

编码RNA的工程化DNA分子

技术领域

本申请涉及生物技术领域，具体涉及一种包含Poly (A)尾的RNA。所述Poly (A)尾使得编码所述RNA的DNA在原核系统中的复制具有较高的稳定性，且可用于调节所述RNA在真核细胞中的表达水平。

背景技术

可翻译mRNA药物分子的一级结构由5'帽子结构、5'非编码区(5'UTR)，编码区，3'非编码区以及多聚腺苷酸尾(poly(A)尾)组成，其中Poly (A)尾已知的作用包括维持mRNA分子的体内稳定性，参与蛋白翻译的起始过程，后者通过Poly (A)结合蛋白(poly A tail binding protein, PABP)与翻译起始复合物的相互作用实现。真核生物细胞内，Poly (A)尾以转录后修饰的方式，在典型Poly (A)聚合酶作用下合成。

体外制备mRNA药物的第一步，是以含有设计好的产品序列的线性化质粒为模板通过体外转录(IVT)合成，Poly (A)尾则通常以共转录的方式添加在3'UTR下游。为实现共转录添加Poly (A)，需要在模板质粒中包含相对应的poly(dA: dT)序列。然而，质粒中poly(dA: dT)重复序列在大肠杆菌中复制过程中不稳定，经常发生缺失突变导致poly (dA: dT)变短。这种现象不利于通过发酵规模化生产体外转录模板质粒的制备工艺，Poly (A)截短对mRNA的体内稳定性和生物学活性产生显著影响。

发明概述

本申请提供一种全新的Poly (A)尾，以提升其在体外制备过程中的保守性。同时，提供一种基于所述Poly (A)尾调节RNA在真核细胞中表达水平的方法。

具体的，本申请的第一方面提供了一种可在细胞中复制的工程化DNA分子，其包含多聚腺苷尾(Poly(A)尾)编码序列，所述Poly (A)尾编码序列包含：

唯一的元件a和至少一个元件b, 以及至少一个元件c;

唯一的元件a和至少一个元件b, 以及至少一个元件d; 或

唯一的元件a和至少一个元件b, 以及至少一个元件c和至少一个元件d,

在所述Poly (A)尾编码序列中, 所述元件a由多个连续的腺嘌呤(A)核苷酸组成, 所述元件a的长度范围为 $\geq 20\text{nt}$;

所述元件b由多个连续的A核苷酸组成, 所述元件b的长度范围为 $3\text{nt} \leq b < 20\text{nt}$;

所述元件c由一个非A的核苷酸组成, 所述核苷酸选自T、C、G核苷酸;

所述元件d由任意两个或更多个连续的核苷酸组成, 所述核苷酸选自A、T、C、G核苷酸, 其中元件d的5'及3'末端的核苷酸不为A核苷酸, 并且其不包含3个以上连续的A核苷酸; 所述元件d的长度范围为 $2\text{nt} \leq d \leq 20\text{nt}$;

其中所述元件a和元件b不相邻、元件c和元件d不相邻,

且所述Poly (A)尾编码序列不包含彼此相邻的元件b、不包含彼此相邻的元件c、且不包含彼此相邻的元件d。

在一些实施方案中, 所述Poly (A)尾编码序列还进一步包含唯一一个元件e, 所述元件e由一个或两个连续的A组成; 且其位于所述Poly (A)尾编码序列的3'末端, 且与元件d或元件c相邻。

在一些实施方案中, 所述Poly (A)尾编码序列不包含元件a、元件b、元件c、及元件d以外的其他元件。

在一些实施方案中, 所述Poly (A)尾编码序列不包含元件a、元件b、元件c、元件d及元件e以外的其他元件。

在一些实施方案中, 所述Poly (A)尾编码序列包含至少2个元件d。在一些实施方案中, 所述Poly (A)尾编码序列包含至少2个元件c。在一些实施方案中, 所述Poly (A)尾编码序列包含至少一个元件d和一个元件c。

在一些实施方案中, 所述元件b的个数为2-10个, 例如3个、4个、5个、6个、7个、8个、或9个。

在一些实施方案中, 所述元件c的个数为0至10个, 例如1个、2个、3个、4个、5个、6个、7个、8个、或9个。

在一些实施方案中, 所述元件d的个数为0至5个, 例如1个、2个、3个、或4个。

在一些实施方案中, 当元件c和元件d同时存在时, 所述元件c和元件d的

个数总和为2-15个，例如3个、4个、5个、6个、7个、8个、9个、10个、11个、12个、13个或14个。

在一些实施方案中，所述元件a的个数为1，所述元件b的个数为3，所述元件c的个数为2，且所述元件d的个数为1。

在一些实施方案中，所述元件a的个数为1，所述元件b的个数为4，所述元件c的个数为4，且所述元件d的个数为1。

在一些实施方案中，所述元件a的个数为1，所述元件b的个数为5，所述元件c的个数为4，且所述元件d的个数为1。

在一些实施方案中，所述元件a的个数为1，所述元件b的个数为3，所述元件c的个数为3，且所述元件d的个数为1。

在一些实施方案中，所述元件a的个数为1，所述元件b的个数为3，所述元件c的个数为2，且所述元件d的个数为1。

在一些实施方案中，所述元件a $\leq$ 80nt。在一些实施方案中，所述元件a为21nt、22nt、23nt、24nt、25nt、26nt、27nt、28nt、29nt、30nt、31nt、32nt、33nt、34nt、35nt、36nt、37nt、38nt、39nt、40nt、41nt、42nt、43nt、44nt、45nt、46nt、47nt、48nt、49nt、50nt、51nt、52nt、53nt、54nt、55nt、56nt、57nt、58nt、59nt、60nt、61nt、62nt、63nt、64nt、65nt、66nt、67nt、68nt、69nt、70nt、71nt、72nt、73nt、74nt、75nt、76nt、77nt、78nt、或79nt。

在一些实施方案中，所述元件b为3nt、4nt、5nt、6nt、7nt、8nt、9nt、10nt、11nt、12nt、13nt、14nt、15nt、16nt、17nt、18nt或19nt。

在一些实施方案中，其中所述元件d为 $2nt \leq d \leq 20nt$ 、3至18nt、5至16nt、4至10nt、或6至12nt，例如2nt、3nt、4nt、5nt、6nt、7nt、8nt、9nt、10nt、11nt、12nt、13nt、14nt、15nt、16nt、17nt、18nt、19nt或20nt，优选6nt。

在一些实施方案中，所述Poly(A)尾编码序列的长度大于40nt，例如41nt、42nt、43nt、44nt、45nt、46nt、47nt、48nt、49nt、50nt、51nt、52nt、53nt、54nt、55nt、56nt、57nt、58nt、59nt、60nt、61nt、62nt、63nt、64nt、65nt、66nt、67nt、68nt、69nt、70nt、71nt、72nt、73nt、74nt、75nt、76nt、77nt、78nt、79nt、80nt、81nt、82nt、83nt、84nt、85nt、86nt、87nt、88nt、89nt、90nt、91nt、92nt、93nt、94nt、95nt、96nt、97nt、98nt、99nt、100nt、101nt、102nt、103nt、104nt、105nt、106nt、107nt、108nt、109nt、110nt、111nt、112nt、113nt、114nt、115nt、116nt、117nt、118nt、119nt、120nt、121nt、122nt、

123nt、124nt、125nt、126nt、127nt、128nt、129nt、130nt、131nt、132nt、
133nt、134nt、135nt、136nt、137nt、138nt、139nt、140nt、141nt、142nt、
143nt、144nt、145nt、146nt、147nt、148nt、149nt、150nt、151nt、152nt、
153nt、154nt、155nt、156nt、157nt、158nt、159nt、160nt、161nt、162nt、
163nt、164nt、165nt、166nt、167nt、168nt、169nt、170nt、171nt、172nt、
173nt、174nt、175nt、176nt、177nt、178nt、179nt、180nt、181nt、182nt、
183nt、184nt、185nt、186nt、187nt、188nt、189nt、190nt、191nt、192nt、
193nt、194nt、195nt、196nt、197nt、198nt、199nt、200nt、201nt、202nt、
203nt、204nt、205nt、206nt、207nt、208nt、209nt、210nt、211nt、212nt、
213nt、214nt、215nt、216nt、217nt、218nt、219nt、220nt、221nt、222nt、
223nt、224nt、225nt、226nt、227nt、228nt、229nt、230nt、231nt、232nt、
233nt、234nt、235nt、236nt、237nt、238nt、239nt、240nt、241nt、242nt、
243nt、244nt、245nt、246nt、247nt、248nt、249nt、250nt、251nt、252nt、
253nt、254nt、255nt、256nt、257nt、258nt、259nt、260nt、261nt、262nt、
263nt、264nt、265nt、266nt、267nt、268nt、269nt、270nt、271nt、272nt、
273nt、274nt、275nt、276nt、277nt、278nt、279nt、280nt、281nt、282nt、
283nt、284nt、285nt、286nt、287nt、288nt、289nt、290nt、291nt、292nt、
293nt、294nt、295nt、296nt、297nt、298nt、299nt、300nt、301nt、302nt、
303nt、304nt、305nt、306nt、307nt、308nt、309nt、310nt、311nt、312nt、
313nt、314nt、315nt、316nt、317nt、318nt、319nt、320nt、321nt、322nt、
323nt、324nt、325nt、326nt、327nt、328nt、329nt、330nt、331nt、332nt、
333nt、334nt、335nt、336nt、337nt、338nt、339nt、340nt、341nt、342nt、
343nt、344nt、345nt、346nt、347nt、348nt、349nt、350nt、351nt、352nt、
353nt、354nt、355nt、356nt、357nt、358nt、359nt、360nt、361nt、362nt、
363nt、364nt、365nt、366nt、367nt、368nt、369nt、370nt、371nt、372nt、
373nt、374nt、375nt、376nt、377nt、378nt、379nt、380nt、381nt、382nt、
383nt、384nt、385nt、386nt、387nt、388nt、389nt、390nt、391nt、392nt、
393nt、394nt、395nt、396nt、397nt、398nt、399nt、400nt等。

在一些实施方案中，所述元件a的50%或以上的多核苷酸位于所述Poly (A)尾编码序列的5'部分或3'部分。在一些实施方案中，所述元件a的50%以上的多核苷酸位于所述Poly (A)尾编码序列的5'部分。在一些实施方案中，所述

元件a的50%以上的多核苷酸位于所述Poly (A)尾编码序列的3'部分。在一些实施方案中, 所述元件a位于所述Poly (A)尾编码序列的3'部分的核苷酸数量和位于所述Poly (A)尾编码序列的5'部分的核苷酸数量相等。

在一些实施方案中, 所述元件c为G、C或T。

在一些实施方案中, 所述元件d包含回文序列。元件d是回文序列。在一些实施方案中, 所述元件d包含选自以下序列: GATATC (SEQ ID NO: 15)、GTATAC (SEQ ID NO: 16)、GAATCT (SEQ ID NO: 17)、GCATATGACT (SEQ ID NO: 18) 及GATATCGTATAC (SEQ ID NO: 19)。在一些实施方案中, 所述元件d为选自以下序列: GATATC (SEQ ID NO: 15)、GTATAC (SEQ ID NO: 16)、GAATCT (SEQ ID NO: 17)、GCATATGACT (SEQ ID NO: 18) 及GATATCGTATAC (SEQ ID NO: 19)。在一些实施方案中, 所述元件d包含如SEQ ID NO:15所示的多核苷酸序列。在一些实施方案中, 所述元件d的多核苷酸序列如SEQ ID NO:15所示。

在一些实施方案中, 所述Poly (A)尾编码序列3'末端的核苷酸为A。在一些实施方案中, 所述Poly (A)尾编码序列3'末端的核苷酸为G。在一些实施方案中, 所述Poly (A)尾编码序列3'末端的核苷酸为C。在一些实施方案中, 所述Poly (A)尾编码序列3'末端的核苷酸为T。

在一些实施方案中, 在所述Poly (A)尾编码序列的3'部分包含一个或多个非A核苷酸。在一些实施方案中, 在所述Poly (A)尾编码序列靠近3'末端的1/2处包含一个或多个非A核苷酸。在一些实施方案中, 在所述Poly (A)尾编码序列靠近3'末端的1/3处包含一个或多个非A核苷酸。在一些实施方案中, 在所述Poly (A)尾编码序列靠近3'末端的1/4处包含一个或多个非A核苷酸。

在一些实施方案中, 所述Poly (A)尾编码序列的结构为:

元件a-元件c-元件b-元件c-元件b-元件c-元件b-元件c-元件b;

元件b-元件c-元件b-元件c-元件a-元件d-元件b-元件c-元件b-元件c-元件b;

元件b-元件c-元件b-元件c-元件b-元件d-元件a-元件c;

元件a-元件d-元件b-元件c-元件b-元件c-元件b; 或

元件b-元件c-元件b-元件c-元件b-元件d-元件a。

在一些实施方案中, 所述Poly (A)尾编码序列的结构为:

元件a-元件c-元件b-元件c-元件b-元件c-元件b-元件c-元件b; 具体为60A-G-19A-G-19A-G-19A-G-3A。

在一些实施方案中，所述Poly (A)尾编码序列的结构为：

7A-C-18A-G-60A-GG-7A-C-18A-G-14A、

19A-G-19A-G-19A-元件d-60A-G、

60A-元件d-19A-G-19A-G-17A、

19A-G-19A-G-19A-元件d-60A、

19A-G-19A-G-19A-元件d-60A、

19A-C-19A-C-19A-元件d-60A、

19A-T-19A-T-19A-元件d-60A、

19A-G-19A-G-19A-元件d-60A、或

19A-G-19A-G-19A-元件d-60A；

且其中元件d由6个或12个核苷酸组成。

在一些实施方案中，所述Poly (A)尾编码序列如SEQ ID NO： 1至10中任一项所示。

在一些实施方案中，所述Poly (A)尾编码序列如SEQ ID NO： 3或SEQ ID NO： 4所示。

在一些实施方案中，所述工程化DNA分子在其Poly (A)尾编码序列的5'端一侧进一步连接目的基因片段，所述目的基因片段与所述Poly (A)尾编码序列共同编码RNA。在一些实施方案中，所述工程化DNA分子在其Poly (A)尾编码序列的5'端一侧进一步连接目的基因片段，所述目的基因片段与所述Poly (A)尾编码序列共同编码mRNA。在一些实施方案中，所述目的基因片段包含蛋白编码序列或非蛋白编码序列例如功能RNA编码序列。在一些实施方案中，所述目的基因片段在所述蛋白编码序列或功能RNA编码序列的5'端一侧还进一步包含5'UTR编码序列。在一些实施方案中，目的基因片段在所述蛋白编码序列或功能RNA编码序列的3'端一侧还进一步包含3'UTR编码序列。在一些实施方案中，目的基因片段在所述蛋白编码序列或功能RNA编码序列的3'端一侧还进一步包含3'UTR编码序列，且在所述蛋白编码序列或功能RNA编码序列的5'端一侧还进一步包含5'UTR编码序列。在一些实施方案中，所述工程化DNA分子进一步包含复制子，例如复制起始位点，例如ORI。在一些实施方案中，所述工程化DNA分子中还进一步包含标记基因以方便对包含所述工程化DNA分子的细胞进行筛选，所述标记基因选自，例如针对抗生素的抗性基因、荧光蛋白等。在一些实施方案中，所述的DNA分子进一步包

含启动子，所述启动子启动所述目的基因片段与所述Poly (A)尾编码序列共同编码RNA的转录。在一些实施方案中，所述启动子为原核启动子。在一些实施方案中，所述启动子为真核启动子。在一些实施方案中，所述的DNA分子进一步包含复制子例如复制起始位点、启动子、5'UTR编码序列、蛋白编码序列、及3'UTR编码序列。在一些实施方案中，所述的DNA分子进一步包含复制子例如复制起始位点、抗性基因、5'UTR编码序列、蛋白编码序列、及3'UTR编码序列。在一些实施方案中，所述的DNA分子进一步包含复制子例如复制起始位点、抗性基因、启动子、5'UTR编码序列、蛋白编码序列、及3'UTR编码序列。在一些实施方案中，所述蛋白编码序列编码HPV病毒抗原蛋白。在一些实施方案中，所述HPV蛋白源自16型和/或18型HPV。在一些实施方案中，所述蛋白编码序列编码HPV的E2、E6或E7蛋白。在一些实施方案中，所述蛋白编码序列编码HPV的E6和E7蛋白的融合蛋白。在一些实施方案中，所述蛋白编码序列编码HPV的E2、E6和E7蛋白的融合蛋白。在一些实施方案中，所述融合蛋白的多肽片段来源于16型和/或18型HPV。在一些实施方案中，所述融合蛋白的多肽片段来源于16型和/或18型HPV E2、E6和E7蛋白。在一些实施方案中，所述蛋白编码序列编码如SEQ NO: 26所示的多肽或其保守取代变体。

在一些实施方案中，所述工程化DNA分子包含如SEQ ID NO: 22-25中任一项所示的多核苷酸序列，或如SEQ ID NO: 22-25中任一项所示的多核苷酸序列的同义突变体，或与SEQ ID NO: 22-25中任一项所示的多核苷酸序列或其同义突变体具有85%以上序列同一性的多核苷酸序列。

在一些实施方案中，所述DNA分子为DNA质粒。在一些实施方案中，所述DNA分子为线性质粒或环状质粒。在一些实施方案中，所述DNA分子为单链或双链。在一些实施方案中，所述质粒为基于pUC、pTZ、pMB1或pColE1的质粒。在一些实施方案中，所述质粒为基于pUC57载体的质粒。

在一些实施方案中，所述细胞为原核细胞。在一些实施方案中，所述细胞为recA-的细菌。在一些实施方案中，所述细胞为大肠杆菌。在一些实施方案中，所述大肠杆菌选自：K-12及其衍生菌株，以及B菌株及其衍生菌株。在一些实施方案中，所述大肠杆菌选自：MG1655、DH5或DH5 α 、DH10B、BL21、DB3.1、HB101、JM109、JM110、MC1061、MG1655、Pir1、Stbl2、Stbl3、Top10、XL1Blue、XL10 Gold、BLR、HMS174、Tuner、Rostetta2、

Lemo21、T7Express、Origami2。

本申请的第二方面公开了包含前述第一方面的DNA分子的细胞。在一些实施方案中，前述第一方面的DNA分子可在所述细胞中复制和/或转录。在一些实施方案中，所述细胞为原核细胞，且前述第一方面的DNA分子可在所述原核细胞中复制。在一些实施方案中，所述细胞为recA-的细菌。在一些实施方案中，所述细胞为大肠杆菌。在一些实施方案中，所述原核细胞为感受态细胞。在一些实施方案中，所述原核细胞为工程化细胞。在一些实施方案中，所述原核细胞为工程化的原核细胞。在一些实施方案中，所述细胞为大肠杆菌，所述大肠杆菌选自：K-12及其衍生菌株，以及B菌株及其衍生菌株。在一些实施方案中，所述大肠杆菌选自：MG1655、DH5或DH5 α 、DH10B、BL21、DB3.1、HB101、JM109、JM110、MC1061、MG1655、Pir1、Stbl2、Stbl3、Top10、XL1Blue、XL10 Gold、BLR、HMS174、Tuner、Rostetta2、Lemo21、T7Express、Origami2。在一些实施方案中，所述细胞为真核细胞，且前述第一方面的DNA分子可在所述真核细胞中转录。在一些实施方案中，所述真核细胞为哺乳动物细胞。在一些实施方案中，所述真核细胞选自酵母菌或霉菌。

本申请的第三方面，提供了一种Poly (A) 尾。所述Poly (A) 尾：

- (1) 由前述第一方面的工程化DNA分子转录获得；
- (2) 由化学合成的方法获得，且与通过所述第一方面的工程化DNA分子转录获得的Poly (A) 尾具有相同的多核苷酸序列；或
- (3) 对前述(1)或(2)中的Poly (A) 尾做进一步的修饰获得。

在一些实施方案中，所述进一步的修饰包含将由(1)或(2)获得的Poly (A) 尾中的一个或多个核糖核苷酸替换为一个或多个脱氧核糖核苷酸。在一些实施方案中，所述一个或多个核糖核苷酸被替换为与所述一个或多个核糖核苷酸相应的脱氧核糖核苷酸，例如将所述Poly (A) 尾的一个或多个核糖核苷酸A替换为脱氧核糖核苷酸A，将其中的一个或多个核糖核苷酸U替换为脱氧核糖核苷酸T，将其中的一个或多个核糖核苷酸C替换为脱氧核糖核苷酸C，将其中的一个或多个核糖核苷酸G替换为脱氧核糖核苷酸G，或将其中的一个或多个核糖核苷酸G替换为核糖核苷酸I(肌苷)或脱氧核糖核苷酸I。在一些实施方案中，所述修饰为化学修饰。在一些实施方案中所述修饰为碱基编辑。在一些实施方案中，所述修饰对由(1)或(2)获得的Poly (A) 尾中的一个或多个核糖核苷酸进行脱氨基处理。

在一些实施方案中，所述Poly (A) 尾包含选自如SEQ ID NO: 1-10中任一项所示的多核苷酸序列。在一些实施方案中，所述Poly (A) 尾的多核苷酸序列如SEQ ID NO: 1-10中任一项所示。

本申请还提供了前述Poly (A) 尾的用途。在一些实施方案中，所述Poly (A) 尾用于使RNA分子更稳定的用途，其中，所述Poly (A) 尾位于所述RNA的3'末端，所述更稳定是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内更加稳定。在一些实施方案中，所述Poly (A) 尾用于使RNA分子稳定性降低的用途，其中，所述Poly (A) 尾位于所述RNA的3'末端，该稳定性的降低是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内的稳定性降低。在一些实施方案中，所述Poly (A) 尾用于使RNA分子在相同时间内的表达量提高的用途，所述表达量提高是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内的表达量提高。在一些实施方案中，所述Poly (A) 尾用于使RNA分子在相同时间内的表达量降低的用途，所述表达量降低是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内的表达量降低。在一些实施方案中，所述Poly (A) 尾用于使RNA分子表达时间延长的用途，所述表达时间延长是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内的表达时间延长。在一些实施方案中，所述Poly (A) 尾用于使RNA分子表达时间缩短的用途，所述表达时间缩短是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内的表达时间缩短。在一些实施方案中，所述Poly (A) 尾用于使RNA分子半衰期延长的用途，所述半衰期延长是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内的半衰期延长。在一些实施方案中，所述Poly (A) 尾用于使RNA分子半衰期缩短的用途，所述半衰期缩短是相对于包含另一Poly (A) 尾的RNA分子，在细胞外、细胞内或动物体内的半衰期缩短。

在一些实施方案中，所述RNA分子为mRNA分子。在一些实施方案中，所述Poly (A) 尾及另一Poly (A) 尾分别属于本申请的第三方面的Poly (A) 尾的两条不同的Poly (A) 尾。在一些实施方案中，所述Poly (A) 尾是属于本申请的第三方面的Poly (A) 尾，所述另一Poly (A) 尾为本申请的第三方面的Poly (A) 尾以外的Poly (A) 尾。在一些实施方案中，所述细胞内为宿主细胞内，所述宿主细胞为真核细胞。在一些实施方案中，所述宿主细胞为

哺乳动物细胞。在一些实施方案中，所述宿主细胞为人细胞。

本申请还提供了编码本申请的第三方面的所述Poly (A)尾的DNA片段或DNA与RNA的杂合分子片段，以及将所述DNA片段或DNA与RNA的杂合分子片段用于使编码RNA的DNA分子或DNA与RNA的杂合分子在宿主细胞中的复制更加保守的用途，在所述用途中，编码本申请的第三方面的所述Poly (A)尾的DNA片段或DNA与RNA的杂合分子片段在所述DNA分子或DNA与RNA的杂合分子中，位于所述RNA编码序列的3'端一侧。在一些实施方案中，所述宿主细胞为原核细胞。在一些实施方案中，所述宿主细胞为recA-的细菌。在一些实施方案中，所述宿主细胞为大肠杆菌。在一些实施方案中，所述大肠杆菌选自：K-12及其衍生菌株，以及B菌株及其衍生菌株。在一些实施方案中，所述大肠杆菌选自：MG1655、DH5或DH5 α 、DH10B、BL21、DB3.1、HB101、JM109、JM110、MC1061、MG1655、Pir1、Stbl2、Stbl3、Top10、XL1Blue、XL10 Gold、BLR、HMS174、Tuner、Rostetta2、Lemo21、T7Express、Origami2。

本申请的第四方面还提供了一种RNA分子，所述RNA分子包含第三方面的Poly (A)尾。在一些实施方案中，所述RNA分子为mRNA分子。在一些实施方案中，所述RNA分子：

(1) 由前述第一方面的工程化DNA分子转录获得；

(2) 由化学合成的方法获得，且与项(1)的RNA分子具有相同的多核苷酸序列；或

(3) 由对前述(1)或(2)中的RNA分子做进一步的修饰获得。

在一些实施方案中，所述进一步的修饰包含将由(1)或(2)获得的RNA分子中的一个或多个核糖核苷酸替换为一个或多个脱氧核糖核苷酸。在一些实施方案中，所述一个或多个核糖核苷酸被替换为与所述一个或多个核糖核苷酸相应的脱氧核糖核苷酸，例如将所述RNA分子的一个或多个核糖核苷酸A替换为脱氧核糖核苷酸A，将其中的一个或多个核糖核苷酸U替换为脱氧核糖核苷酸T，将其中的一个或多个核糖核苷酸C替换为脱氧核糖核苷酸C，将其中的一个或多个核糖核苷酸G替换为脱氧核糖核苷酸G，或将其中的一个或多个核糖核苷酸G替换为核糖核苷酸I（肌苷）或脱氧核糖核苷酸I。在一些实施方案中，所述修饰为化学修饰。在一些实施方案中所述修饰为碱基编辑。在一些实施方案中，所述修饰对由(1)或(2)获得的RNA分子中的一个或

多个核糖核苷酸进行脱氨基处理。在一些实施方案中，所述进一步修饰为转录后修饰。在一些实施方案中，所述进一步修饰包含加帽处理。在一些实施方案中，所述进一步修饰包含剪接。在一些实施方案中，所述进一步修饰包含剪接和加帽处理。

在一些实施方案中，所述RNA分子包括编码RNA(coding RNA)或非编码RNA(non-coding RNA, ncRNA)。在一些实施方案中，所述RNA分子为Pre-mRNA。在一些实施方案中，所述RNA为成熟mRNA。在一些实施方案中，所述RNA分子为长链非编码RNA(long noncoding RNA, lncRNA)。在一些实施方案中，所述RNA分子还进一步包含5'-帽结构。在一些实施方案中，所述RNA分子的多核苷酸序列如SEQ ID NO: 22-25中任一项所示。在一些实施方案中，所述RNA分子包含如SEQ ID NO: 22-25中任一项所示的多核苷酸序列。

此外，本申请还提供了DNA与RNA的杂合分子，其携带与前述第一方面的工程化DNA分子相同的遗传信息，与前述第三方面的Poly (A)尾相同的遗传信息，或与前述第四方面的RNA分子相同的遗传信息。

此外，本申请还提供了核酸分子文库。在一些实施方案中，所述核酸分子文库包含前述第一方面的工程化DNA分子、前述第三方面的Poly (A)尾、编码本申请的第三方面的所述Poly (A)尾的DNA片段或DNA与RNA的杂合分子片段、或前述第四方面的RNA分子。

此外，本申请还提供了调控蛋白表达的方法，包括：将根据前述核酸分子文库中的多种核酸分子在不同时间，和/或以不同的数量比例导入目的细胞。在一些实施方案中，所述核酸分子为前述第一方面的工程化DNA分子。在一些实施方案中，所述核酸分子为前述第四方面的RNA分子。

应当理解，本文描述的本申请的方面和实施方案包括“包含”，“组成”和“基本上由.....组成”的方面和实施方案。以上详细描述了本申请的优选实施方案，但是，本申请并不限于此。在本申请的技术构思范围内，可以对本申请的技术方案进行多种简单变型，包括各个技术特征以任何其它的合适方式进行组合，这些简单变型和组合同样应当视为本申请所公开的内容，均属于本申请的保护范围。

附图说明

图1示出了实施例2中，本申请中的10种poly (A)在大肠杆菌DH5α中的复

制稳定性；

图2示出了实施例2中，本申请中的10种poly(A)在大肠杆菌DH5 α 中的碱基缺失统计；

图3示出了实施例3中poly(A) P1、P2、P3、P4及 poly(A)对照C1、C2在包括HPV抗原序列质粒体系下大肠杆菌DH5 α 中的复制稳定性；

图4示出了实施例3中poly(A) P3、P4及poly(A)对照C1、C2在包括HPV抗原序列质粒体系下大肠杆菌DH5 α 中的碱基缺失统计；

图5示出了实施例3中poly(A) P1、P2、P3及poly(A)对照C1在包括HPV抗原序列质粒体系下大肠杆菌DH5 α 中30°C和37°C两个温度条件下的复制稳定性比较；

图6示出了实施例中通用载体质粒DNA的图谱示例。

图7示出了实施例4中，带有poly(A) P3、P4、P5、P8、P9及对照C2的荧光素酶在小鼠体内表达水平的动物成像结果；

图8示出了实施例4中，带有poly(A) P3、P4、P5、P8、P9及对照C2的荧光素酶在小鼠体内表达水平的动物成像后荧光强度定量结果（ns，无显著性差异；**，显著性差异， $p<0.01$ ）。

发明详述

本申请首先提供了在体外稳定扩增Poly（A）尾转录模板DNA的方法，以减少所述DNA在细胞中大量复制时，其中Poly（A）尾转录模板序列的突变频率。从而，基于所述DNA获得大量包含序列确定的Poly(A)尾的RNA。在此基础上，使得经过工程化设计而具有某种特定功能的Poly（A）尾的RNA，例如mRNA，得以通过体外发酵，实现规模化生产。

此外，本申请还提供了可在体外稳定扩增的包含Poly（A）尾转录模板的DNA，以及由所述DNA转录形成的RNA。进一步的，在满足体外稳定扩增的前提下，本申请还进一步提供了对RNA稳定性和/或表达效率具有不同调节作用的一组Poly（A）尾、包含所述Poly（A）尾的RNA、包含所述Poly（A）尾编码序列的DNA、以及由所述Poly（A）尾、所述RNA或所述DNA组成的文库。

更进一步地，本申请还提供了前述Poly（A）尾、RNA、DNA、以及文库的用途。

术语

如本文所用,“元件a”、“元件b”、“元件c”、“元件d”及“元件e”是Poly(A)可包含的元件类别。元件a由多个连续的腺嘌呤(A)核苷酸组成,所述元件a的长度范围为 $\geq 20\text{nt}$;元件b由多个连续的A核苷酸组成,所述元件b的长度范围为 $3\text{nt} \leq b < 20\text{nt}$;元件c由一个非A的核苷酸组成,所述核苷酸选自T、C、G核苷酸;元件d由任意两个或更多个连续的核苷酸组成,所述核苷酸选自A、T、C、G核苷酸,其中元件d的5'及3'末端的核苷酸不为A核苷酸,并且元件d不包含3个以上连续的A核苷酸,所述元件d的长度范围为 $2\text{nt} \leq d \leq 20\text{nt}$ 。元件e由一个或两个连续的A组成,且当其存在时,其位于Poly(A)尾编码序列的3'末端,且与元件d或元件c相邻。当在一条Poly(A)中,含有两个或多个“元件b”、“元件c”、及“元件d”时,每两个元件b的序列可以相同或不同、每两个元件c的序列可以相同或不同,以及每两个元件d的序列可以相同或不同,只要他们各自均符合前述对元件a、b、c和d的定义即可。在本申请中,Poly(A)尾中的“元件a”、“元件b”、“元件c”、“元件d”、“元件e”等均以术语“元件”指代。

如本文所用,当描述两个或多个元件的位置关系为“不相邻”,则是指两个或多个元件两两之间不毗邻。换言之,即所述两个或多个元件两两之间至少包含所述两个元件的核苷酸以外的一个或多个其他核苷酸或碱基。

在本文中,“编码”是指i)DNA序列中包含可被转录成RNA分子的遗传信息,和/或ii)RNA分子中包含可被翻译成氨基酸序列的遗传信息。因此,如本文所用,“编码序列”可用以指代mRNA前体或成熟mRNA中可以被翻译为蛋白质的核糖核苷酸(RNA)序列或其片段,亦可指代作为模板用以转录所述mRNA前体或成熟mRNA的脱氧核糖核苷酸(DNA)序列的互补序列或其片段。此外,本申请的“编码序列”还可以进一步包含编码蛋白、功能性核酸、或其片段,例如miRNA、shRNA、dsRNA、向导RNA、Poly(A)尾、5'UTR、3'UTR等的多核苷酸序列。其中,包含可被转录成RNA分子的遗传信息的DNA分子称为所述RNA分子的“编码核酸”;包含可被翻译成氨基酸序列的遗传信息的RNA分子称为所述氨基酸序列的“编码核酸”。

在本申请中,所有多核苷酸序列中的核苷酸由5'端向3'端编号,即5'末端的核苷酸为第一个核苷酸,3'末端的核苷酸为最后一个核苷酸。如无特别说明,“5'端”与“5'末端”可互换使用;“3'端”与“3'末端”可互换使用。“5'端”

与“3'端”着重描述同一条核酸序列中，核苷酸之间，核苷酸序列区段之间，或核苷酸与核苷酸序列区段之间的相对位置关系；“5'末端”及“3'末端”则分别用于描述一条核酸序列或一条核酸序列的某一个区段的第一个及最后一个核苷酸所在的位置。“5'端一侧”则用以描述同一条多核苷酸序列中两段相互之间无重叠部分的序列的相对位置关系，当描述一段序列位于另一段序列的5'端一侧，则是指所述一段序列相对于另一段序列更加靠近所述多核苷酸序列的“5'端”。同理，当描述一段序列位于另一段序列的3'端一侧，则是指一段序列相对于所述另一段序列更加靠近所述多核苷酸序列的“3'端”，且所述一段序列与所述另一段序列彼此不包含重叠部分。具体例如“Poly (A)尾的DNA编码序列位于RNA编码序列的3'端”是指所述Poly (A)尾的DNA编码序列作为所述RNA编码序列的组成部分，包含所述RNA编码序列的3'末端的核苷酸。此外，如本文所用，“5'部分”是指以所述多核苷酸序列“中心位置”为界，靠近所述多核苷酸序列5'末端的一半。“3'部分”则指以所述多核苷酸序列中心位置为界，靠近所述多核苷酸序列3'末端的一半。本申请所述“中心位置”至所述5'端和至所述3'端的核苷酸数量相等。

如本文所用，将术语“保守”用于描述核酸分子的复制时，是指所述复制过程中出现突变的概率低。在此语境中，“保守”是个相对的概念。例如当描述“Poly (A)尾的DNA编码序列用于使编码RNA的DNA分子在宿主细胞中的复制更加保守”时，是指编码RNA的亲代DNA分子复制为子代DNA分子后，如果所述RNA分子包含该Poly (A)尾，则相对于不包含所述Poly (A)尾（例如包含某种其他Poly (A)尾的RNA分子）的RNA分子的编码DNA，所述子代DNA分子与所述亲代DNA分子具有100%序列同一性的概率更高；或由所述亲代DNA分子复制获得的多条所述子代DNA分子与所述亲代DNA分子之间的平均序列同一性更高。

在本申请中，当描述“调控”RNA分子表达时，所述“调控”是指：在相同的时长内使所述RNA分子表达的蛋白或功能RNA的总量提高或降低；或使所述RNA可以在更长或更短的时间范围内表达蛋白或功能RNA，所述提高、降低、或更长或更短的时间范围是相较于另一表达相同蛋白或功能RNA的RNA分子而言的。当描述“调控”蛋白表达时，则指调控包含所述蛋白编码序列的RNA分子表达。这里描述的调控作用可以通过将本申请的Poly (A)尾连接在不包含Poly (A)尾的RNA分子的3'末端实现，也可以通过将RNA原有的

Poly (A) 尾替换为本申请的Poly (A) 尾实现。

如本文所用,“同一性”的百分比,例如85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、98.5%、99%、99.5%同一性,是指氨基酸序列之间或核苷酸序列之间,通过序列比对确定的相似程度,是85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、98.5%、99%、99.5%。例如,通过引入空位等方式可以使两条序列在尽可能多的位置上具有相同残基后,确定的具有相同碱基或氨基酸残基的位置数量占位置总数的比例。“同一性”的百分比可以用本领域已知的软件程序来确定。优选的是使用默认参数进行比对。一个优选的比对程序是BLAST。优选的程序是BLASTN和BLASTP。这些程序的细节可以在以下互联网地址找到: ncbi.nlm.nih.gov/cgi-bin/BLAST。

如本文所用,核酸的“互补”是指一条核酸通过传统的Watson-Crick碱基配对与另一条核酸形成氢键的能力。百分比互补性表示核酸分子中可与另一核酸分子形成氢键(即, Watson-Crick碱基配对)的残基的百分比(例如, 10个中的约5、6、7、8、9、10 个分别为约50%, 60%, 70%, 80%, 90%和100%互补)。“完全互补”是指核酸序列的所有连续残基与第二核酸序列中相同数量的连续残基形成氢键。如本文所用,“基本上互补”是指在约40、50、60、70、80、100、150、200、250或更多个核苷酸的区域内,至少约70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%或100%中的任何一个的互补程度,或指在严格条件下杂交的两条核酸。对于单个碱基或单个核苷酸,按照Watson-Crick碱基配对原则, A与T或U、C与G或I配对时,被称为互补或匹配,反之亦然;而除此以外的碱基配对都称为不互补。本申请中某多核苷酸序列的“互补多核苷酸序列”则是指与该某多核苷酸序列完全互补的多核苷酸序列。

如本文所用,某个蛋白、多肽或氨基酸序列的“保守取代变体”是指其中一个或多个氨基酸残基经过氨基酸取代而不改变蛋白质或酶的整体构象和功能,这包括但不限于以前述“保守取代”描述的方式取代亲本蛋白质中氨基酸序列中的氨基酸。因此,相似功能的两个蛋白或氨基酸序列的相似性可能会不同。例如,基于MEGALIGN算法的70%至99%的相似度(同一性)。“保守取代变体”还包括通过BLAST或FASTA算法确定具有60%以上的氨基酸同一性的多肽或酶,若能达75%以上更好,最好能达85%以上,甚至达90%以

上为最佳，并且与天然或亲本蛋白质或酶相比具有相同或基本相似的性质或功能。

在本申请的上下文中，术语“DNA”和“RNA”是指单链或双链DNA或RNA分子。除非另有说明，否则术语“DNA”和“DNA分子”是指由A、C、G和/或T核苷酸构成的双链DNA分子，而术语“RNA”和“RNA分子”是指由A、C、G和/或U核苷酸构成的单链RNA分子。在本文中，所述A、C、G、T和U核苷酸是指包含腺嘌呤、鸟嘌呤、胞嘧啶、胸腺嘧啶和尿嘧啶作为各自的含氮碱基的核苷酸。

RNA分子包括编码RNA(coding RNA)或非编码RNA(non-coding RNA, ncRNA)，例如Pre-mRNA、成熟mRNA或长链非编码RNA(long noncoding RNA, lncRNA)。

如本文所用，所述“DNA与RNA的杂合分子”是一个包含由脱氧核糖核苷酸和核糖核苷酸组成的多核苷酸序列的分子。所述DNA与RNA的杂合分子可以通过以下方式获得：

将DNA中的一个或多个脱氧核糖核苷酸替换为核糖核苷酸；

将RNA中的一个或多个核糖核苷酸替换为脱氧核糖核苷酸；或

通过生物或化学合成的方式，以脱氧核糖核苷酸以及核糖核苷酸为原料从头合成。需注意，获得DNA与RNA的杂合分子的方式不限于以上方式，由任何方式获取的DNA与RNA的杂合分子均属于本申请定义的“DNA与RNA的杂合分子”的范畴。

如本文所用，如果描述两个核酸分子具有“相同的遗传信息”，则是指所述两个核酸分子互补，或包含完全相同的碱基序列，或其中一个核酸分子的碱基序列中一个或多个胸腺嘧啶转变为尿嘧啶后可以获得与另一个核酸分子碱基序列完全相同的核酸分子。因此，DNA、RNA以及DNA与RNA的杂合分子中的任意两种均可具有相同的遗传信息。其中，术语“碱基序列”指多核苷酸分子中碱基的排列顺序。本领域技术人员应当知晓，除非另有说明，否则本申请中所述的碱基序列或多核苷酸序列可在用于描述DNA序列时，以“T”代指胸腺嘧啶，但在所述碱基序列或多核苷酸序列用于描述RNA(例如mRNA)时，“T”将由“U”(尿嘧啶)取代。因此，由本文中的特定序列号(SEQ ID NO)公开的任何DNA也公开与所述DNA互补或对应的RNA(例如mRNA或Poly(A)尾)序列，其中所述DNA序列的每个“T”被“U”取代。

Poly (A) 尾及其用途

如本文所用,术语 "PolyA尾"或"Poly (A) 序列"是指通常位于RNA分子3'-末端的不间断或不中断的腺苷酸残基序列。在RNA中,在有3'-UTR存在的情况下,Poly-A序列与3'-UTR的3'端相连。不间断的poly-A尾的特点是有连续的腺苷酸残基。Poly-A尾可以是任何长度的。在一些实施方案中,Poly-A尾包含,或由至少20、至少30、至少40、至少80或至少100和至多500、至多400、至多300、至多200或至多150个腺苷酸(A)组成,特别是约120个A。通常,PolyA尾中的绝大多数核苷酸都是腺苷,所述绝大多数是指至少75%,至少80%,至少85%,至少90%的核苷酸等,但允许剩余的核苷酸是A以外的核苷酸(非A核苷酸),例如U(尿苷酸)、G(鸟苷酸)或C(胞苷酸)。

在一些实施方案中,所述RNA的体外制备过程是原核发酵过程,即将包含所述Poly (A) 尾的RNA分子的编码核酸导入原核细胞中,通过扩增所述原核细胞,达到扩增所述编码核酸的目的,随后,将扩增后的所述编码核酸转录为所述RNA。在一些实施方案中,所述RNA的体外制备过程是将包含蛋白质编码序列的RNA片段与Poly (A) 尾通过同源重组、酶切连接、或其他非同源重组等方式连接,且所述Poly (A) 尾通过原核发酵过程制备,所述原核发酵过程时将包含所述Poly (A) 尾的编码核酸导入原核细胞中,通过扩增所述原核细胞,达到扩增所述编码核酸的目的,随后,将扩增后的所述编码核酸转录为包含所述Poly (A) 尾的RNA。在一些实施方案中,前述编码核酸是线性的。在一些实施方案中,前述编码核酸是环状的。在一些实施方案中,前述编码核酸是质粒。在一些实施方案中,前述编码核酸是单链或双链。在一些实施方案中,前述编码核酸在导入原核细胞前经过了化学修饰。在一些实施方案中,导入原核细胞前的前述编码核酸通过化学合成。在一些实施方案中,所述编码核酸插入所述原核细胞的类核/拟核基因组DNA中。在一些实施方案中,所述编码核酸游离存在于所述原核细胞的细胞质中或类核/拟核以外。在一些实施方案中,所述原核细胞是大肠杆菌。

基于此,本申请提供了一系列的Poly (A) 尾,所述Poly (A) 尾在RNA的体外制备过程中保守性高。所述Poly (A) 尾在一个或多个位置包含一个或多个非A核苷酸。

在一些实施方案中,所述Poly (A) 尾编码序列包含:

唯一的元件a和至少一个元件b,以及至少一个元件c;

唯一的元件a和至少一个元件b, 以及至少一个元件d; 或

唯一的元件a和至少一个元件b, 以及至少一个元件c和至少一个元件d,
其中所述元件a和元件b不相邻、元件c和元件d不相邻,

且所述Poly (A)尾编码序列不包含彼此相邻的元件b、不包含彼此相邻的元件c、且不包含彼此相邻的元件d。

在一些实施方案中, 所述Poly (A)尾编码序列还进一步包含唯一一个元件e, 所述元件e由一个或两个连续的A组成, 且其位于所述Poly (A)尾编码序列的3'末端, 且与元件d或元件c相邻。

这里的Poly (A)尾可以是一段RNA, 也可以是DNA与RNA的杂合分子。

本申请还提供了可调控蛋白表达水平的Poly (A)尾。并且, 本申请还提供了在体外制备过程中保守性高的同时, 可调控蛋白表达水平的Poly (A)尾。在一些实施方案中, 所述可调控蛋白表达水平的Poly (A)尾; 或所述在体外制备过程中保守性高的同时, 可调控蛋白表达水平的Poly (A)尾的结构选自:

元件a-元件c-元件b-元件c-元件b-元件c-元件b-元件c-元件b;

元件b-元件c-元件b-元件c-元件a-元件d-元件b-元件c-元件b-元件c-元件b;

元件b-元件c-元件b-元件c-元件b-元件d-元件a-元件c;

元件a-元件d-元件b-元件c-元件b-元件c-元件b; 及

元件b-元件c-元件b-元件c-元件b-元件d-元件a。

在一些实施方案中, 所述可调控蛋白表达水平的poly (A)尾; 或所述在体外制备过程中保守性高的同时, 可调控蛋白表达水平的poly (A)尾的结构选自:

60A-G-19A-G-19A-G-19A-G-3A;

7A-C-18A-G-60A-元件d-7A-C-18A-G-14A;

60A-元件d-19A-G-19A-G-17A;

19A-G-19A-G-19A-元件d-60A;

19A-G-19A-G-19A-元件d-60A-G;

19A-G-19A-G-19A-元件d-60A;

19A-C-19A-C-19A-元件d-60A; 及

19A-T-19A-T-19A-元件d-60A。

在一些实施方案中, 所述可调控蛋白表达水平的Poly (A)尾; 或所述在

体外制备过程中保守性高的同时，可调控蛋白表达水平的Poly (A) 尾的结构为：60A-元件d-19A-G-19A-G-17A或19A-G-19A-G-19A-元件d-60A。

其中上述“-”连接的两个元件之间直接相连，及所述两个元件之间不包含任何核苷酸。

上述Poly (A) 尾的结构中以“yA”表示元件a或元件b中连续A的个数，所述y为自然数，例如19A，则表示包含19个连续的A；60A则表示包含60个连续的A。

在一些实施方案中，所述可调控蛋白表达水平的Poly (A) 尾；或所述在体外制备过程中保守性高的同时，可调控蛋白表达水平的Poly (A) 尾的多核苷酸编码序列选自如SEQ ID NO: 1-10中任一项所示的多核苷酸序列。

此外，本申请也提供了将上述Poly (A)尾用于调控蛋白表达的用途，在所述用途中，所述Poly (A) 尾位于所述mRNA的3'端，例如3'UTR的3'端。在一些实施方案中，所述调控蛋白表达使用后文中的调控蛋白表达的方法。

工程化DNA分子及文库

本申请还提供了一种可在细胞中复制的工程化DNA分子，其包含前述Poly A尾的编码序列或其互补序列。本领域技术人员应当知晓，除所述Poly A尾的编码序列以外，所述工程化DNA分子还应包含可使所述DNA分子在细胞中复制，或高效复制所必须的结构元件。使所述工程化DNA分子在细胞中复制或高效复制所述必须的结构元件是本领域已知的，包括例如复制起点(ORI)。在一些实施方案中，所述工程化DNA分子还进一步包含标记基因或其片段和/或报道基因或其片段、和允许插入DNA元件的独特的限制性内切酶位点，优选多克隆位点(MCS)形式的限制性内切酶位点。所述标记基因有利于鉴定含有包含所述标记基因的质粒的细胞，可选自，例如抗生素抗性基因。所述MCS中的每一个限制性内切酶位点均可被不同的限制性内切酶特异性识别。

在一些实施方案中，所述DNA分子是DNA质粒。如本文所用，术语“DNA质粒”是指由双链DNA分子组成的质粒。在一些实施方案中，所述“质粒”是环状DNA分子。在一些实施方案中，所述“质粒”还可以涵盖线性DNA分子。具体的，术语“质粒”还涵盖通过例如用限制性内切酶切割环状质粒，进而使该环状质粒分子转变成线性分子而使该环状质粒线性化所得到的分子，以及可在原核生物中复制的线性分子。质粒可以复制，即在细胞中独立于原核细

胞拟核或类核存储的基因组遗传信息而扩增，并且可以用于克隆，即用于在细菌细胞中扩增遗传信息。优选地，根据本申请的DNA质粒是中拷贝或高拷贝质粒，更优选地是高拷贝质粒。此类高拷贝质粒的实例是这样的载体：其基于pUC、pTZ质粒或包含支持质粒高拷贝的ORI的任意其它质粒(例如pMB1、pCoIE1)等。

在一些实施方案中，所述DNA分子是构成原核生物拟核或类核的DNA分子或其片段，即所述包含前述Poly (A) 尾的编码序列或其互补序列可随原核生物基因组进行复制。

在一些实施方案中，所述DNA分子在所述Poly (A)尾编码序列的5'端一侧进一步连接目的基因片段，所述目的基因片段与所述Poly (A)尾编码序列共同编码RNA。在一些实施方案中，所述目的基因片段与所述Poly (A)尾编码序列共同编码mRNA。所述目的基因片段包含蛋白、多肽或其片段的编码序列。在一些实施方案中，所述目的基因片段还包含转录后可用于启动或调控所述蛋白、多肽或其片段表达的元件的编码序列，所述元件包括但不限于5'UTR、3'UTR等。在一些实施方案中，所述目的基因片段包含至少一个非翻译区(UTR)的编码序列。在一些实施方案中，所述目的基因片段包含至少5'UTR的编码序列和所述蛋白、多肽或其片段的编码序列。在一些实施方案中，所述目的基因片段从5'至3'至少依次包含5'UTR的编码序列，蛋白、多肽或其片段的编码序列以及3'UTR的编码序列。所述蛋白、多肽或其片段的编码序列可被最终翻译为一个或多个蛋白、或一条或多条多肽，例如短肽、寡肽、多肽、融合蛋白、蛋白质及其片段，如已知蛋白质的部分，例如功能性部分。所述功能性部分可以是例如，蛋白质的生物活性部分或可以有效地产生抗体的抗原部分，例如抗原表位。所述蛋白、多肽或其片段的编码序列的两端分别包含起始密码子(5'端)和终止密码子(3'端)，其分别是所述mRNA分子的可被翻译的前三个核苷酸和后三个核苷酸。5'UTR通常包含至少一个核糖体结合位点(RBS)，如原核生物中的Shine-Dalgarno序列，或至少一个翻译起始位点，如真核生物中的Kozak序列。RBS通过在翻译起始时募集核糖体来促进mRNA分子的有效且准确的翻译。可以通过改变给定的RBS或翻译启示位点的长度和序列以及距起始密码子的距离来优化其活性。可选地或任选地，5'UTR包括内部核糖体进入位点或IRES。3'UTR可包含一个或多个调控序列，如增强mRNA分子稳定性的氨基酸序列的结合位点、调控

RNA分子(如miRNA分子)的结合位点、和/或参与mRNA分子的胞内运输的信号序列。

在前述实施方案的基础上,在一些实施方式中,所述目的基因片段还包含一个或多个另外的调控序列,如增强mRNA分子稳定性的氨基酸序列的结合位点、增强mRNA分子翻译的氨基酸序列的结合位点、调节元件(如核糖开关)、调节RNA分子(如miRNA分子)的结合位点、和/或对翻译起始产生积极影响的核苷酸序列。此外,在5'UTR内,优选地不存在功能性的上游开放阅读框、框外上游翻译起始位点、框外上游起始密码子、和/或产生减少或防止翻译的二级结构的核苷酸序列。5'UTR中此类核苷酸序列的存在可对翻译产生负面影响。

所述蛋白、多肽或其片段的编码序列包含可以翻译成氨基酸序列的密码子。所述编码序列包含的全部密码子中,可以全部是天然存在的编码氨基酸的密码子,也可以有部分或全部由人工合成的密码子组成。在一些实施方案中,所述部分或全部密码子经过了密码子优化。在一些实施方案中,所述部分或全部密码子编码非天然氨基酸。

在一些实施方案中,所述DNA分子在所述目的基因片段的5'端一侧还进一步包含可启动或调控所述RNA转录所必须的结构元件,所述结构元件是本领域已知的。在一些实施方案中,所述结构元件至少包含启动子。启动子及其序列是本领域已知的,包括弱启动子、中等强度启动子、强启动子、mini启动子或核心启动子等。在一些特定的实施方案中,所述启动子为强启动子。在一些实施方案中,所述启动子可在原核细胞中启动所述目的基因片段和/或Poly(A)尾转录。在一些实施方案中,所述启动子可以在真核细胞中启动所述目的基因片段和/或Poly(A)尾转录。所述“启动子”包含至少一个转录识别位点及其后的转录因子结合位点。所述识别和结合位点可以与介导或调节转录的氨基酸序列相互作用。与识别位点相比,结合位点更靠近前述目的基因片段。结合位点可以是,例如原核生物中的Pribnow框或真核生物中的TATA框。例如,在一些实施方案中,当使用Pribnow框时,所述转录识别位点可以位于转录起始位点上游约35bp处,而转录因子结合位点可以位于转录起始位点上游约10bp处。在一些实施方案中,所述启动子包含至少一个另外的调控元件,如位于转录起始位点之前约40和/或60个核苷酸处富含AT的上游元件,和/或位于识别位点和结合位点之间的增强启动子活性的另外的调控元件。在

一些实施方式中，所述启动子是强启动子，即所述启动子包含促进前述RNA编码序列转录的序列。强启动子是本领域技术人员已知的，例如来自大肠杆菌的RecA启动子衍生的OXB18、OXB19和OXB20启动子，或者可以通过常规实验室程序鉴定或合成。在一些实施方案中，所述启动子为T7启动子。在一些实施方案中，所述启动子前还包含另外的调控元件，如包含在DNA质粒中可促进前述RNA编码序列转录的增强子。

本申请还提供了包含前述工程化DNA分子的文库。在一些实施方案中，所述文库包含至少两条具有不同Poly (A) 尾编码序列的DNA分子。

此外，本申请还提供了上述工程化DNA分子在稳定扩增Poly (A) 尾编码序列或带有Poly (A) 尾的RNA的编码序列中的用途。在一些实施方案中，所述扩增Poly (A) 尾编码序列或带有Poly (A) 尾的RNA的编码序列的方法如后文在体外稳定扩增Poly (A) 尾转录模板DNA的方法所示。

工程化RNA及文库

本申请提供了一种RNA，其包含前述Poly (A) 尾，以及Poly (A) 尾编码序列5'端一侧的目的基因片段。在一些实施方案中，所述RNA还包含5'-帽结构。在一些实施方案中，所述RNA为mRNA。

如本文所用“mRNA”(信使RNA)是编码至少一种蛋白质、多肽或其片段的，天然存在、非天然存在或经修饰的任何RNA，所述mRNA具备经翻译以在体外、体内、原位或离体产生所编码的蛋白质、多肽或其片段的能力。因此，所述mRNA可以是成熟的mRNA或成熟前的mRNA，其必须包含或选择性包含的元件或结构是本领域已知的。在一些实施方案中，所述mRNA包含多个必要的功能组件的编码序列，以表达、调节、或增强所述蛋白、多肽或其片段的表达水平。所述功能组件包括但不限于5'帽子、5'UTR、3'UTR等。5'UTR和3'UTR两者通常都从基因组DNA转录，并且是成熟前的mRNA就具有的元件。作为成熟mRNA，

术语“5'帽子”位于mRNA的5'最末端，包含甲基化鸟苷酸，所述甲基化鸟苷酸经焦磷酸连接于mRNA的5'末端，与其相邻的核苷酸形成5',5'-三磷酸连接。5'帽子结构通常有三种类型(m⁷G5'ppp5'Np、m⁷G5'ppp5'NmpNp、m⁷G5'ppp5'NmpNmpNp)，分别称为O型、I型和II型。O型指末端核苷酸的核糖未甲基化，I型指末端一个核苷酸的核糖甲基化，II型指末端两个核苷酸的核糖均甲基化。在一些实施方案中，所述5'帽子可根据制造商的方案，使用

以下化学RNA帽类似物在体外转录反应期间同时完成对多核苷酸的5'加帽以产生5'-鸟苷帽结构：3'-O-Me-m7G(5')ppp(5')G[ARCA帽]、G(5')ppp(5')A、G(5')ppp(5')G、m7G(5')ppp(5')A、m7G(5')ppp(5')G(NewEnglandBioLabs、Ipswich、MA)、或m7G(5')ppp(5')(2'-OMeA)pG (CleanCapAG)。例如在一些实施方案中，可使用牛痘病毒加帽酶在转录后完成对经修饰RNA的5'加帽，以产生O型帽子结构：m7G(5')ppp(5')G(New England BioLabs, Ipswich, MA)。可使用牛痘病毒加帽酶和2'-O甲基-转移酶两者来产生I型帽子结构，以产生m7G(5')ppp(5')(2'-OMeA)pG。可从I型帽子结构接着使用2'-O甲基-转移酶对5'-倒数第三核苷酸进行2'-O-甲基化来产生II型帽子结构。可从II型帽子结构接着使用2'-O甲基-转移酶对5'-倒数第四核苷酸进行2'-O-甲基化来产生III型帽子结构。

在一些实施方案中，所述mRNA的部分或全部尿苷为化学修饰的尿苷。

在一些实施方案中，所述mRNA的部分或全部尿苷为假尿苷或1-甲基-假尿苷。

在一些实施方案中，所述mRNA的部分或全部尿嘧啶核苷酸被假尿苷(Ψ)核苷酸或N1-甲基假尿苷(m1 Ψ)核苷酸取代。

在一些实施方案中，所述mRNA还包含稳定化元件。稳定化元件可包括例如组蛋白茎环。在一些实施方案中，所述mRNA包含编码区、至少一个组蛋白茎环和任选地poly(A)序列或多聚腺苷酸化信号。所述poly(A)序列或多聚腺苷酸化信号通常应增强所编码蛋白质的表达水平。在一些实施方案中，所述mRNA包含poly(A)序列或多聚腺苷酸化信号与至少一个组蛋白茎环的组合，尽管两者在自然界中具有替代机制，但其协同作用可使蛋白质表达增加至超过任一单个元件所观察到的水平。poly(A)与至少一个组蛋白茎环的组合的协同效应不依赖于元件的次序或poly(A)序列的长度。在一些实施方案中，所述组蛋白茎环通常源自组蛋白基因，并且包括由间隔区(由短序列组成)分隔的两个相邻部分或完全反向互补序列的分子内碱基配对形成环。未配对的环区通常无法与茎环元件中的任一者进行碱基配对。茎环结构的稳定性通常取决于长度、错配或膨出部的数目以及配对区的碱基组成。在一些实施方案中，可产生摆动碱基配对(非Watson-Crick碱基配对)。在一些实施方案中，所述至少一个组蛋白茎环序列包含15至45个核苷酸长度。

在一些实施方案中，可将所述mRNA的一个或多个富含AU的序列去除。

这些序列有时称为AURES，其是在3'UTR中发现的去稳定化序列。可将AURES从mRNA中去除。或者，可使AURES保留在mRNA中。

在一些实施方案中，所述mRNA被配置于脂质纳米颗粒(LNP)内。在一些实施方案中，脂质于所述mRNA混合形成脂质纳米颗粒。在一些实施方案中，将RNA配制在脂质纳米颗粒中。在一些实施方案中，所述脂质纳米颗粒首先形成空的脂质纳米颗粒，并且在即将施用之前(例如在几分钟至一小时内)与疫苗的mRNA组合或包裹。

所述脂质纳米颗粒通常包含可电离的脂质、非阳离子脂质、固醇和PEG脂质组分以及目标核酸，例如上述mRNA。可使用如本领域中通常已知的组分、组合物和方法来产生本公开的脂质纳米颗粒，参见例如PCT/US2016/052352、PCT/US2016/068300、PCT/US2017/037551、PCT/US2015/027400、PCT/US2016/047406、PCT/US2016000129、PCT/US2016/014280、PCT/US2016/014280、PCT/US2017/038426、PCT/US2014/027077、PCT/US2014/055394、PCT/US2016/52117、PCT/US2012/069610、PCT/US2017/027492、PCT/US2016/059575和PCT/US2016/069491，其全部通过引用整体并入本文。

本申请还提供了包含前述mRNA分子的文库。所述文库中包含至少两条具有不同Poly(A)尾的mRNA分子。

本申请还提供了所述mRNA及所述mRNA文库的用途。使用对mRNA的表达量具有不同影响力梯度的Poly(A)尾的至少两种及以上的mRNA分子，可用于调节前述蛋白质、多肽或其片段的编码序列的表达量。例如通过调节具有所述两种及以上的mRNA分子的文库中不同mRNA分子的比例，或通过在不同时间导入具有相同或不同含量的所述两种及以上的mRNA分子中的一种或多种。

细胞

本申请还提供了包含前述工程化DNA分子的细胞，所述DNA分子可在所述细胞中储存和/或扩增。在一些实施方案中，所述细胞为原核细胞，所述DNA分子可在所述原核细胞中复制。在一些实施方案中，所述细胞为原核细胞，所述DNA分子可在所述原核细胞中复制和/或转录。在一些实施方案中，所述DNA分子为真核细胞，所述DNA分子可在所述细胞中复制。在一些实施方案中，所述DNA分子可在所述DNA细胞中转录和/或复制。

在一些实施方案中，所述细胞为原核细胞。在一些实施方案中，所述细胞为细菌、放线菌、蓝细菌、支原体、立克次氏体和衣原体。在一些实施方案中，所述细胞选自：枯草芽孢杆菌、乳酸杆菌、醋酸杆菌、棒状杆菌、短杆菌、节杆菌、假单胞菌、及小球菌。在一些实施方案中，所述细胞为recA⁻的细菌。在一些实施方案中，所述细胞为大肠杆菌。在一些实施方案中，所述细胞为大肠杆菌，所述大肠杆菌选自：K-12及其衍生菌株，以及B菌株及其衍生菌株。在一些实施方案中，所述大肠杆菌选自：MG1655、DH5或DH5 α 、DH10B、BL21、DB3.1、HB101、JM109、JM110、MC1061、MG1655、Pir1、Stbl2、Stbl3、Top10、XL1Blue、XL10 Gold、BLR、HMS174、Tuner、Rostetta2、Lemo21、T7Express、Origami2等。在一些实施方案中，所述细胞选自链霉菌属、小单孢菌属和诺卡氏菌属。在一些实施方案中，所述细胞为真菌。在一些实施方案中，所述细胞选自酵母菌或霉菌。

方法

本申请提供了在体外稳定扩增Poly (A) 尾转录模板DNA的方法，以减少所述DNA在细胞中大量复制时，其中Poly (A) 尾转录模板序列的突变频率。所述方法，包括：扩增上述包含上述工程化DNA分子的细胞。

在一些实施方案中，在扩增所述细胞之前，还包含将所述工程化DNA分子引入所述细胞。在一些实施方案中，所述引入可以包含化学转化或电转化。在一些实施方案中，所述引入是所述细胞对所述工程化DNA分子的自然胞吞过程。

在一些实施方案中，在扩增所述细胞之后，还包含提取细胞DNA并通过体外转录合成上述RNA。在一些实施方案中，在扩增所述细胞后，还包含诱导所述RNA在所述细胞中转录，然后提取并分离其中的RNA。在一些实施方案中，还包含提取细胞DNA并将其转导入可转录所述RNA的第二细胞中。在一些实施方案中，所述转导包括向人体施用，所述施用选自：通过静脉内，腹膜内，皮下，颅内，鞘内，动脉内(例如经由颈动脉)，肌内及瘤内注射或灌注。

此外，本申请还提供了调控蛋白表达的方法，包括：

将前述工程化DNA分子中的两种或多种在不同时间，和/或以不同的数量比例导入目的细胞；或

将前述RNA分子中的两种或多种在不同时间，和/或以不同的数量比例

导入目的细胞。

其中，所述两种或多种的前述工程化DNA分子，以及两种或多种的前述RNA分子具有不同的Poly(A)尾，所述Poly(A)尾对RNA的表达量具有不同影响力梯度。

在一些实施方案中，本申请还提供了调控蛋白表达的方法，包括将前述工程化DNA分子或前述RNA分子导入目的细胞。在一些实施方案中，所述DNA编码的Poly(A)尾及所述RNA包含的Poly(A)尾的编码序列包含选自下述的结构：

元件a-元件c-元件b-元件c-元件b-元件c-元件b-元件c-元件b；

元件b-元件c-元件b-元件c-元件a-元件d-元件b-元件c-元件b-元件c-元件b；

元件b-元件c-元件b-元件c-元件b-元件d-元件a-元件c；

元件a-元件d-元件b-元件c-元件b-元件c-元件b；或

元件b-元件c-元件b-元件c-元件b-元件d-元件a。

在一些实施方案中，所述DNA编码的Poly(A)尾及所述RNA包含的Poly(A)尾的编码序列包含选自下述的结构：

60A-G-19A-G-19A-G-19A-G-3A；

7A-C-18A-G-60A-元件d-7A-C-18A-G-14A；

60A-元件d-19A-G-19A-G-17A；

19A-G-19A-G-19A-元件d-60A；

19A-G-19A-G-19A-元件d-60A-G；

19A-G-19A-G-19A-元件d-60A；

19A-C-19A-C-19A-元件d-60A；及

19A-T-19A-T-19A-元件d-60A。

在一些实施方案中，所述DNA编码的Poly(A)尾及所述RNA包含的Poly(A)尾的编码序列包含的结构为：60A-元件d-19A-G-19A-G-17A或19A-G-19A-G-19A-元件d-60A。

其中上述“-”连接的两个元件之间直接相连，及所述两个元件之间不包含任何核苷酸。

上述Poly(A)尾的结构中以“yA”表示元件a或元件b中连续A的个数，所述y为自然数，例如19A，则表示包含19个连续的A；60A则表示包含60个连续的A。

在一些实施方案中，所述DNA编码的Poly(A)尾及所述RNA包含的Poly(A)尾的编码序列包含选自下述的多核苷酸序列，或由选自下述的任一个多核苷酸序列组成：SEQ ID NO：1-10。

应当理解，本申请包含本文所描述的各种方面、实施方案以及所述方面和/或实施方案的组合。以上描述以及随后的实施例旨在说明而不是限制本申请的范围。在本申请范围内的其他方面、改进和修改对于本申请所属领域的技术人员将是显而易见的。因此，本领域的普通技术人员应该认识到，本申请的范围还包括对所述方面和实施方案的所述改进和修改。

实施例

实施例1：Poly(A)尾的构建

使用常规基因工程方法，构建下表1所示的Poly(A)尾及编码所述Poly(A)尾的DNA序列。

表1.

设计	具体结构	序列	长度
P1	60A-G-19A-G-19A-G-19A-G-AAA	SEQ ID NO： 1	124nt
P2	7A-C-18A-G-60A-元件d-7A-C-18A-G-14A	SEQ ID NO： 2	130nt
P3	60A-元件d-19A-G-19A-G-17A	SEQ ID NO： 3	123nt
P4	19A-G-19A-G-19A-元件d-60A	SEQ ID NO： 4	125nt
P5	19A-G-19A-G-19A-元件d-60A-G	SEQ ID NO： 5	126nt
P6	19A-G-19A-G-19A-元件d-60A	SEQ ID NO： 6	125nt
P7	19A-G-19A-G-19A-元件d-60A	SEQ ID NO： 7	131nt
P8	19A-C-19A-C-19A-元件d-60A	SEQ ID NO： 8	125nt

P9	19A-T-19A-T-19A-元件d-60A	SEQ ID NO: 9	125nt
P10	19A-G-19A-G-19A-元件d-60A	SEQ ID NO: 10	125nt
C1	A60-元件d-A60	SEQ ID NO: 11	130nt
C2	A30-元件d-A70	SEQ ID NO: 12	110nt
C3	A60-元件c-A60	SEQ ID NO: 13	121nt
C4	A60-元件d-A60	SEQ ID NO: 14	126nt

实施例2：以荧光素酶编码序列为例测试Poly（A）功能

2.1 编码mRNA的DNA分子在原核细胞中的复制稳定性测试

以荧光素酶为蛋白编码区，考察不同poly(A)变体的在大肠杆菌中的稳定性和对荧光素酶在细胞内表达的影响

1) 构建包含荧光素酶蛋白编码区的通用载体

该通用载体以大肠杆菌克隆载体pUC57作为载体骨架，在其多克隆位点的Xba I酶切位点和EcoR I酶切位点之间按顺序排列T7启动子序列(5'-TAATACGACTCACTATAAGG-3')、5' UTR、荧光素酶蛋白、3' UTR以及多聚腺苷酸串poly (dA: dT)。

2) 将通用载体中的多聚腺苷酸串poly (dA: dT)分别替换为本申请的P1-P10以及对照C1-C4（A60-10nt spacer-A60（对照C1）、A30-10nt spacer-A70（对照C2）、A60-1nt spacer-A60（对照C3）或A60-6nt spacer-A60（对照C4）C1 C2来源于文献专利（美国专利号US 10717982B2））。

合成构建P1-P10以及对照C1-C4所需的全部引物，用2个限制性内切酶进行双酶切以便从1)中构建的通用载体中切除poly (dA: dT)，之后用T4 DNA连接酶1将P1-P10以及C1-C4与切除poly (dA: dT)的载体连接，完成了对通用载体中poly (dA: dT)的替换。

3) 检测不同poly(A)变体在大肠杆菌中的复制稳定性

步骤2)中构建的载体质粒经测序确认无误后,转入大肠杆菌DH5 α ,转化后的平板在30°C生长,完成质粒提取和测序工作。测序完成后根据测序结果分析和计算不同poly (A)变体的稳定性和碱基缺失情况;复制稳定性用未发生任何碱基变化的克隆数的占比来表示,占比越高说明质粒在大肠杆菌中的复制稳定性越高。

结果表明(见图1,图2),具体实验结果如下:

对照C1共测100个克隆,15个克隆发生碱基缺失,占比15%;正确克隆数85个,占比85%。

对照C2共测50个克隆,所有克隆均正确,没有任何碱基变化或缺失,正确克隆占比为100%。

对照C3共测50个克隆,9个克隆发生碱基缺失,占比18%;正确克隆数41个,占比82%。

对照C4共测50个克隆,14个克隆发生碱基缺失,占比28%;正确克隆数36个,占比72%。

P1共测100个克隆,9个克隆发生碱基缺失,占比9%;正确克隆数91个,占比91%。

P2共测100个克隆,12个克隆发生碱基缺失,占比12%;正确克隆数88个,占比88%。

P3共测62个克隆,5个克隆发生碱基缺失,占比8%;正确克隆数57个,占比92%。

P4共测50个克隆,其中3个克隆发生碱基缺失,占比6%;正确克隆数47个,占比94%。

P5共测50个克隆,其中4个克隆发生碱基缺失,占比8%;正确克隆数46个,占比92%。

P6共测50个克隆,5个克隆发生碱基缺失,占比10%;正确克隆数45个,占比90%。

P7共测50个克隆,7个克隆发生碱基缺失,占比14%,正确克隆数43个,占比86%。

P8共测50个克隆,3个克隆发生碱基缺失,占比6%,正确克隆数47个,占比94%。

P9共测50个克隆,4个克隆发生碱基缺失,占比8%,正确克隆数46个,

占比92%。

P10共测50个克隆, 5个克隆发生碱基缺失, 占比10%, 正确克隆数45个, 占比90%。

综合正确克隆占比和缺失碱基数两个方面的结果, 本申请设计的poly(A)变体在大肠杆菌细胞内的复制稳定性方面优于现有技术或与现有技术相当。其中poly(A)变体P3、P4、P8的复制稳定性最高, P3、P4、P8的复制稳定性与C2相当, 没有统计学意义的差别($p>0.05$, χ^2 test), 并且P3、P4、P8的复制稳定性优于对照C1、C3、C4, 且差异在统计学上意义显著($p<0.05$, χ^2 test)。

实施例3 以HPV抗原蛋白编码序列为例测试poly(A)功能

1) 以HPV为蛋白编码区的质粒构建

如上, 实施例1中构建了荧光素酶编码基因组合不同poly(A)的载体。在这些载体的基础上, 通过常规分子克隆方法将荧光素酶编码基因替换为HPV编码基因, 主要元件按顺序排列T7启动子序列(5'-TAATACGACTCACTATAAGG-3')、5'UTR、HPV抗原蛋白编码序列、3'UTR以及poly(A)编码序列。

2) 小规模细菌培养, 检测HPV的4个poly(A)变体在大肠杆菌中的稳定性

步骤1)中构建的含有P1、P2、P3和P4的HPV载体质粒经测序确认无误后, 转入大肠杆菌DH5 α , 转化后的平板在30℃生长, 完成质粒提取和测序工作。测序完成后根据测序结果分析和计算不同poly(A)变体的稳定性和碱基缺失情况; 稳定性用未发生任何碱基变化的克隆数的占比来表示, 占比越高越稳定。

结果表明(图3, 图4), 当目的基因更换为HPV抗原编码序列后,

对照C1共测88个克隆, 47个克隆发生碱基缺失, 占比53%, 正确克隆数41个, 占比47%。

对照C2共测50个克隆, 1个克隆发生碱基缺失, 占比2%, 正确克隆数49个, 占比98%。

P1共测101个克隆, 14个克隆发生碱基缺失, 占比14%; 正确克隆数87个, 占比86%。

P2共测100个克隆, 12个克隆发生碱基缺失, 占比12%, 正确克隆数88个, 占比88%。

P3共测70个克隆，4个克隆发生碱基缺失，占比6%，正确克隆数66个，占比94%。

P4共测50个克隆，9个克隆发生多个碱基缺失，占比18%；正确克隆数41个，占比82%。

综合发生突变的比例及平均碱基缺失数目，当蛋白编码基因由实施例1中的荧光素酶换成HPV抗原蛋白后，本申请不同的poly (A)变体仍然保持着高复制稳定性；其中P3与C2稳定性相当，没有统计学上的显著差异 ($p>0.05$, χ^2 test)，与本申请的其他组新变体相比，克隆稳定性最优。且C2发生大片段缺失的几率为 $1/50=2\%$ ，而P3发生大片段缺失的几率为 $1/70=1.4\%$ ，由于poly (A)大片段缺失会影响mRNA产品的体内表达和药效，因此，P3更符合产品需求。以上结果说明本申请设计的poly (A)变体在针对不同蛋白编码区的实施例中的应用具有普遍性。

如上所述，实施例1和实施例2中，大肠杆菌转化后的平板在30℃生化培养箱中培养过夜，得到的克隆进行测序检测以评估复制稳定性。此外，由于大肠杆菌培养温度影响DNA复制速度，也会影响复制稳定性。针对实施例3，本申请还比较了大肠杆菌转化后的平板在37℃生化培养箱中培养过后进行测序检测。结果表明(图5)，37℃培养大大增加了对照C1的碱基缺失比例，由30℃培养时的53%上升至98%。与对照不同，P1、P2和P3在两个温度条件下的突变率无明显差异，说明本申请设计的poly (A)变体在针对大肠杆菌不同温度培养条件的实施例中仍然具有高复制稳定性，其应用具有普遍性。

3) 检测3个poly(A)变体在大规模发酵培养及不同代数之间的稳定性

mRNA药物的生产必须依赖大规模发酵来制备足量的模板质粒，而发酵中质粒的稳定性（本申请中，质粒的稳定性特指poly dA: dT的稳定性）对生产质量均一的mRNA药物至关重要。另一方面，为满足不同批次生产的稳定性需求，需要建立含有目的质粒的菌种库，包括一级库、二级库等不同代数的菌种库，因此需要评估不同代数大肠杆菌中的质粒稳定性。针对上述两个方面的问题，本申请在实施例3中检测了发酵过程的不同代数之间P1、P3的稳定性，根据测序结果，每个poly(A)变体选择4个正确的大肠杆菌克隆分别通过发酵进行传代培养。结果表明，在种子传代的第3、5、7、和第9代，P1、P3的4个克隆的质粒保持稳定，未发生任何碱基变化。

实施例4 荧光素酶mRNA在小鼠体内表达水平测试

真核生物细胞中，一定长度的poly (A)尾巴对保护mRNA 3'端、维持mRNA稳定性和促进蛋白表达至关重要。受体内生理或环境因素的影响，poly (A)逐渐变短，从而引发mRNA降解。本申请考察了不同poly (A)变体对蛋白表达水平的影响。我们采取小鼠体内表达检测用于评估不同polyA变体对荧光素酶活性的影响。本申请实施过程中，含有对照C2、以及P3、P4、P5、P8、P9的荧光素酶mRNA-LNP通过肌肉注射的方式进入小鼠体内，注射6 h后进行动物成像并对荧光强度进行定量比较不同poly (A)对荧光素酶体内活性的影响。具体实验过程如下：

通过体外转录合成荧光素酶mRNA：用II型限制性内切酶BspQ I酶切处理得到线性化DNA，线性化后的DNA 3'端是不同的poly(A)：对照C2、P3、P4、P5、P8或P9。线性化DNA作为模板进行体外转录，100 μ l反应体系中包含1X反应缓冲液，ATP、CTP、N1M-UTP、GTP各5mM(终浓度)，CleanCap AG 4mM(终浓度)，体外转录酶5 μ l。反应混合物充分混匀后，在37 $^{\circ}$ C反应3 h。体外转录合成的mRNA用LiCl沉淀法收集，最终用无酶水溶解。

动物实验：体外合成的荧光素酶mRNA包载成LNP，将得到的mRNA原液分别分散于20mM醋酸溶液(pH 5.0)，得到mRNA浓度为200 μ g/mL的RNA溶液。按照可电离脂：胆固醇：DSPC：DMG-PEG2000 = 50: 38.5: 10: 1.5的摩尔比进行混合成混脂。控制水相和油相流速通过T混流的方式，使mRNA与脂质混合物进行混合，启动注射泵，使mRNA溶液与脂质混合物混合形成LNP。然后用稀释液进行10倍稀释，经过超滤管离心浓缩后进行三次溶液置换。将上述得到的溶液，加入Tris水溶液调pH至7.0~8.0，得到LNP包载的mRNA溶液。LNP即脂质纳米颗粒。利用Ribogreen RNA定量试剂盒(Invitrogen, R11490)和达尔文ZetaSizer粒径仪分别测定LNP包载的mRNA浓度和粒径。4组分LNP中，各个组分的摩尔比为SM102：DSPC：胆固醇：DMG-PEG2000=50: 10: 38.5: 1.5。包载完成后通过测定粒径、包封率、PDI等指标对LNP进行质控，质控结果表明制备的LNP符合粒径范围50 nm-150 nm，PDI<0.3的标准，而且包封率>90%，能够用于后续实验。通过ribogreen法测定LNP中的mRNA含量，然后将LNP稀释到mRNA含量为100 ng/ μ L。

BALB/c小鼠按照体重随机分组，适应性饲养2-3天后进行给药，对照C2以及P3、P4、P5、P8、P9各注射5只小鼠，每只小鼠以肌肉注射的方式给药

100μl (10μg mRNA)，对照组用PBS注射5只小鼠。注射6 h后进行动物成像，并计算荧光值。结果表明（图7-8），与对照C2相比，P3的表达活性显著提高了1.8倍，具有统计学意义的显著性差异（ $p<0.01$, student's t-test），P4、P5、P8、P9与C2表达水平相当，无显著性差异($p>0.05$, student's t-test)。

本申请以上实施例中使用的序列示于如下序列表中。应当理解，以下序列仅为本申请实施方案的示例性序列，而非对本申请方案的任何限制。以下序列表中的核酸序列可表示DNA序列或RNA序列，当其表示RNA序列时，其中的“T”代表尿苷。

序列表：

SEQ ID NO.	名称	序列
核酸序列		
1	P1	AA AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAAA AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAA
2	P2	AAAAAACAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAA AAAG GAAAAAACAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAA
3	P3	AA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAGAA AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAA
4	P4	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAAAA AA
5	P5	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAAAA AAAG
6	P6	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAA AAAAAAAAAAAAAAAAAGTATACAAAAAAAAAAAAAAAAAAAA AA
7	P7	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAA AAAAAAAAAAAAAAAAAGATATCGTATACAAAAAAAAAAAAAA AA
8	P8	AAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAACAAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAAAA AA
9	P9	AAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAATAAAA AAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAAAA AA
10	P10	AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAA AAAAAAAAAAAAAAAAGAATCTAAAAAAAAAAAAAAAAAAAA AA
11	C1	AA AAAAAAAAAAAAAAAAAGCATATGACTAAAAAAAAAAAAAA AA
12	C2	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCATATGACTAAA AA AAAAAAAAAAAAAAAAAAAAAAAAAAAA
13	C3	AA AAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

14	C4	AA AAAAAAAAAAAAAAAAAAAAAAAAATGCATAAAAAAAAAAAAAAAAAA AA
15	元件d示例1	GATATC
16	元件d示例2	GTATAC
17	元件d示例3	GAATCT
18	元件d示例4	GCATATGACT
19	元件d示例5	GATATCGTATAC
20	3'UTR	CTCGCTTCTTGCTGTCCAATTTCTATTAAAGGTTCCCTTGTTCCTA AGTCCAACACTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATC TGGATTCTGCCTGCTCGCTTCTTGCTGTCCAATTTCTATTAAAGGTT CCTTGTTCCTAAGTCCAACACTACTAACTGGGGGATATTATGAAGG GCCTTGAGCATCTGGATTCTGCCT
21	5'UTR	ACATTTGCTTCTGACACAACCTGTGTTCCTAGCAACCTCAAACAGAC ACC
22	荧光素酶编 码基因序列	ATGGAAGACGCCAAAAACATTAAGAAGGGCCAGCGCCATTCTACC CACTCGAAGACGGGACCGCGGAGCAGCTGCACAAAGCCATGA AGCGCTACGCCCTGGTGCCCGGACCATCGCCTTTACCGACGCACA TATCGAGGTGGACATTACCTACGCCGAGTACTTCGAGATGAGCGTT CGGCTGGCAGAAAGCTATGAAGCGCTATGGGCTGAATACAAACCATC GGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCTTCATGCCCGT GTTGGGTGCCCTGTTTCATCGGTGTGGCTGTGGCCCCAGCTAACGAC ATCTACAACGAGCGCGAGCTGCTGAACAGCATGGGCATCAGCCAGC CCACCGTCGTATTCTGTGAGCAAGAAAGGGCTGCAAAAGATCCTCAA CGTGCAAAAGAAGCTACCGATCATAAAAAGATCATCATCATGGAT AGCAAGACCGACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGA CTTCCCATTGGCCACCCGGCTTCAACGAGTACGACTTCGTGCCCGAG AGCTTCGACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTG GCAGTACCGGATTGCCCAAGGGCGTAGCCCTACCGCACCACCGC TTGTGTCCGATTCTAGTCATGCCCCGCGACCCCATCTTCGGCAACCAGA TCATCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTCACCACGGC TTCGGCATGTTACACACGCTGGGCTACTTGATCTGCGGCTTTCGGGT CGTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTTGC AAGACTATAAGATTCAATCTGCCCTGCTGGTGCCACACTATTTAGC TTCTTCGCTAAGAGCACTCTCATCGACAAGTACGACCTAAGCAACTT GCACGAGATCGCCAGCGGCGGGGCGCCGCTCAGCAAGGAGGTAGG TGAGGCGGTGGCCAAACGCTTCCACCTACCAGGCATCCGCCAGGGC TACGGCCTGACAGAAACAACCAGCGCCATTCTGATCACCCCGAAG GGGACGACAAGCCTGGCGCAGTAGGCAAGGTGGTGCCCTTCTTCGA GGCTAAGGTGGTGGACTTGGACACCGGTAAGACACTGGGTGTGAAC CAGCGCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCT ACGTTAACAACCCCGAGGCTACAAACGCTCTCATCGACAAGGACGG CTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAGGACGAGCAC TTCTTCATCGTGGACCGGCTCAAAGCCTGATCAAATACAAGGGCT ACCAGGTAGCCCCAGCCGAACCTGGAGAGCATCTGTGCAACACCC CAACATCTTCGACGCCGGGGTGCCTGGCCTGCCGACGACGATGCC GGCGAGCTGCCCGCCGAGTCGTGCTGTGGAACACGGTAAACCA TGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCAGGTTACAAC CGCCAAGAAGCTGCGCGGTGGTGTGTGTTCTGTGGACGAGGTGCCT AAAGGACTGACCGCAAGTTGGACGCCCCGCAAGATCCGCGAGATTCT TCATTAAGGCCAAGAAGGGCGAAAGATCGCCGTGTAA
23	HPV 蛋白编 码基因序列	ATGGATGCTATGAAACGGGGCCTGTGCTGCGTGCTGCTCCTGTGCG GCGCTGTGTTTGTGAGCCCTAGCATCACCCAGGACTGCTCCTTCCAA CACAGCCCCATCTCTCCGACTTCGCTGTCAAAATCCGTGAGCTGTC TGACTACCTGCTTCAAGATTACCCAGTCACCGTGGCCTCCAACCTGC AGGACGAGGAGCTCTGCGGGGGCCTCTGGCGGCTGGTCTTGGCACA GCGCTGGATGGAGCGGCTCAAGACTGTCGCTGGGTCCAAGATGCAA GGCTTGTGGAGCGCGTGAACACGGAGATACACTTGTCAACCAAT GTGCCTTTCAGCCCCCCCCAGCTGTCTTCGCTTCGTCCAGACCAAC

		ATCTCCCGCCTCCTGCAGGAGACCTCCGAGCAGCTGGTGGCGCTGA AGCCCTGGATCACTCGCCAGAACTTCTCCCGGTGCCTGGAGCTGCA GTGTCAGCCCCGACTCCTCAACCCTGCCACCCCCATGGAGTCCCCGG CCCCCTGGAGGCCACAGCCCCGACAGCCCCGGGCGGCGGCAGCGGC GATATGCACCAGAAGAGAACCGCCATGTTCAGGACCCCTCAGGAGA GACCTAGGAAGCTGCCTCACCTGTGTACAGAGCTCCAGACAACCAT CCACGACATCATCCTGGAGTGCCTGTACTGTAAGCAGCAGCTGCTG AGAAGAGAGGTGTACGACTTCGCCTTCAGAGACCTGTGCATCGTGT ACAGAGACGCAACCCTTACGCCGTGTGCGATAAGTGTCTGAAGTT CTATTCCAAAATCTCCGAATATAGGTACATGCACGGCGACACCCCT ACCCCTGCACGAGTACATGCTGGACCTCCAGCCTGAGACCACAGACC TGTA CTGCTACGAGCAGCTGAACGACAGCTCTGAGGAAGAGGACG AGATTGACGGACCTGCTGGCCAGGCCGAGCCTGACAGAGCCCACTA CAATATCGTGACATTCTGTTGCAAATGCGACTCCACACTGGACAAG TGCCTGAAGTTCTACAGCAAGATCTCTGAGTACAGATACTACTGCT ACTCTGTGTACGGCACCACTGGAGCAGCAGTACAACAAGCCTCT GTGCGACCTCCTGATCCGCTGCATCAACTGCCAGAAGCCTCTGTGCC CTGAGGAGAAGCAGAGACACCTGGACAAGAAGCAGCGGTCCACA ACATCAGAGGCAGATGGACCGGCAGGTGCATGTCCTGCTGTAGATC CTCCAGAACCAGACGGGAGACCCAGCTGCACTACAACATCGTGACC TTCTGCTGCAAGTGCGACTCTACCCTGAGACTGTGCGTGCAGTCTAC CCACGTGGACATCAGAACCCTGGAGGACCTGCTGATGGGCACCCCTG GGCATCGTGTGCCCTATCTGCTCTCAGAAGCCTATGGCCAGGTTCCA GGACCTACAGAAAGACCCCTACAAGCTGCCTGACCTGTGCACCGAG CTGAACACCTCTCTGCAAGACATCGAGATCACCTGCGTGTACTGCA AGACCGTGCTGGAGCTGACCGAGGTGTTGAGTTCGCTTCAAGGA CCTGTTGCTGGTGTACAGAGACAGCATCCCTCACGCTGCCTGCCAC AAGTGCATCGACTTCTATTCCAGGATCAGGGAGCTGCGCTATTACTC CGACTCTGTGATGTACGGCCCCAAGGCCACCCCTCCAGGACATCGTG CTGCACCTGGAGCCTCAGAACGAGATCCCCGTGGACCTGCTGTGCC ACGAGCAGCTGTCTGACTCTGAAGAGGAGAACGACGAGATCGACG GCGTGAACCACCAGCACCTGCCTGCCAGGAGAGCTGAACCCAGCG GCATACCATGCTGTGTATGTGCTTCTACTCTAGGATCAGAGAGCTGA GGTACTACTCTGACTCTGTGTACGGCGACACCCTGGAGAAGCTGAC CAACACCGGCCTGTACAACCTGCTGATCCGGTGCCTGAGGTGCCAG AAGCCTCTGAACCCTGCCGAGAAGCTGAGACACCTGAACGAGAAG AGAAGATTCCACAAGATCGCTGGCCACTACAGAGGCCAGTGCCACT CTTGCTGCAACAGAGCCAGACAGGAGAGACTCCAGCGGAGAAGGG AGACCCAGGTGGCCAGGAGAGCCGAGCCTCAGAGACACACCATGC TGTGCATGTGCTGCAAGTGCGAGGCCAGAAATCGAGCTGGTGGTGA GAGCTCTGCCGACGACCTGAGAGCCTTCCAGCAGCTGTTCTGTCTA CCCTGAGCTTCGTGTGCCCTTGGTGCGCCTCTCAGCAGTAA
24	5' UTR-HPV 蛋白编码基 因 -3' UTR- poly (A) P3	ACATTTGCTTCTGACACAACCTGTGTTCACTAGCAACCTCAAACAGAC ACCGGATCCGCCACCATGGATGCTATGAAACGGGGCCTGTGCTGCG TGCTGCTCCTGTGCGGCGCTGTGTTTGTGAGCCCTAGCATCACCCAG GACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAA AATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTACCCG TGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTGTGGCG GCTGGTCTTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTGCGT GGGTCCAAGATGCAAGGCTTGTGAGGCGCGTGAACACGGAGATA CACTTTGTCAACAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCG CTTCGTCCAGACCAACATCTCCCGCCTCCTGCAGGAGACCTCCGAG CAGCTGGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCC GGTGCCTGGAGCTGCAGTGTGAGCCGACTCCTCAACCCTGCCACC CCCATGGAGTCCCCGGCCCCCTGGAGGCCACAGCCCCGACAGCCCCG GGCGGCGGCAGCGGCGATATGCACCAGAAGAGAACC GCCATGTTT CAGGACCTCAGGAGAGACCTAGGAAGCTGCCTCACCTGTGTACAG AGCTCCAGACAACCATCCACGACATCATCCTGGAGTGCCTGTACTG TAAGCAGCAGCTGCTGAGAAGAGAGGTGTACGACTTCGCCTCAGA GACCTGTGCATCGTGTACAGAGACGGCAACCCTTACGCCGTGTGCG

		ATAAGTGTCTGAAGTTCTATTCCAAAATCTCCGAATATAGGTACATG CACGGCGACACCCCTACCCTGCACGAGTACATGCTGGACCTCCAGC CTGAGACCACAGACCTGTACTGCTACGAGCAGCTGAACGACAGCTC TGAGGAAGAGGACGAGATTGACGGACCTGCTGGCCAGGCCGAGCC TGACAGAGCCCACTACAATATCGTGACATTCTGTTGCAAATGCGAC TCCACACTGGACAAGTGCCCTGAAGTTCTACAGCAAGATCTCTGAGT ACAGATACTACTGCTACTCTGTGTACGGCACCACACTGGAGCAGCA GTACAACAAGCCTCTGTGCGACCTCCTGATCCGCTGCATCAACTGCC AGAAGCCTCTGTGCCCTGAGGAGAAAGCAGAGACACCTGGACAAGA AGCAGCGGTTCCACAACATCAGAGGCAGATGGACCGGCAGGTGCA TGTCTGCTGTAGATCCTCCAGAACCAGACGGGAGACCCAGCTGCA CTACAACATCGTGACCTTCTGCTGCAAGTGCGACTCTACCCTGAGAC TGTGCGTGCAGTCTACCCACGTGGACATCAGAACCCTGGAGGACCT GCTGATGGGCACCCTGGGCATCGTGTGCCCTATCTGCTCTCAGAAG CCTATGGCCAGGTTGAGGACCCTACCAGAAGACCCTACAAGCTGC CTGACCTGTGCACCGAGCTGAACACCTCTCTGCAAGACATCGAGAT CACCTGCGTGTACTGCAAGACCGTGCTGGAGCTGACCGAGGTGTT GAGTTCGCCTTCAAGGACCTGTTGCTGGTGTACAGAGACAGCATCC CTCACGCTGCCTGCCACAAGTGATCGACTTCTATTCCAGGATCAGG GAGCTGCGCTATTACTCCGACTCTGTGATGTACGGCCCCAAGGCCA CCCTCCAGGACATCGTGTGTCACCTGGAGCCTCAGAACGAGATCCC CGTGGACCTGCTGTGCCACGAGCAGCTGTCTGACTCTGAAGAGGAG AACGACGAGATCGACGGCGTGAACCACCAGCACCTGCCTGCCAGG AGAGCTGAACCCCAAGCGGCATACCATGCTGTGTATGTGCTTCTACTC TAGGATCAGAGAGCTGAGGTACTACTCTGACTCTGTGTACGGGAC ACCCTGGAGAAGCTGACCAACACCGGCCTGTACAACCTGCTGATCC GGTGCCTGAGGTGCCAGAAGCCTCTGAACCCTGCCGAGAAGCTGAG ACACCTGAACGAGAAGAGAAGATTCCACAAGATCGCTGGCCACTAC AGAGGCCAGTGCCACTCTTGCTGCAACAGAGCCAGACAGGAGAGA CTCCAGCGGAGAAGGGAGACCCAGGTGGCCAGGAGACCGCAGCT CAGAGACACACCATGCTGTGCATGTGCTGCAAGTGCGAGGCCAGAA TCGAGCTGGTGGTGGAGAGCTCTGCCGACGACCTGAGAGCCTTCCA GCAGCTGTTCTGTCTACCCTGAGCTTCGTGTGCCCTTGGTGCGCCT CTCAGCAGTAAGGCGCGCCGCTCGCTTCTTGCTGTCCAATTTCTAT TAAAGGTTCCCTTGTTCCTAAGTCCAACTACTAACTGGGGGATAT TATGAAGGGCCTTGAGCATCTGGATTCTGCCTGCTCGCTTTCTTGCT GTCCAATTTCTATTAAAGGTTCCCTTGTTCCTAAGTCCAACTACTA AACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTG AATTCAAA AAAAAAAAAAAAAAAAAAAAAAAAAGATCAAAAAAAAAAAAAAAAAA AAGAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAA
25	5' UTR-HPV 蛋白编码基 因 -3' UTR- poly (A) P4	ACATTTGCTTCTGACACAACGTGTTCACTAGCAACCTCAAACAGAC ACCGGATCCGCCACCATGGATGCTATGAAACGGGGCCTGTGCTGCG TGCTGCTCCTGTGCGGCGCTGTGTTTGTGAGCCCTAGCATCACCCAG GACTGCTCCTTCCAACACAGCCCCATCTCCTCCGACTTCGCTGTCAA AATCCGTGAGCTGTCTGACTACCTGCTTCAAGATTACCCAGTCACCG TGGCCTCCAACCTGCAGGACGAGGAGCTCTGCGGGGGCCTTGCGCT GCTGGTCTTGGCACAGCGCTGGATGGAGCGGCTCAAGACTGTGCGT GGGTCCAAGATGCAAGGCTTGTGGAGCGCGTGAACACGGAGATA CACTTTGTACCAAATGTGCCTTTCAGCCCCCCCCCAGCTGTCTTCG CTTCGTCCAGACCAACATCTCCGCGCTCCTGCAGGAGACCTCCGAG CAGCTGGTGGCGCTGAAGCCCTGGATCACTCGCCAGAACTTCTCCC GGTGCCTGGAGCTGCAGTGTGAGCCCGACTCCTCAACCCTGCCACC CCCATGGAGTCCCCGGCCCTGGAGGCCACAGCCCCGACAGCCCCG GGCGGCGGCAGCGCGATATGCACCAGAAGAGAACC GCCATGTTT CAGGACCTCAGGAGAGACCTAGGAAGCTGCCTCACCTGTGTACAG AGCTCCAGACAACCATCCACGACATCATCCTGGAGTGCGTGTACTG TAAGCAGCAGCTGCTGAGAAAGAGAGGTGTACGACTTCGCCTCAGA GACCTGTGCATCGTGTACAGAGACGGCAACCCCTACGCCGTGTGCG

		ATAAGTGTCTGAAGTTCTATTCCAAAATCTCCGAATATAGGTACATG CACGGCGACACCCCTACCCTGCACGAGTACATGCTGGACCTCCAGC CTGAGACCACAGACCTGTACTGCTACGAGCAGCTGAACGACAGCTC TGAGGAAGAGGACGAGATTGACGGACCTGCTGGCCAGGCCGAGCC TGACAGAGCCCACTACAATATCGTGACATTCTGTTGCAAATGCGAC TCCACACTGGACAAGTGCCTGAAGTTCTACAGCAAAGATCTCTGAGT ACAGATACTACTGCTACTCTGTGTACGGCACCACACTGGAGCAGCA GTACAACAAGCCTCTGTGCGACCTCCTGATCCGCTGCATCAACTGCC AGAAGCCTCTGTGCCCTGAGGAGAAGCAGAGACACCTGGACAAGA AGCAGCGGTTCCACAACATCAGAGGCAGATGGACCGGCAGGTGCA TGTCTGCTGTAGATCCTCCAGAACCAGACGGGAGACCCAGCTGCA CTACAACATCGTGACCTTCTGCTGCAAGTGCGACTCTACCCTGAGAC TGTGCGTGCAGTCTACCCACGTGGACATCAGAACCCTGGAGGACCT GCTGATGGGCACCCTGGGCATCGTGTGCCCTATCTGCTCTCAGAAG CCTATGGCCAGGTTGAGGACCCTACCAGAAGACCCTACAAGCTGC CTGACCTGTGCACCGAGCTGAACACCTCTCTGCAAGACATCGAGAT CACCTGCGTGTACTGCAAGACCGTGTGAGCTGACCGAGGTGTTT GAGTTCGCCTTCAAGGACCTGTTCTGTGGTGTACAGAGACAGCATCC CTCACGCTGCCTGCCACAAGTGCATCGACTTCTATTCCAGGATCAGG GAGCTGCGCTATTACTCCGACTCTGTGATGTACGGCCCCAAGGCCA CCCTCCAGGACATCGTGTGCGACCTGGAGCCTCAGAACGAGATCCC CGTGGACCTGCTGTGCCACGAGCAGCTGTCTGACTCTGAAGAGGAG AACGACGAGATCGACGGCGTGAACCACCAGCACCTGCCTGCCAGG AGAGCTGAACCCAGCGGCATACCATGCTGTGTATGTGCTTCTACTC TAGGATCAGAGAGCTGAGGTACTACTCTGACTCTGTGTACGGCGAC ACCCTGGAGAAGCTGACCAACACCGGCCTGTACAACCTGCTGATCC GGTGCCTGAGGTGCCAGAAGCCTCTGAACCCTGCCGAGAAGCTGAG ACACCTGAACGAGAAGAGAAGATTCCACAAGATCGCTGGCCACTAC AGAGGCCAGTGCCACTCTTGCTGCAACAGAGCCAGACAGGAGAGA CTCCAGCGGAGAAGGGAGACCCAGGTGGCCAGGAGAGCCGAGCCT CAGAGACACACCATGCTGTGCATGTGCTGCAAGTGCGAGGCCAGAA TCGAGCTGGTGGTGGAGAGCTCTGCCGACGACCTGAGAGCCTTCCA GCAGCTGTTCTGTCTACCCTGAGCTTCGTGTGCCCTTGGTGCGCCT CTCAGCAGTAAGGCGCGCCGCTCGCTTTCTTGCTGTCCAATTTCTAT TAAAGGTTCCCTTGTTCCTAAGTCCAATACTAACTGGGGGATAT TATGAAGGGCCTTGAGCATCTGGATTCTGCCTGCTCGCTTTCTTGCT GTCCAATTTCTATTAAAGGTTCCCTTGTTCCTAAGTCCAATACTA AACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTG AATTCAA GAAAAAAAAAAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAA AA
26	HPV 蛋白序 列	MDAMKRGLCCVLLLCGAVFVSPSITQDCSFQHSPISDFAVKIRELS DYLLQDYPVTVASNLQDEELCGGLWRLVLAQRWMERLKT VAGSKMQGLLERNTEIHFTVKCAFQPPPSCLRFVQTNISRL LQETSEQLVALKPWTRQNFSRCLELQCQPDSSSTLPPPWSP RPLEATAPTAPGGSGDMHQKRTAMFQDPQERPRKLP HLCTELQTTIHDIILECVYCKQQLLRREYVDFAFRDL CIVYRDGNPYAVCDKCLKFYISKISEYRYMHGDTPTL HEYMLDLQPETDLYCYEQLNDSSEEEDEIDGPAGQAE PDRAHYNIVTFCKCDSTLCLKFYISKISEYRYCYSVY GTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDK KQRFHNIRGRWTGRCMSCCRSSRTRRETQLHYNIVTF CKCDSTLRLRCVQSTHVDIRTLEDLLMGTLGIVCPIC SQKPMARFEDPTRRPYKLPDLCTELNTSLQDIEITC VYCKTVLELFEVFEFAFKDLFVVYRDSIPHAACHKCI DFYSRIRELRYYSDSVMYGPKATLQDIVLHLEPQNEI PVDLLCHEQLSDSEENDEIDGVNHQHLPARRAEPQR HTMLCMCFYSRIELRYYSDSVYGDTLEKLTNTGLYN LLIRCLRCQKPLNPAEKLRLHLEKRRFHKLAGHYRG QCHSCCNRRARQERLQRRRETQVARRAEPQRHTMLC MCKCEARIELVVESSADDLRAFQQLFLSTLSFVCPW CASQQ

权 利 要 求 书

1、一种可在细胞中复制的工程化DNA分子，其包含多聚腺苷尾（Poly A尾）编码序列，所述Poly (A)尾编码序列包含唯一的元件a和至少一个元件b，以及至少一个元件c和/或至少一个元件d：

所述元件a由多个连续的腺嘌呤(A)核苷酸组成，所述元件a的长度范围为 $\geq 20\text{nt}$ ；

所述元件b由多个连续的A核苷酸组成，所述元件b的长度范围为 $3\text{nt} \leq b < 20\text{nt}$ ；

所述元件c由一个非A的核苷酸组成，所述核苷酸选自T、C、G核苷酸；

所述元件d由任意两个或更多个连续的核苷酸组成，所述核苷酸选自A、T、C、G核苷酸，其中元件d的5'及3'末端的核苷酸不为A核苷酸，并且其不包含3个以上连续的A核苷酸；所述元件d的长度范围为 $2\text{nt} \leq d \leq 20\text{nt}$ ；

其中所述元件a和元件b不相邻、元件c和元件d不相邻，

且所述Poly (A)尾编码序列不包含彼此相邻的元件b、不包含彼此相邻的元件c、且不包含彼此相邻的元件d。

2、根据权利要求1所述的DNA分子，所述Poly (A)尾编码序列的长度为101至200nt或101-150nt或120至150nt或130-140nt或120-135nt或123-125nt。

3、根据权利要求1或2所述的DNA分子，所述Poly (A)尾编码序列3'末端的为A核苷酸或非A核苷酸。

4、根据权利要求1-3中任一项所述的DNA分子，其中所述元件a $\leq 80\text{nt}$ 。

5、根据权利要求1-3中任一项所述的DNA分子，其中所述元件a为30至70nt、35至65nt、40至60nt、或45至55nt，优选60nt。

6、根据权利要求1-5中任一项所述的DNA分子，其中所述元件a的50%或以上的多核苷酸位于所述Poly (A)尾编码序列的5'部分或3'部分。

7、根据权利要求1-6中任一项所述的DNA分子，其中所述元件b为3nt至10nt、10至19nt、12至15nt、14nt至17nt或16至19nt，优选19nt。

8、根据权利要求1-7中任一项所述的DNA分子，其中所述元件b的个数为2-10个，优选2-5个，进一步优选3个。

9、根据权利要求1-8中任一项所述的DNA分子，其中所述元件c为G。

10、根据权利要求1-9中任一项所述的DNA分子，其中所述元件c的个数

为2至10个、3个至8个、或4至6个或2-5个，优选2个。

11、根据权利要求1-10中任一项所述的DNA分子，其中所述元件d包含回文序列。

12、根据权利要求1-11中任一项所述的DNA分子，其中所述元件d为3至18nt、5至16nt、4至10nt、或6至12nt，优选6nt。

13、根据权利要求1-12中任一项所述的DNA分子，其中所述元件d选自以下序列中的任一种或多种：GATATC (SEQ ID NO: 15)、GTATAC (SEQ ID NO: 16)、GAATCT (SEQ ID NO: 17)、GCATATGACT (SEQ ID NO: 18) 及GATATCGTATAC (SEQ ID NO: 19)。

14、根据权利要求1-13中任一项所述的DNA分子，其中所述元件d选自如下核苷酸序列中的任一种或多种：SEQ ID NO:15、SEQ ID NO:16、及SEQ ID NO: 17。

15、根据权利要求1-14中任一项所述的DNA分子，其中所述元件d的核苷酸序列如SEQ ID NO: 15所示。

16、根据权利要求1-15中任一项所述的DNA分子，其中所述元件d的个数为0-5，优选1-3个，进一步优选1个。

17、根据权利要求1-16中任一项所述的DNA分子，其中当元件c和元件d同时存在时，所述元件c和元件d的个数总和为2-15个，优选3-5个，进一步优选3个。

18、根据权利要求1-17中任一项所述的DNA分子，其中在Poly (A)尾编码序列的3'部分，优选所述Poly (A) 尾编码序列靠近3'末端的1/2部分，包含一个或多个非A核苷酸。

19、根据权利要求1-18中任一项所述的DNA分子，所述Poly (A)尾编码序列的结构为：

元件a-元件c-元件b-元件c-元件b-元件c-元件b-元件c-元件b；

元件b-元件c-元件b-元件c-元件a-元件d-元件b-元件c-元件b-元件c-元件b；

元件b-元件c-元件b-元件c-元件b-元件d-元件a-元件c；

元件a-元件d-元件b-元件c-元件b-元件c-元件b；或

元件b-元件c-元件b-元件c-元件b-元件d-元件a。

20. 根据权利要求1-19中任一项所述的DNA分子，所述Poly (A)尾编码序列的结构为：

元件a-元件d-元件b-元件c-元件b-元件c-元件b；所述元件a为60nt长度，元件b为16至19nt长度，元件d为6nt长度。

21、根据权利要求1-20中任一项所述的DNA分子，所述Poly (A)尾编码序列的结构为：

元件b-元件c-元件b-元件c-元件b-元件d-元件a；所述元件a为60nt长度，元件b为16至19nt长度，元件d为6nt长度。

22、根据权利要求1-21中任一项所述的DNA分子，所述Poly (A)尾编码序列如SEQ ID NO：1至10中任一项所示。

23、根据权利要求1-22中任一项所述的DNA分子，所述Poly (A)尾编码序列如SEQ ID NO：3或SEQ ID NO：4所示。

24、根据权利要求1-23中任一项所述的DNA分子，其在Poly (A)尾编码序列的5'端一侧进一步连接目的基因片段，所述目的基因片段与所述Poly (A)尾编码序列共同编码RNA。

25、根据权利要求1-24中任一项所述的DNA分子，其进一步包含复制子。

26、根据权利要求1-25中任一项所述的DNA分子，其进一步包含抗性基因。

27、根据权利要求1-26中任一项所述的DNA分子，其进一步包含启动所述RNA转录的启动子。

28、根据权利要求1-27中任一项所述的DNA分子，所述目的基因片段包括5'UTR编码序列。

29、根据权利要求1-28中任一项所述的DNA分子，所述目的基因片段包括蛋白编码序列或非蛋白编码序列。

30、根据权利要求1-29中任一项所述的DNA分子，所述目的基因片段包括3'UTR编码序列。

31、根据权利要求1-30中任一项所述的DNA分子包含复制子、抗生素抗性基因、启动子、5'UTR编码序列、蛋白编码序列、及3'UTR编码序列。

32、根据权利要求1-31的任一项所述的DNA分子，其中所述蛋白编码序列编码HPV（人乳头瘤病毒）蛋白，优选地所述HPV蛋白源自16型和/或18型HPV。

33、根据权利要求1-32中任一项所述的DNA分子，其中所述蛋白编码序列编码HPV的E2、E6或E7蛋白，E6和E7蛋白多肽片段的融合蛋白或E2、E6

和E7蛋白多肽片段的融合蛋白，优选地所述HPV蛋白源自16型和/或18型HPV。

34、根据权利要求1-33中任一项所述的DNA分子，其中所述蛋白编码序列编码如SEQ NO: 26所示的多肽。

35、根据权利要求1-34中任一项所述的DNA分子，其包含如SEQ ID NO: 22-25中任一项所示的多核苷酸序列，或如SEQ ID NO: 22-25中任一项所示的多核苷酸序列的同义突变体，或与SEQ ID NO: 22-25中任一项所示的多核苷酸序列或其同义突变体具有85%以上序列同一性的多核苷酸序列。

36、根据权利要求1-35中任一项所述的DNA分子，所述DNA分子为DNA质粒。

37、包含根据权利要求1-36中任一项所述的DNA分子的细胞。

38、根据权利要求37的细胞为原核细胞。

39、根据权利要求37的细胞为大肠杆菌。

40、根据权利要求1-36中任一项所述的DNA分子编码的RNA分子。

41、根据权利要求40所述的RNA分子，所述RNA还包含5'-帽结构和/或所述RNA的部分或全部尿苷为化学修饰的尿苷；优选的所述RNA的部分或全部尿苷为假尿苷或1-甲基-假尿苷。

42、根据权利要求1-36中任一项所述的Poly (A)尾的DNA编码序列。

43、根据权利要求1-36中任一项所述的Poly (A)尾的DNA编码序列所编码的Poly (A)尾序列。

44、将根据权利要求42的Poly (A)尾的DNA编码序列用于使编码RNA的DNA分子在宿主细胞中的复制更加保守的用途，其中所述Poly (A) 尾位于所述RNA的3'末端。

45、根据权利要求44所述的用途，所述宿主细胞为原核细胞，优选大肠杆菌。

46、根据权利要求43所述的Poly (A)尾序列用于调控RNA分子在宿主细胞中表达的用途，其中所述Poly (A) 尾位于所述RNA的3'末端。

47、根据权利要求46所述的用途，所述宿主细胞为真核细胞，优选哺乳动物细胞，进一步优选人细胞。

48、包含根据权利要求1-36中任一项所述的DNA分子的文库。

49、包含根据权利要求1-36中任一项所述的DNA分子所编码的RNA分子

的文库。

50、调控蛋白表达的方法，包括：

将根据权利要求48的DNA文库中的多种DNA在不同时间，和/或以不同的数量比例导入目的细胞；或

将权利要求49的RNA文库中的多种RNA在不同时间，和/或以不同的数量比例导入目的细胞。

51、DNA与RNA的杂合分子，其携带与根据权利要求1-36中任一项所述的DNA分子相同的遗传信息，与根据权利要求42所述的Poly (A)尾相同的遗传信息，或与根据权利要求40所述的RNA分子相同的遗传信息。