



(12) 发明专利申请

(10) 申请公布号 CN 113994008 A

(43) 申请公布日 2022. 01. 28

(21) 申请号 202080044921.1

(22) 申请日 2020.05.22

(30) 优先权数据

62/852,216 2019.05.23 US

(85) PCT国际申请进入国家阶段日

2021.12.20

(86) PCT国际申请的申请数据

PCT/US2020/034394 2020.05.22

(87) PCT国际申请的公布数据

WO2020/237219 EN 2020.11.26

(71) 申请人 太空飞船七有限责任公司

地址 美国新泽西

(72) 发明人 B·比尔德 D·里克斯

R·普拉巴卡尔

(74) 专利代理机构 中国贸促会专利商标事务所
有限公司 11038

代理人 刘晓东

(51) Int.Cl.

C12N 15/86 (2006.01)

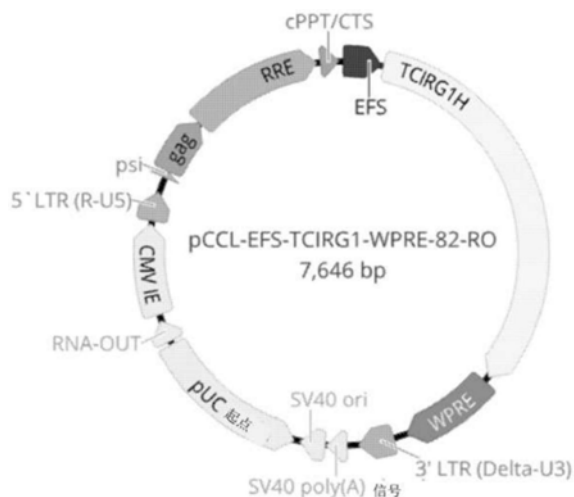
权利要求书4页 说明书22页
序列表26页 附图6页

(54) 发明名称

用于婴儿恶性骨硬化病的基因治疗载体

(57) 摘要

本公开提供了包含编码TCIRG1多肽或其功能性变体的多核苷酸序列的改进的基因治疗载体、其使用方法、药物组合物等。具体地,本公开提供用于治疗婴儿恶性骨硬化病(IMO)的慢病毒载体。



1. 一种转移质粒, 包含表达盒, 所述表达盒包含编码T细胞免疫调节因子1 (TCIRG1) 或其功能性变体的同源异构体的编码多核苷酸和启动子, 其中所述多核苷酸可操作地连接到所述启动子, 并且其中所述转移质粒包含RNA-OUT阻遏物和CMV IE启动子。

2. 根据权利要求1所述的转移质粒, 其中所述RNA-OUT阻遏物与SEQ ID NO:32具有至少95%的同一性或至少99%的同一性。

3. 根据权利要求1或权利要求2所述的转移质粒, 其中所述CMV IE启动子与SEQ ID NO:33具有至少95%的同一性或至少99%的同一性。

4. 根据权利要求1至3中任一项所述的转移质粒, 其中所述转移质粒包含pCCL骨架。

5. 根据权利要求4所述的转移质粒, 其中所述pCCL骨架包含所述RNA-OUT阻遏物。

6. 根据权利要求5所述的转移质粒, 其中所述转移质粒与SEQ ID NO:39具有至少95%或100%的同一性。

7. 根据权利要求1至7中任一项所述的转移质粒, 其中所述启动子是EFS启动子。

8. 根据权利要求7所述的转移质粒, 其中所述EFS启动子与SEQ ID NO:2具有至少95%的同一性。

9. 根据权利要求8所述的转移质粒, 其中所述EFS启动子是SEQ ID NO:2。

10. 根据权利要求1至9中任一项所述的转移质粒, 其中所述编码多核苷酸与SEQ ID NO:3具有至少95%的同一性。

11. 根据权利要求10所述的转移质粒, 其中所述编码多核苷酸与SEQ ID NO:3具有至少99%的同一性。

12. 根据权利要求11所述的转移质粒, 其中所述编码多核苷酸是SEQ ID NO:3。

13. 根据权利要求1至12中任一项所述的转移质粒, 其中所述表达盒包含土拨鼠肝炎病毒(WHP)转录后调控元件(WPRE)。

14. 根据权利要求13所述的转移质粒, 其中所述WPRE是SEQ ID NO:4。

15. 根据权利要求1至14中任一项所述的转移质粒, 其中所述表达盒与SEQ ID NO:1具有至少95%的同一性。

16. 根据权利要求1至15中任一项所述的转移质粒, 其中所述表达盒的两侧是5'长末端重复序列(LTR)和3'LTR。

17. 根据权利要求16所述的转移质粒, 其中所述5'LTR是SEQ ID NO:34和/或所述3'LTR是SEQ ID NO:28。

18. 根据权利要求1至17中任一项所述的转移质粒, 其中所述表达盒与SEQ ID NO:1具有至少95%的同一性。

19. 根据权利要求1至17中任一项所述的转移质粒, 其中所述表达盒是SEQ ID NO:1。

20. 一种慢病毒颗粒, 通过用权利要求1至20中任一项所述的转移质粒转染宿主细胞而产生。

21. 一种表达盒, 包含编码T细胞免疫调节因子1 (TCIRG1) 或其功能性变体的同源异构体的编码多核苷酸和EFS启动子, 其中所述多核苷酸可操作地连接到所述EFS启动子。

22. 根据权利要求1所述的表达盒, 其中所述编码多核苷酸与SEQ ID NO:3具有至少95%的同一性。

23. 根据权利要求2所述的表达盒, 其中所述编码多核苷酸与SEQ ID NO:3具有至少

99%的同一性。

24. 根据权利要求3所述的表达盒,其中所述编码多核苷酸是SEQ ID NO:3。

25. 根据权利要求1至4中任一项所述的表达盒,其中所述EFS启动子与SEQ ID NO:2具有至少95%的同一性。

26. 根据权利要求25所述的表达盒,其中所述EFS启动子是SEQ ID NO:2。

27. 根据权利要求21至26中任一项所述的表达盒,包含土拨鼠肝炎病毒(WHP)转录后调控元件(WPRE)。

28. 根据权利要求27所述的表达盒,其中所述WPRE是SEQ ID NO:4。

29. 根据权利要求21至28中任一项所述的表达盒,其中所述表达盒与SEQ ID NO:1具有至少95%的同一性。

30. 根据权利要求29所述的表达盒,其中所述表达盒是SEQ ID NO:1。

31. 一种重组慢病毒基因组,按5'到3'顺序包含:

- (a) 慢病毒5'长末端重复序列(LTR);
- (b) 权利要求21至30中任一项所述的表达盒;和
- (c) 慢病毒3'LTR,

其中所述重组慢病毒基因组不能复制。

32. 一种转移质粒,包含权利要求31所述的重组慢病毒基因组。

33. 一种慢病毒颗粒,包含权利要求31所述的重组慢病毒基因组。

34. 一种药物组合物,包含权利要求33所述的慢病毒颗粒。

35. 一种经修饰的细胞,包含权利要求21至30中任一项所述的表达盒。

36. 一种经修饰的细胞,包含权利要求31所述的重组慢病毒基因组。

37. 根据权利要求36所述的经修饰的细胞,其中所述经修饰的细胞缺乏内源性功能性TCIRG1基因。

38. 根据权利要求36或37所述的经修饰的细胞,其中所述经修饰的细胞源自患有或疑似患有婴儿恶性骨硬化病(IMO)的受试者。

39. 根据权利要求36至38中任一项所述的经修饰的细胞,其中所述经修饰的细胞表达TCIRG1或其功能性变体,表达水平类似于在具有功能性TCIRG1基因的破骨细胞中观察到的TCIRG1表达水平。

40. 根据权利要求36至39中任一项所述的经修饰的细胞,其中所述经修饰的细胞表达TCIRG1或其功能性变体,表达水平类似于在源自未患有或未疑似患有IMO的受试者的破骨细胞中观察到的TCIRG1表达水平。

41. 根据权利要求36至40中任一项所述的经修饰的细胞,其中所述经修饰的细胞是造血干细胞(HSC)。

42. 根据权利要求36至41中任一项所述的经修饰的细胞,其中所述经修饰的细胞是CD34+祖细胞。

43. 根据权利要求41或42中任一项所述的经修饰的细胞,其中所述经修饰的细胞源自通过单采术从患有或疑似患有IMO的受试者分离的HSC,任选在通过施用G-CSF、plerifaxor,或G-CSF和plerifaxor的组合动员HSC之后。

44. 根据权利要求35至43中任一项所述的经修饰的细胞,其中所述经修饰的细胞源自

通过磁捕获富集CD34+细胞的细胞群。

45. 一种药物组合物, 包含权利要求35至44中任一项所述的经修饰的细胞。

46. 一种修饰患有或疑似患有IMO的受试者的一种或多种细胞的体外方法, 包括:

(a) 通过向所述受试者施用包含G-CSF、plerifaxor或G-CSF和plerifaxor的组合物, 提供从所述受试者动员的外周血单核细胞 (PBMC);

(b) 通过磁分离富集CD34+细胞的所述PBMC以产生富集CD34的细胞群; 和

(c) 将所述富集CD34的细胞与包含重组慢病毒基因组的慢病毒颗粒接触, 所述重组慢病毒基因组按5'到3'顺序包含:

(i) 慢病毒5'长末端重复序列 (LTR);

(ii) 权利要求21至30中任一项所述的表达盒; 和

(iii) 慢病毒3'LTR,

其中所述重组慢病毒基因组不能复制。

47. 一种治疗患有或疑似患有婴儿恶性骨硬化病 (IMO) 受试者的IMO的方法, 包括向所述受试者施用权利要求35至44中任一项所述的经修饰的细胞或权利要求45所述的药物组合物。

48. 根据权利要求47所述的方法, 其中所述方法用表达TCIRG1或其功能性变体的经修饰的细胞重新填充HSC龛。

49. 根据权利要求47或48所述的方法, 其中所述方法用表达TCIRG1或其功能性变体的经修饰的细胞重新填充破骨细胞龛。

50. 根据权利要求47至49中任一项所述的方法, 其中所述方法治疗、改善、预防、减少、抑制或缓解IMO。

51. 根据权利要求47至50中任一项所述的方法, 其中所述方法将接受治疗的受试者的平均总生存期延长至少1、2、3、4、5、6、7、8、9、10或更多年。

52. 根据权利要求47至51中任一项所述的方法, 其中所述方法防止所述受试者因IMO死亡。

53. 根据权利要求47至52中任一项所述的方法, 其中所述受试者是人。

54. 根据权利要求47至53中任一项所述的方法, 其中所述受试者在治疗前表现出IMO的症状。

55. 根据权利要求47至54中任一项所述的方法, 其中所述受试者在治疗前被鉴定为TCIRG1表达降低或检测不到TCIRG1表达。

56. 根据权利要求47至55中任一项所述的方法, 其中所述受试者被鉴定为具有突变的TCIRG1基因。

57. 根据权利要求47至56中任一项所述的方法, 其中所述受试者是婴儿。

58. 根据权利要求47至57中任一项所述的方法, 其中所述方法包括自体治疗。

59. 根据权利要求47至58中任一项所述的方法, 其中通过静脉内输注进行施用。

60. 一种用于制备治疗或预防婴儿恶性骨硬化病 (IMO) 的药物的重组慢病毒基因组, 其中所述慢病毒基因组按5'至3'顺序包含:

(i) 慢病毒5'长末端重复序列 (LTR),

(ii) 权利要求21至30中任一项所述的表达盒, 和

(iii) 慢病毒3' LTR;和

其中所述重组慢病毒基因组不能复制。

61. 一种用于制备治疗或预防婴儿恶性骨硬化病 (IMO) 的药物的慢病毒颗粒, 包含重组慢病毒基因组,

其中所述慢病毒基因组按5'到3'的顺序包含:

(i) 慢病毒5'长末端重复序列 (LTR),

(ii) 权利要求21至50中任一项所述的表达盒, 和

(iii) 慢病毒3' LTR;和

其中所述重组慢病毒基因组不能复制。

62. 一种转移质粒, 包含权利要求24至30任一项所述的表达盒。

63. 根据权利要求62所述的转移质粒, 进一步包含RNA-OUT序列。

64. 根据权利要求63所述的转移质粒, 其中所述RNA-OUT序列是SEQ ID NO:22。

65. 根据权利要求62或63所述的转移质粒, 其中所述RNA-OUT序列被配置为使得所述转移质粒能够在包装细胞系中稳定增殖。

66. 一种产生慢病毒颗粒的方法, 包括: 用权利要求1-19或权利要求62-65中任一项所述的转移质粒转化细菌细胞, 使得所述转移质粒被复制; 分离所复制的转移质粒; 以及用所复制的转移质粒和任选的一种或多种另外的质粒转导包装细胞系, 从而产生所述慢病毒颗粒。

67. 一种产生慢病毒颗粒的方法, 包括: 用权利要求1-19或权利要求62-65中任一项所述的转移质粒和任选的一种或多种另外的质粒转染包装细胞系; 并培养所述包装细胞系。

68. 根据权利要求67所述的方法, 其中所述转移质粒是稳定增殖的。

69. 根据权利要求68所述的方法, 其中在30-37°C下使用摇瓶或发酵使所述转移质粒在细菌宿主中稳定增殖至少1、2、3、4、5、6或7天。

70. 一种慢病毒颗粒, 根据权利要求66-69中任一项所述的方法产生。

用于婴儿恶性骨硬化病的基因治疗载体

[0001] 相关申请的交叉引用

[0002] 本申请要求于2019年5月23日提交的美国临时申请序列号62/852,216的优先权,该申请的内容通过引用整体并入本文。

[0003] 关于序列表的声明

[0004] 与本申请相关的序列表以文本格式提供,而不是纸质副本,并在此通过引用并入说明书中。包含所述序列表的文本文件名称是ROPA_003_01WO_ST25.txt。所述文本文件大约62KB,创建于2020年5月22日,正在通过EFS-Web以电子方式提交。

技术领域

[0005] 本公开总体涉及与T细胞免疫调节因子1,即ATPase H⁺转运V0亚基a3基因(TCIRG1)突变相关疾病的基因治疗。具体地,本公开提供了包括编码TCIRG1蛋白(TCIRG1)的表达盒的基因治疗载体和质粒。

背景技术

[0006] 婴儿恶性骨硬化病(IMO)是一种罕见的隐性遗传病,其特征是由功能失调的破骨细胞造成的骨量增加。该病最常见的原因是T细胞免疫调节因子1,即ATPase H⁺转运V0亚基a3基因(TCIRG1)的突变。TCIRG1涉及破骨细胞的骨吸收能力。

[0007] 破骨细胞功能可以通过慢病毒载体介导的TCIRG1表达恢复。Moscatelli等人Bone 57:1-9 (2013)。进一步的研究表明,尽管由具有组成型生理启动子的慢病毒载体表达,但TCIRG1的慢病毒介导的表达以与内源性基因产物相同的方式受到调控。Thudium等人Calcif Tissue Int.99:638-648 (2016)。此外,他们确定,与所述基因的密码子优化的cDNA相比,即使其mRNA水平要高得多,而天然TCIRG1基因序列导致破骨细胞中蛋白质表达和功能性拯救的水平更高。此外,数据显示仅需要一小部分具有功能性TCIRG1的人类前破骨细胞即可显著增强体外再吸收功能,这可能是由于融合了再吸收破骨细胞和非再吸收破骨细胞,与之前oc/oc鼠骨硬化病模型的结果一致。从疗效和安全性的角度来看,这些发现对进一步开展骨硬化病的基因治疗,令人鼓舞。

[0008] 本领域仍然需要TCIRG1的基因治疗载体和使用这种载体的治疗方法。此外,需要产生此类基因治疗载体的可靠方法。本公开提供了此种基因治疗载体、其制造方法、其使用方法、药物组合物等。

发明内容

[0009] 本公开提供了包含编码TCIRG1多肽或其功能性变体的多核苷酸序列的改进的基因治疗载体、其使用方法、药物组合物等。

[0010] 一方面,本公开提供了包含表达盒的转移质粒,所述表达盒包含编码T细胞免疫调节因子1(TCIRG1)或其功能性变体的同源异构体的编码多核苷酸,以及启动子,其中所述多核苷酸可操作地连接至所述启动子,并且其中所述转移质粒包含RNA-OUT阻遏物和CMV IE

启动子。

[0011] 在一些实施方案中,所述RNA-OUT阻遏物与SEQ ID NO:32具有至少95%的同源性或至少99%的同源性。

[0012] 在一些实施方案中,所述CMV IE启动子与SEQ ID NO:33具有至少95%的同源性或至少99%的同源性。

[0013] 在一些实施方案中,所述转移质粒包含pCCL骨架。

[0014] 在一些实施方案中,所述pCCL骨架包含所述RNA-OUT阻遏物。

[0015] 在一些实施方案中,所述转移质粒与SEQ ID NO:39具有至少95%的同源性。

[0016] 在一些实施方案中,所述转移质粒包含SEQ ID NO:39。

[0017] 在一些实施方案中,所述启动子为EFS启动子。

[0018] 在一些实施方案中,所述EFS启动子与SEQ ID NO:2具有至少95%的同源性。

[0019] 在一些实施方案中,所述EFS启动子为SEQ ID NO:2。

[0020] 在一些实施方案中,所述编码多核苷酸与SEQ ID NO:3具有至少95%的同源性。

[0021] 在一些实施方案中,所述编码多核苷酸与SEQ ID NO:3具有至少99%的同源性。

[0022] 在一些实施方案中,所述编码多核苷酸为SEQ ID NO:3。

[0023] 在一些实施方案中,所述表达盒包含土拨鼠肝炎病毒(WHP)转录后调控元件(WPRE)。

[0024] 在一些实施方案中,所述WPRE是SEQ ID NO:4。

[0025] 在一些实施方案中,所述表达盒与SEQ ID NO:1具有至少95%的同源性。

[0026] 在一些实施方案中,所述表达盒的两侧是5'长末端重复序列(LTR)和3'LTR。

[0027] 在一些实施方案中,所述5'LTR是SEQ ID NO:34和/或所述3'LTR是SEQ ID NO:28。

[0028] 在一些实施方案中,所述表达盒与SEQ ID NO:1具有至少95%的同源性。

[0029] 在一些实施方案中,所述表达盒是SEQ ID NO:1。

[0030] 另一方面,本公开提供了一种表达盒,包含编码T细胞免疫调节因子1(TCIRG1)或其功能性变体的同源异构体的多核苷酸,和EFS启动子,其中可选地所述多核苷酸可操作地连接到所述EFS启动子。

[0031] 在一些实施方案中,所述编码多核苷酸与SEQ ID NO:3具有至少95%的同源性。在一些实施方案中,所述编码多核苷酸与SEQ ID NO:3具有至少99%的同源性。在一些实施方案中,所述编码多核苷酸为SEQ ID NO:3。在一些实施方案中,所述EFS启动子与SEQ ID NO:2具有至少95%的同源性。在一些实施方案中,所述EFS启动子为SEQ ID NO:2。

[0032] 在一些实施方案中,所述表达盒包含土拨鼠肝炎病毒(WHP)转录后调控元件(WPRE)。在一些实施方案中,所述WPRE是SEQ ID NO:4。

[0033] 在一些实施方案中,所述表达盒与SEQ ID NO:1具有至少95%的同源性。在一些实施方案中,所述表达盒为SEQ ID NO:1。

[0034] 另一方面,本公开提供了一种重组慢病毒基因组,包括按5'到3'的顺序,慢病毒5'长末端重复序列(LTR);本文公开的表达盒;和慢病毒3'LTR,其中所述重组慢病毒基因组不能复制。

[0035] 另一方面,本公开提供了一种慢病毒颗粒,包含此种重组慢病毒基因组。

[0036] 另一方面,本公开提供了一种转移质粒,包含此种重组慢病毒基因组。在某些实施

方案中,所述转移质粒包含RNA-OUT序列。在一些实施方案中,所述RNA-OUT序列为SEQ ID NO:22。在一些实施方案中,所述RNA-OUT序列被配置为使得所述转移质粒能够在包装细胞系中稳定增殖。

[0037] 在具体实施方案中,所述转移质粒不包含抗生素抗性基因或不包含氨苄青霉素抗性基因,例如AmpR。

[0038] 在具体实施方案中,所述转移质粒包含SEQ ID NO:23所示的序列。

[0039] 在另一方面,本公开提供了一种产生慢病毒颗粒的方法,包括:用本公开的任何转移质粒以及任选的一种或多种另外质粒转染细胞包装系;以及培养所述包装细胞系。在一些实施方案中,在30-37°C下使用摇瓶或发酵使所述转移质粒在细菌宿主中稳定增殖至少1、2、3、4、5、6或7天。

[0040] 在相关方面,本公开提供了使用本文公开的转移质粒产生的慢病毒颗粒。

[0041] 另一方面,本公开提供了一种药物组合物,包含本公开的任何慢病毒颗粒。

[0042] 另一方面,本公开提供了包含本公开的任何表达盒的经修饰的细胞。

[0043] 另一方面,本公开提供了一种经修饰的细胞,包含本公开的任何重组慢病毒基因组。

[0044] 在一些实施方案中,所述经修饰的细胞缺乏内源性功能性TCIRG1基因。

[0045] 在一些实施方案中,所述经修饰的细胞源自患有或疑似患有婴儿恶性骨硬化病(IMO)的受试者。

[0046] 在一些实施方案中,所述经修饰的细胞表达TCIRG1或其功能性变体,其表达水平类似于在具有功能性TCIRG1基因的破骨细胞中观察到的TCIRG1的表达水平。

[0047] 在一些实施方案中,所述经修饰的细胞表达TCIRG1或其功能性变体,其表达水平类似于在源自未患有或未疑似患有IMO受试者的破骨细胞中观察到的TCIRG1的表达水平。

[0048] 在一些实施方案中,所述经修饰的细胞是造血干细胞(HSC)。

[0049] 在一些实施方案中,所述经修饰的细胞是CD34+祖细胞。

[0050] 在一些实施方案中,所述经修饰的细胞源自通过单采术从患有或疑似患有IMO的受试者中分离出来的HSC。

[0051] 在一些实施方案中,所述经修饰的细胞源自通过施用G-CSF、plerifaxor,或G-CSF和plerifaxor的组合动员HSC之后,通过单采术从患有或疑似患有IMO的受试者分离的HSC。

[0052] 在一些实施方案中,所述经修饰的细胞源自通过磁捕获富集CD34+细胞的细胞群。

[0053] 另一方面,本公开提供了一种药物组合物,包含本公开的任何经修饰的细胞。

[0054] 另一方面,本公开提供了一种修饰患有或疑似患有IMO受试者的一种或多种细胞的体外方法,包括:通过向所述受试者施用包含G-CSF、plerifaxor或G-CSF和plerifaxor组合的组合物,提供从所述受试者动员的外周血单核细胞(PBMC);通过磁分离富集CD34+细胞的所述PBMC以产生富集CD34的细胞群;以及使所述富集CD34的细胞与包含重组慢病毒基因组的慢病毒颗粒接触,所述重组慢病毒基因组包含按5'到3'的顺序:慢病毒5'长末端重复序列(LTR);本公开的任何表达盒;和慢病毒3'LTR,其中所述重组慢病毒基因组不能复制。

[0055] 另一方面,本公开提供了一种治疗患有或疑似患有婴儿恶性骨硬化病(IMO)受试者的IMO的方法,包括施用本公开的任何经修饰的细胞或本公开的任何药物组合物。

- [0056] 在一些实施方案中,所述方法用表达TCIRG1或其功能性变体的经修饰的细胞重新填充HSC龕。
- [0057] 在一些实施方案中,所述方法用表达TCIRG1或其功能性变体的经修饰的细胞重新填充破骨细胞龕。
- [0058] 在一些实施方案中,所述方法治疗、改善、预防、减少、抑制或缓解IMO。
- [0059] 在一些实施方案中,所述方法将接受治疗的受试者的平均总生存期延长至少1、2、3、4、5、6、7、8、9、10年或更长。
- [0060] 在一些实施方案中,所述方法防止所述受试者因IMO死亡。
- [0061] 在一些实施方案中,所述受试者是人。
- [0062] 在一些实施方案中,所述受试者在治疗前表现出IMO的症状。
- [0063] 在一些实施方案中,所述受试者在治疗前被鉴定为TCIRG1表达降低或检测不到TCIRG1表达。
- [0064] 在一些实施方案中,所述受试者被鉴定为具有突变的TCIRG1基因。
- [0065] 在一些实施方案中,所述受试者是婴儿。
- [0066] 在一些实施方案中,所述方法包括自体治疗。
- [0067] 在一些实施方案中,所述施用是通过静脉输注进行的。
- [0068] 另一方面,本公开提供了一种用于制备治疗或预防婴儿恶性骨硬化病(IMO)的药物的重组慢病毒基因组,其中所述慢病毒基因组按5'至3'顺序包含:慢病毒5'长末端重复序列(LTR)、本公开的任何表达盒和慢病毒3'LTR;并且其中所述重组慢病毒基因组不能复制。
- [0069] 另一方面,本公开提供了一种用于制备治疗或预防婴儿恶性骨硬化病(IMO)的药物的慢病毒颗粒,包含重组慢病毒基因组,其中所述慢病毒基因组按5'至3'顺序包含:慢病毒5'长末端重复序列(LTR)、本公开的任何表达盒和慢病毒3'LTR;并且其中所述重组慢病毒基因组不能复制。
- [0070] 本发明的其他特征和优点将从以下详细描述和权利要求中显而易见且被其涵盖。

附图说明

- [0071] 图1提供了用于产生编码TCIRG1(pCCL.PPT.EFS.tcirg1h.wpre)的慢病毒基因治疗载体的转移质粒的示意图。
- [0072] 图2显示了表达盒(SEQ ID NO:1)的基因序列,按5'到3'的顺序包括延伸因子1-短(EFS)启动子(有下划线;SEQ ID NO:2)、编码TCIRG1的多核苷酸(黑底白字;SEQ ID NO:3)以及土拨鼠肝炎病毒(WHP)转录后调控元件(WPRE)(有下划线且加粗;SEQ ID NO:4)。
- [0073] 图3A-3B提供了两种不同慢病毒质粒稳定性的比较。图3A显示用溴化乙锭染色的琼脂糖凝胶照片,显示未用限制酶酶切(“未切割”)或用AflIII (“AflIII”)或AflIII和NarI酶切(“AflIII/NarI”)的质粒pRRL.PPT.EFS.tcirg1h.wpre。图3B显示用溴化乙锭染色的琼脂糖凝胶照片,显示未用限制酶酶切(“未切割”)或用AflIII (“AflIII”)或AflIII和NarI酶切(“AflIII/NarI”)的质粒pCCL.PPT.EFS.tcirg1h.wpre。图3C显示了所述pRRL.PPT.EFS.tcirg1h.wpre和所述pCCL.PPT.EFS.tcirg1h.wpre质粒的示意图。
- [0074] 图4描述了慢病毒颗粒制造的示例性过程。

[0075] 图5A-5B显示了转导后大量CD34⁺细胞液体培养(图5A)6天和(图5B)12天的载体拷贝数(VCN)。在SCGM完全培养基中培养转导的CD34⁺细胞后,通过对提取的gDNA进行qPCR评估VCN。每个供体的VCN和平均值代表每种转导条件。

具体实施方式

[0076] 本发明人已经证实,移植用编码TCIRG1的慢病毒载体转导的自体细胞可有效治疗婴儿恶性骨硬化病(IMO)。此外,在编码TCIRG1的基因治疗载体的表达盒序列中包含特定的序列元件,使对IMO进行安全有效的基因治疗。本公开提供了编码TCIRG1的慢病毒载体和质粒,包括有利于产生所述慢病毒载体的稳定转移质粒。

[0077] 载体和质粒

[0078] 发明人惊奇地发现,通过以下两种方式修饰含有所需表达盒的pRRL质粒,可以改进大规模产生用于TCIRG1基因治疗的慢病毒载体:(i)用pCCL载体骨架替换所述pRRL载体骨架,以及(ii)用RNA-OUT可选择标记替换所述pCCL骨架中的常规抗生素抗性盒。然后将所改进的质粒连同辅助质粒一起转染慢病毒颗粒产生系统,以产生出所需的慢病毒载体。

[0079] 用于TCIRG1基因治疗的所得pCCL/RNA-OUT载体(例如,图1中描述的载体)稳定性有所改进,这反映在基于大肠杆菌的质粒生产中质粒产量更高并且纯化质粒中不利重组产物的水平降低(如实施例1和图3A-3C所示)。所述转移质粒的这种改进使包含所述TCIRG1表达盒的慢病毒颗粒制造的产量足以进行临床测试和使用。本文提供的进一步数据证实,使用本文公开的所述方法和组合物产生的慢病毒颗粒足够有效地转导CD34⁺细胞以达到临床相关的载体拷贝数(VCN)水平。

[0080] 在一些实施方案中,本公开提供了一种转移质粒,所述转移质粒是基于用于第三代慢病毒载体系统的pCCL转移质粒的慢病毒载体。所述pCCL转移质粒包含嵌合巨细胞病毒(CMV)-HIV 5'LTR和载体骨架,其中猿猴病毒40聚腺苷酸化和(无增强子)复制起点已包含在HIV 3'LTR的下游,从而替换从HIV整合位点剩余的大部分人类序列。CCL 5'杂交长末端重复序列(LTR)是与HIV-1LTR的R区相连的巨细胞病毒(CMV)的增强子和启动子(相对于转录起始位点的核苷酸-673至-1;GenBank登录号K03104)。在一些实施方案中,所述转移质粒包括连接到所述TCIRG1基因的EFS启动子,上游是RRE元件和cPPT/CTS元件并且下游是WPRE元件(图1)。在一些实施方案中,所述转移质粒包括连接到所述TCIRG1基因的PGK启动子(SEQ ID NO:24),上游是RRE元件和cPPT/CTS元件并且下游是WPRE元件。在一些实施方案中,所述转移质粒包括RNA-OUT元件。有利地,所述RNA-OUT序列有助于所述转移质粒在包装细胞系中稳定增殖。在一些实施方案中,所述转移质粒不包括抗生素抗性基因,例如AmpR。

[0081] 在一些实施方案中,所述PGK启动子包含与SEQ ID NO:24具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述PGK启动子包含与SEQ ID NO:24具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述PGK启动子具有序列SEQ ID NO:24。

[0082] GGGGTTGGGTTGCGCCTTTTCCAAGGCAGCCCTGGGTTTGCAGGGACGCGGCTGCTCTGGGCGTGGTTCCGGGAAACGCAGCGGCGCCGACCCTGGGTCTCGCACATTCTTACGTCCGTTTCGAGCGTCACCCGGATCTTCGCGCTACCCTTGTGGGCCCCCGGCGACGCTTCTGCTCCGCCCTAAGTCGGGAAGGTTCTTGGGTTTCGCGCGC

TGCCGGACGTGACAAACGGAAGCCGCACGTCTCACTAGTACCCTCGCAGACGGACAGCGCCAGGGAGCAATGGCAGC
GCGCCGACCGCGATGGGCTGTGGCCAATAGCGGCTGCTCAGCAGGGCGCGCCGAGAGCAGCGGCCGGAAGGGGCGG
TGCGGGAGGCGGGGTGTGGGGCGGTAGTGTGGGCCCTGTTCCCTGCCGCGGGTGTTCGCATTCTGCAAGCCTCCG
GAGCGCACGTCCGCGATCGGCTCCCTCGTTGACCGAATCACCGACCTCTCTCCCCAG

[0083] (SEQ ID NO:24)

[0084] 在某些实施方案中,所述转移质粒在大肠杆菌中培养或增殖时比包含相同表达盒的另一种质粒更稳定,从而使得质粒产量更高,这是利于产生载体。在一些实施方案中,所述转移质粒比所述pRRL.PPT.EFS.tcirg1h.wpre转移质粒更稳定。在具体的实施方案中,在相同培养条件下与pRRL.PPT.EFS.tcirg1h.wpre转移质粒产生量相比,产生了至少2倍、至少5倍或至少10倍以上的所述转移质粒。

[0085] 在具体实施方案中,所述转移质粒是pCCL.PPT.EFS.tcirg1h.wpre或其功能性变体,例如本文公开的转移质粒。本公开在具体实施方案中提供了所述转移质粒pCCL.PPT.EFS.tcirg1h.wpre或其功能性变体。所述转移质粒pCCL.PPT.EFS.tcirg1h.wpre可以具有序列SEQ ID NO:23。

[0086] 或者,所述转移质粒pCCL.PPT.EFS.tcirg1h.wpre可能具有序列SEQ ID NO:25,其中删除了序列GATCACGAGACTAGCCTCGAGAAGCTTGATCGATTGGCTCCGGTGCC (SEQ ID NO:26)。

[0087] 序列SEQ ID NO:25代表环状质粒。在碱基对1处以EFS启动子开始排列的相同序列设置为SEQ ID NO:27。

[0088] 在一些实施方案中,所述转移质粒包含与SEQ ID NO:23具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述转移质粒包含与SEQ ID NO:23具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述转移质粒具有序列SEQ ID NO:23。

[0089] 在一些实施方案中,所述转移质粒包含与SEQ ID NO:25具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述转移质粒包含与SEQ ID NO:25具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述转移质粒具有序列SEQ ID NO:25。

[0090] 在一些实施方案中,所述转移质粒包含与SEQ ID NO:27具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述转移质粒包含与SEQ ID NO:27具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述转移质粒具有序列SEQ ID NO:27。

[0091] 在一些实施方案中,所述转移质粒包含表1中所列的一种或多种所述载体元件。

[0092] 表1:pCCL.PPT.EFS.tcirg1h.wpre载体元件

名称	类型	最小-最大 参考序列中的位置 (核苷酸)	长度 (碱基 对)	SEQ ID NO:
EFS	启动子	1-243	243	SEQ ID NO: 2
TCIRG1	CDS	257-2,749	2,493	SEQ ID NO: 3
[0093] WPRE	调控的	2,782-3,384	603	SEQ ID NO: 4
3`LTR	LTR	3,471-3,704	234	SEQ ID NO: 28
SV40 poly(A)	polyA 信号	3,776-3,907	132	SEQ ID NO: 29
SV40 ori	复制起点	3,917-4,076	160	SEQ ID NO: 30
pUC 起点	复制起点	4,115-5,129	1,015	SEQ ID NO: 31
RNA-OUT	阻遏物	5,146-5,284	139	SEQ ID NO: 32
CMV IE	启动子	5,334-5,910	577	SEQ ID NO: 33
5-LTR	LTR	5,933-6,120	188	SEQ ID NO: 34
psi	包装	6,222-6,266	45	SEQ ID NO: 35
[0094] gag	CDS	6,267-6,628	362	SEQ ID NO: 36
RRE	调控的	6,629-7,486	858	SEQ ID NO: 37
cPPT/CTS	多嘌呤区	7,505-7,622	118	SEQ ID NO: 38
-	骨架	3,776-7,622	3,847	SEQ ID NO: 39

[0095] 在一些实施方案中,通过瞬时转染包括pCCL.PPT.EFS.tcirg1h.wpre的第三代慢病毒载体系统产生慢病毒颗粒。

[0096] 在一些实施方案中,所述表达盒包含与SEQ ID NO:1具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述表达盒包含与SEQ ID NO:1具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述表达盒具有序列SEQ ID NO:1。

[0097] 在一些实施方案中,所述表达盒按5'到3'的顺序包含EFS启动子、编码T细胞免疫调节因子1(ATPase H⁺运输V0亚基a3(TCIRG1))或其功能性变体的多核苷酸,以及土拨鼠肝炎病毒(WHP)转录后调节元件(WPRE)。在一些实施方案中,所述EFS启动子可操作连接到编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸。相关实施方案包含转移质粒(包含所述表达盒)和使用所述转移质粒产生的载体。

[0098] 在一些实施方案中,所述EFS启动子包含与SEQ ID NO:2具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述EFS启动子包含与SEQ ID NO:2具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述EFS启动子具有序列SEQ ID NO:2。

[0099] GGCTCCGGTGCCCGTCAGTGGCAGAGCGCACATCGCCACAGTCCCCGAGAAGTTGGGGGAGGGGTC
GGCAATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGTAACTGGGAAAGTGATGTCGTACTGGCTCCGCCTTTT
TCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTCGCAACGGGTTTGCCGCCA
GAACACAGGTGTCGTGACGC

[0100] (SEQ ID NO:2)

[0101] 在一些实施方案中,编码TCIRG1或其功能性变体的同源异构体所述的多核苷酸包含与SEQ ID NO:3具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,编码TCIRG1或其功能性变体的同源异构体所述多核苷酸包含与SEQ ID NO:3具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸具有序列SEQ ID NO:3。

[0102] 在一些实施方案中,所述WPRE包含与SEQ ID NO:4具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述WPRE包含与SEQ ID NO:4具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述WPRE具有序列SEQ ID NO:4。

[0103] ATTCGAGCATCTTACCGCCATTTATACCCATATTTGTTCTGTTTTCTTGATTTGGGTATACATTTAAATGTTAATAAAAACAAAATGGTGGGGCAATCATTACATTTTTAGGGATATGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAACTTTCCCGTTATTTACGCTCTGTTCCCTGTTAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGATATTCTTAACATATGTTGCTCCTTTTACGCTGTGTGGATATGCTGCTTAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTTCGTTTTCTCCTCCTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCAACGTGGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCACTGGCTGGGGCATTGCCACCACCTGTCAACTCCTTTCTGGGACTTTTCGCTTTCCCCCTCCCGATCGCCACGGCAGAAGCTCATCGCCGCCTGCCTTGC CCGCTGCTGGACAGGGGCTAGGTTGCTGGGCACTGATAATCCGTGGTGTGTCGGGGAAGCTGACGTCCTTTTCG

[0104] (SEQ ID NO:4)

[0105] 在一些实施方案中,TCIRG1或其功能性变体的所述同源异构体包含与SEQ ID NO:5具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多肽。在一些实施方案中,TCIRG1或其功能性变体的所述同源异构体包含与SEQ ID NO:5具有至少80%、85%、90%、95%、99%或100%同一性的多肽。在一些实施方案中,TCIRG1或其功能性变体的所述同源异构体具有序列SEQ ID NO:5。

[0106] MGS MFRSEEVALVQLFLPTAAAYTCVSR LGELGLVEFRDLNASVSAFQRRFVVDVRRCEELEKTFTFLQEEVRRAGLVLPKGRPLPAPPRDLLRIQEETERLAQELRDVRGNQQALRAQLHQLQLHAAVLRQGHEPQLAAAHDTGASERTPLLQAPGGPHQDLRVNFVAGAVEPHKAPALERLLWRACRGFLIASFRELEQPLEHPVTGEPATWMTFLISY WGEIQGQKIRKIDCFHCHVFPFLQQEEARLGALQQQLQQSQELQEV LGETERFLSQVLGRVLQLLPPGQVQVHKMKAVYLALNQCSVSTTHKCLIAEAWCSVRDLPALQEALRDSSMEEGVSAVAHRIPCRDMPPTLIRTNRFTASFQGI VDAYVGRYQEVNPAPYTIITFPFLFAVMFGDVGHLLMFLFALAMVLAENRPAVKAAQNEIWQTFFRGRYLLLLMGLFS IYTGFIYNECFSRATSI FPSGWSVAAMANQSGWSDAFLAQHTMLTLDPNVTGVFLGPYPFGIDPIWLSLAANHLSFLN SFKMKMSVILGVVHMAFGVV LGVFNHVHFGQRHRLLETLPELTFLLGLFGYLVFLVIYKWL CVWAARAASAPSIL I HFINMFLFSHSPSNRLLYPRQEVVQATLVVLALAMVPIILLGTPLHLLHRHRRRLRRRPADRQEENKAGLLDLPDAS VNGWSSDEEKAGGLDDEEEAE LVPSEVLMHQAIHTIEFCLGCVSNTASYLRLWALSLAHAQLSEVLWAMVMRIGLGL GREVGVAAVVLVPIFAAFVMTVAILLVMEGLSAFLHALRLHWVEFQNKFYSGTGYKLSPTFAATDD

[0107] (SEQ ID NO:5)

[0108] 在一个实施方案中,编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸是密码子优化的以在人宿主细胞中表达。在一个实施方案中,编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸被修饰,或“密码子优化”,从而通过用更频繁表示的密码子替换不常表示的密码子来增强表达。在一个实施方案中,编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸不是密码子优化的。在一个实施方案中,编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸未被修饰。在一个实施方案中,编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸不是密码子优化的。在一个实施方案中,编码TCIRG1或其功能性变体的同源异构体的所述多核苷酸是天然多核苷酸序列。

[0109] 如本文所用的术语“转基因”是指编码TCIRG1或其功能性变体的同源异构体的多核苷酸。

[0110] 编码序列是编码氨基酸以进行翻译的mRNA序列的部分。在翻译期间,61个三核苷酸密码子中的每一个被翻译成20个氨基酸中的一个,导致遗传密码的简并或冗余。然而,不同的细胞类型和不同的动物物种利用以不同频率编码相同氨基酸的tRNA(每个都带有反密码子)。当基因序列含有很少由相应tRNA表示的密码子时,核糖体翻译机制可能会减慢,从而阻碍有效的翻译。可以针对特定物种通过“密码子优化”来改进表达,其中所述编码序列被改变以编码相同的蛋白质序列,但使用高度表示和/或被高度表达的人蛋白质使用的密码子(Cid-Arregui等人,2003;J.Virol.77:4928)。

[0111] 在一些实施方案中,所述转基因的编码序列被修饰以用在灵长类动物中频繁表达的密码子替换在哺乳动物或灵长类动物中不常表达的密码子。例如,在一些实施方案中,所述转基因编码与参考多肽(例如野生型TCIRG1;SEQ ID NO:3)具有至少85%序列同一性的多肽,例如与所述参考多肽具有至少90%序列同一性、至少95%序列同一性、至少98%同一性,或至少99%同一性,其中所述编码序列的至少一个密码子与上述或本文公开序列中的相应密码子相比,具有更高人tRNA的频率。

[0112] 在一个实施方案中,所述转基因比SEQ ID:3包含更少的替代开放阅读框。在一个实施方案中,所述转基因被修饰以通过终止或去除不编码所需转基因的开放阅读框(ORF)来增强表达。开放阅读框(ORF)是位于起始密码子之后且不包含终止密码子的核酸序列。ORF可能处于正向取向或反向取向,并且与所关注的基因相比可能是处于“框内”或“框外”。这种开放阅读框有可能在表达盒中连同所关注的基因一起被表达,并可能造成不期望的不良影响。在一些实施方案中,所述转基因已被修饰以通过进一步改变密码子使用去除开放阅读框。这是通过以与所需的ORF反向或框外的方式消除一个或多个起始密码子(ATG)和/或引入一个或多个终止密码子(TAG、TAA或TGA)实现的,同时保留被编码的氨基酸序列,并可选地在所关注的基因中保持高度利用的密码子(即,避免使用频率<20%的密码子)。

[0113] 在本公开的变型中,所述转基因编码序列可以通过密码子优化和非转基因ORF的去除中的任一种或使用这两种技术来优化。在某些情况下,可以在密码子优化后去除或最小化非转基因ORF,从而去除在密码子优化期间引入的ORF。

[0114] 在一个实施方案中,所述转基因比SEQ ID:3包含更少的CpG位点。在不受理论约束的情况下,人们相信,多核苷酸序列中CpG位点的存在与宿主对包含所述多核苷酸序列的病毒载体的不良免疫反应有关。在一些实施方案中,所述转基因被设计为减少CpG位点的数

量。美国专利申请公开号US20020065236A1中提供了示例性方法。

[0115] 在一个实施方案中,所述转基因比SEQ ID:3包含更少的隐性剪接位点。对于优化,可以使用GeneArt®软件,例如,增加GC含量和/或去除隐性剪接位点以避免转录沉默,并因此增强转基因表达。或者,可以使用本领域已知的任何优化方法。例如,在国际专利申请公开号W02004015106A1中描述了去除隐性剪接位点。

[0116] 本文还公开了编码TCIRG1(例如,本文公开的TCIRG1序列)的表达盒和基因治疗载体,包括:共有最优Kozak序列、全长聚腺苷酸化(polyA)序列(或取代截短的全长polyA),以及最小或没有上游(即5')起始密码子(即ATG位点)。

[0117] 在一些实施方案中,所述表达盒包含两个或多个5'长末端重复序列(LTR)、增强子/启动子区、共有最优Kozak序列、转基因(例如,编码本文公开的TCIRG1的转基因)、包括全长polyA序列的3'非翻译区和3'LTR。

[0118] 在一个实施方案中,所述表达盒包含可操作地连接到转基因的Kozak序列。在一个实施方案中,所述Kozak序列是包含或由SEQ ID NO:6组成的共有最优Kozak序列。

[0119] GCCGCCACCATGG (SEQ ID NO:6)

[0120] 在各种实施方案中,所述表达盒包含可操作地连接到转基因的替代Kozak序列。在一个实施方案中,所述Kozak序列是包含或由SEQ ID NO:14-18中的任何一个组成的替代Kozak序列。

[0121] (gcc) gccRccAUGG (SEQ ID NO:14)

[0122] AGNNAUGN (SEQ ID NO:15)

[0123] ANNAUGG (SEQ ID NO:16)

[0124] ACCAUGG (SEQ ID NO:17)

[0125] GACACCAUGG (SEQ ID NO:18)

[0126] 在SEQ ID NO:14中,小写字母表示在碱基仍然可以变化位置处的最常见碱基;大写字母表示高度保守的碱基;表示腺嘌呤或鸟嘌呤。在SEQ ID NO:14中,括号(gcc)中的序列是可选的。在SEQ ID NO:15-17中,'N'表示任何碱基。

[0127] 可以使用多种序列代替这种共有的最优Kozak序列作为翻译起始位点,并且鉴定和测试其他序列在所属领域的技术人员的技能范围内。参见Kozak M. An analysis of vertebrate mRNA sequences: intimations of translational control. J. Cell Biol. 115 (4): 887-903 (1991)。

[0128] 在一个实施方案中,所述表达盒包含可操作地连接到所述转基因的全长polyA序列。在一个实施方案中,所述全长polyA序列包含SEQ ID NO:7。

[0129] TGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAG
GACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGC
TGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCTTATTC
CATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTGTTATTTTTTTCTTTAACATCCCTAAA
ATTTTCCTTACATGTTTTACTAGCCAGATTTTCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCT
TATGGAGATC

[0130] (SEQ ID NO:7)

[0131] 本公开的表达盒中可以使用各种替代的polyA序列,包括但不限于牛生长激素聚

腺苷酸化信号 (bGHPA) (SEQ ID NO:19)、SV40早期/晚期聚腺苷酸化信号 (SEQ ID NO:20) 和人生长激素 (HGH) 聚腺苷酸化信号 (SEQ ID NO:21)。

[0132] TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCCGTGCCTTCCTTGACCCTGGAA
GGTGCCACTCCCCTGTCCTTTTCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCT
GGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAGGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT
CTATGGCTTCTG

[0133] (SEQ ID NO:19)

[0134] CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTA
TTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGC
ATTCATTTTATGTTTCAGGTTTCAGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTA

[0135] (SEQ ID NO:20)

[0136] CTGCCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTG
CCCACCAGCCTTGTCTAATAAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAATATTATGGGGTG
GAGGGGGTGGTATGGAGCAAGGGGCCCAAGTTGGGAAGAAACCTGTAGGGCCTGC

[0137] (SEQ ID NO:21)

[0138] 在一些实施方案中,所述表达盒包含polyA序列的活性片段。在特定实施方案中,所述polyA序列的所述活性片段包含或由例如,本文公开的任何polyA序列的小于20个碱基对(bp)、小于50bp、小于100bp或小于150bp组成。

[0139] 在某些情况下,通过确保所述表达盒不包含竞争性ORFs来增强所述转基因的表达。在一个实施方案中,所述表达盒在所述转基因起始密码子5'的20、30、40、50、60、70、80、90、100、200、300、400或500个碱基对内不包含起始密码子。在一个实施方案中,所述表达盒不包含所述转基因的起始密码子的5'的起始密码子。

[0140] 在一个实施方案中,所述表达盒在5'到3'方向上包括可操作地连接的第一反向末端重复序列、增强子/启动子区、内含子、共有最优Kozak序列、所述转基因、包括全长polyA序列的3'非翻译区和第二个反向末端重复序列,其中所述表达盒在所述转基因起始密码子的5'不包含起始密码子。

[0141] 在一实施方案中,增强子/启动子区在5'至3'方向上包含CMV IE增强子和鸡β-肌动蛋白启动子。在一实施方案中,增强子/启动子区包含CAG启动子。如本文所用,“CAG启动子”是指包含CMV早期增强子元件、鸡β-肌动蛋白启动子、鸡β-肌动蛋白基因的第一外显子和第一内含子以及来自兔β-肌红蛋白基因的剪接受体的多核苷酸序列。

[0142] 在一个实施方案中,所述增强子/启动子区包含延伸因子1短启动子(EFS启动子)并且是一个更短的无内含子版本的延伸因子1启动子。如本文所述的“EFS启动子”是指包含短的、无内含子形式的EF1α的多核苷酸序列。所述EFS启动子最近被用于很多临床试验。它是一种细胞衍生的增强子/启动子,减弱了附近启动子的交叉激活,因此假设降低了基因毒性的风险。

[0143] 在一个实施方案中,所述表达盒与选自SEQ ID NO:1的序列具有至少95%同一性。在一个实施方案中,所述表达盒与选自SEQ ID NO:1的序列具有完全同一性,或与选自SEQ ID NO:1的序列具有至少80%、至少85%、至少90%、至少95%、至少98%或至少99%同一性。在某些实施方案中,与选自SEQ ID NO:1的序列相比,所述表达盒包含一种或多种修饰。

在特定实施方案中,一种或多种修饰包括以下中的一种或多种:去除一个或多个(例如,全部)上游ATG序列,用优化的共有的Kozak序列或另一Kozak序列(包括但不限于本文公开的任何序列)替换Kozak序列,和/或用全长聚腺苷酸化序列或另一聚腺苷酸化序列(包括但不限于本文公开的任何序列)替换聚腺苷酸化序列。这些示例性表达盒内的遗传元件的说明性配置描绘于图1中。

[0144] 在相关实施方案中,本公开提供了包含本文公开的表达盒的基因治疗载体。通常,本文所述的基因治疗载体包含表达盒,所述表达盒包含编码TCIRG1的一种或多种同源异构体的多核苷酸,从而表达TCIRG1以部分或完全纠正有需要的受试者(例如,患有婴儿恶性骨硬化病或特征为至少部分由于TCIRG1表达缺陷的破骨细胞形成缺陷的其他病症的受试者)的TCIRG1蛋白表达水平缺陷和/或破骨细胞形成缺陷。在特定实施方案中,所述表达盒包含编码本文公开的TCIRG1的多核苷酸序列,例如SEQ ID NO:3或与SEQ ID NO:s:3中任一种具有至少90%、至少95%、至少98%或至少99%同一性的序列。基因治疗载体可以是病毒或非病毒载体。说明性非病毒载体包括例如裸DNA、阳离子脂质体复合物、阳离子聚合物复合物、阳离子脂质体-聚合物复合物和外来体。病毒载体的实例包括但不限于腺病毒、逆转录病毒、慢病毒、疱疹病毒和腺相关病毒(AAV)载体。

[0145] 可用于本发明的实践中的基因递送病毒载体可以利用分子生物学领域中众所周知的方法来构建。通常,携带转基因的病毒载体由编码转基因的多核苷酸、合适的调节元件和产生病毒蛋白所必需的元件组装,所述病毒蛋白质介导细胞转导。此类重组病毒可以通过所属领域中已知的技术产生,例如通过转染包装细胞或通过用辅助质粒或病毒瞬时转染。病毒包装细胞的典型实例包括但不限于HeLa细胞、SF9细胞(任选地具有杆状病毒辅助载体)、293细胞等。基于疱疹病毒的系统可用于产生AAV载体,如US20170218395A1中所描述。用于产生此类复制缺陷性重组病毒的详细方案可以例如在W095/14785、W096/22378、美国专利号5,882,877、美国专利号6,013,516、美国专利号4,861,719、美国专利号5,278,056和W094/19478中找到,其中的每一个的完整内容在此以引用的方式并入。

[0146] 在一些实施方案中,所述载体是逆转录病毒载体,或更具体地,慢病毒载体。如本文所用,术语“逆转录病毒”(retrovirus)或“逆转录病毒”(retroviral)是指将其基因组RNA逆转录成线性双链DNA拷贝并随后将其基因组DNA共价整合到宿主基因组中的RNA病毒。逆转录病毒载体是基因递送的常用工具(Miller,2000,Nature.357:455-460)。一旦所述病毒被整合到宿主基因组中,它就被称为“前病毒”。所述前病毒作为RNA聚合酶II的模板,并指导所述病毒编码的RNA分子的表达。

[0147] 示例性逆转录病毒(逆转录病毒科)包括但不限于:(1) γ 逆转录病毒属,例如莫洛尼鼠白血病病毒(M-MuLV)、莫洛尼鼠肉瘤病毒(MoMSV)、鼠乳腺肿瘤病毒(MuMTV)、长臂猿白血病病毒(GaLV)和猫白血病病毒(FLV),(2) 泡沫病毒属,如猿猴泡沫病毒,(3) 慢病毒属,如人类免疫缺陷病毒-1和猿猴免疫缺陷病毒。

[0148] 如本文所述,术语“慢病毒”(lentiviral)或“慢病毒”(lentivirus)是指一组(或属)复合逆转录病毒。示例性慢病毒包括但不限于:HIV(人类免疫缺陷病毒,包括HIV 1型和HIV 2型;维斯纳-梅迪病毒(VMV)病毒;山羊关节炎-脑炎病毒(CAEV);马传染性贫血病毒(EIAV);猫免疫缺陷病毒(FIV);牛免疫缺陷病毒(BIV);和猿猴免疫缺陷病毒(SIV)。在一个实施方案中,基于HIV的载体骨架(即HIV顺式作用序列元件)是优选的。

[0149] 逆转录病毒载体,更具体的是慢病毒载体,可用于实施本发明。因此,本文所用的术语“逆转录病毒载体”意在包括“慢病毒载体”;并且本文所用的术语“逆转录病毒”意在包括“慢病毒”。

[0150] 术语病毒载体可以指能够将核酸转移到细胞中的载体或病毒颗粒,或者指所述被转移的核酸本身。病毒载体含有主要来自病毒的结构性遗传元件和/或功能性遗传元件。术语“逆转录病毒载体”是指含有主要来自逆转录病毒的结构性遗传元件和/或功能性遗传元件或其部分的病毒载体。术语“慢病毒载体”是指含有结构性遗传元件和/或功能性遗传元件或其部分的病毒载体,包括主要源自慢病毒的LTR。术语“杂交物”是指包含逆转录病毒(例如慢病毒)序列和非慢病毒病毒序列的载体、LTR或其他核酸。在一个实施方案中,杂合载体是指包含用于逆转录、复制、整合和/或包装的逆转录病毒(例如慢病毒)序列的载体或转移质粒。

[0151] 在具体的实施方案中,术语“慢病毒载体”和“慢病毒表达载体”可用于指代慢病毒转移质粒和/或感染性慢病毒颗粒。当在本文中提及诸如克隆位点、启动子、调控元件、异源核酸等元件时,应理解这些元件的序列在本发明所述的慢病毒颗粒中以RNA形式存在并且在本发明所述的DNA质粒中以DNA形式存在。

[0152] 根据某些具体实施方案,大部分或全部所述病毒载体骨架序列源自慢病毒,例如HIV-1。然而,应当理解,可以使用慢病毒序列的许多不同来源,并且可以在某些慢病毒序列中进行大量替换和改变而不削弱转移载体执行本文所述功能的能力。此外,多种慢病毒载体是本领域已知的,参见Naldini等人,(1996a,1996b和1998);Zufferey等人,(1997);Dull等人,1998,美国专利号6,013,516;和5,994,136,其中许多适用于产生本发明的病毒载体或转移质粒。

[0153] 在制备慢病毒载体时,可以使用用于产生慢病毒载体的任何宿主细胞,包括例如哺乳动物细胞(例如HEK 293T细胞)。宿主细胞也可以是所述慢病毒gag/pol和rev基因在所述宿主细胞中稳定维持的包装细胞,或所述慢病毒载体基因组稳定维持和包装的生产细胞。使用本领域已知的标准技术纯化和配制慢病毒载体。

[0154] 在某些实施方案中,本发明包括包含本文公开的基因表达盒、基因转移盒或重组慢病毒载体的细胞。在相关实施方案中,所述细胞用包含本文公开的表达盒重组慢病毒载体转导,或所述细胞具有整合到细胞基因组中本文公开的表达盒。在某些实施方案中,所述细胞是用于产生重组逆转录病毒载体的细胞,例如包装细胞。

[0155] 在一些实施方案中,所述慢病毒载体是假型化的。例如,包含异源env基因的质粒可以用于假型化。合适的env基因包括但不限于VSV-G。

[0156] 在一些实施方案中,所述转移质粒的骨架包含RNA-OUT序列。RNA-OUT是一种可选择标记系统,它有助于在使用抗生素的情况下选择携带所述转移质粒的细胞,例如在此引入作为参考的美国专利号9,109,012和9,737,620中所述。在一些实施方案中,所述RNA-OUT序列是:

[0157] GTAGAATTGGTAAAGAGAGTCGTGTA AAAATATCGAGTTCGCACATCTTGTTGTCTGATTATTGATTTTT
GGCGAAACCATTTGATCATATGACAAGATGTGTATCTACCTTAACTTAATGATTTTGATAAAAATCATTAGG

[0158] (SEQ ID NO:22)

[0159] 有利地,所述RNA-OUT序列有助于所述转移质粒在包装细胞系中稳定增殖。

[0160] 在一些实施方案中,本公开提供了一种转移表达盒,包含与SEQ ID NO:1具有至少75%、76%、77%、78%、79%、80%、81%、82%、83%、84%、85%、86%、87%、88%、89%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%或100%同一性的多核苷酸。在一些实施方案中,所述表达盒包含与SEQ ID NO:1具有至少80%、85%、90%、95%、99%或100%同一性的多核苷酸。在一些实施方案中,所述表达盒具有序列SEQ ID NO:1。

[0161] AAV是4.7kb的单链DNA病毒。基于AAV的重组载体与出色的临床安全性相关,因为野生型AAV是非致病的,并且与任何已知疾病不具有病因关联。另外,AAV提供了在许多组织中实现高效基因递送和持续转基因表达的能力。“AAV载体”是指源自腺相关病毒血清型的载体,包括但不限于AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAVrh.10、AAVrh.74等。AAV载体可具有全部或部分缺失的一种或多种AAV野生型基因,例如rep和/或cap基因,但保留功能性侧接反向末端重复(ITR)序列。功能性ITR序列是拯救、复制和包装AAV病毒粒子所必需的。因此,本文将AAV载体定义为至少包括顺式复制和包装病毒(例如,功能性ITR)所需的那些序列。ITR不必是野生型核苷酸序列,并且可以例如通过核苷酸的插入、缺失或取代而改变,只要序列提供功能性拯救、复制和包装。AAV载体可包含其他修饰,包括但不限于一种或多种修饰的衣壳蛋白(例如,VP1、VP2和/或VP3)。举例来说,可以修饰衣壳蛋白以改变向性和/或降低免疫原性。使用已知技术构建AAV表达载体,以至少提供作为转录方向上可操作连接组件的控制元件,包括转录起始区、所关注的DNA(即所述TCIRG1基因)和转录终止区。

[0162] 药物组合物和使用方法

[0163] 本公开还提供了包含本文公开的表达盒或载体(例如,基因治疗载体)和一种或多种药学上可接受的载体、稀释剂或赋形剂的药物组合物。在一些实施方案中,所述药物组合物包含慢病毒颗粒,包含本文公开的表达盒,例如,其中所述表达盒包含编码TCIRG1的密码子转基因,例如,SEQ ID NO:3。本发明提供了例如用于预防或治疗以破骨细胞形成缺陷为特征的病症(例如婴儿恶性骨硬化病)的药物组合物,包含治疗有效量的慢病毒颗粒,所述慢病毒颗粒包含编码TCIRG1的一种或多种同源异构体的多核苷酸的核酸序列。

[0164] 在具体的实施方案中,本文公开所述的慢病毒颗粒用于转导源自受试者的自体CD34+造血干细胞(HSC),从而弥补所述遗传缺陷。可在体内或离体状态下发生转导。在一些实施方案中,所述CD34+富集的细胞群在具有重组人细胞因子的CellGenix干细胞生长培养基(SCGM)中培养,并且在5%CO₂和5%O₂中于37°C下温育。在一些实施方案中,所述CD34+富集的细胞群与用于预刺激的相同添加剂(任选地添加转导增强剂),以及包含所述表达盒EFS-TCIRG1-WPRE(例如,MOI 50)的慢病毒颗粒一起温育。在一些实施方案中,在转导之后,洗涤细胞悬浮液,去除一部分细胞及清液用于放行测试,并且冷冻药品以备输注。在一些实施方案中,通过用G-CSF、plerifaxor,或G-CSF和plerifaxor的组合治疗患者来动员HSC。然后通过单采术从所述患者的外周血中收集所述HSC。例如,使用磁捕获(例如,在Miltenyi Biotec ClinMACs系统上)富集CD34+细胞,并且所述CD34+富集的细胞用所述慢病毒颗粒离体转导。在一些实施方案中,转导过程包括使用转导增强剂,例如但不限于聚沙姆和前列腺素E2(PGE2)。

[0165] 在一些实施方案中,然后通过输注至少 2.0×10^6 CD34+细胞/kg将所转导的HSC移植到受试者中,例如人受试者。在一些实施方案中,他们用表达TCIRG1的细胞重新填充HSC龕。

在一些实施方案中,他们用表达TCIRG1的细胞重新填充破骨细胞龛。

[0166] 本发明还提供了例如,用于预防或治疗以破骨细胞形成缺陷为特征的病症(例如婴儿恶性骨硬化病)的药物组合物,包含治疗有效量的经修饰的细胞,所述经修饰的细胞包含编码TCIRG1的一种或多种同源异构体的多核苷酸的核酸序列。在一些实施方案中,所述经修饰的细胞表达TCIRG1或其功能性变体,其表达水平类似于在具有功能性TCIRG1基因的破骨细胞中观察到的TCIRG1表达水平。在一些实施方案中,所述经修饰的细胞表达TCIRG1或其功能性变体,其表达水平类似于在源自未患有或未疑似患有IMO受试者的破骨细胞中观察到的TCIRG1的表达水平。在一些实施方案中,所述经修饰的细胞是造血干细胞(HSC)。在一些实施方案中,所述经修饰的细胞是CD34+祖细胞。在一些实施方案中,所述经修饰的细胞源自通过单采术从患有或疑似患有IMO的受试者中分离出来的HSC。在一些实施方案中,所述经修饰的细胞是所述受试者的自体细胞。在一些实施方案中,所述经修饰的细胞源自通过施用G-CSF、plerifaxor,或G-CSF和plerifaxor的组合动员HSC之后,通过单采术从患有或疑似患有IMO的受试者分离的HSC。在一些实施方案中,所述经修饰的细胞源自通过磁捕获富集CD34+细胞的细胞群。在一些实施方案中,使用本文公开的载体,例如使用本文公开的转移质粒产生的慢病毒载体,转导所述经修饰的细胞。

[0167] 含有所述表达盒或慢病毒颗粒或经修饰的细胞的所述药物组合物的形式可以是任何适合于所选施用模式,例如,用于心室内施用、心肌内施用、冠状动脉内施用、静脉内施用、动脉内施用、肾内施用、尿道内施用、硬膜外施用或肌肉内施用。包含编码一种或多种TCIRG1同源异构体的多核苷酸的基因修饰细胞可以作为与常规药物支持物的混合物,以单位施用形式作为单独的活性剂,或其他活性剂联合向动物和人类施用。在一些实施方案中,药物组合物包含用本公开的任何基因治疗载体离体转导的细胞。

[0168] 在各种实施方案中,药物组合物含有对于能够注射的制剂在药学上可接受的媒介物(例如,载体、稀释剂和赋形剂)。这些媒介物可以具体是等渗的、无菌的、盐溶液(磷酸一钠或磷酸二钠、氯化钠、氯化钾、氯化钙或氯化镁等或这些盐的混合物),或干燥组合物,尤其是冷冻干燥的组合物,其根据情况加入无菌水或生理盐水后,可制成可注射溶液。适合可注射使用的说明性药物形式包括例如无菌水溶液或分散液;包括芝麻油、花生油或水性丙二醇的制剂;和用于临时制备无菌可注射溶液或分散液的无菌粉末。

[0169] 另一方面,本公开提供了预防、减轻、改善、减少、抑制、消除和/或逆转有需要受试者的婴儿恶性骨硬化病(IMO)或另一种病症的一种或多种症状的方法,包括向所述受试者施用本公开的基因治疗载体。术语“婴儿恶性骨硬化病”或“恶性婴儿骨硬化病”或“婴儿常染色体隐性遗传骨硬化病”或“婴儿骨硬化病”或“IMO”是指一种罕见的骨硬化类型的骨骼发育不良,通常出现在婴儿期并且其特征是全身性骨质增生(骨骼过度生长)的独特放射学表现。骨密度的全身性增加特别倾向于涉及髓质部分,而皮质部分相对较少。骨髓间隙的闭塞和随后的细胞功能抑制可导致严重的血液学并发症。继发于骨扩张的视神经萎缩和颅神经损伤可使得发病率显著。未经治疗的病例预后极差,平片为诊断提供了关键信息。临床和放射学相关性也是诊断过程的基础,也经过另外基因检测的确认。

[0170] 在一个实施方案中,所述经修饰的细胞(例如,用本公开的慢病毒颗粒转导自体细胞),通过选自由肠胃外、静脉内、动脉内、心内、冠状动脉内、心肌内、肾内、尿道内、硬膜外和肌肉内组成的组的途径施用。在一些实施方案中,所述经修饰的细胞通过输注(例如静脉

输注)施用。在一个实施方案中,所述经修饰的细胞被多次施用。在一个实施方案中,所述经修饰的细胞通过输注施用。

[0171] 在一个实施方案中,本公开提供了一种治疗有需要的受试者疾病或病症(任选地IMO)的方法,包括将细胞与根据本公开的基因治疗载体接触并且将所述细胞施用于所述受试者。在一实施方案中,细胞是干细胞,任选地多能干细胞。在一个实施方案中,所述干细胞能够分化成骨细胞。在一个实施方案中,所述干细胞能够分化成破骨细胞。在一实施方案中,干细胞是自体的。在一个实施方案中,所述干细胞是CD34+干细胞。

[0172] 在一个实施方案中,所述受试者表现出IMO或其他疾病的症状。在一个实施方案中,所述受试者已被鉴定为TCIRG1表达有所降低或检测不到TCIRG1表达。在一个实施方案中,所述受试者已被鉴定为具有突变的TCIRG1基因。

[0173] 适合使用本文所述方法治疗的受试者/患者包括具有以破骨细胞不足为特征的疾病或病症(例如,IMO以及其他已知的破骨细胞形成病症)风险的个体。在一些实施方案中,所述受试者没有表现出症状。在一些实施方案中,所述受试者目前表现出症状。这样的受试者可能已经被鉴定为具有突变的TCIRG1基因或TCIRG1表达水平降低或检测不到TCIRG1表达水平。所述症状可能是主动表现的症状,或者可能被抑制或控制(例如,通过药物)或处于缓解期。受试者可能已经或可能没有例如由合格的医生诊断患有所述病症。

[0174] 定义

[0175] 术语“T细胞免疫调节因子1(ATPase H⁺转运V0亚基a3)”和“TCIRG1”可互换地指以下核酸和多肽多态性变体、等位基因、突变体和种间同源物:(1)具有氨基酸序列,所述氨基酸序列与由TCIRG1核酸(参见,例如GenBank登录号NM_006019.4(变体1)、NM_006053.3(变体2)、NM_001351059.1(变体3))编码的氨基酸序列或TCIRG1多肽(参见例如GenBank登录号006044.1(同源异构体A)、NP_006044.1(同源异构体B)、NP_001337988.1(同源异构体C))的氨基酸序列具有大于约90%氨基酸序列同一性的氨基酸序列,例如91%、92%、93%、94%、95%、96%、97%、98%或99%或更高的氨基酸序列同一性,优选在至少约25、50、100、200、300、400个或更多氨基酸的区域上,或在全长上;(2)结合到抗体,例如多克隆抗体,其针对包含TCIRG1多肽(例如,本文所述的TCIRG1多肽)的氨基酸序列的免疫原产生;或由TCIRG1核酸(例如,本文所述的TCIRG1多核苷酸)编码的氨基酸序列及其保守修饰的变体;(3)在严格杂交条件下与编码TCIRG1蛋白的核酸序列对应的反义链及其保守修饰变体特异性杂交;(4)具有核酸序列,所述核酸序列与TCIRG1核酸(例如,如本文所述的TCIRG1多核苷酸和如本文所述编码TCIRG1多肽的TCIRG1多核苷酸)具有大于约90%,优选大于约91%、92%、93%、94%、95%、96%、97%、98%、99%或更高核苷酸序列同一性,优选地,在至少约25、50、100、200、500、1000、2000个或更多个核苷酸的区域上,或在全长上。

[0176] 所述TCIRG1基因编码多种蛋白质同源异构体,其中2种主要同源异构体。全长同源异构体a(OC116)编码液泡H⁽⁺⁾-ATPase的A3亚基,其参与调节真核细胞的细胞内隔室和细胞器的pH,包括破骨细胞的细胞内隔室和细胞器的pH。较短的同源异构体b(TIRC7)编码在T淋巴细胞激活和免疫反应中起重要作用的T细胞特异性膜蛋白。

[0177] 在两个或多个核酸序列或多肽序列的语境中,术语“相同的”或百分比“同一性”是指如使用以下序列比较算法之一或通过手动比对和目视检查进行检测,在比较窗口或指定区上进行比较和比对以获得最大对应时,两个或多个相同或具有特定百分比的相同氨基酸

残基或核苷酸的序列或子序列(即在特定区与参考序列(例如,如本文所述的TCIRG1多核苷酸或多肽序列)具有至少约80%的同一性,例如至少约85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%的同一性)。这类序列随后被称为“基本上一致”。此定义也指测试序列的补充。优选地,同一性存在于长度为至少约25个氨基酸或核苷酸的区域上,例如长度为50、100、200、300、400个氨基酸或核苷酸的区域上,或在参考序列的全长上。

[0178] 对于序列比较,通常一个序列充当参考序列,测试序列与参考序列进行比较。当使用序列比较算法时,将测试序列和参考序列输入到计算机中,如果需要,指定子序列坐标,并且指定序列算法程序参数。可以使用默认程序参数,或者可以指定替代参数。随后,序列比较算法基于程序参数计算测试序列相对于参考序列的序列同一性百分比。对于核酸和蛋白质与TCIRG1核酸和蛋白质的序列比较,使用BLAST和BLAST 2.0算法以及默认参数。

[0179] 如本文所用,“比较窗口”包括包括指选自20至600、通常约50至约200、更通常约100至约150组成的组的连续位置数中的任何一个的片段,其中在两个序列最优比对后,可以将序列与相同连续位置数的参考序列进行比较。用于比较的序列的比对方法是所属领域中众所周知的。例如,可用以下方法进行用于比较序列的最佳比对:局部同源算法,Smith&Waterman, *Adv. Appl. Math.* 2:482 (1981), 通过同源比对算法,Needleman&Wunsch, *J. Mol. Biol.* 48:443 (1970), 相似性搜索方法,Pearson&Lipman, *Proc. Nat' l. Acad. Sci. USA* 85:2444 (1988), 通过计算机化实施这些算法(GAP、BESTFIT、FASTA和TFASTA, Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), 或通过手动比对和目视检查(参见,例如, Ausubel等人, 编, *Current Protocols in Molecular Biology* (1995 supplement)) 来进行。适用于确定百分比序列同一性和序列相似性算法的示例是BLAST和BLAST2.0算法,它们分别阐述于Altschul等人, *Nucleic Acids Res.* 25:3389-3402 (1977) 和Altschul等人, *J. Mol. Biol.* 215:403-410 (1990)。用于进行BLAST分析的软件可通过国家生物技术信息中心(万维网ncbi.nlm.nih.gov上)公开获得。

[0180] 两个核酸序列或多肽基本上一致的指示是,由第一核酸编码的多肽与针对由第二核酸编码的多肽产生的抗体免疫上交叉反应,如下所描述。因此,多肽通常与第二多肽基本上一致,例如,其中两个肽仅通过保守取代而不同。两个核酸序列基本上一致的另一个指示是两个分子或其补体在严格条件下彼此杂交。两个核酸序列基本上一致的又一指示是相同的引物可用于扩增序列。

[0181] 如本文所用,“施用”是指局部和全身施用,例如包括肠内、肠胃外、肺部和局部/透皮施用。例如,可用于本文所述方法的化合物(例如,编码一种或多种TCIRG1同源异构体的多核苷酸)的施用途径包括口服(经口(P.O.))施用、鼻腔或吸入施用、作为栓剂施用、局部接触、经皮递送(例如,通过透皮贴剂)、鞘内(IT)施用、静脉内(“iv”)施用、腹膜内(“ip”)施用、肌肉内(“im”)施用、病灶内施用或皮下(“sc”)施用,或植入缓释装置例如微型渗透泵、储库制剂等。于受试者。施用可以通过任何途径进行,包括肠胃外和经粘膜(例如,口服、鼻内、阴道、直肠或透皮)。肠胃外施用包括例如静脉内、肌内、动脉内、肾内、尿道内、心脏内、冠状动脉内、心肌内、皮内、硬膜外、皮下、腹膜内、心室内、离子电渗疗法和颅内。其他递送模式包括但不限于脂质体制剂的使用、静脉内输注、透皮贴剂等。

[0182] 术语“全身施用(systemic administration)”和“全身施用(systemically administered)”是指向哺乳动物施用化合物或组合物以使得化合物或组合物通过循环系

统递送至体内的位点(包括药物作用的靶向位点)的方法。全身施用包括但不限于口服、鼻内、直肠和肠胃外(例如,除了通过消化道之外,例如肌内、静脉内、动脉内、透皮和皮下)施用。

[0183] 术语“共同施用”或“同时施用”,当用于例如化合物(例如,TCIRG1多核苷酸)和/或其类似物和另一种活性剂时,是指施用所述化合物和/或类似物和所述活性剂,使得两者可以同时发挥生理作用。然而,这两种药剂不必一起施用。在某些实施方案中,一种药剂的施用可以先于在一种药剂的施用。同时的生理作用不一定需要在循环中同时存在两种药剂。然而,在某些实施方案中,共施用通常导致两种药剂对于任何给定剂量以其最大血清浓度的显著分数(例如,20%或更大,例如,30%或40%或更大,例如,50%或60%或更大,例如,70%或80%或90%或更大)同时存在于体内(例如,血浆中)。

[0184] 术语“有效量”或“药学有效量”是指产生所预期结果所需的一种或多种组合物(例如基因治疗载体、经修饰的细胞)的量和/或剂量和/或剂量方案,例如,所预期结果是一种或多种TCIRG1同源异构体的表达增强,其量足以降低以自噬受损或缺陷为特征的疾病(例如,IMO)的最终严重程度。

[0185] 短语“导致被施用”是指医学专业人员(例如,医生)或控制受试者的医疗护理的人员所采取的行动,其控制和/或允许向受试者施用有争议的药剂/化合物。导致被施用可以涉及诊断和/或确定适当的治疗或预防方案,和/或为受试者开处特定的药剂/化合物。此类处方可以包括例如起草处方表、注释医疗记录等。

[0186] 如本文所用,术语“治疗”和“治疗”是指延迟所述术语所适用的疾病或病状或所述疾病或病状的一种或多种症状的发作、延迟或逆转其进展、降低其严重程度或减轻或预防其疾病或病状。术语“治疗”和“治疗”还包括预防、减轻、改善、减少、抑制、消除和/或逆转疾病或病状的一种或多种症状。

[0187] 术语“减轻”是指减少或消除所述病理或疾病的一种或多种症状,和/或减少或延迟所述病理或疾病的一种或多种症状的发作率或严重程度,和/或预防所述病理或疾病。在某些实施方案中,减少或消除病理或疾病的一种或多种症状可以包括例如TCIRG1的一种或多种同源异构体的表达水平可测量且持续的增加。

[0188] 如本文所用,短语“基本上由.....组成”是指在方法或组合物中列举的活性药物的种属或种类,并且还包括本身对于所列举的适应症或目的不具有实质性活性的其他药剂。

[0189] 术语“受试者”、“个体”和“患者”可互换地指哺乳动物,优选为人类或非人类灵长类动物,但也指驯养的哺乳动物(例如,犬科动物或猫科动物)、实验室哺乳动物(例如,小鼠、大鼠、兔、仓鼠、豚鼠)和农业哺乳动物(例如,马、牛、猪、绵羊)。在各种实施方案中,受试者可以是人类(例如,成年男性、成年女性、青少年男性、青少年女性、男童、女童)。

[0190] 术语“基因转移”或“基因递送”是指用于将外源DNA可靠地插入宿主细胞内的方法或系统。此类方法可导致非整合的转移DNA的瞬时表达、染色体外复制和转移复制子(例如,附加体)的表达,或将转移的遗传物质整合到宿主细胞的基因组DNA中。

[0191] 术语“载体”在本文中用于指能够转移或转运另一种核酸分子的核酸分子。转移的核酸通常连接至(例如,插入)载体核酸分子内。载体可以包括指导细胞中自主复制或逆转录的序列,或者可以包括足以允许整合到宿主细胞DNA中的序列。“载体”包括基因治疗载

体。如本文所用，术语“基因治疗载体”是指能够用于进行基因治疗（例如，将编码治疗多肽的多核苷酸序列递送到受试者）的载体。基因治疗载体可包含编码治疗活性多肽（例如TCIRG1）的核酸分子（“转基因”），或当被引入受试者时，可用于替代基因治疗的其他基因。有用的载体包括但不限于病毒载体。

[0192] 如本文所用，术语“表达盒”是指能够在适当环境下驱动编码治疗活性多肽（例如，TCIRG1）的多核苷酸（例如，转基因）表达的DNA片段，所述多核苷酸并入所述表达盒。当引入宿主细胞中时，表达盒尤其能够引导细胞的机制将转基因转录成RNA，所述RNA随后通常进一步加工且最终翻译成治疗活性多肽。表达盒可以包含在基因治疗载体中。通常，术语表达盒不包括5'到所述5'LTR和3'到所述3'LTR的多核苷酸序列。

[0193] 本说明书中引用和鉴定的所有专利、专利出版物和其他出版物出于所有目的单独且明确地以全文引用的方式并入本文中。

[0194] 实施例

[0195] 实施例1:转移质粒的稳定增殖

[0196] 检测了包含最小TCIRG1表达盒，EFS-TCIRG1-WPRE (SEQ ID NO:1) 的不同质粒的稳定性。具有氨苄青霉素抗性 (AmpR) 的质粒构建体pRRL.PPT.EFS.tcirg1h.wpre (图3C)，在转染到包装细胞系之前于用于增殖所述质粒的大肠杆菌细胞培养期间，表现出意料之外的不良生长和不稳定性，如质粒产量低和指示不稳定性的降解DNA一般涂片所示 (图3A)。其他各种质粒骨架也表现出不稳定性 (数据未显示)。然而，当最小表达盒EFS-TCIRG1-WPRE (SEQ ID NO:1) 从pRRL载体克隆到具有RNA-OUT序列的pCCL载体中 (图3C) 时，所得质粒构建体，pCCL.PPT.EFS.tcirg1h.wpre (SEQ ID NO:27) 在大肠杆菌中增殖时表现出意料之外的良好生长和稳定性。。这通过质粒的高产量和观察到的限制酶切模式显示出来 (图3B)。完整的载体序列作为SEQ ID NO:27提供，每个载体元件的位置在表2中提供。

[0197] 表2:pCCL.PPT.EFS.tcirg1h.wpre载体元件

名称	类型	最小 - 最大 参考序列中的位置 (核苷酸)	长度 (碱基对)
EFS	启动子	1-243	243
TCIRG1	CDS	257-2,749	2,493
WPRE	调控的	2,782-3,384	603
3'LTR	LTR	3,471-3,704	234
SV40 poly(A)	polyA 信号	3,776-3,907	132

[0198]

[0199]	SV40 ori	复制起点	3,917-4,076	160
	pUC 起点	复制起点	4,115-5,129	1,015
	RNA-OUT	阻遏物	5,146-5,284	139
	CMV IE	启动子	5,334-5,910	577
	5-LTR	LTR	5,933-6,120	188
	psi	包装	6,222-6,266	45
	gag	CDS	6,267-6,628	362
	RRE	调控的	6,629-7,486	858
	cPPT/CTS	多嘌呤区	7,505-7,622	118

[0200] 通过将pCCL/RNA-OUT载体与包装质粒 (pCMV Δ R8.91) 和包膜质粒 (VSV-G pMDG) 瞬时转染到293T细胞中产生慢病毒载体,并根据图4所示的方案产生。

[0201] 实施例2:通过pCCL.PPT.EFS.tcirg1h.wpre恢复IMO患者破骨细胞的再吸收功能

[0202] 这个实施例展示了pCCL.PPT.EFS.tcirg1h.wpre在源自患者的HSC中用于慢病毒介导的TCIRG1基因转移的用途。获得、扩增并用携带实施例1中描述的pCCL.PPT.EFS.tcirg1h.wpre的慢病毒颗粒转导HSC以获得基因修饰的HSC。输注后,所述基因修饰的HSC将分化为破骨细胞。使用的方法基本上阐述于Moscatelli等人Hum.Gene Therap.29:938-949 (2017)。

[0203] 获得IMO患者的外周血样本或正常分娩的脐带血(CB)样本。使用Ficoll密度梯度离心分离这些来源的单核细胞,并且使用磁激活细胞分选(MACS)柱(Miltenyi Biotec, Bergisch Gladbach,Germany)将CD34+细胞从所述单核细胞级分中分离。对于扩增,细胞在SFEM StemSpan培养基(StemCell Technologies,Vancouver,BC)中培养,所述培养基含有人重组细胞因子M-CSF (50ng/ml)、GM-CSF (30ng/ml)、SCF (200ng/ml)、IL-6 (10ng/ml)和Flt3L (50ng/ml) (R&D Systems,Minneapolis MN)。使用24孔细菌培养板将CD34+细胞以 5×10^4 细胞的密度接种在1ml培养基中,并在37°C下温育一周,然后以 1×10^5 /孔的密度收集并重新接种。从第7天开始,每2-3天通过半耗竭的方式更换所述培养基。对于移植,CD34+细胞在SFEM StemSpan培养基(StemCell Technologies,Vancouver,BC)中培养30小时,所述培养基含有所述人重组细胞因子:SCF (100ng/ml)、Flt3L (100ng/ml)和TPO (100ng/ml) (R&D Systems,Minneapolis MN)。

[0204] 转导在涂有RetroNectin (Takara Bio,Otsu,Japan)的24孔板中进行。对于体外实验,CD34+细胞在第3天以30的感染复数(MOI)第一次转导6小时,在第7天以30的MOI第二次转导6小时,然后用骨髓细胞因子混合物培养一周,随后向破骨细胞分化。对于体内实验,使用较短的转导方案以实现有效转导,同时保持所述CD34+群的茎/祖细胞性质。分离单核细胞并用第一次(MOI为30或100)转导过夜,然后在第二天第二次(MOI为30或100)转导6小时,之后准备将所述细胞移植到受试者(小鼠或人类患者)体内。

[0205] 破骨细胞生成可以通过在50ng/ml的M-CSF和50ng/ml的RANKL存在下分化约10天来评估,然后用4%甲醛固定细胞以进行进一步分析或裂解细胞以进行蛋白质印迹分析。通过测定释放c端I型胶原蛋白片段(CTX-I)和释放Ca²⁺到培养基中,以及使用苏木精染色所固定的细胞来观察吸收陷窝的形成,来评估吸收。

[0206] 对于动物研究,将转导的破骨细胞移植到NSG小鼠中。8至15周龄的NSG小鼠接受

300cGy亚致死照射,6小时后通过尾静脉注射移植 1×10^5 未转导的CB CD34⁺细胞或IMO CD34⁺细胞,该细胞转导源自pCCL.PPT.EFS.tcirg1h.wpre的慢病毒颗粒。小鼠通过饮用水给予环丙沙星两周,以避免移植后感染。在不同的时间点收集外周血,并在小鼠死亡后用研钵粉碎股骨来收集骨髓细胞。

[0207] 对移植后9-19周从小鼠采集样本中的全骨髓基因组DNA进行载体拷贝数分析。通过确定huCD45-APC阳性细胞的百分比,对移植的NSG小鼠的外周血和骨髓进行人体重建分析。对于谱系分析,细胞用针对CD33-PeCy7、CD15-PeCy7、CD19-BV605和CD3-PE的抗体染色。

[0208] 上述方法用于确认在慢病毒介导的TCIRG1基因转移和转导CD34⁺细胞的长期移植后,IMO患者的破骨细胞再吸收功能的恢复。

[0209] 实施例3:使用临床建立的转导方案对人CD34⁺富集的细胞进行体外转导

[0210] 本实施例通过以下方式证明了使用实施例1中描述的质粒制备的EFS-TCIRG1-WPRE (SEQ ID NO:25或27) 前GMP批次的适用性:(1) 通过%活CD34⁺细胞和多向分化能力对转导的CD34⁺细胞进行表型表征,和(2) 通过载体拷贝数的转导效率(液体培养和集落中的VCN测定)。

[0211] 用EFS-TCIRG1-WPRE转导后mPB CD34⁺细胞的保留表型和多向能力

[0212] GMP前EFS-TCIRG1-WPRE批次的性能与用于治疗其他疾病的LV(四批次)进行了比较。与预想的IMO临床试验类似,使用动员的PB CD34⁺细胞作为靶细胞。测试了各种MOI。在所有载体和测试的MOI中,转导后20小时均获得高细胞活力(>95%),表明无短期毒性。转导后不久,CD34⁺细胞的百分比在所有测试条件下都非常高(>97%),并且在液体培养2天后,随着时间的推移,在所有条件下都以相当的水平逐渐减少,正如此类培养所预期的那样。

[0213] 通过量化半固体甲基纤维素培养基培养物中分化的CFU来评估转导的CD34⁺细胞的多谱系能力。评估了分化成红细胞系和髓系的总CFU以及BFU-E、CFU-GM和CFU-GEMM。所述EFS-TCIRG1-WPRE的存在不影响CFU的生长,因为与模拟对照相比,在实验条件下它们的总数中没有观察到显著差异。与模拟相比,在任何集落类型中均未观察到差异,从而证实了在用EFS-TCIRG1-WPRE转导后即使在高MOI值下也能维持体外多谱系能力。

[0214] EFS-TCIRG1-WPRE显示出mPB CD34⁺细胞的高转导效率

[0215] 为了确定治疗用的合适转导效率所需的EFS-TCIRG1-WPRE的载体剂量,使用已确立的临床转导方案测试了慢病毒载体。转导的CD34⁺细胞在液体培养中保持长达12天,使得在VCN评估之前细胞清除游离LV基因组拷贝。通过剂量依赖的IMO载体获得了高VCN/细胞值。增加剂量的影响在所有测试的载体中是一致的,并且以相同的MOI转导使得IMO载体的VCN/细胞值高于LAD-I和FA载体的VCN/细胞值(图5A-5B),表明EFS-TCIRG1-WPRE的高转导效率。

[0216] 还根据以下表型在分离的CFU中评估了VCN/细胞:BFU-E、CFU-GM或CFU-GEMM,以确认不同祖细胞的转导情况。EFS-TCIRG1-WPRE在两个供体的细胞中显示出更高的集落VCN值。与液体培养的结果相似,用IMO载体转导使得在甲基纤维素培养基中培养的集落的转导效率和VCN/细胞较高。

[0217] 已发现,不同CFU类型(BFU-E、CFU-GM和CFU-GEMM)中的VCN模式与其他载体相似,并且通常发现在红细胞集落中的值最高,如先前所述在Charrier等人Gene Therapy 18:479-487(2011)。

[0218] 所述IMO载体EFS-TCIRG1-WPRE转导人CD34⁺细胞的水平与临床批次的慢病毒相当。CD34⁺细胞的表型和多谱系能力得到保留,同时实现了高转导效率。IMO载体在与对照载体相比更低的MOI下成功运行,证明其适用于IMO患者的基因治疗。

[0219] 这些研究表明,VCN和转导效率与体内基因修饰造血细胞的矫正水平相当,使该基因疗法能够用于治疗因TCIRG1突变引起的婴儿恶性骨硬化病。

序列表

<110> 火箭制药有限公司(Rocket Pharmaceuticals, LTD)

Beard, Brian

Ricks, David

Prabhakar, Raj

<120> 用于婴儿恶性骨硬化病的基因治疗载体

<130> ROPA-003/01W0 329592-2098

<150> US 62/852,216

<151> 2019-05-23

<160> 39

<170> PatentIn 3.5版

<210> 1

<211> 3384

<212> DNA

<213> 人工序列

<220>

<223> 实验室制备 - 表达盒

<400> 1

```

ggctccggtg cccgtcagtg ggcagagcgc acatcgccca cagtccccga gaagttgggg      60
ggaggggtcg gcaattgaac cggcgcctag agaaggtggc gcggggtaaa ctgggaaagt      120
gatgtcgtgt actggctccg cctttttccc gaggggtgggg gagaaccgta tataagtgca      180
gtagtcgccg tgaacgttct ttttcgcaac gggtttgccg ccagaacaca ggtgtcgtga      240
cgcggggatcc gccaccatgg gctccatggt tcggagcgag gaggtggccc tggcctcagct      300
ctttctgccc acagcggctg cctacacctg cgtgagtcgg ctgggcgagc tgggcctcgt      360
ggagttcaga gacctcaacg cctcgggtgag cgccttcag agacgctttg tggttgatgt      420
tcggcgctgt gaggagctgg agaagacctt caccttctg caggaggagg tgcggcgggc      480
tgggctggtc ctgccccgcg caaaggggag gctgccggca cccccaccc gggacctgct      540
gcgcatccag gaggagacgg agcgcctggc ccaggagctg cgggatgtgc ggggaacca      600
gcaggccctg cgggcccagc tgcaccagct gcagctccac gccgccgtgc tacgccaggg      660
ccatgaacct cagctggcag ccgcccacac agatggggcc tcagagagga cgcccctgct      720
ccaggcccc ggggggcccgc accaggacct gagggtaac tttgtggcag gtgccgtgga      780
gccccacaag gccctgccc tagagcgctt gctctggagg gcctgcagag gcttctctcat      840
tgccagcttc agggagctgg agcagccgct ggagcaccac gtgacgggcg agccagccac      900
gtggatgacc ttctctatct ctaactgggg tgagcagatc ggacagaaga tccgaagat      960
cacggactgc ttccactgcc acgtettccc gtttctgcag caggaggagg cccgcctcgg      1020
ggccctgcag cagctgcaac agcagagcca ggagctgcag gaggtcctcg gggagacaga      1080
gcggttcctg agccagggtc taggccgggt gctgcagctg ctgccgccag ggcaggtgca      1140
ggtccacaag atgaaggccg tgtacctggc cctgaaccag tgcagcgtga gcaccacgca      1200

```

caagtgcctc attgccgagg cctggtgctc tgtgcgagac ctgcccgcc tgcaggaggc	1260
cctgcgggac agctcgatgg aggagggagt gagtgccgtg gctcaccgca tcccctgccg	1320
ggacatgccc cccacactca tccgcaccaa ccgcttcacg gccagcttcc agggcatcgt	1380
ggatgcctac ggcgtgggccc gctaccagga ggtcaacccc gctccctaca ccatcatcac	1440
cttccccctc ctgtttgctg tgatgttcgg ggatgtgggc cacgggctgc tcatgttctt	1500
cttcgccctg gccatggtcc ttgcggagaa ccgaccggct gtgaaggccg cgcagaacga	1560
gatctggcag actttcttca ggggccgcta cctgctcctg cttatgggcc tgttctccat	1620
ctacaccggc ttcatctaca acgagtgett cagtcgcgcc accagcatct tcccctcggg	1680
ctggagtgtg gccgccatgg ccaaccagtc tggtctggagt gatgcattcc tggcccagca	1740
cacgatgctt accctggacc ccaacgtcac cgggtgtctt ctgggaccct acccctttgg	1800
catcgatcct atttgagacc tggtgccaa ccaactgagc ttctcaact ccttcaagat	1860
gaagatgtcc gtcacctcgg gcgtcgtgca catggccttt ggggtggtcc tcggagtctt	1920
caaccacgtg cactttggcc agaggcaccg gctgctgctg gagacgctgc cggagctcac	1980
cttctcgtcgg ggaactctcg gttacctcgt gttcctagtc atctacaagt ggctgtgtgt	2040
ctgggctgcc agggccgccc cggccccag cactctcact cacttcatca acatgttctt	2100
cttctccac agccccagca acagctgct ctacccccgg caggaggtgg tccaggccac	2160
gctggtggtc ctggccttgg ccatggtgcc catcctgctg cttggcacac ccctgcacct	2220
gctgcaccgc caccgccccc gcctgcggag gaggcccgt gaccgacagg aggaaaacaa	2280
ggccgggttg ctggacctgc ctgacgcata tgtgaatggc tggagctccg atgaggaaaa	2340
ggcagggggc ctggatgatg aagaggaggc cgagctcgtc ccctccgagg tgctcatgca	2400
ccaggccatc cacaccatcg agttctgcct gggtcgcgtc tccaacaccg cctcctacct	2460
gcgcctgtgg gccctgagcc tggcccacgc ccagctgtcc gaggttctgt gggccatggt	2520
gatgcgcata ggccctggcc tgggccggga ggtgggctg gcggtctgtg tgctggtccc	2580
catctttgcc gcctttgccg tgatgaccgt ggctatcctg ctggtgatgg agggactctc	2640
agccttctc caccgcccgc ggctgcactg ggtggaattc cagaacaagt tctactcagg	2700
cacgggctac aagctgagtc ccttcacctt cgctgccaca gatgactagt aagtcgacgg	2760
atcccccggg ctgcaggaat tcgagcatct taccgccatt tataccata tttgttctgt	2820
ttttcttgat ttgggtatac atttaaagt taataaaaca aaatggtggg gcaatcattt	2880
acattttttag ggatatgtaa ttactagttc aggtgtattg ccacaagaca aacatgttaa	2940
gaaactttcc cgttatttac gctctgttcc tgttaatcaa cctctggatt acaaaatttg	3000
tgaaagattg actgatattc ttaactatgt tgctcctttt acgctgtgtg gatatgctgc	3060
tttaatgcct ctgtatcatg ctattgett cegtaegget ttctgtttct cctccttgta	3120
taaactcctg ttgctgtctc tttatgagga gttgtggccc gttgtccgct aacgtggcgt	3180
gggtgtctct gtgtttgctg acgcaacccc cactggctgg ggcattgcca ccacctgtca	3240
actcctttct gggactttcg ctttccccct cccgatgcc acggcagaac tcategccgc	3300
ctgccttgcc cgctgctgga caggggctag gttgctgggc actgataatt ccgtggtgtt	3360
gtcggggaag ctgacgtcct ttcg	3384

<210> 2

<211> 243

<212> DNA	
<213> 智人	
<400> 2	
ggctccggtg cccgtcagtg ggcagagcgc acatcgccca cagtccccga gaagttgggg	60
ggaggggtcg gcaattgaac cggcgcctag agaaggtggc gcggggtaaa ctgggaaagt	120
gatgtcgtgt actggctccg cctttttccc gaggggtgggg gagaaccgta tataagtgca	180
gtagtcgccg tgaacgttct ttttcgcaac gggtttgccg ccagaacaca ggtgtcgtga	240
cgc	243
<210> 3	
<211> 2493	
<212> DNA	
<213> 智人	
<400> 3	
atgggctcca tgtttcggag cgaggaggtg gccctggtcc agctctttct gcccacagcg	60
gctgcctaca cctgcgtgag tcggctgggc gagctgggcc tcgtggagtt cagagacctc	120
aacgcctcgg tgagcgcctt ccagagacgc tttgtggttg atgttcggcg ctgtgaggag	180
ctggagaaga ccttcacctt cctgcaggag gaggtgcggc gggctgggct ggtcctgccc	240
ccgccaaagg ggaggctgcc ggcaccccca ccccgggacc tgctgcgcat ccaggaggag	300
acggagcgcc tggcccagga gctgcgggat gtgcggggca accagcaggc cctgcggggc	360
cagctgcacc agctgcagct ccacgccgcc gtgctacgcc agggccatga acctcagctg	420
gcagccgccc acacagatgg ggcctcagag aggaccccc tgctccaggc ccccgggggg	480
ccgcaccagg acctgagggt caactttgtg gcaggtgccg tggagcccca caaggcccct	540
gccctagagc gcctgctctg gagggcctgc agaggettcc tcattgccag cttcagggag	600
ctggagcagc cgctggagca ccccgtagc ggcgagccag ccacgtgat gaccttcctc	660
atctcctact ggggtgagca gatcggacag aagatccgca agatcacgga ctgcttccac	720
tgccacgtct tcccgtttct gcagcaggag gaggcccgc tcggggccct gcagcagctg	780
caacagcaga gccaggagct gcaggaggtc ctcggggaga cagagcggtt cctgagccag	840
gtgctaggcc ggggtgctga gctgctgccg ccagggcagg tgcaggtcca caagatgaag	900
gccgtgtacc tggccctgaa ccagtgcagc gtgagacca cgcacaagtg cctcattgcc	960
gaggcctggg gctctgtgcg agacctgcc gccctgcagg aggccctgcg ggacagctcg	1020
atggaggagg gagtgagtgc cgtggctcac cgcateccct gccgggacat gccccccaca	1080
ctcatccgca ccaaccgctt cacggccagc ttccagggca tcgtggatgc ctacggcgtg	1140
ggccgctacc aggaggtaa ccccgtccc tacaccatea teacctccc ctctctgttt	1200
gctgtgatgt tcggggatgt gggccacggg ctgctcatgt tctctttegc cctggccatg	1260
gtccttgccg agaaccgacc ggetgtgaag gccgcgcaga acgagatctg gcagactttc	1320
ttcaggggcc gctacctget cctgettatg ggectgttct ccactacac cggcttccatc	1380
tacaacgagt gcttcagtcg cgccaccagc atcttcccct cgggctggag tgtggccgcc	1440
atggccaacc agtctggctg gagtgatgca ttcttgcccc agcacacgat gcttacctg	1500
gacccaacg tcaccggtgt ctctctggga ccctaccct ttggcatcga tctattttgg	1560

agcctggctg ccaaccactt gagcttcctc aactcettca agatgaagat gtccgtcatc 1620
 ctgggctg tgcacatggc ctttgggtg gtctcggag tcttcaacca cgtgcacttt 1680
 ggccagaggc accggctgct gctggagacg ctgccggagc tcaccttct gctgggactc 1740
 ttcggttacc tcgtgttctt agtcatctac aagtggctgt gtgtctgggc tgccagggcc 1800
 gcctcggccc ccagcatcct catccacttc atcaacatgt tcctcttctc ccacagcccc 1860
 agcaacaggc tgctctaccc ccggcaggag gtggtccagg ccacgctggt ggtcctggcc 1920
 ttggccatgg tgcccatcct gctgcttggc acaccctgc acctgctgca ccgccaccgc 1980
 cgccgcctgc ggaggaggcc cgctgaccga caggaggaaa acaaggccgg gttgctggac 2040
 ctgcctgacg catctgtgaa tggtgaggc tccgatgagg aaaaggcagg gggcctggat 2100
 gatgaagagg aggccgagct cgtccctec gaggtgctca tgcaccaggc catccacacc 2160
 atcgagttct gcctgggtg cgtctccaac accgctctct acctgcgct gtgggacctg 2220
 agcctggccc acgcccagct gtccgaggtt ctgtgggcca tggatgatgc cataggcctg 2280
 ggctgggccc gggagggtgg cggtgggct gtggtgctgg tccccatctt tgccgcttt 2340
 gccgtgatga ccgtggctat cctgctggtg atggaggac tctcagcctt cctgcacgcc 2400
 ctgcggtgc actgggtgga attccagaac aagttctact caggcacggg ctacaagctg 2460
 agtcccttca ccttcgctgc cacagatgac tag 2493

<210> 4

<211> 606

<212> DNA

<213> 土拨鼠肝炎病毒

<400> 4

attcgagcat cttaccgcca ttataacca tatttgttct gtttttcttg atttgggtat 60
 acatttaaat gtaataaaa caaatggtg gggcaatcat ttacattttt aggatattgt 120
 aattactagt tcagggtgat tgccacaaga caacatggt aagaaacttt cccgttattt 180
 acgctctggt cctgttaatc aacctctgga ttacaaaatt tgtgaaagat tgactgatat 240
 tcttaactat gttgctcctt ttacgctgtg tggatattgt gctttaatgc ctctgtatca 300
 tgctattgct tcccgtacgg ctttcgtttt ctctccttg tataaatcct gtttctgtc 360
 tctttatgag gaggttgtgc ccgttgtccg tcaacgtggc gtggtgtgct ctgtgtttgc 420
 tgacgcaacc cccactggct ggggcattgc caccactgt caactcctt ctgggacttt 480
 cgctttcccc ctcccgatcg ccacggcaga actcatgcc gcctgccttg cccgctgctg 540
 gacaggggct aggttctggt gcaactgataa ttccgtggtg ttgtcgggga agctgacgtc 600
 ctttcg 606

<210> 5

<211> 830

<212> PRT

<213> 智人

<400> 5

Met Gly Ser Met Phe Arg Ser Glu Glu Val Ala Leu Val Gln Leu Phe

1

5

10

15

Leu Pro Thr Ala Ala Ala Tyr Thr Cys Val Ser Arg Leu Gly Glu Leu
 20 25 30
 Gly Leu Val Glu Phe Arg Asp Leu Asn Ala Ser Val Ser Ala Phe Gln
 35 40 45
 Arg Arg Phe Val Val Asp Val Arg Arg Cys Glu Glu Leu Glu Lys Thr
 50 55 60
 Phe Thr Phe Leu Gln Glu Glu Val Arg Arg Ala Gly Leu Val Leu Pro
 65 70 75 80
 Pro Pro Lys Gly Arg Leu Pro Ala Pro Pro Pro Arg Asp Leu Leu Arg
 85 90 95
 Ile Gln Glu Glu Thr Glu Arg Leu Ala Gln Glu Leu Arg Asp Val Arg
 100 105 110
 Gly Asn Gln Gln Ala Leu Arg Ala Gln Leu His Gln Leu Gln Leu His
 115 120 125
 Ala Ala Val Leu Arg Gln Gly His Glu Pro Gln Leu Ala Ala Ala His
 130 135 140
 Thr Asp Gly Ala Ser Glu Arg Thr Pro Leu Leu Gln Ala Pro Gly Gly
 145 150 155 160
 Pro His Gln Asp Leu Arg Val Asn Phe Val Ala Gly Ala Val Glu Pro
 165 170 175
 His Lys Ala Pro Ala Leu Glu Arg Leu Leu Trp Arg Ala Cys Arg Gly
 180 185 190
 Phe Leu Ile Ala Ser Phe Arg Glu Leu Glu Gln Pro Leu Glu His Pro
 195 200 205
 Val Thr Gly Glu Pro Ala Thr Trp Met Thr Phe Leu Ile Ser Tyr Trp
 210 215 220
 Gly Glu Gln Ile Gly Gln Lys Ile Arg Lys Ile Thr Asp Cys Phe His
 225 230 235 240
 Cys His Val Phe Pro Phe Leu Gln Gln Glu Glu Ala Arg Leu Gly Ala
 245 250 255
 Leu Gln Gln Leu Gln Gln Gln Ser Gln Glu Leu Gln Glu Val Leu Gly
 260 265 270
 Glu Thr Glu Arg Phe Leu Ser Gln Val Leu Gly Arg Val Leu Gln Leu
 275 280 285
 Leu Pro Pro Gly Gln Val Gln Val His Lys Met Lys Ala Val Tyr Leu
 290 295 300
 Ala Leu Asn Gln Cys Ser Val Ser Thr Thr His Lys Cys Leu Ile Ala
 305 310 315 320
 Glu Ala Trp Cys Ser Val Arg Asp Leu Pro Ala Leu Gln Glu Ala Leu

	325		330		335
Arg Asp Ser Ser Met Glu Glu Gly Val Ser Ala Val Ala His Arg Ile					
	340		345		350
Pro Cys Arg Asp Met Pro Pro Thr Leu Ile Arg Thr Asn Arg Phe Thr					
	355		360		365
Ala Ser Phe Gln Gly Ile Val Asp Ala Tyr Gly Val Gly Arg Tyr Gln					
	370		375		380
Glu Val Asn Pro Ala Pro Tyr Thr Ile Ile Thr Phe Pro Phe Leu Phe					
385		390		395	400
Ala Val Met Phe Gly Asp Val Gly His Gly Leu Leu Met Phe Leu Phe					
	405		410		415
Ala Leu Ala Met Val Leu Ala Glu Asn Arg Pro Ala Val Lys Ala Ala					
	420		425		430
Gln Asn Glu Ile Trp Gln Thr Phe Phe Arg Gly Arg Tyr Leu Leu Leu					
	435		440		445
Leu Met Gly Leu Phe Ser Ile Tyr Thr Gly Phe Ile Tyr Asn Glu Cys					
	450		455		460
Phe Ser Arg Ala Thr Ser Ile Phe Pro Ser Gly Trp Ser Val Ala Ala					
465		470		475	480
Met Ala Asn Gln Ser Gly Trp Ser Asp Ala Phe Leu Ala Gln His Thr					
	485		490		495
Met Leu Thr Leu Asp Pro Asn Val Thr Gly Val Phe Leu Gly Pro Tyr					
	500		505		510
Pro Phe Gly Ile Asp Pro Ile Trp Ser Leu Ala Ala Asn His Leu Ser					
	515		520		525
Phe Leu Asn Ser Phe Lys Met Lys Met Ser Val Ile Leu Gly Val Val					
	530		535		540
His Met Ala Phe Gly Val Val Leu Gly Val Phe Asn His Val His Phe					
545		550		555	560
Gly Gln Arg His Arg Leu Leu Leu Glu Thr Leu Pro Glu Leu Thr Phe					
	565		570		575
Leu Leu Gly Leu Phe Gly Tyr Leu Val Phe Leu Val Ile Tyr Lys Trp					
	580		585		590
Leu Cys Val Trp Ala Ala Arg Ala Ala Ser Ala Pro Ser Ile Leu Ile					
	595		600		605
His Phe Ile Asn Met Phe Leu Phe Ser His Ser Pro Ser Asn Arg Leu					
	610		615		620
Leu Tyr Pro Arg Gln Glu Val Val Gln Ala Thr Leu Val Val Leu Ala					
625		630		635	640

gagtttggca acatatgccc atatgctggc tgccatgaac aaaggttggc tataaagagg	180
tcatacagtat atgaaacagc cccctgctgt ccattcetta ttccatagaa aagccttgac	240
ttgaggttag atttttttta tattttgitt tgtgttattt ttttctttaa catccctaaa	300
attttcctta catgttttac tagccagatt tttcctctc tctgactac tcccagtcac	360
agctgtccct cttctcttat ggagatc	387
<210> 8	
<400> 8	
000	
<210> 9	
<400> 9	
000	
<210> 10	
<400> 10	
000	
<210> 11	
<400> 11	
000	
<210> 12	
<400> 12	
000	
<210> 13	
<400> 13	
000	
<210> 14	
<211> 13	
<212> RNA	
<213> 人工序列	
<220>	
<223> 共有Kozak序列	
<400> 14	
gccgcccrrca ugg	13
<210> 15	
<211> 8	
<212> RNA	
<213> 人工序列	
<220>	
<223> Kozak序列	
<220>	
<221> misc_feature	

<222>	(2) .. (3)	
<223>	n为A、C、T、G或U	
<220>		
<221>	misc_feature	
<222>	(4) .. (4)	
<223>	n为a、c、g或u	
<220>		
<221>	misc_feature	
<222>	(8) .. (8)	
<223>	n为A、C、T、G或U	
<400>	15	
agnnaugn		8
<210>	16	
<211>	7	
<212>	RNA	
<213>	人工序列	
<220>		
<223>	Kozak序列	
<220>		
<221>	misc_feature	
<222>	(2) .. (3)	
<223>	n为A、C、T、G或U	
<400>	16	
annaugg		7
<210>	17	
<211>	7	
<212>	RNA	
<213>	人工序列	
<220>		
<223>	Kozak序列	
<400>	17	
accaugg		7
<210>	18	
<211>	10	
<212>	RNA	
<213>	人工序列	
<220>		
<223>	n为A、C、T、G或U	
<400>	18	

gacaccaugg	10
<210> 19	
<211> 235	
<212> DNA	
<213> 牛	
<400> 19	
tcgactgtgc cttctagttg ccagccatct gttgtttgcc cctccccctg gccttccttg	60
accctggaag gtgccactcc cactgtcctt tectaataaa atgaggaaat tgcacatgcat	120
tgtctgagta ggtgtcattc tattctgggg ggtgggggtg ggaggacag caagggggag	180
gattgggagg acaatagcag gcatgctggg gatgcgggtg gctctatggc ttctg	235
<210> 20	
<211> 222	
<212> DNA	
<213> 猿猴病毒40	
<400> 20	
cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaa	60
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca	120
ataaacaagt taacaacaac aattgcattc attttatggt tcaggttcag ggggagatgt	180
gggaggtttt ttaaagcaag taaaacctct acaaatgtgg ta	222
<210> 21	
<211> 202	
<212> DNA	
<213> 智人	
<400> 21	
ctgcccgggt ggcatcccctg tgaccctcc ccagtgcctc tcttgccct ggaagttgcc	60
actccagtgc ccaccagcct tgtcctaata aaattaagtt gcatcatttt gtctgactag	120
gtgtccttct ataataattat ggggtggagg ggggtggtat ggagcaagg gccaagttg	180
ggaagaaacc tgtaggcct gc	202
<210> 22	
<211> 141	
<212> DNA	
<213> 大肠杆菌	
<400> 22	
gtagaattgg taaagagagt cgtgtaaaat atcgagttcg cacatcttgt tgtctgatta	60
ttgatttttg gcgaaacat ttgatcatat gacaagatgt gtatctacct taacttaatg	120
atthtgataa aatcattag g	141
<210> 23	
<211> 7660	
<212> DNA	

<213> 人工序列

<220>

<223> 实验室制备 - 质粒构建体

<400> 23

```

acattgatta ttgactagtt attaatagta atcaattacg gggtcattag ttcataagccc      60
atatatggag ttccgcgta cataacttac ggtaaattggc cgcctggct gaccgcccaa      120
cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caataggagc      180
tttccattga cgtcaatggg tggagtattt acggtaaact gccacttgg cagtacatca      240
agtgtatcat atgccaagta cgccccctat tgacgtcaat gacggtaaatt ggccccctg      300
gcattatgcc cagtacatga ccttatggga ctttctact tggcagtaca tctacgtatt      360
agtcacgct attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg      420
gtttgactca cggggatttc caagtctcca cccattgac gtcaatggga gtttgttttg      480
gcacaaaaat caacgggact ttcaaaaatg tcgtaacaac tccgccccat tgacgcaaat      540
gggcggtagg cgtgtacggt gggaggtcta tataagcaga gctcgtttag tgaaccgggg      600
tctctctggg tagaccagat ctgagcctgg gagctctctg gctaactagg gaaccactg      660
cttaagcctc aataaagctt gccttgagtg cttcaagtag tgtgtgcccg tctgttgtgt      720
gactctggta actagagatc cctcagacc ttttagtcag tgtggaaaat ctctagcagt      780
ggcgccccgaa cagggacttg aaagcgaag gaaaccaga ggagctctct cgacgcagga      840
ctcggcttgc tgaagcgcgc acggcaagag gcgaggggcg gcgactggtg agtacgcaa      900
aaattttgac tagcggaggc tagaaggaga gagatgggtg cgagagcgtc agtattaagc      960
gggggagaat tagatcgcga tgggaaaaaa ttcggttaag gccaggggga aagaaaaaat     1020
ataaattaaa acatatagta tgggcaagca gggagctaga acgattcgca gttaatcctg     1080
gcctgttaga aacatcagaa ggctgtagac aaatactggg acagctacaa ccatcccttc     1140
agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc     1200
atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa     1260
acaaaagtaa gaccaccgca cagcaagcgg ccgctgatct tcagacctgg aggaggagat     1320
atgagggaca attggagaag tgaattatat aatataaag tagtaaaaat tgaaccatta     1380
ggagtagcac ccaccaaggc aaagagaaga gtggtgcaga gagaaaaaag agcagtggga     1440
ataggagctt tgttccttgg gttcttggga gcagcaggaa gcactatggg cgcagcgtca     1500
atgacgctga cgttacaggc cagacaatta ttgtctggta tagtgcagca gcagaacaat     1560
ttgctgaggg ctattgaggc gcaacagcat ctggtgcaac tcacagtctg gggcatcaag     1620
cagctccagg caagaatcct ggctgtggaa agatacctaa aggatcaaca gctcctgggg     1680
atitgggggt gctctgaaa actcatttgc accactgctg tgcttggaa tgctagttag     1740
agtaataaat ctctggaaca gatttggaa cacacgacct ggatggagtg ggacagagaa     1800
attaacaatt acacaagctt aatacactcc ttaattgaag aatcgcaaaa ccagcaagaa     1860
aagaatgaac aagaattatt ggaattagat aatgggcaa gtttgtggaa ttggtttaac     1920
ataacaaatt ggctgtggta tataaaatta ttcataatga tagtaggagg cttggtaggt     1980
ttaagaatag tttttgctgt actttctata gtgaatagag ttaggcaggg atattcacca     2040
ttatcgtttc agaccacct cccaaccccg aggggaccg acaggcccga aggaatagaa     2100

```

gaagaaggtg	gagagagaga	cagagacaga	tccattcgat	tagtgaacgg	atctcgacgg	2160
tatcggttaa	cttttaaaag	aaaagggggg	attggggggg	acagtgcagg	ggaaagaata	2220
gtagacataa	tagcaacaga	catacaaaact	aaagaattac	aaaaacaaat	tacaaaaatt	2280
caaaatttta	tcgatcacga	gactagcctc	gagaagcttg	atcgattggc	tccggtgccc	2340
gtcagtgggc	agagcgcaca	tcgcccacag	tccccgagaa	gttgggggga	ggggtcggca	2400
attgaaccgg	tgccntagaga	aggtggcgcg	gggtaaactg	ggaaagtgat	gtcgtgtact	2460
ggctccgcct	ttttcccag	ggtgggggag	aaccgtatat	aagtgcagta	gtcgcctgta	2520
acgttctttt	tcgcaacggg	tttgccgcca	gaacacaggt	gtcgtgacgc	gggatccgcc	2580
accatgggct	ccatgtttcg	gagcggaggag	gtggccctgg	tccagctctt	tctgccaca	2640
gcggctgcct	acacctgcgt	gagtcggctg	ggcgagctgg	gcctcgtgga	gttcagagac	2700
ctcaacgcct	cggtgagcgc	cttcagaga	cgctttgtgg	ttgatgttcg	gcgctgtgag	2760
gagctggaga	agacctcac	cttctgcag	gaggaggtgc	ggcgggctgg	gctggtcctg	2820
ccccgccaa	aggggagget	gccggcacc	ccaccccg	acctgctgcg	catccaggag	2880
gagacggagc	gcctggccca	ggagctgcgg	gatgtgcggg	gcaaccagca	ggccctgcgg	2940
gccagctgc	accagctgca	gctccacgcc	gccgtgctac	gccaggcca	tgaacctcag	3000
ctggcagccg	cccacacaga	tggggcctca	gagaggacgc	ccctgctcca	ggccccggg	3060
gggccgcacc	aggacctgag	ggtcaacttt	gtggcaggtg	ccgtggagcc	ccacaaggcc	3120
cctgccctag	agcgcctgct	ctggagggcc	tgcagaggct	tcctcattgc	cagcttcagg	3180
gagctggagc	agccgctgga	gcaccccg	acggcgagc	cagccacgtg	gatgaccttc	3240
ctcatctcct	actgggggtga	gcagatcgga	cagaagatcc	gcaagatcac	ggactgcttc	3300
cactgccacg	tcttcccgtt	tctgcagcag	gaggaggccc	gcctcggggc	cctgcagcag	3360
ctgcaacagc	agagccagga	gctgcaggag	gtcctcgggg	agacagagcg	gttctgagc	3420
caggtgctag	gccgggtgct	gcagctgctg	ccgccagggc	aggtgcaggt	ccacaagatg	3480
aaggccgtgt	acctggccct	gaaccagtgc	agcgtgagca	ccacgcaca	gtgcctcatt	3540
gccgaggcct	ggtgctctgt	gcgagacctg	cccgcctgc	aggaggcct	gcgggacagc	3600
tcgatggagg	agggagtgag	tgccgtggct	caccgcatcc	cctgccggga	catgcccccc	3660
acactcatcc	gcaccaaccg	cttcacggcc	agcttcagg	gcatcgtgga	tgcttacggc	3720
gtgggccgct	accaggaggt	caaccccgct	ccctacacca	tcatacctt	ccccttctg	3780
tttgctgtga	tgttcgggga	tgtgggccac	ggctgctca	tgcttctctt	cgccctggcc	3840
atggtccttg	cggagaaccg	accggtgtg	aaggccgcgc	agaacgagat	ctggcagact	3900
ttcttcaggg	gccgctacct	gctctgctt	atgggctgt	tctccatcta	caccggcttc	3960
atctacaacg	agtgetttag	tcgcgccacc	agcatcttc	cctcgggctg	gagtggtggc	4020
gcatggcca	accagtctgg	ctggagtgat	gcattcttgg	cccagcacac	gatgcttacc	4080
ctggaccca	acgtcaccgg	tgttctctg	ggacctacc	cctttggcat	cgatcctatt	4140
tggagcctgg	ctgccaacca	cttgagcttc	ctcaactct	tcaagatgaa	gatgtccgtc	4200
atcctgggcg	tcgtgacat	ggcctttggg	gtggtctctg	gagtcttcaa	ccacgtgcac	4260
tttggccaga	ggcaccgget	gctgctggag	acgctgccgg	agctcacctt	cctgctggga	4320
ctcttcgggt	acctcgtgtt	cctagtcctc	tacaagtggc	tgtgtgtctg	ggctgccagg	4380
gccgcctcgg	ccccagcat	cctcatccac	ttcatcaaca	tgttctctt	ctcccacagc	4440

cccagcaaca	ggctgctcta	ccccggcag	gagtggtcc	aggccacgct	ggtggtcctg	4500
gccttggcca	tggtgcccat	cctgctgctt	ggcacacccc	tgcacctgct	gcaccgccac	4560
cgccgccgcc	tgcggaggag	gcccgtgac	cgacaggagg	aaaacaaggc	cgggttgctg	4620
gacctgcctg	acgcatctgt	gaatggctgg	agctccgatg	aggaaaaggc	agggggcctg	4680
gatgatgaag	aggaggccga	gctcgtcccc	tccgaggtgc	tcatgcacca	ggccatccac	4740
accatcgagt	tctgcctggg	ctgcgtctcc	aacaccgctt	cctacctgcg	cctgtgggcc	4800
ctgagcctgg	cccacgcccc	gctgtccgag	gttctgtggg	ccatggtgat	gcgcataggc	4860
ctgggcctgg	gccgggaggt	gggcgtggcg	gctgtggtgc	tggccccat	ctttgccgcc	4920
tttgccgtga	tgaccgtggc	tatcctgctg	gtgatggagg	gactctcagc	cttctgcac	4980
gccctgcggc	tgactgggt	ggaattccag	acaagttct	actcaggcac	gggctacaag	5040
ctgagtcctt	tcaccttcgc	tgccacagat	gactagtaag	tcgacggatc	ccccgggctg	5100
caggaattcg	agcatcttac	cgccatttat	accatattt	gttctgtttt	tcttgatttg	5160
ggtatacatt	taaatgtaa	taaaacaaaa	tggtagggca	atcatttaca	tttttaggga	5220
tatgtaatta	ctagttcagg	tgtattgcca	caagacaaac	atgtaagaa	actttcccgt	5280
tatttacgct	ctgttctctg	taatcaacct	ctggattaca	aaatttgatg	aagattgact	5340
gatattctta	actatgttgc	tccttttacg	ctgtgtggat	atgctgcttt	aatgcctctg	5400
tatcatgcta	ttgcttcccg	tacggctttc	gtttctctct	ccttgataaa	atcctggttg	5460
ctgtctcttt	atgaggagtt	gtggcccgtt	gtccgtcaac	gtggcgtggt	gtgctctgtg	5520
tttctgacg	caacccccac	tggttggggc	attgccacca	cctgtcaact	cctttctggg	5580
actttcgctt	tccccctccc	gatcgccacg	gcagaactca	tcgccgctg	ccttgcccgc	5640
tgctggacag	gggctaggtt	gctgggcact	gataattccg	tgggtttgtc	ggggaagctg	5700
acgtcctttc	gaattcgata	tcaagctgta	cctttaagac	caatgactta	caaggcagct	5760
gtagatctta	gccacttttt	aaaagaaaag	gggggactgg	aagggtaat	tactcccaa	5820
cgaagacaag	atctgctttt	tgcttgact	gggtctctct	ggttagacca	gatctgagcc	5880
tgggagctct	ctggctaact	agggaacca	ctgcttaagc	ctcaataaag	cttgccctga	5940
gtgcttcaag	tagtgtgtgc	ccgtctgttg	tgtgactctg	gtaactagag	atccctcaga	6000
cccttttagt	cagtgtggaa	aatctctagc	agtagtagtt	catgtcatct	tattattcag	6060
tatttataac	ttgcaaagaa	atgaatatca	gagagtgaga	ggaacttggt	tattgcagct	6120
tataatgggt	acaaataaag	caatagcatc	acaatttca	caaataaagc	atttttttca	6180
ctgcattcta	gttgtggttt	gtccaaactc	atcaatgtat	cttatcatgt	ctggctctag	6240
ctatcccgcc	cctaactccg	cccateccgc	ccctaactcc	gcccagttcc	gcccattctc	6300
cgccccatgg	ctgactaatt	ttttttat	atgcagagge	cgaggccgcc	tcggcctctg	6360
agctattcca	gaagtagtga	ggaggttttt	ttggaggcct	aggtagcccc	cctaattgagc	6420
gggctttttt	ttcttaggcc	ttcttccgct	tctctgctca	ctgactcctg	gcgctcggtc	6480
gttcggctgc	ggcgagcggg	atcagctcac	tcaaaggcgg	taatacgggt	atccacagaa	6540
tcaggggata	acgcaggaaa	gaacatgtga	gcaaaaaggc	agcaaaaaggc	caggaaccgt	6600
aaaaaggccg	cgttgctggc	gtttttccat	aggctccgcc	cccctgacga	gcatacaaaa	6660
aatcgacgct	caagtcagag	gtggcgaaac	ccgacaggac	tataaagata	ccaggcgttt	6720
ccccctggaa	gctccctcgt	gcgctctcct	gttccgacct	tgccgcttac	cggatacctg	6780

tccgcctttc tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc	6840
agttcgggtgt aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc	6900
gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccgtaag acacgactta	6960
tcgccactgg cagcagccac tggtaacagg attagcagag cgaggtatgt aggcggtgct	7020
acagagttct tgaagtgggt gcctaactac ggctacacta gaagaacagt atttggtatc	7080
tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa	7140
caaaccaccg ctggtagcgg tggttttttt gtttgcaagc agcagattac gcgcagaaaa	7200
aaaggatctc aagaagatcc tttgatcttt tctacggggt ctgacgctca gtggaacgaa	7260
aactcacggt aagggatfff ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt	7320
ttaaattaaa aatgaagttt taaatcaatc taaagtatat atgagtaaac ttggtctgac	7380
agttaccaat gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcatcc	7440
atagttgcct gactcctgca aaccacgttg tggtagaatt ggtaaagaga gtcgtgtaaa	7500
atatcgagtt cgcacatctt gttgtctgat tattgatttt tggcgaaacc atttgatcat	7560
atgacaagat gtgtatctac cttacttaa tgattttgat aaaaatcatt aggtacctgt	7620
acatttatat tggctcatgt ccaacattac cgccatgttg	7660
<210> 24	
<211> 511	
<212> DNA	
<213> 智人	
<400> 24	
ggggttgggg ttgcgccttt tccaaggcag ccctgggttt gcgcaggac gcggctgctc	60
tgggcgtggg tccgggaaac gcagcggcgc cgaccctggg tctcgcacat tcttcacgtc	120
cgttcgcagc gtcaccgga tcttcgccgc tacccttggt ggcccccg cgacgcttcc	180
tgctccgccc ctaagtcggg aaggttcctt gcggttcgcg gcgtgccgga cgtgacaaac	240
ggaagccgca cgtctcacta gtaccctcgc agacggacag cgccaggag caatggcagc	300
gcgccgaccg cgatgggctg tggccaatag cggctgctca gcagggcgcg ccgagagcag	360
cggccgggaa ggggcggtgc gggaggcggg gtgtggggcg gtagtgtggg ccctgttctc	420
gcccgcgcgg tgttccgcat tctgcaagcc tccggagcgc acgtcggcag tcggctcctc	480
cgttgaccga atcaccgacc tctctcecca g	511
<210> 25	
<211> 7613	
<212> DNA	
<213> 人工序列	
<220>	
<223> 实验室制备 - 质粒构建体	
<400> 25	
acattgatta ttgactagtt attaatagta atcaattac gggtcattag ttcatagccc	60
atatatggag ttccgcgta cataacttac ggtaaatggc ccgcctggct gaccgcccac	120
cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac	180

tttccattga	cgatcaatggg	tggagtattt	acggtaaact	gcccacttgg	cagtacatca	240
agtgtatcat	atgccaaagta	cgccccctat	tgacgtcaat	gacggtaa	ggccccgcctg	300
gcattatgcc	cagtacatga	ccttatggga	ctttcctact	tggcagtaca	tctacgtatt	360
agtcacgcct	attaccatgg	tgatgcgggt	ttggcagtac	atcaatgggc	gtggatagcg	420
gtttgactca	cggggatttc	caagtctcca	ccccattgac	gtcaatggga	gtttgttttg	480
gcaccaaaaat	caacgggact	ttccaaaatg	tcgtaacaac	tccgccccat	tgacgcaaat	540
gggcggtagg	cggtgtacgg	gggagggtcta	tataagcaga	gctcgtttag	tgaaccgggg	600
tctctctggg	tagaccagat	ctgagcctgg	gagctctctg	gctaactagg	gaaccactg	660
cttaagcctc	aataaagctt	gccttgagtg	cttcaagtag	tgtgtgcccc	tctgtttgtg	720
gactctggta	actagagatc	cctcagacc	ttttagtcag	tgtggaaaat	ctctagcagt	780
ggcggccgaa	cagggacttg	aaagcgaag	ggaaaccaga	ggagctctct	cgacgcagga	840
ctcggcttgc	tgaagcgcgc	acggcaagag	gcgagggg	gcgactgggtg	agtacgcca	900
aaattttgac	tagcggaggc	tagaaggaga	gagatgggtg	cgagagcgtc	agtattaagc	960
gggggagaat	tagatcgcca	tgggaaaaaa	ttcggttaag	gccaggggga	aagaaaaat	1020
ataaattaaa	acatatagta	tgggcaagca	gggagctaga	acgattcgca	gttaatcctg	1080
gcctgttaga	aacatcagaa	ggctgttagc	aaatactggg	acagctaca	ccatcccttc	1140
agacaggatc	agaagaactt	agatcattat	ataatacagt	agcaaccctc	tattgtgtgc	1200
atcaaaggat	agagataaaa	gacaccaagg	aagctttaga	caagatagag	gaagagcaaa	1260
acaaaagtaa	gaccaccgca	cagcaagcgg	ccgctgatct	tcagacctgg	aggaggagat	1320
atgagggaca	attggagaag	tgaattatat	aaatataaag	tagtaaaaat	tgaaccatta	1380
ggagtagcac	ccaccaaggc	aaagagaaga	gtggtgcaga	gagaaaaaag	agcagtggga	1440
ataggagcct	tgttccttgg	gttcttggga	gcagcaggaa	gcactatggg	cgagcgtca	1500
atgacgctga	cgggtacaggc	cagacaatta	ttgtctggta	tagtgcagca	gcagaacaat	1560
ttgctgaggg	ctattgaggc	gcaacagcat	ctgttgcaac	tcacagtctg	gggcatcaag	1620
cagctccagg	caagaatcct	ggctgtggaa	agatacctaa	aggatcaaca	gctcctgggg	1680
atttgggggt	gctctggaaa	actcatttgc	accactgctg	tgccctggaa	tgctagtgtg	1740
agtaataaat	ctctggaaca	gatttggaa	cacacgacct	ggatggagtg	ggacagagaa	1800
attaacaatt	acacaagctt	aatacactcc	ttaattgaag	aatcgcaaaa	ccagcaagaa	1860
aagaatgaac	aagaattatt	ggaattagat	aatgggcaa	gtttgtggaa	ttggtttaac	1920
ataacaaatt	ggctgtggta	tataaaatta	ttcataatga	tagtaggagg	cttggtaggt	1980
ttaagaatag	tttttgctgt	actttctata	gtgaatagag	ttaggcaggg	atattcacca	2040
ttatcgtttc	agaccacct	cccaaccccg	aggggaccg	acaggcccga	aggaatagaa	2100
gaagaagggtg	gagagagaga	cagagacaga	tccattcgat	tagtgaacgg	atctcgacgg	2160
tatcggttaa	cttttaaaag	aaaagggggg	attgggggtg	acagtgcagg	ggaaagaata	2220
gtagacataa	tagcaacaga	catacaaaact	aaagaattac	aaaaacaaat	tacaaaaatt	2280
caaaaatttta	tccgtcagtg	ggcagagcgc	acatcgccca	cagtccccga	gaagttgggg	2340
ggaggggtcg	gcaattgaac	cggtgcctag	agaaggtggc	gcggggtaaa	ctgggaaagt	2400
gatgtcgtgt	actggctccg	cctttttccc	gagggtgggg	gagaaccgta	tataagtgca	2460
gtagtcgccg	tgaacgttct	ttttcgcaac	gggtttgccg	ccagaacaca	ggtgtcgtga	2520

cgcgggatcc	gccacatg	gctccatg	tcggagcgag	gaggtggccc	tggtccagct	2580
ctttctgccc	acagcggctg	cctacacctg	ctgagtcgg	ctgggcgagc	tgggcctcgt	2640
ggagttcaga	gacctcaacg	cctcggtag	cgcttcag	agacgcttg	tggttgatgt	2700
tcggcgctgt	gaggagctgg	agaagacctt	caccttcctg	caggaggagg	tgcggcgggc	2760
tgggctggtc	ctgccccgc	caaaggggag	gctgccggca	ccccacccc	gggacctgct	2820
gcgcatccag	gaggagacgg	agcgcctggc	ccaggagctg	cgggatgtgc	ggggcaacca	2880
gcaggccctg	cgggcccagc	tgaccagct	gcagctccac	gccgccgtgc	taccagagg	2940
ccatgaacct	cagctggcag	ccgcccacac	agatggggcc	tcagagagga	cgcccctgct	3000
ccaggccccc	ggggggccgc	accaggacct	gagggtcaac	tttgtggcag	gtgccgtgga	3060
gccccacaag	gcccctgccc	tagagcgctt	gctctggagg	gcctgcagag	gcttcctcat	3120
tgccagcttc	agggagctgg	agcagccgct	ggagcacc	gtgacgggcg	agccagccac	3180
gtggatgacc	ttcctcatct	cctaactggg	tgagcagatc	ggacagaaga	tcgcaagat	3240
cacggactgc	ttcactgcc	acgtctccc	gtttctgcag	caggaggagg	cccgcctcgg	3300
ggccctgcag	cagctgcaac	agcagagcca	ggagctgcag	gaggtcctcg	gggagacaga	3360
gcggttctg	agccaggtgc	taggccgggt	gctgcagctg	ctgccgccag	ggcaggtgca	3420
ggtccacaag	atgaaggccg	tgtacctggc	cctgaaccag	tgacgcgtga	gcaccacgca	3480
caagtgcctc	attgccgagg	cctggtgctc	tgtgcgagac	ctgcccgcc	tgccagaggc	3540
cctgcgggac	agctcgatgg	aggaggaggt	gagtgccgtg	gctcaccgca	tccttgcctg	3600
ggacatgccc	cccacactca	tccgcaccaa	ccgcttcacg	gccagcttcc	agggcacgtt	3660
ggatgcctac	ggcgtggg	gctaccagga	ggtcaacccc	gctccctaca	ccatcatcac	3720
cttccccttc	ctgtttgctg	tgatgttcgg	ggatgtgggc	cacgggctgc	tcatgttctt	3780
cttcgccctg	gccatggtcc	ttgcggagaa	ccgaccggct	gtgaaggccg	cgcagaacga	3840
gatctggcag	actttcttca	ggggccgcta	cctgctcctg	cttatgggcc	tgttctccat	3900
ctacaccggc	ttcatctaca	acgagtgtt	cagtcgcgcc	accagcatct	tccttgcggg	3960
ctggagtgtg	gccgccatgg	ccaaccagtc	tggtctggagt	gatgcattcc	tgcccagca	4020
cacgatgctt	acctggacc	ccaacgtcac	cggtgtcttc	ctgggaccct	accctttgg	4080
catcgatcct	atttgagacc	tggtgccaa	ccacttgagc	ttcctcaact	ccttcaagat	4140
gaagatgtcc	gtcatcctgg	gcgtcgtgca	catggccttt	gggtgtgtcc	tcggagtctt	4200
caaccacgtg	cactttggcc	agaggcaccg	gctgctgctg	gagacgctgc	cggagctcac	4260
cttctcctg	ggactcttcg	gttacctcgt	gttctagtc	atctacaagt	ggctgtgtgt	4320
ctgggctgcc	agggccgctt	cggccccag	catctcctc	cacttcatca	acatgttctt	4380
cttctcccac	agccccagca	acaggtgct	ctacccccg	caggagggtg	tccagccac	4440
gctggtggtc	ctggccttgg	ccatggtgcc	catctgctg	cttggcacac	ccctgcacct	4500
gctgcaccgc	caccgccgc	gctgcccgg	gagggccgct	gaccgacagg	aggaaaacaa	4560
ggccgggttg	ctggacctgc	ctgacgcctc	tgtgaatggc	tgagctccg	atgaggaaaa	4620
ggcagggggc	ctggatgatg	aagaggaggc	cgagctcgtc	ccctccgagg	tgctcatgca	4680
ccaggccatc	cacaccatcg	agttctgcct	gggctgcgtc	tccaacaccg	cctctacct	4740
gcgcctgtgg	gccctgagcc	tgggccacgc	ccagctgtcc	gaggttctgt	gggccatggt	4800
gatgcgcata	ggcctgggcc	tgggccggga	ggtgggcgtg	gcggctgtgg	tgctggtccc	4860

catctttgcc	gcctttgccg	tgatgaccgt	ggctatcctg	ctggtgatgg	agggactctc	4920
agccttcctg	cacgccctgc	ggctgactg	ggtggaattc	cagaacaagt	tctactcagg	4980
cacgggctac	aagctgagtc	ccttcacctt	cgctgccaca	gatgactagt	aagtcgacgg	5040
atcccccggg	ctgcaggaat	tcgagcatct	taccgccatt	tatacccata	tttgttctgt	5100
ttttcttgat	ttgggtatac	atttaaagt	taataaaaca	aatggtggg	gcaatcattt	5160
acatttttag	ggatatgtaa	ttactagttc	aggtgtattg	ccacaagaca	aacatgttaa	5220
gaaactttcc	cgttatttac	gctctgttcc	tgtaaataca	cctctggatt	acaaaatttg	5280
tgaaagattg	actgatattc	ttaactatgt	tgctcctttt	acgctgtgtg	gatatgctgc	5340
tttaatgcct	ctgtatcatg	ctattgcttc	ccgtacggct	ttcgttttct	cctccttgta	5400
taaactcctgg	ttgctgtctc	tttatgagga	gttgtggccc	gttgtccgtc	aacgtggcgt	5460
ggtgtgctct	gtgtttgctg	acgcaacccc	cactggctgg	ggcattgcca	ccacctgtca	5520
actcctttct	gggactttcg	ctttccccct	cccgatcgcc	acggcagaac	tcategccgc	5580
ctgccttgcc	cgctgctgga	caggggctag	gttctgtggc	actgataatt	ccgtggtggt	5640
gtcggggaag	ctgacgtcct	ttcgaattcg	atatcaagct	gtacctttaa	gaccaatgac	5700
ttacaaggca	gctgtagatc	ttagccactt	tttaaagaa	aaggggggac	tggaagggtc	5760
aattcactcc	caacgaagac	aagatctgct	ttttgcttgt	actgggtctc	tctggttaga	5820
ccagatctga	gcctgggagc	tctctggcta	actagggaac	ccactgctta	agcctcaata	5880
aagcttgccct	tgagtgcttc	aagtagtgtg	tgcccgtctg	ttgtgtgact	ctggtaacta	5940
gagatccctc	agaccctttt	agtcagtgtg	gaaaatctct	agcagtagta	gttcatgtca	6000
tcttattatt	cagtatttat	aacttgcaaa	gaaatgaata	tcagagagtg	agaggaactt	6060
gtttattgca	gcttataatg	gttaciaaata	aagcaatagc	atcaciaaatt	tcaciaaata	6120
agcatttttt	tactgcatt	ctagtgtg	ttgtccaaa	ctcatcaatg	tatcttatca	6180
tgtctggctc	tagctatccc	gccctaaact	ccgcccctcc	cgcccctaac	tccgcccagt	6240
tccgcccatt	ctccgcccc	tggtgacta	atTTTTTTT	tttatgcaga	ggccgaggcc	6300
gcctcggcct	ctgagctatt	ccagaagtag	tgaggaggct	TTTTTggagg	cctaggtagc	6360
ccgcctaatg	agcgggcttt	TTTTTcttag	gccttcttcc	gcttctctgc	tactgactc	6420
gctgcgctcg	gtcgttcggc	tgcgcgagc	ggtatcagct	cactcaaagg	cggtaatcag	6480
gttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	6540
ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcgttttct	cataggctcc	gccccctga	6600
cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aaccgacag	gactataaag	6660
ataccaggcg	tttccccctg	gaagctcct	cgtgcgctct	cctgttccga	ccctgccgct	6720
taccggatac	ctgtccgct	ttctccctc	gggaagcgtg	gcgctttctc	atagctcacg	6780
ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	6840
ccccgttcag	cccgaccgct	gcgccttate	cggtaaactat	cgtcttgagt	ccaaccgggt	6900
aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	6960
tgtaggcggt	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaagaac	7020
agtatttggg	atctgcgctc	tgetgaagcc	agttaccttc	ggaaaaagag	ttggtagctc	7080
ttgatccggc	aaacaaacca	ccgctggtag	cggtggtttt	tttgtttgca	agcagcagat	7140
tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	ttttctacgg	ggtctgacgc	7200

tcagtggaac gaaaactcac gttaaggat tttggatcatg agattatcaa aaagatcctt	7260
cacctagatc cttttaaatt aaaaatgaag ttttaaatac atctaaagta tatatgagta	7320
aacttggctc gacagttacc aatgcttaac cagtgaggca cctatctcag cgatctgtct	7380
atttcgttca tccatagttg cctgactcct gcaaacacag ttgtggtaga attggtaaag	7440
agagtcgtgt aaaatatcga gttcgacacat cttgtttgtc gattattgat ttttggcgaa	7500
accatttgat catatgacaa gatgtgtatc taccttaact taatgatttt gataaaaatc	7560
attaggatc tgtacattta tattggctca tgtccaacat taccgcatg ttg	7613
<210> 26	
<211> 47	
<212> DNA	
<213> 人工序列	
<220>	
<223> 实验室制备 - 部分质粒构建体	
<400> 26	
gatcacgaga ctacgctcga gaagcttgat cgattggctc cggtgcc	47
<210> 27	
<211> 7646	
<212> DNA	
<213> 人工序列	
<220>	
<223> 实验室制备 - 质粒构建体	
<400> 27	
ggctccggtg cccgtcagtg ggcagagcgc acatcgccca cagtccccga gaagttgggg	60
ggaggggtcg gcaattgaac cgggtgcctag agaaggtggc gcggggtaaa ctgggaaagt	120
gatgtcgtgt actggctccg cctttttccc gaggggtgggg gagaaccgta tataagtga	180
gtagtcgccg tgaacgttct ttttcgcaac gggtttgccg ccagaacaca ggtgtcgtga	240
cgcgggatcc gccaccatgg gctccatggt tcggagcgag gaggtggccc tggctcagct	300
ctttctgccc acagcggctg cctacacctg cgtgagtcgg ctgggcgagc tgggcctcgt	360
ggagttcaga gacctcaacg cctcggtagag cgccttcag agacgctttg tggttgatgt	420
tcggcgctgt gaggagctgg agaagacctt caccttctc caggaggagg tgcggcgggc	480
tgggctggtc ctgccccgc caaaggggag gctgccgga cccccaccc gggacctgct	540
gcgcatccag gaggagacgg agcgcctggc ccaggagctg cgggatgtgc ggggcaacca	600
gcaggccctg cgggcccagc tgcaccagct gcagctccac gccgccgtgc tacgccaggg	660
ccatgaacct cagctggcag ccgcccacac agatggggcc tcagagagga cgcccctgct	720
ccaggcccc ggggggccc accaggacct gagggtaac tttgtggcag gtgccgtgga	780
gccccacaag gcccctgccc tagagcgctt gctctggagg gcctgcagag gcttctctat	840
tgccagcttc agggagctgg agcagccgct ggagcaccac gtgacgggcg agccagccac	900
gtggatgacc ttctctatct cctactgggg tgagcagatc ggacagaaga tccgcaagat	960
cacggactgc ttccactgcc acgtcttccc gtttctgcag caggaggagg cccgcctcgg	1020

ggccctgcag	cagctgcaac	agcagagcca	ggagctgcag	gaggtcctcg	gggagacaga	1080
gcggttcctg	agccagggtc	taggccgggt	gctgcagctg	ctgccgccag	ggcagggtga	1140
ggtccacaag	atgaaggccg	tgtacctggc	cctgaaccag	tgcagcgtga	gcaccacgca	1200
caagtgcctc	attgccgagg	cctggtgctc	tgtgcgagac	ctgcccgcc	tgcaggaggc	1260
cctgcgggac	agctcgatgg	aggagggagt	gagtgccgtg	gctcaccgca	tcccctgccg	1320
ggacatgccc	cccacactca	tccgcaccaa	ccgcttcacg	gccagcttcc	agggcatcgt	1380
ggatgcctac	ggcgtgggcc	gctaccagga	ggtcaacccc	gctcccctaca	ccatcatcac	1440
cttccccctc	ctgtttgctg	tgatgttcgg	ggatgtgggc	cacgggctgc	tcatgttctc	1500
cttcgccctg	gccatgggtc	ttgcggagaa	ccgaccggct	gtgaaggccg	cgcagaacga	1560
gatctggcag	actttcttca	ggggccgcta	cctgctcctg	cttatgggcc	tgttctccat	1620
ctacaccggc	ttcatctaca	acgagtgett	cagtcgcgcc	accagcatct	tcccctcggg	1680
ctggagtgtg	gccgccatgg	ccaaccagtc	tggtctggagt	gatgcattcc	tggcccagca	1740
cacgatgctt	accctggacc	ccaacgtcac	cggtgtcttc	ctgggacct	acccttttgg	1800
catcgatcct	atttggagcc	tggtgccaa	ccacttgagc	ttcctcaact	ccttcaagat	1860
gaagatgtcc	gtcatcctgg	gcgtcgtgca	catggccttt	ggggtggtcc	tcgagatctt	1920
caaccacgtg	cactttggcc	agaggcaccg	gctgctgctg	gagacgctgc	cggagctcac	1980
cttctgctg	ggactcttcg	gttacctcgt	gttcttagtc	atctacaagt	ggctgtgtgt	2040
ctgggctgcc	agggccgcct	cggccccag	catctcctc	cacttcatca	acatgttctc	2100
cttctcccac	agccccagca	acaggctgct	ctacccccgg	caggaggtgg	tccaggccac	2160
gctggtggtc	ctggccttgg	ccatggtgcc	catcctgctg	cttgccacac	ccctgcacct	2220
gctgcaccgc	caccgccgcc	gcctgcggag	gaggcccgc	gaccgacagg	agggaaaaca	2280
ggccgggttg	ctggacctgc	ctgacgcctc	tgtgaatggc	tggagctccg	atgaggaaaa	2340
ggcagggggc	ctggatgatg	aagaggaggc	cgagctcgtc	ccctccgagg	tgctcatgca	2400
ccaggccatc	cacaccatcg	agttctgcct	gggctgcgtc	tccaacaccg	cctcctacct	2460
gcgcctgtgg	gccctgagcc	tggcccacgc	ccagctgtcc	gaggttctgt	gggccatggt	2520
gatgcgcata	ggcctgggcc	tgggccggga	ggtgggcgtg	gcggctgtgg	tgctggtccc	2580
catctttgcc	gcctttgccg	tgatgacctg	ggctatcctg	ctggtgatgg	agggactctc	2640
agccttcctg	cacgccctgc	ggctgcactg	ggtggaattc	cagaacaagt	tctactcagg	2700
cacgggctac	aagctgagtc	ccttcacctt	cgctgccaca	gatgactagt	aagtcgacgg	2760
atcccccggg	ctgcaggaat	tcgagcatct	taccgccatt	tatacccata	tttgttctgt	2820
ttttcttgat	ttgggtatac	atttaaagt	taataaaaca	aatggtggg	gcaatcattt	2880
acattttttag	ggatatgtaa	ttactagttc	aggtgtattg	ccacaagaca	aacatgttaa	2940
gaaactttcc	cgttatttac	gctctgttcc	tgtaataca	cctctggatt	acaaaatttg	3000
tgaagattg	actgatattc	ttaactatgt	tgctcctttt	acgctgtgtg	gatatgctgc	3060
tttaatgcct	ctgtatcatg	ctattgett	ccgtacggct	ttcgttttct	cctccttgta	3120
taaactcctgg	ttgctgtctc	tttatgagga	gttgtggccc	gttgtccgtc	aacgtggcgt	3180
ggtgtgctct	gtgtttgctg	acgcaacccc	cactggctgg	ggcattgcca	ccacctgtca	3240
actcctttct	gggactttcg	ctttccccct	cccgatcgc	acggcagaac	tcatcgccgc	3300
ctgccttgcc	cgctgctgga	caggggctag	gttgctgggc	actgataatt	ccgtggtgtt	3360

gtcggggaag	ctgacgtcct	ttcgaattcg	atatcaagct	gtacctttaa	gaccaatgac	3420
ttacaaggca	gctgtagatc	ttagccactt	tttaaagaa	aaggggggac	tggaagggct	3480
aattcactcc	caacgaagac	aagatctgct	ttttgcttgt	actgggtctc	tctggttaga	3540
ccagatctga	gcctgggagc	tctctggcta	actaggaac	ccactgctta	agcctcaata	3600
aagcttgcct	tgagtgttc	aagtagtgtg	tgcccgtctg	ttgtgtgact	ctggtaacta	3660
gagatccctc	agaccctttt	agtcagtgtg	gaaaatctct	agcagtagta	gttcatgtca	3720
tcttattatt	cagtatttat	aacttgcaaa	gaaatgaata	tcagagagtg	agaggaactt	3780
gtttattgca	gcttataatg	gttaciaaata	aagcaatagc	atcaciaaatt	tcaciaaataa	3840
agcatttttt	tactgtcatt	ctagttgtgg	ttgtccaaa	ctcatcaatg	tatcttatca	3900
tgtctggctc	tagctatccc	gcccctaact	cgcccatacc	cgcccctaac	tccgcccagt	3960
tccgcccatt	ctccgcccga	tggetgacta	atTTTTTTT	tttatgcaga	ggccgaggcc	4020
gcctcggcct	ctgagctatt	ccagaagtag	tgaggaggct	TTTTTggagg	cctaggtagc	4080
ccgctaatag	agcgggcttt	TTTTTcttag	gccttcttcc	gcttctctgc	tactgactc	4140
gctgcgctcg	gtcgttcggc	tgcggcgagc	ggtatcagct	cactcaaagg	cggtaaatcg	4200
gttatccaca	gaatcagggg	ataacgcagg	aaagaacatg	tgagcaaaag	gccagcaaaa	4260
ggccaggaac	cgtaaaaagg	ccgcgttgct	ggcgTTTTT	cataggctcc	gccccctga	4320
cgagcatcac	aaaaatcgac	gctcaagtca	gaggtggcga	aaccgacag	gactataaag	4380
ataccaggcg	tttccccctg	gaagctccct	cgtgcgctct	cctgttccga	ccctgccgct	4440
taccggatac	ctgtccgctt	ttctccctc	gggaagcgtg	gcgctttctc	atagctcacg	4500
ctgtaggtat	ctcagttcgg	tgtaggtcgt	tcgctccaag	ctgggctgtg	tgcacgaacc	4560
ccccgttcag	cccgaccgct	gcgccttatc	cggtaaactat	cgtcttgagt	ccaaccggt	4620
aagacacgac	ttatcgccac	tggcagcagc	cactggtaac	aggattagca	gagcgaggta	4680
tgtaggcggt	gctacagagt	tcttgaagtg	gtggcctaac	tacggctaca	ctagaagaac	4740
agtatttggg	atctgcgctc	tgctgaagcc	agttaccttc	ggaaaagag	ttggtagctc	4800
ttgatccggc	aaacaaacca	ccgctggtag	cggtggTTTT	tttgtttgca	agcagcagat	4860
tacgcgcaga	aaaaaaggat	ctcaagaaga	tcctttgatc	tttctacgg	ggtctgacgc	4920
tcagtggaac	gaaaactcac	gtaagggat	tttggctatg	agattatcaa	aaaggatctt	4980
cacctagatc	cttttaaatt	aaaaatgaag	ttttaaata	atctaaagta	tatatgagta	5040
aacttggctc	gacagttacc	aatgcttaat	cagtgaggca	cctatctcag	cgatctgtct	5100
atttcgttca	tccatagttg	cctgactcct	gaaaccacg	ttgtggtaga	attggtaaag	5160
agagtcgtgt	aaaatatcga	gttcgcacat	cttgttgtct	gattattgat	ttttggcgaa	5220
accatttgat	catatgacaa	gatgtgtatc	taccttaact	taatgatttt	gataaaaatc	5280
attaggtacc	tgtacattta	tattggtcga	tgtccaacat	taccgcatg	ttgacattga	5340
ttattgacta	gttattaata	gtaatcaatt	acggggctcat	tagttcatag	cccatatatg	5400
gagttccgcg	ttacataact	tacggtaaatt	ggcccgcctg	gctgaccgcc	caacgacccc	5460
cgcccattga	cgtaataat	gacgtatgtt	cccatagtaa	cgccaatagg	gactttccat	5520
tgacgtcaat	gggtggagta	tttacggtaa	actgcccact	tggcagtaca	tcaagtgtat	5580
catatgccaa	gtacgcccc	tattgacgtc	aatgacggta	aatggcccgc	ctggcattat	5640
gcccagtaca	tgaccttatg	ggactttcct	acttggcagt	acatctacgt	attagtcac	5700

gctattacca	tggtgatgcg	gttttggcag	tacatcaatg	ggcgtggata	gcggtttgac	5760
tcacggggat	ttccaagtct	ccacccatt	gacgtcaatg	ggagtttggt	ttggcaccaa	5820
aatcaacggg	actttccaaa	atgtcgtaac	aactccgcc	cattgacgca	aatgggcggt	5880
aggcgtgtac	ggtgggaggt	ctatataagc	agagctcgtt	tagtgaaccg	gggtctctct	5940
ggttagacca	gatctgagcc	tgggagctct	ctggctaact	agggaacca	ctgcttaagc	6000
ctcaataaag	cttgccttga	gtgcttcaag	tagtgtgtgc	ccgtctgttg	tgtgactctg	6060
gtaactagag	atccctcaga	cccttttagt	cagtgtggaa	aatctctagc	agtggcgccc	6120
gaacagggac	ttgaaagcga	aagggaaacc	agaggagctc	tctcgacgca	ggactcggct	6180
tgctgaagcg	cgcacggcaa	gaggcgaggg	gcggcgaactg	gtgagtacgc	caaaaatfff	6240
gactagcgga	ggctagaagg	agagagatgg	gtgcgagagc	gtcagtatta	agcgggggag	6300
aattagatcg	cgatgggaaa	aaattcggtt	aaggccaggg	ggaaagaaaa	aatataaatt	6360
aaaacatata	gtatgggcaa	gcagggagct	agaacgattc	gcagttaatc	ctggcctggt	6420
agaacatca	gaaggtgta	gacaaatact	gggacagcta	caaccatccc	ttcagacagg	6480
atcagaagaa	cttagatcat	tatataatac	agtagcaacc	ctctattgtg	tgcatacaag	6540
gatagagata	aaagacacca	aggaagcttt	agacaagata	gaggaagagc	aaaacaaaag	6600
taagaccacc	gcacagcaag	cggccgctga	tcttcagacc	tggaggagga	gatatgaggg	6660
acaattggag	aagtgaatta	tataaatata	aagtagtaaa	aattgaacca	ttaggagtag	6720
caccaccaa	ggcaaagaga	agagtgggtc	agagagaaaa	aagagcagtg	ggaataggag	6780
ctttgttctt	tgggttcttg	ggagcagcag	gaagcactat	gggcgcagcg	tcaatgacgc	6840
tgacggtaca	ggccagacaa	ttattgtctg	gtatagtgca	gcagcagaac	aatttgctga	6900
gggctattga	ggcgcaacag	catctgttgc	aactcacagt	ctggggcatc	aagcagctcc	6960
aggcaagaat	cctggctgtg	gaaagatacc	taaaggatca	acagctcctg	gggatttggg	7020
gttgctctgg	aaaactcatt	tgcaccactg	ctgtgccttg	gaatgctagt	tggagtaata	7080
aatctctgga	acagatttgg	aatcacacga	cctggatgga	gtgggacaga	gaaattaaca	7140
attacacaag	cttaatacac	tccttaattg	aagaatcgca	aaaccagcaa	gaaaagaatg	7200
aacaagaatt	attggaatta	gataaatggg	caagtttgtg	gaattggttt	aacataacaa	7260
attggctgtg	gtatataaaa	ttattcataa	tgatagtagg	aggcttggtg	ggtttaagaa	7320
tagtttttgc	tgtactttct	atagtgaata	gagttaggca	gggatattca	ccattatcgt	7380
ttcagaccca	cctcccaacc	ccgaggggac	ccgacaggcc	cgaaggaata	gaagaagaag	7440
gtggagagag	agacagagac	agatccattc	gattagttaa	cggatctcga	cggtatcggt	7500
taacttttaa	aagaaaaggg	gggattgggg	ggtacagtgc	aggggaaaga	atagtagaca	7560
taatagcaac	agacatacaa	actaaagaat	tacaaaaaca	aattacaaaa	attcaaaatt	7620
ttatcgatca	cgagactagc	ctcgag				7646
<210>	28					
<211>	234					
<212>	DNA					
<213>	人类免疫缺陷病毒					
<400>	28					
tggaagggct	aattcactcc	caacgaagac	aagatctgct	ttttgcttgt	actgggtctc	60

tctggttaga ccagatctga gcctgggagc tctctggcta actagggaac ccaactgctta	120
agcctcaata aagcttgccct tgagtgttc aagtagtgtg tgcccgtctg ttgtgtgact	180
ctggtaacta gagatccctc agaccctttt agtcagtgtg gaaaatctct agca	234
<210> 29	
<211> 132	
<212> DNA	
<213> 猿猴病毒40	
<400> 29	
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca	60
aataaagcat ttttttcaact gcattctagt tgtggtttgt ccaaactcat caatgtatct	120
tatcatgtct gg	132
<210> 30	
<211> 160	
<212> DNA	
<213> 猿猴病毒40	
<400> 30	
tcccgccct aactccgcc atcccgcc taactccgcc cagttccgcc cattctccgc	60
cccatggctg actaattttt tttatttatg cagaggccga ggccgcctcg gcctctgagc	120
tattccagaa gtagtgagga ggcttttttg gaggcctagg	160
<210> 31	
<211> 1015	
<212> DNA	
<213> 人工序列	
<220>	
<223> 实验室制备	
<400> 31	
tcttccgctt cctcgtcac tgactcgtg cgctcggctg ttcggctgcg gcgagcggtta	60
tcagctcact caaaggcggc aatacggta tccacagaat caggggataa cgcaggaaag	120
aacatgtgag caaaaggcca gaaaaggcc aggaaccgta aaaaggccgc gttgctggcg	180
tttttccata ggctccgcc cctgacgag catcaaaaa atcgacgctc aagtcagagg	240
tggcgaaacc cgacaggact ataaagatac caggcgtttc ccctggaag ctccctcgtg	300
cgctctcctg ttccgacct gccgcttac ggatacctgt ccgcctttct cccttcggga	360
agcgtggcgc tttctcatag ctacgctgt aggtatctca gttcgggtgta ggtcgttcgc	420
tccaagctgg gctgtgtgca cgaaccccc gttcagcccc accgctgcgc cttatccggt	480
aactatcgtc ttgagtccaa cccggtaaga cacgacttat cgccactggc agcagccact	540
ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagtctt gaagtgttg	600
cctaactacg gctacactag aagaacagta tttggtatct gcgctctgct gaagccagtt	660
accttcggaa aaagagttg tagctcttga tccggcaaac aaaccaccgc tggtagcgg	720
ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct	780

ttgatctttt ctacggggtc tgacgctcag tggaacgaaa actcacgtta agggattttg	840
gtcatgagat tatcaaaaag gatcttcacc tagatccttt taaattaaat atgaagtttt	900
aaatcaatct aaagtatata tgagtaaact tggctgaca gttaccaatg cttaatcagt	960
gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg actcc	1015
<210> 32	
<211> 139	
<212> DNA	
<213> 大肠杆菌	
<400> 32	
gtagaattgg taaagagagt cgtgtaaaat atcgagttcg cacatcttgt tgtctgatta	60
ttgatttttg gcgaaacat ttgatcatat gacaagatgt gtatctacct taacttaatg	120
atthttgataa aaatcatta	139
<210> 33	
<211> 577	
<212> DNA	
<213> 人类β疱疹病毒5	
<400> 33	
acattgatta ttgactagtt attaatagta atcaattacg gggtcattag ttcatagccc	60
atatatggag ttccgcgta cataacttac ggtaaattggc ccgcctggct gaccgccc	120
cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caataggac	180
tttccattga cgtcaatggg tggagtattt acggtaaact gccacttg cagtacatca	240
agtgtatcat atgccaagta cgtcccctat tgacgtcaat gacggtaaat ggccccctg	300
gcattatgcc cagtacatga ccttatggga ctttctact tggcagtaca tctacgtatt	360
agtcatcgct attaccatgg tgatgcggtt ttggcagtac atcaatgggc gtggatagcg	420
gtttgactca cggggatttc caagtctcca cccattgac gtcaatggga gtttgttttg	480
gcacccaaat caacgggact ttccaaaatg tcgtaacaac tccgccccat tgacgcaaat	540
ggcggttagg cgtgtacggt gggaggtcta tataagc	577
<210> 34	
<211> 188	
<212> DNA	
<213> 人类免疫缺陷病毒	
<400> 34	
gtctctctgg ttagaccaga tctgagcctg ggagctctct ggctaactag ggaaccact	60
gcttaagcct caataaagct tgcttgagt gcttcaagta gtgtgtgccc gtctgtttgtg	120
tgactctggg aactagagat ccctcagacc cttttagtca gtgtggaaaa tctctagcag	180
tggcgccc	188
<210> 35	
<211> 45	
<212> DNA	

<213> 人类免疫缺陷病毒 1	
<400> 35	
tgagtacgcc aaaaattttg actagcggag gctagaagga gagag	45
<210> 36	
<211> 362	
<212> DNA	
<213> 人类免疫缺陷病毒	
<400> 36	
atgggtgcga gagcgtcagt attaagcggg ggagaattag atcgcgatgg gaaaaaattc	60
ggttaaggcc agggggaaaag aaaaaatata aattaaaca tatagtatgg gcaagcaggg	120
agctagaacg attcgcagtt aatcctggcc tgttagaac atcagaaggc tgtagacaaa	180
tactgggaca gctacaacca tcccttcaga caggatcaga agaacttaga tcattatata	240
atacagtagc aaccctctat tgtgtgcatc aaaggataga gataaaagac accaaggaag	300
ctttagacaa gatagaggaa gagcaaaaca aaagtaagac caccgcacag caagcggccg	360
ct	362
<210> 37	
<211> 858	
<212> DNA	
<213> 人工序列	
<220>	
<223> 实验室制备 - 质粒元件	
<400> 37	
gatcttcaga cctggaggag gagatatgag ggacaattgg agaagtgaat tatataaata	60
taaagtagta aaaattgaac cattaggagt agcaccacc aaggcaaaga gaagagtgg	120
gcagagagaa aaaagagcag tgggaatagg agctttgttc cttgggttct tgggagcagc	180
aggaagcact atgggcgcag cgtcaatgac gctgacggta caggccagac aattattgtc	240
tggtatagtg cagcagcaga acaatttgct gagggctatt gaggcgcaac agcatctgtt	300
gcaactcaca gtctggggca tcaagcagct ccaggcaaga atcctggctg tggaaagata	360
cctaaaggat caacagctcc tggggatttg gggttgetct ggaaaactca tttgcaccac	420
tgctgtgcct tggaatgcta gttggagtaa taaatctctg gaacagattt ggaatcacac	480
gacctggatg gagtgggaca gagaaattaa caattacaca agcttaatac actccttaat	540
tgaagaatcg caaaaccagc aagaaaagaa tgaacaagaa ttattggaat tagataaatg	600
ggcaagtttg tggaattggt ttaacataac aaattggctg tggatatataa aattattcat	660
aatgatagta ggaggcttgg taggtttaag aatagttttt gctgtacttt ctatagtгаа	720
tagagttagg cagggatatt caccattatc gtttcagacc cacctccca ccccgagggg	780
acccgacagg cccgaaggaa tagaagaaga aggtggagag agagacagag acagatccat	840
tcgattagtg aacggatc	858
<210> 38	
<211> 118	

<212>	DNA	
<213>	人类免疫缺陷病毒	
<400>	38	
	ttttaaaaga aaagggggga ttggggggta cagtgcaggg gaaagaatag tagacataat	60
	agcaacagac atacaaacta aagaattaca aaaacaaatt acaaaaattc aaaatttt	118
<210>	39	
<211>	3847	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	实验室制备 - 质粒骨架构建体	
<400>	39	
	aacttgttta ttgcagetta taatggttac aaataaagca atagcatcac aaatttcaca	60
	aataaagcat ttttttcaact gcattctagt tgtggtttgt ccaaactcat caatgtatct	120
	tatcatgtct ggctctaget atcccccccc taactccgcc catccccccc ctaactccgc	180
	ccagttccgc ccattctccg ccccatggct gactaatttt ttttatttat gcagaggccg	240
	aggccgcctc ggctctgag ctattccaga agtagtgagg aggctttttt ggaggcctag	300
	gtagccccgc taatgagcgg gctttttttt cttagcett cttccgcttc ctcgctcact	360
	gactcgtgc gctcggtcgt tcggctgcgg cgagcggat cagctcactc aaaggcggta	420
	atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag	480
	caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc	540
	cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccg gacaggacta	600
	taaagatacc aggcgtttcc ccctggaagc tcctcgtgc gctctcctgt tccgaccctg	660
	ccgcttaccg gatacctgtc cgcttttctc cttcgggaa gcgtggcgct ttctcatagc	720
	tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac	780
	gaaccccccg ttcagcccga ccgctgcgcc ttatccggtta actatcgtct tgagtccaac	840
	ccggtaaagac acgacttatt gccactggca gcagccactg gtaacaggat tagcagagcg	900
	aggtagtag gcggtgctac agagtcttg aagtgggtggc ctaactacgg ctacactaga	960
	agaacagtat ttggtatctg cgctctgctg aagccagtta cttcggaaa aagagttggt	1020
	agctcttgat ccggcaaaaca aaccaccgct ggtagcgggtg gtttttttgt ttgcaagcag	1080
	cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggtct	1140
	gacgctcagt ggaacgaaaa ctacgcttaa gggattttgg tcatgagatt atcaaaaagg	1200
	atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatata	1260
	gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc	1320
	tgtctatttc gttcatccat agttgectga ctctgcaaa ccacgttggt gtagaattgg	1380
	taaagagagt cgtgtaaaat atcgagttcg cacatcttgt tgtctgatta ttgatttttg	1440
	gcgaaaccat ttgatcatat gacaagatgt gtatctacct taacttaatg attttgataa	1500
	aatcatttag gtacctgtac atttatattg gctcatgtcc aacattaccg ccatgttgac	1560
	attgattatt gactagttat taatagtaat caattacggg gtcattagtt catagcccat	1620

atatggagtt	ccgcgttaca	taacttacgg	taaattggccc	gcctggctga	ccgcccacg	1680
acccccgccc	attgacgtca	ataatgacgt	atgttcccat	agtaacgcca	atagggactt	1740
tccattgacg	tcaatgggtg	gagtatttac	ggtaaactgc	ccacttggca	gtacatcaag	1800
tgtatcatat	gccaagtacg	ccccctattg	acgtcaatga	cggtaaattg	cccgcctggc	1860
attatgcccc	gtacatgacc	ttatgggact	ttcctacttg	gcagtacatc	tacgtattag	1920
tcatcgtat	taccatgggtg	atgcggtttt	ggcagtacat	caatgggcgt	ggatagcgg	1980
ttgactcacg	gggatttcca	agtctccacc	ccattgacgt	caatgggagt	ttgttttggc	2040
acaaaaatca	acgggacttt	ccaaaatgtc	gtaacaactc	cgccccattg	acgcaaattg	2100
gcggtaggcg	tgtacgggtg	gaggtctata	taagcagagc	tcgttttagt	aaccggggctc	2160
tctctggtta	gaccagatct	gagcctggga	gctctctggc	taactaggga	accactgct	2220
taagcctcaa	taaagcttgc	cttgagtgtc	tcaagtagtg	tgtgcccgtc	tgttgtgtga	2280
ctctggtaac	tagagatccc	tcagaccctt	ttagtcagtg	tggaaaatct	ctagcagtgg	2340
cgcccgaaca	gggacttgaa	agcgaagg	aaaccagagg	agctctctcg	acgcaggact	2400
cggcttgctg	aagcgcgcac	ggcaagaggc	gaggggcggc	gactggtgag	tacgcaaaa	2460
atthtgacta	gcggaggcta	gaaggagaga	gatgggtgcg	agagcgtcag	tattaagcgg	2520
gggagaatta	gatcgcgatg	ggaaaaaatt	cggttaaggc	cagggggaaa	gaaaaaatat	2580
aaattaatac	atatagtatg	ggcaagcagg	gagctagaac	gattcgcagt	taatcctggc	2640
ctgttagaaa	catcagaagg	ctgtagacaa	atactgggac	agctacaacc	atcccttcag	2700
acaggatcag	aagaacttag	atcattatat	aatacagtag	caaccctcta	ttgtgtgcat	2760
caaaggatag	agataaaaaga	caccaaggaa	gctttagaca	agatagagga	agagcaaaac	2820
aaaagtaaga	ccaccgcaca	gcaagcggcc	gctgatcttc	agacctggag	gaggagatat	2880
gagggacaat	tggagaagtg	aattatataa	atataaagta	gtaaaaattg	aaccattagg	2940
agtagcacc	accaaggcaa	agagaagagt	ggtgcagaga	gaaaaaagag	cagtgggaat	3000
aggagctttg	ttccttgggt	tcttgggagc	agcaggaagc	actatgggcg	cagcgtcaat	3060
gacgctgacg	gtacaggcca	gacaattatt	gtctgggtata	gtgcagcagc	agaacaatth	3120
gctgagggct	attgaggcgc	aacagcatct	gttgcaactc	acagtctggg	gcatcaagca	3180
gctccaggca	agaatcctgg	ctgtggaaag	atacctaaag	gatcaacagc	tcctggggat	3240
ttggggttgc	tctgaaaac	tcatttgcac	cactgctgtg	ccttggaatg	ctagttggag	3300
taataaatct	ctggaacaga	tttggaatca	cacgacctgg	atggagtggg	acagagaaat	3360
taacaattac	acaagcttaa	tacactcctt	aattgaagaa	tcgcaaaacc	agcaagaaaa	3420
gaatgaacaa	gaattattgg	aattagataa	atgggcaagt	ttgtggaatt	ggtttaacat	3480
aacaaattgg	ctgtggtata	taaaattatt	cataatgata	gtaggaggct	tggtaggttt	3540
aagaatagtt	tttgcgttac	tttctatagt	gaatagagtt	aggcagggat	attcaccatt	3600
atcgtttcag	accacctcc	caaccccgag	gggaccgcac	aggcccgaag	gaatagaaga	3660
agaaggtgga	gagagagaca	gagacagatc	cattcgatta	gtgaacggat	ctcgacggta	3720
tcggttaact	tttaaaagaa	aaggggggat	tgggggttac	agtgcagggg	aaagaatagt	3780
agacataata	gcaacagaca	tacaaactaa	agaattacaa	aaacaaatta	caaaaattca	3840
aaattht						3847

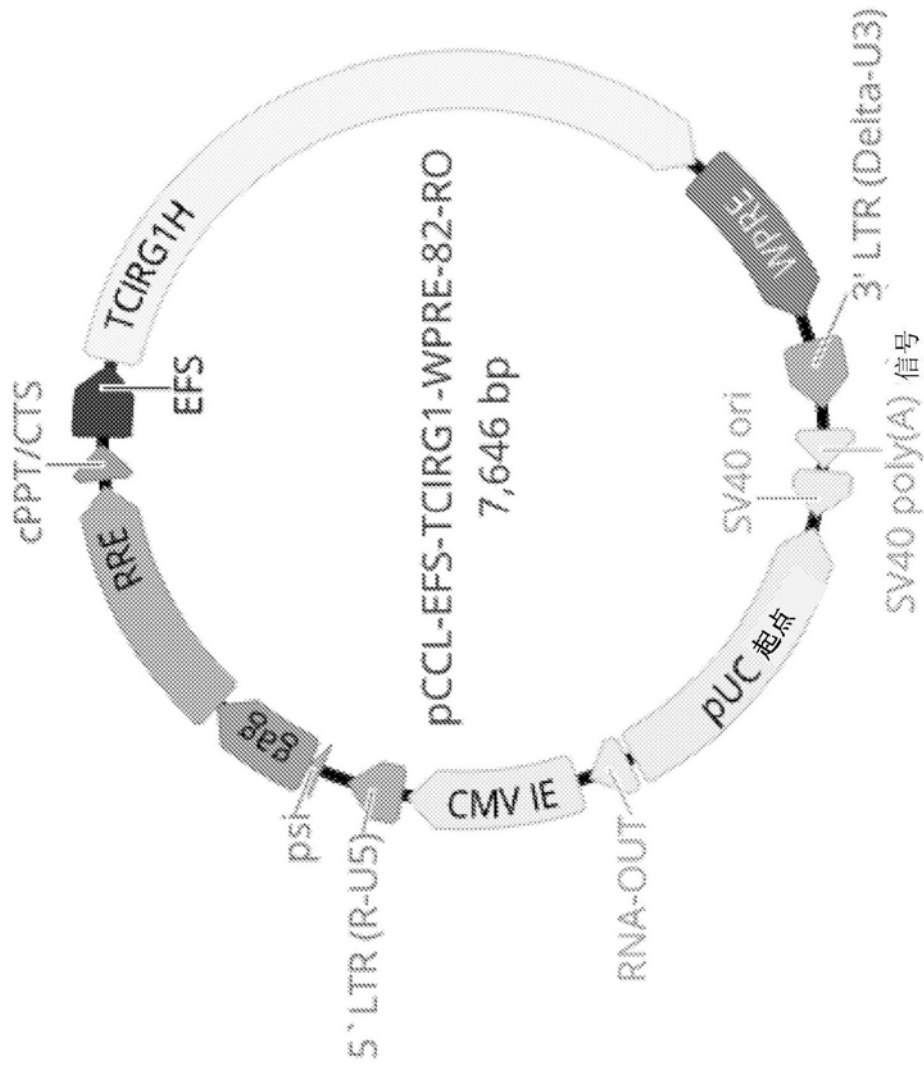


图1



图3A

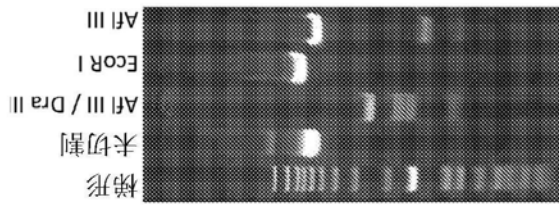


图3B

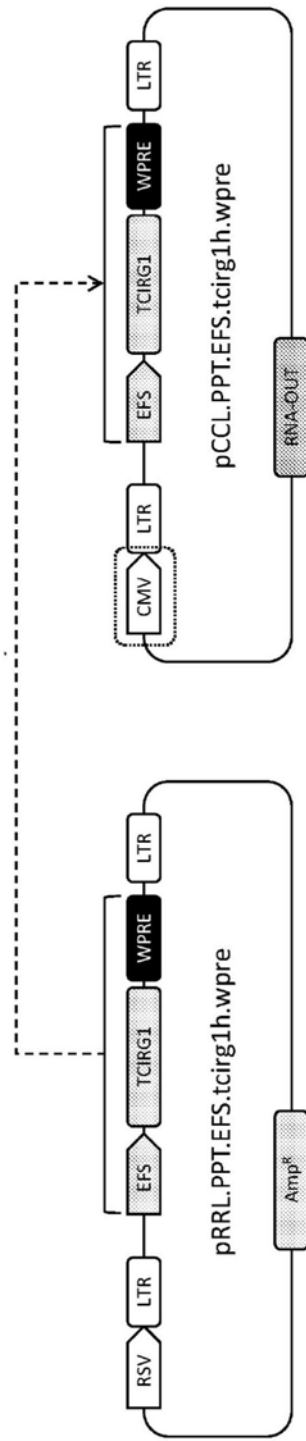


图3C

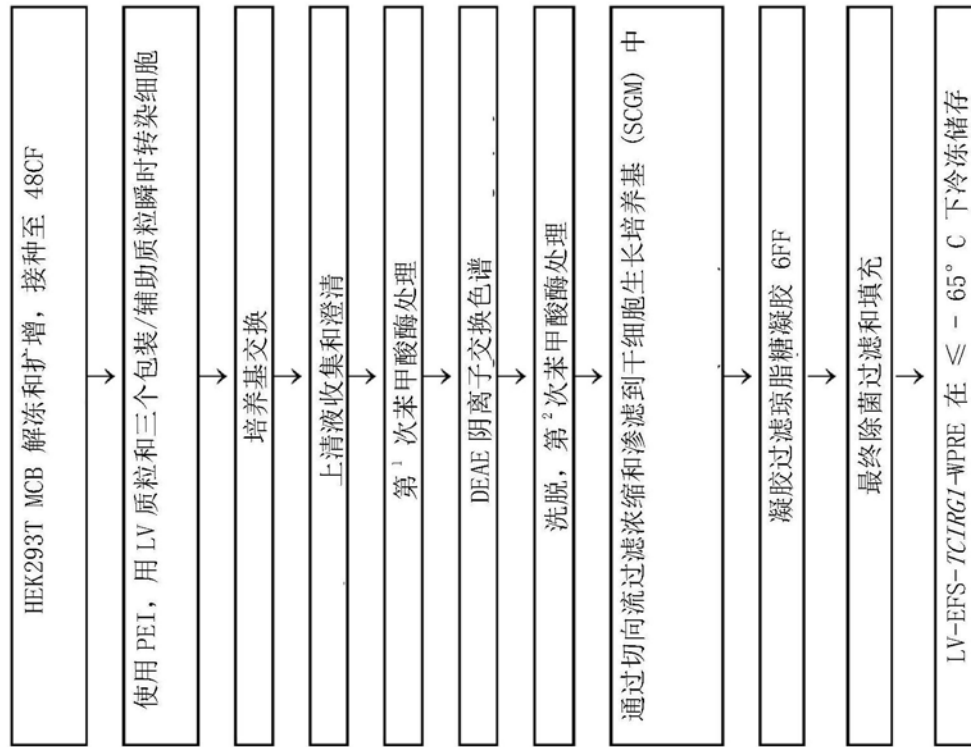


图4

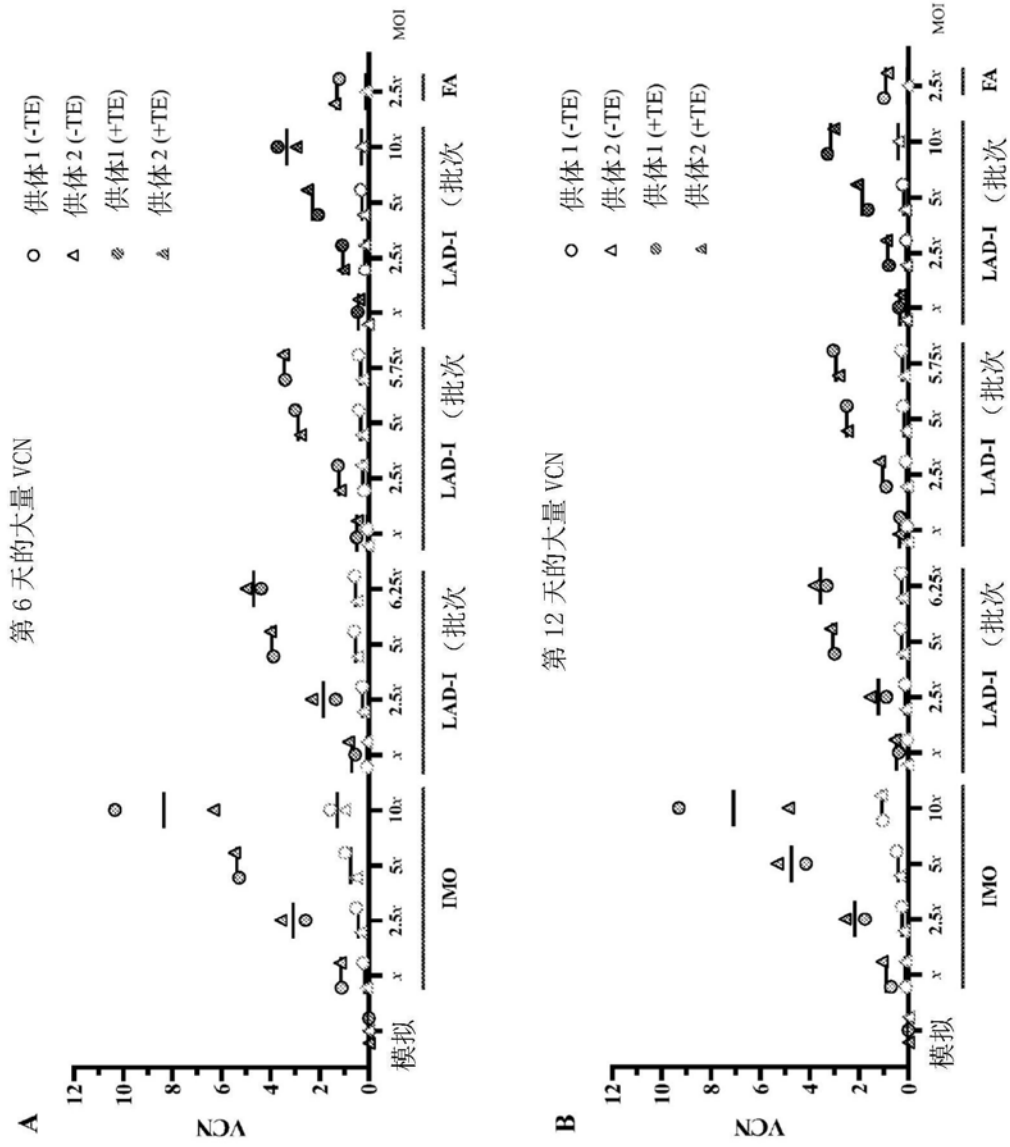


图5A-5B