



(12)发明专利申请

(10)申请公布号 CN 108715602 A

(43)申请公布日 2018. 10. 30

(21)申请号 201810540449.5

(22)申请日 2013.12.05

(30)优先权数据

61/734256 2012.12.06 US

61/758624 2013.01.30 US

61/761046 2013.02.05 US

61/794422 2013.03.15 US

(62)分案原申请数据

201380072477.4 2013.12.05

(71)申请人 西格马-奥尔德里奇有限责任公司

地址 美国密苏里州

(72)发明人 F.陈 G.D.戴维斯

(74)专利代理机构 中国专利代理(香港)有限公司

司 72001

代理人 黄登高 黄希贵

(51)Int.Cl.

C07K 7/06(2006.01)

C07K 14/46(2006.01)

C12N 7/00(2006.01)

C12N 9/22(2006.01)

C12N 9/96(2006.01)

C12N 15/10(2006.01)

C12N 15/11(2006.01)

C12N 15/63(2006.01)

C12N 15/67(2006.01)

C12N 15/85(2006.01)

C12N 15/86(2006.01)

C12N 15/90(2006.01)

权利要求书3页 说明书33页

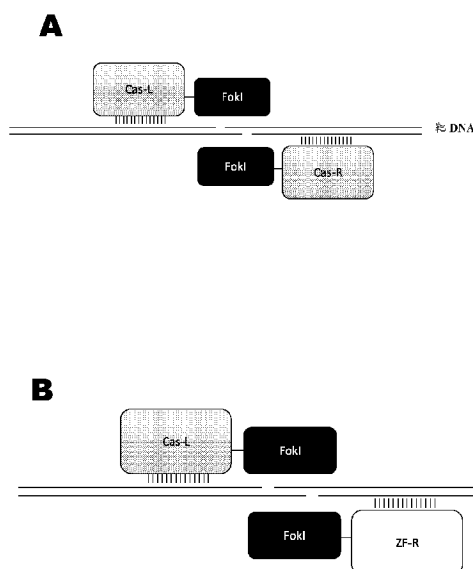
序列表17页 附图7页

(54)发明名称

基于CRISPR的基因组修饰和调控

(57)摘要

本发明提供经工程改造用于在真核细胞或胚胎中表达的RNA指导的核酸内切酶,和在真核细胞或胚胎中将RNA指导的核酸内切酶用于靶基因组修饰的方法。也提供融合蛋白,其中各融合蛋白包含CRISPR/Cas样蛋白或其片段和效应子结构域。效应子结构域可以是剪切结构域、表观遗传修饰结构域、转录激活结构域或转录抑制子结构域。也提供使用融合蛋白修饰染色体序列或调控染色体序列表达的方法。



1. 用于修饰真核细胞中的染色体序列的方法,所述方法包括:

a) 向真核细胞引入两种RNA指导的切口酶系统或编码所述系统的核酸,和任选供体多核苷酸,其中各RNA指导的切口酶系统包含

(i) 经修饰以剪切双链序列的一条链的RNA指导的核酸内切酶;和

(ii) 指导RNA,其包含与染色体序列中的靶位点具有互补性的第一区域和与RNA指导的核酸内切酶相互作用的第二区域,

其中各靶位点紧接着原间隔毗连基序(PAM),并且两种RNA指导的核酸内切酶的靶位点在染色体序列的相对链上;和

b) 培养所述真核细胞使得两种RNA指导的核酸内切酶足够靠近剪切染色体序列的相对链,以在染色体序列中引入双链断裂,并且所述双链断裂通过DNA修复过程修复使得所述染色体序列被修饰,其中

所述方法不包括用于修饰人类的种系遗传身份的过程和,其中

所述方法不包括用于治疗人或动物身体的方法。

2. 用于修饰真核细胞中的染色体序列的离体或体外方法,所述方法包括:

a) 向真核细胞引入两种RNA指导的切口酶系统或编码所述系统的核酸,和任选供体多核苷酸,其中各RNA指导的切口酶系统包含

(i) 经修饰以剪切双链序列的一条链的RNA指导的核酸内切酶;和

(ii) 指导RNA,其包含与染色体序列中的靶位点具有互补性的第一区域和与RNA指导的核酸内切酶相互作用的第二区域,

其中各靶位点紧接着原间隔毗连基序(PAM),并且两种RNA指导的核酸内切酶的靶位点在染色体序列的相对链上;和

b) 培养所述真核细胞使得两种RNA指导的核酸内切酶足够靠近剪切染色体序列的相对链,以在染色体序列中引入双链断裂,并且所述双链断裂通过DNA修复过程修复使得所述染色体序列被修饰,其中

所述方法不包括用于修饰人类的种系遗传身份的过程。

3. 权利要求1或2的方法,其中各RNA指导的核酸内切酶衍生自成簇的规律间隔的短回文重复(CRISPR)/CRISPR相关的(Cas) (CRISPR/Cas) 类型II系统蛋白。

4. 权利要求3的方法,其中所述CRISPR/Cas类型II系统蛋白是Cas9蛋白。

5. 权利要求4的方法,其中所述Cas9蛋白包含RuvC或HNH结构域中的突变。

6. 权利要求1-4中任一项的方法,其中各RNA指导的核酸内切酶进一步包含至少一个选自核定位信号、细胞穿透结构域或标志物结构域的另外的结构域。

7. 权利要求1-5中任一项的方法,其中所述靶位点是Rosa26基因座、HPRT基因座或AAVS1基因座。

8. 权利要求1-7中任一项的方法,其中所述供体多核苷酸包含供体序列,所述供体序列相对于染色体序列中的靶位点附近的染色体序列具有至少一个核苷酸改变。

9. 权利要求8的方法,其中所述供体序列侧接与位于染色体序列中靶位点上游和下游的序列具有基本序列同一性的序列。

10. 权利要求8或9的方法,其中所述供体序列侧接与由RNA指导的核酸内切酶生成的突出端相容的短突出端。

11. 权利要求1-10中任一项的方法, 其中编码各RNA指导的核酸内切酶的核酸是mRNA, 和编码各指导RNA的核酸是DNA。

12. 权利要求1-10中任一项的方法, 其中编码各RNA指导的核酸内切酶的核酸是DNA, 和编码各指导RNA的核酸是DNA。

13. 权利要求12的方法, 其中所述DNA是载体的一部分, 所述载体进一步包含与编码RNA指导的核酸内切酶的核酸可操作连接的启动子控制序列和与编码指导RNA的核酸可操作连接的启动子控制序列。

14. 权利要求1-13中任一项的方法, 其中所述真核细胞是人细胞、非人哺乳动物细胞或非人哺乳动物胚胎。

15. 权利要求1-13中任一项的方法, 其中所述真核细胞是无脊椎动物细胞、昆虫细胞、植物细胞、酵母细胞或单细胞真核生物。

16. 权利要求15的方法, 其中所述真核细胞是植物细胞。

17. 权利要求14的方法, 其中所述人细胞或非人哺乳动物细胞是T细胞。

18. 权利要求1-17中任一项的方法, 其中所述真核细胞在体外。

19. 权利要求1的方法, 其中所述真核细胞在体内。

20. 权利要求1-19中任一项的方法, 其中所述供体多核苷酸不被引入所述真核细胞中, 并且通过非同源末端连接修复过程修复所述双链断裂导致所述染色体序列失活。

21. 权利要求1-19中任一项的方法, 其中将所述供体多核苷酸引入所述真核细胞中, 并且所述双链断裂的修复导致所述染色体序列中至少一个核苷酸的改变。

22. 权利要求21的方法, 其中所述改变包括外源序列的整合。

23. 权利要求1-10或14-22中任一项的方法, 其中所述至少一种指导RNA是化学合成的。

24. 包含至少两种RNA指导的切口酶系统的组合物, 各RNA指导的切口酶系统包含

(i) 经修饰以剪切双链序列的一条链的RNA指导的核酸内切酶; 和

(ii) 指导RNA, 其包含与染色体序列中的靶位点具有互补性的第一区域和与RNA指导的核酸内切酶相互作用的第二区域, 其中各靶位点紧接着原间隔毗连基序(PAM), 并且两种RNA指导的核酸内切酶的靶位点在染色体序列的相对链上且足够靠近以在染色体序列中引入双链断裂。

25. 权利要求24的组合物, 其中各RNA指导的核酸内切酶衍生自成簇的规律间隔的短回文重复(CRISPR)/CRISPR相关的(Cas) (CRISPR/Cas) 类型II系统蛋白。

26. 权利要求25的组合物, 其中所述CRISPR/Cas类型II系统蛋白是Cas9蛋白。

27. 权利要求26的组合物, 其中所述Cas9蛋白包含RuvC或HNH结构域中的突变。

28. 权利要求24-27中任一项的组合物, 其中各RNA指导的核酸内切酶进一步包含至少一个选自核定位信号、细胞穿透结构域或标志物结构域的另外的结构域。

29. 权利要求24-28中任一项的组合物, 其中所述靶位点是Rosa26基因座、HPRT基因座或AAVS1基因座。

30. 权利要求24-29中任一项的组合物, 其中所述至少一种指导RNA是化学合成的。

31. 编码权利要求24-29中任一项的组合物核酸的核酸。

32. 权利要求31的核酸, 其中编码各RNA指导的核酸内切酶的核酸是mRNA, 和编码各指

导RNA的核酸是DNA。

33. 权利要求31的核酸,其中编码各RNA指导的核酸内切酶的核酸是DNA,和编码各指导RNA的核酸是DNA。

34. 包含权利要求33的核酸的载体,其中编码各RNA指导的核酸内切酶的核酸与启动子控制序列可操作连接用于在真核细胞中表达,并且编码各指导RNA的核酸与启动子控制序列可操作连接用于在真核细胞中表达。

35. 包含至少两种RNA指导的切口酶系统或编码所述系统的核酸的试剂盒,各RNA指导的切口酶系统包含 (i) 经修饰以剪切双链序列的一条链的RNA指导的核酸内切酶和 (ii) 指导RNA,其包含与染色体序列中的靶位点具有互补性的第一区域和与RNA指导的核酸内切酶相互作用的第二区域,其中各靶位点紧接着原间隔毗连基序 (PAM),并且两种RNA指导的核酸内切酶的靶位点在染色体序列的相对链上且足够靠近以在染色体序列中引入双链断裂。

36. 权利要求35的试剂盒,进一步包含至少一种供体多核苷酸。

37. 权利要求35或36的试剂盒,其中各RNA指导的核酸内切酶衍生自成簇的规律间隔的短回文重复 (CRISPR) /CRISPR相关的 (Cas) (CRISPR/Cas) 类型II系统蛋白。

38. 权利要求37的试剂盒,其中所述CRISPR/Cas类型II系统蛋白是Cas9蛋白。

39. 权利要求38的试剂盒,其中所述Cas9蛋白包含RuvC或HNH结构域中的突变。

40. 权利要求35-39中任一项的试剂盒,其中各RNA指导的核酸内切酶进一步包含至少一个选自核定位信号、细胞穿透结构域或标志物结构域的另外的结构域。

41. 权利要求35-40中任一项的试剂盒,其中编码各RNA指导的核酸内切酶的核酸是mRNA,和编码各指导RNA的核酸是DNA。

42. 权利要求35-40中任一项的试剂盒,其中编码各RNA指导的核酸内切酶的核酸是DNA,和编码各指导RNA的核酸是DNA。

43. 权利要求42的试剂盒,其中编码各RNA指导的核酸内切酶的核酸与启动子控制序列可操作连接用于在真核细胞中表达,并且编码各指导RNA的核酸与启动子控制序列可操作连接用于在真核细胞中表达。

44. 权利要求36的试剂盒,其中所述供体多核苷酸包含供体序列。

45. 权利要求44的试剂盒,其中所述供体序列侧接与染色体序列中靶位点任一侧的序列具有基本序列同一性的序列。

46. 权利要求44的试剂盒,其中所述供体序列侧接与由RNA指导的核酸内切酶生成的突出端相容的短突出端。

47. 权利要求35-40或44-46中任一项的试剂盒,其中所述至少一种指导RNA是化学合成的。

基于CRISPR的基因组修饰和调控

[0001] 本申请是申请日为2013年12月5日,申请号为201380072477.4,发明名称为“基于CRISPR的基因组修饰和调控”的发明专利的分案申请。

技术领域

[0002] 本公开内容涉及靶基因组修饰。具体而言,本公开涉及RNA指导的核酸内切酶或包含CRISPR/Cas样蛋白的融合蛋白和使用所述蛋白修饰或调控靶染色体序列的方法。

[0003] 发明背景

靶基因组修饰是用于真核细胞、胚胎和动物的基因操作的有力的工具。例如,外源序列可以被整合在靶基因组位置和/或特定的内源染色体序列可以被缺失、失活或修饰。当前的方法依赖于工程化核酸酶的使用,工程化核酸酶例如,锌指核酸酶(ZFNs)或转录激活子样效应子核酸酶(TALENs)。这些嵌合核酸酶含有与非特异性DNA剪切结构域连接的可编程的、序列特异性的DNA结合组件。但是,各个新的基因组靶需要设计包含新的序列特异性的DNA-结合组件的新ZFN或TALEN。因此,这些用户定制设计的核酸酶往往制备昂贵且耗时。此外,ZFNs和TALENs的特异性使得它们能介导脱靶剪切。

[0004] 因此,需要不要求针对每个新的靶基因组位置设计新的核酸酶的靶基因组修饰技术。另外,需要特异性增加的具有很少或没有脱靶作用的技术。

[0005] 发明概述

在本公开内容的不同方面之中提供分离的RNA指导的核酸内切酶,其中所述核酸内切酶包含至少一个核定位信号、至少一个核酸酶结构域和至少一个与指导RNA相互作用以将核酸内切酶靶向用于剪切的特定核苷酸序列的结构域。在一个实施方案中,所述核酸内切酶可衍生自Cas9蛋白。在另一实施方案中,所述核酸内切酶可经修饰而缺失至少一个功能性核酸酶结构域。在其它实施方案中,所述核酸内切酶可进一步包含细胞穿透结构域、标志物结构域,或二者。在进一步的实施方案中,所述核酸内切酶可以是包含指导RNA的蛋白-RNA复合物的一部分。在一些情况下,所述指导RNA可以是包含与靶位点互补的5'区域的单分子。也提供编码任一本公开内容的RNA指导的核酸内切酶的分离的核酸。在一些实施方案中,所述核酸可以是针对在哺乳动物细胞(例如,人细胞)中翻译而最优化的密码子。在其它实施方案中,编码RNA指导的核酸内切酶的核酸序列可以与启动子控制序列可操作性地连接,并任选地可以是载体的一部分。在其它实施方案中,包含可以与启动子控制序列可操作性地连接的编码RNA指导的核酸内切酶的序列的载体也可包含可以与启动子控制序列可操作性地连接的编码指导RNA的序列。

[0006] 本发明的另一方面包括用于在真核细胞或胚胎中修饰染色体序列的方法。该方法包括向真核细胞或胚胎中引入(i)至少一种包含至少一个核定位信号的RNA指导的核酸内切酶或编码至少一种如本文所定义的RNA指导的核酸内切酶的核酸,(ii)至少一种指导RNA或编码至少一种指导RNA的DNA,和任选(iii)至少一种包含供体序列的供体多核苷酸。该方法进一步包括培养所述细胞或胚胎以便各指导RNA将RNA指导的核酸内切酶定向至染色体序列中的靶位点,其中所述RNA指导的核酸内切酶将双链断裂引入所述靶位点,并且所述双

链断裂通过DNA修复过程而修复以便修饰染色体序列。在一个实施方案中,所述RNA指导的核酸内切酶可以衍生自Cas9蛋白。在另一实施方案中,编码引入细胞或胚胎中的RNA指导的核酸内切酶的核酸可以是mRNA。在进一步的实施方案中,其中编码引入细胞或胚胎的RNA指导的核酸内切酶的核酸可以是DNA。在进一步的实施方案中,编码RNA指导的核酸内切酶的DNA可以是另外包含编码指导RNA的序列的载体的一部分。在某些实施方案中,所述真核细胞可以是人细胞、非人哺乳动物细胞、干细胞、非哺乳动物的脊椎动物细胞、无脊椎动物细胞、植物细胞或单细胞真核生物。在某些其它实施方案中,所述胚胎是非人单细胞动物胚胎。

[0007] 本公开内容的另外方面提供包含CRISPR/Cas样蛋白或其片段和效应子结构域的融合蛋白。一般而言,所述融合蛋白包含至少一个核定位信号。融合蛋白的效应子结构域可以是剪切结构域、表观遗传修饰结构域、转录激活结构域或转录抑制子结构域。在一个实施方案中,融合蛋白的CRISPR/Cas样蛋白可以衍生自Cas9蛋白。在一个重复中,所述Cas9蛋白可经修饰而缺失至少一个功能性核酸酶结构域。在备选的重复中,所述Cas9蛋白可以被修饰而缺失全部核酸酶活性。在一个实施方案中,所述效应子结构域可以是剪切结构域,例如,FokI核酸内切酶结构域或修饰的FokI核酸内切酶结构域。在另一实施方案中,一种融合蛋白可与另一融合蛋白形成二聚体。所述二聚体可以是同源二聚体或异源二聚体。在另一实施方案中,所述融合蛋白可与锌指核酸酶形成异源二聚体,其中融合蛋白和锌指核酸酶二者的剪切结构域是FokI核酸内切酶结构域或修饰的FokI核酸内切酶结构域。在仍另一实施方案中,所述融合蛋白包含衍生自经修饰而缺失全部核酸酶活性的Cas9蛋白的CRISPR/Cas样蛋白,并且所述效应子结构域是FokI核酸内切酶结构域或修饰的FokI核酸内切酶结构域。在仍另一实施方案中,所述融合蛋白包含衍生自经修饰而缺失全部核酸酶活性的Cas9蛋白的CRISPR/Cas样蛋白,并且所述效应子结构域可以是表观遗传修饰结构域、转录激活结构域或转录抑制子结构域。在另外的实施方案中,本文公开的融合蛋白中任一种可包含至少一个选自核定位信号、细胞穿透结构域和标志物结构域的另外的结构域。也提供编码本文提供的融合蛋白中任一种的分离的核酸。

[0008] 本公开内容的仍另一方面包括用于在细胞或胚胎中修饰染色体序列或调控染色体序列的表达的方法。所述方法包括向细胞或胚胎中引入:(a)至少一种融合蛋白或编码至少一种融合蛋白的核酸,其中所述融合蛋白包含CRISPR/Cas样蛋白或其片段和效应子结构域,和(b)至少一种指导RNA或编码至少一种指导RNA的DNA,其中所述指导RNA将融合蛋白的CRISPR/Cas样蛋白指导至染色体序列中的靶位点,并且融合蛋白的效应子结构域修饰染色体序列或调控染色体序列的表达。在一个实施方案中,所述融合蛋白的CRISPR/Cas样蛋白可以衍生自Cas9蛋白。在另一实施方案中,所述融合蛋白的CRISPR/Cas样蛋白可经修饰而缺失至少一个功能性核酸酶结构域。在仍另一实施方案中,所述融合蛋白的CRISPR/Cas样蛋白可经修饰而缺失全部核酸酶活性。在其中融合蛋白包含经修饰而缺失全部核酸酶活性的Cas9蛋白和FokI剪切结构域或修饰的FokI剪切结构域的一个实施方案中,该方法可包括向所述细胞或胚胎引入一种融合蛋白或编码一种融合蛋白的核酸和两种指导RNA或编码两种指导RNA的DNA,并且其中一个双链断裂被引入染色体序列。在其中所述融合蛋白包含经修饰而缺失全部核酸酶活性的Cas9蛋白和FokI剪切结构域或修饰的FokI剪切结构域的另一实施方案中,所述方法可包括向所述细胞或胚胎中引入两种融合蛋白或编码两种融合蛋

白的核酸和两种指导RNA或编码两种指导RNA的DNA,并且其中两个双链断裂被引入所述染色体序列。在其中所述融合蛋白包含经修饰而缺失全部核酸酶活性的Cas9蛋白和FokI剪切结构域或修饰的FokI剪切结构域的仍另一实施方案中,所述方法可包括向细胞或胚胎中引入一种融合蛋白或编码一种融合蛋白的核酸、一种指导RNA或编码一种指导RNA的核酸和一种锌指核酸酶或编码一种锌指核酸酶的核酸,其中所述锌指核酸酶包含FokI剪切结构域或修饰的FokI剪切结构域,并且其中一个双链断裂被引入所述染色体序列。在其中所述融合蛋白包含剪切结构域的某些实施方案中,所述方法可进一步包含向细胞或胚胎引入至少一种供体多核苷酸。在其中所述融合蛋白包含选自表观遗传修饰结构域、转录激活结构域或转录抑制子结构域的效应子结构域的实施方案中,所述融合蛋白可包含经修饰而缺失全部核酸酶活性的Cas9蛋白,且所述方法可以包括向细胞或胚胎中引入一种融合蛋白或编码一种融合蛋白的核酸和一种指导RNA或编码一种指导RNA的核酸,并且其中修饰所述靶染色体序列的结构或表达。在某些实施方案中,所述真核细胞可以是人细胞、非人哺乳动物细胞、干细胞、非哺乳动物的脊椎动物细胞、无脊椎动物细胞、植物细胞或单细胞真核生物。在某些其它实施方案中,所述胚胎是非人单细胞动物胚胎。

[0009] 本公开内容的其它方面和重复在下面详述。

[0010] 附图简述

图1图示使用蛋白二聚体的基因组修饰。(A)描绘了由两种融合蛋白组成的二聚体所产生的双链断裂,所述两种融合蛋白的每一种包含用于DNA结合的Cas样蛋白和FokI剪切结构域。(B)描绘了由二聚体所产生的双链断裂,所述二聚体由包含Cas样蛋白和FokI剪切结构域的融合蛋白以及包含锌指(ZF)DNA-结合结构域和FokI剪切结构域的锌指核酸酶组成。

[0011] 图2表明使用包含基因调控结构域的RNA指导的融合蛋白调控基因表达。(A)描绘了包含用于DNA结合的Cas样蛋白和激活或抑制基因表达的“A/R”结构域的融合蛋白。(B)图示包含用于DNA结合的Cas样蛋白和通过邻近的DNA或蛋白的共价修饰而影响表观遗传状态的表观遗传修饰结构域(“Epi-mod”)的融合蛋白。

[0012] 图3图示使用两种RNA指导的核酸内切酶的基因组修饰。

[0013] (A)描绘了由已经转化为切口酶的两RNA指导的核酸内切酶产生的双链断裂。(B)描绘了由具有核酸内切酶活性的两RNA指导的核酸内切酶产生的两个双链断裂。

[0014] 图4显示了用Cas9核酸、Cas9指导的RNA和AAVS1-GFP DNA供体转染的人K562细胞的荧光激活细胞分选术(FACS)。Y轴表示在红色通道的自发荧光强度,和X轴表示绿色荧光强度,(A)用以下转染的K562细胞:10 μ g的用Anti-Reverse帽类似物转录的Cas9mRNA、0.3nmol的预退火的crRNA-tracrRNA双链体和10 μ g的AAVS1-GFP质粒DNA;(B)用以下转染的K562细胞:10 μ g的用Anti-Reverse帽类似物转录的Cas9mRNA、0.3nmol的嵌合RNA和10 μ g的AAVS1-GFP质粒DNA;(C)用以下转染的K562细胞:10 μ g的通过转录后加帽反应加帽的Cas9mRNA、0.3nmol的嵌合RNA和10 μ g的AAVS1-GFP质粒DNA;(D)用10 μ g的Cas9质粒DNA、5 μ g的U6-嵌合RNA质粒DNA和10 μ g的AAVS1-GFP质粒DNA转染的K562细胞;(E)用10 μ g的AAVS1-GFP质粒DNA转染的K562细胞;(F)仅用转染试剂转染的K562细胞。

[0015] 图5显示证明GFP靶向整合至人细胞的AAVS1基因座的连接PCR分析。泳道M:1kb DNA分子标记;泳道A:用以下转染的K562细胞:用Anti-Reverse帽类似物转录的10 μ g的Cas9mRNA、0.3nmol的预退火的crRNA-tracrRNA双链体和10 μ g的AAVS1-GFP质粒DNA;泳道B:

用以下转染的K562细胞：用Anti-Reverse帽类似物转录的10 μ g的Cas9mRNA、0.3nmol的嵌合RNA和10 μ g的AAVS1-GFP质粒DNA；泳道C：用通过转录后加帽反应加帽的10 μ g的Cas9mRNA、0.3nmol的嵌合RNA和10 μ g的AAVS1-GFP质粒DNA转染的K562细胞；泳道D：用10 μ g的Cas9质粒DNA、5 μ g的U6-嵌合RNA质粒DNA和10 μ g的AAVS1-GFP质粒DNA转染的K562细胞；泳道E：用10 μ g的AAVS1-GFP质粒DNA转染的K562细胞；泳道F：仅用转染试剂转染的K562细胞。

[0016] 发明详述

本文提供RNA指导的核酸内切酶，其包含至少一个核定位信号、至少一个核酸酶结构域和至少一个与指导RNA相互作用以将所述核酸内切酶靶向用于剪切的特异性核苷酸序列的结构域。也提供了编码RNA指导的核酸内切酶的核酸，和使用RNA指导的核酸内切酶修饰真核细胞或胚胎的染色体序列的方法。所述RNA指导的核酸内切酶与特异性的指导RNA相互作用，所述特异性的指导RNA的每一种将所述核酸内切酶定向至特定靶位点，在所述位点该RNA指导的核酸内切酶引起双链断裂，其可通过DNA修复过程而修复使得修复染色体序列。因为特异性通过所述指导RNA提供，因此基于RNA的核酸内切酶是通用的，并且可以与不同的指导RNA一起使用以靶向不同基因组序列。本文公开的方法可用于靶向和修饰特定的染色体序列和/或将外源序列引入到细胞或胚胎的基因组中的靶位置。此外，靶向是特异性的，具有有限的脱靶效果。

[0017] 本公开内容提供了融合蛋白，其中融合蛋白包含CRISPR/Cas样蛋白或其片段和效应子结构域。合适的效应子结构域包括但不限于剪切结构域、表观遗传修饰结构域、转录激活结构域和转录抑制子结构域。通过特异性指导RNA将各融合蛋白指导至具体的染色体序列，其中所述效应子结构域介导靶基因组修饰或基因调控。在一个方面，所述融合蛋白可用作二聚体，从而增加靶位点的长度和增加其在基因组中的唯一性的可能性（从而，降低脱靶效果）。例如，内源的CRISPR系统基于大约13-20bp的DNA结合字长度而修饰基因组位置（Cong等人，*Science*, 339:819-823）。在此字长，仅5-7%的靶位点在基因组内是唯一的（Iseli等人，*PLoS One* 2(6):e579）。相比之下，锌指核酸酶的DNA结合字长范围通常为30-36bp，这产生大约85-87%在人基因组内唯一的靶位点。基于CRISPR的系统所利用的较小大小的DNA结合位点限制了在所需位置附近的基于靶向CRISPR的核酸酶的设计并使其复杂化，所述所需位置例如疾病SNPs、小外显子、起始密码子和终止密码子，以及在复杂的基因组中的其它位置。本公开内容不仅提供了用于扩展CRISPR DNA结合字长的方式（即，以便限制脱靶活性），而且进一步提供具有修饰功能性的CRISPR融合蛋白。因此，所公开的CRISPR融合蛋白具有增加的靶特异性和唯一的功能性。本文也提供使用融合蛋白以修饰或调控靶染色体序列的表达的方法。

[0018] (I) RNA指导的核酸内切酶

本公开内容的一个方面提供包含至少一个核定位信号的RNA指导的核酸内切酶，所述核定位信号允许核酸内切酶进入真核细胞和胚胎（例如，非人单细胞胚胎）的胞核。RNA指导的核酸内切酶也包含至少一个核酸酶结构域和至少一个与指导RNA相互作用的结构域。RNA指导的核酸内切酶被指导RNA定向至特定的核酸序列（或靶位点）。所述指导RNA与RNA指导的核酸内切酶和靶位点相互作用使得一旦定向至靶位点，RNA指导的核酸内切酶就能够向靶位点核酸序列中引入双链断裂。因为所述指导RNA提供针对靶向剪切的特异性，RNA指导的核酸内切酶的核酸内切酶是通用的并且可以与不同指导RNA一起使用以剪切不同靶核酸

序列。本文提供了分离的RNA指导的核酸内切酶、编码RNA指导的核酸内切酶的分离的核酸(即,RNA或DNA)、包含编码RNA指导的核酸内切酶的核酸的载体、以及包含RNA指导的核酸内切酶加上指导RNA的蛋白-RNA复合物。

[0019] RNA指导的核酸内切酶可衍生自成簇的规律间隔的短回文重复(CRISPR)/CRISPR相关的(Cas)系统。CRISPR/Cas系统可以是类型I、类型II或类型III系统。合适的CRISPR/Cas蛋白的非限制性实例包括Cas3、Cas4、Cas5、Cas5e(或CasD)、Cas6、Cas6e、Cas6f、Cas7、Cas8a1、Cas8a2、Cas8b、Cas8c、Cas9、Cas10、Cas10d、CasF、CasG、CasH、Csy1、Csy2、Csy3、Cse1(或CasA)、Cse2(或CasB)、Cse3(或CasE)、Cse4(或CasC)、Csc1、Csc2、Csa5、Csn2、Csm2、Csm3、Csm4、Csm5、Csm6、Cmr1、Cmr3、Cmr4、Cmr5、Cmr6、Csb1、Csb2、Csb3、Csx17、Csx14、Csx10、Csx16、CsaX、Csx3、Csz1、Csx15、Csf1、Csf2、Csf3、Csf4和Cu1966。

[0020] 在一个实施方案中, RNA指导的核酸内切酶衍生自类型II CRISPR/Cas系统。在特定实施方案中, RNA指导的核酸内切酶衍生自Cas9蛋白。Cas9蛋白可来自酿脓链球菌(*Streptococcus pyogenes*)、嗜热链球菌(*Streptococcus thermophilus*)、链球菌(*Streptococcus sp.*)、达松维尔拟诺卡氏菌(*Nocardiopsis dassonvillei*)、始旋链霉菌(*Streptomyces pristinaespiralis*)、绿色产色链霉菌(*Streptomyces viridochromogenes*)、绿色产色链霉菌、玫瑰链孢囊菌(*Streptosporangium roseum*)、玫瑰链孢囊菌、酸热脂环酸杆菌(*Alicyclobacillus acidocaldarius*)、假草状芽孢杆菌(*Bacillus pseudomycoides*)、*Bacillus selenitireducens*、*Exiguobacterium sibiricum*、德氏乳杆菌(*Lactobacillus delbrueckii*)、唾液乳杆菌(*Lactobacillus salivarius*)、海洋微颤蓝细菌(*Microscilla marina*)、*Burkholderiales bacterium*、*Polaromonas naphthalenivorans*、*Polaromonas sp.*、*Crocospaera watsonii*、蓝杆藻(*Cyanothece sp.*)、铜绿微囊蓝细菌(*Microcystis aeruginosa*)、聚球菌(*Synechococcus sp.*)、阿拉伯糖醋盐杆菌(*Acetohalobium arabaticum*)、*Ammonifex degensii*、*Caldicelulosiruptor beccarii*、*Candidatus Desulforudis*、肉毒梭菌(*Clostridium botulinum*)、艰难梭菌(*Clostridium difficile*)、大芬戈尔德菌(*Fingoldia magna*)、嗜热盐碱厌氧菌(*Natranaerobius thermophilus*)、丙酸互营细菌(*Pelotomaculum thermopropionicum*)、*Acidithiobacillus caldus*、嗜酸氧化亚铁硫杆菌(*Acidithiobacillus ferrooxidans*)、紫色硫细菌(*Allochromatium vinosum*)、海杆菌(*Marinobacter sp.*)、嗜盐硝化球菌(*Nitrosococcus halophilus*)、*Nitrosococcus watsoni*、*Pseudoalteromonas haloplanktis*、*Ktedonobacter racemifer*、*Methanohalobium evestigatum*、多鱼腥蓝细菌(*Anabaena variabilis*)、产泡沫节球蓝细菌(*Nodularia spumigena*)、*Nostoc sp.*、最大节螺蓝细菌(*Arthrospira maxima*)、*Arthrospiraplatensis*、节螺蓝细菌属(*Arthrospira sp.*)、鞘丝蓝细菌属(*Lyngbya sp.*)、原型体微鞘蓝细菌(*Microcoleus chthonoplastes*)、颤蓝细菌属(*Oscillatoria sp.*)、*Petrogloa mobilis*、非洲栖热腔菌(*Thermosiphon africanus*)或*Acaryochloris marina*。

[0021] 一般而言, CRISPR/Cas蛋白包含至少一个RNA识别和/或RNA结合结构域。RNA识别和/或RNA结合结构域与指导RNA相互作用。CRISPR/Cas蛋白也可包含核酸酶结构域(即, DNA酶或RNA酶结构域)、DNA结合结构域、解螺旋酶结构域、RNA酶结构域、蛋白-蛋白相互作用结构域、二聚结构域和其它结构域。

[0022] CRISPR/Cas样蛋白可以是野生型CRISPR/Cas蛋白、修饰的CRISPR/Cas蛋白或野生型或修饰的CRISPR/Cas蛋白的片段。CRISPR/Cas样蛋白可以经修饰以增加核酸结合亲和力和/或特异性,改变酶活性和/或改变蛋白的另一性质。例如,CRISPR/Cas样蛋白的核酸酶(即,DNA酶、RNA酶)结构域可以被修饰、缺失或失活。备选地,CRISPR/Cas样蛋白可以被截短以除去对于融合蛋白的功能并非必要的结构域。CRISPR/Cas样蛋白也可被截短或修饰以使融合蛋白的效应子结构域的活性最优化。

[0023] 在一些实施方案中,CRISPR/Cas样蛋白可以衍生自野生型Cas9蛋白或其片段。在其它实施方案中,CRISPR/Cas样蛋白可以衍生自修饰的Cas9蛋白。例如,Cas9蛋白的氨基酸序列可以被修饰而改变蛋白的一种或多种性质(例如,核酸酶活性、亲和力、稳定性等)。备选地,不涉及RNA指导的剪切的Cas9蛋白的结构域可以从蛋白中排除以便修饰的Cas9蛋白与野生型Cas9蛋白相比更小。

[0024] 一般而言,Cas9蛋白包含至少两个核酸酶(即,DNA酶)结构域。例如,Cas9蛋白可以包含RuvC样核酸酶结构域和HNH样核酸酶结构域。RuvC和HNH结构域一起起作用以切开单链从而在DNA中产生双链断裂。(Jinek等人,Science,337:816-821)。在一些实施方案中,Cas9衍生的蛋白可以被修饰而仅含有一个功能性核酸酶结构域(RuvC样或HNH样核酸酶结构域)。例如,Cas9衍生的蛋白可以被修饰使得核酸酶结构域的一个被缺失或突变使得它不再具有功能(即,缺少核酸酶活性)。在其中核酸酶结构域之一无活性的一些实施方案中,Cas9衍生的蛋白能够将切口引入双链核酸中(这样的蛋白称为“切口酶”),但并不剪切双链DNA。例如,在RuvC样结构域中天冬氨酸至丙氨酸(D10A)的转换将Cas9衍生的蛋白转化成切口酶。同样,在HNH结构域中组氨酸至丙氨酸(H840A或H839A)的转换将Cas9衍生的蛋白转化成切口酶。各核酸酶结构域可以使用众所周知的方法而修饰,例如位点定向诱变、PCR介导的诱变,和全基因合成以及本领域已知的其它方法。

[0025] 本文公开的RNA指导的核酸内切酶包含至少一个核定位信号。一般而言,NLS包含一段碱性氨基酸序列。核定位信号是本领域已知的(参见,例如,Lange等人,J.Biol.Chem., 2007,282:5101-5105)。例如,在一个实施方案中,NLS可以是单分型序列(monopartite sequence),例如PKKKRKV(SEQ ID NO:1)或PKKKRRV(SEQ ID NO:2)。在另一实施方案中,NLS可以是二分型序列(bipartite sequence)。在仍另一实施方案中,NLS可以是KRPAATKKAGQAKKKK(SEQ ID NO:3)。NLS可以位于RNA指导的核酸内切酶的N-末端、C-末端或内部位置。

[0026] 在一些实施方案中,RNA指导的核酸内切酶可以进一步包含至少一个细胞穿透结构域。在一个实施方案中,细胞穿透结构域可以是衍生自HIV-1TAT蛋白的细胞穿透肽序列。例如,TAT细胞穿透序列可以是GRKKRRQRRRPPQPKKKRKV(SEQ ID NO:4)。在另一实施方案中,细胞穿透结构域可以是TLM(PLSSIFSRIGDPPKKKKRKV;SEQ ID NO:5),衍生自人乙型肝炎病毒的细胞穿透肽序列。在仍另一实施方案中,细胞穿透结构域可以是MPG(GALFLGWLGAAGSTMGAPKKKKRKV;SEQ ID NO:6或GALFLGFLGAAGSTMGAWSQPKKKRKV;SEQ ID NO:7)。在另外的实施方案中,细胞穿透结构域可以是Pep-1(KETWWETWWTEWSQPKKKRKV;SEQ ID NO:8)、VP22,来自单纯疱疹病毒的细胞穿透肽或多精氨酸肽序列。细胞穿透结构域可定位于蛋白的N-末端、C-末端或内部位置。

[0027] 在仍其它的实施方案中,RNA指导的核酸内切酶也可以包含至少一个标志物结构

域。标志物结构域的非限制性实例包括荧光蛋白、纯化标签和表位标签。在一些实施方案中,标志物结构域可以是荧光蛋白。合适的荧光蛋白的非限制性实例包括绿色荧光蛋白(例如,GFP、GFP-2、tagGFP、turboGFP、EGFP、Emerald、Azami Green、Monomeric Azami Green、CopGFP、AceGFP、ZsGreen1)、黄色荧光蛋白(例如YFP、EYFP、Citrine、Venus、YPet、PhiYFP、ZsYellow1)、蓝色荧光蛋白(例如EBFP、EBFP2、Azurite、mKalamal、GFPuv、Sapphire、Tsapphire)、蓝绿色荧光蛋白(例如ECFP、Cerulean、CyPet、AmCyan1、Midoriishi-Cyan)、红色荧光蛋白(mKate、mKate2、mPlum、DsRed monomer、mCherry、mRFP1、DsRed-Express、DsRed2、DsRed-Monomer、HcRed-Tandem、HcRed1、AsRed2、eqFP611、mRaspberry、mStrawberry、Jred)和橙色荧光蛋白(mOrange、mK0、Kusabira-Orange、Monomeric Kusabira-Orange、mTangerine、tdTomato)或任何其它合适的荧光蛋白。在其它实施方案中,标志物结构域可以是纯化标签和/或表位标签。例示性的标签包括但不限于谷胱甘肽-S-转移酶(GST)、几丁质结合蛋白(CBP)、麦芽糖结合蛋白、硫氧还蛋白(TRX)、聚(NANP)、串联亲和力纯化(TAP)标签、myc、AcV5、AU1、AU5、E、ECS、E2、FLAG、HA、nus、Softag 1、Softag 3、Strep、SBP、Glu-Glu、HSV、KT3、S、S1、T7、V5、VSV-G、6xHis、生物素羧基载体蛋白(BCCP)和钙调蛋白。

[0028] 在某些实施方案中, RNA 指导的核酸内切酶可以是包含指导 RNA 的蛋白-RNA 复合物的一部分。指导 RNA 与 RNA 指导的核酸内切酶相互作用以将核酸内切酶定向至特定的靶位点, 其中指导 RNA 的 5' 端与特定的原间隔序列 (protospacer sequence) 碱基配对。

[0029] (II) 融合蛋白

本公开内容的另一方面提供了包含 CRISPR/Cas 样蛋白或其片段和效应子结构域的融合蛋白。通过指导 RNA 将 CRISPR/Cas 样蛋白定向至靶位点, 在该位点效应子结构域可修饰靶核酸序列或使其生效。效应子结构域可以是剪切结构域、表观遗传修饰结构域、转录激活结构域或转录抑制子结构域。融合蛋白可以进一步包含至少一个选自核定位信号、细胞穿透结构域或标志物结构域的另外的结构域。

[0030] (a) CRISPR/Cas 样蛋白

融合蛋白包含 CRISPR/Cas 样蛋白或其片段。CRISPR/Cas 样蛋白在上面的部分 (I) 中详述。CRISPR/Cas 样蛋白可以定位于融合蛋白的 N-末端、C-末端或内部位置。

[0031] 在一些实施方案中, 融合蛋白的 CRISPR/Cas 样蛋白可以衍生自 Cas9 蛋白。Cas9 衍生的蛋白可以是野生型、修饰的或其片段。在一些实施方案中, Cas9 衍生的蛋白可以经修饰以仅包含一个功能性核酸酶结构域 (RuvC 样或 HNH 样核酸酶结构域)。例如, Cas9 衍生的蛋白可以被修饰使得核酸酶结构域之一被删除或突变使得它不再有功能 (即, 缺乏核酸酶活性)。在其中核酸酶结构域之一是无活性的一些实施方案中, Cas9 衍生的蛋白能够将切口引入双链核酸 (这类蛋白被称为“切口酶”), 但不剪切双链 DNA。例如, 在 RuvC 样结构域中天冬氨酸至丙氨酸 (D10A) 的转换将 Cas9 衍生的蛋白转化成切口酶。同样, 在 HNH 结构域中组氨酸至丙氨酸 (H840A 或 H839A) 的转换将 Cas9 衍生的蛋白转化成切口酶。在其它实施方案中, RuvC 样核酸酶结构域和 HNH 样核酸酶结构域两者都可被修饰或消除使得 Cas9 衍生的蛋白不能够在双链核酸上切口或剪切双链核酸。在仍其它的实施方案中, Cas9 衍生的蛋白的全部核酸酶结构域可以被修饰或消除使得 Cas9 衍生的蛋白缺失全部核酸酶活性。

[0032] 在任一上述实施方案中, 任何或所有核酸酶结构域可以通过一个或多个缺失突

变、插入突变和/或置换突变使用众所周知的方法而失活,所述方法例如位点定向诱变、PCR介导的诱变和全基因合成,以及本领域已知的其它方法。在例示性实施方案中,融合蛋白的CRISPR/Cas样蛋白衍生自其中全部核酸酶结构域已经被失活或缺失的Cas9蛋白。

[0033] (b) 效应子结构域

融合蛋白也包含效应子结构域。效应子结构域可以是剪切结构域、表观遗传修饰结构域、转录激活结构域或转录抑制子结构域。效应子结构域可以定位于融合蛋白的N-末端、C-末端或内部位置。

[0034] (i) 剪切结构域

在一些实施方案中,效应子结构域是剪切结构域。本文使用的“剪切结构域”是指剪切DNA的结构域。剪切结构域自任何核酸内切酶或核酸外切酶获得。可自其衍生剪切结构域的核酸内切酶的非限制性实例包括但不限于限制核酸内切酶和归巢核酸内切酶。见,例如, New England Biolabs Catalog或Belfort等人(1997) *Nucleic Acids Res.* 25:3379-3388。剪切DNA的另外的酶是已知的(例如, S1核酸酶; 绿豆核酸酶; 胰DNA酶I; 微球菌核酸酶; 酵母H0核酸内切酶)。也参见Linn等人(eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993。可将这些酶(或其功能片段)中的一种或多种用作剪切结构域的来源。

[0035] 在一些实施方案中,剪切结构域可以衍生自类型II-S核酸内切酶。类型II-S核酸内切酶在通常距离识别位点几个碱基对的位点剪切DNA并且,因此具有可分开的识别和剪切结构域。这些酶通常是单体,其短暂地缔合以形成二聚体以剪切在错开的位置上的DNA的每条链。合适的类型II-S核酸内切酶的非限制性实例包括BfiI、BpmI、BsaI、BsgI、BsmBI、BsmI、BspMI、FokI、MboII和SapI。在例示性的实施方案中,融合蛋白的剪切结构域是FokI剪切结构域或其衍生物。

[0036] 在某些实施方案中,类型II-S剪切可以被修饰以促进两种不同剪切结构域(两种不同剪切结构域的每一种与CRISPR/Cas样蛋白或其片段连接)的二聚化。例如, FokI的剪切结构域可以通过突变某些氨基酸残基而修饰。作为非限制性实例,在FokI剪切结构域的位置446、447、479、483、484、486、487、490、491、496、498、499、500、531、534、537和538的氨基酸残基是用于修饰的靶。例如,形成专性异源二聚体的FokI的修饰的剪切结构域包括这样的配对,其中第一修饰剪切结构域包含在氨基酸位置490和538的突变和第二修饰剪切结构域包含在氨基酸位置486和499的突变(Miller等人, 2007, *Nat. Biotechnol.* 25:778-785; Szczpek等人, 2007, *Nat. Biotechnol.* 25:786-793)。例如,在位置490的Glu(E)可被改变为Lys(K),在位置538的Ile(I)可被改变为在一个结构域(E490K, I538K)的K,在位置486的Gln(Q)可被改变为E,在位置499的I可被改变为在另一剪切结构域(Q486E, I499L)的Leu(L)。在其它实施方案中,修饰的FokI剪切结构域可包括三个氨基酸变化(Doyon等人2011, *Nat. Methods*, 8:74-81)。例如,一个修饰的FokI结构域(其被称为ELD)可包含Q486E、I499L、N496D突变,另一修饰的FokI结构域(其被称为KKR)可包含E490K、I538K、H537R突变。

[0037] 在例示性的实施方案中,融合蛋白的效应子结构域是FokI剪切结构域或修饰的FokI剪切结构域。

[0038] 在其中效应子结构域是剪切结构域且CRISPR/Cas样蛋白衍生自Cas9蛋白的实施方案中,Cas9衍生的蛋白可以如上文所讨论而被修饰使得它的核酸内切酶活性被消除。例如,Cas9衍生的蛋白可以通过突变RuvC和HNH结构域而修饰使得它们不再具有核酸酶活性。

[0039] (ii) 表观遗传修饰结构域

在其它实施方案中,融合蛋白的效应子结构域可以是表观遗传修饰结构域。一般而言,表观遗传修饰结构域改变组蛋白结构和/或染色体结构,而不改变DNA序列。改变组蛋白和/或染色质结构可导致基因表达的改变。表观遗传修饰的实例包括但不限于组蛋白中赖氨酸残基的乙酰化或甲基化,和DNA中的胞嘧啶残基的甲基化。合适的表观遗传修饰结构域的非限制性实例包括组蛋白乙酰基转移酶结构域、组蛋白脱乙酰基酶结构域、组蛋白甲基转移酶结构域、组蛋白脱甲基酶结构域、DNA甲基转移酶结构域和DNA脱甲基酶结构域。

[0040] 在其中效应子结构域是组蛋白乙酰基转移酶 (HAT) 结构域的实施方案中,HAT结构域可以自EP300 (即,E1A结合蛋白p300)、CREBBP (即,CREB结合蛋白)、CDY1、CDY2、CDYL1、CLOCK、ELP3、ESA1、GCN5 (KAT2A)、HAT1、KAT2B、KAT5、MYST1、MYST2、MYST3、MYST4、NCOA1、NCOA2、NCOA3、NCOAT、P/CAF、Tip60、TAFII250或TF3C4衍生。在一个这类实施方案中,HAT结构域是p300。

[0041] 在其中效应子结构域是表观遗传修饰结构域且CRISPR/Cas样蛋白衍生自Cas9蛋白的实施方案中,Cas9衍生的蛋白可以如上文所讨论而被修饰使得它的核酸内切酶活性被消除。例如,Cas9衍生的蛋白可以通过突变RuvC和HNH结构域而修饰使得它们不再具有核酸酶活性。

[0042] (iii) 转录激活结构域

在其它实施方案中,融合蛋白的效应子结构域可以是转录激活结构域。一般而言,转录激活结构域与转录控制元件和/或转录调控蛋白 (即,转录因子, RNA聚合酶等) 相互作用以增加和/或激活基因的转录。在一些实施方案中,转录激活结构域可以是而限于单纯疱疹病毒VP16激活结构域、VP64 (其为VP16的四聚衍生物)、NFκB p65激活结构域、p53激活结构域1和2、CREB (cAMP反应元件结合蛋白) 激活结构域、E2A激活结构域和NFAT (激活的T-细胞的核因子) 激活结构域。在其它实施方案中,转录激活结构域可以是Gal4、Gcn4、MLL、Rtg3、Gln3、Oaf1、Pip2、Pdr1、Pdr3、Pho4和Leu3。转录激活结构域可以是野生型,或它可以是原始转录激活结构域的修饰形式。在一些实施方案中,融合蛋白的效应子结构域是VP16或VP64转录激活结构域。

[0043] 在其中效应子结构域是转录激活结构域且CRISPR/Cas样蛋白衍生自Cas9蛋白的实施方案中,Cas9衍生的蛋白可以如本文讨论而修饰使得它的核酸内切酶活性被消除。例如,Cas9衍生的蛋白可以通过突变RuvC和HNH结构域而修饰使得它们不再具有核酸酶活性。

[0044] (iv) 转录抑制子结构域

在仍其它的实施方案中,融合蛋白的效应子结构域可以是转录抑制子结构域。一般而言,转录抑制子结构域与转录控制元件和/或转录调控蛋白 (即,转录因子、RNA聚合酶等) 相互作用以减少和/或终止基因的转录。合适的转录抑制子结构域的非限制性实例包括可诱导的cAMP早期抑制子 (ICER) 结构域、Kruppel相关的盒A (KRAB-A) 抑制子结构域、YY1富含甘氨酸的抑制子结构域、Sp1-样抑制子、E (sp1) 抑制子、IκB抑制子和MeCP2。

[0045] 在其中效应子结构域是转录抑制子结构域且CRISPR/Cas样蛋白衍生自Cas9蛋白的实施方案中,Cas9衍生的蛋白可以如本文讨论而修饰使得它的核酸内切酶活性被消除。例如,cas9可以通过突变RuvC和HNH结构域而修饰使得它们不具有核酸酶活性。

[0046] (c) 另外的结构域

在一些实施方案中,融合蛋白进一步包含至少一个另外的结构域。合适的另外的结构域的非限制性实例包括核定位信号、细胞穿透或易位结构域和标志物结构域。合适的核定位信号、细胞穿透结构域和标志物结构域的非限制性实例展示于上面的部分(I)中。

[0047] (d) 融合蛋白二聚体

在其中融合蛋白的效应子结构域是剪切结构域的实施方案中,可以形成包含至少一种融合蛋白的二聚体。二聚体可以是同源二聚体或异源二聚体。在一些实施方案中,异源二聚体包含两种不同的融合蛋白。在其它实施方案中,异源二聚体包含一种融合蛋白和一种另外的蛋白。

[0048] 在一些实施方案中,二聚物是同源二聚体,其中两个融合蛋白单体关于一级氨基酸序列是相同的。在其中二聚体是同源二聚体的一个实施方案中,Cas9衍生的蛋白经修饰使得它们的核酸内切酶活性被消除,即,使得它们不具有功能性核酸酶结构域。在其中Cas9衍生的蛋白经修饰使得它们的核酸内切酶活性被消除的某些实施方案中,各个融合蛋白单体包含相同的Cas9样蛋白和相同的剪切结构域。剪切结构域可以是任何剪切结构域,例如本文提供的任一例示性剪切结构域。在一个具体实施方案中,剪切结构域是FokI剪切结构域或修饰的FokI剪切结构域。在这类实施方案中,特异性指导RNA将融合蛋白单体定向至不同但紧密毗连的位点使得在二聚体形成时,两个单体的核酸酶结构域将在靶DNA中产生双链断裂。

[0049] 在其它实施方案中,二聚体是两种不同融合蛋白的异源二聚体。例如,各融合蛋白的CRISPR/Cas样蛋白可以衍生自不同的CRISPR/Cas蛋白或衍生自来自不同细菌种类的直向同源CRISPR/Cas蛋白。例如,每种融合蛋白可包含Cas9-样蛋白,所述Cas9-样蛋白自不同细菌种类衍生。在一些实施方案中,

各融合蛋白会识别不同靶位点(即,由原间隔基和/或PAM序列限定)。例如,指导RNA可以将异源二聚体放到不同但紧密毗连的位点使得它们的核酸酶结构域在靶DNA中产生有效的双链断裂。异源二聚体也可具有具备切口活性的修饰的Cas9蛋白使得切口位置不同。

[0050] 备选地,异源二聚体的两种融合蛋白可具有不同效应子结构域。在其中效应子结构域是剪切结构域的实施方案中,各融合蛋白可以含有不同的修饰的剪切结构域。例如,各融合蛋白可以含有不同的修饰的FokI剪切结构域,如在上面的部分(II)(b)(i)中详细描述。在这些实施方案中,Cas-9蛋白可以经修饰使得它们的核酸内切酶活性被消除。

[0051] 本领域技术人员应该理解的是,形成异源二聚体的两种融合蛋白可以在CRISPR/Cas样蛋白结构域和效应子结构域二方面都不同。

[0052] 在以上描述的实施方案的任一个中,同源二聚体或异源二聚体可以包含选自如上面详述的核定位信号(NLSs)、细胞穿透、易位结构域和标志物结构域的至少一个另外的结构域。

[0053] 在以上描述的实施方案的任一个中,Cas9衍生的蛋白之一或二者可以被修饰使得它们的核酸内切酶活性被消除或修饰。

[0054] 在仍备选的实施方案中,异源二聚体包含一种融合蛋白和一种另外的蛋白。例如,另外的蛋白可以是核酸酶。在一个实施方案中,核酸酶是锌指核酸酶。锌指核酸酶包含锌指DNA结合结构域和剪切结构域。锌指识别并结合三个(3)核苷酸。锌指DNA结合结构域可包含

从约三个锌指至约七个锌指。锌指DNA结合结构域可以衍生自天然存在的蛋白或它可以是工程化的。参见,例如,Beerli等人(2002) *Nat. Biotechnol.* 20:135-141; Pabo等人(2001) *Ann. Rev. Biochem.* 70:313-340; Isalan等人(2001) *Nat. Biotechnol.* 19:656-660; Segal等人(2001) *Curr. Opin. Biotechnol.* 12:632-637; Choo等人(2000) *Curr. Opin. Struct. Biol.* 10:411-416; Zhang等人(2000) *J. Biol. Chem.* 275 (43):33850-33860; Doyon等人(2008) *Nat. Biotechnol.* 26:702-708; 和Santiago等人(2008) *Proc. Natl. Acad. Sci. USA* 105:5809-5814。锌指核酸酶的剪切结构域可以是上面部分(II)(b)(i)中详述的任何剪切结构域。在例示性的实施方案中,锌指核酸酶的剪切结构域是FokI剪切结构域或修饰的FokI剪切结构域。所述锌指核酸酶会与包含FokI剪切结构域或修饰的FokI剪切结构域的融合蛋白二聚化。

[0055] 在一些实施方案中,锌指核酸酶可包含至少一个选自上面详述的核定位信号、细胞穿透或易位结构域的另外的结构域。

[0056] 在某些实施方案中,上面详述的融合蛋白的任一种或包含至少一种融合蛋白的二聚体可以是包含至少一种指导RNA的蛋白-RNA复合物的一部分。指导RNA与融合蛋白的CRISPR-Cas0样蛋白相互作用以将融合蛋白定向至特定的靶位点,其中指导RNA的5'端与特定的原间隔序列碱基配对。

[0057] (III) 编码RNA指导的核酸内切酶或融合蛋白的核酸

本公开内容的另一方面分别提供编码在上面部分(I)和(II)中描述的RNA指导的核酸内切酶或融合蛋白的任一种的核酸。核酸可以是RNA或DNA。在一个实施方案中,编码RNA指导的核酸内切酶或融合蛋白的核酸是mRNA。mRNA可以被5'加帽和/或3'多聚腺苷酸化。在另一实施方案中,编码RNA指导的核酸内切酶或融合蛋白的核酸是DNA。DNA可以存在于载体中(参见以下)。

[0058] 可将编码RNA指导的核酸内切酶或融合蛋白的核酸经密码子最优化用于在目标真核细胞或动物中有效翻译成蛋白。例如,可将密码子最优化用于人、小鼠、大鼠、仓鼠、奶牛、猪、猫、狗、鱼、两栖动物、植物、酵母、昆虫等(参见在www.kazusa.or.jp/codon/的密码子使用数据库)中的表达。用于密码子最优化的程序作为免费软件是可得(例如,在genomes.urv.es/OPTIMIZER的OPTIMIZER;在www.genscript.com/codon_opt.html的来自GenScript的OptimumGene™)。市售密码子最优化程序也是可得。

[0059] 在一些实施方案中,编码RNA指导的核酸内切酶或融合蛋白的DNA可以与至少一个启动子控制序列可操作性地连接。在一些重复中,DNA编码序列可以与启动子控制序列可操作性地连接用于在目标真核细胞或动物中表达。启动子控制序列可以是组成型的、受调控的或组织特异性的。合适的组成型的启动子控制序列包括但不限于巨细胞病毒立即早期启动子(CMV)、猿猴病毒(SV40)启动子、腺病毒主要晚期启动子、劳斯肉瘤病毒(RSV)启动子、小鼠乳腺瘤病毒(MMTV)启动子、磷酸甘油酸激酶(PGK)启动子、延长因子(ED1)- α 启动子、泛素启动子、肌动蛋白启动子、微管蛋白启动子、免疫球蛋白启动子、其片段或以上任一种的组合。合适的受调控的启动子控制序列的实例包括但不限于被热激、金属、类固醇、抗生素或醇调控的那些。组织特异性启动子的非限制性实例包括B29启动子、CD14启动子、CD43启动子、CD45启动子、CD68启动子、结蛋白启动子、弹性蛋白酶-1启动子、内皮糖蛋白启动子、纤连蛋白启动子、Flt-1启动子、GFAP启动子、GPIIb启动子、ICAM-2启动子、INF- β 启动子、Mb

启动子、NphsI启动子、OG-2启动子、SP-B启动子、SYN1启动子和WASP启动子。启动子序列可以是野生型或它可经修饰用于更高效或有效表达。在一个例示性的实施方案中,可将编码DNA与CMV启动子可操作性地连接用于在哺乳动物细胞中的组成型表达。

[0060] 在某些实施方案中,可将编码RNA指导的核酸内切酶或融合蛋白的序列与被用于体外mRNA合成的噬菌体RNA聚合酶识别的启动子序列可操作性地连接。在这类实施方案中,可将体外转录的RNA纯化用于下面部分(IV)和(V)中详述的方法。例如,启动子序列可以是T7、T3或SP6启动子序列或T7、T3或SP6启动子序列的变体。在例示性实施方案中,将编码融合蛋白的DNA与T7启动子可操作性地连接用于使用T7RNA聚合酶的体外mRNA合成。

[0061] 在替代实施方案中,可将编码RNA指导的核酸内切酶或融合蛋白的序列与启动子序列可操作性地连接用于RNA指导的核酸内切酶或融合蛋白在细菌或真核细胞中的体外表达。在这类实施方案中,可将表达的蛋白纯化用于下面部分(IV)和(V)中详述的方法。合适的细菌启动子包括但不限于T7启动子、lac操纵子启动子、trp启动子、其变体和其组合。例示性的细菌启动子是tac,其为trp和lac启动子的杂合物。合适的真核启动子的非限制性实例在下面列出。

[0062] 在另外的方面,编码RNA指导的核酸内切酶或融合蛋白的DNA也可与多聚腺苷酸化信号(例如,SV40polyA信号、牛生长激素(BGH)polyA信号等)和/或至少一个转录终止序列连接。此外,编码RNA指导的核酸内切酶或融合蛋白的序列也可与在上面的部分(I)中详述的编码至少一个核定位信号、至少一个细胞穿透结构域和/或至少一个标志物结构域的序列连接。

[0063] 在不同实施方案中,编码RNA指导的核酸内切酶或融合蛋白的DNA可存在于载体中。合适的载体包括质粒载体、噬菌粒、黏粒、人工/小-染色体、转座子和病毒载体(例如,慢病毒载体、腺伴随病毒载体等)。在一个实施方案中,编码RNA指导的核酸内切酶或融合蛋白的DNA存在于质粒载体中。合适的质粒载体的非限制性实例包括pUC、pBR322、pET、pBluescript和其变体。载体可包含另外的表达控制序列(例如,增强子序列、Kozak序列、多聚腺苷酸化序列、转录终止序列等)、选择标记序列(例如,抗生素抗性基因)、复制起点等。另外的信息可以在以下找到:“Current Protocols in Molecular Biology(分子生物学当前实验设计)”Ausubel等人,John Wiley&Sons,纽约,2003或“Molecular Cloning:A Laboratory Manual(分子克隆:实验室手册)”Sambrook和Russell,Cold Spring Harbor Press,Cold Spring Harbor,NY,第三版,2001。

[0064] 在一些实施方案中,包含编码RNA指导的核酸内切酶或融合蛋白的序列的表达载体可进一步包括编码指导RNA的序列。编码指导RNA的序列通常与至少一个转录控制序列可操作性地连接用于在目的细胞或胚胎中表达该指导RNA。例如,可将编码指导RNA的DNA与被RNA聚合酶III(Pol III)识别的启动子序列可操作性地连接。合适的Pol III启动子的实例包括但不限于哺乳动物U6、U3、H1和7SLRNA启动子。

[0065] (IV) 使用RNA指导的核酸内切酶修饰染色体序列的方法

本公开内容的另一方面包括在真核细胞或胚胎中修饰染色体序列的方法。所述方法包括向真核细胞或胚胎引入:(i)至少一种包含至少一个核定位信号的RNA指导的核酸内切酶或编码至少一种包含至少一个核定位信号的RNA指导的核酸内切酶的核酸、(ii)至少一种指导RNA或编码至少一种指导RNA的DNA和任选(iii)至少一种包含供体序列的供体多核苷

酸。所述方法进一步包括培养细胞或胚胎使得各指导RNA将RNA指导的核酸内切酶定向至染色体序列中的靶位点,其中RNA指导的核酸内切酶将双链断裂引入靶位点,并且双链断裂通过DNA修复过程修复使得染色体序列被修饰。

[0066] 在一些实施方案中,该方法可包括将一种RNA指导的核酸内切酶(或编码核酸)和一种指导RNA(或编码DNA)引入细胞或胚胎,其中所述RNA指导的核酸内切酶将一个双链断裂引入靶染色体序列。在其中不存在任选供体多核苷酸的实施方案中,染色体序列中的双链断裂可通过非同源末端连接(NHEJ)修复过程修复。因为NHEJ是易错的,在断裂的修复期间可发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。因此,靶染色体序列可被修饰或失活。例如,单核苷酸改变(SNP)可导致改变的蛋白产物,或编码序列的阅读框移位可失活或“敲除”序列使得不产生该蛋白产物。在其中存在任选供体多核苷酸的实施方案中,供体多核苷酸中的供体序列在修复双链断裂期间可以与靶位点的染色体序列交换或整合入靶位点的染色体序列。例如,在其中供体序列被分别与染色体序列中的靶位点的上游和下游序列具有基本序列同一性的上游和下游序列侧接的实施方案中,供体序列可以在由同源性指导的修复过程所介导的修复期间与位于靶位点的染色体序列交换或整合入位于靶位点的染色体序列。备选地,在其中供体序列被相容的突出端侧接(或者相容的突出端由RNA指导的核酸内切酶原位生成)的实施方案中,供体序列可在修复双链断裂期间通过非同源性修复过程与剪切的染色体序列直接连接。供体序列交换到染色体序列或供体序列整合到染色体序列修饰靶染色体序列或将外源序列引入细胞或胚胎的染色体序列。

[0067] 在其它实施方案中,所述方法可包括将两种RNA指导的核酸内切酶(或编码核酸)和两种指导RNA(或编码DNA)引入细胞或胚胎,其中所述RNA指导的核酸内切酶将两个双链断裂引入染色体序列。参见图3B。两个断裂可以是在多个碱基对内,几十个碱基对内或可以被数千个碱基对分开。在其中不存在任选供体多核苷酸的实施方案中,生成的双链断裂可通过非同源修复过程修复使得在断裂的修复期间缺失两个剪切位点之间的序列和/或可以发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。在其中存在任选供体多核苷酸的实施方案中,在供体多核苷酸中的供体序列在通过基于同源性的修复过程(例如,其中供体序列被分别与染色体序列中的靶位点的上游和下游序列具有基本序列同一性的上游和下游序列侧接的实施方案中)或非同源修复过程(例如,在其中供体序列通过相容的突出端侧接的实施方案中)的双链断裂修复期间可以与染色体序列交换或整合入染色体序列。

[0068] 在仍其它的实施方案中,所述方法可包括将经修饰以剪切双链序列的一条链的一种RNA-指导的核酸内切酶(或编码核酸)和两种指导RNA(或编码DNA)引入细胞或胚胎,其中各指导RNA将RNA指导的核酸内切酶定向至特异性靶位点,在该位点修饰的核酸内切酶剪切该双链染色体序列的一条链(即,切口),并且其中两个切口是在相对的链上,并且足够靠近而构成双链断裂。参见图3A。在其中不存在任选供体多核苷酸的实施方案中,生成的双链断裂可通过非同源修复过程修复使得在修复断裂期间可发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。在其中存在任选供体多核苷酸的实施方案中,供体多核苷酸中的供体序列在通过基于同源性的修复过程(例如,其中供体序列被分别与染色体序列中的靶位点的上游和下游序列具有基本序列同一性的上游和下游序列

侧接的实施方案中)或非同源修复过程(例如,在其中供体序列通过相容的突出端侧接的实施方案中)的双链断裂的修复期间可以与染色体序列交换或整合入染色体序列。

[0069] (a) RNA指导的核酸内切酶

所述方法包括向细胞或胚胎引入至少一种包含至少一个核定位信号的RNA指导的核酸内切酶或编码至少一种包含至少一个核定位信号的RNA指导的核酸内切酶的核酸。这样的RNA指导的核酸内切酶和编码RNA指导的核酸内切酶的核酸分别在上面部分(I)和(III)中描述。

[0070] 在一些实施方案中,可将RNA指导的核酸内切酶作为分离的蛋白引入细胞或胚胎。在这些实施方案中, RNA指导的核酸内切酶可进一步包含至少一个细胞穿透结构域,其促进蛋白的细胞摄取。在其它实施方案中,可将RNA指导的核酸内切酶作为mRNA分子引入细胞或胚胎。在仍其它的实施方案中,可将RNA指导的核酸内切酶作为DNA分子引入细胞或胚胎。一般而言,将编码融合蛋白的DNA序列与会在目标细胞或胚胎中起作用的启动子序列可操作性地连接。DNA序列可以是线性的,或DNA序列可以是载体的一部分。在仍其它的实施方案中,可将融合蛋白作为包含融合蛋白和指导RNA的RNA-蛋白复合物引入细胞或胚胎。

[0071] 在备选实施方案中,编码RNA指导的核酸内切酶的DNA可进一步包含编码指导RNA的序列。一般而言,将编码RNA指导的核酸内切酶的各序列和指导RNA可操作性地连接到合适的启动子控制序列,其允许在细胞或胚胎中分别表达RNA指导的核酸内切酶和指导RNA。编码RNA指导的核酸内切酶和指导RNA的DNA序列可进一步包括另外的表达控制、调控和/或加工序列。编码RNA指导的核酸内切酶和指导RNA的DNA序列可以为线性的或可以是载体的一部分。

[0072] (b) 指导RNA

所述方法也包括向细胞或胚胎引入至少一种指导RNA或编码至少一种指导RNA的DNA。指导RNA与RNA指导的核酸内切酶相互作用以将核酸内切酶定向至特定的靶位点,在该位点指导RNA的5'端与在染色体序列中的特异性原间隔序列碱基配对。

[0073] 各指导RNA包含三个区域:在与染色体序列中的靶位点互补的5'端的第一个区域、形成茎环结构的第二内部区域和基本上保持单链的第三3'区域。每一种指导RNA的第一区域不同使得每一种指导RNA指导融合蛋白至特异性靶位点。每一个指导RNA的第二和第三区域在全部指导RNA中可以相同。

[0074] 指导RNA的第一区域与在染色体序列中的靶位点的序列(即,原间隔序列)互补使得指导RNA的第一区域可与靶位点碱基配对。在不同实施方案中,指导RNA的第一区域可包含约10个核苷酸至超过约25个核苷酸。例如,指导RNA的第一区域和染色体序列中的靶位点之间的碱基配对的区域长度可以为约10、11、12、13、14、15、16、17、18、19、20、22、23、24、25个或超过25个核苷酸。在例示性实施方案中,指导RNA的第一区域长度是约19、20或21个核苷酸。

[0075] 指导RNA也包含形成二级结构的第二区域。在一些实施方案中,二级结构包含茎(或发夹)和环。环和茎的长度可变。例如,环长度可在约3-约10个核苷酸的范围内,茎长度可在约6-约20个核苷酸的范围内。茎可包含一个或多个1-约10个核苷酸的凸起。因此,第二区域的总体长度可以在约16-约60个核苷酸的范围内。在例示性实施方案中,环长度是约4个核苷酸和茎包含约12个碱基对。

[0076] 指导RNA也包含在3'端的基本上保持单链的第三区域。因此,第三区域不具有与目标细胞中任何染色体序列的互补性并且不具有与指导RNA的余下部分的互补性。第三区域的长度可变化。一般而言,第三区域长度超过约4个核苷酸。例如,第三区域的长度可以在约5-约60个核苷酸的长度范围内。

[0077] 指导RNA的第二和第三区域(也称为通用或支架区域)的组合长度可以在约30-约120个核苷酸的长度范围内。在一个方面,指导RNA的第二和第三区域的组合长度在约70-约100个核苷酸长度的范围内。

[0078] 在一些实施方案中,指导RNA包括包含全部三个区域的单分子。在其它实施方案中,指导RNA可包括两个分开的分子。第一RNA分子可包含指导RNA的第一区域和指导RNA的第二区域的“茎”的一半。第二RNA分子可包含指导RNA的第二区域的“茎”的另一半和指导RNA的第三区域。因此,在此实施方案中,第一和第二RNA分子各包含相互互补的核苷酸序列。例如,在一个实施方案中,第一和第二RNA分子各自包含与另一序列碱基配对的序列(约6-约20个核苷酸)以形成功能性指导RNA。

[0079] 在一些实施方案中,可将指导RNA作为RNA分子引入细胞或胚胎。RNA分子可在体外转录。备选地,RNA分子可以是化学合成的。

[0080] 在其它实施方案中,可将指导RNA作为DNA分子引入细胞或胚胎。在这种情况下,可将编码指导RNA的DNA与启动子控制序列可操作性地连接用于在目标细胞或胚胎中表达指导RNA。例如,可将RNA编码序列与被RNA聚合酶III(Pol III)识别的启动子序列可操作性地连接。合适的Pol III启动子的实例包括但不限于哺乳动物的U6或H1启动子。在例示性的实施方案中,将RNA编码序列与小鼠或人U6启动子连接。在其它例示性的实施方案中,将RNA编码序列与小鼠或人H1启动子连接。

[0081] 编码指导RNA的DNA分子可以是线性的或环形的。在一些实施方案中,编码指导RNA的DNA序列可以是载体的一部分。合适的载体包括质粒载体、噬菌粒、黏粒、人工/小-染色体、转座子和病毒载体。在例示性实施方案中,编码RNA指导的核酸内切酶的DNA存在于质粒载体中。合适的质粒载体的非限制性实例包括pUC、pBR322、pET、pBluescript和其变体。载体可包括另外的表达控制序列(例如,增强子序列、Kozak序列、多腺苷酸化序列、转录终止序列等)、选择标记序列(例如,抗生素抗性基因)、复制起点等等。

[0082] 在其中将RNA指导的核酸内切酶和指导RNA二者作为DNA分子引入细胞的实施方案中,每一种都可以是不同的分子(例如,含有融合蛋白编码序列的一种载体和含有指导RNA编码序列的第二载体)的一部分或二者都可以是相同分子(例如,一种包含融合蛋白和指导RNA二者的编码(和调控)序列的载体)的一部分。

[0083] (c) 靶位点

将与指导RNA连接的RNA指导的核酸内切酶定向至在染色体序列中的靶位点,其中RNA指导的核酸内切酶将双链断裂引入染色体序列。除了紧接着(下游)共有序列之外,靶位点不具有序列限制。此共有序列也被称为原间隔毗连基序(PAM)。PAM的实例包括但不限于NGG、NGGNG和NNAGAAW(其中N被限定为任何核苷酸和W被限定为A或T)。如上面的部分(IV)(b)中详述,指导RNA的第一区域(在5'端)与靶序列的原间隔基互补。通常,指导RNA的第一区域长度是约19-21个核苷酸。因此,在某些方面,染色体序列中靶位点的序列是5'-N₁₉₋₂₁-NGG-3'。PAM呈斜体。

[0084] 靶位点可以在基因的编码区中、在基因的内含子中、在基因的控制区中、在基因之间的非编码区中等。基因可以是蛋白编码基因或RNA编码基因。基因可以是任何目标基因。

[0085] (d) 任选供体多核苷酸

在一些实施方案中,所述方法进一步包括将至少一种供体多核苷酸引入胚胎。供体多核苷酸包含至少一个供体序列。在一些方面,供体多核苷酸的供体序列与内源或天然染色体序列相对应。例如,供体序列可与在靶位点或靶位点附近的染色体序列的一部分基本相同,但其包含至少一个核苷酸改变。因此,供体序列可包含在靶位点的野生型序列的修饰形式,使得当与天然序列整合或交换时,在靶染色体位置的序列包含至少一个核苷酸变化。例如,变化可以是一个或多个核苷酸的插入、一个或多个核苷酸的缺失、一个或多个核苷酸的置换或其组合。作为修饰序列整合的结果,细胞或胚胎/动物可产生相对于靶染色体序列的修饰的基因产物。

[0086] 在其它方面,供体多核苷酸的供体序列与外源序列相对应。本文使用的“外源”序列是指对于细胞或胚胎而言非天然的序列,或其在细胞或胚胎的基因组中的天然位置是在不同位置的序列。例如,外源序列可包含蛋白编码序列,其可以与外源的启动子控制序列可操作性地连接使得,当整合入基因组时,细胞或胚胎/动物能够表达由整合的序列编码的蛋白。备选地,可将外源序列整合入染色体序列使得它的表达由内源启动子控制序列调控。在其它重复中,外源序列可以是转录控制序列、另一表达控制序列、RNA编码序列等。将外源序列整合入染色体序列被称为“敲入”。

[0087] 本领域技术人员可以理解的是,供体序列的长度可以并且会变化。例如,供体序列长度可以从几个核苷酸变化至几百个核苷酸至几十万个核苷酸。

[0088] 包含上游和下游序列的供体多核苷酸。在一些实施方案中,供体多核苷酸中的供体序列被上游序列和下游序列侧接,其分别与位于染色体序列中的靶位点的上游和下游的序列具有基本序列同一性。因为这些序列类似性,供体多核苷酸的上游和下游序列允许供体多核苷酸和靶染色体序列之间的同源重组,使得供体序列可以被整合入染色体序列(或与染色体序列交换)。

[0089] 本文使用的上游序列是指与靶位点上游的染色体序列具有基本序列同一性的核酸序列。类似地,下游序列是指与靶位点下游的染色体序列具有基本序列同一性的核酸序列。本文使用的短语“基本序列同一性”是指具有至少约75%序列同一性的序列。因此,供体多核苷酸中的上游和下游序列可与靶位点上游或下游的序列具有约75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%或99%序列同一性。在例示性实施方案中,供体多核苷酸中的上游和下游序列可具有与靶位点的上游或下游的染色体序列约95%或100%序列同一性。在一个实施方案中,上游序列与紧接位于靶位点上游(即,与靶位点毗连)的染色体序列具有基本序列同一性。在其它实施方案中,上游序列与位于自靶位点上游约一百(100)个核苷酸内的染色体序列具有基本序列同一性。因此,例如,上游序列可与位于自靶位点上游约1-约20、约21-约40、约41-约60、约61-约80或约81-约100个核苷酸的染色体序列具有基本序列同一性。在一个实施方案中,下游序列与紧接位于靶位点下游(即,与靶位点毗连)的染色体序列具有基本序列同一性。在其它实施方案中,下游序列与位于自靶位点下游约一百(100)个核苷酸内的染色体序列具有基本序列同一性。因此,例如,下游序列可以与位于自

靶位点下游约1-约20、约21-约40、约41-约60、约61-约80或约81-约100个核苷酸的染色体序列具有基本序列同一性。

[0090] 各上游或下游序列长度可在约20个核苷酸-约5000个核苷酸的范围内。在一些实施方案中,上游和下游序列可包含约50、100、200、300、400、500、600、700、800、900、1000、1100、1200、1300、1400、1500、1600、1700、1800、1900、2000、2100、2200、2300、2400、2500、2600、2800、3000、3200、3400、3600、3800、4000、4200、4400、4600、4800或5000个核苷酸。在例示性的实施方案中,上游和下游序列长度可以在约50-约1500个核苷酸的范围内。

[0091] 包含具有与靶染色体序列的序列相似性的上游和下游序列的供体多核苷酸可以是线性或环形的。在其中供体多核苷酸是环形的实施方案中,它可以是载体的一部分。例如,载体可以是质粒载体。

[0092] 包含靶剪切位点的供体多核苷酸。在其它实施方案中,供体多核苷酸可以另外包含至少一个被RNA指导的核酸内切酶识别的靶剪切位点。加入供体多核苷酸的靶剪切位点可位于供体序列的上游或下游或上游和下游二者。例如,供体序列可以侧接靶剪切位点,使得当被RNA指导的核酸内切酶剪切时,供体序列与突出端侧接,所述突出端与被RNA指导的核酸内切酶剪切时生成的在染色体序列中那些突出端相容。因此,供体序列可以在通过非同源修复过程修复双链断裂期间与剪切的染色体序列连接。通常,包含靶剪切位点的供体多核苷酸会是环形的(例如,可以为质粒载体的一部分)。

[0093] 包含具有任选突出端的短供体序列的供体多核苷酸。在仍备选的实施方案中,供体多核苷酸可以是包含短供体序列的线性分子,所述线性分子具有与由RNA指导的核酸内切酶生成的突出端相容的任选短突出端。在这类实施方案中,供体序列可以在修复双链断裂期间直接与剪切的染色体序列连接。在一些例子中,供体序列可以是小于约1,000、小于约500、小于约250或小于约100个核苷酸。在某些情况下,供体多核苷酸可以是包含具有平端的短供体序列的线性分子。在其它重复中,供体多核苷酸可以是包含具有5'和/或3'突出端的短供体序列的线性分子。突出端可包含1、2、3、4或5个核苷酸。

[0094] 通常,供体多核苷酸是DNA。DNA可以是单链的或双链的和/或线性的或环形的。供体多核苷酸可以是DNA质粒、细菌人工染色体(BAC)、酵母人工染色体(YAC)、病毒载体、线性的DNA段、PCR片段、裸核酸或与递送载体如脂质体或泊洛沙姆(poloxamer)复合的核酸。在某些实施方案中,包含供体序列的供体多核苷酸可以是质粒载体的一部分。在这些任一种情况下,包含供体序列的供体多核苷酸可进一步包含至少一种另外的序列。

[0095] (e) 引入细胞或胚胎

RNA-靶向的核酸内切酶(或编码核酸),指导RNA(或编码DNA),和任选供体多核苷酸可通过各种方式引入细胞或胚胎。在一些实施方案中,转染细胞或胚胎。合适的转染方法包括磷酸钙-介导的转染、核转染(或电穿孔)、阳离子聚合物转染(例如,DEAE-葡聚糖或聚乙烯亚胺)、病毒转导、病毒颗粒(virosome)转染、病毒粒子(virion)转染、脂质体转染、阳离子脂质体转染、免疫脂质体转染、非脂质体型脂质转染、树形分子(dendrimer)转染、热激转染、磁转染、脂质转染、基因枪递送、穿刺转染(impalefection)、声孔效应、光学转染和专利品(proprietary agent)增强的核酸摄取。转染方法是本领域周知的(参见,例如,“Current Protocols in Molecular Biology(分子生物学当前实验设计)”Ausubel等人,John Wiley&Sons,New York,2003或“Molecular Cloning:ALaboratory Manual(分子克隆:实验

室手册) "Sambrook和Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 第三版, 2001)。在其它实施方案中, 将分子通过显微注射引入细胞或胚胎。通常, 胚胎是目标物种的受精的单细胞期胚胎。例如, 可将分子注射入单细胞胚胎的前核。

[0096] 可将RNA-靶向的核酸内切酶(或编码核酸)、指导RNA(或编码指导RNA的DNA)和任选供体多核苷酸同时地或贯序地引入细胞或胚胎。RNA-靶向的核酸内切酶(或编码核酸)相对于指导RNA(或编码DNA)的比率通常为约化学计量比, 使得它们可以形成RNA-蛋白复合物。在一个实施方案中, 将编码RNA-靶向的核酸内切酶的DNA和编码指导RNA的DNA在质粒载体内一起递送。

[0097] (f) 培养细胞或胚胎

所述方法进一步包含将细胞或胚胎维持在合适的条件下, 使得指导RNA将RNA指导的核酸内切酶定向至在染色体序列中的靶位点, 和RNA指导的核酸内切酶将至少一个双链断裂引入染色体序列。双链断裂可通过DNA修复过程修复使得染色体序列通过至少一个核苷酸的缺失, 至少一个核苷酸的插入, 至少一个核苷酸的置换或其组合而修饰。

[0098] 在其中没有供体多核苷酸被引入细胞或胚胎的实施方案中, 双链断裂可以通过非同源末端连接(NHEJ)修复过程而修复。因为NHEJ是易错的, 因此在断裂修复期间可发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。因此, 在染色体序列的序列可被修饰使得编码区阅读框可被移位并且染色体序列被失活或“敲除”。失活的蛋白编码染色体序列不产生由野生型染色体序列编码的蛋白。

[0099] 在其中将包含上游和下游序列的供体多核苷酸引入细胞或胚胎的实施方案中, 双链断裂可以通过同源性定向的修复(HDR)过程修复使得供体序列被整合入染色体序列。因此, 外源序列可被整合入细胞或胚胎的基因组, 或可通过用修饰的序列替换野生型染色体序列而修饰靶染色体序列。

[0100] 在其中将包含靶剪切位点的供体多核苷酸引入细胞或胚胎的实施方案中, RNA指导的核酸内切酶可剪切靶染色体序列和供体多核苷酸二者。可将线性化的供体多核苷酸通过NHEJ过程通过将供体多核苷酸和剪切的染色体序列之间连接而整合入双链断裂位点的染色体序列。

[0101] 在其中将包含短供体序列的线性的供体多核苷酸引入细胞或胚胎的实施方案中, 短供体序列可通过NHEJ过程整合入双链断裂位点的染色体序列。整合可通过将短供体序列和在双链断裂位点的染色体序列之间的平端连接而进行。或者, 整合可以通过将侧接突出端的短供体序列之间的粘性末端(即, 具有5'或3'突出端)连接而进行, 所述突出端与由RNA-靶向核酸内切酶在剪切的染色体序列中生成的那些突出端是相容的。

[0102] 一般而言, 将细胞维持在适合于细胞生长和/或维持的条件下。合适的细胞培养条件是本领域周知的, 并描述于, 例如, Santiago等人(2008)PNAS 105:5809-5814; Moehle等人(2007)PNAS 104:3055-3060; Urnov等人(2005)Nature 435:646-651; 和Lombardo等人(2007)Nat. Biotechnology 25:1298-1306。本领域技术人员理解的是, 用于培养细胞的方法是本领域已知的并且可以根据细胞类型而改变。在全部情况下, 可使用常规的最优化以确定用于特定细胞种类的最好的技术。

[0103] 胚胎可在体外(例如, 在细胞培养中)培养。通常, 如果必要, 将胚胎于合适的温度下和在合适的培养基中用必要的O₂/CO₂比率培养以允许RNA核酸内切酶和指导RNA的表达。

培养基的合适的非限制性实例包括M2、M16、KSOM、BMOC和HTF培养基。技术人员会理解的是，培养条件可以并会根据胚胎的种类而改变。在全部情况下，可使用常规的最优化以确定用于特定胚胎种类的最好的培养条件。在一些情况下，细胞系可以衍生自体外培养的胚胎（例如，胚胎干细胞系）。

[0104] 或者，可通过将胚胎转移至雌性宿主的子宫在体内培养胚胎。通常而言，雌性宿主来自与胚胎相同或类似的种类。优选地，雌性宿主是假孕的。制备假孕的雌性宿主的方法是本领域已知的。此外，将胚胎转移入雌性宿主的方法是本领域已知的。在体内培养胚胎允许胚胎发育并可以产生由胚胎衍生的动物的活产。这类动物在身体的每一个细胞中包含修饰的染色体序列。

[0105] (g) 细胞和胚胎类型

多种真核细胞和胚胎适合用于该方法。例如，细胞可以是人细胞、非人哺乳动物细胞、非哺乳动物的脊椎动物细胞、无脊椎动物细胞、昆虫细胞、植物细胞、酵母细胞或单细胞真核生物。一般而言，胚胎是非人哺乳动物胚胎。在特定实施方案中，胚胎可以是单细胞非人哺乳动物胚胎。包括单细胞胚胎的例示性的哺乳动物的胚胎包括但不限于小鼠、大鼠、仓鼠、啮齿动物、兔、猫、犬、羊、猪、牛、马和灵长类动物胚胎。在仍其它的实施方案中，细胞可以是干细胞。合适的干细胞包括但不限于胚胎干细胞、ES样干细胞、胎儿干细胞、成体干细胞、多能性干细胞、诱导的多能性干细胞、多能干细胞、寡能干细胞、单能干细胞等等。在例示性的实施方案中，细胞是哺乳动物细胞。

[0106] 合适的哺乳动物细胞的非限制性实例包括中国仓鼠卵巢 (CHO) 细胞、幼仓鼠肾 (BHK) 细胞；小鼠骨髓瘤NS0细胞、小鼠胚胎成纤维细胞3T3细胞 (NIH3T3)、小鼠B淋巴瘤A20细胞；小鼠黑色素瘤B16细胞；小鼠成肌细胞C2C12细胞；小鼠骨髓瘤SP2/0细胞；小鼠胚胎间充质细胞C3H-10T1/2细胞；小鼠癌CT26细胞、小鼠前列腺DuCuP细胞、小鼠乳腺EMT6细胞；小鼠肝癌Hepa1c1c7细胞；小鼠骨髓瘤J5582细胞；小鼠上皮MTD-1A细胞；小鼠心肌MyEnd细胞；小鼠肾RenCa细胞；小鼠胰RIN-5F细胞；小鼠黑色素瘤X64细胞；小鼠淋巴瘤YAC-1细胞；大鼠成胶质细胞瘤9L细胞；大鼠B淋巴瘤RBL细胞；大鼠成神经细胞瘤B35细胞；大鼠肝癌细胞 (HTC)；牛鼠 (buffalo rat) 肝BRL3A细胞；犬肾细胞 (MDCK)；犬乳腺 (CMT) 细胞；大鼠骨肉瘤D17细胞；大鼠单核细胞/巨噬细胞DH82细胞；猴肾SV-40转化成纤维细胞 (COS7) 细胞；猴肾CVI-76细胞；非洲绿猴肾 (VERO-76) 细胞；人胚胎肾细胞 (HEK293, HEK293T)；人颈癌细胞 (HELA)；人肺细胞 (W138)；人肝细胞 (Hep G2)；人U2-OS骨肉瘤细胞、人A549细胞、人A-431细胞和人K562细胞。哺乳动物细胞系的详尽的列表可在美国模式培养物保藏所目录 (ATCC, Mamassas, VA) 中找到。

[0107] (v) 使用融合蛋白以修饰染色体序列或调控染色体序列的表达的方法

本公开内容的另一方面包括一种用于在细胞或胚胎中修饰染色体序列或调控染色体序列表达的方法。所述方法包括向细胞或胚胎中引入 (a) 至少一种融合蛋白或编码至少一种融合蛋白的核酸，其中所述融合蛋白包含CRISPR/Cas样蛋白或其片段和效应子结构域，和 (b) 至少一种指导RNA或编码该指导RNA的DNA，其中所述指导RNA指导该融合蛋白的CRISPR/Cas样蛋白至染色体序列中的靶位点，并且该融合蛋白的效应子结构域修饰染色体序列或调控染色体序列的表达。

[0108] 包含CRISPR/Cas样蛋白或其片段和效应子结构域的融合蛋白在上面的部分 (II)

中详述。一般而言,本文公开的融合蛋白进一步包含至少一个核定位信号。编码融合蛋白的核酸在上面的部分(III)中详述。在一些实施方案中,融合蛋白可以作为分离的蛋白(其可以进一步包含细胞穿透结构域)引入细胞或胚胎。此外,分离的融合蛋白可以是包含指导RNA的蛋白-RNA复合物的一部分。在其它实施方案中,可将融合蛋白作为RNA分子(其可以被加帽和/或多聚腺苷酸化)引入细胞或胚胎。在仍其它的实施方案中,可将融合蛋白作为DNA分子引入细胞或胚胎。例如,可将融合蛋白和指导RNA作为离散的DNA分子或相同的DNA分子的一部分引入细胞或胚胎。这类DNA分子可以是质粒载体。

[0109] 在一些实施方案中,所述方法进一步包括向细胞或胚胎引入至少一种锌指核酸酶。锌指核酸酶在上面的部分(II)(d)中描述。在仍其它的实施方案中,所述方法进一步包括向细胞或胚胎引入至少一种供体多核苷酸。供体多核苷酸在上面的部分(IV)(d)中详述。用于将分子引入细胞或胚胎的方式,和用于培养细胞或胚胎的方式分别在上面的部分(IV)(e)和(IV)(f)中描述。合适的细胞和胚胎在上面的部分(IV)(g)中描述。

[0110] 在其中融合蛋白的效应子结构域是剪切结构域(例如,FokI剪切结构域或修饰的FokI剪切结构域)的某些实施方案中,所述方法可包括向细胞或胚胎引入一种融合蛋白(或编码一种融合蛋白的核酸)和两种指导RNA(或编码两种指导RNA的DNA)。所述两种指导RNA将融合蛋白定向至染色体序列中的两个不同靶位点,其中融合蛋白二聚化(例如,形成同源二聚体)使得所述两个剪切结构域可将双链断裂引入染色体序列。参见图1A。在其中不存在任选供体多核苷酸的实施方案中,染色体序列中的双链断裂可通过非同源末端连接(NHEJ)修复过程而修复。因为NHEJ是易错的,所以在断裂修复期间可发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。因此,靶染色体序列可以被修饰或失活。例如,单核苷酸改变(SNP)可产生改变的蛋白产物,或编码序列阅读框的移位可以失活或“敲除”序列使得不产生该蛋白产物。在其中存在任选供体多核苷酸的实施方案中,供体多核苷酸中的供体序列可在双链断裂的修复期间与靶位点的染色体序列置换或整合入靶位点的染色体序列。例如,在其中供体序列被分别与染色体序列中的靶位点的上游和下游序列具有基本序列同一性的上游和下游序列侧接的实施方案中,供体序列在通过同源性指导修复过程介导的修复期间可与靶位点的染色体序列置换或整合入靶位点的染色体序列。或者,在其中供体序列被相容的突出端侧接(或相容的突出端由RNA指导的核酸内切酶原位生成)的实施方案中,供体序列在双链断裂的修复期间可以通过非同源修复过程与剪切的染色体序列直接连接。供体序列至染色体序列的置换或整合修饰靶染色体序列或将外源序列引入细胞或胚胎的染色体序列。

[0111] 在其中融合蛋白的效应子结构域是剪切结构域(例如,FokI剪切结构域或修饰的FokI剪切结构域)的其它实施方案中,所述方法可以包括向细胞或胚胎中引入两种不同的融合蛋白(或编码两种不同的融合蛋白的核酸)和两种指导RNA(或编码两种指导RNA的DNA)。融合蛋白可以如在上面部分(II)中详述而不同。各指导RNA将融合蛋白定向至染色体序列中的具体靶位点,其中所述融合蛋白二聚化(例如,形成异源二聚体)使得两个剪切结构域可将双链断裂引入染色体序列。在其中不存在任选供体多核苷酸的实施方案中,生成的双链断裂可通过非同源修复过程修复,使得在断裂修复期间可发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。在其中存在任选供体多核苷酸的实施方案中,供体多核苷酸中的供体序列可与在通过基于同源性的修复过程(例如,在

其中供体序列被分别与染色体序列中的靶位点的上游和下游序列具有基本序列同一性的上游和下游序列侧接的实施方案中)或非同源修复过程(例如,在其中供体序列被相容的突出端侧接的实施方案中)的双链断裂的修复期间与染色体序列置换或整合入染色体序列。

[0112] 在其中融合蛋白的效应子结构域是剪切结构域(例如,FokI剪切结构域或修饰的FokI剪切结构域)的仍其它实施方案中,所述方法可包括向细胞或胚胎引入一种融合蛋白(或编码一种融合蛋白的核酸),一种指导RNA(或编码一种指导RNA的DNA)和一种锌指核酸酶(或编码锌指核酸酶的核酸),其中锌指核酸酶包含FokI剪切结构域或修饰的FokI剪切结构域。指导RNA将融合蛋白定向至具体的染色体序列,并且锌指核酸酶被定向至另一染色体序列,其中所述融合蛋白和锌指核酸酶二聚化使得融合蛋白的剪切结构域和锌指核酸酶的剪切结构域可以将双链断裂引入染色体序列。参见图1B。在其中不存在任选供体多核苷酸的实施方案中,生成的双链断裂可以通过非同源修复过程而修复,使得在断裂的修复期间可以发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。在其中存在任选供体多核苷酸的实施方案中,供体多核苷酸中的供体序列在通过基于同源性的修复过程(例如,在其中供体序列被分别与染色体序列中的靶位点的上游和下游序列具有基本序列同一性的上游和下游序列侧接的实施方案中)或非同源修复过程(例如,在其中供体序列被相容的突出端侧接的实施方案中)期间可以与染色体序列置换或整合入染色体序列。

[0113] 在其中融合蛋白的效应子结构域是转录激活结构域或转录抑制子结构域的仍其它实施方案中,所述方法可包括向细胞或胚胎引入一种融合蛋白(或编码一种融合蛋白的核酸)和一种指导RNA(或编码一种指导RNA的DNA)。指导RNA将融合蛋白定向至具体的染色体序列,其中所述转录激活结构域或转录抑制子结构域分别激活或抑制靶染色体序列的表达。参见图2A。

[0114] 在其中融合蛋白的效应子结构域是表观遗传修饰结构域的备选实施方案中,所述方法可包括向细胞或胚胎引入一种融合蛋白(或编码一种融合蛋白的核酸)和一种指导RNA(或编码一种指导RNA的DNA)。指导RNA将融合蛋白定向至具体的染色体序列,其中表观遗传修饰结构域修饰靶染色体序列的结构。参见图2A。表观遗传修饰包括组蛋白的乙酰化、甲基化和/或核苷酸甲基化。在一些例子中,染色体序列的结构修饰导致染色体序列的表达的变化。

[0115] (VI) 遗传修饰的细胞和动物

本公开内容包括遗传修饰的细胞、非人胚胎和非人动物,其包含至少一个使用RNA指导的核酸内切酶-介导的或融合蛋白-介导的过程(例如,使用本文描述的方法)修饰的染色体序列。本公开内容提供细胞,其包含至少一种编码RNA指导的核酸内切酶或靶向至目标染色体序列的融合蛋白或融合蛋白的DNA或RNA分子、至少一种指导RNA,和任选一种或多种供体多核苷酸。本公开内容也提供非人胚胎,其包含至少一种编码RNA指导的核酸内切酶或靶向至目标染色体序列的融合蛋白的DNA或RNA分子,至少一种指导RNA和任选一种或多种供体多核苷酸。

[0116] 本公开内容提供遗传修饰的非人动物、非人胚胎或动物细胞,其包含至少一个修饰的染色体序列。修饰的染色体序列可以被修饰使得它(1)失活,(2)具有改变的表达,或产生改变的蛋白产物,或(3)包含整合的序列。使用本文描述的方法,将染色体序列用RNA指导

的核酸内切酶-介导的或融合蛋白-介导的过程修饰。

[0117] 如讨论的,本公开内容的一个方面提供其中至少一个染色体序列已经被修饰的遗传修饰的动物。在一个实施方案中,遗传修饰的动物包含至少一个失活的染色体序列。修饰的染色体序列可以被失活使得序列不被转录和/或不产生功能性蛋白产物。因此,包含失活的染色体序列的遗传修饰的动物可以称为“敲除”或“条件性敲除”。失活的染色体序列可以包括缺失突变(即,一个或多个核苷酸的缺失),插入突变(即,一个或多个核苷酸的插入)或无义突变(即,用单核苷酸置换另一核苷酸使得终止密码子被引入)。作为突变的结果,靶染色体序列被失活并且不产生功能性蛋白。失活的染色体序列包含非内源引入的序列。本文还包括遗传修饰的动物,其中2、3、4、5、6、7、8、9或10个或更多个染色体序列被失活。

[0118] 在另一实施方案中,修饰的染色体序列可以被改变使得它编码变体蛋白产物。例如,包含修饰的染色体序列的遗传修饰的动物可以包含靶点突变或其它修饰使得产生改变的蛋白产物。在一个实施方案中,染色体序列可以被修饰使得至少一个核苷酸被改变并且该表达的蛋白包含一个改变的氨基酸残基(错义突变)。在另一实施方案中,染色体序列可以被修饰以包含超过一个错义突变使得改变超过一个氨基酸。此外,染色体序列可以被修饰以具有3个核苷酸的缺失或插入使得表达的蛋白含有单氨基酸缺失或插入。改变的或变体蛋白与野生型蛋白相比可以具有改变的性质或活性,例如改变的底物特异性、改变的酶活性、改变的动力学速率等。

[0119] 在另一实施方案中,遗传修饰的动物可以包含至少一个染色体整合的序列。包含整合序列的遗传修饰的动物可以被称为“敲入”或“条件性敲入”。染色体整合的序列可以,例如,编码直向同源蛋白、内源蛋白或两者的组合。在一个实施方案中,编码直向同源蛋白或内源蛋白的序列可以被整合入编码蛋白的染色体序列使得染色体序列失活,但是外源序列表达。在这类情况下,编码直向同源蛋白或内源蛋白的序列可以被可操作性地连接至启动子控制序列。或者,编码直向同源蛋白或内源蛋白的序列可以被整合入染色体序列,而不影响染色体序列的表达。例如,编码蛋白的序列可以被整合入“安全港(safe harbor)”基因座,例如Rosa26基因座、HPRT基因座或AAV基因座。本公开内容也包含遗传修饰的动物,其中2、3、4、5、6、7、8、9或10或更多个序列(包括编码蛋白的序列)被整合入基因组。

[0120] 染色体整合的编码蛋白的序列可以编码目标蛋白的野生型形式或可以编码包含至少一个修饰的蛋白使得产生改变形式的蛋白。例如,染色体整合的编码涉及疾病或病症的蛋白的序列可以包含至少一个修饰,使得产生的改变形式的蛋白引起相关病症或使得所述相关病症成为可能。或者,染色体整合的编码涉及疾病或病症的蛋白的序列可以包含至少一个修饰使得改变形式的蛋白防止所述相关病症的发展。

[0121] 在另外的实施方案中,遗传修饰的动物可以是“人源化”动物,其包含至少一个编码功能性人蛋白的染色体整合的序列。功能性人蛋白在遗传修饰的动物中可以不具有对应的直向同源物。或者,由其衍生出遗传修饰的动物的野生型动物可以包括与功能性人蛋白对应的直向同源物。在这种情况下,“人源化”动物中的直向同源序列被失活使得不产生功能性蛋白,并且“人源化”动物包含至少一个染色体整合的编码人蛋白的序列。

[0122] 在更另一实施方案中,遗传修饰的动物可包含至少一个修饰的编码蛋白的染色体序列使得蛋白的表达模式被改变。例如,控制蛋白表达的调控区域,例如启动子或转录因子结合位点,可以被改变,使得蛋白过量产生,或蛋白的组织特异性或时间性表达被改变,或

其组合。或者,蛋白的表达模式可以使用条件性敲除系统改变。条件性敲除系统的非限制性实例包括Cre-lox重组系统。Cre-lox重组系统包括Cre重组酶、位点-特异性DNA重组酶,其可以催化在核酸分子中的特异性位点(lox位点)之间的核酸序列的重组。使用此系统以产生时间性的和组织特异性表达的方法是本领域已知的。一般而言,遗传修饰的动物用侧接染色体序列的lox位点生成。包含lox-侧接的染色体序列的遗传修饰的动物可以随后与表达Cre重组酶的另一遗传修饰的动物杂交。包含lox-侧接的染色体序列和Cre重组酶的子代动物随后产生,并且lox-侧接的染色体序列被重组,这导致编码该蛋白的染色体序列缺失或倒位。Cre重组酶的表达可以被时间性地和条件性地调控以实现时间性和条件性调控的染色体序列的重组。

[0123] 在这些实施方案的任一个中,本文公开的遗传修饰的动物对于修饰的染色体序列而言可以是杂合的。或者,遗传修饰动物对于修饰的染色体序列而言可以是纯合的。

[0124] 本文公开的遗传修饰的动物可经杂交以生成包含超过一个修饰的染色体序列的动物或生成对于一个或多个修饰的染色体序列而言是纯合的动物。例如,包含相同修饰的染色体序列的两个动物可经杂交以生成对于修饰的染色体序列而言纯合的动物。或者,具有不同修饰的染色体序列的动物可经杂交以生成包含两种修饰的染色体序列的动物。

[0125] 例如,包含失活的染色体序列基因“x”的第一动物可以与包含编码人基因“X”蛋白的染色体整合的序列的第二动物杂交以产生包含失活的基因“x”染色体序列和染色体整合的人基因“X”序列二者的“人源化”基因“X”子代。同样,人源化基因“X”动物可以与人源化基因“Y”动物杂交以生成人源化基因X/基因Y子代。本领域技术人员会理解许多组合是有可能的。

[0126] 在其它实施方案中,包含修饰的染色体序列的动物可经杂交以将修饰的染色体序列和其它遗传背景组合。作为非限制性实例,其它遗传背景可以包括野生型遗传背景、具有缺失突变的遗传背景、具有另一靶整合的遗传背景和具有非靶整合的遗传背景。

[0127] 本文使用的术语“动物”是指非人动物。动物可以是胚胎、幼体或成体。合适的动物包括脊椎动物例如哺乳动物、鸟类、爬行动物、两栖动物、甲壳类动物和鱼类。合适的哺乳动物的实例包括但不限于啮齿动物、陪伴动物、家畜和灵长类动物。啮齿动物的非限制性实例包括小鼠、大鼠、仓鼠、沙鼠和豚鼠。合适的陪伴动物包括但不限于猫、狗、兔、刺猬和雪貂。家畜的非限制性实例包括马、山羊、绵羊、猪、牛、美洲驼和羊驼。合适的灵长类动物包括但不限于僧帽猴、黑猩猩、狐猴、恒河猴、狨、绢毛猴、蜘蛛猴、松鼠猴和长尾猴(vervet monkey)。鸟类的非限制性实例包括鸡、火鸡、鸭和鹅。或者,动物可以是无脊椎动物例如昆虫、线虫等等。昆虫的非限制性实例包括果蝇和蚊子。例示性的动物是大鼠。合适的大鼠品系的非限制性实例包括Dahl Salt-Sensitive、Fischer 344、Lewis、Long Evans Hooded、Sprague-Dawley和Wistar。在一个实施方案中,动物不是遗传修饰的小鼠。在对于本发明合适的动物的上文重述的每一种中,动物不包括内源引入的,随机整合的转座子序列。

[0128] 本公开内容的进一步方面提供包含至少一个修饰的染色体序列的遗传修饰的细胞或细胞系。遗传修饰的细胞或细胞系可以自本文公开的遗传修饰的动物的任一种衍生。或者,染色体序列可以在细胞中如上文所描述(在描述动物中染色体序列修饰的段落中)使用本文描述的方法修饰。本公开内容也包括所述细胞或细胞系的裂解液。

[0129] 一般而言,细胞是真核细胞。合适的宿主细胞包括真菌或酵母,例如毕赤酵母属

(*Pichia*, *Saccharomyces*) 或裂殖酵母属 (*Schizosaccharomyces*) ; 昆虫细胞, 例如来自草地贪夜蛾 (*Spodoptera frugiperda*) 的 SF9 细胞或来自果蝇 (*Drosophila melanogaster*) 的 S2 细胞; 和动物细胞例如小鼠、大鼠、仓鼠、非人灵长类动物或人细胞。例示性的细胞是哺乳动物细胞。哺乳动物细胞可以是原代细胞。一般而言, 可以使用对于双链断裂敏感的任何原代细胞。细胞可以是多种细胞类型, 例如, 成纤维细胞、成肌细胞、T 或 B 细胞、巨噬细胞、上皮细胞等等。

[0130] 当使用哺乳动物细胞系时, 细胞系可以是任何已建立的细胞系或还没描述过的原代细胞系。细胞系可以是附着的或非附着的, 或细胞系可以使用本领域技术人员已知的标准技术, 在促进附着的、附着的或器官型生长的条件下生长。合适的哺乳动物细胞和细胞系的非限制性实例在本文的部分 (IV) (g) 中提供。在仍其它的实施方案中, 细胞可以是干细胞。合适的干细胞的非限制性实例在部分 (IV) (g) 中提供。

[0131] 本公开内容也提供遗传修饰的非人胚胎, 其包含至少一个修饰的染色体序列。染色体序列可以在胚胎中如本文上面所描述 (在描述动物中的染色体序列修饰的段落中) 使用本文描述的方法来修饰。在一个实施方案中, 胚胎是非人受精的目标动物种类的单细胞期胚胎。包括一个细胞胚胎的例示性的哺乳动物胚胎包括但不限于小鼠、大鼠、仓鼠、啮齿动物、兔、猫、犬、羊、猪、牛、马和灵长类动物胚胎。

[0132] 定义

除非另外定义, 本文使用的全部技术和科学术语具有此发明所属领域的技术人员通常理解的含义。以下参考文献给技术人员提供了在本发明中使用的许多术语的通用定义: Singleton 等人, *Dictionary of Microbiology and Molecular Biology* (微生物和分子生物学字典) (第2版, 1994); *The Cambridge Dictionary of Science and Technology* (科学和技术剑桥字典) (Walker ed., 1988); *The Glossary of Genetics* (遗传学词汇表), 第5版, R. Rieger 等人 (eds.), Springer Verlag (1991); 和 Hale 与 Marham, *The Harper Collins Dictionary of Biology* (生物学的 Harper Collins 字典) (1991)。本文使用的以下术语具有属于它们的含义, 除非另有说明。

[0133] 当提出本公开内容或其优选实施方案的要素时, 冠词“一个”、“一种”、“该”和“所述”意指有一个或多个要素。术语“包含”、“含有”和“具有”意图为包容性的, 意指除了列举的要素之外可有另外的要素。

[0134] 本文使用的术语“内源序列”是指对细胞而言是天然的染色体序列。

[0135] 本文使用的术语“外源的”是指对于细胞而言是非天然的序列, 或其在细胞的基因组中的天然位置是在不同染色体位置的染色体序列。

[0136] 本文使用的“基因”是指编码基因产物的 DNA 区域 (其包括外显子和内含子), 以及调控该基因产物的产生的全部 DNA 区域 (无论所述调控序列是否与编码和/或转录序列毗连)。因此, 基因包括但不限于限定于启动子序列、终止子、翻译调控序列例如核糖体结合位点和内部核糖体进入位点、增强子、沉默子、绝缘子、边界元件、复制起点、基质附着位点和基因座控制区域。

[0137] 术语“异源的”是指对于目标细胞而言并非内源或天然的实体。例如, 异源的蛋白是指由内源的来源衍生的或最初由内源的来源衍生的蛋白, 例如内源引入的核酸序列。在一些例子中, 异源的蛋白通常不是由目标细胞产生。

[0138] 术语“核酸”和“多核苷酸”是指脱氧核糖核苷酸或核糖核苷酸聚合物,呈线性的或环形的构象的和呈单链或双链形式的。用于本公开内容的目的,这些术语不应理解为关于聚合物的长度受限制。术语可以包含天然核苷酸的已知类似物,以及在碱基、糖和/或磷酸部分(例如,硫代磷酸酯主链)中修饰的核苷酸。一般而言,特定核苷酸的类似物具有相同的碱基配对特异性;即,A的类似物会与T碱基配对。

[0139] 术语“核苷酸”是指脱氧核糖核苷酸或核糖核苷酸。核苷酸可以是标准核苷酸(即,腺苷、鸟苷、胞苷、胸苷和尿苷)或核苷酸类似物。核苷酸类似物是指具有修饰的嘌呤或嘧啶碱基或修饰的核糖部分的核苷酸。核苷酸模拟物可以是天然存在的核苷酸(例如,肌苷)或非天然存在的核苷酸。对核苷酸的糖或碱基部分的修饰的非限制性实例包括添加(或除去)乙酰基基团、氨基基团、羧基基团、羧甲基基团、羟基基团、甲基基团、磷酰基基团和巯基基团,以及用其它原子取代碱基的碳原子和氮原子(例如,7-脱氮嘌呤)。核苷酸类似物也包括双脱氧核苷酸、2'-O-甲基核苷酸、锁定核酸(LNA)、肽核酸(PNA)和吗啉基寡聚核苷酸(morpholinos)。

[0140] 术语“多肽”和“蛋白”可互换地用于指氨基酸残基的聚合物。

[0141] 用于确定核酸和氨基酸序列同一性的技术是本领域已知的。通常,这些技术包括确定基因的mRNA的核苷酸序列和/或确定由此编码的氨基酸序列,并将这些序列与第二核苷酸或氨基酸序列比较。基因组序列也可以按这种方法确定和比较。一般而言,同一性是指两个多核苷酸或多肽序列各自精确的核苷酸与核苷酸或氨基酸与氨基酸一致性。两个或更多个序列(多核苷酸或氨基酸)可以通过确定它们的百分比同一性而比较。两个序列(无论核酸或氨基酸序列)的百分比同一性是两个比对的序列之间的精确匹配的数量除以较短序列的长度并乘以100。核酸序列的大概比对由Smith和Waterman, *Advances in Applied Mathematics* (应用数学进展) 2:482-489 (1981) 的局部同源性算法提供。此算法可通过使用由Dayhoff, *Atlas of Protein Sequences and Structure* (蛋白质序列和结构的图谱), M.O. Dayhoff ed., 5suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, 开发的并由Gribskov, *Nucl. Acids Res.* 14 (6):6745-6763 (1986) 标准化的得分矩阵而应用于氨基酸序列。此算法用于确定序列百分比同一性的例示性实施,由Genetics Computer Group (Madison, Wis.) 提供于“BestFit (最佳匹配)”实用应用中。其它合适的用于计算序列之间的百分比同一性或相似性的程序通常是本领域已知的,例如,另一比对程序是BLAST,其与默认参数一起使用。例如,BLASTN和BLASTP可以使用以下默认参数而使用:遗传密码=标准;过滤=无;链=二者;截止=60;期望=10;矩阵=BLOSUM62;描述=50个序列;排列方法=高分;数据库=非冗余的,GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR。这些程序的详情可以在GenBank网站上找到。

[0142] 因为在不背离本发明范围的情况下可对上述细胞和方法作出各种改变,意图包含在上面的说明书中和下面给出的实施例中的全部情况应理解为例示性的而不是限制性的。

实施例

[0143] 以下实施例说明了本发明的某些方面。

[0144] 实施例1:修饰Cas9基因用于哺乳动物的表达

来自酿脓链球菌菌株MGAS15252 (登录号YP_005388840.1)的Cas9基因用人密码子偏好性最优化以增强它在哺乳动物细胞中的翻译。Cas9基因也通过将核定位信号PKKKRKV (SEQ ID NO:1)添加到C末端而修饰用于将蛋白靶向至哺乳动物细胞的胞核。表1展示了修饰的Cas9氨基酸序列,其中核定位序列加下划线。表2展示了密码子优化的,修饰的Cas9DNA序列。

表 1. 修饰的 Cas9 氨基酸序列

MDKKYSIGLDIGTNSVGWAVITDDYKVPSPKFKVLGNTDRHSIKKNLIGALLFGSGETAE
 ATRLKRTARRRYTRPKNRICYLGEIFSNEMAKVDDSFHRLVESFLVEEDKKHERHPHF
 GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALANMIKFRGHFLIEGDLNPDNS
 DVDKLFQILVQIYNQLFEENPINASRVDAKILSARLSKSRLENLIAQLPGEKRNGLFG
 NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDOYADFLAAKNLSDA

ILLSDIILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQOLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHOIHLGELH
 AILRQEDFYPLKDNREKIEKILTRIPYVVGPLARGNSRFQAWMTRKSEETITPWNFEE
 VVOKGASAOQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKYVTEGMRKPA
 FLSGEGKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDL
 KIKDKDFLDNEENEDILEDIWLTLTFEDRGMIEERLKYAHLFDDKVMKQLKRRRYTG
 WGRLSRKLINGIRDKQSGKTILDFKSDGFANRNFQMQLIHDDSLTFKEDIQKAQVSGOG
 HSLNEQIANLAGSPAIKKGILOTVKIVDELVKVMGHKPKENIVIEMARENQTTQKGOKNSR
 ERMKRIEIEGKELGSOILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYD
 VDHIVPQSFIDKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKMKMKNYWRQLLNAKLITQRK
 FDNLTKAERGGLSELDAKAGFIKRLVETROITKHVAQILDSPMNTKYDENDKLIREVKVI
 TLKSKLVSDFRKDFQFYKVRINNYHHAHDAYLNAVVGTAIIKKYPKLESEFVYGDYKV
 YDVRKMIKSEQIEGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
 GRDFATVRKVLSPQVNIKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
 DSPTVAYSVLVAKVEKGKSKLKSVELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
 IIKLPKYSLEFENGRIKRLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNE
 QKQLFVEQHKHYLDEIEQISEFSKRVLADANLDKVL SAYNKHRRDKPIREQAENIHLFTL
 TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDPKKRKV
 (SEQ ID NO:9)

表 2. 优化的 Cas9 DNA 序列(5'-3')

ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGG
 CCGTGATCACCAGCACTACAAGGTGCCAGCAAGAAATCAAGGTGCTGGGCAA
 CACCGACCGGCACAGCATCAAGAAGAACCTGATCGGCGCCCTGCTGTTCCGCTCT
 GGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
 AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCA
 AGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCTGGTGGGAAGAGGA
 TAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTAC
 CACGAGAAGTACCCACCATCTACCACCTGAGAAAGAAGCTGGCCGACAGCACCG
 ACAAGGCCGACCTGAGACTGATCTACCTGGCCCTGGCCACATGATCAAGTTCGG
 GGGCCACTTCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAA
 GCTGTTTCATCCAGCTGGTGCAGATCTACAATCAGCTGTTTCGAGGAAAACCCATCA
 ACGCCAGCAGAGTGGACGCCAAGGCCATCCTGAGCGCCAGACTGAGCAAGAGCA
 GACGGCTGGAAAATCTGATCGCCAGCTGCCCGGCGAGAAAGCGGAATGGCCTGT
 TCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCAACTTCAAGAGCAACTT
 CGACCTGGCCGAGGATGCCAACTGCAGCTGAGCAAGGACACCTACGACGACGA
 CCTGGACAACCTGCTGGCCAGATCGGCGACAGTACGCCGACCTGTTTCTGGCC
 GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACAGCG
 AGATCACCAAGGCCCCCTGTCCGCCTCTATGATCAAGAGATACGACGAGCACCA
 CCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTAC
 AAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGCTACATCGATGGCG
 GAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGA
 CGGCACCGAGGAAGTCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCA
 GCGGACCTTCGACAACCGCAGCATCCCCACCAAGATCCACCTGGGAGAGCTGCA


```

CGCCATTCTGCGGCGGCGAGGAAGATTTTTACCCATTCTGAAGGACAACCGGGAA
AAGATCGAGAAGATCCTGACCTTCAGAATCCCCTACTACGTGGGCCCTCTGGCCA
GGGGAACAGCAGATTTCGCTGGATGACCAGAAAGAGCGAGGAAACCATCACCC
CCTGGAACCTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCAGAGCTTCATCG
AGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATAC
GTGACCGAGGGGAATGCGGAAGCCCCGCTTTCTGAGCGGCGAGCAGAAAAAGGCC
ATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAAG
AGGACTACTTCAAGAAAATCGAGTSCCTTCGACAGCGTGGAAATCAGCGGCGTGA
AGATCGGTTCAACGCCTCCCTGGGCGCCTATCAGCATCTGCTGAAAATTATCAAGG
ACAAGGACTTCCTGGACAATGAGGAAAAACGAGGACATTCTGGAAGATATCGTGCT
GACCTTGACACTGTTTGAGGACCGGGGCATGATCGAGGAACGGCTGAAAACTAT
GCCACCTGTTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACC
GGCTGGGCGAGGCTGAGCCGGAAGCTGATCAACGSCATCCGGGACAAGCAGTCC
GGCAAGACAATCCTGGATTTCTGAAGTCCGACGGCTTCGCCAACAGAACTTCAT
GCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCAG
GTGTCCGGCCAGGGACACTCTCTGCACGAGCAGATCGCCAATCTGGCCGGATCC
CCCGCCATTAAAGAGGGCATCCTGCAGACAGTGAAGATTGTGGACGAGCTCGTGA
AAGTGATGGGCCACAAGCCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGA
GGGCATCAAAGAGCTGGGCGAGCAGATCTGAAAGAACACCCCGTGGAAAAACACC
CAGCTGCAGAACGAGAAGCTGTACTGTACTGACCTGCAGAAATGGGCGGGATGT
ACGTGGACCAAGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCACAT
TGTGCCCCAGTCTTCATCAAGGACGACTCCATCGATAACAAAGTGCTGACTCGGA
GCGACAAGAACCAGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGTCTGTGAAGA
AGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACCCAGAGGAA
GTTTCGACAATCTGACCAAGGCCGAGAGAGGGCGGCTGAGCGAACTGGATAAGGC
CGGCTTCATTAAGCGGCAGCTGGTGGAACCCGGCAGATCACAAAGCACGTGBCA
CAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAACTGATCC
GGGAAGTGAAAGTGATCACCCCTGAAGTCCAAGCTGGTGTCCGACTTCAGAAAGGA
TTTCCAGTTTTACAAAGTGCGCGAGATCAACAACCTACCACCACGCCACGACGCT
ACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAG
CGAGTTCTGTGTACGCGGATTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCA
TGAACTTTTTCAAGACCGAGATCACACTGGCCAACGGCGAGATCAGAAAGCGGCC
TCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGA
TTTTGCCACAGTGCGGAAAGTGCTGTCCATGCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACCGGCGGCTTCAGCAAAAGAGTCTATCCTGCCCAAGAGGAACTC
CGACAAGCTGATCGCCAGAAAGAAGGATTGGGACCCCTAAGAAGTACGGCGGCTTT
GACAGCCCCACCGTGGCCTACTCTGTGCTGGTGGTGGCCAAAGTGGAAGAGGGC
AAGTCCAAGAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAA
GAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGA
AGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTTCGAGCTGGAAA
ACGGCCGGAAGCGGATGCTGGCTTCTGCCGGCGAACTGCAGAAGGGAAACGAGC
TGGCCCTGCCCTCCAAATATGTGAACCTTCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACA
AGCACTACCTGGACGAGATCATCGAGCAGATTAGCGAGTTCTCCAAGCGCGTGAT
CCTGGCCGATGCCAACCTGGACAAGGTGCTGAGCGCCTACAACAAGCACCCGGGA
TAAGCCCATCAGAGAGCAGGCGGAGAATATCATCCACCTGTTTACCCTGACCAACC
TGGGAGCCCCCTGCCGCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTA
CACCAGCACCAAGAGGTGCTGGACGCCACCCCTGATCCACCAGAGCATCACCGG
CCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACCCCAAGAAAAAG
CGCAAAGTG (SEQ ID NO:10)

```

[0145] 将修饰的Cas9DNA序列置于巨细胞病毒(CMV)启动子的控制下用于在哺乳动物细胞中的组成性表达。修饰的Cas9DNA序列也被置于T7启动子的控制下用于用T7RNA聚合酶的体外mRNA合成。体外RNA转录通过使用MessageMAX T7ARCA-Capped Message Transcription Kit和T7mScript Standard mRNA Production System(Cellscript)进行。

[0146] 实施例2:靶向Cas9

腺伴随病毒整合位点1(AAVS1)基因座被用作Cas9-介导的人基因组修饰的靶。人AAVS1

基因座位于蛋白磷酸酶1,调控亚基12C (PPP1R12C) 的内含子1 (4427bp)。表3展示了 PPP1R12C 的第一外显子(阴影灰色)和第一内含子。在内含子中的加下划线的>序列是靶修饰位点(即,AAVS1基因座)。

表 3. PPP1R12C 的第一外显子和内含子(5'-3')

GGGGGCGGGCGGTGCGATGTCCGAGAGGATGGCCCGCGGGCTGGCCCGGGG
GCGGCGGGCGGGCTGCCCGGAGCGGCGACGGGAGCAGCTGCGGCAGTGGG
GGGCGCGGGCGGGCGCCGAGCCTGGCCCGGAGAGCGCCCGCGCCCGCACCGT
CCGCTTCGAGCGCGCGCGCGAGTTCTGGCGGGCTGTGCGGGCGGGCGACCTGG
ACGAGGCGCGTCTGATGCTGCGCGCGCGCGACCTGGCCCGGGCGCGGAGGCTC
GACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGGAGTCCACCAACGCGGA
CGGTATCAGCGCCCTGCACCAAGGTGACGCGCCCCCGCGCGGGCTCTCCCGGGG
CAGGTCCACCTCTGCTGCGCCACCTGGGGCATCCTCCTTCCCCGTTGCCAGTCT
CGATCGCGCGCGTGGTCTGGCGCTGGGCTTTGCCACCTATGCTGACACCCCG
TCCAGTCCCCCTTACCATTCCCCCTTCGACCACCCCACTTCCGAATTGGAGCCGT
TCAACTGGCCCTGGGCTTAGCCACTCTGTGCTGACCACTCTGCCCCAGGCCTCT
TACCATCCCCCTTCGACCTACTCTCTTCCGCAATTGGAGTGGCTTTAACTGGCCCTG
GCTTTGGCAGCCTGTGCTGACCCATGCAGTCTCCTTACCATCCCTCCCTCGACTT

CCCTCTTCCGATGTTGAGCCDCTCCAGCCGGTCTGGACTTTGTCTCCTTCCCTG
CCCTGCCCTCTCCTGAACCTGAGCCAGCTCCCATAGCTCAGTCTGGTCTATCTGCC
TGGCCCTGGCCATTGTCACTTTGCGCTGCCCTCCTCTGGCCCGGAGTGCCCTTG
CTGTGCCGCGCGGAACCTCGCCCTCTAACGCTGCCGTCTCTCTCCTGAGTCCGGAC
CACTTTGAGCTCTACTGGCTTCTGCGCGCGCTCTGGCCCACTGTTTCCCCCTTCCCA
GGCAGGTCTGCTTTCTCTGACCTGCATTCTCTCCCTGGGCTGTGCCGCTTTCT
GTCTGCAGCTTGTGGCCTGGGTGACCTCTACGGCTGGCCCGAGATCCTTCCCTGCC
GCCTCCTTCAGGTTCCGTCTTCTCCTCACTCCCTCTTCCCTTGTCTCTGCTGTGT
TGCTGCCCAAGGATGCTCTTTCGGGAGCACTTCTTCTCGGCGCTGCACCAAGT
ATGTCCTCTGAGCGGATCCTCCCGTGTCTGGGTCTCTCCGGGCATCTCTCCTC
CCTCACCCCAAGCCCATGCCGTCTTCACTCGCTGGGTTCCTTTTCTCTCTCTCT
GGGGCCTGTGCCATCTCTCGTTTCTTAGGATGGCCTTCTCCGACGGATGTCTCCCT
TGGCTCCCGCCTCCCTTCTTGTAGGCCTGCATCATCACCGTTTCTTGGACAACC
CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCCTCTTGGCTTCTTTG
CCTGGACACCCCGTTCTCCTGTGGATTGGGTGACCTCTCACTCCTTTCACTTTGGG
CAGTCCCGTACCCCGCTTACCTCTGAGTCTGTGCTAGCTCTTCCAGCCCGCTGT
CATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCTCCCAACCCGGG
CCCCTATGTCCACTTCAGGACAGCATGTTTGCTGCCCTCCAGGGATCCTGTGTCCC
GAGCTGGGACCACTTATATCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGT
ACTTTTATCTGTCCCCTCCACCCCAAGTGGGGCCACTAGGGACAGGATTGGTGA
CAGAAAAGCCCCATCCTTAGGCCTCCTCCTTCTAGTCTCCTGATATTGGGTCTAA
CCCCACCTCCTGTTAGGCAGATTCTTATCTGGTGACACACCCCATTTCTGGA
GCCATCTCTCCTTCCAGAACTCTAAGGTTTGCTTACGATGGAGCCAGAGAGG
ATCCTGGGAGGGAGAGCTTGGCAGGGGGTGGGAGGGAAGGGGGGGATGCGTGA
CCTGCCCGGTTCTCAGTGGCCACCCCTGCGCTACCCCTCTCCAGAACCTGAGCTGC
TCTGACGCGGGCGTCTGGTGGCTTCACTGATCCTGGTGGTGCAGCTTCTTACA
CTTCCCAAGAGGAGAAGCAGTTTGGAAAAACAAATCAGAATAAGTTGGTCTGAG
TTCTAACTTTGGCTCTTCACTTTCTAGTCCCCAATTTATATTGTTCTCCGTGCGT
CAGTTTACCTGTGAGATAAGGCCAGTAGCCAGCCCGTCTTGGCAGGGGTGTGG
TGAGGAGGGGGGTGTCCGTGTGGAAACTCCCTTTGTGAGAATGGTGGTCTTAG
GTGTTACCAAGGTGCTGGCCGCTCTACTCCCTTTCTCTTCTCCATCCTTCTTTCC
TTAAAGAGTCCCCAGTGCTATCTGGGACATATTCTCCGCGCCAGAGCAGGGTCCC
GCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGTTTGGTCTTGGCAAGCCCAAGG
AGAGGCGCTCAGGCTTCCCTGTCCCCCTTCTCCTGTCACCATCTCATGCCCTGG
CTCTCCTGCCCTTCCCTACAGGGGTTCTGGCTCTGCTCTTCACTGAGCCCG
GTTCCCTGCATCCCGTTCCCTGCATCCCGCTTCCCTGCATCCCCCAGAGGC
CCCAGGCCACCTACTTGGCTGGACCCACAGAGAGGCCACCCAGCCCTGTCTA
CCAGGCTGCCCTTTGGGTGGATTCTCCTCCAAGTGTGGGGTGAAGTGTGGCAAA
CTCACTCTTCCGGGTATCCAGGAGGCCTGGAGCATTGGGGTGGGCTGGGGTTC
AGAGAGGAGGGATTCCCTTCTCAGGTTACGTGGCCAAGAAGCAGGGGAGCTGG
TTTGGGTGAGTCTGGGTGTGGGTGACCACTTATGCTGTTTGGCCAGGACAGC
CTAGTTTACGACTGAAACCTCAGTCTAGGAAAAACAGGGATGGTTGGTCACTGT
CTCTGGGTGACTCTTGATTCCCGGCCAGTTTCTCCACCTGGGGCTGTGTTTCTGT
CCTGCATCCTTCTCCAGGCAGGTCCCAAGCATCGCCCCCTGCTGTGGCTGTTC
CCAAGTTCTTAGGGTACCCCAAGTGGGTTTATCAACCACTTGGTGAGGCTGGTACC
CTGCCCCCATCTGACCCCAATTGCCTTAGTGGCTAGGGGGTGGGGGCTAGA


```

GTAGGAGGGGCTGGAGCCAGGATTCTTAGGGCTGAACAGAGAAGAGCTGGGGGC
CTGGGCTCCTGGGTTTGAGAGAGGAGGGGCTGGGGCCTGGACTCCTGGGTCCGA
GGGAGGAGGGGCTGGGGCCTGGACTCCTGGGTCTGAGGGTGGAGGGACTGGGG
GCTGGACTCCTGGGTCCGAGGGAGGAGGGGCTGGGGCCTGGACTCGTGGGTCT
TGAGGGAGGAGGGGCTGGGGGCTGGACTTCTGGGTCTTAGGGAGGCGGGGCT
GGGGCCTGGACCCCTGGGTCTGAATGGGAGAGGCTGGGGGCTGGACTCCTTCA
TCTGAGGGCGGAAGGGCTGGGGCTGGGCTCCTGGGTGAATGGGAGGGGTTG
GGCCTGGACTCTGGAGTCCCTGGTGCCGAGGCTCAGGCATCTTTCACAGGGATG
CCTGTACTGGGCAGGTCTTGAAAGGAAAGGCCCATTTGCTCTCCTTGCCCCCT
CCCTATCGCCATGACAACCTGGGTGGAAATAAACGAGCCGAGTTCATCCGTTCC
CAGGGCACGTGCGGCCCTTACAGGCCGAGTTTCCATGACCTCATGCTCTTGGC
CCTCGTAGCTCCCTCCCGCTCCTCCAGATGGGCAGCTTTGGAGAGGTGAGGGAC
TTGGGGGGTAATTTATCCCGTGGATCTAGGAGTTTAGCTTCACTCCTTCCTCAGCT
CCAGTTCAGGTCCCGAGGCCACCCAGTGTCCACAAGGCCTGGGGCAAGTCCCT
CCTCCGACCCCTGGACTTCGGCTTTTGTCCCCCAAGTTTGGACCCCTAAGGG
AAGAATGAGAAACGGTGGCCCGTGTGAGCCCTGGCTGCAGGGCCCCGTGCAGA
GGGGGCTCAGTGAACCTGGAGTGTGACAGCCTGGGGCCAGGCACACAGGTGTG
CAGCTGTCTCACCCCTCTGGGAGTCCCGCCAGGCCCTGAGTCTGTCCAGCAC
AGGGTGGCCTTCTCCACCTGTCATAGCCCTGGGGCCACGGCTTCGTTCTCGAG
AGTATCTGCTGGGGTGGTTTCCGAGCTTGACCCTTGAAGGACCTGGCTGGGTTT
AAGGCAGGAGGGGCTGGGGGCCAGGACTCCTGGCTCTGAAGGAGGAGGGGCTG
GAACCTCTTCCCTAGTCTGAGCACTGGAAGCGCCACCTGTGGGTGGTGACGGG
GTTTTGCCGTGTCTAACAGGTACCATGTGGGGTTCCCGCACCCAGATGAGAAGCC
CCCTCCCTTCCCGTTCACTTCTGTTGCAGATAGCCAGGAGTCTTTCGTGGTT
TCCACTGAGCACTGAAGGCCTGGCCGGCCTGACCACTGGGCAACCAGGCGTATC
TTAAACAGCCAGTGGCCAGAGGCTGTTGGGTCATTTTCCCACTGTCTAGCACC
GTGTCCCTGGATCTGTTTTCGTGGCTCCCTCTGGAGTCCCGACTTGCTGGGACAC
CGTGGCTGGGGTAGGTGCGGCTGACGGCTGTTTCCACCCCCAG (SEQ ID
NO:11)

```

[0147] Cas9指导RNA被设计用于靶向人AAVS1基因座。制备包含(5'-3')靶识别序列(即,与靶序列的非编码链互补的序列)和原间隔序列的42个核苷酸RNA(在本文中被成为“crRNA”序列);包含与crRNA的3'序列互补的5'序列和另外的发夹序列的85个核苷酸RNA(本文称为“tracrRNA”序列);和包含所述crRNA的核苷酸1-32、GAAA环和所述tracrRNA的核苷酸19-45的嵌合RNA。crRNA由Sigma-Aldrich化学合成。tracrRNA和嵌合RNA通过使用T7-Scribe Standard RNA IVT Kit(Cellscript)用T7RNA聚合酶体外转录而合成。嵌合RNA编码序列也被置于人U6启动子的控制下用于在人细胞中的体内转录。表4展示了指导RNA的序列。

表 4. 指导 RNA		
RNA	5'-3' 序列	SEQ ID NO.
AAVS1-crRNA	ACCCACAGUGGGGCCACUAGUUUUAGAGCUAUGCUGU UUUG	12
tracrRNA	GGAACCAUUCAAAACAGCAUAGCAAGUUAUUAAAGGCU AGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGU GCUUUUUUU	13
嵌合 RNA	ACCCACAGUGGGGCCACUAGUUUUAGAGCUAGAAUA GCAAGUUAUUAAAGGCUAGUCCG	14

[0148] 实施例3:制备供体多核苷酸以监测基因组修饰

将GFP蛋白靶向整合入PPP1R12C的N末端,用于监测Cas9-介导的基因组修饰。为通过同源重组介导整合,制备供体多核苷酸。AAVS1-GFP DNA供体包含5' (1185bp) AAVS1基因座同源臂、RNA剪接受体、turbo GFP编码序列、3' 转录终止子和3' (1217bp) AAVS1基因座同源臂。表5展示RNA剪接受体的序列和GFP编码序列,接着是3' 转录终止子。质粒DNA通过使用GenElute Endotoxin-Free Plasmid Maxiprep Kit(Sigma)而制备。

表 5. AAVSI-GFP DNA 供体序列中的序列		
	5'-3' 序列	SEQ ID NO.
RNA 剪接受体	CTGACCTCTTCTCTTCCCTCCACAG	15
GFP 编码序列和转录终止子	GCCACCATGGACTACAAAGACGATGACGACAAGGTCGACTCTAGAGCTGCAGAGAGCGACGAGAGCGGCCTGCCCGCCA TGGAGATCGAGTGCCGCATCACCGGCACCTGAACGGCGTGGAGTTCGAGCTGGTGGGCGGCGGAGAGGGCACCCCCG AGCAGGGCCGCATGACCAACAAGATGAAGAGCACCAAAAGGCGCCCTGACDTCAGCCCCCTACCTGCTGAGCCACGTGATG GGCTACGGCTTCTACCACTTCGGCACCTACCCAGCGGCT	16
	ACGAGAACCCCTTCCTGCAAGCCATCAACAACGGCGGCTACACCAACACCCGCATCGAGAAGTACGAGGACGGCGGCGT GCTGCAAGTGAAGCTTCAGCTACCGCTACGAGGGCGGCGCGCGTGATCGGCGACTTCAAGGTGATGGGCACCGGCTTCCCG GAGGACAGCGTGATCTTCACCGACAAGATCGTCCGCAGCAACGCCACCGTGGAGCACCTGCACCCCATGGGCGATAACG ATCTGGATGGCAGCTTCACCGCACCTTCAGCCTGCGCGACGGCGGCTACTACAGCTCCGTGGTGGACAGCCACATGCAC TTCAAGAGCGCCATCCACCCAGCATCCTGCAGAACGGGGGCCCCATGTTCCGCTTCCGCGCGTGGAGGAGGATCACA GCAACACCGAGCTGGGCATCGTGGAGTACCAGCACGCCCTTCAAGACCCCGGATGCAGATGCCGGTGAAGAATGAAGATCT CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCGCGTGCCTTCCCTTGACCCTGGAAGGTGCCACTCCACTG TCCTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCCGTGGGCTCTATGGACTCGAGGTTTAAACG TCGACGCGGCCGCGT	

[0149] 靶向的基因整合会产生PPP1R12C的前107个氨基酸和turbo GFP之间的融合蛋白。预期的融合蛋白含有PPP1R12C的前107个氨基酸残基(以灰色突出显示),其来自PPP1R12C的第一外显子和工程改造的剪接受体之间的RNA剪接(参见表6)。

表 6. PPP1R12C-GFP 融合蛋白的预测的氨基酸序列
MSGEDGPAAGPGAAAAAARERRRREQLRQWGARAGAEPGPGERRARTVRFERAAEF LAACAGGDLDEARLMLRAADPGPGAELDPAAPPPARAVALDSTNADGISALHQATMDY KDDDDKVDSRAAESDESGLFAMEIECRITGTLNGVEFELVGGGEGTPEQGRMTNKMK STKGALTFSPYLLSHVMGYGFYHFGTYPSTGYENPFLHAINNGGYTNTRIEKYEDGGVL HVSFSYRYEAGRVIGDFKVMGTGFPEDSVIFTDKIVRSNATVEHLHPMGDNDLDGSFT RTFSLRDGGYYSSVVDSHMHFKSAIHPSILQNGGPMFAFRVEEDHSNTELGIVEYQH AFKTPDADAGEE (SEQ ID NO:17)

[0150] 实施例4: Cas9-介导的靶向整合

转染在人K562细胞上实施。K562细胞系从美国模式培养物保藏所(ATCC)获得并且在使用10%FBS和2mM L-谷氨酰胺补充的Iscove's改良的Dulbecco's培养基中生长。全部培养基和补充物都从Sigma-Aldrich获得。在转染前一天将培养物分开(以在转染前约50万个细胞每毫升)。将细胞用Nucleofector Solution V(Lonza)在Nucleofector(Lonza)上使用T-016程序来转染。各核转染(nucleofection)含有大约60万个细胞。转染处理详述在表7中。细胞在核转染之后立即于37℃和5%CO₂下生长。

表 7.转染处理

处理	修饰的 Cas9	指导 RNA	供体序列
A	用 Anti-Reverse 帽类似物转录的 Cas9 mRNA (10 μ g)	预退火的 crRNA-tracrRNA 双链体(0.3 nmol)	AAVS1-GFP 质粒 DNA (10 μ g)
B	用 Anti-Reverse 帽类似物转录的 Cas9 mRNA (10 μ g)	嵌合 RNA(0.3 nmol)	AAVS1-GFP 质粒 DNA (10 μ g)
C	通过转录后加帽反应加帽的 Cas9 mRNA (10 μ g)	嵌合 RNA (0.3 nmol)	AAVS1-GFP 质粒 DNA (10 μ g)
D	Cas9 质粒 DNA (10 μ g)	U6-嵌合 RNA 质粒 DNA (5 μ g)	AAVS1-GFP 质粒 DNA (10 μ g)
E	无	无	AAVS1-GFP 质粒 DNA (10 μ g)
F	无	无	无

[0151] 荧光激活细胞分选术 (FACS) 在转染之后4天进行。FACS数据展示于图4中。在4个实验处理 (A-D) 的每一个中检测到的GFP百分比大于对照处理 (E,F), 这证实供体序列的整合和融合蛋白的表达。

[0152] 实施例5:PCR证实靶向整合

在转染后12天用GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) 从转染的细胞中提取基因组DNA。随后用定位于AAVS1-GFP质粒供体的5' 同源臂外部的正向引物和定位于 GFP 的 5' 区域的反向引物经 PCR 扩增基因组 DNA。正向引物为 5' -CCACTCTGTGCTGACCACTCT-3' (SEQ ID NO:18) 和反向引物为 5' -GCGGCACTCGATCTCCA-3' (SEQ ID NO:19)。自连接PCR的预期的片段大小是1388bp。扩增用JumpStart Taq ReadyMix (Sigma), 使用以下循环条件实施:98℃维持2分钟用于最初的变性;35个循环:98℃维持15秒,62℃维持30秒和72℃维持1分钟30秒;和最终延伸:72℃维持5分钟。PCR产物在琼脂糖凝胶上分离。

[0153] 采用10 μ g用Anti-Reverse帽类似物转录的Cas9mRNA、0.3nmol预退火的crRNA-tracrRNA双链体和10 μ g AAVS1-GFP质粒DNA转染的细胞显示预期大小的PCR产物 (参见泳道 A,图5)。

[0154] 实施例6:在小鼠胚胎中基于Cas9的基因组编辑

可靶向小鼠Rosa26基因座用于基因组修饰。表8展示了小鼠Rosa26序列的一部分,其中潜在的靶位点以粗体显示。每个靶位点包含原间隔基。

表 8.小鼠 Rosa26 序列

GAGCGGCTGCGGGGGCGGGTGCAGGCACGTTTCCGACTTGAGTTGCCTCAAGAGG
GGCGTGCTGAGCCAGACCTCCATCGCGCACTCCGGGGAGTGGAGGGAAGGAGC
GAGGGCTCAGTTGGGCTGTTTTGGAGGCAGGAAGCACTTGCTCTCCAAAGTCGC
TCTGAGTTGTTATCAGTAAGGGAGCTGCAGTGGAGTAGGCGGGGAGAAGGCCGC
ACCTTCTCCGGAGGGGGGAGGGGAGTGTGCAATACCTTCTGGGAGTTCTCTG
CTGCCTCCTGGCTTCTGAGGACCGCCCTGGGCCTGGGAGAATCCCTTCCCCCTCT
TCCCTCGTGATCTGCAACTCCAGTCTTTCTAGAAGATGGGCGGGAGTCTTCTGGG
CAGGCTTAAAGGCTAACCTGGTGTGTGGCGTTGTCCTGCAGGGGAATTGAACAG
GTGTAAAATTGGAGGGACAAGACTTCCACAGATTTTCGGTTTTGTGGGAAGTTT
TTTAATAGGGGCAAATAAGGAAAATGGGAGGATAGGTAGTCATCTGGGGTTTTATG
CAGCAAACTACAGGTTATTATTGCTTGTGATCCGCCCTCGAGTATTTCCATCGA
GGTAGATTAAAGACATGCTCACCCGAGTTTATACTCTCCTGCTTGAGATCCTTACT
ACAGTATGAAATTACAGTGTGCGGAGTTAGACTATGTAAGCAGAATTTTA (SEQ ID
NO:20)

[0155] 指导RNA被设计以靶向在小鼠Rosa26基因座中的各个靶位点。序列示于表9,每一个长度为42个核苷酸并且5'区域与没有展示于表8中的链(即,与表8中示出的链互补的链)互补。

表 9.小鼠 Rosa26 指导 RNA		
RNA	5'-3'序列	SEQ ID NO:
mRosa26-crRNA-1	CUCCAGUCUUUCUAGAAGAUGUUUUUAGAGCUAU GCUGUUUUG	21
mRosa26-crRNA-2	UGAACAGGUGUAAAAUUGGAGUUUUAGAGCUAU GCUGUUUUG	22
mRosa26-crRNA-3	UGUCGGAAGUUUUUAAUAGUUUUUAGAGCUAU GCUGUUUUG	23

[0156] 化学合成crRNA,并与tracrRNA (SEQ ID NO:13;参见实施例2)预退火。可将预退火的crRNA/tracrRNA和体外转录的编码修饰的Cas9蛋白 (SEQ ID NO:9;参见实施例1)的mRNA显微注射入受精的小鼠胚胎的前核。当通过crRNA指导至靶位点时,Cas9蛋白剪切靶位点,并且生成的双链断裂可以通过非同源末端连接(NHEJ)修复过程而修复。经注射的胚胎可以在37℃,5%CO₂下孵育过夜或孵育多至4天,接着基因分型分析,或者经注射的胚胎可以移植入受者雌性小鼠使得活产动物可以被基因分型。来自活产动物的体外孵育的胚胎或组织可以使用标准方法针对在Rosa基因座的Cas9诱导突变的存在进行筛选。例如,可收获来自胎儿或活产动物的胚胎或组织用于DNA提取和分析。DNA可以使用标准的程序分离。Rosa26基因座的靶向区域可以使用合适的引物经PCR扩增。因为NHEJ是易错的,所以在断裂的修复期间可以发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。突变可以使用基于PCR的基因分型方法(例如Ce1-I错配测定和DNA测序)检测。

[0157] 实施例7:在小鼠胚胎中基于Cas9的基因组修饰

Rosa26基因座可以通过将如在上面部分(IV)(d)中详述的供体多核苷酸与如在上面实施例6中描述的预退火的crRNA/tracrRNA和编码修饰的Cas9的mRNA一起共注射而在小鼠胚胎中被修饰。来自活产动物(如在实施例6中所描述)的体外孵育的胚胎或组织可以针对修饰的Rosa26基因座使用基于PCR的基因分型方法(例如RFLP测定、连接PCR和DNA测序)筛选。

[0158] 实施例8:在大鼠胚胎中基于Cas9的基因组编辑

可靶向大鼠Rosa26基因座用于基因组修饰。表10展示了大鼠序列的一部分,其中潜在的靶位点以粗体示出。每个靶位点包含原间隔基。

表 10. 大鼠 Rosa26 序列

```

GGGATTCCCTCCTTGAGTTGTGGCACTGAGGAACGTGCTGAACAAGACCTACATTG
CACTCCAGGGAGTGGATGAAGGAGTTGGGGCTCAGTCGGGTTGTATTGGAGACAA
GAAGCACTTGCTCTCCAAAAGTCGGTTTGAGTTATCATTAAAGGGAGCTGCAGTGG
GTAGGCGGAGAAAAGSCCGCACCCCTTCTCAGGACGGGGGAGGGGAGTGTTCGAA
TACCTTTCTGGGAGTTCTCTGCTGCCCTCCTGTCTTCTGAGGACCGCCCTGGGCCT
GGAAGATTCCCTTCCCTTCTTCCCTCGTGATCTGCAACTGGAGTCTTTCTGGAA
GATAGGCGGGAGTCTTCTGGGCAGGCTTAAAGGCTAACCTGGTGCGTGGGGCGT
TGTCTGCAGAGGAATTGAACAGGTGTAAAATTGGAGGGGCAAGACTTCCACAG
ATTTTCGATTGTGTTGTTAAGTATTGTAATAGGGGCAAATAAGGGAAATAGACTAGG
CACTCACCTGGGGTTTTATGCAGCAAACTACAGGTTATTATTGCTTGTGATCCGC
CCTGGAGAATTTTTCACCGAGGTAGATTGAAGACATGCCACCCAAATTTTAATATT
CTTCCACTTGCGATCCTTGCTACAGTATGAAA (SEQ ID NO:24)

```

[0159] 指导RNA被设计以靶向大鼠Rosa26基因座中的靶位点。序列示出于表11中,每一个长度为42个核苷酸且5'区域与在表10中没有呈现的链(即,与示出于表10中的链互补的链)互补。

表 11. 大鼠 Rosa26 指导 RNA

RNA	5'-3' 序列	SEQ ID NO.
rRosa26-crRNA-1	AGGGGGAAGGGAAUCUCCAGUUUUAGAGCUA UGCUGUUUUG	25
rRosa26-crRNA-2	UCUGCAACUGGAGUCUUUCUGUUUUAGAGCUA UGCUGUUUUG	26
rRosa26-crRNA-3	AGGCGGGAGUCUUCUGGGCAGUUUUAGAGCUA UGCUGUUUUG	27

[0160] 化学合成crRNA,并与tracrRNA (SEQ ID NO:13;参见实施例2)预退火。可将预退火的crRNA/tracrRNA和体外转录的编码修饰的Cas9蛋白的mRNA (SEQ ID NO.9;参见实施例1)显微注射入受精的大鼠胚胎的前核。在由crRNA指导至靶位点时,Cas9蛋白剪切靶位点,并且生成的双链断裂可以通过非同源末端连接 (NHEJ) 修复过程修复。注射的胚胎可以在37℃,5%CO₂下孵育过夜或孵育多至4天,接着基因分型分析,或者注射的胚胎可被移植入受者雌性小鼠使得活产动物可以被基因分型。来自活产动物的体外孵育的胚胎或组织可以使用标准方法针对在Rosa基因座的Cas9诱导突变的存在进行筛选。例如,可收获来自胎儿活产动物的胚胎或组织用于DNA提取和分析。DNA可以使用标准程序分离。Rosa26基因座的靶向区域可以使用合适的引物经PCR扩增。因为NHEJ是易错的,所以在断裂的修复期间可以发生至少一个核苷酸的缺失、至少一个核苷酸的插入、至少一个核苷酸的置换或其组合。可以使用基于PCR的基因分型方法,例如Cel-I错配测定和DNA测序检测突变。

[0161] 实施例9:在大鼠胚胎中基于Cas9的基因组修饰

Rosa26基因座可以通过将如在上面部分 (IV) (d) 中详述的供体多核苷酸与如在上面实施例8中所描述的预退火的crRNA/tracrRNA和编码修饰Cas9的mRNA一起共注射而在大鼠胚胎中被修饰。来自活产大鼠的体外孵育的胚胎或组织(如在实施例8中所描述)可以针对修饰的Rosa26基因座使用基于PCR的基因分型方法(例如RFLP测定、连接PCR和DNA测序)进行筛选。

- <110> SIGMA-ALDRICH CO. LLC
CHEN, Fuqiang
DAVIS, Gregory D.
KANG, Qiaohua
KNIGHT, Scott W.
- <120> 基于CRISPR的基因组修饰和调控
- <130> 047497-465606
- <150> US 61/734,256
- <151> 2012-12-06
- <150> US 61/758,624
- <151> 2013-01-30
- <150> US 61/761,046
- <151> 2013-02-05
- <150> US 61/794,422
- <151> 2013-03-15
- <160> 27
- <170> PatentIn version 3.5
- <210> 1
- <211> 7
- <212> PRT
- <213> 人工序列
- <220>
- <223> 合成的
- <400> 1
- Pro Lys Lys Lys Arg Lys Val
- 1 5
- <210> 2
- <211> 7
- <212> PRT
- <213> 人工序列
- <220>
- <223> 合成的
- <400> 2
- Pro Lys Lys Lys Arg Arg Val
- 1 5
- <210> 3
- <211> 16
- <212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 3

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 4

<211> 20

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 4

Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Pro Lys Lys

1 5 10 15

Lys Arg Lys Val

20

<210> 5

<211> 19

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 5

Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys Lys

1 5 10 15

Arg Lys Val

<210> 6

<211> 24

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 6

Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly

1 5 10 15

Ala Pro Lys Lys Lys Arg Lys Val

20

<210> 7

<211> 27

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 7

Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly

1 5 10 15

Ala Trp Ser Gln Pro Lys Lys Lys Arg Lys Val

20 25

<210> 8

<211> 21

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 8

Lys Glu Thr Trp Trp Glu Thr Trp Trp Thr Glu Trp Ser Gln Pro Lys

1 5 10 15

Lys Lys Arg Lys Val

20

<210> 9

<211> 1374

<212> PRT

<213> 人工序列

<220>

<223> 合成的

<400> 9

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser

85	90	95
Phe Phe His Arg Leu Glu Glu Ser	Phe Leu Val Glu Glu Asp Lys Lys	
100	105	110
His Glu Arg His Pro Ile Phe Gly Asn	Ile Val Asp Glu Val Ala Tyr	
115	120	125
His Glu Lys Tyr Pro Thr Ile Tyr His	Leu Arg Lys Lys Leu Ala Asp	
130	135	140
Ser Thr Asp Lys Ala Asp Leu Arg Leu	Ile Tyr Leu Ala Leu Ala His	
145	150	155
Met Ile Lys Phe Arg Gly His Phe Leu	Ile Glu Gly Asp Leu Asn Pro	
165	170	175
Asp Asn Ser Asp Val Asp Lys Leu Phe	Ile Gln Leu Val Gln Ile Tyr	
180	185	190
Asn Gln Leu Phe Glu Glu Asn Pro Ile	Asn Ala Ser Arg Val Asp Ala	
195	200	205
Lys Ala Ile Leu Ser Ala Arg Leu Ser	Lys Ser Arg Arg Leu Glu Asn	
210	215	220
Leu Ile Ala Gln Leu Pro Gly Glu Lys	Arg Asn Gly Leu Phe Gly Asn	
225	230	235
Leu Ile Ala Leu Ser Leu Gly Leu Thr	Pro Asn Phe Lys Ser Asn Phe	
245	250	255
Asp Leu Ala Glu Asp Ala Lys Leu Gln	Leu Ser Lys Asp Thr Tyr Asp	
260	265	270
Asp Asp Leu Asp Asn Leu Leu Ala Gln	Ile Gly Asp Gln Tyr Ala Asp	
275	280	285
Leu Phe Leu Ala Ala Lys Asn Leu Ser	Asp Ala Ile Leu Leu Ser Asp	
290	295	300
Ile Leu Arg Val Asn Ser Glu Ile Thr	Lys Ala Pro Leu Ser Ala Ser	
305	310	315
Met Ile Lys Arg Tyr Asp Glu His His	Gln Asp Leu Thr Leu Leu Lys	
325	330	335
Ala Leu Val Arg Gln Gln Leu Pro Glu	Lys Tyr Lys Glu Ile Phe Phe	
340	345	350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly	Tyr Ile Asp Gly Gly Ala Ser	
355	360	365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys	Pro Ile Leu Glu Lys Met Asp	
370	375	380
Gly Thr Glu Glu Leu Leu Val Lys Leu	Asn Arg Glu Asp Leu Leu Arg	
385	390	395
		400

Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile	His	Leu			
																405	410	415
Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr	Pro	Phe			
																420	425	430
Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe	Arg	Ile			
																435	440	445
Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe	Ala	Trp			
																450	455	460
Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe	Glu	Glu			
465																470	475	480
Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe	Ile	Glu	Arg	Met	Thr			
																485	490	495
Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val	Leu	Pro	Lys	His	Ser			
																500	505	510
Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu	Leu	Thr	Lys	Val	Lys			
																515	520	525
Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe	Leu	Ser	Gly	Glu	Gln			
																530	535	540
Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr	Asn	Arg	Lys	Val	Thr			
545																550	555	560
Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys	Ile	Glu	Cys	Phe	Asp			
																565	570	575
Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe	Asn	Ala	Ser	Leu	Gly			
																580	585	590
Ala	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp	Lys	Asp	Phe	Leu	Asp			
																595	600	605
Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile	Val	Leu	Thr	Leu	Thr			
																610	615	620
Leu	Phe	Glu	Asp	Arg	Gly	Met	Ile	Glu	Glu	Arg	Leu	Lys	Thr	Tyr	Ala			
625																630	635	640
His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys	Gln	Leu	Lys	Arg	Arg	Arg	Tyr			
																645	650	655
Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile	Asn	Gly	Ile	Arg	Asp			
																660	665	670
Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu	Lys	Ser	Asp	Gly	Phe			
																675	680	685
Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp	Asp	Ser	Leu	Thr	Phe			
																690	695	700
Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly	Gln	Gly	His	Ser	Leu			

705	710	715	720
His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly			
	725	730	735
Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly			
	740	745	750
His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr			
	755	760	765
Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu			
	770	775	780
Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val			
785	790	795	800
Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln			
	805	810	815
Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu			
	820	825	830
Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp			
	835	840	845
Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly			
	850	855	860
Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn			
865	870	875	880
Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe			
	885	890	895
Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys			
	900	905	910
Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys			
	915	920	925
His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu			
	930	935	940
Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys			
945	950	955	960
Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu			
	965	970	975
Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val			
	980	985	990
Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe Val			
	995	1000	1005
Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys			
1010	1015	1020	

Ser Glu	Gln Glu Ile Gly Lys	Ala Thr Ala Lys Tyr	Phe Phe Tyr
1025	1030	1035	
Ser Asn	Ile Met Asn Phe Phe	Lys Thr Glu Ile Thr	Leu Ala Asn
1040	1045	1050	
Gly Glu	Ile Arg Lys Arg Pro	Leu Ile Glu Thr Asn	Gly Glu Thr
1055	1060	1065	
Gly Glu	Ile Val Trp Asp Lys	Gly Arg Asp Phe Ala	Thr Val Arg
1070	1075	1080	
Lys Val	Leu Ser Met Pro Gln	Val Asn Ile Val Lys	Lys Thr Glu
1085	1090	1095	
Val Gln	Thr Gly Gly Phe Ser	Lys Glu Ser Ile Leu	Pro Lys Arg
1100	1105	1110	
Asn Ser	Asp Lys Leu Ile Ala	Arg Lys Lys Asp Trp	Asp Pro Lys
1115	1120	1125	
Lys Tyr	Gly Gly Phe Asp Ser	Pro Thr Val Ala Tyr	Ser Val Leu
1130	1135	1140	
Val Val	Ala Lys Val Glu Lys	Gly Lys Ser Lys Lys	Leu Lys Ser
1145	1150	1155	
Val Lys	Glu Leu Leu Gly Ile	Thr Ile Met Glu Arg	Ser Ser Phe
1160	1165	1170	
Glu Lys	Asn Pro Ile Asp Phe	Leu Glu Ala Lys Gly	Tyr Lys Glu
1175	1180	1185	
Val Lys	Lys Asp Leu Ile Ile	Lys Leu Pro Lys Tyr	Ser Leu Phe
1190	1195	1200	
Glu Leu	Glu Asn Gly Arg Lys	Arg Met Leu Ala Ser	Ala Gly Glu
1205	1210	1215	
Leu Gln	Lys Gly Asn Glu Leu	Ala Leu Pro Ser Lys	Tyr Val Asn
1220	1225	1230	
Phe Leu	Tyr Leu Ala Ser His	Tyr Glu Lys Leu Lys	Gly Ser Pro
1235	1240	1245	
Glu Asp	Asn Glu Gln Lys Gln	Leu Phe Val Glu Gln	His Lys His
1250	1255	1260	
Tyr Leu	Asp Glu Ile Ile Glu	Gln Ile Ser Glu Phe	Ser Lys Arg
1265	1270	1275	
Val Ile	Leu Ala Asp Ala Asn	Leu Asp Lys Val Leu	Ser Ala Tyr
1280	1285	1290	
Asn Lys	His Arg Asp Lys Pro	Ile Arg Glu Gln Ala	Glu Asn Ile
1295	1300	1305	
Ile His	Leu Phe Thr Leu Thr	Asn Leu Gly Ala Pro	Ala Ala Phe

1310	1315	1320
Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr		
1325	1330	1335
Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly		
1340	1345	1350
Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Pro		
1355	1360	1365
Lys Lys Lys Arg Lys Val		
1370		
<210> 10		
<211> 4122		
<212> DNA		
<213> 人工序列		
<220>		
<223> 合成的		
<400> 10		
atggacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg	60	
atcaccgacg actacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg	120	
cacagcatca agaagaacct gatcggcgcc ctgctgttcg gctctggcga aacagccgag	180	
gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc	240	
tatctgcaag agatcttcag caacgagatg gccaaagtgg acgacagctt ctccacaga	300	
ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc	360	
aacatcgtag acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag	420	
aagctggccg acagcaccga caaggccgac ctgagactga tctacctggc cctggccac	480	
atgatcaagt tccggggcca ctctctgac gagggcgacc tgaaccccga caacagcgac	540	
gtggacaagc tggtcatcca gctggtgcag atctacaatc agctgttcga ggaaaacccc	600	
atcaacgcca gcagagtgga cgccaaggcc atcctgagcg ccagactgag caagagcaga	660	
cggctggaaa atctgatcgc ccagctgccc ggcgagaagc ggaatggcct gttcggcaac	720	
ctgattgccc tgagcctggg cctgacccc aacttcaaga gcaacttcga cctggccgag	780	
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	840	
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc	900	
ctgctgagcg acatcctgag agtgaacagc gagatcacca aggccccct gtccgctct	960	
atgatcaaga gatacgacga gcaccaccag gacctgacct tgctgaaagc tctcgtgcgg	1020	
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc	1080	
ggctacatcg atggcggagc cagccaggaa gagttctaca agttcatcaa gccatcctg	1140	
gaaaagatgg acggcaccga ggaactgtc gtgaagctga acagagagga cctgctgcgg	1200	
aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac	1260	
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc	1320	
gagaagatcc tgaccttcag aatcccctac tacgtgggcc ctctggccag gggaaacagc	1380	

agattcgcct	ggatgaccag	aaagagcgag	gaaaccatca	ccccctggaa	cttcgaggaa	1440
gtggtggaca	agggcgccag	cgcccagagc	ttcatcgagc	ggatgaccaa	cttcgataag	1500
aacctgcccc	acgagaaggt	gctgcccgaag	cacagcctgc	tgtacgagta	cttcaccgtg	1560
tacaacgagc	tgaccaaagt	gaaatacgtg	accgagggaa	tgcggaagcc	cgcctttctg	1620
agcggcgagc	agaaaaaggc	catcgtggac	ctgctgttca	agaccaaccg	gaaagtgacc	1680
gtgaagcagc	tgaaagagga	ctacttcaag	aaaatcgagt	gcttcgacag	cgtggaaatc	1740
agcggcggtg	aagatcggtt	caacgcctcc	ctgggcgcct	atcacgatct	gctgaaaatt	1800
atcaaggaca	aggacttcct	ggacaatgag	gaaaacgagg	acattctgga	agatatcgtg	1860
ctgacctga	cactgtttga	ggaccggggc	atgatcgagg	aacggctgaa	aacctatgcc	1920
cacctgttcg	acgacaaaagt	gatgaagcag	ctgaagcggc	ggagatacac	cggctggggc	1980
aggctgagcc	ggaagctgat	caacggcatc	cgggacaagc	agtccggcaa	gacaatcctg	2040
gatttcctga	agtccgacgg	cttcgccaac	agaaacttca	tgcagctgat	ccacgacgac	2100
agcctgacct	ttaaagagga	catccagaaa	gcccggtgt	ccggccaggg	acactctctg	2160
cacgagcaga	tcgccaatct	ggccggatcc	cccgccatta	agaaggcat	cctgcagaca	2220
gtgaagattg	tggacgagct	cgtgaaagtg	atggggcaca	agcccgagaa	catcgtgac	2280
gaaatggcca	gagagaacca	gaccacccag	aaggacaga	agaacagccg	cgagagaatg	2340
aagcggatcg	aagaggcat	caaagagctg	ggcagccaga	tcctgaaaga	acaccccgtg	2400
gaaaacaccc	agctgcagaa	cgagaagctg	tacctgtact	acctgcagaa	tgggcgggat	2460
atgtacgtgg	accaggaact	ggacatcaac	cggctgtccg	actacgatgt	ggaccacatt	2520
gtgccccagt	ccttcatcaa	ggacgactcc	atcgataaca	aagtgtgac	tcggagcgac	2580
aagaaccggg	gcaagagcga	caacgtgccc	tccgaagagg	tcgtgaagaa	gatgaagaac	2640
tactggcgcc	agctgctgaa	tgccaagctg	attaccaga	ggaagttega	caatctgacc	2700
aaggccgaga	gaggcggcct	gagcgaactg	gataaggccg	gcttcattaa	gcggcagctg	2760
gtggaaaccc	ggcagatcac	aaagcacgtg	gcacagatcc	tggactcccg	gatgaacact	2820
aagtacgacg	agaacgacaa	actgatccgg	gaagtgaag	tgatcacct	gaagtccaag	2880
ctggtgtccg	acttcagaaa	ggatttccag	ttttacaaag	tgcgcgagat	caacaactac	2940
caccacgccc	acgacgccta	cctgaacgcc	gtcgtgggaa	ccgccctgat	caaaaagtac	3000
cctaagctgg	aaagcgagtt	cgtgtacggc	gattacaagg	tgtacgacgt	gcggaagatg	3060
atcgccaaga	gcgagcagga	aatcggcaag	gctaccgcca	agtacttttt	ctacagcaac	3120
atcatgaact	ttttcaagac	cgagatcaca	ctggccaacg	gcgagatcag	aaagcggcct	3180
ctgatcgaga	caaacggcga	aaccggggag	atcgtgtggg	ataagggccg	ggattttgcc	3240
acagtgcgga	aagtgtgtc	catgccccaa	gtgaatatcg	tgaaaaagac	cgaggtgcag	3300
accggcggct	tcagcaaaga	gtctatcctg	ccaagagga	actccgacaa	gctgatcgcc	3360
agaaagaagg	attgggaccc	taagaagtac	ggcggttttg	acagccccac	cgtggcctac	3420
tctgtgctgg	tgggtggcaa	agtggaaaag	ggcaagtcca	agaaactgaa	gagtgtgaaa	3480
gagctgctgg	ggatcaccat	catggaaaga	agcagcttcg	agaagaatcc	catcgacttt	3540
ctggaagcca	agggctacaa	agaagtgaaa	aaggacctga	tcatcaagct	gcctaagtac	3600
tcctgttcg	agctggaaaa	cggccggaag	cggatgctgg	cttctgccgg	cgaactgcag	3660
aagggaacg	agctggccct	gccctccaaa	tatgtgaact	tcctgtacct	ggccagccac	3720

tatgagaagc tgaagggctc ccccaggat aatgagcaga aacagctgtt tgtggaacag	3780
cacaagcact acctggacga gatcatcgag cagattagcg agttctccaa gcgcgtgac	3840
ctggccgatg ccaacctgga caaggtgctg agcgcctaca acaagcaccg ggataagccc	3900
atcagagagc aggccgagaa tatcatccac ctgtttaccc tgaccaacct gggagcccct	3960
gccgccttca agtactttga caccaccatc gaccggaaga ggtacaccag caccaaagag	4020
gtgctggacg ccacctgat ccaccagagc atcaccggcc tgtacgagac acggatcgac	4080
ctgtctcagc tgggaggcga cccaagaaa aagcgcaaag tg	4122
<210> 11	
<211> 4764	
<212> DNA	
<213> 人	
<400> 11	
gcgggcgggc ggtgcgatgt ccggagagga tggcccggcg gctggcccgg gggcgggcgc	60
ggcggtgcc cgggagcggc gacgggagca gctgcggcag tggggggcgc gggcgggcgc	120
cgagcctggc cccggagagc gccgcgccc caccgtccgc ttcgagcgcg ccgccgagtt	180
cctggcgggc tgtgcgggcg gcgacctgga cgaggcgcgt ctgatgctgc gcgccccga	240
ccctggcccc ggcgccgagc tcgaccccc gcgcgccgcg cccgccgcg ccgtgctgga	300
ctccaccaac gccgacggta tcagcgccct gcaccagtc agcgcccccc gcccggcgtc	360
tcccggggcc aggtccaccc tctgctgcgc cacctggggc atcctccttc cccgttgcca	420
gtctcgatcc gccccgtcgt tcctggccct gggetttgcc accctatget gacacccgt	480
cccagtcctt cttaccatc ccttcgacc acccaacttc cgaattggag ccgcttcaac	540
tggccctggg cttagccact ctgtgctgac cactctgccc caggcctcct taccattccc	600
cttcgacctt ctctcttccg cattggagtc gctttaactg gccctggctt tggcagcctg	660
tgtgaccca tgcagtcctt cttaccatcc ctccctcgac tccccctctt ccgatgttga	720
gccccctccag ccggtcctgg actttgtctt ctccctgccc ctgccctctt ctgaacctga	780
gccagctccc atagctcagt ctggctctatc tgccctggccc tggccattgt cactttgcgc	840
tgccctcctt tcgccccga gtgcccttgc tgtgccgcg gaactctgcc ctctaacgt	900
gccgtctctt tcctgagtc ggaccacttt gagctctact ggcttctgcg ccgcctctgg	960
cccactgttt ccccttccca ggcaggtcct gttttctctg acctgcattc tctccccctg	1020
gcctgtgccg ctttctgtct gcagcttggt gcctgggtca cctctacggc tggcccagat	1080
ccttccccctgc cgcctccttc aggttccgtc ttcctccact cctcttccc cttgctctct	1140
gctgtgttgc tgcccaagga tgctctttcc ggagcaactt cttctcggcg ctgcaccacg	1200
tgatgtcctt tgagcggatc ctccccgtgt ctgggtcctt tccgggcac tctctcctt	1260
cacccaaccc catgccgtct tcactcgtg ggttccccct tcttctcct tctggggcct	1320
gtgccatctt tcgtttctta ggatggcctt ctccgacgga tgtctccctt gcgtcccgc	1380
tccccctctt gtaggcctgc atcatcaccg tttttctgga caacccaaa gtaccccgct	1440
tccctggctt tagccacctc tccatcctct tgetttcttt gcctggacac cccgttctcc	1500
tgtggattcg ggtaacctt cactcctttc atttgggcag ctccccctacc ccccttacct	1560
ctctagtctg tgctagtctt tccagcccc tgctcatggca tcttccaggg gtccgagagc	1620

tcagctagtc	ttcttctctc	aaccgggcc	cctatgtcca	cttcaggaca	gcatgtttgc	1680
tgccctcagg	gatcctgtgt	ccccgagctg	ggaccacett	atattcccag	ggccggttaa	1740
tgtggctctg	gttctgggta	cttttatctg	tccctccac	cccacagtgg	ggccactagg	1800
gacaggattg	gtgacagaaa	agccccatcc	ttaggcctcc	tccttcctag	tctcctgata	1860
ttgggtctaa	cccccacctc	ctgttaggca	gattccttat	ctggtgacac	acccccatth	1920
cctggagcca	tctctctcct	tgccagaacc	tctaaggtht	gcttacgatg	gagccagaga	1980
ggatcctggg	agggagagct	tggcaggggg	tgggagggaa	gggggggatg	cgtgacctgc	2040
ccggttctca	gtggccaccc	tgcgtaccc	tctcccagaa	cctgagctgc	tctgacgcgg	2100
ccgtctgggtg	cgtttcactg	atcctgggtgc	tgcagcttcc	ttacacttcc	caagaggaga	2160
agcagtttgg	aaaaacaaaa	tcagaataag	ttggctctga	gttctaactt	tggctcttca	2220
cccttctagt	ccccaattha	tattgttctc	ccgtgcgtca	gttttacctg	ttagataagg	2280
ccagtagcca	gccccgtcct	ggcagggctg	tggtagggag	gggggtgtcc	gtgtggaaaa	2340
ctccctttgt	gagaatgggt	cgtcctaggt	gttcaccagg	tcgtggccgc	ctctactccc	2400
tttctctttc	tccatccttc	tttccttaaa	gagtccccag	tgtatctctg	gacatattec	2460
tccgcccaga	gcagggtccc	gcttccctaa	ggccctgctc	tgggtctctg	ggttttagtc	2520
cttggcaagc	ccaggagagg	cgtcagget	tcctgtccc	ccttctcgt	ccaccatctc	2580
atgccccctg	ctctcctgcc	ccttccctac	aggggttctc	ggctctgctc	ttcagactga	2640
gccccgttcc	cctgcatccc	cgttccccctg	catccccctt	cccctgcate	ccccagaggc	2700
cccaggccac	ctacttggcc	tggacccccc	gagaggccac	cccagccctg	tctaccaggc	2760
tgccttttgg	gtggattctc	ctccaactgt	ggggtgactg	cttggcaaac	tactcttctg	2820
gggtatccca	ggaggcctgg	agcattgggg	tgggtggggg	ttcagagagg	agggattccc	2880
ttctcagggt	acgtggccaa	gaagcagggg	agctgggttt	gggtcaggtc	tgggtgtggg	2940
gtgaccagct	tatgctgttt	gcccaggaca	gcctagtttt	agcactgaaa	ccctcagtec	3000
taggaaaaca	gggatgggtg	gtcactgtct	ctgggtgact	cttgattccc	ggccagtttc	3060
tccacctggg	gctgtgtttc	tcgtcctgca	tccttctcca	ggcagggtccc	caagcatcgc	3120
ccccctgctg	tggctgttcc	caagtcttta	gggtacccca	cgtgggttta	tcaaccactt	3180
ggtgaggctg	gtaccctgcc	cccattcctg	caccccaatt	gccttagtgg	ctaggggggt	3240
gggggctaga	gtaggagggg	ctggagccag	gattcttagg	gctgaacaga	gaagagctgg	3300
gggcctgggc	tcctgggttt	gagagaggag	gggtgggggc	ctggactcct	gggtccgagg	3360
gaggaggggc	tggggcctgg	actcctgggt	ctgagggtgg	agggactggg	ggcctggact	3420
cctgggtccg	agggaggagg	ggctggggcc	tggactcgtg	ggtctgaggg	aggaggggct	3480
gggggcctgg	acttctgggt	cttagggagg	cggggctggg	cctggacccc	tgggtctgaa	3540
tggggagagg	ctgggggcct	ggactccttc	atctgagggc	ggaagggtg	gggcctggcc	3600
tctgggttg	aatggggagg	ggttgggcct	ggactctgga	gtccctgggtg	cccaggcctc	3660
aggcatcttt	cacagggatg	cctgtactgg	gcaggctcct	gaaagggaaa	ggcccattgc	3720
tctccttgcc	ccccctccc	atcgccatga	caactgggtg	gaaataaacg	agccaggttc	3780
atcccgttcc	cagggcacgt	gcggccccct	cacagcccca	gtttccatga	cctcatgctc	3840
ttggccctcg	tagctccctc	ccgcctcctc	cagatgggca	gctttggaga	ggtgagggac	3900
ttggggggta	atttatcccc	tggatctagg	agtttagctt	cactccttcc	tcagctccag	3960


```

ttcagggtccc ggagcccacc cagtgtccac aaggcctggg gcaagtcct cctccgaccc 4020
cctggacttc ggcttttgtc cccccaagtt ttggaccctt aagggaagaa tgagaaacgg 4080
tggtccgtgt cagccccctg ctgcagggcc ccgtgcagag ggggcctcag tgaactggag 4140
tgtgacagcc tggggcccag gcacacaggt gtgcagctgt ctcaccctc tgggagtccc 4200
gcccaggccc ctgagtctgt ccagcacag ggtggccttc ctccaccctg catagccctg 4260
ggcccacggc ttcgttcctg cagagtatct gctggggtgg tttccgagct tgacccttgg 4320
aaggacctgg ctgggtttaa ggaggagggt gctggggggc aggactcctg gctctgaagg 4380
aggaggggct ggaacctctt ccctagtctg agcactggaa gcgccacctg tgggtggtga 4440
cggggggttt gccgtgtcta acaggtacca tgtgggggttc ccgcaccag atgagaagcc 4500
ccctcccttc cccgttcaact tcctgtttgc agatagccag gagtcctttc gtggtttcca 4560
ctgagcactg aaggcctggc cggcctgacc actgggcaac caggcgtatc ttaaacagcc 4620
agtggccaga ggctgttggg tcattttccc cactgtecta gcaccgtgtc cctggatctg 4680
ttttcgtggc tccctctgga gtcccgaact gctgggacac cgtggctggg gtaggtgcgg 4740
ctgacggctg tttcccaccc ccag 4764

```

<210> 12

<211> 42

<212> RNA

<213> 人工序列

<220>

<223> 合成的

<400> 12

```

acccacagu ggggccacua guuuuagagc uaugcuguuu ug 42

```

<210> 13

<211> 86

<212> RNA

<213> 人工序列

<220>

<223> 合成的

<400> 13

```

ggaaccuuu aaaacagcau agcaaguuaa aaauaggcua guccguuau aacuugaaaa 60

```

```

aguggcaccg agucggugcu uuuuuu 86

```

<210> 14

<211> 62

<212> RNA

<213> 人工序列

<220>

<223> 合成的

<400> 14

```

acccacagu ggggccacua guuuuagagc uagaaauagc aaguuaauu aaggcuaguc 60

```

cg	62
<210> 15	
<211> 25	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 15	
ctgacctctt ctcttcctcc cacag	25
<210> 16	
<211> 1009	
<212> DNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 16	
gccacatgg actacaaaga cgatgacgac aaggtcgact ctagagctgc agagagcgac	60
gagagcggcc tgcccgccat ggagatcgag tgccgcatca ccggcaccct gaacggcgtg	120
gagttcgagc tgggtgggagg cggagagggc acccccgagc agggccgcat gaccaacaag	180
atgaagagca ccaaaggcgc cctgaccttc agcccctacc tgctgagcca cgtgatgggc	240
tacggcttct accacttcgg cacctacccc agcggctacg agaaccctt cctgcacgcc	300
atcaacaacg gcggttacac caacacccgc atcgagaagt acgaggacgg cggcgtgctg	360
cacgtgagct tcagctaccg ctacgaggcc ggccgcgtga tcggcgactt caaggtgatg	420
ggcaccggct tccccgagga cagcgtgatc ttcaccgaca agatcgtccg cagcaacgcc	480
accgtggagc acctgcaccc catgggcat aacgatctgg atggcagctt caccgcacc	540
ttcagcctgc gcgacggcgg ctactacagc tccgtgggtg acagccacat gcacttcaag	600
agcgccatcc accccagcat cctgcagaac gggggcccca tgttgcctt ccgccgctg	660
gaggaggatc acagcaacac cgagctgggc atcgtggagt accagcagc cttcaagacc	720
ccggatgcag atgccggtga agaataaga tctctgtgcc ttctagttgc cagccatctg	780
ttgtttgccc ctcccccggtg ctttccttga ccctggaagg tgccactccc actgtccttt	840
cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg	900
gtggggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg	960
atgcggtggg ctctatggac tcgaggttta aacgtcgacg cggccgcgt	1009
<210> 17	
<211> 355	
<212> PRT	
<213> 人工序列	
<220>	
<223> 合成的	

<400> 17

Met	Ser	Gly	Glu	Asp	Gly	Pro	Ala	Ala	Gly	Pro	Gly	Ala	Ala	Ala	Ala
1				5					10					15	
Ala	Ala	Arg	Glu	Arg	Arg	Arg	Glu	Gln	Leu	Arg	Gln	Trp	Gly	Ala	Arg
			20					25					30		
Ala	Gly	Ala	Glu	Pro	Gly	Pro	Gly	Glu	Arg	Arg	Ala	Arg	Thr	Val	Arg
			35					40					45		
Phe	Glu	Arg	Ala	Ala	Glu	Phe	Leu	Ala	Ala	Cys	Ala	Gly	Gly	Asp	Leu
			50				55					60			
Asp	Glu	Ala	Arg	Leu	Met	Leu	Arg	Ala	Ala	Asp	Pro	Gly	Pro	Gly	Ala
65					70					75					80
Glu	Leu	Asp	Pro	Ala	Ala	Pro	Pro	Pro	Ala	Arg	Ala	Val	Leu	Asp	Ser
					85					90				95	
Thr	Asn	Ala	Asp	Gly	Ile	Ser	Ala	Leu	His	Gln	Ala	Thr	Met	Asp	Tyr
				100				105					110		
Lys	Asp	Asp	Asp	Asp	Lys	Val	Asp	Ser	Arg	Ala	Ala	Glu	Ser	Asp	Glu
			115					120					125		
Ser	Gly	Leu	Pro	Ala	Met	Glu	Ile	Glu	Cys	Arg	Ile	Thr	Gly	Thr	Leu
			130				135					140			
Asn	Gly	Val	Glu	Phe	Glu	Leu	Val	Gly	Gly	Gly	Glu	Gly	Thr	Pro	Glu
145					150					155					160
Gln	Gly	Arg	Met	Thr	Asn	Lys	Met	Lys	Ser	Thr	Lys	Gly	Ala	Leu	Thr
					165					170				175	
Phe	Ser	Pro	Tyr	Leu	Leu	Ser	His	Val	Met	Gly	Tyr	Gly	Phe	Tyr	His
			180					185					190		
Phe	Gly	Thr	Tyr	Pro	Ser	Gly	Tyr	Glu	Asn	Pro	Phe	Leu	His	Ala	Ile
			195				200					205			
Asn	Asn	Gly	Gly	Tyr	Thr	Asn	Thr	Arg	Ile	Glu	Lys	Tyr	Glu	Asp	Gly
			210				215					220			
Gly	Val	Leu	His	Val	Ser	Phe	Ser	Tyr	Arg	Tyr	Glu	Ala	Gly	Arg	Val
225					230					235				240	
Ile	Gly	Asp	Phe	Lys	Val	Met	Gly	Thr	Gly	Phe	Pro	Glu	Asp	Ser	Val
					245					250				255	
Ile	Phe	Thr	Asp	Lys	Ile	Val	Arg	Ser	Asn	Ala	Thr	Val	Glu	His	Leu
			260					265					270		
His	Pro	Met	Gly	Asp	Asn	Asp	Leu	Asp	Gly	Ser	Phe	Thr	Arg	Thr	Phe
			275				280						285		
Ser	Leu	Arg	Asp	Gly	Gly	Tyr	Tyr	Ser	Ser	Val	Val	Asp	Ser	His	Met
			290				295					300			

His Phe Lys Ser Ala Ile His Pro Ser Ile Leu Gln Asn Gly Gly Pro
 305 310 315 320
 Met Phe Ala Phe Arg Arg Val Glu Glu Asp His Ser Asn Thr Glu Leu
 325 330 335
 Gly Ile Val Glu Tyr Gln His Ala Phe Lys Thr Pro Asp Ala Asp Ala
 340 345 350
 Gly Glu Glu
 355

<210> 18

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 18

ccactctgtg ctgaccactc t

21

<210> 19

<211> 17

<212> DNA

<213> 人工序列

<220>

<223> 合成的

<400> 19

gcggcactcg atctcca

17

<210> 20

<211> 711

<212> DNA

<213> Mus musculus

<400> 20

gagcggctgc ggggcgggtg caagcacgtt tccgacttga gttgcctcaa gaggggcgtg 60
 ctgagccaga cctccatcgc gcactccggg gagtggaggg aaggagcgag ggctcagttg 120
 ggctgttttg gaggcaggaa gcacttgtc tcccaaagtc gctctgagtt gttatcagta 180
 agggagctgc agtggagtag gcggggagaa ggccgcaccc ttctccggag gggggagggg 240
 agtgttgcaa tacctttctg ggagttctct gctgcctcct ggcttctgag gaccgccctg 300
 ggcttgggag aatcccttcc ccctcttccc tcgtgatctg caactccagt ctttctagaa 360
 gatgggcggg agtcttctgg gcaggettaa aggetaacct ggtgtgtggg cgttgctcctg 420
 caggggaatt gaacagggtg aaaattggag ggacaagact tcccacagat tttcggtttt 480
 gtcgggaagt tttttaatag gggcaaataa ggaaaatggg aggataggta gtcactctggg 540
 gttttatgca gcaaaactac aggttattat tgcttgtgat ccgcctcgga gtattttcca 600

tcgaggtaga ttaaagacat gctcacccga gttttatact ctcttgcttg agatccttac	660
tacagtatga aattacagtg tcgcgagtta gactatgtaa gcagaatttt a	711
<210> 21	
<211> 42	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 21	
cuccagucuu ucuagaagau guuuuagagc uaugcuguuu ug	42
<210> 22	
<211> 42	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 22	
ugaacaggug uaaaauugga guuuuagagc uaugcuguuu ug	42
<210> 23	
<211> 42	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 23	
ugucgggaag uuuuuuaaua guuuuagagc uaugcuguuu ug	42
<210> 24	
<211> 642	
<212> DNA	
<213> Rattus rattus	
<400> 24	
gggattcctc cttgagttgt ggcaactgagg aacgtgctga acaagaccta cattgcactc	60
cagggagtg atgaaggagt tggggetcag tcgggttgta ttggagacaa gaagcacttg	120
ctctccaaaa gtcggtttga gttatcatta aggagctgc agtggagtag gcggagaaaa	180
ggccgcaccc ttctcaggac gggggagggg agtgttgcaa tacctttctg ggagttctct	240
gtgcctcct gtcttctgag gaccgcctg ggcttggaag attcccttcc cccttcttcc	300
ctcgtgatct gcaactggag tctttctgga agataggcgg gagtcttctg ggcaggetta	360
aaggctaacc tgggtgcgtg ggcggtgtcc tgcagaggaa ttgaacaggt gtaaaattgg	420
aggggcaaga cttccacag attttcgatt gtgttgtaa gtattgtaat aggggcaaat	480

aagggaaata gactaggcac tcacctgggg ttttatgcag caaaactaca ggttattatt	540
gcttgtgatc cgccctggag aatttttcac cgaggtagat tgaagacatg cccacccaaa	600
ttttaatatt cttccacttg cgatccttgc tacagtatga aa	642
<210> 25	
<211> 42	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 25	
agggggaagg gaauuuucca guuuuagagc uaugcuguuu ug	42
<210> 26	
<211> 42	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 26	
ucugcaacug gagucuuucu guuuuagagc uaugcuguuu ug	42
<210> 27	
<211> 42	
<212> RNA	
<213> 人工序列	
<220>	
<223> 合成的	
<400> 27	
aggcgggagu cuucugggca guuuuagagc uaugcuguuu ug	42

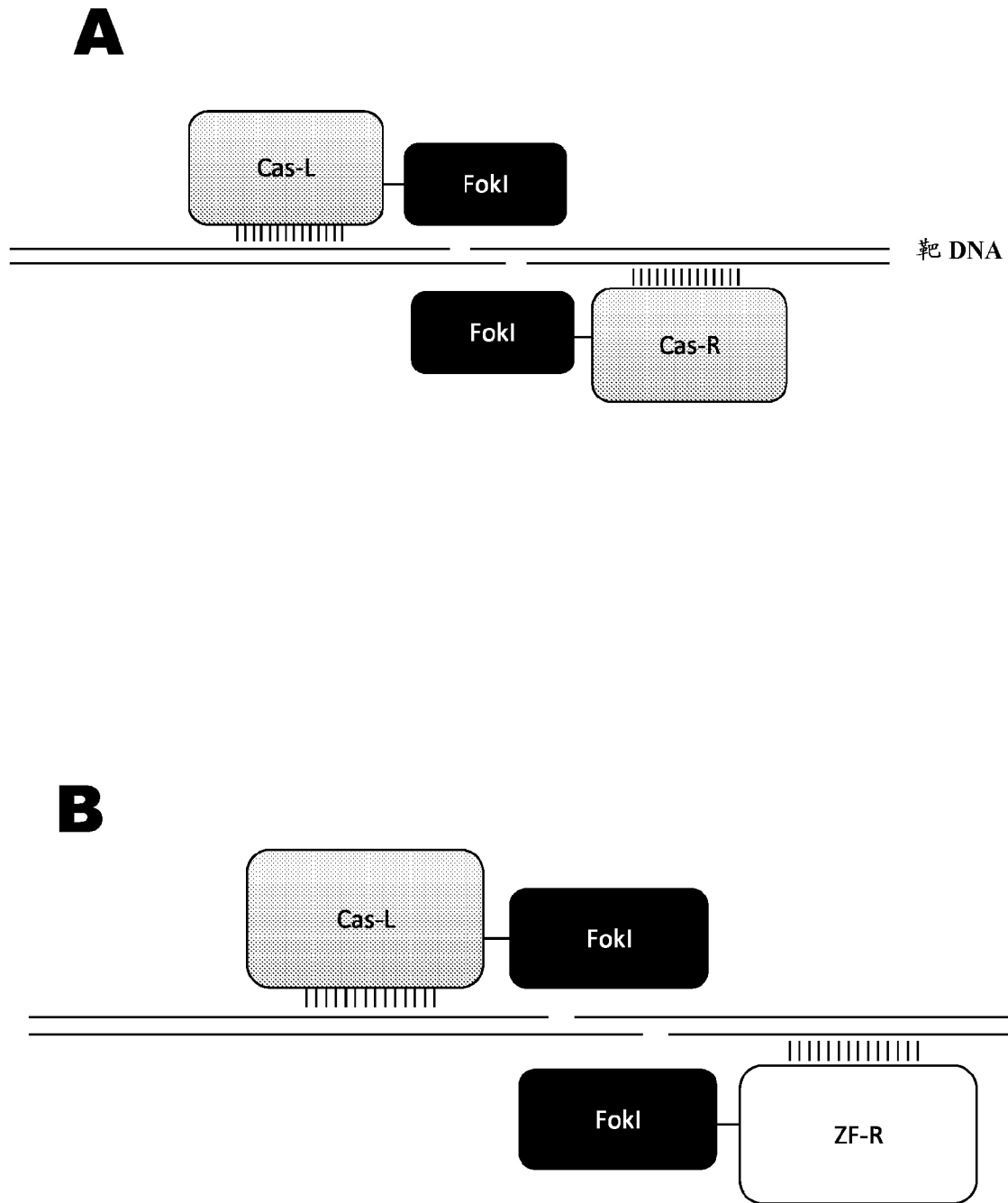


图 1

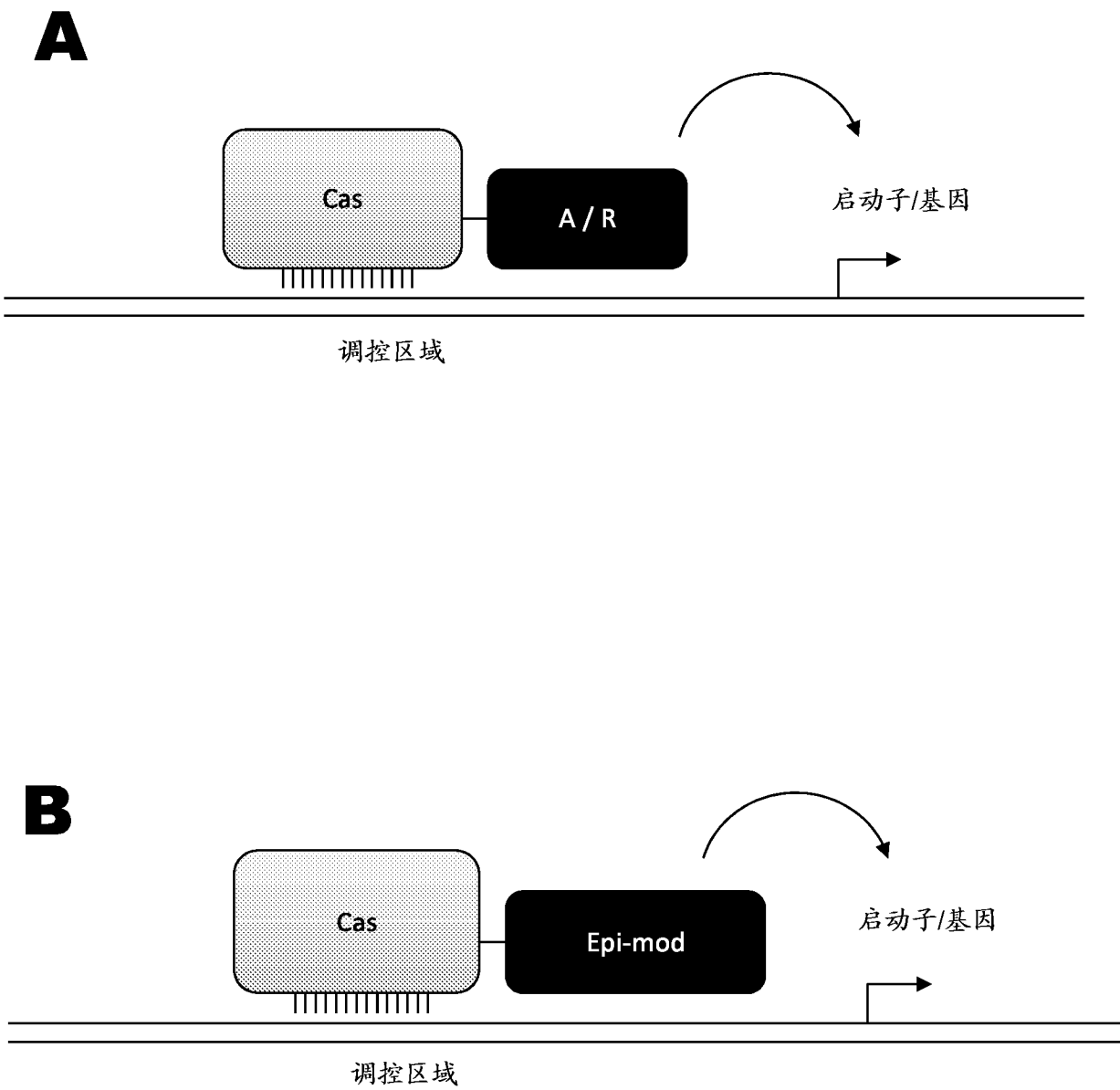


图 2

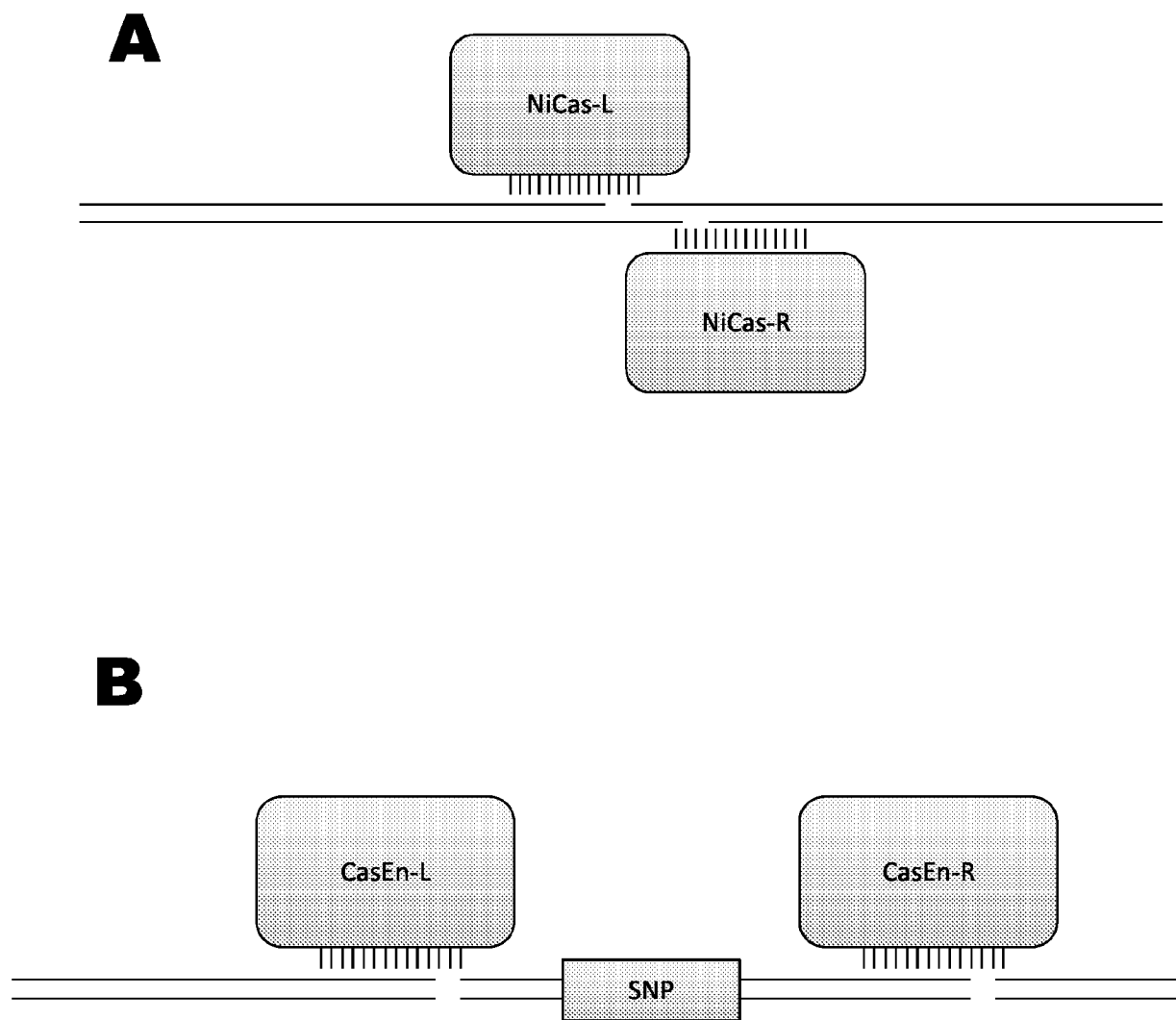
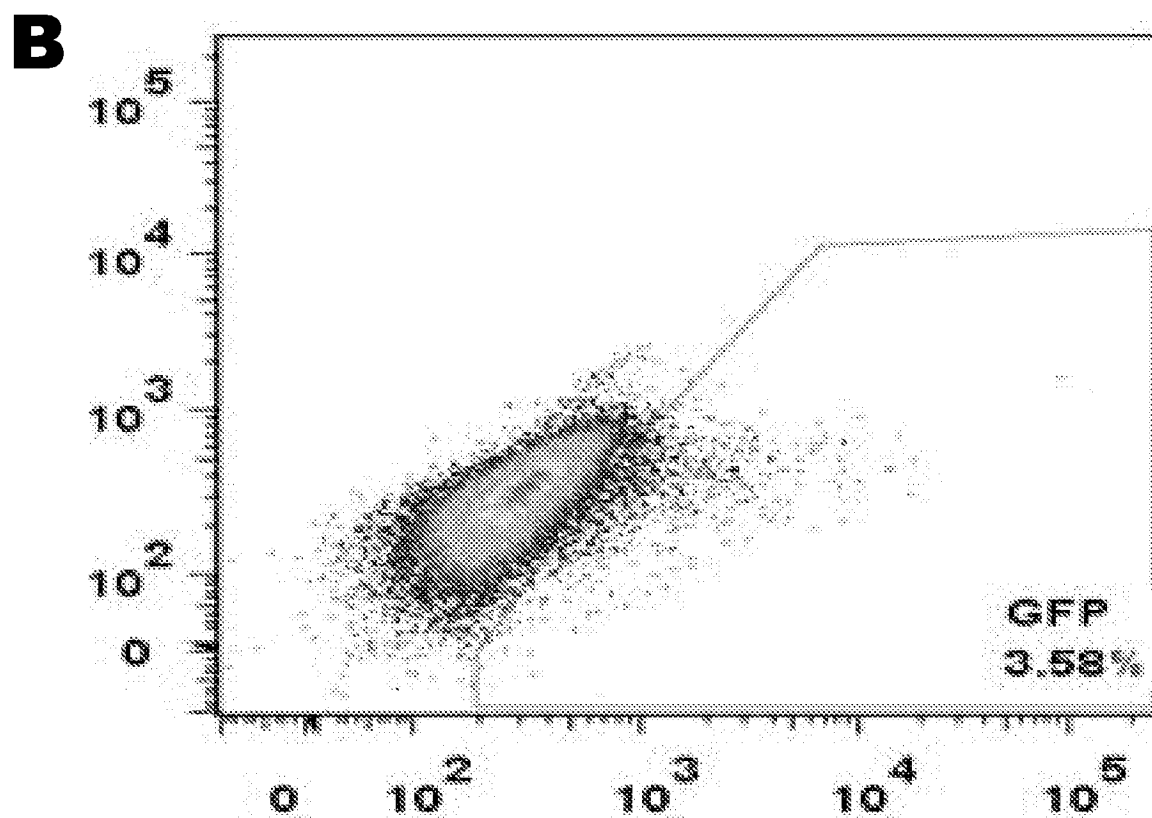
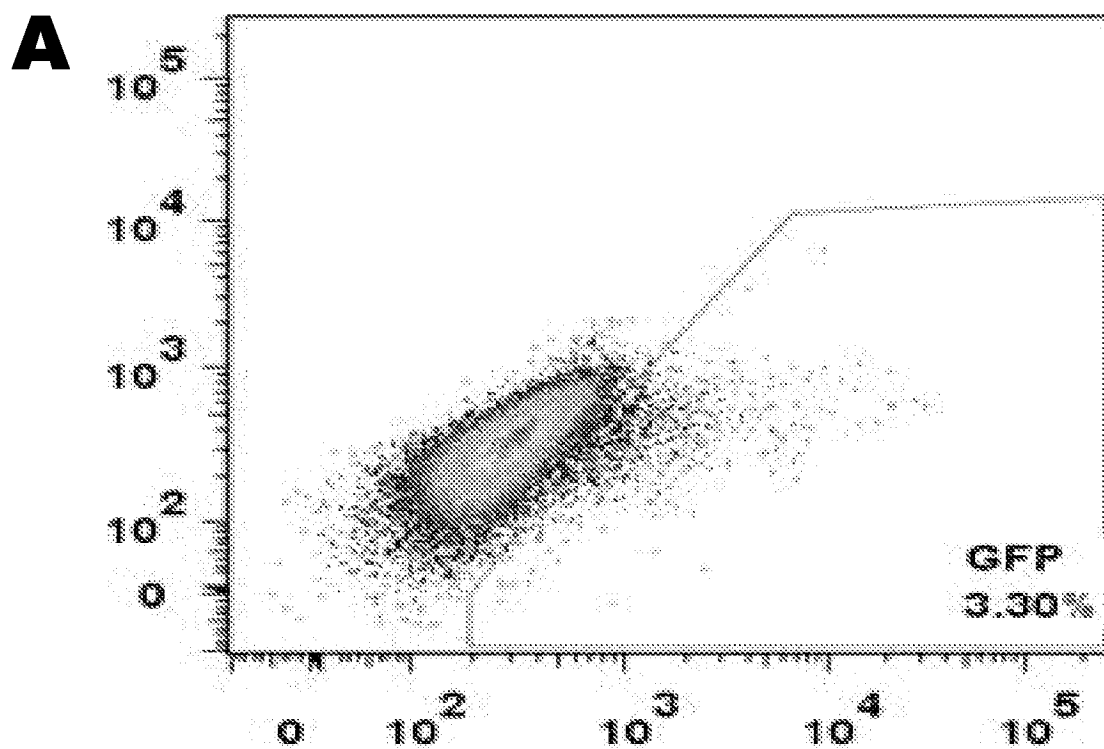
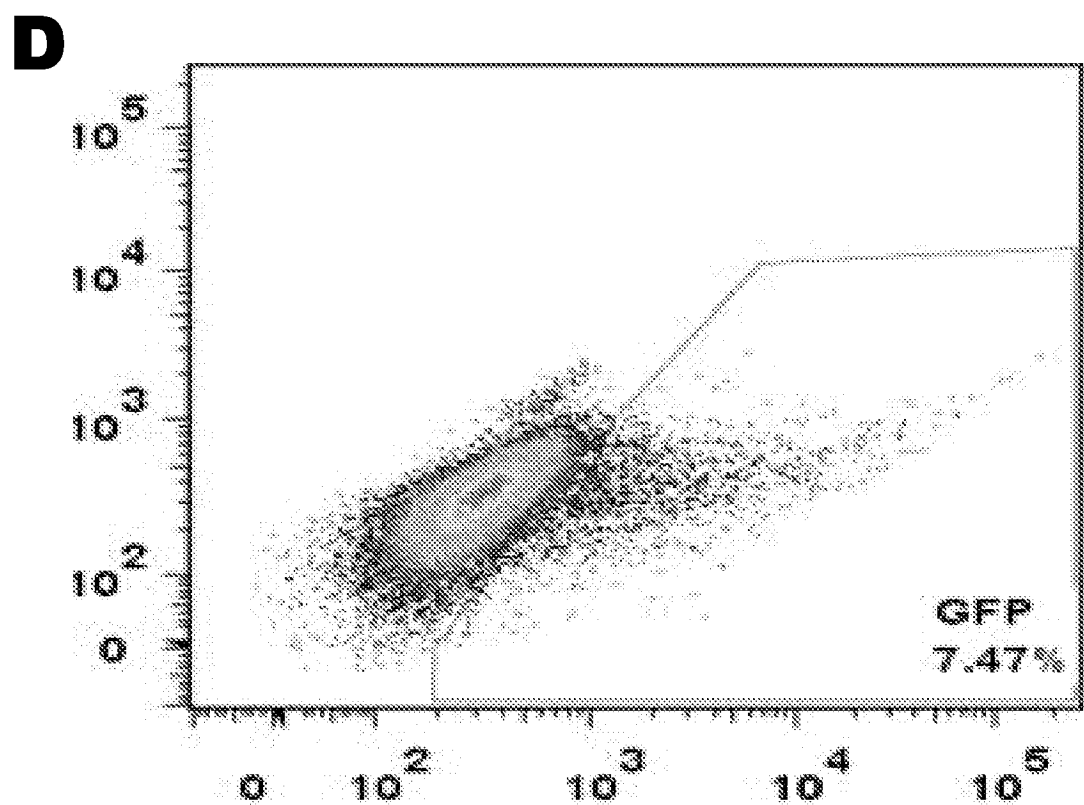
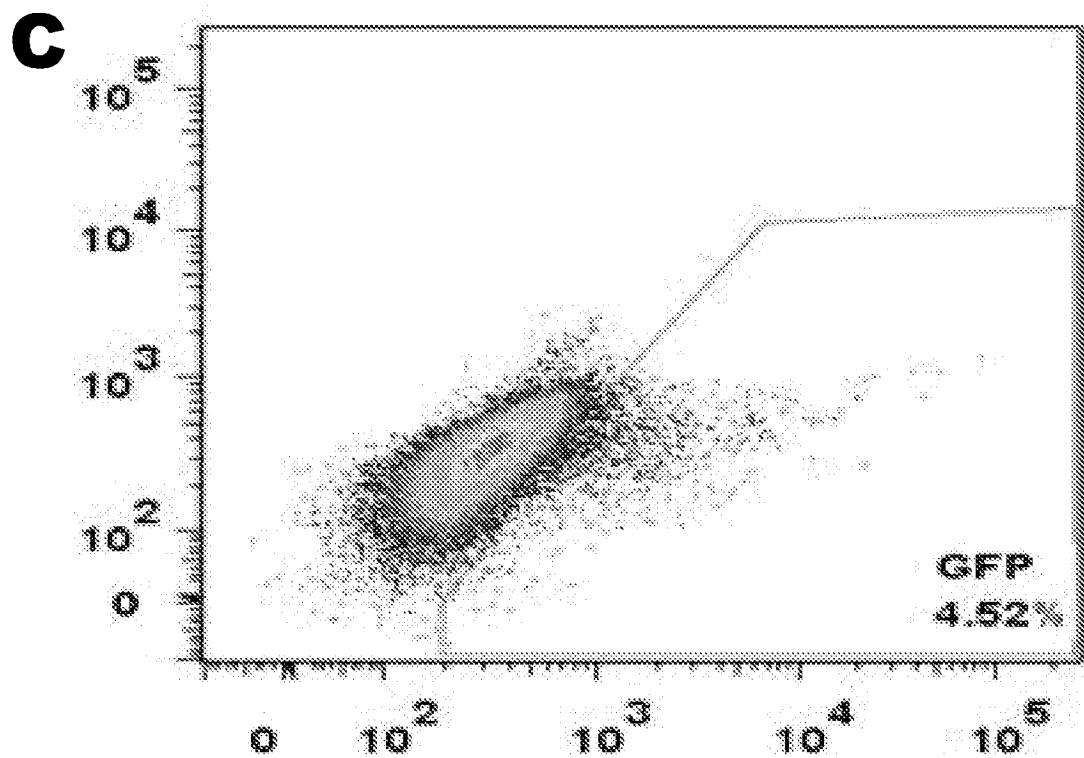


图 3





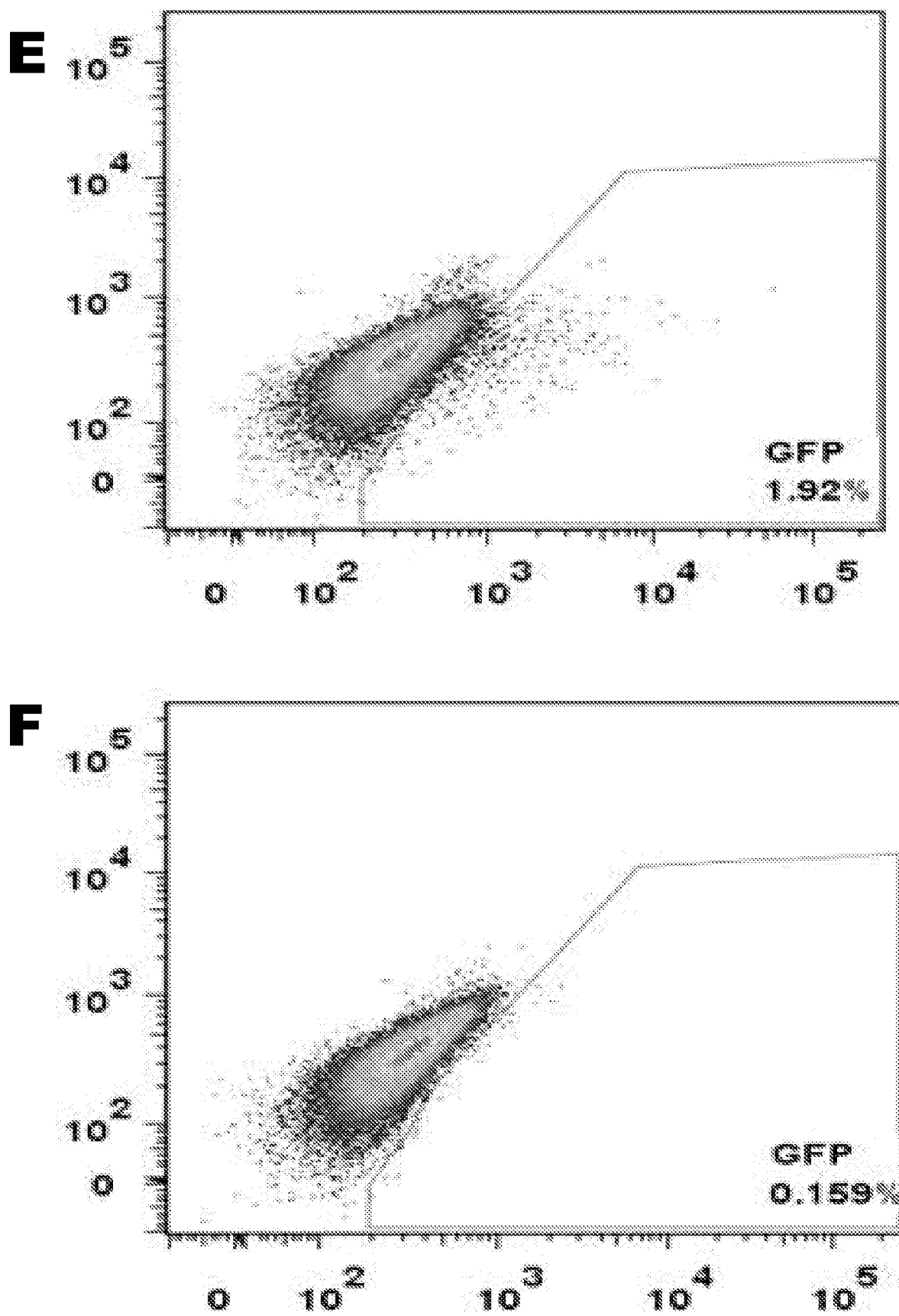


图 4

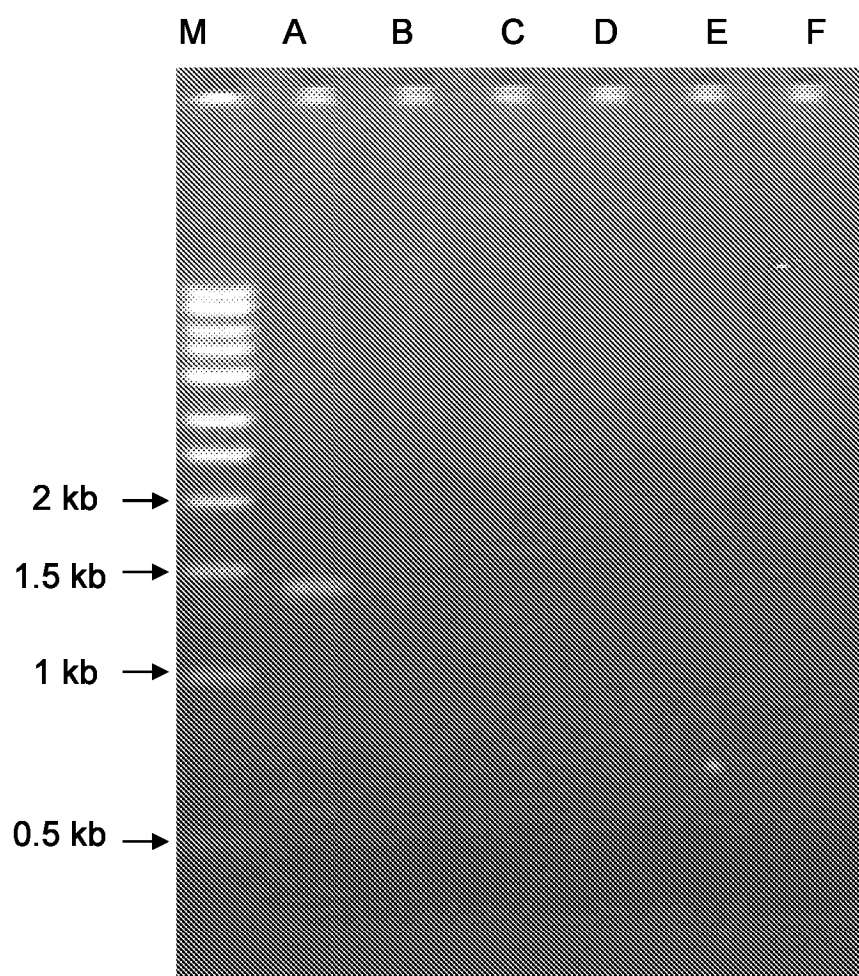


图 5