) 用同源或直系同源的核酸序列替换内源核酸序列;

(b) 缺失内源核酸序列;

(c) 插入外源核酸序列;

(d) 插入包含同源或直系同源的核酸序列的外源核酸序列;

(e) 插入包含人类核酸序列和非人类核酸序列的嵌合核酸序列;

(f) 插入由位点特异性重组酶靶序列侧接的条件性等位基因;

(g) 插入操作性连接在所述人类诱导的多潜能干细胞中具有活性的启动子的可选择标记或报道基因;或

(h) 其组合。

24. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述靶向基因修饰包括缺失所述目标基因组基因座的区,

其中所述缺失至少为40kb;或

其中所述靶向基因修饰还包括在所述目标基因组基因座处插入所述插入核酸和所述目标基因组基因座的区的缺失,其中所述插入至少为30kb,其中所述缺失至少为10kb。

25. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中缺失的所述基因组基因座的所述区为30kb至110kb。

26. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述靶向基因修饰包括缺失所述目标基因组基因座的区和在所述目标基因组基因座处插入所述插入核酸,其中所述缺失至少为30kb,其中所述插入至少为30kb。

27. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述大靶向载体为100kb至300kb,所述5'同源臂和3'同源臂的总和为30kb至150kb,且所述靶向基因修饰包括缺失所述目标基因组基因座的区和在所述目标基因组基因座处插入所述插入核酸,其中所述缺失为30kb至110kb,其中所述插入为40kb至140kb。

28. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述目标基因组基因座为与所述人类诱导的多潜能干细胞内源的。

29. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述目标基因组基因座包含整合到所述人类诱导的多潜能干细胞的所述基因组中的DNA的异源或外源片段。

30. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法,其中所述目标基因组基因座是免疫球蛋白基因座。

31. 如权利要求30所述的方法,其中所述免疫球蛋白基因座编码人免疫球蛋白重链可变区氨基酸序列。

32. 如权利要求30所述的方法,其中所述免疫球蛋白基因座编码人免疫球蛋白轻链可变区氨基酸序列。

33. 如权利要求32所述的方法,其中所述免疫球蛋白基因座包括:(i) 未重排的人 $\lambda$ 轻链



可变区核酸序列；(ii) 未重排的人 $\kappa$ 轻链可变区核酸序列；或(iii) 未重排的人 $\lambda$ 和 $\kappa$ 轻链可变区核酸序列。

34. 如权利要求32所述的方法，其中所述免疫球蛋白基因座包括：(i) 重排的人 $\lambda$ 轻链可变区核酸序列；(ii) 重排的人 $\kappa$ 轻链可变区核酸序列；或(iii) 重排的人 $\lambda$ 和 $\kappa$ 轻链可变区核酸序列。

35. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法，其中所述目标基因组基因座是T细胞受体基因座。

36. 如权利要求35所述的方法，其中所述T细胞受体基因座是T细胞受体 $\alpha$ 基因座。

37. 如权利要求1-4、6-8、10、12-15和21中任一项所述的方法，其中所述目标基因组基因座包括干扰白细胞素-2受体 $\gamma$ 基因座、ApoE基因座、Rag1基因座、Rag2基因座、所述Rag1基因座和所述Rag2基因座两者、Adamts5基因座、Trpa1基因座、Folh1基因座、ErbB4基因座、Lrp5基因座、C5基因座、Ror1基因座或Dpp4基因座。

38. 如权利要求1-4、6-8、10、12-15、21、31-34和36中任一项所述的方法，其中所述插入核酸包含编码人类免疫球蛋白重链可变区氨基酸序列的基因组核酸序列。

39. 如权利要求38所述的方法，其中所述插入核酸包含一个或多个功能性人类 $V_H$ 基因片段，所述功能性人类 $V_H$ 基因片段包含 $V_H1-2$ 、 $V_H1-3$ 、 $V_H1-8$ 、 $V_H1-18$ 、 $V_H1-24$ 、 $V_H1-45$ 、 $V_H1-46$ 、 $V_H1-58$ 、 $V_H1-69$ 、 $V_H2-5$ 、 $V_H2-26$ 、 $V_H2-70$ 、 $V_H3-7$ 、 $V_H3-9$ 、 $V_H3-11$ 、 $V_H3-13$ 、 $V_H3-15$ 、 $V_H3-16$ 、 $V_H3-20$ 、 $V_H3-21$ 、 $V_H3-23$ 、 $V_H3-30$ 、 $V_H3-30-3$ 、 $V_H3-30-5$ 、 $V_H3-33$ 、 $V_H3-35$ 、 $V_H3-38$ 、 $V_H3-43$ 、 $V_H3-48$ 、 $V_H3-49$ 、 $V_H3-53$ 、 $V_H3-64$ 、 $V_H3-66$ 、 $V_H3-72$ 、 $V_H3-73$ 、 $V_H3-74$ 、 $V_H4-4$ 、 $V_H4-28$ 、 $V_H4-30-1$ 、 $V_H4-30-2$ 、 $V_H4-30-4$ 、 $V_H4-31$ 、 $V_H4-34$ 、 $V_H4-39$ 、 $V_H4-59$ 、 $V_H4-61$ 、 $V_H5-51$ 、 $V_H6-1$ 、 $V_H7-4-1$ 、 $V_H7-81$ ，或其组合。

40. 如权利要求38所述的方法，其中所述插入核酸包含一个或多个功能性人类D基因片段，所述人类D基因片段包含D1-1、D1-7、D1-14、D1-20、D1-26、D2-2、D2-8、D2-15、D2-21、D3-3、D3-9、D3-10、D3-16、D3-22、D4-4、D4-11、D4-17、D4-23、D5-12、D5-5、D5-18、D5-24、D6-6、D6-13、D6-19、D6-25、D7-27或其组合。

41. 如权利要求38所述的方法，其中所述插入核酸包含一个或多个功能性 $J_H$ 基因片段，所述功能性 $J_H$ 基因片段包括 $J_H1$ 、 $J_H2$ 、 $J_H3$ 、 $J_H4$ 、 $J_H5$ 、 $J_H6$ 或其组合。

42. 如权利要求1-4、6-8、10、12-15、21、31-34和36中任一项所述的方法，其中所述插入核酸包含编码人类免疫球蛋白轻链可变区氨基酸序列的基因组核酸序列。

43. 如权利要求42所述的方法，其中所述插入核酸包含一个或多个人类 $V_K$ 基因片段，所述人类 $V_K$ 基因片段包含 $V_K4-1$ 、 $V_K5-2$ 、 $V_K7-3$ 、 $V_K2-4$ 、 $V_K1-5$ 、 $V_K1-6$ 、 $V_K3-7$ 、 $V_K1-8$ 、 $V_K1-9$ 、 $V_K2-10$ 、 $V_K3-11$ 、 $V_K1-12$ 、 $V_K1-13$ 、 $V_K2-14$ 、 $V_K3-15$ 、 $V_K1-16$ 、 $V_K1-17$ 、 $V_K2-18$ 、 $V_K2-19$ 、 $V_K3-20$ 、 $V_K6-21$ 、 $V_K1-22$ 、 $V_K1-23$ 、 $V_K2-24$ 、 $V_K3-25$ 、 $V_K2-26$ 、 $V_K1-27$ 、 $V_K2-28$ 、 $V_K2-29$ 、 $V_K2-30$ 、 $V_K3-31$ 、 $V_K1-32$ 、 $V_K1-33$ 、 $V_K3-34$ 、 $V_K1-35$ 、 $V_K2-36$ 、 $V_K1-37$ 、 $V_K2-38$ 、 $V_K1-39$ 、 $V_K2-40$ 或其组合。

44. 如权利要求42所述的方法，其中所述插入核酸包含一个或多个人类 $V_\lambda$ 基因片段，所述人类 $V_\lambda$ 基因片段包含 $V_\lambda3-1$ 、 $V_\lambda4-3$ 、 $V_\lambda2-8$ 、 $V_\lambda3-9$ 、 $V_\lambda3-10$ 、 $V_\lambda2-11$ 、 $V_\lambda3-12$ 、 $V_\lambda2-14$ 、 $V_\lambda3-16$ 、 $V_\lambda2-18$ 、 $V_\lambda3-19$ 、 $V_\lambda3-21$ 、 $V_\lambda3-22$ 、 $V_\lambda2-23$ 、 $V_\lambda3-25$ 、 $V_\lambda3-27$ 或其组合。

45. 如权利要求42所述的方法，其中所述插入核酸包含一个或多个人类 $J_K$ 基因片段，所述人类 $J_K$ 基因片段包含 $J_K1$ 、 $J_K2$ 、 $J_K3$ 、 $J_K4$ 、 $J_K5$ 或其组合。

46. 如权利要求1-4、6-8、10、12-15、21、31-34和36中任一项所述的方法，其中所述插入

核酸包括编码人类T细胞受体的至少一个区的多核苷酸。

47. 如权利要求46所述的方法,其中所述T细胞受体是T细胞受体 $\alpha$ 。

48. 如权利要求1-4、6-8、10、12-15、21、31-34和36中任一项所述的方法,其中所述插入核酸包含至少一个疾病等位基因。

49. 如权利要求48所述的方法,其中所述插入核酸包含人类基因的至少一个人类疾病等位基因。

50. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述CRISPR靶序列不存在于所述插入核酸中。

51. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述CRISPR靶序列位于所述目标基因组基因座处的所述5'靶序列与所述3'靶序列之间的任何地方。

52. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述CRISPR靶序列直接邻近于所述目标基因组基因座处的所述5'靶序列或所述3'靶序列。

53. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述CRISPR靶序列位于内含子、外显子、或调控区中。

54. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述CRISPR靶序列位于基因的编码区内或影响所述基因的表达的调控区内。

55. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述向导RNA包括SEQ ID NO:2、3、4、5、6、7或8。

56. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述人类诱导的多潜能干细胞正维持在包含基础培养基和补充剂的培养基中,其中所述培养基包含:

- (a) 白血病抑制因子多肽;
- (b) 肝糖合成酶激酶抑制剂;和
- (c) MEK抑制剂;

其中所述基础培养基具有180mOsm/kg至250mOsm/kg的渗透压度。

57. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述Cas9蛋白包含核定位信号。

58. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述大靶向载体的所述5'同源臂和所述3'同源臂的总和为至少10kb。

59. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述CRISPR靶序列是对所述人类诱导的多潜能干细胞为内源的天然序列。

60. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述目标基因组基因座对所述人类诱导的多潜能干细胞为天然的。

61. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述Cas9蛋白包含核定位信号,并且

其中所述CRISPR靶序列位于所述5'靶序列和所述3'靶序列之间的任何地方,或者所述CRISPR靶序列直接邻近于所述5'靶序列或所述3'靶序列。

62. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述大靶向载体的所述5'同源臂和所述3'同源臂的总和为至少10kb,

其中所述CRISPR靶序列是对所述人类诱导的多潜能干细胞为内源的天然序列,以及

其中所述CRISPR靶序列位于所述5'靶序列和所述3'靶序列之间的任何地方,或者所述CRISPR靶序列直接邻近于所述5'靶序列或所述3'靶序列。

63. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述大靶向载体的所述5'同源臂和所述3'同源臂的总和为至少10kb,

其中所述目标基因组基因座对所述人类诱导的多潜能干细胞为天然的,以及

其中所述CRISPR靶序列位于所述5'靶序列和所述3'靶序列之间的任何地方,或者所述CRISPR靶序列直接邻近于所述5'靶序列或所述3'靶序列。

64. 如权利要求1-4、6-8、10、12-15、21、31-34、36、39-41、43-45、47和49中任一项所述的方法,其中所述靶向基因修饰包括在所述目标基因组基因座处插入所述插入核酸,或者其中所述靶向基因修饰包括缺失所述目标基因组基因座的所述区,以及在所述目标基因组基因座处插入所述插入核酸。

65. 如权利要求64所述的方法,其中所述CRISPR靶序列是对所述人类诱导的多潜能干细胞为内源的天然序列,且其中所述大靶向载体的所述5'同源臂和所述3'同源臂的总和为10kb至150kb。

## 用于靶向修饰基因组的方法和组合物

[0001] 分案申请说明

[0002] 本申请是申请日为2014年10月15日、申请号为201480074803.X、发明名称为“用于靶向修饰基因组的方法和组合物”的发明专利申请的分案申请。

[0003] 相关申请的交叉引用

[0004] 本申请要求2013年12月11日提交的美国临时专利申请号61/914,768、2014年6月26日提交的美国临时专利申请号62/017,416、2014年7月25日提交的美国临时专利申请号62/029,261、2014年9月19日提交的美国临时专利申请号62/052,906、2014年10月3日提交的美国临时专利申请号62/059,527和2014年10月15日提交的美国临时专利申请号62/064,384的权益,其各自的全部内容都出于所有目的以引用的方式并入本文中。

[0005] 参考经由EFS WEB作为文本文件提交的序列列表

[0006] 序列列表的正式文本经由EFS-Web作为具有名为453460SEQLIST.TXT在2014年10月15日产生且具有27.5千字节的大小的文件的ASCII格式的序列列表电子提交,且与本说明书同时提交。在该ASCII格式文件中包含的序列列表为本说明书的一部分且其全部内容以引用的方式并入本文中。

### 背景技术

[0007] 虽然已经将大鼠视为可概括各种人类疾病的病变的重要动物模型系统,所述人类疾病包括但不限于心血管疾病(例如,高血压)、新陈代谢疾病(例如,肥胖、糖尿病)、神经病学疾病(例如,疼痛病变)和多种癌症,但是与小鼠相比较,大鼠在模仿人类疾病中的用途受限,这部分地归因于无法利用种系可传递的多潜能大鼠细胞,其可在例如一种或多种串行电穿孔的一系列体外基因修饰之后持续其多潜能性;且部分地归因于缺乏允许在多潜能大鼠细胞中引入或缺失大基因组DNA序列或用外源核酸序列替换大内源基因组DNA序列的有效靶向技术。

[0008] 在本领域中需要允许在生物体的基因组中精确靶向改变的组合物和方法,其可开放或扩展靶发现的当前领域并更迅速且容易地证实治疗剂。

[0009] 概述

[0010] 提供经由靶向基因修饰来修饰在真核细胞中的目标基因组基因座的方法。这一方法包括:

[0011] (a) 向所述真核细胞中引入:(i) 包含侧接有5'同源臂和3'同源臂的第一核酸的大靶向载体(LTVEC),其中所述LTVEC为至少10kb,(ii) 包含操作性连接编码Cas蛋白的第二核酸的第一启动子的第一表达构建体,(iii) 包含操作性连接编码包含杂化到靶序列的核苷酸序列的向导RNA(gRNA)和反式激活CRISPR RNA(tracrRNA)的第三核酸的第二启动子的第二表达构建体,其中所述第一启动子和所述第二启动子在所述真核细胞中具有活性;和(b) 鉴定在目标基因组基因座处包括靶向基因修饰的修饰的真核细胞。

[0012] 在一个实施方案中,所述靶向基因修饰为双等位基因基因修饰。

[0013] 在一个实施方案中,所述LTVEC为至少15kb、至少20kb、至少30kb、至少40kb、至少

50kb、至少60kb、至少70kb、至少80kb或至少90kb。在另一实施方案中,所述LTVEC为至少100kb、至少150kb或至少200kb。

[0014] 在一个实施方案中,所述真核细胞为哺乳动物细胞。在一个实施方案中,所述哺乳动物细胞为成纤维细胞。

[0015] 在一个实施方案中,所述真核细胞为多潜能细胞。在一个实施方案中,所述多潜能细胞为人类多潜能细胞。在一个实施方案中,所述人类多潜能细胞为人类胚胎干(ES)细胞或成人干细胞。在另一实施方案中,所述人类多潜能细胞为发育受限的人类祖细胞。在另一实施方案中,所述人类多潜能细胞为人类诱导的多潜能干(iPS)细胞。

[0016] 在一个实施方案中,所述Cas蛋白为Cas9。

[0017] 在一个实施方案中,所述靶序列由原间隔区邻近基序(PAM)序列侧接。在一个实施方案中,所述靶序列在3'端由原间隔区邻近基序(PAM)序列直接侧接。

[0018] 在一些实施方案中,所述5'同源臂和所述3'同源臂的总和为约10kb-约150kb。在一些实施方案中,所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-150kb。

[0019] 所述方法进一步提供包括以下的靶向基因修饰:(a)用同源或直系同源的核酸序列替换内源核酸序列;(b)缺失内源核酸序列;(c)缺失内源核酸序列,其中所述缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb;(d)插入外源核酸序列;(e)插入外源核酸序列,所述外源核酸序列为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb;(f)插入包含同源或直系同源的核酸序列的外源核酸序列;(g)插入包含人类核酸序列和非人类核酸序列的嵌合核酸序列;(h)插入侧接有位点特异性重组酶靶序列的条件性等位基因;(i)插入操作性连接在多潜能细胞中具有活性的第三启动子的可选择标记或报道基因;或(j)其组合。

[0020] 在一个实施方案中,所述目标基因组基因座包含(i)与所述5'同源臂同源的5'靶序列;和(ii)与所述3'同源臂同源的3'靶序列。

[0021] 在一些实施方案中,所述5'靶序列和所述3'靶序列相隔至少5kb,但小于3Mb。在一些实施方案中,所述5'靶序列和所述3'靶序列相隔至少5kb,但小于10kb;至少10kb,但小于20kb;至少20kb,但小于40kb;至少40kb,但小于60kb;至少60kb,但小于80kb;至少约80kb,但小于100kb;至少100kb,但小于150kb;或至少150kb,但小于200kb;至少约200kb,但小于约300kb;至少约300kb,但小于约400kb;至少约400kb,但小于约500kb;至少约500kb,但小于约1Mb;至少约1Mb,但小于约1.5Mb;至少约1.5Mb,但小于约2Mb;至少约2Mb,但小于约2.5Mb;或至少约2.5Mb,但小于约3Mb。

[0022] 在一个实施方案中,所述目标基因组基因座包括干扰白细胞素-2受体 $\gamma$ 基因座、ApoE基因座、Rag1基因座、Rag2基因座或所述Rag1基因座和所述Rag2基因座两者。

[0023] 在一个实施方案中,所述第一表达构建体和所述第二表达构建体在单一核酸分子

上。

[0024] 进一步提供一种用于修饰基因组的方法,其包括在包含至少10kb的核酸序列的大靶向载体(LTVEC)存在下将所述基因组暴露于Cas蛋白和CRISPR RNA,其中在暴露于所述Cas蛋白、所述CRISPR RNA和所述LTVEC之后,所述基因组被修饰以含有至少10kb的核酸序列。

[0025] 在一些这样的方法中,所述LTVEC包含至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb或至少90kb的核酸序列。在一些这样的方法中,所述LTVEC包含至少100kb、至少150kb或至少200kb的核酸序列。

[0026] 进一步提供一种用于修饰基因组的方法,其包括在大靶向载体(LTVEC)存在下使所述基因组与Cas蛋白、杂化到靶序列的CRISPR RNA,和tracrRNA接触,其中所述LTVEC为至少10kb且包含侧接有5'同源臂和3'同源臂的第一核酸,其中在所述LTVEC存在下与所述Cas蛋白、CRISPR RNA,和tracrRNA接触之后,所述基因组在目标基因组基因座处被修饰以含有所述第一核酸。所述靶序列可在所述目标基因组基因座处或在其附近。

[0027] 在一些这样的方法中,所述基因组在真核细胞中,且将所述Cas蛋白、所述CRISPR RNA、所述tracrRNA和所述LTVEC引入所述真核细胞中。一些这样的方法进一步包括鉴定在所述目标基因组基因座处包括靶向基因修饰的修饰的真核细胞。

[0028] 在一些这样的方法中,所述CRISPR RNA和所述tracrRNA以单一向导RNA(gRNA)的形式一起引入。在其它方法中,所述CRISPR RNA和所述tracrRNA单独地引入。

[0029] 在一些这样的方法中,(a)所述Cas蛋白以蛋白质、编码所述Cas蛋白的信使RNA(mRNA)或编码所述Cas蛋白的DNA的形式引入所述真核细胞中;(b)所述CRISPR RNA以编码所述CRISPR RNA的RNA或DNA的形式引入所述真核细胞中;和(c)所述tracrRNA以编码所述tracrRNA的RNA或DNA的形式引入所述真核细胞中。

[0030] 在一些方法中,(a)编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式;(b)编码所述CRISPR RNA的所述DNA呈包含操作性连接编码所述CRISPR RNA的第三核酸的第二启动子的第二表达构建体的形式;且(c)编码所述tracrRNA的所述DNA呈包含操作性连接编码所述tracrRNA的第四核酸的第三启动子的第三表达构建体的形式,其中所述第一启动子、所述第二启动子和所述第三启动子在所述真核细胞中具有活性。任选地,所述第一表达构建体、所述第二表达构建体和/或所述第三表达构建体在单一核酸分子上。

[0031] 在一些方法中,(a)编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式;且(b)编码所述CRISPR RNA的所述DNA和编码所述tracrRNA的所述DNA呈包含操作性连接编码包含所述CRISPR RNA和所述tracrRNA的gRNA的第三核酸的第二启动子的第二表达构建体的形式;其中所述第一启动子和所述第二启动子在所述真核细胞中具有活性。任选地,所述第一表达构建体和所述第二表达构建体在单一核酸分子上。

[0032] 在一些方法中,所述Cas蛋白、所述CRISPR RNA和所述tracrRNA作为蛋白质-RNA复合物引入所述真核细胞中。

[0033] 在一些方法中,所述靶向基因修饰同时包括在所述目标基因组基因座处缺失内源核酸序列和在所述目标基因组基因座处插入所述第一核酸。在一些方法中,所述缺失的内

源核酸序列为约30kb-约110kb,且所述插入的第一核酸为约40kb-约140kb。在一些方法中,所述缺失的内源核酸序列为约38kb-约110kb,且所述插入的第一核酸为约43kb-约134kb。

[0034] 在一些方法中,所述靶向基因修饰为双等位基因基因修饰。任选地,所述双等位基因基因修饰包括在两种同源染色体中在所述目标基因组基因座处内源核酸序列的缺失和所述第一核酸的插入。

[0035] 在一些方法中,所述修饰的真核细胞为在所述目标基因组基因座下杂合的化合物。在一些方法中,所述修饰的真核细胞在所述目标基因组基因座处为半合的。任选地,在一种染色体中在所述目标基因组基因座处的靶向基因修饰包括内源核酸序列的缺失和所述第一核酸的插入。任选地,所述靶向基因修饰包括:(1)在两种同源染色体中在所述目标基因组基因座处缺失内源核酸序列;和(2)在第一染色体中所述第一核酸插入所述目标基因组基因座中和在第二染色体中破坏所述目标基因组基因座。所述第一染色体可为所述两种同源染色体中的一种,且所述第二染色体可为另一同源染色体。

[0036] 在一些方法中,所述LTVEC为至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb或至少90kb。任选地,所述LTVEC为至少100kb、至少150kb或至少200kb。

[0037] 在一些方法中,所述第一核酸为至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少150kb、至少200kb、至少250kb或至少300kb。在一些方法中,所述第一核酸为约40kb-约140kb。在一些方法中,所述第一核酸为约43kb-约134kb。

[0038] 在一些方法中,所述真核细胞为哺乳动物细胞、成纤维细胞、多潜能细胞、非人类多潜能细胞、啮齿动物多潜能细胞、小鼠或大鼠胚胎干(ES)细胞、人类多潜能细胞、人类胚胎干(ES)细胞、人类成人干细胞、发育受限的人类祖细胞或人类诱导的多潜能干(iPS)细胞。

[0039] 在一些方法中,所述Cas蛋白为Cas9。在一些方法中,所述靶序列由原间隔区邻近基序(PAM)序列直接侧接。

[0040] 在一些方法中,所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约150kb。任选地,所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-150kb。

[0041] 在一些方法中,所述靶向基因修饰包括:(a)用同源或直系同源的核酸序列替换内源核酸序列;(b)缺失内源核酸序列;(c)缺失内源核酸序列,其中所述缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb;(d)插入外源核酸序列;(e)插入外源核酸序列,所述外源核酸序列为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb;(f)插入包含同源或直系同源的核酸序列的外源核酸序列;(g)插入包含人类核酸序列和非人类核酸序列的嵌合核酸序列;(h)插入侧接有位点特异性重组酶靶序列的条件性等位基因;(i)插入操作性连接在所述多潜能细胞中具有活性的第三启动子的可

选择标记或报道基因;或(j)其组合。

[0042] 在一些方法中,所述目标基因组基因座包含(i)与所述5'同源臂同源的5'靶序列;和(ii)与所述3'同源臂同源的3'靶序列。任选地,所述5'靶序列和所述3'靶序列相隔至少5kb,但小于3Mb。任选地,所述5'靶序列和所述3'靶序列相隔至少5kb,但小于10kb;至少10kb,但小于20kb;至少20kb,但小于40kb;至少40kb,但小于60kb;至少60kb,但小于80kb;至少约80kb,但小于100kb;至少100kb,但小于150kb;或至少150kb,但小于200kb;至少约200kb,但小于约300kb;至少约300kb,但小于约400kb;至少约400kb,但小于约500kb;至少约500kb,但小于约1Mb;至少约1Mb,但小于约1.5Mb;至少约1.5Mb,但小于约2Mb;至少约2Mb,但小于约2.5Mb;或至少约2.5Mb,但小于约3Mb。任选地,所述5'靶序列和所述3'靶序列相隔至少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。在一些方法中,所述5'靶序列和所述3'靶序列相隔约30kb-约110kb。在一些方法中,所述5'靶序列和所述3'靶序列相隔约38kb-约110kb。

[0043] 在一些方法中,所述目标基因组基因座包括所述干扰白细胞素-2受体 $\gamma$ 基因座、所述ApoE基因座、所述Rag1基因座、所述Rag2基因座或所述Rag1基因座和所述Rag2基因座两者。在其它方法中,所述目标基因组基因座包括Adams5基因座、Trpa1基因座、Folh1基因座或Erbb4基因座。在其它方法中,所述目标基因组基因座包括Lrp5基因座。在又其它方法中,所述目标基因组基因座包括C5(Hc)基因座、Ror1基因座或Dpp4基因座。

[0044] 进一步提供一种用于产生在目标基因组基因座处包括靶向基因修饰的F0代非人类动物的方法,所述方法包括:(a)在大靶向载体(LTVEC)存在下使在非人类ES细胞中的基因组与Cas蛋白、CRISPR RNA,和tracrRNA接触以形成修饰的非人类ES细胞,其中所述LTVEC为至少10kb且包含侧接有5'同源臂和3'同源臂的第一核酸;(b)鉴定在所述目标基因组基因座处具有所述靶向基因修饰的修饰的非人类ES细胞;(c)将所述修饰的非人类ES细胞引入非人类宿主胚胎中;和(d)在代孕母体中孕育所述非人类宿主胚胎,其中所述代孕母体产生在所述目标基因组基因座处包括所述靶向基因修饰的所述F0代非人类动物。

[0045] 在一些这样的方法中,所述CRISPR RNA和所述tracrRNA以单一向导RNA(gRNA)的形式一起引入。在其它这样的方法中,所述CRISPR RNA和所述tracrRNA单独地引入。

[0046] 在一些这样的方法中,(a)所述Cas蛋白以蛋白质、编码所述Cas蛋白的信使RNA(mRNA)或编码所述Cas蛋白的DNA的形式引入所述非人类ES细胞中;(b)所述CRISPR RNA以编码所述CRISPR RNA的RNA或DNA的形式引入所述非人类ES细胞中;且(c)所述tracrRNA以编码所述tracrRNA的RNA或DNA的形式引入所述非人类ES细胞中。

[0047] 在一些这样的方法中,(a)编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式;(b)编码所述CRISPR RNA的所述DNA呈包含操作性连接编码所述CRISPR RNA的第三核酸的第二启动子的第二表达构建体的形式;且(c)编码所述tracrRNA的所述DNA呈包含操作性连接编码所述tracrRNA的第四核酸的第三启动子的第三表达构建体的形式,其中所述第一启动子、所述第二启动子和所述第三启动子在所述非人类ES细胞中具有活性。任选地,所述第一表达构建体、所述第二表达构建体和所述第三表达构建体在单一核酸分子上。

[0048] 在一些这样的方法中,(a)编码所述Cas蛋白的所述DNA呈包含操作性连接编码所



述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式;且(b)编码所述CRISPR RNA的所述DNA和编码所述tracrRNA的所述DNA呈包含操作性连接编码包含所述CRISPR RNA和所述tracrRNA的gRNA的第三核酸的第二启动子的第二表达构建体的形式;其中所述第一启动子和所述第二启动子在所述非人类ES细胞中具有活性。任选地,所述第一表达构建体和所述第二表达构建体在单一核酸分子上。

[0049] 在一些这样的方法中,所述Cas蛋白、所述CRISPR RNA和所述tracrRNA作为蛋白质-RNA复合体引入所述非人类ES细胞中。

[0050] 在一些这样的方法中,所述靶向基因修饰同时包括在所述目标基因组基因座处缺失内源核酸序列和在所述目标基因组基因座处插入所述第一核酸。

[0051] 在一些这样的方法中,所述靶向基因修饰为双等位基因基因修饰。任选地,所述双等位基因基因修饰包括在两种同源染色体中在所述目标基因组基因座处内源核酸序列的缺失和所述第一核酸的插入。

[0052] 在一些这样的方法中,所述修饰的非人类ES细胞为在所述目标基因组基因座处杂合的化合物。在一些这样的方法中,所述修饰的非人类ES细胞在所述目标基因组基因座处为半合的。任选地,在一种染色体中在所述目标基因组基因座处的靶向基因修饰包括内源核酸序列的缺失和所述第一核酸的插入。任选地,所述靶向基因修饰包括:(1)在两种同源染色体中在所述目标基因组基因座处缺失内源核酸序列;和(2)在第一染色体中所述第一核酸插入所述目标基因组基因座中和在第二染色体中破坏所述目标基因组基因座。所述第一染色体可为所述两种同源染色体中的一种,且所述第二染色体可为另一同源染色体。

[0053] 在一些这样的方法中,所述Cas蛋白为Cas9。

[0054] 进一步提供一种用于修饰在真核细胞、小鼠细胞或人类细胞中的目标基因组基因座处的基因组的方法,其包括使所述基因组与Cas蛋白、杂化到在所述目标基因组基因座处的靶序列的CRISPR RNA,和tracrRNA在大靶向载体(LTVEC)存在下接触,其中所述LTVEC为至少10kb且包含侧接有与在所述目标基因组基因座处的5'靶序列同源的5'同源臂和与在所述目标基因组基因座处的3'靶序列同源的3'同源臂的第一核酸,其中所述第一核酸为至少30kb和/或所述5'靶序列和所述3'靶序列相隔至少30kb,其中在所述LTVEC存在下与所述Cas蛋白、所述CRISPR RNA和所述tracrRNA接触之后,所述基因组被修饰以包含包括在所述目标基因组基因座处插入所述第一核酸的靶向基因修饰。

[0055] 上述方法中的任一种可进一步包括将所述Cas蛋白、所述CRISPR RNA、所述tracrRNA和所述LTVEC引入所述真核细胞、所述小鼠细胞或所述人类细胞中。上述方法中的任一种可进一步包括鉴定在所述目标基因组基因座处包括所述靶向基因修饰的所述修饰的真核细胞、所述修饰的小鼠细胞或所述修饰的人类细胞。

[0056] 在一些上述方法中,所述CRISPR RNA和所述tracrRNA以单一转录体的形式一起引入。在一些上述方法中,所述CRISPR RNA和所述tracrRNA单独地引入。

[0057] 在一些上述方法中,(a)所述Cas蛋白以蛋白质、编码所述Cas蛋白的信使RNA(mRNA)或编码所述Cas蛋白的DNA的形式引入所述真核细胞、所述小鼠细胞或所述人类细胞中;(b)所述CRISPR RNA以编码所述CRISPR RNA的RNA或DNA的形式引入所述真核细胞、所述小鼠细胞或所述人类细胞中;且(c)所述tracrRNA以编码所述tracrRNA的RNA或DNA的形式引入所述真核细胞、所述小鼠细胞或所述人类细胞中。在一些上述方法中,所述Cas蛋白、所

述CRISPR RNA和所述tracrRNA作为蛋白质-RNA复合体引入所述真核细胞、所述小鼠细胞或所述人类细胞中。

[0058] 在一些上述方法中, (a) 编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式; (b) 编码所述CRISPR RNA的所述DNA呈包含操作性连接编码所述CRISPR RNA的第三核酸的第二启动子的第二表达构建体的形式; 且 (c) 编码所述tracrRNA的所述DNA呈包含操作性连接编码所述tracrRNA的第四核酸的第三启动子的第三表达构建体的形式; 其中所述第一启动子、所述第二启动子和所述第三启动子在所述真核细胞、所述小鼠细胞或所述人类细胞中具有活性。在一些上述方法中, 所述第一表达构建体、所述第二表达构建体和/或所述第三表达构建体在单一核酸分子上。

[0059] 在一些上述方法中, (a) 编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式; 且 (b) 编码所述CRISPR RNA的所述DNA和编码所述tracrRNA的所述DNA呈包含操作性连接编码包含在单一转录体中的所述CRISPR RNA和所述tracrRNA的gRNA的第三核酸的第二启动子的第二表达构建体的形式; 其中所述第一启动子和所述第二启动子在所述真核细胞、所述小鼠细胞或所述人类细胞中具有活性。在一些上述方法中, 所述第一表达构建体和所述第二表达构建体在单一核酸分子上。

[0060] 在一些上述方法中, 所述LTVEC为至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb或至少90kb。在一些上述方法中, 所述LTVEC为至少100kb、至少150kb或至少200kb。

[0061] 在一些上述方法中, 所述第一核酸为至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少150kb、至少200kb、至少250kb或至少300kb。在一些上述方法中, 所述第一核酸为约40kb-约140kb。

[0062] 在一些上述方法中, 所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约150kb。在一些上述方法中, 所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-150kb。

[0063] 在一些上述方法中, 所述5'靶序列和所述3'靶序列相隔至少5kb, 但小于3Mb。在一些上述方法中, 所述5'靶序列和所述3'靶序列相隔至少5kb, 但小于10kb; 至少10kb, 但小于20kb; 至少20kb, 但小于40kb; 至少40kb, 但小于60kb; 至少60kb, 但小于80kb; 至少约80kb, 但小于100kb; 至少100kb, 但小于150kb; 或至少150kb, 但小于200kb; 至少约200kb, 但小于约300kb; 至少约300kb, 但小于约400kb; 至少约400kb, 但小于约500kb; 至少约500kb, 但小于约1Mb; 至少约1Mb, 但小于约1.5Mb; 至少约1.5Mb, 但小于约2Mb; 至少约2Mb, 但小于约2.5Mb; 或至少约2.5Mb, 但小于约3Mb。在一些上述方法中, 所述5'靶序列和所述3'靶序列相隔至少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。在一些上述方法中, 所述5'靶序列和所述3'靶序列相隔约30kb-约110kb。

[0064] 在一些上述方法中, 所述真核细胞不是大鼠细胞。在一些上述方法中, 所述真核细胞为多潜能细胞、非多潜能细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、啮齿动物

细胞、小鼠细胞、仓鼠细胞、非人类多潜能细胞、人类多潜能细胞、啮齿动物多潜能细胞或成纤维细胞。在一些上述方法中,所述真核细胞为原代细胞或永生化细胞。在一些上述方法中,所述啮齿动物多潜能细胞为小鼠或大鼠胚胎干(ES)细胞。

[0065] 在一些上述方法中,所述小鼠细胞或所述人类细胞为原代细胞或永生化细胞。在一些上述方法中,所述小鼠细胞或所述人类细胞为多潜能细胞。在一些上述方法中,所述小鼠多潜能细胞为小鼠胚胎干(ES)细胞。在一些上述方法中,所述人类多潜能细胞为人类胚胎干(ES)细胞、人类成人干细胞、发育受限的人类祖细胞或人类诱导的多潜能干(iPS)细胞。在一些上述方法中,所述人类iPS细胞维持在包含碱介质和补充液的培养基中,其中所述培养基包含:(a)白血病抑制因子(LIF)多肽;(b)肝糖合成酶激酶(GSK3)抑制剂;和(c)MEK抑制剂;其中所述培养基具有约175mOsm/kg-约280mOsm/kg的渗透压度。

[0066] 在一些上述方法中,所述Cas蛋白为Cas9。在一些上述方法中,所述靶序列由原间隔区邻近基序(PAM)序列直接侧接。

[0067] 在一些上述方法中,所述靶向基因修饰同时包括在单一步骤中在所述目标基因组基因座处缺失内源核酸序列和在所述目标基因组基因座处插入所述第一核酸。在一些上述方法中,所述缺失的内源核酸序列为约30kb-约110kb,且所述插入的第一核酸为约40kb-约140kb。

[0068] 在一些上述方法中,所述靶向基因修饰为双等位基因基因修饰。在一些上述方法中,所述双等位基因基因修饰包括在两种同源染色体中在所述目标基因组基因座处内源核酸序列的缺失和所述第一核酸的插入。在一些上述方法中,所述修饰的真核细胞、所述修饰的小鼠细胞或所述修饰的人类细胞为在所述目标基因组基因座处杂合的化合物。在一些上述方法中,所述修饰的真核细胞、所述修饰的小鼠细胞或所述修饰的人类细胞在所述目标基因组基因座处为半合的。在一些上述方法中,在一种染色体中在目标基因组基因座处的靶向基因修饰包括内源核酸序列的缺失和所述第一核酸的插入。在一些上述方法中,所述靶向基因修饰包括:(1)在第一同源染色体和第二同源染色体中在所述目标基因组基因座处缺失内源核酸序列;和(2)所述第一核酸插入在所述第一同源染色体中的目标基因组基因座中并破坏在所述第二同源染色体中的目标基因组基因座。

[0069] 在一些上述方法中,所述靶向基因修饰包括:(a)用同源或直系同源的核酸序列替换内源核酸序列;(b)缺失内源核酸序列;(c)缺失内源核酸序列,其中所述缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb;(d)插入外源核酸序列;(e)插入外源核酸序列,所述外源核酸序列为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb;(f)插入包含同源或直系同源的核酸序列的外源核酸序列;(g)插入包含人类核酸序列和非人类核酸序列的嵌合核酸序列;(h)插入侧接有位点特异性重组酶靶序列的条件性等位基因;(i)插入操作性连接在所述多潜能细胞中具有活性的启动子的可选择标记或报道基因;或(j)其组合。

[0070] 在一些上述方法中,所述目标基因组基因座包括干扰白细胞素-2受体 $\gamma$ 基因座、

ApoE基因座、Rag1基因座、Rag2基因座、所述Rag1基因座和所述Rag2基因座两者、Adamts5基因座、Trpa1基因座、Folh1基因座、ErbB4基因座、Lrp5基因座、C5(Hc)基因座、Ror1基因座或Dpp4基因座。在一些上述方法中,所述目标基因组基因座包含染色体外DNA。

[0071] 还提供一种用于产生在目标基因组基因座处包括靶向基因修饰的F0代非人类动物或小鼠的方法,所述方法包括:(a)使用上述方法中的任一种修饰非人类或小鼠ES细胞;(b)鉴定在所述目标基因组基因座处包括所述靶向基因修饰的修饰的非人类或小鼠ES细胞;(c)将所述修饰的非人类或小鼠ES细胞引入非人类或小鼠宿主胚胎中;和(d)在代孕母体中孕育所述非人类或小鼠宿主胚胎,其中所述代孕母体产生在所述目标基因组基因座处包括所述靶向基因修饰的所述F0代非人类动物或小鼠。

## 附图说明

[0072] 图1描绘大鼠ESC,其作为通常在培养皿中脱离并漂浮的紧实球形集落生长。

[0073] 图2A至2D描绘由大鼠ESC表达的各种多潜能标记:A描绘Oct-4(绿色);B描绘Sox-2(红色);C描绘DAPI(蓝色);D描绘由rESC表达的多潜能标记的重叠。

[0074] 图3描绘大鼠ESC表达各亮度水平的碱性磷酸酶(多潜能标记)。

[0075] 图4描绘株系DA.2B的核型分析,其为42X,Y。因为大鼠ESC常变成四倍体,所以进行核型分析;株系因此通过计数中期染色体散布来预先筛选,且随后对具有大部分正常计数的株系正式分析核型。

[0076] 图5A-B提供示出ACI.G1大鼠ES细胞系的染色体数的分析的照片。

[0077] 图6A-B提供示出DA.2B大鼠ES细胞系的染色体数的分析的照片。

[0078] 图7A-B提供示出DA.2C大鼠ES细胞系的染色体数的分析的照片。

[0079] 图8描绘图1的大鼠ESC的近视图。

[0080] 图9描绘通过胚胎注射来产生嵌合体以及通过种系传递大鼠ESC基因组。使用亲本ACI.G1大鼠ESC通过胚胎注射产生嵌合体。高百分数嵌合体通常具有白化鼻部。

[0081] 图10描绘由在图9中用星号(\*)标记的ACI/SD嵌合体所生的F1白化刺豚鼠同窝幼仔。

[0082] 图11提供大鼠ApoE基因座的示意图,并且用灰条指示锌指核酸酶(ZFN1和ZFN2)的切割位点。对应于5'同源臂和3'同源臂的基因组区(分别是5kb和5.4kb)通过暗灰色框来指示。ApoE基因的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE基因的三个内含子指示为直线。外显子2和3包含编码区且显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。

[0083] 图12描绘大鼠Rosa26基因座的靶向,如同在小鼠中一样,所述基因座以相同间隔位于Setd5基因与Thumpd3基因之间。版面A示出小鼠Rosa26基因座的结构。小鼠Rosa26转录体由2个或3个外显子组成。版面B描绘大鼠Rosa26基因座的结构;除与小鼠外显子1同源的外显子(Ex1a)之外,大鼠基因座还含有第二外显子1(Ex1b);尚未在大鼠中鉴定出第三外显子。版面C描绘靶向大鼠Rosa26等位基因;使用来自DA rESC的基因组DNA通过PCR来克隆各自具有5kb的同源臂;靶向等位基因含有替换在大鼠Rosa26内含子中的117bp缺失的剪接受体(SA)-lacZ-hUB-neo盒。

[0084] 图13A描绘14周龄野生型大鼠的对照脑,其用X-gal染色。对照脑示出低水平的背

景LacZ染色(背视图)。

[0085] 图13B描绘rRosa26杂合大鼠(14周龄)的脑中的LacZ表达。lacZ报道基因在rRosa26杂合子的整个脑中遍在表达。

[0086] 图13C描绘14周龄野生型大鼠的对照心脏和胸腺(插图),其用X-gal处理。对照心脏和胸腺示出低水平的背景LacZ染色。

[0087] 图13D描绘14周龄rRosa26杂合大鼠的心脏和胸腺(插图)中的LacZ表达。lacZ报道基因在rRosa26杂合子的整个心脏和胸腺中遍在表达。

[0088] 图13E描绘14周龄野生型大鼠的对照肺,其用X-gal处理。对照肺示出低水平的背景LacZ染色。

[0089] 图13F描绘在14周龄rRosa26杂合子大鼠的肺中的LacZ表达。lacZ报道基因在rRosa26杂合子的整个肺中遍在表达。

[0090] 图13G和13H描绘在E12.5大鼠胚胎中的LacZ表达。与显示低水平的背景LacZ染色的野生型对照胚胎(H)形成对比,rRosa26杂合胚胎表现出LacZ报道基因在整个胚胎中遍在表达。

[0091] 图13I和13J描绘在E14.5大鼠胚胎中的LacZ表达。与显示低水平的背景LacZ染色的野生型对照胚胎(J)形成对比,rRosa26杂合大鼠胚胎表现出LacZ报道基因在整个胚胎中遍在表达。

[0092] 图14说明在包含选择盒(lacZ-neo盒)的靶向载体的电穿孔之后在大鼠ES细胞内部发生的同源或非同源重组事件。

[0093] 图15说明基因组编辑核酸内切酶(例如,ZFN和TALEN)在靶基因组序列中引入双链断裂(DSB)以及激活在ES细胞中的非同源末端接合(NHEJ)所依的机理。

[0094] 图16说明利用ZFN/TALEN来改进靶向载体的同源重组的效率的基因靶向技术。DSB代表双链断裂。

[0095] 图17示出通过修饰的大鼠ApoE基因座的嵌合体产生和种系传递产生的ApoE-ZFN-AB5嵌合体。靶向修饰由锌指核酸酶辅助。

[0096] 图18提供与靶向ZFN U和ZFN D的锌指核酸酶组合的IL2r- $\gamma$ 靶向事件的示意图。示出由ZFN U和ZFN D靶向的大鼠IL2r- $\gamma$ 基因座的区(SEQ ID NO:93)。ZFN切割位点在图中指出。

[0097] 图19提供与靶向ZFN U和ZFN D或与gRNA(gRNA1、gRNA2、gRNA3、gRNA4)组合的锌指核酸酶组合的IL2r- $\gamma$ 靶向事件的示意图。示出由ZFN U和ZFN D或gRNA1-4靶向的大鼠IL2r- $\gamma$ 基因座的区且指出ZFN切割位点。

[0098] 图20提供大鼠ApoE基因座和靶向质粒的示意图。上部示意图示出大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区(分别是5kb和5.4kb;暗灰色框)的基因组结构。ApoE基因的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE基因的三个内含子指示为直线。外显子2和3包含编码区且显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。下部版面示出靶向质粒。5'同源臂和3'同源臂(分别是5kb和5.4kb)由暗灰色框指示。靶向载体包含报道基因(lacZ)和由loxP位点(空心箭头)侧接的自缺失性盒。自缺失性盒包含操作性连接Crei基因的小鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0099] 图21A提供使用锌指核酸酶和靶向载体来靶向在大鼠ES细胞中的ApoE基因座的示意图,所述靶向载体包含报道基因(LacZ)和自缺失性盒,所述自缺失性盒包含操作性连接Crei基因的小鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。图21B描绘纯合靶向的ApoE基因座。

[0100] 图22提供大鼠ApoE基因座和大靶向载体(LTVEC)的示意图。上部版面示出大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区(分别是45kb和23kb;暗灰色框)的基因组结构。ApoE的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE基因的三个内含子指示为直线,并且外显子2和3包含编码区并显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。下部版面示出用于修饰大鼠ApoE基因座的LTVEC。5'同源臂和3'同源臂(分别是45kb和23kb)通过暗灰色框来指示。LTVEC包含报道基因(lacZ)和由loxP位点(空心箭头)侧接的自缺失性盒,所述自缺失性盒包含操作性连接Crei基因的小鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0101] 图23提供大鼠ApoE基因座的示意图且用灰条指示连同大靶向载体(LTVEC)一起用于增强靶向载体与靶同源染色体区之间的同源重组的锌指核酸酶(ZFN1和ZFN2)的切割位点。

[0102] 图24描绘已通过3.2kb缺失以及插入报道基因(eGFP)和自缺失性盒来破坏的大鼠IL2r- $\gamma$ 基因座,所述自缺失性盒包含药物选择盒(hUb-neo)和操作性连接小鼠Prm1启动子的Crei基因。

[0103] 图25提供已通过3.2kb缺失以及插入报道基因(eGFP)和自缺失性盒来破坏的大鼠IL2r- $\gamma$ 基因座的另一绘图,所述自缺失性盒包含操作性连接小鼠Prm1启动子的Crei基因和药物选择盒(hUb-neo)。

[0104] 图26提供大鼠Rag2基因座和用于修饰大鼠Rag2基因座的大靶向载体(LTVEC)的示意图。上部版面示出大鼠Rag2基因座和对应于5'同源臂和3'同源臂的同源基因组区(分别是48kb和84kb;暗灰色框)的基因组结构。Rag2包含通过点描灰色阴影来指示的单一外显子。下部版面为LTVEC。5'同源臂和3'同源臂(分别是48kb和84kb)通过暗灰色框来指示。LTVEC包含报道基因(lacZ)和由loxP位点(空心箭头)侧接的自缺失性盒,所述自缺失性盒含有操作性连接Crei基因的大鼠Prm1启动子和含有操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0105] 图27提供大鼠Rag1/Rag2基因座的基因组结构以及通过Rag2靶向(Rag2缺失)或Rag2/Rag1双重靶向(Rag2/Rag1缺失)缺失的基因组区。

[0106] 图28提供大鼠Rag2基因座和Rag1基因座及用于修饰基因座的大靶向载体(LTVEC)的示意图。上部版面示出Rag1和Rag2基因座和对应于5'同源臂和3'同源臂的同源基因组区(分别是48kb和15kb;暗灰色框)的基因组结构。Rag2和Rag1各自包含由点描灰色阴影指示的单一外显子。下部版面为LTVEC。5'同源臂和3'同源臂(分别是48kb和15kb)通过暗灰色框来指示。LTVEC包含报道基因(lacZ)和由loxP位点(空心箭头)侧接的自缺失性盒,所述自缺失性盒包含操作性连接Crei基因的大鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0107] 图29示出对于在来自II2rg<sup>-/-</sup>y嵌合大鼠(版面A-C)和WT DA大鼠(版面D-F)的外周

血液单核细胞 (PBMC) 中的GFP表达和T细胞标记CD3 (版面A和D)、B细胞标记B220 (版面B和E) 和NK细胞标记CD161a (版面C和F) 的流式细胞分析。双重阳性细胞在象限R8中示出。图29示出II2rg<sup>-</sup>/y PBMC不表达成熟淋巴细胞标记。

[0108] 图30示出GFP阳性淋巴细胞在三个II2rg<sup>-</sup>/y嵌合体中的2个中的外周血液中检测到。

[0109] 图31提供大鼠I12rg基因座和用于大鼠I12rg基因座的完全人源化的靶向质粒的示意图。上部版面示出大鼠I12rg基因座和对应于5'同源臂和3'同源臂的同源基因组区 (分别是4.3kb和4.0kb; 暗灰色框) 的基因组结构。下部版面为靶向质粒。5'同源臂和3'同源臂 (分别是4.3kb和4.0kb) 由暗灰色框指示。靶向质粒包含人类IL-2rg基因组区和由loxP位点 (空心箭头) 侧接的缺失盒, 所述缺失盒含有含有操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0110] 图32提供大鼠I12rg基因座和用于大鼠I12rg基因座的胞外结构域人源化的靶向质粒的示意图。上部版面示出大鼠I12rg基因座和对应于5'同源臂和3'同源臂的同源基因组区 (分别是4.3kb和4.0kb; 暗灰色框) 的基因组结构。下部版面为靶向质粒。5'同源臂和3'同源臂 (分别是4.3kb和4.0kb) 由暗灰色框指示。所述靶向质粒包含IL-2Rg基因组区的人类胞外结构域和由loxP位点 (空心箭头) 侧接的自缺失性盒, 所述自缺失性盒含有操作性连接Crei基因的大鼠Prm1启动子和含有操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0111] 图33提供人类IL-2rg蛋白 (SEQ ID NO:20; NP\_000197.1); 大鼠IL-2rg蛋白 (SEQ ID NO:21; NP\_543165.1); 以及包含融合于大鼠IL-2rg蛋白的其余部分的IL-2rg的人类胞外结构域的嵌合IL-2rg蛋白 (SEQ ID NO:22) 的序列比对。在人类IL-2rg和大鼠IL-2rg之间的接合点通过垂直线标注。

[0112] 图34提供小鼠Lrp5基因的CRISPR/Cas9辅助人源化的示意图; LTVEC示于顶部版面中且小鼠Lrp5基因座示于底部版面中。人源化的区为胞外域。箭头指示各gRNA (gA、gB、gB2、gC、gD、gE2、gE、gF) 和ZFN (a-d) 的靶位点。

[0113] 图35描绘增加缺失尺寸的LTVEC靶向基因的%靶向效率 (图35A) 和具有增加尺寸的人类基因插入的LTVEC (图35B)。LTVEC单独使用 (灰色方块或三角形) 或与ZFN组合使用 (黑格方块或三角形)。

[0114] 图36提供小鼠Trpa1基因的整体编码区的CRISPR/Cas9辅助人源化的示意图; LTVEC示于顶部版面中且小鼠Trpa1基因座示于底部版面中。箭头指示各gRNA (gA、gA2、gB、gC、gD、gE2、gE、gF) 的靶位点。

[0115] 图37提供小鼠Fo1h1基因的胞外域 (外显子2-终止密码子) 的CRISPR/Cas9辅助人源化的示意图; LTVEC示于顶部版面中且小鼠Fo1h1基因座示于底部版面中。箭头指示各gRNA (gA、gA2、gB、gC、gD、gE、gE2、gF) 的靶位点。

[0116] 图38提供小鼠C5 (Hc) 基因的从外显子2到终止密码子的区的CRISPR/Cas9辅助人源化的示意图; LTVEC示于顶部版面中且小鼠C5 (Hc) 基因座示于底部版面中。箭头指示各gRNA (gA、gB、gB2、gC、gD、gE2、gE、gF) 的靶位点。

[0117] 图39提供小鼠Adamts5基因的整体编码区的CRISPR/Cas9辅助人源化的示意图; LTVEC示于顶部版面中且小鼠Adamts5基因座示于底部版面中。箭头指示各gRNA (gA、gA2、

gB、gC、gD、gE2、gE、gF)的靶位点。

[0118] 图40提供小鼠ErbB4基因的外显子4-15的CRISPR/Cas9辅助人源化的示意图;LTVEC示于顶部版面中且小鼠ErbB4基因座示于底部版面中。箭头指示各gRNA(gA、gB、gB2、gC、gD、gE2、gE、gF)的靶位点。

[0119] 图41提供小鼠Ror1基因的外显子2-7的CRISPR/Cas9辅助人源化的示意图;LTVEC示于顶部版面中且小鼠Ror1基因座示于底部版面中。箭头指示各gRNA(gA、gB、gC、gD、gE、gF)的靶位点。

[0120] 图42提供小鼠Dpp4基因的从外显子2到终止密码子的区的CRISPR/Cas9辅助人源化的示意图;LTVEC示于顶部版面中且小鼠Dpp4基因座示于底部版面中。箭头指示各gRNA(gA、gB、gB2、gC、gD、gE2、gE、gF)的靶位点。

[0121] 图43示出用X-gal染色的12周龄雌性大鼠脑。图43A-C示出来自野生型大鼠的脑,且图43D-F示出来自ApoE<sup>+/-</sup>大鼠的脑。图43A和D示出背视图,图43B和E示出腹视图,且图43C和F示出近视图。

[0122] 图44示出用X-gal染色的12周龄雌性大鼠心脏(A和C)和血管(B和D)的相应近视图。图44A和44B分别示出来自野生型大鼠的心脏和血管,且图44C和44D分别示出来自ApoE<sup>+/-</sup>大鼠的心脏和血管。染色存在于心脏的心房和一些血管(例如,腔静脉)中。

[0123] 图45示出用X-gal染色的12周龄雌性大鼠肝。图45A和45B示出来自野生型大鼠的肝,且图45C和45D示出来自ApoE<sup>+/-</sup>大鼠的肝。图45B和45D为肝的近视图。

[0124] 图46示出在6周、9周、12周和15周的纯合ApoE靶向大鼠、杂合ApoE靶向大鼠和野生型大鼠中胆固醇、LDL、HDL和三酸甘油酯水平的检测(分别地,图46A-D)。

[0125] 图47示出大鼠ApoE基因座(上部版面)和靶向大鼠ApoE基因座的大靶向载体(LTVEC)(下部版面)的示意图。上部版面示出大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区(分别是45kb和23kb;暗灰色框)的基因组结构。ApoE的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE基因的三个内含子指示为直线,并且外显子2和3包含编码区并显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。指示出ApoE gRNA2(SEQ ID NO:87)和ApoE gRNA3(SEQ ID NO:88)的靶位点。下部版面示出用于修饰大鼠ApoE基因座的LTVEC。5'同源臂和3'同源臂(分别是45kb和23kb)通过暗灰色框来指示。LTVEC包含报道基因(1acZ)和由1oxP位点(空心箭头)侧接的自缺失性盒,所述自缺失性盒包含操作性连接Crei基因的小鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0126] 图48示出大鼠Rag2基因座(上部版面)和靶向大鼠Rag2基因座的大靶向载体(LTVEC)(下部版面)的示意图。上部版面示出大鼠Rag2基因座和对应于5'同源臂和3'同源臂的同源基因组区(分别是48kb和84kb;暗灰色框)的基因组结构。Rag2包含通过点描灰色阴影来指示的单一外显子。指示出Rag2 gRNA1(SEQ ID NO:89)和Rag2 gRNA4(SEQ ID NO:90)的靶位点。下部版面为LTVEC。5'同源臂和3'同源臂(分别是48kb和84kb)通过暗灰色框来指示。LTVEC包含报道基因(1acZ)和由1oxP位点(空心箭头)侧接的自缺失性盒,所述自缺失性盒含有操作性连接Crei基因的大鼠Prm1启动子和含有操作性连接潮霉素抗性基因的人类泛素启动子的药物选择盒。

[0127] 图49示出大鼠I12rg基因座(上部版面)和用于大鼠I12rg基因座的胞外域人源化



的靶向质粒(下部版面)的示意图。上部版面示出大鼠I12rg基因座和对应于5'同源臂和3'同源臂的同源基因组区(分别是4.3kb和4.0kb;暗灰色框)的基因组结构。指示出I12rg gRNA2(SEQ ID NO:91)和I12rg gRNA4(SEQ ID NO:92)的靶位点。下部版面为靶向质粒。5'同源臂和3'同源臂(分别是4.3kb和4.0kb)由暗灰色框指示。所述靶向质粒包含IL-2Rg基因组区的人类胞外结构域和由loxP位点(空心箭头)侧接的自缺失性盒,所述自缺失性盒含有操作性连接Crei基因的大鼠Prm1启动子和含有操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0128] 图50示出大鼠Rag2和Rag1基因座和用于修饰在I12rg靶向大鼠ES细胞(克隆I12rg-CG12)中的基因座的大靶向载体(LTVEC)的示意图。上部版面示出Rag1和Rag2基因座和对应于5'同源臂和3'同源臂的同源基因组区(分别是48kb和15kb;暗灰色框)的基因组结构。Rag2和Rag1各自包含由无阴影的箭头指示的单一外显子。下部版面为LTVEC。5'同源臂和3'同源臂(分别是48kb和15kb)由暗灰色框指示。LTVEC包含报道基因(eGFP)和由内部核糖体进入位点(IRES)间隔且操作性连接肌动蛋白启动子的嘌呤霉素抗性基因。LTVEC进一步包含由loxP位点(空心箭头)侧接的自缺失性盒,所述自缺失性盒包含操作性连接Crei基因的大鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。

[0129] 图51描绘使用LTVEC和在人类iPS细胞中的向导RNA用包含小鼠Adam6a基因座和小鼠Adam6b基因座的核酸替换人类ADAM6基因座的一部分的示意图。向导RNA的靶位点由箭头指示。

[0130] 图52A描绘由在2i培养基中培养8天的人类iPS细胞显示的形态。图52B描绘由在2i培养基中培养12天的人类iPS细胞显示的形态。

[0131] 图53A-53D描绘在mTeSR<sup>TM</sup>-hLIF培养基或低渗透压度VG2i培养基中培养6天的人类iPS细胞的形态。图53A和53B描绘在mTeSR<sup>TM</sup>-hLIF培养基(图53A)或VG2i培养基(图53B)中培养6天的人类iPS细胞的形态。图53C和53D描绘在mTeSR<sup>TM</sup>-hLIF培养基(图53C)或VG2i培养基(图53D)中在新生人类包皮成纤维细胞(NuFF)饲养细胞上培养6天的人类iPS细胞的形态。

[0132] 图54A描绘已经对于碱性磷酸酶染色的在VG2i培养基中培养的重编程人类iPS细胞。图54B和54C描绘已经对于NANOG的表达免疫染色的在VG2i培养基中培养的重编程人类iPS细胞。

[0133] 图55A-55C说明在VG2i培养基中培养的重编程人类iPS细胞的酶分解和传代培养。图55A描绘在缺乏ROCK抑制剂的情况下用胰蛋白酶进行酶分解之前在VG2i培养基中培养的重编程人类iPS细胞。图55B描绘在传代培养之后1天在VG2i培养基中培养的人类iPS细胞。图55C描绘在传代培养之后4天在VG2i培养基中培养的人类iPS细胞。

## 具体实施方式

[0134] 提供组合物和方法以经由在原核细胞中的细菌同源重组(BHR)修饰大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠目标基因组基因座。还提供组合物和方法以使用大靶向载体(LTVEC)以及核酸内切酶基因修饰目标基因组基因座,例如大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物或小鼠目标基因组基因座。还提供组合物和方法以产生包括一种或多种靶向基因修饰的例如大鼠、小鼠、啮齿动物或非大鼠啮齿动物的基因修饰的非人类动物。

还提供分离的人类和非人类全能或多潜能干细胞,尤其是大鼠胚胎干细胞,其能够在一种或多种体外连续基因修饰之后维持多潜能性且能够经由种系将所述靶向基因修饰传递到后代。

[0135] 词汇表

[0136] 本文使用的术语“胚胎干细胞”或“ES细胞”包括在引入胚胎中后能够促进发育胚胎的任何组织的源自胚胎的全能或多潜能细胞。本文使用的术语“多潜能细胞”包括具有发育成多于一种类型的分化细胞的能力的未分化细胞。术语“非多潜能细胞”包括不是多潜能细胞的细胞。

[0137] 本文使用的术语“同源核酸”包括与已知参考序列相同或实质类似的核酸序列。在一个实施方案中,术语“同源核酸”用以表征具有与已知参考序列至少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%或甚至100%同一的氨基酸序列的序列。

[0138] 本文使用的术语“直系同源核酸”包括在功能上与在另一物质中的已知参考序列等效的来自一种物质的核酸序列。

[0139] 本文使用的术语“大靶向载体”或“LTVEC”包括源自克隆基因组DNA的片段的用于真核细胞的大靶向载体,其比由意图在真核细胞中执行同源基因靶向的其它方法通常使用的那些靶向载体大。LTVEC的实例包括但不限于细菌同源染色体(BAC)和酵母菌人造染色体(YAC)。

[0140] 本文使用的术语“等位基因修饰”(MOA)包括在基因组中一种或多种基因或染色体基因座(loci)的一种等位基因的精确DNA序列的修饰。如本文所述的“等位基因修饰(MOA)”的实例包括但不限于缺失、取代或插入仅仅单一核苷酸或跨一种或多种目标基因或染色体基因座缺失上千个碱基以及在这两个末端之间的任何和所有可能的修饰。

[0141] 本文使用的术语“重组位点”包括由位点特异性重组酶识别且可充当重组事件的底物的核苷酸序列。

[0142] “系列”基因修饰包括对细胞(例如,真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或中国仓鼠卵巢(CHO)细胞)独立地实施的两种或多种修饰。第一修饰可通过电穿孔或本领域已知的任何其它方法实现。随后,采用合适的第二核酸构建体对同一细胞基因组进行第二修饰。所述第二修饰可通过第二电穿孔或本领域已知的任何其它方法实现。在各种实施方案中,在同一细胞的第一基因修饰和第二基因修饰之后,可使用例如连续电穿孔或本领域已知的任何其它合适方法(连续地)实现第三基因修饰、第四基因修饰、第五基因修饰、第六基因修饰等连续基因修饰(一种基因修饰跟着另一基因修饰)。

[0143] 本文使用的术语“位点特异性重组酶”包括可促进在“重组位点”之间的重组的一组酶,其中所述两个重组位点在单一核酸分子内或在单独的核酸分子上实体间隔开。“位点特异性重组酶”的实例包括但不限于Cre、Flp和Dre重组酶。

[0144] 关于核酸序列的术语“种系”包括可传到子代的核酸序列。

[0145] 短语“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。除非另有指明,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。典型的重链具有在可变结构域(N-末端到C-末端)之后的 $C_H1$ 结构域、铰链、 $C_H2$ 结构域和 $C_H3$ 结构域。重链的功能片段包括能够特异性识别能够自细胞表达并分泌且包含至少一个CDR的表位(例如,以微摩尔、纳摩尔或皮摩尔范围的 $K_D$ 识别表位)的片段。重链可变结构域由可变区核苷酸序列编码,其通常包含源自在种系中存在的 $V_H$ 、 $D_H$ 和 $J_H$ 片段的清单的 $V_H$ 、 $D_H$ 和 $J_H$ 片段。各种生物体的V、D和J重链片段的序列、定位和命名可在IMGT数据库中见到,其可经由因特网在环球网(www)上在URL“imgt.org.”上访问。

[0146] 短语“轻链”包括来自任何生物体的免疫球蛋白轻链序列,且除非另作说明,否则包括人类卡巴( $\kappa$ )和拉姆达( $\lambda$ )轻链和VpreB以及替代轻链。除非另作说明,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,全长轻链包括从氨基端到羧基端点的包括FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4的可变结构域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,其通常包含源自在种系中存在的轻链V和J基因片段的清单的轻链 $V_L$ 和轻链 $J_L$ 基因片段。各种生物体的轻链V和J基因片段的序列、定位和命名可在IMGT数据库中见到,其可经由因特网在环球网(www)上在URL“imgt.org.”上访问。轻链包括例如不会选择性地结合由在其中呈现其的表位-结合蛋白选择性结合的第一表位或第二表位的那些轻链。轻链还包括结合并识别或帮助重链结合并识别由在其中呈现其的表位-结合蛋白选择性结合的一种或多种表位的那些轻链。

[0147] 短语“操作性连接”包括其中组分以其预定方式操作性连接起作用的关系。在一种情况下,编码蛋白的核酸序列可操作性连接到调控序列(例如,启动子、增强子、沉默子序列等)以保持恰当的转录调控。在一种情况下,免疫球蛋白可变区(或V(D)J片段)的核酸序列可操作性连接到免疫球蛋白恒定区的核酸序列,以允许在序列之间恰当重组到免疫球蛋白重链或轻链序列。

[0148] 1. 包含核酸的靶基因座

[0149] 提供各种方法和组合物,其允许在靶基因座处整合至少一种插入核酸。本文使用的“目标基因组基因座”包括希望整合插入核酸的在基因组内的DNA的任何片段或区。术语“目标基因组基因座”和“目标靶基因组基因座”可互换使用。目标基因组基因座对细胞可为天然的,或者可选地,可包含整合到细胞的基因组中的DNA的异源或外源片段。DNA的所述异源或外源片段可包括转基因、表达盒、编码选择制造体的多核苷酸或基因组DNA的异源或外源区。术语“基因座”在本文中定义为在基因组DNA内的DNA的片段。如本文所述的基因修饰可包括自目标基因座的一种或多种缺失、目标基因座的添加、目标基因座的替换和/或其任意组合。目标基因座可包括编码区或非编码调控区。

[0150] 所述目标基因组基因座可进一步包含靶向整合系统的任何组分,所述组分包括例如识别位点、选择标记、先前整合的插入核酸、编码核酸酶试剂的多核苷酸、启动子等。可选地,所述目标基因组基因座可定位在所述细胞内的染色体外DNA内,例如酵母菌人造染色体(YAC)、细菌人造染色体(BAC)、人类人造染色体或在适当宿主细胞中包含的任何其它工程化的基因组区。在各种实施方案中,所述靶向基因座可包括来自原核生物、真核生物、非大鼠真核生物、酵母菌、细菌、非人类哺乳动物、非人类细胞、啮齿动物、非大鼠啮齿动物、人类、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、鸡、猫、狗、白鼬、灵长类动物(例如,狨猴、恒

河猴)、驯化哺乳动物或农业哺乳动物或任何其它目标生物体或其组合的天然、异源或外源核酸序列。在一些实施方案中,所述目标基因组基因座包含来自人类、小鼠或其组合的核酸序列。

[0151] 在特定的实施方案中,所述靶基因座例如来自真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞。

[0152] 在特定的实施方案中,所述目标基因组基因座包括“大鼠核酸”的靶基因座。这一区域包含整合在细胞的基因组内的来自大鼠的核酸。所述靶基因座的非限制性实例包括编码在B细胞中表达的蛋白质的基因组基因座、表达在未成熟B细胞中的多肽的基因组基因座、表达在成熟B细胞中的多肽的基因组基因座、免疫球蛋白(Ig)基因座或包括例如T细胞受体 $\alpha$ 基因座的T细胞受体基因座。靶基因组基因座的另外实例包括Fcer1a基因座、Tlr4基因座、Pr1r基因座、Notch4基因座、Accn2基因座、Adamts5基因座、Trpa1基因座、Folh1基因座、Lrp5基因座、IL2受体基因座,包括例如IL2受体 $\gamma$ (Il2rg)基因座、ApoE基因座、Rag1基因座、Rag2基因座、Rag1/Rag2基因座和ErbB4基因座。任何这样的靶基因座可来自大鼠或者可来自真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞或非人类哺乳动物细胞。

[0153] 在一个实施方案中,所述靶基因座编码哺乳动物免疫球蛋白重链可变区氨基酸序列。在一个实施方案中,所述靶基因座编码大鼠免疫球蛋白重链可变区氨基酸序列。在一个实施方案中,所述靶基因座包含基因组DNA序列,所述基因组DNA序列包含操作性连接免疫球蛋白重链恒定区核酸序列的未重排的大鼠、小鼠或人类免疫球蛋白重链可变区核酸序列。在一个实施方案中,所述免疫球蛋白重链恒定区核酸序列为选自CH1、铰链、CH2、CH3及其组合的大鼠、小鼠或人类免疫球蛋白重链恒定区核酸序列。在一个实施方案中,所述重链恒定区核酸序列包含CH1-铰链-CH2-CH3。在一个实施方案中,所述靶基因座包含操作性连接免疫球蛋白重链恒定区核酸序列的未重排的大鼠、小鼠或人类免疫球蛋白重链可变区核酸序列。在一个实施方案中,所述免疫球蛋白重链恒定区核酸序列为选自CH1、铰链、CH2、CH3及其组合的大鼠、小鼠或人类免疫球蛋白重链恒定区核酸序列。在一个实施方案中,所述重链恒定区核酸序列包含CH1-铰链-CH2-CH3。

[0154] 在一个实施方案中,所述靶基因座包含编码哺乳动物免疫球蛋白轻链可变区氨基酸序列的基因组DNA序列。在一个实施方案中,所述基因组DNA序列包含未重排的哺乳动物 $\lambda$ 和/或 $\kappa$ 轻链可变区核酸序列。

[0155] 在一个实施方案中,所述基因组DNA序列包含重排的哺乳动物 $\lambda$ 和/或 $\kappa$ 轻链可变区核酸序列。在一个实施方案中,所述未重排的 $\lambda$ 或 $\kappa$ 轻链可变区核酸序列操作性连接选自 $\lambda$ 轻链恒定区核酸序列和 $\kappa$ 轻链恒定区核酸序列的哺乳动物免疫球蛋白轻链恒定区核酸序列。在一个实施方案中,所述哺乳动物免疫球蛋白轻链恒定区核酸序列为大鼠免疫球蛋白轻链恒定区核酸序列。在一个实施方案中,所述哺乳动物免疫球蛋白轻链恒定区核酸序列为小鼠免疫球蛋白轻链恒定区核酸序列。在一个实施方案中,所述哺乳动物免疫球蛋白轻链恒定区核酸序列为人类免疫球蛋白轻链恒定区核酸序列。

[0156] 如在本文中所用,ApoE基因座、干扰白细胞素-2受体 $\gamma$ (Il2rg)基因座、Rag2基因

座、Rag1基因座和/或Rag2/Rag1基因座包含基因组(即,哺乳动物基因组、人类基因组或非人类哺乳动物基因组)的相应区,这些基因或基因组合各自定位在这些相应区中。修饰ApoE基因座、干扰白细胞素-2受体 $\gamma$ (I12rg)基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座(即,哺乳动物、人类或非人类哺乳动物ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或合并的Rag2/Rag1基因座)中的任一个可包括对给定基因座的任何所要改变。修饰给定基因座(即,哺乳动物、人类或非人类哺乳动物基因座)的非限制性实例在本文中更详细地论述。

[0157] 例如,在特定的实施方案中,修饰所述ApoE基因座、干扰白细胞素-2受体 $\gamma$ (I12rg)基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座(即,哺乳动物、人类或非人类哺乳动物ApoE基因座;哺乳动物、人类或非人类哺乳动物干扰白细胞素-2受体 $\gamma$ 基因座;哺乳动物、人类或非人类哺乳动物Rag2基因座和/或Rag2/Rag1基因座)中的一种或多种,使得编码的ApoE蛋白或干扰白细胞素-2受体 $\gamma$ 蛋白或Rag1蛋白或Rag2蛋白或Rag1和Rag2蛋白的组的活性和/或水平减小。在其它实施方案中,所述ApoE蛋白、所述干扰白细胞素-2受体 $\gamma$ 蛋白、所述Rag1蛋白或所述Rag2蛋白或所述Rag1蛋白和所述Rag2蛋白的组的活性不存在。

[0158] “减小”意味着在目标基因座处编码的基因/蛋白的水平或活性的任何程度的减小。例如,活性减小可包括(1)当与适当的对照物相比较时给定蛋白(即,ApoE、干扰白细胞素-2受体 $\gamma$ 、Rag2、Rag2或Rag1和Rag2的组合)的总水平或活性统计上显著减小,包括例如0.5%、1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、120%或更大的水平或活性减小。测定ApoE、干扰白细胞素-2受体 $\gamma$ 、Rag1和Rag2中的任一种的浓度和/或活性减小的方法在本领域中已知。

[0159] 在其它实施方案中,哺乳动物、人类或非人类哺乳动物ApoE基因座,哺乳动物、人类或非人类哺乳动物干扰白细胞素-2受体 $\gamma$ 基因座,哺乳动物、人类或非人类哺乳动物Rag2基因座,哺乳动物、人类或非人类哺乳动物Rag1基因座和/或哺乳动物、人类或非人类哺乳动物Rag2/Rag1基因座中的一种或多种包括如下修饰,其使得编码的ApoE多肽、干扰白细胞素-2受体 $\gamma$ 多肽、Rag2多肽、Rag1多肽或Rag1多肽和Rag2多肽两者的活性和/或水平增加。“增加”意味着在目标基因座处编码的基因/多肽的水平或活性的任何程度的增加。例如,活性增加可包括(1)当与适当的对照物相比较时给定蛋白(即,ApoE、干扰白细胞素-2受体 $\gamma$ 、Rag1、Rag2或Rag1和Rag2)的总水平或活性统计上显著增加,包括例如0.5%、1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、120%或更大的水平或活性增加。测定ApoE、Rag1、Rag2和干扰白细胞素-2受体 $\gamma$ 蛋白中的任一种的浓度和/或活性增加的方法在本领域中已知。

[0160] 对哺乳动物、人类或非人类哺乳动物ApoE基因座,哺乳动物、人类或非人类哺乳动物干扰白细胞素-2受体 $\gamma$ 基因座,哺乳动物、人类或非人类哺乳动物Rag2基因座,哺乳动物、人类或非人类哺乳动物Rag1基因座和/或哺乳动物、人类或非人类哺乳动物Rag2/Rag1基因座的基因修饰可包括在所述基因组基因座处缺失内源核酸序列、在所述基因组基因座处插入外源核酸或其组合。所述缺失和/或插入可在如在本文中的其它地方论述的给定基因座内的任何地方发生。

[0161] 本文提供的其它实施方案包括通过用来自另一生物体的ApoE基因座、干扰白细胞

素-2受体  $\gamma$  基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应同源或直系同源部分替换ApoE基因座、干扰白细胞素-2受体  $\gamma$  (I12rg) 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分修饰哺乳动物、人类或非人类哺乳动物ApoE基因座、干扰白细胞素-2受体  $\gamma$  基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应同源或直系同源部分。

[0162] 在其它实施方案中,哺乳动物、人类或非人类哺乳动物ApoE基因座、干扰白细胞素-2受体  $\gamma$  基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座中的一种或多种的修饰通过用跨其全长与被替换的ApoE基因座、干扰白细胞素-2受体  $\gamma$  (I12rg) 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分共有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%的插入多核苷酸取代ApoE基因座、干扰白细胞素-2受体  $\gamma$  基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分进行。

[0163] 给定插入的多核苷酸和/或基因座的相应缺失区可为编码区、内含子、外显子、未转译区、调控区、启动子或增强子或其任何组合或其任何部分。此外,例如给定插入的多核苷酸和/或基因座的相应缺失区可具有任何所要长度,包括例如10-100核苷酸长度、100-500核苷酸长度、500-1kb核苷酸长度、1kb-1.5kb核苷酸长度、1.5kb-2kb核苷酸长度、2kb-2.5kb核苷酸长度、2.5kb-3kb核苷酸长度、3kb-5kb核苷酸长度、5kb-8kb核苷酸长度、8kb-10kb核苷酸长度或更长。在其它情况下,所述插入或替换的尺寸为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb、约350kb-约400kb、约400kb-约800kb、约800kb-1Mb、约300kb-约400kb、约400kb-约500kb、约500kb-1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb、约2.5Mb-约2.8Mb、约2.8Mb-约3Mb。在其它实施方案中,给定插入的多核苷酸和/或基因座的相应缺失区为至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或更长。在其它实施方案中,给定插入的多核苷酸和/或基因座的相应缺失区为至少10kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少150kb、至少200kb、至少250kb或至少300kb或更长。

[0164] 所述给定插入的多核苷酸可来自任何生物体,包括例如啮齿动物、非大鼠啮齿动物、大鼠、小鼠、仓鼠、哺乳动物、非人类哺乳动物、真核生物、非大鼠真核生物、人类、农业动物或家养动物。

[0165] 如在本文中更详细地论述,提供各种方法以产生任何目标基因座的靶向修饰,包括例如在ApoE基因座、干扰白细胞素-2受体  $\gamma$  (I12rg) 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座中的靶向修饰。进一步提供基因修饰的非人类动物、基因修饰的非人类哺乳动物、基因修饰的非大鼠真核生物、基因修饰的非多潜能细胞或基因修饰的多潜能细胞(例如,多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞或人类iPS细胞),其包括在所述干扰白细胞素-2受体  $\gamma$  基因座处、在所述ApoE基因座处、在所述Rag2基因座处、在所述Rag1基因座处和/或在所述Rag2/Rag1基因座处的缺失、插入、替换和/或其任何组合。这样的基因修饰(包括引起靶基因座的活性的缺乏、减小、增加或调整的那些基因修饰)且还能够经由种系传递。在特定的实施方案中,

所述基因修饰引起所要靶基因座的敲除。例如,这样的非人类动物在如本文中的其它部分中论述的多种实验系统中得到应用。

[0166] 例如,ApoE(脱脂蛋白E)敲除提供研究包括但不限于斑块形成、转录改变(全基因组鸟枪测序(RNA-Seq)和离体功能的内皮功能的动物模型。ApoE为重要的运输分子和可运输例如胆固醇的脂质穿过血流。ApoE还可在神经系统中起作用,例如以从脑中廓清 $\beta$ -淀粉样蛋白。在ApoE中的修饰已经在包括例如动脉粥样硬化、高脂血症和阿兹海默氏病的各种病状中牵涉到。ApoE敲除动物显示出脂蛋白自血液的廓清受损并发展成动脉粥样硬化。因此,ApoE敲除动物提供研究例如内皮功能、斑块形成、转录改变(RNA-Seq)、高脂血症、动脉粥样硬化和阿兹海默氏病的病状和/或过程的模型。测量ApoE活性的测定在本领域中已知。例如,ApoE活性的减小可通过通过免疫测定法、例如通过ELISA或通过免疫印迹技术测定在从受试者中获得的血样中ApoE水平的减小来测量。然而,大型的大鼠促进所有这些测定并改进资料的质量。

[0167] RAG1(重组激活基因1)和RAG2(重组激活基因2)为具有VDJ重组活性的多亚基复合体的一部分且在淋巴细胞中的免疫球蛋白和T细胞受体基因的重排和重组中起重要作用的酶。RAG1和RAG2诱导双链DNA裂解以促进T细胞受体和B细胞受体(即,免疫球蛋白)基因的片段的重组和连接。敲除RAG1和/或RAG2造成在动物中B细胞和T细胞损失,引起严重的免疫缺陷。RAG1和/或RAG2敲除动物例如在异种移植(即,在大鼠中的人类细胞异种移植)、癌症、疫苗研发、自身免疫疾病、传染性疾病和移植物抗宿主疾病(GVHD)的研究中得到应用。测量RAG1和/或RAG2活性的各种测定在本领域中已知且包括例如测量组合效率或测定在受试者中B细胞和/或T细胞的存在与否。

[0168] IL-2受体(IL-2R)在某些免疫细胞的表面上表达且结合细胞因子干扰白细胞素-2(IL-2)。IL-2R为包含至少三个单独的亚基链的膜内在蛋白,所述亚基链包括 $\alpha$ 链(IL-2Ra、CD25)、 $\beta$ 链(IL-2Rb、CD122)和 $\gamma$ 链(IL-2R $\gamma$ ,CD132)。IL-2受体 $\gamma$ (也称作IL-2R $\gamma$ 或IL-2R $\gamma$ )链为包括例如用于IL-2、IL-4、IL-7、IL-9、IL-15和IL-21的受体的各种细胞因子受体共有的常见 $\gamma$ 链。IL-2R $\gamma$ 包含在细胞的细胞外表面上的胞外域,其促进配体、跨膜域和细胞内结构域的结合,其可与各种分子相互作用以诱导细胞内信号转导途径。IL-2R $\gamma$ 基因在哺乳动物中的X染色体上见到且在人类的 $\gamma$ 链基因中的某些突变可导致以明显T细胞缺陷为特征的人类X连接的严重联合免疫缺陷(XSCID)。另外, $\gamma$ 链胞外结构域可自跨膜受体脱落并作为可溶性 $\gamma$ 链受体释放。可溶性 $\gamma$ 链受体可在受试者的血液中检测到且可起到调控细胞因子信号的作用。

[0169] 在一些实施方案中,所述非人类IL-2R $\gamma$ 链用人类IL-2R $\gamma$ 链替换,使得基因修饰的动物表达全人IL-2R $\gamma$ 链。在其它情况下,可能有用的是用人类IL-2R $\gamma$ 链的胞外域仅替换非人类IL-2R $\gamma$ 链的胞外域。在这些情况下,在非人类中表达的所得人源化IL-2R $\gamma$ 链包含人类胞外域,分子的剩余部分来自天然生物体。

[0170] 可使用IL-2R $\gamma$ 的全长人源化,因为具有该修饰基因座的非人类哺乳动物将生成人类IL-2R $\gamma$ 。这将允许用对人类IL-2R $\gamma$ 具有特异性的抗体检测在非人类哺乳动物中的人类IL-2R $\gamma$ 。外人源化(即,用IL-2R $\gamma$ 的人类胞外结构域替换非人类哺乳动物的IL-2R $\gamma$ 的胞外结构域)将产生将结合IL-2R $\gamma$ 的人类配体的IL-2R $\gamma$ 多肽,但是因为细胞质域仍然来自非人类哺乳动物,所以外人源化形式的IL-2R $\gamma$ 也将与非人类哺乳动物信号机构(machinery)

相互作用。

[0171] 2. 修饰靶基因座

[0172] A. 靶向载体并插入核酸

[0173] i. 插入核酸

[0174] 本文使用的“插入核酸”包含希望在靶基因座处整合的DNA的片段。在一个实施方案中,所述插入核酸包含一种或多种目标多核苷酸。在其它实施方案中,所述插入核酸可包含一种或多种表达盒。给定的表达盒可包含目标多核苷酸、编码选择标记和/或报道基因的多核苷酸以及影响表达的各种调控组分。可包含在所述插入核酸内的目标多核苷酸、选择标记和报道基因的非限制性实例在本文中的其它地方详细地论述。

[0175] 在特定的实施方案中,所述插入核酸可包含来自大鼠的核酸,其可包含基因组DNA的片段、cDNA、调控区或其任何部分或组合。在其它实施方案中,所述插入核酸可包括来自真核生物、非大鼠真核生物、哺乳动物、人类、非人类哺乳动物、啮齿动物、非大鼠啮齿动物、人类、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、小鸡、猫、狗、白鼬、灵长类动物(例如,狨猴、恒河猴)、家养哺乳动物或农业哺乳动物或任何其它目标生物体的核酸。如在本文中更详细地概述,在各种方法和组合物中采用的插入核酸可引起目标靶基因座的“人源化”。

[0176] 在一个实施方案中,所述插入核酸包含内源基因的至少一个外显子的敲入等位基因。在一个实施方案中,所述插入核酸包含整体内源基因的敲入等位基因(即,“基因互换敲入”)。

[0177] 在一个实施方案中,所述插入核酸包含调控元件,包括例如启动子、增强子或转录阻遏子-结合元件。

[0178] 在其它实施方案中,所述插入核酸包含条件性等位基因。在一个实施方案中,所述条件性等位基因为多官能等位基因,如在US 2011/0104799中所述,所述专利的全部内容以引用的方式并入本文中。在特定的实施方案中,所述条件性等位基因包含:(a) 相对于靶基因的转录以有义取向的启动序列和以有义或反义取向的药物选择盒;(b) 以反义取向的目标核苷酸序列(NSI)和倒转条件模块(conditional by inversion module)(COIN,其利用外显子-断裂内含子和可倒转的基因诱捕样模块;参见,例如US 2011/0104799,其全部内容以引用的方式并入本文中);和(c) 在暴露于第一重组酶后重组以形成条件性等位基因的可重组单元,所述条件性等位基因(i) 缺乏启动序列和DSC,和(ii) 含有以有义取向的NSI和以反义取向的COIN。

[0179] 所述插入核酸为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb。

[0180] 在一个实施方案中,所述插入核酸包含例如约1kb-约200kb、约2kb-约20kb或约0.5kb-约3Mb的真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞或非人类哺乳动物细胞基因组DNA序列的缺失。在一个实施方案中,所述基因组DNA序列的缺失程度大于5'同源臂和3'同源臂的总长度。在一个实施方案中,所述基因组DNA序列的缺失程度为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约20kb-约30kb、约30kb-约40kb、约40kb-约50kb、约50kb-约60kb、约60kb-约70kb、约70kb-约80kb、约80kb-约90kb、约90kb-约100kb、约100kb-



约110kb、约110kb-约120kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb、约190kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb、约350kb-约400kb、约400kb-约800kb、约800kb-1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb、约2.5Mb-约2.8Mb、约2.8Mb-约3Mb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb。

[0181] 在一个实施方案中,所述插入核酸包含插入同源或直系同源人类核酸序列或用其替换真核、非大鼠真核、哺乳动物、人类或非人类哺乳动物核酸序列。在一个实施方案中,所述插入核酸包含在包含相应DNA序列的内源基因座处插入同源或直系同源人类核酸序列或用其替换DNA序列。

[0182] 在一个实施方案中,所述基因修饰为添加核酸序列。在一个实施方案中,所述添加的核苷酸序列为5kb-200kb。

[0183] 在一个实施方案中,所述插入核酸包含在编码序列中的基因修饰。在一个实施方案中,所述基因修饰包括编码序列的缺失突变。在一个实施方案中,所述基因修饰包括两种内源编码序列的融合。

[0184] 在一个实施方案中,所述插入核酸包含插入同源或直系同源人类核酸序列或用其替换真核、非大鼠真核、哺乳动物、人类或非人类哺乳动物核酸序列。在一个实施方案中,所述插入核酸包含在包含相应大鼠DNA序列的内源大鼠基因座处插入同源或直系同源人类核酸序列或用其替换大鼠DNA序列。

[0185] 在一个实施方案中,所述基因修饰包括非蛋白编码序列的缺失,但不包括蛋白编码序列的缺失。在一个实施方案中,所述非蛋白编码序列的缺失包括调控元件的缺失。在一个实施方案中,所述基因修饰包括启动子的缺失。在一个实施方案中,所述基因修饰包括启动子或调控元件的添加。在一个实施方案中,所述基因修饰包括启动子或调控元件的替换。

[0186] 在一个实施方案中,所述靶向载体的核酸序列可包含在整合到基因组中时将生成哺乳动物、人类或非人类哺乳动物ApoE基因座的区域的基因修饰的多核苷酸,其中在ApoE基因座处的基因修饰引起ApoE活性减小、ApoE活性增加或ApoE活性调整。在一个实施方案中,产生ApoE敲除(“无效等位基因”)。

[0187] 在一个实施方案中,所述靶向载体的核酸序列可包含在整合到基因组中时将生成哺乳动物、人类细胞或非人类哺乳动物干扰白细胞素-2受体基因座的区域的基因修饰的多核苷酸,其中在所述干扰白细胞素-2受体基因座处的基因修饰引起干扰白细胞素-2受体活性减小。在一个实施方案中,产生干扰白细胞素-2受体敲除(“无效等位基因”)。

[0188] 在其它实施方案中,所述插入核酸引起用来自另一生物体的ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应同源或直系同源部分替换哺乳动物、人类细胞或非人类哺乳动物ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座和/或Rag2基因座和/或Rag1基因座和/或Rag2/Rag1基因座的一部分。

[0189] 在其它实施方案中,所述核酸序列包含跨其全长与被替换的ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分共有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%的多核苷酸。

[0190] 给定插入的多核苷酸和/或哺乳动物、人类细胞或非人类哺乳动物基因座的相应

替换区可为编码区、内含子、外显子、未转译区、调控区、启动子或增强子或其任何组合。此外,给定插入的多核苷酸和/或哺乳动物、人类细胞或非人类哺乳动物基因座的缺失区可具有任何所要长度,包括例如10-100核苷酸长度、100-500核苷酸长度、500-1kb核苷酸长度、1Kb-1.5kb核苷酸长度、1.5kb-2kb核苷酸长度、2kb-2.5kb核苷酸长度、2.5kb-3kb核苷酸长度、3kb-5kb核苷酸长度、5kb-8kb核苷酸长度、8kb-10kb核苷酸长度或更长。在其它情况下,所述插入或替换的尺寸为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb、约350kb-约400kb、约400kb-约800kb、约800kb-1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb、约2.5Mb-约2.8Mb、约2.8Mb-约3Mb。在其它实施方案中,给定插入的多核苷酸和/或哺乳动物、人类细胞或非人类哺乳动物基因座的缺失区为至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或更长。

[0191] 在一个实施方案中,所述启动子为组成性活性启动子。

[0192] 在一个实施方案中,所述启动子为可诱导的启动子。在一个实施方案中,所述可诱导的启动子为化学调控启动子。在一个实施方案中,所述化学调控启动子为醇调控启动子。在一个实施方案中,所述醇调控启动子为乙醇脱氢酶(alcA)基因启动子。在一个实施方案中,所述化学调控启动子为四环素调控启动子。在一个实施方案中,所述四环素调控启动子为四环素反应启动子。在一个实施方案中,所述四环素调控启动子为四环素操纵子序列(tetO)。在一个实施方案中,所述四环素调控启动子为tet-On启动子。在一个实施方案中,所述四环素调控启动子为tet-Off启动子。在一个实施方案中,所述化学调控启动子为类固醇调控启动子。在一个实施方案中,所述类固醇调控启动子为大鼠糖皮质激素受体的启动子。在一个实施方案中,所述类固醇调控启动子为雌激素受体的启动子。在一个实施方案中,所述类固醇调控启动子为蜕皮激素受体的启动子。在一个实施方案中,所述化学调控启动子为金属调控启动子。在一个实施方案中,所述金属调控启动子为金属蛋白质启动子。在一个实施方案中,所述可诱导的启动子为物理调控启动子。在一个实施方案中,所述物理调控启动子为温度调控启动子。在一个实施方案中,所述温度调控启动子为热休克启动子。在一个实施方案中,所述物理调控启动子为光调控启动子。在一个实施方案中,所述光调控启动子为光可诱导的启动子。在一个实施方案中,所述光调控启动子为光可阻遏的启动子。

[0193] 在一个实施方案中,所述启动子为组织特异性启动子。在一个实施方案中,所述启动子为神经元特异性启动子。在一个实施方案中,所述启动子为神经胶质特异性启动子。在一个实施方案中,所述启动子为肌细胞特异性启动子。在一个实施方案中,所述启动子为心脏细胞特异性启动子。在一个实施方案中,所述启动子为肾细胞特异性启动子。在一个实施方案中,所述启动子为骨细胞特异性启动子。在一个实施方案中,所述启动子为内皮细胞特异性启动子。在一个实施方案中,所述启动子为免疫细胞特异性启动子。在一个实施方案中,所述免疫细胞启动子为B细胞启动子。在一个实施方案中,所述免疫细胞启动子为T细胞启动子。

[0194] 在一个实施方案中,所述启动子为发育调控启动子。在一个实施方案中,所述发育调控启动子仅在胚胎发育阶段期间具有活性。在一个实施方案中,所述发育调控启动子仅在成人细胞中具有活性。

[0195] 在特定的实施方案中,所述启动子可基于细胞类型选择。因此,各种启动子在真核细胞、非大鼠真核细胞、哺乳动物细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞中得到应用。

[0196] 在一些实施方案中,所述插入核酸包括侧接有位点特异性重组靶序列的核酸。已经认识到,虽然整个插入核酸可由这样的位点特异性重组靶序列侧接,但是在所述插入核酸内的任何区或单个目标多核苷酸也可由这样的位点侧接。所述位点特异性重组酶可通过包括将重组酶多肽引入细胞中或通过将其编码位点特异性重组酶的多核苷酸引入宿主细胞中的任何方法引入细胞中。编码位点特异性重组酶的多核苷酸可定位在插入核酸内或在单独的多核苷酸内。所述位点特异性重组酶可操作性连接在细胞中具有活性的启动子,包括例如可诱导的启动子、与细胞内源的启动子、与细胞异源的启动子、细胞特异性启动子、组织特异性启动子或发育阶段特异性启动子。可侧接插入核酸或在插入核酸中的任何目标多核苷酸的位点特异性重组靶序列可包括但不限于loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attP、att、FRT、rox和其组合。

[0197] 在一些实施方案中,所述位点特异性重组部位侧接在所述插入核酸内包含的编码选择标记和/或报道基因的多核苷酸。在这样的情况下,在靶基因座处整合插入核酸之后,可除去在位点特异性重组位点之间的序列。

[0198] 在一个实施方案中,所述插入核酸包含编码选择标记的多核苷酸。所述选择标记可包含在选择盒中。这样的选择标记包括但不限于新霉素磷酸转移酶(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。在一个实施方案中,编码所述选择标记的所述多核苷酸可操作性连接在细胞、大鼠细胞、多潜能大鼠细胞、ES大鼠细胞、真核细胞、非大鼠真核细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、哺乳动物细胞、非人类哺乳动物细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞中具有活性的启动子。在将目标多核苷酸连续铺装到靶基因座中时,所述选择标记可包含用于核酸酶试剂的识别位点,如上概述。在一个实施方案中,编码所述选择标记的所述多核苷酸侧接有位点特异性重组靶序列。

[0199] 所述插入核酸可进一步包含操作性连接启动子的报道基因,其中编码报道基因蛋白的所述报道基因选自或包含LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(eYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶和/或其组合。这样的报道基因可操作性连接在细胞中具有活性的启动子。这样的启动子可为可诱导的启动子、与报道基因或细胞内源的启动子、与报道基因或细胞异源的启动子、细胞特异性启动子、组织特异性启动子或发育阶段特异性启动子。

[0200] 在一个实施方案中,核酸插入物可包括包含编码在神经系统、骨骼系统、消化系统、循环系统、肌肉系统、呼吸系统、心血管系统、淋巴系统、内分泌系统、泌尿系统、生殖系统或其组合中表达的蛋白的基因组基因座的哺乳动物核酸。在一个实施方案中,所述哺乳

动物核酸包含编码在骨髓或源自骨髓的细胞中表达的蛋白质的基因组基因座。在一个实施方案中,所述核酸包含编码在脾细胞中表达的蛋白质的基因组基因座。

[0201] 在一个实施方案中,所述哺乳动物核酸包含编码在神经系统、骨骼系统、消化系统、循环系统、肌肉系统、呼吸系统、心血管系统、淋巴系统、内分泌系统、泌尿系统、生殖系统或其组合中表达的蛋白质的基因组基因座。在一个实施方案中,所述哺乳动物核酸包含编码在骨髓或源自骨髓的细胞中表达的蛋白质的基因组基因座。在一个实施方案中,所述核酸包含编码在脾细胞中表达的蛋白质的基因组基因座。在一个实施方案中,所述基因组基因座包含小鼠基因组DNA序列、大鼠基因组DNA序列、真核的基因组DNA序列、非大鼠真核基因组DNA序列、哺乳动物基因组DNA序列、人类基因组DNA序列或非人类DNA序列哺乳动物或其组合。在一个实施方案中,所述基因组基因座包含以任何顺序的大鼠和人类基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的小鼠和人类基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的小鼠和大鼠基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的大鼠、小鼠和人类基因组DNA序列。

[0202] 在一个实施方案中,所述基因组基因座包含小鼠基因组DNA序列、大鼠基因组DNA序列、仓鼠基因组DNA序列、人类基因组DNA序列、真核基因组DNA序列、非大鼠真核基因组DNA序列、哺乳动物基因组DNA序列或非人类DNA序列哺乳动物或其组合。在一个实施方案中,所述基因组基因座包含以任何顺序的大鼠和人类基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的小鼠和人类基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的小鼠和大鼠基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的大鼠、小鼠和人类基因组DNA序列。

[0203] 在一个实施方案中,所述基因修饰包含人类基因的至少一种人类疾病等位基因。在一个实施方案中,所述人类疾病为神经病。在一个实施方案中,所述人类疾病为心血管疾病。在一个实施方案中,所述人类疾病为肾病。在一个实施方案中,所述人类疾病为肌肉疾病。在一个实施方案中,所述人类疾病为血液病。在一个实施方案中,所述人类疾病为癌症。在一个实施方案中,所述人类疾病为免疫系统疾病。

[0204] 在一个实施方案中,所述人类疾病等位基因为显性等位基因。在一个实施方案中,所述人类疾病等位基因为隐性等位基因。在一个实施方案中,所述人类疾病等位基因包括单一核苷酸多态性(SNP)等位基因。

[0205] 在一个实施方案中,所述基因修饰生成具有改变的结合特性、改变的定位、改变的表达和/或改变的表达模式的突变体形式的蛋白质。

[0206] 在一个实施方案中,所述插入核酸包含选择盒。在一个实施方案中,所述选择盒包含编码选择性标记的核酸序列,其中所述核酸序列操作性连接在大鼠ES细胞中具有活性的启动子。在一个实施方案中,所述选择性标记选自或包含潮霉素抗性基因或新霉素抗性基因。

[0207] 在一个实施方案中,所述核酸包含编码在B细胞中表达的蛋白质的基因组基因座。在一个实施方案中,所述核酸包含编码在未成熟的B细胞中表达的蛋白质的基因组基因座。在一个实施方案中,所述核酸包含编码在成熟B细胞中表达的蛋白质的基因组基因座。

[0208] 在一个实施方案中,所述插入核酸包含调控元件。在一个实施方案中,所述调控元

件为启动子。在一个实施方案中,所述调控元件为增强子。在一个实施方案中,所述调控元件为转录阻遏子-结合元件。

[0209] 在一个实施方案中,所述基因修饰包括非蛋白编码序列的缺失,但不包括蛋白编码序列的缺失。在一个实施方案中,所述非蛋白编码序列的缺失包括调控元件的缺失。在一个实施方案中,所述基因修饰包括调控元件的缺失。在一个实施方案中,所述基因修饰包括启动子或调控元件的添加。在一个实施方案中,所述基因修饰包括启动子或调控元件的替换。

[0210] ii. 表达盒

[0211] 本文提供包含在本文中提供的靶向基因组整合系统中采用的各种组分(即,核酸酶试剂、识别位点、插入核酸、目标多核苷酸、靶向载体、选择标记及其它组分中的任一种或任何组合)的多核苷酸或核酸分子。

[0212] 术语“多核苷酸”、“多核苷酸序列”、“核苷酸序列”和“核苷酸片段”可在本文中互换使用。这些术语涵盖核苷酸序列等。多核苷酸可为任选含有合成、非天然或改变的核苷酸碱基的单链或双链的RNA或DNA的聚合物。以DNA的聚合物形式的多核苷酸可由cDNA、基因组DNA、合成DNA或其混合物的一种或多种片段构成。多核苷酸可包含脱氧核糖核苷酸和核糖核苷酸,其包括天然存在的分子和合成类似物两者或这些的任何组合。本文提供的多核苷酸还涵盖所有形式的序列,其包括但不限于单链形式、双链形式、发夹、茎和环结构等。

[0213] 进一步提供包含靶向基因组整合系统的各种组分的重组多核苷酸。术语“重组多核苷酸”和“重组DNA”构建体可在本文中互换使用。重组构建体包含例如实际上并未一同发现的调控序列和编码序列的核酸序列的人造和异源组合。在其它实施方案中,重组构建体可包含源自不同来源的调控序列和编码序列,或源自相同来源但以与实际上发现所不同的方式配置的调控序列和编码序列。这一构建体可本身使用或可结合载体使用。如果使用载体,则载体的选择取决于如本领域的技术人员众所周知的用以转化宿主细胞的方法。例如,可使用质粒载体。还提供了用以成功转化、选择和繁殖包含本文提供的分离的核酸片段中的任一种的宿主细胞需要的基因元件。筛选可尤其通过DNA的Southern分析、mRNA表达的Northern分析、蛋白表达的免疫印迹分析或表型分析实现。

[0214] 在特定的实施方案中,本文所述的靶向基因组整合系统的组分中的一种或多种可提供在表达盒中以便在原核细胞、真核细胞、非大鼠真核细胞、细菌、酵母菌细胞或哺乳动物细胞或其它生物体或目标细胞类型中表达。所述盒可以包括操作性连接本文提供的多核苷酸的5'调控序列和3'调控序列。“操作性连接”包括其中组分以其预定方式操作性连接起作用的关系。例如,在目标多核苷酸和调控序列(即,启动子)之间的操作性键为允许表达目标多核苷酸的功能连接。“操作性连接元件”可邻接或不邻接。当用以提到两种蛋白编码区的连接时,操作性连接是指编码区在同一阅读框中。在另一情况下,编码蛋白质的核酸序列可以操作性连接于调控序列(例如,启动子、增强子、沉默子序列等)以保持恰当的转录调控。在一种情况下,免疫球蛋白可变区(或V(D)J片段)的核酸序列可操作性连接到免疫球蛋白恒定区的核酸序列,以允许在序列之间恰当重组到免疫球蛋白重链或轻链序列。

[0215] 所述盒可另外含有将共同引入生物体中的至少一种另外的目标多核苷酸。可选地,所述另外的目标多核苷酸可提供在多个表达盒上。这一表达盒提供有用于插入重组多核苷酸以处于调控区的转录调控下的多个限制位点和/或重组位点。所述表达盒可另外含

有选择标记基因。

[0216] 所述表达盒可包括在5'-3'转录方向上的,在哺乳动物细胞或目标宿主细胞中起作用的转录和转译起动区(即,启动子)、本文提供的重组多核苷酸,以及转录和转译终止区(即,终止区)。所述调控区(即,启动子、转录调控区和转译终止区)和/或本文提供的多核苷酸对于宿主细胞可为天然/类似的或彼此类似。可选地,所述调控区和/或本文提供的多核苷酸可与宿主细胞异源或彼此异源。例如,操作性连接异源多核苷酸的启动子来自与得到多核苷酸的物种不同的物种,或者如果来自相同/类似的物种,一者或两者自其原始形式和/或基因组基因座实质性修饰,或者启动子不是操作性连接的多核苷酸的天然启动子。可选地,本文提供的调控区和/或重组多核苷酸可以是完全合成的。

[0217] 所述终止区关于转录起动区可为天然的,关于操作性连接的重组多核苷酸可为天然的,关于宿主细胞可为天然的,或者关于启动子、重组多核苷酸、宿主细胞或其任何组合,可源自另一来源(即,外来的或异源性)。

[0218] 在制备表达盒的过程中,可操控各种DNA片段,从而提供以恰当取向的DNA序列。为此,可采用接头或连接子以接合DNA片段或者可涉及其它操控以提供适宜的限制位点、除去多余的DNA、除去限制位点等。为此目的,可涉及体外诱变、引物修复、限制、退火、再取代,例如转换和颠换。

[0219] 多种启动子可用于本文提供的表达盒中。启动子可基于所要的结果来选择。应该认识到不同的应用可通过在表达盒中使用不同的启动子来增强,从而调整目标多核苷酸的表达的定时、定位和/或水平。如果需要,则这样的表达构建体还可含有启动子调控区(例如,赋予可诱导的、组成性的、环境或发育调控的或细胞或组织特异性/选择性的表达)、转录起动开始位点、核糖体结合位点、RNA加工信号、转录终止位点和/或多聚腺苷酸化信号。

[0220] 含有本文提供的多核苷酸的表达盒还可包含用于选择转化细胞的选择标记基因。利用可选择的标记基因来选择转化细胞或组织。

[0221] 在适当的情况下,可优化在所述方法和组合物(即,目标多核苷酸、核酸酶试剂等)中采用的序列,以便增加在细胞中的表达。也就是说,所述基因可使用在给定目标细胞中优选的密码子合成以便改进表达,所述密码子包括例如哺乳动物优选的密码子、人类优选的密码子、啮齿动物优选的密码子、非大鼠啮齿动物优选的密码子、小鼠优选的密码子、大鼠优选的密码子、仓鼠优选的密码子等。

[0222] 本文提供的各种方法和组合物可采用选择标记。各种选择标记可用于本文公开的方法和组合物中。所述选择标记可例如给予对例如G418、潮霉素、杀稻瘟菌素、新霉素或嘌呤霉素的抗生素的抵抗力。这样的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>)和杀稻瘟菌素S脱氨酶(bsr<sup>r</sup>)。在其它实施方案中,所述选择标记操作性连接可诱导的启动子且所述选择标记的表达对于细胞具有毒性。这样的选择标记的非限制性实例包括黄嘌呤/鸟嘌呤磷酸核糖基转移酶(gpt)、次黄嘌呤-鸟嘌呤转磷酸核糖基酶(HGPRT)或单纯疱疹病毒胸苷激酶(HSV-TK)。编码所述选择标记的所述多核苷酸操作性连接在细胞中具有活性的启动子。

[0223] iii. 靶向载体

[0224] 采用靶向载体以将插入核酸引入大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸的靶基因座中。所述靶向载体包含

所述插入核酸且进一步包含5'同源臂和3'同源臂,其侧接所述插入核酸。侧接所述插入核酸的同源臂对应于在大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸的靶基因座内的区域。为了便于提及,在靶基因组基因座内的相应同源基因组区在本文中称为“靶位点”。例如,靶向载体可包含由与第一靶位点和第二靶位点互补的第一同源臂和第二同源臂侧接的第一插入核酸。因而,所述靶向载体由此有助于经由在细胞的基因组内的同源臂和互补靶位点之间发生的同源重组事件将插入核酸整合到大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸的靶基因座中。

[0225] 在一个实施方案中,所述大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸的靶基因座包含与5'同源臂互补的第一核酸序列和与3'同源臂互补的第二核酸序列。在一个实施方案中,所述第一核酸序列和所述第二核酸序列相隔至少5kb。在另一实施方案中,所述第一核酸序列和所述第二核酸序列相隔至少5kb,但小于200kb。在一个实施方案中,所述第一核酸序列和所述第二核酸序列相隔至少10kb。在一个实施方案中,所述第一核酸序列和所述第二核酸序列相隔至少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。在更进一步的实施方案中,所述第一核酸序列和所述第二核酸序列相隔至少5kb,但小于10kb;至少5kb,但小于3Mb;至少10kb,但小于20kb;至少20kb,但小于40kb;至少40kb,但小于60kb;至少60kb,但小于80kb;至少约80kb,但小于100kb;至少100kb,但小于150kb;或至少150kb,但小于200kb;至少约200kb,但小于约300kb;至少约300kb,但小于约400kb;至少约400kb,但小于约500kb;至少约500kb,但小于约1Mb;至少约1.5Mb,但小于约2Mb;至少约1Mb,但小于约1.5Mb;至少约2Mb,但小于2.5Mb;至少约2.5Mb,但小于3Mb;或至少约2Mb,但小于约3Mb。

[0226] 所述靶向载体的同源臂可具有足以促进与相应靶位点的同源重组事件的任何长度,包括例如至少5-10kb、5-15kb、10-20kb、20-30kb、30-40kb、40-50kb、50-60kb、60-70kb、70-80kb、80-90kb、90-100kb、100-110kb、110-120kb、120-130kb、130-140kb、140-150kb、150-160kb、160-170kb、170-180kb、180-190kb、190-200kb长或更长。如下文进一步详细概述,大靶向载体可采用更大长度的靶向臂。在一个特定的实施方案中,5'同源臂和3'同源臂的总和为至少10kb,或5'同源臂和3'同源臂的总和为至少约16kb-约100kb或约30kb-约100kb。在其它实施方案中,所述LTVEC的5'同源臂和3'同源臂的总和的尺寸为约10kb-约150kb、约10kb-约100kb、约10kb-约75kb、约20kb-约150kb、约20kb-约100kb、约20kb-约75kb、约30kb-约150kb、约30kb-约100kb、约30kb-约75kb、约40kb-约150kb、约40kb-约100kb、约40kb-约75kb、约50kb-约150kb、约50kb-约100kb、或约50kb-约75kb、约10kb-约30kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-约150kb。在一个实施方案中,所述缺失的尺寸与所述LTVEC的5'同源臂和3'同源臂的总和的尺寸相同或类似。

[0227] 在一个实施方案中,所述目标基因组基因座包含(i)与所述5'同源臂同源的5'靶序列;和(ii)与所述3'同源臂同源的3'靶序列。在一个实施方案中,所述5'靶序列和所述3'靶序列相隔至少5kb,但小于3Mb。在更进一步的实施方案中,所述5'靶序列和所述3'靶序列

相隔至少5kb,但小于10kb;至少5kb,但小于3Mb;至少10kb,但小于20kb;至少20kb,但小于40kb;至少40kb,但小于60kb;至少60kb,但小于80kb;至少约80kb,但小于100kb;至少100kb,但小于150kb;或至少150kb,但小于200kb;至少约200kb,但小于约300kb;至少约300kb,但小于约400kb;至少约400kb,但小于约500kb;至少约500kb,但小于约1Mb;至少约1.5Mb,但小于约2Mb;至少约1Mb,但小于约1.5Mb;至少约2Mb,但小于2.5Mb;至少约2.5Mb,但小于约3Mb;或至少约2Mb,但小于约3Mb。

[0228] 当采用核酸酶试剂时,对应于靶向载体的5'同源臂和3'同源臂的同源基因组区“足够邻近于”核酸酶靶位点定位以促进在识别位点处切口或双链断裂后在同源基因组区和同源臂之间发生同源重组事件。例如,所述核酸酶靶位点可定位在对应于5'同源臂和3'同源臂的同源基因组区之间的任何地方。在特定的实施方案中,识别位点直接邻近于同源基因组区中的至少一个或两个。

[0229] 如本文中所示,在两个区彼此共有足够水平的序列同一性时同源臂和靶位点(即,同源基因组区)彼此“互补”,从而充当用于同源重组反应的底物。“同源性”是指DNA序列与相应或“互补”序列相同或共有序列同一性。在给定靶位点和在靶向载体上发现的相应同源臂之间的序列同一性可为允许同源重组发生的任何程度的序列同一性。例如,靶向载体的同源臂(或其片段)与靶位点(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,因此所述序列经历同源重组。此外,在同源臂和互补靶位点之间同源的互补区可具有足以促进在裂解的识别位点处同源重组的任何长度。例如,给定同源臂和/或互补靶位点可包含同源的互补区,所述互补区为至少5-10kb、5-15kb、10-20kb、20-30kb、30-40kb、40-50kb、50-60kb、60-70kb、70-80kb、80-90kb、90-100kb、100-110kb、110-120kb、120-130kb、130-140kb、140-150kb、150-160kb、160-170kb、170-180kb、180-190kb、190-200kb、200kb-300kb长或更长(例如,如在本文中的其它地方描述的LTVEC载体中所述),因此同源臂与在细胞的基因组内的相应靶位点具有足以进行同源重组的同源性。为了便于提及,同源臂在本文中作为5'同源臂和3'同源臂提到。该术语涉及在靶向载体中同源臂与插入核酸的相对位置。

[0230] 所述靶向载体的同源臂因此设计成与具有靶向基因座的靶位点互补。因此,所述同源臂可与对细胞为天然的基因座互补,或可选地,它们可与整合到细胞的基因组中的DNA的异源或外源片段的区互补,所述区包括但不限于转基因、表达盒或基因组DNA的异源或外源区。可选地,所述靶向载体的同源臂可与人类人造染色体的区或在适当宿主细胞中包含的任何其它工程化的基因组区互补。更进一步,所述靶向载体的同源臂可与BAC文库、粘粒文库或P1噬菌体文库的区互补或来源于BAC文库、粘粒文库或P1噬菌体文库。因此,在特定的实施方案中,所述靶向载体的同源臂与对给定细胞为天然、异源或外源的大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠基因组基因座互补。在其它实施方案中,所述同源臂与无法使用常规方法靶向的大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠基因组基因座互补,或者可在缺乏由核酸酶试剂诱导的切口或双链断裂的情况下仅不正确地靶向或仅以显著低的效率靶向。在一个实施方案中,所述同源臂来源于合成DNA。

[0231] 在其它实施方案中,所述5'同源臂和所述3'同源臂与和靶向基因组相同的基因组



互补。在一个实施方案中,所述同源臂来自相关基因组,例如,靶向基因组为第一株的大鼠基因组,且靶向臂来自第二株的大鼠基因组,其中第一株与第二株不同。在其它实施方案中,所述同源臂来自相同动物的基因组或来自相同株的基因组,例如靶向基因组为第一株的大鼠基因组,且靶向臂来自来自相同大鼠或来自相同株的大鼠基因组。

[0232] 所述靶向载体(例如,大靶向载体)还可包含如在本文中的其它地方论述的选择盒或报道基因。所述选择盒可包含编码选择标记的核酸序列,其中所述核酸序列操作性连接到启动子。所述启动子可在目标原核细胞中具有活性和/或在目标真核细胞中具有活性。这样的启动子可为可诱导的启动子、与报道基因或细胞内源的启动子、与报道基因或细胞异源的启动子、细胞特异性启动子、组织特异性启动子或发育阶段特异性启动子。在一个实施方案中,所述选择标记选自或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)和单纯疱疹病毒胸苷激酶(HSV-k)和/或其组合。所述靶向载体的选择标记可由5'同源臂和3'同源臂侧接或发现5'同源臂或3'同源臂。

[0233] 在一个实施方案中,所述靶向载体(例如,大靶向载体)包含操作性连接启动子的报道基因,其中所述报道基因编码选自由以下蛋白组成的组或包括以下蛋白的报道基因蛋白:LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(EYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶和/或其组合。这样的报道基因可操作性连接在细胞中具有活性的启动子。这样的启动子可为可诱导的启动子、与报道基因或细胞内源的启动子、与报道基因或细胞异源的启动子、细胞特异性启动子、组织特异性启动子或发育阶段特异性启动子。

[0234] 在一个实施方案中,合并使用靶向载体(包括例如大靶向载体)与核酸酶试剂产生与仅使用靶向载体相比较增加的靶向效率。在一个实施方案中,当与在仅使用靶向载体时相比较时,在靶向载体结合核酸酶试剂使用时,靶向载体的靶向效率增加至少两倍、至少三倍或至少4倍。

[0235] 当采用靶向载体时,载体设计可达到允许插入如本文所述的约5kb-约200kb的给定序列的程度。在一个实施方案中,所述插入为约5kb-约10kb、约10kb-约20kb、约20kb-约30kb、约30kb-约40kb、约40kb-约50kb、约50kb-约60kb、约60kb-约70kb、约80kb-约90kb、约90kb-约100kb、约100kb-约110kb、约110kb-约120kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb、或约190kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb。

[0236] 当采用靶向载体时,载体设计可达到允许替换如本文所述的约5kb-约200kb或约5kb-约3.0Mb的给定序列的程度。在一个实施方案中,所述替换为约5kb-约10kb、约10kb-约20kb、约20kb-约30kb、约30kb-约40kb、约40kb-约50kb、约50kb-约60kb、约60kb-约70kb、约80kb-约90kb、约90kb-约100kb、约100kb-约110kb、约110kb-约120kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb、约190kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约

40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb。

[0237] 在一个实施方案中,所述靶向载体包含位点特异性重组酶基因。在一个实施方案中,所述位点特异性重组酶基因编码Cre重组酶。在一个实施方案中,所述Cre重组酶基因为Crei,其中两种编码Cre重组酶的外显子被内含子隔开以防止其在原核细胞中表达。

[0238] 在一个实施方案中,所述Cre重组酶基因进一步包含核定位信号以促进Cre(或任何重组酶或核酸酶试剂)定位到核中(例如,所述基因为NL-Cre基因)。在一个特定的实施方案中,所述Cre重组酶基因进一步包含核定位信号和内含子(例如,NL-Crei)。

[0239] 在各种实施方案中,表达核酸酶试剂(包括上文论述的Cre或Crei重组酶)的合适启动子选自或包括Prm1、Blimp1、Gata6、Gata4、Igf2、Lhx2、Lhx5和/或Pax3。在一个特定的实施方案中,所述启动子为Gata6或Gata4启动子。各种启动子可来自任何生物体,包括例如啮齿动物如小鼠或大鼠、非大鼠啮齿动物、真核生物、非大鼠真核生物、非人类哺乳动物、哺乳动物、人类或仓鼠。在另一特定的实施方案中,所述启动子为Prm1启动子。在另一特定的实施方案中,所述启动子为大鼠Prm1启动子。在另一特定的实施方案中,所述启动子为小鼠Prm1启动子。在另一特定的实施方案中,所述启动子为Blimp1启动子或其片段,例如,Blimp1启动子的1kb或2kb片段。参见,例如美国专利8,697,851和美国申请公布2013-0312129,其两者的全部内容都以引用的方式并入本文中。

[0240] iv. 大靶向载体

[0241] 本文使用的术语“大靶向载体”或“LTVEC”包括如下靶向载体,其包含对应于且来源于大于通常由意欲在细胞中执行同源靶向的其它方法使用的那些的核酸序列的同源臂和/或包含包含大于通常由意欲在细胞中执行同源重组靶向的其它方法使用的那些的核酸序列的插入核酸。例如,所述LTVEC使得可以修饰大基因座,所述大基因组由于其尺寸限制而无法由传统的基于质粒的靶向载体接纳。在特定的实施方案中,所述LTVEC的同源臂和/或插入核酸包含真核细胞或非大鼠真核细胞的基因组序列。所述LTVEC的尺寸太大而无法通过例如southern印迹和广范围(例如,1kb-5kb)PCR的常规测定筛选靶向事件。LTVEC的实例包括但不限于来源于细菌人造染色体(BAC)、人类人造染色体或酵母菌人造染色体(YAC)的载体。LTVEC的非限制性实例及其制造方法例如描述在美国专利号6,586,251、6,596,541、7,105,348和WO 2002/036789(PCT/US01/45375)和US 2013/0137101中,其各自以引用的方式并入本文中。

[0242] 所述LTVEC可具有任何长度,包括但不限于约20kb-约400kb、约20kb-约30kb、约30kb-40kb、约40kb-约50kb、约50kb-约75kb、约75kb-约100kb、约100kb-125kb、约125kb-约150kb、约150kb-约175kb、约175kb-约200kb、约200kb-约225kb、约225kb-约250kb、约250kb-约275kb、或约275kb-约300kb、约200kb-约300kb、约300kb-约350kb、约350kb-约400kb、约350kb-约550kb。在一个实施方案中,所述LTVEC为约100kb。

[0243] 在一些实施方案中,所述LTVEC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或至少200kb。

[0244] 在一些实施方案中,所述LTVEC包含至少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或至少200kb的核酸序列。

[0245] 在一个实施方案中,所述LTVEC包含约5kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约30kb、约0.5kb-约30kb、约0.5kb-约40kb、约30kb-约150kb、约0.5kb-约150kb、约30kb-约40kb、约40kb-约50kb、约60kb-约70kb、约80kb-约90kb、约90kb-约100kb、约100kb-约110kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb、或约190kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb的插入核酸。

[0246] 在一个实施方案中,所述LTVEC包含至少100kb、至少150kb或至少200kb的核酸序列。

[0247] 当采用LTVEC时,载体设计可达到允许替换如本文所述的约5kb-约200kb或约5kb-约3Mb的给定序列的程度。在一个实施方案中,所述替换为约5kb-约10kb、约10kb-约20kb、约20kb-约30kb、约30kb-约40kb、约40kb-约50kb、约50kb-约60kb、约60kb-约70kb、约80kb-约90kb、约90kb-约100kb、约100kb-约110kb、约110kb-约120kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb、约190kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb。

[0248] 在一个实施方案中,所述LTVEC的同源臂来源于BAC文库、粘粒文库或P1噬菌体文库。在其它实施方案中,所述同源臂来源于细胞的靶基因组基因座,且在一些情况下,LTVEC设计用来靶向其的靶基因组基因座无法使用常规方法靶向。在其它实施方案中,所述同源臂来源于合成DNA。

[0249] 在一个实施方案中,在LTVEC中的5'同源臂和3'同源臂的总和为至少10kb。在其它实施方案中,所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约30kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、100kb-约120kb、约120kb-约140kb、约140kb-约160kb、约160kb-约180kb、约180kb-约200kb。在一个实施方案中,所述LTVEC的5'同源臂和3'同源臂的总和为约30kb-约100kb。在其它实施方案中,所述LTVEC的5'同源臂和3'同源臂的总和的尺寸为约10kb-约150kb、约10kb-约100kb、约10kb-约75kb、约20kb-约150kb、约20kb-约100kb、约20kb-约75kb、约30kb-约150kb、约30kb-约100kb、约30kb-约75kb、约40kb-约150kb、约40kb-约100kb、约40kb-约75kb、约50kb-约150kb、约50kb-约100kb、或约50kb-约75kb、约10kb-约30kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-约150kb。在一个实施方案中,所述缺失的尺寸与所述LTVEC的5'同源臂和3'同源臂的总和的尺寸相同或类似。

[0250] 在其它实施方案中,所述5'同源臂为约5kb-约100kb。在一个实施方案中,所述3'同源臂为约5kb-约100kb。在其它实施方案中,所述5'同源臂和所述3'同源臂的总和为约5kb-约10kb、约10kb-约20kb、约20kb-约30kb、约30kb-约40kb、约40kb-约50kb、约50kb-约60kb、约60kb-约70kb、约70kb-约80kb、约80kb-约90kb、约90kb-约100kb、约100kb-约

110kb、约110kb-约120kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb、约190kb-约200kb、或约30kb-约100kb、约10kb-约30kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-约150kb。

[0251] 在一个实施方案中,所述LTVEC包含与由LTVEC同源臂侧接的大鼠核酸序列同源或直系同源的插入核酸。在一个实施方案中,所述插入核酸序列来自除大鼠以外的物种。在一个实施方案中,所述插入核酸序列来自真核生物。在一个实施方案中,与大鼠核酸序列同源或直系同源的插入核酸为哺乳动物核酸。在一个实施方案中,与大鼠核酸序列同源或直系同源的插入核酸为非人类哺乳动物核酸。在一个实施方案中,所述哺乳动物核酸为小鼠核酸。在一个实施方案中,所述哺乳动物核酸为人类核酸。在一个实施方案中,所述哺乳动物核酸为仓鼠核酸。在一个实施方案中,所述插入核酸序列为基因组DNA。在一个实施方案中,所述插入为如上所述的5kb-200kb。

[0252] 在一个实施方案中,所述LTVEC包含选择盒或报道基因。可采用各种形式的选择盒和报道基因在本文中的其它地方论述。

[0253] 如在本文中的其它地方描述,所述LTVEC还可与核酸酶试剂组合用于本文提供的方法中,所述核酸酶试剂促进在靶向载体和在多潜能或非多潜能大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠细胞的大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸的靶基因座之间的同源重组。

[0254] 在一个实施方案中,所述大靶向载体(LTVEC)包含位点特异性重组酶基因。在一个实施方案中,所述位点特异性重组酶基因编码Cre重组酶。在一个实施方案中,所述Cre重组酶基因为Crei,其中两种编码Cre重组酶的外显子被内含子隔开以防止其在原核细胞中表达。在一个实施方案中,所述Cre重组酶基因进一步包含核定位信号以促进Cre(或任何重组酶或核酸酶试剂)定位到核中(例如,所述基因为NL-Cre基因)。在一个特定的实施方案中,所述Cre重组酶基因进一步包含核定位信号和内含子(例如,NL-Crei)。

[0255] 在各种实施方案中,表达核酸酶试剂(包括上文论述的Cre或Crei重组酶)的合适启动子选自或包括Prm1、Blimp1、Gata6、Gata4、Igf2、Lhx2、Lhx5和/或Pax3。在一个特定的实施方案中,所述启动子为Gata6或Gata4启动子。各种启动子可来自任何生物体,包括例如啮齿动物如小鼠或大鼠、非大鼠啮齿动物、真核生物、非大鼠真核生物、非人类哺乳动物、哺乳动物、人类或仓鼠。在另一特定的实施方案中,所述启动子为Prm1启动子。在另一特定的实施方案中,所述启动子为大鼠Prm1启动子。在另一特定的实施方案中,所述启动子为小鼠Prm1启动子。在另一特定的实施方案中,所述启动子为Blimp1启动子或其片段,例如,Blimp1启动子的1kb或2kb片段。参见,例如美国专利8,697,851和美国申请公布2013-0312129,其两者的全部内容都以引用的方式并入本文中。

[0256] 在一个实施方案中,所述LTVEC包含插入核酸,所述插入核酸可生成大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠ApoE基因座、Il2rg基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的区的缺失、添加、替换或其组合,如在本文中的其它部分详细论述。在特定的实施方案中,在ApoE基因座处的基因修饰引起ApoE活性、IL-2Rg活性、Rag2活性、Rag1活性和/或Rag2和Rag1活性减小、增加

或调整。在一个实施方案中,产生ApoE敲除和I12rg敲除、Rag2敲除、Rag1敲除、Rag2/Rag1敲除。如下论述,可采用核酸酶试剂以及任何LTVEC靶向系统以靶向任何目标基因组基因座。

[0257] 在另一实施方案中,所述基因组在包含至少10kb的核酸序列的大靶向载体(LTVEC)存在下暴露于Cas蛋白和CRISPR RNA。在这样的情况下,在暴露于Cas蛋白、CRISPR RNA和LTVEC之后,修饰基因组以含有至少10kb的核酸序列。在特定的实施方案中,所述LTVEC包含至少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或至少200kb的核酸序列。

[0258] v. 核酸酶试剂和核酸酶试剂的识别位点

[0259] 如上文详细概述,核酸酶试剂可在本文公开的方法和组合物中利用以帮助修饰在原核细胞中或在多潜能或非多潜能大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠细胞内的靶基因座。这一核酸酶试剂可促进在靶向载体和靶基因座之间的同源重组。在一个实施方案中,所述核酸酶试剂包括核酸内切酶试剂。

[0260] 本文使用的术语“核酸酶试剂的识别位点”包含在那里因核酸酶试剂诱导切口或双链断裂的DNA序列。核酸酶试剂的识别位点对细胞可为内源(或天然)的或所述识别位点对细胞可为外源的。在特定的实施方案中,所述识别位点对细胞为外源的且由此并不天然存在于细胞的基因组中。在更进一步的实施方案中,所述识别位点对细胞为外源的且对于希望在靶基因组基因座处安置的目标多核苷酸为外源的。在其它实施方案中,所述外源或内源识别位点仅一次存在于宿主细胞的基因组中。在特定的实施方案中,鉴定仅一次存在于基因组内的内源或天然位点。随后可使用这一位点以设计将在内源识别位点处生成切口或双链断裂的核酸酶试剂。

[0261] 所述识别位点的长度可改变且包括例如长至少4、6、8、10、12、14、16、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、41、42、43、44、45、46、47、48、49、50、51、52、53、54、55、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70个或更多个核苷酸的识别位点。在一个实施方案中,所述核酸酶试剂的各单体识别至少9个核苷酸的识别位点。在其它实施方案中,所述识别位点为长约9-约12个核苷酸、长约12-约15个核苷酸、长约15-约18个核苷酸或长约18-约21个核苷酸及这样的子范围的任何组合(例如,9-18个核苷酸)。所述识别位点可为回文的,即在一个链上的序列读出在相反方向上在互补链上的序列。应该认识到,给定核酸酶序列可结合所述识别位点且裂解所述结合位点,或可选地,所述核酸酶试剂可结合与所述结合位点不同的序列。此外,术语识别位点包含核酸酶试剂结合位点和切口/裂解位点两者,而不管切口/裂解位点是在核酸酶试剂结合位点内,还是超出核酸酶试剂结合位点。在另一变体中,由核酸酶试剂引起的裂解可在彼此直接相对的核酸位置处发生,以生成平滑末端切割或在其它情况下切口可错列以生成单链突出,也称作“粘性末端”,其可为5'突出或3'突出。

[0262] 诱导切口或双链断裂到所要识别位点中的任何核酸酶试剂可用于本文公开的方法和组合物中。可采用天然存在或天然的核酸酶试剂,条件是核酸酶试剂在所要识别位点中诱导切口或双链断裂。可选地,可采用修饰或工程化的核酸酶试剂。“工程化核酸试剂”包括由其天然形式工程化(修饰或得到)以在所要识别位点中特定识别并诱导切口或双链断裂的核酸酶。因此,工程化核酸酶试剂可来源于天然、天然存在的核酸酶试剂或者其可人工

地产生或合成。所述核酸酶试剂的修饰可为在蛋白质裂解试剂中的仅仅一个氨基酸或在核酸裂解试剂中的仅仅一个核苷酸。在一些实施方案中,所述工程化核酸酶在识别位点中诱导切口或双链断裂,其中所述识别位点不是将已经通过天然(未工程化或未修饰的)核酸酶试剂识别的序列。在识别位点或其它DNA中生成切口或双链断裂在本文中可称为“切割”或“裂解”识别位点或其它DNA。

[0263] 还提供了例示性识别位点的活性变体和片段。这样的活性变体可包括与给定识别位点至少65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更大的序列同一性,其中所述活性变体保持生物活性且因此能够由核酸酶试剂以序列特异性方式识别并裂解。测量由核酸酶试剂引起的识别位点双链断裂的测定在本领域中已知且通常测量核酸酶切割识别位点的能力。

[0264] 所述核酸酶试剂的识别位点可安置在靶基因座或靠近靶基因座的任何地方。所述识别位点可定位在基因的编码区内或定位在调控区内,其影响基因的表达。因此,所述核酸酶试剂的识别位点可定位在内含子、外显子、启动子、增强子、调控区或任何非蛋白编码区中。

[0265] 在一个实施方案中,所述核酸酶试剂为转录激活子样效应因子核酸酶(TALEN)。TAL效应因子核酸酶为可用以在原核或真核生物体的基因组中的特定靶序列处生成双链断裂的一类序列特异性核酸酶。TAL效应因子核酸酶通过将天然或工程化转录激活子样(TAL)效应因子或其功能部分融合到例如FokI的核酸内切酶的催化结构域中而产生。独特的模块化TAL效应因子DNA结合结构域允许以潜在的任何给定DNA识别特异性设计蛋白质。因此,TAL效应因子核酸酶的DNA结合结构域可工程化以识别特异性DNA靶位点,且因此用以在所靶序列处生成双链断裂。参见WO 2010/079430;Morbiter等人,(2010)PNAS 10.1073/pnas.1013133107;Scholze和Boch(2010)Virulence 1:428-432;Christian等人,Genetics (2010)186:757-761;Li等人,(2010)Nuc.Acids Res.(2010)doi:10.1093/nar/gkq704;和Miller等人,(2011)Nature Biotechnology 29:143-148,其全部以引用的方式并入本文中。

[0266] 合适TAL核酸酶的实例及用于制备合适TAL核酸酶的方法例如公开在美国专利申请号2011/0239315 A1、2011/0269234 A1、2011/0145940 A1、2003/0232410 A1、2005/0208489 A1、2005/0026157 A1、2005/0064474 A1、2006/0188987 A1和2006/0063231 A1(其各自以引用的方式并入本文中)。在各种实施方案中,TAL效应因子核酸酶工程化以在例如目标基因组基因座中的靶核酸序列中或靠近所述靶核酸序列切割,其中所述靶核酸序列处于将由靶向载体修饰的序列处或靠近所述序列。适合供本文提供的各种方法和组合物使用的TAL核酸酶包括特定设计成在将通过如本文所述的靶向载体修饰的靶核酸序列处或所述靶核酸序列附近结合的那些TAL核酸酶。

[0267] 在一个实施方案中,所述TALEN的各单体包含12-25个TAL重复序列,其中各TAL重复序列结合1bp亚基。在一个实施方案中,所述核酸酶试剂为包含操作性连接独立核酸酶的基于TAL重复序列的DNA结合结构域的嵌合蛋白。在一个实施方案中,所述独立核酸酶为FokI核酸内切酶。在一个实施方案中,所述核酸酶试剂包含第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域,其中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域各自操作性连接FokI核酸酶,其

中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域识别由约6bp-约40bp裂解位点间隔的在靶DNA序列的各链中的两个邻接靶DNA序列,且其中所述FokI核酸酶二聚合并靶序列处生成双链断裂。

[0268] 在一个实施方案中,所述核酸酶试剂包含第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域,其中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域各自操作性连接FokI核酸酶,其中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域识别由5bp或6bp裂解位点间隔的在靶DNA序列的各链中的两个邻接靶DNA序列,且其中所述FokI核酸酶二聚合并生成双链断裂。

[0269] 在本文公开的各种方法和组合物中采用的核酸酶试剂可进一步包含锌指核酸酶(ZFN)。在一个实施方案中,所述ZFN的各单体包含3个或更多个基于锌指的DNA结合结构域,其中各个基于锌指的DNA结合结构域结合3bp亚基。在其它实施方案中,所述ZFN为包含操作性连接独立核酸酶的基于锌指的DNA结合结构域的嵌合蛋白。在一个实施方案中,所述独立核酸内切酶为FokI核酸内切酶。在一个实施方案中,所述核酸酶试剂包含第一ZFN和第二ZFN,其中所述第一ZFN和所述第二ZFN各自操作性连接FokI核酸酶,其中所述第一ZFN和所述第二ZFN识别由约6bp-约40bp裂解位点或约5bp-约6bp裂解位点间隔的在靶DNA序列的各链中的两个邻接靶DNA序列,且其中所述FokI核酸酶二聚合并生成双链断裂。参见例如US20060246567;US20080182332;US20020081614;US20030021776;WO/2002/057308A2;US20130123484;US20100291048;和WO/2011/017293A2,其各自以引用的方式并入本文中。

[0270] 在本文提供的方法的一个实施方案中,所述核酸酶试剂包含(a)包含融合到FokI核酸内切酶的基于锌指的DNA结合结构域的嵌合蛋白;或(b)包含融合到FokI核酸内切酶的转录激活子样效应因子核酸酶(TALEN)的嵌合蛋白。

[0271] 在又一实施方案中,所述核酸酶试剂为大范围核酸酶。大范围核酸酶已经基于保守序列基序分成四个家族,所述家族为LAGLIDADG (SEQ ID NO:16)、GIY-YIG、H-N-H和His-Cys box家族。这些基序参与金属离子的配位和磷酸二酯键的水解。HEase因其长识别位点和在其DNA底物中耐受一些序列多态性而值得注意。大范围核酸酶结构域、结构和功能是已知的,参见例如Guhan和Muniyappa (2003) Crit Rev Biochem Mol Biol 38:199-248;Lucas等人,(2001) Nucleic Acids Res 29:960-9;Jurica和Stoddard,(1999) Cell Mol Life Sci 55:1304-26;Stoddard,(2006) Q Rev Biophys 38:49-95;和Moure等人,(2002) Nat Struct Biol 9:764。在一些实施例,使用天然存在的变体和/或工程化衍生物大范围核酸酶。修饰动力学、辅助因子相互作用、表达、最优条件和/或识别位点特异性和筛选活性的方法是已知的,参见例如Epinat等人,(2003) Nucleic Acids Res 31:2952-62;Chevalier等人,(2002) Mol Cell 10:895-905;Gimble等人,(2003) Mol Biol 334:993-1008;Seligman等人,(2002) Nucleic Acids Res 30:3870-9;Sussman等人,(2004) J Mol Biol 342:31-41;Rosen等人,(2006) Nucleic Acids Res 34:4791-800;Chames等人,(2005) Nucleic Acids Res 33:e178;Smith等人,(2006) Nucleic Acids Res 34:e149;Gruen等人,(2002) Nucleic Acids Res 30:e29;Chen和Zhao,(2005) Nucleic Acids Res 33:e154;WO2005105989;WO2003078619;WO2006097854;WO2006097853;WO2006097784;和WO2004031346。

[0272] 任何大范围核酸酶都可在本文中使用,所述大范围核酸酶包括但不限于I-SceI、I-SceII、I-SceIII、I-SceIV、I-SceV、I-SceVI、I-SceVII、I-CeuI、I-CeuAIIP、I-CreI、I-CrepsbIP、I-CrepsbIIP、I-CrepsbIIIP、I-CrepsbIVP、I-TliI、I-PpoI、PI-PspI、F-SceI、F-SceII、F-SuvI、F-TevI、F-TevII、I-AmaI、I-AniI、I-ChuI、I-CmoI、I-CpaI、I-CpaII、I-CsmI、I-CvuI、I-CvuAIP、I-DdiI、I-DdiII、I-DirI、I-DmoI、I-HmuI、I-HmuII、I-HsNIP、I-LlaI、I-MsoI、I-NaaI、I-NanI、I-NcIIP、I-NgrIP、I-NitI、I-NjaI、I-Nsp236IP、I-PakI、I-PboIP、I-PcuIP、I-PcuAI、I-PcuVI、I-PgrIP、I-PobIP、I-PorI、I-PorIIP、I-PbpIP、I-SpBetaIP、I-ScaI、I-SexIP、I-SneIP、I-SpomI、I-SpomCP、I-SpomIP、I-SpomIIP、I-SquIP、I-Ssp6803I、I-SthPhiJP、I-SthPhiST3P、I-SthPhiSTe3bP、I-TdeIP、I-TevI、I-TevII、I-TevIII、I-UarAP、I-UarHGPAIP、I-UarHGPA13P、I-VinIP、I-ZbiIP、PI-MtuI、PI-MtuHIP PI-MtuHIIP、PI-PfuI、PI-PfuII、PI-PkoI、PI-PkoII、PI-Rma43812IP、PI-SpBetaIP、PI-SceI、PI-TfuI、PI-TfuII、PI-ThyI、PI-TliI、PI-TliII或其任何活性变体或片段。

[0273] 在一个实施方案中,所述大范围核酸酶识别12-40个碱基对的双链DNA序列。在一个实施方案中,所述大范围核酸酶识别在基因组中的一个完美匹配的靶序列。在一个实施方案中,所述大范围核酸酶为归巢核酸酶。在一个实施方案中,所述归巢核酸酶为归巢核酸酶的LAGLIDADG (SEQ ID NO:16) 家族。在一个实施方案中,归巢核酸酶的LAGLIDADG (SEQ ID NO:16) 家族选自I-SceI、I-CreI和I-Dm1。

[0274] 核酸酶试剂可进一步包含限制核酸内切酶,其包括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异性识别位点,但通常在与核酸酶结合位点变化的位置裂解,其可距裂解位点(识别位点)数百个碱基对。在II型系统中,限制活性与任何甲基酶活性无关,且裂解通常在结合位点内或结合位点附近的特异性位点处发生。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且在识别位点之外裂解,IIb型酶用在识别位点之外的两个位点处切割序列两次,且II型酶识别不对称的识别位点且在一侧并在距识别位点约1-20个核苷酸的限定距离处裂解。IV型限制酶靶向甲基化DNA。限制酶例如在REBASE数据库(在rebase.neb.com处的网页;Roberts等人,(2003) *Nucleic Acids Res* 31:418-20);Roberts等人,(2003) *Nucleic Acids Res* 31:1805-12;和Belfort等人,(2002)在 *Mobile DNA II* 中,第761-783页,编者Craigie等人,(ASM Press,Washington,DC) 中进一步描述并分类。

[0275] 在各种方法和组合物中采用的核酸酶试剂还可包含CRISPR/Cas系统。这样的系统可采用例如Cas9核酸酶,其在一些情况下为对于其将表达的所要细胞类型而密码子优化的。这样的系统还可采用包含两种单独分子的向导RNA (gRNA)。例示性两分子gRNA包含crRNA-样(“CRISPR RNA”或“靶标-RNA”或“crRNA”或“crRNA重复序列”)分子和相应tracrRNA样(“反向作用CRISPR RNA”或“激活子-RNA”或“tracrRNA”或“支架”)分子。crRNA包含gRNA的DNA靶向片段(单链)和形成gRNA的蛋白结合片段的双链RNA(dsRNA)复体的一半的一段核苷酸两者。相应tracrRNA(激活子-RNA)包含形成gRNA的蛋白结合片段的dsRNA复体的另一半的一段核苷酸。因此,crRNA的一段核苷酸与tracrRNA的一段核苷酸互补并杂化以形成gRNA的蛋白结合结构域的dsRNA复体。因而,各crRNA可说成是具有相应的tracrRNA。所述crRNA另外提供单链DNA靶向片段。相应地,gRNA包含与杂化至靶序列的序列和tracrRNA。因此,crRNA和tracrRNA(作为相应的对)杂化以形成gRNA。如果用于在细胞内修



饰,则可将给定crRNA或tracrRNA分子的确切序列和/或长度设计成对将在其中使用RNA分子的物种具有特异性。

[0276] 编码三种元件(Cas9、tracrRNA和crRNA)的天然存在的基因通常在一种或多种操纵子中组织化。天然存在的CRISPR RNA根据Cas9系统和生物体而不同,但常常含有21-72个核苷酸长、由长度为21-46个核苷酸的两个定向重复序列(DR)侧接的靶向片段(参见例如,WO2014/131833)。在产脓链球菌的情况下,DR为36个核苷酸长且靶向片段30个核苷酸长。3'定位的DR与相应tracrRNA互补并杂化,其又结合Cas9蛋白。

[0277] 可选地,所述系统进一步采用与密码子优化的Cas9一起起作用的融合crRNA-tracrRNA构建体(即,单一转录体)。该单RNA常称为向导RNA或gRNA。在gRNA内,crRNA部分鉴定为给定识别位点的‘靶序列’且tracrRNA常称为‘支架’。简要地讲,含有靶序列的短DNA片段插入向导RNA表达质粒中。所述gRNA表达质粒包含靶序列(在一些实施方案中,约20个核苷酸)、一种形式的tracrRNA序列(支架)以及在细胞中具有活性的合适启动子和在真核细胞中恰当加工的必需元件。许多系统依靠定制的互补oligos,这些oligos退火以形成双链DNA且随后克隆到gRNA表达质粒中。gRNA表达盒和Cas9表达盒随后引入细胞中。参见,例如Mali P等人,(2013) Science 2013年2月15日;339(6121):823-6;Jinek M等人,Science 2012年8月17日;337(6096):816-21;Hwang WY等人,Nat Biotechnol 2013年3月;31(3):227-9;Jiang W等人,Nat Biotechnol 2013年3月;31(3):233-9;和Cong L等人,Science 2013年2月15日;339(6121):819-23,其各自以引用的方式并入本文中。还参见,例如WO/2013/176772A1、WO/2014/065596A1、WO/2014/089290A1、WO/2014/093622A2、WO/2014/099750A2和WO/2013142578A1,其各自以引用的方式并入本文中。

[0278] 在一些实施方案中,所述Cas9核酸酶可以蛋白质形式提供。在一些实施方案中,所述Cas9蛋白可以与gRNA的复合体的形式提供。在其它实施方案中,所述Cas9核酸酶可以编码蛋白质的核酸的形式提供。编码Cas9核酸酶的核酸可为RNA(例如,信使RNA(mRNA))或DNA。

[0279] 在一些实施方案中,所述gRNA可以RNA的形式提供。在其它实施方案中,所述gRNA可以编码RNA的DNA的形式提供。在一些实施方案中,所述gRNA可分别以单独crRNA和tracrRNA分子或编码crRNA和tracrRNA的单独DNA分子的形式提供。

[0280] 在一个实施方案中,所述修饰在细胞中的目标基因组基因座的方法进一步包括向所述细胞中引入:(a)包含操作性连接编码成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白的第一核酸序列的第一启动子的第一表达构建体;(b)包含操作性连接与向导RNA(gRNA)连接的基因组靶序列的第二启动子的第二表达构建体,其中所述基因组靶序列由原间隔区邻近基序侧接。任选地,所述基因组靶序列在3'端由原间隔区邻近基序(PAM)序列侧接。在一个实施方案中,所述细胞包括真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞。

[0281] 在一个实施方案中,所述基因组靶序列包含的GNNNNNNNNNNNNNNNNNNNGG核苷酸序列(GN<sub>1-20</sub>GG;SEQ ID NO:1)。在一个实施方案中,所述基因组靶序列包含SEQ ID NO:23,其中N为1-20个核苷酸长。在另一实施方案中,所述基因组靶序列包含14-20个核苷酸长的

SEQ ID NO:1。

[0282] 在一个实施方案中,所述gRNA包含编码成簇规律间隔短回文重复序列(CRISPR)RNA(crRNA)和反式激活CRISPR RNA(tracrRNA)的第三核酸序列。在特定的实施方案中,所述Cas蛋白为Cas9。

[0283] 在一些实施方案中,所述gRNA包含(a)核酸序列5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'(SEQ ID NO:2)的嵌合RNA;或(b)核酸序列5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAGGCUAGUCCG-3'(SEQ ID NO:3)的嵌合RNA。

[0284] 在另一实施方案中,所述crRNA包含5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAU-3'(SEQ ID NO:4);5'-GUUUUAGAGCUAGAAAUAGCAAGUAAAAUAAG(SEQ ID NO:5);或5'-GAGUCCGAGCAGAAGAAGAAGUUUA-3'(SEQ ID NO:6)。

[0285] 在又一实施方案中,所述tracrRNA包含5'-AAGGCUAGUCCG-3'(SEQ ID NO:7)或5'-AAGGCUAGUCCGU UAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'(SEQ ID NO:8)。

[0286] 在一个实施方案中,所述Cas蛋白为I型Cas蛋白。在一个实施方案中,所述Cas蛋白为II型Cas蛋白。在一个实施方案中,所述II型Cas蛋白为Cas9。在一个实施方案中,所述第一核酸序列编码人类密码子优化的Cas蛋白。

[0287] 在某些实施方案中,所述Cas蛋白为可在靶位点处产生单链断裂(即,“切口”)而不会切割双链DNA(dsDNA)的两个链的“切口酶”。Cas9例如包含两个核酸酶结构域:RuvC样核酸酶结构域和HNH样核酸酶结构域,其负责相反DNA链的裂解。在这些结构域中的任一个中的突变可产生切口酶。产生切口酶的突变的实例可例如在WO/2013/176772A1和WO/2013/142578A1中见到,这些申请各自以引用的方式并入本文中。

[0288] 在某些实施方案中,对在dsDNA的各个链上的靶位点具有特异性的两种单独Cas蛋白(例如,切口酶)可产生与在另一核酸上的突出序列或同一核酸上的单独区互补的突出序列。通过使核酸与对于在dsDNA的两个链上的靶位点具有特异性的两种切口酶接触产生的突出可为5'或3'突出。例如,第一切口酶可在dsDNA的第一链上产生第一链断裂,而第二切口酶可在dsDNA的第二链上产生单链断裂,因此产生突出序列。可选择各切口酶产生单链断裂的靶位点,使得所产生的突出序列与在不同核酸分子上的突出序列互补。这两种不同核酸分子的互补突出可通过本文公开的方法退火。在一些实施方案中,所述切口酶在第一链上的靶位点与所述切口酶在第二链上的靶位点不同。

[0289] 在一个实施方案中,所述第一核酸包含破坏在Cas蛋白中核酸酶活性位点的至少一个氨基酸残基的突变,其中突变体Cas蛋白仅在靶DNA区的一个链上产生断裂,且其中所述突变减小在靶DNA区中的非同源重组。

[0290] 在一个实施方案中,编码Cas蛋白的第一核酸进一步包含核定位信号(NLS)。在一个实施方案中,所述核定位信号为SV40核定位信号。

[0291] 在一个实施方案中,驱动基因组靶序列和向导RNA(gRNA)的表达的第二启动子为RNA聚合酶III启动子。在一个实施方案中,所述RNA聚合酶III启动子为人类U6启动子。在一个实施方案中,所述RNA聚合酶III启动子为大鼠U6聚合酶III启动子。在一个实施方案中,所述RNA聚合酶III启动子为小鼠U6聚合酶III启动子。

[0292] 在一个实施方案中,编码crRNA和tracrRNA的核酸序列经由合成环路连接,其中在

表达后crRNA和tracrRNA形成crRNA:tracrRNA复体。

[0293] 如上所述的CRISPR/Cas系统可与具有以下细胞类型中的任一种的大靶向载体组合使用:真核细胞、非大鼠真核细胞、哺乳动物细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞。

[0294] 在一个实施方案中,所述第一表达构建体和第二表达构建体自相同的质粒表达。

[0295] 在一个实施方案中,所述第一表达构建体和所述第二表达构建体与LTVEC一起引入。在一个实施方案中,所述第一表达构建体和所述第二表达构建体与所述LTVEC经一段时间单独地引入。

[0296] 在一个实施方案中,所述方法包括引入多个第二构建体和多个LTVEC以便多重编辑如本文所述的不同靶基因座。

[0297] 还提供核酸酶试剂(即,工程化核酸酶试剂)的活性变体和片段。这样的活性变体可包括与天然核酸酶试剂至少65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或更大的序列同一性,其中所述活性变体保持在所要识别位点切割的能力且因此保持切口或双链断裂诱导活性。例如,本文所述的核酸酶试剂中的任一种可自天然核酸内切酶序列修饰并设计用来识别并诱导在未被天然核酸酶试剂识别的识别位点处的切口或双链断裂。因此,在一些实施方案中,所述工程化核酸酶具有诱导在与相应天然核酸酶试剂识别位点不同的识别位点处的切口或双链断裂的特异性。对于切口或双链断裂诱导活性的测定是已知的且通常测量核酸内切酶对含有识别位点的DNA底物的总体活性和特异性。

[0298] 所述核酸酶试剂可通过本领域已知的任何方法引入细胞中。编码所述核酸酶试剂的多肽可直接引入细胞中。可选地,编码所述核酸酶试剂的多核苷酸可引入细胞中。当编码所述核酸酶试剂的多核苷酸引入细胞中时,所述核酸酶试剂可在细胞内短暂地、有条件地或组成性地表达。因此,编码所述核酸酶试剂的多核苷酸可包含在表达盒中且操作性连接条件性启动子、可诱导的启动子、组成性启动子或组织特异性启动子。这样的目标启动子在本文中的其它地方进一步详细地论述。可选地,所述核酸酶试剂作为编码或包含核酸酶试剂的mRNA引入细胞中。

[0299] 在一个实施方案中,所述crRNA和所述tracrRNA作为单独的RNA转录体表达。

[0300] 在特定的实施方案中,编码所述核酸酶试剂的多核苷酸稳定地整合到细胞的基因组中且操作性连接在细胞中具有活性的启动子。在其它实施方案中,编码所述核酸酶试剂的多核苷酸在包含插入核酸的相同靶向载体中,而在其它情况下,编码所述核酸酶试剂的多核苷酸在自包含插入核酸的靶向载体中分离的载体或质粒中。

[0301] 当所述核酸酶试剂经由引入编码所述核酸酶试剂的多核苷酸而提供到细胞中时,与编码所述核酸酶试剂的天然存在的多核苷酸序列相比较,可修饰编码核酸酶试剂的这一多核苷酸以取代在目标细胞中具有较高使用频度的密码子。例如,与天然存在的多核苷酸序列相比较,可修饰编码所述核酸酶试剂的多核苷酸以取代在给定的目标原核或真核细胞中具有较高利用频度的密码子,所述原核或真核细胞包括细菌细胞、酵母菌细胞、人类细胞、非人类细胞、非大鼠真核细胞、哺乳动物细胞、啮齿动物细胞、非大鼠啮齿动物细胞、小

鼠细胞、大鼠细胞、仓鼠细胞或任何其它目标宿主细胞。

[0302] 在一个实施方案中,所述核酸内切酶试剂与所述LTVEC一起引入。在一个实施方案中,所述核酸内切酶试剂与所述LTVEC经一段时间单独地引入。在一个实施方案中,所述核酸内切酶试剂在引入所述LTVEC之前引入。在一个实施方案中,所述核酸内切酶试剂在引入所述LTVEC之后引入大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠ES细胞中。

[0303] 在一个实施方案中,所述核酸内切酶试剂为包含编码核酸内切酶的核酸序列的表达构建体,其中所述核酸序列操作性连接启动子。在一个实施方案中,所述启动子为组成性活性启动子。在一个实施方案中,所述启动子为可诱导的启动子。在一个实施方案中,所述启动子在多潜能或非多潜能大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠细胞中具有活性。在一个实施方案中,所述核酸内切酶试剂为编码核酸内切酶的mRNA。

[0304] B. 将目标多核苷酸整合到靶基因座中的方法

[0305] 提供修饰目标靶基因座的方法。在一个实施方案中,靶向在多潜能或非多潜能大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠细胞中的靶基因座以便基因修饰。这一方法包括:(a) 将包含侧接有5'大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠同源臂和3'大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠同源臂的插入核酸的靶向载体引入多潜能或非多潜能大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠细胞中;(b) 鉴定在所述靶基因座处包括靶向基因修饰的基因修饰的多潜能或非多潜能大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠细胞,其中所述靶向基因修饰能够经由种系传递。在特定的实施方案中,所述5'同源臂和所述3'同源臂的总和为至少10kb和/或采用大靶向载体。

[0306] 在其它实施方案中,所述LTVEC的5'同源臂和3'同源臂的总和的尺寸为约10kb-约150kb、约10kb-约100kb、约10kb-约75kb、约20kb-约150kb、约20kb-约100kb、约20kb-约75kb、约30kb-约150kb、约30kb-约100kb、约30kb-约75kb、约40kb-约150kb、约40kb-约100kb、约40kb-约75kb、约50kb-约150kb、约50kb-约100kb、或约50kb-约75kb、约10kb-约30kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-约150kb。在一个实施方案中,所述缺失的尺寸与所述LTVEC的5'同源臂和3'同源臂的总和的尺寸相同或类似。

[0307] 例如大鼠细胞的多潜能细胞可为胚胎干细胞,例如大鼠胚胎干细胞。在一个特定的实施方案中,(a) 大鼠ES细胞来源于DA株或ACI株;或(b) 大鼠ES细胞的特征在于表达包括Oct-4、Sox-2、碱性磷酸酶或其组合的多潜能标记。在其它情况下,所采用的大鼠胚胎干细胞包括如在2014年2月20日提交的美国专利申请号14/185,103中所述的大鼠ES细胞,其全部内容以引用的方式并入本文中。

[0308] 任何多潜能或非多潜能细胞可用于本文提供的方法中。例如,所述多潜能或非多潜能细胞可来自真核生物、非大鼠真核生物、非人类哺乳动物、哺乳动物、啮齿动物、非大鼠啮齿动物、大鼠、小鼠、人类或仓鼠。

[0309] 如在本文中的其它地方描述,所述插入核酸可为任何核酸序列。在非限制性实施方案中,(a)所述插入核酸包含用同源或直系同源哺乳动物核酸序列替换内源大鼠、真核、非大鼠真核、哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸序列;(b)所述插入核酸包含缺失内源大鼠、真核、非大鼠真核、哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸序列;(c)所述插入核酸包含缺失内源大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠核酸序列,其中所述缺失为5kb-200kb或5kb-3Mb(如在本文中的其它地方详细论述);(d)所述插入核酸包含添加外源核酸序列(包括例如如下外源核酸序列,其为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb);(e)所述插入核酸包含包含同源或直系同源的核酸序列的外源核酸序列;(f) (a)的同源或直系同源的核酸序列,其中所述核酸序列为人类核酸序列;(g)所述插入核酸包含(a)的同源或直系同源的核酸序列,其中所述核酸序列为包含人类和大鼠核酸序列的嵌合核酸序列;(h)所述插入核酸包含(e)的外源核酸序列,其中所述插入核酸为约5kb-约200kb;(i)所述插入核酸包含侧接有位点特异性重组酶靶序列的条件性等位基因;(j)所述插入核酸包含操作性连接启动子的报道基因;(k)所述插入核酸包含一个或多个未重排的人类免疫球蛋白重链 $V_H$ 基因片段、一个或多个未重排的人类免疫球蛋白重链D基因片段和一个或多个未重排的人类免疫球蛋白重链 $J_H$ 基因片段,其操作性连接啮齿动物重链恒定区核酸序列;(l)所述插入核酸包含操作性连接啮齿动物重链恒定区核酸序列的重排的人类免疫球蛋白重链可变区核酸序列;(m)所述插入核酸包含一个或多个未重排的人类免疫球蛋白 $V_K$ 或 $V_L$ 基因片段和一个或多个未重排的人类免疫球蛋白 $J_K$ 或 $J_L$ 基因片段,它们操作性连接哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列;(n)所述插入核酸包含操作性连接哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列的重排人类免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链可变区核酸序列;(o) (k)和/或(l)的哺乳动物重链恒定区核酸序列包含大鼠恒定区核酸序列、人类恒定区核酸序列或其组合;或(p) (m)和/或(n)的哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链恒定区核酸包含大鼠恒定区核酸序列、人类恒定区核酸序列或其组合。

[0310] 在一个实施方案中,所述插入核酸包含一个或多个功能性人类 $V_H$ 基因片段,所述功能性人类 $V_H$ 基因片段包括 $V_H1-2$ 、 $V_H1-3$ 、 $V_H1-8$ 、 $V_H1-18$ 、 $V_H1-24$ 、 $V_H1-45$ 、 $V_H1-46$ 、 $V_H1-58$ 、 $V_H1-69$ 、 $V_H2-5$ 、 $V_H2-26$ 、 $V_H2-70$ 、 $V_H3-7$ 、 $V_H3-9$ 、 $V_H3-11$ 、 $V_H3-13$ 、 $V_H3-15$ 、 $V_H3-16$ 、 $V_H3-20$ 、 $V_H3-21$ 、 $V_H3-23$ 、 $V_H3-30$ 、 $V_H3-30-3$ 、 $V_H3-30-5$ 、 $V_H3-33$ 、 $V_H3-35$ 、 $V_H3-38$ 、 $V_H3-43$ 、 $V_H3-48$ 、 $V_H3-49$ 、 $V_H3-53$ 、 $V_H3-64$ 、 $V_H3-66$ 、 $V_H3-72$ 、 $V_H3-73$ 、 $V_H3-74$ 、 $V_H4-4$ 、 $V_H4-28$ 、 $V_H4-30-1$ 、 $V_H4-30-2$ 、 $V_H4-30-4$ 、 $V_H4-31$ 、 $V_H4-34$ 、 $V_H4-39$ 、 $V_H4-59$ 、 $V_H4-61$ 、 $V_H5-51$ 、 $V_H6-1$ 、 $V_H7-4-1$ 、 $V_H7-81$ 或其组合。

[0311] 在一个实施方案中,所述插入核酸包含一个或多个功能性人类D基因片段,所述功能性人类D基因片段包括D1-1、D1-7、D1-14、D1-20、D1-26、D2-2、D2-8、D2-15、D2-21、D3-3、D3-9、D3-10、D3-16、D3-22、D4-4、D4-11、D4-17、D4-23、D5-12、D5-5、D5-18、D5-24、D6-6、D6-13、D6-19、D6-25、D7-27或其组合。

[0312] 在一个实施方案中,所述插入核酸包含一个或多个功能性 $J_H$ 基因片段,所述功能性 $J_H$ 基因片段包括 $J_H1$ 、 $J_H2$ 、 $J_H3$ 、 $J_H4$ 、 $J_H5$ 、 $J_H6$ 或其组合。在一个实施方案中,所述插入核酸包含一个或多个人类 $V_K$ 基因片段,所述人类 $V_K$ 基因片段包括 $V_K4-1$ 、 $V_K5-2$ 、 $V_K7-3$ 、 $V_K2-4$ 、 $V_K1-$

5、Vk1-6、Vk3-7、Vk1-8、Vk1-9、Vk2-10、Vk3-11、Vk1-12、Vk1-13、Vk2-14、Vk3-15、Vk1-16、Vk1-17、Vk2-18、Vk2-19、Vk3-20、Vk6-21、Vk1-22、Vk1-23、Vk2-24、Vk3-25、Vk2-26、Vk1-27、Vk2-28、Vk2-29、Vk2-30、Vk3-31、Vk1-32、Vk1-33、Vk3-34、Vk1-35、Vk2-36、Vk1-37、Vk2-38、Vk1-39、Vk2-40或其组合。

[0313] 在一个实施方案中,所述插入核酸包含一个或多个人类Vλ基因片段,所述人类Vλ基因片段包括Vλ3-1、Vλ4-3、Vλ2-8、Vλ3-9、Vλ3-10、Vλ2-11、Vλ3-12、Vλ2-14、Vλ3-16、Vλ2-18、Vλ3-19、Vλ3-21、Vλ3-22、Vλ2-23、Vλ3-25、Vλ3-27或其组合。

[0314] 在一个实施方案中,所述插入核酸包含一个或多个人类Jκ基因片段,所述人类Jκ基因片段包括Jκ1、Jκ2、Jκ3、Jκ4、Jκ5或其组合。

[0315] 在特定的实施方案中,在修饰在多潜能或非多潜能大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠细胞中的靶基因座后,所述基因修饰经由种系传递。

[0316] 在一个实施方案中,所述插入核酸序列包含多核苷酸,所述多核苷酸在整合到基因组中时将生成大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠ApoE基因座的区的基因修饰,其中在ApoE基因座处的修饰引起ApoE活性减小、ApoE活性增加或ApoE活性调整。在一个实施方案中,产生ApoE敲除。

[0317] 在一个实施方案中,所述插入核酸序列包含多核苷酸,所述多核苷酸在整合到基因组中时将生成大鼠、真核、非大鼠真核、哺乳动物、人类、非人类哺乳动物、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠干扰白细胞素-2受体γ基因座的区的基因修饰,其中在所述干扰白细胞素-2受体γ基因座处的基因修饰引起干扰白细胞素-2受体活性减小、干扰白细胞素-2受体γ活性增加或干扰白细胞素-2受体活性调整。在一个实施方案中,产生干扰白细胞素-2受体敲除。

[0318] 在又一实施方案中,所述插入核酸序列包含多核苷酸,所述多核苷酸在整合到基因组中时将生成大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠Rag1基因座,大鼠、真核、非大鼠真核、非人类哺乳动物、哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠Rag2基因座和/或大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠Rag2/Rag1基因座的区的基因修饰,其中在大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠Rag1、Rag2和/或Rag2/Rag1基因座处的基因修饰引起Rag1、Rag2或Rag1和Rag2蛋白活性减小、Rag1、Rag2或Rag1和Rag2蛋白活性增加或Rag1、Rag2或Rag1和Rag2蛋白活性调整。在一个实施方案中,产生Rag1、Rag2或Rag2/Rag1敲除。

[0319] 在其它实施方案中,所述插入核酸引起大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠ApoE基因座,干扰白细胞素-2受体γ基因座和/或Rag2基因座和/或Rag1基因座和/或Rag2/Rag1基因座的一部分用来自另一生物体的ApoE基因座、干扰白细胞素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应直系同源部分替换。

[0320] 在其它实施方案中,所述插入核酸包含跨其全长与其替换的ApoE基因座、干扰白细胞素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分共有至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%的多核苷酸。

[0321] 给定插入的多核苷酸和/或大鼠、真核、非大鼠真核、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠基因座的相应替换区可为编码区、内含子、外显子、未转译区、调控区、启动子或增强子或其任何组合。此外,给定插入的多核苷酸和/或大鼠、真核、非大鼠真核、哺乳动物、人类、非人类哺乳动物、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠基因座的替换区可具有任何所要长度,所述长度包括例如10-100个核苷酸长、100-500个核苷酸长、500-1kb核苷酸长、1kb-1.5kb核苷酸长、1.5kb-2kb核苷酸长、2kb-2.5kb核苷酸长、2.5kb-3kb核苷酸长、3kb-5kb核苷酸长、5kb-8kb核苷酸长、8kb-10kb核苷酸长或更长。在其它情况下,所述插入或替换的尺寸为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb、约350kb-约400kb、约400kb-约800kb、约800kb-1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb、约2.5Mb-约2.8Mb、约2.8Mb-约3Mb。在其它实施方案中,给定插入的多核苷酸和/或大鼠、真核、非大鼠真核、非人类哺乳动物、哺乳动物、人类、啮齿动物、非大鼠啮齿动物、小鼠或仓鼠基因座的替换区为至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或更长。

[0322] i. 经由细菌同源重组 (BHR) 修饰核酸的靶基因座的方法

[0323] 提供经由在原核细胞中细菌同源重组 (BHR) 修饰真核、非大鼠真核、哺乳动物、人类或非人类哺乳动物核酸的靶基因座的方法和组合物。所述方法利用在原核细胞中细菌同源重组以基因修饰真核、非大鼠真核、哺乳动物、人类或非人类哺乳动物核酸的靶基因座以产生靶向载体中得到应用。包含基因修饰的靶基因座的这一靶向载体可引入真核细胞如真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞中。“同源重组”包括在同源性区内的交叉位点处交换在两个DNA分子之间的DNA片段。因此,“细菌同源重组”或“BHR”包括在细菌中发生的同源重组。

[0324] 提供经由细菌同源重组 (BHR) 修饰来自真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞的核酸的靶基因座的方法。所述方法包括将包含侧接有5'同源臂和3'同源臂的插入核酸的靶向载体引入原核细胞中,其中所述原核细胞包含核酸的靶基因座且能够表达介导在所述靶基因座处的BHR的重组酶。这样的靶向载体可包括本文所述的大靶向载体中的任一种。

[0325] 在一个实施方案中,所述方法包括向原核细胞中引入:(i) 包含具有目标DNA序列的核酸的第一构建体;(ii) 包含侧接有5'同源臂和3'同源臂的插入核酸的第二靶向构建体;和(iii) 编码介导细菌同源重组的重组酶的第三构建体。在一个实施方案中,所述第一构建体、所述第二构建体和所述第三构建体经一段时间单独地引入原核细胞中。在一个实施方案中,所述原核细胞包含编码所述重组酶的核酸,且所述方法不需要引入第三构建体。

在一个实施方案中,所述重组酶在可诱导的启动子的控制下表达。

[0326] 在一个实施方案中,包含所述核酸的第一构建体来源于细菌人造染色体(BAC)或酵母菌人造染色体(YAC)。可选择在靶基因组基因座处包含插入核酸的原核细胞。该方法可如本文公开连续地重复以允许在原核细胞中的靶基因座处引入多个插入核酸。靶核酸基因座一旦“建造”在原核细胞中,包含修饰的靶基因座的靶向载体则可自所述原核细胞中分离并引入在真核细胞、非大鼠真核细胞、哺乳动物细胞、人类细胞、非人类哺乳动物细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类iPS细胞、人类细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞、成纤维细胞或CHO细胞内的靶基因组基因座中。

[0327] 接受靶向载体的优选大鼠细胞描述在2014年2月20日提交的美国申请14/185,703中,其内容在本文中概述。这些大鼠细胞为能够在一种或多种体外靶向基因修饰之后维持其多潜能性且能够经由种系传递靶向基因修饰的多潜能大鼠细胞。

[0328] 电穿孔的多潜能细胞例如以高密度涂铺以便选择包含靶向载体的抗药性细胞。药物选择过程移除大多数的涂铺细胞(约99%),留下个别的集落,其各自为来源于单一细胞的克隆。在剩余细胞之中,大多数细胞(约80-100%)含有在基因组中的随机位置整合的靶向载体(包含药物选择盒)。因此,所述集落个别地拣选并基因分型以鉴定在恰当的基因组位置处藏有靶向载体的ES细胞(例如,使用如下所述的等位基因测定的修饰)。

[0329] 可使用高通量定量测定,即等位基因修饰(MOA)测定,以进行基因分型。这一测定允许在基因修饰之后大规模筛选在亲本染色体中的一种或多种修饰的等位基因。所述MOA测定可经由包括但不限于例如实时PCR(qPCR)的定量PCR的各种分析技术进行。例如,所述实时PCR包括识别靶基因座的第一引物集和识别非靶向参考基因座的第二引物集。另外,所述引物集包括识别扩增序列的荧光探针。在一个实施方案中,所述定量测定经由Invader Probes®进行。在一个实施方案中,所述定量测定经由MMP测定®进行。在一个实施方案中,所述定量测定经由TaqMan®分子信标进行。在一个实施方案中,所述定量测定经由Eclipse™探针技术进行。(参见,例如US2005/0144655,其以引用的方式并入本文中)。

[0330] 包含靶向基因修饰的选择的多潜能细胞(即,非人类多潜能细胞、非人类ES细胞)随后可引入例如前桑葚胚阶段或胚泡阶段胚胎的宿主胚胎中并植入代孕母体的子宫中以产生建群非人类动物(F0动物)。接着,建群动物例如可与野生型动物配种以产生基因修饰杂合性F1子代。杂合F1动物交配可生成基因修饰纯合性子代。杂合F1动物交配可生成基因修饰纯合性子代。在一些实施方案中,本文所述的靶基因座的各种基因修饰可使用如在本文中的其它地方详细描述的大靶向载体(LTVEC)进行。例如,LTVEC可使用VELOCIGENE®基因工程技术来源于细菌人造染色体(BAC)DNA(参见,例如美国专利号6,586,251和Valenzuela,D.M.等人,(2003),High-throughput engineering of the mouse genome coupled with high-resolution expression analysis,Nature Biotechnology 21(6):652-659,其整体以引用的方式并入本文中)。

[0331] 使用细菌同源重组(BHR)以产生大靶向载体(LTVEC)绕过了质粒在容纳大基因组DNA片段中的局限性以及随后将靶向修饰引入在多潜能或非多潜能细胞中的内源基因座中的低效率。一种或多种靶向基因修饰可在产生LTVEC的过程中执行。在原核细胞中生成的例



示性LTVEC可包含载有具有一种或多种基因修饰的基因组序列的插入核酸或由与特异性基因组区互补的同源臂侧接的外源核酸(例如,大鼠核酸的同源物或直系同源物)。

[0332] 还提供了包含本文所述的各种靶向载体的宿主原核细胞。这样的原核细胞包括但不限于细菌,例如大肠杆菌。在一个实施方案中,宿主原核细胞包含靶向载体,所述靶向载体包含侧接有5'同源臂和3'同源臂的插入核酸,其中所述插入核酸为约5kb-约200kb。

[0333] 所述宿主原核细胞可进一步包含编码重组酶多肽的核酸或编码操作性连接可诱导的启动子的重组酶多肽的核酸。

[0334] 进一步提供各种方法和组合物,其采用如本文所述的LTVEC以及原核细胞以生成靶向基因修饰。这样的组合物和方法在本文中的其它地方论述。

[0335] 提供经由细菌同源重组(BHR)修饰核酸的靶基因座的方法,其包括向原核细胞中引入包含侧接有5'同源臂和3'同源臂的插入核酸的靶向载体,其中所述原核细胞包含对应于5'同源臂和3'同源臂的核酸且所述原核细胞能够表达介导在靶基因座处的BHR的重组酶。这样的靶向载体可包括本文所述的大靶向载体中的任一种。这样的方法可采用如在本文中详细论述的LTVEC并且进一步采用如在本文中的其它地方论述的CRISPR/Cas系统。

[0336] 在一个实施方案中,所述CRISPR/Cas系统可通过在例如大肠杆菌的原核细胞中具有活性的启动子控制。

[0337] ii. 修饰在多潜能细胞或非多潜能细胞中的目标靶基因座的方法。

[0338] 进一步提供经由靶向基因修饰来修饰在多潜能细胞或非多潜能细胞中的目标靶基因座的方法,其包括:(a)向所述多潜能细胞或非多潜能细胞中引入包含侧接有5'同源臂和3'同源臂的插入核酸的靶向载体,其中所述5'同源臂和所述3'同源臂的总和为至少10kb;和(b)鉴定在所述目标靶基因座处具有靶向基因修饰的基因修饰的多潜能或非多潜能细胞。在一个实施方案中,所述5'同源臂和所述3'同源臂的总和为至少约16kb-约30kb。在特定的实施方案中所述靶向基因修饰能够经由种系传递。这样的靶向载体可包括本文所述的大靶向载体中的任一种。

[0339] 各种细胞也可用于本文提供的修饰目标靶基因座的方法中。在特定的实施方案中,所述细胞为真核细胞、非大鼠真核细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类诱导的多潜能细胞(iPS)细胞、哺乳动物细胞、人类细胞、成纤维细胞、啮齿动物细胞、非大鼠啮齿动物细胞、小鼠细胞、仓鼠细胞或CHO细胞。

[0340] 一方面,提供经由靶向基因修饰来修饰在多潜能细胞中的目标基因组基因座的方法,其包括:(a)提供能够在其基因组的至少一种靶向基因修饰之后维持其多潜能性且能够传递所述靶向修饰到F1代的种系的多潜能细胞;(b)将大靶向载体(LTVEC)引入所述多潜能细胞中,其中所述LTVEC包含侧接有5'同源臂和3'同源臂的插入核酸,其中所述5'同源臂和所述3'同源臂包含基因组DNA片段;和(c)鉴定包含所述靶向基因修饰的基因修饰的多潜能细胞。

[0341] 可使用各种方法以鉴定具有在目标靶基因座处整合的插入核酸的细胞。在目标靶基因座处插入插入核酸引起“等位基因修饰”。术语“等位基因修饰”和检测修饰的等位基因的方法在本文中的其它地方进一步详细地论述。

[0342] 一方面,提供经由核酸内切酶介导的基因靶向修饰在非多潜能细胞或多潜能细胞

中的目标基因组基因座的方法,所述方法包括:(a)提供能够传递基因修饰的基因组到F1代的种系的分离的非多潜能细胞或分离的多潜能细胞;(b)向所述非多潜能细胞或所述多潜能细胞中引入核酸内切酶试剂;其中所述核酸内切酶试剂在位于所述目标基因组基因座中的靶DNA序列处生成切口或双链断裂,且其中在所述非多潜能细胞或所述多潜能细胞中的靶DNA序列处的切口或双链断裂诱导:(i)切口或双链断裂的非同源端接合(NHEJ)介导的DNA修复,其中所述NHEJ介导的DNA修复产生包含在靶DNA序列处核酸序列的插入或缺失的突变等位基因;或(ii)引起野生型核酸序列复原的同源重组介导的DNA修复;和(c)鉴定修饰的目标基因组基因座。

[0343] 一方面,提供经由核酸酶试剂修饰在分离的胚胎干细胞(ES)中的目标基因组基因座的方法,其包括:(a)提供能够传递靶向基因修饰到F1代的种系的分离的ES细胞;(b)向所述ES细胞中引入:(i)包含侧接有5'同源臂和3'同源臂的插入核酸的大靶向载体(LTVEC),其中所述插入为至少5kb的核酸序列;和(ii)核酸内切酶试剂,其中所述核酸内切酶试剂在位于所述目标基因组基因座中的靶DNA序列处生成切口或双链断裂,且其中所述靶序列在所述插入核酸中不存在;和(c)鉴定在所述胚胎干(ES)细胞中的靶向基因修饰。

[0344] 一方面,提供经由RNA向导的基因组工程化修饰在非多潜能细胞或多潜能细胞中的目标基因组基因座的方法,所述方法包括:(a)提供能够传递基因修饰的基因组到F1代的种系的非多潜能细胞或多潜能细胞;(b)向所述非多潜能细胞或所述多潜能细胞中引入:(i)包含操作性连接编码成簇规律间隔短回文重复序列(CRISPR)相关(Cas)蛋白的第一核酸序列的第一启动子的第一表达构建体,(ii)包含操作性连接与向导RNA(gRNA)连接的基因组靶序列的第二启动子的第二表达构建体,其中所述基因组靶序列由原间隔区邻近基序(PAM)序列侧接。任选地,所述基因组靶序列在3'端由原间隔区邻近基序(PAM)序列侧接。在一个实施方案中,所述Cas蛋白和所述CRISPR RNA和/或tracrRNA并不一起天然存在(例如,Cas蛋白和CRISPR RNA并不一起天然存在)。在一个实施方案中,所述基因组靶序列包含GNNNNNNNNNNNNNNNNNNNGG的核苷酸序列(GN<sub>1-20</sub>GG;SEQ ID NO:1)。在一个实施方案中,所述基因组靶序列包含SEQ ID NO:1,其中N为14-20个核苷酸长。在一个实施方案中,所述gRNA包含编码成簇规律间隔短回文重复序列(CRISPR)RNA(crRNA)的第三核酸序列和编码反式激活CRISPR RNA(tracrRNA)的第四核酸序列。在一个实施方案中,在表达后,所述Cas蛋白形成包括crRNA和tracrRNA的CRISPR-Cas复合体,且所述CRISPR-Cas复合体在位于所述目标基因组基因座中的靶DNA序列处生成切口或双链断裂,且其中在所述非多潜能细胞或所述多潜能细胞中的靶DNA序列处的切口或双链断裂诱导:(i)通过所述CRISPR-Cas复合体产生的切口或双链断裂的非同源端接合(NHEJ)介导的DNA修复,其中所述NHEJ产生包含在所述靶DNA序列处核酸序列的插入或缺失的突变等位基因;或(ii)引起野生型核酸序列复原的同源重组介导的DNA修复;和(c)鉴定修饰的目标基因组基因座。

[0345] 一方面,提供经由RNA向导的基因组工程化修饰在非多潜能细胞或多潜能细胞中的目标基因组基因座的方法,所述方法包括:向能够经由种系传递修饰的基因组的非多潜能细胞或多潜能细胞中引入:(i)成簇规律间隔短回文重复序列(CRISPR)-相关(Cas)蛋白或编码所述Cas蛋白的核酸;和(ii)gRNA或编码所述gRNA的DNA,其中所述gRNA包含杂化到基因组靶序列的核苷酸序列和反式激活CRISPR RNA(tracrRNA);其中所述基因组靶序列由原间隔区邻近基序(PAM)序列侧接。

[0346] 在一些实施方案中,所述Cas蛋白可作为分离的蛋白引入所述非多潜能细胞或所述多潜能细胞中。在一些实施方案中,所述Cas蛋白可进一步包含促进所述蛋白的细胞摄取的细胞穿透结构域。在其它实施方案中,所述Cas蛋白可作为编码所述Cas蛋白的信使RNA (mRNA) 分子引入细胞中。在其它实施方案中,所述Cas蛋白可作为编码所述Cas蛋白的DNA分子引入细胞中。例如,编码所述Cas蛋白的DNA分子可提供于构建体中且操作性连接能够在所述非多潜能细胞或所述多潜能细胞中表达的启动子。在某些实施方案中,编码所述Cas蛋白的核酸为密码子优化以在所述非多潜能细胞或所述多潜能细胞中表达。

[0347] 在一些实施方案中,所述gRNA可作为RNA分子引入所述非多潜能细胞或所述多潜能细胞中。例如,所述gRNA分子可体外转录。在其它实施方案中,所述gRNA可作为编码所述gRNA的DNA分子引入所述非多潜能细胞或所述多潜能细胞中。例如,编码所述gRNA的DNA分子可在构建体中且操作性连接能够在所述非多潜能细胞或所述多潜能细胞中表达所述gRNA的启动子。在其它实施方案中,所述gRNA可化学合成。

[0348] 在一些实施方案中,所述gRNA可作为融合的crRNA-tracrRNA分子(即,单一转录体)引入所述非多潜能细胞或所述多潜能细胞中。在其它实施方案中,所述gRNA可作为单独的crRNA和tracrRNA分子(即,单独的转录体)引入所述非多潜能细胞或所述多潜能细胞中。在其它实施方案中,所述gRNA可作为分别编码crRNA和tracrRNA的单独的DNA分子引入所述非多潜能细胞或所述多潜能细胞中。例如,编码所述crRNA和tracrRNA的单独的DNA分子可在单独的构建体中且操作性连接能够在所述非多潜能细胞或所述多潜能细胞中表达的启动子。在上述实施方案中的任一个中,所述构建体的任何组合可在单独的核酸分子中或一起在单一核酸分子中。

[0349] 在一些实施方案中,所述Cas蛋白和所述gRNA可同时或依次引入所述非多潜能细胞或所述多潜能细胞中。同样,所述gRNA的crRNA和tracrRNA可同时或依次引入所述非多潜能细胞或所述多潜能细胞中。Cas蛋白(或编码核酸)与gRNA(或编码DNA)的比率和/或crRNA与tracrRNA的比率可为大致化学计量的,因此它们可形成RNA-蛋白质复合体。

[0350] 在某些实施方案中,所述Cas蛋白可以与gRNA的复合体的形式引入所述非多潜能细胞或所述多潜能细胞中。

[0351] 在一个实施方案中,所述多潜能细胞为诱导的多潜能干细胞(iPS)。在一个实施方案中,所述多潜能细胞为发育受限的祖细胞。

[0352] 在各种实施方案中,在选择标记内的识别位点中切口或双链断裂的存在增加在靶向载体(例如LTVEC)和目标靶基因座之间的重组的效率和/或频率。在一个实施方案中,所述重组为同源重组。在另一实施方案中,所述重组为通过非同源端接合插入。在各种实施方案中,在存在所述切口或双链断裂的情况下,在靶基因组基因座处靶向载体(例如,LTVEC)的靶向效率为在缺乏切口或双链断裂的情况下(例如,在目标基因组基因座处使用相同的靶向载体和相同的同源臂及相应靶位点,但在缺乏生成切口或双链断裂的加入的核酸酶试剂的情况下)的至少约2倍、至少约3倍、至少约4倍。

[0353] 在一个实施方案中,在靶基因座处的靶向基因修饰为双等位基因。“双等位基因”是指基因的两个等位基因包含靶向基因修饰。所述靶向基因修饰可在各等位基因中相同或不同。例如,双等位基因修饰可由对在相应同源染色体上的相应等位基因产生的相同修饰或对在相应同源染色体上的相应等位基因产生的不同修饰产生。因此,双等位基因修饰可

例如对于在目标基因组基因座处的特异性修饰产生纯合(即,在两种等位基因中的特异性修饰)、在目标基因组基因座处产生复合杂合(例如,在一个等位基因中的特异性修饰和另一基因的失活或破坏)或在目标基因组基因座处产生半合(例如,在一个等位基因中的特异性修饰和另一等位基因的损失)。在某些实施方案中,与仅使用靶向载体相比较,靶向载体(包括例如LTVEC)与核酸酶试剂的组合使用产生在细胞中的目标基因组基因座的双等位基因靶向基因修饰。与在仅使用靶向载体时相比较时,当靶向载体结合核酸酶剂使用时,双等位基因靶向效率增加至少两倍、至少三倍、至少四倍或更多。在其它实施方案中,所述双等位基因靶向效率为至少0.2%、0.3%、0.4%、0.5%、0.6%、0.7%、0.8%、0.9%、1%、2%、3%、4%或5%或更高。

[0354] 在靶基因座处的双等位基因靶向基因修饰可产生纯合基因修饰的细胞。“纯合”是指靶基因座的两个等位基因(即,在两种同源染色体上的等位基因)已经以相同方式修饰。在某些实施方案中,靶向载体(包括例如LTVEC)与核酸酶试剂的组合使用产生在细胞中的目标基因组基因座的双等位基因纯合靶向基因修饰。在一个实施方案中,所述双等位基因基因修饰包括在两种同源染色体(即,一对第一同源染色体和第二同源染色体)中的目标基因组基因座处内源核酸序列的缺失和在两种同源染色体(即,一对第一同源染色体和第二同源染色体)中的目标基因组基因座处插入核酸的插入。在一些实施方案中,所述插入核酸替换在两种同源染色体中的目标基因组基因座处的内源核酸序列。在一个实施方案中,所述插入核酸与缺失的内源核酸序列同源或直系同源。

[0355] 在一个实施方案中,在所述靶基因座处的靶向基因修饰产生半合基因修饰的细胞。“半合”是指仅存在靶基因座的一个等位基因(即,在两种同源染色体之一上的等位基因)或仅一个等位基因能够表达或功能化。在其它实施方案中,所述靶向基因修饰更一般而言产生复合杂合。复合杂合包括其中靶基因座的两个等位基因(即,在两种同源染色体上的等位基因)已经修饰、但它们是以不同方式修饰(例如,在一个等位基因中插入和另一等位基因失活或破坏)的情形。在某些实施方案中,靶向载体(包括例如LTVEC)与核酸酶试剂的组合使用产生在细胞中的目标基因组基因座的半合靶向基因修饰。在某些实施方案中,靶向载体(包括例如LTVEC)与核酸酶试剂是组合使用在细胞中的目标基因组基因座处产生引起复合杂合的靶向基因修饰。在一个实施方案中,在一种染色体中的目标基因组基因座处的靶向基因修饰包括内源核酸序列的缺失和插入核酸的插入。在其它实施方案中,所述靶向基因修饰包括:(1)在两种同源染色体中的目标基因组基因座处缺失内源核酸序列;和(2)将插入核酸插入在第一染色体中的目标基因组基因座中和破坏在第二染色体中的目标基因组基因座。所述第一染色体可为这两种同源染色体中的第一种,且所述第二染色体可为这两种同源染色体中的第二种。在其它实施方案中,所述靶向修饰包括:(1)在目标基因组基因座处缺失内源核酸序列和将插入核酸插入在第一同源染色体中的目标基因组基因座中;和(2)破坏在第二同源染色体中的目标基因组基因座。内源核酸序列的破坏可例如在由核酸酶试剂产生的在目标基因组基因座处的双链断裂通过非同源端接合(NHEJ)介导的DNA修复而修复时产生,其产生包括在目标基因组基因座处核酸序列的插入或缺失的突变等位基因且由此造成目标基因组基因座的破坏。破坏的实例包括在目标基因组基因座处调控元件(例如,启动子或增强子)改变、错义突变、截断突变、零突变或小数量的核苷酸插入或缺失(例如,引起移码突变)。破坏的另一实例为无义突变。破坏可引起失活(即,功能损

失)或等位基因损失。

[0356] 纯合和半合的靶向基因修饰是有利的,因为当使用含有这些突变的基因修饰的细胞以产生如下论述的基因修饰的动物时,产生对于预定靶向基因修饰为非杂合(即,纯合或半合)的基因修饰的动物的方法更有效且不费时,因为需要较少的繁育步骤。出于相同的原因,引起复合杂合或半合的靶向基因修饰(例如,在一个等位基因中插入和另一等位基因失活、破坏或损失)可为有利的。

[0357] 各种细胞类型也可用于上文描述的各种方法中的任一种中,以便经由核酸酶试剂修饰基因组基因座。在特定的实施方案中,所述细胞为真核细胞、非大鼠真核细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类诱导的多潜能细胞(iPS)细胞、哺乳动物细胞、人类细胞、成纤维细胞、啮齿动物细胞、非大鼠啮齿动物细胞、小鼠细胞、仓鼠细胞或CHO细胞。

[0358] 提供组合物,其包含在干扰白细胞素-2受体 $\gamma$ 基因座中或在ApoE基因座中具有靶向基因修饰的基因修饰的非人类动物。本文提供的各种方法和组合物允许这些修饰的基因座经由种系传递。

[0359] 在特定的实施方案中,基因修饰的非人类动物或基因修饰的多潜能或非多潜能细胞包含在干扰白细胞素-2 $\gamma$ 受体基因座中具有靶向基因修饰或在ApoE基因座中具有靶向基因修饰的基因组基因座,其中所述干扰白细胞素-2 $\gamma$ 受体基因组基因座或所述ApoE基因座包含:(i)缺失所述干扰白细胞素-2 $\gamma$ 受体基因座的至少一部分或所述ApoE基因座的至少一部分;(ii)异源核酸序列插入所述ApoE基因座中或插入所述干扰白细胞素-2 $\gamma$ 受体基因座中;或(iii)其组合,其中所述基因修饰的基因组基因座能够经由种系传递。

[0360] 进一步提供了允许产生这样的基因修饰的非人类动物和这样的基因修饰的多潜能细胞的方法。这样的方法包括经由靶向基因修饰修饰在多潜能细胞中的ApoE基因组基因座或干扰白细胞素-2 $\gamma$ 受体基因座的方法。所述方法包括(a)向所述多潜能细胞中引入包含侧接有ApoE基因座的5'同源臂和ApoE基因座的3'同源臂的插入核酸的靶向载体,(b)鉴定在目标ApoE基因组基因座处包括靶向基因修饰的基因修饰的多潜能细胞,其中所述靶向基因修饰能够经由种系传递。

[0361] 另外的方法包括(a)向所述多潜能细胞中引入包含侧接有所述干扰白细胞素-2受体 $\gamma$ 基因座的5'同源臂和所述干扰白细胞素-2受体 $\gamma$ 基因座的3'同源臂的插入核酸的靶向载体,(b)鉴定在所述干扰白细胞素-2受体 $\gamma$ 基因座处包括靶向基因修饰的基因修饰的多潜能细胞,其中所述靶向基因修饰能够经由种系传递。

[0362] iii.在靶基因座处整合多个目标多核苷酸的方法

[0363] 本文提供的各种方法和组合物允许靶向整合多个目标多核苷酸与给定的靶基因座。上文阐述的各种方法可依次重复以允许任何数目的插入核酸靶向整合到给定的靶基因座中。因此,提供各种方法以将至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20个或更多个插入核酸插入靶基因座中。在特定的实施方案中,这样的连续铺装方法允许由真核细胞如非大鼠真核细胞、哺乳动物细胞(即,人类、非人类、啮齿动物、非大鼠啮齿动物、小鼠、猴、大鼠、仓鼠、家养哺乳动物或农业动物)的大基因组区重建成靶基因座。在这样的情况下,包括编码区和非编码区两者的基因组区的转移和重建允许通过至少部分地保留在天然基因组区内发现的编码区、非编码区和拷贝数变异来保持给定区的复杂性。因此,

所述各种方法提供例如在任何真核细胞、任何非大鼠真核细胞、任何哺乳动物细胞或目标动物内、特别是在原核宿主细胞内或在非多潜能细胞、多潜能细胞或ES细胞内产生“杂合”或“外源”基因组区的方法。在一个非限制性实例中,产生在非人类动物(即,大鼠)内的“人源化”基因组区。本文提供在任何细胞内产生基因组区的方法。在特定的实施方案中,所述细胞为真核细胞、非大鼠真核细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类诱导的多潜能细胞(iPS)细胞、哺乳动物细胞、人类细胞、成纤维细胞、啮齿动物细胞、非大鼠啮齿动物细胞、小鼠细胞、仓鼠细胞或CHO细胞。

### [0364] 3. 人源化基因组基因座

[0365] 本文提供包含人源化基因组基因座的各种方法和组合物。本文使用的“人源化”基因组基因座是指包含至少一个人类核酸序列的非人类基因组的区。所述人源化基因组基因座可包含具有插入其中的人类DNA序列的来自任何生物体的DNA的区。在特定的实施方案中,所述生物体为真核生物、非大鼠真核生物、非人类哺乳动物、哺乳动物、人类、啮齿动物、非大鼠啮齿动物、大鼠、小鼠或仓鼠。例如,“人源化大鼠基因座”包含具有插入其中的人类DNA序列的大鼠DNA的区。

[0366] 所述人类DNA序列可为天然存在的人类DNA序列或者其可自其天然形式修饰。在特定的实施方案中,所述人类DNA与天然人类序列共有至少85%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%的序列同一性。如果人类序列不是天然人类序列,则与直系同源非人类序列相比,其与天然类似序列至少具有更大的序列同一性。此外,所述人类DNA序列可包含cDNA、人类基因组DNA的区、非编码调控区或人类DNA的编码、基因组或调控区的任何部分。插入非人类基因座中的人类DNA序列可包含如在本文中的其它地方描述的插入多核苷酸中的任一种。在特定的实施方案中,所述人类DNA序列与所述非人类靶基因座直系同源,而在其它情况下,所述人类DNA序列与所述非人类靶基因座同源。

[0367] 在一个实施方案中,所述靶向基因修饰为插入同源或直系同源人类核酸序列或用其替换内源核酸序列。在一个实施方案中,所述靶向基因修饰包括在包含相应非人类核酸序列的内源基因座处插入同源或直系同源的人类核酸序列或用其替换内源核酸序列。

[0368] 制造人源化基因座的方法包括向包含核酸的靶基因座中引入人类核酸序列。在一个实施方案中,提供制造人源化非人类动物的方法。这一方法包括:(a)用包含插入核酸的靶向载体修饰非人类多潜能细胞或非多潜能细胞的基因组以形成供体细胞,所述插入核酸包含人类核酸序列;(b)将所述供体细胞引入宿主胚胎中;和(c)在代孕母体中孕育所述宿主胚胎;其中所述代孕母体生成包含所述人类核酸序列的子代。在特定的实施方案中,所述人源化基因座能够经由种系传递。在另一实施方案中,所述靶向载体包括大靶向载体(LTVEC)且包含人类核酸序列的所述插入核酸为至少5kb。

[0369] 在其它方法中,所述人源化基因组基因座通过经由细菌同源重组(BHR)修饰核酸的靶基因座制造。所述方法包括向原核细胞中引入包含侧接有5'同源臂和3'同源臂的插入核酸的靶向载体,其中所述插入核酸包含人类核酸序列,且其中所述原核细胞包含核酸且能够表达介导在靶基因座处的BHR的重组酶。

[0370] 所述人源化基因组基因座可包含(a)插入同源或直系同源人类核酸序列;(b)用同源或直系同源人类核酸序列替换内源核酸序列;或(c)其组合。在特定的实施方案中,所述

人源化基因组基因座能够经由种系传递。在其它实施方案中,所述人类直系同源序列替换在非人类基因座中见到的相应序列。

[0371] 任何人类核酸序列都可用于本文提供的方法和组合物中。可用于所述方法和组合物中的人类核酸序列的非限制性实例在本文中的其它地方详细论述。

[0372] 插入目标基因座中的人类核酸序列可为任何尺寸。在一个实施方案中,所述人类核酸序列可为约500个核苷酸-约200kb、约500个核苷酸-约5kb、约5kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约30kb、约30kb-约40kb、约40kb-约50kb、约60kb-约70kb、约80kb-约90kb、约90kb-约100kb、约100kb-约110kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb或约190kb-约200kb。在一个特定的实施方案中,所述人类核酸序列为至少5kb。

[0373] 在一个实施方案中,提供基因组基因座,其中所述同源或直系同源人类核酸序列包含(a)一个或多个未重排的人类免疫球蛋白重链 $V_H$ 基因片段、一个或多个未重排的人类免疫球蛋白重链D基因片段和一个或多个未重排的人类免疫球蛋白重链 $J_H$ 基因片段,其操作性连接哺乳动物重链恒定区核酸序列;(b)操作性连接哺乳动物免疫球蛋白重链恒定区核酸序列的重排的人类免疫球蛋白重链可变区核酸序列;(c)一个或多个未重排的人类免疫球蛋白 $V_K$ 或 $V_L$ 基因片段和一个或多个未重排的人类免疫球蛋白 $J_K$ 或 $J_L$ 基因片段,其操作性连接哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列;或(d)操作性连接哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列的重排的人类免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链可变区核酸序列。

[0374] 在另一实施方案中,提供基因组基因座,其中(a)所述哺乳动物免疫球蛋白重链恒定区核酸序列为恒定区核酸序列、人类恒定区核酸序列或其组合;或(b)所述哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列为小鼠恒定区核酸序列、人类恒定区核酸序列或其组合。

[0375] 在一个特定的实施方案中,提供基因组基因座,其中所述免疫球蛋白重链恒定区核酸序列选自或包含CH1、铰链、CH2、CH3和/或其组合。

[0376] 在一个实施方案中,所述基因组基因座包含一个或多个功能性人类 $V_H$ 基因片段,所述功能性人类 $V_H$ 基因片段包括 $V_H1-2$ 、 $V_H1-3$ 、 $V_H1-8$ 、 $V_H1-18$ 、 $V_H1-24$ 、 $V_H1-45$ 、 $V_H1-46$ 、 $V_H1-58$ 、 $V_H1-69$ 、 $V_H2-5$ 、 $V_H2-26$ 、 $V_H2-70$ 、 $V_H3-7$ 、 $V_H3-9$ 、 $V_H3-11$ 、 $V_H3-13$ 、 $V_H3-15$ 、 $V_H3-16$ 、 $V_H3-20$ 、 $V_H3-21$ 、 $V_H3-23$ 、 $V_H3-30$ 、 $V_H3-30-3$ 、 $V_H3-30-5$ 、 $V_H3-33$ 、 $V_H3-35$ 、 $V_H3-38$ 、 $V_H3-43$ 、 $V_H3-48$ 、 $V_H3-49$ 、 $V_H3-53$ 、 $V_H3-64$ 、 $V_H3-66$ 、 $V_H3-72$ 、 $V_H3-73$ 、 $V_H3-74$ 、 $V_H4-4$ 、 $V_H4-28$ 、 $V_H4-30-1$ 、 $V_H4-30-2$ 、 $V_H4-30-4$ 、 $V_H4-31$ 、 $V_H4-34$ 、 $V_H4-39$ 、 $V_H4-59$ 、 $V_H4-61$ 、 $V_H5-51$ 、 $V_H6-1$ 、 $V_H7-4-1$ 、 $V_H7-81$ 或其组合。

[0377] 在一个实施方案中,所述基因组基因座包含一个或多个功能性人类D基因片段,所述功能性人类D基因片段包括D1-1、D1-7、D1-14、D1-20、D1-26、D2-2、D2-8、D2-15、D2-21、D3-3、D3-9、D3-10、D3-16、D3-22、D4-4、D4-11、D4-17、D4-23、D5-12、D5-5、D5-18、D5-24、D6-6、D6-13、D6-19、D6-25、D7-27或其组合。

[0378] 在一个实施方案中,所述基因组基因座包含一个或多个功能性 $J_H$ 基因片段,所述功能性 $J_H$ 基因片段包括 $J_H1$ 、 $J_H2$ 、 $J_H3$ 、 $J_H4$ 、 $J_H5$ 、 $J_H6$ 和/或其组合。在一个实施方案中,所述插入核酸包含一个或多个人类 $V_K$ 基因片段,所述人类 $V_K$ 基因片段包括 $V_K4-1$ 、 $V_K5-2$ 、 $V_K7-3$ 、 $V_K2-4$ 、 $V_K1-5$ 、 $V_K1-6$ 、 $V_K3-7$ 、 $V_K1-8$ 、 $V_K1-9$ 、 $V_K2-10$ 、 $V_K3-11$ 、 $V_K1-12$ 、 $V_K1-13$ 、 $V_K2-14$ 、 $V_K3-15$ 、

$\kappa 1-16$ 、 $V\kappa 1-17$ 、 $V\kappa 2-18$ 、 $V\kappa 2-19$ 、 $V\kappa 3-20$ 、 $V\kappa 6-21$ 、 $V\kappa 1-22$ 、 $V\kappa 1-23$ 、 $V\kappa 2-24$ 、 $V\kappa 3-25$ 、 $V\kappa 2-26$ 、 $V\kappa 1-27$ 、 $V\kappa 2-28$ 、 $V\kappa 2-29$ 、 $V\kappa 2-30$ 、 $V\kappa 3-31$ 、 $V\kappa 1-32$ 、 $V\kappa 1-33$ 、 $V\kappa 3-34$ 、 $V\kappa 1-35$ 、 $V\kappa 2-36$ 、 $V\kappa 1-37$ 、 $V\kappa 2-38$ 、 $V\kappa 1-39$ 、 $V\kappa 2-40$ 或其组合。

[0379] 在一个实施方案中,所述基因组基因座包含一个或多个人类 $V\lambda$ 基因片段,所述人类 $V\lambda$ 基因片段包括 $V\lambda 3-1$ 、 $V\lambda 4-3$ 、 $V\lambda 2-8$ 、 $V\lambda 3-9$ 、 $V\lambda 3-10$ 、 $V\lambda 2-11$ 、 $V\lambda 3-12$ 、 $V\lambda 2-14$ 、 $V\lambda 3-16$ 、 $V\lambda 2-18$ 、 $V\lambda 3-19$ 、 $V\lambda 3-21$ 、 $V\lambda 3-22$ 、 $V\lambda 2-23$ 、 $V\lambda 3-25$ 、 $V\lambda 3-27$ 或其组合。

[0380] 在一个实施方案中,所述基因组基因座包含一个或多个人类 $J\kappa$ 基因片段,所述人类 $J\kappa$ 基因片段包括 $J\kappa 1$ 、 $J\kappa 2$ 、 $J\kappa 3$ 、 $J\kappa 4$ 、 $J\kappa 5$ 或其组合。

[0381] 在又一实施方案中,提供包括人源化基因组基因座的基因组基因座,所述人源化基因组基因座包含人类干扰白细胞素-2受体(IL2R)核酸序列或其变体或片段。在特定的实施方案中,所述IL2R核酸序列包含干扰白细胞素-2受体 $\alpha$ 、干扰白细胞素-2受体 $\beta$ 或干扰白细胞素-2受体 $\gamma$ 核酸序列或其变体或片段。

[0382] 在其它实施方案中,基因组基因座包括人源化基因组基因座,所述人源化基因组基因座包含替换非人类ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应同源或直系同源部分的人类ApoE基因座、人类干扰白细胞素-2受体 $\gamma$ 基因座、人类Rag2基因座、人类Rag1基因座和/或人类Rag2/Rag1基因座的一部分。在一个实施方案中,IL-2Rg的非人类胞外结构域用人类IL-2Rg的胞外结构域替换,分子的其余部分来自非人类。

[0383] 在另一实施方案中,提供基因修饰的非人类动物,其包含人源化基因组基因座。这样的基因修饰的非人类动物包含(a)插入同源或直系同源人类核酸序列;(b)在内源基因组基因座处用同源或直系同源人类核酸序列替换核酸序列;或(c)其组合,其中所述人源化基因组基因座能够经由种系传递。

[0384] 还提供包含本文提供且如上所述的各种人源化基因组基因座中的任一种的包括非人类动物的基因修饰的动物。

#### [0385] 4. 目标多核苷酸

[0386] 任何目标多核苷酸都可包含在各种插入核酸中且由此在靶基因座处整合。本文公开的方法提供整合到所述靶基因组基因座中的至少1、2、3、4、5、6个或更多个目标多核苷酸。

[0387] 在所述插入核酸内的目标多核苷酸当在所述靶基因组基因座处整合时可将一种或多种基因修饰引入细胞中。所述基因修饰可包括缺失内源核酸序列和/或将外源或异源或直系同源多核苷酸添加到所述靶基因组基因座中。在一个实施方案中,所述基因修饰包括在所述靶基因组基因座处用目标外源多核苷酸替换内源核酸序列。因此,本文提供的方法允许基因修饰产生,所述基因修饰包括敲除、缺失、插入、替换(“敲入”)、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换或其组合。这样的交换可在将第一、第二、第三、第四、第五、第六、第七或任何随后的插入核酸整合到靶基因组基因座中后发生。

[0388] 在插入核酸内的和/或在靶基因座处整合的目标多核苷酸可包括对其引入的细胞为天然的序列;对其引入的细胞可为异源的目标多核苷酸;对其引入的细胞可为外源的目标多核苷酸;对其引入的细胞可为直系同源的目标多核苷酸;或可来自与其引入的细胞不



同的物种的目标多核苷酸。本文中关于在靶基因座处插入的序列使用的“天然”为序列对于具有所述靶基因座的细胞为天然的或对于靶基因座来源于其(即,来自大鼠)的细胞为天然的序列。在本文中关于序列使用的“异源”包括如下序列,其来源于外来物种或如果来源于相同物种,则在本质上不同或通过故意人类干预由其天然形式在组成和/或基因组基因座方面修饰。在本文中关于序列使用的“外源”为序列来源于外来物种。目标多核苷酸可来自任何目标生物体,所述目标生物体包括但不限于非人类、啮齿动物、非大鼠啮齿动物、仓鼠、小鼠、大鼠、人类、猴、农业哺乳动物或非农业哺乳动物。所述目标多核苷酸可进一步包含编码区、非编码区、调控区或基因组DNA。因此,随后插入核酸中的第一、第二、第三、第四、第五、第六、第七插入核酸和/或任一种可包含这样的序列。

[0389] 在一个实施方案中,在所述插入核酸内和/或在所述靶基因座处整合的目标多核苷酸对于小鼠核酸序列、人类核酸、非人类核酸、真核核酸、非大鼠真核核酸、非人类哺乳动物核酸、哺乳动物核酸、啮齿动物核酸、非大鼠啮齿动物核酸、大鼠核酸、仓鼠核酸、猴核酸、农业哺乳动物核酸或非农业哺乳动物核酸为天然的。在更进一步的实施方案中,在所述靶基因座处整合的目标多核苷酸为基因组核酸的片段。在一个实施方案中,所述基因组核酸为小鼠基因组核酸、人类基因组核酸、非人类核酸、真核核酸、非大鼠真核核酸、非人类哺乳动物核酸、哺乳动物核酸、啮齿动物核酸、非大鼠啮齿动物核酸、大鼠核酸、仓鼠核酸、猴核酸、农业哺乳动物核酸或非农业哺乳动物核酸或其组合。

[0390] 在一个实施方案中,如上所述,所述目标多核苷酸可为约500个核苷酸-约200kb。所述目标多核苷酸可为约500个核苷酸-约5kb、约5kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约30kb、约30kb-约40kb、约40kb-约50kb、约60kb-约70kb、约80kb-约90kb、约90kb-约100kb、约100kb-约110kb、约120kb-约130kb、约130kb-约140kb、约140kb-约150kb、约150kb-约160kb、约160kb-约170kb、约170kb-约180kb、约180kb-约190kb、或约190kb-约200kb、约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb。

[0391] 在插入核酸内的和/或在靶基因组基因座处插入的目标多核苷酸可编码多肽,可编码miRNA,或者其可包含任何目标调控区或非编码区,包括例如调控序列、启动子序列、增强子序列、转录阻遏子-结合序列,或缺失非蛋白编码序列,但不包括缺失蛋白-编码序列。另外,在插入核酸内的和/或在靶基因组基因座处插入的目标多核苷酸可编码在神经系统、骨骼系统、消化系统、循环系统、肌肉系统、呼吸系统、心血管系统、淋巴系统、内分泌系统、泌尿系统、生殖系统或其组合中表达的蛋白质。在一个实施方案中,在插入核酸内的和/或在靶基因组基因座处插入的目标多核苷酸编码在骨髓或源自骨髓的细胞中表达的蛋白质。在一个实施方案中,在插入核酸内的和/或在靶基因座处整合的目标多核苷酸编码在脾细胞中表达的蛋白质。在更进一步的实施方案中,在插入核酸内的和/或在靶基因座处插入的目标多核苷酸编码在B细胞中表达的蛋白质,编码在未成熟的B细胞中表达的蛋白质或编码在成熟B细胞中表达的蛋白质。

[0392] 在插入多核苷酸内的目标多核苷酸可包含ApoE基因座、I12rg基因座、Rag1基因座、Rag2基因座和/或Rag2/Rag1基因座的一部分。这些给定基因座的所述部分在本文中的其它地方论述,照此可采用来自任何目标生物体的各种同源和直系同源区。

[0393] 在一个实施方案中,在插入核酸内的和/或在靶基因座处插入的目标多核苷酸包含编码免疫球蛋白重链可变区氨基酸序列的基因组核酸序列。短语“重链”或“免疫球蛋白重链”在本文中的其它地方描述。

[0394] 在一个实施方案中,在插入核酸内的和/或在靶基因座处整合的目标多核苷酸包含编码人类免疫球蛋白重链可变区氨基酸序列的基因组核酸序列。

[0395] 在一个实施方案中,所述基因组核酸序列包含一个或多个未重排的人类免疫球蛋白重链 $V_H$ 基因片段、一个或多个未重排的人类免疫球蛋白重链D基因片段和一个或多个未重排的人类免疫球蛋白重链 $J_H$ 基因片段,其操作性连接哺乳动物重链恒定区核酸序列。在一个实施方案中,所述基因组核酸序列包含操作性连接哺乳动物重链恒定区核酸序列的重排的人类免疫球蛋白重链可变区核酸序列。在一个实施方案中,所述基因组核酸序列包含一个或多个未重排的人类免疫球蛋白 $V_K$ 或 $V_L$ 基因片段和一个或多个未重排的人类免疫球蛋白 $J_K$ 或 $J_L$ 基因片段,其操作性连接哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列。在一个实施方案中,所述基因组核酸序列包含操作性连接哺乳动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列的重排的人类免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链可变区核酸序列。在一个实施方案中,所述重链恒定区核酸序列包含大鼠恒定区核酸序列、人类恒定区核酸序列或其组合。在一个实施方案中,所述免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链恒定区核酸包含大鼠恒定区核酸序列、人类恒定区核酸序列或其组合。

[0396] 在一个实施方案中,所述免疫球蛋白重链恒定区核酸序列选自或包含CH1、铰链、CH2、CH3和/或其组合。在一个实施方案中,所述重链恒定区核酸序列包含CH1-铰链-CH2-CH3。

[0397] 在一个实施方案中,在插入核酸内的和/或在靶基因座处整合的目标多核苷酸包含编码免疫球蛋白轻链可变区氨基酸序列的基因组核酸序列。短语“轻链”包括来自任何生物体的免疫球蛋白轻链序列且在本文中的其它地方描述。

[0398] 在一个实施方案中,在插入核酸内的和/或在靶基因组基因座处整合的目标多核苷酸包含编码人类免疫球蛋白轻链可变区氨基酸序列的基因组核酸序列。

[0399] 在一个实施方案中,所述基因组核酸序列包含一个或多个未重排的人类免疫球蛋白 $V_K$ 或 $V_L$ 基因片段和一个或多个未重排的人类免疫球蛋白 $J_K$ 或 $J_L$ 基因片段,其操作性连接啮齿动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列。在一个实施方案中,所述基因组核酸序列包含操作性连接啮齿动物免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链轻链恒定区核酸序列的重排的人类免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链可变区核酸序列。在一个实施方案中,所述轻链恒定区核酸序列包含大鼠恒定区核酸序列、人类恒定区核酸序列或其组合。在一个实施方案中,所述免疫球蛋白 $\lambda$ 或 $\kappa$ 轻链恒定区核酸包含大鼠恒定区核酸序列、人类恒定区核酸序列或其组合。

[0400] 在插入核酸内的和/或在靶基因座处整合的目标多核苷酸可编码细胞外蛋白质或用于受体的配体。在特定的实施方案中,编码的配体为细胞因子。目标细胞因子包括选自或包括CCL、CXCL、CX3CL和/或XCL的趋化因子。所述细胞因子还可包括肿瘤坏死因子(TNF)。在其它实施方案中,所述细胞因子为白细胞介素(IL)。在一个实施方案中,所述白细胞介素选自或包括IL-1、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、IL-8、IL-9、IL-10、IL-11、IL-12、IL-13、IL-14、IL-15、IL-16、IL-17、IL-18、IL-19、IL-20、IL-21、IL-22、IL-23、IL-24、IL-25、IL-26、IL-27、IL-28、IL-29、IL-30、IL-31、IL-32、IL-33、IL-34、IL-35和/或IL-36。在一个实施

方案中,所述白细胞介素为IL-2。在特定的实施方案中,在插入核酸内的和/或在靶基因组基因座处整合的这样的目标多核苷酸来自人类,且在更特定的实施方案中,其可包含人类基因组序列。

[0401] 在插入核酸内的和/或在靶基因组基因座处整合的目标多核苷酸可编码脱脂蛋白E(ApoE)。

[0402] 在插入核酸内的和/或在靶基因座处整合的目标多核苷酸可编码胞浆蛋白或膜蛋白。在一个实施方案中,所述膜蛋白为受体,例如细胞因子受体、白细胞介素受体、白细胞介素2受体- $\alpha$ 、干扰白细胞素-2受体 $\beta$ 、干扰白细胞素-2受体 $\gamma$ 或受体酪氨酸激酶。在其它情况下,在插入核酸内的和/或在靶基因座处整合的目标多核苷酸可包含靶基因座的直系同源或同源区。

[0403] 在插入核酸内的和/或在靶基因座处整合的目标多核苷酸可包括编码包括T细胞受体 $\alpha$ 的T细胞受体的至少一个区的多核苷酸。在特定方法中,所述插入核酸中的每一种包含T细胞受体基因座(即,T细胞受体 $\alpha$ 基因座)的基因组区,因此在连续整合完成后,基因组T细胞受体基因座的一部分或整体已经在靶基因座处整合。这样的插入核酸可包含T细胞受体基因座(即,T细胞受体 $\alpha$ 基因座)的可变片段或接合片段中的至少一个或多个。在更进一步的实施方案中,编码T细胞受体的区的目标多核苷酸可来自例如编码突变体蛋白的真核生物、非大鼠真核生物、哺乳动物、非人类哺乳动物、啮齿动物、非大鼠啮齿动物、小鼠、大鼠、人类、猴、仓鼠、农业哺乳动物或家养哺乳动物多核苷酸。

[0404] 在其它实施方案中,在靶基因座处整合的目标多核苷酸编码核蛋白。在一个实施方案中,所述核蛋白为核受体。在特定的实施方案中,在插入核酸内的和/或在靶基因座处整合的这样的目标多核苷酸来自人类,且在更特定的实施方案中,其可包含人类基因组序列。

[0405] 在插入核酸内的和/或在靶基因组基因座处整合的目标多核苷酸可包含在编码序列中的基因修饰。这样的基因修饰包括但不限于编码序列的缺失突变或两个编码序列的融合。

[0406] 在插入核酸内的和/或在靶基因座处整合的目标多核苷酸可包含编码包括例如人类突变体蛋白的突变体蛋白的多核苷酸。在一个实施方案中,所述突变体蛋白的特征在于改变的结合特性、改变的定位、改变的表达和/或改变的表达模式。在一个实施方案中,在插入核酸内的和/或在靶基因座处整合的目标多核苷酸包含至少一种疾病等位基因,所述疾病等位基因包括例如神经病等位基因、心血管疾病等位基因、肾病等位基因、肌肉疾病等位基因、血液病等位基因、致癌基因的等位基因或免疫系统疾病等位基因。在这样的情况下,所述疾病等位基因可为显性等位基因或所述疾病等位基因为隐性等位基因。此外,所述疾病等位基因可包括单一核苷酸多态性(SNP)等位基因。编码突变体蛋白的目标多核苷酸可来自任何生物体,其包括但不限于编码突变体蛋白的真核生物、非大鼠真核生物、哺乳动物、非人类哺乳动物、啮齿动物、非大鼠啮齿动物、小鼠、大鼠、人类、仓鼠、猴、农业哺乳动物或家养哺乳动物多核苷酸。

[0407] 在一个实施方案中,所述基因修饰生成具有改变的结合特性、改变的定位、改变的表达和/或改变的表达模式的突变体形式的蛋白质。

[0408] 在一个实施方案中,所述基因修饰生成例如大鼠ApoE基因座的ApoE基因座的区的

缺失、添加、替换或其组合,其中在所述ApoE基因座处的基因修饰引起ApoE活性减小。在一个实施方案中,产生ApoE敲除。

[0409] 在一个实施方案中,所述基因修饰生成例如大鼠Rag1基因座的Rag1基因座的区的缺失、添加、替换或其组合,其中在所述Rag1基因座处的基因修饰引起Rag1活性减小。在一个实施方案中,产生Rag1敲除。在一个实施方案中,所述基因修饰生成例如大鼠Rag2基因座的Rag2基因座的区的缺失、添加、替换或其组合,其中在所述Rag2基因座处的基因修饰引起Rag2活性减小。在一个实施方案中,产生Rag2敲除。在一个实施方案中,所述基因修饰生成例如大鼠Rag1/Rag2基因座的Rag1/Rag2基因座的区的缺失、添加、替换或其组合,其中在所述Rag1/Rag2基因座处的基因修饰引起Rag1活性减小和Rag2活性减小。在一个实施方案中,产生Rag1/Rag2敲除。

[0410] 在一个实施方案中,所述基因修饰生成例如大鼠干扰白细胞素-2受体 $\gamma$ 基因座的干扰白细胞素-2受体 $\gamma$ 基因座的区的缺失、添加、替换或其组合,其中在干扰白细胞素-2受体 $\gamma$ 基因座处的基因修饰引起干扰白细胞素-2受体 $\gamma$ 减小。在一个实施方案中,产生干扰白细胞素-2受体 $\gamma$ 敲除。

[0411] 如在本文中的其它地方论述,本文提供的其它实施方案包括例如大鼠ApoE基因座、大鼠干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座中的一种或多种经由用来自另一生物体的ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应直系同源部分替换大鼠ApoE基因座、干扰白细胞素-2受体 $\gamma$ 基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分修饰。

[0412] 在一个实施方案中,产生多种基因修饰。在一个实施方案中,基因修饰生成例如大鼠干扰白细胞素-2受体 $\gamma$ 基因座的干扰白细胞素-2受体 $\gamma$ 基因座的区的缺失、添加、替换或其组合,其中在所述干扰白细胞素-2受体 $\gamma$ 基因座处的基因修饰引起干扰白细胞素-2受体 $\gamma$ 减小,且第二基因修饰生成大鼠Rag2基因座的区的缺失、添加、替换或其组合,其中在Rag2基因座处的基因修饰引起Rag2活性减小。在一个实施方案中,产生干扰白细胞素-2受体 $\gamma$ /Rag2敲除。这一大鼠具有SCID表型。

[0413] 在一个实施方案中,所述哺乳动物核酸包含编码在神经系统、骨骼系统、消化系统、循环系统、肌肉系统、呼吸系统、心血管系统、淋巴系统、内分泌系统、泌尿系统、生殖系统或其组合中表达的蛋白质的基因组基因座。在一个实施方案中,所述哺乳动物核酸包含编码在骨髓或源自骨髓的细胞中表达的蛋白质的基因组基因座。在一个实施方案中,所述核酸包含编码在脾细胞中表达的蛋白质的基因组基因座。在一个实施方案中,所述基因组基因座包含小鼠基因组DNA序列、大鼠基因组DNA序列、人类基因组DNA序列或其组合。在一个实施方案中,所述基因组基因座包含以任何顺序的大鼠和人类基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的小鼠和人类基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的小鼠和大鼠基因组DNA序列。在一个实施方案中,所述基因组基因座包含以任何顺序的大鼠、小鼠和人类基因组DNA序列。

[0414] 在一个实施方案中,所述插入核酸包含在基因的编码序列中的基因修饰。在一个实施方案中,所述基因修饰包括在编码序列中的缺失突变。在一个实施方案中,所述基因修

饰包括两种内源编码序列的融合。

[0415] 在一个实施方案中,所述基因修饰包括非蛋白编码序列的缺失,但不包括蛋白编码序列的缺失。在一个实施方案中,所述非蛋白编码序列的缺失包括调控元件的缺失。在一个实施方案中,所述基因修饰包括启动子的添加。在一个实施方案中,所述基因修饰包括启动子或调控元件的替换。在一个实施方案中,所述调控元件为增强子。在一个实施方案中,所述调控元件为转录阻遏子-结合元件。

[0416] 在一个实施方案中,所述基因修饰包括放置编码突变体人类蛋白的人类核酸序列。在一个实施方案中,所述基因修饰包含人类基因的至少一种人类疾病等位基因。在一个实施方案中,所述人类疾病为神经病。在一个实施方案中,所述人类疾病为心血管疾病。在一个实施方案中,所述人类疾病为肾病。在一个实施方案中,所述人类疾病为肌肉疾病。在一个实施方案中,所述人类疾病为血液病。在一个实施方案中,所述人类疾病为癌症。在一个实施方案中,所述人类疾病为免疫系统疾病。在一个实施方案中,所述人类疾病等位基因为显性等位基因。在一个实施方案中,所述人类疾病等位基因为隐性等位基因。在一个实施方案中,所述人类疾病等位基因包括单一核苷酸多态性 (SNP) 等位基因。

[0417] 在插入核酸内的和/或在靶基因座处整合的目标多核苷酸还可包含包括例如启动子序列、增强子序列或转录阻遏子-结合序列的调控序列。在特定的实施方案中,在插入核酸内的和/或在靶基因组基因座处整合的目标多核苷酸包括具有非蛋白编码序列缺失、但不包含蛋白编码序列的缺失的多核苷酸。在一个实施方案中,所述非蛋白编码序列的缺失包括调控序列的缺失。在另一实施方案中,所述调控元件的缺失包括启动子序列的缺失。在一个实施方案中,所述调控元件的缺失包括增强子序列的缺失。这一目标多核苷酸可来自任何生物体,其包括但不限于编码突变体蛋白的真核生物、非大鼠真核生物、哺乳动物、非人类哺乳动物、啮齿动物、非大鼠啮齿动物、小鼠、大鼠、人类、猴、农业哺乳动物或家养哺乳动物多核苷酸。

[0418] 5. 引入序列并产生转基因动物的方法

[0419] 如上概述,本文提供允许将一种或多种目标多核苷酸靶向整合到靶基因座中的方法和组合物。这样的系统采用多种组分,且为了便于提及,在本文中术语“靶向整合系统”在属类上包含整合事件中需要的所有组分(即,在非限制性实例中,各种核酸酶试剂、识别位点、插入DNA多核苷酸、靶向载体、靶基因组基因座和/或目标多核苷酸)。

[0420] 本文提供的方法包括向细胞中引入包含靶向基因组整合系统的各种组分的一种或多种多核苷酸或多肽构建体。“引入”以使得序列进入细胞内部的方式而将序列呈递到细胞中(多肽或多核苷酸)。本文提供的方法并不取决于将靶向基因组整合系统的任何组分引入细胞中的特定方法,而只是使多核苷酸进入至少一个细胞的内部。将多核细胞引入各种细胞类型中的方法在本领域中已知且包括但不限于稳定的转染方法、瞬时的转染方法和病毒介导的方法。

[0421] 来自任何生物体的任何细胞都可用于本文提供的方法中。在特定的实施方案中,所述细胞来自真核生物、非大鼠真核生物、哺乳动物、非人类哺乳动物、人类、啮齿动物、非大鼠啮齿动物、大鼠、小鼠或仓鼠。在特定的实施方案中,所述细胞为真核细胞、非大鼠真核细胞、多潜能细胞、非多潜能细胞、非人类多潜能细胞、非人类哺乳动物细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、人类诱导的多潜能细胞(iPS)细

胞、哺乳动物细胞、人类细胞、成纤维细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细胞或CHO细胞。

[0422] 在一些实施方案中,在所述方法和组合物中采用的细胞具有稳定地并入其基因座中的DNA构建体。“稳定地并入”或“稳定地引入”是指将多核苷酸引入细胞中,因此核苷酸序列整合到细胞的基因组中且能够由其子代遗传。可使用用于稳定地并入DNA构建体或靶向基因组整合系统的各种组分的任何方案。

[0423] 可改变转染方案以及将多肽或多核苷酸序列引入细胞中的方案。非限制性转染方法包括基于化学的转染方法,其包括使用脂质体;纳米粒子;磷酸钙(Graham等人,(1973).*Virology* 52(2):456-67;Bacchetti等人,(1977)*Proc Natl Acad Sci USA* 74(4):1590-4;和Kriegler,M(1991).*Transfer and Expression:A Laboratory Manual*.New York:W.H.Freeman and Company.第96-97页);树状体;或阳离子聚合物,例如DEAE-葡聚糖或聚乙烯亚胺。非化学方法包括电穿孔;声孔效应;和光学转染。基于颗粒的转染包括使用基因枪、磁铁辅助转染(Bertram,J.(2006)*Current Pharmaceutical Biotechnology* 7,277-28)。病毒方法也可用于转染。

[0424] 在一个实施方案中,向细胞中引入一种或多种多核苷酸通过电穿孔、胞浆内注射、病毒感染、腺病毒、慢病毒、反转录病毒、转染、脂质介导的转染介导,或者经由Nucleofection™介导。

[0425] 在一个实施方案中,向细胞中引入一种或多种多核苷酸进一步包括:引入包含操作性连接启动子的目标核酸序列的表达构建体。在一个实施方案中,所述启动子为组成性活性启动子。在一个实施方案中,所述启动子为可诱导的启动子。在一个实施方案中,所述启动子在例如胚胎干细胞的干细胞中具有活性。

[0426] 在一个实施方案中,所述表达构建体与所述LTVEC一起引入。在一个实施方案中,所述表达构建体与所述LTVEC单独地经一段时间引入。

[0427] 在一个实施方案中,向所述细胞中引入所述一种或多种多核苷酸可经一段时间执行多次。在一个实施方案中,向所述细胞中引入所述一种或多种多核苷酸可经一段时间执行至少两次,经一段时间执行至少三次,经一段时间执行至少四次,经一段时间执行至少五次,经一段时间执行至少六次,经一段时间执行至少七次,经一段时间执行至少八次,经一段时间执行至少九次,经一段时间执行至少十次,经一段时间执行至少十一次,经一段时间执行至少十二次,经一段时间执行至少十三次,经一段时间执行至少十四次,经一段时间执行至少十五次,经一段时间执行至少十六次,经一段时间执行至少十七次,经一段时间执行至少十八次,经一段时间执行至少十九次或经一段时间执行至少二十次。

[0428] 在一个实施方案中,所述核酸酶试剂与靶向载体或大靶向载体(LTVEC)同时引入细胞中。可选地,所述核酸酶试剂经一段时间与靶向载体或LTVEC单独地引入。在一个实施方案中,所述核酸酶试剂在引入靶向载体或LTVEC之前引入,而在其它实施方案中,所述核酸酶试剂在引入靶向载体或LTVEC之后引入。

[0429] 在一个实施方案中,筛选步骤包括用于测定亲本染色体的等位基因(MOA)的修饰的定量测定。在一个实施方案中,所述定量测定经由定量PCR进行。在一个实施方案中,所述定量PCR为实时PCR(qPCR)。在一个实施方案中,所述实时PCR包括识别靶基因座的第一引物集和识别非靶向参考基因座的第二引物集。在一个实施方案中,所述引物集包括识别扩增

序列的荧光探针。在一个实施方案中,所述定量测定经由荧光介导的原位杂化(FISH)进行。在一个实施方案中,所述定量测定经由对比基因组杂化进行。在一个实施方案中,所述定量测定经由等温DNA扩增进行。在一个实施方案中,所述定量测定经由等温DNA扩增进行。在一个实施方案中,所述定量测定经由对一种或多种固定的探针定量杂化进行。在一个实施方案中,所述定量测定经由Invader **Probes**<sup>®</sup>进行。在一个实施方案中,所述定量测定经由MMP **测定**<sup>®</sup>进行。在一个实施方案中,所述定量测定经由**TaqMan**<sup>®</sup>分子信标进行。在一个实施方案中,所述定量测定经由Eclipse<sup>™</sup>探针技术进行。(参见,例如US2005/0144655,其以引用的方式并入本文中)。

[0430] 进一步提供用于制造人源化非人类动物的方法,其包括:(a)用包含插入核酸的靶向载体修饰多潜能细胞的基因组以形成供体细胞,所述插入核酸包含人类核酸序列;(b)将所述供体细胞引入宿主胚胎中;和(c)在代孕母体中孕育所述宿主胚胎,其中所述代孕母体生成包含所述人类核酸序列的子代。在一个实施方案中,所述供体细胞引入处于胚泡段或处于前桑椹胚阶段(即,4-细胞阶段或8-细胞阶段)的宿主胚胎中。此外,步骤(a)也可用至少5kb长的大靶向载体(LTVEC)和/或人类核酸序列执行。在更进一步的实施方案中,所述基因修饰能够经由种系传递。

[0431] 基因修饰的非人类动物可采用本文公开的各种方法产生。这样的方法包括(1)采用本文公开的方法在多潜能细胞的靶基因座处整合一种或多种目标多核苷酸以产生在靶向基因组基因座中包含插入核酸的基因修饰的多潜能细胞;(2)选择在所述靶基因组基因座处具有所述一种或多种目标多核苷酸的基因修饰的多潜能细胞;(3)将所述基因修饰的多潜能细胞引入宿主胚胎中;和(4)将包含所述基因修饰的多潜能细胞的宿主胚胎植入代孕母体中。产生来自基因修饰的多潜能细胞的子代。在一个实施方案中,所述供体细胞引入处于胚泡段或处于前桑椹胚阶段(即,4-细胞阶段或8-细胞阶段)的宿主胚胎中。产生能够经由种系传递基因修饰的子代。所述多潜能细胞可为如在本文中的其它地方论述的ES细胞。

[0432] 也可使用核移植技术以产生基因修饰的非人类动物。简要地讲,用于核移植的方法包括以下步骤:(1)使卵细胞去核;(2)分离供体细胞或核以与去核的卵细胞合并;(3)将所述细胞或核插入所述去核的卵细胞中以形成重构细胞;(4)将所述重构细胞植入动物的子宫中以形成胚胎;和(5)允许所述胚胎发育。在这样的方法中,卵细胞通常从死动物中找回,不过它们也可自活动物的输卵管和/或卵巢中分离。卵细胞可在去核之前在本领域的普通技术人员已知的多种培养基中成长。卵细胞的去核可以本领域的普通技术人员众所周知的许多方式执行。将供体细胞或核插入去核的卵细胞中以形成重构的细胞通常在融合之前在透明带下微量注射供体细胞进行。融合可通过跨接触/融合平面应用DC电脉冲(电融合)、通过将细胞暴露于促进融合的化学品如聚乙二醇或通过例如仙台病毒(Sendai virus)的灭活病毒诱导。重构细胞通常在核供体和受体卵细胞融合之前、期间和/或之后通过电力和/或非电力方法激活。激活方法包括电脉冲、化学诱导冲击、精液渗透、增加二价阳离子在卵细胞中的水平和降低细胞蛋白在卵细胞中的磷酸化(如通过激酶抑制剂)。激活的重构细胞或胚胎通常在本领域的普通技术人员众所周知的培养基中培养且随后移植到动物的子宫中。参见,例如US20080092249、WO/1999/005266A2、US20040177390、WO/2008/017234A1和美国专利7,612,250号,其各自以引用的方式并入本文中。

[0433] 一方面,提供一种用于制造基因修饰的非人类动物的方法,其包括采用核酸内切酶介导的基因靶向修饰在多潜能细胞中的目标基因组基因座以在目标基因组基因座处引入修饰以形成修饰的多潜能细胞,在足以维持多潜能性的条件下维持所述修饰的多潜能细胞,采用所述修饰的多潜能细胞作为在宿主胚胎中的供体细胞,和在代孕母体中孕育包含所述修饰的多潜能细胞的宿主胚胎,其中所述宿主胚胎通过所述代孕母体孕育且生出基因修饰的子代。

[0434] 在一个实施方案中,所述靶序列位于内含子中。在一个实施方案中,所述靶序列位于外显子中。在一个实施方案中,所述靶序列位于启动子中。在一个实施方案中,所述靶序列位于启动子调控区中。在一个实施方案中,所述靶序列位于增强子区中。

[0435] 在一个实施方案中,引入步骤使用识别不同靶序列的多种核酸内切酶经一段时间执行多次。在一个实施方案中,步骤使用识别不同靶序列的多种核酸内切酶经一段时间执行至少两次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少三次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少四次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少五次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少六次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少七次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少八次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少九次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十一次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十二次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十三次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十四次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十五次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十六次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十七次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十八次,使用识别不同靶序列的多种核酸内切酶经一段时间执行至少十九次或使用识别不同靶序列的多种核酸内切酶经一段时间执行至少二十次。

[0436] 在一个实施方案中,引入步骤通过电穿孔、胞浆内注射、腺病毒、慢病毒、反转录病毒、转染、脂质介导的转染介导,或者经由Nucleofection™介导。

[0437] 在一个实施方案中,所述方法进一步包括将外源核酸引入所述基因修饰的多潜能细胞中。在一个实施方案中,所述外源核酸为转基因。在一个实施方案中,所述外源核酸引入内源基因座中。在一个实施方案中,所述外源核酸异位引入(例如,在与其内源基因座不同的基因座处)。

[0438] 一方面,提供一种用于制造基因修饰的非人类动物的方法,其包括采用RNA向导基因组工程化修饰在多潜能细胞中的目标基因组基因座以在目标基因组基因座处引入修饰以形成修饰的多潜能细胞,在足以维持多潜能性的条件下维持所述修饰的多潜能细胞,采用所述修饰的多潜能细胞作为在宿主胚胎中的供体细胞,和在代孕母体中孕育包含所述修饰的多潜能细胞的宿主胚胎,其中所述宿主胚胎通过所述代孕母体孕育且生出基因修饰的子代。

[0439] 在一个实施方案中,所述方法具有约2%-约80%的靶向率。

[0440] 在一个实施方案中,所述方法包括共同引入包含不同基因组靶序列的多种第二表



达构建体以便多重编辑不同的基因组基因座。在一个实施方案中,所述方法包括经一段时间引入包含不同基因组靶序列的多种第二表达构建体以便多重编辑不同的基因组基因座。

[0441] 在一个实施方案中,引入步骤经一段时间执行多次。在一个实施方案中,引入步骤(b)经一段时间执行至少两次,经一段时间执行至少三次,经一段时间执行至少四次,经一段时间执行至少五次,经一段时间执行至少六次,经一段时间执行至少七次,经一段时间执行至少八次,经一段时间执行至少九次,经一段时间执行至少十次,经一段时间执行至少十一次,经一段时间执行至少十二次,经一段时间执行至少十三次,经一段时间执行至少十四次,经一段时间执行至少十五次,经一段时间执行至少十六次,经一段时间执行至少十七次,经一段时间执行至少十八次,经一段时间执行至少十九次、经一段时间执行至少二十次。

[0442] 在一个实施方案中,所述第一表达构建体和第二表达构建体自相同的质粒表达。

[0443] 在一个实施方案中,引入步骤通过电穿孔、胞浆内注射、腺病毒、慢病毒、反转录病毒、转染、脂质介导的转染介导,或者经由Nucleofection™介导。

[0444] 在一个实施方案中,所述方法进一步包括将外源核酸引入包含突变等位基因的多潜能细胞中。

[0445] 在一个实施方案中,所述外源核酸为转基因。在一个实施方案中,所述外源核酸引入内源基因座中。在一个实施方案中,所述外源核酸异位放置(例如,在与其内源基因座不同的基因座处)。

[0446] 在一个实施方案中,所述方法进一步包括将外源核酸引入所述基因修饰的多潜能细胞中。在一个实施方案中,所述外源核酸为转基因。在一个实施方案中,所述外源核酸引入内源基因座中。在一个实施方案中,所述外源核酸异位引入(例如,在与其内源基因座不同的基因座处)。

[0447] 一方面,提供一种用于制造人源化非人类动物的方法,其包括用包含至少5kb的人类序列的插入物的LTVEC修饰多潜能细胞的基因组,和采用所述多潜能细胞作为供体细胞,将所述供体细胞引入宿主胚胎中,和在代孕母体中孕育所述宿主胚胎,其中所述代孕母体生出包含人源化的子代。

[0448] 提供用于制造在其种系中包含一种或多种如本文所述的基因修饰的基因修饰的非人类动物的其它方法,其包括:(a)采用本文所述的各种方法修饰在原核细胞中包含的靶基因座;(b)选择在所述靶基因座处包括基因修饰的修饰的原核细胞;(c)自所述修饰的原核细胞的基因组中分离基因修饰的靶向载体;(d)将所述基因修饰的靶向载体引入多潜能细胞中以产生在所述靶基因组基因座处包含插入核酸的基因修饰的多潜能细胞;(e)选择所述基因修饰的多潜能细胞;(f)将所述基因修饰的多潜能细胞引入处于前桑椹胚阶段的宿主胚胎中;和(g)将包含所述基因修饰的多潜能细胞的宿主胚胎植入代孕母体中,以产生来源于所述基因修饰的多潜能细胞的F0代。在这样的方法中,所述靶向载体可包括大靶向载体。所述多潜能细胞可为ES细胞。在其它方法中,分离步骤(c)还包括(c1)使所述基因修饰的靶向载体(即,基因修饰的LTVEC)线性化。在更进一步的实施方案中,引入步骤(d)进一步包括(d1)将如本文所述的核酸酶试剂引入多潜能细胞中。在一个实施方案中,选择步骤(b)和/或(e)通过将如本文所述的可选择的试剂应用到所述原核细胞或所述多潜能细胞来进行。在一个实施方案中,选择步骤(b)和/或(e)经由如本文所述的等位基因修饰(MOA)测定

进行。

[0449] 提供经由在原核细胞中的细菌同源重组 (BHR) 修饰哺乳动物细胞的靶基因组基因座的另外方法且其包括: (a) 提供包含包含核酸的靶基因座的原核细胞, (b) 向所述原核细胞中引入包含侧接有5'同源臂和3'同源臂的插入核酸的靶向载体, 其中所述插入核酸包含哺乳动物区 (包括例如来自人类的DNA插入物), 和 (c) 选择在所述靶基因座处包含插入核酸的靶向原核细胞, 其中所述原核细胞能够表达介导所述BHR的重组酶。步骤 (a1) 可包括提供包含包含核酸的靶基因座的原核细胞, 所述核酸包含包含用于第一核酸酶试剂的第一识别位点的第一多核苷酸, 且步骤 (b1) 可进一步包括在所述原核细胞中表达在所述第一识别位点处或其附近生成切口或双链断裂的核酸酶试剂。步骤 (a) - (c) 可如本文公开连续地重复以允许在原核细胞中的靶基因座处引入多个插入核酸。靶基因组基因座一旦用原核细胞“建造”, 则包含修饰的靶基因座的靶向载体可自原核细胞分离并引入在多潜能细胞内的靶基因组基因座中。包含修饰的基因组基因座的多潜能细胞 (即, ES细胞) 因此可制造成基因修饰的非人类动物。

[0450] 在一些实施方案中, 本文所述的靶基因组基因座的各种基因修饰可通过使用来源于细菌人造染色体 (BAC) DNA的LTVEC使用 **VELOCIGENE®** 基因工程技术在细菌细胞中的一系列同源重组反应 (BHR) 进行 (参见, 例如美国专利号6,586,251和Valenzuela, D.M. 等人, (2003), High-throughput engineering of the mouse genome coupled with high-resolution expression analysis, Nature Biotechnology 21 (6):652-659, 其整体以引用的方式并入本文中)。

[0451] 在一些实施方案中, 包含如本文所述的各种基因修饰的靶向ES细胞作为插入物ES细胞使用并经由 **VELOCIMOUSE®** 方法引入来自相应生物体的前桑椹胚阶段胚胎如8-细胞阶段小鼠胚胎中 (参见, 例如US 7,576,259、US 7,659,442、US 7,294,754和US 2008-0078000 A1, 其所有的整体内容都以引入的方式并入本文中)。孵育包含基因修饰的ES细胞的胚胎直至胚泡阶段, 且随后将其植入代孕母体中以产生F0。载有基因修饰的基因组基因座的动物可经由如本文所述的等位基因修饰 (MOA) 测定鉴定。来源于基因修饰的ES细胞的所得F0代非人类动物与野生型非人类动物交叉以得到F1代后代。在与特异性引物和/或探针基因分型之后, 对于基因修饰的基因组基因座杂合的F1非人类动物彼此交叉以生成对于基因修饰的基因组基因座纯合的动物。可选地, 可使各自具有基因修饰的F0雌性非人类动物和F0雄性非人类动物交叉以得到对于基因修饰纯合的F1非人类动物。

[0452] 一方面, 提供例如基因修饰的大鼠基因组, 其包含用来自另一生物体的同源或直系同源的核酸序列靶向修饰内源核酸序列。

[0453] 在一个实施方案中, 所述同源或直系同源的核酸序列具有约5kb-约200kb的长度。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约5kb-约10kb。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约10kb-约20kb。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约20kb-约30kb。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约30kb-约40kb。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约40kb-约50kb。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约50kb-约60kb。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约60kb-约70kb。在一个实施方案中, 所述同源或直系同源非大鼠核酸序列为约70kb-约80kb。在一个实施方案

中,所述同源或直系同源非大鼠核酸序列为约80kb-约90kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约90kb-约100kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约100kb-约110kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约110kb-约120kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约120kb-约130kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约140kb-约150kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约150kb-约160kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约160kb-约170kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约170kb-约180kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约180kb-约190kb。在一个实施方案中,所述同源或直系同源非大鼠核酸序列为约190kb-约200kb。可在插入核酸中采用的各种目标多核苷酸在本文中的其它地方描述。

[0454] 提供靶向基因组修饰非人类动物的另外方法。这样的方法可包括(a)根据本文提供的各种方法中的任一种修饰在非人类多潜能细胞中的目标基因组基因座以便修饰目标基因组基因座,由此生成包含靶向基因组修饰的基因修饰的非人类多潜能细胞;(b)将步骤(a)的修饰的非人类多潜能细胞引入非人类宿主胚胎中;和(c)在代孕母体中孕育包含所述修饰的多潜能细胞的非人类宿主胚胎,其中所述代孕母体生成包含所述靶向基因组修饰的F0子代,且其中所述靶向基因组修饰能够经由种系传递。

[0455] 在一些实施方案中,所述靶向基因组修饰同时包括在目标基因组基因座处缺失内源核酸序列和在目标基因组基因座处插入外源核酸(即,在单一步骤中缺失和插入)。在一些实施方案中,所述靶向基因组修饰包括双等位基因基因修饰。所述双等位基因基因修饰可包括在两种同源染色体(即,一对第一同源染色体和第二同源染色体)中的目标基因组基因座处缺失内源核酸序列和插入外源核酸。

[0456] 在其它实施方案中,所述靶向基因组修饰产生在目标基因组基因座处复合杂合的修饰的多潜能细胞。在其它实施方案中,所述靶向基因组修饰产生在目标基因组基因座处半合的修饰的多潜能细胞。在一些实施方案中,在一种染色体中的目标基因组基因座处的靶向基因修饰包括内源核酸序列的缺失和外源核酸的插入。例如,所述靶向基因修饰可包括:(1)在两种同源染色体中的目标基因组基因座处缺失内源核酸序列;和(2)外源核酸插入在第一染色体中的目标基因组基因座中和破坏在第二染色体中的目标基因组基因座。所述第一染色体可为这两种同源染色体中的第一种,且所述第二染色体可为这两种同源染色体中的第二种。

#### [0457] 6. 细胞

[0458] 本文所述的各种方法和组合物采用在细胞中的基因组基因座靶向系统。在一个实施方案中,细胞是多潜能细胞。在一个实施方案中,所述细胞为非多潜能细胞。在一个实施方案中,所述多潜能细胞为非人类多潜能细胞。在一个实施方案中,所述非人类多潜能细胞为哺乳动物多潜能细胞。在一个实施方案中,所述多潜能细胞为人类诱导的多潜能干(iPS)细胞。

[0459] 在其它实施方案中,所述细胞为真核细胞、非大鼠真核细胞、人类多潜能细胞、人类ES细胞、人类成人干细胞、发育受限的人类祖细胞、非人类哺乳动物细胞、哺乳动物细胞、人类细胞、成纤维细胞、啮齿动物细胞、非大鼠啮齿动物细胞、大鼠细胞、小鼠细胞、仓鼠细

胞或CHO细胞。

[0460] 在一个实施方案中,真核细胞为原代细胞。原代细胞包括从生物体、器官或组织中直接分离的细胞或细胞的培养物。原代细胞包括既不转化也不永生的细胞。它们包括从生物体、器官或组织中获得的所有细胞,其先前并未在组织培养物中传代,或者先前已经在组织培养物中传代,但不能在组织培养物中无限期地传代。这样的细胞可通过常规方法分离且包括例如造血细胞、内皮细胞、上皮细胞、成纤维细胞、间充质细胞、角质细胞、黑素细胞、单核细胞(monocyte)、单核细胞(mononuclear cell)、脂肪细胞、前成脂肪细胞、神经元、胶质细胞、肝细胞、骨骼成肌细胞和平滑肌细胞。在一些实施方案中,原代细胞来源于结缔组织、肌肉组织、神经系统组织或上皮组织。

[0461] 在另一实施方案中,真核细胞为永生化细胞。永生化细胞包括来自多细胞生物体的细胞,其由于突变或改变通常将不会无限期地繁殖,具有逃避的正常细胞衰老且代之以可保持经历分裂。这样的突变或改变可自然地发生或有意诱导地发生。永生细胞的实例包括中国仓鼠卵巢(CHO)细胞、人类胚肾细胞(例如,HEK 293细胞)和小鼠胚胎成纤维细胞(例如,3T3细胞)。许多类型的永生化细胞在本领域中众所周知。

[0462] 在一些实施方案中,永生化细胞来源于癌症细胞。在另一实施方案中,原代或永生化细胞为通常用于培养或用于表达重组基因或蛋白质的细胞。

[0463] 在其它实施方案中,所述多潜能细胞能够在其基因组的至少一种靶向基因修饰之后维持其多潜能性且能够传递所述靶向修饰到F1代的种系。

[0464] 在一个实施方案中,所述多潜能细胞为处于单细胞阶段的非人类受精卵。在一个实施方案中,所述非人类受精卵为哺乳动物受精卵。在一个实施方案中,所述哺乳动物受精卵为处于单细胞阶段的啮齿动物受精卵。在一个实施方案中,所述哺乳动物受精卵为处于单细胞阶段的大鼠或小鼠受精卵。

[0465] 在本文公开的方法和组合物中采用的各种细胞还可包括原核细胞,例如包括大肠杆菌的细菌细胞。在特定的实施方案中,所述原核细胞为大肠杆菌的重组感受态株。在一个实施方案中,所述原核细胞包含编码重组酶的核酸,而在其它情况下,所述原核细胞不包含编码所述重组酶的所述核酸,且将编码所述重组酶的所述核酸引入所述原核细胞中。在一个实施方案中,编码所述重组酶的所述核酸包含DNA或mRNA。在一些实施方案中,编码所述重组酶的所述核酸为pABG。在一个实施方案中,所述重组酶在可诱导的启动子的控制下表达。在一个实施方案中,所述重组酶的表达通过阿拉伯糖控制。

[0466] A. 制造并维持人类诱导的多潜能干细胞的低渗透压度培养基

[0467] 提供细胞培养基以便在本发明的方法和组合物中使用。在一个实施方案中,所述培养基适合生成人类iPS细胞群。在另一实施方案中,所述培养基适合在培养物中维持人类iPS细胞。在一些实施方案中,所述人类iPS细胞为天然的或天然外观的。

[0468] 本文提供的培养基包含至少一种基础培养基、补充物、白血病抑制因子(LIF)多肽、肝糖合成酶激酶3(GSK3)抑制剂和MEK抑制剂。

[0469] 本发明的培养基为低渗透压度培养基。在一个实施例中,所述渗透压度为约175-280mOsm/kg。在另外的实施例中,所述培养基的渗透压度为约180-270mOsm/kg、约200-250mOsm/kg、约220-240mOsm/kg或约225-235mOsm。在一个特定的实施方案中,所述培养基的渗透压度为约233mOsm/kg。

[0470] 为本发明提供的基础培养基为向其中加入补充物的低渗透压度基础培养基。本发明的基础培养基与通常用以在培养物中维持人类iPS细胞的基础培养基不同,其包括达尔伯克改良伊格尔培养基(DMEM),以各种形式(例如,Invitrogen DMEM,目录号1 1971-025)和作为KO-DMEM<sup>TM</sup>购得的低盐DMEM(Invitrogen目录号10829-018)。

[0471] 本文提供的基础培养基为低渗透压度培养基,但其表现出不限于低渗透压度的特性。例如,在表A中示出的DMEM制剂可通过改变如本文提供的氯化钠和/或碳酸氢钠浓度而使其适合本发明的目的,将其产生与标准DMEM基础培养基或在表A中示出的低盐DMEM基础培养基(KO-DMEM)相比较不同的渗透压度。

[0472] 表A:DMEM基础培养基制剂。

	<b>组分</b>	<b>Mg/L</b>	<b>mM</b>
[0473]	甘氨酸	30	0.4
	L-精氨酸·HCl	84	0.398
	L-胱氨酸·2HCl	63	0.201

[0474]

L-谷氨酰胺	584	4
L-组氨酸·HCl·H <sub>2</sub> O	42	0.2
L-异亮氨酸	105	0.802
L-亮氨酸	105	0.802
L-赖氨酸·HCl	146	0.798
L-蛋氨酸	30	0.201
L-苯基丙氨酸	66	0.4
L-丝氨酸	42	0.4
L-苏氨酸	95	0.798
L-色氨酸	16	0.0784
L-酪氨酸二钠盐二水合物	104	0.398
L-缬氨酸	94	0.803
氯化胆碱	4	0.0286
D-泛酸钙	4	$8.39 \times 10^{-3}$
叶酸	4	$9.07 \times 10^{-3}$
烟酰胺	4	0.0328
吡哆醇·HCl	4	0.0196
核黄素	0.4	$1.06 \times 10^{-3}$
硫胺素·HCl	4	0.0119
i-肌醇	7.2	0.04
氯化钙(CaCl <sub>2</sub> )(无水)	200	1.8
硝酸铁(Fe(NO <sub>3</sub> ) <sub>3</sub> ·9H <sub>2</sub> O)	0.1	$2.48 \times 10^{-4}$
硫酸镁(MgSO <sub>4</sub> )(无水)	97.67	0.814
氯化钾(KCl)	400	5.33
D-葡萄糖(右旋糖)	4500	25
苯酚红	15	0.0399
<b>DMEM 的 NaCl/NaHCO<sub>3</sub> 含量</b>		
碳酸氢钠(NaHCO <sub>3</sub> )	3700	44.05
氯化钠(NaCl)	6400	110.34
<b>低盐 DMEM(KO-DMEM)的 NaCl/NaHCO<sub>3</sub> 含量</b>		
碳酸氢钠(NaHCO <sub>3</sub> )	2200	26
氯化钠(NaCl)	5100	87.7
<b>低渗透压度 DMEM 的 NaCl/NaHCO<sub>3</sub> 含量</b>		

[0475]	碳酸氢钠(NaHCO <sub>3</sub> )	2200	26
	氯化钠(NaCl)	3000	50

[0476] 本发明的基础培养基可包含碱金属和卤素的盐,例如氯化钠(NaCl)。在所述基础培养基中NaCl的例示性浓度包括 $50 \pm 5\text{mM}$ 或约 $3\text{mg/mL}$ 。

[0477] 在另一实施方案中,所述基础培养基表现出一定浓度的碳酸盐。所述碳酸盐可为钠盐。在这一实施例中,所述钠盐可为碳酸氢钠。在一个特定的实施方案中,碳酸氢钠在所述基础培养基中以约 $26 \pm 5\text{mM}$ 或约 $2.2\text{mg/mL}$ 的浓度存在。

[0478] 在又一实施方案中,所述基础培养基为低渗透压度基础培养基。所述基础培养基的渗透压度可在约 $175\text{-}280\text{mOsm/kg}$ 、约 $180\text{-}250\text{mOsm/kg}$ 、约 $190\text{-}225\text{mOsm/kg}$ 或约 $195\text{-}205\text{mOsm/kg}$ 范围内。所述基础培养基的例示性渗透压度可为 $200$ 、 $214$ 、 $216$ 或 $218\text{mOsm/kg}$ 。在一特定的实施例中,所述基础培养基的渗透压度为 $200\text{mOsm/kg}$ 。渗透压度可在细胞在不同浓度的 $\text{CO}_2$ 中培养时测定。在一些实施例中,细胞在 $3\%\text{CO}_2$ 或 $5\%\text{CO}_2$ 下培养。

[0479] 在一个优选的实施方案中,所述基础培养基包含以 $3.0\text{mg/mL}$ 浓度的NaCl、以约 $2.2\text{mg/mL}$ 浓度的碳酸氢钠,且具有 $200\text{mOsm/kg}$ 的渗透压度。

[0480] 用本发明的基础培养基配制的补充物适合制造、维持或富集本文公开的人类iPS细胞群。这样的补充物在本公开内容中指示为“补充物”或“+补充物”。术语“补充物”或短语“+补充物”包含加到在表A中描述的基础培养基的组分中的一种或多种添加元素。例如,补充物可包括而不限于**F-12®**培养基(Gibco)、**N2®**补充物(Gibco;100X溶液)、**NEUROBASAL®**培养基(Gibco)、**B-27®**补充物(Gibco;50X溶液)、L-谷氨酰胺、葡萄糖、2-巯基乙醇、白血病抑制因子(LIF)多肽、肝糖合成酶激酶3抑制剂、MEK抑制剂或其任何组合。

[0481] 在一个特定的实施方案中,所述LIF多肽为人类LIF(hLIF)多肽。在一些实施例中,hLIF多肽在约 $1\text{-}1000$ 单位/毫升、约 $20\text{-}800$ 单位/毫升、约 $50\text{-}500$ 单位/毫升、约 $75\text{-}250$ 单位/毫升或约 $100$ 单位/毫升的浓度下使用。

[0482] 在另一特定的实施方案中,所述GSK3抑制剂包括CHIR99021。在一些实施例中,CHIR99021以约 $0.1\text{-}10\mu\text{M}$ 、约 $1\text{-}5\mu\text{M}$ 、约 $2\text{-}4\mu\text{M}$ 或约 $3\mu\text{M}$ 的浓度使用。

[0483] 在另一特定的实施方案中,所述MEK抑制剂包括PD0325901。在一些实施例中,PD0325901以约 $0.1\text{-}5\mu\text{M}$ 、约 $0.2\text{-}1\mu\text{M}$ 、约 $0.3\text{-}0.7\mu\text{M}$ 或约 $0.5\mu\text{M}$ 的浓度使用。

[0484] 例示性培养基包含约 $24.75\%$  (v/v)的本文所述的低渗透压度基础培养基、约 $24.75\%$  (v/v)的F-12培养基、约 $0.5\%$  (v/v)的N2补充物、约 $49\%$  (v/v)的NEUROBASAL培养基、约 $1\%$  (v/v)的B-27补充物、约 $2\text{mM}$ 的L-谷氨酰胺、约 $0.1\text{mM}$ 的2-巯基乙醇、约 $100$ 单位/毫升的hLIF、约 $3\mu\text{M}$ 的CHIR99021和约 $0.5\mu\text{M}$ 的PD0325901。

[0485] 在另一特定的实施方案中,所述培养基可能包含或可能不含基本纤维母细胞生长因子(bFGF,也称作FGF2或FGF- $\beta$ )。优选本发明的培养基不包含bFGF。

[0486] B. 人类诱导的多潜能干细胞

[0487] 本发明提供用于生成人类iPS细胞群的方法和组合物。进一步提供在培养物中维持人类iPS细胞的方法和组合物。还提供在培养物中生成或维持的人类iPS细胞。

[0488] 术语“多潜能细胞”或“多潜能干细胞”包括具有发育成多于一种类型的分化细胞

的能力的未分化细胞。这样的多潜能细胞可例如为哺乳动物胚胎干 (ES细胞) 细胞或哺乳动物诱导的多潜能干细胞 (iPS细胞)。多潜能细胞的实例包括人类iPS细胞。

[0489] 术语“胚胎干细胞”或“ES细胞”是指来源于胚胎的内细胞团的源自胚胎的全能或多潜能干细胞,其可在合适的条件下在体外培养物中维持。ES细胞能够分化成三种脊椎动物胚层如内胚层、外胚层或中胚层中的任一种的细胞。ES细胞的特征还在于其能够在合适的体外培养条件下无限期地繁殖。参见,例如Thomson等人,(*Science* (1998),第282(5391)卷,第1145-1147页)。

[0490] 术语“诱导的多潜能干细胞”或“iPS细胞”包括可直接来源于分化的成人细胞的多潜能干细胞。人类iPS细胞可通过将特定的重新编程因子集引入非多潜能细胞中而产生,所述重新编程因子可包括例如Oct3/4,Sox族转录因子(例如,Sox1、Sox2、Sox3、Sox15);Myc族转录因子(例如,c-Myc、l-Myc、n-Myc);Krüppel样族(KLF)转录因子(例如,KLF1、KLF2、KLF4、KLF5)和/或相关的转录因子,例如NANOG、LIN28和/或Glis1。人类iPS细胞也可例如通过使用miRNA或谱系特异性分子产生,所述miRNA为模拟转录因子的作用的小分子。人类iPS细胞的特征在于其能够分化成三种脊椎动物胚层如内胚层、外胚层或中胚层的任何细胞。人类iPS细胞的特征还在于其能够在合适的体外培养条件下无限期地繁殖。参见,例如Takahashi和Yamanaka(*Cell* (2006),第126(4)卷,第663-676页)。

[0491] 术语“天然”和“始发态”鉴定人类iPS细胞的不同多潜能性状态。术语“天然外观”鉴定表达表现出天然多潜能细胞的一种或多种特征的多潜能状态的细胞。天然外观人类iPS细胞也可称为“天然样”人类iPS细胞。在一些实施方案中,天然外观人类iPS细胞表现出天然人类iPS细胞的一种或多种形态特征,例如以紧实的圆顶形集落为特征的形态。在一些实施方案中,天然外观的人类iPS细胞表达本文所述的多潜能标记中的一种或多种。在一些实施方案中,天然或天然外观的人类iPS细胞为天然的人类iPS细胞。在其它实施方案中,天然或天然外观的人类iPS细胞为天然外观的iPS细胞。

[0492] 天然或始发态iPS细胞的特征在本领域中描述。参见,例如Nichols和Smith(*Cell Stem Cell* (2009),第4(6)卷,第487-492页)。天然的人类iPS细胞表现出与植入前胚胎的内细胞团的ES细胞类似的多潜能状态。这样的天然细胞对于谱系规范和定型而言不是始发态的。雌性天然iPS细胞以两种活性X染色体为特征。在培养中,天然人类iPS细胞的自更新取决于白血病抑制因子(LIF)及其它抑制剂。培养的天然人类iPS细胞显示以圆形圆顶形集落和缺乏顶面-底侧极性为特征的无性系形态。培养的天然细胞可进一步显示如在本文中的其它地方描述的一种或多种多潜能标记(maker)。在适当的条件下,天然人类iPS细胞在培养物中的倍增时间可为16-24小时。

[0493] 始发态人类iPS细胞表达与植入后外胚层细胞类似的多潜能状态。这样的细胞对于谱系规范和定型而言是始发态的。雌性始发态iPS细胞以一种活性X染色体和一种非活性X染色体为特征。在培养中,始发态人类iPS细胞的自更新取决于成纤维细胞生长因子(FGF)和活化素。培养的始发态人类iPS细胞显示以上皮单层为特征的克隆形态且显示顶面-底侧极性。在适当的条件下,始发态人类iPS细胞在培养物中的倍增时间可为24小时或更久。

[0494] 在一个实施方案中,人类iPS细胞可来源于转化以表达多潜能状态的非多潜能细胞。这样的转化细胞包括例如已经转化以表达诱导多潜能性的重新编程基因的细胞。多潜能状态可包括例如本文所述的多潜能标记中的一种或多种的表达。这样的细胞(例如,人类



包皮成纤维细胞)可通过本领域已知的任何方法转化以表达重新编程基因或任何另外的目标基因。参见,例如Takahashi和Yamanaka (Cell (2006),第126(4)卷,第663-676页)。例如,它们可使用一种或多种质粒慢病毒载体或逆转录病毒载体引入细胞中。在一些情况下,所述载体整合到基因组中且可在重新编程完成之后除去。在特定的实施方案中,所述非多潜能细胞使用包括Oct4、Sox2、Klf4、Myc或其组合的重新编程基因转化。在一些实施例中,所述转化的细胞包括始发态人类iPS细胞。

[0495] 在一些实施方案中,在本文所述的低渗透压度培养基中培养的人类iPS细胞表达天然状态的一种或多种表型、基因表达谱或标记特征。在一个实施例中,所述人类iPS细胞表达其表达指示天然状态的一种或多种多潜能标记。这样的多潜能标记可包括碱性磷酸酶、NANOG、5T4、ABCG2、活化素R1B/ALK-4、活化素R1IB、E-钙粘蛋白、Cb<sub>x</sub>2、CD9、CD30/TNFRSF8、CD117/c-试剂盒、CDX2、CHD1、Cripto、DNMT3B、DPPA2、DPPA4、DPPA5/ESG1、EpCAM/TROP1、ERR $\beta$ /NR3B2、ESGP、F-盒蛋白15/FBXO15、FGF-4、FGF-5、FoxD3、GBX2、GCMF/NR6A1、GDF-3、Gi24/VISTA/B7-H5、整合素 $\alpha$ 6/CD49f、整合素 $\alpha$ 6 $\beta$ 1、整合素 $\alpha$ 6 $\beta$ 4、整合素 $\beta$ 1/CD29、KLF4、KLF5、L1TD1、Lefty、Lefty-1、Lefty-A、LIN-28A、LIN-28B、LIN-41、cMaf、cMyc、Oct-3/4、Oct-4A、糖萼蛋白(Podocalyxin)、Rex-1/ZFP42、Smad2、Smad2/3、SOX2、SSEA-1、SSEA-3、SSEA-4、STAT3、Stella/Dppa3、SUZ12、TBX2、TBX3、TBX5、TERT、TEX19、TEX19.1、THAP11、TRA-1-60(R)、TROP-2、UTF1和/或ZIC3。在一个特定的实施例中,表达的多潜能标记为碱性磷酸酶、NANOG或两者。

[0496] 在另一实施方案中,在本文所述的低渗透压度培养基中培养的人类iPS细胞显示指示天然状态的形态特征。例示性形态以在培养物中具有紧实圆顶形集落的细胞为特征。

[0497] 在另一实施方案中,可将在本文所述的低渗透压度培养基中培养的人类iPS细胞机械地或酶促地分离成单细胞悬浮液、传代和/或传代培养。在一个实施例中,酶促分解可使用胰蛋白酶执行。当在本发明的低渗透压度培养基中培养时,人类iPS细胞可由于向单细胞悬浮液的分解增强而提供更大的转化效率。在用通常用以在培养物中维持人类iPS细胞的其它类型的培养基(例如,mTeSR<sup>TM</sup>培养基或2i培养基)的情况下,人类iPS细胞的分解必须机械地或用例如不如胰蛋白酶苛刻的胶原酶的酶执行。因此,所述细胞不会有效地或完全地分解。相比之下,在用本发明的低渗透压度培养基的情况下,可使用胰蛋白酶来分解所述细胞,且增强的分解产生增加的转化效率。此外,与通常用以在培养物中维持人类iPS细胞的其它类型的培养基(例如,mTeSR<sup>TM</sup>培养基或2i培养基)不同,用本发明的低渗透压度培养基(优选不包含bFGF的低渗透压度培养基)培养的人类iPS细胞的酶促分解可在缺乏通常对于这样的细胞的传代所必需的一种或多种抑制剂的情况下执行。可省略的例示性抑制剂为 $\rho$ 相关的蛋白激酶(ROCK)抑制剂。ROCK抑制剂在传代人类iPS细胞以抑制前细胞凋亡途径的激活时通常是必需的。

[0498] 在另一实施方案中,在本文所述的低渗透压度培养基中培养的传代培养的人类iPS细胞可在酶促分解和传代培养之后维持天然或天然外观状态。在一些实施例中,传代培养的人类iPS细胞可持续显示以紧实的圆顶形集落为特征的形态。传代培养的人类iPS细胞也可持续表达如本文所述一种或潜能标记。

[0499] C. 制造和维持人类诱导的多潜能干细胞群的方法

[0500] 提供在体外培养物中生成人类iPS细胞的方法和组合物。进一步提供在体外培养

物中维持人类iPS细胞的方法和组合物。

[0501] 术语“制造”包括在合适的条件下培养非多潜能细胞,使其转化以表达一种或多种如本文所述的重新编程因子,以诱导在细胞表型、基因表达或两者的变化,使得细胞显示天然或天然外观状态,即,表达天然人类iPS细胞的一种或多种特性。天然或天然外观状态可响应特定的培养条件如在如本文所述的低渗透压度培养基中培养而表达。在一些实施例中,表达天然或天然外观状态的细胞的比例为在培养物中至少约30%、40%、50%、60%、70%、80%、90%、95%和高达100%的细胞。

[0502] 在一个实施方案中,所述方法富集用于天然或天然外观人类iPS细胞的体外培养物。在这一实施方案中,天然或天然外观的人类iPS细胞可在培养物中优先于不表达天然或天然外观状态的细胞繁殖。在另一实施方案中,天然或天然外观的人类iPS细胞可选自酶促分解并传代培养以生成富集的天然或天然外观人类iPS细胞群的培养物。

[0503] 在一个实施方案中,转化以表达多潜能状态的非多潜能细胞在适合诱导表达天然或天然外观状态的本文提供的培养基中离体培养至少1、2、5、7、10、14、21或28天或足以诱导在培养物中表达天然或天然外观状态的任何时间。转化的细胞可在本发明培养基中培养至少1、2、3或4周。有时,将转化的细胞培养1-4周。天然或天然外观状态的表达可通过观察形态特征或多潜能标记的表达、天然或天然外观状态的特征来确定,这在本文中的其它地方描述。

[0504] 在一个实施方案中,转化以表达多潜能状态的非多潜能细胞在本发明低渗透压度培养基中培养,直至它们表达天然或天然外观状态的特征。细胞随后可在本发明培养基中培养以维持天然或天然外观状态。在另一实施方案中,转化以表达多潜能状态的非多潜能细胞在本发明低渗透压度培养基中培养之前首先在高渗透压度培养基中培养。这样的高渗透压度培养基表现出高于本发明低渗透压度培养基的渗透压度且可包含bFGF。某一高渗透压度培养基包含牛血清蛋白、bFGF、转化生长因子 $\beta$  (TGF $\beta$ )、氯化锂、六氢吡啶羧酸和 $\gamma$ -氨基丁酸(GABA)中的一种或多种。高渗透压度培养基的实例包括mTeSR™培养基(Stemcell Technologies)。

[0505] 在一些实施方案中,转化以表达多潜能状态的非多潜能细胞可首先在包含bFGF的高渗透压度培养基中培养,直至它们开始表达天然或天然外观状态的特征,此时将细胞在本发明的低渗透压度培养基中培养。在一个实施例中,细胞可在包含bFGF的高渗透压度培养基中培养至少1、2、5、10、30、60或90天的时间、1、2、4、8或12周的时间或者1天-3个月的时间。在包含bFGF的高渗透压度培养基中培养的例示性时间为2个月。

[0506] 在其它实施方案中,转化以表达多潜能状态的非多潜能细胞可首先在包含bFGF的高渗透压度培养基中培养,直至它们开始显示以三维细胞团块为特征的形态,此时将细胞在本发明的低渗透压度培养基中培养。在这样的实施方案中,可将显示三维细胞团块的细胞选择、分解(例如,用胰蛋白酶)并转移到在本文所述的低渗透压度培养基的新培养物中。

[0507] 术语“维持”包括保存本文所述的人类iPS细胞的特征或表型中的至少一种或多种。这样的特征可包括维持天然细胞的多潜能性、细胞形态性、基因表达谱和/或其它功能特征。术语“维持”还可涵盖细胞的繁殖和/或培养的天然细胞的数目增加。所述术语包括防止细胞转变成始发态或非多潜能状态的培养条件。所述术语进一步包括容许细胞保留多潜能和/或天然的培养条件,而所述细胞可能持续或可能不持续分裂且数目增加。

[0508] 在一个实施方案中,人类iPS细胞在适合维持所述细胞处于天然或天然外观状态的本文提供的培养基中离体培养。在一个特定的实施例,人类iPS细胞可在合适的培养基中培养1、2、5、7、10、14、21或28天的时间、或约2周、约3周、约4周或更久的时间,条件是培养的细胞维持处于天然或天然外观状态。细胞可培养至少1、2、3或4周。有时,将细胞培养1-4周。可将人类iPS细胞维持例如足以使细胞在培养物中的繁殖、细胞的基因修饰和/或细胞的传代培养的任何时间。

[0509] 在另一实施方案中,转化以表达多潜能状态的人类iPS细胞或非多潜能细胞可在适合体外培养的底物或喂养细胞层上培养。在一个特定的实施例,细胞在MATRIGEL™ (BD Biosciences) 上培养。在另一实施例,细胞在新生的人类包皮成纤维细胞 (NuFF) 喂养细胞上培养。在另一实施例,细胞在GELTREX™ (Life Technologies) 上培养。

[0510] 在另一实施方案中,与转化以表达多潜能状态的始发态人类iPS细胞或非多潜能细胞相比较,在本发明的低渗透压度培养基中培养的人类iPS细胞的倍增时间减少。在一个特定的实施例中,本发明的人类iPS细胞的倍增时间为约16-24小时。

[0511] 7. 序列同一性

[0512] 本文提供的方法和组合物采用靶向基因组整合系统的多种不同组分(即,核酸酶试剂、识别位点、插入核酸、目标多核苷酸、靶向载体、选择标记及其它组分)。在本说明书各处认定靶向基因组整合系统的一些组分可具有活性变体和片段。这样的组分包括例如核酸酶试剂(即,工程化核酸酶试剂)、核酸酶试剂识别位点、目标多核苷酸、靶位点和靶向载体的相应同源臂。这些组分中的每一种的生物活性在本文中的其它地方描述。

[0513] 如本文在两种多核苷酸或多肽序列的上下文中使用的“序列同一性”或“同一性”指在特定比较窗上对齐以实现最大对应性时在这两种序列中相同的残基。当在关于蛋白质使用百分比序列同一性时,应该认识到并不相同的残基位置常常因保守氨基酸取代而不同,其中氨基酸残基用具有类似化学性质(例如,电荷或疏水性)的其它氨基酸残基取代且因此不改变分子的功能性质。当序列在保守取代上存在差别时,可向上调节百分比序列同一性以校正取代的保守特性。在这样的保守取代方面存在差别的序列被说成是具有“序列相似性”或“相似性”。本领域的技术人员众所周知进行该调节的方法。通常,这涉及将保守取代作为一部分而不作为完全错配来评分,由此使序列同一性百分数增加。因此,例如,若一个相同的氨基酸的分数给定为1且一个非保守取代的分数给定为零,则一个保守取代的分数给定为0-1。保守取代的分数例如如在程序PC/GENE (Intelligenetics, Mountain View, California, USA) 中所执行来计算。

[0514] 本文使用的“序列同一性百分数”是指通过比较窗上比较两个最佳对齐的序列确定的值,其中,与参考序列相比较,多核苷酸序列在比较窗中的部分可包含添加或缺失(即,间隙),以便最佳对齐这两个序列。该百分数通过以下方式,即确定其中相同的核酸碱基或氨基酸残基在两个序列中出现的位置的数目以产生匹配位置的数目,将匹配位置的数目除以在比较窗口中的位置总数并且将结果乘以100以产生序列同一性百分数。

[0515] 除非另作说明,否则本文提供的序列同一性/相似性值是指使用GAP版本10使用以下参数获得的值:核苷酸序列的%同一性和%相似性使用缺口权重(GAP Weight)50和长度权重3和nws gapdna. cmp评分矩阵;氨基酸序列的%同一性或%相似性使用缺口权重8和长度权重2和BLOSUM62评分矩阵;或其任何等同程序。“等同程序”指任何序列比较程序,其为

所讨论的任何两个序列产生这样的比对,当与由GAP版本10产生的对应比对相比较时,该比对具有相同的核苷酸或氨基酸残基匹配和相同的百分比序列同一性。

[0516] 除非另外定义,否则在本文中使用的所有技术和科学术语均具有与本发明所属领域的一般技术人员通常理解的含义相同的含义。尽管在本发明的实践或试验中也可使用类似或等同于本文所描述的那些方法和材料的任何方法和材料,但是现在描述优选的方法和材料。为了公开并描述公布在引用时所涉及的方法和/或材料,在本文中提到的所有公布均以引用方式并入本文。

[0517] 还应该注意,除非上下文另外明确规定,否则如在本文中和在所附权利要求中所用,单数形式“一个/种”和“所述”包括复数个/种提及物。在本文中使用的所有技术和科学术语都具有相同的含义。

[0518] 提供在本文中论述的公布仅仅是针对它们在本申请的提交日期之前的公开。在本文中绝不应解释为承认所描述的发明无权占先于在先发明的这类公布。此外,所提供的公布日期可不同于实际公布日期,其可能需要单独地证实。

[0519] 所描述的发明可在不脱离其精神和必要特性的情况下以其它具体形式体现,且相应地可参考指示本发明的范围的所附权利要求,而不是上述说明书。

[0520] 非限制性实施方案包括:

[0521] 1.一种用于靶向修饰在多潜能大鼠细胞中的目标基因组基因座的方法,其包括:(a)将包含侧接有5'大鼠同源臂和3'大鼠同源臂的插入核酸的大靶向载体(LTVEC)引入所述多潜能大鼠细胞中,其中所述5'同源臂和所述3'同源臂的总和为至少10kb,但小于150kb;和(b)鉴定在所述目标基因组基因座处包括所述靶向基因修饰的基因修饰的多潜能大鼠细胞,其中所述靶向基因修饰能够经由种系传递。

[0522] 2.实施方案1的方法,其中所述靶向基因修饰为双等位基因的。

[0523] 3.实施方案1或2的方法,其中所述多潜能大鼠细胞为大鼠胚胎干(ES)细胞。

[0524] 4.实施方案1、2或3的方法,其中所述多潜能大鼠细胞来源于DA株或ACI株。

[0525] 5.实施方案1-4中任一项的方法,其中所述多潜能大鼠细胞的特征在于表达包括Dnmt3L、Eras、Err- $\beta$ 、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1的至少一种多潜能标记或其组合。

[0526] 6.实施方案1-4中任一项的方法,其中所述多潜能大鼠细胞的特征在于以下特征中的一种或多种:

[0527] (a)缺乏包括c-Myc、Ecat1和/或Rexo1的一种或多种多潜能标记的表达;(b)缺乏包括Brachyury和/或Bmpr2的中胚层标记的表达;(c)缺乏包括Gata6、Sox17和/或Sox7的一种或多种内胚层标记的表达;或(d)缺乏包括Nestin和/或Pax6的一种或多种神经标记的表达。

[0528] 7.实施方案1-6中任一项的方法,其中所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约30kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-150kb。

[0529] 8.实施方案1-6中任一项的方法,其中所述LTVEC的5'同源臂和3'同源臂的总和为约16kb-约150kb。

[0530] 9.实施方案1-8中任一项的方法,其中所述靶向基因修饰包括:(a)用同源或直系

同源的核酸序列替换内源大鼠核酸序列; (b) 缺失内源大鼠核酸序列; (c) 缺失内源大鼠核酸序列, 其中所述缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb; (d) 外源核酸序列为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb; (e) 包含同源或直系同源的核酸序列的外源核酸序列; (f) 包含人类核酸序列和大鼠核酸序列的嵌合核酸序列; (g) 侧接有位点特异性重组酶靶序列的条件性等位基因; 或 (h) 操作性连接在大鼠细胞中具有活性的启动子的报道基因。

[0531] 10. 实施方案1-9中任一项的方法, 其中所述目标基因组基因座包含 (i) 与所述5'大鼠同源臂互补的第一核酸序列; 和 (ii) 与所述3'大鼠同源臂互补的第二核酸序列。

[0532] 11. 实施方案10的方法, 其中所述第一核酸序列和所述第二核酸序列相隔至少5kb, 但小于3Mb。

[0533] 12. 实施方案10的方法, 其中所述第一核酸序列和所述第二核酸序列相隔至少5kb, 但小于10kb; 至少10kb, 但小于20kb; 至少20kb, 但小于40kb; 至少40kb, 但小于60kb; 至少60kb, 但小于80kb; 至少约80kb, 但小于100kb; 至少100kb, 但小于150kb; 或至少150kb, 但小于200kb; 至少约200kb, 但小于约300kb; 至少约300kb, 但小于约400kb; 至少约400kb, 但小于约500kb; 至少约500kb, 但小于约1Mb; 至少约1Mb, 但小于约1.5Mb; 至少约1.5Mb, 但小于约2Mb; 至少约2Mb, 但小于约2.5Mb; 或至少约2.5Mb, 但小于约3Mb。

[0534] 13. 实施方案1-12中任一项的方法, 其中引入步骤 (a) 进一步包括引入编码促进在所述多潜能大鼠细胞中在所述靶向构建体和所述目标基因组基因座之间的同源重组的核酸酶试剂的第二核酸。

[0535] 14. 实施方案13的方法, 其中所述核酸酶试剂包含 (a) 包含融合到FokI核酸内切酶的基于锌指的DNA结合结构域的嵌合蛋白; 或 (b) 包含融合到FokI核酸内切酶的转录激活因子样效应因子核酸酶 (TALEN) 的嵌合蛋白。

[0536] 15. 实施方案1-12中任一项的方法, 其中引入步骤 (a) 进一步包括向所述多潜能大鼠细胞中引入: (i) 包含操作性连接编码成簇规律间隔短回文重复序列 (CRISPR) 相关 (Cas) 蛋白的第一核酸序列的第一启动子的第一表达构建体, (ii) 包含操作性连接与向导RNA (gRNA) 连接的基因组靶序列的第二启动子的第二表达构建体, 其中所述基因组靶序列在3'端由原间隔区邻近基序 (PAM) 序列直接侧接。

[0537] 16. 实施方案15的方法, 其中所述目标基因组基因座包含SEQ ID NO:1的核苷酸序列。

[0538] 17. 实施方案15或16的方法, 其中所述gRNA包含编码成簇规律间隔短回文重复序列 (CRISPR) RNA (crRNA) 和反式激活CRISPR RNA (tracrRNA) 的第三核酸序列。

[0539] 18. 实施方案15、16或17的方法, 其中所述Cas蛋白为Cas9。

[0540] 19. 实施方案15、16、17或18的方法, 其中所述gRNA包含: (a) SEQ ID NO:2的核酸序列的嵌合RNA; 或 (b) SEQ ID NO:3的核酸序列的嵌合RNA。

[0541] 20. 实施方案17的方法, 其中所述crRNA包含SEQ ID NO:4; SEQ ID NO:5; 或SEQ ID

NO:6。

[0542] 21. 实施方案17的方法,其中所述tracrRNA包含SEQ ID NO:7或SEQ ID NO:8。

[0543] 22. 一种修饰的大鼠基因组基因座,其包括:(i) 插入同源或直系同源人类核酸序列;(ii) 用同源或直系同源人类核酸序列替换内源大鼠核酸序列;或(iii) 其组合,其中所述修饰的大鼠基因组基因座能够经由种系传递。

[0544] 23. 实施方案22的修饰的大鼠基因组基因座,其中所述插入或替换的尺寸为约5kb-约400kb。

[0545] 24. 实施方案22的大鼠基因组基因座,其中所述插入或替换的尺寸为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb。

[0546] 25. 一种用于制造人源化大鼠的方法,其包括:(a) 用包含人类核酸的靶向构建体靶向在多潜能大鼠细胞中的目标基因组基因座以形成基因修饰的多潜能大鼠细胞;(b) 将所述基因修饰的多潜能大鼠细胞引入宿主大鼠胚胎中;和(c) 在代孕母体中孕育所述宿主大鼠胚胎,其中所述代孕母体生成包含修饰的基因组基因座的大鼠子代,所述修饰的基因组基因座包括:(i) 插入人类核酸序列;(ii) 在所述目标基因组基因座处用同源或直系同源人类核酸序列替换所述大鼠核酸序列;(iii) 包含人类核酸序列和大鼠核酸序列的嵌合核酸序列;或(iv) 其组合,其中所述修饰的基因组基因座能够经由种系传递。

[0547] 26. 实施方案25的方法,其中所述靶向构建体为大靶向载体(LTVEC),且所述LTVEC的5'同源臂和3'同源臂的总和为至少10kb,但小于150kb。

[0548] 27. 实施方案26的方法,其中所述靶向构建体的5'同源臂和3'同源臂的总和为约10kb-约30kb、约20kb-40kb、约40kb-约60kb、约60kb-约80kb、或约80kb-约100kb、约100kb-约120kb或约120kb-150kb。

[0549] 28. 实施方案25、26或27的方法,其中所述人类核酸序列为至少5kb,但小于400kb。

[0550] 29. 实施方案25、26或27的方法,其中所述人类核酸序列为至少5kb,但小于10kb;至少10kb,但小于20kb;至少20kb,但小于40kb;至少40kb,但小于60kb;至少60kb,但小于80kb;至少约80kb,但小于100kb;至少100kb,但小于150kb;至少150kb,但小于200kb;至少200kb,但小于250kb;至少250kb,但小于300kb;至少300kb,但小于350kb;或至少350kb,但小于400kb。

[0551] 30. 实施方案25-29中任一项的方法,其中所述多潜能大鼠细胞为大鼠胚胎干(ES)细胞。

[0552] 31. 实施方案25-30中任一项的方法,其中所述多潜能大鼠细胞来源于DA株或ACI株。

[0553] 32. 实施方案25-31中任一项的方法,其中所述多潜能大鼠细胞的特征在于表达包括Dnmt3L、Eras、Err- $\beta$ 、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1的至少一种多潜能标记或其组合。

[0554] 33. 实施方案25-31中任一项的方法,其中所述多潜能大鼠细胞的特征在于以下特点中的一种或多种:(a) 缺乏包括c-Myc、Ecat1和/或Rexo1的一种或多种多潜能标记的表达;(b) 缺乏包括Brachyury和/或Bmpr2的一种或多种中胚层标记的表达;(c) 缺乏包括

Gata6、Sox17和/或Sox7的一种或多种内胚层标记的表达;或(d)缺乏包括Nestin和/或Pax6的一种或多种神经标记的表达。

[0555] 34.一种包含人源化基因组基因座的修饰的大鼠,其中所述人源化基因组基因座包含:(i)插入同源或直系同源人类核酸序列;(ii)在内源基因组基因座处用同源或直系同源人类核酸序列替换大鼠核酸序列;(iii)包含人类核酸序列和大鼠核酸序列的嵌合核酸序列;或(iv)其组合,其中所述人源化基因组基因座能够经由种系传递。

[0556] 35.一种在其基因组基因座中包含靶向基因修饰的大鼠或大鼠细胞,其中所述基因组基因座为干扰白细胞素-2受体 $\gamma$ 基因座、ApoE基因座、Rag1基因座、Rag2基因座或Rag2/Rag1基因座,其中所述靶向基因修饰包括:(a)在所述基因组基因座处缺失内源大鼠核酸序列;(b)插入同源核酸、直系同源核酸或包含人类核酸序列和大鼠核酸序列的嵌合核酸;或(c)其组合,其中所述靶向基因修饰可经由所述大鼠或自所述大鼠细胞繁殖的大鼠的种系传递。

[0557] 36.实施方案35的大鼠或大鼠细胞,其中(a)在所述基因组基因座处所述内源大鼠核酸的缺失为至少约10kb;或(b)在所述基因组基因座处所述内源大鼠核酸的缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb;或(c)在所述基因组基因座处所述外源核酸序列的插入为至少约5kb;或(d)在所述基因组基因座处所述外源核酸序列的插入为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb。

[0558] 37.实施方案35或36的大鼠或大鼠细胞,其中(a)在所述干扰白细胞素-2受体 $\gamma$ 基因座处的靶向基因修饰引起干扰白细胞素-2受体 $\gamma$ 蛋白活性减小或缺乏;(b)在所述ApoE基因座处的靶向基因修饰引起ApoE蛋白活性减小或缺乏;(c)在所述Rag1基因座处的靶向基因修饰引起Rag1蛋白活性减小或缺乏;(d)在所述Rag2基因座处的靶向基因修饰引起Rag2蛋白活性减小或缺乏;或(e)在所述Rag2/Rag1基因座处的靶向基因修饰引起Rag2蛋白活性和Rag1活性减小或缺乏。

[0559] 38.实施方案35、36或37的大鼠或大鼠细胞,其中所述干扰白细胞素-2受体 $\gamma$ 基因座的靶向基因修饰包括:(a)缺失整体大鼠干扰白细胞素-2受体 $\gamma$ 编码区或其部分;(b)用人类干扰白细胞素-2受体 $\gamma$ 编码区或其部分替换所述整体大鼠干扰白细胞素-2受体 $\gamma$ 编码区或其部分;(c)用人类干扰白细胞素-2受体 $\gamma$ 的胞外结构域替换所述大鼠干扰白细胞素-2受体 $\gamma$ 编码区的胞外结构域;或(d)所述干扰白细胞素-2受体 $\gamma$ 基因座的至少3kb缺失。

[0560] 39.实施方案35-37中任一项的大鼠或大鼠细胞,其中所述ApoE基因座的靶向基因修饰包括:(a)整体ApoE编码区或其部分的缺失;或(b)包含所述ApoE编码区的ApoE基因座的至少1.8kb缺失。

[0561] 40.实施方案35-37中任一项的大鼠或大鼠细胞,其中所述Rag2基因座的靶向基因修饰包括:(a)整体Rag2编码区或其部分的缺失;或(b)包含所述Rag2编码区的Rag2基因座的至少5.7kb缺失。

[0562] 41. 实施方案35-37中任一项的大鼠或大鼠细胞,其中所述Rag2/Rag1基因座的靶向基因修饰包括:(a)所述整体Rag2编码区或其部分的缺失和所述整体Rag1编码区或其部分的缺失;或(b)包含所述Rag2编码区的Rag2/Rag1基因座的至少16kb缺失。

[0563] 42. 实施方案35-41中任一项的大鼠或大鼠细胞,其中所述靶向基因修饰包括在所述干扰白细胞素-2受体 $\gamma$ 基因座、所述ApoE基因座、所述Rag1基因座、所述Rag2基因座或所述Rag2/Rag1基因座处插入包含选择性标记的表达盒。

[0564] 43. 实施方案42中任一项的大鼠或大鼠细胞,其中所述表达盒包含操作性连接在所述基因组基因座处的内源启动子和操作性连接选择性标记的人类泛素启动子的lacZ基因。

[0565] 44. 实施方案35-43中任一项的大鼠或大鼠细胞,其中在所述干扰白细胞素-2受体 $\gamma$ 基因座、所述ApoE基因座、所述Rag1基因座、所述Rag2基因座或所述Rag2/Rag1基因座中的靶向基因修饰包括自缺失选择盒的插入。

[0566] 45. 实施方案44的大鼠或大鼠细胞,其中所述自缺失选择盒包含操作性连接在所述大鼠细胞中具有活性的启动子的选择性标记基因和操作性连接雄性生殖细胞特异性启动子的重组酶基因,其中所述自缺失盒由通过所述重组酶识别的重组识别位点侧接。

[0567] 46. 实施方案45的大鼠或大鼠细胞,其中(a)所述雄性生殖细胞特异性启动子为鱼精蛋白-1启动子;或(b)所述重组酶基因编码Cre,且所述重组识别位点为loxP位点。

[0568] 47. 实施方案35-46中任一项的大鼠或大鼠细胞,其中在所述基因组基因座处所述外源核酸序列的插入包含操作性连接内源干扰白细胞素-2受体 $\gamma$ 启动子、内源ApoE启动子、内源Rag1启动子或内源Rag2启动子的报道基因核酸。

[0569] 48. 实施方案47的大鼠或大鼠细胞,其中所述报道基因核酸编码包含 $\beta$ -半乳糖苷酶、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(EYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶或其组合的报道基因。

[0570] 49. 实施方案35-48中任一项的大鼠细胞,其中所述大鼠细胞为多潜能大鼠细胞或大鼠胚胎干(ES)细胞。

[0571] 50. 实施方案49的大鼠细胞,其中所述多潜能大鼠细胞或所述大鼠胚胎干(ES)细胞(a)来源于DA株或ACI株;(b)以表达包括Dnmt3L、Eras、Err- $\beta$ 、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1的至少一种多潜能标记或其组合为特征;或(c)以一种或多种以下特征为特征:(i)缺乏包括c-Myc、Ecat1和/或Rexo1的一种或多种多潜能标记的表达;(ii)缺乏包括Brachyury和/或Bmpr2的中胚层标记的表达;(iii)缺乏包括Gata6、Sox17和/或Sox7的一种或多种内胚层标记的表达;或(iv)缺乏包括Nestin和/或Pax6的一种或多种神经标记的表达。

[0572] 51. 一种用于修饰在多潜能大鼠细胞中在干扰白细胞素-2受体 $\gamma$ 基因座、ApoE基因座、Rag1基因座、Rag2基因座或Rag2/Rag1基因座中的靶基因组基因座的方法,所述方法包括:(a)向所述多潜能大鼠细胞中引入包含侧接有与所述靶基因组基因座同源的5'大鼠同源臂和3'大鼠同源臂的插入核酸的靶向载体,(b)鉴定在所述靶基因组基因座处包括靶向基因修饰的基因修饰的多潜能大鼠细胞,其中所述靶向基因修饰能够经由自所述多潜能大鼠细胞繁殖的大鼠的种系传递。



[0573] 52. 实施方案51的方法,其中所述靶向载体为大靶向载体(LTVEC),其中所述5'大鼠同源臂和所述3'大鼠同源臂的总和为至少约10kb,但小于约150kb。

[0574] 53. 实施方案51或52的方法,其中向所述多潜能大鼠细胞中引入所述靶向载体引起:(i)在所述靶基因组基因座处缺失内源大鼠核酸序列;(ii)在所述靶基因组基因座处插入外源核酸序列;或(iii)其组合。

[0575] 54. 实施方案53的方法,其中(a)在所述基因组基因座处所述内源大鼠核酸的缺失为至少约10kb;或(b)在所述基因组基因座处所述内源大鼠核酸的缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb;(c)在所述基因组基因座处所述外源核酸序列的插入为至少约5kb;或(d)在所述基因组基因座处所述外源核酸序列的插入为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb。

[0576] 55. 实施方案51-54中任一项的方法,其中(a)在所述干扰白细胞素-2受体 $\gamma$ 基因座处的靶向基因修饰引起干扰白细胞素-2受体 $\gamma$ 蛋白活性减小或缺乏;(b)在所述ApoE基因座处的靶向基因修饰引起ApoE蛋白活性减小或缺乏;(c)在所述Rag1基因座处的靶向基因修饰引起Rag1蛋白活性减小或缺乏;(d)在所述Rag2基因座处的靶向基因修饰引起Rag2蛋白活性减小或缺乏;或(e)在所述Rag2/Rag1基因座处的靶向基因修饰引起Rag2蛋白活性和Rag1蛋白活性减小或缺乏。

[0577] 56. 实施方案51-54中任一项的方法,其中所述干扰白细胞素-2受体 $\gamma$ 基因座的靶向基因修饰包括(a)整体大鼠干扰白细胞素-2受体 $\gamma$ 编码区或其部分的缺失;(b)用人类干扰白细胞素-2受体 $\gamma$ 编码区或其部分替换所述整体大鼠干扰白细胞素-2受体 $\gamma$ 编码区或其部分;(c)用人类干扰白细胞素-2受体 $\gamma$ 的胞外结构域替换所述大鼠干扰白细胞素-2受体 $\gamma$ 编码区的胞外结构域;或(d)包含所述干扰白细胞素-2受体 $\gamma$ 编码区的所述干扰白细胞素-2受体 $\gamma$ 基因座的至少3kb缺失。

[0578] 57. 实施方案51-55中任一项的方法,其中所述ApoE基因座的靶向基因修饰包括:(a)整体ApoE编码区或其部分的缺失;或(b)包含所述ApoE编码区的所述ApoE基因座的至少1.8kb缺失。

[0579] 58. 实施方案51-55中任一项的方法,其中所述Rag2基因座的靶向基因修饰包括:(a)整体Rag2编码区或其部分的缺失;或(b)包含所述Rag2编码区的所述Rag2基因座的至少5.7kb缺失。

[0580] 59. 实施方案51-55中任一项的方法,其中所述Rag1/Rag2基因座的靶向基因修饰包括:(a)所述整体Rag2编码区或其部分的缺失和所述整体Rag1编码区或其部分的缺失;或(b)包含所述Rag2和Rag1编码区的所述Rag2/Rag1基因座的至少16kb缺失。

[0581] 60. 实施方案51-59中任一项的方法,其中所述插入核酸包含包含编码选择性标记的多核苷酸的表达盒。

[0582] 61. 实施方案60的方法,其中所述表达盒包含操作性连接在所述基因组基因座处的内源启动子和操作性连接选择性标记基因的人类泛素启动子的lacZ基因。

- [0583] 62. 实施方案51-60中任一项的方法,其中所述插入核酸包含自缺失选择盒。
- [0584] 63. 实施方案62的方法,其中所述自缺失选择盒包含操作性连接在所述大鼠多潜能细胞中具有活性的启动子的选择性标记和编码操作性连接雄性生殖细胞特异性启动子的重组酶的多核苷酸,其中所述自缺失盒由通过所述重组酶识别的重组识别位点侧接。
- [0585] 64. 实施方案63的方法,其中 (a) 所述雄性生殖细胞特异性启动子为鱼精蛋白-1启动子;或 (b) 所述重组酶基因编码Cre,且所述重组识别位点为loxP位点。
- [0586] 65. 实施方案53的方法,其中在所述基因组基因座处所述外源核酸序列的插入包含操作性连接内源干扰白细胞素-2受体  $\gamma$  启动子、内源ApoE启动子、内源Rag1启动子或内源Rag2启动子的报道基因核酸序列。
- [0587] 66. 实施方案65的方法,其中所述报道基因核酸序列编码包括 $\beta$ -半乳糖苷酶、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白 (EYFP)、Emerald、增强的绿色荧光蛋白 (EGFP)、CyPet、青色荧光蛋白 (CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶或其组合的报道基因。
- [0588] 67. 实施方案51-66中任一项的方法,其中所述多潜能大鼠细胞为大鼠胚胎干 (ES) 细胞。
- [0589] 68. 实施方案51-67中任一项的方法,其中所述多潜能大鼠细胞 (a) 来源于DA株或ACI株;或 (b) 以表达包含Oct-4、Sox-2、碱性磷酸酶或其组合的多潜能标记为特征;或 (c) 以一种或多种以下特征为特征: (i) 缺乏包括c-Myc、Ecat1和/或Rexol1的一种或多种多潜能标记的表达; (ii) 缺乏包括Brachyury和/或Bmpr2的中胚层标记的表达; (iii) 缺乏包括Gata6、Sox17和/或Sox7的一种或多种内胚层标记的表达;或 (iv) 缺乏包括Nestin和/或Pax6的一种或多种神经标记的表达。
- [0590] 69. 实施方案51-68中任一项的方法,进一步包括鉴定在所述靶基因组基因座处的靶向基因修饰,其中所述鉴定步骤采用用于评定在所述靶基因组基因座处的等位基因修饰 (MOA) 的定量测定。
- [0591] 70. 实施方案51-69中任一项的方法,其中引入步骤 (a) 进一步包括引入编码促进在所述多潜能大鼠细胞中在所述靶向载体和所述靶基因组基因座之间的同源重组的核酸酶试剂的第二核酸。
- [0592] 71. 实施方案70的方法,其中所述核酸酶试剂包含包含融合到FokI核酸内切酶的基于锌指的DNA结合结构域的嵌合蛋白。
- [0593] 72. 实施方案71的方法,其中所述方法产生所述靶基因组基因座的双等位修饰。
- [0594] 73. 实施方案51-70中任一项的方法,其中引入步骤 (a) 进一步包括向所述多潜能大鼠细胞中引入: (i) 包含操作性连接编码成簇规律间隔短回文重复序列 (CRISPR) 相关 (Cas) 蛋白的第一核酸序列的第一启动子的第一表达构建体, (ii) 包含操作性连接与向导RNA (gRNA) 连接的基因组靶序列的第二启动子的第二表达构建体,其中所述基因组靶序列在3'端由原间隔区邻近基序 (PAM) 序列直接侧接。
- [0595] 74. 实施方案73的方法,其中所述目标基因组基因座包含SEQ ID NO:1的核苷酸序列。
- [0596] 75. 实施方案73或74的方法,其中所述gRNA包含编码成簇规律间隔短回文重复序列 (CRISPR) RNA (crRNA) 和反式激活CRISPR RNA (tracrRNA) 的第三核酸序列。

- [0597] 76. 实施方案73的方法,其中所述Cas蛋白为Cas9。
- [0598] 77. 实施方案73、74或75的方法,其中所述gRNA包含:(a) SEQ ID NO:2的核酸序列的嵌合RNA;或(b) SEQ ID NO:3的核酸序列的嵌合RNA。
- [0599] 78. 实施方案75的方法,其中所述crRNA包含SEQ ID NO:4;SEQ ID NO:5;或SEQ ID NO:6。
- [0600] 79. 实施方案75的方法,其中所述tracrRNA包含SEQ ID NO:7或SEQ ID NO:8。
- [0601] 80. 实施方案35-50中任一项的大鼠或大鼠细胞,其中所述大鼠或大鼠细胞包含在所述干扰白细胞素-2受体 $\gamma$ 基因座、所述ApoE基因座、所述Rag1基因座、所述Rag2基因座和/或所述Rag2/Rag1基因座处的靶向基因修饰。
- [0602] 81. 实施方案80的大鼠或大鼠细胞,其中所述大鼠或大鼠细胞包含在所述干扰白细胞素-2受体 $\gamma$ 基因座和所述Rag2/Rag1基因座处的靶向基因修饰。
- [0603] 另外的非限制性实施方案包括:
- [0604] 1. 一种修饰在真核细胞中的目标基因组基因座的方法,其包括:(a) 向所述真核细胞中引入:(i) 包含侧接有5'同源臂和3'同源臂的第一核酸的大靶向载体(LTVEC),其中所述LTVEC为至少10kb,(ii) 包含操作性连接编码Cas蛋白的第二核酸的第一启动子的第一表达构建体,(iii) 包含操作性连接编码包含杂化到靶序列的核苷酸序列的向导RNA(gRNA)和反式激活CRISPR RNA(tracrRNA)的第三核酸的第二启动子的第二表达构建体,其中所述第一启动子和所述第二启动子在所述真核细胞中具有活性;和(b) 鉴定在所述目标基因组基因座处包括靶向基因修饰的修饰的真核细胞。
- [0605] 2. 实施方案1的方法,其中所述靶向基因修饰为双等位基因基因修饰。
- [0606] 3. 实施方案1的方法,其中所述LTVEC为至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb或至少90kb。
- [0607] 4. 实施方案1的方法、其中所述LTVEC为至少100kb、至少150kb或至少200kb。
- [0608] 5. 实施方案1的方法,其中所述真核细胞为哺乳动物细胞。
- [0609] 6. 实施方案5的方法,其中所述哺乳动物细胞为成纤维细胞。
- [0610] 7. 实施方案1的方法,其中所述真核细胞为多潜能细胞。
- [0611] 8. 实施方案7的方法,其中所述多潜能细胞为人类多潜能细胞。
- [0612] 9. 实施方案8的方法,其中所述人类多潜能细胞为人类胚胎干(ES)细胞或人类成人干细胞。
- [0613] 10. 实施方案8的方法,其中所述人类多潜能细胞为发育受限的人类祖细胞。
- [0614] 11. 实施方案8的方法,其中所述人类多潜能细胞为人类诱导的多潜能干(iPS)细胞。
- [0615] 12. 实施方案1的方法,其中所述Cas蛋白为Cas9。
- [0616] 13. 实施方案1的方法,其中所述靶序列在3'端由原间隔区邻近基序(PAM)序列直接侧接。
- [0617] 14. 实施方案1的方法,其中所述5'同源臂和所述3'同源臂的总和为约10kb-约150kb。
- [0618] 15. 实施方案1的方法,其中所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb

或约120kb-150kb。

[0619] 16. 实施方案1的方法,其中所述靶向基因修饰包括:(a)用同源或直系同源的核酸序列替换内源核酸序列;(b)缺失内源核酸序列;(c)缺失内源核酸序列,其中所述缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb;(d)插入外源核酸序列;(e)插入外源核酸序列,所述外源核酸序列为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb;(f)插入包含同源或直系同源的核酸序列的外源核酸序列;(g)插入包含人类核酸序列和非人类核酸序列的嵌合核酸序列;(h)插入侧接有位点特异性重组酶靶序列的条件性等位基因;(i)插入操作性连接在所述多潜能细胞中具有活性的第三启动子的可选择标记或报道基因;或(j)其组合。

[0620] 17. 实施方案1的方法,其中所述目标基因组基因座包含(i)与所述5'同源臂同源的5'靶序列;和(ii)与所述3'同源臂同源的3'靶序列。

[0621] 18. 实施方案17的方法,其中所述5'靶序列和所述3'靶序列相隔至少5kb,但小于3Mb。

[0622] 19. 实施方案17的方法,其中所述5'靶序列和所述3'靶序列相隔至少5kb,但小于10kb;至少10kb,但小于20kb;至少20kb,但小于40kb;至少40kb,但小于60kb;至少60kb,但小于80kb;至少约80kb,但小于100kb;至少100kb,但小于150kb;或至少150kb,但小于200kb;至少约200kb,但小于约300kb;至少约300kb,但小于约400kb;至少约400kb,但小于约500kb;至少约500kb,但小于约1Mb;至少约1Mb,但小于约1.5Mb;至少约1.5Mb,但小于约2Mb;至少约2Mb,但小于约2.5Mb;或至少约2.5Mb,但小于约3Mb。

[0623] 20. 实施方案1的方法,其中所述目标基因组基因座包括干扰白细胞素-2受体 $\gamma$ 基因座、ApoE基因座、Rag1基因座、Rag2基因座或所述Rag1基因座和所述Rag2基因座两者。

[0624] 21. 实施方案1的方法,其中所述第一表达构建体和所述第二表达构建体在单一核酸分子上。

[0625] 22. 一种用于修饰基因组的方法,其包括在包含至少10kb的核酸序列的大靶向载体(LTVEC)存在下将所述基因组暴露于Cas蛋白和CRISPR RNA,其中在暴露于所述Cas蛋白、所述CRISPR RNA和所述LTVEC之后,所述基因组被修饰以含有至少10kb核酸序列。

[0626] 23. 实施方案22的方法,其中所述LTVEC包含至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb或至少90kb的核酸序列。

[0627] 24. 实施方案22的方法,其中所述LTVEC包含至少100kb、至少150kb或至少200kb的核酸序列。

[0628] 25. 一种用于修饰基因组的方法,其包括在大靶向载体(LTVEC)存在下使所述基因组与Cas蛋白、杂化到靶序列的CRISPR RNA,和tracrRNA接触,其中所述LTVEC为至少10kb且包含侧接有5'同源臂和3'同源臂的第一核酸,其中在所述LTVEC存在下与所述Cas蛋白、CRISPR RNA,和tracrRNA接触之后,所述基因组在目标基因组基因座处修饰以含有所述第一核酸。

- [0629] 26. 实施方案25的方法,其中所述基因组在真核细胞中,且将所述Cas蛋白、所述CRISPR RNA、所述tracrRNA和所述LTVEC引入所述真核细胞中。
- [0630] 27. 实施方案26的方法,其进一步包括鉴定在所述目标基因组基因座处包括靶向基因修饰的修饰的真核细胞。
- [0631] 28. 实施方案26或27的方法,其中所述CRISPR RNA和所述tracrRNA以单一向导RNA (gRNA)的形式一起引入。
- [0632] 29. 实施方案26或27的方法,其中所述CRISPR RNA和所述tracrRNA单独地引入。
- [0633] 30. 实施方案26-29中任一项的方法,其中:(a)所述Cas蛋白以蛋白质、编码所述Cas蛋白的信使RNA (mRNA) 或编码所述Cas蛋白的DNA的形式引入所述真核细胞中;(b)所述CRISPR RNA以编码所述CRISPR RNA的RNA或DNA的形式引入所述真核细胞中;和(c)所述tracrRNA以编码所述tracrRNA的RNA或DNA的形式引入所述真核细胞中。
- [0634] 31. 实施方案30的方法,其中所述Cas蛋白、所述CRISPR RNA和所述tracrRNA作为蛋白质-RNA复合体引入所述真核细胞中。
- [0635] 32. 实施方案30的方法,其中:(a) 编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式;(b) 编码所述CRISPR RNA的所述DNA呈包含操作性连接编码所CRISPR RNA的第三核酸的第二启动子的第二表达构建体的形式;且(c) 编码所述tracrRNA的所述DNA呈包含操作性连接编码所述tracrRNA的第四核酸的第三启动子的第三表达构建体的形式,其中所述第一启动子、所述第二启动子和所述第三启动子在所述真核细胞中具有活性。
- [0636] 33. 实施方案32的方法,其中所述第一表达构建体、所述第二表达构建体和/或所述第三表达构建体在单一核酸分子上。
- [0637] 34. 实施方案30的方法,其中:(a) 编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式;且(b) 编码所述CRISPR RNA的所述DNA和编码所述tracrRNA的所述DNA呈包含操作性连接编码包含所述CRISPR RNA和所述tracrRNA的gRNA的第三核酸的第二启动子的第二表达构建体的形式;其中所述第一启动子和所述第二启动子在所述真核细胞中具有活性。
- [0638] 35. 实施方案34的方法,其中所述第一表达构建体和所述第二表达构建体在单一核酸分子上。
- [0639] 36. 实施方案27-35中任一项的方法,其中所述靶向基因修饰同时包括在所述目标基因组基因座处缺失内源核酸序列和在所述目标基因组基因座处插入所述第一核酸。
- [0640] 37. 实施方案27-36中任一项方法,其中所述靶向基因修饰为双等位基因基因修饰。
- [0641] 38. 实施方案37的方法,其中所述双等位基因基因修饰包括在两种同源染色体中的目标基因组基因座处缺失内源核酸序列和插入所述第一核酸。
- [0642] 39. 实施方案27-36中任一项的方法,其中所述修饰的真核细胞在所述目标基因组基因座处为半合的。
- [0643] 40. 实施方案39的方法,其中在一种染色体中在目标基因组基因座处的靶向基因修饰包括缺失内源核酸序列和插入所述第一核酸。
- [0644] 41. 实施方案39的方法,其中所述靶向基因修饰包括:(1) 在两种同源染色体中在

目标基因组基因座处缺失内源核酸序列;和(2)在第一染色体中所述第一核酸插入所述目标基因组基因座中和在第二染色体中破坏所述目标基因组基因座。

[0645] 42. 实施方案25-41中任一项的方法,其中所述LTVEC为至少15kb、至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb或至少90kb。

[0646] 43. 实施方案25-42中任一项的方法,其中所述LTVEC为至少100kb、至少150kb或至少200kb。

[0647] 44. 实施方案25-43中任一项的方法,其中所述第一核酸为至少20kb、至少30kb、至少40kb、至少50kb、至少60kb、至少70kb、至少80kb、至少90kb、至少100kb、至少150kb、至少200kb、至少250kb或至少300kb。

[0648] 45. 实施方案26-44中任一项的方法,其中所述真核细胞为哺乳动物细胞。

[0649] 46. 实施方案45的方法,其中所述哺乳动物细胞为成纤维细胞。

[0650] 47. 实施方案26-43中任一项的方法,其中所述真核细胞为多潜能细胞。

[0651] 48. 实施方案47的方法,其中所述多潜能细胞为非人类多潜能细胞。

[0652] 49. 实施方案48的方法,其中所述非人类多潜能细胞为啮齿动物多潜能细胞。

[0653] 50. 实施方案49的方法,其中所述啮齿动物多潜能细胞为小鼠或大鼠胚胎干(ES)细胞。

[0654] 51. 实施方案47的方法,其中所述多潜能细胞为人类多潜能细胞。

[0655] 52. 实施方案51的方法,其中所述人类多潜能细胞为人类胚胎干(ES)细胞或人类成人干细胞。

[0656] 53. 实施方案51的方法,其中所述人类多潜能细胞为发育受限的人类祖细胞。

[0657] 54. 实施方案51的方法,其中所述人类多潜能细胞为人类诱导的多潜能干(iPS)细胞。

[0658] 55. 实施方案25-54中任一项的方法,其中所述Cas蛋白为Cas9。

[0659] 56. 实施方案25-55中任一项的方法,其中所述靶序列由原间隔区邻近基序(PAM)序列直接侧接。

[0660] 57. 实施方案25-56中任一项的方法,其中所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约150kb。

[0661] 58. 实施方案25-57中任一项的方法,其中所述LTVEC的5'同源臂和3'同源臂的总和为约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约120kb或约120kb-150kb。

[0662] 59. 实施方案27-58中任一项的方法,其中所述靶向基因修饰包括:(a)用同源或直系同源的核酸序列替换内源核酸序列;(b)缺失内源核酸序列;(c)缺失内源核酸序列,其中所述缺失为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、或约150kb-约200kb、约200kb-约300kb、约300kb-约400kb、约400kb-约500kb、约500kb-约1Mb、约1Mb-约1.5Mb、约1.5Mb-约2Mb、约2Mb-约2.5Mb或约2.5Mb-约3Mb;(d)插入外源核酸序列;(e)插入外源核酸序列,所述外源核酸序列为约5kb-约10kb、约10kb-约20kb、约20kb-约40kb、约40kb-约60kb、约60kb-约80kb、约80kb-约100kb、约100kb-约150kb、约150kb-约200kb、约200kb-约250kb、约250kb-约300kb、约300kb-约350kb或约350kb-约400kb;(f)插入包含同源或直系同源的核酸序列的

外源核酸序列；(g) 插入包含人类核酸序列和非人类核酸序列的嵌合核酸序列；(h) 插入侧接有位点特异性重组酶靶序列的条件性等位基因；(i) 插入操作性连接在所述多潜能细胞中具有活性的第三启动子的可选择标记或报道基因；或 (j) 其组合。

[0663] 60. 实施方案25-59中任一项的方法,其中所述目标基因组基因座包含(i)与所述5'同源臂同源的5'靶序列;和(ii)与所述3'同源臂同源的3'靶序列。

[0664] 61. 实施方案60的方法,其中所述5'靶序列和所述3'靶序列相隔至少5kb,但小于3Mb。

[0665] 62. 实施方案60的方法,其中所述5'靶序列和所述3'靶序列相隔至少5kb,但小于10kb;至少10kb,但小于20kb;至少20kb,但小于40kb;至少40kb,但小于60kb;至少60kb,但小于80kb;至少约80kb,但小于100kb;至少100kb,但小于150kb;或至少150kb,但小于200kb;至少约200kb,但小于约300kb;至少约300kb,但小于约400kb;至少约400kb,但小于约500kb;至少约500kb,但小于约1Mb;至少约1Mb,但小于约1.5Mb;至少约1.5Mb,但小于约2Mb;至少约2Mb,但小于约2.5Mb;或至少约2.5Mb,但小于约3Mb。

[0666] 63. 实施方案60的方法,其中所述5'靶序列和所述3'靶序列相隔至少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。

[0667] 64. 实施方案25-63中任一项的方法,其中所述目标基因组基因座包括干扰白细胞素-2受体 $\gamma$ 基因座、ApoE基因座、Rag1基因座、Rag2基因座或所述Rag1基因座和所述Rag2基因座两者。

[0668] 65. 实施方案25-63中任一项的方法,其中所述目标基因组基因座包括Adamt5基因座、Trpa1基因座、Folh1基因座或ErbB4基因座。

[0669] 66. 实施方案25-63中任一项的方法,其中所述目标基因组基因座包括Lrp5基因座。

[0670] 67. 一种用于产生在目标基因组基因座处包括靶向基因修饰的F0代非人类动物的方法,所述方法包括:(a)在大靶向载体(LTVEC)存在下使在非人类ES细胞中的基因组与Cas蛋白、CRISPR RNA,和tracrRNA接触以形成修饰的非人类ES细胞,其中所述LTVEC为至少10kb且包含侧接有5'同源臂和3'同源臂的第一核酸;(b)鉴定在所述目标基因组基因座含所述靶向基因修饰的修饰的非人类ES细胞;(c)将所述修饰的非人类ES细胞引入非人类宿主胚胎中;和(d)在代孕母体中孕育所述非人类宿主胚胎,其中所述代孕母体产生在所述目标基因组基因座处包括所述靶向基因修饰的所述F0代非人类动物。

[0671] 68. 实施方案67的方法,其中所述CRISPR RNA和所述tracrRNA以单一向导RNA(gRNA)的形式一起引入。

[0672] 69. 实施方案67的方法,其中所述CRISPR RNA和所述tracrRNA单独地引入。

[0673] 70. 实施方案67-69中任一项的方法,其中:(a)所述Cas蛋白以蛋白质、编码所述Cas蛋白的信使RNA(mRNA)或编码所述Cas蛋白的DNA的形式引入所述非人类ES细胞中;(b)所述CRISPR RNA以编码所述CRISPR RNA的RNA或DNA的形式引入所述非人类ES细胞中;且(c)所述tracrRNA以编码所述tracrRNA的RNA或DNA的形式引入所述非人类ES细胞中。

[0674] 71. 实施方案70的方法,其中所述Cas蛋白、所述CRISPR RNA和所述tracrRNA作为

蛋白质-RNA复合体引入所述非人类ES细胞中。

[0675] 72. 实施方案70的方法, 其中: (a) 编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式; (b) 编码所述CRISPR RNA的所述DNA呈包含操作性连接编码所CRISPR RNA的第三核酸的第二启动子的第二表达构建体的形式; 且 (c) 编码所述tracrRNA的所述DNA呈包含操作性连接编码所述tracrRNA的第四核酸的第三启动子的第三表达构建体的形式, 其中所述第一启动子、所述第二启动子和所述第三启动子在所述非人类ES细胞中具有活性。

[0676] 73. 实施方案72的方法, 其中所述第一表达构建体、所述第二表达构建体和/或所述第三表达构建体在单一核酸分子上。

[0677] 74. 实施方案70的方法, 其中: (a) 编码所述Cas蛋白的所述DNA呈包含操作性连接编码所述Cas蛋白的第二核酸的第一启动子的第一表达构建体的形式; 且 (b) 编码所述CRISPR RNA的所述DNA和编码所述tracrRNA的所述DNA呈包含操作性连接编码包含所述CRISPR RNA和所述tracrRNA的gRNA的第三核酸的第二启动子的第二表达构建体的形式; 其中所述第一启动子和所述第二启动子在所述非人类ES细胞中具有活性。

[0678] 75. 实施方案74的方法, 其中所述第一表达构建体和所述第二表达构建体在单一核酸分子上。

[0679] 76. 实施方案67-75中任一项的方法, 其中所述靶向基因修饰同时包括在所述目标基因组基因座处缺失内源核酸序列和在所述目标基因组基因座处插入所述第一核酸。

[0680] 77. 实施方案67-76中任一项的方法, 其中所述靶向基因修饰为双等位基因基因修饰。

[0681] 78. 实施方案77的方法, 其中所述双等位基因基因修饰包括在两种同源染色体中的目标基因组基因座处内源核酸序列的缺失和所述第一核酸的插入。

[0682] 79. 实施方案67-76中任一项的方法, 其中所述修饰的非人类ES细胞在所述目标基因组基因座处为半合的。

[0683] 80. 实施方案79的方法, 其中在一种染色体中在目标基因组基因座处的靶向基因修饰包括内源核酸序列的缺失和所述第一核酸的插入。

[0684] 81. 实施方案79的方法, 其中所述靶向基因修饰包括: (1) 在两种同源染色体中在目标基因组位点处缺失内源核酸序列; 和 (2) 在第一染色体中所述第一核酸插入所述目标基因组基因座中和在第二染色体中破坏所述目标基因组基因座。

[0685] 82. 实施方案67-81中任一项的方法, 其中所述Cas蛋白为Cas9。

[0686] 实施例

[0687] 提出以下实施例以便向本领域的普通技术人员提供对如何进行和使用本发明的完全公开和描述, 并不意欲限制本发明人看待其发明的范围, 也不意欲表示以下实验是执行的全部或仅有的实验。已经进行尝试以确保关于所使用的数值 (例如, 量、温度等) 的准确性, 但应该考虑一些实验误差和偏差。除非另外指出, 否则份数是重量份, 分子量是重均分子量, 温度是摄氏度并且压力是大气压或接近大气压。

[0688] 实施例1. 大鼠ES细胞衍生和表征

[0689] 1.1. 大鼠ES细胞表征

[0690] 如在图1中所示, 大鼠ESC生长成紧实的球形集落, 其通常在培养皿中脱离并浮动



(特写,图8)。大鼠ESC表达包括Oct-4(图2A)和Sox2(图2B)的多潜能标记,且表达高水平的碱性磷酸酶(图3)。株系DA.2B的核型为42X,Y(图4)。大鼠ESC常变成四倍体;因此,株系通过计数中期染色体散布来预筛选;随后正式分析具有大部分正常计数的株系的核型。

[0691] ACI胚泡从商业获得的超排卵雌性收集。DA胚泡自商业获得的冷冻的8-细胞胚胎培养。透明带用酸性台氏液(Acid Tyrode)移除;且将胚泡涂铺到有丝分裂失活的MEF上。生长晕使用标准方法挑选并扩增。所有胚泡使用2i培养基涂铺、培养并扩增(Li等人,(2008) Germline competent embryonic stem cells derived from rat blastocysts,Cell 135:1299-1310;其整体通过引用的方式并入本文中)。

	ACI	DA
<b>胚胎来源</b>	胚泡(超排卵)	培养成胚泡的冷冻的 8-细胞胚胎
<b>涂铺的胚泡:</b>	107	22
[0692] <b>生长晕:</b>	32(30%的母细胞)	10(45%的母细胞)
<b>株系:</b>	16(50%的生长晕)	9(90%的生长晕)
<b>核型分析:</b>	3; 都是 42X,Y	6: 3 42X,X 3 42X,Y
<b>验证的 GLT:</b>	1(ACI.G1)	1 42X,X(DA.2C) 1 42X,Y(DA.2B)

[0693] 1.2.:大鼠产生

[0694] 嵌合大鼠通过胚泡注射和传送大鼠ESC基因组来产生。通过使用亲本ACI.G1大鼠ESC进行胚泡微量注射产生的嵌合体示于图9中。由在图9中的星号(\*)标记的ACI/SD嵌合体所生的F1白化刺豚鼠同窝幼仔示于图10中。

[0695] 亲本大鼠ESC的种系传递。

[0696] 三个整倍体大鼠ESC株系的多潜能性通过向白化SD胚泡中微量注射来评价。嵌合体通过指示大鼠ESC贡献的刺鼠毛色来鉴定(参见图10)。对于各株系,大多数嵌合体向F1后代传递rESC基因组(表2)。

株系 (核型)	配种的嵌 合体	种系传递 体	来自 GLT 嵌 合体的总幼仔	rESC-源性 幼仔	GLT 效率 (%)
[0697] <b>ACI.G1 (XY)</b>	5	<b>3 (60%)</b>	103	11	11
<b>DA.2B (XY)</b>	5	<b>4 (80%)</b>	129	11	9
<b>DA.2C (XX)</b>	3	<b>2 (66%)</b>	45	7	16

[0698] 1.3.:大鼠胚胎干细胞的衍生。

[0699] 超排卵方案,大鼠

[0700] 第0天:用妊娠母马血清注射:IP,20U(0.4ml)。

- [0701] 第1天:不采取行动
- [0702] 第2天:(46小时后):注射hCG,IP,50U(1ml)。
- [0703] -设置单一雌性交配。
- [0704] 第3天:检查栓塞。将雌性堵塞。这是第0.5天。
- [0705] 第6天(e3.5):使雌性安乐死,并冲洗胚胎。
- [0706] ES细胞衍生方案(超排卵)
- [0707] 第0天:
- [0708] 1) 用CO<sub>2</sub>使雌性大鼠安乐死。
- [0709] 2) 用70%乙醇擦拭腹侧腹部;使用剪刀打开腹侧体壁以暴露内脏。
- [0710] 3) 解剖出输卵管和子宫角,并将它们放置到含有温热N2B27培养基的组织培养皿中。洗出尽可能多的血液,并转移到具有N2B27的新培养皿中。
- [0711] 4) 使用1ml注射器和钝27g针,使培养基冲洗经过子宫角和输卵管以使胚泡排出到培养基中。
- [0712] 5) 用口吸移管收集胚泡,并且转移到含有KSOM+2i (1 $\mu$ M PD0325901、3 $\mu$ M CHIR99021)的胚胎培养皿中。KSOM为由Millipore生产的培养基。目录号为MR-106-D。
- [0713] 6) 在37°下培养过夜;7.5%CO<sub>2</sub>。
- [0714] ES细胞衍生方案(冷冻胚胎)
- [0715] 第0天:
- [0716] 1) 使冷冻的8细胞胚胎(商业上获得)解冻到M2培养基中。在室温下培养10分钟。
- [0717] 2) 转移到KSOM+2i中并培养过夜。
- [0718] ES细胞衍生方案(两者相同)
- [0719] 第1天:
- [0720] 1) 将空腔化胚胎转移到2i培养基中并培养过夜。
- [0721] 2) 在KSOM+2i中继续培养未空腔化的胚胎。
- [0722] 第2天:
- [0723] 1) 转移所有剩余的胚胎到2i培养基中(无论它们是否已空腔化)。
- [0724] 2) 培养过夜;在2i培养基中继续培养较早期的胚胎。
- [0725] 第3天:
- [0726] 1) 用酸性台氏液转移胚胎30-60秒以移除透明带。
- [0727] 2) 在2i培养基中洗涤胚胎3次以移除酸性台氏液。
- [0728] 3) 将各胚胎沉积到96孔饲养板的单独孔中(所述孔含有单层有丝分裂失活的小鼠胚胎成纤维细胞(MEF))。
- [0729] 4) 在2i培养基中培养过夜。
- [0730] 第4-5天:
- [0731] 1) 监测涂铺的胚胎的生长晕(无定形未分化的细胞团)是否存在。当生长晕的大小是涂铺胚胎的约两倍时,所述生长晕即准备用于转移。
- [0732] 2) 每天:用微量吸移管移除用过的培养基,并且用新鲜2i培养基替换。
- [0733] 3) 转移生长晕到新饲养孔中:
- [0734] a. 移除用过的培养基并用PBS温和地洗涤孔。

- [0735] b. 移除PBS, 并加入30 $\mu$ l 0.05%胰蛋白酶; 孵育10分钟。
- [0736] c. 通过加入30 $\mu$ l 2i+10%FBS来终止胰蛋白酶反应。
- [0737] d. 用微量移液器温和地离散细胞, 并转移孔的整个内含物到24孔饲养板的新孔中。这是第1代(P1)。
- [0738] e. 在2i培养基中培养过夜。
- [0739] 第5-8天: (时机取决于各株系扩增的快速程度)
- [0740] 1) 每天更换培养基(2i培养基)并监测具有ESC形态的集落是否存在。
- [0741] 2) 当集落出现时, 继续培养直至集落扩增至约50%汇合。
- [0742] 3) 如前将集落用胰蛋白酶处理并传代; 涂铺在饲养器上, 在6孔培养皿中, 每个株系1个孔。这是第2代(P2)。
- [0743] 进行中的事务:
- [0744] 1) 继续饲养并监测各株系, 直至约50%汇合。
- [0745] 2) 照常用胰蛋白酶处理细胞。
- [0746] 3) 用2i+10%FBS终止胰蛋白酶; 细胞通过离心(5', 在Beckman-Coulter台式离心机中1200rpm)粒化。
- [0747] 4) 吸出上清液, 且将细胞温和地再悬浮于400 $\mu$ l冷冻培养基(70%2i、20%FBS、10%DMSO)中。
- [0748] 5) 将细胞分配到2个小瓶中并在-80°下冷冻。这是第3代(P3)。
- [0749] 6) 对于长期储存, 转移小瓶到液N<sub>2</sub>储存库中。
- [0750] 2i培养基如下表3制备。

试剂	供应商	浓度
DMEM/F12 基础培养基	Invitrogen/Life Technologies	1x
神经基础培养基	Invitrogen/Life Technologies	1x
青霉素/链霉素	Invitrogen/Life Technologies	1%
L-谷氨酰胺	Invitrogen/Life Technologies	4 mM
[0751] 2-巯基乙醇	Invitrogen/Life Technologies	0.1 mM
N2 补充剂	Invitrogen/Life Technologies	1x
B27 补充剂	Invitrogen/Life Technologies	1x
LIF	Millipore	100 U/ml
PD0325901(MEK 抑制剂)	Stemgent	1 uM
CHIR99021(GSK 抑制剂)	Stemgent	3 uM

- [0752] 材料: 妊娠母马的血清促性腺激素 (PMSG)
- [0753] 人类妊娠绒毛膜促性腺激素 (HCG)
- [0754] 雌性大鼠 (5-12周龄)
- [0755] 雄性大鼠 (12周至8月龄), 每笼一只

- [0756] 注射器/针头
- [0757] 动物室在6:00-18:00进行光照
- [0758] 程序:
- [0759] 第1天:8:00-10:00AM
- [0760] 用20IU PMSG(0.4ml)注射雌性,IP
- [0761] 丢弃未使用的PMSG。
- [0762] 第3天:8:00-10:00AM(在PMSG注射之后48小时)
- [0763] 用50IU HCG(1ml)注射雌性,IP
- [0764] 在交配笼中每个雄性放置一个雌性。
- [0765] 丢弃未使用的HCG。
- [0766] 第4天:8:00-10:00AM(在HCG注射之后24小时)
- [0767] 检查雌性的栓塞。
- [0768] 激素供应商
- [0769] PMSG:Sigma#G-4877(1000IU)。在PBS中再悬浮到最终[ ]50IU/ml。以1ml等分试样在-20°下储存。
- [0770] HCG:Sigma#CG-5(5000IU)。在PBS中再悬浮到最终[ ]50IU/ml。以1ml等分试样在-20°下储存。
- [0771] 1.4.:大鼠胚胎干细胞系的核型分析
- [0772] 分析本文产生的大鼠ES细胞系的核型,且结果概述在表4-7中。
- [0773] 表4

ACLG1 核型分析结果	细胞的数目
分析核型的细胞的数目	7
分析的细胞的数目	20
42, XY 细胞的数目	18
异常细胞的数目	2
40, XY,-5,-9	1
41, XY,-14	1
42, XY	18
其它注解: 两个分析的细胞丢失不同的常染色体,这可由于技术性人为因素而偶尔发生。90%的分析细胞具有正常雄性 42, XY 核型。	
图 5 提供示出 ACLG1 大鼠 ES 细胞系的染色体数的分析的照片。	

- [0774]
- [0775] 表5

<b>DA.2B 核型分析结果</b>		<b>细胞的数目</b>
[0776]	分析核型的细胞的数目	6
	分析的细胞的数目	20
	42, XY 细胞的数目	20
	异常细胞的数目	0
	42, XY	20
	其它注解: 所有分析的细胞具有正常二倍体 42, XY 核型。	
	图 6 提供示出 DA.2B 大鼠 ES 细胞系的染色体数的分析的照片。	

[0777] 表6

<b>DA.2C 核型分析结果</b>		<b>细胞的数目</b>
[0778]	分析核型的细胞的数目	5
	分析的细胞的数目	20
	42, XX 细胞的数目	20
	异常细胞的数目	0
	42, XX	
	其它注解: 100%的分析的细胞具有正常雌性 XX 大鼠核型。	
	图 7 提供示出 DA.2C 大鼠 ES 细胞系的染色体数的分析的照片。	

[0779] 表7

株	涂铺的胚泡	确定的株系	分析核型的株系	核型	
[0780]	BN x SD F1	41	8 (20%)	5	所有株系都具有高复合多倍体%
	ACI	27	16 (60%)	3	G1: 90% 42 XY; 其它是 70-85%整倍体
	DA	20	9 (45%)	6	2B: 100% 42 XY; 2C: 100% 42 XX; 其它是 95-100%整倍体
	F344	4	1 (25%)	0	
	<b>总计</b>	92	34 (37%)		

[0781] 1.5.:载体电穿孔到大鼠胚胎干细胞中

[0782] 1. 在电穿孔之前24-48小时使大鼠ES细胞传代。

[0783] 2. 在电穿孔之前24小时将培养基更换成RVG2i+ROCKi (10 $\mu$ M Y-27632)。

[0784] 3. 在胰蛋白酶处理之前30' 更换培养基。

[0785] 4. 将待电穿孔的DNA等分。

[0786] 5. 使DNA在室温下温热&gt;10分钟。

[0787] 6. 在62 $^{\circ}$ C下加热DNA 5'。将DNA放置在冰上。

[0788] 7. 用胰蛋白酶处理细胞:

- [0789] a. 收集浮动集落。将板洗涤以收集尽可能多的浮动物。
- [0790] b. 使集落粒化:在750rpm下3'。
- [0791] c. 用5-10ml PBS洗涤粒料1次,并且再旋转/粒化。
- [0792] d. 吸出上清液;加入500 $\mu$ l胰蛋白酶、0.05%+1%鸡血清。
- [0793] i. 每个管不要汇集超过1个10cm板的集落。如果在胰蛋白酶处理期间有太多集落堆积到管的底部,那么它们将结块,并且大多数细胞将被损失。
- [0794] e. 在37°下4'。吸移集落数次以使结块最小。
- [0795] f. 重复步骤1-2次:在37°下4'。
- [0796] g. 用500 $\mu$ lRVG2i+10%FBS终止胰蛋白酶。
- [0797] 8. 使细胞粒化:在1200rpm下5'。
- [0798] 9. 使细胞在10ml PBS中再悬浮。对两个20 $\mu$ l等分试样计数以测定总细胞数目。
- [0799] 10. 使细胞粒化(5'/1200rpm);计算总细胞数目和总再悬浮体积以获得正确的细胞浓度(靶数目/75 $\mu$ l EP缓冲液)。
- [0800] 11. 在最小体积的EP缓冲液中再悬浮;测量总体积,并用EP缓冲液调整至靶体积。电穿孔缓冲液由Millipore销售。目录号是ES-003-D。参见Valenzuela等人,(2003)Nature Biotechnology 21:652-659,其通过引用的方式并入本文中。
- [0801] 12. 将75 $\mu$ l细胞加到50 $\mu$ lDNA中;将125 $\mu$ l细胞/DNA溶液转移到BTX 48孔比色皿的一个孔中。
- [0802] a. 用125 $\mu$ lEP缓冲液填充在同一列中的空孔。
- [0803] 13. 在BTX电穿孔仪中对比色皿施加脉冲一次:
- [0804] a. 设置:400V; $\Omega$ ;100 $\mu$ F(设置可变化)
- [0805] 14. 将比色皿放置在冰上15'以便恢复。
- [0806] 15. 将细胞移除到5ml RVG2i+10 $\mu$ M ROCKi中。
- [0807] 16. 添加到具有20ml RVG2i+10 $\mu$ M ROCKi的15cm板中。板具有2x neoR MEF(或其它MEF,视项目而定)。neoR可选择标记为Beck等人,(1982)Gene,19:327-36的或在美国专利号7,205,148或6,596,541中的新霉素磷酸转移酶(neo)基因,所述文献和专利各自以引用的方式并入本文中。
- [0808] 17. 在37°下孵育。48小时后开始选择。
- [0809] 所使用的ROCK抑制剂为Y-27632。
- [0810] 1.6:选择在大鼠胚胎干细胞中的靶向基因修饰。
- [0811] 1. 在电穿孔之前24-48小时使细胞传代。
- [0812] 2. 在电穿孔之前24小时将培养基更换成RVG2i+ROCKi(10 $\mu$ M Y-27632)。
- [0813] 3. 在胰蛋白酶处理之前30'更换培养基。
- [0814] 4. 将待电穿孔的DNA等分。
- [0815] 5. 使DNA在室温下温热>10分钟。
- [0816] 6. 在62°C下加热DNA 5'。将DNA放置在冰上。
- [0817] 7. 用胰蛋白酶处理细胞:
- [0818] a. 收集浮动集落。将板洗涤以收集尽可能多的浮动物。
- [0819] b. 使集落粒化:在750rpm下3'。

- [0820] c. 用5-10ml PBS洗涤粒料1次,并且再旋转/粒化。
- [0821] d. 吸出上清液;加入500 $\lambda$ 胰蛋白酶、0.05%+1%鸡血清。
- [0822] i. 每个管不要汇集超过1个10cm板的集落。如果在胰蛋白酶处理期间有太多集落堆积到管的底部,那么它们将结块,并且大多数细胞将被损失。
- [0823] e. 在37°下4'。吸移集落数次以使结块最小。
- [0824] f. 重复步骤1-2次:在37°下4'。
- [0825] g. 用500 $\lambda$ RVG2i+10%FBS终止胰蛋白酶。
- [0826] 8. 使细胞粒化:在1200rpm下5'。
- [0827] 9. 使细胞在10ml PBS中再悬浮。对两个20 $\lambda$ 等分试样计数以测定总细胞数目。
- [0828] 10. 使细胞粒化(5'/1200rpm);计算总细胞数目和总再悬浮体积以获得正确的细胞浓度(靶数目/75 $\mu$ l EP缓冲液)。
- [0829] 11. 在最小体积的EP缓冲液中再悬浮;测量总体积,并用EP缓冲液调整至靶体积。
- [0830] 12. 将75 $\lambda$ 细胞加到50 $\lambda$ DNA中;将125 $\lambda$ 细胞/DNA溶液转移到BTX 48孔比色皿的一个孔中。
- [0831] a. 用125 $\lambda$ EP缓冲液填充在同一列中的空孔。
- [0832] 13. 在BTX电穿孔仪中对比色皿施加脉冲一次:
- [0833] a. 设置:400V;100 $\mu$ F(设置可变化)。
- [0834] 14. 将比色皿放置在冰上15'以便恢复。
- [0835] 15. 将细胞移除到5ml RVG2i+10 $\mu$ M ROCKi中。
- [0836] 16. 添加到具有20ml RVG2i+10 $\mu$ M ROCKi的15cm板中。板具有2x neoR MEF(或其它MEF,视项目而定)。
- [0837] 17. 在37°下孵育。48小时后开始选择。
- [0838] 18. G418选择方案如下:
- [0839] a. 第2天(在EP之后第2天):在2i培养基+75 $\mu$ g/ml G418中孵育细胞。
- [0840] b. 第3天:在无G418的2i培养基中孵育细胞。
- [0841] c. 第4天:在2i培养基+75 $\mu$ g/ml G418中孵育细胞。
- [0842] d. 第5天:在无G418的2i培养基中孵育细胞。
- [0843] e. 第6天:在2i培养基+75 $\mu$ g/ml G418中孵育细胞。
- [0844] f. 第7天:在无G418的2i培养基中孵育细胞。
- [0845] g. 第8天:在2i培养基+75 $\mu$ g/ml G418中孵育细胞。
- [0846] h. 第9天:在无G418的2i培养基中孵育细胞。
- [0847] i. 第10天:在2i培养基+75 $\mu$ g/ml G418中孵育细胞。
- [0848] j. 第11天:在无G418的2i培养基中孵育细胞。
- [0849] k. 第12天:挑选集落以进行扩增以便筛选。各集落在0.05%胰蛋白酶+1%鸡血清中离散10分钟,且随后涂铺到96孔饲养板的1个孔中。
- [0850] 19. 在2i培养基中扩增集落3天。
- [0851] 20. 以1:1使克隆传代至新96孔饲养板中。
- [0852] 21. 在2i培养基中扩增克隆3天。
- [0853] 22. 对于各克隆,在胰蛋白酶中离散集落。冷冻2/3的各克隆,并在-80°下储存;将

剩余1/3涂铺到层粘蛋白板(用10 $\mu$ g/ml层粘蛋白涂布的96孔板)上。

[0854] 23.当层粘蛋白板汇合时,传至筛选实验室以分析克隆的基因分型。

[0855] 1.7.大鼠胚胎干细胞的分子标签

[0856] 在表8中所列的基因在大鼠ES细胞中比相应基因在小鼠ES细胞中的表达低20倍。在表9中所列的基因在大鼠ES细胞中比相应基因在小鼠ES细胞中的表达水平高20倍。

[0857] 如下产生在表8和9中的微阵列数据。在2i培养基中培养大鼠ES细胞(AC1.G2和DA.2B)和小鼠ES细胞(F1H4)3代直至汇合。在不存在饲养器的情况下在明胶涂布的板上培养F1H4细胞。F1H4小鼠ES细胞来源于129S6/SvEvTac和C57BL/6NTac杂合胚胎(参见,例如美国专利号7,294,754和Poueymirou,W.T.,Auerbach,W.,Frendewey,D.,Hickey,J.F.,Escaravage,J.M.,Esau,L.,Dore,A.T.,Stevens,S.,Adams,N.C.,Dominguez,M.G.,Gale,N.W.,Yancopoulos,G.D.,DeChiara,T.M.,Valenzuela,D.M.(2007),其以引用的方式并入本文中)。

[0858] 以下方案用于样品制备:1.5mL Eppendorf管用样品识别符标记。将在板上生长的细胞在37 $^{\circ}$ C磷酸盐缓冲盐水(PBS)中冲洗。移除PBS且加入300u1 **Trizol**<sup>®</sup>。使用刮削器以破坏在**Trizol**<sup>®</sup>(Life Technology)中的细胞。将裂解的细胞收集于1.5mL Eppendorf管中的**Trizol**<sup>®</sup>。对于悬浮生长的细胞,在37 $^{\circ}$ C PBS中冲洗细胞,并且收集在1.5mL管中。使细胞旋转减慢;移除PBS;并将300u1 **Trizol**<sup>®</sup>加到细胞中。细胞膜通过吸移来破坏。分选样品以用于以10 $\cdot$ 10<sup>5</sup>个细胞进行FACS,浓缩体积至小于100 $\mu$ L。加入4体积的RNA裂解缓冲液,并且通过吸移来混合。对于样品,将320 $\mu$ L RNA裂解缓冲液加到80 $\mu$ L样品中。将样品在-20 $^{\circ}$ C下储存。

[0859] 使用RNA-Seq以测量小鼠和大鼠基因的表达水平。通过Tophat来使测序读数定位到小鼠和大鼠参照基因组,并且计算小鼠和大鼠基因的RPKM(定位的每百万片段中外显子的每千碱基的片段数)。选择基于基因符号的同源性基因,且随后使用t检验来比较在小鼠与大鼠之间各基因的表达水平。miR-32处于大鼠ESC中最高表达的前10名,但不在小鼠ES细胞中表达。尽管不存在来自miR-632的比较数据,但基于它相较于在大鼠ESC中表达的其它基因的表达水平以及它们在胚胎发育方面的已知功能,选择miR-632作为大鼠ES细胞的标记。

[0860] 表8.所列基因在大鼠ES细胞中比相应基因在小鼠ES细胞中的表达水平低20倍。



[0861]

识别符	符号	Entrez 基因名称	位置	类型
Abcb1b	Abcb1b	ATP 结合盒, 亚家族 B(MDR/TAP), 成员 1B	质膜	转运体
Acta2	ACTA2	主动脉平滑肌 $\alpha$ 2 肌动蛋白	细胞质	其它
Actg2	ACTG2	肠道平滑肌 $\gamma$ 2 肌动蛋白	细胞质	其它
Aebp1	AEBP1	AE 结合蛋白 1	核	肽酶
Angptl2	ANGPTL2	促血管生成素样 2	细胞外间隙	其它
Ankrd1	ANKRD1	锚蛋白重复结构域 1(心肌)	细胞质	转录调控子
Anxa1	ANXA1	膜联蛋白 A1	质膜	其它
Anxa6	ANXA6	膜联蛋白 A6	质膜	其它
Anxa8	ANXA8L2	膜联蛋白 A8-样 2	质膜	其它
Arhgef25	ARHGEF25	Rho 鸟嘌呤鸟苷酸核苷酸交换因子 (GEF) 25	细胞质	其它
Axl	AXL	AXL 受体酪氨酸激酶	质膜	激酶
Basp1	BASP1	脑中充足的膜连接信号蛋白 1	核	转录调控子
Bgn	BGN	双糖链蛋白聚糖	细胞外间隙	其它
Bst2	BST2	骨髓基质细胞抗原	质膜	其它

[0862]

识别符	符号	Entrez 基因名称	位置	类型
		2		
Btf3	BTF3	碱性转录因子 3	核	转录调控子
Btg2	BTG2	BTG 家族, 成员 2	核	转录调控子
Capsl	CAPSL	钙磷蛋白样	其它	其它
Cav1	CAV1	窖蛋白 1, 胞膜窖蛋白, 22 kDa	质膜	跨膜受体
Ccdc80	CCDC80	含有卷曲螺旋结构域的 80	核	其它
Cend2	CCND2	周期素 D2	核	其它
Cd248	CD248	CD248 分子内皮唾酸蛋白	质膜	其它
Cd44	CD44	CD44 分子(印度血型)	质膜	酶
Cd97	CD97	CD97 分子	质膜	G 蛋白偶联受体
Cdc42ep5	CDC42EP5	CDC42 效应子蛋白(Rho GTPase 结合)5	细胞质	其它
Cdh11	CDH11	钙粘蛋白 11, 2 型, OB-钙粘蛋白(成骨细胞)	质膜	其它
Cdkn2a	CDKN2A	周期素依赖性激酶抑制剂 2A	核	转录调控子
Cdo1	CDO1	1 型半胱氨酸双加氧酶	细胞质	酶
Clip3	CLIP3	含有连接子蛋白 3 的 CAP-GLY 结构域	细胞质	其它
Cln5	CLN5	神经元蜡样脂褐素沉积病 5	细胞质	其它
Cnn1	CNN1	平滑肌碱性钙调蛋白 1	细胞质	其它
Col1a1	COL1A1	$\alpha 1$ I 型胶原蛋白	细胞外间隙	其它
Col1a2	COL1A2	$\alpha 2$ I 型胶原蛋白	细胞外间隙	其它
Col3a1	COL3A1	$\alpha 1$ III 型胶原蛋白	细胞外间隙	其它
Col5a2	COL5A2	$\alpha 2$ V 型胶原蛋白	细胞外间隙	其它
Col6a2	COL6A2	$\alpha 2$ VI 型胶原蛋白	细胞外间隙	其它
Cryab	CRYAB	$\alpha$ B 晶体蛋白	核	其它
Csfl	CSF1	集落刺激因子 1(巨噬细胞)	细胞外间隙	细胞因子
Cth	CTH	胱硫醚酶(胱硫醚 $\gamma$ -裂解酶)	细胞质	酶

[0863]

识别符	符号	Entrez 基因名称	位置	类型
Cthrc1	CTHRC1	含有胶原蛋白三螺旋重复的 1	细胞外间隙	其它
Ctsc	CTSC	组织蛋白酶 C	细胞质	肽酶
Cyr61	CYR61	富含半胱氨酸的血管生成诱导因子, 61	细胞外间隙	其它
Ddx58	DDX58	DEAD(Asp-Glu-Ala-Asp)框多肽 58	细胞质	酶
Dkk3	DKK3	dickkopf WNT 信号传导通路抑制剂 3	细胞外间隙	细胞因子
Dmcl	DMC1	DNA 减数分裂重组酶 1	核	酶
Dpysl3	DPYSL3	二氢嘧啶酶样 3	细胞质	酶
Dse	DSE	硫酸皮肤素差向异构酶	细胞质	酶
Dusp1	DUSP1	双重特异性磷酸酯酶 1	核	磷酸酶
Dusp27	DUSP27	双重特异性磷酸酯酶 27(推定)	其它	磷酸酶
Dusp9	DUSP9	双重特异性磷酸酯酶 9	核	磷酸酶
Ece2	ECE2	内皮素转化酶 2	质膜	肽酶
Ecm1	ECM1	细胞外基质蛋白 1	细胞外间隙	转运体
Egr1	EGR1	早期生长应答 1	核	转录调控子
Emp1	EMP1	上皮膜蛋白 1	质膜	其它
Emp3	EMP3	上皮膜蛋白 3	质膜	其它
Ephx2	EPHX2	细胞质环氧化物水解酶 2	细胞质	酶
F3	F3	凝血因子 III(促凝血酶原激酶, 组织因子)	质膜	跨膜受体
Fau	FAU	遍在表达的 Finkel-Biskis-Reilly 鼠科肉瘤病毒 (FBR-MuSV)	细胞质	其它
Fbn1	FBN1	肌原纤维蛋白 1	细胞外间隙	其它
Fbxo15	FBXO15	F 框蛋白 15	其它	转录调控子
Fhl2	FHL2	四个半 LIM 结构域 2	核	转录调控子
Flnc	FLNC	$\gamma$ 细丝蛋白 C	细胞质	其它
Fos	FOS	FBJ 鼠科骨肉瘤病毒致癌基因同源物	核	转录调控子

[0864]

识别符	符号	Entrez 基因名称	位置	类型
Fundc2	FUNDC2	含有 FUN14 结构域的 2	细胞质	其它
Gjb3	GJB3	间隙连接蛋白 $\beta$ 3, 31 kDa	质膜	转运体
Gpa33	GPA33	糖蛋白 A33(跨膜)	质膜	其它
Gpbp111	GPBP1L1	富含 GC 的启动子结合蛋白 1-样 1	其它	其它
Gpc3	GPC3	磷脂酰肌醇蛋白聚糖 3	质膜	其它
Grb10	GRB10	生长因子受体结合蛋白 10	细胞质	其它
Gstm1	GSTM5	谷胱甘肽 S-转移酶 $\mu$ 5	细胞质	酶
Hap1	HAP1	亨廷顿蛋白相关蛋白 1	细胞质	其它
Hist1h2bc	HIST2H2BE(包括其它)	组蛋白簇 2, H2be	核	其它
Hmga2	HMGA2	高迁移率组 AT-钩 2	核	酶
Hmgn3	Hmgn3	高迁移率组核小体结合结构域 3	核	其它
Hormad1	HORMAD1	含有 HORMA 结构域的 1	核	其它
Hsd17b14	HSD17B14	羟基类固醇(17- $\beta$ )脱氢酶 14	细胞质	酶
Hspb1	HSPB1	热击 27 kDa 蛋白 1	细胞质	其它
Hspb8	HSPB8	热击 22 kDa 蛋白 8	细胞质	激酶
Htra1	HTRA1	Htra 丝氨酸蛋白酶 1	细胞外间隙	肽酶
Ifi204	Ifi204(包括其它)	干扰素活化的基因 204	核	转录调控子
Ifi44	IFI44	干扰素诱导的蛋白 44	细胞质	其它
Ifit1	IFIT1B	具有三十四肽重复的干扰素诱导的蛋白 1B	细胞质	其它
Ifitm3	IFITM2	干扰素诱导的跨膜蛋白 2	细胞质	其它
Igf2	IGF2	胰岛素样生长因子 2(生长调节素 A)	细胞外间隙	生长因子
Igfbp7	IGFBP7	胰岛素样生长因子结合蛋白 7	细胞外间隙	转运体
Il1rl1	IL1RL1	白细胞介素 1 受体	质膜	跨膜受体

[0865]

识别符	符号	Entrez 基因名称	位置	类型
		样 1		
Inhba	INHBA	抑制素 $\beta$ A	细胞外间隙	生长因子
Inhbb	INHBB	抑制素 $\beta$ B	细胞外间隙	生长因子
Irf7	IRF7	干扰素调控因子 7	核	转录调控子
Isg15	ISG15	ISG15 泛素样修饰剂	细胞外间隙	其它
Itga5	ITGA5	整合素 $\alpha$ 5(粘连蛋白受体, $\alpha$ 多肽)	质膜	跨膜受体
Jun	JUN	jun 原致癌基因	核	转录调控子
Junb	JUNB	jun B 原癌基因	核	转录调控子
Lgals3bp	LGALS3BP	半乳糖苷结合性可溶性凝集素 3 结合蛋白	质膜	跨膜受体
Lgals9	LGALS9	半乳糖苷结合性可溶性凝集素 9	细胞外间隙	其它
Lmna	LMNA	核纤层蛋白 A/C	核	其它
Lox	LOX	赖氨酰氧化酶	细胞外间隙	酶
Loxl2	LOXL2	赖氨酰氧化酶样 2	细胞外间隙	酶
Loxl3	LOXL3	赖氨酰氧化酶样 3	细胞外间隙	酶
Lrp1	LRP1	低密度脂蛋白受体相关蛋白 1	质膜	跨膜受体
Mageb16	MAGEB16	黑色素瘤抗原家族 B, 16	其它	其它
Mcam	MCAM	黑素瘤细胞粘附分子	质膜	其它
Mgp	MGP	基质 Gla 蛋白	细胞外间隙	其它
Mmp2	MMP2	基质金属肽酶 2(明胶酶 A, 72 kDa 明胶酶, 72 kDa IV 型胶原酶)	细胞外间隙	肽酶
Mxra8	MXRA8	基质重塑相关的 8	其它	其它
Myl9	MYL9	调控性肌球蛋白轻链 9	细胞质	其它
Mylpf	MYLPF	快速骨骼肌可磷酸化的肌球蛋白轻链	细胞质	其它
Nab2	NAB2	NGFI-A 结合蛋白 2(EGR1 结合蛋白 2)	核	转录调控子
Ndufb4	NDUFB4	NADH 脱氢酶(泛醌) $1\beta$ 亚复合物, 4, 15 kDa	细胞质	转运体
Npm1	NPM1	核磷酸蛋白(核仁磷蛋白 B23, 核基	核	转录调控子

识别符	符号	Entrez 基因名称	位置	类型
		质蛋白)		
Nr0b1	NR0B1	核受体亚家族 0, B 组, 成员 1	核	配体依赖性核受体
Nr4a1	NR4A1	核受体亚家族 4, A 组, 成员 1	核	配体依赖性核受体
Nrp2	NRP2	神经纤毛蛋白 2	质膜	激酶
Oas1a	OAS1	2'-5'-寡腺苷酸合成酶 1, 40/46 kDa	细胞质	酶
Oasl2	Oasl2	2'-5'寡腺苷酸合成酶样 2	其它	酶
P4ha2	P4HA2	脯氨酰基 4-羟化酶 $\alpha$ 多肽 II	细胞质	酶
Parp3	PARP3	聚(ADP-核糖)聚合酶家族, 成员 3	核	酶
Pcolce	PCOLCE	原胶原 C-内肽酶增强子	细胞外间隙	其它
Pcyt1b	PCYT1B	磷酸胆碱胞苷酰基转移酶 1 $\beta$	细胞质	酶
Pdgfc	PDGFC	血小板源性生长因子 C	细胞外间隙	生长因子
Phlda1	PHLDA1	普列克底物蛋白同源样结构域, 家族 A, 成员 1	细胞质	其它
Phlda2	PHLDA2	普列克底物蛋白同源样结构域, 家族 A, 成员 2	细胞质	其它
Pla2g1b	PLA2G1B	IB 组磷脂酶 A2(胰腺)	细胞外间隙	酶
Pla2g4a	PLA2G4A	IVA 组磷脂酶 A2(胞质, 钙依赖性)	细胞质	酶
Porcn	PORCN	豪猪同源物(果蝇)	细胞质	其它
Postn	POSTN	骨膜蛋白, 成骨细胞特异性因子	细胞外间隙	其它
Prrx1	PRRX1	成对的相关同源框 1	核	转录调控子
Prss23	PRSS23	丝氨酸蛋白酶 23	细胞外间隙	肽酶
Psmb8	PSMB8	$\beta$ 型蛋白酶体(前体, 巨蛋白因子)亚基 8	细胞质	肽酶
Ptgs2	PTGS2	前列腺素内过氧化物合酶 2(前列腺素 G/H 合成酶和环氧酶)	细胞质	酶

[0866]

[0867]

识别符	符号	Entrez 基因名称	位置	类型
Ptn	PTN	多效生长因子	细胞外间隙	生长因子
Ptrf	PTRF	聚合酶 I 和转录体释放因子	核	转录调控子
Rarg	RARG	视黄酸受体 $\gamma$	核	配体依赖性核受体
Rgs16	RGS16	G 蛋白信号传导调控子 16	细胞质	其它
Rn45s	Rn45s	45S 前核糖体 RNA	其它	其它
Rpl10a	RPL10A	核糖体蛋白 L10a	其它	其它
Rpl31	RPL31	核糖体蛋白 L31	其它	其它
Rpl37a	RPL37A	核糖体蛋白 L37a	细胞质	其它
Rps10	RPS10-N UDT3	RPS10-NUDT3 通读蛋白	细胞质	其它
Rps14	RPS14	核糖体蛋白 S14	细胞质	转译调控子
Rps20	Rps20	核糖体蛋白 S20	细胞质	其它
Rps26	RPS26	核糖体蛋白 S26	细胞质	其它
Rps9	RPS9	核糖体蛋白 S9	细胞质	转译调控子
S100a4	S100A4	S100 钙结合蛋白 A4	细胞质	其它
S100a6	S100A6	S100 钙结合蛋白 A6	细胞质	转运体
Schip1	SCHIP1	神经鞘瘤素相互作用蛋白 1	细胞质	其它
Sdc2	SDC2	粘结蛋白聚糖 2	质膜	其它
Serpine1	SERPIN E1	丝氨酸蛋白酶肽酶抑制剂, 分化体 E(微管连接蛋白, 1 型纤溶酶原激活物抑制剂), 成员 1	细胞外间隙	其它
Serpine2	SERPIN E2	丝氨酸蛋白酶肽酶抑制剂, 分化体 E(微管连接蛋白, 1 型纤溶酶原激活物抑制剂), 成员 2	细胞外间隙	其它
Serpinf1	SERPIN F1	丝氨酸蛋白酶肽酶抑制剂, 分化体 F( $\alpha$ -2 抗纤溶酶, 色素上皮源性因子), 成员 1	细胞外间隙	其它
Sh3gl2	SH3GL2	SH3-结构域 GRB2 样 2	质膜	酶
Slc19a2	SLC19A 2	溶质载体家族 19(硫胺转运体), 成员	质膜	转运体

识别符	符号	Entrez 基因名称	位置	类型
		2		
Slc25a5	SLC25A5	溶质载体家族 25(线粒体载体; 腺嘌呤核苷酸易位体), 成员 5	细胞质	转运体
Slc29a1	SLC29A1	溶质载体家族 29(平衡核苷转运体), 成员 1	质膜	转运体
Slc35f2	SLC35F2	溶质载体家族 35, 成员 F2	其它	其它
Snrpn	SNRPN	小核核糖核酸蛋白多肽 N	核	其它
Snx22	SNX22	分选微管连接蛋白 22	其它	转运体
Sparc	SPARC	酸性富含半胱氨酸的分泌蛋白(骨粘连蛋白)	细胞外间隙	其它
Spp1	SPP1	分泌的磷蛋白 1	细胞外间隙	细胞因子
Sult4a1	SULT4A1	磺基转移酶家族 4A, 成员 1	细胞质	酶
[0868] Tagln	TAGLN	转凝蛋白	细胞质	其它
Tcea3	TCEA3	转录延伸因子 A(SII), 3	核	转录调控子
Tgfb3	TGFB3	转化生长因子, $\beta$ 3	细胞外间隙	生长因子
Thbs1	THBS1	凝血酶敏感蛋白 1	细胞外间隙	其它
Thbs2	THBS2	凝血酶敏感蛋白 2	细胞外间隙	其它
Tm4sf1	TM4SF1	跨膜 4L6 家族成员 1	质膜	其它
Tmbim1	TMBIM1	含有跨膜 BAX 抑制剂基序的 1	细胞质	其它
Tmem176b	TMEM176B	跨膜蛋白 176B	其它	其它
Tnc	TNC	腱生蛋白 C	细胞外间隙	其它
Tpd52l1	TPD52L1	肿瘤蛋白 D52 样 1	细胞质	其它
Tpm2	TPM2	原肌球蛋白 2( $\beta$ )	细胞质	其它
Usp18	USP18	泛素特异性肽酶 18	细胞质	肽酶
Vim	VIM	波形蛋白	细胞质	其它
Wfdc2	WFDC2	WAP 四-二硫化物核心结构域 2	细胞外间隙	其它
Wisp2	WISP2	WNT1 可诱导的信号传导途径蛋白 2	细胞外间隙	生长因子
[0869] 识别符	符号	Entrez 基因名称	位置	类型
Ybx1	YBX1	Y 框结合蛋白 1	核	转录调控子



[0870] 表9.所列基因在大鼠ES细胞中比相应基因在小鼠ES细胞中的表达水平高20倍。

识别符	符号	Entrez 基因名称	位置	类型
Ajap1	Ajap1	粘附蛋白接合相关蛋白 1	其它	其它
Amd1	AMD1	腺苷甲硫氨酸脱羧酶 1	细胞质	酶
Ankrd2	ANKRD2	锚蛋白重复结构域 2(拉伸应答性肌肉)	核	转录调控子
Arhgef9	ARHGEF9	Cdc42 鸟苷酸核苷酸交换因子(GEF) 9	细胞质	其它
Atp5h	Atp5h	ATP 合酶, H <sup>+</sup> 转运线粒体, F <sub>0</sub> 复合体, 亚基 d	细胞质	酶
Btg3	BTG3	BTG 家族, 成员 3	核	其它
Car6	CA6	碳酸酐酶 VI	细胞外间隙	酶
Camk4	CAMK4	钙/钙调蛋白依赖性蛋白激酶 IV	核	激酶
Capn12	CAPN12	钙蛋白酶 12	其它	肽酶
Cct6b	CCT6B	含有伴侣蛋白的 TCP1, 亚基 6B(ξ2)	细胞质	转运体
Cdx2	CDX2	尾型同源框 2	核	转录调控子
Cldn5	CLDN5	密封蛋白 5	质膜	其它
Clec3a	CLEC3A	C 型凝集素结构域家族 3, 成员 A	其它	其它
Clic6	CLIC6	氯化细胞内通道 6	质膜	离子通道
Dhrsx	DHRSX	X-连锁脱氢酶/还原酶 (SDR 家族)	其它	酶
Dpysl2	DPYSL2	二氢嘧啶酶样 2	细胞质	酶
Dusp26	DUSP26	双重特异性磷酸酯酶 26(推定)	细胞质	酶
Eci3	Eci3	烯酰基-辅酶 Aδ 异构酶 3	其它	酶
Eef2k	EEF2K	真核延伸因子-2 激酶	细胞质	激酶
Efna1	EFNA1	肝配蛋白-A1	质膜	其它
Epha4	EPHA4	EPH 受体 A4	质膜	激酶

[0871]

[0872]

识别符	符号	Entrez 基因名称	位置	类型
Fank1	FANK1	III 型纤连蛋白和锚蛋白重复结构域 1	核	转录调控子
Fhit	FHIT	脆弱的组氨酸三联体	细胞质	酶
Filip1	FILIP1	细丝蛋白 A 相互作用蛋白 1	细胞质	其它
Fmod	FMOD	纤调蛋白聚糖	细胞外间隙	其它
Foxe1	FOXE1	叉头框 E1(甲状腺转录因子 2)	核	转录调控子
Fry	FRY	毛皮同源物(果蝇)	细胞外间隙	其它
Gjb5	GJB5	间隙连接蛋白 $\beta 5$ , 31.1 kDa	质膜	转运体
Gpx2	GPX2	谷胱甘肽过氧化物酶 2(胃肠)	细胞质	酶
Grxcr2	GRXC R2	富含半胱氨酸的谷氧还蛋白 2	其它	其它
Hecw2	HECW 2	含有 HECT、C2 和 WW 结构域的 E3 泛素蛋白连接酶 2	细胞外间隙	酶
Hey2	HEY2	与 YRPW 基序相关的毛状/分裂增强子 2	核	转录调控子
Icos	Icos	可诱导的 T 细胞共刺激剂	质膜	其它
Ifitm1	IFITM1	干扰素诱导的跨膜蛋白 1	质膜	跨膜受体
Il1f8	IL1F8(I L36B)	白细胞介素-1 家族成员(白细胞介素 36 $\beta$ )	细胞外间隙	细胞因子
Il28ra	IL-28R A	白细胞介素 28 受体 $\alpha$	质膜	细胞因子受体
Igfbp1	IGFBP L1	胰岛素样生长因子结合蛋白样 1	其它	其它
Ipcef1	IPCEF1	细胞粘附蛋白交换因子相互作用蛋白 1	细胞质	酶
Lctl	Lctl	乳糖酶样	细胞质	其它
Ldhd	LDHD	乳酸脱氢酶 D	细胞质	酶
Lef1	LEF1	淋巴增强子结合因子 1	核	转录调控子
Lefty1	LEFTY 1	左-右决定因子 1	细胞外间隙	生长因子
Lifr	LIFR	白血病抑制因子受体 $\alpha$	质膜	跨膜受

[0873]

识别符	符号	Entrez 基因名称	位置	类型
				体
Lpar2	LPAR2	溶血磷脂酸受体 2	质膜	G 蛋白偶联受体
Mog	MOG	髓鞘质少突胶质细胞糖蛋白	细胞外间隙	其它
Morn5	MORN5	含 MORN 重复序列的 5	其它	其它
Pigz	NCBP2	核帽结合蛋白亚基 2, 20 kDa	核	其它
Nptxr	NPTXR	神经元穿透素受体	质膜	跨膜受体
Ntm	NTM	神经营养因子	质膜	其它
Nutf2	NUTF2	核转运因子 2	核	转运体
Ocln	OCLN	封闭蛋白	质膜	酶
Olr1	OLR1	氧化低密度脂蛋白(凝集素样)受体 1	质膜	跨膜受体
Pabpc4	PABPC4	细胞质多聚(A)结合蛋白 4(可诱导形式)	细胞质	转译调控子
Pde11a	PDE11A	磷酸二酯酶 11A	细胞质	酶
Pdyn	PDYN	强啡肽原	细胞外间隙	转运体
Per3	PER3	时间昼夜节律钟 3	核	其它
Plip	PLLP	浆脂蛋白	质膜	转运体
Ppp1r14c	PPP1R14C	蛋白磷酸酯酶 1 调控性(抑制剂)亚基 14C	细胞质	其它
Pramel6	PRAMEL6	在黑素瘤中优先表达的抗原样 6	其它	其它
Ptpn18	PTPN18	非受体型蛋白酪氨酸磷酸酯酶 18(脑源性)	核	磷酸酶
Pycr1	PYCR1	吡咯啉-5-羧酸还原酶 1	细胞质	酶
Rab26	RAB26	RAB26、RAS 致癌基因家族成员	质膜	酶
Ramp2	RAMP2	受体(G 蛋白偶联)活性修饰蛋白 2	质膜	转运体
Rbm24	RBM24	RNA 结合基序蛋白 24	其它	其它
Rhag	RHAG	Rh 相关糖蛋白	质膜	肽酶
Rpl3	RPL3	核糖体蛋白 L3	细胞质	其它
Sall3	SALL3	sal 样 3(果蝇)	核	其它
Satb1	SATB1	SATB 同源框 1	核	转录调控子
Scg2	SCG2	分泌粒蛋白 II	细胞外间隙	细胞因子

[0874]

识别符	符号	Entrez 基因名称	位置	类型
Slc15a1	SLC15A1	溶质载体家族 15(寡肽转运体), 成员 1	质膜	转运体
Slc1a1	SLC1A1	溶质载体家族 1(神经元/上皮高亲和性谷氨酸转运体, 系统 Xag), 成员 1	质膜	转运体
Slc24a5	Slc24a5	溶质载体家族 24(钠/钾/钙交换剂), 成员 5	其它	其它
Slc37a2	SLC37A2	溶质载体家族 37(葡萄糖-6-磷酸转运体), 成员 2	其它	转运体
40424	SNTB1	互养蛋白 $\beta$ 1(肌萎缩蛋白相关的蛋白 A1, 59 kDa, 基本组分 1)	质膜	其它
St6galnac3	ST6GALNAC3	ST6( $\alpha$ -N-乙酰基-神经氨基-2,3- $\beta$ -半乳糖基-1,3)-N-乙酰氨基半乳糖苷 $\alpha$ -2,6-唾液酰基转移酶 3	细胞质	酶
Tex12	TEX12	睾丸表达的 12	核	其它
Tex15	TEX15	睾丸表达的 15	细胞外间隙	其它
Tfap2a	TFAP2A	转录因子 AP-2 $\alpha$ (活化增强子结合蛋白 2 $\alpha$ )	核	转录调控子
Tmc1	TMC1	跨膜通道样 1	质膜	其它
Tmem130	TMEM130	跨膜蛋白 130	其它	其它
Tmem30b	TMEM30B	跨膜蛋白 30B	其它	其它
Tomm20	TOMM20	外部线粒体膜 20 同源物的转位酶(酵母)	细胞质	转运体
Tox3	TOX3	TOX 高迁移率组框家族成员 3	其它	其它
Ttc25	TTC25	三十四肽重复结构域 25	细胞质	其它
Tymp	TYMP	胸苷磷酸化酶	细胞外间隙	生长因子
Ubb	Ubb	泛素 B	细胞质	其它
Vamp7	VAMP7	囊泡相关的膜蛋白 7	细胞质	转运体
Wfdc12	Wfdc12	WAP 四-二硫化物核心结构域 12	细胞外间隙	其它
Wfdc15a	Wfdc15a	WAP 四-二硫化物核心结构域 15A	其它	其它

识别符	符号	Entrez 基因名称	位置	类型
[0875] Wfdc6a	Wfdc6a	WAP 四-二硫化物核心 结构域 6A	其它	其它

[0876] 表10. 来自表9的基因亚组, 其在大鼠ES细胞中比相应基因在小鼠ES细胞中的表达水平高20倍。

识别符	Entrez基因名称
Ajap1	粘附蛋白接合相关蛋白
Cldn5	密封蛋白5
Arhgef9	Cdc42鸟苷酸核苷酸交换因子9
Camk4	钙/钙调蛋白依赖性蛋白激酶IV
Efnal	肝配蛋白-A1
Epha4	EPH受体A4
Gjb5	间隙连接蛋白β5
Igfbpl1	胰岛素样生长因子结合蛋白样1
Il1f8	白细胞介素36β
Il28ra	白细胞介素28受体α
Lefty1	左-右决定因子1
Lifr	白血病抑制因子受体α
Lpar2	溶血磷脂酸受体2
Ntm	神经元穿透素受体
Ptpn18	非受体型蛋白酪氨酸磷酸酯酶18
Cdx2	尾型同源框2
Fank1	III型纤连蛋白和锚蛋白重复结构域1
Foxe1	叉头框E1 (甲状腺转录因子2)
Hey2	与YRPW基序相关的毛状/分裂增强子2
Lef1	淋巴增强子结合因子1
Sal13	Sal样3 (果蝇)
Satb1	SATB同源框1

[0878] 还已开发采用大鼠ES细胞的多潜能标记/基因的另一分子标签。表11提供基因清单和它们根据RNA剖析数据的表达分级。mRNA从大鼠ES细胞分离, 并且相对于彼此比较各种标记的表达水平。术语“分级”是指个别基因的比较表达水平: 分级越高(1是最高), 表达越高。例如, Oct4的分级为13是指在所有测定的基因中, 其表达高于除12个基因之外的所有基因。在该实验中的本底为低于30的任何表达值; 6107个基因具有30或更高的表达值。

[0879] 表11. 采用各种多潜能性、中胚层、内胚层、神经和滋养外胚层标记/基因的大鼠ES细胞分子标签。

多潜能性	多潜能性分级	中胚层	中胚层分级	内胚层	内胚层分级	神经	神经分级	滋养外胚层	滋养外胚层分级
c-Myc	8248	短尾畸形	7542	Gata6	11195	巢蛋白	7761	Cdx2	739
Dnmt3L	127	Flk1	未测出	Sox17	11418	Pax6	13570		
Dppa2	未测出	Nodal	3050	Hhex1	4571	Sox2	681		
Dppa5	未测出	Bmp4	3072	Nodal	3050				
Ecat1	9714	Bmpr2	6382	Ext1	6091				
Eras	2541			Sox7	10284				
Err-β	1368								
Fbxo15	1369								
Fgf4	3440								
Fthl17	未测出								
[0880] Gdf3	2771		分级 > 6107 = 本底表达						
Klf4	836								
Lef1	1313								
LIF受体	724								
Lin28	828								
Nanog	774								
Oct4	13								
Rexo1	6119								
Sox15	4524								
Sox2	681								
SSEA1	未测出								
SSEA4	未测出								
Stella	未测出								
Tcl1	未测出								
Utf1	1501								

[0881] 实施例2:使在大鼠中的基因组基因座失活

[0882] 2.1:使用核酸内切酶试剂使内源基因组基因座失活

[0883] 为了在内源大鼠基因组基因座处引入突变等位基因,本文所述的大鼠ES细胞用表达ZFN1和ZFN2(或TALEN1和TALEN2)的表达载体(或mRNA)电穿孔。这些蛋白质在相对链上结合它们的相隔约6bp-约40bp的靶序列。在靶基因座内形成细胞试图通过非同源端接合(NHEJ)修复的双链断裂。在许多情况下,NHEJ导致产生常破坏基因功能(最常通过产生移码突变)的缺失。为了鉴定包含突变等位基因的阳性克隆,以低密度涂铺电穿孔的细胞,因为未进行药物选择。挑选出集落,并且在靶位点处进行测定以观察是否产生突变(例如使用上述等位基因修饰(MOA)测定)。随后将包含突变等位基因的所选择的ES细胞引入宿主大鼠胚

胎(例如,前桑椹胚阶段或胚泡阶段的大鼠胚胎)中,并且植入代孕母体的子宫中以产生建群大鼠(F0大鼠)。接着,使建群大鼠与野生型大鼠配种以产生突变等位基因杂合性F1子代。杂合F1大鼠交配可产生突变等位基因纯合性子代。

[0884] 2.2.:大鼠ESC靶向以使用锌指核酸酶使大鼠载脂蛋白E(ApoE)基因失活

[0885] 锌指核酸酶使用序列特异性模块化DNA结合结构域以将核酸内切酶活性定向到在基因组中的独特靶序列。ZFN工程化成一对单体。各单体含有融合到3个或更多个锌指DNA结合结构域的来自FokI核酸内切酶的非特异性裂解结构域。各锌指结合3bp亚位点,且特异性通过两种单体的组合靶位点来实现。ZFN在DNA中产生双链断裂(DSB),并且突变(插入或缺失)在非同源端接合(NHEJ)期间频繁地发生。图15说明例如ZFN和TALEN的编辑基因组的核酸内切酶在靶基因组序列中引入双链断裂并在细胞中激活NHEJ的机制。如果供体序列提供有ZFN,则DSB也通过同源重组激发同源性定向修复(HDR)。

[0886] 这样的ZFN与本文所述的各种方法和组合物组合采用以改进靶向效率。如在实施例3.2(a)(i)中所述靶向大鼠载脂蛋白E(ApoE)基因座,例外的是也将表达ZFN1和ZFN2的表达载体引入大鼠ES细胞中。参见图11,其提供与rTZFN1P和rTZFN2P组合的ApoE靶向事件的示意图。如以下在实施例5中论述来确定靶向效率且结果示于表12中。为了筛选以便杂合靶向、纯合靶向和“混合”双重靶向(例如,复合杂合靶向),使用特异性引物和探针来确定基因分型。令人惊讶地,靶向效率增长8-10倍。

[0887] 表12.大鼠ApoE ZFN:改进的靶向效率。

	DNA	集落	筛选的	杂合靶向的	纯合靶向的	“混合”双重的	切割/未靶向的
	载体	330	184	15 (8.2%)	0	0	N/A
[0888]	载体 + ZFN 1	560	192	132 (68.8%)	6 (3.1%)	18 (9.4%)	17 (8.9%)
	载体 + ZFN 2	410	192	136 (70.8%)	2 (1.0%)	6 (3.1%)	18 (9.4%)

[0889] 构建具有自我缺失的药物选择盒和作为报道基因的lacZ基因的质粒靶向载体(参见图14,其说明可在包含选择盒的靶向载体电穿孔后发生的同源和非同源重组事件)。实现良好的靶向效率,并且产生高%嵌合体。也与靶向载体组合试验锌指核酸酶(ZFN)以考查它对改进靶向效率的影响(参见图16,其说明利用ZFN或TALEN改进靶向载体的同源重组的效率的基因靶向技术)。使靶向载体与切割ApoE基因座的2个ZFN对的表达载体共表达。用靶向载体与一组ZFN两者电穿孔的大鼠ESC克隆显示靶向效率比仅用靶向载体电穿孔的大鼠ESC克隆的靶向效率高8-10倍。此外,在我们的克隆中的约2%中检测到双等位基因纯合靶向。自这些靶向克隆中的两种获得高%嵌合体。

[0890] 将ApoE靶向(在ZFN辅助下)大鼠ESC克隆微量注射到SD胚泡中,随后使用标准技术将其转移到假妊娠SD受体雌性中。嵌合体通过毛色来鉴定(参见图17,其示出ApoE-ZFN-AB5嵌合体(即,ApoE<sup>-/-</sup>嵌合体);使雄性F0嵌合体与SD雌性配种。对于是否存在靶向ApoE等位基因来对种系F1幼仔基因分型(表13)。自这些靶向克隆中的两种获得高%嵌合体。

[0891] 表13.微量注射结果。

		嵌合体(%嵌合体)
[0892]	ApoE-ZFN1-AB5(纯合靶向)	12 4 (90, 90, 80, 80)
	ApoE-ZFN1-AE5(纯合靶向)	6 3 (90, 80, 70)

[0893] ApoE敲除大鼠提供研究各种类型的病症和疾病的手段。在人类中,载脂蛋白在乳糜微粒、HDL、LDL和VLDL中见到。ApoE对于富含甘油三酯的脂蛋白成分的正常分解代谢是不可少的。APOE的缺陷导致众多疾病状态,包括例如家族性高胆固醇血症、高脂血症、 $\beta$ 脂蛋白血症、家族性异常 $\beta$ 脂蛋白血症、III型高脂蛋白血症(HLP III)、冠状动脉疾病风险。一种同种型(ApoE4)与迟发型家族性和偶发性阿尔茨海默氏病相关,可能也与MS相关。

[0894] 在小鼠中,ApoE主要见于HDL中;转运胆固醇,如同在人类中一样。ApoE缺乏性小鼠(2种独立K0)具有5倍于正常的血浆胆固醇;截至3个月龄在它们的近端主动脉中产生富含泡沫细胞的沉积物(与人类综合征相当)。

[0895] 在大鼠中敲除ApoE提供研究内皮功能(包括但不限于斑块形成)、转录改变(RNA-Seq)、离体功能的动物模型。此外,大鼠的身材较大将有助于所有这些测定,并且潜在地改进RNA-Seq数据的质量。

[0896] 2.3. 使用锌指核酸酶使大鼠干扰白细胞素-2受体 $\gamma$ (IL2r- $\gamma$ )基因座失活

[0897] 如在实施例3.3(a)中所述,靶向大鼠干扰白细胞素-2受体 $\gamma$ (IL2r- $\gamma$ 或I12rg)基因座,例外之处在于还将表达ZFN U(ZFN上游)和ZFN D(ZFN下游)的表达载体引入大鼠ES细胞中。图18提供与ZFN U和ZFN D组合的IL2r- $\gamma$ 靶向事件的示意图。这些锌指结合的IL2r- $\gamma$ 基因座的序列在图18中在SEQ ID NO:93内指出。如以下在实施例3.3(a)中论述来确定靶向效率且结果示于表14中。简要地讲,纯合靶向克隆通过PCR证实。对于ZFN1成对:筛选192种突变克隆中的173种突变克隆(90%),且对于ZFN2对,筛选192种克隆中的162种克隆(84%)。

[0898] 表14. 大鼠IL2r- $\gamma$ 基因座的靶向。

板	筛选的集落	靶向	效率
7/18:仅载体	96	4	4.2%
7/18:载体+ZFN	96	3	3.1%

[0900] 将IL2r- $\gamma$ 靶向(在ZFN辅助下)大鼠ESC克隆微量注射到SD胚泡中,随后使用标准技术将其转移到假妊娠SD受体雌性中。通过毛色来鉴定嵌合体;使雄性F0嵌合体与SD雌性配种。针对是否存在靶向IL2r- $\gamma$ 等位基因来对种系F1幼仔进行基因分型。

[0901] 2.4.:使用CRISPR/Cas9使大鼠干扰白细胞素-2受体 $\gamma$ (IL2r- $\gamma$ )失活

[0902] 如在实施例3.3(a)中所述靶向大鼠IL2r- $\gamma$ 基因座,例外之处在于还将CRISPR/Cas9系统引入大鼠ES细胞中以有助于靶向效率。采用SBI: System Biosciences Cas9“SmartNuclease”全合一载体,并且Cas9表达由CAG、EF1a、PGK或CMV启动子驱动。使定制的gRNA接合到载体中,并通过H1启动子来表达。设计针对I12rg的4种gRNA。通过gRNA1-4靶向的大鼠IL2r- $\gamma$ 基因座的区示于图19中。为了筛选以便靶向(例如,杂合靶向、纯合靶向和复合杂合靶向),使用特异性引物和探针来确定基因分型。当采用各种向导RNA时的靶向结果示于表15中。“强”和“弱”是指集落具有靶向修饰的基于筛选的证据的力度。



[0903] 表15.用向导RNA靶向大鼠I12rg基因座

	一种或多种构建体	DNA(ug)	集落	候选物 (潜在地靶向)
[0904]	I12rg 质粒载体	6 ug	30	3 弱
	质粒 + SBI gRNA1	6 ug/40 ug	22	1 强, 1 弱
	质粒 + SBI gRNA2	6 ug/40 ug	45	2 强, 1 弱
[0905]	质粒 + SBI gRNA3	6 ug/40 ug	66	1 强, 2 弱
	质粒 + SBI gRNA4	6 ug/40 ug	59	0

[0906] 2.5.:使用CRISPR/Cas9使小鼠次黄嘌呤鸟嘌呤磷酸核糖基转移酶(Hprt)基因失活

[0907] 小鼠Hprt基因座在小鼠ES细胞中仅使用LTVEC或与CRISPR/Cas9组合靶向。靶向32.9kb完整Hprt编码序列以便缺失和用pCAGG-Puro嘌呤霉素抗性选择盒替换,其也表达eGFP。缺失端点为起始密码子和终止密码子。所使用的向导RNA序列为5'-GACCCGACAGUCCAGCGUCG-3'(SEQ ID NO:84),其靶向小鼠Hprt基因的外显子1。预测的靶位点裂解位置为距缺失的5'端的22个碱基对。在ES细胞中观察到的Cas9/gRNA靶上裂解效率 $\geq 93\%$ 。概述示于表16中。与仅使用LTVEC相比,使用CRISPR/Cas9以帮助靶向完整32.9kb Hprt基因座引起靶向增强5倍。

[0908] 表16.Hprt基因的CRISPR辅助的缺失的概述

<i>Hprt</i> 基因的 CRISPR 辅助的缺失的概述							
				靶向效率(%)			
靶向 基因	缺失 (kb)	5'同源 臂(kb)	3'同源臂 (kb)	仅 LTVEC	LTVEC + CRISPR/Cas9	增强 倍数	
<i>Hprt</i>	32.9	88	66	5.0	25.4	5.1	

[0910] 实施例3:大鼠基因组基因座的靶向修饰

[0911] 3.1:大鼠ESC靶向:大鼠Rosa26基因座。

[0912] 如在小鼠中一样,大鼠Rosa26基因座以相同间隔位于Setd5基因和Thumpd3基因之间。大鼠Rosa26基因座(图12,版面B)不同于小鼠Rosa26基因座(图12,版面A)。小鼠Rosa26转录体由2个或3个外显子组成。除与小鼠外显子1同源的外显子(Ex1a)之外,大鼠基因座还含有第二外显子1(Ex1b)。尚未在大鼠中鉴定出第三外显子。对大鼠Rosa26等位基因的靶向描绘在图12C中,其中使用来自DA大鼠ESC的基因组DNA,通过PCR来克隆各自5kb的同源臂。靶向等位基因含有替换在大鼠Rosa26内含子中的117bp缺失的SA(剪接受体)-lacZ-hUbe-neo盒。

[0913] 确定在大鼠Rosa26基因座处的靶向效率(表17)。将线性化载体电穿孔到DA或ACI大鼠ESC中,并且转染的集落使用标准技术在2i培养基+G418中培养。挑选个别的集落,并且使用等位基因损失(LOA)测定加以筛选(Valenzuela,D.等人,(2003)High-throughput engineering of the mouse genome coupled with high-resolution expression analysis,Nature Biotech.21:652-660,通过引用的方式并入本文中)。

表 17. 大鼠 Rosa26 靶向效率

细胞系	挑选的集落	再次证实阳性	靶向效率 (%)
DA.2B	192	4	2.1
ACL.G1	96	4	4.2

[0915] 使用Rosa26-靶向大鼠ESC克隆的嵌合体产生和种系传递。将再次证实的Rosa26-靶向大鼠ESC克隆微量注射到SD胚泡中,随后使用标准技术将其转移到假妊娠SD受体雌性中。通过毛色来鉴定嵌合体;使雄性F0嵌合体与SD雌性配种。针对是否存在靶向Rosa26等位基因对种系(刺鼠)F1幼仔进行基因分型;22个刺鼠幼仔中的9个被基因分型为在Rosa26基因座处是杂合的(表18)。

表 18. 使用靶向的 Rosa26 rESC 种系传递

细胞系	注射的 R26 克隆	产生嵌合体的克隆	种系传递克隆	总幼仔	rESC 源性幼仔	ESC 源性幼仔 (%)
DA.2B	4	3	2	AH7: 64 AE3: 112	AH7: 41 AE3: 6	AH7: 63 AE3: 3
ACL.G1	4	4	1	DE9: 39	DE9: 4	10

[0917] 为了证实在Rosa26基因座处的基因修饰的等位基因经由种系传递,lacZ表达通过在杂合Rosa26-靶向大鼠中的X-gal染色证实。来自14周龄杂合Rosa26-靶向大鼠的脑、心脏和胸腺及肺的X-gal染色示出lacZ的表达(分别地,图13B、13D和13F),而周龄匹配的野生型对照物示出低水平的本底X-gal染色(分别地,图13A、13C和13E)。在E12.5和E14.5杂合Rosa26-靶向大鼠胚胎中的X-gal染色示出lacZ的遍在表达(分别地,图13G和13I),而对照大鼠胚胎示出低水平的本底X-gal染色(分别地,图13H和13J)。

[0918] 3.2. (a) (i):大鼠载脂蛋白E(ApoE)基因座的靶向。

[0919] 靶向大鼠载脂蛋白E(ApoE)基因座以破坏ApoE功能。ApoE基因座的靶向使用包含侧接有与ApoE基因座同源的5'同源臂和3'同源臂的lacZ-hUb-neo盒的靶向载体进行。图20描绘已经通过1.8kb缺失和lacZ-hUb-neo盒插入破坏的基因修饰的大鼠ApoE基因座,其进一步包括包含由鱼精蛋白启动子驱动的Crei基因的自缺失Cre盒。电穿孔条件如下:6 $\mu$ g DNA;2.05x 10<sup>6</sup>个细胞;400V;200 $\mu$ F:342V,593微秒;在2i+10 $\mu$ M ROCKi中涂铺15cm 2x密集neoR MEF上。

[0920] 确定在ApoE基因座处的靶向效率且将其示于表19中。将线性化载体电穿孔到源于DA株的DA.2B大鼠ESC中,并且转染的集落使用标准技术培养。挑选个别的集落,并且使用等位基因损失(LOA)测定加以筛选。

表 19. 大鼠 ApoE 靶向效率

细胞系	载体	挑选的集落	靶向	靶向效率 (%)
DA.2B	ApoE-mSDC	192	7	3.7
DA.2B	ApoE-mSDC	192	15	7.8

[0922] 执行使用ApoE-靶向大鼠ESC克隆的嵌合体产生和种系传递。将ApoE-靶向大鼠ESC

克隆微量注射到SD胚泡中,随后使用标准技术将其转移到假妊娠SD受体雌性中。通过毛色来鉴定嵌合体;使雄性F0嵌合体与SD雌性配种。实现种系传递。对于是否存在靶向ApoE等位基因来对F1幼仔基因分型(表20)。

[0923] 表20. 微量注射结果

实施例	克隆	幼仔	嵌合体 (%嵌合体)
1	ApoE-AF5	4	3 (90, 90, 90)
2	ApoE-BC4	5	0

[0925] 由内源ApoE启动子驱动的LacZ表达通过在12周龄ApoE<sup>+/-</sup>雌性大鼠中在脑、血管和肝中的X-gal染色证实(分别地,图43-45)。图43-45示出反映内源ApoE的表达模式的lacZ的表达模式。年龄匹配的野生型对照物示出低水平的本底X-gal染色。

[0926] 进一步研究ApoE-缺失的大鼠的表型。执行纵向血清化学研究以每隔三周测量胆固醇、LDL、HDL和三酸甘油酯水平。图46A-D示出6周、9周、12周和15周龄的纯合靶向、杂合靶向和野生型大鼠中的血清胆固醇、LDL、HDL和三酸甘油酯水平。对由2只野生型大鼠、7只杂合大鼠和8只纯合大鼠组成的年龄匹配群执行眼睛放血(Eye bleed)。在雄性和雌性之间没有见到显著区别。纯合的ApoE-缺失大鼠示出升高的胆固醇和LDL水平及减小的HDL水平。与ApoE<sup>-/-</sup>小鼠不同,在ApoE-缺失的大鼠中没有观察到三酸甘油酯的显著增加。

[0927] 执行的其它表型的分析包括对于主动脉弓斑块形成的组织学/离体成像、对于主动脉弓斑块形成的体内成像和对于主动脉弓内皮的转录改变(全基因组鸟枪测序(RNA-Seq))。这些测定的时机取决于斑块形成的时间线。斑块可在ApoE<sup>-/-</sup>小鼠中在24周检测。

[0928] ApoE的其它靶向数据也提供在表22中。

[0929] 3.2. (a) (ii). 用靶向载体靶向在大鼠中的ApoE

[0930] 图20提供大鼠ApoE基因座和靶向质粒的示意图。图20的上部示意图示出大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区(分别是5kb和5.4kb;暗灰色框)的基因组结构。ApoE的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE的3个内含子指示为直线,且外显子2和3包含编码区并显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。

[0931] 图20的下部示意图为靶向载体。5'同源臂和3'同源臂(分别是5kb和5.4kb)通过暗灰色框来指示。靶向载体包含报道基因(lacZ)和由loxP位点(空心箭头)侧接的自缺失性盒。自缺失性盒包含操作性连接小鼠Prm1启动子的Crei基因和包含操作性连接人类泛素启动子的新霉素抗性基因的选择盒。

[0932] Crei基因包含两个编码Cre重组酶的外显子,其由内含子(Crei)分隔以防止它在原核细胞中表达。参见,例如美国专利8,697,851和美国申请公布2013-0312129,其详细地描述自缺失性盒并且以引用的方式整体并入本文中。通过采用Prm1启动子,自缺失性盒可在F0大鼠的雄性生殖细胞中被特异性缺失。将靶向载体电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集新霉素抗性MEF上。如在实施例1中所述来培养、选择并维持转化的大鼠ES细胞。

[0933] 如在表44中所示,筛选384个集落,并获得23个靶向克隆。靶向效率为5.99%。如本文在实施例1中所述将3个克隆注射到胚泡中。获得3个产生嵌合体的克隆,并且所述克隆中的1个经由种系传递靶向修饰。

[0934] 3.2. (a) (iii). 与锌指核酸酶组合用靶向载体靶向大鼠中的ApoE

[0935] 在实施例3.2 (a) (ii) 中采用的靶向载体与锌指核酸酶组合用于靶向大鼠ApoE基因座。表21提供大鼠ApoE基因座的基因组结构的概述。在表21中所示的位置取自大鼠基因组的参考序列 (ENSMBL) 的5.0构建版。ApoE在(-)链上的染色体1上。

[0936] 表21. 大鼠ApoE基因座以及锌指核酸酶结合位点和切割位点的位置的概述。

特征	起始	终止	长度	注释
外显子 1	8188111 0	8188118 2	73	5'非编码
外显子 2	8188026 9	8188033 2	64	含有 ATG
ATG	8188030 9	8188031 1	3	起始密码子
外显子 3	8187960 7	8187977 5	169	
ZFN1a 结合位点	8187970 7	8187969 3	15	CAGGCCCTGAACCGC(SEQ ID NO: 10)
ZFN1 切割位点	8187969 2	8187968 7	6	TTCTGG(SEQ ID NO: 11)
ZFN1b 结合位点	8187968 6	8187967 1	16	GATTACCTGCGCTGGG(SEQ ID NO: 12)
内含子 3-4	8187977 6	8187920 7	400	
ZF21a 结合位点	8187959 1	8187957 7	15	TTCACCCTCCGCACC(SEQ ID NO: 13)
ZFN2 切割位点	8187957 6	8187957 0	7	TGCTGAG(SEQ ID NO: 14)
ZF21b 结合位点	8187956 9	8187955 2	18	TATCCAGATCCAGGGGTT(SEQ ID NO: 15)
外显子 4	8187837 1	8187920 8	838	含有 TGA
TGA	8187848 2	8187848 4	3	
ApoE 缺失	8187848 2	8188031 1	183 0	

[0938] 图11提供大鼠ApoE基因座的示意图,并且用灰条指示ZFN1和ZFN2的切割位点。ZFN1的切割位点在外显子3中,并且ZFN2的切割位点在内含子3中。两个ZFN位点的精确位置阐述在表21中。对应于5'同源臂和3'同源臂的基因组区(分别是5kb和5.4kb)通过暗灰色框来指示。ApoE的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE基因的三个内含子指示为直线,并且外显子2和3包含编码区并显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。

[0939] 所采用的靶向载体与实施例3.2 (a) (ii) 中以及示于图20中的靶向载体相同,且图21A提供使用锌指核酸酶和在图20中描绘的靶向载体靶向在大鼠ES细胞中的ApoE基因座的

示意图。将ZFN作为两个表达质粒形式引入,每个质粒针对ZFN对的各一半。使用20ug针对ZFN1的质粒和20ug针对ZFN2的质粒。ZFN购自Sigma。各ZFN的表达由CMV启动子驱动。

[0940] 将靶向载体电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集neoR MEF上。如在实施例1中所述来培养、选择并维持转化的大鼠ES细胞。

[0941] 如在表22和表44中所示,筛选384个集落并获得290个靶向克隆。靶向效率为75.52%。如本文在实施例1中所述将2个克隆注射到胚泡中。获得2个产生嵌合体的克隆,并且所述克隆中的一个经由种系传递靶向修饰。

[0942] 此外,采用ZFN1和ZFN2,以效率2.08%产生8个双等位基因靶向克隆。

[0943] 表22. ApoE基因座的靶向。

	DNA	杂合靶向的	纯合靶向	微量注射	嵌合体 (%嵌合)	配种
	仅载体	15/192 (8%)	0			
[0944]	载体 + ZFN 对 1	156/192 (81%)	6/192 (3%)	2 个克隆	7 (70-90%)	
	载体 + ZFN 对 2	134/192 (70%)	2/192 (1%)			

[0945] 3.2. (b) (i) :使用大靶向载体 (LTC) 来靶向修饰大鼠载脂蛋白E (ApoE) 基因座。

[0946] 使用包含侧接有ApoE基因座的约45kb 5'同源臂和ApoE基因座的约23Kb 3'同源臂的lacZ-小鼠Prm1-Crei盒的大靶向载体 (LTVEC) 对ApoE基因座进行靶向。图22描绘大鼠ApoE基因座,其中ApoE基因座已通过1.83kb缺失以及插入lacZ基因和包含mPrm1-Crei盒和hUb-neo选择盒的自缺失性盒来破坏。在实施例3.2(a) (i) 中采用的方法可用以将该载体引入大鼠ES细胞中。

[0947] 实施例3.2. (b) (ii) .用大靶向载体 (LTVEC) 靶向大鼠ApoE基因座

[0948] 图22提供大鼠ApoE基因座和大靶向载体 (LTVEC) 的示意图。图22的上部示意图示出大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区 (分别是45kb和23kb;暗灰色框) 的基因组结构。ApoE的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE的3个内含子指示为直线,且外显子2和3包含编码区并显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。

[0949] 图22的下部示意图为LTVEC。5'同源臂和3'同源臂 (分别是45kb和23kb) 通过暗灰色框来指示。靶向载体包含报道基因 (lacZ) 和由loxP位点 (空心箭头) 侧接的自缺失性盒,所述自缺失性盒包含操作性连接小鼠Prm1启动子的Crei基因和包含操作性连接人类泛素启动子的新霉素抗性基因的药物选择盒。Crei基因包含两个编码Cre重组酶的外显子,其由内含子 (Crei) 分隔以防止它在原核细胞中表达。参见,例如美国专利8,697,851和美国申请公布2013-0312129,其详细地描述自缺失性盒并且以引用的方式整体并入本文中。通过采用Prm1启动子,自缺失性盒可在F0大鼠的雄性生殖细胞中被特异性缺失。

[0950] 将LTVEC电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集neoR MEF上。如在实施例1中所述来培养、选择并维持转化的大鼠ES细胞。

[0951] 如在表44中所示,筛选288个集落,并获得8个靶向克隆。靶向效率为2.78%。如本文在实施例2中所述将3个克隆注射到处于胚泡阶段的宿主胚胎中以产生嵌合大鼠(F0)。此外,产生一个双等位基因靶向克隆,从而提供0.35%的双等位基因效率。

[0952] 3.2. (b) (iii). 与锌指核酸酶组合用大靶向载体(LTVEC)靶向在大鼠中的ApoE

[0953] 在实施例3.2. (b) (ii) 中采用的LTVEC与锌指核酸酶组合用于靶向大鼠ApoE基因座。表21提供大鼠ApoE基因座的基因组结构的概述,并且所示位置取自大鼠基因组参照序列(ENSMBL)的5.0构建版。

[0954] 图23提供大鼠ApoE基因座的示意图,并且用灰条指示ZFN1和ZFN2的切割位点。ZFN1的切割位点在外显子3中,并且ZFN2的切割位点在内含子3中。两个ZFN位点的精确位置阐述在表21中。5'同源臂和3'同源臂(分别是45kb和23kb)通过暗灰色框来指示。ApoE基因的外显子1是非编码的,并且显示为最靠近5'同源臂的空心框。ApoE基因的三个内含子指示为直线。外显子2和3包含编码区且显示为点描灰色框。外显子4含有编码序列和非编码序列两者,如通过点描灰色阴影和空心框所指示。

[0955] 所采用的LTVEC与在实施例3.2(b) (ii) 中且示于图22中的LTVEC相同。将ZFN作为两个表达质粒形式引入,每个质粒针对ZFN对各一半。使用20ug针对ZFN1的质粒和20ug针对ZFN2的质粒。ZFN购自Sigma。各ZFN的表达由CMV启动子驱动。

[0956] 将靶向载体电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集neoR MEF上。如在实施例1中所述来培养、选择并维持转化的大鼠ES细胞。

[0957] 如在表44中所示,筛选288个集落,并获得16个靶向克隆。靶向效率为5.56%。如本文在实施例2中所述将1个克隆注射到胚泡中。

[0958] 此外,采用ZFN1和ZFN2以0.35%的效率产生一个双等位基因靶向克隆。

[0959] 3.2. (b) (iv). 与CRISPR/Cas9组合用大靶向载体(LTVEC)靶向在大鼠中的ApoE

[0960] 在实施例3.2. (b) (ii) 中采用的LTVEC与CRISPR/Cas9组合用于靶向大鼠ApoE基因座。表23示出其中仅使用ApoE LTVEC来靶向大鼠ApoE基因座或与CRISPR/Cas9核酸酶组合用以靶向大鼠ApoE基因座的实验的结果的比较。在各实验中,电穿孔的细胞以高密度涂铺且进行药物选择以发现抗药性的集落。挑选抗药性集落且使用如本文所述的等位基因修饰(MOA)测定针对靶向修饰加以筛选。具体地讲,  $4 \times 10^6$  个细胞在400V的电压、100uF的电容和0的抗性下用2ug ApoE LTVEC电穿孔。在后一实验中,还将6ug Cas9表达质粒和3ug ApoE gRNA2或3ug ApoE gRNA3电穿孔。使用75ug/mL的G418进行选择。ApoE gRNA2具有GCAGGCCCTGAACCGCTTCTGG的序列(SEQ ID NO:87)且靶向距大鼠ApoE外显子3的起始的区67bp 3'。ApoE gRNA3具有CCTGCGCTGGGTGCAGACGCTTT的序列(SEQ ID NO:88)且靶向大鼠ApoE外显子3的起始的区97bp 3' (参见图47)。如在表23中所示,当Cas9和gRNA中的任一种与ApoE LTVEC一起引入细胞中,靶向效率增加(从43%增加到53%或47%)。在与ApoE gRNA2或3组合用ApoE LTVEC靶向的5个集落中观察到双等位基因靶向,但在仅用ApoE LTVEC的情况下没有观察到双等位基因靶向。

[0961] 表23. 在有和没有CRISPR/Cas9的情况下Rag2 LTVEC靶向的比较

载体	Cas9	gRNA	筛选的集落	靶向的克隆	靶向的双等位基因	靶向效率
[0962] ApoE LTVEC	否	否	75	32	0	43%
ApoE LTVEC	是	ApoE gRNA2	80	42	1	53%
ApoE LTVEC	是	ApoE gRNA3	89	42	4	47%

[0963] 3.3(a):大鼠干扰白细胞素-2受体  $\gamma$  (IL2r- $\gamma$ ) 基因座的靶向

[0964] 靶向大鼠干扰白细胞素-2受体  $\gamma$  (IL2r- $\gamma$  或 I12rg) 基因座以破坏 IL2r- $\gamma$  功能。IL2r- $\gamma$  对于通过 IL-2、IL-4、IL-7、IL-9、IL-15、IL-21 进行的信号传导起重要作用,并且在 IL2r- $\gamma$  中的突变与重度 T、B 和 NK 细胞发育缺陷相关。

[0965] 使用包含侧接有与 IL2r- $\gamma$  基因座同源的 5' 同源臂和 3' 同源臂的 eGFP-hUb-neo 盒的靶向载体对 IL2r- $\gamma$  基因座进行靶向,如图 24 中所描绘。图 25 描绘大鼠 IL2r- $\gamma$  基因座的基因组结构,其中 IL2r- $\gamma$  基因座已通过 3.2kb 缺失来破坏。靶向的 IL2r- $\gamma$  基因座也包含 eGFP 基因和自缺失性盒,所述自缺失性盒含有操作性连接小鼠鱼精蛋白 1 启动子的 Crei 和包含操作性连接新霉素抗性基因的 hUb 启动子的药物选择盒。

[0966] 测定在 IL2r- $\gamma$  基因座处的靶向效率且示于表 24 中。将线性化载体电穿孔到 DA.2B 大鼠 ESC 中,并且转染的集落使用标准技术培养。挑选个别的集落,并且使用等位基因损失 (LOA) 测定加以筛选。

细胞系	载体	挑选的集落	靶向	靶向效率 (%)	嵌合体 (%嵌合)
[0967] DA.2B	I12rg-floxed neo	136	1	0.7	5 (70-90%)
DA.2B	I12rg-mSDC	96	4	4.2	

[0968] 执行使用 IL2r- $\gamma$  - 靶向大鼠 ESC 克隆的嵌合体产生和种系传递。将 IL2r- $\gamma$  - 靶向大鼠 ESC 克隆微量注射到 SD 胚泡中,接着使用标准技术将所述胚泡转移到假妊娠 SD 受体雌性中。通过毛色来鉴定嵌合体;使雄性 F0 嵌合体与 SD 雌性配种。针对是否存在靶向的 IL2r- $\gamma$  等位基因来对种系 F1 幼仔进行基因分型 (表 25)。在用克隆 I12rg-CG12 的另一微量注射实验中,还通过毛色和基因分型证实种系传递。

[0969] 表 25. 微量注射结果

实施例	克隆	幼仔	嵌合体 (%嵌合体)
[0970] 1	I12rg-AA1	5	2 (90, 70)
2	I12rg-AA1	10	3 (90, 90, 80)
3	I12rg-CG12	11	7 (95, 90, 90, 90, 80, 80, 80)

[0971] 进一步研究 I12rg<sup>-Y</sup> 嵌合体 #3 的表型。外周血液单核细胞 (PBMC) 用识别若干淋巴谱系中的抗原的抗体染色。从 2 个嵌合体中检测到 GFP 阳性 PBMC, 如在图 30 中所示。此外, GFP + 细胞对 T 细胞标记 CD3 为阴性的 (图 29A), 并且大部分对于 B 细胞标记 B220 和 NK 细胞标记 CD161a 为阴性的 (分别地, 图 29B 和 29C)。将来自野生型大鼠的 PBMC 用作 GFP 表达的阴性对照。参见图 29D-F。小双重阳性群体与在小鼠中公开的 I12rg 敲除表型一致。这些数据从含有

IL2受体  $\gamma$ -阳性细胞的嵌合大鼠获得,并且这可使表型分析复杂化。还可对来自骨髓和脾的细胞群体执行流式细胞分析以揭示淋巴细胞数目的相应减少。参见, Mashimo等人, (2010) PLoS One 5(1):e8870.

[0972] 3.3(b):大鼠干扰白细胞素-2受体  $\gamma$  (IL2r- $\gamma$ ) 基因座的靶向修饰

[0973] 靶向大鼠干扰白细胞素-2受体  $\gamma$  (IL2r- $\gamma$ ) 基因座以破坏在大鼠中的IL2r- $\gamma$  功能。图25示出大鼠I12rg基因座的基因组结构(图25的上部版面)和引入基因座中的靶向载体(图25的下部版面)。eGFP被选作报道基因以使得可使用FACS考查遗传修饰的大鼠的免疫表型。自缺失性盒(hUb-Neo;Prm1-Cre)用于在F0大鼠的雄性生殖细胞中特异性缺失药物选择盒和Cre基因。另外,靶向载体设计成缺失大鼠I12rg基因的整个编码区(约3.2kb)。

[0974] 通过使用对大鼠I12rg基因座具有特异性的引物进行PCR来证实在大鼠ESC中的缺失的尺寸。在将靶向克隆微量注射到处于胚泡阶段的宿主胚胎中,获得高百分数的嵌合体。已创建那些嵌合体用于配种。为了确定靶向是否如预期起作用,在配种之前收集来自嵌合体的外周血液,并且经由FACS分析在外周血液中的免疫细胞的表型。如在图30中所示,在考查的3个嵌合体中的2个中的外周血液中检测到GFP阳性细胞,并且嵌合大鼠含有针对GFP(即,I12rg KO细胞)是阳性的小于1%的T细胞、小于1%的B细胞和小于1%的NK细胞(图29A-C)。

[0975] 3.4(a)(i).用大靶向载体(LTVEC)靶向在大鼠中的Rag2基因座

[0976] 表26提供大鼠Rag2基因座的基因组结构的概述,并且所示位置取自大鼠基因组参照序列(ENSEMBL)的5.0构建版。Rag2在(+)链上的染色体3上。

[0977] 表26.大鼠Rag2基因座的基因组结构概述。

特征	起始	终止	长度	注释
外显子1	97,851,317	97,851,448	132	
外显子2	97,854,635	97,854,693	59	
外显子3	97,858,260	97,859,615	1,356	含有整体编码序列
ATG	97,856,286	97,856,288	3	起始密码子
TGA	97,857,867	97,857,869	3	终止密码子
Rag2缺失	97,856,289	97,859,784	3,496	

[0979] 图26提供大鼠Rag2基因座和大靶向载体(LTVEC)的示意图。LTVEC为140kb且针对缺失靶向大鼠Rag2基因座的约5.7kb部分。图26的上部示意图示出大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区(分别是48kb和84kb;暗灰色框)的基因组结构。Rag2包含通过点描灰色阴影来指示的单一外显子。

[0980] 图26的下部示意图为LTVEC。5'同源臂和3'同源臂(分别是48kb和84kb)通过暗灰色框来指示。LTVEC包含报道基因(lacZ)和由loxP位点(空心箭头)侧接的自缺失性盒。自缺失性盒包含操作性连接Crei基因的小鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。产生LTVEC的另一变型,其中新霉素抗性基因用潮霉素抗性基因替换,从而能够重新靶向I12rg靶向的大鼠ES细胞。Crei基因包含两个编码Cre重组酶的外显子,其由内含子(Crei)分隔以防止它在原核细胞中表达。参见,例如美国专利8,697,



851和美国申请公布2013-0312129,其详细地描述自缺失性盒并且以引用的方式整体并入本文中。通过采用小鼠Prm1启动子,自缺失性盒可在F0大鼠的雄性生殖细胞中被特异性缺失。

[0981] 将LTVEC电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集neoR MEF上。如在实施例1中所述来培养并维持转化的大鼠ES细胞。

[0982] 如在本文中其它地方所述来筛选集落并获得靶向的克隆。随后如在本文中其它地方所述,将靶向的克隆注射到宿主胚胎中以产生F0大鼠。

[0983] 3.4(a) (ii).用大靶向载体 (LTVEC) 和CRISPR/Cas9靶向在大鼠中的Rag2基因座

[0984] 表27示出其中仅使用具有潮霉素抗性基因的Rag2 LTVEC的变型以靶向大鼠Rag2基因座(参见图48)或与CRISPR/Cas9核酸酶组合使用以靶向大鼠Rag2基因座的实验的结果的比较。在各实验中,电穿孔的细胞以高密度涂铺且进行药物选择以发现抗药性的集落。挑选抗药性集落且使用如本文所述的等位基因修饰(MOA)测定针对靶向修饰加以筛选。具体地讲,4x 10<sup>6</sup>个细胞在400V的电压、100uF的电容和0的抗性下用2ug Rag2 LTVEC电穿孔。在后一实验中,还将6ug Cas9表达质粒和3ug Rag2 gRNA1或3ug Rag2 gRNA4电穿孔。使用75ug/mL的G418进行选择。Rag2 gRNA1具有CCAGCTACTTGCTCGTACAA的序列(SEQ ID NO:89)并靶向大鼠Rag2起始密码子(ATG)的区219bp 3'。Rag2 gRNA4具有CCCCTCAGATTCACGTGCGT的序列(SEQ ID NO:90)且靶向大鼠Rag2终止密码子(TAG)的区12bp 3'(参见图48)。如在表27中所示,当Cas9和gRNA中的任一种与Rag2 LTVEC一起引入细胞中,靶向效率增加(从0增加到10%或38%)。在一个集落上观察到双等位基因靶向。

[0985] 表27.在有和没有CRISPR/Cas9的情况下Rag2 LTVEC靶向的比较

载体	Cas9	gRNA	筛选的集落	靶向的克隆	靶向的双等位基因	靶向效率
[0986] Rag2 LTVEC	否	否	36	0	0	0
Rag2 LTVEC	是	Rag2 gRNA1	23	5	1	22%
Rag2 LTVEC	是	Rag2 gRNA4	16	1	0	6%

[0987] 3.4. (b) (i):靶向在大鼠中的Rag1基因座和Rag2基因座

[0988] 图27提供大鼠Rag1/Rag2基因座的基因组结构。CDS指示编码序列且灰色框代表外显子。Rag2在“+”链上,其中向右进行转录。Rag1在“-”链上,其中向左进行转录。Mbp=百万碱基对。

[0989] 表28提供大鼠Rag2基因座和Rag1基因座的基因组结构的概述,并且所示位置取自大鼠基因组参照序列(ENSEMBL)的5.0构建版。Rag1在(-)链上的染色体3上。

[0990] 表28.大鼠Rag1基因座的基因组结构概述。

特征	起始	终止	长度	注释
外显子 1	97,877,145	97,877,066	80	
[0991] 外显子 2	97,872,503	97,866,047	6,457	含有整体编码序列
ATG	97,872,489	97,872,487	3	起始密码子
TAA	97,869,369	97,869,367	3	终止密码子
Rag1-2 缺失	97,856,289	97,872,486	16,198	

[0992] 图28提供大鼠Rag2基因座和Rag1基因座及大靶向载体 (LTVEC) 的示意图。LTVEC为约70kb且靶向包含缺失的Rag1和Rag2基因座的约16.6kb大鼠基因组基因座。图28的上部示意图示出Rag1和Rag2基因座和对应于5'同源臂和3'同源臂的基因组区(分别是48kb和15kb;暗灰色框)的基因组结构。Rag2和Rag1各自包含由点描灰色阴影指示的单一外显子。在图28中的下部示意图为LTVEC。5'同源臂和3'同源臂(分别是48kb和15kb)通过暗灰色框来指示。LTVEC包含报道基因(lacZ)和由loxP位点(空心箭头)侧接的自缺失性盒。自缺失性盒包含操作性连接Crei基因的大鼠Prm1启动子和包含操作性连接新霉素抗性基因的人类泛素启动子的药物选择盒。产生LTVEC的另一变型,其中新霉素抗性基因用潮霉素抗性基因替换,从而能够重新靶向I12rg靶向的大鼠ES细胞。Crei基因包含两个编码Cre重组酶的外显子,其由内含子(Crei)分隔以防止它在原核细胞中表达。参见,例如美国专利8,697,851和美国申请公布2013-0312129,其详细地描述自缺失性盒并且以引用的方式整体并入本文中。通过采用驱动Crei在雄性生殖细胞中特异性表达的大鼠Prm1启动子,自缺失性盒可从F0大鼠的雄性生殖细胞中缺失。

[0993] 将LTVEC电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集neoR MEF上。如在实施例1中所述来培养并维持转化的大鼠ES细胞。

[0994] 如在本文中其它地方所述来筛选集落并获得靶向的克隆。随后如在本文中其它地方所述,将靶向的克隆注射到宿主胚胎中以产生F0大鼠。

[0995] 3.4. (b) (ii): 靶向在其中已经靶向I12rg基因座的大鼠ES细胞中的Rag1和Rag2基因座

[0996] 制备如在图50中的LTVEC以针对缺失靶向Rag1基因座和Rag2基因座。LTVEC的全长为72kb。将LTVEC电穿孔到如在实施例3.3中已经靶向I12rg基因座缺失的大鼠ES细胞中。具体地讲,大鼠ES细胞来自克隆I12rg-CG12,其种系传递在实施例3.3(a)中证实。如在实施例1中所述来培养并维持转化的大鼠ES细胞。如在本文中其它地方所述来筛选双重靶向克隆并获得靶向的克隆。I12rg-CG12细胞以85%的效率再靶向,且I12rg突变仍存在于靶向的克隆中。电穿孔如在本文中其它地方描述,且使用1.5ug/ml嘌呤霉素进行抗生素选择。随后如在本文中其它地方所述,将靶向的克隆注射到宿主胚胎中以产生F0大鼠。再靶向是有利的,因为其比用I12rg靶向大鼠异种交配Rag1/Rag2-靶向大鼠快。

[0997] 实施例4. 人源化

[0998] 4.1. 大鼠基因组基因座的人源化

[0999] 采用本文所述的大鼠ES细胞进行大鼠基因组基因座的人源化,所述细胞能够在一次或多次体外电穿孔之后持续它们的多潜能性,并且能够向后代传递靶向基因修饰。另外,

为了规避质粒在容纳大基因组DNA片段方面的限制,以及为了克服向在大鼠ES细胞中的内源基因座中引入靶向基因修饰的低效率,通过利用细菌同源重组(BHR)以及采用大靶向载体(LTVEC)在例如大肠杆菌的细菌中进行一次或多次靶向基因修饰。本文所述的LTVEC例如包括内源大鼠基因组序列的具有一种或多种修饰的大片段,或包含侧接有与特定基因组区互补的大鼠同源臂的外源核酸(例如,同源或直系同源人类核酸)。

#### [1000] 4.2. 大鼠免疫球蛋白基因座的人源化

[1001] 通过以下方式进行内源大鼠免疫球蛋白重链基因座的人源化:移除一个或多个内源大鼠免疫球蛋白重链核酸序列(例如,一个或多个内源 $V_H$ 基因区段、一个或多个人类D基因区段和一个或多个人类 $J_H$ 基因区段);以及向修饰的免疫球蛋白基因座中引入靶向载体,例如包含以下的大靶向载体(LTVEC):(i)一个或多个未重排的人类可变区核酸序列(例如,一个或多个人类 $V_H$ 基因区段、一个或多个人类D基因区段和一个或多个人类 $J_H$ 基因区段)或一个或多个重排的人类可变区核酸序列(例如,一个或多个人类重排的V-D-J基因区段);(ii)选择盒(例如,侧接有loxP位点的新霉素抗性基因);和(iii)5'和3'大鼠同源臂。

[1002] 简要地讲,通过用由大鼠同源臂侧接的选择盒靶向内源大鼠免疫球蛋白重链基因座来移除或失活在大鼠BAC克隆中的一个或多个内源大鼠免疫球蛋白重链可变区基因区段(即,一个或多个 $V_H$ 基因区段、一个或多个人类D基因区段和一个或多个人类 $J_H$ 基因区段)。更具体地讲,构建靶向载体以含有侧接有与靶大鼠基因组序列(例如,涵盖一个或多个大鼠 $V_H$ 基因区段、一个或多个人类D基因区段和一个或多个人类 $J_H$ 基因区段的上游和下游大鼠基因组DNA序列)互补的5'和3'大鼠同源臂的选择盒(例如,侧接有loxP位点的新霉素抗性基因)。

[1003] 紧接着,选择含有涵盖大鼠免疫球蛋白重链基因座的大鼠基因组DNA片段的细菌细胞,并且用编码操作性连接短暂可诱导的启动子的重组酶的质粒(例如,PABG)引入。随后将如上构造的靶向载体引入重组感受态细菌细胞中。在电穿孔之后,细菌细胞用诱导因子(例如,阿拉伯糖苷)处理以引发靶向载体与在BAC克隆中的靶大鼠基因组序列之间的同源重组。转化细胞以高密度涂铺,并且进行药物选择以发现抗药性的集落。挑选抗药性集落,并且针对靶向修饰加以筛选。

[1004] 为了促进靶向基因修饰的鉴定,采用高通量定量测定,即,等位基因修饰(MOA)测定,其允许在基因修饰之后大规模筛选在亲本染色体中的一种或多种修饰等位基因。所述MOA测定可经由包括但不限于例如实时PCR(qPCR)的定量PCR的各种分析技术进行。例如,所述实时PCR包括识别靶基因座的第一引物集和识别非靶向参考基因座的第二引物集。另外,引物集可包括识别扩增序列的荧光探针。可选地,定量测定可经由多种分析技术来进行,所述技术包括但不限于荧光介导的原位杂化(FISH)、比较基因组杂化、等温DNA扩增、与固定的探针定量杂化、Invader Probes®、MMP测定®、TaqMan®分子信标和Eclipse™探测技术。(参见,例如US2005/0144655,其以引用的方式并入本文中)。

[1005] 接着用包含以下的大靶向载体(LTVEC)电穿孔包含修饰的大鼠BAC克隆(即,含有其中一个或多个内源重链可变区基因区段( $V_H$ 、D和/或 $J_H$ 基因区段)已缺失或失活的大鼠基因组DNA序列的BAC克隆)的细菌细胞:(i)一个或多个未重排的人类可变区核酸序列(例如,一个或多个未重排的人类 $V_H$ 基因区段、一个或多个人类D基因区段和一个或多个人类 $J_H$ 基因区段)或一个或多个重排的人类可变区核酸序列(例如,一个或多个重排的人类V-D-J基因

区段)。

[1006] 如上所述进行在细菌细胞中同源重组的引发以及阳性克隆的选择。未重排或重排的人类免疫球蛋白重链可变区核酸序列在靶向内源免疫球蛋白重链基因座中时变成操作性连接内源大鼠免疫球蛋白重链恒定区核酸序列。可选地,内源大鼠重链恒定区基因座可例如通过从内源重链恒定区基因座缺失一个或多个大鼠重链恒定区基因区段(CH)来失活,并且可被人类重链恒定区核酸序列替换。

[1007] 同样,通过以下方式进行内源大鼠免疫球蛋白 $\kappa$ 或 $\lambda$ 轻链基因座的人源化:移除一个或多个内源大鼠免疫球蛋白 $\kappa$ 和/或 $\lambda$ 轻链可变区核酸序列(例如,一个或多个内源大鼠 $V_{\kappa}$ 基因片段和一个或多个内源大鼠 $J_{\kappa}$ 基因片段);以及用靶向载体如包含以下的大靶向载体(LTVEC)靶向修饰的免疫球蛋白轻链基因座:(i)一个或多个未重排的人类免疫球蛋白轻链可变区核酸序列(例如,一个或多个人类 $V_{\kappa}$ 基因片段和一个或多个人类 $J_{\kappa}$ 基因片段)或一个或多个重排的人类可变区核酸序列(例如,一个或多个人类重排的 $V_{\kappa}$ - $J_{\kappa}$ 基因片段);(ii)选择盒(例如,侧接有loxP位点的新霉素抗性基因);和(iii)5'和3'大鼠同源臂。

[1008] 未重排或重排的人类免疫球蛋白轻链可变区核酸序列在靶向内源免疫球蛋白轻链基因座中时变成操作性连接内源大鼠免疫球蛋白轻链恒定区核酸序列。

[1009] 在细菌细胞中这样产生的LTVEC包含例如含有人源化大鼠免疫球蛋白重链或轻链基因座的插入核酸,其中一个或多个内源大鼠重链或轻链可变区基因片段已被一个或多个人类重链或轻链可变区基因片段替换;以及与特定基因组靶序列互补的大鼠同源臂(例如,在5kb-150kb的范围内)。接着使上述包含基因修饰的LTVEC线性化,并且电穿孔到大鼠ES细胞中。电穿孔的大鼠ES细胞以高密度涂铺以选择包含靶向载体的抗药性ES细胞。药物选择过程移除大多数的涂铺细胞(约99%),留下个别的集落,其各自为来源于单一细胞的克隆。在剩余细胞之中,大多数细胞(约80-100%)含有在基因组中的随机位置整合的靶向载体。因此,挑选集落并个别地基因分型以鉴定在正确基因组位置处包含靶向载体的大鼠ES细胞(例如,使用上述等位基因修饰(MOA)测定)。

[1010] 为了增加靶向基因修饰的效率,大鼠ES细胞用表达ZFN1和ZFN2(或TALEN1和TALEN2)的表达载体(或mRNA)连同LTVEC一起电穿孔。靶向载体的同源臂位于ZFN靶位点外部,因此,靶向载体不被ZFN裂解。由ZFN产生的双链断裂刺激源性定向修复(HDR),其另外占据通常在哺乳动物细胞中发生的极小百分数的修复(相较于非同源末端接合;NHEJ)。

[1011] 可选地,如本文所述的含有II型CRISPR相关核酸酶(例如,Cas9)、向导RNA(包括CRISPR-RNA(cr-RNA)和反式激活CRISPR RNA(tracrRNA))的表达载体可连同LTVEC一起引入细菌细胞中以增加在靶基因组基因座处同源重组的效率。电穿孔的细胞以高密度涂铺,并且进行药物选择以发现抗药性的集落。挑选抗药性集落且使用如本文所述的等位基因修饰(MOA)测定针对靶向修饰加以筛选。在这些程序之后,可实现靶向效率的改进。例如,改进的量可较小(例如,从10%改进到15%)或较大(例如,从10%改进到80%)。

[1012] 随后将包含靶向基因修饰的所选择的ES细胞引入宿主大鼠胚胎(例如,前桑椹胚阶段或胚泡阶段的大鼠胚胎)中,并且植入代孕母体的子宫中以产生建群大鼠(F0大鼠)。随后,使建群大鼠与野生型大鼠配种以产生对于基因修饰杂合的F1子代。杂合F1大鼠的交配可产生对于基因修饰纯合的子代。

[1013] 4.3(a).用人类IL2受体 $\gamma$ 替换大鼠IL2rg

[1014] 表29提供大鼠白细胞介素2受体  $\gamma$  基因座的基因组结构的概述,并且所示位置取自大鼠基因组参照序列(ENSEMBL)的5.0构建版。Il2rg在(-)链上的染色体X上。

[1015] 表29. 大鼠Il2rg基因座的基因组结构的概述

特征	起始	终止	长度	注释
外显子 1	72,021,388	72,021,516	129	含有 ATG
ATG	72,017,500	72,017,502	3	起始密码子
外显子 2	72,021,007	72,021,160	154	
ZFN1a 结合位点	72,021,014	72,021,028	15	CAGGCCCTGAACCGC(SEQ ID NO: 17)
ZFN1 切割位点	72,021,008	72,021,013	6	TTCTGG(SEQ ID NO: 18)
ZFN1b 结合位点	72,020,993	72,021,007	15	GATTACCTGCGCTGGG(SEQ ID NO: 20)
外显子 3	72,020,606	72,020,790	185	
外显子 4	72,020,274	72,020,413	140	
外显子 5	72,019,662	72,019,824	163	
外显子 6	72,019,101	72,019,197	97	
外显子 7	72,018,844	72,018,910	67	
外显子 8	72,017,856	72,018,506	651	含有 TGA
TGA	72,018,321	72,018,323	3	终止密码子
Il2rg 缺失	72,018,323	72,021,502	3,180	

[1016] 在图25中的下部示意图为Il2rg 3.2kb缺失的靶向载体。靶向载体包含操作性连接内源启动子的报道基因(eGFP)和由loxP位点(空心箭头)侧接的自缺失性盒。自缺失性盒包含操作性连接小鼠Prm1启动子的Crei基因和包含操作性连接人类泛素启动子的新霉素抗性基因的选择盒。

[1017] Crei基因包含两个编码Cre重组酶的外显子,其由内含子(Crei)分隔以防止它在原核细胞中表达。参见,例如美国专利8,697,851和美国申请公布2013-0312129,其详细描述自缺失性盒并且以引用的方式整体并入本文中。通过采用小鼠Prm1启动子,Cre表达盒和药物选择盒可在F0大鼠的雄性生殖细胞中特异性缺失。将靶向载体电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集新霉素抗性MEF上。如在实施例1中所述来培养、选择并维持转化的大鼠ES细胞。

[1018] 如在图31中所示,构建质粒靶向载体以用全长人类白细胞介素2受体  $\gamma$  编码区替换全长大鼠白细胞介素2受体  $\gamma$  编码区。将靶向载体电穿孔到在实施例1中获得的大鼠ES细

胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集新霉素抗性MEF上。具体地讲,4x 10<sup>6</sup>个细胞在400V的电压、100uF的电容和0的抗性下用2ug I12rg全长人源化载体电穿孔。使用75ug/mL的G418进行选择。如在实施例1中所述来培养、选择并维持转化的大鼠ES细胞。

[1020] 如在表44中所示,筛选168个集落,并获得6个靶向克隆。靶向效率为3.57%。如在实施例1中所述将一个克隆注射到胚泡中,并且获得一个产生嵌合体的克隆。

[1021] 如本文在实施例1中所述将克隆注射到胚泡中。获得产生F0嵌合大鼠的克隆。胚泡使用标准技术转移到假妊娠受体雌性,且获得嵌合F0大鼠。获得经由种系传递靶向修饰的F0大鼠。

[1022] 4.3(b) (i). 用人类IL2rg胞外结构域替换大鼠IL2rg胞外结构域

[1023] IL 2受体 $\gamma$ 的全长人源化是有用的,因为具有该修饰的基因座的大鼠将产生人类I12rg;并且这将允许用对人类I12rg具有特异性的抗体检测在大鼠中的人类I12rg。

[1024] 外人源化(即,用I12rg的人类胞外结构域替换I12rg的大鼠胞外结构域)将产生将结合I12rg的人类配体的I12rg多肽,但因为细胞质结构域仍然是大鼠的,所以I12rg的外人源化形式也将与大鼠信号传导机构相互作用。图33提供人类IL-2rg蛋白(SEQ ID NO:20; NP\_000197.1);大鼠IL-2rg蛋白(SEQ ID NO:21;NP\_543165.1);以及包含融合于大鼠IL-2rg蛋白的其余部分的IL-2rg的人类胞胞外结构域的嵌合IL-2rg蛋白(SEQ ID NO:22)的序列比对。在人类IL-2rg和大鼠IL-2rg之间的接合点通过垂直线标注。

[1025] 表30提供大鼠白细胞介素2受体 $\gamma$ 基因座的基因组结构的概述,并且所示位置取自大鼠基因组参照序列(ENSEMBL)的5.0构建版。I12rg在(-)链上的染色体X上。进一步标注的是I12rg的胞外结构域的位置。

[1026] 表30. 大鼠I12rg基因座的基因组结构的概述

特征	起始	终止	长度	注释
外显子 1	71,111,444	71,111,543	100	含有 ATG
ATG	71,111,537	71,111,539	3	起始密码子
外显子 2	71,110,897	71,111,050	154	
外显子 3	71,110,504	71,110,688	185	
外显子 4	71,110,156	71,110,295	140	
外显子 5	71,109,228	71,109,390	163	
外显子 6	71,108,599	71,108,645	47	含有跨膜域
外显子 7	71,108,277	71,108,346	70	
外显子 8	71,107,404	71,107,921	518	含有 TGA
TGA	71,108,736	71,108,738	3	终止密码子
全长人源化:	71,107,404	71,111,539	4,136	(ATG 至 TGA 加 3'多聚-A)
外人源化	71,108,679	71,111,539	2,861	(ATG 至跨膜结构域开始)

[1028] 构建质粒靶向载体以如图32中所示用人类胞外结构域替换白细胞介素2受体 $\gamma$ 编码区的大鼠胞外结构域。将靶向载体电穿孔到在实施例1中获得的大鼠ES细胞中,并将细胞在2i+10uM ROCKi中涂铺在15cm 2x密集新霉素抗性MEF上。如在实施例1中所述来培养、选

择并维持转化的大鼠ES细胞。

[1029] 如在表44中所示,筛选192个集落,并获得13个靶向克隆。靶向效率为6.77%。

[1030] 如本文所述在实施例1中所述将两个克隆注射到胚泡中,且获得两个产生F0大鼠的克隆。获得产生F0嵌合大鼠的克隆。获得经由种系传递靶向修饰的F0大鼠。

[1031] 4.3 (b) (ii) .与CRISPR/Cas9组合使用质粒用人类IL2rg胞外结构域替换大鼠IL2rg胞外结构域

[1032] 表31示出其中仅使用在图32中示出的I12rg胞外结构域人源化载体的变型以靶向大鼠I12rg基因座或与CRISPR/Cas9核酸酶组合使用以靶向大鼠I12rg基因座的实验的结果的比较。在各实验中,电穿孔的细胞以高密度涂铺且进行药物选择以发现抗药性的集落。挑选抗药性集落且使用如本文所述的等位基因修饰(MOA)测定针对靶向修饰加以筛选。具体地讲,  $4 \times 10^6$  个细胞在400V的电压、100uF的电容和0的抗性下用2ug I12rg胞外结构域人源化载体电穿孔。在后一实验中,还将6ug Cas9表达质粒和3ug I12rg gRNA2或3ug I12rg gRNA4电穿孔。使用75ug/mL的G418进行选择。I12rg gRNA2具有GAAGCTCTTCTATACAATCTGG的序列(SEQ ID NO:91)且靶向大鼠I12rg外显子1的区190bp 3'。I12rg gRNA4具有CCCCGAAAGGAGGAGCCCTAGG的序列(SEQ ID NO:92)且靶向大鼠I12rg终止密码子的区80bp 5' (TGA) (参见图49)。

[1033] 表31. 在有和没有CRISPR/Cas9的情况下I12rg胞外结构域人源化载体靶向的比较

	载体	Cas9	gRNA	筛选的集落	靶向的克隆	靶向效率
[1034]	I12rg 质粒载体	否	否	77	46	60%
	I12rg 质粒载体	是	I12rg gRNA2	84	54	64%
[1035]	I12rg 质粒载体	是	I12rg gRNA4	88	50	57%

[1036] 4.4 (a) .具有同时人类基因替换的大非人类动物基因缺失的通过CRISPR/Cas9核酸内切酶实现的增强靶向

[1037] 例如全人抗体的用于人类疾病病状的新近研发的药物对于在人类细胞和组织中的靶常具有高度特异性且并不识别在啮齿动物中的同源靶。该高水平的选择性使得药物在人类中首次使用之前不可能试验其在啮齿动物中的功效和作用机制。

[1038] 该问题的一种非常有效的解决方案是产生基因修饰的小鼠或大鼠,其中编码药物靶的人类基因替换啮齿动物同源物。在啮齿动物中产生这一人源化等位基因的一种方法是首先在胚胎干(ES)细胞中缺失啮齿动物基因,且随后在第二基因修饰事件中在缺失的基因座处精确地插入人类基因。随后将ES细胞注射到啮齿动物胚胎中,并且植入代孕母体啮齿动物的子宫中,其接着生出携带人源化等位基因的基因修饰的幼仔。

[1039] 产生人源化基因修饰的一种更有效的方法是使用同时定向啮齿动物基因的缺失和用其人类对应物替换的大靶向载体(LTVEC)。通过采用VELOCIGENE<sup>®</sup>基因工程方法,当啮齿动物基因缺失和人类基因插入小于约20千碱基对(kb)时,这一单步骤人源化可在相对较高的效率下实现。要求大于100kb的缺失和替换的较大单步人源化可用LTVEC和例如

VELOCIGENE<sup>®</sup>基因工程方法的基因工程方法实现,但是由于有时与非常大的修饰冲突的较低靶向效率,成功常需要筛选或成百上千个ES细胞克隆来寻找携带所要基因修饰的克隆。

[1040] 为了改进大人源化的效率,我们已经研发了组合LTVEC基因靶向与成簇规律间隔短回文重复序列RNA-向导Cas9核酸内切酶(CRISPR/Cas9)的方法。CRISPR/Cas9核酸酶为由结合CRISPR RNA的细菌Cas9 DNA核酸内切酶构成的核糖核蛋白酶,所述细菌Cas9 DNA核酸内切酶向导Cas9以通过在向导RNA和靶向DNA的一个链之间的沃森-克里克(Watson-Crick)碱基配对在特异性DNA序列处裂解。由于所述靶向机制的简单性,设计定向在几乎任何基因组基因座处的双链断裂的CRISPR/Cas9核酸内切酶是容易的。双链断裂通过非同源端接合(NHEJ)途径诱导细胞基因组修复,其易于出错且常在双链断裂的位点处产生缺失或插入。修复双链断裂的替代机制为同源性定向修复(HDR),其中与断裂位点共有序列同一性或相似性的内源或外源段的DNA在细胞同源重组机构的作用下准确无误地修复断裂端。HDR可产生修复在断裂位点处的原始序列的完全修复,或者其可用以定向设计的修饰,例如在双链断裂的位点处序列的缺失、插入或替换。CRISPR/Cas9核酸酶可通过定向在预定基因修饰的位点处的精确双链裂解大大增强工程化HDR事件的速率。

[1041] 为了实现啮齿动物基因的全部或部分的精确、单步缺失和用其人类同源物的全部或部分同时替换,我们通过电穿孔向啮齿动物ES细胞中引入三种核酸分子:(1)LTVEC;(2)编码Cas9核酸内切酶的质粒或mRNA;和(3)编码CRISPR单向导RNA(sgRNA)或sgRNA本身的质粒。LTVEC包含编码由设计成定向缺失啮齿动物基因并插入人类基因的HR事件的啮齿动物DNA的同源臂侧接的基因产物(蛋白或RNA)的人类基因的全部或部分。人源化LTVEC还携带定向赋予抗生素药物(例如,G418)抗性的酶(例如,新霉素磷酸转移酶)的表达的药物选择盒。吸收LTVEC并将其并入其基因组中的ES细胞能够在皮氏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。因为我们引入与LTVEC分子相比多500-1,000倍的编码CRISPR/Cas9的核酸分子,但是大部分的含LTVEC的抗药性集落也至少短暂地含有CRISPR/Cas9组分。我们挑选抗药性集落并通过等位基因丢失方法对其加以筛选(Valenzuela,D.等人,(2003)High-throughput engineering of the mouse genome coupled with high-resolution expression analysis,Nature Biotech.21:652-660;Frendewey,D.等人,(2010)The loss-of-allele assay for ES cell screening and mouse genotyping,Methods Enzymol.476:295-307;其整体以引用的方式并入本文中)以鉴定具有正确靶向的人源化等位基因的克隆。

[1042] 在一种特定的实验中,LTVEC设计用来产生小鼠Lrp5(低密度脂蛋白受体-相关蛋白5)基因的68kb缺失和用同源人类LRP5基因的91kb片段的同时替换(图34)。LTVEC包含由含有来源于侧接意欲缺失的小鼠Lrp5基因的68kb序列的小鼠Lrp5基因座的部分的7kb和33kb基因组DNA的同源臂侧接的人类LRP5基因的91-kb片段。在单独的实验中,我们组合Lrp5人源化LTVEC与编码Cas9的质粒和编码8种sgRNA(gA、gB、gB2、gC、gD、gE2、gE、gF)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠Lrp5基因的区内产生双链断裂。sgRNA设计用来避免在人类LRP5基因的插入部分中的任何序列的识别。

[1043] Lrp5基因的CRISPR/Cas9辅助的人源化的结果示于表32中。当仅将LTVEC引入ES细胞中时,我们发现1.0%的筛选出的抗药性克隆携带正确靶向的单等位基因杂合人源化等



位基因。相比之下,组合LTVEC与由8种试验的sgRNA中的7种(sgRNA-5'A、sgRNA-5'B、sgRNA-5'B2、sgRNA-C、sgRNA-D、sgRNA-3'E2和sgRNA-3'F;序列提供于表33中)向导的Cas9核酸内切酶以2.1-7.3%的效率产生正确靶向的单等位基因杂合突变,这代表着与未受帮助的LTVEC相比较单步人源化基因靶向增强2-9倍。除了单等位基因靶向之外,对于由sgRNA-5'B2引起的Cas9-向导裂解,我们检测到以1%的频率的双等位基因纯合人源化。纯合Lrp5人源化ES细胞可通过VELOCIMOUSE<sup>®</sup>基因工程方法(Poueymirou,W.T.等人,(2007) F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses, Nature Biotech.25:91-99,其整体以引用的方式并入本文中)直接转变成准备用于表型和药物功效研究的完全源自ES细胞的小鼠。

[1044] 表32.Lrp5基因的CRISPR/Cas9辅助人源化的筛选结果。

实验	筛选的克隆	CRISPR活性(%)	单等位基因杂合突变频率(%)	双等位基因复合杂合突变频率(%)	双等位基因纯合突变频率(%)
仅 LTVEC	96	N/A	1.0 (1/96)	0	0
LTVEC + Cas9 + sgRNA-5'A	96	75.6	7.3 (7/96)	0	0
LTVEC + Cas9 + sgRNA-5'B	96	79.5	4.2 (4/96)	0	0
LTVEC + Cas9 + sgRNA-5'B2	96	60.5	6.2 (6/96)	0	1.0 (1/96)
LTVEC + Cas9 + sgRNA-C	96	未测定	4.2 (4/96)	0	0
LTVEC + Cas9 + sgRNA-D	96	未测定	7.3 (7/96)	0	0
LTVEC + Cas9 + sgRNA-3'E2	96	84.5	2.1 (2/96)	0	0
LTVEC + Cas9 + sgRNA-3'E	96	52.4	0	0	0
LTVEC + Cas9 + sgRNA-3'F	96	79.8	6.2 (6/96)	0	0

[1046] 表33.靶向小鼠Lrp5基因的六种sgRNA的向导部分的序列。

	sgRNA	距缺失终点的近似距离(bp)	向导序列(5'至3')
[1047]	sgRNA-5'A	50	GGGAACCCACAGCATACTCC(SEQ ID NO: 24)
	sgRNA-5'B	500	GAATCATGCACGGCTACCCC(SEQ ID NO: 25)
	sgRNA-5'B2	1000	TGCTCCTATGGGGAGGCGCG(SEQ ID NO: 26)
	sgRNA-C	29900/ 38430	ACTGAGATCAATGACCCCGA(SEQ ID NO: 85)
	sgRNA-D	29950/ 38380	GGGTCGCCCGGAACCTCTAC(SEQ ID NO: 86)
[1048]	sgRNA-3'E2	1000	CTTGGATAACATTGATACCC(SEQ ID NO: 27)
	sgRNA-3'E	500	GGGGCAGAGCCCTTATATCA(SEQ ID NO: 28)
	sgRNA-3'F	50	TCGCTCACATTAATCCCTAG(SEQ ID NO: 29)

[1049] 当与用锌指核酸酶 (ZFN) 执行的等价实验相比较时,通过CRISPR/Cas9核酸内切酶实现的大Lrp5人源化的增强靶向是显著的。我们获得设计用以在针对缺失靶向的小鼠Lrp5基因的区内的位点处生成双链断裂的四种ZFN(图34)。一种ZFN靶向在缺失的5'端附近的序列(a),一种ZFN靶向在缺失中间的序列(b)且两种ZFN靶向在缺失的3'端附近的序列(c,d)。在单独的实验中,我们组合Lrp5人源化LTVEC与编码设计用来在针对缺失靶向的小鼠Lrp5基因的区内产生双链断裂的四种ZFN(a-d)中的一种的质粒。我们确定所有ZFN都具有活性且能够诱导在Lrp5基因中的NHEJ突变(数据没有示出),但当与LTVEC组合时,HDR介导的基因靶向与仅LTVEC相比较没有增强。

[1050] 当与一系列ZFN辅助的人源化实验相比较时,由CRISPR/Cas9核酸内切酶实现的大Lrp5人源化的增强的靶向效率也是显著的。在这些实验中,执行一系列ZFN辅助的人源化,其中小鼠靶基因缺失和人类基因插入通常具有增加的尺寸(表34;图35)。图35A描绘在增加的缺失尺寸下LTVEC靶向基因的%靶向效率。LTVEC单独使用(灰色方块)或与ZFN组合使用(黑色方块)。图35B描绘在增加尺寸的人类基因插入下LTVEC的%靶向效率。再次,LTVEC单独使用(灰色三角形)或与ZFN组合使用(黑色三角形)。如在表34和图35中所示,当小鼠靶基因缺失的尺寸大于24.7kb时和当人类基因插入的尺寸大于22.2kb时,ZFN介导DNA裂解以增强LTVEC靶向效率的能力消失(表34;图35A)。相比之下,CRISPR/Cas9能够增强Lrp5基因的LTVEC靶向效率,其涉及68.3kb的小鼠基因缺失和91.0kb的人类基因插入(表32;图34)。这指示,在其它核酸酶(例如,锌指核酸酶)不能增强LTVEC靶向效率的情形下,CRISPR/Cas9核酸内切酶能够增强LTVEC靶向效率。

[1051] 表34. ZFN辅助人源化的概述

靶向基因	小鼠基因缺失 (kb)	人类基因插入 (kb)	5'同源臂 (kb)	3'同源臂 (kb)	ZFN裂解效率 (%)	靶向效率 (%)		增强倍数
						仅LTVEC	LTVEC + ZFN	
<i>Fcer1a</i>	4.1	4.2	10.9	76.8	22.9	5.20	32.81	6.3
<i>Tlr4</i>	7.1	5.7	67.6	85.5	12.5	5.20	22.39	4.3
<i>Prrr</i>	8.7	18.0	49.6	112.9	30.7	1.56	24.48	15.7
<i>Notch4</i>	18.7	22.2	50.1	34.9	27.1	10.41	12.50	1.2
<i>Accn2</i>	24.7	18.9	57.8	60.1	20.8	4.17	8.33	2.0
<i>Adamts5</i>	37.6	39.6	83.3	61.5	4.2	0.00	0.00	n.a.
<i>Trpa1</i>	45.3	53.9	41.3	57.8	8.8	0.52	0.00	0
<i>Folh1</i>	55.1	61.3	18.4	114.7	8.8	1.04	0.00	0
<i>Lrp5</i>	68.3	91.0	6.9	33.4	35.9	2.08	1.04	(0.5)
<i>ErbB4</i>	101.6	126.7	47.8	26.0	n.d.	0.00	0.00	n.a.

[1053] n.d. = 未测

[1054] n.a. = 不适用

[1055] () = 与没有ZFN的情况相比,在用ZFN的情况下靶向效率较低

[1056] 对于其它小鼠基因的人源化,执行可比较的实验。在一个实验中,LTVEC设计用来产生小鼠Trpa1(瞬时受体电位阳离子通道亚族A成员1)基因的45kb缺失和用同源人类TRPA1基因的55kb片段的同时替换(图36)。LTVEC包含由含有来源于侧接意欲缺失的小鼠Trpa1基因的45kb序列的小鼠Trpa1基因座的部分的41kb和58kb的基因组DNA的同源臂侧接的人类TRPA1基因的55kb片段。在单独的实验中,我们组合Trpa1人源化LTVEC与编码Cas9的质粒和编码8种sgRNA(gA、gA2、gB、gC、gD、gE、gE2和gF)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠Trpa1基因的区内产生双链断裂。sgRNA设计用来避免在人类TRPA1基因的插入部分中的任何序列的识别。

[1057] Trpa1基因的CRISPR/Cas9辅助的人源化的结果示于表35中。当仅将LTVEC引入ES细胞中时,我们发现1.0%的筛选出的抗药性克隆携带正确靶向的单等位基因杂合人源化等位基因。相比之下,组合LTVEC与由8种试验的sgRNA中的6种(A、A2、B、C、D和F;序列提供于表43中)向导的Cas9核酸内切酶以1.0-3.1%的效率产生正确靶向的单等位基因杂合突变或双等位基因复合杂合或纯合突变。对于由gRNA A和gRNA F实现的Cas9向导裂解,我们检测到以1.0%的频率的复合杂合突变。

[1058] 表35.Trpa1基因的CRISPR/Cas9辅助人源化的筛选结果。

sgRNA 位置	距缺失终点的近似距离(bp)	gRNA	CRISPR 活性(%)	筛选的克隆	杂合靶向的	复合杂合	纯合靶向
5'	100	gRNA A	30.9	96	0	1	0
5'	500	gRNA A2	未测定	96	2	0	0
5'	1000	gRNA B	42.8	96	3	0	0
[1059] 中间	25600 / 19740	gRNA C	未测定	96	1	0	0
中间	26970 / 18370	gRNA D	未测定	96	2	0	0
3'	1000	gRNA E2	未测定	96	0	0	0
3'	500	gRNA E	22.6	96	0	0	0
3'	100	gRNA F	28.6	96	1	1	0
N/A	N/A	无	N/A	96	1	0	0

[1060] 在另一实验中,LTVEC设计用来产生小鼠Folh1(谷氨酸羧肽酶2)基因的55kb缺失和用同源人类FOLH1基因的61kb片段的同时替换(图37)。LTVEC包含由含有来源于侧接意欲缺失的小鼠Folh1基因的55kb序列的小鼠Folh1基因座的部分的22kb和46kb的基因组DNA的同源臂侧接的人类FOLH1基因的61kb片段。在单独的实验中,我们组合Folh1人源化LTVEC与编码Cas9的质粒和编码6种sgRNA(gA、gA2、gC、gD、gE和gE2)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠Folh1基因的区内产生双链断裂。sgRNA设计用来避免在人类FOLH1基因的插入部分中的任何序列的识别。

[1061] Folh1基因的CRISPR/Cas9辅助的人源化的结果示于表36中。当仅将LTVEC引入ES细胞中时,我们发现96个筛选的抗药性克隆中一个也没有携带正确靶向的单等位基因杂合人源化等位基因。相比之下,组合LTVEC与由6种试验的sgRNA中的3种(A、D和E2;序列提供于表43中)向导的Cas9核酸内切酶以1.0-3.1%的效率产生正确靶向的单等位基因杂合突变。

[1062] 表36.Folh1基因的CRISPR/Cas9辅助人源化的筛选结果。

sgRNA 位置	距缺失终点的近似距离(bp)	gRNA	CRISPR 活性(%)	筛选的克隆	杂合靶向的	复合杂合	纯合靶向
5'	100	gRNA A	45.2	96	2	0	0
5'	500	gRNA A2	61.9	96	0	0	0
[1063] 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'	500	gRNA E2	未测定	96	1	0	0
3'	100	gRNA E	1.2	96	0	0	0
N/A	N/A	无	N/A	96	0	0	0

[1064] 在另一实验中,LTVEC设计用来产生对于互补组分5(C5或Hc)的小鼠基因的76kb缺

失和用同源人类C5基因的97kb片段的同时替换(图38)。LTVEC包含由含有来源于侧接意欲缺失的小鼠C5(Hc)基因的76kb序列的小鼠C5(Hc)基因座的部分的34.1kb和31.2kb的基因组DNA的同源臂侧接的人类C5基因的97kb片段。在单独的实验中,我们组合C5(Hc)人源化LTVEC与编码Cas9的质粒和编码6种sgRNA(gA、gB、gC、gD、gE和gE2)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠C5(Hc)基因的区内产生双链断裂。sgRNA设计用来避免在人类C5基因的插入部分中的任何序列的识别。

[1066] C5(Hc)基因的CRISPR/Cas9辅助的人源化的结果示于表37中。当仅将LTVEC引入ES细胞中时,我们发现1.0%的筛选出的抗药性克隆携带正确靶向的单等位基因杂合人源化等位基因。相比之下,组合LTVEC与由所有6种试验的sgRNA(A、B、C、D、E和E2;序列提供于表43中)向导的Cas9核酸内切酶以4.2-16.7%的效率产生正确靶向的单等位基因杂合突变或双等位基因复合杂合或纯合突变。对于gRNA A和gRNA E实现的Cas9-向导的裂解,我们检测到分别以5.2%和4.2%的频率的复合杂合突变。

[1066] 表37. C5(Hc)基因的CRISPR/Cas9辅助的人源化的筛选结果。

sgRNA 位置	距缺失终点的近似距离(bp)	gRNA	CRISPR 活性(%)	筛选的克隆	杂合靶向的	复合杂合	纯合靶向
5'	100	gRNA A	64.3	96	11	5	0
5'	500	gRNA B	72.6	96	14	0	0
[1067] 中间	38200 / 37500	gRNA C	47.6	96	11	0	0
中间	43500 / 32200	gRNA D	47.6	96	7	0	0
3'	500	gRNA E	25.0	96	0	4	0
3'	100	gRNA E2	27.4	96	6	0	0
N/A	N/A	无	N/A	96	1	0	0

[1068] 在另一实验中,LTVEC设计用来产生小鼠Adamts5(具有凝血蛋白基序5的去整合素和金属蛋白酶)基因的38kb缺失和用同源人类ADAMTS5基因的43kb片段的同时替换(图39)。LTVEC包含由含有来源于侧接意欲缺失的小鼠Adamts5基因的38kb序列的小鼠Adamts5基因座的部分的22kb和46kb的基因组DNA的同源臂侧接的人类ADAMTS5基因的43kb片段。在单独的实验中,我们组合Adamts5人源化LTVEC与编码Cas9的质粒和编码8种sgRNA(gA、gA2、gB、gC、gD、gE、gE2和gF)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠Adamts5基因的区内产生双链断裂。sgRNA设计用来避免在人类ADAMTS5基因的插入部分中的任何序列的识别。

[1069] Adamts5基因的CRISPR/Cas9辅助的人源化的结果示于表38中。当仅将LTVEC引入ES细胞中时,我们发现96个筛选的抗药性克隆中一个也没有携带正确靶向的单等位基因杂合人源化等位基因。相比之下,组合LTVEC与由8种试验的sgRNA中的2种(B和F;序列提供于表43中)向导的Cas9核酸内切酶以1.0%的效率产生正确靶向的单等位基因杂合突变或双等位基因复合杂合突变。对于由gRNA E2实现的Cas9向导裂解,我们检测到以1.0%的频率的复合杂合突变。

[1070] 表38. Adamts5基因的CRISPR/Cas9辅助人源化的筛选结果。

[1071]

sgRNA 位置	距缺失终点的近似距离(bp)	gRNA	CRISPR 活性 (%)	筛选的克隆	杂合靶向的	复合杂合	纯合靶向
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
N/A	N/A	无	N/A	96	0	0	0

[1072] 在另一实验中,LTVEC设计用来产生小鼠ErbB4(受体酪氨酸-蛋白激酶erbB-4)基因的102kb缺失和用同源人类ERBB4基因的127kb片段的同时替换(图40)。LTVEC包含由含有来源于侧接意欲缺失的小鼠ErbB4基因的102kb序列的小鼠ErbB4基因座的部分的48kb和26kb的基因组DNA的同源臂侧接的人类ERBB4基因的127kb片段。在单独的实验中,我们组合ErbB4人源化LTVEC与编码Cas9的质粒和编码8种sgRNA(gA、gB、gB2、gC、gD、gE、gE2和gF)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠ErbB4基因的区内产生双链断裂。sgRNA设计用来避免在人类ERBB4基因的插入部分中的任何序列的识别。

[1073] ErbB4基因的CRISPR/Cas9辅助的人源化的结果示于表39中。当仅将LTVEC引入ES细胞中时,我们发现96个筛选的抗药性克隆中一个也没有携带正确靶向的单等位基因杂合人源化等位基因。相比之下,组合LTVEC与由8种试验的sgRNA中的一种(D;序列提供于表43中)向导的Cas9核酸内切酶以1.0%的效率产生正确靶向的单等位基因杂合突变或双等位基因复合杂合突变。对于由gRNA D实现的Cas9向导裂解,我们检测到以1%的频率的复合杂合突变。

[1074] 表39.ErbB4基因的CRISPR/Cas9辅助人源化的筛选结果。

[1075]

sgRNA 位置	距缺失 终点的近似 距离(bp)	gRNA	CRISPR 活性(%)	筛选的 克隆	杂合 靶向的	复合 杂合	纯合 靶向
5'	100	gRNA A	25.0	96	0	0	0
5'	500	gRNA B	未测定	96	0	0	0
5'	1000	gRNA B2	47.6	96	0	0	0
中间	50200 / 51350	gRNA C	20.2	96	0	0	0
中间	50230 / 51320	gRNA D	42.8	96	0	1	0
3'	1000	gRNA F	15.5	96	0	0	0
3'	500	gRNA E	89.2	96	0	0	0
3'	100	gRNA E2	14.3	96	0	0	0
N/A	N/A	无	N/A	96	0	0	0

[1076] 在另一实验中,LTVEC设计用来产生小鼠Ror1(酪氨酸-蛋白激酶跨膜受体ROR1)基因的110kb缺失和用同源人类ROR1基因的134kb片段的同时替换(图41)。LTVEC包含由含有来源于侧接意欲缺失的小鼠Ror1基因的110kb序列的小鼠Ror1基因座的部分的41.8kb和96.4kb的基因组DNA的同源臂侧接的人类ROR1基因的134kb片段。在单独的实验中,我们组合Ror1人源化LTVEC与编码Cas9的质粒和编码6种sgRNA(gA、gB、gC、gD、gE和gF)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠Ror1基因的区内产生双链断裂。sgRNA设计用来避免在人类ROR1基因的插入部分中的任何序列的识别。

[1077] Ror1基因的CRISPR/Cas9辅助的人源化的结果示于表40中。当仅将LTVEC引入ES细胞中时,我们发现96个筛选的抗药性克隆中一个也没有携带正确靶向的单等位基因杂合人源化等位基因。I相比之下,组合LTVEC与由6种试验的sgRNA中的2种(D和F;序列提供于表43中)向导的Cas9核酸内切酶以1.0%的效率产生正确靶向的单等位基因杂合突变或双等位基因突变。对于由gRNA F实现的Cas9向导裂解,我们也检测到以1%的频率的复合杂合突变。

[1078] 表40.Ror1基因的CRISPR/Cas9辅助人源化的筛选结果。

sgRNA 位置	距缺失终点的近似距离 (bp)	gRNA	CRISPR 活性(%)	筛选的克隆	杂合靶向的	复合杂合	纯合靶向
5'	200	gRNA A	未测定	96	0	0	0
5'	1000	gRNA B	未测定	96	0	0	0
[1079] 中间	54300 / 55500	gRNA D	未测定	96	1	0	0
中间	54500 / 55300	gRNA C	未测定	96	0	0	0
3'	1000	gRNA E	未测定	96	0	0	0
3'	200	gRNA F	未测定	96	0	1	0
N/A	N/A	无	N/A	96	0	0	0

[1080] 在另一实验中,LTVEC设计用来产生小鼠Dpp4(二肽基肽酶4)基因的79kb缺失和用同源人类DPP4基因的82kb片段的同时替换(图42)。LTVEC包含由各自含有来源于侧接意欲缺失的小鼠Dpp4基因的79kb序列的小鼠Dpp4基因座的部分的46kb的基因组DNA的5'同源臂和3'同源臂侧接的人类DPP4基因的82kb片段。在单独的实验中,我们组合Dpp4人源化LTVEC与编码Cas9的质粒和编码8种sgRNA(gA、gB、gB2、gC、gD、gE、gE2和gF)中的一种的第二质粒,设计用来在针对缺失靶向的小鼠Dpp4基因的区内产生双链断裂。sgRNA设计用来避免在人类DPP4基因的插入部分中的任何序列的识别。

[1081] Dpp4基因的CRISPR/Cas9辅助的人源化的结果示于表41中。当仅将LTVEC引入ES细胞中时,我们发现2.1%的筛选的抗药性克隆携带正确靶向的单等位基因杂合人源化等位基因。相比之下,组合LTVEC与由8种试验的sgRNA(A、B、B2、C、D、E、E2和F;序列提供于表43中)中的任一种向导的Cas9核酸内切酶以2.1-7.3%的效率产生正确靶向的单等位基因杂合突变。

[1082] 表41.Dpp4基因的CRISPR/Cas9辅助人源化的筛选结果。



[1083]

sgRNA 位置	距缺失终点的近似距离(bp)	gRNA	CRISPR 活性(%)	筛选的克隆	杂合靶向的	复合杂合	纯合靶向
5'	50	gRNA A	未测定	96	7	0	0
5'	400	gRNA B	未测定	96	2	0	0
5'	900	gRNA B2	未测定	96	5	0	0
中间	38800 / 40200	gRNA C	未测定	96	3	0	0
中间	40800 / 38100	gRNA D	未测定	96	3	0	0
3'	900	gRNA E2	未测定	96	2	0	0
3'	500	gRNA E	未测定	96	6	0	0
3'	200	gRNA F	未测定	96	5	0	0
N/A	N/A	无	N/A	96	2	0	0

[1084] 概述各种小鼠基因的CRISPR/Cas9辅助的人源化的结果的表提供于表42中。第一行指示靶向的基因基因座。第二行指示内源小鼠基因座的缺失尺寸 (Del) 和相应人类基因座的插入尺寸 (Ins)。剩余行示出对于具有正确靶向的单等位基因杂合突变、双等位基因复合杂合突变或双等位基因纯合突变的各情况的集落数目 (在96个集落之中)。“无gRNA”代表仅LTVEC, 而其它行代表LTVEC加相应gRNA (由在缺失基因座内的相对位置指示)。

[1085] 表42. 小鼠基因的CRISPR/Cas9辅助人源化的概述。

[1086]

	<i>Lrp5</i>	<i>Trpa1</i>	<i>Folh1</i>	<i>C5(Hc)</i>	<i>Adamts5</i>	<i>ErbB4</i>	<i>Ror1</i>	<i>Dpp4</i>
<b>Del/Ins(kb)</b>	68/91	45/55	55/61	76/97	38/43	102/127	110/134	79/82
<b>大部分 5'</b>	7	1	2	16	0	0	0	7
<b>5'</b>	4	2	0	14	0	0	0	2
<b>5'</b>	7	3	N/A	N/A	1	0	N/A	5
<b>中间</b>	4	1	0	11	0	0	1	3
<b>中间</b>	7	2	1	7	0	1	0	3
<b>3'</b>	2	0	N/A	N/A	1	0	N/A	2
<b>3'</b>	0	0	1	4	0	0	0	6
<b>大部分 3'</b>	6	2	0	6	0	0	1	5
<b>无 gRNA</b>	1	1	0	1	0	0	0	2

[1087] 表43. 用于小鼠基因的CRISPR/Cas9辅助人源化的向导RNA序列

[1088]

<b>gRNA</b>	<b>向导序列(5'至3')</b>	<b>SEQ ID NO</b>
<i>Trpa1</i> gRNA A	GTACTGGGGAATCGGTGGTC	30
<i>Trpa1</i> gRNA A2	CACGCACTCCAAATTTATCC	31
<i>Trpa1</i> gRNA B	CTAAGTGTGTATCAGTACAT	32
<i>Trpa1</i> gRNA C	TGCCCTGCACAATAAGCGCA	33
<i>Trpa1</i> gRNA D	ACTCATTGAAACGTTATGGC	34
<i>Trpa1</i> gRNA E2	AGTAAGGGTGGATTAATTC	35
<i>Trpa1</i> gRNA E	GCCATCTAGATTCATGTAAC	36
<i>Trpa1</i> gRNA F	GACTAGAAATGTTCTGCACC	37
<b> </b>		
<i>Folh1</i> gRNA A	TGAACCAATTGTGTAGCCTT	38
<i>Folh1</i> gRNA A2	AATAGTGGTAAAGCACCATG	39
<i>Folh1</i> gRNA B	GTGTGCTAAGGATCGAAGTC	40
<i>Folh1</i> gRNA C	CACCGAGATGCTTGGGTATT	41
<i>Folh1</i> gRNA D	TGTAACCGCCCTGAATGACC	42
<i>Folh1</i> gRNA E	AAAAGGGCATCATAAATCCC	43
<i>Folh1</i> gRNA E2	TCAAAAATAGTCATACACCT	44
<i>Folh1</i> gRNA F	GGTCTCTAGTACATTGTAGA	45
<b> </b>		
<i>C5 (Hc)</i> gRNA A	ATCACAAACCAGTTAACCGG	46
<i>C5 (Hc)</i> gRNA B	TTTCAGACGAGCCGACCCGG	47
<i>C5 (Hc)</i> gRNA B2	CTGTCAACAGTGCCGCGTTT	48
<i>C5 (Hc)</i> gRNA C	TGTGTGTCATAGCGATGTCG	49
<i>C5 (Hc)</i> gRNA D	AACAGGTACCCTATCCTCAC	50
<i>C5 (Hc)</i> gRNA E2	TCGTGGTTGCATGCGCACTG	51
<i>C5 (Hc)</i> gRNA E	GGCCCGGACCTAGTCTCTCT	52
<i>C5 (Hc)</i> gRNA F	AGTCTGTAAAGTTAGCAGTC	53
<b> </b>		
<i>Adamts5</i> gRNA A	GGTGGTGGTGCTGACGGACA	54
<i>Adamts5</i> gRNA A2	TATGAGATCAACACTCGCTA	55
<i>Adamts5</i> gRNA B	CCAAGGACTTCCCCACGTTA	56
<i>Adamts5</i> gRNA C	TGCTTCCCTTATGCAAGATT	57
<i>Adamts5</i> gRNA D	TTAGGTACCCTATTTGAATA	58
<i>Adamts5</i> gRNA E2	TGCAGTGGGTGACAGGTCCA	59
<i>Adamts5</i> gRNA E	AGGGTTATACTGACGTTGTG	60
<i>Adamts5</i> gRNA F	TGCTTTCAAGGAGGGCTAC	61

	<i>ErbB4</i> gRNA A	TGATGTGCAGTCAGACAAAG	62
	<i>ErbB4</i> gRNA B	TGCACTATGGTTGACTATGA	63
	<i>ErbB4</i> gRNA B2	GGAATATTCTAATAGGAAGT	64
	<i>ErbB4</i> gRNA C	AAGTGCTGTACCATTCTAGC	65
	<i>ErbB4</i> gRNA D	TAATCAATAGACAACCTCGT	66
	<i>ErbB4</i> gRNA E2	TCATTCCTAATGGTATTATA	67
	<i>ErbB4</i> gRNA E	AGGGTACATAGATGGCATCG	68
	<i>ErbB4</i> gRNA F	CTCTTTAACAATTACCACTT	69
	<i>Ror1</i> gRNA A	TGTGGGCCTTTGCTGATCAC	70
	<i>Ror1</i> gRNA B	AATCTATGATCCTATGGCCT	71
[1089]	<i>Ror1</i> gRNA D	TGCCAATAGCAGTGACTTGA	72
	<i>Ror1</i> gRNA C	GGGAAGAATGGGCTATTGTC	73
	<i>Ror1</i> gRNA E	GGTTGTTTGTGCTGATGACG	74
	<i>Ror1</i> gRNA F	CCGTCTAGGCCTTCTACGT	75
	<i>Dpp4</i> gRNA A	ACTAGTAGACCTGAGGGGTT	76
	<i>Dpp4</i> gRNA B	GCTCCAGTGTTTAGGCCTTG	77
	<i>Dpp4</i> gRNA B2	GGCAAGCTGAAAACGCATGC	78
	<i>Dpp4</i> gRNA C	GTAGATCGCTTTCCACTACC	79
	<i>Dpp4</i> gRNA D	GAACTCCACTGCTCGTGAGC	80
	<i>Dpp4</i> gRNA E2	ATAGGTGGGCACTATTGAAG	81
	<i>Dpp4</i> gRNA E	ATGGGAAGGTTTATACCAGC	82
	<i>Dpp4</i> gRNA F	CGGTGTAAAAACAACGGGAA	83

[1090] 实施例5.大鼠基因组基因座的靶向修饰的概述

[1091] 表44.用在实施例3和4中论述的各种载体类型和核酸酶试剂进行的大鼠靶向的概述

[1092]

表 44. 大鼠靶向概述

实施例编号	基因座	载体	筛选的集落	靶向的克隆	靶向效率	双等位基因靶向	双等位基因效率	注射的克隆	产生嵌合体的克隆	经由种系传递的克隆	注释
3.2(a) (ii)	ApoE	质粒	384	23	5.99%	0	0	3	3	1	
3.2(a) (iii)	ApoE + ZFN	质粒	384	290	75.52%	8	2.08%	2	2	1	这两个克隆为双等位基因靶向的
3.3(a)	Il2rg	质粒	232	5	2.16%	N/A	N/A	6	5	1	
3.2(b) (ii)	ApoE LTVEC	LTVEC	288	8	2.78%	1	0.35%	3	1	0	
3.2(b) (iii)	ApoE LTVEC + ZFN	LTVEC	288	16	5.56%	1	0.35%	1	N/A	0	该克隆为双等位基因靶向的
3.2(b) (iv)	ApoE LTVEC	LTVEC	75	32	42.67%	0	0				
3.2(b) (iv)	ApoE LTVEC + CRISPR/Cas9	LTVEC	169	84	50%	5	3%	0	0	0	
4.3(a)	Il2rg 人源化 1	质粒	168	6	3.57%	N/A	N/A	1	1	0	用人类 Il2rg 替换整体大鼠 Il2rg
4.3(b) (i)	Il2rg 人源化 2	质粒	192	13	6.77%	N/A	N/A	2	2	0	用人类 Il2rg 胞外结构域替换大鼠 Il2rg 胞外结构域
4.3(b) (ii)	Il2rg 人源化 2	质粒	77	46	59.74%						
4.3(b) (ii)	Il2rg 人源化 2 + CRISPR/Cas9	质粒	172	104	60.47%	N/A	N/A	0	0	0	用人类 Il2rg 胞外结构域替换大鼠 Il2rg 胞外结构域

[1093]

3.4(a) (i)	Rag2	LTVE C	270	0	0	0	0	0	0	0	0	0	外结构域 预测的 5.7 KB 缺失
3.4(a) (ii)	Rag2 LTVEC	LTVE C	36	0	0	0	0	0	0	0	0	0	
3.4(a) (ii)	Rag2 LTVEC + CRISPR/Cas9	LTVE C	39	6	15.38%	1	2.5%	1	1	1	0	0	
3.4(b) (i)	Rag1-2	LTVE C	256	1	0.39%	0	0	1	1	1	0	0	预测的 16.2 kb 缺失
3.4(b) (ii)	Rag1-2	LTVE C	94	80	85%	0	0	0	0	0	0	0	将 Rag1-2 LTVEC 电穿孔 到 Il2rg-CG12 突 变体克隆中(再 靶向)

表 45 示出用质粒或 LTVEC 与 CRISPR/Cas9 组合靶向大鼠 ES 细胞的概述。两种 gRNA 对于各靶基因座: *Rag2*、*ApoE* 和 *Il2rg* 单独地试验。在所有三种基因座处, CRISPR/Cas9 的裂解效率 > 20%。当 CRISPR/Cas9 与靶向质粒和 LTVEC 组合使用时, 观察到增加的靶向效率和增加的双等位基因靶向。

表 45. 与 CRISPR/Cas9 组合用质粒或 LTVEC 进行的大鼠 ES 细胞靶向的概述

条件	靶向效率	双等位基因靶向
Rag2(LTVEC)	0	0
Rag2(LTVEC+ CRISPR)	6-22%	0-4%
ApoE(LTVEC)	43%	0
ApoE(LTVEC + CRISPR)	47-53%	1-4%
Il2rg 人源化(质粒载体)	60%	N/A(X-连锁)

[1094]

II2rg 人源化(质粒+CRISPR)	57-64%	N/A(X-连锁)
----------------------	--------	-----------

表 46 示出关于大鼠基因组座的靶向修饰的种系传递数据的概述。种系传递对于 *ApoE* 靶向的大鼠和 *II2rg* 靶向的大鼠证实。大鼠 ES 细胞为 XY(雄性)且杂合靶向的。因此，当靶向的 ES 细胞捐献种系时，约 50%的来源于 ES 细胞的精液将携带突变等位基因且将产生杂合的 F1 幼仔。

表 46. 大鼠基因组基因座的靶向修饰的种系传递数据

靶向的基因	微量注射的克隆	产生嵌合体的克隆	实现种系传递的克隆	种系幼仔/总幼仔	杂合 F1 幼仔*	缺失新盒的杂合幼仔
<i>ApoE</i>	3	3	1	7/79(9%)	4	4
<i>II2rg</i>	5	5	1	11/257(5%)	5	5

[1095] 实施例6.产生、维持并靶向人类诱导的多潜能干细胞

[1096] 6.1.人类iPS细胞的产生

[1097] 该实施例描述由非多潜能人类细胞产生人类iPS细胞。使用RED和BLUE GeneIn™转染试剂(GlobalStem)将PiggyBac(System Biosciences)载体(PB-600A\_CAGGS Bst XI (0.64μg/μL)和包含编码操作性连接CM7启动子的四种重新编程因子(hOct4、hSox2、hKLF-4、hMYC)的基因的PB-200(0.99μg/μL)引入初生儿人类包皮成纤维细胞中。转染的细胞在E7培养基中在NuFF1饲养细胞上孵育以允许载体并入和重新编程因子表达。E7培养基包含DMEM/F-12、NaHCO<sub>3</sub>、L-抗坏血酸、胰岛素、转铁蛋白、硒和FGF-2。

[1098] 嘌呤霉素选择在使用2μg/mL嘌呤霉素在E7培养基中转染之后10天开始。在第21天,选择集落并在mTeSR™培养基中培养,其包含DMEM/F-12、NaHCO<sub>3</sub>、L-抗坏血酸、胰岛素、转铁蛋白、硒、FGF-2、TGF-β1、谷胱甘肽、L-谷氨酰胺、限定的脂质、硫胺素、痕量元素B和C、β-巯基乙醇、牛血清蛋白、六氢吡啶羧酸、氯化锂和GABA。在第29-57天,细胞在mTeSR™培养基中繁殖并传代直至达到约50%的汇合度。在第65-73天,使用mTeSR™培养基和Gentle Cell Dissociation Reagent(Stem Cell Technologies)继续繁殖并传代。在第76天,培养基变为低渗透压度VG2i培养基以便进一步繁殖、传代并维持包含天然或天然外观的hiPSC的细胞。

[1099] 6.2. 在人类iPS细胞中LTVEC靶向

[1100] 该实施例描述在人类iPS细胞中使用LTVEC靶向。如在图51中所示,我们通过电穿孔到在VG2i培养基中繁殖的人类iPS细胞引入以下核酸分子:(1)LTVEC(0.67μg);(2)编码Cas9核酸内切酶的质粒(5μg);和(3)编码CRISPR单向导RNA(gRNA)的质粒(10μg)。在一组样品中,Cas9和gRNA排除在外。具体地讲,3x 10<sup>6</sup>个细胞在700V的电压、25uF的电容和400Ω的电阻下电穿孔。LTVEC包含包含由含有来源于侧接意欲缺失的人类ADAM6基因座的4.1kb序列的基因组区的34kb和105kb的基因组DNA的同源臂侧接的小鼠Adam6a和Adam6b基因的16.7kb核酸。LTVEC也携带定向赋予抗生素药物(潮霉素)抗性的酶的表达的药物的选择盒。所使用的人类ADAM6 gRNA具有以下序列:GTATAGCCCTGTTACACATT(SEQ ID NO:94)。

[1101] 吸收LTVEC并将其并入其基因组中的细胞能够在GELTREX™涂布的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。因为我们引入与LTVEC分子相比多500-1,000倍的编码CRISPR/Cas9的核酸分子,但是大部分的含LTVEC的抗药性集落也至少短暂地含有CRISPR/Cas9组分。我们挑选抗药性集落并通过等位基因丢失方法加以筛选(Valenzuela等人,(2003) Nat. Biotech. 21:652-660;Frendewey等人,(2010) Methods Enzymol. 476:295-307;其整体以引用的方式并入本文中)以鉴定具有正确靶向的等位基因的克隆。

[1102] ADAM6基因座的CRISPR/Cas9辅助的LTVEC靶向的结果示于表47中。

[1103] 表47. CRISPR/Cas9辅助的LTVEC靶向

[1104]

	靶向效率
仅LTVEC	3.1%
LTVEC+CRISPR	7.3%

[1105] 当仅将LTVEC引入人类iPS细胞中时,观察到3.1%的靶向效率。相比之下,组合LTVEC与由ADAM6 gRNA向导的Cas9产生7.3%的靶向效率。

[1106] 6.3. 低渗透压度培养基对人类iPS细胞形态的影响

[1107] 该实施例描述盐浓度、离子强度和/或渗透压度对在培养中的人类iPS细胞的多潜

能状态的影响。人类iPS细胞在表48中描述的培养基中或在mTeSR™-hLIF培养基中在MATRIGEL™或GELTREX™底物上培养。

[1108] 表48. 用于iPS细胞培养的培养基。

组分	量(v/v)
基础培养基	24.75
F-12 培养基	24.75
N2®补充剂	0.5
Neurobasal®培养基	49
B-27®补充剂	1
青霉素/链霉素	1
L-谷氨酰胺(200 mM)	1
2-巯基乙醇(55 mM)	0.1836
hLIF(1 x 10 <sup>4</sup> 单位/毫升)	0.001
CHIR99021(10 mM)	0.03
PD0325901(10 mM)	0.005

[1110] 当所使用的基础培养基为DMEM时,该培养基被称为2i培养基。当所使用的基础培养基为VG-DMEM时,该低渗透压度培养基被称为VG2i培养基。VG2i培养基的渗透压度(233mOsm/kg)低于传统2i培养基的渗透压度(261mOsm/kg)。

[1111] 如在图52中所示,在2i培养基中在MATRIGEL™上培养8天的时间(图52A)或12天的时间(图52B)的人类iPS细胞显示处于始发态的iPS细胞的形态特征,特别是在上皮单层中生长和顶面-底侧极性的外观。

[1112] mTeSR-hLIF培养基和VG2i培养基对于其对人类iPS细胞的形态和多潜能状态的影响进一步评价。在该研究中,人类iPS细胞在mTeSR™-hLIF培养基(图53A和53C)中或在VG2i培养基(图53B和53D)中在MATRIGEL™或NuFF饲养细胞上培养6天的时间。当在mTeSR™-hLIF培养基中在MATRIGEL™或NuFF饲养细胞上培养时,人类iPS细胞显示始发多潜能状态的形态特征,特别是在上皮单层中生长和顶面-底侧极性的外观。在mTeSR™-hLIF培养基中培养的一些细胞开始显示以三维结块为特征的形态。相反,当在VG2i培养基中在MATRIGEL™或NuFF饲养细胞上培养时,人类iPS细胞显示天然多潜能状态的形态特征,特别是在圆形圆顶形集落中生长且缺乏顶面-底侧极性。

[1113] 6.4. 低渗透压度培养基对多潜能标记在人类iPS细胞中的表达的影响

[1114] 该实施例描述盐浓度、离子强度和/或渗透压度对多潜能标记在已经自始发状态重编程到天然状态的人类iPS细胞中的表达的影响。在MATRIGEL™底物上在VG2i培养基中培养24天之后,对于碱性磷酸酶或NANOG的表达,将重编程的天然人类iPS细胞染色。观察到重编程的细胞强烈地表达碱性磷酸酶(图54A)和NANOG(图54B和54C)两者,这指示天然多潜能状态。

[1115] 6.5. 低渗透压度培养基对人类iPS细胞的酶分解和传代培养的影响

[1116] 在该实施例中,使用低渗透压度VG2i培养基重编程到天然状态的人类iPS细胞使用胰蛋白酶来酶分解以产生单一细胞悬浮液(图55A)。使细胞悬浮液传代到新的GELTREX™-涂覆板上以便在VG2i培养基中传代培养。在1天(图55B)和4天(图55C)之后观察到传代培养的细胞持续显示细胞处于天然多潜能状态的形态特征。特定地讲,细胞作为圆形顶形集落



生长且并未表现出顶面-底侧极性。值得注意的是酶分解可在缺乏ROCK抑制剂的情况下执行,而ROCK抑制剂通常是防止前细胞凋亡途径激活所必需的。这提出前细胞凋亡途径在本文鉴定的条件下培养的天然人类iPS细胞中在酶分解和传代培养期间并未那样强烈地激活。

[1117] 在说明书中提及的所有出版物和专利申请都指示本发明所属领域的技术人员的水平。所有出版物和专利申请都以引用的方式并入本文中,就如同特别地且单独地指示将各个别出版物或专利申请以引用的方式并入一般。除非另外自任何实施方案的上下文显而易见,否则本发明的方面、步骤或特征可与任何其它方面、步骤或特征组合使用。提及某一范围包括在所述范围内的任何整数、在所述范围内的任何子范围。提及多个范围包括所述范围的复合。

- [0001] 序列表
- [0002] <110> D·弗伦德维
- [0003] W·奥尔巴克
- [0004] K·V·莱
- [0005] 久野淳子
- [0006] D·M·瓦伦泽拉
- [0007] G·D·扬科普洛斯
- [0008] <120> 用于靶向修饰基因组的方法和组合物
- [0009] <130> 057766/453460
- [0010] <150> US 61/914,768
- [0011] <151> 2013-12-11
- [0012] <150> US 62/017,416
- [0013] <151> 2014-06-26
- [0014] <150> US 62/029,261
- [0015] <151> 2014-07-25
- [0016] <150> US 62/052,906
- [0017] <151> 2014-09-19
- [0018] <150> US 62/059,527
- [0019] <151> 2014-10-03
- [0020] <150> US 62/064,384
- [0021] <151> 2014-10-15
- [0022] <160> 94
- [0023] <170> FastSEQ for Windows 4.0版
- [0024] <210> 1
- [0025] <211> 23
- [0026] <212> DNA
- [0027] <213> 人工序列
- [0028] <220>
- [0029] <223> 连接向导RNA(gRNA)的基因组靶序列
- [0030] <220>
- [0031] <221> misc\_feature
- [0032] <222> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,
- [0033] 19, 20, 21
- [0034] <223> n = A,T,C 或 G
- [0035] <400> 1
- [0036] gnnnnnnnnnn nnnnnnnnnn ngg 23
- [0037] <210> 2
- [0038] <211> 80

- [0039] <212> RNA  
[0040] <213> 人工序列  
[0041] <220>  
[0042] <223> 向导RNA (gRNA)  
[0043] <400> 2  
[0044] guuuuagagc uagaaauagc aaguuaaaaau aaggcuaguc cguuaucaac uugaaaaagu 60  
[0045] ggcaccgagu cggugcuuuu 80  
[0046] <210> 3  
[0047] <211> 42  
[0048] <212> RNA  
[0049] <213> 人工序列  
[0050] <220>  
[0051] <223> 向导RNA (gRNA)  
[0052] <400> 3  
[0053] guuuuagagc uagaaauagc aaguuaaaaau aaggcuaguc cg 42  
[0054] <210> 4  
[0055] <211> 30  
[0056] <212> RNA  
[0057] <213> 人工序列  
[0058] <220>  
[0059] <223> crRNA  
[0060] <400> 4  
[0061] guuuuagagc uagaaauagc aaguuaaaaau 30  
[0062] <210> 5  
[0063] <211> 33  
[0064] <212> RNA  
[0065] <213> 人工序列  
[0066] <220>  
[0067] <223> crRNA  
[0068] <400> 5  
[0069] guuuuagagc uagaaauagc aaguuaaaaau aag 33  
[0070] <210> 6  
[0071] <211> 26  
[0072] <212> RNA  
[0073] <213> 人工序列  
[0074] <220>  
[0075] <223> crRNA  
[0076] <400> 6  
[0077] gaguccgagc agaagaagaa guuuua 26

[0078] <210> 7  
 [0079] <211> 12  
 [0080] <212> RNA  
 [0081] <213> 人工序列  
 [0082] <220>  
 [0083] <223> tracrRNA  
 [0084] <400> 7  
 [0085] aaggcuaguc cg 12  
 [0086] <210> 8  
 [0087] <211> 50  
 [0088] <212> RNA  
 [0089] <213> 人工序列  
 [0090] <220>  
 [0091] <223> tracrRNA  
 [0092] <400> 8  
 [0093] aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 50  
 [0094] <210> 9  
 [0095] <211> 203  
 [0096] <212> PRT  
 [0097] <213> 小家鼠  
 [0098] <400> 9  
 [0099] Met Lys Val Leu Ala Ala Gly Ile Val Pro Leu Leu Leu Leu Val Leu  
 [0100] 1 5 10 15  
 [0101] His Trp Lys His Gly Ala Gly Ser Pro Leu Pro Ile Thr Pro Val Asn  
 [0102] 20 25 30  
 [0103] Ala Thr Cys Ala Ile Arg His Pro Cys His Gly Asn Leu Met Asn Gln  
 [0104] 35 40 45  
 [0105] Ile Lys Asn Gln Leu Ala Gln Leu Asn Gly Ser Ala Asn Ala Leu Phe  
 [0106] 50 55 60  
 [0107] Ile Ser Tyr Tyr Thr Ala Gln Gly Glu Pro Phe Pro Asn Asn Val Glu  
 [0108] 65 70 75 80  
 [0109] Lys Leu Cys Ala Pro Asn Met Thr Asp Phe Pro Ser Phe His Gly Asn  
 [0110] 85 90 95  
 [0111] Gly Thr Glu Lys Thr Lys Leu Val Glu Leu Tyr Arg Met Val Ala Tyr  
 [0112] 100 105 110  
 [0113] Leu Ser Ala Ser Leu Thr Asn Ile Thr Arg Asp Gln Lys Val Leu Asn  
 [0114] 115 120 125  
 [0115] Pro Thr Ala Val Ser Leu Gln Val Lys Leu Asn Ala Thr Ile Asp Val  
 [0116] 130 135 140

[0117]	Met Arg Gly Leu Leu Ser Asn Val Leu Cys Arg Leu Cys Asn Lys Tyr
[0118]	145 150 155 160
[0119]	Arg Val Gly His Val Asp Val Pro Pro Val Pro Asp His Ser Asp Lys
[0120]	165 170 175
[0121]	Glu Ala Phe Gln Arg Lys Lys Leu Gly Cys Gln Leu Leu Gly Thr Tyr
[0122]	180 185 190
[0123]	Lys Gln Val Ile Ser Val Val Val Gln Ala Phe
[0124]	195 200
[0125]	<210> 10
[0126]	<211> 15
[0127]	<212> DNA
[0128]	<213> 人工序列
[0129]	<220>
[0130]	<223> ZFN1a结合位点
[0131]	<400> 10
[0132]	caggccctga accgc 15
[0133]	<210> 11
[0134]	<211> 6
[0135]	<212> DNA
[0136]	<213> 人工序列
[0137]	<220>
[0138]	<223> ZFN1切割位点
[0139]	<400> 11
[0140]	ttctgg 6
[0141]	<210> 12
[0142]	<211> 16
[0143]	<212> DNA
[0144]	<213> 人工序列
[0145]	<220>
[0146]	<223> ZFN1b结合位点
[0147]	<400> 12
[0148]	gattacctgc gctggg 16
[0149]	<210> 13
[0150]	<211> 15
[0151]	<212> DNA
[0152]	<213> 人工序列
[0153]	<220>
[0154]	<223> ZF21a结合位点
[0155]	<400> 13

- [0156] ttcaccctcc gcacc 15  
[0157] <210> 14  
[0158] <211> 7  
[0159] <212> DNA  
[0160] <213> 人工序列  
[0161] <220>  
[0162] <223> ZFN2切割位点  
[0163] <400> 14  
[0164] tgctgag 7  
[0165] <210> 15  
[0166] <211> 18  
[0167] <212> DNA  
[0168] <213> 人工序列  
[0169] <220>  
[0170] <223> ZF21b结合位点  
[0171] <400> 15  
[0172] tatccagatc caggggtt 18  
[0173] <210> 16  
[0174] <211> 9  
[0175] <212> PRT  
[0176] <213> 人工序列  
[0177] <220>  
[0178] <223> 归巢核酸酶家族的保守结构域  
[0179] <400> 16  
[0180] Leu Ala Gly Leu Ile Asp Ala Asp Gly  
[0181] 1 5  
[0182] <210> 17  
[0183] <211> 15  
[0184] <212> DNA  
[0185] <213> 人工序列  
[0186] <220>  
[0187] <223> ZFN1a结合位点  
[0188] <400> 17  
[0189] caggccctga accgc 15  
[0190] <210> 18  
[0191] <211> 6  
[0192] <212> DNA  
[0193] <213> 人工序列  
[0194] <220>

[0195] <223> ZFN1切割位点  
 [0196] <400> 18  
 [0197] ttctgg 6  
 [0198] <210> 19  
 [0199] <211> 16  
 [0200] <212> DNA  
 [0201] <213> 人工序列  
 [0202] <220>  
 [0203] <223> ZFN1b结合位点  
 [0204] <400> 19  
 [0205] gattacctgc gctggg 16  
 [0206] <210> 20  
 [0207] <211> 369  
 [0208] <212> PRT  
 [0209] <213> 智人  
 [0210] <400> 20  
 [0211] Met Leu Lys Pro Ser Leu Pro Phe Thr Ser Leu Leu Phe Leu Gln Leu  
 [0212] 1 5 10 15  
 [0213] Pro Leu Leu Gly Val Gly Leu Asn Thr Thr Ile Leu Thr Pro Asn Gly  
 [0214] 20 25 30  
 [0215] Asn Glu Asp Thr Thr Ala Asp Phe Phe Leu Thr Thr Met Pro Thr Asp  
 [0216] 35 40 45  
 [0217] Ser Leu Ser Val Ser Thr Leu Pro Leu Pro Glu Val Gln Cys Phe Val  
 [0218] 50 55 60  
 [0219] Phe Asn Val Glu Tyr Met Asn Cys Thr Trp Asn Ser Ser Ser Glu Pro  
 [0220] 65 70 75 80  
 [0221] Gln Pro Thr Asn Leu Thr Leu His Tyr Trp Tyr Lys Asn Ser Asp Asn  
 [0222] 85 90 95  
 [0223] Asp Lys Val Gln Lys Cys Ser His Tyr Leu Phe Ser Glu Glu Ile Thr  
 [0224] 100 105 110  
 [0225] Ser Gly Cys Gln Leu Gln Lys Lys Glu Ile His Leu Tyr Gln Thr Phe  
 [0226] 115 120 125  
 [0227] Val Val Gln Leu Gln Asp Pro Arg Glu Pro Arg Arg Gln Ala Thr Gln  
 [0228] 130 135 140  
 [0229] Met Leu Lys Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu  
 [0230] 145 150 155 160  
 [0231] Thr Leu His Lys Leu Ser Glu Ser Gln Leu Glu Leu Asn Trp Asn Asn  
 [0232] 165 170 175  
 [0233] Arg Phe Leu Asn His Cys Leu Glu His Leu Val Gln Tyr Arg Thr Asp





[0273]	Gln Pro Thr Asn Leu Thr Met His Tyr Arg Tyr Lys Gly Ser Asp Asn
[0274]	85 90 95
[0275]	Asn Thr Phe Gln Glu Cys Ser His Tyr Leu Phe Ser Lys Glu Ile Thr
[0276]	100 105 110
[0277]	Ser Gly Cys Gln Ile Gln Lys Glu Asp Ile Gln Leu Tyr Gln Thr Phe
[0278]	115 120 125
[0279]	Val Val Gln Leu Gln Asp Pro Gln Lys Pro Gln Arg Arg Ala Glu Gln
[0280]	130 135 140
[0281]	Lys Leu Asn Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu
[0282]	145 150 155 160
[0283]	Thr Leu Tyr Asn Leu Ser Glu Ser Gln Val Glu Leu Arg Trp Lys Ser
[0284]	165 170 175
[0285]	Arg Tyr Ile Glu Arg Cys Leu Gln Tyr Leu Val Gln Tyr Arg Ser Asn
[0286]	180 185 190
[0287]	Arg Asp Arg Ser Trp Thr Glu Gln Ile Val Asp His Glu Pro Arg Phe
[0288]	195 200 205
[0289]	Ser Leu Pro Ser Val Asp Glu Gln Lys Leu Tyr Thr Phe Arg Val Arg
[0290]	210 215 220
[0291]	Ser Arg Phe Asn Pro Ile Cys Gly Ser Thr Gln Gln Trp Ser Lys Trp
[0292]	225 230 235 240
[0293]	Ser Gln Pro Ile His Trp Gly Ser His Thr Ala Glu Glu Asn Pro Ser
[0294]	245 250 255
[0295]	Leu Phe Ala Leu Glu Ala Val Leu Ile Pro Val Gly Thr Met Gly Leu
[0296]	260 265 270
[0297]	Ile Ile Thr Leu Ile Phe Val Tyr Cys Trp Leu Glu Arg Met Pro Arg
[0298]	275 280 285
[0299]	Ile Pro Ala Ile Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His Gly
[0300]	290 295 300
[0301]	Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Thr Glu Ser Leu
[0302]	305 310 315 320
[0303]	Gln Pro Asp Tyr Ser Glu Arg Phe Cys His Val Ser Glu Ile Pro Pro
[0304]	325 330 335
[0305]	Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Gly Ser Pro Cys Ser Leu
[0306]	340 345 350
[0307]	His Ser Pro Tyr Trp Pro Pro Pro Cys Tyr Ser Leu Lys Pro Glu Ala
[0308]	355 360 365
[0309]	<210> 22
[0310]	<211> 368
[0311]	<212> PRT



[0350] Ile Ile Thr Leu Ile Phe Val Tyr Cys Trp Leu Glu Arg Met Pro Arg  
 [0351] 275 280 285  
 [0352] Ile Pro Ala Ile Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His Gly  
 [0353] 290 295 300  
 [0354] Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Thr Glu Ser Leu  
 [0355] 305 310 315 320  
 [0356] Gln Pro Asp Tyr Ser Glu Arg Phe Cys His Val Ser Glu Ile Pro Pro  
 [0357] 325 330 335  
 [0358] Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Gly Ser Pro Cys Ser Leu  
 [0359] 340 345 350  
 [0360] His Ser Pro Tyr Trp Pro Pro Pro Cys Tyr Ser Leu Lys Pro Glu Ala  
 [0361] 355 360 365  
 [0362] <210> 23  
 [0363] <211> 23  
 [0364] <212> DNA  
 [0365] <213> 人工序列  
 [0366] <220>  
 [0367] <223> 连接向导RNA (gRNA)的基因组靶序列  
 [0368] <220>  
 [0369] <221> misc\_feature  
 [0370] <222> (2) ... (21)  
 [0371] <223> n= A, T, C 或 G  
 [0372] <220>  
 [0373] <221> misc\_feature  
 [0374] <222> (0) ... (0)  
 [0375] <223> n可为1-20个核苷酸  
 [0376] <400> 23  
 [0377] gnnnnnnnnnn nnnnnnnnnn ngg 23  
 [0378] <210> 24  
 [0379] <211> 20  
 [0380] <212> DNA  
 [0381] <213> 人工序列  
 [0382] <220>  
 [0383] <223> 向导序列  
 [0384] <400> 24  
 [0385] gggaaccac agcatactcc 20  
 [0386] <210> 25  
 [0387] <211> 20  
 [0388] <212> DNA

- [0389] <213> 人工序列  
[0390] <220>  
[0391] <223> 向导序列  
[0392] <400> 25  
[0393] gaatcatgca cggctacccc 20  
[0394] <210> 26  
[0395] <211> 20  
[0396] <212> DNA  
[0397] <213> 人工序列  
[0398] <220>  
[0399] <223> 向导序列  
[0400] <400> 26  
[0401] tgctcctatg gggaggecg 20  
[0402] <210> 27  
[0403] <211> 20  
[0404] <212> DNA  
[0405] <213> 人工序列  
[0406] <220>  
[0407] <223> 向导序列  
[0408] <400> 27  
[0409] cttgataac attgataccc 20  
[0410] <210> 28  
[0411] <211> 20  
[0412] <212> DNA  
[0413] <213> 人工序列  
[0414] <220>  
[0415] <223> 向导序列  
[0416] <400> 28  
[0417] ggggcagagc ccttatatca 20  
[0418] <210> 29  
[0419] <211> 20  
[0420] <212> DNA  
[0421] <213> 向导序列  
[0422] <220>  
[0423] <223> 向导序列  
[0424] <400> 29  
[0425] tcgctecat taatccctag 20  
[0426] <210> 30  
[0427] <211> 20

- [0428] <212> DNA  
[0429] <213> 人工序列  
[0430] <220>  
[0431] <223> 向导序列  
[0432] <400> 30  
[0433] gtactgggga atcggtaggc 20  
[0434] <210> 31  
[0435] <211> 20  
[0436] <212> DNA  
[0437] <213> 人工序列  
[0438] <220>  
[0439] <223> 向导序列  
[0440] <400> 31  
[0441] cacgcactcc aaatttatcc 20  
[0442] <210> 32  
[0443] <211> 20  
[0444] <212> DNA  
[0445] <213> 人工序列  
[0446] <220>  
[0447] <223> 向导序列  
[0448] <400> 32  
[0449] ctaagtgtgt atcagtagat 20  
[0450] <210> 33  
[0451] <211> 20  
[0452] <212> DNA  
[0453] <213> 人工序列  
[0454] <220>  
[0455] <223> 向导序列  
[0456] <400> 33  
[0457] tgccctgcac aataagcgca 20  
[0458] <210> 34  
[0459] <211> 20  
[0460] <212> DNA  
[0461] <213> 人工序列  
[0462] <220>  
[0463] <223> 向导序列  
[0464] <400> 34  
[0465] actcattgaa acgtagatgc 20  
[0466] <210> 35

- [0467] <211> 20  
[0468] <212> DNA  
[0469] <213> 人工序列  
[0470] <220>  
[0471] <223> 向导序列  
[0472] <400> 35  
[0473] agtaagggtg gattaaattc 20  
[0474] <210> 36  
[0475] <211> 20  
[0476] <212> DNA  
[0477] <213> 人工序列  
[0478] <220>  
[0479] <223> 向导序列  
[0480] <400> 36  
[0481] gccatctaga ttcattgtaac 20  
[0482] <210> 37  
[0483] <211> 20  
[0484] <212> DNA  
[0485] <213> 人工序列  
[0486] <220>  
[0487] <223> 向导序列  
[0488] <400> 37  
[0489] gactagaaat gttctgcacc 20  
[0490] <210> 38  
[0491] <211> 20  
[0492] <212> DNA  
[0493] <213> 人工序列  
[0494] <220>  
[0495] <223> 向导序列  
[0496] <400> 38  
[0497] tgaaccaatt gtgtagcctt 20  
[0498] <210> 39  
[0499] <211> 20  
[0500] <212> DNA  
[0501] <213> 人工序列  
[0502] <220>  
[0503] <223> 向导序列  
[0504] <400> 39  
[0505] aatagtggta aagcacatg 20

- [0506] <210> 40  
[0507] <211> 20  
[0508] <212> DNA  
[0509] <213> 人工序列  
[0510] <220>  
[0511] <223> 向导序列  
[0512] <400> 40  
[0513] gtgtgctaag gatcgaagtc 20  
[0514] <210> 41  
[0515] <211> 20  
[0516] <212> DNA  
[0517] <213> 人工序列  
[0518] <220>  
[0519] <223> 向导序列  
[0520] <400> 41  
[0521] caccgagatg cttgggtatt 20  
[0522] <210> 42  
[0523] <211> 20  
[0524] <212> DNA  
[0525] <213> 人工序列  
[0526] <220>  
[0527] <223> 向导序列  
[0528] <400> 42  
[0529] tgtaaccgcc ctgaatgacc 20  
[0530] <210> 43  
[0531] <211> 20  
[0532] <212> DNA  
[0533] <213> 人工序列  
[0534] <220>  
[0535] <223> 向导序列  
[0536] <400> 43  
[0537] aaaagggcat cataaatccc 20  
[0538] <210> 44  
[0539] <211> 20  
[0540] <212> DNA  
[0541] <213> 人工序列  
[0542] <220>  
[0543] <223> 向导序列  
[0544] <400> 44

[0545] tcaaaaatag tcatacacct 20  
[0546] <210> 45  
[0547] <211> 20  
[0548] <212> DNA  
[0549] <213> 人工序列  
[0550] <220>  
[0551] <223> 向导序列  
[0552] <400> 45  
[0553] ggtctctagt acattgtaga 20  
[0554] <210> 46  
[0555] <211> 20  
[0556] <212> DNA  
[0557] <213> 人工序列  
[0558] <220>  
[0559] <223> 向导序列  
[0560] <400> 46  
[0561] atcaciaaac agttaaccgg 20  
[0562] <210> 47  
[0563] <211> 20  
[0564] <212> DNA  
[0565] <213> 人工序列  
[0566] <220>  
[0567] <223> 向导序列  
[0568] <400> 47  
[0569] tttcagacga gccgaccgg 20  
[0570] <210> 48  
[0571] <211> 20  
[0572] <212> DNA  
[0573] <213> 人工序列  
[0574] <220>  
[0575] <223> 向导序列  
[0576] <400> 48  
[0577] ctgtcaacag tgccgcgttt 20  
[0578] <210> 49  
[0579] <211> 20  
[0580] <212> DNA  
[0581] <213> 人工序列  
[0582] <220>  
[0583] <223> 向导序列



[0584] <400> 49  
[0585] tgtgtgtcat agcgatgtcg 20  
[0586] <210> 50  
[0587] <211> 20  
[0588] <212> DNA  
[0589] <213> 人工序列  
[0590] <220>  
[0591] <223> 向导序列  
[0592] <400> 50  
[0593] aacaggtacc ctatcctcac 20  
[0594] <210> 51  
[0595] <211> 20  
[0596] <212> DNA  
[0597] <213> 人工序列  
[0598] <220>  
[0599] <223> 向导序列  
[0600] <400> 51  
[0601] tcgtggttgc atgcgcactg 20  
[0602] <210> 52  
[0603] <211> 20  
[0604] <212> DNA  
[0605] <213> 人工序列  
[0606] <220>  
[0607] <223> 向导序列  
[0608] <400> 52  
[0609] ggcccggacc tagtctctct 20  
[0610] <210> 53  
[0611] <211> 20  
[0612] <212> DNA  
[0613] <213> 人工序列  
[0614] <220>  
[0615] <223> 向导序列  
[0616] <400> 53  
[0617] agtctgtaaa gttagcagtc 20  
[0618] <210> 54  
[0619] <211> 20  
[0620] <212> DNA  
[0621] <213> 人工序列  
[0622] <220>

- [0623] <223> 向导序列  
[0624] <400> 54  
[0625] ggtggtggtg ctgacggaca 20  
[0626] <210> 55  
[0627] <211> 20  
[0628] <212> DNA  
[0629] <213> 人工序列  
[0630] <220>  
[0631] <223> 向导序列  
[0632] <400> 55  
[0633] tatgagatca acactcgcta 20  
[0634] <210> 56  
[0635] <211> 20  
[0636] <212> DNA  
[0637] <213> 人工序列  
[0638] <220>  
[0639] <223> 向导序列  
[0640] <400> 56  
[0641] ccaaggactt ccccacgtta 20  
[0642] <210> 57  
[0643] <211> 20  
[0644] <212> DNA  
[0645] <213> 人工序列  
[0646] <220>  
[0647] <223> 向导序列  
[0648] <400> 57  
[0649] tgcttcctt atgcaagatt 20  
[0650] <210> 58  
[0651] <211> 20  
[0652] <212> DNA  
[0653] <213> 人工序列  
[0654] <220>  
[0655] <223> 向导序列  
[0656] <400> 58  
[0657] ttaggtaccc tatttgaata 20  
[0658] <210> 59  
[0659] <211> 20  
[0660] <212> DNA  
[0661] <213> 人工序列

[0662] <220>  
[0663] <223> 向导序列  
[0664] <400> 59  
[0665] tgcagtgggt gacaggtcca 20  
[0666] <210> 60  
[0667] <211> 20  
[0668] <212> DNA  
[0669] <213> 人工序列  
[0670] <220>  
[0671] <223> 向导序列  
[0672] <400> 60  
[0673] agggttatac tgacgttgtg 20  
[0674] <210> 61  
[0675] <211> 20  
[0676] <212> DNA  
[0677] <213> 人工序列  
[0678] <220>  
[0679] <223> 向导序列  
[0680] <400> 61  
[0681] tgtctttcaa ggagggtac 20  
[0682] <210> 62  
[0683] <211> 20  
[0684] <212> DNA  
[0685] <213> 人工序列  
[0686] <220>  
[0687] <223> 向导序列  
[0688] <400> 62  
[0689] tgatgtgcag tcagacaaag 20  
[0690] <210> 63  
[0691] <211> 20  
[0692] <212> DNA  
[0693] <213> 人工序列  
[0694] <220>  
[0695] <223> 向导序列  
[0696] <400> 63  
[0697] tgcactatgg ttgactatga 20  
[0698] <210> 64  
[0699] <211> 20  
[0700] <212> DNA

- [0701] <213> 人工序列  
[0702] <220>  
[0703] <223> 向导序列  
[0704] <400> 64  
[0705] ggaatattct aataggaagt 20  
[0706] <210> 65  
[0707] <211> 20  
[0708] <212> DNA  
[0709] <213> 人工序列  
[0710] <220>  
[0711] <223> 向导序列  
[0712] <400> 65  
[0713] aagtgctgta ccattctagc 20  
[0714] <210> 66  
[0715] <211> 20  
[0716] <212> DNA  
[0717] <213> 人工序列  
[0718] <220>  
[0719] <223> 向导序列  
[0720] <400> 66  
[0721] taatcaatag acaacctcgt 20  
[0722] <210> 67  
[0723] <211> 20  
[0724] <212> DNA  
[0725] <213> 人工序列  
[0726] <220>  
[0727] <223> 向导序列  
[0728] <400> 67  
[0729] tcattcctaa tggtattata 20  
[0730] <210> 68  
[0731] <211> 20  
[0732] <212> DNA  
[0733] <213> 人工序列  
[0734] <220>  
[0735] <223> 向导序列  
[0736] <400> 68  
[0737] agggtacata gatggcatcg 20  
[0738] <210> 69  
[0739] <211> 20

- [0740] <212> DNA  
[0741] <213> 人工序列  
[0742] <220>  
[0743] <223> 向导序列  
[0744] <400> 69  
[0745] ctctttaaca attaccactt 20  
[0746] <210> 70  
[0747] <211> 20  
[0748] <212> DNA  
[0749] <213> 人工序列  
[0750] <220>  
[0751] <223> 向导序列  
[0752] <400> 70  
[0753] tgtgggcctt tgctgatcac 20  
[0754] <210> 71  
[0755] <211> 20  
[0756] <212> DNA  
[0757] <213> 人工序列  
[0758] <220>  
[0759] <223> 向导序列  
[0760] <400> 71  
[0761] aatctatgat cctatggcct 20  
[0762] <210> 72  
[0763] <211> 20  
[0764] <212> DNA  
[0765] <213> 人工序列  
[0766] <220>  
[0767] <223> 向导序列  
[0768] <400> 72  
[0769] tgccaatagc agtgacttga 20  
[0770] <210> 73  
[0771] <211> 20  
[0772] <212> DNA  
[0773] <213> 人工序列  
[0774] <220>  
[0775] <223> 向导序列  
[0776] <400> 73  
[0777] gggaagaatg ggctattgct 20  
[0778] <210> 74

- [0779] <211> 20  
[0780] <212> DNA  
[0781] <213> 人工序列  
[0782] <220>  
[0783] <223> 向导序列  
[0784] <400> 74  
[0785] ggttgtttgt gctgatgacg 20  
[0786] <210> 75  
[0787] <211> 20  
[0788] <212> DNA  
[0789] <213> 人工序列  
[0790] <220>  
[0791] <223> 向导序列  
[0792] <400> 75  
[0793] ccgtcctagg ccttctacgt 20  
[0794] <210> 76  
[0795] <211> 20  
[0796] <212> DNA  
[0797] <213> 人工序列  
[0798] <220>  
[0799] <223> 向导序列  
[0800] <400> 76  
[0801] actagtagac ctgagggggtt 20  
[0802] <210> 77  
[0803] <211> 20  
[0804] <212> DNA  
[0805] <213> 人工序列  
[0806] <220>  
[0807] <223> 向导序列  
[0808] <400> 77  
[0809] gctccagtgt ttaggccttg 20  
[0810] <210> 78  
[0811] <211> 20  
[0812] <212> DNA  
[0813] <213> 人工序列  
[0814] <220>  
[0815] <223> 向导序列  
[0816] <400> 78  
[0817] ggcaagctga aaacgcatgc 20

- [0818] <210> 79  
[0819] <211> 20  
[0820] <212> DNA  
[0821] <213> 人工序列  
[0822] <220>  
[0823] <223> 向导序列  
[0824] <400> 79  
[0825] gtagatcgct ttccactacc 20  
[0826] <210> 80  
[0827] <211> 20  
[0828] <212> DNA  
[0829] <213> 人工序列  
[0830] <220>  
[0831] <223> 向导序列  
[0832] <400> 80  
[0833] gaactccact gctcgtgagc 20  
[0834] <210> 81  
[0835] <211> 20  
[0836] <212> DNA  
[0837] <213> 人工序列  
[0838] <220>  
[0839] <223> 向导序列  
[0840] <400> 81  
[0841] ataggtgggc actattgaag 20  
[0842] <210> 82  
[0843] <211> 20  
[0844] <212> DNA  
[0845] <213> 人工序列  
[0846] <220>  
[0847] <223> 向导序列  
[0848] <400> 82  
[0849] atgggaaggt ttataaccagc 20  
[0850] <210> 83  
[0851] <211> 20  
[0852] <212> DNA  
[0853] <213> 人工序列  
[0854] <220>  
[0855] <223> 向导序列  
[0856] <400> 83

[0857] cggtgtaaaa acaacgggaa 20  
[0858] <210> 84  
[0859] <211> 20  
[0860] <212> RNA  
[0861] <213> 人工序列  
[0862] <220>  
[0863] <223> 向导序列  
[0864] <400> 84  
[0865] gacccgcagu cccagcgucg 20  
[0866] <210> 85  
[0867] <211> 20  
[0868] <212> DNA  
[0869] <213> 人工序列  
[0870] <220>  
[0871] <223> 向导序列  
[0872] <400> 85  
[0873] actgagatca atgaccccga 20  
[0874] <210> 86  
[0875] <211> 20  
[0876] <212> DNA  
[0877] <213> 人工序列  
[0878] <220>  
[0879] <223> 向导序列  
[0880] <400> 86  
[0881] gggtcgcccg gaacctctac 20  
[0882] <210> 87  
[0883] <211> 23  
[0884] <212> DNA  
[0885] <213> 人工序列  
[0886] <220>  
[0887] <223> 向导序列  
[0888] <400> 87  
[0889] gcaggccctg aaccgcttct tgg 23  
[0890] <210> 88  
[0891] <211> 23  
[0892] <212> DNA  
[0893] <213> 人工序列  
[0894] <220>  
[0895] <223> 向导序列



- [0896] <400> 88  
[0897] cctgcgctgg gtgcagacgc ttt 23  
[0898] <210> 89  
[0899] <211> 20  
[0900] <212> DNA  
[0901] <213> 人工序列  
[0902] <220>  
[0903] <223> 向导序列  
[0904] <400> 89  
[0905] ccagctactt gctcgtacaa 20  
[0906] <210> 90  
[0907] <211> 20  
[0908] <212> DNA  
[0909] <213> 人工序列  
[0910] <220>  
[0911] <223> 向导序列  
[0912] <400> 90  
[0913] cccctcagat tcacgtgcgt 20  
[0914] <210> 91  
[0915] <211> 23  
[0916] <212> DNA  
[0917] <213> 人工序列  
[0918] <220>  
[0919] <223> 向导序列  
[0920] <400> 91  
[0921] gaagctcttt ctatacaatc tgg 23  
[0922] <210> 92  
[0923] <211> 23  
[0924] <212> DNA  
[0925] <213> 人工序列  
[0926] <220>  
[0927] <223> 向导序列  
[0928] <400> 92  
[0929] cccccgaaag gaggagccct agg 23  
[0930] <210> 93  
[0931] <211> 56  
[0932] <212> DNA  
[0933] <213> 褐家鼠  
[0934] <400> 93

- 
- [0935] ttctgagcct cagccgacca acctcactat gcactatagg tatgagaagg gggagg 56  
[0936] <210> 94  
[0937] <211> 20  
[0938] <212> DNA  
[0939] <213> a人工序列  
[0940] <220>  
[0941] <223> 向导序列  
[0942] <400> 94  
[0943] gtatagccct gttacacatt 20

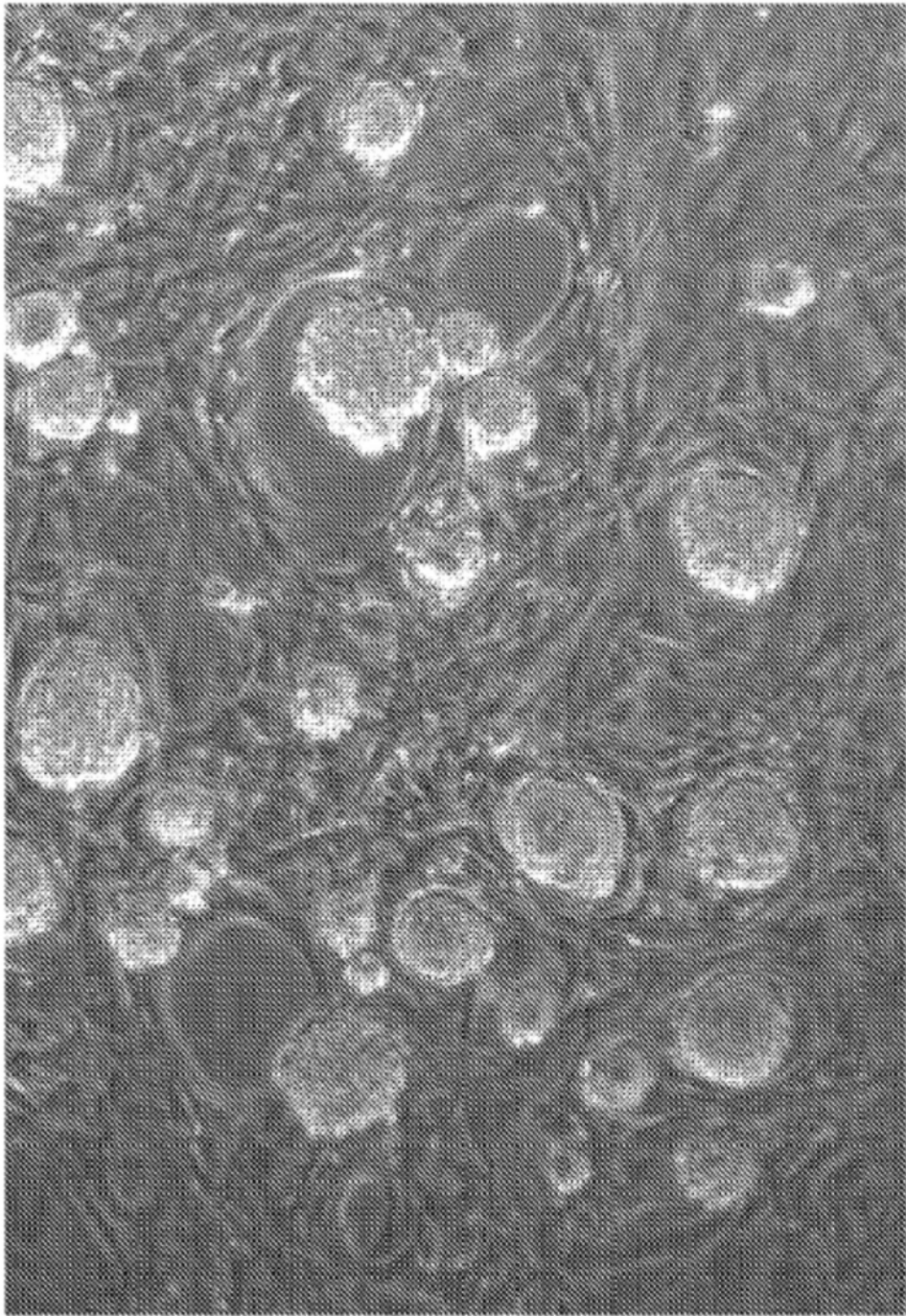


图1

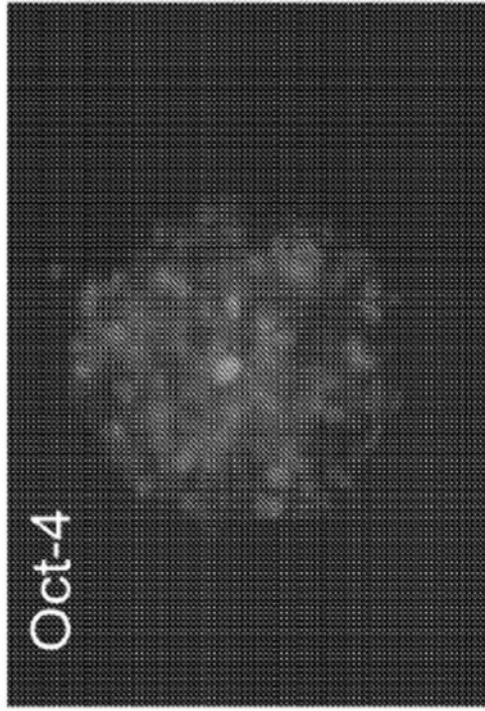


图2A

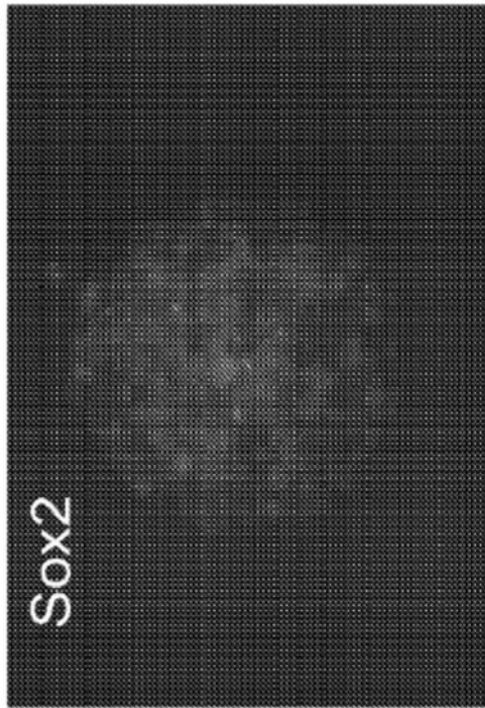


图2B

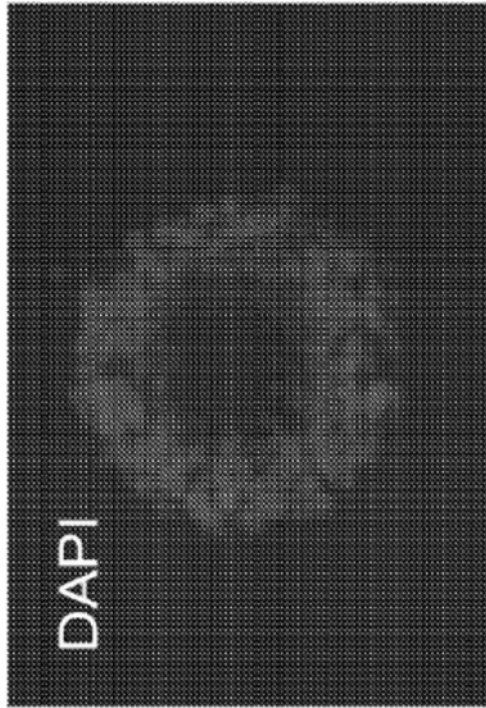


图2C

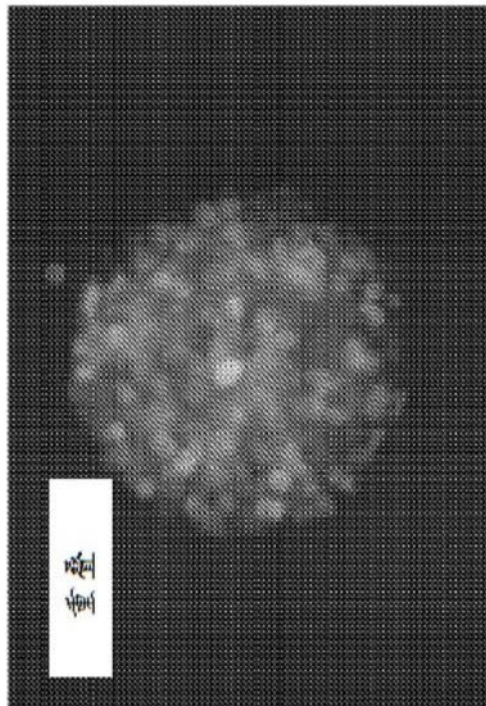


图2D

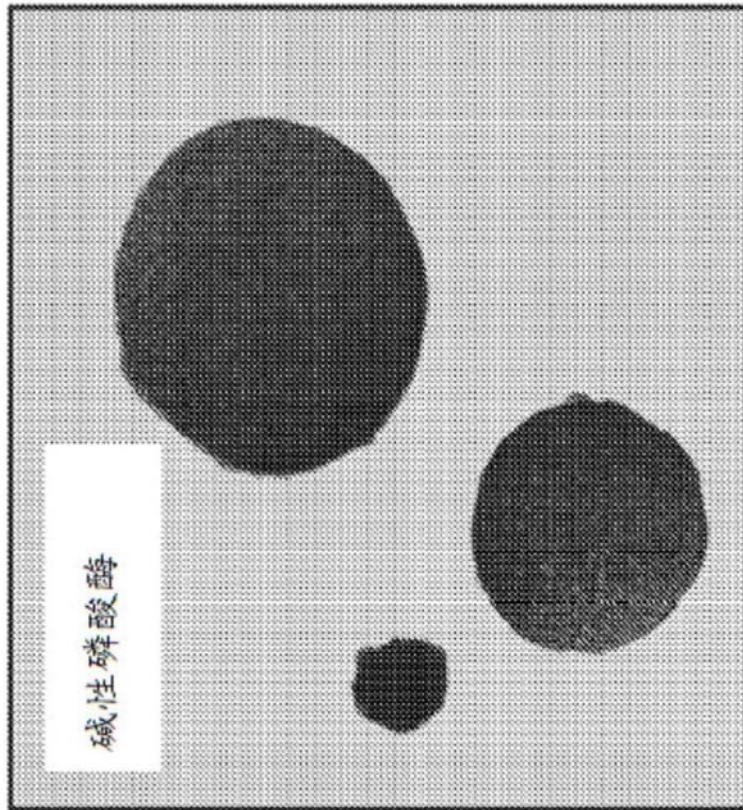


图3



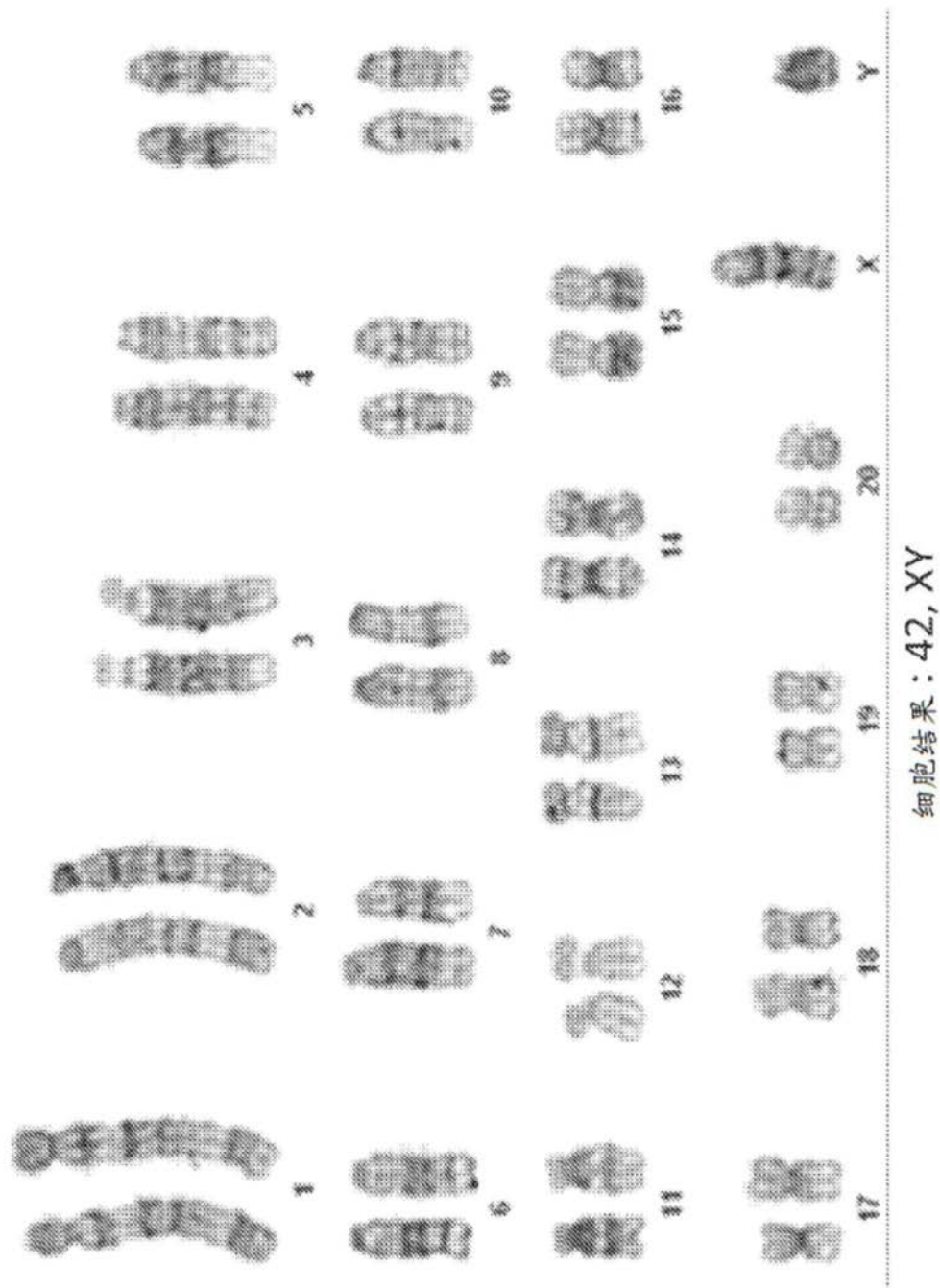


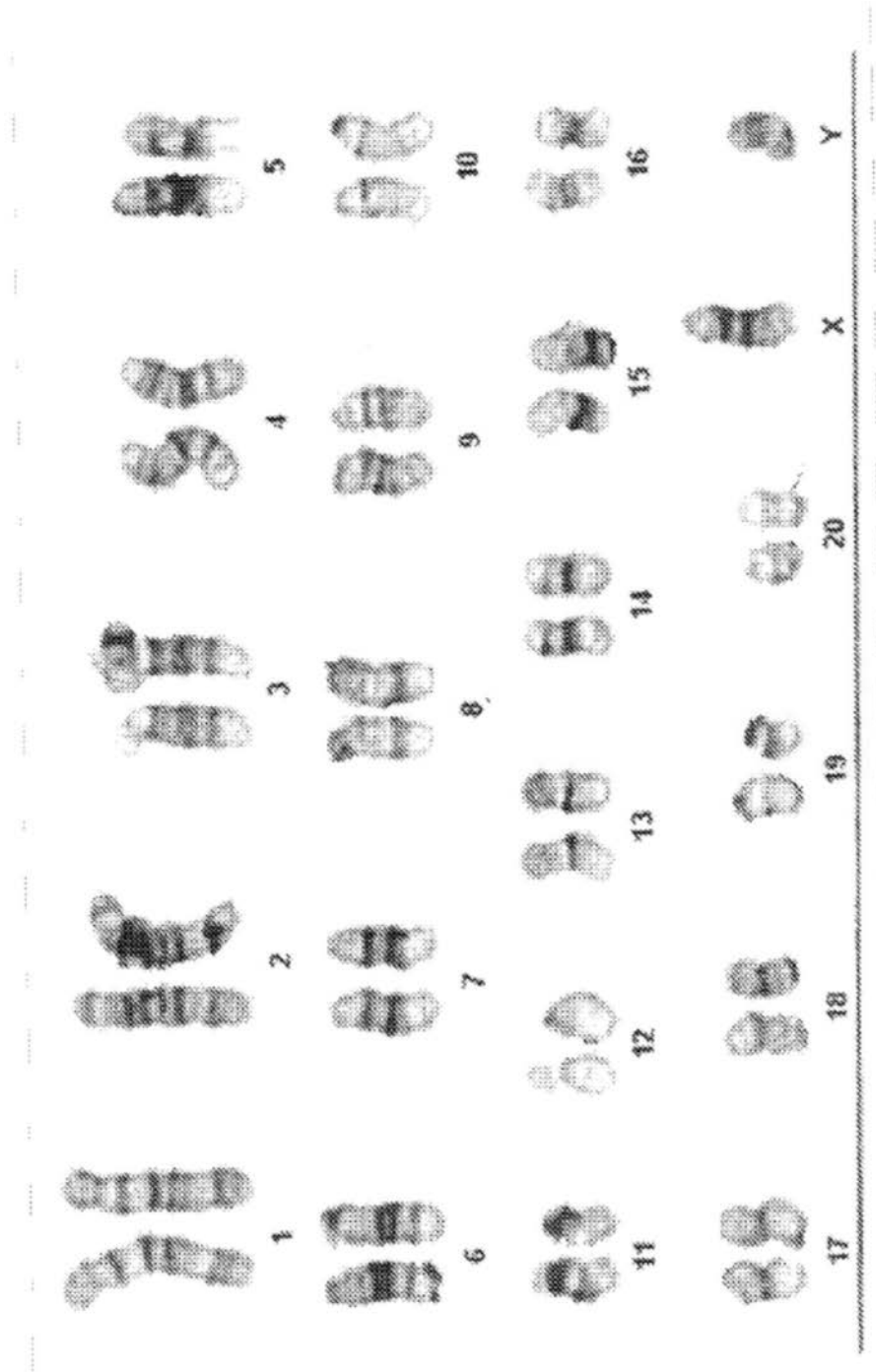
图4



细胞结果：42, XY

图5A





细胞结果：42, XY

图5B



细胞结果：42, XY

图6A

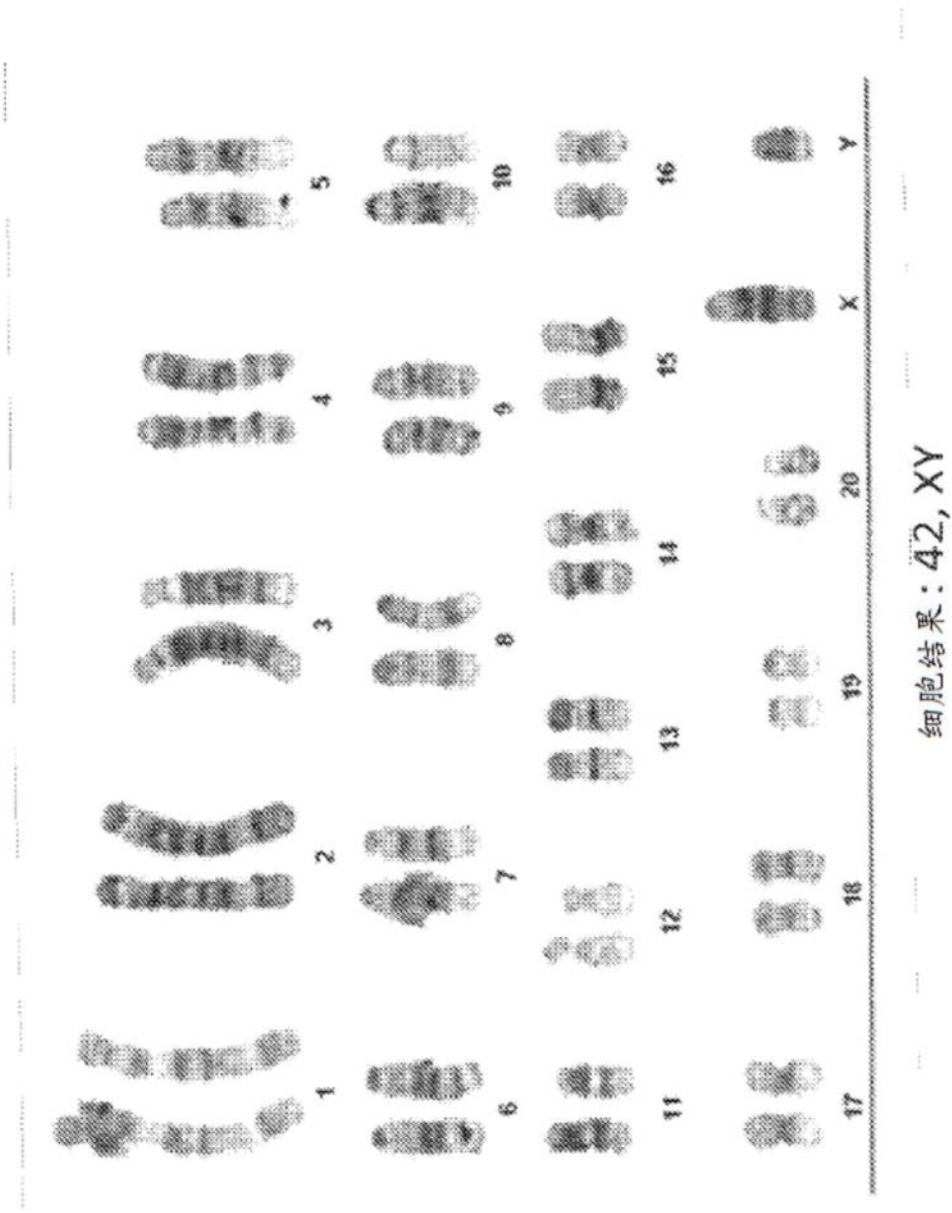
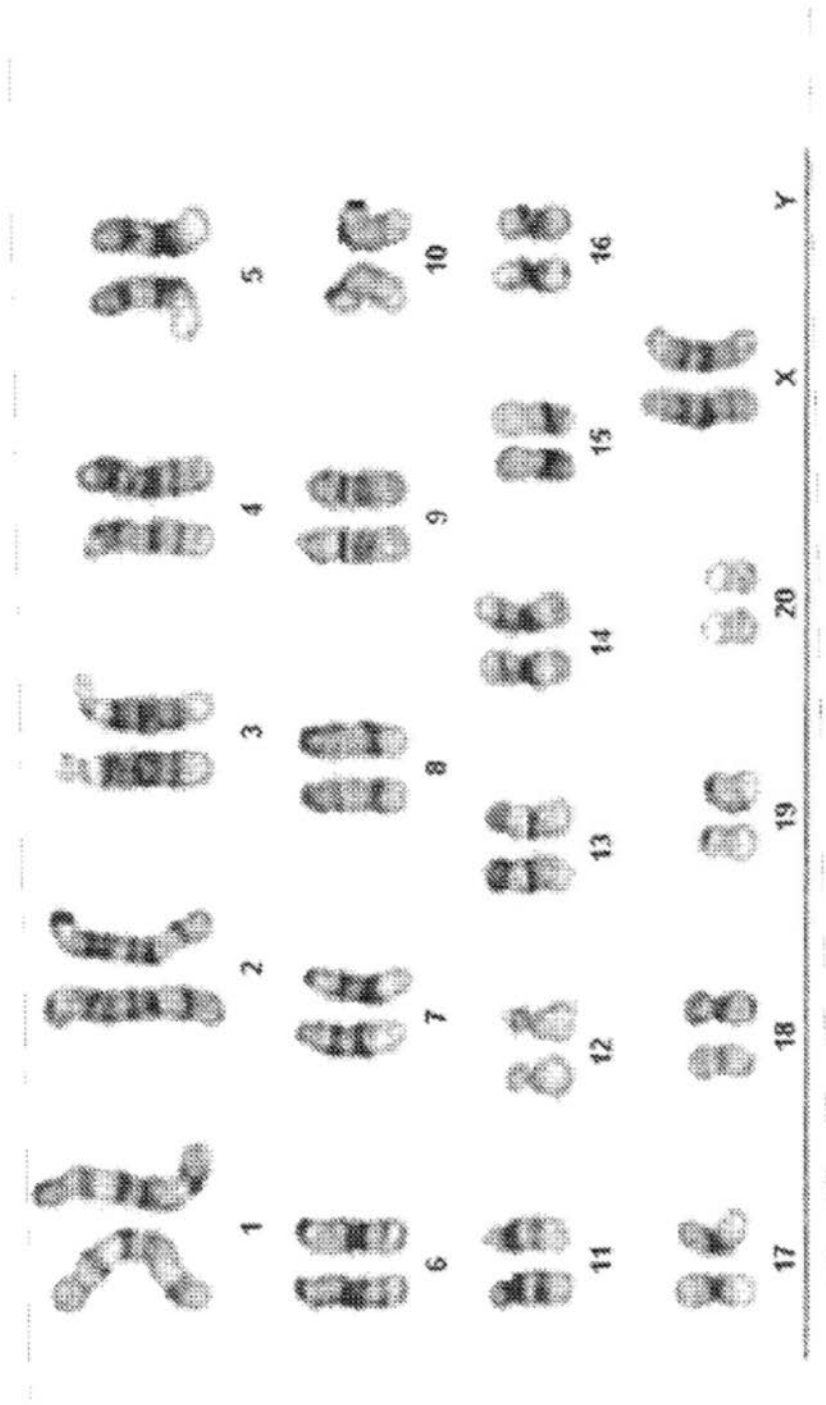


图6B



图7A



细胞结果: 42, XX

图7B

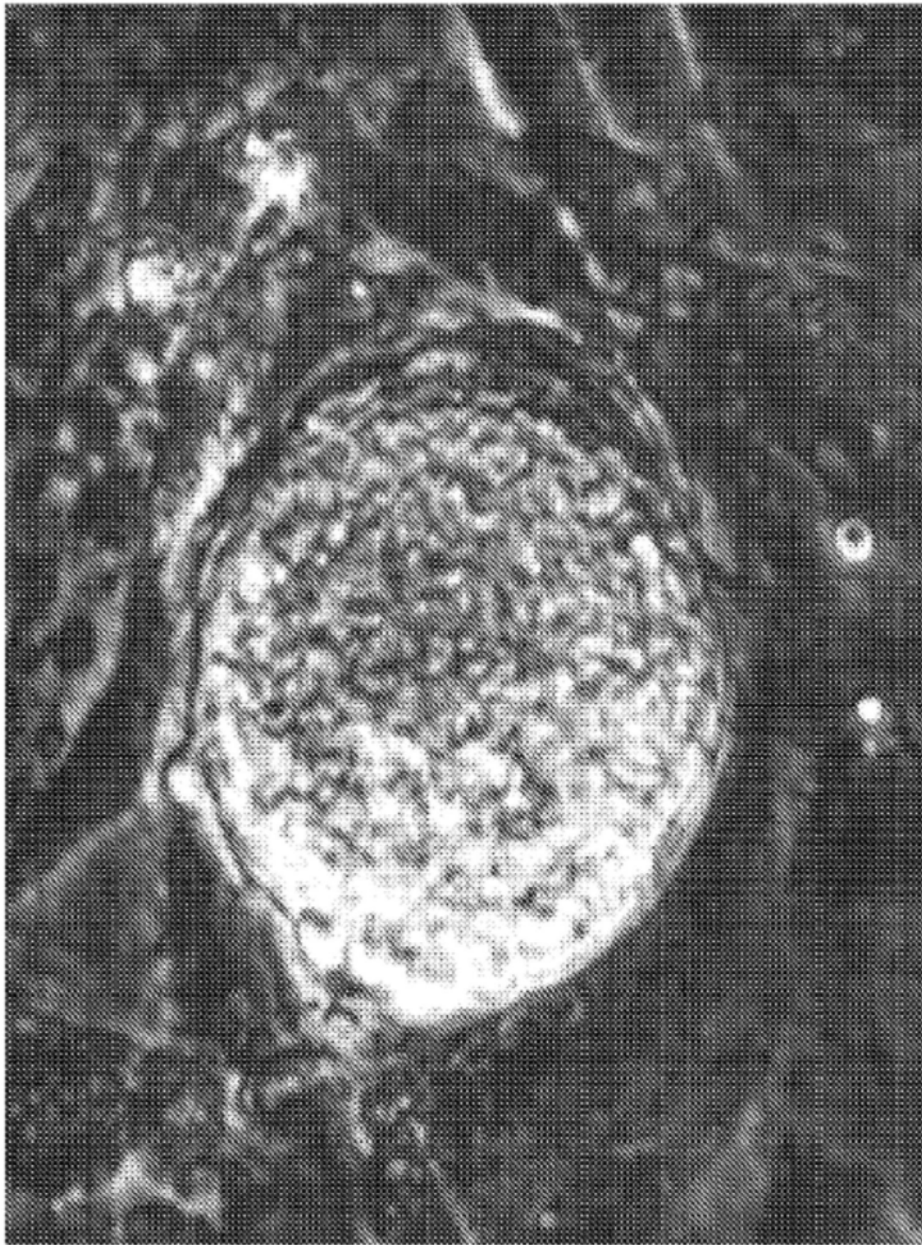


图8





图9

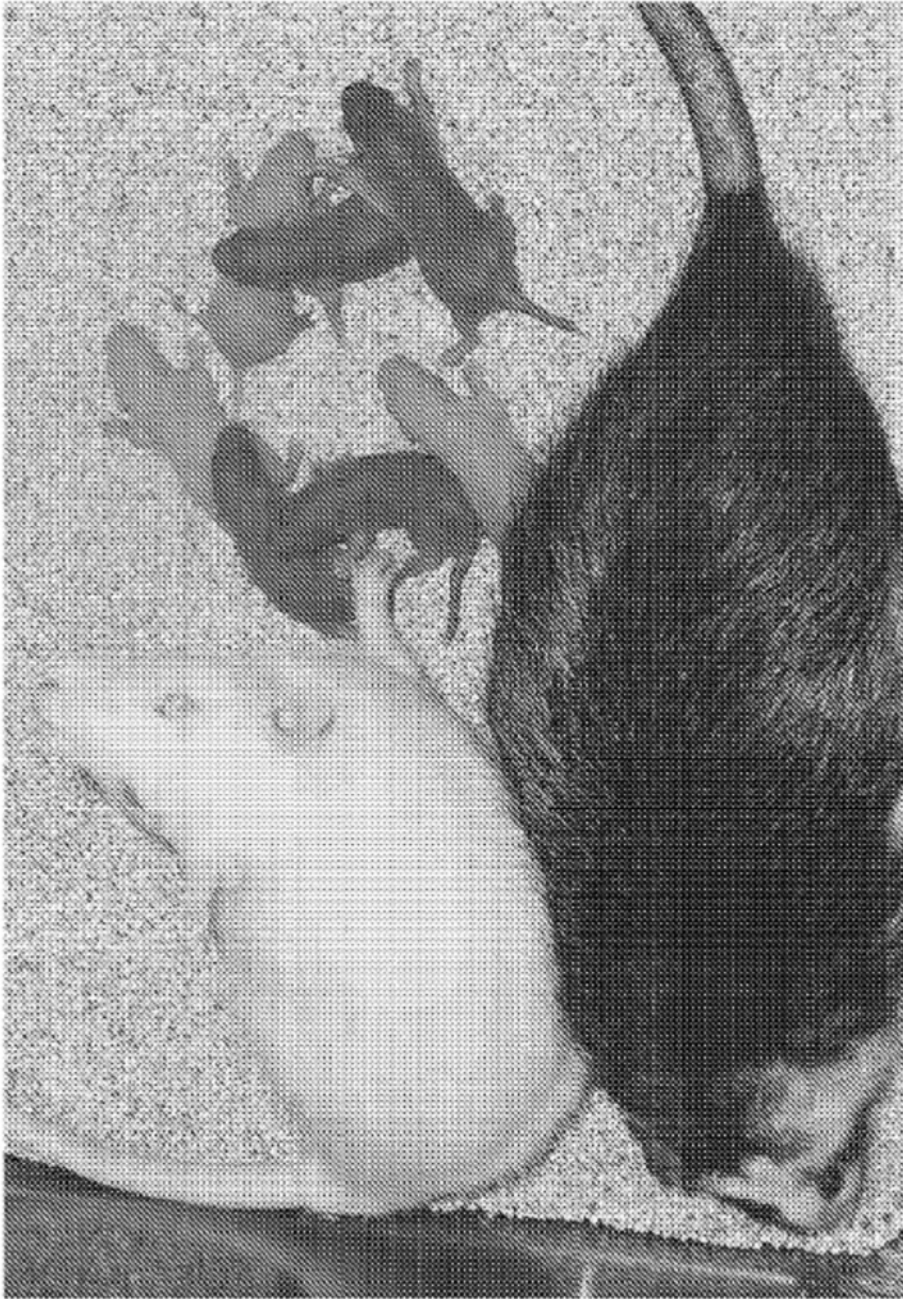


图10



大鼠 ApoE 基因座和锌指切割位点的位置

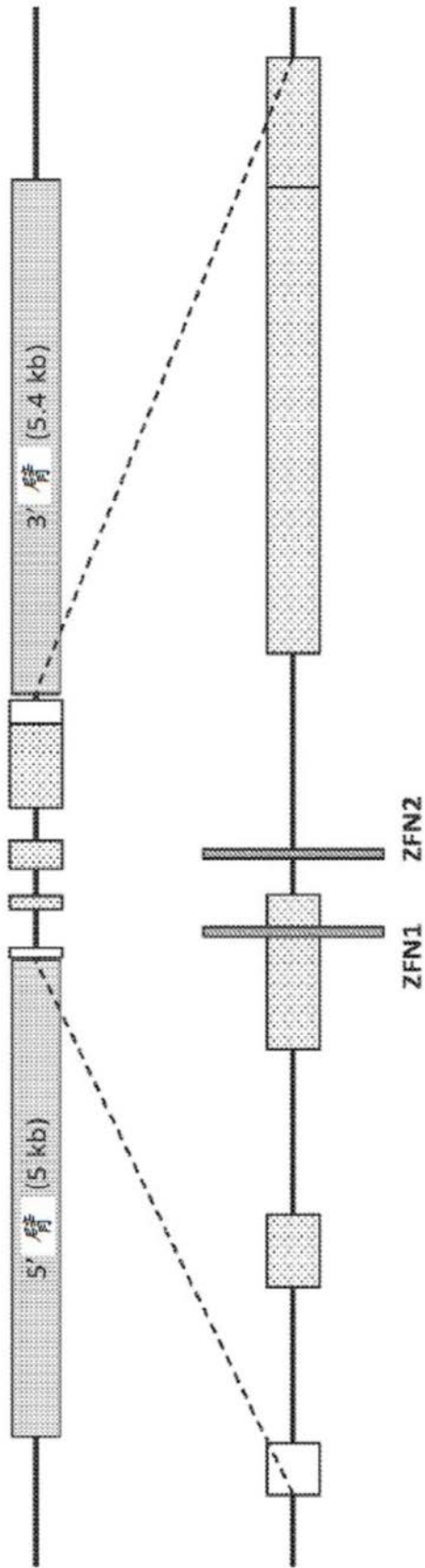


图11



图12A

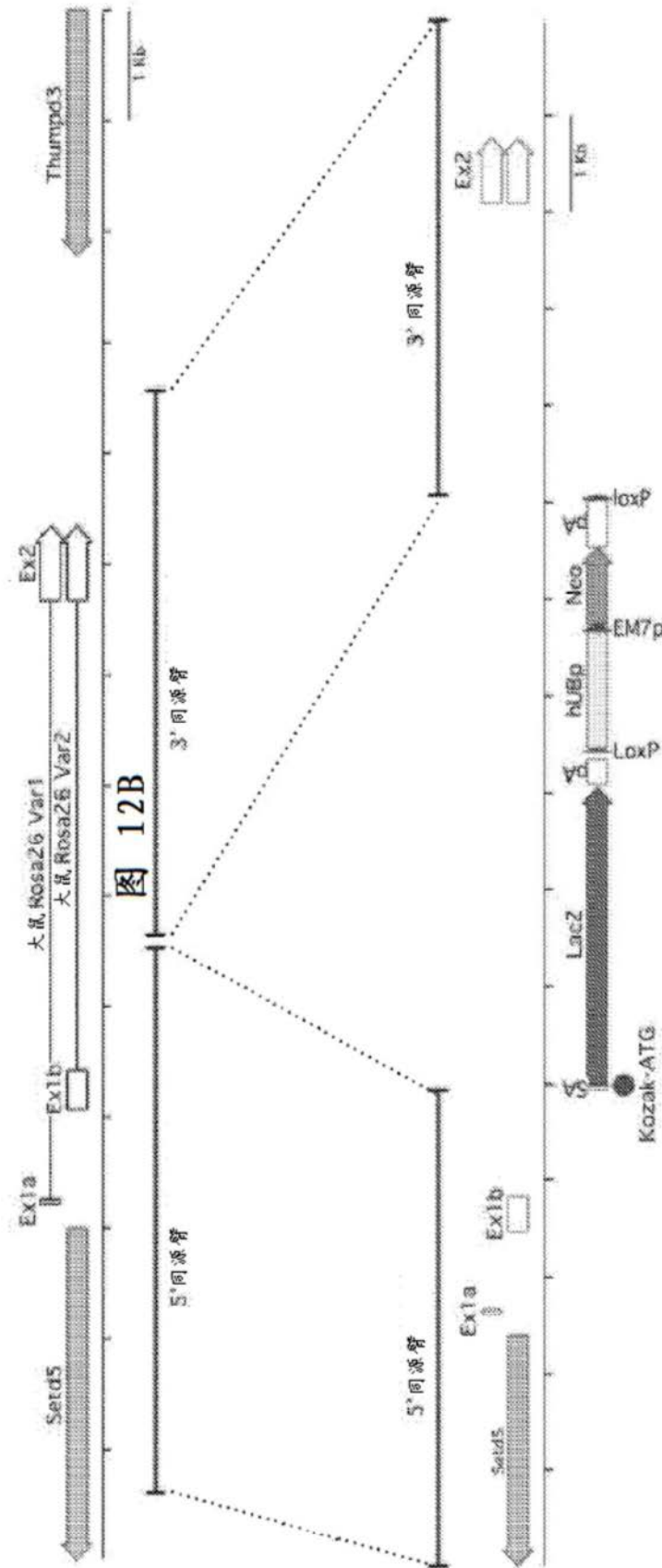


图 12C

图 12B

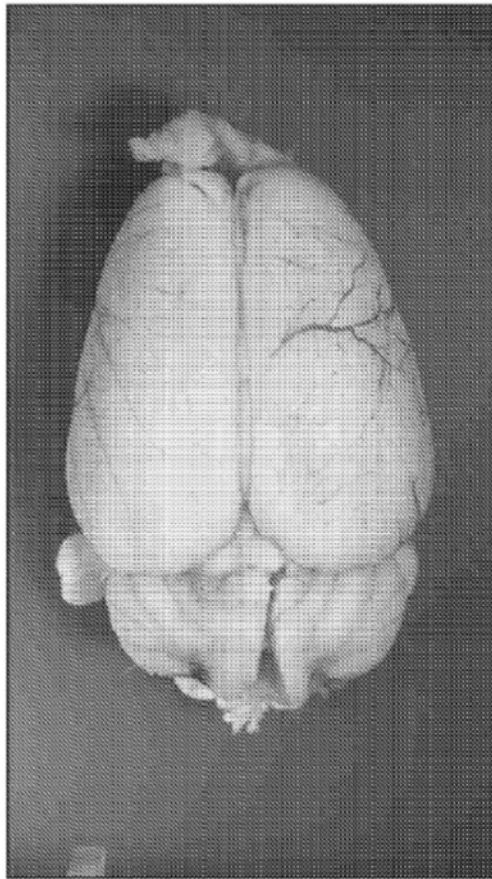


图13A

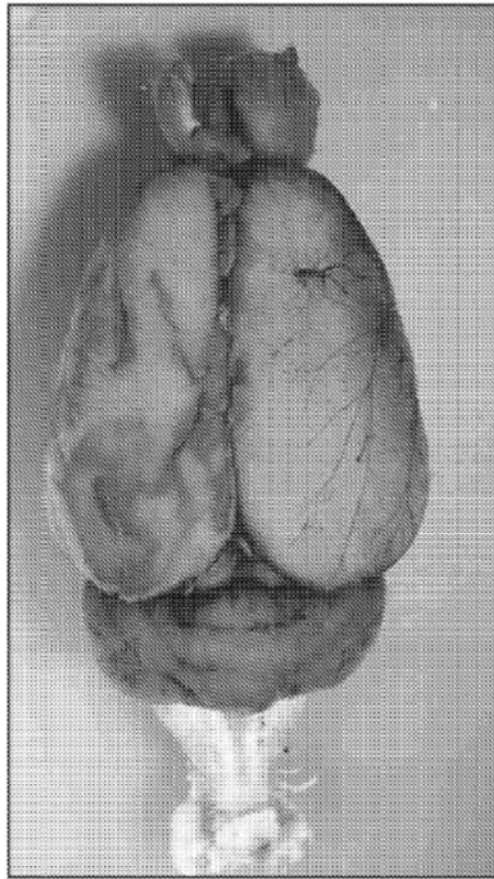


图13B

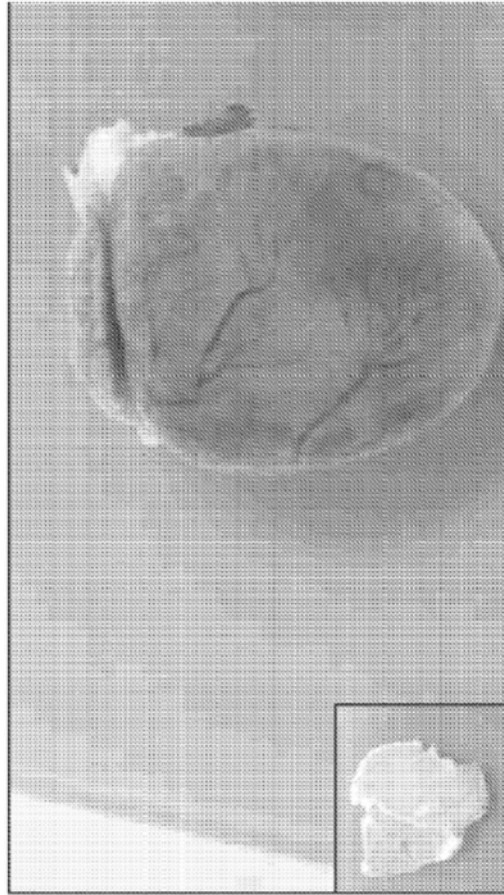


图13C

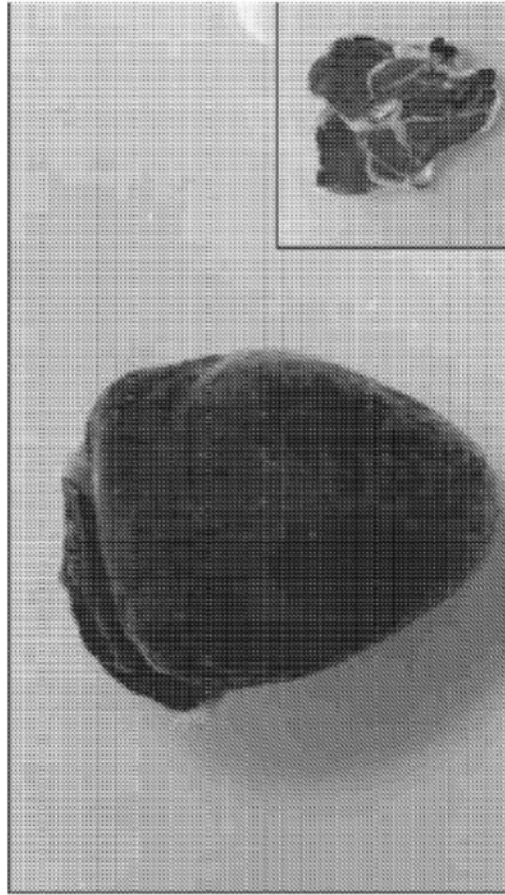


图13D



图13E



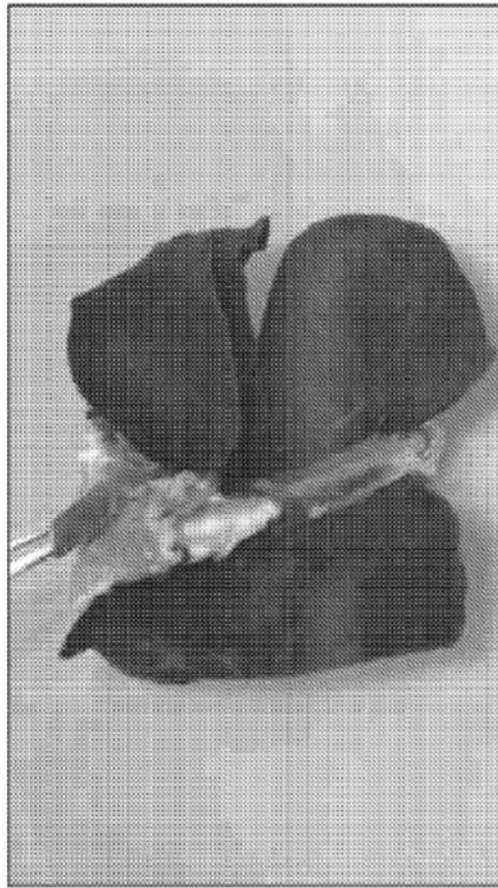


图13F



图13G

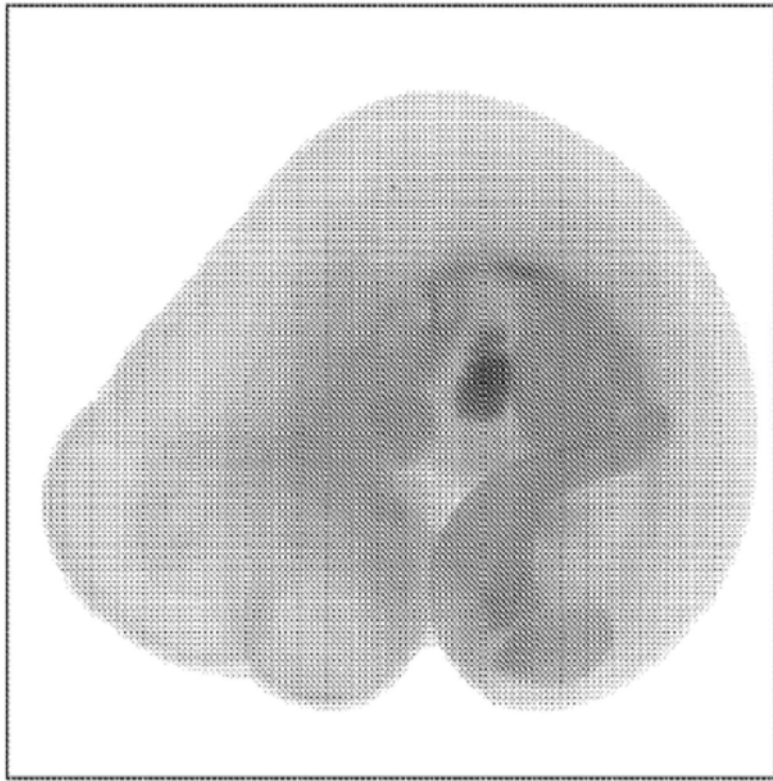


图13H

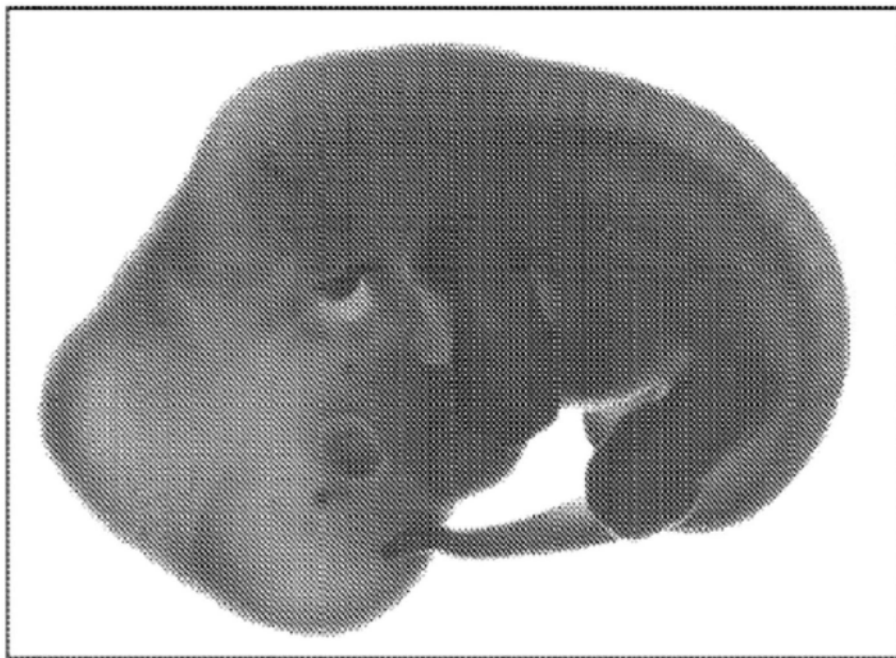


图13I

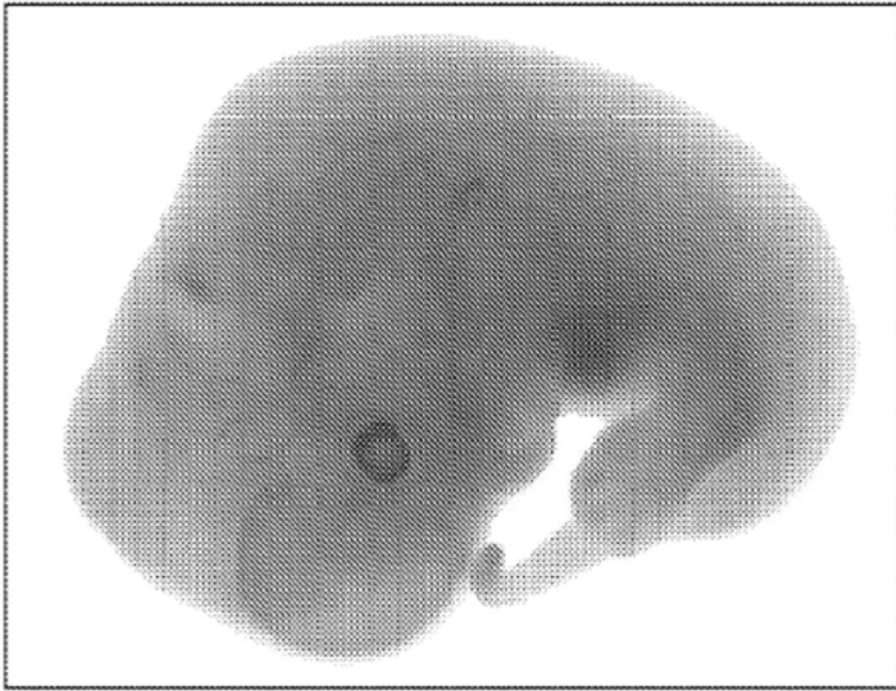


图13J

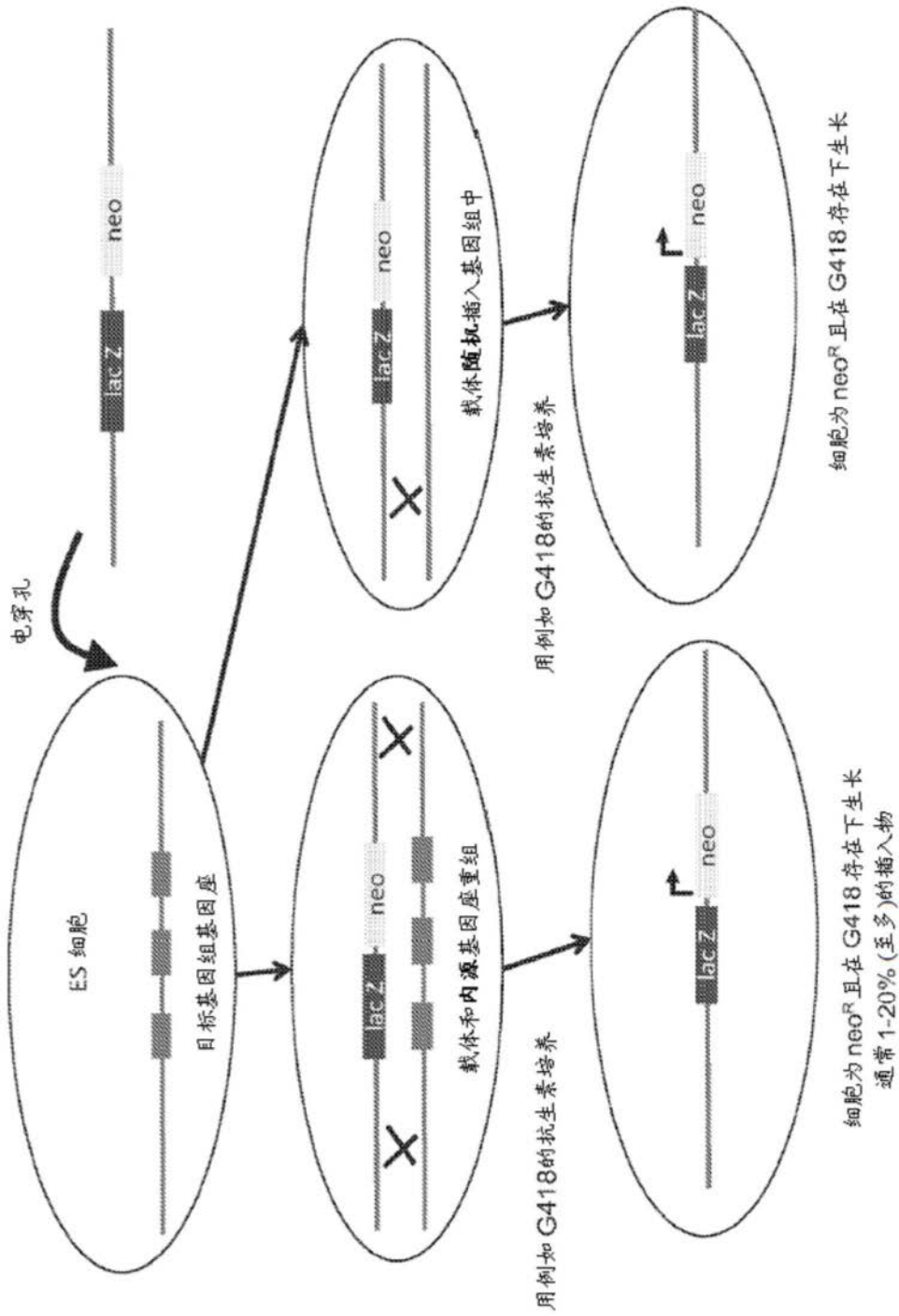


图14

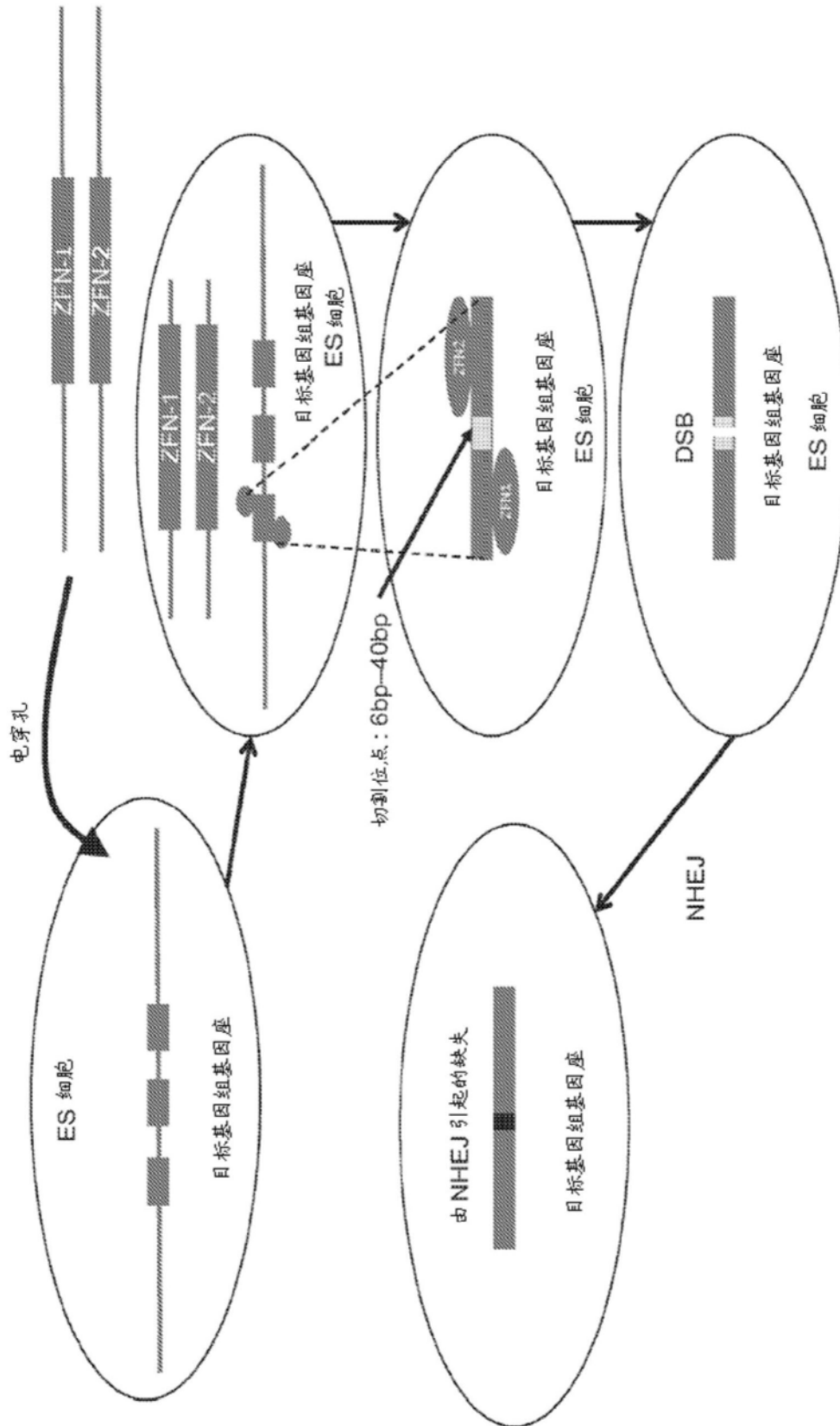


图15

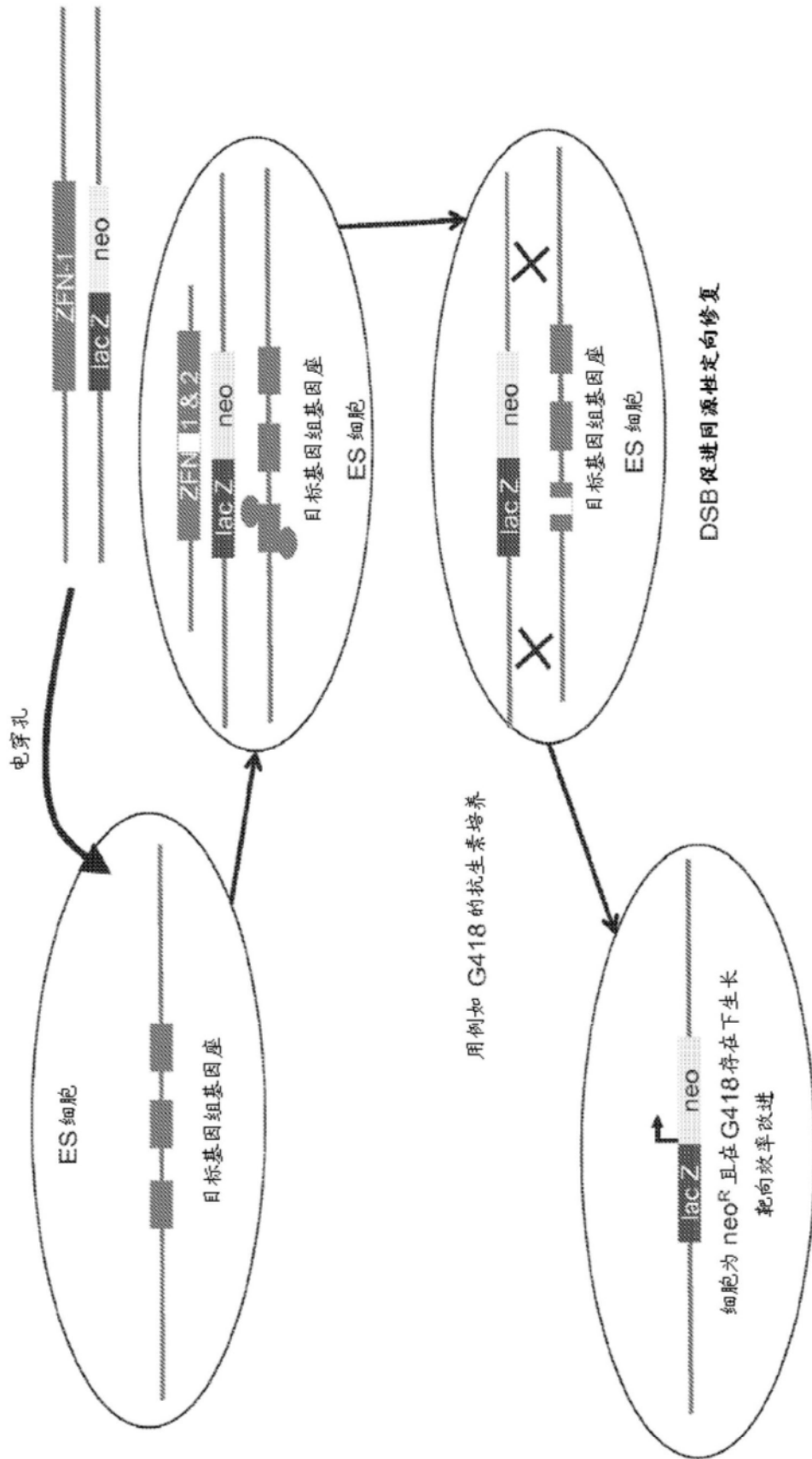


图16



图17

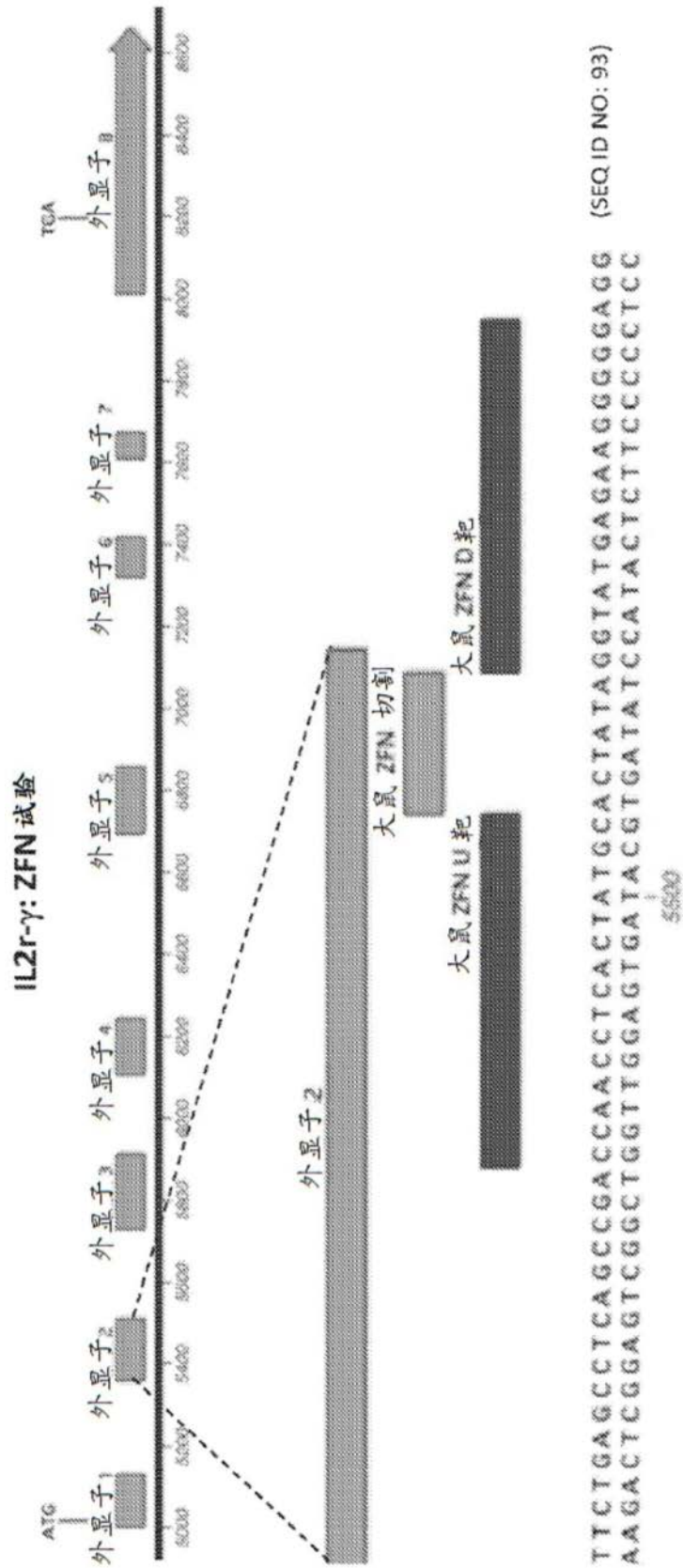


图18





图19

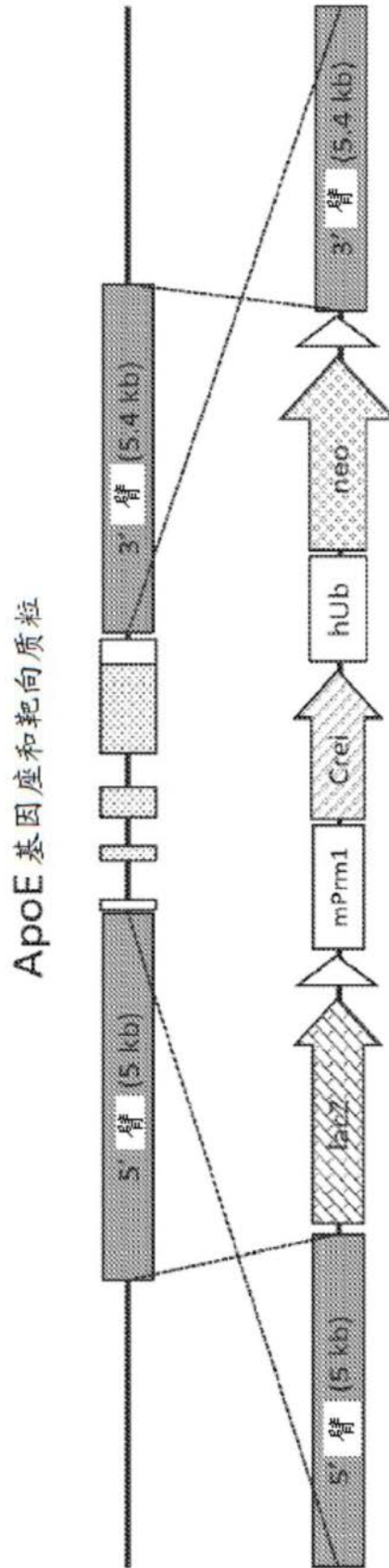


图20

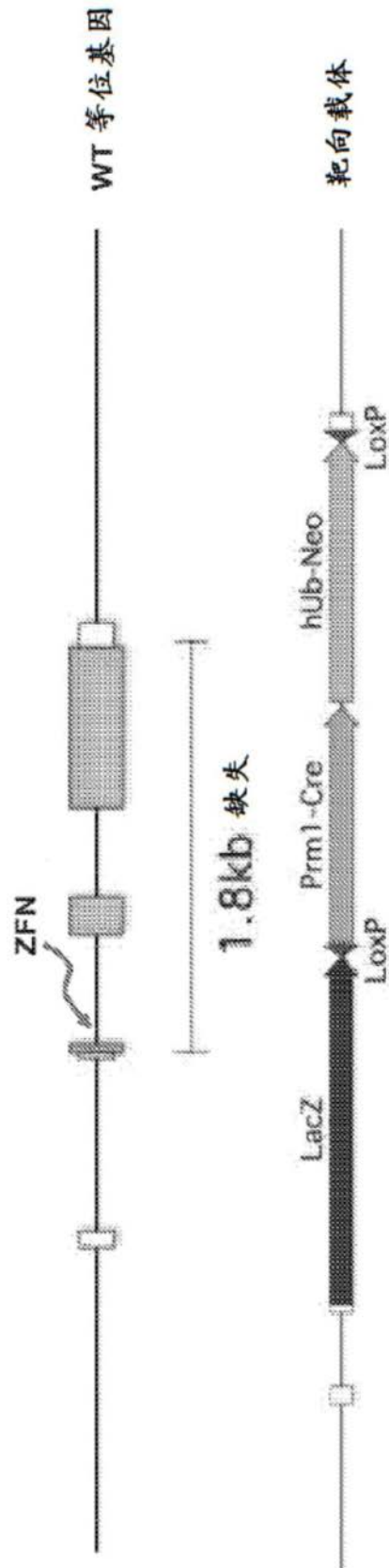


图21A

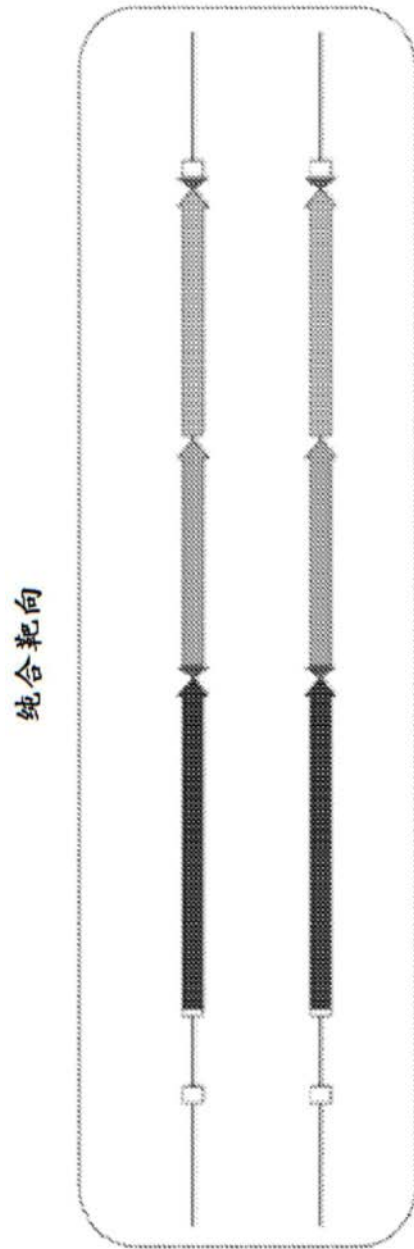


图21B

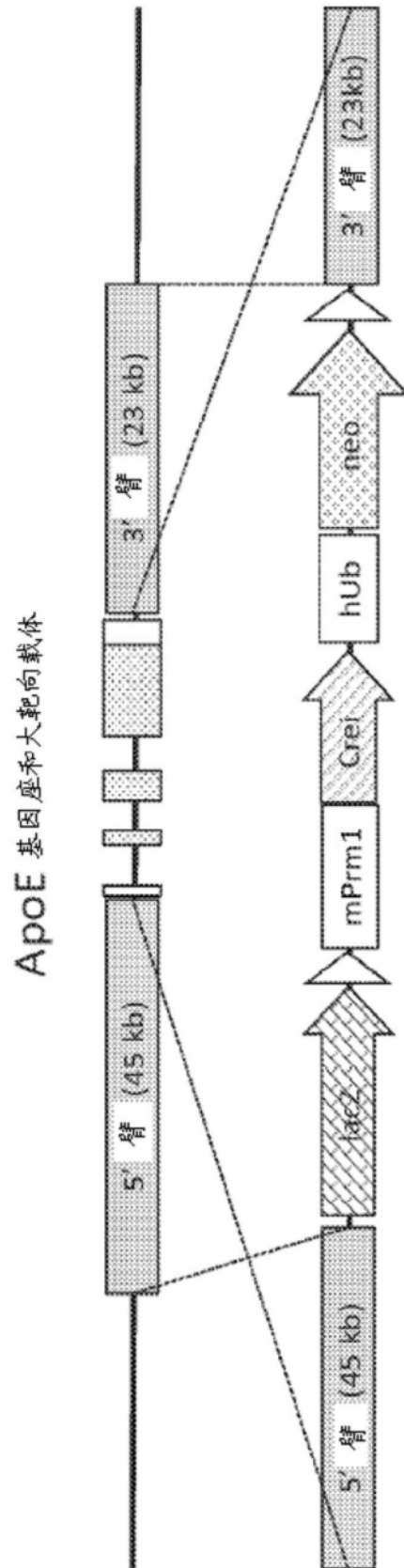


图22

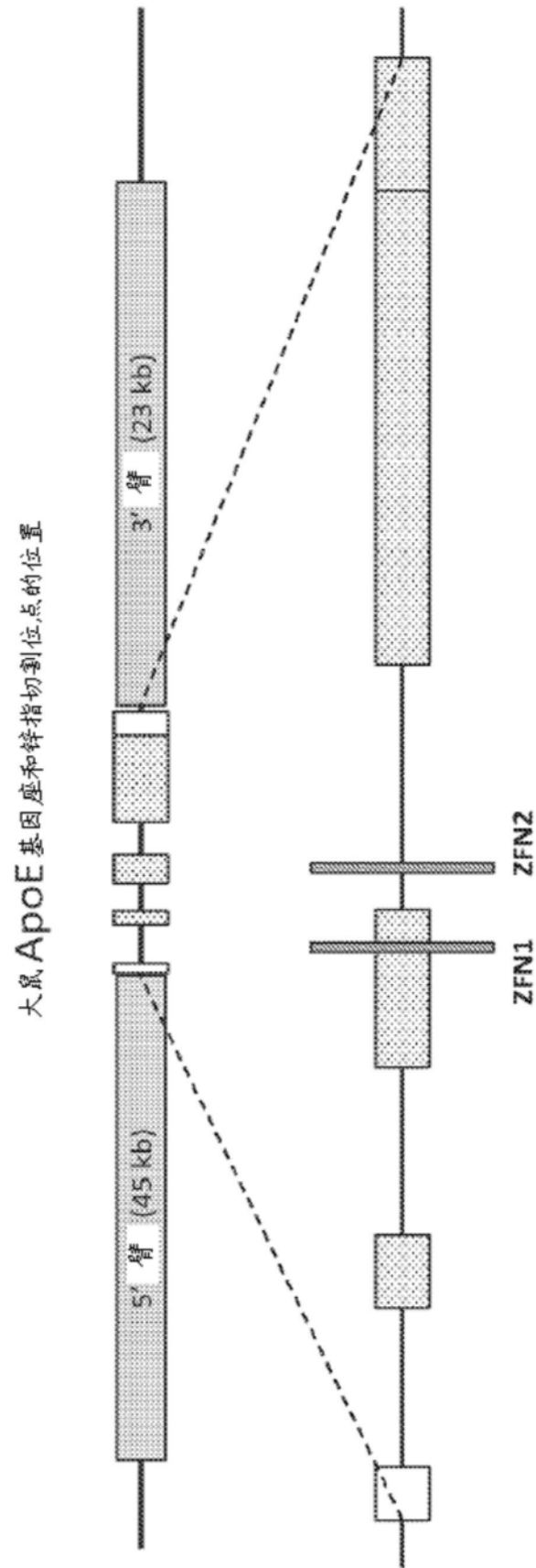


图23

Il2rg 缺失

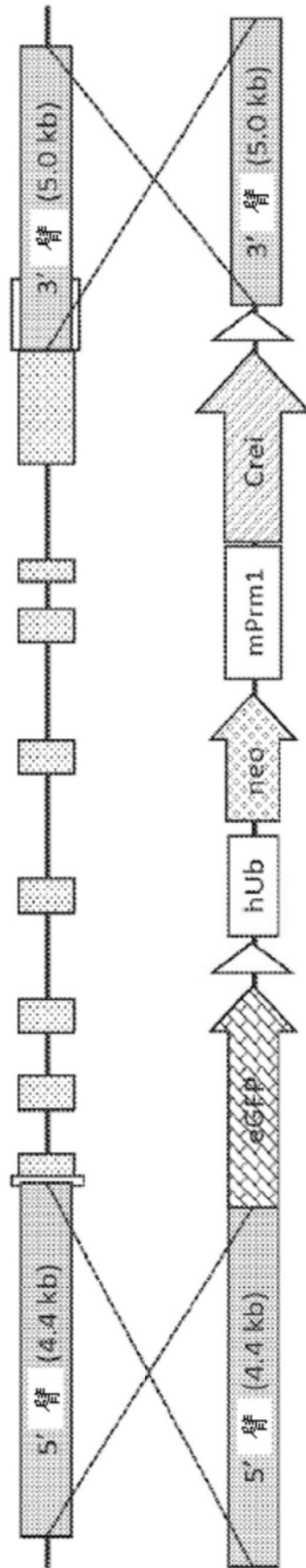


图24

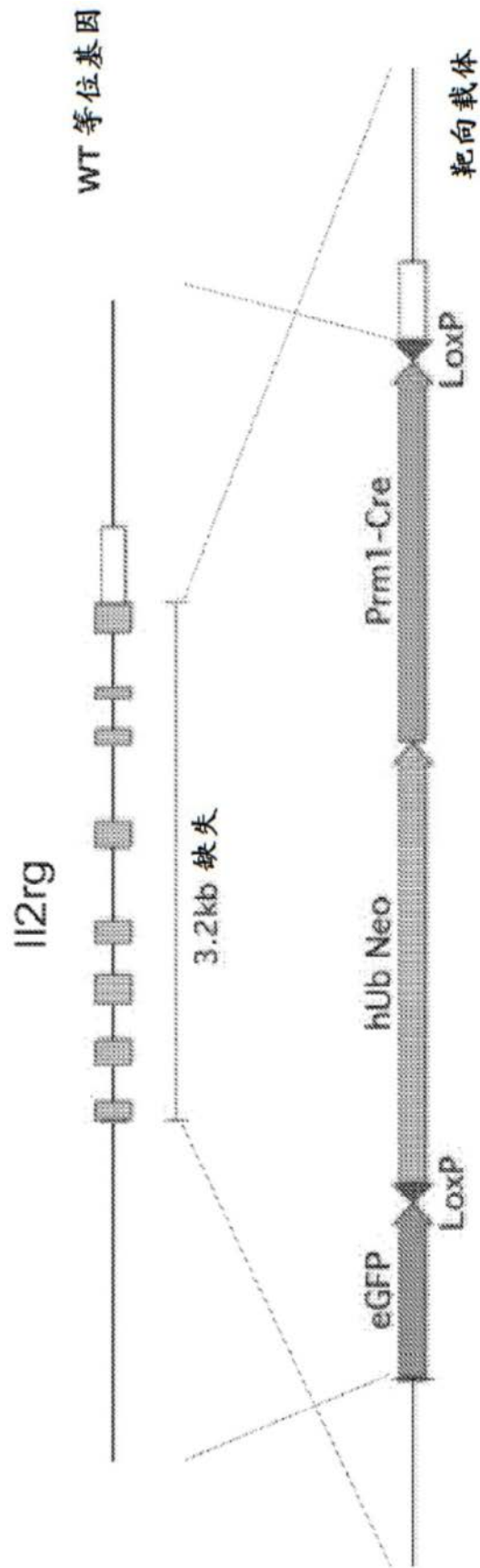


图25



大鼠Rag2基因座和大靶向载体

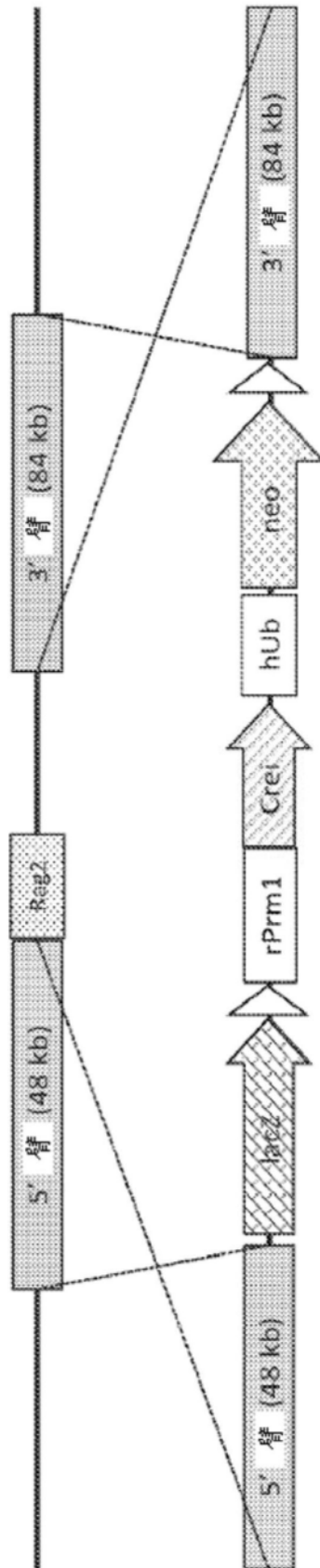


图26

### Rag2/Rag1 靶向

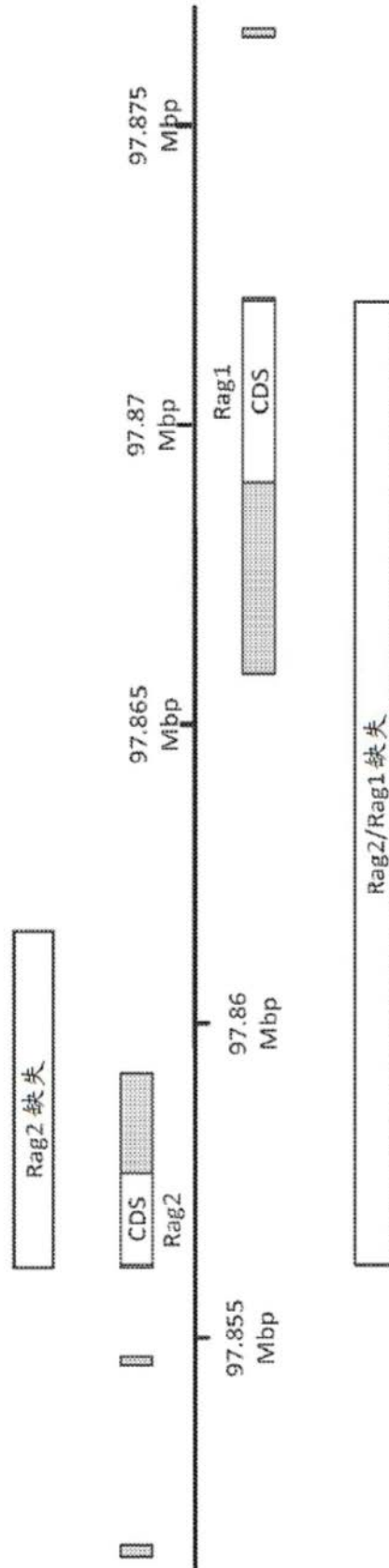


图27

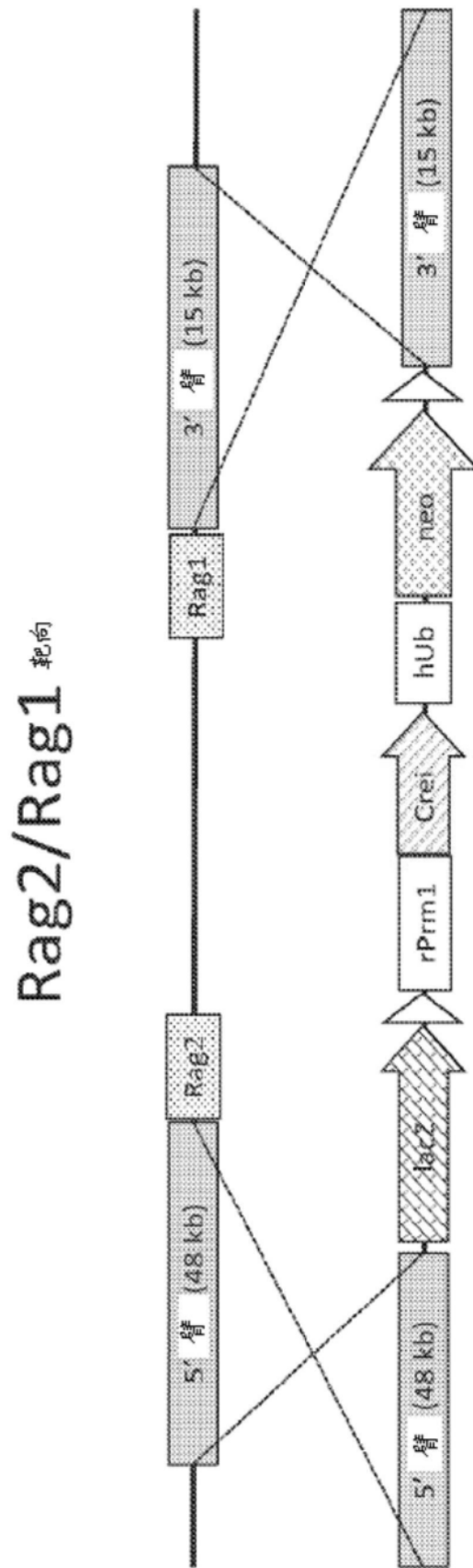


图28

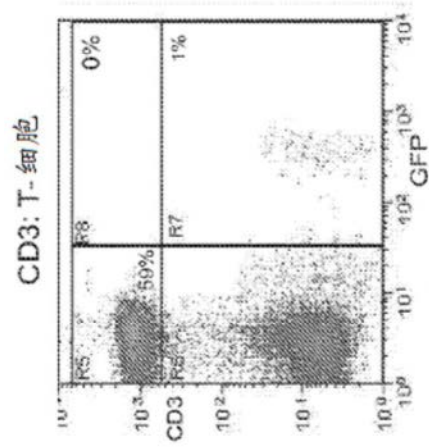


图29A

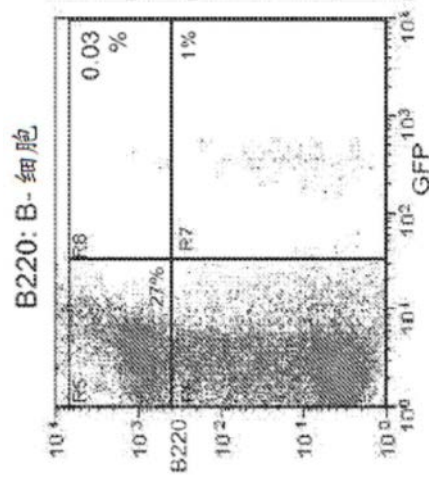


图29B

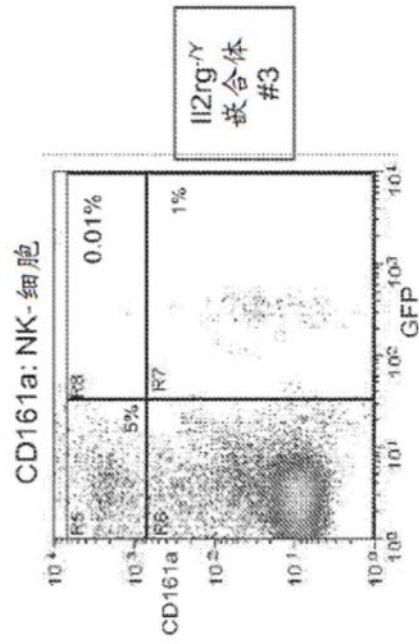


图29C

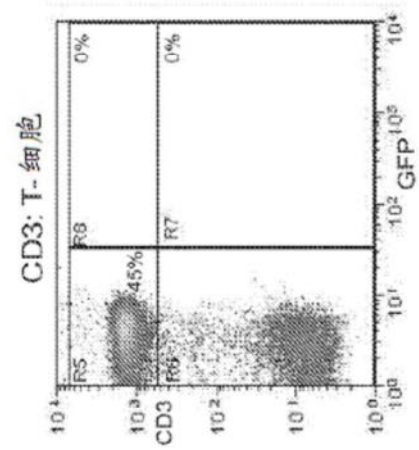


图29D

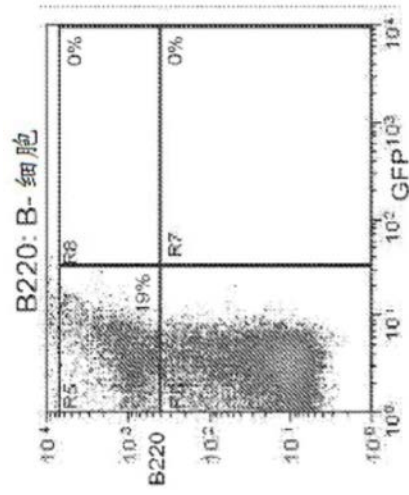


图29E

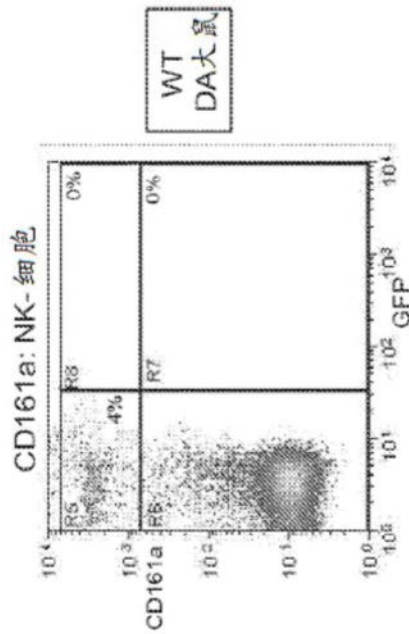


图29F

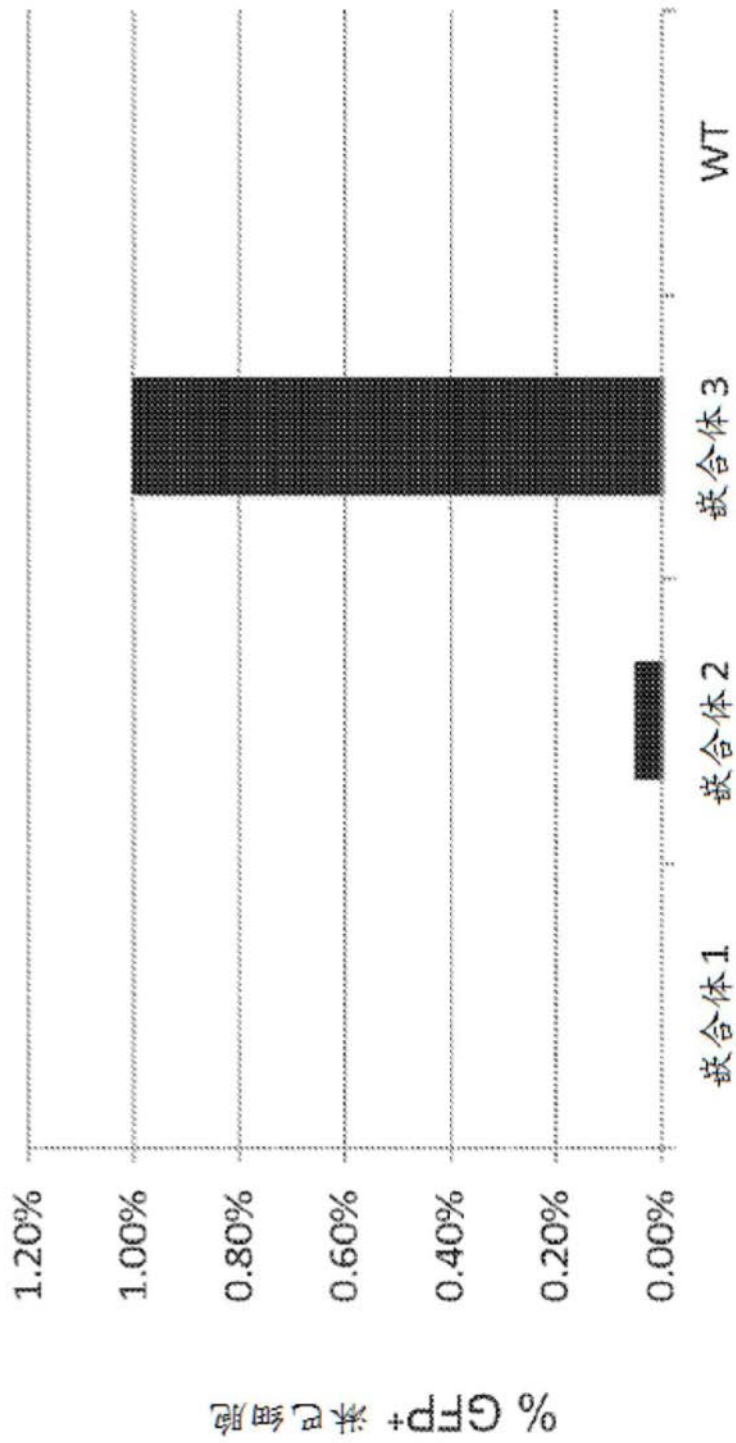


图30

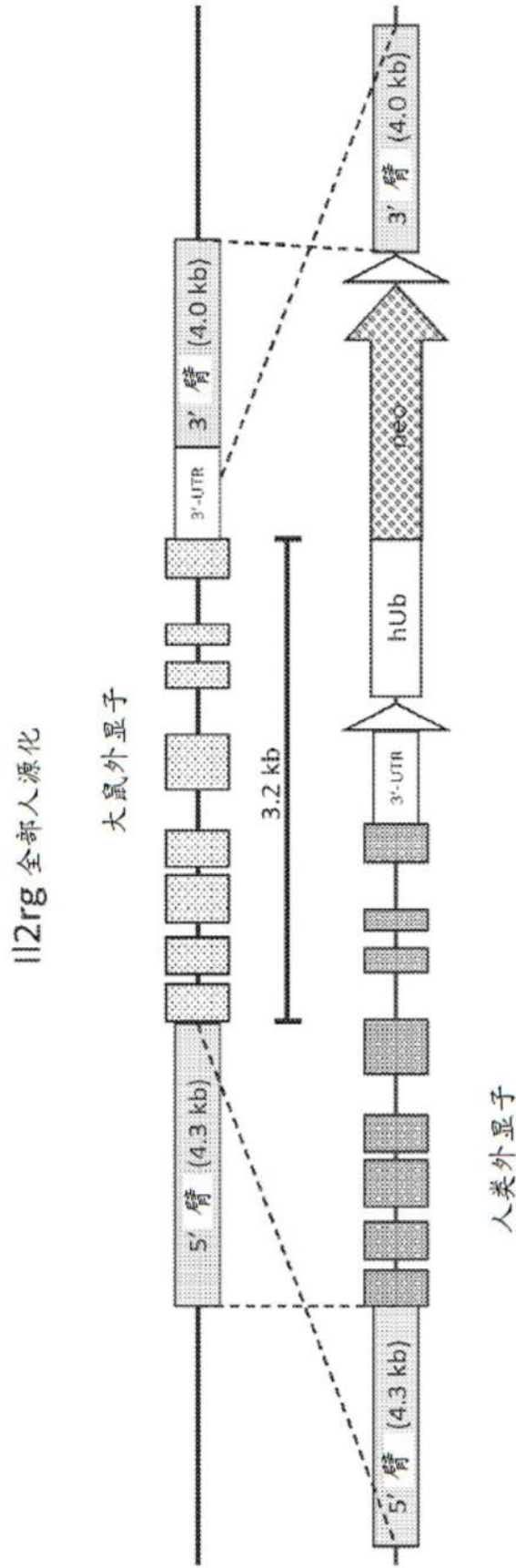


图31



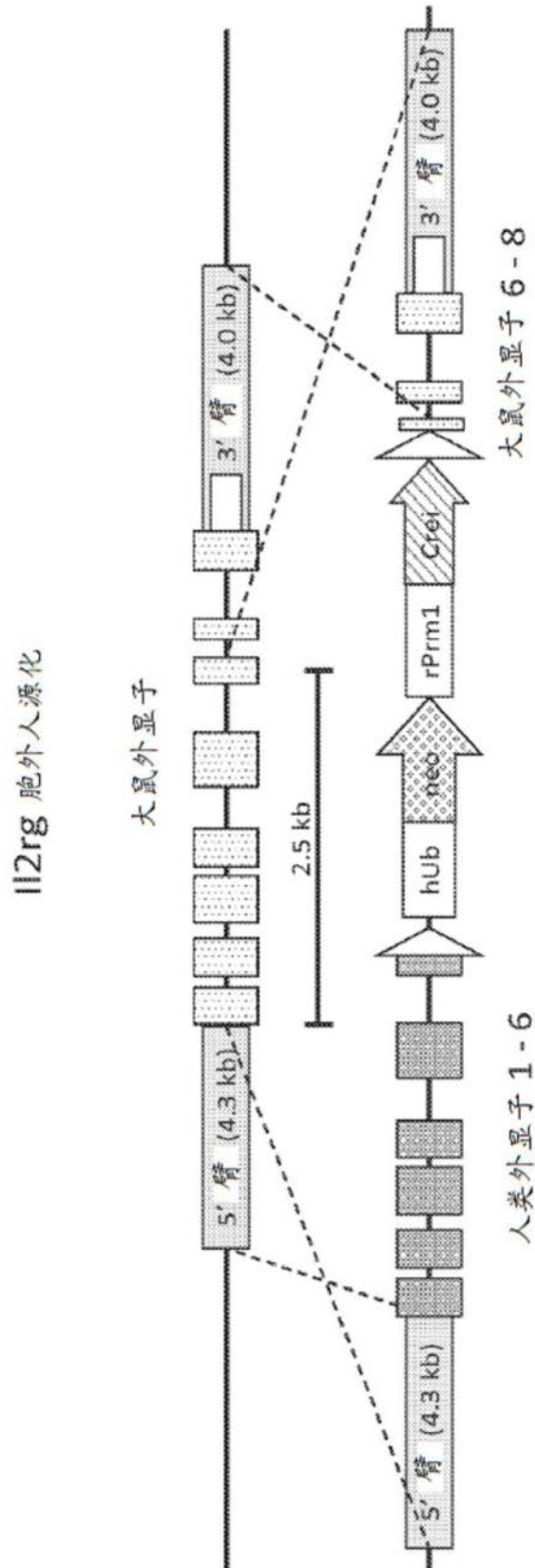


图32



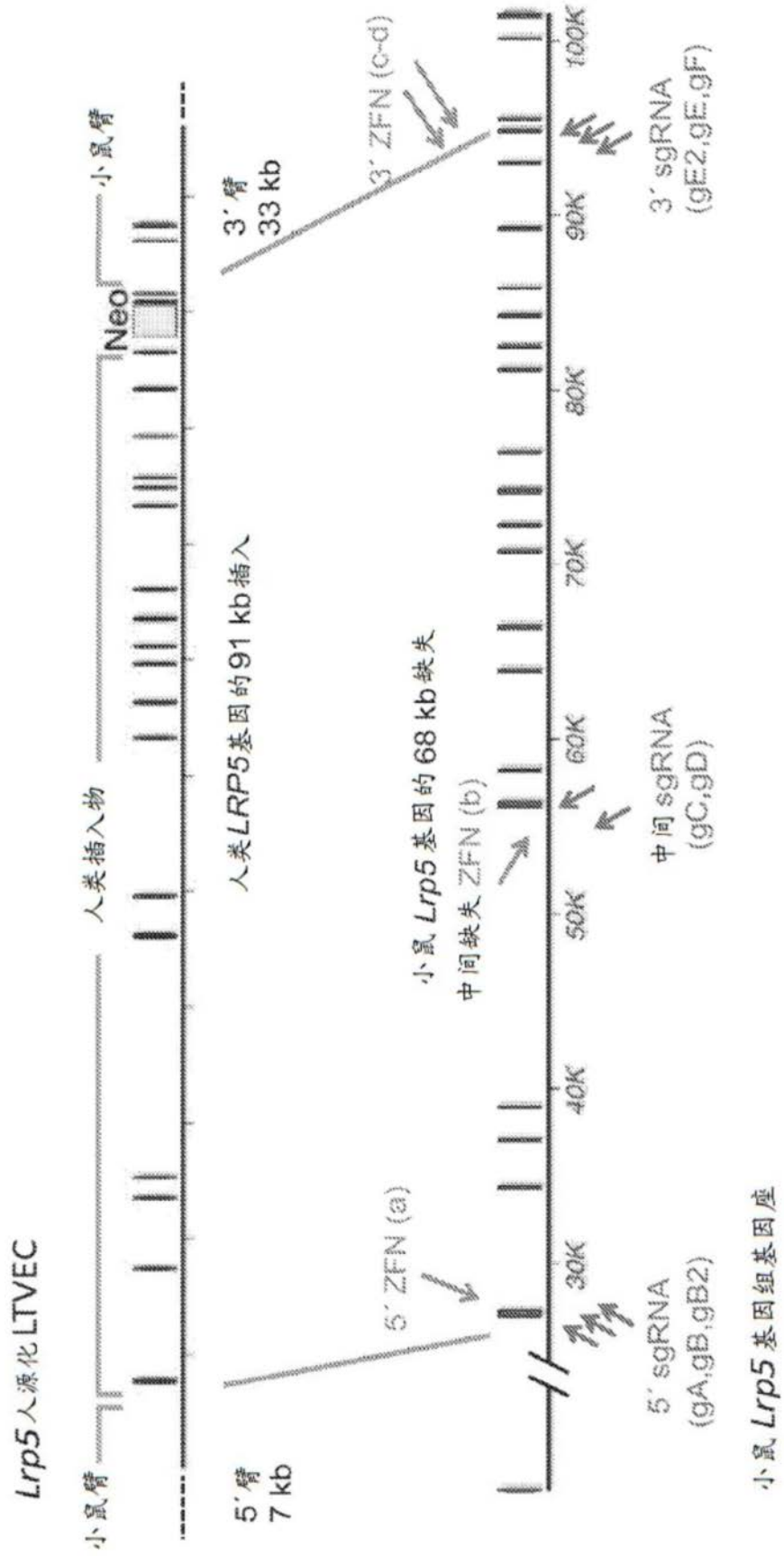


图34

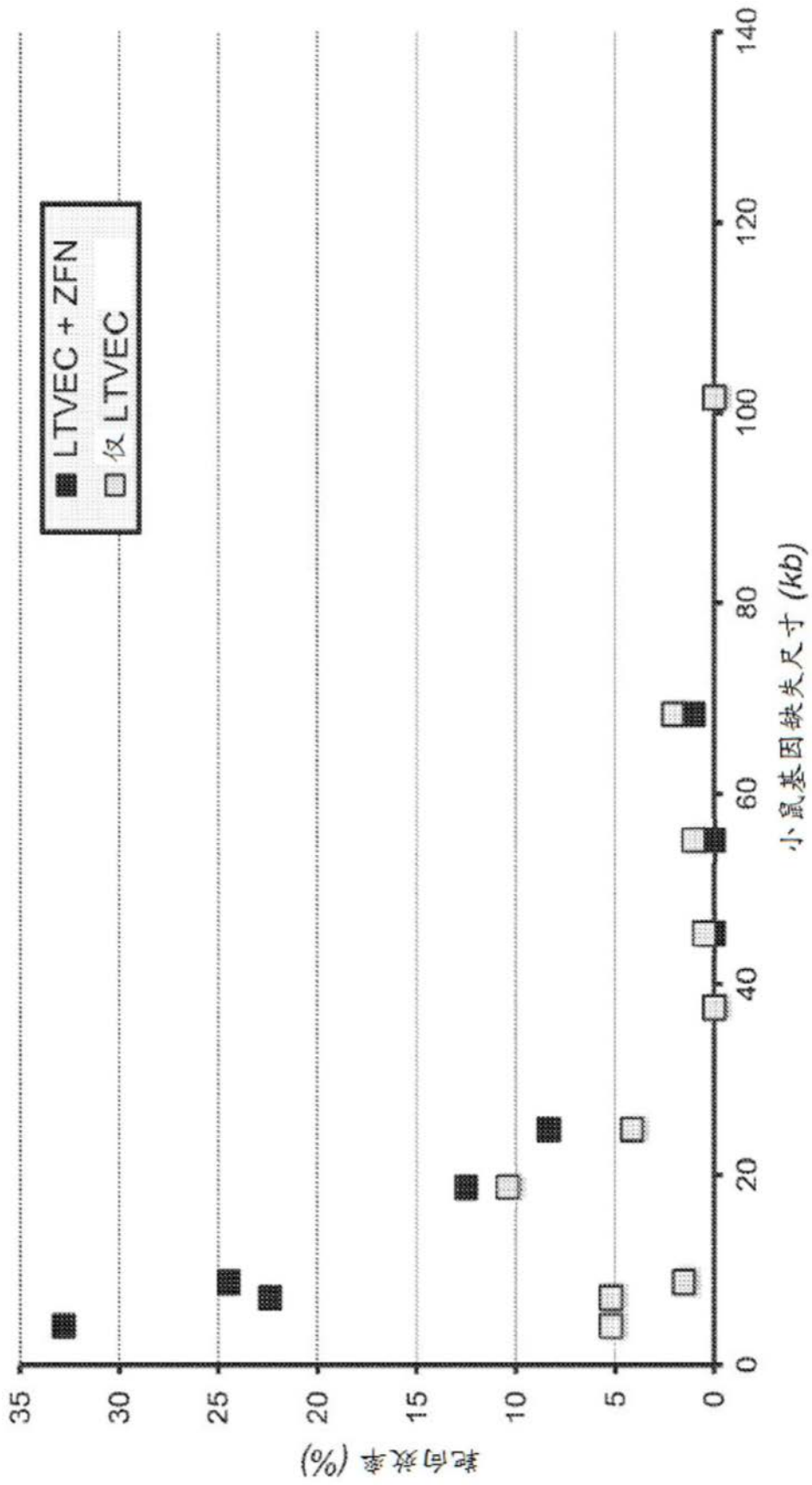


图35A

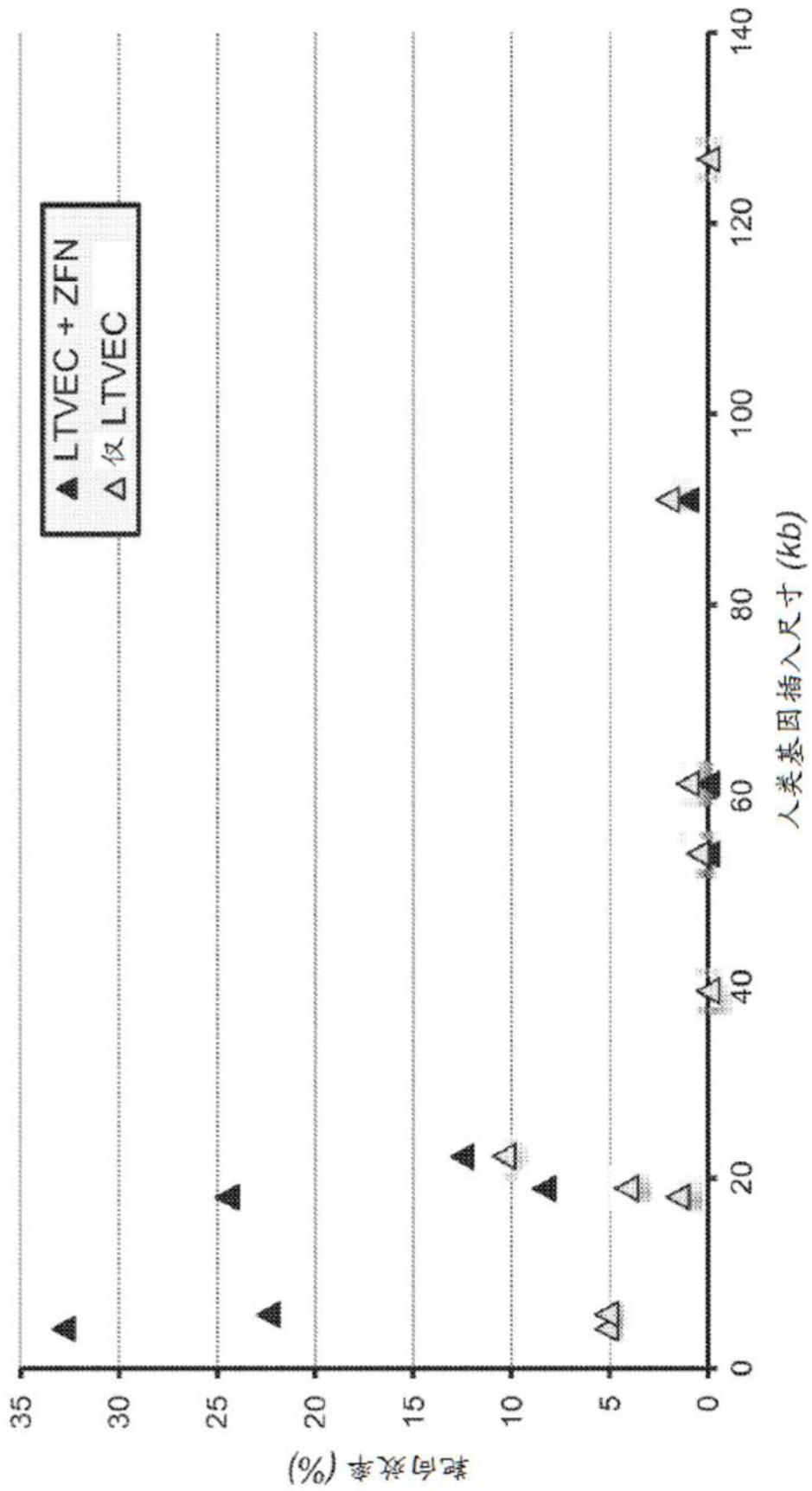


图35B



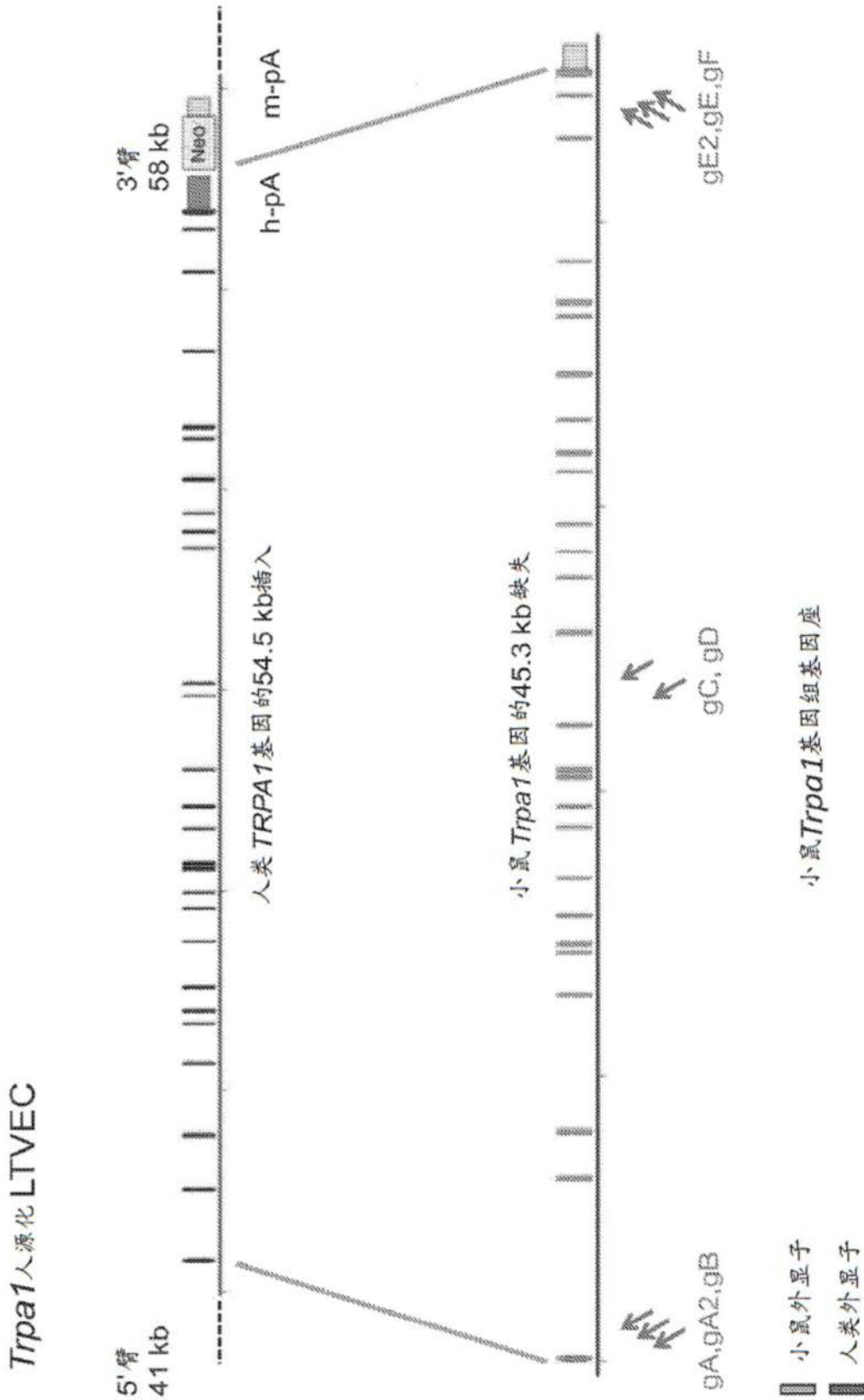


图36

Folh1人源化 LTVEC

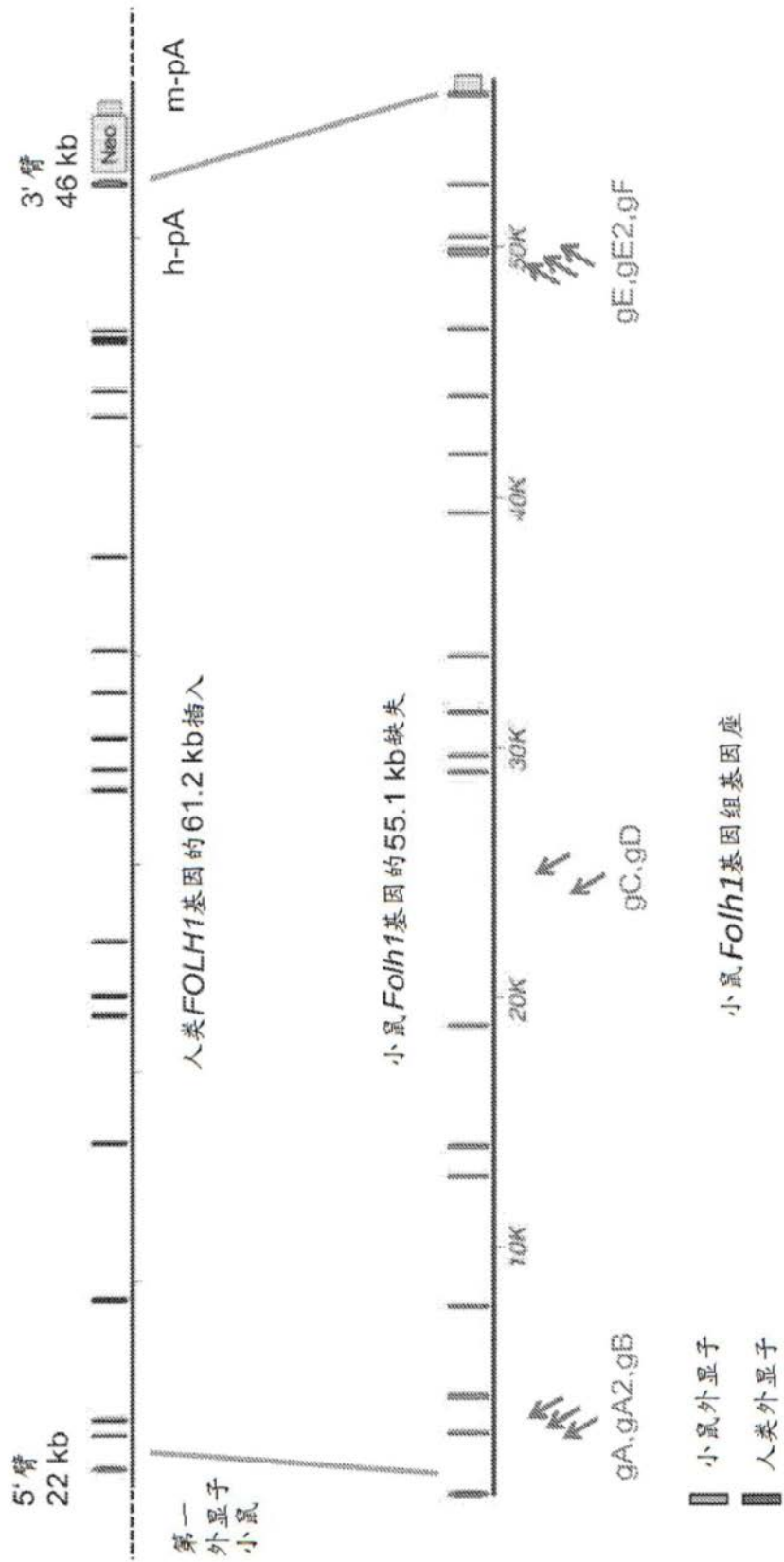


图37

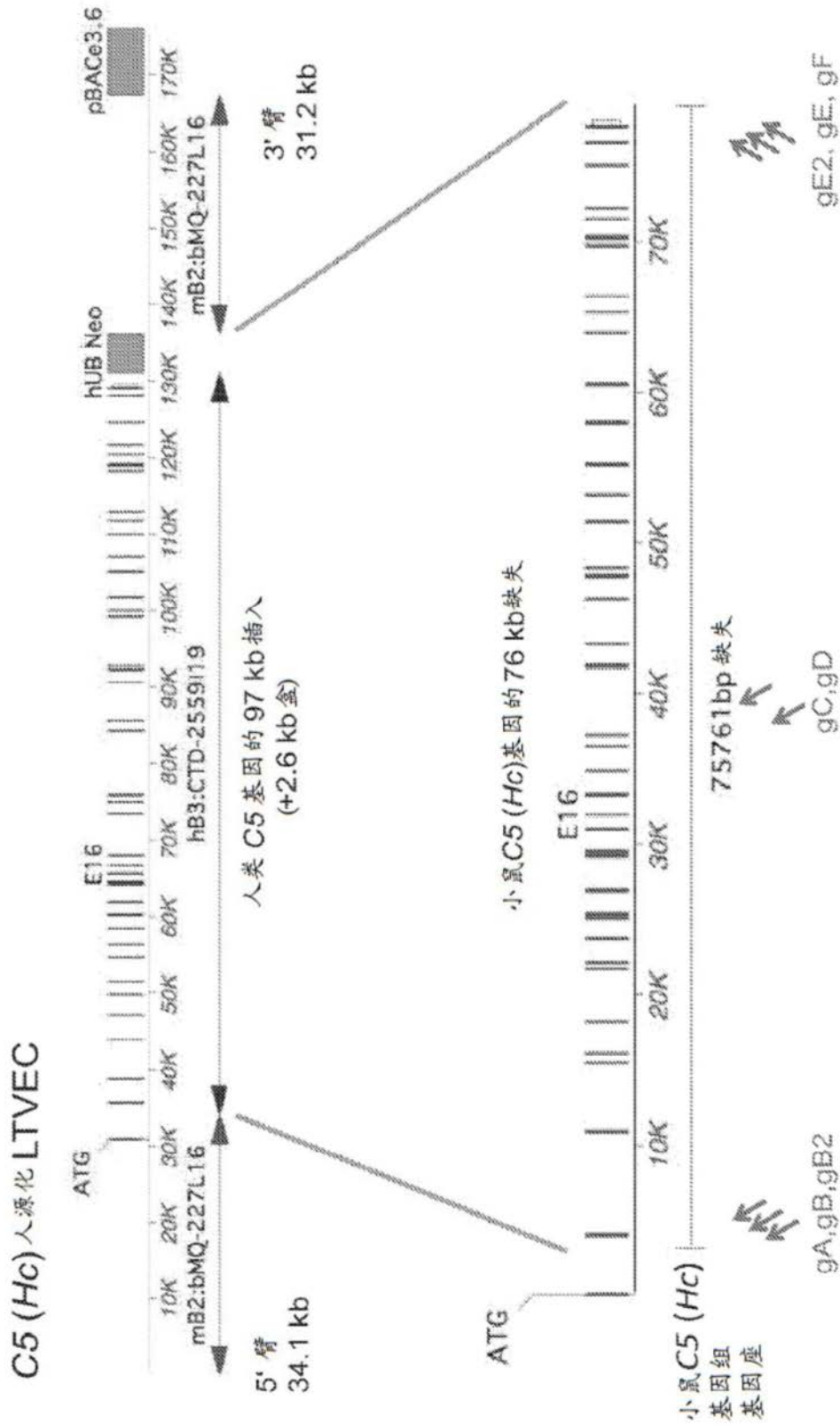


图38



Adamts5人源化LTVEC

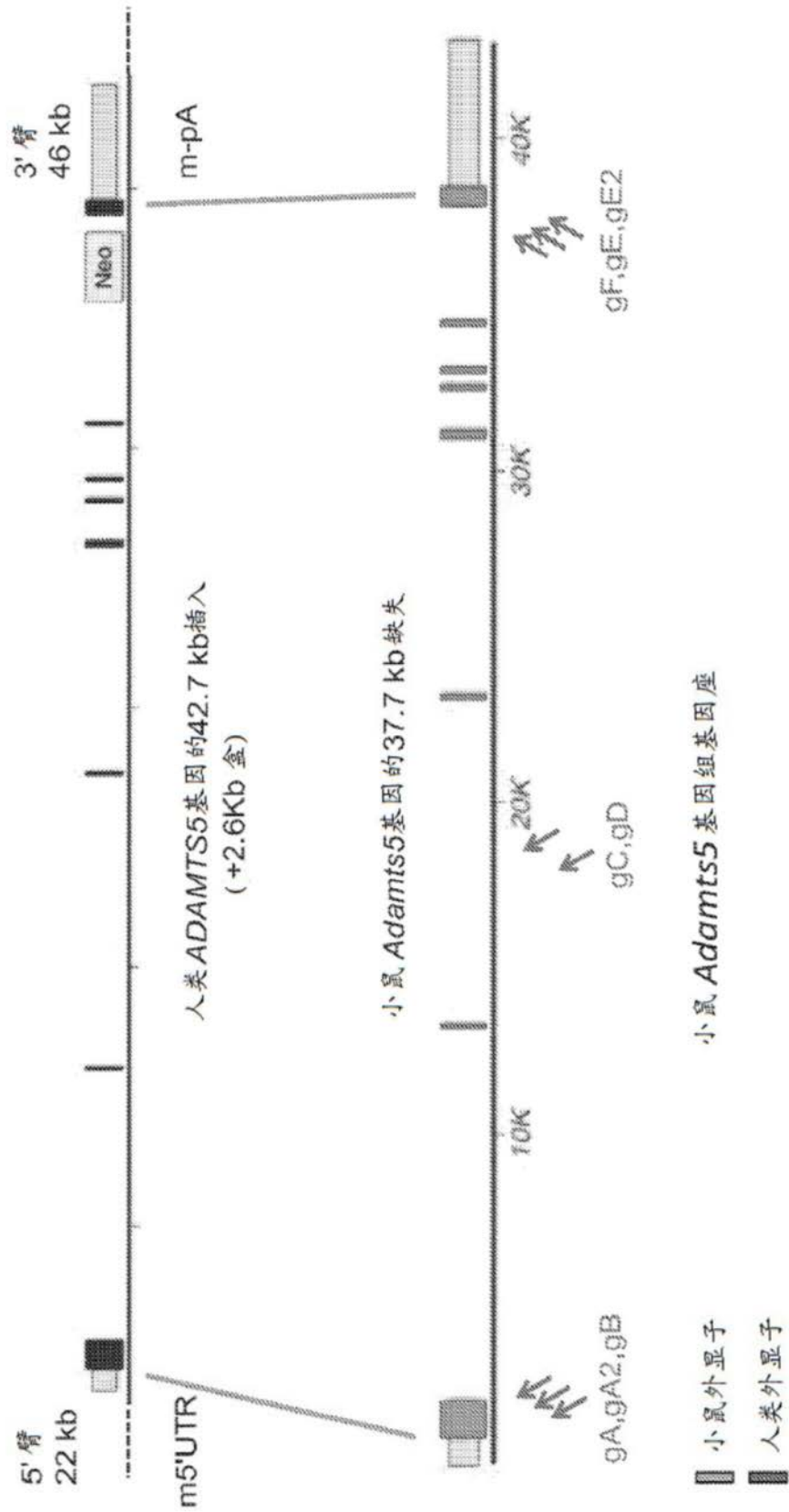


图39

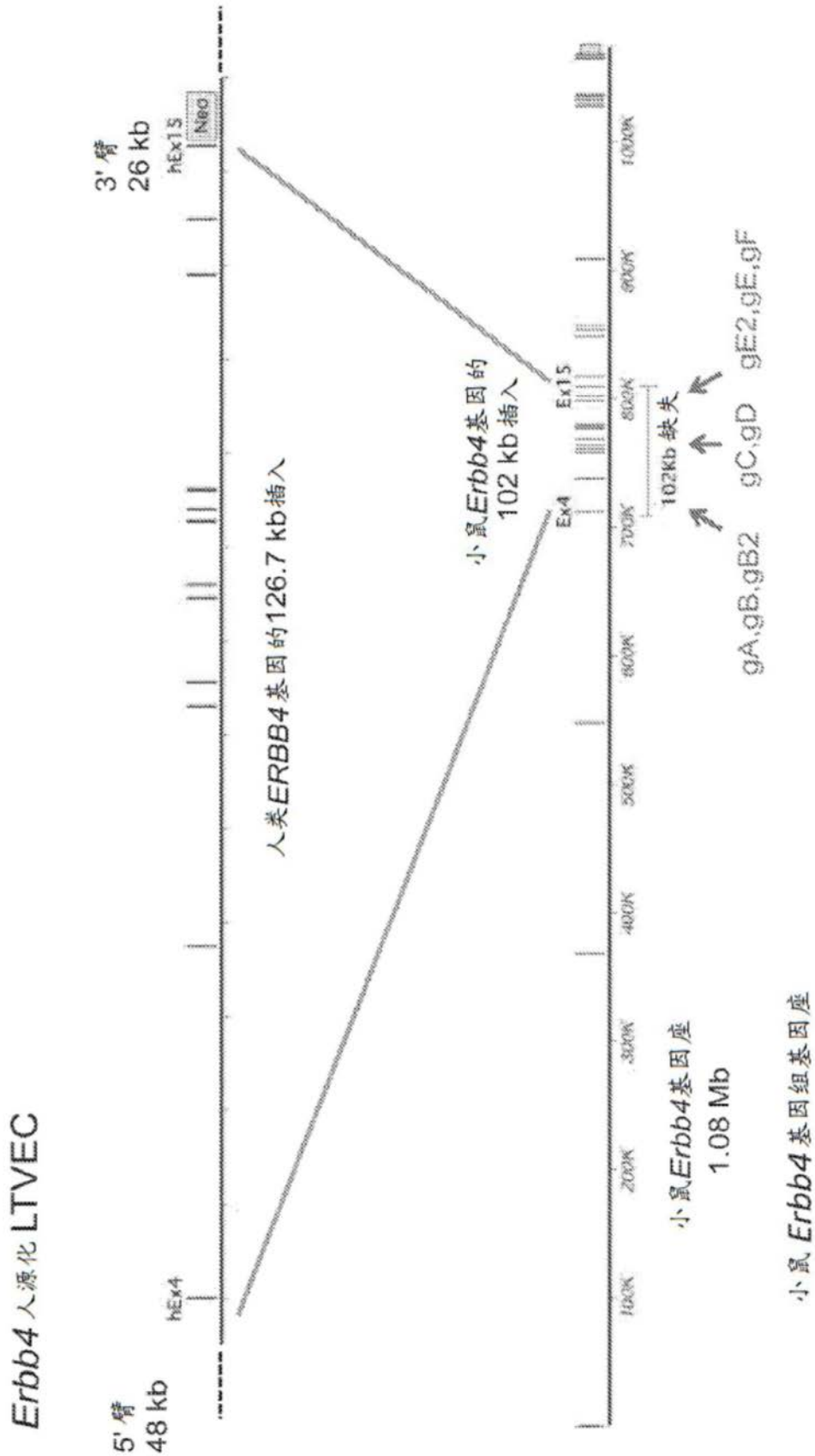


图40

Ror1 人源化 LTVEC

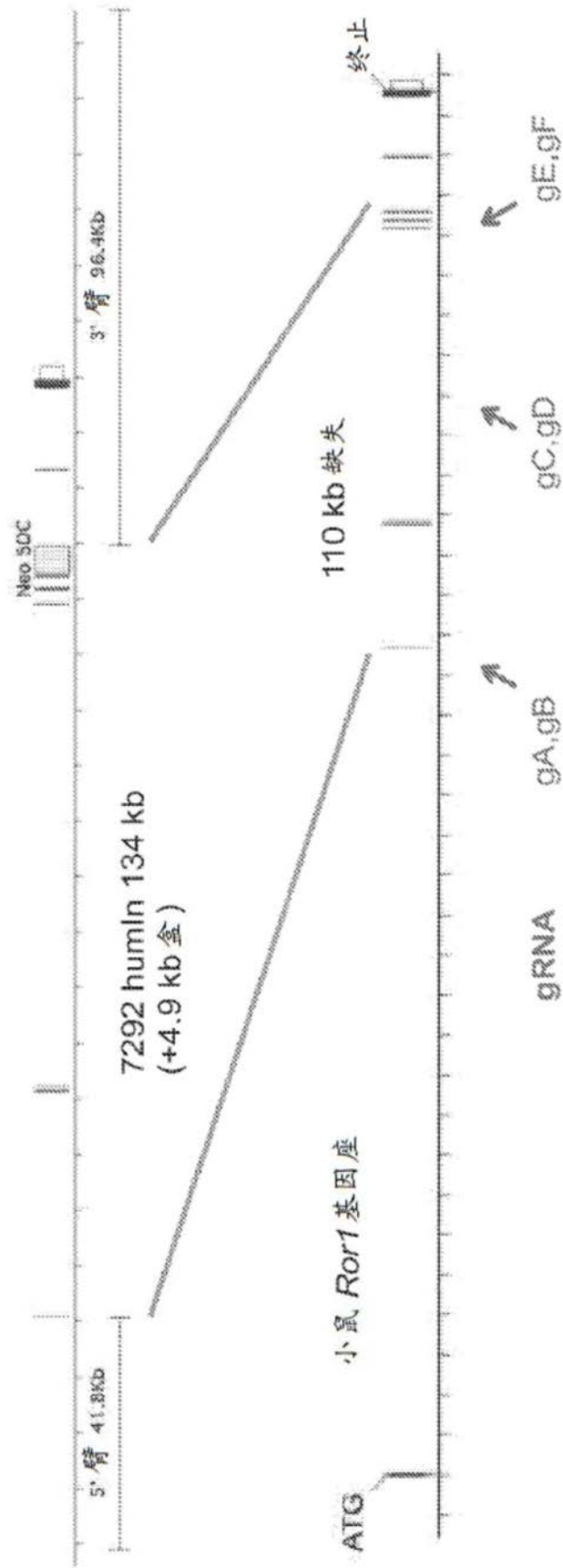


图41

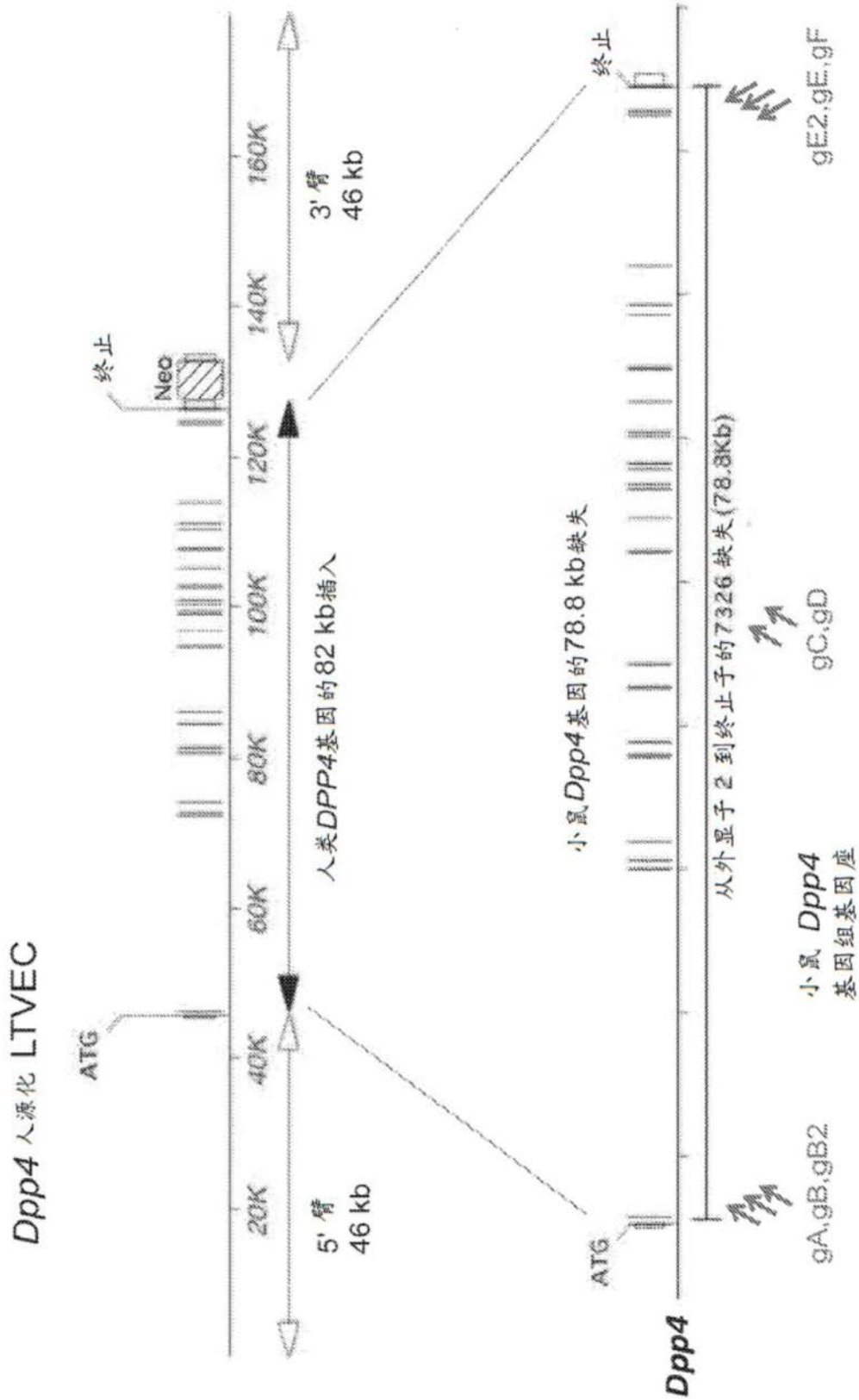


图42

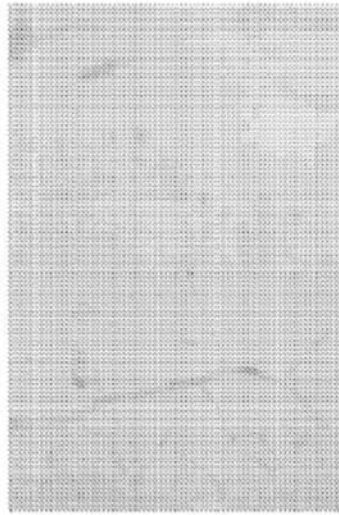


图 43C

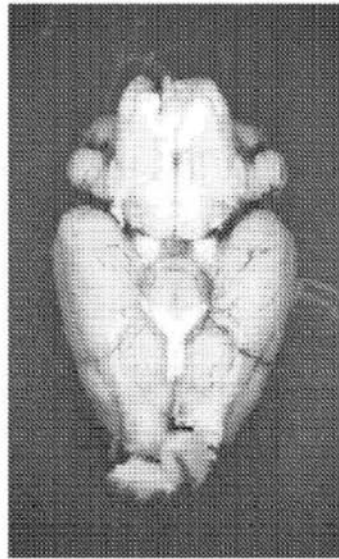


图 43B



图 43A

浮耳蛙

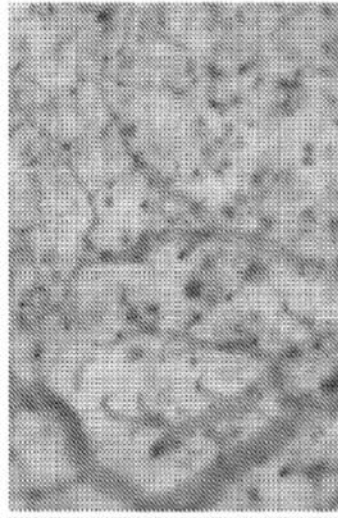


图 43F

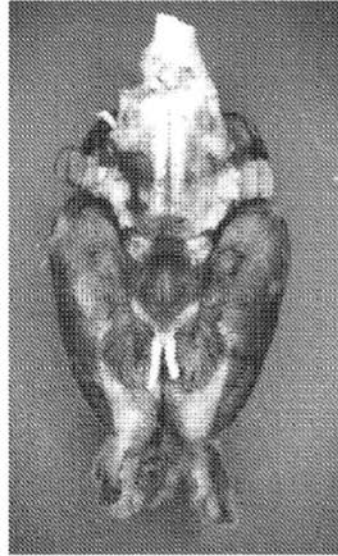


图 43E

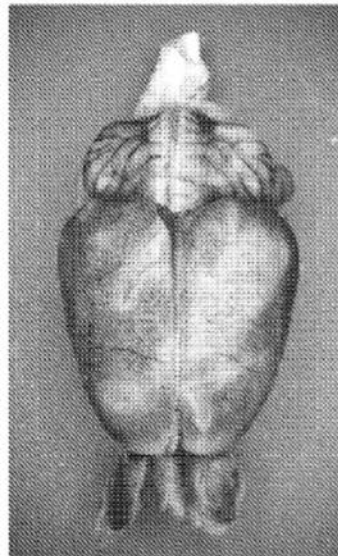


图 43D

ApoE<sup>+/-</sup>



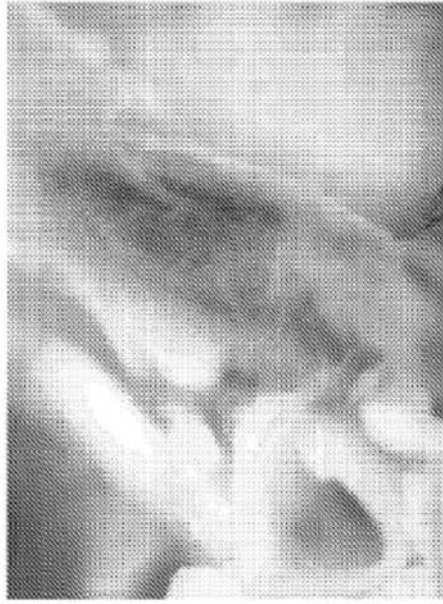


图 44B

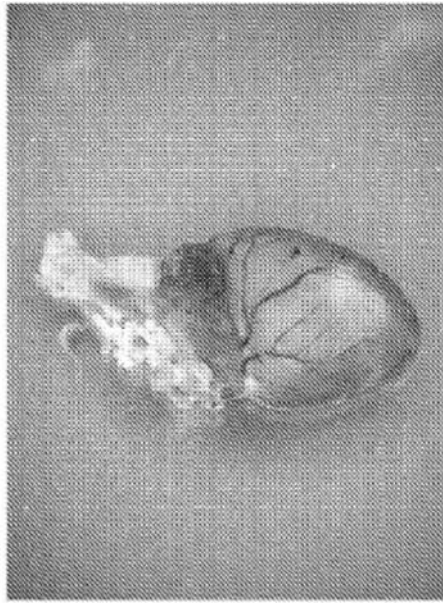


图 44A

野生型

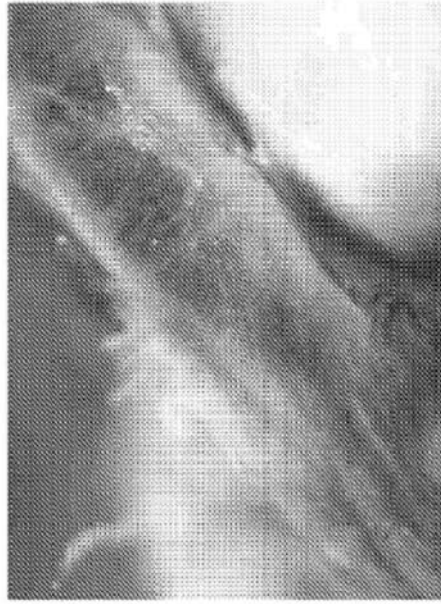


图 44D

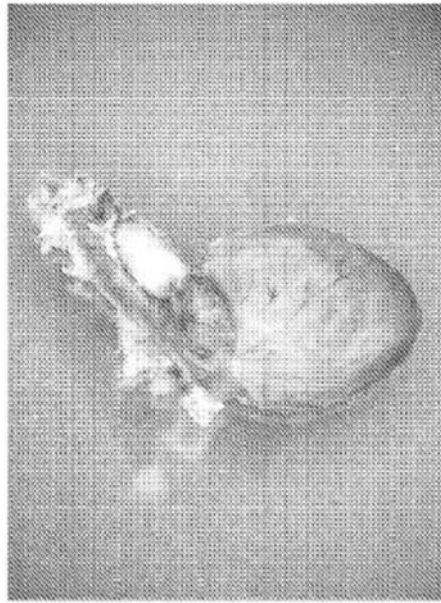


图 44C

ApoE<sup>+/-</sup>



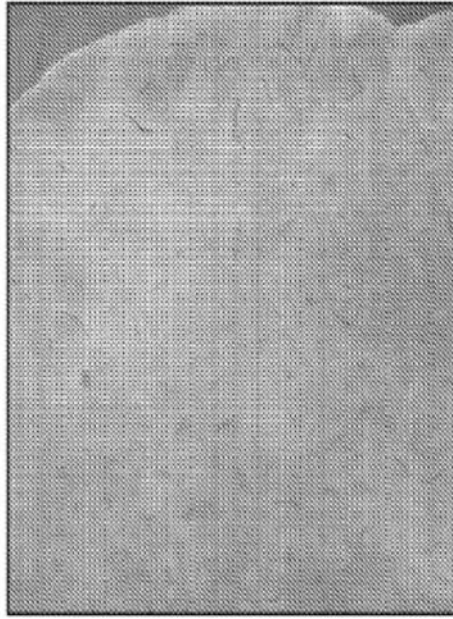


图 45B

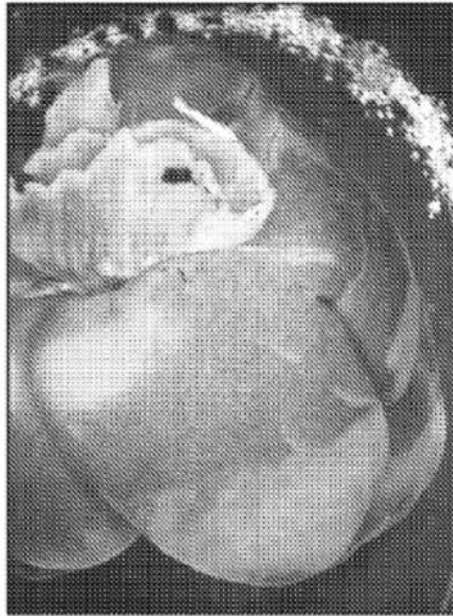


图 45A

月季花

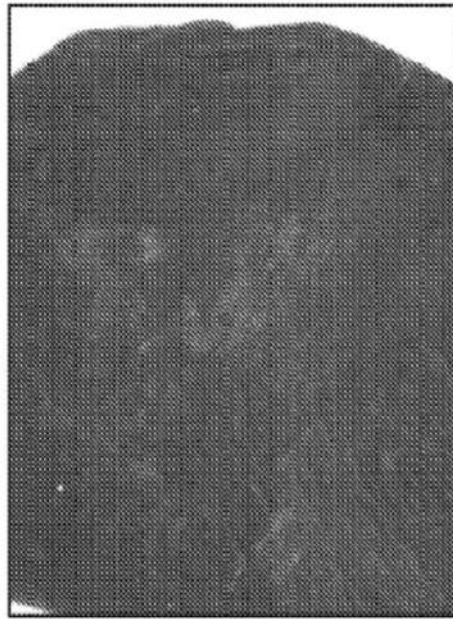


图 45D



图 45C

ApoE+/-

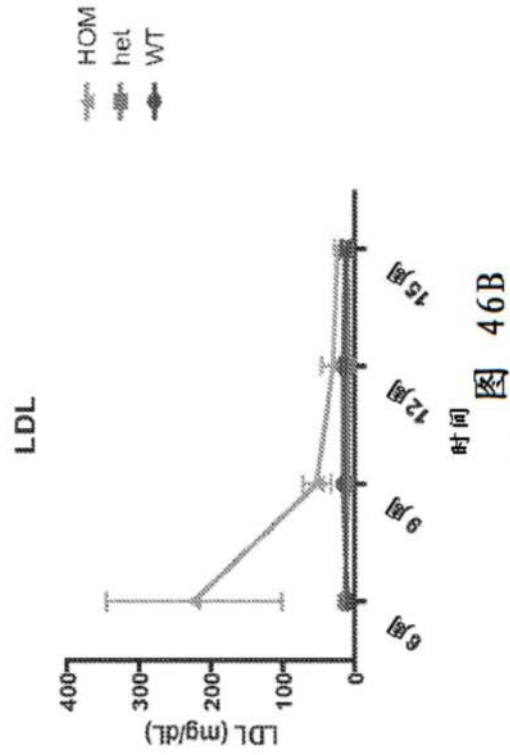


图 46B

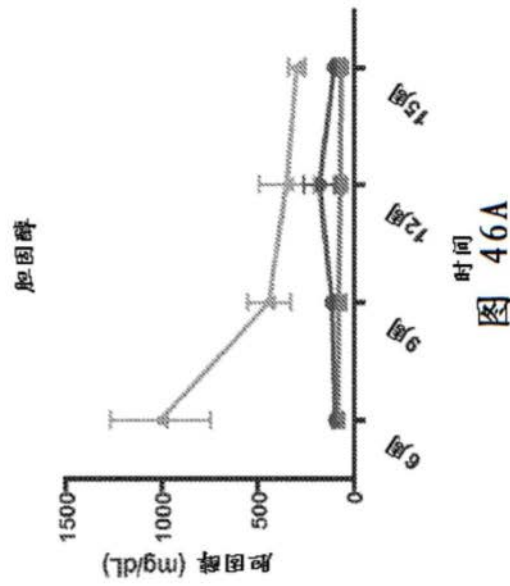


图 46A

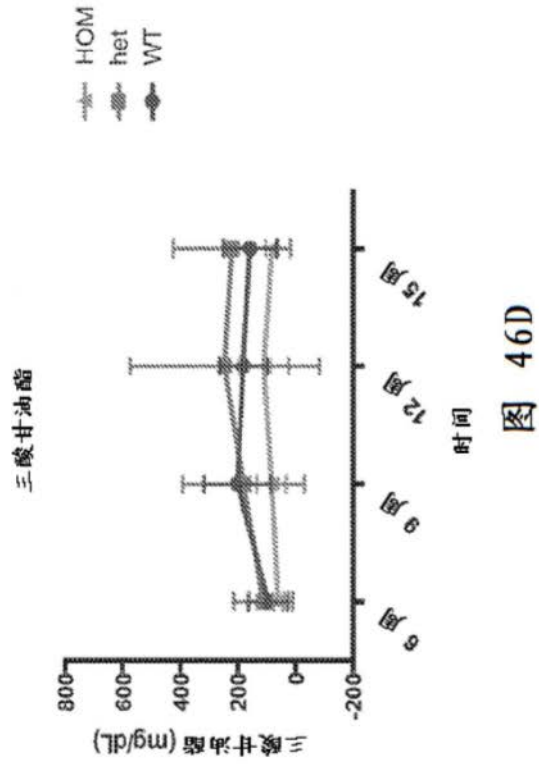


图 46D

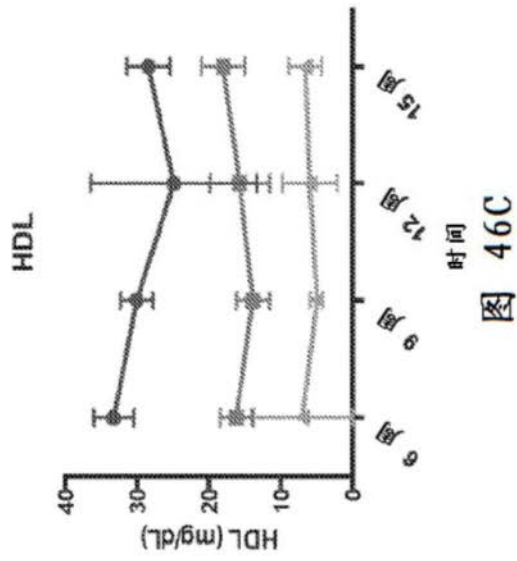


图 46C

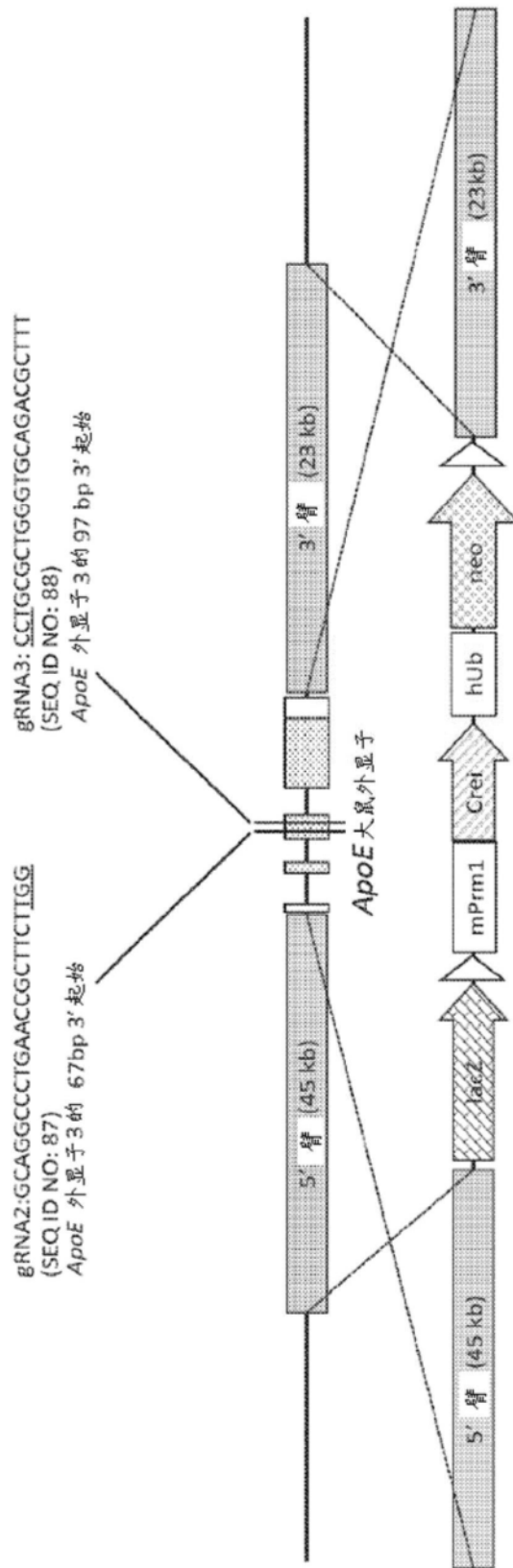


图47

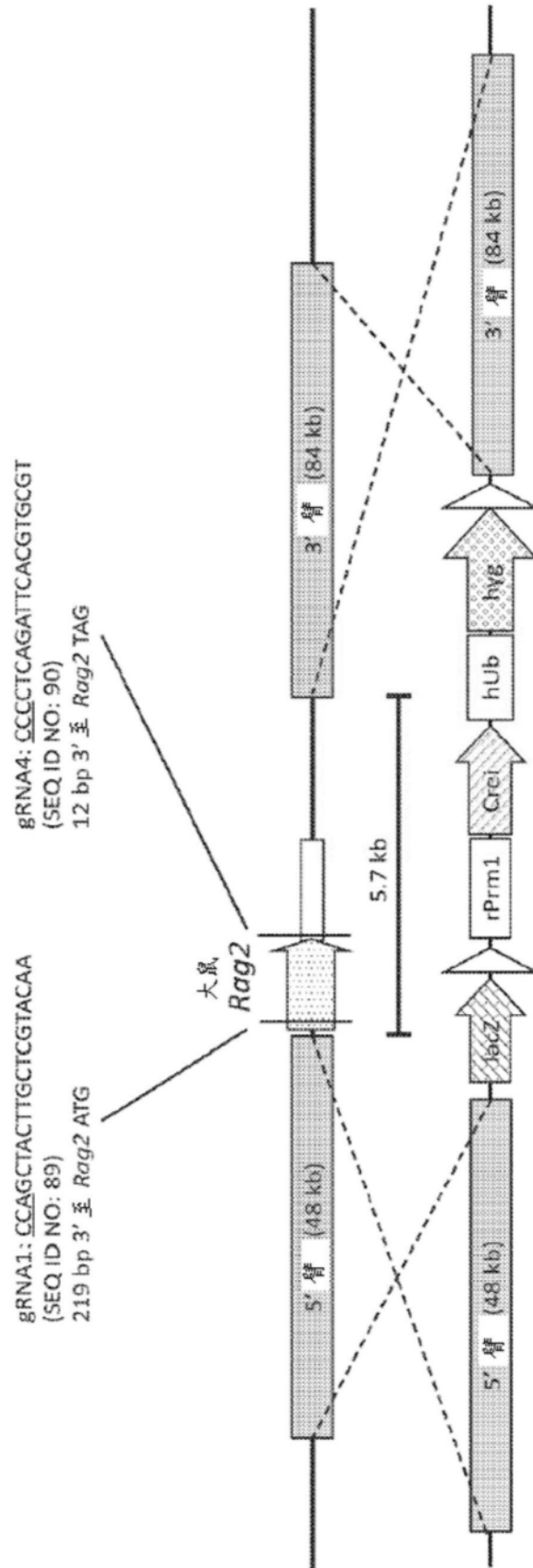


图48

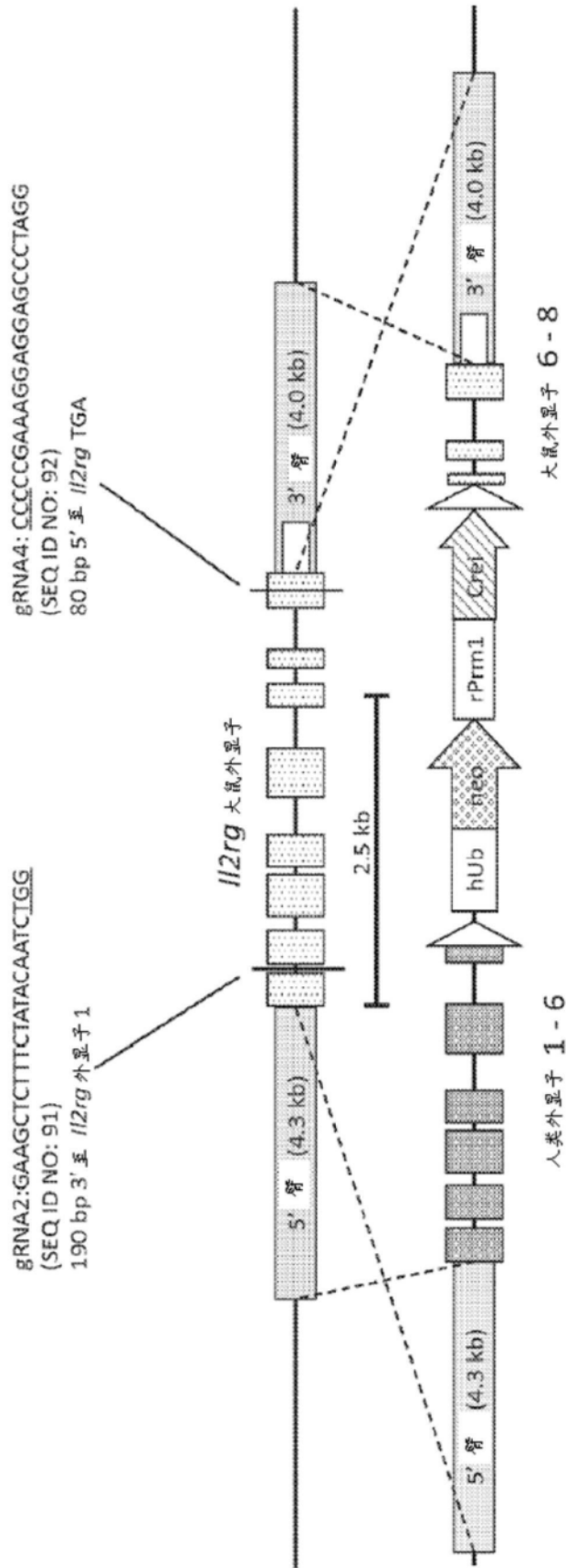


图49

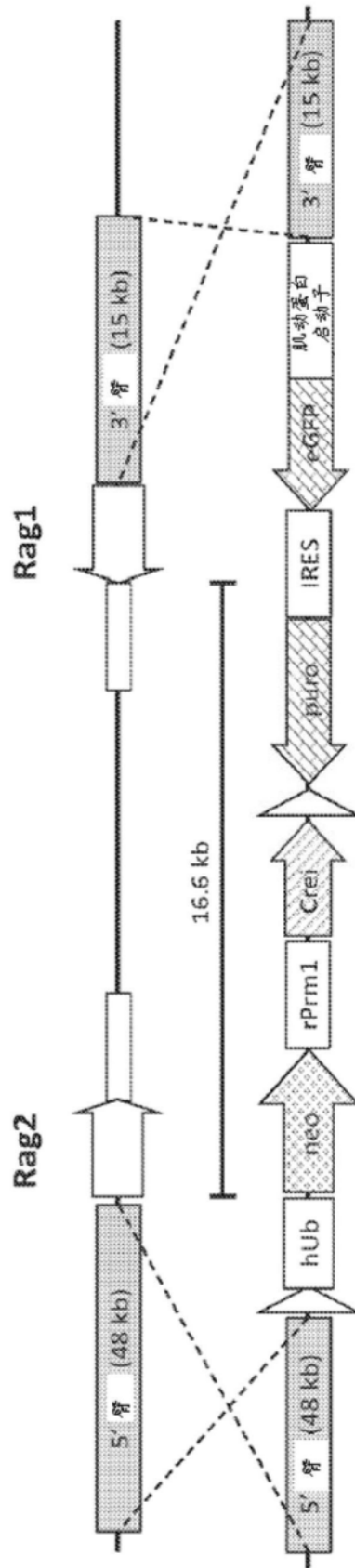


图50



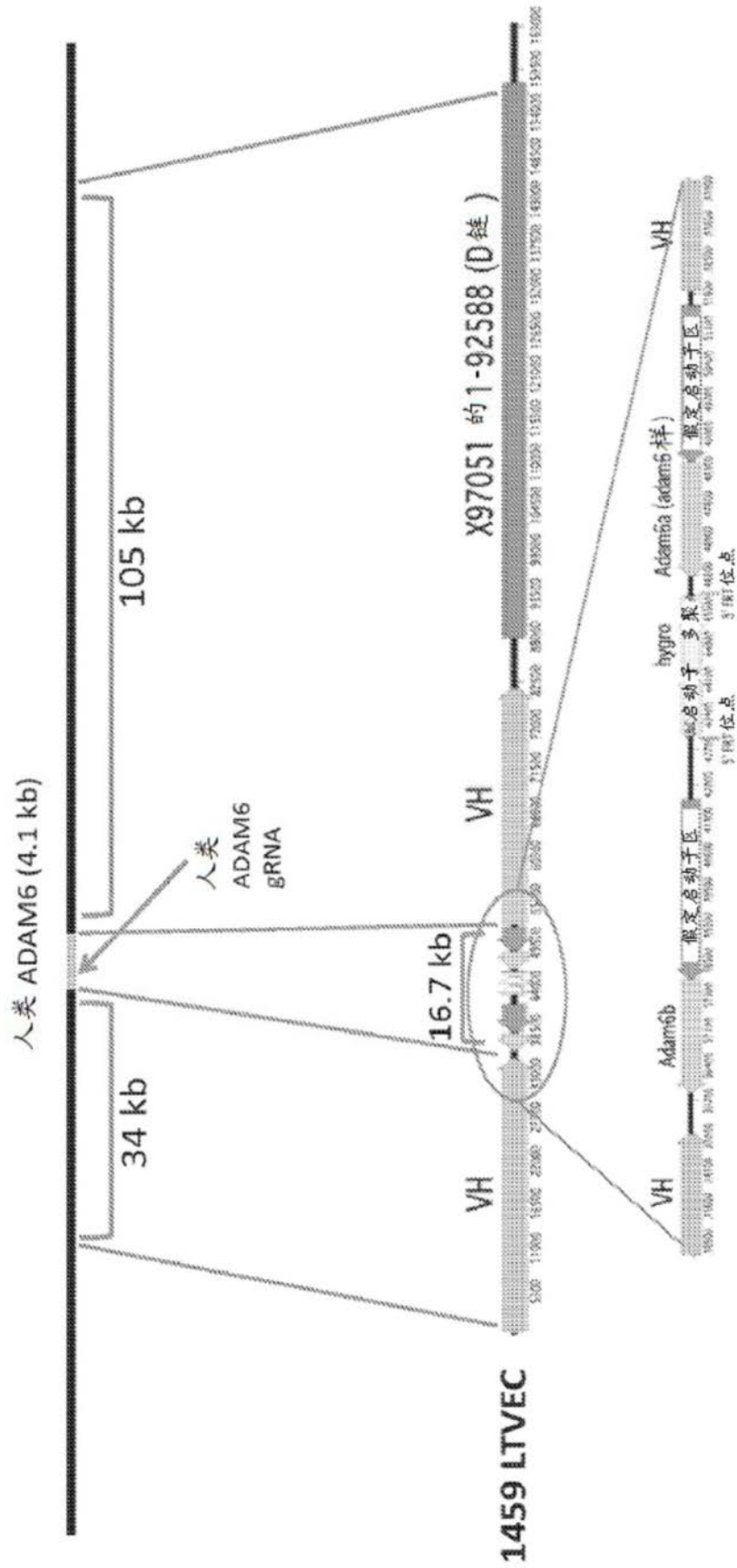
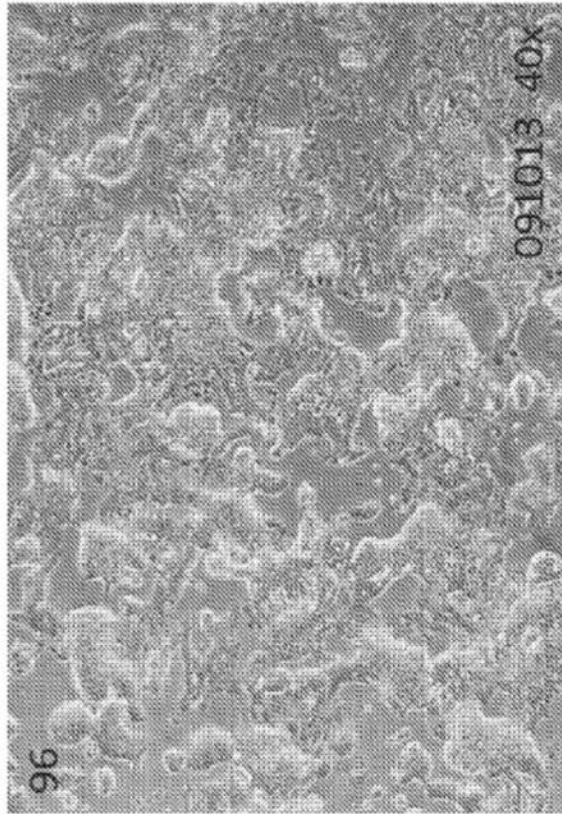
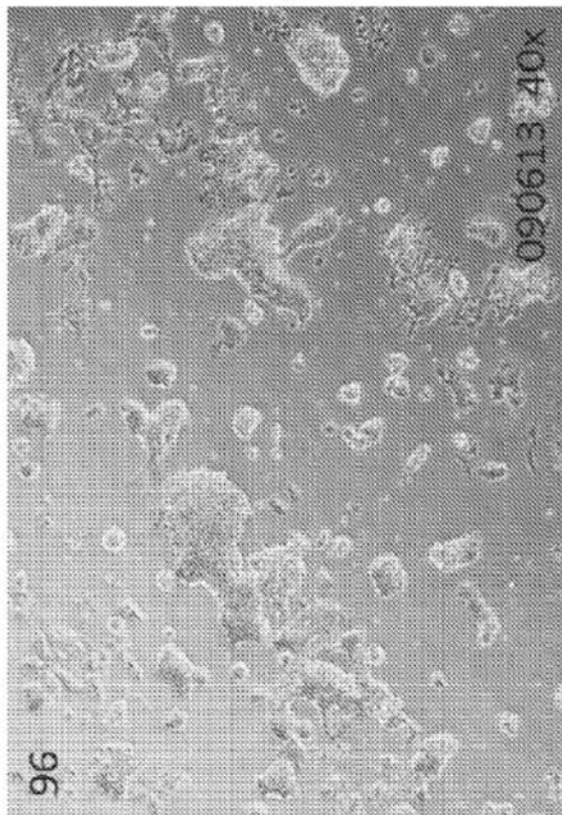


图51



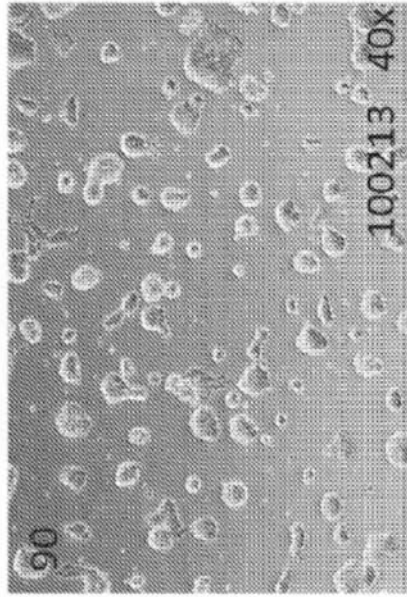
在 2i 中 12 天

图 52B

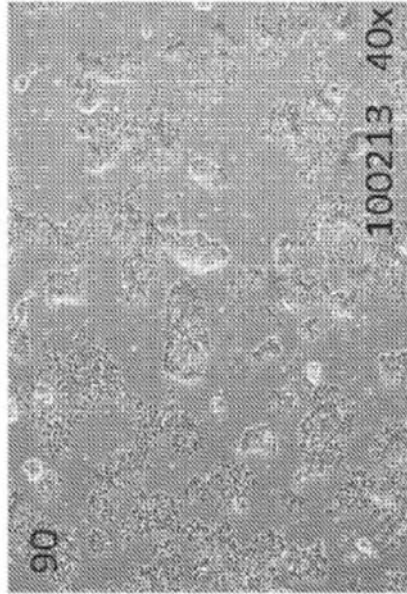


在 2i 中 8 天

图 52A

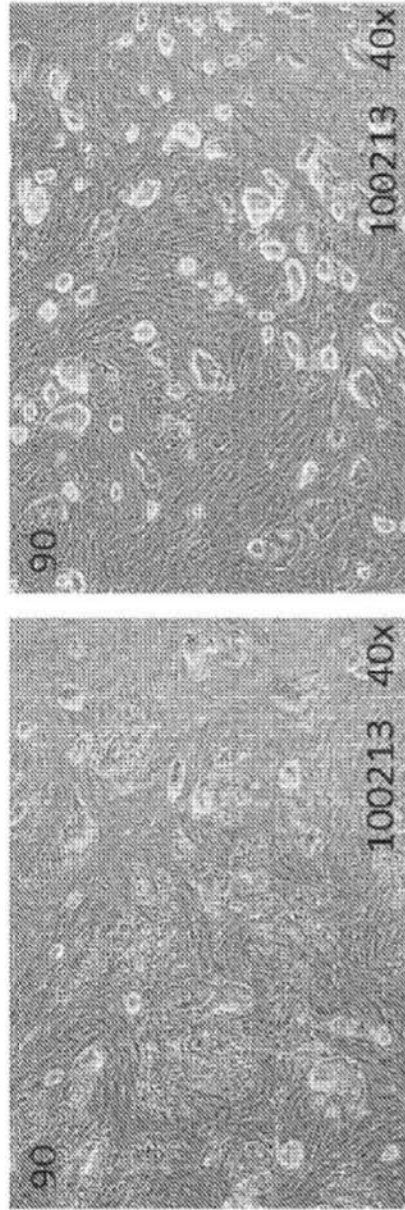


hVG2i  
图 53B



mTeSR-hLIF  
Matrigel  
图 53A





mTeSR-hLIF NuFF 何养液 hVG2i

图 53C 图 53D

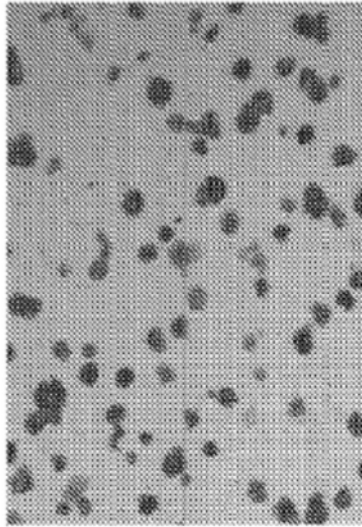


图54A

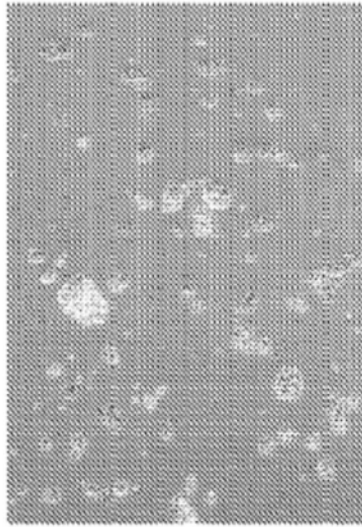


图54B

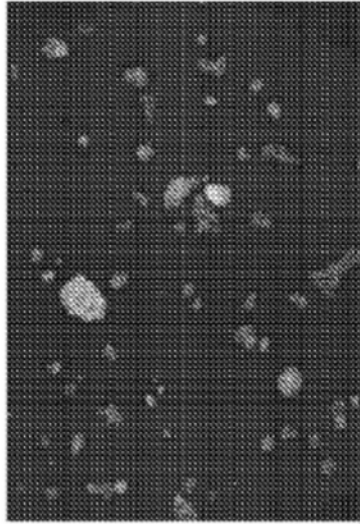


图54C

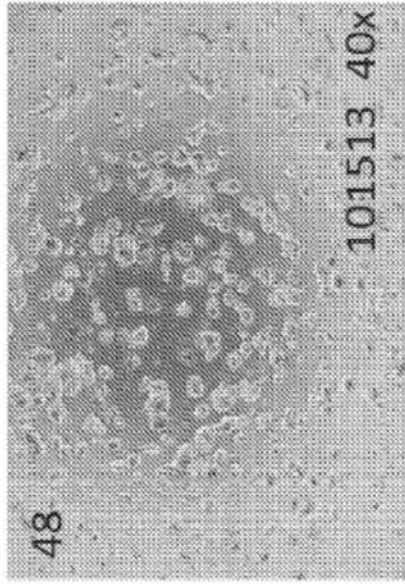


图 55A

在没有 ROCK 抑制剂的  
情况下通过胰蛋白酶  
单细胞传代

A large black arrow points from this text block towards the right, indicating the direction of the cell passage process.

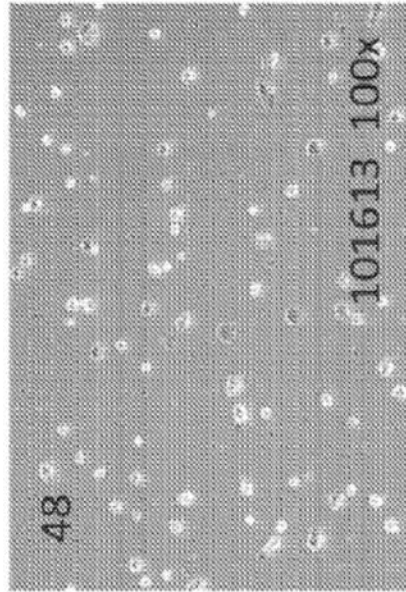


图 55B

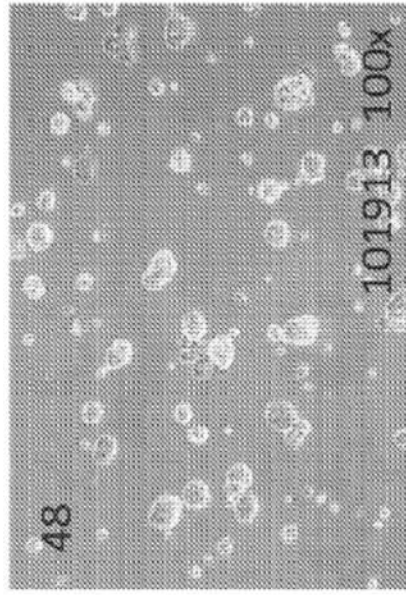


图 55C