



(12) 发明专利申请

(10) 申请公布号 CN 113272428 A

(43) 申请公布日 2021.08.17

(21) 申请号 201980078035.8

(22) 申请日 2019.10.18

(30) 优先权数据

62/747,393 2018.10.18 US

62/840,343 2019.04.29 US

(85) PCT国际申请进入国家阶段日
2021.05.27

(86) PCT国际申请的申请数据

PCT/US2019/057084 2019.10.18

(87) PCT国际申请的公布数据

W02020/082041 EN 2020.04.23

(71) 申请人 英特利亚治疗股份有限公司

地址 美国马萨诸塞

申请人 瑞泽恩制药公司

(72) 发明人 J·D·芬恩 H-R·黄

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038

代理人 李程达

(51) Int.Cl.

C12N 15/11 (2006.01)

C12N 15/10 (2006.01)

C12N 15/86 (2006.01)

权利要求书4页 说明书86页 附图27页

(54) 发明名称

核酸构建体和使用方法

(57) 摘要

提供了允许插入和/或表达感兴趣的序列诸如转基因的核酸构建体。还提供了使用此类构建体来表达例如多肽或治疗剂的组合物和方法。

1. 一种双向核酸构建体,所述双向核酸构建体包含:
 - a) 第一区段,所述第一区段包含诸如多肽的剂的编码序列;以及
 - b) 第二区段,所述第二区段包含剂的编码序列的反向补体,其中所述构建体不包含驱动诸如多肽的剂的表达的启动子。
2. 一种双向核酸构建体,所述双向核酸构建体包含:
 - a) 第一区段,所述第一区段包含诸如第一多肽的第一剂的编码序列;以及
 - b) 第二区段,所述第二区段包含诸如第二多肽的第二剂的编码序列的反向补体,其中所述构建体不包含驱动所述一种或多种剂的表达的启动子。
3. 如权利要求1或2所述的双向核酸构建体,其中所述第二区段在所述第一区段的3'。
4. 如权利要求1至3中任一项所述的双向核酸构建体,其中所述第二区段中反向补体的编码序列采用与所述第一区段的第一编码序列不同的密码子使用。
5. 如权利要求1至4中任一项所述的双向核酸构建体,其中所述第二区段包含核苷酸序列,其与所述第一区段中的编码序列具有至少约30%、约35%、约40%、约45%、约50%、约55%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%、约97%或约99%互补性。
6. 如权利要求1至5中任一项所述的双向核酸构建体,其中所述第二区段的编码序列使用由所述第一区段中的编码序列所编码的一个或多个氨基酸的一个或多个替代密码子编码多肽。
7. 如权利要求1至6中任一项所述的双向核酸构建体,其中所述第二区段包含所述第一区段的编码序列或其片段的反向补体。
8. 如权利要求7所述的双向核酸构建体,其中所述反向补体:
 - a. 基本上不与所述第一区段的编码序列互补;
 - b. 基本上不与所述第一区段的编码序列的片段互补;
 - c. 与所述第一区段的编码序列高度互补;
 - d. 与所述第一区段的编码序列的片段高度互补;
 - e. 与所述第一区段的编码序列的反向补体至少60%相同;
 - f. 与所述第一区段的编码序列的反向补体至少70%相同;
 - f. 与所述第一区段的编码序列的反向补体至少90%相同;
 - g. 与所述第一区段的编码序列的反向补体50%至80%相同;且/或
 - h. 与所述第一区段的编码序列的反向补体60%至100%相同。
9. 如权利要求1至8中任一项所述的双向核酸构建体,其中所述构建体不包含同源臂。
10. 如权利要求1至9中任一项所述的双向核酸构建体,其中所述第一区段通过接头连接到所述第二区段。
11. 如权利要求10所述的双向核酸构建体,其中所述接头的长度为5、10、20、30、40、50、60、70、80、90、100、150、200、250、300、500、1000、1500、2000个核苷酸。
12. 如权利要求1至11中任一项所述的双向核酸构建体,其中所述第一区段和所述第二区段中的每个包含聚腺苷酸化序列,诸如聚腺苷酸化信号序列或聚腺苷酸化尾序列。
13. 如权利要求1至12中任一项所述的双向核酸构建体,其中所述构建体包含剪接受体位点。

14. 如权利要求13所述的双向核酸构建体,其中所述构建体包含所述第一区段的第一剪接受体位点5'和所述第二区段的第二剪接受体位点3'。

15. 如权利要求1至14中任一项所述的双向核酸构建体,其中所述构建体为双链的,任选双链DNA。

16. 如权利要求1至15中任一项所述的双向核酸构建体,其中所述构建体为单链的,任选单链DNA。

17. 如权利要求1至16中任一项所述的双向核酸构建体,其中编码所述多肽的序列为密码子优化的。

18. 如权利要求1至17中任一项所述的双向核酸构建体,其中所述构建体包含以下末端结构中的一个或多个:发夹、环、反向末端重复序列(ITR)或螺旋管。

19. 如权利要求1至18中任一项所述的双向核酸构建体,其中所述构建体包含一个、两个或三个反向末端重复序列(ITR)。

20. 如权利要求1至19中任一项所述的双向核酸构建体,其中所述构建体包含不多于两个ITR。

21. 如权利要求1至20中任一项所述的双向核酸构建体,其中所述多肽为分泌型多肽。

22. 如权利要求1至20中任一项所述的双向核酸构建体,其中所述多肽 y 为细胞内多肽。

23. 如权利要求2至22中任一项所述的双向核酸构建体,其中所述第一多肽和所述第二多肽是不同的。

24. 如权利要求1至23中任一项所述的双向核酸构建体,其中所述多肽为变体多肽。

25. 如权利要求1至24中任一项所述的双向核酸构建体,其中所述多肽为肝脏蛋白。

26. 如权利要求1至24中任一项所述的双向核酸构建体,其中所述多肽为非肝脏蛋白。

27. 如权利要求1至26中任一项所述的双向核酸构建体,其中所述构建体为同源非依赖性构建体。

28. 如权利要求1至27中任一项所述的双向核酸构建体,其中所述多肽在表达时包含异源信号肽。

29. 如权利要求1至28中任一项所述的双向核酸构建体,其中所述核酸编码异源信号肽。

30. 如权利要求1至27中任一项所述的双向核酸构建体,其中所述核酸不编码信号肽。

31. 如权利要求1至27中任一项所述的双向核酸构建体,其中所述多肽在表达时包含其自身的信号肽。

32. 如权利要求1至31中任一项所述的双向核酸构建体,其中所述核酸编码异源肽。

33. 如权利要求32所述的双向核酸构建体,其中所述异源肽为2A。

34. 一种载体,所述载体包含如权利要求1至33中任一项所述的构建体。

35. 如权利要求34所述的载体,其中所述载体为腺相关病毒(AAV)载体。

36. 如权利要求34或35所述的载体,其中所述AAV包含单链基因组(ssAAV)或自互补基因组(scAAV)。

37. 如权利要求34或35所述的载体,其中所述AAV载体选自由以下组成的组:AAV1、AAV2、AAV3、AAV3B、AAV4、AAV5、AAV6、AAV6.2、AAV7、AAVrh.64R1、AAVhu.37、AAVrh.8、AAVrh.32.33、AAV8、AAV9、AAV-DJ、AAV2/8、AAVrh10、AAVLK03、AV10、AAV11、AAV12、rh10及其

杂交体。

38. 一种病毒载体, 所述病毒载体包含含有编码多肽的核苷酸序列的自互补(或双链)核酸构建体, 其中所述载体不包含驱动所述多肽的表达的启动子。

39. 如权利要求38所述的载体, 其中所述载体不包含同源臂。

40. 一种脂质纳米颗粒, 所述脂质纳米颗粒包含如权利要求1至33中任一项所述的构建体。

41. 一种宿主细胞, 所述宿主细胞包含如权利要求1至33中任一项所述的构建体。

42. 如权利要求41所述的宿主细胞, 其中所述宿主细胞为肝脏细胞、肌肉细胞或神经元细胞。

43. 如权利要求41所述的宿主细胞, 其中所述宿主细胞为非分裂细胞类型。

44. 如权利要求41至43中任一项所述的宿主细胞, 其中所述宿主细胞表达由所述构建体编码的多肽。

45. 如权利要求41至44中任一项所述的宿主细胞, 其中所述宿主细胞为肝细胞。

46. 一种修饰靶基因座的方法, 所述方法包括向细胞提供如权利要求1至33中任一项所述的构建体、如权利要求34至39中任一项所述的载体或如权利要求40所述的LNP。

47. 一种将构建体引入细胞的方法, 所述方法包括向所述细胞提供根据权利要求1至33中任一项所述的构建体、根据权利要求34至39中任一项所述的载体或如权利要求40所述的LNP。

48. 一种在细胞中表达多肽的方法, 所述方法包括向所述细胞提供如权利要求1至33中任一项所述的构建体、根据权利要求34至39中任一项所述的载体或如权利要求40所述的LNP。

49. 一种增加细胞中多肽表达的方法, 所述方法包括向所述细胞提供如权利要求1至33中任一项所述的构建体、根据权利要求34至39中任一项所述的载体或如权利要求40所述的LNP。

50. 如权利要求46至49中任一项所述的方法, 其中所述构建体、载体或LNP在体内提供。

51. 如权利要求46至49中任一项所述的方法, 其中所述构建体、载体或LNP在体外提供。

52. 一种修饰靶基因座的离体方法, 所述方法包括向细胞提供根据权利要求1至33中任一项所述的构建体、根据权利要求34至39中任一项所述的载体或如权利要求40所述的LNP。

53. 如权利要求46至52中任一项所述的方法, 其中所述细胞为非分裂细胞类型。

54. 如权利要求46至53中任一项所述的方法, 所述方法还包括向所述细胞提供核酸酶, 其中所述核酸酶能够引入位点特异性ssDNA或dsDNA断裂。

55. 如权利要求46至54中任一项所述的方法, 所述方法包括向所述细胞提供RNA指导的DNA结合剂和指导RNA(gRNA)。

56. 如权利要求55所述的方法, 其中所述gRNA为单个RNA(sgRNA)。

57. 如权利要求55或56所述的方法, 其中同时向所述细胞提供所述构建体、RNA指导的DNA结合剂和所述gRNA。

58. 如权利要求55或56所述的方法, 其中以任何顺序依次向所述细胞提供所述构建体、RNA指导的DNA结合剂和所述gRNA。

59. 如权利要求58所述的方法, 其中在提供所述gRNA和RNA指导的DNA结合剂之前将所

述构建体提供给所述细胞。

60. 如权利要求58所述的方法, 其中在提供所述构建体之前, 将所述RNA指导的DNA结合剂或RNA指导的DNA结合剂和gRNA的组合提供给所述细胞。

61. 如权利要求54至60中任一项所述的方法, 其中所述构建体以载体的形式提供。

62. 如权利要求54至60中任一项所述的方法, 其中所述构建体以脂质纳米颗粒的形式提供。

63. 如权利要求54至62中任一项所述的方法, 其中所述RNA指导的DNA结合剂作为编码RNA指导的DNA结合剂的核酸提供。

64. 如权利要求54至62中任一项所述的方法, 其中所述RNA指导的DNA结合剂作为Cas酶或编码RNA指导的DNA结合剂的mRNA提供。

65. 如权利要求54至62中任一项所述的方法, 其中所述RNA指导的DNA结合剂为Cas核酸酶, 诸如Cas9核酸酶。

66. 如权利要求65所述的方法, 其中所述Cas核酸酶为2类Cas核酸酶或其变体。

67. 如权利要求54至66中任一项所述的方法, 其中所述CAS核酸酶为酿脓链球菌Cas9或其变体。

68. 如权利要求54至67中任一项所述的方法, 其中所述Cas核酸酶为切口酶。

69. 如权利要求54至68中任一项所述的方法, 其中所述基因座为白蛋白基因座。

70. 如权利要求54至69中任一项所述的方法, 其中所述细胞为非分裂细胞类型。

71. 如权利要求54至70中任一项所述的方法, 其中所述细胞为肝脏细胞、神经元细胞或肌肉细胞。

72. 一种细胞, 所述细胞通过如权利要求46至71中任一项所述的方法制备。

73. 如权利要求72所述的细胞, 其中所述细胞为宿主细胞。

74. 如权利要求73所述的宿主细胞, 其中所述宿主细胞为肝脏细胞、肌肉细胞或神经元细胞。

75. 如权利要求72至74中任一项所述的细胞, 其中所述宿主细胞为非分裂细胞类型。

76. 如权利要求72至75中任一项所述的细胞, 其中所述宿主细胞表达由所述构建体编码的多肽。

77. 如权利要求72至76中任一项所述的细胞, 其中所述宿主细胞为肝细胞。

78. 如权利要求1至33中任一项所述的双向核酸构建体, 其中所述剂为治疗剂。

79. 如权利要求1至33中任一项所述的双向核酸构建体, 其中所述剂为功能性RNA、mRNA或多肽中的任一种或多种。

80. 如权利要求1至33中任一项所述的双向核酸构建体, 其中所述剂为多肽。

81. 如权利要求1至33中任一项所述的双向核酸构建体, 其中所述剂为治疗性多肽。

核酸构建体和使用方法

[0001] 本申请要求2018年10月18日提交的美国临时申请号62/747,393和2019年4月29日提交的美国临时申请号62/840,343的优先权权益。上述申请中的每个的说明书以引用的方式整体并入本文。

[0002] 基因治疗方法中的基因组编辑源于这样的观点：外源性引入缺失或以其他方式受损的遗传物质可纠正遗传疾病。长期以来，基因治疗因其在从业者如何处理和治疗人疾病方面的巨大潜力而被认可。不是依赖药物或手术，具有潜在遗传因素的患者可通过直接靶向潜在病因进行治疗。此外，通过靶向潜在的遗传病因，基因治疗可能有效治愈患者。但是，现有方法的临床应用仍需要在几个方面进行改进。

[0003] 本公开提供了双向核酸构建体，所述双向核酸构建体允许增强插入和表达例如编码治疗剂（诸如多肽）的感兴趣的核酸序列。如本文所述，双向构建体包含至少两个核酸区段，其中一个区段（第一区段）包含编码感兴趣的剂的编码序列（编码序列在本文中可称为“转基因”或第一转基因），而另一区段（第二区段）包含其中序列的补体编码感兴趣的剂的序列（或第二转基因）。在一些实施方案中，构建体包含至少两个核酸区段，其中一个区段（第一区段）包含编码感兴趣的多肽的编码序列，而另一区段（第二区段）包含其中序列的补体编码感兴趣的多肽的序列。当与基因编辑系统组合使用时，核酸构建体的双向性允许构建体在靶插入位点内的任一方向上插入（不限于在一个方向上插入），从而允许从以下中的任一个表达感兴趣的多肽：a) 一个区段的编码序列，或2) 另一（第二）区段的补体，从而增强插入和表达效率，如本文所例示。

附图说明

[0004] 图1示出了如AAV基因组中所表示的构建体形式。SA=剪接受体；pA=polyA信号序列；HA=同源臂；LHA=左同源臂；RHA=右同源臂。

[0005] 图2示出了没有同源臂的载体在永生化肝脏细胞系（Hepa1-6）中无效。衍生自包含200bp同源臂的质粒P00204的scAAV导致此细胞系中可检测到的hFIX表达。使用衍生自P00123（缺乏同源臂的scAAV）和P00147（缺乏同源臂的ssAAV双向构建体）的AAV载体不会导致可检测到的hFIX表达。

[0006] 图3A和图3B示出了使用衍生自P00123、P00147或P00204的载体对具有和不具有同源臂的插入模板进行体内测试的结果。图3A示出了在用包含CRISPR/Cas9系统组分的LNP治疗的每组动物中检测到如通过插入缺失形成测量为约60%的肝脏编辑水平。图3B示出了接受不具有同源臂的ssAAV载体（衍生自P00147）与LNP治疗组合的动物，导致血清中hFIX表达的最高水平。

[0007] 图4A和图4B示出了对具有和不具有同源臂的ssAAV插入模板进行体内测试的结果。图4A将使用衍生自质粒P00350、P00356、P00362（具有如所示的不对称同源臂）和P00147（如图4B所示的双向构建体）的载体的靶向插入进行比较。图4B将使用衍生自质粒P00353、P00354（具有如所示的对称同源臂）和P00147的载体来靶向的第二位点中的插入进行比较。

[0008] 图5A至图5D示出了在原始小鼠肝细胞中跨20个靶位点通过三个双向构建体进行

靶向插入的结果。图5A示出了所测试的载体中的每个的示意图。图5B示出了针对所测试的每个组合中每个治疗组如通过插入缺失形成来测量的编辑。图5C和图5D示出了显著水平的编辑(在特定靶位点处)不一定导致转基因的更有效插入或表达。在此项靶向插入研究中,所测试的构建体有效导致转基因表达。hSA=人F9剪接受体;mSA=小鼠白蛋白剪接受体;HiBit=用于基于荧光素酶的检测的标签;pA=polyA信号序列;Nluc=纳米荧光素酶报告基因;GFP=绿色荧光报告基因。

[0009] 图6示出了使用衍生自P00147的ssAAV跨10个靶位点靶向插入双向构建体的体内筛选的结果。如图所示,显著水平的编辑不一定导致高水平的转基因表达。

[0010] 图7A至图7D示出了使用衍生自P00147的ssAAV跨20个靶位点对双向构建体进行体内筛选的结果。图7A示出了针对所测试的每个LNP/载体组合的每个治疗组检测到的编辑。图7B提供了相应的靶向插入数据。结果显示编辑与双向构建体的插入/表达之间较差的相关性(图7B和图7D),以及体外和体内结果之间的正相关性(图7C)。

[0011] 图8A和图8B示出了使用原位杂交方法,双向构建体在细胞水平上的插入,所述原位杂交方法使用可检测hFIX转基因与小鼠白蛋白外显子1序列之间的连接的探针(图8A)。循环hFIX水平与杂交转录物呈阳性的细胞数相关(图8B)。

[0012] 图9a示出了hFIX在体内表达的持久性。图9b证明了来自白蛋白内含子1的表达得以持续。

[0013] 图10A至图10B示出了变化的AAV或LNP剂量可在体内调节来自白蛋白基因内含子1的hFIX的表达量。

[0014] 图11A至图11C示出了跨原代食蟹猴肝细胞中的靶位点筛选双向构建体的结果。图11A示出了针对每个样品检测到的如通过插入缺失形成测量的不同水平的编辑。图11B和图11C示出了显著水平的插入缺失形成不预测双向构建体插入或表达白蛋白内含子1中。

[0015] 图12A至图12C示出了跨原代人肝细胞中的靶位点筛选双向构建体的结果。图12A示出了针对每个样品检测到的如通过插入缺失形成测量的编辑。图12B、图12C和图12D示出了显著水平的插入缺失形成不预测双向构建体插入或表达白蛋白基因内含子1中。

[0016] 图13示出了体内研究的结果,其中向非人灵长类动物给药LNP以及双向hFIX插入模板(衍生自P00147)。仅在用LNP和AAV治疗的动物中实现了全身性hFIX水平,其中单独使用AAV或LNP检测不到hFIX。

具体实施方式

[0017] 现在将详细参考本发明的某些实施方案,附图中示出了所述实施方案的实例。虽然结合各种实施方案来描述本教导,但并不旨在将本教导限制于那些实施方案。相反,如本领域技术人员将理解的,本教导涵盖各种替代方案、修改和等效物。

[0018] 在详细描述本教导之前,应当理解,本公开不限于特定的组合物或处理步骤,因为它们可变化。应当指出的是,如本说明书和所附权利要求中所用,除非上下文另有规定,否则单数形式“一种/个(a/an)”和“所述(the)”包括复数指示物。因此,例如,对“缀合物”的引用包括多个缀合物,并且对“细胞”的引用包括多个细胞等。如本文所用,术语“包括”及其语法变体旨在是非限制性的,使得列表中项目的列举不排除可被替换或添加到所列项目的其他类似项目。

[0019] 数值范围将限定范围的数量包括在内。考虑到有效数字和与测量相关的误差,将测量值和可测量值理解为近似值。此外,“包含(comprise/comprises/comprising)”、“含有(contain/contains/containing)”和“包括(include/includes和including)”的使用并不旨在是限制性的。应当理解,上述一般描述和详细描述均仅为示例性和解释性的并且不限制本教导。

[0020] 除非在说明书中特别指出,否则说明书中列举“包含”各种组件的实施方案也被设想为“由所列举的组件组成”或“基本上由所列举的组件组成”;说明书中列举“由各种组件组成”的实施方案也被设想为“包含”所列举的组件或“基本上由所列举的组件组成”;并且说明书中列举“基本上由各种组件组成”的实施方案也被设想为“由所列举的组件组成”或“包含”所列举的组件(这种可互换性不适用于权利要求中这些术语的使用)。除非上下文另外明确指出,否则术语“或”以包括性含义使用,即等同于“和/或”。术语“约”在列表之前使用时,修饰列表的每个成员。术语“约”或“大约”意指如通过本领域普通技术人员确定的特定值的可接受误差,这部分取决于如何测量或确定所述值。

[0021] 本文所使用的小节标题仅出于组织性目的并且不解释为以任何方式限制所需主题。如果通过引用并入的任何材料与本说明书中定义的任何术语或本说明书的任何其他明示内容矛盾,则以本说明书为准。

[0022] I. 定义

[0023] 除非另有说明,否则如本文所用的以下术语和短语旨在具有以下含义:

[0024] 如本文所用,“多核苷酸”和“核酸”是指包含核苷或核苷类似物的多聚化合物,所述核苷或核苷类似物具有沿着骨架连接在一起的含氮杂环碱基或碱基类似物,包括常规RNA、DNA、混合RNA-DNA和作为其类似物的聚合物。核酸“骨架”可由多种键联构成,包括以下中的一个或多个:糖-磷酸二酯键联、肽-核酸键(“肽核酸”或PNA;PCT号W0 95/32305)、硫代磷酸酯键联、甲基磷酸酯键联或其组合。核酸的糖部分可为核糖、脱氧核糖或具有任选的取代基(例如,2'甲氧基或2'卤化物取代基)的类似化合物。含氮碱基可为常规碱基(A、G、C、T、U)、其类似物(例如,修饰的尿苷,诸如5-甲氧基尿苷、假尿苷或N1-甲基假尿苷或其他);肌苷;嘌呤或嘧啶的衍生物(例如,N⁴-甲基脱氧鸟苷、脱氮嘌呤或氮杂嘌呤、脱氮嘧啶或氮杂嘧啶、在5或6位具有取代基的嘧啶碱基(例如,5-甲基胞嘧啶)、在2、6或8位具有取代基的嘌呤碱基、2-氨基-6-甲基氨基嘌呤、0⁶-甲基鸟嘌呤、4-硫代-嘧啶、4-氨基-嘧啶、4-二甲基胍-嘧啶和0⁴-烷基-嘧啶;美国专利号5,378,825和PCT号W0 93/13121)。一般性讨论参见The Biochemistry of the Nucleic Acids 5-36,Adams等人,编,第11版,1992)。核酸可包括一个或多个“脱碱基”残基,其中骨架在聚合物的一个或多个位置不包括含氮碱基(美国专利号5,585,481)。核酸可仅包含常规RNA或DNA糖、碱基和键联,或可包括常规组分和取代基(例如,具有2'甲氧基取代基的常规核苷,或含有常规核苷酸和一个或多个核苷酸类似物的聚合物)。核酸包括含有一种或多种LNA核苷酸单体的类似物“锁核酸”(LNA),其中双环呋喃糖单元被锁定在模仿糖构象的RNA中,这增强了对互补RNA和DNA序列的杂交亲和力(Vester和Wengel,2004,Biochemistry 43(42):13233-41)。RNA和DNA具有不同的糖部分,并且可因RNA中存在尿嘧啶或其类似物且DNA中存在胸腺嘧啶或其类似物而不同。

[0025] “指导RNA”、“gRNA”和简单的“指导”在本文中可互换地用于指包含指导序列的指导,例如,crRNA(也称为CRISPR RNA)或者crRNA和trRNA的组合(也称为tracrRNA)。crRNA和

trRNA可作为单个RNA分子(单指导RNA,sgRNA)缔合或例如在两个单独的RNA分子(双指导RNA,dgRNA)中。“指导RNA”或“gRNA”是指每种类型。trRNA可为天然存在的序列,或可为与天然存在的序列相比具有修饰或变异的trRNA序列。指导RNA,诸如sgRNA或dgRNA,可包括如本文所述的修饰的RNA。

[0026] 如本文所用,“指导序列”是指在指导RNA内与靶序列互补并起到将指导RNA引导到靶序列以通过RNA指导的DNA结合剂结合或修饰(例如,裂解)的作用的序列。“指导序列”也可称为“靶向序列”或“间隔区序列”。指导序列的长度可为20个碱基对,例如在酿脓链球菌(即,Spy Cas9)和相关Cas9同源物/直向同源物的情况下。更短或更长的序列也可用作指导,例如,长度为15、16、17、18、19、21、22、23、24或25个核苷酸。在一些实施方案中,靶序列例如位于基因中或染色体上,并且与指导序列互补。在一些实施方案中,指导序列与其相应的靶序列之间的互补性或同一性程度可为至少约75%、80%、85%、90%、95%、96%、97%、98%、99%或100%。在一些实施方案中,指导序列和靶区域可为100%互补或相同的。在其他实施方案中,指导序列和靶区域可含有至少一个错配。例如,指导序列和靶序列可含有1、2、3或4个错配,其中靶序列的总长度为至少17、18、19、20或更多个碱基对。在一些实施方案中,指导序列和靶区域可含有1至4个错配,其中指导序列包含至少17、18、19、20或更多个核苷酸。在一些实施方案中,指导序列和靶区域可含有1、2、3或4个错配,其中指导序列包含20个核苷酸。

[0027] RNA指导的DNA结合剂的靶序列包括基因组DNA的正链和负链(即给定的序列和序列的反向补体),因为RNA指导的DNA结合剂的核酸底物为双链核酸。因此,在指导序列被称为“与靶序列互补”的情况下,应当理解,指导序列可引导指导RNA与靶序列的正义链和反义链(例如反向补体)结合。因此,在一些实施方案中,在指导序列结合靶序列的反向补体的情况下,指导序列与靶序列(例如,不包括PAM的靶序列)的某些核苷酸相同,除了在指导序列中用U取代T。

[0028] 如本文所用,“RNA指导的DNA结合剂”意指具有RNA和DNA结合活性的多肽或多肽复合物,或此种复合物的DNA结合亚基,其中DNA结合活性是序列特异性的并且取决于RNA的序列。术语RNA指导的DNA结合剂还包括编码此类多肽的核酸。示例性RNA指导的DNA结合剂包括Cas裂解酶/切口酶。示例性RNA指导的DNA结合剂可包括其失活形式(“dCas DNA结合剂”),例如如果那些剂被修饰以例如通过与FokI裂解酶结构域融合来允许DNA裂解。如本文所用,“Cas核酸酶”涵盖Cas裂解酶和Cas切口酶。Cas裂解酶和Cas切口酶包括III型CRISPR系统的Csm或Cmr复合物、其Cas10、Csm1或Cmr2亚基、I型CRISPR系统的Cascade复合物、其Cas3亚基和2类Cas核酸酶。如本文所用,“2类Cas核酸酶”为具有RNA指导的DNA结合活性的单链多肽。2类Cas核酸酶包括2类Cas裂解酶/切口酶(例如,H840A、D10A或N863A变体),其还具有RNA指导的DNA裂解酶或切口酶活性,以及2类dCas DNA结合剂,其中裂解酶/切口酶活性失活),如果那些剂被修饰以允许DNA裂解。2类Cas核酸酶包括例如Cas9、Cpf1、C2c1、C2c2、C2c3、HF Cas9(例如,N497A、R661A、Q695A、Q926A变体)、HypaCas9(例如,N692A、M694A、Q695A、H698A变体)、eSPCas9(1.0)(例如,K810A、K1003A、R1060A变体)和eSPCas9(1.1)(例如,K848A、K1003A、R1060A变体)蛋白及其修饰。Cpf1蛋白,Zetsche等人,Cell,163:1-13(2015),也含有RuvC样核酸酶结构域。Zetsche的Cpf1序列以引用的方式整体并入。参见例如,Zetsche,表S1和表S3。参见例如,Makarova等人,Nat Rev Microbiol,13

(11):722-36 (2015);Shmakov等人,Molecular Cell,60:385-397 (2015)。如本文所用,RNA 指导的DNA结合剂(例如,Cas核酸酶、Cas9核酸酶或酿脓链球菌Cas9核酸酶)的递送包括多肽或mRNA的递送。

[0029] 如本文所用,“核糖核蛋白”(RNP)或“RNP复合物”是指指导RNA与RNA指导的DNA结合剂,诸如Cas核酸酶,例如Cas裂解酶、Cas切口酶、Cas9裂解酶或Cas9切口酶。在一些实施方案中,指导RNA将RNA指导的DNA结合剂诸如Cas9指导到靶序列,并且指导RNA与靶序列杂交且所述剂与靶序列结合;并且结合之后可以裂解或切口。

[0030] 如本文所用,如果第一序列与第二序列的比对显示第二序列整体上X%或更多的位置与第一序列匹配,则认为第一序列“包含与第二序列具有至少X%同一性的序列”。例如,序列AAGA包含与序列AAG具有100%同一性的序列,因为比对将给出100%同一性,因为存在与第二序列的所有三个位置的匹配。RNA与DNA之间的差异(通常为尿苷与胸苷的交换,或反之亦然)以及核苷类似物(诸如修饰的尿苷)的存在不会导致多核苷酸之间的同一性或互补性差异,只要相关核苷酸(诸如胸苷、尿苷或修饰的尿苷)具有相同的补体(例如,对于所有的胸苷、尿苷或修饰的尿苷为腺苷;另一个实例为胞嘧啶和5-甲基胞嘧啶,两者均具有鸟苷或修饰的鸟苷作为补体)。因此,例如,序列5'-AXG(其中X为任何修饰的尿苷诸如假尿苷、N1-甲基假尿苷或5-甲氧基尿苷)被认为与AUG 100%相同,因为两者与相同的序列(5'-CAU)完全互补。示例性比对算法是本领域众所周知的Smith-Waterman算法和Needleman-Wunsch算法。本领域技术人员将理解哪种算法和参数设置的选择适合于给定的一对待比对的序列;对于长度一般相似且氨基酸的预期同一性>50%或核苷酸的预期同一性>75%的序列,使用EBI在www.ebi.ac.uk网页服务器上提供的Needleman-Wunsch算法界面的默认设置的Needleman-Wunsch算法通常是适当的。

[0031] 如本文所用,如果第一序列X%的碱基与第二序列进行碱基配对,则第一序列被认为与第二序列“X%互补”。例如,第一序列5' AAGA3' 与第二序列3' TTCT5' 100%互补,并且第二序列与第一序列100%互补。在一些实施方案中,第一序列5' AAGA3' 与第二序列3' TTCTGTGA5' 100%互补,而第二序列与第一序列50%互补。

[0032] “mRNA”在本文中用于指多核苷酸,其全部或主要是RNA或修饰的RNA,并且包含可翻译成多肽的开放阅读框(即,可充当由核糖体和氨基酰化tRNA翻译的底物)。mRNA可包含包括核糖残基或其类似物(例如2'-甲氧基核糖残基)的磷酸盐-糖骨架。在一些实施方案中,mRNA磷酸盐-糖骨架的糖基本上由核糖残基、2'-甲氧基核糖残基或其组合组成。mRNA的碱基可为修饰的碱基,诸如假尿苷、N-1-甲基-假尿苷或其他天然存在或非天然存在的碱基。

[0033] 可用于本文所述的组合物和方法的示例性指导序列在表1和整个申请中示出。

[0034] 如本文所用,“插入缺失”是指由在靶核酸中的双链断裂(DSB)位点处插入或缺失的多个核苷酸组成的插入/缺失突变。

[0035] 如本文所用,“多肽”是指野生型或变体蛋白(例如,突变体、片段、融合体或其组合)。变体多肽可具有野生型多肽的至少或约5%、10%、15%、20%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%或100%的功能活性。在一些实施方案中,变体与野生型多肽的序列至少70%、75%、80%、85%、90%、92%、93%、94%、95%、96%、97%、98%或99%相同。在一些实施方案中,变体多肽可为活动过度的变体。在

某些情况下,变体具有野生型多肽的约80%至约120%、140%、160%、180%、200%、300%、400%、500%或更多的功能活性。

[0036] 如本文所用,“异源基因”是指已作为外部来源引入宿主细胞基因组内的位点(例如,在基因组基因座处,诸如包括白蛋白内含子1位点的安全港基因座处)的基因。也就是说,引入的基因相对于其插入位点是异源的。由此种异源基因表达的多肽称为“异源多肽”。异源基因可为天然存在或工程改造的,并且可为野生型或变体。异源基因可包括编码异源多肽的序列以外的核苷酸序列(例如,内部核糖体进入位点)。异源基因可为在宿主基因组中作为野生型或变体(例如,突变体)天然存在的基因。例如,尽管宿主细胞含有感兴趣的基因(作为野生型或作为变体),但是可将相同的基因或其变体作为外部来源引入,以用于例如在高度表达的基因座处表达。异源基因也可为在宿主基因组中非天然存在的基因,或表达在宿主基因组中非天然存在的异源多肽的基因。“异源基因”、“外源基因”和“转基因”可互换使用。在一些实施方案中,异源基因或转基因包括外源核酸序列,例如对于受体细胞不是内源的核酸序列。在某些实施方案中,异源基因在受体细胞中非天然存在。例如,异源基因相对于其插入位点和相对于其受体细胞可为异源的。

[0037] 如本文所用,“靶序列”是指靶基因中与gRNA的指导序列具有互补性的核酸序列。靶序列和指导序列的相互作用引导RNA指导的DNA结合剂在靶序列内结合,并潜在地切口或裂解(取决于剂的活性)。

[0038] 如本文所用,“双向核酸构建体”(在本文中可互换地称为“双向构建体”)包含至少两个核酸区段,其中一个区段(第一区段)包含编码感兴趣的剂的编码序列(编码序列在本文中可称为“转基因”或第一转基因),而另一区段(第二区段)包含其中序列的补体编码感兴趣的剂的序列(或第二转基因)。剂可为治疗剂,诸如多肽、功能性RNA、mRNA等。转基因可编码诸如多肽、功能性RNA或mRNA的剂。在一些实施方案中,双向核酸构建体包含至少两个核酸区段,其中一个区段(第一区段)包含编码感兴趣的多肽的编码序列,而另一区段(第二区段)包含其中序列的补体编码感兴趣的多肽的序列(或第二转基因)。也就是说,至少两个区段可编码相同或不同的多肽或相同或不同的剂。当两个区段编码相同的多肽时,第一区段的编码序列不需要与第二区段的序列的补体相同。在一些实施方案中,第二区段的序列为第一区段的编码序列的反向补体。双向构建体可为单链或双链的。本文公开的双向构建体涵盖能够表达任何感兴趣的多肽的构建体。双向构建体可用于基因组插入转基因序列,特别是靶向插入转基因。

[0039] 如本文所用,“反向补体”是指作为参考序列的补体序列的序列,其中补体序列以反向方向写入。例如,对于假设序列5' CTGGACCGA3' (SEQ ID NO:500),“完美”补体序列为3' GACCTGGCT 5' (SEQ ID NO:501),并且“完美”反向补体写为5' TCGGTCCAG 3' (SEQ ID NO:502)。反向补体序列不需要是“完美的”,并且仍可编码与参考序列相同的多肽或相似的多肽。由于密码子使用冗余,反向补体可与编码相同多肽的参考序列不同。如本文所用,“反向补体”还包括例如与参考序列的反向补体序列至少30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%相同的序列。

[0040] 在一些实施方案中,双向核酸构建体包含第一区段,所述第一区段包含编码第一多肽的编码序列(第一转基因),以及第二区段,所述第二区段包含其中所述序列的补体编

码第二多肽的序列(第二转基因)。在一些实施方案中,第一多肽和第二多肽为至少50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%相同。在一些实施方案中,第一多肽和第二多肽包含例如跨50、100、200、500、1000或更多个氨基酸残基的至少80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%相同的氨基酸序列。

[0041] II. 双向核酸构建体

[0042] 本文描述了促进增强插入(例如,增强有效插入)和表达感兴趣的基因的双向核酸构建体。简而言之,本文公开的各种双向构建体包含至少两个核酸区段,其中一个区段(第一区段)包含编码感兴趣的剂的编码序列(例如异源基因)(编码序列在本文中可称为“转基因”或第一转基因),而另一区段(第二区段)包含其中序列的补体编码感兴趣的剂的序列(或第二转基因)。剂可为治疗剂,诸如多肽、功能性RNA、mRNA等。转基因可编码诸如多肽、功能性RNA、mRNA或转录因子的剂。在一些实施方案中,编码序列编码治疗剂,诸如多肽或功能性RNA。至少两个区段可编码相同或不同的多肽或相同或不同的剂。在一些实施方案中,本文公开的双向构建体包含至少两个核酸区段,其中一个区段(第一区段)包含编码感兴趣的多肽的编码序列,而另一区段(第二区段)包含其中序列的补体编码感兴趣的多肽的序列。

[0043] 在一个实施方案中,双向构建体包含顺式中的至少两个核酸区段,其中一个区段(第一区段)包含编码序列(有时在本文中可互换地称为“转基因”),而另一区段(第二区段)包含其中序列的补体编码转基因的序列。第一转基因和第二转基因可为相同或不同的。双向构建体可包含顺式中的至少两个核酸区段,其中一个区段(第一区段)包含在一个方向上编码异源基因的编码序列,而另一区段(第二区段)包含其中序列的补体在另一方向上编码异源基因的序列。也就是说,第一区段是第二区段的补体(不一定是完全补体);第二区段的补体是第一区段的反向补体(虽然两者编码相同的异源蛋白,但不一定是完全反向补体)。双向构建体可包含编码与剪接受体连接的异源基因的第一编码序列以及第二编码序列,其中补体以其他方向编码也与剪接受体连接的异源基因。

[0044] 如本文所用,此种构建体有时称为“供体构建体/模板”。在一些实施方案中,构建体为DNA构建体。设计和对供体构建体进行各种功能/结构修饰的方法是本领域已知的。在一些实施方案中,构建体可包含聚腺苷酸化尾序列、聚腺苷酸化信号序列、剪接受体位点或选择标记中的任何一个或多个。在一些实施方案中,聚腺苷酸化尾序列在编码序列的3'端被编码为例如“poly-A”段。

[0045] 当与如本文所述的基因编辑系统组合使用时,核酸构建体的双向性允许构建体在靶插入位点内的任一方向上插入(不限于在一个方向上插入),从而允许从以下中的任一个表达感兴趣的多肽:a)一个区段的编码序列(例如,左侧区段编码图1左上方ssAAV构建体中的“人F9”)或b)另一区段的补体(例如,右侧区段的补体编码左上方ssAAV构建体图1中倒转指示的“人F9”),从而增强插入和表达效率,如本文所例示。通过基因编辑系统靶向裂解可促进构建体整合和/或转基因表达。各种已知的基因编辑系统可在本公开的实践中使用,所述基因编辑系统包括例如,包括CRISPR/Cas系统的位点特异性DNA裂解系统;锌指核酸酶(ZFN)系统;转录激活因子样效应子核酸酶(TALEN)系统。

[0046] 在一些实施方案中,双向核酸构建体不包含驱动剂或多肽表达的启动子。例如,多肽的表达由宿主细胞的启动子(例如,当转基因整合到宿主细胞的白蛋白基因座中时为内

源白蛋白启动子)驱动。

[0047] 在一些实施方案中,双向核酸构建体包含第一区段,所述第一区段包含多肽的编码序列;以及第二区段,所述第二区段包含多肽的编码序列的反向补体。非多肽剂也是如此。因此,第一区段中的编码序列能够表达多肽,而第二区段中的反向补体的补体也能够表达多肽。如本文所用,当提及包含反向补体序列的第二区段时,“编码序列”是指第二区段的互补(编码)链(即,第二区段中反向补体序列的补体编码序列)。

[0048] 在一些实施方案中,第一区段中编码多肽A的编码序列与也编码多肽A的编码序列的反向补体小于100%互补。也就是说,在一些实施方案中,第一区段包含多肽A的编码序列(1),并且第二区段为多肽A的编码序列(2)的反向补体,其中编码序列(1)与编码序列(2)不同。例如,编码多肽A的编码序列(1)和/或编码序列(2)可利用不同的密码子。在一些实施方案中,可对一个或两个序列进行密码子优化,使得编码序列(1)和编码序列(2)的反向补体具有100%或小于100%的互补性。在一些实施方案中,第二区段的编码序列使用由第一区段中的编码序列所编码的相同多肽的一个或多个氨基酸的一个或多个替代密码子编码多肽。如本文所用,“替代密码子”是指给定氨基酸的密码子使用的变化,并且对于给定的表达系统,可以是或可以不是优选或优化的密码子(密码子优化的)。优选的密码子使用或在给定的表达系统中良好耐受的密码子是本领域已知的。

[0049] 在一些实施方案中,第二区段包含反向补体序列,其采用与第一区段的编码序列不同的密码子使用,以减少发夹的形成。此种反向补体形成的碱基对具有少于第一区段中编码序列的所有核苷酸,但是它任选地编码相同的多肽。在此类情况下,例如第一区段的多肽A的编码序列可与例如双向构建体的后半部分的多肽A的编码序列同源但不相同。在一些实施方案中,第二区段包含与第一区段中的编码序列基本上不互补(例如,不超过70%互补)的反向补体序列。在一些实施方案中,第二区段包含与第一区段中的编码序列高度互补(例如,至少90%互补)的反向补体序列。在一些实施方案中,第二区段包含反向补体序列,其与第一区段中的编码序列具有至少约30%、约35%、约40%、约45%、约50%、约55%、约60%、约65%、约70%、约75%、约80%、约85%、约90%、约95%、约97%或约99%互补性。

[0050] 在一些实施方案中,第二区段包含与第一区段中的编码序列具有100%互补性的反向补体序列。也就是说,第二区段中的序列为第一区段中的编码序列的完全反向补体。以举例的方式,第一区段包含假设序列5' CTGGACCGA 3' (SEQ ID NO:500),并且第二区段包含SEQ ID NO:1的反向补体,即5' TCGGTCCAG 3' (SEQ ID NO:502)。

[0051] 在一些实施方案中,双向核酸构建体包含第一区段,所述第一区段包含多肽或剂(例如第一多肽)的编码序列;以及第二区段,所述第二区段包含多肽或剂(例如第二多肽)的编码序列的反向补体。在一些实施方案中,如上所述,第一多肽和第二多肽是相同的。在一些实施方案中,如上所述,第一治疗剂和第二治疗剂是相同的。在一些实施方案中,第一多肽和第二多肽是不同的。在一些实施方案中,第一治疗剂和第二治疗剂是不同的。例如,第一多肽为多肽A并且第二多肽为多肽B。作为进一步的实例,第一多肽为多肽A并且第二多肽为多肽A的变体(例如,片段(诸如官能片段)、突变体、融合物(包括在多肽末端只添加一个氨基酸)或其组合)。编码多肽的编码序列可任选地包含一个或多个另外的序列,诸如编码氨基或羧基末端氨基酸序列的序列,诸如信号序列、标记序列(例如HiBit)或与多肽连接的异源功能序列(例如核定位序列(NLS)或自裂解)。编码多肽的编码序列可任选地包含编

码一个或多个氨基末端信号肽序列的序列。这些另外的序列中的每个在构建体的第一区段和第二区段中可为相同或不同的。

[0052] 本文所述的双向构建体可根据本文公开的方法用于表达任何多肽。在一些实施方案中,多肽为分泌型多肽。在一些实施方案中,多肽是其功能作为分泌型多肽正常实现(例如功能活性)的多肽。如本文所用,“分泌型多肽”是指由细胞分泌和/或作为可溶性细胞外蛋白具有功能活性的蛋白质。

[0053] 在一些实施方案中,多肽为细胞内多肽。在一些实施方案中,多肽是其功能在细胞内正常实现(例如功能活性)的多肽。如本文所用,“细胞内多肽”是指不由细胞分泌的蛋白质,包括可溶性胞质多肽。

[0054] 在一些实施方案中,多肽为野生型多肽。

[0055] 在一些实施方案中,多肽为肝脏蛋白或其变体。如本文所用,“肝脏蛋白”是例如在肝脏中内源性产生和/或在肝脏中具有功能活性的蛋白质。在一些实施方案中,肝脏蛋白是由肝脏产生的循环蛋白或其变体。在一些实施方案中,肝脏蛋白是在肝脏中具有功能活性的蛋白质或其变体。在一些实施方案中,与一种或多种其他组织类型相比,肝脏蛋白在肝脏中表现出升高的表达。在一些实施方案中,多肽为非肝脏蛋白。在一些实施方案中,多肽包括但不限于因子IX及其变体。

[0056] 在一些实施方案中,双向核酸构建体为线性的。例如,第一区段和第二区段通过接头序列以线性方式连接。在一些实施方案中,包含反向补体序列的第二区段的5'端连接至第一区段的3'端。在一些实施方案中,第一区段的5'端连接至包含反向补体序列的第二区段的3'端。在一些实施方案中,接头序列的长度为约1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、25、30、35、40、45、50、55、60、65、70、75、80、85、90、95、100、150、200、250、300、500、1000、1500、2000或更多个核苷酸。如本领域技术人员将理解的,除了接头序列之外或代替接头序列的其他结构元件可插入第一区段与第二区段之间。

[0057] 可修饰本文公开的构建体,以包括任何特定用途所需的和/或赋予一种或多种所需功能的任何合适的结构特征。在一些实施方案中,本文公开的双向核酸构建体不包含同源臂。在一些实施方案中,本文公开的双向核酸构建体为同源非依赖性供体构建体。在一些实施方案中,部分由于核酸构建体的双向功能,双向构建体可以如本文所述的任一方向(取向)插入到基因组基因座中,以允许有效插入和/或表达感兴趣的多肽。在一些实施方案中,双向核酸构建体包括第一区段和第二区段,每个区段各自在转基因上游具有剪接受体。在某些实施方案中,剪接受体与宿主细胞的安全港位点的剪接供体序列相容,例如人白蛋白基因内含子1的剪接供体。

[0058] 在一些实施方案中,本文所述的组合物包含一个或多个内部核糖体进入位点(IRES)。IRES首先被鉴定为特征性小核糖核酸病毒RNA,在5'帽结构不存在的情况下,IRES在启动蛋白质合成中起重要作用。IRES可作为唯一的核糖体结合位点,或可充当多核苷酸的多个核糖体结合位点中的一个。含有多于一个功能性核糖体结合位点的构建体可编码由核糖体独立翻译的若干肽或多肽(“多顺反子核酸分子”)。替代地,构建体可包含IRES以表达未与内源性多肽(即白蛋白信号肽)融合的异源蛋白质。可利用的IRES序列的实例包括但不限于来自以下的那些:小核糖核酸病毒(例如FMDV)、瘟病毒(CFFV)、脊髓灰质炎病毒(PV)、脑心肌炎病毒(ECMV)、口蹄疫病毒(FMDV)、丙型肝炎病毒(HCV)、经典猪瘟病毒

(CSFV)、鼠白血病病毒 (MLV)、猴免疫缺陷病毒 (SIV) 或蜂麻痹病毒 (CrPV)。

[0059] 在一些实施方案中,核酸构建体包含编码自裂解肽的序列,诸如2A序列或2A样序列。在一些实施方案中,自裂解肽位于感兴趣的多肽的上游。在一个实施方案中,编码2A肽的序列可用于分离两个或更多个感兴趣的多肽的编码区。在另一个实施方案中,此序列可用于将编码序列与构建体分离,并将编码序列与内源基因座(即内源性白蛋白信号序列)分离。作为非限制性实例,编码2A肽的序列可在区域A与区域B之间(A-2A-B)。2A肽的存在将导致一个长蛋白被裂解成蛋白质A、蛋白质B和2A肽。蛋白质A和蛋白质B可为相同或不同的感兴趣的多肽。

[0060] 在一些实施方案中,第一区段和第二区段中的一个或两个包含开放阅读框下游的聚腺苷酸化尾序列和/或聚腺苷酸化信号序列。在一些实施方案中,聚腺苷酸化尾序列在第一区段和/或第二区段的3'端被编码为例如“poly-A”段。在一些实施方案中,由于在第一区段和/或第二区段的3'端处或附近编码的聚腺苷酸化信号序列的结果,共转录地提供聚腺苷酸化尾序列。在一些实施方案中,poly-A尾包含至少20、30、40、50、60、70、80、90或100个腺嘌呤,任选多至300个腺嘌呤。在一些实施方案中,poly-A尾包含95、96、97、98、99或100个腺嘌呤核苷酸。设计合适的聚腺苷酸化尾序列和/或聚腺苷酸化信号序列的方法是本领域众所周知的。本文公开并例示了合适的剪接受体序列,包括小鼠白蛋白和人FIX剪接受体位点。在一些实施方案中,聚腺苷酸化信号序列AAUAAA (SEQ ID NO:800) 通常用于哺乳动物系统,尽管诸如UAUAAA (SEQ ID NO:801) 或AU/GUAAA (SEQ ID NO:802) 的变体已被鉴定。参见例如,NJ Proudfoot, *Genes&Dev.* 25 (17):1770-82, 2011。在一些实施方案中,包括polyA尾序列。

[0061] 在一些实施方案中,本文公开的构建体可为单链、双链或部分单链和部分双链的DNA或RNA,并且可以线性或环状(例如,小环)形式引入到宿主细胞中。参见例如,美国专利公布号2010/0047805、2011/0281361、2011/0207221。如果以线性形式引入,供体序列的端部可通过本领域技术人员已知的方法来保护(例如,免受核酸外切降解)。例如,将一个或多个双脱氧核苷酸残基添加至线性分子的3'末端和/或将自互补寡核苷酸连接至一端或两端。参见例如,Chang等人(1987) *Proc. Natl. Acad. Sci. USA* 84:4959-4963; Nehls等人(1996) *Science* 272:886-889。用于保护外源多核苷酸免受降解的另外方法包括但不限于添加一个或多个末端氨基和使用修饰的核苷酸间键联,例如像硫代磷酸酯、氨基磷酸酯和O-甲基核糖或脱氧核糖残基。

[0062] 在一些实施方案中,可插入构建体,使得其表达由插入位点的内源启动子(例如,当供体整合到宿主细胞的白蛋白基因座中时为内源白蛋白启动子)驱动。在此类情况下,转基因可能缺乏驱动其表达的控制元件(例如,启动子和/或增强子)(例如,无启动子的构建体)。尽管如此,显而易见的是,在其他情况下,构建体可包含启动子和/或增强子,例如组成型启动子或在整合时驱动功能蛋白表达的诱导型或组织特异性(例如,肝或血小板特异性)启动子。构建体可包含编码异源蛋白的序列,所述序列在编码信号肽的信号序列的下游并且可操作地连接到所述信号序列。

[0063] 在一些实施方案中,核酸构建体在同源非依赖性插入编码异源多肽的核酸中起作用。在一些实施方案中,核酸构建体在非分裂细胞中起作用,例如在其中NHEJ而不是HR是修复双链DNA断裂的主要机制的细胞中。核酸可为同源非依赖性供体构建体。例如,构建体可

为单链或双链DNA。在一些实施方案中,可以如本文所述修饰核酸(例如,使用核苷类似物)。

[0064] 在一些实施方案中,本文公开的构建体包含构建体的任一端或两端上的剪接受体位点,例如,第一区段和/或第二区段中的开放阅读框的5'或一个或两个转基因序列的5'。在一些实施方案中,剪接受体位点包含NAG。在其他实施方案中,剪接受体位点由NAG组成。在一些实施方案中,剪接受体为白蛋白剪接受体,例如,用于将白蛋白的外显子1和2剪接在一起的白蛋白剪接受体。在一些实施方案中,剪接受体衍生自人白蛋白基因。在一些实施方案中,剪接受体衍生自小鼠白蛋白基因。在一些实施方案中,剪接受体为F9(或“FIX”)剪接受体,例如,用于将F9的外显子1和2剪接在一起的F9剪接受体。在一些实施方案中,剪接受体衍生自人F9基因。在一些实施方案中,剪接受体衍生自小鼠F9基因。在真核生物中有用的另外的合适的剪接受体位点(包括人工剪接受体)是已知的,并且可从本领域得到。参见例如,Shapiro等人,1987,Nucleic Acids Res.,15,7155-7174,Burset等人,2001,Nucleic Acids Res.,29,255-259。

[0065] 在一些实施方案中,本文公开的构建体可在一端或两端上进行修饰,以根据需要包括一个或多个合适的结构特征,和/或赋予一个或多个功能益处。例如,结构修饰可根据用于将本文公开的构建体递送到宿主细胞的一个或多个方法而变化,例如,使用病毒载体递送或包装成脂质纳米颗粒以进行递送。此类修饰包括但不限于例如末端结构,诸如反向末端重复序列(ITR)、发夹、环和其他结构诸如螺旋管(toroid)。在一些实施方案中,本文公开的构建体包含一个、两个或三个ITR。在一些实施方案中,本文公开的构建体包含不超过两个ITR。结构修饰的各种方法是本领域已知的。

[0066] 在一些实施方案中,构建体的一端或两端可通过本领域已知的方法来保护(例如,免受核酸外切降解)。例如,将一个或多个双脱氧核苷酸残基添加至线性分子的3'末端和/或将自互补寡核苷酸连接至一端或两端。参见例如,Chang等人(1987)Proc.Natl.Acad.Sci.USA84:4959-4963;Nehls等人(1996)Science 272:886-889。用于保护构建体免受降解的另外方法包括但不限于添加一个或多个末端氨基和使用修饰的核苷酸间键联,例如像硫代磷酸酯、氨基磷酸酯和O-甲基核糖或脱氧核糖残基。

[0067] 在一些实施方案中,可将本文公开的构建体引入到细胞中作为载体的一部分,所述载体具有另外的序列例如像复制起点、启动子和编码抗生素耐药性的基因。构建体可省略病毒元件。在一些实施方案中,构建体可作为裸露核酸、作为与剂(诸如脂质体、聚合物或泊洛沙姆)复合的核酸引入,或可通过病毒载体(例如,腺病毒、AAV、疱疹病毒、逆转录病毒、慢病毒)来递送。

[0068] 在一些实施方案中,虽然不为表达所需,但是本文公开的构建体还可包括转录或翻译调控序列,例如启动子、增强子、绝缘子、内部核糖体进入位点、编码肽的序列和/或聚腺苷酸化信号。

[0069] 在一些实施方案中,包含感兴趣的多肽的编码序列的构建体可包括以下修饰中的一个或多个:密码子优化(例如,对人密码子进行的)和/或添加一个或多个糖基化位点。参见例如,McIntosh等人(2013)Blood (17):3335-44。

[0070] III. 基因编辑系统

[0071] 各种已知的基因编辑系统可在本公开的实践中使用,所述基因编辑系统包括例如,CRISPR/Cas系统;锌指核酸酶(ZFN)系统;和转录激活因子样效应子核酸酶(TALEN)系

统。一般来讲,这些方法可涉及使用工程改造的裂解系统来在靶DNA序列中诱导双链断裂(DSB)或切口(例如,单链断裂或SSB)。裂解或切口可通过使用特定的核酸酶(诸如工程改造的ZFN、TALEN)或使用CRISPR/Cas系统来发生,所述CRISPR/Cas系统具有工程改造的指导RNA来指导靶DNA序列的特异性裂解或切口。此外,例如基于Argonaute系统,已开发了靶向核酸酶,并且正在开发另外的核酸酶(例如,来自嗜热栖热菌(*T. thermophilus*),称为“TtAgo”,参见Swarts等人(2014) *Nature* 507 (7491):258-261),其也可具有用于基因组编辑和基因治疗的潜力。

[0072] 在一些实施方案中,CRISPR/Cas系统可用于在宿主基因组内的所需基因座处创建插入位点,在所述位点可插入本文公开的双向构建体(例如,双向构建体)来表达一个或多个感兴趣的多肽。设计靶向宿主基因组的任何所需基因座的合适的指导RNA以供插入的方法在本领域中是众所周知的。包含转基因的双向构建体相对于其插入位点可为异源的,例如,将异源转基因插入“安全港”基因座。包含转基因的双向构建体相对于其插入位点可为非异源的,例如,将野生型转基因插入其内源基因座。

[0073] “安全港”基因座是在基因组内的基因座,其中外源核酸可被插入而不对宿主细胞(例如肝细胞)产生显著的有害影响,例如,不引起细胞凋亡、坏死和/或衰老,或不引起与对照细胞相比超过5%、10%、15%、20%、25%、30%或40%的细胞凋亡、坏死和/或衰老。参见例如,Hsin等人,“Hepatocyte death in liver inflammation, fibrosis, and tumorigenesis,”2017。在一些实施方案中,安全港基因座允许外源核酸(例如外源基因)表达而不对宿主细胞或细胞群(诸如肝细胞或肝脏细胞)产生显著的有害影响,例如,不引起细胞凋亡、坏死和/或衰老,或不引起与对照细胞或细胞群相比超过5%、10%、15%、20%、25%、30%或40%的细胞凋亡、坏死和/或衰老。安全港可在白蛋白基因内,诸如人白蛋白基因内。安全港可在白蛋白内含子1区域内,例如人白蛋白内含子1内。安全港可为例如针对肝脏组织或肝细胞宿主细胞的人安全港。一个或多个核酸酶靶向的安全港基因座的非限制性实例包括CCR5、HPRT、AAVS1、Rosa、白蛋白、AAVS1 (PPP1 R12C)、AngptiS、ApoC3、ASGR2、FIX (F9)、G6PC、Gys2、HGD、Lp(a)、Pcsk9、SERPINA1、TF和TTR。参见例如,美国专利号7,951,925和8,110,379;美国公布号2008/0159996;2010/00218264;2012/0017290;2011/0265198;2013/0137104;2013/0122591;2013/0177983;2013/0177960;和WO 2017093804。如本文所例示,在一些实施方案中,指导RNA可设计成靶向人或小鼠白蛋白基因座(例如,内含子1)。表5至表10示出了本文例示的指导RNA的实例。应当理解,根据本发明方法,可靶向任何其他基因座以插入包含转基因的双向构建体。

[0074] 在一些实施方案中,可将异源基因插入安全港基因座中并使用安全港基因座的内源性信号序列,例如,由外显子1编码的白蛋白信号序列。例如,可将编码序列插入人白蛋白内含子1中,使得其位于人白蛋白外显子1的信号序列的下游并与所述信号序列融合。

[0075] 在一些实施方案中,基因可包含其自身的信号序列,可被插入安全港基因座中,并且可进一步使用安全港基因座的内源性信号序列。例如,可将包含天然信号序列的编码序列插入人白蛋白内含子1中,使得其位于由外显子1编码的人白蛋白的信号序列的下游并与所述信号序列融合。

[0076] 在一些实施方案中,基因可包含其自身的信号序列和内部核糖体进入位点(IRES),可被插入安全港基因座中,并且可进一步使用安全港基因座的内源性信号序列。例

如,可将包含天然信号序列和IRES序列的编码序列插入人白蛋白内含子1中,使得其位于由外显子1编码的人白蛋白的信号序列的下游并与所述信号序列融合。

[0077] 在一些实施方案中,基因可包含其自身的信号序列和IRES,可被插入安全港基因座中,并且不使用安全港基因座的内源性信号序列。例如,可将包含天然信号序列和IRES序列的编码序列插入人白蛋白内含子1中,使得其不与由外显子1编码的人白蛋白的信号序列融合。在这些实施方案中,蛋白质从IRES位点翻译并且不是嵌合的(例如,与异源蛋白融合的白蛋白信号肽),其可有利地为非免疫原性或低免疫原性。在一些实施方案中,蛋白质不在细胞外分泌和/或转运。

[0078] 在一些实施方案中,基因可被插入安全港基因座中,并且可包含IRES,并且不使用任何信号序列。例如,可将包含IRES序列且不包含天然信号序列的编码序列插入人白蛋白内含子1中,使得其不与由外显子1编码的人白蛋白的信号序列融合。在一些实施方案中,蛋白质从IRES位点翻译而不需要任何信号序列。在一些实施方案中,蛋白质不在细胞外分泌和/或转运。

[0079] 还应当理解,Cas核酸酶(诸如可在本发明方法中使用的Cas9核酸酶)的指导RNA可包括各种已知的变化和修饰(例如,化学修饰)中的任一种,包括一种或多种非天然和/或天然存在的组分或构型的存在,所述组分或构型用于替代或补充标准A、G、C和U残基。例如,本文例示的指导序列中的每个(表5至表10)还可包含另外的核苷酸以从例如SpyCas9 CRISPR/Cas系统形成crRNA、指导RNA和/或sgRNA。例如,本文例示的指导序列中的每个(表5至表10)还可包含另外的核苷酸以形成crRNA和/或sgRNA,其中以下示例性核苷酸序列在指导序列3'端之后:GUUUUAGAGCUAUGCUGUUUUG (SEQ ID NO:200),呈5'至3'方向。在sgRNA的情况下,指导序列(诸如表5至表10中列出的指导序列)还可包含另外的核苷酸以形成sgRNA,例如,其中以下示例性核苷酸序列(SpyCas9指导序列)在指导序列的3'端之后:GUUUUAGAGCUAGAAAUAAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU (SEQ ID NO:201)或GUUUUAGAGCUAGAAAUAAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCCCGAGUCGGUGC (SEQ ID NO:202),呈5'至3'方向。

[0080] 指导RNA可任选地包含trRNA。在本文所述的每个组合物和方法实施方案中,crRNA和trRNA可作为单个RNA (sgRNA) 缔合或可在单独的RNA (dgRNA) 上。在sgRNA的情况下,crRNA和trRNA组分可例如通过磷酸二酯键或其他共价键共价连接。在一些实施方案中,sgRNA在核苷酸之间包含不是磷酸二酯键联的一个或多个键联。在本文所述的组合物、用途和方法实施方案中的每个中,指导RNA可包含两个RNA分子作为“双指导RNA”或“dgRNA”。dgRNA包含第一RNA分子,所述第一RNA分子包含含有例如表5至表10中任一个所示的指导序列的crRNA,以及第二RNA分子,所述第二RNA分子包含trRNA。第一RNA分子和第二RNA分子可以不共价连接,但是可通过crRNA和trRNA的部分之间的碱基配对形成RNA双链体。

[0081] 在一些实施方案中,本文公开的指导RNA结合至原间隔子相邻基序(PAM)上游的区域。如本领域技术人员将理解的,PAM序列发生在与含有靶序列的链相反的链上。也就是说,PAM序列在靶链(含有指导RNA结合的靶序列的链)的补体链上。在一些实施方案中,PAM选自由NGG、NNGRRT、NNGRR (N)、NNAGAAW、NNNNG (A/C) TT和NNNNRYAC组成的组。在一些实施方案中,PAM为NGG。

[0082] 在一些实施方案中,本文提供的指导RNA序列与邻近PAM序列的序列互补。

[0083] 在一些实施方案中,指导RNA序列包含与根据人参考基因组hg38中的坐标选自本文表中的基因组区域内的序列互补的序列。在一些实施方案中,指导RNA序列包含与包含选自表5至表10中的基因组区域内的5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个连续核苷酸的序列互补的序列。在一些实施方案中,指导RNA序列包含与包含跨越选自表5至表10中的5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个连续核苷酸的序列互补的序列。

[0084] 本文公开的指导RNA介导靶特异性切割,从而导致双链断裂(DSB)。本文公开的指导RNA介导靶特异性切割,从而导致单链断裂(SSB或切口)。

[0085] 使用各种RNA指导的DNA结合剂(例如核酸酶,诸如Cas核酸酶,例如Cas9)的方法在本领域中也是众所周知的。虽然本文例示了将双向核酸与CRISPR/Cas系统一起使用,但是应当理解,也可使用系统的合适变体。应当理解,根据上下文,可将RNA指导的DNA结合剂提供为核酸(例如,DNA或mRNA)或蛋白质。在一些实施方案中,本发明方法可在已经包含和/或表达RNA指导的DNA结合剂的宿主细胞中实施。

[0086] 在一些实施方案中,RNA指导的DNA结合剂(诸如Cas9核酸酶)具有裂解酶活性,其也可称为双链核酸内切酶活性。在一些实施方案中,RNA指导的DNA结合剂(诸如Cas9核酸酶)具有切口酶活性,其也可称为单链核酸内切酶活性。在一些实施方案中,RNA指导的DNA结合剂包含Cas核酸酶。Cas核酸酶的实例包括酿脓链球菌、金黄色葡萄球菌和其他原核生物的II型CRISPR系统的那些(参见例如,下一段中的列表),以及它们的变体或突变体(例如,工程改造的、非天然存在的、天然存在的或其他变体)形式。参见例如,US2016/0312198A1;US 2016/0312199 A1。

[0087] Cas核酸酶可衍生自的非限制性示例性菌种包括酿脓链球菌、嗜热链球菌、链球菌某种、金黄色葡萄球菌、无害李斯特氏菌(*Listeria innocua*)、加氏乳杆菌(*Lactobacillus gasserii*)、新凶手弗朗西丝菌(*Francisella novicida*)、产琥珀酸沃林氏菌(*Wolinella succinogenes*)、华德萨特菌(*Sutterella wadsworthensis*)、 γ -变形杆菌(*Gammaproteobacterium*)、脑膜炎奈瑟氏菌、空肠弯曲杆菌、多杀巴斯德氏菌(*Pasteurella multocida*)、产琥珀酸丝状杆菌(*Fibrobacter succinogene*)、深红红螺菌(*Rhodospirillum rubrum*)、达氏拟诺卡氏菌(*Nocardiopsis dassonvillei*)、始旋链霉菌(*Streptomyces pristinaespiralis*)、绿色产色链霉菌(*Streptomyces viridochromogenes*)、绿色产色链霉菌、粉红链孢囊菌(*Streptosporangium roseum*)、粉红链孢囊菌、酸热脂环酸芽孢杆菌(*Alicyclobacillus acidocaldarius*)、假真菌样芽孢杆菌(*Bacillus pseudomycoides*)、还原硒酸盐芽孢杆菌(*Bacillus selenitireducens*)、西伯利亚微小杆菌(*Exiguobacterium sibiricum*)、德氏乳杆菌(*Lactobacillus delbrueckii*)、唾液乳杆菌(*Lactobacillus salivarius*)、布氏乳杆菌(*Lactobacillus buchneri*)、齿垢密螺旋体(*Treponema denticola*)、海洋微颤菌(*Microscilla marina*)、伯克霍尔德氏菌(*Burkholderiales bacterium*)、食萘极地单胞菌(*Polaromonas naphthalenivorans*)、极地单胞菌某种(*Polaromonas sp.*)、瓦氏鳄球藻(*Crocospaera watsonii*)、蓝杆藻某种(*Cyanothece sp.*)、铜绿微囊藻(*Microcystis aeruginosa*)、聚球藻某种(*Synechococcus sp.*)、阿拉伯糖醋盐杆菌(*Acetohalobium arabaticum*)、丹氏制氮菌(*Ammonifex degensii*)、热解纤维素菌(*Caldicelulosiruptor beccsii*)、金矿菌

(*Candidatus Desulforudis*)、肉毒杆菌(*Clostridium botulinum*)、艰难梭菌(*Clostridium difficile*)、大芬戈尔德菌(*Fingoldia magna*)、嗜热盐碱厌氧菌(*Natranaerobius thermophilus*)、嗜热丙酸降解菌(*Pelotomaculum thermopropionicum*)、喜温嗜酸硫杆菌(*Acidithiobacillus caldus*)、氧化亚铁嗜酸硫杆菌(*Acidithiobacillus ferrooxidans*)、酒色别样着色菌(*Allochromatium vinosum*)、海杆菌某种(*Marinobacter sp.*)、嗜盐亚硝化球菌(*Nitrosococcus halophilus*)、瓦氏亚硝化球菌(*Nitrosococcus watsoni*)、游海假交替单胞菌(*Pseudoalteromonas haloplanktis*)、消旋纤线杆菌(*Ktedonobacter racemifer*)、调查甲烷盐菌(*Methanohalobium evestigatum*)、多变鱼腥藻(*Anabaena variabilis*)、泡沫节球藻(*Nodularia spumigena*)、念珠藻某种(*Nostoc sp.*)、极大节旋藻(*Arthrospira maxima*)、钝顶节旋藻(*Arthrospira platensis*)、节旋藻某种、鞘丝藻某种(*Lyngbya sp.*)、原型微鞘藻(*Microcoleus chthonoplastes*)、颤藻某种(*Oscillatoria sp.*)、运动石袍菌(*Petrogloia mobilis*)、非洲栖热腔菌(*Thermosiphon africanus*)、巴氏链球菌(*Streptococcus pasteurianus*)、灰色奈瑟氏菌(*Neisseria cinerea*)、红嘴鸥弯曲杆菌(*Campylobacter lari*)、食清洁剂细小棒菌(*Parvibaculum lavamentivorans*)、白喉棒状杆菌(*Corynebacterium diphtheriae*)、氨基酸球菌某种(*Acidaminococcus sp.*)、毛螺科菌ND2006(*Lachnospiraceae bacterium ND2006*)和深海单细胞蓝藻(*Acaryochloris marina*)。

[0088] 在一些实施方案中,CAS核酸酶是来自酿脓链球菌的Cas9核酸酶。在一些实施方案中,CAS核酸酶是来自嗜热链球菌的Cas9核酸酶。在一些实施方案中,CAS核酸酶是来自脑膜炎奈瑟氏菌的Cas9核酸酶。在一些实施方案中,CAS核酸酶是来自金黄色葡萄球菌的Cas9核酸酶。在一些实施方案中,Cas核酸酶是来自新凶手弗朗西丝菌的Cpf1核酸酶。在一些实施方案中,Cas核酸酶是来自氨基酸球菌某种的Cpf1核酸酶。在一些实施方案中,Cas核酸酶是来自毛螺科菌ND2006的Cpf1核酸酶。在其他实施方案中,Cas核酸酶是来自土拉热弗朗西丝菌(*Francisella tularensis*)、毛螺科菌、瘤胃溶纤维丁酸弧菌(*Butyrivibrio proteoclasticus*)、异域菌门菌(*Peregrinibacteria bacterium*)、Parcubacteria门菌、史密斯氏菌属(*Smithella*)、氨基酸球菌属(*Acidaminococcus*)、候选白蚁甲烷枝原体(*Candidatus Methanoplasma termitum*)、挑剔真杆菌(*Eubacterium eligens*)、牛眼莫拉氏菌(*Moraxella bovoculi*)、稻田钩端螺旋体(*Leptospira inadai*)、犬口腔卟啉单胞菌(*Porphyromonas crevioricanis*)、解糖脲普雷沃菌(*Prevotella disiens*)或猕猴卟啉单胞菌(*Porphyromonas macacae*)的Cpf1核酸酶。在某些实施方案中,Cas核酸酶是来自氨基酸球菌或毛螺菌科的Cpf1核酸酶。

[0089] 在一些实施方案中,gRNA与RNA指导的DNA结合剂一起被称为核糖核蛋白复合物(RNP)。在一些实施方案中,RNA指导的DNA结合剂为Cas核酸酶。在一些实施方案中,gRNA与Cas核酸酶一起被称为Cas RNP。在一些实施方案中,RNP包含I型、II型或III型组分。在一些实施方案中,Cas核酸酶是来自II型CRISPR/Cas系统的Cas9蛋白。在一些实施方案中,gRNA与Cas9一起被称为Cas9RNP。

[0090] 野生型Cas9具有两个核酸酶结构域:RuvC和HNH。RuvC结构域裂解非靶DNA链,并且HNH结构域裂解DNA的靶链。在一些实施方案中,Cas9蛋白包含不止一个RuvC结构域和/或不

止一个HNH结构域。在一些实施方案中,Cas9蛋白为野生型Cas9。在组合物、用途和方法实施方案的每个中,Cas诱导靶DNA中的双链断裂。

[0091] 在一些实施方案中,使用嵌合Cas核酸酶,其中蛋白质的一个结构域或区域被不同蛋白质的一部分替换。在一些实施方案中,可来自不同核酸酶诸如FokI的结构域替换Cas核酸酶结构域。在一些实施方案中,Cas核酸酶可为修饰的核酸酶。

[0092] 在其他实施方案中,Cas核酸酶可来自I型CRISPR/Cas系统。在一些实施方案中,Cas核酸酶可为I型CRISPR/Cas系统的级联复合物的组分。在一些实施方案中,Cas核酸酶可为Cas3蛋白。在一些实施方案中,Cas核酸酶可来自III型CRISPR/Cas系统。在一些实施方案中,Cas核酸酶可具有RNA裂解活性。

[0093] 在一些实施方案中,RNA指导的DNA结合剂具有单链切口酶活性,即可切割一条DNA链以产生单链断裂,也称为“切口”。在一些实施方案中,RNA指导的DNA结合剂包含Cas切口酶。切口酶是在dsDNA中创建切口的酶,即切割DNA双螺旋的一条链而不切割另一条。在一些实施方案中,Cas切口酶是其中核酸内切活性位点例如通过催化结构域中的一个或多个改变(例如,点突变)而失活的Cas核酸酶(例如,上面讨论的Cas核酸酶)的形式。关于Cas切口酶和示例性催化结构域改变的讨论,参见例如美国专利号8,889,356。在一些实施方案中,Cas切口酶诸如Cas9切口酶具有失活的RuvC或HNH结构域。

[0094] 在一些实施方案中,对RNA指导的DNA结合剂进行修饰以仅含有一个功能性核酸酶结构域。例如,可修饰剂蛋白,使得核酸酶结构域中的一个突变或完全或部分缺失,以降低其核酸裂解活性。在一些实施方案中,使用具有活性降低的RuvC结构域的切口酶。在一些实施方案中,使用具有非活性RuvC结构域的切口酶。在一些实施方案中,使用具有活性降低的HNH结构域的切口酶。在一些实施方案中,使用具有非活性HNH结构域的切口酶。

[0095] 在一些实施方案中,Cas蛋白核酸酶结构域内的保守氨基酸被取代以降低或改变核酸酶活性。在一些实施方案中,Cas核酸酶可包含RuvC或RuvC样核酸酶结构域中的氨基酸取代。RuvC或RuvC样核酸酶结构域中的示例性氨基酸取代包括D10A(基于酿脓链球菌Cas9蛋白)。参见例如,Zetsche等人(2015) Cell Oct 22:163(3):759-771。在一些实施方案中,Cas核酸酶可包含HNH或HNH样核酸酶结构域中的氨基酸取代。HNH或HNH样核酸酶结构域中的示例性氨基酸取代包括E762A、H840A、N863A、H983A和D986A(基于酿脓链球菌Cas9蛋白)。参见例如,Zetsche等人(2015)。其他示例性氨基酸取代包括D917A、E1006A和D1255A(基于新凶手弗朗西丝菌U112 Cpf1(FnCpf1)序列(UniProtKB-A0Q7Q2(CPF1_FRATN)))。

[0096] 在一些实施方案中,将切口酶与分别与靶序列的正义链和反义链互补的一对指导RNA组合提供。在此实施方案中,指导RNA将切口酶引导到靶序列并通过在靶序列的相反链上产生切口(即,双切口)来引入DSB。在一些实施方案中,切口酶与靶向DNA的相反链的两个单独的指导RNA一起使用,以在靶DNA中产生双切口。在一些实施方案中,切口酶与选择为非常接近的两个单独的指导RNA一起使用,以在靶DNA中产生双切口。

[0097] 在一些实施方案中,RNA指导的DNA结合剂包含一个或多个异源功能结构域(例如,是或包含融合多肽)。

[0098] 在一些实施方案中,异源功能结构域可促进RNA指导的DNA结合剂转运到细胞核中。例如,异源功能结构域可为核定位信号(NLS)。在一些实施方案中,RNA指导的DNA结合剂可与1至10个NLS融合。在一些实施方案中,RNA指导的DNA结合剂可与1至5个NLS融合。在一

些实施方案中, RNA指导的DNA结合剂可与一个NLS融合。当使用一个NLS时, NLS可在RNA指导的DNA结合剂序列的N末端或C末端连接。也可将其插入到RNA指导的DNA结合剂序列中。在其他实施方案中, RNA指导的DNA结合剂可与不止一个NLS融合。在一些实施方案中, RNA指导的DNA结合剂可与2、3、4或5个NLS融合。在一些实施方案中, RNA指导的DNA结合剂可与两个NLS融合。在某些情况下, 两个NLS可为相同(例如, 两个SV40 NLS) 或不同的。在一些实施方案中, RNA指导的DNA结合剂融合到在羧基末端连接的两个SV40 NLS序列。在一些实施方案中, RNA指导的DNA结合剂可与两个NLS融合, 一个在N末端连接, 并且一个在C末端连接。在一些实施方案中, RNA指导的DNA结合剂可与3个NLS融合。在一些实施方案中, RNA指导的DNA结合剂可不与NLS融合。在一些实施方案中, NLS可为单分型序列(monopartite sequence), 例如像SV40 NLS, PKKKRKV (SEQ ID NO:600) 或PKKKRRV (SEQ ID NO:601)。在一些实施方案中, NLS可为二分型序列(bipartite sequence), 诸如核质蛋白的NLS, KRPAATKKAGQAKKKK (SEQ ID NO:602)。在特定的实施方案中, 单个PKKKRKV (SEQ ID NO:600) NLS可在RNA指导的DNA结合剂的C末端连接。融合位点任选地包括一个或多个接头。

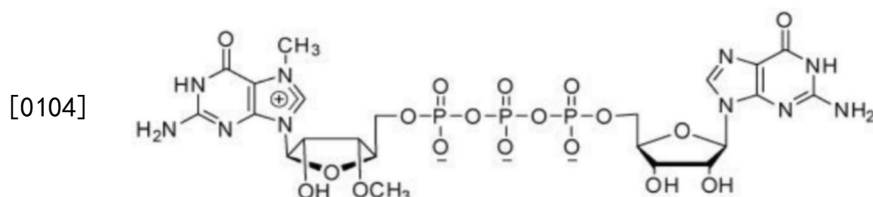
[0099] 如上所述, RNA指导的DNA结合剂可为编码RNA指导的DNA结合多肽的核酸。在一些实施方案中, RNA指导的DNA结合剂包含mRNA, 所述mRNA包含编码RNA指导的DNA结合剂诸如本文所述的Casintegrate核酸酶的开放阅读框(ORF)。在一些实施方案中, 提供、使用或施用包含编码RNA指导的DNA结合剂诸如Cas核酸酶的ORF的mRNA。如下所述, 包含Cas核酸酶的mRNA可包含Cas9核酸酶, 诸如具有裂解酶、切口酶和/或位点特异性DNA结合活性的酿脓链球菌Cas9核酸酶。在一些实施方案中, 编码RNA指导的DNA核酸酶的ORF为“修饰的RNA指导的DNA结合剂ORF”或简单的“修饰的ORF”, 其用作指示ORF被修饰的简写。

[0100] 本文提供了Cas9 ORF, 包括修饰的Cas9 ORF, 并且其是本领域已知的。作为一个实例, 可对Cas9 ORF进行密码子优化, 使得编码序列包括一个或多个氨基酸的一个或多个替代密码子。如本文所用, “替代密码子”是指给定氨基酸的密码子使用的变化, 并且对于给定的表达系统, 可以是或可以不是优选或优化的密码子(密码子优化的)。优选的密码子使用或在给定的表达系统中良好耐受的密码子是本领域已知的。W02013/176772、W02014/065596、W02016/106121和W02019/067910的Cas9编码序列、Cas9 mRNA和Cas9蛋白序列特此以引用的方式并入。特别地, W02019/067910第[0449]段的表中的ORF和Cas9氨基酸序列以及W02019/067910第[0214]至[0234]段中的Cas9 mRNA和ORF特此以引用的方式并入。

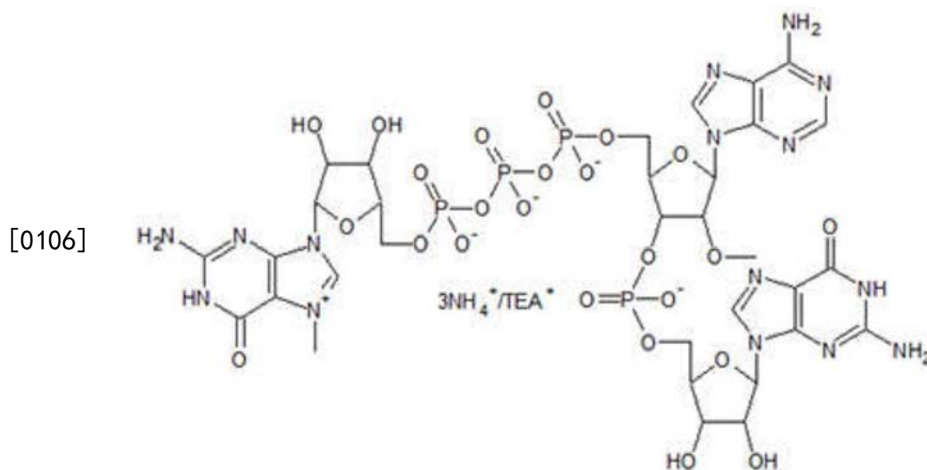
[0101] 在一些实施方案中, 修饰的ORF可包含至少一个、多个或所有尿苷位置处的修饰的尿苷。在一些实施方案中, 修饰的尿苷是在5位例如被卤素、甲基或乙基修饰的尿苷。在一些实施方案中, 修饰的尿苷是在1位例如被卤素、甲基或乙基修饰的假尿苷。修饰的尿苷可为例如假尿苷、N1-甲基-假尿苷、5-甲氧基尿苷、5-碘尿苷或其组合。在一些实施方案中, 修饰的尿苷为5-甲氧基尿苷。在一些实施方案中, 修饰的尿苷为5-碘尿苷。在一些实施方案中, 修饰的尿苷为假尿苷。在一些实施方案中, 修饰的尿苷为N1-甲基-假尿苷。在一些实施方案中, 修饰的尿苷为假尿苷和N1-甲基-假尿苷的组合。在一些实施方案中, 修饰的尿苷为假尿苷和5-甲氧基尿苷的组合。在一些实施方案中, 修饰的尿苷为N1-甲基假尿苷和5-甲氧基尿苷的组合。在一些实施方案中, 修饰的尿苷为5-碘尿苷和N1-甲基-假尿苷的组合。在一些实施方案中, 修饰的尿苷为假尿苷和5-碘尿苷的组合。在一些实施方案中, 修饰的尿苷为5-碘尿苷和5-甲氧基尿苷的组合。

[0102] 在一些实施方案中,本文公开的mRNA包含5'帽,诸如Cap0、Cap1或Cap2。5'帽通常是通过5'-三磷酸将7-甲基鸟嘌呤核糖核苷酸(其可被进一步修饰,如下面例如关于ARCA所讨论的)连接到mRNA的5'至3'链的第一个核苷酸的5'位置,即第一帽近端核苷酸。在Cap0中,mRNA的第一帽近端核苷酸和第二帽近端核苷酸的核糖均包含2'-羟基。在Cap1中,mRNA的第一转录核苷酸和第二转录核苷酸的核糖分别包含2'-甲氧基和2'-羟基。在Cap2中,mRNA的第一帽近端核苷酸和第二帽近端核苷酸的核糖均包含2'-甲氧基。参见例如,Katibah等人(2014) *Proc Natl Acad Sci USA* 111(33):12025-30;Abbas等人(2017) *Proc Natl Acad Sci USA* 114(11):E2106-E2115。大多数内源性高等真核mRNA,包括哺乳动物mRNA诸如人mRNA,均包含Cap1或Cap2。Cap0和不同于Cap1和Cap2的其他帽结构可在哺乳动物诸如人中具有免疫原性,这是由于先天免疫系统的组分(诸如IFIT-1和IFIT-5)将其识别为“非自身”,这可能导致包括I型干扰素的细胞因子水平升高。先天免疫系统的组分(诸如IFIT-1和IFIT-5)也可能与eIF4E竞争,以便与具有Cap1或Cap2以外的帽的mRNA结合,从而可能抑制mRNA的翻译。

[0103] 可共转录地包括帽。例如,ARCA(抗反向帽类似物;Thermo Fisher Scientific目录号AM8045)是帽类似物,其包含与可在起始时体外并入转录的鸟嘌呤核糖核苷酸的5'位置连接的7-甲基鸟嘌呤3'-甲氧基-5'-三磷酸酯。ARCA产生Cap0帽,其中第一帽近端核苷酸的2'位置为羟基。参见例如,Stepinski等人,(2001) “Synthesis and properties of mRNAs containing the novel ‘anti-reverse’ cap analogs 7-methyl (3'-O-methyl) GpppG and 7-methyl (3'-deoxy) GpppG,” *RNA* 7:1486-1495。ARCA结构如下所示。



[0105] CleanCap™ AG (m7G (5') ppp (5') (2'OMeA) pG; TriLink Biotechnologies目录号N-7113) 或CleanCap™ GG (m7G (5') ppp (5') (2'OMeG) pG; TriLink Biotechnologies目录号N-7133) 可用于共转录地提供Cap1结构。CleanCap™ AG和CleanCap™ GG的3'-O-甲基化形式也可分别作为目录号N-7413和N-7433购自TriLink Biotechnologies。CleanCap™ AG结构如下所示。



[0107] 替代地,可在转录后将帽添加到RNA上。例如,牛痘病毒加帽酶是可商购获得的(New England Biolabs目录号M2080S),并且具有由其D1亚基提供的RNA三磷酸酶和鸟嘌呤基转移酶活性以及由其D12亚基提供的鸟嘌呤甲基转移酶。因此,在S-腺苷甲硫氨酸和GTP存在的情况下,它可以将7-甲基鸟嘌呤添加到RNA,从而得到Cap0。参见例如,Guo,P.和Moss,B. (1990) Proc.Natl.Acad.Sci.USA 87,4023-4027;Mao,X.和Shuman,S. (1994) J.Biol.Chem.269,24472-24479。

[0108] 在一些实施方案中,mRNA还包含聚腺苷酸化(poly-A)尾。在一些实施方案中,poly-A尾包含至少20、30、40、50、60、70、80、90或100个腺嘌呤,任选多至300个腺嘌呤。在一些实施方案中,poly-A尾包含95、96、97、98、99或100个腺嘌呤核苷酸。

[0109] IV.递送方法

[0110] 可使用本领域中可用的各种已知且合适的方法将本文公开的核酸构建体在体内或离体递送至宿主细胞或受试者。如本文所述,核酸构建体可与合适的基因编辑系统的组分(例如,RNA指导的DNA结合剂,诸如Cas核酸酶及其相应的指导RNA)一起递送。

[0111] 常规的基于病毒和非病毒的基因递送方法可用于在细胞(例如,哺乳动物细胞)和靶组织中引入本文公开的构建体和基因编辑系统的组分。如本文进一步提供的,非病毒载体递送系统包括核酸诸如非病毒载体、质粒载体以及例如与递送载体诸如脂质体、脂质纳米颗粒(LNP)或泊洛沙姆复合的核酸。病毒载体递送系统包括DNA和RNA病毒。

[0112] 用于非病毒递送核酸的方法和组合物包括电穿孔、脂质体转染、显微注射、基因枪(biolistics)、病毒体、脂质体、免疫脂质体、LNP、聚阳离子或脂质:核酸缀合物、裸露核酸(例如裸露DNA/RNA)、人工病毒体和DNA的剂增强的摄取。使用例如Sonitron 2000系统(Rich-Mar)的声致穿孔(sonoporation)也可用于递送核酸。

[0113] 另外的示例性核酸递送系统包括由AmaxaBiosystems(Cologne,Germany)、Maxcyte,Inc.(Rockville,Md.)、BTX Molecular Delivery Systems(Holliston,Ma.)和Copernicus Therapeutics Inc.提供的那些(参见例如美国专利号6,008,336)。脂质体转染描述于例如美国专利号5,049,386;4,946,787;和4,897,355)中,并且脂质体转染试剂在商业上销售(例如,TransfectamTM和LipofectinTM)。脂质:核酸复合物(包括靶向脂质体,诸如免疫脂质复合物)的制备是本领域众所周知的,并且如本文所述。

[0114] 也可将含有双向构建体和/或基因编辑组分(例如,指导RNA和Cas)的各种递送系统(例如,载体、脂质体、LNP)施用到生物体以体内递送至细胞或将其离体施用到细胞或细胞培养物。通过通常用于使分子与血液、液体或细胞最终接触的任何途径进行施用,所述途径包括但不限于注射、输注、局部应用和电穿孔。施用此类核酸的合适方法是可用的,并且是本领域技术人员众所周知的。

[0115] 在某些实施方案中,本公开提供了包含本文公开的双向核酸构建体的载体以递送至宿主细胞。在某些实施方案中,基因编辑系统的组分(例如,RNA指导的DNA结合剂和指导RNA)也作为载体的一部分递送至宿主细胞。在某些实施方案中,病毒载体可用于将双向核酸构建体、指导RNA和/或RNA指导的DNA结合剂中的任何一种或多种递送至宿主细胞。

[0116] 在一些实施方案中,本文提供了用于将本文公开的双向核酸构建体递送至宿主细胞或受试者的组合物和方法,其中构建体是如本文所述的载体系统的一部分。在一些实施方案中,载体系统包含另外的组分,诸如基因编辑系统的组分(例如,指导RNA和/或RNA指导

的DNA结合剂)。

[0117] 在一些实施方案中,提供了包含本文公开的双向核酸构建体的载体组合物。在一些实施方案中,组合物还包含基因编辑系统的组分(例如,指导RNA和/或RNA指导的DNA结合剂)。

[0118] 在一些实施方案中,载体可为环状。在其他实施方案中,载体可为线性的。在一些实施方案中,载体可通过脂质纳米颗粒、脂质体、非脂质纳米颗粒或病毒衣壳来递送。非限制性示例性载体包括质粒、噬菌粒、粘粒、人工染色体、微型染色体、转座子、病毒载体和表达载体。

[0119] 在一些实施方案中,载体系统可能能够驱动细胞中一种或多种核酸酶组分的表达。在一些实施方案中,任选地作为载体系统的一部分的双向构建体可包含能够驱动细胞中编码序列的表达的启动子。在一些实施方案中,细胞可为真核细胞,例如像酵母、植物、昆虫或哺乳动物细胞。在一些实施方案中,真核细胞可为哺乳动物细胞。在一些实施方案中,真核细胞可为啮齿动物细胞。在一些实施方案中,真核细胞可为人细胞。驱动不同类型的细胞中表达的合适启动子是本领域已知。在一些实施方案中,启动子可为野生型。在其他实施方案中,可修饰启动子以更有效或更有功效地表达。在其他实施方案中,启动子可被截短但仍保留其功能。例如,启动子可具有适合将载体适当包装成病毒的正常大小或减小的大小。在一些实施方案中,载体不包含驱动细胞中一个或多个编码序列的表达的启动子(例如,一旦插入靶内源基因座中,则编码序列的表达由内源启动子驱动)。

[0120] 在一些实施方案中,载体可为病毒载体。在一些实施方案中,病毒载体可从其野生型对应物进行遗传修饰。例如,病毒载体可包含一个或多个核苷酸的插入、缺失或取代以促进克隆或使得载体的一个或多个特性改变。此类特性可包括包装能力、转导效率、免疫原性、基因组整合、复制、转录和翻译。在一些实施方案中,可删除病毒基因组的一部分,使得病毒能够包装具有较大尺寸的外源序列。在一些实施方案中,病毒载体可具有增强的转导效率。在一些实施方案中,由病毒在宿主中诱导的免疫应答可降低。在一些实施方案中,可使促进病毒序列整合到宿主基因组中的病毒基因(例如像整合酶)突变,使得病毒变为非整合。在一些实施方案中,病毒载体可为复制缺陷的。在一些实施方案中,病毒载体可包含外源转录或翻译控制序列,以驱动编码序列在载体上的表达。在一些实施方案中,病毒可为辅助依赖型。例如,病毒可能需要一个或多个辅助病毒来提供将载体扩增并包装成病毒颗粒所需的病毒组分(例如像病毒蛋白)。在此种情况下,可将一个或多个辅助组分(包括编码病毒组分的一个或多个载体)连同本文所述的载体系统一起引入到宿主细胞中。在其他实施方案中,病毒可为无辅助的。例如,病毒可能能够在没有辅助病毒的情况下扩增并包装载体。在一些实施方案中,本文所述的载体系统还可编码病毒扩增和包装所需的病毒组分。

[0121] 使用基于RNA或DNA病毒的系统以递送核酸利用高度进化的过程将病毒靶向体内的特定细胞,并将病毒有效载荷运输到细胞核。病毒载体可直接施用于受试者(体内),或可将其用于体外治疗细胞。在一些实施方案中,将体外修饰的细胞施用于受试者(例如,作为来源于受试者或来自供体来源的细胞的离体操作)。非限制性示例性病毒载体包括腺相关病毒(AAV)载体、慢病毒载体、腺病毒载体、辅助依赖型腺病毒载体(HDAAd)、单纯疱疹病毒(HSV-1)载体、噬菌体T4、杆状病毒载体和逆转录病毒载体。在例如逆转录病毒、慢病毒和腺相关病毒基因转移方法的情况下,宿主基因组中的整合是可能的,通常导致插入的转基因

的长期表达。另外,已在许多不同的细胞类型和靶组织中观察到高转导效率。

[0122] 逆转录病毒的向性可通过并入外源包膜蛋白来改变,从而扩大靶细胞的潜在靶细胞群。慢病毒载体是能够转导或感染非分裂细胞的逆转录病毒载体,并且通常产生高病毒滴度。逆转录病毒基因转移系统的选择取决于靶组织。逆转录病毒载体由顺式作用的长末端重复序列组成,其中包装容量高达6kb至10kb的外源序列。最小顺式作用的LTR足以复制和包装载体,随后将其用于将包含转基因的双向构建体整合到靶细胞中以提供永久转基因表达。广泛使用的逆转录病毒载体包括基于以下的那些:鼠白血病病毒(MuLV)、长臂猿白血病病毒(GaLV)、猴免疫缺陷病毒(SIV)、人免疫缺陷病毒(HIV)及其组合(参见例如,Buchscher等人,J.Virol.66:2731-2739(1992);Johann等人,J.Virol.66:1635-1640(1992);Sommerfelt等人,Virol.176:58-59(1990);Wilson等人,J.Virol.63:2374-2378(1989);Miller等人,J.Virol.65:2220-2224(1991);PCT/US94/05700)。

[0123] 在一些实施方案中,可使用基于腺病毒的系统。基于腺病毒的载体能够在许多细胞类型中具有非常高的转导效率,并且不需要细胞分裂。在此类载体的情况下,获得了高滴度和高水平的表达。此载体可在相对简单的系统中大量产生。复制缺陷型重组腺病毒载体可以高滴度产生,并且容易感染许多不同的细胞类型。大多数腺病毒载体为工程改造的,使得转基因替换Ad E1a、E1b和/或E3基因;随后,复制缺陷载体在反式提供缺失的基因功能的人293细胞中繁殖。Ad载体可在体内转导多种类型的组织,包括非分裂的分化细胞,诸如肝脏、肾和肌肉中发现的细胞。常规的Ad载体具有较大的承载能力。在临床试验中使用Ad载体的实例涉及通过肌肉内注射进行抗肿瘤免疫的多核苷酸治疗(Sterman等人,Hum.Gene Ther.7:1083-9(1998))。在临床试验中使用腺病毒载体进行基因转移的另外的实例包括Rosenecker等人,Infection 24:15-10(1996);Sterman等人,Hum.Gene Ther.9:71083-1089(1998);Welsh等人,Hum.Gene Ther.2:205-18(1995);Alvarez等人,Hum.Gene Ther.5:597-613(1997);Topf等人,Gene Ther.5:507-513(1998);Sterman等人,Hum.Gene Ther.7:1083-1089(1998)。

[0124] 在一些实施方案中,腺相关病毒(AAV)载体用于递送本文提供的双向核酸构建体。AAV载体是众所周知的,并且已被用于用靶核酸转导细胞,例如在体外生产核酸和肽,以及用于体内和体外基因治疗程序(参见例如,West等人,Virology 160:38-47(1987);美国专利号4,797,368;WO 93/24641;Kotin,Human Gene Therapy 5:793-801(1994);Muzyczka,J.Clin.Invest.94:1351(1994)。重组AAV载体的构建描述于许多出版物中,包括美国专利号5,173,414;Tratschin等人,Mol.Cell.Biol.5:3251-3260(1985);Tratschin等人,Mol.Cell.Biol.4:2072-2081(1984);Hermonat&Muzyczka,PNAS81:6466-6470(1984);和Samulski等人,J.Virol.63:03822-3828(1989)。在一些实施方案中,病毒载体可为AAV载体。在一些实施方案中,AAV载体例如AAV1、AAV2、AAV3、AAV3B、AAV4、AAV5、AAV6、AAV6.2、AAV7、AAVrh.64R1、AAVhu.37、AAVrh.8、AAVrh.32.33、AAV8、AAV9、AAV-DJ、AAV2/8、AAVrh10或AAVLK03,以及任何新的AAV血清型也可根据本发明使用。AAV载体重组腺相关病毒载体是有前途的替代核酸递送系统,例如基于缺陷的和非致病性细小病毒腺相关2型病毒的那些核酸递送系统。

[0125] 如本文所用,“AAV”是指所有血清型、亚型和天然存在的AAV以及重组AAV。“AAV”可用于指病毒本身或其衍生物。术语“AAV”包括AAV1、AAV2、AAV3、AAV3B、AAV4、AAV5、AAV6、

AAV6.2、AAV7、AAVrh.64R1、AAVhu.37、AAVrh.8、AAVrh.32.33、AAV8、AAV9、AAV-DJ、AAV2/8、AAVrh10、AAVLK03、AV10、AAV11、AAV12、rh10及其杂交体、鸟类AAV、牛AAV、犬科AAV、马AAV、灵长类AAV、非灵长类AAV和绵羊AAV。AAV的各种血清型的基因组序列,以及天然末端重复序列(TR)、Rep蛋白和衣壳亚基的序列是本领域已知的。此类序列可在文献中或在诸如GenBank的公共数据库中找到。如本文所用,“AAV载体”是指包含非AAV来源的异源序列(即,与AAV异源的核酸序列)的AAV载体,其通常包含编码感兴趣的异源多肽的序列。构建体可包含AAV1、AAV2、AAV3、AAV3B、AAV4、AAV5、AAV6、AAV6.2、AAV7、AAVrh.64R1、AAVhu.37、AAVrh.8、AAVrh.32.33、AAV8、AAV9、AAV-DJ、AAV2/8、AAVrh10、AAVLK03、AV10、AAV11、AAV12、rh10及其杂交体、鸟类AAV、牛AAV、犬科AAV、马AAV、灵长类AAV、非灵长类AAV和绵羊AAV衣壳序列。一般来讲,异源核酸序列(转基因)侧接至少一个,并且通常侧接两个AAV反向末端重复序列(ITR)。AAV载体可为单链(ssAAV)或自互补的(scAAV)。

[0126] 在其他实施方案中,病毒载体可为慢病毒载体。在一些实施方案中,慢病毒可为非整合的。在一些实施方案中,病毒载体可为腺病毒载体。在一些实施方案中,腺病毒可为高克隆能力或“无肠”腺病毒,其中从病毒中删除除5'和3'反向末端重复序列(ITR)和包装信号('I')之外的所有编码病毒区域以增加其包装能力。在其他实施方案中,病毒载体可为HSV-1载体。在一些实施方案中,基于HSV-1的载体为辅助依赖型,而在其他实施方案中,它不依赖辅助。例如,仅保留包装序列的扩增子载体需要具有用于包装的结构组分的辅助病毒,而去除非必需病毒功能的删除30kb的HSV-1载体则不需要辅助病毒。在另外的实施方案中,病毒载体可为噬菌体T4。在一些实施方案中,当排空病毒头时,噬菌体T4可能能够包装任何线性或环状DNA或RNA分子。在其他实施方案中,病毒载体可为杆状病毒载体。在其他实施方案中,病毒载体可为逆转录病毒载体。在使用具有较小克隆能力的AAV或慢病毒载体的实施方案中,可能需要使用不止一个载体来递送如本文公开的载体系统的所有组分。例如,一个AAV载体可含有编码RNA指导的DNA结合剂诸如Cas蛋白(例如,Cas9)的序列,而第二AAV载体可含有一个或多个指导序列。

[0127] 包装细胞用于形成能够感染宿主细胞的病毒颗粒。此类细胞包括293细胞,其可包装腺病毒和AAV;以及 ϕ 2细胞或PA317细胞,其包装逆转录病毒。用于基因治疗的病毒载体通常是由将核酸载体包装成病毒颗粒的生产细胞系产生的。载体通常含有包装所需的最小病毒序列,其他病毒序列被编码待表达蛋白质的序列替换。缺失的病毒功能由包装细胞系反式提供。例如,用于基因治疗的AAV载体通常仅具有AAV基因组中包装所需的反向末端重复(ITR)序列。病毒DNA被包装在细胞系中,所述细胞系含有编码其他AAV基因的辅助质粒,即rep和cap,但缺乏ITR序列。细胞系也可能感染腺病毒作为辅助。辅助病毒促进AAV载体的复制和辅助质粒中AAV基因的表达。

[0128] 如下文所述,基因治疗载体可通过向个体患者施用而在体内递送,通常通过全身施用(例如,静脉内、腹腔内、肌肉内、皮下或颅内输注)或局部施用。替代地,载体可被离体递送至细胞,诸如从个体患者(例如,淋巴细胞、骨髓抽吸物、组织活检)移植的细胞或通用供体造血干细胞,之后通常在选择并入载体的细胞之后将细胞重新植入患者。

[0129] 在一些实施方案中,除了本文公开的双向核酸构建体之外,载体系统还可包含编码核酸酶的核酸。在一些实施方案中,除了本文公开的双向核酸构建体之外,载体系统还可包含编码指导RNA的核酸和/或编码RNA指导的DNA结合剂(其可为Cas蛋白诸如Cas9)的核

酸。在一些实施方案中,编码指导RNA的核酸和/或编码RNA指导的DNA结合剂或核酸酶的核酸各自或两者都在与包含本文公开的双向构建体的载体分开的载体上。在任何实施方案中,载体系统可包括其他序列,其包括但不限于如本文所述的启动子、增强子、调节序列。在一些实施方案中,载体系统内的启动子不驱动双向构建体的转基因的表达。在一些实施方案中,载体系统包含编码crRNA、trRNA或crRNA和trRNA的一个或多个核苷酸序列。在一些实施方案中,载体包含编码sgRNA的一个或多个核苷酸序列和编码RNA指导的DNA结合剂(其可为Cas核酸酶(例如,Cas9))的mRNA。在一些实施方案中,载体系统包含编码crRNA、trRNA的一个或多个核苷酸序列和编码RNA指导的DNA结合剂(其可为Cas核酸酶诸如Cas9)的mRNA。在一些实施方案中,Cas9来自酿脓链球菌(即Spy Cas9)。在一些实施方案中,编码crRNA、trRNA或crRNA和trRNA(其可为sgRNA)的核苷酸序列包含侧接来自天然存在的CRISPR/Cas系统的全部或部分重复序列的指导序列或由其组成。载体系统可包含含有crRNA、trRNA或crRNA和trRNA或由其组成的核酸,其中载体系统包含不与crRNA、trRNA或crRNA和trRNA一起天然存在的核酸或由其组成。本文所述的任何载体可通过脂质体、纳米颗粒、外来体、微囊泡和/或脂质纳米颗粒(LNP)来递送。一个或多个指导RNA、RNA结合的DNA结合剂(例如mRNA)或包含编码异源蛋白的序列的供体构建体可单独地或以任何组合通过脂质体、纳米颗粒、外来体或微囊泡来递送。一个或多个指导RNA、RNA结合的DNA结合剂(例如mRNA)或包含编码异源蛋白的序列的供体构建体可单独地或以任何组合通过LNP来递送。本文所述的LNP和LNP制剂中的任一种都适合于递送指导。

[0130] 脂质纳米颗粒(LNP)是众所周知的用于递送核苷酸和蛋白质货物的手段,并且可用于递送本文公开的双向核酸构建体。在一些实施方案中,LNP可用于递送基因编辑系统的组分。在一些实施方案中,LNP将核酸(例如,DNA或RNA)、蛋白质(例如,RNA指导的DNA结合剂)或核酸与蛋白质一起递送。

[0131] 在一些实施方案中,本文提供了一种用于将本文公开的双向核酸构建体递送至宿主细胞或受试者的方法,其中构建体通过LNP递送。在一些实施方案中,本文提供了一种用于将本文公开的双向核酸构建体递送至宿主细胞或受试者的方法,其中基因编辑系统的一个或多个组分(诸如CRISPR/Cas核酸酶系统)通过LNP递送。在一些实施方案中,LNP包含双向构建体和/或基因编辑系统的一个或多个组分(例如,指导RNA和/或RNA指导的DNA结合剂或编码RNA指导的DNA结合剂的mRNA)。

[0132] 在一些实施方案中,本文提供了一种包含本文公开的双向核酸构建体和LNP的组合物。在一些实施方案中,组合物还包含基因编辑系统的组分(例如,指导RNA和/或RNA指导的DNA结合剂,诸如Cas9或能够编码所述Cas9的载体系统)。在一些实施方案中,本文提供了一种包含本文公开的双向核酸构建体和包含指导RNA和/或编码RNA指导的DNA结合剂(诸如Cas9)的mRNA的LNP的组合物。

[0133] 在一些实施方案中,LNP包含可生物降解的可电离的脂质。在一些实施方案中,LNP包含十八碳-9,12-二烯酸(9Z,12Z)-3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙酯(也称为(9Z,12Z)-十八碳-9,12-二烯酸3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙酯)或另一种可电离的脂质。参见例如,PCT/US2018/053559(2018年9月28日提交)、W0/2017/173054、W02015/095340和W02014/136086以及本文提供的参考文献的脂质。在一些实施方案中,在

LNP脂质的上下文中,术语阳离子的和可电离的是可互换的,例如,其中可电离的脂质根据pH是阳离子的。

[0134] 电穿孔是众所周知的用于递送货物的手段,并且任何电穿孔方法可用于递送本文公开的双向构建体。在一些实施方案中,电穿孔可用于递送本文公开的双向构建体,任选地具有通过相同或不同手段递送的指导RNA和/或RNA指导的DNA结合剂(例如,Cas9)或编码RNA指导的DNA结合剂(例如,Cas9)的mRNA。

[0135] 在一些实施方案中,本公开包括一种用于将本文公开的双向构建体在体外递送至细胞的方法,其中双向构建体通过LNP递送。在一些实施方案中,双向构建体通过非LNP手段递送,诸如通过AAV系统,并且指导RNA和/或RNA指导的DNA结合剂(例如,Cas9)或编码RNA指导的DNA结合剂(例如,Cas9)的mRNA通过LNP递送。

[0136] 在一些实施方案中,将本文所述的双向构建体单独或作为载体的一部分在脂质纳米颗粒中配制或通过脂质纳米颗粒施用;参见例如,W0/2017/173054,其内容特此以引用的方式整体并入。

[0137] 本文所述的任何载体可通过LNP递送。本文所述的LNP和LNP制剂中的任一种适合于递送本文公开的gRNA、Cas核酸酶或编码Cas核酸酶的mRNA、其组合和/或双向构建体。在一些实施方案中,涵盖的LNP组合物包含:RNA组分和脂质组分,其中脂质组分包含胺脂质,诸如可生物降解的可电离的脂质;并且其中RNA组分包含指导RNA和/或编码Cas核酸酶的mRNA。

[0138] 在一些情况下,脂质组分包含可生物降解的可电离的脂质、胆固醇、DSPC和PEG-DMG。

[0139] 显而易见的是,基因编辑系统的组分(例如,指导RNA和/或RNA指导的DNA结合剂)和双向构建体可使用相同或不同的系统来递送。例如,指导RNA、RNA指导的DNA结合剂序列和双向构建体可由相同载体(例如AAV载体)携带或在一种或多种LNP组合物中配制。替代地,RNA指导的DNA结合剂(作为蛋白质或mRNA)和/或gRNA可由LNP携带或与其相关,而双向构建体可由载体携带,或反之亦然。此外,不同的递送系统可通过相同或不同的路径来施用。

[0140] 不同的递送系统可同时或以任何相继次序在体外或体内递送。在一些实施方案中,双向构建体、指导RNA和RNA指导的DNA结合剂可同时在体外或体内递送,例如在一个载体、两个载体、单独的载体、一个LNP、两个LNP、单独的LNP或其组合中。在一些实施方案中,双向构建体可在递送单独作为载体和/或与LNP缔合或一起作为核糖核蛋白(RNP)的指导RNA和/或RNA指导的DNA结合剂(例如,约1、2、3、4、5、6、7、8、9、10、11、12、13、14或更多天)之前作为载体和/或与LNP缔合在体内或在体外递送。在一些实施方案中,供体构建体可以多次施用递送,例如每天、每两天、每三天、每四天、每周、每两周、每三周或每四周。在一些实施方案中,供体构建体可以一周的间隔递送,例如,在第1周、第2周和第3周等。作为进一步的实例,指导RNA和/或RNA指导的DNA结合剂可在递送作为载体和/或与LNP缔合的双向构建体(例如,约1、2、3、4、5、6、7、8、9、10、11、12、13、14或更多天)之前单独作为载体和/或与LNP缔合或一起作为核糖核蛋白(RNP)在体内或在体外递送。在一些实施方案中,白蛋白指导RNA可以多次施用递送,例如每天、每两天、每三天、每四天、每周、每两周、每三周或每四周。在一些实施方案中,白蛋白指导RNA可以一周的间隔递送,例如,在第1周、第2周和第3周等。

在一些实施方案中,Cas核酸酶可以多次施用递送,例如,可以每天、每两天、每三天、每四天、每周、每两周、每三周或每四周递送。在一些实施方案中,Cas核酸酶可以一周的间隔递送,例如,在第1周、第2周和第3周等。

[0141] V. 使用方法

[0142] 本公开提供了在各种应用中使用本文所述的双向核酸构建体的方法。在一些实施方案中,在各种应用中使用本文所述的双向核酸构建体的方法包括使用如本文所述的诸如CRISPR/Cas系统的基因编辑系统。

[0143] 在一些实施方案中,本文提供了一种修饰靶基因座(例如,在基因座内的靶位点处插入转基因)的体外或体内方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体、如本文所述的指导RNA和RNA指导的DNA结合剂(例如,Cas核酸酶诸如Cas9)。在一些实施方案中,本文提供了一种修饰靶基因座的体外或体内方法,所述方法包括在宿主细胞中裂解靶序列,以及插入本文所述的双向核酸构建体,任选地利用如本文所述的指导RNA和RNA指导的DNA结合剂(例如,Cas核酸酶诸如Cas9)以用于裂解步骤。

[0144] 在一些实施方案中,本文提供了一种将构建体引入宿主细胞的体外或体内方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体、指导RNA和如本文所述的RNA指导的DNA结合剂(例如,Cas核酸酶诸如Cas9)。在一些实施方案中,本文提供了一种将构建体引入宿主细胞的体外或体内方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体和诸如ZFN、TALEN或CRISPR/Cas9系统的基因编辑系统。

[0145] 在一些实施方案中,本文提供了一种增加宿主细胞中多肽表达的体外或体内方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体、指导RNA和如本文所述的RNA指导的DNA结合剂(例如,Cas核酸酶诸如Cas9)。在一些实施方案中,本文提供了一种增加宿主细胞中多肽表达的体外或体内方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体和诸如ZFN、TALEN或CRISPR/Cas9系统的基因编辑系统。多肽可为细胞外的。

[0146] 双向构建体可通过诸如核酸载体的载体来施用。指导RNA和RNA指导的DNA结合剂可单独施用或以任何组合施用,例如通过包含指导RNA和编码RNA指导的DNA结合剂的mRNA的LNP来施用。向宿主细胞的施用和递送可通过本文所述的任何递送方法来实现。

[0147] 在一些实施方案中,本文提供了一种在靶基因座处表达由转基因编码的多肽的体外或体内方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体、如本文所述的指导RNA和RNA指导的DNA结合剂(例如,Cas核酸酶诸如Cas9)。在一些实施方案中,本文提供了一种在靶基因座处表达由转基因编码的多肽的体外或体内方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体和诸如ZFN、TALEN或CRISPR/Cas9系统的基因编辑系统。在一些实施方案中,本文提供了一种制备用于表达多肽的宿主细胞的方法,所述方法包括向宿主细胞施用或递送本文所述的双向核酸构建体和诸如ZFN、TALEN或CRISPR/Cas9系统的基因编辑系统。

[0148] 例如,如本文所述,双向构建体、指导RNA和RNA指导的DNA结合剂可单独施用或以任何组合施用。在一些实施方案中,双向构建体、指导RNA和RNA指导的DNA结合剂可同时或顺序地递送,例如在一个载体、两个载体、单独的载体、一个LNP、两个LNP、单独的LNP或其组合中。向宿主细胞的施用和递送可通过本文所述的任何递送方法来实现。

[0149] 另外,在一些实施方案中,方法涉及例如使用包含选自表5、表6、表7、表8、表9和表10中的任一个的序列的指导RNA插入白蛋白基因座(诸如白蛋白内含子1)中。在涉及插入白蛋白基因座中的某些实施方案中,个体的循环白蛋白水平正常。方法可包括将个体的循环白蛋白水平维持在正常循环白蛋白水平的 $\pm 5\%$ 、 $\pm 10\%$ 、 $\pm 15\%$ 、 $\pm 20\%$ 或 $\pm 50\%$ 内。在某些实施方案中,与未经治疗的个体的白蛋白水平相比,个体的白蛋白水平至少在第4周、第8周、第12周或第20周没有变化。在某些实施方案中,个体的白蛋白水平瞬时下降,然后恢复正常水平。特别地,方法可包括检测血浆白蛋白水平无显著改变。

[0150] 在一些实施方案中,本发明包括一种修饰白蛋白基因(诸如人白蛋白基因)(例如,在其中创建双链断裂)的方法或用途,所述方法或用途包括将本文所述的gRNA、供体构建体(例如,包含编码因子IX的序列的双向构建体)和RNA指导的DNA结合剂(例如,Cas核酸酶)中的任何一种或多种施用或递送至宿主细胞或宿主细胞群。在一些实施方案中,本发明包括一种修饰白蛋白内含子1区域(诸如人白蛋白内含子1)(例如,在其中创建双链断裂)的方法或用途,所述方法或用途包括将本文所述的gRNA、供体构建体(例如,包含编码异源多肽的核酸的双向构建体)和RNA指导的DNA结合剂(例如,Cas核酸酶或编码Cas核酸酶的核酸)中的任何一种或多种施用或递送至宿主细胞或宿主细胞群。在一些实施方案中,本发明包括一种修饰人安全港(诸如肝脏组织或肝细胞宿主细胞)(例如,在其中创建双链断裂)的方法或用途,所述方法或用途包括将本文所述的gRNA、供体构建体(例如,包含编码异源多肽的序列的双向构建体)和RNA指导的DNA结合剂(例如,Cas核酸酶或编码Cas核酸酶的核酸)中的任何一种或多种施用或递送至宿主细胞或宿主细胞群。

[0151] 如本文所述,转基因的插入和/或表达可在其同源基因座处(例如,将野生型转基因插入内源基因座中)或进入非同源基因座(例如,安全港基因座,诸如白蛋白)中。

[0152] 在一些实施方案中,宿主细胞为非分裂细胞类型。如本文所用,“非分裂细胞”是指终末分化而不分裂的细胞,以及不分裂但保留重新进入细胞分裂和增殖的能力的静止细胞。例如,肝脏细胞保留分裂的能力(例如,当受伤或切除时),但通常不分裂。在有丝分裂细胞分裂期间,同源重组是保护基因组并修复双链断裂的机制。在一些实施方案中,“非分裂”细胞是指这样的细胞,其中例如与对照分裂细胞相比,同源重组(HR)不是修复细胞中双链DNA断裂的主要机制。在一些实施方案中,“非分裂”细胞是指这样的细胞,其中例如与对照分裂细胞相比,非同源末端连接(NHEJ)是修复细胞中双链DNA断裂的主要机制。非分裂细胞类型已例如通过活性NHEJ双链DNA断裂修复机制在文献中描述。参见例如,Iyama, DNA Repair (Amst.) 2013, 12(8):620-636。在一些实施方案中,宿主细胞包括但不限于肝细胞、肌肉细胞或神经元细胞。在一些实施方案中,宿主细胞为肝细胞,诸如小鼠、食蟹猴或人肝细胞。在一些实施方案中,宿主细胞为肌细胞,诸如小鼠、食蟹猴或人肌细胞。在一些实施方案中,本文提供了一种上述宿主细胞,其包含本文公开的双向构建体。在一些实施方案中,宿主细胞表达由本文公开的双向构建体编码的转基因多肽。在一些实施方案中,本文提供了一种通过本文公开的方法制备的宿主细胞。在某些实施方案中,通过向宿主细胞施用或递送本文所述的双向核酸构建体和诸如ZFN、TALEN或CRISPR/Cas9系统的基因编辑系统来制备宿主细胞。

[0153] 还提供了一种从本文所述的双向构建体表达多肽的方法。类似地,包含本文所述的双向构建体的宿主细胞可表达由构建体编码的多肽。在一些实施方案中,多肽为分泌型

多肽。在一些实施方案中,多肽是其功能作为分泌型多肽正常实现(例如功能活性)的多肽。如本文所用,“分泌型多肽”是指由细胞分泌的蛋白质。在一些实施方案中,多肽为细胞内多肽。在一些实施方案中,多肽是其功能在细胞内正常实现(例如功能活性)的多肽。如本文所用,“细胞内多肽”是指不由细胞分泌的蛋白质,包括可溶性胞质多肽。在一些实施方案中,多肽为野生型多肽。在一些实施方案中,多肽为突变体多肽(例如野生型多肽的活动过度的突变体)。在一些实施方案中,多肽为肝脏蛋白。在一些实施方案中,多肽为非肝脏蛋白。在一些实施方案中,多肽包括但不限于因子IX及其变体。在一些实施方案中,肝脏多肽为例如治疗肝脏病症的多肽,所述病症诸如但不限于酪氨酸血症、威尔逊病(Wilson's disease)、泰-萨克斯病(Tay-Sachs disease)、高胆红素血症(Crigler-Najjar)、急性间歇性卟啉症、瓜氨酸血症1型、进行性肝内胆汁淤积症或枫糖尿病。

[0154] 在一些实施方案中,方法还包括达到持久效果,例如至少1个月、2个月、6个月、1年或2年效果。在一些实施方案中,方法还包括以持久和持续的方式达到治疗效果,例如至少1个月、2个月、6个月、1年或2年效果。在一些实施方案中,异源多肽活性和/或水平的水平稳定至少1个月、2个月、6个月、1年或更长时间。在一些实施方案中,达到多肽的稳态活性和/或水平至少7天、至少14天或至少28天。在另外的实施方案中,方法包括在单次给药双向构建体后维持异源多肽活性和/或蛋白质水平至少1、2、4或6个月,或至少1、2、3、4或5年。

[0155] 在一些实施方案中,相对于未施用包含转基因的构建体的宿主细胞对照所表达的水平,宿主细胞对多肽的表达(无论是体外还是体内)增加至少2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%或更多。在一些实施方案中,宿主细胞对多肽的表达(无论是体外还是体内)增加到已知正常水平(例如,健康受试者中的多肽水平)的至少2%,2%、3%、4%、5%、6%、7%、8%、9%、10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%或更多。在一些实施方案中,如在例如受试者的细胞、血浆和/或血清中确定的,宿主细胞对多肽的表达(无论是体外还是体内)增加到至少约1 μ g/ml、2 μ g/ml、3 μ g/ml、4 μ g/ml、5 μ g/ml、6 μ g/ml、7 μ g/ml、8 μ g/ml、9 μ g/ml、10 μ g/ml、15 μ g/ml、20 μ g/ml、25 μ g/ml、30 μ g/ml、35 μ g/ml、40 μ g/ml、45 μ g/ml、50 μ g/ml、55 μ g/ml、60 μ g/ml、65 μ g/ml、70 μ g/ml、75 μ g/ml、80 μ g/ml、85 μ g/ml、90 μ g/ml、95 μ g/ml、100 μ g/ml、120 μ g/ml、140 μ g/ml、160 μ g/ml、180 μ g/ml、200 μ g/ml、225 μ g/ml、250 μ g/ml、275 μ g/ml、300 μ g/ml、325 μ g/ml、350 μ g/ml、400 μ g/ml、450 μ g/ml、500 μ g/ml、550 μ g/ml、600 μ g/ml、650 μ g/ml、700 μ g/ml、750 μ g/ml、800 μ g/ml、850 μ g/ml、900 μ g/ml、1000 μ g/ml、1100 μ g/ml、1200 μ g/ml、1300 μ g/ml、1400 μ g/ml、1500 μ g/ml、1600 μ g/ml、1700 μ g/ml、1800 μ g/ml、1900 μ g/ml、2000 μ g/ml或更多。

[0156] 在一些实施方案中,本文提供了一种根据本文所述的方法治疗肝脏相关病症的方法。如本文所用,“肝脏相关疾病”是指直接导致肝脏组织损伤的病症、由肝脏组织损伤引起的病症和/或由肝脏缺陷引起的非肝脏器官或组织的病症。

[0157] 在一些实施方案中,双向构建体、指导RNA和RNA指导的DNA结合剂单独或以任何组合局部或全身施用,例如静脉内施用。在一些实施方案中,将双向构建体、指导RNA和RNA指导的DNA结合剂单独或以任何组合施用到肝循环中。

[0158] 在一些实施方案中,宿主或受试者为哺乳动物。在一些实施方案中,宿主或受试者为人。在一些实施方案中,宿主或受试者为灵长类。在一些实施方案中,宿主或受试者为啮

齿动物(例如,小鼠、大鼠)、奶牛、猪、猴、绵羊、狗、猫、鱼或家禽。

[0159] 此描述和示例性实施方案不应被视为限制性的。为了本说明书和所附权利要求的目的,除非另外指出,否则在所有情况下,就没有被术语“约”修饰来说,应将表示数量、百分比或比例的所有数字以及在说明书和权利要求中使用的其他数值理解为由所述术语修饰。因此,除非有相反的指示,否则以下说明书和所附权利要求中所阐述的数值参数为近似值,其可根据要寻求获得的所需性质而变化。在最低限度并且不试图限制对权利要求范围的等同原则的应用,每个数值参数均应当至少根据报道的有效数字的数值和通过应用普通四舍五入技术来解读。

[0160] 实施例

[0161] 提供以下实施例来例示某些公开的实施方案,并且不应以任何方式解释为限制本公开的范围。

[0162] 实施例1-材料和方法

[0163] 克隆和质粒制备

[0164] 合成由ITR侧接的双向插入构建体,并由商业供应商将其克隆到pUC57-Kan中。将所得构建体(P00147)用作其他载体的亲本克隆载体。其他插入构建体(不含ITR)也被商业合成并克隆到pUC57中。用BglIII限制酶(New England BioLabs,目录号R0144S)消化纯化的质粒,并将插入构建体克隆到亲本载体中。质粒在Stb13TM化学感受态大肠杆菌(Thermo Fisher,目录号C737303)中繁殖。

[0165] AAV生产

[0166] HEK293细胞中的三重转染用于用感兴趣的用于AAV8和AAVDJ生产的构建体包装基因组,并且通过碘克沙醇梯度超速离心法从裂解的细胞和培养基中纯化所得载体(参见例如,Lock等人.,Hum Gene Ther.2010年10月;21(10):1259-71)。在三重转染中使用的质粒含有具有感兴趣的构建体的基因组,在实施例中通过“PXXXX”号引用,也参见例如表11。将分离的AAV在储存缓冲液(含0.001%Pluronic F68的PBS)中透析。使用位于ITR区域内的引物/探针通过qPCR确定AAV滴度。

[0167] 核酸酶mRNA的体外转录(“IVT”)

[0168] 使用线性化质粒DNA模板和T7 RNA聚合酶通过体外转录产生含有N1-甲基假U的加帽和聚腺苷酸化酿脓链球菌(“Spy”)Cas9mRNA。一般来讲,通过在37°C下与XbaI孵育以完全消化,之后在65°C下对XbaI进行热灭活,将含有T7启动子和100nt聚(A/T)区的质粒DNA线性化。从酶和缓冲盐中纯化线性化质粒。将产生Cas9修饰的mRNA的IVT反应在以下条件下于37°C孵育4小时:50ng/μL线性化质粒;GTP、ATP、CTP和N1-甲基假UTP(Trilink)各自为2mM;10mM ARCA(Trilink);5U/μL T7 RNA聚合酶(NEB);1U/μL鼠RNA酶抑制剂(NEB);0.004U/μL无机大肠杆菌焦磷酸酶(NEB);和1x反应缓冲液。添加TURBO RNA酶(ThermoFisher)至最终浓度为0.01U/μL,并将反应再孵育30分钟以去除DNA模板。根据制造商的方案(ThermoFisher),使用MegaClear转录清洁试剂盒纯化Cas9 mRNA。替代地,使用LiCl沉淀、乙酸铵沉淀和乙酸钠沉淀或使用LiCl沉淀方法之后通过切向流过滤进一步纯化来纯化Cas9 mRNA。通过测量260nm处的吸光度(Nanodrop)来确定转录物浓度,并且通过借助Bioanalyzer(Agilent)的毛细管电泳来分析所述转录物。

[0169] 下面的Cas9 mRNA包含Cas9 ORF SEQ ID NO:703或SEQ ID NO:704或PCT/US2019/

053423 (其特此以引用的方式并入) 的表24的序列。

[0170] 用于递送Cas9 mRNA和gRNA的脂质制剂

[0171] 利用脂质制剂将Cas9 mRNA和gRNA递送至细胞和动物,所述脂质制剂包含可电离的脂质十八碳-9,12-二烯酸(9Z,12Z)-3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙酯(也称为(9Z,12Z)-十八碳-9,12-二烯酸3-((4,4-双(辛氧基)丁酰基)氧基)-2-(((3-(二乙基氨基)丙氧基)羰基)氧基)甲基)丙酯)、胆固醇、DSPC和PEG2k-DMG。

[0172] 对于利用预混合脂质制剂(在本文中称为“脂质包”)的实验,将组分以50:38:9:3的可电离脂质:胆固醇:DSPC:PEG2k-DMG的摩尔比在100%乙醇中重构,之后将其以约6.0的脂质胺与RNA磷酸(N:P)摩尔比与RNA货物(例如,Cas9 mRNA和gRNA)混合,如本文进一步所述。

[0173] 对于利用配制为脂质纳米颗粒(LNP)的组分的实验,将所述组分以各种摩尔比溶于100%乙醇中。将RNA货物(例如,Cas9 mRNA和gRNA)溶解在25mM柠檬酸盐、100mM NaCl,pH 5.0中,导致RNA货物的浓度为大约0.45mg/mL。

[0174] 对于实施例2中描述的实验,根据制造商的方案,使用Precision Nanosystems NanoAssemblr™ Benchtop仪器通过微流体混合脂质和RNA溶液来形成LNP。在混合期间,使用不同的流速保持2:1的水与有机溶剂的比率。混合后,收集LNP,在水(大约1:1体积/体积)中稀释,在室温下保持1小时,并且在最终缓冲液交换之前用水(大约1:1体积/体积)进一步稀释。用PD-10脱盐柱(GE)完成将最终缓冲液交换到50mM Tris、45mM NaCl、5%(重量/体积)蔗糖,pH 7.5(TSS)中。如果需要,通过用Amicon 100kDa离心过滤器(Millipore)浓缩制剂。然后使用0.2μm无菌过滤器过滤所得混合物。将最终LNP储存在-80℃下直至进一步使用。以45:44:9:2的可电离的脂质:胆固醇:DSPC:PEG2k-DMG的摩尔比配制LNP,其中脂质胺与RNA磷酸酯(N:P)摩尔比为约4.5,并且gRNA与mRNA的重量比为1:1。

[0175] 对于其他实施例中描述的实验,使用交叉流技术制备LNP,所述技术利用将乙醇中的脂质与两体积的RNA溶液和一体积的水冲击喷射混合。通过混合十字管将乙醇中的脂质与两体积的RNA溶液混合。将第四股水流通过内联三通与十字管的出口流混合(参见W02016010840,图2)。将LNP在室温下放置1小时,并且用水(大约1:1体积/体积)进一步稀释。使用切向流过滤在平板滤筒(Sartorius,100kD MWCO)上浓缩稀释的LNP,然后通过渗滤将缓冲液交换到50mM Tris、45mM NaCl、5%(重量/体积)蔗糖,pH 7.5(TSS)中。替代地,用PD-10脱盐柱(GE)完成将最终缓冲液交换到TSS中。如果需要,通过用Amicon 100kDa离心过滤器(Millipore)浓缩制剂。然后使用0.2μm无菌过滤器过滤所得混合物。将最终LNP储存在4℃或-80℃下直至进一步使用。以50:38:9:3的可电离的脂质:胆固醇:DSPC:PEG2k-DMG的摩尔比配制LNP,其中脂质胺与RNA磷酸酯(N:P)摩尔比为约6.0,并且gRNA与mRNA的重量比为1:1。

[0176] Cas9 mRNA、gRNA和插入构建体的细胞培养和体外递送

[0177] Hepa1-6细胞

[0178] 将Hepa 1-6细胞以10,000个细胞/孔的密度接种在96孔板中。24小时后,用LNP和AAV处理细胞。在处理之前,将培养基从孔中吸出。将LNP在DMEM+10%FBS培养基中稀释至4ng/ul,并在10%FBS(DMEM中)中进一步稀释至2ng/ul,并在37℃下孵育10min(最终浓度为

5%FBS)。AAV的靶MOI为 $1e6$ ，在DMEM+10%FBS培养基中稀释。将50 μ l上述稀释的LNP以2ng/ μ l添加到细胞中(递送总共100ng RNA货物)，之后添加50 μ l AAV。LNP和AAV的治疗间隔数分钟。细胞中培养基的总体积为100 μ l。在处理72小时和处理后30天，如下所述收集来自这些经处理的细胞的上清液以用于人FIX ELISA分析。

[0179] 原代肝细胞

[0180] 将原代小鼠肝细胞(PMH)、原代猕猴肝细胞(PCH)和原代人肝细胞(PHH)解冻，并重悬于含补充剂(ThermoFisher)的肝细胞解冻培养基中，之后进行离心。丢弃上清液，并将沉淀的细胞重悬于肝细胞平板培养基加补充包(ThermoFisher)中。对细胞计数，并将其以对于PHH为33,000个细胞/孔和对于PCH为50,000个细胞/孔和对于PMH为15,000个细胞/孔的密度接种于Bio-coat胶原蛋白I包被的96孔板上。使接种的细胞在37℃和5%CO₂气氛下的组织培养箱中沉降并粘附5小时。孵育后，检查细胞的单层形成，并用先前的肝细胞维持液洗涤三次，并在37℃下孵育。

[0181] 对于利用脂质包递送的实验，分别将Cas9 mRNA和gRNA在维持培养基中各自稀释至2mg/ml，并将各自的2.9 μ l添加到含有12.5 μ l的50mM柠檬酸盐、200mM氯化钠，pH 5和6.9 μ l的水的孔(96孔Eppendorf板)中。然后添加12.5 μ l脂质包制剂，之后添加12.5 μ l水和150 μ l TSS。使用肝细胞维持培养基将每个孔稀释至20ng/ μ l(相对于总RNA含量)，然后用6%的新鲜小鼠血清稀释至10ng/ μ l(相对于总RNA含量)。转染前从细胞中吸出培养基，并将40 μ l脂质包/RNA混合物添加到细胞中，之后以 $1e5$ 的MOI添加AAV(在维持培养基中稀释)。如本文所述，在处理72小时收集培养基以进行分析，并收获细胞以进行进一步的分析。

[0182] 萤光素酶测定

[0183] 对于涉及细胞培养基中NanoLuc检测的实验，将一体积的**Nano-Glo®**萤光素酶测定底物与50体积的**Nano-Glo®**萤光素酶测定缓冲液组合。使用1:10稀释的样品(50 μ l试剂+40 μ l水+10 μ l细胞培养基)在Promega Glomax运行器上以0.5秒的积分时间进行测定。

[0184] 对于涉及检测细胞培养基中HiBit标签的实验，将LgBiT蛋白和Nano-GloR HiBiT细胞外底物分别在室温下的Nano-GloR HiBiT细胞外缓冲液中以1:100和1:50稀释。使用1:10稀释的样品(50 μ l试剂+40 μ l水+10 μ l细胞培养基)在Promega Glomax运行器上以1.0秒的积分时间进行测定。

[0185] LNP和/或AAV的体内递送

[0186] 通过侧尾静脉向小鼠给药AAV、LNP、AAV和LNP、或载体(对于AAV载体为PBS+0.001%Pluronic，对于LNP载体为TSS)。以每只动物0.1mL的体积施用AAV，其量(载体基因组/只小鼠，“vg/ms”)如本文所述。将LNP在TSS中稀释，并以本文指出的量以约5 μ l/克体重施用。通常，如果适用，首先给小鼠注射AAV，然后注射LNP。在治疗后的各个时间点，收集血清和/或肝脏组织用于某些分析，如下文进一步所述。

[0187] 人因子IX(hFIX)ELISA分析

[0188] 对于体外研究，根据制造商的方案使用人因子IX ELISA试剂盒(Abcam，目录号ab188393)确定细胞培养基中分泌的总人因子IX水平。使用4参数对数拟合将分泌的hFIX水平从标准曲线上定量，并表示为ng/mL培养基。

[0189] 对于体内研究，如所示收集血液并分离血清。根据制造商的方案使用人因子IX ELISA试剂盒(Abcam，目录号ab188393)确定总人因子IX血清水平。使用4参数对数拟合将血

清hFIX水平从标准曲线上定量,并表示为 $\mu\text{g/mL}$ 血清。

[0190] 下一代测序(“NGS”)和中靶裂解效率分析

[0191] 利用深度测序来鉴定由基因编辑例如在白蛋白内含子1内引入的插入和缺失的存在。在靶位点周围设计PCR引物,并扩增感兴趣的基因组区域。引物序列设计是按照本领域的标准进行的。

[0192] 根据制造商的方案(Illumina)进行另外的PCR,以添加用于测序的化学物质。在Illumina MiSeq仪器上对扩增子进行测序。在消除具有低质量得分的那些读段之后,将读段与参考基因组比对。将含有读段的所得文件映射到参考基因组(BAM文件),其中选择与感兴趣的靶区域重叠的读段,并对野生型读段的数量与含有插入或缺失(插入缺失)的读段的数量进行计算。

[0193] 编辑百分比(例如,“编辑效率”或“百分比编辑”)被定义为相对于包括野生型的序列读段总数的具有插入或缺失(“插入缺失”)的序列读段总数。

[0194] 原位杂交分析

[0195] BaseScope (ACDbio, Newark, CA) 是专门的RNA原位杂交技术,其可例如在含有插入转基因(hFIX)和来自插入位点(白蛋白的外显子1)的编码序列的杂交mRNA转录物中提供外显子连接的特异性检测。BaseScope用于测量表达杂交mRNA的肝脏细胞的百分比。

[0196] 为了检测杂交mRNA,ACDbio (Newark, CA) 设计了针对可能在插入双向构建体后出现的杂交mRNA的两种探针。一种探针被设计来检测由于在一个方向上插入构建体而产生的杂交mRNA,而另一种探针被设计来检测由于在另一方向上插入构建体而产生的杂交mRNA。收集不同组小鼠的肝脏并对其进行新鲜冷冻切片。根据制造商的方案,使用单个探针或合并的探针进行BaseScope测定。通过HALO软件对载片进行扫描和分析。此测定的背景(盐水处理的组)为0.58%。

[0197] 实施例2-具有和不具有同源臂的插入模板的体外测试

[0198] 在本实施例中,在存在或不存在LNP递送Cas9 mRNA和G000551的情况下,例如,如实施例1(n=3)中所述,培养Hepa1-6细胞并用携带各种形式的插入模板(例如,具有单链基因组(“ssAAV”)或自身互补基因组(“scAAV”)的AAV处理。如实施例1中所述制备AAV和LNP。处理后,如实施例1中所述收集培养基以用于转基因表达(例如,人因子IX水平)。

[0199] Hepa1-6细胞是在培养中继续分裂的永生化小鼠肝脏细胞系。如图2所示(治疗后72小时的时间点),仅包含200bp同源臂的载体(衍生自质粒P00204的scAAV)导致可检测到的hFIX表达。在本实验中使用衍生自P00123(缺乏同源臂的scAAV)和P00147(缺乏同源臂的ssAAV双向构建体)的AAV载体不会导致任何可检测到的hFIX表达。将细胞保持在培养中,并且在处理后30天重新测定证实这些结果(数据未示出)。

[0200] 实施例3-具有和不具有同源臂的插入模板的体内测试

[0201] 在本实施例中,用衍生自与实施例2中体外测试相同的质粒(P00123、P00204和P00147)的AAV处理小鼠。如实施例1中所述制备给药材料并给药。向C57B1/6小鼠(每组n=5)各自给药 3×10^{11} 个载体基因组(vg/ms),之后以4mg/kg的剂量(相对于总RNA货物含量)给药包含G000551(“G551”)的LNP。给药后四周,对动物实施安乐死,并收集肝脏组织和血清分别用于编辑和转基因(例如hFIX)表达。

[0202] 如图3A和表12所示,在用包含靶向鼠白蛋白内含子1的gRNA的LNP治疗的每组动物

中,检测到肝脏编辑水平为约60%。然而,尽管每个治疗组中的编辑水平稳健且一致,但接受无同源臂的ssAAV载体(衍生自P00147的ssAAV载体)与LNP治疗组合的动物导致在血清中的hFIX表达水平最高(图3B和表13)。

[0203] 表12:%插入缺失

[0204]	模板	平均插入缺失(%)	标准差插入缺失(%)
	scAAV平端(P00123)	66.72	4.09
	ssAAV平端(P00147)	68.10	2.27
	ssAAV HR(P00204)	70.16	3.68
	仅LNP	68.24	6.47
	载体	0.28	0.08

[0205] 表13:因子IX水平

	模板	平均因子 IX (ug/mL)	标准差因子 IX (ug/mL)
	scAAV 平端(P00123)	0.75	0.28
[0206]	ssAAV 平端(P00147)	2.92	1.04
	ssAAV HR (P00204)	0.96	0.35
	仅 LNP	0	0
	载体	0	0

[0207] 实施例4-具有和不具有同源臂的ssAAV插入模板的体内测试

[0208] 在本实施例中描述的实验检查在体内将同源臂并入ssAAV载体的效果。

[0209] 制备在本实验中使用的给药材料并如实施例1中所述给药。向C57B1/6小鼠(每组n=5)给药 3×10^{11} vg/ms,之后以0.5mg/kg的剂量(相对于总RNA货物含量)给药包含G000666(“G666”)或G000551(“G551”)的LNP。给药后四周,收集动物血清用于转基因(例如hFIX)表达。

[0210] 如图4A和表14所示,使用用于插入G551靶向的位点中的具有不对称同源臂的ssAAV载体(对于衍生自质粒P00350、P00356和P00362的载体分别为300/600bp臂、300/2000bp臂和300/1500bp臂)导致循环hFIX水平低于测定的检测下限。然而,使用不具有同源臂且在双向方向上具有两个hFIX开放阅读框(ORF)的ssAAV载体(衍生自P00147)导致每只动物中可检测到的hFIX循环水平。

[0211] 类似地,与使用不具有同源臂的双向载体(衍生自P00147)相比,使用用于插入G666靶向的位点中的具有对称同源臂的ssAAV载体(对于衍生自质粒P00353和P00354的载体分别为500bp臂和800bp臂)导致较低但可检测到的水平(参见图4B和表15)。

[0212] 表14:血清FIX水平

	AAV	平均血清 FIX (ug/mL)	标准差血清 FIX (ug/mL)
[0213]	P00147	5.13	1.31
	P00350	-0.22	0.08
	P00356	-0.23	0.04
	P00362	-0.09	0.16

[0214] 表15:血清FIX水平

[0215]	AAV	平均血清 FIX (ug/mL)	标准差血清 FIX (ug/mL)
	P00147	7.72	4.67
[0216]	P00353	0.20	0.23
	P00354	0.46	0.26

[0217] 实施例5-原代小鼠肝细胞中跨靶位点的双向构建体的体外筛选

[0218] 在证明缺乏同源臂的双向构建体表现优于具有其他构型的载体后,本实施例中描述的实验检查了改变双向构建体的模块(此处为ORF和剪接受体)以及改变靶向CRISPR/Cas9介导的插入的gRNA的效果。利用原代小鼠肝细胞(PMH)中靶向鼠白蛋白内含子1的20种不同gRNA,在一组靶位点上测试这些不同的双向构建体。

[0219] 如实施例1中所述制备本实施例中测试的ssAAV和脂质包递送材料并将其递送至PMH,其中AAV的MOI为 $1e5$ 。处理后,收集分离的基因组DNA和细胞培养基分别用于编辑和转基因表达分析。每个载体包含可通过如实施例1中所述的基于萤光素酶的荧光检测来测量的报告基因,在图5C中绘制为相对萤光素酶单位(“RLU”)。例如,包含hFIX ORF的AAV载体含有在其3'端融合的HiBit肽,并且仅包含报告基因的AAV载体包含NanoLuc ORF(除了GFP之外)。图5A提供了所测试的载体中的每个的示意图。测试的gRNA显示在图5B和图5C中,对于表4中列出的那些使用缩短的数字(例如,其中省略前导零,例如其中“G551”对应于表4中的“G000551”)。

[0220] 如图5B和表16所示,在每个测试的组合中,对于每个治疗组,检测到一致但不同水平的编辑。使用模板和指导RNA的各种组合进行的转基因表达在图5C和表17中示出。如图5D所示,显著水平的插入缺失形成不一定导致更有效的转基因表达。使用P00411和P00418衍生的模板,当不包括编辑低于10%的指导时, R^2 值分别为0.54和0.37。小鼠白蛋白剪接受体和人FIX剪接受体各自导致有效的转基因表达。有趣的是,尽管ORF和剪接受体不同,但RLU中测量的相对表达水平在所测试的三个载体之间是一致的,这表明双向构建体系统的鲁棒性、再现性和模块性(参见图5C)。

[0221] 表16: %插入缺失

[0222]

指导编号	P00411		P00418		P00415	
	平均插入缺失(%)	标准差插入缺失(%)	平均插入缺失(%)	标准差插入缺失(%)	平均插入缺失(%)	标准差插入缺失(%)
G000551	67.4	1.42	70.67	2.29	66.73	4.90
G000552	90.93	0.15	91.10	2.43	90.37	1.01
G000553	77.80	3.83	77.47	1.87	80.50	0.85
G000554	72.37	6.49	70.53	3.16	70.60	2.91
G000555	35.37	2.63	35.77	9.34	40.47	4.75
G000666	62.47	3.87	50.90	19.41	65.90	3.99
G000667	30.57	2.73	25.30	3.67	31.67	2.29
G000668	63.60	2.02	66.65	4.60	68.30	4.90
G000669	19.10	2.51	19.33	1.53	18.70	1.25
G000670	47.80	3.27	49.10	4.42	51.97	2.06
G011722	4.20	0.72	4.27	1.20	4.20	0.26
G011723	5.63	1.27	6.07	0.15	5.93	0.15
G011724	6.10	1.28	8.50	2.69	7.13	1.27
G011725	1.93	0.29	2.60	0.79	2.53	0.65
G011726	10.73	1.46	11.70	0.50	12.43	1.33
G011727	14.20	1.56	14.80	2.36	16.20	2.69
G011728	10.55	1.20	13.65	0.92	15.50	1.56
G011729	5.00	0.10	5.63	0.25	6.00	1.01
G011730	7.83	0.97	9.13	0.59	7.33	0.59
G011731	23.70	0.66	25.27	1.21	24.87	1.01
仅 AAV	0.15	0.07	0.05	0.07	0.10	0.00

[0223]

表17: 萤光素酶水平

[0224]

指导编号	P00411		P00418		P00415	
	平均萤光素酶(RLU)	标准差萤光素酶(RLU)	平均萤光素酶(RLU)	标准差萤光素酶(RLU)	平均萤光素酶(RLU)	标准差萤光素酶(RLU)
G000551	58000.00	4331.28	41800.00	2165.64	78633.33	20274.70
G000552	95700.00	10573.08	80866.67	27911.35	205333.33	30664.86
G000553	205333.33	52993.71	177333.33	32929.22	471666.67	134001.00
G000554	125333.33	55949.38	91933.33	19194.10	232666.67	67002.49
G000555	59933.33	11566.04	77733.33	11061.80	155666.67	15947.83
G000666	88500.00	28735.87	93266.67	30861.19	313000.00	15394.80
G000667	75333.33	22653.11	68966.67	27222.11	153000.00	30805.84
G000668	164000.00	56320.51	133400.00	65111.29	429000.00	120751.80
G000669	28933.33	11636.29	22033.33	2413.16	46466.67	6543.19
G000670	162666.67	32959.57	200000.00	33867.39	424666.67	36473.73
G011722	16766.67	3384.28	8583.33	4103.10	24000.00	8915.16
G011723	22733.33	7252.82	17133.33	4905.44	26100.00	8109.87

[0225]	G011724	17300.00	2400.00	28033.33	9091.94	30933.33	3365.02
	G011725	8253.33	1163.20	8890.00	1429.27	20366.67	13955.05
	G011726	12223.33	3742.54	11610.00	2490.44	14950.00	8176.03
	G011727	35600.00	8128.35	36300.00	12301.22	86700.00	5023.94
	G011728	14900.00	5011.99	22466.67	7130.45	38166.67	13829.08
	G011729	10460.00	2543.95	11223.33	2220.28	26966.67	16085.50
	G011730	14833.33	2307.24	21700.00	8681.59	41233.33	25687.03
	G011731	16433.33	3274.65	22566.67	2205.30	20756.67	13096.20
	仅 AAV	217.00	15.56	215.00	15.56	207.00	1.41

[0226] 实施例6-跨靶位点的双向构建体的体内筛选

[0227] 如实施例1所述制备本实施例中测试的ssAAV和LNP并将其递送至C57B1/6小鼠,以评估体内跨靶位点的双向构建体的性能。给药后四周,对动物实施安乐死,并收集肝脏组织和血清分别用于编辑和转基因(例如hFIX)表达。

[0228] 在最初的实验中,将含有10种不同的靶向白蛋白内含子1的gRNA的10种不同的LNP制剂与衍生自P00147的ssAAV一起递送到小鼠。分别以 3×10^{11} vg/ms和4mg/kg(相对于总RNA货物含量)递送AAV和LNP。本实验中测试的gRNA示于图6和表18中。如图6所示和如体外所观察到的,显著水平的插入缺失形成不能预测转基因的插入或表达。

[0229] 在单独的实验中,在体内测试了实施例5中体外测试的靶向20个不同靶位点的全组20个gRNA。为此,将含有20个靶向白蛋白内含子1的gRNA的20种LNP制剂与衍生自P00147的ssAAV一起递送到小鼠。分别以 3×10^{11} vg/ms和1mg/kg(相对于总RNA货物含量)递送AAV和LNP。表4中列出的本实验中测试的gRNA均使用缩短的数字显示在图7A和图7B以及表19和表20中。

[0230] 如图7A所示,在每个测试的LNP/载体组合中,对于每个治疗组,检测到不同水平的编辑。然而,如图7B所示,并且与实施例5中所述的体外数据一致,较高水平的编辑不一定导致转基因在体内较高水平的表达,这指示编辑和插入/表达双向构建体之间缺乏相关性。实际上,如在图7D中提供的图中所看到的,实现的编辑量与转基因(hFIX)表达量之间几乎没有相关性。特别地,当从分析中去除实现少于10%编辑的那些gRNA时,在本实验的编辑和表达数据集之间计算出 R^2 值仅为0.34。有趣的是,如图7C所示,提供相关图,将实施例5的体外实验以RLU为单位测得的表达水平与本实验中检测到的体内转基因表达水平进行比较, R^2 值为0.70,这表明原代细胞筛选与体内治疗之间正相关。

[0231] 为了评估双向构建体在细胞水平上的插入,使用原位杂交方法(BaseScope)对来自治疗动物的肝脏组织进行测定,例如,如实施例1中所述。此测定利用了可检测作为杂交转录物的hFIX转基因与小鼠白蛋白外显子1序列之间的连接的探针。如图8A所示,在同时接受AAV和LNP的动物中检测到对杂交转录物呈阳性的细胞。特别地,当施用单独的AAV时,不到1.0%的细胞对杂交转录物呈阳性。在施用包含G011723、G000551或G000666的LNP的情况下,4.9%、19.8%或52.3%的细胞对杂交转录物呈阳性。另外,如图8B和表14所示,循环hFIX水平与对杂交转录物呈阳性的细胞数量相关。最后,所述测定利用可检测双向构建体在任一方向上的插入的合并的探针。然而,当使用仅检测单个方向的单个探针时,对杂交转录物呈阳性的细胞数量为约使用合并探针检测到的细胞数量的一半(在一个实施例中,为4.46%对9.68%),这表明双向构建体确实能够以任一方向插入,从而产生与蛋白质水平上

的转基因表达量相关的表达的杂交转录物。

[0232] 表18:因子IX水平和%插入缺失

指导	平均插入缺失 (%)	标准差插入缺 失(%)	平均萤光素酶 (RLU)	标准差萤光素 酶(RLU)
G000551	75.02	1.27	3.82	3.38
G000555	51.18	1.19	32.56	9.05
G000553	62.78	2.64	25.07	4.04
G000667	52.96	4.96	32.03	6.74
G000554	55.24	2.28	29.48	7.34
G000552	67.56	1.73	14.79	5.34
G000668	43.14	5.78	26.72	7.97
G000669	50.68	2.97	10.70	4.43

G000666	64.62	1.34	26.19	5.56
G000670	55.90	1.30	30.96	8.44

[0235] 表19:%肝脏编辑

指导	平均肝脏编辑 (%)	标准差肝脏编辑 (%)
G000551	59.48	4.02
G000555	58.72	3.65
G000553	51.26	2.81
G000554	33.04	8.76
G000555	12.72	4.46
G000666	53.60	4.92
G000667	26.74	4.98
G000668	39.22	3.04
G000669	33.34	4.77
G000670	47.50	5.58
G011722	10.34	1.68
G011723	4.02	0.84
G011724	2.46	0.64
G011725	8.26	1.24
G011726	6.90	1.01
G011727	13.33	6.43
G011728	35.78	9.34
G011729	4.62	1.46
G011730	12.68	3.14
G011731	26.70	1.86

[0237] 表20:FIX水平

指导	第 1 周		第 2 周		第 4 周		
	平均 FIX (ug/mL)	标准差 FIX (ug/mL)	平均 FIX (ug/mL)	标准差 FIX (ug/mL)	平均 FIX (ug/mL)	标准差 FIX (ug/mL)	
[0238]	G000551	10.88	2.74	10.25	2.51	9.39	3.48
	G000555	13.34	2.09	12.00	2.75	12.43	2.57
	G000553	17.64	4.34	20.27	6.35	15.31	2.43
	G000554	12.79	4.99	14.29	6.09	12.74	4.93
	G000555	11.94	5.79	11.99	5.76	8.61	4.02
	G000666	21.63	1.32	20.65	1.55	17.23	0.62
	G000667	16.77	2.86	12.35	2.85	12.57	5.60
	G000668	21.35	1.51	18.20	3.18	17.72	2.25
	G000669	5.76	2.10	6.72	2.93	3.39	0.78
	G000670	18.18	2.17	19.16	3.05	15.49	3.61
	G011722	8.07	1.74	7.74	2.41	8.07	1.74
	G011723	2.11	0.28	1.65	0.28	2.11	0.28
	G011724	0.92	0.43	0.60	0.30	0.92	0.43
[0239]	G011725	1.75	0.77	1.14	0.67	1.75	0.77
	G011726	0.59	0.30	1.01	0.64	0.59	0.30
	G011727	6.71	2.80	6.90	3.68	6.71	2.80
	G011728	11.77	3.12	12.29	3.43	11.77	3.12
	G011729	0.94	0.35	0.89	0.29	0.94	0.35
	G011730	5.93	1.77	6.33	1.73	5.93	1.77
	G011731	3.56	0.87	3.78	0.50	3.56	0.87
	仅 AAV	0.00	0.00	0.00	0.00	0.00	0.00
	载体	0.00	0.00	0.00	0.00	0.00	0.00
	人血清	3.63	0.32	3.61	0.35	3.28	0.03

[0240] 实施例7-hFIX在体内表达的持久性

[0241] 在本实施例中评估了hFIX表达在经治疗的动物中随时间的持久性。为此,作为一年持久性研究的一部分,在给药后在治疗的动物血清中测量hFIX。

[0242] 如实施例1中所述制备本实施例中测试的ssAAV和LNP并将其递送至C57B1/6小鼠。LNP制剂含有G000551,并且ssAAV衍生自P00147。以 3×10^{11} vg/ms递送AAV,并且以0.25mg/kg或1.0mg/kg (相对于总RNA货物含量) 递送LNP (每组n=5)。

[0243] 如图9A和图9B以及表21和表22所示,两组分别在41周或52周的每个评估时间点,hFIX表达都得以维持。在图9A中,在8周时观察到的水平下降被认为是由于ELISA分析的可变性。在第2周和第41周通过ELISA测量血清白蛋白水平,这表明循环白蛋白水平在整个研究中保持不变。

[0244] 表21:hFIX水平

星期	剂量			
	0.25 mpk LNP		1 mpk LNP	
	平均 hFIX (ug/mL)	标准差 hFIX (ug/mL)	平均 hFIX (ug/mL)	标准差 hFIX (ug/mL)
[0245] 2	0.48	0.21	2.24	1.12
4	0.55	0.18	2.82	1.67
8	0.40	0.17	1.72	0.77
12	0.48	0.20	2.85	1.34
20	0.48	0.27	2.45	1.26
[0246] 41	0.79	0.49	4.63	0.95

[0247] 表22:hFIX水平

星期	剂量			
	0.25 mpk LNP		1 mpk LNP	
	平均 hFIX (ug/mL)	标准差 hFIX (ug/mL)	平均 hFIX (ug/mL)	标准差 hFIX (ug/mL)
[0248] 2	0.87	0.15	4.02	1.75
8	0.99	0.15	4.11	1.41
12	0.93	0.14	4.15	1.35
20	0.83	0.22	4.27	1.54
41	0.83	0.37	4.76	1.62
52	0.82	0.25	4.72	1.54

[0249] 实施例8-不同剂量的AAV和LNP在体内调节hFIX表达的效果

[0250] 在本实施例中,在C57B1/6小鼠中评估了改变AAV和LNP的剂量以调节hFIX表达的效果。

[0251] 如实施例1中所述制备本实施例中测试的ssAAV和LNP并将其递送至小鼠。LNP制剂含有G000553,并且ssAAV衍生自P00147。以1e11 vg/ms、3e11 vg/ms、1e12 vg/ms或3e12 vg/ms递送AAV,并且以0.1mg/kg、0.3mg/kg或1.0mg/kg (相对于总RNA货物含量)递送LNP (每组n=5)。给药后两周,对动物实施安乐死。在两个时间点收集血清用于hFIX表达分析。

[0252] 如图10A (1周)、图10B (2周) 和表23所示,改变AAV或LNP的剂量可调节体内hFIX表达的量。

[0253] 表23:血清hFIX

时间点	RNP 剂量 (mg/kg)	AAV 剂量 (MOI)	平均值 FIX (ng/ml)	SD	N
[0254] 第 1 周	0.1	1E+11	0.08	0.02	2
		3E+11	0.11	0.04	5
		1E+12	0.41	0.15	5
		3E+12	0.61	0.17	5

时间点	RNP 剂量 (mg/kg)	AAV 剂量 (MOI)	平均值 FIX (ng/ml)	SD	N
	0.3	1E+11	0.36	0.14	5
		3E+11	0.67	0.26	5
		1E+12	1.76	0.14	5
		3E+12	4.70	2.40	5
	1.0	1E+11	3.71	0.31	4
		3E+11	8.00	0.51	5
		1E+12	14.17	1.38	5
		3E+12	20.70	2.79	5
	人血清 1:1000		6.62	-	1
	第 2 周	0.1	1E+11	0.12	0.01
3E+11			0.26	0.07	5
1E+12			0.83	0.24	5
3E+12			1.48	0.35	5
0.3		1E+11	0.70	0.26	4
		3E+11	1.42	0.37	5
		1E+12	3.53	0.49	5
		3E+12	8.94	4.39	5
1.0		1E+11	5.40	0.47	4
		3E+11	12.31	2.45	5
		1E+12	17.89	1.95	5
		3E+12	25.52	3.62	5
人血清 1:1000		4.47	-	1	

[0256] 实施例9-原代食蟹猴和原代人肝细胞中跨靶位点的双向构建体的体外筛选

[0257] 在本实施例中,利用分别靶向原代cyno (PCH) 和原代人肝细胞 (PHH) 中食蟹猴 (“cyno”) 和人白蛋白的内含子1的gRNA,在一组靶位点上测试包含双向构建体的ssAAV载体。

[0258] 如实施例1中所述制备本实施例中测试的ssAAV和脂质包递送材料并将其递送至PCH和PHH。处理后,收集分离的基因组DNA和细胞培养基分别用于编辑和转基因表达分析。每个载体包含可通过如实施例1中所述的基于萤光素酶的荧光检测来测量的报告基因(衍生自P00415),在图11B和图12B中绘制为相对萤光素酶单位(“RLU”)。例如,AAV载体含有NanoLuc ORF (除了GFP)。图11B和图12B中提供了所测试的载体的示意图表1和表7中列出的测试的gRNA均使用缩短的数字显示在每个图中。

[0259] 如针对PCH的图11A和针对PHH的图12A所示,对于所测试的每种组合,检测到不同水平的编辑(由于用于基于扩增子的测序的某些引物对的失败,在PCH实验中测试的一些组合的编辑数据未在图11A和表1中报告)。以图形显示在图11A和图12A中的编辑数据在下表1和表2中以数字然而,如图11B、图11C以及图12B和图12C所示,显著水平的插入缺失形成不能预测转基因的插入或表达,这指示编辑和分别在PCH和PHH中双向构建体的插入/表达之间几乎没有相关性作为一个量度,在图11C中计算出的 R^2 值为0.13,并且图12D的 R^2 值为0.22。

[0260] 表1:递送至原代食蟹猴肝细胞的sgRNA的白蛋白内含子1编辑和转基因表达数据

指导编号	平均%编辑	标准差%编辑	平均 RLU	标准差 RLU
G009867	25.05	0.21	10650.67	1455.97
G009866	18.7	3.96	75556.67	12182.98
G009876	14.85	4.88	27463.33	10833.53
G009875	12.85	2.33	51660.00	6362.36
G009874	28.25	6.01	270433.30	133734.10
G009873	42.65	5.59	178600.00	87607.25
G009865	59.15	0.21	301666.70	18610.03
G009872	48.15	3.46	320233.30	63517.43
G009871	46.5	5.23	211966.70	65852.44
G009864	33.2	8.34	210033.30	61201.33
G009863	54.8	12.45	69853.33	15216.92
G009862	44.6	7.21	508666.70	119876.30
G009861	28.65	0.21	178666.70	15821.93
G009860	33.2	7.07	571333.30	52728.87
G009859	0.05	0.07	258333.30	79052.73
G009858	14.65	1.77	402333.30	25579.94
G009857	23	0.99	312333.30	73036.52
G009856	14.8	0.99	95900.00	21128.42
G009851	1.5	0.42	105766.70	27048.91
G009868	12.15	2.47	43033.33	9141.85
G009850	63.45	13.93	228200.00	101542.10
G009849	57.55	8.27	225400.00	46001.30
G009848	33	5.37	156333.30	20647.84
G009847	66.75	7	100866.70	22159.72
G009846	61.85	5.02	31766.67	10107.59
G009845	54.4	7.5	43020.00	11582.23
G009844	47.15	2.05	110466.70	32031.44

[0263] 表2:递送至原代人肝细胞的sgRNA的白蛋白内含子1编辑和转基因表达数据

[0264]

指导编号	平均%编辑	标准差%编辑	平均 RLU	标准差 RLU
G009844	19.07	2.07	268333.30	80432.17
G009851	0.43	0.35	18033.33	2145.54
G009852	47.20	3.96	18400.00	2251.67
G009857	0.10	0.14	71100.00	14609.24
G009858	8.63	9.16	32000.00	18366.55
G009859	3.07	3.50	59500.00	16014.99
G009860	18.80	4.90	190333.30	54307.76
G009861	10.27	2.51	62233.33	9865.26
G009866	13.60	13.55	96200.00	46573.81
G009867	12.97	3.04	3916.67	1682.03
G009868	0.63	0.32	10176.67	2037.80
G009874	49.13	0.60	318000.00	114118.40
G012747	3.83	0.23	51000.00	6161.17
G012748	1.30	0.35	17433.33	2709.86
G012749	9.77	1.50	75066.67	11809.04
G012750	42.73	4.58	5346.67	2977.35
G012751	7.77	1.16	32066.67	18537.62
G012752	32.93	2.27	402000.00	83144.45
G012753	21.20	2.95	71800.00	32055.73
G012754	0.60	0.10	16933.33	4254.80
G012755	1.10	0.10	13833.33	3685.56
G012756	2.17	0.40	35600.00	6055.58
G012757	1.07	0.25	13993.33	6745.08
G012758	0.90	0.10	34900.00	15308.82
G012759	2.60	0.35	30566.67	15287.36
G012760	39.10	6.58	6596.67	2133.13
G012761	36.17	2.43	467666.70	210965.20
G012762	8.50	0.57	217000.00	13000.00
G012763	47.07	3.07	142333.30	37581.02
G012764	44.57	5.83	1423333.00	261023.60
G012765	19.90	1.68	179666.70	57011.69
G012766	8.50	0.28	243333.30	17473.79

[0265] 另外,利用靶向原代人肝细胞 (PHH) 中人白蛋白的内含子1的单个指导gRNA,在一组靶位点上测试包含双向构建体的ssAAV载体。

[0266] 如实施例1中所述制备ssAAV和LNP材料并将其递送至PHH。处理后,收集分离的基因组DNA和细胞培养基分别用于编辑和转基因表达分析。如上所述,每个载体包含可通过如实施例1中所述的基于萤光素酶的荧光检测来测量的报告基因(衍生自质粒P00415),在图12C中绘制并在表23中示出为相对萤光素酶单位(“RLU”)。例如,AAV载体含有NanoLuc ORF (除了GFP)。图11B和图12B中提供了所测试的载体的示意图。表1和表7中列出的测试的gRNA均使用缩短的数字显示在图12C中。

[0267] 表23:递送至原代人肝细胞的sgRNA的白蛋白内含子1转基因表达数据

[0268]

指导	平均荧光素酶(RLU)	标准差荧光素酶(RLU)
G009844	3,700,000	509,117
G009852	281,000	69,296
G009857	1,550,000	127,279
G009858	551,000	108,894
G009859	1,425,000	77,782
G009860	2,240,000	183,848
G009861	663,500	238,295
G009866	274,000	11,314
G009867	44,700	566
G009874	2,865,000	431,335
G012747	651,000	59,397

[0269]

G012749	867,000	93,338
G012752	4,130,000	268,701
G012753	1,145,000	162,635
G012757	579,000	257,387
G012760	129,000	36,770
G012761	4,045,000	728,320
G012762	2,220,000	127,279
G012763	1,155,000	205,061
G012764	11,900,000	1,555,635
G012765	1,935,000	134,350
G012766	2,050,000	169,706
LNP	8,430	212

[0270] 实施例10-来自替代的安全港基因座的因子IX表达的体内测试

[0271] 在本实施例中,评估了包含双向hFIX构建体的ssAAV在替代的安全港位点的插入。为了测试插入到替代的安全港基因座中,如上所述制备AAV。以 3×10^{11} vg/只小鼠的剂量给小鼠施用AAV,之后立即以0.3mg/kg的剂量施用由Cas9 mRNA和指导RNA配制的LNP。给药后4周将动物处死,并收集肝脏和血液样品。通过NGS确定肝脏样品中的编辑。通过ELISA确定血清中的人hFIX水平。NGS和ELISA数据显示hFIX在替代的安全港基因座中有效插入和表达。

[0272] 实施例11-非人灵长类动物中人因子IX基因插入的体内测试

[0273] 在本实施例中,进行8周的研究,以通过在各种指导下施用腺相关病毒(AAV)和/或脂质纳米颗粒(LNP)来评估食蟹猴中人因子IX基因插入和hFIX蛋白表达。用如上所述制备的LNP制剂和AAV制剂进行此研究。每个LNP制剂都含有Cas9 mRNA和指导RNA(gRNA),其中mRNA:gRNA的重量比为2:1。ssAAV来源于P00147。

[0274] 以n=3的队列处理雄性食蟹猴。以表3所述的剂量通过缓慢推注注射或输注向动物给药AAV。AAV处理后,动物通过缓慢推注或输注接受如表3所述的缓冲液或LNP。

[0275] 给药后两周,通过单次超声引导下穿刺活检收集肝脏标本。每个活检标本在液氮中快速冷冻,并储存在-86℃至-60℃下。如前所述,通过NGS测序对肝脏标本进行编辑分析。

[0276] 对于因子IX ELISA分析,在给药后7天、14天、28天和56天从动物收集血液样品。在抽血后收集血液样品并处理成血浆,并保存在-86℃至-60℃下直至分析。

[0277] 通过ELISA从血浆样品中确定总人因子IX水平。简而言之,Reacti-Bind 96孔微板(VWR目录号PI15041)涂覆有浓度为1 μ g/ml的捕获抗体(抗人因子IX抗体的小鼠mAB(HTI,目录号AHIX-5041)),然后使用含5%牛血清白蛋白的1x PBS封闭。接下来在单个孔中孵育稀释在食蟹猴血浆中的纯化的人因子IX蛋白(ERL,目录号HFIX1009,批次号HFIX4840)的测试样品或标准品。以100ng/ml的浓度吸附检测抗体(绵羊抗人因子9多克隆抗体,Abcam,目录号ab128048)。以100ng/mL使用二抗(具有HRP的驴抗绵羊IgG pAb,Abcam,目录号ab97125)。使用TMB底物试剂组(BD OptEIA目录号555214)显影板。在微板读取器(Molecular Devices i3系统)上于450nm处分光光度法评估光密度,并使用SoftMax pro 6.4进行分析。

[0278] 检测到插入缺失形成,从而确认发生编辑。NGS数据显示有效的插入缺失形成。通过ELISA测量NHP中白蛋白基因座中hFIX的表达,并在表4和图13中进行描述。hFIX的血浆水平达到了先前描述为治疗有效的水平(George等人,NEJM 377 (23),2215-27,2017)。

[0279] 根据测量,通过八周的研究,循环hFIX蛋白水平得以维持(参见图13,显示第7天、第14天、第28天和第56天的平均水平分别为约135ng/m、约140ng/m、约150ng/m和约110ng/mL),使蛋白质水平达到约75ng/mL至约250ng/mL。血浆hFIX水平是使用比R338L功能过度的hFIX变体高约8倍的比活性来计算的(Simioni等人,NEJM 361 (17),1671-75,2009)(其报道了hFIX-R338L的蛋白比活性为 390 ± 28 U/毫克,并且野生型因子IX的蛋白比活性为 45 ± 2.4 U/毫克。计算本实施例中测试的功能过度的因子IX变体的功能标准化因子IX活性,实验在8周的研究中达到了NHP中稳定的人因子IX蛋白水平,相当于野生型因子IX活性的约20-40%(范围跨越野生型因子IX活性的12-67%)。

[0280] 表3:肝脏中的编辑

动物 编号	指导编 号	F9-AAV (vg/kg)	F9-AAV 体积 (mL/kg)	LNP (mg/kg)	LNP 体 积 (mL/kg)
4001	G009860	3E+13	1	3	2
4002	G009860	3E+13	1	3	2
4003	G009860	3E+13	1	3	2
5001	TSS	3E+13	1	0	0
5002	TSS	3E+13	1	0	0
5003	TSS	3E+13	1	0	0
6001	G009862	0	0	3	2
6002	G009862	0	0	3	2
6003	G009862	0	0	3	2

[0282] 表4:hFIX表达

动物 编号	第 7 天 因子 IX (ng/mL)	第 14 天 因子 IX (ng/mL)	第 28 天 因子 IX (ng/mL)	第 56 天 因子 IX (ng/mL)
[0283]	4001	122.84/+2.85	94.93/+0.56	105.65/+1.94
	4002	149.77/+13.5	222.92/+9.61	252.49/+6.46
	4003	134.06/+6.17	107.04/+6.46	95.30/+3.18
	5001	ND	ND	ND
	5002	ND	ND	ND
[0284]	5003	ND	ND	ND
	6001	ND	ND	ND
	6002	ND	ND	ND
	6003	ND	ND	ND

[0285] 实施例12-非人灵长类动物中因子IX插入的体内测试

[0286] 在本实施例中,进行研究,以在各种指导(包括G009860和各种LNP组分)下施用衍生自P00147的ssAAV和/或CRISPR/Cas9脂质纳米颗粒(LNP)后评估食蟹猴中因子IX基因插入和hFIX蛋白表达。

[0287] 通过NGS测量插入缺失形成,从而确认发生编辑。使用抗人因子IX抗体的小鼠mAb (HTI,目录号AHIX-5041)、绵羊抗人因子9多克隆抗体(Abcam,目录号ab128048)和具有HRP的驴抗绵羊IgG pAb(Abcam,目录号ab97125)通过ELISA从血浆样品中确定总人因子IX水平,如实施例11中所述。使用替代的CRISPR/Cas9 LNP从双向模板获得比实施例13的实验中达到的人FIX蛋白水平高>3倍的人FIX蛋白水平。在研究中,ELISA分析结果表明,至少在第14天和第28天的时间点在GHP中使用G009860达到处于或高于人FIX水平的正常范围(3-5ug/mL;Amiral等人,Clin.Chem.,30(9),1512-16,1984)的循环hFIX蛋白水平。初始数据表明,单次给药后第14天的循环人FIX蛋白水平为约3-4μg/mL,在研究的前28天维持水平(约3-5μg/mL)。

[0288] 通过ELISA测量循环白蛋白水平,这指示基线白蛋白水平维持在28天。未经处理的动物中测试的白蛋白水平在研究中变化了±约15%。在治疗的动物中,循环白蛋白水平变化很小且没有超出正常范围,并且所述水平在一个月内恢复到基线。

[0289] 还通过具有更大动态范围的夹心免疫测定法确定循环人FIX蛋白水平。简而言之,用1%ECL封闭剂(Sigma,GERPN2125)封闭MSD GOLD 96孔链霉亲和素SECTOR板(Meso Scale Diagnostics,目录L15SA-1)。取出封闭溶液后,将生物素化的捕获抗体(Sino Biological,11503-R044)固定在板上。重组人FIX蛋白(酶研究实验室,hFIX 1009)用于制备0.5%ECL封闭剂中的校准标准品。洗涤后,将校准标准品和血浆样品添加到板中并孵育。洗涤后,将与碘基标签标记缀合的检测抗体(Haematologic Technologies,AHIX-5041)添加到孔中并孵育。在洗去任何未结合的检测抗体后,将读取缓冲器T施加到孔中。在没有任何另外的孵育的情况下,使用MSD Quick Plex SQ120仪器对板进行成像,并使用Discovery Workbench 4.0软件包(Meso Scale Discovery)分析数据。浓度表示为平均计算浓度,以ug/ml为单位。对于样品,除非用星号指示(在所述情况下,N=2),否则N=3。表24描绘了如通过MSD ELISA

测量的在治疗的研究组中来自白蛋白基因座的hFIX表达。

[0290] 表24:血清人因子IX蛋白水平

	平均计算浓度 (ug/mL)		
	3001	3002	3003
第 7 天	7.85	5.63	11.20
第 14 天	8.65	11.06	14.70
第 28 天	9.14	14.12	10.85
第 42 天	9.03	33.12*	13.22
第 56 天	10.24	16.72	33.84*

[0292] 实施例13-白蛋白人指导的脱靶分析

[0293] 生化方法(参见例如,Cameron等人,Nature Methods.6,600-606;2017)用于确定由靶向白蛋白的Cas9裂解的潜在脱靶基因组位点。在此实验中,使用分离的HEK293基因组DNA筛选了13种靶向人白蛋白的sgRNA和具有已知脱靶特征的两个对照指导。表26示出了在生化分析中使用16nM指导浓度检测到的潜在脱靶位点的数量。所述测定识别所测试的sgRNA的潜在脱靶位点。

[0294] 表25:脱靶分析

gRNA 编号	靶	指导序列(SEQ ID NO:)	脱靶位点计数
---------	---	------------------	--------

[0296]

G012753	白蛋白	GACUGAAACUUCACAGAAUA (SEQ ID NO: 20)	62
G012761	白蛋白	AGUGCAAUGGAUAGGUCUUU (SEQ ID NO: 28)	75
G012752	白蛋白	UGACUGAAACUUCACAGAAU (SEQ ID NO: 19)	223
G012764	白蛋白	CCUCACUCUUGUCUGGGCAA (SEQ ID NO: 31)	3985
G012763	白蛋白	UGGGCAAGGGAAGAAAAAA (SEQ ID NO: 30)	5443
G009857	白蛋白	AUUUAUGAGAUCAACAGCAC (SEQ ID NO: 5)	131
G009859	白蛋白	UUAAAUAAAGCAUAGUGCAA (SEQ ID NO: 7)	91
G009860	白蛋白	UAAAGCAUAGUGCAAUGGAU (SEQ ID NO: 8)	133
G012762	白蛋白	UGAUUCCUACAGAAAAACUC (SEQ ID NO: 29)	68
G009844	白蛋白	GAGCAACCUCACUCUUGUCU (SEQ ID NO: 2)	107
G012765	白蛋白	ACCUCACUCUUGUCUGGGCA (SEQ ID NO: 32)	41
G012766	白蛋白	UGAGCAACCUCACUCUUGUC (SEQ ID NO: 33)	78
G009874	白蛋白	UAAUAAAAUCAAACAUCU (SEQ ID NO: 13)	53
G000644	EMX1	GAGUCCGAGCAGAAGAAGAA (SEQ ID NO: 1129)	304
G000645	VEGFA	GACCCCCUCCACCCCGCCUC (SEQ ID NO: 1130)	1641

[0297] 在已知的脱靶检测分析(诸如上面使用的生化方法)中,通常通过设计回收大量潜在的脱靶位点,以便为可在其他情况下(例如在感兴趣的原始细胞中)验证的潜在位点“广泛撒网”。例如,生化方法通常过多地表示潜在的脱靶位点数量,因为所述测定利用纯化的高分子量基因组DNA(不含细胞环境),并且取决于所用的Cas9 RNP剂量。因此,可使用所鉴定的潜在脱靶位点的靶向测序来验证通过这些方法鉴定的潜在脱靶位点。

[0298] 实施例14-分泌或非分泌蛋白表达的构建体的构建

[0299] 构建体(诸如双向构建体)可被设计成使得其表达分泌或非分泌蛋白。对于分泌蛋白的产生,构建体可包含有助于将多肽易位到内质网腔的信号序列。替代地,构建体可利用宿主细胞的内源性信号序列(例如,当转基因整合到宿主细胞的白蛋白基因座中时为内源白蛋白信号序列)。

[0300] 相反,用于表达非分泌蛋白的构建体可被设计成使得其不包含信号序列并且使得其不利用宿主细胞的内源性信号序列。可实现这个的一些方法包括将内部核糖体进入位点(IRES)序列并入构建体中。IRES序列(诸如EMCV IRES)允许从IRES所定位的紧靠mRNA下游的任何位置开始翻译。这将允许从含有上游信号序列的插入位点表达缺乏宿主细胞内源性信号序列的蛋白质(例如,在白蛋白基因座的外显子1中发现的信号序列将不包括在表达的

蛋白质中)。在信号序列不存在的情况下,蛋白质不将被分泌。可用于构建体的IRES序列的实例包括来自以下的那些:小核糖核酸病毒(例如,FMDV)、瘟病毒(CFFV)、脊髓灰质炎病毒(PV)、脑心肌炎病毒(ECMV)、口蹄疫病毒(FMDV)、丙型肝炎病毒(HCV)、经典猪瘟病毒(CSFV)、鼠白血病病毒(MLV)、猴免疫缺陷病毒(SIV)或蜂麻痹病毒(CrPV)。

[0301] 用于表达非分泌蛋白的替代方法是将一个或多个自裂解肽包括在构建体中感兴趣的多肽的上游。自裂解肽(诸如2A或2A样序列)充当核糖体跳跃信号,以从单个mRNA转录本产生多个单独的蛋白质。如表11中的质粒编号P00415所示,自裂解肽(诸如P2A)可用于产生表达两个转基因(例如,纳米荧光素酶和GFP)的双顺反子载体。替代地,自裂解肽可用于表达缺乏宿主细胞的内源性信号序列的蛋白质(例如,位于感兴趣的蛋白质的上游的2A序列将导致内源性白蛋白信号序列与感兴趣的蛋白质之间的裂解)。表12示出了可利用的代表性2A肽。另外,如表12所示,可将(GSG)残基添加到肽的5'端以提高裂解效率。

表 26: 用于构建体的自裂解肽	
肽	氨基酸序列
[0302] T2A (SEQ ID NO: 1131)	EGRGSLLTCGDVEENPGP
P2A (SEQ ID NO: 1132)	ATNFSLLKQAGDVEENPGP
E2A (SEQ ID NO: 1133)	QCTNYALLKLAGDVESNPGP
F2A (SEQ ID NO: 1134)	VKQTLNFDLLKLAGDVESNPGP
[0303] 具有 GSG 残基的 T2A (SEQ ID NO: 1135)	GSGEGRGSLLTCGDVEENPGP
具有 GSG 残基的 P2A (SEQ ID NO: 1136)	GSGATNFSLLKQAGDVEENPGP
具有 GSG 残基的 E2A (SEQ ID NO: 1137)	GSGQCTNYALLKLAGDVESNPGP
具有 GSG 残基的 F2A (SEQ ID NO: 1138)	GSGVKQTLNFDLLKLAGDVESNPGP

[0304] 表5:人指导RNA序列

[0305]

指导编号	指导序列	基因组坐标	SEQ ID NO:
G009844	GAGCAACCUCACUCUUGUCU	chr4:73405113-73405133	2
G009851	AUGCAUUUGUUUCAAAAUAU	chr4:73405000-73405020	3
G009852	UGCAUUUGUUUCAAAAUAUU	chr4:73404999-73405019	4
G009857	AUUUAUGAGAUCAACAGCAC	chr4:73404761-73404781	5
G009858	GAUCAACAGCACAGGUUUUG	chr4:73404753-73404773	6
G009859	UUAAAUAAGCAUAGUGCAA	chr4:73404727-73404747	7
G009860	UAAAGCAUAGUGCAAUGGAU	chr4:73404722-73404742	8
G009861	UAGUGCAAUGGAUAGGUCUU	chr4:73404715-73404735	9
G009866	UACUAAAACUUUAUUUUACU	chr4:73404452-73404472	10
G009867	AAAGUUGAACAAUAGAAAAA	chr4:73404418-73404438	11
G009868	AAUGCAUAAUCUAAGUCAAA	chr4:73405013-73405033	12
G009874	UAAUAAAAUUCAAACAUCU	chr4:73404561-73404581	13
G012747	GCAUCUUUAAAGAAUUAUUU	chr4:73404478-73404498	14
G012748	UUUGGCAUUUAUUUCUAAAA	chr4:73404496-73404516	15
G012749	UGUAUUUGUGAAGUCUUACA	chr4:73404529-73404549	16
G012750	UCCUAGGUAAAAAAAAAAAAA	chr4:73404577-73404597	17
G012751	UAAUUUUUCUUUUGCGCACUA	chr4:73404620-73404640	18
G012752	UGACUGAAACUUCACAGAAU	chr4:73404664-73404684	19
G012753	GACUGAAACUUCACAGAAUA	chr4:73404665-73404685	20
G012754	UUCAUUUUAGUCUGUCUUCU	chr4:73404803-73404823	21
G012755	AUUAUCUAAGUUUGAAUAUA	chr4:73404859-73404879	22
G012756	AAUUUUUAAAAUAGUAUUCU	chr4:73404897-73404917	23
G012757	UGAAUUAUUCUUCUGUUUAA	chr4:73404924-73404944	24
G012758	AUCAUCCUGAGUUUUUCUGU	chr4:73404965-73404985	25
G012759	UUACUAAAACUUUAUUUUAC	chr4:73404453-73404473	26
G012760	ACCUUUUUUUUUUUUUUACCU	chr4:73404581-73404601	27
G012761	AGUGCAAUGGAUAGGUCUUU	chr4:73404714-73404734	28
G012762	UGAUUCCUACAGAAAAACUC	chr4:73404973-73404993	29
G012763	UGGGCAAGGGAAGAAAAAAA	chr4:73405094-73405114	30
G012764	CCUCACUCUUGUCUGGGCAA	chr4:73405107-73405127	31
G012765	ACCUCACUCUUGUCUGGGCA	chr4:73405108-73405128	32
G012766	UGAGCAACCUCACUCUUGUC	chr4:73405114-73405134	33

[0306] 表6:小鼠指导RNA序列

[0307]

指导编号	指导序列	基因组坐标	SEQ ID NO:
G000551	AUUUGCAUCUGAGAACCCUU	chr5:90461148-90461168	98
G000552	AUCGGGAACUGGCAUCUUC	chr5:90461590-90461610	99
G000553	GUUACAGGAAAUCUGAAGG	chr5:90461569-90461589	100
G000554	GAUCGGGAACUGGCAUCUUC	chr5:90461589-90461609	101
G000555	UGCAUCUGAGAACCCUUAGG	chr5:90461151-90461171	102
G000666	CACUCUUGUCUGUGGAAACA	chr5:90461709-90461729	103
G000667	AUCGUUACAGGAAAUCUGA	chr5:90461572-90461592	104
G000668	GCAUCUUCAGGGAGUAGCUU	chr5:90461601-90461621	105
G000669	CAAUCUUUAAAUAUGUUGUG	chr5:90461674-90461694	106
G000670	UCACUCUUGUCUGUGGAAAC	chr5:90461710-90461730	107
G011722	UGCUGUAUUUUUCUAGUAA	chr5:90461039-90461059	108
G011723	GUAAAUUACUACUAAGACAA	chr5:90461425-90461445	109
G011724	UUUUUCUAGUAAUGGAAGCC	chr5:90461047-90461067	110
G011725	UUAUAUUUAUUGAUUAUUUU	chr5:90461174-90461194	111
G011726	GCACAGAUUAAACACUUA	chr5:90461480-90461500	112
G011727	CACAGAUUAAACACUUAAC	chr5:90461481-90461501	113
G011728	GGUUUUAAAAUAUAUAUGU	chr5:90461502-90461522	114
G011729	UCAGAUUUUCCUGUACGAU	chr5:90461572-90461592	115
G011730	CAGAUUUUCCUGUACGAUC	chr5:90461573-90461593	116
G011731	CAAUGGUAAAUAAGAAUAA	chr5:90461408-90461428	117
G013018	GGAAAUCUGAAGGUGGCAA	chr5:90461563-90461583	118
G013019	GGCGAUCUCACUCUUGUCUG	chr5:90461717-90461737	119

[0308] 表7:食蟹猴指导RNA序列

[0309]

指导编号	指导序列	基因组坐标	SEQ ID NO:
G009844	GAGCAACCUCACUCUUGUCU	chr5:61198711-61198731	164
G009845	AGCAACCUCACUCUUGUCUG	chr5:61198712-61198732	165
G009846	ACCUCACUCUUGUCUGGGGA	chr5:61198716-61198736	166
G009847	CCUCACUCUUGUCUGGGGAA	chr5:61198717-61198737	167
G009848	CUCACUCUUGUCUGGGGAAG	chr5:61198718-61198738	168
G009849	GGGGAAGGGGAGAAAAAAA	chr5:61198731-61198751	169
G009850	GGGAAGGGGAGAAAAAAA	chr5:61198732-61198752	170
G009851	AUGCAUUUGUUCAAAUAU	chr5:61198825-61198845	171
G009852	UGCAUUUGUUCAAAUAU	chr5:61198826-61198846	172
G009853	UGAUUCCUACAGAAAAGUC	chr5:61198852-61198872	173
G009854	UACAGAAAAGUCAGGAUAA	chr5:61198859-61198879	174
G009855	UUUCUUCUGCCUUAAACAG	chr5:61198889-61198909	175
G009856	UUAUAGUUUAUAUUCAAAC	chr5:61198957-61198977	176

[0310]

指导编号	指导序列	基因组坐标	SEQ ID NO:
G009857	AUUUAUGAGAUC AACAGCAC	chr5:61199062-61199082	177
G009858	GAUCAACAGCACAGGUUUUG	chr5:61199070-61199090	178
G009859	UUAAAUAAGCAUAGUGCAA	chr5:61199096-61199116	179
G009860	UAAAGCAUAGUGCAAUGGAU	chr5:61199101-61199121	180
G009861	UAGUGCAAUGGAUAGGUCUU	chr5:61199108-61199128	181
G009862	AGUGCAAUGGAUAGGUCUUA	chr5:61199109-61199129	182
G009863	UUACUUUGCACUUUCCUUAG	chr5:61199186-61199206	183
G009864	UACUUUGCACUUUCCUUAGU	chr5:61199187-61199207	184
G009865	UCUGACCUUUUAUUUUACCU	chr5:61199238-61199258	185
G009866	UACUAAAACUUUAUUUUACU	chr5:61199367-61199387	186
G009867	AAAGUUGAACAAUAGAAAAA	chr5:61199401-61199421	187
G009868	AAUGCAUAAUCUAAGUCAAA	chr5:61198812-61198832	188
G009869	AUUAUCCUGACUUUUUCUGU	chr5:61198860-61198880	189
G009870	UGAAUUAUUCUCUGUUUAA	chr5:61198901-61198921	190
G009871	UAAUUUUCUUUUGCCCACUA	chr5:61199203-61199223	191
G009872	AAAAGGUCAGAAUUGUUUAG	chr5:61199229-61199249	192
G009873	AACAUCCUAGGUAAAAUAAA	chr5:61199246-61199266	193
G009874	UAAUAAAAUUCAAACAUCU	chr5:61199258-61199278	194
G009875	UUGUCAUGUAUUUCUAAAAU	chr5:61199322-61199342	195
G009876	UUUGUCAUGUAUUUCUAAAA	chr5:61199323-61199343	196

[0311] 表8. 人白蛋白sgRNA和修饰模式

[0312]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G009844	GAGCAACCUCACUCUUGUCUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	34	mG*mA*mG*CAACCUCACUCUUG UCUGU UUUAGAmGmCmUmAmGmAmAm AmUm AmGmCAAGUUAAAAUAGGCUA GUCC GUUAUCAmAmCmUmUmGmAmA mAmAm AmGmUmGmGmCmAmCmCmGmA mGmUm CmGmGmUmGmCmU*mU*mU*mU	66
G009851	AUGCAUUUGUUUAAAAUAUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	35	mA*mU*mG*CAUUUGUUUAAAA UAUG UUUUAGAmGmCmUmAmGmAmA mAmUm AmGmCAAGUUAAAAUAGGCUA GUCCG UUAUCAmAmCmUmUmGmAmAm AmAmAm GmUmGmGmCmAmCmCmGmAmG mUmCm GmGmUmGmCmU*mU*mU*mU	67
G009852	UGCAUUUGUUUAAAAUAUG UUUU AGAGCUAGAAUAGCAAGUUA	36	mU*mG*mC*AUUUGUUUAAAAU AUUGU UUUAGAmGmCmUmAmGmAmAm	68

[0313]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
	AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU		AmUmAm GmCAAGUUAUUUAAGGCUAGU CCGUUA UCAmAmCmUmUmGmAmAmAmA mAmGmUmGmGmCmAmCmCmGm AmGmUmCmGmGmUmGmCmU*mU *mU*mU	
G009857	AUUUAUGAGAUCAACAGCACG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	37	mA*mU*mU*UAUGAGAUCAACAG CACGU UUUAGAmGmCmUmAmGmAmAm AmUmAm GmCAAGUUAUUUAAGGCUAGU CCGUUA UCAmAmCmUmUmGmAmAmAmA mAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	69
G009858	GAUCAACAGCACAGGUUUUGG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	38	mG*mA*mU*CAACAGCACAGGUU UUGGU UUUAGAmGmCmUmAmGmAmAm AmUmAm GmCAAGUUAUUUAAGGCUAGU CCGUUA UCAmAmCmUmUmGmAmAmAmA mAmGm UmGmGmCmAmCmCmGmAmGmU mCmGm GmUmGmCmU*mU*mU*mU	70
G009859	UUAAUAAAGCAUAGUGCAAG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	39	mU*mU*mA*AAUAAAGCAUAGUG CAAGUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	71
G009860	UAAAGCAUAGUGCAAUGGAUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	40	mU*mA*mA*AGCAUAGUGCAAUG GAUGUUUAGAmGmCmUmAmG mAmAmAmUmAmGmCAAGUUAAA AAUAAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	72
G009861	UAGUGCAAUGGAUAGGUCUUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	41	mU*mA*mG*UGCAAUGGAUAGGU CUUGUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	73
G009866	UACUAAAACUUUAUUUUACUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	42	mU*mA*mC*UAAAACUUUAUUUU ACUGUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	74
G009867	AAAGUUGAACAAUAGAAAAAG UUUU	43	mA*mA*mA*GUUGAACAAUAGAA AAAGUUUAGAmGmCmUmAmG	75

[0314]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
	AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU		mAmAmAmUmAmGmCAAGUUA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	
G009868	AAUGCAUAAUCUAAGUCAAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	44	mA*mA*mU*GCAUAAUCUAAGUC AAAGUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	76
G009874	UAAUAAAAUCAAACAUCUG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	45	mU*mA*mA*UAAAAUCAAACAU CCUGUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUA AUAAGGCUAGUCCGUUAUCAm mCmUmUmGmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU* U	77
G012747	GCAUCUUUAAAGAAUUAUUUG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	46	mG*mC*mA*UCUUUAAAGAAUUA UUUGUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	78
G012748	UUUGGCAUUUAUUUCUAAAAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	47	mU*mU*mU*GGCAUUUAUUUCUA AAAGUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	79
G012749	UGUAUUUGUGAAGUCUACAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	48	mU*mG*mU*AUUUGUGAAGUCUU ACAGUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUA AUAAGGCUAGUCCGUUAUCAm mCmUmUmGmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU* U	80
G012750	UCCUAGGUAAAAAAAAAAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	49	mU*mC*mC*UAGGUAAAAAAAAA AAAGUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	81
G012751	UAAUUUUUUUUUGCGCACUAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	50	mU*mA*mA*UUUUUUUUUGCGCA CUAGUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUA AUAAGGCUAGUCCGUUAUCAm mCmUmUmGmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU* U	82
G012752	UGACUGAAACUUCACAGAAUG	51	mU*mG*mA*CUGAAACUUCACAG	83

[0315]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
	UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU		AAUGUUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUAA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	
G012753	GACUGAAACUUCACAGAAUAG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	52	mG*mA*mC*UGAAACUUCACAGA AUAGUUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUAA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	84
G012754	UUCAUUUUAGUCUGUCUUCUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	53	mU*mU*mC*AUUUUAGUCUGUCU UCUGUUUUAGAmGmCmUmAmGm AmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAm mCmUmUmGmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	85
G012755	AUUAUCUAAGUUUGAAUAUAG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	54	mA*mU*mU*AUCUAAGUUUGAAU AUAGUUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUAA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	86
G012756	AAUUUUUAAAAUAGUAUUCUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	55	mA*mA*mU*UUUUAAAAUAGUAU UCUGUUUUAGAmGmCmUmAmGm AmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAm mCmUmUmGmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	87
G012757	UGAAUUAUUCUUCUGUUUAAG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	56	mU*mG*mA*AUUAUUCUUCUGUU UAAGUUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUAA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	88
G012758	AUCAUCCUGAGUUUUUCUGUG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	57	mA*mU*mC*AUCCUGAGUUUUUC UGUGUUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUAA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	89
G012759	UUACUAAAACUUUAUUUUACG UUUU AGAGCUAGAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	58	mU*mU*mA*CUAAAACUUUAUUU UACGUUUUAGAmGmCmUmAmGm AmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAm mCmUmUmGmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	90

[0316]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G012760	ACCUUUUUUUUUUUUUUACCUU UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	59	mA*mC*mC*UUUUUUUUUUUUUA CCUGUUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	91
G012761	AGUGCAAUGGAUAGGUCUUUG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	60	mA*mG*mU*GCAAUGGAUAGGUC UUUGUUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUAA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	92
G012762	UGAUUCCUACAGAAAAACUCG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	61	mU*mG*mA*UUCUACAGAAAAA CUCGUUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	93
G012763	UGGGCAAGGGAAGAAAAAAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	62	mU*mG*mG*GCAAGGGAAGAAAA AAAGUUUUAGAmGmCmUmAmG mAmAmUmAmGmCAAGUUAA AAUAAGGCUAGUCCGUUAUCAm AmCmUmUmGmAmAmAmAmAmG mUmGmGmCmAmCmCmGmAmGm UmCmGmGmUmGmCmU*mU*mU* mU	94
G012764	CCUCACUCUUGUCUGGGCAAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	63	mC*mC*mU*CACUCUUGUCUGGG CAAGUUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	95
G012765	ACCUCACUCUUGUCUGGGCAG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	64	mA*mC*mC*UCACUCUUGUCUGG GCAGUUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	96
G012766	UGAGCAACCUCACUCUUGUCG UUUU AGAGCUAGAAAUAGCAAGUUA AAAU AAGGCUAGUCCGUUAUCAACU UGAA AAAGUGGCACCGAGUCGGUGC UUUU	65	mU*mG*mA*GCAACCUCACUCUU GUCGUUUUAGAmGmCmUmAmGm AmAmAmUmAmGmCAAGUUAAA AUAAGGCUAGUCCGUUAUCAmA mCmUmUmGmAmAmAmAmAmGm UmGmGmCmAmCmCmGmAmGmU mCmGmGmUmGmCmU*mU*mU*m U	97

[0317] 表9. 小鼠白蛋白指导sgRNA和修饰模式

[0318]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G000551	AUUUGCAUCUGAGAACC CUUGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	120	mA*mU*mU*UGCAUCUGAG AACCCUUGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUUAUAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	142
G000552	AUCGGAACUGGCAUCU UCAGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	121	mA*mU*mC*GGGAACUGGC AUCUUCAGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUUAUAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	143
G000553	GUUACAGGAAAAUCUGA AGGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	122	mG*mU*mU*ACAGGAAAA UCUGAAGGUUUUAGAmG mCmUmAmGmAmAmAmUm AmGmCAAGUUAUAAUAAG GCUAGUCCGUUAUCAmAm CmUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm GmAmGmUmCmGmGmUmG mCmU*mU*mU*mU	144
G000554	GAUCGGGAACUGGCAUC UUCGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	123	mG*mA*mU*CGGGAACUGG CAUCUUCGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUUAUAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	145
G000555	UGCAUCUGAGAACCCU AGGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	124	mU*mG*mC*AUCUGAGAAC CCUAGGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUUAUAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	146
G000666	CACUCUUGUCUGUGGAA ACAGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	125	mC*mA*mC*UCUUGUCUGU GGAAACAGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUUAUAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	147

[0319]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G000667	AUCGUUACAGGAAAAUC UGAGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	126	mA*mU*mC*GUUACAGGAA AAUCUGAGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	148
G000668	GCAUCUUCAGGGAGUAG CUUGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	127	mG*mC*mA*UCUUCAGGGA GUAGCUUGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	149
G000669	CAAUCUUUAAAAUAUGUU GUGGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	128	mC*mA*mA*UCUUUAAAAUA UGUUGUGGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	150
G000670	UCACUCUUGUCUGUGGA AACGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	129	mU*mC*mA*CUCUUGUCUG UGGAAACGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	151
G011722	UGCUGUAUUUUUCUAG UAAGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	130	mU*mG*mC*UUGUAUUUU UCUAGUAAGUUUUAGAmG mCmUmAmGmAmAmAmUm AmGmCAAGUAAAAUAAG GCUAGUCCGUUAUCAmAm CmUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm GmAmGmUmCmGmGmUmG mCmU*mU*mU*mU	152
G011723	GUAAAUAUCUACUAAGA CAAGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	131	mG*mU*mA*AAUAUCUACU AAGACAAGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	153

[0320]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G011724	UUUUUCUAGUAAUGGAA GCCGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	132	mU*mU*mU*UUCUAGUAA UGGAAGCCGUUUUAGAmG mCmUmAmGmAmAmUm AmGmCAAGUAAAAUAAG GCUAGUCCGUUAUCAmAm CmUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm GmAmGmUmCmGmGmUmG mCmU*mU*mU*mU	154
G011725	UUAUAUUUAUUGAUUAU UUUGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	133	mU*mU*mA*UAUUAUUGA UAUAUUUUUGUUUAGAmG mCmUmAmGmAmAmUm AmGmCAAGUAAAAUAAG GCUAGUCCGUUAUCAmAm CmUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm GmAmGmUmCmGmGmUmG mCmU*mU*mU*mU	155
G011726	GCACAGAUUAACACU UAAGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	134	mG*mC*mA*GAGUAUAAA CACUUAAGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	156
G011727	CACAGAUUAACACU AACGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	135	mC*mA*mC*AGAUUAAAC ACUUAACGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	157
G011728	GGUUUUAAAAUAAUAA UGUGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	136	mG*mG*mU*UUUAAAAAU AAUAAUGUGUUUUAGAmG mCmUmAmGmAmAmAmUm AmGmCAAGUAAAAUAAG GCUAGUCCGUUAUCAmAm CmUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm GmAmGmUmCmGmGmUmG mCmU*mU*mU*mU	158
G011729	UCAGAUUUUCCUGUAAC GAUGUUUUAGAGCUAGA AAUAGCAAGUAAAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	137	mU*mC*mA*GAUUUCCUG UAACGAUGUUUUAGAmGm CmUmAmGmAmAmAmUmA mGmCAAGUAAAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmAm GmUmGmGmCmAmCmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	159

[0321]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G011730	CAGAUUUUCCUGUAACG AUCGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	138	mC*mA*mG*AUUUUCCUGU AACGAUCGUUUUAGAmGm CmUmAmGmAmAmUmA mGmCAAGUUAUAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmG mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	160
G011731	CAAUGGUAAAUAAGAAA UAAGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	139	mC*mA*mA*UGGUAAAUA AGAAAUAAGUUUUAGAmG mCmUmAmGmAmAmUm AmGmCAAGUUAUAAUAAG GCUAGUCCGUUAUCAmAm CmUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm GmAmGmUmCmGmGmUmG mCmU*mU*mU*mU	161
G013018	GGAAAAUCUGAAGGUGG CAAGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	140	mG*mG*mA*AAAUCUGAA GGUGGCAAGUUUUAGAmG mCmUmAmGmAmAmUm AmGmCAAGUUAUAAUAAG GCUAGUCCGUUAUCAmAm CmUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm GmAmGmUmCmGmGmUmG mCmU*mU*mU*mU	162
G013019	GGCGAUCUCACUCUUGU CUGGUUUUAGAGCUAGA AAUAGCAAGUUAUAAUA AGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCAC CGAGUCGGUGCUUUU	141	mG*mG*mC*GAUCUCACUC UUGUCUGGUUUUAGAmGm CmUmAmGmAmAmUmA mGmCAAGUUAUAAUAAGG CUAGUCCGUUAUCAmAmC mUmUmGmAmAmAmAmA mGmUmGmGmCmAmCmCm mAmGmUmCmGmGmUmGm CmU*mU*mU*mU	163

[0322] 表10:食蟹猴sgRNA和修饰模式

[0323]

指导编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G009844	GAGCAACCUCACUC UUGUCU GUUUUAGAGCUAG AAAUAGC AAGUUAUAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	197	mG*mA*mG*CAACCUCACUCUUGUC UGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmA mCmUmUm GmAmAmAmAmGmUmGmGmCmA mCmCmGm AmGmUmCmGmGmUmGmCmU*mU* mU*mU	230

[0324]

指导 编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G0098 45	AGCAACCUCACUCU UGUCUG GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	198	mA*mG*mC*AACCUCACUCUUGUCU GGUUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAAUAAAGGCUAGUCCGUUAUCAmAmCmUmUm GmAmAmAmAmAmGmUmGmGmCmA mCmCmGm AmGmUmCmGmGmUmGmCmU*mU* mU*mU	231
G0098 46	ACCUCACUCUUGUC UGGGGA GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	199	mA*mC*mC*UCACUCUUGUCUGGGG AGUUUUU AGAmGmCmUmAmGmAmAmUmAm mGmCAA GUUAAAAUAAGGCUAGUCCGUUAU CAmAmCm UmUmGmAmAmAmAmAmGmUmGmG mCmAmCm CmGmAmGmUmCmGmGmUmGmCmU *mU*mU*mU	232
G0098 47	CCUCACUCUUGUCU GGGGAA GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	200	mC*mC*mU*CACUCUUGUCUGGGGA AGUUUUU GAmGmCmUmAmGmAmAmUmAm GmCAAGU UAAAAUAAGGCUAGUCCGUUAUCA mAmCmUm UmGmAmAmAmAmAmGmUmGmGmC mAmCmCm GmAmGmUmCmGmGmUmGmCmU*m U*mU*mU	233
G0098 48	CUCACUCUUGUCUG GGGAAG GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	201	mC*mU*mC*ACUCUUGUCUGGGGA AGGUUUU AGAmGmCmUmAmGmAmAmUmAm mGmCAA GUUAAAAUAAGGCUAGUCCGUUAU CAmAmCm UmUmGmAmAmAmAmAmGmUmGmG mCmAmCm CmGmAmGmUmCmGmGmUmGmCmU *mU*mU*mU	234
G0098 49	GGGGAAGGGGAGA AAAAAAA GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	202	mG*mG*mG*GAAGGGGAGAAAAAA AAGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmA mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	235
G0098 50	GGGAAGGGGAGAA AAAAAAA GUUUUAGAGCUAG AAAUAGC	203	mG*mG*mG*AAGGGGAGAAAAAA AAGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA	236

[0325]

指导 编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
	AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU		AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	
G0098 51	AUGCAUUUGUUUC AAAAUAU GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	204	mA*mU*mG*CAUUUGUUUCAAUAU AUGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	237
G0098 52	UGCAUUUGUUUCA AAAAUAU GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	205	mU*mG*mC*AUUUGUUUCAAUAU UUGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	238
G0098 53	UGAUUCCUACAGA AAAAGUC GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	206	mU*mG*mA*UUCUACAGAAAAAG UCGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	239
G0098 54	UACAGAAAAAGUC AGGAUAA GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	207	mU*mA*mC*AGAAAAAGUCAGGAU AAGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	240
G0098 55	UUUCUUCUGCCUU UAAACAG GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU	208	mU*mU*mU*CUUCUGCCUUUAAAC AGGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm	241

[0326]

指导 编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
	GGCACCGAGUCGG UGCUIUUU		UmCmGmGmUmGmCmU*mU*mU*mU	
G0098 56	UUUAUAGUUUAUA UUCAAAC GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	209	mU*mU*mA*UAGUUUAUAUUCAA ACGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	242
G0098 57	AUUUAUGAGAUA ACAGCAC GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	210	mA*mU*mU*UAUGAGAUAACAGC ACGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	243
G0098 58	GAUCAACAGCACA GGUUUUU GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	211	mG*mA*mU*CAACAGCACAGGUUU UGGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	244
G0098 59	UUAAAUAAGCAU AGUGCAA GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	212	mU*mU*mA*AAUAAAGCAUAGUGC AAGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	245
G0098 60	UAAAGCAUAGUGC AAUGGAU GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	213	mU*mA*mA*AGCAUAGUGCAAUGG AUGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	246
G0098 61	UAGUGCAAUGGAU AGGUCUU	214	mU*mA*mG*UGCAAUGGAUAGGUC UUGUUUUAG	247

[0327]

指导 编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
	GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU		AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	
G0098 62	AGUGCAAUGGAUA GGUCUUA GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	215	mA*mG*mU*GCAAUGGAUAGGUCU UAGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	248
G0098 63	UUACUUUGCACUU UCCUUAG GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	216	mU*mU*mA*CUUUGCACUUUCCUU AGGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	249
G0098 64	UACUUUGCACUUU CCUUAGU GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	217	mU*mA*mC*UUUGCACUUUCCUUA GUGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	250
G0098 65	UCUGACCUUUUAU UUUACCU GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	218	mU*mC*mU*GACCUUUUAUUUUAC CUGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	251
G0098 66	UACUAAAACUUUA UUUUACU GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC	219	mU*mA*mC*UAAAACUUUAUUUUUA CUGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm	252

[0328]

指导 编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
	CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU		AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	
G0098 67	AAAGUUGAACAAU AGAAAAA GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	220	mA*mA*mA*GUUGAACAAUAGAAA AAGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	253
G0098 68	AAUGCAUAAUCUA AGUCAA GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	221	mA*mA*mU*GCAUAAUCUAAGUCA AAGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	254
G0098 69	AUUAUCCUGACUU UUUCUGU GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	222	mA*mU*mU*AUCCUGACUUUUUCU GUGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	255
G0098 70	UGAAUUAUCCUC UGUUUAA GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	223	mU*mG*mA*AUUAUCCUCUGUUU AAGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	256
G0098 71	UAAUUUUCUUUUG CCCACUA GUUUUAGAGCUAG AAAUAGC AAGUUAUUUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUIUUU	224	mU*mA*mA*UUUUCUUUUGCCCAC UAGUUUUAG AmGmCmUmAmGmAmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUm GmAmAmAmAmAmGmUmGmGmCmA mCmCmGm AmGmUmCmGmGmUmGmCmU*mU* mU*mU	257

指导 编号	全序列	SEQ ID NO:	修饰的全序列	SEQ ID NO:
G0098 72	AAAAGGUCAGAAU UGUUUAG GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	225	mA*mA*mA*AGGUCAGAAUUGUUU AGGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	258
G0098 73	AACAUCCUAGGUA AAAUAAA GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	226	mA*mA*mC*AUCCUAGGUAUUA AAGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	259
G0098 74	UAAUAAAAUCAA ACAUCCU GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	227	mU*mA*mA*UAAAAUCAAACAUC CUGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	260
G0098 75	UUGUCAUGUAUUU CUAAAAU GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	228	mU*mU*mG*UCAUGUAUUUCUAAA AUGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	261
G0098 76	UUUGUCAUGUAUU UCUAAAA GUUUUAGAGCUAG AAAUAGC AAGUUAAAAUAAG GCUAGUC CGUUAUCAACUUG AAAAAGU GGCACCGAGUCGG UGCUUUU	229	mU*mU*mU*GUCAUGUAUUUCUAA AAGUUUUAG AmGmCmUmAmGmAmAmUmAmG mCAAGUUA AAUAAGGCUAGUCCGUUAUCAmAm CmUmUmGm AmAmAmAmAmGmUmGmGmCmAmC mCmGmAmGm UmCmGmGmUmGmCmU*mU*mU*mU	262

[0329]

[0330] 表11:载体组分和序列

[0331]

质粒 编号	5' ITR	剪接受体 (第1方向)	转基因 (第1方向)	Poly-A(第1方向)	Poly-A(第2方向)	转基因 (第2方向)	剪接受体 (第2方向)	3' ITR
P00147	(SEQ ID NO: 263)	小鼠白蛋白剪接受体(SEQ ID NO: 264)	人因子 IX (R338L) (SEQ ID NO: 265)	SEQ ID NO: 266	SEQ ID NO: 267	人因子 IX (R338L) (SEQ ID NO: 268)	小鼠白蛋白剪接受体 (SEQ ID NO: 269)	(SEQ ID NO: 270)
P00411	(SEQ ID NO: 263)	人因子 IX 剪接受体 (SEQ ID NO: 271)	人因子 IX (R338L)- HiBit (SEQ ID NO: 272)	SEQ ID NO: 266	SEQ ID NO: 267	人因子 IX (R338L)- HiBit (SEQ ID NO: 273)	人因子 IX 剪接受体 (SEQ ID NO: 274)	(SEQ ID NO: 270)
P00415	(SEQ ID NO: 263)	小鼠白蛋白剪接受体(SEQ ID NO: 264)	Nluc-P2A -GFP (SEQ ID NO: 275)	SEQ ID NO: 266	SEQ ID NO: 267	Nluc-P2A -GFP (SEQ ID NO: 276)	小鼠白蛋白剪接受体 (SEQ ID NO: 269)	(SEQ ID NO: 270)
P00418	(SEQ ID NO: 263)	小鼠白蛋白剪接受体(SEQ ID NO: 264)	人因子 IX (R338L)- HiBit (SEQ ID NO: 272)	SEQ ID NO: 266	SEQ ID NO: 267	人因子 IX (R338L)- HiBit (SEQ ID NO: 273)	小鼠白蛋白剪接受体 (SEQ ID NO: 269)	(SEQ ID NO: 270)

[0332] 人白蛋白内含子1: (SEQ ID NO:1)

[0333] GTAAGAAATCCATTTTTCTATTGTTCAACTTTTATTCTATTTTCCCAGTAAAATAAAGTTTTAGTAAAC
TCTGCATCTTTAAAGAATTATTTTGGCATTATTTCTAAAAATGGCATAGTATTTTGTATTTGTGAAGTCTTACAAGG
TTATCTTATTAATAAAATTCAAACATCCTAGGTAAAAAAGGTCAGAATTGTTTAGTGACTGTAATTTTCT
TTTGCGCACTAAGGAAAGTGCAAAGTAAGTCTAGAGTGACTGAACTTCACAGAATAGGGTTGAAGATTGAATTCATA
ACTATCCCAAAGACCTATCCATTGCACTATGCTTTATTTAAAAACCACAAAACCTGTGCTGTTGATCTCATAAATAG
AACTTGTATTTATATTTATTTTCAATTTAGTCTGTCTTCTTGGTTGCTGTTGATAGACACTAAAAGAGTATTAGATA
TTATCTAAGTTTGAATATAAGGCTATAAATATTTAATAATTTTAAAAATAGTATTCTTGGTAATTGAATTATTCTTC
TGTTTAAAGGCAGAAGAAATAATTGAACATCATCCTGAGTTTTTCTGTAGGAATCAGAGCCCAATATTTTGAACAA
ATGCATAATCTAAGTCAAATGGAAAGAAATATAAAAAAGTAACATTATTACTTCTTGTGTTTCTTCAGTATTTAACAAT
CCTTTTTTTTCTTCCCTTGCCAG

[0334] 5' ITR序列 (SEQ ID NO:263):

[0335] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGTCGCCGACGCCCG
GGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTC

[0336] 小鼠白蛋白剪接受体(第1方向) (SEQ ID NO:264):

[0337] TAGGTCAGTGAAGAGAAGAACAAGCAGCATATTACAGTTAGTTGTCTTCATCAATCTTTAAATATG
TTGTGTGTTTTTCTCTCCCTGTTCCACAG

[0338] 人因子IX (R338L), 第1方向 (SEQ ID NO:265):

[0339] TTTCTTGATCATGAAACGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAA
GAGTTTGTTCAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTGA
AAACACTGAAAGAACAAGTGAATTTTGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATG
GCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTA
GATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTGGTTTGCTC
CTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAGTGCCATTTCCATGTGGAAGAGTTT

CTGTTTCACAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTCTGATGTGGACTATGTAAATTCTACTGAAGCT
GAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTCGGGTGTGGTGGAGAAGATGC
CAAACCAGGTCAATTCCCTTGGCAGGTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATG
AAAAATGGATTGTAAGTCTGCTGCCCCTGTGTTGAACTGGTGTAAAAATTACAGTTGTGCGAGGTGAACATAATATT
GAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGCAATTATTCCTCACCACAACATAATGCAGCTATTAA
TAAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCCTAGTGCTAAACAGCTACGTTACACCTATTTGCA
TTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTTCCAC
AAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGAGCCACATGTCTTCTATCTACAAA
GTTCCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGG
GACCCCATGTTACTGAAGTGAAGGGACCAGTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAA
GGCAAATATGGAATATATACCAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAAACAAGCTCACTTAA

[0340] Poly-A (第1方向) (SEQ ID NO:266) :

[0341] CCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTGCCCCCTCCCCCGTGCCTTCCTTGACCCTGG
AAGGTGCCACTCCCCTGTCTTTTCTTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATT
CTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGG
CTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCC

[0342] Poly-A (第2方向) (SEQ ID NO:267) :

[0343] AAAAAACCTCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTT
TATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTCACAAATAAAGCATTTTTTTCACTGCATT
CTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTG

[0344] 人因子IX (R338L), 第2方向 (SEQ ID NO:268) :

[0345] TTAGGTGAGCTTAGTCTTTTCTTTTATCCAATTCACGTAGCGAGAGACCTTCGTATAGATGCCATATTT
CCCCTTCATCGCACATTCTCCCCCACTTATTATCCCGGTCAAGAACTTGTTCCCTCGACTTCAGTGACGTGTG
GTCCACCTGAATCACCTTGGCATGAGTCGCGACCGCCCTCGTGAAACCCAGCACAAAACATGTTATTGTAAATCGTA
AATTTCTGTGGACAGAAGACAGGTCGCTCTATCGACCAACGGGACGCGCAAATATTGCAGAACGAGGGCTGATCGACC
TTTGTGGAAGACCCGCCCCCACCCTCACATATCCGCTCCCAAATTTCAAGAAGATATTTGTATATTCTTTATCGG
CTATACAAATCGGGGTAACATAGGAGTTAAGTACGAGTGGCTCGTCCAGCTCCAGGAGGGCTATATCATGGTTGTAC
TTGTTTATAGCGGCATTATAATTGTGATGGGGTATGATCCTGATAACATTCCTTTTCTGTTTCAGTATGCTCAGTTTC
TTCAATGTTGTGTTCCGACGCCACGACCGTAATCTTAACCCCCGTCTCGACACAGTGTGCGGCCGTTACAATCCACT
TTTCATTGACTATGGAGCCCCCAGAAAACGCGTCGACTTTTCCGTTGAGCACCACCTGCCATGGAAATTGGCCAGGT
TTAGCGTCTCGCCCCCGACAACCTAGTAAAGTCATTAAATGACTGTGTGGATTGTGTTATATTATCAAGAATCGT
TTCGGCTTCAGTAGAGTTAACGTAGTCCACATCGGGAAAAAAGTGTCTCGGCCCTTGTCAACTTTGATGTCTGGGACA
CACTTACCCGACCGCACGGGAAGGGACCGCCGGTTCACAGCTCTTTTGATTCTCAGCGAGCCGGTAGCCCTCAGTG
CAACTACACACAACCTTTGTTGTGCGCGGAATTTTACAGAATTGCTCGCATCGTCCATTTTAAATGTTGCAGGTGAC
GTCCAACCTCGCAGTTTTTCTTCAAAAACAAAAGGGCACCAACACTCGTAGGAATTTATATCGTCTTTACAACCTCC
CCCCATTGACATGGATTAGATTGCGATTGGTCCCCATCGACATATTGCTTCCAGAACTCAGTGGTCCGTTCTGTA
TTCTCAAACACCTCGCGCGCTTCTTCAAACTGCATTTTCTCCATACACTCTCGCTCCAAGTTCCCTTGCACGAA
TTCTTCAAGCTTTCTGAGTTATACCTTTTAGGCCGGTTAAGTATCTTATTCGCGTTTTCGTGGTCCAGAAA

[0346] 小鼠白蛋白剪接受体 (第2方向) (SEQ ID NO:269) :

[0347] CTGTGGAAACAGGGAGAGAAAAACACACAACATATTTAAAGATTGATGAAGACAACCTAACTGTAATATGCTGCTTTTTGTTCTTCTTCACTGACCTA

[0348] 3' ITR序列(第2方向)(SEQ ID NO:270):

[0349] AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAA

[0350] 人因子IX剪接受体(第1方向)(SEQ ID NO:271):

[0351] GATTATTTGGATTAAAAACAAAGACTTTCTTAAGAGATGTAAAATTTTCATGATGTTTTCTTTTTTGCTAAACTAAAGAATTATTCTTTTACATTTTACAG

[0352] 人因子IX(R338L)-HiBit(第1方向)(SEQ ID NO:272):

[0353] TTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAAGTTTTGTTCAAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTGAAACTGAAAGAACAACCTGAATTTTGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTGGTTTGCTCTGTACTGAGGGATATCGACTTGCGAGAAAACCGAAGTCCTGTGAACCAGCAGTGCCATTTCCATGTGGAAGAGTTCTGTTTCACAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTCTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACCTCGGGTGTGGTGGAGAAGATGCCAAACCAGGTCAATTCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAACCTGCTGCCACTGTGTTGAACTGGTGTAAAAATTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGCAATTATTCCTCACCACAACCTACAATGCAGCTATTAAATAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACCTATTTGCAATGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGAGCCACATGTCTTCTATCTACAAAGTTCAACATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGAAGGGACCAGTTTCTTAAGTGAATATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGGTCTCCCGGTATGTCAACTGGATTAAGGAAAAACAAAGCTCACTGTCAGCGGATGGAGACTGTTCAAGAAGATCAGCTAA

[0354] 人因子IX(R338L)-HiBit(第2方向)(SEQ ID NO:273):

[0355] TTAGGAAATCTTCTTAACAGCCGCCAGCCGCTCACGGTGAGCTTAGTCTTTTCTTTTATCCAATTCACGTAGCGAGAGACCTTCGTATAGATGCCATATTTCCCTTCATCGCACATTCCTCCCCCAACTATTATCCCGGTCAAGAACTTGTTCTTCGACTTCAGTGACGTGTGGTCCACCTGAATCACCTTGGCATGAGTCGCGACCGCCCTCGTGAACCCAGCACAAAACATGTTATTGTAAATCGTAAATTCGTGGACAGAAGACAGGTCGCTCTATCGACCAACGGGACGCGCAAATATTGCAGAACGAGGGCTGATCGACCTTTGTGGAAGACCCGCCCCCACCCTCACATATCCGCTCCCAAATTTCAAGAAGATATTTGTATATTCTTTATCGGCTATACAAATCGGGGTAAACATAGGAGTTAAGTACGAGTGGCTCGTCCAGCTCCAGGAGGGCTATATCATGGTTGTACTTGTATAGCGGCATTATAATTGTGATGGGGTATGATCCTGATAACATTCCTTTTCTGTTTCAGTATGCTCAGTTTCTTCAATGTTGTGTTTCGCCAGCCACGACCGTAATCTTAACCCCGTCTCGACACAGTGTGCGGCCGTTACAATCCACTTTTCATTGACTATGGAGCCCCACAAAACGCGTCGACTTTTCCGTTGAGCACCACTGCCATGGAAATTGGCCAGGTTTAGCGTCCTCGCCCCGACAACCCTAGTAAAGTCATTAAATGACTGTGTGGATTGTGTTATATTATCAAGAATCGTTTCGGCTTCAGTAGAGTTAACGTAGTCCACATCGGGAAAACTG

TCTCGGCCCTTGTCAACTTTGATGTCTGGGACACACTTACCCGACCGCACGGGAAGGGCACCGCCGGTTCACAGCTC
TTTTGATTCTCAGCGAGCCGGTAGCCCTCAGTGCAACTACACACAACCTTTGTTGTGCGGCGAATTTTACAGAATTG
CTCGCATCGTCCATTTTAAATGTTGCAGGTGACGTCCAACTCGCAGTTTTTTCCTTCAAAACCAAAAGGGCACCAAC
ACTCGTAGGAATTTATATCGTCTTTACAACCTCCCCCATTTCAGACATGGATTAGATTGCGATTGGTCCCCATCGACA
TATTGCTTCCAGAACTCAGTGGTCCGTTCTGTATTCTCAAACACCTCGCGCGCTTCTTCAAACACTGCATTTTTCTC
CATACACTCTCGCTCCAAGTTCCTTGACGAATTCTTCAAGCTTTCCTGAGTTATACCTTTTAGGCCGGTTAAGTA
TCTTATTCGCGTTTTCTGTGGTCCAGAAA

[0356] 人因子IX剪接受体(第2方向)(SEQ ID NO:274):

[0357] CTGAAATGTAAAGAATAATTCTTTAGTTTTAGCAAAAAAGAAAACATCATGAAAATTTTACATCTCTT
AAGAAAGTCTTTGTTTTTAATCCAAATAATC

[0358] Nluc-P2A-GFP(第1方向)(SEQ ID NO:275):

[0359] TTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAA
GAGTTTGTTCAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCAGTATTCACTTTGGAA
GGACTTTGTGCGGTGACTGGAGGCAAACCGCTGGTTATAATCTCGACCAAGTACTGGAACAGGGCGGGGTAAGTTCCC
TCTTTTCAGAATTTGGGTGTAAGCGTTCACCAATCCAGCGGATTGTGTTGTCTGGAGAGAACGGACTCAAAATTGAC
ATCCATGTTATCATTCCATATGAAGGTCTCAGTGGAGACCAAATGGGGCAGATCGAGAAGATTTTCAAGGTAGTTTA
CCCAGTCGACGATCACCATTCAAAGTCATTCTCCACTATGGCACACTTGTTATCGACGGAGTAACTCCTAATATGA
TTGATTACTTTGGTGCCTCGTATGAGGGCATCGCAGTGTGATGGCAAAAAGATCACCGTAACAGGAACGTTGTGG
AATGGGAACAAGATAATCGACGAGAGATTGATAAATCCAGACGGGTCACCTCTGTTTCAAGGTTACAATTAACGGCGT
CACAGGATGGAGACTCTGTGAACGAATACTGGCCACAAATTTTCACTCCTGAAGCAGGCCGGAGACGTGGAGGAAA
ACCCAGGGCCCGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTA
AACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTG
CACCACCGGCAAGCTGCCCCGTGCCCTGGCCACCCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCT
ACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTC
TTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCT
GAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCT
ATATCATGGCCGACAAGCAGAAGACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTG
CAGCTCGCCGACCACTACCAGCAGAACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCCACAACCACTACCTGAG
CACCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCG
GGATCACTCTCGGCATGGACGAGCTGTACAAGGGAGGAGGAAGCCGAAGAAGAAGAGAAAAGGTCTAA

[0360] Nluc-P2A-GFP(第2方向)(SEQ ID NO:276):

[0361] TTACACCTTCCTCTTCTTCTTGGGGCTGCCGCCGCCCTTGACAGCTCGTCCATGCCAGGGTGATGCC
GGCGGCGGTACGAACTCCAGCAGCACCATGTGGTCCCTCTTCTCGTTGGGGTCCTTGCTCAGGGCGCTCTGGGTGC
TCAGGTAGTGGTTGTGCGGCAGCAGCAGGGGCCGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTTCGGCCAGCTGC
ACGCTGCCGTCTCGATGTTGTGCCTGATCTTGAAGTTCACCTTGATGCCGTTCTTCTGCTTGTGCGCCATGATGTA
CACGTTGTGGCTGTTGTAGTTGTACTCCAGCTTGTTGGCCAGGATGTTGCCGTCTCCTTGAAGTCGATGCCCTTCA
GCTCGATCCTGTTACACAGGGTGTGCCCTCGAACTTCACCTCGGCCCTGGTCTTGTAGTTGCCGTCGTCCTTGAAG
AAGATGGTCCTCTCCTGCACGTAGCCCTCGGGCATGGCGCTCTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTA
CCTGCTGAAGCACTGCACGCCGTAGGTCAGGGTGGTCACCAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGC

AGATGAACTTCAGGGTCAGCTTGCCGTAGGTGGCGTCGCCCTCGCCCTCGCCGCTCACGCTGAACTTGTGGCCGTTTC
ACGTCGCCGTCCAGCTCCACCAGGATGGGCACCACGCCGGTGAACAGCTCCTCGCCCTTGCTCACGGGGCCGGGGTT
CTCCTCCACGTCGCCGGCTGCTTCAGCAGGCTGAAGTTGGTGGCCAGGATCCTCTGCACAGCCTCCAGCCGGTCA
CGCCGTTGATGGTCACCCTGAACAGCAGGCTGCCGTGCGGGTTGATCAGCCTCTCGTCGATGATCTTGTGGCCGTTTC
CACAGGGTGCCGGTCACGGTGATCTTCTTGCCGTGGAACACGGCGATGCCCTCGTAGGGCCTGCCGAAGTAGTCGAT
CATGTTGGGGGTACGCCGTGATCACCAGGGTGCCGTAGTGCAGGATCACCTTGAAGTGGTGGTCGTCCACGGGGT
ACACCACCTTGAAAATCTTCTCGATCTGGCCCATCTGGTCGCCGCTCAGGCCCTCGTAGGGGATGATCACGTGGATG
TCGATCTTCAGGCCGTTCTCGCCGCTCAGCACGATCCTCTGGATGGGGGTACGCTCACGCCAGGTTCTGGAACAG
GCTGCTCACGCCGCCCTGCTCCAGCACCTGGTCCAGGTTGTAGCCGGCGGTCTGCCTCCAGTCGCCCACGAAGTCTT
CCAGGGTGAACACGGCCTCCTCGAAGCTGCACTTCTCCTCCATGCACTCCCTCTCCAGGTTGCCCTGCACGAACTCC
TCCAGCTTGCCGCTGTTGTACCTCTTGGGCCTGTTCAGGATCTTGTGGCGTTCTCGTGGTCCAGGAA

[0362] P00147全序列(从ITR到ITR):(SEQ ID NO:277)

[0363] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGTCGCCCCACGCCCG
GGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTA
GATCTCTTAGGTCACTGAAGAGAAGAACAAGAGCAGCATATTACAGTTAGTTGTCTTCATCAATCTTTAAATATGT
TGTGTGGTTTTTCTCTCCCTGTTTCCACAGTTTTTCTTGATCATGAAAACGCCAACAATAATTCTGAATCGGCCAAAG
AGGTATAATTACAGGTAAATTGGAAGAGTTTGTTCAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTT
TGAAGAAGCACGAGAAGTTTTTGAAGCACTGAAAGAACAACTGAATTTTGAAGCAGTATGTTGATGGAGATCAGT
GTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGA
TTTGAAGGAAAGAACTGTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAG
TGCTGATAACAAGGTGGTTTGTCTCTGTAAGGATATCGACTTGCAGAAAACCAGAAGTCTGTGAACCAGCAG
TGCCATTTCCATGTGGAAGAGTTTCTGTTTCAAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTCTCTGATGTG
GACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCAATCATTTAATGACTTCAC
TCGGGTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGATGCAT
TCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAAGTGTGCTGCCACTGTGTTGAAACTGGTGTAAATTTACA
GTTGTGCGAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGCAATTATTCCTCA
CCACAACCTACAATGCAGCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCCTTAGTGCTAA
ACAGCTACGTTACACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTA
AGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCG
AGCCACATGTCTTCTATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAG
ATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAAGGACCAGTTTCTTAAGTGAATTATTAGC
TGGGGTGAAGAGTGTGCAATGAAAGGCAAAATATGGAATATATACCAAGGTATCCCGGTATGTCAACTGGATTAAGGA
AAAAACAAAGCTCACTTAACCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTC
CTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGT
GTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGG
GATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGGGGTATCCCCAAAAACCTCCCA
CACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTA
CAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAAC
TCATCAATGTATCTTATCATGTCTGTTAGGTGAGCTTAGTCTTTTCTTTTATCCAATTCAGTAGCGAGAGACCTTC

GTATAGATGCCATATTTCCCTTCATCGCACATTCCCTCCCCCAACTTATTATCCCGGTCAAGAACTTGTTCTTC
GACTTCAGTGACGTGTGGTCCACCTGAATCACCTTGGCATGAGTCGCGACCGCCCTCGTGAACCCAGCACAAAACA
TGTTATTGTAAATCGTAAATTTCTGTGGACAGAAGACAGGTCGCTCTATCGACCAACGGGACGCGCAAATATTGCAGA
ACGAGGGCTGATCGACCTTTGTGGAAGACCCGCCCCACCCACTCACATATCCGCTCCCAAATTTCAAGAAGATATT
TGTATATTCTTTATCGGCTATACAAATCGGGGTAACATAGGAGTTAAGTACGAGTGGCTCGTCCAGCTCCAGGAGGG
CTATATCATGGTTGTA CTGTGTTATAGCGGCATTATAATTGTGATGGGGTATGATCCTGATAACATTCCTTTTCTGT
TCAGTATGCTCAGTTTCTTCAATGTTGTGTTTCGCCAGCCACGACCGTAATCTTAACCCCGTCTCGACACAGTGTGC
GGCCGTTACAATCCACTTTTCATTGACTATGGAGCCCCACAAAACGCGTCGACTTTTCCGTTGAGCACCACTGCC
ATGGAAATTGGCCAGGTTTAGCGTCCTCGCCCCGACAACCCTAGTAAAGTCATTAAATGACTGTGTGGATTGTGTT
ATATTATCAAGAATCGTTTCGGCTTCAGTAGAGTTAACGTAGTCCACATCGGGAAAACTGTCTCGGCCCTTGTCAA
CTTTGATGTCTGGGACACACTTACCCGACCGCACGGGAAGGGCACCGCCGTTTACAGCTCTTTTGATTCTCAGCGA
GCCGGTAGCCCTCAGTGCAACTACACACAACTTTGTTGTGCGCGGAATTTTTACAGAATTGCTCGCATCGTCCATTT
TTAATGTTGCAGGTGACGTCCAACCTCGCAGTTTTTTCCTTCAAAACAAAAGGGCACCAACACTCGTAGGAATTTAT
ATCGTCTTTACAACCTCCCCCATTCAGACATGGATTAGATTGCGATTGGTCCCCATCGACATATTGCTTCCAGAACT
CAGTGGTCCGTTCTGTATTCTCAACACCTCGCGCGTTCTTCAAACTGCATTTTTCTCCATACACTCTCGCTCC
AAGTTCCCTTGACGAATTCTTCAAGCTTTCCTGAGTTATACCTTTTAGGCCGGTTAAGTATCTTATTCGCGTTTTC
GTGGTCCAGAAAACTGTGGAAACAGGGAGAGAAAAACCACACAACATATTTAAAGATTGATGAAGACAATACTG
TAATATGCTGCTTTTTGTTCTTCTTCACTGACCTAAGAGATCTAGGAACCCCTAGTGATGGAGTTGGCCACTCCC
TCTCTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTC
AGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGCCAA

[0364] P00411全序列(从ITR到ITR):(SEQ ID NO:278)

[0365] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGTCGCCCCAGCCCCG
GGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTA
GATCTCTGATTATTTGGATTAAAAACAAAGACTTTCTTAAGAGATGTAAAATTTTCATGATGTTTTCTTTTTTGCTA
AACTAAAGAATTATTCTTTTACATTTTCAGTTTTTCTTGATCATGAAAACGCCAACAAAATTCGAATCGGCCAAAG
AGGTATAATTACAGGTAAATTGGAAGAGTTTGTTCAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTT
TGAAGAAGCACGAGAAGTTTTGAAAACACTGAAAAGCAACTGAATTTTGAAGCAGTATGTTGATGGAGATCAGT
GTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGA
TTTGAAGGAAAGAACTGTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAG
TGCTGATAACAAGGTGGTTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAG
TGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTCTCTGATGTG
GACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCAC
TCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGATGCAT
TCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAAGTCTGCCCCTGTGTTGAACTGGTGTAAAAATTACA
GTTGTGCGCAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGCAATTATTCCTCA
CCACAACATAATGCAGCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCCTTAGTGCTAA
ACAGCTACGTTACACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCTCAAATTTGGATCTGGCTATGTA
AGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCG
AGCCACATGTCTTCTATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAG

ATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGGAAGGGACCAGTTTCTTAACTGGAATTATTAGC
TGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGGTCTCCCGGTATGTCAACTGGATTAAGGA
AAAAACAAAGCTCACTGTCAGCGGATGGAGACTGTTCAAGAAGATCAGCTAACCTCGACTGTGCCTTCTAGTTGCCA
GCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCCTGGAAGGTGCCACTCCCCTGTCTTTCTTAATAAA
ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGG
GAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTG
GGGCTCTAGGGGTATCCCCAAAAAACCTCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGT
TGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAAATAAGCAATAGCATCACAAATTCACAAATAAAGCATTTT
TTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTTAGGAAATCTTCTTAAAC
AGCCGCCAGCCGCTCACGGTGAGCTTAGTCTTTTCTTTTATCCAATTCACGTAGCGAGAGACCTTCGTATAGATGCC
ATATTTCCCCTTCATCGCACATTCTCCCCCAACTTATTATCCCGTCAAGAACTTGTTCTTCGACTTCAGTGA
CGTGTGGTCCACCTGAATCACCTTGGCATGAGTCGCGACCGCCCTCGTGAAACCCAGCACAAAACATGTTATTGTAA
ATCGTAAATTTCTGTTGACAGAAGACAGGTCGCTCTATCGACCAACGGGACGCGCAAATATTGCAGAACGAGGGCTGA
TCGACCTTTGTGGAAGACCCGCCCCCACCCTCACATATCCGCTCCCAAATTTCAAGAAGATATTTGTATATTCTT
TATCGGCTATACAAATCGGGGTAAACATAGGAGTTAAGTACGAGTGGCTCGTCCAGCTCCAGGAGGGCTATATCATGG
TTGTACTTGTTTATAGCGGCATTATAATTGTGATGGGGTATGATCCTGATAACATTCCTTTTCTGTTTCACTATGCTC
AGTTTCTTCAATGTTGTGTTTCCGACCCACGACCGTAATCTTAACCCCGTCTCGACACAGTGTGCGGCCGTTACAA
TCCACTTTTCATTGACTATGGAGCCCCCAGAAACGCGTCGACTTTTCCGTTGAGCACCACCTGCCATGGAAATTGG
CCAGGTTTAGCGTCTCGCCCCGACAACCTAGTAAAGTCATTAATGACTGTGTGGATTGTGTTATATTATCAAG
AATCGTTTCGGCTTCAGTAGAGTTAACGTAGTCCACATCGGGAAAACTGTCTCGGCCCTTGTCAACTTTGATGTCT
GGGACACACTTACCCGACCGCACGGGAAGGGACCGCCGGTTCACAGCTCTTTTGATTCTCAGCGAGCCGGTAGCCC
TCAGTGCAACTACACACAACCTTTGTTGTGCGCGGAATTTTACAGAATTGCTCGCATCGTCCATTTTAAATGTTGCA
GGTGACGTCCAACCTCGCAGTTTTTCTTCAAAAACCAAAAGGGCACCAACACTCGTAGGAATTTATATCGTCTTTAC
AACTCCCCCATTCAGACATGGATTAGATTTCGCATTGGTCCCCATCGACATATTGCTTCCAGAACTCAGTGGTCCGT
TCTGTATTCTCAAAACACCTCGCGCGCTTCTTCAAACTGCATTTTCCCTCCATACACTCTCGCTCCAAGTTCCCTTG
CACGAATTCTTCAAGCTTTCTGAGTTATACCTTTTAGGCCGGTTAAGTATCTTATTCGCGTTTTCTGTTGTCAGAA
AACTGAAATGTAAAAGAATAATTCTTTAGTTTTAGCAAAAAAGAAAACATCATGAAAATTTACATCTCTTAAGAA
AGTCTTTGTTTTTAATCCAAATAATCAGAGATCTAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGC
TCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCTCAGTGAGCGAGC
GAGCGCGCAGAGAGGGAGTGGCCAA

[0366] P00415全序列(从ITR到ITR):(SEQ ID NO:279)

[0367] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCC
GGGCTTTGCCCCGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCC
TAGATCTCTTAGGTCACTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCATCAATCTTTAAATA
TGTTGTGTGGTTTTTCTCTCCCTGTTTCCACAGTTTTTCTTGATCATGAAAACGCCAACAAAATCTGAATCGGCC
AAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAGGGAACTTGAGAGAGAATGTATGGAAGAAAAGTGT
AGTTTTGAAGAAGCAGTATTCATTTGGAGGACTTTGTGCGTGACTGGAGGCAAACCGCTGGTTATAATCTCGACC
AAGTACTGGAACAGGGCGGGGTAAGTTCCCTCTTTCAGAATTTGGGTGTAAGCGTCACACCAATCCAGCGGATTGT
GTTGTCTGGAGAGAACGGACTCAAAATTGACATCCATGTTATCATTCCATATGAAGGTCTCAGTGGAGACCAAATG

GGGCAGATCGAGAAGATTTTCAAGGTAGTTTACCCAGTCGACGATCACCATTCAAAGTCATTCTCCACTATGGCA
CACTTGTTATCGACGGAGTAACCTCCTAATATGATTGATTACTTTGGTCGCCCCGTATGAGGGCATCGCAGTGTTGA
TGGCAAAAAGATCACCGTAACAGGAACGTTGTGGAATGGGAACAAGATAATCGACGAGAGATTGATAAATCCAGAC
GGGTCACTCCTGTTTCAGGGTTACAATTAACGGCGTCACAGGATGGAGACTCTGTGAACGAATACTGGCCACAAATT
TTTCACTCCTGAAGCAGGCCGGAGACGTGGAGGAAAACCCAGGGCCCGTGAGCAAGGGCGAGGAGCTGTTACCGG
GGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC
GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCCCTCG
TGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTC
CGCCATGCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG
GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
TGGGGCACAAGCTGGAGTACAATAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAA
GGTGAACCTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACG
AGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTGTACAA
GGGAGGAGGAAGCCCGAAGAAGAAGAGAAAGGTCTAACCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTCTCCTAATAAAATGAGGAAATTGC
ATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAA
GACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGG
GGTATCCCCAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGTTGTTAACTTG
TTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTTTTTTCACTGC
ATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTTACACCTTCCTCTTCTTCTGGGGCT
GCCGCCGCCCTTGACAGCTCGTCCATGCCAGGGTGATGCCGGCGGCGGTACGAACTCCAGCAGCACCATGTGG
TCCCTCTTCTCGTTGGGGTTCCTTGCTCAGGGCGCTCTGGGTGCTCAGGTAGTGGTTGTGGGCAGCAGCAGGGGC
CGTCGCCGATGGGGGTGTTCTGCTGGTAGTGGTCGGCCAGCTGCACGCTGCCGTCCCTCGATGTTGTGCTGATCTT
GAAGTTCACCTTGATGCCGTTCTTCTGCTTGTCGGCCATGATGTACACGTTGTGGCTGTTGTAGTTGTACTCCAGC
TTGTGGCCCAGGATGTTGCCGTCTCCTTGAAGTCGATGCCCTTCAGCTCGATCCTGTTACCAGGGTGTGCCCT
CGAACTTCACCTCGGCCCTGGTCTTGTAGTTGCCGTCTCCTTGAAGAAGATGGTCCTCTCCTGCACGTAGCCCTC
GGGCATGGCGCTCTTGAAGAAGTCGTGCTGCTTCATGTGGTCGGGGTACCTGCTGAAGCACTGCACGCCGTAGGTC
AGGGTGGTCACCAGGGTGGGCCAGGGCACGGGCAGCTTGCCGGTGGTGCAGATGAACCTCAGGGTCAGCTTGCCGT
AGGTGGCGTCGCCCTCGCCCTCGCCGCTCACGCTGAACCTTGTGGCCGTTACGTCGCCGTCCAGCTCCACCAGGAT
GGGCACCACGCCGGTGAACAGCTCCTCGCCCTTGCTCACGGGGCCGGGGTTCTCCTCCACGTCGCCGGCCTGCTTC
AGCAGGCTGAAGTTGGTGGCCAGGATCCTCTCGCACAGCCTCCAGCCGGTCACGCCGTTGATGGTCACCCTGAACA
GCAGGCTGCCGTGGGGTTGATCAGCCTCTCGTCGATGATCTTGTGCGGTTCCACAGGGTGGCGGTACGGTGAT
CTTCTTGCCGTGGAACACGGCGATGCCCTCGTAGGGCTGCCGAAGTAGTCGATCATGTTGGGGGTACGCCGTG
ATCACCAGGGTGCCGTAGTGCAGGATCACCTTGAAGTGGTGGTCGTCCACGGGGTACACCACCTTGAATATCTTCT
CGATCTGGCCCATCTGGTCGCCGCTCAGGCCCTCGTAGGGGATGATCACGTGGATGTCGATCTTACGGCCGTTCTC
GCCGCTCAGCACGATCCTCTGGATGGGGGTACGCTCACGCCAGGTTCTGGAACAGGCTGCTCACGCCGCCCTGC
TCCAGCACCTGGTCCAGGTTGTAGCCGGCGGTCTGCCTCCAGTCGCCACGAAGTCTCCAGGGTGAACACGGCCT
CCTCGAAGCTGCACTTCTCCTCCATGCACTCCCTCTCCAGGTTGCCCTGCACGAACTCCTCCAGCTTGCCGCTGTT

GTACCTCTTGGGCCTGTTTCAGGATCTTGTGGCGTTCTCGTGGTCCAGGAAP00418全序列(从ITR到ITR):
(SEQ ID NO:280)

[0368] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCACGCCCG
GGCTTTGCCCCGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTA
GATCTCTTAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCATCAATCTTTAAATATGT
TGTGTGGTTTTTCTCTCCCTGTTTCCACAGTTTTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAG
AGGTATAATTACAGGTAAATTGGAAGAGTTTGTTCAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTT
TGAAGAAGCACGAGAAGTTTTTGAAAACTGAAAGAACAACCTGAATTTTGAAGCAGTATGTTGATGGAGATCAGT
GTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGA
TTTGAAGGAAAGAACTGTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAG
TGCTGATAACAAGGTGGTTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAG
TGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTCTCTGATGTG
GACTATGTAAATTCTACTGAAGCTGAAACCATTTTGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCAC
TCGGGTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGATGCAT
TCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAAGTGTGCTGCCACTGTGTTGAAACTGGTGTAAAAATTACA
GTTGTGCGCAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGAATTATTCTCA
CCACAACCTACAATGCAGCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCCTTAGTGCTAA
ACAGCTACGTTACACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTA
AGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCG
AGCCACATGTCTTCTATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAG
ATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGAAGGGACCAGTTTCTTAACTGGAATTATTAGC
TGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGGTCTCCCGGTATGTCAACTGGATTAAGGA
AAAAACAAAGCTCACTGTCAGCGGATGGAGACTGTTCAAGAAGATCAGCTAACCTCGACTGTGCCTTCTAGTTGCCA
GCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTCTCTAATAAA
ATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGG
GAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTG
GGGCTCTAGGGGTATCCCCAAAAACCTCCACACCTCCCCCTGAACCTGAAACATAAAATGAATGCAATTGTTGT
TGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTACAAATAAAGCATTTT
TTTCACTGCATTCTAGTTGTGGTTTGTCCAACTCATCAATGTATCTTATCATGTCTGTTAGGAAATCTTCTTAAAC
AGCCGCCAGCCGCTACGGTGAGCTTAGTCTTTTCTTTTATCCAATTCACGTAGCGAGAGACCTTCGTATAGATGCC
ATATTTCCCCTTCATCGCACATTCTCCCCCAACTTATTATCCCGGTCAAGAACTTGTTCCCTTCGACTTCAGTGA
CGTGTGGTCCACCTGAATCACCTTGGCATGAGTCGCGACCGCCCTCGTGAAACCCAGCACAAAACATGTTATTGTAA
ATCGTAAATTTCTGTGGACAGAAGACAGGTCGCTCTATCGACCAACGGGACGCGCAAATATTGCAGAACGAGGGCTGA
TCGACCTTTGTGGAAGACCCGCCCCCACCCTCACATATCCGCTCCCAAATTTCAAGAAGATATTTGTATATTCTT
TATCGGCTATACAAATCGGGGTAACATAGGAGTTAAGTACGAGTGGCTCGTCCAGCTCCAGGAGGGCTATATCATGG
TTGTACTTGTTTATAGCGGCATTATAATTGTGATGGGGTATGATCCTGATAACATTCCTTTTCTGTTCACTATGCTC
AGTTTCTTCAATGTTGTGTTGCGCCAGCCACGACCGTAATCTTAACCCCGTCTCGACACAGTGTGCGGCCGTTACAA
TCCACTTTTTCATTGACTATGGAGCCCCACAAAACGCGTCGACTTTTCCGTTGAGCACCACCTGCCATGGAAATTGG
CCAGGTTTAGCGTCTCGCCCCGACAACCTAGTAAAGTCATTAAATGACTGTGTGGATTGTGTTATATTATCAAG

AATCGTTTCGGCTTCAGTAGAGTTAACGTAGTCCACATCGGGAAAACTGTCTCGGCCCTTGTCAACTTTGATGTCT
GGGACACACTTACCCGACCGCACGGGAAGGGCACCGCCGGTTCACAGCTCTTTTGATTCTCAGCGAGCCGGTAGCCC
TCAGTGCAACTACACACAACCTTTGTTGTCGGCGGAATTTTTACAGAATTGCTCGCATCGTCCATTTTTAATGTTGCA
GGTGACGTCCAACCTCGCAGTTTTTCTTCAAAAACCAAAAGGGCACCAACACTCGTAGGAATTTATATCGTCTTTAC
AACTCCCCCATTCAGACATGGATTAGATTGCGATTGGTCCCCATCGACATATTGCTTCCAGAACTCAGTGGTCCGT
TCTGTATTCTCAAACACCTCGCGCGCTTCTTCAAACTGCATTTTTCTCCATACACTCTCGCTCCAAGTTCCCTTG
CACGAATTCTTCAAGCTTCTGAGTTATACCTTTTAGGCCGGTTAAGTATCTTATTGCGTTTTTCGTGGTCCAGAA
AACTGTGGAACAGGGAGAGAAAAACACACAACATATTTAAAGATTGATGAAGACAATACTGTAATATGCTGC
TTTTTGTTCTTCTTCACTGACCTAAGAGATCTAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGC
TCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCTCAGTGAGCGAGC
GAGCGCGCAGAGAGGGAGTGGCCAA

[0369] P00123全序列(从ITR到ITR):(SEQ ID NO:281)

[0370] GGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCGGG
CTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTGA
GGGGTGGAGTCGTGATAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCATCAATCTTT
AAATATGTTGTGTGGTTTTTCTCTCCCTGTTTCCACAGTTTTTCTTGATCATGAAAACGCCAACAATAATTCTGAATC
GGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAGGGAACTTGAGAGAGAATGTATGGAAGAAAAG
TGATGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGAAAGAACAATACTGAATTTTGAAGCAGTATGTTGATGG
AGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTC
CCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGT
AAAAATAGTGCTGATAACAAGGTGGTTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGA
ACCAGCAGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAACTTCTAAGCTCAGCCGTGCTGAGACTGTTTTTC
CTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCAATCATTTAAT
GACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCTTGGCAGGTTGTTTTGAATGGTAAAGT
TGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAAGTGTGCTGCCCAGTGTGTTGAAACTGGTGTTA
AAATTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGCAATT
ATTCTCACCACAACATAATGCAGCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCCTT
AGTGCTAAACAGCTACGTTACACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCTCAAATTTGGATCTG
GCTATGTAAGTGGCTGGGGAAGAGTCTTCCACAAAGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTT
GTTGACCGAGCCACATGTCTTCTATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGG
AGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGAAGGGACAGTTTCTTAACTGGAA
TTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAAATATGGAATATATACCAAGGTATCCCGGTATGTCAACTGG
ATTAAGGAAAAACAAAGCTCACTTAACCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCC
GTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTTCTAATAAAATGAGGAAATTGCATCGCATTGTCT
GAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGC
ATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCACTAG
TCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCG
GGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGA

[0371] P00204全序列(从ITR到ITR):(SEQ ID NO:282)

[0372] GGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCAGCCCCGGGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTGAGGGGTGGAGTCGTGACCTAGGTCGTCTCCGGCTCTGCTTTTTCCAGGGGTGTGTTTCGCCGAGAAGCACGTAAGAGTTTATGTTTTTTCATCTCTGCTTGTATTTTTCTAGTAATGGAAGCCTGGTATTTTAAAATAGTTAAATTTTCCTTAGTGCTGATTTCTAGATTATTACTGTTGTTGTTGTTATTATTGTCATTATTTGCATCTGAGAACTAGGTCAGTAGAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCATCAATCTTTAAATATGTTGTGTGGTTTTTCTCTCCCTGTTTTCCACAGTTTTTCTTGATCATGAAAACGCCAACAAAATCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTGAAAACACTGAAAGAACAAGTGAATTTTGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTGGTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTCGGGTGTGGTGGAGAAGATGCCAAACCAGGTCAATTCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAAGTGTGCCCCACTGTGTTGAACTGGTGTAAAAATTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGCAATTATTCCTCACCACAACATAATGCAGCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTGACCGAGCCACATGTCTTCTATCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATGTCAAGGAGATGTGGGGGACCCCATGTTACTGAAGTGAAGGGACCAGTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATATGGAATATATACCAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAACAAAGCTCACTTAACCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCACTCCCCTGTCTTTCTTAATAAAATGAGGAAATGTCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCCTTAGGTGGTTATATTATTGATATATTTTGTATCTTTGATGACAATAATGGGGGATTTTGAAAGCTTAGCTTTAAATTTCTTTTAATTAATAAAAAAATGCTAGGCAGAATGACTCAAATTACGTTGGATACAGTTGAATTTATTACGGTCTCATAGGGCCTGCCTGCTCGACCATGCTATACTAAAAATTAAGTGTACTAGTCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCGCCGACGCCCCGGGCTTTGCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGA

[0373] P00353全序列(从ITR到ITR):(SEQ ID NO:283)

[0374] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCAGCCCCGGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTAGATCTGATTTTGAAAGCTTAGCTTTAAATTTCTTTTAATTAATAAAAAAATGCTAGGCAGAATGACTCAAATTACGTTGGATACAGTTGAATTTATTACGGTCTCATAGGGCCTGCCTGCTCGACCATGCTATACTAAAAATTAAGTGTGTGTACTAATTTTATAAATGGAGTTTCCATTTATTTACCTTTATTTCTTATTTACCATTTGTCTTAGTAGATATTTACAACATGACAGAAACACTAAATCTTGAGTTTGAATGCACAGATATAAACTTAACGGGTTTTAAAAATAATAATGTTGGTGAAAAAATATAACTTTGAGTGTAGCAGAGAGGAACCATTGCCACCTTCAGATTTTCTGTAAACGATCGGGAAC

GGCATCTTCAGGGAGTAGCTTAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCATCAA
TCTTTAAATATGTTGTGTGGTTTTTCTCTCCCTGTTTCCACAGTTTTTCTTGATCATGAAAACGCCAACAAAATTCT
GAATCGGCCAAAGAGGTTTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTACAGGTA
AATTGGAAGAGTTTGTTCAAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAA
GTTTTTGAAAACACTGAAAAGAACAACCTGAATTTTGGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATG
TTTAAATGGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACT
GTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTG
GTTTGCTCCTGTACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAGTGCCATTTCCATGTGG
AAGAGTTTCTGTTTCACAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCTTGATGTGGACTATGTAAATTCTA
CTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTCGGGTTGTTGGTGGGA
GAAGATGCCAAACCAGGTCAATTCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTAT
CGTTAATGAAAAATGGATTGTAAGTGTGCCCCTGTGTTGAACTGGTGTAAAATTACAGTTGTGCGCAGGTGAAC
ATAATATTGAGGAGACAGAACATACAGAGCAAAAGCGAAATGTGATTGCAATTATTCCTCACCACAACCTACAATGCA
GCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAAGTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACC
TATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAG
TCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGAGCCACATGTCTTCTA
TCTACAAAGTTCACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATGTCAAGGAGA
TAGTGGGGGACCCCATGTTACTGAAGTGAAGGGACCAGTTTCTTAAGTGAATTATTAGCTGGGGTGAAGAGTGTG
CAATGAAAGGCAAATATGGAATATATACCAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAAACAAAGCTCACT
TAACCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGT
GCCACTCCCCTGTCTTTCTAATAAAATGAGGAAATGTCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGG
GGGTGGGGTGGGGCAGGACAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTA
TGGCTTCTGAGGCGGAAAGAACCAGCTGGGGCTCTAGGGGGTATCCCCGTGAGATCGCCCATCGGTATAATGATTTG
GGAGAACAACATTTCAAAGGCCTGTAAGTTATAATGCTGAAAGCCCACTTAATATTTCTGGTAGTATTAGTTAAAGT
TTTAAACACCTTTTTCCACCTTGAGTGTGAGAATTGTAGAGCAGTGCTGTCCAGTAGAAATGTGTGCATTGACAGA
AAGACTGTGGATCTGTGCTGAGCAATGTGGCAGCCAGAGATCACAAGGCTATCAAGCACTTGCACATGGCAAGTGT
AACTGAGAAGCACACATTCAAATAATAGTTAATTTTAATTGAATGTATCTAGCCATGTGTGGCTAGTAGCTCCTTTC
CTGGAGAGAGAATCTGGAGCCACATCTAACTTGTTAAGTCTGGAATCTTATTTTTTATTTCTGGAAAGGTCTATGA
ACTATAGTTTTGGGGCAGCTCACTTACTAACTTTAATGCAATAAGAATCTCATGGTATCTTGAGAACATTATTTT
GTCTCTTTGTAGATCTAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGC
CGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAG
TGGCCAA

[0375] P00354全序列(从ITR到ITR):(SEQ ID NO:284)

[0376] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCGACGCCG
GGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTA
GATCTTAGCCTCTGGCAAAATGAAGTGGGTAAACCTTTCTCCTCCTCCTCTTCGTCTCCGGCTCTGCTTTTTCCAGGG
GTGTGTTTCGCCGAGAAGCACGTAAGAGTTTTATGTTTTTTCATCTCTGCTTGTATTTTTCTAGTAATGGAAGCCTG
GTATTTTAAATAGTTAAATTTTCTTTAGTGCTGATTTCTAGATTATTATTACTGTTGTTGTTGTTATTATTGTCA
TTATTTGCATCTGAGAACCCTTAGGTGGTTATATTATTGATATATTTTTGGTATCTTTGATGACAATAATGGGGGAT

TTTGAAAGCTTAGCTTTAAATTTCTTTTAATTAATAAAAAAATGCTAGGCAGAATGACTCAAATTACGTTGGATACAG
TTGAATTTATTACGGTCTCATAGGGCTGCCTGCTCGACCATGCTATACTAAAAATTAAAAGTGTGTGTTACTAATT
TTATAAATGGAGTTTCCATTTATATTTACCTTTATTTCTTATTTACCATTGTCTTAGTAGATATTTACAAACATGAC
AGAAACACTAAATCTTGAGTTTGAATGCACAGATATAAACACTTAACGGGTTTTAAAAATAATAATGTTGGTGAAAA
AATATAACTTTGAGTGTAGCAGAGAGGAACCATTGCCACCTTCAGATTTTCCCTGTAACGATCGGGAACCTGGCATCTT
CAGGGAGTAGCTTAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCATCAATCTTTAAA
TATGTTGTGTGGTTTTTCTCTCCCTGTTTCCACAGTTTTTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGC
CAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTTCAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGT
AGTTTTGAAGAAGCACGAGAAGTTTTTGAAGAACTGAAAGAACAACTGAATTTTGAAGCAGTATGTTGATGGAGA
TCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTTGCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCT
TTGGATTTGAAGGAAAGAACTGTGAATTAGATGTAACATGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAA
AATAGTGCTGATAACAAGGTGGTTTGTCTCTGACTGAGGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACC
AGCAGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCACAACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCTGT
ATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATTTTGGATAACATCACTCAAAGCACCCAATCATTTAATGAC
TTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGGTCAATTCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGA
TGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGATTGTAAGTCTGCCCCTGTGTTGAACTGGTGTAAAA
TTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACAGAACATACAGAGCAAAGCGAAATGTGATTGCAATTATT
CCTCACCACAACCTACAATGCAGCTATTAATAAGTACAACCATGACATTGCCCTTCTGGAACCTGGACGAACCTTAGT
GCTAAACAGCTACGTTACACCTATTTGCATTGCTGACAAGGAATACACGAACATCTTCCTCAAATTTGGATCTGGCT
ATGTAAGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGATCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTT
GACCGAGCCACATGTCTTCTATCTACAAAGTTTACCATCTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGG
TAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATGTTACTGAAGTGAAGGGACCAGTTTCTTAAGTGAATTA
TTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAAATATGGAATATATACCAAGGTATCCCGGTATGTCAACTGGATT
AAGGAAAAAACAAGCTCACTTAACCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTG
CCTTCCTTGACCCTGGAAGGTGCCACTCCCCTGTCTTTTCCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAG
TAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACAGCAAGGGGAGGATTGGGAAGACAATAGCAGGCATG
CTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAACAGCTGGGGCTCTAGGGGTATCCCCGTGAGATC
GCCCCTCGGTATAATGATTTGGGAGAACAACATTTCAAAGGCCTGTAAGTTATAATGCTGAAAGCCCACTTAATATT
TCTGGTAGTATTAGTTAAAGTTTTAAAAACACCTTTTTCCACCTTGAGTGTGAGAATTGTAGAGCAGTGCTGTCCAGT
AGAAATGTGTGCATTGACAGAAAGACTGTGGATCTGTGCTGAGCAATGTGGCAGCCAGAGATCACAAGGCTATCAAG
CACTTTGCACATGGCAAGTGTAAGTGAAGACACACATTCAAATAATAGTTAATTTTAATTGAATGTATCTAGCCAT
GTGTGGCTAGTAGCTCCTTTTCTGGAGAGAGAATCTGGAGCCACATCTAACTTGTTAAGTCTGGAATCTTATTTTT
TATTTCTGGAAAGGTCTATGAACTATAGTTTTGGGGGCAGCTCACTTACTAACTTTTAATGCAATAAGAATCTCATG
GTATCTTGAGAACATTATTTTGTCTCTTTGTAGTACTGAAACCTTATACATGTGAAGTAAGGGGTCTATACTTAAGT
CACATCTCCAACCTTAGTAATGTTTTAATGTAGTAAAAAATGAGTAATTAATTTATTTTGAAGGTCAATAGTAT
CATGTATTCCAAATAACAGAGGTATATGGTTAGAAAAAGAAACAATTCAAAGGACTTATATAATATCTAGCCTTGACA
ATGAATAAATTTAGAGAGTAGTTTGCCTGTTTGCCTCATGTTCATAAATCTATTGACACATATGTGCATCTGCACTT
CAGCATGGTAGAAGTCCATATTAGATCTAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCT
CGCTCACTGAGGCCGCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCG

CGCAGAGAGGGAGTGGCCAA

[0377] P00350:300/600bp HA F9构建体(对于G551) (SEQ ID NO:285)

[0378] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCAGCCCCG
GGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCCTA
GATCTAAGTATATTAGAGCGAGTCTTTCTGCACACAGATCACCTTTCCTATCAACCCCACTAGCCTCTGGCAAAATG
AAGTGGGTAACCTTTCTCCTCCTCTCTCGTCTCCGGCTCTGCTTTTTCCAGGGGTGTGTTTCGCCGAGAAGCACG
TAAGAGTTTTATGTTTTTTCATCTCTGCTTGATTTTTCTAGTAATGGAAGCCTGGTATTTTAAATAGTTAAATTT
TCCTTTAGTGCTGATTTCTAGATTATTACTGTTGTTGTTGTTATTATTGTCATTATTTGCATCTGAGAACCTTT
TTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTT
CAAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTGAAGAAGCACGAGAAGTTTTTGAAGAACTGA
AAGAACAACCTGAATTTTGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTT
GCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATGTAACA
TGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAATAATAGTGCTGATAACAAGGTGGTTTGCTCCTGTACTGA
GGGATATCGACTTGCAGAAAACCAGAAGTCTGTGAACCAGCAGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCAC
AACTTCTAAGCTCACCCGTGCTGAGACTGTTTTCTCTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATT
TTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGG
TCAATTCCTTGGCAGGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGA
TTGTAAGTGTGCCCCTGTGTTGAACTGGTGTAAAAATTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACA
GAACATACAGAGCAAAAGCGAAATGTGATTTCGAATTATTCCTCACCACAACCTACAATGCAGCTATTAATAAGTACAA
CCATGACATTGCCCTTCTGGAACCTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACCTATTTGCATTGCTGACA
AGGAATACACGAACATCTTCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGA
TCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGAGCCACATGTCTTCTATCTACAAAGTTCACCAT
CTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGACCCCATG
TTACTGAAGTGAAGGGACCAGTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATAT
GGAATATATACCAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAACAAAGCTCACTTAACCTCGACTGTGCCT
TCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCT
TTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG
ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAA
AGAACCAGCTGGGGCTCTAGGGGGTATCCCCCTTAGGTGGTTATATTATTGATATATTTTTGGTATCTTTGATGACA
ATAATGGGGGATTTTGAAGCTTAGCTTTAAATTTCTTTTAATTAATAAAAAAATGCTAGGCAGAATGACTCAAATTA
CGTTGGATACAGTTGAATTTATTACGGTCTCATAGGGCCTGCCTGCTCGACCATGCTATACTAAAAATTAAGTGT
GTGTTACTAATTTTATAATGGAGTTTCCATTTATTTACCTTTATTTCTTATTTACCATTGTCTTAGTAGATATT
TACAAACATGACAGAAACACTAAAGATCTAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCT
CGCTCACTGAGGCCGCCGGGCAAAGCCCGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCG
CGCAGAGAGGGAGTGGCCAA

[0379] P00356:300/2000bp HA F9构建体(对于G551) (SEQ ID NO:286)

[0380] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCAGCCCCG
GGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCCTA
GATCTAAGTATATTAGAGCGAGTCTTTCTGCACACAGATCACCTTTCCTATCAACCCCACTAGCCTCTGGCAAAATG

AAGTGGGTAACCTTTCTCCTCCTCCTCTTCGTCTCCGGCTCTGCTTTTTCCAGGGGTGTGTTTCGCCGAGAAGCACG
TAAGAGTTTTATGTTTTTTCATCTCTGCTTGATTTTTCTAGTAATGGAAGCCTGGTATTTTAAATAGTTAAATTT
TCCTTTAGTGCTGATTTCTAGATTATTATTACTGTTGTTGTTGTTATTATTGTCATTATTTGCATCTGAGAACCTTT
TTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGTT
CAAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTGAAAACACTGA
AAGAACAACCTGAATTTTGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTT
GCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATGTAACA
TGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTGGTTTGCTCCTGTACTGA
GGGATATCGACTTGAGAGAAAACAGAAGTCTGTGAACCAGCAGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCAC
AACTTCTAAGCTCACCCGTGCTGAGACTGTTTTCTGTATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATT
TTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGG
TCAATTCCTTGCGCAGGTTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGA
TTGTAACCTGCTGCCACTGTGTTGAAACTGGTGTTAAATTTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACA
GAACATACAGAGCAAAAGCGAAATGTGATTGCAATTATTCCTCACCACAACCTACAATGCAGCTATTAATAAGTACAA
CCATGACATTGCCCTTCTGGAAGTGGACGAACCCCTTAGTGCTAAACAGCTACGTTACACCTATTTGCATTGCTGACA
AGGAATACACGAACATCTTCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGA
TCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTTGTTGACCGAGCCACATGTCTTCTATCTACAAAGTTTACCAT
CTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATG
TTACTGAAGTGAAGGGACAGTTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATAT
GGAATATATACCAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAACAAAGCTCACTTAACCTCGACTGTGCCT
TCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCGGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTCT
TTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGG
ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAA
AGAACCAGCTGGGGCTCTAGGGGGTATCCCCCTTAGGTGGTTATATTATTGATATATTTTTGGTATCTTTGATGACA
ATAATGGGGGATTTTGAAAGCTTAGCTTTAAATTTCTTTTAAATTAATAAAAAAATGCTAGGCAGAATGACTCAAATTA
CGTTGGATACAGTTGAATTTATTACGGTCTCATAGGGCCTGCCTGCTCGACCATGCTATACTAAAAATTAAGTGT
GTGTTACTAATTTTATAAATGGAGTTTCCATTTATTTTACCTTTATTTCTTATTTACCATTGTCTTAGTAGATATT
TACAAACATGACAGAAACACTAAATCTTGAGTTTGATGCACAGATATAAACACTTAACGGGTTTTAAAAATAATAA
TGTTGGTGAAAAAATATAACTTTGAGTGATGAGAGGAACCATTGCCACCTTCAGATTTTCTGTAAACGATCGGG
AACTGGCATCTTCAGGGAGTAGCTTAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCA
TCAATCTTTAAATATGTTGTGTGGTTTTTCTCTCCCTGTTTCCACAGACAAGAGTGAGATCGCCCATCGGTATAATG
ATTTGGGAGAACAAACATTTCAAAGGCCTGTAAGTTATAATGCTGAAAGCCCACTTAATATTTCTGGTAGTATTAGTT
AAAGTTTTTAAACACCTTTTTCCACCTTGAGTGTGAGAATTGTAGAGCAGTGCTGTCCAGTAGAAATGTGTGCATTG
ACAGAAAGACTGTGGATCTGTGCTGAGCAATGTGGCAGCCAGAGATCACAAGGCTATCAAGCACTTTGCACATGGCA
AGTGTAACCTGAGAAGCACACATTCAAATAATAGTTAATTTTAAATTGAATGTATCTAGCCATGTGTGGCTAGTAGCTC
CTTTCCTGGAGAGAGAATCTGGAGCCCACATCTAACTTGTTAAGTCTGGAATCTTATTTTTTATTTCTGGAAAGGTC
TATGAACTATAGTTTTGGGGGCAGCTCACTTACTAACTTTTAAATGCAATAAGATCCATGGTATCTTGAGAACATTAT
TTTGTCTCTTTGTAGTACTGAAACCTTATACATGTGAAGTAAGGGGTCTATACTTAAGTCACATCTCCAACCTTAGT
AATGTTTTAATGTAGTAAAAAATGAGTAATTAATTTATTTTAGAAGGTCAATAGTATCATGTATTCCAAATAACA

GAGGTATATGGTTAGAAAAGAAACAATTCAAAGGACTTATATAATATCTAGCCTTGACAATGAATAAATTTAGAGAG
TAGTTTGCCTGTTTGCCTCATGTTTCATAAATCTATTGACACATATGTGCATCTGCACTTCAGCATGGTAGAAGTCCA
TATTCCTTTGCTTGGAAGGCAGGTGTTCCCATACGCCTCAGAGAATAGCTGACGGGAAGAGGCTTTCTAGATAGT
TGTATGAAAGATATACAAAATCTCGCAGGTATACACAGGCATGATTTGCTGGTTGGGAGAGCCACTTGCCTCATACT
GAGGTTTTTGTGTCTGCTTTTCAGAGTCCTGATTGCCTTTTCCCAGTATCTCCAGAAATGCTCATACGATGAGCATG
CCAAATTAGTGAGGAAGTAACAGACTTTGCAAAGACGTGTGTTGCCGATGAGTCTGCCGCCAACTGTGACAAATCC
CTTGTGAGTACCTTCTGATTTTGTGGATCTACTTTCCTGCTTTCTGGAACCTCTGTTTCAAAGCCAATCATGACTCCA
TCACTTAAGGCCCCGGGAACACTGTGGCAGAGGGCAGCAGAGAGATTGATAAAGCCAGGGTGATGGGAATTTTCTGT
GGGACTCCATTTTCATAGTAATTGCAGAAGCTACAATACACTCAAAAAGTCTCACCACATGACTGCCCAAATGGGAGC
TTGACAGTGACAGTGACAGTAGATATGCCAAAGTGGATGAGGGAAAGACCACAAGAGCTAAACCCTGTAAAAAGAAC
TGTAGGCAACTAAGGAATGCAGAGAGAAAGATCTAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGC
TCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCGGCCTCAGTGAGCGAGC
GAGCGCGCAGAGAGGGAGTGGCCAA

[0381] P00362:300/1500bp HA F9构建体 (对于G551) (SEQ ID NO:287)

[0382] TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCG
GGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTTA
GATCTAAGTATATTAGAGCGAGTCTTTCTGCACACAGATCACCTTTCCTATCAACCCCACTAGCCTCTGGCAAAATG
AAGTGGGTAACCTTTCTCCTCCTCTCTCGTCTCCGGCTCTGCTTTTCCAGGGGTGTGTTTCGCCGAGAAGCACG
TAAGAGTTTTATGTTTTTTCATCTCTGCTTGATTTTTCTAGTAATGGAAGCCTGGTATTTTAAATAGTTAAATTT
TCCTTTAGTGCTGATTTCTAGATTATTACTGTTGTTGTTGTTATTATTGTCATTATTTGCATCTGAGAACCTTT
TTCTTGATCATGAAAACGCCAACAAAATTCTGAATCGGCCAAAGAGGTATAATTCAGGTAAATTGGAAGAGTTTGT
CAAGGGAACCTTGAGAGAGAATGTATGGAAGAAAAGTGTAGTTTTGAAGAAGCACGAGAAGTTTTTGAAAACACTGA
AAGAACAACCTGAATTTTGAAGCAGTATGTTGATGGAGATCAGTGTGAGTCCAATCCATGTTTAAATGGCGGCAGTT
GCAAGGATGACATTAATTCCTATGAATGTTGGTGTCCCTTTGGATTTGAAGGAAAGAACTGTGAATTAGATGTAACA
TGTAACATTAAGAATGGCAGATGCGAGCAGTTTTGTAAAAATAGTGCTGATAACAAGGTGGTTTGCTCCTGTACTGA
GGGATATCGACTTGCAGAAAACCAGAAGTCCTGTGAACCAGCAGTGCCATTTCCATGTGGAAGAGTTTCTGTTTCAC
AACTTCTAAGCTCACCCGTGCTGAGACTGTTTTTCTGATGTGGACTATGTAAATTCTACTGAAGCTGAAACCATT
TTGGATAACATCACTCAAAGCACCCAATCATTTAATGACTTCACTCGGGTTGTTGGTGGAGAAGATGCCAAACCAGG
TCAATTCCTTGGCAGGTGTTTTGAATGGTAAAGTTGATGCATTCTGTGGAGGCTCTATCGTTAATGAAAAATGGA
TTGTAAGTGTGCCCCTGTGTTGAAACTGGTGTAAATTTACAGTTGTGCGAGGTGAACATAATATTGAGGAGACA
GAACATACAGAGCAAAAGCGAAATGTGATTTCGAATTATTCCTCACCACAACCTACAATGCAGCTATTAATAAGTACAA
CCATGACATTGCCCTTCTGGAAGTGGACGAACCCTTAGTGCTAAACAGCTACGTTACACCTATTTGCATTGCTGACA
AGGAATACACGAACATCTTCTCAAATTTGGATCTGGCTATGTAAGTGGCTGGGGAAGAGTCTTCCACAAAGGGAGA
TCAGCTTTAGTTCTTCAGTACCTTAGAGTTCCACTGTTGACCGAGCCACATGTCTTCTATCTACAAAGTTTACCAT
CTATAACAACATGTTCTGTGCTGGCTTCCATGAAGGAGGTAGAGATTCATGTCAAGGAGATAGTGGGGGACCCCATG
TTACTGAAGTGAAGGGGACCAGTTTCTTAACTGGAATTATTAGCTGGGGTGAAGAGTGTGCAATGAAAGGCAAATAT
GGAATATATACCAAGGTATCCCGGTATGTCAACTGGATTAAGGAAAAACAAAGCTCACTTAACCTCGACTGTGCCT
TCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCGTGCCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCT
TTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG

ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAA
AGAACCAGCTGGGGCTCTAGGGGGTATCCCCCTTAGGTGGTTATATTATTGATATATTTTTGGTATCTTTGATGACA
ATAATGGGGGATTTTGAAAGCTTAGCTTTAAATTTCTTTTAATTAATAAAAAAATGCTAGGCAGAATGACTCAAATTA
CGTTGGATACAGTTGAATTTATTACGGTCTCATAGGGCTGCCTGCTCGACCATGCTATACTAAAAATTAAAGTGT
GTGTTACTAATTTTATAAATGGAGTTTCCATTTATATTTACCTTTATTTCTTATTTACCATTGTCTTAGTAGATATT
TACAAACATGACAGAAACACTAAATCTTGAGTTTGAATGCACAGATATAAACACTTAACGGGTTTTAAAAATAATAA
TGTTGGTGAAAAAATATAACTTTGAGTGTAGCAGAGAGGAACCATTGCCACCTTCAGATTTTCTGTAAACGATCGGG
AACTGGCATCTTCAGGGAGTAGCTTAGGTCAGTGAAGAGAAGAACAAAAAGCAGCATATTACAGTTAGTTGTCTTCA
TCAATCTTTAAATATGTTGTGTGGTTTTTCTCTCCCTGTTTCCACAGACAAGAGTGAGATCGCCCATCGGTATAATG
ATTTGGGAGAACAAACATTTCAAAGGCCTGTAAGTTATAATGCTGAAAGCCCACTTAATATTTCTGGTAGTATTAGTT
AAAGTTTTTAAACACCTTTTTCCACCTTGAGTGTGAGAATTGTAGAGCAGTGCTGTCCAGTAGAAATGTGTGCATTG
ACAGAAAGACTGTGGATCTGTGCTGAGCAATGTGGCAGCCAGAGATCACAAGGCTATCAAGCACTTTGCACATGGCA
AGTGTAACTGAGAAGCACACATTCAAATAATAGTTAATTTTAATTGAATGTATCTAGCCATGTGTGGCTAGTAGCTC
CTTTCCTGGAGAGAGAATCTGGAGCCACATCTAACTGTTAAGTCTGGAATCTTATTTTTTATTTCTGGAAAGGTC
TATGAACATAGTTTTGGGGGCAGCTCACTTACTAACTTTTAATGCAATAAGATCCATGGTATCTTGAGAACATTAT
TTTGTCTCTTTGTAGTACTGAAACCTTATACATGTGAAGTAAGGGGTCTATACTTAAGTCACATCTCCAACCTTAGT
AATGTTTTAATGTAGTAAAAAATGAGTAATTAATTTATTTTTAGAAAGTCAATAGTATCATGTATTCCAAATAACA
GAGGTATATGGTTAGAAAAGAAACAATTCAAAGGACTTATATAATATCTAGCCTTGACAATGAATAAATTTAGAGAG
TAGTTTGCTGTTTGCTCATGTTTATAAATCTATTGACACATATGTGCATCTGCACTTCAGCATGGTAGAAGTCCA
TATTCCTTTGCTTGGAAGGCAGGTGTTCCCATACGCCTCAGAGAATAGCTGACGGGAAGAGGCTTTCTAGATAGT
TGTATGAAAGATATACAAAATCTCGCAGGTATACACAGGCATGATTTGCTGGTTGGGAGAGCCACTTAGATCTAGGA
ACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGC
GTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAA

[0383] Cas9 ORF (SEQ ID NO:703)

[0384] ATGGATAAGAAGTACTCAATCGGGCTGGATATCGGAACTAATTCCGTGGGTGGGCAGTGATCACGGAT
GAATACAAAGTGCCGTCCAAGAAGTTCAAGGTCCTGGGGAACACCGATAGACACAGCATCAAGAAAAATCTCATCGG
AGCCCTGCTGTTTGACTCCGGCGAAACCGCAGAAGCGACCCGGCTCAAACGTACCGCGAGGCGACGCTACACCCGGC
GGAAGAATCGCATCTGCTATCTGCAAGAGATCTTTTCGAACGAAATGGCAAAGGTCGACGACAGCTTCTTCCACCGC
CTGGAAGAATCTTTCCTGGTGGAGGAGGACAAGAAGCATGAACGGCATCCTATCTTTGGAACATCGTCGACGAAGT
GGCGTACCACGAAAAGTACCCGACCATCTACCATCTGCGGAAGAAGTTGGTTGACTCAACTGACAAGGCCGACCTCA
GATTGATCTACTTGGCCCTCGCCCATATGATCAAATTCGCGGACACTTCCTGATCGAAGGCGATCTGAACCCTGAT
AACTCCGACGTGGATAAGCTTTTCATTCAACTGGTGCAGACCTACAACCAACTGTTCAAGAAAACCCAATCAATGC
TAGCGGCGTCGATGCCAAGGCCATCCTGTCCGCCCGGCTGTCAAGTCGCGGCGCCTCGAAAACCTGATCGCACAGC
TGCCGGGAGAGAAAAAGAACGGACTTTTCGGCAACTTGATCGCTCTCTCACTGGGACTCACTCCCAATTTCAAGTCC
AATTTTGACCTGGCCGAGGACGCGAAGCTGCAACTCTCAAAGGACACCTACGACGACGACTTGGACAATTTGCTGGC
ACAAATTGGCGATCAGTACGCGGATCTGTTCCCTGCCGTAAGAACCTTTCGGACGCAATCTTGCTGTCCGATATCC
TGCGCGTGAACACCGAAATAACCAAAGCGCCGCTTAGCGCCTCGATGATTAAGCGGTACGACGAGCATCACCAGGAT
CTCACGCTGCTCAAAGCGCTCGTGAGACAGCAACTGCCTGAAAAGTACAAGGAGATCTTCTTCGACCAGTCCAAGAA
TGGGTACGCAGGGTACATCGATGGAGGCGCTAGCCAGGAAGAGTTCTATAAGTTCATCAAGCCAATCCTGGAAAAGA

TGGACGGAACCGAAGAAGCTGCTGGTCAAGCTGAACAGGGAGGATCTGCTCCGAAACAGAGAACCTTTGACAACGGA
TCCATTCCCCACCAGATCCATCTGGGTGAGCTGCACGCCATCTTGCGGCGCCAGGAGGACTTTTACCCATTCTCTCAA
GGACAACCGGGAAAAGATCGAGAAAATTCTGACGTTCCGCATCCCGTATTACGTGGGCCCACTGGCGCGCGGCAATT
CGCGCTTCGCGTGGATGACTAGAAAATCAGAGGAAACCATCACTCCTTGAATTTTCGAGGAAGTTGTGGATAAGGGA
GCTTCGGCACAAAGCTTCATCGAACGAATGACCAACTTCGACAAGAATCTCCCAAACGAGAAGGTGCTTCCTAAGCA
CAGCCTCCTTTACGAATACTTCACTGTCTACAACGAACTGACTAAAGTGAAATACGTTACTGAAGGAATGAGGAAGC
CGGCCTTTTCTGTCCGGAGAACAGAAGAAAGCAATTGTGATCTGCTGTTCAAGACCAACCGCAAGGTGACCGTCAAG
CAGCTTAAAGAGGACTACTTCAAGAAGATCGAGTGTTTCGACTCAGTGGAATCAGCGGGGTGGAGGACAGATTCAA
CGCTTCGCTGGGAACCTATCATGATCTCTGAAGATCATCAAGGACAAGGACTTCCTTGACAACGAGGAGAACGAGG
ACATCCTGGAAGATATCGTCTGACCTTGACCTTTTCGAGGATCGCGAGATGATCGAGGAGAGGCTTAAGACCTAC
GCTCATCTCTTCGACGATAAGGTCATGAAACAACCTCAAGCGCCGCCGTACACTGGTTGGGGCCGCCTCTCCCGCAA
GCTGATCAACGGTATTCGCGATAAACAGAGCGGTAAAACTATCCTGGATTTCTCAAATCGGATGGCTTCGCTAATC
GTAACCTTCATGCAATTGATCCACGACGACAGCCTGACCTTTAAGGAGGACATCCAAAAGCACAAGTGTCGGACAG
GGAGACTCACTCCATGAACACATCGCGAATCTGGCCGGTTCGCCGGCGATTAAGAAGGGAATTCTGCAAATGTGAA
GGTGGTCGACGAGCTGGTGAAGGTCATGGGACGGCACAAACCGGAGAATATCGTGATTGAAATGGCCCCGAGAAAACC
AGACTACCCAGAAGGGCCAGAAAACTCCCGCGAAAGGATGAAGCGGATCGAAGAAGGAATCAAGGAGCTGGGCAGC
CAGATCCTGAAAGAGCACCCGGTGGAAAAACGCGAGCTGCAGAACGAGAAGCTCTACCTGTACTATTTGCAAAATGG
ACGGGACATGTACGTGGACCAAGAGCTGGACATCAATCGGTTGTCTGATTACGACGTGGACCACATCGTTCCACAGT
CCTTTCTGAAGGATGACTCGATCGATAACAAGGTGTTGACTCGCAGCGACAAGAACAGAGGGAAGTCAGATAATGTG
CCATCGGAGGAGGTCGTGAAGAAGATGAAGAATTACTGGCGGCAGCTCCTGAATGCGAAGCTGATTACCCAGAGAAA
GTTTGACAATCTCACTAAAGCCGAGCGCGCGGACTCTCAGAGCTGGATAAGGCTGGATTCATCAAACGGCAGCTGG
TCGAGACTCGGCAGATTACCAAGCACGTGGCGCAGATCTTGGACTCCCGCATGAACACTAAATACGACGAGAACGAT
AAGCTCATCCGGAAGTGAAGGTGATTACCCTGAAAAGCAAACCTGTGTGCGACTTTCGGAAGGACTTTCAGTTTAA
CAAAGTGAGAGAAATCAACAACCTACCATCACGCGCATGACGCATACCTCAACGCTGTGGTCGGTACCGCCCTGATCA
AAAAGTACCCTAAACTTGAATCGGAGTTTGTGTACGGAGACTACAAGGTCTACGACGTGAGGAAGATGATAGCCAAG
TCCGAACAGGAAATCGGGAAAGCAACTGCGAAATACTTCTTTTACTCAAACATCATGAACTTTTTCAAGACTGAAAT
TACGCTGGCCAATGGAGAAATCAGGAAGAGGCCACTGATCGAAACTAACGGAGAAACGGGCGAAATCGTGTGGGACA
AGGGCAGGGACTTCGCAACTGTTGCAAAAGTGCTCTCTATGCCGCAAGTCAATATTGTGAAGAAAACCGAAGTGCAA
ACCGGCGGATTTTCAAAGGAATCGATCCTCCCAAAGAGAAATAGCGACAAGCTCATTGCACGCAAGAAAGACTGGGA
CCCGAAGAAGTACGGAGGATTCGATTGCGCGACTGTCGCATACTCCGTCCTCGTGGTGGCCAAGGTGGAGAAGGGAA
AGAGCAAAAAGCTCAAATCCGTCAAAGAGCTGCTGGGGATTACCATCATGGAACGATCCTCGTTCGAGAAGAACCCG
ATTGATTTCTCGAGGCGAAGGGTTACAAGGAGGTGAAGAAGGATCTGATCATCAAACCTCCCAAGTACTCACTGTT
CGAACTGGAAAATGGTCGGAAGCGCATGCTGGCTTCGGCCGGAGAACTCCAAAAGGAAATGAGCTGGCCTTGCCTA
GCAAGTACGTCAACTTCTCTATCTTGCTTCGCACTACGAAAACTCAAAGGGTCACCGGAAGATAACGAACAGAAG
CAGCTTTTCTGTGGAGCAGCACAAGCATTATCTGGATGAAATCATCGAACAATCTCCGAGTTTTCAAAGCGCGTGAT
CCTCGCCGACGCCAACCTCGACAAAGTCTGTGCGCTACAATAAGCATAGAGATAAGCCGATCAGAGAACAGGCCG
AGAACATTATCCACTTGTTACCCCTGACTAACCTGGGAGCCCCAGCCGCTTCAAGTACTTCGATACTACTATCGAT
CGCAAAAGATACACGTCCACCAAGGAAGTTCTGGACGCGACCCTGATCCACCAAAGCATCACTGGACTCTACGAAAC
TAGGATCGATCTGTGCGAGCTGGGTGGCGAT

[0385] U-dep Cas9 ORF (SEQ ID NO:704)

[0386] ATGGACAAGAAGTACAGCATCGGACTGGACATCGGAACAAACAGCGTCGGATGGGCAGTCATCACAGAC
GAATACAAGGTCCCGAGCAAGAAGTTCAAGGTCCTGGGAAACACAGACAGACACAGCATCAAGAAGAACCTGATCGG
AGCACTGCTGTTTCGACAGCGGAGAAAACAGCAGAAGCAACAAGACTGAAGAGAACAGCAAGAAGAAGATACACAAGAA
GAAAGAACAGAATCTGCTACCTGCAGGAAATCTTCAGCAACGAAATGGCAAAGGTCGACGACAGCTTCTTCCACAGA
CTGGAAGAAAGCTTCTGGTTCGAAGAAGACAAGAAGCACGAAAAGACACCCGATCTTCGGAACATCGTCGACGAAGT
CGCATACACGAAAAGTACCCGACAATCTACCACCTGAGAAAGAAGCTGGTCGACAGCACAGACAAGGCAGACCTGA
GACTGATCTACCTGGCACTGGCACACATGATCAAGTTCAGAGGACACTTCCTGATCGAAGGAGACCTGAACCCGGAC
AACAGCGACGTCGACAAGCTGTTTCATCCAGCTGGTCCAGACATACAACCAGCTGTTTGAAGAAAACCCGATCAACGC
AAGCGGAGTCGACGCAAAGGCAATCCTGAGCGCAAGACTGAGCAAGAGCAGAAGACTGGAAAACCTGATCGCACAGC
TGCCGGGAGAAAAGAAGAACGGACTGTTTCGGAACCTGATCGCACTGAGCCTGGGACTGACACCGAACTTCAAGAGC
AACTTCGACCTGGCAGAAGACGCAAAGCTGCAGCTGAGCAAGGACACATACGACGACGACCTGGACAACCTGCTGGC
ACAGATCGGAGACCAGTACGCAGACCTGTTCTGGCAGCAAAGAACCTGAGCGACGCAATCCTGCTGAGCGACATCC
TGAGAGTCAACACAGAAATCACAAGGCACCGCTGAGCGCAAGCATGATCAAGAGATACGACGAACACCACCAGGAC
CTGACACTGCTGAAGGCACTGGTCAGACAGCAGCTGCCGAAAAGTACAAGGAAATCTTCTTCGACCAGAGCAAGAA
CGGATACGCAGGATACATCGACGGAGGAGCAAGCCAGGAAGAATTCTACAAGTTCATCAAGCCGATCCTGGAAAAGA
TGGACGGAACAGAAGAAGTCTGCTGCTCAAGCTGAACAGAGAAGACCTGCTGAGAAAGCAGAGAACATTCGACAACGGA
AGCATCCCGCACCAGATCCACCTGGGAGAACTGCACGCAATCCTGAGAAGACAGGAAGACTTCTACCCGTTCTGAA
GGACAACAGAGAAAAGATCGAAAAGATCCTGACATTCAGAATCCCGTACTACGTCGGACCGCTGGCAAGAGGAAACA
GCAGATTCGCATGGATGACAAGAAAGAGCGAAGAAACAATCACACCGTGGAACCTCGAAGAAGTCGTCGACAAGGGA
GCAAGCGCACAGAGCTTCATCGAAAAGATGACAAACTTCGACAAGAACCTGCCGAACGAAAAGGTCCTGCCGAAGCA
CAGCCTGCTGTACGAATACTTCACAGTCTACAACGAACTGACAAAGGTCAAGTACGTCACAGAAGGAATGAGAAAGC
CGGCATTCTGAGCGGAGAACAGAAGAAGGCAATCGTCGACCTGCTGTTCAAGACAAACAGAAAGGTCACAGTCAAG
CAGCTGAAGGAAGACTACTTCAAGAAGATCGAATGCTTCGACAGCGTCGAAATCAGCGGAGTCGAAGACAGATTCAA
CGCAAGCCTGGGAACATACCACGACCTGCTGAAGATCATCAAGGACAAGGACTTCCTGGACAACGAAGAAAACGAAG
ACATCCTGGAAGACATCGTCTGACACTGACACTGTTTCGAAGACAGAGAAATGATCGAAGAAAGACTGAAGACATAC
GCACACCTGTTTCGACGACAAGGTCATGAAGCAGCTGAAGAGAAGAAGATACACAGGATGGGGAAGACTGAGCAGAAA
GCTGATCAACGGAATCAGAGACAAGCAGAGCGGAAAGACAATCCTGGACTTCCTGAAGAGCGACGGATTTCGAAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACATTCAAGGAAGACATCCAGAAGGCACAGGTCAGCGGACAG
GGAGACAGCCTGCACGAACACATCGCAAACCTGGCAGGAAGCCCGCAATCAAGAAGGGAATCCTGCAGACAGTCAA
GGTCGTCGACGAACTGGTCAAGGTCATGGGAAGACACAAGCCGGAACATCGTCATCGAAATGGCAAGAGAAAACC
AGACAACACAGAAGGGACAGAAGAACAGCAGAGAAAAGATGAAGAGAATCGAAGAAGGAATCAAGGAAGTGGGAAGC
CAGATCCTGAAGGAACACCCGGTCGAAAACACACAGCTGCAGAACGAAAAGCTGTACCTGTACTACCTGCAGAACGG
AAGAGACATGTACGTCGACCAGGAAGTGGACATCAACAGACTGAGCGACTACGACGTCGACCACATCGTCCCGCAGA
GCTTCCTGAAGGACGACAGCATCGACAACAAGGTCCTGACAAGAAGCGACAAGAACAGAGGAAAGAGCGACAACGTC
CCGAGCGAAGAAGTCGTCAGAAGATGAAGAATACTGGAGACAGCTGCTGAACGCAAAGCTGATCACACAGAGAAA
GTTTCGACAACCTGACAAAGGCAGAGAGAGGAGGACTGAGCGAACTGGACAAGGCAGGATTCATCAAGAGACAGCTGG
TCGAAACAAGACAGATCACAAGCACGTCGCACAGATCCTGGACAGCAGAATGAACACAAAGTACGACGAAAACGAC
AAGCTGATCAGAGAAGTCAAGGTCATCACACTGAAGAGCAAGCTGGTCAGCGACTTCAGAAAGGACTTCCAGTTCTA

CAAGGTCAGAGAAATCAACAACCTACCACCACGCACACGACGCATACCTGAACGCAGTCGTCGGAACAGCACTGATCA
AGAAGTACCCGAAGCTGGAAAGCGAATTCGTCTACGGAGACTACAAGGTCTACGACGTCAGAAAGATGATCGCAAAG
AGCGAACAGGAAATCGGAAAGGCAACAGCAAAGTACTTCTTCTACAGCAACATCATGAACTTCTTCAAGACAGAAAT
CACACTGGCAAACGGAGAAATCAGAAAGAGACCGCTGATCGAAACAAACGGAGAAACAGGAGAAATCGTCTGGGACA
AGGGAAGAGACTTCGCAACAGTCAGAAAGGTCTGAGCATGCCGCAGGTCAACATCGTCAAGAAGACAGAAGTCCAG
ACAGGAGGATTTCAGCAAGGAAAGCATCCTGCCGAAGAGAAAACAGCGACAAGCTGATCGCAAGAAAGAAGGACTGGGA
CCCGAAGAAGTACGGAGGATTTCGACAGCCCGACAGTCGCATACAGCGTCTTGGTCGTCGCAAAGGTGCGAAAAGGGAA
AGAGCAAGAAGCTGAAGAGCGTCAAGGAACTGCTGGGAATCACAATCATGGAAAGAAGCAGCTTCGAAAAGAACCCG
ATCGACTTCTGGAAGCAAAGGGATACAAGGAAGTCAAGAAGGACCTGATCATCAAGCTGCCGAAGTACAGCCTGTT
CGAACTGGAAAACGGAAGAAAGAGAATGCTGGCAAGCGCAGGAGAACTGCAGAAGGGAAACGAACTGGCACTGCCGA
GCAAGTACGTCAACTTCTGTACCTGGCAAGCCACTACGAAAAGCTGAAGGGAAGCCCGGAAGACAACGAACAGAAG
CAGCTGTTTCGTGCAACAGCACAAGCACTACCTGGACGAAATCATCGAACAGATCAGCGAATTCAGCAAGAGAGTCAT
CCTGGCAGACGCAAACCTGGACAAGGTCTGAGCGCATACAACAAGCACAGAGACAAGCCGATCAGAGAACAGGCAG
AAAACATCATCCACCTGTTTCACTGACAAAACCTGGGAGCACCGGCAGCATTCAAGTACTTCGACACAACAATCGAC
AGAAAGAGATACACAAGCACAAAGGAAGTCTTGGACGCAACACTGATCCACCAGAGCATCACAGGACTGTACGAAAC
AAGAATCGACCTGAGCCAGCTGGGAGGAGACGGAGGAGGAAGCCCGAAGAAGAAGAGAAAGGTCTAG

[0387] 包含U dep Cas9的mRNA (SEQ ID NO:705)

[0388] GGGUCCCGCAGUCGGCGUCCAGCGGCUCUGCUUGUUCGUGUGUGUCGUUGCAGGCCUUAUUCGGAUC
CGCCACCAUGGACAAGAAGUACAGCAUCGGACUGGACAUCGGAACAAACAGCGUCGGAUGGGCAGUCAUCACAGACG
AAUACAAGGUCCCGAGCAAGAAGUUAAGGUCCUGGGAAACACAGACAGACACAGCAUCAAGAAGAACCUGAUCGGA
GCACUGCUGUUCGACAGCGGAGAAACAGCAGAAGCAACAAGACUGAAGAGAACAGCAAGAAGAAGAUACACAAGAAG
AAAGAACAGAAUCUGCUACCUGCAGGAAAUCUUCAGCAACGAAAUGGCAAAGGUCGACGACAGCUUCUCCACAGAC
UGGAAGAAAGCUUCCUGGUCGAAGAAGACAAGAAGCACGAAAGACACCCGAUCUUCGGAACAUUCGUCGACGAAGUC
GCAUACCACGAAAAGUACCCGACAAUCUACCACCUGAGAAAAGAAGCUGGUCGACAGCACAGACAAGGCAGACCUGAG
ACUGAUCUACCUGGCACUGGCACACAUGAUAAGUUCAGAGGACACUUCUGAUCGAAGGAGACCUGAACCCGGACA
ACAGCGACGUCGACAAGCUGUUAUCCAGCUGGUCCAGACAUACAACCAGCUGUUCGAAGAAAACCCGAUCAACGCA
AGCGGAGUCGACGCAAAGGCAAUCCUGAGCGCAAGACUGAGCAAGAGCAGAAGACUGGAAAACCCUGAUCGCACAGCU
GCCGGGAGAAAAGAAGAACGGACUGUUCGGAACCCUGAUCGCACUGAGCCUGGGACUGACACCCGAACUUAAGAGCA
ACUUCGACCUGGCAGAAGACGCAAAGCUGCAGCUGAGCAAGGACACAUACGACGACGACCUGGACAACCUGCUGGCA
CAGAUCCGAGACCAGUACGACGACCUGUUCUGGCAGCAAAGAACCUGAGCGACGCAAUCCUGCUGAGCGACAUCU
GAGAGUCAACACAGAAAUCACAAAGGCACCGCUGAGCGCAAGCAUGAUAAGAGAUACGACGAACACCACCAGGACC
UGACACUGCUGAAGGCACUGGUCAGACAGCAGCUGCCGAAAAGUACAAGGAAAUUCUUCUUCGACCAGAGCAAGAAC
GGAUACGCAGGAUACAUCGACGGAGGAGCAAGCCAGGAAGAAUUCUACAAGUUAUCAAGCCGAUCCUGGAAAAGAU
GGACGGAACAGAAGAACUGCUGGUCAAGCUGAACAGAGAAGACCUGCUGAGAAAGCAGAGAACAUUCGACAACGGAA
GCAUCCCGCACCCAGAUCCACCUGGGAGAACUGCACGCAAUCCUGAGAAGACAGGAAGACUUCUACCCGUUCCUGAAG
GACAACAGAGAAAAGAUCGAAAAGAUCCUGACAUCAGAAUCCCGUACUACGUCGGACCGCUGGCAAGAGGAAACAG
CAGAUUCGAUGGAUGACAAGAAAGAGCGAAGAAACAAUCACACCGUGGAACUUCGAAGAAGUCGUCGACAAGGGAG
CAAGCGCACAGAGCUUAUCGAAAGAAUGACAAACUUCGACAAGAACCUGCCGAACGAAAAGGUCCUGCCGAAGCAC
AGCCUGCUGUACGAAUACUUCACAGUCUACAACGAACUGACAAAGGUCAAGUACGUCACAGAAGGAUAGAGAAAGCC

GGCAUUCUGAGCGGAGAACAGAAGAAGGCAAUCGUCGACCUGCUGUUCAAGACAAACAGAAAGGUCACAGUCAAGC
AGCUGAAGGAAGACUACUUAAGAAGAUUCGAAUGCUUCGACAGCGUCGAAAUCAGCGGAGUCGAAGACAGAUUCAAC
GCAAGCCUGGGAACAUACCACGACCUGCUGAAGAUAUCAAGGACAAGGACUUCUGGACAACGAAGAAAACGAAGA
CAUCCUGGAAGACAUCGUCCUGACACUGACACUGUUCGAAGACAGAGAAAUGAUCGAAGAAAGACUGAAGACAUCG
CACACCUGUUCGACGACAAGGUCAUGAAGCAGCUGAAGAGAAGAAGAUACACAGGAUGGGGAAGACUGAGCAGAAAG
CUGAUAACGGAAUCAGAGACAAGCAGAGCGGAAAAGACAAUCCUGGACUUCUGAAGAGCGACGGAUUCGCAAACAG
AAACUUAUGCAGCUGAUCCACGACGACAGCCUGACAUAUUAAGGAAGACAUCAGGAAGGCACAGGUCAGCGGACAGG
GAGACAGCCUGCAGAACACAUCGCAAACCUGGCAGGAAGCCCGGCAAUCAAGAAGGGAAUCCUGCAGACAGUCAAG
GUCGUCGACGAACUGGUCAAGGUCAUGGGAAGACACAAGCCGGAAAACAUUGUAUCGAAAUGGCAAGAGAAAACCA
GACAACACAGAAGGGACAGAAGAACAGCAGAGAAAAGAAUGAAGAGAAUCGAAGAAGGAUUAAGGAACUGGGAAGCC
AGAUCUGAAGGAACACCCGGUUGAAAACACACAGCUGCAGAACGAAAAGCUGUACCUGUACUACCUGCAGAACGGA
AGAGACAUGUACGUCGACCAGGAACUGGACAUAACAGACUGAGCGACUACGACGUCGACCACAUCGUCCCGCAGAG
CUUCUGAAGGACGACAGCAUCGACAACAAGGUCCUGACAAGAAGCGACAAGAACAGAGGAAAGAGCGACAACGUCC
CGAGCGAAGAAGUCGUAAGAAGAUAGAAGACUACUGGAGACAGCUGCUGAACGCAAAGCUGAUCACACAGAGAAAG
UUCGACAACCUGACAAAGGCAGAGAGAGGAGGACUGAGCGAACUGGACAAGGCAGGAUUAUCAAGAGACAGCUGGU
CGAAACAAGACAGAUACAAAGCACGUCGCACAGAUCCUGGACAGCAGAAUGAACACAAAGUACGACGAAAACGACA
AGCUGAUCAGAGAAGUCAAGGUCAUCACACUGAAGAGCAAGCUGGUCAGCGACUUCAGAAAGGACUCCAGUUCUAC
AAGGUCAGAGAAAUAACAACUACCACCACGCACACGACGCAUACCUGAACGCAGUCGUCGGAACAGCACUGAUCAA
GAAGUACCCGAAGCUGGAAAGCGAAUUCGUCUACGGAGACUACAAGGUCUACGACGUCAGAAAGAUGAUCGCAAAGA
GCGAACAGGAAUUCGAAAGGCAACAGCAAAGUACUUCUUCUACAGCAACAUAUGAACUUCUUAAGACAGAAAUC
ACACUGGCAAACGGAGAAAUCAGAAAGAGACCGCUGAUCGAAACAAACGGAGAAACAGGAGAAAUCGUCUGGGACAA
GGGAAGAGACUUCGCAACAGUCAGAAAGGUCCUGAGCAUGCCGCAGGUCAACAUCGUAAGAAGACAGAAGUCCAGA
CAGGAGGAUUCAGCAAGGAAAGCAUCCUGCCGAAGAGAAACAGCGACAAGCUGAUCGCAAGAAAGAAGGACUGGGAC
CCGAAGAAGUACGGAGGAUUCGACAGCCCGACAGUCGCAUACAGCGUCCUGGUCGUCGCAAAGGUCGAAAAGGGAAA
GAGCAAGAAGCUGAAGAGCGUCAAGGAACUGCUGGGAUUCACAAUCAUGGAAAGAAGCAGCUUCGAAAAGAACCCGA
UCGACUUCUGGAAGCAAAGGGGAUACAAGGAAGUCAAGAAGGACCUGAUAUCAAGCUGCCGAAGUACAGCCUGUUC
GAACUGGAAAACGGAAGAAAGAGAAUUCUGGCAAGCGCAGGAGAACUGCAGAAGGGAAACGAACUGGCACUGCCGAG
CAAGUACGUAACUUCUGUACCUGGCAAGCCACUACGAAAAGCUGAAGGGAAGCCCGGAAGACAACGAACAGAAGC
AGCUGUUCGUCGAACAGCACAAGCACUACCUGGACGAAAUCAUCGAACAGAUACAGCGAAUUCAGCAAGAGAGUCAUC
CUGGCAGACGCAAACCUGGACAAGGUCCUGAGCGCAUACAACAAGCACAGAGACAAGCCGAUCAGAGAACAGGCAGA
AAACAUAUCCACCUGUUCACACUGACAAACCUGGGAGCACCAGCAGCAUUAAGUACUUCGACACAACAAUCGACA
GAAAGAGAUACACAAGCACAAAGGAAGUCCUGGACGCAACACUGAUCCACCAGAGCAUCACAGGACUGUACGAAACA
AGAAUCGACCUGAGCCAGCUGGGAGGAGACGGAGGAGGAAGCCCGAAGAAGAAGAGAAAGGUCUAGCUAGCCAUCAC
AUUUAAAAGCAUCUCAGCCUACCAUGAGAAUUAAGAGAAAGAAAAUGAAGAUCAAUAGCUUAUUAUCUCUUUUUCU
UUUCGUUGGUGUAAAGCCAACACCCUGUCUAAAAACAUAUUUUUUAUUAUUUUGCCUCUUUUUCUCUGUGCUU
CAAUUAAUAAAAAUGGAAAGAACCUCGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA
AAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG

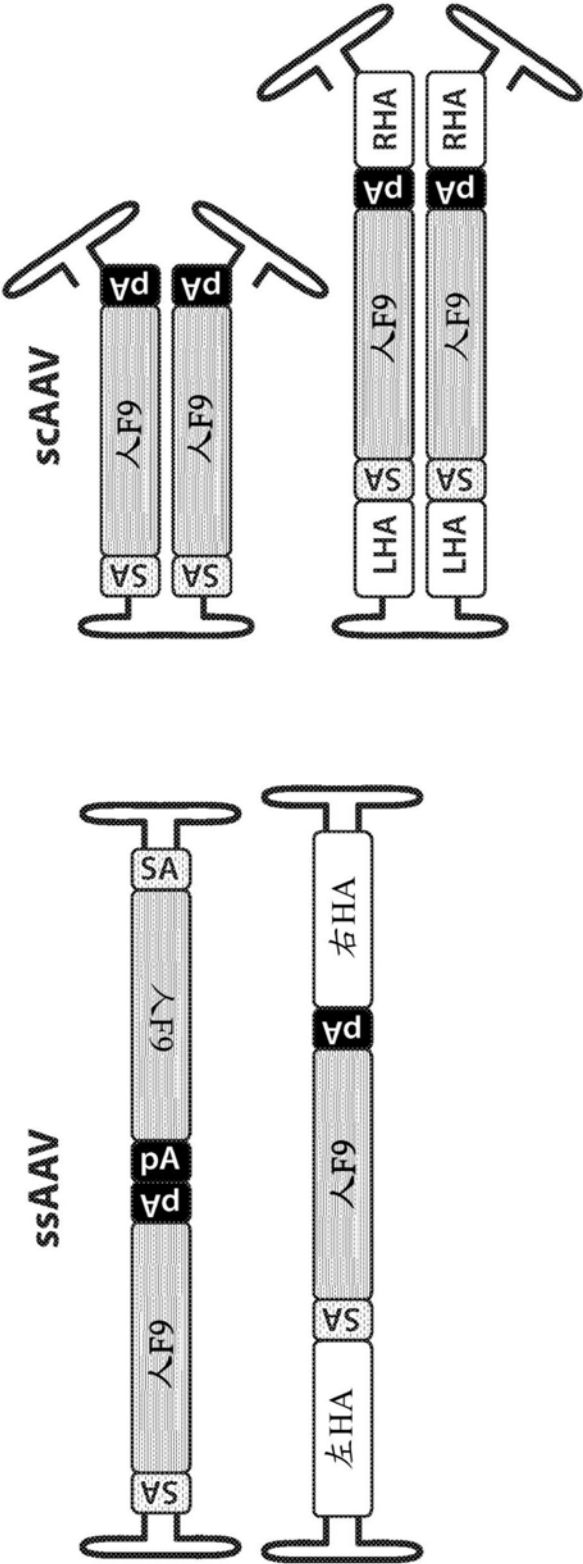


图1

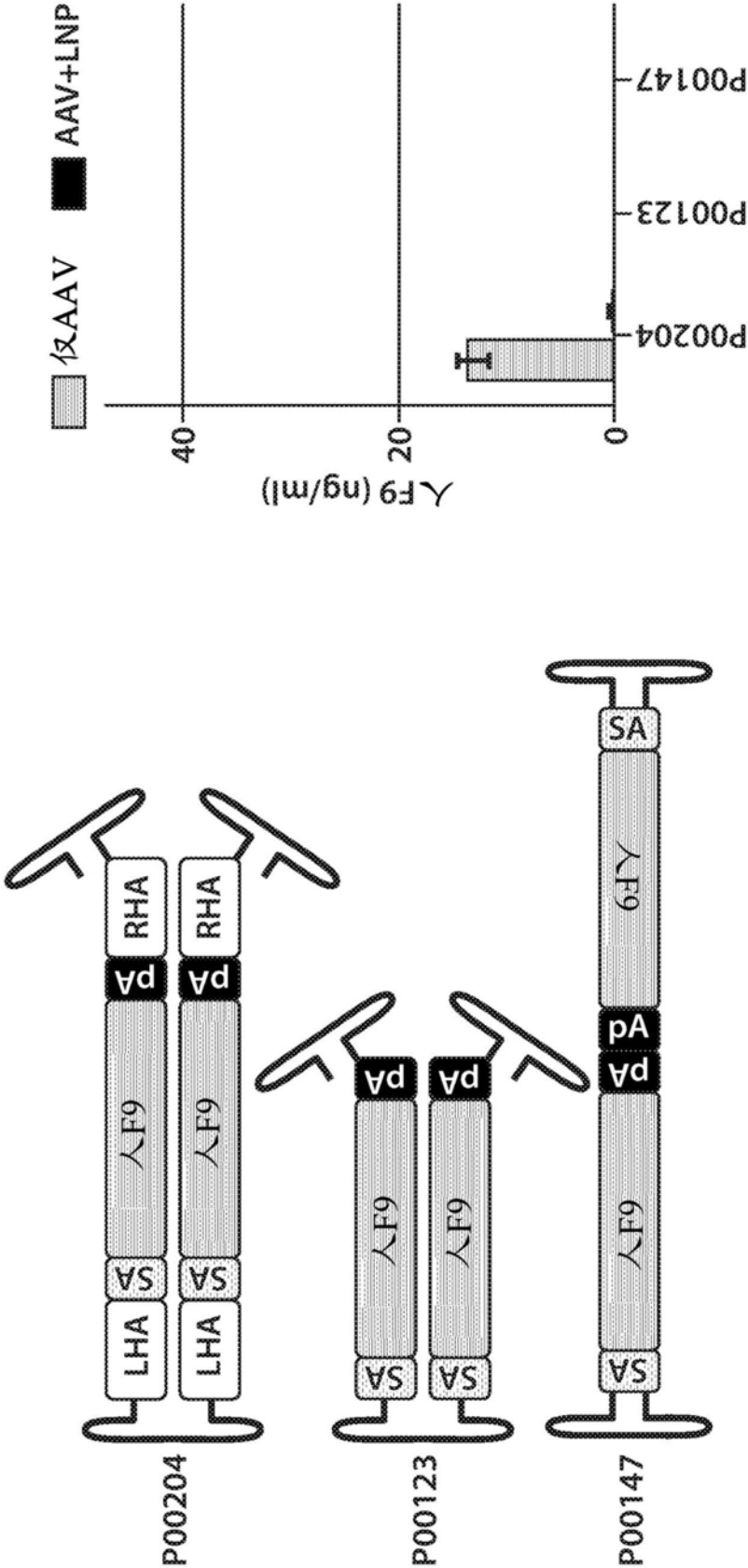


图2

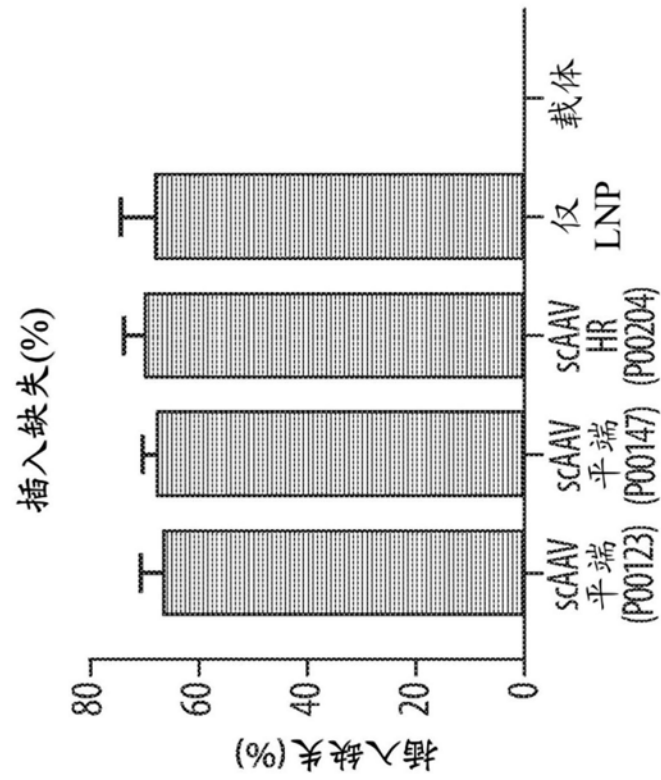


图3A

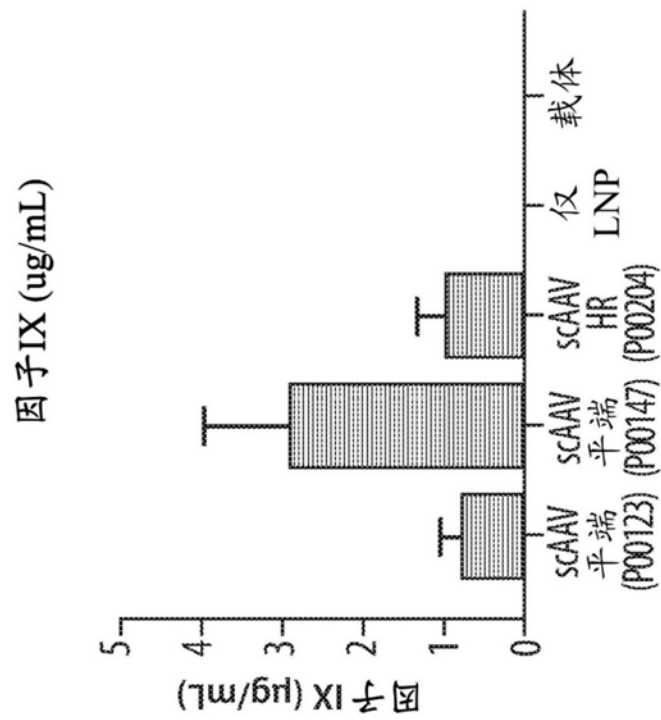


图3B

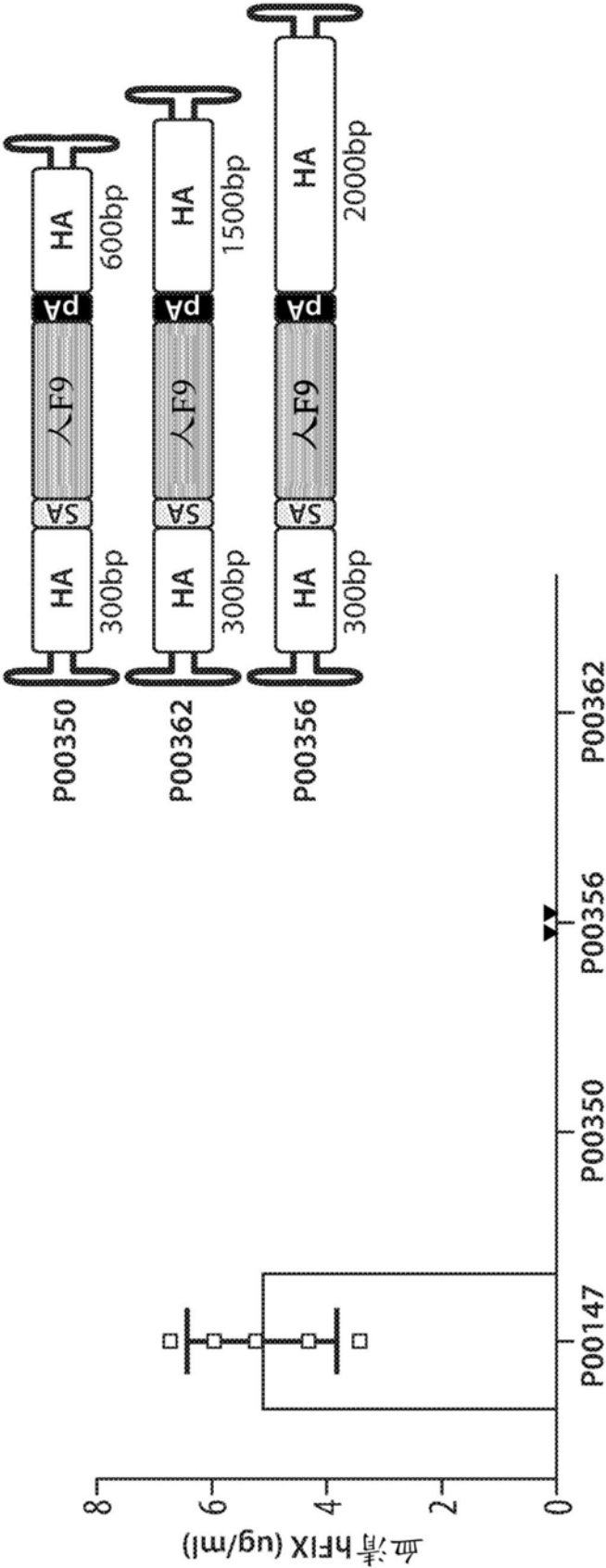


图4A

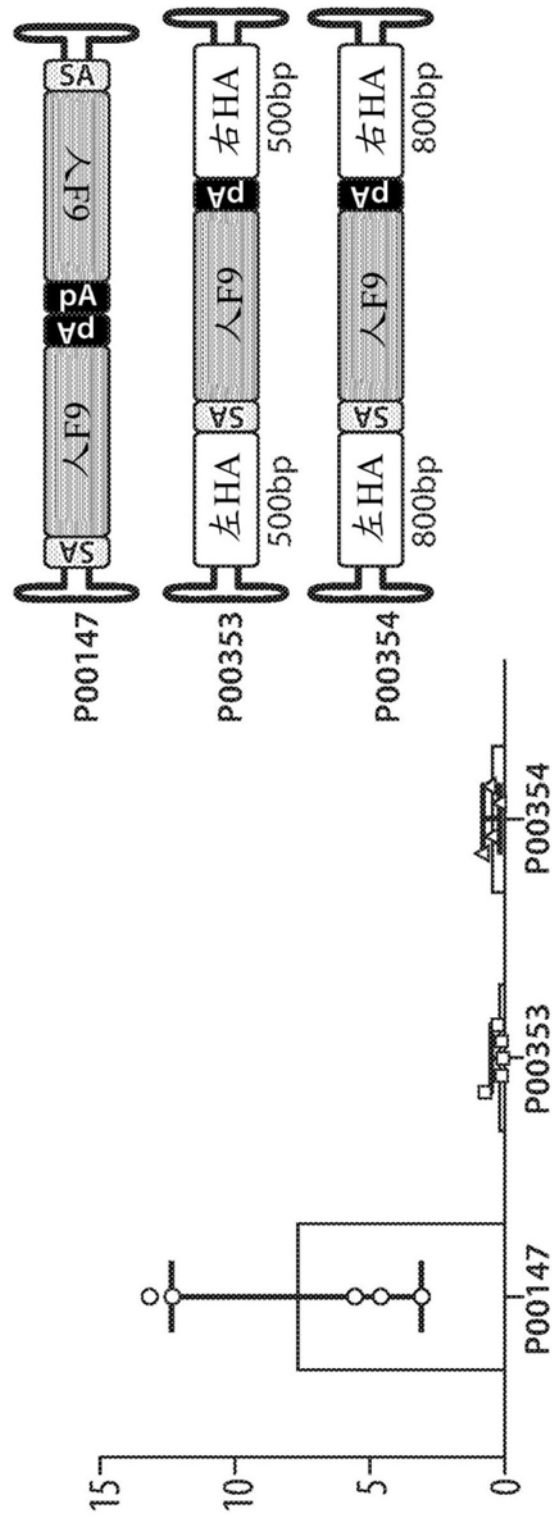


图4B

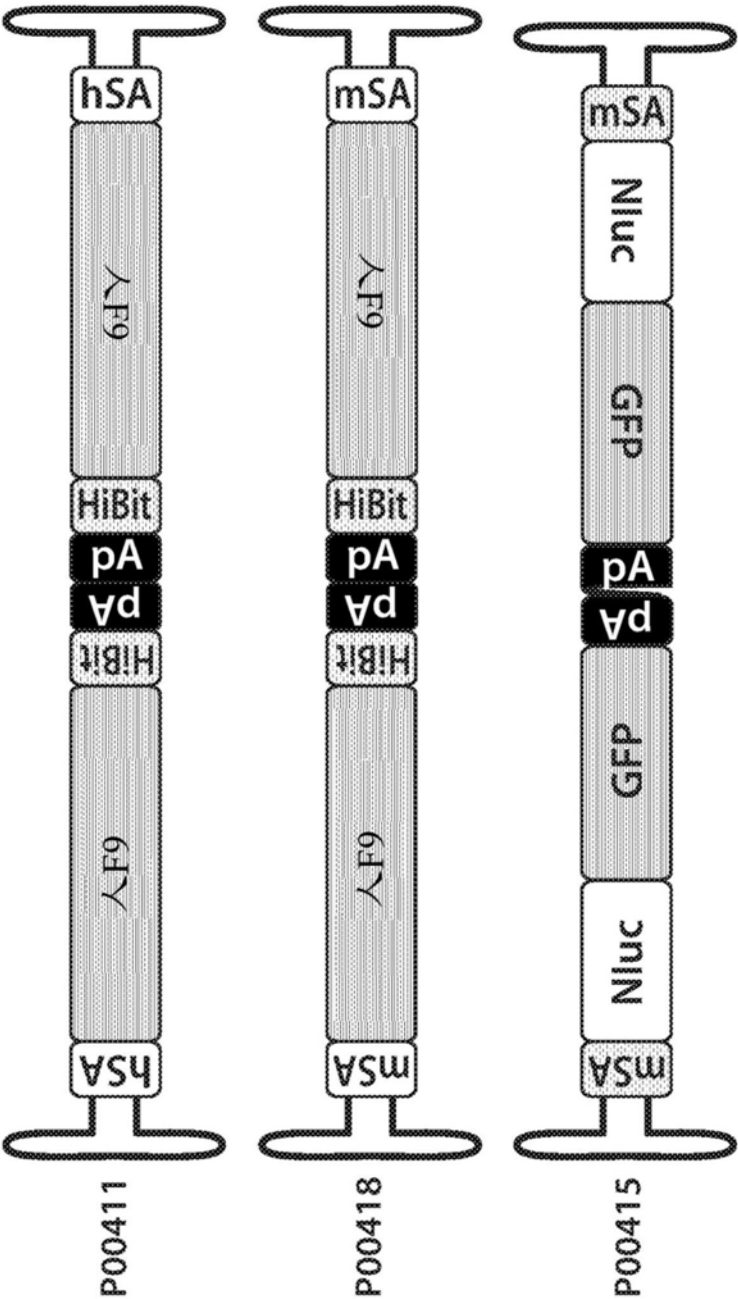


图5A

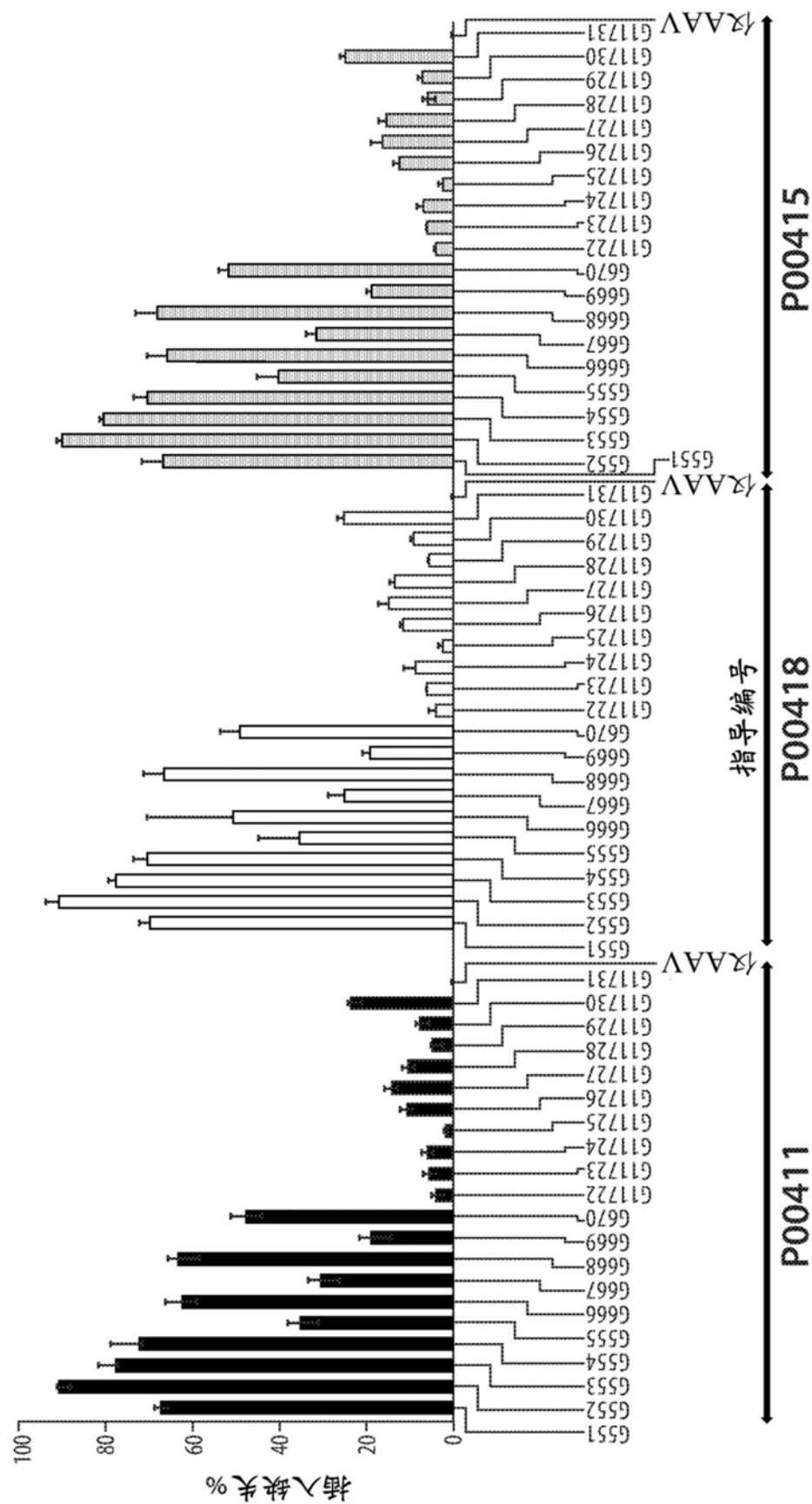


图5B

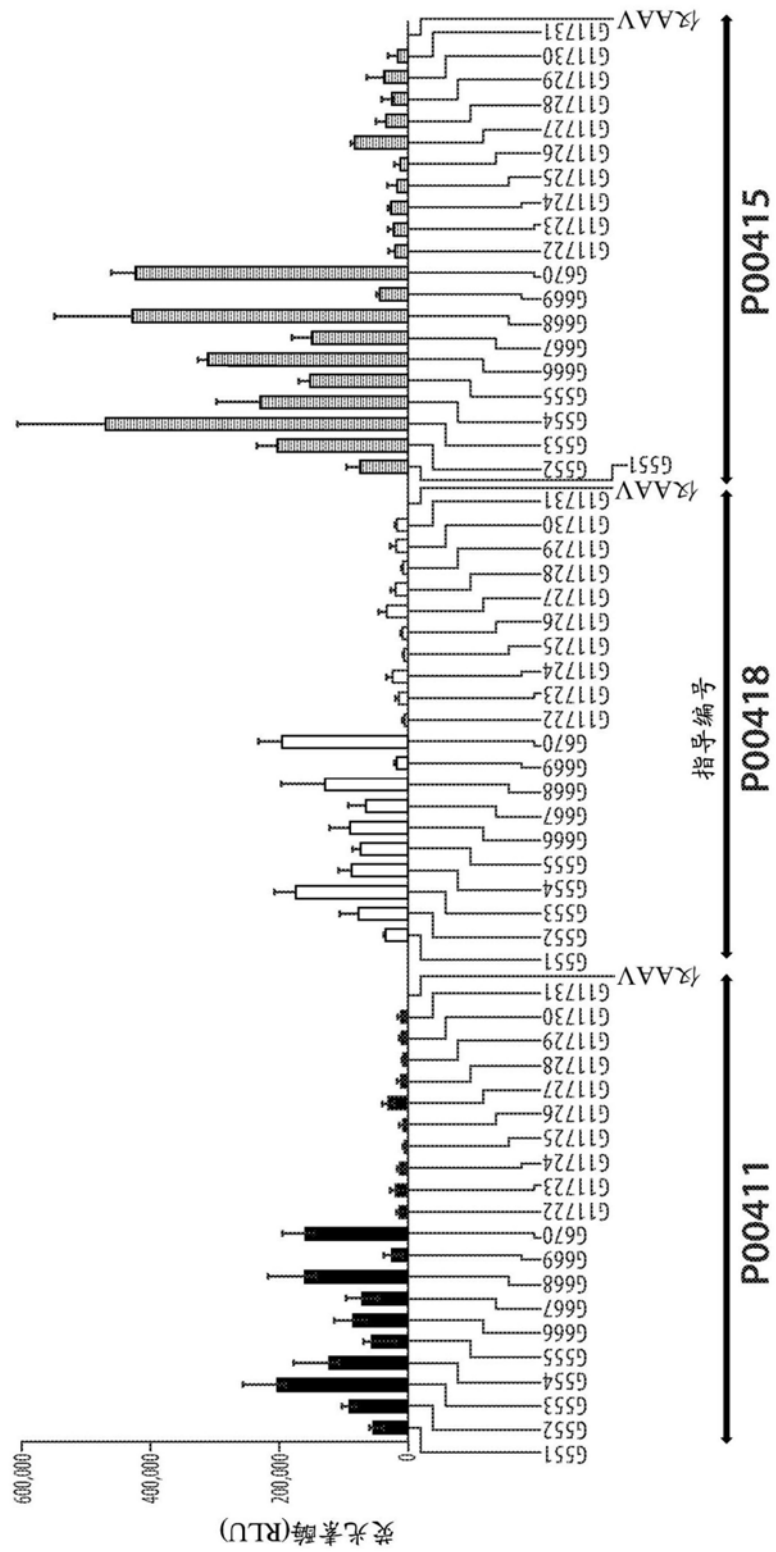


图5C

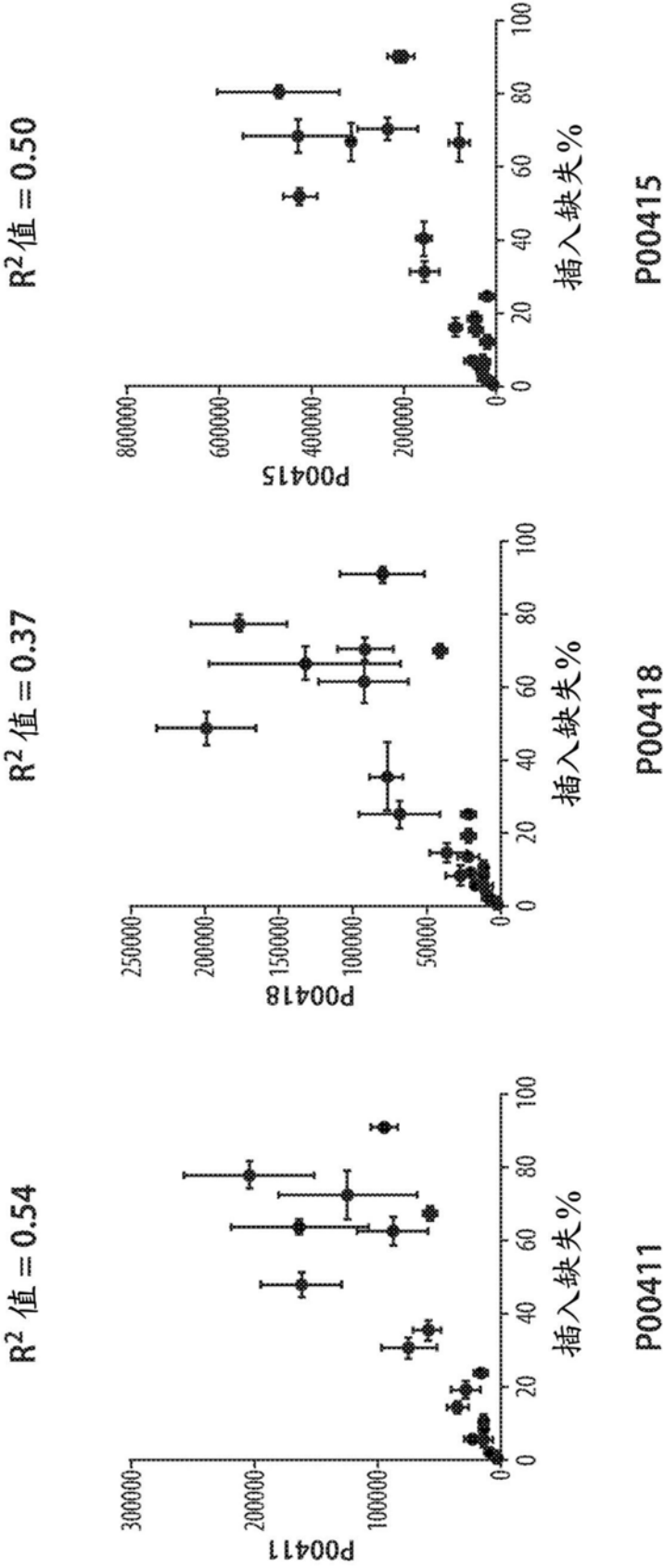


图5D

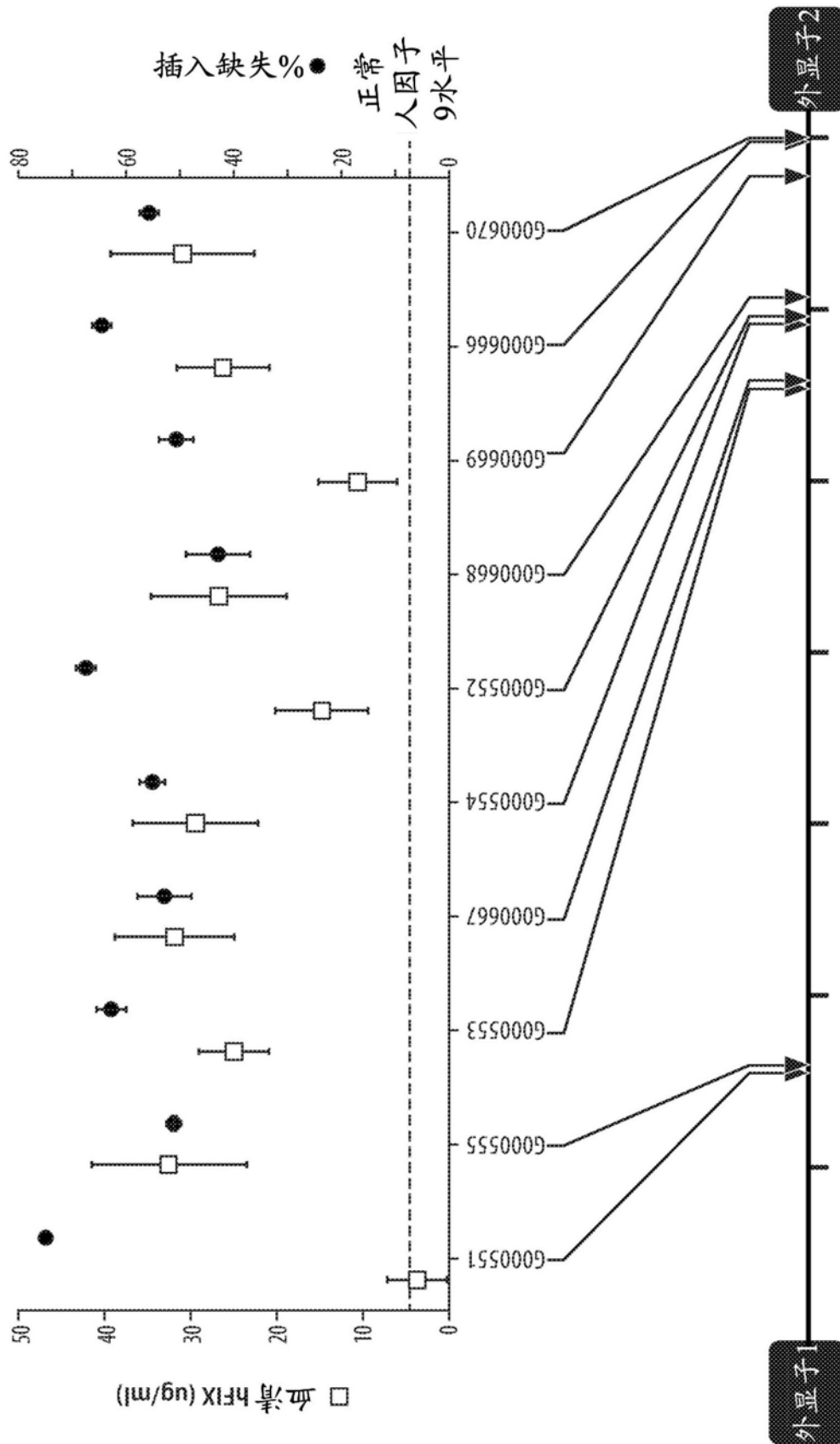


图6

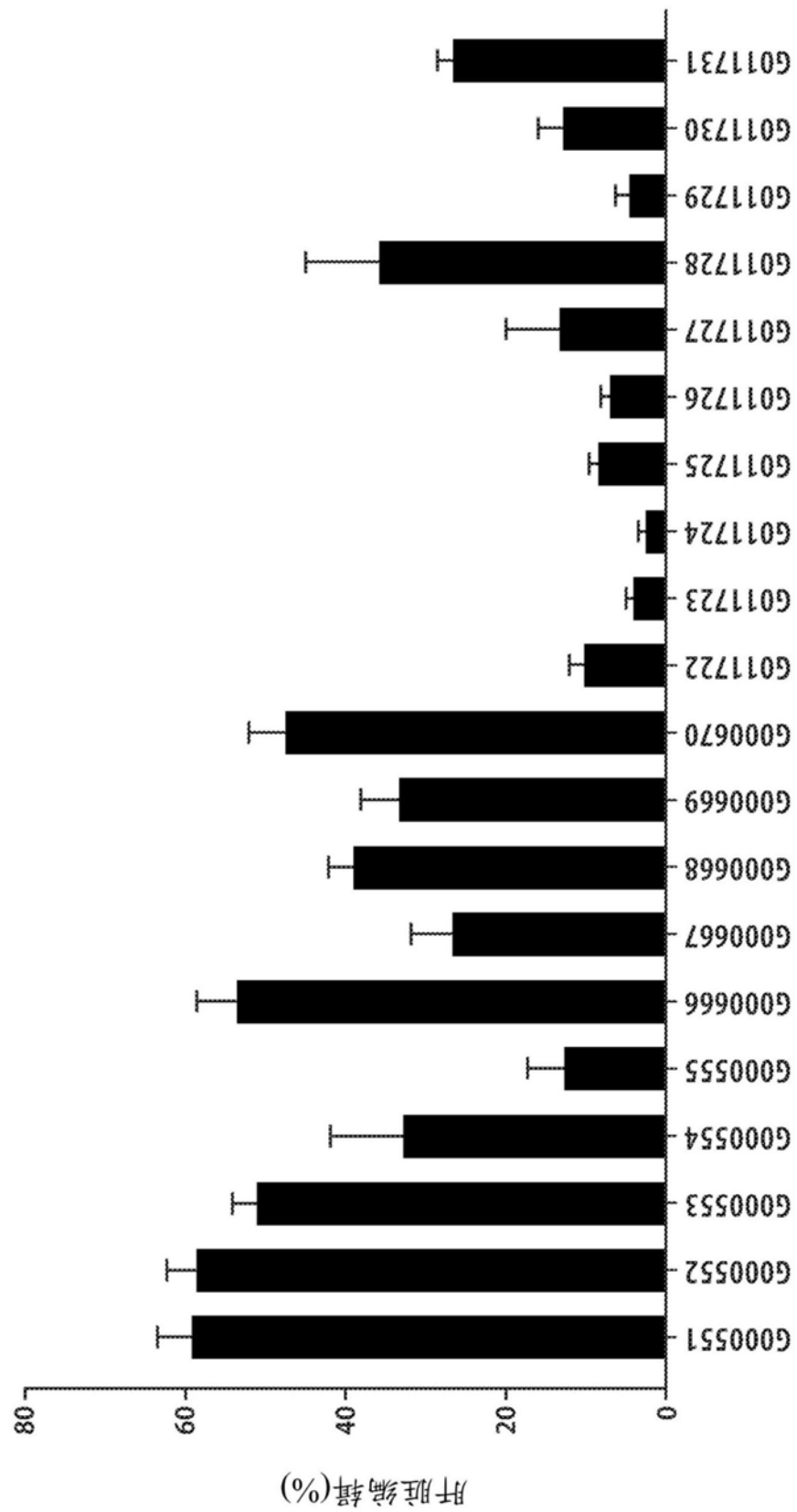


图7A

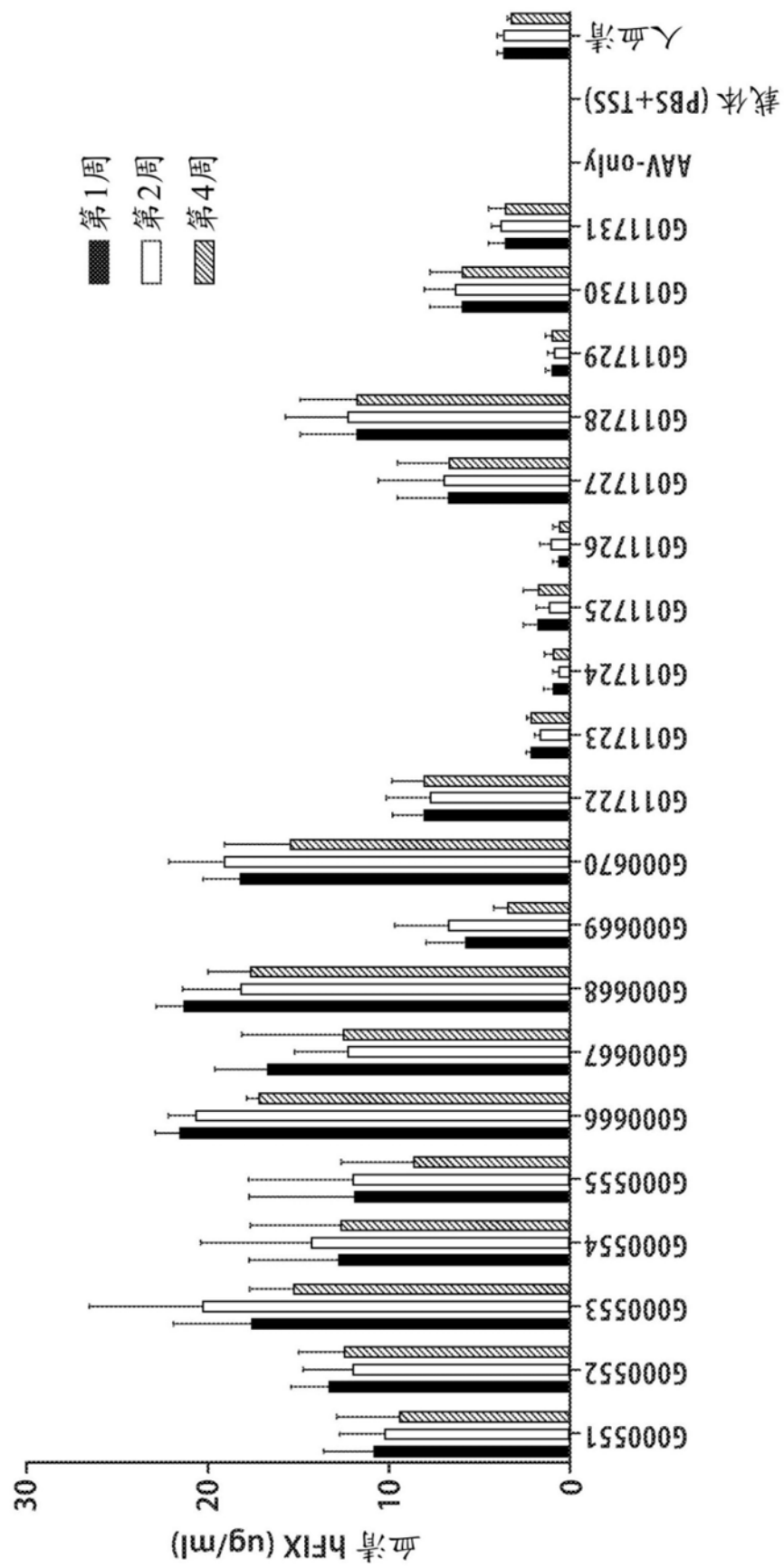


图7B

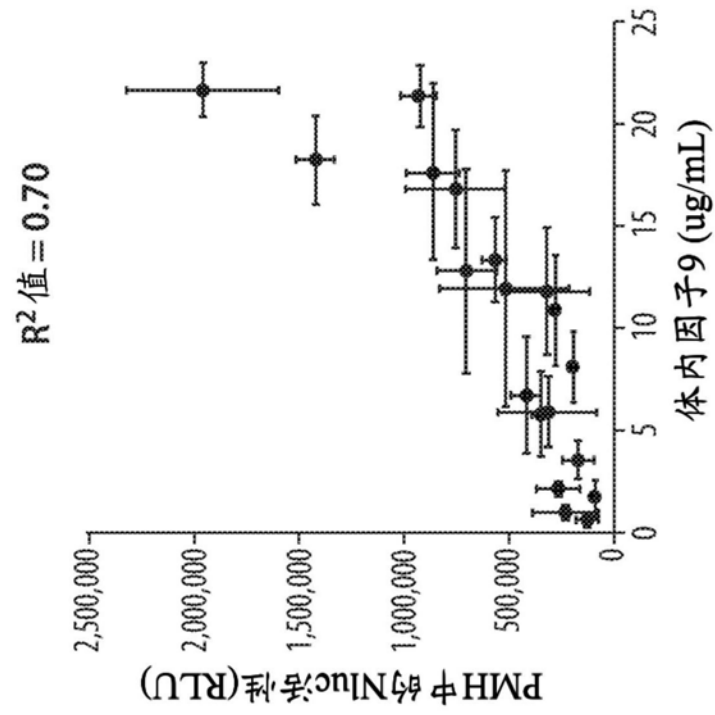


图7C

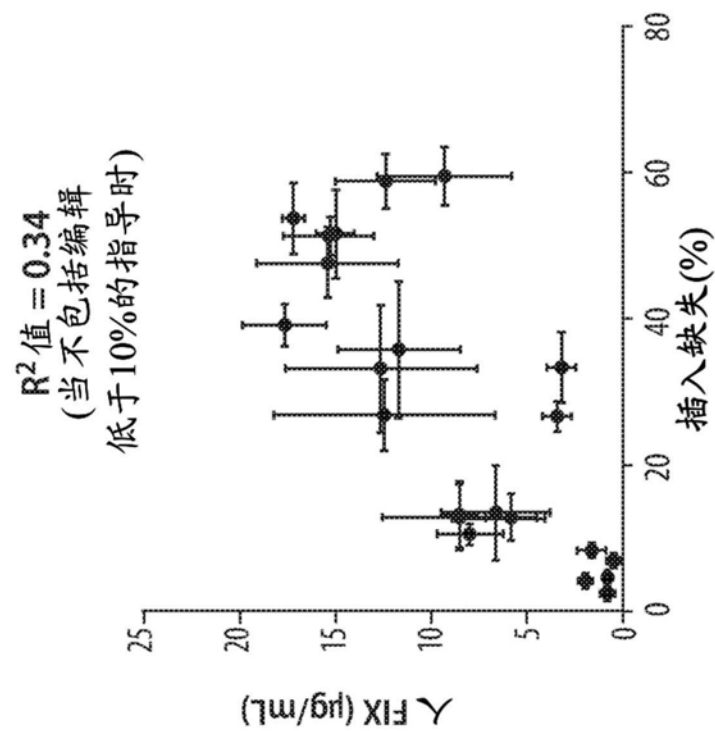


图7D

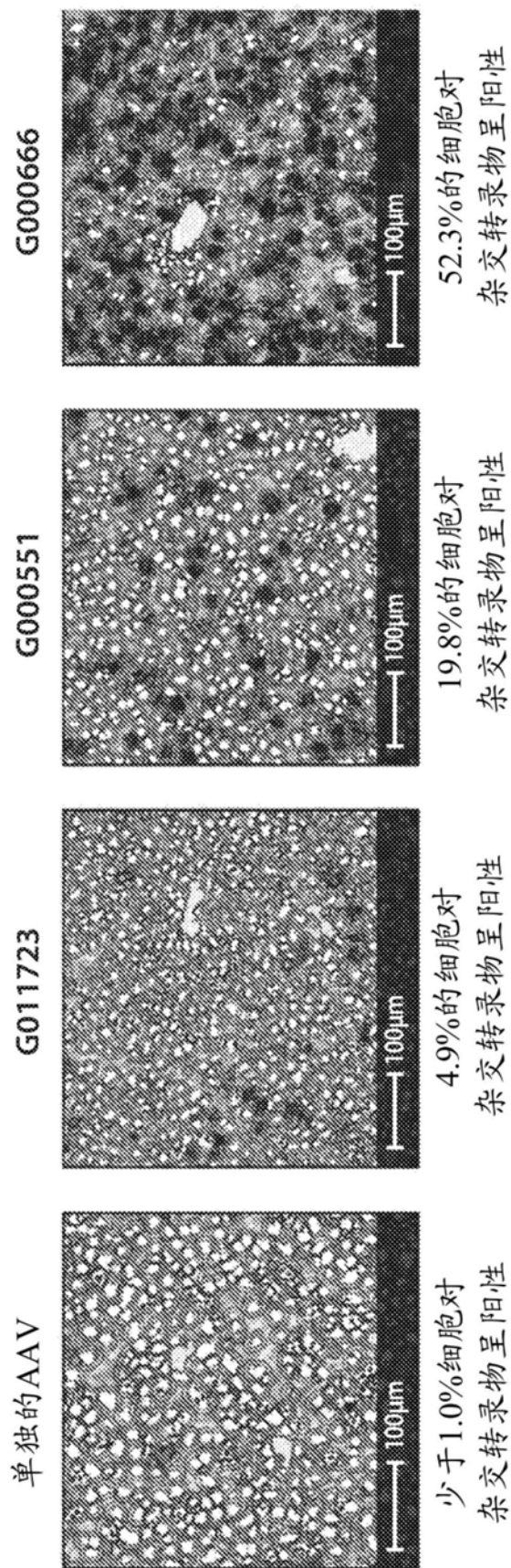


图8A

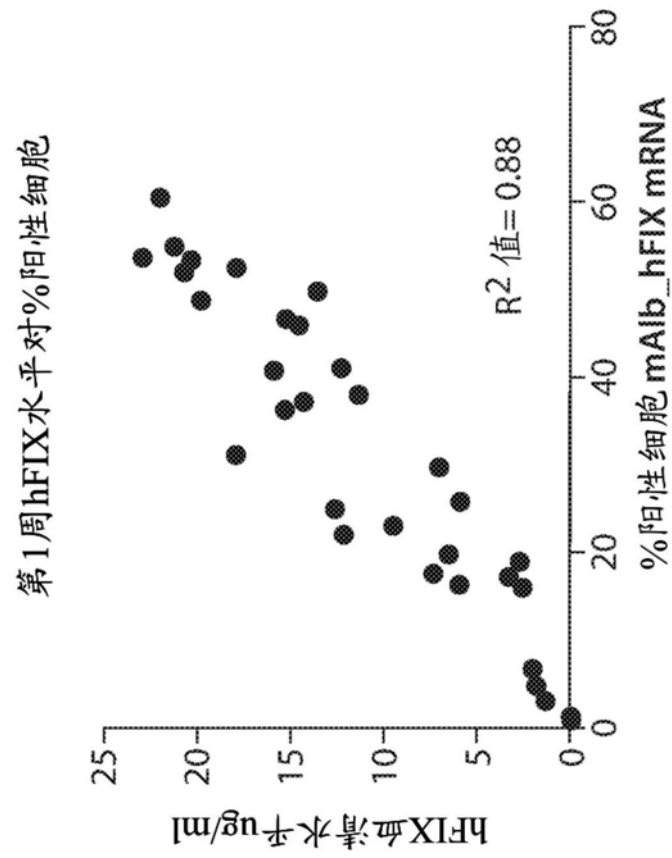


图8B

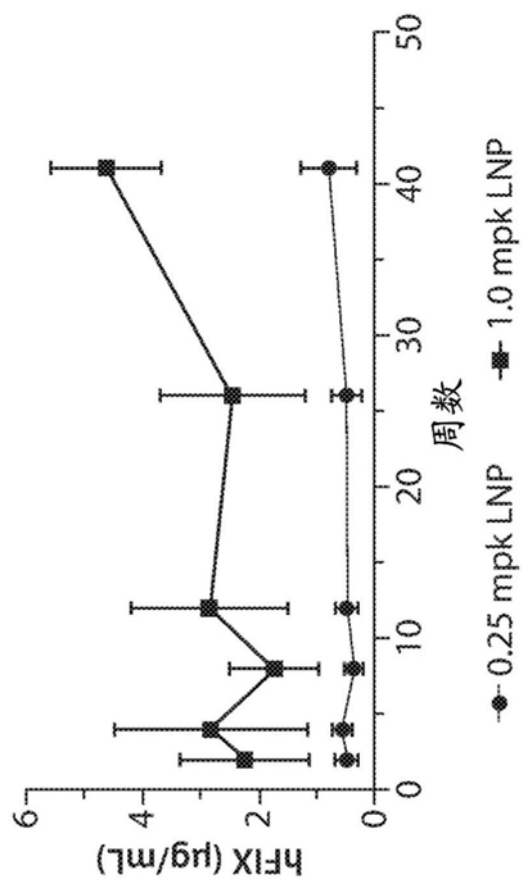


图9A

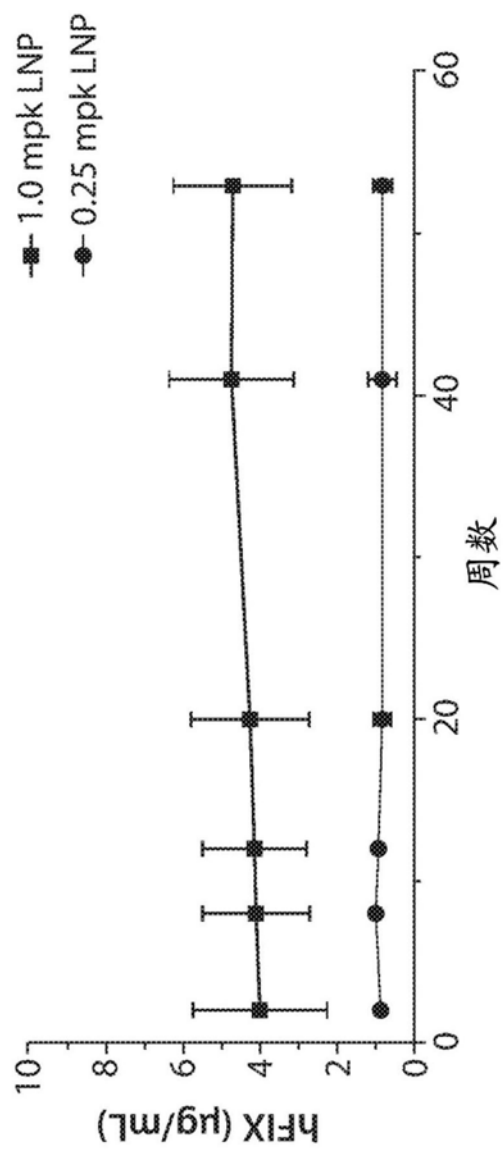


图9B

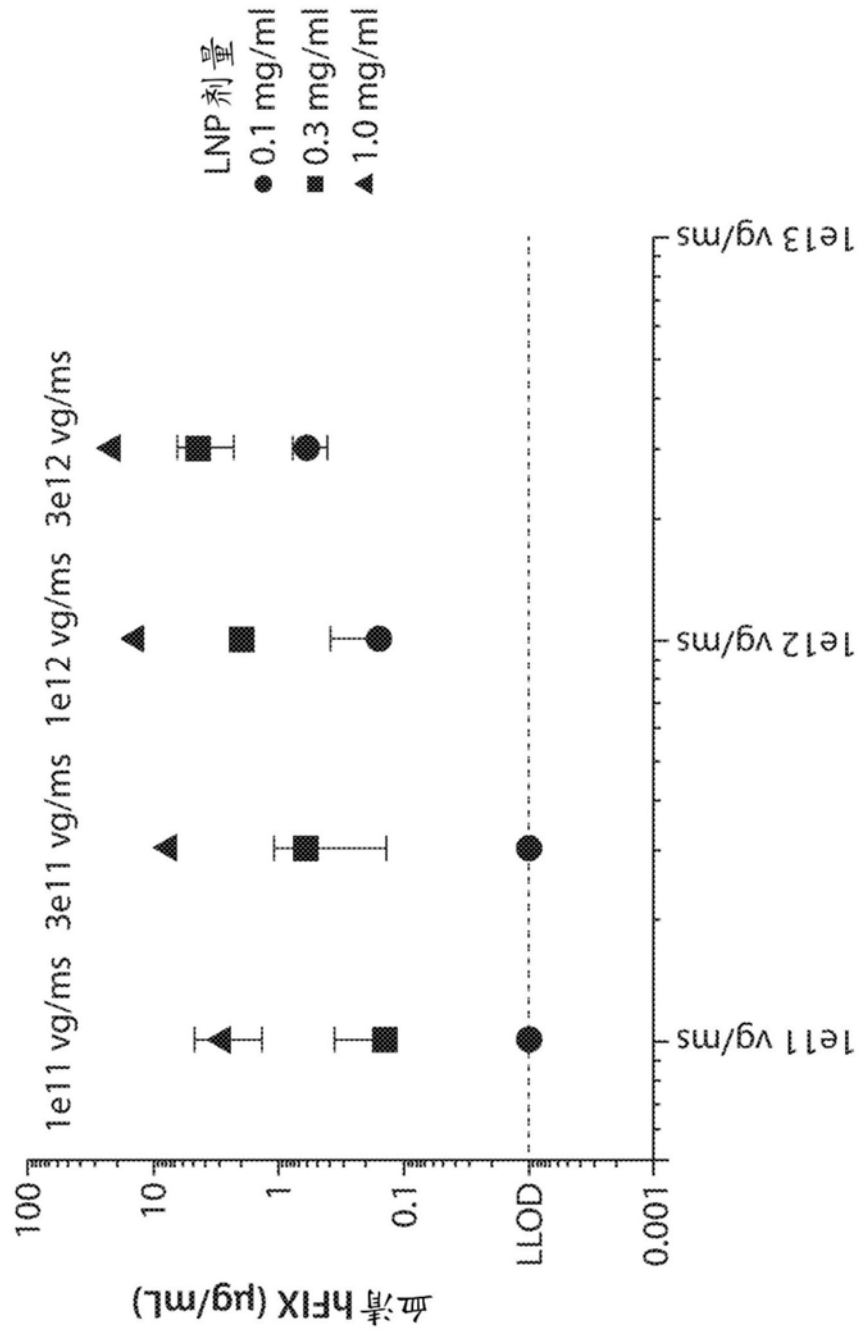


图10A

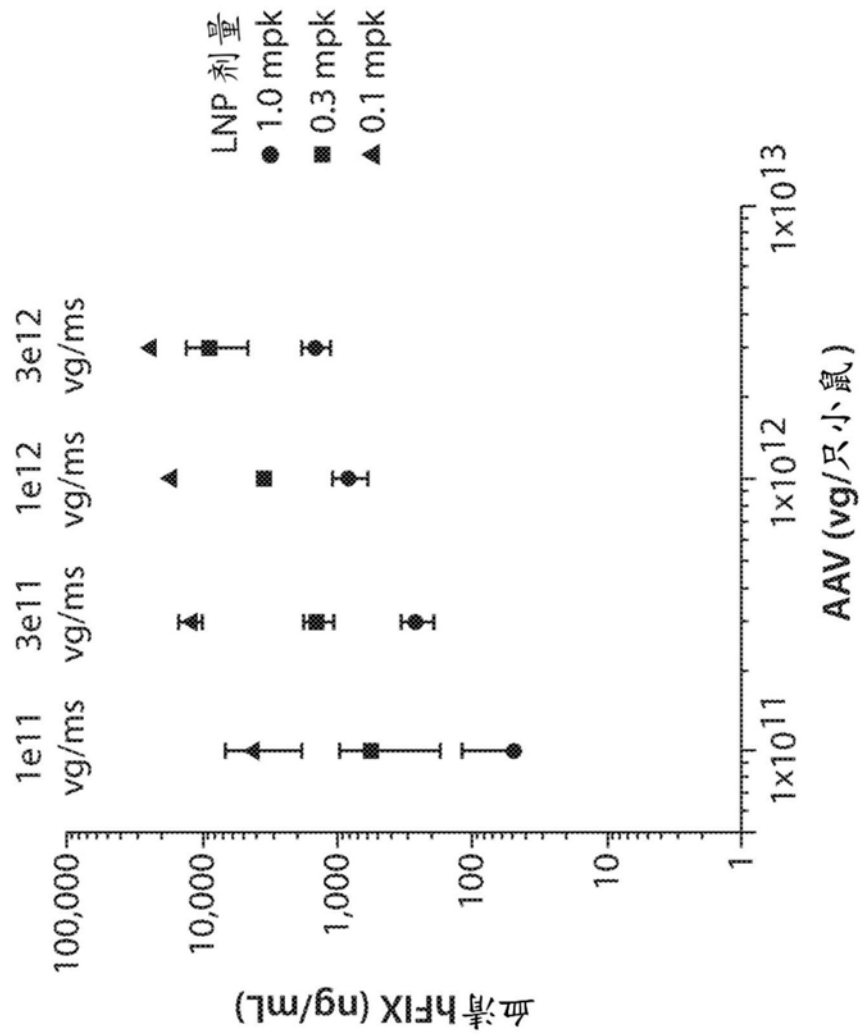


图10B

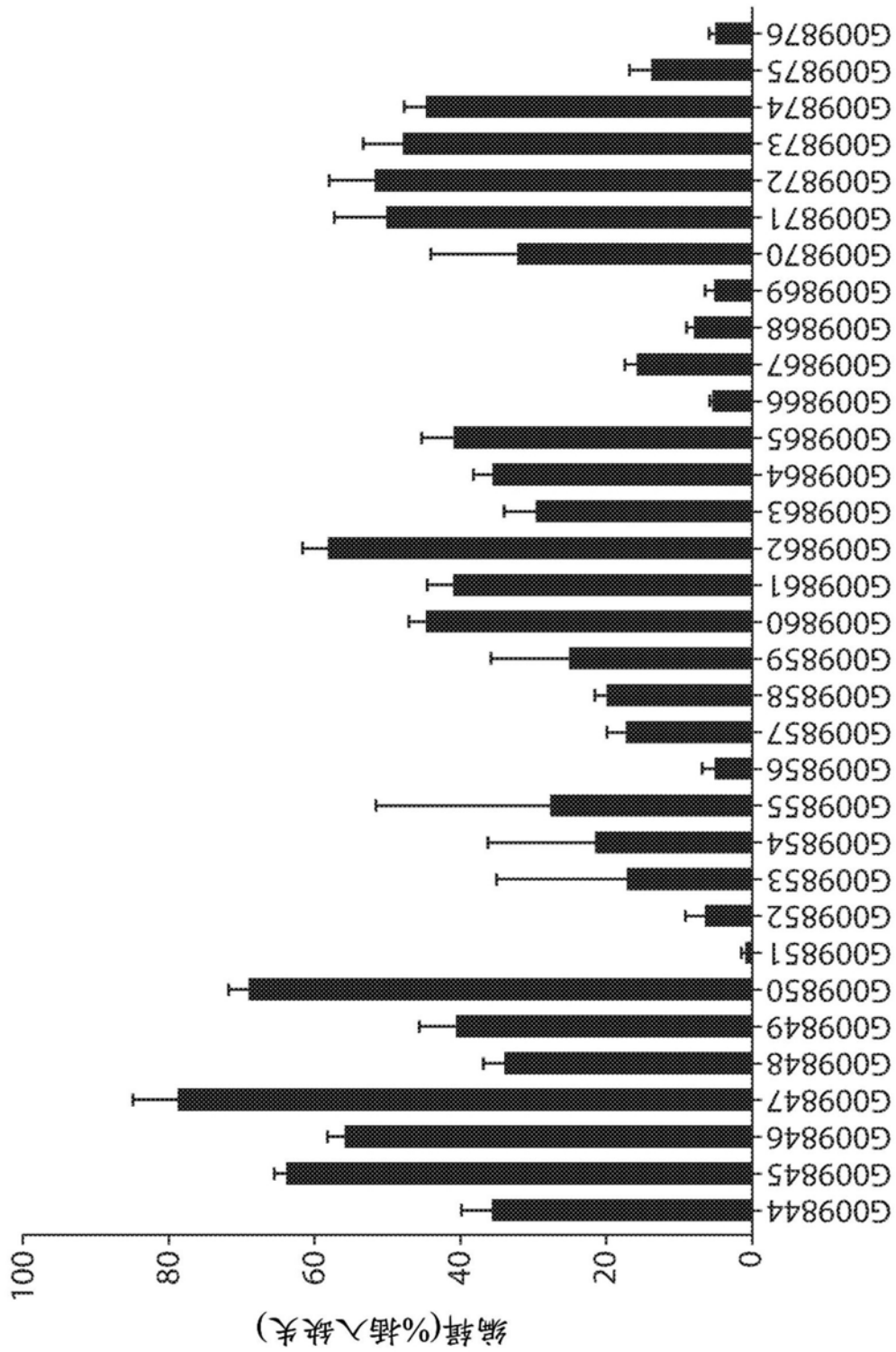


图11A

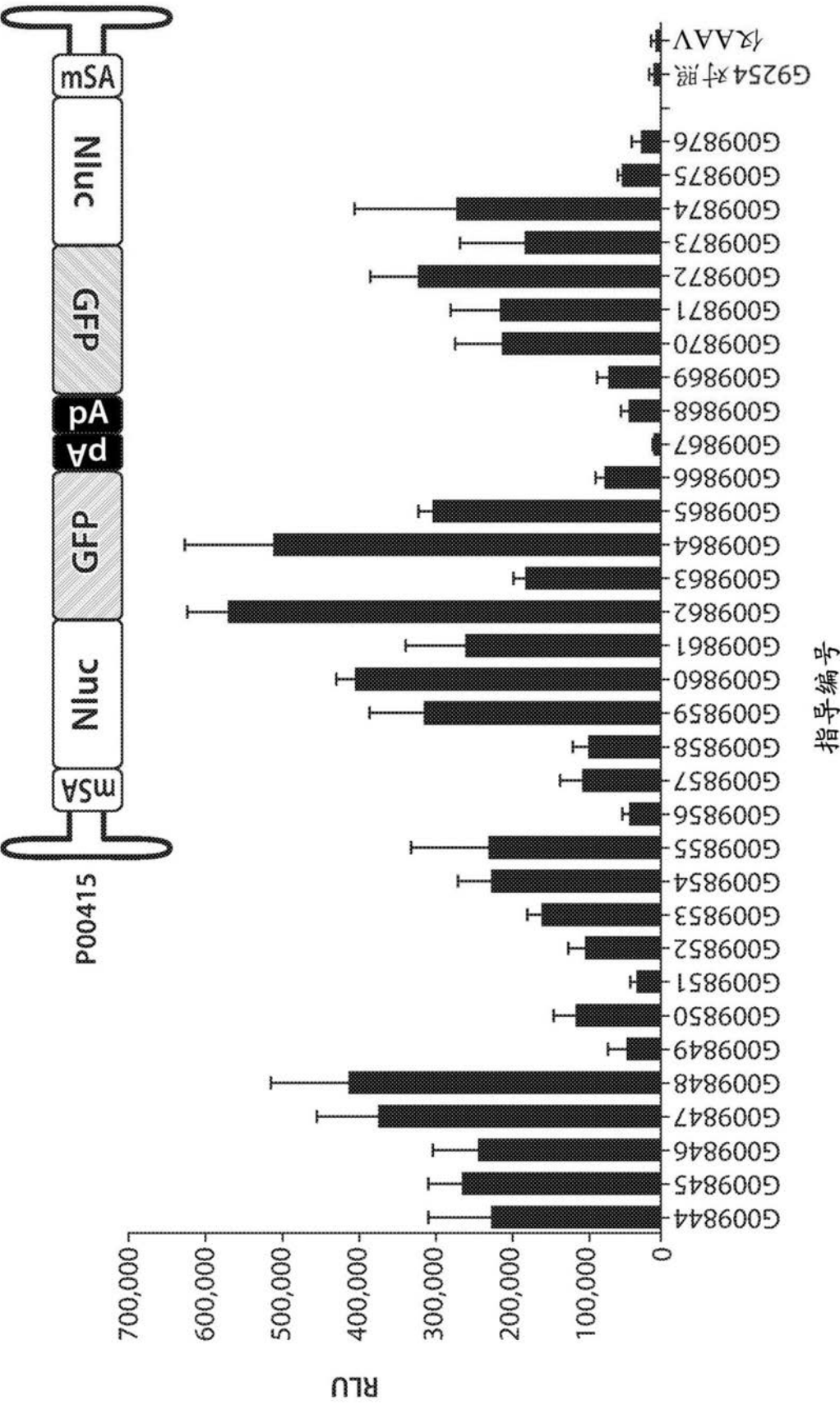


图11B

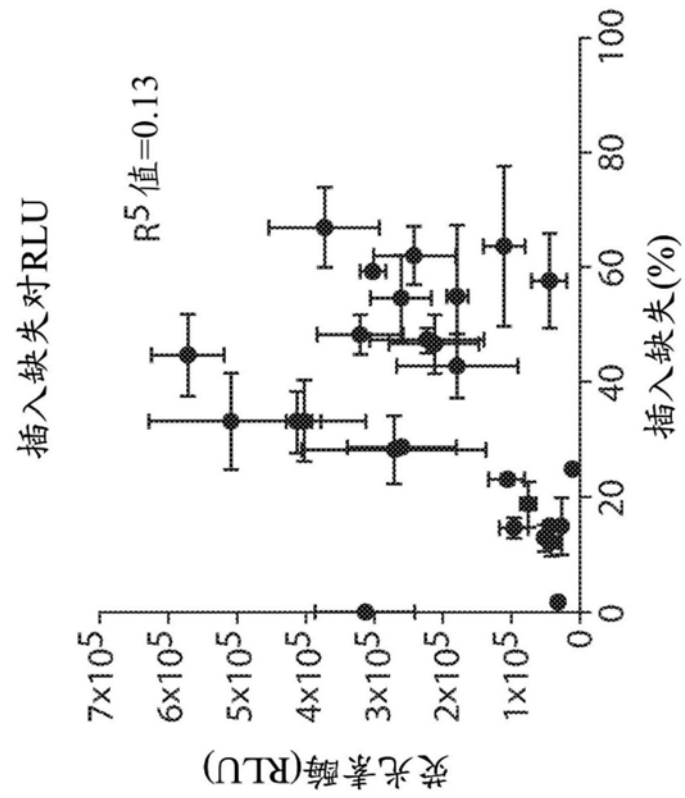


图11C

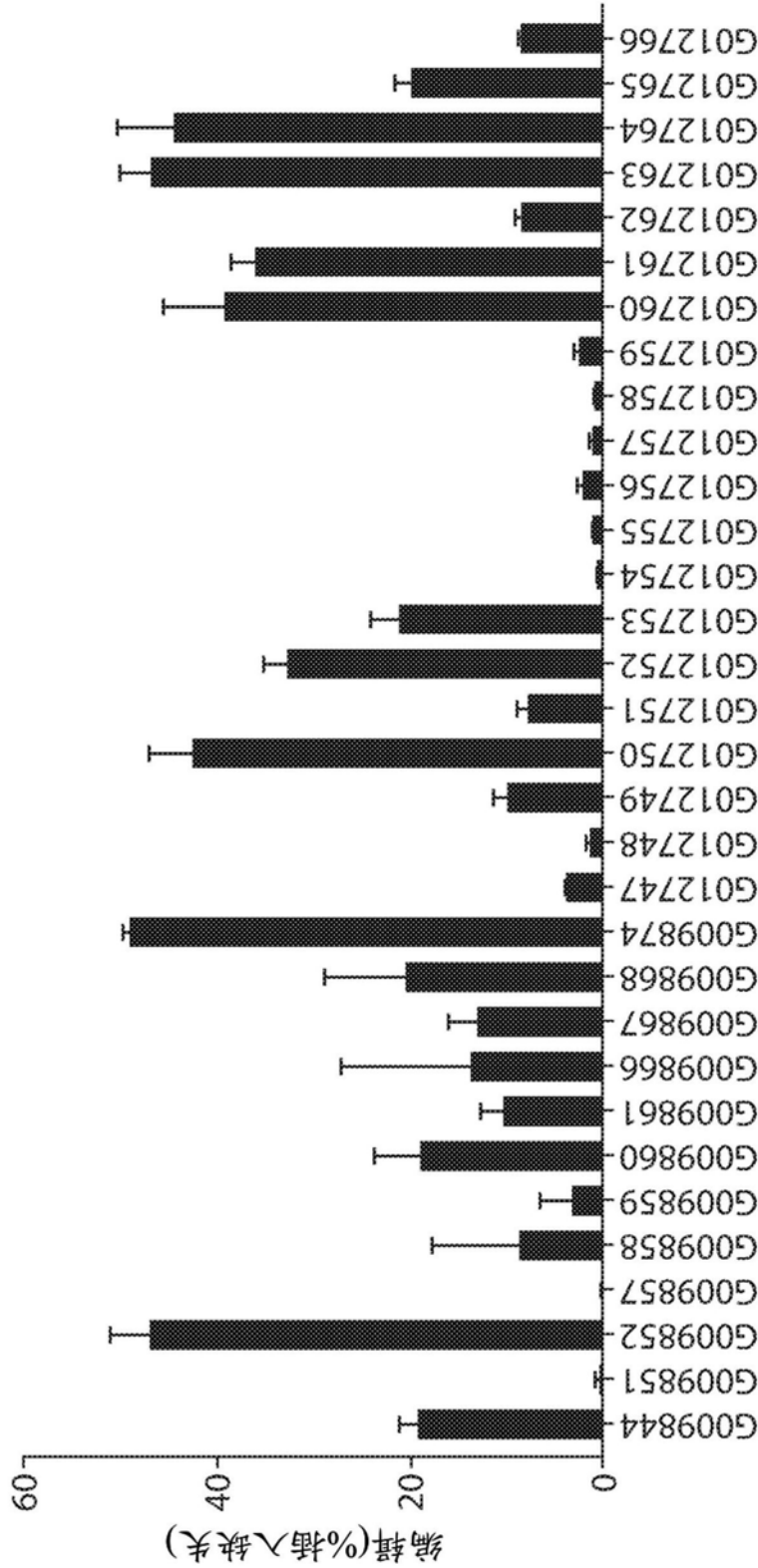


图12A

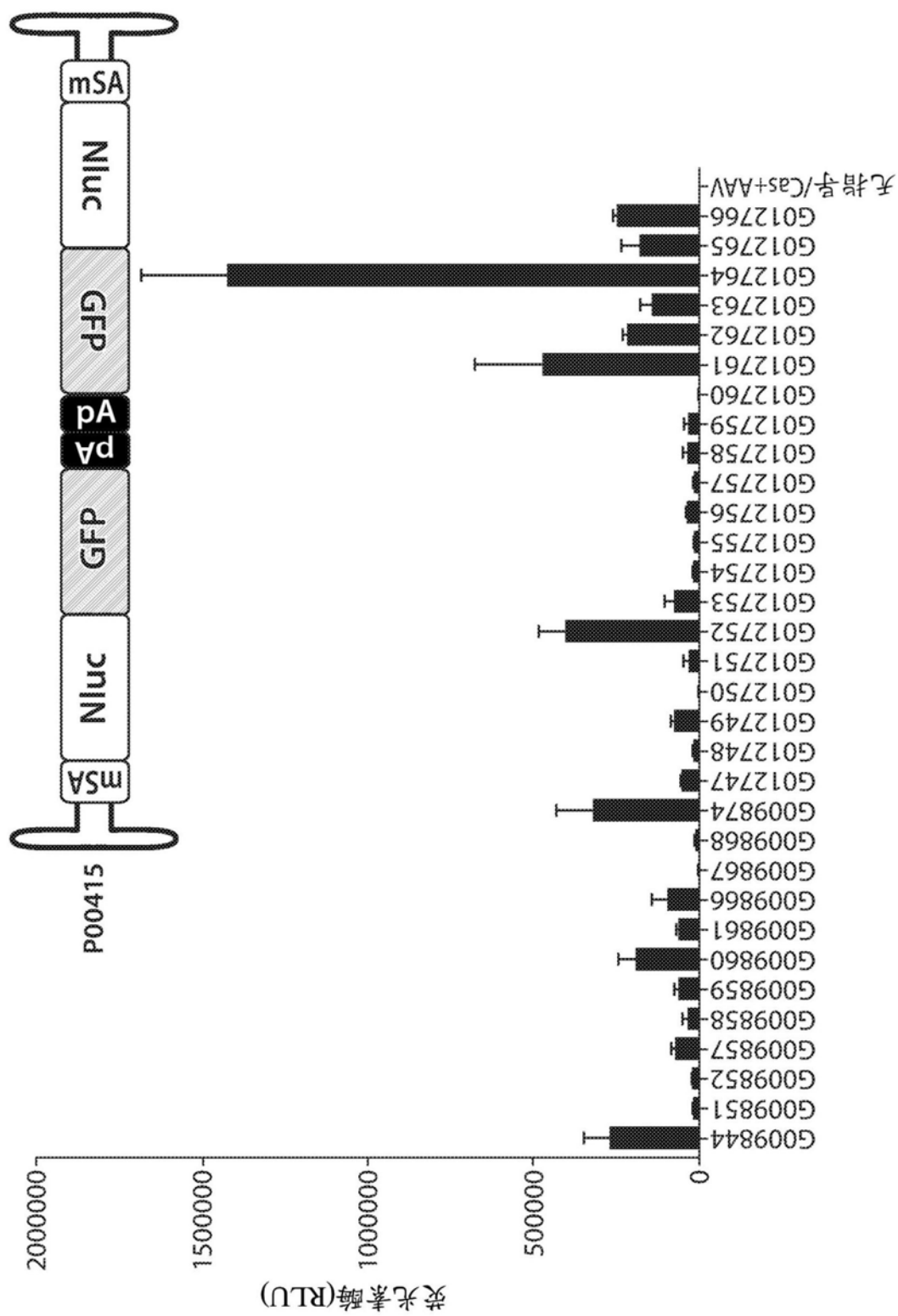


图12B

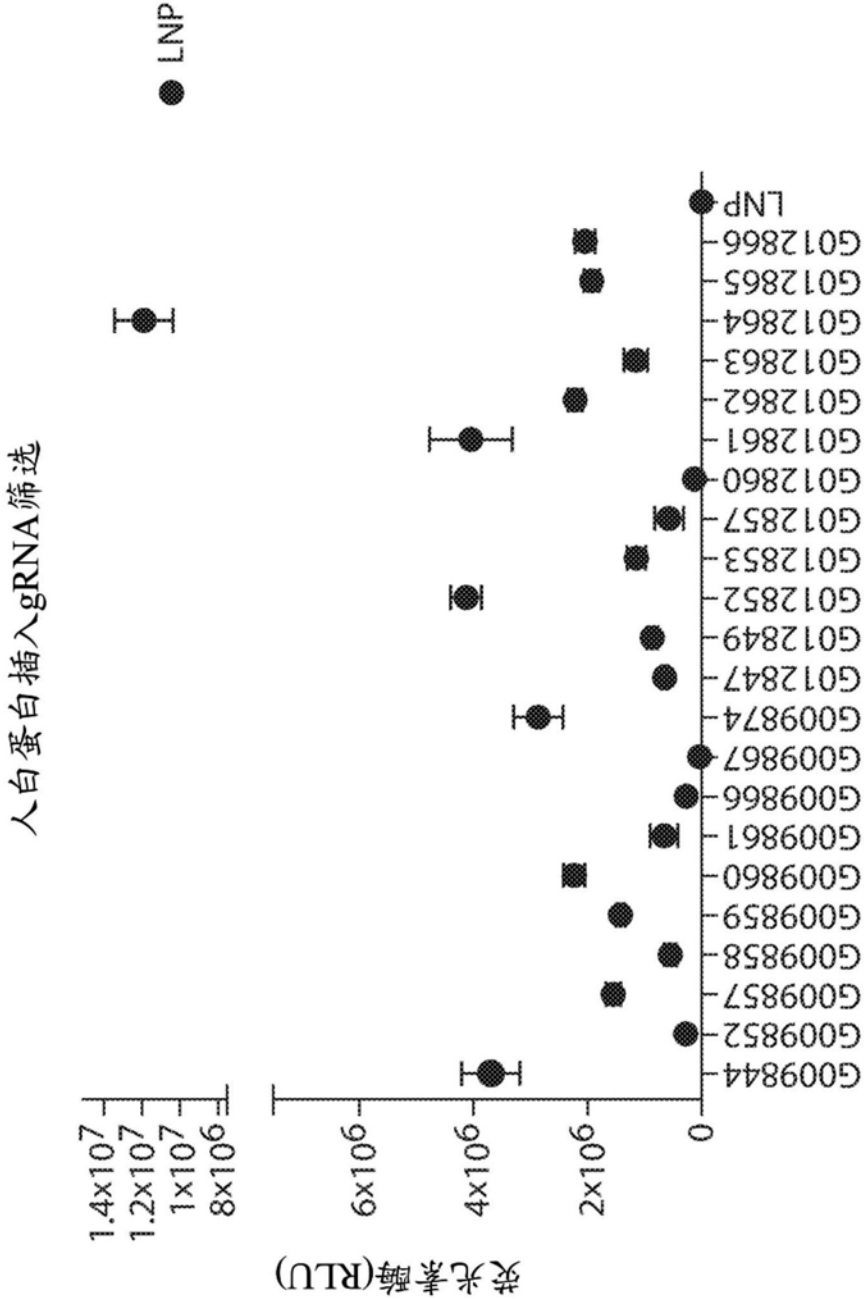


图12C

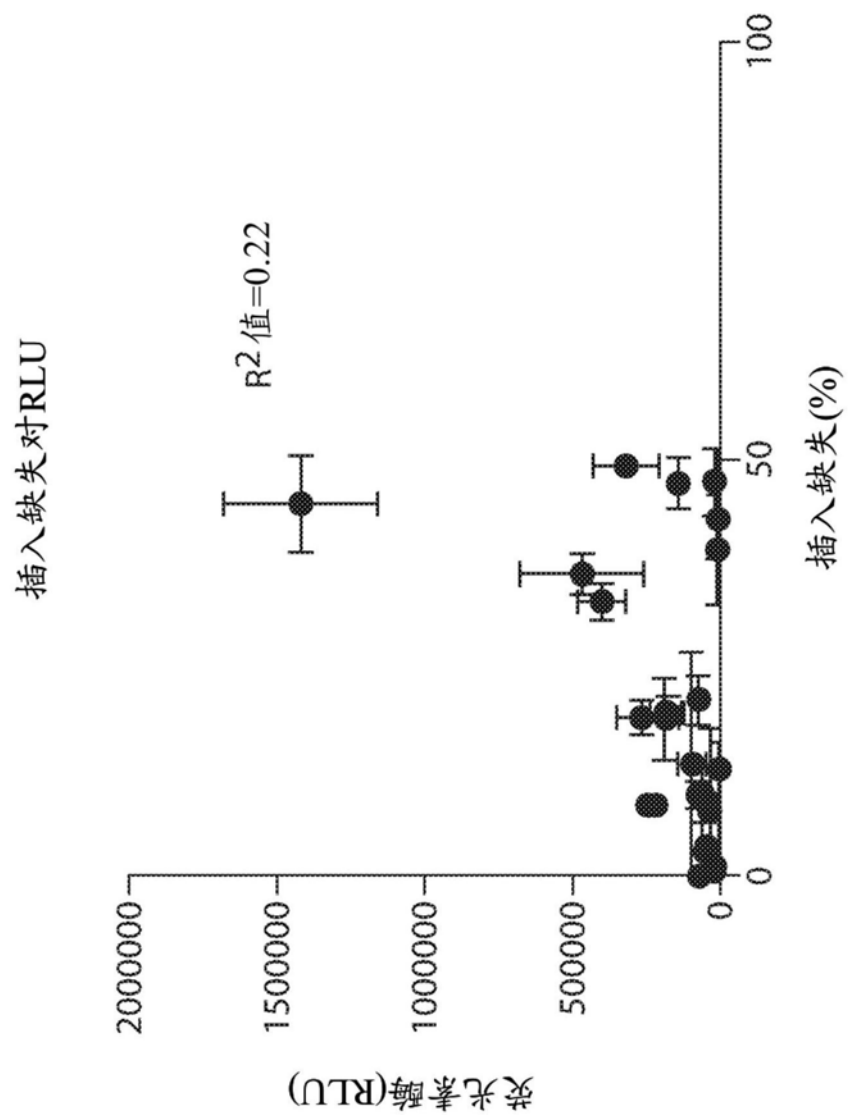


图12D

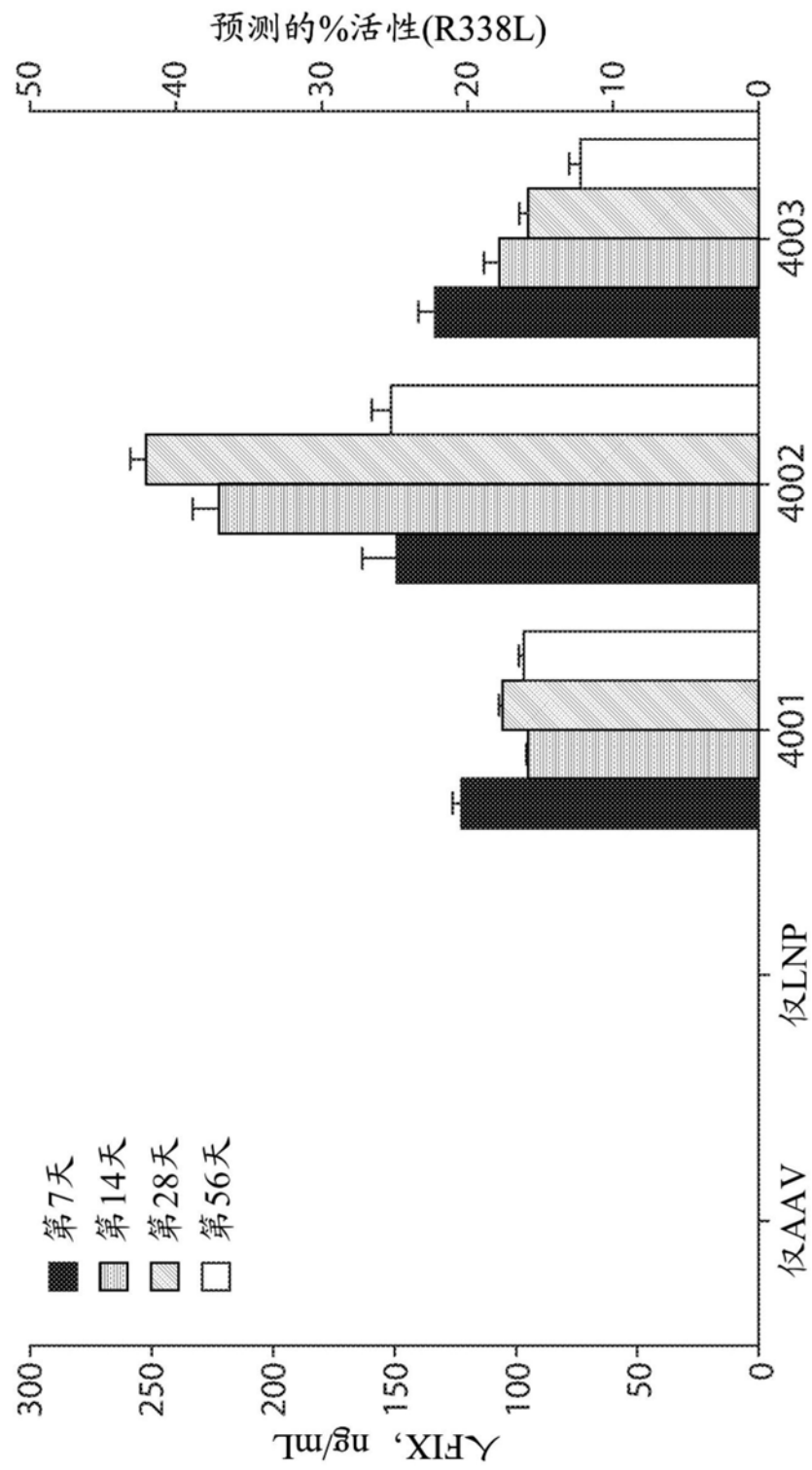


图13