(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105308184 B
(45)授权公告日 2020.06.02

(21)申请号 201480033094.0

(22)申请日 2014.04.16

(65)同一申请的已公布的文献号
申请公布号 CN 105308184 A

(43)申请公布日 2016.02.03

(30)优先权数据
61/812,319 2013.04.16 US
61/914,768 2013.12.11 US

(85)PCT国际申请进入国家阶段日
2015.12.10

(86)PCT国际申请的申请数据
PCT/US2014/034412 2014.04.16

(87)PCT国际申请的公布数据
WO2014/172489 EN 2014.10.23

(73)专利权人 瑞泽恩制药公司
地址 美国纽约州

(72)发明人 J·D·李 A·O·穆希卡
W·奥尔巴克 K·V·莱
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
代理人 郝传鑫

(51)Int.Cl.
C12N 15/85(2006.01)
C12N 15/12(2006.01)
A01K 67/027(2006.01)
C12N 5/10(2006.01)

审查员 代月函

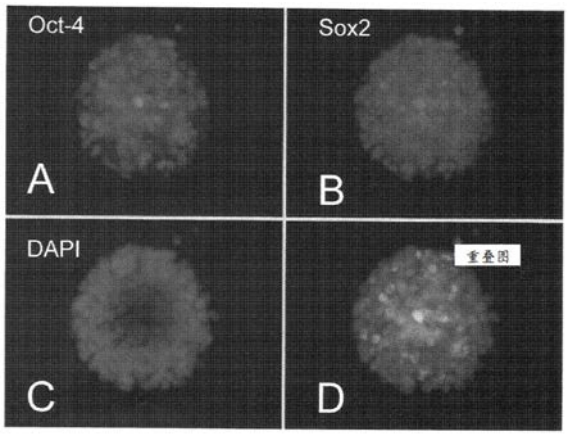
权利要求书5页 说明书116页
序列表11页 附图43页

(54)发明名称

大鼠基因组的靶向修饰

(57)摘要

提供用于使用如本文所述的包含各种内源性或外源性核酸序列的大靶向载体(LTVEC)来修饰目标大鼠基因组基因座的组合物和方法。也提供用于产生在它们的种系中包含一种或多种靶向遗传修饰的经遗传修饰的大鼠的组合物和方法。提供包括经遗传修饰的大鼠或大鼠细胞的组合物和方法,所述经遗传修饰的大鼠或大鼠细胞在大鼠白介素-2受体γ基因座、大鼠ApoE基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或大鼠Rag2/Rag1基因座中包含靶向遗传修饰。本文提供的各种方法和组合物允许这些经修饰的基因座通过种系传递。



CN 105308184 B

1.一种用于靶向修饰多能大鼠细胞群体中的目标基因组基因座的方法,其包括

(a)提供多能大鼠细胞群体,所述多能大鼠细胞群体通过在饲养细胞层上使用包括N2补充剂、B27补充剂、50U/ml至150U/ml白血病抑制因子(LIF)以及由MEK抑制剂和GSK3抑制剂组成的抑制剂组合物的培养基来培养大鼠胚胎干细胞获得,其中所述培养基不包含FGF受体抑制剂,所述饲养细胞层未被遗传修饰以表达LIF;

其中所述多能大鼠细胞缺乏c-Myc的表达;

其中在培养中所述多能大鼠细胞形成球样的、自由浮动的集落;

其中所述多能大鼠细胞为正常核型;并且

其中所述多能大鼠细胞具有种系胜任性;

(b)向所述多能大鼠细胞中引入包含侧接有与所述目标基因组基因座处的第一核酸序列互补的5'同源臂和与所述目标基因组基因座处的第二核酸序列互补的3'同源臂的插入核酸的大靶向载体(LTVEC),其中所述5'同源臂和所述3'同源臂的总和是至少10kb;以及

(c)鉴定在所述目标基因组基因座处包含靶向遗传修饰的经修饰的多能大鼠细胞,

其中在所述目标基因组基因座的修饰之后,一个或多个所述多能大鼠细胞保留种系胜任性。

2.一种用于制备人源化大鼠的方法,其包括:

(a)根据权利要求1所述的方法靶向多能大鼠细胞群体中的目标基因组基因座以形成经遗传修饰的多能大鼠细胞群体,其中所述插入核酸包括人核酸,可选地,所述人核酸序列至少5kb但小于400kb;

(b)向宿主大鼠胚胎中引入一个或多个所述经遗传修饰的多能大鼠细胞;以及

(c)在代孕母体中孕育所述宿主大鼠胚胎;其中所述代孕母体产生包含经修饰的基因组基因座的大鼠子代,所述经修饰的基因组基因座包含:

(i)包括所述人核酸序列的所述插入核酸的插入;

(ii)在所述目标基因组基因座处用人核酸序列替换大鼠核酸序列,其中所述被替换的大鼠核酸序列与所述人核酸序列同源或直系同源;

(iii)包含人核酸序列和大鼠核酸序列的嵌合核酸序列的插入;

或

(iv)其组合,

其中所述经修饰的基因组基因座能够通过种系传递。

3.如权利要求1或2所述的方法,其中所述靶向遗传修饰是双等位基因修饰。

4.如权利要求1或2所述的方法,其中所述多能大鼠细胞是大鼠胚胎干(ES)细胞。

5.如权利要求1或2所述的方法,其中所述多能大鼠细胞源于DA品系或ACI品系。

6.如权利要求1或2所述的方法,其中所述多能大鼠细胞具有如下一个或多个特征:

(a)表达至少一种选自下组的多能性标志物:Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2及Utf1;

(b)缺乏所述多能性标志物Ecat1和Rexo1的一种或多种的表达;

(c)缺乏中胚层标志物Brachyury和Bmpr2的一种或多种的表达;

(d)缺乏内胚层标志物Gata6、Sox17和Sox7的一种或多种的表达;或

(e)缺乏神经标志物巢蛋白和Pax6的一种或多种的表达。

7.如权利要求1或2所述的方法，其中所述培养基中LIF的浓度在75U/ml至125U/ml之间。

8.如权利要求7中所述的方法，其中所述培养基中LIF的浓度在90U/ml至110U/ml之间。

9.如权利要求8中所述的方法，其中所述培养基中LIF的浓度为100U/ml。

10.如权利要求1、2、8、9中任一项所述的方法，其中所述MEK抑制剂是PD0325901和/或所述GSK3抑制剂是CHIR99021。

11.如权利要求10中所述的方法，其中所述MEK抑制剂是浓度为0.8μM至1.2μM的PD0325901，并且所述GSK3抑制剂是浓度为2.5μM至3μM或3μM至3.5μM的CHIR99021。

12.如权利要求11中所述的方法，其中所述培养基中的LIF浓度为100U/ml，所述MEK抑制剂是浓度为1μM的PD0325901，并且所述GSK3抑制剂是浓度为3μM的CHIR99021。

13.如权利要求12中所述的方法，其中所述培养基包括如表3中所示的组分。

14.如权利要求1、2、8、9、11～13中任一项所述的方法，其中：

(a)所述LTVEC的所述5'同源臂和所述3'同源臂的总和是10kb至150kb；且/或

(b)所述5'同源臂是5kb至100kb或所述3'同源臂是5kb至100kb；且/或

(c)所述LTVEC是20kb至400kb。

15.如权利要求1、2、8、9、11～13中任一项所述的方法，其中所述靶向遗传修饰包括：

(a)用同源或直系同源核酸序列替换内源性大鼠核酸序列；

(b)内源性大鼠核酸序列的缺失；

(c)内源性大鼠核酸序列的缺失，

其中所述缺失在5kb至10kb、10kb至20kb、20kb至40kb、40kb至60kb、60kb至80kb、80kb至100kb、100kb至150kb、或150kb至200kb、200kb至300kb、300kb至400kb、400kb至500kb、500kb至1Mb、1Mb至1.5Mb、1.5Mb至2Mb、2Mb至2.5Mb、或2.5Mb至3Mb的范围内；

(d)外源性核酸序列的插入，所述外源性核酸序列在5kb至10kb、10kb至20kb、20kb至40kb、40kb至60kb、60kb至80kb、80kb至100kb、100kb至150kb、150kb至200kb、200kb至250kb、250kb至300kb、300kb至350kb、或350kb至400kb的范围内；

(e)外源性核酸序列的插入，所述外源性核酸序列包含与内源性大鼠核酸序列同源或直系同源的核酸序列；

(f)包含人核酸序列和大鼠核酸序列的嵌合核酸序列的插入；

(g)侧接有位点特异性重组酶靶序列的条件性等位基因的插入；或，

(h)报告基因的插入，所述报告基因可操作地连接至在大鼠细胞中具有活性的启动子。

16.如权利要求1、2、8、9、11～13中任一项所述的方法，其中所述靶向遗传修饰包括：

(i)与大鼠核酸序列同源或直系同源的人核酸序列在内源性基因座处的插入，可选地，其中所述插入的大小是5kb至500kb；

(ii)在所述内源性基因座处用所述同源或直系同源的人核酸序列替换所述大鼠核酸序列，可选地，其中所述插入的大小是5kb至500kb；

(iii)包含人核酸序列和大鼠核酸序列的嵌合核酸序列的插入；或

(iv)其组合。

17.如权利要求16所述的方法，其中所述插入或替换为5kb至10kb、10kb至20kb、20kb至40kb、40kb至60kb、60kb至80kb、80kb至100kb、100kb至150kb、150kb至200kb、200kb至

250kb、250kb至300kb、300kb至350kb、或350kb至400kb。

18.如权利要求1、2、8、9、11～13、17中任一项所述的方法，其中引入步骤(a)进一步包括引入核酸酶试剂或编码促进所述LTVEC与所述多能大鼠细胞中的所述目标基因组基因座之间的同源重组的所述核酸酶试剂的核酸。

19.如权利要求18所述的方法，其中所述核酸酶试剂包括

(a)锌指核酸酶；或，

(b)转录激活因子样效应物核酸酶(TALEN)。

20.如权利要求18所述的方法，其中所述核酸酶试剂包括成簇规律间隔短回文重复序列(CRISPR)相关蛋白(Cas蛋白)和导向RNA(gRNA)，所述导向RNA(gRNA)包含识别基因组靶序列和反式活化CRISPR RNA(tracrRNA)的CRISPR RNA(crRNA)，

其中所述基因组靶序列通过前间区序列邻近基序(PAM)序列以紧邻方式侧接在所述3'末端上。

21.如权利要求20所述的方法，其中所述目标基因组基因座包含核苷酸序列SEQ ID NO:1。

22.如权利要求20或21所述的方法，其中所述Cas蛋白是Cas9。

23.如权利要求22所述的方法，其中：

所述gRNA包括：SEQ ID NO:2，SEQ ID NO:3，SEQ ID NO:4，SEQ ID NO:5，SEQ ID NO:6，SEQ ID NO:7，或SEQ ID NO:8。

24.如权利要求1、2、8、9、11～13、17、19～21、23中任一项所述的方法，其中所述鉴定步骤采用用于评估所述目标基因组基因座处的等位基因修饰(MOA)的定量测定。

25.如权利要求1、2、8、9、11～13、17、19～21、23中任一项所述的方法，其中当在体外将所述经修饰的多能大鼠细胞在饲养细胞层上培养时其保留形成球样集落的能力，从而在所述目标基因组基因座的修饰之后，当所述经修饰的多能大鼠细胞体外涂铺于饲养细胞层上时形成的所述集落的50％具有球形样形态。

26.如权利要求1、2、8、9、11～13、17、19～21、23中任一项所述的方法，其中所述目标基因组基因座：

(i)是白介素-2受体γ基因座、ApoE基因座、Rag1基因座、Rag2基因座、或Rag2/Rag1基因座；

(ii)是免疫球蛋白基因座，

可选地，其中所述目标基因组基因座编码哺乳动物免疫球蛋白重链可变区氨基酸序列，或者

可选地，其中所述目标基因组基因座包括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，可选地，其中所述基因组DNA序列包含未重排哺乳动物λ和/或κ轻链可变区核酸序列；或者

(iii)是T细胞受体基因座，可选地，所述T细胞受体基因座是T细胞受体α基因座。

27.如权利要求1、2、8、9、11～13、17、19～21、23中任一项所述的方法，其中向所述多能大鼠细胞中引入所述LTVEC导致：(i)在所述目标基因组基因座处的内源性大鼠核酸序列缺失；(ii)外源性核酸序列插入在所述目标基因组基因座处；或(iii)其组合。

28.如权利要求27所述的方法，其中

(a)在所述目标基因组基因座处的所述内源性大鼠核酸序列的所述缺失是至少10kb;或,

(b)在所述目标基因组基因座处的所述内源性大鼠核酸序列的所述缺失是5kb至10kb、10kb至20kb、20kb至40kb、40kb至60kb、60kb至80kb、80kb至100kb、100kb至150kb、或150kb至200kb、200kb至300kb、300kb至400kb、400kb至500kb、500kb至1Mb、1Mb至1.5Mb、1.5Mb至2Mb、2Mb至2.5Mb、或2.5Mb至3Mb;

(c)在所述目标基因组基因座处的所述外源性核酸序列的所述插入是至少5kb;或,

(d)在所述目标基因组基因座处的所述外源性核酸序列的所述插入是5kb至10kb、10kb至20kb、20kb至40kb、40kb至60kb、60kb至80kb、80kb至100kb、100kb至150kb、150kb至200kb、200kb至250kb、250kb至300kb、300kb至350kb、或350kb至400kb。

29.如权利要求26所述的方法,其中

(a)所述靶向遗传修饰在所述白介素-2受体γ基因座处并且所述靶向遗传修饰导致白介素-2受体γ蛋白活性降低或不存在;

(b)所述靶向遗传修饰在所述ApoE基因座处并且所述靶向遗传修饰导致ApoE蛋白活性降低或不存在;

(c)所述靶向遗传修饰在所述Rag1基因座处并且所述靶向遗传修饰导致Rag1蛋白活性降低或不存在;

(d)所述靶向遗传修饰在所述Rag2基因座处并且所述靶向遗传修饰导致Rag2蛋白活性降低或不存在;或,

(e)所述靶向遗传修饰在所述Rag2/Rag1基因座处的所述靶向遗传修饰导致Rag2蛋白活性和Rag1蛋白活性降低或不存在。

30.如权利要求26所述的方法,其中所述靶向遗传修饰在所述白介素-2受体γ基因座处并且所述靶向遗传修饰包括

(a)整个大鼠白介素-2受体γ编码区或其一部分的缺失;

(b)用人白介素-2受体γ编码区或其一部分替换所述整个大鼠白介素-2受体γ编码区或其一部分;

(c)用所述人白介素-2受体γ的外结构域替换所述大鼠白介素-2受体γ编码区的外结构域;或,

(d)包含所述白介素-2受体γ编码区的所述白介素-2受体γ基因座的至少3kb的缺失。

31.如权利要求26所述的方法,其中所述靶向遗传修饰在所述ApoE基因座处并且所述靶向遗传修饰包括:

(a)整个ApoE编码区或其一部分的缺失;或,

(b)包含所述ApoE编码区的所述ApoE基因座的至少1.8kb的缺失。

32.如权利要求26所述的方法,其中所述靶向遗传修饰在所述Rag2基因座处并且所述靶向遗传修饰包括:

(a)整个Rag2编码区或其一部分的缺失;或,

(b)包含所述Rag2编码区的所述Rag2基因座的至少5.7kb的缺失。

33.如权利要求26所述的方法,其中所述靶向遗传修饰在所述Rag1/Rag2基因座处并且所述靶向遗传修饰包括:

(a)缺失整个Rag2编码区或其一部分以及缺失整个Rag1编码区或其部分;或,

(b)缺失包含所述Rag2和所述Rag1编码区的所述Rag2/Rag1基因座的至少16kb。

34.如权利要求1、2、8、9、11～13、17、19～21、23、28～33中任一项所述的方法,其中所述插入核酸包含表达盒,所述表达盒包含编码选择性标志物的多核苷酸。

35.如权利要求34所述的方法,其中所述表达盒包含可操作地连接至所述目标基因组基因座处的内源性启动子的lacZ基因和可操作地连接至选择性标志物基因的人泛素启动子。

36.如权利要求1、2、8、9、11～13、17、19～21、23、28～33、35中任一项所述的方法,其中所述插入核酸包含自我缺失性选择盒。

37.如权利要求36所述的方法,其中所述自我缺失性选择盒包含可操作地连接至在所述多能大鼠细胞中具有活性的启动子的选择性标志物和可操作地连接至雄性生殖细胞特异性启动子的编码重组酶的多核苷酸,其中所述自我缺失性盒由被所述重组酶识别的重组识别位点侧接。

38.如权利要求37所述的方法,其中

(a)所述雄性生殖细胞特异性启动子是鱼精蛋白-1启动子;或,

(b)所述重组酶基因编码Cre,并且所述重组识别位点是loxP位点。

39.如权利要求26所述的方法,其中在所述目标基因组基因座处的所述外源性核酸序列的插入包含可操作地连接至内源性白介素-2受体γ启动子、内源性ApoE启动子、内源性Rag1启动子或内源性Rag2启动子的报告核酸序列。

40.如权利要求39所述的方法,其中所述报告核酸序列编码包括以下的报告蛋白:β-半乳糖苷酶、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强黄色荧光蛋白(EYFP)、Emerald、增强绿色荧光蛋白(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。

大鼠基因组的靶向修饰

[0001] 通过EFS网以文本文件形式提交的序列表的引用

[0002] 序列表的正式文本通过EFS网以文件名为444701SEQLIST.TX T,于2014年4月16日创建,并且具有15千字节大小的ASCII格式的序列表形式电子提交,并且与说明书一同提交。这个ASCII格式的文件中含有的序列表是说明书的一部分,并且以引用的方式整体并入本文。

发明领域

[0003] 能够在一次或多次连续体外遗传修饰之后维持多能性,并且能够通过种系向后代传递靶向遗传修饰的分离的非人全能或多能干细胞,特别是大鼠胚胎干细胞。用于经由原核细胞中的细菌同源重组(BHR)来修饰目标大鼠基因组基因座的组合物和方法。用于使用与核酸内切酶组合的大靶向载体(LTVEC)来遗传修饰目标大鼠基因组基因座的组合物和方法。用于产生包含一种或多种靶向遗传修饰的经遗传修饰的大鼠的组合物和方法。

[0004] 发明背景

[0005] 尽管大鼠已被认为是一种可重演各种人类疾病(包括但不限于心血管疾病(例如高血压)、代谢疾病(例如肥胖症、糖尿病)、神经疾病(例如疼痛病变)和多种癌症)的病理学的重要动物模型系统,但相较于小鼠,大鼠在人疾病建模方面的使用受到限制,这部分是由于可在一系列体外遗传修饰(例如一次或多次连续电穿孔)之后维持它们多能性的种系可传递多能大鼠细胞的不可用性,并且部分是由于缺乏允许在多能大鼠细胞中引入或缺失大基因组DNA序列,或用外源性核酸序列替换大内源性基因组DNA序列的高效靶向技术。

[0006] 本领域中需要允许在大鼠的基因组中实现精确靶向改变的组合物和方法,此可展开或扩展当前靶标发现领域,并且更快速和容易地验证治疗剂。

[0007] 概述

[0008] 提供用于通过靶向遗传修饰来修饰多能细胞中的目标基因组基因座的方法。所述方法包括(a)向多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大靶向载体(LTVEC);以及(b)鉴定在目标基因组基因座处包含靶向遗传修饰的遗传修饰的多能细胞,其中靶向遗传修饰能够通过种系传递。

[0009] 在一个实施方案中,多能细胞源于非人动物(包括但不限于啮齿动物)、人、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、鸡、猫、狗、雪貂、灵长类动物(例如狨猿、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。

[0010] 在一个实施方案中,多能细胞是非人多能细胞。在一个实施方案中,非人多能细胞是哺乳动物多能细胞。在一个实施方案中,哺乳动物多能细胞是啮齿动物多能细胞。在一个实施方案中,啮齿动物多能细胞是大鼠或小鼠多能细胞。在一个实施方案中,多能细胞是人诱导多能干(iPS)细胞。

[0011] 在一个实施方案中,多能细胞是处于单细胞阶段的非人受精卵。在一个实施方案中,非人受精卵是哺乳动物受精卵。在一个实施方案中,哺乳动物受精卵是处于单细胞阶段的啮齿动物受精卵。在一个实施方案中,哺乳动物受精卵是处于单细胞阶段的大鼠或小鼠

受精卵。

[0012] 在一些实施方案中，LTVEC的5'同源臂和3'同源臂的总和是至少10kb。在一些实施方案中，LTVEC的5'同源臂和3'同源臂的总和是至少10kb但小于100kb，或LTVEC的5'同源臂和3'同源臂的总和是至少10kb但小于150kb。在其它实施方案中，LTVEC的5'同源臂和3'同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、或约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中，缺失的大小与LTVEC的5'同源臂和3'同源臂的总和大小相同或类似。

[0013] 在一些所述实施方案中，靶向遗传修饰是双等位基因修饰。

[0014] 在一些实施方案中，多能细胞是多能大鼠细胞。在一个实施方案中，多能大鼠细胞是大鼠胚胎干细胞。在一个实施方案中，多能大鼠细胞源于DA品系或ACI品系。在一些实施方案中，多能大鼠细胞的特征在于表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1或其组合。在一些所述方法中，多能大鼠细胞的特征在于以下特征中的一个或多个：(a)缺乏一种或多种包括c-Myc、Ecat1和/或Rexo1的多能性标志物的表达；(b)缺乏包括Brachyury和/或Bmpr2的中胚层标志物的表达；(c)缺乏一种或多种包括Gata6、Sox17和/或Sox7的内胚层标志物的表达；或(d)缺乏一种或多种包括巢蛋白(Nestin)和/或Pax6的神经标志物的表达。所述方法规定LTVEC的5'同源臂和3'同源臂的总和是约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、或约80kb至约100kb、约100kb至约120kb、约120kb至约150kb、或约10kb但小于约150kb。在一些实施方案中，LTVEC的5'同源臂和3'同源臂的总和是约16Kb至约100Kb。在其它实施方案中，LTVEC的5'同源臂和3'同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中，缺失的大小与LTVEC的5'同源臂和3'同源臂的总和大小相同或类似。

[0015] 方法进一步规定靶向遗传修饰(a)包括用同源或直系同源哺乳动物核酸序列替换内源性大鼠核酸序列；(b)包括缺失内源性大鼠核酸序列；(c)包括缺失内源性大鼠核酸序列，其中所述缺失在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb的范围内；(d)包括在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb的范围内的外源性核酸序列；(e)包括包含同源或直系同源核酸序列的外源性核酸序列；(f)包括包含人和大鼠核酸序列的嵌合核酸序列；(g)在约5kb

至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb的范围内；(h)包括侧接有位点特异性重组酶靶序列的条件性等位基因；或(i)包括可操作地连接至在大鼠细胞中具有活性的启动子的报告基因。

[0016] 进一步提供一种用于通过靶向遗传修饰来修饰多能大鼠细胞中的目标基因组基因座的方法，其中所述目标基因组基因座包含(i)与5'大鼠同源臂互补的第一核酸序列；和(ii)与3'大鼠同源臂互补的第二核酸序列。在一些所述实施方案中，第一核酸序列和第二核酸序列相隔至少5kb。在一些实施方案中，第一核酸序列和第二核酸序列相隔至少5kb但小于3Mb。在一些所述方法中，第一核酸序列和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、至少约2.5Mb但小于约3Mb、至少约1Mb但小于约2Mb、至少约2Mb但小于约3Mb。

[0017] 在一些实施方案中，引入步骤进一步包括引入编码促进靶向构建体与多能大鼠细胞中的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。在一些所述实施方案中，核酸酶试剂包括(a)包含融合于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)包含融合于FokI核酸内切酶的转录激活因子样效应物核酸酶(TALEN)的嵌合蛋白。

[0018] 在一些方法中，引入步骤进一步包括向多能大鼠细胞中引入：(i)包含可操作地连接至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)包含可操作地连接至编码基因组靶序列的第二核酸序列的第二启动子的第二表达构建体，所述基因组靶序列可操作地连接至导向RNA(gRNA)，其中所述基因组靶序列通过前间区序列邻近基序(PAM)序列以紧邻方式侧接在3'末端上。在一个实施方案中，目标基因组基因座包含核苷酸序列SEQ ID NO:1。在一个实施方案中，gRNA包含编码成簇规律间隔短回文重复序列(CRISPR)RNA(crRNA)和反式活化CRISPR RNA(tracrRNA)的第三核酸序列。在另一实施方案中，多能大鼠细胞的基因组包含与基因组靶序列互补的靶DNA区域。在一些所述方法中，Cas蛋白是Cas9。在一些所述方法中，gRNA包括(a)具有核酸序列SEQ ID NO:2的嵌合RNA；或(b)具有核酸序列SEQ ID NO:3的嵌合RNA。在一些所述方法中，crRNA包含SEQ ID NO:4、SEQ ID NO:5或SEQ ID NO:6中所示的序列。在一些所述方法中，tracrRNA包含SEQ ID NO:7或SEQ ID NO:8中所示的序列。

[0019] 进一步提供一种大鼠基因组基因座，其包含(i)同源或直系同源人核酸序列的插入；(ii)内源性大鼠核酸序列被同源或直系同源人核酸序列的替换；或(iii)其组合，其中所述大鼠基因组基因座能够通过种系传递。在某一所述大鼠基因组基因座中，插入或替换的大小是约5kb至约400kb。在某一所述大鼠基因组基因座中，插入或替换的大小是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约800kb、约800kb至1Mb、约1Mb至约1.5Mb、

约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约2.8Mb、约2.8Mb至约3Mb、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11Mb但小于约2Mb、至少约2Mb但小于约3Mb。

[0020] 进一步提供一种用于制备人源化大鼠的方法，其包括：(a)用包含人插入核酸的靶向构建体靶向多能大鼠细胞中的目标基因组基因座以形成经遗传修饰的多能大鼠细胞；(b)向宿主大鼠胚胎中引入所述经遗传修饰的多能大鼠细胞；以及(c)在代孕母体中孕育所述宿主大鼠胚胎；其中所述代孕母体产生包含经修饰的基因组基因座的大鼠子代，所述经修饰的基因组基因座包含：(i)人核酸序列的插入；(ii)在所述目标基因组基因座处的大鼠核酸序列被同源或直系同源人核酸序列的替换；(iii)包含人核酸序列和大鼠核酸序列的嵌合核酸序列；或(iv)其组合，其中所述经修饰的基因组基因座能够通过种系传递。

[0021] 在一些所述方法中，靶向构建体是大靶向载体(LTVEC)，并且LTVEC的5'同源臂和3'同源臂的总和是至少10kb但小于100kb，或LTVEC的5'同源臂和3'同源臂的总和是至少10kb但小于150kb。在一些所述方法中，靶向构建体的5'同源臂和3'同源臂的总和是约10kb至约30kb、约20kb至40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一些所述方法中，人核酸序列是至少5kb但小于400kb。在一些所述方法中，人核酸序列是至少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。在其它实施方案中，LTVEC的5'同源臂和3'同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中，缺失的大小与LTVEC的5'同源臂和3'同源臂的总和大小相同或类似。

[0022] 在一些用于制备人源化大鼠的方法中，多能大鼠细胞是大鼠胚胎干(ES)细胞。在一些所述方法中，多能大鼠细胞源于DA品系或ACI品系。在一些所述方法中，多能大鼠细胞的特征在于表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1和/或其组合。在一些所述方法中，多能大鼠细胞的特征在于以下特征中的一个或多个：(a)缺乏一种或多种包括c-Myc、Ecat1和/或Rexo1的多能性标志物的表达；(b)缺乏一种或多种包括Brachyury和/或Bmpr2的中胚层标志物的表达；(c)缺乏一种或多种包括Gata6、Sox17和/或Sox7的内胚层标志物的表达；或(d)缺乏一种或多种包括巢蛋白和/或Pax6的神经标志物的表达。

[0023] 进一步提供一种包含人源化基因组基因座的经遗传修饰的大鼠，其中所述经遗传修饰的大鼠包含：(i)同源或直系同源人核酸序列的插入；(ii)在内源性基因组基因座处大鼠核酸序列被同源或直系同源人核酸序列的替换；(iii)包含人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通过种系传递。在一些所述经遗传修饰的大鼠中，人源化基因组基因座包含嵌合核酸序列，其包含人核酸序列和大鼠核酸序列。

[0024] 也提供用于通过细菌同源重组(BHR)来修饰大鼠的靶基因组基因座的方法，并且包括：向原核细胞中引入包含侧接有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)，其中所述原核细胞包含大鼠核酸，并且能够表达介导在靶基因座处的BHR的重组酶，并且其中LTVEC的5’同源臂和3’同源臂的总和是至少10kb但小于100kb，或LTVEC的5’同源臂和3’同源臂的总和是至少10kb但小于150kb。在其它实施方案中，LTVEC的5’同源臂和3’同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、或约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中，缺失的大小与LTVEC的5’同源臂和3’同源臂的总和大小相同或类似。

[0025] 在一些所述方法中，大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。在一些所述方法中，第一核酸序列和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11Mb但小于约2Mb、至少约2Mb但小于约3Mb。

[0026] 在一些所述方法中，向原核细胞中引入靶向载体导致：(i)内源性大鼠核酸序列从靶基因组基因座缺失；(ii)外源性核酸序列添加在靶基因组基因座处；(iii)在靶基因座处内源性大鼠核酸序列被外源性核酸序列替换；或(iv)其组合。在一些所述方法中，插入核酸包含(a)与靶基因组基因座处的大鼠核酸序列同源或直系同源的多核苷酸；或(b)侧接有位点特异性重组识别序列的条件性等位基因。

[0027] 进一步提供一种包含靶向载体的宿主原核细胞，所述靶向载体包含侧接有5’大鼠同源臂和3’大鼠同源臂的插入核酸，其中所述插入核酸在约5k至约400kb的范围内。在一些宿主原核细胞中，插入核酸的大小是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或350kb至约400kb。在一些宿主原核细胞中，原核细胞包含可操作地连接至组成型活性启动子或诱导型启动子的重组酶基因。

[0028] 也提供用于通过靶向遗传修饰来修饰细胞中的目标基因组基因座的方法，其包括向所述细胞中引入

[0029] (a)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LTVEC的所述5’同源臂和所述3’同源臂的总和是至少10kb；和

[0030] (b)(i)包含可操作地连接至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)包含可操作地连接至编码基因组靶序列的第二核酸序列的第二启动子的第二表达构建体，所述基因组靶序列可操作地连接至导向RNA(gRNA)；以及鉴定在目标基因组基因座处包含靶向遗传修饰的遗传修饰的多能细胞。

[0031] 在一个实施方案中，目标基因组基因座包含SEQ ID NO:1中所示的核苷酸序列，其

中gRNA包含编码成簇规律间隔短回文重复序列(CRISPR)RNA(crRNA)和反式活化CRISPR RNA(tracrRNA)的第三核酸序列,并且其中细胞的基因组包含与基因组靶序列互补的靶DNA区域。在一些所述方法中,Cas蛋白是Cas9。在所述方法中,细胞可为多能细胞(如胚胎干细胞)或原核细胞。在一个实施方案中,多能细胞来自非人动物、非人哺乳动物、啮齿动物、人、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、鸡、猫、狗、雪貂、灵长类动物(例如狨猿、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。在另一实施方案中,原核细胞来自细菌,如大肠杆菌(E.coli)。

[0032] 在其它实施方案中,LTVEC的5'同源臂和3'同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、或约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中,缺失的大小与LTVEC的5'同源臂和3'同源臂的总和大小相同或类似。

[0033] 在一个实施方案中,多能细胞是非人多能细胞。在一个实施方案中,非人多能细胞是哺乳动物多能细胞。在一个实施方案中,哺乳动物多能细胞是啮齿动物多能细胞。在一个实施方案中,啮齿动物多能细胞是大鼠或小鼠多能细胞。在一个实施方案中,多能细胞是人诱导多能干(iPS)细胞。

[0034] 在一个实施方案中,多能细胞是处于单细胞阶段的非人受精卵。在一个实施方案中,非人受精卵是哺乳动物受精卵。在一个实施方案中,哺乳动物受精卵是处于单细胞阶段的啮齿动物受精卵。在一个实施方案中,哺乳动物受精卵是处于单细胞阶段的大鼠或小鼠受精卵。

[0035] 进一步提供一种在它的基因组基因座中包含靶向遗传修饰的大鼠或大鼠细胞,其中所述基因组基因座是白介素-2受体γ基因座、ApoE基因座、Rag1基因座、Rag2基因座或Rag2/Rag1基因座,其中所述靶向遗传修饰包括:(a)缺失所述基因组基因座处的内源性大鼠核酸序列;(b)插入同源核酸、直系同源核酸或包含人核酸序列和大鼠核酸序列的嵌合核酸;或(c)其组合。在所述大鼠或大鼠细胞中,靶向遗传修饰可通过所述大鼠或从所述大鼠细胞繁殖的大鼠的种系传递。

[0036] 在一些所述大鼠或大鼠细胞中,在基因组基因座处的内源性大鼠核酸的缺失是至少约10kb,或在基因组基因座处的外源性核酸序列的插入是至少约5kb。

[0037] 进一步提供一种大鼠或大鼠细胞,其中(a)在白介素-2受体γ基因座处的靶向遗传修饰导致白介素-2受体γ蛋白活性降低或不存在;(b)在ApoE基因座处的靶向遗传修饰导致ApoE蛋白活性降低或不存在;(c)在Rag1基因座处的靶向遗传修饰导致Rag1蛋白活性降低或不存在;(d)在Rag2基因座处的靶向遗传修饰导致Rag2蛋白活性降低或不存在;或(e)在Rag2/Rag1基因座处的靶向遗传修饰导致Rag2蛋白活性和Rag1活性降低或不存在。

[0038] 在一些实施方案中,白介素-2受体γ基因座的靶向遗传修饰包括:(a)缺失整个大鼠白介素-2受体γ编码区或其一部分;(b)用人白介素-2受体γ编码区或其一部分替换整个大鼠白介素-2受体γ编码区或其一部分;(c)用人白介素-2受体γ的外结构域替换大鼠白介素-2受体γ编码区的外结构域;或(d)缺失白介素-2受体γ基因座的至少3kb。在其它

所述大鼠或大鼠细胞中，ApoE基因座的靶向遗传修饰包括：(a)缺失整个ApoE编码区或其一部分；或(b)缺失包含ApoE编码区的ApoE基因座的至少1.8kb。

[0039] 进一步提供一种大鼠或大鼠细胞，其中Rag2基因座的靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分；或(b)缺失包含Rag2编码区的Rag2基因座的至少5.7kb。在一些实施方案中，Rag2/Rag1基因座的靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分以及缺失整个Rag1编码区或其部分；或(b)缺失包含Rag2编码区的Rag2/Rag1基因座的至少16kb。

[0040] 进一步提供一种大鼠或大鼠细胞，其中靶向遗传修饰包括在白介素-2受体γ基因座、ApoE基因座、Rag1基因座、Rag2基因座或Rag2/Rag1基因座处插入包含选择性标志物的表达盒。在一些所述大鼠或大鼠细胞中，表达盒包含可操作地连接至基因组基因座处的内源性启动子的lacZ基因和可操作地连接至选择性标志物的人泛素启动子。

[0041] 进一步提供一种大鼠或大鼠细胞，其中白介素-2受体γ基因座、ApoE基因座、Rag1基因座、Rag2基因座或Rag2/Rag1基因座中的靶向遗传修饰包括插入自我缺失性选择盒。在一些所述大鼠或大鼠细胞中，自我缺失性选择盒包含可操作地连接至在大鼠细胞中具有活性的启动子的选择性标志物基因和可操作地连接至雄性生殖细胞特异性启动子的重组酶基因，其中自我缺失性盒由被重组酶识别的重组识别位点侧接。在一些所述大鼠或大鼠细胞中，雄性生殖细胞特异性启动子是鱼精蛋白(Protamine)-1启动子；重组酶基因编码Cre，并且重组识别位点是loxP位点。在一个实施方案中，鱼精蛋白-1启动子是小鼠或大鼠鱼精蛋白-1启动子。

[0042] 进一步提供一种大鼠或大鼠细胞，其中在基因组基因座处的外源性核酸序列的插入包含可操作地连接至内源性白介素-2受体γ启动子、内源性ApoE启动子、内源性Rag1启动子或内源性Rag2启动子的报告核酸。在一些所述大鼠或大鼠细胞中，报告核酸编码包括以下的报告蛋白：β-半乳糖苷酶、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强黄色荧光蛋白(EYFP)、Emerald、增强绿色荧光蛋白(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。

[0043] 进一步提供一种大鼠细胞，其中所述大鼠细胞是多能大鼠细胞或大鼠胚胎干(ES)细胞。在一些所述大鼠细胞中，多能大鼠细胞或大鼠胚胎干(ES)细胞(a)源于DA品系或ACI品系；(b)特征在于表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1或其组合；或(c)特征在于以下特征中的一个或多个：(i)缺乏一种或多种包括c-Myc、Ecat1和Rexo1的多能性标志物的表达；(ii)缺乏包括Brachyury和Bmpr2的中胚层标志物的表达；(iii)缺乏一种或多种包括Gata6、Sox17和Sox7的内胚层标志物的表达；或(iv)缺乏一种或多种包括巢蛋白和Pax6的神经标志物的表达。

[0044] 进一步提供一种用于修饰多能大鼠细胞中白介素-2受体γ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'和3'大鼠同源臂的插入核酸的靶向载体；以及(b)鉴定在所述靶基因组基因座处包含靶向遗传修饰的经遗传修饰的多能大鼠细胞，其中所述靶向遗传修饰能够通过从所述多能大鼠细胞繁殖的大

鼠的种系传递。在一些所述方法中，靶向载体是大靶向载体(LTVEC)，其中5’大鼠同源臂和3’大鼠同源臂的总和是至少约10kb。在一些实施方案中，5’大鼠同源臂和3’大鼠同源臂的总和是至少10kb但小于150kb。在一些实施方案中，5’大鼠同源臂和3’大鼠同源臂的总和是至少约10kb但小于约100kb。在一些实施方案中，向多能大鼠细胞中引入靶向载体导致：(i)靶基因组基因座处的内源性大鼠核酸序列缺失；(ii)外源性核酸序列插入在靶基因组基因座处；或(iii)其组合。

[0045] 在一些实施方案中，在基因组基因座处的内源性大鼠核酸的缺失是至少约10kb；在基因组基因座处的内源性大鼠核酸序列的缺失在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb的范围内；在基因组基因座处的外源性核酸序列的插入是至少约5kb；或外源性核酸序列的插入在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb的范围内。

[0046] 在一些实施方案中，(a)在白介素-2受体γ基因座处的靶向遗传修饰导致白介素-2受体γ蛋白活性降低或不存在；(b)在ApoE基因座处的靶向遗传修饰导致ApoE蛋白活性降低或不存在；(c)在Rag1基因座处的靶向遗传修饰导致Rag1蛋白活性降低或不存在；(d)在Rag2基因座处的靶向遗传修饰导致Rag2蛋白活性降低或不存在；或(e)在Rag2/Rag1基因座处的靶向遗传修饰导致Rag2蛋白活性和Rag1蛋白活性降低或不存在。

[0047] 在一些实施方案中，在白介素-2受体γ基因座处的靶向遗传修饰包括(a)缺失整个大鼠白介素-2受体γ编码区或其一部分；(b)用人白介素-2受体γ编码区或其一部分替换整个大鼠白介素-2受体γ编码区或其一部分；(c)用人白介素-2受体γ的外结构域替换大鼠白介素-2受体γ编码区的外结构域；或(d)缺失包含白介素-2受体γ编码区的白介素-2受体γ基因座的至少3kb。

[0048] 在一些实施方案中，在ApoE基因座处的靶向遗传修饰包括：(a)缺失整个ApoE编码区或其一部分；或(b)缺失包含ApoE编码区的ApoE基因座的至少1.8kb。

[0049] 在一些实施方案中，在Rag2基因座处的靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分；或(b)缺失包含Rag2编码区的Rag2基因座的至少5.7kb。在其它方法中，Rag1/Rag2基因座的靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分以及缺失整个Rag1编码区或其部分；或(b)缺失包含Rag2和Rag1编码区的Rag2/Rag1基因座的至少16kb。

[0050] 在一些用于修饰靶基因组基因座的所述实施方案中，插入核酸包含表达盒，其包含编码选择性标志物的多核苷酸。在一些所述实施方案中，表达盒包含可操作地连接至基因组基因座处的内源性启动子的lacZ基因和可操作地连接至选择性标志物基因的人泛素启动子。

[0051] 在一些实施方案中，插入核酸包含自我缺失性选择盒。在一些所述实施方案中，自我缺失性选择盒包含可操作地连接至在大鼠多能细胞中具有活性的启动子的选择性标志物和可操作地连接至雄性生殖细胞特异性启动子的编码重组酶的多核苷酸，其中自我缺失性盒由被重组酶识别的重组识别位点侧接。在一些所述实施方案中，雄性生殖细胞特异性

启动子是鱼精蛋白-1启动子;或重组酶基因编码Cre,并且重组识别位点是loxP位点。在一些实施方案中,鱼精蛋白-1启动子是小鼠或大鼠鱼精蛋白-1启动子。

[0052] 在其它方法中,在基因组基因座处的外源性核酸序列的插入包含可操作地连接至内源性白介素-2受体γ启动子、内源性ApoE启动子、内源性Rag1启动子或内源性Rag2启动子的报告核酸序列。在一些所述实施方案中,报告核酸序列编码包括以下的报告蛋白:β-半乳糖苷酶、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强黄色荧光蛋白(EYFP)、Emerald、增强绿色荧光蛋白(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。

[0053] 在一些用于修饰靶基因组基因座的实施方案中,多能大鼠细胞是大鼠胚胎干(ES)细胞。在一些所述实施方案中,多能大鼠细胞(a)源于DA品系或ACI品系;(b)特征在于表达包括Oct-4、Sox-2、碱性磷酸酶或其组合的多能性标志物;或(c)特征在于以下特征中的一个或多个:(i)缺乏一种或多种包括c-Myc、Ecat1和Rexo1的多能性标志物的表达;(ii)缺乏包括Brachyury和Bmpr2的中胚层标志物的表达;(iii)缺乏一种或多种包括Gata6、Sox17和Sox7的内胚层标志物的表达;或(iv)缺乏一种或多种包括巢蛋白和Pax6的神经标志物的表达。

[0054] 在一些实施方案中,方法进一步包括鉴定靶基因组基因座处的靶向遗传修饰,其中鉴定步骤采用用于评估靶基因组基因座处的等位基因修饰(MOA)的定量测定。

[0055] 在一些实施方案中,引入步骤进一步包括引入编码促进靶向载体与多能大鼠细胞中的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。在一些所述实施方案中,核酸酶试剂包括包含融合于FokI核酸内切酶的基于锌指的DNA结合结构域的嵌合蛋白。一些所述方法导致靶基因组基因座的双等位基因修饰。

[0056] 在一些实施方案中,方法的引入步骤进一步包括向多能大鼠细胞中引入:包含可操作地连接至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,和包含可操作地连接至编码基因组靶序列的第二核酸序列的第二启动子的第二表达构建体,所述基因组靶序列可操作地连接至导向RNA(gRNA),其中所述基因组靶序列通过前间区序列邻近基序(PAM)序列以紧邻方式侧接在3'末端上。在一个实施方案中,基因组靶序列包含SEQ ID NO:1中所示的核苷酸序列。在一个实施方案中,gRNA包含编码成簇规律间隔短回文重复序列(CRISPR)RNA(crRNA)和反式活化CRISPR RNA(tacrRNA)的第三核酸序列。在一些所述实施方案中,Cas蛋白是Cas9。在一些所述实施方案中,(a)gRNA是具有SEQ ID NO:2中所示的核酸序列的嵌合RNA;(b)gRNA是具有SEQ ID NO:3中所示的核酸序列的嵌合RNA;(c)crRNA包含SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6中所示的序列;或(d)tracrRNA包含SEQ ID NO:7和/或SEQ ID NO:8中所示的序列。

[0057] 附图简述

[0058] 本专利或申请文件含有至少一个以彩色制作的附图。带有彩色附图的本专利或专利申请公布的复本将在请求以及支付必要费用后由专利局提供。

[0059] 图1描绘大鼠ESC,其以通常在培养皿中脱离和浮动的紧实球形集落形式生长。

[0060] 图2A至2D描绘由大鼠ESC表达的各种多能性标志物: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表达的多能性标志物的重叠。

[0061] 图3描绘大鼠ESC表达各亮度水平的碱性磷酸酶(多能性标志物)(左侧),并且株系

DA.2B的核型是42X,Y(右侧)。因为大鼠ESC常常变为四倍体,所以进行核型分析;因此,通过计数中期染色体散布来预先筛选株系;并且接着正式分析具有大部分正常计数的株系的核型。

[0062] 图4A-B提供显示对ACI.G1大鼠ES细胞系的染色体数目的分析的照片。

[0063] 图5A-B提供显示对DA.2B大鼠ES细胞系的染色体数目的分析的照片。

[0064] 图6A-B提供显示对DA.C2大鼠ES细胞系的染色体数目的分析的照片。

[0065] 图7描绘图1的大鼠ESC的近视图。

[0066] 图8描绘通过胚泡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,通过胚泡注射来产生嵌合体;高百分比嵌合体通常具有白化鼻部。

[0067] 图9描绘由在图8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。

[0068] 图10提供大鼠ApoE基因座的示意图,并且用灰条表示锌指核酸酶(ZFN1和ZFN2)的切割位点。对应于5'同源臂和3'同源臂的基因组区域(分别是5kb和5.4kb)通过暗灰色框来表示。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基因的三个内含子表示为直线。外显子2和3包含编码区,并且显示为点描灰色框。外显子4含有编码序列与非编码序列两者,如通过点描灰色阴影和空心框所表示。

[0069] 图11提供在锌指核酸酶(ZFN1或ZFN2)存在下进行时的ApoE靶向效率的概述。

[0070] 图12描绘大鼠Rosa 26基因座的靶向,如同小鼠中一样,所述基因座以相同间隔位于Setd5与Thumpd3基因之间。图A显示小鼠Rosa 26基因座的结构。小鼠Rosa26转录物由2或3个外显子组成。图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。图C描绘靶向大鼠Rosa26等位基因;通过使用来自DA rESC的基因组DNA进行PCR来克隆各自具有5kb的同源臂;靶向等位基因含有替换大鼠Rosa26内含子中的117bp缺失的剪接接受体(SA)-lacZ-hUB-neo盒。

[0071] 图13A描绘14周龄野生型大鼠的对照脑,其用X-gal染色。对照脑显示低水平的背景LacZ染色(背视图)。

[0072] 图13B描绘rRosa26杂合性大鼠(14周龄)的脑中的LacZ表达。lacZ报告蛋白在rRosa26杂合子的整个脑中遍在表达。

[0073] 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。对照心脏和胸腺显示低水平的背景LacZ染色。

[0074] 图13D描绘14周龄rRosa26杂合性大鼠的心脏和胸腺(插图)中的LacZ表达。lacZ报告蛋白在rROSA26杂合子的整个心脏和胸腺中遍在表达。

[0075] 图13E描绘14周龄野生型大鼠的对照肺,其用X-gal处理。对照肺显示低水平的背景LacZ染色。

[0076] 图13F描绘14周龄rRosa26杂合子大鼠的肺中的LacZ表达。lacZ报告蛋白在rRosa26杂合子的整个肺中遍在表达。

[0077] 图13G和13H描绘E12.5大鼠胚胎中的LacZ表达。不同于显示低水平的背景LacZ染色的野生型对照胚胎(H),rRosa26杂合性胚胎展现LacZ报告蛋白在整个胚胎中遍在表达。

[0078] 图13I和13J描绘E14.5大鼠胚胎中的LacZ表达。不同于显示低水平的背景LacZ染

色的野生型对照胚胎(J),rRosa26杂合性大鼠胚胎展现LacZ报告蛋白在整个胚胎中遍在表达。

[0079] 图14说明在包含选择盒(lacZ-neo盒)的靶向载体的电穿孔之后发生在大鼠ES细胞内部的同源或非同源重组事件。

[0080] 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)以及活化ES细胞中的非同源末端接合(NHEJ)所依的机理。

[0081] 图16说明利用ZFN/TALEN来提高靶向载体的同源重组效率的基因靶向技术。DSB代表双链断裂。

[0082] 图17提供经修饰的大鼠ApoE基因座的嵌合体产生和种系传递的概述。通过锌指核酸酶来辅助靶向修饰。

[0083] 图18提供与靶向ZFN U和ZFN D的锌指核酸酶组合的IL2r-γ靶向事件的示意图。ZFN切割位点指示在图中。

[0084] 图19提供当与CRISPR/Cas9系统组合靶向IL2r-γ时的靶向效率。

[0085] 图20提供大鼠ApoE基因座和靶向质粒的示意图。上部示意图显示大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区域(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基因的三个内含子表示为直线。外显子2和3包含编码区,并且显示为点描灰色框。外显子4含有编码序列与非编码序列两者,如通过点描灰色阴影和空心框所表示。下部图显示靶向质粒。5'同源臂和3'同源臂(分别是5kb和5.4kb)通过暗灰色框来表示。靶向载体包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒。自我缺失性盒包含可操作地连接至Crei基因的小鼠Prm1启动子和包含可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。

[0086] 图21提供使用锌指核酸酶和靶向载体来靶向大鼠ES细胞中的ApoE基因座的示意图,所述靶向载体包含报告基因(LacZ)和自我缺失性盒,所述自我缺失性盒包含可操作地连接至Crei基因的小鼠Prm1启动子和包含可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。

[0087] 图22提供大鼠ApoE基因座和大靶向载体(LTVEC)的示意图。上部图显示大鼠ApoE基因座和对应于5'同源臂和3'同源臂的基因组区域(分别是45kb和23kb;暗灰色框)的基因组组构。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基因的三个内含子被表示为直线,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含有编码序列与非编码序列两者,如通过点描灰色阴影和空心框所表示。下部图显示用于修饰大鼠ApoE基因座的LTVEC。5'同源臂和3'同源臂(分别是45kb和23kb)通过暗灰色框来表示。LTVEC包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒,所述自我缺失性盒包含可操作地连接至Crei基因的小鼠Prm1启动子和包含可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。

[0088] 图23提供大鼠ApoE基因座的示意图,并且用灰条表示连同大靶向载体(LTVEC)一起用于增强靶向载体与靶同源染色体区域之间的同源重组的锌指核酸酶(ZFN1和ZFN2)的切割位点。

[0089] 图24描绘已通过3.2kb缺失以及插入报告基因(eGFP)和自我缺失性盒来破坏的大

鼠IL2r-γ基因座，所述自我缺失性盒包含药物选择盒(hUb-neo)和可操作地连接至小鼠Prm1启动子的Crei基因。

[0090] 图25提供种系传递性可靶向大鼠胚胎干细胞系的概述。

[0091] 图26提供对大鼠IL2r-γ基因座的另一描绘，所述基因座已通过3.2kb缺失以及插入报告基因(eGFP)和自我缺失性盒来破坏，所述自我缺失性盒包含可操作地连接至小鼠Prm1启动子的Crei基因和药物选择盒(hUb-Neo)。

[0092] 图27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。上部图显示大鼠Rag2基因座和对应于5'同源臂和3'同源臂的同源基因组区域(分别是48kb和15kb;暗灰色框)的基因组组构。Rag2包含通过点描灰色阴影来表示的单一外显子。下部图是LTVEC。5'同源臂和3'同源臂(分别是48kb和15kb)通过暗灰色框来表示。LTVEC包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒，所述自我缺失性盒含有可操作地连接至Crei基因的大鼠Prm1启动子和含有可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。

[0093] 图28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)缺失的基因组区域。

[0094] 图29提供大鼠Rag2和Rag1基因座以及用于修饰基因座的大靶向载体(LTVEC)的示意图。上部图显示Rag1和Rag2基因座以及对应于5'同源臂和3'同源臂的同源基因组区域(分别是48kb和84kb;暗灰色框)的基因组组构。Rag2和Rag1各自包含通过点描灰色阴影来表示的单一外显子。下部图是LTVEC。5'同源臂和3'同源臂(分别是48kb和84kb)通过暗灰色框来表示。LTVEC包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒，所述自我缺失性盒包含可操作地连接至Crei基因的大鼠Prm1启动子和包含可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。

[0095] 图30显示II2rg-/y PBMC不表达成熟淋巴细胞标志物。在3个嵌合体中的2个中的外周血液中检测到GFP阳性淋巴细胞。

[0096] 图31提供大鼠IL-2rg基因座以及用于使大鼠IL-2rg基因座完全人源化的靶向质粒的示意图。上部图显示大鼠IL-2rg基因座以及对应于5'同源臂和3'同源臂的同源基因组区域(分别是4.4kb和5.0kb;暗灰色框)的基因组组构。下部图是靶向质粒。5'同源臂和3'同源臂(分别是4.4kb和5.0kb)通过暗灰色框来表示。靶向质粒包含人IL-2rg基因组区域、报告基因(GFP)和由loxP位点(空心箭头)侧接的自我缺失性盒，所述自我缺失性盒含有可操作地连接至Crei基因的小鼠Prm1启动子和含有可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。

[0097] 图32提供大鼠IL-2rg基因座以及用于达成大鼠IL-2rg基因座的外结构域人源化的靶向质粒的示意图。上部图显示大鼠IL-2rg基因座以及对应于5'同源臂和3'同源臂的同源基因组区域(分别是4.4kb和5.0kb;暗灰色框)的基因组组构。下部图是靶向质粒。5'同源臂和3'同源臂(分别是4.4kb和5.0kb)通过暗灰色框来表示。靶向质粒包含IL-2Rg基因组区域的人外结构域、报告基因(GFP)和由loxP位点(空心箭头)侧接的自我缺失性盒，所述自我缺失性盒含有可操作地连接至Crei基因的小鼠Prm1启动子和含有可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。

[0098] 图33提供人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的人外结构域的嵌合IL-2rg蛋白(SEQ ID NO:22)的序列比对。人IL-2rg与大鼠IL-2rg之间的接合通过垂直线来指示。

[0099] 发明详述

[0100] 术语

[0101] 如本文所用的术语“胚胎干细胞”或“ES细胞”包括在向胚胎中引入后能够促成发育胚胎的任何组织的胚胎源性全能或多能细胞。如本文所用的术语“多能细胞”包括具有发育成超过一种分化细胞类型的能力的未分化细胞。

[0102] 如本文所用的术语“同源核酸”包括与已知参照序列同一或大致上类似的核酸序列。在一个实施方案中,术语“同源核酸”用于表征与已知参照序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或甚至100%同一性的氨基酸序列的序列。

[0103] 如本文所用的术语“直系同源核酸”包括来自一个物种的在功能上等效于另一物种中的已知参照序列的核酸序列。

[0104] 如本文所用的术语“大靶向载体”或“LTVEC”包括源于克隆的基因组DNA的片段的用于真核细胞的大靶向载体,所述片段大于通常由意图在真核细胞中进行同源基因靶向的其它方法使用的那些。LTVEC的实例包括但不限于细菌同源染色体(BAC)和酵母人工染色体(YAC)。

[0105] 如本文所用的术语“等位基因修饰”(MOA)包括对基因组中基因或染色体基因座的一个等位基因的精确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、取代或插入少至单一核苷酸,或缺失跨越目标基因或染色体基因座的数千碱基,以及这两种极端之间的任何和所有可能修饰。

[0106] 如本文所用的术语“重组位点”包括由位点特异性重组酶识别,并且可充当重组事件的底物的核苷酸序列。

[0107] “连续”遗传修饰包括对例如大鼠ES细胞的两次或更多次独立进行的修饰。举例来说,采用适合第一核酸构建体对大鼠ES细胞基因组进行第一修饰。第一修饰可通过电穿孔或本领域中已知的任何其它方法来实现。接着采用适合的第二核酸构建体对同一大鼠ES细胞基因组进行第二修饰。第二修饰可通过第二电穿孔或本领域中已知的任何其它方法来实现。在各种实施方案中,在对同一大鼠ES细胞的第一和第二遗传修饰之后,可使用例如连续电穿孔或本领域中已知的任何其它适合方法(连续)实现第三、第四、第五、第六等等连续遗传修饰(一个接着另一个)。

[0108] 如本文所用的术语“位点特异性重组酶”包括一组可促进“重组位点”之间的重组的酶,其中两个重组位点在单一核酸分子内或在单独核酸分子上被物理分隔。“位点特异性重组酶”的实例包括但不限于Cre、Flp和Dre重组酶。

[0109] 关于核酸序列的术语“种系”包括可传至子代的核酸序列。

[0110] 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。除非另外规定,否则重链可变结构域包括三个重链CDR和四个FR区。重链的片段包括CDR、CDR和FR、及其组合。在可变结构域之后(从N末端至C末端),典型重链具有C_H1结构域、铰链、C_H2结构域和C_H3结构域。重链的功能性片段包括能够特异性识

别表位(例如以在微体积摩尔浓度、纳体积摩尔浓度或皮体积摩尔浓度的范围内的K_D识别表位),能够由细胞表达并分泌,以及包含至少一个CDR的片段。重链可变结构域由可变区核苷酸序列编码,所述序列通常包含源于种系中存在的V_H、D_H和J_H区段的谱系的V_H、D_H和J_H区段。各种生物体的V、D和J重链区段的序列、位置和命名可见于IMGT数据库中,所述数据库可通过因特网在万维网(www)上以URL“imgt.org.”访问。

[0111] 短语“轻链”包括来自任何生物体的免疫球蛋白轻链序列,并且除非另外规定,否则包括人κ和λ轻链和VpreB以及替代轻链。除非另外规定,否则轻链可变结构域通常包括三个轻链CDR和四个框架(FR)区。通常,全长轻链从氨基末端至羧基末端包括可变结构域(其包括FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4)和轻链恒定区氨基酸序列。轻链可变结构域由轻链可变区核苷酸序列编码,所述序列通常包含源于种系中存在的轻链V和J基因区段的谱系的轻链V_L和轻链J_L基因区段。各种生物体的轻链V和J基因区段的序列、位置和命名可见于IMGT数据库中,所述数据库可通过因特网在万维网(www)上以URL“imgt.org.”访问。轻链包括例如不选择性结合由它们出现于其中的表位结合蛋白选择性结合的第一或第二表位的那些。轻链也包括结合以及识别一个或多个由它们出现于其中的表位结合蛋白选择性结合的表位,或协助重链结合以及识别所述表位的那些。

[0112] 短语“可操作地连接”包括其中可操作地连接的组分以它们的预定方式起作用的关系。在一种情况下,编码蛋白质的核酸序列可可操作地连接至调控序列(例如启动子、增强子、沉默子序列等)以便保持适当转录调控。在一种情况下,免疫球蛋白可变区(或V(D)J区段)的核酸序列可可操作地连接至免疫球蛋白恒定区的核酸序列以便允许在序列之间适当重组成免疫球蛋白重链或轻链序列。

[0113] 1.包含大鼠核酸的靶基因座

[0114] 提供允许在靶基因座处整合至少一种插入核酸的各种方法和组合物。如本文所用,“目标基因组基因座”包含基因组内需要整合插入核酸的任何DNA区段或区域。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。目标基因组基因座可为细胞天然的,或者可包含整合至细胞基因组中的异源性或外源性DNA区段。所述异源性或外源性DNA区段可包括转基因、表达盒、编码选择标志物的多核苷酸、或基因组DNA的异源性或外源性区域。术语“基因座”在本文中定义为基因组DNA内的DNA区段。如本文所述的遗传修饰可包括一次或多次从目标基因座进行缺失,向目标基因座进行添加,替换目标基因座和/或其任何组合。目标基因座可包含编码区或非编码调控区。

[0115] 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,包括例如识别位点、选择标志物、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。或者,目标基因组基因座可位于酵母人工染色体(YAC)、细菌人工染色体(BAC)、人人工染色体或适当宿主细胞中含有的任何其它工程化基因组区域内。在各种实施方案中,靶向基因座可包含来自原核生物、真核生物、酵母、细菌、非人哺乳动物、非人细胞、啮齿动物、人、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、鸡、猫、狗、雪貂、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、异源性或外源性核酸序列。

[0116] 在特定实施方案中,目标基因组基因座包括“大鼠核酸”的靶基因座。所述区域包含来自大鼠的整合在细胞基因组内的核酸。

[0117] 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

达不成熟B细胞中的多肽的基因组基因座、表达成熟B细胞中的多肽的基因组基因座、免疫球蛋白(Ig)基因座、或T细胞受体基因座，包括例如T细胞受体α基因座。靶基因组基因座的其它实例包括FcER1a基因座、TLR4基因座、PRLR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γ(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。任何所述靶基因座都可来自大鼠。

[0118] 在一个实施方案中，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，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，靶基因座包含基因组DNA序列，所述序列包含可操作地连接至免疫球蛋白重链恒定区核酸序列的未重排大鼠、小鼠或人免疫球蛋白重链可变区核酸序列。在一个实施方案中，免疫球蛋白重链恒定区核酸序列是选自CH1、铰链、CH2、CH3及其组合的大鼠、小鼠或人免疫球蛋白重链恒定区核酸序列。在一个实施方案中，重链恒定区核酸序列包括CH1-铰链-CH2-CH3。在一个实施方案中，靶基因座包含可操作地连接至免疫球蛋白重链恒定区核酸序列的重排大鼠、小鼠或人免疫球蛋白重链可变区核酸序列。在一个实施方案中，免疫球蛋白重链恒定区核酸序列是选自CH1、铰链、CH2、CH3及其组合的大鼠、小鼠或人免疫球蛋白重链恒定区核酸序列。在一个实施方案中，重链恒定区核酸序列包括CH1-铰链-CH2-CH3。

[0119] 在一个实施方案中，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，基因组DNA序列包含未重排哺乳动物λ和/或κ轻链可变区核酸序列。

[0120] 在一个实施方案中，基因组DNA序列包含重排哺乳动物λ和/或κ轻链可变区核酸序列。在一个实施方案中，未重排λ或κ轻链可变区核酸序列可操作地连接至选自λ轻链恒定区核酸序列和κ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，哺乳动物免疫球蛋白轻链恒定区核酸序列是大鼠免疫球蛋白轻链恒定区核酸序列。在一个实施方案中，哺乳动物免疫球蛋白轻链恒定区核酸序列是小鼠免疫球蛋白轻链恒定区核酸序列。在一个实施方案中，哺乳动物免疫球蛋白轻链恒定区核酸序列是人免疫球蛋白轻链恒定区核酸序列。

[0121] 如本文所用，大鼠ApoE基因座、大鼠白介素-2受体γ(Il-2rg)基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或大鼠Rag2/Rag1基因座包含大鼠基因组的其中定位有这些基因或基因组合中的各个的相应区域。修饰大鼠ApoE基因座、大鼠白介素-2受体γ基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或组合大鼠Rag2/Rag1基因座中的任一个都可包括对给定基因座进行任何所需改变。对给定大鼠基因座的修饰的非限制性实例在本文中进一步详细讨论。

[0122] 举例来说，在特定实施方案中，修饰大鼠ApoE基因座、大鼠白介素-2受体γ基因座、Rag2基因座和/或Rag2/Rag1基因座中的一个或多个以使编码的ApoE蛋白、或白介素-2受体γ蛋白、或Rag1蛋白、或Rag2蛋白、或Rag1蛋白和Rag2蛋白的组合的活性和/或水平降低。在其它实施方案中，不存在ApoE蛋白、白介素-2受体γ蛋白、Rag1蛋白、或Rag2蛋白、或Rag1蛋白和Rag2蛋白的组合的活性。

[0123] 就“降低”来说，其意指在目标基因座处的基因/编码的蛋白质的水平或活性的任何降低。举例来说，活性降低可包括(1)当相较于适当对照时，给定蛋白质(即ApoE、白介素-

2受体γ、Rag2、Rag2、或Rag1和Rag2的组合)的总体水平或活性统计显著降低,包括例如水平或活性降低0.5%、1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、120%或大于120%。用以测定ApoE、白介素-2受体γ、Rag1和Rag2中的任一个的浓度和/或活性降低的方法在本领域中是已知的。

[0124] 在其它实施方案中,大鼠ApoE基因座、大鼠白介素-2受体γ基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或大鼠Rag2/Rag1基因座中的一个或多个包含修饰以使编码的ApoE多肽、白介素-2受体γ多肽、Rag2多肽、Rag1多肽、或Rag1多肽与Rag2多肽两者的活性和/或水平增加。就"增加"来说,其意指在目标基因座处的基因/编码的多肽的水平或活性的任何增加。举例来说,活性增加可包括(1)当相较于适当对照时,给定蛋白质(即ApoE、白介素-2受体γ、Rag1、Rag2、或Rag1和Rag2)的总体水平或活性统计显著增加,包括例如水平或活性增加0.5%、1%、5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、100%、120%或大于120%。用以测定ApoE、Rag1、Rag2和白介素-2受体γ蛋白中的任一个的浓度和/或活性增加的方法在本领域中是已知的。

[0125] 对大鼠ApoE基因座、大鼠白介素-2受体γ基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或大鼠Rag2/Rag1基因座的遗传修饰可包括缺失基因组基因座处的内源性大鼠核酸序列,在基因组基因座处插入外源性核酸,或其组合。缺失和/或插入可发生在给定基因座内的任何地方,如本文其它地方所讨论。

[0126] 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、白介素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应同源或直系同源部分替换大鼠ApoE基因座、白介素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分来修饰大鼠ApoE基因座、大鼠白介素-2受体γ基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或大鼠Rag2/Rag1基因座中的一个或多个。

[0127] 其它实施方案,通过用插入多核苷酸替换大鼠ApoE基因座、大鼠白介素-2受体γ基因座和/或大鼠Rag2基因座和/或Rag1基因座和/或Rag2/Rag1基因座的一部分来对大鼠ApoE基因座、大鼠白介素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座中的一个或多个进行修饰,所述插入多核苷酸跨越它的全长与它替换的ApoE基因座、白介素-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%。

[0128] 给定插入多核苷酸和/或大鼠基因座的相应缺失区域可为编码区、内含子、外显子、非翻译区、调控区、启动子、或增强子或其任何组合或其任何部分。此外,给定插入多核苷酸和/或大鼠基因座的缺失区域可具有任何希望的长度,包括例如长度在10-100个核苷酸、100-500个核苷酸、500-1kb核苷酸、1Kb至1.5kb核苷酸、1.5kb至2kb核苷酸、2kb至2.5kb核苷酸、2.5kb至3kb核苷酸、3kb至5kb核苷酸、5kb至8kb核苷酸、8kb至10kb核苷酸之间或大于10kb核苷酸。在其它情况下,插入或替换的大小是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约800kb、约800kb至1Mb、约300kb至约400kb、约400kb至约500kb、约500kb至1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约2.8Mb、约2.8Mb至约3Mb。在其它实施方案中,给定插入多核苷酸和/或大鼠基因座的缺失区域是至少

100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或大于16kb。

[0129] 给定插入多核苷酸可来自任何生物体，包括例如啮齿动物、大鼠、小鼠、仓鼠、哺乳动物、非人哺乳动物、人、农业动物或家养动物。

[0130] 如本文进一步详细讨论，提供用以产生任何目标大鼠基因座的靶向修饰的各种方法，所述靶向修饰包括例如大鼠ApoE基因座、大鼠白介素-2受体γ基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或大鼠Rag2/Rag1基因座中的靶向修饰。进一步提供在白介素-2受体γ基因座处，在ApoE基因座处，在大鼠Rag2基因座处，在大鼠Rag1基因座处，和/或在大鼠Rag2/Rag1基因座处包含缺失、插入、替换和/或其任何组合的经遗传修饰的大鼠或经遗传修饰的多能大鼠细胞(例如大鼠ES细胞)。所述遗传修饰(包括导致靶基因座的活性不存在、降低、增加或受调节的那些)，并且也能够通过种系传递。在特定实施方案中，遗传修饰导致敲除所需靶基因座。所述大鼠适用于多种实验系统中，如本文其它地方所讨论。

[0131] 举例来说，在大鼠中敲除ApoE(载脂蛋白E(Apolipoprotein E))提供一种用以研究内皮功能(包括但不限于斑块形成)、转录变化(全转录物组鸟枪测序(RNA-Seq))和离体功能的动物模型。此外，大鼠的身材较大有助于所有这些测定，并且潜在改进RNA-Seq数据的质量。ApoE是一种重要转运分子，并且可通过血流转运脂质，如胆固醇。ApoE也可在神经系统中起作用，例如从脑清除β-淀粉状蛋白(amyloid)。ApoE中的修饰已牵涉于各种病状中，包括例如动脉粥样硬化、高脂质血症和阿尔茨海默氏病(Alzheimer's disease)。ApoE敲除动物显示脂蛋白从血液的清除受损，并且显现动脉粥样硬化。因此，ApoE敲除动物提供一种用以研究例如像内皮功能、斑块形成、转录变化(RNA-Seq)、高脂质血症、动脉粥样硬化和阿尔茨海默氏病的病状和/或过程的模型。用以测量ApoE活性的测定在本领域中是已知的。举例来说，ApoE活性降低可通过免疫测定，如通过ELISA或通过免疫印迹技术测定从受试者获得的血液样品中的ApoE水平降低来测量。此外，大鼠的身材较大有助于所有这些测定，并且改进数据的质量。

[0132] RAG1(重组活化基因1)和RAG2(重组活化基因2)是作为具有VDJ重组活性的多亚基复合物的一部分的酶，并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。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)基因的区段的重组和接合。敲除RAG1和/或RAG2导致动物中的B细胞和T细胞损失，从而导致重度免疫缺陷。RAG1和/或RAG2敲除动物适用于例如异种移植物(即大鼠中的人细胞异种移植物)、癌症、疫苗开发、自体免疫疾病、感染性疾病和移植物抗宿主疾病(GVHD)的研究中。用以测量RAG1和/或RAG2活性的各种测定在本领域中是已知的，并且包括例如测量重组效率或测定受试者中存在或不存在B细胞和/或T细胞。此外，大鼠的身材较大有助于所有这些测定，并且潜在改进数据的质量。

[0133] IL-2受体(IL-2R)在某些免疫细胞的表面上表达，并且结合细胞因子白介素-2(IL-2)。IL-2R是一种包含至少三个单独亚基链的整合膜蛋白，所述亚基链包括α链(IL-2Ra、CD25)、β链(IL-2Rb、CD122)和γ链(IL2-Rg、CD132)。IL-2受体γ(也被称为IL2r-γ或IL2Rg)链是由各种细胞因子受体共有的共同γ链，所述受体包括例如IL-2、IL-4、IL-7、IL-9、IL-15和IL-21的受体。IL-2Rg包含有助于结合配体的在细胞的细胞外表面上的外结构域、跨膜结构域、和可与各种分子相互作用以诱导细胞内信号转导路径的细胞内结构域。

I12rg基因在哺乳动物中见于X染色体上，并且人的γ链基因中的某些突变可导致特征在于深度T细胞缺陷的人X-连锁重度联合免疫缺陷(XSCID)。此外，γ链外结构域可脱离跨膜受体，并且以可溶性γ链受体形式释放。可溶性γ链受体可在受试者的血液中检测到，并且可起调控细胞因子信号传导的作用。

[0134] 在一些实施方案中，用人IL2-Rg链替换大鼠IL-2Rg链以使大鼠表达完全人IL-2Rg链。在其它情况下，可适用的是仅用人IL-2Rg链的外结构域替换大鼠IL-2Rg链的外结构域。在所述情况下，在大鼠中表达的所得人源化IL-2Rg链包含人外结构域，其中分子的其余部分来自大鼠。

[0135] IL-2Rg的全长人源化是适用的，因为具有这个经修饰的基因座的大鼠将产生人IL-2Rg。这将允许用对人IL-2Rg具有特异性的抗体检测大鼠中的人IL-2Rg。外人源化(即用IL-2Rg的人外结构域替换IL-2Rg的大鼠外结构域)将产生将结合IL2-Rg的人配体的IL-2Rg多肽，但因为细胞质结构域仍然是大鼠的，所以IL-2Rg的外人源化形式也将与大鼠信号传导机构相互作用。

[0136] 2.修饰大鼠靶基因座

[0137] A.靶向载体和插入核酸

[0138] i.插入核酸

[0139] 如本文所用，“插入核酸”包含希望整合在靶基因座处的DNA区段。在一个实施方案中，插入核酸包含一个或多个目标多核苷酸。在其它实施方案中，插入核酸可包含一个或多个表达盒。给定表达盒可包含目标多核苷酸、编码选择标志物和/或报告基因的多核苷酸以及影响表达的各种调控组分。可包括在插入核酸内的目标多核苷酸、选择标志物和报告基因的非限制性实例在本文其它地方详细讨论。

[0140] 在特定实施方案中，插入核酸可包含来自大鼠的核酸，其可包括基因组DNA的区段、cDNA、调控区或其任何部分或组合。在其它实施方案中，插入核酸可包括来自非人哺乳动物、啮齿动物、人、大鼠、小鼠、仓鼠、兔、猪、牛、鹿、绵羊、山羊、鸡、猫、狗、雪貂、灵长类动物(例如狨猿、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的核酸。如本文进一步详细概述，用于各种方法和组合物中的插入核酸可导致包含大鼠核酸的靶基因座的“人源化”。

[0141] 在一个实施方案中，插入核酸包含内源性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，插入核酸包含整个内源性基因的敲入等位基因(即“基因调换敲入(gene-swap knock-in)”)。

[0142] 在一个实施方案中，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。

[0143] 在其它实施方案中，插入核酸包含条件性等位基因。在一个实施方案中，条件性等位基因是如US 2011/0104799中所述的多功能等位基因，所述专利以引用的方式整体并入本文。在特定实施方案中，条件性等位基因包含：(a)关于靶标基因的转录呈正义定向的触动序列和呈正义或反义定向的药物选择盒；(b)呈反义定向的目标核苷酸序列(NSI)和倒位条件性模块(COIN，其利用外显子分割性内含子和可倒位基因诱捕样模块；参见例如以引用的方式整体并入本文的US2011/0104799)；和(c)在暴露于第一重组酶后重组以形成(i)缺乏触动序列和DSC，以及(ii)含有呈正义定向的NSI和呈反义定向的COIN的条件性等位基因

的可重组单元。

[0144] 插入核酸在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb的范围内。

[0145] 在一个实施方案中，插入核酸包含在约1kb至约200kb、约2kb至约20kb、或约0.5kb至约3Mb的范围内的大鼠基因组DNA序列的缺失。在一个实施方案中，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，基因组DNA序列的缺失程度在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约70kb至约80kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约110kb至约120kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、约190kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约800kb、约800kb至1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约2.8Mb、约2.8Mb至约3Mb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb的范围内。

[0146] 在一个实施方案中，插入核酸包含插入或用同源或直系同源人核酸序列替换大鼠核酸序列。在一个实施方案中，插入核酸包含在包含相应大鼠DNA序列的内源性大鼠基因座处插入或用同源或直系同源人核酸序列替换大鼠DNA序列。

[0147] 在一个实施方案中，遗传修饰是添加核酸序列。在一个实施方案中，添加的核苷酸序列在5kb至200kb的范围内。

[0148] 在一个实施方案中，插入核酸在编码序列中包含遗传修饰。在一个实施方案中，遗传修饰包括编码序列的缺失突变。在一个实施方案中，遗传修饰包括融合两个内源性编码序列。

[0149] 在一个实施方案中，插入核酸包含插入或用同源或直系同源人核酸序列替换大鼠核酸序列。在一个实施方案中，插入核酸包含在包含相应大鼠DNA序列的内源性大鼠基因座处插入或用同源或直系同源人核酸序列替换大鼠DNA序列。

[0150] 在一个实施方案中，遗传修饰包括缺失非蛋白质编码序列，但不包括缺失蛋白质编码序列。在一个实施方案中，非蛋白质编码序列的缺失包括调控元件的缺失。在一个实施方案中，遗传修饰包括缺失启动子。在一个实施方案中，遗传修饰包括添加启动子或调控元件。在一个实施方案中，遗传修饰包括替换启动子或调控元件。

[0151] 在一个实施方案中，靶向载体的核酸序列可包含当整合至基因组中时将产生对大鼠ApoE基因座的区域的遗传修饰的多核苷酸，其中在ApoE基因座处的遗传修饰导致ApoE活性降低、ApoE活性增加或ApoE活性受调节。在一个实施方案中，产生ApoE敲除(“无效等位基因”)。

[0152] 在一个实施方案中，靶向载体的核酸序列可包含当整合至基因组中时将产生对大鼠白介素-2受体基因座的区域的遗传修饰的多核苷酸，其中在白介素-2受体基因座处的遗传修饰导致白介素-2受体活性降低。在一个实施方案中，产生白介素-2受体敲除(“无效等

位基因”)。

[0153] 在其它实施方案中,插入核酸导致大鼠ApoE基因座、白介素-2受体γ基因座和/或Rag2基因座和/或Rag1基因座和/或Rag2/Rag1基因座的一部分被来自另一生物体的ApoE基因座、白介素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的相应同源或直系同源部分替换。

[0154] 其它实施方案,插入核酸包含跨越它的全长与它替换的ApoE基因座、白介素-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。

[0155] 给定插入多核苷酸和大鼠基因座的相应替换区域可为编码区、内含子、外显子、非翻译区、调控区、启动子、或增强子或其任何组合。此外,给定插入多核苷酸和/或大鼠基因座的缺失区域可具有任何所需长度,包括例如长度在10-100个核苷酸、100-500个核苷酸、500-1kb核苷酸、1Kb至1.5kb核苷酸、1.5kb至2kb核苷酸、2kb至2.5kb核苷酸、2.5kb至3kb核苷酸、3kb至5kb核苷酸、5kb至8kb核苷酸、8kb至10kb核苷酸之间或大于10kb核苷酸。在其它情况下,插入或替换的大小是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约800kb、约800kb至1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约2.8Mb、约2.8Mb至约3Mb。在其它实施方案中,给定插入多核苷酸和/或大鼠基因座的缺失区域是至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或大于16kb。

[0156] 在一个实施方案中,启动子是组成型活性启动子。

[0157] 在一个实施方案中,启动子是诱导型启动子。在一个实施方案中,诱导型启动子是化学调控启动子。在一个实施方案中,化学调控启动子是醇调控启动子。在一个实施方案中,醇调控启动子是醇脱氢酶(alcA)基因启动子。在一个实施方案中,化学调控启动子是四环素调控启动子。在一个实施方案中,四环素调控启动子是四环素响应性启动子。在一个实施方案中,四环素调控启动子是四环素操纵子序列(tetO)。在一个实施方案中,四环素调控启动子是tet开启启动子。在一个实施方案中,四环素调控启动子是tet关闭启动子。在一个实施方案中,化学调控启动子是类固醇调控启动子。在一个实施方案中,类固醇调控启动子是大鼠糖皮质素受体的启动子。在一个实施方案中,类固醇调控启动子是雌激素受体的启动子。在一个实施方案中,类固醇调控启动子是蜕皮激素(ecdysone)受体的启动子。在一个实施方案中,化学调控启动子是金属调控启动子。在一个实施方案中,金属调控启动子是金属蛋白质启动子。在一个实施方案中,诱导型启动子是物理调控启动子。在一个实施方案中,物理调控启动子是温度调控启动子。在一个实施方案中,温度调控启动子是热激启动子。在一个实施方案中,物理调控启动子是光调控启动子。在一个实施方案中,光调控启动子是光诱导型启动子。在一个实施方案中,光调控启动子是光阻遏性启动子。

[0158] 在一个实施方案中,启动子是组织特异性启动子。在一个实施方案中,启动子是神经元特异性启动子。在一个实施方案中,启动子是神经胶质特异性启动子。在一个实施方案中,启动子是肌肉细胞特异性启动子。在一个实施方案中,启动子是心脏细胞特异性启动子。在一个实施方案中,启动子是肾细胞特异性启动子。在一个实施方案中,启动子是骨细

胞特异性启动子。在一个实施方案中，启动子是内皮细胞特异性启动子。在一个实施方案中，启动子是免疫细胞特异性启动子。在一个实施方案中，免疫细胞启动子是B细胞启动子。在一个实施方案中，免疫细胞启动子是T细胞启动子。

[0159] 在一个实施方案中，启动子是发育调控启动子。在一个实施方案中，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，发育调控启动子仅在成体细胞中具有活性。

[0160] 在一些实施方案中，插入核酸包含侧接有位点特异性重组靶序列的核酸。应认识到尽管整个插入核酸可由所述位点特异性重组靶序列侧接，但插入核酸内的任何目标区域或个别多核苷酸都也可由所述位点侧接。位点特异性重组酶可通过任何手段引入细胞中，包括通过向细胞中引入重组酶多肽或通过向宿主细胞中引入编码位点特异性重组酶的多核苷酸。编码位点特异性重组酶的多核苷酸可位于插入核酸内或单独多核苷酸内。位点特异性重组酶可可操作地连接至在细胞中具有活性的启动子，包括例如诱导型启动子、细胞内源性启动子、细胞异源性启动子、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及其组合。

[0161] 在一些实施方案中，位点特异性重组位点侧接插入核酸内含有的编码选择标志物和/或报告基因的多核苷酸。在所述情况下，在插入核酸整合在靶向基因座处之后，可移除位点特异性重组位点之间的序列。

[0162] 在一个实施方案中，插入核酸包含编码选择标志物的多核苷酸。选择标志物可含于选择盒中。所述选择标志物包括但不限于新霉素磷酸转移酶(neo^r)、潮霉素B磷酸转移酶(hyg^r)、嘌呤霉素-N-乙酰基转移酶($puro^r$)、杀稻瘟菌素S脱氨酶(bsr^r)、黄嘌呤/鸟嘌呤磷酸核糖基转移酶(gpt)、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，编码选择标志物的多核苷酸可操作地连接至在细胞、大鼠细胞、多能大鼠细胞或ES大鼠细胞中具有活性的启动子。当向靶向基因座中连续堆积目标多核苷酸时，选择标志物可包含核酸酶试剂的识别位点，如上所概述。在一个实施方案中，编码选择标志物的多核苷酸侧接有位点特异性重组靶序列。

[0163] 插入核酸可进一步包含可操作地连接至启动子的报告基因，其中所述报告基因编码选自由以下组成的组或包括以下的报告蛋白：LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强黄色荧光蛋白(EYFP)、Emerald、增强绿色荧光蛋白(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。所述报告基因可可操作地连接至在细胞中具有活性的启动子。所述启动子可为诱导型启动子、报告基因或细胞内源性启动子、报告基因或细胞异源性启动子、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。

[0164] 在一个实施方案中，核酸插入可包括哺乳动物核酸，其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，哺乳动物核酸包含编码在骨髓或骨髓源性细胞中表达的蛋白质的基因组基因座。在一个实施方案中，核酸包含编码在脾细胞中表达的蛋白质的基因组基因座。

[0165] 在一个实施方案中，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，哺乳动物核酸包含编码在骨髓或骨髓源性细胞中表达的蛋白质的基因组基因座。在一个实施方案中，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，基因组基因座包含小鼠基因组DNA序列、大鼠基因组DNA序列、人基因组DNA序列或其组合。在一个实施方案中，基因组基因座以任何顺序包含大鼠和人基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含小鼠和人基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含小鼠和大鼠基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含大鼠、小鼠和人基因组DNA序列。

[0166] 在一个实施方案中，基因组基因座包含小鼠基因组DNA序列、大鼠基因组DNA序列、人基因组DNA序列或其组合。在一个实施方案中，基因组基因座以任何顺序包含大鼠和人基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含小鼠和人基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含小鼠和大鼠基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含大鼠、小鼠和人基因组DNA序列。

[0167] 在一个实施方案中，遗传修饰包括人基因的至少一个人疾病等位基因。在一个实施方案中，人疾病是神经疾病。在一个实施方案中，人疾病是心血管疾病。在一个实施方案中，人疾病是肾病。在一个实施方案中，人疾病是肌肉疾病。在一个实施方案中，人疾病是血液疾病。在一个实施方案中，人疾病是癌症。在一个实施方案中，人疾病是免疫系统疾病。

[0168] 在一个实施方案中，人疾病等位基因是显性等位基因。在一个实施方案中，人疾病等位基因是隐性等位基因。在一个实施方案中，人疾病等位基因包括单核苷酸多态性(SNP)等位基因。

[0169] 在一个实施方案中，遗传修饰产生蛋白质的具有改变的结合特征、改变的定位、改变的表达和/或改变的表达样式的突变形式。

[0170] 在一个实施方案中，插入核酸包含选择盒。在一个实施方案中，选择盒包含编码选择性标志物的核酸序列，其中所述核酸序列可操作地连接至在大鼠ES细胞中具有活性的启动子。在一个实施方案中，选择性标志物选自或包括潮霉素抗性基因或新霉素抗性基因。

[0171] 在一个实施方案中，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，核酸包含编码在不成熟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，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。

[0172] 在一个实施方案中，插入核酸包含调控元件。在一个实施方案中，调控元件是启动子。在一个实施方案中，调控元件是增强子。在一个实施方案中，调控元件是转录阻遏子结合元件。

[0173] 在一个实施方案中，遗传修饰包括缺失非蛋白质编码序列，但不包括缺失蛋白质编码序列。在一个实施方案中，非蛋白质编码序列的缺失包括调控元件的缺失。在一个实施方案中，遗传修饰包括缺失调控元件。在一个实施方案中，遗传修饰包括添加启动子或调控元件。在一个实施方案中，遗传修饰包括替换启动子或调控元件。

[0174] ii.表达盒

[0175] 本文提供包含用于本文提供的靶向基因组整合系统中的各种组分(即核酸酶试

剂、识别位点、插入核酸、目标多核苷酸、靶向载体、选择标志物和其它组分中的任一个或任何组合)的多核苷酸或核酸分子。

[0176] 术语“多核苷酸”、“多核苷酸序列”、“核酸序列”和“核酸片段”在本文中可互换使用。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RNA或DNA的聚合物。呈DNA的聚合物形式的多核苷酸可包括cDNA、基因组DNA、合成DNA或其混合物的一个或多个区段。多核苷酸可包括脱氧核糖核苷酸和核糖核苷酸,包括天然存在的分子与合成类似物两者及这些的任何组合。本文提供的多核苷酸也涵盖序列的所有形式,包括但不限于单链形式、双链形式、发夹、茎和环结构等。

[0177] 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”在本文中可互换使用。重组构建体包含人工或异源性核酸序列(例如不共同见于自然界中的调控和编码序列)组合。在其它实施方案中,重组构建体可包含源于不同来源的调控序列和编码序列,或源于相同来源,但以与在自然界中所见的方式不同的方式排列的调控序列和编码序列。所述构建体可单独使用,或可连同载体一起使用。如果使用载体,那么对载体的选择取决于如为本领域技术人员所熟知的用于转化宿主细胞的方法。举例来说,可使用质粒载体。也提供为成功转化、选择和繁殖包含任何本文提供的分离的核酸片段的宿主细胞所需的遗传元件。筛选可通过DNA的Southern分析、mRNA表达的Northern分析、蛋白质表达的免疫印迹分析、或表型分析以及其它分析来实现。

[0178] 在特定实施方案中,本文所述的靶向基因组整合系统的一种或多种组分可提供在用于在原核细胞、真核细胞、细菌、酵母细胞、或哺乳动物细胞或其它目标生物体或细胞类型中表达的表达盒中。所述盒可包括可操作地连接至本文提供的多核苷酸的5’和3’调控序列。“可操作地连接”包括其中可操作地连接的组分以它们的预定方式起作用的关系。举例来说,目标多核苷酸与调控序列(即启动子)之间的可操作连接是允许表达目标多核苷酸的功能性连接。可操作地连接的元件可为连续的或非连续的。当用于指代接合两个蛋白质编码区时,可操作地连接意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可操作地连接至调控序列(例如启动子、增强子、沉默子序列等)以便保持适当转录调控。在一种情况下,免疫球蛋白可变区(或V(D)J区段)的核酸序列可可操作地连接至免疫球蛋白恒定区的核酸序列以便允许在序列之间适当重组成免疫球蛋白重链或轻链序列。

[0179] 盒可另外含有至少一种待向生物体中共同引入的其它目标多核苷酸。或者,其它目标多核苷酸可提供在多个表达盒上。所述表达盒具有多个用于插入将受调控区的转录调控的重组多核苷酸的限制位点和/或重组位点。表达盒可另外含有选择标志物基因。

[0180] 表达盒可在5’-3’转录方向上包括在哺乳动物细胞或目标宿主细胞中具有功能性的转录和翻译起始区(即启动子)、本文提供的重组多核苷酸、以及转录和翻译终止区(即终止区)。本文提供的调控区(即启动子、转录调控区和翻译终止区)和/或多核苷酸可为宿主细胞或彼此天然的/类似的。或者,本文提供的调控区和/或多核苷酸可为宿主细胞或彼此异源性的。举例来说,可操作地连接至异源性多核苷酸的启动子来自与多核苷酸所源于的物种不同的物种,或如果来自相同/类似物种,那么一者或两者从它们的原始形式和/或基因组基因座被实质上修饰,或启动子不是可操作地连接的多核苷酸的天然启动子。或者,本文提供的调控区和/或重组多核苷酸可完全是合成的。

[0181] 终止区可关于转录起始区是天然的,可关于可操作地连接的重组多核苷酸是天然

的，可关于宿主细胞是天然的，或可相对于启动子、重组多核苷酸、宿主细胞或其任何组合源于另一来源(即外来或异源性)。

[0182] 在制备表达盒时，可操作各种DNA片段以便提供呈适当定向的DNA序列。为此，适体或连接子可用于接合DNA片段，或可涉及其它操作以提供适宜限制位点，移除多余DNA，移除限制位点等。出于这个目的，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。

[0183] 许多启动子可用于本文提供的表达盒中。可基于希望的结果选择启动子。应认识到通过在表达盒中使用不同启动子来调节目标多核苷酸的时序、位置和/或表达水平，可增强不同应用。必要时，所述表达构建体也可含有启动子调控区(例如，赋予诱导性、组成性、环境或发育调控、或细胞或组织特异性/选择性表达的启动子调控区)、转录起始开始位点、核糖体结合位点、RNA加工信号、转录终止位点和/或聚腺苷酸化信号。

[0184] 含有本文提供的多核苷酸的表达盒也可包含用于选择转化的细胞的选择标志物基因。可选择标志物基因用于选择转化的细胞或组织。

[0185] 当适当时，可优化用于方法和组合物中的序列(即目标多核苷酸、核酸酶试剂等)以达成在细胞中的表达增加。也就是说，可使用在给定目标细胞中优选的密码子(包括例如哺乳动物优选密码子、人优选密码子、啮齿动物优选密码子、小鼠优选密码子、大鼠优选密码子等)合成基因以提高表达。

[0186] 本文提供的各种方法和组合物可采用选择标志物。各种选择标志物可用于本文公开的方法和组合物中。所述选择标志物可例如赋予对抗生素，如G418、潮霉素、杀稻瘟菌素、新霉素或嘌呤霉素的抗性。所述选择标志物包括新霉素磷酸转移酶(neo^r)、潮霉素B磷酸转移酶(hyg^r)、嘌呤霉素-N-乙酰基转移酶($puro^r$)和杀稻瘟菌素S脱氨酶(bsr^r)。在其它实施方案中，选择标志物可操作地连接至诱导型启动子，并且选择标志物的表达对细胞具有毒性。所述选择标志物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-鸟嘌呤磷酸核糖基转移酶(HGPRT)或单纯疱疹病毒胸苷激酶(HSV-TK)。编码选择标志物的多核苷酸可操作地连接至在细胞中具有活性的启动子。

[0187] iii.靶向载体

[0188] 靶向载体用于向大鼠核酸的靶基因座中引入插入核酸。靶向载体包含插入核酸，并且进一步包含侧接插入核酸的5'和3'同源臂。侧接插入核酸的同源臂对应于大鼠核酸的靶基因座内的区域。为便于引用，靶向基因组基因座内的相应同源基因组区域在本文中被称为“靶位点”。举例来说，靶向载体可包含第一插入核酸，其由与第一和第二靶位点互补的第一和第二同源臂侧接。因此，靶向载体由此有助于通过发生在同源臂与细胞基因组内的互补靶位点之间的同源重组事件使插入核酸整合至大鼠核酸的靶基因座中。

[0189] 在一个实施方案中，大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，第一核酸序列和第二核酸序列相隔至少5kb。在另一实施方案中，第一核酸序列和第二核酸序列相隔至少5kb但小于200kb。在一个实施方案中，第一核酸序列和第二核酸序列相隔至少10kb。在一个实施方案中，第一核酸序列和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。在其它实施方案中，第

一核酸序列和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。

[0190] 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，包括例如长度是至少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或大于200kb。如以下进一步详细概述，大靶向载体可采用长度更大的靶向臂。在一特定实施方案中，5'同源臂和3'同源臂的总和是至少10kb，或5'同源臂和3'同源臂的总和是至少约16kb至约100kb或约30kb至约100kb。在其它实施方案中，LTVEC的5'同源臂和3'同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、或约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中，缺失的大小与LTVEC的5'同源臂和3'同源臂的总和大小相同或类似。

[0191] 当采用核酸酶试剂时，对应于靶向载体的5'和3'同源臂的同源基因组区域与核酸酶靶位点“足够邻近定位”以便促进在识别位点处获得切口或双链断裂后在同源基因组区域与同源臂之间发生同源重组事件。举例来说，核酸酶靶位点可定位在对应于5'和3'同源臂的同源基因组区域之间的任何地方。在特定实施方案中，识别位点紧邻同源基因组区域中的至少一个或两个。

[0192] 如本文所用，同源臂和靶位点(即同源基因组区域)“互补”，或彼此“互补”，此时两个区域彼此共有足以充当同源重组反应的底物的序列同一性水平。就“同源性”来说，其意指DNA序列与相应或“互补”序列是同一的，或与相应或“互补”序列共有序列同一性。给定靶位点与见于靶向载体上的相应同源臂之间的序列同一性可为允许发生同源重组的任何程度的序列同一性。举例来说，由靶向载体的同源臂(或其片段)和靶位点(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，以使序列经受同源重组。此外，同源臂与互补靶位点之间具有同源性的互补区域可具有足以促进在裂解的识别位点处进行同源重组的任何长度。举例来说，给定同源臂和/或互补靶位点可包含具有同源性的互补区域，其长度是至少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或大于200kb(如本文其它地方所述的LTVEC载体中所述)以使同源臂具有足以经受与细胞基因组内的相应靶位点的同源重组的同源性。为便于引用，同源臂在本文中被称为5'和3'同源臂。这个术语涉及同源臂相对于靶向载体内的插入核酸的相对位

置。

[0193] 因此，靶向载体的同源臂被设计以与靶向基因座的靶位点互补。因此，同源臂可与细胞天然基因座互补，或者，它们可与整合至细胞基因组中的异源性或外源性DNA区段的区域互补，所述区段包括但不限于转基因、表达盒或基因组DNA的异源性或外源性区域。或者，靶向载体的同源臂可与人人工染色体的区域或适当宿主细胞中含有的任何其它工程化基因组区域互补。更进一步，靶向载体的同源臂可与BAC文库、粘粒文库或P1噬菌体文库的区域互补或源于此。因此，在特定实施方案中，靶向载体的同源臂与对于给定细胞来说是天然、异源性或外源性的大鼠基因组基因座互补。在其它实施方案中，同源臂与不可使用常规方法靶向，或在不存在由核酸酶试剂诱导的切口或双链断裂下，仅可不正确靶向或仅具有显著低效率靶向的大鼠基因组基因座互补。在一个实施方案中，同源臂源于合成DNA。

[0194] 在其它实施方案中，5'和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，同源臂来自相关基因组，例如靶向基因组是第一品系的大鼠基因组，并且靶向臂来自第二品系的大鼠基因组，其中所述第一品系和所述第二品系不同。在其它实施方案中，同源臂来自相同动物的基因组或来自相同品系的基因组，例如靶向基因组是第一品系的大鼠基因组，并且靶向臂来自相同大鼠或相同品系的大鼠基因组。

[0195] 靶向载体(如大靶向载体)也可包含如本文其它地方讨论的选择盒或报告基因。选择盒可包含编码选择标志物的核酸序列，其中所述核酸序列可操作地连接至启动子。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。所述启动子可为诱导型启动子、报告基因或细胞内源性启动子、报告基因或细胞异源性启动子、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，选择标志物选自或包括新霉素磷酸转移酶(neo^r)、潮霉素B磷酸转移酶(hyg^r)、嘌呤霉素-N-乙酰基转移酶(puro^r)、杀稻瘟菌素S脱氨酶(bsr^r)、黄嘌呤/鸟嘌呤磷酸核糖基转移酶(gpt)、和单纯疱疹病毒胸苷激酶(HSV-k)和/或其组合。靶向载体的选择标志物可由5'和3'同源臂侧接或见于同源臂的5'或3'。

[0196] 在一个实施方案中，靶向载体(如大靶向载体)包含可操作地连接至启动子的报告基因，其中所述报告基因编码选自由以下组成的组或包括以下的报告蛋白：LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强黄色荧光蛋白(EYFP)、Emerald、增强绿色荧光蛋白(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。所述报告基因可可操作地连接至在细胞中具有活性的启动子。所述启动子可为诱导型启动子、报道基因或细胞内源性启动子、报告基因或细胞异源性启动子、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。

[0197] 在一个实施方案中，相较于单独使用靶向载体，组合使用靶向载体(包括例如大靶向载体)与核酸酶试剂导致靶向效率增加。在一个实施方案中，当相较于单独使用靶向载体时，当靶向载体连同核酸酶试剂一起使用时，靶向载体的靶向效率增加至少2倍、至少3倍、或至少4倍。

[0198] 当采用靶向载体时，载体设计可为如此以致允许插入如本文所述的约5kb至约200kb给定序列。在一个实施方案中，插入是约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约80kb至约

90kb、约90kb至约100kb、约100kb至约110kb、约110kb至约120kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、或约190kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb。

[0199] 当采用靶向载体时，载体设计可为如此以致允许替换如本文所述的约5kb至约200kb或约5kb至约3.0Mb给定序列。在一个实施方案中，替换是约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约110kb至约120kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、约190kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb。

[0200] 在一个实施方案中，靶向载体包含位点特异性重组酶基因。在一个实施方案中，位点特异性重组酶基因编码Cre重组酶。在一个实施方案中，Cre重组酶基因是Crei，其中编码Cre重组酶的两个外显子由内含子分隔以防止它在原核细胞中表达。

[0201] 在一个实施方案中，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)定位于核(例如基因是NL-Cre基因)。在一特定实施方案中，Cre重组酶基因进一步包含核定位信号和内含子(例如NL-Crei)。

[0202] 在各种实施方案中，适于表达核酸酶试剂(包括以上讨论的Cre或Crei重组酶)的启动子选自或包括Prm1、Blimp1、Gata6、Gata4、Igf2、Lhx2、Lhx5和/或Pax3。在一特定实施方案中，启动子是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。在另一特定实施方案中，启动子是Prm1启动子。在另一特定实施方案中，启动子是大鼠Prm1启动子。在另一特定实施方案中，启动子是小鼠Prm1启动子。在另一特定实施方案中，启动子是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，其两者均以引用的方式整体并入本文。

[0203] iv.大靶向载体

[0204] 如本文所用的术语“大靶向载体”或“LTVEC”包括以下大靶向载体：所述载体包含对应于以及源于大于通常由意图在细胞中进行同源靶向的其它方法使用的核酸序列的核酸序列的同源臂；和/或包含含有大于通常由意图在细胞中进行同源重组靶向的其它方法使用的核酸序列的核酸序列的插入核酸。举例来说，LTVEC使得有可能修饰由于它们的大小限制而不能被传统基于质粒的靶向载体容纳的大基因座。在特定实施方案中，LTVEC的同源臂和/或插入核酸包含真核细胞的基因组序列。LTVEC的大小太大以致不能通过常规测定，例如DNA印迹和长程(例如1kb-5kb)PCR来筛选靶向事件。LTVEC的实例包括但不限于源于细菌人工染色体(BAC)、人人工染色体或酵母人工染色体(YAC)的载体。LTVEC和用于制备它们

的方法的非限制性实例例如描述于美国专利No.6,586,251、6,596,541、7,105,348以及WO 2002/036789(PCT/US01/45375)和US 2013/0137101中,所述专利各自以引用的方式并入本文。

[0205] LTVEC可具有任何长度,包括但不限于约20kb至约400kb、约20kb至约30kb、约30kb至40kb、约40kb至约50kb、约50kb至约75kb、约75kb至约100kb、约100kb至125kb、约125kb至约150kb、约150kb至约175kb、约175kb至约200kb、约200kb至约225kb、约225kb至约250kb、约250kb至约275kb、或约275kb至约300kb、约200kb至约300kb、约300kb至约350kb、约350kb至约400kb、约350kb至约550kb。在一个实施方案中,LTVEC是约100kb。

[0206] 在一个实施方案中,LTVEC包含在约5kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约0.5kb至约30kb、约0.5kb至约40kb、约30kb至约150kb、约0.5kb至约150kb、约30kb至约40kb、约40kb至约50kb、约60kb至约70kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、或约190kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb的范围内的插入核酸。

[0207] 当采用LTVEC时,载体设计可为如此以致允许替换如本文所述的约5kb至约200kb或约5kb至约3Mb给定序列。在一个实施方案中,替换是约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约110kb至约120kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、约190kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb。

[0208] 在一个实施方案中,LTVEC的同源臂源于BAC文库、粘粒文库或P1噬菌体文库。在其它实施方案中,同源臂源于细胞的靶向基因组基因座,并且在一些情况下,LTVEC被设计以靶向的靶基因组基因座不可使用常规方法靶向。在其它实施方案中,同源臂源于合成DNA。

[0209] 在一个实施方案中,LTVEC中的5'同源臂和3'同源臂的总和是至少10kb。在其它实施方案中,LTVEC的5'同源臂和3'同源臂的总和是约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、100kb至约120kb、约120kb至约140kb、约140kb至约160kb、约160kb至约180kb、约180kb至约200kb。在一个实施方案中,LTVEC的5'同源臂和3'同源臂的总和是约30kb至约100kb。在其它实施方案中,LTVEC的5'同源臂和3'同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、或约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中,缺失的大小与LTVEC的5'同源臂和3'同源臂的总和大小相同或类似。

[0210] 在其它实施方案中,5'同源臂在约5kb至约100kb的范围内。在一个实施方案中,3'同源臂在约5kb至约100kb的范围内。在其它实施方案中,5'同源臂和3'同源臂的总和是约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约50kb至约60kb、约60kb至约70kb、约70kb至约80kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约110kb至约120kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、约190kb至约200kb、或约30kb至约100kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。

[0211] 在一个实施方案中,LTVEC包含由LTVEC同源臂侧接的与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,插入核酸序列来自除大鼠以外的物种。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是哺乳动物核酸。在一个实施方案中,哺乳动物核酸是小鼠核酸。在一个实施方案中,哺乳动物核酸是人核酸。在一个实施方案中,插入核酸是基因组DNA。在一个实施方案中,插入物是如上所述的5kb至200kb。

[0212] 在一个实施方案中,LTVEC包含选择盒或报告基因。可采用的各种形式的选择盒和报告基因在本文其它地方讨论。

[0213] 如本文其它地方所述,LTVEC也可与促进靶向载体与多能大鼠细胞中的大鼠核酸的靶基因座之间的同源重组的核酸酶试剂组合用于本文提供的方法中。

[0214] 在一个实施方案中,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,位点特异性重组酶基因编码Cre重组酶。在一个实施方案中,Cre重组酶基因是Crei,其中编码Cre重组酶的两个外显子由内含子分隔以防止它在原核细胞中表达。在一个实施方案中,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)定位于核(例如基因是NL-Cre基因)。在一特定实施方案中,Cre重组酶基因进一步包含核定位信号和内含子(例如NL-Crei)。

[0215] 在各种实施方案中,适于表达核酸酶试剂(包括以上讨论的Cre或Crei重组酶)的启动子选自或包括Prm1、Blimp1、Gata6、Gata4、Igf2、Lhx2、Lhx5和/或Pax3。在一特定实施方案中,启动子是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。在另一特定实施方案中,启动子是Prm1启动子。在另一特定实施方案中,启动子是大鼠Prm1启动子。在另一特定实施方案中,启动子是小鼠Prm1启动子。在另一特定实施方案中,启动子是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,其两者均以引用的方式整体并入本文。

[0216] 在一个实施方案中,LTVEC包含可产生大鼠ApoE基因座、IL-2Rg基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的区域的缺失、添加、替换或其组合的插入核酸,如本文其它地方详细讨论。在特定实施方案中,在ApoE基因座处的遗传修饰导致ApoE活性、IL-2Rg活性、Rag2活性、Rag1活性和/或Rag2和Rag1活性降低、增加或受调节。在一个实施方案中,产生ApoE敲除、和IL-2Rg敲除、Rag2敲除、Rag1敲除、Rag2/Rag1敲除。如下所讨论,核酸酶试剂可与任何LTVEC靶向系统一起用于靶向任何目标基因组基因座。

[0217] v.核酸酶试剂和核酸酶试剂的识别位点

[0218] 如以上详细概述,核酸酶试剂可在本文公开的方法和组合物中用于协助修饰原核

细胞中或多能大鼠细胞内的靶基因座。所述核酸酶试剂可促进靶向载体与靶基因座之间的同源重组。在一个实施方案中，核酸酶试剂包括核酸内切酶试剂。

[0219] 如本文所用，术语“核酸酶试剂的识别位点”包含核酸酶试剂诱导切口或双链断裂所处的DNA序列。核酸酶试剂的识别位点可为细胞内源性的(或天然的)，或识别位点可为细胞外源性的。在特定实施方案中，识别位点是细胞外源性的，并且由此不天然存在于细胞基因组中。在其它实施方案中，识别位点是细胞和希望定位在靶基因组基因座处的目标多核苷酸外源性的。在其它实施方案中，外源性或内源性识别位点在宿主细胞的基因组中仅存在一次。在特定实施方案中，鉴定在基因组内仅出现一次的内源性或天然位点。所述位点可接着用于设计将在内源性识别位点处产生切口或双链断裂的核酸酶试剂。

[0220] 识别位点的长度可变化，并且包括例如长度是至少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或大于70个核苷酸的识别位点。在一个实施方案中，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，识别位点的长度是约9至约12个核苷酸、约12至约15个核苷酸、约15至约18个核苷酸、或约18至约21个核苷酸以及所述子范围的任何组合(例如9-18个核苷酸)。识别位点可为回文的，也就是说一条链上的序列与互补链在相反方向上的读取相同。应认识到给定核酸酶试剂可结合识别位点并裂解那个结合位点，或者，核酸酶试剂可结合不同于识别位点的序列。此外，术语识别位点包括核酸酶试剂结合位点与切口/裂解位点两者而不考虑切口/裂解位点在核酸酶试剂结合位点内部或外部。在另一变化形式中，由核酸酶试剂进行的裂解可发生在彼此紧密相对的核苷酸位置处以产生钝性末端切割，或在其它情况下，切口可交错以产生也称为“粘性末端”的单链突出部分，其可为5’突出部分或3’突出部分。

[0221] 在所需识别位点中诱导切口或双链断裂的任何核酸酶试剂都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核酸酶试剂，只要所述核酸酶试剂在所需识别位点中诱导切口或双链断裂即可。或者，可采用经修饰的或工程化核酸酶试剂。“工程化核酸酶试剂”包括从它的天然形式工程化(修饰或衍生)以在所需识别位点中特异性识别并诱导切口或双链断裂的核酸酶。因此，工程化核酸酶试剂可源于天然、天然存在的核酸酶试剂，或它可被人工创建或合成。对核酸酶试剂的修饰可少至蛋白质裂解试剂中的一个氨基酸或核酸裂解试剂中的一个核苷酸。在一些实施方案中，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由天然(非工程化或非修饰)核酸酶试剂识别的序列。在识别位点或其它DNA中产生切口或双链断裂可在本文中称为“切割”或“裂解”所述识别位点或其它DNA。

[0222] 也提供例示的识别位点的活性变体和片段。所述活性变体可包含与给定识别位点的至少65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或大于99％序列同一性，其中活性变体保留生物活性，并且因此能够被核酸酶试剂以序列特异性方式识别和裂解。用以测量由核酸酶试剂对识别位点达成的双链断裂的测定在本领域中是已知的，并且通常测量核酸酶切割识别位点的能力。

[0223] 核酸酶试剂的识别位点可定位在靶基因座中或附近的任何地方。识别位点可位于基因的编码区内，或影响所述基因的表达的调控区内。因此，核酸酶试剂的识别位点可位于

内含子、外显子、启动子、增强子、调控区、或任何非蛋白质编码区中。

[0224] 在一个实施方案中，核酸酶试剂是转录激活因子样效应物核酸酶(TALEN)。TAL效应物核酸酶是一类可用于在原核或真核生物体的基因组中的特定靶序列处产生双链断裂的序列特异性核酸酶。TAL效应物核酸酶通过使天然或工程化转录激活因子样(TAL)效应物或其功能性部分融合于核酸内切酶(例如像FokI)的催化结构域来创建。独特模块化TAL效应物DNA结合结构域允许设计具有潜在任何给定DNA识别特异性的蛋白质。因此，TAL效应物核酸酶的DNA结合结构域可被工程化以识别特定DNA靶位点，并且因此用于在所需靶序列处产生双链断裂。参见WO 2010/079430；Morbitzer等(2010)PNAS 10.1073/pnas.1013133107；Scholze和Boch(2010)Virulence1:428-432；Christian等Genetics(2010)186:757-761；Li等(2010)Nuc.Acids Res.(2010)doi:10.1093/nar/gkq704；以及Miller等(2011)Nature Biotechnology 29:143-148；其全部以引用的方式并入本文。

[0225] 适合的TAL核酸酶和用于制备适合的TAL核酸酶的方法的实例例如公开于美国专利申请号2011/0239315A1、2011/0269234A1、2011/0145940A1、2003/0232410A1、2005/0208489A1、2005/0026157A1、2005/0064474A1、2006/0188987A1和2006/0063231A1(各自据此以引用的方式并入本文)中。在各种实施方案中，使TAL效应物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附近进行切割，其中所述靶核酸序列在待通过靶向载体修饰的序列处或附近。适于与本文提供的各种方法和组合物一起使用的TAL核酸酶包括被特定设计以在待通过如本文所述的靶向载体修饰的靶核酸序列处或附近进行结合的那些。

[0226] 在一个实施方案中，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子位点。在一个实施方案中，核酸酶试剂是包含可操作地连接至独立核酸酶的基于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，独立核酸酶是FokI核酸内切酶。在一个实施方案中，核酸酶试剂包含第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域，其中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域各自可操作地连接至FokI核酸酶，其中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域识别靶DNA序列的各链中由约6bp至约40bp裂解位点分隔的两个连续靶DNA序列，并且其中FokI核酸酶二聚化并在靶序列处产生双链断裂。

[0227] 在一个实施方案中，核酸酶试剂包含第一基于TAL重复序列的DNA结合结构域和第二基于TAL重复序列的DNA结合结构域，其中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域各自可操作地连接至FokI核酸酶，其中所述第一基于TAL重复序列的DNA结合结构域和所述第二基于TAL重复序列的DNA结合结构域识别靶DNA序列的各链中由5bp或6bp裂解位点分隔的两个连续靶DNA序列，并且其中FokI核酸酶二聚化并产生双链断裂。

[0228] 用于本文公开的各种方法和组合物中的核酸酶试剂可进一步包括锌指核酸酶(ZFN)。在一个实施方案中，ZFN的各单体包含3个或大于3个基于锌指的DNA结合结构域，其中各基于锌指的DNA结合结构域结合3bp子位点。在其它实施方案中，ZFN是包含可操作地连接至独立核酸酶的基于锌指的DNA结合结构域的嵌合蛋白。在一个实施方案中，独立核酸内切酶是FokI核酸内切酶。在一个实施方案中，核酸酶试剂包含第一ZFN和第二ZFN，其中所述

第一ZFN和所述第二ZFN各自可操作地连接至FokI核酸酶，其中所述第一ZFN和所述第二ZFN识别靶DNA序列的各链中由约6bp至约40bp裂解位点或约5bp至约6bp裂解位点分隔的两个连续靶DNA序列，并且其中FokI核酸酶二聚化并产生双链断裂。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。

[0229] 在本文提供的方法的一个实施方案中，核酸酶试剂包括(a)包含融合于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)包含融合于FokI核酸内切酶的转录激活因子样效应物核酸酶(TALEN)的嵌合蛋白。

[0230] 在另一实施方案中，核酸酶试剂是兆碱基大范围核酸酶。兆碱基大范围核酸酶已基于保守序列基序被分类成四个家族，所述家族是LAGLIDADG(SEQ ID NO:16)、GIY-YIG、H-N-H和His-Cys框家族。这些基序参与金属离子的配位和磷酸二酯键的水解。HE酶以它们的长识别位点以及以容许它们的DNA底物中的一些序列多态性而著称。兆碱基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，(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；以及Moure等，(2002)Nat Struct Biol 9:764。在一些实施例中，使用天然存在的变体和/或工程化衍生兆碱基大范围核酸酶。用于修改动力学、辅因子相互作用、表达、最优条件和/或识别位点特异性、以及活性筛选的方法是已知的，参见例如Epinat等，(2003)Nucleic Acids Res 31:2952-62；Chevalier等，(2002)Mol Cell 10:895-905；Gimble等，(2003)Mol Biol 334:993-1008；Seligman等，(2002)Nucleic Acids Res30:3870-9；Sussman等，(2004)J Mol Biol 342:31-41；Rosen等，(2006)Nucleic Acids Res 34:4791-800；Chames等，(2005)Nucleic Acids Res33:e178；Smith等，(2006)Nucleic Acids Res 34:e149；Gruen等，(2002)Nucleic Acids Res 30:e29；Chen和Zhao，(2005)Nucleic Acids Res33:e154；WO2005105989；WO2003078619；WO2006097854；WO2006097853；WO2006097784；以及WO2004031346。

[0231] 在本文中可使用任何兆碱基大范围核酸酶，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-CmoeI、I-CpaI、I-CpaII、I-CsmI、I-CvuI、I-CvuAIP、I-DdiI、I-DdiII、I-DirI、I-DmoI、I-HmuI、I-HmuII、I-HsNIP、I-LlaI、I-MsoI、I-NaaI、I-NanI、I-NcIIP、I-NgrIP、I-NitI、I-NjaI、I-Nsp236IP、I-PakI、I-PboIP、I-PcuIP、I-PcuAI、I-PcuVI、I-PgrIP、I-PobIP、I-PorI、I-PorIIP、I-PbpIP、I-SpBetaIP、I-ScaI、I-SexIP、I-SneIP、I-SpomI、I-SpomCP、I-SpomIP、I-SpomIIP、I-SquIP、I-Ssp6803I、I-SthPhiJP、I-SthPhiST3P、I-SthPhiSTe3bP、I-TdeIP、I-TevI、I-TevII、I-TevIII、I-UarAP、I-UarHGPAIP、I-UarHGPA13P、I-VinIP、I-ZbiIP、PI-MtuI、PI-MtuHIP PI-MtuHIIP、PI-PfuI、PI-PfuII、PI-PkoI、PI-PkoII、PI-Rma43812IP、PI-SpBetaIP、PI-SceI、PI-TfuI、PI-TfuII、PI-ThyI、PI-TliI、PI-TliII或其任何活性变体或片段。

[0232] 在一个实施方案中，兆碱基大范围核酸酶识别12至40个碱基对的双链DNA序列。在一个实施方案中，兆碱基大范围核酸酶识别基因组中的一个完全匹配靶序列。在一个实施

方案中，兆碱基大范围核酸酶是归巢核酸酶。在一个实施方案中，归巢核酸酶是归巢核酸酶的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-Dmol。

[0233] 核酸酶试剂可进一步包括限制核酸内切酶，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定识别位点，但通常在离开核酸酶结合位点的可变位置处裂解，所述结合位点可远离裂解位点(识别位点)数百个碱基对。在II型系统中，限制活性不依赖于任何甲基酶活性，并且裂解通常发生在结合位点内或附近的特定位点处。大多数II型酶切割回文序列，然而，IIa型酶识别非回文识别位点并在识别位点外部裂解，IIb型酶切割序列两次，其中两个位点均在识别位点外部，并且IIs型酶识别不对称识别位点并在一侧以及在离开识别位点约1-20个核苷酸的确定距离处裂解。IV型限制酶靶向甲基化DNA。例如于REBASE数据库(网页在rebase.neb.com处;Roberts等,(2003)Nucleic Acids Res 31:418-20);Roberts等,(2003)Nucleic Acids Res 31:1805-12;以及Belfort等,(2002),Mobile DNA II,第761-783页,Craigie等编,(ASM Press,Washington,DC)中进一步描述以及分类限制酶。

[0234] 用于各种方法和组合物中的核酸酶试剂也可包括CRISPR/Cas系统。所述系统可采用例如Cas9核酸酶，其在一些情况下针对它将被表达所处的所需细胞类型加以密码子优化。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。这个单一RNA常被称为导向RNA或gRNA。在gRNA内，crRNA部分被鉴定为给定识别位点的‘靶序列’，而tracrRNA常被称为‘骨架’。简要来说，将含有靶序列的短DNA片段插入导向RNA表达质粒中。gRNA表达质粒包含靶序列(在一些实施方案中约20个核苷酸)、某一形式的tracrRNA序列(骨架)以及在细胞中具有活性的适合启动子和为在真核细胞中适当加工所必需的元件。许多系统依赖于退火以形成双链DNA，接着克隆至gRNA表达质粒中的定制互补寡聚物。接着将gRNA表达盒和Cas9表达盒引入细胞中。参见例如Mali P等(2013)Science2013年2月15日;339(6121):823-6;Jinek M等Science 2012年8月17日;337(6096):816-21;Hwang WY等Nat Biotechnol 2013年3月;31(3):227-9;Jiang W等Nat Biotechnol 2013年3月;31(3):233-9;以及Cong L等Science 2013年2月15日;339(6121):819-23,其各自以引用的方式并入本文。

[0235] 在一个实施方案中，用于修饰多能大鼠细胞中的目标基因组基因座的方法进一步包括向所述多能大鼠细胞中引入:(a)包含可操作地连接至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)包含可操作地连接至基因组靶序列的第二启动子的第二表达构建体，所述基因组靶序列连接至导向RNA(gRNA)，其中所述基因组靶序列通过前间区序列邻近基序(PAM)序列侧接在3’末端上。在一个实施方案中，基因组靶序列包含核苷酸序列GNNNNNNNNNNNNNNNNNNNG G(GN_{1-20}GG;SEQ ID NO:1)。在一个实施方案中，基因组靶序列包含SEQ ID NO:23，其中N的长度在1与20个核苷酸之间。在另一实施方案中，基因组靶序列包含SEQ ID NO:1的14与20个核苷酸之间的长度。

[0236] 在一个实施方案中，gRNA包含编码成簇规律间隔短回文重复序列(CRISPR)RNA(crRNA)和反式活化CRISPR RNA(tracrRNA)的第三核酸序列。在特定实施方案中，Cas蛋白是Cas9。

[0237] 在一些实施方案中，gRNA包括(a)具有核酸序列5'-GUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUA UCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3'(SEQ I D NO:2)的嵌合RNA;或(b)具有核酸序列5'-GUUUUAGAGC UAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCG-3'(SEQ ID N O:3)的嵌合RNA。

[0238] 在另一实施方案中，crRNA包含5'-GUUUUAGAGCUAGAAAU AGCAAGUUAAAAU-3'(SEQ ID NO:4);5'-GUUUUAGAGCUAG AAAUAGCAAGUUAAAAUAAG(SEQ ID NO:5);或5'-GAGUCCGAGCAGAAGAAGAAGUUUUA-3'(SEQ ID NO:6)。

[0239] 在其它实施方案中，tracrRNA包含5'-AAGGCUAGUCCG-3'(S EQ ID NO:7)或5'-AAGGCUAGUCCGU UAUCAACUUGAAAAA GUGGCACCGAGUCGGUGCUUUU-3'(SEQ ID NO:8)。

[0240] 在一个实施方案中，Cas蛋白是I型Cas蛋白。在一个实施方案中，Cas蛋白是II型Cas蛋白。在一个实施方案中，II型Cas蛋白是Cas9。在一个实施方案中，第一核酸序列编码人密码子优化的Cas蛋白。

[0241] 在一个实施方案中，第一核酸包含破坏Cas蛋白中的核酸酶活性位点的至少一个氨基酸残基的突变，其中突变Cas蛋白仅在靶DNA区域的一个链中产生断裂，并且其中所述突变减弱靶DNA区域中的非同源重组。

[0242] 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，核定位信号是SV40核定位信号。

[0243] 在一个实施方案中，驱动基因组靶序列和导向RNA(gRNA)的表达的第二启动子是RNA聚合酶III启动子。在一个实施方案中，RNA聚合酶III启动子是人U6启动子。在一个实施方案中，RNA聚合酶III启动子是大鼠U6聚合酶III启动子。在一个实施方案中，RNA聚合酶III启动子是小鼠U6聚合酶III启动子。

[0244] 在一个实施方案中，编码crRNA的核酸序列和编码tracrRNA的核酸序列通过合成环连接，其中在表达后crRNA和tracrRNA形成crRNA:tracrRNA双链体。

[0245] 在一个实施方案中，第一表达构建体和第二表达构建体从同一质粒表达。

[0246] 在一个实施方案中，连同LTVEC一起引入第一和第二表达构建体。在一个实施方案中，历经一段时期与LTVEC分开引入第一和第二表达构建体。

[0247] 在一个实施方案中，方法包括引入多个第二构建体和多个LTVEC以对如本文所述的不同靶基因座进行多路编辑(multiplex editing)。

[0248] 也提供核酸酶试剂的活性变体和片段(即工程化核酸酶试剂)。所述活性变体可包含与天然核酸酶试剂的至少65％、70％、75％、80％、85％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或大于99％序列同一性，其中活性变体保留在所需识别位点处切割的能力，并且因此保留切口或双链断裂诱导活性。举例来说，任何本文所述的核酸酶试剂都可从天然核酸内切酶序列被修饰，并且被设计以在不由天然核酸酶试剂识别的识别位点处识别并诱导切口或双链断裂。因此，在一些实施方案中，工程化核酸酶具有用以在不同于相应天然核酸酶试剂识别位点的识别位点处诱导切口或双链断裂的特异性。针对切口或双链断裂诱导活性的测定是已知的，并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。

[0249] 可通过本领域中已知的任何手段将核酸酶试剂引入细胞中。编码核酸酶试剂的多肽可被直接引入细胞中。或者，可将编码核酸酶试剂的多核苷酸引入细胞中。当将编码核酸

酶试剂的多核苷酸引入细胞中时，核酸酶试剂可在细胞内短暂、条件性或组成型表达。因此，编码核酸酶试剂的多核苷酸可含于表达盒中，并且可操作地连接至条件性启动子、诱导型启动子、组成型启动子或组织特异性启动子。所述目标启动子在本文其它地方进一步详细讨论。或者，将核酸酶试剂以编码或包含核酸酶试剂的mRNA形式引入细胞中。

[0250] 在一个实施方案中，cRNA和tracrRNA被表达为单独RNA转录物。

[0251] 在特定实施方案中，编码核酸酶试剂的多核苷酸被稳定整合在细胞基因组中，并且可操作地连接至在细胞中具有活性的启动子。在其它实施方案中，编码核酸酶试剂的多核苷酸在包含插入核酸的同一靶向载体中，而在其它情况下，编码核酸酶试剂的多核苷酸在与包含插入核酸的靶向载体分开的载体或质粒中。

[0252] 当通过引入编码核酸酶试剂的多核苷酸向细胞提供核酸酶试剂时，所述编码核酸酶试剂的多核苷酸可被修饰以用相较于天然存在的编码核酸酶试剂的多核苷酸序列，在目标细胞中具有较高使用频率的密码子进行取代。举例来说，编码核酸酶试剂的多核苷酸可被修饰以用相较于天然存在的多核苷酸序列，在给定目标原核或真核细胞(包括细菌细胞、酵母细胞、人细胞、非人细胞、哺乳动物细胞、啮齿动物细胞、小鼠细胞、大鼠细胞或任何其它目标宿主细胞)中具有较高使用频率的密码子进行取代。

[0253] 在一个实施方案中，连同LTVEC一起引入核酸内切酶试剂。在一个实施方案中，历经一段时期与LTVEC分开引入核酸内切酶试剂。在一个实施方案中，在引入LTVEC之前引入核酸内切酶试剂。在一个实施方案中，在引入LTVEC之后将核酸内切酶试剂引入大鼠ES细胞中。

[0254] 在一个实施方案中，核酸内切酶试剂是包含编码核酸内切酶的核酸序列的表达构建体，其中所述核酸序列可操作地连接至启动子。在一个实施方案中，启动子是组成型活性启动子。在一个实施方案中，启动子是诱导型启动子。在一个实施方案中，启动子在多能大鼠细胞中具有活性。在一个实施方案中，核酸内切酶试剂是编码核酸内切酶的mRNA。

[0255] B.用于使目标多核苷酸整合至靶基因座中的方法

[0256] 提供用于修饰目标靶基因座的方法。在一个实施方案中，靶向多能大鼠细胞中的靶基因座以达成遗传修饰。所述方法包括：(a)向多能大鼠细胞中引入包含侧接有5'大鼠同源臂和3'大鼠同源臂的插入核酸的靶向载体；以及(b)鉴定在靶基因座处包含靶向遗传修饰的经遗传修饰的多能大鼠细胞，其中靶向遗传修饰能够通过种系传递。在特定实施方案中，5'同源臂和3'同源臂的总和是至少10kb，和/或采用大靶向载体。

[0257] 在其它实施方案中，LTVEC的5'同源臂和3'同源臂的总体的总和大小是约10kb至约150kb、约10kb至约100kb、约10kb至约75kb、约20kb至约150kb、约20kb至约100kb、约20kb至约75kb、约30kb至约150kb、约30kb至约100kb、约30kb至约75kb、约40kb至约150kb、约40kb至约100kb、约40kb至约75kb、约50kb至约150kb、约50kb至约100kb、或约50kb至约75kb、约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至约150kb。在一个实施方案中，缺失的大小与LTVEC的5'同源臂和3'同源臂的总和大小相同或类似。

[0258] 多能大鼠细胞可为大鼠胚胎干细胞。在一特定实施方案中，(a)大鼠ES细胞源于DA品系或ACI品系；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能性标志物。在其它情况下，采用的大鼠胚胎干细胞包括如2014年2月20日提交，以引

用的方式整体并入本文的美国专利申请号14/185,103中所述的大鼠ES细胞。

[0259] 如本文其它地方所述，插入核酸可为任何核酸序列。在非限制性实施方案中，(a)插入核酸包含内源性大鼠核酸序列被同源或直系同源哺乳动物核酸序列的替换；(b)插入核酸包含内源性大鼠核酸序列的缺失；(c)插入核酸包含内源性大鼠核酸序列的缺失，其中所述缺失在5kb至200kb或5kb至3Mb的范围内(如本文其它地方详细讨论)；(d)插入核酸包含外源性核酸序列(包括例如在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb的范围内的外源性核酸序列)的添加；(e)插入核酸包含包括同源或直系同源核酸序列的外源性核酸序列；(f)(a)的同源或直系同源核酸序列，其中核酸序列是人核酸序列；(g)插入核酸包含(a)的同源或直系同源核酸序列，其中核酸序列是包含人核酸序列和大鼠核酸序列的嵌合核酸序列；(h)插入核酸包含(e)的外源性核酸序列，其中插入核酸在约5kb至约200kb的范围内；(i)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)插入核酸包含可操作地连接至启动子的报告基因；(k)插入核酸包含一个或多个未重排人免疫球蛋白重链V_H基因区段、一个或多个未重排人免疫球蛋白重链D基因区段和一个或多个未重排人免疫球蛋白重链J_H基因区段，其可操作地连接至啮齿动物重链恒定区核酸序列；(l)插入核酸包含可操作地连接至啮齿动物重链恒定区核酸序列的重排人免疫球蛋白重链可变区核酸序列；(m)插入核酸包含一个或多个未重排人免疫球蛋白V_κ或V_λ基因区段和一个或多个未重排人免疫球蛋白J_κ或J_λ基因区段，其可操作地连接至哺乳动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列；(n)插入核酸包含可操作地连接至哺乳动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列的重排人免疫球蛋白λ或κ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括大鼠恒定区核酸序列、人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λ或κ轻链恒定区核酸包括大鼠恒定区核酸序列、人恒定区核酸序列或其组合。

[0260] 在一个实施方案中，插入核酸包含一个或多个功能性人V_H基因区段，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_H 3-30-5、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。

[0261] 在一个实施方案中，插入核酸包含一个或多个功能性人D基因区段，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。

[0262] 在一个实施方案中，插入核酸包含一个或多个功能性J_H基因区段，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，插入核酸包含一个或多个人V_κ基因区段，包括V_κ4-1、V_κ5-2、V_κ7-3、V_κ2-4、V_κ1-5、V_κ1-6、V_κ3-7、V_κ1-8、V_κ1-9、V_κ2-10、V_κ3-11、V_κ1-12、V_κ1-13、V_κ2-14、V_κ3-15、V_κ1-16、V_κ1-17、V_κ2-18、V_κ2-19、V_κ3-20、V_κ6-21、V_κ1-22、V_κ1-23、V_κ2-24、V_κ3-25、V_κ2-26、V_κ1-27、V_κ2-28、V_κ2-29、V_κ2-30、V_κ3-31、V_κ1-32、V_κ1-33、V_κ3-34、

Vκ1-35、Vκ2-36、Vκ1-37、Vκ2-38、Vκ1-39、Vκ2-40或其组合。

[0263] 在一个实施方案中，插入核酸包含一个或多个人Vλ基因区段，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。

[0264] 在一个实施方案中，插入核酸包含一个或多个人Jκ基因区段，包括Jκ1、Jκ2、Jκ3、Jκ4、Jκ5或其组合。

[0265] 在特定实施方案中，在修饰多能大鼠细胞中的靶基因座后，遗传修饰通过种系传递。

[0266] 在一个实施方案中，插入核酸序列包含当整合至基因组中时将产生对大鼠ApoE基因座的区域的遗传修饰的多核苷酸，其中在ApoE基因座处的遗传修饰导致ApoE活性降低、ApoE活性增加或ApoE活性受调节。在一个实施方案中，产生ApoE敲除。

[0267] 在一个实施方案中，插入核酸序列包含当整合至基因组中时将产生对大鼠白介素-2受体γ基因座的区域的遗传修饰的多核苷酸，其中在白介素-2受体γ基因座处的遗传修饰导致白介素-2受体活性降低、白介素-2受体γ活性增加或白介素-2受体活性受调节。在一个实施方案中，产生白介素-2受体敲除。

[0268] 在另一实施方案中，插入核酸序列包含当整合至基因组中时将产生对大鼠Rag1基因座、大鼠Rag2基因座和/或大鼠Rag2/Rag1基因座的区域的遗传修饰的多核苷酸，其中在大鼠Rag1、Rag2和/或Rag2/Rag1基因座处的遗传修饰导致Rag1、Rag2或Rag1和Rag2蛋白活性降低、Rag1、Rag2或Rag1和Rag2蛋白活性增加、或Rag1、Rag2或Rag1和Rag2蛋白活性受调节。在一个实施方案中，产生Rag1、Rag2或Rag2/Rag1敲除。

[0269] 在其它实施方案中，插入核酸导致大鼠ApoE基因座、白介素-2受体γ基因座和/或Rag2基因座、和/或Rag1基因座和/或Rag2/Rag1基因座的一部分被来自另一生物体的ApoE基因座、白介素-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。

[0270] 其它实施方案，插入核酸包含跨越它的全长与它替换的ApoE基因座、白介素-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。

[0271] 给定插入多核苷酸和大鼠基因座的相应替换区域可为编码区、内含子、外显子、非翻译区、调控区、启动子、或增强子或其任何组合。此外，给定插入多核苷酸和/或大鼠基因座的替换区域可具有任何所需长度，包括例如长度在10-100个核苷酸、100-500个核苷酸、500-1kb核苷酸、1Kb至1.5kb核苷酸、1.5kb至2kb核苷酸、2kb至2.5kb核苷酸、2.5kb至3kb核苷酸、3kb至5kb核苷酸、5kb至8kb核苷酸、8kb至10kb核苷酸之间或大于10kb核苷酸。在其它情况下，插入或替换的大小是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、约350kb至约400kb、约400kb至约800kb、约800kb至1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、约2.5Mb至约2.8Mb、约2.8Mb至约3Mb。在其它实施方案中，给定插入多核苷酸和/或大鼠基因座的替换区域是至少100、200、300、400、500、600、700、800或900个核苷酸或至少1kb、2kb、3kb、4kb、5kb、6kb、7kb、8kb、9kb、10kb、11kb、12kb、13kb、14kb、15kb、16kb或大于16kb。

[0272] i.用于通过细菌同源重组(BHR)来修饰大鼠核酸的靶基因座的方法

[0273] 提供用于通过原核细胞中的细菌同源重组(BHR)来修饰大鼠核酸的靶基因座的方法和组合物。所述方法适用于利用原核细胞中的细菌同源重组来遗传修饰大鼠核酸的靶基因座以创建靶向载体。可将所述包含遗传修饰的靶基因座的靶向载体引入真核细胞,例如多能大鼠细胞中。“同源重组”包括在具有同源性的区域内的交叉位点处在两个DNA分子之间交换DNA片段。因此,“细菌同源重组”或“BHR”包括发生在细菌中的同源重组。

[0274] 提供用于通过细菌同源重组(BHR)来修饰大鼠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’大鼠同源臂和3’大鼠同源臂的插入核酸的靶向载体,其中所述原核细胞包含大鼠核酸,并且能够表达介导在靶基因座处的BHR的重组酶。所述靶向载体可包括任何本文所述的大靶向载体。

[0275] 在一个实施方案中,方法包括向原核细胞中引入:(i)包含具有目标DNA序列的大鼠核酸的第一构建体;(ii)包含侧接有大鼠5’同源臂和大鼠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,历经一段时期将第一、第二和第三构建体分别引入原核细胞中。在一个实施方案中,原核细胞包含编码重组酶的核酸,并且方法不需要引入第三构建体。在一个实施方案中,重组酶在诱导型启动子控制下表达。

[0276] 在一个实施方案中,包含大鼠核酸的第一构建体源于细菌人工染色体(BAC)或酵母人工染色体(YAC)。

[0277] 可选择在靶基因组基因座处包含插入核酸的原核细胞。这个方法可如本文所公开来连续重复以使得在原核细胞中在靶向大鼠基因座处引入多个插入核酸。一旦在原核细胞内“构建”靶大鼠核酸基因座,即可从原核细胞分离包含经修饰的大鼠靶基因座的靶向载体,并且引入哺乳动物细胞(即大鼠细胞、多能大鼠细胞或大鼠胚胎干细胞)内的靶基因组基因座中。

[0278] 用于接受靶向载体的优选大鼠细胞描述于2014年2月20日提交的美国申请14/185,703中,所述申请的内容概述于本文中。这些大鼠细胞是能够在一次或多次体外靶向遗传修饰之后持续它们的多能性,并且能够通过种系传递靶向遗传修饰的多能大鼠细胞。

[0279] 在高密度下涂铺电穿孔多能大鼠细胞以选择包含靶向载体的药物抗性细胞。药物选择过程移除大多数涂铺细胞(约99%),留下个别集落,其各自是源于单细胞的克隆。在剩余细胞中,大多数细胞(约80-100%)含有整合在基因组中的随机位置处的靶向载体(包含药物选择盒)。因此,个别地挑选集落并分析基因型以鉴定在正确基因组位置处具有靶向载体的大鼠ES细胞(例如使用下述等位基因修饰测定)。

[0280] 高通量定量测定,即等位基因修饰(MOA)测定,可用于分析基因型。所述测定允许在遗传修饰之后大规模筛选亲本染色体中的修饰等位基因。MOA测定可通过各种分析技术来进行,所述技术包括但不限于定量PCR,例如实时PCR(qPCR)。举例来说,实时PCR包括识别靶基因座的第一引物组和识别非靶向参照基因座的第二引物组。此外,引物组包含识别扩增的序列的荧光探针。在一个实施方案中,定量测定通过Invader Probes®来进行。在一个实施方案中,定量测定通过MMP assays®来进行。在一个实施方案中,定量测定通过TaqMan®分子信标来进行。在一个实施方案中,定量测定通过Eclipse™探测技术来进行。

(参见例如US2005/0144655,其以引用的方式整体并入本文)。

[0281] 可接着将所选包含靶向遗传修饰的多能大鼠细胞或大鼠ES细胞引入宿主大鼠胚胎(例如前桑椹胚阶段或胚泡阶段大鼠胚胎)中,并且植入代孕母体的子宫中以产生首建大鼠(F0大鼠)。随后,首建大鼠可与野生型大鼠配种以创建遗传修饰杂合性F1子代。杂合性F1大鼠的交配可产生遗传修饰纯合性子代。杂合性F1大鼠的交配可产生遗传修饰纯合性子代。在一些实施方案中,本文所述的靶基因座的各种遗传修饰可使用如本文其它地方详细所述的大靶向载体(LTVEC)进行。举例来说,LTVEC可使用VELOCIGENE®遗传工程技术从细菌人工染色体(BAC)DNA获得(参见例如美国专利No.6,586,251以及Valenzuela,D.M.等(2003),High-throughput engineering of the mouse genome coupled with high-resolution expression analysis,Nature Biotechnology 21(6):652-659,其以引用的方式整体并入本文)。

[0282] 使用细菌同源重组(BHR)来产生大靶向载体(LTVEC)会规避质粒在容纳大基因组DNA片段方面的限制以及随之发生的向多能大鼠细胞中的内源性基因座中引入靶向修饰的低效率。可在产生LTVEC时进行一次或多次靶向遗传修饰。在原核细胞中产生的一示例性LTVEC可包含携带具有一个或多个遗传修饰的大鼠基因组序列或外源性核酸(例如大鼠核酸的同源物或直系同源物)的插入核酸,其由与特定基因组区域互补的大鼠同源臂侧接。

[0283] 也提供包含本文所述的各种靶向载体的宿主原核细胞。所述原核细胞包括但不限于细菌,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,其包含侧接有5'大鼠同源臂和3'大鼠同源臂的插入核酸,其中所述插入核酸在约5kb至约200kb的范围内。

[0284] 宿主原核细胞可进一步包含编码重组酶多肽的核酸,或编码重组酶多肽的核酸可操作地连接至诱导型启动子。

[0285] 进一步提供各种方法和组合物,其采用与原核细胞组合的如本文所述的LTVEC来产生靶向遗传修饰。所述组合物和方法在本文其它地方讨论。

[0286] 提供用于通过细菌同源重组(BHR)来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所述5'同源臂和所述3'同源臂的核酸,并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。所述靶向载体可包括任何本文所述的大靶向载体。所述方法可采用如本文详细讨论的LTVEC,并且进一步采用如本文其它地方讨论的CRISPR/Cas系统。

[0287] ii.用于修饰多能大鼠细胞中的目标靶基因座的方法

[0288] 进一步提供一种用于通过靶向遗传修饰来修饰多能大鼠细胞中的目标靶基因座的方法,其包括(a)向所述多能大鼠细胞中引入包含侧接有5'大鼠同源臂和3'大鼠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'同源臂的总和是至少10kb;以及(b)鉴定在所述目标靶基因座处包含靶向遗传修饰的经遗传修饰的多能大鼠细胞。在一个实施方案中,5'同源臂和3'同源臂的总和是至少约16kb至约30kb。在特定实施方案中,靶向遗传修饰能够通过种系传递。所述靶向载体可包括任何本文所述的大靶向载体。

[0289] 在一个方面,提供一种用于通过靶向遗传修饰来修饰多能大鼠细胞中的目标基因组基因座的方法,其包括:(a)提供多能大鼠细胞,其能够在对它的基因组进行至少一次靶向遗传修饰之后持续它的多能性,并且能够向F1代的种系传递靶向修饰;(b)向所述多能大鼠细胞中引入大靶向载体(LTVEC),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’同源臂包含大鼠基因组DNA片段；以及(c)鉴定包含靶向遗传修饰的经遗传修饰的多能大鼠细胞。

[0290] 各种方法可用于鉴定具有整合在目标靶基因座处的插入核酸的细胞。插入核酸插入在目标靶基因座处会产生“等位基因修饰”。术语“等位基因修饰”以及用于检测修饰的等位基因的方法在本文其它地方进一步详细讨论。

[0291] 在一个方面，提供一种用于通过核酸内切酶介导的基因靶向来修饰多能大鼠细胞中的目标基因组基因座的方法，所述方法包括：(a)提供能够向F1代的种系传递遗传修饰基因组的分离的多能大鼠细胞；(b)向所述多能大鼠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目标基因组基因座中的靶DNA序列处产生切口或双链断裂，并且其中在大鼠ES细胞中的所述靶DNA序列处的所述切口或所述双链断裂诱导：(i)对所述切口或所述双链断裂的非同源末端接合(NHEJ)介导的DNA修复，其中所述NHEJ介导的DNA修复产生在所述靶DNA序列处包含核酸序列的插入或缺失的突变等位基因；或(ii)导致恢复野生型核酸序列的同源重组介导的DNA修复；以及(c)鉴定修饰的目标基因组基因座。

[0292] 在一个方面，提供一种用于通过核酸酶试剂来修饰分离的大鼠胚胎干细胞(ES)中的目标基因组基因座的方法，其包括：(a)提供能够向F1代的种系传递靶向遗传修饰的分离的大鼠ES细胞；(b)向大鼠ES细胞中引入：(i)包含侧接有大鼠5’同源臂和大鼠3’同源臂的插入核酸的大靶向载体(LTVEC)，其中插入物是至少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目标基因组基因座中的靶DNA序列处产生切口或双链断裂，并且其中靶序列不存在于所述插入核酸中；以及(c)鉴定大鼠胚胎干(ES)细胞中的靶向遗传修饰。

[0293] 在一个方面，提供一种用于通过RNA导向基因组工程化来修饰多能大鼠细胞中的目标基因组基因座的方法，所述方法包括：(a)提供能够向F1代的种系传递遗传修饰基因组的多能大鼠细胞；(b)向所述多能大鼠细胞中引入：(i)包含可操作地连接至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)包含可操作地连接至基因组靶序列的第二启动子的第二表达构建体，所述基因组靶序列连接至导向RNA(gRNA)，其中所述基因组靶序列通过前间区序列邻近基序(PAM)序列侧接在3’末端上。在一个实施方案中，基因组靶序列包含核苷酸序列GNNNNNNNNNNNNNNNNNNNGG(GN$_{1-20}$GG；SEQ ID NO:1)。在一个实施方案中，基因组靶序列包含SEQ ID NO:1，其中N的长度在14与20个核苷酸之间。在一个实施方案中，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活化CRISPR RNA(tracrRNA)的第四核酸序列。在一个实施方案中，在表达后，Cas蛋白形成包含crRNA和tracrRNA的CRISPR-Cas复合物，并且CRISPR-Cas复合物在位于目标基因组基因座中的靶DNA序列处产生切口或双链断裂，并且其中在多能大鼠细胞中的靶DNA序列处的切口或双链断裂诱导：(i)对由CRISPR-Cas复合物创建的切口或双链断裂的非同源末端接合(NHEJ)介导的DNA修复，其中NHEJ产生在靶DNA序列处包含核酸序列的插入或缺失的突变等位基因；或(ii)导致恢复野生型核酸序列的同源重组介导的DNA修复；以及(c)鉴定修饰的目标基因组基因座。

[0294] 在一个实施方案中，多能大鼠细胞是诱导的大鼠多能干细胞(iPS)。在一个实施方案中，多能大鼠细胞是发育限制的祖细胞。

[0295] 在各种实施方案中，在选择标志物内的识别位点中存在切口或双链断裂使靶向载体(如LTVEC)与目标靶向基因座之间的重组效率和/或频率增加。在一个实施方案中，重组是同源重组。在另一实施方案中，重组是通过非同源末端接合达成的插入。在各种实施方案中，在切口或双链断裂存在下，靶向载体(如LTVEC)在靶基因组基因座处的靶向效率比在不存在切口或双链断裂下(使用例如相同靶向载体和相同同源臂以及在目标基因组基因座处的相应靶位点，但不存在添加的会产生切口或双链断裂的核酸酶试剂)高至少约2倍、至少约3倍、至少约4倍。

[0296] 在一个实施方案中，在靶基因座处的靶向遗传修饰是双等位基因修饰。就“双等位基因”来说，其意指基因的两个等位基因均包含靶向遗传修饰。在某些实施方案中，相较于单独使用靶向载体，组合使用靶向载体(包括例如LTVEC)与核酸酶试剂导致对细胞中的目标基因组基因座的双等位基因靶向遗传修饰。相较于当单独使用靶向载体时，当靶向载体连同核酸酶试剂一起使用时，双等位基因靶向效率增加至少2倍、至少3倍、至少4倍或大于4倍。在其它实施方案中，双等位基因靶向效率是至少0.2％、0.3％、0.4％、0.5％、0.6％、0.7％、0.8％、0.9％、1％、2％、3％、4％或5％或高于5％。

[0297] 提供包含在白介素-2受体γ基因座中或在ApoE基因座中具有靶向遗传修饰的经遗传修饰的大鼠的组合物。本文提供的各种方法和组合物允许这些修饰基因座通过种系传递。

[0298] 在特定实施方案中，经遗传修饰的大鼠或经遗传修饰的多能大鼠细胞包含在白介素-2γ受体基因座中具有靶向遗传修饰或在ApoE基因座中具有靶向遗传修饰的基因组基因座，其中所述白介素-2γ受体基因组基因座或所述ApoE基因座包含：(i)所述白介素-2γ受体基因座的至少一部分或所述ApoE基因座的至少一部分的缺失；(ii)异源性核酸序列向所述ApoE基因座中或向所述白介素-2γ受体基因座中的插入；或(iii)其组合，其中遗传修饰基因组基因座能够通过种系传递。

[0299] 进一步提供允许制备所述经遗传修饰的大鼠和所述经遗传修饰的多能大鼠细胞的方法。所述方法包括一种用于通过靶向遗传修饰来修饰多能大鼠细胞中的ApoE基因组基因座或白介素-2γ受体基因座的方法。方法包括(a)向多能大鼠细胞中引入包含侧接有ApoE基因座的5’大鼠同源臂和ApoE基因座的3’大鼠同源臂的插入核酸的靶向载体，(b)鉴定在目标ApoE基因组基因座处包含靶向遗传修饰的经遗传修饰的多能大鼠细胞，其中靶向遗传修饰能够通过种系传递。

[0300] 其它方法包括(a)向多能大鼠细胞中引入包含侧接有白介素-2受体γ基因座的5’大鼠同源臂和白介素-2受体γ基因座的3’大鼠同源臂的插入核酸的靶向载体，(b)鉴定在白介素-2受体γ基因座处包含靶向遗传修饰的经遗传修饰的多能大鼠细胞，其中靶向遗传修饰能够通过种系传递。

[0301] iii.在靶向基因座处整合多个目标多核苷酸的方法

[0302] 本文提供的各种方法和组合物允许使多个目标多核苷酸与给定靶基因座靶向整合。上述各种方法可依序重复以允许使任何数目的插入核酸靶向整合至给定靶向基因座中。因此，各种方法提供使至少1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20或大于20个插入核酸插入靶基因座中。在特定实施方案中，所述依序堆积方法允许将来自哺乳动物细胞(即人、非人、啮齿动物、小鼠、猴、大鼠、仓鼠、驯化哺乳动物或农业动物)的大

基因组区域重构至靶向基因座中。在所述情况下，包括编码区与非编码区两者的基因组区域的转移和重构允许给定区域的复杂性得以通过至少部分保留见于天然基因组区域内的编码区、非编码区和拷贝数变化而保持。因此，各种方法提供例如用以在任何目标哺乳动物细胞或动物内，特别是在原核宿主细胞内或在多能大鼠细胞或大鼠ES细胞内产生“异源性”或“外源性”基因组区域的方法。在一个非限制性实施例中，产生在非人动物内(即在大鼠内)的“人源化”基因组区域。

[0303] 3.人源化基因组基因座

[0304] 本文提供包含人源化大鼠基因座的各种方法和组合物。如本文所用，就“人源化”基因组基因座来说，其意指非人基因组的包含至少一个人核酸序列的区域。“人源化大鼠基因座”包含大鼠DNA的具有插入其中的人DNA序列的区域。人DNA序列可为天然存在的人DNA序列，或它可从它的天然形式被修饰。在特定实施方案中，人DNA与天然人序列共有至少85％、90％、91％、92％、93％、94％、95％、96％、97％、98％或99％序列同一性。如果人序列不是天然人序列，那么它与天然人序列的序列同一性至少大于它与直系同源大鼠序列的序列同一性。此外，人DNA序列可包括cDNA、人基因组DNA的区域、非编码调控区、或人DNA的编码区、基因组区域或调控区的任何部分。插入大鼠基因座中的人DNA序列可包含如本文其它地方所述的任何插入多核苷酸。在特定实施方案中，人DNA序列与大鼠靶基因座直系同源，而在其它情况下，人DNA序列与大鼠靶基因座同源。

[0305] 在一个实施方案中，靶向遗传修饰是插入或用同源或直系同源人核酸序列替换内源性大鼠核酸序列。在一个实施方案中，靶向遗传修饰包括在包含相应大鼠核酸序列的内源性大鼠基因座处插入或用同源或直系同源人核酸序列替换内源性大鼠核酸序列。

[0306] 用于制备人源化大鼠基因座(或包含人源化大鼠基因座的大鼠或大鼠细胞)的方法包括向包含大鼠核酸的靶基因座中引入人核酸序列。在一个实施方案中，提供一种制备人源化大鼠的方法。所述方法包括(a)用包含插入核酸的靶向载体修饰多能大鼠细胞的基因组以形成供体细胞，所述插入核酸包含人核酸序列；(b)将所述供体细胞引入宿主大鼠胚胎中；以及(c)在代孕母体中孕育所述宿主大鼠胚胎；其中所述代孕母体产生包含所述人核酸序列的大鼠子代。在特定实施方案中，人源化大鼠基因座能够通过种系传递。在另一实施方案中，靶向载体包括大靶向载体(LTVEC)，并且包含人核酸序列的插入核酸是至少5kb。

[0307] 在其它方法中，通过经由细菌同源重组(BHR)修饰大鼠核酸的靶基因座来制备人源化大鼠基因座。方法包括向原核细胞中引入包含侧接有5’大鼠同源臂和3’大鼠同源臂的插入核酸的靶向载体，其中所述插入核酸包含人核酸序列，并且其中所述原核细胞包含大鼠核酸，并且能够表达介导在靶基因座处的BHR的重组酶。

[0308] 人源化大鼠基因组基因座可包含(a)同源或直系同源人核酸序列的插入；(b)内源性大鼠核酸序列被同源或直系同源人核酸序列的替换；或(c)其组合。在特定实施方案中，人源化大鼠基因组基因座能够通过种系传递。在其它实施方案中，人直系同源序列替换见于大鼠中的相应序列。

[0309] 任何人核酸序列都可用于本文提供的方法和组合物中。可用于方法和组合物中的人核酸序列的非限制性实例在本文其它地方详细讨论。

[0310] 用于插入目标大鼠基因座中的人核酸序列可为任何大小。在一个实施方案中，人核酸序列可为约500个核苷酸至约200kb、约500个核苷酸至约5kb、约5kb至约200kb、约5kb

至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约60kb至约70kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、或约190kb至约200kb。在一特定实施方案中，人核酸序列是至少5kb。

[0311] 在一个实施方案中，提供一种大鼠基因组基因座，其中同源或直系同源人核酸序列包含(a)一个或多个未重排人免疫球蛋白重链V_H基因区段、一个或多个未重排人免疫球蛋白重链D基因区段和一个或多个未重排人免疫球蛋白重链J_H基因区段，其可操作地连接至哺乳动物重链恒定区核酸序列；(b)可操作地连接至哺乳动物免疫球蛋白重链恒定区核酸序列的重排人免疫球蛋白重链可变区核酸序列；(c)一个或多个未重排人免疫球蛋白V_κ或V_λ基因区段和一个或多个未重排人免疫球蛋白J_κ或J_λ基因区段，其可操作地连接至哺乳动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列；或(d)可操作地连接至哺乳动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列的重排人免疫球蛋白λ或κ轻链可变区核酸序列。

[0312] 在另一实施方案中，提供一种大鼠基因组基因座，其中(a)哺乳动物免疫球蛋白重链恒定区核酸序列是大鼠恒定区核酸序列、人恒定区核酸序列或其组合；或(b)哺乳动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列是大鼠恒定区核酸序列、人恒定区核酸序列或其组合。

[0313] 在一特定实施方案中，提供一种大鼠基因组基因座，其中免疫球蛋白重链恒定区核酸序列选自或包括CH1、铰链、CH2、CH3和/或其组合。

[0314] 在一个实施方案中，大鼠基因组基因座包含一个或多个功能性人V_H基因区段，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_H 3-30-5、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。

[0315] 在一个实施方案中，大鼠基因组基因座包含一个或多个功能性人D基因区段，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。

[0316] 在一个实施方案中，大鼠基因组基因座包含一个或多个功能性J_H基因区段，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，插入核酸包含一个或多个人V_κ基因区段，包括V_κ4-1、V_κ5-2、V_κ7-3、V_κ2-4、V_κ1-5、V_κ1-6、V_κ3-7、V_κ1-8、V_κ1-9、V_κ2-10、V_κ3-11、V_κ1-12、V_κ1-13、V_κ2-14、V_κ3-15、V_κ1-16、V_κ1-17、V_κ2-18、V_κ2-19、V_κ3-20、V_κ6-21、V_κ1-22、V_κ1-23、V_κ2-24、V_κ3-25、V_κ2-26、V_κ1-27、V_κ2-28、V_κ2-29、V_κ2-30、V_κ3-31、V_κ1-32、V_κ1-33、V_κ3-34、V_κ1-35、V_κ2-36、V_κ1-37、V_κ2-38、V_κ1-39、V_κ2-40或其组合。

[0317] 在一个实施方案中，大鼠基因组基因座包含一个或多个人V_λ基因区段，包括V_λ3-1、V_λ4-3、V_λ2-8、V_λ3-9、V_λ3-10、V_λ2-11、V_λ3-12、V_λ2-14、V_λ3-16、V_λ2-18、V_λ3-19、V_λ3-21、V_λ3-22、V_λ2-23、V_λ3-25、V_λ3-27或其组合。

[0318] 在一个实施方案中，大鼠基因组基因座包含一个或多个人J_κ基因区段，包括J_κ1、J

κ2、Jκ3、Jκ4、Jκ5或其组合。

[0319] 在另一实施方案中，提供大鼠基因组基因座，其包括包含人白介素-2受体(IL2R)核酸序列或其变体或片段的人源化基因组基因座。在特定实施方案中，IL2R核酸序列包括白介素-2受体α、白介素-2受体β或白介素-2受体γ核酸序列或其变体或片段。

[0320] 在其它实施方案中，大鼠基因组基因座包括人源化基因组基因座，其包含替换大鼠ApoE基因座、大鼠白介素-2受体γ基因座、大鼠Rag2基因座、大鼠Rag1基因座和/或大鼠Rag2/Rag1基因座的相应同源或直系同源部分的人ApoE基因座、人白介素-2受体γ基因座、人Rag2基因座、人Rag1基因座和/或人Rag2/Rag1基因座的一部分。在一个实施方案中，大鼠IL-2Rg的外结构域被人IL-2Rg的外结构域替换，其中分子的其余部分来自大鼠。

[0321] 在另一实施方案中，提供一种包含人源化基因组基因座的经遗传修饰的大鼠。所述经遗传修饰的大鼠包含(a)同源或直系同源人核酸序列的插入；(b)在内源性基因组基因座处大鼠核酸序列被同源或直系同源人核酸序列的替换；或(c)其组合，其中人源化基因组基因座能够通过种系传递。

[0322] 也提供包含本文提供以及以上所述的各种人源化基因组基因座中的任一个的经遗传修饰的大鼠。

[0323] 4.目标多核苷酸

[0324] 任何目标多核苷酸都可含于各种插入核酸中，并且由此整合在靶基因座处。本文公开的方法提供待整合至靶向基因组基因座中的至少1、2、3、4、5、6或大于6个目标多核苷酸。

[0325] 插入核酸内的目标多核苷酸在整合在靶基因组基因座处时可将一种或多种遗传修饰引入细胞中。遗传修饰可包括缺失内源性核酸序列和/或向靶基因组基因座中添加外源性或异源性或直系同源多核苷酸。在一个实施方案中，遗传修饰包括在靶基因组基因座处用外源性目标多核苷酸替换内源性核酸序列。因此，本文提供的方法允许产生包括敲除、缺失、插入、替换(“敲入”)、点突变、结构域调换、外显子调换、内含子调换、调控序列调换、基因调换或其组合的遗传修饰。所述修饰可在使第一、第二、第三、第四、第五、第六、第七或任何后续插入核酸整合至靶基因组基因座中后发生。

[0326] 在插入核酸内和/或整合在靶基因座处的目标多核苷酸可包含对它被引入所处的细胞的天然的序列；目标多核苷酸可为它被引入所处的细胞的异源性的；目标多核苷酸可为它被引入所处的细胞的外源性的；目标多核苷酸可为它被引入所处的细胞的直系同源的；或目标多核苷酸可来自不同于它被引入所处的细胞的物种。如本文关于插入在靶基因座处的序列所用的“天然”是具有靶基因座的细胞的天然序列或靶基因座所源于的细胞(即来自大鼠)的天然序列。如本文所用，关于序列的“异源性”包括序列源于外来物种，或如果来自相同物种，那么由于故意人为干预而在组成和/或基因组基因座方面实质上不同于它的天然形式或从它的天然形式被修饰。如本文所用，关于序列的“外源性”是源于外来物种的序列。目标多核苷酸可来自任何目标生物体，包括但不限于非人、啮齿动物、仓鼠、小鼠、大鼠、人、猴、农业哺乳动物或非农业哺乳动物。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，第1、第2、第3、第4、第5、第6、第7和/或任何后续插入核酸可包含所述序列。

[0327] 在一个实施方案中，在插入核酸内和/或整合在靶基因座处的目标多核苷酸对小

鼠核酸序列、人核酸、非人核酸、啮齿动物核酸、大鼠核酸、仓鼠核酸、猴核酸、农业哺乳动物核酸或非农业哺乳动物核酸是天然的。在其它实施方案中，整合在靶基因座处的目标多核苷酸是基因组核酸的片段。在一个实施方案中，基因组核酸是小鼠基因组核酸、人基因组核酸、非人核酸、啮齿动物核酸、大鼠核酸、仓鼠核酸、猴核酸、农业哺乳动物核酸或非农业哺乳动物核酸或其组合。

[0328] 在一个实施方案中，目标多核苷酸可在如上所述的约500个核苷酸至约200kb的范围内。目标多核苷酸可为约500个核苷酸至约5kb、约5kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约30kb、约30kb至约40kb、约40kb至约50kb、约60kb至约70kb、约80kb至约90kb、约90kb至约100kb、约100kb至约110kb、约120kb至约130kb、约130kb至约140kb、约140kb至约150kb、约150kb至约160kb、约160kb至约170kb、约170kb至约180kb、约180kb至约190kb、或约190kb至约200kb、约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb。

[0329] 在插入核酸内和/或插入在靶基因组基因座处的目标多核苷酸可编码多肽，可编码miRNA，或它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、或非蛋白质编码序列的缺失，但不包含蛋白质编码序列的缺失。此外，在插入核酸内和/或插入在靶基因组基因座处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，在插入核酸内和/或插入在靶基因组基因座处的目标多核苷酸编码在骨髓或骨髓源性细胞中表达的蛋白质。在一个实施方案中，在插入核酸内和/或整合在靶基因座处的目标多核苷酸编码在脾细胞中表达的蛋白质。在其它实施方案中，在插入核酸内和/或插入在靶基因座处的目标多核苷酸编码在B细胞中表达的蛋白质，编码在不成熟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。

[0330] 在插入多核苷酸内的目标多核苷酸可包含ApoE基因座、IL-2-Rg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其它地方讨论，可采用的来自任何目标生物体的各种同源和直系同源区域也在本文其它地方讨论。

[0331] 在一个实施方案中，在插入核酸内和/或插入在靶基因座处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”在本文其它地方描述。

[0332] 在一个实施方案中，在插入核酸内和/或整合在靶基因座处的目标多核苷酸包含编码人免疫球蛋白重链可变区氨基酸序列的基因组核酸序列。

[0333] 在一个实施方案中，基因组核酸序列包含一个或多个未重排人免疫球蛋白重链V_H基因区段、一个或多个未重排人免疫球蛋白重链D基因区段和一个或多个未重排人免疫球蛋白重链J_H基因区段，其可操作地连接至哺乳动物重链恒定区核酸序列。在一个实施方案中，基因组核酸序列包含可操作地连接至哺乳动物重链恒定区核酸序列的重排人免疫球蛋白重链可变区核酸序列。在一个实施方案中，基因组核酸序列包含一个或多个未重排人免疫球蛋白V_κ或V_λ基因区段和一个或多个未重排人免疫球蛋白J_κ或J_λ基因区段，其可操作地

连接至哺乳动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列。在一个实施方案中，基因组核酸序列包含可操作地连接至哺乳动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列的重排人免疫球蛋白λ或κ轻链可变区核酸序列。在一个实施方案中，重链恒定区核酸序列包括大鼠恒定区核酸序列、人恒定区核酸序列或其组合。在一个实施方案中，免疫球蛋白λ或κ轻链恒定区核酸包括大鼠恒定区核酸序列、人恒定区核酸序列或其组合。

[0334] 在一个实施方案中，免疫球蛋白重链恒定区核酸序列选自或包括CH1、铰链、CH2、CH3和/或其组合。在一个实施方案中，重链恒定区核酸序列包括CH1-铰链-CH2-CH3。

[0335] 在一个实施方案中，在插入核酸内和/或整合在靶基因座处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，并且在本文其它地方描述。

[0336] 在一个实施方案中，在插入核酸内和/或整合在靶基因组基因座处的目标多核苷酸包含编码人免疫球蛋白轻链可变区氨基酸序列的基因组核酸序列。

[0337] 在一个实施方案中，基因组核酸序列包含一个或多个未重排人免疫球蛋白V_κ或V_λ基因区段和一个或多个未重排人免疫球蛋白J_κ或J_λ基因区段，其可操作地连接至啮齿动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列。在一个实施方案中，基因组核酸序列包含可操作地连接至啮齿动物免疫球蛋白λ或κ轻链轻链恒定区核酸序列的重排人免疫球蛋白λ或κ轻链可变区核酸序列。在一个实施方案中，轻链恒定区核酸序列包括大鼠恒定区核酸序列、人恒定区核酸序列或其组合。在一个实施方案中，免疫球蛋白λ或κ轻链恒定区核酸包括大鼠恒定区核酸序列、人恒定区核酸序列或其组合。

[0338] 在插入核酸内和/或整合在靶基因座处的目标多核苷酸可编码细胞外蛋白质或受体的配体。在特定实施方案中，编码的配体是细胞因子。目标细胞因子包括选自或包括CCL、CXCL、CX3CL和/或XCL的趋化因子。细胞因子也可包括肿瘤坏死因子(TNF)。在其它实施方案中，细胞因子是白介素(IL)。在一个实施方案中，白介素选自或包括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，白介素是IL-2。在特定实施方案中，所述在插入核酸内和/或整合在靶基因组基因座处的目标多核苷酸来自人，并且在更特定实施方案中，可包含人基因组序列。

[0339] 在插入核酸内和/或整合在靶基因组基因座处的目标多核苷酸可编码载脂蛋白E(ApoE)。

[0340] 在插入核酸内和/或整合在靶基因座处的目标多核苷酸可编码细胞质蛋白或膜蛋白。在一个实施方案中，膜蛋白是受体，如细胞因子受体、白介素受体、白介素2受体-α、白介素-2受体β、白介素-2受体γ或受体酪氨酸激酶。在其它情况下，在插入核酸内和/或整合在靶基因座处的目标多核苷酸可包含靶基因座的直系同源或同源区域。

[0341] 在插入核酸内和/或整合在靶基因座处的目标多核苷酸可包括编码T细胞受体(包括T细胞受体α)的至少某一区域的多核苷酸。在特定方法中，各插入核酸包含T细胞受体基因座(即T细胞受体α基因座)的基因组区域以使在完成连续整合后，基因组T细胞受体基因座的一部分或整体已被整合在靶基因座处。所述插入核酸可包含T细胞受体基因座(即T细胞受体α基因座)的至少一个或多个可变区段或接合区段。在其它实施方案中，编码T细胞受

体的区域的目标多核苷酸可来自例如编码突变蛋白质的哺乳动物、非人哺乳动物、啮齿动物、小鼠、大鼠、人、猴、农业哺乳动物或家养哺乳动物多核苷酸。

[0342] 在其它实施方案中，整合在靶基因座处的目标多核苷酸编码核蛋白质。在一个实施方案中，核蛋白质是核受体。在特定实施方案中，所述在插入核酸内和/或整合在靶基因座处的目标多核苷酸来自人，并且在更特定实施方案中，可包含人基因组序列。

[0343] 在插入核酸内和/或整合在靶基因组基因座处的目标多核苷酸可在编码序列中包含遗传修饰。所述遗传修饰包括但不限于编码序列的缺失突变或两个编码序列的融合。

[0344] 在插入核酸内和/或整合在靶基因座处的目标多核苷酸可包括编码突变蛋白质(包括例如人突变蛋白质)的多核苷酸。在一个实施方案中，突变蛋白质的特征在于结合特征改变、定位改变、表达改变和/或表达样式改变。在一个实施方案中，在插入核酸内和/或整合在靶基因座处的目标多核苷酸包含至少一个疾病等位基因，包括例如神经疾病的等位基因、心血管疾病的等位基因、肾病的等位基因、肌肉疾病的等位基因、血液疾病的等位基因、致癌基因的等位基因或免疫系统疾病的等位基因。在所述情况下，疾病等位基因可为显性等位基因，或疾病等位基因是隐性等位基因。此外，疾病等位基因可包括单核苷酸多态性(SNP)等位基因。编码突变蛋白质的目标多核苷酸可来自任何生物体，包括但不限于编码突变蛋白质的哺乳动物、非人哺乳动物、啮齿动物、小鼠、大鼠、人、猴、农业哺乳动物或家养哺乳动物多核苷酸。

[0345] 在一个实施方案中，遗传修饰产生蛋白质的具有改变的结合特征、改变的定位、改变的表达和/或改变的表达样式的突变形式。

[0346] 在一个实施方案中，遗传修饰产生大鼠ApoE基因座的区域的缺失、添加、替换或其组合，其中在ApoE基因座处的遗传修饰导致ApoE活性降低。在一个实施方案中，产生ApoE敲除。

[0347] 在一个实施方案中，遗传修饰产生大鼠Rag1基因座的区域的缺失、添加、替换或其组合，其中在Rag1基因座处的遗传修饰导致Rag1活性降低。在一个实施方案中，产生Rag1敲除。在一个实施方案中，遗传修饰产生大鼠Rag2基因座的区域的缺失、添加、替换或其组合，其中在Rag2基因座处的遗传修饰导致Rag2活性降低。在一个实施方案中，产生Rag2敲除。在一个实施方案中，遗传修饰产生大鼠Rag1/Rag2基因座的区域的缺失、添加、替换或其组合，其中在Rag1/Rag2基因座处的遗传修饰导致Rag1活性降低和Rag2活性降低。在一个实施方案中，产生Rag1/Rag2敲除。

[0348] 在一个实施方案中，遗传修饰产生大鼠白介素-2受体γ基因座的区域的缺失、添加、替换或其组合，其中在白介素-2受体γ基因座处的遗传修饰导致白介素-2受体γ降低。在一个实施方案中，产生白介素-2受体γ敲除。

[0349] 如本文其它地方所讨论，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、白介素-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大鼠ApoE基因座、白介素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座的一部分来修饰大鼠ApoE基因座、大鼠白介素-2受体γ基因座、Rag2基因座、Rag1基因座和/或Rag2/Rag1基因座中的一个或多个。

[0350] 在一个实施方案中，产生多个遗传修饰。在一个实施方案中，遗传修饰产生大鼠白介素-2受体γ基因座的区域的缺失、添加、替换或其组合，其中在白介素-2受体γ基因座处

的遗传修饰导致白介素-2受体γ降低，并且第二遗传修饰产生大鼠Rag2基因座的区域的缺失、添加、替换或其组合，其中在Rag2基因座处的遗传修饰导致Rag2活性降低。在一个实施方案中，产生白介素-2受体γ/Rag2敲除。所述大鼠具有SCID表型。

[0351] 在一个实施方案中，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，哺乳动物核酸包含编码在骨髓或骨髓源性细胞中表达的蛋白质的基因组基因座。在一个实施方案中，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，基因组基因座包含小鼠基因组DNA序列、大鼠基因组DNA序列、人基因组DNA序列或其组合。在一个实施方案中，基因组基因座以任何顺序包含大鼠和人基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含小鼠和人基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含小鼠和大鼠基因组DNA序列。在一个实施方案中，基因组基因座以任何顺序包含大鼠、小鼠和人基因组DNA序列。

[0352] 在一个实施方案中，插入核酸在基因的编码序列中包含遗传修饰。在一个实施方案中，遗传修饰包括编码序列中的缺失突变。在一个实施方案中，遗传修饰包括两个内源性编码序列的融合。

[0353] 在一个实施方案中，遗传修饰包括缺失非蛋白质编码序列，但不包括缺失蛋白质编码序列。在一个实施方案中，非蛋白质编码序列的缺失包括调控元件的缺失。在一个实施方案中，遗传修饰包括添加启动子。在一个实施方案中，遗传修饰包括替换启动子或调控元件。在一个实施方案中，调控元件是增强子。在一个实施方案中，调控元件是转录阻遏子结合元件。

[0354] 在一个实施方案中，遗传修饰包括放置编码突变人蛋白质的人核酸序列。在一个实施方案中，遗传修饰包括人基因的至少一个人疾病等位基因。在一个实施方案中，人疾病是神经疾病。在一个实施方案中，人疾病是心血管疾病。在一个实施方案中，人疾病是肾病。在一个实施方案中，人疾病是肌肉疾病。在一个实施方案中，人疾病是血液疾病。在一个实施方案中，人疾病是癌症。在一个实施方案中，人疾病是免疫系统疾病。在一个实施方案中，人疾病等位基因是显性等位基因。在一个实施方案中，人疾病等位基因是隐性等位基因。在一个实施方案中，人疾病等位基因包括单核苷酸多态性(SNP)等位基因。

[0355] 在插入核酸内和/或整合在靶基因座处的目标多核苷酸也可包含调控序列，包括例如启动子序列、增强子序列或转录阻遏子结合序列。在特定实施方案中，在插入核酸内和/或整合在靶基因组基因座处的目标多核苷酸包括具有非蛋白质编码序列的缺失，但不包含蛋白质编码序列的缺失的多核苷酸。在一个实施方案中，非蛋白质编码序列的缺失包括调控序列的缺失。在另一实施方案中，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，调控元件的缺失包括增强子序列的缺失。所述目标多核苷酸可来自任何生物体，包括但不限于编码突变蛋白质的哺乳动物、非人哺乳动物、啮齿动物、小鼠、大鼠、人、猴、农业哺乳动物或家养哺乳动物多核苷酸。

[0356] 5.引入序列以及产生转基因动物的方法

[0357] 如上所概述，本文提供用以允许使一个或多个目标多核苷酸靶向整合至靶基因座中的方法和组合物。所述系统采用多种组分，并且为便于引用，在本文中术语“靶向整合系

统”一般地包含为整合事件所需的所有组分(即在非限制性实施例中,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)。

[0358] 本文提供的方法包括向细胞中引入一种或多种包含靶向基因组整合系统的各种组分的多核苷酸或多肽构建体。“引入”意指以使序列进入细胞内部的方式向细胞递呈序列(多肽或多核苷酸)。本文提供的方法不取决于用于向细胞中引入靶向基因组整合系统的任何组分的特定方法,只是为了多核苷酸进入至少一个细胞的内部。用于将多核苷酸引入各种细胞类型中的方法在本领域中是已知的,并且包括但不限于稳定转染方法、短暂转染方法和病毒介导的方法。

[0359] 在一些实施方案中,用于方法和组合物中的细胞具有稳定并入它们的基因组中的DNA构建体。“稳定并入”或“稳定引入”意指向细胞中引入多核苷酸以使核苷酸序列整合至细胞基因组中,并且能够通过其子代遗传。任何方案都可用于稳定并入DNA构建体或靶向基因组整合系统的各种组分。

[0360] 转染方案以及用于将多肽或多核苷酸序列引入细胞中的方案可变化。非限制性转染方法包括基于化学的转染方法,包括使用脂质体;纳米粒子;磷酸钙(Graham等(1973).Virology 52(2):456-67;Bacchetti等(1977)Proc Natl Acad Sci USA 74(4):1590-4;以及Kriegler,M(1991).Transfer and Expression:A Laboratory Manual.New York:W.H.Freeman and Company.第96-97页);树枝状聚合物;或阳离子聚合物,如DEAE-葡聚糖或聚乙烯亚胺。非化学方法包括电穿孔;声致穿孔;和光学转染。基于粒子的转染包括使用基因枪、磁体辅助转染(Bertram,J.(2006)Current Pharmaceutical Biotechnology 7,277-28)。病毒方法也可用于转染。

[0361] 在一个实施方案中,一个或多个多核苷酸向细胞中的引入通过电穿孔、细胞质内注射、病毒性感染、腺病毒、慢病毒、逆转录病毒、转染、脂质介导的转染来介导,或通过Nucleofection™来介导。

[0362] 在一个实施方案中,向细胞中引入一个或多个多核苷酸进一步包括:引入包含可操作地连接至启动子的目标核酸序列的表达构建体。在一个实施方案中,启动子是组成型活性启动子。在一个实施方案中,启动子是诱导型启动子。在一个实施方案中,启动子在大鼠胚胎干细胞中具有活性。

[0363] 在一个实施方案中,连同LTVEC一起引入表达构建体。在一个实施方案中,历经一段时期与LTVEC分开引入表达构建体。

[0364] 在一个实施方案中,可历经一段时期多次进行一个或多个多核苷酸向细胞中的引入。在一个实施方案中,历经一段时期至少两次;历经一段时期至少三次;历经一段时期至少四次;历经一段时期至少五次;历经一段时期至少六次;历经一段时期至少七次;历经一段时期至少八次;历经一段时期至少九次;历经一段时期至少十次;至少十一次;历经一段时期至少十二次;历经一段时期至少十三次;历经一段时期至少十四次;历经一段时期至少十五次;历经一段时期至少十六次;历经一段时期至少十七次;历经一段时期至少十八次;历经一段时期至少十九次;或历经一段时期至少二十次进行一个或多个多核苷酸向细胞中的引入。

[0365] 在一个实施方案中,核酸酶试剂与靶向载体或大靶向载体(LTVEC)同时引入细胞中。或者,历经一段时期,与靶向载体或LTVEC分开引入核酸酶试剂。在一个实施方案中,在

引入靶向载体或LTVEC之前引入核酸酶试剂，而在其它实施方案中，在引入靶向载体或LTVEC之后引入核酸酶试剂。

[0366] 在一个实施方案中，筛选步骤包括用于评估亲本染色体的等位基因修饰(MOA)的定量测定。在一个实施方案中，定量测定通过定量PCR来进行。在一个实施方案中，定量PCR是实时PCR(qPCR)。在一个实施方案中，实时PCR包括识别靶基因座的第一引物组和识别非靶向参照基因座的第二引物组。在一个实施方案中，引物组包含识别扩增的序列的荧光探针。在一个实施方案中，定量测定通过荧光介导的原位杂交(FISH)来进行。在一个实施方案中，定量测定通过比较基因组杂交来进行。在一个实施方案中，定量测定通过等温DNA扩增来进行。在一个实施方案中，定量测定通过等温DNA扩增来进行。在一个实施方案中，定量测定通过定量杂交于固定的探针来进行。在一个实施方案中，定量测定通过Invader Probes®来进行。在一个实施方案中，定量测定通过MMP assays®来进行。在一个实施方案中，定量测定通过TaqMan®分子信标来进行。在一个实施方案中，定量测定通过Eclipse™探测技术来进行。(参见例如US2005/0144655，其以引用的方式整体并入本文)。

[0367] 进一步提供一种用于制备人源化大鼠的方法，其包括：(a)用包含插入核酸的靶向载体修饰多能大鼠细胞的基因组以形成供体细胞，所述插入核酸包含人核酸序列；(b)将所述供体细胞引入宿主大鼠胚胎中；以及(c)在代孕母体中孕育所述宿主大鼠胚胎；其中所述代孕母体产生包含所述人核酸序列的大鼠子代。在一个实施方案中，将供体细胞引入处于胚泡阶段或处于前桑椹胚阶段(即4细胞阶段或8细胞阶段)的宿主大鼠胚胎中。此外，步骤(a)也可用大靶向载体(LTVEC)和/或长度是至少5Kb的人核酸序列进行。在其它实施方案中，遗传修饰能够通过种系传递。

[0368] 可采用本文公开的各种方法产生经遗传修饰的大鼠。所述方法包括(1)采用本文公开的方法在多能大鼠细胞的靶基因座处整合一个或多个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经遗传修饰的多能大鼠细胞；(2)选择在靶基因组基因座处具有一个或多个目标多核苷酸的经遗传修饰的多能大鼠细胞；(3)将经遗传修饰的多能大鼠细胞引入大鼠宿主胚胎中；以及(4)将包含经遗传修饰的多能大鼠细胞的宿主大鼠胚胎植入代孕母体中。产生来自经遗传修饰的多能大鼠细胞的子代。在一个实施方案中，将供体细胞引入处于胚泡阶段或处于前桑椹胚阶段(即4细胞阶段或8细胞阶段)的大鼠宿主胚胎中。产生能够通过种系传递遗传修饰的子代。多能大鼠细胞可为如本文其它地方讨论的大鼠ES细胞。

[0369] 核转移技术也可用于产生经遗传修饰的大鼠。简要来说，用于核转移的方法包括以下步骤：(1)使卵母细胞去核；(2)分离待与去核卵母细胞组合的供体细胞或核；(3)将细胞或核插入去核卵母细胞中以形成重构细胞；(4)将重构细胞植入动物的子宫中以形成胚胎；以及(5)使胚胎发育。在所述方法中，卵母细胞通常从死亡动物撷取，但它们也可自活动物的输卵管和/或卵巢分离。可在去核之前使卵母细胞在为本领域普通技术人员所知的多种培养基中成熟。可以为本领域普通技术人员所熟知的许多方式进行卵母细胞去核。通常通过在融合之前将供体细胞显微注射在透明带下来将供体细胞或核插入去核卵母细胞中以形成重构细胞。融合可通过跨越接触/融合平面施加DC电脉冲(电融合)；通过使细胞暴露于融合促进化学物质(如聚乙二醇)；或通过失活病毒(如仙台病毒(Sendai virus))的方式

来诱导。重构细胞通常在融合核供体和接受者卵母细胞之前、期间和/或之后通过电学和/或非电学手段来活化。活化方法包括电脉冲、化学诱导冲击、通过精子来渗透、增加卵母细胞中的二价阳离子的水平、以及降低卵母细胞中的细胞蛋白质的磷酸化(如通过激酶抑制剂的方式)。活化的重构细胞或胚胎通常在为本领域普通技术人员所熟知的培养基中培养,接着转移至动物的子宫中。参见例如US20080092249、WO/1999/005266A2、US20040177390、WO/2008/017234A1和美国专利No.7,612,250,其各自以引用的方式并入本文。

[0370] 在一个方面,提供一种用于制备经遗传修饰的大鼠的方法,其包括采用核酸内切酶介导的基因靶向来修饰多能大鼠细胞中的目标基因组基因座以在目标大鼠基因组基因座处引入修饰来形成修饰多能大鼠细胞,在足以维持多能性的条件下维持修饰多能大鼠细胞,采用修饰多能大鼠细胞作为大鼠宿主胚胎中的供体细胞,以及在代孕母体中孕育包含经修饰的多能大鼠细胞的宿主胚胎,其中宿主胚胎由代孕母体孕育,并且生育经遗传修饰的大鼠子代。

[0371] 在一个实施方案中,靶序列位于内含子中。在一个实施方案中,靶序列位于外显子中。在一个实施方案中,靶序列位于启动子中。在一个实施方案中,靶序列位于启动子调控区中。在一个实施方案中,靶序列位于增强子区域中。

[0372] 在一个实施方案中,历经一段时期使用识别不同靶序列的多种核酸内切酶多次进行引入步骤。在一个实施方案中,历经一段时期使用识别不同靶序列的多种核酸内切酶至少两次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少三次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少四次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少五次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少六次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少七次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少八次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少九次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十一次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十二次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十三次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十四次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十五次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十六次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十七次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十八次;历经一段时期使用识别不同靶序列的多种核酸内切酶至少十九次;或历经一段时期使用识别不同靶序列的多种核酸内切酶至少二十次进行步骤。

[0373] 在一个实施方案中,引入步骤通过电穿孔、细胞质内注射、腺病毒、慢病毒、逆转录病毒、转染、脂质介导的转染来介导,或通过Nucleofection™来介导。

[0374] 在一个实施方案中,方法进一步包括将外源性核酸引入经遗传修饰的多能大鼠细胞中。在一个实施方案中,外源性核酸是转基因。在一个实施方案中,将外源性核酸引入内源性基因座中。在一个实施方案中,外源性核酸被异位引入(例如在不同于它的内源性基因座的基因座处)。

[0375] 在一个方面,提供一种用于制备经遗传修饰的大鼠的方法,其包括采用RNA导向基因组工程化修饰多能大鼠细胞中的目标基因组基因座以在目标大鼠基因组基因座处引入

修饰来形成经修饰的多能大鼠细胞，在足以维持多能性的条件下维持经修饰的多能大鼠细胞，采用经修饰的多能大鼠细胞作为大鼠宿主胚胎中的供体细胞，以及在代孕母体中孕育包含经修饰的多能大鼠细胞的宿主胚胎，其中宿主胚胎由代孕母体孕育，并且生育经遗传修饰的大鼠子代。

[0376] 在一个实施方案中，方法具有在约2％至约80％的范围内的靶向率。

[0377] 在一个实施方案中，方法包括共同引入多个包含不同基因组靶序列以对不同基因组基因座进行多路编辑的第二表达构建体。在一个实施方案中，方法包括历经一段时期引入多个包含不同基因组靶序列以对不同基因组基因座进行多路编辑的第二表达构建体。

[0378] 在一个实施方案中，历经一段时期多次进行引入步骤。在一个实施方案中，历经一段时期至少两次；历经一段时期至少三次；历经一段时期至少四次；历经一段时期至少五次；历经一段时期至少六次；历经一段时期至少七次；历经一段时期至少八次；历经一段时期至少九次；历经一段时期至少十次；历经一段时期至少十一次；历经一段时期至少十二次；历经一段时期至少十三次；历经一段时期至少十四次；历经一段时期至少十五次；历经一段时期至少十六次；历经一段时期至少十七次；历经一段时期至少十八次；历经一段时期至少十九次；历经一段时期至少二十次进行引入步骤(b)。

[0379] 在一个实施方案中，第一表达构建体和第二表达构建体从同一质粒表达。

[0380] 在一个实施方案中，引入步骤通过电穿孔、细胞质内注射、腺病毒、慢病毒、逆转录病毒、转染、脂质介导的转染来介导，或通过Nucleofection™来介导。

[0381] 在一个实施方案中，方法进一步包括将外源性核酸引入包含突变等位基因的多能大鼠细胞中。

[0382] 在一个实施方案中，外源性核酸是转基因。在一个实施方案中，将外源性核酸引入内源性基因座中。在一个实施方案中，外源性核酸被异位放置(例如在不同于它的内源性基因座的基因座处)。

[0383] 在一个实施方案中，方法进一步包括将外源性核酸引入经遗传修饰的多能大鼠细胞中。在一个实施方案中，外源性核酸是转基因。在一个实施方案中，将外源性核酸引入内源性基因座中。在一个实施方案中，外源性核酸被异位引入(例如在不同于它的内源性基因座的基因座处)。

[0384] 在一个方面，提供一种用于制备人源化大鼠的方法，其包括用包含插入物(其包含至少5kb的人序列)的LTVEC修饰多能大鼠细胞的基因组，以及采用多能大鼠细胞作为供体细胞，将供体细胞引入宿主胚胎中，以及在代孕母体中孕育宿主胚胎，其中代孕母体分娩包含人源化的大鼠子代。

[0385] 提供其它用于制备在它的种系中包含一种或多种如本文所述的遗传修饰的经遗传修饰的大鼠的方法，其包括：(a)采用本文所述的各种方法修饰原核细胞中含有的靶向大鼠基因座；(b)选择在靶向大鼠基因座处包含遗传修饰的修饰的原核细胞；(c)自修饰的原核细胞的基因组分离遗传修饰的靶向载体；(d)将遗传修饰的靶向载体引入多能大鼠细胞中以产生在靶向基因组基因座处包含插入核酸的遗传修饰的多能细胞；(e)选择经遗传修饰的大鼠多能细胞；(f)将经遗传修饰的多能大鼠细胞引入处于前桑椹胚阶段的宿主大鼠胚胎中；以及(g)将包含遗传修饰的多能大鼠细胞的宿主大鼠胚胎植入代孕母体中以产生源于经遗传修饰的多能大鼠细胞的F0代。在所述方法中，靶向载体可包括大靶向载体。多能

大鼠细胞可为大鼠ES细胞。在其它方法中，分离步骤(c)进一步包括(c1)使遗传修饰靶向载体(即遗传修饰LTVEC)线性化。在其它实施方案中，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能大鼠细胞中。在一个实施方案中，通过向原核细胞或多能大鼠细胞施用如本文所述的可选择试剂来进行选择步骤(b)和/或(e)。在一个实施方案中，通过如本文所述的等位基因修饰(MOA)测定来进行选择步骤(b)和/或(e)。

[0386] 提供其它用于通过原核细胞中的细菌同源重组(BHR)来修饰哺乳动物细胞的靶基因组基因座的方法，并且包括：(a)提供包含靶基因座的原核细胞，所述靶基因座包含大鼠核酸，(b)向原核细胞中引入包含侧接有5'大鼠同源臂和3'大鼠同源臂的插入核酸的靶向载体，其中插入核酸包含哺乳动物区域(包括例如来自人的DNA插入物)，以及(c)选择在靶大鼠基因座处包含插入核酸的靶向原核细胞，其中原核细胞能够表达介导BHR的重组酶。步骤(a1)可包括提供包含靶基因座的原核细胞，所述靶基因座包含大鼠核酸，所述大鼠核酸包含第一多核苷酸，所述第一多核苷酸包含第一核酸酶试剂的第一识别位点，并且步骤(b1)可进一步包括在原核细胞中表达在第一识别位点处或附近产生切口或双链断裂的核酸酶试剂。步骤(a)-(c)可如本文所公开来连续重复以使得在原核细胞中在靶向大鼠基因座处引入多个插入核酸。一旦用原核细胞“构建”靶向基因组基因座，即可从原核细胞分离包含经修饰的靶大鼠基因座的靶向载体，并且引入多能大鼠细胞内的靶基因组基因座中。可接着将包含经修饰的基因组基因座的多能大鼠细胞(即大鼠ES细胞)制备成经遗传修饰的大鼠。

[0387] 在一些实施方案中，对本文所述的靶基因组基因座的各种遗传修饰可通过使用利用VELOCIGENE®遗传工程技术从细菌人工染色体(BAC)DNA获得的LTVEC，在细菌细胞中进行一系列同源重组反应(BHR)来进行(参见例如美国专利No.6,586,251以及Valenzuela,D.M.等(2003),High-throughput engineering of the mouse genome coupled with high-resolution expression analysis,Nature Biotechnology21(6):652-659,其以引用的方式整体并入本文)。

[0388] 在一些实施方案中，包含如本文所述的各种遗传修饰的靶向大鼠ES细胞用作插入ES细胞，并且通过VELOCIMOUSE®方法被引入来自相应生物体的前桑椹胚阶段胚胎(例如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US2008-0078000A1，其全部以引用的方式整体并入本文)。孵育包含经遗传修饰的大鼠ES细胞的大鼠胚胎直至胚泡阶段，接着植入代孕母体中以产生F0。可通过如本文所述的等位基因修饰(MOA)测定来鉴定携带遗传修饰基因组基因座的大鼠。使源于遗传修饰ES大鼠细胞的所得F0代大鼠与野生型大鼠杂交以获得F1代后代。在用特定引物和/或探针分析基因型之后，使遗传修饰基因组基因座杂合性F1大鼠彼此杂交以产生遗传修饰基因组基因座纯合性大鼠。或者，可使各自具有遗传修饰的F0雌性大鼠和F0雄性大鼠杂交以获得遗传修饰纯合性F1大鼠。

[0389] 在一个方面，提供一种经遗传修饰的大鼠基因组，其包含内源性大鼠核酸序列被同源或直系同源非大鼠核酸序列的靶向修饰。

[0390] 在一个实施方案中，同源或直系同源非大鼠核酸序列具有约5kb至约200kb的长度。在一个实施方案中，同源或直系同源非大鼠核酸序列在约5kb至约10kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约10kb至约20kb的范围内。在一个实施

方案中，同源或直系同源非大鼠核酸序列在约20kb至约30kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约30kb至约40kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约40kb至约50kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约50kb至约60kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约60kb至约70kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约70kb至约80kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约80kb至约90kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约90kb至约100kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约100kb至约110kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约110kb至约120kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约120kb至约130kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约140kb至约150kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约150kb至约160kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约160kb至约170kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约170kb至约180kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约180kb至约190kb的范围内。在一个实施方案中，同源或直系同源非大鼠核酸序列在约190kb至约200kb的范围内。可用于插入核酸中的各种目标多核苷酸在本文其它地方描述。

[0391] 6.细胞

[0392] 本文所述的各种方法和组合物在细胞中采用基因组基因座靶向系统。在一个实施方案中，细胞是多能细胞。在一个实施方案中，多能细胞是非人多能细胞。在一个实施方案中，非人多能细胞是哺乳动物多能细胞。在一个实施方案中，多能细胞是人诱导多能干(iPS)细胞。

[0393] 在一个实施方案中，多能细胞是多能大鼠细胞。在一个实施方案中，多能大鼠细胞是大鼠胚胎干(ES)细胞。在一个实施方案中，多能大鼠细胞是诱导多能干(iPS)细胞或是发育限制的祖细胞。在其它实施方案中，多能大鼠细胞能够在对它的基因组进行至少一次靶向遗传修饰之后持续它的多能性，并且能够向F1代的种系传递靶向修饰。

[0394] 在一个实施方案中，多能细胞是处于单细胞阶段的非人受精卵。在一个实施方案中，非人受精卵是哺乳动物受精卵。在一个实施方案中，哺乳动物受精卵是处于单细胞阶段的啮齿动物受精卵。在一个实施方案中，哺乳动物受精卵是处于单细胞阶段的大鼠或小鼠受精卵。

[0395] 用于本文公开的方法和组合物中的各种细胞也可包括原核细胞，如细菌细胞，包括大肠杆菌。在特定实施方案中，原核细胞是大肠杆菌的重组感受态菌株。在一个实施方案中，原核细胞包含编码重组酶的核酸，而在其它情况下，原核细胞不包含编码重组酶的核酸，并且编码重组酶的核酸被引入原核细胞中。在一个实施方案中，编码重组酶的核酸包括DNA或mRNA。在一些实施方案中，编码重组酶的核酸是pABG。在一个实施方案中，重组酶在诱导型启动子控制下表达。在一个实施方案中，重组酶的表达受阿拉伯糖控制。

[0396] A.大鼠胚胎干(ES)细胞

[0397] 如以上详细概述，本文提供的各种组合物和方法可采用来自大鼠的胚胎干(ES)细胞。在一个实施方案中，多能大鼠细胞是大鼠ES细胞。在一个实施方案中，大鼠ES细胞源于

是Wistar大鼠、LEA品系、Sprague Dawley品系、Fischer品系、F344、F6和Dark Agouti或ACI的大鼠品系。在一个实施方案中，大鼠品系是两个或更多个选自由以下组成的组的品系的混合：Wistar、LEA、Sprague Dawley、Fischer、F344、F6和Dark Agouti。在一个实施方案中，大鼠ES细胞源于近交品系。在一个实施方案中，大鼠ES细胞源于选自DA品系和ACI品系的品系。在一特定实施方案中，大鼠ES细胞源于ACI品系。在一个实施方案中，大鼠ES细胞源于大鼠胚泡。

[0398] 在其它实施方案中，大鼠ES细胞的特征在于表达至少一种多能性标志物。在特定实施方案中，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能性标志物。在一个实施方案中，大鼠ES细胞是雄性(XY)大鼠ES细胞或雌性(XX)大鼠ES细胞。

[0399] 在一个实施方案中，在1至15次连续遗传修饰之后，经遗传修饰的大鼠ES细胞在暴露于分化培养基后能够分化成多种细胞类型。

[0400] 在一个实施方案中，在1至15次连续遗传修饰之后，经遗传修饰的大鼠ES细胞能够在培养时被维持未分化状态。在一个实施方案中，呈未分化状态的遗传修饰以及培养的大鼠ES细胞在用作大鼠宿主胚胎中的供体细胞时在胚胎中定殖并且形成包含1至15个遗传修饰的胚泡。在一个实施方案中，胚泡在适于孕育的条件下植入代孕母体中时发育成包含1至15个遗传修饰的F0大鼠子代。

[0401] 在一个方面，提供一种能够在一次或多次体外遗传修饰之后持续多能性，并且能够向F1代的种系传递遗传修饰基因组的分离的大鼠ES细胞。

[0402] 在一个实施方案中，大鼠ES细胞在一次或多次连续体外遗传修饰(例如两次、三次、四次、五次或六次或大于六次连续遗传修饰)之后维持它的发育成多种细胞类型的多能性。在一个实施方案中，遗传修饰通过电穿孔、细胞质内注射、病毒性感染、腺病毒、慢病毒、逆转录病毒、转染、脂质介导的转染、或Nucleofaction™来介导。

[0403] 在一个实施方案中，大鼠ES细胞在用外源性核酸进行单轮电穿孔之后维持它的发育成多种细胞类型的多能性。在一个实施方案中，大鼠ES细胞在用外源性核酸进行第2、第3、第4、第5、第6、第7、第8、第9、第10、第11、第12、第13、第14或第15轮电穿孔之后维持它的发育成多种细胞类型的多能性。

[0404] 在其它实施方案中，采用的大鼠ES细胞是2014年2月20日提交，并且以引用的方式整体并入本文的美国申请号14/185,703中所述的那些。

[0405] 用于本文公开的各种方法和组合物中的多能大鼠细胞的特征可在于表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1和/或其组合。在其它情况下，用于本文公开的各种方法和组合物中的多能大鼠细胞的特征在于以下特征中的一个或多个：(a)缺乏一种或多种包括c-Myc、Ecat1和/或Rexo1的多能性标志物的表达；(b)缺乏一种或多种包括Brachyury和/或Bmpr2的中胚层标志物的表达；(c)缺乏一种或多种包括Gata6、Sox17和/或Sox7的内胚层标志物的表达；或(d)缺乏一种或多种包括巢蛋白和/或Pax6的神经标志物的表达。如本文所用，“缺乏表达”在它涉及多能性标志物的表达时意指多能性标志物的表达处于或低于如对于各个别实验测定的实验本底。

[0406] 在一个非限制性实施方案中，本文提供的大鼠ES细胞具有任何以下性质中的一个或多个：

[0407] (a)具有种系胜任性，意味着当将大鼠ES细胞植入大鼠宿主胚胎中时，大鼠ES细胞系的基因组向后代中传递；

[0408] (b)在至少一次靶向遗传修饰之后具有种系胜任性，意味着当将具有靶向遗传修饰的大鼠ES细胞植入大鼠宿主胚胎中时，大鼠ES细胞系的基因组内的靶向遗传修饰向后代中传递；

[0409] (c)具有体外多能性；

[0410] (d)具有体外全能性；

[0411] (e)当体外培养时松散粘附于饲养细胞层；

[0412] (f)当体外培养时，当体外涂铺在饲养细胞层上时形成球样集落；

[0413] (g)当在包括未被遗传修饰以表达白血病抑制因子(LIF)的饲养细胞层的条件下体外培养时维持多能性，其中培养基包含足够浓度的LIF；

[0414] (h)当在包括饲养细胞层的条件下体外培养时维持多能性，其中培养基包含小鼠LIF或其活性变体或片段；

[0415] (i)包含特征在于以下的分子标签

[0416] i)表达一种或多种大鼠ES细胞特异性基因，包括粘附接合相关蛋白(Ajap1)、密封蛋白(Claudin)5(Cldn5)、Cdc42鸟嘌呤核苷酸交换因子9(Arhgef9)、钙/钙调蛋白依赖性蛋白质激酶IV(Camk4)、肝配蛋白(ephrin)-A1(Efna1)、EPH受体A4(Epha4)、间隙连接蛋白β5(Gjb5)、胰岛素样生长因子结合蛋白样1(Igfbpl1)、白介素36β(Il1f8)、白介素28受体α(Il28ra)、左-右决定因子1(Lefty1)、白血病抑制因子受体α(Lifr)、溶血磷脂酸受体2(Lpar2)、神经元穿透素(pentraxin)受体(Ntm)、蛋白质酪氨酸磷酸酶非受体类型18(Ptpn18)、尾型同源框2(Cdx2)、III型纤连蛋白和锚蛋白重复结构域1(Fank1)、叉头框E1(甲状腺转录因子2)(Foxe1)、发状/分裂增强子相关含YRPW基序2(Hey2)、叉头框E1(甲状腺转录因子2)(Foxe1)、发状/分裂增强子相关含YRPW基序2(Hey2)、淋巴增强子结合因子1(Lef1)、Sal样3(果蝇)(Sall3)、SATB同源框1(Satb1)、miR-632或其组合；

[0417] ii)表达包括以下的大鼠ES细胞特异性基因中的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25个或大于25个：粘附接合相关蛋白(Ajap1)、密封蛋白5(Cldn5)、Cdc42鸟嘌呤核苷酸交换因子9(Arhgef9)、钙/钙调蛋白依赖性蛋白质激酶IV(Camk4)、肝配蛋白-A1(Efna1)、EPH受体A4(Epha4)、间隙连接蛋白β5(Gjb5)、胰岛素样生长因子结合蛋白样1(Igfbpl1)、白介素36β(Il1f8)、白介素28受体α(Il28ra)、左-右决定因子1(Lefty1)、白血病抑制因子受体α(Lifr)、溶血磷脂酸受体2(Lpar2)、神经元穿透素受体(Ntm)、蛋白质酪氨酸磷酸酶非受体类型18(Ptpn18)、尾型同源框2(Cdx2)、III型纤连蛋白和锚蛋白重复结构域1(Fank1)、叉头框E1(甲状腺转录因子2)(Foxe1)、发状/分裂增强子相关含YRPW基序2(Hey2)、叉头框E1(甲状腺转录因子2)(Foxe1)、发状/分裂增强子相关含YRPW基序2(Hey2)、淋巴增强子结合因子1(Lef1)、Sal样3(果蝇)(Sall3)、SATB同源框1(Satb1)、miR-632或其组合；

[0418] iii)当相较于F1H4小鼠ES细胞时，如表9中阐述的一种或多种大鼠ES细胞特异性基因的表达增加至少20倍；

[0419] iv)当相较于F1H4小鼠ES细胞时，如表9中阐述的大鼠ES细胞特异性基因中的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25个或大于25个的

表达增加至少20倍；

[0420] v)表达如表10中阐述的一种或多种大鼠ES细胞特异性基因；

[0421] vi)表达如表10中阐述的大鼠ES细胞特异性基因中的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50个或大于50个；

[0422] vii)当相较于F1H4小鼠ES细胞时，如表10中阐述的一种或多种大鼠ES细胞特异性基因的表达增加至少20倍；

[0423] viii)当相较于F1H4小鼠ES细胞时，如表10中阐述的大鼠ES细胞特异性基因中的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50个或大于50个的表达增加至少20倍；

[0424] ix)当相较于F1H4小鼠ES细胞时，如表8中阐述的一种或多种大鼠ES细胞特异性基因的表达降低至少20倍；

[0425] x)当相较于F1H4小鼠ES细胞时，如表8中阐述的大鼠ES细胞特异性基因中的至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、30、35、40、45、50个或大于50个的表达降低至少20倍；

[0426] xi)部分(i)-(x)的大鼠ES细胞特异性基因的表达的任何组合；

[0427] xii)至少2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17或18种所列多能性标志物具有如表11中所示的多能性标志物的相对表达水平。参见表11的针对相对表达水平的多能性分级列；

[0428] xiii)至少2、3或4种所列中胚层标志物具有如表11中所示的中胚层标志物的相对表达水平。参见表11中针对相对表达水平的中胚层分级列；

[0429] xiv)至少2、3、4、5或6种所列内胚层标志物具有如表11中所示的内胚层标志物的相对表达水平。参见表11中针对相对表达水平的内胚层分级列；

[0430] xv)至少2和3种所列神经标志物具有如表11中所示的神经标志物的相对表达水平。参见表11中针对相对表达水平的神经分级列；

[0431] xvi)所列滋养外胚层标志物具有如表11中所示的滋养外胚层标志物的相对表达水平。参见表11中针对相对表达水平的滋养外胚层分级列；

[0432] xvii)具有表11中阐述的多能性标志物、中胚层标志物、内胚层标志物、神经标志物和/或滋养外胚层标志物中的一个或多个(2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29或30个)的任何相对表达水平；

[0433] xviii)具有表11中阐述的各标志物的相对表达水平；

[0434] xix)xii-xiix中阐述的标签的任何组合；和/或

[0435] xx)i-xiix中阐述的标签的任何组合；

[0436] (j)能够产生F0大鼠；

[0437] (k)能够次培养并维持未分化状态；

[0438] (l)与正常大鼠细胞具有相同染色体数目；

[0439] (m)在无需旁分泌LIF信号传导下维持体外多能性；

[0440] (n)具有自我更新性，意味着它们在维持多能性下无限分裂；

[0441] (o)大鼠ES细胞表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1和/或其组

合；

[0442] (p)大鼠ES细胞不表达一种或多种包括c-Myc、Ecat1和/或Rex01的分化标志物；

[0443] (q)大鼠ES细胞不表达一种或多种包括Brachyury、Bmpr2和/或其组合的中胚层标志物；

[0444] (r)大鼠ES细胞不表达一种或多种包括Gata6、Sox17、Sox7和/或其组合的内胚层标志物；和/或

[0445] (s)大鼠ES细胞不表达一种或多种包括巢蛋白、Pax6和/或其组合的神经标志物。

[0446] (a)-(s)中概述的一种或多种特征可存在于用于本文提供的方法和组合物中的大鼠ES细胞、大鼠ES细胞群体或大鼠ES细胞系中，其中大鼠ES细胞尚未经受靶向遗传修饰。此外，在如上详细所述对大鼠靶基因座进行一次或多次遗传修饰之后，(a)-(s)中概述的一种或多种特征可在靶基因座的遗传修饰之后保留在大鼠ES细胞中。

[0447] 在一个实施方案中，大鼠ES细胞展现至少2％、至少3％、至少4％、至少5％、至少6％、至少7％、至少8％、至少9％、至少10％、至少11％、至少12％、至少13％、至少14％、至少15％、至少20％、至少25％、至少30％、至少35％、至少40％、至少45％、至少50％、至少55％、至少60％、至少65％、至少70％、至少75％、或至少80％的同源重组效率。

[0448] 在一个实施方案中，采用大鼠ES细胞的同源重组效率大于4％。

[0449] 在一个实施方案中，大鼠ES细胞具有在24小时至36小时的范围内的倍增时间。在一个实施方案中，大鼠ES细胞具有25小时的倍增时间。

[0450] 在一个实施方案中，大鼠ES细胞可在2i培养基(Millipore目录SF016-200)中传代多达至少15次。在一个实施方案中，大鼠ES细胞可在2i培养基(Millipore目录号SF016-200)中传代至少14次。在一个实施方案中，大鼠ES细胞可在2i培养基中传代至少13、12、11、10、9、8、7、6或5次。

[0451] 在一个实施方案中，当移植至前桑椹胚阶段大鼠胚胎中时，大鼠ES细胞可促成F0代中至少90％的细胞。在一个实施方案中，当移植至前桑椹胚阶段大鼠胚胎中时，大鼠ES细胞可促成F0代中至少95％、96％、97％、98％或99％的细胞。

[0452] 在特定实施方案中，用于本文提供的各种方法和组合物中的各种大鼠ES细胞和细胞系用于在靶基因座处产生靶向修饰。具有这些靶向遗传修饰的大鼠ES细胞可具有种系胜任性，意味着当将具有靶向遗传修饰的大鼠ES细胞植入大鼠宿主胚胎中时，大鼠ES细胞的靶向遗传修饰被传递至后代(即F1群体)中。因此，在各个方面，各种方法和组合物中的大鼠ES细胞用于获得来自已经受靶向遗传修饰的大鼠ES细胞的大鼠细胞基因组的高频率或高效率种系传递。在各种实施方案中，种系传递的频率大于1:600、大于1:500、大于1:400、大于1:300、大于1:200以及大于1:100。在各种实施方案中，种系传递的频率大于1％、大于2％、大于3％、大于4％、大于5％、大于6％、大于7％、大于8％、大于9％、大于10％、多达约16％、大于25％、大于50％、大于60％、大于65％、大于70％、大于75％或更大。在各种实施方案中，种系传递的频率在9％至16％的范围内。在各个方面，F1代中供体rESC源性子代的百分比是1％或大于1％、2％或大于2％、3％或大于3％、10％或大于10％、20％或大于20％、30％或大于30％、40％或大于40％、50％或大于50％、60％或大于60％、3％至约10％或大于10％；3％或大于3％至约63％、约10％至约30％、约10％至约50％、约30％至约70％、约30％至约60％、约20％至约40％、约20％至65％、或约40％至70％。因此，具有靶向遗传修饰的大

鼠ES细胞能够向F1群体中传递它们的基因组。

[0453] 具有靶向遗传修饰的大鼠ES细胞可为多能的和/或全能的。各种方法可用于确定大鼠ES细胞是否是多能的。举例来说，可测定ES细胞的各种多能标志物的表达，所述标志物包括但不限于Oct-4、Sox2、碱性磷酸酶或其组合。对于测定所述标志物的存在性或水平的各种方法，参见例如Okamoto,K.等,Cell,60:461-472(1990);Scholer,H.R.等,EMBO J.9:2185-2195(1990))以及Nanog(Mitsui,K.等,Cell,113:631-642(2003);Chambers,I.等,Cell,113:643-655(2003)。也参见本文提供的图2和3。其它多能性标志物包括例如存在至少1、2、3、4或5种包括Nanog、Klf4、Dppa2、Fgf4、Rex1、Eras、Err-β和/或Sall3的多能性标志物。其它多能性标志物包括例如不存在至少1、2、3、4、5、6、7、8、9或10种包括T/Brachyury、Flk1、Nodal、Bmp4、Bmp2、Gata6、Sox17、Hhex1、Sox7和/或Pax6的多能性标志物。

[0454] 在特定实施方案中，可使用RT-PCR测定这些标志物的表达和/或表达水平。各种试剂盒可用于测定碱性磷酸酶的水平和/或存在，包括例如ALP组织染色试剂盒(Sigma)和Vector红色碱性磷酸酶底物试剂盒I(Funakoshi)等。其它测定包括原位杂交、免疫组织化学、免疫荧光。在特定实施方案中，大鼠ES细胞的特征在于表达至少一种多能性标志物，包括例如表达Oct-4、Sox2、碱性磷酸酶或其组合，并且优选表达这些标志物中的全部三种。

[0455] 用于本文提供的方法和组合物中的各种大鼠ES细胞能够在于体外培养条件下维持时维持多能性和/或全能性。因此，在一些实施方案中，本文提供的各种大鼠ES细胞可被次培养，同时仍然维持未分化状态。培养大鼠ES细胞的各种方法在本文其它地方以及在2014年2月20日提交，以引用的方式整体并入本文的美国专利申请号14/185,103中进一步详细讨论。

[0456] 在一些实施方案中，本文采用的大鼠胚胎干细胞已采用各种分离、纯化和培养扩展技术从大鼠胚胎分离，所述技术详细讨论于2014年2月20日提交，以引用的方式整体并入本文的美国专利申请号14/185,103中的美国申请中。

[0457] “分离的”大鼠ES细胞或大鼠胚胎已从它的天然环境移除。术语“分离的”可意指不含某一组分以它的天然状态与其一起被发现的组分的70％、80％、90％、95％、96％、97％、98％或99％。如本文所用，大鼠ES“细胞系”包含由单一大鼠ES细胞产生的分离的大鼠细胞的群体，并且因此除在繁殖期间或在靶向遗传修饰期间发生的突变或核型变化以外，给定细胞系内的细胞群体具有均一遗传组成。举例来说，大鼠ES细胞的特征可在于具有高整倍性水平。然而，在一些细胞系中，由于种系从单细胞繁殖时的核型变化，整倍性水平小于100％。此外，大鼠ES细胞的给定群体可包含至少1×10^3、1×10^4、1×10^5、1×10^6、1×10^7、1×10^8、1×10^9或1×10^{10}个细胞或大于1×10^{10}个细胞。一些细胞群体具有足以允许选择所需修饰细胞，而非过度更大数目的细胞，以便降低突变或核型变化产生在细胞系中的可能性。举例来说，一些细胞群体具有1×10^3至1×10^6个细胞。

[0458] 如本文其它地方所讨论，提供用于靶向遗传修饰大鼠ES细胞系的各种方法。当执行所述方法时，大鼠ES细胞系内的至少一个细胞含有靶向遗传修饰。通过各种培养和/或选择技术，产生具有一个或多个所需靶向遗传修饰的大鼠ES细胞系。

[0459] 在特定实施方案中，大鼠ES细胞、大鼠ES细胞群体或大鼠ES细胞系(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)是整倍体的，并且因此具有是单倍体数目的精确多倍的染色体数目。在其它实施方案中，大鼠ES细胞、大鼠ES细胞群体或大鼠ES细胞系(其尚未

经受靶向遗传修饰和/或具有靶向遗传修饰)是二倍体的,并且因此具有两个单倍体组的同源染色体。当涉及大鼠ES细胞群体或来自给定大鼠ES细胞群体或大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)时,给定群体的至少约50%、60%、65%、70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的细胞是整倍体的和/或二倍体的。在其它情况下,当涉及大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)时,给定群体内至少约50%至95%、约60%至90%、约60%至95%、约60%至85%、约60%至80%、约70%至80%、约70%至85%、约70%至约90%、约70%至约95%、约70%至约100%、约80%至约100%、约80%至约95%、约80%至约90%、约90%至约100%、约90%至约99%、约90%至约98%、约90%至约97%、约90%至约95%的细胞是整倍体的和/或二倍体的。

[0460] 在其它实施方案中,大鼠ES细胞、大鼠ES细胞群体或大鼠ES细胞系(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)具有42个染色体。当涉及大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)时,给定群体的至少约50%、60%、65%、70%、75%、80%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%的细胞具有42个染色体。在其它情况下,当涉及大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)时,给定群体内至少约50%至95%、约60%至90%、约60%至95%、约60%至85%、约60%至80%、约70%至80%、约70%至85%、约70%至约90%、约70%至约95%、约70%至约100%、约80%至约100%、约80%至约95%、约80%至约90%、约90%至约100%、约90%至约99%、约90%至约98%、约90%至约97%、约90%至约95%的细胞具有42个染色体。

[0461] 在其它实施方案中,本文提供的大鼠ES细胞、大鼠ES细胞群体或大鼠ES细胞系(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)在体外涂铺在饲养细胞层上时形成球样集落。“球样”形态是指所培养的大鼠ES细胞集落的形状,而非个别ES细胞的形状。大鼠ES细胞集落是球形样的。松散附着于饲养细胞的集落看起来是圆形的(具有圆形样形态)。自由浮动的集落是球形样的。大鼠ES细胞集落是球形样的,并且极其紧实,这意味着:极难观察到细胞之间的边界。集落的边缘看起来是鲜明且清晰的。个别核难以区分,因为细胞极小(以致核占据细胞的大部分体积)。小鼠ES细胞形成细长集落,并且强烈附着于饲养细胞。mESC形态可随品系而变化;例如B6集落比F1H4集落更圆并且更呈穹状凸起,但仍然比rESC更细长。人ES细胞集落比mESC集落更平坦并且更铺展。本发明大鼠ES集落不是平坦的,并且不类似于人ES细胞集落。

[0462] 在其它实施方案中,大鼠ES细胞、大鼠ES细胞群体或大鼠ES细胞系(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)具有圆形形态。以下提供圆形的形态量表,其中得分10代表理想圆形,而得分1代表椭圆。

[0463] 圆形的形态量表:

[0464] 10=具有以下结构的圆形,所述结构具有穿过结构中心,并且具有相等长度的纵向轴和垂直轴。

[0465] 9=具有穿过结构中心的纵向轴和垂直轴的结构,其中一个轴在另一轴的长度的

0.9999至0.9357之间。

[0466] 8＝具有穿过结构中心的纵向轴和垂直轴的结构，其中一个轴在另一轴的长度的0.9357至0.875之间。

[0467] 7＝具有穿过结构中心的纵向轴和垂直轴的结构，其中一个轴在另一轴的长度的0.875至约0.8125之间。

[0468] 6＝具有穿过结构中心的纵向轴和垂直轴的结构，其中一个轴在另一轴的长度的0.8125至0.750之间。

[0469] 5＝具有穿过结构中心的纵向轴和垂直轴的结构，其中一个轴在另一轴的长度的0.750至0.6875之间。

[0470] 4＝具有穿过结构中心的纵向轴和垂直轴的结构，其中一个轴在另一轴的长度的0.6875至0.625之间。

[0471] 3＝具有穿过结构中心的纵向轴和垂直轴的结构，其中一个轴在另一轴的长度的0.625至0.5625之间。

[0472] 2＝具有穿过圆形中心的纵向轴和垂直轴的结构，其中一个轴在另一轴的长度的0.5625至0.523之间。

[0473] 1＝椭圆，其被定义为具有穿过结构中心的纵向轴和垂直轴，其中一个轴在另一轴的长度的0.523至0.500之间。

[0474] 在一个非限制性实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有给定群体的至少约50％、60％、65％、70％、75％、80％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的细胞具有圆形形态得分10、9或8。在其它实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有给定群体内至少约50％至95％、约60％至90％、约60％至95％、约60％至85％、约60％至80％、约70％至80％、约70％至85％、约70％至约90％、约70％至约95％、约70％至约100％、约80％至约100％、约80％至约95％、约80％至约90％、约90％至约100％、约90％至约99％、约90％至约98％、约90％至约97％、约90％至约95％的细胞具有圆形形态得分10、9或8。

[0475] 在另一非限制性实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有给定群体的至少约50％、60％、65％、70％、75％、80％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的细胞具有圆形形态得分7、6、5、4或3。在其它非限制性实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有给定群体内至少约50％至95％、约60％至90％、约60％至95％、约60％至85％、约60％至80％、约70％至80％、约70％至85％、约70％至约90％、约70％至约95％、约70％至约100％、约80％至约100％、约80％至约95％、约80％至约90％、约90％至约100％、约90％至约99％、约90％至约98％、约90％至约97％、约90％至约95％的细胞具有圆形形态得分7、6、5、4或3。

[0476] 在其它实施方案中，当将大鼠ES细胞(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)体外涂铺在饲养细胞层上时，形成球样集落。以下提供球的形态量表，其中得分10

代表理想球，而得分1代表三维椭圆结构。

[0477] 球样结构的形态量表：

[0478] 10＝球，其是具有各自穿过结构中心，并且具有相等长度的X轴和Y轴以及Z轴的结构。

[0479] 9＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴在至少一个其它轴的长度的0.9999至0.9357之间。

[0480] 8＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴在其它轴中的至少一个或两个的长度的0.9357至0.875之间。

[0481] 7＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴在其它轴中的至少一个或两个的长度的0.875至0.8125之间。

[0482] 6＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴在其它轴中的至少一个或两个的长度的0.8125至0.750之间。

[0483] 5＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴是其它轴中的至少一个或两个的长度的0.750至0.6875。

[0484] 4＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴是其它轴中的至少一个或两个的长度的0.6875至0.625。

[0485] 3＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴在其它轴中的至少一个或两个的长度的0.625至0.5625之间。

[0486] 2＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴在其它轴中的至少一个或两个的长度的0.5625至0.523之间。

[0487] 1＝具有穿过结构中心的X轴和Y轴以及Z轴的结构，其中一个轴在其它轴中的至少一个或两个的长度的0.523至0.500之间。

[0488] 在一个非限制性实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有至少约50％、60％、65％、70％、75％、80％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的在将细胞体外涂铺在饲养细胞层上时形成的集落具有球样形态10、9或8。在其它实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有至少约50％至95％、约60％至90％、约60％至95％、约60％至85％、约60％至80％、约70％至80％、约70％至85％、约70％至约90％、约70％至约95％、约70％至约100％、约80％至约100％、约80％至约95％、约80％至约90％、约90％至约100％、约90％至约99％、约90％至约98％、约90％至约97％、约90％至约95％的在将细胞体外涂铺在饲养细胞层上时形成的集落具有球样形态10、9或8。

[0489] 在另一非限制性实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有至少约50％、60％、65％、70％、75％、80％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％或100％的在将细胞体外涂铺在饲养细胞层上时形成的集落具有球样形态7、6、5、4或3。在其它实施方案中，大鼠ES细胞群体或来自给定大鼠ES细胞系的细胞的群体(其尚未经受靶向遗传修饰和/或具有靶向遗传修饰)中有至少约50％至95％、约60％至90％、约60％至95％、约60％至85％、约60％至80％、约70％至80％、约70％至85％、约70％

至约90％、约70％至约95％、约70％至约100％、约80％至约100％、约80％至约95％、约80％至约90％、约90％至约100％、约90％至约99％、约90％至约98％、约90％至约97％、约90％至约95％的在将细胞体外涂铺在饲养细胞层上时形成的集落具有球样形态7、6、5、4或3。

[0490] 用于本文提供的各种方法和组合物中的给定大鼠ES细胞可为雄性(XY)大鼠ES细胞、雄性(XY)大鼠ES细胞群体、或雄性(XY)大鼠ES细胞系。在其它实施方案中，本文采用的大鼠ES细胞群体或大鼠ES细胞系可为雌性(XX)大鼠ES细胞、雌性(XX)大鼠ES细胞群体、或雌性(XX)大鼠ES细胞系。任何所述大鼠ES细胞、大鼠ES细胞群体或大鼠ES细胞系都可包含如上所述的整倍性和/或二倍性。

[0491] 用于方法和组合物中的各种大鼠ES细胞可来自任何大鼠品系，包括但不限于ACI大鼠品系、Dark Agouti(DA)大鼠品系、Wistar大鼠品系、LEA大鼠品系、Sprague Dawley(SD)大鼠品系、或Fischer大鼠品系，如Fisher F344或Fisher F6。各种大鼠ES细胞也可从源于两个或更多个上述品系的混合的品系获得。在一个实施方案中，大鼠ES细胞源于选自DA品系和ACI品系的品系。在一特定实施方案中，大鼠ES细胞源于ACI品系。ACI大鼠品系被表征为具有黑色环纹(agouti)，具有白色腹部和足部以及RT1av1单倍型。所述品系可从包括Harlan Laboratories的多种来源获得。在其它实施方案中，各种大鼠ES细胞来自Dark Agouti(DA)大鼠品系，其被表征为具有环纹皮毛和RT1av1单倍型。所述大鼠可从包括Charles River和Harlan Laboratories的多种来源获得。在另一实施方案中，本文采用的各种大鼠ES细胞来自近交大鼠品系。

[0492] 在特定实施方案中，大鼠ES细胞系来自ACI大鼠，并且包括如2014年2月20日提交，以引用的方式整体并入本文的美国专利申请号14/185,103中详细所述的ACI.G1大鼠ES细胞。在另一实施方案中，大鼠ES细胞系来自DA大鼠，并且包括如2014年2月20日提交，以引用的方式整体并入本文的美国专利申请号14/185,103中详细所述的DA.2B大鼠ES细胞系或DA.2C大鼠ES细胞系。

[0493] 本文提供的给定大鼠ES细胞可从处于各种大鼠胚胎发育阶段的大鼠胚胎获得。用于获得大鼠ES细胞的大鼠胚胎可为桑椹胚阶段胚胎、胚泡阶段胚胎或处于桑椹胚阶段胚胎与胚泡阶段胚胎之间的发育阶段的大鼠胚胎。因此，在特定实施方案中，采用的大鼠胚胎处于Witschi阶段5和7或其之间。在其它实施方案中，采用的大鼠胚胎处于Witschi阶段5、6或7。

[0494] 在一个实施方案中，大鼠ES细胞从大鼠胚泡获得。在其它实施方案中，大鼠ES细胞从来自超排卵大鼠的胚泡获得。在其它实施方案中，大鼠ES细胞从8细胞阶段胚胎获得，其接着在体外培养直至它发育成桑椹胚阶段、胚泡阶段、在Witschi阶段5与7之间的胚胎、或处于Witschi阶段5、6或7的胚胎。此时，接着涂铺胚胎。桑椹胚阶段胚胎包含不具有内部空腔的紧实细胞球。胚泡阶段胚胎具有可见内部空腔(囊胚腔)，并且含有内部细胞团(ICM)。ICM细胞形成ES细胞。

[0495] B.获得和繁殖大鼠胚胎干(ES)细胞

[0496] 获得和繁殖大鼠胚胎干细胞的方法在本领域中是已知的，并且例如公开于2014年2月20日提交，以引用的方式整体并入本文的美国专利申请号14/185,103中。在特定实施方案中，所述方法包括(a)提供体外培养物，其包含饲养细胞层和分离的大鼠胚胎干(ES)细胞的群体；(b)在足以维持分离的大鼠ES细胞的多能性和/或全能性的条件下体外培养。所述

方法由此允许繁殖大鼠ES细胞群体和/或大鼠ES细胞系。

[0497] 提供用于培养大鼠胚胎干细胞系的方法。所述方法包括体外培养饲养细胞层和大鼠ES细胞系，其中培养条件维持大鼠ES细胞的多能性，并且包括具有小鼠白血病抑制因子(LIF)或其活性变体或片段的培养基。方法可进一步包括体外传代和培养大鼠ES细胞系的细胞，其中各后续体外培养包括在维持大鼠ES细胞的多能性的条件下在饲养细胞层上培养大鼠ES细胞，并且包括具有小鼠LIF或其活性变体或片段的培养基。

[0498] 用于各种方法和组合物中的培养基可维持大鼠ES细胞。术语“维持(maintaining/maintenance)”是指稳定保持本文概述的大鼠ES细胞的特征或表型中的至少一个或多个。所述表型可包括维持本文所述的大鼠干细胞的多能性和/或全能性、细胞形态、基因表达谱和其它功能特征。术语“维持”也可涵盖繁殖干细胞，或增加所培养的干细胞的数目。所述术语进一步涵盖允许干细胞保持多能，而干细胞可或可不继续分裂并增加数目的培养条件。

[0499] 术语“饲养细胞”或“饲养细胞层”包括在体外生长，并且向培养基中分泌至少一种用于支持培养物中另一目标细胞的生长的因子的细胞的培养物。本文采用的饲养细胞有助于维持大鼠ES细胞的多能性，并且在特定实施方案中，有助于维持一种或多种本文所述的其它特征或表型。可使用各种饲养细胞，包括例如小鼠胚胎成纤维细胞，包括在妊娠的第12天与第16天之间获得的小鼠胚胎成纤维细胞。在特定实施方案中，饲养细胞层包括单层有丝分裂失活小鼠胚胎成纤维细胞(MEF)。

[0500] 大鼠ES细胞的体外培养物进一步包含有效量的白血病抑制因子(LIF)或其活性变体或片段。白血病抑制因子(LIF)属于IL-6受体家族。LIF结合由LIF特异性亚基gp190或LIFR、和与IL-6家族的其它成员共有的亚基gp130组成的异二聚膜受体。LIF抑制小鼠中的胚胎干细胞的分化，并且有助于干细胞自我更新。人LIF和小鼠LIF共有79％序列同源性，并且展现交叉物种活性。大鼠LIF(rtLIF)是一种含有202个氨基酸残基的22.1kDa蛋白质，其展现与鼠类LIF的91％氨基酸序列同一性(Takahama等1998)。存在六个在来自各种物种的LIF多肽之间保守的可能天冬酰胺连接糖基化(N-糖基化)位点以及为大鼠LIF所特有的另一Asn150位点。小鼠LIF的三级结构和它的功能进一步详细描述于Aikawa等(1998)Biosci.Biotechnol.Biochem.62 1318-1325以及Senturk等(2005)Immunology of Pregnancy,Gil Mor.编;美国专利No.5,750,654以及D P Gearing(1987)EMBO Journal 1987-12-20中，其各自以引用的方式整体并入本文。部分小鼠LIF序列以登记号P09056报道于SwissProt网站上。

[0501] 通过小鼠LIF诱导M1骨髓性白血病细胞分化的能力来评估它的活性。比活性是1×10^6个单位/毫升(目录号03-0011，来自Stemgent)和1×10^7个单位/毫升(目录号03-0011-100，来自Stemgent)，其中50个单位定义为诱导1ml培养基中50％M1集落的分化所需的小鼠LIF的量。也参见Williams,R.L.等(1988)Nature 336:684-687.;Metcalf,D.等(1988)Leukemia 2:216-221;Niwa,H.等(2009)Nature460:118-122;Xu,J.等(2010)Cell Biol Int.34:791-797;Fukunaga,N.等(2010)Cell Reprogram.12:369-376;以及Metcalf D.(2003)Stem Cells 21:5-14，其各自以引用的方式整体并入本文。“LIF的有效量”包括LIF的允许体外培养物的大鼠ES细胞保持未分化多能状态的浓度。可用于测定保持多能状态的细胞的各种标志物在本文其它地方讨论。

[0502] 用于本文提供的各种方法和组合物中的LIF多肽可来自任何生物体，包括来自哺

乳动物、啮齿动物、人、大鼠或小鼠。在一个实施方案中,LIF多肽来自小鼠。在其它实施方案中,小鼠LIF多肽包含SwissProt登记号:P09056中所示的氨基酸序列,其以引用的方式整体并入本文,并且也显示于SEQ ID NO:9中。

[0503] 在其它实施方案中,可使用如SEQ ID NO:9中或SwissProt登记号P09056中所示的小鼠LIF多肽的活性变体或片段。所述活性变体和片段(包括与SEQ ID NO:9具有至少75%、80%、85%90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的活性变体)在本文其它地方进一步详细讨论。

[0504] 可以多种方式向体外培养物提供LIF多肽或其活性变体或片段。在一个实施方案中,添加有效量的LIF多肽或其活性变体或片段至培养基中。在其它实施方案中,饲养细胞已被遗传修饰以过度表达LIF多肽或其活性变体或片段。所述饲养细胞包括从表达基质相关LIF的γ照射或丝裂霉素-C处理的DIA-M小鼠成纤维细胞制备的饲养细胞。产生以及使用所述遗传修饰饲养细胞的方法可例如见于Buehr等(2003)Biol Reprod 68:222-229;Rathjen等(1990)Cell 621105-1115;以及Buehr等(2008)Cell 135:1287-1298中,其各自以引用的方式并入本文。在饲养细胞中表达的异源性LIF可来自与饲养细胞相同的生物体,或来自与饲养细胞的生物体不同的生物体。此外,在饲养细胞中表达的异源性LIF可来自与饲养层支持的ES细胞相同或不同的生物体。

[0505] 在其它实施方案中,用于本文公开的各种方法中的饲养细胞未被遗传修饰以表达异源性LIF多肽或其活性变体或片段。因此,在特定实施方案中,用于方法中的单层有丝分裂失活小鼠胚胎成纤维细胞尚未被遗传修饰以表达异源性LIF多肽。

[0506] 在其它实施方案中,添加LIF多肽或其活性变体或片段至培养基中。当添加LIF至培养基中时,LIF可来自任何生物体,包括来自哺乳动物、啮齿动物、人、大鼠或小鼠。在一个实施方案中,培养基中存在的LIF来自小鼠。在其它实施方案中,小鼠LIF多肽包含SEQ ID NO:9中所示的氨基酸序列。在其它实施方案中,可使用如SEQ ID NO:9中所示的小鼠LIF多肽的活性变体或片段。所述活性变体和片段(包括与SEQ ID NO:9具有至少75%、80%、85%90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性的活性变体)在本文其它地方进一步详细讨论。

[0507] 在特定实施方案中,本文提供的大鼠ES细胞和大鼠ES细胞系在无需旁分泌LIF信号传导下维持体外多能性。

[0508] 在特定实施方案中,LIF或其活性变体或片段在维持大鼠ES细胞的任何浓度下存在于培养基中。LIF多肽或其活性变体或片段以约25U/ml至约50U/ml、约50U/ml至约100U/ml、约100U/ml至约125U/ml、约125U/ml至约150U/ml、约150U/ml至约175U/ml、约175U/ml至约200U/ml、约200U/ml至约225U/ml、约225U/ml至约250U/ml、约250U/ml至约300U/ml,至约300U/ml至约325U/ml、约325U/ml至约350U/ml、约350U/ml至约400U/ml、约400U/ml至约425U/ml、约425U/ml至约450U/ml、约450U/ml至约475U/ml、约475U/ml至约500U/ml、约75U/ml至约500U/ml或大于500U/ml存在于培养基中。在其它实施方案中,LIF多肽或其活性变体或片段以约25U/ml至约50U/ml、约25U/ml至约100U/ml、约75U/ml至约125U/ml、约50U/ml至约150U/ml、约90U/ml至约125U/ml、约90U/ml至约110U/ml、约80U/ml至约150U/ml、或约80U/ml至约125U/ml存在于培养基中。在一特定实施方案中,LIF多肽或其活性变体或片段以约100U/ml存在于培养基中。

[0509] 当采用小鼠LIF时，小鼠LIF多肽或其活性变体或片段以维持大鼠ES细胞的任何浓度存在于培养基中。小鼠LIF多肽或其活性变体或片段以约25U/ml至约50U/ml、约50U/ml至约100U/ml、约100U/ml至约125U/ml、约125U/ml至约150U/ml、约150U/ml至约175U/ml、约175U/ml至约200U/ml、约200U/ml至约225U/ml、约225U/ml至约250U/ml、约250U/ml至约300U/ml，至约300U/ml至约325U/ml、约325U/ml至约350U/ml、约350U/ml至约400U/ml、约400U/ml至约425U/ml、约425U/ml至约450U/ml、约450U/ml至约475U/ml、约475U/ml至约500U/ml、约75U/ml至约500U/ml或大于500U/ml存在。在其它实施方案中，小鼠LIF多肽或其活性变体或片段以约25U/ml至约50U/ml、约25U/ml至约100U/ml、约75U/ml至约125U/ml、约50U/ml至约150U/ml、约90U/ml至约125U/ml、约90U/ml至约110U/ml、约80U/ml至约150U/ml、或约80U/ml至约125U/ml存在。在一特定实施方案中，小鼠LIF多肽或其活性变体或片段以约100U/ml存在于培养基中。

[0510] 采用的培养基维持大鼠ES细胞。因此，在特定实施方案中，用于各种方法和组合物中的培养基将持续至少5、10或15代的时期维持细胞系中所有或大多数(即超过50％)大鼠ES细胞的多能性。在一个实施方案中，培养基包含一种或多种有助于维持多能性的化合物。在一个实施方案中，培养基包含MEK路径抑制剂和糖原合成酶激酶-3(GSK-3)抑制剂。培养基可进一步包含有助于维持ES细胞的其它组分，包括例如FGF受体抑制剂、ROCK抑制剂和/或ALK(TGFb受体)抑制剂。FGF受体抑制剂的一非限制性实例包括PD184352。ROCK抑制剂的一非限制性实例包括Y-27632，并且ALK(TGFb受体)抑制剂的非限制性实例包括A-83-01。在特定实施方案中，当解冻冷藏保存的rESC时或当在用胰蛋白酶解离之后再涂铺rESC时，2i培养基与10u M ROCKi一起使用。

[0511] 在其它实施方案中，培养基包含由MEK路径抑制剂和糖原合成酶激酶-3(GSK-3)抑制剂组成的抑制剂的组合。

[0512] 在一个非限制性实施方案中，培养基包含包括CHIR99021的GSK-3抑制剂和/或包含包括PD0325901的MEK抑制剂。在其它实施方案中，培养基包含由CHIR99021和PD0325901组成的抑制剂的组合。这些化合物中的任一个都可例如从Stemgent获得。在特定实施方案中，CHIR99021以约0.5μ至约3μM、约0.5μ至约3.5μM、约0.5μM至约4μM、约0.5μM至约1μM、约1μM至约1.5μM、约1.5μM至约2μM、约2μM至约2.5μM、约2.5至约3μM、3μM至约3.5μM的浓度下存在于培养基中。在其它实施方案中，CHIR99021在约3μM的浓度下存在于培养基中。在其它实施方案中，PD0325901在约0.4μM至约1uM、约0.4μM至约1.5uM、约0.4μM至约2μM、约0.4μM至约0.8μM、0.8μM至约1.2μM、约1.2至约1.5μM的浓度存在于培养基中。在其它实施方案中，PD0325901以约1μM的浓度存在于培养基中。在特定实施方案中，CHIR99021以约3μM的浓度存在于培养基中，并且PD0325901以约1μM的浓度存在。

[0513] 在一个非限制性实施方案中，用于本文公开的各种方法和组合物中的培养基是2i培养基，其包含：DMEM/F12基础培养基(以1×(50％)的浓度)；神经基础培养基(以1×(50％)的浓度)；青霉素/链霉素(以1％的浓度)；L-谷氨酰胺(以4mM的浓度)；2-巯基乙醇(以0.1mM的浓度)；N2补充剂(以1×的浓度)；B27补充剂(以浓度1×)；LIF(以100U/ml的浓度)；PD0325901(MEK抑制剂)(以1μM的浓度)和CHIR99021(GSK抑制剂)(以3μM的浓度)。

[0514] 可采用的其它培养基包括Li等(2008)Cell 135:1299-1310；Yamamoto等(2012)Transgenic Rats 21:743-755；Ueda等(2008)PLoS ONE 3(6):e2800；Meek等(2010)PLoS

ONE 4(12):e14225;Tong等(2010)Nature 467:211-213;美国专利公布2012/0142092;Buehr等(2008)Cell 135:1287-1298;Li等(135)Cell 1299-1310中公开的那些,所述文献和公布各自以引用的方式整体并入本文。当采用所述培养基时,LIF的浓度和来源可如本文概述加以改进。在特定实施方案中,各种培养基与小鼠LIF或其活性变体或片段组合使用,并且在其它实施方案中,各种培养基以约50U/ml至约100U/ml、约50U/ml至约150U/ml、或约100U/ml的浓度包含小鼠LIF或其活性变体或片段。

[0515] 通常用于产生ES细胞系以及用于培养和维持ES株系的大鼠ES细胞培养的温度在约35℃至约37.5℃下。在特定实施方案中,温度是37.0℃。培养通常在7.5%CO_2下进行。

[0516] 7.序列同一性

[0517] 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、选择标志物和其它组分)。应认识到在整篇描述中,靶向基因组整合系统的一些组分可具有活性变体和片段。所述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。这些组分中的各个的生物活性在本文其它地方描述。

[0518] 如本文所用,在两个多核苷酸或多肽序列的情形下的“序列同一性”或“同一性”涉及当历经指定比较窗比对以达成最大对应性时两个序列中相同的残基。当关于蛋白质使用序列同一性百分比时,应认识到不同一的残基位置常常差异在于保守性氨基酸取代,其中氨基酸残基被取代成具有类似化学性质(例如电荷或疏水性)的其它氨基酸残基,并且因此不改变分子的功能性质。当序列在保守性取代方面有差异时,序列同一性百分比可被调高以针对保守性取代性质进行校正。差异在于所述保守性取代的序列被称为具有“序列类似性”或“类似性”。用于进行这个调整的手段为本领域技术人员所熟知。通常,这涉及将保守性取代作为部分而非完全错配进行评分,由此增加序列同一性百分比。因此,举例来说,当同一氨基酸被给予得分1,并且非保守性取代被给予得分零时,保守性取代被给予在零与1之间的得分。例如如程序PC/GENE(Intelligenetics,Mountain View,California)中所实施来计算保守性取代得分。

[0519] 如本文所用,“序列同一性百分比”意指通过历经比较窗比较两个最优对准序列来确定的值,其中相较于参照序列(其不包含添加或缺失),比较窗中的多核苷酸序列的部分可包含添加或缺失(即空位)以达成两个序列的最优比对。通过以下方式来计算百分比:确定同一核酸碱基或氨基酸残基存在于两个序列中所处的位置的数目以产生匹配位置的数目,用匹配位置的数目除以比较窗中的位置总数,并且将结果乘以100以产生序列同一性百分比。

[0520] 除非另外陈述,否则本文提供的序列同一性/类似性值是指使用第10版GAP利用以下参数获得的值:核苷酸序列的同一性%和类似性%,使用GAP权重50和长度权重3以及nwsgapdna.cmp评分矩阵;氨基酸序列的同一性%和类似性%,使用GAP权重8和长度权重2以及BLOSUM62评分矩阵;或其任何等效程序。“等效程序”意指对于任何两个所论述的序列,当相较于由第10版GAP产生的相应比对时,产生具有相同核苷酸或氨基酸残基匹配和相同序列同一性百分比的比对的任何序列比较程序。

[0521] 除非另外定义,否则本文所用的所有技术和科学术语都具有与由本发明所属领域中的普通技术人员通常理解相同的含义。尽管在实施或测试所述发明时也可使用与本文所

述的那些方法和材料类似或等效的任何方法和材料，但现在描述优选方法和材料。本文提及的所有出版物都以引用的方式并入本文以公开和描述与出版物被引用的内容关联的方法和/或材料。

[0522] 必须注意，除非上下文另外明确规定，否则如本文以及随附权利要求中所用，单数形式“一(a/an)”和“所述”包括复数个指示物。本文所用的所有技术和科学术语都具有相同含义。

[0523] 本文讨论的出版物仅由于它们的公开内容在本申请的提交日期之前而加以提供。本文中没有内容应解释为承认所述发明由于先前发明而无权先于所述出版物。此外，提供的公布日期可不同于可能需要独立确认的实际公布日期。

[0524] 所述发明可在不脱离其精神或基本属性下以其它特定形式体现，并且因此，应将随附权利要求而非前述说明书提及为指示本发明的范围。

[0525] 非限制性实施方案包括：

[0526] 1.一种用于靶向修饰多能大鼠细胞中的目标基因组基因座的方法，其包括(a)向所述多能大鼠细胞中引入包含侧接有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)，其中所述5’同源臂和所述3’同源臂的总和是至少10kb但小于150kb；以及(b)鉴定在所述目标基因组基因座处包含靶向遗传修饰的经遗传修饰的多能大鼠细胞，其中所述靶向遗传修饰能够通过种系传递。

[0527] 2.如实施方案1所述的方法，其中所述靶向遗传修饰是双等位基因修饰。

[0528] 3.如实施方案1或2所述的方法，其中所述多能大鼠细胞是大鼠胚胎干(ES)细胞。

[0529] 4.如实施方案1、2或3所述的方法，其中所述多能大鼠细胞源于DA品系或ACI品系。

[0530] 5.如实施方案1-4中任一个所述的方法，其中所述多能大鼠细胞的特征在于表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1或其组合。

[0531] 6.如实施方案1-4中任一个所述的方法，其中所述多能大鼠细胞的特征在于以下特征中的一个或多个：

[0532] (a)缺乏一种或多种包括c-Myc、Ecat1和/或Rex01的多能性标志物的表达；(b)缺乏包括Brachyury和/或Bmpr2的中胚层标志物的表达；(c)缺乏一种或多种包括Gata6、Sox17和/或Sox7的内胚层标志物的表达；或(d)缺乏一种或多种包括巢蛋白和/或Pax6的神经标志物的表达。

[0533] 7.如实施方案1-6中任一个所述的方法，其中所述LTVEC的所述5’同源臂和所述3’同源臂的总和是约10kb至约30kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约120kb、或约120kb至150kb。

[0534] 8.如实施方案1-6中任一个所述的方法，其中所述LTVEC的所述5’同源臂和所述3’同源臂的总和是约16Kb至约150Kb。

[0535] 9.如实施方案1-8中任一个所述的方法，其中所述靶向遗传修饰包括：(a)用同源或直系同源核酸序列替换内源性大鼠核酸序列；(b)缺失内源性大鼠核酸序列；(c)缺失内源性大鼠核酸序列，其中所述缺失在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb

至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb的范围内；(d)在约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb的范围内的外源性核酸序列；(e)包含同源或直系同源核酸序列的外源性核酸序列；(f)包含人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)可操作地连接至在大鼠细胞中具有活性的启动子的报告基因。

[0536] 10.如实施方案1-9中任一个所述的方法，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。

[0537] 11.如实施方案10所述的方法，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。

[0538] 12.如实施方案10所述的方法，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。

[0539] 13.如实施方案1-12中任一个所述的方法，其中引入步骤(a)进一步包括引入编码促进靶向构建体与所述多能大鼠细胞中的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。

[0540] 14.如实施方案13所述的方法，其中所述核酸酶试剂包括(a)包含融合于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)包含融合于FokI核酸内切酶的转录激活因子样效应物核酸酶(TALEN)的嵌合蛋白。

[0541] 15.如实施方案1-12中任一个所述的方法，其中引入步骤(a)进一步包括向所述多能大鼠细胞中引入：(i)包含可操作地连接至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)包含可操作地连接至基因组靶序列的第二启动子的第二表达构建体，所述基因组靶序列连接至导向RNA(gRNA)，其中所述基因组靶序列通过前间区序列邻近基序(PAM)序列以紧邻方式侧接在3'末端上。

[0542] 16.如实施方案15所述的方法，其中所述目标基因组基因座包含核苷酸序列SEQ ID NO:1。

[0543] 17.如实施方案15或16所述的方法，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)和反式活化CRISPR RNA(tracrRNA)的第三核酸序列。

[0544] 18.如实施方案15、16或17所述的方法，其中所述Cas蛋白是Cas9。

[0545] 19.如实施方案15、16、17或18所述的方法，其中所述gRNA包括：(a)具有核酸序列SEQ ID NO:2的嵌合RNA；或(b)具有核酸序列SEQ ID NO:3的嵌合RNA。

[0546] 20.如实施方案17所述的方法，其中所述crRNA包含SEQ ID NO:4；SEQ ID NO:5；或SEQ ID NO:6。

[0547] 21.如实施方案17所述的方法，其中所述tracrRNA包含SEQ ID NO:7或SEQ ID NO:8。

[0548] 22.一种经修饰的大鼠基因组基因座，其包含：(i)同源或直系同源人核酸序列的插入；(ii)内源性大鼠核酸序列被所述同源或直系同源人核酸序列的替换；或(iii)其组合，其中所述经修饰的大鼠基因组基因座能够通过种系传递。

[0549] 23.如实施方案22所述的经修饰的大鼠基因组基因座，其中所述插入或替换的大小是约5kb至约400kb。

[0550] 24.如实施方案22所述的大鼠基因组基因座，其中所述插入或替换的大小是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb。

[0551] 25.一种用于制备人源化大鼠的方法，其包括：(a)用包含人核酸的靶向构建体靶向多能大鼠细胞中的目标基因组基因座以形成经遗传修饰的多能大鼠细胞；(b)向宿主大鼠胚胎中引入所述经遗传修饰的多能大鼠细胞；以及(c)在代孕母体中孕育所述宿主大鼠胚胎；其中所述代孕母体产生包含经修饰的基因组基因座的大鼠子代，所述经修饰的基因组基因座包含：(i)人核酸序列的插入；(ii)在所述目标基因组基因座处的大鼠核酸序列被同源或直系同源人核酸序列的替换；(iii)包含人核酸序列和大鼠核酸序列的嵌合核酸序列；或(iv)其组合，其中所述经修饰的基因组基因座能够通过种系传递。

[0552] 26.如实施方案25所述的方法，其中所述靶向构建体是大靶向载体(LTVEC)，并且所述LTVEC的5'同源臂和3'同源臂的总和是至少10kb但小于150kb。

[0553] 27.如实施方案26所述的方法，其中所述靶向构建体的所述5'同源臂和所述3'同源臂的总和是约10kb至约30kb、约20kb至40kb、约40kb至约60kb、约60kb至约80kb、或约80kb至约100kb、约100kb至约120kb、或约120kb至150kb。

[0554] 28.如实施方案25、26或27所述的方法，其中所述人核酸序列是至少5kb但小于400kb。

[0555] 29.如实施方案25、26或27所述的方法，其中所述人核酸序列是至少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。

[0556] 30.如实施方案25-29中任一个所述的方法，其中所述多能大鼠细胞是大鼠胚胎干(ES)细胞。

[0557] 31.如实施方案25-30中任一个所述的方法，其中所述多能大鼠细胞源于DA品系或ACI品系。

[0558] 32.如实施方案25-31中任一个所述的方法，其中所述多能大鼠细胞的特征在于表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1或其组合。

[0559] 33.如实施方案25-31中任一个所述的方法，其中所述多能大鼠细胞的特征在于以下特征中的一个或多个：(a)缺乏一种或多种包括c-Myc、Ecat1和/或Rexo1的多能性标志物

的表达；(b)缺乏一种或多种包括Brachyury和/或Bmpr2的中胚层标志物的表达；(c)缺乏一种或多种包括Gata6、Sox17和/或Sox7的内胚层标志物的表达；或(d)缺乏一种或多种包括巢蛋白和/或Pax6的神经标志物的表达。

[0560] 34.一种包含人源化基因组基因座的经修饰的大鼠，其中所述人源化基因组基因座包含：(i)同源或直系同源人核酸序列的插入；(ii)在内源性基因组基因座处的大鼠核酸序列被同源或直系同源人核酸序列的替换；(iii)包含人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通过种系传递。

[0561] 35.一种在它的基因组基因座中包含靶向遗传修饰的大鼠或大鼠细胞，其中所述基因组基因座是白介素-2受体γ基因座、ApoE基因座、Rag1基因座、Rag2基因座、或Rag2/Rag1基因座，其中所述靶向遗传修饰包括：(a)缺失所述基因组基因座处的内源性大鼠核酸序列；(b)插入同源核酸、直系同源核酸或包含人核酸序列和大鼠核酸序列的嵌合核酸，或(c)其组合，其中所述靶向遗传修饰可通过所述大鼠或从所述大鼠细胞繁殖的大鼠的种系传递。

[0562] 36.如实施方案35所述的大鼠或大鼠细胞，其中(a)在所述基因组基因座处的所述内源性大鼠核酸的所述缺失是至少约10kb；或(b)在所述基因组基因座处的所述内源性大鼠核酸的所述缺失是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb；(c)在所述基因组基因座处的所述外源性核酸序列的所述插入是至少约5kb；或(d)在所述基因组基因座处的所述外源性核酸序列的所述插入是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb。

[0563] 37.如实施方案35或36所述的大鼠或大鼠细胞，其中(a)在所述白介素-2受体γ基因座处的所述靶向遗传修饰导致白介素-2受体γ蛋白活性降低或不存在；(b)在所述ApoE基因座处的所述靶向遗传修饰导致ApoE蛋白活性降低或不存在；(c)在所述Rag1基因座处的所述靶向遗传修饰导致Rag1蛋白活性降低或不存在；(d)在所述Rag2基因座处的所述靶向遗传修饰导致Rag2蛋白活性降低或不存在；或(e)在所述Rag2/Rag1基因座处的所述靶向遗传修饰导致Rag2蛋白活性和Rag1活性降低或不存在。

[0564] 38.如实施方案35、36或37所述的大鼠或大鼠细胞，其中所述白介素-2受体γ基因座的所述靶向遗传修饰包括：(a)缺失整个大鼠白介素-2受体γ编码区或其一部分；(b)用人白介素-2受体γ编码区或其一部分替换所述整个大鼠白介素-2受体γ编码区或其一部分；(c)用人白介素-2受体γ的外结构域替换所述大鼠白介素-2受体γ编码区的外结构域；或(d)缺失所述白介素-2受体γ基因座的至少3kb。

[0565] 39.如实施方案35-37中任一个所述的大鼠或大鼠细胞，其中所述ApoE基因座的所述靶向遗传修饰包括：(a)缺失整个ApoE编码区或其一部分；或(b)缺失包含所述ApoE编码区的所述ApoE基因座的至少1.8kb。

[0566] 40.如实施方案35-37中任一个所述的大鼠或大鼠细胞，其中所述Rag2基因座的所述靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分；(b)缺失包含所述Rag2编码区

的所述Rag2基因座的至少5.7kb。

[0567] 41.如实施方案35-37中任一个所述的大鼠或大鼠细胞，其中所述Rag2/Rag1基因座的所述靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分以及缺失整个Rag1编码区或其部分；或(b)缺失包含所述Rag2编码区的所述Rag2/Rag1基因座的至少16kb。

[0568] 42.如实施方案35-41中任一个所述的大鼠或大鼠细胞，其中所述靶向遗传修饰包括在所述白介素-2受体γ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志物的表达盒。

[0569] 43.如实施方案42中任一个所述的大鼠或大鼠细胞，其中所述表达盒包含可操作地连接至所述基因组基因座处的内源性启动子的lacZ基因和可操作地连接至选择性标志物的人泛素启动子。

[0570] 44.如实施方案35-43中任一个所述的大鼠或大鼠细胞，其中所述白介素-2受体γ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基因座中的所述靶向遗传修饰包括插入自我缺失性选择盒。

[0571] 45.如实施方案44所述的大鼠或大鼠细胞，其中所述自我缺失性选择盒包含可操作地连接至在所述大鼠细胞中具有活性的启动子的选择性标志物基因和可操作地连接至雄性生殖细胞特异性启动子的重组酶基因，其中所述自我缺失性盒由被重组酶识别的重组识别位点侧接。

[0572] 46.如实施方案45所述的大鼠或大鼠细胞，其中(a)所述雄性生殖细胞特异性启动子是鱼精蛋白-1启动子；或(b)所述重组酶基因编码Cre，并且所述重组识别位点是loxP位点。

[0573] 47.如实施方案35-46中任一个所述的大鼠或大鼠细胞，其中在所述基因组基因座处的所述外源性核酸序列的插入物包含可操作地连接至内源性白介素-2受体γ启动子、内源性ApoE启动子、内源性Rag1启动子或内源性Rag2启动子的报告核酸。

[0574] 48.如实施方案47所述的大鼠或大鼠细胞，其中所述报告核酸编码包括以下的报告蛋白：β-半乳糖苷酶、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强黄色荧光蛋白(EYFP)、Emerald、增强绿色荧光蛋白(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。

[0575] 49.如实施方案35-48中任一个所述的大鼠细胞，其中所述大鼠细胞是多能大鼠细胞或大鼠胚胎干(ES)细胞。

[0576] 50.如实施方案49所述的大鼠细胞，其中所述多能大鼠细胞或所述大鼠胚胎干(ES)细胞(a)源于DA品系或ACI品系；(b)特征在于表达至少一种包括以下的多能性标志物：Dnmt3L、Eras、Err-β、Fbxo15、Fgf4、Gdf3、Klf4、Lef1、LIF受体、Lin28、Nanog、Oct4、Sox15、Sox2、Utf1或其组合；或(c)特征在于以下特征中的一个或多个：(i)缺乏一种或多种包括c-Myc、Ecat1和/或Rexo1的多能性标志物的表达；(ii)缺乏包括Brachyury和/或Bmpr2的中胚层标志物的表达；(iii)缺乏一种或多种包括Gata6、Sox17和/或Sox7的内胚层标志物的表达；或(iv)缺乏一种或多种包括巢蛋白和/或Pax6的神经标志物的表达。

[0577] 51.一种用于修饰多能大鼠细胞中的白介素-2受体γ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'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含靶向遗传修饰的经遗传修饰的多能大鼠细胞，其中所述靶向遗传修饰能够通过从所述多能大鼠细胞繁殖的大鼠的种系传递。

[0578] 52.如实施方案51所述的方法，其中所述靶向载体是大靶向载体(LTVEC)，其中所述5'大鼠同源臂和所述3'大鼠同源臂的总和是至少约10kb但小于约150kb。

[0579] 53.如实施方案51或52所述的方法，其中向所述多能大鼠细胞中引入所述靶向载体导致：(i)在所述靶基因组基因座处的内源性大鼠核酸序列缺失；(ii)外源性核酸序列插入在所述靶基因组基因座处；或(iii)其组合。

[0580] 54.如实施方案53所述的方法，其中(a)在所述基因组基因座处的所述内源性大鼠核酸的所述缺失是至少约10kb；或(b)在所述基因组基因座处的所述内源性大鼠核酸的所述缺失是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、或约150kb至约200kb、约200kb至约300kb、约300kb至约400kb、约400kb至约500kb、约500kb至约1Mb、约1Mb至约1.5Mb、约1.5Mb至约2Mb、约2Mb至约2.5Mb、或约2.5Mb至约3Mb；(c)在所述基因组基因座处的所述外源性核酸序列的所述插入是至少约5kb；或(d)在所述基因组基因座处的所述外源性核酸序列的所述插入是约5kb至约10kb、约10kb至约20kb、约20kb至约40kb、约40kb至约60kb、约60kb至约80kb、约80kb至约100kb、约100kb至约150kb、约150kb至约200kb、约200kb至约250kb、约250kb至约300kb、约300kb至约350kb、或约350kb至约400kb。

[0581] 55.如实施方案51-54中任一个所述的方法，其中(a)在所述白介素-2受体γ基因座处的所述靶向遗传修饰导致白介素-2受体γ蛋白活性降低或不存在；(b)在所述ApoE基因座处的所述靶向遗传修饰导致ApoE蛋白活性降低或不存在；(c)在所述Rag1基因座处的所述靶向遗传修饰导致Rag1蛋白活性降低或不存在；(d)在所述Rag2基因座处的所述靶向遗传修饰导致Rag2蛋白活性降低或不存在；或(e)在所述Rag2/Rag1基因座处的所述靶向遗传修饰导致Rag2蛋白活性和i Rag1蛋白活性降低或不存在。

[0582] 56.如实施方案51-54中任一个所述的方法，其中所述白介素-2受体γ基因座的所述靶向遗传修饰包括(a)缺失整个大鼠白介素-2受体γ编码区或其一部分；(b)用人白介素-2受体γ编码区或其一部分替换所述整个大鼠白介素-2受体γ编码区或其一部分；(c)用人白介素-2受体γ的外结构域替换所述大鼠白介素-2受体γ编码区的外结构域；或(d)缺失包含所述白介素-2受体γ编码区的所述白介素-2受体γ基因座的至少3kb。

[0583] 57.如实施方案51-55中任一个所述的方法，其中所述ApoE基因座的所述靶向遗传修饰包括：(a)缺失整个ApoE编码区或其一部分；或(b)缺失包含所述ApoE编码区的所述ApoE基因座的至少1.8kb。

[0584] 58.如实施方案51-55中任一个所述的方法，其中所述Rag2基因座的所述靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分；或(b)缺失包含所述Rag2编码区的所述Rag2基因座的至少5.7kb。

[0585] 59.如实施方案51-55中任一个所述的方法，其中所述Rag1/Rag2基因座的所述靶向遗传修饰包括：(a)缺失整个Rag2编码区或其一部分以及缺失整个Rag1编码区或其部分；或(b)缺失包含所述Rag2编码区和所述Rag1编码区的所述Rag2/Rag1基因座的至少16kb。

[0586] 60.如实施方案51-59中任一个所述的方法，其中所述插入核酸包含表达盒，所述表达盒包含编码选择性标志物的多核苷酸。

[0587] 61.如实施方案60所述的方法，其中所述表达盒包含可操作地连接至所述基因组基因座处的内源性启动子的lacZ基因和可操作地连接至选择性标志物基因的人泛素启动子。

[0588] 62.如实施方案51-60中任一个所述的方法，其中所述插入核酸包含自我缺失性选择盒。

[0589] 63.如实施方案62所述的方法，其中所述自我缺失性选择盒包含可操作地连接至在所述大鼠多能细胞中具有活性的启动子的选择性标志物和可操作地连接至雄性生殖细胞特异性启动子的编码重组酶的多核苷酸，其中所述自我缺失性盒由被所述重组酶识别的重组识别位点侧接。

[0590] 64.如实施方案63所述的方法，其中(a)所述雄性生殖细胞特异性启动子是鱼精蛋白-1启动子；或(b)所述重组酶基因编码Cre，并且所述重组识别位点是loxP位点。

[0591] 65.如实施方案53所述的方法，其中在所述基因组基因座处的所述外源性核酸序列的插入包含可操作地连接至内源性白介素-2受体γ启动子、内源性ApoE启动子、内源性Rag1启动子或内源性Rag2启动子的报告核酸序列。

[0592] 66.如实施方案65所述的方法，其中所述报告核酸序列编码包括以下的报告蛋白：β-半乳糖苷酶、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强黄色荧光蛋白(EYFP)、Emerald、增强绿色荧光蛋白(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。

[0593] 67.如实施方案51-66中任一个所述的方法，其中所述多能大鼠细胞是大鼠胚胎干(ES)细胞。

[0594] 68.如实施方案51-67中任一个所述的方法，其中所述多能大鼠细胞(a)源于DA品系或ACI品系；或(b)特征在于表达包括Oct-4、Sox-2、碱性磷酸酶或其组合的多能性标志物；或(c)特征在于以下特征中的一个或多个：(i)缺乏一种或多种包括c-Myc、Ecat1和/或Rexo1的多能性标志物的表达；(ii)缺乏包括Brachyury和/或Bmpr2的中胚层标志物的表达；(iii)缺乏一种或多种包括Gata6、Sox17和/或Sox7的内胚层标志物的表达；或(iv)缺乏一种或多种包括巢蛋白和/或Pax6的神经标志物的表达。

[0595] 69.如实施方案51-68中任一个所述的方法，其进一步包括鉴定所述靶基因组基因座处的所述靶向遗传修饰，其中鉴定步骤采用用于评估所述靶基因组基因座处的等位基因修饰(MOA)的定量测定。

[0596] 70.如实施方案51-69中任一个所述的方法，其中引入步骤(a)进一步包括引入编码促进所述靶向载体与所述多能大鼠细胞中的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。

[0597] 71.如实施方案70所述的方法，其中所述核酸酶试剂包括包含融合于FokI核酸内切酶的基于锌指的DNA结合结构域的嵌合蛋白。

[0598] 72.如实施方案71所述的方法，其中所述方法导致所述靶基因组基因座的双等位基因修饰。

[0599] 73.如实施方案51-70中任一个所述的方法，其中引入步骤(a)进一步包括向所述

多能大鼠细胞中引入：(i)包含可操作地连接至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)包含可操作地连接至基因组靶序列的第二启动子的第二表达构建体，所述基因组靶序列连接至导向RNA(gRNA)，其中所述基因组靶序列通过前间区序列邻近基序(PAM)序列以紧邻方式侧接在3'末端上。

[0600] 74.如实施方案73所述的方法，其中所述目标基因组基因座包含核苷酸序列SEQ ID NO:1。

[0601] 75.如实施方案73或74所述的方法，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)和反式活化CRISPR RNA(tracrRNA)的第三核酸序列。

[0602] 76.如实施方案73所述的方法，其中所述Cas蛋白是Cas9。

[0603] 77.如实施方案73、74或75所述的方法，其中所述gRNA包括：(a)具有核酸序列SEQ ID NO:2的嵌合RNA；或(b)具有核酸序列SEQ ID NO:3的嵌合RNA。

[0604] 78.如实施方案75所述的方法，其中所述crRNA包含SEQ ID NO:4；SEQ ID NO:5；或SEQ ID NO:6。

[0605] 79.如实施方案75所述的方法，其中所述tracrRNA包含SEQ ID NO:7或SEQ ID NO:8。

实施例

[0606] 提出以下实施例以便向本领域普通技术人员提供对如何进行和使用本发明的完整公开和描述，并且不意图限制发明者所视为的他们的发明的范围，也不意图表示以下实验是进行的所有或唯一实验。已努力确保关于所用数字(例如量、温度等)的准确性，但应考虑一些实验误差和偏差。除非另外指示，否则份是重量份，分子量是重量平均分子量，温度以摄氏度计，并且压力是大气压下或接近大气压。

[0607] 实施例1.大鼠ES细胞衍生和表征

[0608] 1.1.大鼠ES细胞表征

[0609] 如图1中所示，大鼠ESC生长成紧实球形集落，其通常在培养皿中脱离和浮动(特写，图7)。大鼠ESC表达包括Oct-4(图2A)和Sox2(图2B)的多能性标志物，并且表达高水平的碱性磷酸酶(图3，左侧图)。株系DA.2B的核型是42X,Y(图3，右侧图)。大鼠ESC常常变为四倍体；因此，通过对中期染色体散布计数来预先筛选株系；接着正式分析具有大部分正常计数的株系的核型。

[0610] 从商购获得的超排卵雌性收集ACI胚泡。从商购获得的冷冻8细胞胚胎培养DA胚泡。用酸性台氏液(Acid Tyrodes)移除透明带；并且将胚泡涂铺于有丝分裂失活的MEF上。挑选生长晕，并且使用标准方法扩增。使用2i培养基涂铺、培养并扩增所有胚泡(Li等(2008)Germline competent embryonic stem cells derived from rat blastocysts, Cell 135:1299-1310；以引用的方式整体并入本文)。

[0611]

表 1. 大鼠 ES 细胞衍生		
	ACI	DA
胚胎来源	胚泡（超排卵）	冷冻 8 细胞胚胎培养成胚泡
涂铺的胚泡：	107	22
生长晕：	32（30%的母细胞）	10（45%的母细胞）
株系：	16（50%的生长晕）	9（90%的生长晕）
核型分析：	3；都是 42X,Y	6：3 42X,X 3 42X,Y
验证的 GLT：	1（ACI.G1）	1 42X,X（DA.2C） 1 42X,Y（DA.2B）

[0612] 1.2.:大鼠产生

[0613] 通过胚泡注射和传递大鼠ESC基因组来产生嵌合大鼠。通过使用亲本ACI.G1大鼠ESC进行胚泡显微注射产生的嵌合体显示于图8中。由在图8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显示于图9中。

[0614] 亲本大鼠ESC的种系传递.

[0615] 通过向白化SD胚泡中显微注射来评估三个整倍体大鼠ESC株系的多能性。通过指示大鼠ESC作用的环纹毛色来鉴定嵌合体。对于各株系,大多数嵌合体向F1后代传递rESC基因组(表2)。

[0616]

表 2. 亲本 rESC 的种系传递					
株系	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（%）
ACI.G1	5	3（60%）	103	11	11
DA.2B	5	4（80%）	129	11	9
DA.2C（XX）	3	2（66%）	45	7	16

[0617] 1.3.:大鼠胚胎干细胞的衍生.

[0618] 超排卵方案,大鼠

[0619] 第0天:用妊娠母马血清注射:IP,20U(0.4ml)。

[0620] 第1天:不采取行动

[0621] 第2天:(46小时后):注射hCG,IP,50U(1ml)。

[0622] -设置单一雌性交配。

[0623] 第3天:检查栓塞。将雌性堵塞。这是第0.5天。

[0624] 第6天(e3.5):使雌性安乐死,并且冲洗胚胎。

[0625] ES细胞衍生方案(超排卵)

[0626] 第0天:

[0627] 1)用CO_2使雌性大鼠安乐死。

[0628] 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。

[0629] 3)解剖出输卵管和子宫角,并且将它们放置至含有温热N2B27培养基的组织培养皿中。洗出尽可能多的血液,并且转移至具有N2B27的新培养皿中。

[0630] 4)使用1ml注射器和钝27g针,使培养基通过子宫角和输卵管进行冲洗以使胚泡排出至培养基中。

[0631] 5)用口吸移管收集胚泡,并且转移至含有KSOM+2i(1μM PD0325901、3μM CHIR99021)的胚胎培养皿中。KSOM是由Millipore生产的培养基。目录号是MR-106-D。

[0632] 6)在37℃下培养过夜;7.5％CO_2。

[0633] ES细胞衍生方案(冷冻胚胎)

[0634] 第0天:

[0635] 1)使冷冻8细胞胚胎(商购获得)解冻至M2培养基中。在室温下培养10分钟。

[0636] 2)转移至KSOM+2i中,并且培养过夜。

[0637] ES细胞衍生方案(两者相同)

[0638] 第1天:

[0639] 1)转移空腔化胚胎至2i培养基中,并且培养过夜。

[0640] 2)在KSOM+2i中继续培养未空腔化胚胎

[0641] 第2天:

[0642] 1)转移所有剩余胚胎至2i培养基中(无论它们是否已空腔化)。

[0643] 2)培养过夜;在2i培养基中继续培养较早期胚胎。

[0644] 第3天:

[0645] 1)用酸性台氏液转移胚胎30-60秒以移除透明带。

[0646] 2)在2i培养基中洗涤胚胎3次以移除酸性台氏液。

[0647] 3)将各胚胎放置至96孔饲养板的单独孔中(所述孔含有单层有丝分裂失活小鼠胚胎成纤维细胞(MEF))。

[0648] 4)在2i培养基中培养过夜。

[0649] 第4-5天:

[0650] 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的胚胎的约两倍时,所述生长晕即准备用于转移。

[0651] 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。

[0652] 3)转移生长晕至新饲养孔中:

[0653] a.移除用过的培养基,并且用PBS温和洗涤孔。

[0654] b.移除PBS,并且添加30μl 0.05％胰蛋白酶;孵育10分钟。

[0655] c.通过添加30μl 2i+10％FBS来终止胰蛋白酶反应。

[0656] d.用微量移液器温和离散细胞,并且转移孔的整个内含物至24孔饲养板的新孔中。这是第1代(P1)。

[0657] e.在2i培养基中培养过夜。

[0658] 第5-8天:(时机取决于各株系扩增的快速程度)

[0659] 1)每天更换培养基(2i培养基),并且监测具有ESC形态的集落是否存在。

[0660] 2)当集落出现时,继续培养直至集落扩增至约50％汇合。

[0661] 3)如前所述用胰蛋白酶处理和传代集落;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)。

[0662] 进行中的事务:

[0663] 1)继续饲养和监测各株系直至约50%汇合。

[0664] 2)照常用胰蛋白酶处理细胞。

[0665] 3)用2i+10%FBS终止胰蛋白酶;通过离心(于Beckman-Coulter台式离心机中1200rpm 5分钟)来沉淀细胞。

[0666] 4)吸出上清液,并且将细胞温和再混悬于400μl冷冻培养基(70%2i、20%FBS、10%DMSO)中。

[0667] 5)将细胞分配至2个小瓶中,并且在-80°下冷冻。这是第3代(P3)。

[0668] 6)对于长期储存,转移小瓶至液N_2储存库中。

[0669] 如以下表3制备2i培养基。

[0670]

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

[0671] 材料:妊娠母马的血清促性腺激素(PMSG)

[0672] 人妊娠尿绒毛膜促性腺激素(HCG)

[0673] 雌性大鼠(5-12周龄)

[0674] 雄性大鼠(12周至8月龄),每笼一只

[0675] 注射器/针头

[0676] 动物室在6:00-18:00进行光照

[0677] 程序:

[0678] 第1天:8:00-10:00AM

[0679] 用20IU PMSG(0.4ml)注射雌性,IP

[0680] 丢弃未使用的PMSG。

[0681] 第3天:8:00-10:00AM(在PMSG注射之后48小时)

[0682] 用50IU HCG(1ml)注射雌性,IP

[0683] 在交配笼中每个雄性放置一个雌性。

[0684] 丢弃未使用的HCG。

[0685] 第4天:8:00-10:00AM(在HCG注射之后24小时)

[0686] 检查雌性的栓塞。

[0687] 激素供应商

[0688] PMSG:Sigma#G-4877(1000IU)。于PBS中再混悬至最终[ ]50IU/ml。以1ml等分试样在-20°下储存。

[0689] HCG:Sigma#CG-5(5000IU)。于PBS中再混悬至最终[ ]50IU/ml。以1ml等分试样在-20°下储存。

[0690] 1.4.:大鼠胚胎干细胞系的核型分析

[0691] 分析本文产生的大鼠ES细胞系的核型,并且结果概述于表4-7中。

[0692] 表4

[0693]

ACI.G1 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常染色体, 此可由于技术性人为因素而偶发。90%的分析细胞具有正常雄性 42, XY 核型。	
图 4 提供显示对 ACI.G1 大鼠 ES 细胞系的染色体数目的分析的照片。	

[0694] 表5

[0695]

DA.2B 核型分析结果	细胞的数目
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 细胞都具有正常二倍体 42, XY 核型。	
图 5 提供显示对 DA.2B 大鼠 ES 细胞系的染色体数目的分析的照片。	

[0696] 表6

[0697]

DA.C2 核型分析结果	细胞的数目
分析核型的细胞的数目	5
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, XX	
其它注解: 100%分析的细胞具有正常雌性 XX 大鼠核型。	
图 6 提供显示对 DA.C2 大鼠 ES 细胞系的染色体数目的分析的照片。	

[0698] 表7.

[0699]

品系	涂铺的胚泡	确定的株系	分析核型的株系	核型
BN×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	
总计	92	34（37%）		

[0700] 1.5.:将载体电穿孔至大鼠胚胎干细胞中

[0701] 1.在电穿孔之前24-48小时使大鼠ES细胞传代。

[0702] 2.在电穿孔之前24小时将培养基更换成RVG2i+ROCKi(10μM Y-27632)

[0703] 3.在胰蛋白酶处理之前30分钟更换培养基。

[0704] 4.等分待进行电穿孔的DNA。

[0705] 5.使DNA在室温下温热>10分钟。

[0706] 6.在62℃下加热DNA 5分钟。将DNA放置在冰上。

[0707] 7.用胰蛋白酶处理细胞：

[0708] a.收集浮动集落。将板洗涤以收集尽可能多的浮动物。

[0709] b.沉淀集落:在750rpm下3分钟。

[0710] c.用5-10ml PBS洗涤沉淀1次,并且再旋转/沉淀

[0711] d.吸出上清液;添加500λ胰蛋白酶(0.05%)+1%鸡血清。

[0712] 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至管的底部,那么它们将结块,并且大多数细胞将被损失。

[0713] e.在37℃下4分钟。吸移集落数次以使结块最小。

[0714] f.重复步骤1-2次:在37℃下4分钟。

[0715] g.用500λRVG2i+10%FBS终止胰蛋白酶。

[0716] 8.沉淀细胞:在1200rpm下5分钟。

[0717] 9.于10ml PBS中再混悬细胞。对两个20λ等分试样计数以测定总细胞数目。

[0718] 10.沉淀细胞(5分钟/1200rpm);计算总细胞数目和总再混悬体积以获得正确细胞浓度(靶数目/75μl EP缓冲液)。

[0719] 11.于最小体积的EP缓冲液中再混悬;测量总体积,并且用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(2003)Nature Biotechnology 21:652-659,其以引用的方式并入本文。

[0720] 12.添加75λ细胞至50λDNA中;转移125λ细胞/DNA溶液至BTX 48孔比色皿的一个孔中。

[0721] a.用125λEP缓冲液填充同一列中的空孔。

[0722] 13.在BTX电穿孔仪中对比色皿施加脉冲一次:

[0723] a.设置:400V;Ω;100μF(设置可变化)

[0724] 14.将比色皿放置在冰上15分钟以进行恢复。

[0725] 15.移除细胞至5ml RVG2i+10μM ROCKi中。

[0726] 16.添加至具有20ml RVG2i+10μM ROCKi的15cm板中。板具有2×neoR MEF(或其它MEF,视项目而定)。neoR可选择标志物是Beck等(1982)Gene,19:327-36的或美国专利No.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。

[0727] 17.在37℃下孵育。48小时后开始选择。

[0728] 所用ROCK抑制剂是Y-27632。

[0729] 1.6:选择大鼠胚胎干细胞中的靶向遗传修饰.

[0730] 1.在电穿孔之前24-48小时使细胞传代。

[0731] 2.在电穿孔之前24小时将培养基更换成RVG2i+ROCKi(10μM Y-27632)

[0732] 3.在胰蛋白酶处理之前30分钟更换培养基。

[0733] 4.等分待进行电穿孔的DNA。

[0734] 5.使DNA在室温下温热>10分钟。

[0735] 6.在62℃下加热DNA 5分钟。将DNA放置在冰上。

[0736] 7.用胰蛋白酶处理细胞:

[0737] a.收集浮动集落。将板洗涤以收集尽可能多的浮动物。

[0738] b.沉淀集落:在750rpm下3分钟。

[0739] c.用5-10ml PBS洗涤沉淀1次,并且再旋转/沉淀

[0740] d.吸出上清液;添加500λ胰蛋白酶(0.05%)+1%鸡血清。

[0741] 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

堆积至管的底部，那么它们将结块，并且大多数细胞将被损失。

[0742] e.在37℃下4分钟。吸移集落数次以使结块最小

[0743] f.重复1-2次：在37℃下4分钟。

[0744] g.用500λRVG2i+10％FBS终止胰蛋白酶。

[0745] 8.沉淀细胞：在1200rpm下5分钟。

[0746] 9.于10ml PBS中再混悬细胞。对两个20λ等分试样计数以测定总细胞数目。

[0747] 10.沉淀细胞(5分钟/1200rpm)；计算总细胞数目和总再混悬体积以获得正确细胞浓度(靶数目/75μl EP缓冲液)。

[0748] 11.于最小体积的EP缓冲液中再混悬；测量总体积，并且用EP缓冲液调整至靶体积。

[0749] 12.添加75λ细胞至50λDNA中；转移125λ细胞/DNA溶液至BTX 48孔比色皿的一个孔中。

[0750] a.用125λEP缓冲液填充同一列中的空孔。

[0751] 13.在BTX电穿孔仪中对比色皿施加脉冲一次：

[0752] a.设置：400V；100μF(设置可变化)

[0753] 14.将比色皿放置在冰上15分钟以进行恢复。

[0754] 15.移除细胞至5ml RVG2i+10μM ROCKi中。

[0755] 16.添加至具有20ml RVG2i+10μM ROCKi的15cm板中。板具有2×neoR MEF(或视项目而定其它MEF)。

[0756] 17.在37℃下孵育。48小时后开始选择。

[0757] 18.G418选择方案如下：

[0758] a.第2天(在EP之后第2天)：在2i培养基+75μg/ml G418中孵育细胞。

[0759] b.第3天：在无G418的2i培养基中孵育细胞

[0760] c.第4天：在2i培养基+75μg/ml G418中孵育细胞。

[0761] d.第5天：在无G418的2i培养基中孵育细胞

[0762] e.第6天：在2i培养基+75μg/ml G418中孵育细胞。

[0763] f.第7天：在无G418的2i培养基中孵育细胞

[0764] g.第8天：在2i培养基+75μg/ml G418中孵育细胞。

[0765] h.第9天：在无G418的2i培养基中孵育细胞

[0766] i.第10天：在2i培养基+75μg/ml G418中孵育细胞。

[0767] j.第11天：在无G418的2i培养基中孵育细胞

[0768] k.第12天：挑选集落以进行扩增用于筛选。各集落于0.05％胰蛋白酶+1％鸡血清中离散10分钟，接着涂铺至96孔饲养板的1个孔中。

[0769] 19.在2i培养基中扩增集落3天。

[0770] 20.以1:1使克隆传代至新96孔饲养板中。

[0771] 21.在2i培养基中扩增克隆3天。

[0772] 22.对于各克隆，于胰蛋白酶中离散集落。冷冻2/3的各克隆，并且在-80℃下储存；将剩余1/3涂铺于层粘连蛋白板(用10μg/ml层粘连蛋白涂布的96孔板)上。

[0773] 23.当层粘连蛋白板汇合时，传至筛选实验室以分析克隆的基因型。

[0774] 1.7.大鼠胚胎干细胞的分子标签

[0775] 表8中所列的基因在大鼠ES细胞中比相应基因在小鼠ES细胞中表达低20倍。表9中所列的基因在大鼠ES细胞中比相应基因在小鼠ES细胞中表达水平高20倍。

[0776] 如下产生表8和9中的微阵列数据。在2i培养基中培养大鼠ES细胞(ACI.G2和DA.2B)和小鼠ES细胞(F1H4)3代直至汇合。在不存在饲养器下在明胶涂布的板上培养F1H4细胞。F1H4小鼠ES细胞源于129S6/SvEvTac和C57BL/6NTac杂合胚胎(参见例如美国专利No.7,294,754以及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),其以引用的方式整体并入本文)。

[0777] 以下方案用于样品制备:1.5mL Eppendorf管用样品识别符标记。于37℃磷酸盐缓冲盐水(PBS)中冲洗生长在板上的细胞。移除PBS,并且添加300ul Trizol®。刮削器用于破坏Trizol® (Life Technology)中的细胞。将裂解的细胞收集于1.5mL Epperdorf管中的Trizol®中。对于混悬生长的细胞,于37℃PBS中冲洗细胞,并且收集在1.5mL管中。短暂离心细胞;移除PBS;并且添加300ul Trizol®至细胞中。通过吸移来破坏细胞膜。分选样品以用于以10至105个细胞进行FACS,浓缩体积至小于100uL。添加4体积的RNA裂解缓冲液,并且通过吸移来混合。对于样品,添加320uL RNA裂解缓冲液至80uL样品中。在-20℃下储存样品。

[0778] RNA-Seq用于测量小鼠和大鼠基因的表达水平。通过Tophat来使测序读数定位于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,接着使用t检验来比较小鼠与大鼠之间各基因的表达水平。miR-632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细胞的标志物。

[0779] 表8.所列基因在大鼠ES细胞中比相应基因在小鼠ES细胞中表达水平低20倍。

[0780]

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 α	细胞质	其它

[0781]

识别符	符号	Entrez 基因名称	位置	类型
		2 肌动蛋白		
Actg2	ACTG2	肠平滑肌 γ2 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 2	质膜	其它
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，胞膜窖蛋白质，22 kDa	质膜	跨膜受体
Ccdc80	CCDC80	含有卷曲螺旋结构域的 80	核	其它
Ccnd2	CCND2	周期素 D2	核	其它
Cd248	CD248	CD248 分子，内皮唾液酸蛋	质膜	其它

[0782]

识别符	符号	Entrez 基因名称	位置	类型
		白		
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 效应物蛋白（Rho GTP 酶结合）5	细胞质	其它
Cdh11	CDH11	钙粘着蛋白 11，2 型，OB-钙粘着蛋白（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	神经元蜡样质脂褐素质沉积病 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	α1 I 型胶原蛋白	细胞外间隙	其它
Col1a2	COL1A2	α2 I 型胶原蛋白	细胞外间隙	其它
Col3a1	COL3A1	α1 III 型胶原蛋白	细胞外间隙	其它
Col5a2	COL5A2	α2 V 型胶原蛋白	细胞外间隙	其它
Col6a2	COL6A2	α2 VI 型胶原蛋白	细胞外间隙	其它
Cryab	CRYAB	αB 晶体蛋白	核	其它
Csf1	CSF1	集落刺激因子 1	细胞外间隙	细胞因子

[0783]

识别符	符号	Entrez 基因名称	位置	类型
		（巨噬细胞）		
Cth	CTH	胱硫醚酶（胱硫醚γ裂解酶）	细胞质	酶
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
Dmc1	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	双重特异性磷酸酶 1	核	磷酸酶
Dusp27	DUSP27	双重特异性磷酸酶 27（推定）	其它	磷酸酶
Dusp9	DUSP9	双重特异性磷酸酶 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

[0784]

识别符	符号	Entrez 基因名称	位置	类型
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 鼠类肉瘤病毒（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	γ 细丝蛋白 C	细胞质	其它
Fos	FOS	FBJ 鼠类骨肉瘤病毒致癌基因同源物	核	转录调控子
Fundc2	FUNDC2	含有 FUN14 结构域的 2	细胞质	其它
Gjb3	GJB3	间隙连接蛋白 β 3，31 kDa	质膜	转运体
Gpa33	GPA33	糖蛋白 A33（跨膜）	质膜	其它
Gpbp1l1	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	生长因子受体结合的蛋白 10	细胞质	其它
Gstm1	GSTM5	谷胱甘肽 S 转移酶 μ5	细胞质	酶
Hap1	HAP1	亨廷顿蛋白相关蛋白 1	细胞质	其它
Hist1h2bc	HIST2H2BE	组蛋白簇 2，H2be	核	其它

[0785]

识别符	符号	Entrez 基因名称	位置	类型
	（包括其它）			
Hmga2	HMGA2	高迁移率族 AT 钩 2	核	酶
Hmgn3	Hmgn3	高迁移率族核小体结合结构域 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-β）脱氢酶 14	细胞质	酶
Hspb1	HSPB1	热激 27 kDa 蛋白 1	细胞质	其它
Hspb8	HSPB8	热激 22 kDa 蛋白 8	细胞质	激酶
Htra1	HTRA1	HtrA 丝氨酸肽酶 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	白介素 1 受体样 1	质膜	跨膜受体
Inhba	INHBA	抑制素 βA	细胞外间隙	生长因子
Inhbb	INHBB	抑制素 βB	细胞外间隙	生长因子

[0786]

识别符	符号	Entrez 基因名称	位置	类型
Irf7	IRF7	干扰素调控因子 7	核	转录调控子
Isg15	ISG15	ISG15 泛素样调节剂	细胞外间隙	其它
Itga5	ITGA5	整联蛋白 α5（纤连蛋白受体，α 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 原致癌基因	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	赖氨酰基氧化酶	细胞外间隙	酶
Loxl2	LOXL2	赖氨酰基氧化酶样 2	细胞外间隙	酶
Loxl3	LOXL3	赖氨酰基氧化酶样 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 型胶原蛋白酶）	细胞外间隙	肽酶

[0787]

识别符	符号	Entrez 基因名称	位置	类型
Mxra8	MXRA8	基质重塑相关 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	15 kDa NADH 脱氢酶（泛醌）1β 亚复合物 4	细胞质	转运体
Npm1	NPM1	核磷酸蛋白（核仁磷蛋白 B23，核基质蛋白）	核	转录调控子
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	40/46 kDa 2'-5'-寡腺苷酸合成酶 1	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 α 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	PCYT1	磷酸胆碱胞苷	细胞质	酶

[0788]

识别符	符号	Entrez 基因名称	位置	类型
	B	酰基转移酶 1β		
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性样结构域 A 家族成员 1	细胞质	其它
Phlda2	PHLDA2	普列克底物蛋白同源性样结构域 A 家族成员 2	细胞质	其它
Pla2g1b	PLA2G1B	IB 组磷酸脂酶 A2（胰腺）	细胞外间隙	酶
Pla2g4a	PLA2G4A	IVA 组磷酸脂酶 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	β 型蛋白酶体（蛋白酶体，巨蛋白因子）亚基 8	细胞质	肽酶
Ptgs2	PTGS2	前列腺素-内过氧化物合成酶 2（前列腺素 G/H 合成酶和环加氧酶）	细胞质	酶
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	核	转录调控子

[0789]

识别符	符号	Entrez 基因名称	位置	类型
		物释放因子		
Rarg	RARG	视黄酸受体γ	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-NUDT3	RPS10-NUDT3 通读蛋白	细胞质	其它
Rps14	RPS14	核糖体蛋白 S14	细胞质	翻译调控子
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	翻译调控子
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	SERPINE1	丝氨酸蛋白酶抑制因子肽酶抑制因子，E 分枝（微管连接蛋白，1 型纤溶酶原激活剂抑制剂），成员 1	细胞外间隙	其它
Serpine2	SERPINE2	丝氨酸蛋白酶抑制因子肽酶抑制因子，E 分枝（微管连接	细胞外间隙	其它

[0790]

识别符	符号	Entrez 基因名称	位置	类型
		蛋白，1 型纤溶酶原激活剂抑制剂），成员 2		
Serpinf1	SERPINF1	丝氨酸蛋白酶抑制因子肽酶抑制因子，F 分枝（α-2 抗血纤维蛋白溶酶，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	SLC19A2	溶质载体家族 19（硫胺转运体），成员 2	质膜	转运体
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

[0791]

识别符	符号	Entrez 基因名称	位置	类型
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 β 3	细胞外间隙	生长因子
Thbs1	THBS1	血小板反应蛋白 1	细胞外间隙	其它
Thbs2	THBS2	血小板反应蛋白 2	细胞外间隙	其它
Tm4sf1	TM4SF1	跨膜 4 L6 家族成员 1	质膜	其它
Tmbim1	TMBIM1	含有跨膜 BAX 抑制因子基序的 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（β）	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 诱导性信号传导通路蛋白 2	细胞外间隙	生长因子
Ybx1	YBX1	Y 框结合蛋白 1	核	转录调控子

[0792] 表9.所列基因在大鼠ES细胞中比相应基因在小鼠ES细胞中表达水平高20倍。

[0793]

识别符	符号	Entrez 基因名称	位置	类型
Ajap1	Ajap1	粘附接合相关蛋白 1	其它	其它

[0794]

识别符	符号	Entrez 基因名称	位置	类型
Amd1	AMD 1	腺苷甲硫氨酸脱羧酶 1	细胞质	酶
Ankrd2	ANK RD2	锚蛋白重复结构域 2（拉伸应答性肌肉）	核	转录调控子
Arhgef9	ARH GEF9	Cdc42 鸟嘌呤核苷酸交换因子（GEF）9	细胞质	其它
Atp5h	Atp5h	ATP 合成酶，H+转运，线粒体 F0 复合物，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	CAM K4	钙/钙调蛋白依赖性蛋白质激酶 IV	核	激酶
Capn12	CAPN 12	钙蛋白酶 12	其它	肽酶
Cct6b	CCT6 B	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	CLDN 5	密封蛋白 5	质膜	其它
Clec3a	CLEC 3A	C 型凝集素结构域家族 3，成员 A	其它	其它
Clic6	CLIC 6	氯离子细胞内通道 6	质膜	离子通道
Dhrsx	DHRS X	X-连锁脱氢酶/还原酶（SDR 家族）	其它	酶
Dpysl2	DPYS L2	二氢嘧啶酶样 2	细胞质	酶
Dusp26	DUSP 26	双重特异性磷酸酶 26（推定）	细胞质	酶
Eci3	Eci3	烯酰基-辅酶 A δ 异构酶 3	其它	酶
Eef2k	EEF2 K	真核延伸因子-2 激酶	细胞质	激酶
Efna1	EFNA	肝配蛋白-A1	质膜	其它

[0795]

识别符	符号	Entrez 基因名称	位置	类型
	1			
Epha4	EPHA 4	EPH 受体 A4	质膜	激酶
Fank1	FANK 1	III 型纤连蛋白和锚蛋白重复结构域 1	核	转录调控子
Fhit	FHIT	脆弱组氨酸三联体	细胞质	酶
Filip1	FILIP 1	细丝蛋白 A 相互作用蛋白 1	细胞质	其它
Fmod	FMO D	纤维调节素	细胞外间隙	其它
Foxe1	FOXE 1	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	31.1 kDa 间隙连接蛋白 β 5	质膜	转运体
Gpx2	GPX2	谷胱甘肽过氧化物酶 2（胃肠）	细胞质	酶
Grxcr2	GRX CR2	富含半胱氨酸的谷氧还蛋白 2	其它	其它
Hecw2	HEC W2	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	发状/分裂增强子相关含 YRPW 基序 2	核	转录调控子
Icos	Icos	诱导性 T 细胞共刺激剂	质膜	其它
Ifitm1	IFITM 1	干扰素诱导的跨膜蛋白 1	质膜	跨膜受体
Il1f8	IL1F8 (IL36 B)	白介素-1 家族成员（白介素 36β）	细胞外间隙	细胞因子
Il28ra	IL-28 RA	白介素 28 受体 α	质膜	细胞因子受体
Igfbpl1	IGFB PL1	胰岛素样生长因子结合蛋白样 1	其它	其它

[0796]

识别符	符号	Entrez 基因名称	位置	类型
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	LEFTY1	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 α	质膜	跨膜受体
Lpar2	LPAR2	溶血磷脂酸受体 2	质膜	G 蛋白偶联受体
Mog	MOG	髓鞘少突胶质细胞糖蛋白	细胞外间隙	其它
Morn5	MORN5	含有 MORN 重复的 5	其它	其它
Pigz	NCBP2	20 kDa 核帽结合蛋白亚基 2	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	翻译调控子
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	时期昼夜节律时钟 3	核	其它
Pllp	PLLP	浆脂蛋白	质膜	转运体

[0797]

识别符	符号	Entrez 基因名称	位置	类型
Ppp1r14c	PPP1R14C	蛋白质磷酸酶 1 调控性（抑制剂）亚基 14C	细胞质	其它
Pramel6	Pramel6	黑素瘤中优先表达的抗原样 6	其它	其它
Ptpn18	PTPN18	非受体型蛋白质酪氨酸磷酸酶 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）活性调节蛋白 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
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/上皮高亲和力谷氨酸转运体，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	其它	转运体

[0798]

识别符	符号	Entrez 基因名称	位置	类型
		体），成员 2		
40424	SNTB1	互养蛋白 β1（肌营养不良蛋白相关蛋白 A1，59 kDa，基本组分 1）	质膜	其它
St6galnac3	ST6GALNAC3	ST6（α-N-乙酰基-神经氨基-2,3-β-半乳糖基-1,3）-N-乙酰基氨基半乳糖苷 α-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-2α（活化增强子结合蛋白 2α）	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 高迁移率族框家族成员 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	其它	其它

[0799]

识别符	符号	Entrez 基因名称	位置	类型
		心结构域 15A		
Wfdc6a	Wfdc6a	WAP 四二硫化物核心结构域 6A	其它	其它

[0800] 表10.来自表9的一个亚组的基因，其在大鼠ES细胞中比相应基因在小鼠ES细胞中表达水平高20倍。

[0801]

识别符	Entrez基因名称
Ajap1	粘附接合相关蛋白
Cldn5	密封蛋白5
Arhgef9	Cdc42鸟嘌呤核苷酸交换因子9
Camk4	钙/钙调蛋白依赖性蛋白质激酶IV
Efna1	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	白介素36β
Il28ra	白介素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	非受体型蛋白质酪氨酸磷酸酶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	发状/分裂增强子相关含YRPW基序2
Lef1	淋巴增强子结合因子1
Sall3	Sal样3(果蝇)
Satb1	SATB同源框1

[0802] 也已开发采用大鼠ES细胞的多能性标志物/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志物的表达水平。术语“分级”意指个别基因的比较表达水平：分级越高(1是最高)，表达越高。举例来说，Oct4的分级为13意指在所有测定基因中，它的表达高于除12个基因之外的所有基因。这个实验中的本底是低于30的任何表达值；6107个基因具有30或更高的表达值。

[0803] 表11.采用各种多能性、中胚层、内胚层、神经和滋养外胚层标志物/基因的大鼠ES细胞分子标签。

[0804]

多能性	多能性分级	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								
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								

[0805]

Stella	未测出								
Tcl1	未测出								
Utf1	1501								

[0806] 实施例2:使大鼠中的基因组基因座失活

[0807] 2.1:使用核酸内切酶试剂使内源性基因组基因座失活

[0808] 为在内源性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ZFN 1和ZFN 2(或TALEN 1和TALEN 2)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由约6bp至约40bp分隔的靶序列。在靶基因座内形成细胞试图通过非同源末端接合(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。为鉴定包含突变等位基因的阳性克隆,以低密度涂铺电穿孔细胞,因为未进行药物选择。挑选集落,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)。接着将所选包含突变等位基因的ES细胞引入宿主大鼠胚胎(例如前桑椹胚阶段或胚泡阶段大鼠胚胎)中,并且植入代孕母体的子宫中以产生首建大鼠(F0大鼠)。随后,使首建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大鼠的交配可产生突变等位基因纯合性子代。

[0809] 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

[0810] 锌指核酸酶使用序列特异性模块化DNA结合结构域来将核酸内切酶活性导向基因组中的独特靶序列。ZFN被工程改造成一对单体。各单体含有融合于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子位点,并且特异性通过两个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)常发生在非同源末端接合(NHEJ)期间。如果供体序列具有ZFN,那么DSB也刺激通过同源重组进行的同源性定向修复(HDR)。

[0811] 所述ZFN与本文所述的各种方法和组合物组合用于改进靶向效率。如实施例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ZFN 1和ZFN 2的表达载体引入大鼠ES细胞中。参见图10,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6中所讨论来测定靶向效率,并且结果显示于图11中。惊人地,靶向效率增长8-10倍。

[0812] 构建具有自我缺失性药物选择盒和作为报告基因的lacZ基因的质粒靶向载体。实现良好靶向效率,并且产生高%嵌合体。也与靶向载体组合测试锌指核酸酶(ZFN)以考查它对改进靶向效率的影响。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用单独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。从这些靶向克隆中的两个获得高%嵌合体。

[0813] 将ApoE靶向(在ZFN辅助下)大鼠ESC克隆显微注射至SD胚泡中,接着使用标准技术将所述胚泡转移至假妊娠SD接受者雌性中。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ApoE等位基因来分析种系F1幼仔的基因型(图17)。存在来自这些靶向克隆中的两个的高%嵌合体。

[0814] ApoE敲除大鼠提供一种用以研究各种类型的病症和疾病的手段。在人中,载脂蛋白见于乳糜微粒、HDL、LDL和VLDL中。ApoE为富含甘油三酯的脂蛋白组分的正常分解代谢所必需。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、β脂蛋白血症、家族性异常β脂蛋白血症、III型高脂蛋白血症(HLP III)、冠状动脉疾病风险。一种亚型(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。

[0815] 在小鼠中,ApoE主要见于HDL中;转运胆固醇,如同人中一样。ApoE缺乏性小鼠(2个独立KO)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(与人综合征相当)。

[0816] 在大鼠中敲除ApoE提供一种用以研究内皮功能(包括但不限于斑块形成)、转录变化(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,并且潜在改进RNA-S eq数据的质量。

[0817] 2.3.使用锌指核酸酶使大鼠白介素-2受体γ(IL2r-γ)基因座失活

[0818] 如实施例3.3(a)中所述,靶向大鼠白介素-2受体γ(IL2r-γ)基因座,例外之处是也将表达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-γ靶向事件的示意图。这些锌指结合的IL2r-γ基因座的序列表示在图18中。如以下在实施例3.3(a)中所讨论来测定靶向效率,并且结果显示于图18中。简要来说,通过PCR来确认纯合靶向克隆。对于ZFN1对:筛选192个克隆中的173个突变克隆(90%),并且对于ZFN2对:筛选192个克隆中的162个克隆(84%)。

[0819] 将IL2r-γ靶向(在ZFN辅助下)大鼠ESC克隆显微注射至SD胚泡中,接着使用标准技术将所述胚泡转移至假妊娠SD接受者雌性中。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-γ等位基因来分析种系F1幼仔的基因型。

[0820] 2.4.:使用CRISPR/Cas9使大鼠白介素-2受体γ(IL2r-γ)失活

[0821] 如实施例3.3(a)中所述靶向大鼠IL2r-γ基因座,例外之处是也将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gRNA连接至载体中,并且通过H1启动子来表达。设计针对I12rg的4个gRNA。当采用各种导向RNA时的靶向效率显示于图19中。

[0822] 实施例3:大鼠基因组基因座的靶向修饰

[0823] 3.1:大鼠ESC靶向:大鼠Rosa 26基因座。

[0824] 如同小鼠中一样,大鼠Rosa26基因座以相同间隔位于Setd5与Thumpd3基因之间。大鼠Rosa 26基因座(图12,图B)不同于小鼠Rosa 26基因座(图12,图A)。小鼠Rosa26转录物由2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,大鼠基因座也含有第2外显子1(Ex1b)。尚未在大鼠中鉴定出第3外显子。对大鼠Rosa26等位基因的靶向描绘于图12(底部)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大鼠Rosa26内含子中的117bp缺失的SA(剪接接受体)-lacZ-hUb-neo盒。

[0825] 测定在大鼠Rosa 26基因座处的靶向效率(表12)。将线性化载体电穿孔至DA或ACI大鼠ESC中,并且使用标准技术在2i培养基+G418中培养转染集落。挑选个别集落,并且使用等位基因损失(LOA)测定加以筛选(Valenzuela,D.等(2003)High-throughput engineering of the mouse genome coupled with high-resolution expression analysis,Nature Biotech.21:652-660,其以引用的方式并入本文)。

[0826]

表 12. 大鼠 Rosa26 靶向效率

细胞系	挑选的集落	再次确认的阳性物	靶向效率(%)
DA.2B	192	4	2.1
ACI.G1	96	4	4.2

[0827] 使用Rosa26靶向大鼠ESC克隆的嵌合体产生和种系传递。将再次确认的Rosa26靶向大鼠ESC克隆显微注射至SD胚泡中,接着使用标准技术将所述胚泡转移至假妊娠SD接受

者雌性中。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等位基因来分析种系(刺豚鼠)F1幼仔的基因型;22个刺豚鼠幼仔中的9个的基因型被分析为在Rosa26基因座处是杂合的(表13)。

[0828]

表 13. 使用靶向 Rosa26 rESC 的种系传递

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
ACI.G1	4	4	1	DE9: 39	DE9: 4	10

[0829] 3.2.(a)(i):大鼠载脂蛋白E(ApoE)基因座的靶向.

[0830] 靶向大鼠载脂蛋白E(ApoE)基因座以破坏ApoE功能。使用包含侧接有与ApoE基因座同源的5'和3'同源臂的lacZ-hUb-neo盒的靶向载体对ApoE基因座进行靶向。图20描绘已通过1.8kb缺失以及插入lacZ-hUb-neo盒来破坏的经遗传修饰的大鼠ApoE基因座,其进一步包括包含由鱼精蛋白启动子驱动的Crei基因的自我缺失性Cre盒。电穿孔条件如下:6ug DNA;2.05×10^6个细胞;400V;200uF:342V,593微秒;于2i+10uM ROCKi中涂铺在15cm 2×密集neoR MEF上。

[0831] 测定在ApoE基因座处的靶向效率,并且显示于表14中。将线性化载体电穿孔至源于DA品系的DA.2B大鼠ESC中,并且使用标准技术培养经转染的集落。挑选个别集落,并且使用等位基因损失(LOA)测定加以筛选。

[0832]

表 14. 大鼠 ApoE 靶向效率

[0833]

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

[0834] 使用ApoE靶向大鼠ESC克隆进行嵌合体产生和种系传递。将ApoE靶向大鼠ESC克隆显微注射至SD胚泡中,接着使用标准技术将所述胚泡转移至假妊娠SD接受者雌性中。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ApoE等位基因来分析F1幼仔的基因型(表15)。

[0835] 表15显微注射结果

[0836]

实验	克隆	幼仔	嵌合体
1	ApoE-AF5	4	3(90,90,90)
2	ApoE-BC4	5	0

[0837] ApoE的其它靶向数据也提供在图21中。

[0838] 3.2.(a)(ii).用靶向载体靶向大鼠中的ApoE

[0839] 图20提供大鼠ApoE基因座和靶向质粒的示意图。图20的上部示意图显示大鼠ApoE基因座和对应于5’和3’同源臂的基因组区域(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个内含子被表示为直线,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含有编码和非编码序列两者,如通过点描灰色阴影和空心框所表示。

[0840] 图20中的下部示意图是靶向载体。5’和3’同源臂(分别是5kb和5.4kb)通过暗灰色框来表示。靶向载体包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒。自我缺失性盒包含可操作地连接至小鼠Prm1启动子的Crei基因和包含可操作地连接至人泛素启动子的新霉素抗性基因的选择盒。

[0841] 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,其详细描述自我缺失性盒并且据此以引用的方式整体并入。通过采用Prm1启动子,自我缺失性盒可在F0大鼠的雄性生殖细胞中被特异性缺失。将靶向载体电穿孔至在实施例1中获得的大鼠ES细胞中,并且将细胞于2i+10uM ROCKi中涂铺在15cm 2×密集新霉素抗性MEF上。如实施例1中所述来培养、选择和维持转化的大鼠ES细胞。

[0842] 如表23中所示,筛选384个集落,并且获得23个靶向克隆。靶向效率是5.99%。如本文在实施例1中所述将3个克隆注射至胚泡中。获得3个产生嵌合体的克隆,并且所述克隆中的2个通过种系传递靶向修饰。

[0843] 3.2.(a)(iii).与锌指核酸酶组合用靶向载体靶向大鼠中的ApoE

[0844] 用于实施例3.2(a)(ii)中的靶向载体与锌指核酸酶组合用于靶向大鼠ApoE基因座。表16提供大鼠ApoE基因座的基因组组构的概述。表16中所示的位置取自大鼠基因组参照序列(ENSMBL)的5.0构建版。ApoE在(-)链上的染色体1上。

[0845] 表16.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。

[0846]

特征	起始	终止	长度	注释
外显子 1	81881110	81881182	73	5'非编码
外显子 2	81880269	81880332	64	含有 ATG
ATG	81880309	81880311	3	起始密码子
外显子 3	81879607	81879775	169	

[0847]

ZFN1a 结合位点	81879707	81879693	15	CAGGCCCTGAACCGC（SEQ ID NO: 10）
ZFN1 切割位点	81879692	81879687	6	TTCTGG（SEQ ID NO: 11）
ZFN1b 结合位点	81879686	81879671	16	GATTACCTGCGCTGGG（SEQ ID NO: 12）
内含子 3-4	81879776	81879207	400	
ZF21a 结合位点	81879591	81879577	15	TTCACCCTCCGCACC（SEQ ID NO: 13）
ZFN2 切割位点	81879576	81879570	7	TGCTGAG（SEQ ID NO: 14）
ZF21b 结合位点	81879569	81879552	18	TATCCAGATCCAGGGGTT（SEQ ID NO: 15）
外显子 4	81878371	81879208	838	含有 TGA
TGA	81878482	81878484	3	
ApoE 缺失	81878482	81880311	1830	

[0848] 图10提供大鼠ApoE基因座的示意图，并且用灰条表示ZFN1和ZFN2的切割位点。ZFN1的切割位点在外显子3中，并且ZNF2的切割位点在内含子3中。两个ZFN位点的精确位置阐述于表16中。对应于5’和3’同源臂的基因组区域(分别是5kb和5.4kb)通过暗灰色框来表示。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基因的三个内含子被表示为直线，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含有编码和非编码序列两者，如通过点描灰色阴影和空心框所表示。

[0849] 采用的靶向载体与实施例3.2(a)(ii)中以及显示于图20中的靶向载体相同。将ZFN以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。每个ZFN的表达由CMV启动子驱动。

[0850] 将靶向载体电穿孔至在实施例1中获得的大鼠ES细胞中，并且将细胞于2i+10uM ROCKi中涂铺在15cm 2×密集neoR MEF上。如实施例1中所述来培养、选择和维持转化的大鼠ES细胞。

[0851] 如表23中所示，筛选384个集落，并且获得290个靶向克隆。靶向效率是75.52％。如本文在实施例1中所述将2个克隆注射至胚泡中。获得两个产生嵌合体的克隆，并且所述克隆中的一个通过种系传递靶向修饰。

[0852] 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。

[0853] 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。

[0854] 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选择盒的自我缺失性盒来破坏。用于实施例3.2(a)(i)中的方法可用于将这个载体引入大鼠ES细胞中。

[0855] 实施例3.2.(b)(ii).用大靶向载体(LTVEC)靶向大鼠ApoE基因座

[0856] 图22提供大鼠ApoE基因座和大靶向载体(LTEVC)的示意图。图22的上部示意图显示大鼠ApoE基因座和对应于5'和3'同源臂的基因组区域(分别是45kb和23kb;暗灰色框)的基因组组构。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个内含子被表示为直线,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含有编码和非编码序列两者,如通过点描灰色阴影和空心框所表示。

[0857] 图22中的下部示意图是LTVEC。5'和3'同源臂(分别是45kb和23kb)通过暗灰色框来表示。靶向载体包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒,所述自我缺失性盒包含可操作地连接至小鼠Prm1启动子的Crei基因和包含可操作地连接至人泛素启动子的新霉素抗性基因的药物选择盒。Crei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,其详细描述自我缺失性盒并且据此以引用的方式整体并入。通过采用小鼠Prm1启动子,自我缺失性盒可在F0大鼠的雄性生殖细胞中被特异性缺失。

[0858] 将LTVEC电穿孔至在实施例1中获得的大鼠ES细胞中,并且将细胞于2i+10uM ROCKi中涂铺在15cm 2×密集neoR MEF上。如实施例1中所述来培养、选择和维持转化的大鼠ES细胞。

[0859] 如表23中所示,筛选288个集落,并且获得8个靶向克隆。靶向效率是2.78%。如本文在实施例2中所述将3个克隆注射至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。

[0860] 3.2.(b)(iii).与锌指核酸酶组合用大靶向载体(LTVEC)靶向大鼠中的ApoE

[0861] 用于实施例3.2.(b)(ii)中的LTVEC与锌指核酸酶组合用于靶向大鼠ApoE基因座。表16提供大鼠ApoE基因座的基因组组构的概述,并且所示位置取自大鼠基因组参照序列(ENSMBL)的5.0构建版。

[0862] 图23提供大鼠ApoE基因座的示意图,并且用灰条表示ZFN1和ZFN2的切割位点。ZFN1的切割位点在外显子3中,并且ZNF2的切割位点在内含子3中。两个ZFN位点的精确位置阐述于表16中。5'和3'同源臂(分别是45kb和23kb)通过暗灰色框来表示。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基因的三个内含子表示为直线。外显子2和3包含编码区,并且显示为点描灰色框。外显子4含有编码序列与非编码序列两者,如通过点描灰色阴影和空心框所表示。

[0863] 采用的LTVEC与实施例3.2(b)(ii)中以及显示于图22中的LTVEC相同。将ZFN以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ZFN 1的质粒和20ug针对ZFN2的质粒。ZFN购自Sigma。每个ZFN的表达由CMV启动子驱动。

[0864] 将靶向载体电穿孔至在实施例1中获得的大鼠ES细胞中,并且将细胞于2i+10uM ROCKi中涂铺在15cm 2×密集neoR MEF上。如实施例1中所述来培养、选择和维持转化的大鼠ES细胞。

[0865] 如表23中所示，筛选288个集落，并且获得16个靶向克隆。靶向效率是5.56％。如本文在实施例2中所述将一个克隆注射至胚泡中。

[0866] 此外，采用ZFN1和ZFN2以效率0.35％产生一个双等位基因靶向克隆。

[0867] 3.3(a)：大鼠白介素-2受体γ(IL2r-γ)基因座的靶向

[0868] 靶向大鼠白介素-2受体γ(IL2r-γ)基因座以破坏IL2r-γ功能。IL2r-γ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IL2r-γ中的突变与重度T、B和NK细胞发育缺陷相关。

[0869] 使用包含侧接有与IL2r-γ基因座同源的5’和3’同源臂的eGFP-hUb-neo盒的靶向载体对IL2r-γ基因座进行靶向。图26描绘大鼠IL2r-γ基因座的基因组结构，其中IL2r-γ基因座已通过3.2kb缺失来破坏。靶向IL2r-γ基因座也包含eGFP基因和自我缺失性盒，所述自我缺失性盒含有可操作地连接至小鼠鱼精蛋白1启动子的Crei和包含可操作地连接至新霉素抗性基因的hUb启动子的药物选择盒。

[0870] 测定在IL2r-γ基因座处的靶向效率，并且显示于表17中。将线性化载体电穿孔至DA.2B大鼠ESC中，并且使用标准技术培养转染集落。挑选个别集落，并且使用等位基因损失(LOA)测定加以筛选。

[0871]

表 17. 大鼠 IL2r-γ 靶向效率

细胞系	载体	挑选的集落	靶向	靶向效率（%）
DA.2B	Il2rg-floxed neo	136	1	0.7
DA.2B	Il2rg-mSDC	96	4	4.2

[0872] 使用IL2r-γ靶向大鼠ESC克隆进行嵌合体产生和种系传递。将IL2r-γ靶向大鼠ESC克隆显微注射至SD胚泡中，接着使用标准技术将所述胚泡转移至假妊娠SD接受者雌性中。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-γ等位基因来分析种系F1幼仔的基因型(表18)。

[0873] 表18显微注射结果

[0874]

实验	克隆	幼仔	嵌合体
1	Il2rg-AA1	5	2(90,70)
2	Il2rg-AA1	10	3(90,90,80)

[0875] 进一步研究3号$Il2rg^{-/Y}$嵌合体的表型。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个嵌合体检测到GFP阳性PBMC。此外，GFP+细胞是T细胞标志物CD3阴性的，并且大部分是B细胞标志物B220和NK细胞标志物CD161a阴性的。参见图30。小双重阳性群体与小鼠中公开的Il2rg敲除表型一致。这些数据从含有IL2受体γ阳性细胞的嵌合大鼠获得，并且这可使表型分析复杂化。

[0876] 3.3(b)：大鼠白介素-2受体γ(IL2r-γ)基因座的靶向修饰

[0877] 靶向大鼠白介素-2受体γ(IL2r-γ)基因座以破坏大鼠中的IL2r-γ功能。图26显示大鼠Il2rg基因座的基因组结构和引入基因座中的靶向载体。eGFP被选作报告蛋白以使

可使用FACS考查经遗传修饰的大鼠的免疫表型。自我缺失性盒(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,靶向载体被设计以缺失大鼠Il2rg基因的整个编码区(约3.2kb)。

[0878] 通过使用对大鼠Il2rg基因座具有特异性的引物进行PCR来确认大鼠ESC中的缺失大小。在将靶向克隆显微注射至处于胚泡阶段的宿主胚胎中后,获得高嵌合体百分比。那些嵌合体已被创建来进行配种。为确定靶向是否如预期来起作用,在配种之前收集来自嵌合体的外周血液,并且经由FACS分析外周血液中的免疫细胞的表型。如图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(上部右侧图),并且嵌合大鼠含有针对GFP(即Il2rg 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。

[0879] 3.4(a).用大靶向载体(LTVEC)靶向大鼠中的Rag2基因座

[0880] 表19提供大鼠Rag2基因座的基因组组构的概述,并且所示位置取自大鼠基因组参照序列(ENSMBL)的5.0构建版。Rag2在(+)链上的染色体3上。

[0881] 表19.大鼠Rag2基因座的基因组组构概述.

[0882]

特征	起始	终止	长度	注释
外显子 1	97,851,317	97,851,448	132	

[0883]

外显子 2	97,854,635	97,854,693	59	
外显子 3	97,858,260	97,859,615	1,356	含有整个编码序列
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	

[0884] 图27提供大鼠Rag2基因座和大靶向载体(LTVEC)的示意图。图27的上部示意图显示大鼠ApoE基因座和对应于5'和3'同源臂的基因组区域(分别是48Kb和15Kb;暗灰色框)的基因组组构。Rag2包含通过点描灰色阴影来表示的单一外显子。

[0885] 图27中的下部示意图是LTVEC。5'和3'同源臂(分别是48kb和15kb)通过暗灰色框来表示。LTVEC包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒。自我缺失性盒包含可操作地连接至Crei基因的小鼠Prm1启动子和包含可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。Crei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,其详细描述自我缺失性盒并且据此以引用的方式整体并入。通过采用小鼠Prm1启动子,自我缺失性盒可在F0大鼠的雄性生殖细胞中被特异性缺失。

[0886] 将LTVEC电穿孔至在实施例1中获得的大鼠ES细胞中,并且将细胞于2i+10uM ROCKi中涂铺在15cm 2×密集neoR MEF上。如实施例1中所述来培养和维持转化的大鼠ES细胞。如本文其它地方所述来筛选集落,并且获得靶向克隆。接着如本文其它地方所述来将靶向克隆注射至宿主胚胎中以产生F0大鼠。

[0887] 3.4.(b).靶向大鼠中的Rag1和Rag 2基因座

[0888] 图28提供大鼠Rag1/Rag2基因座的基因组结构。CDS表示编码序列，并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。

[0889] 表20提供大鼠Rag2和Rag1基因座的基因组组构的概述，并且所示位置取自大鼠基因组参照序列(ENSMBL)的5.0构建版。R ag1在(-)链上的染色体3上。

[0890] 表20.大鼠Rag1基因座的基因组组构概述.

[0891]

特征	起始	终止	长度	注释
外显子1	97,877,145	97,877,066	80	
外显子2	97,872,503	97,866,047	6,457	含有整个编码序列
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	

[0892] 图29提供大鼠Rag2和Rag1基因座以及大靶向载体(LTVEC)的示意图。图29的上部示意图显示Rag1和Rag2基因座以及对应于5’和3’同源臂的基因组区域(分别是48kb和84kb;暗灰色框)的基因组组构。Rag2和Rag 1各自包含通过点描灰色阴影来表示的单一外显子。图29中的下部示意图是LTVEC。5’和3’同源臂(分别是48kb和84kb)通过暗灰色框来表示。LTVEC包含报告基因(lacZ)和由loxP位点(空心箭头)侧接的自我缺失性盒。自我缺失性盒包含可操作地连接至Crei基因的大鼠Prm1启动子和包含可操作地连接至新霉素抗性基因的人泛素启动子的药物选择盒。Crei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，其详细描述自我缺失性盒并且据此以引用的方式整体并入。通过采用特定驱动Crei在雄性生殖细胞中表达的大鼠Prm1启动子，自我缺失性盒可从F0大鼠的雄性生殖细胞缺失。

[0893] 将LTVEC电穿孔至在实施例1中获得的大鼠ES细胞中，并且将细胞于2i+10uM ROCKi中涂铺在15cm 2×密集neoR MEF上。如实施例1中所述来培养和维持转化的大鼠ES细胞。

[0894] 如本文其它地方所述来筛选集落，并且获得靶向克隆。接着如本文其它地方所述来将靶向克隆注射至宿主胚胎中以产生F0大鼠。

[0895] 实施例4.人源化

[0896] 4.1.大鼠基因组基因座的人源化

[0897] 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能性，并且能够向后代传递靶向遗传修饰。此外，为规避质粒在容纳大基因组DNA片段方面的限制，以及为克服向大鼠ES细胞中的内源性基因座中引入靶向遗传修饰的低效率，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遗传修饰。本文所述的LTVEC例如包括内源性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域互补的大鼠同源臂的外源性核酸(例如同源或直系同源人核酸)。

[0898] 4.2.大鼠免疫球蛋白基因座的人源化

[0899] 通过以下方式来进行内源性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性大鼠免疫球蛋白重链核酸序列(例如一个或多个内源性V_H基因区段、一个或多个人D基因区段和一个或多个人J_H基因区段);以及向经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)一个或多个未重排人可变区核酸序列(例如一个或多个人V_H基因区段、一个或多个人D基因区段和一个或多个人J_H基因区段)或一个或多个重排人可变区核酸序列(例如一个或多个人重排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。

[0900] 简要来说,通过用由大鼠同源臂侧接的选择盒靶向内源性大鼠免疫球蛋白重链基因座来移除或失活大鼠BAC克隆中的一个或多个内源性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D基因区段和一个或多个人J_H基因区段)。更具体来说,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D基因区段和一个或多个人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)。

[0901] 接着,选择含有涵盖大鼠免疫球蛋白重链基因座的大大鼠基因组DNA片段的细菌细胞,并且用编码可操作地连接至短暂诱导型启动子的重组酶的质粒(例如pABG)引入。接着将以上构建的靶向载体引入重组感受态细菌细胞中。在电穿孔之后,细菌细胞用诱导剂(例如阿拉伯糖苷)处理以引发靶向载体与BAC克隆中的靶大鼠基因组序列之间的同源重组。以高密度涂铺转化的细胞,并且使其经受药物选择以发现药物抗性集落。挑选药物抗性集落,并且针对靶向修饰加以筛选。

[0902] 为有助于鉴定靶向遗传修饰,采用高通量定量测定,即等位基因修饰(MOA)测定,其允许在遗传修饰之后大规模筛选亲本染色体中的修饰等位基因。MOA测定可经由各种分析技术来进行,所述技术包括但不限于定量PCR,例如实时PCR(qPCR)。举例来说,实时PCR包括识别靶基因座的第一引物组和识别非靶向参照基因座的第二引物组。此外,引物组可包含识别扩增的序列的荧光探针。或者,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交(FISH)、比较基因组杂交、等温DNA扩增、定量杂交于固定的探针、Invader Probes®、MMP测定®、 TaqMan®分子信标和Eclipse™探测技术。(参见例如US2005/0144655,其以引用的方式整体并入本文)。

[0903] 接着用包含以下的大靶向载体(LTVEC)电穿孔包含经修饰的大鼠BAC克隆(即含有其中一个或多个内源性重链可变区基因区段(V_H、D和/或J_H基因区段)已被缺失或失活的大鼠基因组DNA序列的BAC克隆)的细菌细胞:(i)一个或多个未重排人可变区核酸序列(例如一个或多个未重排人V_H基因区段、一个或多个人D基因区段和一个或多个人J_H基因区段)或一个或多个重排人可变区核酸序列(例如一个或多个重排人V-D-J基因区段)。

[0904] 如上所述进行细菌细胞中同源重组的引发以及阳性克隆的选择。未重排或重排的人免疫球蛋白重链可变区核酸序列在靶向内源性免疫球蛋白重链基因座中时变为可操作地连接至内源性大鼠免疫球蛋白重链恒定区核酸序列。或者,内源性大鼠重链恒定区基因座可例如通过从内源性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,并且可被人重链恒定区核酸序列替换。

[0905] 同样,通过以下方式来进行内源性大鼠免疫球蛋白κ或λ轻链基因座的人源化:移

除一个或多个内源性大鼠免疫球蛋白κ和/或λ轻链可变区核酸序列(例如一个或多个内源性大鼠$V_κ$基因区段和一个或多个内源性大鼠$J_κ$基因区段);以及用靶向载体,例如包含以下的大靶向载体(LTVEC)靶向经修饰的免疫球蛋白轻链基因座:(i)一个或多个未重排人免疫球蛋白轻链可变区核酸序列(例如一个或多个人$V_κ$基因区段和一个或多个人$J_κ$基因区段)或一个或多个重排人可变区核酸序列(例如一个或多个人重排$V_κ$-$J_κ$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。

[0906] 未重排或重排人免疫球蛋白轻链可变区核酸序列在靶向内源性免疫球蛋白轻链基因座中时变为可操作地连接至内源性大鼠免疫球蛋白轻链恒定区核酸序列。

[0907] 如此在细菌细胞中产生的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性大鼠重链或轻链可变区基因区段已被一个或多个人重链或轻链可变区基因区段替换;以及与特定基因组靶序列互补的大鼠同源臂(例如在5kb至150kb的范围内)。接着使上述包含遗传修饰的LTVEC线性化,并且电穿孔至大鼠ES细胞中。以高密度涂铺电穿孔大鼠ES细胞以选择包含靶向载体的药物抗性ES细胞。药物选择过程移除大多数涂铺细胞(约99%),留下个别集落,其各自是源于单细胞的克隆。在剩余细胞中,大多数细胞(约80-100%)含有整合在基因组中的随机位置处的靶向载体。因此,挑选菌落并单独地分析基因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)。

[0908] 为增加靶向遗传修饰的效率,大鼠ES细胞用表达ZFN 1和ZFN2(或TALEN 1和2)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,靶向载体不由ZFN裂解。由ZFN产生的双链断裂刺激同源性定向修复(HDR),其另外占据通常在哺乳动物细胞中发生的极小百分比的修复(相较于非同源末端接合;NHEJ)。

[0909] 或者,如本文所述的含有II型CRISPR相关核酸酶(例如Cas9)、导向RNA(包括CRISPR-RNA(cr-RNA)和反式活化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。以高密度涂铺电穿孔细胞,并且使其经受药物选择以发现药物抗性集落。挑选药物抗性集落,并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,可实现靶向效率改进。举例来说,改进量可为较小的(例如从10%改进至15%)或较大的(例如从10%改进至80%)。

[0910] 接着将所选包含靶向遗传修饰的大鼠ES细胞引入宿主大鼠胚胎(例如前桑椹胚阶段或胚泡阶段大鼠胚胎)中,并且植入代孕母体的子宫中以产生首建大鼠(F0大鼠)。随后,使首建大鼠与野生型大鼠配种以创建遗传修饰杂合性F1子代。杂合性F1大鼠的交配可产生遗传修饰纯合性子代。

[0911] 4.3(a).用人IL2受体γ替换大鼠IL2rg

[0912] 表21提供大鼠白介素2受体γ基因座的基因组组构的概述,并且所示位置取自大鼠基因组参照序列(ENSMBL)的5.0构建版。I L2rg在(-)链上的染色体X上。

[0913] 表21.大鼠Il2rg基因座的基因组组构的概述

[0914]

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

[0915]

Il2rg 缺失	72,018,323	72,021,502	3,180	

[0916] 图26中的下部示意图是IL2rg 3.2kb缺失的靶向载体。靶向载体包含可操作地连接至内源性启动子的报告基因(eGFP)和由loxP位点(空心箭头)侧接的自我缺失性盒。自我缺失性盒包含可操作地连接至小鼠Prm1启动子的Crei基因和包含可操作地连接至人泛素启动子的新霉素抗性基因的选择盒。

[0917] 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,其详细描述自我缺失性盒并且据此以引用的方式整体并入。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被特异性缺失。将靶向载体电穿孔至在实施例1中获得的大鼠ES细胞中,并且将细胞于2i+10uM ROCKi中涂铺在15cm 2×密集新霉素抗性MEF

上。如实施例1中所述来培养、选择和维持转化的大鼠ES细胞。

[0918] 如表23中所示，筛选168个集落，并且获得6个靶向克隆。靶向效率是3.57％。

[0919] 如本文在实施例1中所述将克隆注射至胚泡中。获得产生F0大鼠的克隆，并且获得通过种系传递靶向修饰的F0大鼠。

[0920] 实施例4.3(b).用人IL2rg外结构域替换大鼠IL2rg外结构域

[0921] IL 2受体γ的全长人源化是有用的，因为具有这个经修饰的基因座的大鼠将产生人Il2rg；并且这将允许用对人Il2rg具有特异性的抗体检测大鼠中的人Il2rg。

[0922] 外人源化(即用Il2rg的人外结构域替换Il2rg的大鼠外结构域)将产生将结合Il2rg的人配体的Il2rg多肽，但因为细胞质结构域仍然是大鼠的，所以Il2rg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的人外结构域的嵌合IL-2rg蛋白(SEQ ID NO:22)的序列比对。人IL-2rg与大鼠IL-2rg之间的接合通过垂直线来指示。

[0923] 表22提供大鼠白介素2受体γ基因座的基因组组构的概述，并且所示位置取自大鼠基因组参照序列(ENSMBL)的5.0构建版。I L2rg在(-)链上的染色体X上。进一步指示的是IL2rg的外结构域的位置。

[0924] 表22.大鼠Il2rg基因座的基因组组构的概述

[0925]

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	含有跨膜结构域
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'多聚腺苷酸)
外人源化	71,108,679	71,111,539	2,861	(ATG 至跨膜结构域开始)

[0926] 构建质粒靶向载体以如图31中所示用人外结构域替换白介素2受体γ编码区的大鼠外结构域。将靶向载体电穿孔至在实施例1中获得的大鼠ES细胞中，并且将细胞于2i+10uM ROCKi中涂铺在15cm2×密集新霉素抗性MEF上。如实施例1中所述来培养、选择和维持

转化的大鼠ES细胞。

[0927] 如表23中所示，筛选192个集落，并且获得13个靶向克隆。靶向效率是6.77％。

[0928] 如本文在实施例1中所述将克隆注射至胚泡中。获得产生F0大鼠的克隆，并且获得通过种系传递靶向修饰的F0大鼠。

[0929] 实施例5.概述

[0930] 表23.用实施例3和4中讨论的各种载体类型和核酸酶试剂进行大鼠靶向的概述。

[0931] 表23靶向概述

[0932]

实施例号	基因座	载体	筛选的集落	靶向克隆	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	通过种系进行传递的克隆	注释
3.2(a)(ii)	ApoE	质粒	384	23	5.99%			3	3	2	
3.2(a)(iii)	ApoE+ZFN	质粒	384	290	75.52%	8	2.08%	2	2	1	这2个克隆是双等位基因靶向的
3.3(a)	Il2rg	质粒	232	5	2.16%			6	5		
3.2(b)(ii)	ApoE LTVEC	LTVEC	288	8	2.78%	1	0.35%	3	1		
3.2	ApoE	LT	288	16	5.56	1	0.35	1	N/A		这个

[0933]

(b)(iii)	LTVEC + ZFN	VEC			%		%				克隆是双等位基因靶向的
4.3 (a)	Il2rg 人源化 1	质粒	168	6	3.57 %						用人 Il2rg 替换整个大鼠 Il2rg
4.3 (b)	Il2rg 人源化 2	质粒	192	13	6.77 %			2	N/A		用人 Il2rg 外结构域替换大鼠 Il2rg 外结构域
3.4 (a)	Rag2	LTVEC	270	N/A							预测的 5.7 KB 缺失
3.4 (b)	Rag1-2	LTVEC	256	N/A							预测的 16.2 kb 缺失

[0934] 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，所述引用的程度就如同已特定地和个别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。除非另外根据上下文是明显的，否则本发明的任何实施方案、方面、步骤或特征可与任何其它实施方案、方面、步骤或特征组合使用。提及某一范围包括所述范围内的任何整数、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。

序列表

<110> Lee, Jeffrey D.
Mujica, Alexander O.
Auerbach, Wojtek
Lai, Venus Ka-Man
Valenzuela, David M.
Vancopoulos, George D.

<120> 大鼠基因组的靶向修饰

<130> 57766-444071

<160> 23

<170> 针对 Windows 的 FastSEQ 4.0 版

[0001]

<210> 1
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 连接至导向 RNA（gRNA）的基因组靶序列

<220>
<221> misc_feature
<222> 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
<223> n = A、T、C 或 G

<400> 1
gnnnnnnnnn nnnnnnnnnn ngg 23

<210> 2
<211> 80
<212> DNA
<213> 人工序列

<220>
<223> 导向 RNA（gRNA）

<400> 2
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60
ggcaccgagu cggugcuuuu 80

<210> 3
<211> 42
<212> DNA
<213> 人工序列

<220>
<223> 导向 RNA（gRNA）

<400> 3
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cg 42

<210> 4
<211> 30
<212> DNA
<213> 人工序列

<220>
<223> crRNA

[0002]

<400> 4
guuuuagagc uagaaauagc aaguuaaaau 30

<210> 5
<211> 33
<212> DNA
<213> 人工序列

<220>
<223> crRNA

<400> 5
guuuuagagc uagaaauagc aaguuaaaau aag 33

<210> 6
<211> 26
<212> DNA
<213> 人工序列

<220>
<223> crRNA

<400> 6
gaguccgagc agaagaagaa guuuua 26

<210> 7
<211> 12
<212> DNA
<213> 人工序列

<220>
<223> tracrRNA

<400> 7
aaggcuaguc cg 12

<210> 8
<211> 50
<212> DNA
<213> 人工序列

<220>
<223> tracrRNA

[0003]

<400> 8
aaggcuaguc cguuaucaac uugaaaaagu ggcaccgagu cggugcuuuu 50

<210> 9
<211> 203
<212> PRT
<213> 小家鼠

<400> 9

Met	Lys	Val	Leu	Ala	Ala	Gly	Ile	Val	Pro	Leu	Leu	Leu	Leu	Val	Leu
1				5				10					15		
His	Trp	Lys	His	Gly	Ala	Gly	Ser	Pro	Leu	Pro	Ile	Thr	Pro	Val	Asn
			20					25					30		
Ala	Thr	Cys	Ala	Ile	Arg	His	Pro	Cys	His	Gly	Asn	Leu	Met	Asn	Gln
		35					40					45			
Ile	Lys	Asn	Gln	Leu	Ala	Gln	Leu	Asn	Gly	Ser	Ala	Asn	Ala	Leu	Phe
	50					55					60				
Ile	Ser	Tyr	Tyr	Thr	Ala	Gln	Gly	Glu	Pro	Phe	Pro	Asn	Asn	Val	Glu
65					70					75					80
Lys	Leu	Cys	Ala	Pro	Asn	Met	Thr	Asp	Phe	Pro	Ser	Phe	His	Gly	Asn
				85					90					95	

Gly Thr Glu Lys Thr Lys Leu Val Glu Leu Tyr Arg Met Val Ala Tyr
100 105 110
Leu Ser Ala Ser Leu Thr Asn Ile Thr Arg Asp Gln Lys Val Leu Asn
115 120 125
Pro Thr Ala Val Ser Leu Gln Val Lys Leu Asn Ala Thr Ile Asp Val
130 135 140
Met Arg Gly Leu Leu Ser Asn Val Leu Cys Arg Leu Cys Asn Lys Tyr
145 150 155 160
Arg Val Gly His Val Asp Val Pro Pro Val Pro Asp His Ser Asp Lys
165 170 175
Glu Ala Phe Gln Arg Lys Lys Leu Gly Cys Gln Leu Leu Gly Thr Tyr
180 185 190
Lys Gln Val Ile Ser Val Val Val Gln Ala Phe
195 200

<210> 10
<211> 15
<212> DNA
<213> 人工序列

[0004]
<220>
<223> ZFN1a 结合位点

<400> 10
caggccctga accgc 15

<210> 11
<211> 6
<212> DNA
<213> 人工序列

<220>
<223> ZFN1 切割位点

<400> 11
ttctgg 6

<210> 12
<211> 16
<212> DNA
<213> 人工序列

<220>

<223> ZFN1b 结合位点

<400> 12
gattacctgc gctggg 16

<210> 13
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> ZF21a 结合位点

<400> 13
ttcaccctcc gcacc 15

<210> 14
<211> 7
<212> DNA
<213> 人工序列

[0005]

<220>
<223> ZFN2 切割位点

<400> 14
tgctgag 7

<210> 15
<211> 18
<212> DNA
<213> 人工序列

<220>
<223> ZF21b 结合位点

<400> 15
tatccagatc caggggtt 18

<210> 16
<211> 9
<212> PRT
<213> 人工序列

<220>

<223> 归巢核酸酶家族的保守结构域

<400> 16
Leu Ala Gly Leu Ile Asp Ala Asp Gly
1 5

<210> 17
<211> 15
<212> DNA
<213> 人工序列

<220>
<223> ZFN1a 结合位点

<400> 17
caggccctga accgc 15

<210> 18
<211> 6
<212> DNA
<213> 人工序列

[0006]

<220>
<223> ZFN1 切割位点

<400> 18
ttctgg 6

<210> 19
<211> 16
<212> DNA
<213> 人工序列

<220>
<223> ZFN1b 结合位点

<400> 19
gattacctgc gctggg 16

<210> 20
<211> 369
<212> PRT
<213> 智人

<400> 20

Met Leu Lys Pro Ser Leu Pro Phe Thr Ser Leu Leu Phe Leu Gln Leu
1 5 10 15
Pro Leu Leu Gly Val Gly Leu Asn Thr Thr Ile Leu Thr Pro Asn Gly
20 25 30
Asn Glu Asp Thr Thr Ala Asp Phe Phe Leu Thr Thr Met Pro Thr Asp
35 40 45
Ser Leu Ser Val Ser Thr Leu Pro Leu Pro Glu Val Gln Cys Phe Val
50 55 60
Phe Asn Val Glu Tyr Met Asn Cys Thr Trp Asn Ser Ser Ser Glu Pro
65 70 75 80
Gln Pro Thr Asn Leu Thr Leu His Tyr Trp Tyr Lys Asn Ser Asp Asn
85 90 95
Asp Lys Val Gln Lys Cys Ser His Tyr Leu Phe Ser Glu Glu Ile Thr
100 105 110
Ser Gly Cys Gln Leu Gln Lys Lys Glu Ile His Leu Tyr Gln Thr Phe
115 120 125
Val Val Gln Leu Gln Asp Pro Arg Glu Pro Arg Arg Gln Ala Thr Gln
130 135 140
Met Leu Lys Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu
145 150 155 160

[0007] Thr Leu His Lys Leu Ser Glu Ser Gln Leu Glu Leu Asn Trp Asn Asn
165 170 175
Arg Phe Leu Asn His Cys Leu Glu His Leu Val Gln Tyr Arg Thr Asp
180 185 190
Trp Asp His Ser Trp Thr Glu Gln Ser Val Asp Tyr Arg His Lys Phe
195 200 205
Ser Leu Pro Ser Val Asp Gly Gln Lys Arg Tyr Thr Phe Arg Val Arg
210 215 220
Ser Arg Phe Asn Pro Leu Cys Gly Ser Ala Gln His Trp Ser Glu Trp
225 230 235 240
Ser His Pro Ile His Trp Gly Ser Asn Thr Ser Lys Glu Asn Pro Phe
245 250 255
Leu Phe Ala Leu Glu Ala Val Val Ile Ser Val Gly Ser Met Gly Leu
260 265 270
Ile Ile Ser Leu Leu Cys Val Tyr Phe Trp Leu Glu Arg Thr Met Pro
275 280 285
Arg Ile Pro Thr Leu Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His
290 295 300
Gly Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Ala Glu Ser
305 310 315 320
Leu Gln Pro Asp Tyr Ser Glu Arg Leu Cys Leu Val Ser Glu Ile Pro
325 330 335

Pro Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Ala Ser Pro Cys Asn
340 345 350
Gln His Ser Pro Tyr Trp Ala Pro Pro Cys Tyr Thr Leu Lys Pro Glu
355 360 365
Thr

<210> 21
<211> 368
<212> PRT
<213> 褐家鼠

<400> 21
Met Leu Lys Pro Leu Leu Pro Ser Arg Ser Phe Leu Leu Leu Gln Leu
1 5 10 15
Leu Leu Leu Arg Val Gly Trp Ser Ser Lys Val Leu Met Ser Ser Gly
20 25 30
Asn Glu Asp Thr Lys Ser Asp Leu Leu Leu Thr Ser Met Asp Leu Lys
35 40 45
His Leu Ser Val Pro Thr Leu Pro Leu Pro Glu Val Gln Cys Phe Val
50 55 60

[0008]

Phe Asn Val Glu Tyr Met Asn Cys Thr Trp Asn Ser Ser Ser Glu Pro
65 70 75 80
Gln Pro Thr Asn Leu Thr Met His Tyr Arg Tyr Lys Gly Ser Asp Asn
85 90 95
Asn Thr Phe Gln Glu Cys Ser His Tyr Leu Phe Ser Lys Glu Ile Thr
100 105 110
Ser Gly Cys Gln Ile Gln Lys Glu Asp Ile Gln Leu Tyr Gln Thr Phe
115 120 125
Val Val Gln Leu Gln Asp Pro Gln Lys Pro Gln Arg Arg Ala Glu Gln
130 135 140
Lys Leu Asn Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu
145 150 155 160
Thr Leu Tyr Asn Leu Ser Glu Ser Gln Val Glu Leu Arg Trp Lys Ser
165 170 175
Arg Tyr Ile Glu Arg Cys Leu Gln Tyr Leu Val Gln Tyr Arg Ser Asn
180 185 190
Arg Asp Arg Ser Trp Thr Glu Gln Ile Val Asp His Glu Pro Arg Phe
195 200 205
Ser Leu Pro Ser Val Asp Glu Gln Lys Leu Tyr Thr Phe Arg Val Arg
210 215 220
Ser Arg Phe Asn Pro Ile Cys Gly Ser Thr Gln Gln Trp Ser Lys Trp
225 230 235 240

```
Ser Gln Pro Ile His Trp Gly Ser His Thr Ala Glu Glu Asn Pro Ser
                245                 250                 255
Leu Phe Ala Leu Glu Ala Val Leu Ile Pro Val Gly Thr Met Gly Leu
            260                 265                 270
Ile Ile Thr Leu Ile Phe Val Tyr Cys Trp Leu Glu Arg Met Pro Arg
        275                 280                 285
Ile Pro Ala Ile Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His Gly
    290                 295                 300
Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Thr Glu Ser Leu
305                 310                 315                 320
Gln Pro Asp Tyr Ser Glu Arg Phe Cys His Val Ser Glu Ile Pro Pro
                325                 330                 335
Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Gly Ser Pro Cys Ser Leu
            340                 345                 350
His Ser Pro Tyr Trp Pro Pro Pro Cys Tyr Ser Leu Lys Pro Glu Ala
        355                 360                 365
```

<210> 22
<211> 368
<212> PRT
<213> 人工序列

[0009]

<220>
<223> 包含具有来自人的 IL-2γ 受体的外结构域的大鼠 IL-2 受体γ蛋白的嵌合 IL-2 受体γ

<400> 22

```
Met Leu Lys Pro Ser Leu Pro Phe Thr Ser Leu Leu Phe Leu Gln Leu
 1               5                  10                  15
Pro Leu Leu Gly Val Gly Leu Asn Thr Thr Ile Leu Thr Pro Asn Gly
            20                  25                  30
Asn Glu Asp Thr Thr Ala Asp Phe Phe Leu Thr Thr Met Pro Thr Asp
        35                  40                  45
Ser Leu Ser Val Ser Thr Leu Pro Leu Pro Glu Val Gln Cys Phe Val
    50                  55                  60
Phe Asn Val Glu Tyr Met Asn Cys Thr Trp Asn Ser Ser Ser Glu Pro
65                  70                  75                  80
Gln Pro Thr Asn Leu Thr Leu His Tyr Trp Tyr Lys Asn Ser Asp Asn
                85                  90                  95
Asp Lys Val Gln Lys Cys Ser His Tyr Leu Phe Ser Glu Glu Ile Thr
            100                 105                 110
Ser Gly Cys Gln Leu Gln Lys Lys Glu Ile His Leu Tyr Gln Thr Phe
        115                 120                 125
Val Val Gln Leu Gln Asp Pro Arg Glu Pro Arg Arg Gln Ala Thr Gln
```

[0010]

```
    130                 135                 140
Met Leu Lys Leu Gln Asn Leu Val Ile Pro Trp Ala Pro Glu Asn Leu
145                 150                 155                 160
Thr Leu His Lys Leu Ser Glu Ser Gln Leu Glu Leu Asn Trp Asn Asn
                165                 170                 175
Arg Phe Leu Asn His Cys Leu Glu His Leu Val Gln Tyr Arg Thr Asp
            180                 185                 190
Trp Asp His Ser Trp Thr Glu Gln Ser Val Asp Tyr Arg His Lys Phe
        195                 200                 205
Ser Leu Pro Ser Val Asp Gly Gln Lys Arg Tyr Thr Phe Arg Val Arg
    210                 215                 220
Ser Arg Phe Asn Pro Leu Cys Gly Ser Ala Gln His Trp Ser Glu Trp
225                 230                 235                 240
Ser His Pro Ile His Trp Gly Ser Asn Thr Ser Lys Glu Asn Pro Phe
                245                 250                 255
Leu Phe Ala Leu Glu Ala Val Leu Ile Pro Val Gly Thr Met Gly Leu
            260                 265                 270
Ile Ile Thr Leu Ile Phe Val Tyr Cys Trp Leu Glu Arg Met Pro Arg
        275                 280                 285
Ile Pro Ala Ile Lys Asn Leu Glu Asp Leu Val Thr Glu Tyr His Gly
    290                 295                 300
Asn Phe Ser Ala Trp Ser Gly Val Ser Lys Gly Leu Thr Glu Ser Leu
305                 310                 315                 320
Gln Pro Asp Tyr Ser Glu Arg Phe Cys His Val Ser Glu Ile Pro Pro
                325                 330                 335
Lys Gly Gly Ala Leu Gly Glu Gly Pro Gly Gly Ser Pro Cys Ser Leu
            340                 345                 350
His Ser Pro Tyr Trp Pro Pro Pro Cys Tyr Ser Leu Lys Pro Glu Ala
        355                 360                 365
```

<210> 23
<211> 23
<212> DNA
<213> 人工序列

<220>
<223> 连接至导向 RNA（gRNA）的基因组靶序列

<220>
<221> misc_feature
<222> (2)...(21)
<223> n= A、T、C 或 G

<220>

<221> misc_feature

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

[0011] <223> n 可为 1-20 个核苷酸

<400> 23

gnnnnnnnnn nnnnnnnnnn ngg 23

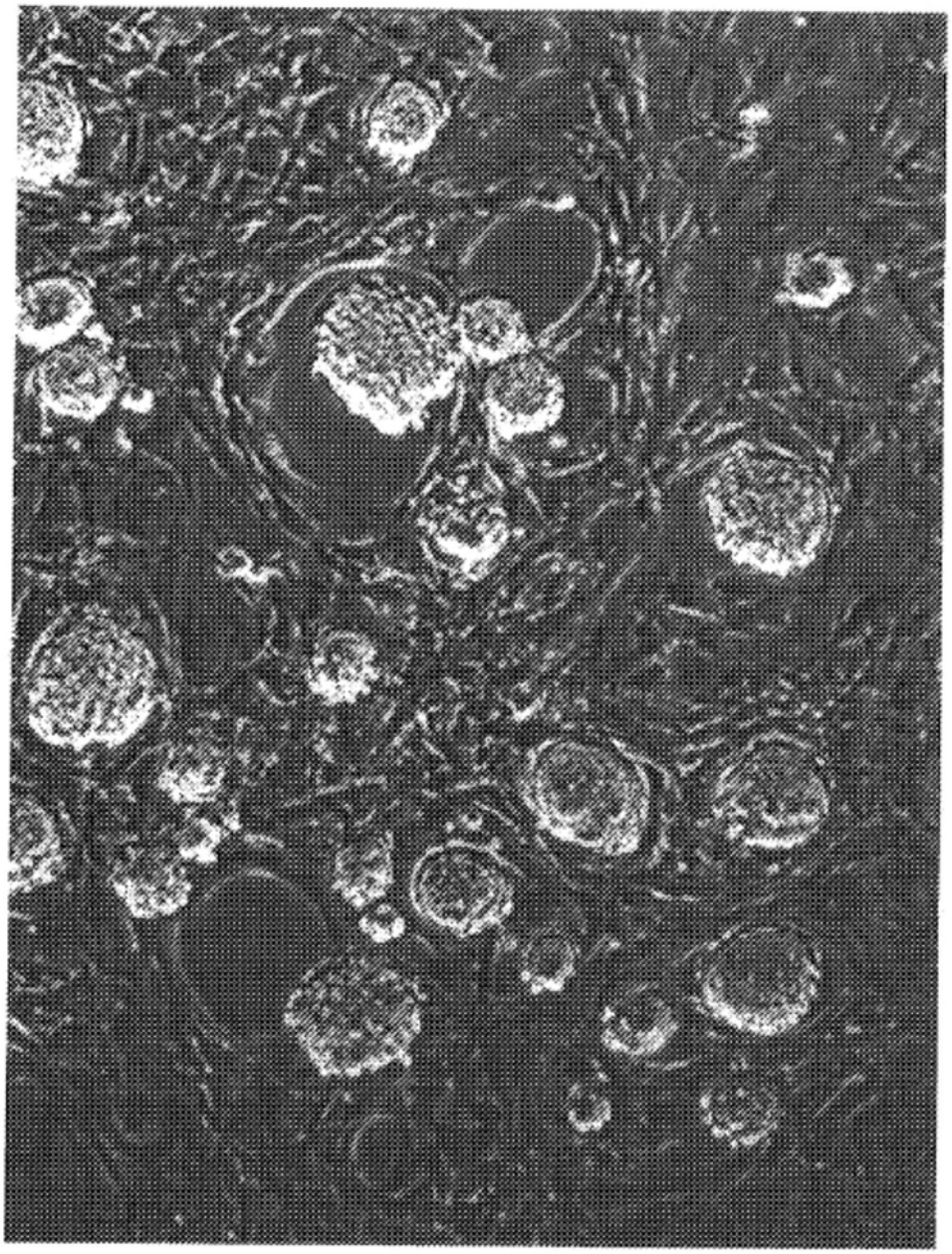


图1

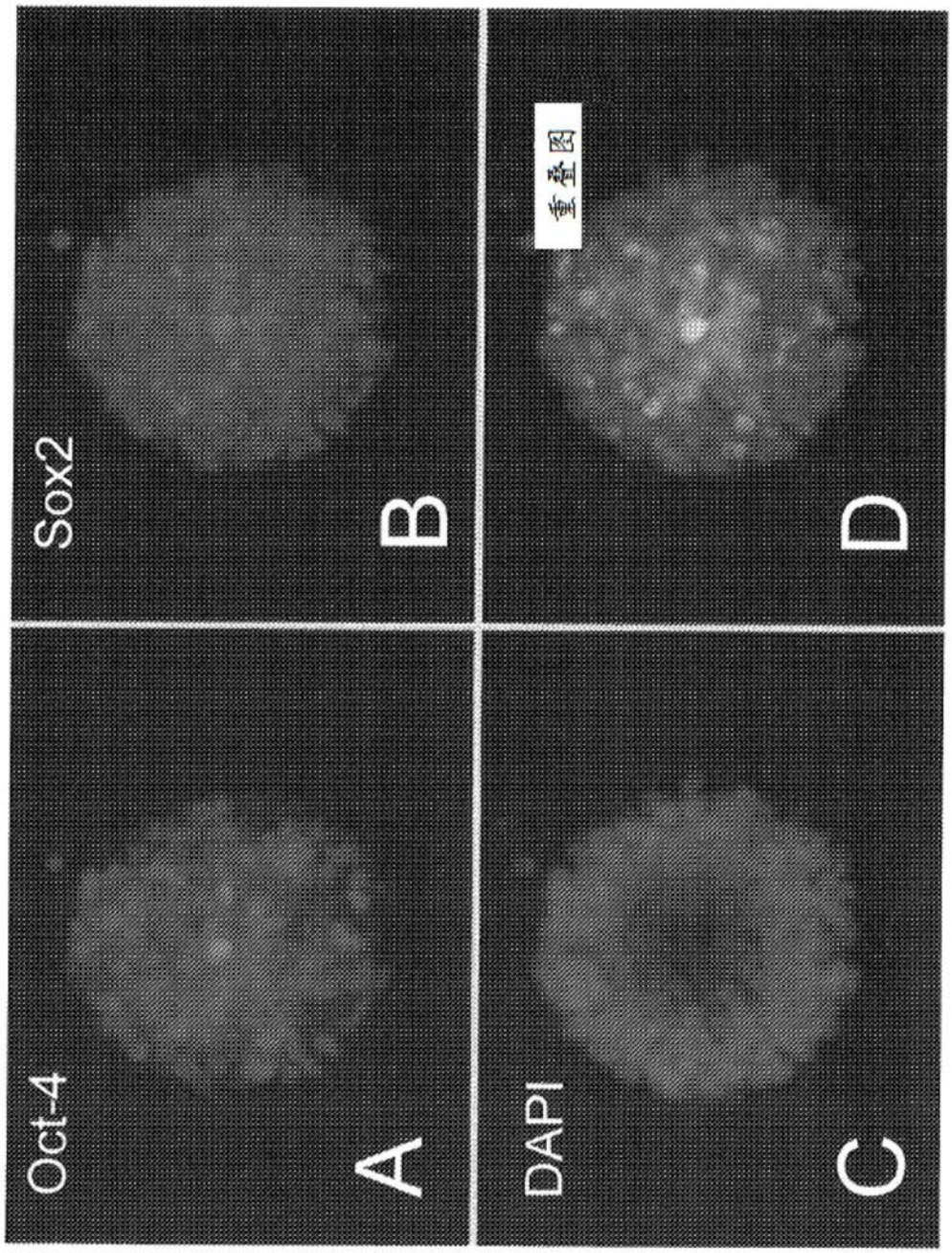


图2

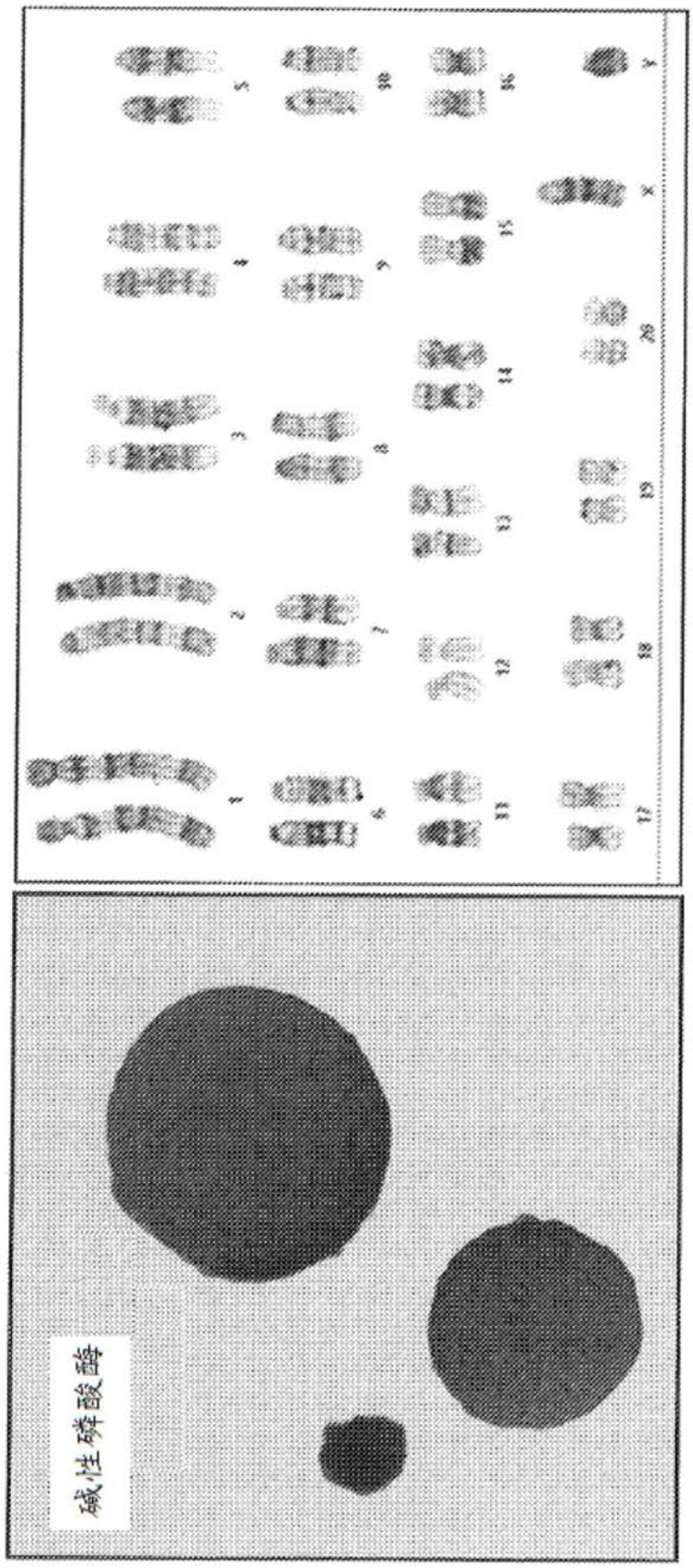


图3



图4a

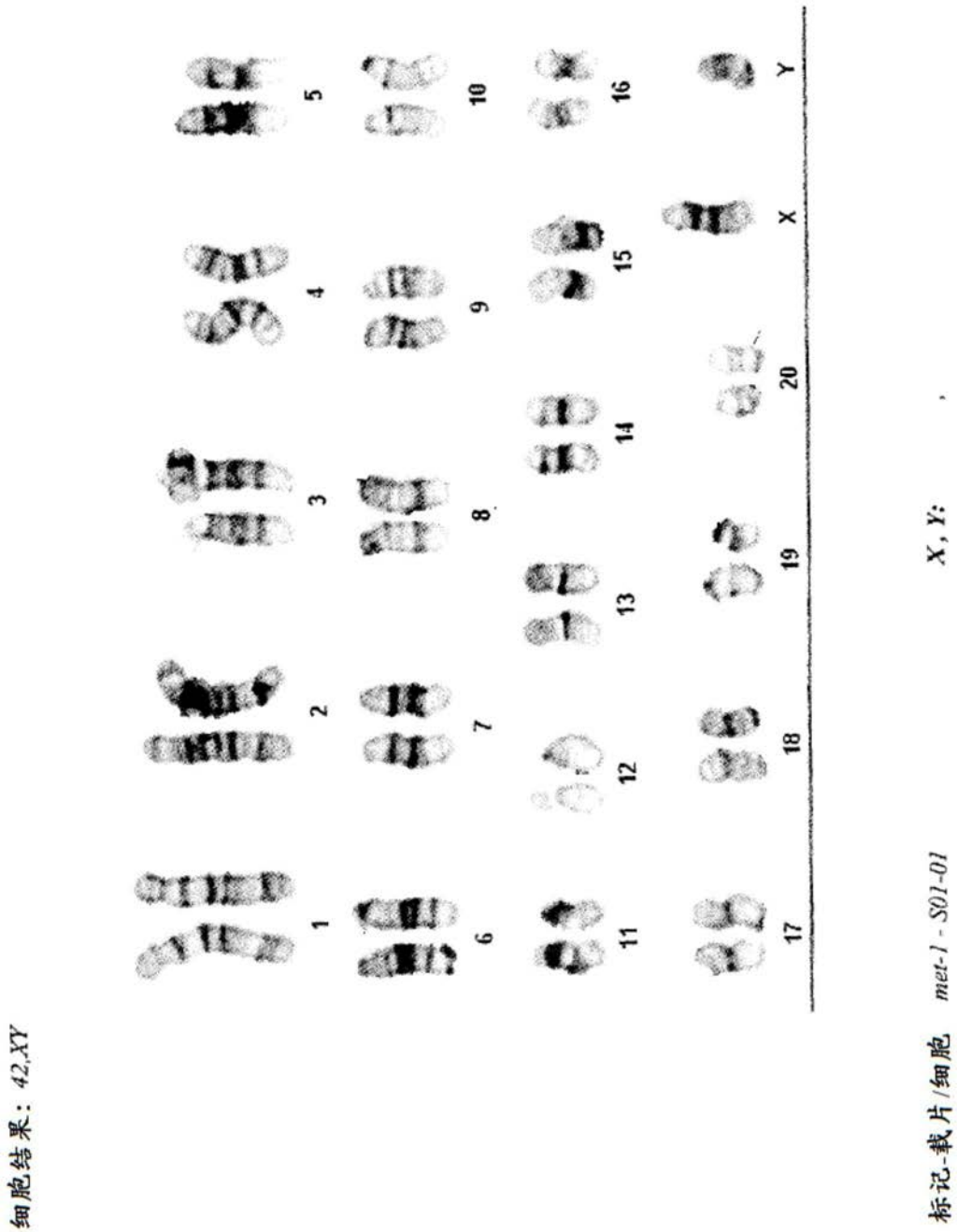

细胞结果：42,XY
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
X
Y
X, Y:
标记-载片/细胞 met-1 - S01-01

图4b

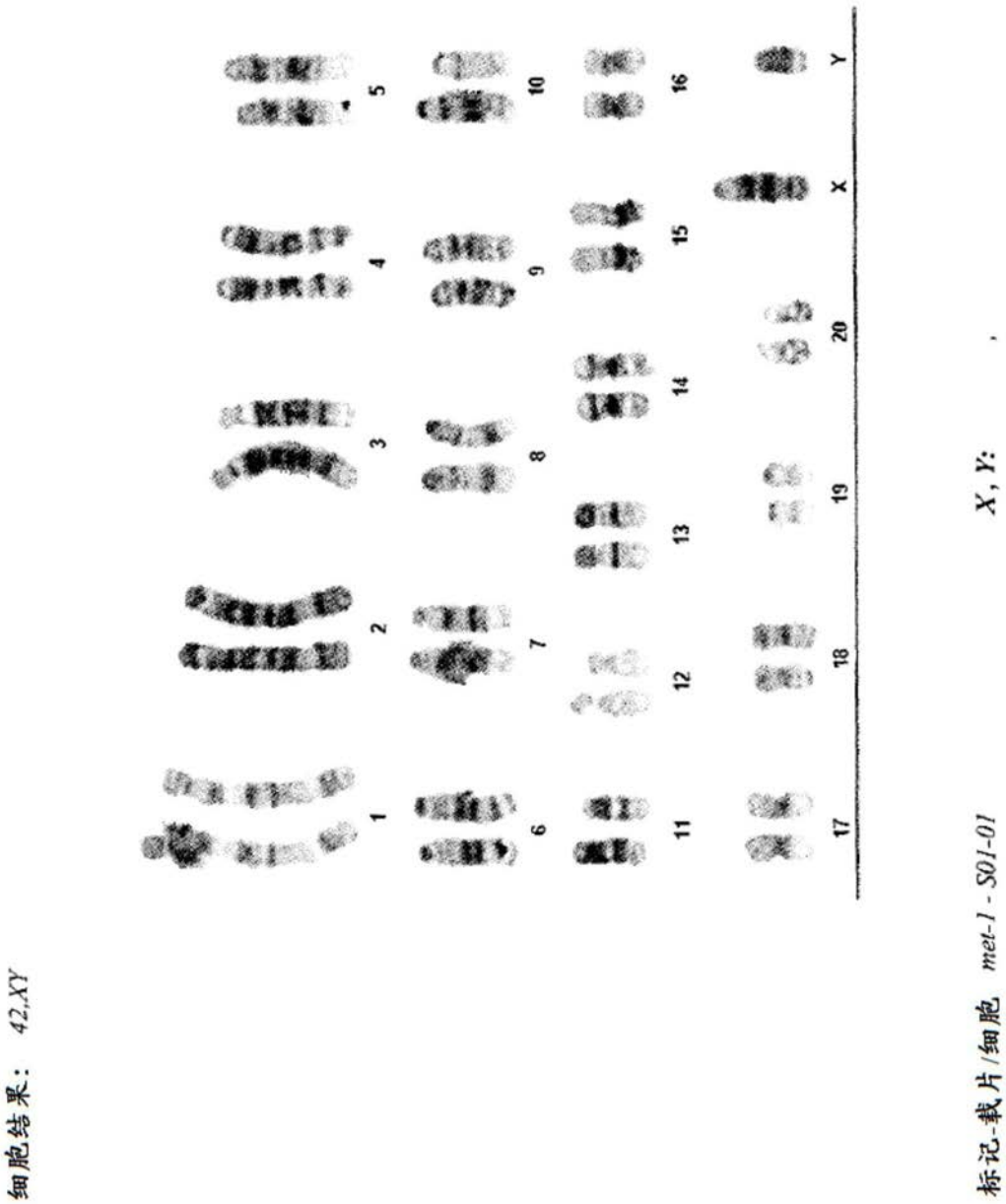

细胞结果： 42,XY
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
X
Y
X, Y:
标记-载片/细胞 met-1 - S01-01

图5a

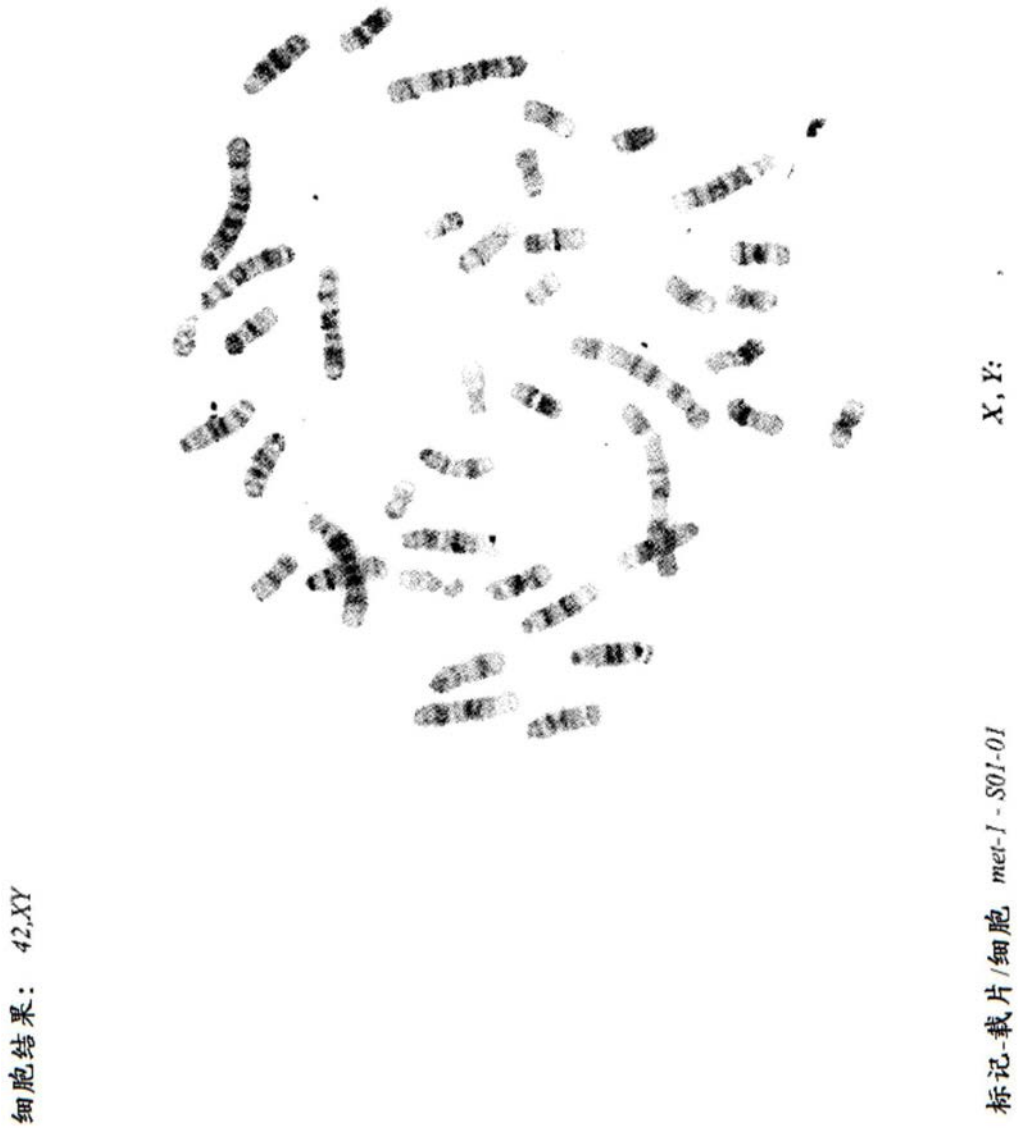


图5b

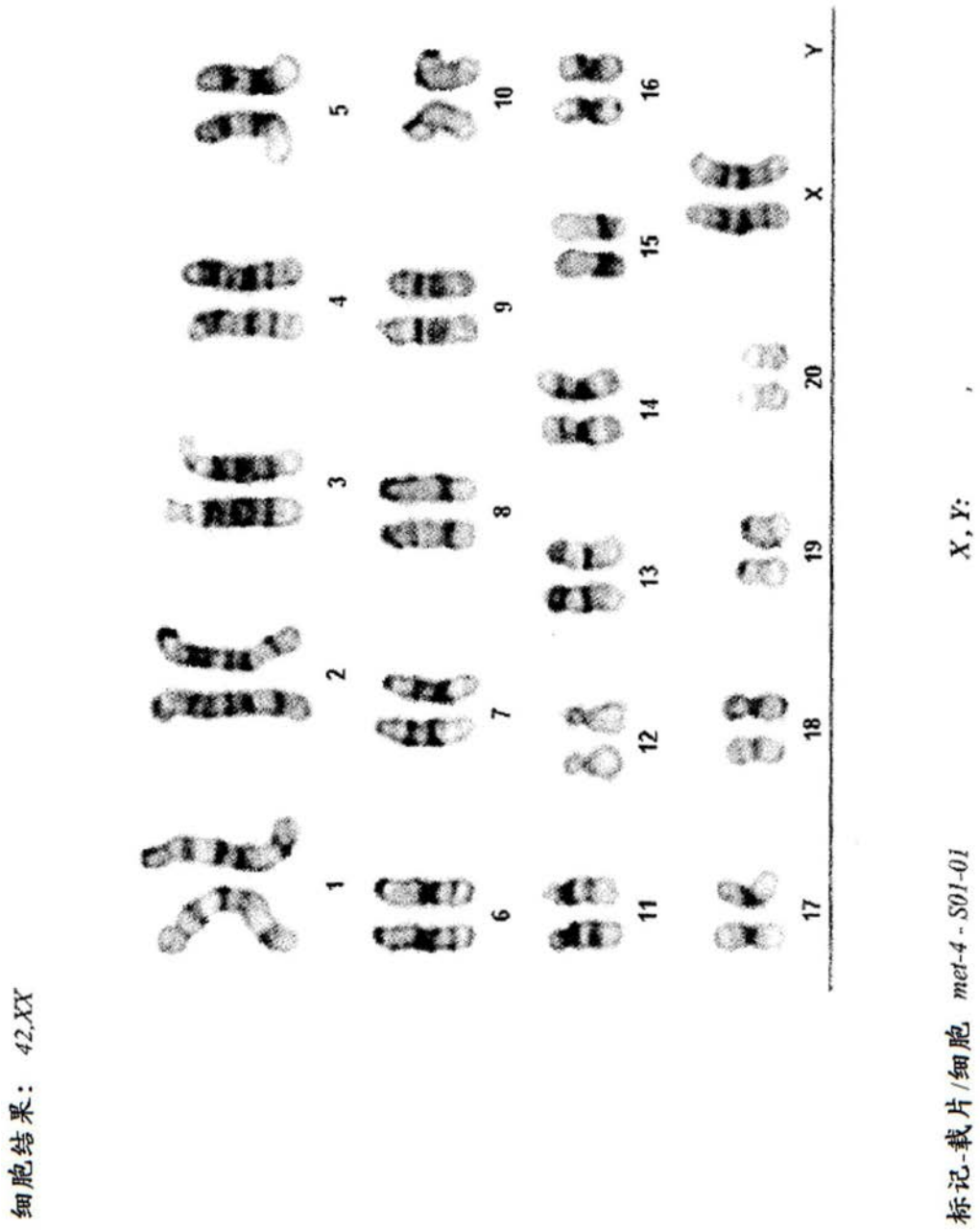

细胞结果： 42,XX
1 2 3 4 5
6 7 8 9 10
11 12 13 14 15 16
17 18 19 20 X Y
X, Y:
标记-载片/细胞 met-4 - S01-01

图6a

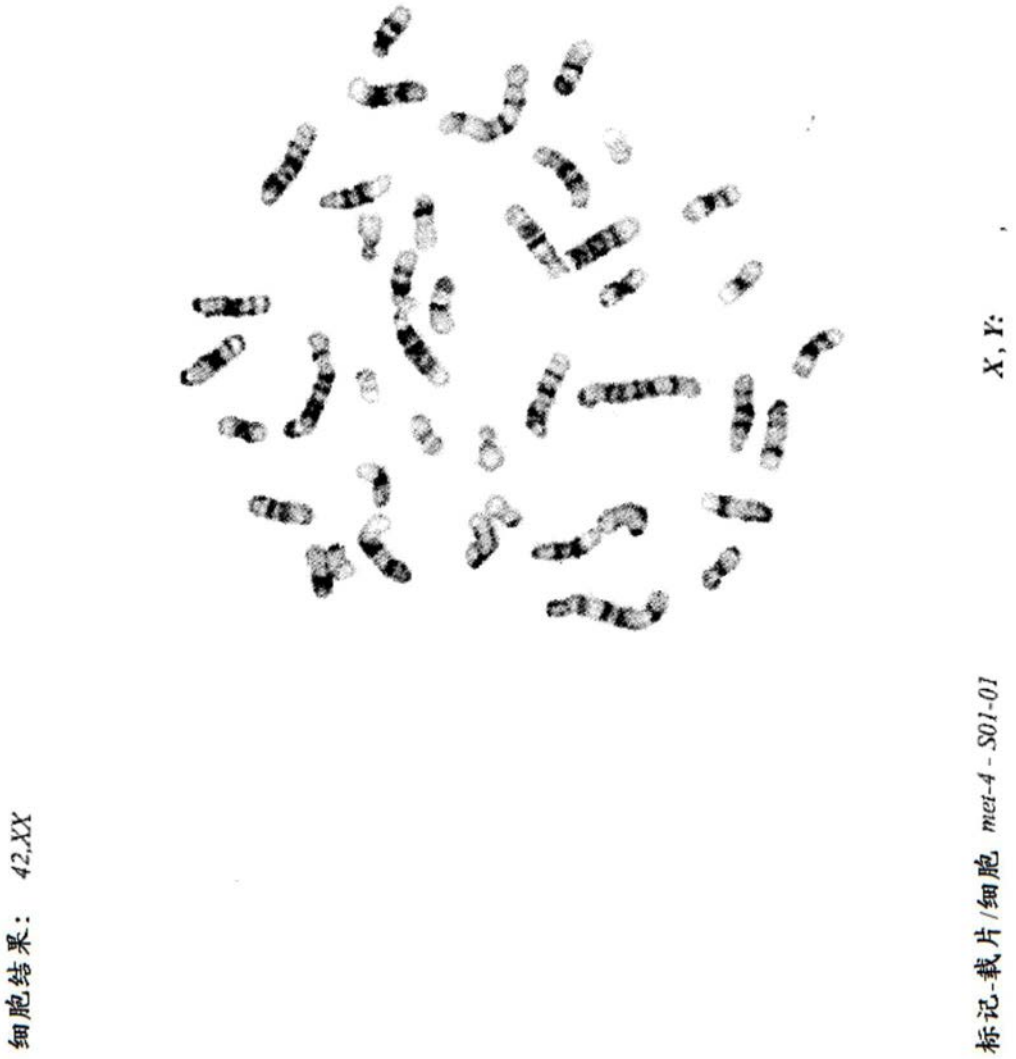


图6b

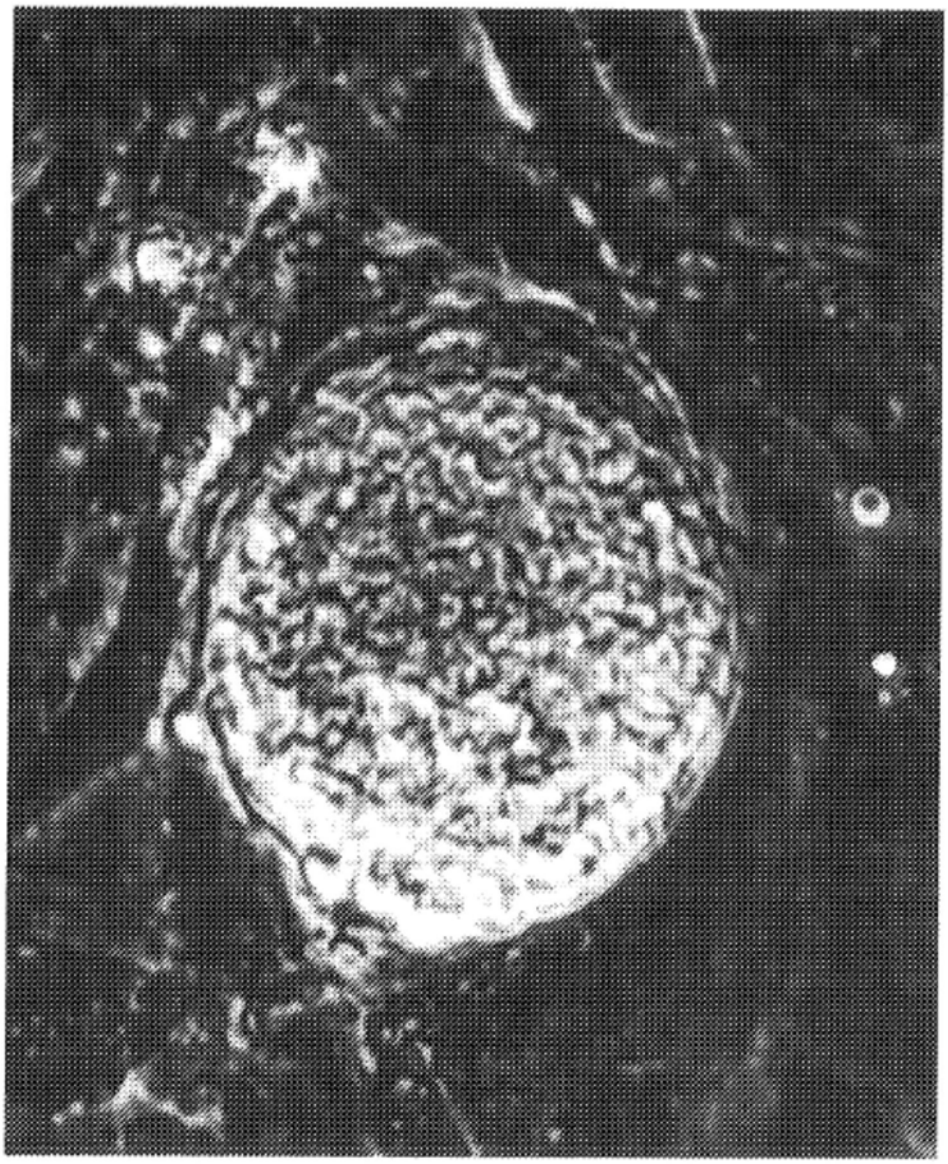


图7

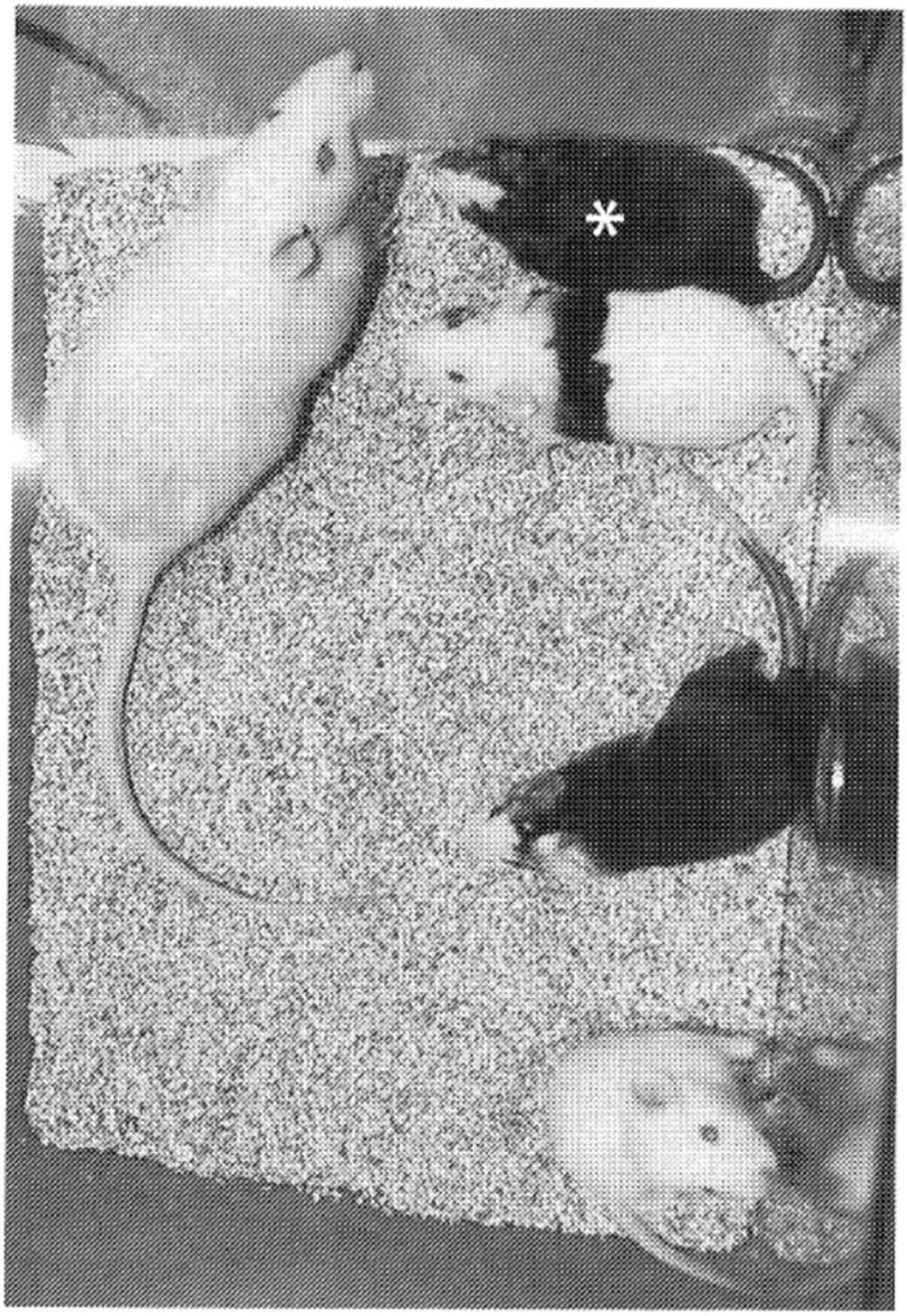


图8

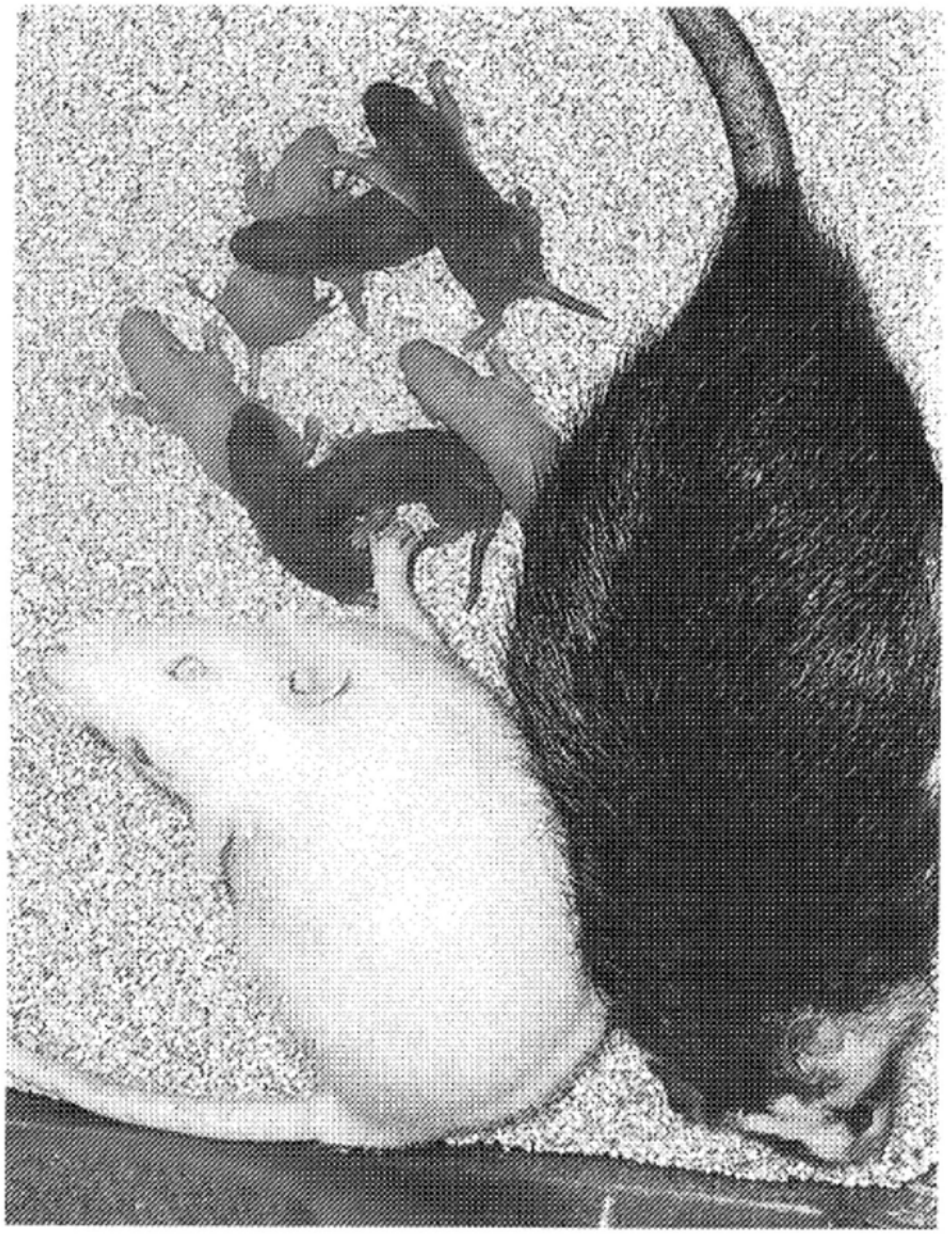


图9

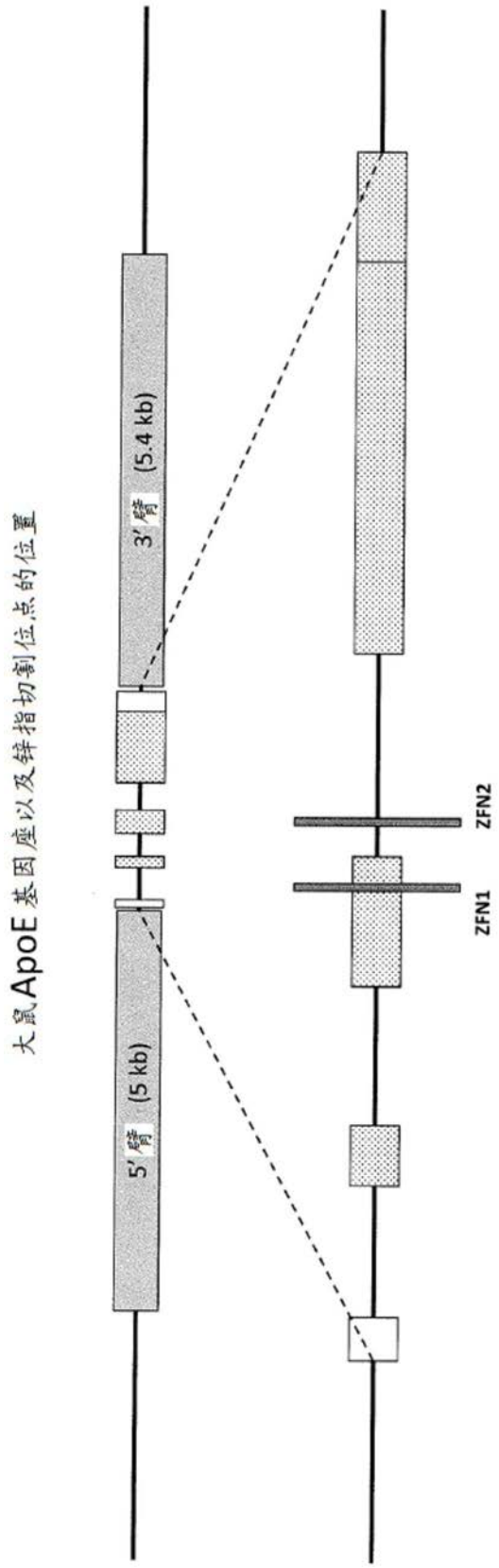
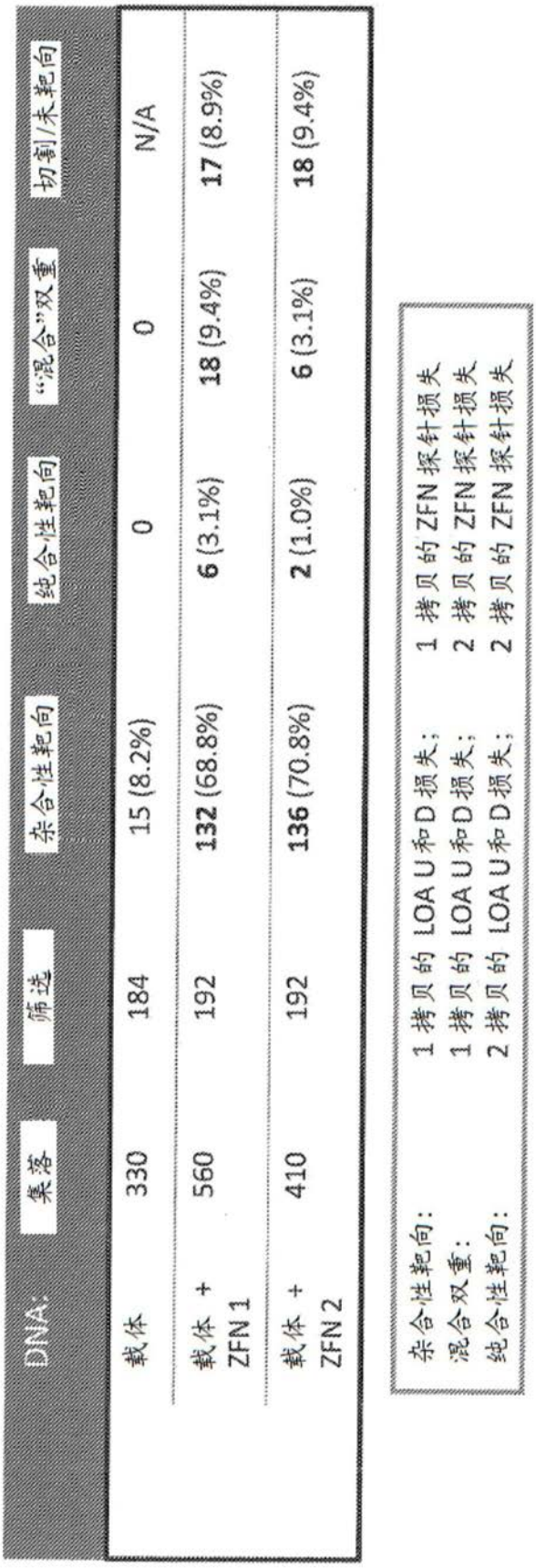


图10

大鼠 ApoE ZFN：改进的靶向效率



DNA:	集落	筛选	杂合性靶向	纯合性靶向	“混合”双重	切割/未靶向
载体	330	184	15 (8.2%)	0	0	N/A
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%)

杂合性靶向：1 拷贝的 LOA U 和 D 损失；1 拷贝的 ZFN 探针损失
混合双重：1 拷贝的 LOA U 和 D 损失；2 拷贝的 ZFN 探针损失
纯合性靶向：2 拷贝的 LOA U 和 D 损失；2 拷贝的 ZFN 探针损失

图11

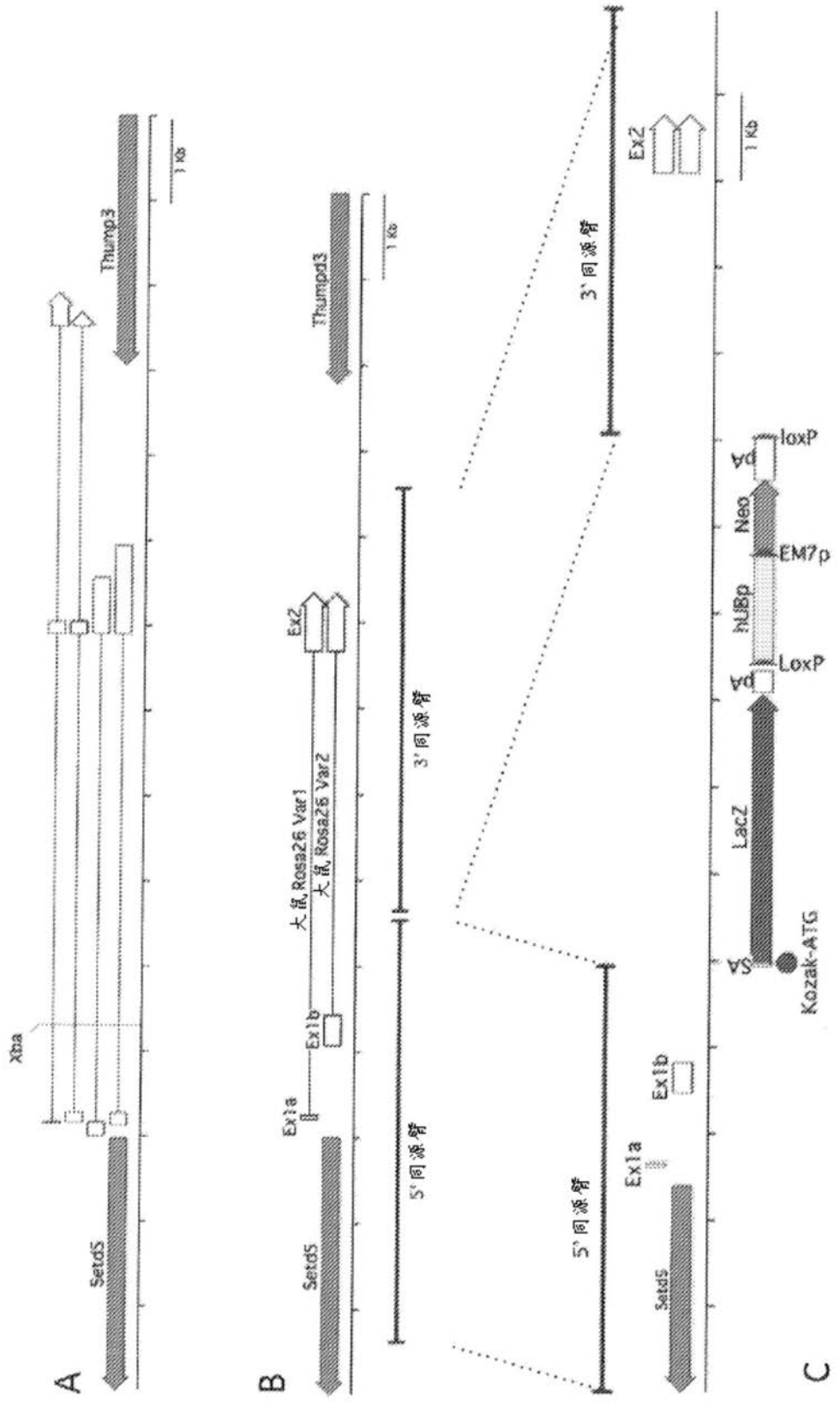

A
Xba
Setd5
Thumpd3
1 Kb
B
Ex1a
Ex1b
Ex2
大鼠Rosa26 Var1
大鼠Rosa26 Var2
5'同源臂
3'同源臂
C
loxP
Neo
EM7p
NEBp
LoxP
LacZ
Kozak-ATG

图12

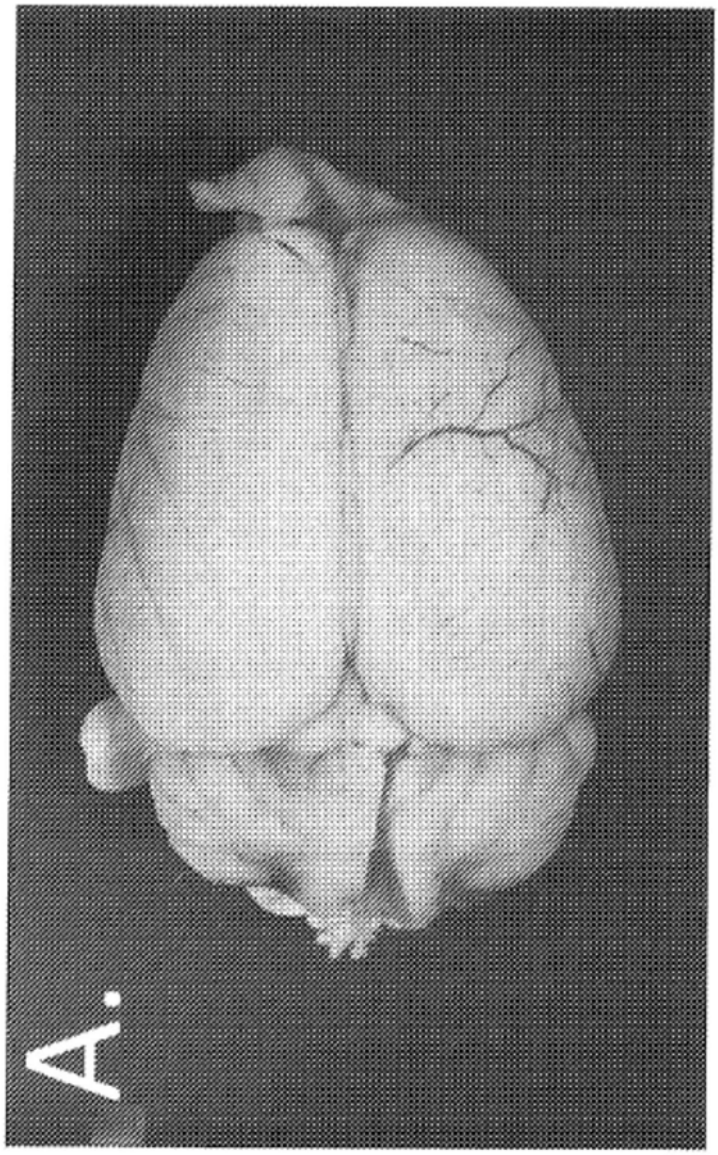


图13a

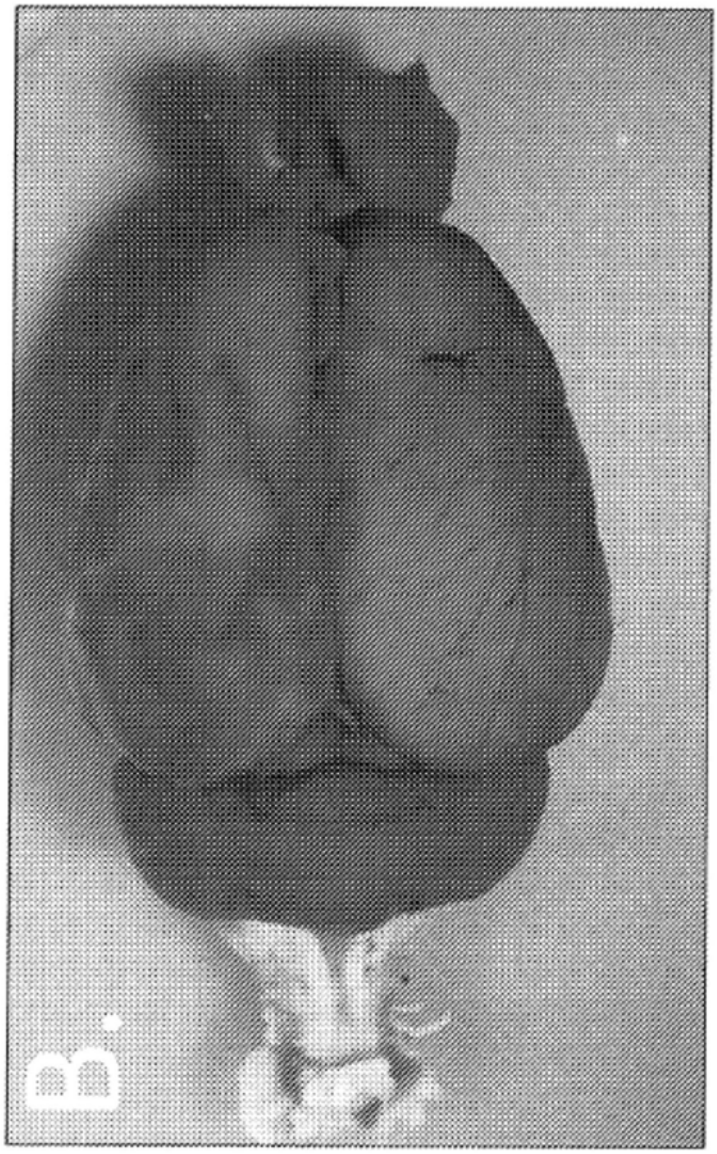


图13b

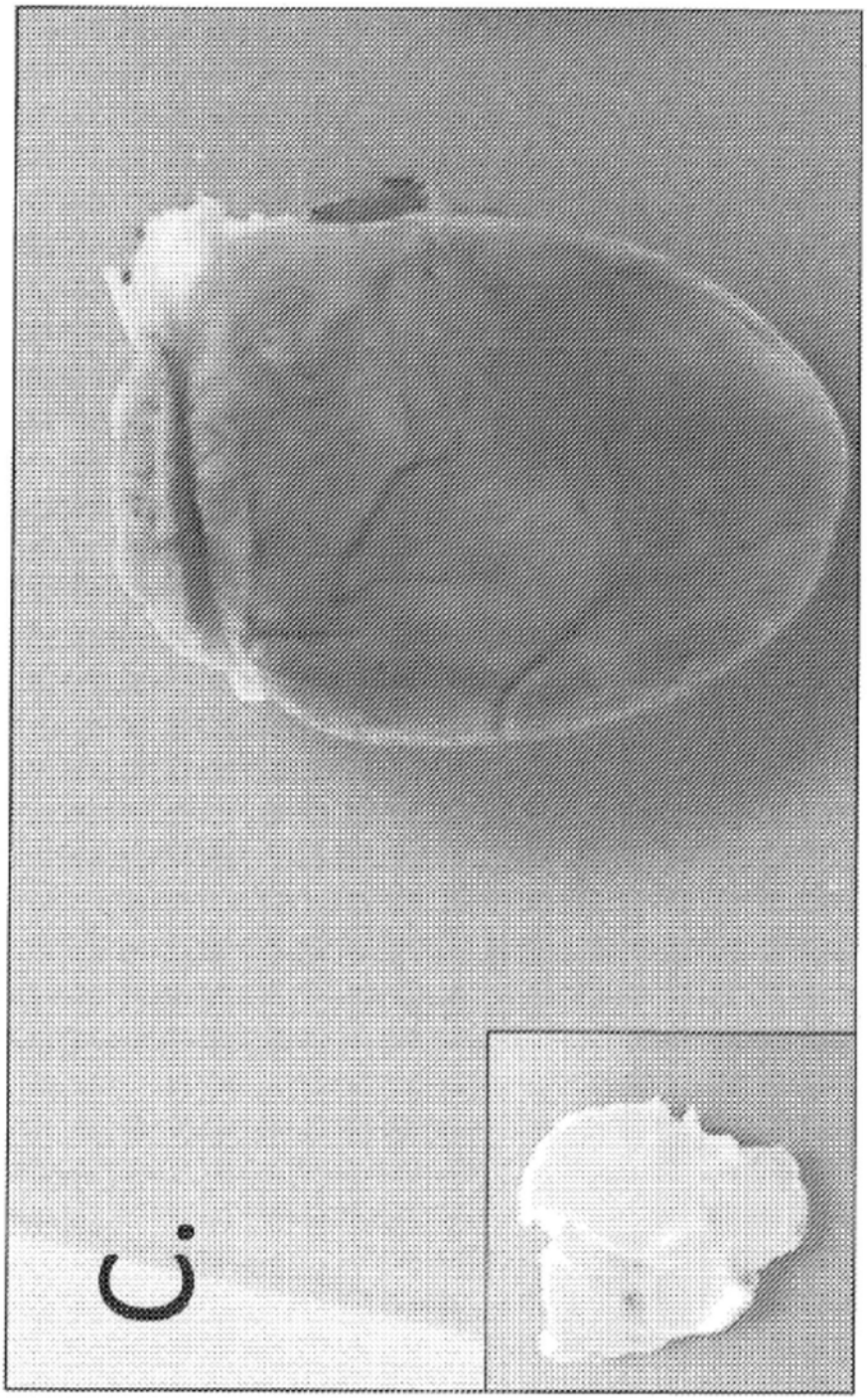


图13c

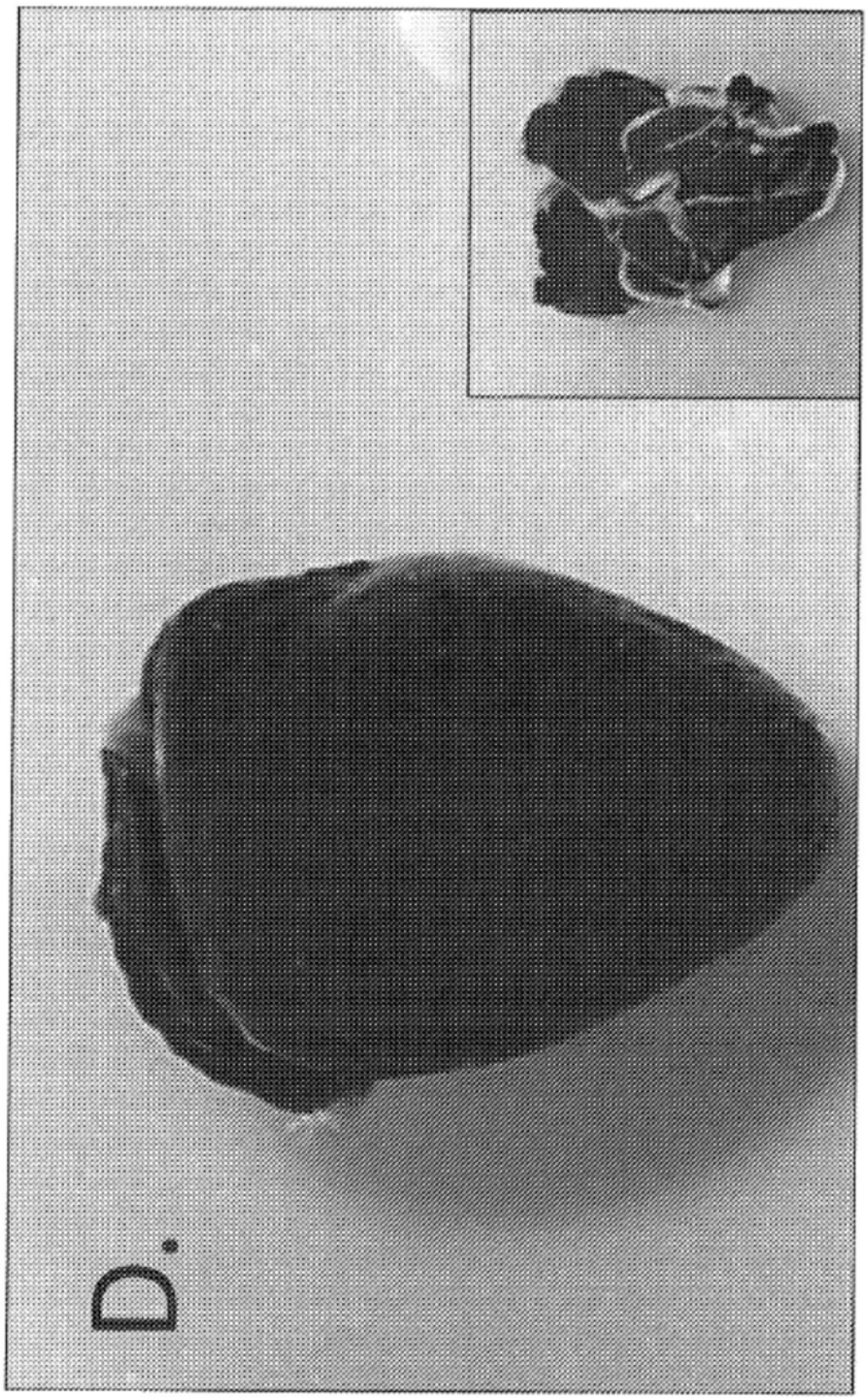


图13d

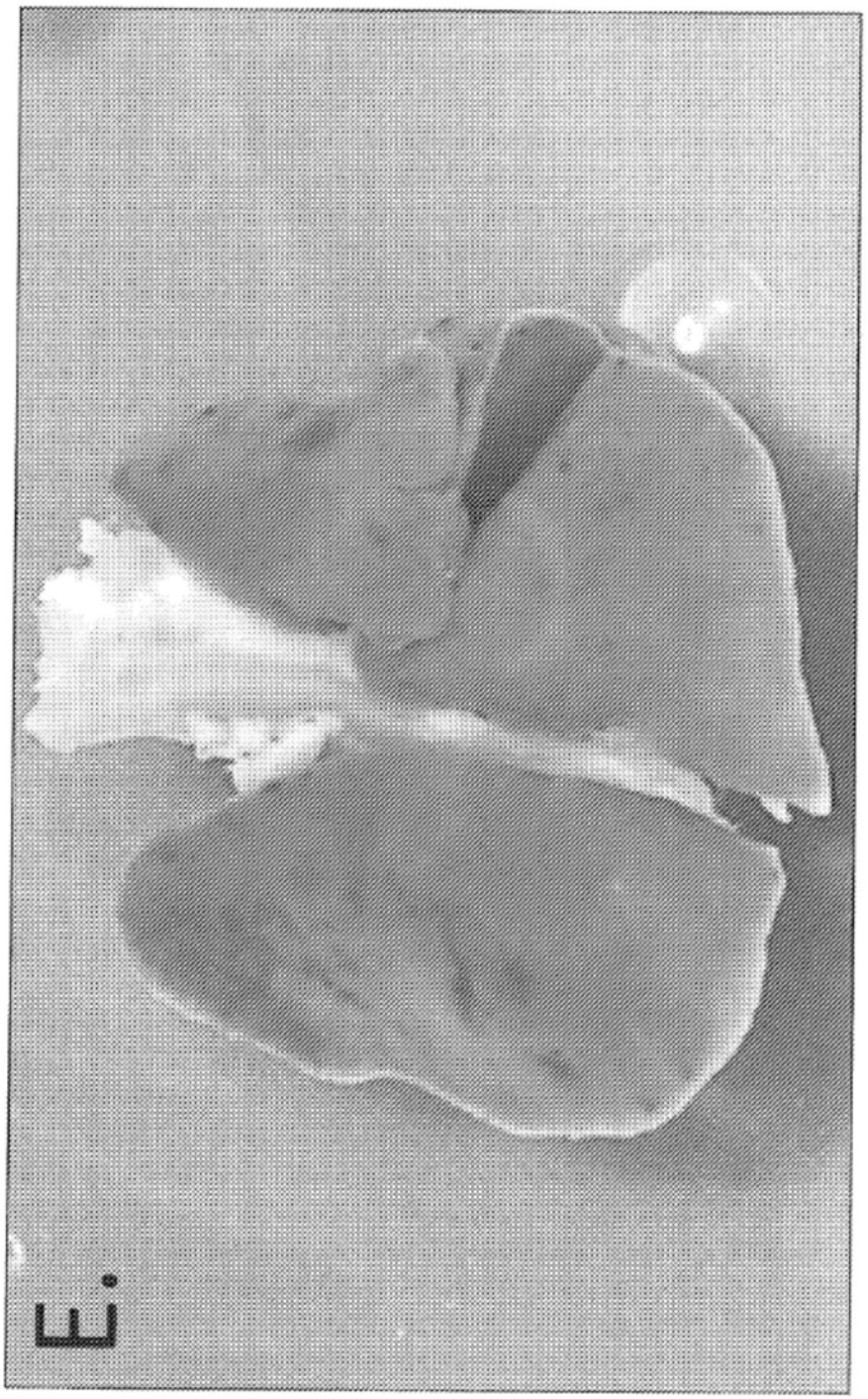


图13e

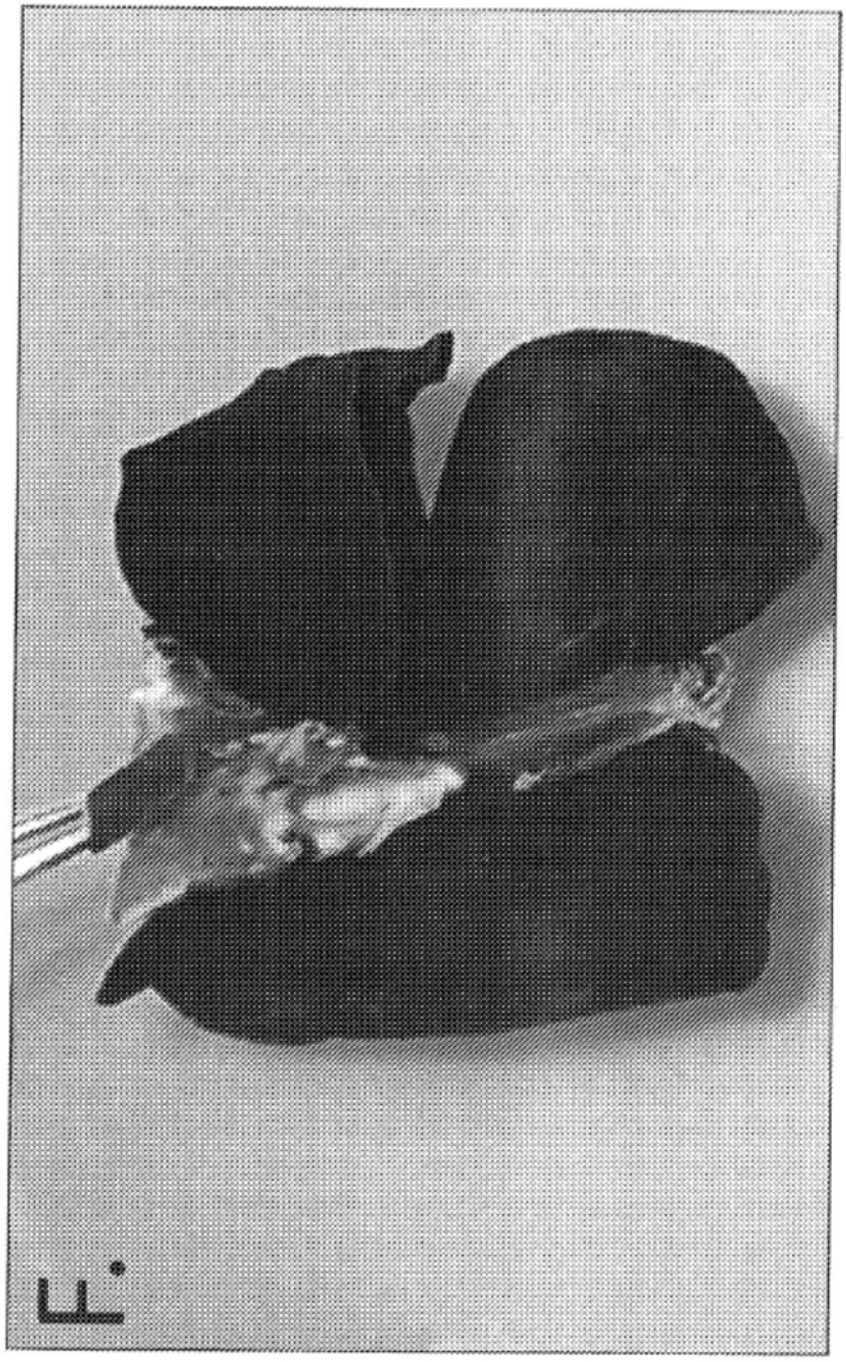
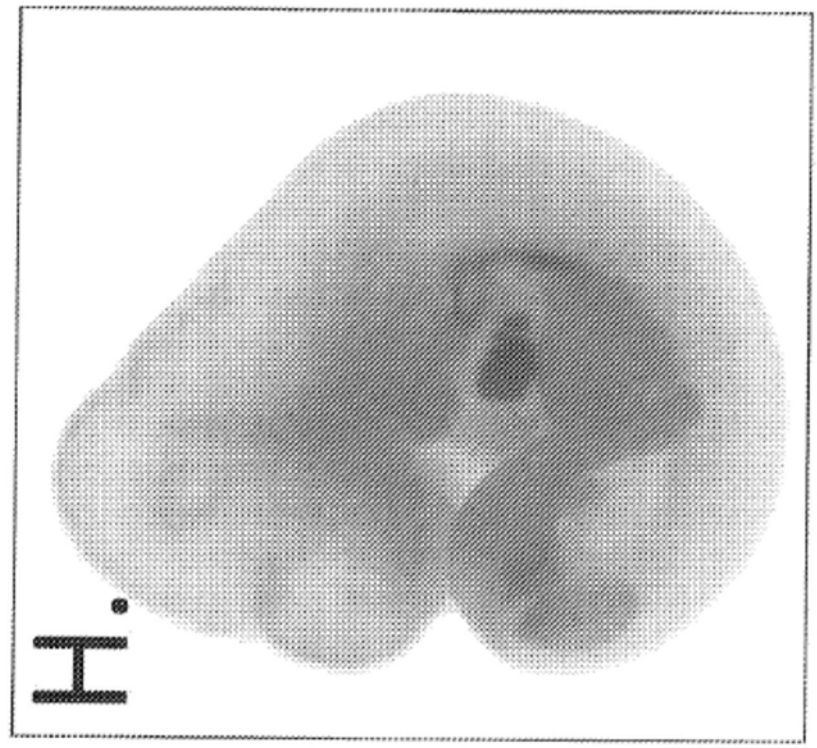


图13F



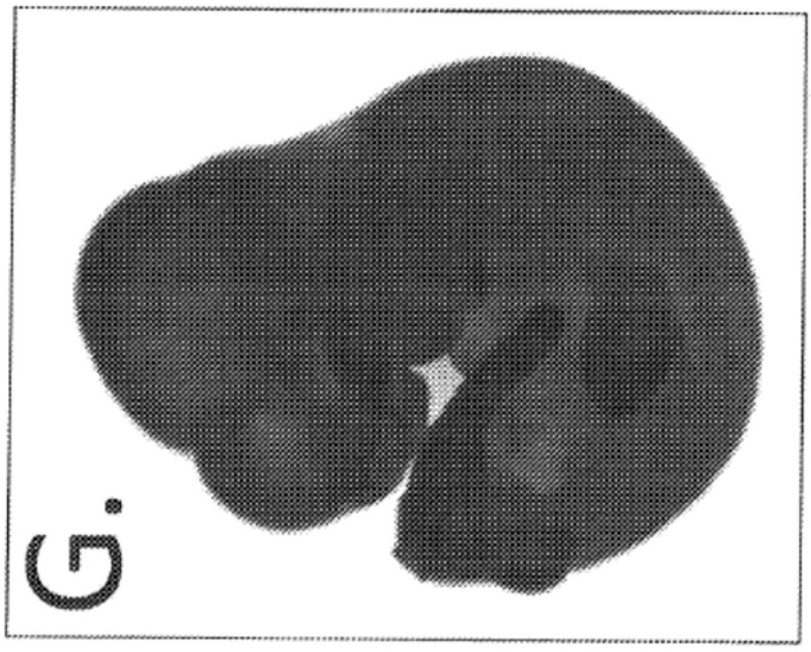


图13g,h

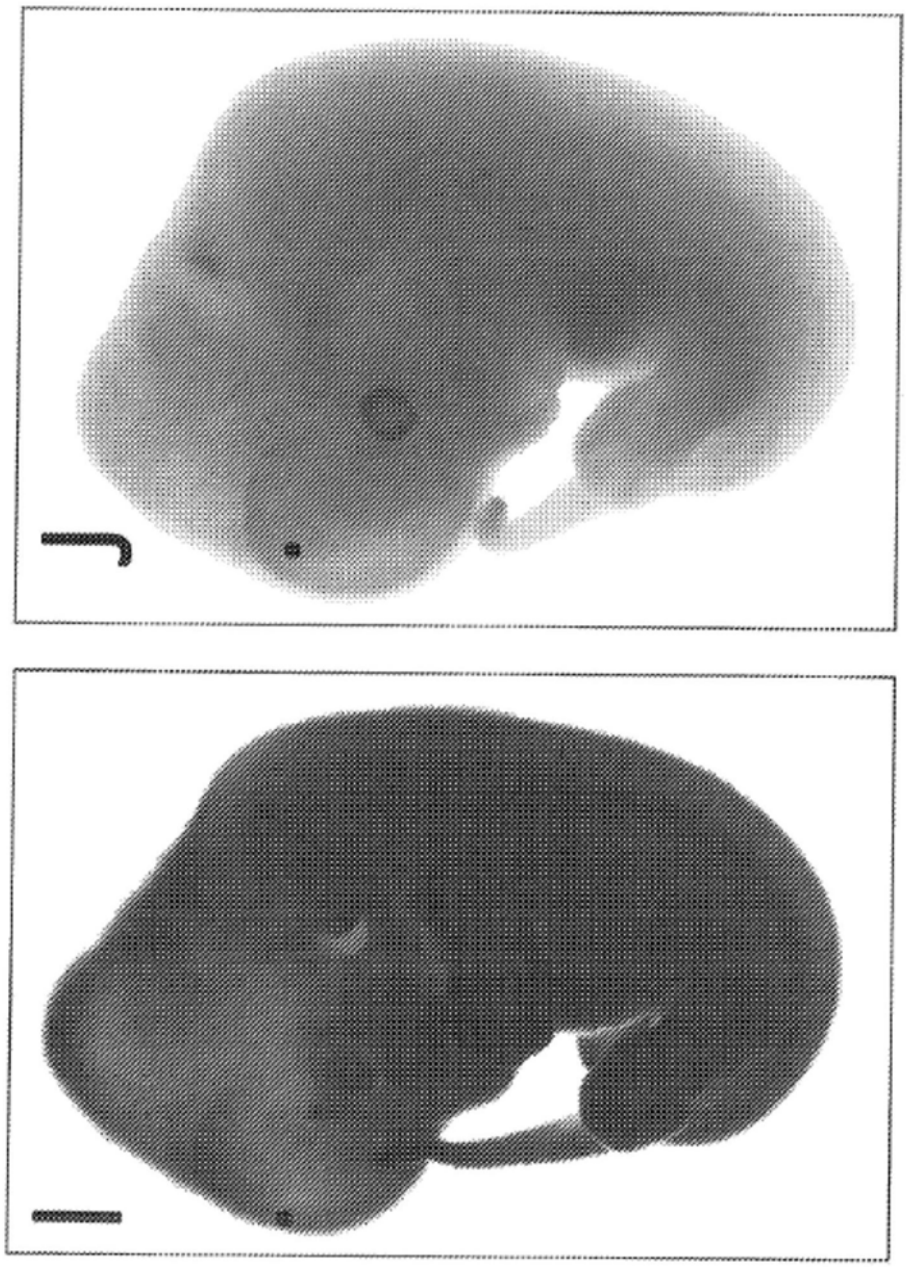


图13i,j

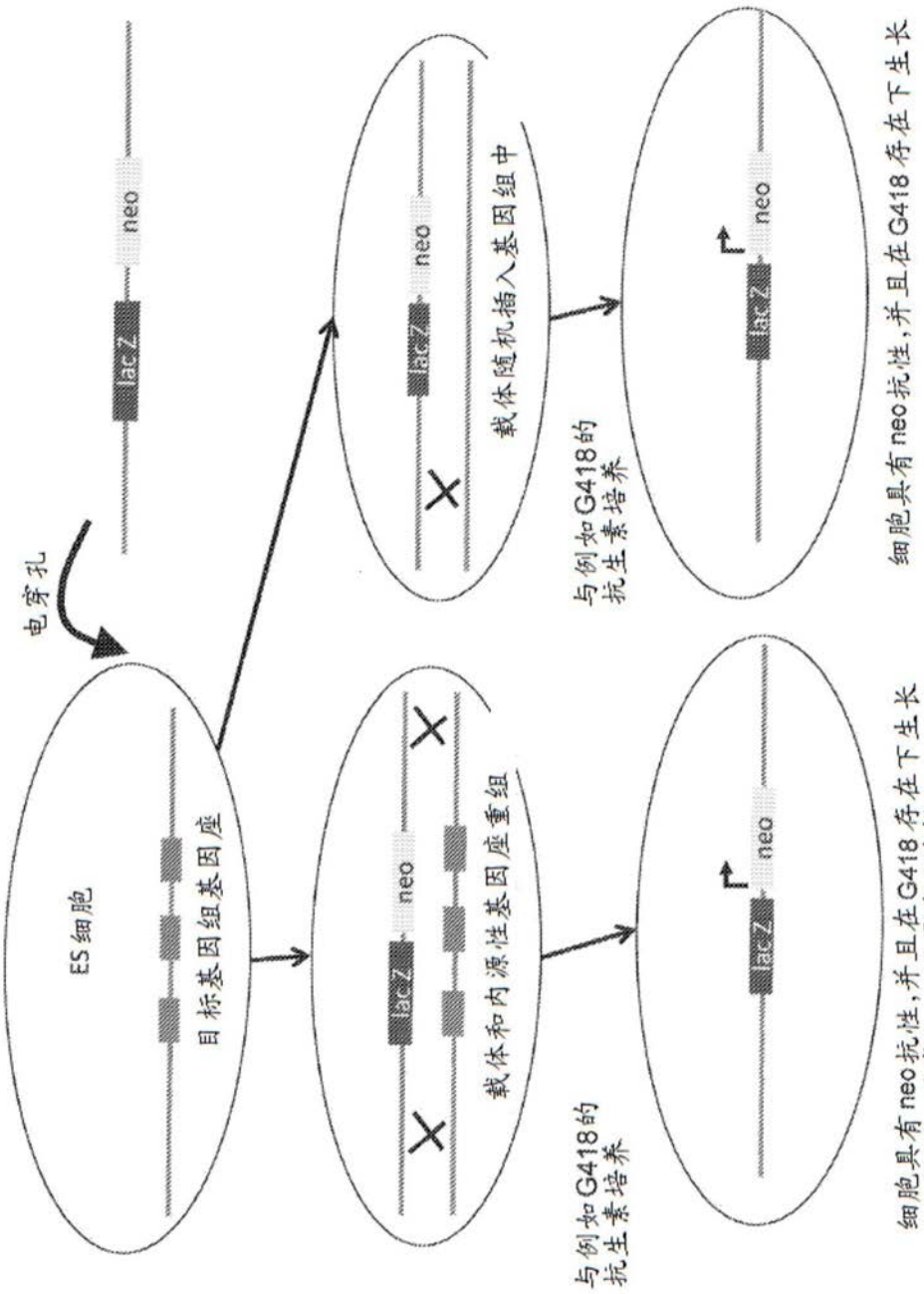

neo
lac Z
电穿孔
ES细胞
目标基因组基因座
载体随机插入基因组中
载体和内源性基因座重组
与例如G418的
抗生素培养
细胞具有neo抗性，并且在G418存在下生长
细胞具有neo抗性，并且在G418存在下生长
通常1-20%（最多）的插入

图14

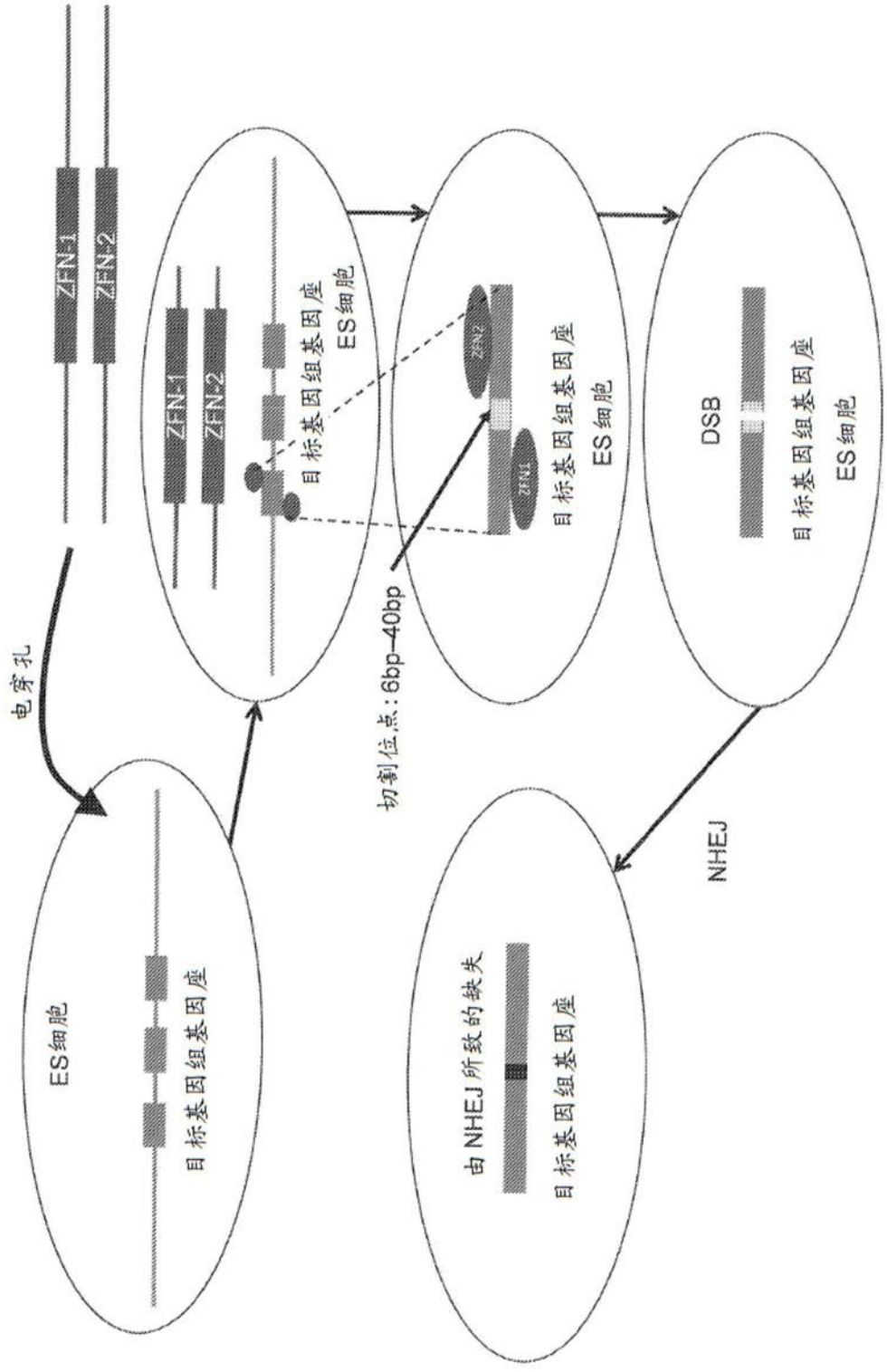

ZFN-1
ZFN-2
电穿孔
ES细胞
目标基因组基因座
ZFN-1
ZFN-2
目标基因组基因座
ES细胞
切割位点：6bp–40bp
ZFN2
ZFN1
目标基因组基因座
ES细胞
DSB
目标基因组基因座
ES细胞
NHEJ
由NHEJ所致的缺失
目标基因组基因座

图15

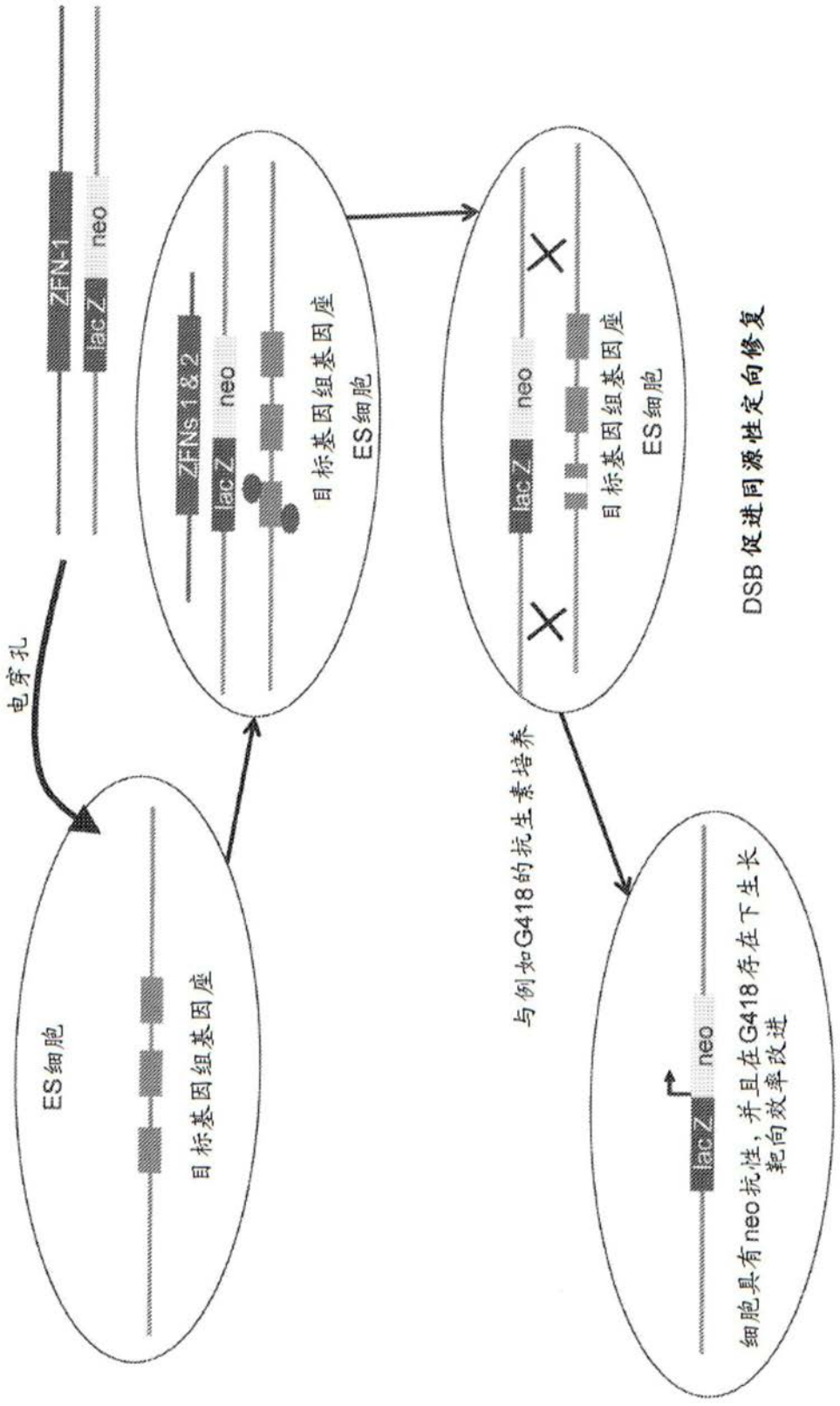

ZFN-1
lac Z
neo
电穿孔
ES细胞
目标基因组基因座
ZFNs 1 & 2
lac Z
neo
目标基因组基因座
ES细胞
DSB促进同源性定向修复
lac Z
neo
目标基因组基因座
ES细胞
与例如G418的抗生素培养
lac Z
neo
细胞具有neo抗性，并且在G418存在下生长
靶向效率改进

图16

克隆		幼仔	嵌合体
ApoE-ZFN1-AB5（纯合性靶向）		12	4 (90, 90, 80, 80)
ApoE-ZFN1-AE5（纯合性靶向）		6	3 (90, 80, 70)



ApoE-ZFN1-AB5
嵌合体

图17

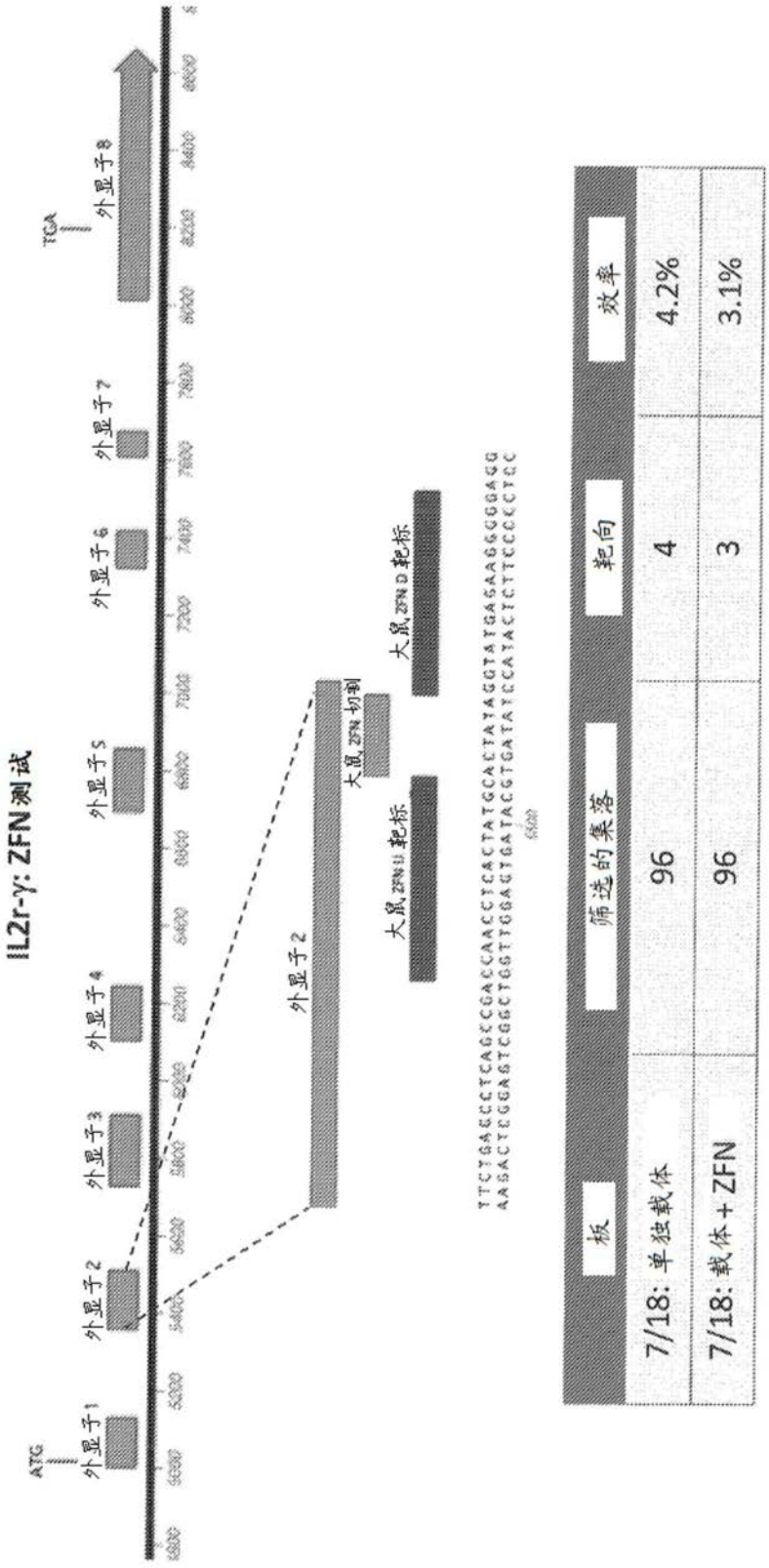
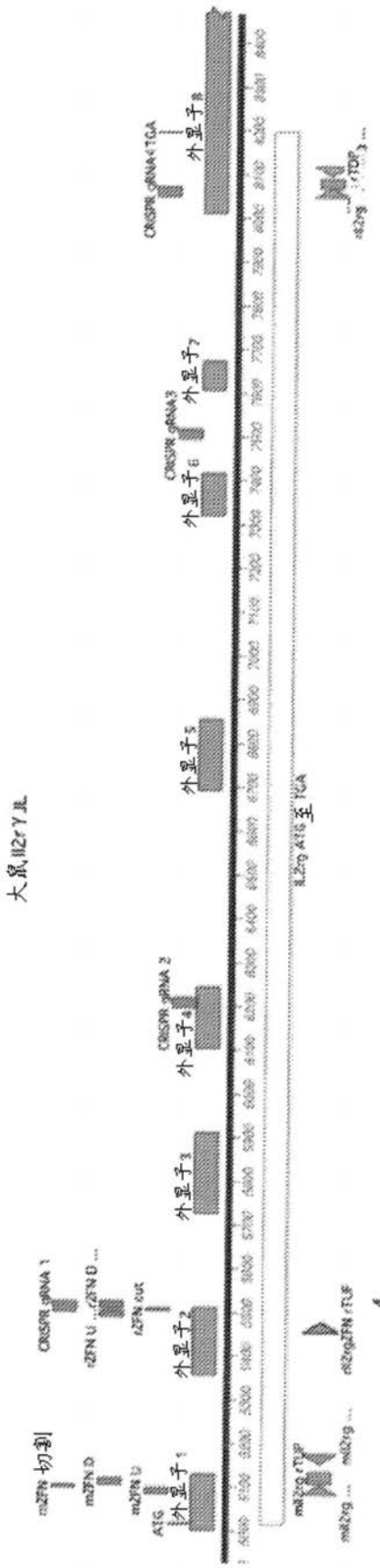

IL2r-γ: ZFN测试
ATG
TGA
外显子1
外显子2
外显子3
外显子4
外显子5
外显子6
外显子7
外显子8
外显子2
大鼠ZFN切割
大鼠ZFN D靶标
大鼠ZFN U靶标
板
筛选的集落
靶向
效率
7/18: 单独载体
96
4
4.2%
7/18: 载体 + ZFN
96
3
3.1%

图18



构建体	DNA (μg)	集落	候选物
Il2rg 质粒载体	6 μg	30	3（全都"弱"）
质粒 + SBI gRNA1	6 μg/40 μg	22	1强，1弱
质粒 + SBI gRNA2	6 μg/40 μg	45	2强，1弱
质粒 + SBI gRNA3	6 μg/40 μg	66	1强，2弱
质粒 + SBI gRNA4	6 μg/40 μg	59	0

图19

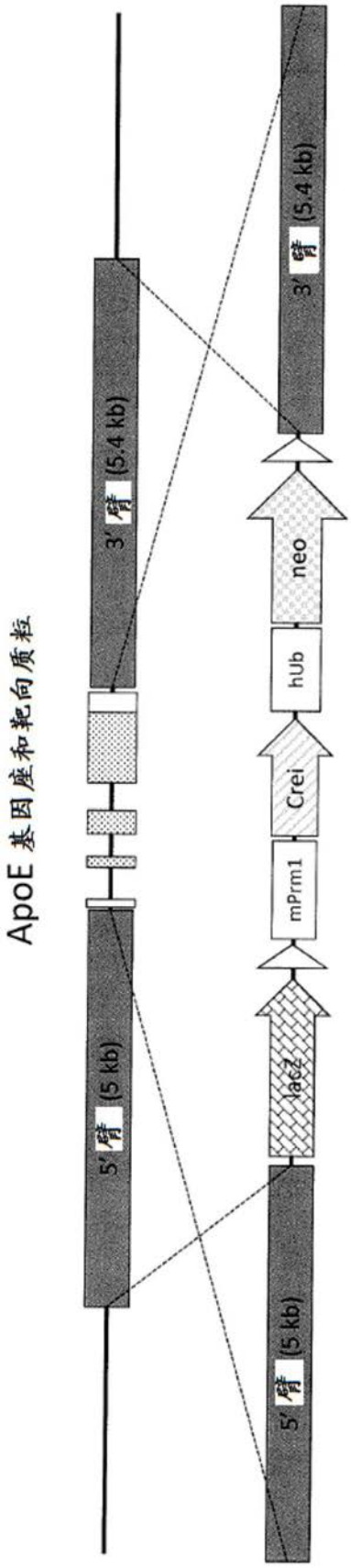

ApoE 基因座和靶向质粒
5′臂 (5 kb)
3′臂 (5.4 kb)
lacZ
mPrm1
Crei
hUb
neo
5′臂 (5 kb)
3′臂 (5.4 kb)

图20

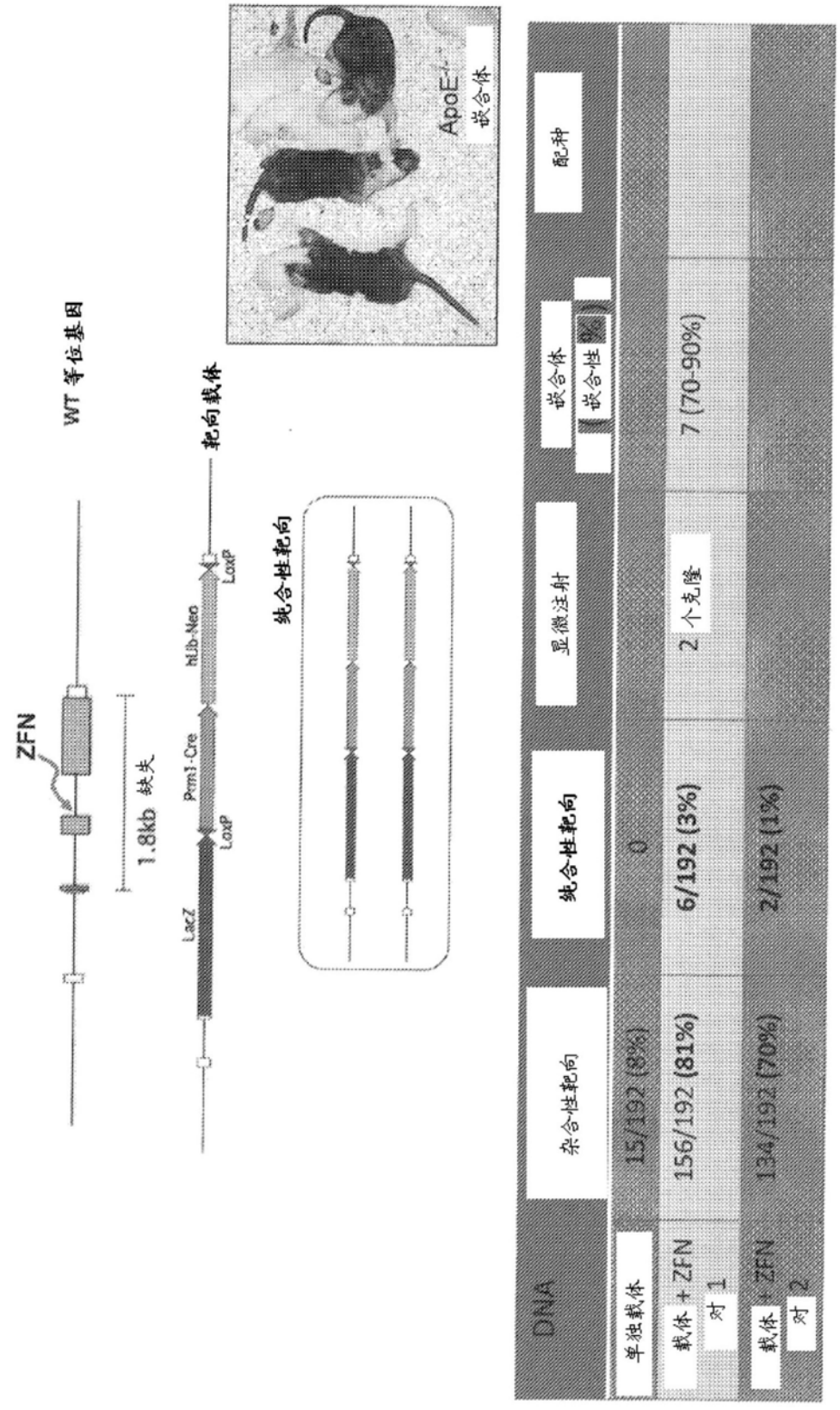

WT 等位基因
ZFN
1.8kb 缺失
LacZ
Prm1-Cre
hUb-Neo
LoxP
LoxP
LoxP
靶向载体
纯合性靶向
ApoE-/-
嵌合体
DNA
杂合性靶向
纯合性靶向
显微注射
嵌合体
嵌合性 %
配种
单独载体
0
载体 + ZFN 对 1
15/192 (8%)
156/192 (81%)
6/192 (3%)
2 个克隆
7 (70-90%)
载体 + ZFN 对 2
134/192 (70%)
2/192 (1%)

图21

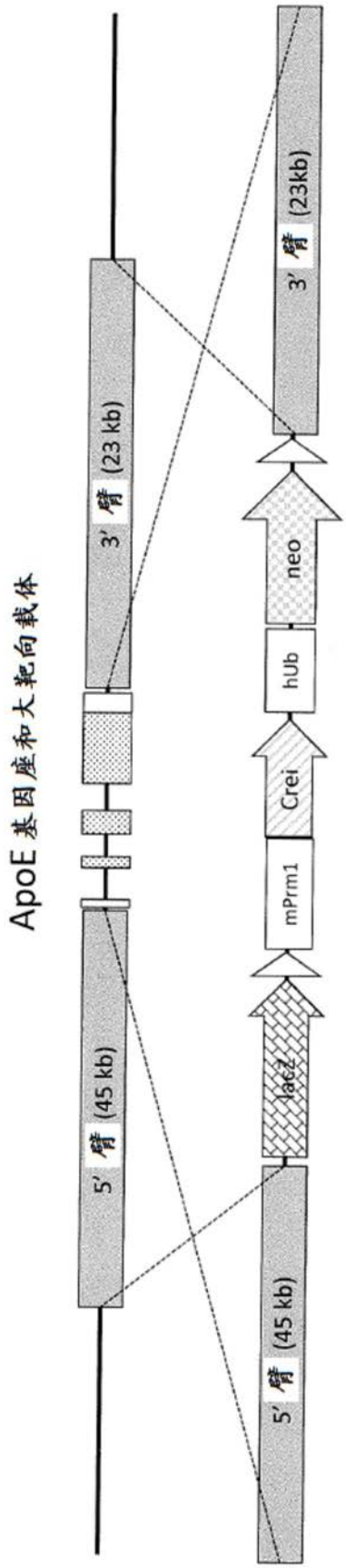

ApoE 基因座和大靶向载体
5′臂 (45 kb)
3′臂 (23 kb)
lacZ
mPrm1
Crei
hUb
neo
5′臂 (45 kb)
3′臂 (23kb)

图22

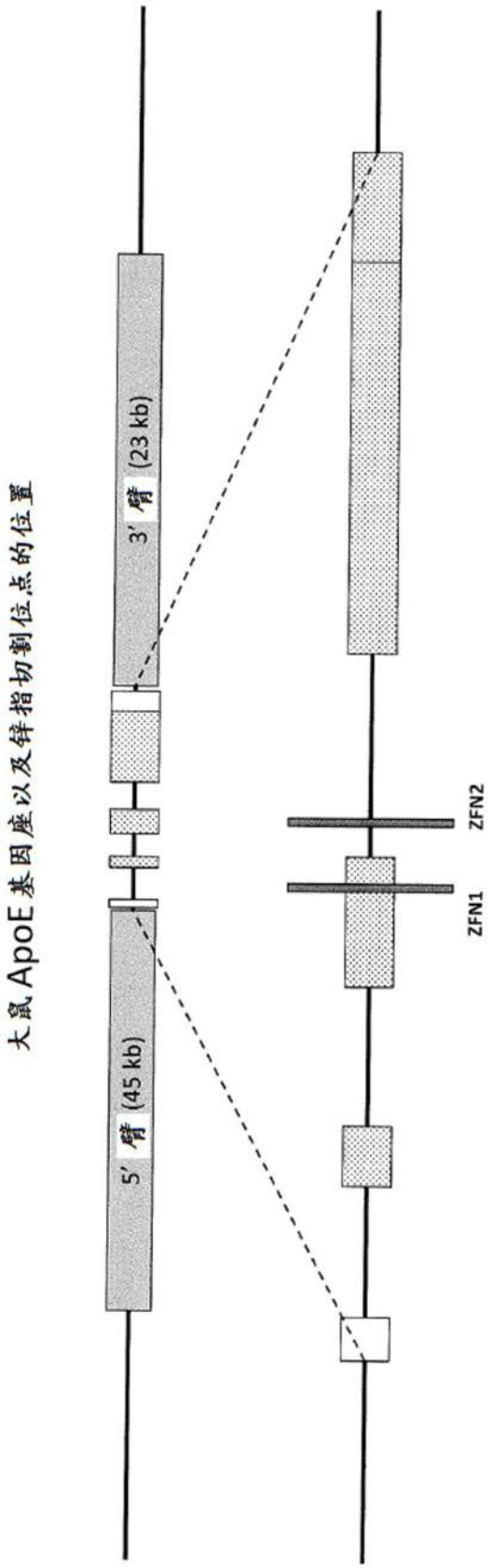


图23

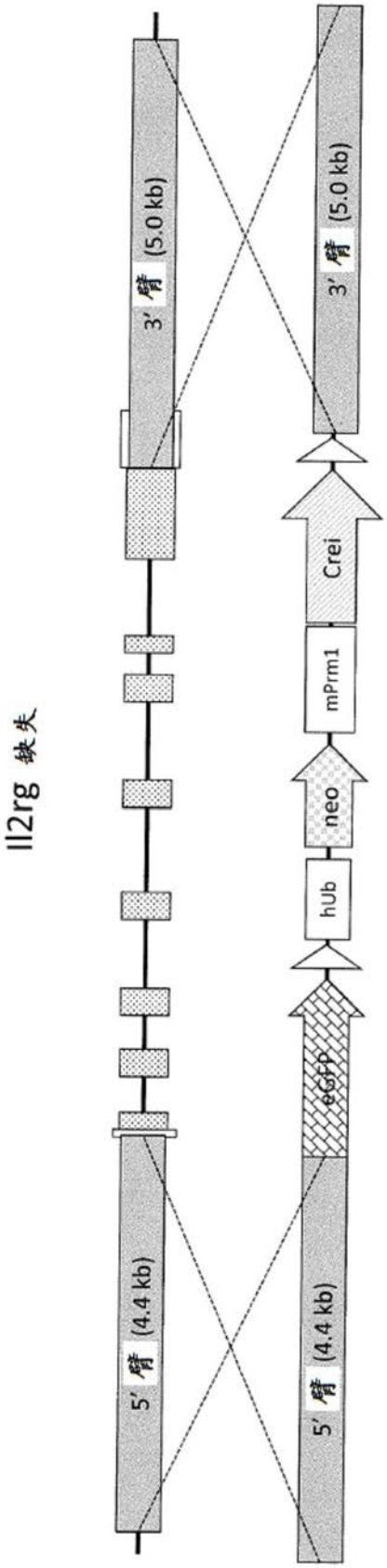

Il2rg 缺失
3′臂 (5.0 kb)
Crei
mPrm1
neo
hUb
eGFP
5′臂 (4.4 kb)
3′臂 (5.0 kb)
5′臂 (4.4 kb)

图24

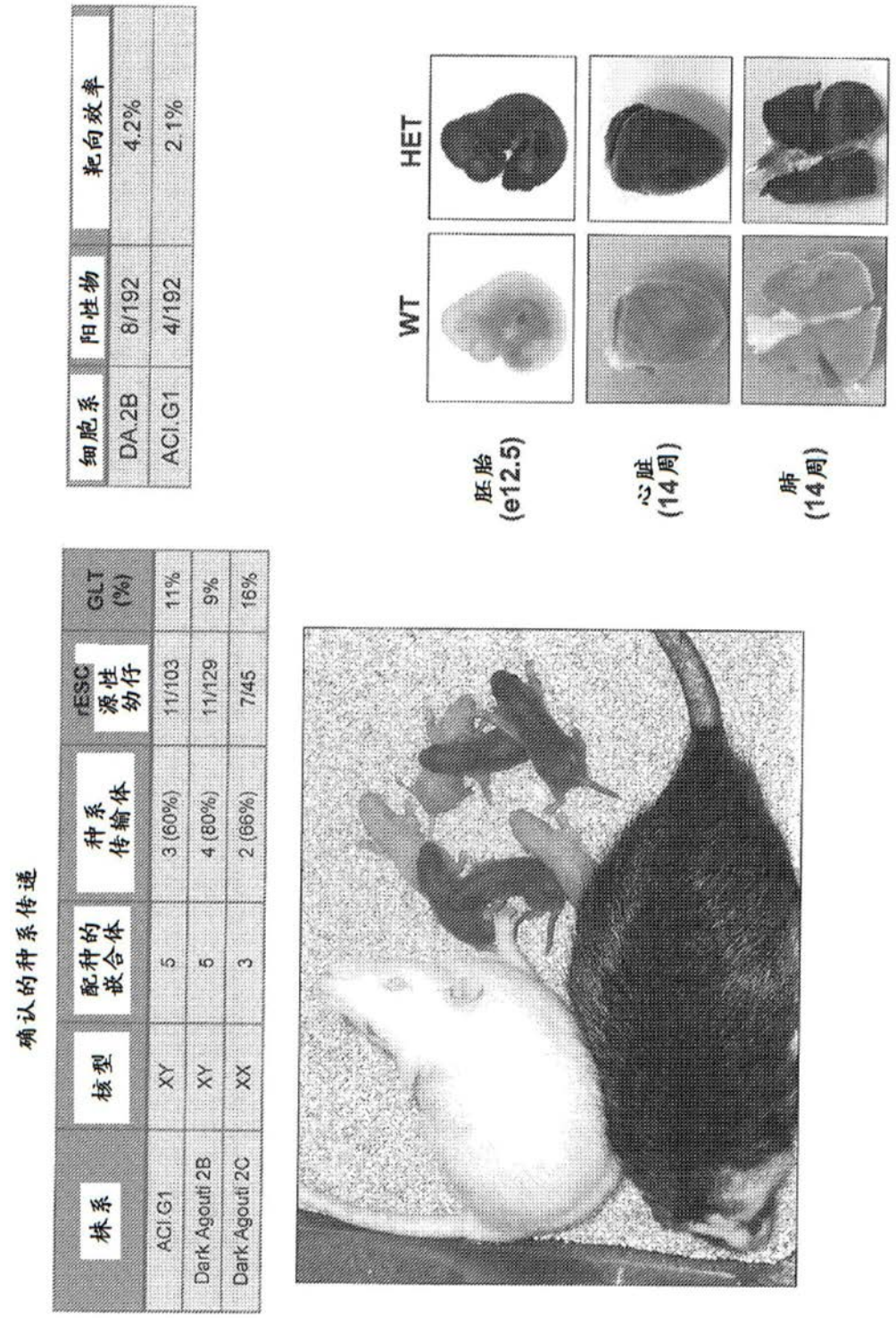

确认的种系传递
株系
核型
配种的嵌合体
种系传输体
rESC源性幼仔
GLT (%)
ACI.G1 XY 5 3 (60%) 11/103 11%
Dark Agouti 2B XY 5 4 (80%) 11/129 9%
Dark Agouti 2C XX 3 2 (66%) 7/45 16%
细胞系
阳性物
靶向效率
DA.2B 8/192 4.2%
ACI.G1 4/192 2.1%
WT
HET
胚胎 (e12.5)
心脏 (14周)
肺 (14周)

图25

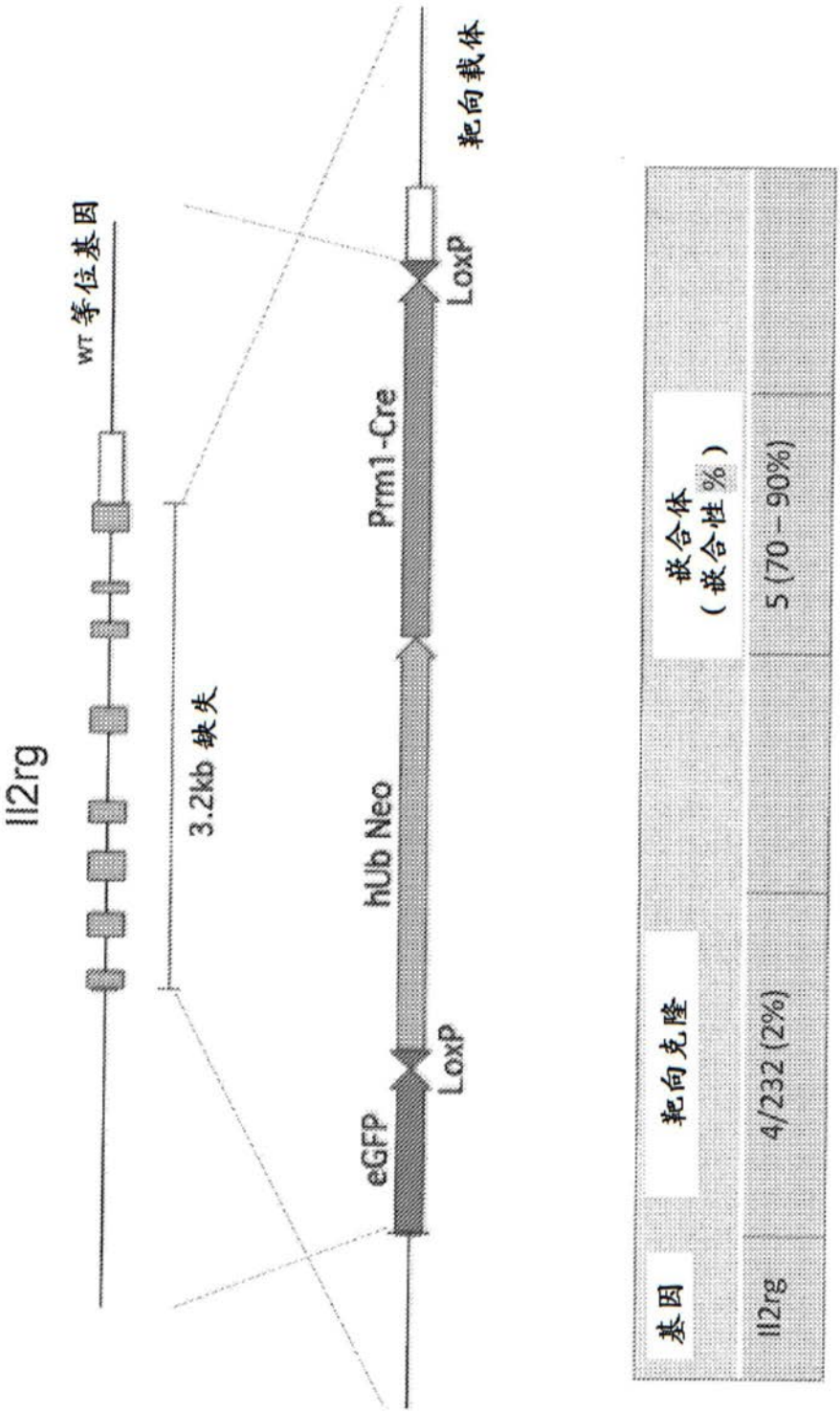

Il2rg
WT 等位基因
3.2kb 缺失
hUb Neo
Prm1-Cre
eGFP
LoxP
LoxP
靶向载体
基因
靶向克隆
嵌合体
（嵌合性 %）
Il2rg
4/232 (2%)
5 (70－90%)

图26

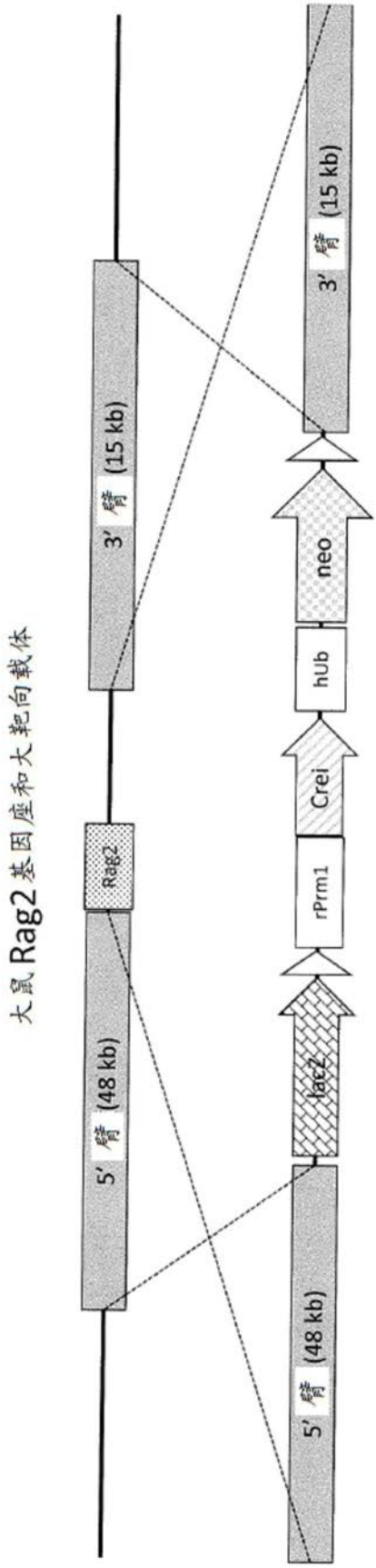

大鼠 Rag2 基因座和大靶向载体
5′ 臂 (48 kb)
Rag2
3′ 臂 (15 kb)
lacZ
rPrm1
Crei
hUb
neo
5′ 臂 (48 kb)
3′ 臂 (15 kb)

图27

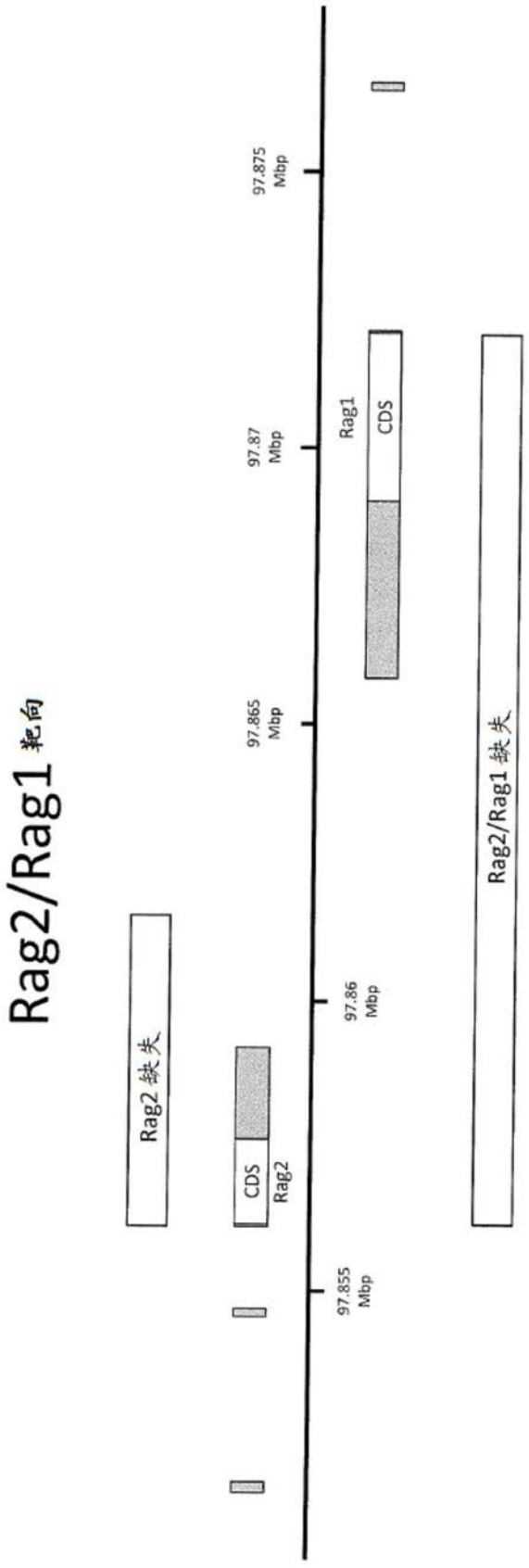

Rag2/Rag1 靶向
Rag2 缺失
CDS
Rag2
Rag1
CDS
Rag2/Rag1 缺失
97.875 Mbp
97.87 Mbp
97.865 Mbp
97.86 Mbp
97.855 Mbp

图28

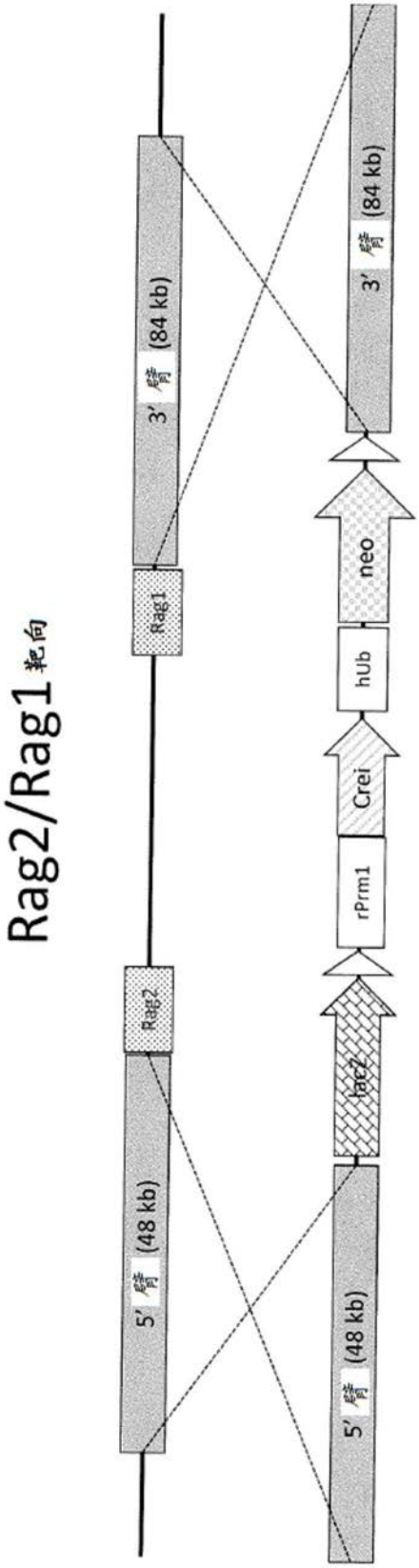

Rag2/Rag1靶向
5′臂 (48 kb)
Rag2
3′臂 (84 kb)
Rag1
5′臂 (48 kb)
lacZ
rPrm1
Crei
hUb
neo
3′臂 (84 kb)

图29

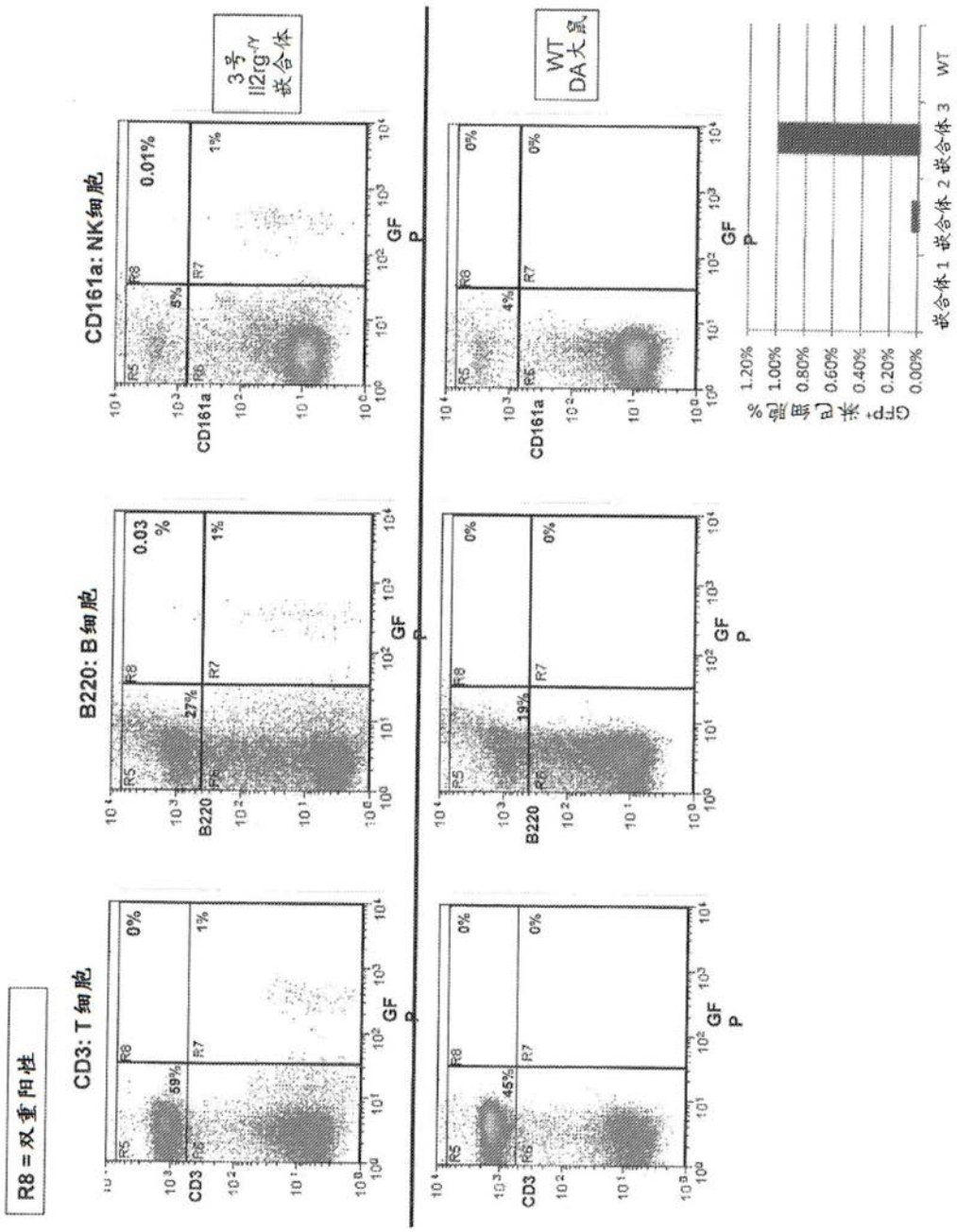

R8 = 双重阳性
CD3: T 细胞
B220: B 细胞
CD161a: NK细胞
3号 Il2rg-/- 嵌合体
WT DA大鼠
CD3
B220
CD161a
GFP
R5
R6
R7
R8
59%
45%
27%
19%
5%
4%
0.03 %
0.01%
0%
1%
嵌合体 1 嵌合体 2 嵌合体 3 WT
GFP+细胞百分比
1.20%
1.00%
0.80%
0.60%
0.40%
0.20%
0.00%

图30

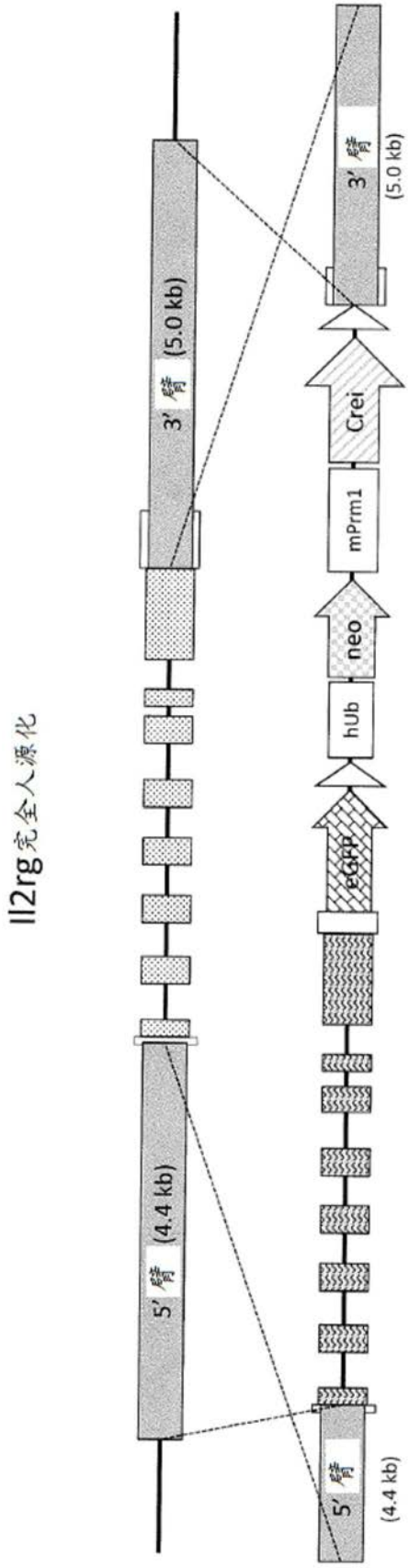

Il2rg完全人源化
3′臂 (5.0 kb)
5′臂 (4.4 kb)
3′臂 (5.0 kb)
Crei
mPrm1
neo
hUb
eGFP
5′臂 (4.4 kb)

图31

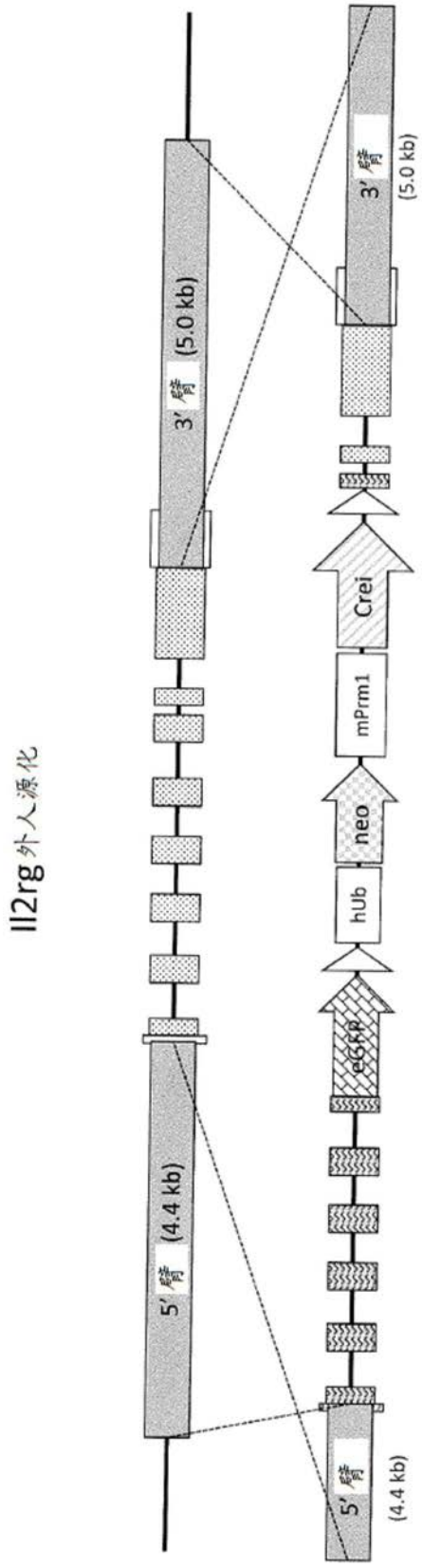

Il2rg外人源化
3′臂 (5.0 kb)
5′臂 (4.4 kb)
eGFP
hUb
neo
mPrm1
Crei
3′臂 (5.0 kb)
5′臂 (4.4 kb)

图32


Hsa_Il2rg_NP_000197.1
Rno_Il2rg_NP_543165.1
Rogn_Il2rg_humin
人/大鼠 Il2rg 蛋白之间的接合

图33