(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118103393 A
(43)申请公布日 2024.05.28

(21)申请号 202280068810.3

(22)申请日 2022.08.19

(30)优先权数据
PCT/US2021/047207 2021.08.23 US
63/310,042 2022.02.14 US

(85)PCT国际申请进入国家阶段日
2024.04.11

(86)PCT国际申请的申请数据
PCT/US2022/075187 2022.08.19

(87)PCT国际申请的公布数据
WO2023/028441 EN 2023.03.02

(71)申请人 比奥维拉迪维治疗股份有限公司
地址 美国马萨诸塞州

(72)发明人 A·马戈迪亚 P·扎卡斯

(74)专利代理机构 北京坤瑞律师事务所 11494
专利代理师 封新琴

(51)Int.Cl.
C07K 14/755(2006.01)
C12N 9/64(2006.01)
C12N 15/86(2006.01)
C07K 14/75(2006.01)

权利要求书6页 说明书54页
序列表（电子公布） 附图19页

(54)发明名称

工程化的ITR序列及其使用方法

(57)摘要

本公开文本提供了核酸分子,其包含第一反向末端重复序列(ITR)、第二ITR和编码靶序列的基因盒。在一些实施方案中,所述第一ITR和/或所述第二ITR是非腺相关病毒(AAV)的ITR。还公开了在基因疗法应用中使用所述核酸分子的方法。

CN 118103393 A

1.一种核酸分子,所述核酸分子包含第一反向末端重复序列(ITR)和第二ITR,所述第一反向末端重复序列(ITR)和第二ITR侧接包含异源多核苷酸序列的基因盒,其中:

所述第一ITR包含与以下至少约75%相同的多核苷酸序列:

SEQ ID NO:1的核苷酸1-49、50-58和59-125,

SEQ ID NO:15的核苷酸1-27和50-114,或

SEQ ID NO:25;并且

所述第二ITR包含与以下至少约75%相同的多核苷酸序列:

SEQ ID NO:2的核苷酸1-67、68-76和77-125,

SEQ ID NO:16的核苷酸1-65和88-114,或

SEQ ID NO:26。

2.根据权利要求1所述的核酸分子,其中所述第一ITR包含与以下至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%相同的多核苷酸序列:

SEQ ID NO:1的核苷酸1-49、50-58和59-125,

SEQ ID NO:15的核苷酸1-27和50-114,或

SEQ ID NO:25。

3.根据权利要求1或2所述的核酸分子,其中所述第一ITR包含SEQ ID NO:1的核苷酸1-49、50-58和59-125,SEQ ID NO:15的核苷酸1-27和50-114,或SEQ ID NO:25。

4.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:1所示的多核苷酸序列。

5.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:3所示的多核苷酸序列。

6.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:5所示的多核苷酸序列。

7.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:9所示的多核苷酸序列。

8.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:13所示的多核苷酸序列。

9.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:15所示的多核苷酸序列。

10.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:17所示的多核苷酸序列。

11.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:19所示的多核苷酸序列。

12.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:21所示的多核苷酸序列。

13.根据权利要求3所述的核酸分子,其中所述第一ITR包含SEQ ID NO:25所示的多核苷酸序列。

14.根据权利要求1-13中任一项所述的核酸分子,其中所述第二ITR包含与以下至少约

80％、至少约85％、至少约90％、至少约95％、至少约96％、至少约97％、至少约98％、至少约99％相同的多核苷酸序列：

SEQ ID NO:2的核苷酸1-67、68-76和77-125，

SEQ ID NO:16的核苷酸1-65和88-114，或

SEQ ID NO:26。

15.根据权利要求12或14所述的核酸分子，其中所述第二ITR包含SEQ ID NO:2的核苷酸1-67、68-76和77-125，SEQ ID NO:16的核苷酸1-65和88-114，或SEQ ID NO:26。

16.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:2所示的多核苷酸序列。

17.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:4所示的多核苷酸序列。

18.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:6所示的多核苷酸序列。

19.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:10所示的多核苷酸序列。

20.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:14所示的多核苷酸序列。

21.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:16所示的多核苷酸序列。

22.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:18所示的多核苷酸序列。

23.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:20所示的多核苷酸序列。

24.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:22所示的多核苷酸序列。

25.根据权利要求15所述的核酸分子，其中所述第二ITR包含SEQ ID NO:26所示的多核苷酸序列。

26.一种核酸分子，所述核酸分子包含第一反向末端重复序列(ITR)和第二ITR，所述第一反向末端重复序列(ITR)和第二ITR侧接包含异源多核苷酸序列的基因盒，其中所述第一ITR选自SEQ ID NO:1、3、5、9、13、15、17、19、21、23或25所示的多核苷酸序列，并且所述第二ITR选自SEQ ID NO:2、4、6、10、14、16、18、20、22、24或26所示的多核苷酸序列。

27.根据权利要求1-26中任一项所述的核酸分子，其中所述基因盒还包含启动子。

28.根据权利要求27所述的核酸分子，其中所述启动子是组织特异性启动子。

29.根据权利要求27或28所述的核酸分子，其中所述启动子驱动所述异源多核苷酸序列在器官或组织中的表达，其中所述器官或组织包括肌肉、中枢神经系统(CNS)、眼睛、肝脏、心脏、肾脏、胰腺、肺、皮肤、膀胱、泌尿道、脾脏、髓系细胞谱系和淋巴样细胞谱系或其任何组合。

30.根据权利要求27-29中任一项所述的核酸分子，其中所述启动子驱动所述异源多核苷酸序列在以下细胞中的表达：肝细胞、上皮细胞、内皮细胞、心肌细胞、骨骼肌细胞、窦状

细胞、传入神经元、传出神经元、中间神经元、神经胶质细胞、星形胶质细胞、少突胶质细胞、小神经胶质细胞、室管膜细胞、肺上皮细胞、许旺细胞、卫星细胞、感光细胞、视网膜神经节细胞、T细胞、B细胞、NK细胞、巨噬细胞、树突细胞或其任何组合。

31.根据权利要求27-30中任一项所述的核酸分子,其中所述启动子定位于所述异源多核苷酸序列的5'。

32.根据权利要求27-31中任一项所述的核酸分子,其中所述启动子是小鼠甲状腺素转运蛋白启动子(mTTR)、天然人因子VIII启动子、人α-1-抗胰蛋白酶启动子(hAAT)、人白蛋白最小启动子、小鼠白蛋白启动子、三重四脯氨酸(TTP)启动子、CASI启动子、CAG启动子、巨细胞病毒(CMV)启动子、α1-抗胰蛋白酶(AAT)、肌肉肌酸激酶(MCK)、肌球蛋白重链α(αMHC)、肌红蛋白(MB)、结蛋白(DES)、SPc5-12、2R5Sc5-12、dMCK、tMCK或磷酸甘油酸激酶(PGK)启动子。

33.根据权利要求1至32中任一项所述的核酸分子,其中所述基因盒还包含内含子序列。

34.根据权利要求33所述的核酸分子,其中所述内含子序列定位于所述异源多核苷酸序列的5'。

35.根据权利要求33或34所述的核酸分子,其中所述内含子序列定位于所述启动子的3'。

36.根据权利要求33-35中任一项所述的核酸分子,其中所述内含子序列包含合成内含子序列。

37.根据权利要求1至36中任一项所述的核酸分子,其中所述基因盒还包含转录后调节元件。

38.根据权利要求37所述的核酸分子,其中所述转录后调节元件定位于所述异源多核苷酸序列的3'。

39.根据权利要求37或38所述的核酸分子,其中所述转录后调节元件包含突变的土拨鼠肝炎病毒调节元件(WPRE)、微RNA结合位点、DNA核靶向序列、TLR9抑制序列或其任何组合。

40.根据权利要求1至39中任一项所述的核酸分子,其中所述基因盒还包含3'UTR多(A)尾序列。

41.根据权利要求40所述的核酸分子,其中所述3'UTR多(A)尾序列选自bGH多(A)、肌动蛋白多(A)、血红蛋白多(A)及其任何组合。

42.根据权利要求1至41中任一项所述的核酸分子,其中所述基因盒还包含增强子序列。

43.根据权利要求42所述的核酸分子,其中所述增强子序列定位于所述第一ITR与所述第二ITR之间。

44.根据权利要求1至43中任一项所述的核酸分子,其中所述核酸分子从5'至3'包含:所述第一ITR、所述基因盒和所述第二ITR;其中所述基因盒包含组织特异性启动子序列、内含子序列、所述异源多核苷酸序列、转录后调节元件和3'UTR多(A)尾序列。

45.根据权利要求44所述的核酸分子,其中所述基因盒从5'至3'包含:组织特异性启动子序列、内含子序列、所述异源多核苷酸序列、转录后调节元件和3'UTR多(A)尾序列。

46.根据权利要求1至45中任一项所述的核酸分子,其中所述基因盒是单链核酸。

47.根据权利要求1至45中任一项所述的核酸分子,其中所述基因盒是双链核酸。

48.根据权利要求1至47中任一项所述的核酸分子,其中所述异源多核苷酸序列编码治疗性蛋白质。

49.根据权利要求1至48中任一项所述的核酸分子,其中所述异源多核苷酸序列编码凝血因子、生长因子、激素、细胞因子、抗体、其片段或其任何组合。

50.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码生长因子。

51.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码激素。

52.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码细胞因子。

53.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码抗体或其片段。

54.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码X连锁抗肌萎缩蛋白、MTM1(肌微管素)、酪氨酸羟化酶、AADC、环水解酶、SMN1、FXN(共济蛋白)、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA(MPS I)、IDS(MPS II)、PAH、GAA(酸性α-葡糖苷酶)、GALT、OTC、CMD1A、LAMA2或其任何组合。

55.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码微RNA(miRNA)。

56.根据权利要求55所述的核酸分子,其中所述miRNA下调包含以下的靶基因的表达:SOD1、HTT、RHO、CD38或其任何组合。

57.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码凝血因子,其中所述凝血因子包括因子I(FI)、因子II(FII)、因子III(FIII)、因子IV(FIV)、因子V(FV)、因子VI(FVI)、因子VII(FVII)、因子VIII(FVIII)、因子IX(FIX)、因子X(FX)、因子XI(FXI)、因子XII(FXII)、因子XIII(FXIII)、血管性血友病因子(VWF)、前激肽释放酶、高分子量激肽原、纤连蛋白、抗凝血酶III、肝素辅因子II、蛋白质C、蛋白质S、蛋白质Z、蛋白质Z相关蛋白酶抑制剂(ZPI)、纤溶酶原、α2-纤溶酶抑制剂、组织纤溶酶原激活物(tPA)、尿激酶、纤溶酶原激活物抑制物-1(PAI-1)、纤溶酶原激活物抑制物-2(PAI-2)或其任何组合。

58.根据权利要求1至57中任一项所述的核酸分子,其中所述异源多核苷酸序列经密码子优化。

59.根据权利要求58所述的核酸分子,其中所述异源多核苷酸序列经密码子优化以用于在人体内表达。

60.根据权利要求1至59中任一项所述的核酸分子,其中将所述核酸分子与递送剂一起配制。

61.根据权利要求60所述的核酸分子,其中所述递送剂包括脂质纳米颗粒(LNP)。

62.根据权利要求60所述的核酸分子,其中所述递送剂包括脂质体、非脂质聚合物分子、内体或其任何组合。

63.根据权利要求1至62中任一项所述的核酸分子,其中所述核酸分子被配制用于静脉

内、经皮、皮内、神经内、眼内、鞘内、皮下、肺部、口服施用或其任何组合。

64.根据权利要求63所述的核酸分子,其中所述核酸分子被配制用于静脉内施用。

65.根据权利要求1至62中任一项所述的核酸分子,其中所述核酸分子被配制用于通过原位注射施用。

66.根据权利要求1至62中任一项所述的核酸分子,其中所述核酸分子被配制用于通过吸入施用。

67.一种载体,所述载体包含根据权利要求1至59中任一项所述的核酸分子。

68.一种宿主细胞,所述宿主细胞包含根据权利要求1至59中任一项所述的核酸分子或根据权利要求67所述的载体。

69.一种药物组合物,所述药物组合物包含根据权利要求1至66中任一项所述的核酸。

70.一种药物组合物,所述药物组合物包含根据权利要求67所述的载体和药学上可接受的赋形剂。

71.一种药物组合物,所述药物组合物包含根据权利要求68所述的宿主细胞和药学上可接受的赋形剂。

72.一种试剂盒,所述试剂盒包含根据权利要求1至66中任一项所述的核酸分子和用于向有需要的受试者施用所述核酸分子的说明书。

73.一种杆状病毒系统,所述杆状病毒系统用于产生根据权利要求1至59中任一项所述的核酸分子。

74.根据权利要求73所述的杆状病毒系统,其中根据权利要求1至59中任一项所述的核酸分子是在昆虫细胞中产生的。

75.一种纳米颗粒递送系统,所述纳米颗粒递送系统包含根据权利要求1至66中任一项所述的核酸分子。

76.一种在有需要的受试者体内表达异源多核苷酸序列的方法,所述方法包括向所述受试者施用根据权利要求1至66中任一项所述的核酸分子、根据权利要求67所述的载体或根据权利要求69-71中任一项所述的药物组合物。

77.一种治疗有需要的受试者的疾病或障碍的方法,所述方法包括向所述受试者施用根据权利要求1至66中任一项所述的核酸分子、根据权利要求67所述的载体或根据权利要求69-71中任一项所述的药物组合物。

78.根据权利要求76或77所述的方法,其中所述核酸分子是静脉内、经皮、皮内、神经内、眼内、鞘内或其任何组合方式施用的。

79.根据权利要求78所述的方法,其中所述核酸分子是静脉内施用的。

80.根据权利要求78所述的方法,其中所述核酸分子是通过原位注射施用。

81.根据权利要求78所述的方法,其中所述核酸分子是通过吸入施用。

82.根据权利要求76至81中任一项所述的方法,其中所述受试者是哺乳动物。

83.根据权利要求76至81中任一项所述的方法,其中所述受试者是人。

84.根据权利要求45所述的核酸分子,其中所述基因盒包含SEQ ID NO:27的核苷酸序列。

85.根据权利要求1至49中任一项所述的核酸分子,其中所述异源多核苷酸序列编码凝血因子。

86.根据权利要求68所述的宿主细胞,其中所述宿主细胞是昆虫细胞。

87.根据权利要求1-26中任一项所述的核酸分子,其中与所述第二ITR相比,所述第一ITR衍生自不同的病毒基因组。

工程化的ITR序列及其使用方法

相关申请的交叉引用

[0001] 本申请要求2021年8月23日提交的国际申请号PCT/US2021/047207和2022年2月14日提交的美国临时专利申请号63/310,042的优先权,将所述申请的公开内容通过引用以其整体特此并入。

对电子提交的序列表的引用

[0002] 以ASCII文本文件电子提交的序列表(名称:732871_SA9_478BPC_ST26.xml;大小90.1KB;创建日期:2022年8月17日)的内容通过引用以其整体并入本文。

背景技术

[0003] 基因疗法提供了治疗多种疾病的持久手段的可能性。过去,许多基因疗法治疗通常依赖于使用病毒。有多种可以选择用于这个目的的病毒剂,每种病毒剂具有不同特性,这使得它们更适合于或更不适合于基因疗法。然而,一些病毒载体的不期望的特性已经导致了临床安全性问题并且限制了它们的治疗用途。

[0004] 腺相关病毒(AAV)是一种常见的基因疗法载体,但是它并非没有缺点。AAV基因组的编码序列侧接反向末端重复序列(ITR),所述反向末端重复序列是病毒复制和包装以及转基因表达所需。AAV ITR的T形发夹结构易于由宿主细胞蛋白质结合,抑制AAV载体中的转基因表达。需要提供靶序列的有效且持久的表达,同时避免现有AAV载体技术的局限性。

发明内容

[0005] 本文公开了核酸分子及其用途,所述核酸分子包含经修饰的第一反向末端重复序列(ITR)和/或经修饰的第二ITR,所述经修饰的第一反向末端重复序列(ITR)和/或经修饰的第二ITR侧接包含异源多核苷酸序列的基因盒。本文公开的经修饰的ITR为野生型ITR提供了更短的和/或可替代的ITR,同时保留了野生型ITR的功能特性。本文公开的经修饰ITR还为野生型ITR提供了更短的和/或可替代的ITR,同时保留了由基因盒产生的蛋白质的功能特性。

[0006] 在一方面,本文提供了一种核酸分子,所述核酸分子包含第一反向末端重复序列(ITR)和第二ITR,所述第一反向末端重复序列(ITR)和第二ITR侧接包含异源多核苷酸序列的基因盒,其中所述第一ITR包含与以下至少约75%相同的多核苷酸序列:SEQ ID NO:1的核苷酸1-49、50-58和59-125,或SEQ ID NO:15的核苷酸1-27和50-114;并且所述第二ITR包含与以下至少约75%相同的多核苷酸序列:SEQ ID NO:2的核苷酸1-67、68-76和77-125,或SEQ ID NO:16的核苷酸1-65和88-114,或SEQ ID NO 25或26。

[0007] 在一些实施方案中,所述第一ITR包含与以下至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%相同的多核苷酸序列:SEQ ID NO:1的核苷酸1-49、50-58和59-125,或SEQ ID NO:15的核苷酸1-27和50-114,或SEQ ID NO:25。在一些实施方案中,所述第一ITR包含SEQ ID NO:1的核苷酸1-49、50-58和59-125,或SEQ ID NO:15的核苷酸1-27和50-114,或SEQ ID NO:25。

[0008] 在一些实施方案中,所述第一ITR包含SEQ ID NO:1所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:3所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:5所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:9所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:13所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:15所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:17所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:19所示的多核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:21所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:23的核苷酸序列。在一些实施方案中,所述第一ITR包含SEQ ID NO:25所示的多核苷酸序列。

[0009] 在一些实施方案中,所述第二ITR包含与以下至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%相同的多核苷酸序列:SEQ ID NO:2的核苷酸1-67、68-76和77-125,或SEQ ID NO:16的核苷酸1-65和88-114,或SEQ ID NO:26。在一些实施方案中,所述第二ITR包含SEQ ID NO:2的核苷酸1-67、68-76和77-125,或SEQ ID NO:16的核苷酸1-65和88-114,或SEQ ID NO:26。在一些实施方案中,所述第二ITR包含SEQ ID NO:24的核苷酸序列。

[0010] 在一些实施方案中,所述第二ITR包含SEQ ID NO:2所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:4所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:6所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:10所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:14所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:16所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:18所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:20所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:22所示的多核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:24的核苷酸序列。在一些实施方案中,所述第二ITR包含SEQ ID NO:26所示的多核苷酸序列。

[0011] 在一些实施方案中,所述第一ITR选自SEQ ID NO:1、3、5、9、13、15、17、19、21、23或25所示的多核苷酸序列,并且所述第二ITR选自SEQ ID NO:2、4、6、10、14、16、18、20、22、24或26所示的多核苷酸序列。

[0012] 在一些实施方案中,所述核酸分子还包含启动子。在一些实施方案中,所述启动子是组织特异性启动子。在一些实施方案中,所述启动子驱动所述异源多核苷酸序列在器官或组织中的表达,其中所述器官或组织包括肌肉、中枢神经系统(CNS)、眼睛、肝脏、心脏、肾脏、胰腺、肺、皮肤、膀胱、泌尿道、脾脏、髓系细胞谱系和淋巴样细胞谱系或其任何组合。在一些实施方案中,所述启动子驱动所述异源多核苷酸序列在以下细胞中的表达:肝细胞、上皮细胞、内皮细胞、心肌细胞、骨骼肌细胞、窦状细胞、传入神经元、传出神经元、中间神经元、神经胶质细胞、星形胶质细胞、少突胶质细胞、小神经胶质细胞、室管膜细胞、肺上皮细胞、许旺细胞、卫星细胞、感光细胞、视网膜神经节细胞、T细胞、B细胞、NK细胞、巨噬细胞、树突细胞或其任何组合。在一些实施方案中,所述启动子定位于所述异源多核苷酸序列的5'。在一些实施方案中,所述启动子是小鼠甲状腺素转运蛋白启动子(mTTR)、天然人因子VIII

启动子、人α-1-抗胰蛋白酶启动子(hAAT)、人白蛋白最小启动子、小鼠白蛋白启动子、三重四脯氨酸(TTP)启动子、CASI启动子、CAG启动子、巨细胞病毒(CMV)启动子、α1-抗胰蛋白酶(AAT)、肌肉肌酸激酶(MCK)、肌球蛋白重链α(αMHC)、肌红蛋白(MB)、结蛋白(DES)、SPc5-12、2R5Sc5-12、dMCK、tMCK或磷酸甘油酸激酶(PGK)启动子。在一些实施方案中,所述启动子包含SEQ ID NO:31的核酸序列。

[0013] 在一些实施方案中,所述异源多核苷酸序列还包含内含子序列。在一些实施方案中,所述内含子序列定位于所述异源多核苷酸序列的5'。在一些实施方案中,所述内含子序列定位于所述启动子的3'。在一些实施方案中,所述内含子序列包含合成内含子序列。在一些实施方案中,所述内含子序列包含SEQ ID NO:32的核酸序列。

[0014] 在一些实施方案中,所述基因盒还包含转录后调节元件。在一些实施方案中,所述转录后调节元件定位于所述异源多核苷酸序列的3'。在一些实施方案中,所述调节元件包含突变的土拨鼠肝炎病毒转录后调节元件(WPRE)、微RNA结合位点、DNA核靶向序列、TLR9抑制序列或其任何组合。在一些实施方案中,所述转录后调节元件包含SEQ ID NO:33的核酸序列。

[0015] 在一些实施方案中,所述基因盒还包含3'UTR多(A)尾序列。在一些实施方案中,所述3'UTR多(A)尾序列选自bGH多(A)、肌动蛋白多(A)、血红蛋白多(A)及其任何组合。

[0016] 在一些实施方案中,所述基因盒还包含增强子序列。在一些实施方案中,所述增强子序列定位于第一ITR与第二ITR之间。在一些实施方案中,所述增强子包含SEQ ID NO:30的核酸序列。

[0017] 在一些实施方案中,所述核酸分子从5'至3'包含:所述第一ITR、所述基因盒和所述第二ITR;其中所述基因盒包含组织特异性启动子序列、内含子序列、所述异源多核苷酸序列、转录后调节元件和3'UTR多(A)尾序列。

[0018] 在一些实施方案中,所述基因盒从5'至3'包含:组织特异性启动子序列、内含子序列、所述异源多核苷酸序列、转录后调节元件和3'UTR多(A)尾序列。

[0019] 在一些实施方案中,所述基因盒包含单链核酸。在一些实施方案中,所述基因盒包含双链核酸。

[0020] 在一些实施方案中,所述异源多核苷酸序列编码治疗性蛋白质。

[0021] 在一些实施方案中,所述异源多核苷酸序列编码凝血因子、生长因子、激素、细胞因子、抗体、其片段或其任何组合。在一些实施方案中,所述异源多核苷酸序列编码生长因子。在一些实施方案中,所述异源多核苷酸序列编码激素。在一些实施方案中,所述异源多核苷酸序列编码细胞因子。

[0022] 在一些实施方案中,所述异源多核苷酸序列编码抗体或其片段。

[0023] 在一些实施方案中,所述异源多核苷酸序列编码X连锁抗肌萎缩蛋白、MTM1(肌微管素)、酪氨酸羟化酶、AADC、环水解酶、SMN1、FXN(共济蛋白(frataxin))、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA(MPS I)、IDS(MPS II)、PAH、GAA(酸性α-葡糖苷酶)、GALT、OTC、CMD1A、LAMA2或其任何组合。

[0024] 在一些实施方案中,所述异源多核苷酸序列编码微RNA(miRNA)。在一些实施方案中,所述miRNA下调包括以下的靶基因的表达:SOD1、HTT、RHO、CD38或其任何组合。

[0025] 在一些实施方案中,所述异源多核苷酸序列编码凝血因子,其中所述凝血因子包

含因子I(FI)、因子II(FII)、因子III(FIII)、因子IV(FIV)、因子V(FV)、因子VI(FVI)、因子VII(FVII)、因子VIII(FVIII)、因子IX(FIX)、因子X(FX)、因子XI(FXI)、因子XII(FXII)、因子XIII(FXIII)、血管性血友病因子(VWF)、前激肽释放酶、高分子量激肽原、纤连蛋白、抗凝血酶III、肝素辅因子II、蛋白质C、蛋白质S、蛋白质Z、蛋白质Z相关蛋白酶抑制剂(ZPI)、纤溶酶原、α2-纤溶酶抑制剂、组织纤溶酶原激活物(tPA)、尿激酶、纤溶酶原激活物抑制物-1(PAI-1)、纤溶酶原激活物抑制物-2(PAI2)或其任何组合。

[0026] 在一些实施方案中,所述异源多核苷酸序列经密码子优化。在一些实施方案中,所述异源多核苷酸序列经密码子优化以用于在人体内表达。

[0027] 在一些实施方案中,将所述核酸分子与递送剂一起配制。在一些实施方案中,所述递送剂包括脂质纳米颗粒。在一些实施方案中,所述递送剂包括脂质体、非脂质聚合物分子、内体或其任何组合。

[0028] 在一些实施方案中,所述核酸分子被配制用于静脉内、经皮、皮内、皮下、肺部、神经内、眼内、鞘内、或口服施用或其任何组合。在一些实施方案中,所述核酸分子被配制用于静脉内施用。在一些实施方案中,所述核酸分子被配制用于通过原位注射施用。在一些实施方案中,所述核酸分子被配制用于通过吸入施用。

[0029] 在另一方面,本文提供了一种载体,所述载体包含本文所述的核酸分子。

[0030] 在另一方面,本文提供了一种宿主细胞,所述宿主细胞包含本文所述的核酸分子或本文所述的载体。

[0031] 在另一方面,本文提供了一种药物组合物,所述药物组合物包含本文所述的核酸分子。

[0032] 在另一方面,本文提供了一种药物组合物,所述药物组合物包含本文所述的载体和药学上可接受的赋形剂。

[0033] 在另一方面,本文提供了一种药物组合物,所述药物组合物包含本文所述的宿主细胞和药学上可接受的赋形剂。

[0034] 在另一方面,本文提供了一种试剂盒,所述试剂盒包含本文所述的核酸分子和用于向有需要的受试者施用所述核酸分子的说明书。

[0035] 在另一方面,本文提供了一种杆状病毒系统,所述杆状病毒系统用于产生本文所述的核酸分子。

[0036] 在一些实施方案中,所述核酸分子是在昆虫细胞中产生的。

[0037] 在另一方面,本文提供了一种纳米颗粒递送系统,所述纳米颗粒递送系统包含本文所述的核酸分子。

[0038] 在另一方面,本文提供了一种在有需要的受试者体内表达异源多核苷酸序列的方法,所述方法包括向所述受试者施用本文所述的核酸分子、本文所述的载体或本文所述的药物组合物。

[0039] 在另一方面,本文提供了一种治疗有需要的受试者的疾病或障碍的方法,所述方法包括向所述受试者施用本文所述的核酸分子、本文所述的载体或本文所述的药物组合物。

[0040] 在一些实施方案中,将所述核酸分子通过静脉内、经皮、皮内、皮下、口服、肺部、神经内、眼内、鞘内或其任何组合施用。在一些实施方案中,所述核酸分子是静脉内施用的。在

一些实施方案中,所述核酸分子是通过原位注射施用。在一些实施方案中,所述核酸分子是通过吸入施用。

[0041] 在一些实施方案中,所述受试者是哺乳动物。在一些实施方案中,所述受试者是人。

附图说明

[0042] 图1A:野生型B19和截短型B19Δ135ITR的预测结构的图形描绘。阴影区域指示B19 Rep结合元件(RBE)。为每个ITR序列提供了吉布斯(Gibbs)自由能(ΔG)。注意:图形不代表ITR内的RBE的精确位置。

[0043] 图1B:野生型GPV和截短型GPVΔ162ITR的预测结构的图形描绘。阴影区域指示GPV Rep结合元件(RBE)。为每个ITR序列提供了吉布斯(Gibbs)自由能(ΔG)。注意:图形不代表ITR内的RBE的精确位置。

[0044] 图2A:野生型B19、B19Δ151、B19Δ223和B19_最小型截短型ITR的预测结构的图形描绘。阴影区域指示B19 Rep结合元件(RBE)。为每个ITR序列提供了吉布斯(Gibbs)自由能(ΔG)。注意:图形不代表ITR内的RBE的精确位置。

[0045] 图2B:野生型GPV、GPVΔ120、GPVΔ186和GPV_最小型截短型ITR的预测结构的图形描绘。阴影区域指示GPV Rep结合元件(RBE)。为每个ITR序列提供了吉布斯(Gibbs)自由能(ΔG)。注意:图形不代表ITR内的RBE的精确位置。

[0046] 图3:在接受34.7μg的侧接B19_min经修饰的ITR的FVIII表达构建体的ssFVIII-DNA(圆形)或dsFVIII-DNA(正方形)的HemA小鼠中测量的FVIII活性。在注射后3天和7天采集血液样品。

[0047] 图4-图9:根据SEQ ID NO:1-22的经修饰的ITR的预测DNA结构的图形描绘。使用Geneious软件(Biomatters Ltd.,新西兰奥克兰)产生预测结构。

[0048] 图10A-图10D:根据本发明的实施方案,具有经修饰的细小病毒ITR的经修饰的FVIIIXTEN表达盒的示意图。图10A示出了侧接AAV2 WT ITR(描述于USPTO申请号63/069,073中)的经修饰的FVIIIXTEN表达盒的示意性线性图谱。图10B示出了侧接HBoV1WT ITR的经修饰的FVIIIXTEN表达盒的示意性线性图谱。图10C示出了侧接B19 WT(描述于USPTO申请号63/069,073中)或B19最小型(SEQ ID:1、SEQ ID:2)ITR的经修饰的FVIIIXTEN表达盒的示意性线性图谱。图10D示出了侧接GPVΔ186(SEQ ID:9、SEQ ID:10)、GPVΔ120(SEQ ID:13、SEQ ID:14)或GPV最小型(SEQ ID:15、SEQ:16)ITR的经修饰的FVIIIXTEN表达盒的示意性线性图谱。

[0049] 图11:用于产生ssDNA的方法的示意图,其中将侧接细小病毒ITR的FVIIIXTEN表达盒用识别ITR相关序列的限制性酶来消化并且产生平端DNA。经消化的双链DNA产物(FVIII表达盒和质粒骨架)在95℃下热变性(变性),然后在4℃下冷却(复性)以允许回文ITR序列形成发夹。所得ssFVIIIXTEN(ssDNA)用于在HemA小鼠中经由流体动力学尾静脉注射而全身性施用。

[0050] 图12:通过ChromogenixCoatest® SP因子VIII生色测定测量的血浆FVIII活性水平的图示。以不同的间隔从hFVIIIR593C$^{+/+}$/HemA小鼠收集血浆样品,所述小鼠经由流体动力学尾静脉注射被全身注射200、800或1600μg/kg如所指示的两侧为人博卡病毒(HBoV1)、人

红病毒(B19)、鹅细小病毒(GPV)或它们的ITR变体或ITR组合(“杂合”)的单链V2.0 ssFVIIIXTEN(ssDNA)。误差条表示标准差。

[0051] 图13A-图13B:从杆状病毒系统中获得的纯化的ceFVIIIXTEN(ceDNA)及其体内功效研究的表示。图13A示出了从连续洗脱电泳方法获得的侧接AAV2 WT或HBoV1WT ITR的纯化的ceFVIIIXTEN(ceDNA)的琼脂糖凝胶图像。示出了与起始材料(SM)相比的纯度,其中箭头指示对应于FVIIIXTEN ceDNA载体(ceDNA)、杆状病毒DNA(vDNA)和Sf9细胞基因组DNA(gDNA)的大小的DNA条带。图13B示出了通过Chromogenix Coatest® SP因子VIII生色测定测量的血浆FVIII活性水平的图示。以不同间隔收集来自hFVIIIR593C$^{+/+}$/HemA小鼠的血浆样品,所述小鼠经由流体动力学尾静脉注射被全身性注射80、40或12μg/kg的如所指示的侧接AAV2或HBoV1 ITR的ceFVIIIXTEN(ceDNA)。误差条表示标准差。

具体实施方式

[0052] 本文公开了核酸分子及其用途,所述核酸分子包含经修饰的第一反向末端重复序列(ITR)和/或经修饰的第二ITR,所述经修饰的第一反向末端重复序列(ITR)和/或经修饰的第二ITR侧接包含异源多核苷酸序列的基因盒。在一些实施方案中,所述第一和/或第二ITR衍生自细小病毒B19或鹅细小病毒(GPV)。本文公开的经修饰的ITR为野生型ITR提供了更短的和/或可替代的ITR,同时保留了野生型ITR的功能特性。本文公开的经修饰ITR还为野生型ITR提供了更短的和/或可替代的ITR,同时保留了由基因盒产生的蛋白质的功能特性。

[0053] 在附图和序列表中展示了本公开文本的示例性构建体。为了提供说明书和权利要求的清晰理解,下文提供了以下定义。

I.定义

[0054] 应注意,术语“一个/一种(a)”或“一个/一种(an)”实体是指一个/一种或多个/多种所述实体:例如,“一个核苷酸序列”应理解为代表一个或多个核苷酸序列。类似地,“一种治疗性蛋白质”和“一种miRNA”应理解为分别代表一种或多种治疗性蛋白质和一种或多种miRNA。因此,术语“一个/一种(a)”(或“一个/一种(an)”)、“一个/一种或多个/多种”和“至少一个/一种”在本文中可互换使用。

[0055] 术语“约”在本文中用于意指大约、大致、大约地或在……左右。在术语“约”结合数字范围使用时,其通过扩展所述数值上下的边界来修饰所述范围。通常,术语“约”在本文中用于向上或向下(较高或较低)以10%的差异修饰高于和低于所述值的数值。

[0056] 同样如本文所用,“和/或”是指并且涵盖一个或多个相关列示项目的任何和所有可能组合,以及在替代方案(“或”)中解释时组合的缺少。

[0057] “核酸”、“核酸分子”、“核苷酸”、“一个或多个核苷酸的序列”和“多核苷酸”可互换使用并且是指呈单链形式或双链螺旋的核糖核苷(腺苷、鸟苷、尿苷或胞苷;“RNA分子”)或脱氧核糖核苷(脱氧腺苷、脱氧鸟苷、脱氧胸苷或脱氧胞苷;“DNA分子”)的磷酸酯聚合形式或其任何磷酸酯类似物,如硫代磷酸酯和硫代酸酯。单链核酸序列是指单链DNA(ssDNA)或单链RNA(ssRNA)。双链DNA-DNA、DNA-RNA和RNA-RNA螺旋是可能的。术语核酸分子、特别是DNA或RNA分子仅仅是指分子的一级和二级结构,并且不将其限制为任何特定三级形式。因此,这个术语包括尤其在线性或环状DNA分子(例如,限制性片段)、质粒、超螺旋化DNA和染

色体中发现的双链DNA。在讨论特定双链DNA分子的结构时,在本文中可以根据常规习惯描述序列,仅沿DNA的非转录链(即,具有与mRNA同源的序列的链)以5’至3’方向给出序列。“重组DNA分子”是已经经历分子生物学操纵的DNA分子。DNA包括但不限于cDNA、基因组DNA、质粒DNA、合成DNA和半合成DNA。本公开文本的“核酸组合物”包含如本文所述的一种或多种核酸。

[0058] 如本文所用,“反向末端重复序列”(或“ITR”)是指位于单链核酸序列的5’端或3’端的核酸子序列,其包含一组核苷酸(初始序列),之后下游是其反向互补体,即回文序列。初始序列与反向互补体之间的间插核苷酸序列可以具有任何长度,包括零。在一个实施方案中,可用于本公开文本的ITR包含一个或多个“回文序列”。ITR可以具有任何数目的功能。在一些实施方案中,本文所述的ITR形成发夹结构。在一些实施方案中,ITR形成T形发夹结构。在一些实施方案中,ITR形成非T形发夹结构,例如U形发夹结构。在一些实施方案中,ITR促进核酸分子在细胞的细胞核中的长期存活。在一些实施方案中,ITR促进核酸分子在细胞的细胞核中的永久存活(例如,持续细胞的整个寿命)。在一些实施方案中,ITR促进核酸分子在细胞的细胞核中的稳定性。在一些实施方案中,ITR促进核酸分子在细胞的细胞核中的保留。在一些实施方案中,ITR促进核酸分子在细胞的细胞核中的持久性。在一些实施方案中,ITR抑制或防止核酸分子在细胞的细胞核中的降解。

[0059] 在一个实施方案中,ITR的初始序列和/或反向互补体包含约2-600个核苷酸、约2-550个核苷酸、约2-500个核苷酸、约2-450个核苷酸、约2-400个核苷酸、约2-350个核苷酸、约2-300个核苷酸或约2-250个核苷酸。在一些实施方案中,初始序列和/或反向互补体包含约5-600个核苷酸、约10-600个核苷酸、约15-600个核苷酸、约20-600个核苷酸、约25-600个核苷酸、约30-600个核苷酸、约35-600个核苷酸、约40-600个核苷酸、约45-600个核苷酸、约50-600个核苷酸、约60-600个核苷酸、约70-600个核苷酸、约80-600个核苷酸、约90-600个核苷酸、约100-600个核苷酸、约150-600个核苷酸、约200-600个核苷酸、约300-600个核苷酸、约350-600个核苷酸、约400-600个核苷酸、约450-600个核苷酸、约500-600个核苷酸或约550-600个核苷酸。在一些实施方案中,初始序列和/或反向互补体包含约5-550个核苷酸、约5至500个核苷酸、约5-450个核苷酸、约5至400个核苷酸、约5-350个核苷酸、约5至300个核苷酸或约5-250个核苷酸。在一些实施方案中,初始序列和/或反向互补体包含约10-550个核苷酸、约15-500个核苷酸、约20-450个核苷酸、约25-400个核苷酸、约30-350个核苷酸、约35-300个核苷酸或约40-250个核苷酸。在某些实施方案中,初始序列和/或反向互补体包含约225个核苷酸、约250个核苷酸、约275个核苷酸、约300个核苷酸、约325个核苷酸、约350个核苷酸、约375个核苷酸、约400个核苷酸、约425个核苷酸、约450个核苷酸、约475个核苷酸、约500个核苷酸、约525个核苷酸、约550个核苷酸、约575个核苷酸或约600个核苷酸。在特定实施方案中,初始序列和/或反向互补体包含约400个核苷酸。

[0060] 在其他实施方案中,ITR的初始序列和/或反向互补体包含约2-200个核苷酸、约5-200个核苷酸、约10-200个核苷酸、约20-200个核苷酸、约30-200个核苷酸、约40-200个核苷酸、约50-200个核苷酸、约60-200个核苷酸、约70-200个核苷酸、约80-200个核苷酸、约90-200个核苷酸、约100-200个核苷酸、约125-200个核苷酸、约150-200个核苷酸或约175-200个核苷酸。在其他实施方案中,初始序列和/或反向互补体包含约2-150个核苷酸、约5-150个核苷酸、约10-150个核苷酸、约20-150个核苷酸、约30-150个核苷酸、约40-150个核苷酸、

约50-150个核苷酸、约75-150个核苷酸、约100-150个核苷酸或约125-150个核苷酸。在其他实施方案中,初始序列和/或反向互补体包含约2-100个核苷酸、约5-100个核苷酸、约10-100个核苷酸、约20-100个核苷酸、约30-100个核苷酸、约40-100个核苷酸、约50-100个核苷酸或约75-100个核苷酸。在其他实施方案中,初始序列和/或反向互补体包含约2-50个核苷酸、约10-50个核苷酸、约20-50个核苷酸、约30-50个核苷酸、约40-50个核苷酸、约3-30个核苷酸、约4-20个核苷酸或约5-10个核苷酸。在另一个实施方案中,初始序列和/或反向互补体由两个核苷酸、三个核苷酸、四个核苷酸、五个核苷酸、六个核苷酸、七个核苷酸、八个核苷酸、九个核苷酸、十个核苷酸、11个核苷酸、12个核苷酸、13个核苷酸、14个核苷酸、15个核苷酸、16个核苷酸、17个核苷酸、18个核苷酸、19个核苷酸或20个核苷酸组成。在其他实施方案中,初始序列与反向互补体之间的插入核苷酸是(例如,由以下组成)0个核苷酸、1个核苷酸、两个核苷酸、三个核苷酸、四个核苷酸、五个核苷酸、六个核苷酸、七个核苷酸、八个核苷酸、九个核苷酸、10个核苷酸、11个核苷酸、12个核苷酸、13个核苷酸、14个核苷酸、15个核苷酸、16个核苷酸、17个核苷酸、18个核苷酸、19个核苷酸或20个核苷酸。

[0061] 因此,如本文所用的“ITR”可以折叠回于自身并形成双链区段。例如,在折叠形成双螺旋时,序列GATCXXXXGATC包含GATC的初始序列及其互补体(3'CTAG5')。在一些实施方案中,ITR包含初始序列与反向互补体之间的连续回文序列(例如,GATCGATC)。在一些实施方案中,ITR包含初始序列与反向互补体之间的中断的回文序列(例如,GATCXXXXGATC)。在一些实施方案中,连续或中断的回文序列的互补部分彼此相互作用以形成“发夹环”结构。如本文所用,在单链核苷酸分子上的至少两个互补序列碱基配对形成双链部分时,产生“发夹环”结构。在一些实施方案中,仅ITR的一部分形成发夹环。在其他实施方案中,整个ITR形成发夹环。在一些实施方案中,ITR保留衍生出它的野生型ITR的Rep结合元件(RBE)。RBE的保留对于ITR的稳定性和制造目的可能是重要的。

[0062] 如本文所用的术语“细小病毒”涵盖细小病毒科,包括但不限于自主复制的细小病毒和依赖病毒。自主细小病毒包括例如以下属的成员:博卡病毒属(Bocavirus)、依赖病毒属、红病毒属、阿留申病毒属(Amdovirus)、细小病毒属(Parvovirus)、浓核病毒属(Densovirus)、重复病毒属(Iteravirus)、康特拉病毒属(Contravirus)、禽细小病毒属(Aveparvovirus)、反刍类细小病毒属(Copiparvovirus)、原细小病毒属(Protoparvovirus)、四型细小病毒属(Tetraparvovirus)、双义浓核病毒属(Ambidensovirus)、短浓核病毒属(Brevidensovirus)、肝胰浓核病毒属(Hepandensovirus)和对虾浓核病毒属(Penstyldensovirus)。示例性自主细小病毒包括但不限于猪细小病毒、小鼠微小病毒、犬细小病毒、水貂肠炎病毒、牛细小病毒、鸡细小病毒、猫瘟病毒、猫细小病毒、鹅细小病毒(GPV)、H1细小病毒、番鸭细小病毒、蛇细小病毒和B19病毒。其他自主细小病毒是本领域技术人员已知的。参见例如FIELDS等人VIROLOGY,第2卷,第69章(第4版,Lippincott-Raven Publishers)。

[0063] 如本文所用的术语“非AAV”涵盖来自细小病毒科的核酸、蛋白质和病毒,不包括细小病毒科的任何腺相关病毒(AAV)。“非AAV”包括但不限于以下属的自主复制成员:博卡病毒属、依赖病毒属、红病毒属、阿留申病毒属、细小病毒属、浓核病毒属、重复病毒属、康特拉病毒属、禽细小病毒属、反刍类细小病毒属、原细小病毒属、四型细小病毒属、双义浓核病毒属、短浓核病毒属、肝胰浓核病毒属和对虾浓核病毒属。

[0064] 如本文所用,术语“腺相关病毒”(AAV)包括但不限于AAV 1型、AAV 2型、AAV 3型(包括3A型和3B型)、AAV 4型、AAV 5型、AAV 6型、AAV 7型、AAV 8型、AAV 9型、AAV 10型、AAV 11型、AAV 12型、AAV 13型、蛇AAV、禽AAV、牛AAV、犬AAV、马AAV、绵羊AAV、山羊AAV、虾AAV、Gao等人(J.Virol.78:6381(2004))和Moris等人(Virol.33:375(2004))所披露的那些AAV血清型和进化枝以及目前已知或将来发现的任何其他AAV。参见例如FIELDS等人VIROLOGY,第2卷,第69章(第4版,Lippincott-Raven Publishers)。

[0065] 如本文所用,术语“衍生自”是指组分从指定分子或生物体分离或使用指定分子或生物体制备,或者信息(例如,氨基酸或核酸序列)来自指定分子或生物体。例如,衍生自第二核酸序列(例如,ITR)的核酸序列(例如,ITR)可以包括与第二核酸序列的核苷酸序列相同或基本上相似的核苷酸序列。在核苷酸或多肽的情况下,衍生的物质可以通过例如天然存在的诱变、人工定向诱变或人工随机诱变来获得。用于衍生核苷酸或多肽的诱变可以是有意定向的或有意随机的,或者是各自的混合。诱变核苷酸或多肽以产生衍生自第一核苷酸或多肽的不同核苷酸或多肽可以是随机事件(例如,由于聚合酶失真而引起),并且所衍生的核苷酸或多肽的鉴定可以通过适当的筛选方法(例如,如本文所讨论)来进行。多肽的诱变通常需要操纵编码多肽的多核苷酸。

[0066] “无衣壳(capsid-free)”或“无衣壳(capsid-less)”载体或核酸分子是指不具有衣壳的载体构建体。

[0067] 如本文所用,“编码区”或“编码序列”是多核苷酸中由可翻译成氨基酸的密码子组成的部分。虽然“终止密码子”(TAG、TGA或TAA)通常不翻译成氨基酸,但是可以认为其是编码区的一部分,但是任何侧接序列(例如启动子、核糖体结合位点、转录终止子、内含子等)不是编码区的一部分。编码区的边界通常由5’末端的起始密码子(编码所得多肽的氨基末端)和3’末端的翻译终止密码子(编码所得多肽的羧基末端)来决定。两个或更多个编码区可以存在于单一多核苷酸构建体中(例如,在单一载体上),或者存在于单独的多核苷酸构建体中(例如,在单独的(不同的)载体上)。然后,结果就是单一载体可以仅含有单个编码区,或包含两个或更多个编码区。

[0068] 哺乳动物细胞分泌的某些蛋白质与分泌信号肽相关,一旦生长中的蛋白质链跨越糙面内质网的输出已经起始所述分泌信号肽便从成熟蛋白质被切割下来。本领域普通技术人员知道,信号肽通常融合至多肽的N末端,并且从完整或“全长”多肽被切割下来以产生多肽的分泌或“成熟”形式。在某些实施方案中,天然信号肽或该序列的保留指导多肽分泌的能力的功能衍生物与所述多肽可操作地缔合。可替代地,可以使用异源哺乳动物信号肽(例如,人组织纤溶酶原激活物(TPA)或小鼠β-葡糖醛酸糖苷酶信号肽)或其功能衍生物。

[0069] 术语“下游”是指核苷酸序列位于参考核苷酸序列的3’。在某些实施方案中,下游核苷酸序列是指转录起点之后的序列。例如,基因的翻译起始密码子位于转录起始位点的下游。

[0070] 术语“上游”是指核苷酸序列位于参考核苷酸序列的5’。在某些实施方案中,上游核苷酸序列是指位于编码区或转录起点的5’侧的序列。例如,大多数启动子位于转录起始位点的上游。

[0071] 如本文所用,术语“基因盒”或“表达盒”意指能够指导特定多核苷酸序列在适当的宿主细胞中表达的DNA序列,其包含与感兴趣多核苷酸序列可操作地连接的启动子。基因盒

可以涵盖位于编码区上游(5'非编码序列)、内部或下游(3'非编码序列),且影响相关编码区的转录、RNA加工、稳定性或翻译的核苷酸序列。如果意图在真核细胞中表达编码区,则多腺苷酸化信号和转录终止序列通常将位于编码序列的3'。在一些实施方案中,基因盒包含编码基因产物的多核苷酸。在一些实施方案中,基因盒包含编码miRNA的多核苷酸。在一些实施方案中,基因盒包含异源多核苷酸序列。

[0072] 编码产物(例如,miRNA或基因产物(例如,多肽,如治疗性蛋白质))的多核苷酸可以包括与一个或多个编码区可操作地缔合的启动子和/或其他表达(例如,转录或翻译)控制序列。在可操作缔合中,以将基因产物的表达置于一个或多个调节区的影响或控制下的方式将基因产物(例如,多肽)的编码区与所述一个或多个调节区缔合。例如,如果启动子功能的诱导导致编码基因产物(由编码区编码)的mRNA的转录,并且如果启动子与编码区之间的连接的性质确实不会干扰启动子指导基因产物表达的能力或干扰DNA模板转录的能力,则编码区和启动子是"可操作地缔合"。除了启动子以外的其他表达控制序列(例如,增强子、操纵子、阻遏物和转录终止信号)也可以与编码区可操作地缔合以指导基因产物表达。

[0073] "表达控制序列"是指提供编码序列在宿主细胞中的表达的调节核苷酸序列,如启动子、增强子、终止子等。表达控制序列通常涵盖有助于与其可操作地连接的编码核酸的有效转录和翻译的任何调节核苷酸序列。表达控制序列的非限制性实例包括启动子、增强子、翻译前导序列、内含子、多腺苷酸化识别序列、RNA加工位点、效应子结合位点或茎环结构。多种表达控制序列是本领域技术人员已知的。这些包括而不限于在脊椎动物细胞中起作用的表达控制序列,如但不限于来自巨细胞病毒的启动子和增强子片段(即早期启动子,与内含子A结合)、猿猴病毒40(早期启动子)和逆转录病毒(如劳斯肉瘤病毒)。其他表达控制序列包括衍生自脊椎动物基因的那些,如肌动蛋白、热休克蛋白、牛生长激素和兔β珠蛋白,以及能够控制真核细胞中基因表达的其他序列。另外的合适的表达控制序列包括组织特异性启动子和增强子以及淋巴因子诱导型启动子(例如可由干扰素或白介素诱导的启动子)。其他表达控制序列包括内含子序列、转录后调节元件和多腺苷酸化信号。在本公开文本的其他地方讨论了另外的示例性表达控制序列。

[0074] 类似地,多种翻译控制元件是本领域普通技术人员已知的。这些翻译控制元件包括但不限于核糖体结合位点、翻译起始和终止密码子以及衍生自小核糖核酸病毒的元件(特别是内部核糖体进入位点,或IRES)。

[0075] 如本文所用的术语"表达"是指多核苷酸产生基因产物(例如RNA或多肽)的过程。它包括而不限于将多核苷酸转录成信使RNA(mRNA)、转移RNA(tRNA)、小发夹RNA(shRNA)、小干扰RNA(siRNA)或任何其他RNA产物以及将mRNA翻译为多肽。表达产生"基因产物"。如本文所用,基因产物可以是核酸,例如通过基因转录产生的信使RNA,或者是从转录物翻译的多肽。本文描述的基因产物还包括经过转录后修饰(例如多腺苷酸化或剪接)的核酸,或经过翻译后修饰(例如甲基化、糖基化、脂质的添加、与其他蛋白质亚基缔合或蛋白水解切割)的多肽。如本文所用,术语"产量"是指通过基因表达产生的多肽的量。

[0076] "载体"是指用于将核酸克隆和/或转移到宿主细胞中的任何媒介物。载体可以是复制子,另一核酸区段可以与其连接以实现所连接区段的复制。"复制子"是指在体内起自主复制单元的作用(即,能够在其自身控制下复制)的任何基因元件(例如,质粒、噬菌体、粘粒、染色体、病毒)。术语"载体"包括在体外、离体或在体内将核酸引入细胞中的媒介物。很

多载体是本领域已知和使用的,包括例如质粒、经修饰的真核病毒或经修饰的细菌病毒。通过将适当的多核苷酸片段连接到具有互补粘性末端的选择载体中,可以实现将多核苷酸插入合适的载体中。

[0077] 可以对载体进行工程化以编码选择性标记或报告物,其提供对已经并入载体的细胞的选择或鉴定。选择性标记或报告物的表达允许鉴定和/或选择并入并表达含于载体上的其他编码区的宿主细胞。本领域中已知并使用的选择性标记基因的例子包括:提供针对氨苄青霉素、链霉素、庆大霉素、卡那霉素、潮霉素、双丙氨磷除草剂、磺酰胺等的抗性的基因;和用作表型标记的基因,即花色苷调节基因、异戊烷基转移酶基因等。本领域中已知并使用的报告物的例子包括:萤光素酶(Luc)、绿色荧光蛋白(GFP)、氯霉素乙酰转移酶(CAT)、β-半乳糖苷酶(LacZ)、β-葡糖醛酸糖苷酶(Gus)等。还可以将选择性标记视为报告物。

[0078] 如本文所用的术语“宿主细胞”是指例如可以或已经用作ssDNA或载体的接受者的微生物、酵母细胞、昆虫细胞和哺乳动物细胞。所述术语包括已经转导的初始细胞的后代。因此,如本文所用的“宿主细胞”通常是指已经用外源DNA序列转导的细胞。应理解,由于天然、偶然或刻意突变,单个亲代细胞的后代的形态或基因组或总DNA互补体可能不一定与初始亲代完全相同。在一些实施方案中,宿主细胞可以是体外宿主细胞。

[0079] 术语“选择性标记”是指能够基于标记基因的效应(即,对抗生素的抗性、对除草剂的抗性、比色标记、酶、荧光标记等)加以选择的鉴定因子(通常是抗生素或化学抗性基因),其中所述效应用于跟踪感兴趣核酸的遗传和/或鉴定已经遗传感兴趣核酸的细胞或生物体。本领域中已知并使用的选择性标记基因的例子包括:提供针对氨苄青霉素、链霉素、庆大霉素、卡那霉素、潮霉素、双丙氨磷除草剂、磺酰胺等的抗性的基因;和用作表型标记的基因,即花色苷调节基因、异戊烷基转移酶基因等。

[0080] 术语“报告基因”是指编码能够基于报告基因的效应加以鉴定的鉴定因子的核酸,其中所述效应用于跟踪感兴趣核酸的遗传性、鉴定已经遗传感兴趣核酸的细胞或生物体和/或测量基因表达诱导或转录。本领域中已知并使用的报告基因的例子包括:萤光素酶(Luc)、绿色荧光蛋白(GFP)、氯霉素乙酰转移酶(CAT)、β-半乳糖苷酶(LacZ)、β-葡糖醛酸糖苷酶(Gus)等。也可以将选择性标记基因视为报告基因。

[0081] “启动子”与“启动子序列”可互换使用并且是指能够控制编码序列或功能RNA的表达的DNA序列。通常,编码序列位于启动子序列的3'。启动子可以整体衍生自天然基因,或者由衍生自在自然界中发现的不同启动子的不同元件构成,或者甚至包含合成DNA区段。本领域技术人员应理解,不同的启动子可以指导基因在不同组织或细胞类型中、或在不同发育阶段、或响应不同环境或生理条件而表达。使基因在大多数细胞类型中在大多数时间表达的启动子一般被称为“组成型启动子”。使基因在特定细胞类型中表达的启动子一般被称为“细胞特异性启动子”或“组织特异性启动子”。使基因在发育或细胞分化的特定阶段表达的启动子一般被称为“发育特异性启动子”或“细胞分化特异性启动子”。在用诱导启动子的试剂、生物分子、化学品、配体、光等暴露或处理细胞后被诱导并且使基因表达的启动子一般被称为“诱导型启动子”或“可调型启动子”。还认识到,由于在大多数情况下,尚未完全界定调节序列的确切边界,所以不同长度的DNA片段可以具有相同的启动子活性。在本公开文本的其他地方讨论了另外的示例性启动子。

[0082] 启动子序列通常在其3'末端以转录起始位点为界,并向上游(5'方向)延伸以包括

以高于背景的可检测水平起始转录所必需的最小数目的碱基或元件。在启动子序列内将发现转录起始位点(例如通过用核酸酶S1标示(mapping)方便地界定)以及负责RNA聚合酶的结合的蛋白质结合结构域(共有序列)。

[0083] 在一些实施方案中,核酸分子包含组织特异性启动子。在某些实施方案中,组织特异性启动子驱动治疗性蛋白质在肝脏中、在肝细胞和/或内皮细胞中的表达。在一个特定实施方案中,启动子包含TTP启动子。在一个特定实施方案中,启动子包含mTTR启动子。

[0084] 术语“限制性内切核酸酶”与“限制性酶”可互换使用并且是指在双链DNA内的特定核苷酸序列内结合并切割的酶。

[0085] 术语“质粒”是指染色体外元件,其通常携带不是细胞中心代谢的一部分的基因,并且通常是环状双链DNA分子的形式。此类元件可以是衍生自任何来源的单链或双链DNA或RNA的线性、环状或超螺旋化的自主复制序列、基因组整合序列、噬菌体或核苷酸序列,其中多个核苷酸序列已经接合或重组至独特构建体中,所述构建体能够将用于所选基因产物的启动子片段和DNA序列以及适当的3’非翻译序列引入细胞中。

[0086] 可以使用的真核病毒载体包括但不限于腺病毒载体、逆转录病毒载体、腺相关病毒载体、痘病毒(例如,痘苗病毒载体)、杆状病毒载体或疱疹病毒载体。非病毒载体包括质粒、脂质体、带电脂质(细胞转染剂)、DNA-蛋白质复合物和生物聚合物。

[0087] “克隆载体”是指“复制子”,其是单位长度的连续复制的核酸并且其包含复制起点,如质粒、噬菌体或粘粒,另一核酸区段可以与所述复制子连接以实现所连接区段的复制。某些克隆载体能够在一种细胞类型(例如,细菌)中复制,并在另一细胞类型(例如,真核细胞)中表达。克隆载体通常包含一个或多个可以用于选择包含载体的细胞的序列和/或一个或多个用于插入感兴趣核酸序列的多克隆位点。

[0088] 术语“表达载体”是指设计为使得所插入核酸序列在插入宿主细胞中之后能够表达的媒介物。将所插入核酸序列以如上所述的与调节区可操作地缔合地放置。

[0089] 通过本领域熟知的方法将载体引入宿主细胞中,所述方法是例如转染、电穿孔、显微注射、转导、细胞融合、DEAE葡聚糖、磷酸钙沉淀、脂质体转染(溶酶体融合)、使用基因枪或DNA载体转运蛋白。如本文所用,“培养”(“culture”,“to culture”和“culturing”)意指在体外条件下孵育细胞,所述条件允许细胞生长或分裂或使细胞维持存活状态。如本文所用,“培养的细胞”是指体外繁殖的细胞。

[0090] 如本文所用,术语“多肽”旨在涵盖单数“多肽”以及复数“多肽”,并且是指由通过酰胺键(也称为肽键)线性连接的单体(氨基酸)构成的分子。术语“多肽”是指两个或更多个氨基酸的任何一条或多条链,并且不涉及产物的具体长度。因此,肽、二肽、三肽、寡肽、“蛋白质”、“氨基酸链”或用于指两个或更多个氨基酸的一条或多条链的任何其他术语包括于“多肽”的定义内,并且术语“多肽”可以代替这些术语中的任一个术语或与其互换使用。术语“多肽”还意图指多肽的表达后修饰的产物,所述修饰包括但不限于糖基化、乙酰化、磷酸化、酰胺化、通过已知的保护/阻断基团衍生、蛋白质水解切割或通过非天然存在的氨基酸修饰。多肽可以源自天然生物来源或由重组技术产生,但不一定从指定的核酸序列翻译。其能以任何方式产生,包括通过化学合成产生。

[0091] 术语“氨基酸”包括丙氨酸(Ala或A);精氨酸(Arg或R);天冬酰胺(Asn或N);天冬氨酸(Asp或D);半胱氨酸(Cys或C);谷氨酰胺(Gln或Q);谷氨酸(Glu或E);甘氨酸(Gly或G);组

氨酸(His或H);异亮氨酸(Ile或I);亮氨酸(Leu或L);赖氨酸(Lys或K);甲硫氨酸(Met或M);苯丙氨酸(Phe或F);脯氨酸(Pro或P);丝氨酸(Ser或S);苏氨酸(Thr或T);色氨酸(Trp或W);酪氨酸(Tyr或Y);和缬氨酸(Val或V)。非传统氨基酸也在本公开文本的范围内,并且包括正亮氨酸、鸟氨酸、正缬氨酸、高丝氨酸和其他氨基酸残基类似物,如Ellman等人Meth.Enzym.202:301-336(1991)中所述的那些。为了产生此类非天然存在的氨基酸残基,可以使用Noren等人Science 244:182(1989)和Ellman等人(同上)的程序。简单来说,这些程序涉及用非天然存在的氨基酸残基化学激活抑制剂tRNA,之后进行RNA的体外转录和翻译。引入非传统氨基酸也可以使用本领域中已知的肽化学来实现。如本文所用,术语"极性氨基酸"包括具有零个净电荷,但是在其侧链的不同部分具有非零部分电荷的氨基酸(例如,M、F、W、S、Y、N、Q、C)。这些氨基酸可以参与疏水相互作用和静电相互作用。如本文所用,术语"带电氨基酸"包括在其侧链上具有非零净电荷的氨基酸(例如,R、K、H、E、D)。这些氨基酸可以参与疏水相互作用和静电相互作用。

[0092] 本公开文本还包括多肽的片段或变体及其任何组合。术语"片段"或"变体"在涉及本公开文本的多肽结合结构域或结合分子时包括保留参考多肽的至少一些特性(例如,对FcRn结合结构域或Fc变体的FcRn结合亲和力、对FVIII变体的凝结活性或对VWF片段的FVIII结合活性)的任何多肽。除了本文中其他地方讨论的特定抗体片段以外,多肽片段包括蛋白质水解片段以及缺失片段,但是不包括天然存在的全长多肽(或成熟多肽)。本公开文本的多肽结合结构域或结合分子的变体包括如上所述的片段,以及由于氨基酸取代、缺失或插入而具有改变的氨基酸序列的多肽。变体可以是天然存在的或非天然存在的。非天然存在的变体可以使用本领域中已知的诱变技术来产生。变体多肽可以包含保守或非保守氨基酸取代、缺失或添加。

[0093] "保守氨基酸取代"是用具有类似侧链的氨基酸残基替代氨基酸残基的取代。本领域已经定义了具有相似侧链的氨基酸残基家族,包括碱性侧链(例如赖氨酸、精氨酸、组氨酸)、酸性侧链(例如天冬氨酸、谷氨酸)、不带电荷的极性侧链(例如甘氨酸、天冬酰胺、谷氨酰胺、丝氨酸、苏氨酸、酪氨酸、半胱氨酸)、非极性的侧链(例如丙氨酸、缬氨酸、亮氨酸、异亮氨酸、脯氨酸、苯丙氨酸、甲硫氨酸、色氨酸)、β-分支侧链(例如苏氨酸、缬氨酸、异亮氨酸)、以及芳香族侧链(例如酪氨酸、苯丙氨酸、色氨酸、组氨酸)。因此,如果多肽中的氨基酸被来自相同侧链家族的另一种氨基酸替代,则将取代视为保守的。在另一个实施方案中,氨基酸串可以用结构上类似但侧链家族成员的顺序和/或组成不同的串来保守地替代。

[0094] 如本领域中已知的术语"同一性百分比"是两个或更多个多肽序列或者两个或更多个多核苷酸序列之间的关系,如通过比较序列所确定。在本领域中,根据具体情况,"同一性"也意指多肽或多核苷酸序列之间的序列关联性程度,如通过此类序列的串之间的匹配所确定。"同一性"可以通过已知方法容易地计算,所述已知方法包括但不限于:Computational Molecular Biology(Lesk,A.M.编)Oxford University Press,New York(1988);Biocomputing:Informatics and Genome Projects(Smith,D.W.编)Academic Press,New York(1993);Computer Analysis of Sequence Data,Part I(Griffin,A.M.和Griffin,H.G.编)Humana Press,New Jersey(1994);Sequence Analysis in Molecular Biology(von Heinje,G.编)Academic Press(1987);以及Sequence Analysis Primer(Gribskov,M.和Devereux,J.编)Stockton Press,New York(1991)。确定同一性的优选方

法设计为给出所测试序列之间的最佳匹配。确定同一性的方法被编纂于可公开获得的电脑程序中。序列比对和同一性百分比计算可以使用序列分析软件来进行,所述序列分析软件是例如LASERGENE生物信息学计算套件的Megalign程序(DNASTAR Inc.,Madison,WI)、GCG程序套件(Wisconsin Package 9.0版,Genetics Computer Group(GCG),Madison,WI)、BLASTP、BLASTN、BLASTX(Altschul等人,J.Mol.Biol.215:403(1990))和DNASTAR(DNASTAR,Inc.1228S.Park St.Madison,WI 53715USA)。在本申请的上下文中将理解,如果使用序列分析软件进行分析,除非另外指定,否则分析结果将基于所参考程序的"缺省值"。如本文所用的"缺省值"将意指在首次初始化时最初用软件加载的任一组值或参数。出于确定查询序列(例如核酸序列)与参考序列之间的百分比同一性的目的,仅使用与参考序列中的核苷酸匹配的查询序列中的核苷酸来计算同一性百分比。因此,在确定查询序列或其指定部分(例如,核苷酸1-522)与参考序列之间的同一性百分比,通过将匹配的核苷酸的数量除以完全查询序列中的核苷酸的总数来计算同一性百分比。

[0095] 如本文所用,对应于本公开文本的特定序列中的核苷酸的核苷酸是通过比对本公开文本的序列以最大化与参考序列的同一性来鉴定。用于鉴定参考序列中的等效氨基酸的编号是基于用于鉴定本公开文本的序列中的相应氨基酸的编号。

[0096] 本文所用的治疗(treat/treatment/treating)是指,例如,疾病或病症严重程度的降低;病程持续时间的缩短;与疾病或病症有关的一种或多种症状的改善;为患有疾病或病症的受试者提供有益的效果而不一定治愈所述疾病或病症或与疾病或病症相关的一种或多种症状的预防。

[0097] 如本文所用,"施用"意指经由药学上可接受的途径将本公开文本的药学上可接受的核酸分子、从所述核酸分子表达的多肽或包含所述核酸分子的载体给予至受试者。施用途径可以是静脉内的,例如静脉内注射和静脉内输注。另外的施用途径包括例如皮下、肌内、口服、经鼻和经肺施用。核酸分子、多肽和载体可以作为包含至少一种赋形剂的药物组合物的一部分来施用。

[0098] 如本文所用的术语"药学上可接受的"是指分子实体和组合物在施用至人时在生理上可耐受并且通常不会产生毒性或过敏反应或类似的不良反应(如胃部不适、头晕等)。任选地,如本文所用,术语"药学上可接受的"意指由联邦政府或州政府的管理机构批准或者在美国药典或其他公认药典中列示用于动物,更特别是用于人。

[0099] 如本文所用,词组"有需要的受试者"包括将受益于本公开文本的核酸分子、多肽或载体的施用的受试者(如哺乳动物受试者)。在一些实施方案中,受试者是人受试者。在一些实施方案中,受试者是患有血友病的个体。受试者可以是成年人或未成年人(例如,小于12岁)。

[0100] 如本文所用,术语"治疗性蛋白质"是指本领域中已知可以向受试者施用的任何多肽。在一些实施方案中,治疗性蛋白质包含选自以下的蛋白质:凝血因子、生长因子、抗体、其功能片段、或它们的组合。如本文所用,术语"凝血因子"是指天然存在的或重组产生的分子或其类似物,其预防受试者的出血发作或缩短出血发作的持续时间。换句话说,其意指具有促凝血活性(即,负责将纤维蛋白原转化为不溶性纤维蛋白的网状物,从而引起血液凝结或凝血)的分子。如本文所用的"凝血因子"包括激活的凝血因子、其酶原或可激活的凝血因子。"可激活的凝血因子"是呈无活性形式(例如,呈其酶原形式)的凝血因子,其能够被转化

为活性形式。术语“凝血因子”包括但不限于因子I(FI)、因子II(FII)、因子III(FIII)、因子IV(FIV)、因子V(FV)、因子FVI(FVI)、因子FVII(FVII)、因子FVIII(FVIII)、因子FIX(FIX)、因子X(FX)、因子XI(FXI)、因子XII(FXII)、因子XIII(FXIII)、血管性血友病因子(VWF)、前激肽释放酶、高分子量激肽原、纤连蛋白、抗凝血酶III、肝素辅因子II、蛋白质C、蛋白质S、蛋白质Z、蛋白质Z相关蛋白酶抑制剂(ZPI)、纤溶酶原、α2-纤溶酶抑制剂、组织纤溶酶原激活物(tPA)、尿激酶、纤溶酶原激活物抑制物-1(PAI-1)、纤溶酶原激活物抑制物-2(PAI-2)、其酶原、其活化形式或其任何组合。

[0101] 如本文所用的凝血活性意指参与生化反应级联的能力,所述级联以形成纤维蛋白凝块告终和/或降低出血症或出血发作的严重程度、持续时间或频率。

[0102] 如本文所用,“生长因子”包括本领域中已知的包括细胞因子和激素在内的任何生长因子。

[0103] 在一些实施方案中,治疗性蛋白质由选自以下的基因编码:X连锁抗肌萎缩蛋白、MTM1(肌微管素)、酪氨酸羟化酶、AADC、环水解酶、SMN1、FXN(共济蛋白)、GUCY2D、RS1、CFH、HTRA、ARMS、CFB/CC2、CNGA/CNGB、Prf65、ARSA、PSAP、IDUA(MPS I)、IDS(MPS II)、PAH、GAA(酸性α-葡糖苷酶)或其任何组合。

[0104] 如本文所用,术语“异源的”或“外源的”是指此类分子通常在给定背景下(例如,在细胞中或在多肽中)未发现。例如,可以将外源或异源分子引入细胞中并且仅在例如通过转染或其他形式的遗传工程化操纵细胞之后存在,或者异源氨基酸序列可以存在于其天然不存在的蛋白质中。

[0105] 如本文所用,关于核苷酸序列的术语“优化”是指编码多肽的多核苷酸序列,其中所述多核苷酸序列已经突变以增强所述多核苷酸序列的特性。在一些实施方案中,进行优化以增加转录水平、增加翻译水平、增加稳态mRNA水平、增加或减少调节蛋白(如通用转录因子)的结合、增加或减少剪接或增加多核苷酸序列产生的多肽的产量。可以对多核苷酸序列进行以使其优化的变化的例子包括密码子优化、G/C含量优化、去除重复序列、去除富含AT的元件、去除隐蔽剪接位点、去除阻遏转录或翻译的顺式作用元件、添加或去除多T或多A序列、在转录起始位点周围添加增强转录的序列(如Kozak共有序列)、去除可以形成茎环结构的序列、去除去稳定序列及其两个或更多个组合。

II.核酸分子

[0106] 本公开文本的某些方面旨在克服AAV载体用于基因疗法的缺陷。具体地,本公开文本的某些方面涉及一种核酸分子,其包含第一ITR、第二ITR和基因盒(例如,编码治疗性蛋白质和/或miRNA)。在一些实施方案中,第一ITR和第二ITR侧接包含异源多核苷酸序列的基因盒。在一些实施方案中,核酸分子不包含编码衣壳蛋白、复制蛋白和/或组装蛋白的基因。在一些实施方案中,基因盒编码治疗性蛋白质。在一些实施方案中,治疗性蛋白质包含凝血因子。在一些实施方案中,基因盒编码miRNA。在某些实施方案中,基因盒定位于第一ITR与第二ITR之间。在一些实施方案中,核酸分子还包含一个或多个非编码区。在某些实施方案中,所述一个或多个非编码区包含启动子序列、内含子、转录后调节元件、3'UTR多(A)序列或其任何组合。

[0107] 在一个实施方案中,基因盒是单链核酸。在另一个实施方案中,基因盒是双链核酸。在另一个实施方案中,基因盒是封闭端双链核酸(ceDNA)。

[0108] 在一些实施方案中,核酸分子包含:(a)第一ITR,其是衍生自细小病毒科的非AAV家族成员的ITR(例如,B19或GPV ITR);(b)组织特异性启动子序列,例如TTP或TTR启动子;(c)内含子,例如合成内含子;(d)编码miRNA或治疗性蛋白质(例如凝血因子)的核苷酸;(e)转录后调节元件,例如WPRE;(f)3'UTR多(A)尾序列,例如bGHpA;(g)第二ITR,其是衍生自细小病毒科的非AAV家族成员的ITR(例如,B19或GPV ITR)。在一些实施方案中,核酸分子包含:(a)第一ITR,其是衍生自细小病毒科的非AAV家族成员的ITR(例如,B19或GPV ITR);(b)组织特异性启动子序列,例如mTTR;(c)内含子,例如合成内含子;(d)编码miRNA或治疗性蛋白质(例如凝血因子)的核苷酸;(e)转录后调节元件,例如WPRE;(f)3'UTR多(A)尾序列,例如bGHpA;(g)第二ITR,其是衍生自细小病毒科的非AAV家族成员的ITR(例如,B19或GPV ITR)。

[0109] 在一些实施方案中,核酸分子包含第一ITR、第二ITR和编码靶序列的基因盒,其中所述靶序列编码治疗性蛋白质,其中所述治疗性蛋白质包含因子VIII(FVIII)多肽。

[0110] 在一些实施方案中,基因盒包含由mTTR启动子驱动、编码密码子优化的FVIII的核苷酸序列。在一些实施方案中,mTTR启动子包含SEQ ID NO:31的核酸序列。在一些实施方案中,基因盒还包含A1MB2增强子元件。在一些实施方案中,A1MB2增强子元件包含SEQ ID NO:30的核酸序列。在一些实施方案中,基因盒还包含嵌合或合成内含子。在一些实施方案中,嵌合内含子由鸡β-肌动蛋白/兔β-珠蛋白内含子组成并且已经修饰以消除现有五个ATG序列,从而减少假翻译起始。在一些实施方案中,内含子序列定位于编码FVIII多肽的核酸序列的5'。在一些实施方案中,嵌合内含子定位于启动子序列如mTTR启动子的5'。在一些实施方案中,嵌合内含子包含SEQ ID NO:32的核酸序列。在一些实施方案中,基因盒还包含土拨鼠转录后调节元件(WPRE)。在一些实施方案中,WPRE包含SEQ ID NO:33的核酸序列。在一些实施方案中,基因盒还包含牛生长激素多腺苷酸化(bGHpA)信号。在一些实施方案中,bGHpA信号包含SEQ ID NO:34的核酸序列。在一些实施方案中,基因盒包含与SEQ ID NO:27具有至少70%、75%、80%、85%、90%、95%或100%序列同一性的核苷酸序列。在一些实施方案中,基因盒包含SEQ ID NO:27的核苷酸序列。

[0111] 在一些实施方案中,本文公开了编码FVIII蛋白的分离的核酸分子。在一些实施方案中,本文公开了编码FVIII蛋白的分离的核酸分子,并且包含与SEQ ID NO:28至少约75%相同的核苷酸序列。在一些实施方案中,本文公开了编码FVIII蛋白的分离的核酸分子,并且包含如SEQ ID NO:28所示的核苷酸序列。

[0112] 在一些实施方案中,本文公开了编码FVIII蛋白的分离的核酸分子。在一些实施方案中,本文公开了编码FVIII蛋白的分离的核酸分子,并且包含与SEQ ID NO:29至少约75%相同的核苷酸序列。在一些实施方案中,本文公开了编码FVIII蛋白的分离的核酸分子,并且包含如SEQ ID NO:29所示的核苷酸序列。

A.反向末端重复序列(ITR)

[0113] 本公开文本的某些方面涉及一种核酸分子,其包含第一ITR(例如,5'ITR)和第二ITR(例如,3'ITR)。通常,ITR参与细小病毒(例如,AAV)从原核质粒进行DNA复制和挽救或切除(Samulski等人,1983,1987;Senapathy等人,1984;Gottlieb和Muzyczka,1988)。另外,ITR显现是AAV原病毒整合和将AAV DNA包装至病毒粒子中所需的最小序列(McLaughlin等人,1988;Samulski等人,1989)。这些元件是细小病毒基因组的有效操纵必需的。假设对于

ITR功能不可或缺的最小定义元件是Rep结合位点和末端解离位点加允许发夹形成的可变回文序列。回文核苷酸区域通常作为DNA复制的起点以及作为病毒的包装信号以顺式一起作用。ITR中的互补序列在DNA复制期间折叠成发夹结构。在一些实施方案中,ITR折叠成发夹T形结构。在其他实施方案中,ITR折叠成非T形发夹结构,例如折叠成U形发夹结构。数据表明,AAV ITR的T形发夹结构可以抑制侧接ITR的转基因的表达。参见例如,Zhou等人,Scientific Reports 7:5432(2017年7月14日)。通过利用不形成T形发夹结构的ITR,可以避免这种形式的抑制。因此,在某些方面中,与包含形成T形发夹的AAV ITR的多核苷酸相比,包含非AAV ITR的多核苷酸具有改善的转基因表达。

[0114] 在一些实施方案中,ITR包含天然存在的ITR,例如ITR包含衍生自细小病毒科的成员的ITR的全部或一部分。在一些实施方案中,ITR包含合成序列。在一个实施方案中,第一ITR或第二ITR包含合成序列。在另一个实施方案中,第一ITR和第二ITR各自包含合成序列。在一些实施方案中,第一ITR或第二ITR包含天然存在的序列。在另一个实施方案中,第一ITR和第二ITR各自包含天然存在的序列。

[0115] 在一些实施方案中,ITR包含天然存在的ITR的一部分(例如,截短型ITR)或由所述部分组成。在一些实施方案中,ITR包含天然存在的ITR的片段或由所述片段组成,其中所述片段包含至少约5个核苷酸、至少约10个核苷酸、至少约15个核苷酸、至少约20个核苷酸、至少约25个核苷酸、至少约30个核苷酸、至少约35个核苷酸、至少约40个核苷酸、至少约45个核苷酸、至少约50个核苷酸、至少约55个核苷酸、至少约60个核苷酸、至少约65个核苷酸、至少约70个核苷酸、至少约75个核苷酸、至少约80个核苷酸、至少约85个核苷酸、至少约90个核苷酸、至少约95个核苷酸、至少约100个核苷酸、至少约125个核苷酸、至少约150个核苷酸、至少约175个核苷酸、至少约200个核苷酸、至少约225个核苷酸、至少约250个核苷酸、至少约275个核苷酸、至少约300个核苷酸、至少约325个核苷酸、至少约350个核苷酸、至少约375个核苷酸、至少约400个核苷酸、至少约425个核苷酸、至少约450个核苷酸、至少约475个核苷酸、至少约500个核苷酸、至少约525个核苷酸、至少约550个核苷酸、至少约575个核苷酸或至少约600个核苷酸;其中ITR保留天然存在的ITR的功能特性。在某些实施方案中,ITR包含天然存在的ITR的片段或由所述片段组成,其中所述片段包含至少约129个核苷酸;其中ITR保留天然存在的ITR的功能特性。在某些实施方案中,ITR包含天然存在的ITR的片段或由所述片段组成,其中所述片段包含至少约102个核苷酸;其中ITR保留天然存在的ITR的功能特性。在一些实施方案中,ITR保留衍生出它的野生型ITR的Rep结合元件(RBE)。在一些实施方案中,ITR保留衍生出它的野生型ITR的RBE中的至少一个。在一些实施方案中,ITR保留衍生出它的野生型ITR的RBE或其功能部分中的至少一个。RBE的保留对于ITR的稳定性和制造目的可能是重要的。

[0116] 在一些实施方案中,ITR包含天然存在的ITR的一部分或由所述部分组成,其中所述片段包含天然存在的ITR的长度的至少约5%、至少约10%、至少约15%、至少约20%、至少约25%、至少约30%、至少约35%、至少约40%、至少约45%、至少约50%、至少约55%、至少约60%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%或至少约99%;其中所述片段保留天然存在的ITR的功能特性。

[0117] 在某些实施方案中,ITR包含如下序列或由如下序列组成,所述序列在正确比对时

与天然存在的ITR的同源部分具有至少50％、至少51％、至少52％、至少53％、至少54％、至少55％、至少56％、至少57％、至少58％、至少59％、至少60％、至少61％、至少62％、至少63％、至少64％、至少65％、至少66％、至少67％、至少68％、至少69％、至少70％、至少71％、至少72％、至少73％、至少74％、至少75％、至少76％、至少77％、至少78％、至少79％、至少80％、至少81％、至少82％、至少83％、至少84％、至少85％、至少86％、至少87％、至少88％、至少89％、至少90％、至少91％、至少92％、至少93％、至少94％、至少95％、至少96％、至少97％、至少98％、至少99％或100％的序列同一性;其中ITR保留天然存在的ITR的功能特性。在其他实施方案中,ITR包含如下序列或由如下序列组成,所述序列在正确比对时与天然存在的ITR的同源部分具有至少90％的序列同一性;其中ITR保留天然存在的ITR的功能特性。在一些实施方案中,ITR包含如下序列或由如下序列组成,所述序列在正确比对时与天然存在的ITR的同源部分具有至少80％的序列同一性;其中ITR保留天然存在的ITR的功能特性。在一些实施方案中,ITR包含如下序列或由如下序列组成,所述序列在正确比对时与天然存在的ITR的同源部分具有至少70％的序列同一性;其中ITR保留天然存在的ITR的功能特性。在一些实施方案中,ITR包含如下序列或由如下序列组成,所述序列在正确比对时与天然存在的ITR的同源部分具有至少60％的序列同一性;其中ITR保留天然存在的ITR的功能特性。在一些实施方案中,ITR包含如下序列或由如下序列组成,所述序列在正确比对时与天然存在的ITR的同源部分具有至少50％的序列同一性;其中ITR保留天然存在的ITR的功能特性。

[0118] 在一些实施方案中,ITR包含来自AAV基因组的ITR。在一些实施方案中,ITR是选自以下的AAV基因组的ITR:AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10及其任何组合。在一些实施方案中,ITR是本领域技术人员已知的任何AAV基因组的ITR,包括天然分离物,例如天然人分离物。在一个特定实施方案中,ITR是AAV2基因组的ITR。在另一个实施方案中,ITR是经基因工程化以在其5′和3′端包括衍生自一个或多个AAV基因组的ITR的合成序列。

[0119] 在一些实施方案中,ITR不是衍生自AAV基因组(即ITR衍生自不是AAV的病毒)。在一些实施方案中,ITR是非AAV的ITR。在一些实施方案中,ITR是来自选自但不限于以下的病毒科细小病毒科的非AAV基因组的ITR:博卡病毒属、依赖病毒属、红病毒属、阿留申病毒属、细小病毒属、浓核病毒属、重复病毒属、康特拉病毒属、禽细小病毒属、反刍类细小病毒属、原细小病毒属、四型细小病毒属、双义浓核病毒属、短浓核病毒属、肝胰浓核病毒属、对虾浓核病毒属及其任何组合。在某些实施方案中,ITR衍生自红病毒属细小病毒B19(人病毒)。在另一个实施方案中,ITR衍生自番鸭细小病毒(MDPV)株。在某些实施方案中,MDPV株被减弱,例如MDPV株FZ91-30。在其他实施方案中,MDPV株是病原性的,例如MDPV株YY。在一些实施方案中,ITR衍生自猪细小病毒,例如猪细小病毒U44978。在一些实施方案中,ITR衍生自小鼠微小病毒,例如小鼠微小病毒U34256。在一些实施方案中,ITR衍生自犬细小病毒,例如犬细小病毒M19296。在一些实施方案中,ITR衍生自水貂肠炎病毒,例如水貂肠炎病毒D00765。在一些实施方案中,ITR衍生自依赖细小病毒(Dependoparvovirus)。在一个实施方案中,依赖细小病毒是依赖病毒属鹅细小病毒(GPV)株。在具体实施方案中,GPV株被减弱,例如GPV株82-0321V。在另一个具体实施方案中,GPV株是病原性的,例如GPV株B。

[0120] 核酸分子的第一ITR和第二ITR可以衍生自相同基因组(例如,衍生自相同病毒的基因组),或衍生自不同基因组(例如,衍生自两种或更多种不同病毒基因组的基因组(也称

为“杂合”ITR))。在某些实施方案中,第一ITR和第二ITR衍生自相同的AAV基因组。在一个具体实施方案中,存在于本发明的核酸分子中的这两个ITR是相同的,并且特别地可以是AAV2 ITR。在其他实施方案中,第一ITR衍生自AAV基因组并且第二ITR并非衍生自AAV基因组(例如,非AAV基因组)。在其他实施方案中,第一ITR并非衍生自AAV基因组(例如,非AAV基因组)并且第二ITR衍生自AAV基因组。在再其他实施方案中,第一ITR和第二ITR二者都并非衍生自AAV基因组(例如,非AAV基因组)。在一个特定实施方案中,第一ITR和第二ITR是相同的。

[0121] 在一些实施方案中,第一ITR衍生自非AAV基因组,并且第二ITR衍生自非AAV基因组,其中第一ITR和第二ITR衍生自相同的基因组。非AAV病毒基因组的非限制性例子来自:博卡病毒属、依赖病毒属、红病毒属、阿留申病毒属、细小病毒属、浓核病毒属、重复病毒属、康特拉病毒属、禽细小病毒属、反刍类细小病毒属、原细小病毒属、四型细小病毒属、双义浓核病毒属、短浓核病毒属、肝胰浓核病毒属和对虾浓核病毒属。在一些实施方案中,第一ITR衍生自非AAV基因组,并且第二ITR衍生自非AAV基因组,其中第一ITR和第二ITR衍生自不同的病毒基因组(也称为“杂合”ITR)。在一些实施方案中,第一ITR衍生自B19而第二ITR衍生自GPV。在一些实施方案中,第一ITR衍生自GPV而第二ITR衍生自B19。

[0122] 在一些实施方案中,第一ITR衍生自AAV基因组,并且第二ITR衍生自红病毒属细小病毒B19(人病毒)。在其他实施方案中,第二ITR衍生自AAV基因组,并且第一ITR衍生自红病毒属细小病毒B19(人病毒)。

[0123] 在一些实施方案中,第一ITR包含衍生自AAV或非AAV基因组的ITR的全部或一部分或由所述ITR的全部或一部分组成,并且第二ITR包含衍生自AAV或非AAV基因组的ITR的全部或一部分或由所述ITR的全部或一部分组成。在一些实施方案中,衍生自AAV或非AAV基因组的ITR的一部分是衍生自AAV或非AAV基因组的天然存在的ITR的截短形式。在一些实施方案中,衍生自AAV或非AAV基因组的ITR的一部分包含衍生自AAV或非AAV基因组的天然存在的ITR的部分。例如,衍生自AAV或非AAV基因组的ITR的一部分包含衍生自AAV或非AAV基因组的天然存在的ITR的部分,其中至少一个RBE或其功能部分被保留。

[0124] 在某些实施方案中,第一ITR和/或第二ITR包含衍生自B19的ITR的全部或一部分,或由所述ITR的全部或一部分组成。在某些实施方案中,第一ITR和/或第二ITR包含衍生自B19的ITR的全部或一部分,或由所述ITR的全部或一部分组成。在一些实施方案中,第二ITR是第一ITR的反向互补体。在一些实施方案中,第一ITR是第二ITR的反向互补体。在一些实施方案中,衍生自B19的第一ITR和/或第二ITR能够形成发夹结构。在某些实施方案中,发夹结构不包含T形发夹。

[0125] 在一些实施方案中,第一ITR和/或第二ITR包含如下核苷酸序列或由如下核苷酸序列组成,所述核苷酸序列与SEQ ID NO:1-8、17或18所示的核苷酸序列至少约50%、至少约55%、至少约60%、至少约65%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%相同,其中第一ITR和/或第二ITR保留衍生出它的B19 ITR的功能特性。在一些实施方案中,第一ITR和/或第二ITR包含如下核苷酸序列或由如下核苷酸序列组成,所述核苷酸序列与选自SEQ ID NO:1-8、17或18的核苷酸序列至少约50%、至少约55%、至少约60%、至少约65%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%相同,其中第一ITR和/

或第二ITR能够形成发夹结构。在某些实施方案中,发夹结构不包含T形发夹。

[0126] 在一些实施方案中,第一ITR和/或第二ITR包含选自SEQ ID NO:1-8、17或18的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:1所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:2所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:3所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:4所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:5所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:6所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:7所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:8所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:17所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:18所示的核苷酸序列或由所述核苷酸序列组成。

[0127] 在一些实施方案中,第一ITR衍生自AAV基因组,并且第二ITR衍生自鹅细小病毒(GPV)。在其他实施方案中,第二ITR衍生自AAV基因组,并且第一ITR衍生自鹅细小病毒(GPV)。

[0128] 在某些实施方案中,第一ITR和/或第二ITR包含衍生自GPV的ITR的全部或一部分,或由所述ITR的全部或一部分组成。在某些实施方案中,第一ITR和/或第二ITR包含衍生自GPV的ITR的全部或一部分,或由所述ITR的全部或一部分组成。在一些实施方案中,第二ITR是第一ITR的反向互补体。在一些实施方案中,第一ITR是第二ITR的反向互补体。在一些实施方案中,衍生自GPV的第一ITR和/或第二ITR能够形成发夹结构。在某些实施方案中,发夹结构不包含T形发夹。

[0129] 在一些实施方案中,第一ITR和/或第二ITR包含如下核苷酸序列或由如下核苷酸序列组成,所述核苷酸序列与SEQ ID NO:9-16或19-22所示的核苷酸序列至少约50%、至少约55%、至少约60%、至少约65%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%相同,其中第一ITR和/或第二ITR保留衍生出它的GPV ITR的功能特性。在一些实施方案中,第一ITR和/或第二ITR包含如下核苷酸序列或由如下核苷酸序列组成,所述核苷酸序列与选自SEQ ID NO:9-16或19-22的核苷酸序列至少约50%、至少约55%、至少约60%、至少约65%、至少约65%、至少约70%、至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%相同,其中第一ITR和/或第二ITR能够形成发夹结构。在某些实施方案中,发夹结构不包含T形发夹。

[0130] 在一些实施方案中,第一ITR和/或第二ITR包含选自SEQ ID NO:9-16或19-22的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:9所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:10所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:11所示的核苷酸序列或由所述核苷酸序列组成。在一

些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:12所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:13所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:14所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:15所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:16所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:19所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:20所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:21所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:22所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:23所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:24所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:25所示的核苷酸序列或由所述核苷酸序列组成。在一些实施方案中,第一ITR和/或第二ITR包含SEQ ID NO:26所示的核苷酸序列或由所述核苷酸序列组成。

[0131] 在一些实施方案中,第一ITR和/或第二ITR包含与以下至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%相同的多核苷酸序列:SEQ ID NO:1的核苷酸1-49、50-58和59-125,或SEQ ID NO:15的核苷酸1-27和50-114,SEQ ID NO:23或SEQ ID NO:25。在一些实施方案中,第一ITR和/或第二ITR包含与以下至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%相同的多核苷酸序列:SEQ ID NO:2的核苷酸1-67、68-76和77-125,或SEQ ID NO:16的核苷酸1-65和88-114,SEQ ID NO:23或SEQ ID NO:26。

[0132] 对于本领域技术人员将理解,本文所述的任何第一ITR序列可以与本文所述的任何第二ITR序列匹配。在一些实施方案中,本文所述的第一ITR序列是5'ITR序列。在一些实施方案中,本文所述的第二ITR序列是3'ITR序列。在一些实施方案中,本文所述的第二ITR序列是5'ITR序列。在一些实施方案中,本文所述的第一ITR序列是3'ITR序列。本领域技术人员将能够确定关于基因盒的架构的第一和第二ITR的合适的取向。

[0133] 在另一个特定实施方案中,ITR是经基因工程化以在其5'和3'端包括并非衍生自AAV基因组的ITR的合成序列。在另一个特定实施方案中,ITR是经基因工程化以在其5'和3'端包括衍生自一个或多个非AAV基因组的ITR的合成序列。存在于本发明的核酸分子中的两个ITR可以是相同或不同的非AAV基因组。具体地,ITR可以衍生自相同的非AAV基因组。在一个具体实施方案中,存在于本发明的核酸分子中的这两个ITR是相同的,并且特别地可以是AAV2 ITR。

[0134] 在一些实施方案中,ITR序列包含一个或多个回文序列。本文所公开的ITR的回文序列包括但不限于天然回文序列(即,在自然界中发现的序列)、合成序列(即,未在自然界中发现的序列)(如伪回文序列)及其组合或经修饰形式。"伪回文序列"是包括不完美的回文序列在内的回文DNA序列,其与形成二级结构的天然AAV或非AAV回文序列中的序列共享

小于80％(包括小于70％、60％、50％、40％、30％、20％、10％或5％或无)核酸序列同一性。天然回文序列可以从本文所公开的任何基因组获得或衍生。合成回文序列可以基于本文所公开的任何基因组。

[0135] 回文序列可以是连续或中断的。在一些实施方案中,回文序列是中断的,其中回文序列包含第二序列的插入。在一些实施方案中,第二序列包含启动子、增强子、整合酶的整合位点(例如,Cre或Flp重组酶的位点)、基因产物的开放阅读框或其组合。

[0136] 在一些实施方案中,ITR形成发夹环结构。在一个实施方案中,第一ITR形成发夹结构。在另一个实施方案中,第二ITR形成发夹结构。仍然在另一个实施方案中,第一ITR和第二ITR二者都形成发夹结构。在一些实施方案中,第一ITR和/或第二ITR不形成T形发夹结构。在某些实施方案中,第一ITR和/或第二ITR形成非T形发夹结构。在一些实施方案中,非T形发夹结构包含U形发夹结构。

[0137] 在一些实施方案中,本文所述的核酸分子中的ITR可以是转录激活的ITR。转录激活的ITR可以包含已经通过包括至少一个转录活性元件而被转录激活的野生型ITR的全部或一部分。多种类型的转录活性元件适用于这个背景。在一些实施方案中,转录活性元件是组成型转录活性元件。组成型转录活性元件提供持续水平的基因转录,并且在需要持续表达转基因时是优选的。在其他实施方案中,转录活性元件是诱导型转录活性元件。诱导型转录活性元件通常在诱导物(或诱导条件)不存在下展现低活性,并且在诱导物存在下(或切换至诱导条件)上调。当仅在某些时间或在某些位置需要表达时或当需要使用诱导剂逐步增加表达水平时,诱导型转录活性元件可能是优选的。转录活性元件也可以具有组织特异性;也就是说,其仅在某些组织或细胞类型中展现活性。

[0138] 可以以多种方式将转录活性元件并入ITR中。在一些实施方案中,将转录活性元件并入ITR的任何部分的5'或ITR的任何部分的3'。在其他实施方案中,转录激活的ITR的转录活性元件位于两个ITR序列之间。如果转录活性元件包含两个或更多个必须被间隔开的元件,则那些元件可以与ITR的部分交替。在一些实施方案中,ITR的发夹结构缺失并且用转录元件的反向重复来替代。这后一种排列将产生模拟结构中的缺失部分的发夹。多个串联转录活性元件也可以存在于转录激活的ITR中,并且这些元件可以相邻或间隔开。另外,可以将蛋白质结合位点(例如,Rep结合位点)引入转录激活的ITR的转录活性元件中。转录活性元件可以包含使得能够通过RNA聚合酶受控转录DNA以形成RNA的任何序列,并且可以包含例如如下文所定义的转录活性元件。

[0139] 转录激活的ITR向具有相对受限的核苷酸序列长度的核酸分子提供转录激活和ITR功能二者,这有效地使可以携带并从核酸分子表达的转基因的长度最大化。将转录活性元件并入ITR中可以以多种方式来完成。对ITR序列和转录活性元件的序列需求的比较可以提供对编码ITR内的元件的方式的了解。例如,可以通过在复制转录活性元件的功能元件的ITR序列中引入特定变化将转录活性添加至ITR。本领域中存在多种技术可有效添加、缺失和/或改变特定位点的具体核苷酸序列(参见例如,Deng和Nickoloff(1992)Anal.Biochem.200:81-88)。产生转录激活的ITR的另一方式涉及在ITR中的所需位置引入限制位点。另外,可以使用本领域中已知的方法将多个转录活性元件并入转录激活的ITR中。

[0140] 通过说明的方式,可以通过包括一个或多个诸如以下的转录活性元件来产生转录

激活的ITR:TATA盒、GC盒、CCAAT盒、Sp1位点、Inr区、CRE(cAMP调节元件)位点、ATF-1/CRE位点、APBβ盒、APBα盒、CArG盒、CCAC盒或如本领域中已知参与转录的任何其他元件。

[0141] 本公开文本的多个方面提供了一种克隆本文所述的核酸分子的方法,所述方法包括将能够形成复杂二级结构的核酸分子插入合适的载体中,并且将所得载体引入合适的细菌宿主菌株中。如本领域中已知,核酸的复杂二级结构(例如,长回文区)可能不稳定并且难以在细菌宿主菌株中克隆。例如,本公开文本的包含第一ITR和第二ITR(例如,非AAV细小病毒ITR,例如B19或GPV ITR)的核酸分子可能难以使用常规方法来克隆。长DNA回文序列抑制DNA复制并且在大肠杆菌(E.coli)、芽孢杆菌属(Bacillus)、链球菌属(Steptococcus)、链霉菌属(Streptomyces)、酿酒酵母(S.cerevisiae)、小鼠和人的基因组中不稳定。这些效应是由于通过链内碱基配对形成发夹或十字形结构所致。在大肠杆菌中,在SbcC或SbcD突变体中可以显著克服对DNA复制的抑制。SbcD是核酸酶亚基,并且SbcC是SbcCD复合物的ATP酶亚基。大肠杆菌SbcCD复合物是负责防止长回文序列复制的外切核酸酶复合物。SbcCD复合物是具有ATP依赖性双链DNA外切核酸酶活性和ATP非依赖性单链DNA内切核酸酶活性的核复合物。SbcCD可以识别DNA回文序列并通过攻击所产生的发夹结构来瓦解复制叉。

[0142] 在某些实施方案中,合适的细菌宿主菌株不能拆分十字形DNA结构。在某些实施方案中,合适的细菌宿主菌株包含SbcCD复合物中的破坏。在一些实施方案中,SbcCD复合物中的破坏包含SbcC基因和/或SbcD基因中的基因破坏。在某些实施方案中,SbcCD复合物中的破坏包含SbcC基因中的基因破坏。包含SbcC基因中的基因破坏的各种细菌宿主菌株是本领域中已知的。例如而不限于,细菌宿主菌株PMC103包含基因型sbcC、recD、mcrA、ΔmcrBCF;细菌宿主菌株PMC107包含基因型recBC、recJ、sbcBC、mcrA、ΔmcrBCF;并且细菌宿主菌株SURE包含基因型recB、recJ、sbcC、mcrA、ΔmcrBCF、umuC、uvrC。因此,在一些实施方案中,克隆本文所述的核酸分子的方法包括将能够形成复杂二级结构的核酸分子插入合适的载体中,并且将所得载体引入宿主菌株PMC103、PMC107或SURE中。在某些实施方案中,克隆本文所述的核酸分子的方法包括将能够形成复杂二级结构的核酸分子插入合适的载体中,并且将所得载体引入宿主菌株PMC103中。

[0143] 合适的载体是本领域中已知的并且描述于本文中的其他地方。在某些实施方案中,用于本公开文本的克隆方法中的合适的载体是低拷贝载体。在某些实施方案中,用于本公开文本的克隆方法中的合适的载体是pBR322。

[0144] 因此,本公开文本提供了一种克隆核酸分子的方法,所述方法包括将能够形成复杂二级结构的核酸分子插入合适的载体中,和将所得载体引入包含SbcCD复合物中的破坏的细菌宿主菌株中,其中核酸分子包含第一反向末端重复序列(ITR)和第二ITR,其中第一ITR和/或第二ITR包含与SEQ ID NO:1至22所示的核苷酸序列至少约75%、至少约80%、至少约85%、至少约90%、至少约95%、至少约96%、至少约97%、至少约98%、至少约99%或100%相同的核苷酸序列或其功能衍生物。

B.治疗性蛋白质

[0145] 本公开文本的某些方面涉及一种包含第一ITR、第二ITR和编码靶序列的基因盒的核酸分子,其中靶序列编码治疗性蛋白质。在一些实施方案中,基因盒编码一种治疗性蛋白质。在一些实施方案中,基因盒编码多于一种治疗性蛋白质。在一些实施方案中,基因盒编码相同治疗性蛋白质的两个或更多个拷贝。在一些实施方案中,基因盒编码相同治疗性蛋

白质的两种或更多种变体。在一些实施方案中,基因盒编码两种或更多种不同的治疗性蛋白质。

[0146] 本公开文本的某些实施方案涉及一种包含第一ITR、第二ITR和编码治疗性蛋白质的基因盒的核酸分子,其中治疗性蛋白质包含凝血因子。在一些实施方案中,凝血因子选自FI、FII、FIII、FIV、FV、FVI、FVII、FVIII、FIX、FX、FXI、FXII、FXIII、VWF、前激肽释放酶、高分子量激肽原、纤连蛋白、抗凝血酶III、肝素辅因子II、蛋白质C、蛋白质S、蛋白质Z、蛋白质Z相关蛋白酶抑制剂(ZPI)、纤溶酶原、α2-纤溶酶抑制剂、组织纤溶酶原激活物(tPA)、尿激酶、纤溶酶原激活物抑制物-1(PAI-1)、纤溶酶原激活物抑制物-2(PAI2)、其任何酶原、其任何活性形式及其任何组合。在一个实施方案中,凝血因子包含FVIII或其变体或片段。在另一个实施方案中,凝血因子包含FIX或其变体或片段。在另一个实施方案中,凝血因子包含FVII或其变体或片段。在另一个实施方案中,凝血因子包含VWF或其变体或片段。

[0147] 在一些实施方案中,核酸分子包含第一ITR、第二ITR和编码靶序列的基因盒,其中靶序列编码治疗性蛋白质,其中治疗性蛋白质包含因子VIII多肽。除非另外指定,否则如本文所用,“因子VIII”在整个本申请中缩写为“FVIII”,意指具有其在凝结中的正常作用的功能性FVIII多肽。因此,术语FVIII包括功能性变体多肽。“FVIII蛋白”可与FVIII多肽(或蛋白质)或FVIII互换使用。FVIII功能的例子包括但不限于激活凝结的能力、作为因子IX的辅因子作用的能力或在Ca^{2+}和磷脂存在下与因子IX形成因子X酶(tenase)复合物的能力,所述复合物然后将因子X转化为激活形式Xa。

[0148] 本文所用的治疗性蛋白质中的FVIII部分具有FVIII活性。FVIII活性可以通过本领域中任何已知方法来测量。多种测试可用于评估凝结系统的功能:活化部分凝血活酶时间(aPTT)测试、生色测定、ROTEM测定、凝血酶原时间(PT)测试(也用于确定INR)、纤维蛋白原测试(通常通过克劳斯法进行)、血小板计数、血小板功能测试(通常通过PFA-100进行)、TCT、出血时间、混合测试(如果将患者的血浆与正常血浆混合,是否矫正异常)、凝结因子测定、抗磷脂抗体、D-二聚体、遗传测试(例如,因子V Leiden、凝血酶原突变G20210A)、稀释鲁塞尔蝰蛇毒时间(dRVVT)、杂项血小板功能测试、凝血弹性描记法(TEG或Sonoclot)、凝血弹性测量法(TEM®,例如ROTEM®)或优球蛋白溶解时间(ELT)。

[0149] aPTT测试是测量“固有”(也称为接触激活途径)和常见凝结途径二者的功效的性能指示物。这个测试一般用于测量市售重组凝血因子(例如,FVIII)的凝血活性。其与测量外在途径的凝血酶原时间(PT)结合使用。

[0150] ROTEM分析提供关于止血的整体动力学的信息:凝血时间、凝块形成、凝块稳定性和溶解。凝血弹性测量法中的不同参数依赖于血浆凝结系统的活性、血小板功能、纤维蛋白溶解或影响这些相互作用的许多因素。这个分析可以提供次级止血的全面见解。

[0151] 生色测定机制是基于血液凝结级联的原理,其中在激活的因子IX、磷脂和钙离子的存在下,激活的FVIII加速因子X转化为因子Xa。因子Xa活性是通过水解对因子Xa具有特异性的对硝基苯胺(pNA)底物来评估。在405nM下测量的对硝基苯胺的初始释放速率与因子Xa活性成正比,并且因此与样品中的FVIII活性成正比。生色测定得到了国际血栓形成与止血学会(ISTH)的科学标准化委员会(SSC)的FVIII和因子IX小组委员会的推荐。从1994年以来,生色测定也已经是欧洲药典中用于指定FVIII浓缩物效力的参考方法。

[0152] 在一些实施方案中,基因盒包含编码FVIII多肽的核苷酸序列,其中核苷酸序列经

密码子优化。在一些实施方案中,基因盒包含由mTTR启动子驱动、编码密码子优化的FVIII的核苷酸序列和合成内含子。在一些实施方案中,基因盒包含国际申请号PCT/US 2017/015879中披露的核苷酸序列,将所述申请通过引用以其整体并入。在一些实施方案中,基因盒是如PCT/US2017/015879中所述的“hFVIIIco6XTEN”基因盒。

[0153] 在一些实施方案中,核酸分子包含第一ITR、第二ITR和编码靶序列的基因盒,其中所述靶序列编码治疗性蛋白质,并且其中所述治疗性蛋白质包含生长因子。生长因子可以选自本领域中已知的任何生长因子。在一些实施方案中,生长因子是激素。在其他实施方案中,生长因子是细胞因子。在一些实施方案中,生长因子是趋化因子。

[0154] 在一些实施方案中,生长因子是肾上腺髓质素(AM)。在一些实施方案中,生长因子是血管生成素(Ang)。在一些实施方案中,生长因子是自分泌运动因子。在一些实施方案中,生长因子是骨形态发生蛋白(BMP)。在一些实施方案中,BMP选自BMP2、BMP4、BMP5和BMP7。在一些实施方案中,生长因子是睫状神经营养因子家族成员。在一些实施方案中,睫状神经营养因子家族成员选自睫状神经营养因子(CNTF)、白血病抑制因子(LIF)、白介素-6(IL-6)。在一些实施方案中,生长因子是集落刺激因子。在一些实施方案中,集落刺激因子选自巨噬细胞集落刺激因子(m-CSF)、粒细胞集落刺激因子(G-CSF)和粒细胞巨噬细胞集落刺激因子(GM-CSF)。在一些实施方案中,生长因子是表皮生长因子(EGF)。在一些实施方案中,生长因子是肝配蛋白。在一些实施方案中,肝配蛋白选自肝配蛋白A1、肝配蛋白A2、肝配蛋白A3、肝配蛋白A4、肝配蛋白A5、肝配蛋白B1、肝配蛋白B2和肝配蛋白B3。在一些实施方案中,生长因子是促红细胞生成素(EPO)。在一些实施方案中,生长因子是成纤维细胞生长因子(FGF)。在一些实施方案中,FGF选自FGF1、FGF2、FGF3、FGF4、FGF5、FGF6、FGF7、FGF8、FGF9、FGF10、FGF11、FGF12、FGF13、FGF14、FGF15、FGF16、FGF17、FGF18、FGF19、FGF20、FGF21、FGF22和FGF23。在一些实施方案中,生长因子是胎牛促体素(FBS)。在一些实施方案中,生长因子是GDNF家族成员。在一些实施方案中,GDNF家族成员选自神经胶质细胞系衍生的神经营养因子(GDNF)、神经秩蛋白、普塞芬蛋白和阿特敏蛋白。在一些实施方案中,生长因子是生长分化因子-9(GDF9)。在一些实施方案中,生长因子是肝细胞生长因子(HGF)。在一些实施方案中,生长因子是肝细胞瘤衍生的生长因子(HDGF)。在一些实施方案中,生长因子是胰岛素。在一些实施方案中,生长因子是胰岛素样生长因子。在一些实施方案中,胰岛素样生长因子是胰岛素样生长因子-1(IGF-1)或IGF-2。在一些实施方案中,生长因子是白介素(IL)。在一些实施方案中,IL选自IL-1、IL-2、IL-3、IL-4、IL-5、IL-6和IL-7。在一些实施方案中,生长因子是角质形成细胞生长因子(KGF)。在一些实施方案中,生长因子是迁移刺激因子(MSF)。在一些实施方案中,生长因子是巨噬细胞刺激蛋白(MSP或肝细胞生长因子样蛋白(HGFLP))。在一些实施方案中,生长因子是肌生成抑制蛋白(GDF-8)。在一些实施方案中,生长因子是神经调节蛋白。在一些实施方案中,神经调节蛋白选自神经调节蛋白1(NRG1)、NRG2、NRG3和NRG4。在一些实施方案中,生长因子是神经营养蛋白。在一些实施方案中,生长因子是脑源性神经营养因子(BDNF)。在一些实施方案中,生长因子是神经生长因子(NGF)。在一些实施方案中,NGF是神经营养蛋白-3(NT-3)或NT-4。在一些实施方案中,生长因子是胎盘生长因子(PGF)。在一些实施方案中,生长因子是血小板源性生长因子(PDGF)。在一些实施方案中,生长因子是肾胺酶(RNLS)。在一些实施方案中,生长因子是T细胞生长因子(TCGF)。在一些实施方案中,生长因子是血小板生成素(TPO)。在一些实施方案中,生长因子

是转化生长因子。在一些实施方案中,转化生长因子是转化生长因子α(TGF-α)或TGF-β。在一些实施方案中,生长因子是肿瘤坏死因子-α(TNF-α)。在一些实施方案中,生长因子是血管内皮生长因子(VEGF)。

C.表达控制序列

[0155] 在一些实施方案中,本公开文本的核酸分子还包含至少一个表达控制序列。如本文所用的表达控制序列是有助于与其可操作地连接的编码核酸的有效转录和翻译的任何调节核苷酸序列,如启动子序列或启动子-增强子组合。例如,本公开文本的核酸分子可以可操作地连接至至少一个转录控制序列。基因表达控制序列可以是例如哺乳动物或病毒启动子,如组成型或诱导型启动子。

[0156] 组成型哺乳动物启动子包括但不限于以下基因的启动子:次黄嘌呤磷酸核糖基转移酶(HPRT)、腺苷脱氨酶、丙酮酸激酶、β-肌动蛋白启动子和其他组成型启动子。在真核细胞中组成性地发挥作用的示例性病毒启动子包括例如来自以下病毒的启动子:巨细胞病毒(CMV)、猿猴病毒(例如,SV40)、乳头瘤病毒、腺病毒、人类免疫缺陷病毒(HIV)、劳斯肉瘤病毒、巨细胞病毒、莫洛尼氏白血病病毒的长末端重复(LTR)和其他逆转录病毒、以及单纯疱疹病毒的胸苷激酶启动子。其他组成型启动子是本领域普通技术人员已知的。可用作本公开文本的基因表达序列的启动子还包括诱导型启动子。诱导型启动子是在诱导剂的存在下表达。例如,在某些金属离子的存在下诱导金属硫蛋白启动子以促进转录和翻译。其他诱导型启动子是本领域普通技术人员已知的。

[0157] 在一个实施方案中,本公开文本包括转基因在组织特异性启动子和/或增强子控制下的表达。在另一个实施方案中,启动子或其他表达控制序列选择性地增强转基因在肝细胞中的表达。在某些实施方案中,启动子或其他表达控制序列选择性地增强转基因在肝细胞、窦状细胞和/或内皮细胞中的表达。在一个特定实施方案中,启动子或其他表达控制序列选择性地增强转基因在内皮细胞中的表达。在某些实施方案中,启动子或其他表达控制序列选择性地增强转基因在以下中的表达:肌肉细胞、中枢神经系统、眼睛、肝脏、心脏或其任何组合。肝脏特异性启动子的例子包括但不限于小鼠甲状腺素启动子(mTTR)、内源人因子VIII启动子、人α-1-抗胰蛋白酶启动子(hAAT)、人白蛋白最小启动子和小鼠白蛋白启动子。在特定实施方案中,启动子包含mTTR启动子。mTTR启动子描述于R.H.Costa等人,1986,Mol.Cell.Biol.6:4697中。FVIII启动子描述于Figueiredo和Brownlee,1995,J.Biol.Chem.270:11828-11838中。在一些实施方案中,启动子选自肝脏特异性启动子(例如,α1-抗胰蛋白酶(AAT))、肌肉特异性启动子(例如,肌肉肌酸激酶(MCK)、肌球蛋白重链α(αMHC)、肌红蛋白(MB)和结蛋白(DES))、合成启动子(例如,SPc5-12、2R5Sc5-12、dMCK和tMCK)及其任何组合。

[0158] 在一个实施方案中,启动子选自小鼠甲状腺素转运蛋白启动子(mTTR)、天然人因子VIII启动子、人α-1-抗胰蛋白酶启动子(hAAT)、人白蛋白最小启动子、小鼠白蛋白启动子、TTPp、CASI启动子、CAG启动子、巨细胞病毒(CMV)启动子、α1-抗胰蛋白酶(AAT)、肌肉肌酸激酶(MCK)、肌球蛋白重链α(αMHC)、肌红蛋白(MB)、结蛋白(DES)、SPc5-12、2R5Sc5-12、dMCK和tMCK、磷酸甘油酸激酶(PGK)启动子及其任何组合。在一些实施方案中,启动子是TTP启动子。在一些实施方案中,启动子是小鼠甲状腺素转运蛋白启动子(mTTR)启动子。

[0159] 可以使用一种或多种增强子进一步增强表达水平以实现治疗功效。一种或多种增

强子可以单独提供,或与一种或多种启动子元件一起提供。通常,表达控制序列包含多个增强子元件和组织特异性启动子。在一个实施方案中,增强子包含一个或多个拷贝的α-1-微球蛋白/双库尼茨抑制剂(bikunin)增强子(Rouet等人,1992,J.Biol.Chem.267:20765-20773;Rouet等人,1995,Nucleic Acids Res.23:395-404;Rouet等人,1998,Biochem.J.334:577-584;Ill等人,1997,Blood Coagulation Fibrinolysis 8:S23-S30)。在另一个实施方案中,增强子衍生自肝脏特异性转录因子结合位点,如EBP、DBP、HNF1、HNF3、HNF4、HNF6和Enh1,包含HNF1、(有义)-HNF3、(有义)-HNF4、(反义)-HNF1、(反义)-HNF6、(有义)-EBP、(反义)-HNF4(反义)。

[0160] 在一个实施方案中,本公开文本的核酸分子还包含内含子序列。在一些实施方案中,内含子序列定位于编码FVIII多肽的核酸序列的5'。在一些实施方案中,内含子序列是天然存在的内含子序列。在一些实施方案中,内含子序列是合成序列。在一些实施方案中,内含子序列衍生自天然存在的内含子序列。在某些实施方案中,内含子序列包含SV40小T内含子。

[0161] 在一些实施方案中,核酸分子还包含转录后调节元件。在某些实施方案中,调节元件包含土拨鼠肝炎病毒调节元件(WPRE)。在一些实施方案中,WPRE被突变。

[0162] 在一些实施方案中,核酸分子包含微RNA(miRNA)结合位点。在一个实施方案中,miRNA结合位点是miR-142-3p的miRNA结合位点。在其他实施方案中,miRNA结合位点选自Rennie等人,RNA Biol.13(6):554-560(2016)所披露的miRNA结合位点以及可在http://sfold.wadsworth.org/starmirDB.php获得的STarMirDB,将上述文献通过引用以其整体并入本文。

[0163] 在一些实施方案中,核酸分子包含一个或多个DNA核靶向序列(DTS)。DTS促进含有此类序列的DNA分子易位至核中。在某些实施方案中,DTS包含SV40增强子序列。在某些实施方案中,DTS包含c-Myc增强子序列。在一些实施方案中,DTS介于第一ITR与第二ITR之间。在一些实施方案中,DTS在第一ITR的3'和治疗性蛋白质的5'。在其他实施方案中,DTS在治疗性蛋白质的3'和第二ITR的5'。

[0164] 在一些实施方案中,核酸分子还包含3'UTR多(A)尾序列。在一个实施方案中,3'UTR多(A)尾序列包含bGH多(A)。在一个实施方案中,3'UTR多(A)尾包含肌动蛋白多(A)位点。在一个实施方案中,3'UTR多(A)尾包含血红蛋白多(A)位点。

[0165] 在一些实施方案中,转基因表达靶向肝脏。在某些实施方案中,转基因表达靶向肝细胞。在其他实施方案中,转基因表达靶向内皮细胞。在一个特定实施方案中,转基因表达靶向天然表达内源FVIII的任何组织。

[0166] 在一些实施方案中,转基因表达靶向中枢神经系统。在某些实施方案中,转基因表达靶向神经元。在一些实施方案中,转基因表达靶向传入神经元。在一些实施方案中,转基因表达靶向传出神经元。在一些实施方案中,转基因表达靶向中间神经元。在一些实施方案中,转基因表达靶向神经胶质细胞。在一些实施方案中,转基因表达靶向星形胶质细胞。在一些实施方案中,转基因表达靶向少突胶质细胞。在一些实施方案中,转基因表达靶向小神经胶质细胞。在一些实施方案中,转基因表达靶向室管膜细胞。在一些实施方案中,转基因表达靶向许旺细胞。在一些实施方案中,转基因表达靶向卫星细胞。

[0167] 在一些实施方案中,转基因表达靶向肌肉组织。在一些实施方案中,转基因表达靶

向平滑肌。在一些实施方案中,转基因表达靶向心肌。在一些实施方案中,转基因表达靶向骨骼肌。

[0168] 在一些实施方案中,转基因表达靶向眼睛。在一些实施方案中,转基因表达靶向感光细胞。在一些实施方案中,转基因表达靶向视网膜神经节细胞。

III.宿主细胞

[0169] 本公开文本还提供了一种包含本公开文本的核酸分子或载体的宿主细胞。如本文所用,术语"转化"应在广义上使用,是指将DNA引入受体宿主细胞中,这改变了基因型并因此导致受体细胞的变化。

[0170] "宿主细胞"是指已经用使用重组DNA技术构建并且编码至少一种异源基因的载体转化的细胞。本公开文本的宿主细胞优选地具有哺乳动物来源;最优选地具有人或小鼠来源。相信本领域技术人员有能力优先地确定最适合于其目的的特定宿主细胞系。示例性宿主细胞系包括但不限于CHO、DG44和DUXB11(中国仓鼠卵巢系,DHFR-)、HELA(人宫颈癌)、CVI(猴肾系)、COS(具有SV40 T抗原的CVI的衍生物)、R1610(中国仓鼠成纤维细胞)、BALBC/3T3(小鼠成纤维细胞)、HAK(仓鼠肾系)、SP2/0(小鼠骨髓瘤)、P3x63-Ag8.653(小鼠骨髓瘤)、BFA-1c1BPT(牛内皮细胞)、RAJI(人淋巴细胞)、PER.C6®、NS0、CAP、BHK21和HEK 293(人肾)。在一个特定实施方案中,宿主细胞选自:CHO细胞、HEK293细胞、BHK21细胞、PER.C6®细胞、NS0细胞、CAP细胞及其任何组合。在一些实施方案中,本公开文本的宿主细胞具有昆虫来源。在一个特定实施方案中,宿主细胞是SF9细胞。宿主细胞系典型地可从商业服务、美国组织培养物保藏中心(American Tissue Culture Collection)或公开的文献中获得。

[0171] 将本公开文本的核酸分子或载体引入宿主细胞中可以通过本领域技术人员熟知的多种技术来完成。这些技术包括但不限于转染(包括电泳和电穿孔)、原生质体融合、磷酸钙沉淀、与包膜DNA的细胞融合、显微注射和完整病毒感染。参见Ridgway,A.A.G."Mammalian Expression Vectors"第24.2章,第470-472页Vectors,Rodriguez和Denhardt编(Butterworths,Boston,Mass.1988)。最优选地,通过电穿孔将质粒引入宿主中。使转化的细胞在适合于产生轻链和重链的条件下生长,并且测定重链和/或轻链蛋白质合成。示例性测定技术包括酶联免疫吸附测定(ELISA)、放射免疫测定(RIA)或荧光激活细胞分选仪分析(FACS)、免疫组织化学等。

[0172] 使包含本公开文本的分离的核酸分子或载体的宿主细胞在适当生长培养基中生长。如本文所用,术语"适当生长培养基"是指含有细胞生长所需的营养物的培养基。细胞生长所需的营养素可以包括碳源、氮源、必需氨基酸、维生素、矿物质和生长因子。任选地,培养基可以含有一种或多种选择因子。任选地,培养基可以含有小牛血清或胎牛血清(FCS)。在一个实施方案中,培养基基本上不含IgG。生长培养基通常将例如通过药物选择或必需营养素的缺乏来选择含有DNA构建体的细胞,所述营养素通过DNA构建体上或与DNA构建体共转染的可选标记来补充。所培养的哺乳动物细胞通常在市售的含血清或无血清培养基(例如,MEM、DMEM、DMEM/F12)中生长。在一个实施方案中,培养基是CDoptiCHO(Invitrogen,加利福尼亚州卡尔斯巴德)。在另一个实施方案中,培养基是CD17(Invitrogen,加利福尼亚州卡尔斯巴德)。适合于所用特定细胞系的培养基的选择在本领域普通技术人员的水平内。

IV.药物组合物

[0173] 含有本公开文本的核酸分子、核酸分子编码的多肽、载体或宿主细胞的组合物可

以含有合适的药学上可接受的载体。例如,所述组合物可以含有有助于将活性化合物加工成设计用于递送至作用位点的制剂的赋形剂和/或助剂。

[0174] 在一个实施方案中,本公开文本涉及一种药物组合物,其包含(a)本文所公开的核酸分子、载体、多肽或宿主细胞;和(b)药学上可接受的赋形剂。

[0175] 在一些实施方案中,药物组合物还包含递送剂。在某些实施方案中,递送剂包括脂质纳米颗粒(LNP)。在其他实施方案中,药物组合物还包含脂质体、其他聚合分子和外体。

[0176] 如本文所用,"脂质纳米颗粒"是指包含多个通过分子间力相互物理缔合的脂质分子的纳米颗粒。脂质纳米颗粒可以是例如微球体(包括单层和多层囊泡,例如脂质体)、乳液中的分散相、胶束或悬浮液中的内相。

[0177] 在一些实施方案中,本公开文本提供了一种包封的核酸分子组合物,其可以包括包封本发明的核酸分子的脂质纳米颗粒主体。脂质纳米颗粒可以包含一种或多种脂质(例如,阳离子脂质、非阳离子脂质和PEG修饰的脂质)。在某些实施方案中,本公开文本的脂质纳米颗粒被配制为将本发明的一种或多种核酸分子递送至一种或多种靶细胞。合适的脂质的例子包括但不限于磷脂酰基化合物(例如,磷脂酰基乙醇胺、鞘脂、磷脂酰基胆碱、磷脂酰基丝氨酸、磷脂酰基甘油、神经节苷脂和脑苷脂)。"阳离子脂质"是指在某一pH(例如,生理pH)下携带净正电荷的任何脂质种类。

[0178] 在某些实施方案中,本公开文本的脂质纳米颗粒具有某一N/P比率。如本文所用,"N/P比率"或"NP比率"是指可带正电的聚合物氨基与带负电的核酸磷酸基的比率。脂质纳米颗粒/核酸分子复合物的N/P特征可以影响诸如表面净电荷、稳定性和大小等特性。如本文所述的脂质纳米颗粒的NP比率可以是约1、约2、约3、约4、约5、约6、约7、约8、约9、约10、约15、约20、约25、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100和其间的任何比率。例如,如本文所述的脂质纳米颗粒的NP比率可以是约18、约36或约72。

[0179] 因此,在某些实施方案中,药物组合物包含包封于脂质纳米颗粒中的本公开文本的核酸分子以及药学上可接受的赋形剂。

[0180] 药物组合物可以被配制用于通过推注注射的肠胃外施用(即静脉内、皮下或肌内)。注射用配制品能以单位剂型存在,例如在添加有防腐剂的安瓿中或多剂量容器中。组合物可以采取诸如于油性或水性媒介物中的混悬剂、溶液或乳剂等形式,并且含有诸如助悬剂、稳定剂和/或分散剂等配制剂。可替代地,活性成分可以呈粉末形式,用于用合适的媒介物(例如,无热原水)来构造。

[0181] 适合于肠胃外施用的配制品还包括呈水溶性形式(例如,水溶性盐)的活性化合物的水溶液。另外,可以施用作为适当油性注射混悬剂的活性化合物的混悬剂。合适的亲脂溶剂或媒介物包括脂肪油(例如,芝麻油)或合成脂肪酸酯(例如,油酸乙酯或甘油三酯)。水性注射混悬剂可以含有增加混悬剂粘度的物质,包括例如羧甲基纤维素钠、山梨醇和葡聚糖。任选地,混悬剂还可以含有稳定剂。脂质体还可以用于包封本公开文本的分子,用于递送至细胞或间质空间中。示例性药学上可接受的载体是生理上相容的溶剂、分散介质、包衣、抗细菌和抗真菌剂、等渗和吸收延迟剂、水、盐水、磷酸盐缓冲盐水、右旋糖、甘油、乙醇等。在一些实施方案中,组合物包含等渗剂,例如糖、多元醇(如甘露醇、山梨醇)或氯化钠。在其他实施方案中,组合物包含增强活性成分的贮存期限或有效性的药学上可接受的物质(如润

湿剂)或少量辅助物质(如润湿剂或乳化剂、防腐剂或缓冲液)。

[0182] 本公开文本的组合物可以呈多种形式,包括例如液体(例如,可注射和可输注的溶液)、分散液、混悬剂、半固体和固体剂型。优选形式取决于施用方式和治疗应用。

[0183] 组合物可以配制为溶液、微乳液、分散液、脂质体或其他适合于高药物浓度的有序结构。无菌可注射溶液可以通过以下方式制备:将活性成分以所需量并入视需要具有上文所列举成分中的一种或组合的适当溶剂中,之后过滤灭菌。通常,分散液是通过以下方式来制备:将活性成分并入无菌媒介物中,所述无菌媒介物含有基础分散介质和来自上文所列举的那些成分的所需其他成分。在用于制备无菌可注射溶液的无菌粉末的情况下,优选制备方法是真空干燥和冷冻干燥,这些方法从先前无菌过滤的溶液产生活性成分和任何其他所需成分的粉末。可以例如通过以下方式维持溶液的适当流动性:通过使用诸如卵磷脂等包衣,在分散液的情况下通过维持所需粒度,以及通过使用表面活性剂。可注射组合物的延长吸收可以通过以下方式来实现:在组合物中包括延迟吸收的药剂,例如单硬脂酸盐和明胶。

[0184] 活性成分可以用受控释放配制品或装置来配制。此类配制品和装置的例子包括植入物、经皮贴剂和微囊化递送系统。可以使用生物可降解的生物相容性聚合物,例如乙烯乙酸乙烯酯、聚酐、聚乙醇酸、胶原蛋白、聚原酸酯和聚乳酸。制备此类配制品和装置的方法是本领域已知的。参见例如,Sustained and Controlled Release Drug Delivery Systems,J.R.Robinson编,Marcel Dekker,Inc.,New York,1978。

[0185] 可注射长效配制品可以通过以下方式来制备:形成药物于生物可降解聚合物(如聚丙交酯-聚乙交酯)中的微囊化基质。根据药物与聚合物的比率以及所采用聚合物的性质,可以控制药物释放速率。其他示例性生物可降解聚合物是聚原酸酯和聚酐。可注射长效配制品还可以通过将药物捕集于脂质体或微乳液中来制备。

[0186] 可以将补充性活性化合物并入组合物中。在一个实施方案中,将本公开文本的核酸分子与凝血因子或其变体、片段、类似物或衍生物一起配制。例如,凝血因子包括但不限于因子V、因子VII、因子VIII、因子IX、因子X、因子XI、因子XII、因子XIII、凝血酶原、纤维蛋白原、血管性血友病因子或重组可溶性组织因子(rsTF)或前述任一种的激活形式。凝血因子或止血剂还可以包括抗纤维蛋白溶解药,例如ε-氨基己酸、氨甲环酸。

[0187] 可以调节剂量方案以提供最佳所需反应。例如,可以施用单次推注,可以随时间施用若干次分开剂量,或者可以如治疗情况的紧迫性所示的成比例地减少或增加剂量。有利的是以剂量单位形式配制肠胃外组合物,以便于施用和剂量均匀。参见例如,Remington's Pharmaceutical Sciences(Mack Pub.Co.,Easton,Pa.1980)。

[0188] 除了活性化合物以外,液体剂型还可以含有惰性成分,如水、乙醇、碳酸乙酯、乙酸乙酯、苯甲醇、苯甲酸苄酯、丙二醇、1,3-丁二醇、二甲基甲酰胺、油、甘油、四氢糠醇、聚乙二醇和山梨聚糖的脂肪酸酯。

[0189] 合适的药物载体的非限制性例子也描述于E.W.Martin的Remington'sPharmaceutical Sciences中。赋形剂的一些例子包括淀粉、葡萄糖、乳糖、蔗糖、明胶、麦芽、水稻、面粉、白垩、硅胶、硬脂酸钠、单硬脂酸甘油酯、滑石、氯化钠、脱脂乳粉、甘油、丙烯、二醇、水、乙醇等。组合物还可以含有pH缓冲试剂以及润湿剂或乳化剂。

[0190] 对于口服施用,药物组合物可以采取通过常规手段制备的片剂或胶囊的形式。还

可以将组合物制备为液体,例如糖浆或混悬剂。液体可以包括助悬剂(例如,山梨醇糖浆、纤维素衍生物或氢化食用脂肪)、乳化剂(卵磷脂或阿拉伯胶)、非水性媒介物(例如,扁桃仁油、油性酯、乙醇或分馏植物油)和防腐剂(例如,对羟基苯甲酸甲酯或对羟基苯甲酸丙酯或山梨酸)。制剂还可以包括调味剂、着色剂和甜味剂。可替代地,组合物可以作为干产品存在,用于用水或另一种合适的媒介物来构造。

[0191] 对于经颊施用,组合物可以采取根据常规方案的片剂或锭剂的形式。

[0192] 对于通过吸入施用,根据本公开文本使用的化合物便捷地以含有或不含赋形剂的雾化气溶胶的形式或以气溶胶喷雾剂的形式从任选地具有推进剂的加压包或雾化器递送,所述推进剂是例如二氯二氟甲烷、三氯氟甲烷、二氯四氟甲烷、二氧化碳或其他合适的气体。在加压气溶胶的情况下,剂量单位可以通过提供递送计量量的阀来确定。例如用于吸入器或吹入器中的明胶的胶囊和药筒可以配制含有化合物与诸如乳糖或淀粉等合适的粉末基质的粉末混合物。

[0193] 药物组合物还可以被配制用于作为例如含有常规栓剂基质(如可可脂或其他甘油酯)的栓剂或保留灌肠剂经直肠施用。

[0194] 在一些实施方案中,组合物是通过选自以下的途径来施用:局部施用、眼内施用、肠胃外施用、鞘内施用、硬膜下施用和口服施用。肠胃外施用可以是静脉内或皮下施用。

V.治疗方法

[0195] 在一些方面中,本公开文本涉及治疗有需要的受试者的疾病或病症的方法,其包括施用本文所公开的核酸分子、载体、多肽或药物组合物。

[0196] 在一些实施方案中,本公开文本涉及治疗出血障碍的方法。在一些实施方案中,本公开文本涉及治疗血友病A的方法。

[0197] 分离的核酸分子、载体或多肽可以用以下方式施用:静脉内、皮下、肌内或通过任何粘膜表面,例如口服、舌下、经颊、舌下、经鼻、经直肠、经阴道或通过经肺途径。可以将凝血因子蛋白植入生物聚合物固相支持物内或与所述固相支持物连接,从而允许将嵌合蛋白缓慢释放至所需部位。

[0198] 对于口服施用,药物组合物可以采取通过常规手段制备的片剂或胶囊的形式。还可以将组合物制备为液体,例如糖浆或混悬剂。液体可以包括助悬剂(例如,山梨醇糖浆、纤维素衍生物或氢化食用脂肪)、乳化剂(卵磷脂或阿拉伯胶)、非水性媒介物(例如,扁桃仁油、油性酯、乙醇或分馏植物油)和防腐剂(例如,对羟基苯甲酸甲酯或对羟基苯甲酸丙酯或山梨酸)。制剂还可以包括调味剂、着色剂和甜味剂。可替代地,组合物可以作为干产品存在,用于用水或另一种合适的媒介物来构造。

[0199] 对于经颊和舌下施用,组合物可以采取根据常规方案的片剂、锭剂或速溶薄膜的形式。

[0200] 对于通过吸入施用,根据本公开文本使用的具有凝血因子活性的多肽便捷地以气溶胶喷雾剂的形式从具有合适的推进剂(例如,二氯二氟甲烷、三氯氟甲烷、二氯四氟甲烷、二氧化碳或其他合适的气体)的加压包或雾化器(例如,于PBS中)递送。在加压气溶胶的情况下,剂量单位可以通过提供递送计量量的阀来确定。例如用于吸入器或吹入器中的明胶的胶囊和药筒可以配制含有化合物与诸如乳糖或淀粉等合适的粉末基质的粉末混合物。

[0201] 在一个实施方案中,分离的核酸分子、载体或多肽的施用途径是肠胃外的。如本文

所用的术语肠胃外包括静脉内、动脉内、腹膜内、肌内、皮下、直肠或阴道施用。肠胃外施用的静脉内形式是优选的。虽然所有这些施用形式都明确地考虑在本公开文本的范围内，但是施用形式可以是注射用溶液，特别是用于静脉内或动脉内注射或滴注。通常，用于注射的合适的药物组合物可以包含缓冲液（例如乙酸盐、磷酸盐或柠檬酸盐缓冲液）、表面活性剂（例如聚山梨醇酯）、任选地稳定剂（例如人白蛋白）等。然而，在与本文的传授内容相容的其他方法中，可以将分离的核酸分子、载体或多肽直接递送至不良细胞群的部位，由此增加患病组织对治疗剂的暴露。

[0202] 用于肠胃外施用的制剂包括无菌的水性或非水性溶液、混悬剂和乳剂。非水性溶剂的例子是丙二醇、聚乙二醇、植物油如橄榄油、以及可注射的有机酯如油酸乙酯。水性载体包括水、醇性/水性溶液、乳液或悬浮液，包括盐水和缓冲介质。在本公开文本中，药学上可接受的载体包括但不限于0.01-0.1M且优选地0.05M磷酸盐缓冲液或0.8％盐水。其他常见的肠胃外媒介物包括磷酸钠溶液、林格氏右旋糖、右旋糖和氯化钠、乳酸林格氏液或固定油。静脉内媒介物包括流体和营养补充剂、电解质补充剂（如基于林格氏右旋糖的那些）等。也可以存在防腐剂和其他添加剂，例如抗微生物剂、抗氧化剂、螯合剂和惰性气体等。

[0203] 更具体地，适合于注射使用的药物组合物包括无菌水溶液（水溶性的）或分散液，以及用于临时制备无菌可注射溶液或分散液的无菌粉末。在此类情况下，组合物必须是无菌的并且应当是易于注射的程度的流体。它在制造和储存条件下应当稳定，并且将优选地抵抗微生物（如细菌和真菌）的污染作用而保存。载体可以是溶剂或分散介质，其含有例如水、乙醇、多元醇（例如，甘油、丙二醇和液体聚乙二醇等）及其合适的混合物。可以例如通过以下方式维持适当流动性：通过使用包衣（如卵磷脂），在分散液情况下通过维持所需粒度，以及通过使用表面活性剂。

[0204] 药物组合物还可以被配制用于作为例如含有常规栓剂基质（如可可脂或其他甘油酯）的栓剂或保留灌肠剂经直肠施用。

[0205] 用于治疗病症的本公开文本的组合物的有效剂量根据许多不同因素而变化，所述因素包括施用方式、靶位点、患者的生理状态、患者是人还是动物、施用的其他药物以及治疗是预防性的还是治疗性的。通常，患者是人，但是也可以治疗非人哺乳动物，包括转基因哺乳动物。治疗剂量可以使用本领域技术人员已知的常规方法逐步调整，以优化安全性和功效。

[0206] 本公开文本的核酸分子、载体或多肽可以任选地与在需要治疗（例如，预防性或治疗性）的障碍或病症的治疗中有效的其他药剂组合施用。

[0207] 如本文所用，与辅助疗法结合或组合的本公开文本的分离的核酸分子、载体或多肽的施用意指依序、同时、同延、并行、伴随或同期施用或应用所述疗法和所公开的多肽。本领域技术人员将理解，可以定时施用或施加组合的治疗性方案的各种组分以增强治疗的总体有效性。基于所选辅助疗法和本说明书的传授内容，技术人员（例如，医师）在不进行过度实验的情况下便能够容易地辨别有效的组合治疗方案。

[0208] 应进一步了解，本公开文本的分离的核酸分子、载体或多肽可以与一种或多种药剂结合或组合使用（例如，以提供组合治疗方案）。可以与本公开文本的多肽或多核苷酸组合的示例性药剂包括代表用于所治疗的特定障碍的当前护理标准的药剂。此类药剂在本质上可以是化学的或生物的。术语“生物剂（biologic）”或“生物剂（biologic agent）”是指预

期用作治疗剂的从活的生物体和/或其产物制备的任何药学活性剂。

[0209] 要与本公开文本的多核苷酸或多肽组合使用的药剂的量可以随受试者变化,或者可以根据本领域的知识来施用。参见例如,Bruce A Chabner等人,Antineoplastic Agents,于Goodman&Gilman's The Pharmacological Basis of Therapeutics 1233-1287(Joel G.Hardman等人编,第9版1996)。在另一个实施方案中,施用符合护理标准的量的这种药剂。

[0210] 在一个实施方案中,本文还公开了一种试剂盒,所述试剂盒包含本文公开的核酸分子以及向有需要的受试者施用所述核酸分子的说明书。在另一个实施方案中,本文公开了一种用于产生本文所提供的核酸分子的杆状病毒系统。核酸分子是在昆虫细胞中产生的。在另一个实施方案中,提供了一种针对表达构建体的纳米颗粒递送系统。表达构建体包含本文公开的核酸分子。

VI.基因疗法

[0211] 本公开文本的某些方面提供了一种在受试者体内表达基因表达构建体的方法,其包括向有需要的受试者施用本公开文本的分离的核酸分子。在一些方面中,本公开文本提供了一种增加受试者体内多肽的表达的方法,所述方法包括向有需要的受试者施用本公开文本的分离的核酸分子。

[0212] 已经探索出体基因疗法作为多种病症的可能的治疗,所述病症包括但不限于血友病A。基因疗法是血友病的非常有吸引力的治疗,因为其可能在载体的单次施用后通过连续内源性产生凝血因子(例如,FVIII)治愈所述疾病。血友病A非常适合于基因替代方法,因为其临床表现可完全归因于缺少以微小量(200ng/ml)在血浆中循环的单一基因产物(例如,FVIII)。

[0213] 本文所述的所有各个方面、实施方案和选择都能以任何和所有变化组合。

[0214] 本说明书中所提到的所有出版物、专利和专利申请都通过引用并入本文,并入程度如同指示每个单独出版物、专利或专利申请明确且单独地通过引用并入一般。

实施例

[0215] 在提供了上述公开文本之后,可以通过参考本文提供的实施例获得进一步的理解。这些实施例仅用于说明的目的,而不旨在是限制性的。

实施例1:经修饰的GPV和B19 ITR的设计和构建

[0216] 分析了AAV2(Gene Bank登录号NC_001401.2)、依赖病毒属GPV(Gene Bank登录号U25749.1)和红病毒属B19(Gene Bank登录号KY940273.1)的ITR序列。基于此分析,设计了野生型GPV和B19 ITR的经修饰的衍生物,以研究需要GPV和B19 ITR的核酸序列的哪一种以用于用带有经修饰的ITR的基因构建体持续转导真核细胞。表2中提供了示例性经修饰的ITR的核酸序列。图4-图9中示出了每个经修饰的ITR的预测DNA结构。

[0217] SEQ ID NO:1至8的经修饰的ITR序列是野生型B19 ITR的截短型衍生物。图1A和图2A中示出了截短型B19 ITR衍生物的预测结构的图形描绘。SEQ ID NO:9至16的经修饰的ITR序列是野生型GPV ITR的截短型衍生物。图1B和图2B中示出了截短型GBV ITR衍生物的预测结构的图形描绘。这些截短型ITR被设计为维持野生型ITR的发夹结构,并且被设计为在序列内保留一个或多个REP结合元件(RBE)。在这些序列中去除的碱基是位于RBE与二分体(dyad)之间的核苷酸连同所述核苷酸在发夹另一侧的相应的结合配偶体。SEQ ID NO:1

至16的经修饰的ITR序列具有一定范围的发夹长度和热力学稳定性,预测可以促进体内稳定性和功效以及改善的可制造性和产品稳定性。

[0218] SEQ ID NO:1和2(B19_min)的经修饰的ITR序列还含有B19 ITR序列的截短。图2A中示出了衍生物B19_min ITR的预测结构的图形描绘。

[0219] SEQ ID NO:15和16(GPV_min)的经修饰的ITR序列还含有GPV ITR序列的截短,所述截短去除RBE之间但是对RBE无贡献的核苷酸。图2B中示出了衍生物GPV_min ITR的预测结构的图形描绘。

[0220] SEQ ID NO:17至22的经修饰的ITR序列含有在二分体处插入了另外的核苷酸(以改善从头合成和随后的克隆工作量)的截短型ITR序列。

[0221] 假设由于ITR内RBE的保留存在,带有这些经修饰的衍生物ITR的基因表达构建体将保持功能性并且有效地转导真核细胞。

实施例2.具有经修饰的ITR的FVIII表达构建体的产生。

[0222] 使用编码FVIII基因盒的质粒创建基因构建体,所述基因盒包含具有mTTR启动子和合成内含子的经密码子优化的FVIII。

[0223] 将产生新的构建体,所述构建体包含在5'或3'ITR位置的任一处或二者处侧接SEQ ID NO:1至22所示的经修饰的ITR的经密码子优化的FVIII基因盒。表1中示出了示例性构建体和经修饰的ITR对。

[0224] 将这些新构建体在含有缺失的sbcC基因的大肠杆菌PMC103株中生长和维持,所述sbcC基因编码识别和消除十字形DNA结构的核酸外切酶。在优化细菌培养物的温度和生长条件后,PMC103大肠杆菌能够支持具有经修饰的ITR的经密码子优化的FVIII构建体的生长。将所有新构建体在氨苄青霉素抗性板上进行选择,并且通过限制性酶作图(mapping)进行筛选以确认正确的基因结构。在限制性消化和凝胶提取后,还对经修饰的ITR进行Sanger测序。

表1.示例性FVIII构建体和ITR对。

FVIII 构建体	ITR的描述	5' ITR (SEQ ID NO.)	3' ITR (SEQ ID NO.)
构建体-1	B19_min	1	2
构建体-2	B19_Δ223	3	4
构建体-3	B19_Δ151	5	6
构建体-4	B19_Δ135	7	8
构建体-4	GPV_Δ186	9	10
构建体-5	GPV_Δ162	11	12
构建体-6	GPV_Δ120	13	14
构建体-7	GPV_min	15	16
构建体-8	B19_Δ151_延伸_环	17	18
构建体-9	GPV_Δ186_延伸_环	19	20
构建体-10	GPV_Δ120_延伸_环	21	22

实施例3.含有侧接经修饰的ITR的FVIII表达盒的单链DNA片段的制备。

[0225] 假设侧接经密码子优化的FVIII表达盒的ITR区内发夹结构的形成会驱动靶细胞的持久转导。对于每种质粒构建体,具有形成的发夹ITR结构的单链(ss)DNA片段是通过以下方式来产生:在95℃下使PvuII或LguI消化的双链DNA片段产物(FVIII表达盒和质粒骨

架)变性,并且然后在4℃下冷却以允许回文ITR序列折叠。

实施例4.来自B19最小型ITR的ssDNA和dsDNA构建体的FVIII表达的体内评价

[0226] 为了验证带有经修饰的ITR区的表达构建体在体内介导持久转基因表达的能力,将pFVIII.B19_min构建体(其含有侧接B19最小型ITR(5'ITR:SEQ ID NO:1和3'ITR:SEQ ID NO:2)的经密码子优化的FVIII转基因)的34.7ug DNA经由流体动力学注射(HDI)全身性递送至8-12周龄雄性血友病(HemA)小鼠中。以不具有预先形成的发夹的双链(dsDNA)形式,或以其中预先形成ITR发夹的单链DNA(ssDNA)(如实施例3所制备)施用DNA。在注射后3天和7天从实验动物收集血浆样品。通过生色FVIII活性测定分析血液中的FVIII血浆活性。

[0227] 如图3所示,注射后3天和7天血浆中FVIII水平的分析揭示,与dsDNA相比,ssDNA产生更高水平的FVIII。在第7天,来自ssDNA的FVIII水平保持稳定,但是来自dsDNA的FVIII水平降低了约30%。在第14天,ssDNA和dsDNA二者的FVIII水平都接近零。第14天的样品的ELISA揭示了在所有小鼠中存在抗FVIII抗体(抑制剂)。不受理论的束缚,第14天的接近零的FVIII水平可能是由于抗FVIII抗体的存在,所述抗体可能由于FVIII表达的超生理学水平而形成。此实施例证明,含有经修饰的ITR(诸如B19 ITR)的ssDNA和dsDNA构建体可以表达FVIII。

实施例5:来自ssDNA构建体的体内基因表达的评价

[0228] 为了验证包含侧接经修饰的ITR的经密码子优化的FVIII表达盒的ssDNA构建体的功能,将在hFVIIIR593C$^{+/+}$/HemA小鼠中以与ssFVIII.B19_min构建体类似的方式(参见实施例4)测试每个ssDNA构建体。表2中SEQ ID NO:1至22列出了示例性ITR序列。将产生和测试GPV与B19野生型ITR和经修饰的ITR的组合。

[0229] 将如实施例3中产生来自具有侧接的经修饰的ITR的经密码子优化的FVIII表达构建体的单链DNA,并在hFVIIIR593C$^{+/+}$/HemA小鼠(5-12周龄)中测试在经密码子优化的FVIII表达盒中由mTTR启动子驱动的肝定向FVIII表达。经由流体动力学注射来对人FVIIIR593C$^{+/+}$/HemA小鼠注射10、20、37、50μg或其他预定量的含有前述表达盒的ssDNA,并且从在注射后第1、3、7、14、21和28天或其他预定的时间间隔收集的鼠类血浆测量FVIII。将在施用这些具有经修饰的ITR的表达盒的小鼠中的FVIII表达和寿命与施用具有B19Δ135ITR、GPVΔ162ITR和/或相应的野生型ITR表达盒的ssDNA构建体的小鼠中的FVIII表达和寿命进行直接比较。将通过生色FVIII活性测定分析血液中的FVIII活性。

实施例6.使用杆状病毒表达系统产生带有经修饰的ITR的ceDNA表达构建体

[0230] 已经证实封闭端DNA(ceDNA)表达盒的全身性递送可建立肝细胞的持久转导,并且驱动在肝脏中的稳定长期转基因表达。为了测试转导效率和基因表达,WO 2020/033863的实施例1D中描述的杆状病毒表达系统可用于在昆虫细胞中产生呈封闭端DNA(ceDNA)分子的形式的带有经修饰的ITR的基因构建体的FVIII表达。此杆状病毒表达系统基于Li等人,PLoS ONE 8(8):e69879(2013)中描述的系统。

实施例7.在HemA小鼠中带有经修饰的ITR的ceDNA构建体的体内基因表达的评价

[0231] 将在hFVIIIR593C$^{+/+}$/HemA小鼠中以与pFVIII.B19_min构建体类似的方式(参见实施例4)测试含有侧接经修饰的5'ITR和经修饰的3'ITR的优化的FVIII表达盒的ceDNA构建体。在表2的SEQ ID NO:1至22中列出了示例性ITR序列。如实施例6中所述产生ceDNA构建体。

[0232] 在hFVIIIR593C$^{+/+}$/HemA小鼠(5-12周龄)中测试来自具有侧接的经修饰的ITR的经密码子优化的FVIII表达构建体的ceDNA的肝指向FVIII表达。经由流体动力学注射来对人FVIIIR593C$^{+/+}$/HemA小鼠注射10、20、37、50μg或其他预定量的含有前述表达盒的ceDNA,并且从在注射后第1、3、7、14、21和28天或其他预定的时间间隔收集的鼠类血浆测量FVIII。将与施用具有B19Δ135ITR、GPVΔ162ITR和/或相应的野生型ITR表达盒的ceDNA构建体的小鼠的FVIII表达和寿命直接比较在施用这些具有经修饰的ITR的表达盒的小鼠中的FVIII表达和寿命。将通过生色FVIII活性测定分析血液中的FVIII活性。

实施例8.带有B19或GPV来源的ITR的报告基因构建体的产生。

[0233] 为了证实基于经修饰的ITR的基因表达系统作为平台在基因疗法应用中的一般使用的实用性,用侧接经修饰的5'ITR和经修饰的3'ITR的绿色荧光蛋白(GFP)或萤光素酶(luc)产生包含表达盒的报告基因构建体。使用与实施例2相同的程序。然而,通过常规分子克隆技术用GFP或luc的ORF替代在经密码子优化的FVIII表达盒中FVIII的开放阅读框(ORF)。

[0234] 还产生含有鼠类苯丙氨酸羟化酶(PAH)转基因的侧接经修饰的ITR的表达盒,使用所述表达盒在苯丙酮尿症的相关小鼠模型中评价PAH表达和血液苯丙氨酸浓度的降低。使用此模型,将PKU小鼠经由流体动力学注射施用200μg的侧接经修饰的ITR的ssDNA用于肝表达。将在第3、7、14、28、42、56、70和81天收集血液样品,并且将分离血浆用于苯丙氨酸浓度测定。为了确认鼠类PAH蛋白在肝脏中的存在,对在注射后第81天取自所处理小鼠的肝脏裂解物进行蛋白质印迹。

实施例9.带有经修饰ITR的ssDNA报告基因构建体的制备。

[0235] 如实施例3中所述制备ssDNA报告基因或PAH构建体。简单来说,用LguI、MscI和Eco53kI限制性酶消化质粒。具有形成的发夹ITR结构的ssDNA片段将通过以下方式来产生:使限制性酶消化的双链DNA片段产物(报告基因表达盒和质粒骨架)在95℃下变性,并且然后在4℃下冷却以允许回文ITR序列折叠。可以在小鼠中通过检测报告基因或PAH来测试所得ssDNA构建体建立肝脏、肌肉组织、眼睛中的光感受器和中枢神经系统(CNS)或其他组织的持久转导的能力。

实施例10.ssDNA介导的报告基因表达的体内评价。

[0236] 为了验证实施例13中所述的ssDNA报告基因构建体在体内介导持久转基因表达的能力,对5-12周龄小鼠(至少4只动物/组)用5、10或20μg/小鼠的报告基因ssDNA全身性注射和/或局部注射至靶组织。在预定的时间点收集血液样品,并且可以使用常规技术检测和/或测量报告转基因的水平。

实施例11:具有工程化细小病毒ITR的经修饰的V2.0 FVIIIXTEN表达盒

[0237] 假设可以通过针对靶向宿主对cDNA进行密码子优化来提高转基因表达水平。来自V1.0 FVIIIco6XTEN表达盒的生理水平的FVIII表达已在如美国公开号20190185543中所述的先前研究中得到证明。然而,为了进一步改善转基因表达水平并降低免疫原性,对FVIIIXTEN cDNA进行密码子优化并缺失CpG重复序列,以逃避针对编码具有细小病毒ITR的FVIIIXTEN表达盒的DNA载体产生的固有免疫应答。产生了经修饰的V2.0FVIIIXTEN表达盒,其包含在肝脏特异性修饰小鼠甲状腺素转运蛋白(mTTR)启动子(mTTR482)以及增强子元件(A1MB2)调节下的与XTEN 144肽(FVIIIXTEN)融合的B结构域缺失(BDD)的经密码子优化的

人因子VIII(BDDcoFVIII)、杂合的合成内含子(嵌合内含子)、土拨鼠转录后调节元件(WPRE)和牛生长激素多腺苷酸化(bGHpA)信号。V2.0 FVIIIXTEN表达盒包含SEQ ID NO:27的核苷酸序列。图10中示出了示例性V2.0 FVIIIXTEN表达盒的图形描绘。

[0238] 初始的体内功效研究显示,与V1.0 FVIIIXTEN表达盒相比FVIII活性的显著改善(数据未示出)。因此,克隆了包括AAV2 WT(图10A)、HBoV1 WT(SEQ ID NO:25、SEQ ID NO:26)(图10B)、B19 WT(SEQ ID NO:23、SEQ ID NO:24)、B19最小型(SEQ ID:1、SEQ ID:2)(图10C)和GPVΔ186(SEQ ID:9、SEQ ID:10)、GPVΔ120(SEQ ID NO:13、SEQ ID NO:14)或GPV最小型(SEQ ID NO:15、SEQ ID NO:16)ITR(图10D)的工程化细小病毒ITR,使得ITR侧接V2.0 FVIIIXTEN表达盒。还用不同的细小病毒ITR,以单链(ss)或封闭端(ce)DNA形式,通过在hFVIIIR593C$^{+/+}$/HemA小鼠中经由流体动力学尾静脉注射的全身性施用,证实了经修饰的V2.0 FVIIIXTEN表达盒的体内功能,如下所示。

实施例12:单链FVIIIXTEN(ssFVIIIXTEN)DNA的体内评价

[0239] 假设在ITR区域内形成的发夹驱动了转基因的更高水平的长期持久表达。为了验证经修饰的FVIIIXTEN表达盒的体内功能,在hFVIIIR593C$^{+/+}$/HemA小鼠中测试了包含具有工程化细小病毒ITR的V2.0 FVIIIXTEN基因盒的单链DNA(ssNA)。这些小鼠含有人FVIII-R593C转基因,所述转基因被设计为具有鼠类白蛋白(Alb)启动子,所述启动子驱动含有在轻度血友病A患者中经常观察到的突变的改变的人凝结因子VIII(FVIII)cDNA的表达。这些小鼠还携带FVIII基因的敲除并且缺乏内源性FVIII蛋白质。这些双突变小鼠耐受人FVIII注射,并且没有FVIII活性。它们在用人FVIII处理后产生非常少的抑制性抗体并且缺乏FVIII应答性T细胞或B细胞。hFVIIIR593C$^{+/+}$/HemA小鼠进一步描述于Bril,等人(2006)Thromb.Haemost.95(2):341-7中。

[0240] 通过用识别ITR相关序列并且产生平端DNA的限制性酶消化质粒DNA构建体产生具有不同细小病毒ITR的ssFVIIIXTEN。经消化的双链DNA产物(FVIII表达盒和质粒骨架)在95℃下热变性(变性),然后在4℃下冷却(复性)以允许回文ITR序列形成发夹(图11)。然后在hFVIIIR593C$^{+/+}$/HemA小鼠中以200、800或1600μg/kg经由流体动力学尾静脉注射全身性施用所得ssFVIIIXTEN(ssDNA)。以指定的间隔持续5.5个月从注射的小鼠收集血浆样品,并且根据制造商的说明通过ChromogenixCoatest® SP因子VIII生色测定测量FVIII活性。

[0241] 对于ssFVIIIXTEN注射的动物,归一化为占正常值百分比的血浆FVIII活性示于图12中。结果示出了尽管在不同水平下,但在所有测试的细小病毒ITR中的长期持久的FVIIIXTEN表达。与其他细小病毒ITR相比,测试的所有GPV ITR变体或杂合GPV ITR均示出了FVIIIXTEN表达水平持续下降。相比之下,HBoV1和B19 ITR示出了FVIIIXTEN初始下降直至第56天并且然后至第168天稳定,这表明体内FVIIIXTEN转基因的ITR依赖性持久性。不像GPV ITR,B19和HBoV1 ITR二者都显示出显著更高的FVIII表达水平,而不管测试的变体如何。在测试的不同细小病毒ITR中,HBoV1 ITR在hFVIIIR593C$^{+/+}$/HemA小鼠中显示出占正常FVIII活性的显著更高水平(>1000%)。(图10B)。这些结果验证了具有不同细小病毒ITR的经修饰的FVIIIXTEN表达盒的功能,并且证明了ITR依赖性稳定性以及体内转基因表达的持久性。

实施例13:封闭端FVIIIXTEN(ceFVIIIXTEN)DNA的体内评价

[0242] 虽然ssFVIIIXTEN(ssDNA)在体内表达经修饰的FVIIIXTEN表达盒方面是有效的,

仍存在与有待用作非病毒基因疗法载体的ssDNA相关的若干限制。其中一个是由于用于产生质粒DNA的原核宿主(大肠杆菌)引起的内毒素污染水平,所述质粒DNA还含有在大肠杆菌中选择和扩增所需的外来序列,如抗生素抗性基因和原核复制起点。为了解决这些挑战和局限,开发了一种基于真核细胞的系统,以产生处于包含具有细小病毒ITR的FVIIIXTEN表达盒的封闭端DNA(ceDNA)形式的DNA治疗性药物物质。ceDNA的遗传构成类似于重组AAV载体DNA,但是构象不同。

[0243] 为了产生这种DNA载体,利用杆状病毒昆虫细胞系统,所述系统广泛用于生物制剂制造并且是FDA批准的用于流感疫苗制造的唯一平台。在杆状病毒系统中采用三种不同的ceDNA产生方法,如美国专利申请号63/069,073中所述。图13A中示出了与起始材料(SM)相比,编码侧接AAV2 WT或HBoV1 WT ITR的经修饰V2.0 FVIIIXTEN的纯化ceDNA的示例性琼脂糖凝胶图像。

[0244] 为了验证如从ceDNA中表达的经修饰的FVIIIXTEN的功能,将纯化的ceFVIIIXTEN以0.3μg、1.0μg或2.0μg/小鼠(分别等同于12μg、40μg和80μg/kg)经由流体动力学尾静脉注射而全身性注射在hFVIIIR593C$^{+/+}$/HemA小鼠中。以指定的间隔收集来自注射的小鼠的血浆样品,并且通过生色测定测量FVIII活性,如上所述。

[0245] 对于ceFVIIIXTEN注射的动物,归一化为占正常值百分比的血浆FVIII活性示于图13B中。结果显示出HemA小鼠中的剂量依赖性反应,其中在测试的最高剂量的AAV2或HBoV1 ceDNA中观察到FVIII表达的超生理水平(>正常值的500%)。然而,在ceFVIIIXTEN AAV2 ITR注射的群组中,观察到FVIII表达水平逐渐下降直至注射后第140天,之后水平稳定。有趣的是,对于ceFVIIIXTEN HBoV1 ITR观察到的FVIII表达水平显示出与用ceFVIIIXTEN AAV2 ITR注射的群组中观察到的类似的趋势。

[0246] 这些体内功效研究验证了ceFVIIIXTEN DNA的功能,并且证明了细小病毒ITR可用于在杆状病毒昆虫细胞系统中产生编码感兴趣转基因的功能性非病毒基因疗法载体的概念。

序列

表2:另外的序列

SEQ ID NO和描述	核苷酸或氨基酸序列
SEQ ID NO: 1：5'_B19_min	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAGCGCGCCGCTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 2：3'_B19_min	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAGCGGCGCGCTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG
SEQ ID NO: 3：5'_B19_Δ223	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTAGCGGCGCGCCGCTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 4：3'_B19_Δ223	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTAGCGGCGCGCCGCTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG
SEQ ID NO: 5：5'_B19_Δ151	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCACGCGGCGCGCCGCGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 6：3'_B19_Δ151	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCACGCGGCGCGCCGCGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG
SEQ ID NO: 7：5'_B19_Δ135	CTCTGGGCCAGCTTGCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAGTGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 8：3'_B19_Δ135	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCGGCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAGAG
SEQ ID NO: 9：5'_GPV_Δ186	CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCGCATGCGAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTG
SEQ ID NO: 10：3'_GPV_Δ186	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCGCATGCGAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAG
SEQ ID NO:	CGGTGACGTGTTTCCGGCTGTTAGGTTGACCACGCGCATGCCGCGCGGTC

11： 5'_GPV_ Δ162	AGCCCAATAGTTAAGCCGGAAACACGTCACCGGAAGTCACATGACCGGA AGTCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTAC GTCACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACT TCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGAC TCAAGGACAAGAGGATATTTTGCGCGCCAGGAAGTG
SEQ ID NO: 12： 3'_GPV_ Δ162	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGG TTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTT CCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCAC GTGCTTCCTGTCACGTGTTTCCGGTCACGTGACTTCCGGTCATGTGACTTC CGGTGACGTGTTTCCGGCTTAACTATTGGGCTGACCGCGCGGCATGCGCG TGGTCAACCTAACAGCCGGAAACACGTCACCG
SEQ ID NO: 13： 5'_GPV_Δ120	CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAA GTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGT AGTTCCGGTCACGTGCTTCCTGTCACGTGTTTCCGGTCGCATGCTCACGT GACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTCACCGG ATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCC CCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCAAGGAC AAGAGGATATTTTGCGCGCCAGGAAGTG
SEQ ID NO: 14： 3'_GPV_Δ120	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGG TTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTT CCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCAC GTGCTTCCTGTCACGTGTTTCCGGTCACGTGAGCATGCGACCGGAAACAC GTGACAGGAAGCACGTGACCGGAACTACGTCACCGGATGTGCGTCACCG GAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGG TTCGAACGAACGAACCCTCCAATGAG
SEQ ID NO: 15： 5'_GPV_最小型	CTCATTGGAGGGTTCGTTCGTTCGAACGTTCGTTCGCATGCGAACGAACG TTCGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATATTTT GCGCGCCAGGAAGTG
SEQ ID NO: 16： 3'_GPV_最小型	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGG TTCGTTCGTTCGAACGTTCGTTCGCATGCGAACGAACGTTCGAACGAACG AACCCTCCAATGAG
SEQ ID NO: 17： 5'_B19_Δ151_ 延伸_环	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACA AGATGGCGGACAATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTC ACGCGGCGCGCCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAA TCCCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTCAACCTACATC CTCAATCCCGATCTATCCGAGGCGCGCCGCGTGACGTCACAGGAAATGA CGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGC CTACCGGCGGCGACCGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 18： 3'_B19_Δ151_ 延伸_环	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTA GGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGA CGTCATTTCCTGTGACGTCACGCGGCGCGCCAGCGCCTATCGCCAGGTAT TACTCCAATCCCGAATATCCGAGATCGGGATCACCCGAGAGAAGTTCAA CCTACATCCTCAATCCCGATCTATCCGAGATCCGGCGCGCCGCGTGACGT CACAGGAAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTA CCGGAAGTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG
SEQ ID NO: 19： 5'_GPV_Δ186_ 延伸_环	CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAA GTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGT AGTTCGCATGCCTGTCTATCGCCTACCCATCCCTGTCTGAGATCAAGGGC GTGATCGTGCACAGACTGGAGAGCGTGTCCTATAATATCGGCTCTCAGG AGTGGAGCACCACAGTGCCCAGATACGTGGCCACCCAGGGCTATCTGAT CTCCAACTTCGACGCATGCGAACTACGTCACCGGATGTGCGTCACCGGA AGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTT CGAACGAACGAACCCTCCAATGAGACTCAAGGACAAGAGGATATTTTGC GCGCCAGGAAGTG

SEQ ID NO: 20： 3'_GPV_Δ186_延伸_环	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGG TTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTT CCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCGCATGCC TGTCTATCGCCTACCCATCCCTGTCTGAGATCAAGGGCGTGATCGTGCAC AGACTGGAGAGCGTGTCCTATAATATCGGCTCTCAGGAGTGGAGCACCA CAGTGCCCAGATACGTGGCCACCCAGGGCTATCTGATCTCCAACTTCGAC GCATGCGAACTACGTCACCGGATGTGCGTCACCGGAAGCATGTGACCGG AACTTGCGTCACTTCCCCCTCCCCTGATTGGCTGGTTCGAACGAACGAAC CCTCCAATGAG
SEQ ID NO: 21： 5'_GPV_Δ120_延伸_环	CTCATTGGAGGGTTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAA GTGACGCAAGTTCCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGT AGTTCCGGTCACGTGCTTCCTGTCACGTGTTTCCGGTCGCATGCCTGTCTA TCGCCTACCCATCCCTGTCTGAGATCAAGGGCGTGATCGTGCACAGACTG GAGAGCGTGTCCTATAATATCGGCTCTCAGGAGTGGAGCACCACAGTGC CCAGATACGTGGCCACCCAGGGCTATCTGATCTCCAACTTCGACGCATGC TCACGTGACCGGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTC ACCGGATGTGCGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCC CCCTCCCCTGATTGGCTGGTTCGAACGAACGAACCCTCCAATGAGACTCA AGGACAAGAGGATATTTTGCGCGCCAGGAAGTG
SEQ ID NO: 22： 3'_GPV_Δ120_延伸_环	CACTTCCTGGCGCGCAAAATATCCTCTTGTCCTTGAGTCTCATTGGAGGG TTCGTTCGTTCGAACCAGCCAATCAGGGGAGGGGGAAGTGACGCAAGTT CCGGTCACATGCTTCCGGTGACGCACATCCGGTGACGTAGTTCCGGTCAC GTGCTTCCTGTCACGTGTTTCCGGTCACGTGAGCATGCCTGTCTATCGCCT ACCCATCCCTGTCTGAGATCAAGGGCGTGATCGTGCACAGACTGGAGAG CGTGTCCTATAATATCGGCTCTCAGGAGTGGAGCACCACAGTGCCCAGA TACGTGGCCACCCAGGGCTATCTGATCTCCAACTTCGACGGCATGCGACC GGAAACACGTGACAGGAAGCACGTGACCGGAACTACGTCACCGGATGTG CGTCACCGGAAGCATGTGACCGGAACTTGCGTCACTTCCCCCTCCCCTGA TTGGCTGGTTCGAACGAACGAACCCTCCAATGAG
SEQ ID NO: 23： B19 WT 5'	CCAAATCAGATGCCGCCGGTCGCCGCCGGTAGGCGGGACTTCCGGTACA AGATGGCGGACAATTACGTCATTTCCTGTGACGTCATTTCCTGTGACGTC ACTTCCGGTGGGCGGGACTTCCGGAATTAGGGTTGGCTCTGGGCCAGCTT GCTTGGGGTTGCCTTGACACTAAGACAAGCGGCGCGCCGCTTGATCTTAG TGGCACGTCAACCCCAAGCGCTGGCCCAGAGCCAACCCTAATTCCGGAA GTCCCGCCCACCGGAAGTGACGTCACAGGAAATGACGTCACAGGAAATG ACGTAATTGTCCGCCATCTTGTACCGGAAGTCCCGCCTACCGGCGGCGAC CGGCGGCATCTGATTTGGTGTCTTCTTTTAAATTTT
SEQ ID NO: 24： B19 WT 3'	AAAATTTAAAAGAAGACACCAAATCAGATGCCGCCGGTCGCCGCCGGTA GGCGGGACTTCCGGTACAAGATGGCGGACAATTACGTCATTTCCTGTGA CGTCATTTCCTGTGACGTCACTTCCGGTGGGCGGGACTTCCGGAATTAGG GTTGGCTCTGGGCCAGCGCTTGGGGTTGACGTGCCACTAAGATCAAGCG GCGCGCCGCTTGTCTTAGTGTCAAGGCAACCCCAAGCAAGCTGGCCCAG AGCCAACCCTAATTCCGGAAGTCCCGCCCACCGGAAGTGACGTCACAGG AAATGACGTCACAGGAAATGACGTAATTGTCCGCCATCTTGTACCGGAA GTCCCGCCTACCGGCGGCGACCGGCGGCATCTGATTTGG
SEQ ID NO: 25 HBoV1 WT ITR 5'	TTGTTGTTGTACATGCGCCATCTTAGTTTTATATCAGCTGGCGCCTTAGTT ATATAACATGCATGTTATATAACTAAGGCGCCAGCTGATATAAAACTAA GATGGCGCATGTACAACAACAACACATTAAAAGATATAGAGTTTCGCGA TTGC
SEQ ID NO: 26 HBoV1 WT ITR 3'	TATATGTGACGTGGTTGTACAGACGCCATCTTGGAATCCAATATGTCTGC CGGCGATTAGATCATGCGCGCGCGCAGCGCGCTGCGCGCAGCGCAGGCA TGACTGAGCCGGCAGACATATTGGATTCCAAGATGGCGTCTGTACAACC AC

SEQ ID NO: 27 V2.0 表达盒 mTTR482-内含子-coBDDFVIIIXTEN (V2.0)-WPRE-bGHPolyA	GGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTG GCAGCATTTACTCTCTCTATTGACTTTGGTTAATAATCTCAGGAGCACAA ACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTG GGCCTCTCCCCACCTTCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCA AAGGTCAAAGTGGCCCTTGGCAGCATTTACTCTCTCTATTGACTTTGGTT AATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAAC ATCCTGGACTTATCCTCTGGGCCTCTCCCCACCGATATCTACCTGCTGATC GCCCGGCCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGCAAAGGTCG GCAGTAGTTTTCCATCTTACTCAACATCCTCCCAGTGTACGTAGGATCCT GTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCG GGGCAAAGGTCGTATTGACTTAGGTTACTTATTCTCCTTTTGTTGACTAA GTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGC AGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAGC CGTCACACAGATCCACAAGCTCCTGCTAGGAATTCTCAGGAGCACAAAC ATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTGGG CCTCTCCCCACCGATATCTACCTGCTGATCGCCCGGCCCCTGTTCAAACA TGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAGTTTTCCATCTTAC TCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTA GAGCGAGTGTTCCGATACTCTAATCTCCCGGGGCAAAGGTCGTATTGACT TAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGCA GGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGG GGTATAAAAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGC TCCTGCTAGAGTCGCTGCGCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCG CCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCGTTACTCCCACAG GTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGT TTATTGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAGCCTTGAGGGGC TCCGGGAAGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGGTGCGTGCGT GTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCCGGCGGC TGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTGC GCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCT GCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAG CAGGGGGTGTGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCC TCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACG GGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGGCGGCAGGTG GGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCCGGGGAGGGCTCGGGG GAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCGAGGCGCGGCG AGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGGCGCAGGGAC TTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGGCGCCGCCGC ACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGGAAG GAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTC TCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGG GGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCT AGAGCCTCTGCTAACCTTGTTCTTGCCTTCTTCTTTTTCCTACAGCTCCTG GGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAAGAATTACT CGAGGCCACCATGCAGATTGAACTGTCCACTTGCTTCTTCCTGTGCCTCC TGCGGTTTTGCTTCTCGGCCACCCGCCGGTATTACTTAGGTGCTGTGGAA CTGAGCTGGGACTACATGCAGTCCGACCTGGGAGAACTGCCGGTGGACG CGAGATTCCCACCTAGAGTCCCGAAGTCCTTCCCATTCAACACCTCCGTG GTCTACAAAAAGACCCTGTTCGTGGAGTTCACTGACCACCTTTTCAATAT TGCCAAGCCGCGCCCCCCCTGGATGGGCCTGCTTGGTCCTACGATCCAAG CAGAGGTCTACGACACCGTGGTCATCACACTGAAGAACATGGCCTCACA CCCCGTGTCGCTGCATGCTGTGGGAGTGTCCTACTGGAAGGCCTCAGAG GGTGCCGAATATGATGACCAGACCAGCCAGAGGGAAAAGGAGGATGAC AAAGTGTTCCCGGGTGGCAGCCACACTTACGTGTGGCAAGTGCTGAAGG AAAACGGGCCTATGGCGTCGGACCCCCTATGCCTGACCTACTCCTACCTG TCCCATGTGGACCTTGTGAAGGATCTCAACTCGGGACTGATCGGCGCCCT CTTGGTGTGCAGAGAAGGCAGCCTGGCGAAGGAAAAGACTCAGACCCTG CACAAGTTCATTCTGTTGTTTGCTGTGTTCGATGAAGGAAAGTCCTGGCA CTCAGAAACCAAGAACTCGCTGATGCAGGATAGAGATGCGGCCTCGGCC AGAGCCTGGCCTAAAATGCACACCGTCAACGGATATGTGAACAGGTCGC

	TCCCTGGCCTCATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGATC GGCATGGGTACTACTCCGGAAGTGCATAGTATCTTTCTGGAGGGCCATAC CTTCTTGGTGCGCAACCACAGACAGGCCTCGCTGGAAATCTCGCCTATCA CTTTCTTGACTGCGCAGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGT TCTGTCACATCAGCTCCCATCAGCATGATGGGATGGAGGCCTATGTCAAA GTGGACTCCTGCCCTGAGGAGCCACAGCTCCGGATGAAGAACAATGAGG AAGCGGAGGATTACGACGACGACCTGACTGACAGCGAAATGGACGTCGT GCGATTCGATGACGACAACAGCCCGTCCTTCATCCAAATTAGATCAGTG GCGAAGAAGCACCCCAAGACCTGGGTGCACTACATTGCCGCCGAGGAAG AGGACTGGGACTACGCGCCGCTGGTGCTGGCGCCAGACGACAGGAGCTA CAAGTCCCAGTACCTCAACAACGGGCCGCAGCGCATTGGCAGGAAGTAC AAGAAAGTCCGCTTCATGGCCTACACTGATGAAACCTTCAAGACGAGGG AAGCCATCCAGCACGAGTCAGGCATCCTGGGACCGCTCCTTTACGGCGA AGTCGGGGATACCCTGCTCATCATTTTCAAGAACCAGGCATCGCGGCCCT ACAACATCTACCCTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGC CGGCTGCCCAAGGGAGTGAAGCACCTGAAGGATTTTCCCATCCTGCCGG GAGAAATCTTCAAGTACAAGTGGACCGTGACTGTGGAAGATGGCCCTAC CAAGTCGGACCCTCGCTGTCTGACCCGGTACTATTCCTCGTTTGTGAACA TGGAGCGCGACCTGGCCTCGGGGCTGATTGGTCCGCTGCTGATCTGCTAC AAGGAGTCCGTGGACCAGCGCGGGAACCAGATCATGTCCGACAAGCGCA ACGTGATCCTGTTCTCTGTCTTTGATGAAAACAGATCGTGGTACTTGACT GAGAATATCCAGCGGTTCCTGCCCAACCCAGCGGGAGTGCAACTGGAGG ACCCGGAGTTCCAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTG TTCGACTCGCTCCAACTGAGCGTGTGCCTGCATGAAGTGGCATACTGGTA CATTCTGTCCATCGGAGCCCAGACCGACTTCCTGTCCGTGTTCTTCTCCG GATACACCTTCAAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTC CCATTTTCCGGAGAAACTGTGTTCATGTCAATGGAAAACCCGGGCTTGTG GATTCTGGGTTGCCATAACTCGGACTTCCGGAATAGAGGGATGACCGCC CTGCTGAAAGTGTCCAGCTGTGACAAGAATACCGGCGATTACTACGAGG ACAGCTATGAGGACATCTCCGCTTATCTGCTGTCCAAGAACAACGCCATT GAACCCAGGTCCTTCTCCCAAAACGGTGCACCGACCTCCGAAAGCGCCA CCCCAGAGTCAGGACCTGGCTCGGAACCGGCTACCTCGGGCTCAGAGAC ACCGGGGACTTCCGAGTCCGCAACCCCCGAGAGTGGACCCGGATCCGAA CCAGCAACCTCAGGATCAGAAACCCCGGGAACTTCGGAATCCGCCACTC CCGAGTCGGGACCAGGCACCTCCACTGAGCCTTCCGAGGGAAGCGCCCC CGGATCCCCTGCTGGATCCCCTACCAGCACTGAAGAAGGCACCTCAGAA TCCGCGACCCCTGAGTCCGGCCCTGGAAGCGAACCCGCCACCTCCGGTTC CGAAACCCCTGGGACTAGCGAGAGCGCCACTCCGGAATCGGGCCCAGGA AGCCCTGCCGGATCCCCGACCAGCACCGAGGAGGGAAGCCCCGCCGGGT CACCGACTTCCACTGAGGAGGGAGCCTCATCCCCCCCCGTGCTGAAGCG GCATCAAAGAGAGATCACCAGGACCACTCTCCAGTCCGATCAGGAAGAA ATTGACTACGACGATACTATCAGCGTGGAGATGAAGAAGGAGGACTTCG ACATCTACGATGAGGATGAGAACCAGTCCCCTCGGAGCTTTCAGAAGAA AACCCGCCACTACTTCATCGCTGCCGTGGAGCGGCTGTGGGATTACGGG ATGTCCAGCTCACCGCATGTGCTGCGGAATAGAGCGCAGTCAGGATCGG TGCCCCAGTTCAAGAAGGTCGTGTTCCAAGAGTTCACCGACGGGTCCTTC ACTCAACCCCTGTACCGGGGCGAACTCAACGAACACCTGGGACTGCTTG GGCCGTATATCAGGGCAGAAGTGGAAGATAACATCATGGTCACCTTCCG CAACCAGGCCTCCCGGCCGTACAGCTTCTACTCTTCACTGATCTCCTACG AGGAAGATCAGCGGCAGGGAGCCGAGCCCCGGAAGAACTTCGTCAAGC CTAACGAAACTAAGACCTACTTTTGGAAGGTCCAGCATCACATGGCCCC GACCAAAGACGAGTTCGACTGTAAAGCCTGGGCCTACTTCTCCGATGTG GACCTGGAGAAGGACGTGCACTCGGGACTCATTGGCCCGCTCCTTGTGT GCCATACTAATACCCTGAACCCTGCTCACGGTCGCCAAGTCACAGTGCA GGAGTTCGCCCTCTTCTTCACCATCTTCGATGAAACAAAGTCCTGGTACT TTACTGAGAACATGGAACGCAATTGCAGGGCACCCTGCAACATCCAGAT GGAAGATCCCACCTTCAAGGAAAACTACCGGTTTCATGCCATTAACGGC TACATAATGGACACGTTGCCAGGACTGGTCATGGCCCAGGACCAGAGAA TCCGGTGGTATCTGCTCTCCATGGGCTCCAACGAAAACATTCACAGCATT CATTTTTCCGGCCATGTGTTCACCGTCCGGAAGAAGGAAGAGTACAAGA

	TGGCTCTGTACAACCTCTACCCTGGAGTGTTCGAGACTGTGGAAATGCTG CCTAGCAAGGCCGGCATTTGGAGAGTGGAATGCCTGATCGGAGAGCATT TGCACGCCGGAATGTCCACCCTGTTTCTTGTGTACTCCAACAAGTGCCAG ACCCCGCTGGGAATGGCCTCAGGTCATATTAGGGATTTCCAGATCACTGC TTCGGGGCAGTACGGGCAGTGGGCACCTAAGTTGGCCCGGCTGCACTAC TCTGGCTCCATCAATGCCTGGTCCACCAAGGAACCCTTCTCCTGGATTAA GGTGGACCTCCTGGCCCCAATGATTATTCACGGTATTAAGACCCAGGGTG CCCGACAGAAGTTCTCCTCACTCTACATCTCGCAATTCATCATAATGTAC AGCCTGGATGGGAAGAAGTGGCAGACCTACCGGGGAAACTCCACTGGA ACGCTCATGGTGTTTTTCGGCAACGTGGACTCCTCCGGCATTAAGCACAA CATCTTCAACCCTCCGATCATTGCTCGGTACATCCGGCTGCACCCAACTC ACTACAGCATCCGGTCCACCCTGCGGATGGAACTGATGGGTTGTGACCT GAACTCCTGCTCCATGCCCCTTGGGATGGAATCCAAGGCCATTAGCGATG CACAGATCACCGCCTCTTCATACTTCACCAACATGTTCGCGACCTGGTCC CCGTCGAAGGCCCGCCTGCACCTCCAAGGTCGCTCCAATGCGTGGCGGC CTCAAGTGAACAACCCCAAGGAGTGGCTCCAGGTCGACTTCCAAAAGAC CATGAAGGTCACCGGAGTGACCACCCAGGGCGTGAAGTCCCTGCTGACC TCTATGTACGTTAAGGAGTTCCTCATCTCCTCAAGCCAAGACGGACATCA GTGGACCCTGTTCTTCCAAAACGGAAAAGTCAAAGTATTCCAGGGCAAC CAGGACTCCTTCACCCCTGTGGTCAACAGCCTGGACCCCCCATTGCTGAC CCGCTACCTCCGCATCCACCCCCAAAGCTGGGTCCACCAGATCGCACTGC GCATGGAGGTCCTTGGATGCGAAGCCCAAGATCTGTACTAAGCGGCCGC TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTC TTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATA AATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAA CGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGG CATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCC TATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATC ATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGG GACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTC CCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCC CTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTGCCTAGGCGA CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTT CCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAG GAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCA TGCTGGGGAAGACCATGGGCGCGCCAGGCCTGTCGACGCCCGGGCGGTA CCGCGATCGCTCGCGACGCATAAAG
SEQ ID NO: 28 编码 coBDDFVIIIXTEN (V2.0)的核苷酸序列	ATGCAGATTGAACTGTCCACTTGCTTCTTCCTGTGCCTCCTGCGGTTTTGC TTCTCGGCCACCCGCCGGTATTACTTAGGTGCTGTGGAACTGAGCTGGGA CTACATGCAGTCCGACCTGGGAGAACTGCCGGTGGACGCGAGATTCCCA CCTAGAGTCCCGAAGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAA GACCCTGTTCGTGGAGTTCACTGACCACCTTTTCAATATTGCCAAGCCGC GCCCCCCCTGGATGGGCCTGCTTGGTCCTACGATCCAAGCAGAGGTCTAC GACACCGTGGTCATCACACTGAAGAACATGGCCTCACACCCCGTGTCGC TGCATGCTGTGGGAGTGTCCTACTGGAAGGCCTCAGAGGGTGCCGAATA TGATGACCAGACCAGCCAGAGGGAAAAGGAGGATGACAAAGTGTTCCC GGGTGGCAGCCACACTTACGTGTGGCAAGTGCTGAAGGAAAACGGGCCT ATGGCGTCGGACCCCCTATGCCTGACCTACTCCTACCTGTCCCATGTGGA CCTTGTGAAGGATCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCA GAGAAGGCAGCCTGGCGAAGGAAAAGACTCAGACCCTGCACAAGTTCAT TCTGTTGTTTGCTGTGTTCGATGAAGGAAAGTCCTGGCACTCAGAAACCA AGAACTCGCTGATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCC TAAAATGCACACCGTCAACGGATATGTGAACAGGTCGCTCCCTGGCCTC ATCGGCTGCCACAGAAAGTCCGTGTATTGGCATGTGATCGGCATGGGTA CTACTCCGGAAGTGCATAGTATCTTTCTGGAGGGCCATACCTTCTTGGTG CGCAACCACAGACAGGCCTCGCTGGAAATCTCGCCTATCACTTTCTTGAC

	TGCGCAGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCACA TCAGCTCCCATCAGCATGATGGGATGGAGGCCTATGTCAAAGTGGACTC CTGCCCTGAGGAGCCACAGCTCCGGATGAAGAACAATGAGGAAGCGGA GGATTACGACGACGACCTGACTGACAGCGAAATGGACGTCGTGCGATTC GATGACGACAACAGCCCGTCCTTCATCCAAATTAGATCAGTGGCGAAGA AGCACCCCAAGACCTGGGTGCACTACATTGCCGCCGAGGAAGAGGACTG GGACTACGCGCCGCTGGTGCTGGCGCCAGACGACAGGAGCTACAAGTCC CAGTACCTCAACAACGGGCCGCAGCGCATTGGCAGGAAGTACAAGAAA GTCCGCTTCATGGCCTACACTGATGAAACCTTCAAGACGAGGGAAGCCA TCCAGCACGAGTCAGGCATCCTGGGACCGCTCCTTTACGGCGAAGTCGG GGATACCCTGCTCATCATTTTCAAGAACCAGGCATCGCGGCCCTACAACA TCTACCCTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTG CCCAAGGGAGTGAAGCACCTGAAGGATTTTCCCATCCTGCCGGGAGAAA TCTTCAAGTACAAGTGGACCGTGACTGTGGAAGATGGCCCTACCAAGTC GGACCCTCGCTGTCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGC GCGACCTGGCCTCGGGGCTGATTGGTCCGCTGCTGATCTGCTACAAGGA GTCCGTGGACCAGCGCGGGAACCAGATCATGTCCGACAAGCGCAACGTG ATCCTGTTCTCTGTCTTTGATGAAAACAGATCGTGGTACTTGACTGAGAA TATCCAGCGGTTCCTGCCCAACCCAGCGGGAGTGCAACTGGAGGACCCG GAGTTCCAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTGTTCGA CTCGCTCCAACTGAGCGTGTGCCTGCATGAAGTGGCATACTGGTACATTC TGTCCATCGGAGCCCAGACCGACTTCCTGTCCGTGTTCTTCTCCGGATAC ACCTTCAAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATT TTCCGGAGAAACTGTGTTCATGTCAATGGAAAACCCGGGCTTGTGGATTC TGGGTTGCCATAACTCGGACTTCCGGAATAGAGGGATGACCGCCCTGCT GAAAGTGTCCAGCTGTGACAAGAATACCGGCGATTACTACGAGGACAGC TATGAGGACATCTCCGCTTATCTGCTGTCCAAGAACAACGCCATTGAACC CAGGTCCTTCTCCCAAAACGGTGCACCGACCTCCGAAAGCGCCACCCCA GAGTCAGGACCTGGCTCGGAACCGGCTACCTCGGGCTCAGAGACACCGG GGACTTCCGAGTCCGCAACCCCCGAGAGTGGACCCGGATCCGAACCAGC AACCTCAGGATCAGAAACCCCGGGAACTTCGGAATCCGCCACTCCCGAG TCGGGACCAGGCACCTCCACTGAGCCTTCCGAGGGAAGCGCCCCCGGAT CCCCTGCTGGATCCCCTACCAGCACTGAAGAAGGCACCTCAGAATCCGC GACCCCTGAGTCCGGCCCTGGAAGCGAACCCGCCACCTCCGGTTCCGAA ACCCCTGGGACTAGCGAGAGCGCCACTCCGGAATCGGGCCCAGGAAGCC CTGCCGGATCCCCGACCAGCACCGAGGAGGGAAGCCCCGCCGGGTCACC GACTTCCACTGAGGAGGGAGCCTCATCCCCCCCCGTGCTGAAGCGGCAT CAAAGAGAGATCACCAGGACCACTCTCCAGTCCGATCAGGAAGAAATTG ACTACGACGATACTATCAGCGTGGAGATGAAGAAGGAGGACTTCGACAT CTACGATGAGGATGAGAACCAGTCCCCTCGGAGCTTTCAGAAGAAAACC CGCCACTACTTCATCGCTGCCGTGGAGCGGCTGTGGGATTACGGGATGTC CAGCTCACCGCATGTGCTGCGGAATAGAGCGCAGTCAGGATCGGTGCCC CAGTTCAAGAAGGTCGTGTTCCAAGAGTTCACCGACGGGTCCTTCACTCA ACCCCTGTACCGGGGCGAACTCAACGAACACCTGGGACTGCTTGGGCCG TATATCAGGGCAGAAGTGGAAGATAACATCATGGTCACCTTCCGCAACC AGGCCTCCCGGCCGTACAGCTTCTACTCTTCACTGATCTCCTACGAGGAA GATCAGCGGCAGGGAGCCGAGCCCCGGAAGAACTTCGTCAAGCCTAACG AAACTAAGACCTACTTTTGGAAGGTCCAGCATCACATGGCCCCGACCAA AGACGAGTTCGACTGTAAAGCCTGGGCCTACTTCTCCGATGTGGACCTGG AGAAGGACGTGCACTCGGGACTCATTGGCCCGCTCCTTGTGTGCCATACT AATACCCTGAACCCTGCTCACGGTCGCCAAGTCACAGTGCAGGAGTTCG CCCTCTTCTTCACCATCTTCGATGAAACAAAGTCCTGGTACTTTACTGAG AACATGGAACGCAATTGCAGGGCACCCTGCAACATCCAGATGGAAGATC CCACCTTCAAGGAAAACTACCGGTTTCATGCCATTAACGGCTACATAATG GACACGTTGCCAGGACTGGTCATGGCCCAGGACCAGAGAATCCGGTGGT ATCTGCTCTCCATGGGCTCCAACGAAAACATTCACAGCATTCATTTTTCC GGCCATGTGTTCACCGTCCGGAAGAAGGAAGAGTACAAGATGGCTCTGT ACAACCTCTACCCTGGAGTGTTCGAGACTGTGGAAATGCTGCCTAGCAA GGCCGGCATTTGGAGAGTGGAATGCCTGATCGGAGAGCATTTGCACGCC GGAATGTCCACCCTGTTTCTTGTGTACTCCAACAAGTGCCAGACCCCGCT

	GGGAATGGCCTCAGGTCATATTAGGGATTTCCAGATCACTGCTTCGGGGC AGTACGGGCAGTGGGCACCTAAGTTGGCCCGGCTGCACTACTCTGGCTC CATCAATGCCTGGTCCACCAAGGAACCCTTCTCCTGGATTAAGGTGGACC TCCTGGCCCCAATGATTATTCACGGTATTAAGACCCAGGGTGCCCGACAG AAGTTCTCCTCACTCTACATCTCGCAATTCATCATAATGTACAGCCTGGA TGGGAAGAAGTGGCAGACCTACCGGGGAAACTCCACTGGAACGCTCATG GTGTTTTTCGGCAACGTGGACTCCTCCGGCATTAAGCACAACATCTTCAA CCCTCCGATCATTGCTCGGTACATCCGGCTGCACCCAACTCACTACAGCA TCCGGTCCACCCTGCGGATGGAACTGATGGGTTGTGACCTGAACTCCTGC TCCATGCCCCTTGGGATGGAATCCAAGGCCATTAGCGATGCACAGATCA CCGCCTCTTCATACTTCACCAACATGTTCGCGACCTGGTCCCCGTCGAAG GCCCGCCTGCACCTCCAAGGTCGCTCCAATGCGTGGCGGCCTCAAGTGA ACAACCCCAAGGAGTGGCTCCAGGTCGACTTCCAAAAGACCATGAAGGT CACCGGAGTGACCACCCAGGGCGTGAAGTCCCTGCTGACCTCTATGTAC GTTAAGGAGTTCCTCATCTCCTCAAGCCAAGACGGACATCAGTGGACCCT GTTCTTCCAAAACGGAAAAGTCAAAGTATTCCAGGGCAACCAGGACTCC TTCACCCCTGTGGTCAACAGCCTGGACCCCCCATTGCTGACCCGCTACCT CCGCATCCACCCCCAAAGCTGGGTCCACCAGATCGCACTGCGCATGGAG GTCCTTGGATGCGAAGCCCAAGATCTGTACTAA
SEQ ID NO: 29: HBoV1-5'ITR-mTTR482-内含子-coBDDFVIIIXTEN (V2.0)-WPRE-bGHPolyA-HBoV1-3'ITR	GTATACCTGCAGGCTAGCCACGTGTTGTTGTTGTACATGCGCCATCTTAG TTTTATATCAGCTGGCGCCTTAGTTATATAACATGCATGTTATATAACTA AGGCGCCAGCTGATATAAAACTAAGATGGCGCATGTACAACAACAACAC ATTAAAAGATATAGAGTTTCGCGATTGCAAGCTTGGCCCCAGGTTAATTT TTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTGGCAGCATTTACTCTCTC TATTGACTTTGGTTAATAATCTCAGGAGCACAAACATTCCTGGAGGCAGG AGAAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTCCCCACCTTCG ATGGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCT TGGCAGCATTTACTCTCTCTATTGACTTTGGTTAATAATCTCAGGAGCAC AAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTC TGGGCCTCTCCCCACCGATATCTACCTGCTGATCGCCCGGCCCCTGTTCA AACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTAGTTTTCCATC TTACTCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTCTGCACATTTC GTAGAGCGAGTGTTCCGATACTCTAATCTCCCGGGGCAAAGGTCGTATTG ACTTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAATCAGAATCA GCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGA GGGGGTATAAAAGCCCCTTCACCAGGAGAAGCCGTCACACAGATCCACA AGCTCCTGCTAGGAATTCTCAGGAGCACAAACATTCCTGGAGGCAGGAG AAGAAATCAACATCCTGGACTTATCCTCTGGGCCTCTCCCCACCGATATC TACCTGCTGATCGCCCGGCCCCTGTTCAAACATGTCCTAATACTCTGTCG GGGCAAAGGTCGGCAGTAGTTTTCCATCTTACTCAACATCCTCCCAGTGT ACGTAGGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTCCGATAC TCTAATCTCCCGGGGCAAAGGTCGTATTGACTTAGGTTACTTATTCTCCTT TTGTTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCAGCTTGGC AGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCCCCTTCAC CAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGCTAGAGTCGCTGCG CGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCC CCGGCTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGC CCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTTATTGACGGCTTGTTTCT TTTCTGTGGCTGCGTGAAAGCCTTGAGGGGCTCCGGGAAGGCCCTTTGTG CGGGGGGAGCGGCTCGGGGGGTGCGTGCGTGTGTGTGTGCGTGGGGAGC GCCGCGTGCGGCTCCGCGCTGCCCGGCGGCTGTGAGCGCTGCGGGCGCG GCGCGGGGCTTTGTGCGCTCCGCAGTGTGCGCGAGGGGAGCGCGGCCGG GGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGAGGGGAACAAAGGCTG CGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTGTGGGCGCGTCG GTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGAGCACGG CCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGCC GTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGG CCGCCTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGG AGCGCCGGCGGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTAT

	GGTAATCGTGCGAGAGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGC GGAGCCGAAATCTGGGAGGCGCCGCCGCACCCCCTCTAGCGGGCGCGGG GCGAAGCGGTGCGGCGCCGGCAGGAAGGAAATGGGCGGGGAGGGCCTT CGTGCGTCGCCGCGCCGCCGTCCCCTTCTCCCTCTCCAGCCTCGGGGCTG TCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGGGGCAGGGCGGGGTTC GGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTAACCTTGTTCT TGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATTGTGCT GTCTCATCATTTTGGCAAAGAATTACTCGAGGCCACCATGCAGATTGAAC TGTCCACTTGCTTCTTCCTGTGCCTCCTGCGGTTTTGCTTCTCGGCCACCC GCCGGTATTACTTAGGTGCTGTGGAACTGAGCTGGGACTACATGCAGTCC GACCTGGGAGAACTGCCGGTGGACGCGAGATTCCCACCTAGAGTCCCGA AGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAAGACCCTGTTCGTG GAGTTCACTGACCACCTTTTCAATATTGCCAAGCCGCGCCCCCCCTGGAT GGGCCTGCTTGGTCCTACGATCCAAGCAGAGGTCTACGACACCGTGGTC ATCACACTGAAGAACATGGCCTCACACCCCGTGTCGCTGCATGCTGTGG GAGTGTCCTACTGGAAGGCCTCAGAGGGTGCCGAATATGATGACCAGAC CAGCCAGAGGGAAAAGGAGGATGACAAAGTGTTCCCGGGTGGCAGCCA CACTTACGTGTGGCAAGTGCTGAAGGAAAACGGGCCTATGGCGTCGGAC CCCCTATGCCTGACCTACTCCTACCTGTCCCATGTGGACCTTGTGAAGGA TCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCAGAGAAGGCAGC CTGGCGAAGGAAAAGACTCAGACCCTGCACAAGTTCATTCTGTTGTTTGC TGTGTTCGATGAAGGAAAGTCCTGGCACTCAGAAACCAAGAACTCGCTG ATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCCTAAAATGCACA CCGTCAACGGATATGTGAACAGGTCGCTCCCTGGCCTCATCGGCTGCCAC AGAAAGTCCGTGTATTGGCATGTGATCGGCATGGGTACTACTCCGGAAG TGCATAGTATCTTTCTGGAGGGCCATACCTTCTTGGTGCGCAACCACAGA CAGGCCTCGCTGGAAATCTCGCCTATCACTTTCTTGACTGCGCAGACCCT CCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCACATCAGCTCCCATC AGCATGATGGGATGGAGGCCTATGTCAAAGTGGACTCCTGCCCTGAGGA GCCACAGCTCCGGATGAAGAACAATGAGGAAGCGGAGGATTACGACGA CGACCTGACTGACAGCGAAATGGACGTCGTGCGATTCGATGACGACAAC AGCCCGTCCTTCATCCAAATTAGATCAGTGGCGAAGAAGCACCCCAAGA CCTGGGTGCACTACATTGCCGCCGAGGAAGAGGACTGGGACTACGCGCC GCTGGTGCTGGCGCCAGACGACAGGAGCTACAAGTCCCAGTACCTCAAC AACGGGCCGCAGCGCATTGGCAGGAAGTACAAGAAAGTCCGCTTCATGG CCTACACTGATGAAACCTTCAAGACGAGGGAAGCCATCCAGCACGAGTC AGGCATCCTGGGACCGCTCCTTTACGGCGAAGTCGGGGATACCCTGCTC ATCATTTTCAAGAACCAGGCATCGCGGCCCTACAACATCTACCCTCACGG GATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTGCCCAAGGGAGTG AAGCACCTGAAGGATTTTCCCATCCTGCCGGGAGAAATCTTCAAGTACA AGTGGACCGTGACTGTGGAAGATGGCCCTACCAAGTCGGACCCTCGCTG TCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGCGCGACCTGGCCT CGGGGCTGATTGGTCCGCTGCTGATCTGCTACAAGGAGTCCGTGGACCA GCGCGGGAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGTTCTCT GTCTTTGATGAAAACAGATCGTGGTACTTGACTGAGAATATCCAGCGGTT CCTGCCCAACCCAGCGGGAGTGCAACTGGAGGACCCGGAGTTCCAGGCC TCAAACATTATGCACTCTATCAACGGCTATGTGTTCGACTCGCTCCAACT GAGCGTGTGCCTGCATGAAGTGGCATACTGGTACATTCTGTCCATCGGAG CCCAGACCGACTTCCTGTCCGTGTTCTTCTCCGGATACACCTTCAAGCAT AAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATTTTCCGGAGAAAC TGTGTTCATGTCAATGGAAAACCCGGGCTTGTGGATTCTGGGTTGCCATA ACTCGGACTTCCGGAATAGAGGGATGACCGCCCTGCTGAAAGTGTCCAG CTGTGACAAGAATACCGGCGATTACTACGAGGACAGCTATGAGGACATC TCCGCTTATCTGCTGTCCAAGAACAACGCCATTGAACCCAGGTCCTTCTC CCAAAACGGTGCACCGACCTCCGAAAGCGCCACCCCAGAGTCAGGACCT GGCTCGGAACCGGCTACCTCGGGCTCAGAGACACCGGGGACTTCCGAGT CCGCAACCCCCGAGAGTGGACCCGGATCCGAACCAGCAACCTCAGGATC AGAAACCCCGGGAACTTCGGAATCCGCCACTCCCGAGTCGGGACCAGGC ACCTCCACTGAGCCTTCCGAGGGAAGCGCCCCCGGATCCCCTGCTGGATC CCCTACCAGCACTGAAGAAGGCACCTCAGAATCCGCGACCCCTGAGTCC

	GGCCCTGGAAGCGAACCCGCCACCTCCGGTTCCGAAACCCCTGGGACTA GCGAGAGCGCCACTCCGGAATCGGGCCCAGGAAGCCCTGCCGGATCCCC GACCAGCACCGAGGAGGGAAGCCCCGCCGGGTCACCGACTTCCACTGAG GAGGGAGCCTCATCCCCCCCCGTGCTGAAGCGGCATCAAAGAGAGATCA CCAGGACCACTCTCCAGTCCGATCAGGAAGAAATTGACTACGACGATAC TATCAGCGTGGAGATGAAGAAGGAGGACTTCGACATCTACGATGAGGAT GAGAACCAGTCCCCTCGGAGCTTTCAGAAGAAAACCCGCCACTACTTCA TCGCTGCCGTGGAGCGGCTGTGGGATTACGGGATGTCCAGCTCACCGCA TGTGCTGCGGAATAGAGCGCAGTCAGGATCGGTGCCCCAGTTCAAGAAG GTCGTGTTCCAAGAGTTCACCGACGGGTCCTTCACTCAACCCCTGTACCG GGGCGAACTCAACGAACACCTGGGACTGCTTGGGCCGTATATCAGGGCA GAAGTGGAAGATAACATCATGGTCACCTTCCGCAACCAGGCCTCCCGGC CGTACAGCTTCTACTCTTCACTGATCTCCTACGAGGAAGATCAGCGGCAG GGAGCCGAGCCCCGGAAGAACTTCGTCAAGCCTAACGAAACTAAGACCT ACTTTTGGAAGGTCCAGCATCACATGGCCCCGACCAAAGACGAGTTCGA CTGTAAAGCCTGGGCCTACTTCTCCGATGTGGACCTGGAGAAGGACGTG CACTCGGGACTCATTGGCCCGCTCCTTGTGTGCCATACTAATACCCTGAA CCCTGCTCACGGTCGCCAAGTCACAGTGCAGGAGTTCGCCCTCTTCTTCA CCATCTTCGATGAAACAAAGTCCTGGTACTTTACTGAGAACATGGAACG CAATTGCAGGGCACCCTGCAACATCCAGATGGAAGATCCCACCTTCAAG GAAAACTACCGGTTTCATGCCATTAACGGCTACATAATGGACACGTTGCC AGGACTGGTCATGGCCCAGGACCAGAGAATCCGGTGGTATCTGCTCTCC ATGGGCTCCAACGAAAACATTCACAGCATTCATTTTTCCGGCCATGTGTT CACCGTCCGGAAGAAGGAAGAGTACAAGATGGCTCTGTACAACCTCTAC CCTGGAGTGTTCGAGACTGTGGAAATGCTGCCTAGCAAGGCCGGCATTT GGAGAGTGGAATGCCTGATCGGAGAGCATTTGCACGCCGGAATGTCCAC CCTGTTTCTTGTGTACTCCAACAAGTGCCAGACCCCGCTGGGAATGGCCT CAGGTCATATTAGGGATTTCCAGATCACTGCTTCGGGGCAGTACGGGCA GTGGGCACCTAAGTTGGCCCGGCTGCACTACTCTGGCTCCATCAATGCCT GGTCCACCAAGGAACCCTTCTCCTGGATTAAGGTGGACCTCCTGGCCCCA ATGATTATTCACGGTATTAAGACCCAGGGTGCCCGACAGAAGTTCTCCTC ACTCTACATCTCGCAATTCATCATAATGTACAGCCTGGATGGGAAGAAGT GGCAGACCTACCGGGGAAACTCCACTGGAACGCTCATGGTGTTTTTCGG CAACGTGGACTCCTCCGGCATTAAGCACAACATCTTCAACCCTCCGATCA TTGCTCGGTACATCCGGCTGCACCCAACTCACTACAGCATCCGGTCCACC CTGCGGATGGAACTGATGGGTTGTGACCTGAACTCCTGCTCCATGCCCCT TGGGATGGAATCCAAGGCCATTAGCGATGCACAGATCACCGCCTCTTCA TACTTCACCAACATGTTCGCGACCTGGTCCCCGTCGAAGGCCCGCCTGCA CCTCCAAGGTCGCTCCAATGCGTGGCGGCCTCAAGTGAACAACCCCAAG GAGTGGCTCCAGGTCGACTTCCAAAAGACCATGAAGGTCACCGGAGTGA CCACCCAGGGCGTGAAGTCCCTGCTGACCTCTATGTACGTTAAGGAGTTC CTCATCTCCTCAAGCCAAGACGGACATCAGTGGACCCTGTTCTTCCAAAA CGGAAAAGTCAAAGTATTCCAGGGCAACCAGGACTCCTTCACCCCTGTG GTCAACAGCCTGGACCCCCCATTGCTGACCCGCTACCTCCGCATCCACCC CCAAAGCTGGGTCCACCAGATCGCACTGCGCATGGAGGTCCTTGGATGC GAAGCCCAAGATCTGTACTAAGCGGCCGCTCATAATCAACCTCTGGATT ACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTT ACGCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCC CGTATGGCTTTCATTTTCTCCTCCTTGTATAAATCCTGGTTGCTGTCTCTTT ATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGCACTGT GTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGC TCCTTTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAACTCA TCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTCGGCTGTTGGGCACT GACAATTCCGTGGTGTTGTCGGGGAAATCATCGTCCTTTCCTTGGCTGCT CGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCC CTTCGGCCCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCT CTGCGGCCTCTTCCGCGTCTTCGCCTTCGCCCTCAGACGAGTCGGATCTC CCTTTGGGCCGCCTCCCCGCTGCCTAGGCGACTGTGCCTTCTAGTTGCCA GCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGC CACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTC

	TGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAA GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGAAGACCATGG GCGCGCCAGGCCTGTCGACGCCCGGGCGGTACCGCGATCGCTCGCGACG CATAAAGTATATGTGACGTGGTTGTACAGACGCCATCTTGGAATCCAATA TGTCTGCCGGCGATTAGATCATGCGCGCGCGCAGCGCGCTGCGCGCAGC GCAGGCATGACTGAGCCGGCAGACATATTGGATTCCAAGATGGCGTCTG TACAACCACGTGCTTAAGCTGCAGACTAGTGAGCTCGTTAAC
SEQ ID NO: 30 A1MB2增强子	GGCCCCAGGTTAATTTTTAAAAAGCAGTCAAAGGTCAAAGTGGCCCTTG GCAGCATTTACTCTCTCTATTGACTTTGGTTAATAATCTCAGGAGCACAA ACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGGACTTATCCTCTG GGCCTCTCCCCACCTTCGATGGCCCCAGGTTAATTTTTAAAAAGCAGTCA AAGGTCAAAGTGGCCCTTGGCAGCATTTACTCTCTCTATTGACTTTGGTT AATAATCTCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAAC ATCCTGGACTTATCCTCTGGGCCTCTCCCCACC
SEQ ID NO: 31 mTTR启动子	GATATCTACCTGCTGATCGCCCGGCCCCTGTTCAAACATGTCCTAATACT CTGTCGGGGCAAAGGTCGGCAGTAGTTTTCCATCTTACTCAACATCCTCC CAGTGTACGTAGGATCCTGTCTGTCTGCACATTTCGTAGAGCGAGTGTTC CGATACTCTAATCTCCCGGGGCAAAGGTCGTATTGACTTAGGTTACTTAT TCTCCTTTTGTTGACTAAGTCAATAATCAGAATCAGCAGGTTTGGAGTCA GCTTGGCAGGGATCAGCAGCCTGGGTTGGAAGGAGGGGGTATAAAAGCC CCTTCACCAGGAGAAGCCGTCACACAGATCCACAAGCTCCTGCTAG
SEQ ID NO: 32 嵌合内含子	TCAGGAGCACAAACATTCCTGGAGGCAGGAGAAGAAATCAACATCCTGG ACTTATCCTCTGGGCCTCTCCCCACCGATATCTACCTGCTGATCGCCCGG CCCCTGTTCAAACATGTCCTAATACTCTGTCGGGGCAAAGGTCGGCAGTA GTTTTCCATCTTACTCAACATCCTCCCAGTGTACGTAGGATCCTGTCTGTC TGCACATTTCGTAGAGCGAGTGTTCCGATACTCTAATCTCCCGGGGCAAA GGTCGTATTGACTTAGGTTACTTATTCTCCTTTTGTTGACTAAGTCAATAA TCAGAATCAGCAGGTTTGGAGTCAGCTTGGCAGGGATCAGCAGCCTGGG TTGGAAGGAGGGGGTATAAAAGCCCCTTCACCAGGAGAAGCCGTCACAC AGATCCACAAGCTCCTGCTAGAGTCGCTGCGCGCTGCCTTCGCCCCGTGC CCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGGCTCTGACTGACCGCG TTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAA TTAGCGCTTGGTTTATTGACGGCTTGTTTCTTTTCTGTGGCTGCGTGAAAG CCTTGAGGGGCTCCGGGAAGGCCCTTTGTGCGGGGGGAGCGGCTCGGGG GGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGC TGCCCGGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCT CCGCAGTGTGCGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGC GGGGGGGGCTGCGAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTG GGGGGGTGAGCAGGGGGTGTGGGCGCGTCGGTCGGGCTGCAACCCCCCC TGCACCCCCCTCCCCGAGTTGCTGAGCACGGCCCGGCTTCGGGTGCGGG GCTCCGTACGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGGTGG CGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGCCTCGGGCCGGGGAG GGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGCGGCTGTCGA GGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAGAGGG CGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAG GCGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCC GGCAGGAAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCC GTCCCCTTCTCCCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGC CTTCGGGGGGGGACGGGGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCG GCGGCTCTAGAGCCTCTGCTAACCTTGTTCTTGCCTTCTTCTTTTTCCTAC AGCTCCTGGGCAACGTGCTGGTTATTGTGCTGTCTCATCATTTTGGCAAA GAATTA
SEQ ID NO: 33 WPRE	TCATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTC TTAACTATGTTGCTCCTTTTACGCTATGTGGATACGCTGCTTTAATGCCTT TGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCTCCTCCTTGTATA AATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAA CGTGGCGTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGG CATTGCCACCACCTGTCAGCTCCTTTCCGGGACTTTCGCTTTCCCCCTCCC

	TATTGCCACGGCGGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAG GGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTTGTCGGGGAAATC ATCGTCCTTTCCTTGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGG GACGTCCTTCTGCTACGTCCCTTCGGCCCTCAATCCAGCGGACCTTCCTTC CCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCCTTCGCC CTCAGACGAGTCGGATCTCCCTTTGGGCCGCCTCCCCGCTG
SEQ ID NO: 34 bGHpA	CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAAT GAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGG TGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAG GCATGCTGGGGA
SEQ ID NO: 35 编码 BDD-co6FVIII (V1.0)的核苷酸 序列 (无XTEN)	GCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGACTACAT GCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCCCCCCGC GTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAAAACCCT CTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGCGCCCAC CTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTACGACAC CGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCCCTGCATG CGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGTACGACGA CCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCCCGGGCGG CTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACCTATGGCA TCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGACCTCGTG AAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCCGCGAAG GTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCATCCTTTTG TTCGCTGTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTAAGAACT CGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCCTAAAAT GCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTCATCGGT TGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCACTACGC CTGAAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTGCGCAAC CACCGCCAGGCCTCTCTGGAAATCTCCCCGATTACCTTTCTGACCGCCCA GACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACATCTCCA GCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTCATGCCC GGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGAGGACTA TGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTCGATGAC GACAACAGCCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGAAGCACC CCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTGGGACTA CGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCCCAGTATC TGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAGTGCGGTT CATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCATTCAACAT GAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTCGGCGATACCC TGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACATCTACCCT CACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGCCGAAGG GCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAATCTTCAA GTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAGCGATCCT AGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACGGGACCT GGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAGTCGGTG GATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGATCCTGT TCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAACATCCAG AGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTGAGTTTC AGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGACTCGCTG CAGCTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCTGTCCAT CGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACACCTTTA AGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTCTCCGGC GAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCTGGGAT GCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTGAAGGT GTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCTACGAG GATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCCGCGCA GCTTCAGCCAGAACCCGCCTGTGCTGAAGAGGCACCAGCGAGAAATTAC CCGGACCACCCTCCAATCGGATCAGGAGGAAATCGACTACGACGACACC

	ATCTCGGTGGAAATGAAGAAGGAAGATTTCGATATCTACGACGAGGACG AAAATCAGTCCCCTCGCTCATTCCAAAAGAAAACTAGACACTACTTTATC GCCGCGGTGGAAAGACTGTGGGACTATGGAATGTCATCCAGCCCTCACG TCCTTCGGAACCGGGCCCAGAGCGGATCGGTGCCTCAGTTCAAGAAAGT GGTGTTCCAGGAGTTCACCGACGGCAGCTTCACCCAGCCGCTGTACCGG GGAGAACTGAACGAACACCTGGGCCTGCTCGGTCCCTACATCCGCGCGG AAGTGGAGGATAACATCATGGTGACCTTCCGTAACCAAGCATCCAGACC TTACTCCTTCTATTCCTCCCTGATCTCATACGAGGAGGACCAGCGCCAAG GCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCAACGAGACTAAGACCTA CTTCTGGAAGGTCCAACACCATATGGCCCCGACCAAGGATGAGTTTGAC TGCAAGGCCTGGGCCTACTTCTCCGACGTGGACCTTGAGAAGGATGTCC ATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTCACACCAACACCCTGAAC CCAGCGCATGGACGCCAGGTCACCGTCCAGGAGTTTGCTCTGTTCTTCAC CATTTTTGACGAAACTAAGTCCTGGTACTTCACCGAGAATATGGAGCGA AACTGTAGAGCGCCCTGCAATATCCAGATGGAAGATCCGACTTTCAAGG AGAACTATAGATTCCACGCCATCAACGGGTACATCATGGATACTCTGCC GGGGCTGGTCATGGCCCAGGATCAGAGGATTCGGTGGTACTTGCTGTCA ATGGGATCGAACGAAAACATTCACTCCATTCACTTCTCCGGTCACGTGTT CACTGTGCGCAAGAAGGAGGAGTACAAGATGGCGCTGTACAATCTGTAC CCCGGGGGTGTTCGAAACTGTGGAGATGCTGCCGTCCAAGGCCGGCATCT GGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCACGCGGGGATGTCCAC CCTCTTCCTGGTGTACTCGAATAAGTGCCAGACCCCGCTGGGCATGGCCT CGGGCCACATCAGAGACTTCCAGATCACAGCAAGCGGACAATACGGCCA ATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCGGATCGATCAACGCA TGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGTGGACCTCCTGGCCCC TATGATTATCCACGGAATTAAGACCCAGGGCGCCAGGCAGAAGTTCTCC TCCCTGTACATCTCGCAATTCATCATCATGTACAGCCTGGACGGGAAGAA GTGGCAGACTTACAGGGGAAACTCCACCGGCACCCTGATGGTCTTTTTCG GCAACGTGGATTCCTCCGGCATTAAGCACAACATCTTCAACCCACCGATC ATAGCCAGATATATTAGGCTCCACCCCACTCACTACTCAATCCGCTCAAC TCTTCGGATGGAACTCATGGGGTGCGACCTGAACTCCTGCTCCATGCCGT TGGGGATGGAATCAAAGGCTATTAGCGACGCCCAGATCACCGCGAGCTC CTACTTCACTAACATGTTCGCCACCTGGAGCCCCTCCAAGGCCAGGCTGC ACTTGCAGGGACGGTCAAATGCCTGGCGGCCGCAAGTGAACAATCCGAA GGAATGGCTTCAAGTGGATTTCCAAAAGACCATGAAAGTGACCGGAGTC ACCACCCAGGGAGTGAAGTCCCTTCTGACCTCGATGTATGTGAAGGAGT TCCTGATTAGCAGCAGCCAGGACGGGCACCAGTGGACCCTGTTCTTCCA AAACGGAAAGGTCAAGGTGTTCCAGGGGAACCAGGACTCGTTCACACCC GTGGTGAACTCCCTGGACCCCCCACTGCTGACGCGGTACTTGAGGATTCA TCCTCAGTCCTGGGTCCATCAGATTGCATTGCGAATGGAAGTCCTGGGCT GCGAGGCCCAGGACCTGTACTGA
SEQ ID NO: 36 编码 coBDDFVIII (V2.0)的核苷酸 序列 (无XTEN)	GCCACCCGCCGGTATTACTTAGGTGCTGTGGAACTGAGCTGGGACTACAT GCAGTCCGACCTGGGAGAACTGCCGGTGGACGCGAGATTCCCACCTAGA GTCCCGAAGTCCTTCCCATTCAACACCTCCGTGGTCTACAAAAAGACCCT GTTCGTGGAGTTCACTGACCACCTTTTCAATATTGCCAAGCCGCGCCCCC CCTGGATGGGCCTGCTTGGTCCTACGATCCAAGCAGAGGTCTACGACAC CGTGGTCATCACACTGAAGAACATGGCCTCACACCCCGTGTCGCTGCATG CTGTGGGAGTGTCCTACTGGAAGGCCTCAGAGGGTGCCGAATATGATGA CCAGACCAGCCAGAGGGAAAAGGAGGATGACAAAGTGTTCCCGGGTGG CAGCCACACTTACGTGTGGCAAGTGCTGAAGGAAAACGGGCCTATGGCG TCGGACCCCCTATGCCTGACCTACTCCTACCTGTCCCATGTGGACCTTGT GAAGGATCTCAACTCGGGACTGATCGGCGCCCTCTTGGTGTGCAGAGAA GGCAGCCTGGCGAAGGAAAAGACTCAGACCCTGCACAAGTTCATTCTGT TGTTTGCTGTGTTCGATGAAGGAAAGTCCTGGCACTCAGAAACCAAGAA CTCGCTGATGCAGGATAGAGATGCGGCCTCGGCCAGAGCCTGGCCTAAA ATGCACACCGTCAACGGATATGTGAACAGGTCGCTCCCTGGCCTCATCG GCTGCCACAGAAAGTCCGTGTATTGGCATGTGATCGGCATGGGTACTACT CCGGAAGTGCATAGTATCTTTCTGGAGGGCCATACCTTCTTGGTGCGCAA CCACAGACAGGCCTCGCTGGAAATCTCGCCTATCACTTTCTTGACTGCGC

	AGACCCTCCTTATGGACCTTGGACAGTTCCTGCTGTTCTGTCACATCAGC TCCCATCAGCATGATGGGATGGAGGCCTATGTCAAAGTGGACTCCTGCC CTGAGGAGCCACAGCTCCGGATGAAGAACAATGAGGAAGCGGAGGATT ACGACGACGACCTGACTGACAGCGAAATGGACGTCGTGCGATTCGATGA CGACAACAGCCCGTCCTTCATCCAAATTAGATCAGTGGCGAAGAAGCAC CCCAAGACCTGGGTGCACTACATTGCCGCCGAGGAAGAGGACTGGGACT ACGCGCCGCTGGTGCTGGCGCCAGACGACAGGAGCTACAAGTCCCAGTA CCTCAACAACGGGCCGCAGCGCATTGGCAGGAAGTACAAGAAAGTCCGC TTCATGGCCTACACTGATGAAACCTTCAAGACGAGGGAAGCCATCCAGC ACGAGTCAGGCATCCTGGGACCGCTCCTTTACGGCGAAGTCGGGGATAC CCTGCTCATCATTTTCAAGAACCAGGCATCGCGGCCCTACAACATCTACC CTCACGGGATCACAGACGTGCGCCCGCTCTACTCCCGCCGGCTGCCCAA GGGAGTGAAGCACCTGAAGGATTTTCCCATCCTGCCGGGAGAAATCTTC AAGTACAAGTGGACCGTGACTGTGGAAGATGGCCCTACCAAGTCGGACC CTCGCTGTCTGACCCGGTACTATTCCTCGTTTGTGAACATGGAGCGCGAC CTGGCCTCGGGGCTGATTGGTCCGCTGCTGATCTGCTACAAGGAGTCCGT GGACCAGCGCGGGAACCAGATCATGTCCGACAAGCGCAACGTGATCCTG TTCTCTGTCTTTGATGAAAACAGATCGTGGTACTTGACTGAGAATATCCA GCGGTTCCTGCCCAACCCAGCGGGAGTGCAACTGGAGGACCCGGAGTTC CAGGCCTCAAACATTATGCACTCTATCAACGGCTATGTGTTCGACTCGCT CCAACTGAGCGTGTGCCTGCATGAAGTGGCATACTGGTACATTCTGTCCA TCGGAGCCCAGACCGACTTCCTGTCCGTGTTCTTCTCCGGATACACCTTC AAGCATAAGATGGTGTACGAGGACACTCTGACCCTCTTCCCATTTTCCGG AGAAACTGTGTTCATGTCAATGGAAAACCCGGGCTTGTGGATTCTGGGTT GCCATAACTCGGACTTCCGGAATAGAGGGATGACCGCCCTGCTGAAAGT GTCCAGCTGTGACAAGAATACCGGCGATTACTACGAGGACAGCTATGAG GACATCTCCGCTTATCTGCTGTCCAAGAACAACGCCATTGAACCCAGGTC CTTCTCCCAAAACGGTGCACCGGCCTCATCCCCCCCCCGTGCTGAAGCGGC ATCAAAGAGAGATCACCAGGACCACTCTCCAGTCCGATCAGGAAGAAAT TGACTACGACGATACTATCAGCGTGGAGATGAAGAAGGAGGACTTCGAC ATCTACGATGAGGATGAGAACCAGTCCCCTCGGAGCTTTCAGAAGAAAA CCCGCCACTACTTCATCGCTGCCGTGGAGCGGCTGTGGGATTACGGGATG TCCAGCTCACCGCATGTGCTGCGGAATAGAGCGCAGTCAGGATCGGTGC CCCAGTTCAAGAAGGTCGTGTTCCAAGAGTTCACCGACGGGTCCTTCACT CAACCCCTGTACCGGGGCGAACTCAACGAACACCTGGGACTGCTTGGGC CGTATATCAGGGCAGAAGTGGAAGATAACATCATGGTCACCTTCCGCAA CCAGGCCTCCCGGCCGTACAGCTTCTACTCTTCACTGATCTCCTACGAGG AAGATCAGCGGCAGGGAGCCGAGCCCCGGAAGAACTTCGTCAAGCCTAA CGAAACTAAGACCTACTTTTGGAAGGTCCAGCATCACATGGCCCCGACC AAAGACGAGTTCGACTGTAAAGCCTGGGCCTACTTCTCCGATGTGGACCT GGAGAAGGACGTGCACTCGGGACTCATTGGCCCGCTCCTTGTGTGCCAT ACTAATACCCTGAACCCTGCTCACGGTCGCCAAGTCACAGTGCAGGAGT TCGCCCTCTTCTTCACCATCTTCGATGAAACAAAGTCCTGGTACTTTACTG AGAACATGGAACGCAATTGCAGGGCACCCTGCAACATCCAGATGGAAGA TCCCACCTTCAAGGAAAACTACCGGTTTCATGCCATTAACGGCTACATAA TGGACACGTTGCCAGGACTGGTCATGGCCCAGGACCAGAGAATCCGGTG GTATCTGCTCTCCATGGGCTCCAACGAAAACATTCACAGCATTCATTTTT CCGGCCATGTGTTCACCGTCCGGAAGAAGGAAGAGTACAAGATGGCTCT GTACAACCTCTACCCTGGAGTGTTCGAGACTGTGGAAATGCTGCCTAGCA AGGCCGGCATTTGGAGAGTGGAATGCCTGATCGGAGAGCATTTGCACGC CGGAATGTCCACCCTGTTTCTTGTGTACTCCAACAAGTGCCAGACCCCGC TGGGAATGGCCTCAGGTCATATTAGGGATTTCCAGATCACTGCTTCGGGG CAGTACGGGCAGTGGGCACCTAAGTTGGCCCGGCTGCACTACTCTGGCT CCATCAATGCCTGGTCCACCAAGGAACCCTTCTCCTGGATTAAGGTGGAC CTCCTGGCCCCAATGATTATTCACGGTATTAAGACCCAGGGTGCCCGACA GAAGTTCTCCTCACTCTACATCTCGCAATTCATCATAATGTACAGCCTGG ATGGGAAGAAGTGGCAGACCTACCGGGGAAACTCCACTGGAACGCTCAT GGTGTTTTTCGGCAACGTGGACTCCTCCGGCATTAAGCACAACATCTTCA ACCCTCCGATCATTGCTCGGTACATCCGGCTGCACCCAACTCACTACAGC ATCCGGTCCACCCTGCGGATGGAACTGATGGGTTGTGACCTGAACTCCTG

	CTCCATGCCCCTTGGGATGGAATCCAAGGCCATTAGCGATGCACAGATC ACCGCCTCTTCATACTTCACCAACATGTTCGCGACCTGGTCCCCGTCGAA GGCCCGCCTGCACCTCCAAGGTCGCTCCAATGCGTGGCGGCCTCAAGTG AACAACCCCAAGGAGTGGCTCCAGGTCGACTTCCAAAAGACCATGAAGG TCACCGGAGTGACCACCCAGGGCGTGAAGTCCCTGCTGACCTCTATGTAC GTTAAGGAGTTCCTCATCTCCTCAAGCCAAGACGGACATCAGTGGACCCT GTTCTTCCAAAACGGAAAAGTCAAAGTATTCCAGGGCAACCAGGACTCC TTCACCCCTGTGGTCAACAGCCTGGACCCCCCATTGCTGACCCGCTACCT CCGCATCCACCCCCAAAGCTGGGTCCACCAGATCGCACTGCGCATGGAG GTCCTTGGATGCGAAGCCCAAGATCTGTACTAA
SEQ ID NO: 37 V1.0 表达盒 TTP-内含子-BDDFVIIIco6XTEN (V1.0)-WPRE-bGHPolyA 表达盒	ATGCAGATTGAGCTGTCCACTTGTTTCTTCCTGTGCCTCCTGCGCTTCTGT TTCTCCGCCACTCGCCGGTACTACCTTGGAGCCGTGGAGCTTTCATGGGA CTACATGCAGAGCGACCTGGGCGAACTCCCCGTGGATGCCAGATTCCCC CCCCGCGTGCCAAAGTCCTTCCCCTTTAACACCTCCGTGGTGTACAAGAA AACCCTCTTTGTCGAGTTCACTGACCACCTGTTCAACATCGCCAAGCCGC GCCCACCTTGGATGGGCCTCCTGGGACCGACCATTCAAGCTGAAGTGTA CGACACCGTGGTGATCACCCTGAAGAACATGGCGTCCCACCCCGTGTCC CTGCATGCGGTCGGAGTGTCCTACTGGAAGGCCTCCGAAGGAGCTGAGT ACGACGACCAGACTAGCCAGCGGGAAAAGGAGGACGATAAAGTGTTCC CGGGCGGCTCGCATACTTACGTGTGGCAAGTCCTGAAGGAAAACGGACC TATGGCATCCGATCCTCTGTGCCTGACTTACTCCTACCTTTCCCATGTGGA CCTCGTGAAGGACCTGAACAGCGGGCTGATTGGTGCACTTCTCGTGTGCC GCGAAGGTTCGCTCGCTAAGGAAAAGACCCAGACCCTCCATAAGTTCAT CCTTTTGTTCGCTGTGTTCGATGAAGGAAAGTCATGGCATTCCGAAACTA AGAACTCGCTGATGCAGGACCGGGATGCCGCCTCAGCCCGCGCCTGGCC TAAAATGCATACAGTCAACGGATACGTGAATCGGTCACTGCCCGGGCTC ATCGGTTGTCACAGAAAGTCCGTGTACTGGCACGTCATCGGCATGGGCA CTACGCCTGAAGTGCACTCCATCTTCCTGGAAGGGCACACCTTCCTCGTG CGCAACCACCGCCAGGCCTCTCTGGAAATCTCCCCGATTACCTTTCTGAC CGCCCAGACTCTGCTCATGGACCTGGGGCAGTTCCTTCTCTTCTGCCACA TCTCCAGCCATCAGCACGACGGAATGGAGGCCTACGTGAAGGTGGACTC ATGCCCGGAAGAACCTCAGTTGCGGATGAAGAACAACGAGGAGGCCGA GGACTATGACGACGATTTGACTGACTCCGAGATGGACGTCGTGCGGTTC GATGACGACAACAGCCCCAGCTTCATCCAGATTCGCAGCGTGGCCAAGA AGCACCCCAAAACCTGGGTGCACTACATCGCGGCCGAGGAAGAAGATTG GGACTACGCCCCGTTGGTGCTGGCACCCGATGACCGGTCGTACAAGTCC CAGTATCTGAACAATGGTCCGCAGCGGATTGGCAGAAAGTACAAGAAAG TGCGGTTCATGGCGTACACTGACGAAACGTTTAAGACCCGGGAGGCCAT TCAACATGAGAGCGGCATTCTGGGACCACTGCTGTACGGAGAGGTCGGC GATACCCTGCTCATCATCTTCAAAAACCAGGCCTCCCGGCCTTACAACAT CTACCCTCACGGAATCACCGACGTGCGGCCACTCTACTCGCGGCGCCTGC CGAAGGGCGTCAAGCACCTGAAAGACTTCCCTATCCTGCCGGGCGAAAT CTTCAAGTATAAGTGGACCGTCACCGTGGAGGACGGGCCCACCAAGAGC GATCCTAGGTGTCTGACTCGGTACTACTCCAGCTTCGTGAACATGGAACG GGACCTGGCATCGGGACTCATTGGACCGCTGCTGATCTGCTACAAAGAG TCGGTGGATCAACGCGGCAACCAGATCATGTCCGACAAGCGCAACGTGA TCCTGTTCTCCGTGTTTGATGAAAACAGATCCTGGTACCTCACTGAAAAC ATCCAGAGGTTCCTCCCAAACCCCGCAGGAGTGCAACTGGAGGACCCTG AGTTTCAGGCCTCGAATATCATGCACTCGATTAACGGTTACGTGTTCGAC TCGCTGCAACTGAGCGTGTGCCTCCATGAAGTCGCTTACTGGTACATTCT GTCCATCGGCGCCCAGACTGACTTCCTGAGCGTGTTCTTTTCCGGTTACA CCTTTAAGCACAAGATGGTGTACGAAGATACCCTGACCCTGTTCCCTTTC TCCGGCGAAACGGTGTTCATGTCGATGGAGAACCCGGGTCTGTGGATTCT GGGATGCCACAACAGCGACTTTCGGAACCGCGGAATGACTGCCCTGCTG AAGGTGTCCTCATGCGACAAGAACACCGGAGACTACTACGAGGACTCCT ACGAGGATATCTCAGCCTACCTCCTGTCCAAGAACAACGCGATCGAGCC GCGCAGCTTCAGCCAGAACGGCGCGCCAACATCAGAGAGCGCCACCCCT GAAAGTGGTCCCGGGAGCGAGCCAGCCACATCTGGGTCGGAAACGCCAG GCACAAGTGAGTCTGCAACTCCCGAGTCCGGACCTGGCTCCGAGCCTGC

	CACTAGCGGCTCCGAGACTCCGGGAACTTCCGAGAGCGCTACACCAGAA AGCGGACCCGGAACCAGTACCGAACCTAGCGAGGGCTCTGCTCCGGGCA GCCCAGCCGGCTCTCCTACATCCACGGAGGAGGGCACTTCCGAATCCGC CACCCCGGAGTCAGGGCCAGGATCTGAACCCGCTACCTCAGGCAGTGAG ACGCCAGGAACGAGCGAGTCCGCTACACCGGAGAGTGGGCCAGGGAGC CCTGCTGGATCTCCTACGTCCACTGAGGAAGGGTCACCAGCGGGCTCGC CCACCAGCACTGAAGAAGGTGCCTCGAGCCCGCCTGTGCTGAAGAGGCA CCAGCGAGAAATTACCCGGACCACCCTCCAATCGGATCAGGAGGAAATC GACTACGACGACACCATCTCGGTGGAAATGAAGAAGGAAGATTTCGATA TCTACGACGAGGACGAAAATCAGTCCCCTCGCTCATTCCAAAAGAAAAC TAGACACTACTTTATCGCCGCGGTGGAAAGACTGTGGGACTATGGAATG TCATCCAGCCCTCACGTCCTTCGGAACCGGGCCCAGAGCGGATCGGTGC CTCAGTTCAAGAAAGTGGTGTTCCAGGAGTTCACCGACGGCAGCTTCAC CCAGCCGCTGTACCGGGGAGAACTGAACGAACACCTGGGCCTGCTCGGT CCCTACATCCGCGCGGAAGTGGAGGATAACATCATGGTGACCTTCCGTA ACCAAGCATCCAGACCTTACTCCTTCTATTCCTCCCTGATCTCATACGAG GAGGACCAGCGCCAAGGCGCCGAGCCCCGCAAGAACTTCGTCAAGCCCA ACGAGACTAAGACCTACTTCTGGAAGGTCCAACACCATATGGCCCCGAC CAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTCTCCGACGTGGAC CTTGAGAAGGATGTCCATTCCGGCCTGATCGGGCCGCTGCTCGTGTGTCA CACCAACACCCTGAACCCAGCGCATGGACGCCAGGTCACCGTCCAGGAG TTTGCTCTGTTCTTCACCATTTTTGACGAAACTAAGTCCTGGTACTTCACC GAGAATATGGAGCGAAACTGTAGAGCGCCCTGCAATATCCAGATGGAAG ATCCGACTTTCAAGGAGAACTATAGATTCCACGCCATCAACGGGTACAT CATGGATACTCTGCCGGGGCTGGTCATGGCCCAGGATCAGAGGATTCGG TGGTACTTGCTGTCAATGGGATCGAACGAAAACATTCACTCCATTCACTT CTCCGGTCACGTGTTCACTGTGCGCAAGAAGGAGGAGTACAAGATGGCG CTGTACAATCTGTACCCCGGGGTGTTCGAAACTGTGGAGATGCTGCCGTC CAAGGCCGGCATCTGGAGAGTGGAGTGCCTGATCGGAGAGCACCTCCAC GCGGGGATGTCCACCCTCTTCCTGGTGTACTCGAATAAGTGCCAGACCCC GCTGGGCATGGCCTCGGGCCACATCAGAGACTTCCAGATCACAGCAAGC GGACAATACGGCCAATGGGCGCCGAAGCTGGCCCGCTTGCACTACTCCG GATCGATCAACGCATGGTCCACCAAGGAACCGTTCTCGTGGATTAAGGT GGACCTCCTGGCCCCTATGATTATCCACGGAATTAAGACCCAGGGCGCC AGGCAGAAGTTCTCCTCCCTGTACATCTCGCAATTCATCATCATGTACAG CCTGGACGGGAAGAAGTGGCAGACTTACAGGGGAAACTCCACCGGCACC CTGATGGTCTTTTTCGGCAACGTGGATTCCTCCGGCATTAAGCACAACAT CTTCAACCCACCGATCATAGCCAGATATATTAGGCTCCACCCCACTCACT ACTCAATCCGCTCAACTCTTCGGATGGAACTCATGGGGTGCGACCTGAAC TCCTGCTCCATGCCGTTGGGGATGGAATCAAAGGCTATTAGCGACGCCC AGATCACCGCGAGCTCCTACTTCACTAACATGTTCGCCACCTGGAGCCCC TCCAAGGCCAGGCTGCACTTGCAGGGACGGTCAAATGCCTGGCGGCCGC AAGTGAACAATCCGAAGGAATGGCTTCAAGTGGATTTCCAAAAGACCAT GAAAGTGACCGGAGTCACCACCCAGGGAGTGAAGTCCCTTCTGACCTCG ATGTATGTGAAGGAGTTCCTGATTAGCAGCAGCCAGGACGGGCACCAGT GGACCCTGTTCTTCCAAAACGGAAAGGTCAAGGTGTTCCAGGGGAACCA GGACTCGTTCACACCCGTGGTGAACTCCCTGGACCCCCCACTGCTGACGC GGTACTTGAGGATTCATCCTCAGTCCTGGGTCCATCAGATTGCATTGCGA ATGGAAGTCCTGGGCTGCGAGGCCCAGGACCTGTACTGA

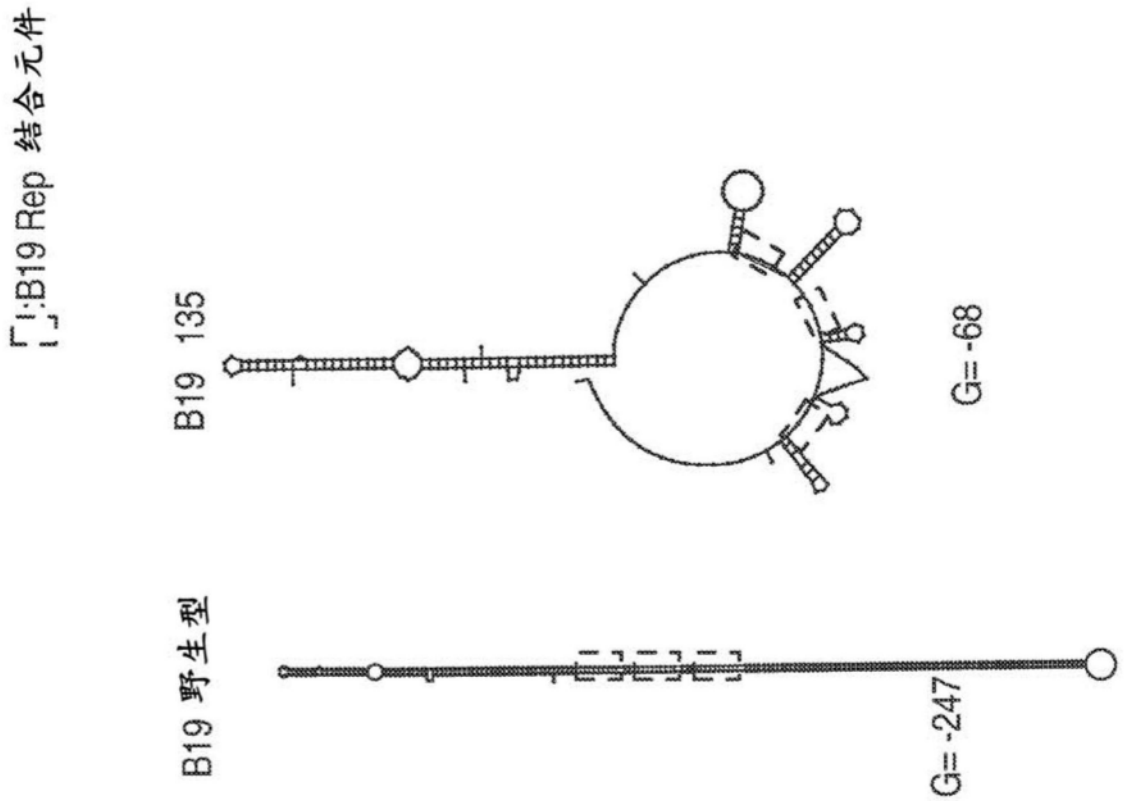


图1A

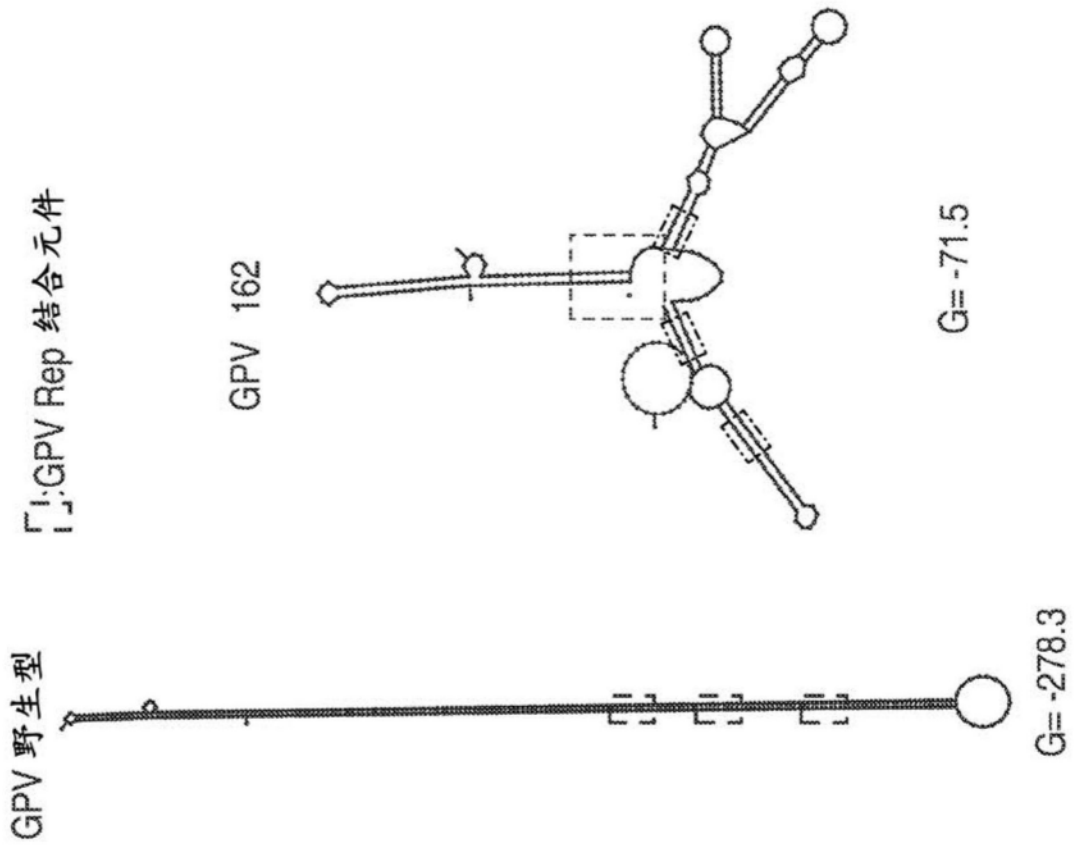


图1B

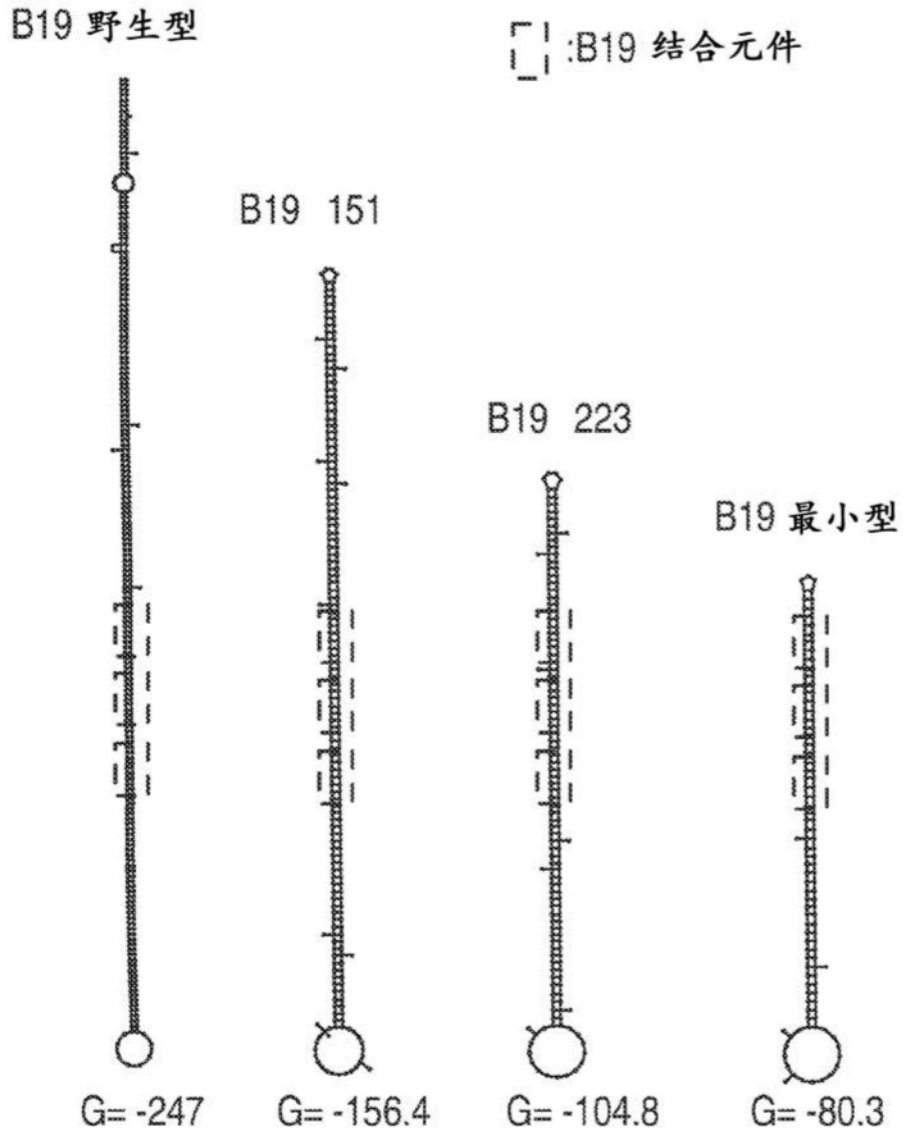


图2A

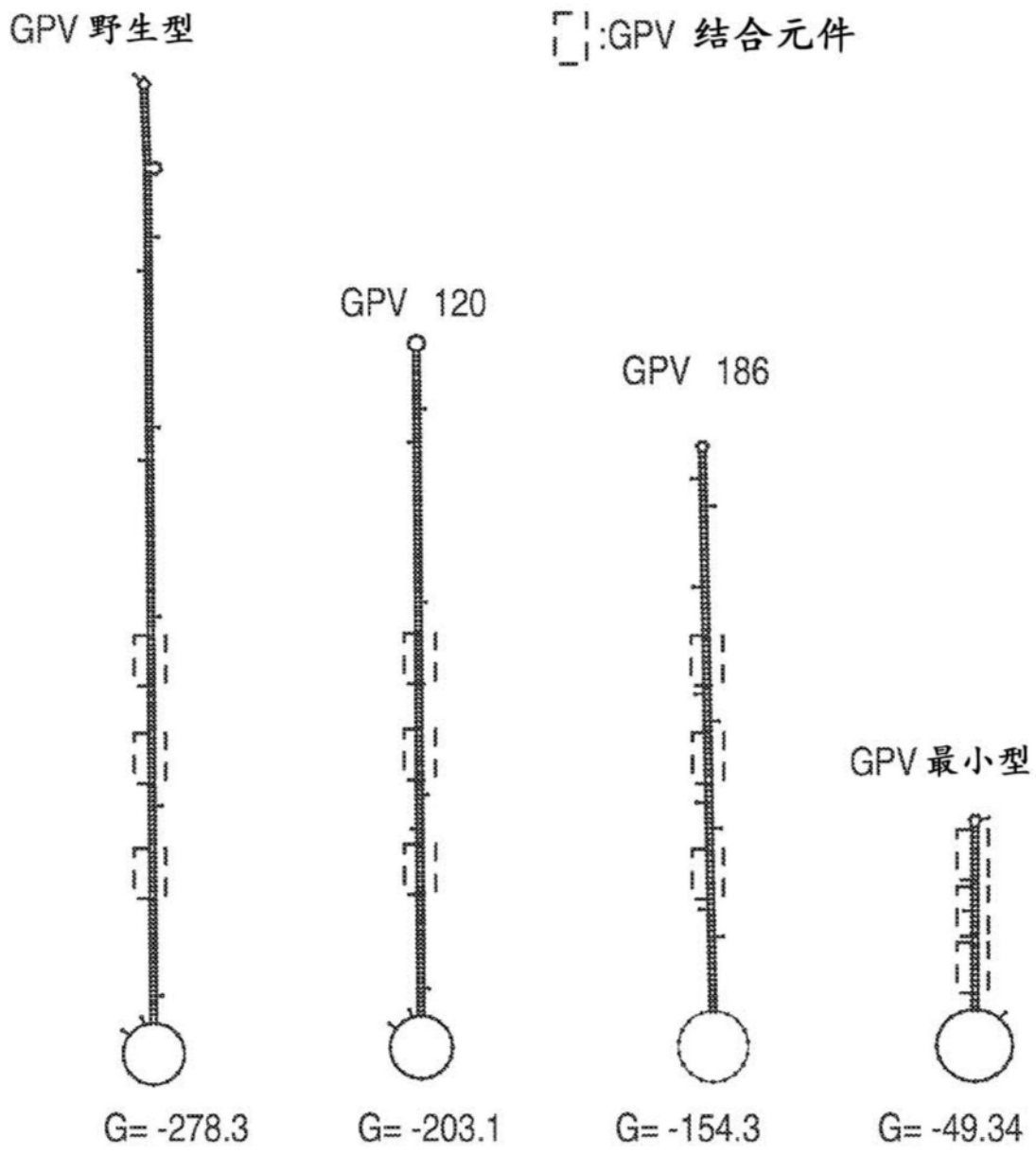

GPV 野生型
:GPV 结合元件
GPV 120
GPV 186
GPV 最小型
G= -278.3
G= -203.1
G= -154.3
G= -49.34

图2B

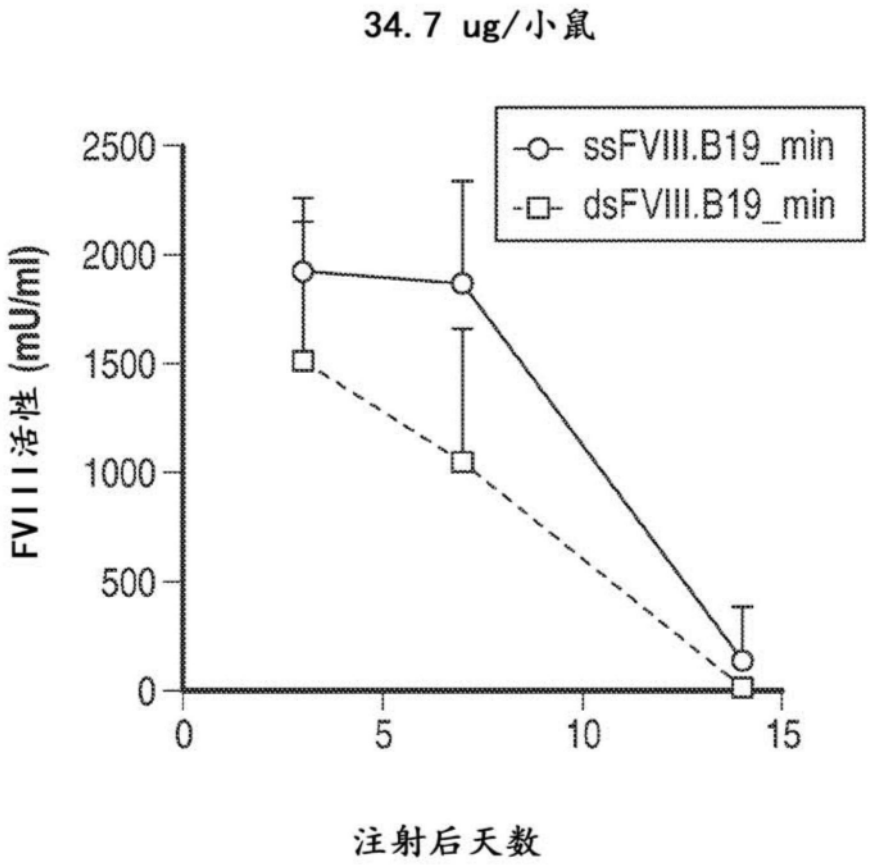

34.7 ug/小鼠
ssFVIII.B19_min
dsFVIII.B19_min
FVIII活性 (mU/ml)
2500
2000
1500
1000
500
0
0
5
10
15
注射后天数

图3

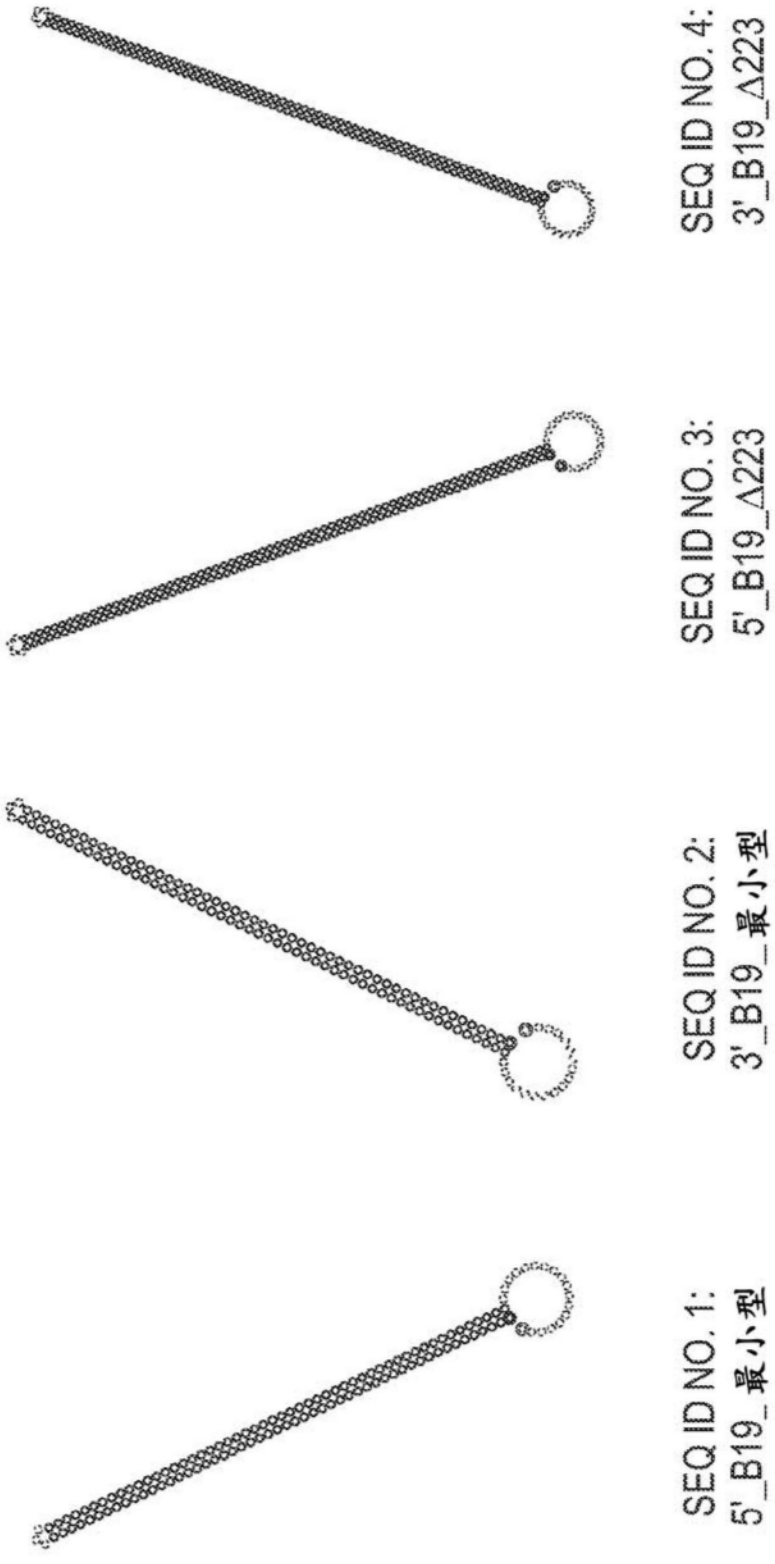

SEQ ID NO. 4:
3'_B19_Δ223
SEQ ID NO. 3:
5'_B19_Δ223
SEQ ID NO. 2:
3'_B19_最小型
SEQ ID NO. 1:
5'_B19_最小型

图4

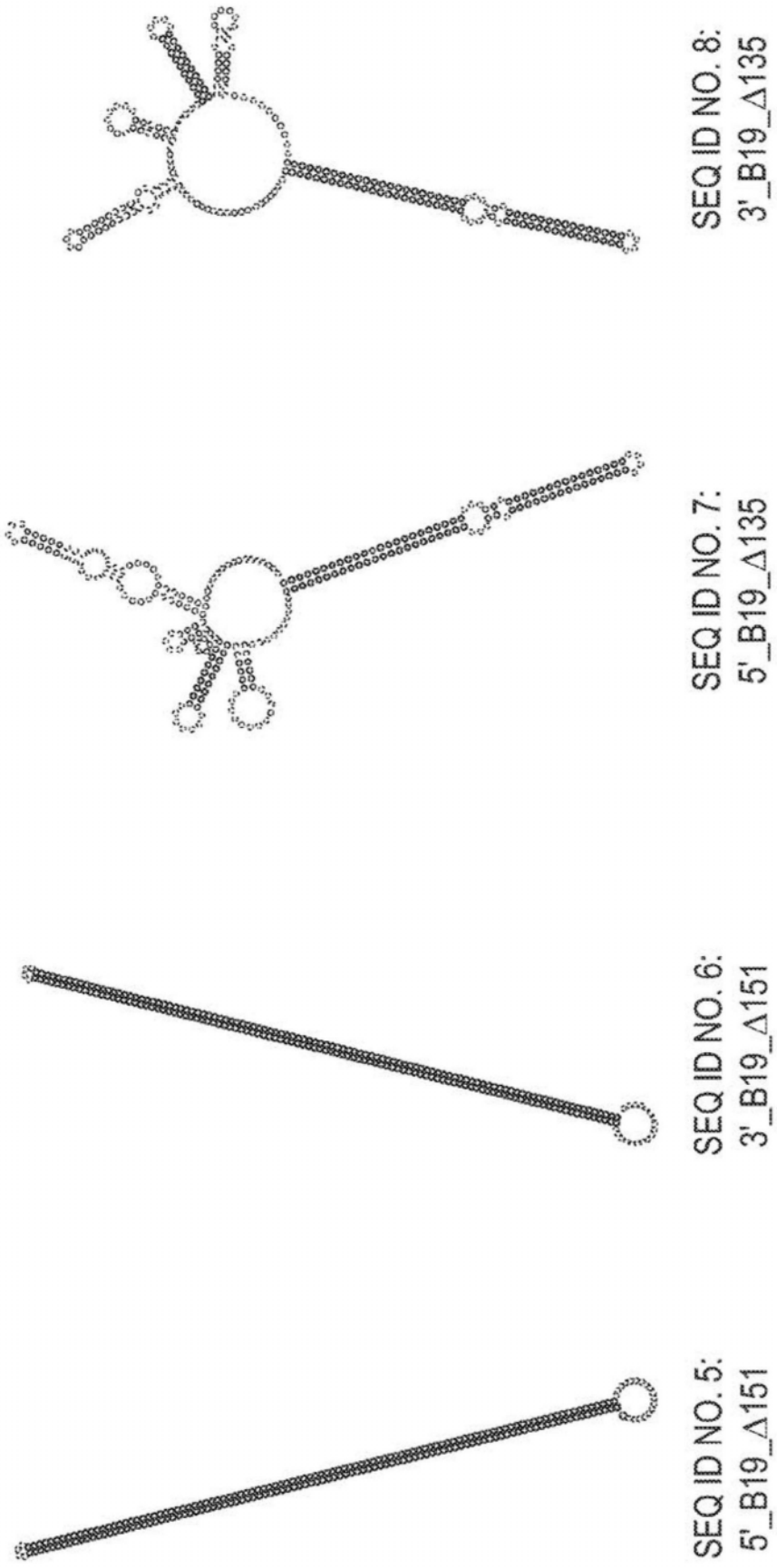

SEQ ID NO. 8:
3'_B19_Δ135
SEQ ID NO. 7:
5'_B19_Δ135
SEQ ID NO. 6:
3'_B19_Δ151
SEQ ID NO. 5:
5'_B19_Δ151

图5

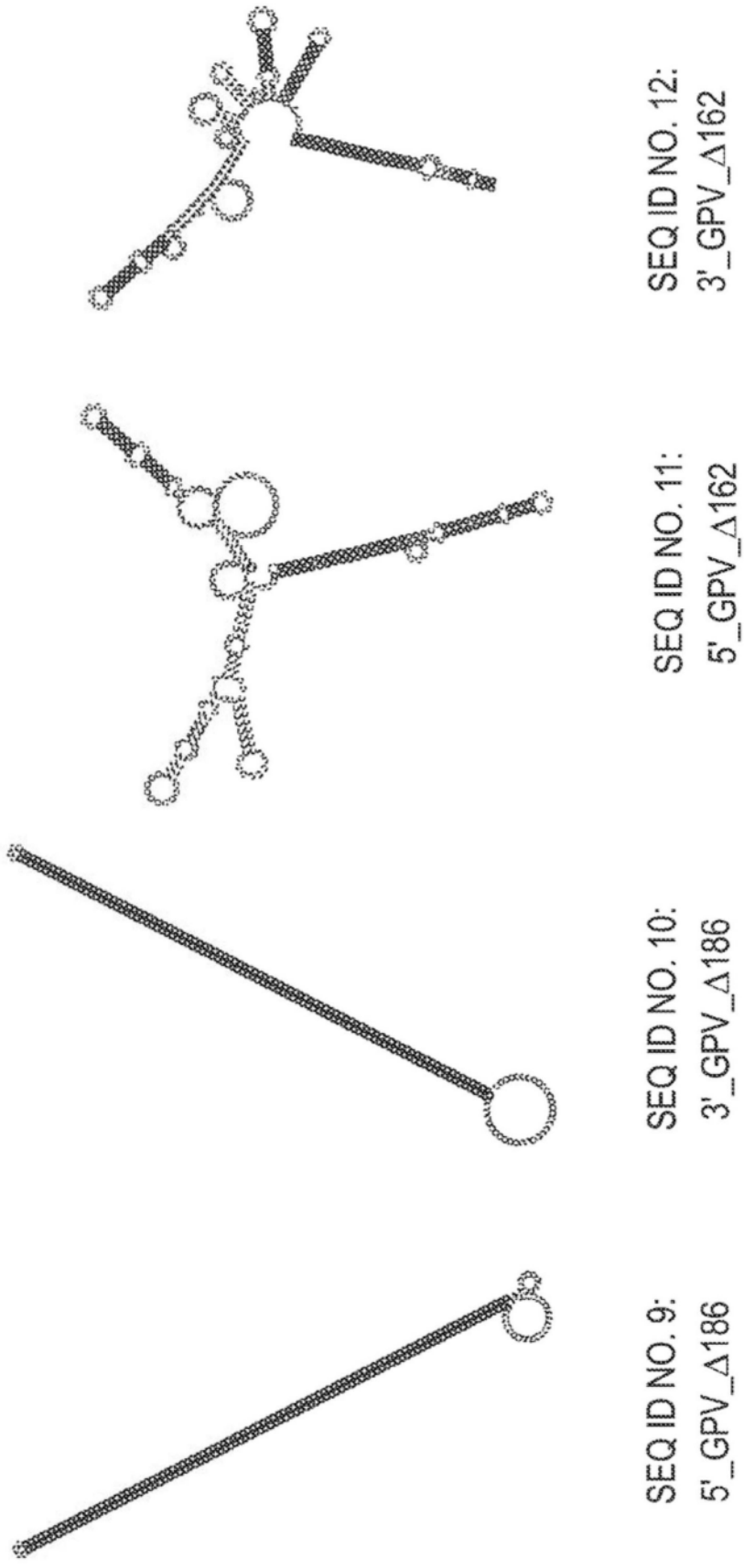

SEQ ID NO. 12:
3'_GPV_Δ162
SEQ ID NO. 11:
5'_GPV_Δ162
SEQ ID NO. 10:
3'_GPV_Δ186
SEQ ID NO. 9:
5'_GPV_Δ186

图6

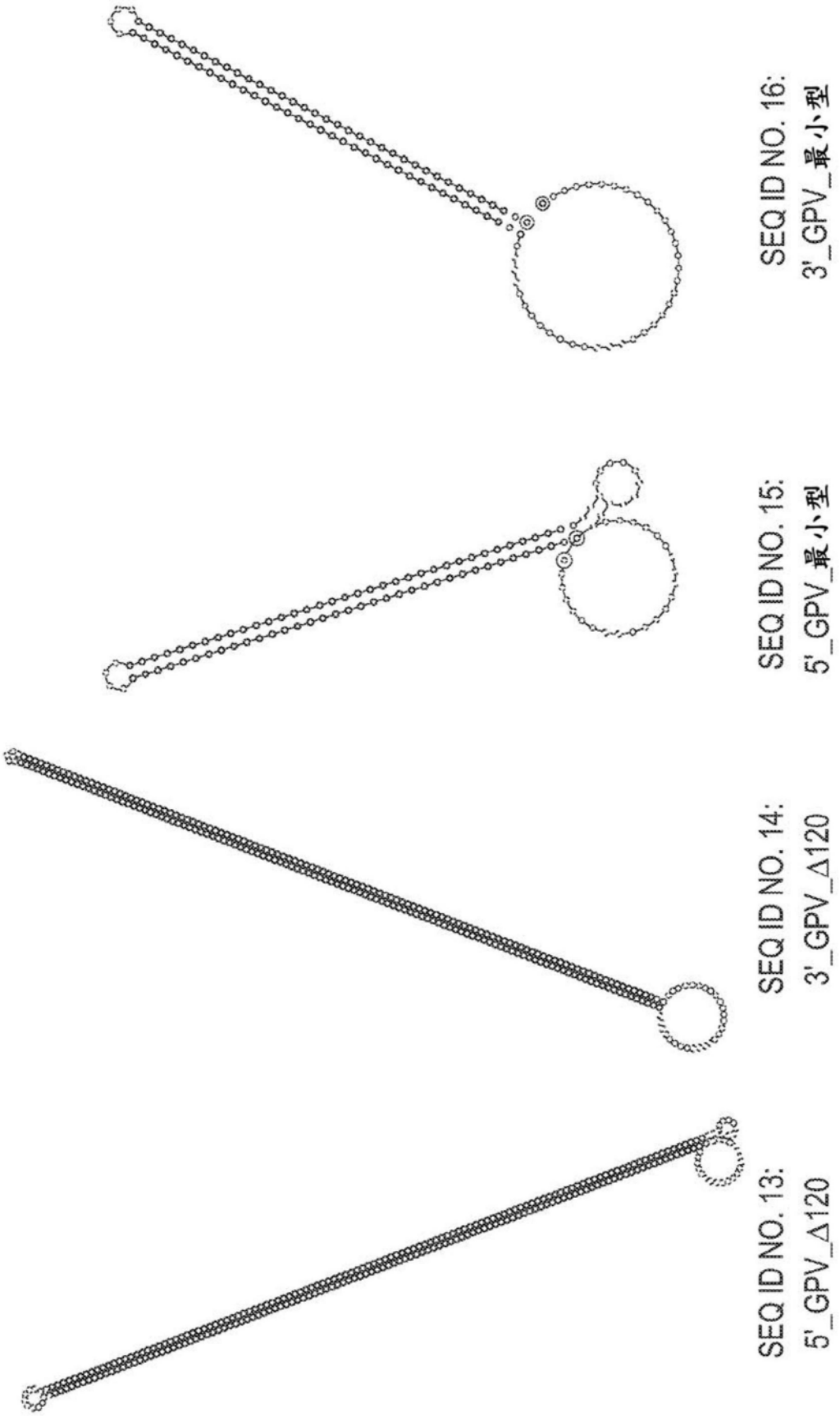

SEQ ID NO. 16:
3'_GPV_最小型
SEQ ID NO. 15:
5'_GPV_最小型
SEQ ID NO. 14:
3'_GPV_Δ120
SEQ ID NO. 13:
5'_GPV_Δ120

图7

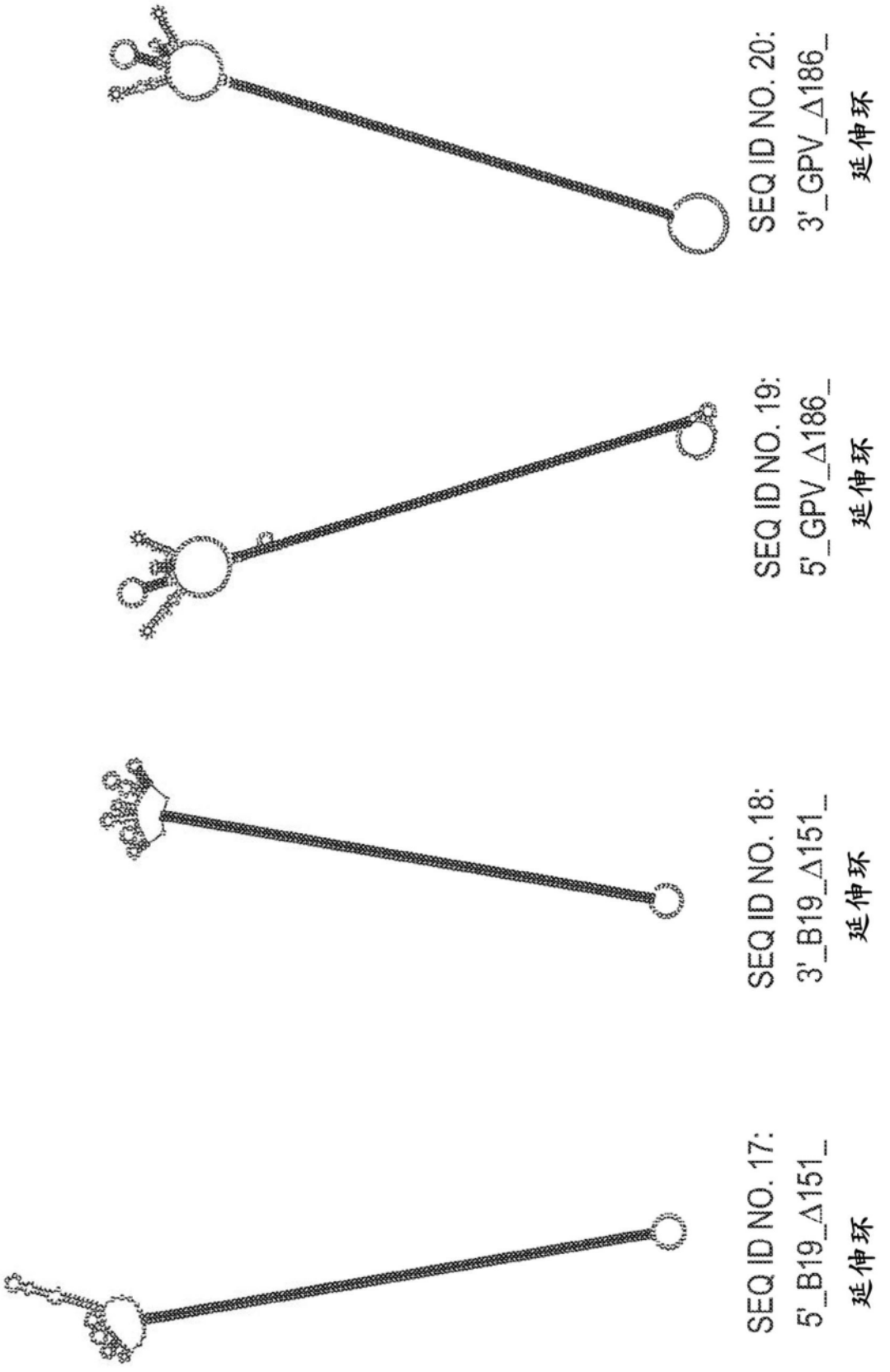

SEQ ID NO. 17:
5'_B19_Δ151_
延伸环
SEQ ID NO. 18:
3'_B19_Δ151_
延伸环
SEQ ID NO. 19:
5'_GPV_Δ186_
延伸环
SEQ ID NO. 20:
3'_GPV_Δ186_
延伸环

图8

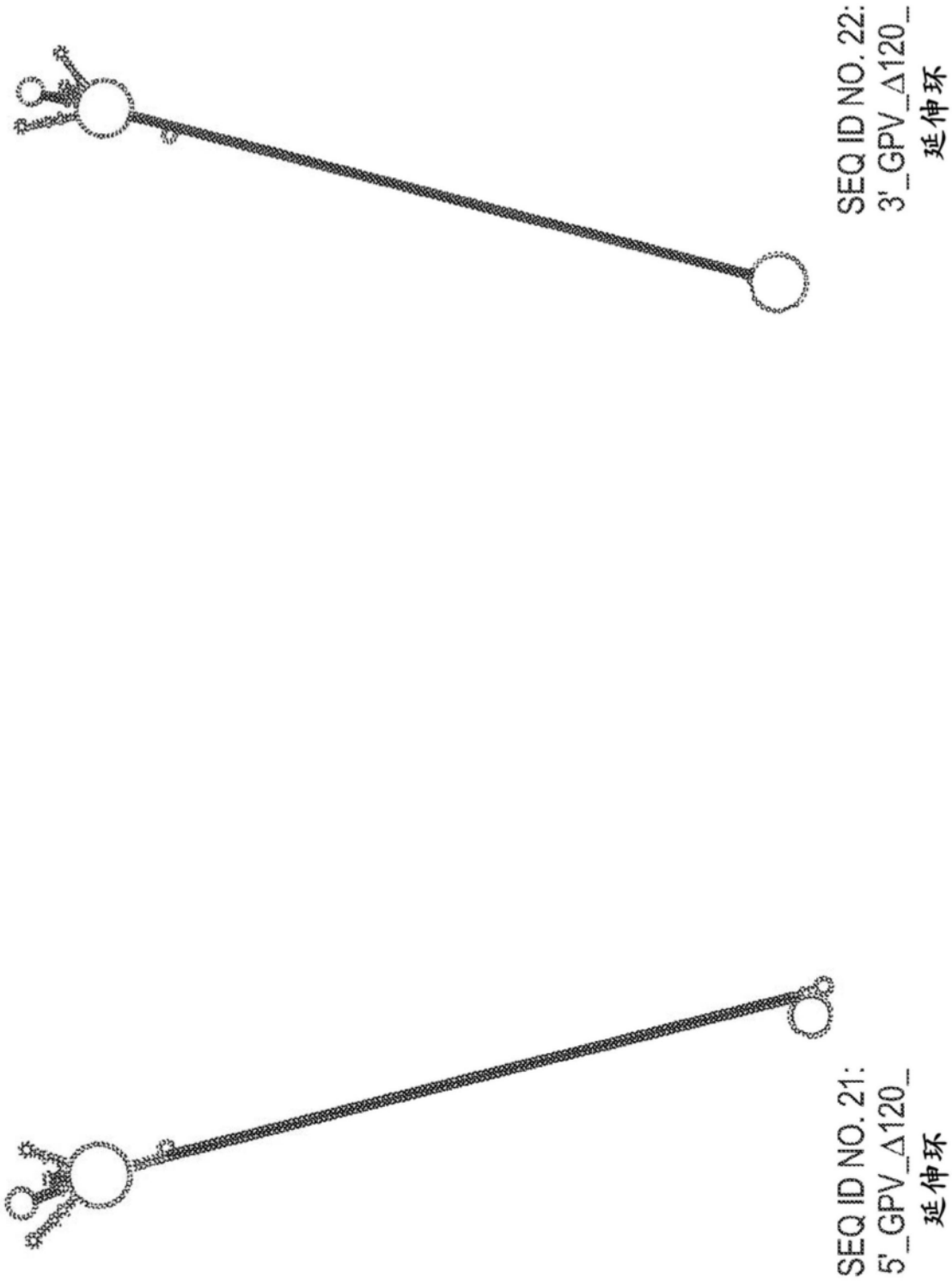


图9

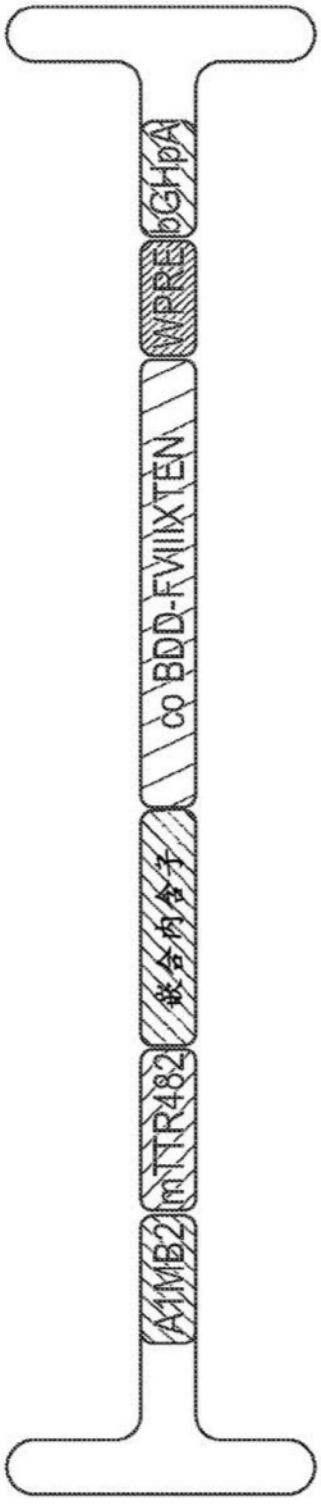


图10A

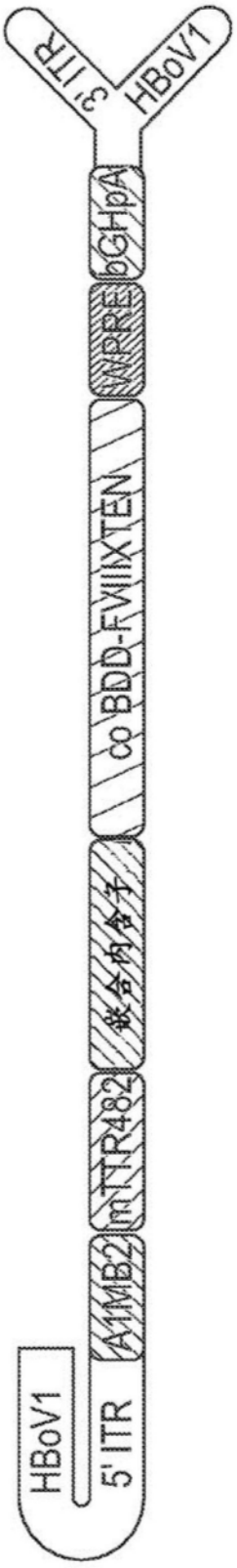

3' ITR
HBoV1
bGHpA
WPRE
co BDD-FVIIIXTEN
嵌合的含子
mTTR482
A1MB2
HBoV1
5' ITR

图10B

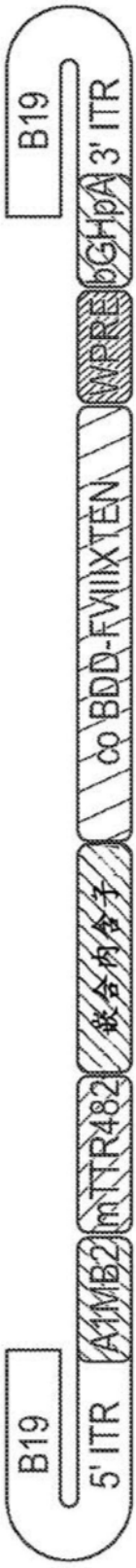

B19
3' ITR
bGHpA
WPRE
co BDD-FVIIIXTEN
嵌合内含子
mTTR482
A1MB2
5' ITR
B19

图10C

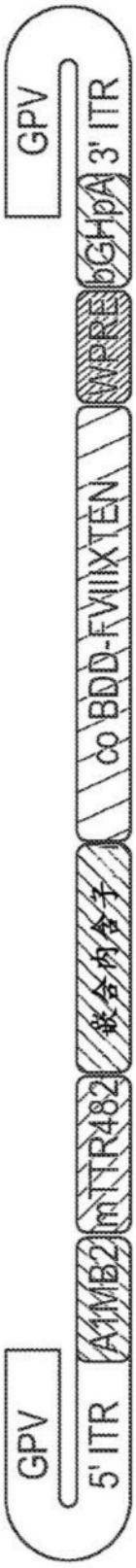

GPV
3' ITR
bGHpA
WPRE
Co BDD-FVIIIXTEN
嵌合内含子
mTTR482
A1MB2
5' ITR
GPV

图10D

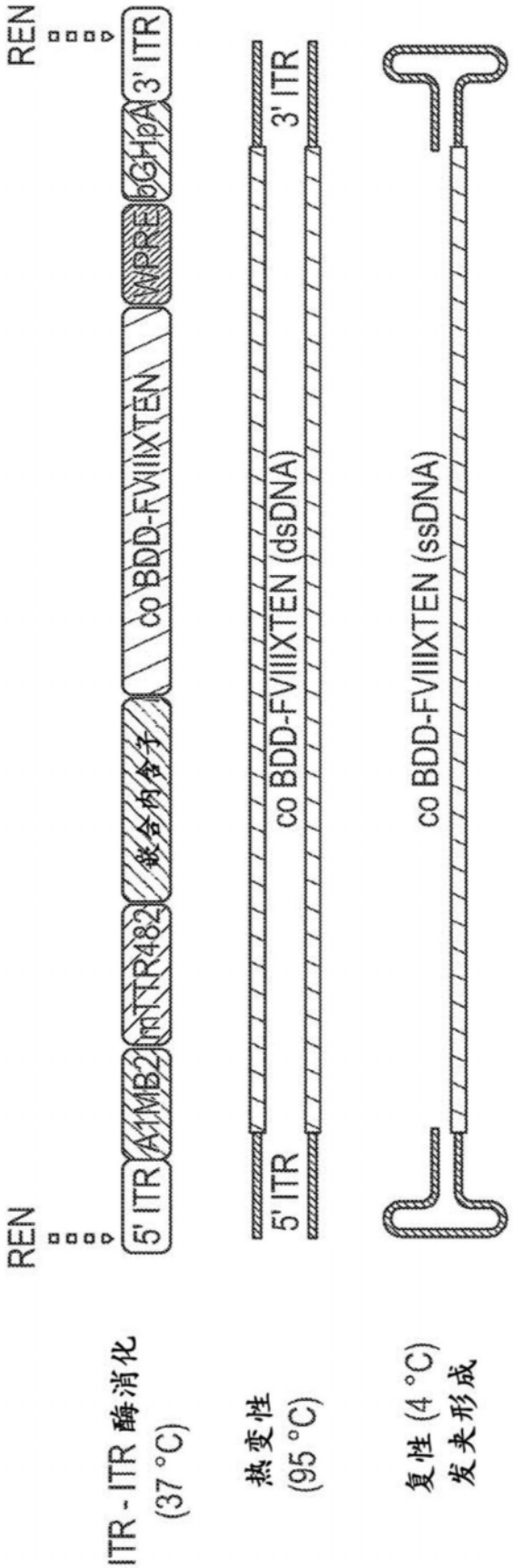

REN
5' ITR
A1MB2
mTTR482
嵌合内含子
co BDD-FVIIIXTEN
MPRE
bGHpA
3' ITR
ITR - ITR 酶消化
(37 °C)
热变性
(95 °C)
复性 (4 °C)
发夹形成
co BDD-FVIIIXTEN (dsDNA)
co BDD-FVIIIXTEN (ssDNA)

图11

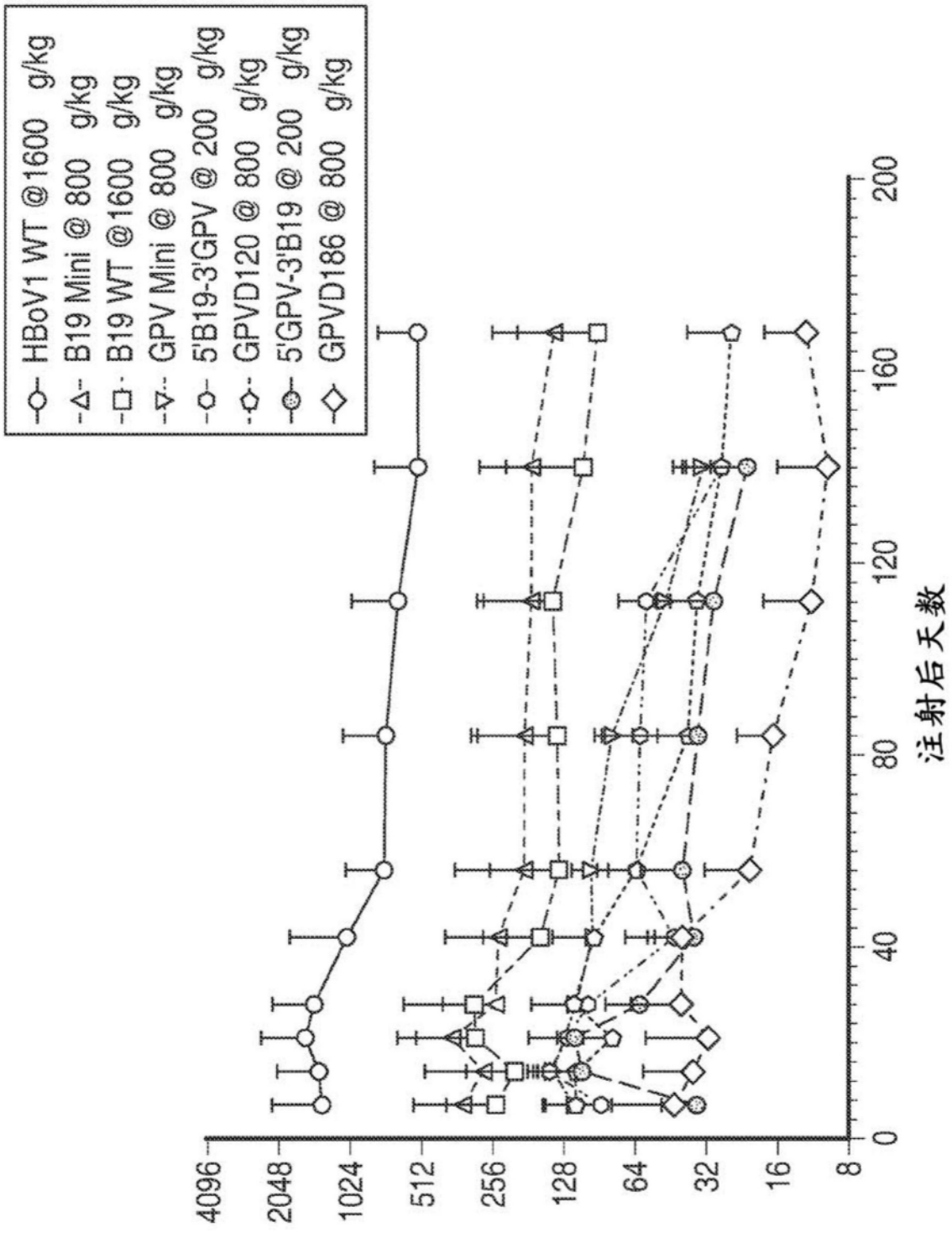

HBoV1 WT @1600 g/kg
B19 Mini @ 800 g/kg
B19 WT @1600 g/kg
GPV Mini @ 800 g/kg
5'B19-3'GPV @ 200 g/kg
GPVD120 @ 800 g/kg
5'GPV-3'B19 @ 200 g/kg
GPVD186 @ 800 g/kg
4096
2048
1024
512
256
128
64
32
16
8
0
40
80
120
160
200
注射后天数

图12

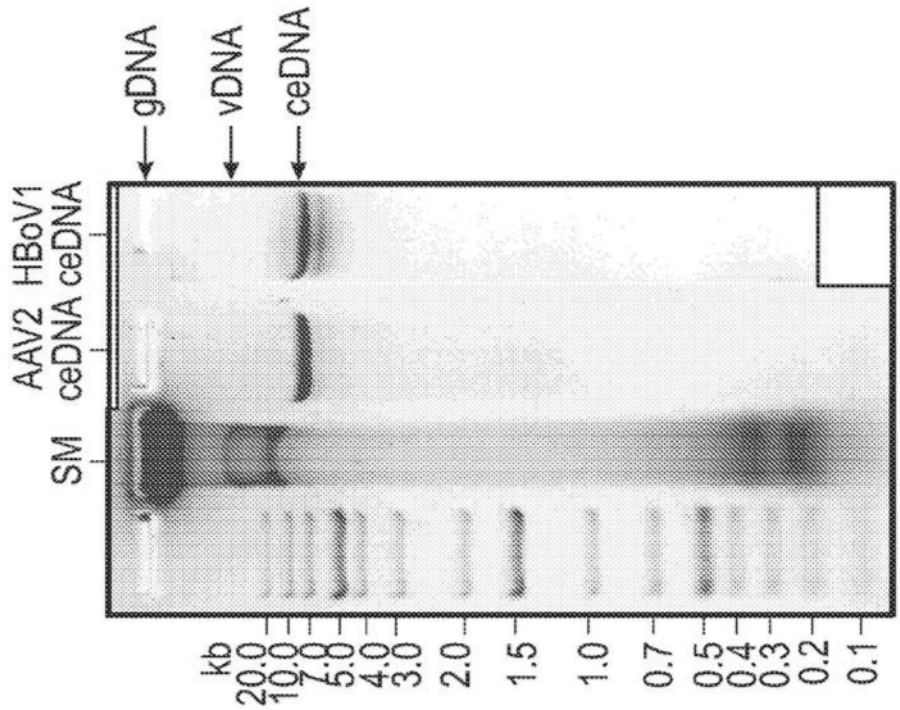

gDNA
vDNA
ceDNA
HBoV1 ceDNA
AAV2 ceDNA
SM
kb
20.0
10.0
7.8
5.8
4.0
3.0
2.0
1.5
1.0
0.7
0.5
0.4
0.3
0.2
0.1

图13A

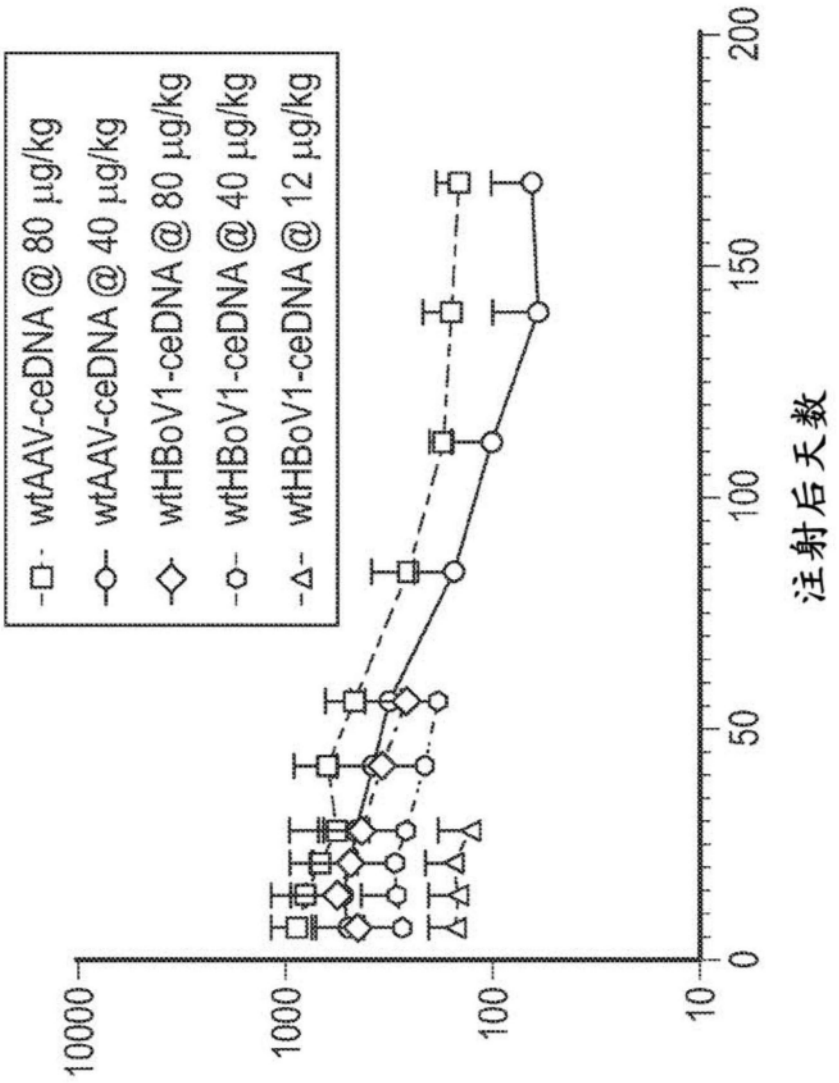

wtAAV-ceDNA @ 80 μg/kg
wtAAV-ceDNA @ 40 μg/kg
wtHBoV1-ceDNA @ 80 μg/kg
wtHBoV1-ceDNA @ 40 μg/kg
wtHBoV1-ceDNA @ 12 μg/kg
10000
1000
100
10
0
50
100
150
200
注射后天数

图13B