



(12)发明专利申请

(10)申请公布号 CN 109536526 A

(43)申请公布日 2019.03.29

(21)申请号 201811024942.8

(51)Int.Cl.

(22)申请日 2013.04.25

C12N 15/85(2006.01)

(30)优先权数据

C12N 15/90(2006.01)

61/638,267 2012.04.25 US

C12N 15/12(2006.01)

(62)分案原申请数据

201380030738.6 2013.04.25

(71)申请人 瑞泽恩制药公司

地址 美国纽约州

(72)发明人 D·弗伦德维 W·奥尔巴克

D·M·瓦伦泽拉

G·D·扬科普洛斯 K·V·莱

(74)专利代理机构 北京市君合律师事务所

11517

代理人 赵昊 孙倩

权利要求书2页 说明书28页

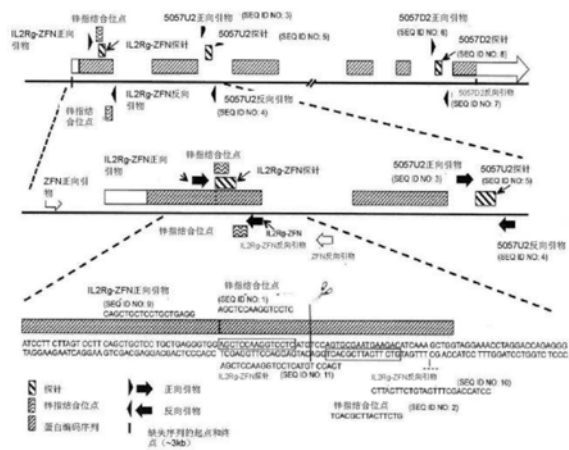
序列表3页 附图1页

(54)发明名称

核酸酶介导的使用大靶向载体的靶向

(57)摘要

本申请提供了通过使用同源重组在基因组的靶基因座制备一种或多种经靶向的基因修饰的组合物和方法,其中通过在所述靶基因组的基因座或其附近的单链或双链的断裂促进所述同源重组。本申请还提供了使用工程改造的核酸酶在原核或真核细胞中提高LTVEC和基因组的靶基因座之间同源重组效率的组合物和方法。



1. 一种通过在小鼠胚胎干 (ES) 细胞中进行同源重组以修饰基因组的靶基因座的方法, 所述方法包括:

(a) 向所述小鼠ES细胞中引入:

(i) 锌指核酸酶 (ZFN), 所述锌指核酸酶在基因组的靶基因座或其附近产生单链或双链断裂, 和

(ii) 大靶向载体 (LTVEC), 所述大靶向载体包含翼侧为上游同源臂和下游同源臂的插入核酸; 其中所述插入核酸的长度范围为约5kb至约30kb, 并且所述上游同源臂和所述下游同源臂的总和是至少10kb;

(b) 检测所述小鼠ES细胞中所述插入核酸整合进入所述基因组的靶基因座的情况, 其中所述整合使得在所述靶基因座的基因座处内源性核酸序列缺失并被所述插入核酸取代; 以及

(c) 选择经靶向的小鼠ES细胞, 所述小鼠ES细胞包含在所述基因组的靶基因座中的所述插入核酸, 其中与单独使用LTVEC相比, 将所述LTVEC与所述ZFN联合使用使得靶向效率增加。

2. 根据权利要求1所述的方法, 其中与单独使用LTVEC相比, 所述LTVEC的靶向效率增加至少2倍。

3. 根据权利要求1所述的方法, 其中所述ZFN是表达构建体, 所述表达构建体包含编码ZFN蛋白的核酸序列, 并且其中所述核酸与在小鼠细胞中有活性的启动子可操作地连接。

4. 根据权利要求1所述的方法, 其中所述ZFN是编码ZFN蛋白的mRNA。

5. 根据权利要求1所述的方法, 其中所述ZFN的靶序列位于所述基因组的靶基因座的内含子、外显子、启动子、启动子调节区或增强子区中。

6. 根据权利要求1所述的方法, 其中所述插入核酸包含选择性表达盒。

7. 根据权利要求1所述的方法, 其中所述插入核酸包含报告基因。

8. 根据权利要求1所述的方法, 其中将所述插入核酸整合进入所述基因组的靶基因座使得在所述基因组的靶基因座缺失内源性基因。

9. 根据权利要求1所述的方法, 其中将所述插入核酸整合进入所述基因组的靶基因座导致敲除、敲入、点突变、结构域交换、外显子交换、内含子交换、调节序列交换、基因交换或其组合。

10. 根据权利要求1所述的方法, 其中所述插入核酸包含人源核酸序列。

11. 根据权利要求1所述的方法, 其中检测步骤 (b) 通过等位基因修饰 (MOA) 检测实施。

12. 根据权利要求1所述的方法, 其中所述插入核酸包含与在所述基因组的靶基因座处的被取代的核酸序列具有同源性的核酸序列。

13. 根据权利要求1所述的方法, 其中所述插入核酸包含与在所述基因组的靶基因座处的被取代的核酸序列具有直系同源性的核酸序列。

14. 根据权利要求1所述的方法, 其中所述插入核酸包含来自非小鼠的种属的核酸序列。

15. 根据权利要求1所述的方法, 其中所述LTVEC的长度范围为约50kb至约300kb。

16. 根据权利要求15所述的方法, 其中所述LTVEC的长度范围为约50kb至约100kb。

17. 根据权利要求15所述的方法, 其中所述LTVEC的长度范围为约100kb至约200kb。

18. 根据权利要求15所述的方法,其中所述LTVEC的长度范围为约200kb至约300kb。
19. 根据权利要求1所述的方法,其中所述上游同源臂和所述下游同源臂的总和的范围为约10kb至约100kb。
20. 根据权利要求1所述的方法,其中所述上游同源臂和所述下游同源臂的总和的范围为约10kb至约200kb。
21. 根据权利要求1所述的方法,其中所述插入核酸包含条件型等位基因。
22. 根据权利要求1所述的方法,其中所述插入核酸包含编码人源免疫球蛋白重链可变区氨基酸序列的基因组核酸序列。
23. 根据权利要求1所述的方法,其中所述插入核酸包含编码人源免疫球蛋白轻链可变区氨基酸序列的基因组核酸序列。
24. 根据权利要求23所述的方法,其中所述基因组核酸序列包含未经重排的人源 κ 和/或 λ 轻链可变区核酸序列。
25. 根据权利要求23所述的方法,其中所述基因组核酸序列包含重排的人源 κ 和/或 λ 轻链可变区核酸序列。
26. 根据权利要求1所述的方法,其中所述插入核酸包含人疾病等位基因。
27. 根据权利要求22所述的方法,其中所述插入核酸包含重排的人源免疫球蛋白重链可变区核酸序列,所述重排的人源免疫球蛋白重链可变区核酸序列与免疫球蛋白重链恒定区核酸序列可操作地连接。
28. 根据权利要求22所述的方法,其中所述插入核酸包含未经重排的人源免疫球蛋白重链可变区核酸序列,所述未经重排的人源免疫球蛋白重链可变区核酸序列与免疫球蛋白重链恒定区核酸序列可操作地连接。
29. 根据权利要求27所述的方法,其中所述免疫球蛋白重链恒定区序列是小鼠免疫球蛋白重链恒定区序列、人源免疫球蛋白重链恒定区序列或者其组合。
30. 根据权利要求27所述的方法,其中所述免疫球蛋白重链恒定区核酸序列选自C_H1、 α 链、C_H2、C_H3及其组合。
31. 根据权利要求24所述的方法,其中所述插入核酸包含未经重排的人源 κ 或 λ 轻链可变区核酸序列,所述未经重排的人源 κ 或 λ 轻链可变区核酸序列与小鼠或人源免疫球蛋白轻链恒定区核酸序列可操作地连接,所述小鼠或人源免疫球蛋白轻链恒定区核酸序列选自 κ 轻链恒定区核酸序列和 λ 轻链恒定区核酸序列。
32. 根据权利要求25所述的方法,其中所述插入核酸包含重排的人源 κ 或 λ 轻链可变区核酸序列,所述重排的人源 κ 或 λ 轻链可变区核酸序列与小鼠或人源免疫球蛋白轻链恒定区核酸序列可操作地连接,所述小鼠或人源免疫球蛋白轻链恒定区核酸序列选自 κ 轻链恒定区核酸序列和 λ 轻链恒定区核酸序列。
33. 根据权利要求28所述的方法,其中所述免疫球蛋白重链恒定区序列是小鼠免疫球蛋白重链恒定区序列、人源免疫球蛋白重链恒定区序列或者其组合。
34. 根据权利要求28所述的方法,其中所述免疫球蛋白重链恒定区核酸序列选自C_H1、 α 链、C_H2、C_H3及其组合。

核酸酶介导的使用大靶向载体的靶向

[0001] 本申请为申请日为2013年4月25日,申请号为201380030738.6,标题为“核酸酶介导的使用大靶向载体的靶向”的中国发明专利申请的分案申请,其全部内容均通过引用并入本申请。

[0002] 相关申请的交叉引用

[0003] 本申请要求2012年4月25日提交的美国临时申请号61/638,267的优先权,其全部内容通过引用并入本申请。

发明领域

[0004] 核酸酶和DNA构建体包括靶向载体(例如,大靶向载体,“LTVEC”)以用于实现在基因组的靶基因座的同源重组。本申请提供了通过使用同源重组制备经靶向的基因修饰的组合物和方法,其中通过在基因组的靶基因座或其附近的单链或双链断裂促进所述同源重组。本申请提供了使用工程改造的核酸酶在原核或真核细胞中提高LTVEC和基因组的靶基因座之间同源重组效率的组合物和方法。

背景技术

[0005] 使用靶向载体的同源重组是一种在非人动物中获得所需基因组修饰的流行方法,所述靶向载体被特别设计以便在基因组的基因座加入、缺失或取代特定的核酸序列。被特异性工程改造以便在基因组的靶基因座或其附近引入单链或双链断裂的核酸酶能够与靶向载体一起使用以提高在基因组的靶基因座同源重组的效率。

[0006] 尽管通过同源重组进行基因组修饰的技术在最近二十年中已经相当先进,但是在很多情况下使用非常大的靶向载体LTVEC达到可接受的靶向频率仍存在困难,例如当使用较多的人源基因组片段取代较大部分的啮齿动物基因组时,或者靶向某些细胞类型时存在困难,例如成纤维细胞或其他体细胞。在本领域中需要用于使用LTVEC修饰真核基因组的较大基因组基因座的更多的和改进的方法。

[0007] 发明概述

[0008] 本申请提供了使用大靶向载体(LTVEC)联合核酸酶试剂修饰基因组中感兴趣的基因座的组合物和方法,所述组合物和方法使得在基因组中感兴趣的基因座上能够有效进行较大的核酸序列的缺失、加入(例如插入)和/或取代。

[0009] 本申请提供了使用LTVEC和核酸酶试剂通过细菌同源重组(BHR)在原核细胞中修饰哺乳动物基因组的靶基因座的组合物和方法,其中通过所述核酸酶试剂在基因组的靶基因座或其附近产生的单链或双链断裂促进所述BHR。本申请提供了包含LTVEC和核酸酶试剂的原核细胞,所述核酸酶试剂在表达后能够在靶位点或其附近引入单链或双链断裂。本申请提供了用于通过使用本申请所述的各种LTVEC和核酸酶在发生重组的原核细胞中使用外源性的核酸序列(例如,同源的或直系同源的人源基因组核酸序列)取代非人动物较大的基因组基因座的组合物和方法。

[0010] 本申请提供了在各种多能性哺乳动物细胞中使用LTVEC和核酸酶试剂修饰基因组

中感兴趣的基因座的组合物和方法。本申请提供了用于通过使用本申请所述的各种LTVEC和核酸酶在非人动物的多能细胞中使用外源性的核酸序列(例如,同源的或直系同源的人源基因组核酸序列)取代非人动物较大的基因组基因座的组合物和方法。

[0011] 本申请提供了非人动物的多能细胞,所述细胞包含本申请所述的各种LTVEC和核酸酶试剂。

[0012] 本申请提供了用于生成经基因修饰的非人动物的组合物和方法,所述经基因修饰的非人动物包含如本申请所述的一个或多个经靶向的基因修饰。

[0013] 在一个方面,本申请提供了一种原核细胞,所述原核细胞包含大靶向载体(LTVEC),所述载体包含定向至靶基因座的同源臂,和编码核酸酶试剂的核酸序列,所述核酸酶试剂在所述靶基因座或其附近形成单链或双链断裂。

[0014] 在一个实施方式中,所述原核细胞能表达介导细菌同源重组(BHR)的重组酶。在一个实施方式中,所述原核细胞是大肠杆菌的具有重组能力的菌株。

[0015] 在一个实施方式中,所述LTVEC的范围为约50kb至约300kb。在一个实施方式中,所述LTVEC的范围为约50kb至约75kb。在一个实施方式中,所述LTVEC的范围为约75kb至约100kb。在一个实施方式中,所述LTVEC的范围为约100kb至约125kb。在一个实施方式中,所述LTVEC的范围为约125kb至约150kb。在一个实施方式中,所述LTVEC的范围为约150kb至约175kb。在一个实施方式中,所述LTVEC的范围为约175kb至约200kb。在一个实施方式中,所述LTVEC的范围为约200kb至约225kb。在一个实施方式中,所述LTVEC的范围为约225kb至约250kb。在一个实施方式中,所述LTVEC的范围为约250kb至约275kb。在一个实施方式中,所述LTVEC的范围为约275kb至约300kb。

[0016] 在一个实施方式中,所述靶向载体的同源臂来源于BAC文库、粘粒文库或P1噬菌体文库。在一个实施方式中,所述同源臂来源于使用常规方法不能靶向的非人动物基因组基因座。在一个实施方式中,所述同源臂来源于合成的DNA。

[0017] 在一个实施方式中,上游同源臂和下游同源臂的总和是至少10kb。在一个实施方式中,所述上游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述下游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约5kb至约10kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约10kb至约20kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约20kb至约30kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约30kb至约40kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约40kb至约50kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约50kb至约60kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约60kb至约70kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约70kb至约80kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约80kb至约90kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约90kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约100kb至约110kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约110kb至约120kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约120kb至约130kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约130kb至约140kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约140kb至约150kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约150kb至约160kb。在一个实施方式

中,所述上游和所述下游同源臂的范围为约160kb至约170kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约170kb至约180kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约180kb至约190kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约190kb至约200kb。

[0018] 在一个实施方式中,所述LTVEC包含选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在原核细胞中具有活性。在一个实施方式中,所述启动子在原核和真核细胞中均具有活性。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其组合。

[0019] 在一个实施方式中,所述LTVEC包含范围为约5kb至约200kb的插入核酸。在一个实施方式中,所述插入核酸为约5kb至约10kb。在一个实施方式中,所述插入核酸为约10kb至约20kb。在一个实施方式中,所述插入核酸为约20kb至约30kb。在一个实施方式中,所述插入核酸为约30kb至约40kb。在一个实施方式中,所述插入核酸为约40kb至约50kb。在一个实施方式中,所述插入核酸为约60kb至约70kb。在一个实施方式中,所述插入核酸为约80kb至约90kb。在一个实施方式中,所述插入核酸为约90kb至约100kb。在一个实施方式中,所述插入核酸为约100kb至约110kb。在一个实施方式中,所述插入核酸为约120kb至约130kb。在一个实施方式中,所述插入核酸为约130kb至约140kb。在一个实施方式中,所述插入核酸为约140kb至约150kb。在一个实施方式中,所述插入核酸为约150kb至约160kb。在一个实施方式中,所述插入核酸为约160kb至约170kb。在一个实施方式中,所述插入核酸为约170kb至约180kb。在一个实施方式中,所述插入核酸为约180kb至约190kb。在一个实施方式中,所述插入核酸为约190kb至约200kb。

[0020] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有翼侧为位点特异性重组靶序列的核酸。在一个实施方式中,所述核酸含有基因组核酸。在一个实施方式中,所述基因组核酸来源于小鼠、人或其组合。在一个实施方式中,所述位点特异性重组靶序列选自下组:loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attP、att、FRT、rox及其组合。

[0021] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有条件型等位基因。在一个实施方式中,所述条件型等位基因是如在US 2011/0104799中所描述的多功能等位基因,其全部内容通过引用并入本申请。在一个实施方式中,所述条件型等位基因包含:(a)在相对于靶基因转录而言有意义方向的驱动序列,和在有意义或反义方向的药物选择性表达盒;(b)在反义方向的感兴趣的核苷酸序列(NSI)和条件型倒置模块(COIN,其利用外显子分割的内含子和可逆的基因捕获样分子;参见例如US 2011/0104799,其全部内容通过引用并入和(c)可重组单元,其在与第一重组酶接触后重组形成(i)缺乏驱动序列和DSC以及(ii)在有意义方向含有NSI和在反义方向含有COIN的条件型等位基因。

[0022] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在原核细胞中具有活性。在一

个实施方式中,所述核酸在原核和真核细胞中均具有活性。在一个实施方式中,所述选择性表达盒的翼侧具有位点特异性重组靶序列。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其组合。

[0023] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有与启动子可操作地连接的报告基因,其中所述报告基因编码选自下组的报告蛋白:LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(EYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶及其组合。在一个实施方式中,所述报告基因在诱导型启动子的控制下表达。在一个实施方式中,所述报告基因在内源性启动子的控制下表达。在一个实施方式中,所述报告基因在外源性启动子的控制下表达。在一个实施方式中,所述报告基因在特定的细胞类型中表达。在一个实施方式中,所述报告基因以组织特异性的方式表达。在一个实施方式中,所述报告基因以发育阶段特异性的方式表达。

[0024] 在一个方面,本申请提供了一种真核细胞,所述真核细胞包含大靶向载体,所述靶向载体含有定向至所述真核细胞基因组中的靶基因座的同源臂和编码核酸酶试剂的核酸序列,所述核酸酶试剂在靶基因座或其附近形成单链或双链断裂。

[0025] 在一个实施方式中,所述真核细胞是多能细胞。在一个实施方式中,所述多能细胞是胚胎干(ES)细胞。在一个实施方式中,所述多能细胞是非人ES细胞。在一个实施方式中,所述多能细胞是诱导多能性干(iPS)细胞。在一个实施方式中,所述诱导多能性(iPS)细胞来源于成纤维细胞。在一个实施方式中,所述诱导多能性(iPS)细胞来源于人成纤维细胞。在一个实施方式中,所述多能细胞是造血干细胞(HSC)。在一个实施方式中,所述多能细胞神经干细胞(NSC)。在一个实施方式中,所述多能细胞是外胚层干细胞。在一个实施方式中,所述多能细胞是发育受限的祖细胞。在一个实施方式中,所述多能细胞是啮齿动物多能细胞。在一个实施方式中,所述啮齿动物多能细胞是大鼠多能细胞。在一个实施方式中,所述大鼠多能细胞是大鼠ES细胞。在一个实施方式中,所述啮齿动物多能细胞是小鼠多能细胞。在一个实施方式中,所述多能细胞是小鼠胚胎干(ES)细胞。

[0026] 在一个实施方式中,所述真核细胞是永生化的小鼠或大鼠细胞。在一个实施方式中,所述真核细胞是永生化的来源细胞。在一个实施方式中,所述真核细胞是人成纤维细胞。在一个实施方式中,所述真核细胞是癌细胞。在一个实施方式中,所述真核细胞是人癌细胞。

[0027] 在一个实施方式中,所述LTVEC的范围为约50kb至约300kb。在一个实施方式中,所述LTVEC的范围为约50kb至约75kb。在一个实施方式中,所述LTVEC的范围为约75kb至约100kb。在一个实施方式中,所述LTVEC的范围为约100kb至125kb。在一个实施方式中,所述LTVEC的范围为约125kb至约150kb。在一个实施方式中,所述LTVEC的范围为约150kb至约175kb。在一个实施方式中,所述LTVEC的范围为约175kb至约200kb。在一个实施方式中,所述LTVEC的范围为约200kb至约225kb。在一个实施方式中,所述LTVEC的范围为约225kb至约250kb。在一个实施方式中,所述LTVEC的范围为约250kb至约275kb。在一个实施方式中,所述LTVEC的范围为约275kb至约300kb。

[0028] 在一个实施方式中,所述靶向载体的同源臂来源于BAC文库、粘粒文库或P1噬菌体文库。在一个实施方式中,所述同源臂来源于使用常规方法不能靶向的非人动物的基因组基因座。在一个实施方式中,所述同源臂来源于合成的DNA。

[0029] 在一个实施方式中,上游同源臂和下游同源臂的总和为至少10kb。在一个实施方式中,所述上游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述下游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约5kb至约10kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约10kb至约20kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约20kb至约30kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约30kb至约40kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约40kb至约50kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约50kb至约60kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约60kb至约70kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约70kb至约80kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约80kb至约90kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约90kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约100kb至约110kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约110kb至约120kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约120kb至约130kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约130kb至约140kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约140kb至约150kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约150kb至约160kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约160kb至约170kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约170kb至约180kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约180kb至约190kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约190kb至约200kb。

[0030] 在一个实施方式中,所述LTVEC包含选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在原核细胞中具有活性。在一个实施方式中,所述核酸在原核和真核细胞中均具有活性。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其组合。

[0031] 在一个实施方式中,所述LTVEC包含范围为约5kb至约200kb的插入核酸。在一个实施方式中,所述插入核酸为约5kb至约10kb。在一个实施方式中,所述插入核酸为约10kb至约20kb。在一个实施方式中,所述插入核酸为约20kb至约30kb。在一个实施方式中,所述插入核酸为约30kb至约40kb。在一个实施方式中,所述插入核酸为约40kb至约50kb。在一个实施方式中,所述插入核酸为约60kb至约70kb。在一个实施方式中,所述插入核酸为约80kb至约90kb。在一个实施方式中,所述插入核酸为约90kb至约100kb。在一个实施方式中,所述插入核酸为约100kb至约110kb。在一个实施方式中,所述插入核酸为约120kb至约130kb。在一个实施方式中,所述插入核酸为约130kb至约140kb。在一个实施方式中,所述插入核酸为约140kb至约150kb。在一个实施方式中,所述插入核酸为约150kb至约160kb。在一个实施方式

中,所述插入核酸为约160kb至约170kb。在一个实施方式中,所述插入核酸为约170kb至约180kb。在一个实施方式中,所述插入核酸为约180kb至约190kb。在一个实施方式中,所述插入核酸为约190kb至约200kb。

[0032] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有翼侧为位点特异性重组靶序列的核酸。在一个实施方式中,所述核酸含有基因组核酸。在一个实施方式中,所述基因组核酸来源于小鼠、人或其组合。在一个实施方式中,所述位点特异性重组靶序列选自下组:loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attP、att、FRT、rox及其组合。

[0033] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有条件型等位基因。在一个实施方式中,所述条件型等位基因是如在US 2011/0104799中所描述的多功能等位基因,其全部内容通过引用并入本申请。在一个实施方式中,所述条件型等位基因包含:(a)在相对于靶基因转录而言有意义方向的驱动序列,和在有意义或反义方向的药物选择性表达盒;(b)在反义方向的感兴趣的核苷酸序列(NSI)和条件倒置模块(COIN,其利用外显子分割的内含子和可逆的基因捕获样分子;参见例如US 2011/0104799,其全部内容通过引用并入)和(c)可重组单元,其在与第一重组酶接触后重组形成(i)缺乏驱动序列和DSC以及(ii)在有意义方向含有NSI和在反义方向含有COIN的条件型等位基因。

[0034] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在原核细胞中具有活性。在一个实施方式中,所述核酸在原核和真核细胞中均具有活性。在一个实施方式中,所述选择性表达盒的翼侧具有位点特异性重组靶序列。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其组合。

[0035] 在一个实施方式中,所述LTVEC包含插入核酸,所述插入核酸含有与启动子可操作地连接的报告基因,其中所述报告基因编码选自下组的报告蛋白:LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(EYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶及其组合。在一个实施方式中,所述报告基因在诱导型启动子的控制下表达。在一个实施方式中,所述报告基因在内源性启动子的控制下表达。在一个实施方式中,所述报告基因在外源性启动子的控制下表达。在一个实施方式中,所述报告基因在特定的细胞类型中表达。在一个实施方式中,所述报告基因以组织特异性的方式表达。在一个实施方式中,所述报告基因以发育阶段特异性的方式表达。

[0036] 在一个方面,本申请提供了一种在原核细胞中通过细菌同源重组(BHR)修饰哺乳动物细胞基因组的靶基因座的方法,

[0037] 所述方法包括:

[0038] (a) 引入包含哺乳动物基因组的靶基因座的原核细胞:

[0039] (i) 靶向载体,所述靶向载体包含翼侧为第一上游同源臂和第一下游同源臂的插入核酸,和

[0040] (ii) 核酸酶试剂,所述核酸酶试剂在所述基因组的靶基因座或其附近产生单链或双链断裂,和

[0041] (i) 选择经靶向的原核细胞,所述原核细胞包含所述插入核酸,

[0042] 其中所述原核细胞能够表达介导所述BHR的重组酶。

[0043] 在一个实施方式中,所述基因组的靶基因座选自FcERIa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基因座和ERBB4基因座。

[0044] 在一个实施方式中,所述基因组的靶基因座存在于大靶向载体(LTVEC)中,所述大靶向载体包含第二上游同源臂和第二下游同源臂。在一个实施方式中,所述第二上游同源臂和所述第二下游同源臂的总和为至少10kb。在一个实施方式中,所述第二上游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述第二下游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约5kb至约10kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约10kb至约20kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约20kb至约30kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约30kb至约40kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约40kb至约50kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约50kb至约60kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约60kb至约70kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约70kb至约80kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约80kb至约90kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约90kb至约100kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约100kb至约110kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约110kb至约120kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约120kb至约130kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约130kb至约140kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约140kb至约150kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约150kb至约160kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约160kb至约170kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约170kb至约180kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约180kb至约190kb。在一个实施方式中,所述第二上游和所述下游同源臂的范围为约190kb至约200kb。

[0045] 在一个实施方式中,所述哺乳动物是人。在一个实施方式中,所述哺乳动物是人并且所述靶向是离体的人细胞。在一个实施方式中,所述哺乳动物是啮齿动物。在一个实施方式中,所述啮齿动物选自小鼠、大鼠和仓鼠。

[0046] 在一个实施方式中,所述核酸酶试剂与所述靶向载体一起引入。在一个实施方式中,在一段时间中所述核酸酶试剂与所述靶向载体分别引入。在一个实施方式中,所述核酸酶试剂先于所述靶向载体引入。在一个实施方式中,所述核酸酶试剂在所述靶向载体引入之后引入。

[0047] 在一个实施方式中,将所述靶向载体与所述核酸酶试剂联用与单独使用所述靶向载体相比使靶向效率提高。在一个实施方式中,当将所述靶向载体与所述核酸酶试剂联用

时,与当将所述靶向载体单独使用时相比,所述靶向载体的靶向效率提高至少2倍。在一个实施方式中,当将所述靶向载体与所述核酸酶试剂联用时,与当将所述靶向载体单独使用时相比,所述靶向载体的靶向效率提高至少3倍。在一个实施方式中,当将所述靶向载体与所述核酸酶试剂联用时,与当将所述靶向载体单独使用时相比,所述靶向载体的靶向效率提高至少4倍。

[0048] 在一个实施方式中,所述原核细胞是具有重组能力的大肠杆菌菌株。在一个实施方式中,所述原核细胞包含编码所述重组酶的核酸。在一个实施方式中,所述原核细胞不包含编码所述重组酶的核酸,并且编码所述重组酶的核酸被引入所述原核细胞中。在一个实施方式中,所述核酸包含编码所述重组酶的DNA或mRNA。在一个实施方式中,编码所述重组酶的所述核酸是pABG。在一个实施方式中,所述重组酶在诱导型启动子的控制下表达。在一个实施方式中,所述重组酶的表达由阿拉伯糖控制。

[0049] 在一个实施方式中,所述核酸酶试剂是表达构建体,所述构建体包含编码核酸酶的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子是组成型激活的启动子。在一个实施方式中,所述启动子是诱导型启动子。在一个实施方式中,所述启动子在所述原核细胞中具有活性。在一个实施方式中,所述核酸酶试剂是编码核酸酶的mRNA。在一个实施方式中,所述核酸酶试剂是编码核酸内切酶的mRNA。

[0050] 在一个实施方式中,所述核酸酶试剂是锌指核酸酶(ZFN)。在一个实施方式中,所述ZFN的各个单体均含有3个或更多个基于锌指的DNA结合结构域,其中各基于锌指的DNA结合结构域均结合至3bp子位点。在一个实施方式中,所述ZFN是嵌合蛋白,所述嵌合蛋白包含与独立的核酸酶可操作地连接的基于锌指的DNA结合结构域。在一个实施方式中,所述独立的核酸内切酶是FokI核酸内切酶。在一个实施方式中,所述核酸酶试剂包含第一ZFN和第二ZFN,其中各个所述第一ZFN和所述第二ZFN均与FokI核酸酶可操作地连接,其中所述第一和所述第二ZFN识别各条靶DNA序列链中的被约6bp至约40bp的切割位点所分隔的两个相邻的靶DNA序列,并且其中所述FokI核酸酶二聚化且形成双链断裂。

[0051] 在一个实施方式中,所述核酸酶试剂是转录激活因子样效应物核酸酶(TALEN)。在一个实施方式中,所述TALEN的各个单体均包含12-25个TAL重复,其中各个TAL重复均与1bp子位点结合。在一个实施方式中,所述核酸酶试剂是嵌合蛋白,所述嵌合蛋白包含与独立的核酸酶可操作地连接的基于TAL重复的DNA结合结构域。在一个实施方式中,所述独立的核酸酶是FokI核酸内切酶。在一个实施方式中,所述核酸酶试剂包含第一基于TAL重复的DNA结合结构域和第二基于TAL重复的DNA结合结构域,其中各个所述第一和所述第二基于TAL重复的DNA结合结构域均与FokI核酸酶可操作地连接,其中所述第一和所述第二基于TAL重复的DNA结合结构域识别各条靶DNA序列链中的被约6bp至约40bp的切割位点所分隔的两个相邻的靶DNA序列,并且其中所述FokI核酸酶二聚化且在靶序列形成双链断裂。

[0052] 在一个实施方式中,所述核酸酶的各个单体均识别至少9个核苷酸的靶序列。在一个实施方式中,所述靶序列的长度为约9至约12个核苷酸。在一个实施方式中,所述靶序列的长度为约12至约15个核苷酸。在一个实施方式中,所述靶序列的长度为约15至约18个核苷酸。在一个实施方式中,所述靶序列的长度为约18至约21个核苷酸。

[0053] 在一个实施方式中,所述核酸酶试剂的靶序列位于内含子中。在一个实施方式中,所述靶序列位于外显子中。在一个实施方式中,所述靶序列位于启动子中。在一个实施方式

中,所述靶序列在非蛋白编码区中。在一个实施方式中,所述非蛋白编码区是调控区。在一个实施方式中,所述靶序列位于启动子调控区。在一个实施方式中,所述靶序列位于增强子区中。

[0054] 在一个实施方式中,所述核酸酶试剂是大范围核酸酶。在一个实施方式中,所述大范围核酸酶识别12至40个碱基对的双链DNA序列。在一个实施方式中,所述大范围核酸酶识别在所述基因组中的一个极佳匹配的靶序列。在一个实施方式中,所述大范围核酸酶是归巢核酸酶。在一个实施方式中,所述归巢核酸酶是LAGLIDADG家族的归巢核酸酶。在一个实施方式中,所述LAGLIDADG家族的归巢核酸酶选自I-SceI、I-CreI和I-DmO1。

[0055] 在一个实施方式中,所述靶向载体是大靶向载体(LTVEC)。

[0056] 在一个实施方式中,所述LTVEC的范围为约50kb至约300kb。在一个实施方式中,所述LTVEC的范围为约50kb至约75kb。在一个实施方式中,所述LTVEC的范围为约75kb至约100kb。在一个实施方式中,所述LTVEC的范围为约100kb至125kb。在一个实施方式中,所述LTVEC的范围为约125kb至约150kb。在一个实施方式中,所述LTVEC的范围为约150kb至约175kb。在一个实施方式中,所述LTVEC的范围为约175kb至约200kb。在一个实施方式中,所述LTVEC的范围为约200kb至约225kb。在一个实施方式中,所述LTVEC的范围为约225kb至约250kb。在一个实施方式中,所述LTVEC的范围为约250kb至约275kb。在一个实施方式中,所述LTVEC的范围为约275kb至约300kb。

[0057] 在一个实施方式中,所述靶向载体的同源臂来源于BAC文库、粘粒文库或P1噬菌体文库。在一个实施方式中,所述同源臂来源于使用常规方法不能靶向的非人动物的基因组基因座。在一个实施方式中,所述同源臂来源于合成的DNA。

[0058] 在一个实施方式中,上游同源臂和下游同源臂的总和为至少10kb。在一个实施方式中,所述上游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述下游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约5kb至约10kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约10kb至约20kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约20kb至约30kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约30kb至约40kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约40kb至约50kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约50kb至约60kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约60kb至约70kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约70kb至约80kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约80kb至约90kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约90kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约100kb至约110kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约110kb至约120kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约120kb至约130kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约130kb至约140kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约140kb至约150kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约150kb至约160kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约160kb至约170kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约170kb至约180kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约180kb至约190kb。在一个实施方式中,所述上游和所述下游同源臂的范

围为约190kb至约200kb。

[0059] 在一个实施方式中,所述靶向载体包含选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在原核细胞中具有活性。在一个实施方式中,所述核酸在原核和真核细胞中均具有活性。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其组合。

[0060] 在一个实施方式中,所述插入核酸为约5kb至约300kb。在一个实施方式中,所述插入核酸为约5kb至约200kb。在一个实施方式中,所述插入核酸为约5kb至约10kb。在一个实施方式中,所述插入核酸为约10kb至约20kb。在一个实施方式中,所述插入核酸为约20kb至约30kb。在一个实施方式中,所述插入核酸为约30kb至约40kb。在一个实施方式中,所述插入核酸为约40kb至约50kb。在一个实施方式中,所述插入核酸为约60kb至约70kb。在一个实施方式中,所述插入核酸为约80kb至约90kb。在一个实施方式中,所述插入核酸为约90kb至约100kb。在一个实施方式中,所述插入核酸为约100kb至约110kb。在一个实施方式中,所述插入核酸为约120kb至约130kb。在一个实施方式中,所述插入核酸为约130kb至约140kb。在一个实施方式中,所述插入核酸为约140kb至约150kb。在一个实施方式中,所述插入核酸为约150kb至约160kb。在一个实施方式中,所述插入核酸为约160kb至约170kb。在一个实施方式中,所述插入核酸为约170kb至约180kb。在一个实施方式中,所述插入核酸为约180kb至约190kb。在一个实施方式中,所述插入核酸为约190kb至约200kb。

[0061] 在一个实施方式中,所述插入核酸含有翼侧为位点特异性重组靶序列的核酸。在一个实施方式中,所述核酸含有基因组核酸。在一个实施方式中,所述基因组核酸来源于小鼠、人或其组合。在一个实施方式中,所述位点特异性重组靶序列选自下组:loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attP、att、FRT、rox及其组合。

[0062] 在一个实施方式中,所述插入核酸包含条件等位基因。在一个实施方式中,所述条件等位基因是如在US 2011/0104799中所描述的多功能等位基因,其全部内容通过引用并入本申请。在一个实施方式中,所述条件等位基因包含:(a)在相对于靶基因转录的有意义方向的驱动序列,和在有意义或反义方向的药物选择性表达盒;(b)在反义方向的感兴趣的核苷酸序列(NSI)和条件倒置模块(COIN,其利用外显子分割的内含子和可逆的基因捕获分子;参见例如US 2011/0104799,其全部内容通过引用并入)和(c)可重组的单元,其在与第一重组酶接触后重组形成条件等位基因,所述条件等位基因(i)缺乏驱动序列和DSC并且(ii)在有意义方向含有NSI和在反义方向含有COIN。

[0063] 在一个实施方式中,所述插入核酸包含选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在原核细胞中具有活性。在一个实施方式中,所述核酸在原核和真核细胞中均具有活性。在一个实施方式中,所述选择性表达盒的翼侧具有位点特异性重组靶序列。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其

组合。

[0064] 在一个实施方式中,所述插入核酸含有与启动子可操作地连接的报告基因,其中所述报告基因编码选自下组的报告蛋白:LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(EYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶及其组合。在一个实施方式中,所述报告基因在诱导型启动子的控制下表达。在一个实施方式中,所述报告基因在内源性启动子的控制下表达。在一个实施方式中,所述报告基因在外源性启动子的控制下表达。在一个实施方式中,所述报告基因在特定的细胞类型中表达。在一个实施方式中,所述报告基因以组织特异性的方式表达。在一个实施方式中,所述报告基因以发育阶段特异性的方式表达。

[0065] 在一个实施方式中,所述插入核酸整合进入所述基因组的靶基因座引入如本申请所述的一个或多个基因修饰。在一个实施方式中,所述基因修饰是内源性核酸序列的缺失。在一个实施方式中,所述基因修饰是将外源性核酸序列加入所述基因组的靶基因座中。在一个实施方式中,所述基因修饰是在所述基因组的靶基因座使用外源性核酸序列取代内源性核酸序列。在一个实施方式中,所述外源性核酸序列是非小鼠核酸序列。在一个实施方式中,所述外源性核酸序列是人源核酸序列。在一个实施方式中,所述基因修饰是敲除、缺失、插入、取代(“敲入”)、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换或其组合。

[0066] 在一个实施方式中,所述插入核酸与小鼠核酸序列是同源的。在一个实施方式中,所述插入核酸序列是人源核酸。在一个实施方式中,所述插入核酸是基因组核酸的片段。在一个实施方式中,所述基因组核酸是小鼠基因组核酸、人源基因组核酸或其组合。在一个实施方式中,如上文所述的插入核酸的范围为约5kb至约200kb。

[0067] 在一个实施方式中,所述插入核酸与小鼠核酸序列是直系同源的。在一个实施方式中,所述插入核酸是人源核酸。在一个实施方式中,所述插入核酸是基因组核酸的片段。在一个实施方式中,所述基因组核酸是小鼠基因组核酸、人源基因组核酸或其组合。在一个实施方式中,如上文所述的插入核酸的范围为约5kb至约200kb。

[0068] 在一个实施方式中,所述插入核酸包含基因组基因座,所述基因座编码在神经系统、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。在一个实施方式中,所述插入核酸包含基因组基因座,所述基因座编码在骨髓或骨髓来源细胞中表达的蛋白。在一个实施方式中,所述插入核酸包含基因组基因座,所述基因座编码在脾细胞中表达的蛋白。在一个实施方式中,所述基因组基因座包含小鼠基因组核酸序列、人源基因组核酸序列或其组合。在一个实施方式中,所述核酸包含基因组基因座,所述基因座编码在B细胞中表达的蛋白。在一个实施方式中,所述核酸包含基因组基因座,所述基因座编码在未成熟的B细胞中表达的蛋白。在一个实施方式中,所述核酸包含基因组基因座,所述基因座编码在成熟的B细胞中表达的蛋白。

[0069] 在一个实施方式中,所述插入核酸包含基因组核酸序列,所述核酸序列编码人源免疫球蛋白重链可变区氨基酸序列。在一个实施方式中,所述插入核酸包含与免疫球蛋白重链恒定区核酸序列可操作地连接的未经重排的人源免疫球蛋白重链可变区核酸序列。

[0070] 在一个实施方式中,所述基因组核酸序列包含与免疫球蛋白重链恒定区核酸序列可操作地连接的未经重排的人源免疫球蛋白重链可变区核酸序列。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列是小鼠免疫球蛋白重链恒定区核酸序列或人源免疫球蛋白重链恒定区核酸序列或者其组合。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列选自C_H1、铰链、C_H2、C_H3及其组合。在一个实施方式中,所述重链恒定区核酸序列包含C_H1-铰链-C_H2-C_H3。在一个实施方式中,所述基因组核酸序列包含与免疫球蛋白重链恒定区核酸序列可操作地连接的重排的人源免疫球蛋白重链可变区核酸序列。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列是小鼠免疫球蛋白重链恒定区核酸序列或人源免疫球蛋白重链恒定区核酸序列或者其组合。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列选自C_H1、铰链、C_H2、C_H3及其组合。在一个实施方式中,所述重链恒定区核酸序列包含C_H1-铰链-C_H2-C_H3。

[0071] 在一个实施方式中,所述插入核酸包含基因组核酸序列,所述核酸序列编码人源免疫球蛋白轻链可变区氨基酸序列。在一个实施方式中,所述基因组核酸序列包含未经重排的人源λ和/或κ轻链可变区核酸序列。在一个实施方式中,所述插入核酸包含与小鼠或人源免疫球蛋白轻链恒定区核酸序列可操作地连接的未经重排的人源λ或κ轻链可变区核酸序列,所述小鼠或人源免疫球蛋白轻链恒定区核酸序列选自λ轻链恒定区核酸序列和κ轻链恒定区核酸序列。在一个实施方式中,所述基因组核酸序列包含重排的人源λ和/或κ轻链可变区核酸序列。在一个实施方式中,所述未经重排的或重排的λ和/或κ轻链可变区核酸序列可操作地连接至小鼠、大鼠或人源免疫球蛋白轻链恒定区核酸序列,所述轻链恒定区核酸序列选自λ轻链恒定区核酸序列和κ轻链恒定区核酸序列。在一个实施方式中,所述插入核酸包含与小鼠或人源免疫球蛋白轻链恒定区核酸序列可操作地连接的重排的人源λ或κ轻链可变区核酸序列,所述小鼠或人源免疫球蛋白轻链恒定区核酸序列选自λ轻链恒定区核酸序列和κ轻链恒定区核酸序列。

[0072] 在一个实施方式中,所述插入核酸包含来自不同物种的核酸序列。

[0073] 在一个实施方式中,所述插入核酸包含人源核酸序列。在一个实施方式中,所述人源核酸序列编码胞外蛋白。在一个实施方式中,所述人源核酸序列编码受体的配体。在一个实施方式中,所述配体是细胞因子。在一个实施方式中,所述细胞因子是选自下述的趋化因子:CCL、CXCL、CX3CL和XCL。在一个实施方式中,所述细胞因子是肿瘤坏死因子(TNF)。在一个实施方式中,所述细胞因子是白介素(IL)。在一个实施方式中,所述白介素选自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IL-36。在一个实施方式中,所述白介素是IL-2。在一个实施方式中,所述人源基因组核酸序列编码细胞质蛋白。在一个实施方式中,所述人源基因组核酸序列编码膜蛋白。在一个实施方式中,所述膜蛋白是受体。在一个实施方式中,所述受体是细胞因子受体。在一个实施方式中,所述细胞因子受体是白介素受体。在一个实施方式中,所述白介素受体是白介素2受体α。在一个实施方式中,所述白介素受体是白介素2受体β。在一个实施方式中,所述白介素受体是白介素2受体γ。在一个实施方式中,所述人源基因组核酸序列编码核蛋白。在一个实施方式中,所述核蛋白是核受体。

[0074] 在一个实施方式中,所述插入核酸包含在编码序列中的基因修饰。在一个实施方

式中,所述基因修饰包含编码序列的缺失突变。在一个实施方式中,所述基因修饰包含两个内源性编码序列的融合。

[0075] 在一个实施方式中,所述插入核酸包含编码突变的人源蛋白的人源核酸序列。在一个实施方式中,所述突变的人源蛋白的特征在于具有改变的结合特性、改变的位置、改变的表达和/或改变的表达模式。在一个实施方式中,所述人源核酸序列包含至少一个人类疾病等位基因。在一个实施方式中,所述人类疾病等位基因是神经系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是心血管疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肾脏疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肌肉疾病等位基因。在一个实施方式中,所述人类疾病等位基因是血液疾病等位基因。在一个实施方式中,所述人类疾病等位基因是致癌基因等位基因。在一个实施方式中,所述人类疾病等位基因是免疫系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是显性等位基因。在一个实施方式中,所述人类疾病等位基因是隐性等位基因。在一个实施方式中,所述人类疾病等位基因包含单核苷酸多态性 (SNP) 等位基因。

[0076] 在一个实施方式中,所述插入核酸包含调控序列。在一个实施方式中,所述调控序列是启动子序列。在一个实施方式中,所述调控序列是增强子序列。在一个实施方式中,所述调控序列是转录阻遏结合序列。在一个实施方式中,所述插入核酸包含人源核酸序列,其中所述人源核酸序列包含非蛋白编码序列的缺失,但是不包含蛋白编码序列的缺失。在一个实施方式中,所述非蛋白编码序列的缺失包含调控序列的缺失。在一个实施方式中,所述调控元件的缺失包括启动子序列的缺失。在一个实施方式中,所述调控元件的缺失包括增强子序列的缺失。

[0077] 在一个方面,本申请提供了在哺乳动物细胞中修饰基因组的靶基因座的方法,所述方法包括 (a) 引入哺乳动物细胞: (i) 核酸酶试剂,所述核酸酶试剂在所述基因组的靶基因座或其附近形成单链或双链断裂,和 (ii) 大靶向载体 (LTVEC),所述大靶向载体包含翼侧为上游同源臂和下游同源臂的插入核酸。在一些实施方式中,所述方法进一步包括 (b) 选择经靶向的哺乳动物细胞,所述哺乳动物细胞包含在所述基因组的靶基因座中的所述插入核酸。在一些实施方式中,在选择步骤 (b) 通过等位基因修饰 (MOA) 检测实施。

[0078] 在一个实施方式中,所述哺乳动物细胞是多能细胞。在一个实施方式中,所述多能细胞是胚胎干 (ES) 细胞。在一个实施方式中,所述多能细胞是非人ES细胞。在一个实施方式中,所述多能细胞是诱导多能干 (iPS) 细胞。在一个实施方式中,所述诱导多能 (iPS) 细胞来源于成纤维细胞。在一个实施方式中,所述诱导多能 (iPS) 细胞来源于人成纤维细胞。在一个实施方式中,所述多能细胞是造血干细胞 (HSC)。在一个实施方式中,所述多能细胞是神经干细胞 (NSC)。在一个实施方式中,所述多能细胞是外胚层干细胞。在一个实施方式中,所述多能细胞是发育受限的祖细胞。

[0079] 在一个实施方式中,所述多能细胞是啮齿动物多能细胞。在一个实施方式中,所述啮齿动物多能细胞是大鼠多能细胞。在一个实施方式中,所述大鼠多能细胞是大鼠ES细胞。在一个实施方式中,所述啮齿动物多能细胞是小鼠多能细胞。在一个实施方式中,所述多能细胞是小鼠胚胎干 (ES) 细胞。

[0080] 在一个实施方式中,所述哺乳动物细胞是永生化的小鼠或大鼠细胞。在一个实施方式中,所述哺乳动物细胞是永生化的

人成纤维细胞。在一个实施方式中,所述哺乳动物细胞是癌细胞。在一个实施方式中,所述哺乳动物细胞是人癌细胞。

[0081] 在一个实施方式中,所述哺乳动物细胞是从患有疾病的患者中分离的人源细胞。在一个实施方式中,所述人源细胞包含编码突变蛋白的人源核酸序列。在一个实施方式中,所述突变的人源蛋白的特征在于具有改变的结合特性、改变的位置、改变的表达和/或改变的表达模式。在一个实施方式中,所述哺乳动物细胞是分离自患有疾病的患者的人源细胞,并且其中所述人源细胞在其基因组中包含至少一个人疾病等位基因。在一个实施方式中,所述人源核酸序列包含至少一个人类疾病等位基因。在一个实施方式中,所述人源核酸序列包含至少一个人类疾病等位基因。在一个实施方式中,所述人类疾病等位基因是神经系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是心血管疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肾脏疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肌肉疾病等位基因。在一个实施方式中,所述人类疾病等位基因是血液疾病等位基因。在一个实施方式中,所述人类疾病等位基因是致癌基因等位基因。在一个实施方式中,所述人类疾病等位基因是免疫系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是显性等位基因。在一个实施方式中,所述人类疾病等位基因是隐性等位基因。在一个实施方式中,所述人类疾病等位基因包含单核苷酸多态性 (SNP) 等位基因。

[0082] 在一个实施方式中,所述基因组的靶基因座选自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基因座和ERBB4基因座。

[0083] 在一个实施方式中,所述基因组的靶基因座包含人源基因组序列。在一个实施方式中,所述基因组的靶基因座包含非人动物的基因组核酸序列。在一个实施方式中,所述非人动物是啮齿动物。在一个实施方式中,所述啮齿动物选自小鼠、大鼠和仓鼠。

[0084] 在一个实施方式中,位于所述基因组的靶基因座的上文所述的至少一个人类疾病等位基因被所述插入核酸取代。在一个实施方式中,所述人类疾病等位基因的取代是由敲除、缺失、插入、取代(“敲入”)、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换或其组合所介导。

[0085] 在一个实施方式中,所述核酸酶试剂与所述大靶向载体(LTVEC)一起引入。在一个实施方式中,在一段时间中所述核酸酶试剂与所述LTVEC分别引入。在一个实施方式中,所述核酸酶试剂先于所述LTVEC引入。在一个实施方式中,所述核酸酶试剂在所述LTVEC引入之后引入。

[0086] 在一个实施方式中,与单独使用所述LTVEC相比,将所述LTVEC与所述核酸酶试剂联用具有提高的靶向效率。在一个实施方式中,与单独使用所述LTVEC相比,当将所述LTVEC与所述核酸酶试剂联用时,所述LTVEC的靶向效率提高至少两倍。在一个实施方式中,与单独使用所述LTVEC相比,当将所述LTVEC与所述核酸酶试剂联用时,所述LTVEC的靶向效率提高至少三倍。在一个实施方式中,与单独使用所述LTVEC相比,当将所述LTVEC与所述核酸酶试剂联用时,所述LTVEC的靶向效率提高至少四倍。

[0087] 在一个实施方式中,所述核酸酶试剂是表达构建体,所述表达构建体包含编码核酸酶的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子是组成型激活的启动子。在一个实施方式中,所述启动子是诱导型启动子。在一个实施方

式中,所述启动子在哺乳动物细胞中具有活性。在一个实施方式中,所述核酸酶试剂是编码核酸酶的mRNA。在一个实施方式中,所述核酸酶试剂是编码核酸内切酶的mRNA。

[0088] 在一个实施方式中,所述核酸酶试剂是锌指核酸酶(ZFN)。在一个实施方式中,所述ZFN的各个单体均含有3个或更多个基于锌指的DNA结合结构域,其中各基于锌指的DNA结合结构域均结合至3bp子位点。在一个实施方式中,所述ZFN是嵌合蛋白,所述嵌合蛋白包含与独立的核酸酶可操作地连接的基于锌指的DNA结合结构域。在一个实施方式中,所述独立的核酸内切酶是FokI核酸内切酶。在一个实施方式中,所述核酸酶试剂包含第一ZFN和第二ZFN,其中各个所述第一ZFN和所述第二ZFN均与FokI核酸酶可操作地连接,其中所述第一和所述第二ZFN识别各条靶DNA链中的被约6bp至约40bp的切割位点所分隔的两个相邻的靶DNA序列,并且其中所述FokI核酸酶二聚化并且形成双链断裂。

[0089] 在一个实施方式中,所述核酸酶试剂是转录激活因子样效应物核酸酶(TALEN)。在一个实施方式中,所述TALEN的各个单体均包含12-25个TAL重复,其中各个TAL重复均与1bp子位点结合。在一个实施方式中,所述核酸酶试剂是嵌合蛋白,所述嵌合蛋白包含与独立的核酸酶可操作地连接的基于TAL重复的DNA结合结构域。在一个实施方式中,所述独立的核酸酶是FokI核酸内切酶。在一个实施方式中,所述核酸酶试剂包含第一基于TAL重复的DNA结合结构域和第二基于TAL重复的DNA结合结构域,其中各个所述第一和所述第二基于TAL重复的DNA结合结构域与FokI核酸酶可操作地连接,其中所述第一和所述第二基于TAL重复的DNA结合结构域识别各条靶DNA序列中的被约6bp至约40bp的切割位点所分隔的两个相邻的靶DNA序列,并且其中所述FokI核酸酶二聚化并且形成双链断裂。

[0090] 在一个实施方式中,所述核酸酶的各个单体识别至少9个核苷酸的靶序列。在一个实施方式中,所述靶序列的长度为约9至约12个核苷酸。在一个实施方式中,所述靶序列的长度为约12至约15个核苷酸。在一个实施方式中,所述靶序列的长度为约15至约18个核苷酸。在一个实施方式中,所述靶序列的长度为约18至约21个核苷酸。

[0091] 在一个实施方式中,所述核酸酶试剂的靶核酸序列位于内含子中。在一个实施方式中,所述靶核酸序列位于外显子中。在一个实施方式中,所述靶核酸序列位于启动子中。在一个实施方式中,所述靶核酸序列在非蛋白编码区中。在一个实施方式中,所述非蛋白编码区是调控区。在一个实施方式中,所述靶核酸序列位于启动子调控区。在一个实施方式中,所述靶核酸序列位于增强子区中。

[0092] 在一个实施方式中,所述核酸酶试剂是大范围核酸酶。在一个实施方式中,所述大范围核酸酶识别12至40个碱基对的双链DNA序列。在一个实施方式中,所述大范围核酸酶识别在所述基因组中的一个极佳匹配的靶序列。在一个实施方式中,所述大范围核酸酶是归巢核酸酶。在一个实施方式中,所述归巢核酸酶是LAGLIDADG家族的归巢核酸酶。在一个实施方式中,所述LAGLIDADG家族的归巢核酸酶选自I-SceI、I-CreI和I-DmO1。

[0093] 在一个实施方式中,所述LTVEC的范围为约50kb至约300kb。在一个实施方式中,所述LTVEC的范围为约50kb至约75kb。在一个实施方式中,所述LTVEC的范围为约75kb至约100kb。在一个实施方式中,所述LTVEC的范围为约100kb至约125kb。在一个实施方式中,所述LTVEC的范围为约125kb至约150kb。在一个实施方式中,所述LTVEC的范围为约150kb至约175kb。在一个实施方式中,所述LTVEC的范围为约175kb至约200kb。在一个实施方式中,所述LTVEC的范围为约200kb至约225kb。在一个实施方式中,所述LTVEC的范围为约225kb至约

250kb。在一个实施方式中,所述LTVEC的范围为约250kb至约275kb。在一个实施方式中,所述LTVEC的范围为约275kb至约300kb。

[0094] 在一个实施方式中,所述LTVEC的同源臂来源于BAC文库、粘粒文库或P1噬菌体文库。在一个实施方式中,所述同源臂来源于使用常规方法不能靶向的非人动物基因组基因座。在一个实施方式中,所述同源臂来源于合成的DNA。

[0095] 在一个实施方式中,上游同源臂和下游同源臂的总和是至少10kb。在一个实施方式中,所述上游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述下游同源臂的范围为约5kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约5kb至约10kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约10kb至约20kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约20kb至约30kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约30kb至约40kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约40kb至约50kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约50kb至约60kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约60kb至约70kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约70kb至约80kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约80kb至约90kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约90kb至约100kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约100kb至约110kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约110kb至约120kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约120kb至约130kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约130kb至约140kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约140kb至约150kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约150kb至约160kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约160kb至约170kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约170kb至约180kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约180kb至约190kb。在一个实施方式中,所述上游和所述下游同源臂的范围为约190kb至约200kb。

[0096] 在一个实施方式中,所述靶向载体包含选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在原核细胞中具有活性。在一个实施方式中,所述启动子在原核和真核细胞中均具有活性。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其组合。

[0097] 在一个实施方式中,所述插入核酸为约5kb至约300kb。在一个实施方式中,所述插入核酸为约5kb至约200kb。在一个实施方式中,所述插入核酸为约5kb至约10kb。在一个实施方式中,所述插入核酸为约10kb至约20kb。在一个实施方式中,所述插入核酸为约20kb至约30kb。在一个实施方式中,所述插入核酸为约30kb至约40kb。在一个实施方式中,所述插入核酸为约40kb至约50kb。在一个实施方式中,所述插入核酸为约60kb至约70kb。在一个实施方式中,所述插入核酸为约80kb至约90kb。在一个实施方式中,所述插入核酸为约90kb至约100kb。在一个实施方式中,所述插入核酸为约100kb至约110kb。在一个实施方式中,所述

插入核酸为约120kb至约130kb。在一个实施方式中,所述插入核酸为约130kb至约140kb。在一个实施方式中,所述插入核酸为约140kb至约150kb。在一个实施方式中,所述插入核酸为约150kb至约160kb。在一个实施方式中,所述插入核酸为约160kb至约170kb。在一个实施方式中,所述插入核酸为约170kb至约180kb。在一个实施方式中,所述插入核酸为约180kb至约190kb。在一个实施方式中,所述插入核酸为约190kb至约200kb。

[0098] 在一个实施方式中,所述插入核酸含有翼侧为位点特异性重组靶序列的核酸。在一个实施方式中,所述核酸含有基因组核酸。在一个实施方式中,所述基因组核酸来源于小鼠、人或其组合。在一个实施方式中,所述位点特异性重组靶序列选自下组:loxP、lox511、lox2272、lox66、lox71、loxM2、lox5171、FRT、FRT11、FRT71、attP、att、FRT、rox及其组合。

[0099] 在一个实施方式中,所述插入核酸含有条件等位基因。在一个实施方式中,所述条件等位基因是如在US 2011/0104799中所描述的多功能等位基因,其全部内容通过引用并入本申请。在一个实施方式中,所述条件等位基因包含:(a)在相对于靶基因转录的有意义方向的驱动序列,和在有意义或反义方向的药物选择性表达盒;(b)在反义方向的感兴趣的核苷酸序列(NSI)和条件倒置模块(COIN,其利用外显子分割的内含子和可逆的基因捕获样分子;参见例如US 2011/0104799,其全部内容通过引用并入)和(c)可重组的单元,其在与第一重组酶接触后重组形成条件等位基因,所述条件等位基因(i)缺乏驱动序列和DSC并且(ii)在有意义方向含有NSI和在反义方向含有COIN。

[0100] 在一个实施方式中,所述插入核酸包含选择性表达盒。在一个实施方式中,所述选择性表达盒包含编码选择性标记物的核酸序列,其中所述核酸序列与启动子可操作地连接。在一个实施方式中,所述启动子在哺乳动物细胞中具有活性。在一个实施方式中,所述核酸在原核细胞中具有活性。在一个实施方式中,所述选择性表达盒的翼侧具有位点特异性重组靶序列。在一个实施方式中,所述选择性标记物选自新霉素磷酸转移酶(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/鸟嘌呤磷酸核糖转移酶(gpt)和单纯疱疹病毒胸腺嘧啶激酶(HSV-k)及其组合。

[0101] 在一个实施方式中,所述插入核酸包含报告基因。在一个实施方式中,所述插入核酸含有与启动子可操作地连接的报告基因。在一些实施方式中,所述报告基因编码选自下组的报告蛋白:LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(EYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶及其组合。在一个实施方式中,所述报告基因在诱导型启动子的控制下表达。在一个实施方式中,所述报告基因在内源性启动子的控制下表达。在一个实施方式中,所述报告基因在外源性启动子的控制下表达。在一个实施方式中,所述报告基因在特定的细胞类型中表达。在一个实施方式中,所述报告基因以组织特异性的方式表达。在一个实施方式中,所述报告基因以发育阶段特异性的方式表达。

[0102] 在一个实施方式中,所述插入核酸整合进入所述基因组的靶基因座引入如本申请所述的一个或多个基因修饰。在一个实施方式中,所述基因修饰是内源性核酸序列的缺失。在一个实施方式中,所述基因修饰是将外源性核酸序列加入所述基因组的靶基因座中。在一个实施方式中,所述基因修饰是在所述基因组的靶基因座使用外源性核酸序列取代内源性核酸序列。在一个实施方式中,所述外源性核酸序列是非小鼠核酸序列。在一个实施方式

中,所述外源性核酸序列是人源核酸序列。在一个实施方式中,所述基因修饰是敲除、缺失、插入、取代(“敲入”)、点突变、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换或其组合。

[0103] 在一个实施方式中,所述插入核酸与小鼠核酸序列是同源的。在一个实施方式中,所述插入核酸序列是人源核酸。在一个实施方式中,所述插入核酸是基因组核酸的片段。在一个实施方式中,所述基因组核酸是小鼠基因组核酸、人源基因组核酸或其组合。在一个实施方式中,如上文所述的插入核酸的范围为约5kb至约200kb。

[0104] 在一个实施方式中,所述插入核酸与小鼠核酸序列是直系同源的。在一个实施方式中,所述插入核酸是人源核酸。在一个实施方式中,所述插入核酸是基因组核酸的片段。在一个实施方式中,所述基因组核酸是小鼠基因组核酸、人源基因组核酸或其组合。在一个实施方式中,如上文所述的插入核酸的范围为约5kb至约200kb。

[0105] 在一个实施方式中,所述插入核酸包含基因组基因座,所述基因座编码在神经系统、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。在一个实施方式中,所述插入核酸包含基因组基因座,所述基因座编码在骨髓或骨髓来源细胞中表达的蛋白。在一个实施方式中,所述插入核酸包含基因组基因座,所述基因座编码在脾细胞中表达的蛋白。在一个实施方式中,所述基因组基因座包含小鼠基因组核酸序列、人源基因组核酸序列或其组合。在一个实施方式中,所述核酸包含基因组基因座,所述基因座编码在B细胞中表达的蛋白。在一个实施方式中,所述核酸包含基因组基因座,所述基因座编码在未成熟的B细胞中表达的蛋白。在一个实施方式中,所述核酸包含基因组基因座,所述基因座编码在成熟的B细胞中表达的蛋白。

[0106] 在一个实施方式中,所述插入核酸包含基因组核酸序列,所述核酸序列编码人源免疫球蛋白重链可变区氨基酸序列。

[0107] 在一个实施方式中,所述基因组核酸序列包含与免疫球蛋白重链恒定区核酸序列可操作地连接的未经重排的人源免疫球蛋白重链可变区核酸序列。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列是小鼠免疫球蛋白重链恒定区核酸序列或人源免疫球蛋白重链恒定区核酸序列或者其组合。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列选自C_H1、铰链、C_H2、C_H3及其组合。在一个实施方式中,所述重链恒定区核酸序列包含C_H1-铰链-C_H2-C_H3。在一个实施方式中,所述基因组核酸序列包含与免疫球蛋白重链恒定区核酸序列可操作地连接的重排的人源免疫球蛋白重链可变区核酸序列。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列是小鼠免疫球蛋白重链恒定区核酸序列或人源免疫球蛋白重链恒定区核酸序列或者其组合。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列选自C_H1、铰链、C_H2、C_H3及其组合。在一个实施方式中,所述重链恒定区核酸序列包含C_H1-铰链-C_H2-C_H3。

[0108] 在一个实施方式中,所述插入核酸包含基因组核酸序列,所述核酸序列编码人源免疫球蛋白轻链可变区氨基酸序列。在一个实施方式中,所述基因组核酸序列包含未经重排的人源λ和/或κ轻链可变区核酸序列。在一个实施方式中,所述基因组核酸序列包含重排的人源λ和/或κ轻链可变区核酸序列。在一个实施方式中,所述未经重排的或重排的λ和/或κ轻链可变区核酸序列可操作地连接至小鼠或人源免疫球蛋白轻链恒定区核酸序列,所述

疫球蛋白轻链恒定区核酸序列选自 λ 轻链恒定区核酸序列和 κ 轻链恒定区核酸序列。

[0109] 在一个实施方式中,所述插入核酸包含人源核酸序列。在一个实施方式中,所述人源核酸序列编码胞外蛋白。在一个实施方式中,所述人源核酸序列编码受体的配体。在一个实施方式中,所述配体是细胞因子。在一个实施方式中,所述细胞因子是选自下述的趋化因子:CCL、CXCL、CX3CL和XCL。在一个实施方式中,所述细胞因子是肿瘤坏死因子(TNF)。在一个实施方式中,所述细胞因子是白介素(IL)。在一个实施方式中,所述白介素选自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IL-36。在一个实施方式中,所述白介素是IL-2。在一个实施方式中,所述人源基因组核酸序列编码细胞质蛋白。在一个实施方式中,所述人源基因组核酸序列编码膜蛋白。在一个实施方式中,所述膜蛋白是受体。在一个实施方式中,所述受体是细胞因子受体。在一个实施方式中,所述细胞因子受体是白介素受体。在一个实施方式中,所述白介素受体是白介素2受体 α 。在一个实施方式中,所述白介素受体是白介素2受体 β 。在一个实施方式中,所述白介素受体是白介素2受体 γ 。在一个实施方式中,所述人源基因组核酸序列编码核蛋白。在一个实施方式中,所述核蛋白是核受体。

[0110] 在一个实施方式中,所述插入核酸包含在编码序列中的基因修饰。在一个实施方式中,所述基因修饰包含编码序列的缺失突变。在一个实施方式中,所述基因修饰包含两个内源性编码序列的融合。

[0111] 在一个实施方式中,所述插入核酸包含编码突变的人源蛋白的人源核酸序列。在一个实施方式中,所述突变的人源蛋白的特征在于具有改变的结合性质、改变的位置、改变的表达和/或改变的表达模式。在一个实施方式中,所述人源核酸序列包含至少一个人类疾病等位基因。在一个实施方式中,所述人类疾病等位基因是神经系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是心血管疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肾脏疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肌肉疾病等位基因。在一个实施方式中,所述人类疾病等位基因是血液疾病等位基因。在一个实施方式中,所述人类疾病等位基因是致癌基因等位基因。在一个实施方式中,所述人类疾病等位基因是免疫系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是显性等位基因。在一个实施方式中,所述人类疾病等位基因是隐性等位基因。在一个实施方式中,所述人类疾病等位基因包含单核苷酸多态性(SNP)等位基因。

[0112] 在一个实施方式中,所述插入核酸包含调控序列。在一个实施方式中,所述调控序列是启动子序列。在一个实施方式中,所述调控序列是增强子序列。在一个实施方式中,所述调控序列是转录阻遏蛋白结合序列。在一个实施方式中,所述插入核酸包含人源核酸序列,其中所述人源核酸序列包含非蛋白编码序列的缺失,但是不包含蛋白编码序列的缺失。在一个实施方式中,所述非蛋白编码序列的缺失包含调控序列的缺失。在一个实施方式中,所述调控元件的缺失包括启动子序列的缺失。在一个实施方式中,所述调控元件的缺失包括增强子序列的缺失。

[0113] 在一个方面,本申请提供了一种由如本申请所述的方法制备的哺乳动物细胞。在一个实施方式中,所述哺乳动物细胞包含插入核酸,所述插入核酸包含在基因组的靶基因座的一个或多个如本申请所述的基因修饰。

[0114] 在一个实施方式中,所述哺乳动物细胞是多能细胞。在一个实施方式中,所述多能细胞是胚胎干(ES)细胞。在一个实施方式中,所述多能细胞是诱导多能干(iPS)细胞。在一个实施方式中,所述诱导多能(iPS)细胞来源于成纤维细胞。在一个实施方式中,所述诱导多能(iPS)细胞来源于人成纤维细胞。在一个实施方式中,所述多能细胞是造血干细胞(HSC)。在一个实施方式中,所述多能细胞神经干细胞(NSC)。在一个实施方式中,所述多能细胞是外胚层干细胞。在一个实施方式中,所述多能细胞是发育受限的祖细胞。

[0115] 在一个实施方式中,所述多能细胞是小鼠多能细胞。在一个实施方式中,所述多能细胞是小鼠胚胎干(ES)细胞。

[0116] 在一个实施方式中,所述哺乳动物细胞是永生化的小鼠或大鼠细胞。在一个实施方式中,所述哺乳动物细胞是永生化的来源细胞。在一个实施方式中,所述哺乳动物细胞是人成纤维细胞。在一个实施方式中,所述哺乳动物细胞是癌细胞。在一个实施方式中,所述哺乳动物细胞是人癌细胞。

[0117] 在一个实施方式中,所述基因组的靶基因座选自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基因座和ERBB4基因座。

[0118] 在一个实施方式中,所述基因组的靶基因座包含一个或多个如本申请所述的基因修饰。在一个实施方式中,所述基因修饰是内源性核酸序列的缺失。在一个实施方式中,所述基因修饰是将外源性核酸序列加入所述基因组的靶基因座中。在一个实施方式中,所述基因修饰是在所述基因组的靶基因座使用外源性核酸序列取代内源性核酸序列。在一个实施方式中,所述外源性核酸序列是非小鼠核酸序列。在一个实施方式中,所述外源性核酸序列是人源核酸序列。在一个实施方式中,所述基因组的靶基因座包含选自敲除、缺失、插入、取代(“敲入”)、结构域交换、外显子交换、内含子交换、调控序列交换、基因交换及其组合的修饰。

[0119] 在一个实施方式中,所述基因组的靶基因座包含插入核酸,所述插入核酸与小鼠核酸序列是同源的。在一个实施方式中,所述插入核酸序列是人源核酸。在一个实施方式中,所述插入核酸是基因组核酸的片段。在一个实施方式中,所述基因组核酸是小鼠基因组核酸、人源基因组核酸或其组合。在一个实施方式中,如上文所述的插入核酸的范围为约5kb至约200kb。

[0120] 在一个实施方式中,所述基因组的靶基因座包含插入核酸,所述插入核酸与小鼠核酸序列是直系同源的。在一个实施方式中,所述插入核酸是人源核酸。在一个实施方式中,所述插入核酸是基因组核酸的片段。在一个实施方式中,所述基因组核酸是小鼠基因组核酸、人源基因组核酸或其组合。在一个实施方式中,如上文所述的插入核酸的范围为约5kb至约200kb。

[0121] 在一个实施方式中,所述基因组的靶基因座包含条件等位基因。在一个实施方式中,所述条件等位基因是如在US 2011/0104799中所描述的多功能的等位基因,其全部内容通过引用并入本申请。在一个实施方式中,所述条件等位基因包含:(a)在相对于靶基因转录的有意义方向的驱动序列,和在有意义或反义方向的药物选择性表达盒;(b)在反义方向的感兴趣的核苷酸序列(NSI)和条件倒置模块(COIN,其利用外显子分割的内含子和可逆的基因捕获样分子;参见例如US 2011/0104799,其全部内容通过引用并入)和(c)可重组的单

元,其在与第一重组酶接触后重组形成条件等位基因,所述条件等位基因(i)缺乏驱动序列和DSC以及(ii)在有意义方向含有NSI并在反义方向含有COIN。

[0122] 在一个实施方式中,所述插入核酸含有与启动子可操作地连接的报告基因,其中所述报告基因编码选自下组的报告蛋白:LacZ、mPlum、mCherry、tdTomato、mStrawberry、J-Red、DsRed、mOrange、mKO、mCitrine、Venus、YPet、增强的黄色荧光蛋白(EYFP)、Emerald、增强的绿色荧光蛋白(EGFP)、CyPet、青色荧光蛋白(CFP)、Cerulean、T-Sapphire、荧光素酶、碱性磷酸酶及其组合。在一个实施方式中,所述报告基因在诱导型启动子的控制下表达。在一个实施方式中,所述报告基因在内源性启动子的控制下表达。在一个实施方式中,所述报告基因在外源性启动子的控制下表达。在一个实施方式中,所述报告基因在特定的细胞类型中表达。在一个实施方式中,所述报告基因以组织特异性的方式表达。在一个实施方式中,所述报告基因以发育阶段特异性的方式表达。

[0123] 在一个实施方式中,所述插入核酸包含人源核酸序列,所述人源核酸序列编码在神经系统、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。在一个实施方式中,所述人源核酸序列编码在骨髓或骨髓来源细胞中表达的蛋白。在一个实施方式中,所述小鼠ES细胞的基因组包含人源基因组的基因座,所述人源基因组的基因座编码在脾细胞中表达的蛋白。在一个实施方式中,所述人源核酸编码在B细胞中表达的蛋白。在一个实施方式中,所述人源核酸编码在未成熟的B细胞中表达的蛋白。在一个实施方式中,所述人源核酸编码在成熟的B细胞中表达的蛋白。

[0124] 在一个实施方式中,所述插入核酸包含人源核酸序列,所述人源核酸序列编码人源免疫球蛋白重链可变区氨基酸序列。在一个实施方式中,所述人源核酸序列包含未经重排的人源免疫球蛋白重链可变区核酸序列。在一个实施方式中,所述未经重排的人源免疫球蛋白重链可变区核酸序列与小鼠免疫球蛋白重链恒定区核酸序列、人源免疫球蛋白重链恒定区核酸序列或其组合可操作地连接。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列选自C_H1、铰链、C_H2、C_H3及其组合。在一个实施方式中,所述重链恒定区核酸序列包含C_H1-铰链-C_H2-C_H3。在一个实施方式中,所述人源核酸序列包含重排的人源免疫球蛋白重链可变区核酸序列。在一个实施方式中,所述人源免疫球蛋白重链可变区核酸序列与小鼠免疫球蛋白重链恒定区核酸序列、人源免疫球蛋白重链恒定区核酸序列或其组合可操作地连接。在一个实施方式中,所述免疫球蛋白重链恒定区核酸序列选自C_H1、铰链、C_H2、C_H3及其组合。在一个实施方式中,所述重链恒定区核酸序列包含C_H1-铰链-C_H2-C_H3。

[0125] 在一个实施方式中,所述插入核酸包含人源核酸序列,所述核酸序列编码人源免疫球蛋白轻链可变区氨基酸序列。在一个实施方式中,所述人源核酸序列包含未经重排的人源λ和/或κ轻链可变区核酸序列。在一个实施方式中,所述人源核酸序列包含重排的人源λ和/或κ轻链可变区核酸序列。在一个实施方式中,所述未经重排的λ和/或κ轻链可变区核酸序列与选自λ轻链恒定区核酸序列和κ轻链恒定区核酸序列的小鼠或人源免疫球蛋白轻链恒定区核酸序列可操作地连接。

[0126] 在一个实施方式中,所述插入核酸包含人源核酸序列。在一个实施方式中,所述人源核酸序列编码胞外蛋白。在一个实施方式中,所述人源核酸序列编码受体的配体。在一个实施方式中,所述配体是细胞因子。在一个实施方式中,所述细胞因子是选自下述的趋化因

子:CCL、CXCL、CX3CL和XCL。在一个实施方式中,所述细胞因子是肿瘤坏死因子(TNF)。在一个实施方式中,所述细胞因子是白介素(IL)。在一个实施方式中,所述白介素选自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IL-36。在一个实施方式中,所述白介素是IL-2。在一个实施方式中,所述人源基因组核酸序列编码细胞质蛋白。在一个实施方式中,所述人源基因组核酸序列编码膜蛋白。在一个实施方式中,所述膜蛋白是受体。在一个实施方式中,所述受体是细胞因子受体。在一个实施方式中,所述细胞因子受体是白介素受体。在一个实施方式中,所述白介素受体是白介素2受体 α 。在一个实施方式中,所述白介素受体是白介素2受体 β 。在一个实施方式中,所述白介素受体是白介素2受体 γ 。在一个实施方式中,所述人源基因组核酸序列编码核蛋白。在一个实施方式中,所述核蛋白是核受体。

[0127] 在一个实施方式中,所述插入核酸包含在编码序列中的基因修饰。在一个实施方式中,所述基因修饰包含编码序列的缺失突变。在一个实施方式中,所述基因修饰包含两个内源性编码序列的融合。

[0128] 在一个实施方式中,所述插入核酸包含编码突变的人源蛋白的人源核酸序列。在一个实施方式中,所述突变的人源蛋白的特征在于具有改变的结合性质、改变的位置、改变的表达和/或改变的表达模式。在一个实施方式中,所述人源核酸序列包含至少一个人类疾病等位基因。在一个实施方式中,所述人类疾病等位基因是神经系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是心血管疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肾脏疾病等位基因。在一个实施方式中,所述人类疾病等位基因是肌肉疾病等位基因。在一个实施方式中,所述人类疾病等位基因是血液疾病等位基因。在一个实施方式中,所述人类疾病等位基因是致癌基因等位基因。在一个实施方式中,所述人类疾病等位基因是免疫系统疾病等位基因。在一个实施方式中,所述人类疾病等位基因是显性等位基因。在一个实施方式中,所述人类疾病等位基因是隐性等位基因。在一个实施方式中,所述人类疾病等位基因包含单核苷酸多态性(SNP)等位基因。

[0129] 在一个实施方式中,所述插入核酸包含调控序列。在一个实施方式中,所述调控序列是启动子序列。在一个实施方式中,所述调控序列是增强子序列。在一个实施方式中,所述调控序列是转录阻遏蛋白结合序列。在一个实施方式中,所述插入核酸包含人源核酸序列,其中所述人源核酸序列包含非蛋白编码序列的缺失,但是不包含蛋白编码序列的缺失。在一个实施方式中,所述非蛋白编码序列的缺失包含调控序列的缺失。在一个实施方式中,所述调控元件的缺失包括启动子序列的缺失。在一个实施方式中,所述调控元件的缺失包括增强子序列的缺失。

[0130] 在一个方面,本申请提供了一种制备非人动物的方法,所述非人动物在其种系中包含如本申请所述的一个或多个基因修饰,所述方法包括:

[0131] (a) 在原核细胞中使用大靶向载体(LTVEC)和核酸酶试剂修饰非人动物基因组的感兴趣的基因座,所述核酸酶试剂在基因组的感兴趣的基因座中或其附近产生单链或双链断裂,其中所述LTVEC包含翼侧为上游和下游同源臂的插入核酸,并且所述原核细胞能够表达重组酶;

[0132] (b) 选择包含经基因修饰的LTVEC的经修饰的原核细胞;

[0133] (c) 分离所述经基因修饰的LTVEC;

[0134] (d) 将所述经基因修饰的LTVEC引入所述非人动物的多能细胞以生成经基因修饰的多能细胞,所述经基因修饰的多能细胞在基因组的感兴趣的基因座中包含所述插入核酸;

[0135] (e) 选择所述经基因修饰的多能细胞;

[0136] (f) 将所述经基因修饰的多能细胞在前桑葚胚期引入所述非人动物的宿主胚胎;
和

[0137] (g) 将包含经基因修饰的多能细胞的所述宿主胚胎引入代孕母体以生成来源于所述经基因修饰的多能细胞的F0代。

[0138] 在一个实施方式中,所述非人动物是哺乳动物。在一个实施方式中,所述哺乳动物是啮齿动物。在一个实施方式中,所述啮齿动物选自小鼠、大鼠和仓鼠。在一个实施方式中,所述非人动物是小鼠,并且所述多能细胞是小鼠ES细胞。在一个实施方式中,所述非人动物是大鼠,并且所述多能细胞是大鼠ES细胞。

[0139] 在一个实施方式中,所述基因组的靶基因座包含如本申请所述的一个或多个基因修饰。

[0140] 在一个实施方式中,分离步骤(c)还包含(c)'线性化所述经基因修饰的LTVEC。

[0141] 在一个实施方式中,引入步骤(d)还包含(d)'将如本申请所述的核酸酶试剂引入所述多能细胞。在一个实施方式中,所述核酸酶试剂是锌指核酸酶(ZFN)。在一个实施方式中,所述核酸酶试剂是转录激活因子样效应物核酸酶(TALEN)。

[0142] 在一个实施方式中,通过应用如本申请所述的对所述原核细胞或所述多能细胞而言可选择的试剂进行选择步骤(b)和/或(e)。

[0143] 在一个实施方式中,通过如本申请所述的等位基因修饰(MOA)检测进行选择步骤(b)和/或(e)。

[0144] 附图简述

[0145] 图1描述了在小鼠112rg基因上的锌指核酸酶(ZFN)的切割位点。该图没有按照比例绘制。在图下部的加框的序列代表ZFN靶序列。

[0146] 发明详述

[0147] 本发明不限于特定方法,并且所描述的实验条件如方法和条件可以改变。还应理解本申请中使用的术语仅用于描述特定实施方式的目的,其并非旨在进行限制,因为本发明的范围是由权利要求所限定的。

[0148] 除非另有定义,本申请使用的所有术语和短语包括所述术语和短语在本领域中已有的含义,除非有明确相反的指示或者从使用该术语或短语的上下文中有明确相反的显示。尽管与本申请所述的那些类似或等价的任意方法和材料都能够用于实施或检测本发明,但是在此还是描述了特定的方法和材料。提及的所有出版物均通过引用并入本申请。

[0149] 定义

[0150] 本申请使用的术语“胚胎干细胞”或“ES细胞”包括胚胎来源的全能性或多能细胞,所述细胞能够在体外进行不分化的增殖,并且能够通过引入胚胎中促成发育胚胎的任何组织。本申请使用的术语“多能细胞”包括具有发育成一种以上分化的细胞类型的能力的、未经分化的细胞。

[0151] 术语“种系”指包括能够传递给后代的核酸序列的免疫球蛋白核酸序列。

[0152] 短语“重链”或“免疫球蛋白重链”包括来自任意生物体的包括免疫球蛋白重链恒定区序列的免疫球蛋白重链序列。除非另有说明,重链可变结构域包括三个重链CDR和四个框架(FR)区。重链的片段包括CDR、CDR和FR、及其组合。典型的重链具有下述可变结构域(从N末端到C末端): 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的片段(例如以微摩、纳摩或皮摩范围的 K_D 识别表位),所述片段能够由细胞表达和分泌,并且其包含至少一个CDR。重链可变结构域由可变区核苷酸序列编码,其通常包括来源于在种系中存在的 V_H 、 D_H 和 J_H 区段库的 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、位置和名称能够在IMGT数据库中找到,该数据库可以通过因特网在万维网(www)上的URL“imgt.org.”进行访问。

[0153] 短语“轻链”包括来自任意生物体的免疫球蛋白轻链序列,并且除非另有说明包括人源 κ 和 λ 轻链和 V_{preB} ,以及替代轻链。除非另有说明,轻链可变结构域通常包括三个轻链CDR和四个FR。通常情况下,全长的轻链从氨基末端到羧基末端包括包含FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4的可变结构域和轻链恒定区氨基酸序列。轻链可变结构域由轻链可变区核苷酸序列编码,其通常包括来源于在种系中存在的轻链V和J基因区段库的轻链 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、位置和名称能够在IMGT数据库中找到,该数据库可以通过因特网在万维网(www)上的URL“imgt.org.”进行访问。轻链包括那些例如存在于与第一或第二表位选择性地结合的表位结合蛋白中但即不与第一表位又不与第二表位选择性地结合的那些。轻链还包括存在于与第一或第二表位选择性地结合的表位结合蛋白中,并且结合和识别或者辅助重链结合和识别一个或多个表位的那些。

[0154] 本申请中使用的术语“同源性核酸”包括与已知的对照序列相同或基本相似的核酸序列。在一个实施方式中,术语“同源性核酸”用于表征与已知的对照序列相比具有至少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%或者甚至100%一致性的DNA或RNA序列。

[0155] 本申请中使用的术语“直系同源性核酸”包括来自一个物种的核酸序列,所述核酸序列与已知的对照序列在另一个物种中的功能等效。

[0156] 本申请中使用的术语“大靶向载体”或“LTVEC”包括来源于克隆的基因组核酸片段的用于原核细胞的大靶向载体,其大于通常用于在原核细胞中通过其他方法旨在进行同源靶向的那些。所述LTVEC的尺寸过大以至于不能通过常规检测筛选靶向事件,例如southern印迹和长范围(例如1kb-5kb)PCR。所述LTVEC的示例包括但不限于来源于细菌人工染色体(BAC)和酵母人工染色体(YAC)的载体。

[0157] 术语“等位基因修饰”或“MOA”包括在基因组中的基因或染色体基因座(loci)的一个等位基因的精确DNA序列的修饰。“等位基因修饰(MOA)”的示例包括但不限于缺失、取代或插入至少至单个的核苷酸或缺失几千碱基跨距的感兴趣的基因或染色体基因座(loci),以及在这两个端值之间的任意和全部可能的修饰。

[0158] 本申请中使用的术语“核酸酶”包括在核酸序列中诱导断裂(例如,在双链DNA序列中的单链或双链断裂)的试剂。核酸酶包括结合预先选定的或特定序列,并且在预先选定的或特定序列处或其附近切断的那些,例如经工程改造的锌指核酸酶和经工程改造的TAL效应物核酸酶。核酸酶不限于ZFN和TAL效应物核酸酶,而是可以是适用于LTVEC以实现改善的

靶向效率的任意核酸酶。非限制性示例包括其他的基于锌指的核酸酶和经工程改造的大范围核酸酶,其在预先选定的或特定序列处切断。

[0159] 适用于本发明的TAL效应物核酸酶包括本领域公知的任意TAL核酸酶。适宜的TAL核酸酶和用于制备适宜的TAL核酸酶的方法的示例在例如美国专利申请号2011/0239315 A1、2011/0269234 A1、2011/0145940 A1、2003/0232410 A1、2005/0208489 A1、2005/0026157 A1、2005/0064474 A1、2006/0188987 A1和2006/0063231 A1(其均通过引用并入本申请)中所公开的。在各种实施方式中,TAL效应物核酸酶被工程改造,使其在例如感兴趣的基因组中的靶核酸序列或其附近切断,其中所述靶核酸序列位于被LTVEC修饰的序列或在其附近。TAL效应物核酸酶是蛋白,所述蛋白包含核酸内切酶结构域和一个或多个TAL效应物DNA结合结构域,其中所述一个或多个TAL效应物DNA结合结构域包含识别预先选定的或特定核酸序列的序列。适用于本发明的TAL核酸酶包括如本申请所述的特别设计为在被LTVEC修饰的靶核酸序列或其附近结合的那些。

[0160] 短语“可操作地连接”包括某种关系,其中所述可操作地连接的组件以其预期的方式发挥功能。在一个例子中,编码蛋白的核酸序列可以与调控序列(例如启动子、增强子、沉默子序列等)可操作地连接以保持适当的转录调控。在一个例子中,免疫球蛋白可变区(或V(D)J区段)的核酸序列可以与免疫球蛋白恒定区的核酸序列可操作地连接以使得所述序列之间适当地重组成为免疫球蛋白的重链或轻链序列。

[0161] 本申请中使用的术语“启动子”和“启动子调控元件”等包括在核酸片段或基因中的控制该基因表达的核苷酸序列元件。

[0162] 本申请中使用的术语“重组位点”包括被位点特异性重组酶识别并且能够作为重组事件的底物的核苷酸序列。

[0163] 本申请中使用的术语“位点特异性重组酶”包括一组酶,所述酶能够促进“重组位点”间的重组,这两个重组位点在单一核酸分子中是物理上隔离的或在彼此分离的核酸分子上。“位点特异性重组酶”的示例包括但不限于Cre、Flp和Dre重组酶。

[0164] 使用LTVEC和核酸酶试剂的基因组基因座的修饰

[0165] 尽管对非人动物的各种基因组基因座的靶向已经取得了进步,但是仍有很多的基因组基因座或细胞类型不能使用常规的靶向策略靶向。失败的原因可能不同,但是在本申请中使用时原因包括基因座或细胞使用常规的靶向方法根本不能成功靶向或者不能正确靶向或效率非常低。常规的靶向方法包括通过常规靶向载体使用同源重组靶向。难于靶向的基因座包括甚至单独使用LTVEC不能靶向的基因座,即在缺乏以重组的单链或双链断裂的形式的辅助的情况下,或者在缺乏重组的单链或双链断裂的情况下使用LTVEC不能正确靶向或靶向基因座的效率较低的基因座。

[0166] 本申请提供了用于靶向核酸序列的组合物和方法,其使用能够形成重组的单链或双链断裂的核酸酶试剂以及大靶向载体或LTVEC,其中所述经靶向的核酸序列(或经靶向的核酸序列附近的序列)被所述LTVEC修饰。所述组合物和方法能用于修饰那些使用常规的靶向策略甚至单独使用LTVEC难于或无法修饰的基因组核酸序列。

[0167] 在各种方面,本申请提供了使用LTVEC以便在例如基因组的基因座进行靶核酸修饰的组合物和方法,其中所述靶核酸包含拟通过所述LTVEC的序列修饰的靶序列(通过使用所述LTVEC的靶核酸的同源重组),其中在所述靶核酸处或所述靶序列附近形成单链或双链

断裂。

[0168] 在各种实施方式中,在所述靶核酸或其附近存在单链或双链断裂增加LTVEC和靶核酸之间的重组效率和/或频率。在一个实施方式中,所述重组是同源重组。在另一个实施方式中,所述重组是通过非同源末端连接的插入。在各种实施方式中,当存在所述单链或双链断裂时,LTVEC序列在基因组靶基因座的靶向效率与不存在单链或双链断裂时(使用例如相同的LTVEC和相同的包含相同靶序列的靶核酸,但不存在形成单链或双链断裂的附加核酸酶)相比高至少约2倍、至少约3倍、至少约4倍。

[0169] 适用于本发明的LTVEC以及制备其的方法在例如美国专利号6,586,251、6,596,541、7,105,348和WO 2002/036789 (PCT/US01/45375) 中进行了描述。

[0170] 尽管对直接将LTVEC引入小鼠多能细胞例如小鼠ES细胞的实施方式进行了大量地讨论,但是本申请还提供了将LTVEC引入各种哺乳动物细胞类型的其他方法。此类哺乳动物细胞包括能够根据如本申请所公开的方法被基因修饰的任意哺乳动物细胞,包括例如小鼠细胞、大鼠细胞、家兔细胞、猪细胞、牛细胞、鹿细胞、绵羊细胞、山羊细胞、鸡细胞、猫细胞、犬细胞、雪貂细胞、灵长类(例如绒猴、恒河猴)细胞等。在一些实施方式中,对于适宜的可基因修饰的多能细胞不易获得的那些哺乳动物而言,使用其他方法将体细胞重编程成多能细胞,例如通过将多能诱导因子的组合引入体细胞,所述多能性诱导因子包括但不限于Oct3/4、Sox2、KLF4、Myc、Nanog、LIN28和Glis 1。

[0171] 在一个实施方式中,所述上游和下游同源臂来自与所述经靶向的基因组相同的基因组。在一个实施方式中,所述同源臂来自相关的基因组,例如所述经靶向的基因组是第一个品系的小鼠基因组,而所述靶向臂来自第二个品系的小鼠基因组,其中所述第一个品系和所述第二个品系是不同的。在一个实施方式中,所述靶向臂来源于BAC文库、粘粒文库或P1噬菌体文库。在一个实施方式中,所述同源臂来源于合成的DNA。在一个实施方式中,所述同源臂来源于使用常规方法不能靶向的基因。在一个特定的实施方式中,所述同源臂来自于基因,所述基因在不存在由核酸酶试剂诱导的单链或双链断裂的情况下使用常规的靶向技术不能被靶向,或仅能够被不正确的靶向或仅能够被以非常低的效率靶向。

[0172] 在各种实施方式中,为了便于鉴定经靶向的修饰,使用了一种高通量定量检测(即:等位基因修饰(MOA)检测)。本申请所述的MOA检测允许在基因修饰后对母体染色体中的经修饰的等位基因进行大规模筛选。所述MOA检测能够通过各种分析技术实施,包括但不限于定量PCR,例如实时PCR(qPCR)。例如,所述实时PCR包括识别所述靶基因座的第一引物集合和识别非靶向对照基因座的第二引物集合。此外,所述引物集合包含识别所述扩增序列的荧光探针。所述定量检测还能够通过各种分析技术实施,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、与固定探针的定量杂交、Invader探针®、MMP检测®、TaqMan®分子信标和Eclipse™探针技术。(参见例如US2005/0144655,其通过引用整体并入本申请)。

[0173] 在一些实施方式中,本申请所述的靶基因组基因座的各种基因修饰能够通过使用VELOCIGENE®基因工程技术利用来源于细菌人工染色体(BAC) DNA的LTVEC在细菌细胞中进行一系列同源重组反应(BHR) 来实施(参见例如美国专利号6,586,251和Valenzuela, D.M.等(2003), High-throughput engineering of the mouse genome coupled with high-resolution expression analysis(与高分辨表达分析偶联的小鼠基因组的高通量

工程改造), Nature Biotechnology 21 (6):652-659, 其通过引用整体并入本申请)。

[0174] 在一些实施方式中, 使用包含如本申请所述的各种基因修饰的经靶向的小鼠ES细胞作为供体ES细胞并通过VELOCIMOUSE®法将其引入前桑葚胚阶段的小鼠胚胎中例如8-细胞阶段的小鼠胚胎中(参见例如US 7,576,259、US 7,659,442、US 7,294,754和US 2008-0078000 A1, 其均通过引用整体并入本申请)。对包含经基因修饰的ES细胞的小鼠胚胎进行孵育直至胚泡阶段, 然后将其植入代孕母体中以产生F0小鼠。可以通过如本申请所述的等位基因修饰(MOA)检测对携带所述经基因修饰的基因组基因座的小鼠进行鉴定。将所得到的来源于所述经基因修饰的ES细胞的F0代小鼠与野生型小鼠交配以获得F1代子代。在使用特定的引物和/或探针进行基因分型后, 将针对所述经基因修饰的基因组基因座是杂合子的F1代幼仔彼此之间交配以产生针对所述经基因修饰的基因组基因座是纯合子的小鼠。

[0175] 必须注意的是, 除非上下文中另有明示, 在本申请中和所附的权利要求书使用的单数形式的“一个(a)”、“一种(an)”和“所述(the)”包括复数形式。本申请中使用的所有技术和科学术语具有相同的含义。

[0176] 提供本申请所讨论的出版物仅为了其先于本申请的申请日所公开的内容。本申请中没有任何内容可以被解释为承认由于在先发明所述发明没有资格早于此类出版物。此外, 所提供的出版物的日期可能不同于实际出版日期, 可能需要对其进行独立地确认。

[0177] 在不脱离其精神或基本属性的前提下, 可以将所述发明以其他特定的形式体现, 并且应参照所附的权利要求而非前述的说明书表示本发明的范围。

实施例

[0178] 列出下述实施例以便向那些本领域的普通技术人员完整地公开和描述如何实现和使用本发明, 并非旨在限制发明人所认为的其发明的范围, 也不是旨在表示下述实验是所进行的全部或唯一的实验。已做出努力来确保所使用数字(例如数量、温度等)的准确性, 但是应考虑一些实验误差和偏差。除非另有明示, 份数以重量份计, 分子量是重均分子量, 温度是摄氏度, 并且压力是处于或接近大气压。

[0179] 实施例1: 通过锌指核酸酶(ZFN)增强的LTVEC靶向

[0180] 为了检测利用锌指核酸酶(ZFN)在基因中诱导双链断裂是否能够增强LTVEC(基于BAC的大靶向载体)对相同基因的靶向, 使用靶向112rg基因的LTVEC和编码以下述组合中ZFN对的每一半的质粒在F1H4ES细胞中进行三次电穿孔: (1)单独的1.5 μ g 112rgLTVEC; (2) 20 μ g 112rg ZFN-1+20 μ g 112rg LTVEC ZFN-2+1.5 μ g 112rg LTVEC和(3) 20 μ g 112rg ZFN-1+20 μ g 112rg LTVEC ZFN-2无LTVEC。

[0181] 对本申请中使用的ZFN-1和ZFN-2进行设计以含有(1)锌指DNA结合结构域, 所述结构域识别在所述靶序列的各条链中由6bp的切割位点分隔的两个相邻的靶DNA序列和(2)FokI核酸酶, 所述核酸酶二聚化并且在所述靶位点形成双链断裂。更具体地, 将ZFN-1的锌指结构域设计为识别在外显子1的有意义链中的5'-AGCTCCAAGGTCCTC-3'(SEQ ID NO:1); 并且将ZFN-2的锌指结构域设计为识别在112rg基因的外显子1的反义链中的5'-GTCTTCATTCGCACT-3'(SEQ ID NO:2)。

[0182] 将所述LTVEC(VelociGene MAID 5057L1)设计为缺失112rg基因的约3,000个碱基

对(3kb),包括全部IL-2受体 γ 链的编码序列,并且使用表达潮霉素磷酸转移酶的选择性表达盒取代内源性序列,这使其具有潮霉素B的抗性。从市售的BAC文库(BAC ESRelease 2; Incyte Genomics)中分离来源于母体BAC克隆290o15的LTVEC 5057L1,并且其含有约90kb(同源臂1)和约8kb(同源臂2)的两个较大的同源臂。

[0183] 针对电穿孔(1)和(2),分离具有潮霉素B抗性的集落。针对电穿孔(3),允许在不使用药物选择的情况下形成集落。从各次电穿孔中挑取集落并且在96孔板中培养,随后通过等位基因缺失(LOA)检测筛选靶向事件,所述检测设计为检测由LTVEC正确靶向112rg基因形成的3kb缺失。还使用LOA检测检测ZFN在外显子1中诱导的切割事件。

[0184] 表1:针对经靶向的3kb缺失的LOA检测

5057U2试验(内含子2)		
正向引物	5'-GGAGGGTAGCACGGGAAGAAG-3'	SEQ ID NO: 3
反向引物	5'-GCTGGCTACCCACTTGATTGG-3'	SEQ ID NO: 4
TaqMan探针	5'-TCAAGCAGTCTCTCCCAGCTAACCTCCCT-3'	SEQ ID NO: 5
5057D2试验(内含子7)		
正向引物	5'-CAGGATGTGGCTGACCAAATG-3'	SEQ ID NO: 6
反向引物	5'-GGCTCCTAATGCCCTGTAGTTTC-3'	SEQ ID NO: 7
TaqMan探针	5'-CCGTCTCTGCGCTAGCCACCCT-3'	SEQ ID NO: 8

[0186] 表2:针对ZFN切割的LOA检测

正向引物	5'-CAGCTGCTCCTGCTGAGG-3'	SEQ ID NO: 9
反向引物	5'-CCTACCAGCTTTGATGTCTTCATTC-3'	SEQ ID NO: 10
TaqMan探针	5'-AGCTCCAAGGTCCTCATGTCCAGT-3'	SEQ ID NO: 11

[0188] 表3:LTVEC靶向效率比较

	LTVEC	ZFN-1	ZFN-2	检测的克隆数	通过LTVEC正确靶向的克隆	靶向效率(%)	使用ZFN切割的克隆	ZFN切割效率(%)	
[0189]	EP	(μ g)	(μ g)	(μ g)					
	1	1.5	0	0	26	5	19.2	ND ¹	ND ¹
	2	1.5	20	20	102	76	74.5	7 ²	6.9
	3	0	20	20	192	0	0	6	3.1

[0190] ¹未测定

[0191] ²在LTVEC未正确靶向的克隆中

[0192] 如表3所示,当将编码设计为切割靶基因中位点的ZFN的质粒与靶向相同基因的LTVEC联用时(EP#2),与单独的LTVEC(EP#1)相比实现了靶向效率的显著增加(在所描述的实验中增加约4倍)。

[0193] 尽管参照其特定的具体实施方式对所述发明进行了描述,本领域技术人员应理解在不脱离本发明真实精神和范围的前提下可以对其进行多种改变和可以使用等效物替代。此外,可以采用特定的情况、材料、物质组成、工艺、工艺步骤或步骤对所述发明的客观精神和范围进行多种修订。所有的这种修订均旨在包括在所附权利要求的范围内。

序列表

<110> 瑞恩泽制药公司

D·弗伦德维

W·奥尔巴克

D·M·瓦伦泽拉

G·D·扬科普洛斯

K·V·莱

<120> 核酸酶介导的使用大靶向载体的靶向

<130> 57766-443601

<150> 61/638,267

<151> 2012-04-25

<160> 11

<170> PatentIn 3.5版

<210> 1

<211> 15

<212> DNA

<213> 人工的

<220>

<223> 合成的

<400> 1

agctccaagg tcctc 15

<210> 2

<211> 15

<212> DNA

<213> 人工的

<220>

<223> 合成的

<400> 2

gtcttcattc gcaact 15

<210> 3

<211> 21

<212> DNA

<213> 人工的

<220>

<223> 合成的

<400> 3

ggagggtagc acgggaagaa g 21

<210> 4

<211> 21
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 4
gctggctacc cacttgattg g 21
<210> 5
<211> 29
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 5
tcaagcagtc tctcccagct aacctcct 29
<210> 6
<211> 21
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 6
caggatgtgg ctgaccaaat g 21
<210> 7
<211> 23
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 7
ggctcctaata gccctgtagt ttc 23
<210> 8
<211> 24
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 8
ccgtctctct gcctagccca ccct 24

<210> 9
<211> 18
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 9
cagctgctcc tgctgagg 18
<210> 10
<211> 25
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 10
cctaccagct ttgatgtctt cattc 25
<210> 11
<211> 24
<212> DNA
<213> 人工的
<220>
<223> 合成的
<400> 11
agctccaagg tcctcatgtc cagt 24

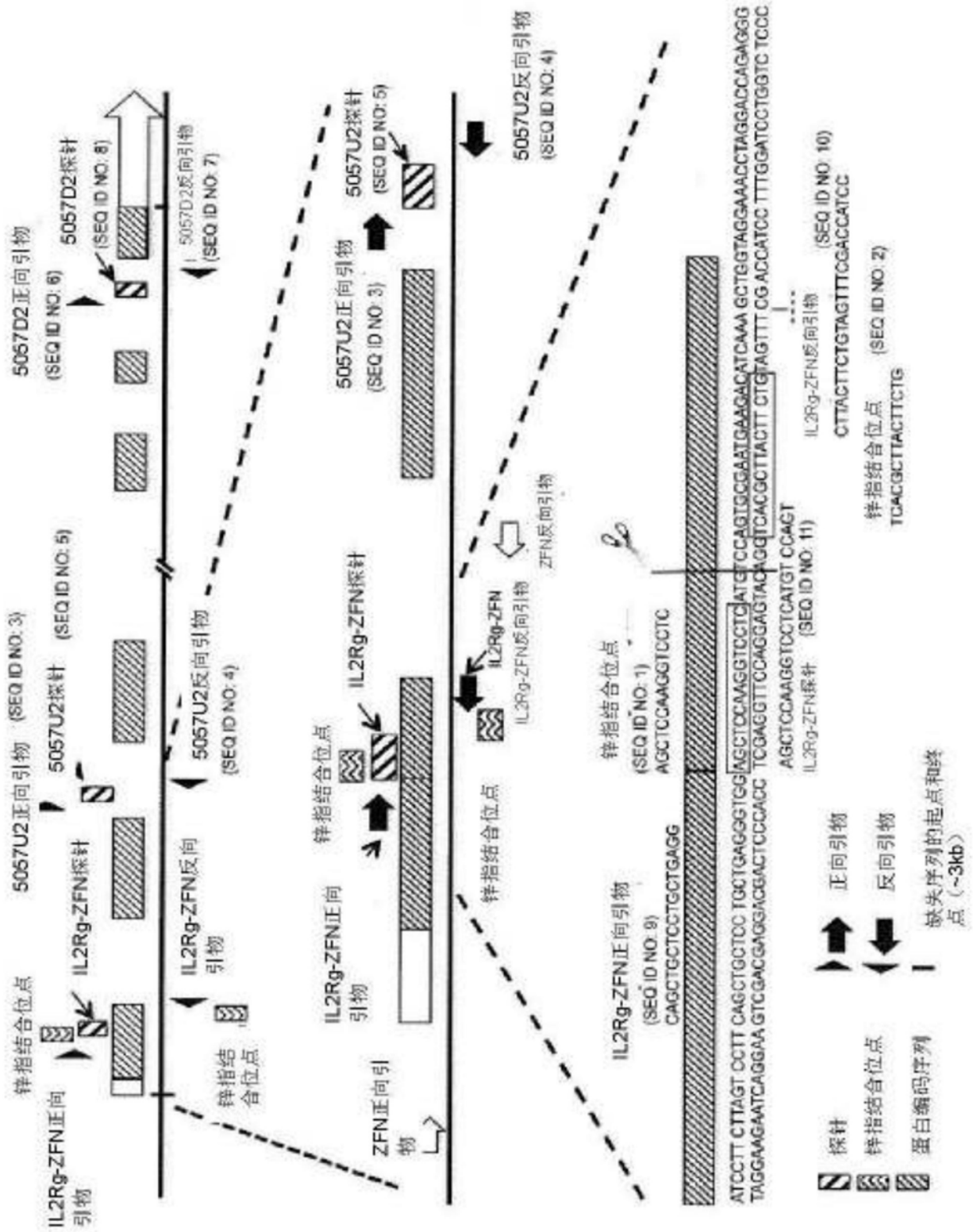


图1