促进 RNA 翻译的 UTR

技术领域

本发明涉及生物医药领域，特别涉及 RNA 药物领域。具体而言，本发明涉及利用人工智能技术获得的促进 RNA 翻译的 UTR，以及包含所述 UTR 的 RNA 分子及其用途。

发明背景

RNA (如 mRNA 或环形 RNA)药物分子具有独特优势。相较于传统的 DNA 基因治疗方法，RNA (如 mRNA 或环形 RNA)药物不会生产外源基因组整合至宿主基因组的安全性风险。此外，RNA (如 mRNA 或环形 RNA)还可以提供即时和可调控的蛋白质表达，使其成为治疗特定基因缺失导致疾病的理想选择。同时，RNA (如 mRNA 或环形 RNA)还能够作为抗原，引起免疫应答，为疫苗的开发提供了新的途径。

mRNA 的 UTR 对于 mRNA 的表达发挥了重要的调控作用。不同的 UTR 序列可以通过影响 mRNA 稳定性、翻译效率，直接影响 mRNA 编码蛋白的表达水平。因此，对 5'UTR 序列进行优化和筛选，以提高 mRNA 所编码蛋白的表达水平，对于增强治疗效果和激活疫苗免疫应答至关重要。

当前的 mRNA 药物中的 UTR 序列通常是从天然存在的基因组中筛选获得的。然而，这种筛选方法存在一些局限性，限制了 mRNA 的表达水平和治疗效果的进一步提升。首先，天然存在的 UTR 序列通常具有较为保守的特征，其演化受到多种选择压力的影响，包括转录调控、翻译效率和 RNA 稳定性等。因此，这些序列在调控机制和翻译效率方面可能存在一定的局限性，无法满足对高水平基因表达的需求。其次，天然存在的 UTR 序列通常是针对特定基因和物种的调控进行优化的。将这些序列应用于其他基因或物种时，可能会遇到表达水平不佳、不稳定或引起过强免疫反应等问题。这限制了 UTR 序列的通用性和应用范围。另外，当前的筛选方法主要依赖于实验室实际操作和传统生物学方法，这需要耗费大量时间和资源，并且存在一定的随机性和局限性，使得传统的筛选方法变得繁琐且难以高效地实现。

发明简述

本发明通过应用人工智能(Artificial Intelligence, AI)技术，系统地筛选和设计具有优化翻译效率的 UTR 序列。AI 算法能够分析目前已公开的人类 RNA-seq 组学信息和 Ribo-seq 组学信息，并建立深度学习预测模型来预测不同 UTR 序列的翻译效率。在此基础上，使用人工智能分析策略，包括对抗生成网络(Generative Adversarial Network, GAN)和长短期记忆递归神经网络(Long Short-Term Memory, LSTM)等技术，以人类基因组中的

UTR 作为训练集，构建生成式学习模型，产生非天然的 UTR 序列。进一步地，通过体外转录(in vitro transcription)的技术在体外进行实验验证，测定目标蛋白的表达水平，筛选出实际具有增强目标蛋白表达效果的 UTR。

通过本发明的筛选和优化过程，可获得一系列具有增强目标蛋白的表达水平的 5'UTR 和 3'UTR 序列，可在 RNA 治疗中增强治疗效果，作为疫苗时增强免疫激活。为治疗特定基因缺失疾病和疫苗研发领域带来新的突破。

因此，本发明一方面涉及一种 5'UTR (5'非翻译区)，其包含与 SEQ ID NO:1-13 和 41-60 之一具有至少 80%、至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%序列相同性的核苷酸序列或其互补序列，优选地，所述 5'UTR 包含 SEQ ID NO:1-13 和 41-60 之一所示的核苷酸序列或其互补序列，更优选地，所述 5'UTR 包含 SEQ ID NO:1、3-5、7-11、13、41-46、48、50-53、55-57、59-60 之一所示的核苷酸序列或其互补序列，更优选地，所述 5'UTR 包含 SEQ ID NO:4-5、7-10、41、43-45、48、50-51、53、55、57、59 之一所示的核苷酸序列或其互补序列。

在另一方面，本发明涉及一种 3'UTR (3'非翻译区)，其

i) 包含与 SEQ ID NO:15-28 和 61-80 之一具有至少 80%、至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%序列相同性的核苷酸序列或其互补序列，优选地，所述 3'UTR 包含 SEQ ID NO:15-28 和 61-80 之一所示的核苷酸序列或其互补序列，更优选地，所述 3'UTR 包含 SEQ ID NO:15-20、22-23、25-26、28、61-71、73-80 之一所示的核苷酸序列或其互补序列，更优选地，所述 3'UTR 包含 SEQ ID NO:15-16、18-19、22、25、28、61、63-66、68、70-71、73-80 之一所示的核苷酸序列或其互补序列；或

ii) 包含具有插入的一个或多个额外序列的 SEQ ID NO:15-28 和 61-80 之一所示的核苷酸序列或其互补序列，优选所述额外序列为 miRNA 结合位点，更优选额外序列选自全长的 microRNA 反向互补序列或其种子序列的反向互补序列，更优选所述全长的 microRNA 反向互补序列长度为 19-25 nt 或种子序列的反向互补序列长度为 7-8 nt。

在另一方面，本发明涉及一种用于表达感兴趣的多肽的 RNA 分子，其包含本发明的 5'UTR 和/或本发明的 3'UTR。

在另一方面，本发明涉及一种核酸载体，其包含本发明的 RNA 分子的编码序列。

在另一方面，本发明涉及包含本发明的 RNA 分子或本发明的核酸载体的细胞。

在另一方面，本发明涉及一种提高细胞中感兴趣多肽的表达的方法，所述方法包括向所述细胞导入本发明的 RNA 分子和/或本发明的核酸载体。

在另一方面，本发明涉及一种药物组合物，其包含本发明的 RNA 分子、和/或本发明的核酸载体和/或本发明的宿主细胞，以及药物可接受的载体。

在另一方面，本发明涉及本发明的 5'UTR 和/或本发明的 3'UTR 用于提高包含感兴趣的多肽的编码序列的 RNA 分子中所述感兴趣多肽的翻译效率的用途。

附图简述

图 1、包含本发明 UTR 的示例性的 mRNA 序列。A：5'UTR；B：3'UTR。

图 2、5200 片段分析仪(Agilent)检测 mRNA 分子(图 1 所示序列)的长度和完整度。

图 3、包含候选 5'UTR 的 mRNA 转染的细胞的裂解液中萤火虫荧光素酶相对光单位。根据 student t-test 统计差异。

图 4、流式细胞技术检测用包含候选 5'UTR 的 mRNA 转染的细胞裂解液中红色荧光蛋白的相对光强，并根据 student t-test 统计差异。

图 5、用包含候选 3'UTR 的 mRNA 转染的细胞的裂解液中萤火虫荧光素酶相对光单位

图 6、流式细胞技术检测用包含候选 3'UTR 的 mRNA 转染的细胞中红色荧光蛋白的相对光强。

图 7、荧光定量分析包含 3 个 microRNA 结合位点的本发明 3'UTR 的 mRNA 转染 293T 细胞后绿色荧光蛋白的相对光强，以及在 24 小时内的变化趋势。

图 8、荧光定量分析包含 3 个 microRNA 结合位点的本发明 3'UTR 的 mRNA 转染 Hela 细胞后绿色荧光蛋白的相对光强，以及在 24 小时内的变化趋势。

发明详述

在本发明中，除非另有说明，否则本文中使用的科学和技术名词具有本领域技术人员所通常理解的含义。并且，本文中所用的蛋白质和核酸化学、分子生物学、细胞和组织培养、微生物学、免疫学相关术语和实验室操作步骤均为相应领域内广泛使用的术语和常规步骤。

如本文所用，术语“和/或”涵盖由该术语连接的项目的所有组合，应视作各个组合已经单独地在本文列出。例如，“A 和/或 B”涵盖了“A”、“A 和 B”以及“B”。例如，“A、B 和/或 C”涵盖“A”、“B”、“C”、“A 和 B”、“A 和 C”、“B 和 C”以及“A 和 B 和 C”。

“多核苷酸”、“核酸序列”、“核苷酸序列”或“核酸片段”可互换使用并且是单链或双链 RNA 或 DNA 聚合物，任选地可含有合成的、非天然的或改变的核苷酸碱基。核苷酸通过如下它们的单个字母名称来指代:“A”为腺苷或脱氧腺苷(分别对应 RNA 或 DNA)，“C”表示胞苷或脱氧胞苷，“G”表示鸟苷或脱氧鸟苷，“U”表示尿苷，“T”表示脱氧胸苷，“R”表示嘌呤(A 或 G)，“Y”表示嘧啶(C 或 T)，“K”表示 G 或 T，“H”表示 A 或 C 或 T，“I”表示肌苷，并且“N”表示任何核苷酸。尽管本文中的核苷酸序列可能以 DNA 序列表示(包含 T)，但在提及 RNA 时，本领域技术人员可以容易地确定相应的 RNA 序列(即用 U 替换 T)。

“多肽”、“肽”、和“蛋白质”在本发明中可互换使用，指氨基酸残基的聚合物。该术语适用于其中一个或多个氨基酸残基是相应的天然存在的氨基酸的人工化学类似物的氨基酸聚合物，以及适用于天然存在的氨基酸聚合物。术语“多肽”、“肽”、“氨

基酸序列”和“蛋白质”还可包括修饰形式，包括但不限于糖基化、脂质连接、硫酸盐化、谷氨酸残基的γ羧化、羟化和 ADP-核糖基化。

“包含”一词在本文中用于描述蛋白质或核酸的序列时，所述蛋白质或核酸可以是由所述序列组成，或者在所述蛋白质或核酸的一端或两端可以具有额外的氨基酸或核苷酸，但仍然具有本发明所述的活性。

两个多肽序列或两个多核苷酸序列之间的“序列相同性”指的是所述序列之间的相同的氨基酸或核苷酸的百分比。评估多肽或多核苷酸序列之间的序列相同性水平的方法是本领域已知的。序列相同性可以使用已知的各种序列分析软件评估。例如，序列相同性可以通过 EMBL-EBI 的在线比对工具来评估(https://www.ebi.ac.uk/Tools/psa/)。两个序列之间的序列相同性可以使用 Needleman-Wunsch 算法，使用默认参数来评估。序列相同性可以是沿着给定序列的全长上的序列相同性。

在一方面，本发明涉及一种 5’UTR(5’非翻译区)，其包含与 SEQ ID NO:1-13 和 41-60 之一具有至少 80%、至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%序列相同性的核苷酸序列或其互补序列。在一些实施方案中，所述 5’UTR 包含 SEQ ID NO:1-13 和 41-60 之一所示的核苷酸序列或其互补序列。在一些优选实施方案中，所述 5’UTR 包含 SEQ ID NO:1、3-5、7-11、13、41-46、48、50-53、55-57、59-60 之一所示的核苷酸序列或其互补序列。在一些更优选的实施方案中，所述 5’UTR 包含 SEQ ID NO:4-5、7-10、41、43-45、48、50-51、53、55、57、59 之一所示的核苷酸序列或其互补序列。

“5’UTR”通常是指 mRNA 分子从 5’末端到翻译起始密码子之间的序列，其能够募集核糖体复合物并启动 mRNA 的翻译。5’UTR 通过与转录因子、核糖体和其他转录调控蛋白相互作用，调节转录后修饰、翻译起始复合物的形成和稳定性等过程。

在本发明各方面的一些实施方案中，所述 5’UTR 是非天然的 5’UTR。在一些实施方案中，所述 5’UTR 是人工智能设计的 5’UTR。

在另一方面，本发明涉及一种 3’UTR(3’非翻译区)，其包含与 SEQ ID NO:15-28 和 61-80 之一具有至少 80%、至少 85%、至少 90%、至少 95%、至少 96%、至少 97%、至少 98%、至少 99%序列相同性的核苷酸序列或其互补序列。在一些实施方案中，所述 3’UTR 包含 SEQ ID NO:15-28 和 61-80 之一所示的核苷酸序列或其互补序列。在一些优选实施方案中，所述 3’UTR 包含 SEQ ID NO:15-20、22-23、25-26、28、61-71、73-80 之一所示的核苷酸序列或其互补序列。在一些更优选的实施方案中，所述 3’UTR 包含 SEQ ID NO:15-16、18-19、22、25、28、61、63-66、68、70-71、73-80 之一所示的核苷酸序列或其互补序列。

如本文所用，“3’UTR”是指 mRNA 中多肽编码序列的终止密码子与多聚(A)(poly(A))序列之间的序列。3’UTR 可以通过与 mRNA 结合蛋白、miRNA 等相互作用来调控 mRNA 的翻译。3’UTR 的序列和结构特征可以影响 mRNA 的稳定性、核糖体的扫描和翻译终止复合物的形成等，从而影响蛋白质的表达水平。

本发明的 3’UTR 具有一定的稳定性，可以容忍一定程度的序列(如 microRNA 结合位点)插入，而不影响其促进 mRNA 分子中编码区翻译的能力，亦不影响 mRNA 分子的稳定性。因此，在一些实施方案中，本发明的 3’UTR 还具有插入的一或多个(例如 1-5 个)额外序列。

在一些实施方案中，所述一或多个额外序列插入至 SEQ ID NO:15-28 和 61-80 之一的序列或其互补序列中。在一些实施方案中，所述一或多个额外序列随机插入至 SEQ ID NO:15-28 和 61-80 之一的序列或其互补序列中。在一些实施方案中，所述一或多个额外序列随机插入至 SEQ ID NO:15-28 和 61-80 之一的序列或其互补序列中的不同位置。

在一些优选的实施方案中，所述额外序列为 miRNA 结合位点。在一些优选的实施方案中，额外序列选自全长 microRNA 的反向互补序列或 microRNA 的种子序列的反向互补序列。在一些优选的实施方案中，所述全长 microRNA 的反向互补序列长度为 19-25 nt 或更优选 microRNA 的种子序列的反向互补序列长度为 7-8 nt。

在一些具体实施方案中，所述 microRNA 是 hsa-microRNA-122-5p。hsa-microRNA-122-5p 例如具有 SEQ ID NO:36 所示核苷酸序列，其反向互补序列示于 SEQ ID NO:37。在一些具体实施方案中，本发明所述具有插入的一或多个额外序列的 3’UTR 包含 SEQ ID NO:38 所示核苷酸序列。

在本发明各方面的一些实施方案中，所述 3’UTR 是非天然的 3’UTR。在一些实施方案中，所述 3’UTR 是人工智能设计的 3’UTR

在一方面，本发明提供一种 RNA 分子，其包含本发明的 5’UTR 和/或本发明的 3’UTR。

在一些实施方案中，所述 RNA 分子包含本发明的 5’UTR。本发明的 5’UTR 可以与不同的 3’UTR 组合，实现增强 RNA 分子的翻译的效果。所述 3’UTR 可以是本领域已知的 3’UTR，例如来自天然 RNA 分子的 3’UTR，也可以是非天然的，例如人工智能设计的 3’UTR。

在一些实施方案中，所述 RNA 分子包含本发明的 3’UTR。本发明的 3’UTR 可以与不同的 5’UTR 组合，实现增强 RNA 分子的翻译的效果。所述 5’UTR 可以是本领域已知的 5’UTR，例如来自天然 RNA 分子的 5’UTR，也可以是非天然的，例如人工智能设计的 5’UTR。

在一些实施方案中，所述 RNA 分子包含本发明的 5’UTR 和本发明的 3’UTR。本发明的 5’UTR 可以与本发明的 3’UTR 组合，实现增强 RNA 分子的翻译的效果。

在一些实施方案中，所述 RNA 分子是信使 RNA(mRNA)。mRNA 分子通常是线性的 RNA 分子。

在一些实施方案中，所述 RNA 分子是环形 RNA 分子。环形 RNA 分子是共价闭合的 RNA 分子。

在一些实施方案中，所述 RNA 分子用于在细胞中表达感兴趣的多肽。在一些实施方案中，所述 RNA 分子包含感兴趣的多肽的编码序列，其与所述 5’UTR 和/或所述 3’UTR 可操作地连接。可操作地连接指给定元件能够有效地调控细胞中所述感兴趣的多肽从所

述 RNA 分子的翻译。

在一些实施方案中，所述 RNA 分子还包括多聚(A)序列。在一些实施方案中，所述多聚(A)序列与感兴趣的多肽的编码序列可操作地连接。

多聚(A)序列通常包含多个腺嘌呤核苷酸。多聚(A)序列的添加有助于 mRNA 的稳定性和转运，防止其降解，并在转录后修饰过程中发挥重要作用。多聚(A)序列可以是纯腺嘌呤核苷酸的连续链，但是也可以是包含非腺嘌呤的核苷酸的变体，只要其功能等同于常规的多聚(A)序列，即能够提供与天然多聚(A)序列相似的生物学功能，如影响 mRNA 的稳定性、翻译效率或核糖体结合。已知的多聚(A)序列如人生长激素 (hGH) 多聚(A)序列和猴病毒 40 (SV40) 多聚(A)序列等。这些变体在核苷酸组成上可能有所不同，但在功能上被认定为与传统的多聚(A)序列等效。

在本发明的一些实施方案中，所述多聚(A)序列包含大约 20 个至大约 500 个连续的腺嘌呤核苷酸(A)，例如，包含大约 25 个、大约 50 个、大约 100 个、大约 150 个、大约 175 个、大约 200 个、大约 300 个、大约 400 个、大约 500 个连续的腺嘌呤核苷酸(A)。

在一些实施方案中，所述 RNA 分子从 5’至 3’方向按以下顺序包含

1) 本发明所述 5’UTR；和

2) 感兴趣的多肽的编码序列。

在一些实施方案中，所述 RNA 分子从 5’至 3’方向按以下顺序包含

1) 本发明的 5’UTR；

2) 感兴趣的多肽的编码序列；和

3) 3’UTR。

在一些实施方案中，所述 RNA 分子从 5’至 3’方向按以下顺序包含

1) 本发明的 5’UTR；

2) 感兴趣的多肽的编码序列；和

3) 3’UTR；和

4) 多聚(A)序列。

所述 3)中所述 3’UTR 可以是本领域已知的 3’UTR，例如来自天然 RNA 分子的 3’UTR，也可以是非天然的，例如人工智能设计的 3’UTR。在一些实施方案中，所述 3’UTR 是本发明的 3’UTR。

在一些实施方案中，所述 RNA 分子从 5’至 3’方向按以下顺序包含

1] 感兴趣的多肽的编码序列；和

2] 本发明的 3’UTR。

在一些实施方案中，所述 RNA 分子从 5’至 3’方向按以下顺序包含

1] 5’UTR；

2] 感兴趣的多肽的编码序列；和

3] 本发明的 3’UTR。

在一些实施方案中，所述 RNA 分子从 5’至 3’方向按以下顺序包含

1] 5’UTR；

2] 感兴趣的多肽的编码序列；

3] 本发明的 3’UTR；和

4] 多聚(A)序列。

所述 1]中所述 5’UTR 可以是本领域已知的 5’UTR，例如来自天然 RNA 分子的 5’UTR，也可以是非天然的，例如人工智能设计的 5’UTR。在一些实施方案中，所述 5’UTR 是本发明的 5’UTR。

本文所述“感兴趣的多肽”可以是待在目标细胞中表达的任何多肽。

感兴趣的多肽可以是真核、原核或者病毒来源的多肽。在某些实施方案中，感兴趣的多肽可以是用于治疗、预防或诊断用途的任何多肽。例如，感兴趣的多肽可以是抗原、抗体、基因编辑酶如 CRISPR 核酸酶等。感兴趣的多肽还可以是嵌合抗原受体、免疫调节蛋白、转录因子等。所述多肽的实例包括但不限于：荧光素酶、红色/绿色荧光蛋白、人促红细胞生成素、β-半乳糖苷酶。

所述感兴趣的多肽的编码序列可以针对待表达的目标细胞进行密码子优化。

密码子优化是指通过用在宿主细胞的基因中更频繁地或者最频繁地使用的密码子代替天然序列的至少一个密码子(例如约或多于约 1、2、3、4、5、10、15、20、25、50 个或更多个密码子)同时维持其编码氨基酸序列而修饰核酸序列以便增强在感兴趣宿主细胞中的表达的方法。不同的物种对于特定氨基酸的某些密码子展示出特定的偏好。密码子偏好性(在生物之间的密码子使用的差异)经常与 mRNA 的翻译效率相关，而该翻译效率则被认为依赖于被翻译的密码子的性质和特定的转运 RNA(tRNA)分子的可用性。细胞内选定的 tRNA 的优势一般反映了最频繁用于肽合成的密码子。因此，可以将基因定制为基于密码子优化在给定生物中的最佳基因表达。密码子利用率表可以容易地获得。参见，Nakamura Y.等，“Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucl. Acids Res.，28:292 (2000)。

在一些实施方案中，所述 RNA 分子是化学合成的。在一些实施方案中，所述 RNA 分子是通过体外转录获得的。

在一些实施方案中，所述 RNA 分子是 mRNA 分子，其还包括 5’帽结构。

如本文所用，对于 RNA 而言的“5’帽”包括存在于天然 mRNA 上的 5’帽结构以及其类似物。天然 mRNA 上的 5’帽结构是指甲基化鸟苷酸经焦磷酸与 RNA 的 5’末端核苷酸相连，形成 5’,5’-三磷酸连接(5’,5’-triphosphate linkage)。5’帽通常有三种类型(m7G5’ppp5’Np、m7G5’ppp5’NmpNp、 m7G5’ppp5’NmpNmpNp)，分别称为 Cap0、Cap1 和 Cap2。Cap0 指末端核苷酸的核糖未甲基化，Cap1 指末端一个核苷酸的核糖甲基化，Cap2 指末端两个核苷酸的核糖均甲基化。在本发明一些实施方案中，所述 5’帽结构是 Cap1 帽结构。

对 mRNA 分子加帽的方法是本领域已知的。所述 mRNA 分子的 5’帽结构可以在化学合成或体外转录获得 mRNA 分子后使用通过酶促反应添加(例如通过包含牛痘加帽酶

和 mRNA 帽结构 2'-O-甲基转移酶的商品化试剂盒)。然而，也可以通过在体外转录过程中直接将带有帽结构的核苷酸类似物作为第一核苷酸掺入转录物，生产带有帽结构的mRNA。

在一些实施方案中，本发明的 RNA 分子还可以包含至少一种核苷酸修饰。所述至少一种核苷酸修饰包括但不限于胞苷修饰、尿苷修饰或腺苷修饰。在一些实施方案中，至少一种核苷修饰包括但不限于 5-甲基胞嘧啶(m5C)、N6-甲基腺苷(m6A)、假尿苷(ψ)、N1-甲基假尿苷(m1ψ)和 5-甲氧基尿苷(5moU)。在一些实施方案中，本发明的 RNA 分子可以包含至少一个修饰核苷酸，优选地，所述修饰核苷酸选自假尿苷、N1-甲基-假尿苷、5-甲基胞苷或其组合。

在本发明的一些实施方案中，与相应的包含对照 5'UTR(如 SEQ ID NO:14)的 RNA 分子相比，包含本发明的 5'UTR 的 RNA 分子在细胞中(例如在 HEK293T 或 HCT116 细胞中)导致所述感兴趣的多肽相当的表达或增加的表达，优选地，所述感兴趣的多肽的表达增加大约 10%、大约 20%、大约 30%、大约 40%、大约 50%、大约 60%、大约 70%、大约 80%、大约 90%、大约 100%或更多。优选的本发明的 5'UTR 能导致感兴趣的多肽的表达增加至少 20%。更优选的本发明的 5'UTR 能导致感兴趣的多肽的表达增加至少 50%。

在本发明的一些实施方案中，与相应的包含对照 3'UTR(如 SEQ ID NO:29)的 RNA 分子相比，包含本发明的 3'UTR 的 RNA 分子在细胞中(例如在 HEK293T 或 HCT116 细胞中)所述感兴趣的多肽相当的表达或增加的表达，优选地，所述感兴趣的多肽的表达增加大约 10%、大约 20%、大约 30%、大约 40%、大约 50%、大约 60%、大约 70%、大约 80%、大约 90%、大约 100%或更多。优选的本发明的 3'UTR 能导致感兴趣的多肽的表达增加至少 20%。更优选的本发明的 3'UTR 能导致感兴趣的多肽的表达增加至少 50%。

在本发明的一些实施方案中，与相应的包含对照 5'UTR(如 SEQ ID NO:14)和对照 3'UTR(如 SEQ ID NO:29)的 RNA 分子相比，包含本发明的 5'UTR 和本发明的 3'UTR 的 RNA 分子在细胞中(例如在 HEK293T 或 HCT116 细胞中)导致所述感兴趣的多肽相当的表达或增加的表达，优选地，所述感兴趣的多肽的表达增加大约 10%、大约 20%、大约 30%、大约 40%、大约 50%、大约 60%、大约 70%、大约 80%、大约 90%、大约 100%或更多。

在一方面，本发明提供一种核酸载体，所述载体包含本发明的 RNA 分子的编码序列。在一些实施方案中，所述核酸载体用于生成本发明的 RNA 分子。

如本文所用，"载体"是指从病毒、质粒或高等生物的细胞中提取的一段 DNA，可以将外来 DNA 片段插入或已经插入其中以进行克隆和/或表达目的。在某些实施方案中，载体可以稳定地维持在生物体中。载体可包含，例如，复制起点、选择性标记或报告基因，如抗生素抗性或 GFP，和/或多克隆位点(MCS)。该术语包括线性 DNA 片段(例如，PCR 产物、线性的质粒片段)、质粒载体、病毒载体、粘粒、细菌人工染色体(BAC)、酵母人工染色体(YAC)等。

在一些实施方案中，所述核酸载体还包含与所述 RNA 分子的编码序列可操作地连接的 RNA 聚合酶启动子序列。所述可操作地连接的启动子允许所述 RNA 分子的体内和/或体外转录。所述启动子例如是 T7 RNA 聚合酶启动子、T6 病毒 RNA 聚合酶启动子、SP6 病毒 RNA 聚合酶启动子、T3 病毒 RNA 聚合酶启动子或 T4 病毒 RNA 聚合酶启动子。

在一些实施方案中，所述核酸载体是质粒载体。在一些实施方案中，所述核酸载体在所述 RNA 分子的编码序列的 3'侧翼包含限制性核酸内切酶位点，例如 IIS 型限制性核酸内切酶位点。合适的限制性核酸内切酶包括但不限于 BsmBI、BsaI、SapI 等。所述限制性核酸内切酶位点可以用于对核酸载体进行线性化以进行体外转录。

从核酸载体体外转录获得 RNA 分子的方法是本领域已知的，例如可以使用商品化试剂盒进行体外转录。

在另一方面，本发明提供一种提高细胞中感兴趣多肽的表达的方法，所述方法包括向所述细胞导入本发明所述 RNA 分子和/或本发明所述核酸载体。

本发明所述 RNA 分子和/或本发明所述核酸载体向细胞的导入可用过本发明已知的方法进行，例如显微注射或脂质体介导的转染或电转法。

在另一方面，本发明提供包含本发明的 RNA 分子或本发明的核酸载体的细胞。

在另一方面，本发明提供药物组合物，其包含本发明所述 RNA 分子和/或本发明所述核酸载体和/或本发明的细胞，以及药物可接受的载体。所述组合物的具体用途取决于所述感兴趣的多肽。

在一些实施方案中，所述药物组合物用于在对象中治疗和/或预防疾病。所述治疗和/或预防的具体疾病取决于所述感兴趣的多肽。当所述感兴趣的多肽是抗原性多肽时，所述药物组合物可以是疫苗。

药学上可接受的载体可以包括但不限于缓冲剂、赋形剂、稳定剂或防腐剂。药学上可接受的载体的实例是生理上相容的溶剂、分散介质、包衣、抗细菌和抗真菌剂、等渗剂和吸收延迟剂等，如盐、缓冲液、糖类、抗氧化剂、水性或非水性载体、防腐剂、润湿剂、表面活性剂或乳化剂或其组合。可以基于载体的活性和制剂的所需特性，如稳定性和/或最小氧化，通过实验确定药物组合物中的药学上可接受的载体的量。

本文中所述“细胞”可以是哺乳动物细胞，包括但不限于啮齿类动物细胞如小鼠细胞、大鼠细胞；灵长类动物细胞如猴细胞、人细胞。优选地，所述细胞是人细胞。

本文中所述“对象”可以是哺乳动物，包括但不限于啮齿类动物如小鼠或大鼠，灵长类动物如猴或人。优选地，所述对象是人。

实施例

通过参考本文中所述的具体实施例可以进一步理解本发明，这些实施例仅用于说明本发明，并不意图限制本发明的范围。显然，本领域技术人员知晓可以对本发明进行各种修改和变化而不脱离本发明的精神，因此这样的修改和变化也属于本发明的范围。

实施例 1、数据训练和生成非天然 UTR 序列的方法

使用公共数据库中已公开发表的人类细胞 Ribo-seq 数据和 RNA-seq 数据，训练了深度学习预测模型，该模型能够基于 mRNA 5'UTR 序列预测蛋白质翻译效率。随后，采用人工智能分析策略，包括对抗生成网络(GAN)和长短期记忆递归神经网络(LSTM)等技术，使用人类基因组中的 5'UTR 序列进行训练，构建生成式学习模型，从中产生非天然、新的 5'UTR 序列。将生成模型与预测模型结合，对生成的非天然 5'UTR 序列进行评估，预测由这些 5'UTR 及蛋白编码 ORF 组成的 RNA 翻译效率，从中筛选出一批可能实现较高蛋白翻译效率的 5'UTR 序列。然后通过湿试验排除实际表达中蛋白翻译效率不高的序列，最终得到的候选 5'UTR 序列见表 1。

利用类似方法筛选出候选 3'UTR，其序列见表 2。

表 1、候选 5'UTR

5'UTR	序列	SEQ ID NO:
seq1	ATAAGCCCGCAGGTTTTTGTTTTGTCGGTTAGGGGTTCCGTGACTCTAAGGCCACC	1
seq2	AGAGGTGTGTTTTCGGTACACCGGAGCTTGACCGGGTACGAGTGGGTGCAGCCACC	2
seq3	AACGAGCTGTGGTTTTGCGACCGGAGCCCGACTTCCTGGTTCGCCCAAGGGCCACC	3
seq4	AGTACTGCTGGTGTTGTACGGTTGTAGACGGCGGACTAAGTGTCGGAACCGCCACC	4
seq5	GTCCTTTCCTTTTCTTGCGGGCTCTCCGGGCCGCTCCAGCGGCTGCCTGTGCCACC	5
seq6	TCCTGACTGCCTGTTTTCGACTCGGGTTCTCTCTCGGCTTGGTGACGTTTGCCACC	6
seq7	TCTATTTTGGTTTTTGAGGAGGAGCTTTCGTTGTCGATGCGGCCGTGCCGGCCACC	7
seq8	GTGAAATCTTTTTTGCGATTTTGCAGTGGTGAGGAAGCTTTGGCGGCTGTGCCACC	8
seq9	GGAAGTAGAACTGGCTTGTGTCAGCTGTTTGGTGAACTGGCGGCGCGCAGGCCACC	9
seq10	GCTCCTTCGCGTCAGGTTCTCCTCTCTTTCTATCCTTCTTTCTCCTGTCTGCCACC	10
seq11	TGCAGGACCAGAAGCGCTCGCACAAGACGCAAAAAGCTCGCAGCGGGAGCGCCACC	11
seq12	TCGGCTGCTTGGGTTTTTCTTGGAGGAGTTCGCCTTTTTGCTGGGGGGTGGCCACC	12
seq13	GCTCTTTTGAATTGCTTTGTCTCTGGAAACATTTCTGGCCGGGATGGGCGGCCACC	13
seq28	GGACTTGGTTAGACTTGGAAGCAGTTCGGGTTTTTCTGGAGAGTCTCGCGGCCACC	41
seq29	TGCCTTTCAACGCAAGACTTTTCGAAAGTTTTTAGTAGTGCGAGTAAGTGGCCACC	42
seq30	GCAGAAGCTTTCCGCAGCGGAGTAGGTCTTGCGTGTGTTTGAAGCGGAGTGCCACC	43
seq31	GAAACACACGTGAGTTTCAGTCTAGTTTTTGGAGTGGTGCGAATTTCTCAGCCACC	44
seq32	GGAAACTTTTCGTGTTATTTGTTTGCGGGAGTTGTGTTAACTCTTGCGGAGCCACC	45
seq33	GAAAAAATTTTCTGTTTCGGCAGCAGTTGTTTTTGGGTTTCACAGCGAAGGCCACC	46

seq34	CCGATTCGTAGTCTTTTTGAGTGAGTTTTGCGGAGGAGTTAGTTTCTGAGGCCACC	47
seq35	GAACTTTTTTCGAGTACTTAGACTTGTGTGGGTTGTGTGTTTTTCCGGAAGCCACC	48
seq36	TCAAACAAAATTGTTGTGTTTTTGGAGTAGTGTTGTGTTTGTTTTTGAGCGCCACC	49
seq37	TTAGTAGTTTTGAGTGGGTGGAAGCGGAGCTTTTTGAGTTGGAAGTTTGCGCCACC	50
seq38	GACTTCCACAGCGGCTTCCCTCGACCCGTCAGCGAGCGTACCAGCTGTCGGCCACC	51
seq39	GAAGAACGCGCCGCTTCGCGAGCTCGCCGCACCCAGTCACCCCCGAAACGGCCACC	52
seq40	CGAATTCGTTCCCGCACGACTCAGCGCGAACCCCGACGCTGTCTCTTCCAGCCACC	53
seq41	CATACTTGCACGTCTCGCTCTTCCGTTGTCGCAGCGAACCACACTCCACCGCCACC	54
seq42	GCTTTGCTCTCAGCCACAGCAACCGCGCTCTCACGCTCCCGCGCCCTTCTGCCACC	55
seq43	CTTTCCAGTCCCTTCCTCACTCCCGTTCTTCGGCACCACTCGGACGCGAAGCCACC	56
seq44	CCGCTCTCAGCCCGCGCAAGCTCCGCGCGCTGCAACCCCGAACAACTCTCGCCACC	57
seq45	TCAGAAACCCCAACCCGCCACCTTCGGTCTCCCCACTCACGCGCGCAACTGCCACC	58
seq46	GCGCCTCTCGCCAGTTCACCCGTTCAGCAGCCGTCTGCGCAGCGTCTTACGCCACC	59
seq47	CTGCGTTGAGCGTCGCCGGTTGCGTTTCTCAGCAGCGAGCGAGCGAGACTGCCACC	60
Moderna_ctrl_5’	GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC	14

表 2、候选 3’UTR

3’UTR	序列	SEQ ID NO:
Seq14	TTAGAGTAAAGATAGCATGTTTACTTCCATCTACCGTACGGGAAAGCCTAACTTATGCTGACATTTGGAATGTATGTAAATAAAGTGGAGGAAGATAGAA	15
Seq15	GTTATATTCTTCCAGTACTGCTCTCTCATGTTTGTAAACTTGTGTTTATAATAAATTTGTAAATTTGCTTTCATATTGCAATAAACATGGAAATCTTAAA	16
Seq16	AGGTATCTGAGTTTGAATTTAACATAATGGAAATAGAAAGTACACTTTTAAACTTGTATAGTTGTAATGATATTGTATTTAATTTATTTTACATCTATAA	17
Seq17	CAACTTAACAATAATGCGTGCTACAGTATACATCTACTGCAACTCCATCTTGACCTTTGTATTTACTAGCATATTTCTAAATATACTATTGATACTCTCT	18
Seq18	GGATGCCTCCGCCTAGACGCCCGCCTCAGCCTTCACTGAATTTGTGTCAATCCACTCCTCAAACTCTTTGAGGACCTTTCTTTCTGTCTGTTTTGCCCTCTTTCTCCCTC	19
Seq19	GCCCGCTTTTTCCTTGTCTCTCCTCCCCTCACCCTCTGCTCGGTTTATTTGCATCTGCCTTCCAAGTCTTCCAAGAATCTGAATAGCTGAGGATCTCTACTCCCCCCTTT	20
Seq20	GCCCTCTCGCCTCTTCTCTTCGCCCTCAGCTAACAGGCCCTCCTGTGTGGCACCGCCCCTGAGGGGAACCCCACTTTGCATCTTTTCCGGCCGCAGGACTGCAGCTGGAG	21
Seq21	GGGCCCCGGCCTTCCTCTTTCTCATGTGCCCTTTTCTTCCTCCTGGTGCTGTCTGGGGCGGCGATTTCCTTGTCTCTCTGCATCCTTGTTCATCCCCTTACCTCCGTCCA	22

Seq22	GGCACCTGCCTCGGACCACCCACCCTCTCGCCCACTCTGGACCTGGTTTTCTGCCCATCATCTCTTCCCGCGCCGAAGGCCAGCCTCTGAAATGCTCACCTTTACTTCT	23
Seq23	TCCTTCCTGCTGCTGCCCACAACCGGAAGCCTTTTCTCTCCTTTCCTTCTGTAGGATATATTTGCTGGACGAGTTTGCTCCTGAAAGTAGAACACAAGATTGCCAAAGTC	24
Seq24	AGCTTCCACCAGTCTCCTGCTCTGCGGCCTGTCTTCCACATCCTCGACTCCCTCGGCTGTGTCCCTGCTGTGGTGCCACTCGCCTTCCCTCGAAGTTTCTGTTTTTCTCA	25
Seq25	GCTCCTCCGCCCTTCAGCACTCTGCAGCCGCCCCTCCCTCCTCTCCCCCTGCCCTTGTGGTGCCACAGCCCCAGTTGGCCTCCCTGACCTCTCCCCCTGTCCTCGTCGTC	26
Seq26	AAACTCTCCTTCCACCATGCCTCCGAAAGATGGGATACTATCGCTGCGCAATTCTGAGGAACCTCTGGAATACACTCTGCTGTGTGTCAAACAATCTCTTCGAATACTCC	27
Seq27	GCAGACCTCTACGCCTAGCTGCCTCTGTTCCCCGCTTCTCTCCCTTTCTCCGCACCCCTTTCTGAGCTGGCTCCTACCCATCCTGCCTGACCCGTTTTGACATTCTTCTA	28
seq48	ACTGCACTTCGTCATGGGCTACTAGTTGTGCACTTGGAAGAGGAGGGATCCCTTCAGAAAGCTTCTAACCTGTGAATACAAGAAGAATTTCATGGCGAATAAATCTTCTG	61
seq49	AGAATGGAGTGAGAAAAGATGAAAAGATCCATTCTATTGCTATTAAATCCAACGTTACTCTATGGTGTTTAAAGAATTTCATTTTCCAAATATCTGGAATGTTTTGCAAG	62
seq50	TGACAAGGAAGTGGGGGTGGGCAATAAAGTTCACTGAAACTAACAAAAACAGACCTTTGTATAGGTGATCTCAAATATTTCTGGTCTTTATCTGAATTATAGTGTAGTAT	63
seq51	AAAGATCTCTATGTTGATGACAAAAAGGAACAGCTACCCCAAACAAGGAGATCAGGTTGTTAAATGTTTGTAGATTGTAATTTGTTAAAAATAATTATGTGCTTTTTTAG	64
seq52	ACTGTATGTCAGCATGTGGAGCTGTAGTGACATTGACTAACTGTCTGTGCACATTGTCAAAGTTCACAAAAAATGGCTTTTCCAGGATTGCTTTACCAGGTCTTAATCGT	65
seq53	TGACTCAGAGAACCTTGACATCCATGTTTCTTGTGACCATGACCTCAGTGTACAAGTTATCGCAGGTGGTGTTAGCTTCAGATCATTTACTCAAGGAAGAGACATATTAT	66
seq54	TAGATAAAAGGCTGTGATTGGGTGAAAGGGGAAAATGAAGGAATAAAGGAAAGATCTAAAGCATTTCTATTACTTTATTTTTTCCTAAATGTTATATCAGTTTATATTCA	67
seq55	AATACTTGCCAAGAGTGAGCTAGCAGCGCAACTCGCCACCAGTGGCTGCAAATCGAAAGAAAACAACCAGGAGTTCTACTGTACACTGGTCATGGCATGTTCATTTTGGC	68
seq56	ACTGATCCACTCTTCTGTCACATCAATCAAGAAAACAGCCAGGCATCTTTACCAATATTTGTATTATATTAAATATAATATATGATTTAATTTAATTGTGAGTTGTGTAT	69
seq57	ACTCTTCCGAGCTGTGCTCATGCGATCTTGCTTGTGGTGTCTGTGGAGGATATCTGCCTGCTGGCAGTAAACTAGTTGCTAGCAATTTTGTGTTCGAGACTGCCTTGAAA	70
seq58	AGGGCTTGGCATCTTCAGTTTCTTCTTCTTTCTCTTGCAATCCCTTTATGATCTATTGTTAAGCAAATTCCTCAACATGAAAAATCAAAAGATCGTCTGTGAATCTTCTG	71
seq59	ATGTTCCACTTCAATCCTTACAGTTCTCAGCTCTGAATCTTTCTTATCTGGTTGTGTCGTTTGGGAAGAGAATCTTGTATTTGAATATGCTTCCCCTGCTTTTGTCTCTT	72

seq60	AGTTCTTTGTCTTCTGACTTCTTTTCCTAATCCTCTCCCTTGCTTAAATGTTTGCTTAAACACTTTAACACAGTAAATGGGTAGTCATTTTTTAGGAAATAATTTATGAA	73
seq61	AGTTCCTCGCTCATCTCTCTCATTTCTTCTCTTGCAATTATGAGCTTAGATGGCCATTTTGGGAATACTACTGTTAAATTACTTTTTCAAGCTACCCACCTTAAATTCAA	74
seq62	AGTACTTAACCTCTTTCACTTTGCTTCTGCCTTTGCTGCATCTGCTGCTTTAGCCACAGCTTTCTTTTGGTCTTCACTGAGTTAAATGAAACCATGGTGAGAAGAAAGTA	75
seq63	ACACTGTACTCTTTGCTGCTAGAAATTTCTCAGGGCTTCCTCCATCTTCTCCTCTTCTGTATATTCACGTGGCTTTGAGTTGTAAAACTGTTTTTACACAGTTGTTATTC	76
seq64	AGTGTGCTTCTTCACTATTCAAATGCCAGCCACATCTTCTGACTGCAGTCCTTTGTCAAATTGTTACCAAAGTTTTTTAACATTCTGAATTACCAGTAATAAAGTCTAGA	77
seq65	AGTCCCCTTTACTGATGTCTAGGATTTCACCAACTGACCTCAGGCTTCATCCTTTTATTTCTTATCAGAATTGTAATAGATGATTTAAAATTTTCATATTTGAAAAACCT	78
seq66	AAGGATTCTCTCTCTTTCATGATTCATTTTTACTTCCTGATAAGGCTTCTGCCACAACTTAATTTATGTCGTCATGCACTTCCACAGTAGAACTCAGACTCTATTGTTTC	79
seq67	TCATTGAATTCACTTCCGTTTACTATAACATGTTCCTGGATGTCTCAGTGTTGCTCTTCTCTATTATGGCATTCAGTTCTGTTCTTACATCTTCTCAACTACTTCTGGAA	80
Moderna_ctrl_3'	TGATAATAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGC	29

实施例 2、包含候选 UTR 序列元件的载体构建

为了测试候选 5'UTR，体外合成包含 T7 启动子、候选 5'UTR 或对照 5'UTR、编码萤火虫荧光素酶或红色荧光蛋白的序列(表 3)、3'UTR (SEQ ID NO:29)、包含 100 个 A 核苷酸残基的多聚(A)序列、和 IIS 型限制性核酸内切酶切割位点的核酸片段，并克隆到体外转录载体中(pIVTRup, Addgene plasmid #101362)。作为对照的 5'UTR 序列元件(SEQ ID NO:14)来源于 Moderna mRNA1273 expression vector, complete sequence(GenBank: OR134578.1)。3'UTR 序列元件(SEQ ID NO:29)来源于 Moderna mRNA1273 expression vector, complete sequence(GenBank: OR134578.1)。

为了测试候选 3'UTR，体外合成包含 T7 启动子、5'UTR(SEQ ID NO:14)、编码萤火虫荧光素酶或红色荧光蛋白的序列(表 3)、候选 3'UTR 或对照 3'UTR、包含 100 个 A 核苷酸残基的多聚(A)序列、和 IIS 型限制性核酸内切酶切割位点的核酸片段，并克隆到体外转录载体中。5'UTR 序列元件(SEQ ID NO:14)来源于 Moderna mRNA1273 expression vector, complete sequence(GenBank: OR134578.1)。作为对照的 3'UTR 序列元件(SEQ ID NO:29) 来源于 Moderna mRNA1273 expression vector, complete sequence(GenBank: OR134578.1)。

表 3、萤火虫荧光素酶和红色荧光蛋白序列

名称	序列	SEQ ID NO:

萤火虫荧光素酶氨基酸序列	MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTIAFTDAHIEVDI TYAEYFEMSVRLAEAMKRYGLNTNHRIVVCSENSLQFFMPVLGALFIGVAV APANDIYNERELLNSMGISQPTVVFVSKKGLQKILNVQKKLPIIQKIIIMDSK TDYQGFQSMYTFVTSHLPPGFNEYDFVPESFDRDKTIALIMNSSGSTGLPK GVALPHRTACVRFSHARDPIFGNQIIPDTAILSVVPFHHGFGMFTTLGYLIC GFRVVLMYRFEEELFLRSLQDYKIQSALLVPTLFSFFAKSTLIDKYDLSNLHEI ASGGAPLSKEVGEAVAKRFHLPGIRQGYGLTETTSAILITPEGDDKPGAVGK VVPFFEAKVVDLDTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALID KDGWLHSGDIAYWDEDEHFFIVDRLKSLIKYKGYQVAPAELESILLQHPNIF DAGVAGLPDDDAGELPAAVVVLEHGKTMTEKEIVDYVASQVTTAKKLRG GVVFVDEVPKGLTGKLDARKIREILIKAKKGGKIAV*	30
萤火虫荧光素酶核苷酸序列	ATGGAGGACGCCAAGAACATCAAGAAGGGCCCCGCCCCCTTCTACCC CCTGGAGGACGGCACCGCCGGCGAGCAGCTGCACAAGGCCATGAAG CGGTACGCCCTGGTGCCCGGCACCATCGCCTTCACCGACGCCCACAT CGAGGTGGACATCACCTACGCCGAGTACTTCGAGATGAGCGTGCGGC TGGCCGAGGCCATGAAGCGGTACGGCCTGAACACCAACCACCGGAT CGTGGTGTGCAGCGAGAACAGCCTGCAGTTCTTCATGCCCGTGCTGG GCGCCCTGTTCATCGGCGTGGCCGTGGCCCCCGCCAACGACATCTAC AACGAGCGGGAGCTGCTGAACAGCATGGGCATCAGCCAGCCCACCG TGGTGTTCGTGAGCAAGAAGGGCCTGCAGAAGATCCTGAACGTGCAG AAGAAGCTGCCCATCATCCAGAAGATCATCATCATGGACAGCAAGAC CGACTACCAGGGCTTCCAGAGCATGTACACCTTCGTGACCAGCCACC TGCCCCCCGGCTTCAACGAGTACGACTTCGTGCCCGAGAGCTTCGAC CGGGACAAGACCATCGCCCTGATCATGAACAGCAGCGGCAGCACCG GCCTGCCCAAGGGCGTGGCCCTGCCCCACCGGACCGCCTGCGTGCG GTTCAGCCACGCCCGGGACCCCATCTTCGGCAACCAGATCATCCCCG ACACCGCCATCCTGAGCGTGGTGCCCTTCCACCACGGCTTCGGCATGT TCACCACCCTGGGCTACCTGATCTGCGGCTTCCGGGTGGTGCTGATGT ACCGGTTCGAGGAGGAGCTGTTCCTGCGGAGCCTGCAGGACTACAAG ATCCAGAGCGCCCTGCTGGTGCCCACCCTGTTCAGCTTCTTCGCCAAG AGCACCCTGATCGACAAGTACGACCTGAGCAACCTGCACGAGATCGC CAGCGGCGGCGCCCCCCTGAGCAAGGAGGTGGGCGAGGCCGTGGCC AAGCGGTTCCACCTGCCCGGCATCCGGCAGGGCTACGGCCTGACCGA GACCACCAGCGCCATCCTGATCACCCCCGAGGGCGACGACAAGCCC GGCGCCGTGGGCAAGGTGGTGCCCTTCTTCGAGGCCAAGGTGGTGGA CCTGGACACCGGCAAGACCCTGGGCGTGAACCAGCGGGGCGAGCTG TGCGTGCGGGGCCCCATGATCATGAGCGGCTACGTGAACAACCCCGA GGCCACCAACGCCCTGATCGACAAGGACGGCTGGCTGCACAGCGGC GACATCGCCTACTGGGACGAGGACGAGCACTTCTTCATCGTGGACCG GCTGAAGAGCCTGATCAAGTACAAGGGCTACCAGGTGGCCCCCGCC GAGCTGGAGAGCATCCTGCTGCAGCACCCCAACATCTTCGACGCCGG CGTGGCCGGCCTGCCCGACGACGACGCCGGCGAGCTGCCCGCCGCC GTGGTGGTGCTGGAGCACGGCAAGACCATGACCGAGAAGGAGATCG TGGACTACGTGGCCAGCCAGGTGACCACCGCCAAGAAGCTGCGGGG CGGCGTGGTGTTCGTGGACGAGGTGCCCAAGGGCCTGACCGGCAAG CTGGACGCCCGGAAGATCCGGGAGATCCTGATCAAGGCCAAGAAGG GCGGCAAGATCGCCGTGTGA	31
红色荧光蛋白氨基酸序列	MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFQYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFED GGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASTER MYPEDGALKGEIKMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKL	32

	DITSHNEDYTIVEQYERAEGRHSTGA*	
红色荧光蛋白核苷酸序列	ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCGCTTCAAGGT GCGCATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGC GAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGA AGGTGACCAAGGGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCC CCTCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGA CATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGA GCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAG GACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCG CGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCA TGGGCTGGGAGGCCTCCACCGAGCGGATGTACCCCGAGGACGGCGC CCTGAAGGGCGAGATCAAGATGAGGCTGAAGCTGAAGGACGGCGGC CACTACGACGCCGAGGTCAAGACCACCTACATGGCCAAGAAGCCCGT GCAGCTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACATCACCT CCCACAACGAGGACTACACCATCGTGGAACAGTACGAGCGCGCCGA GGGCCGCCACTCCACCGGCGCCTAA	33

实施例 3、包含候选 UTR 序列元件的 mRNA 分子的产生

将实施例 2 获得的载体进行线性化处理，并使用 T7-RNA 聚合酶进行体外转录生产 mRNA 分子，并同时添加 5’-帽结构。5’-帽结构通过共转录加帽的方式，在体外转录过程中将帽类似物作为第一核苷酸掺入转录本，直接生产带有 Cap1 结构的 mRNA 分子。图 1 示出包含本发明的候选 UTR 元件的示例性 mRNA 的序列。

将这样得到的 mRNA 分子纯化并重悬在水中。使用 5200 片段分析仪(Agilent)对 mRNA 分子进行质检来检测 mRNA 分子长度和完整度(数值来源于预期长度片段的曲线下面积占比)符合要求(见图 2)，即可用于后续对于不同候选 UTR 序列元件的测试。

实施例 4、通过荧光素酶报告系统鉴定候选 5’UTR 序列元件对翻译效率的作用

本实施例测试候选的 5’UTR 序列元件是否对翻译具有正向的调控作用，通过如下方法进行：

1) 用含有候选 5’UTR 序列元件的 RNA 分子(编码蛋白序列为萤火虫荧光素酶)转染哺乳动物细胞 293T；

2) 在转染后的特定时间点(24 小时)检测荧光素酶的化学发光光吸收值，代表其蛋白表达量；

3) 用包含候选 5’UTR 序列元件实验组的光吸收值读数减去背景组的光吸收值，得到该实验组的相对光单位。

每个候选 5’UTR 序列元件测得的相对光单位可指征该候选 5’UTR 序列元件的翻译调控能力。

用于确定翻译效率的荧光素酶化学发光光吸收值测定方式：

将 293T 人胚肾细胞以 5×10^4 个细胞/孔的密度接种在 48 孔板中。次日，将细胞在 Opti-MEM 中洗涤，并随后用 200 ng/孔的在 Opti-MEM 中的 Lipofectamine2000-复合的

包含候选 UTR 序列元件的编码萤火虫荧光素酶的 mRNA 转染。没有加入任何 RNA 分子的细胞用作为背景组。转染 6 小时后，吸去混合培养基，替换为完全培养基。转染 24 小时后，吸去培养基，加入 100 μl 的裂解缓冲液(Promega)，裂解 5 分钟后，进行检测读取吸光值。

在多孔板读数器(Agilent)中以相对光单位(RLU)测量荧光素酶活性。在荧光素酶测定中从单个样品顺序测量荧光素酶的活性。吸取 20 μl 裂解液，加入 50 μl 含有萤火虫荧光素酶底物的缓冲液，震板混匀，检测光吸收值。

5'UTR 测试结果见下表 4 和图 3。

表 4、候选 5'UTR 的 RLU 比值

5'UTR	RLU Mean ratio(n=3)
Moderna-ctrl	1.00
seq1	1.35
seq2	1.03
seq3	1.42
seq4	2.83
seq5	1.67
seq6	1.12
seq7	1.85
seq8	1.68
seq9	1.55
seq10	1.60
seq11	1.37
seq12	1.08
seq13	1.27
seq28	2.22
seq29	1.30
seq30	2.00
seq31	2.14
seq32	1.58
seq33	1.28
seq34	1.11
seq35	1.71
seq36	1.03
seq37	1.77
seq38	3.47
seq39	1.22
seq40	2.34
seq41	1.10
seq42	3.10
seq43	1.49
seq44	1.92
seq45	1.10
seq46	2.14
seq47	1.39

实施例 5、通过红色荧光蛋白报告系统鉴定候选 5'UTR 序列元件对翻译效率的作用

本实施例利用荧光蛋白实验测试候选的 5'UTR 序列元件是否对翻译具有正向的调控作用，通过如下方法进行：

1) 用含有一个候选 5'UTR 序列元件的 RNA 分子(编码蛋白序列为红色荧光蛋白)转染哺乳动物细胞 293T；

2) 在转染后的特定时间点(24 小时)通过流式细胞技术检测荧光蛋白光强，代表其蛋白表达量；

3) 用红色荧光蛋白的光强的除以内参(绿色荧光蛋白光强)，得到实验组的相对光强。每个候选 5'UTR 序列元件得到的相对光强可指征该候选 5'UTR 序列元件的翻译调控能力。

用于确定翻译效率的流式细胞技术测定荧光光强方式：

将 293T 人胚肾细胞以 2×10^5 个细胞/孔的密度接种在 24 孔板中。次日，将细胞在 Opti-MEM 中洗涤，并随后用 200 ng/孔的在 Opti-MEM 中的 Lipofectamine2000-复合的包含候选 5'UTR 序列元件的编码红色荧光蛋白的 mRNA 转染，并共转染等量的绿色荧光蛋白的 mRNA 作为内参。转染 6 小时后，吸去混合培养基，替换为完全培养基。转染 24 小时后，吸去培养基，加入胰酶消化 3 分钟，中和并收细胞样品。以 PBS 为悬液，在流式细胞仪中进行细胞分析，读取红色荧光、绿色荧光的光强。

作为内参的绿色荧光蛋白的氨基酸序列如下所示(SEQ ID NO:34)：

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPT
LVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN
RIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQ
QNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK*

作为内参的绿色荧光蛋白的序列如下所示(SEQ ID NO:35)：

*GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC*ATGGTGAGCAAGG
GCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACC
CTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACC
CTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTC
TTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGAC
GGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCAT
CGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGG
TGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACC
AGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCA
CCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGT
TCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA

5'UTR 测试结果见下表 5 和图 4。

表 5、候选 5'UTR 的相对光强

5'UTR	RFP/GFP 荧光强度 mean (n=3)

Moderna_ctrl	81.09
seq1	131.64
seq2	135.11
seq3	108.61
seq4	131.32
seq5	118.13
seq6	N/A
seq7	N/A
seq8	170.33
seq9	129.33
seq10	167.56
seq11	N/A
seq12	N/A
seq13	N/A

N/A：未进行实验

本发明的候选 5’UTR 相比对照 5’UTR 能够获得相当或更优的翻译效率。

实施例 6、通过荧光素酶报告系统鉴定候选 3’UTR 序列元件对翻译效率的作用

本实施例测试候选的 3’UTR 序列元件是否对翻译具有正向的调控作用，通过如下方法进行：

1) 用含有候选 3’UTR 序列元件的 RNA 分子(编码蛋白序列为萤火虫荧光素酶)转染有代表性的人细胞系人胚胎肾细胞系 HKE293T 和人结肠癌细胞系 HCT116；

2) 在转染后的特定时间点(24 小时)检测荧光素酶的化学发光光吸收值，代表其蛋白表达量；

3) 用包含候选 3’UTR 序列元件实验组的光吸收值读数减去背景组的光吸收值，得到该实验组的相对光单位。

每个候选 3’UTR 序列元件测得的相对光单位可指征该候选 3’UTR 序列元件的翻译调控能力。

用于确定翻译效率的荧光素酶化学发光光吸收值测定方式：

将 HEK293T 人胚肾细胞或 HCT116 人结肠癌细胞系，以 5×10^4 个细胞/孔的密度接种在 48 孔板中。次日，将细胞在 Opti-MEM 中洗涤，并随后用 200 ng/孔的在 Opti-MEM 中的 Lipofectamine2000-复合的包含候选 3’UTR 序列元件的编码萤火虫荧光素酶的 mRNA 转染。没有加入任何 RNA 分子的细胞用作为背景组。转染 6 小时后，吸去混合培养基，替换为完全培养基。转染 24 小时后，吸去培养基，加入 100 μl 的裂解缓冲液(Promega)，裂解 5 分钟后，进行检测读取吸光值。

在多孔板读数器(Agilent)中以相对光单位(RLU)测量荧光素酶活性。在荧光素酶测定中从单个样品顺序测量荧光素酶的活性。吸取 20 μl 裂解液，加入 50 μl 含有萤火虫荧光素酶底物的缓冲液，震板混匀，检测光吸收值。

3’UTR 测试结果见下表 6A、表 6B 和图 5A 和图 5B。

表 6A、候选 3’UTR 的 RLU 比值(HEK293T 细胞系)

3'UTR	RLU mean ratio(n=3)
Moderna_ctrl	1.00
Seq14	1.59
Seq15	1.62
Seq16	1.28
Seq17	1.69
Seq18	1.60
Seq19	1.19
Seq20	0.93
Seq21	1.59
Seq22	0.97
Seq23	0.89
Seq24	1.40
Seq25	1.28
Seq26	1.05
Seq27	1.57
seq48	0.95
seq49	1.26
seq50	1.58
seq51	2.51
seq52	1.78
seq53	1.56
seq54	1.21
seq55	1.62
seq56	1.36
seq57	1.51
seq58	1.27
seq59	1.04
seq60	2.20
seq61	2.30
seq62	2.53
seq63	3.18
seq64	2.99
seq65	1.00
seq66	2.50
seq67	2.69

表 6B、候选 3'UTR 的 RLU 比值(HCT116 细胞系)

3'UTR	RLU mean ratio(n=3)
Moderna_ctrl	1.00
Seq18	1.95
Seq19	1.23
Seq20	1.08
Seq21	1.48
Seq22	1.33
Seq23	1.12
Seq24	1.51
Seq25	1.26

Seq26	1.00
Seq48	1.61
Seq58	2.57
Seq60	2.48
Seq61	1.30
Seq62	2.32
Seq63	2.03
Seq64	2.18
Seq65	2.08
Seq66	1.30
Seq67	3.00

实施例 7、通过红色荧光蛋白报告系统鉴定候选 3'UTR 序列元件对翻译效率的作用

本实施例利用荧光蛋白实验测试候选的 3'UTR 序列元件是否对翻译具有正向的调控作用，通过如下方法进行：

1) 用含有一个候选 3'UTR 序列元件的 RNA 分子(编码蛋白序列为红色荧光蛋白)转染哺乳动物细胞 293T；

2) 在转染后的特定时间点(24 小时)通过流式细胞技术检测荧光蛋白光强，代表其蛋白表达量；

3) 用红色荧光蛋白的光强的除以内参(绿色荧光蛋白光强)，得到实验组的相对光强。

每个候选 3'UTR 序列元件得到的相对光强可指征该候选 3'UTR 序列元件的翻译调控能力。

用于确定翻译效率的流式细胞技术测定荧光光强方式：

将 293T 人胚肾细胞以 2×10^5 个细胞/孔的密度接种在 24 孔板中。次日，将细胞在 Opti-MEM 中洗涤，并随后用 200 ng/孔的在 Opti-MEM 中的 Lipofectamine2000-复合的包含候选 UTR 序列元件的编码红色荧光蛋白的 mRNA 转染，并共转染等量的绿色荧光蛋白的 mRNA(SEQ ID NO:35)作为内参。转染 6 小时后，吸去混合培养基，替换为完全培养基。转染 24 小时后，吸去培养基，加入胰酶消化 3 分钟，中和并收细胞样品。以 PBS 为悬液，在流式细胞仪中进行细胞分析，读取红色荧光、绿色荧光的光强。

3'UTR 测试结果见下表 7 和图 6。

表 7、候选 3'UTR 的相对光强

3'UTR	RFP/GFP 荧光强度 mean (n=3)
Moderna_ctrl	81.09
Seq14	175.08
Seq15	188.50
Seq16	145.14
Seq17	131.64
Seq18	118.13
Seq19	N/A
Seq20	N/A
Seq21	148.96
Seq22	N/A

Seq23	N/A
Seq24	175.92
Seq25	N/A
Seq26	N/A
Seq27	N/A

本发明的候选 3’UTR 相比对照 3’UTR 能够获得相当或更优的翻译效率。

实施例 8、向候选 3’UTR 中插入 microRNA 结合位点不影响其功能

在自然界中，许多 mRNA 分子的 3’UTR 区域含有 microRNA 结合位点，通过这些位点 microRNA 能够实现对 mRNA 的精准调控，包括翻译抑制和促进 mRNA 降解。鉴于 microRNA 在不同组织和疾病阶段表现出的显著表达差异，它们常被视为高度特异的生物标志物。因此，将 microRNA 结合位点插入到 3’UTR 区域是调控 mRNA 功能的常用策略之一。理论上，miRNA 与其靶 mRNA 的相互作用主要依赖于 miRNA 的种子序列(通常为 2 到 8 个核苷酸)与 mRNA 的 3’UTR 中相应位点的完全互补。这种互补配对足以引导 RNA 诱导的沉默复合体(RISC)到目标 mRNA 上，从而抑制其翻译或促进其降解。大量研究证实除了种子序列匹配外，miRNA 序列与 mRNA 之间额外的互补配对也能增强 miRNA 介导的基因调控效果。这种额外的互补配对通常发生在 miRNA 种子序列之外的区域，这些区域的匹配可以提高 miRNA 与其靶标的亲和力，从而增加 mRNA 抑制的效率。例如，Moderna 公司采用的一种策略是在 3’UTR 区域插入 miRNA 的全长反向互补序列。这种设计使得 miRNA 能够与其靶 mRNA 发生更广泛的互补配对，不仅限于种子序列，从而增强了 miRNA 对 mRNA 的调控效力。

本实施例利用在细胞内快速降解的荧光蛋白测试候选 3’UTR 序列元件在改造后功能是否发生改变，确认其对翻译具有正向调控功能，以及 mRNA 分子稳定性的调控作用。方法如下：

1) 向候选 3’UTR(SEQ ID NO: 25)中的 3 个随机位点插入全长的 hsa-microRNA-122-5p 反向互补序列，得到新的 3’UTR(SEQ ID NO:38)。
2) 用含有新的 3’UTR 及原始 3’UTR 序列元件的 RNA 分子 (均编码 1 小时降解的绿色荧光蛋白，d1EGFP，该蛋白寿命约 1 小时，当 mRNA 不再翻译为蛋白的 1 小时后，则不存在绿色荧光，因此蛋白荧光可以直接指征 mRNA 的稳定性)，转染哺乳动物细胞系，如 293T、Hela、raw264.7 等。
3) 在转染后的 24 小时内，每间隔一小时对细胞进行拍摄，检测荧光蛋白光强，代表其蛋白表达量；
4) 用绿色荧光蛋白的光强的除以内参(红色荧光蛋白光强)，得到实验组或对照组的相对光强。相对光强可指征该 3’UTR 序列的翻译调控能力，且在 24 小时内的荧光强度变化可指征该 3’UTR 序列对于 mRNA 分子稳定性的影响。

全长的 hsa-microRNA-122-5p 的序列如下所示(SEQ ID NO:36)：

TGGAGTGTGACAATGGTGTTTG

全长的 hsa-microRNA-122-5p 的反向互补序列如下所示(SEQ ID NO:37)：

CAAACACCATTGTCACACTCCA

向 3'UTR-seq24 中随机三个位点插入全长的 hsa-microRNA-122-5p 的反向互补序列后得到的新的 3'UTR 的序列如下所示(SEQ ID NO:38)：

CAAACACCATTGTCACACTCCAAGCTTCCACCAGTCTCCTGCTCTGCGGCCCAAACACCATTGTCACACTC
CATGTCTTCCACATCCTCGACTCCCTCGGCTGTGTCCCTGCTGTGGTGCCACTCGCCCAAACACCATTGTC
ACACTCCATTCCCTCGAAGTTTCTGTTTTTCTCA

1 小时降解的绿色荧光蛋白(d1EGFP)的氨基酸序列如下所示(SEQ ID NO:39)

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQC
FSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKL
EYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN
EKRDHMVLLEFVTAAGITLGMDELYKKLSHGFPPAVAAQDDGTLPMSCAQESGMDRHPAACASARINV*

包含 1 小时降解的绿色荧光蛋白，和向 3'UTR-seq24 中随机三个位点插入全长的 hsa-microRNA-122-5p 的反向互补序列后得到的新的 3'UTR 的序列的 mRNA 的序列如下所示(SEQ ID NO:40)：

GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACCATGGTGAGCAAGGGCGAGGAGCTG
TTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGG
CGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG
TGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATG
AAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGA
CGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGA
AGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAAC
GTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG
CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTT
CCCGCCGGCGGTGGCGGCGCAGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACC
GTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGTAGCAAACACCATTGTCACACTCCAAGCTTCCA
CCAGTCTCCTGCTCTGCGGCCCAAACACCATTGTCACACTCCATGTCTTCCACATCCTCGACTCCCTCGGC
TGTGTCCCTGCTGTGGTGCCACTCGCCCAAACACCATTGTCACACTCCATTCCCTCGAAGTTTCTGTTTTT
CTCA

实验结果表明，向候选 3'UTR 中插入 microRNA 结合位点不影响其功能(图 7 和 8)。

权 利 要 求 书

1. 一种5'UTR (5'非翻译区),其包含与SEQ ID NO:1-13和41-60之一具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列相同性的核苷酸序列或其互补序列，优选地，所述5'UTR包含SEQ ID NO:1-13和41-60之一所示的核苷酸序列或其互补序列；更优选地，所述5'UTR包含SEQ ID NO:1、3-5、7-11、13、41-46、48、50-53、55-57、59-60之一所示的核苷酸序列或其互补序列；更优选地，所述5'UTR包含SEQ ID NO:4-5、7-10、41、43-45、48、50-51、53、55、57、59之一所示的核苷酸序列或其互补序列。

2. 一种3'UTR (3'非翻译区)，其

i) 包含与SEQ ID NO:15-28和61-80之一具有至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%序列相同性的核苷酸序列或其互补序列，优选地，所述3'UTR包含SEQ ID NO:15-28和61-80之一所示的核苷酸序列或其互补序列；更优选地，所述3'UTR包含SEQ ID NO:15-20、22-23、25-26、28、61-71、73-80之一所示的核苷酸序列或其互补序列；更优选地，所述3'UTR包含SEQ ID NO:15-16、18-19、22、25、28、61、63-66、68、70-71、73-80之一所示的核苷酸序列或其互补序列；或

ii) 包含具有插入的一个或多个额外序列的SEQ ID NO:15-28和61-80之一所示的核苷酸序列或其互补序列，优选所述额外序列为miRNA结合位点，更优选额外序列选自全长的microRNA反向互补序列或其种子序列的反向互补序列，更优选所述全长的microRNA反向互补序列长度为19-25 nt或种子序列的反向互补序列长度为7-8 nt。

3. 一种用于表达感兴趣的多肽的RNA分子，其包含权利要求1的5'UTR和/或权利要求2的3'UTR。

4. 权利要求3的RNA分子，其是信使RNA(mRNA)分子或环形RNA分子。

5. 权利要求3或4的RNA分子，其中所述RNA分子从5'至3'方向按以下顺序包含

1) 权利要求1的5'UTR；和

2) 感兴趣的多肽的编码序列。

6. 权利要求3或4的RNA分子，其中所述RNA分子从5'至3'方向按以下顺序包含

1) 权利要求1的5'UTR；

2) 感兴趣的多肽的编码序列；和

3) 3'UTR。

7. 权利要求6的RNA分子，其中所述3'UTR是权利要求2所述的3'UTR。

8. 权利要求3或4的RNA分子，其中所述RNA分子从5'至3'方向按以下顺序包含

1] 感兴趣的多肽的编码序列；和

2] 权利要求 2 的 3’UTR。

9. 权利要求 3 或 4 的 RNA 分子，其中所述 RNA 分子从 5’至 3’方向按以下顺序包含

1] 5’UTR；

2] 感兴趣的多肽的编码序列；和

3] 权利要求 2 的 3’UTR。

10. 权利要求 9 的 RNA 分子，其中所述 5’UTR 是权利要求 1 所述的 5’UTR。

11. 权利要求 3-10 中任一项的 RNA 分子，其还包含多聚(A)序列。

12. 权利要求 11 的 RNA 分子，其中所述多聚(A)序列包含大约 20 个至大约 500 个腺嘌呤核苷酸(A)，例如，包含大约 25 个、大约 50 个、大约 100 个、大约 150 个、大约 175 个、大约 200 个、大约 300 个、大约 400 个、大约 500 个腺嘌呤核苷酸(A)。

13. 权利要求 3-12 中任一项的 RNA 分子，其中所述 RNA 分子是 mRNA 分子，其还包括 5’帽结构，例如 Cap1 结构。

14. 权利要求 3-13 中任一项的 RNA 分子，其中所述 RNA 分子是化学合成的或所述 RNA 分子是通过体外转录获得的。

15. 权利要求 3-14 中任一项的 RNA 分子，其中所述 RNA 分子还可以包含至少一种核苷酸修饰，优选地，核苷酸修饰选自 5-甲基胞嘧啶(m5C)、N6-甲基腺苷(m6A)、假尿苷(ψ)、N1-甲基假尿苷(m1ψ)和 5-甲氧基尿苷(5moU)中的一种或其组合。

16. 权利要求 3-15 中任一项的 RNA 分子，与相应的包含对照 UTR 的 RNA 分子相比，所述 RNA 在细胞中导致所述感兴趣的多肽相当的表达或增加的表达，优选地，所述感兴趣的多肽的表达增加大约 10%、大约 20%、大约 30%、大约 40%、大约 50%、大约 60%、大约 70%、大约 80%、大约 90%、大约 100%或更多。

17. 一种核酸载体，其包含权利要求 3-16 中任一项的 RNA 分子的编码序列。

18. 权利要求 17 的核酸载体，其中所述核酸载体还包含与所述 RNA 分子的编码序列可操作地连接的 RNA 聚合酶启动子序列。

19. 权利要求 18 的核酸载体，其中所述启动子选自 T7 RNA 聚合酶启动子、T6 病毒 RNA 聚合酶启动子、SP6 病毒 RNA 聚合酶启动子、T3 病毒 RNA 聚合酶启动子或 T4 病毒 RNA 聚合酶启动子。

20. 包含权利要求 3-16 中任一项的 RNA 分子或权利要求 17-19 中任一项的核酸载体的细胞。

21. 一种提高细胞中感兴趣多肽的表达的方法，所述方法包括向所述细胞导入权利要求 3-16 中任一项所述的 RNA 分子和/或权利要求 17-19 中任一项所述核酸载体。

22. 一种药物组合物，其包含权利要求 3-16 中任一项所述的 RNA 分子、和/或权利要求 17-19 中任一项所述核酸载体和/或权利要求 20 所述宿主细胞，以及药物可接受的载体。

23. 权利要求 1 的 5’UTR 和/或权利要求 2 的 3’UTR 用于提高包含感兴趣的多肽的编码序列的 RNA 分子中所述感兴趣多肽的翻译效率的用途。