



(10) 授权公告号 CN 110073002 B

(45) 授权公告日 2024. 06. 11

(21) 申请号 201780075817.7

(22) 申请日 2017.10.03

(65) 同一申请的已公布的文献号  
申请公布号 CN 110073002 A

(43) 申请公布日 2019.07.30

(30) 优先权数据  
62/409,228 2016.10.17 US

(85) PCT国际申请进入国家阶段日  
2019.06.06

(86) PCT国际申请的申请数据  
PCT/US2017/054928 2017.10.03

(87) PCT国际申请的公布数据  
W02018/075235 EN 2018.04.26

(73) 专利权人 杨森制药公司  
地址 美国新泽西州

(72) 发明人 库尔特·艾弗·卡姆鲁德

(74) 专利代理机构 永新专利商标代理有限公司  
72002  
专利代理师 黄歆

(51) Int.Cl.  
C12N 15/86 (2006.01)

(56) 对比文件  
W0 2014170493 A2, 2014.10.23

US 2014079734 A1, 2014.03.20

Kulasegaran-Shylini, R. 等. Structural and functional elements of the promoter encoded by the 5' untranslated region of the venezuelan equine encephalitis virus genome.《Journal of Virology》.2009,第83卷(第17期),

Kim D.Y. 等. Enhancement of protein expression by alphavirus replicons by designing self-replicating subgenomic RNAs.《PNAS》.2014,第111卷(第29期),

Kulasegaran-Shylini R. 等. The 5' UTR-specific mutation in VEEV TC-83 genome has a strong effect on RNA replication and subgenomic RNA synthesis, but not on translation of the encoded proteins.《Virology》.2009,

Maruggi G. 等. Engineered alphavirus replicon vaccines based on known attenuated viral mutants show limited effects on immunogenicity.《Virology》.2013,

审查员 蔡玉品

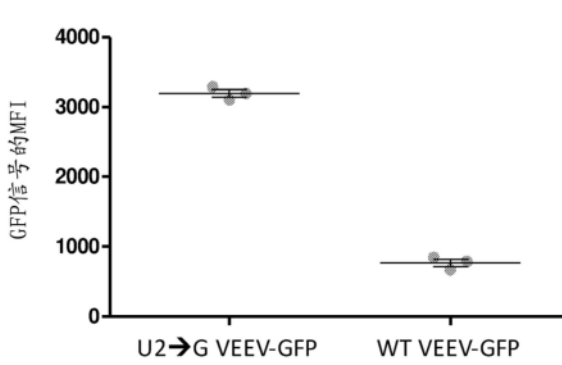
权利要求书2页 说明书23页  
序列表15页 附图9页

(54) 发明名称  
重组病毒复制子体系及其用途

(57) 摘要

本公开整体涉及根据基于病毒的表达体系,其适用于制备感兴趣的分子。本公开涉及核酸构建体,诸如表达载体,其包含经修饰的复制子RNA,所述经修饰的复制子RNA包含经修饰的5'-非翻译区(5'-UTR),并且任选地已经缺失编码结构化蛋白质的初始病毒序列中的至少一些。本发明还公开了用于制备感兴趣的多肽的方法。

平均荧光强度



1. 一种包含经修饰的甲病毒复制子RNA的核酸分子,其用于制备感兴趣的多肽,其中所述经修饰的甲病毒复制子RNA包含甲病毒的经修饰的5'-UTR和编码生物活性甲病毒非结构化蛋白质的序列,并且缺乏编码病毒结构化蛋白质的核酸序列的至少一部分,其中所述经修饰的5'-UTR的修饰是位置2处的U->G核苷酸置换,其中所述甲病毒为委内瑞拉马脑炎病毒(VEEV)。

2. 根据权利要求1所述的核酸分子,其中所述经修饰的甲病毒复制子RNA还包含亚基因组启动子RNA、顺式复制所需的3'病毒序列以及聚腺苷酸核束。

3. 根据权利要求1所述的核酸分子,其中所述经修饰的甲病毒复制子RNA缺乏相当大部分的编码病毒结构化蛋白质的核酸序列。

4. 根据权利要求1所述的核酸分子,其中所述经修饰的甲病毒复制子RNA不包含编码病毒结构化蛋白质的核酸序列。

5. 根据权利要求1所述的核酸分子,其还包含一个或多个表达盒,其中所述表达盒中的每一个包含可操作地连接至异源核酸序列的启动子。

6. 根据权利要求1至5中任一项所述的核酸分子,其中所述经修饰的甲病毒复制子RNA可操作地连接至异源调控元件。

7. 一种包含经修饰的VEEV基因组或VEEV复制子RNA的核酸分子,其用于制备感兴趣的多肽,其中所述核酸分子的序列为在5'-非翻译区(5'-UTR)的位置2处含有U->G置换的SEQ ID NO:1的相应RNA核酸序列。

8. 一种包含经修饰的VEEV基因组或VEEV复制子RNA的核酸分子,其用于制备感兴趣的多肽,其中所述经修饰的VEEV基因组或VEEV复制子RNA包含5'-UTR,其中所述5'-UTR的序列为在5'-UTR的位置2处含有U->G置换的SEQ ID NO:15的相应RNA核酸序列,其中所述经修饰的VEEV基因组或VEEV复制子RNA包含编码生物活性VEEV非结构化蛋白质的序列,并且其中所述经修饰的VEEV基因组或VEEV复制子RNA缺乏编码病毒结构化蛋白质的序列的至少一部分。

9. 根据权利要求8所述的核酸分子,其中所述经修饰的VEEV基因组或VEEV复制子RNA缺乏相当大部分的编码病毒结构化蛋白质的核酸序列。

10. 根据权利要求7至9中任一项所述的核酸分子,其中所述经修饰的VEEV基因组或VEEV复制子RNA不包含编码病毒结构化蛋白质的核酸序列。

11. 一种重组细胞,其包含根据权利要求1至10中任一项所述的核酸分子。

12. 根据权利要求11所述的重组细胞,其中所述重组细胞为原核细胞或真核细胞。

13. 根据权利要求12所述的重组细胞,其中所述重组细胞为动物细胞。

14. 一种用于制备感兴趣的多肽的方法,所述方法包括培养包含核酸分子的宿主细胞,所述核酸分子包含经修饰的VEEV复制子RNA,其中所述经修饰的VEEV复制子RNA包含VEEV的经修饰的5'-UTR以及编码生物活性VEEV非结构化蛋白质的序列,并且缺乏编码病毒结构化蛋白质的核酸序列的至少一部分,其中所述经修饰的5'-UTR的修饰是位置2处的U->G核苷酸置换。

15. 包含经修饰的VEEV复制子RNA的核酸分子在制备用于在受试者中产生感兴趣的多肽的组合物中的用途,其中所述经修饰的VEEV复制子RNA包含VEEV的经修饰的5'-UTR以及编码生物活性VEEV非结构化蛋白质的序列,并且缺乏编码病毒结构化蛋白质的核酸序列的

至少一部分,其中所述经修饰的5'-UTR的修饰是位置2处的U->G核苷酸置换。

16.根据权利要求15所述的用途,其中所述受试者为人、马、猪、灵长类动物、小鼠、牛、绵羊、兔子、猫、狗、鸟、鱼、山羊、驴或仓鼠。

17.根据权利要求15所述的用途,其中所述受试者为野牛。

18.根据权利要求15至17中任一项所述的用途,其中所述经修饰的VEEV复制子RNA还包含亚基因组启动子RNA、顺式复制所需的3' 病毒序列以及聚腺苷酸核束。

19.根据权利要求15至17中任一项所述的用途,其中所述经修饰的VEEV复制子RNA缺乏相当大部分的编码病毒结构化蛋白质的核酸序列。

20.根据权利要求15至17中任一项所述的用途,其中所述经修饰的VEEV复制子RNA不包含编码病毒结构化蛋白质的核酸序列。

21.根据权利要求15至17中任一项所述的用途,其中所述核酸分子还包含一个或多个表达盒,其中所述表达盒中的每一个均包含可操作地连接至异源核酸序列的启动子。

22.根据权利要求15至17中任一项所述的用途,其中所述经修饰的VEEV复制子RNA可操作地连接至异源调控元件。

23.根据权利要求15至17中任一项所述的用途,其中所述包含经修饰的VEEV复制子RNA的核酸分子包含5'-UTR,其中所述5'-UTR的序列为在5'-UTR的位置2处含有U->G置换的SEQ ID NO:15的相应RNA核酸序列。

24.一种组合物,其包含根据权利要求1至10中任一项所述的核酸分子和药学上可接受的载体。

25.一种组合物,其包含根据权利要求11至13中任一项所述的重组细胞和药学上可接受的载体。

## 重组病毒复制子体系及其用途

[0001] 相关申请

[0002] 本专利申请根据119 USC§35要求2016年10月17日提交的美国临时专利申请62/409228的优先权,该文献全文以引用方式明确地并入本文。

[0003] 并入序列列表

[0004] 随附的序列列表中的材料据此以引用方式并入本专利申请中。名称为SGI011WO\_SeqListing.txt的随附序列列表文本文件在2017年10月3日创建,并且为34KB。

### 技术领域

[0005] 本公开涉及分子生物学领域,其包括包含经修饰的病毒复制子的核酸分子以及此类核酸分子用于在细胞培养物中或活体中的合适宿主细胞中产生期望产物的用途。

### 背景技术

[0006] 近年来,若干不同的动物病毒组已通过同源重组或对其基因组的直接工程化进行了基因操纵。DNA和RNA病毒的反向遗传学系统的可用性为重组病毒的使用创造了新的前景,例如,作为疫苗、表达载体、抗肿瘤剂、基因治疗载体和药物递送载体。

[0007] 例如,许多基于病毒的表达载体已被部署用于在培养的重组细胞中表达异源蛋白质。具体地讲,经修饰的病毒载体在宿主细胞中的基因表达的应用继续扩大。这方面的最新进展包括进一步开发用于生产多亚基蛋白质复合物的技术和体系,以及蛋白质修饰酶的共表达以改善异源蛋白质生产。关于病毒表达载体技术的其他最新进展包括用于控制基因表达、制备病毒载体、体内基因治疗应用和产生疫苗递送载体的许多先进的基因组工程化应用。

[0008] 然而,仍然需要用于在重组表达体系中表达感兴趣的基因的更有效的方法和体系。

### 发明内容

[0009] 本段提供了对本专利申请的一般概述,并且不包括其全部范围或其所有特征。

[0010] 在一个方面,本文公开了一种包括经修饰的复制子RNA的核酸分子,其中该经修饰的复制子RNA包括经修饰的5'-UTR并且缺乏编码病毒结构化蛋白质的核酸序列的至少一部分。在本公开的该方面和其他方面的各种实施方案中,本文所公开的核酸分子可包括下列特征中的一个或多个。在一些实施方案中,经修饰的复制子RNA为经修饰的甲病毒复制子RNA。在一些实施方案中,经修饰的甲病毒复制子RNA包括经修饰的甲病毒基因组。在一些实施方案中,经修饰的5'-UTR包括在位置1、2、4、或它们的组合处的一个或多个核苷酸置换。在一些实施方案中,该核苷酸置换中的至少一个为在经修饰的5'-UTR的位置2处的核苷酸置换。在一些实施方案中,在经修饰的5'-UTR的位置2处的核苷酸置换是U->G置换。

[0011] 在一些实施方案中,核酸分子包括缺乏相当大部分的编码病毒结构化蛋白质的核酸序列的经修饰的复制子RNA。在一些实施方案中,本文所公开的经修饰的甲病毒基因组或

复制子RNA不包括编码病毒结构化蛋白质的核酸序列。

[0012] 在本公开的这个方面和其他方面的各种实施方案中,核酸分子还包括一个或多个表达盒,其中表达盒中的每一个包括可操作地连接到异源核酸序列的启动子。在一些实施方案中,核酸分子包括至少两个、三个、四个、五个或六个表达盒。在一些实施方案中,表达盒中的至少一个的启动子为或包含26S亚基因组启动子。

[0013] 在一些实施方案中,本文所公开的表达盒中的至少一个的异源核酸序列包括感兴趣的基因(GOI)的编码序列。在一些实施方案中,GOI编码选自下列的多肽:治疗性多肽、预防性多肽、诊断性多肽、营养性多肽、工业酶和报告多肽。在一些实施方案中,GOI编码选自下列的多肽:抗体、抗原、免疫调节剂、和细胞因子。在一些具体的实施方案中,GOI的编码序列被优化用于以高于参考编码序列的表达水平的水平进行表达。

[0014] 在一些实施方案中,核酸分子包括经修饰的复制子RNA,所述经修饰的复制子RNA包括病毒的经修饰的基因组或复制子RNA,所述病毒属于披盖病毒(Togaviridae)家族的甲病毒属。在一些实施方案中,经修饰的基因组或复制子RNA具有属于VEEV/EEEV组、或SF组或SIN组的甲病毒。在一些实施方案中,甲病毒选自:东部马脑炎病毒(EEEV)、委内瑞拉马脑炎病毒(VEEV)、沼泽地病毒(EVEV)、穆坎布病毒(MUCV)、那皮舒纳病毒(PIXV)、米德堡病毒(MIDV)、奇昆古尼亚病毒病(CHIKV)、阿尼昂-尼昂病毒(ONNV)、罗斯河病毒(RRV)、巴马森林病毒(BF)、盖塔病毒(GET)、鹭山病毒(SAGV)、贝巴鲁病毒(BEBV)、马雅罗病毒(MAYV)、乌纳病毒(UNAV)、辛德毕斯病毒(SINV)、奥拉病毒(AURAV)、沃达罗河病毒(WHAV)、巴班肯病毒(BABV)、孜拉加奇病毒(KYZV)、西部马脑炎病毒(WEEV)、高地J病毒(HJV)、摩根堡病毒(FMV)、恩杜穆病毒(NDUV)、和博吉河病毒。在一些实施方案中,甲病毒为委内瑞拉马脑炎病毒(VEEV)。

[0015] 一些实施方案提供核酸分子,该核酸分子包括可操作地连接至异源调控元件的经修饰的甲病毒基因组或复制子RNA。在一些实施方案中,该异源调控元件包括启动子序列。在一些实施方案中,该启动子序列包括T7启动子序列。在一些实施方案中,该异源调控元件包括转录终止序列。在一些实施方案中,该转录终止序列为或包含T7终止序列。

[0016] 在一些实施方案中,如本文所公开的核酸分子包括经修饰的甲病毒基因组或复制子RNA,其中所述核酸分子表现出与SEQ ID NO:1的核酸序列的至少80%序列同一性,其中所述经修饰的甲病毒基因组或复制子RNA在5'-非翻译区(5'-UTR)的位置2处包含U->G置换,并且缺乏编码病毒结构化蛋白质的序列的至少一部分。在一些实施方案中,该核酸分子表现出与SEQ ID NO:1的核酸序列的至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性。

[0017] 在一些实施方案中,本文所公开的核酸分子包括经修饰的甲病毒基因组或复制子RNA,其中所述经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与SEQ ID NO:2-18中的至少一者的核酸序列的至少80%序列同一性,和在所述5'-UTR的位置2处的U->G置换,并且其中所述经修饰的甲病毒基因组或复制子RNA缺乏编码病毒结构化蛋白质的序列的至少一部分。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与SEQ ID NO:2-18中的至少一者的核酸序列的至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性。在某些实施方案中,经修饰的甲病毒基因组或复制子RNA缺乏相当大一部分的编码病毒结构化蛋白质的核酸序

列。在某些实施方案中,经修饰的甲病毒基因组或复制子RNA不包含编码病毒结构化蛋白质的核酸序列。

[0018] 在一个方面,本文所公开的一些实施方案涉及包含本文所述的核酸分子的重组细胞。在一些实施方案中,该重组细胞为原核细胞或真核细胞。在一些实施方案中,该重组细胞为动物细胞。在一些实施方案中,该重组细胞为脊椎动物细胞或无脊椎动物细胞。在一些实施方案中,该重组细胞选自:马肺动脉内皮细胞、马真皮细胞、幼仓鼠肾(BHK)细胞、兔肾细胞、小鼠肌细胞、小鼠结缔组织细胞、人子宫颈细胞、人表皮喉细胞、中国仓鼠卵巢细胞(CHO)、人HEK-293细胞、小鼠3T3细胞、Vero细胞、Madin-Darby犬肾上皮细胞(MDCK)、原代鸡成纤维细胞、HuT78细胞、A549肺细胞、HeLa细胞、**PER.C6<sup>®</sup>**细胞、WI-38细胞、MRC-5细胞、FRhL-2和CEM T细胞。本文所公开的一些实施方案提供了包含如本文所公开的至少一种重组细胞的细胞培养物。

[0019] 在一个方面,一些实施方案提供了用于产生感兴趣的多肽的方法,所述方法涉及培养包含核酸分子的宿主细胞,所述核酸分子包含经修饰的5'-UTR并且缺乏编码病毒结构化蛋白质的核酸序列的至少一部分。在一些实施方案中,重组细胞为脊椎动物细胞或无脊椎动物细胞。

[0020] 在另一方面,一些实施方案提供了用于在受试者中产生感兴趣的多肽的方法,该方法涉及向受试者施用核酸分子,所述核酸分子包含经修饰的5'-UTR并且缺乏编码病毒结构化蛋白质的核酸序列的至少一部分。在一些实施方案中,所述受试者为人类、马、猪、灵长类动物、小鼠、牛、猪、绵羊、兔子、猫、狗、鸟、鱼、山羊、驴、仓鼠或野牛。

[0021] 根据本公开的方法的实施方案的具体实施可包括以下特征中的一个或多个。在一些实施方案中,经修饰的复制子RNA为经修饰的甲病毒复制子RNA。在一些实施方案中,经修饰的甲病毒复制子RNA包括经修饰的甲病毒基因组。在一些实施方案中,经修饰的5'-UTR包括在位置1、2、4、或它们的组合处的一个或多个核苷酸置换。在一些实施方案中,该核苷酸置换中的至少一个为在经修饰的5'-UTR的位置2处的核苷酸置换。在一些实施方案中,在经修饰的5'-UTR的位置2处的核苷酸置换是U->G置换。在某些实施方案中,经修饰的复制子RNA缺乏相当大一部分的编码病毒结构化蛋白质的核酸序列。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA不包含编码病毒结构化蛋白质的核酸序列。

[0022] 在一些实施方案中,如本文所述的核酸分子还包括一个或多个表达盒,其中该表达盒中的每一个包括可操作地连接至异源核酸序列的启动子。在一些实施方案中,核酸分子包括至少两个、至少三个、至少四个、至少五个、或至少六个表达盒。在一些实施方案中,该表达盒中的至少一个的启动子包括26S亚基因组启动子。在一些具体实施方案中,该表达盒中的至少一个的启动子包括甲病毒26S亚基因组启动子。优选地,该启动子包含委内瑞拉马脑炎(VEEV)26S亚基因组启动子。在某些实施方案中,该表达盒中的至少一个的异源核酸序列包括感兴趣的基因(GOI)的编码序列。在一些实施方案中,GOI的编码序列被优化用于以高于参考编码序列的表达水平的水平进行表达。在一些实施方案中,可操作地连接至异源核酸序列的启动子包含异源启动子序列。合适的异源启动子包括但不限于来源于脑心肌炎病毒(EMCV)、牛病毒性腹泻病毒(BVDV)、脊髓灰质炎病毒、口蹄疫病毒(FMD)、肠病毒71型或丙型肝炎病毒的内部核糖体进入位点(IRES)的调控元件。

[0023] 在一些实施方案中,经修饰的复制子RNA包括病毒的经修饰的基因组或复制子

RNA,所述病毒属于披盖病毒(Togaviridae)家族的甲病毒属。在一些实施方案中,经修饰的基因组或复制子RNA具有属于VEEV/EEEV组、或SF组或SIN组的甲病毒。在一些实施方案中,甲病毒选自:东部马脑炎病毒(EEEV)、委内瑞拉马脑炎病毒(VEEV)、沼泽地病毒(EVEV)、穆坎布病毒(MUCV)、那皮舒纳病毒(PIXV)、米德堡病毒(MIDV)、奇昆古尼亚病毒病(CHIKV)、阿尼昂-尼昂病毒(ONNV)、罗斯河病毒(RRV)、巴马森林病毒(BF)、盖塔病毒(GET)、鹭山病毒(SAGV)、贝巴鲁病毒(BEBV)、马雅罗病毒(MAYV)、乌纳病毒(UNAV)、辛德毕斯病毒(SINV)、奥拉病毒(AURAV)、沃达罗河病毒(WHAV)、巴班肯病毒(BABV)、孜拉加奇病毒(KYZV)、西部马脑炎病毒(WEEV)、高地J病毒(HJV)、摩根堡病毒(FMV)、恩杜穆病毒(NDUV)、和博吉河病毒。在一些实施方案中,甲病毒为委内瑞拉马脑炎病毒(VEEV)。

[0024] 在一些实施方案中,经修饰的甲病毒基因组或复制子RNA可操作地连接至异源调控元件。在一些实施方案中,该异源调控元件包括启动子序列。在一些实施方案中,该启动子序列包括T7启动子序列。在一些实施方案中,该异源调控元件包含转录终止序列。在一些实施方案中,该转录终止序列为或包含T7终止序列。

[0025] 在一些实施方案中,核酸分子包括经修饰的甲病毒基因组或复制子RNA,其中所述经修饰的甲病毒基因组或复制子RNA包括5'-UTR,所述5'-UTR表现出与SEQ ID NO:1的核酸序列的至少80%序列同一性,和在所述5'-UTR的位置2处的U->G置换,并且其中所述经修饰的甲病毒基因组或复制子RNA缺乏编码病毒结构化蛋白质的序列的至少一部分。在一些实施方案中,该核酸分子表现出与SEQ ID NO:1的核酸序列的至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性。在一些实施方案中,核酸分子包括经修饰的甲病毒基因组或复制子RNA,其中所述经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与SEQ ID NO:2-18中的至少一者的核酸序列的至少80%序列同一性,和在所述5'-UTR的位置2处的U->G置换,并且其中所述经修饰的甲病毒基因组或复制子RNA缺乏编码病毒结构化蛋白质的序列的至少一部分。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与SEQ ID NO:2-18中的至少一者的核酸序列的至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性。

[0026] 在另一个方面,本文所公开的一些实施方案提供了通过根据本文所公开的一个或多个实施方案的方法制备的重组多肽。

[0027] 在一个方面,本文所公开的一些实施方案涉及一种组合物,其包含如本文所述的重组多肽。在一些实施方案中,该组合物为包含药学上可接受的载体的药物组合物。在一些实施方案中,该组合物为预防性组合物、营养性组合物、药物组合物、或它们的组合。

[0028] 在另一个方面,本文所公开的一些实施方案涉及一种组合物,其包含如本文所公开的核酸分子。在一些实施方案中,该组合物为包含药学上可接受的载体的药物组合物。在一些实施方案中,该组合物为预防性组合物、营养性组合物、药物组合物、或它们的组合。

[0029] 在又一个方面,本文所公开的一些实施方案涉及一种组合物,其包含如本文所述的重组细胞。在一些实施方案中,该组合物为包含药学上可接受的载体的药物组合物。在一些实施方案中,该组合物为预防性组合物、营养性组合物、药物组合物、或它们的组合。

[0030] 上述发明内容仅是示例性的,并且不旨在以任何方式进行限制性。除了本文所述的示例性实施例和特征之外,本申请的其他方面、实施方案、对象和特征将由附图和具体实

施方式以及权利要求变得完全显而易见。

## 附图说明

[0031] 图1A示出来自相应的甲病毒的5' 未翻译区(5' -UTR)的序列对比:奥拉病毒(AURAV;SEQ ID NO:2)、奇昆古尼亚病毒病(CHIKV,SEQ ID NO:3)、阿尼昂-尼昂病毒(SEQ ID NO:4,ONNV)、贝巴鲁病毒(SEQ ID NO:5,BEBV)、塞姆利基森林病毒(SEQ ID NO:6,SFV)、马雅罗病毒(SEQ ID NO:7,MAYV)、盖塔病毒(SEQ ID NO:8,GETV)、鹭山病毒(SEQ ID NO:9,SAGV)、恩杜穆病毒(SEQ ID NO:10,NDUV)、米德堡病毒(SEQ ID NO:11,MIDV)、东部马脑炎病毒(SEQ ID NO:12,EEEV)、摩根堡病毒(SEQ ID NO:13,FMV)、布吉溪病毒(SEQ ID NO:14,Buggy)、委内瑞拉马脑炎病毒(SEQ ID NO:15,VEEV)、瓦塔罗阿病毒(SEQ ID NO:16,WHAV)、辛德毕斯病毒(SEQ ID NO:17,SINV)和比巴鲁病毒(SEQ ID NO:18,BEBV)。图1A的序列比对使用具有默认设置的程序MUSCLE 3.6形成。在本文所示的序列比对中,由程序MUSCLE 3.6产生的用于最佳比对的比对序列中的破折号表示间隙,即该位置处缺乏核苷酸。如下文所详述的,已在该序列比较分析中鉴定了若干保守核苷酸残基。星号标识比对序列之间的相同核苷酸残基。图1B以加权一致性示出共有序列的图示,其中命名给定氨基酸的字母的大小与用于产生基序的不同序列中残基的保守性成正比(该字母的大小表示在比对序列中该位置处的残基的相对频率)。字符的大小反映了以位为单位测量的信息内容。

[0032] 图2示出了非限制性示例性碱基一价VEEV复制子设计,Rep- $\alpha$ 的结构示意图,其包括T7启动子序列,来自具有如本文所述的U2->G置换的VEEV的5' UTR序列,甲病毒基因组的非结构化多肽的编码序列nsp1、nsp2、nsp3和nsp4。碱基一价VEEV复制子Rep- $\alpha$ 还包含26S亚基因组启动子序列、3'UTR序列、T7终止序列、聚腺苷酸化序列PolyA、以及被工程化以促进附加组分插入复制子中的多个独特的限制性位点。

[0033] 图3和图4图示了两种非限制性示例性一价VEEV复制子设计的结构,其中可操作地掺入载体中的感兴趣的基因(GOI)分别为A越南型1203HA基因(图3)和增强的绿色荧光蛋白质(eGFP)报告基因(图4)。

[0034] 图5图示总结了示例性实验的结果,其示出了经修饰的VEEV-HA复制子中的5' UTR的位置2处的U2->G修饰不影响经修饰的复制子的生物活性。对用表达流感血凝素-HA基因的经修饰的U2->G VEEV-HA复制子电穿孔的细胞进行流式细胞术分析(FACS)(对于复制子的结构组织也参见图3)。将野生型VEEV-HA复制子(即在位置2处包含U残基)用作对照。

[0035] 图6图示总结了进行示例性流式细胞术分析的结果以展示出来自携带5' UTR的U2->G修饰的甲病毒复制子的表达不限于任何感兴趣的特定基因(GOI)。在该实验中,将经修饰的甲病毒复制子工程化以表达绿色荧光蛋白(GFP)报告基因(对于复制子的结构组织,也参见图4)。经修饰的VEEV-GFP复制子的5' UTR中的U2->G修饰示出相对于从野生型对照复制子检测到的表达,GFP基因的表达增强了3倍。

[0036] 图7图示总结了评估来自经修饰的VEEV复制子的红色萤火虫报告基因的表达的另一示例性流式细胞术分析的结果。在该实验中,经修饰的VEEV-rFF复制子的5' UTR中的U2->G修饰示出相对于从野生型对照复制子检测到的表达,红色萤火虫基因的表达增强了2倍。

[0037] 图8示意性地示出了非限制性示例性甲病毒基因组结构和基因组表达(改编自Strauss等人,Microbiological Reviews,第491-562页,1994年9月)。示出了SIN病毒的基

基因组组织。给出了非结构化基因和结构化蛋白基因的名称。基因和蛋白质的命名的参考可见于Strauss等人,出处同上,1994。49S基因组RNA在中心示意性地示出,其翻译的ORF示为空心框。小黑框是保守序列元件;空心菱形表示渗漏乳白终止密码子。非结构化多聚蛋白及其加工产物如上所示。在乳白密码子处的终止产生P123,据信其在复制中的主要功能是作为蛋白酶,所述蛋白酶以反式起作用以加工活性RNA复制酶中的多聚蛋白;该蛋白酶结构域存在于nsP2区域中。读取乳白终止密码子产生P1234,其可形成活性复制酶。下文扩展26S亚基因组mRNA以示出结构化ORF及其翻译产物。存在于病毒粒子中的多肽以阴影示出。vcRNA为基因组RNA的负链互补物。

[0038] 本公开的上述和其他特征将结合附图由以下说明书和所附权利要求而变得更加显而易见。应当理解这些附图仅描述了根据本公开的若干实施方案,并且不应当被认为是对其范围的限制;本公开将通过使用附图由附加的特异性和细节来描述。

### 具体实施方式

[0039] 本公开整体涉及具有优异表达潜力的病毒表达体系,其适于在重组细胞中表达异源分子,例如疫苗和治疗性多肽。例如,本公开的一些实施方案涉及核酸分子,诸如表达构建体和载体,其包含经修饰的复制子RNA,所述经修饰的复制子RNA包含经修饰的5'-非翻译区(5'-UTR),并且任选地已经缺失编码结构化蛋白质的初始病毒序列中的至少一些。还包括在根据本公开的一些实施方案中的是基于病毒的表达载体,其包括编码异源多肽的一个或多个表达盒。因此,经遗传修饰以包括本文所公开的一种或多种核酸分子的重组细胞,以及来源于此类细胞的生物材料和重组产物也在本专利申请的范围内。本公开的具体方面还提供了包含本文所公开的分子和/或重组细胞中的一者或多者的组合物。

[0040] 基于RNA病毒(例如,甲病毒)的自扩增RNA(复制子)可用作稳健表达体系。例如,对野生型病毒(例如,甲病毒)5'非翻译区(UTR)的修饰可允许包含RNA复制和RNA转录两者中所涉及的启动子元件的关键RNA核苷酸的分离。通过操纵5'UTR序列来开发增强的病毒(例如,甲病毒)表达体系代表了复制子平台开发中的重要进步。不受任何具体理论的限制,据信使用甲病毒作为病毒表达载体的非限制性优点是其可指导重组宿主细胞中大量异源蛋白的合成。具体地讲,除其他优点之外,在一些实施方案中,本文所公开的甲病毒复制子平台体系还能够表达通常由甲病毒复制子表达的蛋白质量的最多三倍。这种改善是显著的,因为已注意到标准甲病毒复制子体系的天然的高表达水平,并且赋予该能力的5'UTR突变先前被认为是对甲病毒复制和转录的几乎致命的突变。例如,如果在体内以高水平表达,则多肽诸如治疗性单链抗体可能是最有效的。此外,为了产生从培养物(体外)中的细胞纯化的重组抗体,来自复制子RNA的高蛋白表达可增加抗体产物的总收率。此外,如果表达的蛋白质是疫苗抗原,则高水平表达可诱导体内最稳健的免疫应答。

[0041] 在以下具体实施方式中,参考形成本发明一部分的附图。在附图中,除非上下文另有指示,否则相同的符号通常标识相同的组件。在具体实施方式、附图和权利要求书中所述的示例性另选的替代方案不旨在进行限制。在不脱离本文提出的主题的实质或范围的情况下,可以使用其他另选的替代方案,并且可作出其他改变。应当容易理解,如大致在本文中所述的并且在附图中示出的各方面可以各种不同的构型来布置、取代、组合和设计,其全部均被明确地设想到并且成为本专利申请的一部分。

[0042] 除非另有定义,否则本文所用的所有技术术语、符号和其他科学术语或用辞均旨在具有本专利申请所属领域的技术人员通常理解的含义。在一些情况下,为了清楚起见和/或为了便于参考,本文定义了具有通常理解的含义的术语,并且本文中包括此类定义不一定解释为代表与本领域一般所理解的实质性差异。本文中所描述或引用的许多技术和程序本领域技术人员使用常规方法很好理解和通常采用的。

[0043] 一些定义

[0044] 除非上下文另有明确说明,否则单数形式“一种”、“一个”和“该”包括复数指代。例如,术语“细胞”包括一个或多个细胞,包括它们的混合物。“A和/或B”在本文中用于包括所有下列替代方案:“A”、“B”、“A或B”以及“A和B”。

[0045] 如本文所用,术语“约”具有其为大约的通常含义。在范围为如果上下文中没有明确的近似程度,则“约”意指在所提供值的正负10%内,或四舍五入到最接近的有效数字,在所有情况下均包括所提供的值。当提供范围的情况下,它们包括边界值。

[0046] 如本文所用,术语“细胞”、“细胞培养物”、“细胞系”、“重组宿主细胞”、“受体细胞”和“宿主细胞”包括原代受试者细胞和它们任何子代,但不考虑传递次数。应当理解,并非所有子代与亲本细胞都完全相同(由于有意或无意的突变或环境差异);然而,这些改变的子代包括在这些术语中,只要子代保留与最初转化的细胞相同的功能即可。

[0047] 如本文所用,术语“构建体”旨在意指来源于能够进行基因组整合或自主复制的任何源的任何重组核酸分子,诸如表达盒、质粒、粘粒、病毒、自主复制多核苷酸分子、噬菌体、或直链或环状的单链或双链DNA或RNA多核苷酸分子,其包括核酸分子,其中一个或多个核酸序列已经以功能上操作性方式连接,例如可操作地连接。

[0048] 术语“基因”广义地用于指编码蛋白质或可转录成功能性RNA的核酸分子的任何片段。基因可包括转录但不是最终、成熟和/或功能性RNA转录物的一部分的序列,并且编码蛋白质的基因还可包含转录但未翻译的序列,例如5'非翻译区、3'非翻译区、内含子等。此外,基因可任选地还包含其表达所需的调控序列,并且此类序列可以是例如未转录或翻译的序列。基因可从多种来源获得,包括从感兴趣的源克隆或由已知的或预测的序列信息合成,并且可包括被设计成具有所需参数的序列。

[0049] 当提及多核苷酸、基因或核酸分子使用时,术语“异源的”是指不来源于宿主物种的多核苷酸、基因或核酸分子。例如,如本文所用,“异源性基因”或“异源核酸序列”是指基因或核酸序列,其来自与所述基因或核酸序列被引入其中的宿主生物的物种不同的物种。当提及基因调控序列或用于操纵基因序列表达的辅助核酸序列(例如5'非翻译区、3'非翻译区、poly A加成序列等)或编码蛋白质结构域或蛋白质定位序列的核酸序列时,“异源的”意指调控序列或辅助序列或编码蛋白质结构域或定位序列的序列来自与基因不同的源,该调控核酸序列或辅助核酸序列或编码蛋白质结构域或定位序列的核酸序列与上所述基因一起并置于基因组中。因此,与在其天然状态下不可操作地连接的基因可操作地连接的启动子(例如,在非遗传工程生物体的基因组中)在本文中称为“异源启动子”,即使该启动子可来自与其所连接的基因相同的物种(或在一些情况下,相同的生物体)。例如,在本文所公开的一些实施方案中,感兴趣的异源基因(GOI)的编码序列不与处于其天然状态的EAV复制子序列连接。在一些实施方案中,编码GOI序列来源于另一生物体,诸如另一种病毒、细菌、真菌、人细胞(例如,肿瘤Ag)、寄生生物(例如,疟疾)等。

[0050] 术语“核酸分子”和“多核苷酸”在本文中可互换使用,并且是指RNA和DNA分子两者,其包括核酸分子,所述核酸分子包括cDNA、基因组DNA、合成DNA、和包含核酸类似物的DNA或RNA分子。核酸分子可具有任何三维结构。核酸分子可以是双链的或单链的(例如,有义链或反义链)。核酸分子的非限制性示例包括基因、基因片段、外显子、内含子、信使RNA(mRNA)、转移RNA、核糖体RNA、siRNA、微RNA、tracrRNA、crRNA、向导RNA、核糖酶、cDNA、重组多核苷酸、支链多核苷酸、核酸探针和核酸引物。核酸分子可包含非常规的或修饰的核苷酸。如本文所用,术语“多核苷酸序列”和“核酸序列”可互换地指多核苷酸分子的序列。本文使用如37CFR§1.822中所示的核苷酸碱基的命名。

[0051] 核酸分子可以是任何长度的核酸分子,包括但不限于以下长度的核酸分子:介于约3Kb和约50Kb之间,例如介于约3Kb和约40Kb之间,介于约3Kb和约40Kb之间,介于约3Kb和约30Kb之间,介于约3Kb和约20Kb之间,介于约5Kb和约40Kb之间,介于约5Kb和约40Kb之间,介于约5Kb和约30Kb之间,介于约5Kb和约20Kb,或介于约10Kb和约50Kb之间,例如介于约15Kb至30Kb之间,介于约20Kb和约50Kb之间,介于约20Kb和约40Kb,约5Kb和约25Kb,或约30Kb和约50Kb之间。核酸分子也可例如大于50kb。

[0052] 关于结构属性,诸如核酸与另一种核酸杂交的能力,或者多核苷酸序列能够被转录因子和/或核酸聚合酶识别和结合的能力,本公开的多核苷酸可以是“生物活性的”。

[0053] 如本文所用,术语“重组的”或“工程化的”核酸分子是指已通过人干预改变的核酸分子。作为非限制性示例,cDNA为重组DNA分子,如同已经通过体外聚合酶反应产生的任何核酸分子,或已附接连接子的任何核酸分子,或已整合到载体(诸如克隆载体或表达载体)中的任何核酸分子一样。作为非限制性示例,重组核酸分子:1)已在体外合成或修饰,例如,使用核酸分子的化学或酶促技术(例如,通过使用化学核酸合成,或通过使用酶来复制、聚合、核酸外切消化、核酸内切消化、连接、逆转录、转录、碱基修饰(包括,例如甲基化)、或重组(包括同源重组和位点特异性重组);2)包括在自然界中不结合的结合核苷酸序列,3)已使用分子克隆技术进行工程化,使得其相对于天然存在的核酸分子序列缺乏一个或多个核苷酸,和/或4)已使用分子克隆技术进行操纵,使得其相对于天然存在的核酸序列具有一个或多个序列改变或重排。作为非限制性示例,cDNA为重组DNA分子,如同已经通过体外聚合酶反应产生的任何核酸分子,或已附接连接子的任何核酸分子,或已整合到载体(诸如克隆载体或表达载体)中的任何核酸分子一样。

[0054] 如本文所用,“大部分”编码病毒结构化多肽的核酸序列可包含足够的编码病毒结构化多肽的核酸序列,以通过本领域技术人员手动评估序列,或通过使用诸如BLAST的算法的计算机自动序列比较和鉴定来提供该多肽的推定的同一性(参见,例如,“Basic Local Alignment Search Tool”;Altschul SF等人,J.Mol.Biol.215:403-410,1993)。一般来讲,本领域的技术人员将认识到,为了推定地鉴定与已知蛋白质或基因同源的多肽或核酸序列,需要十个或更多个连续氨基酸或者三十个或更多个核苷酸的序列。此外,关于核苷酸序列,包含15-30个连续核苷酸的基因特异性寡核苷酸探针可用于基因鉴定(例如,Southern杂交)和分离(例如,细菌菌落或噬菌体斑的原位杂交)的序列依赖性方法中。此外,10-15个碱基的短寡核苷酸可用作PCR中的扩增引物,以获得包含引物的特定核酸片段。因此,“大部分”核苷酸序列包含足够的序列以提供包含该序列的核酸片段的特异性鉴定和分离。本公开提供了缺乏编码一种或多种病毒结构多肽的部分或完整核酸序列的核酸分子。由于如本

文所公开的序列的有益效果,技术人员可容易地将所有或相当大部分的本发明所公开的序列用于本领域技术人员已知的目的。因此,本专利申请包括如本文公开的完整序列,例如,所附序列表中列出的那些序列,以及如上定义的那些序列中的大部分。

[0055] 如本领域普通技术人员将理解的,出于任何和所有目的,诸如就提供书面说明而言,本文公开的所有范围还涵盖任何和所有可能的子范围及其子范围的组合。任何所列的范围可容易地被识别为充分描述并使得相同的范围被分解成至少相等的一半、三分之一、四分之一、五分之一、十分之一等。作为非限制性示例,本文所讨论的每个范围均可容易地被分解成下限三分之一、中间三分之一和上限三分之一等。如本领域技术人员还将理解的,所有语言,诸如“最多至”、“至少”、“大于”、“小于”等包括所列举的数字并且是指可随后分解成如上所述的子范围的范围。最后,如本领域技术人员将理解的,范围包括每个单独的成员。因此,例如,具有1-3个制品的组是指具有1个、2个或3个制品的组。类似地,具有1-5个制品的在是指具有1个、2个、3个、4个或5个制品的组等等。

[0056] 本文所提供的一般方法的讨论本文仅旨在用于示意性的目的。在阅读本专利申请后,其他替代方法和另选的替代方案对于本领域技术人员而言是显而易见的,并且应当包括在本申请的实质和范围内。

#### [0057] 甲病毒

[0058] 甲病毒是IV族披膜病毒科的在遗传上、结构上和血清学上相关的病毒的属,其包括至少30个成员,各自具有正极性单链RNA基因组,其被包封在被包含病毒刺突蛋白的包壳包围的核壳体中。目前,甲病毒属包括辛德毕斯病毒(SIN)、塞姆利基森林病毒(SFV)、罗斯河病毒(RRV)、委内瑞拉马脑炎病毒(VEEV)和东部马脑炎病毒(EEEV),其全部密切相关并且能够感染各种脊椎动物,诸如哺乳动物、啮齿类、鱼类、禽类,和大型哺乳动物诸如人和马,以及无脊椎动物诸如昆虫。物种和个体之间的传播主要通过蚊子进行,使得甲病毒成为收集虫媒病毒或节肢动物传播的病毒的贡献者。具体地讲,已广泛研究了辛德毕斯病毒和塞姆利基森林病毒,并且这些病毒的生命周期、复制模式等被充分表征。具体地讲,已经示出甲病毒在动物细胞中非常有效地复制,这使得其作为用于在此类细胞中产生蛋白质和核酸的载体是有价值的。

[0059] 甲病毒颗粒被包封,具有70nm直径,趋于为球形(尽管稍微是多形性的),并且具有大约40nm等距核壳体。图8示出了典型的甲病毒基因组结构和基因组表达(改编自Strauss等人, *Microbiological Reviews*, 第491-562页, 1994年9月)。甲病毒基因组是长度约为11-12kb的正极性单链RNA,其包含5'帽、3' poly-A尾和两个开放阅读框,其中第一框编码具有酶促功能的非结构化蛋白质,并且第二框编码病毒结构化蛋白质(例如,衣壳蛋白C、E1糖蛋白、E2糖蛋白类、E3蛋白和6K蛋白)。

[0060] 甲病毒基因组的5'三分之二编码转录和复制病毒RNA所必需的多个非结构化蛋白质。这些蛋白质直接从RNA翻译,并且与细胞蛋白一起形成RNA依赖性RNA聚合酶,该RNA依赖性RNA聚合酶是病毒基因组复制和亚基因组RNA转录所必需的。四种非结构化蛋白质(nsP1-4)以单一多聚蛋白形式产生,其构成病毒的复制机制。多聚蛋白的加工以高度调控方式进行,其中P2/3连接处的裂解影响基因组复制期间RNA模板的使用。该位点位于狭窄裂口的基部并且不容易触及。一旦裂解,nsP3就形成环绕nsP2的环结构。这两种蛋白质具有广泛的界面。nsP2的突变在P2/P3界面区域处产生非细胞病变型病毒或温度敏感表型簇。与nsP2非细

胞病变型突变的位置相对的P3突变阻止P2/3的有效裂解。这继而可影响RNA感染,从而改变病毒RNA生产水平。

[0061] 基因组的3'三分之一包含亚基因组RNA,其充当用于翻译形成病毒颗粒所需的所有结构化蛋白质的模板:核心核衣壳蛋白C,以及作为异源二聚体缔合的薄膜蛋白P62和E1。病毒膜锚定表面糖蛋白起到受体识别作用并通过膜融合进入靶细胞。亚基因组RNA由存在于编码nsp4蛋白的RNA序列的3'端处的p26S亚基因组启动子转录。P62到E2和E3的蛋白分解成熟导致病毒表面的改变。E1、E2、以及有时E3、糖蛋白“突出”一起形成E1/E2二聚体或E1/E2/E3三聚体,其中E2从中心延伸至顶点,E1填充顶点之间的空间,并且E3(如果存在的话)位于该突出的远端处。当病毒暴露于核内体的酸性时,E1从E2解离以形成E1同源三聚体,这是用于一起驱动细胞和病毒膜的融合步骤所必需的。甲病毒糖蛋白E1是II类病毒融合蛋白,其在结构上不同于存在于流感病毒和HIV中的I类融合蛋白。E2糖蛋白用于通过其细胞质结构域与核衣壳相互作用,然而其胞外域负责结合细胞受体。大多数甲病毒失去周边蛋白E3,然而在塞姆利基森林病毒,其保持与病毒表面缔合。

[0062] 据报道,甲病毒复制在细胞的细胞质中发生。在感染周期的第一步中,基因组RNA的5'末端被翻译成具有RNA聚合酶活性的多聚蛋白(nsP1-4),其产生与基因组RNA互补的负链。在第二步中,将负链用作分别产生两个RNA的模板:(1)对应于二级病毒基因组的正基因组RNA,其通过翻译产生其他nsp蛋白并充当病毒的基因组;和(2)亚基因组RNA,其编码形成感染性颗粒的病毒的结构化蛋白质。正基因组RNA/亚基因组RNA比率通过多聚蛋白到nsp1、nsp2、nsp3和nsp4的蛋白水解自动裂解来调节。实际上,病毒基因表达分两阶段进行。在第一阶段,主要合成正基因组链和负链。在第二阶段期间,亚基因组RNA的合成实际上是排他性的,因此导致大量结构化蛋白质的产生。

[0063] 先前对甲病毒的5'-非翻译区(5'-UTR)的详细分析已经显露极端5'核苷酸支持RNA复制和转录的绝对重要性。具体地讲,如图1所示,在所有甲病毒中均注意到AU二核苷酸分别在5'UTR序列的核苷酸位置1和2处的保守性,从而表明了这些核苷酸的重要性。如本文所用,“A1”是指在5'-UTR的核苷酸位置1处的保守A核苷酸(例如,甲病毒5'-UTR),并且“U2”指在5'-UTR的核苷酸位置2处的保守U核苷酸(例如,甲病毒5'-UTR)。此外,对于委内瑞拉马脑炎病毒(VEEV),已经进行了紧接该序列后发现的5'最多三个核苷酸以及茎环区域的详细分析。具体地讲,先前已确定了保留5'UTR的位置2处的U残基的重要性(Kulasegaran-Shylini等人,J.Virol.83:17,第8327-8339页,2009a;和Kulasegaran-Shylini等人,J.Virol.83:17,第8327-8339页,2009b)。具体地讲,与来自野生型VEE/SINV感染性克隆体的体外转录的RNA相比,来自名称为(G2) VEE/SINV的全长感染性克隆体(其包含5'UTR中的单个U2->G变化)的体外转录的RNA展示出感染性的仅三个对数损失。该报告有力地表明,位置2处的U对于RNA复制是至关重要的并且不能被G置换。然而,如本文详细描述,在5'UTR中具有U2->G变化的VEEV复制子与这个先前的报告令人惊讶地直接矛盾,其不仅完全能够进行稳健复制,而且导致VEEV复制子的表达潜力相比于位置2处包含U残基的野生型5'UTR为三倍。

[0064] 大多数RNA病毒的极端5'和3'序列是高度受约束的,并且几乎没有任何变化被容忍;大多数修饰导致RNA复制的高度致死或致死的结果。Kulasegaran-Shylini等人完成了对嵌合VEEV/SINV感染性克隆体的RNA复制至关重要的5'核苷酸序列的深入分析,该克隆体

是所有甲病毒的代表 (Kulasegaran-Shylini 等人, 2009a, 同上)。该报告基于许多研究人员在 25 年期间所进行的分析, 其明确支持在任何特定甲病毒的 5' 端处可能发生的 RNA 序列变异的限制。Kulasegaran-Shylini 等人, 2009b 论文 (J. Virol. 83:17, 第 8327-8339 页, 2009) 具体地陈述/示出将 5' UTR 中的核苷酸 2 从 U 残基改变成 G 残基 (U2->G) 可显著降低该传染性克隆体 RNA 的活力。即, 5' -UTR 的该特异性变化使感染性克隆体 RNA 的生物活性降低了近 3 个数量级。如本文所公开, 掺入到 VEEV (菌株 TC-83) 复制子 RNA 中的 5' -UTR 的变化 (例如, U2->G 变化) 不仅不使复制子的复制致残, 而且实际上可增加复制子的生物活性。例如, 在一些实施方案中, 包含 U2->G 置换的复制子可导致感兴趣的蛋白质的表达超过表达相同蛋白质的野生型复制子多达三倍。该结果是令人惊讶的, 并且不能预测携带 U2->G 变化的复制子的增加的生物活性。该经修饰的复制子具有作为支持疫苗和治疗应用的优异 RNA 表达平台的潜力。

[0065] 在所有的甲病毒亚型的基因组 RNA 中均观察到 5' 最多 2 个核苷酸的保守性。保守的 AU 二核苷酸 (A1 和 U2) 也已示出是 RNA 复制所关键必需的 (Kulasegaran Shylini 等人, 2009a 和 2009b, 同上)。表明在极端 5' 端处携带 AG 二核苷酸的甲病毒复制子 RNA 不仅是完全功能性的, 而且表现出增强的生物活性, 这是令人惊讶的, 并且完全违背了该领域的教条。

[0066] 如本文所公开的, 单基因或多基因甲病毒表达体系可通过使用具有表达/翻译增强活性的经修饰的复制子 RNA 产生, 例如包含修饰的 5' -UTR 的复制子 RNA。在一些实施方案中, 如本文所述的病毒 (例如, 甲病毒) 表达体系还缺乏一种或多种病毒结构化蛋白质的部分或整个编码区。例如, 甲病毒表达体系可能缺乏病毒衣壳蛋白 C、E1 糖蛋白、E2 糖蛋白、E3 蛋白和 6K 蛋白中的一者或多者的一部分或整个编码序列。在一些实施方案中, 与具有野生型 5' UTR 序列的 VEEV 复制子相比, 委内瑞拉马脑炎病毒 (VEEV) 5' UTR 序列的 cDNA 拷贝中的位置 2 处的核苷酸从胸腺嘧啶 (T) 核苷酸到鸟嘌呤 (G) 核苷酸 (T2->G 突变) 的修饰赋予该复制子显著更高的蛋白表达潜力。

[0067] 在一些实施方案中, 相对于从相应的未修饰复制子 (例如具有野生型 5' UTR 的复制子) 检测到的表达水平, 如本文所公开的经修饰复制子 RNA 的表达水平和/或翻译增强活性为至少 1.1 倍、1.2 倍、1.3 倍、1.4 倍、1.5 倍、1.6 倍、1.7 倍、1.8 倍、1.9 倍、2 倍 (2 倍)、3 倍、4 倍、5 倍、6 倍、7 倍、8 倍、或更多倍。不受任何特定理论的限制, 增强的翻译可由于转录的增强, 这导致增加水平的转录物可用于翻译和/或可不依赖于转录并且是由于例如增强的核糖体结合。增强活性的水平可通过本领域已知的任何便利的方法和技术测量, 其包括但不限于转录物水平、蛋白质的量、蛋白活性等 (参见例如下文实施例 1、3-5)。

#### [0068] 核酸分子

[0069] 在一个方面, 本文公开了包括经修饰复制子 RNA 的新型核酸分子。例如, 经修饰的复制子 RNA 可包含亲本复制子 RNA 的一个或多个原始基因组区域中的突变、缺失、置换和/或插入 (例如, 开放阅读框 (ORF)) 和/或非编码区 (例如, 启动子序列)。在一些实施方案中, 经修饰的复制子 RNA 包括经修饰的 5' -非翻译区 (5' -UTR)。在一些实施方案中, 经修饰的 5' -UTR 包括在位置 1、2、4、或它们的组合处的一个或多个核苷酸置换。在一些实施方案中, 该核苷酸置换中的至少一个为在经修饰的 5' -UTR 的位置 1 处的核苷酸置换。在一些实施方案中, 该核苷酸置换中的至少一个为在经修饰的 5' -UTR 的位置 2 处的核苷酸置换。在一些实施方案中, 该核苷酸置换中的至少一个为在经修饰的 5' -UTR 的位置 4 处的核苷酸置换。在一些实

施方案中,在经修饰的5'-UTR的位置2处的核苷酸置换是U->G置换。在一些实施方案中,在经修饰的5'-UTR的位置2处的核苷酸置换是U->A置换。在一些实施方案中,在经修饰的5'-UTR的位置2处的核苷酸置换是U->C置换。

[0070] 如本文所用,术语“复制子RNA”是指包含在允许细胞内指导其自身扩增或自我复制所需的所有遗传信息的RNA。为了指导其自身的复制, RNA分子1) 编码聚合酶、复制酶或其他蛋白质,其可以与病毒或宿主细胞来源的蛋白质、核酸或核蛋白相互作用以催化RNA扩增过程;以及2) 包含亚基因组复制子编码的RNA的复制和转录所需的顺式作用RNA序列。这些序列可在复制过程中结合到其自编码蛋白质,或非自编码的细胞衍生的蛋白质、核酸或核糖核酸、或这些组分中的任一种之间的复合物。出于本公开的目的,甲病毒来源的复制子RNA分子通常包含以下有序元件:顺式复制所需的5'病毒或缺陷型干扰RNA序列、编码生物活性甲病毒非结构化蛋白质的序列(例如,nsP1、nsP2、nsP3和nsP4)、亚基因组RNA的启动子、顺式复制所需的3'病毒序列、以及聚腺苷酸核束。此外,术语复制子RNA通常是指具有正极性或“信息”意义的分子,并且复制子RNA的长度可以与任何已知的天然存在的甲病毒的长度不同。在本公开的一些实施方案中,复制子RNA不包含结构化病毒蛋白质中的至少一个的序列;编码结构基因的序列可被异源序列取代。在将复制子RNA包装到重组甲病毒颗粒中的那些情况下,其必须包含一个或多个序列,所谓的包装信号,其用于引发与导致颗粒形成的甲病毒结构化蛋白质的相互作用。

[0071] 如本文所用,“亚基因组RNA”是指长度或尺寸小于其来源的基因组RNA的RNA分子。甲病毒亚基因组RNA应当从内部启动子转录,所述内部启动子的序列位于基因组RNA或其互补序列内。甲病毒亚基因组RNA的转录可由基于宿主细胞编码的蛋白质、核蛋白或它们的组合缔合的病毒编码聚合酶介导。在本公开的一些实施方案中,亚基因组RNA由如本文所公开的经修饰的复制子RNA产生,并且编码或表达一种或多种感兴趣的基因(GOI)。代替天然亚基因组启动子,可以将亚基因组RNA置于来源于下列病毒的内部核糖体进入位点(IRES)的控制下:脑心肌炎病毒(EMCV)、牛病毒性腹泻病毒(BVDV)、脊髓灰质炎病毒、(FMD),口蹄疫病毒(FMD)、肠病毒71型或丙型肝炎病毒。

[0072] 因此,在一些实施方案中,一种或多种病毒结构化蛋白的一部分或整个编码序列在本文所公开的核酸分子中不存在和/或被修饰。因此,在一些具体实施方案中,如本文所公开的经修饰的复制子RNA包括经修饰的5'-UTR并且缺乏编码一种或多种病毒结构化蛋白质的核酸序列的至少一部分,例如,缺乏编码病毒结构化蛋白质的核酸序列中的前一个、两个、三个、四个、五个、六个、七个、八个、九个、十个或更多个核苷酸。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA可缺乏约10%、20%、30%、40%、50%、60%、70%、80%、90%、95%、或更多的编码结构化多肽E1、E2、E3、6K和衣壳蛋白C中的一种或多种的序列。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA缺乏编码结构化多肽E1、E2、E3、6K和衣壳蛋白C中的一种或多种的相当大部分或整个序列。如本文所用,“相当大部分”编码病毒结构化蛋白质的核酸序列包含足够的编码病毒结构化蛋白质的核酸序列,以通过本领域技术人员手动评估序列,或通过使用诸如BLAST的算法(参见,例如,Altschul SF等人,1993,同上)的计算机自动序列比较和鉴定来提供该蛋白质的推定的同一性。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA缺乏编码结构化多肽E1、E2、E3、6K和衣壳蛋白C中的一种或多种的整个序列。

[0073] 在本专利申请的一些具体实施方案中,如本文所公开的核酸分子包括经修饰的甲病毒基因组或复制子RNA,其中所述核酸分子表现出与SEQ ID NO:1的核酸序列的至少80%序列同一性,其中所述经修饰的甲病毒基因组或复制子RNA在5'-非翻译区(5'-UTR)的位置2处包含U->G置换,并且缺乏编码病毒结构化蛋白质的序列的至少一部分。在一些实施方案中,该核酸分子表现出与SEQ ID NO:1的核酸序列的至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性。在一些实施方案中,核酸分子表现出与SEQ ID NO:1的核酸序列的100%序列同一性。

[0074] 在一些实施方案中,本文所公开的核酸分子包括经修饰的甲病毒基因组或复制子RNA,其中所述经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与至少一种本文所公开的5'-UTR的核酸序列的至少80%序列同一性,和在该5'-UTR的位置2处的U->G置换,并且其中所述经修饰的甲病毒基因组或复制子RNA缺乏编码病毒结构化蛋白质的序列的至少一部分。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与SEQ ID NO:2-18中所示序列中至少一种的至少80%序列同一性。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与SEQ ID NO:2-18中所示序列中至少一种的至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%序列同一性。在一些实施方案中,经修饰的甲病毒基因组或复制子RNA包含5'-UTR,所述5'-UTR表现出与序列列表中的SEQ ID NO:2-18中所示序列中的至少一种的100%序列同一性。

[0075] 可通过使用本文所鉴定的序列(例如SEQ ID NO:1-18)或任何其他甲病毒5' UTR(如本领域已知的),例如,具有基因库/NCBI登录号J02363、NC\_001547、U38305、L04653、NC\_001449、U38304、X04129、NC\_003215和TR339基因组序列的序列,来鉴定和/或分离核酸分子,所述核酸分子与本文所公开的5' UTR具有高度序列同一性(例如,至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%)(Klimstra等人,J.Virol.72:7357,1988;McKnight等人,J.Virol.70:1981,1996),其通过基因组序列分析、杂交和/或PCR,利用来自相应甲病毒基因组中鉴定的序列的简并引物或基因特异性引物来进行。如本文所用的,“序列同一性”是指两个最佳比对的多核苷酸在整个组分(例如核苷酸)比对窗口中不变的程度。测试序列和参考序列的比对区段的“同一性分数”是由两个比对序列共享的相同组分数除以参考序列区段中的组分总数,例如整个参考序列或参考序列的较小限定部分。

[0076] 本文所公开的一些实施方案涉及包含经修饰的复制子RNA的核酸分子,其中经修饰的复制子RNA包含具有翻译增强活性的经修饰甲病毒碱基序列,例如5'UTR。此类经修饰的复制子RNA可用于实现编码期望产物的异源核酸序列(例如,DNA或cDNA)的增强的表达水平。在一些实施方案中,经修饰的复制子RNA用于在细胞中引入经修饰的复制子后,实现编码期望产物的异源核酸序列(例如,DNA或cDNA)的增强的表达水平,所述细胞可以为例如细胞培养中的细胞或可以为活体中的细胞。

[0077] 此外,在一些实施方案中,所述核酸分子可包括经修饰的甲病毒基因组或复制子RNA,其包含一个或多个减毒突变,以便增加病毒操纵和/或施用的安全性。如本文所用,短语“减毒突变”是指根据本领域的标准术语,核苷酸突变或根据此类突变编码的氨基酸,其导致引起其宿主疾病的概率降低(即,丧失毒力),无论该突变是置换突变还是框内缺失或

插入突变。减毒突变可处于甲病毒基因组的编码区或非编码区(例如,5'UTR)中。如本领域的技术人员已知,短语“减毒突变”不包括可对病毒而言致死的突变或突变的组合。此外,本领域的技术人员将会知道,在不存在第二位点阻遏突变的情况下,一些减毒突变可能是致死的。

[0078] 构建和表征这些新核酸分子的分子技术和方法在本专利申请的实施例中更全面地描述。在实施例部分中,已经使用了委内瑞拉马脑炎病毒(VEEV)来说明本文所公开的组合物和方法。

[0079] 在一些实施方案中,核酸分子是重组核酸分子。如本文所用,术语重组意指任何分子(例如DNA、RNA、多肽),即,或通过人操纵间接产生。作为非限制性示例,cDNA为重组DNA分子,如同已经通过体外聚合酶反应产生的任何核酸分子,或已附接连接子的任何核酸分子,或已整合到载体(诸如克隆载体或表达载体)中的任何核酸分子一样。作为非限制性示例,重组核酸分子:1)已在体外合成或修饰,例如,使用核酸分子的化学或酶促技术(例如,通过使用化学核酸合成,或通过使用酶来复制、聚合、核酸外切消化、核酸内切消化、连接、逆转录、转录、碱基修饰(包括,例如甲基化)、或重组(包括同源重组和位点特异性重组);2)包括在自然界中不结合的结合核苷酸序列;3)已使用分子克隆技术进行工程化,使得其相对于天然存在的核苷酸序列缺乏一个或多个核苷酸;和/或4)已使用分子克隆技术进行操纵,使得其相对于天然存在的核苷酸序列具有一个或多个序列改变或重排。

[0080] 在一些实施方案中,本文所公开的核酸分子使用重组DNA技术(例如聚合酶链反应(PCR)扩增、克隆等)或化学合成来产生。如本文所公开的核酸分子包括天然核酸分子及其同源物,其包括但不限于天然等位变体和经修饰的核酸分子,其中已经以以下方式插入、缺失和/或置换一个或多个核苷酸残基,使得此类修饰提供实现如本文所述生物活性的所需特性。

[0081] 核酸分子(包括天然存在的核酸序列的变体)可使用本领域技术人员已知的多种方法制备(参见例如Sambrook等人,In:Molecular Cloning,ALaboratory Manual,第2版,Cold Spring Harbor Press,Cold Spring Harbor,N.Y. (1989))。核酸分子的序列可使用多种技术相对于其所来源的天然存在的序列进行修饰,所述技术包括但不限于经典诱变技术和重组DNA技术,诸如但不限于位点-定向诱变、核酸分子的化学处理以诱导突变、核酸片段的限制酶裂解、核酸片段的连接、核酸序列的选定区域的PCR扩增和/或诱变,重组克隆,以及化学合成,其包括寡核苷酸混合物的化学合成和连接混合物基团以“构建”核酸分子的混合物,以及它们的组合。通过筛选蛋白质或核酸分子编码的复制子的功能和/或通过与野生型基因或其片段杂交,或通过PCR,使用与靶或野生型核酸分子或序列具有同源性的引物,可从经修饰的核酸分子的混合物中选择核酸分子同源物。

[0082] 在本文所公开的各种实施例中,本文所公开的核酸分子可包括下列特征中的一个或多个。在一些实施方案中,经修饰的复制子RNA为经修饰的甲病毒复制子RNA。在一些实施方案中,经修饰的甲病毒复制子RNA包括经修饰的甲病毒基因组。在一些实施方案中,经修饰的5'-UTR包括在位置1、2、4、或它们的组合处的一个或多个核苷酸置换。在某些实施方案中,该核苷酸置换中的至少一个为在经修饰的5'-UTR的位置2处的核苷酸置换。在一些具体实施方案中,在经修饰的5'-UTR的位置2处的核苷酸置换是U→G置换。

[0083] 在本文所公开的一些实施方案中,经修饰的甲病毒基因组或复制子RNA可操作地

连接至异源调控元件。如本文所用,“调控元件”、“调控序列”或“调控元件序列”是指位于编码序列,例如多肽编码序列或功能性RNA-编码序列的上游(5')、内或下游(3')的核苷酸序列。编码序列的转录和/或由编码序列的转录产生的RNA分子的翻译通常受存在或不存在调控元件影响。这些调控元件可包括启动子、顺式元件、增强子、终止子或内含子。本领域的技术人员将会知道,本文所述的调控元件可存在于嵌合或杂交调控表达元件中。在一些实施方案中,异源调控元件为或包含启动子序列。异源启动子序列可以为任何异源启动子序列,例如SP6启动子、T3启动子、或T7启动子、或它们的组合。在一些具体实施方案中,启动子序列包括T7启动子序列。

[0084] 另外,在一些实施方案中,经修饰的甲病毒基因组或复制子RNA可包括一种或多种异源转录终止信号序列。如本文中可互换使用的,术语“转录终止信号”、“终止子”或“终止子序列”或“转录终止子”是指导致RNA聚合酶停止转录的基因序列的调控部分。异源转录终止信号序列通常可以为任何异源转录终止信号序列,并且例如为SP6终止信号序列、T3终止信号序列、T7终止信号序列、或它们的变体。因此,根据本公开的一些实施方案的核酸分子可包括一种或多种异源转录终止信号序列中的至少一种,所述异源转录终止信号序列选自SP6终止信号序列、T3终止信号序列、T7终止信号序列、或它们的变体。在一些具体实施方案中,转录终止序列包括T7终止信号序列。

[0085] 在一些实施方案中,本文所公开的核酸分子可包括一个或多个表达盒。原则上,本文所公开的核酸分子通常可包括任何数目的表达盒。在一些具体实施方案中,本文所公开的核酸分子可包括至少两个、至少三个、至少四个、至少五个、或至少六个表达盒。如本文所用,术语“表达盒”是指遗传物质的构建体,其含有编码序列和足够的调控信息,以在体内和/或体外指导受体细胞中编码序列的正确转录和/或翻译。可将表达盒插入载体中以靶向期望的宿主细胞和/或靶向到受试者中。此外,术语表达盒可与术语“表达构建体”互换使用。在一些实施方案中,术语“表达盒”指核酸构建体,其包括编码可操作地连接至调控元件(例如启动子和/或终止信号)的蛋白质或功能性RNA的基因,以及任选地,影响基因转录或翻译的其他核酸序列中的任一种或组合。

[0086] 如本文所用,术语“可操作地连接”表示两个或更多个序列之间的功能性连接。例如,感兴趣的多核苷酸和调控序列(例如启动子)之间的可操作地连接是允许表达感兴趣的多核苷酸的功能性连接。在这个意义上,术语“可操作地连接”是指调节区和待转录的编码序列的定位,使得调节区对于调节感兴趣的编码序列的转录或翻译是有效的。在本文所公开的一些实施方案中,术语“可操作地连接”表示其中将调控序列相对于编码多肽或功能性RNA的序列置于适当位置处的构型,使得该控制序列指导或调控编码多肽的mRNA、多肽和/或功能性RNA的表达或细胞定位。因此,如果启动子能够介导核酸序列的转录,则其与核酸序列可操作地连接。可操作地连接的元件可以是连续的或不连续的。

[0087] 将两个或更多个DNA序列可操作地连接在一起的基础技术是技术人员所熟悉的,并且此类方法已在多个标准分子生物学操作的教科书中描述(参见,例如,Maniatis等人,“Molecular Cloning:A Laboratory Manual”,第2版,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.;以及Gibson等人,Nature Methods 6:343-45,2009)。

[0088] 因此,本文所提供的核酸分子可用作例如表达载体,当包括与异源核酸序列可操作地连接的调控元件时,所述表达载体可影响异源核酸序列的表达。在一些实施方案中,异

源核苷酸序列包括感兴趣的基因 (GOI) 的编码序列。在一些实施方案中,GOI 的编码序列被优化用于以高于参考编码序列的表达水平的水平进行表达。在一些实施方案中,参考编码序列是尚未优化的序列。在一些实施方案中,GOI 编码序列的优化可包括序列优化。关于核苷酸序列的序列优化,遗传密码的简并性提供了用不同碱基置换基因的蛋白质编码序列的至少一个碱基的可能性,但不导致由该基因产生的多肽的氨基酸序列被改变。因此,本专利申请申请的核酸分子还可具有任何碱基序列,其根据遗传密码的简并性通过置换从本文所公开的任何多核苷酸序列改变。描述密码子使用的参考文献容易公开获得。在一些实施方案中,出于多种原因可产生多核苷酸序列变体,例如用于优化特定宿主的表达 (例如,将甲病毒 mRNA 中的密码子使用改变成其他生物体诸如人、仓鼠、小鼠或猴优选的那些)。

[0089] 由 GOI 编码的多肽一般可以为任何多肽,并且可以为例如治疗性多肽、预防性多肽、诊断性多肽、营养性多肽或工业酶。在一些实施方案中,GOI 编码选自下列的多肽:抗体、抗原、免疫调节剂、和细胞因子。

[0090] 在一些实施方案中,GOI 的编码序列针对所期望的特性进行优化。在一些实施方案中,GOI 的编码序列被优化用于以高于参考编码序列的表达水平的水平进行表达。

[0091] 在一些实施方案中,本文所公开的经修饰的基因组或复制子 RNA 为甲病毒的基因组或复制子 RNA,诸如属于披盖病毒 (Togaviridae) 家族的甲病毒属的病毒物种的基因组或复制子 RNA。在一些实施方案中,经修饰的基因组或复制子 RNA 具有属于 VEEV/EEEV 组、或 SF 组或 SIN 组的甲病毒 (综述参见例如, Strauss 和 Strauss, Microbiol. Rev. 58:3, 第 492-562 页, 1994)。Sf 族甲病毒的非限制性示例包括:塞姆利基森林病毒、阿尼昂-尼昂病毒、罗斯河病毒、米德尔堡病毒、奇昆古尼亚病毒、巴马森林病毒、盖塔病毒、马雅罗病毒、鹭山病毒、比巴鲁病毒和乌纳病毒。SIN 族甲病毒的非限制性示例包括辛德毕斯病毒、Girdwood S.A. 病毒、南非虫媒病毒 86 号、奥克尔博病毒、奥拉病毒、巴班肯病毒、瓦塔罗阿病毒和孜拉加奇病毒。VEEV/EEEV 族甲病毒的非限制示例包括:东部马脑炎病毒 (EEEV)、委内瑞拉马脑炎病毒 (VEEV)、沼泽地病毒 (EVEV)、穆坎布病毒 (MUCV)、那皮舒纳病毒 (PIXV)、米德堡病毒 (MIDV)、奇昆古尼亚病毒病 (CHIKV)、阿尼昂-尼昂病毒 (ONNV)、罗斯河病毒 (RRV)、巴马森林病毒 (BF)、盖塔病毒 (GET)、鹭山病毒 (SAGV)、贝巴鲁病毒 (BEBV)、马雅罗病毒 (MAYV) 和乌纳病毒 (UNAV)。

[0092] 甲病毒物种的非限制性示例包括:东部马脑炎病毒 (EEEV)、委内瑞拉马脑炎病毒 (VEEV)、沼泽地病毒 (EVEV)、穆坎布病毒 (MUCV)、那皮舒纳病毒 (PIXV)、米德堡病毒 (MIDV)、奇昆古尼亚病毒病 (CHIKV)、阿尼昂-尼昂病毒 (ONNV)、罗斯河病毒 (RRV)、巴马森林病毒 (BF)、盖塔病毒 (GET)、鹭山病毒 (SAGV)、贝巴鲁病毒 (BEBV)、马雅罗病毒 (MAYV)、乌纳病毒 (UNAV)、辛德毕斯病毒 (SINV)、奥拉病毒 (AURAV)、沃达罗河病毒 (WHAV)、巴班肯病毒 (BABV)、孜拉加奇病毒 (KYZV)、西部马脑炎病毒 (WEEV)、高地 J 病毒 (HJV)、摩根堡病毒 (FMV)、恩杜穆病毒 (NDUV)、和博吉河病毒。毒性和无毒性甲病毒菌株均是合适的。在一些具体实施方案中,经修饰的基因组或复制子 RNA 为辛德毕斯病毒 (SIN)、塞姆利基森林病毒 (SFV)、罗斯河病毒 (RRV)、委内瑞拉马脑炎病毒 (VEEV) 或东部马脑炎病毒 (EEEV)。在一些实施方案中,经修饰的基因组或复制子 RNA 具有委内瑞拉马脑炎病毒 (VEEV)。

[0093] 重组细胞

[0094] 在一个方面,本文所公开的一些实施方案涉及转化细胞的方法,所述方法包括在

宿主细胞,诸如动物细胞中引入本文所提供的核酸分子,并且选择或筛选转化的细胞。术语“宿主细胞”和“重组宿主细胞”在本文中可互换使用。应当理解,此类术语不仅涉及具体受试者细胞,而且涉及此类细胞的子代或潜在子代。因为某些修饰可由于突变或环境影响而在子代中发生,所以此类子代实际上可与亲本细胞不同,但仍包括在本文所用术语的范围内。在一些实施方案中,核酸分子通过电穿孔程序或基因枪程序引入宿主细胞中。

[0095] 在一个相关的方面,一些实施方案涉及重组宿主细胞,例如包含本文所述的核酸分子的重组动物细胞。核酸分子可稳定地整合在宿主基因组中,或者可以游离复制,或作为用于稳定或瞬时表达的微环表达载体存在于重组宿主细胞中。因此,本文所公开的一些实施方案中,核酸分子作为附加型单元在重组宿主细胞中被维持和复制。在一些实施方案中,核酸分子稳定地整合到重组细胞的基因组中。稳定整合可使用经典随机基因组重组技术或利用更精确的基因组编辑技术来完成,诸如使用向导RNA引导的CRISPR/Cas9,或DNA引导的内切核酸酶基因组编辑NgAgo(格氏嗜盐碱杆菌Argonaute),或TALEN基因组编辑(转录激活因子样效应子核酸酶)。在一些实施方案中,核酸分子作为用于稳定表达或瞬时表达的微环表达载体存在于重组宿主细胞中。

[0096] 在一些实施方案中,宿主细胞可利用例如本专利申请的载体构建体进行遗传工程化(例如,转导或转化或转染),所述载体构建体可以是例如用于同源重组的载体,其包括与宿主细胞的基因组的一部分同源的核酸序列,或者可以为用于表达感兴趣的基因中任一种或组合的表达载体。载体可以为例如呈质粒、病毒粒子、噬菌体等的形式。在一些实施方案中,用于表达感兴趣的多肽的载体也可被设计用于例如通过同源重组整合到宿主中。包含如本文所述的多核苷酸序列的载体,例如包含经修饰的甲病毒基因组或复制子RNA的核酸分子,以及任选地选择性标记或报告基因,可用于转化适当的宿主细胞。

[0097] 本文所公开的方法和组合物可被部署用于任何物种的遗传工程,包括但不限于原核物种和真核物种。使用根据本公开的组合物和方法修饰的合适的宿主细胞可包括但不限于藻类细胞、细菌细胞、长短鞭毛体、真菌细胞、衣原体细胞、小型真菌、微藻和动物细胞。在一些实施方案中,动物细胞是无脊椎动物细胞。在一些实施方案中,脊椎动物细胞是哺乳动物细胞。宿主细胞可以为未转化的细胞或已经利用至少一个核酸分子转染的细胞。

[0098] 本文所公开的方法和组合物可与例如受试者或宿主细胞一起使用,所述受试者或宿主细胞对于水产养殖、农业、畜牧业和/或治疗性和药物应用而言是重要或感兴趣的,包括生产用于制造疫苗、药物产品、工业产品、化学品等的多肽。在一些实施方案中,本文所公开的组合物和方法可以与来自作为甲病毒天然宿主的物种的宿主细胞一起使用,诸如啮齿动物、小鼠、鱼、鸟和大型哺乳动物,诸如人、马、猪、猴和猿,以及无脊椎动物。在本专利申请的一些实施方案中,特别优选的物种是脊椎动物物种和无脊椎动物物种。原则上,通常可使用任何动物物种,并且可以为例如人、狗、鸟、鱼、马、猪(pig)、灵长类动物、小鼠、牛、猪(swine)、绵羊、兔子、猫、山羊、驴、仓鼠或水牛。合适的鸟类物种的非限制性示例包括鸡、鸭、鹅、火鸡、鸵鸟、鹑、天鹅、孔雀、野鸡、鹧鸪和珍珠鸡。在一些具体的实施方案中,鱼类物种为鲑鱼物种。原代哺乳动物细胞和连续/永生化细胞类型也是合适的。合适的动物细胞的非限制性示例包括但不限于,马肺动脉内皮细胞、马真皮细胞、幼仓鼠肾(BHK)细胞、兔肾细胞、小鼠肌细胞、小鼠结缔组织细胞、人子宫颈细胞、人表皮喉细胞、中国仓鼠卵巢细胞(CHO)、人HEK-293细胞、小鼠3T3细胞、Vero细胞、Madin-Darby犬肾上皮细胞(MDCK)、原代鸡

成纤维细胞、HuT78细胞、A549肺细胞、HeLa细胞、**PER.C6<sup>®</sup>**细胞、WI-38细胞、MRC-5细胞、FRhL-2和CEM T细胞。在一些实施方案中,宿主细胞是幼仓鼠肾细胞。在一些实施方案中,幼仓鼠肾细胞为BHK-21细胞。

[0099] 用于转化多种上述宿主细胞和物种的技术是本领域已知的并且在技术文献和科学文献中有所描述。因此,包括如本文所公开的至少一种重组细胞的细胞培养物也在本专利申请的范围内。适用于产生和维持细胞培养物的方法和体系是本领域已知的。

[0100] 用于产生多肽的方法

[0101] 本公开的宿主细胞诸如原核宿主细胞或真核宿主细胞可用于产生(即表达)在如本文所公开的感兴趣的基因(GOI)的开放阅读框中编码的感兴趣的分子,例如多肽。因此,本专利申请还提供了使用本公开的宿主细胞生产感兴趣的分子(例如多肽)的方法,所述宿主细胞可以为例如细胞培养物中的细胞或可以为活体中的细胞。

[0102] 因此,本文所公开的一些实施方案提供了用于在宿主细胞中产生感兴趣的多肽的方法。此类方法包括培养重组宿主细胞,包括根据前述方面和实施方案中的任一项所述的核酸分子。在一些实施方案中,所述方法包括在合适的培养基中培养本发明的宿主细胞(其中已引入编码感兴趣的分子的重组表达载体),使得感兴趣的分子产生。在一些实施方案中,所述方法还包括将感兴趣的分子与培养基或宿主细胞分离。

[0103] 在另一个方面,在一些实施方案涉及用于在受试者中产生感兴趣的多肽的方法,所述方法包括向受试者施用根据前述方法和实施方案中任一项所述的核酸分子。

[0104] 用于本文所公开的方法和组合物中的合适的宿主细胞和/或受试者包括但不限于原核物种和真核物种。使用根据本公开的组合物和方法修饰的合适的宿主细胞可包括但不限于藻类细胞、细菌细胞、长短鞭毛体、真菌细胞、衣原体细胞、小型真菌、微藻和动物细胞。在一些实施方案中,动物细胞是无脊椎动物细胞。在一些实施方案中,脊椎动物细胞是哺乳动物细胞。宿主细胞可以为未转化的宿主细胞或已经利用至少一个核酸分子转染的细胞。因此,根据前述方面和实施方案中任一项所述的重组细胞的生物样品、生物质和子代也在本专利申请的范围内。因此,如下文所更详细地讨论的,通过根据本专利申请的这一方面的方法产生的多肽也在本专利申请的范围内。

[0105] 在一些实施方案中,该重组细胞为动物细胞。小规模和大规模的治疗性蛋白质生产是制药工业中的重要发展领域,因为据信动物细胞中产生的蛋白质通常具有适当的加工、翻译后修饰,因此具有用于治疗生理病症的足够的活性。原则上,通常可使用任何动物物种,并且可以为例如人、狗、鸟、鱼、马、猪(pig)、灵长类动物、小鼠、牛、猪(swine)、绵羊、兔子、猫、山羊、驴、仓鼠或水牛。合适的鸟类物种的非限制性示例包括鸡、鸭、鹅、火鸡、鸵鸟、鹌、天鹅、孔雀、野鸡、鹧鸪和珍珠鸡。在一些具体的实施方案中,鱼类物种为鲑鱼物种。原代哺乳动物细胞和连续/永生化细胞类型也是合适的。合适的动物宿主细胞的非限制性示例包括但不限于,马肺动脉内皮细胞、马真皮细胞、幼仓鼠肾(BHK)细胞、兔肾细胞、小鼠肌细胞、小鼠结缔组织细胞、人子宫颈细胞、人表皮喉细胞、中国仓鼠卵巢细胞(CHO)、人HEK-293细胞、小鼠3T3细胞、Vero细胞、Madin-Darby犬肾上皮细胞(MDCK)、原代鸡成纤维细胞、HuT78细胞、A549肺细胞、HeLa细胞、**PER.C6<sup>®</sup>**细胞、WI-38细胞、MRC-5细胞、FRhL-2和CEM T细胞。在一些实施方案中,宿主细胞是幼仓鼠肾细胞。在一些实施方案中,幼仓鼠肾细

胞为BHK-21细胞。

#### [0106] 重组多肽

[0107] 本文所公开的一些实施方案涉及通过根据本文所述的一个或多个实施方案的方法产生的重组多肽。本专利申请的重组多肽一般可以为任何重组多肽,并且可以为例如下列中的一种或多种:治疗性多肽、预防性多肽、诊断性多肽、营养性多肽、工业酶和报告多肽。在一些实施方案中,重组多肽可以为下列中的一种或多种:抗体、抗原、免疫调节剂和细胞因子。在一些实施方案中,感兴趣的多肽可具有治疗活性或预防活性。

#### [0108] 组合物

[0109] 本文所公开的一些实施方案涉及包含本文所述的任何重组多肽的组合物。所述组合物可以例如为营养性组合物、预防性组合物、包含药学上可接受的载体的药物组合物、或它们的混合物。在一些实施方案中,本专利申请的组合物可用作疫苗。

[0110] 本文所公开的一些实施方案涉及包含本文所述的核酸分子中任一种的组合物。所述组合物可以例如为营养性组合物、预防性组合物、包含药学上可接受的载体的药物组合物、或它们的混合物。在一些实施方案中,本专利申请的组合物可用作疫苗。

[0111] 本文所公开的一些实施方案涉及包含本文所述的重组细胞中的任一种的组合物。所述组合物可以例如为营养性组合物、预防性组合物、包含药学上可接受的载体的药物组合物、或它们的混合物。在一些实施方案中,本专利申请的组合物可用作疫苗。

[0112] 如本文所用,术语“药学上可接受的载体”是指可用于制备药物组合物或制剂的载体,所述药物组合物或制剂通常是安全的、无毒性的,并且既不是生物学上不可取的也不是其他方面不可取的,并且包括兽医学用途以及人类药物用途可接受的载体。在一些实施方案中,药学上可接受的载体与水一样简单,但其也可包括例如生理盐浓度的溶液。在一些实施方案中,药学上可接受的载体可以为或可包括稳定剂、稀释剂和缓冲剂。合适的稳定剂例如为SPGA、碳水化合物(诸如奶粉、血清白蛋白或酪蛋白)或它们的降解产物。合适的缓冲剂例如为碱金属磷酸盐。稀释剂包括水、含水缓冲剂(诸如缓冲盐水)、醇和多元醇(诸如甘油)。对于施用于动物或人而言,根据本专利申请的组合物可通过喷雾、皮内、皮下、口服、通过气溶胶或肌肉内给药,尤其是鼻内给药。

[0113] 本公开提及的所有出版物和专利申请以引用方式并入本文,其程度如同每个单独的出版物或专利申请被具体地和单独地指示以引用方式并入本文那样。

[0114] 不承认本文所引用的任何参考文献构成现有技术。对参考文献的讨论陈述了其作者所主张的内容,并且发明人保留对所引用的文献的准确性和相关性提出质疑的权利。应当清楚地理解,虽然本文提及了多种信息来源,包括科学期刊文章、专利文献和教科书;但该参考文献并不构成对这些文件中的任一个形成本领域公知常识的一部分的承认。

[0115] 本文所给出的一般方法的讨论本文仅旨在用于示意性的目的。在阅读本公开后,其他替代方法和另选的替代方案对于本领域技术人员而言是显而易见的,并且应当包括在本专利申请的实质和范围内。

#### [0116] 实施例

[0117] 在以下实施例中进一步详细公开了附加的替代方案,这些实施例不旨在以任何方式限制权利要求的范围。

#### [0118] 实施例1

### [0119] 一般实验程序

#### [0120] 体外转录

[0121] 使用1 $\mu$ g如上所述制备的DNA模板,经过在37℃下温育一小时,在20 $\mu$ l反应中进行体外转录 (IVT) 反应 (NEB目录号E2065S)。然后将由供应商提供的1单位DNA酶直接加入IVT反应中,并在37℃下温育另外15分钟。然后将反应置于冰上,并使用制造商建议的方法纯化 (Qiagen目录号74104)。然后使用NanoDrop 2000c UV-Vis分光光度计定量纯化的RNA。在进行电穿孔之前,通过电泳穿过0.8%琼脂糖凝胶 (Life Technologies目录号G5018-08),并与Millennium RNA标记 (Ambion目录号AM7150) 进行比较来观察RNA。

[0122] 从在补充有50ng/ml氨基甲氧苄青霉素 (Teknova目录号NC9730116) 的LB肉汤培养基 (Teknova目录号L8000 06) 中生长的300mL饱和大肠杆菌TransforMax Epi300 (Epicentre目录号EC300105) 培养物中,纯化质粒DNA模板 (Qiagen目录号12163)。通过Not-I消化 (New England Biolabs NEB目录号R3189S) 将质粒DNA在37℃线性化一小时。然后将线性模板DNA再次纯化 (Zymo目录号D4003),并通过0.8%琼脂糖凝胶 (Life Technologies目录号G5018-08) 针对商业2-log DNA ladder (New England Biolabs, NEB目录号N3200S) 进行分析。在进行体外转录之前,确认每个样品中存在单一条带,这对应于线性DNA模板的预期片段大小。

#### [0123] 转染和分析

[0124] 在典型的细胞转染实验中,通过使用用于4D-Nucleofector™系统 (Lonza) 的SF细胞系Nucleofector™试剂盒进行电穿孔将复制子RNA引入BHK-21细胞中。使用0.25%胰蛋白酶收获BHK-21细胞并用冷PBS洗涤一次。将细胞以 $1 \times 10^6$ 个细胞/20 $\mu$ l电穿孔反应的细胞密度重新悬浮于SF缓冲液中。将三微克RNA在16孔比色测试条中一式三份地电穿孔到细胞中,并且在室温下温育10分钟。将经电穿孔的细胞回收到包含含有10%胎牛血清的Dulbecco改良Eagle培养基的板中,然后在标准细胞培养条件下温育16-18小时。

[0125] 通过流式细胞术进行复制子转染效率和蛋白质制备的细胞内分析。将经转染的BHK-21细胞固定并使用固定/灌注浓缩物和透化缓冲液 (eBioscience) 透化。将细胞与抗体一起温育,用于与R-藻红蛋白 (Innova Biosciences) 缀合的双链RNA制备 (J2抗-dsRNA IgG2A单克隆抗体, English&Scientific Company)。通过与缀合有PE-Cy5 (Innova Biosciences) 的抗原特异性抗体 (例如对于绿色海肾、红色萤火虫、HA或RSV-F0 (Abeam) 的抗体) 附加温育来评估抗原产生。然后将细胞洗涤一次并使用FACS Aria™融合细胞分选仪 (BD Bioscience) 或FACS Aria™II细胞分选仪 (BD Bioscience) 进行分析。在样品收集之前,运行用单一颜色染色经转染的BHK-21细胞以用于补偿对照。使用FACSDiva (BBD Biosciences) 收集数据,并使用FlowJo软件进一步分析。使用正向和侧面散射图进行初始选通以排除死细胞和碎片。进行进一步的选通以鉴定对于dsRNA (R-PE-阳性) 和蛋白质表达两者均是阳性的细胞群 (对于GFP表达而言,为PE-Cy5阳性或FITC阳性)。收集频率和平均荧光强度并用于构建体比较和优化。

#### [0126] 实施例2

##### [0127] 5'-UTR序列修饰

[0128] 该实施例描述了实验结果,其中修饰5'-UTR序列以增强在VEEV复制子RNA载体中编码的感兴趣的基因的表达。在包含VEEV复制子载体的cDNA拷贝的质粒上进行定点诱变

(SDM)。SDM引物被设计用于将5'UTR的位置2处的胸腺嘧啶(T)残基改变为G残基。

[0129] 在这些实验中,将各自表达rFF荧光素酶基因、A越南型1203流感血凝素基因(HA)或绿色荧光蛋白(GFP)基因的VEEV复制子用作模板以进行定点诱变。将野生型VEEV(菌株TC-83)序列的5'UTR中的位置2处的核苷酸T改变成G。引物“VEE 5' T->G nt 2F”和“VEE 5' T->G nt 2R”用于经由来自Agilent的定点诱变试剂盒诱导核苷酸2变化。引物通过Agilent网站设计。鉴定出阳性克隆体,并且表达rFF、HA和GFP的代表性克隆体经序列确认是完全正确的。

[0130] 用于制备 $\alpha$ -R-T2G复制子的SDM引物如下。

[0131] 正向引物:VEE 5' T->G nt 2F (SEQ ID NO:23):  
cgactcactatagaGaggcggcgcatgag。反向引物:VEE 5' T->G nt 2R (SEQ ID NO:24):  
ctcatgcgcgcctCtctatagttagtg。

[0132] 在对VEEV复制子cDNA中的T2->G变化进行序列确认后,使用线性化质粒DNA上的T7 RNA聚合酶通过体外转录产生RNA。纯化体外转录的RNA并将其用于经电穿孔的BHK-21细胞。通过FACS分别使用抗-dsRNA特异性抗体和GOI特异性抗体监测复制和GOI表达。将来自U2→GVEEV复制子的复制效率和GOI表达直接与表达相同GOI的野生型VEEV复制子进行比较。包含红色萤火虫报告基因的 $\alpha$ -R-rFF-T2G复制子的序列作为序列表中的SEQ ID NO:19提供,其具有T7启动子和带40个A残基的polyA尾。还示出了T7启动子序列之后在位置2处的突变核苷酸。

[0133] 实施例3

[0134] 5'UTR中的U2->G置换不影响经修饰的VEEV-HA复制子的生物活性

[0135] 该实施例描述了评估经修饰的甲病毒复制子的5'UTR中的U2->G置换对A越南型1203流感血凝集素基因(HA)报告基因的表达的影响的实验结果。

[0136] 为了证明在5'UTR中包含U2->G变化的甲病毒复制子确实能够表达蛋白质,由携带5'UTR中的U2->G变化的载体(U2->G VEEV-HA)体外转录复制子RNA。作为表达的阳性对照,由携带野生型5'UTR的载体(WT VEEV-HA)体外转录RNA。用3 $\mu$ g野生型VEEV-HA RNA的U2->GVEEV-HA RNA电穿孔幼仓鼠肾(BHK-21)细胞。图5中示出了用于该比较的流式细胞术分析的示例。利用HA特异性抗体通过FACS分析细胞,以证明经表达的HA蛋白的存在和基于每个细胞的经表达HA的相对量(平均荧光强度-MFI)。相对于从具有野生型5'UTR的复制子检测到的活性,不存在来自携带U2->G 5'UTR变化的复制子RNA的生物活性的丧失。

[0137] 该实施例中呈现的实验数据指示,U2->G VEEV-HA复制子不仅能够表达HA蛋白,而且表达水平等同于野生型VEEV-HA复制子的表达水平。考虑到携带相同U2->G 5'UTR变化的VEEV/SINV感染性克隆体具有生物活性的仅三个对数损失,该结果是意料不到的。在此示出,携带U2->G 5'UTR变化的复制子生物活性完全没有降低。

[0138] 实施例4

[0139] 5'UTR中的U2->G置换使GFP报道基因的表达增强3倍

[0140] 该实施例描述了评估经修饰的甲病毒复制子的5'UTR中的U2->G置换对绿荧光蛋白(GFP)报告基因的表达的影响的实验结果。

[0141] 为了证明来自5'UTR中包含U2->G变化的甲病毒复制子的表达不限于任何特定的GOI,以相似的方式比较表达GFP基因的复制子载体。复制子RNA由携带5'UTR中的U2->G变化

的载体 (U2->G VEEV-GFP) 体外转录。作为表达的阳性对照,由携带野生型5' UTR的载体 (WT VEEV-GFP) 体外转录RNA。用3 $\mu$ g的U2->G VEEV-GFP RNA或野生型VEEV-GFP RNA电穿孔BHK细胞。图6中示出了用于该比较的流式细胞术分析的示例。通过FACS分析细胞的GFP表达,以证明经表达的GFP蛋白的存在和基于每个细胞的经表达GFP的相对量(平均荧光强度-MFI)。不仅不存在来自携带U2->G 5' UTR的复制子RNA的生物活性的损失,而且相对于从具有野生型5' UTR的复制子检测到的表达,U2->G 5' UTR变化实际上使表达增强了3倍。

[0142] 该实施例中呈现的实验数据再次表明携带U2->G 5' UTR变化的复制子载体能够表达蛋白质(这次是GFP)。可能比简单的蛋白质表达甚至更出乎意料的是U2->G VEEV-GFP复制子比野生型VEEV-GFP复制子表达多三倍的GFP。再次,由于5' UTR中的U2->G变化而预料的生物活性的预期损失未实现,并且该结果证明5' UTR中的U2->G变化实际上可显著增强GOI复制子表达。

#### [0143] 实施例5

##### [0144] 5' UTR中的U2->G置换使rFF报告基因的表达增强2倍

[0145] 该实施例描述了评估经修饰的甲病毒复制子的5' UTR中的U2->G置换对红萤火虫 (FF) 报告基因的表达的影响的实验结果。

[0146] 在这些实验中,以相似的方式比较来自包含5' UTR中的U2->G变化的甲病毒复制子,表达rFF基因的复制子载体的表达的另一示例。复制子RNA由携带5' UTR中的U2->G变化的载体 (U2->G VEEV-rFF) 体外转录。作为表达的阳性对照,由携带野生型5' UTR的载体 (WT VEEV-rFF) 体外转录RNA。用3 $\mu$ g的U2->G VEEV-rFF RNA或野生型VEEV-rFF RNA电穿孔BHK细胞。图7示出了荧光素酶蛋白表达的示例。在该实验中,将经修饰以携带5' UTR中的U2->G变化的甲病毒复制子 (U2->GarFF) 的能力与具有野生型5' UTR的甲病毒复制子 ( $\alpha$ rFF) 进行比较。用等同量的来自任一复制子的体外转录RNA对BHK细胞进行电穿孔,然后分析细胞的rFF荧光素酶表达。提供了基于每个细胞表达的荧光素酶的量(表示为相对光单位 (RLU))。不仅不存在来自携带U2->G 5' UTR的复制子RNA的生物活性的损失,而且相对于从具有野生型5' UTR的复制子检测到的表达,U2->G 5' UTR变化实际上使表达增强了约2倍。

[0147] 该实施例中呈现的实验数据再次表明携带U2->G 5' UTR变化的复制子载体能够表达蛋白质(这次是rFF)。可能比简单的蛋白质表达甚至更出乎意料的是U2->G VEEV-rFF复制子比野生型VEEV-rFF复制子表达多~两倍的rFF。再次,由于5' UTR中的U2->G变化而预料的生物活性的预期损失未实现,并且该结果证明5' UTR中的U2->G变化实际上可显著增强GOI复制子表达。

#### [0148] 实施例6

##### [0149] 多价VEEV复制子设计

[0150] 本实施例描述了为构建和评估多价VEEV复制子而进行的实验,其随后被部署用于在重组细胞中表达至少两种不同的多肽。在一些实验中,多价VEEV复制子以5'至3'的顺序包括 (i) 非结构化蛋白质介导的扩增所需的5' 序列, (ii) 编码VEEV非结构化蛋白质nsP1、nsP2、nsP3和nsP4的核苷酸序列, (iii) 至少两种启动子,其各自均可操作地连接至异源核酸序列,其中该异源核酸序列置换VEEV结构化蛋白质基因中的一个或全部, (iv) 非结构化蛋白质介导的扩增所需的3' 序列,以及 (v) 聚腺苷酸核束。

[0151] 虽然已经公开了本公开的特定另选的替代方案,但应当理解,各种修改和组合是

可能的并且在所附权利要求的真正实质和范围内设想到。因此,不旨在对本文所提出的确切摘要和公开内容进行限制。

[0001]	序列表
[0002]	<110> 合成基因组公司
[0003]	<120> 重组病毒复制子体系及其用途
[0004]	<130> SGI.011W0
[0005]	<150> 62/409,228
[0006]	<151> 2016-10-17
[0007]	<160> 24
[0008]	<170> PatentIn version 3.5
[0009]	<210> 1
[0010]	<211> 8364
[0011]	<212> DNA
[0012]	<213> 人工序列(Artificial Sequence)
[0013]	<220>
[0014]	<223> 合成多核苷酸
[0015]	<220>
[0016]	<221> misc_feature
[0017]	<223> Rep- $\alpha$ 复制子
[0018]	<220>
[0019]	<221> misc_feature
[0020]	<223> T7启动子,修饰的VEEV 5'UTR和非结构基因
[0021]	<400> 1
[0022]	taatacgact cactatagag aggcggcgca tgagagaagc ccagaccaat tacctaccca 60
[0023]	aaatggagaa agttcacgtt gacatcgagg aagacagccc attcctcaga gctttgcagc 120
[0024]	ggagcttccc gcagtttgag gtagaagcca agcaggtcac tgataatgac catgctaattg 180
[0025]	ccagagcgtt ttcgcactctg gcttcaaaac tgatcgaaac ggaggtggac ccatccgaca 240
[0026]	cgatccttga cattggaagt gcgccccgcc gcagaatgta ttctaagcac aagtatcatt 300
[0027]	gtatctgtcc gatgagatgt gcggaagatc cggacagatt gtataagtat gcaactaagc 360
[0028]	tgaagaaaaa ctgtaaggaa ataactgata aggaattgga caagaaaatg aaggagctcg 420
[0029]	ccgccgtcat gagecgacct gacctggaaa ctgagactat gtgcctccac gacgacgagt 480
[0030]	cgtgtcgcta cgaagggcaa gtcgctgttt accaggatgt atacgcggtt gacggaccga 540
[0031]	caagtctcta tcaccaagcc aataaggag ttagagtcgc ctactggata ggctttgaca 600
[0032]	ccaccctttt tatgtttaag aacttggtcg gagcatatcc atcatactct accaactggg 660
[0033]	ccgacgaaac cgtgttaacg gtcgtaaca taggcctatg cagctctgac gttatggagc 720
[0034]	ggtcacgtag agggatgtcc attcttagaa agaagtattt gaaaccatcc aacaatgttc 780
[0035]	tattctctgt tggtcgacc atctaccacg agaagaggga cttactgagg agctggcacc 840
[0036]	tgccgtctgt atttactta cgtggcaagc aaaattacac atgtcgggtg gagactatag 900
[0037]	ttagttgcga cgggtacgtc gttaaaagaa tagctatcag tccaggcctg tatgggaagc 960
[0038]	cttcaggcta tgctgtacg atgcaccgag agggattctt gtgctgcaaa gtgacagaca 1020
[0039]	cattgaacgg ggagagggtc tcttttcccg tgtgcacgta tgtgccagct acattgtgtg 1080
[0040]	accaaattgac tggcatactg gcaacagatg tcagtgcgga cgacgcgcaa aaactgctgg 1140
[0041]	ttgggctcaa ccagcgtata gtcgtcaacg gtcgcacca gagaaacacc aataccatga 1200

[0042]	aaaattacct	tttgcccgt	gtggcccagg	catttgctag	gtgggcaaag	gaatataagg	1260
[0043]	aagatcaaga	agatgaaagg	ccactaggac	tacgagatag	acagtttagt	atgggggtgt	1320
[0044]	gttgggcttt	tagaaggcac	aagataacat	ctattttata	gcgcccggat	acccaaacca	1380
[0045]	tcatcaaagt	gaacagcgat	ttccactcat	tcgtgctgcc	caggataggc	agtaacacat	1440
[0046]	tggagatcgg	gctgagaaca	agaatcagga	aaatgttaga	ggagcacaag	gagccgtcac	1500
[0047]	ctctcattac	cgccgaggac	gtacaagaag	ctaagtgcgc	agccgatgag	gctaaggagg	1560
[0048]	tgcgtgaagc	cgaggagtgt	cgcgacgtc	taccaccttt	ggcagctgat	gttgaggagc	1620
[0049]	ccactctgga	agccgatgtc	gacttgatgt	tacaagaggc	tggggccggc	tcagtggaga	1680
[0050]	cacctcgtgg	cttgataaag	gttaccagct	acgatggcga	ggacaagatc	ggctcttacg	1740
[0051]	ctgtgctttc	tccgcaggct	gtactcaaga	gtgaaaaatt	atcttgcatc	caccctctcg	1800
[0052]	ctgaacaagt	catagtata	acacactctg	gccgaaaagg	gcgttatgcc	gtggaacat	1860
[0053]	accatggtaa	agtagtggtg	ccagagggac	atgcaatacc	cgtccaggac	tttcaagctc	1920
[0054]	tgagtgaag	tgccaccatt	gtgtacaacg	aacgtgagtt	cgtaaacagg	tacctgcacc	1980
[0055]	atattgccac	acatggagga	gcgctgaaca	ctgatgaaga	atattacaaa	actgtcaagc	2040
[0056]	ccagcgagca	cgacggcgaa	tacctgtacg	acatcgacag	gaaacagtgc	gtcaagaaag	2100
[0057]	aactagtcac	tgggctaggg	ctcacaggcg	agctgggtga	tcctcccttc	catgaattcg	2160
[0058]	cctacgagag	tctgagaaca	cgaccagccg	ctccttacca	agtaccaacc	ataggggtgt	2220
[0059]	atggcgtgcc	aggatcaggc	aagtctggca	tcattaaaag	cgcagtcacc	aaaaaagatc	2280
[0060]	tagtggtgag	cgccaagaaa	gaaaactgtg	cagaaattat	aagggacgtc	aagaaaatga	2340
[0061]	aagggtgga	cgtcaatgcc	agaactgtgg	actcagtgtc	cttgaatgga	tgcaaacacc	2400
[0062]	ccgtagagac	cctgtatatt	gacgaagctt	ttgcttgta	tgcaggtact	ctcagagcgc	2460
[0063]	tcatagccat	tataagacct	aaaaaggcag	tgctctgcgg	ggatcccaaa	cagtgcggtt	2520
[0064]	tttttaacat	gatgtgcctg	aaagtgcatt	ttaaccacga	gatttgacac	caagtcttcc	2580
[0065]	acaaaagcat	ctctcgccgt	tgcactaaat	ctgtgacttc	ggtcgtctca	accttgtttt	2640
[0066]	acgacaaaaa	aatgagaacg	acgaatccga	aagagactaa	gattgtgatt	gacactaccg	2700
[0067]	gcagtaccaa	acctaagcag	gacgatctca	ttctcacttg	tttcagaggg	tgggtgaagc	2760
[0068]	agttgcaaat	agattacaaa	ggcaacgaaa	taatgacggc	agctgcctct	caagggtga	2820
[0069]	cccgtaaagg	tgtgtatgcc	gttcgggtaca	aggtgaatga	aaatcctctg	tacgcacca	2880
[0070]	cctctgaaca	tgtgaacgtc	ctactgaccc	gcacggagga	ccgcatcgtg	tggaaaacac	2940
[0071]	tagccggcga	cccatggata	aaaacactga	ctgccaaagta	ccctgggaat	ttcactgcca	3000
[0072]	cgatagagga	gtggcaagca	gagcatgatg	ccatcatgag	gcacatcttg	gagagaccgg	3060
[0073]	accctaccga	cgtcttccag	aataaggcaa	acgtgtgttg	ggccaaggct	ttagtgccgg	3120
[0074]	tgtctgaagac	cgctggcata	gacatgacca	ctgaacaatg	gaacactgtg	gattattttg	3180
[0075]	aaacggacaa	agctcactca	gcagagatag	tattgaacca	actatgcgtg	aggttctttg	3240
[0076]	gactcgatct	ggactccggg	ctattttctg	caccactgt	tccgttatcc	attaggaata	3300
[0077]	atcactggga	taactccccg	tcgcctaaca	tgtacgggct	gaataaagaa	gtgggtccgtc	3360
[0078]	agctctctcg	caggtagcca	caactgcctc	gggcagttgc	cactggaaga	gtctatgaca	3420
[0079]	tgaacactgg	tacactgcgc	aattatgatc	cgcgcataaa	cctagtacct	gtaaacagaa	3480
[0080]	gactgcctca	tgttttagtc	ctccaccata	atgaacaccc	acagagtgc	ttttcttcat	3540
[0081]	tcgtcagcaa	attgaagggc	agaactgtcc	tgggtggtcgg	ggaaaagtgt	tccgtccag	3600
[0082]	gcaaaatggg	tgactgggtg	tcagaccggc	ctgaggctac	cttcagagct	cggctggatt	3660
[0083]	taggcacccc	aggtgatgtg	cccaaataatg	acataatatt	tgtaaatgtg	aggaccccat	3720

[0084]	ataaatacca	tcactatcag	cagtgtgaag	accatgccat	taagcttagc	atgttgacca	3780
[0085]	agaaagcttg	tctgcatctg	aatccccggc	gaacctgtgt	cagcataggt	tatggttacg	3840
[0086]	ctgacagggc	cagcgaaagc	atcattgttg	ctatagcgcg	gcagttcaag	ttttccccgg	3900
[0087]	tatgcaaacc	gaaatcctca	cttgaagaga	cggaagttct	gtttgtattc	attgggtacg	3960
[0088]	atcgcaaggc	ccgtacgcac	aatccttaca	agctttcatc	aaccttgacc	aacatttata	4020
[0089]	caggttccag	actccacgaa	gccggatgtg	cacctcata	tcatgtgggtg	cgaggggata	4080
[0090]	ttgccacggc	caccgaagga	gtgattataa	atgctgctaa	cagcaaagga	caacctggcg	4140
[0091]	gaggggtgtg	cggagcgctg	tataagaaat	tcccgaaaag	cttcgattta	cagccgatcg	4200
[0092]	aagtaggaaa	agcgcgactg	gtcaaagggtg	cagctaaaca	tatcattcat	gccgtaggac	4260
[0093]	caaacttcaa	caaagtttcg	gaggttgaag	gtgacaaaca	gttggcagag	gcttatgagt	4320
[0094]	ccatcgctaa	gattgtcaac	gataacaatt	acaagtcagt	agcgattcca	ctgttgtcca	4380
[0095]	ccggcatctt	ttccgggaac	aaagatcgac	taaccaatc	attgaaccat	ttgctgacag	4440
[0096]	ctttagacac	cactgatgca	gatgtagcca	tatactgcag	ggacaagaaa	tgggaaatga	4500
[0097]	ctctcaagga	agcagtggct	aggagagaag	cagtggagga	gatatgcata	tccgacgact	4560
[0098]	cttcagtgc	agaacctgat	gcagagctgg	tgagggtgca	tccgaagagt	tctttggctg	4620
[0099]	gaaggaaggg	ctacagcaca	agcgatggca	aaactttctc	atatttgga	gggaccaagt	4680
[0100]	ttcaccaggc	ggccaaggat	atagcagaaa	ttaatgccat	gtggcccgtt	gcaacggagg	4740
[0101]	ccaatgagca	ggtatgcatg	tatatcctcg	gagaaagcat	gagcagtatt	aggtcgaaat	4800
[0102]	gccccgtcga	agagtcggaa	gcctccacac	cacctagcac	gctgccttgc	ttgtgcatcc	4860
[0103]	atgccatgac	tccagaaaga	gtacagcgcc	taaaagcctc	acgtccagaa	caaattactg	4920
[0104]	tgtgtcatc	ctttccattg	ccgaagtata	gaatcactgg	tgtgcagaag	atccaatgct	4980
[0105]	cccagcctat	attgtttctca	ccgaaagtgc	ctgcgtatat	tcatccaagg	aagtatctcg	5040
[0106]	tggaaacacc	accggtagac	gagactccgg	agccatcggc	agagaaccaa	tccacagagg	5100
[0107]	ggacacctga	acaaccacca	cttataaccg	aggatgagac	caggactaga	acgcctgagc	5160
[0108]	cgatcatcat	cgaagaggaa	gaagaggata	gcataagttt	gctgtcagat	ggccccgacc	5220
[0109]	accaggtgct	gcaagtcgag	gcagacattc	acgggcccgc	ctctgtatct	agctcatcct	5280
[0110]	ggtccattcc	tcatgcatcc	gactttgatg	tggacagttt	atccatactt	gacaccctgg	5340
[0111]	agggagctag	cgtgaccagc	ggggcaacgt	cagccgagac	taactcttac	ttcgcaaaga	5400
[0112]	gtatggagtt	tctggcgcg	ccggtgcctg	cgcctcgaa	agtattcagg	aacctccac	5460
[0113]	atcccgtctc	gcgcacaaga	acaccgtcac	ttgcaccag	cagggcctgc	tcgagaacca	5520
[0114]	gcctagtttc	caccccgcca	ggcgtgaata	gggtgatcac	tagagaggag	ctcgaggcgc	5580
[0115]	ttaccccgtc	acgcactcct	agcaggtcgg	tctcgagaac	cagcctggtc	tccaaccgc	5640
[0116]	caggcgtaaa	tagggtgatt	acaagagagg	agtttgaggc	gttcgtagca	caacaacaat	5700
[0117]	gacggtttga	tgcgggtgca	tacatctttt	cctccgacac	cgggtcaagg	catttacaac	5760
[0118]	aaaaatcagt	aaggcaaacg	gtgctatccg	aagtgggtgt	ggagaggacc	gaattggaga	5820
[0119]	tttctgatgc	cccgccctc	gaccaagaaa	agaagaatt	actacgcaag	aaattacagt	5880
[0120]	taaatccac	acctgctaac	agaagcagat	accagtccag	gaaggtggag	aacatgaaag	5940
[0121]	ccataacagc	tagacgtatt	ctgcaaggcc	tagggcatta	tttgaaggca	gaagggaaa	6000
[0122]	tggagtgc	ccgaacctg	catcctgttc	ctttgtattc	atctagtgtg	aacctgcct	6060
[0123]	tttcaagccc	caaggtcgca	gtggaagcct	gtaacgcat	gttgaaagag	aactttccga	6120
[0124]	ctgtggcttc	ttactgtatt	attccagagt	acgatgccta	tttgacatg	gttgacggag	6180
[0125]	cttcatgctg	cttagacact	gccagttttt	gcctgcaaa	gctgcgcagc	tttccaaaga	6240

[0126]	aacactccta tttggaaccc acaatacgat cggcagtgcc ttcagcgatc cagaacacgc 6300
[0127]	tccagaacgt cctggcagct gccacaaaaa gaaattgcaa tgtcacgcaa atgagagaat 6360
[0128]	tgcccgattt ggattcggcg gcctttaatg tggaaatgctt caagaaatat gcgtgtaata 6420
[0129]	atgaatattg ggaaacgttt aaagaaaacc ccatcaggct tactgaagaa aacgtggtaa 6480
[0130]	attacattac caaattaaaa ggacccaaaag ctgctgctct ttttgcaag acacataatt 6540
[0131]	tgaatatgtt gcaggacata ccaatggaca ggtttgtaat ggacttaaag agagacgtga 6600
[0132]	aagtgactcc aggaacaaaa catactgaag aacggcccaa ggtacagggtg atccaggctg 6660
[0133]	ccgatccgct agcaacagcg tatctgtgcg gaatccaccg agagctggtt aggagattaa 6720
[0134]	atgcggtcct gcttccgaac attcatacac tgtttgatat gtcggctgaa gactttgacg 6780
[0135]	ctattatagc cgagcacttc cagcctgggg attgtgttct ggaaactgac atcgcgtcgt 6840
[0136]	ttgataaaag tgaggacgac gccatggctc tgaccgcgtt aatgattctg gaagacttag 6900
[0137]	gtgtggacgc agagctgttg acgtgattg aggcggcttt cggcgaaatt tcatcaatac 6960
[0138]	atttgccac taaaactaaa tttaaattcg gagccatgat gaaatctgga atgttctca 7020
[0139]	cactgtttgt gaacacagtc attaacattg taatcgcaag cagagtgtg agagaacggc 7080
[0140]	taaccggatc accatgtgca gcattcattg gagatgacaa tatcgtgaaa ggagtcaaat 7140
[0141]	cggacaaatt aatggcagac aggtgcgcca cctggttgaa tatggaagtc aagattatag 7200
[0142]	atgctgtggt gggcgagaaa gcgccttatt tctgtggagg gtttattttg tgtgactccg 7260
[0143]	tgaccggcac agcgtgccgt gtggcagacc ccctaaaaag gctgtttaag cttggcaaac 7320
[0144]	ctctggcagc agacgatgaa catgatgatg acaggagaag ggcattgcat gaagagtcaa 7380
[0145]	cacgttgcaa ccgagtgggt attctttcag agctgtgcaa ggcagtagaa tcaaggatatg 7440
[0146]	aaaccgtagg aacttccatc atagtattgg ccatgactac tctagctagc agtgttaaat 7500
[0147]	cattcagcta cctgagaggg gccctataa ctctctacgg ctaacctgaa tggactacga 7560
[0148]	catagtctag tccgccaaga tatcatgat acagcagcaa ttggcaagct gcttacatag 7620
[0149]	aaggcgcgcc gtttaaacgg ccggccttaa ttaagtaacg atacagcagc aattggcaag 7680
[0150]	ctgcttacat agaactcgcg gcgattggca tgccgcttta aaatTTTTAT tttatttttc 7740
[0151]	TTTTCTTTC cgaatcgat tttgttttta atatttcaaa aaaaaaaaaa aaaaaaaaaa 7800
[0152]	aaaaaaaaa aaaaaaaccc ctctctaaac ggaggggttt ttttcagcgt aactggactg 7860
[0153]	gccacagtta ggcgccgcg catgttcac atcagtaacc cgtatcgtga gcacacctc 7920
[0154]	tcgtttcatc ggtatcatta cctccatgaa cagaaatccc cttacacgg aggcacagct 7980
[0155]	gaccaaacag gaaaaaacg cccttaacat ggcccgttt atcagaagcc agacattaac 8040
[0156]	gcttctggag aaactcaacg agctggacgc ggatgaacag gcagacatct gtgaatcgct 8100
[0157]	tcacgaccac gctgatgagc tttaccgcag ctgcctcgcg cgtttcggtg atgacggtga 8160
[0158]	aaacctctga cacatgcagc tcccggagac ggtcacagct tgtctgtaag cggatgccgg 8220
[0159]	gagcagacaa gcccgtcagg gcgcgtcagc ggggtgttggc ggggtgtcggg gcgcagccat 8280
[0160]	gaccagtcga cgtagcgata gcggagtgt tactggctta actatgcgc atcagagcag 8340
[0161]	attgtactga gagtgcacca tatg 8364
[0162]	<210> 2
[0163]	<211> 50
[0164]	<212> DNA
[0165]	<213> 奥拉病毒(Aura virus)
[0166]	<220>
[0167]	<221> misc_feature

[0168] <223> 5'UTR  
[0169] <400> 2  
[0170] atagcggacg gactagtact tgtactacag aattaactgc cgtgtgccgc 50  
[0171] <210> 3  
[0172] <211> 50  
[0173] <212> DNA  
[0174] <213> 奇昆古尼亚病毒病(Chikungunya virus)  
[0175] <220>  
[0176] <221> misc\_feature  
[0177] <223> 5'UTR  
[0178] <400> 3  
[0179] atggctgctg gagacacacg tagcctacca gtttcttact gctctactct 50  
[0180] <210> 4  
[0181] <211> 50  
[0182] <212> DNA  
[0183] <213> 阿尼昂-尼昂病毒(O'Nyong-Nyong virus)  
[0184] <220>  
[0185] <221> misc\_feature  
[0186] <223> 5'UTR  
[0187] <400> 4  
[0188] atagctgctg gatacacaca cgcagcttac gggtttcata ctgctctact 50  
[0189] <210> 5  
[0190] <211> 50  
[0191] <212> DNA  
[0192] <213> 贝巴鲁病毒(Bebaru virus)  
[0193] <220>  
[0194] <221> misc\_feature  
[0195] <223> 5'UTR  
[0196] <400> 5  
[0197] atggcgctg tgtgacacac gagccgtcga tttcaacctt cttgctccct 50  
[0198] <210> 6  
[0199] <211> 50  
[0200] <212> DNA  
[0201] <213> 塞姆利基森林病毒(Semliki Forest virus)  
[0202] <220>  
[0203] <221> misc\_feature  
[0204] <223> 5'UTR  
[0205] <400> 6  
[0206] atggcgcatg tgtgacatac acgacgcaa aagattttgt tccagctcct 50  
[0207] <210> 7  
[0208] <211> 50  
[0209] <212> DNA

[0210] <213> 马雅罗病毒 (Mayaro virus)  
[0211] <220>  
[0212] <221> misc\_feature  
[0213] <223> 5'UTR  
[0214] <400> 7  
[0215] atggcgggca agtgacactt gttccgccgg tcgtctctaa gctcttcctc 50  
[0216] <210> 8  
[0217] <211> 50  
[0218] <212> DNA  
[0219] <213> 盖塔病毒 (Getah virus)  
[0220] <220>  
[0221] <221> misc\_feature  
[0222] <223> 5'UTR  
[0223] <400> 8  
[0224] atggcggacg tgtgacatca ccgttcgctc tttctaggat cctttgctac 50  
[0225] <210> 9  
[0226] <211> 50  
[0227] <212> DNA  
[0228] <213> 鹭山病毒 (Sagiyama virus)  
[0229] <220>  
[0230] <221> misc\_feature  
[0231] <223> 5'UTR  
[0232] <400> 9  
[0233] atggcggacg tgtgacatca ccgttcgctc tttctaggat cctttgctac 50  
[0234] <210> 10  
[0235] <211> 50  
[0236] <212> DNA  
[0237] <213> 恩杜穆病毒 (Ndumu virus)  
[0238] <220>  
[0239] <221> misc\_feature  
[0240] <223> 5'UTR  
[0241] <400> 10  
[0242] atggtgcgga gttgagagac gaagcaccaa acaactacgc ggctcaccat 50  
[0243] <210> 11  
[0244] <211> 50  
[0245] <212> DNA  
[0246] <213> 米德堡病毒 (Middleburg virus)  
[0247] <220>  
[0248] <221> misc\_feature  
[0249] <223> 5'UTR  
[0250] <400> 11  
[0251] attggtggtt acgtacacgt gccaccaccc cccaccctcc aagcgatcca 50

---

[0252]	<210> 12
[0253]	<211> 50
[0254]	<212> DNA
[0255]	<213> 东部马脑炎病毒(Eastern equine encephalitis virus)
[0256]	<220>
[0257]	<221> misc_feature
[0258]	<223> 5'UTR
[0259]	<400> 12
[0260]	atagggtacg gtgtagaggc aaccacccta tttccaccta tccaaaatgg 50
[0261]	<210> 13
[0262]	<211> 50
[0263]	<212> DNA
[0264]	<213> 摩根堡病毒(Fort Morgan virus)
[0265]	<220>
[0266]	<221> misc_feature
[0267]	<223> 5'UTR
[0268]	<400> 13
[0269]	atagggtatg gtttagaggc gcctacccta cttaaccgat ccaaaccatgg 50
[0270]	<210> 14
[0271]	<211> 50
[0272]	<212> DNA
[0273]	<213> 布吉溪病毒(Buggy Creek virus)
[0274]	<220>
[0275]	<221> misc_feature
[0276]	<223> 5'UTR
[0277]	<400> 14
[0278]	atagggtatg gtttagaggc gcctacccta cttaaccgat ccaaaccatgg 50
[0279]	<210> 15
[0280]	<211> 50
[0281]	<212> DNA
[0282]	<213> 委内瑞拉马脑炎病毒(Venezuelan equine encephalitis virus)
[0283]	<220>
[0284]	<221> misc_feature
[0285]	<223> 5'UTR
[0286]	<400> 15
[0287]	atgggcgcgcg caagagagaa gcccaaacca attacctacc caaatggag 50
[0288]	<210> 16
[0289]	<211> 50
[0290]	<212> DNA
[0291]	<213> 沃达罗河病毒(Whataroa virus)
[0292]	<220>
[0293]	<221> misc_feature

[0294] <223> 5'UTR  
 [0295] <400> 16  
 [0296] attggcggca tagtacatac tatataaaag aaacagccga ccaattgcac 50  
 [0297] <210> 17  
 [0298] <211> 50  
 [0299] <212> DNA  
 [0300] <213> 辛德毕斯病毒(Sindbis virus)  
 [0301] <220>  
 [0302] <221> misc\_feature  
 [0303] <223> 5'UTR  
 [0304] <400> 17  
 [0305] attgacggcg tagtacacac tattgaatca aacagccgac caattgcact 50  
 [0306] <210> 18  
 [0307] <211> 50  
 [0308] <212> DNA  
 [0309] <213> 比巴鲁病毒(Bebaru virus)  
 [0310] <220>  
 [0311] <221> misc\_feature  
 [0312] <223> 5'UTR  
 [0313] <400> 18  
 [0314] attggcggcg tagtacacac tattgaatca aacagccgac caattgcact 50  
 [0315] <210> 19  
 [0316] <211> 9476  
 [0317] <212> DNA  
 [0318] <213> 人工序列(Artificial Sequence)  
 [0319] <220>  
 [0320] <223>  $\alpha$ -R-rFF-T2G  
 [0321] <220>  
 [0322] <221> misc\_feature  
 [0323] <222> (1) .. (18)  
 [0324] <223> T7启动子  
 [0325] <220>  
 [0326] <221> misc\_feature  
 [0327] <222> (20) .. (20)  
 [0328] <223> T到G置换  
 [0329] <220>  
 [0330] <221> misc\_feature  
 [0331] <222> (9437) .. (9476)  
 [0332] <223> Poly A  
 [0333] <400> 19  
 [0334] taatacgact cactatagag aggcggcgca tgagagaagc ccagaccaat tacctaccca 60  
 [0335] aatggagaa agttcacgtt gacatcgagg aagacagccc attcctcaga gctttgcage 120

[0336]	ggagcttccc gcagtttgag gtagaagcca agcaggtcac tgataatgac catgctaattg	180
[0337]	ccagagcgtt ttgcgcatctg gcttcaaaac tgatcgaaac ggaggtggac ccatccgaca	240
[0338]	cgatccttga cattggaagt gcgcccgcgc gcagaatgta ttctaagcac aagtatcatt	300
[0339]	gtatctgtcc gatgagatgt gcggaagatc cggacagatt gtataagtat gcaactaagc	360
[0340]	tgaagaaaaa ctgtaaggaa ataactgata aggaattgga caagaaaatg aaggagctcg	420
[0341]	ccgccgtcat gagcgaccct gacctggaaa ctgagactat gtgcctccac gacgacgagt	480
[0342]	cgtgtcgcta cgaagggcaa gtctgtgttt accaggatgt atacgcggtt gacggaccga	540
[0343]	caagtctcta tcaccaagcc aataaggag ttagagtcgc ctactggata ggctttgaca	600
[0344]	ccaccccttt tatgtttaag aacttggtg gagcatalcc atcatactct accaactggg	660
[0345]	ccgacgaaac cgtgttaacg gctcgttaaca taggcctatg cagctctgac gttatggagc	720
[0346]	ggtcacgtag agggatgtcc attcttagaa agaagtattt gaaaccatcc aacaatgttc	780
[0347]	tattctctgt tggctcgacc atctaccacg agaagaggga cttactgagg agctggcacc	840
[0348]	tgccgtctgt atttcaacta cgtggcaagc aaaattacac atgtcgggtg gagactatag	900
[0349]	ttagttgcga cgggtacgtc gttaaaagaa tagctatcag tccaggcctg tatgggaagc	960
[0350]	cttcaggcta tgctgtacg atgcaccgcg agggattctt gtgctgcaa gtgacagaca	1020
[0351]	cattgaacgg ggagagggtc tcttttccc tgtgcacgta tgtgccagct acattgtgtg	1080
[0352]	accaaagac tggcatactg gcaacagatg tcagtgcgga cgacgcgcaa aaactgctgg	1140
[0353]	ttgggctcaa ccagcgata gtctgcaacg gtgcaccca gagaaacacc aataccatga	1200
[0354]	aaaattacct ttgcccgtg gtggcccagg catttgctag gtgggcaaag gaataaagg	1260
[0355]	aagatcaaga agatgaaagg ccactaggac tacgagatag acagttagtc atgggggtgt	1320
[0356]	gttgggcttt tagaaggcac aagataacat ctatttataa gcgcccggat acccaaacca	1380
[0357]	tcatcaaagt gaacagcgat ttccactcat tcgtgctgcc caggataggc agtaacacat	1440
[0358]	tggagatcgg gctgagaaca agaatacagga aaatgttaga ggagcacaag gagccgtcac	1500
[0359]	ctctcattac cgccgaggac gtacaagaag ctaagtgcgc agccgatgag gctaaggagg	1560
[0360]	tgcgtgaagc cgaggagtgt cgcgcagctc taccacctt ggcagctgat gttgaggagc	1620
[0361]	ccactctgga agccgatgtc gacttgatgt tacaagaggc tggggccggc tcagtggaga	1680
[0362]	cacctcgtgg cttgataaag gttaccagct acgatggcga ggacaagatc ggctcttacg	1740
[0363]	ctgtgctttc tccgcaggct gtactcaaga gtgaaaaatt atcttgcatc caccctctcg	1800
[0364]	ctgaacaagt catagtgata acacactctg gccgaaaagg gcgttatgcc gtggaacct	1860
[0365]	accatggtaa agtagtggtg ccagagggac atgcaatacc cgtccaggac tttcaagctc	1920
[0366]	tgagtgaag tgccaccatt gtgtacaacg aacgtgagtt cgtaaacagg tacctgcacc	1980
[0367]	atattgccac acatggagga gcgctgaaca ctgatgaaga atattacaaa actgtcaagc	2040
[0368]	ccagcgagca cgacggcgaa tacctgtacg acatcgacag gaaacagtgc gtcaagaaag	2100
[0369]	aactagtcac tgggctaggg ctcacaggcg agctgggtga tcttccttc catgaattcg	2160
[0370]	cctacgagag tctgagaaca cgaccagccg ctccttacca agtaccaacc ataggggtgt	2220
[0371]	atggcgtgcc aggatcaggc aagtctggca tcattaaaag cgcagtcacc aaaaaagatc	2280
[0372]	tagtggtgag cgccaagaaa gaaaactgtg cagaaattat aaggacgtc aagaaaatga	2340
[0373]	aagggtgga cgtcaatgcc agaactgtgg actcagtgtc cttgaatgga tgcaaacc	2400
[0374]	ccgtagagac cctgtatatt gacgaagctt ttgctgtca tgcaggtact ctcagagcgc	2460
[0375]	tcatagccat tataagacct aaaaaggcag tgctctgcgg ggatcccaa cagtgcggtt	2520
[0376]	ttttaacat gatgtgcctg aaagtgcatt ttaaccacga gatttgcaca caagtcttc	2580
[0377]	acaaaagcat ctctgccgt tgcactaaat ctgtgacttc ggtcgtctca acctgtttt	2640

[0378]	acgacaaaaa aatgagaacg acgaatccga aagagactaa gattgtgatt gacactaccg	2700
[0379]	gcagtaccaa acctaagcag gacgatctca ttctcacttg tttcagaggg tgggtgaagc	2760
[0380]	agttgcaaat agattacaaa ggcaacgaaa taatgacggc agctgcctct caagggtga	2820
[0381]	cccgtaaagg tgtgtatgcc gttcggtaca aggtgaatga aaatcctctg tacgcacca	2880
[0382]	cctctgaaca tgtgaacgtc ctactgaccc gcacggagga ccgcatcgtg tggaaaacac	2940
[0383]	tagccggcga cccatggata aaaacactga ctgccaaagta ccctgggaat ttcactgcc	3000
[0384]	cgatagagga gtggcaagca gagcatgatg ccatcatgag gcacatcttg gagagaccgg	3060
[0385]	accctaccga cgtcttccag aataaggcaa acgtgtgttg ggccaaggct ttagtgccgg	3120
[0386]	tgctgaagac cgctggcata gacatgacca ctgaacaatg gaacactgtg gattattttg	3180
[0387]	aaacggacaa agctcactca gcagagatag tattgaacca actatgcgtg aggttctttg	3240
[0388]	gactcgatct ggactccggt ctattttctg caccactgt tccgttatcc attaggaata	3300
[0389]	atcactggga taactccccg tcgcctaaca tgtacgggct gaataaagaa gtgggtccgtc	3360
[0390]	agctctctcg caggtacca caactgcctc gggcagttgc cactggaaga gtctatgaca	3420
[0391]	tgaacactgg taaactgcgc aattatgatc cgcgcataaa cctagtacct gtaaacagaa	3480
[0392]	gactgcctca tgcttttagtc ctccaccata atgaacaccc acagagtgc ttttcttcat	3540
[0393]	tcgtcagcaa attgaaggc agaactgtcc tgggtggtcg ggaaaagttg tccgtccag	3600
[0394]	gcaaaatggt tgactggttg tcagaccggc ctgaggctac cttcagagct cggctggatt	3660
[0395]	taggcacccc aggtgatgtg cccaaatag acataatatt tgtaaatgtg aggacccat	3720
[0396]	ataaatacca tcactatcag cagtgtgaag accatgccat taagcttagc atgttgacca	3780
[0397]	agaaagcttg tctgcatctg aatcccgcg gaacctgtgt cagcataggt tatggttacg	3840
[0398]	ctgacagggc cagcgaagc atcattgtg ctatagcgc gcagttcaag ttttccggg	3900
[0399]	tatgcaaacc gaaatcctca cttgaagaga cggaagttct gtttgtattc attgggtacg	3960
[0400]	atcgcaaggc ccgtacgcac aatccttaca agctttcacc aaccttgacc aacatttata	4020
[0401]	caggttccag actccacgaa gccgatgtg caccctcata tcatgtggtg cgaggggata	4080
[0402]	ttgccacggc caccgaagga gtgattataa atgctgctaa cagcaaagga caacctggcg	4140
[0403]	gaggggtgtg cggagcgtg tataagaaat tcccgaaag cttcgattta cagccgatcg	4200
[0404]	aagtaggaaa agcgcgactg gtcaaagggt cagctaaaca tatcattcat gccgtaggac	4260
[0405]	caaacttcaa caaagtttcg gaggttgaag gtgacaaaca gttggcagag gcttatgagt	4320
[0406]	ccatcgctaa gattgtcaac gataacaatt acaagtcagt agcgattcca ctgttgctca	4380
[0407]	ccggcatctt ttccgggaac aaagatcgac taaccaatc attgaaccat ttgctgacag	4440
[0408]	ctttagacac cactgatgca gatgtagcca tatactgcag ggacaagaaa tgggaaatga	4500
[0409]	ctctcaagga agcagtggct aggagagaag cagtggagga gatatgcata tccgacgact	4560
[0410]	cttcagtgc agaacctgat gcagagctgg tgagggtgca tccgaagagt tctttggctg	4620
[0411]	gaaggaaggc ctacagcaca agcgatggca aaactttctc atatttgga gggaccaagt	4680
[0412]	ttcaccaggc ggccaaggat atagcagaaa ttaatgcat gtggcccgtt gcaacggagg	4740
[0413]	ccaatgagca ggtatgcatg tatatcctcg gagaaagcat gagcagtatt aggtcgaaat	4800
[0414]	gccccgtcga agagtcggaa gcctccacac cacctagcac gctgccttgc ttgtgcatcc	4860
[0415]	atgccatgac tccagaaaga gtacagcgcc taaaagcctc acgtccagaa caaattactg	4920
[0416]	tgtgtctatc ctttccattg ccgaagtata gaatcactgg tgtgcagaag atccaatgct	4980
[0417]	cccagcctat attgttctca ccgaaagtgc ctgcgtatat tcatccaagg aagtatctcg	5040
[0418]	tggaaacacc accggtagac gagactccgg agccatcggc agagaaccaa tccacagagg	5100
[0419]	ggacacctga acaaccacca cttataaccg aggatgagac caggactaga acgcctgagc	5160

[0420]	cgatcatcat cgaagaggaa gaagaggata gcataagttt gctgtcagat ggcccgaccc	5220
[0421]	accagggtgct gcaagtcgag gcagacattc acgggccgcc ctctgtatct agctcatcct	5280
[0422]	ggtccattcc tcatgcatcc gactttgatg tggacagttt atccatactt gacaccctgg	5340
[0423]	agggagctag cgtgaccagc ggggcaacgt cagccgagac taactcttac ttcgcaaaga	5400
[0424]	gtatggagtt tctggcgcga ccggtgcctg cgcctcgaac agtattcagg aaccctccac	5460
[0425]	atcccgcctcc gcgcacaaga acaccgtcac ttgcacccag cagggcctgc tcgagaacca	5520
[0426]	gcctagtttc caccgcgcca ggcgtaata gggatgatcac tagagaggag ctcgaggcgc	5580
[0427]	ttaccccgtc acgcactcct agcaggtcgg tctcgagaac cagcctggtc tccaaccgcg	5640
[0428]	caggcgtaaa tagggtgatt acaagagagg agtttgaggc gttcgtagca caacaacaat	5700
[0429]	gacggtttga tgcgggtgca tacatctttt cctccgacac cggtaaggg catttacaac	5760
[0430]	aaaaatcagt aaggcaaacg gtgctatccg aagtgggtgt ggagaggacc gaattggaga	5820
[0431]	tttcgtatgc cccgcgcctc gaccaagaaa aagaagaatt actacgcaag aaattacagt	5880
[0432]	taaatccac acctgctaac agaagcagat accagtccag gaaggtggag aacatgaaag	5940
[0433]	ccataacagc tagacgtatt ctgcaaggcc tagggcatta tttgaaggca gaaggaaaag	6000
[0434]	tggagtgcga ccgaaccctg catcctgttc ctttgtattc atctagtgtg aaccgtgcct	6060
[0435]	tttcaagccc caaggtcgca gtggaagcct gtaacgccat gttgaaagag aactttccga	6120
[0436]	ctgtggcttc ttactgtatt attccagagt acgatgccta tttggacatg gttgacggag	6180
[0437]	cttcatgctg cttagacact gccagttttt gccctgcaa gctgcgcagc tttccaaaag	6240
[0438]	aacactccta tttggaaccc acaatacgat cggcagtgcc ttcagcgatc cagaacacgc	6300
[0439]	tccagaacgt cctggcagct gccacaaaaa gaaattgcaa tgtcacgcaa atgagagaat	6360
[0440]	tgcccgtatt ggattcggcg gcctttaatg tggaatgctt caagaaatat gcgtgtaata	6420
[0441]	atgaatattg ggaaacgttt aaagaaaacc ccatcaggct tactgaagaa aacgtggtaa	6480
[0442]	attacattac caaatataaa ggacaaaaag ctgctgtctt ttttgcaag acacataatt	6540
[0443]	tgaatatgtt gcaggacata ccaatggaca ggtttgtaat ggacttaaag agagacgtga	6600
[0444]	aagtgactcc aggaacaaaa catactgaag aacggcccaa ggtacaggatg atccaggctg	6660
[0445]	ccgatccgct agcaacagcg tatctgtgcg gaatccaccg agagctggtt aggagattaa	6720
[0446]	atgcggctct gcttccgaac attcatacac tgtttgatat gtcggctgaa gactttgacg	6780
[0447]	ctattatagc cgagcacttc cagcctgggg attgtgttct ggaaactgac atcgcgtcgt	6840
[0448]	ttgataaaag tgaggacgac gccatggctc tgaccgcgtt aatgattctg gaagacttag	6900
[0449]	gtgtggacgc agagctgttg acgctgattg aggcggcttt cggcgaaatt tcatcaatac	6960
[0450]	atttgcacac taaaactaaa tttaaattcg gagccatgat gaaatctgga atgttcctca	7020
[0451]	cactgtttgt gaacacagtc attaacattg taatcgcaag cagagtgttg agagaacggc	7080
[0452]	taaccggatc accatgtgca gcattcattg gagatgacaa tatcgtgaaa ggagtcaaat	7140
[0453]	cggacaaatt aatggcagac aggtgcgcca cctggttgaa tatggaagtc aagattatag	7200
[0454]	atgctgtggt gggcgagaaa gcgccttatt tctgtggagg gtttattttg tgtgactccg	7260
[0455]	tgaccggcac agcgtgccgt gtggcagacc ccataaaaag gctgtttaag cttggcaaac	7320
[0456]	ctctggcagc agacgatgaa catgatgatg acaggagaag ggcattgcat gaagagtcaa	7380
[0457]	cacgtcgtaa ccgagtgggt attctttcag agctgtgcaa ggcagtagaa tcaaggtatg	7440
[0458]	aaaccgtagg aacttccatc atagtattgg ccatgactac tctagctagc agtgttaaat	7500
[0459]	cattcagcta cctgagaggg gccctataa ctctctacgg ctaacctgaa tggactacga	7560
[0460]	catagtctag tccgccaaga tatcgacca tggaaaatat ggaaaacgac gagaacatcg	7620
[0461]	tggtgggccc caagcccttc taccatcg aggaaggcag cgccggcacc cagctgcgga	7680

[0462]	agtacatgga aagatacgcc aagctgggcg ccattgcctt caccaacgcc gtgaccggcg	7740
[0463]	tggactacag ctacgccgag tacctggaaa agagctgctg cctgggcaag gctctgcaga	7800
[0464]	actacggcct ggtggtggac ggccggatcg ccctgtgcag cgagaactgc gaggaattct	7860
[0465]	tcatccccgt gatcgccggc ctgttcatcg gcgtgggctg ggctcccacc aacgagatct	7920
[0466]	acacctgcg ggagctggtg cacagcctgg gcatcagcaa gccaccatc gtgttcagca	7980
[0467]	gcaagaaggg cctggacaaa gtcacaccg tgcagaaaac cgtgaccacc atcaagacca	8040
[0468]	tcgtgatacct ggacagcaag gtggactacc ggggctacca gtgcctggac accttcatca	8100
[0469]	agcggaacac cccccctggc ttccaggcca gcagcttcaa gaccgtggag gtggaccgga	8160
[0470]	aagaacaggt ggccctgata atgaacagca gcggcagcac cggcctgccc aagggcgtgc	8220
[0471]	agctgacca cgagaacacc gtgaccgggt tcagccacgc cagggacccc atctacggca	8280
[0472]	accagtgct ccccgccacc gccgtgctga ccgtgggtgc cttccaccac ggcttcggca	8340
[0473]	tgttcaccac cctgggctac ctgatactcg gcttccgggt ggtgatgctg accaagtctg	8400
[0474]	acgaggaaac cttcctgaaa accctgcagg actacaagtg cacctacgtg attctggtgc	8460
[0475]	ccaccctgtt cgccatcctg aacaagagcg agctgctgaa caagtacgac ctgagcaacc	8520
[0476]	tggtggagat cgccagcggc ggagcccccc tgagcaaaga agtgggagag gccgtcgcca	8580
[0477]	ggcggttcaa tctgcccggc gtgcggcagg gctacggcct gaccgagaca accagcgcca	8640
[0478]	tcatcatcac ccccgagggc gacgacaagc ctggagccag cggcaagggt gtgcccctgt	8700
[0479]	tcaaggccaa agtgatcgac ctggacacca agaagagcct gggccccaac agacggggcg	8760
[0480]	aagtgtgcgt gaaggggccc atgctgatga agggctacgt gaacaacccc gaggccacca	8820
[0481]	aagagctgat cgacgaagag ggctggctgc acaccggcg catcggtac tacgacgaag	8880
[0482]	agaagcactt cttcatctg gaccggctga agagcctgat caagtacaag ggctatcagg	8940
[0483]	tgccccctgc cgagctggaa agcgtcctgc tgcagcacc cagcatcttc gacgccggcg	9000
[0484]	tggccggggt gccagatcct gtggccggcg agctgcctgg cgccgtggtg gtgctggaat	9060
[0485]	ccggcaagaa catgaccgag aaagaagtga tggactacgt cgccagccag gtgtccaacg	9120
[0486]	ccaagcggt gagaggcggc gtgagattcg tggacgaagt gccaaagggc ctgaccggca	9180
[0487]	agatcgacgg cagggccatc cgggagatcc tgaagaaacc cgtggccaag atgtgattaa	9240
[0488]	ttgatcgata cagcagcaat tggcaagctg cttacataga aggcgcgccg tttaaaccgc	9300
[0489]	cggccttaat taagtaacga tacagcagca attggcaagc tgcttacata gaactcgcg	9360
[0490]	cgattggcat gccgttttaa aatttttatt ttatttttct tttcttttcc gaatcggtat	9420
[0491]	ttgtttttaa tatttcaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaa	9476
[0492]	<210>	20
[0493]	<211>	1647
[0494]	<212>	DNA
[0495]	<213>	人工序列(Artificial Sequence)
[0496]	<220>	
[0497]	<223>	合成多核苷酸
[0498]	<220>	
[0499]	<221>	misc_feature
[0500]	<223>	红色萤火虫报告基因 - rFF
[0501]	<400>	20
[0502]	atggaaaata tggaacacga cgagaacatc gtggtgggcc ccaagccctt ctaccccatc	60
[0503]	gaggaaggca gcgccggcac ccagctgcgg aagtacatgg aaagatacgc caagctgggc	120

[0504]	gccattgcct tcaccaacgc cgtgaccggc gtggactaca gctacgccga gtacctggaa	180
[0505]	aagagctgct gcctgggcaa ggctctgcag aactacggcc tgggtggtgga cggccggatc	240
[0506]	gccctgtgca gcgagaactg cgaggaattc ttcatccccg tgatcgccgg cctgttcatc	300
[0507]	ggcgtgggcg tggctcccac caacgagatc tacaccctgc gggagctggt gcacagcctg	360
[0508]	ggcatcagca agcccacat cgtgttcagc agcaagaagg gcctggacaa agtcatcacc	420
[0509]	gtgcagaaaa ccgtgaccac catcaagacc atcgtgatcc tggacagcaa ggtggactac	480
[0510]	cggggctacc agtgcttga caccctcatc aagcgggaca cccccctgg cttccaggcc	540
[0511]	agcagcttca agaccgtgga ggtggaccgg aaagaacagg tggccctgat catgaacagc	600
[0512]	agcggcagca ccggcctgcc caagggcgtg cagctgaccc acgagaacac cgtgaccgg	660
[0513]	ttcagccacg ccagggaccc catctacggc aaccaggtgt cccccggcac cgccgtgctg	720
[0514]	accgtggtgc cttccacca cggcttcggc atgttcacca ccctgggcta cctgatctgc	780
[0515]	ggcttccggg tggatgatgt gaccaagtgc gacgaggaaa ctttctgaa aacctgcag	840
[0516]	gactacaagt gcacctacgt gattctggtg cccaccctgt tcgccatcct gaacaagagc	900
[0517]	gagctgctga acaagtacga cctgagcaac ctggtggaga tcgccagcgg cggagcccc	960
[0518]	ctgagcaaaag aagtgggaga ggccgtcgcc aggcggttca atctgcccg cgtgcggcag	1020
[0519]	ggctacggcc tgaccgagac aaccagcgcc atcatcatca ccccgaggg cgacgacaag	1080
[0520]	cctggagcca gcggcaaggt ggtgcccctg ttcaaggcca aagtgatcga cctggacacc	1140
[0521]	aagaagagcc tgggccccaa cagacggggc gaagtgtgcg tgaaggggcc catgctgatg	1200
[0522]	aagggtacg tgaacaaccc cgaggccacc aaagagctga tcgacgaaga gggctggctg	1260
[0523]	cacaccggcg acatcggtc ctacgacgaa gagaagcact ttttcatcgt ggaccggtg	1320
[0524]	aagagcctga tcaagtacaa gggctatcag gtgccccctg ccgagctgga aagcgtcctg	1380
[0525]	ctgcagcacc ccagcatctt cgacgccggc gtggccgggg tgccagatcc tgtggccggc	1440
[0526]	gagctgcctg gcgccgtggt ggtgtctgaa tccggcaaga acatgaccga gaaagaagt	1500
[0527]	atggactacg tcgccagcca ggtgtccaac gccaaagcgc tgagaggcgg cgtgagattc	1560
[0528]	gtggacgaag tgccaaagg cctgaccggc aagatcgacg gcagggccat ccgggagatc	1620
[0529]	ctgaagaaac ccgtggccaa gatgtga	1647
[0530]	<210> 21	
[0531]	<211> 729	
[0532]	<212> DNA	
[0533]	<213> 人工序列(Artificial Sequence)	
[0534]	<220>	
[0535]	<223> 合成多核苷酸	
[0536]	<220>	
[0537]	<221> misc_feature	
[0538]	<223> EGFP	
[0539]	<400> 21	
[0540]	atgggaagag ccggcgtgag caagggcgag gagctgttca ccgggggtggt gcccatcctg	60
[0541]	gtcagctggt acggcgacgt aaacggccac aagttcagcg tgtccggcga gggcgagggc	120
[0542]	gatgccacct acggcaagct gaccctgaag ctgatctgca ccaccggcaa gctgcccgtg	180
[0543]	ccctggccca ccctcgtgac caccctgggc tacggcctgc agtgcttcgc ccgctacccc	240
[0544]	gaccacatga agcagcacga cttcttcaag tccgccatgc ccgaaggcta cgtccaggag	300
[0545]	cgcaccatct tcttcaagga cgacggcaac tacaagaccc gcgccgaggt gaagttcgag	360

[0546]	ggcgacaccc tggatgaaccg catcgagctg aagggcacgc acttcaagga ggacggcaac	420
[0547]	atcctggggc acaagctgga gtacaactac aacagccaca acgtctatat caccgccgac	480
[0548]	aagcagaaga acggcatcaa ggccaacttc aagatccgcc acaacatcga ggacggcggc	540
[0549]	gtgcagctcg ccgaccacta ccagcagaac acccccatcg gcgacggccc cgtgctgctg	600
[0550]	cccgaacacc actacctgag ctaccagtcc gccctgagca aagaccccaa cgagaagcgc	660
[0551]	gatcacatgg tctgctgga gttcgtgacc gccgccggga tcactctcgg catggacgag	720
[0552]	ctgtacaag	729
[0553]	<210>	22
[0554]	<211>	1707
[0555]	<212>	DNA
[0556]	<213>	人工序列(Artificial Sequence)
[0557]	<220>	
[0558]	<223>	合成多核苷酸
[0559]	<220>	
[0560]	<221>	misc_feature
[0561]	<223>	HA甲型流感病毒
[0562]	A-越南型-1203-2004 (H5N1) - GenBank	
[0563]	<400>	22
[0564]	atggagaaaa tagtgcttct ttttgcaata gtcagtcttg ttaaaagtga tcagatttgc	60
[0565]	attggttacc atgcaacaa ctcgacagag caggttgaca caataatgga aaagaacgtt	120
[0566]	actgttacac atgccaaga catactggaa aagaaacaca acgggaagct ctgcgatcta	180
[0567]	gatggagtga agcctctaata tttgagagat tgtagcgtag ctggatggct cctcgaaac	240
[0568]	ccaatgtgtg acgaattcat caatgtccg gaatggtctt acatagtgga gaaggccaat	300
[0569]	ccagtcaatg acctctgtta cccaggggat ttcaatgact atgaagaatt gaaacaccta	360
[0570]	ttgagcagaa taaaccattt tgagaaaatt cagatcatcc ccaaaagttc ttggtccagt	420
[0571]	catgaagcct cattaggggt gagctcagca tgtccatacc agggaaagtc ctcctttttc	480
[0572]	agaaatgtgg tatggcttat caaaaagaac agtacatacc caacaataaa gaggagctac	540
[0573]	aataatacca accaagaaga tcttttggtg ctgtggggga ttcaccatcc taatgatgag	600
[0574]	gcagagcaga caaagctcta tcaaaaccca accacctata ttccggttg gacatcaaca	660
[0575]	ctaaaccaga gattgtgacc aagaatagct actagatcca aagtaaagcg gcaaagtgga	720
[0576]	aggatggagt tcttctggac aattttaaag ccgaatgatg caatcaactt cgagagtaat	780
[0577]	ggaaatttca ttgctccaga atatgcatac aaaattgtca agaaagggga ctcaacaatt	840
[0578]	atgaaaagtg aattggaata tggtaactgc aacaccaagt gtcaaactcc aatgggggcg	900
[0579]	ataaactcta gcatgccatt ccacaatata caccctctca ccattgggga atgccccaaa	960
[0580]	tatgtgaaat caaacagatt agtccttgcg actgggctca gaaatagccc tcaaagagag	1020
[0581]	agaagaagaa aaaagagagg attatttga gctatagcag gttttataga gggaggatgg	1080
[0582]	cagggaatgg tagatggttg gtatgggtac caccatagca atgagcaggg gagtgggtac	1140
[0583]	gtcgagaca aagaatccac tcaaaaggca atagatggag tcaccaataa ggtcaactcg	1200
[0584]	atcattgaca aatgaacac tcagtttgag gccgttgga gggaatttaa caacttagaa	1260
[0585]	aggagaatag agaatttaa caagaagatg gaagacgggt tcctagatgt ctggacttat	1320
[0586]	aatgctgaac ttctggttct catggaaaat gagagaactc tagactttca tgactcaaat	1380
[0587]	gtcaagaacc tttagacaa ggtccgacta cagcttaggg ataatgcaa ggagctgggt	1440

[0588] aacggttggt tcgagttcta tcataaatgt gataatgaat gtatggaaag tgtaagaaat 1500  
[0589] ggaacgtatg actacccgca gtattcagaa gaagcgagac taaaagaga ggaaataagt 1560  
[0590] ggagtaaaat tggaatcaat aggaatttac caataactgt caatttattc tacagtggcg 1620  
[0591] agttccctag cactggcaat catggtagct ggtctatcct tatggatgtg ctccaatggg 1680  
[0592] tcgttacaat gcagaatttg catttaa 1707  
[0593] <210> 23  
[0594] <211> 29  
[0595] <212> DNA  
[0596] <213> 人工序列(Artificial Sequence)  
[0597] <220>  
[0598] <223> 合成多核苷酸  
[0599] <220>  
[0600] <223> VEE 5' T-G nt 2 F - Forward Primer  
[0601] <400> 23  
[0602] cgactcacta tagagaggcg gcgcatgag 29  
[0603] <210> 24  
[0604] <211> 29  
[0605] <212> DNA  
[0606] <213> 人工序列(Artificial Sequence)  
[0607] <220>  
[0608] <223> 合成多核苷酸  
[0609] <220>  
[0610] <223> VEE 5' T-G nt 2 R -反向引物  
[0611] <400> 24  
[0612] ctcatgcgcc gcctctctat agtgagtcg 29

SEQ ID NO: 2 (AURAV)	ATAGCGGACGGACTAGTACTTGTACTACAGATTAACTGCGGTGTGCGC-----
SEQ ID NO: 3 (CHIKV)	ATGGCTG-CGTGAGA-----CACACGTAGCCTACAGTTTCTTACTGCTCTACTCT-----
SEQ ID NO: 4 (ONNV)	ATAGCTG-CGTGATA-----CACACAGCAGCTTACGGTTTCATACTGCTCTACT-----
SEQ ID NO: 5 (BEBV)	ATGGCGGCTGTGTGA-CACACGAGCCGTCGATTTCAA-CCTTCTTGCTCCCT-----
SEQ ID NO: 6 (SFV)	ATGGCGGATGTGTGA-CATACACGACGCCAAAGATT-TTGTCCAGCTCCT-----
SEQ ID NO: 7 (MAYV)	ATGGCGGGCAACTGA-CACTTGTTCGCGCTCTTCTAG-GATCCTTTGCTAC-----
SEQ ID NO: 8 (GETV)	ATGGCGGACGTGTGACATCACCCTTCTTCTAG-GATCCTTTGCTAC-----
SEQ ID NO: 9 (SAGV)	ATGGCGGACGTGTGACATCACCCTTCTTCTAG-GATCCTTTGCTAC-----
SEQ ID NO: 10 (NDUV)	ATGG--TGCGGAGTT-GA---GAGACGA-AGCACCA-ACAACACGCGGCTCACC-AT-----
SEQ ID NO: 11 (MIDV)	ATTGGTGGTTACCTA---CAGTGCCACCCACCCCCCA-CCTCCAGCGATCCA-----
SEQ ID NO: 12 (EEEV)	ATAGGTACGGTGA-----GAGCAACCCACCTAT---TTCCACCTATCCAAATGG-----
SEQ ID NO: 13 (FMV)	ATAGGTATGGTTA-----GAGCGCCTACCTAC-----TTAACCGATCCAAACATGG-----
SEQ ID NO: 14 (Buggy)	ATAGGTATGGTTA-----GAGCGCCTACCTAC-----TTAACCGATCCAAACATGG-----
SEQ ID NO: 15 (VEEV)	ATGGCGGCGCAACA-----GAGAAGCCCAACCAA-----TTACCTACCCAAATGGAG-----
SEQ ID NO: 16 (WHAV)	ATTGGCGGCATAGTA-----CATACTATATAAAGAAACAGCGACCAATTGCAC-----
SEQ ID NO: 17 (SINV)	ATTGACGGCGTAGTA-----CACACTAT-TGAATCAAAACAGCGACCAATTGCAC-----
SEQ ID NO: 18 (BABV)	ATTGGCGGCGTAGTA-----CACACTAT-TGAATCAAAACAGCGACCAATTGCAC-----

\*\* \*

图1A

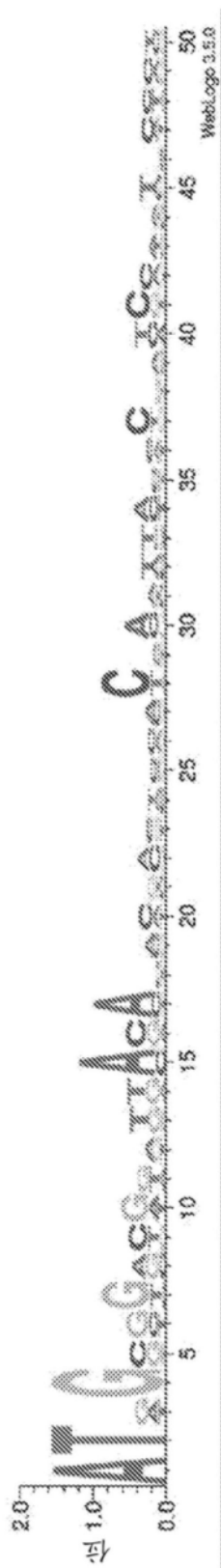


图1B

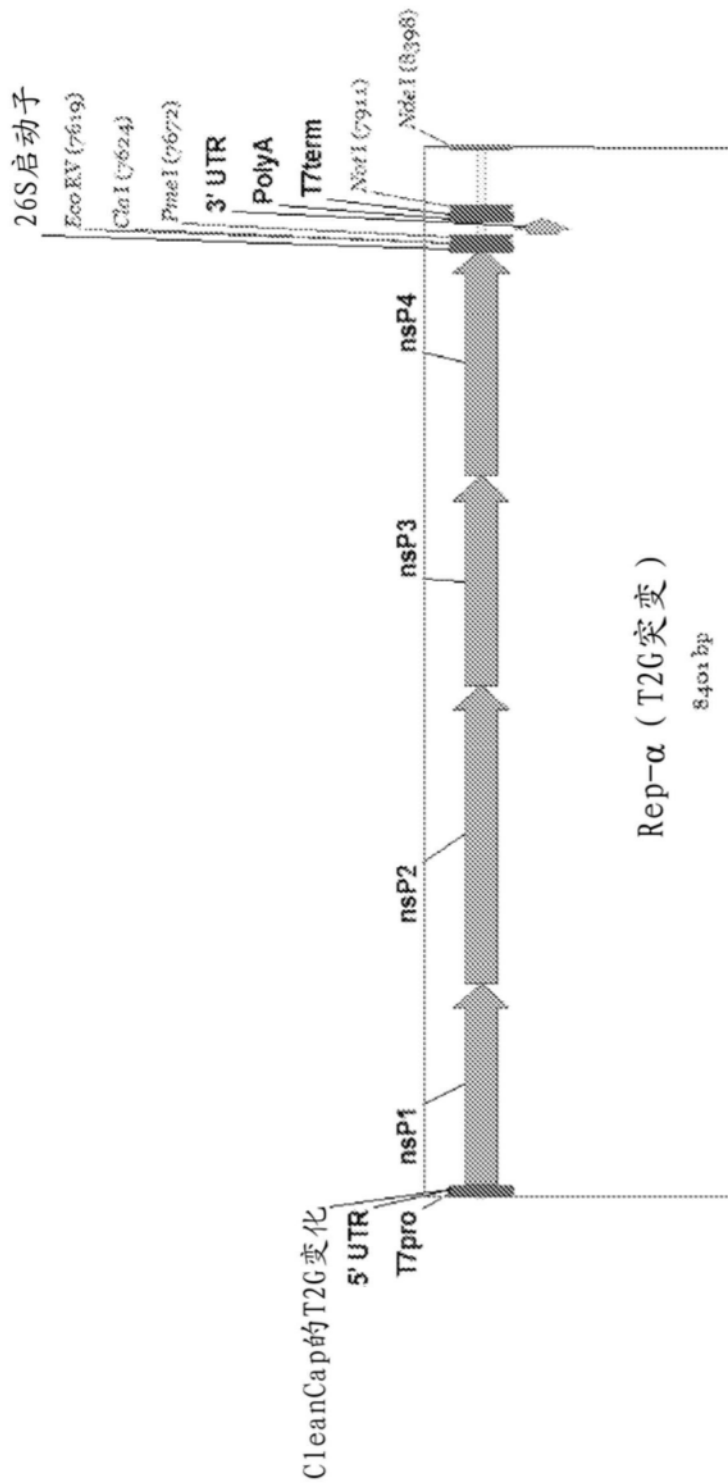


图2

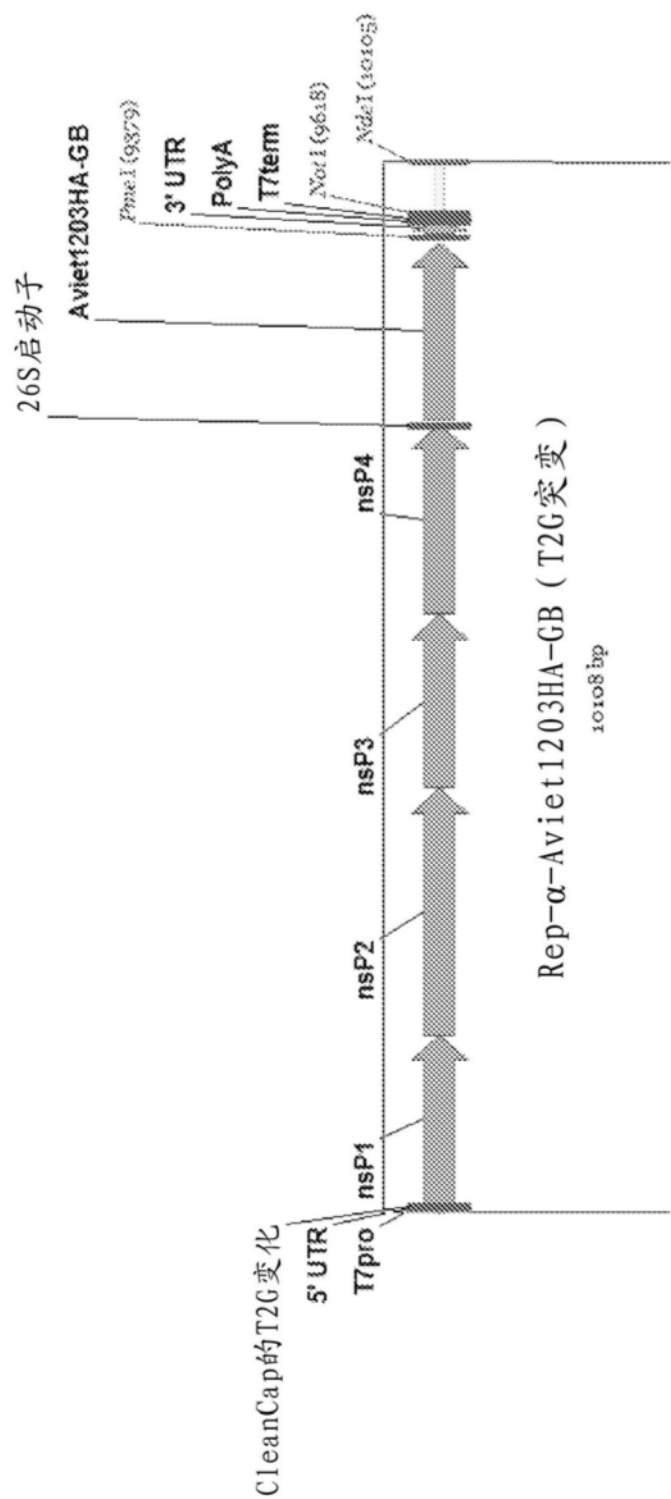


图3

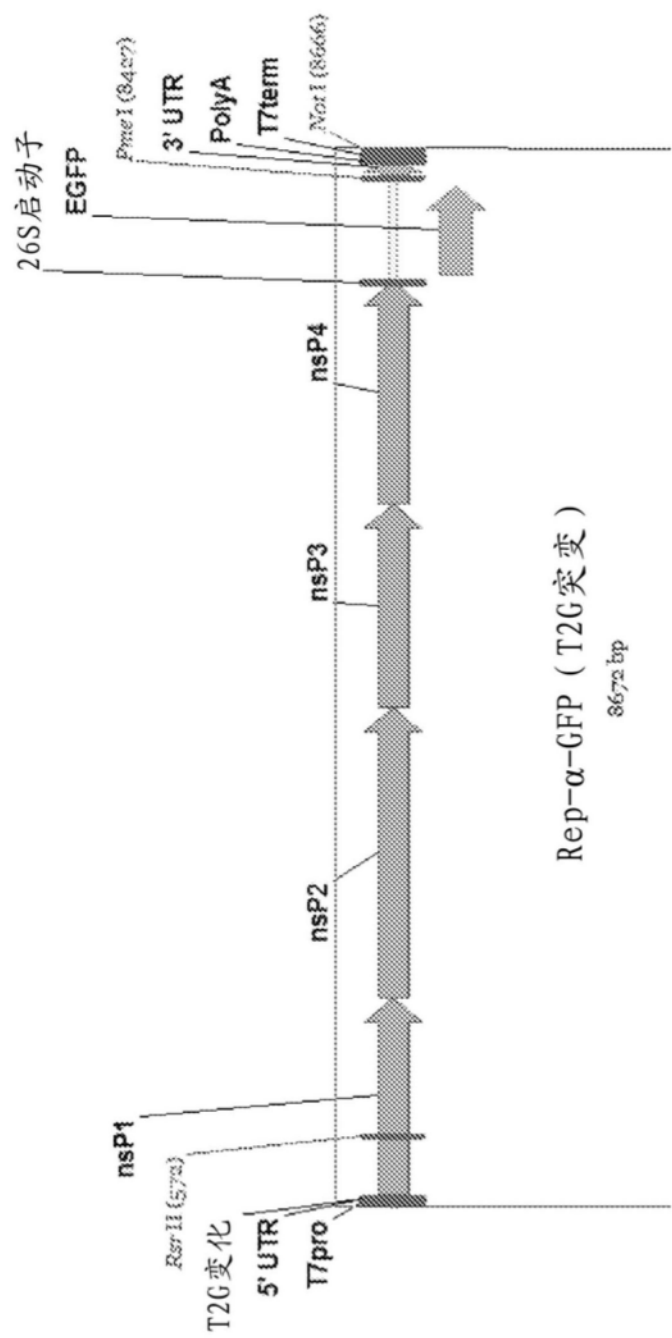


图4

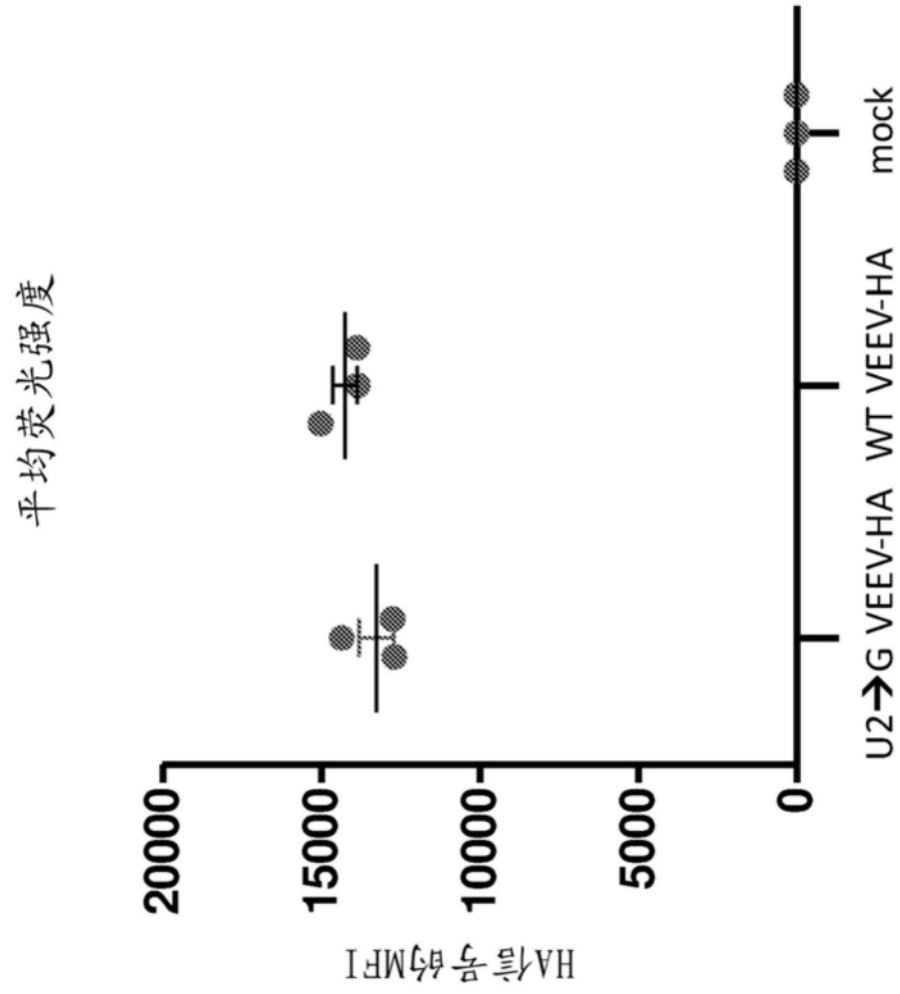


图5

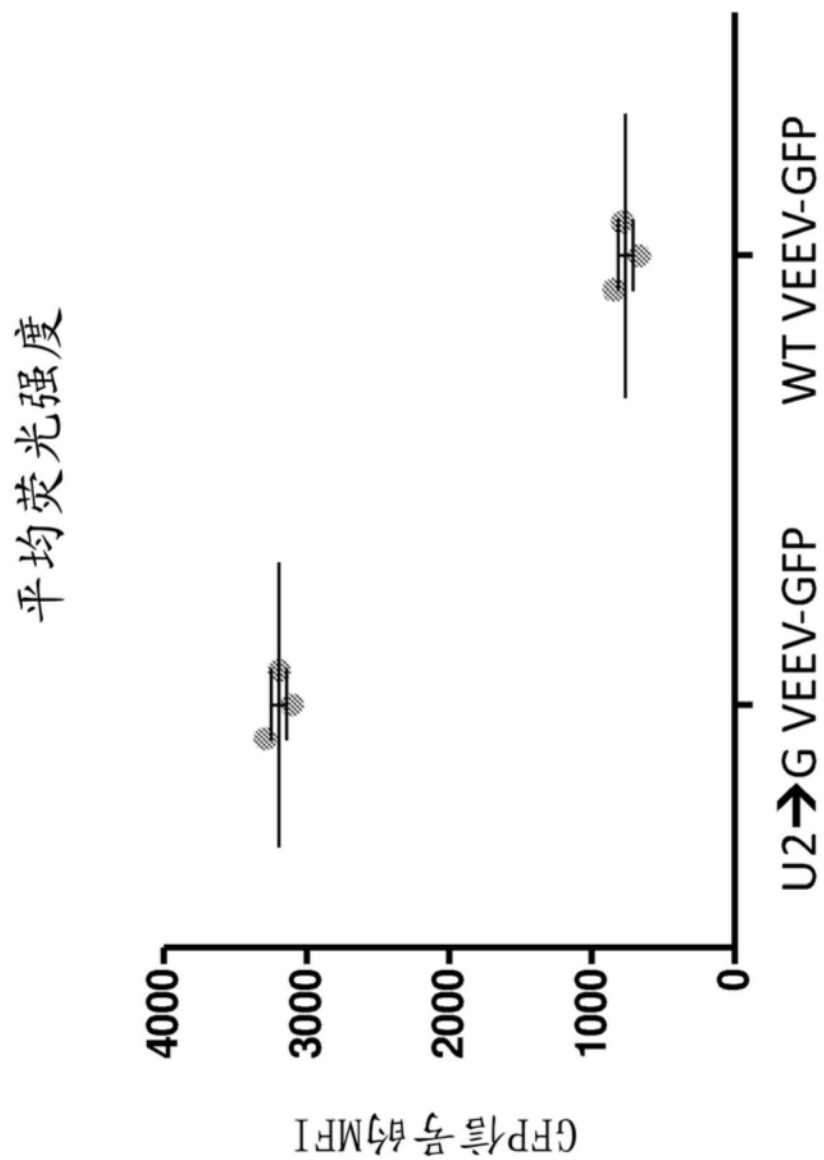


图6

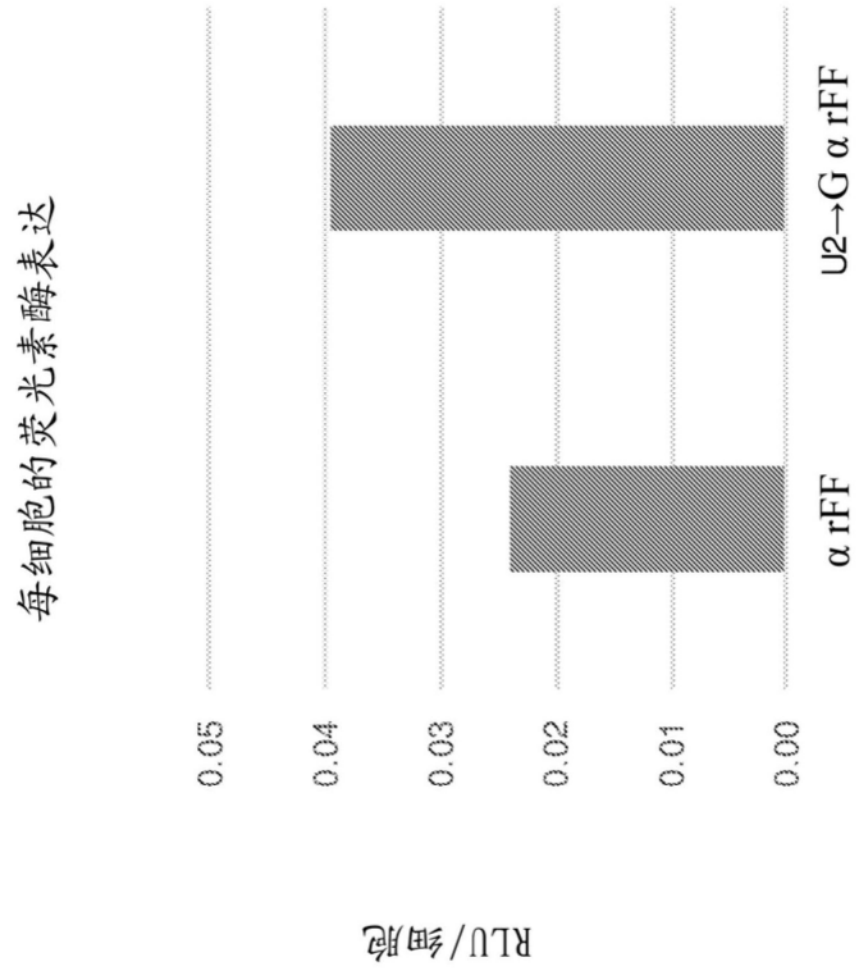


图7

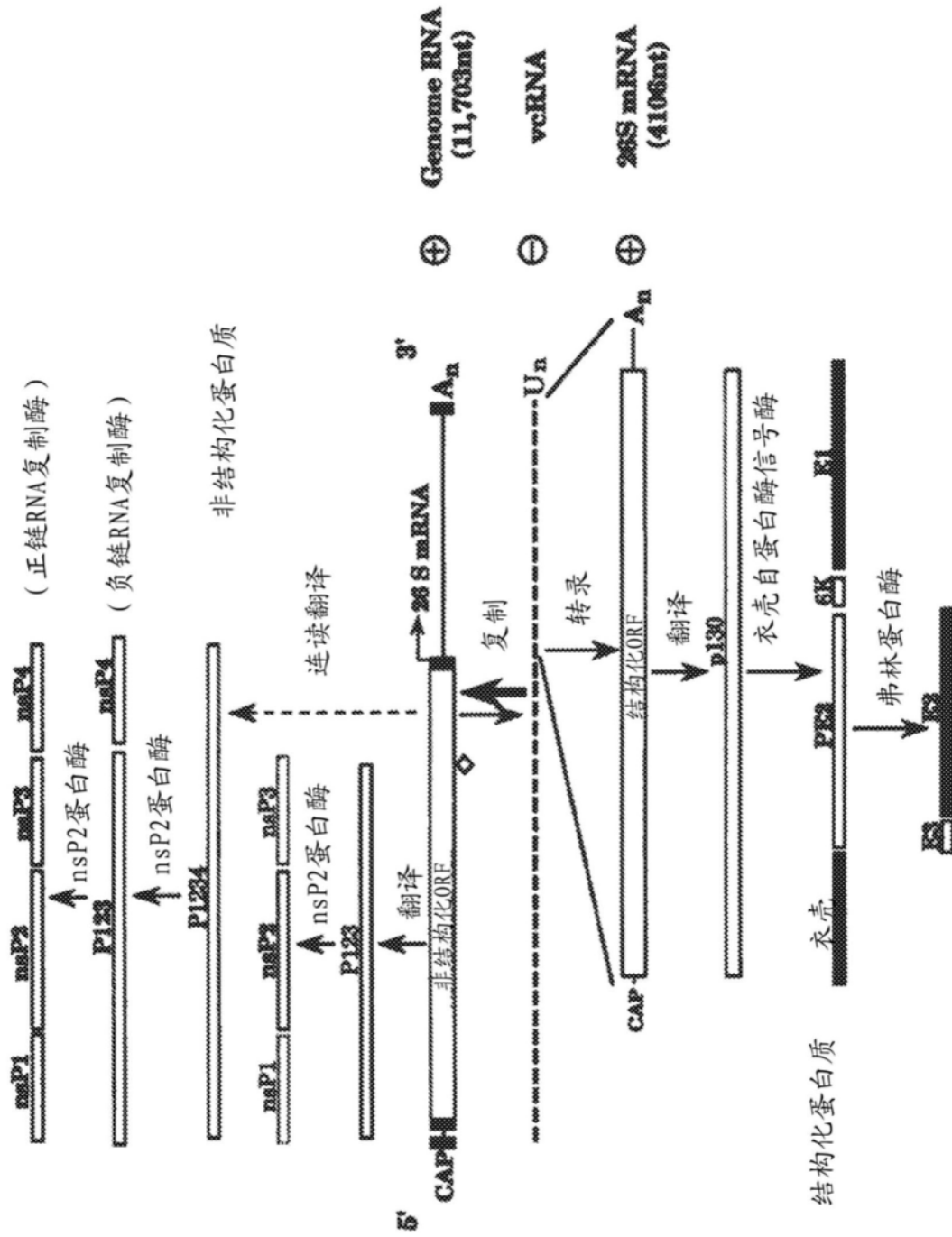


图8