UTR FOR PROMOTING RNA TRANSLATION

TECHNICAL FIELD

The present disclosure relates to the field of biomedicine, and in particular, to the field of RNA drugs. In particular, the present disclosure relates to a UTR obtained using artificial intelligence technology for promoting RNA translation, an RNA molecule comprising the UTR, and use of the UTR.

BACKGROUND

RNA (e.g., mRNA or circular RNA) drug molecules have unique advantages. Compared with the conventional DNA gene therapy methods, RNA (e.g., mRNA or circular RNA) drugs do not pose the safety risk of exogenous genome integration into the host genome. In addition, RNA (e.g., mRNA or circular RNA) can also provide immediate and regulatable protein expression, making it an ideal choice for treating diseases caused by the deficiency of particular genes. Meanwhile, RNA (e.g., mRNA or circular RNA) can also be used as an antigen to induce immune responses, providing a new approach for the development of vaccines.

The UTR of mRNA plays an important regulatory role in mRNA expression. Different UTR sequences can directly affect the expression level of proteins encoded by mRNA by affecting mRNA stability and translation efficiency. Therefore, optimization and screening of the 5' UTR sequences to increase the expression level of the proteins encoded by mRNA are critical for enhancing therapeutic effects and activating vaccine immune responses.

The UTR sequences in current mRNA drugs are generally obtained by screening from naturally occurring genomes. However, this screening method has some limitations that limit the further improvement of mRNA expression levels and therapeutic effects. First, naturally occurring UTR sequences generally have relatively conserved characteristics, and their evolution is affected by a variety of selection pressures, including transcriptional regulation, translation efficiency, RNA stability, and the like. Therefore, these sequences may have certain limitations in terms of regulatory mechanisms and translation efficiency and cannot meet the requirement for high-level gene expression. Second, naturally occurring UTR sequences are generally optimized for regulation of particular genes and species. When these sequences are applied to other genes or species, problems such as poor expression levels, instability, or the induction of excessive immune responses may be encountered. This limits the universality and application scope of UTR sequences. In addition, the current screening methods mainly rely on practical laboratory operations and conventional biological methods,

which require a significant investment of time and resources, and have certain randomness and limitations, so that the conventional screening methods become cumbersome and difficult to implement efficiently.

SUMMARY

In the present disclosure, by applying artificial intelligence (AI) technology, UTR sequences with optimized translation efficiency are systematically screened and designed. The AI algorithm can analyze the published human RNA-seq omics information and Ribo-seq omics information, and establish a deep learning prediction model to predict the translation efficiency of different UTR sequences. On this basis, a generative learning model is constructed by using an artificial intelligence analysis strategy, including technologies such as the generative adversarial network (GAN) and long short-term memory (LSTM) recurrent neural networks, and by using UTRs in the human genome as a training set, to generate non-natural UTR sequences. Further, experimental verification is performed *in vitro* by *in vitro* transcription technology. The expression level of the target protein is determined, and UTRs that actually have the effect of enhancing the expression of the target protein are screened.

Through the screening and optimization processes of the present disclosure, a series of 5' UTR and 3' UTR sequences with enhanced expression levels of the target protein can be obtained, which can enhance the therapeutic effect in RNA therapy and enhance immune activation when used as a vaccine. This brings a new breakthrough in the fields of treatment of diseases caused by the deficiency of particular genes and vaccine research and development.

Thus, one aspect of the present disclosure relates to a 5' UTR (5' untranslated region) comprising a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to one of SEQ ID NOs: 1-13 and 41-60 or a complementary sequence thereof, preferably, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 1-13 and 41-60 or a complementary sequence thereof; more preferably, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 1, 3-5, 7-11, 13, 41-46, 48, 50-53, 55-57, and 59-60 or a complementary sequence thereof; more preferably, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 4-5, 7-10, 41, 43-45, 48, 50-51, 53, 55, 57, and 59 or a complementary sequence thereof.

In another aspect, the present disclosure relates to a 3' UTR (3' untranslated region), wherein i) the 3' UTR comprises a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity

to one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof; preferably, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof; more preferably, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-20, 22-23, 25-26, 28, 61-71, and 73-80 or a complementary sequence thereof; more preferably, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-16, 18-19, 22, 25, 28, 61, 63-66, 68, 70-71, and 73-80 or a complementary sequence thereof; or

ii) the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-28 and 61-80 having one or more additional sequences inserted or a complementary sequence thereof; preferably, the additional sequences are miRNA binding sites; more preferably, the additional sequences are selected from a full-length microRNA reverse complementary sequence and a reverse complementary sequence of a seed sequence thereof; more preferably, the full-length microRNA reverse complementary sequence is 19-25 nt in length, or the reverse complementary sequence of the seed sequence is 7-8 nt in length.

In another aspect, the present disclosure relates to an RNA molecule for expressing a polypeptide of interest, which comprises the 5' UTR of the present disclosure and/or the 3' UTR of the present disclosure.

In another aspect, the present disclosure relates to a nucleic acid vector comprising a coding sequence of the RNA molecule of the present disclosure.

In another aspect, the present disclosure relates to a cell comprising the RNA molecule of the present disclosure or a nucleic acid vector of the present disclosure.

In another aspect, the present disclosure relates to a method for increasing expression of a polypeptide of interest in a cell, which comprises introducing the RNA molecule of the present disclosure and/or the nucleic acid vector of the present disclosure into the cell.

In another aspect, the present disclosure relates to a pharmaceutical composition comprising the RNA molecule of the present disclosure, the nucleic acid vector of the present disclosure, and/or the host cell of the present disclosure, and a pharmaceutically acceptable carrier.

In another aspect, the present disclosure relates to use of the 5' UTR of the present disclosure and/or the 3' UTR of the present disclosure for increasing translation efficiency of a polypeptide of interest in an RNA molecule comprising a coding sequence of the polypeptide of interest.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows the sequences of an exemplary mRNA comprising the UTRs of the present disclosure; A: 5' UTR; B: 3' UTR.

FIG. 2 shows the length and completeness of the mRNA molecule (the sequences shown in FIG. 1) as measured by a 5200 fragment analyzer (Agilent).

FIG. 3 shows the relative light units of firefly luciferase in the lysates of cells transfected with each mRNA comprising the candidate 5' UTR. Differences were statistically analyzed based on the Student's t-test.

FIG. 4 shows the relative light intensity of the red fluorescent protein in the lysates of cells transfected with each mRNA comprising the candidate 5' UTR measured by flow cytometry, with differences statistically analyzed based on the Student's t-test.

FIG. 5 shows the relative light units of firefly luciferase in the lysates of cells transfected with each mRNA comprising the candidate 3' UTR.

FIG. 6 shows the relative light intensity of the red fluorescent protein in cells transfected with each mRNA comprising the candidate 3' UTR measured by flow cytometry.

FIG. 7 shows the relative light intensity of the green fluorescent protein after the 293T cells were transfected with each mRNA comprising the 3' UTR of the present disclosure comprising 3 microRNA binding sites, and the trend of change within 24 h determined by quantitative fluorescence analysis.

FIG. 8 shows the relative light intensity of the green fluorescent protein after the Hela cells were transfected with each mRNA comprising the 3' UTR of the present disclosure comprising 3 microRNA binding sites, and the trend of change within 24 h determined by quantitative fluorescence analysis.

DETAILED DESCRIPTION OF INVENTION

In the present disclosure, unless otherwise defined, the scientific and technical terms used herein have the meanings generally understood by those skilled in the art. In addition, the terms and laboratory procedures related to protein and nucleic acid chemistry, molecular biology, cell and tissue culture, microbiology, and immunology used herein are the terms and conventional procedures widely used in the corresponding fields.

As used herein, the term "and/or" encompasses all combinations of the items connected by the term, and each combination should be deemed to have been individually listed herein. For example, "A and/or B" encompasses "A", "A and B", and "B". For example, "A, B, and/or C" encompasses "A", "B", "C", "A and B", "A and C", "B and C", and "A and B and C".

"Polynucleotide", "nucleic acid sequence", "nucleotide sequence", and "nucleic acid fragment" are used interchangeably and are single- or double-stranded RNA or DNA polymers, optionally containing synthetic, non-natural, or altered nucleotide bases.

Nucleotides are referred to by their single-letter designations as follows: “A” is adenosine or deoxyadenosine (corresponding to RNA or DNA, respectively), “C” represents cytidine or deoxycytidine, “G” represents guanosine or deoxyguanosine, “U” represents uridine, “T” represents deoxythymidine, “R” represents purine (A or G), “Y” represents pyrimidine (C or T), “K” represents G or T, “H” represents A or C or T, “I” represents inosine, and “N” represents any nucleotide. Although the nucleotide sequences herein may be represented as DNA sequences (containing T), when referring to RNA, those skilled in the art can readily determine the corresponding RNA sequences (i.e., by replacing T with U).

“Polypeptide”, “peptide”, and “protein” are used interchangeably in the present disclosure to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residues are artificial chemical analogs of corresponding naturally occurring amino acids, as well as to naturally occurring amino acid polymers. The terms “polypeptide”, “peptide”, “amino acid sequence”, and “protein” may also include modifications including, but not limited to, glycosylation, lipid attachment, sulfation, γ-carboxylation of glutamic acid residues, hydroxylation, and ADP-ribosylation.

When the term “comprise” or “comprising” is used to describe the sequence of a protein or nucleic acid herein, the protein or nucleic acid may consist of the sequence, or may have additional amino acids or nucleotides at one or both ends of the protein or nucleic acid while still retaining the activity described in the present disclosure.

The “sequence identity” between two polypeptide sequences or two polynucleotide sequences refers to the percentage of amino acids or nucleotides that are identical between the sequences. Methods for assessing the level of sequence identity between polypeptide or polynucleotide sequences are known in the art. Sequence identity can be assessed using a variety of known sequence analysis software. For example, sequence identity can be assessed by the online alignment tool of EMBL-EBI (https://www.ebi.ac.uk/Tools/psa/). Sequence identity between two sequences can be assessed using the Needleman-Wunsch algorithm with default parameters. Sequence identity may be the identity along the full length of a given sequence.

In one aspect, the present disclosure relates to a 5' UTR (5' untranslated region) comprising a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to one of SEQ ID NOs: 1-13 and 41-60 or a complementary sequence thereof. In some embodiments, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 1-13 and 41-60 or a complementary sequence thereof. In some preferred embodiments, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 1, 3-5, 7-11, 13, 41-46, 48, 50-53,

55-57, and 59-60 or a complementary sequence thereof. In some more preferred embodiments, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 4-5, 7-10, 41, 43-45, 48, 50-51, 53, 55, 57, and 59 or a complementary sequence thereof.

"5' UTR" generally refers to the sequence between the 5' end and the translation initiation codon in an mRNA molecule and is capable of recruiting ribosomal complexes and initiating translation of the mRNA. The 5' UTR regulates post-transcriptional modification, formation and stability of translation initiation complexes, and other processes by interacting with transcription factors, ribosomes, and other transcriptional regulatory proteins.

In some embodiments of various aspects of the present disclosure, the 5' UTR is a non-natural 5' UTR. In some embodiments, the 5' UTR is a 5' UTR designed by artificial intelligence.

In another aspect, the present disclosure relates to a 3' UTR (3' untranslated region) comprising a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof. In some embodiments, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof. In some preferred embodiments, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-20, 22-23, 25-26, 28, 61-71, and 73-80 or a complementary sequence thereof. In some more preferred embodiments, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-16, 18-19, 22, 25, 28, 61, 63-66, 68, 70-71, and 73-80 or a complementary sequence thereof.

As used herein, "3' UTR" refers to the sequence between the termination codon of a polypeptide coding sequence and a poly(A) sequence in an mRNA. The 3' UTR can regulate the translation of mRNA by interacting with mRNA binding proteins, miRNAs, and the like. Sequence and structural characteristics of the 3' UTR can affect the mRNA stability, the ribosome scanning, the formation of translation termination complexes, etc., and thereby the protein expression levels.

The 3' UTR of the present disclosure has certain stability and can tolerate a certain degree of sequence (e.g., microRNA binding site) insertion without affecting its ability to promote the translation of the coding region in the mRNA molecule and the stability of the mRNA molecule. Thus, in some embodiments, the 3' UTR of the present disclosure further has one or more (e.g., 1-5) additional sequences inserted.

In some embodiments, the one or more additional sequences are inserted into the sequence of one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof. In some

embodiments, the one or more additional sequences are inserted randomly into the sequence of one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof. In some embodiments, the one or more additional sequences are inserted randomly at different positions in the sequence of one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof.

In some preferred embodiments, the additional sequences are miRNA binding sites. In some preferred embodiments, the additional sequences are selected from a full-length microRNA reverse complementary sequence and a reverse complementary sequence of a seed sequence of microRNA. In some preferred embodiments, the full-length microRNA reverse complementary sequence is 19-25 nt in length, or more preferably, the reverse complementary sequence of the seed sequence of the microRNA is 7-8 nt in length.

In some specific embodiments, the microRNA is hsa-microRNA-122-5p. hsa-microRNA-122-5p, for example, has the nucleotide sequence set forth in SEQ ID NO: 36, and its reverse complementary sequence is set forth in SEQ ID NO: 37. In some specific embodiments, the 3' UTR having one or more additional sequences inserted described in the present disclosure comprises the nucleotide sequence set forth in SEQ ID NO: 38.

In some embodiments of various aspects of the present disclosure, the 3' UTR is a non-natural 3' UTR. In some embodiments, the 3' UTR is a 3' UTR designed by artificial intelligence.

In one aspect, the present disclosure provides an RNA molecule comprising the 5' UTR of the present disclosure and/or the 3' UTR of the present disclosure.

In some embodiments, the RNA molecule comprises the 5' UTR of the present disclosure. The 5' UTR of the present disclosure can be combined with different 3' UTRs to achieve the effect of enhancing the translation of the RNA molecule. The 3' UTR may be a 3' UTR known in the art, e.g., a 3' UTR from a natural RNA molecule, or may be non-natural, e.g., a 3' UTR designed by artificial intelligence.

In some embodiments, the RNA molecule comprises the 3' UTR of the present disclosure. The 3' UTR of the present disclosure can be combined with different 5' UTRs to achieve the effect of enhancing the translation of the RNA molecule. The 5' UTR may be a 5' UTR known in the art, e.g., a 5' UTR from a natural RNA molecule, or may be non-natural, e.g., a 5' UTR designed by artificial intelligence.

In some embodiments, the RNA molecule comprises the 5' UTR of the present disclosure and the 3' UTR of the present disclosure. The 5' UTR of the present disclosure can be combined with the 3' UTR of the present disclosure to achieve the effect of enhancing the translation of the RNA molecule.

In some embodiments, the RNA molecule is messenger RNA (mRNA). An mRNA molecule is generally a linear RNA molecule.

In some embodiments, the RNA molecule is a circular RNA molecule. A circular RNA molecule is a covalently closed RNA molecule.

In some embodiments, the RNA molecule is used to express a polypeptide of interest in a cell. In some embodiments, the RNA molecule comprises a coding sequence of a polypeptide of interest operably linked to the 5' UTR and/or the 3' UTR. “Operably linked” means that the given element is capable of effectively regulating the translation of the polypeptide of interest from the RNA molecule in a cell.

In some embodiments, the RNA molecule further comprises a poly(A) sequence. In some embodiments, the poly(A) sequence is operably linked to a coding sequence of a polypeptide of interest.

The poly(A) sequence typically comprises a plurality of adenine nucleotides. The addition of the poly(A) sequence contributes to the stability and transport of mRNA, prevents its degradation, and plays an important role in post-transcriptional modification. The poly(A) sequence may be a continuous strand of only adenine nucleotides, but may also be a variant comprising non-adenine nucleotides, as long as it is functionally equivalent to a conventional poly(A) sequence, i.e., is capable of providing biological functions similar to those of the natural poly(A) sequence, such as affecting the stability, translation efficiency or ribosome binding of mRNA. Known poly(A) sequences include, for example, the human growth hormone (hGH) poly(A) sequence and the simian virus 40 (SV40) poly(A) sequence. These variants may differ in nucleotide composition, but are functionally equivalent to the conventional poly(A) sequence.

In some embodiments of the present disclosure, the poly(A) sequence comprises about 20 to about 500 consecutive adenine nucleotides (A), e.g., comprises about 25, about 50, about 100, about 150, about 175, about 200, about 300, about 400, or about 500 consecutive adenine nucleotides (A).

In some embodiments, the RNA molecule comprises sequentially, in the 5' to 3' direction, the following:

1) the 5' UTR described in the present disclosure; and

2) a coding sequence of a polypeptide of interest.

In some embodiments, the RNA molecule comprises sequentially, in the 5' to 3' direction, the following:

1) the 5' UTR of the present disclosure;

2) a coding sequence of a polypeptide of interest; and

3) a 3' UTR.

In some embodiments, the RNA molecule comprises sequentially, in the 5' to 3' direction, the following:

1) the 5' UTR of the present disclosure;

2) a coding sequence of a polypeptide of interest; and

3) a 3' UTR; and

4) a poly(A) sequence.

The 3' UTR described in 3) may be a 3' UTR known in the art, e.g., a 3' UTR from a natural RNA molecule, or may be non-natural, e.g., a 3' UTR designed by artificial intelligence. In some embodiments, the 3' UTR is the 3' UTR of the present disclosure.

In some embodiments, the RNA molecule comprises sequentially, in the 5' to 3' direction, the following:

1] a coding sequence of a polypeptide of interest; and

2] the 3' UTR of the present disclosure.

In some embodiments, the RNA molecule comprises sequentially, in the 5' to 3' direction, the following:

1] a 5' UTR;

2] a coding sequence of a polypeptide of interest; and

3] the 3' UTR of the present disclosure.

In some embodiments, the RNA molecule comprises sequentially, in the 5' to 3' direction, the following:

1] a 5' UTR;

2] a coding sequence of a polypeptide of interest;

3] the 3' UTR of the present disclosure; and

4] a poly(A) sequence.

The 5' UTR described in 1] may be a 5' UTR known in the art, e.g., a 5' UTR from a natural RNA molecule, or may be non-natural, e.g., a 5' UTR designed by artificial intelligence. In some embodiments, the 5' UTR is the 5' UTR of the present disclosure.

“Polypeptide of interest” described herein may be any polypeptide to be expressed in a target cell.

The polypeptide of interest may be a polypeptide of eukaryotic, prokaryotic, or viral origin. In certain embodiments, the polypeptide of interest may be any polypeptide for therapeutic, prophylactic, or diagnostic use. For example, the polypeptide of interest may be an antigen, an antibody, a gene editing enzyme such as a CRISPR nuclease, or the like. The polypeptide of interest may also be a chimeric antigen receptor, an immunomodulatory protein, a

transcription factor, or the like. Examples of the polypeptide include, but are not limited to: luciferase, a red/green fluorescent protein, human erythropoietin, and β-galactosidase.

The coding sequence of the polypeptide of interest may be codon optimized for the target cell to be expressed.

"Codon optimization" refers to a method for modifying a nucleic acid sequence to enhance expression in a host cell of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the natural sequence with codons that are more frequently or most frequently used in the genes of the host cell while maintaining its coding amino acid sequence. Different species exhibit a particular bias for certain codons of a particular amino acid. Codon bias (the difference in codon usage between organisms) is often correlated with the efficiency of translation of mRNA, which is in turn believed to be dependent on the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons that are used most frequently in peptide synthesis. Thus, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available. See Nakamura Y. et al., "Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucl. Acids Res., 28:292 (2000).

In some embodiments, the RNA molecule is chemically synthesized. In some embodiments, the RNA molecule is obtained by *in vitro* transcription.

In some embodiments, the RNA molecule is an mRNA molecule and further comprises a 5' cap structure.

As used herein, the "5' cap" with respect to RNA includes 5' cap structures and analogs thereof that are present on natural mRNA. The 5' cap structure on the natural mRNA means that methylated guanosine is linked to the 5'-end nucleotide of RNA via pyrophosphate to form a 5',5'-triphosphate linkage. The 5' cap typically exists in three types (m7G5'ppp5'Np, m7G5'ppp5'NmpNp, and m7G5'ppp5'NmpNmpNp), referred to as Cap0, Cap1, and Cap2, respectively. Cap0 means that the ribose of the terminal nucleotides is unmethylated, Cap1 means that the ribose of one terminal nucleotide is methylated, and Cap2 means that the ribose of both terminal nucleotides is methylated. In some embodiments of the present disclosure, the 5' cap structure is a Cap1 structure.

Methods for capping mRNA molecules are known in the art. The 5' cap structure of the mRNA molecule can be added by enzymatic reactions (e.g., by a commercial kit comprising vaccinia capping enzyme and mRNA cap structure 2'-O-methyltransferase) after the mRNA molecule is obtained by chemical synthesis or *in vitro* transcription. However, it is also

possible to produce mRNA with a cap structure by directly incorporating a nucleotide analog with a cap structure as the first nucleotide into the transcript during *in vitro* transcription.

In some embodiments, the RNA molecule of the present disclosure may further comprise at least one nucleotide modification. The at least one nucleotide modification includes, but is not limited to, a cytidine modification, a uridine modification, or an adenosine modification. In some embodiments, the at least one nucleoside modification includes, but is not limited to, 5-methylcytosine (m5C), N6-methyladenosine (m6A), pseudouridine (ψ), N1-methylpseudouridine (m1ψ), and 5-methoxyuridine (5moU). In some embodiments, the RNA molecule of the present disclosure may comprise at least one modified nucleotide; preferably, the modified nucleotide is selected from pseudouridine, N1-methyl-pseudouridine, 5-methylcytidine, and a combination thereof.

In some embodiments of the present disclosure, an RNA molecule comprising the 5' UTR of the present disclosure results in comparable expression or increased expression of the polypeptide of interest in a cell (e.g., in an HEK293T cell or an HCT116 cell) compared with a corresponding RNA molecule comprising a control 5' UTR (e.g., SEQ ID NO: 14); preferably, the expression of the polypeptide of interest is increased by about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 100%, or more. Preferably, the 5' UTR of the present disclosure is capable of increasing the expression of the polypeptide of interest by at least 20%. More preferably, the 5' UTR of the present disclosure is capable of increasing the expression of the polypeptide of interest by at least 50%.

In some embodiments of the present disclosure, an RNA molecule comprising the 3' UTR of the present disclosure results in comparable expression or increased expression of the polypeptide of interest in a cell (e.g., in an HEK293T cell or an HCT116 cell) compared with a corresponding RNA molecule comprising a control 3' UTR (e.g., SEQ ID NO: 29); preferably, the expression of the polypeptide of interest is increased by about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 100%, or more. Preferably, the 3' UTR of the present disclosure is capable of increasing the expression of the polypeptide of interest by at least 20%. More preferably, the 3' UTR of the present disclosure is capable of increasing the expression of the polypeptide of interest by at least 50%.

In some embodiments of the present disclosure, an RNA molecule comprising the 5' UTR of the present disclosure and the 3' UTR of the present disclosure results in comparable expression or increased expression of the polypeptide of interest in a cell (e.g., in an HEK293T cell or an HCT116 cell) compared with a corresponding RNA molecule

comprising a control 5' UTR (e.g., SEQ ID NO: 14) and a control 3' UTR (e.g., SEQ ID NO: 29); preferably, the expression of the polypeptide of interest is increased by about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 100%, or more.

In one aspect, the present disclosure provides a nucleic acid vector comprising a coding sequence of the RNA molecule of the present disclosure. In some embodiments, the nucleic acid vector is used to generate the RNA molecule of the present disclosure.

As used herein, "vector" refers to a DNA fragment which is extracted from a virus, a plasmid, or a cell of a higher organism and into which a foreign DNA fragment may be or has been inserted for cloning and/or expression purposes. In certain embodiments, the vector may be stably maintained in the organism. The vector may comprise, for example, an origin of replication, a selectable marker or reporter gene, such as antibiotic resistance or GFP, and/or a multiple cloning site (MCS). The term includes linear DNA fragments (e.g., PCR products and linear plasmid fragments), plasmid vectors, viral vectors, cosmids, bacterial artificial chromosomes (BAC), yeast artificial chromosomes (YAC), and the like.

In some embodiments, the nucleic acid vector further comprises an RNA polymerase promoter sequence operably linked to the coding sequence of the RNA molecule. The operably linked promoter allows for *in vivo* and/or *in vitro* transcription of the RNA molecule. The promoter is, for example, a T7 RNA polymerase promoter, a T6 viral RNA polymerase promoter, an SP6 viral RNA polymerase promoter, a T3 viral RNA polymerase promoter, or a T4 viral RNA polymerase promoter.

In some embodiments, the nucleic acid vector is a plasmid vector. In some embodiments, the nucleic acid vector comprises a restriction endonuclease site, e.g., a type IIS restriction endonuclease site, on the 3' flank of the coding sequence of the RNA molecule. Suitable restriction endonucleases include, but are not limited to, BsmBI, BsaI, SapI, and the like. The restriction endonuclease sites may be used to linearize nucleic acid vectors for *in vitro* transcription.

Methods for obtaining RNA molecules by *in vitro* transcription from nucleic acid vectors are known in the art. For example, commercial kits can be used for *in vitro* transcription.

In another aspect, the present disclosure provides a method for increasing expression of a polypeptide of interest in a cell, which comprises introducing the RNA molecule described in the present disclosure and/or the nucleic acid vector described in the present disclosure into the cell.

The introduction of the RNA molecule described in the present disclosure and/or the nucleic acid vector described in the present disclosure into the cell can be performed by methods

known in the present disclosure, such as microinjection, liposome-mediated transfection, or electroporation.

In another aspect, the present disclosure provides a cell comprising the RNA molecule of the present disclosure or a nucleic acid vector of the present disclosure.

In another aspect, the present disclosure provides a pharmaceutical composition comprising the RNA molecule described in the present disclosure, the nucleic acid vector described in the present disclosure, and/or the cell of the present disclosure, and a pharmaceutically acceptable carrier. The specific use of the composition depends on the polypeptide of interest. In some embodiments, the pharmaceutical composition is used for treating and/or preventing a disease in a subject. The specific disease to be treated and/or prevented depends on the polypeptide of interest. When the polypeptide of interest is an antigenic polypeptide, the pharmaceutical composition may be a vaccine.

The pharmaceutically acceptable carrier may include, but is not limited to, a buffer, an excipient, a stabilizer, or a preservative. Examples of the pharmaceutically acceptable carrier are physiologically compatible solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic agents, absorption delaying agents, and the like, such as salts, buffers, saccharides, antioxidants, aqueous or nonaqueous carriers, preservatives, wetting agents, surfactants, or emulsifiers, or combinations thereof. An amount of a pharmaceutically acceptable carrier in a pharmaceutical composition can be determined experimentally based on the activity of the carrier and desired properties of the formulation, such as stability and/or minimum oxidation.

The “cell” described herein may be a mammalian cell, including but not limited to rodent cells such as mouse cells and rat cells; primate cells such as monkey cells and human cells. Preferably, the cell is a human cell.

The “subject” described herein may be a mammal, including but not limited to, rodents such as mice or rats, and primates such as monkeys or humans. Preferably, the subject is a human.

Examples

The present disclosure can be further understood by reference to the specific examples described herein, which are provided solely to illustrate the present disclosure and are not intended to limit the scope of the present disclosure. Obviously, those skilled in the art know that various modifications and variations can be made to the present disclosure without departing from the spirit of the present disclosure. Therefore, such modifications and variations also fall within the scope of the present disclosure.

Example 1. Methods for Data Training and Generation of Non-Natural UTR Sequences

Using published human cell Ribo-seq data and RNA-seq data in public databases, a deep learning prediction model was trained, which is capable of predicting protein translation efficiency based on mRNA 5' UTR sequences. Subsequently, a generative learning model was constructed by using an artificial intelligence analysis strategy, including technologies such as the generative adversarial network (GAN) and long short-term memory (LSTM) recurrent neural networks, and by using 5' UTR sequences in the human genome for training, to produce new and non-natural 5' UTR sequences. The generated model was combined with the prediction model to evaluate the generated non-natural 5' UTR sequences, predict the translation efficiency of the RNAs consisting of these 5' UTRs and protein-coding ORFs, and screen a batch of 5' UTR sequences that are likely to achieve relatively high protein translation efficiency. Then, sequences whose protein translation efficiency was not high in the actual expression were excluded by wet testing, and the finally obtained candidate 5' UTR sequences are shown in Table 1.

Candidate 3' UTRs were screened using a similar method. Their sequences are shown in Table 2.

Table 1. Candidate 5' UTRs

5'UTR	Sequence	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

Table 2. Candidate 3' UTRs

3’UTR	Sequence	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	GGCACCTGCCTCGGACCACCCACCCTCTCGCCCACTCTGG ACCTGGTTTTTCTGCCCATCATCTCTTCCCGCGCCGAAGGC CAGCCTCTGAAATGCTCACCTTTACTTCT	23
Seq23	TCCTTCCTGCTGCTGCCCACAACCGGAAGCCTTTTCTCTCC TTTCCTTCTGTAGGATATATTTGCTGGACGAGTTTGCTCCTG AAAGTAGAACACAAGATTGCCAAAGTC	24
Seq24	AGCTTCCACCAGTCTCCTGCTCTGCGGCCTGTCTTCCACAT CCTCGACTCCCTCGGCTGTGTCCCTGCTGTGGTGCCACTCG CCTTCCCTCGAAGTTTCTGTTTTTCTCA	25
Seq25	GCTCCTCCGCCCTTCAGCACTCTGCAGCCGCCCCTCCCTCC TCTCCCCCTGCCCTTGTGGTGCCACAGCCCCAGTTGGCCTC CCTGACCTCTCCCCCTGTCCTCGTCGTC	26
Seq26	AAACTCTCCTTCCACCATGCCTCCGAAAGATGGGATACTAT CGCTGCGCAATTCTGAGGAACCTCTGGAATACACTCTGCT GTGTGTCAAACAATCTCTTCGAATACTCC	27
Seq27	GCAGACCTCTACGCCTAGCTGCCTCTGTTCCCCGCTTCTCT CCCTTTCTCCGCACCCCTTTCTGAGCTGGCTCCTACCCATC CTGCCTGACCCGTTTTGACATTCTTCTA	28
seq48	ACTGCACTTCGTCATGGGCTACTAGTTGTGCACTTGGAAGA GGAGGGATCCCTTCAGAAAGCTTCTAACCTGTGAATACAA GAAGAATTTCATGGCGAATAAATCTTCTG	61
seq49	AGAATGGAGTGAGAAAAGATGAAAAGATCCATTCTATTGC TATTAAATCCAACGTTACTCTATGGTGTTTAAAGAATTTCAT TTTCCAAATATCTGGAATGTTTTGCAAG	62
seq50	TGACAAGGAAGTGGGGGTGGGCAATAAAGTTCACTGAAA CTAACAAAAACAGACCTTTGTATAGGTGATCTCAAATATTT CTGGTCTTTATCTGAATTATAGTGTAGTAT	63
seq51	AAAGATCTCTATGTTGATGACAAAAAGGAACAGCTACCCC AAACAAGGAGATCAGGTTGTTAAATGTTTGTAGATTGTAAT TTGTTAAAAATAATTATGTGCTTTTTTAG	64
seq52	ACTGTATGTCAGCATGTGGAGCTGTAGTGACATTGACTAAC TGTCTGTGCACATTGTCAAAGTTCACAAAAAATGGCTTTTC CAGGATTGCTTTACCAGGTCTTAATCGT	65
seq53	TGACTCAGAGAACCTTGACATCCATGTTTCTTGTGACCATG ACCTCAGTGTACAAGTTATCGCAGGTGGTGTTAGCTTCAGA TCATTTACTCAAGGAAGAGACATATTAT	66

seq54	TAGATAAAAGGCTGTGATTGGGTGAAAGGGGAAAATGAA GGAATAAAGGAAAGATCTAAAGCATTTCTATTACTTTATTTT TTCCTAAATGTTATATCAGTTTATATTCA	67
seq55	AATACTTGCCAAGAGTGAGCTAGCAGCGCAACTCGCCACC AGTGGCTGCAAATCGAAAGAAAACAACCAGGAGTTCTACT GTACACTGGTCATGGCATGTTCATTTTGGC	68
seq56	ACTGATCCACTCTTCTGTCACATCAATCAAGAAAACAGCCA GGCATCTTTACCAATATTTGTATTATATTAAATATAATATAT GATTTAATTTAATTGTGAGTTGTGTAT	69
seq57	ACTCTTCCGAGCTGTGCTCATGCGATCTTGCTTGTGGTGTC TGTGGAGGATATCTGCCTGCTGGCAGTAAACTAGTTGCTA GCAATTTTGTGTTCGAGACTGCCTTGAAA	70
seq58	AGGGCTTGGCATCTTCAGTTTCTTCTTCTTTCTCTTGCAATC CCTTTATGATCTATTGTTAAGCAAATTCCTCAACATGAAAA ATCAAAAGATCGTCTGTGAATCTTCTG	71
seq59	ATGTTCCACTTCAATCCTTACAGTTCTCAGCTCTGAATCTTT CTTATCTGGTTGTGTCGTTTGGGAAGAGAATCTTGTATTTG AATATGCTTCCCCTGCTTTTGTCTCTT	72
seq60	AGTTCTTTGTCTTCTGACTTCTTTTCCTAATCCTCTCCCTTGC TTAAATGTTTGCTTAAACACTTTAACACAGTAAATGGGTAG TCATTTTTTAGGAAATAATTTATGAA	73
seq61	AGTTCCTCGCTCATCTCTCTCATTTCTTCTCTTGCAATTATG AGCTTAGATGGCCATTTTGGGAATACTACTGTTAAATTACT TTTTCAAGCTACCCACCTTAAATTCAA	74
seq62	AGTACTTAACCTCTTTCACTTTGCTTCTGCCTTTGCTGCATC TGCTGCTTTAGCCACAGCTTTCTTTTGGTCTTCACTGAGTTA AATGAAACCATGGTGAGAAGAAAGTA	75
seq63	ACACTGTACTCTTTGCTGCTAGAAATTTCTCAGGGCTTCCT CCATCTTCTCCTCTTCTGTATATTCACGTGGCTTTGAGTTGT AAAACTGTTTTTACACAGTTGTTATTC	76
seq64	AGTGTGCTTCTTCACTATTCAAATGCCAGCCACATCTTCTG ACTGCAGTCCTTTGTCAAATTGTTACCAAAGTTTTTTAACAT TCTGAATTACCAGTAATAAAGTCTAGA	77
seq65	AGTCCCCTTTACTGATGTCTAGGATTTCACCAACTGACCTC AGGCTTCATCCTTTTATTTCTTATCAGAATTGTAATAGATGA TTTAAAATTTTCATATTTGAAAAACCT	78

seq66	AAGGATTCTCTCTCTTTCATGATTCATTTTTACTTCCTGATAA GGCTTCTGCCACAACTTAATTTATGTCGTCATGCACTTCCA CAGTAGAACTCAGACTCTATTGTTTC	79
seq67	TCATTGAATTCACTTCCGTTTACTATAACATGTTCCTGGATG TCTCAGTGTTGCTCTTCTCTATTATGGCATTCAGTTCTGTTCT TACATCTTCTCAACTACTTCTGGAA	80
Moderna_ctrl_3'	TGATAATAGGCTGGAGCCTCGGTGGCCATGCTTCTTGCCC CTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCACCCG TACCCCCGTGGTCTTTGAATAAAGTCTGAGTGGGCGGC	29

Example 2. Construction of Vectors Comprising Candidate UTR Sequence Elements

To test candidate 5' UTRs, a nucleic acid fragment comprising a T7 promoter, a candidate 5' UTR or a control 5' UTR, a sequence encoding firefly luciferase or a red fluorescent protein (Table 3), a 3' UTR (SEQ ID NO: 29), a poly(A) sequence comprising 100 A nucleotide residues, and a type IIS restriction endonuclease cleavage site was synthesized *in vitro* and cloned into an *in vitro* transcription vector (pIVTRup, Addgene plasmid #101362). The 5' UTR sequence element (SEQ ID NO: 14) as a control was from the Moderna mRNA1273 expression vector (complete sequence, GenBank: OR134578.1). The 3' UTR sequence element (SEQ ID NO: 29) was from the Moderna mRNA1273 expression vector (complete sequence, GenBank: OR134578.1).

To test candidate 3' UTRs, a nucleic acid fragment comprising a T7 promoter, a 5' UTR (SEQ ID NO: 14), a sequence encoding firefly luciferase or a red fluorescent protein (Table 3), a candidate 3' UTR or a control 3' UTR, a poly(A) sequence comprising 100 A nucleotide residues, and a type IIS restriction endonuclease cleavage site was synthesized *in vitro* and cloned into an *in vitro* transcription vector. The 5' UTR sequence element (SEQ ID NO: 14) was from the Moderna mRNA1273 expression vector (complete sequence, GenBank: OR134578.1). The 3' UTR sequence element (SEQ ID NO: 29) as a control was from the Moderna mRNA1273 expression vector (complete sequence, GenBank: OR134578.1).

Table 3. Firefly luciferase and red fluorescent protein sequences

Name	**Sequence**	**SEQ ID NO:**
Firefly luciferase amino acid sequence	MEDAKNIKKGPAPFYPLEDGTAGEQLHKAMKRYALVPGTI AFTDAHIEVDITYAEYFEMSVRLAEAMKRYGLNTNHRIVVC SENSLQFFMPVLGALFIGVAVAPANDIYNERELLNSMGISQ PTVVFVSKKGLQKILNVQKKLPIIQKIIIMDSKTDYQGFQSM YTFVTSHLPPGFNEYDFVPESFDRDKTIALIMNSSGSTGLPK	30

	GVALPHRTACVRFSHARDPIFGNQIIPDTAILSVVPFHHGFG MFTTLGYLICGFRVVLMYRFEEELFLRSLQDYKIQSALLVPTL FSFFAKSTLIDKYDLSNLHEIASGGAPLSKEVGEAVAKRFHLP GIRQGYGLTETTSAILITPEGDDKPGAVGKVVPFFEAKVVDL DTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALIDKD GWLHSGDIAYWDEDEHFFIVDRLKSLIKYKGYQVAPAELESI LLQHPNIFDAGVAGLPDDDAGELPAAVVVLEHGKTMTEKE IVDYVASQVTTAKKLRGGVVFVDEVPKGLTGKLDARKIREILI KAKKGGKIAV*	
Firefly luciferase nucleotide sequence	ATGGAGGACGCCAAGAACATCAAGAAGGGCCCCGCCC CCTTCTACCCCCTGGAGGACGGCACCGCCGGCGAGCA GCTGCACAAGGCCATGAAGCGGTACGCCCTGGTGCCC GGCACCATCGCCTTCACCGACGCCCACATCGAGGTGGA CATCACCTACGCCGAGTACTTCGAGATGAGCGTGCGGC TGGCCGAGGCCATGAAGCGGTACGGCCTGAACACCAA CCACCGGATCGTGGTGTGCAGCGAGAACAGCCTGCAGT TCTTCATGCCCGTGCTGGGCGCCCTGTTCATCGGCGTGG CCGTGGCCCCCGCCAACGACATCTACAACGAGCGGGA GCTGCTGAACAGCATGGGCATCAGCCAGCCCACCGTGG TGTTCGTGAGCAAGAAGGGCCTGCAGAAGATCCTGAAC GTGCAGAAGAAGCTGCCCATCATCCAGAAGATCATCAT CATGGACAGCAAGACCGACTACCAGGGCTTCCAGAGC ATGTACACCTTCGTGACCAGCCACCTGCCCCCCGGCTTC AACGAGTACGACTTCGTGCCCGAGAGCTTCGACCGGGA CAAGACCATCGCCCTGATCATGAACAGCAGCGGCAGCA CCGGCCTGCCCAAGGGCGTGGCCCTGCCCCACCGGAC CGCCTGCGTGCGGTTCAGCCACGCCCGGGACCCCATCT TCGGCAACCAGATCATCCCCGACACCGCCATCCTGAGC GTGGTGCCCTTCCACCACGGCTTCGGCATGTTCACCACC CTGGGCTACCTGATCTGCGGCTTCCGGGTGGTGCTGAT GTACCGGTTCGAGGAGGAGCTGTTCCTGCGGAGCCTGC AGGACTACAAGATCCAGAGCGCCCTGCTGGTGCCCACC CTGTTCAGCTTCTTCGCCAAGAGCACCCTGATCGACAAG TACGACCTGAGCAACCTGCACGAGATCGCCAGCGGCG GCGCCCCCCTGAGCAAGGAGGTGGGCGAGGCCGTGGC CAAGCGGTTCCACCTGCCCGGCATCCGGCAGGGCTACG GCCTGACCGAGACCACCAGCGCCATCCTGATCACCCCC GAGGGCGACGACAAGCCCGGCGCCGTGGGCAAGGTG GTGCCCTTCTTCGAGGCCAAGGTGGTGGACCTGGACAC CGGCAAGACCCTGGGCGTGAACCAGCGGGGCGAGCTG TGCGTGCGGGGCCCCATGATCATGAGCGGCTACGTGAA CAACCCCGAGGCCACCAACGCCCTGATCGACAAGGAC	31

	GGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACG AGGACGAGCACTTCTTCATCGTGGACCGGCTGAAGAGC CTGATCAAGTACAAGGGCTACCAGGTGGCCCCCGCCGA GCTGGAGAGCATCCTGCTGCAGCACCCCAACATCTTCG ACGCCGGCGTGGCCGGCCTGCCCGACGACGACGCCGG CGAGCTGCCCGCCGCCGTGGTGGTGCTGGAGCACGGC AAGACCATGACCGAGAAGGAGATCGTGGACTACGTGG CCAGCCAGGTGACCACCGCCAAGAAGCTGCGGGGCGG CGTGGTGTTCGTGGACGAGGTGCCCAAGGGCCTGACCG GCAAGCTGGACGCCCGGAAGATCCGGGAGATCCTGAT CAAGGCCAAGAAGGGCGGCAAGATCGCCGTGTGA	
Red fluorescent protein amino acid sequence	MASSEDVIKEFMRFKVRMEGSVNGHEFEIEGEGEGRPYEGT QTAKLKVTKGGPLPFAWDILSPQFQYGSKAYVKHPADIPDY LKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQDGEFIYKV KLRGTNFPSDGPVMQKKTMGWEASTERMYPEDGALKGEI KMRLKLKDGGHYDAEVKTTYMAKKPVQLPGAYKTDIKLDIT SHNEDYTIVEQYERAEGRHSTGA*	32
Red fluorescent protein nucleotide sequence	ATGGCCTCCTCCGAGGACGTCATCAAGGAGTTCATGCG CTTCAAGGTGCGCATGGAGGGCTCCGTGAACGGCCACG AGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTA CGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAG GGCGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCC TCAGTTCCAGTACGGCTCCAAGGCCTACGTGAAGCACC CCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCG AGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGAC GGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCA GGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGC ACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAA GACCATGGGCTGGGAGGCCTCCACCGAGCGGATGTAC CCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGATGA GGCTGAAGCTGAAGGACGGCGGCCACTACGACGCCGA GGTCAAGACCACCTACATGGCCAAGAAGCCCGTGCAG CTGCCCGGCGCCTACAAGACCGACATCAAGCTGGACAT CACCTCCCACAACGAGGACTACACCATCGTGGAACAGT ACGAGCGCGCCGAGGGCCGCCACTCCACCGGCGCCTA A	33

Example 3. Production of mRNA Molecules Comprising Candidate UTR Sequence Elements

The vector obtained in Example 2 was linearized and subjected to *in vitro* transcription using

a T7-RNA polymerase to produce an mRNA molecule, with simultaneous addition of a 5'-cap structure. The 5'-cap structure was added by co-transcriptional capping, where a cap analog was incorporated as the first nucleotide into the transcript during *in vitro* transcription to directly produce an mRNA molecule with the Cap1 structure. FIG. 1 shows the sequences of an exemplary mRNA comprising candidate UTR elements of the present disclosure.

The mRNA molecules thus obtained were purified and resuspended in water. The mRNA molecules were tested for quality using a 5200 fragment analyzer (Agilent) to determine whether the length and completeness of the mRNA molecules (values were from the percentage of area under the curve for fragments of expected length) met the requirements (see FIG. 2), and the mRNA molecules that met the requirements could be used for subsequent testing of different candidate UTR sequence elements.

Example 4. Identification of Effects of Candidate 5' UTR Sequence Elements on Translation Efficiency by Luciferase Reporter System

In this example, whether the candidate 5' UTR sequence elements had a positive regulatory effect on translation was tested by the following method:

1) Mammalian 293T cells were transfected with the RNA molecule containing the candidate 5' UTR sequence element (the encoded protein sequence was firefly luciferase).

2) The chemiluminescence absorbance value of luciferase was measured at a particular time point (24 h) after transfection, which represented the protein expression level.

3) The absorbance value of the background group was subtracted from the absorbance reading of the experimental group containing the candidate 5' UTR sequence element to obtain the relative light unit of the experimental group.

The measured relative light unit of each candidate 5' UTR sequence element could indicate the translational regulatory ability of the candidate 5' UTR sequence element.

Measurement method of luciferase chemiluminescence absorbance value for determining translation efficiency:

293T human embryonic kidney cells were seeded in a 48-well plate at a density of 5×10^4 cells/well. On the next day, the cells were washed in Opti-MEM and transfected with 200 ng/well of the Lipofectamine 2000-complexed mRNA encoding firefly luciferase containing the candidate UTR sequence element in Opti-MEM. Cells without any addition of RNA molecules were used as the background group. 6 h after transfection, the mixed medium was aspirated and replaced with a complete medium. 24 h after transfection, the medium was aspirated, and 100 μL of a lysis buffer (Promega) was added. After lysis for 5 min, the absorbance value was measured and read.

The luciferase activity was measured in relative light units (RLU) in a multi-well plate reader (Agilent). The luciferase activity of individual samples was measured sequentially in the luciferase assay. 20 μL of the lysis buffer was taken and added to 50 μL of a buffer containing a firefly luciferase substrate. The plate was shaken, and the mixture was mixed well. The absorbance value was measured.

The results of the 5' UTR test are shown in Table 4 and FIG. 3 below.

Table 4. RLU ratios of candidate 5' UTRs

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

Example 5. Identification of Effects of Candidate 5' UTR Sequence Elements on Translation Efficiency by Red Fluorescent Protein Reporter System

In this example, whether the candidate 5' UTR sequence elements had a positive regulatory effect on translation was tested by the fluorescent protein experiment as follows:

1) Mammalian 293T cells were transfected with the RNA molecule containing one candidate 5' UTR sequence element (the encoded protein sequence was a red fluorescent protein).

2) The light intensity of the fluorescent protein was measured by flow cytometry at a particular time point (24 h) after transfection, which represented the protein expression level.

3) The light intensity of the red fluorescent protein was divided by that of the internal control (the light intensity of the green fluorescent protein) to obtain the relative light intensity of the experimental group.

The obtained relative light intensity of each candidate 5' UTR sequence element could indicate the translational regulatory ability of the candidate 5' UTR sequence element.

Measurement method of fluorescence intensity by flow cytometry for determining translation efficiency:

293T human embryonic kidney cells were seeded in a 24-well plate at a density of 2×10^5 cells/well. On the next day, the cells were washed in Opti-MEM, transfected with 200 ng/well of the Lipofectamine 2000-complexed mRNA encoding the red fluorescent protein containing the candidate 5' UTR sequence element in Opti-MEM, and co-transfected with an equal amount of mRNA of the green fluorescent protein as the internal control. 6 h after

transfection, the mixed medium was aspirated and replaced with a complete medium. 24 h after transfection, the medium was aspirated, and the cells were digested with trypsin for 3 min. The digestion was then neutralized, and cell samples were collected. With PBS as a suspension, cell analysis was performed in a flow cytometer, and the light intensity of the red fluorescence and green fluorescence was read.

The amino acid sequence of the green fluorescent protein as the internal control is shown below (SEQ ID NO: 34):

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP
VPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKV
NFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHM
VLLEFVTAAGITLGMDELYK*

The sequence of the green fluorescent protein as the internal control is shown below (SEQ ID NO: 35):

*GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACC*ATGGTGAGC
AAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGG
CGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGT
GCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCC
GCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAA
GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGAC
CCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG
AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGT
ACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG
CATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAG
CTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCT
GCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACG
AGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACT
CTCGGCATGGACGAGCTGTACAAGTAA

The results of the 5' UTR test are shown in Table 5 and FIG. 4 below.

Table 5. Relative light intensity of candidate 5' UTRs

5'UTR	**Mean RFP/GFP fluorescence intensity (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: The experiment was not performed.

The candidate 5' UTRs of the present disclosure can achieve comparable or superior translation efficiency compared with the control 5' UTR.

Example 6. Identification of Effects of Candidate 3' UTR Sequence Elements on Translation Efficiency by Luciferase Reporter System

In this example, whether the candidate 3' UTR sequence elements had a positive regulatory effect on translation was tested by the following method:

1) Representative human cell lines, namely human embryonic kidney cell line HKE293T and human colon cancer cell line HCT116, were each transfected with the RNA molecule containing the candidate 3' UTR sequence element (the encoded protein sequence was firefly luciferase).

2) The chemiluminescence absorbance value of luciferase was measured at a particular time point (24 h) after transfection, which represented the protein expression level.

3) The absorbance value of the background group was subtracted from the absorbance reading of the experimental group containing the candidate 3' UTR sequence element to obtain the relative light unit of the experimental group.

The measured relative light unit of each candidate 3' UTR sequence element could indicate the translational regulatory ability of the candidate 3' UTR sequence element.

Measurement method of luciferase chemiluminescence absorbance value for determining translation efficiency:

HEK293T human embryonic kidney cells or HCT116 human colon cancer cells were seeded in a 48-well plate at a density of 5×10^4 cells/well. On the next day, the cells were washed in Opti-MEM and transfected with 200 ng/well of the Lipofectamine 2000-complexed

mRNA encoding firefly luciferase containing the candidate 3' UTR sequence element in Opti-MEM. Cells without any addition of RNA molecules were used as the background group. 6 h after transfection, the mixed medium was aspirated and replaced with a complete medium. 24 h after transfection, the medium was aspirated, and 100 μL of a lysis buffer (Promega) was added. After lysis for 5 min, the absorbance value was measured and read. The luciferase activity was measured in relative light units (RLU) in a multi-well plate reader (Agilent). The luciferase activity of individual samples was measured sequentially in the luciferase assay. 20 μL of the lysis buffer was taken and added to 50 μL of a buffer containing a firefly luciferase substrate. The plate was shaken, and the mixture was mixed well. The absorbance value was measured.

The results of the 3' UTR test are shown in Tables 6A and 6B and FIGs. 5A and 5B below.

Table 6A. RLU ratios of candidate 3' UTRs (HEK293T cell line)

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

Table 6B. RLU ratios of candidate 3' UTRs (HCT116 cell line)

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

Example 7. Identification of Effects of Candidate 3' UTR Sequence Elements on Translation Efficiency by Red Fluorescent Protein Reporter System

In this example, whether the candidate 3' UTR sequence elements had a positive regulatory effect on translation was tested by the fluorescent protein experiment as follows:

1) Mammalian 293T cells were transfected with the RNA molecule containing one candidate 3' UTR sequence element (the encoded protein sequence was a red fluorescent protein).

2) The light intensity of the fluorescent protein was measured by flow cytometry at a particular time point (24 h) after transfection, which represented the protein expression level.

3) The light intensity of the red fluorescent protein was divided by that of the internal control (the light intensity of the green fluorescent protein) to obtain the relative light intensity of the experimental group.

The obtained relative light intensity of each candidate 3' UTR sequence element could indicate the translational regulatory ability of the candidate 3' UTR sequence element.

Measurement method of fluorescence intensity by flow cytometry for determining translation efficiency:

293T human embryonic kidney cells were seeded in a 24-well plate at a density of 2×10^5 cells/well. On the next day, the cells were washed in Opti-MEM, transfected with 200 ng/well of the Lipofectamine 2000-complexed mRNA encoding the red fluorescent protein containing the candidate UTR sequence element in Opti-MEM, and co-transfected with an equal amount of mRNA (SEQ ID NO: 35) of the green fluorescent protein as the internal control. 6 h after transfection, the mixed medium was aspirated and replaced with a complete medium. 24 h after transfection, the medium was aspirated, and the cells were digested with trypsin for 3 min. The digestion was then neutralized, and cell samples were collected. With PBS as a suspension, cell analysis was performed in a flow cytometer, and the light intensity of the red fluorescence and green fluorescence was read.

The results of the 3' UTR test are shown in Table 7 and FIG. 6 below.

Table 7. Relative light intensity of candidate 3' UTRs

3'UTR	Mean RFP/GFP fluorescence intensity (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

The candidate 3' UTRs of the present disclosure can achieve comparable or superior translation efficiency compared with the control 3' UTR.

Example 8. Insertion of microRNA Binding Sites into Candidate 3' UTRs Without Affecting Functions Thereof

In nature, the 3' UTR regions of many mRNA molecules contain microRNA binding sites, through which microRNAs can achieve precise regulation of mRNAs, including translation inhibition and promotion of mRNA degradation. Given that microRNAs exhibit significant expression differences in different tissues and disease stages, they are often considered highly specific biomarkers. Therefore, the insertion of microRNA binding sites into the 3' UTR region is one of the common strategies for regulating mRNA functions. Theoretically, the interaction of miRNA with its target mRNA mainly depends on the complete complementarity of the seed sequence (usually 2 to 8 nucleotides) of the miRNA with the corresponding site in the 3' UTR of the mRNA. This complementary pairing is sufficient to guide the RNA-induced silencing complex (RISC) onto the target mRNA, thereby inhibiting

its translation or promoting its degradation. Numerous studies have confirmed that in addition to seed sequence matching, additional complementary pairing between miRNA sequences and mRNAs can also enhance miRNA-mediated gene regulation effects. This additional complementary pairing generally occurs in regions outside the miRNA seed sequence, and the matching in these regions can improve the affinity of the miRNA for its target, thereby increasing the efficiency of mRNA inhibition. For example, one strategy employed by Moderna is to insert the full-length reverse complementary sequence of a miRNA into the 3' UTR region. This design allows for broader complementary pairing of the miRNA with its target mRNA and is not limited to seed sequences, thereby enhancing the regulatory potency of the miRNA on the mRNA.

In this Example, a fluorescent protein that degrades rapidly within cells was utilized to test whether the functions of the candidate 3' UTR sequence elements changed after engineering, confirming their positive regulatory function on translation and their regulatory effect on the stability of mRNA molecules. The method was as follows:

1) The full-length reverse complementary sequence of hsa-microRNA-122-5p was inserted into 3 random sites in the candidate 3' UTR (SEQ ID NO: 25) to obtain a new 3' UTR (SEQ ID NO: 38).

2) Mammalian cell lines such as 293T, Hela, and raw264.7 were transfected with RNA molecules containing the new 3' UTR and original 3' UTR sequence elements (both elements encode a green fluorescent protein, d1EGFP, that degrades within 1 h and has a life of about 1 h; 1 h after translation of the mRNA into protein has stopped, there is no green fluorescence, so the protein fluorescence can directly indicate the stability of the mRNA).

3) Within 24 h after transfection, the cells were photographed at an interval of one hour, and the light intensity of the fluorescent protein was measured to represent the protein expression level.

4) The light intensity of the green fluorescent protein was divided by that of the internal control (the light intensity of the red fluorescent protein) to obtain the relative light intensity of the experimental group or the control group. The relative light intensity could indicate the translational regulatory ability of the 3' UTR sequence, and the change in the fluorescence intensity within 24 h could indicate the effect of the 3' UTR sequence on the stability of the mRNA molecule.

The full-length sequence of hsa-microRNA-122-5p is shown below (SEQ ID NO: 36):

TGGAGTGTGACAATGGTGTTTG

The full-length reverse complementary sequence of hsa-microRNA-122-5p is shown below (SEQ ID NO: 37):

CAAACACCATTGTCACACTCCA

The sequence of the new 3' UTR obtained by inserting the full-length reverse complementary sequence of hsa-microRNA-122-5p into three random sites in 3' UTR-seq24 is shown below (SEQ ID NO: 38):

CAAACACCATTGTCACACTCCAAGCTTCCACCAGTCTCCTGCTCTGCGGCCCAA
ACACCATTGTCACACTCCATGTCTTCCACATCCTCGACTCCCTCGGCTGTGTCC
CTGCTGTGGTGCCACTCGCCCAAACACCATTGTCACACTCCATTCCCTCGAAGT
TTCTGTTTTTCTCA

The amino acid sequence of the green fluorescent protein (d1EGFP) degrading within 1 h is shown below (SEQ ID NO: 39):

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLP
VPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKV
NFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHM
VLLEFVTAAGITLGMDELYKKLSHGFPPAVAAQDDGTLPMSCAQESGMDRHPAA
CASARINV*

The sequence of the mRNA comprising the sequences of the green fluorescent protein degrading within 1 h and the sequence of the new 3' UTR obtained by inserting the full-length reverse complementary sequence of hsa-microRNA-122-5p into three random sites in 3' UTR-seq24 is shown below (SEQ ID NO: 40):

GGGAAATAAGAGAGAAAAGAAGAGTAAGAAGAAATATAAGAGCCACCATGG
TGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG
GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG
ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC
AGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCC
CGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACA
AGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA
GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTG
GAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGA
ACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGT
GCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGC
TGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCC
AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGAT
CACTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGG

CGGTGGCGGCGCAGGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGC
GGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCTAGGATCAATGTGTAGCA
AACACCATTGTCACACTCCAAGCTTCCACCAGTCTCCTGCTCTGCGGCCCAAAC
ACCATTGTCACACTCCATGTCTTCCACATCCTCGACTCCCTCGGCTGTGTCCCT
GCTGTGGTGCCACTCGCCCAAACACCATTGTCACACTCCATTCCCTCGAAGTTT
CTGTTTTTCTCA

The experimental results show that the insertion of microRNA binding sites into candidate 3' UTRs does not affect their functions (FIGs. 7 and 8).

CLAIMS

1. A 5' UTR (5' untranslated region), comprising a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to one of SEQ ID NOs: 1-13 and 41-60 or a complementary sequence thereof, preferably, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 1-13 and 41-60 or a complementary sequence thereof; more preferably, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 1, 3-5, 7-11, 13, 41-46, 48, 50-53, 55-57, and 59-60 or a complementary sequence thereof; more preferably, the 5' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 4-5, 7-10, 41, 43-45, 48, 50-51, 53, 55, 57, and 59 or a complementary sequence thereof.

2. A 3' UTR (3' untranslated region), wherein

i) the 3' UTR comprises a nucleotide sequence having at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof; preferably, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-28 and 61-80 or a complementary sequence thereof; more preferably, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-20, 22-23, 25-26, 28, 61-71, and 73-80 or a complementary sequence thereof; more preferably, the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-16, 18-19, 22, 25, 28, 61, 63-66, 68, 70-71, and 73-80 or a complementary sequence thereof; or

ii) the 3' UTR comprises the nucleotide sequence set forth in one of SEQ ID NOs: 15-28 and 61-80 having one or more additional sequences inserted or a complementary sequence thereof; preferably, the additional sequences are miRNA binding sites; more preferably, the additional sequences are selected from a full-length microRNA reverse complementary sequence and a reverse complementary sequence of a seed sequence thereof; more preferably, the full-length microRNA reverse complementary sequence is 19-25 nt in length, or the reverse complementary sequence of the seed sequence is 7-8 nt in length.

3. An RNA molecule for expressing a polypeptide of interest, comprising the 5' UTR according to claim 1 and/or the 3' UTR according to claim 2.

4. The RNA molecule according to claim 3, wherein the RNA molecule is a messenger RNA (mRNA) molecule or a circular RNA molecule.

5. The RNA molecule according to claim 3 or 4, comprising sequentially, in the 5' to 3' direction, the following:

1) the 5' UTR according to claim 1; and

2) a coding sequence of a polypeptide of interest.

6. The RNA molecule according to claim 3 or 4, comprising sequentially, in the 5' to 3' direction, the following:

1) the 5' UTR according to claim 1;

2) a coding sequence of a polypeptide of interest; and

3) a 3' UTR.

7. The RNA molecule according to claim 6, wherein the 3' UTR is the 3' UTR according to claim 2.

8. The RNA molecule according to claim 3 or 4, comprising sequentially, in the 5' to 3' direction, the following:

1] a coding sequence of a polypeptide of interest; and

2] the 3' UTR according to claim 2.

9. The RNA molecule according to claim 3 or 4, comprising sequentially, in the 5' to 3' direction, the following:

1] a 5' UTR;

2] a coding sequence of a polypeptide of interest; and

3] the 3' UTR according to claim 2.

10. The RNA molecule according to claim 9, wherein the 5' UTR is the 5' UTR according to claim 1.

11. The RNA molecule according to any one of claims 3-10, further comprising a poly(A) sequence.

12. The RNA molecule according to claim 11, wherein the poly(A) sequence comprises about 20 to about 500 adenine nucleotides (A), e.g., about 25, about 50, about 100, about 150, about 175, about 200, about 300, about 400, or about 500 adenine nucleotides (A).

13. The RNA molecule according to any one of claims 3-12, wherein the RNA molecule is an mRNA molecule and further comprises a 5' cap structure, e.g., a Cap1 structure.

14. The RNA molecule according to any one of claims 3-13, wherein the RNA molecule is chemically synthesized or obtained by *in vitro* transcription.

15. The RNA molecule according to any one of claims 3-14, further comprising at least one nucleotide modification, wherein preferably, the nucleotide modification is selected from one of 5-methylcytosine (m5C), N6-methyladenosine (m6A), pseudouridine (ψ), N1-methylpseudouridine (m1ψ), and 5-methoxyuridine (5moU), and a combination thereof.

16. The RNA molecule according to any one of claims 3-15, wherein the RNA results in

comparable expression or increased expression of the polypeptide of interest in a cell compared with a corresponding RNA molecule comprising a control UTR; preferably, the expression of the polypeptide of interest is increased by about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 100%, or more.

17. A nucleic acid vector, comprising a coding sequence of the RNA molecule according to any one of claims 3-16.

18. The nucleic acid vector according to claim 17, further comprising an RNA polymerase promoter sequence operably linked to the coding sequence of the RNA molecule.

19. The nucleic acid vector according to claim 18, wherein the promoter is selected from a T7 RNA polymerase promoter, a T6 viral RNA polymerase promoter, an SP6 viral RNA polymerase promoter, a T3 viral RNA polymerase promoter, and a T4 viral RNA polymerase promoter.

20. A cell comprising the RNA molecule according to any one of claims 3-16 or the nucleic acid vector according to any one of claims 17-19.

21. A method for increasing expression of a polypeptide of interest in a cell, comprising introducing the RNA molecule according to any one of claims 3-16 and/or the nucleic acid vector according to any one of claims 17-19 into the cell.

22. A pharmaceutical composition, comprising the RNA molecule according to any one of claims 3-16, the nucleic acid vector according to any one of claims 17-19, and/or the host cell according to claim 20, and a pharmaceutically acceptable carrier.

23. Use of the 5' UTR according to claim 1 and/or the 3' UTR according to claim 2 for increasing translation efficiency of a polypeptide of interest in an RNA molecule comprising a coding sequence of the polypeptide of interest.