



# (12)发明专利申请

(10)申请公布号 CN 111093715 A

(43)申请公布日 2020.05.01

(21)申请号 201880057176.7

(22)申请日 2018.09.18

(30)优先权数据

62/559,751 2017.09.18 US

(85)PCT国际申请进入国家阶段日

2020.03.03

(86)PCT国际申请的申请数据

PCT/US2018/051590 2018.09.18

(87)PCT国际申请的公布数据

WO2019/056015 EN 2019.03.21

(71)申请人 儿童医院医疗中心

地址 美国俄亥俄州

(72)发明人 P·马利克

(74)专利代理机构 上海一平知识产权代理有限公司 31266

代理人 徐迅 王正君

(51)Int.Cl.

A61K 48/00(2006.01)

C12N 15/86(2006.01)

权利要求书3页 说明书39页

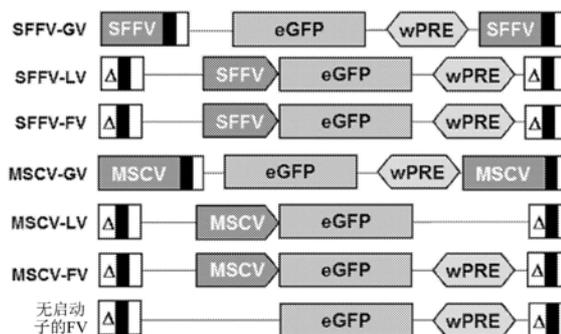
序列表26页 附图16页

## (54)发明名称

强绝缘子和其在基因递送中的用途

## (57)摘要

一种来自泡沫病毒的强绝缘子片段,其可以用于使转基因的表达绝缘,并且减少包括此类绝缘子的整合载体的基因毒性。所述绝缘子片段也可以用于基因编辑中的基因靶向性构建体中。



1. 一种用于基因递送的核酸构建体,所述核酸构建体包括由绝缘子片段侧接的至少一个目的基因(GOI),其中所述绝缘子片段包括与AAGGGAGACATCTAGTGATATAAGTGTGAACTACAC(SEQ ID NO:2)至少90%相同的核苷酸序列或其互补序列。

2. 根据权利要求1所述的核酸构建体,其中所述绝缘子片段与所述核酸构建体的至少一个片段异源。

3. 根据权利要求1或权利要求2所述的核酸构建体,其中所述绝缘子片段包括与SEQ ID NO:2至少95%相同的核苷酸序列或其所述互补序列。

4. 根据权利要求3所述的核酸构建体,其中所述绝缘子片段包括SEQ ID NO:2的核苷酸序列或其所述互补序列。

5. 根据权利要求1到4中任一项所述的核酸构建体,其中所述绝缘子片段包括SEQ ID NO:2或其所述互补序列的多个拷贝。

6. 根据权利要求1到5中任一项所述的核酸构建体,其中所述GOI编码目的试剂,所述目的试剂是蛋白质或核酸。

7. 根据权利要求6所述的核酸构建体,其中所述目的试剂是治疗性蛋白质。

8. 根据权利要求7所述的核酸构建体,其中所述治疗性蛋白质是抗体、生长因子、细胞因子、凝血因子、酶或血红蛋白。

9. 根据权利要求6所述的核酸构建体,其中所述目的试剂是核酸,所述核酸是干扰RNA、反义寡核苷酸或微小RNA。

10. 根据权利要求1到9中任一项所述的核酸构建体,其中所述绝缘子片段位于所述GOI的上游或所述GOI的下游。

11. 根据权利要求1到9中任一项所述的核酸构建体,其中所述核酸构建体包括所述绝缘子片段的至少两个拷贝,一个拷贝位于所述GOI的上游,并且另一个拷贝位于所述GOI的下游。

12. 根据权利要求1到11中任一项所述的核酸构建体,其中所述核酸构建体包括两个GOI,并且至少一个绝缘子片段位于所述两个GOI之间。

13. 根据权利要求1到12中任一项所述的核酸构建体,其中所述核酸构建体是载体。

14. 根据权利要求13所述的核酸构建体,其中所述载体是病毒载体。

15. 根据权利要求14所述的核酸构建体,其中所述病毒载体是逆转录病毒载体、腺病毒载体或腺相关病毒载体。

16. 根据权利要求15所述的核酸构建体,其中所述载体是逆转录病毒载体,其包括5'末端重复序列(LTR)和3'LTR,其中所述绝缘子片段位于所述5'LTR内部、所述3'LTR内部、或所述5'LTR和3'LTR两者内部,并且其中所述绝缘子与所述5'LTR和/或所述3'LTR是异源的。

17. 根据权利要求15或权利要求16所述的核酸构建体,其中所述绝缘子片段位于所述5'LTR或所述3'LTR内部。

18. 根据权利要求15或权利要求16所述的核酸构建体,其中所述绝缘子片段位于所述5'LTR和所述3'LTR两者内部。

19. 根据权利要求15到18中任一项所述的核酸构建体,其中所述核酸构建体包括所述绝缘子片段的多个拷贝,所述多个拷贝中的至少一个位于所述5'LTR或所述3'LTR内部,并且所述多个拷贝中的至少一个位于所述GOI的上游或下游附近。

20. 根据权利要求15到19中任一项所述的核酸构建体,其中所述病毒载体是逆转录病毒载体,所述逆转录病毒载体是慢病毒载体或 $\gamma$ 逆转录病毒载体。

21. 根据权利要求20所述的核酸构建体,其中所述逆转录病毒载体是人免疫缺陷病毒(HIV)载体、禽白血病病毒(ALV)载体、鼠白血病病毒(MLV)载体、鼠乳腺肿瘤病毒(MMTV)载体、鼠干细胞病毒或人T细胞白血病病毒(HTLV)载体。

22. 根据权利要求15到21中任一项所述的核酸构建体,其中所述3'LTR包括上游聚腺苷酸化(polyA)增强子信号序列。

23. 根据权利要求22所述的核酸构建体,其中所述上游聚腺苷酸化(polyA)增强子信号序列是衍生自SV40晚期polyA信号序列的上游序列元件(USE)。

24. 根据权利要求22或权利要求23所述的核酸构建体,其中所述上游polyA增强子信号序列替代所述3'LTR中的U3区。

25. 根据权利要求15到24中任一项所述的核酸构建体,其进一步包括以下元件中的一种或多种:

(ix) psi ( $\Psi$ ) 包装信号;

(x) rev应答元件(RRE);

(xi) gag元件;

(xii) env剪接受体序列;

(xiii) 所述3'LTR下游的异源polyA信号序列的一个或多个拷贝;

(xiv) 一种或多种染色质绝缘子元件;

(xv) 中央多嘌呤区(cPPT);和

(xvi) 转录后调节元件(PRE)。

26. 根据权利要求25所述的核酸构建体,其中所述一种或多种染色质绝缘子元件是鸡超敏位点4元件(cHS4),和/或所述PRE是土拨鼠肝炎病毒PRE。

27. 根据权利要求15到26中任一项所述的核酸构建体,其是自灭活(SIN)逆转录病毒载体。

28. 根据权利要求1到12中任一项所述的核酸构建体,其中所述核酸构建体是用于基因编辑的基因靶构建体。

29. 一种用于将目的试剂递送到宿主细胞的方法,所述方法包括使宿主细胞与有效量的根据权利要求1到28中任一项所述的核酸构建体接触。

30. 一种用于将目的试剂递送到受试者的方法,所述方法包括向有需要的受试者施用包括由权利要求15到27中任一项所述的逆转录病毒载体产生的遗传物质的病毒颗粒。

31. 一种用于将外源基因整合到宿主细胞的基因组中的方法,所述方法包括(i)将DNA内切核酸酶和权利要求28所述的基因靶向性构建体递送到宿主细胞中,和(ii)在允许在所述宿主细胞的染色体的位点处切割,并且将所述基因靶向性构建体整合到所述染色体中的条件下,温育所述宿主细胞。

32. 根据权利要求31所述的方法,其中所述DNA内切核酸酶是CRISPR相关的RNA引导的内切核酸酶、锌指核酸酶、转录激活因子样效应子核酸酶(TALEN)或大范围核酸酶。

33. 根据权利要求31或权利要求32所述的方法,其中所述基因靶向性构建体包括侧接所述一个或多个目的基因和所述绝缘子片段的同源臂,并且其中所述同源臂与所述染色体

的切割位点同源。

34. 根据权利要求29、31和32中任一项所述的方法,其中所述接触步骤是通过向有需要的受试者施用所述核酸构建体或所述基因靶向性构建体来进行的。

35. 根据权利要求30或权利要求34所述的方法,其中所述受试者是人类患者。

36. 一种用于减少载体的基因毒性的方法,所述方法包括:

(a) 提供包括目的基因(GOI)或克隆所述GOI的位点的载体;和

(b) 插入侧接所述GOI或克隆所述GOI的所述位点的绝缘子片段,其中所述绝缘子片段包括与SEQ ID NO:2至少90%相同的核苷酸序列或其互补序列。

37. 根据权利要求36所述的方法,其中所述绝缘子片段包括与SEQ ID NO:2至少95%相同的核苷酸序列或其所述互补序列。

38. 根据权利要求37所述的方法,其中所述绝缘子片段包括SEQ ID NO:2的核苷酸序列或其所述互补序列。

39. 根据权利要求36到38中任一项所述的方法,其中所述绝缘子片段与所述载体的至少一个片段异源。

40. 根据权利要求36到39中任一项所述的方法,其中所述绝缘子片段包括SEQ ID NO:2或其所述互补序列的多个拷贝。

41. 根据权利要求36到39中任一项所述的方法,其中所述载体是病毒载体。

42. 根据权利要求40所述的方法,其中所述病毒载体是逆转录病毒载体、腺病毒载体或腺相关病毒载体。

43. 根据权利要求41所述的方法,其中所述病毒载体是包括5'LTR和3'LTR的逆转录病毒载体,并且其中所述绝缘子片段位于所述5'LTR或所述3'LTR内部。

44. 根据权利要求41所述的方法,其中所述绝缘子片段位于所述5'LTR和所述3'LTR两者内部。

45. 根据权利要求43或权利要求44所述的方法,其中所述逆转录病毒载体是慢病毒载体或 $\gamma$ 逆转录病毒载体。

46. 根据权利要求45所述的方法,其中所述逆转录病毒载体是人免疫缺陷病毒(HIV)载体、禽白血病病毒(ALV)载体、鼠白血病病毒(MLV)载体、鼠乳腺肿瘤病毒(MMTV)载体、鼠干细胞病毒或人T细胞白血病病毒(HTLV)载体。

47. 根据权利要求43到46中任一项所述的方法,其中所述逆转录病毒载体是自灭活(SIN)逆转录病毒载体。

48. 根据权利要求43到47中任一项所述的方法,其中步骤(b)是通过将所述绝缘子片段的多个拷贝插入到所述载体中来进行的,所述多个拷贝中的至少一个插入所述GOI或克隆所述GOI的所述位点的上游或下游附近,并且所述多个拷贝中的至少一个插入所述5'LTR或所述3'LTR内部。

49. 一种核酸,其包括与SEQ ID NO:2至少90%相同的核苷酸序列或其所述互补序列,其中所述核酸是绝缘子。

50. 根据权利要求49所述的核酸,其中所述核酸包括SEQ ID NO:2的核苷酸序列或其所述互补序列。

## 强绝缘子和其在基因递送中的用途

[0001] 相关申请的交叉引用

[0002] 本申请根据35U.S.C. §119(e) 要求于2017年9月18日提交的美国临时申请第62/559,751号的权益,所述申请的内容通过引用以其整体并入本文。

### 背景技术

[0003] 逆转录病毒载体(RV)已经成为基因治疗中的宝贵工具,因为它们能够进入人类宿主细胞,并且将目的基因,特别是难以递送的DNA,整合到宿主细胞中,以进行稳定的基因转移和表达。逆转录病毒载体已经有效地用于治疗性转基因的递送和整合。迄今为止,已经使用这种方法成功地治疗了多种基因疾病,例如,X连锁严重综合免疫缺陷症(X-SCID)、慢性肉芽肿病(CGD)、腺嘌呤脱氨酶缺乏症(ADA-SCID)和Wiskott-Aldrich综合征(WAS)(Hacein-Bey-Abina等人,《新英格兰医学期刊(N.Engl.J.Med.)》363:355-64,2010;Hacein-Bey-Abina等人,《临床研究期刊(J.Clin.Invest.)》118:3132-42,2008,Howe等人,《临床研究期刊》118:3143-50,2008,Stein等人,《自然医学(Nat.Med.)》16:198-204,2010,Ott等人,《自然医学》12:401-9,2006,Boztug等人,《新英格兰医学期刊》363:1918-27,2010)。

[0004] 逆转录病毒疗法的一个主要挑战是增强子介导的基因毒性或插入性肿瘤发生的高风险。如 $\gamma$ 逆转录病毒(GV)载体和慢病毒(LV)载体等逆转录病毒载体在其长末端重复序列(LTR)的U3区中具有强大的增强子,由于宿主细胞中原癌基因的转录开始位点附近的非随机整合,这可能引起原癌基因激活。这也称为插入性激活(Maruggi等人,《分子治疗(Mol.Ther.)》17:851-6,2009)。具有3'LTR U3增强子/启动子缺失和驱动转基因表达的内部较弱的细胞/内源基因启动子的自灭活(SIN)GV和LV载体已经被用于规避通过病毒增强子的插入性肿瘤发生的风险。然而,折衷是不足以用于有效的转基因治疗或表型校正的转基因表达。

[0005] 因此,非常感兴趣的是开发减少逆转录病毒载体的基因毒性的新方法,所述逆转录病毒载体用于将目的基因递送到宿主细胞。

### 发明内容

[0006] 本公开至少部分地基于从泡沫病毒(FV)的长末端报道(LTR)中发现的强绝缘子片段,当插入到逆转录病毒载体的LTR区中时,其显著降低了基因毒性。

[0007] 因此,本公开的一方面提供了一种用于基因递送的核酸构建体,所述核酸构建体包括由绝缘子片段侧接的至少一个目的基因(GOI),其中所述绝缘子片段包括与AAGGGAGACATCTAGTGATATAAGTGTGAA CTACAC(SEQ ID NO:2)至少90%(例如,至少95%)相同的核苷酸序列或其互补序列(GTGTAGTTCACA CTTATATCACTAGATGTCTCCCTT;SEQ ID NO:1)。在一些实施例中,所述绝缘子片段与所述核酸构建体的至少一个片段异源。

[0008] 在一些实施例中,绝缘子片段可以包括SEQ ID NO:2的核苷酸序列或其互补序列。例如,绝缘子片段可以包括SEQ ID NO:2或其互补序列的单个拷贝;或SEQ ID NO:2或其互

补序列的多个拷贝。

[0009] GOI可以编码目的试剂,其可以是蛋白质或核酸。在一些实施例中,目的试剂是治疗性蛋白质。实例包含但不限于抗体、生长因子、细胞因子、凝血因子、酶或血红蛋白。在其它实施例中,目的试剂可以是核酸。实例包含但不限于干扰RNA、反义寡核苷酸或微小RNA。

[0010] 在一些实施例中,绝缘子片段可以位于GOI的上游或GOI的下游。在其它实施例中,本文所述的核酸构建体可以包括绝缘子片段的至少两个拷贝。绝缘子的一个拷贝可以位于GOI的上游,并且另一个拷贝可以位于GOI的下游。在一些实例中,核酸构建体可以包括两个GOI,并且至少一个绝缘子片段位于两个GOI之间。

[0011] 在一些实施例中,本文所述的核酸构建体可以是载体,例如病毒载体(例如,逆转录病毒载体、腺病毒载体或腺相关病毒载体)。在一些情况下,病毒载体是自灭活(SIN)病毒载体。

[0012] 在一些实例中,载体是逆转录病毒载体(例如,慢病毒载体或 $\gamma$ 逆转录病毒载体),其可以包括5'长末端重复序列(LTR)和3'LTR。绝缘子片段可以位于5'LTR内部,3'LTR内部,或5'LTR和3'LTR两者内部。在一些情况下,绝缘子与5'LTR和/或3'LTR异源。示例性逆转录病毒载体包含但不限于人免疫缺陷病毒(HIV)载体、禽白血病病毒(ALV)载体、鼠白血病病毒(MLV)载体、鼠乳腺肿瘤病毒(MMTV)载体、鼠干细胞病毒或人T细胞白血病病毒(HTLV)载体。在一些情况下,本文所述的逆转录病毒载体中的任一个可以包括所述绝缘子片段的多个拷贝,所述多个拷贝中的至少一个位于所述5'LTR或所述3'LTR内部,并且所述多个拷贝中的至少一个位于所述GOI的上游或下游附近。

[0013] 本文所述的逆转录病毒载体中的任一个可以包括3'LTR,其包括上游聚腺苷酸化(polyA)增强子信号序列,例如,衍生自SV40晚期polyA信号序列的上游序列元件(USE)。在一些实例中,上游polyA增强子信号序列可以替代3'LTR中的U3区。可替代地或另外,本文公开的逆转录病毒载体可以进一步包括以下元件中的一种或多种:

[0014] (i) psi ( $\Psi$ ) 包装信号;

[0015] (ii) rev应答元件(RRE);

[0016] (iii) gag元件;

[0017] (iv) env剪接受体序列;

[0018] (v) 所述3'LTR下游的异源polyA信号序列的一个或多个拷贝;

[0019] (vi) 一种或多种染色质绝缘子元件(例如,鸡超敏位点4元件(cHS4));

[0020] (vii) 中央多嘌呤区(cPPT);和

[0021] (viii) 转录后调节元件(PRE),如土拨鼠肝炎病毒PRE。

[0022] 在其它实施例中,核酸构建体可以是用于基因编辑的基因靶构建体。

[0023] 另一个方面,本文提供了一种用于将目的试剂递送到宿主细胞的方法,所述方法包括使宿主细胞与有效量的本文所述的核酸构建体中的任一种接触。本文还提供了一种用于将目的试剂递送到受试者的方法,所述方法包括向有需要的受试者施用包括由如本文所述的逆转录病毒载体产生的遗传物质的病毒颗粒。

[0024] 进一步,本申请提供了一种用于将外源基因整合到宿主细胞的基因组中的方法,所述方法包括(i)将DNA内切核酸酶和如本文所述的基因靶向性构建体递送到宿主细胞中,和(ii)在允许在所述宿主细胞的染色体的位点处切割,并且将所述基因靶向性构建体整合

到所述染色体中的条件下,温育所述宿主细胞。用于本文描述的方法的示例性DNA内切核酸酶包含但不限于CRISPR相关的RNA引导的内切核酸酶、锌指核酸酶、转录激活因子样效应子核酸酶 (TALEN) 或大范围核酸酶。在一些实施例中,基因靶向性构建体可以包括侧接一个或多个目的基因和绝缘子片段的同源臂。同源臂与染色体的切割位点同源。

[0025] 本文所述的方法中的任一个可以通过向有需要的受试者施用所述核酸构建体或所述基因靶向性构建体来进行的。在一些实施例中,受试者是人类患者。

[0026] 又另一个方面,本公开提供了一种用于减少载体的基因毒性的方法,所述方法包括:(a) 提供包括目的基因 (GOI) 或克隆所述GOI的位点的载体;和 (b) 插入侧接所述GOI或克隆所述GOI的所述位点的本文所述的绝缘子片段中的任一个。绝缘子片段可以与载体的至少一个片段异源。在一些情况下,绝缘子片段包括SEQ ID NO:2或其互补序列的多个拷贝。在一些实施例中,载体可以是病毒载体,例如,逆转录病毒载体、腺病毒载体或腺相关病毒载体,如本文所述的那些。绝缘子片段可以在载体内部位于本文也描述的位置。

[0027] 在一些实施例中,本文所述的方法的步骤 (b) 可以通过将所述绝缘子片段的多个拷贝插入到逆转录病毒载体中进行,所述多个拷贝中的至少一个插入所述GOI或克隆所述GOI的所述位点的上游或下游附近,并且所述多个拷贝中的至少一个插入所述5'LTR或所述3'LTR内部。

[0028] 在下面的描述中阐述了本发明的一个或多个实施例的细节。从以下附图和几个实施例的具体实施方式以及所附权利要求书,本发明的其它特征或优点将变得显而易见。

## 附图说明

[0029] 以下附图形成本说明书的一部分,并且被包含以进一步说明本公开的某些方面,通过参考这些附图中的一个或多个并结合本文呈现的具体实施例的具体实施方式,可以更好地理解本公开的某些方面。为了清楚起见,并非每个部件都可以在每个附图中标记。应当理解,附图中示出的数据决不限本公开的范围。在附图中:

[0030] 图1A-1C包含示出与慢病毒 (LV) 和  $\gamma$  逆转录病毒 (GV) 载体相比,泡沫病毒 (FV) 载体的永生化和重铺效率显著较低的图。图1A:载体的前病毒形式的示意图,包含从上到下, (i) 形成脾脏病灶的病毒 (SFFV, 也缩写为SF) 长末端重复序列 (LTR) 驱动的  $\gamma$  逆转录病毒 (GV) 载体RSP91.eGFP.pre (SFFV-GV), 其含有U3 LTR区,所述区含有SFFV的启动子/增强子元件,并且驱动eGFP cDNA的表达; (ii) SIN慢病毒 (LV) 载体RRL.ppt.SFFV.eGFP.pre (SFFV-LV), 其具有400-bp LTR缺失,并且由来自eGFP的上游,置于内部的SFFV U3 LTR区的增强子-启动子元件驱动; (iii) SIN泡沫病毒载体 (SFFV-FV),  $\Delta \Phi$  SF.eGFP, 其由来自eGFP的上游,置于内部的SFFV U3 LTR区的增强子-启动子元件驱动, (iv) 鼠干细胞病毒 (MSCV) LTR驱动的GV, MSCV.eGFP.pre (MSCV-GV), 其含有U3 LTR区,所述区含有MSCV的启动子/增强子元件,并且驱动eGFP cDNA的表达; (v) SIN LV载体RRL.ppt.MSCV.eGFP.pre (MSCV-LV), 其由来自eGFP的上游,置于内部的MSCV U3 LTR区的内部MSCV启动子/增强子元件驱动; (vi) SIN FV载体  $\Delta \Phi$  MSCV.eGFP (MSCV-FV), 其由来自eGFP的上游,置于内部的MSCV U3 LTR区的增强子-启动子元件驱动;以及 (vii) 无启动子的泡沫病毒载体  $\Delta \Phi$ .eGFP (无Pr.的FV)。 $\Delta \Phi$  系列载体代表具有582-bp LTR缺失的FV载体,并且由衍生自SFFV和MSCV的内部增强子-启动子元件驱动。所有载体编码eGFP cDNA。除MSCV-LV载体 (RRL.ppt.MSCV.eGFP.pre) 以外,所

有载体的土拨鼠肝炎病毒转录后调节元件 (wPRE) 都存在于eGFP的下游。 $\Delta$ 代表具有U3缺失的LTR。实心(黑)框和空(白)框分别代表LTR中的R和U5区。图1B-1C:示出在第2周和第5周评估的永生化克隆(重铺频率)形成的图。x轴代表在2周(左图)和5周(右图)归一化到载体拷贝数(VCN)的重铺频率。y轴代表使用体外永生化(IVIM)测定测试的载体。将SFFV-FV(图1B)和MSCV-FV(图1C)的永生化潜力与SFFV-GV的永生化潜力进行比较。也将SFFV-LV(图1B)和MSCV-LV(图1C)的重铺频率与SFFV-GV进行比较。中线由黑线表示。载体主链之间的统计显著性由p值指示。\*\*\*表示 $p < 0.001$ 。

[0031] 图2A-2C包含示出CRISPR/Cas9促进GV、LV和FV前病毒序列插入到先前示出增加LM02的表达的已知的基因座中的图。图2A:示出将前病毒序列的CRISPR/Cas9插入到LM02基因中的示例性轮廓的示意图。“HDR”表示同源指导修复。图2B:示出示例性gRNA靶序列的图,所述靶序列靶向在临床上发现与插入性LM02反式激活和白血病发生相关的插入位点附近的LM02基因座,包含gRNA1(SEQ ID NO:34),gRNA2(SEQ ID NO:37),gRNA3(SEQ ID NO:40),gRNA4(SEQ ID NO:43)和gRNA5(SEQ ID NO:46)。前病毒序列的插入发生在由箭头指示的位置。图2C:示出示例性质粒构建体作为同源重组(HR)的供体模板的图,所述供体模板用于将GV、LV和FV载体序列插入到此基因座处的LM02基因中。病毒序列由对应于插入位点两侧上的区域的同源臂(HA)侧接。每个HA标签的长度为 $\sim 600$ -bp。如所指出的,所有三个载体利用SFFV启动子/增强子,并且编码eGFP cDNA。 $\Delta$ 代表具有U3缺失的LTR。实心(黑)框和空(白)框分别代表LTR中的R和U5区。

[0032] 图3A-3D包含示出在转染和通过CRISPR/Cas9介导的HR成功插入前病毒序列后,HeLa LM02克隆的建立的图,如由GFP表达所示。图3A:示出了在转染后2周通过GFP表达评估的HeLa细胞的编辑效率的图。GFP阳性细胞被分为单个细胞以建立克隆。图3B:示出对HeLa细胞克隆进行的两个示例性PCR反应以检测同源定向修复的示意图。图3C:示出跨5'同源臂(HA)桥接的PCR 1扩增子的照片。图3D:示出跨3'HA桥接的PCR 2扩增子的照片。

[0033] 图4A-4D包含示出FV诱导LM02 mRNA表达的程度小于GV或LV的图。描绘了使用两种不同的引物/探针组和两种不同的内源性对照的数据。图4A:示出Hs001534473\_m1引物/探针组和GAPDH内源性对照的图。 $N=8$ GV,9LV和11FV克隆。图4B:示出Hs001534473\_m1引物/探针组和PPIA内源性对照的图。 $N=7$ GV,8LV和8FV克隆。图4C:示出Hs00277106\_m1引物/探针组和GAPDH内源性对照的图。 $N=8$ GV,9LV和11FV克隆。图4D:示出Hs00277106\_m1引物/探针组和PPIA内源性对照的图。 $N=7$ GV,8LV和8FV克隆。

[0034] 图5是示出FV诱导LM02蛋白表达的程度小于GV或LV的照片。(−)和(+)分别表示阴性对照和阳性对照。

[0035] 图6A-6B包含示出FV的序列中大量CCCTC结合因子(CTCF)绝缘子结合位点的图。图6A:示出对GV、LV和FV的前病毒序列(SFFV、eGFP和wPRE序列除外)进行计算机分析,用于潜在的CTCF结合基序的示意图。预测的结合基序的大概位置在构建体上方标有星号。图6B:示出了预测的CTCF结合基序的位置权重矩阵(PWM)得分的图。

[0036] 图7A-7E包含示出FV前病毒序列的CTCF结合的图。图7A:示出FV克隆(FVA2)的CTCF-染色质免疫沉淀(CTFC-ChIP)和随后的定性PCR的照片,其被执行以询问CTCF与预测的结合位点的细胞内结合。图7B:示出对应于FV和LV前病毒序列中预测的CTCF结合位点的荧光标记探针结合重组CTCF蛋白,并且通过电泳迁移率分析(EMSA)解析的照片,证明了FV2

探针与CTCF之间的结合。图7C:示出FV2探针与未标记的H19探针之间的竞争性结合测定的照片。以指示的摩尔过量提供H19。图7D:示出具有预测的CTCF结合位点TGTAGTTCA, TATATCACTA (SEQ ID NO:144) 和GATGTCTCCC (SEQ ID NO:145) 的FV2探针的序列(顶部)的图。列出了突变体FV2探针(1-6),带下划线的文本表示突变的区域。图7E:示出利用突变体探针的EMSA的照片。原始的FV2探针用作阳性对照。

[0037] 图8是示出当去除绝缘子(ins.)时,FV对LM02 mRNA表达的诱导增加,并且当将绝缘子添加到LV LTR时,LV对LM02 mRNA表达的诱导减少的图。

[0038] 图9A-9C包含示出对LM02拷贝数的分析的图。图9A:示出如通过荧光原位杂交(FISH)测定确定的,与靶向LM02基因座重叠的区域的照片,揭示了HeLa对照细胞中的4个LM02等位基因。图9B:示出跨LM02(内含子1)的插入位点进行的拷贝数分析的图。HeLa对照样品由白色条形表示。图9C:示出桥接gRNA靶位点的区域的PCR扩增的照片。

## 具体实施方式

[0039] 了解病毒载体的基因毒性潜力对于设计用于基因治疗的安全且有效的载体非常重要。本公开至少部分地基于独特的绝缘子片段的意外发现,当将其插入到逆转录病毒载体中时,其显著降低了所得逆转录病毒载体的基因毒性。如本文所用,术语“基因毒性”是指逆转录病毒载体在引入逆转录病毒载体的宿主细胞的遗传物质中引起损害的性质。基因毒性包含插入基因毒性或增强子介导的基因毒性,这可能导致细胞原癌基因活化(例如,插入性肿瘤发生),从而导致癌症发展。另外,当将本文所述的绝缘子片段置于侧接转基因基因(上游或下游)时,有望减少或消除附近基因对转基因的表达的影响,或反之亦然。

[0040] 因此,本文提供了核酸构建体(例如,基因靶向性构建体或如病毒载体的载体),其包括侧接目的基因(GOI)的强绝缘子片段。当插入病毒载体,如逆转录病毒载体中时(例如在5'LTR和/或3'LTR中),绝缘子片段可以显著降低病毒载体的基因毒性潜力。此类病毒载体将是用于将目的基因递送到宿主细胞中的安全且有效的工具,使得由GOI编码的治疗剂可以发挥期望的治疗效果。可替代地或另外,当将绝缘子片段放置在GOI附近,以插入到宿主细胞的基因组中时,绝缘子片段可以减少或消除由于附近的内源基因而对GOI的表达的影响,或者减少或消除对插入GOI的位置附近的内源基因的表达的影响。在另一个实例中,当置于两个GOI之间时,绝缘子可以减少或消除由另一个基因引起的对一个基因的表达的影响。

[0041] 术语

[0042] I. 绝缘子片段

[0043] 如本文所用,术语“绝缘子片段”是指一种类型的顺式调节元件(核苷酸片段),其含有DNA结合蛋白(例如,序列特异性DNA结合蛋白)的聚集结合位点,并且介导内部和/或染色体间的相互作用。绝缘子保护基因免受不适当的顺式调节信号的影响(Ziebarth等人,《核酸研究(Nucleic Acids Res.)》41:D188-94,2013),在病毒载体的情况下,所述信号是增强子。绝缘子的非限制性实例包含CCCTC结合因子(CTCF)绝缘子、吉普赛绝缘子和 $\beta$ -球蛋白基因座。

[0044] 本文所述的绝缘子片段可以保护基因免受不适当的顺式调节信号的影响。因此,当置于两个基因之间时,绝缘子片段可以减少或消除由另一个基因引起的对一个基因的表

达的影响。另外,本文所述的绝缘子片段能够通过任何合适的机制降低基因转移载体的基因毒性,包含例如通过阻断逆转录病毒LTR中增强子的活性。除了增强子阻断活性,绝缘子还可以减少位置效应,这意味着绝缘子将防止周围的染色质影响整合的转基因表达或通过异染色质的入侵而使转基因沉默。绝缘子具有以下双重功能:防止a)载体内的增强子激活周围的细胞基因和/或b)整合载体周围的染色质中的增强子或阻遏子影响载体转基因表达。绝缘子还可以防止异染色质向载体扩散。异染色质在转录上是沉默的,并且将使由载体携带的转基因的表达失活。

[0045] 绝缘子片段可以用作增强子-阻断子和/或屏障。增强子阻断性绝缘子是位置依赖性的,并且通常起到阻断调节元件之间的通信的作用。当置于启动子与增强子之间时,增强子阻断性绝缘子是有效的,并且因此防止了由所述增强子诱导的转录(例如,通读转录)。可替代地或另外,绝缘子可以是绝缘子,其通过对异染色质形成的破坏来防止基因的沉默(West、Gaszner和Felsenfeld,《基因与发育(Genes&Development)》,16(3):271-88,2002)。

[0046] 本文公开的绝缘子可以包括与AAGGGAGACATCTAGTGATATAAGTGTG AACTACAC (SEQ ID NO:2)或其互补序列,例如GTGTAGTTCACA CTTATATCACTA GATGTCTCCCTT (SEQ ID NO:1)至少90%(例如,至少93%、至少95%、至少97%、至少98%、至少99%或以上)相同的核苷酸序列。使用Karlin和Altschul,《美国国家科学院院刊(Proc.Natl.Acad.Sci.USA)》87:2264-68,1990的算法确定两种核酸的“同一性百分比”,修改为Karlin和Altschul,《美国国家科学院院刊》90:5873-77,1993。此类算法并入到Altschul等人,《分子生物学杂志(J.Mol.Biol.)》215:403-10,1990的NBLAST程序(2.0版)中。BLAST核苷酸搜索可以用NBLAST程序进行,得分=100,字长-12,以获得与本文公开的核酸分子同源的核苷酸序列。在两个序列之间存在缺口的情况下,可以如Altschul等人,《核酸研究》25(17):3389-3402,1997中所述的利用带缺口的BLAST。当使用BLAST和Gapped BLAST程序时,可以使用相应程序(例如,NBLAST)的默认参数。

[0047] 在一些实施例中,本文所述的绝缘子片段包括SEQ ID NO:2的核苷酸序列或其互补序列。此类绝缘子片段可以含有一个拷贝的SEQ ID NO:2或其互补序列、其功能变体(例如,与如本文所述的SEQ ID NO:2具有至少90%的序列同一性)或其互补序列。可替代地,其可以含有多个拷贝(例如2个、3个、4个或5个拷贝),所述拷贝中的每两个拷贝可以以串联重复的形式直接连接,或通过核酸连接子间接连接。

[0048] 本文所述的绝缘子片段的最大长度可以为约500个碱基对(bps),例如,约450bps、约400bps、约350bps、约300bps、约250bps、约200bps、约150bps、约100bps或约50bps。在一个实例中,本文所述的绝缘子片段由SEQ ID NO:2或其互补序列组成。通常,术语“约”和“大约”是指在特定值的可接受的误差范围内,如本领域的普通技术人员所确定的。“约”可以表示小于给定值的 $\pm 30\%$ ,优选地小于 $\pm 20\%$ ,更优选地小于 $\pm 10\%$ ,更优选地小于 $\pm 5\%$ ,并且仍更优选地小于 $\pm 1\%$ 的范围。

[0049] 本文所述的绝缘子可能能够结合CCCTC结合因子(CTCF),所述CCCTC结合因子在基因绝缘和增强子阻断(Schmidt等人,《细胞(Cell)》148:335-48,2012),以及提供脊椎动物中的染色质屏障功能(Cuddapah等人,《基因组研究(Genome Res.)》19:24-32,2009)中起重要作用。通过11个锌指的各种组合,CTCF结合不同的DNA序列(Ziebarth等人,《核酸研究》41:D188-94,2013)。除了其绝缘子和屏障功能外,CTCF结合位点已经被证明减少病毒载体

的基因毒性,而不降低病毒滴度(Liu等人,《自然生物技术(Nat.Biotechnol.)》33:198-203,2015)。不受理论的束缚,本文公开的绝缘子片段可以提供CTCF的结合位点,其在绝缘子位点处与逆转录病毒载体结合后,将充当增强子阻断子,从而降低逆转录病毒载体的基因毒性潜力。已经报道了CTCF用于控制包含HPV和HTLV-1的各种类型的病毒中病毒启动子的活性,Goodman等人,《病毒学期刊(J.Virology)》92(1):e01639-17,2018,其相关的公开内容出于本文引用的目的或主题通过引用并入。因此,预期绝缘子片段在各种类型的逆转录病毒载体,如本文所述的那些中表现出预期的绝缘子功能。

[0050] 本文所述的绝缘子片段可以包括SEQ ID NO:2的基序的一个拷贝,其功能变体(例如,与如本文所述的SEQ ID NO:2具有至少90%序列同一性)或其互补序列。本文公开的绝缘子片段中的任一个也都在本公开的范围之内。

[0051] II. 包括绝缘子片段的核酸构建体

[0052] 本文公开的绝缘子片段中的任一个可以用于制备核酸构建体,其可以携带一个或多个目的基因(GOI)。包含绝缘子片段可以保护GOI免受不适当的顺式调节信号的干扰,例如位于GOI或携带此类的载体插入的基因组基因座处的那些,或者保护局部内源性免受控制GOI的表达的顺式调节信号的影响。当核酸构建体携带多个GOI时,在两个GOI之间包含绝缘子片段也可以使一个不受另一个影响。

[0053] 因此,本文还公开了用于递送由核酸构建体携带的转基因(目的基因或GOI)的核酸构建体,使得转基因的表达不受转基因将插入的染色体位点附近的内源性基因的影响,或反之亦然。当核酸构建体含有多个转基因时,将绝缘子片段置于两个转基因之间还可以减少或消除一个基因的表达对另一个的影响。

[0054] 本文公开的核酸构建体可以是任何类型的核酸分子,其包括一个或多个GOI和侧接至少一个GOI的一个或多个绝缘子片段,例如,GOI的上游或GOI的下游,或使一个位于GOI的上游和一个位于GOI的下游。绝缘子片段可以直接连接到GOI。可替代地,可以将绝缘子片段通过核酸间隔子连接到GOI。此类核酸间隔子的范围可以为10bps到2kbps,例如50bps-2kbps、100bp-1.5kbps、150bps-1kbps、200bps-500bps或500bps-1kbps。

[0055] 在一些情况下,绝缘子片段位于GOI附近,GOI的上游和/或下游。绝缘子片段可以直接连接到GOI或通过短核酸间隔子连接到GOI,其可以含有5bp-200bp,例如5-100bps、10-150bps、20-100bps、10-50bps、20-50bps或20-30bps。

[0056] 本文所述的核酸构建体可以含有一个GOI,其由也公开于本文的GOI上游或下游的一个绝缘子片段侧接。可替代地,核酸构建体可以含有两个或更多个GOI,并且如本文所述的绝缘子片段可以位于两个GOI之间。此类核酸构建体可以进一步含有侧接下游和/或上游的GOI的额外的绝缘子片段。

[0057] 在一些实施例中,本文公开的核酸构建体可以是基因靶构建体,其可以通过基因编辑插入由基因靶构建体携带的GOI。基因靶构建体可以是线性核酸分子,其可以进一步在整个构建体的5'和3'末端两者或侧接GOI处包括同源臂。同源臂含有与基因靶构建体旨在插入的基因组位点同源的序列,以便于通过同源重组插入基因靶构建体。本文公开的基因靶构建体可以含有多个GOI,并且一个绝缘子片段可以位于两个GOI之间。

[0058] 在一些实施例中,本文公开的核酸构建体可以是载体。“载体”是指能够转运已经与其连接的另一个核酸的核酸分子。载体能够自主复制或整合到宿主细胞的基因组中。在

一些情况下,本文公开的载体可以是病毒载体,其含有衍生自病毒基因组的元件(天然存在或经修饰的)。病毒载体可以用于将遗传物质(例如,转基因)递送到合适的宿主细胞中。病毒载体的非限制性实例包含但不限于逆转录病毒载体(例如慢病毒载体或 $\gamma$ 逆转录病毒载体)、腺病毒载体、腺相关病毒载体(AAV)和杂种载体(含有来自不同的病毒基因组的组分)。在美国专利第5,698,443号、美国专利第5,650,309号和美国专利第5,827,703号中提供了病毒载体的额外实例,出于本文引用的目的和主题,所述专利中的每一个的相关公开内容通过引用并入本文。

[0059] 在一些实例中,本文公开的核酸构建体是逆转录病毒载体。在一些情况下,可以将如本文公开的至少一个绝缘子片段插入到逆转录病毒载体中的5'LTR、3'LTR或两者中,以减少逆转录病毒载体的基因毒性。逆转录病毒载体是DNA分子,其含有可以容纳目的基因的前病毒序列(例如,LTR序列、Psi( $\psi$ )序列和/或启动子/增强子序列),以允许两者并入到靶细胞中。前病毒序列衍生自病毒基因组,并且经修饰使得它们可以用作携带和转移遗传物质的质粒媒介。还对前病毒序列进行了修饰,以去除必需的病毒基因和安全问题。通常,逆转录病毒载体不能自我增殖和/或包装,以在不提供必需的病毒蛋白/基因的辅助病毒的情况下,产生病毒颗粒。

[0060] 本文所述的逆转录病毒载体包括5'长末端重复序列(LTR)、3'LTR和本文所述的绝缘子片段中的任一个,可以将其插入到LTR区中的一个或两个中。另外,逆转录病毒载体可以包括额外的病毒或非病毒元件,以促进如本文所述的预期的病毒载体功能。

[0061] 绝缘子片段可以位于本文所述的逆转录病毒载体中的5'LTR、3'LTR或5'LTR和3'LTR两者的内部。例如,当逆转录病毒载体是泡沫逆转录病毒载体时,绝缘子片段可以与5'LTR、3'LTR或两者异源。如本文所用,术语“异源的”是指衍生自不同的病毒物种的两个序列。换句话说,绝缘子序列和LTR序列中的一个或两个可能不是衍生自相同的逆转录病毒基因组。

[0062] LTR区通常位于逆转录病毒载体的相对两端上,所述逆转录病毒载体可以是线性DNA分子。在一些实施例中,逆转录病毒载体的LTR包括U3区、R区和U5区。在一些情况下,5'LTR、3'LTR或两者中的U3区可以包括增强子/启动子元件,其可以驱动逆转录病毒载体内基因的表达。这些增强子/启动子元件可以用作增强子、启动子或两者。此类逆转录病毒载体通常被称为LTR驱动的载体(Maetzig等人,《病毒(Viruses)》3(6):677-713,2011)。在其它情况下,5'LTR、3'LTR或两者可以具有缺失的U3区、R区和U5区中的一个或多个(例如,自灭活的载体,如下文所述的那些)。

[0063] 在一些实施例中,3'-LTR可以进一步包括聚腺苷酸化(poly A)增强子信号序列,其位于切割/聚腺苷酸化(polyA)位点(例如,AAUAAA)的上游,并且起增加polyA位点效率的作用,并且因此增加聚腺苷酸化效率。示例性的聚腺苷酸化增强子信号序列包含来自合适的病毒基因,例如猿猴病毒40(SV40)晚期基因的上游序列元件(USE)。以下提供了SV40晚期基因的示例性USE元件的核苷酸序列:

[0064] TTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAAACCG (SEQ ID NO:3)。

[0065] 包含此类polyA增强子信号序列可以促进转录终止,并且减少载体转录物的通读,并且改善包装效率,这将导致病毒滴度增加。

[0066] 除了本文所述的绝缘子片段和LTR之外,逆转录病毒载体还包括逆转录病毒载体

的基本功能所必需的组分,例如,能够被复制、包装到病毒颗粒中和/或能够驱动从而宿主细胞中携带的目的基因的表达。用于构建逆转录病毒载体的此类必需元件是本领域技术人员众所周知的。

[0067] 在一些实施例中,本文所述的逆转录病毒载体可以包括以下组分中的一种或多种:(i) psi ( $\Psi$ ) 包装信号;(ii) rev应答元件(RRE);(iii) gag元件;(iv) env剪接受体序列;(v) 3'LTR下游的异源polyA信号序列的一个或多个拷贝;(vi) 一种或多种染色质绝缘子元件;(vii) 中央多嘌呤区(cPPT);以及(viii) 转录后调节元件(PRE)。

[0068] psi ( $\Psi$ ) 包装信号,也被称为衣壳化序列,调节复制期间逆转录病毒RNA进入到病毒衣壳中的包装。通常将其置于逆转录病毒载体中5'长末端重复序列的下游,以有效地包装和递送由逆转录病毒载体携带的转基因。以下提供了示例性 $\Psi$ 包装信号的核苷酸序列:

[0069] TCGACGCAGGACTCGGCTTGCTGAAGCGGCACGGCAAGAGGCGAGGGCGGCGACTGGTGAGTACGC CAAAAATTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCAGTATTAAGCGGGGAGAA. (SEQ ID NO:4)

[0070] rev应答元件(RRE)是位于env区中的结构域。RRE在“env基因”内最多可以具有360个核苷酸长。Rev蛋白结合RRE,以调节病毒基因的表达。Rev/RRE系统有助于mRNA的核输出。以下提供了env基因中示例性RRE的核苷酸序列:

[0071] GATCTTCAGACCTGGAGGAGGAGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAG TAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGG AATAGGAGCTTTGTTCCCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTA CAGGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATC TGTTGCAACTCACAGTCTGGGCGATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGAAAAGATACCTAAAGGATCA ACAGCTCCTGGGGATTTGGGGTTGCTCTGAAAACACTCATTTGCACCACTGCTGTGCCTTGAATGCTAGTTGGAGT AATAAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCT TAATACACTCCTTAATTGAAGAATCGAAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAATG GGCAAGTTTGTGGAATTGGTTTAACATAACAAATTGGCTGTGGTATATAAAATTTATTCATAATGATAGTAGGAGGC TTGGTAGGTTTAAGAATAGTTTTTGTGCTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGT TTCAGACCCACCTCCCAACCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAG AGACAGATCCATTCGATTAGTGAACGGATC. (SEQ ID NO:5)。

[0072] gag (组特定抗原) 元件编码逆转录病毒,即,基质、衣壳和核衣壳组分的结构蛋白(或其部分)。在一些情况下,本文所述的逆转录病毒载体可以含有gag片段,其是gag基因的5'片段。此类片段可以含有250-650bps(例如,约360bps或600bps)。含有此类短gag片段可以增强携带大目的基因(例如,球蛋白基因)的逆转录病毒载体的病毒滴度。参见例如US20150316511,相关公开内容通过引用并入本文。在其它情况下,本文所述的逆转录病毒载体可以不含任何gag片段。以下提供了示例性gag片段的核苷酸序列:

[0073] ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGAGAATTAGATCGCGATGGGAAAAAATTCGGTTAAGG CCAGGGGGAAAGAAAAAATATAAATTTAAACATATAGTATGGGCAAGCAGGGAGCTAGAACCATTTCAGGTTAATC CTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGA AGAACTTAGATCATTATATAATACAGTAGCAACCCTCTATTGTGTGCATCAAAGGATAGAGATAAAAGACACCAAG GAAGCTTTAGACAAGATAGAGGAAGAGCAAAAACAAAAGTAAGACCACCGCACAGCAAGC. (SEQ ID NO:6)

[0074] env剪接受体序列是逆转录病毒基因组中pol1编码区的3'末端附近的核苷酸序列。剪接受体序列调节转录物的剪接。它还使得能够表达env编码区。剪接受体以粗体突出显示,并且在下面的env序列中带有下划线。

[0075] GATCTTCAGACCTGGAGGAGGATATGAGGGACAATTGGAGAAGTGAATTATATAAATATAAAGTAGTAAAAATTGAACCATTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGGTGCAGAGAGAAAAAGAGCAGTGGGAATAGGAGCTTTGTTCTTGGGTTCTTGGGAGCAGCAGGAAGCACTATGGGCGCAGCGTCAATGACGCTGACGGTACAGCCAGACAATTATTGTCTGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAGCATCTGTTGCAACTCACAGTCTGGGGCATCAAGCAGCTCCAGGCAAGAATCCTGGCTGTGGAAGATACCTAAAGGATCAACAGCTCCTGGGGATTTGGGGTTGCTCTGGAAAACCTCATTTGCACCACTGCTGTGCCTTGAATGCTAGTTGGAGTAAATAATCTCTGGAACAGATTTGGAATCACACGACCTGGATGGAGTGGGACAGAGAAATTAACAATTACACAAGCTTAATACACTCCTTAATTGAAGAATCGCAAACCAGCAAGAAAAGAATGAACAAGAATTATTGGAATTAGATAAATGGCAAGTTTGTGGAATTGGTTAACATAACAAATTGGCTGTGGTATATAAAATTTATTCATAATGATAGTAGGAGGCTTGGTAGGTTTAAAGAATAGTTTTGCTGTACTTTCTATAGTGAATAGAGTTAGGCAGGGATATTCACCATTATCGTTTCAGACCCACCTCCCAACCCCGAGGGGACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGAGAGACAGAGACAGATCCATTCGATTAGTGAACGGATC. (SEQ ID NO:5)。

[0076] 在一些情况下,逆转录病毒载体可以包括一个或多个异源polyA信号传导位点,其可以位于3'LTR的下游。此类异源polyA信号传导位点可能不是病毒来源的(例如,来自非病毒基因,如β珠蛋白基因)。可替代地,异源polyA信号传导位点可以衍生自与含有异源polyA信号传导位点的逆转录病毒载体不同的病毒物种的病毒基因。包含此类异源polyA信号传导位点可以增强聚腺苷酸化效率,从而进一步减少载体转录物的通读,并且改善包装效率,这将导致病毒滴度增加。以下提供了来自牛生长激素基因的示例性polyA信号传导序列的核苷酸序列:

[0077] CCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG (SEQ ID NO:7)。

[0078] 在一些实施例中,逆转录病毒载体可以包含一种或多种染色质绝缘子元件。染色质绝缘子是抵抗异染色质形成的启动子或增强子序列。在一些实施例中,染色质绝缘子可以是长度为约1kb的片段,其阻断增强子的转录激活。如本文所述,其可以充当屏障元件,以尤其防止异染色质的扩散和基因沉默,减少染色质位置效应,并且具有增强子阻断活性。这些特性对于通过随机地整合的载体实现一致的可预测表达和安全转基因递送而言是理想的。绝缘载体具有减少的染色质位置效应,并且提供一致的并且因此改善的总体表达。

[0079] 在一些实例中,本文所述的逆转录病毒载体中的所述一种或多种染色质绝缘子元件可以是鸡超敏位点4元件(cHS4),其是来自鸡β-珠蛋白基因座控制区的染色质绝缘子。Arumugam等人,《公共科学图书馆·综合(PLoS ONE)》4(9):e6995,2009。在一些情况下,可以将来自鸡p-珠蛋白基因座的一个或多个全长染色质绝缘子(约1.2kb)的超敏位点4(cHS4)插入3'LTR中,以使其复制到γ逆转录病毒和慢病毒载体中的5'LTR中。在其它情况下,包括~250bp核心的截短的cHS4片段可以用于本文所述的逆转录病毒载体中。此类核心片段可以与来自cHS4元件的3'~400-bp片段组合。在一个实例中,包含核心序列和3'-片段的约650个碱基对的功能性长度减小的绝缘子可以用于构建本文所述的逆转录病毒载体。

此类cHS4来源的绝缘子序列描述于US20150316511中,相关公开内容通过引用并入本文。以下提供了示例性全长cHS4的核苷酸序列:

[0080] GAGCTCACGGGGACAGCCCCCCCCAAAGCCCCCAGGGATGTAATTACGTCCCTCCCCCGCTAGGGGG  
CAGCAGCGAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCCCCGAGCCGGCAGCGTGCGGGGACA  
GCCCGGGCACGGGAAGGTGGCACGGGATCGCTTTCCTCTGAACGCTTCTCGCTGCTCTTTGAGCCTGCAGACACC  
TGGGGGGATACGGGAAAAAGCTTTAGGCTGAAAGAGAGATTTAGAATGACAGAATCATAGAACGGCCTGGGTTGC  
AAAGGAGCACAGTGCTCATCCAGATCCAACCCCTGCTATGTGCAGGGTCATCAACCAGCAGCCCAGGCTGCCAG  
AGCCACATCCAGCCTGGCCTTGAATGCCTGCAGGGATGGGGCATCCACAGCCTCCTGGGCAACCTGTTTCAGTGCG  
TCACCACCCTCTGGGGAAAAACTGCCTCCTCATATCCAACCCAAACCTCCCCTGTCTCAGTGTAAGCCATTCCC  
CCTTGTCTATCAAGGGGGAGTTTGTGTGACATTGTTGGTCTGGGGTGACACATGTTTGCCAATTCAGTGCATCA  
CGGAGAGGCAGATCTTGGGGATAAGGAAGTGCAGGACAGCATGGACGTGGGACATGCAGGTGTTGAGGGCTCTGGG  
ACACTCTCCAAGTCACAGCGTTCAGAACAGCCTTAAGGATAAGAAGATAGGATAGAAGGACAAAGAGCAAGTTAAA  
ACCCAGCATGGAGAGGAGCACAAAAAGGCCACAGACACTGCTGGTCCCTGTGTCTGAGCCTGCATGTTTGATGGTG  
TCTGGATGCAAGCAGAAGGGGTGGAAGAGCTTGCCTGGAGAGATACAGCTGGGTGAGTAGGACTGGGACAGGCAGC  
TGGAGAATTGCCATGTAGATGTTTCATACAATCGTCAAATCATGAAGGCTGGAAAAGCCCTCCAAGATCCCCAAGAC  
CAACCCCAACCCACCCACCGTGCCACTGGCCATGTCCCTCAGTGCCACATCCCCACAGTTCTTCATCACCTCCAG  
GGACGGTGACCCCCCACCTCCGTGGGACAGCTGTGCCACTGCAGCACCGCTCTTTGGAGAAGGTAAATCTTGCTAA  
ATCCAGCCCGACCCTCCCCTGGCACAACGTAAGGCCATTATCTCTCATCCAACCTCCAGGACGGAGTCAGTGAGAAT  
ATT (SEQ ID NO:8)

[0081] 以下是来自cHS4的250-bp核心和3'~400-bp片段的核苷酸序列:

[0082] 250-bp核心:

[0083] GGGACAGCCCCCCCCAAAGCCCCCAGGGATGTAATTACGTCCCTCCCCCGCTAGGGGGCAGCAGCGA  
GCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCCCCGAGCCGGCAGCGTGCGGGACAGCCCGGGCA  
CGGGGAAGGTGGCACGGGATCGCTTTCCTCTGAACGCTTCTCGCTGCTCTTTGAGCCTGCAGACACCTGGGGGGAT  
ACGGGGAAAAAGCTTTAGGCTGAAAGAGAT (SEQ ID NO:9)

[0084] 3'~400-bp片段:

[0085] CTGAGCCTGCATGTTTGATGGTGTCTGGATGCAAGCAGAAGGGGTGGAAGAGCTTGCCTGGAGAGATA  
CAGCTGGGTGAGTAGGACTGGGACAGGCAGCTGGAGAATTGCCATGTAGATGTTTCATACAATCGTCAAATCATGAA  
GGCTGGAAAAGCCCTCCAAGATCCCCAAGACCAACCCCAACCCACCCACCGTGCCACTGGCCATGTCCCTCAGTG  
CCACATCCCCACAGTTCTTCATCACCTCCAGGGACGGTGACCCCCCACCTCCGTGGGACAGCTGTGCCACTGCAGC  
ACCGCTCTTTGGAGAAGGTAAATCTTGCTAAATCCAGCCCGACCCTCCCCTGGCACAACGTAAGGCCATTATCTCT  
CATCCAACCTCCAGGACGGAGTCAGTGAG (SEQ ID NO:10)

[0086] 其它染色质绝缘子的非限制性实例包含ArsI (衍生自海胆芳香基硫酸酯酶基因位点)、sns5 (衍生自海胆H2A早期组蛋白基因)、锚蛋白1基因启动子元件和果蝇吉普赛元件 (Emery,《人类基因治疗 (Human Gene Therapy)》22 (6):761-74,2011)。

[0087] 中央多嘌呤区 (cPPT) 指导病毒颗粒穿透核膜。在逆转录病毒复制中,它充当合成正链DNA的引物。已经示出当掺入到逆转录病毒载体中时,增加转导效率和转基因表达。示例性cPPT的核苷酸序列如下所示:AAAAGAAAAGGGGGGA (SEQ ID NO:11)。

[0088] 转录后调节元件 (PRE) 是一种序列,在转录时增强转基因在病毒载体中的表达。已

经示出当掺入到逆转录病毒载体中时,增加转导效率和转基因表达。以下提供了示例性PRE的核苷酸序列:

[0089] TCGACAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGATAAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCAC TGTGTTTGTGACGCAACCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCTTCCGGGACTTTCGCTTTC CCCCTCCCTATTGCCACGGCGGAATCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCA CTGACAATTCCGTGGTGTGTCGGGGAAGCTGACGTCCTTTCCATGGCTGCTCGCCTGTGTTGCCACCTGGATTCT GCGCGGGACGTCCTTCTGCTACGTCCTTTCGGCCCTCAATCCAGCGGACCTTCTTCCCGCGGCCTGCTGCCGGCT CTGCGGCCTCTTCCGCGTCTTCGCTTTCGCCCTCAGACGAGTCGGATCTCCCTTGGGC (SEQ ID NO:12)

[0090] 在一些实施例中,逆转录病毒载体中使用的PRE是来自乙型肝炎病毒 (HPRE) 的PRE或来自土拨鼠肝炎病毒 (WPRE) 的PRE。在一些实施例中,逆转录病毒载体中有不止一个PRE,并且不止一个PRE可以是HPRE、WPRE或其混合物。在一个实施例中,逆转录病毒载体不包含PRE。

[0091] 本文所述的逆转录病毒载体可以进一步包括如本领域中已知的额外功能元件,以解决安全性和/或改善载体功能,如包装效率和/或病毒滴度。可以在US20150316511和W02015/117027中找到额外信息,所述文献中的每一个的相关公开内容出于本文引用的目的和主题通过引用并入本文。

[0092] 本文所述的逆转录病毒载体可以包括来自一种或多种合适的逆转录病毒的病毒元件,如本文所述的那些,所述逆转录病毒载体是具有单链正义RNA分子的RNA病毒。逆转录病毒包括逆转录酶和整合酶。进入到靶细胞中后,逆转录病毒利用其逆转录酶将其RNA分子转录为DNA分子。随后,整合酶用于将DNA分子整合到宿主细胞基因组中。在整合到宿主细胞基因组中时,来自逆转录病毒的序列被称为前病毒(例如,前病毒序列或前病毒序列)。这种有效的基因转移机制已经使逆转录病毒载体成为基因治疗中非常有价值的工具,因为它们可以用于宿主细胞中的长期转基因表达。

[0093] 逆转录病毒包含7个家族: $\alpha$ 逆转录病毒(禽白血病病毒)、 $\beta$ 逆转录病毒(小鼠乳腺肿瘤病毒)、 $\gamma$ 逆转录病毒(鼠白血病病毒)、 $\delta$ 逆转录病毒(牛白血病病毒)、 $\epsilon$ 逆转录病毒(角膜皮肤肉瘤病毒)、慢病毒(人类免疫缺陷病毒1)和泡沫病毒(人类泡沫病毒)。美国专利第7,901,671号提供了逆转录病毒的六个额外的实例。来自合适的逆转录病毒的病毒元件,如本文所述的那些,可以用于构建本文所述的逆转录病毒载体。本文所述的逆转录病毒载体可以是慢病毒载体或 $\gamma$ 逆转录病毒载体。逆转录病毒载体的非限制性实例包含人免疫缺陷病毒(HIV)载体、禽白血病病毒(ALV)载体、鼠白血病病毒(MLV)载体、鼠乳腺肿瘤病毒(MMTV)载体、鼠干细胞病毒和人T细胞白血病病毒(HTLV)载体。这些逆转录病毒载体包括来自对应的逆转录病毒的前病毒序列。

[0094] 本文所述的逆转录病毒载体可以通过常规重组技术制备。在一些实例中,可以将如本文所述的那些的绝缘子片段插入到逆转录病毒载体的合适位置中,以降低所得逆转录病毒载体的基因毒性。例如,可以通过常规技术,将绝缘子片段插入5'LTR内部、3'LTR内部或5'LTR和3'LTR两者内部。当需要时,可以将额外的绝缘子片段插入逆转录病毒载体内部的合适位点处,例如,邻近由逆转录病毒载体携带的转基因。如本文所用,术语“插入”是指

通过使用例如限制性消化和连接或重组,将核苷酸的序列添加到逆转录病毒载体的过程。用于将序列插入到逆转录病毒载体中的技术对本领域技术人员而言是显而易见的。

[0095] 在一些情况下,通过常规克隆技术,将绝缘子片段插入到逆转录病毒载体中。《分子生物学实验指南(Current Protocols in Molecular Biology)》,Ausubel,F.M.等人,纽约:约翰威立国际出版公司,2006;《分子克隆实验室手册(Molecular Cloning:A Laboratory Manual)》,Green,M.R.和Sambrook J.,纽约:冷泉港实验室出版社,2012;Gibson,D.G.等人,《自然方法(Nature Methods)》6(5):343-345(2009),其关于分子克隆的教导通过引用并入本文。可替代地,可以通过基因编辑,例如通过CRISPR来实施绝缘子片段的插入。

[0096] 在常规基因治疗中,已经越来越多地使用具有3'LTR的自灭活(SIN)GV和LV载体来规避通过病毒增强子的插入性肿瘤发生的风险。这些SIN GV和LV载体具有U3增强子/启动子缺失和驱动转基因表达的内部较弱的细胞/内源基因启动子。这使长末端重复序列(LTR)的U3区中普遍存在的活性增强子缺失。与体外和体内两者实验系统中的LTR-完整GV载体相比,这些SIN“LTR少”或“增强子少”载体示出降低的基因毒性(Modlich等人,《血液(Blood)》108:2545-53,2006,Zychlinski等人,《分子治疗》16:718-25,2008,Montini等人,《临床研究杂志(J Clin Invest)》119:964-75,2006)。然而,转基因的表达通常不稳健,并且疾病类型的成功和完全校正很大程度上取决于每个细胞引入大量转导/载体拷贝数(VCN),但是除了适中水平的转基因表达足以校正的疾病。

[0097] 在本公开的一个实施例中,逆转录病毒载体是自灭活(SIN)逆转录病毒载体。如本文所用,SIN载体是逆转录病毒载体,其U3区中的至少一个已经缺失。在一个实施例中,SIN载体的5'LTR的U3区已经缺失。在替代性实施例中,SIN载体的3'LTR的U3区已经缺失。在又一个实施例中,SIN载体的5'LTR和3'LTR两者中的U3区已经缺失。如图1A所示,逆转录病毒载体的一些实施例具有内部启动子区。在一些实施例中,所述内部启动子区可操作地连接到GOI。在一些实施例中,内部启动子区在非SIN载体中。在一些实施例中,内部启动子区是内源或异源的U3区。如在内部启动子的上下文中所用,术语“内源的”是指内部启动子区来自相同的逆转录病毒载体或来自相同物种的逆转录病毒载体。如在内部启动子的上下文中所用,术语“异源的”是指所述内部启动子来自不同的逆转录病毒载体或不同物种的逆转录病毒载体。在一些实施例中,U3区被异源U3区取代。在此特定实施例中,术语“异源的”是指经取代的U3区是与逆转录病毒载体的R和/或U5区不同的物种。

[0098] 本领域普通技术人员将知道,构建体中绝缘子元件与启动子和报告基因的距离可以根据逆转录病毒载体中使用的GOI或内部启动子的相对大小而变化。

[0099] III. 绝缘子片段在有效递送转基因中的应用

[0100] 当引入到宿主细胞中,例如插入到宿主细胞的基因组中时,本文所述的核酸构建体中的任一个还可以发挥本文所述的绝缘子功能。因此,本文提供了用于在体外或体内,将可以编码目的试剂的GOI递送到宿主细胞中的方法。所公开的方法可以应用于实验、兽医和医学领域。

[0101] 在一些实施例中,将本文所述的核酸构建体中的任一个(例如基因靶构建体或病毒载体,如逆转录病毒载体)转染到合适的宿主细胞中,以产生病毒颗粒。用于将核酸构建体转导到宿主细胞,如哺乳动物细胞中的技术在本领域中是成熟的。美国专利第5,399,346

号提供了一些实例。核酸转染的方法在本领域中是成熟的,并且其范围为从化学、生物学到物理方法。化学方法包含但不限于,磷酸钙转染、阳离子聚合物转染、脂转染、FUGENE®和DEAE-葡聚糖介导的转染。其它转染的方法包含但不限于电穿孔、声纳穿孔、细胞挤压、变质、光学转染、原生质体融合、磁转染™和粒子轰击。

[0102] 当核酸构建体是病毒载体,如逆转录病毒载体时,宿主细胞可以是表达病毒结构和/或辅助蛋白(例如逆转录病毒结构和/或辅助蛋白)的包装细胞,例如gag、pol、env、tat、rev、vif、vpr、vpu、vpx和/或nef。病毒包膜蛋白(env)确定了病毒颗粒可以感染的宿主细胞的范围,并且通过由包装细胞系产生的重组逆转录病毒进行转化。在慢病毒,如HIV-1、HIV-2、SIV、FIV和EIV的情况下,env蛋白包含gp41和gp120。在一些情况下,编码病毒env蛋白的基因可以与编码病毒gag和pol的那些在单独的载体上。在其它情况下,编码env、pol和gag的基因可以位于同一载体上。可以将此类载体转染到合适的宿主细胞中,以稳定表达病毒蛋白。

[0103] 包装细胞在其遗传物质中不含有包装信号,并且能够表达(例如,稳定地)病毒结构蛋白、复制酶(例如,gag、pol和env)以及其它包装病毒颗粒必需的蛋白。任何合适的细胞系,例如哺乳动物细胞系,可以用于制备包装细胞。实例包含CHO细胞、BHK细胞、MDCK细胞、COS细胞、VERO细胞、3T3细胞、NIH3T3细胞、HepG2细胞、HeLa细胞、293细胞、293T细胞或A549细胞。

[0104] 由包装细胞制备病毒储备溶液的方法在本领域中是已知的,并且例如由Y. Soneoka等人,《核酸研究》23:628-633,1995和N.R. Landau等人,《病毒学期刊(J. Virol.)》66:5110-5113,1992中所示。可以使用常规技术,从包装细胞中收集传染性病毒颗粒。例如,如本领域已知的,可以通过细胞裂解或细胞培养物的上清液的收集,收集传染性颗粒。如果需要,可以使用常规技术,纯化收集的病毒颗粒。

[0105] 由此产生的包括从本文所述的逆转录病毒载体中的任一个转录的RNA分子的病毒颗粒可以用于感染合适的宿主细胞,从而递送由逆转录病毒载体携带的目的基因。在一些实例中,可以使病毒颗粒与细胞培养物中的宿主细胞接触,用于表达由体外目的基因编码的目的试剂。在其它实例中,可以通过合适的途径(例如静脉内注射或局部注射)向需要治疗的受试者(例如人类受试者)施用病毒颗粒。在一些实例中,可以通过以下向需要治疗的受试者施用病毒颗粒:静脉内、皮内、动脉内、病变内、肿瘤内、颅内、关节内、前列腺内、胸膜内、气管内、鼻内、玻璃体内、阴道内、直肠内、局部、肌肉内、腹膜内、皮下、结膜下、囊内、粘膜、心包内或脐带内;或通过其它方法或前述方法的任何组合,如本领域普通技术人员已知的(参见,例如,《雷明顿氏药物科学(Remington's Pharmaceutical Sciences)》(1990))。

[0106] 可以通过合适的途径,将核酸构建体或携带衍生自核酸构建体的遗传物质(例如,从逆转录病毒载体转录的RNA分子)的病毒颗粒中的任一个递送到需要治疗的受试者。在其它实施例中,本文所述的核酸构建体中的任一个可以用于体外或离体操作细胞,其中可以向需要的受试者施用经操作的细胞。

[0107] 在一些实施例中,受试者是哺乳动物。在一些实施例中,受试者是人类或人类患者。在一些实施例中,受试者是动物(例如动物模型)。在其它实施例中,受试者是小鼠。受试者还包含动物,如家庭宠物(例如狗、猫、兔子、雪貂等)、牲畜或农场动物(例如牛、猪、绵羊、鸡和其它家禽)、马,如纯种马、实验动物(例如大鼠、兔子等)等。

[0108] 在一些实施例中,本文公开的逆转录病毒载体用于递送目的试剂,其由逆转录病毒载体中包含的目的基因编码。递送的示例性病毒方法包括但不限于重组逆转录病毒(参见,例如,PCT公开第WO 90/07936号;第WO 94/03622号;第WO 93/25698号;第WO 93/25234号;第WO 93/11230号;第WO 93/10218号;第WO 91/02805号;美国专利第5,219,740号和第4,777,127号;GB专利第2,200,651号和EP专利第0345242号)、基于 $\alpha$ 病毒的载体和腺相关病毒(AAV)载体(参见,例如,PCT公开第WO 94/12649号、第WO 93/03769号;第WO 93/19191号;第WO 94/28938号;第WO 95/11984号和第WO 95/00655号)。

[0109] 由本文所述的核酸构建体携带的GOI可以编码目的试剂,其可以是治疗剂。当递送到受试者时,GOI可以表达目的试剂,其将显示出用于治疗受试者的靶标疾病或病症的治疗活性。治疗剂可以是治疗性蛋白质,例如抗体、生长因子、细胞因子、凝血因子、酶或血红蛋白。

[0110] 抗体(也称为免疫球蛋白)是大蛋白,其响应于抗原产生,并且主要起到中和病原体的作用。具有治疗目的的抗体的非限制性实例可以在美国专利申请第2011/0293705 A1号中找到。

[0111] 生长因子是调节细胞分裂、细胞存活和炎症以及其它细胞功能的蛋白质。生长因子的非限制性实例包含肾上腺髓质素(AM)、血管生成素(Ang)、自分泌运动因子、骨形态发生蛋白(BMP)、睫状神经营养因子(CNTF)、白血病抑制因子(LIF)、白介素6(IL-6)、巨噬细胞集落刺激因子(m-CSF)、粒细胞集落刺激因子(G-CSF)、粒细胞巨噬细胞集落刺激因子(GM-CSF)、表皮生长因子(EGF)、肝配蛋白A1、肝配蛋白A2、肝配蛋白A3、肝配蛋白A4、肝配蛋白A5、肝配蛋白B1、肝配蛋白B2、肝配蛋白B3、促红细胞生成素(EPO)、成纤维细胞生长因子1(FGF1)、成纤维细胞生长因子2(FGF2)、成纤维细胞生长因子3(FGF3)、成纤维细胞生长因子4(FGF4)、成纤维细胞生长因子5(FGF5)、成纤维细胞生长因子6(FGF6)、成纤维细胞生长因子7(FGF7)、成纤维细胞生长因子8(FGF8)、成纤维细胞生长因子9(FGF9)、成纤维细胞生长因子10(FGF10)、成纤维细胞生长因子11(FGF11)、成纤维细胞生长因子12(FGF12)、成纤维细胞生长因子13(FGF13)、成纤维细胞生长因子14(FGF14)、成纤维细胞生长因子15(FGF15)、成纤维细胞生长因子16(FGF16)、成纤维细胞生长因子17(FGF17)、成纤维细胞生长因子18(FGF18)、成纤维细胞生长因子19(FGF19)、成纤维细胞生长因子20(FGF20)、成纤维细胞生长因子21(FGF21)、成纤维细胞生长因子22(FGF22)、成纤维细胞生长因子23(FGF23)、胎牛生长激素(FBS)、胶质细胞源性神经营养因子(GDNF)、神经营养素、Persephin、Artemin、生长分化因子9(GDF9)、肝细胞生长因子(HGF)、肝癌衍生生长因子(HDGF)、胰岛素、胰岛素样生长因子1(IGF-1)、胰岛素样生长因子2(IGF-2)、白介素1(IL-1)、IL-2、IL-3、IL-4、IL-5、IL-6、IL-7、角质形成细胞生长因子(KGF)、迁移刺激因子(MSF)、巨噬细胞刺激蛋白(MSP)、肌生长抑制素(GDF-8)、神经调节蛋白1(NRG1)、神经调节蛋白2(NRG2)、神经调节蛋白3(NRG3)、神经调节蛋白4(NRG4)、脑源性神经营养因子(BDNF)、神经生长因子(NGF)、神经营养蛋白3(NT-3)、神经营养蛋白4(NT-4)、胎盘生长因子(PGF)、血小板源性生长因子(PDGF)、肾素酶(RNLS)、T细胞生长因子(TCGF)、血小板生成素(TPO)、转化生长因子 $\alpha$ (TGF- $\alpha$ )、转化生长因子 $\beta$ (TGF- $\beta$ )、肿瘤坏死因子 $\alpha$ (TNF- $\alpha$ )和血管内皮生长因子(VEGF)。

[0112] 细胞因子是已知增强细胞免疫反应的小分子。细胞因子的非限制性实例包含TNF

$\alpha$ 、IFN- $\gamma$ 、IFN- $\alpha$ 、TGF- $\beta$ 、IL-1、IL-2、IL-4、IL-10、IL-13、IL-17、IL-18和趋化因子。趋化因子的非限制性实例包含CCL14、CCL19、CCL20、CCL21、CCL25、CCL27、CXCL12、CXCL13、CXCL-8、CCL2、CCL3、CCL4、CCL5、CCL11和CXCL10。

[0113] 凝血因子调节凝血。凝血因子的非限制性实例包含纤维蛋白原、凝血酶原、组织因子或组织凝血活酶、IV(钙)、促凝血球蛋白原(不稳定因子)、VI、稳定因子(前转变素)、抗血友病因子A、抗血友病因子B或圣诞节因子、Stuart-Prower因子、血浆凝血活酶前体、接触因子、纤维蛋白稳定因子、血管假性血友病因子、前激肽释放酶(Fletcher因子)、高分子量激肽原(HMWK)(Fitzgerald因子)、纤连蛋白、抗凝血酶III、肝素辅因子II、蛋白C、蛋白S、蛋白Z、蛋白Z相关蛋白酶抑制剂(ZPI)、纤溶酶原、 $\alpha$ 2抗纤溶酶、组织纤溶酶原激活物(tPA)、尿激酶、纤溶酶原激活物抑制剂1(PAI1)、纤溶酶原激活物抑制剂2(PAI2)和癌症促凝剂。

[0114] 酶是充当生物催化剂的多肽。在美国专利第7,807,618号中提供了治疗性酶的非限制性实例。

[0115] 血红蛋白是在红细胞中发现的一种蛋白质,其充当氧转运金属蛋白。

[0116] 在美国申请第2011/0294114A1号中提供了本公开的上下文中的目的基因。在一些实施例中,编码目的试剂的基因是 $\beta$ 珠蛋白或 $\gamma$ 球蛋白,其可以用于治疗贫血,例如,镰状细胞性贫血或 $\beta$ 地中海贫血。

[0117] 在一些实施例中,目的试剂是荧光蛋白(例如,mCherry、eGFP等)。荧光蛋白的非限制性实例包含wt-GFP、绿色荧光蛋白(例如EGFP、Emerald、Superfolder GFP、Azami Green、mWasabi、TagGFP、TurboGFP、AcGFP、ZsGreen、T-Sapphire等)、蓝色荧光蛋白(例如EBFP、EBFP2、Azurite、mTagBFP等)、青色荧光蛋白(例如ECFP、mECFP、Cerulean、mTurquoise、CyPet、AmCyan1、Midori-Ishi Cyan、TagCFP、mTFP1(Teal)等)、黄色荧光蛋白(例如EYFP、Topaz、Venus、mCitrine、YPet、TagYFP、PhiYFP、ZsYellow1、mBanana等)、橙色荧光蛋白(例如Kusabira Orange、Kusabira Orange2、mOrange、mOrange2、dTomato、dTomato-Tandem、TagRFP、TagRFP-T、DsRed、DsRed2、DsRed-Express(T1)、DsRed-Monomer、mTangerine等)或红色荧光蛋白(例如mRuby、mApple、mStrawberry、AsRed2、mRFP1、JRed、mCherry、HcRed1、mRaspberry、dKeima-Tandem、HcRed-Tandem、mPlum、AQ143等)。

[0118] 在其它实施例中,目的试剂可以是核酸。实例包含干扰RNA(例如,siRNA或shRNA)、反义寡核苷酸或微小RNA。

[0119] IV. 绝缘子片段在基因编辑中的应用

[0120] 本文公开的绝缘子片段中的任一个也可以用于基因编辑中的绝缘目的,例如本领域已知的那些。例如,绝缘子片段可以在用于基因编辑的基因靶向性构建体中,侧接待整合到宿主细胞中合适的染色体位点中的基因(在一侧或在两侧)。在另一个实例中,用于基因编辑中的基因靶向性构建体可以含有两个或更多个待插入到宿主细胞的基因组中的基因,并且一个或多个绝缘子片段可以插入成侧接两个或更多个基因。此类基因靶向性构建体可以用于合适的基因编辑方法中,通过所述方法,基因靶向构建体或其一部分可以例如在特定位点处被整合到宿主细胞的基因组中。在一些实施例中,基因靶构建体可以在5'和3'末端两者处含有与基因靶构建体旨在整合的靶位点同源的臂。绝缘子片段可以起到使由基因靶构建体携带的两个或更多个基因的表达绝缘的作用(例如,一个基因的表达对另一个基因的表达影响很小或没有影响)。可替代地或另外,绝缘子片段还可以起到在靶基因构建体

或其一部分被整合的染色体位点处隔离由基因靶构建体携带的一个基因和附近基因的作用(例如,靶基因的表达对附近天然基因的表达影响较低或没有影响,或反之亦然)。

[0121] 为了进行本文所述的基因编辑方法,可以将如本文公开的基因靶构建体与内切核酸酶一起递送到合适的宿主细胞。内切核酸酶可以是限制性内切核酸酶或DNA内切核酸酶,从而允许基因靶向性构建体的位点特异性插入。DNA内切核酸酶和通过基因编辑靶向基因插入的技术是本领域众所周知的。DNA内切核酸酶的非限制性实例包含锌指核酸酶、转录激活因子样效应子核酸酶(TALEN®)、基于CRISPR/Cas的RNA引导的DNA内切核酸酶和归巢内切核酸酶(亦称,大范围核酸酶)。在Gaj等人,《生物技术趋势(Trends Biotechnol.)》31(7):397-405,2013和Certo和Morgan,《分子治疗》24(3):422-29,2016中提供了这些内切核酸酶中的一些的实例,所述文献中的每一个的相关公开内容出于本文引用的目的和主题通过引用并入本文。

[0122] 在一个实例中,本文所述的基因靶构建体用于CRISPR方法中,其可以涉及编码gRNA和CRISPR核酸酶的质粒,如cas9,以及基因靶构建体作为DNA模板。可以基于期望的靶位点设计gRNA,在所述靶位点将整合靶基因构建体或其一部分。基因靶构建体可以由同源臂侧接,其将通过靶细胞中靶位点处的同源指导修复(HDR)促进靶基因构建体的整合。本文所述的这种基于CRISPR/Cas的方法是一种灵活、极其有效的一步法,其中Cas9/gRNA质粒和模板靶基因构建体可以一起转染,并且在转染后的短时间内(例如,2-3周)衍生出克隆,并且易于在多种细胞类型中实施。

[0123] 在一些实例中,基因靶构建体可以包括前病毒序列,并且本文描述的基于CRISPR/Cas(例如,CRISPR/Cas9)的方法可以用于检测目的原癌基因附近的病毒载体序列的存在。从一种原癌基因转换为另一种原癌基因仅需要克隆供体质粒,其前病毒序列由相关的同源臂侧接,并且改变gRNA序列。无需进一步阐述,据信本领域的技术人员可以基于以上描述,最大程度地利用本发明。因此,以下具体实施例被解释为仅是说明性的,而不以任何方式限制本公开的其余部分。本文引用的所有出版物出于本文引用的目的或主题通过引用并入。

#### [0124] 实施例

[0125]  $\gamma$  逆转录病毒载体中强大的病毒增强子已经引起细胞原癌基因活化和白血病,因此需要在“无增强子”的自灭活整合载体中使用细胞启动子。然而,细胞启动子导致相对较低的转基因表达,通常导致疾病校正不足。衍生自非致病性逆转录病毒泡沫病毒的载体示出比 $\gamma$  逆转录病毒/慢病毒更优先选择非基因整合,并且优先转录起始位点,如 $\gamma$  逆转录病毒附近的整合。在此,结果发现泡沫病毒载体的基因毒性显著降低;远远低于其对整合位点偏好的预期。进一步,结果发现,与携带相同增强子/启动子的类似 $\gamma$  逆转录病毒/慢病毒载体相比,置于泡沫病毒载体中的强病毒增强子/启动子对原代小鼠造血干/祖细胞的永生生化极低;与 $\gamma$  逆转录病毒/慢病毒载体相比,泡沫病毒对非基因区域的适度插入性位点偏好不能完全解释这种效果。使用CRISPR/Cas9介导的将类似的前病毒序列靶向插入到LM02基因中,并且然后测量LM02表达,结果证明泡沫病毒具有序列特异性作用,与插入性偏倚无关,有助于降低基因毒性。这些结果表明,这种作用是由位于泡沫病毒长末端重复序列(LTR)区中的36bp绝缘子介导的,所述绝缘子对CCCTC结合因子(CTCF)具有高亲和力结合。使用LM02活化测定,当从泡沫病毒中去除此绝缘子时,LM02表达显著增加,而当将此绝缘子插入到慢病毒LTR中时,LM02表达则显著下降。这些结果阐明了泡沫病毒的低基因毒性的潜在机制,

识别了新颖独特的绝缘子,并且支持将泡沫病毒用作基因治疗的载体,特别是在需要增强子/启动子时。

[0126] 材料和方法

[0127] 载体设计和生产

[0128] 先前已经描述了本公开中使用的载体SFFV-GV (RSF91.eGFP.pre) 和SFFV-LV (RRL.ppt.SF.eGFP.pre) (Modlich等人,《血液》108:2545-53,2006;Schambach等人,《分子治疗》13:391-400,2006)。已经描述了MSCV-GV eGFP.pre (Bauer等人,《血液》108:3313-20,2006)。MSCV-LV (RRL.ppt MSCV.eGFP.pre) 示出在图1A中。MSCV启动子的388-bp区(与 $\Delta \Phi$  MSCV CD18中的MSCV LTR启动子/增强子在序列上相同 (Bauer等人,《自然医学》14:93-7,2008) 是通过与人工XhoI和AgeI末端进行PCR扩增获得的,并且然后克隆到XhoI/AgeI限制酶消化的pRRLSIN.cPPT.PGK.eGFP.WPRE (Addgene, 剑桥, 马萨诸塞州) 中,以创建pRRL.ppt.MSCV.eGFP.pre (M.J.Hunter、D.D.Hickstein, 未出版的结果)。所有FV都在 $\Delta \Phi$  主链中。已经描述了FV  $\Delta \Phi$  MSCV.eGFP (Trobridge等人,《分子治疗》6:321-8,2002,Kiem等人,《血液》109:65-70,2007)。SFFV启动子替换了 $\Delta \Phi$  MSCV.eGFP FV载体中的MSCV启动子,以创建 $\Delta \Phi$  SF.eGFP。所有载体编码eGFP cDNA。通过除去MSCV增强子/启动子并且重新连接,无启动子的FV衍生自 $\Delta \Phi$  MSCV.eGFP载体。

[0129] 通过瞬时转染,在293T细胞中产生亲热GV上清液,如(Arumugam等人,2009)中所述,并且在NIH 3T3成纤维细胞上确定滴度(美国典型培养物保藏中心(ATCC))。病毒滴度处于106-107个感染单位/mL的范围内。通过瞬时共转染293T细胞(ATCC),产生LV载体SFFV-LV和MSCV-LV,如(Puthenveetil等人,2004)中所述。通过四个质粒[pCiES (Env)、pCiGS  $\Delta \Psi$  (Gag)、pCiPs (Pol) 和载体(p $\Delta \Phi$ )]瞬时转染,产生SFFV-FV、MSCV-FV和无启动子的(无Pr) FV,如(Trobridge等人,《分子治疗》6:321-8,2002)中所述。pCiGS  $\Delta \Psi$  是Gag表达盒(Russell DW;未出版),在包装信号中有更完整的缺失。p $\Delta \Phi$  是缺失的FV主链,多连接器插入转基因盒。

[0130] 将FV重新悬浮于含有2%热灭活的胎牛血清(FBS;Hyclone,Logan,UT)和5%DMSO (Sigma,St.Louis,MO)的Stemspan (Stem Cell Technologies,Vancouver,BC,Canada)中,并且在5%DMSO中冷冻保存直到使用。通过使用识别wPRE的引物,通过定量实时PCR(qRT-PCR)测量转导的HT1080细胞的基因组拷贝数来确定无Pr FV的滴度,而通过荧光激活细胞分选(FACS)定量GFP表达来确定其它载体的滴度。FV的滴度处于 $3 \times 10^7$  IU/ml到 $1 \times 10^8$  IU/ml的范围内。

[0131] 谱系阴性(Lin<sup>-</sup>)细胞的分离

[0132] 使用生物素化的谱系特异性抗体(谱系细胞衰竭试剂盒;BD生物科学,加利福尼亚州,圣何塞)使用(Arumugam等人,《分子治疗》17:1929-37,2009,Modlich等人,《分子治疗》17:1919-28,2009)中所述的方法,将来自C57BL/6J小鼠的骨髓用于分离Lin细胞。将生物素标记的Lin细胞与抗生物素微珠(Miltenyi Biotech)一起温育,随后对未标记的Lin细胞进行磁性分选。在含有1%青霉素/链霉素、50ng/ml mSCF、100ng/ml hIL-11和10ng/ml mIL-3的Stemspan培养基(Stem Cell Technologies)中,预刺激分离的Lin细胞进行病毒转导。

[0133] 体外永生化测定

[0134] 在含有1%青霉素/链霉素、50ng/ml mSCF、100ng/ml hIL-11和10ng/ml mIL-3的

Stemspan培养基中,将Lin细胞预刺激过夜。在第2天,100000个Lin细胞用于每个LV载体转导。使用浓缩的LV载体上清液,以感染复数(MOI)为20,每8小时间隔两次转导Lin细胞。对于GV转导,在Stemspan细胞因子混合物中,将谱系阴性细胞预刺激2天。在以MOI为20,包被了预装有逆转录病毒载体SFFV-GV和MSCV-GV的24孔皿的RetroNectin重组人纤连蛋白片段(Takara Bio Inc.,Kusatsu,Shiga,Japan)上,在第3天和第4天进行GV转导。在最后的转导之后,洗涤转导的Lin细胞,并且在Stemspan细胞因子混合物中作为批量培养物扩增19天。

[0135] 通过添加温热的预刺激培养基快速融化FV储备液,并且在48孔板中,用FV转导Lin细胞一次,在MOI 50上用8 $\mu$ g/cm<sup>2</sup>浓度的RetroNectin重组人纤连蛋白片段预涂所述Lin细胞。与用慢病毒载体进行的两次转导相比,泡沫病毒的转导仅进行了一次,因此选择了更高的MOI进行转导。16小时之后,洗涤细胞,并且在Stemspan细胞因子混合物中扩增。即使在转导期间最终浓度 $\leq 1\%$ 的DMSO中,也观察到了泡沫病毒载体转导后Lin细胞的毒性(50-60%的活力)。在最终转导之后的第4天,使用流式细胞术分析了转导的批量培养物中的转基因表达。在具有较低基因转移效率的批量培养物中,使用BD FACS Aria II(BD Biosciences)对GFP+细胞进行分选,并且扩增直到铺板。还对具有较高基因转移的批量培养物进行了GFP+细胞分选,并且比较了分选库与未分选转导库之间的重铺频率。在扩增期间,转导的批量培养物保持在 $2-5 \times 10^5$ 细胞/mL的浓度。扩增之后,将细胞以100个细胞/孔的密度接种在96孔板中。铺板2周之后,检查96孔板,并且对具有增殖细胞群的孔(阳性/永生化孔)的存在进行评分。在这些条件下,模拟细胞几乎无法生存。模拟转导是不添加病毒的转导,并且是每个实验的阴性对照。也包含无启动子的FV作为阴性对照。阳性孔进一步扩增用于分子分析。在第5周,在第2周扩增的克隆中的一些最终已经分化并且在第5周死亡。使用L-Calc软件(Stem Cell Technologies),基于泊松统计来计算在2周和5周测试的每个载体的重铺频率。将重铺频率归一化为重铺之前Lin批量培养种群的平均载体拷贝数(VCN)。

[0136] 衍生自SFFV-GV的永生化克隆可以在2周和5周以相同的频率重铺,从而可以比较载体与高度基因毒性的SFFV-GV载体的相对基因毒性。相反,衍生自具有低基因毒性潜力的载体的克隆在2周时显示出最初的生长和重铺潜力,但是此后最终分化,并且在5周时失去其重铺频率。为了能够比较分选和未分选种群的永生化频率/VCN,对从SFFV-GV、MSCV-GV、SFFV-LV、MSCV-LV、SFFV-FV和MSCV-FV转导的Lin细胞中转导的批量培养物的一部分进行分选,以用于GFP表达,并且在分选之前和之后显示成比例的永生化,验证此修饰以给出类似的永生化读数。在对SFFV-GV、SFFV-LV和SFFV-FV转导的细胞进行分选之前,永生化频率/VCN分别为0.001755、0.000176和0.000006。分选之后,频率分别为0.002016、0.000128和0.000023。对于载体转导的重铺克隆阴性的组,计算基于假设:如果接种了97孔而不是96孔,则检测到重铺克隆(Zychlinski等人,《分子治疗》16:718-25,2008)。

#### [0137] 永生化克隆的表型分析

[0138] 用识别来自BD Biosciences的细胞表面标记Sca-1 PE(克隆D7,目录号553108)和c-Kit APC(克隆2B8,目录号553356)的抗体标记永生化克隆,并且使用荧光激活细胞分选仪(FACS)Canto(BD Biosciences)进行分析。

#### [0139] 载体拷贝数分析

[0140] 进行定量实时PCR,以评估GV、LV和FV载体转导的批量培养物的基因转移效率。对于GV载体、LV载体和FV载体,使用识别wPRE区的引物来测量批量培养物中的VCN。来自携带

MM13载体的单拷贝的单拷贝NIH 3T3细胞克隆的基因组DNA用作拷贝数分析的标准。先前已经描述了MM13质粒(Will等人,《细胞周期(Cell Cycle)》5:14-22,2006)。FV主链中的引物也用于测量拷贝数。FV主链正向引物:5'-AATCCTTTACATGGAGAAGTTATAGGTCTT-3'(SEQ ID NO:13),反向引物:5'-TGGCCAAATCCATAGCCTTAGA-3'(SEQ ID NO:146)。使用Taqman探针:5'-ATCTGAAATCTCTCAATTTGTCCCCACCA-3'(SEQ ID NO:14),以四甲基-6-羧基若丹明染料为猝灭剂,进行PCR反应。将FV或wPRE特异性信号归一化为每个样品中的小鼠ApoB。用未转导的MEL DNA稀释来自FV转导的单拷贝鼠红细胞白血病(MEL)克隆的基因组DNA(50ng),以产生拷贝数标准。使用Applied Biosystems 7900HT实时PCR系统(Thermo Fisher,Grand Island,NY)进行定量PCR,根据制造商的说明,使用用于96孔板的热循环仪协议。

[0141] 连接扩增介导的PCR,以确定永生化克隆中的插入位点

[0142] 对于LAM-PCR,使用0.25pmol FV特异性5'生物素化引物[5'-GAACCTTGTGTCTCTCATCCC-3'](SEQ ID NO:15)和2.5个单位的Qiagen Taq聚合酶(Qiagen,希尔登,德国),初始变性的循环条件是在95°C下持续3分钟,进行50个扩增循环(95°C下持续30秒,55°C下持续30秒,72°C下持续1分钟),并且最后在72°C下延伸3分钟,用来自FV永生化克隆的100ng基因组DNA,将病毒LTR与小鼠基因组之间的连接序列线性扩增两次。在生物素化的DNA的DNA富集之后,使用Klenow(Promega,Madison,WI)进行六核苷酸引物延伸,并且用TasI[新英格兰生物实验室(New England Biolabs)(NEB),伊普斯威奇,马萨诸塞州]消化引物延伸产物。TasI消化后,将DNA连接到TasI特异性双链连接子5'-GACCCGGGAGATCTGAATTCAGTGGCAC AGCAGTTAGG-3'(SEQ ID NO:16)/5'-AATTCCTAACTGCTGTGCCAC TGAATTCAGATC-3'(SEQ ID NO:17)。使用12.5pmol的FV特异性引物[5'-GTCTATGAGGAGCAGGAGTA-3'](SEQ ID NO:18)和连接子盒特异性引物[5'-GACCCGGGAGATCTGAATTC-3'](SEQ ID NO:19)中的每一种进行连接产物的第一次指数扩增。然后使用12.5pmol的嵌套FV特异性引物[5'-CCTCCTCCCTGTAATACTC-3'](SEQ ID NO:20)和嵌套连接子盒特异性引物[5'-AGTGGCACAGCAGTTAGG-3'](SEQ ID NO:21)中的每一种,使用与第一次PCR相同的条件,将百分之八的第一指数PCR反应用作第二指数嵌套PCR扩增的模板。为了检测来自MSCV GV永生化克隆的插入位点,使用富集有Dynabeads M-280链霉抗生物素蛋白(Thermo Fisher),用Tsp5091(NEB)消化并连接到不对称连接子盒(5'-AATTCTCTAGTATGCTACTCGCACCG ATTATCTCCGCTGTGTCAGT-3'(SEQ ID NO:23)和5'-ACTGACAGCGGAGATAA TCGGTGCGAGTAGCATACTAGAG-3'(SEQ ID NO:24))的MSCV LTR特异性5'-生物素化引物(LTR1):5'-CTGGGGACCATCTGTTCTTGGCCCT-3'(SEQ ID NO:22),线性扩增100ng基因组DNA。然后用LTR和连接子特异性引物LTR2(5'-GACTTGTGGTCTCGCTGTTT CTTGG-3')(SEQ ID NO:25)和连接子盒引物LC1(5'-ACTGACAGCGGAG ATAATCG-3')(SEQ ID NO:26)(第1次指数PCR)扩增连接产物。用引物LTR3:(5'-GGTCTCCTCTGAGTGATTGACTACC-3'(SEQ ID NO:27)和LC2:(5'-GTGCGAGTAGCATACTAGAG-3')(SEQ ID NO:28)进行第二次指数PCR(Shou等人,《美国国家科学院院刊》103:11730-11735,2006)。

[0143] 连接扩增介导的(LAM)PCR产物的下一代测序

[0144] 来自第二次指数PCR的产物在辛辛那提儿童医院医学中心(CCHMC)DNA测序核心中进行了下一代DNA测序。使用QIAquick PCR纯化试剂盒(Qiagen)纯化LAM第2次指数PCR产物,然后在10mM dNTP(Thermo Fisher)的存在下,通过具有T4 DNA聚合酶、Klenow和T4 PNK

(Promega)的末端修复结束钝化。通过在室温下用T4快速连接酶(NEB)处理15分钟,将末端钝化产物随机合并。使用Nextera体外转座试剂盒(Epicentre, Madison, WI),按照制造商的建议,制备了与Illumina系统兼容的下一代测序文库,并且使用每个样品的不同分子条形码进行扩增。在另一轮PCR纯化之后,所有10个文库均在安捷伦生物分析仪(Agilent,加利福尼亚州,圣克拉拉)上进行了质量检查,然后在单库中进行等量混合。测序是在Illumina HiSeq2000(Illumina,加利福尼亚州,圣地亚哥)的具有索引的单读模式上进行的,产生100个碱基长的序列。

[0145] 在对库中的所有序列进行多路分配,并且分配了它们相应的样品之后,使用CASAVA 1.8程序包,处理读段,并且将其与mm9小鼠参考组合件进行比对。结果以QSEQ SORTED文件格式生成,因此可以使用Illumina的Genome Studio软件的ChIP-SEQ模块可视化比对。通过LAM PCR检测到的插入位点的特征在于插入上游的LTR序列和下游的衔接子序列。虽然比对仪被配置成定位仅含有小鼠基因组序列的读段,但是也定位了含有一些LTR或一些衔接子序列以及大多数小鼠序列的读段。通过放大Genome Studio(Illumina)中的基本级别显示,可以确定覆盖区域的边缘和与衔接子序列匹配的侧面以及与LTR序列匹配的侧面,从而确定插入点和前病毒整合的方向。将所有识别出的插入与国家生物技术信息中心(NCBI)小鼠第37组基因组数据库([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov))进行比较。

[0146] 前病毒序列的CRISPR/Cas9插入

[0147] gRNA发展

[0148] 用于LM02整合位点(p12-14.1的智人染色体11克隆RP1-22J9图谱,基因库#AL135799.8)的初始说明的参考序列(Natkunam等人,《血液》109:1636-1642,2007)是从NCBI获得的。这对应于GRCh38.p2 chr11:33890271。从Jurkat细胞分离基因组DNA,并且使用Q5聚合酶(NEB)对插入位点周围的区域进行PCR扩增,并且通过CCHMC DNA测序和基因分型核心进行测序。PCR引物是LM02 FWD PCR

[0149] (5'-TTTAGGTTGCCCTGAAAAGGTG-3')(SEQ ID NO:29)和LM02 REV PCR(5'-GCCAAACTCCTAGGCTCTTG-3')(SEQ ID NO:30)。测序引物是LM02 FWD PCR、LM02 REV PCR和LM02 seq1(5'-GTCTCTCGCAGCCACATGGG-3')(SEQ ID NO:31)。使用Benchling CRISPR设计程序(<https://benchling.com>)对插入位点周围的区域进行了潜在的gRNA靶位点分析。根据与计划插入位点的接近度和较低的脱靶效应,选择了五个gRNA。含有gRNA和Cas9-T2A-eGFP表达盒(pX458m)两者的质粒(转基因动物和基因组编辑核心,CCHMC)eGFP cDNA首先被mCherry报告基因(pX458m-mCherry)替换。使用QuikChange II XL定点诱变试剂盒(Agilent)进行定点诱变,以去除mCherry序列(引物5'-CCCCTAATGCAGAAGAAAACCATGGGCTGGAGGC-3'(SEQ ID NO:32)和5'-GCCTCCCAGCCCATGGTTTTCTTCTGCATTACGGG-3')(SEQ ID NO:33)内的BbsI位点。设计了用于将靶序列克隆到pX458m-mCherry载体中的DNA寡核苷酸,并且从Integrated DNA Technologies(Coralville,IA)获得。用于产生具有靶向性序列GATACCAATAGATATCAATC(SEQ ID NO:34)的gRNA 1的寡核苷酸是LM02 gRNA 1FWD(5'-CACCGGGATACCAATAGATATCAATC-3')(SEQ ID NO:35)和LM02 gRNA 1REV(5'-AAACGATTGATATCTATTGGTATCCC-3')(SEQ ID NO:36)。用于产生具有靶向性序列ATCACCAGATTGATATCTAT(SEQ ID NO:37)的gRNA 2的寡核苷酸是LM02 gRNA 2FWD(5'-CACCGGGATCACCAGATTGATATCTAT-3')(SEQ ID NO:38)和LM02 gRNA 2REV(5'-

AAACATAGATATCAATCTGGTGATCCC-3') (SEQ ID NO:39)。用于产生具有靶向性序列 AATTGCATAGTCGTGAAGTC (SEQ ID NO:40) 的 gRNA 3 的寡核苷酸是 LM02 gRNA 3FWD (5'-CACCGGGAATTGCATAGTCGTGAAGTC-3') (SEQ ID NO:41) 和 LM02 gRNA 3REV (5'-AAACGACTTCACGACTATGCAATCCC-3') (SEQ ID NO:42)。用于产生具有靶向性序列 ATTGCATAGTCGT GAAGTCA (SEQ ID NO:43) 的 gRNA 4 的寡核苷酸是 LM02 gRNA 4FWD (5'-CACCGGGATTGCATAGTCGTGAAGTCA-3') (SEQ ID NO:44) 和 LM02 gRNA 4REV (5'-AAACTGACTTCACGACTATGCAATCCC-3') (SEQ ID NO:45)。用于产生具有靶向性序列 TCGTGAAGTCAGGGCTTCTA (SEQ ID NO:46) 的 gRNA 5 的寡核苷酸是 LM02 gRNA 5FWD (5'-CACCGGGTCGTGAAGTCAGGGCTTCTA-3') (SEQ ID NO:47) 和 LM02 gRNA 5REV (5'-AAACTAGAAGCCCTGACTTCACGACCC-3') (SEQ ID NO:48)。

[0150] 用 FastDigest BbsI (Thermo Fisher) 消化 pX458m-mCherry, 并且同时用 FastAP (Thermo Fisher) 脱磷酸。然后将消化的产物进行凝胶纯化。寡核苷酸对在 100 $\mu$ M 的每种寡核苷酸和 T4 多核苷酸激酶 (NEB) 的反应中被磷酸化和退火, 并且放置在 Veriti 96 孔快速热循环仪 (Thermo Fisher) 中 37 $^{\circ}$ C 持续 30 分钟, 95 $^{\circ}$ C 持续 5 分钟, 然后以 5 $^{\circ}$ C/min 降到 25 $^{\circ}$ C。然后将退火的寡核苷酸连接到切割的 pX458m 质粒中, 并且转化到 PX5- $\alpha$  感受态细胞 (蛋白表达, 俄亥俄州辛辛那提) 中。随后使用无内源性质粒 Maxi 试剂盒 (Qiagen) 制备质粒。按照制造商的规程, 使用 Lipofectamine 3000 (Thermo Fisher), 将 2 $\mu$ g 的每种 gRNA/Cas9 质粒转染到 24 孔板中的  $2.4 \times 10^5$  个白血病细胞中。在第 7 天, 收获细胞, 并且纯化基因组 DNA。使用 Q5 聚合酶扩增靶位点周围的区域并测序 (PCR 引物: LM02 FWD v3 5'-GCTTGGGTTTTACAGTCTTC-3' (SEQ ID NO:49) 和 LM02 REV v3 5'-TCAGCTAGAAAACAAGTACTTGC-3' (SEQ ID NO:50), 测序引物 LM02 seq1 5'-GTCTCTCGCAGCCACATGGG-3') (SEQ ID NO:147)。使用通过分解来追踪插入缺失 (TIDE) 测定来确定 gRNA 效率 (Shou 等人, 《美国国家科学院院刊》103: 11730-11735, 2006)。

[0151] 用于同源性定向修复 (HDR) 的供体载体模板

[0152] 在白血病细胞中对 LM02 区进行测序之后, 设计了约 600-bp 同源臂, 在所选插入位点处具有多个克隆位点区域。从 GenScript 订购了同源载体作为 pUC57 主链中的质粒。用 XbaI 和 EcoNI 在 LTR 处切割泡沫病毒 p $\Delta\Phi$ .SF.eGFP.PRE, 并且插入 NheI 与 EcoNI 之间。用 BsaI 和 PsiI 在 LTR 处切割慢病毒 pRRL.PPT.SF.eGFP.PRE, 并且插入 BbsI 与 NaeI 之间。在 XhoI 与 HindIII 之间的 LTR 处切割逆转录病毒 pRSF91.eGFP.PRE, 并且插入 BsmFI 与 XhoI 之间。通过限制性消化检查所得克隆, 并且测序以验证。对于 GV, 克隆序列中含有含有 SFFV 启动子/增强子的整个 LTR 序列。为了促进 LV 和 FV 的克隆, 从病毒产生质粒中克隆了从 5'LTR 的 R 区到整个 3'LTR 的序列。对于 LV 载体, 5' $\Delta$ U3 区的 40-bp (LTR 中缺失 U3 启动子/增强子后剩下) 不是克隆的“前病毒”构建体的一部分。这个小区不具有增强子/启动子活性, 并且因此对于研究基因毒性而言是无关紧要的。类似地, 在 FV 载体的情况下, 省略的 5'U3 区含有 582-bp 缺失, 从而去除了 LTR 的 U3 TATA 框和转录增强子位点, 导致 LTR 沉默 (Trobridge 等人, 《酶学方法 (Methods Enzymol.)》346:628-48, 2002)。如本文所用, 术语“前病毒”是指这些 LV 和 FV 构建体。

[0153] 在识别 FV LTR 中的 36-bp 绝缘子区后, 在 5'LTR 和 3'LTR 处修饰含有 p $\Delta\Phi$ .SF.eGFP.PRE 的 LM02 供体, 以去除已识别的 CTCF 结合位点。此区的初始序列是 AGT AAA AGG

ATT TGT ATA TTA GCC TTG CTA AGG GAG ACA TCT AGT GAT ATA AGT GTG AAC TAC ACT TAT CTT AAA TGA TG (SEQ ID NO:51) 到 AGT AAA AGG ATT TGT ATA TTA GCC TTG CTA AGC ACA TTC GAT AGT GAT ATA AGA GGC TTT ATA TCT TAT CTT AAA TGA TG (SEQ ID NO:52) (带下划线的是绝缘子序列)。对于前病毒序列的3'LTR,从IDT订购了含有经修饰的绝缘子序列的基因块,以替换EcoNI与MluI之间的~800-bp区。对于5'LTR,从GenScript订购了基因合成产品,以替换PacI与AvrII之间的~550-bp区。通过测序确认所得质粒。

[0154] 为了向慢病毒LTR中添加36-bp绝缘子序列,在含有pRRL.PPT.SF.eGFP.PRE的LM02供体中,在LTR的R区之前添加以下序列:AAG GGA GAC ATC TAG TGA TAT AAG TGT GAA CTA CAC (SEQ ID NO:53)。从IDT订购了两个基因块,以替换BsiWI与MluI之间涵盖3'LTR的~1kb区和BspEI与MfeI之间涵盖5'LTR的~900-bp区。通过测序确认所得质粒。

#### [0155] HeLa克隆的产生

[0156] 在第-1天,将 $5 \times 10^4$ 个细胞接种到24孔板中。用LM02 gRNA 5质粒以及三个(GV、LV或FV) LM02供体质粒中的一个转染细胞。转染500ng总DNA,LM02 gRNA 5质粒与供体质粒的摩尔比约为1:2。根据制造商的推荐,使用1.5 $\mu$ L Lipofectamine 3000 (Thermo Fisher)进行转染。在第2天,通过使用FACS Canto (BD Biosciences)分析一部分细胞的eGFP (供体质粒)和mCherry (LM02 gRNA 5质粒)两者的表达来验证成功的转染。在第2周,通过FACS重新分析细胞的eGFP和mCherry。通过CCHMC的流式细胞术核心研究,使用BD FACS Aria II,将GFP阳性细胞和mCherry阴性细胞作为单细胞分选到96孔板中(图3A)。

[0157] 在达到至少80%汇合之后,收获一部分细胞。通过将细胞沉淀物重悬于20 $\mu$ L QuickExtract DNA提取液(Epicentre)中,并且在65 $^{\circ}$ C下温育15分钟,在68 $^{\circ}$ C下温育15分钟,并且在98 $^{\circ}$ C下温育10分钟来纯化DNA。然后使用侧接同源臂的引物对,通过PCR筛选经纯化的DNA,以正确整合供体序列。用于确保正确的5'同源性的第一PCR使用了引物LM02 FWD v3和病毒FWD (5'-CGAGCGTTGTAAGAGAAGC-3') (SEQ ID NO:54)。用于确保正确的3'同源性的第二PCR (PCR 2) 使用了不同的FV FWD引物。用于确保正确的3'同源性的第二PCR使用了引物LM02 REV v3和GV和LV的病毒REV1 (5'-GAGATCTGTCCCGCTAGCA-3') (SEQ ID NO:55) 或FV的病毒REV2 (5'-GGATAATTTACAAATAAACCCGAC TTATATTCG-3') (SEQ ID NO:56)。只有正确地整合的序列产生扩增子(图3B-3D)。对于两个PCR均具有正确条带的克隆被HDR认为具有正确地插入的病毒序列。已经报道HeLa细胞含有LM02基因所在的染色体11p的3-4个拷贝。估计靶向等位基因拷贝数的步骤如下:i) 通过荧光原位杂交(FISH)确认未编辑的/WT HeLa细胞具有4个LM02等位基因(图9A),ii) 使用侧接Cas9 DSB/前病毒插入位点的LM02基因中的引物,确定经编辑的非靶向或WT LM02等位基因的数目,因此仅检测到WT/经编辑的非靶向LM02,并且含有前病毒序列插入的LM02基因座不会扩增-WT/非靶向LM02拷贝数分析表明,所有可评估的克隆均具有1-2个WT/非靶向LM02拷贝(图9B);以及iii) 通过PCR扩增跨gRNA靶位点,进一步询问克隆,随后测序(PCR/测序引物与用于TIDE测定的那些相同)以评估每个克隆中大的缺失和插入缺失的存在(图9C)。使用识别染色体11 (Chr11) (33,736,494-33,907,488)的RP11-1006P23 FISH探针(帝国基因组学(Empire Genomics),水牛城,纽约)进行FISH。相对于未编辑的HeLa细胞,计算了非靶向/WT LM02等位基因的数量。克隆之间存在相对类似的非靶向/WT LM02拷贝数。每个克隆的N=3。

#### [0158] 通过qRT-PCR分析LM02表达

[0159] 通过在RNA Stat-60 (AMS Biotechnology, 英国, 阿宾登) 中裂解细胞, 并且通过QIAshredder柱 (Qiagen) 来制备RNA。通过氯仿相分离分离RNA。用异丙醇沉淀水层, 并且用75%乙醇洗涤所得沉淀物。将RNA沉淀物重新悬浮在无核酸酶的分子级水中, 并且通过在55℃下温育而溶解。使用NanoDrop 1000分光光度计 (Thermo Fisher) 对RNA进行定量。使用大容量cDNA反转录试剂盒 (Thermo Fisher) 用纯化的RNA制备cDNA。从经LM02修饰的克隆产生cDNA, 所述克隆含有FV、LV、无绝缘子的FV和具有LTR中放置的FV绝缘子的LV。使用qRT-PCR, 确定LM02 mRNA表达。使用Hs001534473\_m1引物/探针组和PPIA内源性对照来获取数据。N分别=5个、6个、17个和9个克隆 (图8)。还从GV、LV和FV克隆产生cDNA, 并且使用RT-PCR确定LM02 mRNA表达。通过具有两个探针和引物组的RT-PCR, 对应于LM02 cDNA中两个不同区域的LM02 TaqMan基因组测定Hs00277106和Hs00153473 (Thermo Fisher), 并且使用HeLa细胞、人类PPIA和GAPDH的两个不同的经验证的加载对照 (Thermo Fisher) 来确定HeLa克隆中LM02 mRNA的表达 (Stein等人, 《自然医学 (Nat Med)》16:198-204, 2010)。在所有报道的转录物变体中发现了扩增的LM02区。探针/引物组两者桥接外显子, 因此仅mRNA会被扩增。对于LM02转录变体1 (NM\_005574.3), 探针/引物组分别设置桥外显子4-5和5-6。用iTaq Universal Probes Supermix (Bio-Rad, Hercules, CA) 制备qRT-PCR反应。使用ABI 7900HT实时PCR系统运行qRT-PCR。使用 $2^{-\Delta\Delta CT}$ 方法通过相对定量来分析来自RT-PCR的数据 (Schmittgen等人, 《自然实验手册 (Nat. Protocols)》3:1101-1108, 2008) 使用人类GAPDH/PPIA归一化结果, 并且确定诱导倍数。未经编辑的HeLa样品用作校准物。

#### [0160] WT LM02拷贝数分析

[0161] 从HeLa克隆中分离基因组DNA。用iTaq Universal Probes Supermix (Bio-Rad) 制备反应。人类ApoB用作内源性对照基因。使用CFX Connect实时PCR检测系统 (Bio-Rad) 运行qRT-PCR。使用 $2^{-\Delta\Delta CT}$ 方法通过相对定量来分析来自RT-PCR的数据 (Schmittgen等人, 《自然实验手册》3:1101-1108, 2008) 使用K562细胞系和未经编辑的HeLa细胞归一化结果, 并且确定拷贝数。引物和探针组如下: LM02-CN-FW (5'-TGGGGAACAAGTACAATTTTGTG-3') (SEQ ID NO:57), LM02-CN-RV2 (5'-CAATGTGGTGATATCAATCTGGTG-3') (SEQ ID NO:58), LM02-CN-探针 (5'-ACAAGCGTAAATTGCATAGTCGTGA-3') (SEQ ID NO:59), hApoB-CN-FW (5'-CTTGGTTTATGAATCTGGCTC-3') (SEQ ID NO:60), hApoB-CN-RV (5'-GCCTTTAGCAGTTAGAACAC-3') (SEQ ID NO:61), hApoB-CN-探针 (5'-ACATGCTGGGAATCGACTTGTGAT-3') (SEQ ID NO:62)。

#### [0162] LM02蛋白表达的蛋白质印迹分析

[0163] 在SFFV-GV、SFFV-LV和SFFV-FV克隆上进行LM02表达的蛋白质印迹分析。所用的克隆由图5中每行上方的字母和数字标记表示。未转导的HeLa细胞 (-) 和K562细胞 (+) 分别用作阴性和阳性对照。内源GAPDH表达用作加载对照。首先, 将细胞在RIPA裂解缓冲液 (圣克鲁斯, 达拉斯, 德克萨斯州) 中裂解, 并且在4-15%的Mini-PROTEAN TGX预制蛋白凝胶 (Bio-Rad) 上分离20μL蛋白质裂解物 (每个样品中含有介于30微克与87微克之间的蛋白质), 并且转移到硝酸纤维素PVDF膜上 (Bio-Rad)。用Odyssey封闭缓冲液 (LI-COR Biosciences, Lincoln, NE) 封闭膜, 并且用一级人LM02抗体和二级抗山羊IRDye 800CW抗体 (LI-COR Biosciences) 探测。然后用一级人GAPDH抗体 (Fitzgerald, Acton, MA) 和二级抗小鼠IRDye 680LT (LI-COR Biosciences) 再次探测膜。使用Odyssey 9120红外成像仪 (LI-COR

Biosciences) 可视化信号。

[0164] 电泳迁移率变动测定

[0165] 设计寡核苷酸,以预测LV和FV前病毒序列中的CTCF结合位点。从集成DNA技术公司(Integrated DNA Technologies (IDT)) 订购了在5'末端标记有IRDye®700的寡核苷酸。LV1探针,通过使寡核苷酸5'-ACAAGATAGAGGAAGAGCAAAACAAAAGTAAGACCACCGCACAGCAAGCGGCCGCTGATCTTCAGACCTGGAGGAGATATGAGGGA-3'(SEQ ID NO:63)和5'-TCCCTCATATCTCCTCCTCCAGGTCTGAAGATCAGCGGCCGCTTGCTGTGCGGTGGTCTTACTTTTGTGTTTGCTCTCCTCTATCTTGT-3'(SEQ ID NO:64)退火制造,含有GGAAGAGCA和CTCCTCCTCCAGGT(SEQ ID NO:65)序列基序。LV2探针,通过使寡核苷酸5'-GATACCTAAAGGATCAACAGCTCCTGGGGATTGGGGTTGCTCTGGAAAAC TCATTTGCACCACTGCTGTGCCTTGAATGCT-3'(SEQ ID NO:66)和5'-AGCATTCCAAGGCACAGCAGTGTGCAAATGAGTTTTCCAGAGCAACCCCAAATCCCAGGAGCTGTTGATCCTTTAGGTATC-3'(SEQ ID NO:67)退火制造,含有TCCCAGGAGCTGTTGATCC(SEQ ID NO:68)和GGCACAGCA序列基序。LV3探针,通过使寡核苷酸5'-GTCGGGAAGCTGACGTCCTTTCGAATTCGATATCAAGCTGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATC-3'(SEQ ID NO:69)和5'-GATCTACAGCTGCCTTGTAAGTCATTGGTCTTAAAGGTACAGCTTGATATCGAATTCGAAAGGACGTCAGCTTCCCGAC-3'(SEQ ID NO:70)退火制造,含有GGTACAGCT序列基序。FV1探针,通过使寡核苷酸5'-TCCATTAACACTCTGCTTATAGATTGTAA GGGTGATTGCAATGCTTTCTGCATAAACTTTGGTTTTCTTGTTAATCAAT-3'(SEQ ID NO:71)和5'-ATT GATTAACAAGAAAACCAAAGTTTTATGCAGAAAGCATTGCAATCACCTTACAATCTATAAGCAGAGTGTTAATG GA-3'(SEQ ID NO:72)退火制造,含有AGCATTGCA序列基序。FV2探针,通过使寡核苷酸5'-A GTAAAAGGATTTGTATATTAGCCTTGCTAAGGGAGACATCTAGTGATATAAGTGTAAGTACTACTTATCTTAAATG ATG-3'(SEQ ID NO:73)和5'-CATCATTTAAGATAAGTGTAGTTCACACTTATATCACTAGATGTCTCCCTT AGCAAGGCTAATATACAAATCCTTTTACT-3'(SEQ ID NO:74)退火制造,含有 ATATCACTAGATGTCTCCCT(SEQ ID NO:75)和重叠序列基序。FV3探针,通过使寡核苷酸5'-TC GGGTTTATTTGTAAATTATCCCTAGGGACCTCCGAGCATAGCGGGAGGCATATAAAAGCCAATAGACAATGGCTAG CA-3'(SEQ ID NO:76)和5'-TGCTAGCCATTGTCTATTGGCTTTTATATGCCTCCCGCTATGCTCGGAGGTC CCTAGGGATAATTTACAAATAAACCCGA-3'(SEQ ID NO:77)退火制造,含有AGCATAGCG序列基序。FV4探针,通过使寡核苷酸5'-GGCATCAGCCTACAAATACCAGTATTCATACTGAAGGCAATGCCCTAGCAG ATAAGCTTGCCACCCAAGGAAGTTATGTA-3'(SEQ ID NO:78)和5'-TACATAACTTCCTTGGGTGGCAAGC TTATCTGCTAGGGCATTGCCTTCAGTATGAATACTGGTATTTGTAGGCTGATGCC-3'(SEQ ID NO:79)退火 制造,含有GGCATTGCC序列基序。FV5探针,通过使寡核苷酸5'-CGCAACTGTAAATCTCTCAATGT ACTCACTAGTATTGCAATTCAAAGGTGATTCACTCTGATCAAGGTGCAGCATTCA-3'(SEQ ID NO:80)和 5'-TGAATGCTGCACCTTGATCAGAGTGAATCACCTTTGGAATTGCAATACTAGTGAGTACATTGAGAGATTTAAC AGTTGCG-3'(SEQ ID NO:81)退火制造,含有GGAATTGCA序列基序。FV6探针,通过使寡核苷酸 5'-CTCGTTCCTGGTCTCCTGTTGTTGGCCAATTGGTCCAGGAGAGGGTGGCTAGGCCTGCTTCTTTGAGACCTCG TTGGCAT-3'(SEQ ID NO:82)和5'-ATGCCAACGAGGTCTCAAAGAAGCAGGCCTAGCCACCCTCTCCTGGA CCAATTGGCCAACAACAGGAGACCAGGAACGAG-3'(SEQ ID NO:83)退火制造,含有 TGGTCCAGGAGAGGGT GGCT(SEQ ID NO:84)和重叠序列基序。FV7探针,通过使寡核苷酸5'-A TGAGGCACTTCAGAATACAACAAGTGTGACTGAACAGCAGAAGGAACAAATTATACTGGACATTCAAATGAAGAA GTA-3'(SEQ ID NO:85)和5'-TACTTCTTCATTTTGAATGTCCAGTATAATTTGTTCCCTTCTGCTGTTTCAGT

CACAGTTGTTGTATTCTGAAGTGCCTCAT-3' (SEQ ID NO:86) 退火制造, 含有TGAACAGCAG AAGGAACAAA (SEQ ID NO:87) 和重叠序列基序。FV8探针, 通过使寡核苷酸5'-TATGGAAGCTTA TGGACCTCAGAGAGGAAGTAACGAGGAGAGGGTGTGGTGAATGCCACTAGAAACCAGGAAAACAAG-3' (SEQ ID NO:88) 和5'-CTTGTTCCTGGTTTCTAGTGGCATTCCACCACACCCTCTCCTCGTTACTTCCCTCTCTGA GGTCCATAAGCTTCCATA-3' (SEQ ID NO:89) 退火制造, 含有TAACGAGGAGAGGGTGTGGT (SEQ ID NO:90)、GGCATTCCA和重叠序列基序。类似地, FV2突变体1探针通过使寡核苷酸5'-AGTAAAA GGATTTGTATATTAGCCTTGCTAAGGGAGACATCTAGTGATATAAGaggctttatatacTTATCTTAAATGATG-3' (SEQ ID NO:91) 和5'-CATCATTTAAGATAAagatataaagcctCTTATATCACTAGATGTCTCCCTTAGCAAG GCTAATATACAAATCCTTTTACT-3' (SEQ ID NO:92) 退火制成。FV2突变体2探针通过使寡核苷酸 5'-AGTAAAAGGATTTGTATATTAGCCTTGCTAAGGGAGACATCaggctttatatacTGTGAACTACACTTATCTTA AATGATG-3' (SEQ ID NO:93) 和5'-CATCATTTAAGATAAGTGTAGTTCACagatataaagcctGATGTCTC CCTTAGCAAGGCTAATATACAAATCCTTTTACT-3' (SEQ ID NO:94) 退火制成。FV2突变体3探针通过 使寡核苷酸5'-AGTAAAAGGATTTGTATATTAGCCTTGCTaggctttatatacTAGTGATATAAGTGTGAACTAC ACTTATCTTAAATGATG-3' (SEQ ID NO:95) 和5'-CATCATTTAAGATAAGTGTAGTTCACACTTATATCAC TAgatataaagcctAGCAAGGCTAATATACAAATCCTTTTACT-3' (SEQ ID NO:96) 退火制成。FV2突变 体4探针通过使寡核苷酸5'-AGTAAAAGGATTTGTATaggctttatatacAAGGGAGACATCTAGTGATATAA GTGTGAACTACACTTATCTTAAATGATG-3' (SEQ ID NO:97) 和5'-CATCATTTAAGATAAGTGTAGTTCAC ACTTATATCACTAGATGTCTCCCTTgatataaagcctATACAAATCCTTTTACT-3' (SEQ ID NO:98) 退火制 成。FV2突变体5探针通过使寡核苷酸5'-AGTAAAAGGATTTGTATATTAGCCTTGCTaagcacattcgaT AGTGATATAAGaggctttatatacTTATCTTAAATGATG-3' (SEQ ID NO:99) 和5'-CATCATTTAAGATAAa gataaagcctCTTATATCACTAtcgaatgtgcttAGCAAGGCTAATATACAAATCCTTTTACT-3' (SEQ ID NO:100) 退火制成。FV2突变体6探针通过使寡核苷酸5'-AGTAAAAGGATTTGTATATTAGCCTTGCT aagcacattcgaaggctttatatacTGTGAACTACACTTATCTTAAATGATG-3' (SEQ ID NO:101) 和5'-CA TCATTTAAGATAAGTGTAGTTCACagatataaagccttcgaatgtgcttAGCAAGGCTAATATACAAATCCTTTTACT-3' (SEQ ID NO:102) 退火制成。

[0166] 需要注意, 未询问一个LV预测的结合位点。先前示出结合CTCF的对应于H19-Igf2 基因座 (H19) 的探针用作阳性对照 (Hark等人, 《自然 (Nature)》405:486-489, 2000)。未标记 的H19探针用于竞争测定。将寡核苷酸在双缓冲液 (IDT) 中退火。将纯化的全长人重组CTCF 蛋白 (Abnova, 台湾, 台北市) 和标记的寡核苷酸在室温下在20mM HEPES (pH 7.5)、50mM KCl、5mM MgCl<sub>2</sub>、3.3μM ZnSO<sub>4</sub>、1mM二硫苏糖醇、0.3mg/ml BSA、0.5μg聚 (dI:dC)、5%甘油和 0.5% 氘核X-100中温育30分钟 (Spencer等人, 《遗传学 (Genetics)》189:441-454, 2011)。然 后使用0.5X TBE运行缓冲液 (Thermo Fisher) 在6%Novex TBE凝胶 (Thermo Fisher) 上解 析结合反应。使用Odyssey 9120红外成像仪, 对凝胶成像。

#### [0167] 染色质免疫沉淀 (ChIP)

[0168] 对于HeLa对照细胞和FV A2克隆, 对ChIP输入进行PCR, 并且对于FV A2克隆, 对 ChIP产物进行PCR。简而言之, 用甲醛 (终浓度1%) 处理来自转染的FV、LV和未转染的对照 HeLa细胞的HeLa细胞克隆 (1-2 × 10<sup>7</sup> 个细胞), 并且在37°C下温育10分钟以将组蛋白交联到 DNA。在室温下用2.5M甘氨酸 (终浓度0.25M) 中和甲醛, 持续5-10分钟, 并且在2000rpm下离 心5分钟。沉淀细胞, 并且保存在-80°C下。对于ChIP, 将细胞解冻, 并且将沉淀物重新悬浮于

200 $\mu$ l SDS裂解缓冲液 (Millipore, Billerica, 马萨诸塞州) 中, 并且在冰上温育10分钟, 并且将蛋白酶抑制剂 (Pierce蛋白酶抑制剂, Thermo Fisher) (1mM苯甲酰甲基磺酰氟 (PMSF)、1 $\mu$ g/mL抑肽酶和1 $\mu$ g/mL胃抑素A) 加入到细胞裂解液中。在最适条件下 [在Covaris TM S220 (Covaris, Woburn, Massachusetts) 中, 在4 $^{\circ}$ C下, 在峰值功率:105, 占空比:10和周期/迸发:200] 下剪切裂解物, 持续70秒, 以产生200-1000个碱基对长度的交联的DNA片段。将超声处理过的样品在4 $^{\circ}$ C下以13,000rpm离心10分钟。用ChIP稀释缓冲液 (Millipore) 将上清液稀释10倍, 并且如上所述加入蛋白酶抑制剂。保留一部分样品 (1%) 作为输入样品。然后, 在搅拌下, 在4 $^{\circ}$ C下用75 $\mu$ l鲑鱼精DNA/蛋白A琼脂糖50%浆液 (Millipore) 预澄清经稀释的样品, 持续30分钟。通过短暂离心, 将琼脂糖沉淀, 并且将2 $\mu$ l抗CTCF抗体 (Millipore) 添加到2ml预澄清的上清液中, 并且在4 $^{\circ}$ C下恒速温育过夜。第二天, 添加60 $\mu$ l鲑鱼精DNA/蛋白A琼脂糖50%浆液, 在4 $^{\circ}$ C下旋转一小时, 以收集抗体/组蛋白复合物。通过在1000 $\times$ g下温和离心1分钟, 使琼脂糖沉淀, 并且弃去具有未结合的染色质的上清液。依次用1ml低盐免疫复合物 (Millipore)、高盐免疫复合物 (Millipore)、LiCl免疫复合物 (Millipore) 中的每一种洗涤蛋白A琼脂糖/抗体/染色质复合物, 并且在4 $^{\circ}$ C下, 用TE缓冲液 (Millipore) 旋转洗涤两次, 持续5分钟。去除TE洗涤缓冲液, 并且将蛋白A琼脂糖/抗体/染色质复合物重新悬浮于250 $\mu$ l新鲜洗脱缓冲液 (1%SDS, 0.1M NaHCO<sub>3</sub>) 中, 并且在室温下旋转温育15分钟。将琼脂糖珠旋转下来, 并且将过程重复两次, 并且合并洗脱液。通过向合并的洗脱液中添加20 $\mu$ l 5M NaCl (Millipore) 并且在65 $^{\circ}$ C下加热4小时, 随后向洗脱液中添加10 $\mu$ l 0.5M EDTA (Millipore)、20 $\mu$ l 1M Tris-HCl, pH 6.5 (Millipore) 和2 $\mu$ l 10mg/ml蛋白酶K来使交联反向。将此混合物在45 $^{\circ}$ C下温育一小时。通过使用PCR纯化试剂盒 (Qiagen), 从此洗脱液中回收DNA。ChIP纯化后, 通过定性PCR分析洗脱产物。利用了对应于预测的CTCF结合位点的以下引物: FV1 (5'-CGAGACTCTCCAGGTTTGGTAA-3' (SEQ ID NO:103) 和5'-GGTTCTCGAATCAAGTCGGTTT-3' (SEQ ID NO:104)), FV2 (5'-AACCGACTTGATTCGAGAACCCT-3' (SEQ ID NO:105) 和5'-GTTGGGCGCCAATTGTCAT-3' (SEQ ID NO:106)), FV5 (5'-ACTAAGGCTCCTTCTACTAGCG-3' (SEQ ID NO:107) 和5'-GTTGAAGAAGTGAATGCTGCAC-3' (SEQ ID NO:108)), FV6 (5'-TTATACCATCCATCCACCCCTC-3' (SEQ ID NO:109) 和5'-GTTTATGCCAACGA GGTCTCAA-3' (SEQ ID NO:110)), 和FV7 (5'-GCATGAGGCACTTCAGAATACA-3' (SEQ ID NO:111) 和5'-AGGCCAATACTCTTGAGCTAGT-3' (SEQ ID NO:112))。使用引物5'-CCCATCTTGCTGAC CTCAC-3' (SEQ ID NO:113) 和5'-AGACCTGGGACGTTTCTGTG-3' (SEQ ID NO:114), 再次将H19基因座用作阳性对照。H19、FV1、FV2、FV5、FV6和FV7的扩增子的大小分别为165-bp、157-bp、188-bp、110-bp、115-bp和155-bp。测定了用于HeLa对照细胞和HeLa FV克隆 (FVA2) 的ChIP输入, 以及用于HeLa FV克隆的ChIP产物。

#### [0169] 统计分析

[0170] 使用GraphPad软件 (GraphPad Software, Inc., La Jolla, CA), 使用两尾学生的未配对t测试来计算永生化测定中各组之间的统计差异。比较了SFFV和MSCV系列FV和LV载体的永生化频率与SFFV-GV的重铺频率。数据以平均值 $\pm$ 平均值的标准误差 (SEM) 表示, 并且p < 0.05的差异被认为具有统计学意义。使用单尾曼恩-惠特尼 (Mann-Whitney) U测试, 分析RT-PCR数据。

#### [0171] 结果

[0172] (A) 对比基因毒性的载体设计

[0173] 先前已经报道了脾脏病灶形成病毒 (SFFV) 载体GV在体外永生化 (IVIM) 测定中产生高频率的永生化克隆 (Modlich等人,《血液》108:2545-53,2006,Zychlinski等人,《分子治疗》16:718-25,2008),这与小鼠白血病的发病相关 (Zychlinski等人,《分子治疗》16:718-25,2008),并且在CGD和WAS基因治疗试验中MDS和白血病的发病率很高(80-100%)。在此,构建了类似的增强绿色荧光蛋白 (eGFP),其编码携带来自SFFV LTR的U3区的内部增强子/启动子的FV和LV载体,或携带来自鼠干细胞病毒 (MSCV) 的另一个GV LTR的内部增强子/启动子的FV和LV载体:(1)  $\Delta \Phi$  SF.eGFP携带内部SFFV增强子/启动子 (SFFV-FV); (2)  $\Delta \Phi$  MSCV.eGFP携带内部MSCV增强子/启动子 (MSCV-FV)。将这些载体与携带内部SFFV增强子/启动子 (SFFV-LV) 和由SFFV LTR驱动的GV, SF91-eGFP.pre (SFFV-GV) 的类似LV载体 RRL.ppt.SF.eGFP.pre进行比较。产生了携带MSCV LTR的GV, MSCV.eGFP.pre (MSCV-GV) 和携带内部MSCV增强子/启动子的LV载体 RRL.ppt.MSCV.eGFP.pre (MSCV-LV), 作为具有已知高基因毒性潜力的阳性对照 (图1A)。包含无启动子的FV载体  $\Delta \Phi$ .eGFP和模拟转导作为阴性对照。

[0174] (B) FV载体示出原代小鼠HSPC的永生化显著低于GV和LV载体。

[0175] 为了比较基因毒性潜力,将上述病毒载体用于体外永生化 (IVIM) 测定中。此测定广泛用作临床前筛查工具,并且对髓系谱系相关的基因毒性的相对定量检测特别敏感 (Du等人,《血液》106:2498-505,2005)。

[0176] 使用每个载体的优化转导方案,用SFFV和MSCV GV、LV和FV载体转导来自WT (C57BL/6J) 小鼠的骨髓的谱系阴性 (Lin<sup>-</sup>) 细胞 (Ott等人,《自然医学》12:401-9,2006, Bauer等人,《自然医学》14:93-7,2008,Puthenveetil等人,《血液》104:3445-53,2004)。如先前所述,将细胞扩增2周,并且然后克隆 (Arumugam等人,《分子治疗》17:1929-37,2009)。到2周时,未转导/模拟的Lin细胞最终分化并死亡。如果载体插入具有增殖潜力,则会发生克隆生长,从而产生“永生化”克隆。在2周和5周时评估用GV、LV和FV载体转导的细胞的重铺频率 (永生化频率)。挑选具有永生化细胞的所有孔进行扩增。用SFFV驱动的GV或LV载体转导的细胞即使在5周时仍能强劲扩增。在SFFV-GV、MSCV-GV、SFFV-LV、MSCV-LV、SFFV-FV、MSCV-FV转导的Lin细胞中, Lin细胞的转导效率 (通过GFP标记测量) 分别为91%、70%、89%、99%、93%、75%。用SFFV-GV、MSCV-GV、SFFV-LV、MSCV-LV、SFFV-FV和MSCV-FV载体转导的Lin细胞的平均VCN分别为 $8 \pm 2$ 、 $9 \pm 0.3$ 、 $10 \pm 2$ 、 $27 \pm 2$ 、 $7 \pm 0.8$ 、 $8.5 \pm 0.7$  (平均值  $\pm$  平均值的标准误差) (表1)。

[0177] 表1. IVIM测定突变体的频率

[0178]

载体	独立转导次数	S/接种的US细胞	每96孔IVIM孔(2周时)*	2周重铺频率**	载体拷贝数(VCN)	5周重铺频率***	5周重铺频率/VCN****
SFFV-GV	转导 1	US	15	0.00200	1.0	0.00200	0.002000
		S	58	0.01000	5.7	0.00900	0.001579
	转导 2	US	17	0.00200	3.0	0.00200	0.000667
		S	60	0.01000	5.4	0.01000	0.001852
	转导 3	S	96	0.04500	12	0.04500	0.003750
	转导 4	US	93	0.03400	14	0.03400	0.002429
	转导 5	US	92	0.03200	15	0.03200	0.002133
	转导 6	US	70	0.01300	8.0	0.01300	0.001625
	转导 7	S	82	0.01800	7.3	0.01800	0.002466

[0179]

平均			65	0.01844 (54分之1) #	7.9	0.01844 (54分之1)	0.002311 (452分之1)
SFFV-LV	转导 1	US	13	0.00100	11.5	0.00100	0.000087
		S	20	0.00110	8.6	0.00110	0.000128
	转导 2	US	11	0.00120	15.0	0.00120	0.000080
		S	20	0.00110	8.6	0.00110	0.000128
	转导 3	US	7	0.00080	6.0	0.00080	0.000133
	转导 4	US	25	0.00300	14.0	0.00300	0.000214
	转导 5	US	27	0.00300	10.0	0.00300	0.000300
	转导 6	US	21	0.00260	10.5	0.00260	0.000248
平均			18	0.00181 (552分之1)	10.8	0.00181 (552分之1)	0.000168 (5,952分之1)
SFFV-FV	转导 1	US	1	0.00003	3.7	0.00003	0.000008
		S	2	0.00020	8.6	0.00020	0.000023
	转导 2	US	1	0.00002	6.8	0.00002	0.000003
		S	2	0.00020	8.3	0.00010	0.000012
	转导 3	S	5	0.00026	9	0.00020	0.000022
	转导 4	US	4	0.00008	6	0.00003	0.000005
		S	3	0.00030	9.3	0.00003	0.000003
平均			3	0.00015 (6667分之1)	7.4	0.00009 (11,111分之1)	0.000012 (83,333分之1)
无启动子的FV	转导 1	US	0	0.00005	3.8	0.00000	0.000000
		US	1	0.00010	16.8	0.00001	0.000001
		US	0	0.00005	6	0.00000	0.000000
		US	1	0.00010	12	0.00001	0.000001
平均			0.5	0.00008 (12,500分之1)	9.6	0.00001 (100,000分之1)	0.000001 (1,000,000分之1)
MSCV-GV	转导 1	US	48	0.00680	9.6	0.00680	0.000708
		S	36	0.00469	9.2	0.00469	0.000510
	转导 2	US	28	0.00344	8.2	0.00344	0.000420
		S	65	0.0110	9.6	0.0110	0.001146
平均			44	0.00648 (154分之1)	9.2	0.00648 (154分之1)	0.000708 (1,412分之1)
MSCV-LV	转导 1	US	40	0.00530	25	0.0053	0.000212
		US	23	0.00270	29	0.0027	0.000093
平均			32	0.00400 (250分之1)	27	0.00400 (250分之1)	0.000148 (6,757分之1)
MSCV-FV	转导 1	S	8	0.00087	9	0.00032	0.000036
		S	9	0.00098	9	0.00021	0.000023
		US	0	0.00000	6.4	0.00000	0.000000

		S	3	0.00015	10	0.00005	0.000005
	转导 4	US	2	0.00010	6.5	0.00005	0.000008
		S	3	0.00015	10	0.00005	0.000005
	<b>平均</b>		<b>4</b>	<b>0.00038</b> (2631 分之 1)	<b>8.5</b>	<b>0.00011</b> (9091 分之 1)	<b>0.000013</b> (76,923 分之 1)
[0180]	<b>模拟</b>	不可用	0	0	0	0	0
		不可用	0	0	0	0	0
		不可用	0	0	0	0	0
		不可用	0	0	0	0	0
		不可用	0	0	0	0	0
		不可用	0	0	0	0	0
	<b>平均</b>	<b>不可用</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

[0181] 第3列示出了通过FACS为eGFP+细胞分类的转导库。NA-不适用,US-未分类,S-已分类

[0182] \*第4列描绘了在96孔板中以100个细胞/孔的极限稀释度转导、扩增和铺板100,000个Lin-细胞之后,永生化细胞的孔数。

[0183] 第5列:代表使用L-Calc软件使用Poisson统计数据计算的第2周永生细胞(重铺频率)的孔。每个载体的单独转导的平均重铺频率以阴影线表示。#括号中的数据是列出通过载体插入而永生化的Lin细胞的频率的另一种方式。例如,0.001的重铺频率意味着通过所述载体使1000个Lin细胞中的1个永生化。

[0184] \*\*对于载体转导的重铺克隆阴性的孔,计算基于假设:如果接种了97孔而不是96孔,则检测到重铺克隆(Zychlinski等人,《分子治疗》16:718-25,2008)。

[0185] 第6列代表在96孔中铺板之前,转导池的载体拷贝数(VCN)。

[0186] 第7列代表5周的重铺频率。\*\*\*由于SFFV-GV和SFFV-LV组存在大量永生化克隆,基于我们先前报道的这些克隆在2周和5周时的相等适应性,因此只有2周永生化的亚基与这些载体一起扩增5周。然而,将来自泡沫病毒载体组的所有永生化克隆扩增5周以进行进一步分析。

[0187] \*\*\*\*第8列:对载体拷贝数的永生化频率进行了归一化(假设一个永生化事件是由一个载体插入引起的,并且永生化频率与载体拷贝数呈线性相关)。

[0188] 在第5周,用具有SFFV/MSCV增强子的GV或LV载体转导之后,永生化克隆的适应性(即被重铺和扩增的能力)与2周时类似。具有SFFV-GV和SFFV-LV的永生化克隆的数目与先前报道的研究一致(Modlich等人,《分子治疗》17:1919-28,2009)。值得注意的是,与类似的SFFV-GV和MSCV-GV相比,SFFV-FV和MSCV-FV的永生化频率显著降低了两个数量级以上( $p < 0.01$ )。与先前的报道一致,与SFFV-GV相比,类似的SFFV-LV和MSCV-LV显示出永生化频率降低了10-14倍(Modlich等人,《分子治疗》17:1919-28,2009)。此外,衍生自FV转导的永生化克隆不如衍生自LV或GV转导的那些永生化克隆适应,因为与具有SFFV-GV和SFFV-LV插入的克隆相比,它们具有较低的扩增潜力,并且因此较低的5周重铺频率。与SFFV-GV、SFFV-LV和MSCV-LV克隆(分别为89.0%、76.6%和70.5%)相比,SFFV-FV和MSCV-FV克隆的平均Sca-1和c-Kit表达趋于降低(分别为69.6和57.5%),但是差异在统计学上并不显著(Mann-Whitney,一位尾注)。FV和LV载体的唯一插入相对于基因转录单位和非基因/重复序列的百

分比证明,与LV插入相比,FV具有唯一整合特性。FV在基因内、进入到非基因/重复序列和不可分配的整合位点分别为27%、62%和11%。LV分别为60%、6%和34%。SFFV-FV和MSCV FV的唯一插入数分别为121和65。为MSCV LV识别的唯一插入数为270。

[0189] 图1B-1C中的每个符号表示使用指定的载体,从一项独立的转导实验为VCN归一化的重铺频率。水平线左侧的数据点表示没有重铺克隆的独立转导。永生突变体的频率的倍数减少在图1B-1C中示出。为了排除由于VCN的永生频率差异,将每个载体的永生频率归一化到VCN(图1B-1C),可以比较分析相对基因毒性。当对VCN进行归一化时,与SFFV-GV载体相比,MSCV-GV载体的永生潜力低3倍。重要的是,与SFFV-GV载体相比,在2周时,SFFV-FV和MSCV-FV载体的永生潜力分别降低了110倍和156倍。到5周时,FV载体的永生潜力甚至进一步下降,导致与SFFV-GV载体相比,SFFV-FV和MSCV-FV载体的永生潜力分别降低了155倍和414倍。与先前的报道一致,与SFFV-GV载体相比,SFFV-LV和MSCV-LV载体显示出永生频率降低了12倍和14倍(Zychlinski等人,《分子治疗》16:718-25,2008,Modlich等人,《分子治疗》17:1919-28,2009)。支持此测定中的永生发生在载体整合之后:i)在模拟转导的祖细胞培养物中没有可检测到的永生克隆,并且ii)无启动子的FV在2周时有活孔,但是含有的细胞数量较少,并且到5周时失去,这表明它们并未真正获得永生。

[0190] (C)如通过CRISPR/Cas9介导的靶向插入确定,SFFV前病毒显示的LM02表达明显低于LV和GV前病毒序列

[0191] SFFV增强子已经在IVIM测定、小鼠和人体试验中显示为最具基因毒性的增强子之一(Ott等人,《自然医学》12:401-9,2006;Montini等人,《临床研究杂志》119:964-75,2009,Modlich等人,《分子治疗》17:1919-28,2009)。如此处所示,SFFV/MSCV增强子在FV载体中的IVIM测定中的基因毒性显著降低(150到400倍)不能由报道的FV在非基因区域中的整合的高2倍的可能性完全解释,尤其是当FV在TSS附近趋于整合时,像GV。这些结果表明,FV主链可能具有增强子阻断/绝缘子作用。为了评估载体主链的潜在增强子阻断功能,而没有启动子/增强子、转基因或整合位点的混杂影响,将前病毒形式的SFFV-GV、SFFV-LV和SFFV-FV靶向到已知导致多例继发性白血病(1-3)的逆转录病毒整合位点(RIS)处的LM02基因中。为了从整合位点效应中分离出病毒载体主链序列的基因毒性作用,设计了一种基于CRISPR/Cas9的测定,所述测定允许整合GV、LV和FV的前病毒序列,所述前病毒序列在LM02内完全相同的基因座处,并且以相同方向全部编码由SFFV启动子增强子驱动的eGFP转基因(图2A)。病毒载体序列的插入位点基于使用GV载体,对SCID-X1进行基因治疗后患者继发性白血病的先前报道(Hacein-Bey-Abina等人,《科学(Science)》302:415-9,2003)。如图2A所示,gRNA/Cas9核糖核蛋白复合物在插入位点附近产生了双链断裂(DSB)。通常通过非同源末端连接(NHEJ)或通过同源重组(HR)来修复DSB,如果提供了编码设计的通过同源臂侧接的基因修饰的供体DNA。

[0192] 识别了五个潜在的gRNA靶序列,每个序列均具有较低的预测脱靶活性,并且与插入位点非常接近(图2B)。LM02 gRNA 5效率最高,插入缺失生成效率为24.1%(gRNA 1-4的效率分别为0、13.3%、0%和1.5%),并且用于后续实验。克隆了含有针对GV、LV或FV的前病毒盒的供体质粒(图2C)。以相反的朝向构建序列,以匹配先前描述的插入的方向性(Hacein-Bey-Abina等人,《科学》302:415-9,2003)。同时用gRNA/Cas9质粒和供体(含前病

毒)质粒中的一种转染HeLa细胞,所述HeLa细胞具有非常低的LMO2 mRNA表达和LMO2蛋白缺失(Natkunam等人,《血液》109:1636-1642,2007)。转染之后大约2周,当GFP阳性细胞分选成单个细胞时,9.7%、28.2%和12.1%的GV、LV和FV转染的库显示GFP转基因表达(图3A)。然后收获克隆,并且通过PCR筛选前病毒的同源性定向插入(图3B-D)。

[0193] 当用CRISPR/Cas9编辑时,可以在给定克隆中编辑多于一个等位基因,尤其是在高度可转染的细胞系中。这将导致前病毒序列整合到给定HeLa细胞中的多个LMO2等位基因(前病毒靶向的等位基因)中,或者通过非同源末端连接修复双链断裂(DSB),形成小的插入缺失(此后称为“编辑的非靶向”等位基因)。由于HeLa细胞的LMO2mRNA表达非常低,不表达LMO2蛋白,并且前病毒被靶向LMO2基因的内含子,因此LMO2表达仅在前病毒靶向的等位基因中受到病毒增强子的影响。此外,LMO2(和其mRNA表达)的阅读框架不会被编辑的非靶向等位基因改变。前病毒质粒还可以随机整合到HeLa细胞的基因组中,导致GFP表达,但是不会影响LMO2表达。由于潜在的随机整合,因此无法通过前病毒序列的qPCR确定克隆具有的前病毒靶向的LMO2等位基因的数量。因此,间接计算了靶向等位基因拷贝数(参见材料和方法以及表2)。简而言之,对照HeLa细胞上LMO2基因座的FISH显示了四个LMO2等位基因。接下来,使用拷贝数分析检测未编辑的等位基因和含有小插入缺失的等位基因。使用较大的扩增子跨靶位点进行PCR,随后进行凝胶电泳和测序,以检测通过拷贝数分析无法检测到的较大插入缺失。对PCR产物进行测序后,在LV克隆中的三个中发现了一个261-bp缺失的LMO2等位基因,所述LV克隆最初在拷贝数分析中仅显示了一个非靶向/WT等位基因(由图9B中的星号和图9C中的箭头所示)。一个LV克隆和三个FV克隆未扩增(图9C)。给定克隆的靶向等位基因的数量等于LMO2基因座的数量减去通过拷贝数分析和PCR检测到的非靶向等位基因的数量,并且列于表2中。总体而言,用FV、LV或GV靶向了相当数量的LMO2等位基因(大多数克隆中的2-3个靶向等位基因)。

[0194] 使用没有明显内源LMO2表达的HeLa细胞系的优势在于,由于定向前病毒插入事件,废除基因表达的编辑事件不会显著影响LMO2表达的总体增加。通过用两个探针和引物组的qRT-PCR,确定生成的HeLa克隆中LMO2 mRNA的表达。选择引物以从修饰的或WT等位基因两者中检测从LMO2启动子表达的所有剪接的转录变体。来自使用两种不同的装载对照的两个引物组的数据非常类似(图4A-4D)。总体而言,结果发现,GV中的SFFV增强子表现出LMO2 mRNA表达的最大倍数增加(中值比未修饰的HeLa细胞增加 $280 \pm 23$ 倍),其次是LV中的SFFV增强子(中值增加 $200 \pm 27$ 倍)。需要注意,SFFV-GV前病毒在任一LTR处都具有两个拷贝的增强子,而SFFV-LV(和SFFV-FV)只有一个拷贝的SFFV增强子。然而,FV中相同的SFFV增强子显示LMO2 mRNA表达的增加显著降低(中值 $45 \pm 7$ 倍),表达分别比LV和GV中的SFFV增强子降低了4倍和6倍。随后,进行了蛋白质印迹分析,以检测来自用于qRT-PCR的三个代表性克隆中的LMO2蛋白的表达(图5和表2)。未检测到SFFV-FV克隆中的LMO2表达,这与模拟(未经编辑)的HeLa细胞中的基线类似。然而,在GV和LV克隆中可检测到明显更高的LMO2蛋白。总而言之,qRT-PCR和蛋白质印迹分析证实FV主链/顺式元件具有强大的增强子阻断或绝缘子作用,这可能导致SFFV增强子上调LMO2表达的能力降低。

[0195] 表2. 每个克隆中的数目或WT或插入缺失等位基因

克隆	WT 或插入缺失等位基因的数目	[计算的靶向前病毒的等位基因]**	用于蛋白质印迹
[0196] GV A7	2	2	x
GV B1	2	2	x
GV B2	2	2	x
GV B3	2	2	
GV B4	2	2	
GV B12	3	1	
LV A2*	2	2	
LV A7 <sup>#</sup>	1	3	x
LV A9	1	3	x
LV A11	1	3	x
LV B1	2	2	
LV B3	1	3	
LV B8	2	2	
LV B11*	2	2	
LV C2*	2	2	
[0197] FV A2	2	2	x
FV A3	2	2	x
FV A7	2	2	x
FV A8 <sup>#</sup>	2	2	
FV A10	1	3	
FV A11	2	2	
FV B1 <sup>#</sup>	1	3	
FV B8 <sup>##</sup>	2	2	
FV C1	2	2	
FV C2	2	2	
FV C4 <sup>#</sup>	1	3	

[0198] \*克隆LV A2、LV B11和LV C2含有具有通过拷贝数分析检测不到的261-bp内含子缺失的等位基因。\*\*假设所有克隆均具有4个LM02等位基因,如其衍生的亲代HeLa细胞。#当通过PCR评估大缺失时,克隆LV A7、FV A8、FV B1和FV C4没有PCR扩增子。##由于缺少样品材料,未测定到克隆FV B8的大缺失。对于克隆LV A7、FV A8、FV B1、FV B8和FV C4,仅在拷贝数分析中计算计算的前病毒靶向的等位基因。

[0199] (D) 在计算机绝缘子分析中,与GV和LV主链相比,FV主链中识别了更多的CTCF结合基序。

[0200] 基于结果,观察到的FV主链的增强子阻断作用的可能机制是它含有一个或多个绝缘子元件。为了测试这一点,对CTCF结合位点,脊椎动物的主要绝缘子,进行了计算机分析。使用CTCFBSDB 2.0数据库(<http://insulatordb.uthsc.edu/>),分析了预测的CTCF结合位点或共有序列的GV、LV和FV载体的前病毒序列(不包含SFFV增强子/启动子,eGFP和WPRES序列)(Ziebarth等人,《核酸研究》41:D188-94,2013),以识别CTCF结合的核心基序,以位置权重矩阵(PWM)表示。算法搜索CTCF结合位点的已识别核心基序,并且将基序表示为PWM。PWM

得分对应于由基序生成与由背景生成的观测序列的对数奇数。PWM得分>3.0表示存在显著匹配。预测工具的局限性在于,它只能返回序列中给定PWM的最佳匹配。因此,没有揭示在相同分析序列内给定基序的其它推定CTCF共有序列。为了部分地说明这一点,将序列分为几个片段,并且分别分析每个片段。

[0201] 对载体主链序列的分析分别识别了GV、LV和FV的总共4个、6个和26个PWM>3的基序。载体主链中基序的位置描绘在图6A和表3中,并且其PWM得分描绘于图6B中。除了重要的PWM数量外,GV的PWM得分通常较低,并且范围为3.3到3.4,而LV的PWM得分范围为5.1到9.9,并且FV的PWM得分范围为3.4到12.6。与GV相比,FV的基序分散在LTR和顺式序列两者中。总体而言,计算机分析表明,与GV和LV相比,FV具有更多的预测CTCF结合基序和具有较高PWM得分的基序。

[0202] 表3. 预测的CTCF结合基序

[0203]

载体	基序 PWM	基序序列	得分	距 3' HA 的位置	基序朝向	相关特征
GV	MIT_LM2 3	AAACCTACAGGTGGGGTCT T (SEQ ID NO: 115)	3.4027 8	88	-	5' LTR
	EMBL_M1	CCCCACCTGTAGGT (SEQ ID NO: 116)	3.2515	92	+	5' LTR
	MIT_LM2 3	AAACCTACAGGTGGGGTCT T (SEQ ID NO: 117)	3.4027 8	2105	-	3' LTR
	EMBL_M1	CCCCACCTGTAGGT (SEQ ID NO: 118)	3.2515	2109	+	3' LTR
LV	EMBL_M1	TTCCCCCTGGCCTT (SEQ ID NO: 119)	5.3095 8	427	-	Psi 之后
	EMBL_M2	GGAAGAGCA	8.6871 6	679	+	Psi 之后
	EMBL_M1	CTCCTCCTCCAGGT (SEQ ID NO: 120)	7.1454 2	734	-	Psi 之后
	MIT_LM7	TCCCCAGGAGCTGTTGATCC (SEQ ID NO: 121)	5.0822 4	1091	-	Psi 之后
	EMBL_M2	GGCACAGCA	9.9271 9	1145	-	Psi 之后
	EMBL_M2	GGTACAGCT	8.2519 2	3579	-	3' LTR 之前
FV	EMBL_M2	AGCATTGCA	9.5685	142	-	5' LTR
	MIT_LM2 3	ATATCACTAGATGTCTCCCT (SEQ ID NO: 122)	4.1486 1	273	-	5' LTR
	MIT_LM7	ATATCACTAGATGTCTCCCT (SEQ ID NO: 123)	8.7346 1	273	-	5' LTR
	MIT_LM2	ATATCACTAGATGTCTCCC (SEQ ID NO: 124)	8.0292 3	274	-	5' LTR
	EMBL_M1	AGACATCTAGTGAT (SEQ ID NO: 125)	5.8952 6	277	+	5' LTR

[0204]

EMBL_M2	AGCATAGCG	3.4264 9	429	+	gag 之前
EMBL_M2	GGCATTGCC	9.5646 1	1304	-	Pro-Pol
EMBL_M2	GGAATTGCA	12.351 2	1882	-	Pro-Pol/整合酶
REN_20	TGGTCCAGGAGAGGGTGGCT (SEQ ID NO: 126)	9.1435 1	2319	+	Pro-Pol/整合酶
MIT_LM2 3	GGTCCAGGAGAGGGTGGCT A (SEQ ID NO: 127)	3.4839 2	2320	+	Pro-Pol/整合酶
MIT_LM2	GGTCCAGGAGAGGGTGGCT (SEQ ID NO: 128)	5.5485 9	2320	+	Pro-Pol/整合酶
MIT_LM7	GGTCCAGGAGAGGGTGGCT A (SEQ ID NO: 129)	9.1009	2320	+	Pro-Pol/整合酶
EMBL_M1	CACCCTCTCCTGGA (SEQ ID NO: 130)	12.556 9	2322	-	Pro-Pol/整合酶
MIT_LM2	TGAACAGCAGAAGGAACA A (SEQ ID NO: 131)	4.2609 8	2570	+	ENV
MIT_LM2 3	TGAACAGCAGAAGGAACA AA (SEQ ID NO: 132)	4.5199 9	2570	+	ENV
MIT_LM7	TGAACAGCAGAAGGAACA AA (SEQ ID NO: 133)	6.2621 8	2570	+	ENV
EMBL_M2	TGAACAGCA	7.7698 1	2570	+	ENV
EMBL_M1	TTCCTTCTGCTGTT (SEQ ID NO: 134)	9.1354 2	2572	-	ENV
MIT_LM7	TAACGAGGAGAGGGTGTGG T (SEQ ID NO: 135)	4.8262 6	4898	+	BEL3
EMBL_M1	CACCCTCTCCTCGT (SEQ ID NO: 136)	5.6871 9	4900	-	BEL3
EMBL_M2	GGCATTCCA	5.7963 8	4917	-	BEL3
EMBL_M2	AGCATTGCA	9.5685	5234	-	3' LTR
MIT_LM2 3	ATATCACTAGATGTCTCCCT (SEQ ID NO: 137)	4.1486 1	5365	-	3' LTR
MIT_LM7	ATATCACTAGATGTCTCCCT (SEQ ID NO: 138)	8.7346 1	5365	-	3' LTR
MIT_LM2	ATATCACTAGATGTCTCCC (SEQ ID NO: 139)	8.0292 3	5366	-	3' LTR
EMBL_M1	AGACATCTAGTGAT (SEQ ID NO: 140)	5.8952 6	5369	+	3' LTR

[0205] (E) 在FV主链中映射了独特的36-bp增强子阻断元件

[0206] 为了评估CTCF与HeLa细胞内的前病毒序列的结合,进行了CTCF结合的DNA的ChIP纯化,随后针对FV前病毒序列内的预测结合位点进行了定性PCR(图7A)。使用HeLa对照细胞,对ChIP输入材料的分析显示仅扩增了H19-Igf2基因座,已知的CTCF结合位点。然而,输入材料以及来自FV克隆之一(FVA2)的ChIP纯化DNA两者均显示了五个测试位点的扩增,并且具有高水平的预测结合。虽然ChIP分析显示存在CTCF结合,但是通过此ChIP-PCR测定,测定区域的紧密接近限制了位点之间的分辨率。无论如何,ChIP-PCR测定证明了CTCF与FV前病毒序列的细胞内结合。

[0207] 然后使用电泳迁移率变动分析 (EMSA) 绘制LV和FV内预测的CTCF结合位点。将对应于通过计算机分析预测的CTCF结合位点,用荧光染料在两个5'末端处进行标记的80-90-bp DNA片段用作EMSA探针。探测了LV中六个预测的结合位点中的五个,并且探针LV1、LV2和FV8含有两个预测的CTCF结合基序序列。使用重组人CTCF进行EMSA。阳性对照是含有已知以高亲和力结合CTCF的共有蛋白的H19寡核苷酸。

[0208] LV探针均未显示出与CTCF的任何结合。然而,对应于序列ATATCACTAGATGTCTCCCT (SEQ ID NO:141)的探针FV2(位于LTR中,并且含有PWM得分为4.1、8.7、8.0和5.9的四个基序)表现出明显的带移(图7B)。另外,标记的探针可以与未标记的H19探针竞争(图7C)。在计算机上分析了FV2探针的序列,以预测CTCF结合位点。除了先前预测的位点,还识别了第二个位点5'-TG TAGTTCA-3',得分为6.8。FV2探针的中央区域分为四个区域(1-4)。区域1含有新识别的预测结合位点。区域2-3含有原始的预测结合位点。设计了六个突变体探针,其用加扰的DNA序列代替了四个区域中的一个或多个(图7D)。分析序列以确保没有创建新的预测的CTCF结合位点。突变区1、2或3降低了CTCF结合(EMSA带强度,图7E)。突变区4似乎没有作用。将区域1和3一起突变,或将2和3一起突变进一步降低了CTCF的结合。因此,结果发现CTCF结合了由FV2探针的区域1-3定义的36-bp序列。使用定义的序列进行blast搜索(<https://blast.ncbi.nlm.nih.gov/>)并未显示与除泡沫病毒以外的任何序列的匹配。

[0209] (F) 通过靶向插入到LM02基因中来验证绝缘子功能

[0210] 为了验证所定义的CTCF结合序列的绝缘子功能:1)从前病毒SFFV-FV序列中精确切除了36-bp序列,其余序列保持完整;以及2)将建议的绝缘子插入到SFFV-LV前病毒序列的两个LTR中。然后,像以前一样,使用基于CRISPR/Cas9的靶向插入性基因毒性测定,将修饰的前病毒序列插入到LM02基因中。使用Hs001534473\_m1引物/探针组和PP1A内源对照,通过qPCR确定相对于对照HeLa细胞以及先前使用的LV和FV克隆的LM02的表达(图8)。从FV LTR中移除绝缘子导致相对LM02表达增加5倍以上(9.4到47.4,  $p=0.001$ )。将序列插入到LV LTR中导致相对LM02表达下降13倍以上(21.0到1.6,  $p=0.0002$ )。有趣的是,将绝缘子置于SFFV-LV的LTR中导致LM02表达仅比对照HeLa细胞高1.6倍,并且显著低于含有绝缘子的原始SFFV-FV ( $p=0.002$ )。与先前使用GV、LV和FV的测定相比,此测定中看到的总体较低的相对表达是由于在对照HeLa细胞中观察到的LM02表达较高。基于先前的测定,插入的前病毒序列的数量与LM02表达没有很好的相关性。仅通过qPCR确定的拷贝数预测,没有绝缘子的所有新FV克隆均具有三个正确放置的前病毒序列。在含有绝缘子6、1和2的新LV克隆中,克隆分别含有前病毒序列的1个、2个和3个拷贝。同样,插入的前病毒序列的数目似乎与表达水平不相关。

[0211] 两者合计,结果表明FV LTR含有强大的36-bp CTCF结合基序,所述基序具有有效的CTCF结合,其产生增强子阻断作用,并且用于保护附近的基因免受已递送的转基因的增强子活性的影响。这些数据为FV载体中SFFV(和MSCV)的非常低的永生化潜力提供了新颖的见解,证明了先前报道的和重要的机制有助于降低携带强病毒增强子的FV的基因毒性。

[0212] 讨论

[0213] 载体驱动基因毒性主要是由于使用了(LTR中)的强增强子(Modlich等人,《血液》108:2545-53,2006,Zychlinski等人,《分子治疗》16:718-25,2008,Kustikova等人,《科学》308:1171-4,2005,Maruggi等人,《分子治疗》17:851-6,2009)和载体的整合位点偏好引

起的 (Montini 等人,《临床研究杂志》119:964-75,2009, Montini 等人,《自然生物技术》24:687-96,2006)。GV LTR 增强子普遍存在并强烈增强转基因的表达,从而导致治疗性纠正,而且还增强了侧接转基因插入位点的基因的表达,这可能导致白血病。此外,GV 和 LV 整合酶两者将前病毒靶向基因富集区域的时间接近 60-70%,而 GV 载体倾向于在 TSS 附近整合,并且 LV 载体非常倾向于在活性转录单元的内含子中整合,并且因此具有较低的激活细胞基因转录的倾向 (Modlich 等人,《血液》108:2545-53,2006, Zychlinski 等人,《分子治疗》16:718-25,2008, Kustikova 等人,《科学》308:1171-4,2005, Maruggi 等人,《分子治疗》17:851-6,2009)。总体而言,当与携带相同增强子的 GV 载体相比时,携带 GV LTR 增强子的 LV 载体在体外和体内实验系统中的潜在基因毒性趋于降低大约 10-20 倍 (Modlich 等人,《血液》108:2545-53,2006, Montini 等人,《临床研究杂志》119:964-75,2009)。

[0214] 通过 qRT-PCR 和蛋白质印迹测定 LM02 靶向的 GV、LV 和 FV 克隆的 LM02 表达反映了永生测定结果:如与 GV 相比,观察到 FV 的 SFFV 增强子作用降低了 6 倍。这可以部分地通过以下事实解释:GV 载体具有两个 SFFV 增强子,而 FV 载体只有一个。然而,当与 LV 相比时,FV 的增强子效果也降低了 4 倍,其也只在内部放置了单个 SFFV 增强子,与 FV 类似。类似地,对 LM02 蛋白表达的蛋白质印迹分析表明,与 GV 克隆相比,LV 克隆中 LM02 表达降低了 3 倍。FV 克隆无法检测到 LM02 表达,表明 FV 载体序列非常有效地阻断了 SFFV 增强子。

[0215] 如 SFFV 等强增强子的效果的这种显著降低导致在 FV 载体顺式序列中意外发现了绝缘子。在 CTCF 的结合位点的计算机分析中,主要脊椎动物绝缘子蛋白显示,仅在增强子/启动子之前,每个 LTR 的 U3 区开始处 GV 的 CTCF 位置权重矩阵 (PWM),但是所有 PWM 得分均低于 3.5 (表 3 列出了基序的位置)。LV 具有 6 个唯一的 CTCF PWM,得分范围为 5.3 到 9.9。FV 具有 26 个 CTCF PWM,其 SIN LTR 中存在的得分范围为 3.5 到 12.6,并且在 FV 载体中重新训练了 gag、pro-pol、env 和 bel 顺式序列的部分,其中十一个是不重叠的 CTCF PWM,并且两个处于 LTR 的 R 区域内。如果使用更严格的 PWM 得分截止值,例如 3.5,GV 失去其所有 CTCF 结合位点,而 LV 和 FV CTCF 结合位点中的仅 2 个均不会失去。然而,未观察到可检测的 CTCF 与测试的计算机预测的 LV CTCF 结合位点的结合。

[0216] 使用 EMSA 证明了 CTCF 与 FV LTR 序列内预测的结合位点之一的结合,并且将 CTCF 结合定位到 36-bp 区域。LTR 内的位置很明显,因为这将在前病毒序列的两端提供绝缘。最近,实验性插入到 LV 或 FV LTR 中的 CTCF 结合位点已经显示出降低了基因毒性 (Liu 等人,《自然生物技术》33:198-203,2015, Browning 等人,《人类基因治疗》27:255-266,2015, Browning 等人,《基因治疗 (Gene Ther.)》24:187-198,2017) 我们的研究表明 FV LTR 内已经存在固有的 CTCF 结合位点。通过从 FV 前病毒序列中去除 CTCF 结合位点,可以消除 LM02 激活测定中观察到的基因毒性降低。相反,在同一测定中,将此位点插入到 LV LTR 中将基因毒性显著降低到仅略高于对照的水平。

[0217] 重要的是,这些研究导致识别了具有强增强子阻断活性的新的绝缘子序列,使得这种强绝缘子片段的使用成为构建核酸构建体,如病毒载体中有吸引力的工具,用于表达需要强增强子的转基因。另外,此序列可以转座到其它载体中,有可能实现绝缘子功能,如降低基因毒性。据报道,由于它们的增强子活性弱/可忽略,在通过延长因子-1EF-1 (EFS) 和磷酸甘油酸激酶 (PGK) 启动子表达转基因的 SIN GV 和 LV 载体中,插入性永生化风险已经大大降低 (Zychlinski 等人,《分子治疗》16:718-25,2008)。然而,使用细胞启动子的主要限

制是获得和维持一致高水平的转基因表达的问题。在如LAD (Hunter等人,《人类基因治疗》22:689-96,2011)、CGD (Barde等人,《基因治疗》18:1087-97,2011,Chiriaco等人,《分子治疗》22:1472-83,2014)和WAS (Aiuti等人,《科学》341:1233151,2013,Hacein-Bey Abina等人,《美国医学会杂志(JAMA)》313:1550-63,2015)等疾病中,在需要高水平的转基因表达实现治疗效果的情况下,除非存在高VCN并且使用强增强子两者,否则细胞启动子不足以介导治疗效果。实际上,我们最近已经报告,LV中来自细胞或内源性启动子的穿孔素的表达仅部分纠正了噬血细胞淋巴组织细胞增生症(HLH)表型,并且强病毒增强子对于完全纠正是必需的(Tiwari等人,《人类基因治疗》27(10):847-859,2016)。

[0218] 此公开显示,在FV载体内的强病毒LTR增强子是可应用的,因为整合在非基因区域中的组合趋势与固有的绝缘子作用组合可以显著降低基因毒性。

[0219] 总而言之,直接比较由类似的GV、LV和FV载体诱导的基因毒性潜力显示,通过插入携带某些最具基因毒性的病毒增强子的FV载体显著降低HSPC的永生化潜力。在特定的LM02基因座上,使用CRISPR/Cas9介导的GV、LV和FV病毒序列的靶向整合来询问载体主链对基因毒性的唯一影响揭示了,这种基因毒性降低的机制在很大程度上是由于紧密结合CTCF的FV LTR中的增强子阻断性绝缘子作用。本公开提供了对FV载体的基因毒性有价值的见解,并且与临床载体设计有关。识别的小(36-bp)绝缘子序列可能在多种载体中具有潜在用途,尤其是在添加强增强子对疾病纠正至关重要的情况下。

[0220] 其它实施例

[0221] 本说明书中公开的所有特征可以以任何组合组合。本说明书中公开的每个特征可以由具有相同、等同或类似目的的替代性特征替代。因此,除非另有明确说明,否则所公开的每个特征仅是一系列等同或类似特征的实例。

[0222] 根据以上描述,本领域技术人员可以容易地确定本发明的基本特性,并且在不脱离其精神和范围的情况下,可以对本发明进行各种改变和修改,以使其适应各种用途和条件。因此,其它实施例也在权利要求内。

[0223] 等同物

[0224] 尽管本文已经描述和展示了若干个发明实施例,但本领域普通技术人员将容易想到用于执行本文描述的功能和/或获得这些结果和/或这些优点中的一个或多个优点的各种其它装置和/或结构,并且此类变型和/或修改中的每一个被认为是在本文所述的发明实施例的范围内。更一般来讲,本领域的技术人员将容易认识到,本文中描述的所有参数、尺寸、材料以及配置意味着是示例性的,并且实际参数、尺寸、材料和/或配置将取决于本发明传授内容所用于的一种或多种具体应用。仅仅使用常规实验,本领域的技术人员将认识到或能够确认本文所述的具体发明实施例的许多等效物。因此,应理解,前述实施例是仅通过实例方式来介绍的,并且在所附权利要求和其等效物的范围内,本发明实施例可以按与具体描述和要求不同的方式来实践。本公开的发明实施例涉及本文中描述的每个单独的特征、系统、物品、材料、试剂盒和/或方法。此外,两个或更多个此类特征、系统、物品、材料、试剂盒和/或方法的任何组合,如果此类特征、系统、物品、材料、试剂盒和/或方法并不相互矛盾,被包含在本公开的发明范围内。

[0225] 如本文所定义和使用,所有定义应被理解为字典定义上的控制、通过引用并入的文档中的定义和/或所定义术语的普通含义。

[0226] 本文所公开的所有参考文献、专利和专利申请相对于其各自被引用的主题通过引用并入,在一些情况下,其可以涵盖文档的整体。

[0227] 如本文在说明书中使用的,不定冠词“一(a)”和“一个(an)”除非明确作出相反指示,否则应当理解为是指“至少一个”。

[0228] 如本文在说明书中使用的,短语“和/或”应当理解为是指这样联合的要素中的“任一个或两个”,即,要素在一些情况下共同存在而在其它情况下分开存在。用“和/或”列出的多个要素应以相同的方式解释,即,如此连接的要素中的“一个或多个”。除了用“和/或”短语具体标识的要素,其它要素可以任选地存在,无论是与具体标识的那些要素相关还是不相关。因此,作为非限制性实例,在一个实施例中,当与如“包括”等开放式语言结合使用时,对“A和/或B”的引用可以仅指A(任选地包含除B以外的要素);在另一个实施例中,仅指B(任选地包含除A以外的要素);在又一个实施例中,指A和B两者(任选地包含其它要素);等。

[0229] 如本文中在本说明书和权利要求中所使用的,“或”应被理解为具有与如上所定义的“和/或”相同的含义。例如,当将列表中的用于分开时,“或”或“和/或”应被解释为包容性的,即包括多个要素或要素清单中的至少一个要素、而且还包括多于一个要素,以及可选地其它未列出的项。仅仅清楚地指示相反的用语,如“……中的仅一个”或“……中的确切一个”或者在权利要求中使用“由……组成”将指包括多个要素或要素清单中的恰好一个要素。一般而言,当之前有排他性术语、比如“任一个”、“……之一”、“……中的仅一个”、“……中的确切一个”时,本文中使用的用语“或”应当仅被解释为指示排他性替代品(即,“一个或另一个、而不是两个”)。当在权利要求中使用“基本上由……组成”应当具有如在专利法领域中所使用的普通含义。

[0230] 如在本说明书和权利要求中所使用的,关于一个或多个要素的清单的短语“至少一个”应被理解为是指选自要素清单中的任一个或多个要素的至少一个要素、但不一定包括要素清单内具体列出的每一个要素的至少一个、并且不排除要素清单中要素的任何组合。这个定义还允许可以可选地存在除了在短语“至少一个”所指的要素清单内具体标识的要素之外的要素,而无论是否与具体标识的那些元素相关还是不相关。因此,作为非限制性实例,在一个实施例中,“A和B中的至少一个”(或等效地,“A或B中的至少一个”,或等效地“A和/或B中的至少一个”)可以指任选地包含多于一个A,不存在B(并且任选地包含除B以外的元素)的至少一个;在另一个实施例中,可以指任选地包含多于一个B,不存在A(并且任选地包含除A以外的元素)的至少一个;在又一个实施例中,可以指任选地包含多于一个A的至少一个,以及任选地包含多于一个B(以及任选地包含其它元素)的至少一个;等。

[0231] 还应该理解,除非明确相反地指出,否则在本文要求保护的包含多于一个步骤或动作的任何方法中,方法的步骤或动作的顺序不一定限于叙述方法的步骤或动作的顺序。

## 序列表

<110> 儿童医院医学中心(Children's Hospital Medical Center)

<120> 泡沫病毒载体携带强绝缘子

<130> C1367.70030W000

<150> US 62/559,751

<151> 2017-09-18

<160> 147

<170> SIPOSequenceListing 1.0

<210> 1

<211> 36

<212> DNA

<213> 人工序列(Artificial sequence)

<400> 1

gtgtagttca cacttatatc actagatgtc tccett 36

<210> 2

<211> 36

<212> DNA

<213> 人工序列(Artificial sequence)

<400> 2

aagggagaca tctagtgata taagtgtgaa ctacac 36

<210> 3

<211> 44

<212> DNA

<213> 人工序列(Artificial sequence)

<400> 3

tttatttgtg aaatttgtga tgctattgct ttatttgtaa accg 44

<210> 4

<211> 140

<212> DNA

<213> 人工序列(Artificial sequence)

<400> 4

tcgacgcagg actcggettg ctgaagcgcg cacggcaaga ggcgaggggc ggcgactggt 60

gagtacgcca aaaattttga ctacgaggagg ctagaaggag agagatgggt gcgagagcgt 120

cagtattaag cgggggagaa 140

<210> 5

<211> 858

<212> DNA

<213> 人工序列(Artificial sequence)

<400> 5

gatcttcaga cctggaggag gagatatgag ggacaattgg agaagtgaat tatataaata 60  
 taaagtagta aaaattgaac cattaggagt agcaccacc aaggcaaaga gaagagtgg 120  
 gcagagagaa aaaagagcag tgggaatagg agctttgttc cttgggttct tgggagcagc 180  
 aggaagcact atgggcgag cgtcaatgac gctgacgta caggccagac aattattgtc 240  
 tggtatagtg cagcagcaga acaatttgct gagggctatt gaggcgcaac agcatctgtt 300  
 gcaactcaca gtctggggca tcaagcagct ccaggcaaga atcctggctg tggaaagata 360  
 cctaaaggat caacagctcc tggggatttg gggttgctct ggaaaactca tttgcaccac 420  
 tgctgtgcct tggaaatgcta gttggagtaa taaatctctg gaacagattt ggaatcacac 480  
 gacctggatg gagtgggaca gagaaattaa caattacaca agcttaatac actccttaat 540  
 tgaagaatcg caaaaccagc aagaaaagaa tgaacaagaa ttattggaat tagataaatg 600  
 ggcaagtttg tggaaattgg ttaacataac aaattggctg tggatatata aattattcat 660  
 aatgatagta ggagccttg taggtttaag aatagttttt gctgtacttt ctatagtga 720  
 tagagttagg cagggatatt caccattatc gtttcagacc cacctccca ccccgagggg 780  
 acccgacagg cccgaaggaa tagaagaaga aggtggagag agagacagag acagatccat 840  
 tcgattagtg aacggatc 858

<210> 6

<211> 355

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 6

atgggtgcga gagcgtcagt attaagcggg ggagaattag atcgcgatgg gaaaaaattc 60  
 ggtaaggcc agggggaaag aaaaaatata aattaaaca tatagtatgg gcaagcaggg 120  
 agctagaacg attcgcagtt aatcctggcc tgtagaaac atcagaaggc tgtagacaaa 180  
 tactgggaca gctacaacca tcccttcaga caggatcaga agaacttaga tcattatata 240  
 atacagtagc aaccctctat tgtgtgcatc aaaggataga gataaaagac accaaggaag 300  
 cttagacaa gatagaggaa gagcaaaaca aaagtaagac caccgcacag caagc 355

<210> 7

<211> 205

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 7

ccatctgttg tttgcccctc ccccgctgct tecttgacc tggaaggtgc cactcccact 60  
 gtcctttcct aataaaatga ggaaattgca tcgattgtc tgagtaggtg tcattctatt 120  
 ctgggggggtg ggggtggggca ggacagcaag ggggaggatt gggaagacaa tagcaggcat 180  
 gctgggggatg cgggtgggctc tatgg 205

<210> 8

<211> 1211

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 8

```

gagctcacgg ggacagcccc cccccaaagc cccagggat gtaattacgt ccctccccg 60
ctagggggca gcagcgagcc gcccggggct ccgctccggt ccggcgctcc ccccgcatcc 120
ccgagccggc agcgtgcggg gacagcccgg gcacggggaa ggtggcacgg gatcgctttc 180
ctctgaacgc ttctcgctgc tctttgagcc tgcagacacc tggggggata cggggaaaaa 240
gctttaggct gaaagagaga tttagaatga cagaatcata gaacggcctg gtttgcaaag 300
gagcacagtg ctcatccaga tccaaccccc tgctatgtgc aggtcatca accagcagcc 360
caggctgccc agagccacat ccagcctggc ctgtaatgcc tgcagggatg gggcatccac 420
agcctccttg ggcaacctgt tcagtgcgtc accacctct gggggaaaaa ctgcctcctc 480
atatccaacc caaacctccc ctgtctcagt gtaaagccat tcccccttgt cctatcaagg 540
gggagtttgc tgtgacattg ttggtctggg gtgacacatg tttgccaatt cagtgcata 600
cggagaggca gatcttgggg ataaggaagt gcaggacagc atggacgtgg gacatgcagg 660
tgttgagggc tctgggacac tctccaagtc acagcgttca gaacagcctt aaggataaga 720
agataggata gaaggacaaa gagcaagtta aaaccagca tggagaggag cacaaaaagg 780
ccacagacac tgctggtccc tgtgtctgag cctgcatggt tgatggtgtc tggatgcaag 840
cagaaggggt ggaagagctt gcctggagag atacagctgg gtcagtagga ctgggacagg 900
cagctggaga attgccatgt agatgttcat acaatcgtca aatcatgaag gctggaaaag 960
ccctccaaga tccccaaagc caacccaac ccaccaccg tgcctactgg ccatgtcctt 1020
cagtgccaca tccccacagt tcttcatcac ctccaggac ggtgaccccc ccacctcctt 1080
gggcagctgt gccactgcag caccgctctt tggagaaggt aatcttgct aatccagcc 1140
cgacctccc ctggcacaac gtaaggccat tatctctcat ccaactccag gacggagtca 1200
gtgagaatat t 1211

```

<210> 9

<211> 250

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 9

```

gggacagccc cccccaaaag cccccaggga tgtaattacg tccctcccc gctagggggc 60
agcagcgagc cgcccggggc tccgctccgg tccggcgctc ccccgcatc cccgagccgg 120
cagcgtgcgg ggacagcccg ggcacgggga aggtggcacg ggatcgcttt cctctgaacg 180
cttctcgctg ctctttgagc ctgcagacac ctggggggat acggggaaaa agcttttaggc 240
tgaaagagat 250

```

<210> 10

<211> 400

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 10

```

ctgagcctgc atgtttgatg gtgtctggat gcaagcagaa ggggtggaag agcttgctctg 60

```

gagagataca gctgggtcag taggactggg acaggcagct ggagaattgc catgtagatg 120  
 ttcatacaat cgtcaaatca tgaaggctgg aaaagccctc caagatcccc aagaccaacc 180  
 ccaaccacc caccgtgccc actggccatg tccctcagtg ccacatcccc acagttcttc 240  
 atcacctcca gggacggtga cccccccacc tccgtgggca gctgtgccac tgcagcaccg 300  
 ctctttggag aaggtaaate ttgctaaate cagccccgacc ctcccctggc acaacgtaag 360  
 gccattatct ctcatccaac tccaggacgg agtcagttag 400

<210> 11

<211> 16

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 11

aaaagaaaag ggggga 16

<210> 12

<211> 583

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 12

tcgacaatca acctctggat taaaaattt gtgaaagatt gactggtatt cttactatg 60  
 ttgctccttt tacgctatgt ggatacgtg ctttaatgcc tttgtatcat gctattgctt 120  
 cccgtatggc tttcattttc tctccttgt ataaatcctg gttgctgtct ctttatgagg 180  
 agttgtggcc cgttgtcagg caacgtggcg tgggtgacac tgtgtttgct gacgcaacc 240  
 ccaactggttg gggcattgcc accacctgtc agctccttc cgggactttc gctttcccc 300  
 tccctattgc cacggcggaa ctcatcgccg cctgccttgc ccgctgctgg acaggggctc 360  
 ggctgttggg cactgacaat tccgtggtgt tgcggggaa gctgacgtcc tttccatggc 420  
 tgctcgccctg tgttgccacc tggattctgc gcgggacgtc cttctgctac gtcccttcgg 480  
 ccctcaatcc agcggacctt cctcccgcg gctgctgcc ggctctgcgg cctcttccgc 540  
 gtcttcgcct tcgccctcag acgagtcgga tctcccttg ggc 583

<210> 13

<211> 30

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 13

aatcctttac atggagaagt tataggtctt 30

<210> 14

<211> 29

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 14

atctgaaate tctcaatttg tccccacca 29

<210> 15  
<211> 21  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 15  
gaaccttggtg tctctcatcc c 21  
<210> 16  
<211> 38  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 16  
gacccgggag atctgaattc agtggcacag cagttagg 38  
<210> 17  
<211> 33  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 17  
aattcctaac tgcttgcca ctgaattcag atc 33  
<210> 18  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 18  
gtctatgagg agcaggagta 20  
<210> 19  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 19  
gacccgggag atctgaattc 20  
<210> 20  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 20  
cctccttccc tgtaatactc 20  
<210> 21  
<211> 18  
<212> DNA

<213> 人工序列 (Artificial sequence)  
<400> 21  
agtggcacag cagttagg 18  
<210> 22  
<211> 25  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 22  
ctggggacca tctgttcttg gccct 25  
<210> 23  
<211> 44  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 23  
aattctctag tatgctactc gcaccgatta tctccgctgt cagt 44  
<210> 24  
<211> 40  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 24  
actgacagcg gagataatcg gtgcgagtag catactagag 40  
<210> 25  
<211> 25  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 25  
gacttgtggt ctcgctgttc cttgg 25  
<210> 26  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 26  
actgacagcg gagataatcg 20  
<210> 27  
<211> 25  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 27  
ggtctcctct gagtgattga ctacc 25

<210> 28  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 28  
gtgcgagtag catactagag 20  
<210> 29  
<211> 22  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 29  
tttaggttgc cctgaaaagg tg 22  
<210> 30  
<211> 22  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 30  
gccaaacact cctaggctct tg 22  
<210> 31  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 31  
gtctctcgca gccacatggg 20  
<210> 32  
<211> 35  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 32  
cccgtaatgc agaagaaaac catgggctgg gaggc 35  
<210> 33  
<211> 35  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 33  
gcctcccage ccatggtttt cttctgcatt acggg 35  
<210> 34  
<211> 20  
<212> DNA

<213> 人工序列 (Artificial sequence)  
<400> 34  
gataccaata gatatcaatc 20  
<210> 35  
<211> 26  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 35  
caccgggata ccaatagata tcaatc 26  
<210> 36  
<211> 26  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 36  
aaacgattga tatctattgg tatccc 26  
<210> 37  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 37  
atcaccagat tgatatctat 20  
<210> 38  
<211> 27  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 38  
caccgggatc accagattga tatctat 27  
<210> 39  
<211> 27  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 39  
aaacatagat atcaatctgg tgatccc 27  
<210> 40  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 40  
aattgcatag tcgtgaagtc 20

<210> 41  
<211> 27  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 41  
caccgggaat tgcatagtcg tgaagtc 27  
<210> 42  
<211> 27  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 42  
aaacgacttc acgactatgc aattccc 27  
<210> 43  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 43  
attgcatagt cgtgaagtca 20  
<210> 44  
<211> 27  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 44  
caccgggatt gcatagtcgt gaagtca 27  
<210> 45  
<211> 27  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 45  
aaactgactt cacgactatg caatccc 27  
<210> 46  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 46  
tcgtgaagtc agggcttcta 20  
<210> 47  
<211> 27  
<212> DNA

<213> 人工序列 (Artificial sequence)  
<400> 47  
caccgggtcg tgaagtcagg gcttcta 27  
<210> 48  
<211> 27  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 48  
aaactagaag ccctgacttc acgaccc 27  
<210> 49  
<211> 21  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 49  
gcttgggttt tacacgtctt c 21  
<210> 50  
<211> 23  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 50  
tcagctagaa aacaagtact tgc 23  
<210> 51  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 51  
agtaaaagga tttgtatatt agccttgcta agggagacat ctagtgatat aagtgtgaac 60  
tacacttatc ttaaattgatg 80  
<210> 52  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 52  
agtaaaagga tttgtatatt agccttgcta agcacattcg atagtgatat aagaggcttt 60  
atatcttatc ttaaattgatg 80  
<210> 53  
<211> 36  
<212> DNA  
<213> 人工序列 (Artificial sequence)

<400> 53  
aagggagaca tctagtgata taagtgtgaa ctacac 36  
<210> 54  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 54  
cgagcgttgg taagagaagc 20  
<210> 55  
<211> 19  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 55  
gagatctgtc ccgctagca 19  
<210> 56  
<211> 33  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 56  
ggataattta caaataaacc cgacttatat tcg 33  
<210> 57  
<211> 23  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 57  
tggggaacaa gtacaatttt gtg 23  
<210> 58  
<211> 24  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 58  
caatgtggtg atatcaatct ggtg 24  
<210> 59  
<211> 25  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 59  
acaagcgtaa attgcatagt cgtga 25  
<210> 60

<211> 21  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 60  
cttggtttat gaatctggct c 21  
<210> 61  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 61  
gcctttagca gtagaacac 20  
<210> 62  
<211> 24  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 62  
acatgctggg aatcgacttg tgat 24  
<210> 63  
<211> 89  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 63  
acaagataga ggaagagcaa aacaaaagta agaccaccgc acagcaagcg gccgctgac 60  
ttcagacctg gaggaggaga tatgaggga 89  
<210> 64  
<211> 89  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 64  
tcctcatat ctctctctcc aggtctgaag atcagcgcc gcttgctgtg cgggtgtctt 60  
acttttgttt tgctcttctct ctatcttgt 89  
<210> 65  
<211> 14  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 65  
ctctctctcc aggt 14  
<210> 66  
<211> 83

<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 66  
gatacctaaa ggatcaacag ctctgggga tttggggtg ctctggaaa ctcatttgca 60  
ccactgctgt gccttgaat gct 83  
<210> 67  
<211> 83  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 67  
agcattccaa ggcacagcag tgggtcaaat gagttttcca gagcaacccc aaatccccag 60  
gagctgttga tcctttaggt atc 83  
<210> 68  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 68  
tccccaggag ctgttgatcc 20  
<210> 69  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 69  
gtcggggaag ctgacgtcct ttcgaattcg atatcaagct gtacctttaa gaccaatgac 60  
ttacaaggca gctgtagatc 80  
<210> 70  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 70  
gatctacagc tgccttgtaa gtcattggtc ttaaaggtag agcttgatat cgaattcgaa 60  
aggacgtcag cttccccgac 80  
<210> 71  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 71  
tccattaaca ctctgcttat agattgtaag ggtgattgca atgctttctg cataaaactt 60  
tggttttctt gttaatcaat 80

<210> 72

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 72

attgattaac aagaaaacca aagttttatg cagaaagcat tgcaatcacc cttacaatct 60  
ataagcagag tgtaaatgga 80

<210> 73

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 73

agtaaaagga tttgtatatt agccttgcta agggagacat ctagtgatat aagtgtgaac 60  
tacacttatc ttaaattgatg 80

<210> 74

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 74

catcatttaa gataagtga gttcacactt atatcactag atgtctccct tagcaaggct 60  
aatatacaaa tccttttact 80

<210> 75

<211> 20

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 75

atatcactag atgtctccct 20

<210> 76

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 76

tcgggtttat ttgtaaatta tcctagga cctccgagca tagcgggagg catataaaag 60  
ccaatagaca atggctagca 80

<210> 77

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 77

tgctagccat tgtctattgg cttttatatg cctcccgcta tgctcggagg tccttagga 60  
taatttacaataaaacccga 80

<210> 78

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 78

ggcatcagcc tacaaatacc agtattcata ctgaaggcaa tgccttagca gataagcttg 60  
ccaccaagg aagttatgta 80

<210> 79

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 79

tacataactt ccttgggtgg caagcttate tgctagggca ttgccttcag tatgaatact 60  
ggtatttgta ggctgatgcc 80

<210> 80

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 80

cgcaactggt aaatctctca atgtactcac tagtattgca attccaaagg tgattcactc 60  
tgatcaaggt gcagcattca 80

<210> 81

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 81

tgaatgctgc accttgatca gagtgaatca ctttgggaat tgcaatacta gtgagtacat 60  
tgagagattt aacagttgag 80

<210> 82

<211> 80

<212> DNA

<213> 人工序列 (Artificial sequence)

<400> 82

ctcgttctcg gtctctgtt gttggccaat tggccagga gagggtggct aggctgctt 60  
ctttgagacc tcgttggeat 80

<210> 83

<211> 80

<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 83  
atgccaacga ggtctcaaag aagcaggcct agccaccctc tcctggacca attggccaac 60  
aacaggagac caggaacgag 80  
<210> 84  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 84  
tggtcagga gaggtggct 20  
<210> 85  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 85  
atgaggcact tcagaatata acaactgtga ctgaacagca gaaggaacaa attatactgg 60  
acattcaaaa tgaagaagta 80  
<210> 86  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 86  
tacttcttca ttttgaatgt ccagtataat ttgttccttc tgctgttcag tcacagttgt 60  
tgtattctga agtgcctcat 80  
<210> 87  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 87  
tgaacagcag aaggaacaaa 20  
<210> 88  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 88  
tatggaagct tatggacctc agagaggaag taacgaggag agggtgtggt ggaatgccac 60  
tagaaaccag ggaaaacaag 80  
<210> 89

<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 89  
cttgttttcc ctggtttcta gtggcattcc accacacct ctccctcgta cttcctctct 60  
gaggtccata agcttccata 80  
<210> 90  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 90  
taacgaggag agggtgtggt 20  
<210> 91  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 91  
agtaaaagga tttgtatatt agccttgcta agggagacat ctagtgatat aagaggcttt 60  
atatcttatac ttaaatagatg 80  
<210> 92  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 92  
catcatttaa gataagatat aaagcctctt atatcactag atgtctccct tagcaaggct 60  
aatatacaaaa tccttttact 80  
<210> 93  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 93  
agtaaaagga tttgtatatt agccttgcta agggagacat caggctttat atctgtgaac 60  
tacacttatac ttaaatagatg 80  
<210> 94  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 94  
catcatttaa gataagtga gttcacagat ataaagcctg atgtctccct tagcaaggct 60

aatatacaaa tccttttact 80  
<210> 95  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 95  
agtaaaagga tttgtatatt agccttgcta ggctttatat ctagtgatat aagtgtgaac 60  
tacacttatac ttaaatagatg 80  
<210> 96  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 96  
catcatttaa gataagtga gttcacactt atatacactag atataaagcc tagcaaggct 60  
aatatacaaa tccttttact 80  
<210> 97  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 97  
agtaaaagga tttgtatagg ctttatatca agggagacat ctagtgatat aagtgtgaac 60  
tacacttatac ttaaatagatg 80  
<210> 98  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 98  
catcatttaa gataagtga gttcacactt atatacactag atgtctcct tgatataaag 60  
cctatacaaa tccttttact 80  
<210> 99  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 99  
agtaaaagga tttgtatatt agccttgcta agcacattcg atagtgatat aagaggcttt 60  
atatcttatac ttaaatagatg 80  
<210> 100  
<211> 80  
<212> DNA

<213> 人工序列 (Artificial sequence)  
<400> 100  
catcatttaa gataagatat aaagcctctt atatcactat cgaatgtgct tagcaaggct 60  
aatatacaaaa tccttttact 80  
<210> 101  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 101  
agtaaaagga tttgtatatt agccttgcta agcacattcg aaggctttat atctgtgaac 60  
tacacttatac ttaaagtatg 80  
<210> 102  
<211> 80  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 102  
catcatttaa gataagtga gttcacagat ataaagcctt cgaatgtgct tagcaaggct 60  
aatatacaaaa tccttttact 80  
<210> 103  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 103  
cgagactctc caggtttggg aa 22  
<210> 104  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 104  
ggttctcgaa tcaagtcggg tt 22  
<210> 105  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 105  
aaccgacttg attcgagaac ct 22  
<210> 106  
<211> 19  
<212> DNA

<213> 人工序列 (Artificial sequence)  
<400> 106  
gttgggcgcc aattgtcat 19  
<210> 107  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 107  
actaaggctc cttctactag cg 22  
<210> 108  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 108  
gttgaagaag tgaatgctgc ac 22  
<210> 109  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 109  
ttataccatc catccacccc tc 22  
<210> 110  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 110  
gtttatgcca acgaggtctc aa 22  
<210> 111  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 111  
gcatgaggca cttcagaata ca 22  
<210> 112  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 112  
aggccaatac tcttgagcta gt 22

<210> 113  
<211> 19  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 113  
cccatcttgc tgacctcac 19  
<210> 114  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 114  
agacctggga cgtttctgtg 20  
<210> 115  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 115  
aaacctacag gtgggtctt 20  
<210> 116  
<211> 14  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 116  
ccccacctgt aggt 14  
<210> 117  
<211> 20  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 117  
aaacctacag gtgggtctt 20  
<210> 118  
<211> 14  
<212> DNA  
<213> 人工序列(Artificial sequence)  
<400> 118  
ccccacctgt aggt 14  
<210> 119  
<211> 14  
<212> DNA

<213> 人工序列 (Artificial sequence)  
<400> 119  
ttccccctgg cctt 14  
<210> 120  
<211> 14  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 120  
ctcctcctcc aggt 14  
<210> 121  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 121  
tccccaggag ctgttgatcc 20  
<210> 122  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 122  
atatcactag atgtctccct 20  
<210> 123  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 123  
atatcactag atgtctccct 20  
<210> 124  
<211> 19  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 124  
atatcactag atgtctccc 19  
<210> 125  
<211> 14  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 125  
agacatctag tgat 14

<210> 126  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 126  
tgggccagga gaggtggct 20  
<210> 127  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 127  
ggtccaggag aggtggcta 20  
<210> 128  
<211> 19  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 128  
ggtccaggag aggtggct 19  
<210> 129  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 129  
ggtccaggag aggtggcta 20  
<210> 130  
<211> 14  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 130  
caccctctcc tgga 14  
<210> 131  
<211> 19  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 131  
tgaacagcag aaggaaca 19  
<210> 132  
<211> 20  
<212> DNA

<213> 人工序列 (Artificial sequence)  
<400> 132  
tgaacagcag aaggaacaaa 20  
<210> 133  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 133  
tgaacagcag aaggaacaaa 20  
<210> 134  
<211> 14  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 134  
ttccttctgc tggt 14  
<210> 135  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 135  
taacgaggag aggggtgtggt 20  
<210> 136  
<211> 14  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 136  
caccctctcc tcgt 14  
<210> 137  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 137  
atatcactag atgtctccct 20  
<210> 138  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 138  
atatcactag atgtctccct 20

<210> 139  
<211> 19  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 139  
atatcactag atgtctccc 19  
<210> 140  
<211> 14  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 140  
agacatctag tgat 14  
<210> 141  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 141  
atatcactag atgtctccct 20  
<210> 142  
<211> 79  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 142  
caagcgtaaa ttgcatagtc gtgaagtcag ggcttctaag gtatctatca ccagattgat 60  
atctattggg atctatcac 79  
<210> 143  
<211> 79  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 143  
gttcgcattt aacgtatcag cacttcagtc ccgaagattc catagatagt ggtctaacta 60  
tagataacca tagatagtg 79  
<210> 144  
<211> 10  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 144  
tatatcacta 10  
<210> 145

<211> 10  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 145  
gatgtctccc 10  
<210> 146  
<211> 22  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 146  
tggccaaatc catagcetta ga 22  
<210> 147  
<211> 20  
<212> DNA  
<213> 人工序列 (Artificial sequence)  
<400> 147  
gtctctcgca gccacatggg 20

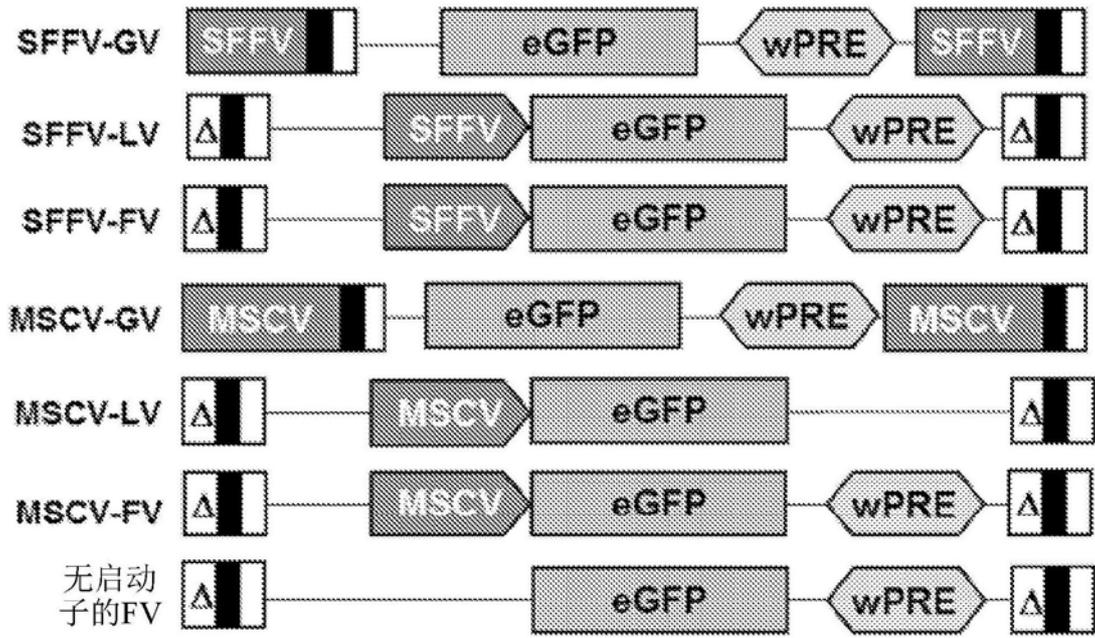


图1A

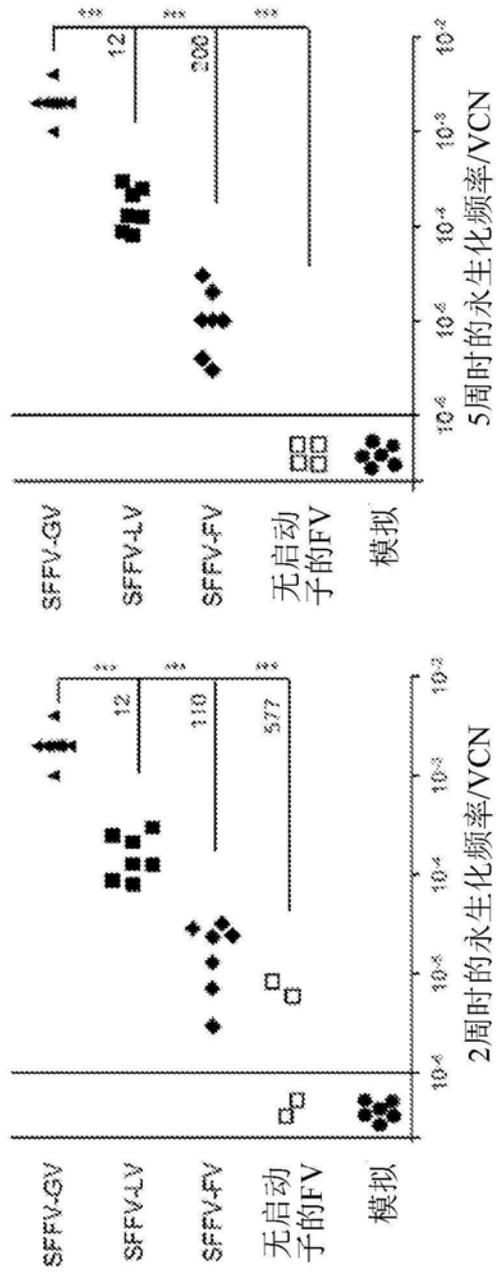


图1B

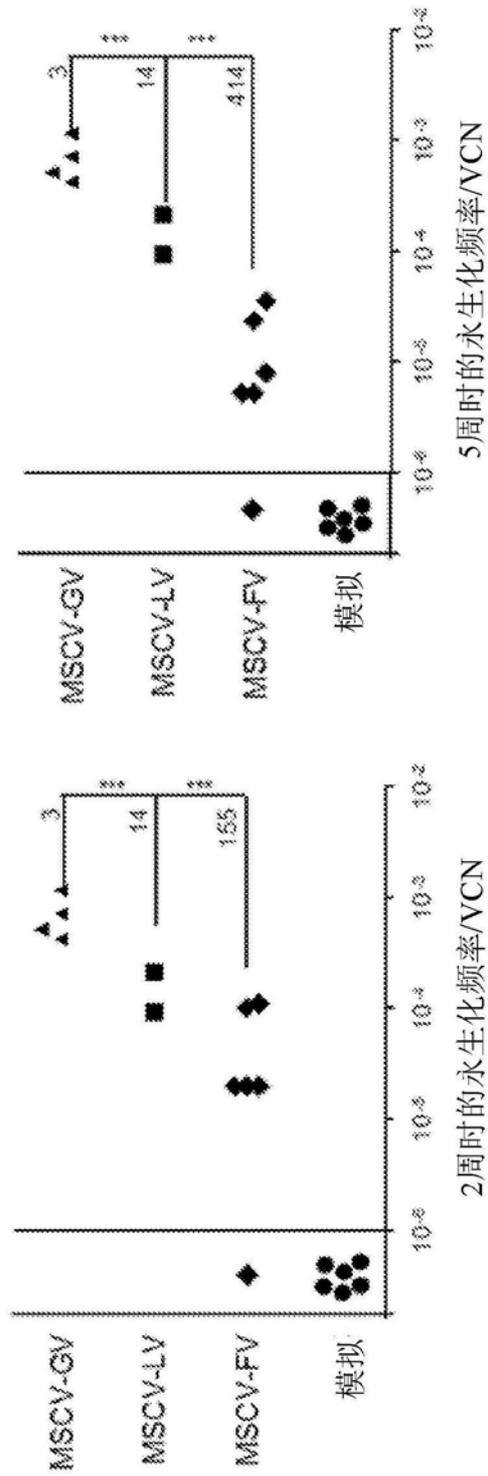


图1C

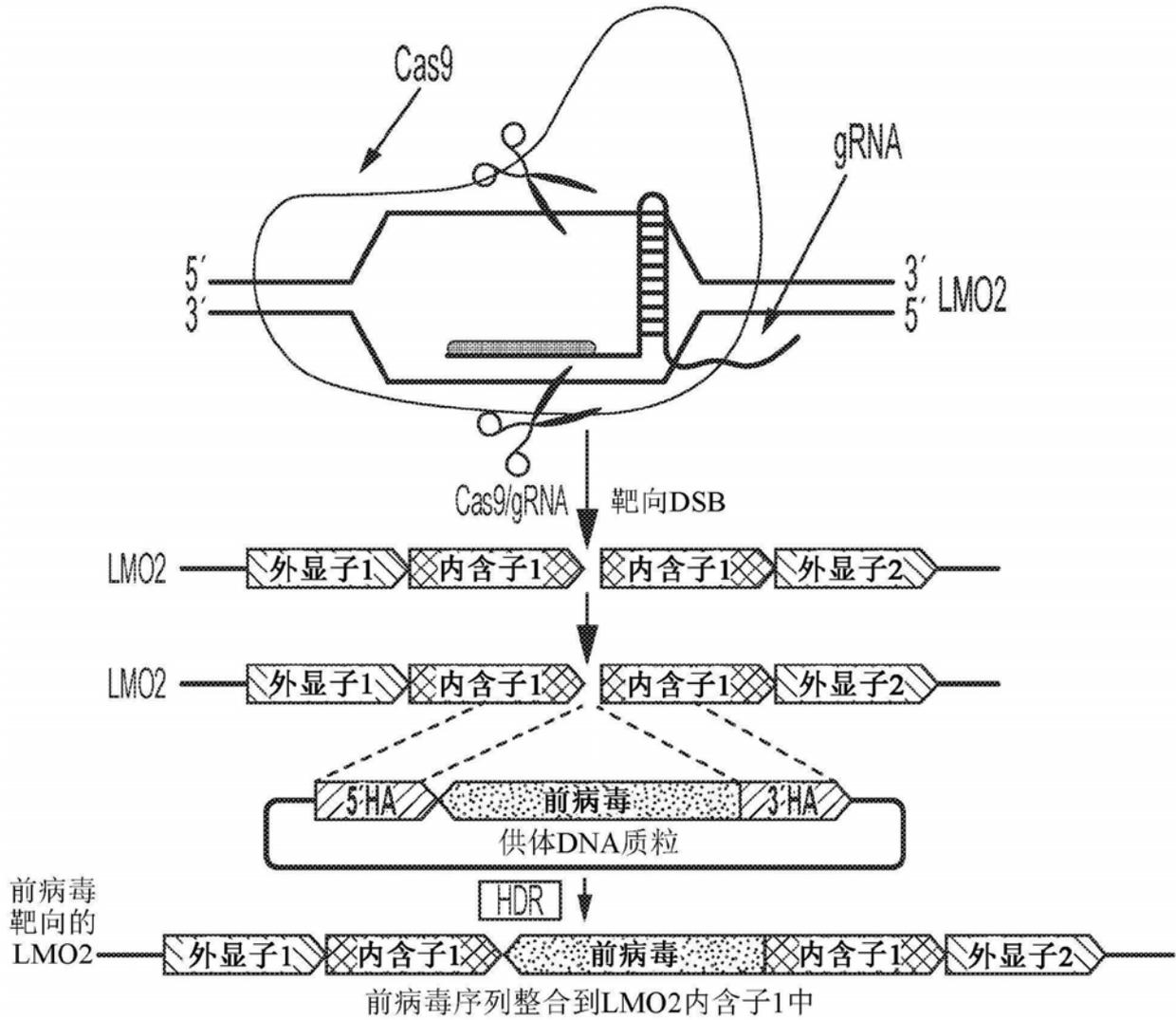


图2A

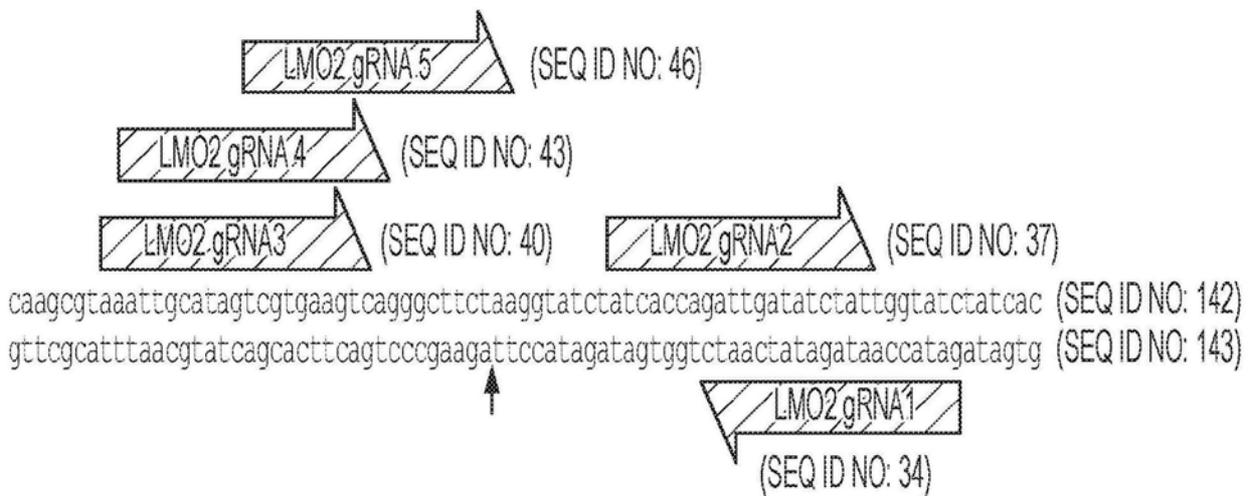


图2B

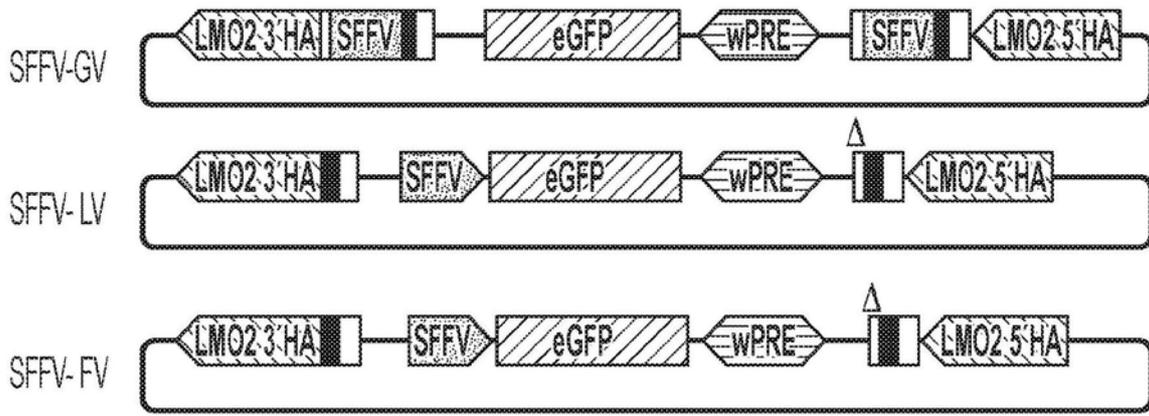


图2C

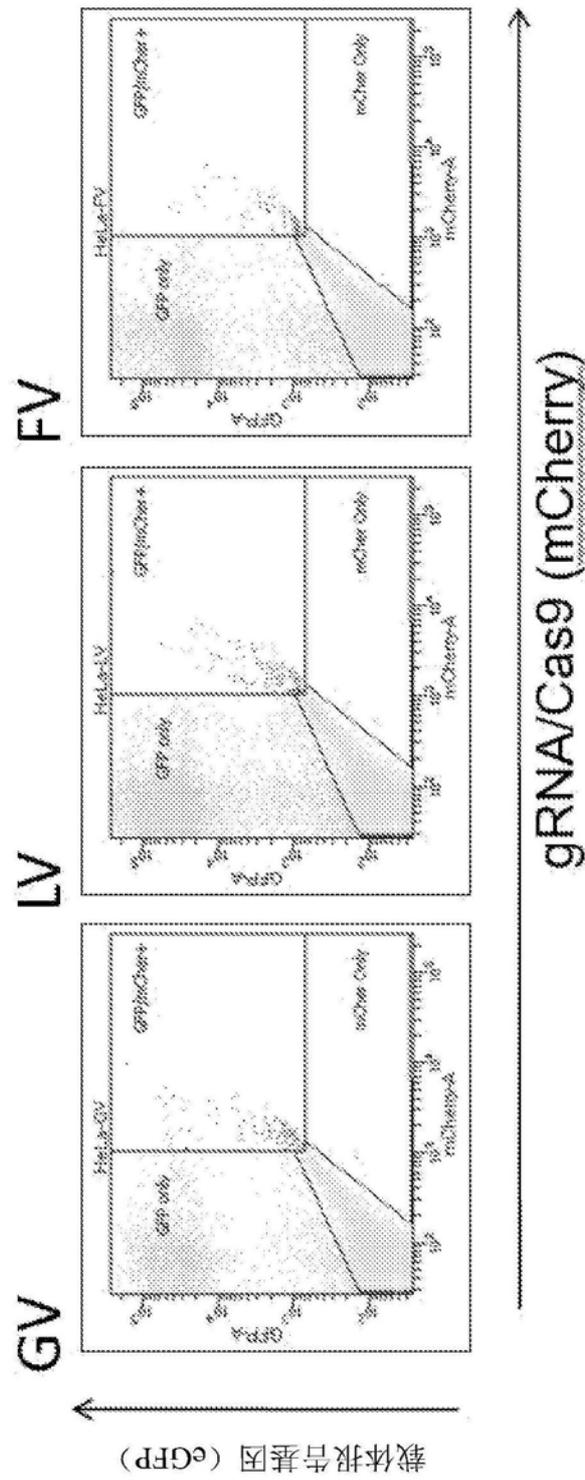


图3A

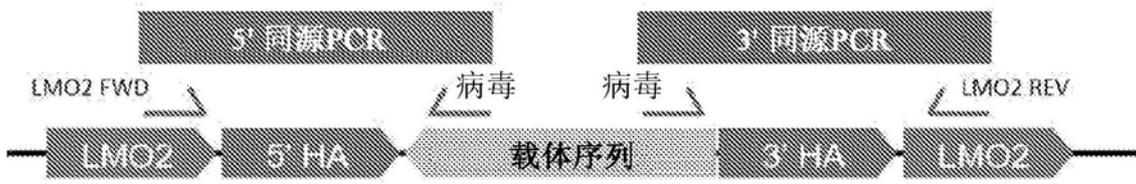


图3B

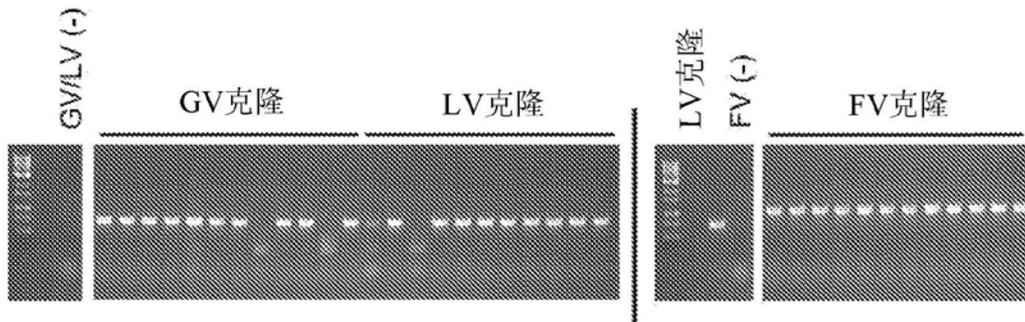


图3C

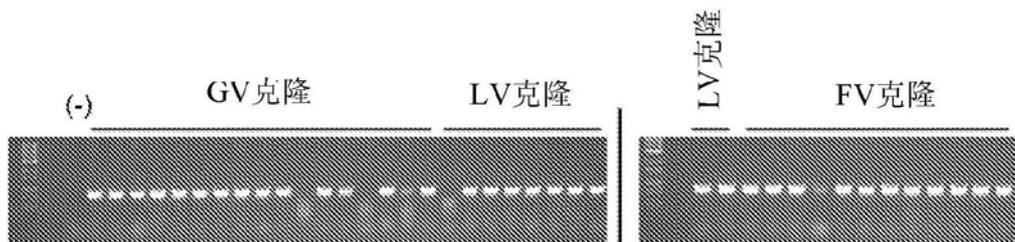


图3D

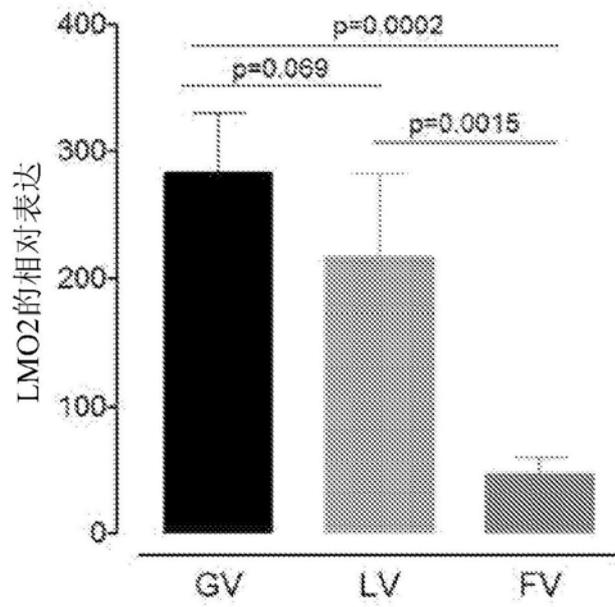


图4A

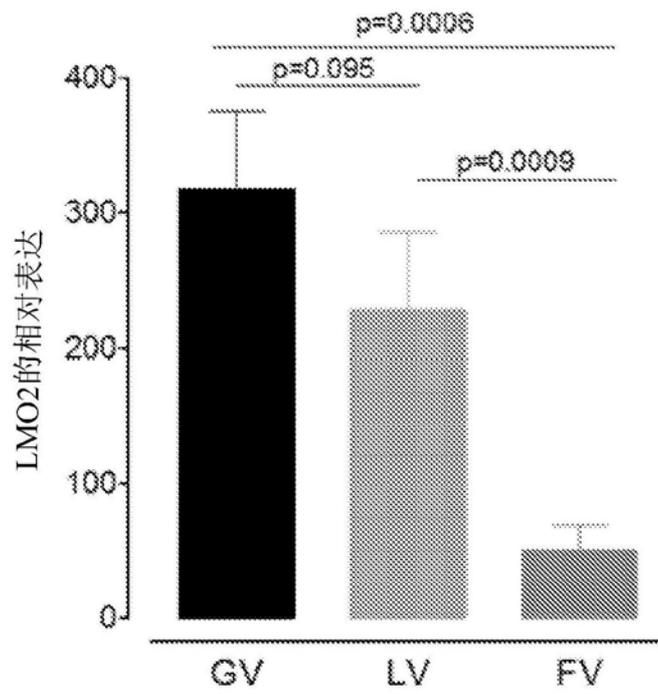


图4B

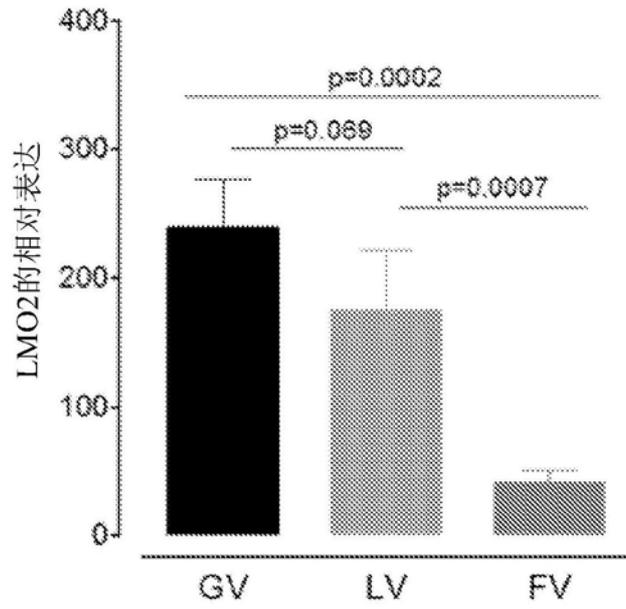


图4C

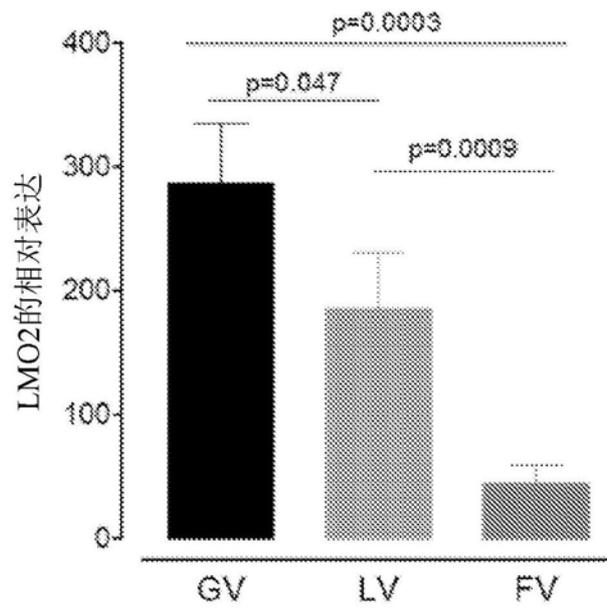


图4D

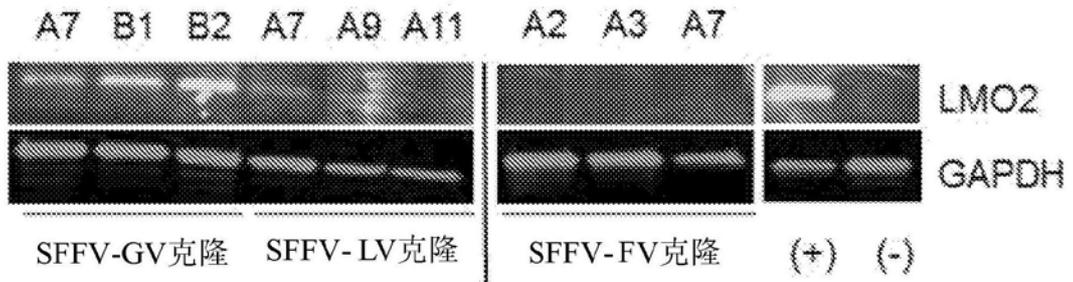


图5

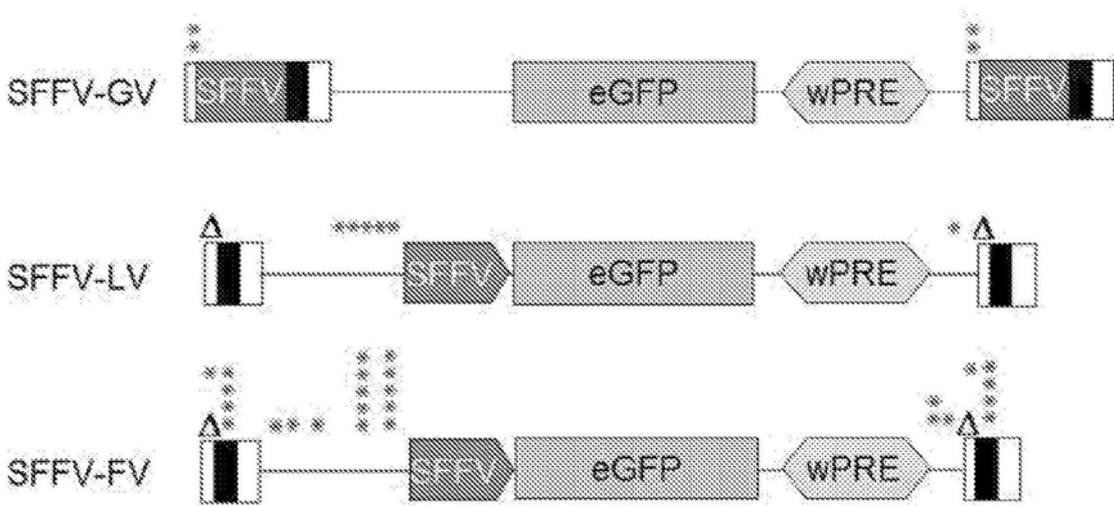


图6A

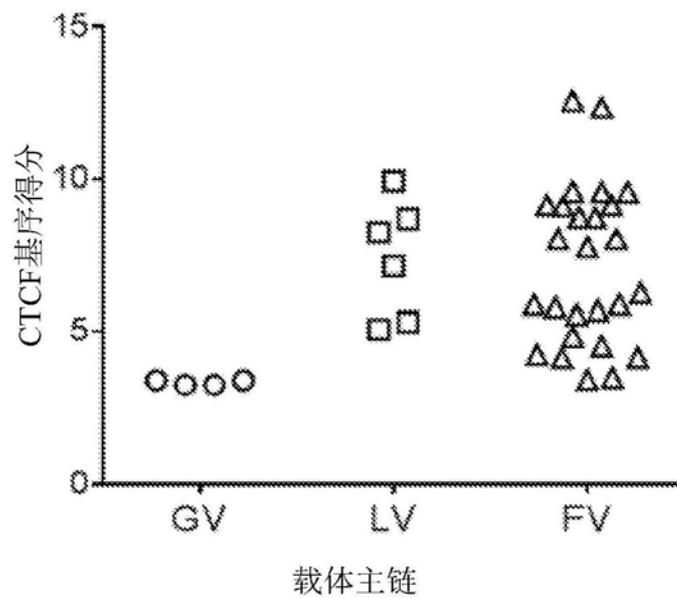


图6B

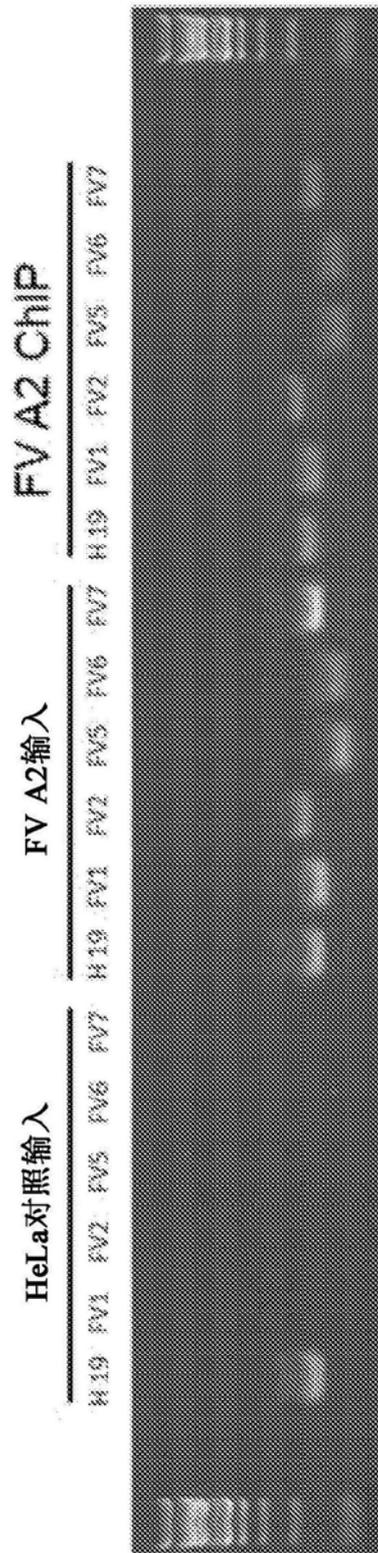


图7A

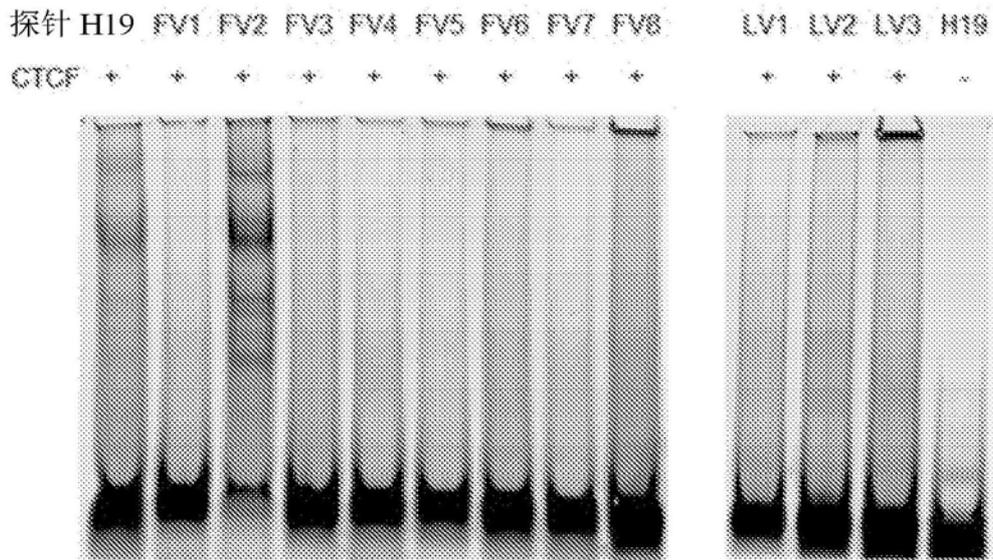


图7B

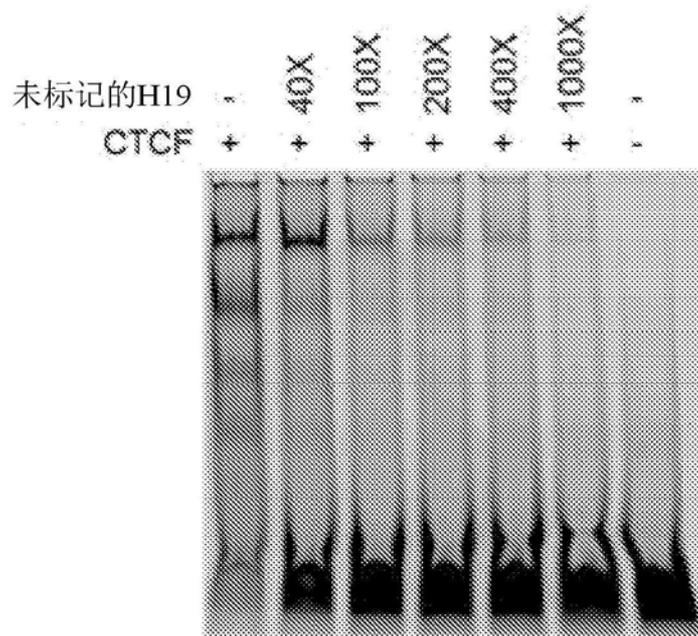


图7C

	突变的区域 (12 bp/区域)				
	1	2	3	4	
CATCATTAAAGATAA	<u>GTCTAGTTCACA</u>	CTTATATCACTA	GATGTCTCCCTT	AGCAAGGCTAAT	ATACAAATCCTTTTACT (SEQ ID NO: 74)
<b>突变体1</b>					
CATCATTAAAGATAA	<u>GATATAAAGCCT</u>	CTTATATCACTA	GATGTCTCCCTT	AGCAAGGCTAAT	ATACAAATCCTTTTACT (SEQ ID NO: 92)
<b>突变体2</b>					
CATCATTAAAGATAA	GTGTAGTTCACA	<u>GATATAAAGCCT</u>	GATGTCTCCCTT	AGCAAGGCTAAT	ATACAAATCCTTTTACT (SEQ ID NO: 94)
<b>突变体3</b>					
CATCATTAAAGATAA	GTGTAGTTCACA	CTTATATCACTA	<u>GATATAAAGCCT</u>	AGCAAGGCTAAT	ATACAAATCCTTTTACT (SEQ ID NO: 96)
<b>突变体4</b>					
CATCATTAAAGATAA	GTGTAGTTCACA	CTTATATCACTA	GATGTCTCCCTT	<u>GATATAAAGCCT</u>	ATACAAATCCTTTTACT (SEQ ID NO: 98)
<b>突变体5</b>					
CATCATTAAAGATAA	<u>GATATAAAGCCT</u>	CTTATATCACTA	<u>TCGAATGTGCTT</u>	AGCAAGGCTAAT	ATACAAATCCTTTTACT (SEQ ID NO: 100)
<b>突变体6</b>					
CATCATTAAAGATAA	GTGTAGTTCACA	<u>GATATAAAGCCT</u>	<u>TCGAATGTGCTT</u>	AGCAAGGCTAAT	ATACAAATCCTTTTACT (SEQ ID NO: 102)

图7D

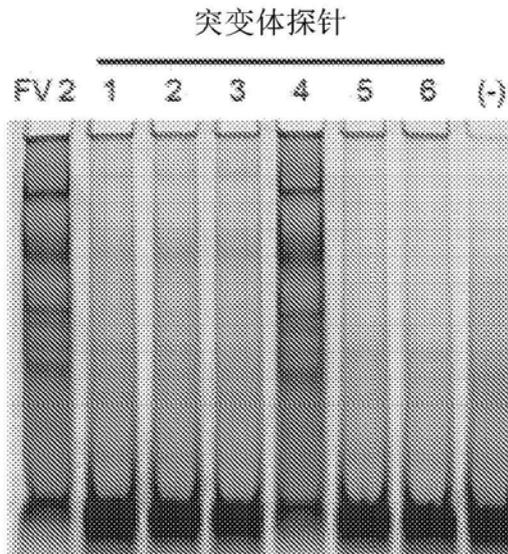


图7E

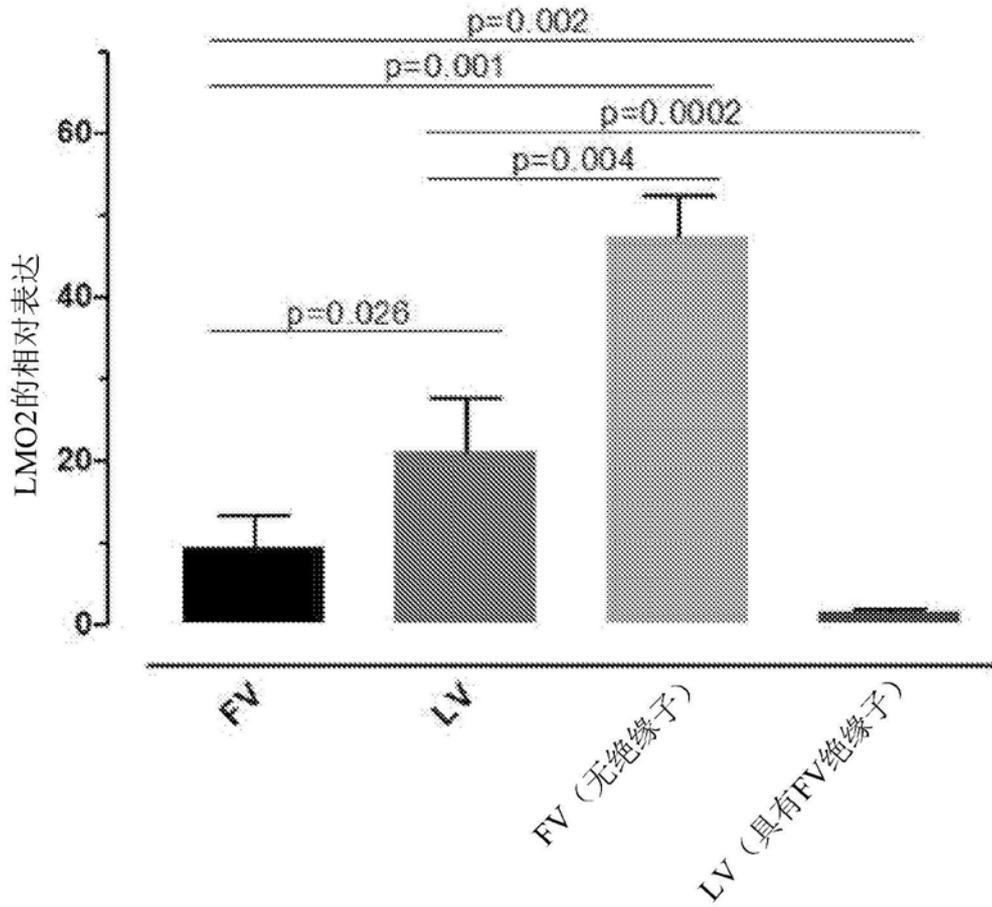


图8

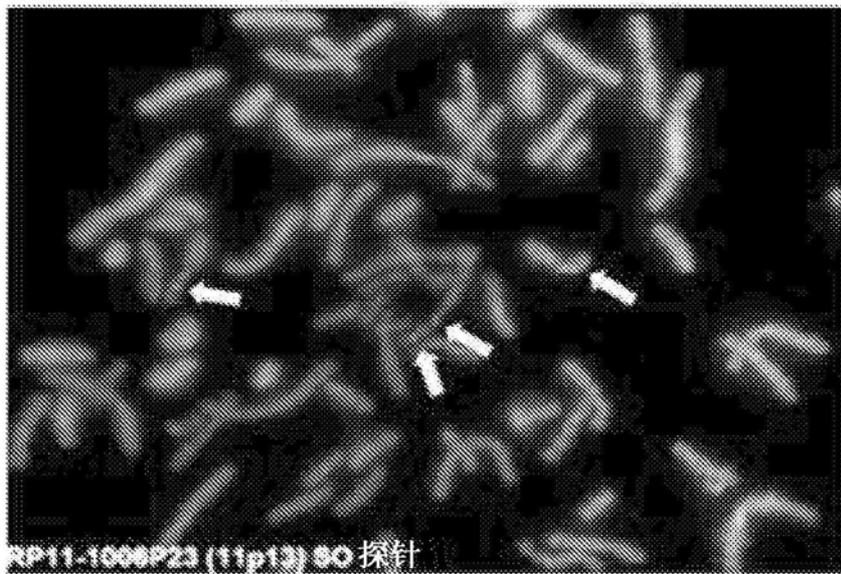


图9A

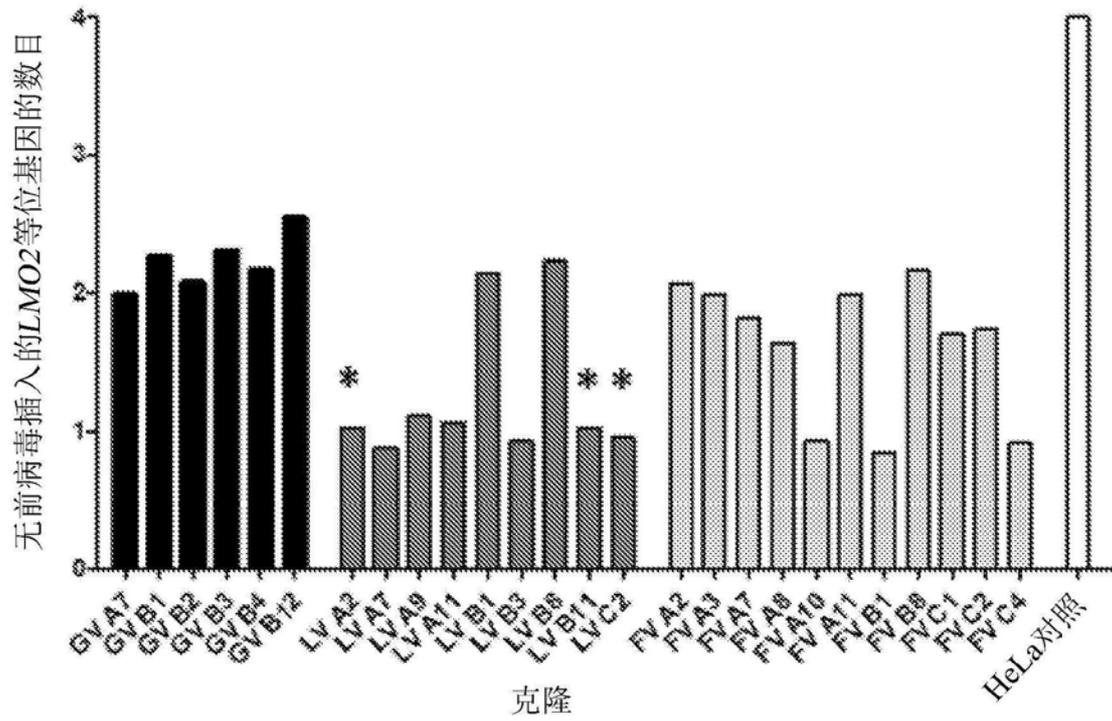


图9B

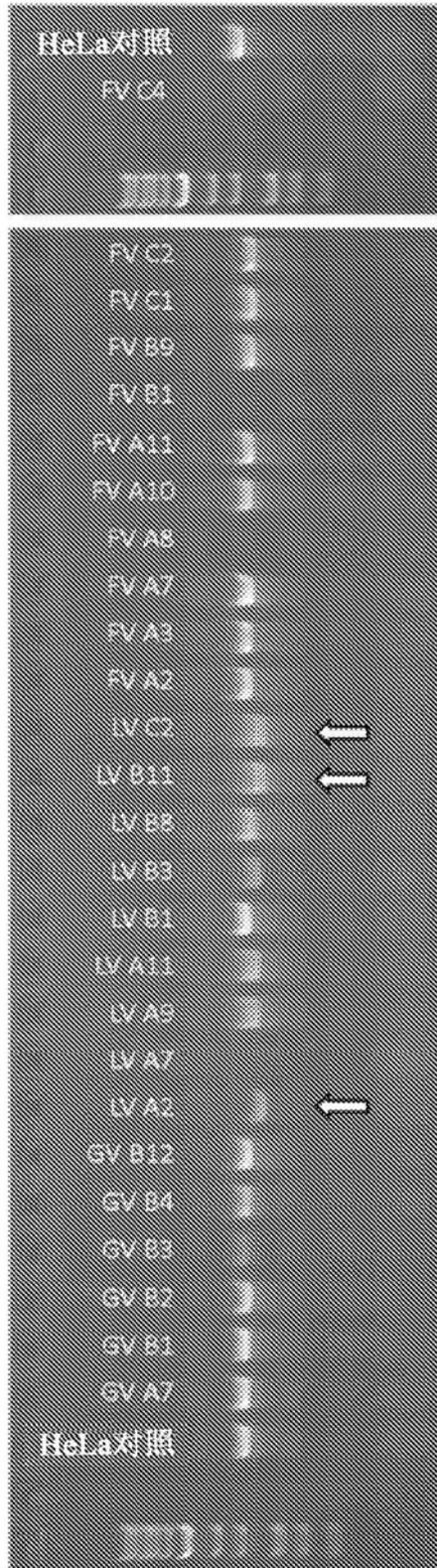


图9C