(19) Octrooicentrum Nederland

(11) 2033707

(12) B1 OCTROOI

(21) Aanvraagnummer: **2033707**

(22) Aanvraag ingediend: **9 december 2022**

(62)

(51) Int. Cl.:
C12N 15/63 (2023.01) **C12N 15/67** (2023.01) **C12N 15/113** (2023.01) **C12N 15/86** (2023.01)

(30) Voorrang:
-

(41) Aanvraag ingeschreven:
14 juni 2024

(43) Aanvraag gepubliceerd:
-

(47) Octrooi verleend:
14 juni 2024

(45) Octrooischrift uitgegeven:
2 augustus 2024

(73) Octrooihouder(s):
Stichting Sanquin Bloedvoorziening te AMSTERDAM

(72) Uitvinder(s):
Monika Christine Wolkers te AMSTERDAM
Benoît Pierre Nicolet te AMSTERDAM

(74) Gemachtigde:
ir. A. Blokland c.s. te Eindhoven

(54) **Altering RNA sequences in non-coding regions for modification of protein expression**

(57) This invention pertains in general to short sequence elements that can be incorporated in an RNA sequence increase or decrease protein from said RNA. Particularly the invention relates to sequence elements that can be included in the untranslated region of an RNA. The invention further provides methods for modulating protein expression in a cell and therapeutic methods based on modulating protein expression in a subject or a cell.

NL B1 2033707

Title: **Altering RNA sequences in non-coding regions for modification of protein expression**

FIELD OF THE INVENTION

[001] This invention pertains in general to short sequence elements that can be incorporated in an RNA sequence to increase or decrease protein expression from said RNA. Particularly the invention relates to sequences elements that can be included in the untranslated region of an RNA. The invention further provides methods for modulating protein expression in a cell and therapeutic methods based on modulating protein expression in a subject or a cell.

BACKGROUND OF THE INVENTION

[002] The background description includes information that may be useful in understanding the present invention. It is not an admission that any of the information provided herein is prior art or relevant to the presently claimed invention, or that any publication specifically or implicitly referenced is prior art.

[003] Sequence motifs are central factors in eukaryotic mRNA and protein expression control. Motifs in the DNA allow e.g. transcription factors (TF) to bind for gene-specific regulation. In the mRNA, codon motifs in the coding region (CDS) define the tRNA usage for protein assembly. RNA-binding proteins (RBP) and microRNAs (miR) primarily interact with motifs within 5' and 3' untranslated regions (UTR) and define the mRNA stability and protein output. Similarly, sequence motif-guided RNA modifications and RNA editing modify mRNA degradation and the translation efficiency. Sequence-intrinsic factors such as length, GC-content or codon usage also define the mRNA and/or protein abundance. Owing to this plethora of regulatory mechanisms, the mRNA and protein expression levels poorly correlate. To decipher the sequence features that define mRNA and protein expression, large-scale expression measurements combined with powerful mathematical models are pivotal. .

[004] The ability to predict and particularly modulate protein expression based on sequence motifs has broad applications in protein expression in general as well as medical application. For example, there is an ever existing need to increase protein production of specific product in bioreactors. Furthermore, efficacy of RNA based

vaccines could be improved by increasing protein expression. Moreover, many diseases are in some way linked to either over- or under expression of a protein, and could therefore benefit from method to increase or decrease said protein expression. Accordingly, the technical problem underlying the present invention can been seen in the provision of such products, compositions, methods and uses for complying with any of the aforementioned needs. The technical problem is solved by the embodiments characterized in the claims and herein below.

SUMMARY OF THE INVENTION

[005] The inventors have discovered sequence elements present in the untranslated regions (UTR) of the RNA which allow modulation of protein expression. Using a machine learning tool, sequence elements were identified that displayed high predictive capacity for protein expression. The inventors further found that introducing these sequence elements into the UTR of a designed RNA construct allowed for increasing or decreasing the protein expression.

[006] As embodied and broadly described herein, the present invention is directed to the surprising finding that modifying the UTR of an RNA with sequence elements as defined herein, allows for the modulation of protein expression. That is, the protein expression may be increased or decreased, depending on the sequence element used.

[007] Therefore, in a first aspect the invention relates to an RNA molecule having a modified nucleotide sequence, wherein the modified nucleotide sequence comprises one or more sequence elements selected from Table 1 and/or Table 2 in the UTR of the RNA molecule, and wherein the RNA molecule encodes a protein.

[008] In a second aspect the invention relates to a DNA molecule encoding the RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention.

[009] In a third aspect the invention relates to a virus particle encoding the RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention.

[010] In a fourth aspect the invention relates to a cell comprising the RNA molecule having a modified nucleotide sequence according to the first aspect of the invention

or the DNA molecule according to the second aspect of the invention or the virus particle according to the third aspect of the invention.

[011] In a fifth aspect the invention relates to an *in vitro* or *ex vivo* method for modulating the expression of a protein in a cell, the method comprising one or more of:

a) introducing the RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention in the cell;

b) introducing the DNA molecule according to the second aspect of the invention in a cell and allowing the DNA molecule to be transcribed;

c) infecting the cell with the virus particle according to the third aspect of the invention;

d) modifying a DNA molecule in a cell such that the modified DNA molecule transcribes to an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention; and/or

e) modifying an RNA molecule in the cell to obtain an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention.

[012] In a sixth aspect the invention relates to an *in vitro* or *ex vivo* method for expressing a protein in a cell, the method comprising: introducing an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention, a DNA molecule as defined in the second aspect of the invention, the virus particle as defined in the third aspect of the invention in a cell, or obtaining a cell as defined in the fourth aspect of the invention, and allowing the translation of the protein from the RNA molecule, and optionally isolating or obtaining the protein.

[013] In a seventh aspect the invention relates to an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention, a DNA molecule as defined in the second aspect of the invention, the virus particle as defined in the third aspect of the invention, the cell as defined in the fourth aspect of the invention, or the cell obtained or obtainable by the method according to the fifth or the sixth aspect of the invention, for use as a medicament.

BRIEF DESCRIPTION OF THE DRAWINGS

[014] Embodiments of the invention are further described hereinafter with reference to the accompanying drawings, in which:

Fig. 1: Machine learning modelling identifies the sequence features contributing to the protein expression.

[015] Separate machine learning models were trained to learn the protein expression of indicated peripheral blood derived human immune cells and cell lines. The top 100 of the most contributing features are represented for each cell type in a heatmap, indicated for the regions of the mRNA i.e., 5'UTR, CDS, 3'UTR, and for full mRNA/protein features). The feature contribution is expressed as feature importance (grey scale gradient).

Fig. 2: mRNA regions distinctively contribute to gene expression models in human cells.

[016] Feature importance for all sequence features used in the protein abundance models as in Fig. 1, of all tested immune cells and cell lines compiled. Feature importance was extracted and grouped per mRNA region of origin (i.e., 5'UTR, CDS, 3'UTR, full mRNA/protein features).

Fig. 3: Sequence features identified by XGboost models can modulate protein expression

[017] (**A**) Schematic overview of the used strategy, the 3' UTR of GFP was modified by introducing 6 repeated sequence motifs separated by a spacer sequence as indicated, the construct was inserted retroviral expression vector and cells were transduced. (**B**) GFP expression levels normalized to scrambled control assessed by flow cytometry for reporter constructs using 3'UTRs containing 6 occurrences of the indicated motif in $CD8^+$ T cell. (**C**) predicted feature importance (left panel) and absolute fold-change of GFP mean fluorescent intensity (MFI) in indicated cell type (right panel). The Pearson's correlation coefficient (R^2) was computed for each cell type between the absolute GFP fold-change and the feature importance of the corresponding cell type. (**D**) GFP expression levels normalized to scrambled control assessed by flow cytometry for reporter constructs in $CD8^+$ T cells using 3'UTR containing indicated number of occurrences (#), of the motif *CUUUCUU*. Data shown are compiled for 3-5 biological replicates.

Fig. 4: 3'UTR sequence motif modulate protein expression in primary human T cells

[018] GFP expression levels normalized to scrambled control 3'UTRs assessed by flow cytometry for reporter constructs in $CD4^+$ T cells (here indicated as CD8 negative populations) and $CD8^+$ T cells using 3'UTR containing 6 occurrences of the indicated motifs. Data shown are representative of 5 biological replicates.

Fig. 5: 3'UTR sequence motif modulates protein expression in human cells

[019] GFP expression levels normalized to scrambled control assessed by flow cytometry for reporter constructs using 3'UTR containing 6 occurrences of the indicated motifs in $CD4^+$ T cells, K562, HeLa of HEK293T cell lines. Data shown are compiled for 3-5 biological replicates.

Fig. 6: Sequence motif occurrence in 3'UTR modulates the protein expression

GFP expression levels normalized to scrambled control assessed by flow cytometry for reporter constructs in $CD4^+$ and $CD8^+$ T cells using 3'UTR containing indicated number of occurrences (#), of the motif *CUUUCUU*. Data shown are representative of 5 biological replicates.

DESCRIPTION

Definitions

[020] A portion of this disclosure contains material that is subject to copyright protection (such as, but not limited to, diagrams, device photographs, or any other aspects of this submission for which copyright protection is or may be available in any jurisdiction). The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or patent disclosure, as it appears in the Patent Office patent file or records, but otherwise reserves all copyright rights whatsoever.

[021] Various terms relating to the methods, compositions, uses and other aspects of the present invention are used throughout the specification and claims. Such terms are to be given their ordinary meaning in the art to which the invention pertains, unless otherwise indicated. Other specifically defined terms are to be construed in a manner consistent with the definition provided herein. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the present invention, the preferred materials and methods are described herein.

[022] For purposes of the present invention, the following terms are defined below.

[023] As used herein, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise. For example, a method for administrating a pharmaceutical agent includes the administrating of a plurality of molecules (e.g., 10's, 100's, 1000's, 10's of thousands, 100's of thousands, millions, or more molecules).

[024] As used herein, “about” and “approximately", when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of ±20% or ±10%, more preferably ±5%, even more preferably ±1%, and still more preferably ±0.1% from the specified value, as such variations are appropriate to perform the disclosed invention. Unless otherwise clear from context, all numerical values provided herein include numerical values modified by the term “about.”

[025] As used herein, “and/or” refers to a situation wherein one or more of the stated cases may occur, alone or in combination with at least one of the stated cases, up to with all of the stated cases.

[026] As used herein, "at least" a particular value means that particular value or more. For example, "at least 2" is understood to be the same as "2 or more" i.e., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ..., etc. As used herein, the term "at most" a particular value means that particular value or less. For example, "at most 5" is understood to be the same as "5 or less" i.e., 5, 4, 3, ...-10, -11, etc.

[027] As used herein, “comprising” or “to comprise” is construed as being inclusive and open ended, and not exclusive. Specifically, the term and variations thereof mean the specified features, steps or components are included. These terms are not to be interpreted to exclude the presence of other features, steps or components. It also encompasses the more limiting “to consist of”.

[028] As used herein, “conventional techniques” or “methods known to the skilled person” refer to a situation wherein the methods of carrying out the conventional techniques used in methods of the invention will be evident to the skilled worker. The practice of conventional techniques in molecular biology, biochemistry, cell culture, genomics, sequencing, medical treatment, pharmacology, immunology and related fields are well-known to those of skill in the art and are discussed, in various handbooks and literature references.

[029] As used herein, "exemplary" or “for example” means "serving as an example, instance, or illustration," and should not be construed as excluding other configurations, including those disclosed herein.

[030] Throughout this disclosure, various aspects of the invention can be presented in a range format. It should be understood that the description in range format is merely for convenience and should not be construed as a limitation on the scope of the invention. The description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range including both integers and non-integers. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 2.7, 3, 4, 5, 5.3, 6 etc. This applies regardless of the breadth of the range.

[031] As used herein, "cancer" refers to the physiological condition in mammals that is typically characterized by unregulated cell growth. The terms "cancer," "neoplasm," and "tumor," are often used interchangeably to describe cells that have undergone a malignant transformation that makes them pathological to the host organism. Primary cancer cells can be distinguished from non-cancerous cells by techniques known to the skilled person. A cancer cell, as used herein, includes not only primary cancer cells, but also cancer cells derived from such primary cancer cell, including metastasized (secondary) cancer cells, and cell lines derived from cancer cells. Examples include solid tumors and non-solid tumors or blood tumors. Treatment of a cancer in a subject includes the treatment of a tumor in the subject.

[032] Drugs, therapeutic agents, medicaments and pharmaceutical compositions according to the present invention may be formulated for administration by a number of routes, including but not limited to, parenteral, intravenous, intra-arterial, intramuscular, intratumoral and oral. Drugs, therapeutic agents, medicaments and compositions may be formulated in fluid or solid form. Fluid formulations may be formulated for administration by injection to a selected region of the human or animal body. Preferably the cells according to the invention, when used as a medicament, are formulated for administration in fluid formulations, preferably suitable for injection, for

example for intravenous, intra-arterial, intramuscular, or intratumoral delivery or injection.

[033] As used herein, the term “pharmaceutical composition” refers to a composition formulated in pharmaceutically-acceptable or physiologically-acceptable compositions for administration to a cell or subject. The compositions of the invention may be administered in combination with other agents as well, provided that the additional agents do not adversely affect the ability of the composition to deliver the intended therapy. The pharmaceutical composition often comprise, in addition to a pharmaceutical active agent, one or more pharmaceutical acceptable carriers (or excipients).

[034] As used herein, a "subject" is to indicate the organism to be treated e.g., to which administration is contemplated. The subject may be any subject in accordance with the present invention, including, but not limited to humans, males, females, infants, children, adolescents, adults, young adults, middle-aged adults or senior adults and/or other primates or mammals. Preferably the subject is a human patient. In some embodiments, the subject may have been diagnosed with cancer, an immune related disorder, a bleeding disorder, a disorder related to over expression of a protein, a disorder related to under expression of a protein. In some embodiments the subject may be at risk of developing a disease or disorder which may be prevented or ameliorated by vaccination.

[035] As used herein, a “T-cell”, also referred to as “T cell”, can be selected from, for example, the group consisting of inflammatory T-lymphocytes, cytotoxic T-lymphocytes, regulatory T-lymphocytes or helper T- lymphocytes. In another embodiment, said cell can be derived from the group consisting of CD4+ T-lymphocytes and CD8+ T-lymphocytes. They can be extracted from blood or derived from stem cells. T-cells can be obtained from a number of non-limiting sources, including peripheral blood mononuclear cells, bone marrow, lymph node tissue, cord blood, and tumors. In certain embodiments of the present invention, any T cell line available and known to those skilled in the art, may be used. In another embodiment, said cell can be derived from a healthy donor, from a patient diagnosed with cancer or from a patient diagnosed with an infection. In another embodiment, said cell is part of a mixed population of cells which present different phenotypic characteristics.

[036] As used herein, "treatment", "treating", "palliating", “alleviating” and "ameliorating" in the context of a subject to be treated, all refer to an approach for obtaining beneficial or desired results including, but not limited to, therapeutic benefit. By therapeutic benefit is meant eradication or amelioration of the underlying disorder being treated. Also, a therapeutic benefit is achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disorder such that an improvement is observed in the patient, notwithstanding that the patient can still be afflicted with the underlying disorder. As used herein, “prevention" and "preventing" refers to an approach for reducing in part or in full the change of developing adverse effects, for example normally associated with the use of a particular drug or agent. Within the context of the current invention, for example, the terms may refer to preventing, treating or reducing the effects of cancer, an immune related disorder, a bleeding disorder, a disorder related to over expression of a protein, a disorder related to under expression of a protein. In some embodiments the preventing may also refer to the effects of preventing a disease by vaccination.

[037] As used herein the term “nucleic acid” or “polynucleotide” refers to any polymers or oligomers of (contiguous) nucleotides. The nucleic acid may be DNA or RNA, or a mixture thereof, and may exist permanently or transitionally in single-stranded or double-stranded form, including homoduplex, heteroduplex, and hybrid states. The present invention contemplates any deoxyribonucleotide, ribonucleotide or peptide nucleic acid component, and any chemical variants thereof, such as methylated, hydroxymethylated or glycosylated forms of these bases, and the like. The polymers or oligomers may be heterogeneous or homogenous in composition, and may be isolated from naturally occurring sources or may be artificially or synthetically produced. The term “isolated” thus means isolated from naturally occurring sources or artificially or synthetically produced.

[038] As used herein, “Percent (%) nucleic acid sequence identity” with respect to a reference polynucleotide sequence is defined as the percentage of nucleic acid residues in a candidate sequence that are identical with the nucleic acid residues in the reference polynucleotide sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Alignment for

purposes of determining percent nucleic acid sequence identity can be achieved in various ways that are within the skill in the art, for instance, using publicly available computer software such as BLAST, BLAST-2, ALIGN or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for aligning sequences, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared. For purposes herein, however, % nucleic acid sequence identity values are generated using the sequence comparison computer program ALIGN-2. The ALIGN-2 sequence comparison computer program was authored by Genentech, Inc., and the source code has been filed with user documentation in the U.S. Copyright Office, Washington D.C., 20559, where it is registered under U.S. Copyright Registration No. TXU510087. The ALIGN-2 program is publicly available from Genentech, Inc., South San Francisco, Calif., or may be compiled from the source code. The ALIGN-2 program should be compiled for use on a UNIX operating system, including digital UNIX V4.0D. All sequence comparison parameters are set by the ALIGN-2 program and do not vary. In situations where ALIGN-2 is employed for nucleic acid sequence comparisons, the % nucleic acid sequence identity of a given nucleic acid sequence A to, with, or against a given nucleic acid sequence B (which can alternatively be phrased as a given nucleic acid sequence A that has or comprises a certain % nucleic acid sequence identity to, with, or against a given nucleic acid sequence B) is calculated as follows:

100 times the fraction X/Y

where X is the number of nucleic acid residues scored as identical matches by the sequence alignment program ALIGN-2 in that program's alignment of A and B, and where Y is the total number of nucleic acid residues in B. It will be appreciated that where the length of nucleic acid sequence A is not equal to the length of nucleic acid sequence B, the % nucleic acid sequence identity of A to B will not equal the % nucleic acid sequence identity of B to A. Unless specifically stated otherwise, all % nucleic acid sequence identity values used herein are obtained as described in the immediately preceding paragraph using the ALIGN-2 computer program.

[039] As used herein, the terms “protein” and “polypeptide” refer to molecules consisting of a chain of amino acids, without reference to a specific mode of action, size, three dimensional structure or origin. A “fragment” or “portion” or “part” of a

polypeptide may thus still be referred to as a "polypeptide". An "isolated protein" or isolated polypeptide" is used to refer to a protein or polypeptide which is no longer in its natural environment, for example in vitro or in a recombinant host cell.

[040] As used herein the term "construct" or "nucleic acid construct" or "vector" refers to a man-made nucleic acid molecule resulting from the use of recombinant DNA technology and which is used to deliver exogenous DNA into a host cell, often with the purpose of expression in the host cell of a DNA region comprised on the construct. The vector backbone of a construct may for example be a plasmid into which a (chimeric) gene is integrated or, if a suitable transcription regulatory sequence is already present, only a desired nucleic acid sequence (e.g. a coding sequence) is integrated downstream of the transcription regulatory sequence. Vectors may comprise further genetic elements to facilitate their use in molecular cloning, such as e.g. selectable markers, multiple cloning sites and the like.

Detailed description

[041] The invention is defined herein, and in particular in the accompanying claims. Subject-matter which is not encompassed by the scope of the claims does not form part of the present claimed invention.

[042] It is contemplated that any method, use or composition described herein can be implemented with respect to any other method, use or composition described herein. Embodiments or preferences discussed in the context of methods, use and/or compositions of the invention may likewise be employed with respect to any other method, use or composition described herein. Thus, an embodiment or preference pertaining to one method, use or composition may be applied to other methods, uses and compositions of the invention as well.

[043] Any references in the description to methods of treatment refer to the compounds, pharmaceutical compositions and medicaments of the present invention for use in a method for treatment of the human (or animal) body by therapy.

[044] As embodied and broadly described herein, the present invention is directed to the surprising finding that the UTR of RNA can be edited to increase or decrease protein expression. Particularly the inventors found that sequence elements can be

included in the UTR to modulate protein expression. Exemplary sequence elements are listed herein below in Table 1 and Table 2.

[045] Machine learning (ML) can be used to understand complex biological questions such as modelling mRNA abundance from DNA sequence alone, or modelling protein production in massively parallel reporter assays in cell lines. While these synthetic reporter assays are greatly insightful, their library size and complexity, and the use of a strong promoter to drive reporter protein expression limits their use to decipher the endogenous regulation. In addition, reporter assays only include a small sequence, lacks the sequence context of a gene, and omits possible interaction of sequence in different mRNA regions. Thereby, it is paramount to study the code for gene expression in its endogenous context.

[046] To overcome these limitations, some ML models use large-scale observations of gene expression as source of information. To date, these models primarily focus on transcriptional regulation, and have neglected the exquisite interplay between transcriptional, post-transcriptional and post-translational regulation. Furthermore, the current models fail to reveal the sequence features (SF) outside of promoter regions that dictate the intricate regulation of gene expression.

[047] Here, the inventors show that transcriptional regulation information are not required to make accurate prediction of mRNA or protein abundance, highlighting that transcriptional output synergizes with post-transcriptional regulation. Indeed, models using mRNA-level SFs information can partially replace mRNA and proteomics measurements. Yet the extreme complexity of gene expression regulation warrants integration of models for all regulatory mechanisms to further improve prediction in future work.

[048] Therefore, in a first aspect the invention relates to an RNA molecule having a modified nucleotide sequence, wherein the modified nucleotide sequence comprises one or more sequence elements selected from Table 1 and/or Table 2 in the UTR of the RNA molecule, and wherein the RNA molecule encodes a protein.

[049] Tables 1 and 2 describe the sequence elements that the inventors identified using a ML algorithm to correlated mRNA expression levels and sequence data to protein levels. The inventors found that up to 61% of observed mRNA levels and up to 63% of protein expression levels could be explained based on the presence of

sequence elements as defined herein. Table 1 describes sequence elements that are specifically identified in the 5' UTR and thus anticipated to correlate with a modulated protein expression when introduced in the 5' UTR of the RNA molecule. Table 2 describes sequence elements that are specifically identified in the 3' UTR and thus anticipated to correlate with a modulated protein expression when introduced in the 3' UTR of the RNA molecule. Therefore in an embodiment the invention relates to an RNA molecule having a modified nucleotide sequence, wherein the modified nucleotide sequence comprises one or more sequence elements selected from Table 1 and/or Table 2 in the UTR of the RNA molecule, and wherein the RNA molecule encodes a protein, wherein one or more sequence elements selected from Table 1 are comprised in the 5' UTR of the RNA molecule, and/or wherein one or more sequence elements selected from Table 2 are comprised in the 3' UTR of the RNA molecule. In an embodiment the invention relates to an RNA molecule having a modified nucleotide sequence, wherein the modified nucleotide sequence comprises one or more sequence elements selected from Table 1 in the 5' UTR of the RNA molecule, and wherein the RNA molecule encodes a protein. In an embodiment the invention relates to an RNA molecule having a modified nucleotide sequence, wherein the modified nucleotide sequence comprises one or more sequence elements selected from Table 2 in the 3' UTR of the RNA molecule, and wherein the RNA molecule encodes a protein.

[050] When used herein the term RNA molecule refers to a polymeric molecule consisting of ribonucleic acids. The molecule has a sequence defined by the order of the ribonucleic acids. The RNA molecule when used herein encodes a protein, and therefore at least part of the sequence defines a protein coding sequence. The RNA molecule as defined herein further comprises non translated regions, herein referred to as UTR. It is understood that the RNA molecule may comprise a 5' UTR and/or a 3'UTR. Preferably the RNA comprises both a 5' UTR and a 3' UTR. The RNA is preferably an mRNA molecule, but other RNA molecules not limited to synthetic RNA molecules are also comprised within the invention. The invention does not exclude for example the use of synthetic nucleotides or a combination of conventional and synthetic nucleotides.

[051] The RNA molecule of the invention is an RNA molecule having a modified nucleotide sequence. When used herein, modified nucleotide sequence should be interpreted as a sequence that has been deliberately changed. For example, in an mRNA encoding a naturally occurring protein the term modified should be interpreted as the sequence is modified with respect to the naturally occurring sequence to include one or more sequence element as defined herein. Thus, compared to the unmodified RNA, e.g. mRNA, the modified RNA comprises one or more additional sequence elements in the UTR. Thus, it is understood that a sequence element may already be naturally present in the UTR of an RNA, in which case modified RNA intends to mean that one or more additional sequence elements are introduced in the UTR of the RNA. In the case of a sequence encoding an engineered protein, such as a recombinant protein, the term modified should be interpreted that deliberate choices were made in the selection of the RNA sequence encoding the recombinant protein with respect to the sequence of said RNA, in particular the UTRs. Particularly the UTRs were chosen to include sequence elements that modulate protein expression levels.

[052] When used herein the term “sequence element” refers to a short sequence of RNA that, when present in the UTR modulates protein expression from said RNA. The sequence element as defined herein is selected from Table 1 and/or Table 2. Therefore exemplary sequence elements are represented with numbers 1 to 3670 in Tables 1 and 2. For example the sequence element may be 4, 5, 6, 7, 8, or 9 nucleotides long. Without wishing to be bound by theory, the sequence elements identified by the inventors may for example change RNA stability, modulate translation efficiency, thus modulating protein expression. For example, the sequence elements may render the RNA, e.g. an mRNA, more stable or more prone to degradation, thereby increasing or decreasing its half-life. As a further example, the sequence element may help or interfere with engaging the ribosome, direct the RNA to a target site in the cell or engage with agonists or antagonists of translation, thus modulating protein translation.

[053] When used herein the term modulating protein expression is intended to refer to the protein translated from the RNA with the modified UTR. The term modulating may imply that the protein has increased or decreased expression due to the modification of the RNA sequence. Increased or decreased expression when used

herein should be interpreted as with respect to expression of the same protein from an RNA, where no modification in the otherwise identical RNA is present.

[054] The modification as described herein is made to the non-coding part of an RNA molecule, such as an mRNA molecule. Therefore, the modification is preferably in the 3' UTR, or in the 5' UTR or in both the 3' UTR and the 5' UTR. Therefore, in an embodiment the RNA molecule having a modified nucleotide sequence as defined herein has a modified nucleotide sequence comprising one or more sequence elements selected from Table 1 and/or Table 2 in the 3' UTR and/or the 5' UTR of the RNA molecule. In an embodiment the RNA molecule having a modified nucleotide sequence as defined herein has a modified nucleotide sequence comprising one or more sequence elements selected from Table 1 in the 5' UTR of the RNA molecule. In an embodiment the RNA molecule having a modified nucleotide sequence as defined herein has a modified nucleotide sequence comprising one or more sequence elements selected from Table 2 in the 3' UTR of the RNA molecule. Thus in an embodiment the invention relates to an RNA molecule which is modified to have one or more sequence elements selected from Table 1 and/or Table 2 in the UTR of the RNA molecule, and wherein the RNA molecule encodes a protein.

[055] It is understood that the sequence elements described herein in Table 1 and Table 2, when introduced in the UTR of an RNA encoding a protein, result in increased or decreased protein expression compared to an otherwise identical RNA encoding said protein which lacks the sequence elements in the UTR. Therefore, in an embodiment, the RNA molecule having a modified nucleotide sequence having the one or more sequence elements selected from Table 1 and/or Table 2 in the UTR of the RNA molecule result in increased expression of the protein translated from the RNA molecule having a modified nucleotide sequence when introduced in a cell, compared to the same protein translated from an RNA molecule not having a modified nucleotide sequence. Therefore, in an alternative embodiment, the RNA molecule having a modified nucleotide sequence having the one or more sequence elements selected from Table 1 and/or Table 2 in the UTR of the RNA molecule result in decreased expression of the protein translated from the RNA molecule having a modified nucleotide sequence when introduced in a cell, compared to the same protein translated from an RNA molecule not having a modified nucleotide sequence.

[056] It is understood that multiple sequence elements can be included. The inventors demonstrate that the effect of the sequence elements may be increased by including two, three, four or more sequence elements in the UTR. It is further understood that multiple identical sequence elements may be introduced or a combination of different sequence elements can be made. Therefore, in an embodiment, the RNA molecule having a modified nucleotide sequence as defined herein, has an UTR comprising 2, 3, 4, 5, 6, 7, 8, 9, or 10 sequence elements selected from Table 1 and/or Table 2. For example, the RNA molecule having a modified nucleotide sequence as defined herein, has a 5' UTR comprising 2, 3, 4, 5, 6, 7, 8, 9, or 10 sequence elements selected from Table 1 or the RNA molecule having a modified nucleotide sequence as defined herein, has a 3' UTR comprising 2, 3, 4 5, 6, 7, 8, 9, or 10 sequence elements selected from Table 2. Alternatively, the RNA molecule having a modified nucleotide sequence as defined herein, has an UTR comprising 2, 3, 4, 5, 6, 7, 8, 9, or 10 instances of the same sequence element selected from Table 1 and/or Table 2. For example the RNA molecule having a modified nucleotide sequence as defined herein, has a 5' UTR comprising 2, 3, 4, 5, 6, 7, 8, 9, or 10 instances of the same sequence element selected from Table 1 or the RNA molecule having a modified nucleotide sequence as defined herein, has a 3' UTR comprising 2, 3, 4, 5, 6, 7, 8, 9, or 10 instances of the same sequence element selected from Table 2. It is further envisioned that combinations of identical and different sequence elements are used, for example two or more (e.g. 2, 3, 4, 5, 6, or more) of the same sequence elements may be combined with one or more (e.g. 1, 2, 3, 4, 5, 6, or more) additional but different sequence element.

[057] The inventors further found that, when including multiple sequence elements in the UTR of the RNA, an improved effect on protein expression modulation can be achieved when the sequence elements are separated by a spacer. When used herein a spacer is a short sequence of nucleotides, which does not comprise a sequence element as defined herein, nor comprises a repeat sequence. Therefore, the spacer is preferably a randomized sequence which, when included in the mRNA has no or minimal effect on translation or transcription of the mRNA. The spacer is preferably 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more nucleotides long, such as for example 4 to 40 nucleotides, or preferably

6 to 35 or more preferably 8 to 32 nucleotides long. Therefore in an embodiment the RNA molecule having a modified nucleotide sequence as defined herein has an UTR comprising 2 or more sequence elements selected from Table 1 and/or Table 2, wherein each sequence element is separated by a spacer sequence. In an embodiment the RNA molecule having a modified nucleotide sequence as defined herein has a 5' UTR comprising 2 or more sequence elements selected from Table 1, wherein each sequence element is separated by a spacer sequence. In an embodiment the RNA molecule having a modified nucleotide sequence as defined herein has a 3' UTR comprising 2 or more sequence elements selected from Table 2, wherein each sequence element is separated by a spacer sequence.

[058] A particularly suitable application of the present invention is in the expression of recombinant proteins. Thus, in an embodiment the invention describes an RNA molecule having a modified nucleotide sequence as defined herein, wherein the protein encoded by the RNA molecule is a recombinant protein. Particularly preferred is that the modified nucleotide sequence increases expression of the recombinant protein.

[059] The invention further extends to DNA molecules encoding the RNA molecule as broadly defined herein. When used herein, the term "encodes", when used in the context of a DNA molecule encoding an RNA molecule, should be interpreted as able to transcribe an RNA molecule as described herein. Therefore, the DNA should at least comprise the reverse complement sequence of the encoded RNA. Thus, the DNA preferably provides the reverse complement sequence of the RNA sequence which is translated to the protein, and the reverse complement sequence of the UTR, e.g. the 5' UTR and/or the 3' UTR. The DNA may optionally further comprise elements that allow translation of the RNA, such as but not limited to a promoter sequence.

[060] Therefore, in a second aspect the invention relates to a DNA molecule encoding the RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention. In an embodiment the DNA molecule is a single stranded DNA, a double stranded DNA, or a circular DNA, preferably wherein the DNA molecule is a vector, more preferably an expression vector or a viral vector. The DNA molecule may also refer to genomic DNA, for example genomic DNA in a cell. The genomic DNA is preferably altered to encode the modified RNA as described herein.

[061] The DNA when used herein does not exclude the incorporation of non-canonical bases, use of synthetic nucleotides or nucleotide modifications such as but not limited to methylation or acetylation.

[062] The DNA as broadly described herein may be used to express the RNA molecule as broadly described herein. Thus, by encoding the RNA molecule, the RNA molecule can be expressed in a cell. By selecting where the DNA is introduced, or be using a promoter which allows control over transcription (e.g. by using an inducible promoter), transcription of the DNA into the RNA molecule can be controlled, allowing specific modulated protein expression in a specific cell or cells. The term modulating when used in the context of protein translation, generally is intended to mean increased or decreased translation of the protein (e.g. compared to an RNA molecule encoding the same protein but lacking the modification in the UTR as defined herein).

[063] A viral vector when used herein refers to a nucleotide-based means of gene transfer to modify a cell. The viral vector may refer to the nucleotide (the viral genomic material) or the virus particle comprising the genomic material, meaning the genomic material encapsulated in the virus components such as proteins and/or membrane. The virus particle may be a modified virus such as an attenuated virus.

[064] Further envisioned are virus particles encoding the RNA molecule as broadly defined herein. Therefore, in a third aspect the invention relates to a virus particle encoding the RNA molecule having a modified nucleotide sequence as defined in in the first aspect of the invention. When used herein, a virus particle refers to a virus, viroid or attenuated virus comprising a nucleotide that encodes the RNA molecule as broadly defined herein. It is understood that the virus particle may comprise single stranded or double stranded DNA, or single stranded or double stranded RNA, each of which are herein referred to as viral genome. The term “encode” when used in the context of a virus particle should be interpreted as capable of introducing the RNA molecule in a cell when infected by the virus. For example:

- the RNA molecule may be directly comprised in the virus particle and released in the cell upon infection;
- The RNA molecule may be encoded in the viral genome and expressed through reverse transcription and translation, as is the case in for example positive-sense single-stranded RNA viruses such as retroviruses;

- The RNA molecule may be encoded and directly translated from the viral genome. The viral genome may be RNA or DNA based.

The skilled person is aware how viruses or viral vectors may be used to transfer a cell and express an RNA, therefore the above examples should not be construed as limiting. Upon introducing the RNA molecule in the cell, protein may be translated from the RNA, and thus the modified UTR (e.g. the 5' UTR and/or the 3' UTR) of the RNA may modulate protein expression compared to when similar RNA without the modified UTR would have been introduced – preferably in the same way – in the cell.

[065] It is clear to the skilled person that virus particle or viral vectors may be used *in vitro* or *ex vivo* to transfect cells, or may also be used medically to introduce the RNA molecule as defined herein in the cell of a subject.

[066] The invention further extends to cells expressing or comprising the RNA molecule as broadly defined herein. Therefore, in a fourth aspect, the invention relates to a cell comprising the RNA molecule having a modified nucleotide sequence according to the first aspect of the invention or the DNA molecule according to the second aspect of the invention or the virus particle of the third aspect of the invention. The cell is preferably an isolated cell such as a cell or tissue culture, or a blood cell.

[067] In a fifth aspect the invention relates to an *in vitro* or *ex vivo* method for modulating the expression of a protein in a cell, the method comprising one or more of:

a) introducing the RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention in the cell;

b) introducing the DNA molecule according the second aspect of the invention in a cell and allowing the DNA molecule to be transcribed;

c) infecting the cell with the virus particle according to the third aspect of the invention;

d) modifying a DNA molecule in a cell such that the modified DNA molecule transcribes to an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention; and/or

e) modifying an RNA molecule in the cell to obtain an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention.

[068] It is understood that protein expression in a cell may be modulated by introducing an RNA molecule having a modified nucleotide sequence into a cell, or modifying a cell such that an RNA molecule having a modified nucleotide sequence is expressed by the cell. Introduction an RNA molecule having a modified nucleotide sequence into the cell is broadly envisioned as: directly introducing the RNA molecule into the cell or by introducing a genetic carrier into the cell allowing expression of the RNA molecule. the genetic carrier may for example be a DNA molecule (e.g. an expression vector) or a viral particle able to express the RNA into the cell. Methods of introducing DNA, including plasmids and viral vectors, in a cell are known and available to the skilled artisan, such as but not limited to transformation, transduction and transfection. It is further known that RNA may also be introduced through transfection techniques or through a viral particle intermediate.

[069] Cells may be modified such that an RNA molecule having a modified nucleotide sequence is expressed by the cell by editing the DNA (e.g. genomic or mitochondrial DNA) or RNA present in the cell. Methods to edit the genome are known and available to the skilled artisan. Non-limiting examples include restriction enzymes, zinc-finger nucleases (ZNF), Transcription activator-like effector nucleases (TALEN), CRISPR based gene editing technologies such as CRISPR-Cas9, prime editing, and Programmable Addition via Site-specific Targeting Elements (PASTE).

[070] In an embodiment the modulating of the in vitro or ex vivo method results in increased protein expression. In an embodiment the modulating of the in vitro or ex vivo method results in decreased protein expression. It is understood that depending of the context it may be desirable to increase or decrease protein expression, and that the sequence element may be selected accordingly.

[071] In a sixth aspect the invention relates to an in vitro or ex vivo method for expressing a protein in a cell, the method comprising: introducing an RNA molecule having a modified nucleotide sequence as defined in in the first aspect of the invention, a DNA molecule as defined in the second aspect of the invention, the virus particle as defined in the third aspect of the invention in a cell, or obtaining a cell as defined in the fourth aspect of the invention, and allowing the translation of the protein from the RNA molecule, and optionally isolating or obtaining the protein.

[072] It is envisioned the present technology finds application in biotechnological processes where production of large amount of protein in a bioreactor. There is an ever-existing need to further optimize these processes to further increase yield and efficiency. Further applications are in the expression or production of recombinant proteins.

[073] In a seventh aspect the invention relates to an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention, a DNA molecule as defined in the second aspect of the invention, the virus particle as defined in the third aspect of the invention, the cell as defined in the fourth aspect of the invention, or the cell obtained or obtainable by the method according to the fifth aspect of the invention, for use as a medicament.

[074] In an embodiment the invention relates to an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention, a DNA molecule as defined in the second aspect of the invention, the virus particle as defined in the third aspect of the invention, the cell as defined in the fourth aspect of the invention, or the cell obtained or obtainable by the method according to the fifth aspect of the invention for use in treating, preventing or ameliorating a disease in a subject, the use comprising administering the RNA molecule having a modified nucleotide sequence or the DNA molecule or the viral particle or the cell to the subject.

[075] Alternatively, the invention relates to a method of treating, preventing or ameliorating a disease, the method comprising administering an RNA molecule having a modified nucleotide sequence as defined in the first aspect of the invention, a DNA molecule as defined in the second aspect of the invention, the virus particle as defined in the third aspect of the invention, the cell as defined in the fourth aspect of the invention, or the cell obtained or obtainable by the method according to the fifth aspect of the invention to a subject in need thereof.

[076] In an embodiment the invention relates to the RNA molecule having a modified nucleotide sequence, DNA molecule, virus particle or cell for use according to the seventh aspect of the invention, wherein the use comprises increasing or decreasing the expression of a protein in a cell of the subject. Alternatively, the invention relates to a method of treating, preventing or ameliorating a disease comprising administering the RNA molecule having a modified nucleotide sequence, DNA molecule, virus

particle or cell as broadly described herein to a subject in need thereof, wherein such administering results in the increased or decreased expression of a protein in the subject.

[077] In an embodiment the invention relates to the RNA molecule having a modified nucleotide sequence, DNA molecule, virus particle or cell for use according to the seventh aspect of the invention, wherein the disease is selected from: cancer, an immune related disorder, a bleeding disorder, a disorder related to over expression of a protein, a disorder related to under expression of a protein.

[078] In an embodiment the invention relates to the RNA molecule having a modified nucleotide sequence, DNA molecule, virus particle or cell for use according to the seventh aspect of the invention, wherein the RNA molecule having a modified nucleotide sequence, DNA molecule, virus particle or cell is a vaccine.

[079] The technology broadly described herein enables sequence design for biotech application like recombinant protein production, driving cost down by more efficient production. In addition, the invention can speed up therapeutic application such as HLA-ligand prediction, or patient-specific modulation of CAR-T expression to either prevent lethal 'cytokine storm' or to boost efficacy of immune therapies.

[080] A particular interesting application of the technology described herein may be the generation of CAR-T cells, where the sequence elements described herein may be used to increase expression of the chimeric antigen receptor, thus possibly increasing efficacy of such cells. Therefore, the cells as broadly described herein expressing an RNA molecule having a modified nucleotide sequence may be a CAR-T cell, or the *in vitro* or *ex vivo* method as broadly described herein may be used in the preparation of CAR-T cells.

[081] It will be understood that all details, embodiments and preferences discussed with respect to one aspect of embodiment of the invention is likewise applicable to any other aspect or embodiment of the invention and that there is therefore not need to detail all such details, embodiments and preferences for all aspect separately.

[082] Having now generally described the invention, the same will be more readily understood through reference to the following examples which is provided by way of illustration and is not intended to be limiting of the present invention. Further aspects and embodiments will be apparent to those skilled in the art.

EXAMPLES

Example 1

[083] The foregoing description of the specific embodiments will so fully reveal the general nature of the invention that others can, by applying knowledge within the skill of the art (including the contents of the references cited herein), readily modify and/or adapt for various applications such specific embodiments, without undue experimentation, without departing from the general concept of the present invention. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed embodiments, based on the teaching and guidance presented herein.

[084] All references cited herein, including journal articles or abstracts, published or corresponding patent applications, patents, or any other references, are entirely incorporated by reference herein, including all data, tables, figures, and text presented in the cited references. Additionally, the entire contents of the references cited within the references cited herein are also entirely incorporated by references.

[085] It is to be understood that the phraseology or terminology herein is for the purpose of description and not of limitation, such that the terminology or phraseology of the present specification is to be interpreted by the skilled artisan in light of the teachings and guidance presented herein, in combination with the knowledge of one of ordinary skill in the art.

[086] Sequence motifs are central factors in eukaryotic mRNA and protein expression control. Motifs in the DNA allow e.g. transcription factors (TF) to bind for gene-specific regulation. In the mRNA, codon motifs in the coding region (CDS) define the tRNA usage for protein assembly. RNA-binding proteins (RBP) and microRNAs (miR) primarily interact with motifs within 5' and 3' untranslated regions (UTR) and define the mRNA stability and protein output. Similarly, sequence motif-guided RNA modifications (REF) and RNA editing modify mRNA degradation and the translation efficiency (1, 2). Sequence-intrinsic factors such as length, GC-content or codon usage also define the mRNA and/or protein abundance (3, 4). Owing to this plethora of regulatory mechanisms, the mRNA and protein expression levels poorly correlate

(5, 6). To decipher the sequence features that define mRNA and protein expression, large-scale expression measurements combined with powerful mathematical models are pivotal.

[087] Machine learning substantially advanced our understanding of complex biological questions. ML strategies have successfully been used to model mRNA and protein expression, although primarily focused on DNA-level information such as transcription factor binding site or promoter regions (7–9). Reporter assay modelling has highlighted the SFs contribution to mRNA stability (10, 11) and translation efficiency in cell lines (12), or mRNA to protein ratio in tissues (13). Although post-transcriptional regulation is a critical step of mRNA and protein expression, its sequence code to control mRNA and protein abundance in mammalian cells remains unknown.

[088] To extract the endogenous sequence context, ML models can use large-scale observations of gene expression as source of information (9, 14). These ML models are unbiased by strong promoters of reporter assay, can capture the complexity and the code of long sequences, and are able to capture long range feature association. Drawing from these powerful models, we developed an ML strategy to capture the endogenous code for mRNA and protein expression in human immune cells and cell lines.

Materials and methods

Cloning. We first generated a randomly scrambled 3'UTR sequence of 180nt with a GC content of 44.4%. Known 3'UTR motifs with a high contribution to the mRNA or protein models of $CD4^+$ and $CD8^+$ T cell subsets were iteratively removed from the sequence, while maintaining an identical GC content.

To validate the biological effect of specific 3'UTR motifs, 6 occurrences of sequence motifs of interest (*CUUUCUU* - SEQ ID NO: 3, *CUCAGGU* – SEQ ID NO: 2, *UAUUUA* – SEQ ID NO: 5, *AGAAGA* – SEQ ID NO: 4) were inserted into the scrambled sequence. Gene blocks with the 3′ UTR sequences (IDT-Technologies) were cloned into the NotI and BamHI site of pRETRO-SUPER_GFP at the 3'end of the eGFP gene. Sequences were confirmed by Sanger sequencing.

Virus production. FLYRD18 retroviral packaging cells (ECACC 95091902) were cultured in culture-supplemented IMDM (Gibco-BRL), containing 10% fetal bovine serum (FBS; Bodinco), 2 mM L-glutamine, 20 IU/mL penicillin G sodium salts, 20 µg/mL streptomycin sulfate (all Sigma Aldrich) in a humidified incubator at 37°C + 5% CO_2. Cells were plated at a density of 100.000 cells per well in a 6-well plate 16h prior to transfection, after which they were transfected with pRETRO-SUPER_GFP containing the respective 3'UTRs using GeneJammer™ (Agilent Cat #204130) according to the manufacturer's protocol. Cells were incubated at 32ºC + 5% CO_2 for 48h after which retroviral supernatants were harvested. Virus was used immediately or snap-frozen in liquid nitrogen and stored at -80ºC for later use.

T cell culture and retroviral transduction. Studies with human T cells from anonymized healthy donors were performed in accordance with the Declaration of Helsinki (Seventh Revision, 2013) after written informed consent (Sanquin). For T cell activation, 24-well plates were precoated overnight with 0.5 µg/mL rat α-mouse IgG2a (clone MW1483; Sanquin) at 4°C, washed once with PBS, coated with 1 µg/mL αCD3 (clone Hit3a, eBioscience) for at least 3 h at 37°C, and washed once with PBS. Cryopreserved human Peripheral blood mononuclear cells (PBMCs) were defrosted, 1.3×10^6 PBMCs were added per well in IMDM medium supplemented with 1 µg/mL αCD28 (clone CD28.2; eBioscience) and cultured for 48h. Cells were collected and retroviral transduction was performed with Retronectin (Takara) as described by the manufacturer. Briefly, non-tissue cultured treated 24 well plates were coated with 50µg/mL Retronectin (Takara) overnight and washed once with PBS. 400µL/well of viral supernatant was added to the plate, and plates were centrifuged for 45min at 4°C at 4500 rpm (2820g). 1.3×10^6 activated PBMCs were added per well, plates were centrifuged for 5min at 1000rpm, and incubated overnight at 37°C. Media was refreshed after 20h. Cells were cultured with 100 IU/mL recombinant human (rh) IL-2 (Proleukin, Novartis), and 10 ng/mL rhIL-15 (Peprotech). Cells were cultured at a density of $0.5\text{-}1 \times 10^6$ cells/mL. Medium was refreshed every 2 days.

HeLa, HEK293 and K562 culture and transduction. HeLa (ECACC 93021013) and HEK293 (ECACC 85120602) cells were cultured in DMEM (Gibco-BRL), and K562

cells (ECACC 89121407), were cultured in IMDM (Gibco-BRL). Culture media were supplemented with 10% fetal bovine serum (FBS; Bodinco), 2 mM L-glutamine, 20 IU/mL penicillin G sodium salts, 20 µg/mL streptomycin sulfate (all Sigma Aldrich). Cells were cultured in a humidified incubator at 37°C + 5% CO_2. All three cell lines were split every 2 days to keep them at a confluency of max 80% and cultured according to the manufacturers' protocol. K652 cells were spin-occulated with Retronectin (Takara) as described above for T cells. HeLa and HEK293 cells were plated at a confluency of 250.000 per well (in a 6 well plate) and left overnight at 37°C + 5% CO_2. The next morning, medium was refreshed and 400µL virus was added to each well, along with 10mg/mL polybrene (Merck Millipore). One day later, medium was refreshed and cells were cultured for an additional 4 days until the readout was performed.

Flow Cytometry. Four days post retroviral transduction, T cells (or other cells) were harvested, washed with FACS buffer (PBS, containing 1% FBS and 2 mM EDTA), labeled with monoclonal antibodies α-CD4 (clone SK3), α-CD8 (clone SK1) (all Biolegend), and analyzed for GFP expression. Near-IR live/dead marker (Life Technologies) was used to exclude dead cells from analysis. Expression levels were acquired using Symphony flow cytometer (BD Biosciences) and data were analyzed using FlowJo (FlowJo LLC, version 10). Data were analyzed with GraphPad PRISM version 9.

Results:

Sequence features can predict mRNA abundance

[089] We hypothesized that mRNA abundance could be modelled using machine learning and SFs. We first searched through the entire human coding transcriptome (94,348 mRNA isoforms) for SFs, including GC content, transcript length, RNA-binding protein motifs, predicted miR seed score and codon usage (see Methods). The SFs were annotated in the entire mRNA transcript (termed "full mRNA"), or in mRNA regions (i.e. the CDS, 5' and 3'UTR) which combined, yielded 7112 features. We tested the SF library on published mRNA expression data from HEK293T, HeLa and K562 cells (15), and from peripheral blood-derived human T cell, B cell, dendritic cell

(DC), monocyte and granulocyte subsets (16). Prediction accuracy ranged from 31% to 57%.

[090] Auditing the ML models revealed that 145, 602 and 2165 SFs were required to reach 50%, 75% and 95% of the model's full predictive capacity, respectively. Importantly, while XGBoost models cannot disentangle the direct or indirect contribution of a feature, they identify the relative contribution of each feature by quantifying the effect of a feature's loss. Therefore, our models reveal their individual and coordinated contribution to mRNA abundance prediction in human cells.

Effective prediction of protein abundance

[091] We next trained XGBoost models to predict protein abundance. The prediction accuracy of protein abundance models averaged 41.2% for primary immune cells, and reached a striking prediction accuracy of >60% in HeLa and HEK293T cell lines (data from (20, 21)). The SF usage was sufficient to recapitulate the origins of the cell types. Most of the feature contributing to protein levels were sequence motifs, including G-rich (5'UTR), CT-rich (CDS, 3'UTR), AU-rich motifs (3'UTR). Surprisingly, the importance of RBP-motifs in the CDS was higher than this in the 5' and 3'UTR, in accordance with the mapping of 100s of RBP. Other substantial contributors were the occurrence of PTM sites (e.g. acetylation and ubiquitination). mRNA and protein abundance models alike rely on a complex mixture of SFs for accurate prediction.

SF models capture dynamic gene expression during T cell differentiation and activation

[092] We next sought to understand how SFs contribute to the protein expression control in human T cells. T cells are in particular an attractive model because they (i) can be precisely activated, and (ii) rapidly remodel their gene expression landscape through transcriptional and post-transcriptional control (16, 17). We therefore determined the capacity of SFs models to predict mRNA abundance, translation rate, and protein abundance of blood-derived naive CD4+ T cells activated with α-CD3/α-CD28 (data from (17)). The prediction accuracy of mRNA and protein models improved upon T cell activation, while this of translation rate models decreased. Clustering revealed that the contribution of individual SFs alters during T cell activation.

Interestingly, codon usage also contributed to mRNA abundance models, suggesting that models captured translation-dependent and translation-independent mRNA decay processes. All T cell activation models captured the preparedness state for rapid activation of T cells, exemplified by the high importance of the Lysine-encoding AAG tRNA in the early activation of T cells, as observed in murine T cells. Thus, our models prove useful to decipher the complex gene regulation during T cell activation.

[093] We next determined whether SFs differentially contribute to the models of mRNA abundance, protein translation rate and protein abundance in T cells. mRNA modification sites (m1A, m5C, m6A, m7G) belonged to the most important features in all 3 models. Translation rate and protein models preferentially used CDS parameters such as length, GC content and TOP-like motifs (*CUUC* and *CUUUUC*). Combined, integration of models trained on different cellular measurements (mRNA, translation rate, protein) captures the biological effect of SFs.

[094] Our ML pipeline also allowed us to (I) isolate the feature usage that is associated with gene expression pattern throughout T cell activation; (II) define SFs that predict mRNA and protein abundance during T cell differentiation from naïve T cells to effector and memory T cells, and to (III) isolate SFs associated with mRNA abundance of dysfunctional tumor-infiltrating CD8+ T cells from melanoma. In conclusion, alterations in feature importance during T cell activation and differentiation reveals how regulatory mechanisms are rearranged to permit timely protein production.

ML-aided 3'UTR design to manipulate protein expression

[095] Having identified SFs that predict protein expression, we questioned whether SFs could be exploited to modulate protein expression. To this end, we focused on SFs found in the 3'UTRs. The 3'UTR region was consistently found to comprise highly important features alongside CDS, yet 3'UTR features could manipulate expression level without requiring codon re-arrangement. We isolated the following motifs: *CUUU*, *AGAA*, *CUCAGG*, and *AUUUA* (AU-rich element, strong negative effect (18)), which were amongst the highest contributing features in protein models (Fig. 1, 2). As RNA sequence motifs may not have a strict motif requirement for interpretation by post-transcriptional regulators (i.e. for RBP binding (19)), we isolated derivatives of the 4

above-mentioned sequence motifs. For all four motifs, we found a high heterogeneity in feature importance between sequence relatives. We thereby selected for sequence motifs of >= 6nt that contained the above-mentioned motifs and had a high feature importance in protein models of CD4+ and CD8+ T cells. *CUUUCUU*, *AGAAGA*, *CUCAGGU*, and *UAUUUA* motifs were selected for validation. We created a 180nt-long 3'UTR scrambled sequence devoid of strong regulatory motifs, and inserted 6 occurrences of the 4 motifs to be validated. We then fused the 3'UTR sequences at the 3'end of the eGFP reporter gene and retrovirally introduced the GFP reporter genes containing different 3'UTR sequences into human primary T cells to measure the GFP protein levels in primary T cells.

[096] Whereas 3'UTRs containing *AGAAGA* motifs (SEQ ID NO: 4) or *UAUUUA* motifs (SEQ ID NO: 5) suppressed the GFP protein expression, the *CUCAGGU* and *CUUUCUU* motifs (SEQ ID NO: 1 and SEQ ID NO: 2) in the 3'UTR augmented GFP protein levels compared to the scrambled control (SEQ ID NO: 1) 3'UTR (Fig. 3A). When using 3'UTR constructs with 1, 2, 4 or 6 *CUUUCUU* motifs (SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8 and SEQ ID NO: 3 respectively) in the 3'UTR we found that the increase in protein expression depended on the number of motifs (Fig. 3B).

[097] Next, to validate the cell-type SF specificity revealed in XGBoost models, we assessed the expression of the GFP reporters in 3 cells lines alongside T cell subsets (Fig. 3B). While *CUUUCUU* motifs induced higher GFP expression in all cell types, yet to a different extent, all 3 other motifs induced cell types-specific modulation in expression, correlating with model predictions (Fig. 3B).

Exemplary sequences that have been used in the experiments are described below. Sequence elements, where present are underlined. Sequence elements are separated by exemplary spacer motifs.

SEQ ID NO: 1
Scrambled_no_important_motifs_GC_44.4%
CUAUUCCGGUAUAUACUAUAUCCGAUCAUAUACUAAACUGUACGCCUCUGCAACAUAUCGCCUAGU
AGCUAUAGACUGAGCGAUCAAGUAUCGAGUAUCAAGUGGUACAGUCGCGAGUGCGUAUAGUGCCAG
CUAGAGUACUCUGAUCUGACCGCCAAUCGUGAAAGUUACGAUCUCAUC

SEQ ID NO: 2
CUCAGGUx6_Scrambled_no_important_motifs_GC_46.7%
CUAUUCCGGCUCAGGUCUAUAUCCGAUCACUCAGGUAACUGUACGCCUCUCAGGUAUAUCGCCUAG
UAGCUAUAGACUCUCAGGUCAAGUAUCGAGUAUCAAGUGGUACACUCAGGUGUGCGUAUAGUGCCA
GCUAGAGUACUCUGAUCUCUCAGGAAUCGUGAAAGUUACGAUCUCAUC

SEQ ID NO: 3
CUUUCUUx6_Scrambled_no_important_motifs_GC_39.4%
CUAUUCCGGCUUUCUUUAUAUCCGAUCACUUUCUUAACUGUACGCCUCUUUCUUAUAUCGCCUAGU
AGCUAUAGACUCUUUCUUCAAGUAUCGAGUAUCAAGUGGUACACUUUCUUGUGCGUAUAGUGCCAG
CUAGAGUACUCUGAUCUCUUUCUUAAUCGUGAAAGUUACGAUCUCAUC

SEQ ID NO: 4
AGAAGAx6_Scrambled_no_important_motifs_GC_41.7%
CUAUUCCGGUAGAAGAUAUAUCCGAUCAUAGAAGAAACUGUACGCCUAGAAGACAUAUCGCCUAGU
AGCUAUAGACUGAGAAGACAAGUAUCGAGUAUCAAGUGGUACAGAGAAGAGUGCGUAUAGUGCCAG
CUAGAGUACUCUGAUCUAGAAGACAAUCGUGAAAGUUACGAUCUCAUC

SEQ ID NO: 5
UAUUUAx6_Scrambled_no_important_motifs_GC_35%
CUAUUCCGGUUAUUUAUAUAUCCGAUCAUUAUUUAAACUGUACGCCUUAUUUACAUAUCGCCUAGU
AGCUAUAGACUGUAUUUACAAGUAUCGAGUAUCAAGUGGUACAGUAUUUAGUGCGUAUAGUGCCAG
CUAGAGUACUCUGAUCUUAUUUACAAUCGUGAAAGUUACGAUCUCAUC

SEQ ID NO: 6
Scrambled_CUUUCUUx1
gaucaCUAUUCCGGCUUUCUUUAUAUCCGAUCAUAUACUAAACUGUACGCCUCUGCAACAUAU
CGCCUAGUAGCUAUAGACUGAGCGAUCAAGUAUCGAGUAUCAAGUGGUACAGUCGCGAGUGCG
UAUAGUGCCAGCUAGAGUACUCUGAUCUGACCGCCAAUCGUGAAAGUUACGAUCUCAUCgc

SEQ ID NO: 7
Scrambled_CUUUCUUx2
gaucaCUAUUCCGGCUUUCUUUAUAUCCGAUCACUUUCUUAACUGUACGCCUCUGCAACAUAU
CGCCUAGUAGCUAUAGACUGAGCGAUCAAGUAUCGAGUAUCAAGUGGUACAGUCGCGAGUGCG
UAUAGUGCCAGCUAGAGUACUCUGAUCUGACCGCCAAUCGUGAAAGUUACGAUCUCAUCgc

SEQ ID NO: 8
Scrambled_CUUUCUUx4

gaucaCUAUUCCGGCUUUCUUUAUAUCCGAUCACUUUCUUAACUGUACGCCUCUUUCUUAUAU
CGCCUAGUAGCUAUAGACUCUUUCUUCAAGUAUCGAGUAUCAAGUGGUACAGUCGCGAGUGCG
UAUAGUGCCAGCUAGAGUACUCUGAUCUGACCGCCAAUCGUGAAAGUUACGAUCUCAUCgc

REFERENCES

1. W.Gilbert, et al., Science (1979).
2. H. Shi, et al., Mol Cell. 74, 640–650 (2019).
3. P. Anderson, Nat Rev Immunol. 10, 24–35 (2010).
4. B. P. Nicolet, et al., Immunol Rev. 304, 10–29 (2021).
5. C. Buccitelli, et al., Nat Rev Genet. 21, 630–644 (2020).
6. B. Schwanhäusser, Nature. 473, 337–342 (2011).
7. J. Zrimec, Front Mol Biosci. 8 (2021),
8. E. D. Vaishnav, Nature. 603, 455–463 (2022).
9. J. Zrimec, Nat Commun. 11, 6141 (2020).
10. M. Rabani, Mol Cell. 68, 1083-1094.e5 (2017).
11. A. J. Litterman, Genome Research . 29, 896–906 (2019).
12. P. J. Sample, Nat Biotechnol. 37, 803–809 (2019).
13. B. Eraslan, Mol Syst Biol. 15, 1–25 (2019).
14. J. T. Cuperus, Genome Res. 27, 2015–2024 (2017).
15. T. F. Martinez, Nat Chem Biol. 16, 458–468 (2020).
16. G. Monaco, Cell Rep. 26, 1627-1640.e7 (2019).
17. M. A. Cassatella. Trends Immunol. 40, 648–664 (2019).
18. S. Mitschka,.Elife. 10, e65700 (2021).
19. C. C. Thoreen, L.. Nature. 485, 109–113 (2012).
20. J. C. Rieckmann,. Nat Immunol. 18, 583–593 (2017).
21. T. Geiger, Molecular & Cellular Proteomics, doi:10.1074/mcp.M111.014050.

TABLE 1

5' UTR MOTIFS

no.	motif
1	CUCUUU
2	GCGCUG
3	AUUUG
4	GGUG
5	CGGGC
6	CUGAGUU
7	UGCUG
8	AAAGGAG
9	GGAUUC
10	GGACAAG
11	AGAGG
12	AGGCAGU
13	CGGGCA
14	AGCC
15	CGGGGGG
16	CAGACAA
17	CCAC
18	UCUGCUGC
19	CCGCGCA
20	CCGCGGG
21	CUAC
22	CGGCCCG
23	AUUU
24	CAUCCAG
25	ACAA
26	GAUC
27	AGCCCCCA
28	GGGG
29	CCCUCC
30	CCGCUG
31	UGGAGAU
32	UGUGGA
33	CGGGA
34	UUUUUGU
35	GCACACA
36	UUCAG
37	CUUUUUG
38	AAGAGGA
39	UCACAGG
40	GGGGC
41	GACUGGA
42	CUGU
43	UGGGC
44	CUCAA
45	AGAA
46	ACAAAUC
47	GCAGC
48	UCAC
49	UGGG
50	CUUG
51	UCGG
52	UGCU
53	AUGCGCG
54	UGUGU
55	CUUUG
56	GUGG
57	CCGC
58	UAGGCA
59	CCAG
60	CUUGU
61	CUGA
62	ACAGUCA
63	UUAGGGU
64	AGAGAGG
65	UCAU
66	CCUG
67	GGGUAA
68	CUCCCUCC
69	ACCCUU
70	ACCAAA
71	ACACAAA
72	CCUUA
73	GCGAGCG
74	UAGGUA
75	UCUCAGG
76	AGAC
77	UGUU
78	CUCAGGU
79	CUAG
80	UCCG
81	GGCUUU
82	CUGAG
83	UCUAC
84	CCUGU
85	CUCUUC
86	CGGAGCG
87	CCCCCGG
88	UCUCA
89	CAGGGA
90	ACCCCCAC
91	UUGGG
92	GAAG
93	UAGG
94	AGGAGAG
95	GAAA
96	ACAU
97	UGGAGAA
98	ACCGC
99	CAGUCAC
100	UGGAC
101	CGGCGCC
102	CCAU
103	ACUUGC
104	GCUGC
105	UGGGAAG
106	UACA
107	ACCAC
108	AAAA
109	GCGGGUC
110	CUCC
111	UGGGA
112	UGGUGG
113	CCGAGGG
114	GGGA
115	UCCGCUUU
116	UCGGGC
117	UUGA
118	GAGACAU
119	AUGAGAG
120	CAUCAUC
121	CUUCAAC
122	UUUCUCG
123	AGGG
124	CCGCUU
125	ACUUUC
126	AACCUG
127	GGAAGCA
128	UACAUC
129	GAGC
130	CGCC
131	UUUCUGG
132	CUCUUCAC
133	UACGUC
134	GGUUCUU
135	CACCUUCC
136	AGUAG
137	AUAGAA
138	GGGAGCC
139	AGAGGUA
140	UGCAGC
141	CUCUC
142	CUCA
143	CCCUCCU
144	CCAGCGC
145	UGAUGGC
146	CCUCCCC
147	CCCCCUCC
148	GAGUGC
149	CGCUGGG
150	ACAC
151	CCUGA
152	CCCAGCC

153	CCAGCGA	193	UGAUGAC	233	GCAC	273	AAGACUA
154	UUUCAGA	194	CACGGA	234	ACUCU	274	AGAGUGA
155	CCGGCCA	195	AAAGAAG	235	AAGAAGAAG	275	ACGGC
156	AGCAG	196	CUACAGC	236	CUUU	276	UCCUGC
157	CGCCCCA	197	UCUUG	237	GCUU	277	AAAACUA
158	CUUUCUU	198	CCCCGCG	238	CCACACA	278	CCUUU
159	AUGAAAA	199	UUGCCA	239	AAGAA	279	CCCACCA
160	CUGCCC	200	CGGGGCG	240	UAGGCU	280	AGGGC
161	UGCGGUC	201	UACUAAG	241	AAGGUG	281	AGGACAA
162	UUCUUG	202	UCCUCC	242	CCUCC	282	CUUUUUC
163	CCUUACC	203	CUGC	243	AGCAGAGGA	283	AGGAUUA
164	GACAGA	204	ACCCAU	244	CCCGGCG	284	AGGUAU
165	CAAAUCA	205	CCAAACA	245	AUUC	285	UGAUGC
166	GGAGGG	206	UAGACAG	246	UUGAUUG	286	CACG
167	CUUGGAA	207	CCCCCGA	247	CACC	287	GGACGAC
168	CAUC	208	GCCGGAG	248	AGAGGAA	288	CUAAU
169	UCUGC	209	UGGGGU	249	UUGGUGG	289	ACUUCC
170	GGGGG	210	AUAAAA	250	CCCUGGA	290	UCUCC
171	UGUUU	211	CGAAGGAC	251	UUCAU	291	AAUCUUG
172	AAUCAAA	212	AGACAAG	252	CUUCACG	292	CACGUA
173	GGGUGC	213	AGCAUGA	253	AGAGA	293	UCUGCUGU
174	GAAGAA	214	CCACCCA	254	ACCCCCAA	294	GGAGAA
175	CUCCGGA	215	GUGGUUU	255	CCUUUCU	295	ACCGGG
176	UCGUGC	216	UUGCCG	256	ACCUUC	296	AGAAG
177	ACCUCC	217	UCUU	257	AUGAAAG	297	GCUGGCC
178	CCGGGGA	218	CGCCGGA	258	CCCGGC	298	AUCAUC
179	AAGAC	219	AGCGAGC	259	CCGCAGG	299	CGACGCA
180	UGGGG	220	UAGGG	260	UGGAGCU	300	UCCUGU
181	GCGGGCC	221	AGAGAAG	261	CACAGCA	301	UUGCCU
182	CCUUUU	222	UGUA	262	UGCC	302	AUUUU
183	GUGGU	223	AUUGCACC	263	CACAUC	303	CGUUCUU
184	GUUUA	224	CAACAGC	264	CGGGCGG	304	AAAUAAA
185	ACUUUCU	225	GGAGGAA	265	GGCG	305	CUCG
186	GCGC	226	UGAAGGC	266	CUUAG	306	AUUUUCU
187	CACAGUG	227	AGGGA	267	AGAGAGC	307	CGGCCGG
188	AAAGGGG	228	CUUC	268	CGCGGGG	308	GCGCGGG
189	UGUUC	229	GAAGGAC	269	AACA	309	CGGGCCG
190	AAGUGUU	230	CGGUGG	270	UGGUGAC	310	GGAUAC
191	CACAGGA	231	AGGCA	271	UCCGC	311	ACACC
192	AUUUAUUU	232	CCUGG	272	UCCCAC	312	CUCU

313	GGUUCGU	353	GGUGUGG	393	UGUGUGG	433	CCUCACG
314	ACAGC	354	UCGCUG	394	AGGAGC	434	CCUUUC
315	GAGGAGGA	355	CGCAUC	395	AAGAAG	435	AUUUC
316	AGGGU	356	CCUUCCA	396	GGAUGCC	436	GGGAGGC
317	AUGUG	357	UGGAAAG	397	GACCCCCA	437	AUGACCA
318	AGGUCU	358	GAAGGGG	398	UUUUUUG	438	GGCUGC
319	UGGAGU	359	CCUC	399	CUAAAA	439	CCUAC
320	ACAAUCA	360	AGGGG	400	UGGUGAA	440	CUGCUU
321	UACUAC	361	CCUUGU	401	CAGAGGA	441	UUGCUG
322	AAGACAG	362	AGGAG	402	ACCUCU	442	UAUGU
323	CGCU	363	UUGUA	403	AUUUUUU	443	UCCCAA
324	UGGAGCC	364	GCGCG	404	CUUCAC	444	CACAGC
325	GAGUAC	365	GGCACGG	405	UUACUUG	445	ACAGG
326	GGGUGU	366	CCCUGC	406	GGACAGG	446	GCCGGA
327	UCAG	367	UCCUCU	407	ACAAC	447	CGCUGCG
328	GGAAGCC	368	AAAGGGA	408	GACUCCCG	448	GGGUGCA
329	AGGAU	369	ACUUUU	409	AGGAA	449	CCCAGCA
330	CAGCGGA	370	UUUU	410	GGCCCU	450	AGGACAG
331	AGAAC	371	GGCUCCCG	411	UAGGGC	451	UAAG
332	UCGGG	372	CCGCCGA	412	ACGCUU	452	UGGAU
333	CUGCGGU	373	CGUGU	413	CACCGGA	453	CAUU
334	CUCGGAC	374	CCCUGU	414	ACCACC	454	CCGGGCG
335	AGGAGCA	375	CCCGUUU	415	AAAAUCA	455	GAUCAAG
336	GGGUGCC	376	UGGUGGA	416	GCAGCGC	456	CCACC
337	GUAGGGU	377	ACGCGAG	417	GUUU	457	UCUUUC
338	GGCCACCG	378	CGCGGCG	418	GGCCACCA	458	AUAAAU
339	GGAGAU	379	CUUUUG	419	GCGGGAA	459	GGCUUAA
340	AGAGAAC	380	GAGGAAC	420	CGGACAG	460	CGGAGGA
341	AAAGG	381	AUUA	421	CGCAGCG	461	UCUCACG
342	AAAUUCAC	382	CUGCCG	422	AGACGGG	462	UUUC
343	GGGGA	383	CGGACGG	423	CAGGAC	463	ACUUGU
344	CGCG	384	GGGAGCA	424	UGUGGGG	464	UUGUGG
345	ACUUCU	385	ACGCUC	425	CCUCG	465	CGCAGGU
346	AGUAGCC	386	ACCCCAA	426	AAACUCA	466	UUGUG
347	AUUUAU	387	UUAUUA	427	GUAACG	467	CGCGGA
348	CCGCUC	388	CACAUA	428	CGCAGA	468	AGAGGAG
349	AUCCCC	389	GCGCGC	429	AUCG	469	UCUUU
350	CGGUG	390	UGGGGAG	430	AACACC	470	GGCCCCUG
351	CUUUC	391	UUAAUGU	431	CCGUCGA	471	UCUCUGG
352	UUGCCC	392	CGCUGCA	432	CGACGCG	472	AAGAAC

473	CAGCGGG	513	CGUCG	553	GUGGUGU	593	CACAGA
474	AAAGC	514	AGGACA	554	AGGUAA	594	AGAGGGG
475	UUUAUU	515	CUGCUG	555	UCUUC	595	GACAGGG
476	GGACAGA	516	CCUUCCC	556	GUGGUUG	596	CGCACGG
477	CAGAGGG	517	UGGAGGAC	557	ACUUUUU	597	UCGCUC
478	ACGCGCG	518	CCGAGCA	558	CUGGGGU	598	CGCGCGG
479	GGGGGCUG	519	UACUGC	559	CACCGGU	599	CUCCUUCC
480	CCUUCCU	520	CUGAACA	560	AGGUUU	600	CCGCCGG
481	CGGAGAA	521	CUAA	561	GCGUGC	601	CCCGCCG
482	UGGUGA	522	GAGGAG	562	UCUAG	602	GGGGGG
483	UCUCU	523	UUUGCUGU	563	GGUUCCCA	603	CGCAGC
484	AGGU	524	AAAAAAA	564	CCGGCGG	604	AUCAUCAU
485	AAAAU	525	CGCUCGG	565	GAGGGAG	605	UUUUUG
486	CUACACG	526	UUUUU	566	CGGUGGG	606	UGCGUA
487	CGAAGCA	527	CCCCUUCC	567	CAUCUUG	607	GCAGG
488	GAGUAGG	528	UUGAUGU	568	AGACAGA	608	CUACUCG
489	GCCCAG	529	ACUGC	569	GUGAUGG	609	AGACGGA
490	GGAGGGA	530	GUGCUG	570	CCCUAA	610	AUGACA
491	GUCCUCAG	531	UUGUU	571	CCACAGG	611	UCACC
492	CGCUUU	532	UGAAGGA	572	CGCCCCG	612	GAGAGGA
493	GCGGGCA	533	UCUUCAC	573	GAUGAAG	613	AUCUUCAU
494	GUUGU	534	CAGUGA	574	CUAAG	614	UCUCCG
495	AAUAAA	535	CAGCCGA	575	ACGGG	615	CCUAA
496	GAAAGAAG	536	AGAAAAC	576	AGUAGG	616	CCUUGC
497	UCAAC	537	CAUCGC	577	CCAACCC	617	GGUGGU
498	AGCACC	538	GAAAGAA	578	CUUA	618	GACUUAA
499	AAAGAAA	539	UCCGCUGC	579	CAGCCUCC	619	AUUUAA
500	CAGUGG	540	UGUCGC	580	UUAGAG	620	UUGCUU
501	GAAGGA	541	GGGU	581	ACCGUC	621	UGGGGAU
502	AGGAACG	542	UGCGUC	582	CACGGC	622	AUCUGAC
503	GCAAACA	543	AAUGAAU	583	GCGCUU	623	CCUUC
504	AUAUAA	544	CCCUCCA	584	GGGUGA	624	CCUUCC
505	CUGG	545	UCUUCA	585	CCGGACG	625	AUCAC
506	UUAG	546	CGGAG	586	CCUUG	626	CUAU
507	CUUUCUUUU	547	UCGCACA	587	AAAAGA	627	AGGAGCG
508	CAUCUG	548	AAAAA	588	CCCAGCG	628	CUUCUCUCU
509	CGCUGCGGC	549	GUCUUU	589	CCGCGGA	629	AAAAACA
510	UCUUCC	550	GCUGGAC	590	GCGCGGU	630	CGCGGC
511	CCGUGCG	551	CCGUGCA	591	CCCUUCCC	631	CAUCG
512	GGGAGGG	552	GACGUC	592	AAUCUUA	632	GCCUGCG

633	GGAGGGG	673	CUCUUG	713	CACUUCAU	753	ACCUGU
634	UUUUUC	674	UGGUGU	714	UGUGGUG	754	AUCCCA
635	UCCUUC	675	UGGAGGG	715	UUGCACA	755	UGGGAAA
636	AGGAUAG	676	UGUGUUU	716	CGACCGA	756	ACGCUCA
637	UCUCUCG	677	GGAUAU	717	CUCGCUGU	757	UGUUGUU
638	CGAGCGA	678	CAACUCA	718	CUGCCUCC	758	GUUUUG
639	CAGCCGU	679	AUGACG	719	CCCAGGG	759	GACUAAC
640	UCAAA	680	UGCAUG	720	CCAAGCA	760	GCGCGCU
641	GGUUGGCG	681	AUGACAG	721	UGAAGCA	761	CCAAUCA
642	GCCCCUA	682	CUCUCU	722	ACCCGC	762	AAAAAA
643	UGGGU	683	CCCGACU	723	GUUCAAG	763	AAAAAAAA
644	AGGAAAA	684	CCGCGCC	724	GGUCCCCA	764	AAAAAAAAA
645	UUUAAU	685	CCUGCUGU	725	AUGCCC	765	AAAAAAC
646	UUAGGA	686	CGCAGGG	726	UUGCACG	766	AAAAAAG
647	UAGAGA	687	UCUAU	727	GGGGU	767	AAAAACC
648	GUGCUC	688	GUCUGU	728	AGGAACA	768	AAAAAU
649	UUCUUU	689	UACAGC	729	CUGACUA	769	AAAACGA
650	CCGUGGG	690	UGACUG	730	GAUUUAA	770	AAAAGAA
651	GCGAGCA	691	UCUGG	731	UUAGGG	771	AAAAUA
652	GAAGGAG	692	CCCU	732	UGCUGUU	772	AAAAUU
653	GUUUUC	693	CGGCGGG	733	AAGAAGAA	773	AAACAAA
654	GCUGGUC	694	AGAGCAGG	734	UCGUAC	774	AAACACA
655	GGGCGCC	695	CCUCA	735	AACUAAG	775	AAACCAAA
656	CGCCUUCC	696	UCCCUU	736	GCAUAC	776	AAAGAAGAA
657	GAGGUU	697	AUUAAC	737	GCCCACU	777	AAAGACAAA
658	CUAAAU	698	ACCUGC	738	UGUGUGUG	778	AAAGAGA
659	GCACUCA	699	AAUACAG	739	CCAAACC	779	AAAGAGG
660	CCGCCC	700	UCUGU	740	ACCUUU	780	AAAGGAA
661	CCGCGCG	701	UGCGGC	741	CCUAG	781	AAAUAA
662	UCCUUU	702	CCCUCCC	742	UGGGAAU	782	AAAUACU
663	CUGCUC	703	CAUAUA	743	GGCUCCCCC	783	AAAUAU
664	ACUAGCU	704	CCGCCCA	744	AAAUACA	784	AAAUUA
665	UGGUAA	705	CGGUGCG	745	GGACAAA	785	AAAUUCA
666	CCUCUUCC	706	UCUCG	746	CGCUGGA	786	AAAUUCCA
667	CGAGU	707	UUUUUGG	747	UGCUGAC	787	AAAUUCCC
668	UCUUCU	708	GUUUUA	748	GGAGCGC	788	AAAUUU
669	CGCGGAA	709	UUGCUC	749	AGGAGGAU	789	AACAACA
670	UUUGCUUU	710	GAAACUA	750	GAUUCCCA	790	AACAACG
671	GAUCUG	711	UCUUGC	751	GAUCAUU	791	AACAACU
672	AGGAC	712	UAGGA	752	UAGGUU	792	AACAUC

793	AACAUCA	833	ACACAAG	873	ACUUAU	913	AGGACCA
794	AACAUCG	834	ACACACA	874	ACUUUAA	914	AGGACCG
795	AACAUCU	835	ACACG	875	AGAAAAA	915	AGGAGA
796	AACUACG	836	ACACGAA	876	AGAAAAU	916	AGGAGAA
797	AACUUUAU	837	ACACGAC	877	AGAAAGU	917	AGGAGAAC
798	AAGACGA	838	ACACGAG	878	AGAAAUU	918	AGGAGAU
799	AAGGACC	839	ACACUCA	879	AGAAGA	919	AGGAGAUAA
800	AAGGGGAA	840	ACAUAAA	880	AGAAGAAC	920	AGGAGGA
801	AAGU	841	ACAUACA	881	AGAAGAGC	921	AGGAGUA
802	AAUAA	842	ACAUUAAA	882	AGAAGGAC	922	AGGAUAA
803	AAUAAAA	843	ACAUUCA	883	AGACAA	923	AGGAUGA
804	AAUAAAC	844	ACCCAA	884	AGACAAA	924	AGGAUGG
805	AAUAAU	845	ACCCGA	885	AGACAGG	925	AGGAUUG
806	AAUACCA	846	ACCCGU	886	AGACGAA	926	AGGGAA
807	AAUACGA	847	ACCUUAAA	887	AGACGAG	927	AGGGAU
808	AAUACUA	848	ACGAACC	888	AGACGAU	928	AGGGUAU
809	AAUAUA	849	ACGAAUA	889	AGACGUA	929	AGGUAC
810	AAUAUU	850	ACGAAUC	890	AGACGUU	930	AGGUAG
811	AAUCAAG	851	ACGACA	891	AGAGAAA	931	AGGUUA
812	AAUCAAU	852	ACGACG	892	AGAGACA	932	AGUACAG
813	AAUCAUA	853	ACGAGAG	893	AGAGACC	933	AGUGUGA
814	AAUCAUG	854	ACGAGAU	894	AGAGACG	934	AGUGUGG
815	AAUCAUU	855	ACGCGCA	895	AGAGAGA	935	AGUGUGU
816	AAUCCAA	856	ACGCGCC	896	AGAGGGA	936	AUAAAAA
817	AAUCCAG	857	ACGCGCU	897	AGAGUGG	937	AUAAAAC
818	AAUGAAA	858	ACGCUA	898	AGAGUGU	938	AUAAAAG
819	AAUGAAG	859	ACGCUG	899	AGAUAAA	939	AUAAAC
820	AAUGAUA	860	ACUAAC	900	AGAUACU	940	AUAAAG
821	AAUGAUG	861	ACUAACA	901	AGAUAU	941	AUAAAGG
822	AAUGAUU	862	ACUAACC	902	AGAUCC	942	AUAAUA
823	AAUUAA	863	ACUAACU	903	AGAUGGAC	943	AUAAUU
824	AAUUAU	864	ACUAAU	904	AGCAC	944	AUAAUUA
825	AAUUUA	865	ACUAAUA	905	AGCAUG	945	AUAAUUG
826	AAUUUU	866	ACUAAUC	906	AGCAUGC	946	AUACAUG
827	ACAAACA	867	ACUAGAG	907	AGCCUCAG	947	AUACAUU
828	ACAAACC	868	ACUAGCC	908	AGCGGA	948	AUACUUG
829	ACAAAUA	869	ACUAGCG	909	AGCUCCCA	949	AUACUUU
830	ACAAGAC	870	ACUAGGA	910	AGCUGUU	950	AUAGCA
831	ACAAGGAC	871	ACUGCA	911	AGGAAAG	951	AUAGGGA
832	ACACAAC	872	ACUGG	912	AGGACC	952	AUAGGGU

953	AUAUAGA	993	AUUUA	1033	CAGACGU	1073	CCCACGA
954	AUAUAU	994	AUUUAC	1034	CAGAGAGAC	1074	CCCACGG
955	AUAUAUA	995	AUUUAUUUU	1035	CAGAGGU	1075	CCCAGGA
956	AUAUAUAU	996	AUUUUA	1036	CAGCGGU	1076	CCCAUGA
957	AUAUAUC	997	AUUUUC	1037	CAGCUUCC	1077	CCCCCCA
958	AUAUUA	998	AUUUUG	1038	CAGUAGA	1078	CCCCCCC
959	AUAUUU	999	AUUUUU	1039	CAGUCGG	1079	CCCCCCG
960	AUCAAA	1000	AUUUUUA	1040	CAGUGC	1080	CCCCCCU
961	AUCAAC	1001	AUUUUUC	1041	CAGUGU	1081	CCCCGAC
962	AUCACC	1002	AUUUUUG	1042	CAUACAG	1082	CCCCGCA
963	AUCGUA	1003	AUUUUUUUA	1043	CAUCAAC	1083	CCCCGGA
964	AUCUUC	1004	AUUUUUUUC	1044	CAUCAU	1084	CCCCGGG
965	AUCUUCAC	1005	CAAAACA	1045	CAUCCAA	1085	CCCGCCA
966	AUGAACA	1006	CAAAAU	1046	CAUCUUA	1086	CCCGCGA
967	AUGACAA	1007	CAACACA	1047	CAUUUUU	1087	CCCGCGG
968	AUGACCG	1008	CAACCGG	1048	CCAACCA	1088	CCCGCUGC
969	AUGAGAA	1009	CAAGGAG	1049	CCAACCG	1089	CCCGCUGU
970	AUGAGCA	1010	CAAGGGG	1050	CCAACGA	1090	CCCGCUUC
971	AUGAGCG	1011	CAAUACA	1051	CCAACGC	1091	CCCGCUUU
972	AUGCACA	1012	CAAUACU	1052	CCAAGGA	1092	CCCGGCA
973	AUGCACG	1013	CAAUUCA	1053	CCAAUCC	1093	CCCGGGA
974	AUGCCA	1014	CACA	1054	CCAAUGG	1094	CCCGGGG
975	AUGCCG	1015	CACAA	1055	CCACACG	1095	CCCUCCG
976	AUGCGCA	1016	CACACGU	1056	CCACAGA	1096	CCCUCGA
977	AUGCGCC	1017	CACAGAC	1057	CCACCGA	1097	CCCUCGG
978	AUGCGCU	1018	CACAGCG	1058	CCACGCA	1098	CCCUCU
979	AUGCUC	1019	CACAGGU	1059	CCACGCU	1099	CCCUGAC
980	AUGCUU	1020	CACCCGA	1060	CCACGGA	1100	CCCUGCA
981	AUGUAUAU	1021	CACCCGU	1061	CCACUCA	1101	CCCUGCG
982	AUGUGU	1022	CACCCUCC	1062	CCACUCG	1102	CCCUGGG
983	AUGUGUAU	1023	CACCGAC	1063	CCACUGG	1103	CCCUUC
984	AUGUGUGU	1024	CACCGCA	1064	CCAUACA	1104	CCCUUU
985	AUGUU	1025	CACGGAC	1065	CCAUACC	1105	CCGACAG
986	AUUAAA	1026	CACGUC	1066	CCAUCCC	1106	CCGACCA
987	AUUAAAC	1027	CACUACAC	1067	CCAUGGU	1107	CCGACCG
988	AUUAAU	1028	CACUGAC	1068	CCAUUAAA	1108	CCGACGA
989	AUUAUA	1029	CACUGCA	1069	CCAUUC	1109	CCGACGG
990	AUUAUU	1030	CACUUCAG	1070	CCAUUCA	1110	CCGAGCG
991	AUUCAAG	1031	CAGACAG	1071	CCAUUCC	1111	CCGAGGA
992	AUUGCAC	1032	CAGACGA	1072	CCCACCG	1112	CCGAUGA

1113	CCGCCCG	1153	CGACCCA	1193	CGGAAGA	1233	CUACAAA
1114	CCGCCUCC	1154	CGACGGA	1194	CGGAC	1234	CUACAAC
1115	CCGCUA	1155	CGAGCGC	1195	CGGACCA	1235	CUACAAU
1116	CCGCUAU	1156	CGAGCGG	1196	CGGACCG	1236	CUACAGG
1117	CCGCUUCC	1157	CGAGGCA	1197	CGGACGA	1237	CUACGAA
1118	CCGGCCG	1158	CGAUGGA	1198	CGGACGU	1238	CUACGAC
1119	CCGGCGA	1159	CGCACCA	1199	CGGAGAGAA	1239	CUACUGG
1120	CCGGGCA	1160	CGCACCG	1200	CGGAGCA	1240	CUAUAGA
1121	CCGGGGG	1161	CGCACGU	1201	CGGAGGG	1241	CUAUAUA
1122	CCGUCCA	1162	CGCAGCA	1202	CGGAGGU	1242	CUAUAUC
1123	CCGUCCG	1163	CGCAGGA	1203	CGGAGUA	1243	CUCGCUGC
1124	CCGUCGG	1164	CGCAGGC	1204	CGGCACA	1244	CUCGCUUC
1125	CCGUGGA	1165	CGCAUA	1205	CGGCAGA	1245	CUCGCUUU
1126	CCUAACC	1166	CGCCCGA	1206	CGGCCCA	1246	CUCGGAA
1127	CCUACCC	1167	CGCCCGG	1207	CGGCCGA	1247	CUCGUGU
1128	CCUAGGG	1168	CGCCCGU	1208	CGGCCGU	1248	CUCUUA
1129	CCUAU	1169	CGCCCUCC	1209	CGGCCUCC	1249	CUCUUCUCU
1130	CCUAUCC	1170	CGCCCUUG	1210	CGGCGCA	1250	CUGCCA
1131	CCUCAGG	1171	CGCCGCA	1211	CGGCGCG	1251	CUGCCU
1132	CCUCCCA	1172	CGCCGCG	1212	CGGCGGA	1252	CUGCUA
1133	CCUCCCU	1173	CGCCGGG	1213	CGGCGGU	1253	CUGCUUCC
1134	CCUCGGA	1174	CGCCGGU	1214	CGGCUUCC	1254	CUGGAUU
1135	CCUCGUCC	1175	CGCCGUA	1215	CGGGCCA	1255	CUGUAUAU
1136	CCUCU	1176	CGCGCAA	1216	CGGGCGA	1256	CUGUAUGU
1137	CCUCUCG	1177	CGCGCCA	1217	CGGGGCA	1257	CUGUCUG
1138	CCUCUGG	1178	CGCGCCC	1218	CGGGGGA	1258	CUGUGUGU
1139	CCUCUUU	1179	CGCGCCG	1219	CGGUAA	1259	CUUCAGG
1140	CCUGC	1180	CGCGCGA	1220	CGGUAG	1260	CUUCUCG
1141	CCUGCGG	1181	CGCGGAC	1221	CGGUAU	1261	CUUCUGG
1142	CCUGCUGC	1182	CGCGGAG	1222	CGGUCCA	1262	CUUGCUGC
1143	CCUGCUUC	1183	CGCGGCA	1223	CGGUCCG	1263	CUUGCUGU
1144	CCUGCUUU	1184	CGCGGGA	1224	CGGUCGG	1264	CUUGCUUC
1145	CCUUCCUU	1185	CGCGGUU	1225	CGGUCU	1265	CUUGCUUU
1146	CCUUCU	1186	CGCGUA	1226	CGGUGA	1266	CUUGGAC
1147	CCUUCUGUG	1187	CGCGUC	1227	CGGUGCA	1267	CUUUA
1148	CCUUUCC	1188	CGCGUCA	1228	CGGUGGA	1268	CUUUUA
1149	CCUUUCCC	1189	CGCUCCA	1229	CGUACAG	1269	CUUUUC
1150	CCUUUUU	1190	CGCUCCG	1230	CGUUCGU	1270	CUUUUCU
1151	CGAACCA	1191	CGCUCGA	1231	CGUUGGU	1271	CUUUUU
1152	CGAAGGA	1192	CGCUGAC	1232	CGUUGUU	1272	CUUUUUA

1273	CUUUUUU	1313	GACUCCCA	1353	GAUUAGUG	1393	GGACGGA
1274	CUUUUUUC	1314	GACUCCUG	1354	GAUUCCCG	1394	GGACGGC
1275	GAAAAC	1315	GACUGC	1355	GAUUCCCU	1395	GGACGGG
1276	GAAAACA	1316	GACUUCUU	1356	GAUUCCUA	1396	GGACUGGGA
1277	GAAAACC	1317	GAGACGA	1357	GAUUCCUG	1397	GGACUGUA
1278	GAAACGA	1318	GAGACUA	1358	GCAAUCA	1398	GGAGAG
1279	GAAAGAAU	1319	GAGCACUG	1359	GCAGCGA	1399	GGAGC
1280	GAAAGGAGA	1320	GAGCGAG	1360	GCAUAA	1400	GGAGCGA
1281	GAAAUCA	1321	GAGCGGA	1361	GCAUACA	1401	GGAGGA
1282	GAACACA	1322	GAGGAAG	1362	GCAUGC	1402	GGAGGAAC
1283	GAACUCA	1323	GAGGAAGAA	1363	GCAUUCA	1403	GGAGGAC
1284	GAAGAAGA	1324	GAGGCAU	1364	GCCCAA	1404	GGAGGAG
1285	GAAGAAGAA	1325	GAGGGAA	1365	GCCCAC	1405	GGAGGGC
1286	GAAGAGGAA	1326	GAGGGUA	1366	GCCGUC	1406	GGAGUA
1287	GAAGGAA	1327	GAGGGUG	1367	GCCUAC	1407	GGAGUAGCU
1288	GAAGGUA	1328	GAGUAGA	1368	GCCUGC	1408	GGAGUG
1289	GAAGGUG	1329	GAGUAGU	1369	GCGCGCA	1409	GGAGUU
1290	GAAUAC	1330	GAGUCAU	1370	GCGCGCC	1410	GGAUCCGG
1291	GAAUACA	1331	GAUAAAA	1371	GCGCGCG	1411	GGAUGC
1292	GAAUCCCG	1332	GAUAAAC	1372	GCGCGGC	1412	GGAUGCA
1293	GAAUGA	1333	GAUACUA	1373	GCGCUC	1413	GGAUGGA
1294	GAAUGC	1334	GAUAUAC	1374	GCGGGAC	1414	GGAUGGAG
1295	GAAUGG	1335	GAUCAAA	1375	GCGGGUA	1415	GGAUGGC
1296	GAAUGU	1336	GAUCAAC	1376	GCGUAC	1416	GGAUUAA
1297	GAAUUA	1337	GAUCAAU	1377	GCUAGUA	1417	GGCCAAA
1298	GAAUUAA	1338	GAUCACC	1378	GCUGGAA	1418	GGCCACUG
1299	GAAUUCA	1339	GAUCACCA	1379	GCUGGCA	1419	GGCCCACA
1300	GACAAA	1340	GAUCACUG	1380	GCUGGUA	1420	GGCCCCCA
1301	GACCACCA	1341	GAUCAUA	1381	GCUUGC	1421	GGCCCCCG
1302	GACCACCG	1342	GAUCAUG	1382	GGAAAGAC	1422	GGCCCCUA
1303	GACCCCCG	1343	GAUCCCCA	1383	GGAACAGC	1423	GGCCGCAG
1304	GACCCCUA	1344	GAUCCCCG	1384	GGAAGAAC	1424	GGCCUCC
1305	GACCCCUG	1345	GAUCCCUA	1385	GGAAGAAG	1425	GGCUAC
1306	GACCCGCA	1346	GAUCCCUG	1386	GGAAGGA	1426	GGCUCCAA
1307	GACCCGG	1347	GAUCGCUG	1387	GGAAGGC	1427	GGCUCCCA
1308	GACCUGCG	1348	GAUGAAA	1388	GGAAUGAC	1428	GGCUCCUA
1309	GACGAAC	1349	GAUGAAU	1389	GGACAA	1429	GGCUCCUG
1310	GACUAC	1350	GAUGAUA	1390	GGACGAA	1430	GGCUGAUG
1311	GACUACCA	1351	GAUGAUG	1391	GGACGAG	1431	GGCUGCAA
1312	GACUCAAG	1352	GAUGAUU	1392	GGACGCUG	1432	GGGACGU

1433	GGGAGGA	1473	GUCCUCCG	1513	UAAAUAA	1553	UAGUAU
1434	GGGCGCG	1474	GUCGGAA	1514	UAAAUU	1554	UAGUAUA
1435	GGGGGA	1475	GUCGGAC	1515	UAAGAC	1555	UAGUAUG
1436	GGGGGAA	1476	GUCUUA	1516	UAAUAA	1556	UAGUGA
1437	GGGGGAC	1477	GUGAAUGA	1517	UAAUAC	1557	UAGUGAA
1438	GGGGGAG	1478	GUGACGA	1518	UAAUAU	1558	UAGUGC
1439	GGGGGGG	1479	GUGAUGU	1519	UAAUGC	1559	UAGUUAA
1440	GGGUAC	1480	GUGAUUG	1520	UAAUUA	1560	UAGUUGG
1441	GGGUGG	1481	GUGAUUU	1521	UAAUUU	1561	UAGUUUA
1442	GGGUGGA	1482	GUGCGCC	1522	UACAGA	1562	UAGUUUG
1443	GGGUGGC	1483	GUGCGCG	1523	UACAUA	1563	UAUAAA
1444	GGUCACCG	1484	GUGCGCU	1524	UACGGA	1564	UAUAAU
1445	GGUCACUG	1485	GUGCUA	1525	UACGGC	1565	UAUACAG
1446	GGUCAGUG	1486	GUGCUU	1526	UACGUA	1566	UAUAUA
1447	GGUCCCCG	1487	GUGGUGG	1527	UACUAAC	1567	UAUAUAC
1448	GGUCCCUG	1488	GUGUAGA	1528	UACUAAU	1568	UAUAUAG
1449	GGUCGCCG	1489	GUGUAGG	1529	UACUAGC	1569	UAUAUU
1450	GGUCUGC	1490	GUGUAGU	1530	UACUGAC	1570	UAUAUUUU
1451	GGUGGUGGU	1491	GUGUAUAU	1531	UAGACA	1571	UAUCAAC
1452	GGUGUGU	1492	GUGUAUGU	1532	UAGAGAGG	1572	UAUCCAA
1453	GGUGUUG	1493	GUGUG	1533	UAGAGU	1573	UAUCCAG
1454	GGUGUUU	1494	GUGUGU	1534	UAGCUGUG	1574	UAUGUG
1455	GGUUCCCG	1495	GUGUGUAU	1535	UAGGAA	1575	UAUGUU
1456	GGUUCCUA	1496	GUGUGUGU	1536	UAGGAAA	1576	UAUUAA
1457	GGUUCCUG	1497	GUUCCAGGU	1537	UAGGAAG	1577	UAUUAU
1458	GGUUGGU	1498	GUUGGAA	1538	UAGGAGG	1578	UAUUAUUUU
1459	GGUUGUU	1499	GUUGGAC	1539	UAGGAU	1579	UAUUUA
1460	GGUUGUUG	1500	GUUGUGC	1540	UAGGAUA	1580	UAUUUAUA
1461	GGUUUAA	1501	GUUGUU	1541	UAGGAUG	1581	UAUUUAUU
1462	GUAAUGG	1502	GUUGUUC	1542	UAGGGA	1582	UAUUUU
1463	GUAAUGU	1503	GUUUAGA	1543	UAGGGU	1583	UCAAAA
1464	GUAAUUG	1504	GUUUC	1544	UAGGUAA	1584	UCAACA
1465	GUAAUUU	1505	GUUUG	1545	UAGGUGG	1585	UCACAAC
1466	GUAGGGA	1506	GUUUUGC	1546	UAGGUUA	1586	UCACACG
1467	GUAGUGG	1507	GUUUUUA	1547	UAGGUUG	1587	UCACUCG
1468	GUAGUGU	1508	GUUUUUC	1548	UAGUAAA	1588	UCACUGG
1469	GUAGUUG	1509	GUUUUUG	1549	UAGUAAC	1589	UCAGUCAC
1470	GUAGUUU	1510	UAAAAA	1550	UAGUAAG	1590	UCAUAC
1471	GUAUGU	1511	UAAAAU	1551	UAGUAC	1591	UCAUC
1472	GUCCCCUG	1512	UAAAUA	1552	UAGUAGG	1592	UCAUGC

1593	UCCAGCAA	1633	UGAAGCU	1673	UGGAAGAAC	1713	UUAAAA
1594	UCCAGU	1634	UGAAGGAC	1674	UGGACAA	1714	UUAAAU
1595	UCCCAU	1635	UGAAGGU	1675	UGGACAG	1715	UUAAUA
1596	UCCCCA	1636	UGACAAA	1676	UGGAGAC	1716	UUAAUGG
1597	UCCCCAA	1637	UGACAAG	1677	UGGAGCA	1717	UUAAUU
1598	UCCCCAU	1638	UGACAGA	1678	UGGAGGA	1718	UUAAUUA
1599	UCCCCCAA	1639	UGACAGG	1679	UGGAGGC	1719	UUAAUUG
1600	UCCCCCAU	1640	UGACGAA	1680	UGGAGGU	1720	UUAAUUU
1601	UCCCCUA	1641	UGACGAG	1681	UGGAGUA	1721	UUACACG
1602	UCCCCUU	1642	UGACGGA	1682	UGGGGAA	1722	UUACAGG
1603	UCCCGA	1643	UGACGGG	1683	UGGUAC	1723	UUACAUG
1604	UCCCUA	1644	UGAGUGA	1684	UGGUGAU	1724	UUACAUU
1605	UCCGCUGU	1645	UGAGUGG	1685	UGGUGC	1725	UUACUCG
1606	UCCGCUUC	1646	UGAGUGU	1686	UGGUGCA	1726	UUACUGG
1607	UCCGGA	1647	UGAUAC	1687	UGGUGCC	1727	UUACUUU
1608	UCCUAC	1648	UGAUGAA	1688	UGGUGCU	1728	UUAGGGA
1609	UCCUCUUC	1649	UGAUGAAC	1689	UGGUGGC	1729	UUAGUGG
1610	UCGAGAU	1650	UGAUGAU	1690	UGGUGGU	1730	UUAGUGU
1611	UCGCUA	1651	UGAUGCA	1691	UGUAAAUA	1731	UUAGUUG
1612	UCGCUU	1652	UGAUGCC	1692	UGUACAG	1732	UUAGUUU
1613	UCGGGCA	1653	UGAUGCU	1693	UGUACAU	1733	UUAUAA
1614	UCUAA	1654	UGAUGGA	1694	UGUAUAU	1734	UUAUAU
1615	UCUAUCU	1655	UGAUGGU	1695	UGUCAAC	1735	UUAUAUUUU
1616	UCUCUC	1656	UGCAGA	1696	UGUCCAG	1736	UUAUUU
1617	UCUGA	1657	UGCAGAGUG	1697	UGUGAGG	1737	UUAUUUAAA
1618	UCUGCUUC	1658	UGCAGGCA	1698	UGUGAUG	1738	UUAUUUAUU
1619	UCUGCUUU	1659	UGCAUA	1699	UGUGGG	1739	UUAUUUU
1620	UCUUA	1660	UGCAUC	1700	UGUGUG	1740	UUCAC
1621	UCUUCUCU	1661	UGCAUGA	1701	UGUGUGA	1741	UUCGCUGC
1622	UCUUCUU	1662	UGCAUGC	1702	UGUGUGU	1742	UUCGCUGU
1623	UCUUGU	1663	UGCAUGU	1703	UGUGUGUGU	1743	UUCGCUUC
1624	UCUUUCU	1664	UGCCGCC	1704	UGUGUGUU	1744	UUCUCCAA
1625	UCUUUU	1665	UGCCGUU	1705	UGUGUUG	1745	UUCUUC
1626	UCUUUUU	1666	UGCGGA	1706	UGUUACUA	1746	UUCUUCAA
1627	UGAACCA	1667	UGCGUG	1707	UGUUCGU	1747	UUCUUCAU
1628	UGAAGAA	1668	UGCUAAC	1708	UGUUCUU	1748	UUCUUUU
1629	UGAAGAAC	1669	UGCUAC	1709	UGUUG	1749	UUGAUGG
1630	UGAAGAC	1670	UGCUGC	1710	UGUUGGU	1750	UUGAUUU
1631	UGAAGAU	1671	UGCUGCC	1711	UGUUGU	1751	UUGCUA
1632	UGAAGCC	1672	UGCUGUUUU	1712	UGUUUGU	1752	UUGGUGU

1753 UUGGUUG
1754 UUGGUUU
1755 UUGUAUGU
1756 UUGUGC
1757 UUGUGU
1758 UUGUGUAU
1759 UUGUGUGU
1760 UUGUUUU
1761 UUUAAA
1762 UUUAAUUUU
1763 UUUAG
1764 UUUAGUU
1765 UUUAUA
1766 UUUAUUU
1767 UUUAUUUA
1768 UUUAUUUU
1769 UUUCACG
1770 UUUCAGG
1771 UUUCAUUUU
1772 UUUCUUU
1773 UUUGAUUUU
1774 UUUGCUGC
1775 UUUGCUUC
1776 UUUGGUU
1777 UUUGUUU
1778 UUUUA
1779 UUUUAA
1780 UUUUAAGUU
1781 UUUUAC
1782 UUUUAU
1783 UUUUAUUGA
1784 UUUUAUUGU
1785 UUUUAUUUA
1786 UUUUAUUUU
1787 UUUUUA
1788 UUUUUCC
1789 UUUUUCCC
1790 UUUUUCG
1791 UUUUUCU
1792 UUUUUU
1793 UUUUUUC
1794 UUUUUUU
1795 UUUUUUUUU

TABLE 2

3' UTR MOTIFS

no.	motif
1796	CCUUUC
1797	AAGUGUU
1798	CCACC
1799	GAGGAG
1800	AUAGAA
1801	UAUGU
1802	GCGC
1803	AAACACA
1804	UGGUGC
1805	ACCUUU
1806	ACGAGAG
1807	CUGU
1808	CGCCCGG
1809	AAAA
1810	CCUC
1811	CCUCC
1812	UCUUG
1813	CGCCCUCC
1814	ACUCU
1815	AGGG
1816	CUGA
1817	CUUUG
1818	GCUU
1819	GCUGC
1820	ACAA
1821	UAUUUA
1822	GAAACUA
1823	UGUUU
1824	CCCU
1825	UUCAU
1826	GAUUCCCA
1827	AGGAGA
1828	CCAU
1829	GUUU
1830	CCCUGU
1831	CCAUGGU
1832	AACA
1833	GCCUGCG
1834	UAUCCAG
1835	AUCGUA
1836	UCUU
1837	GAGC
1838	GGCCACUA
1839	UGCU
1840	CACC
1841	CUGAG
1842	CCCUCU
1843	AGGU
1844	AGAGGAG
1845	AUUUG
1846	UUGA
1847	AAUGAUA
1848	UUAG
1849	UGUU
1850	ACAAC
1851	UUUUU
1852	ACUGC
1853	GCUGGUC
1854	UAUUUAUA
1855	UAUAUUUU
1856	UCAUAC
1857	AGAAG
1858	UAGUUGG
1859	AGCC
1860	CUCC
1861	ACCCCCAA
1862	CGCAGC
1863	CUGG
1864	CCGCUUCC
1865	CCUUUCCC
1866	ACUGG
1867	AGAA
1868	CUUUC
1869	CUCAA
1870	AAAAACC
1871	UGUGU
1872	ACAU
1873	UGGUGA
1874	UGCC
1875	AUUGCAC
1876	UAGUAUA
1877	CCCUGAC
1878	AAAAAA
1879	CACAGC
1880	AAGAA
1881	CUACUCG
1882	AGCAG
1883	GAUGAAG
1884	AAUCAUG
1885	UGUA
1886	CUUG
1887	UGCUGUU
1888	GAAA
1889	AAACUCA
1890	UUCAG
1891	AUUU
1892	UUUUUU
1893	CUGC
1894	UCAAAA
1895	UCCCUA
1896	UCUGG
1897	CCUAG
1898	UUUUAC
1899	GGGUGA
1900	GAAGAAGAA
1901	AAUAAAA
1902	UCUUUU
1903	CCUUA
1904	AUUUUUG
1905	ACACC
1906	ACUUGC
1907	CCGAGGG
1908	CUAG
1909	CCGACAG
1910	UUUUA
1911	CCACUGG
1912	CCUCU
1913	AGCACC
1914	UGAAGGA
1915	AAUACCA
1916	UUAUAU
1917	UAGUAAA
1918	GUUUG
1919	UACAUC
1920	CGGACAG
1921	UAUGUG
1922	CUUA
1923	AAAGAAA
1924	CAUCG
1925	UUUUUCCC
1926	CCUAA
1927	AUUUU
1928	UUGUGU
1929	UACAUA
1930	CCGGGGG
1931	CCUAACC
1932	GUUUUGC
1933	AUUUAA
1934	UUGUAUGU
1935	ACGGG
1936	GCUGGAC
1937	AAAAAAAA
1938	CAGAGGG
1939	UACAGC
1940	AACACC
1941	CUUUCUUUU
1942	AUAAAG
1943	CUCUUC
1944	UUUAUA
1945	CUAAU
1946	AGGGG
1947	CGCUCCA
1948	CCUCGGA
1949	GAUGGAGGA
1950	UUGUU
1951	CCAUUAAA

1952	UCCAGU	1992	CAUC	2032	AUCUUC	2072	UUGAUGG
1953	GGGAGCC	1993	AGGAAAA	2033	CGCAGCA	2073	UCCUCC
1954	AGAGG	1994	GUGG	2034	AGAGA	2074	CAAGGAG
1955	AAAUAU	1995	CCUCCCU	2035	AUUAUA	2075	UAGGUA
1956	GGCCCCUG	1996	GGAAGCA	2036	AAAUAAA	2076	AGGAACG
1957	UAAAAA	1997	UAGGA	2037	GUGGU	2077	UCUUGU
1958	CUCAGGU	1998	GGGA	2038	GGGGA	2078	AUAAAA
1959	CCAGCGC	1999	AGGACCA	2039	AUAAAC	2079	CUGCCU
1960	GUUUUA	2000	GGGGGAA	2040	UCUCU	2080	AAGAC
1961	ACAC	2001	UAAUAA	2041	UAUGUU	2081	UAUUUU
1962	CAUCAUC	2002	AGAAC	2042	AGGGC	2082	GGGUAA
1963	GGGGGA	2003	AGACGUU	2043	CCAUUCA	2083	CACUUCAG
1964	GGGUGG	2004	UGUCCAG	2044	ACAAUCA	2084	CACA
1965	GUAUGU	2005	AUUUUG	2045	GACAAA	2085	CCCACCG
1966	AGAGGGA	2006	ACUAGCC	2046	UUAAAU	2086	UUUUUC
1967	AAUCAAU	2007	UUAGGA	2047	AAUAA	2087	GGCCCCCA
1968	ACCUCC	2008	GAGUAGU	2048	GAGAGGA	2088	AUUUC
1969	AGAAAAA	2009	UAUAUU	2049	GAUACGA	2089	UACUGAC
1970	AGGAGAU	2010	GAGACAU	2050	UUUUAU	2090	AUGCCG
1971	AGGUUU	2011	ACUUUCU	2051	AUUUUU	2091	CUUCAC
1972	UGCGUG	2012	CCUUUU	2052	CCCUUCCC	2092	CUUUUG
1973	AGGAUGG	2013	AUGACA	2053	CCCUCCC	2093	AAUUUA
1974	AUCG	2014	GAGGGUG	2054	UCAUC	2094	GAUCAUU
1975	UGGGGU	2015	CCCUGC	2055	UCCUUU	2095	CUAU
1976	CACUUCAU	2016	UGAUGGC	2056	CCUGCUGC	2096	UCUUC
1977	AAACAAA	2017	AGAGAAG	2057	UCCCCUU	2097	UGGUGCC
1978	UCUAU	2018	UCGCUG	2058	CUGUCUG	2098	GCAC
1979	AAUAUA	2019	CAAUUCA	2059	AGGUCU	2099	AGAGAAA
1980	UCCUGU	2020	AGGAGC	2060	UGGUGCA	2100	ACGGC
1981	GCCCAG	2021	UUGUGG	2061	AUGUU	2101	GGGGGG
1982	UUUUUUU	2022	GAUCACUA	2062	CCGUGGG	2102	CGCCCCG
1983	UCUUA	2023	UCUUU	2063	UCUCC	2103	UCUAG
1984	UCUUCA	2024	UCCCCCAA	2064	UGAUGCU	2104	CUGGGGU
1985	AUUC	2025	CCACCCA	2065	GAGGGUA	2105	GGAGGG
1986	AAUCUUA	2026	CCCUUU	2066	AGGAG	2106	CCGC
1987	UAUUAU	2027	UCAU	2067	GAAUAC	2107	UGUUCUU
1988	UUUU	2028	AUAAAU	2068	CCUGU	2108	UCUCA
1989	UACA	2029	CCGCUC	2069	AAUUAA	2109	UUUUAA
1990	CUCU	2030	UGACAAG	2070	CGAAGCA	2110	GGACAAA
1991	AAAUUA	2031	GGUGGU	2071	GGAGGGA	2111	GAAAACA

2112	CUGUGUAU	2152	ACUUUU	2192	AUACAUU	2232	UCUGCUUU
2113	UGUUG	2153	CUAA	2193	AUGCCC	2233	AAAAGAA
2114	UGUACAG	2154	CUCUCU	2194	CGCACGU	2234	UCUGCUUC
2115	AGGAC	2155	CCAG	2195	GGGU	2235	CUAAAA
2116	UUAAUU	2156	GGUUGGU	2196	UCAACA	2236	GGAUGGC
2117	AUGCGCA	2157	UGACAGA	2197	AGGUAG	2237	UGCGGC
2118	AAUAAA	2158	CUUUUUA	2198	GGGGC	2238	AUCAAA
2119	CUAC	2159	CUUGGAC	2199	GCUGGCC	2239	CUCUUU
2120	CUUC	2160	GGCCUCC	2200	GAAUGC	2240	GAGCACUG
2121	CGCG	2161	UUAGAG	2201	CCCGACU	2241	UUGUG
2122	CUUUUUU	2162	UGUGUGG	2202	GAUCAUA	2242	CCUGA
2123	CCCAGCC	2163	UAGGG	2203	ACUUCU	2243	AAUAAAC
2124	UGUGUG	2164	AAAGAGA	2204	UUCUUU	2244	GGAUAC
2125	ACACAAC	2165	UGGGG	2205	AGGGU	2245	ACCUUC
2126	CCCGGC	2166	ACUUGU	2206	GUGAAUGA	2246	UAGUGA
2127	CCUUU	2167	CCUUCCA	2207	AAAGG	2247	GUUUC
2128	AUAUAUA	2168	GUAGGGU	2208	CAGGAC	2248	GUUUUUG
2129	ACAGG	2169	UUAUUUAUA	2209	UUGGUGU	2249	UCAC
2130	CUUU	2170	GUGGUGG	2210	GGCCAAA	2250	UCUUUC
2131	GUUGU	2171	UUCAC	2211	UCCG	2251	ACUAACA
2132	CCUUCCU	2172	UUACAUG	2212	GGGGG	2252	CCCUUC
2133	UAGAGA	2173	UCAAA	2213	ACUUAU	2253	UCGGGC
2134	AAAGGAA	2174	UAAAAU	2214	UCAG	2254	UAUCCAA
2135	GGGUGU	2175	UUAUUA	2215	UGGAC	2255	GAUCACUG
2136	UGUUC	2176	AAAUUACC	2216	GCAUAC	2256	UUGCCG
2137	CCUGC	2177	CGUCG	2217	UAAAUAA	2257	GAGUAC
2138	GUGCUA	2178	CCUGG	2218	UAAG	2258	UCCCAC
2139	AAAGAAG	2179	CUGCCUCC	2219	UAGAGU	2259	UGCAUA
2140	CUGCUG	2180	AGGAGAA	2220	AUGAAAG	2260	AUCAAC
2141	GAGAGGAAC	2181	CAGACAG	2221	UGAACCA	2261	UGCUGCC
2142	GUGUGU	2182	UUUUAAUUU	2222	CGGAG	2262	CUCA
2143	UGUGAUG	2183	AGAGAGG	2223	AUUAAC	2263	CUUCAAC
2144	GUUUUG	2184	CCUCG	2224	UAAUUA	2264	CUAAG
2145	CCAC	2185	CAGAGGA	2225	CCCUCCA	2265	UGAAGGC
2146	CAGUGG	2186	AAUAUU	2226	GGCCCU	2266	UUUUUUUUU
2147	CAAAAU	2187	GUGUG	2227	UGGGGAU	2267	AUUUA
2148	UGGUGGA	2188	CUACAAU	2228	AGCAUG	2268	UUAUUUU
2149	UACAGA	2189	CUUCAGG	2229	GACAGA	2269	UGGAGGA
2150	CUCUUCAC	2190	UGUUGU	2230	UGCGUA	2270	CCUCCCC
2151	CUUGU	2191	AAAAA	2231	UUUUUG	2271	UAGACA

2272	UAGGAU	2312	ACCCUU	2352	UGAGUGG	2392	GAUGAAU
2273	GGGGGCUG	2313	CUAUAUC	2353	AUCAUC	2393	UCAUGC
2274	UGGAGAC	2314	GUUGGAC	2354	UGCUAC	2394	UGGAGAA
2275	GUUGGAA	2315	ACCUCU	2355	UGGAGU	2395	GCAGCGC
2276	UCCUUC	2316	ACUUUUU	2356	ACAGC	2396	CUUUA
2277	GGUG	2317	AAAAACA	2357	AGAAAAU	2397	CCUCUCG
2278	CCGCCC	2318	UUAUAA	2358	GAAAUCA	2398	UAUUAA
2279	UGGAGCC	2319	GAAGGAC	2359	AGACGGG	2399	CUCUUG
2280	UUAAUA	2320	GGAUGAAC	2360	UGGAGCA	2400	CCUUCU
2281	UAUUUAUU	2321	GAAGGA	2361	ACUAAUC	2401	UCUAUCU
2282	UCUAA	2322	AAAAGA	2362	UCCCCUA	2402	UGGGA
2283	GGGUGGA	2323	GCUGGAA	2363	AUAAAGG	2403	UGGGC
2284	UGGGAAG	2324	CCGUGGA	2364	GGUUGUU	2404	GUUUAGA
2285	GCUGGCA	2325	UUUGGUU	2365	GAAG	2405	AAAAAAC
2286	UUAGGGA	2326	GUGCGCU	2366	CGGAC	2406	AGGGUAU
2287	GGUAAAGGU	2327	CACCGAC	2367	AAUGAAU	2407	UUUUUCC
2288	UUGCUC	2328	UCACAGG	2368	CCGCCCA	2408	UUCUCCAA
2289	GAAUGG	2329	UUUAAU	2369	UGAAGCC	2409	UCCCCA
2290	CUGUAUGU	2330	AGAC	2370	UUAGUGG	2410	UGGAAGAAC
2291	UCUUCC	2331	AGAGACA	2371	UUGUGC	2411	CGAGU
2292	CCGCCUCC	2332	CCUUGU	2372	UGCGGA	2412	AAAUAA
2293	GUGCUU	2333	CAUCUUA	2373	CCAUACC	2413	UGGG
2294	CGGAGAA	2334	UCCCAA	2374	GCAUAA	2414	UUGGG
2295	AAAGGGG	2335	GUAAUUU	2375	CGCCCCA	2415	CCUG
2296	GAAACGA	2336	AGGGAU	2376	UUGCCA	2416	ACAAAUC
2297	UUACUUU	2337	UCCUCUUC	2377	UUUC	2417	GUGGUUU
2298	UCUGA	2338	AUACUUG	2378	CUCUUCUCU	2418	UCUGC
2299	CCACACA	2339	UAGGCU	2379	UCCCAU	2419	GGAGGGC
2300	UGAUAC	2340	CAGUAGA	2380	CUUUUUC	2420	CUACAGC
2301	AUAUAU	2341	UAUAUA	2381	UUACAGG	2421	ACCUGU
2302	GGACAAG	2342	UCUAC	2382	AUAAUU	2422	GAGGAAG
2303	CAUUUUU	2343	GCAUGC	2383	CUUUUU	2423	AGCCUCAG
2304	UGACUG	2344	AAAUUCCA	2384	CAUCUUG	2424	CCAACCG
2305	ACCAAA	2345	UUGCUU	2385	CAUCUG	2425	GGAGGAA
2306	GGCUUAA	2346	UUUCUGG	2386	CCACAGA	2426	AAAAU
2307	CGGCGGA	2347	CCCGCGA	2387	UUGUGUGU	2427	CAAGGGG
2308	GUUGUUC	2348	GCAAACA	2388	AACAUC	2428	UUUUUA
2309	CUACUGG	2349	UUGGUUG	2389	AUCCCC	2429	GGUUGUUG
2310	ACUAAU	2350	GAUCUG	2390	GAUCAAG	2430	CACCGCA
2311	AGCAC	2351	UGUGUGU	2391	AUGCACA	2431	AUGUG

2432	AUAGCA	2472	UCUUGC	2512	UUUCAUUUU	2552	AAAUUCA
2433	CUCG	2473	UAGGUGG	2513	UCUCG	2553	CCAAACA
2434	UAGG	2474	ACACG	2514	UGGAGGC	2554	AUAAAAC
2435	CCUUUUU	2475	GAGGCAU	2515	UUUGCUUU	2555	AGAGAGA
2436	UAGGAUA	2476	AAAACUA	2516	GGUUCUU	2556	GGAGUG
2437	GAAUGU	2477	CUCUUA	2517	UUUUUUC	2557	AGAGGGG
2438	UGGGU	2478	CACAGUG	2518	AAUGAUG	2558	CCCCGCA
2439	GCCCAA	2479	AGGGAA	2519	CAAUUAAA	2559	GACUCCCG
2440	CCACACG	2480	AAAAAU	2520	CCCUCCG	2560	CCCCCUCC
2441	CCCAGCG	2481	UGGUGAC	2521	UGUACAU	2561	AGCAGUGUU
2442	UCCUCU	2482	UUAAUUA	2522	CGAGCGC	2562	AGAGGAA
2443	GACCCCCG	2483	CCUUG	2523	CCCGCCA	2563	UUUCUUU
2444	AUGCUC	2484	UGCUG	2524	UCACUCG	2564	CUUUUCU
2445	GACCACUG	2485	CCGCGCC	2525	CGGCGCA	2565	CGGUGG
2446	AUUUAU	2486	AUAGGGA	2526	CUACAAC	2566	AGGAU
2447	UGACAAA	2487	GGAAGAAC	2527	GCAAUCA	2567	CCCCCCG
2448	GGAGGA	2488	AAUAAU	2528	UUGAUGU	2568	GCAGG
2449	UGCUGAC	2489	UAAAUA	2529	AGUAG	2569	GUGAUGG
2450	CAAAUCA	2490	GCCUGC	2530	ACCCGC	2570	GGAGGAC
2451	CAUU	2491	ACCCAA	2531	ACCAC	2571	CCUUUCC
2452	AUUAAAC	2492	CCCCCCU	2532	ACCCAU	2572	UUAAUGU
2453	UCCCUU	2493	CGUGU	2533	UGGGGAA	2573	CUGUAUAU
2454	AUGACAA	2494	CUGCUU	2534	UUUUUCU	2574	GAAUUAA
2455	UUUAAA	2495	CCAACCC	2535	AGGAGGA	2575	CGCUUCAU
2456	GGCUUU	2496	CGGUCU	2536	GCCCAC	2576	GACUCCUG
2457	ACAAACC	2497	GGAGAG	2537	AGGAGGAU	2577	UAAAUU
2458	AACCUG	2498	UUGCACA	2538	GAGGGAG	2578	CCACAGG
2459	UGUGGUG	2499	GGAGC	2539	CUUUUC	2579	UUAUUUAUU
2460	UGCGUC	2500	UUGUGUAU	2540	AUGCGCU	2580	CCCGGGA
2461	AGAAAGU	2501	UACUAAU	2541	CCUCCCA	2581	AGAUAAA
2462	AGGCA	2502	ACCUGC	2542	GGUGUGU	2582	CAACUCA
2463	AAAGAGG	2503	AGAUACU	2543	ACACGAC	2583	CGGAAGA
2464	AAGACAG	2504	UGCUAAC	2544	UCGGGCA	2584	UAAUAU
2465	AAAAAAAAA	2505	GCCGGAG	2545	UCUUCUU	2585	AUUAUU
2466	CUUAG	2506	AAUGAAA	2546	UUUAAUUUU	2586	AGGAGAG
2467	CCAUUC	2507	AAAUUU	2547	CCCUCCU	2587	AUUAAA
2468	CGGUG	2508	GAUCAAC	2548	GCACUCA	2588	AGUAGG
2469	ACCGC	2509	AGGAAAG	2549	UGGACAG	2589	CGCUGCG
2470	AGGAA	2510	AAUGAAG	2550	ACGCUCA	2590	GACCGGUG
2471	AAGGACC	2511	ACUUCC	2551	UUAAUUU	2591	CUGCUA

2592	GUUUA	2632	CCUUACC	2672	UUAAAA	2712	UACUGC
2593	GCACACA	2633	ACUUUC	2673	GAAAGAA	2713	AUAUAA
2594	AGGUAU	2634	CCUUUCU	2674	AUUUUCU	2714	GGAAGAAG
2595	AAGAGGA	2635	AUGUGU	2675	CUUUUUUC	2715	GGGAGGC
2596	CUAUAUA	2636	GCAGC	2676	AGUGUGU	2716	CAGUGA
2597	GGUGUUU	2637	UAGGAAA	2677	CAAUACU	2717	AGGAUAA
2598	GUGGUGU	2638	CCCAGGA	2678	GUCUUA	2718	CAGUGC
2599	GGGUAC	2639	UAUAAA	2679	UUCUUG	2719	AUGAGCA
2600	AGAGACC	2640	ACGCUC	2680	CCUUGC	2720	AUUA
2601	GAAUACA	2641	UGAGUGA	2681	UUUAG	2721	CAUCCAA
2602	ACUAGGA	2642	UAUAUAC	2682	GUGAUUU	2722	AGACAGG
2603	CUACAGG	2643	ACAUUAGCC	2683	CCAAGGA	2723	UCUCCG
2604	UGUAUAU	2644	AGGACCG	2684	AGGACC	2724	GGUGUGG
2605	UGAAGAA	2645	AAGAAG	2685	GGACAA	2725	GGCUGC
2606	CUCUC	2646	AUAAAAA	2686	UUUGCUUC	2726	CCCUUAAA
2607	AUCCCA	2647	UUUUUGG	2687	UGGAU	2727	GGCCACCA
2608	CACCGGA	2648	GUGUAGU	2688	GACAGGG	2728	CUGCCG
2609	UAAGAC	2649	UGCAGC	2689	GUGCUG	2729	GUGGUUG
2610	UGGAGGG	2650	UUGAUUU	2690	AUGAACA	2730	GGAGAA
2611	GAGGGAA	2651	GCAUUCA	2691	CACAUC	2731	UAAUGC
2612	UGAUGAA	2652	GCGCGGG	2692	AAUGAUU	2732	CGGUGA
2613	AUUUUUU	2653	UCUUUCU	2693	AAUCAAG	2733	AGAAAAC
2614	ACACAAG	2654	UUAGGGU	2694	AUUAAU	2734	AGACAAG
2615	UAGGUUA	2655	UAGUUUA	2695	GGACAGA	2735	GAGGUU
2616	CUCGGAA	2656	CCCACGG	2696	UCUCAGG	2736	CCUACCC
2617	AUUUAC	2657	CCCGGCG	2697	CUUGCUUC	2737	GGAUAU
2618	CACUGCA	2658	UGUUGGU	2698	CCCCUUCC	2738	UAGGCA
2619	ACGCUU	2659	CGGGC	2699	ACAUUCA	2739	AAGU
2620	CAGUGU	2660	AUUUUA	2700	CGCU	2740	UGCUGC
2621	ACAAACA	2661	GCUUGC	2701	GGGAGCA	2741	CUUUCUU
2622	CACACGU	2662	AUAAUA	2702	UAGGAA	2742	CUGUGUGU
2623	UGCAGA	2663	UGAGUGU	2703	UGUGAGG	2743	CAGCCUCC
2624	CCACCGA	2664	UGUGUUG	2704	ACCACC	2744	GGAUGCC
2625	GACUUAA	2665	AAAUUCAC	2705	GUUUUUA	2745	CGCCGGA
2626	ACUGCA	2666	GGGAGGA	2706	CCUUCCUU	2746	UAUAUUUA
2627	CGGCCGA	2667	UAGUUUG	2707	AAAGGAG	2747	UCAAC
2628	AUGUGUAU	2668	UGUAAUAUU	2708	CACAGCA	2748	AAAAUCA
2629	UUAGUUU	2669	AAGACGA	2709	ACAUUAAA	2749	UCACC
2630	CCUCAGG	2670	GAUUCCCU	2710	AGAGAAC	2750	ACCUUAAA
2631	CUAUAGC	2671	AUGUGUGU	2711	GUGACGA	2751	UUGCCU

2752	AGAGAGC	2792	CGCAGGC	2832	AACACGA	2872	ACCCCAA
2753	UAUAAU	2793	AGACAAA	2833	AACAUCA	2873	ACCCCCAC
2754	GCAUACA	2794	CACGUA	2834	AACAUCG	2874	ACCCGA
2755	GAGCGAG	2795	GUGCUC	2835	AACAUCU	2875	ACCCGU
2756	CAUAUA	2796	AUUAGGCA	2836	AACCUUGCC	2876	ACCGGG
2757	GGGGU	2797	GAUC	2837	AACUAAG	2877	ACGAACA
2758	UUUUUUG	2798	CUUGCUUU	2838	AACUACG	2878	ACGAACC
2759	UUGUA	2799	CAGACAA	2839	AACUUUAU	2879	ACGAAUA
2760	GGAGAU	2800	CAGUCAC	2840	AAGAAC	2880	ACGAAUC
2761	UCCGC	2801	ACCGUC	2841	AAGAAGAA	2881	ACGACA
2762	GGAGGAG	2802	GGGAGGG	2842	AAGAAGAAG	2882	ACGACG
2763	ACAUACA	2803	GAAUUA	2843	AAGACUA	2883	ACGAGAU
2764	UUCUUUU	2804	AGGAGAGUA	2844	AAGGCGAA	2884	ACGCGCA
2765	CCCCCCC	2805	CCGCCCG	2845	AAGGGGAA	2885	ACGCGCC
2766	AAAACGA	2806	AUUUUUUUC	2846	AAGGUG	2886	ACGCGCG
2767	CCCAUGA	2807	CAGCCGU	2847	AAGUAACAA	2887	ACGCGCU
2768	GAUCAUG	2808	GCGCUC	2848	AAUAAGAAG	2888	ACGCUA
2769	AUGCUU	2809	UGAAGGAC	2849	AAUACAG	2889	ACUAAC
2770	UCUUCU	2810	UAGUAU	2850	AAUACGA	2890	ACUAACC
2771	AGCUCCCA	2811	GAUCGCUG	2851	AAUACUA	2891	ACUAACG
2772	UUGCCC	2812	UGGGAAA	2852	AAUCAAA	2892	ACUAACU
2773	UCUGU	2813	AGGAGCG	2853	AAUCAUA	2893	ACUAAGUA
2774	GACUCAAG	2814	UCACUGG	2854	AAUCAUU	2894	ACUAAUA
2775	UUACGAC	2815	CCAAUCC	2855	AAUCCAA	2895	ACUAGAG
2776	UUAAUGG	2816	UUGAAGCAA	2856	AAUCCAG	2896	ACUAGCG
2777	GUUUUUC	2817	CGGCCGU	2857	AAUCUUG	2897	ACUAGCU
2778	AAAAUU	2818	AAAAAAA	2858	AAUUAU	2898	ACUUUAA
2779	CCCGGCA	2819	AAAAAAG	2859	AAUUUU	2899	AGAAAUU
2780	CCAACCA	2820	AAAAUA	2860	ACAAAUA	2900	AGAAGA
2781	GUGUAGG	2821	AAACCAAA	2861	ACAAGAC	2901	AGAAGAAC
2782	UUGUUUU	2822	AAAGAAGAA	2862	ACAAGGAC	2902	AGAAGAGC
2783	AGGAGAAUA	2823	AAAGACAAA	2863	ACACAAA	2903	AGAAGGAC
2784	UGGUGCU	2824	AAAGC	2864	ACACACA	2904	AGACAA
2785	CGCGCCC	2825	AAAGGGA	2865	ACACCAGAC	2905	AGACAGA
2786	AGGAACA	2826	AAAUACA	2866	ACACGAA	2906	AGACGAA
2787	CCCCGGA	2827	AAAUACU	2867	ACACGAG	2907	AGACGAG
2788	ACGCUG	2828	AAAUUCCC	2868	ACACUCA	2908	AGACGAU
2789	UGACGAG	2829	AACAACA	2869	ACAGUCA	2909	AGACGGA
2790	CACG	2830	AACAACG	2870	ACAUAAA	2910	AGACGUA
2791	CGAGGCA	2831	AACAACU	2871	ACAUUCAU	2911	AGAGACG

2912	AGAGCAGG	2952	AGUGUGA	2992	AUUUAUAUA	3032	CAGCCGA
2913	AGAGGUA	2953	AGUGUGG	2993	AUUUAUUU	3033	CAGCGGA
2914	AGAGUGA	2954	AUAAAAG	2994	AUUUAUUUA	3034	CAGCGGG
2915	AGAGUGG	2955	AUAAUUA	2995	AUUUAUUUU	3035	CAGCGGU
2916	AGAGUGU	2956	AUAAUUG	2996	AUUUUC	3036	CAGCUUCC
2917	AGAUAU	2957	AUACAUG	2997	AUUUUUA	3037	CAGGGA
2918	AGAUCC	2958	AUACUUU	2998	AUUUUUC	3038	CAGUCGG
2919	AGAUGGAC	2959	AUAGGGU	2999	AUUUUUUUA	3039	CAUACAG
2920	AGCAGAGAU	2960	AUAUAGA	3000	CAAAACA	3040	CAUCAAC
2921	AGCAGAGGA	2961	AUAUAGC	3001	CAACACA	3041	CAUCAU
2922	AGCAGAGUA	2962	AUAUAUAU	3002	CAACAGC	3042	CAUCCAG
2923	AGCAGAGUU	2963	AUAUAUC	3003	CAACCACAA	3043	CAUCGC
2924	AGCAUGA	2964	AUAUUA	3004	CAACCGG	3044	CCAAACC
2925	AGCAUGC	2965	AUAUUU	3005	CAAUACA	3045	CCAACGA
2926	AGCCCCCA	2966	AUCAC	3006	CACAA	3046	CCAACGC
2927	AGCGAGC	2967	AUCACC	3007	CACACCACA	3047	CCAAGCA
2928	AGCGGA	2968	AUCAUCAU	3008	CACACGA	3048	CCAAUCA
2929	AGCUGUU	2969	AUCCCGUG	3009	CACAGA	3049	CCAAUGG
2930	AGGACA	2970	AUCUGAC	3010	CACAGAC	3050	CCACGCA
2931	AGGACAA	2971	AUCUUCAC	3011	CACAGCG	3051	CCACGCU
2932	AGGACAG	2972	AUCUUCAU	3012	CACAGGA	3052	CCACGGA
2933	AGGAGAAC	2973	AUGAAAA	3013	CACAGGU	3053	CCACUCA
2934	AGGAGAAUG	2974	AUGAACG	3014	CACAUA	3054	CCACUCG
2935	AGGAGAGGU	2975	AUGACAG	3015	CACAUCAU	3055	CCAGCGA
2936	AGGAGAUAA	2976	AUGACCA	3016	CACCCGA	3056	CCAUACA
2937	AGGAGCA	2977	AUGACCG	3017	CACCCGU	3057	CCAUCCC
2938	AGGAGUA	2978	AUGACG	3018	CACCCUCC	3058	CCAUUCC
2939	AGGAGUGAC	2979	AUGAGAA	3019	CACCGGU	3059	CCCACCA
2940	AGGAUAG	2980	AUGAGAG	3020	CACCUCCG	3060	CCCACGA
2941	AGGAUGA	2981	AUGAGCG	3021	CACCUUCC	3061	CCCAGCA
2942	AGGAUUA	2982	AUGCACG	3022	CACGGA	3062	CCCAGGG
2943	AGGAUUG	2983	AUGCCA	3023	CACGGAC	3063	CCCCCCA
2944	AGGCAGU	2984	AUGCGCC	3024	CACGGC	3064	CCCCCGA
2945	AGGGA	2985	AUGCGCG	3025	CACGUC	3065	CCCCCGG
2946	AGGUAA	2986	AUGUAUAU	3026	CACUACAC	3066	CCCCGAC
2947	AGGUAC	2987	AUGUAUGU	3027	CACUGAC	3067	CCCCGCG
2948	AGGUUA	2988	AUUCAAG	3028	CAGACGA	3068	CCCCGGG
2949	AGUACAG	2989	AUUGCACC	3029	CAGACGU	3069	CCCGCCG
2950	AGUAGCC	2990	AUUUACUUU	3030	CAGAGAGAC	3070	CCCGCGG
2951	AGUCGAU	2991	AUUUAUACU	3031	CAGAGGU	3071	CCCGCUGC

3072	CCCGCUGU	3112	CCGUCCG	3152	CGCCGCA	3192	CGGCCUCC
3073	CCCGCUUC	3113	CCGUCGG	3153	CGCCGCG	3193	CGGCGCC
3074	CCCGCUUU	3114	CCGUGCA	3154	CGCCGGG	3194	CGGCGGG
3075	CCCGGGG	3115	CCGUGCG	3155	CGCCGGU	3195	CGGCGGU
3076	CCCGUUU	3116	CCUAC	3156	CGCCUGUAG	3196	CGGCUUCC
3077	CCCUAA	3117	CCUAGGG	3157	CGCCUUCC	3197	CGGGA
3078	CCCUCC	3118	CCUAU	3158	CGCGCCA	3198	CGGGCA
3079	CCCUCGA	3119	CCUAUCC	3159	CGCGCCG	3199	CGGGCCA
3080	CCCUCGG	3120	CCUCA	3160	CGCGCGG	3200	CGGGCCG
3081	CCCUGCA	3121	CCUCACG	3161	CGCGGA	3201	CGGGCGA
3082	CCCUGCG	3122	CCUCGUCC	3162	CGCGGAG	3202	CGGGCGG
3083	CCCUGGA	3123	CCUCUGG	3163	CGCGGC	3203	CGGGGCA
3084	CCCUGGG	3124	CCUCUUCC	3164	CGCGGCA	3204	CGGGGCG
3085	CCGACCA	3125	CCUCUUU	3165	CGCGGCG	3205	CGGGGGA
3086	CCGACGA	3126	CCUGCGG	3166	CGCGGGA	3206	CGGGGGG
3087	CCGAGAUCA	3127	CCUGCUGU	3167	CGCGGGG	3207	CGGUAA
3088	CCGAGCA	3128	CCUGCUUC	3168	CGCGUA	3208	CGGUAG
3089	CCGAGCG	3129	CCUGCUUU	3169	CGCGUC	3209	CGGUAU
3090	CCGAGGA	3130	CCUUC	3170	CGCUCCG	3210	CGGUCCA
3091	CCGAUGA	3131	CCUUCC	3171	CGCUCGG	3211	CGGUCGG
3092	CCGCAGG	3132	CCUUCCC	3172	CGCUGAC	3212	CGGUGCA
3093	CCGCCGA	3133	CCUUCUGUG	3173	CGCUGCA	3213	CGGUGCG
3094	CCGCCGG	3134	CGAACCA	3174	CGCUGGA	3214	CGGUGGA
3095	CCGCGCA	3135	CGAAGGA	3175	CGCUGGG	3215	CGGUGGG
3096	CCGCGCG	3136	CGACCCA	3176	CGCUUU	3216	CGUACAG
3097	CCGCGGA	3137	CGAUACU	3177	CGGACCA	3217	CGUUCUU
3098	CCGCGGG	3138	CGAUGGA	3178	CGGACCG	3218	CGUUGGU
3099	CCGCUA	3139	CGCACCA	3179	CGGACGG	3219	CGUUGUU
3100	CCGCUAU	3140	CGCACCG	3180	CGGACGU	3220	CUAAAU
3101	CCGCUG	3141	CGCAGA	3181	CGGAGCA	3221	CUACAAA
3102	CCGCUU	3142	CGCAGCG	3182	CGGAGCG	3222	CUACACG
3103	CCGGACG	3143	CGCAGGA	3183	CGGAGGA	3223	CUACGAA
3104	CCGGCCA	3144	CGCAGGG	3184	CGGAGGG	3224	CUAUAGA
3105	CCGGCCG	3145	CGCAGGU	3185	CGGAGGU	3225	CUCCAUCUU
3106	CCGGCGA	3146	CGCAUA	3186	CGGAGUA	3226	CUCCCUCC
3107	CCGGCGG	3147	CGCAUC	3187	CGGCACA	3227	CUCCGGA
3108	CCGGGCA	3148	CGCC	3188	CGGCAGA	3228	CUCCUUCC
3109	CCGGGCG	3149	CGCCCGA	3189	CGGCCCA	3229	CUCGCUGC
3110	CCGGGGA	3150	CGCCCGU	3190	CGGCCCG	3230	CUCGCUGU
3111	CCGUCCA	3151	CGCCCUUG	3191	CGGCCGG	3231	CUCGCUUC

3232	CUCGCUUU	3272	GACCACCA	3312	GAUCCCCA	3352	GCGGGUC
3233	CUCGGAC	3273	GACCACUA	3313	GAUCCCUA	3353	GCGUAC
3234	CUCGUGU	3274	GACCCCCA	3314	GAUCCCUG	3354	GCGUGC
3235	CUGAACA	3275	GACCCCUA	3315	GAUGAAA	3355	GCUAGUA
3236	CUGACUA	3276	GACCCCUG	3316	GAUGAAGAG	3356	GCUGGUA
3237	CUGAGAGAC	3277	GACCCGG	3317	GAUGAUA	3357	GGAAAGAC
3238	CUGAGUU	3278	GACCUGCG	3318	GAUGAUG	3358	GGAACAGC
3239	CUGCCA	3279	GACGUC	3319	GAUGAUU	3359	GGAAGCC
3240	CUGCCC	3280	GACUAAC	3320	GAUUACCA	3360	GGAAGGA
3241	CUGCGGU	3281	GACUAC	3321	GAUUACUA	3361	GGAAGGC
3242	CUGCUC	3282	GACUACCA	3322	GAUUACUG	3362	GGAAUGAC
3243	CUGCUUCC	3283	GACUACUA	3323	GAUUAGUG	3363	GGACAGG
3244	CUGGAUU	3284	GACUACUG	3324	GAUUCCCG	3364	GGACGAA
3245	CUUCACG	3285	GACUAGAA	3325	GAUUCCUA	3365	GGACGAC
3246	CUUCUCG	3286	GACUAGCA	3326	GAUUCCUG	3366	GGACGAG
3247	CUUCUCUCU	3287	GACUCCCA	3327	GAUUGAUG	3367	GGACGCUG
3248	CUUCUGG	3288	GACUCCUA	3328	GAUUUAA	3368	GGACGGA
3249	CUUGCUGC	3289	GACUGC	3329	GCACUAGAU	3369	GGACGGC
3250	CUUGCUGU	3290	GACUGGA	3330	GCAGCGA	3370	GGACGGG
3251	CUUGGAA	3291	GACUUCAU	3331	GCCCACU	3371	GGACUGGGA
3252	CUUUAAUUU	3292	GACUUCUU	3332	GCCCCUA	3372	GGACUGUA
3253	CUUUUA	3293	GAGACGA	3333	GCCGGA	3373	GGAGCGA
3254	CUUUUUG	3294	GAGACUA	3334	GCCGUC	3374	GGAGCGC
3255	GAAAAC	3295	GAGCGGA	3335	GCCUAC	3375	GGAGGAAC
3256	GAAAACC	3296	GAGGAAC	3336	GCGAGCA	3376	GGAGGGG
3257	GAAAGAAG	3297	GAGGAAGAA	3337	GCGAGCG	3377	GGAGUA
3258	GAAAGAAU	3298	GAGGAGGA	3338	GCGCG	3378	GGAGUAGCC
3259	GAAAGGAGA	3299	GAGUAGA	3339	GCGCGC	3379	GGAGUAGCU
3260	GAACACA	3300	GAGUAGG	3340	GCGCGCA	3380	GGAGUU
3261	GAACUCA	3301	GAGUCAU	3341	GCGCGCC	3381	GGAUGC
3262	GAAGAA	3302	GAGUGC	3342	GCGCGCG	3382	GGAUGCA
3263	GAAGAAGA	3303	GAUAAAA	3343	GCGCGGC	3383	GGAUGGA
3264	GAAGAGGAA	3304	GAUAAAC	3344	GCGCGGU	3384	GGAUGGAG
3265	GAAGGAA	3305	GAUACUA	3345	GCGCUG	3385	GGAUUAA
3266	GAAGGAG	3306	GAUAUAC	3346	GCGCUU	3386	GGAUUC
3267	GAAGGGG	3307	GAUAUAG	3347	GCGGGAA	3387	GGCACGG
3268	GAAGGUA	3308	GAUCAAA	3348	GCGGGAC	3388	GGCCACCG
3269	GAAGGUG	3309	GAUCAAU	3349	GCGGGCA	3389	GGCCACUG
3270	GAAUGA	3310	GAUCACC	3350	GCGGGCC	3390	GGCCCACA
3271	GAAUUCA	3311	GAUCACCA	3351	GCGGGUA	3391	GGCCCCCG

3392	GGCCCCUA	3432	GGUUAAUG	3472	GUUGUU	3512	UAUAUAG
3393	GGCCGCAG	3433	GGUUACCA	3473	GUUUUC	3513	UAUCAAC
3394	GGCG	3434	GGUUACUA	3474	GUUUUUUUC	3514	UAUUAUUUU
3395	GGCUAC	3435	GGUUACUG	3475	UAAUAC	3515	UCACAAC
3396	GGCUACCA	3436	GGUUCCCA	3476	UAAUUU	3516	UCACACG
3397	GGCUACCG	3437	GGUUCCCG	3477	UACAUCAC	3517	UCAGAGAUU
3398	GGCUACUA	3438	GGUUCCUA	3478	UACGAAC	3518	UCAGUCAC
3399	GGCUACUG	3439	GGUUCCUG	3479	UACGGA	3519	UCCAGCAA
3400	GGCUCCAA	3440	GGUUCGU	3480	UACGGC	3520	UCCCCAA
3401	GGCUCCCA	3441	GGUUUAA	3481	UACGUA	3521	UCCCCAU
3402	GGCUCCCCC	3442	GUAACG	3482	UACGUC	3522	UCCCCCAU
3403	GGCUCCCG	3443	GUAAUGG	3483	UACUAAC	3523	UCCCCCCAA
3404	GGCUCCUA	3444	GUAAUGU	3484	UACUAAG	3524	UCCCGA
3405	GGCUCCUG	3445	GUAAUUG	3485	UACUAC	3525	UCCGCUUU
3406	GGCUCGUG	3446	GUAGGGA	3486	UACUAGC	3526	UCCGGA
3407	GGCUGAUG	3447	GUAGUGG	3487	UACUUCAA	3527	UCCUAC
3408	GGCUGCAA	3448	GUAGUGU	3488	UAGACAG	3528	UCCUGC
3409	GGGACGU	3449	GUAGUUG	3489	UAGAGAGAC	3529	UCCUUCUUG
3410	GGGCGCC	3450	GUAGUUU	3490	UAGAGAGG	3530	UCGAGAU
3411	GGGCGCG	3451	GUCCCCUG	3491	UAGCUGUG	3531	UCGCACA
3412	GGGG	3452	GUCCUCAG	3492	UAGGAAG	3532	UCGCUA
3413	GGGGGAC	3453	GUCCUCCG	3493	UAGGAGG	3533	UCGCUC
3414	GGGGGAG	3454	GUCGGAA	3494	UAGGAUG	3534	UCGCUU
3415	GGGGGGG	3455	GUCUGU	3495	UAGGGA	3535	UCGG
3416	GGGUGC	3456	GUCUUU	3496	UAGGGC	3536	UCGGG
3417	GGGUGCA	3457	GUGAUGU	3497	UAGGGU	3537	UCGUAC
3418	GGGUGCC	3458	GUGAUUG	3498	UAGGUAA	3538	UCGUGC
3419	GGGUGGC	3459	GUGCGCC	3499	UAGGUAG	3539	UCUAAGGAA
3420	GGUCACCA	3460	GUGCGCG	3500	UAGGUU	3540	UCUCACG
3421	GGUCACCG	3461	GUGUAGA	3501	UAGGUUG	3541	UCUCUC
3422	GGUCACUA	3462	GUGUAUAU	3502	UAGUAAC	3542	UCUCUCG
3423	GGUCACUG	3463	GUGUAUGU	3503	UAGUAAG	3543	UCUCUGG
3424	GGUCAGUG	3464	GUGUGUAU	3504	UAGUAC	3544	UCUGCUGC
3425	GGUCCCCA	3465	GUGUGUGU	3505	UAGUAGG	3545	UCUGCUGU
3426	GGUCCCCG	3466	GUUCAAG	3506	UAGUAUG	3546	UCUUAUAUU
3427	GGUCCCUA	3467	GUUCCAGAU	3507	UAGUGAA	3547	UCUUCAC
3428	GGUCCCUG	3468	GUUCCAGGU	3508	UAGUGC	3548	UCUUCUCU
3429	GGUCUGC	3469	GUUCCAGUU	3509	UAGUUAA	3549	UCUUUUU
3430	GGUGGUGGU	3470	GUUCUGUA	3510	UAGUUAG	3550	UGAAGAAC
3431	GGUGUUG	3471	GUUGUGC	3511	UAUACAG	3551	UGAAGAC

3552	UGAAGAU	3582	UGGAGAU	3612	UGUGUUU	3642	UUGGUGG
3553	UGAAGCA	3583	UGGAGCU	3613	UGUUACUA	3643	UUGGUUU
3554	UGAAGCU	3584	UGGAGGAC	3614	UGUUCGU	3644	UUGUAUAU
3555	UGAAGGU	3585	UGGAGGU	3615	UGUUGUU	3645	UUGUAUUGU
3556	UGACAGG	3586	UGGAGUA	3616	UGUUUGU	3646	UUUAAUGUU
3557	UGACGAA	3587	UGGGAAU	3617	UUAAUUG	3647	UUUAAUUUG
3558	UGACGGA	3588	UGGGGAG	3618	UUACACG	3648	UUUAGUU
3559	UGACGGG	3589	UGGUAA	3619	UUACAUU	3649	UUUAUU
3560	UGAUGAAC	3590	UGGUAC	3620	UUACUCG	3650	UUUAUUU
3561	UGAUGAC	3591	UGGUGAA	3621	UUACUGG	3651	UUUAUUUA
3562	UGAUGAU	3592	UGGUGAU	3622	UUACUUG	3652	UUUAUUUU
3563	UGAUGC	3593	UGGUGG	3623	UUAGGG	3653	UUUCACG
3564	UGAUGCA	3594	UGGUGGC	3624	UUAGUGU	3654	UUUCAGA
3565	UGAUGCC	3595	UGGUGGU	3625	UUAGUUG	3655	UUUCAGG
3566	UGAUGGA	3596	UGGUGU	3626	UUAUACUUU	3656	UUUCUCG
3567	UGAUGGAGA	3597	UGUAAAUA	3627	UUAUAUUUU	3657	UUUGAUUUU
3568	UGAUGGU	3598	UGUACAUA	3628	UUAUUU	3658	UUUGCUGC
3569	UGCAGAGUG	3599	UGUACAUC	3629	UUAUUUAAA	3659	UUUGCUGU
3570	UGCAGGCA	3600	UGUAGAUA	3630	UUAUUUAAU	3660	UUUGUUU
3571	UGCAUC	3601	UGUAGGAC	3631	UUCAUCAC	3661	UUUUAAGUU
3572	UGCAUG	3602	UGUAUAUA	3632	UUCGCUGU	3662	UUUUAGUUA
3573	UGCAUGA	3603	UGUCAAC	3633	UUCGCUUC	3663	UUUUAUGUU
3574	UGCAUGC	3604	UGUCGC	3634	UUCGCUUU	3664	UUUUAUUGA
3575	UGCAUGU	3605	UGUGGA	3635	UUCUUC	3665	UUUUAUUGU
3576	UGCCGCC	3606	UGUGGG	3636	UUCUUCAA	3666	UUUUAUUUA
3577	UGCCGUU	3607	UGUGGGG	3637	UUCUUCAU	3667	UUUUAUUUG
3578	UGCGGUC	3608	UGUGUGA	3638	UUGAUUG	3668	UUUUAUUUU
3579	UGCUGUUUU	3609	UGUGUGUG	3639	UUGCACG	3669	UUUUUCG
3580	UGGAAAG	3610	UGUGUGUGU	3640	UUGCUA	3670	UUUUUGU
3581	UGGACAA	3611	UGUGUGUU	3641	UUGCUG		

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ST26SequenceListing PUBLIC "-//WIPO//DTD Sequence Listing 1.3//EN"
"ST26SequenceListing_V1_3.dtd">
<ST26SequenceListing dtdVersion="V1_3" fileName="P095624NL SEQUENCE LISTING
ST26.xml" softwareName="WIPO Sequence" softwareVersion="2.2.0"
productionDate="2022-12-08">
    <ApplicantFileReference>P095624NL </ApplicantFileReference>
    <ApplicantName languageCode="en">Stichting Sanquin
    Bloedvoorziening</ApplicantName>
    <InventionTitle languageCode="en">Modified RNA for modulated protein
    expression</InventionTitle>
    <SequenceTotalQuantity>8</SequenceTotalQuantity>
    <SequenceData sequenceIDNumber="1">
        <INSDSeq>
            <INSDSeq_length>180</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
                <INSDFeature>
                    <INSDFeature_key>source</INSDFeature_key>
                    <INSDFeature_location>1..180</INSDFeature_location>
                    <INSDFeature_quals>
                        <INSDQualifier>
                            <INSDQualifier_name>mol_type</INSDQualifier_name>
                            <INSDQualifier_value>other RNA</INSDQualifier_value>
                        </INSDQualifier>
                        <INSDQualifier id="q2">
                            <INSDQualifier_name>organism</INSDQualifier_name>
                            <INSDQualifier_value>synthetic
                            construct</INSDQualifier_value>
                        </INSDQualifier>
                    </INSDFeature_quals>
                </INSDFeature>
            </INSDSeq_feature-table>

            <INSDSeq_sequence>ctattccggtatatactatatccgatcatatactaaactgtacgcctctgcaacat
            atcgcctagtagctatagactgagcgatcaagtatcgagtatcaagtggtacagtcgcgagtgcgtatagtgcc
            agctagagtactctgatctgaccgccaatcgtgaaagttacgatctcatc</INSDSeq_sequence>
        </INSDSeq>
    </SequenceData>
    <SequenceData sequenceIDNumber="2">
        <INSDSeq>
            <INSDSeq_length>180</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
                <INSDFeature>
                    <INSDFeature_key>source</INSDFeature_key>
                    <INSDFeature_location>1..180</INSDFeature_location>
                    <INSDFeature_quals>
                        <INSDQualifier>
                            <INSDQualifier_name>mol_type</INSDQualifier_name>
                            <INSDQualifier_value>other RNA</INSDQualifier_value>
                        </INSDQualifier>
                        <INSDQualifier id="q4">
                            <INSDQualifier_name>organism</INSDQualifier_name>
                            <INSDQualifier_value>synthetic
                            construct</INSDQualifier_value>
                        </INSDQualifier>
                    </INSDFeature_quals>
                </INSDFeature>
            </INSDSeq_feature-table>

            <INSDSeq_sequence>ctattccggctcaggtctatatccgatcactcaggtaactgtacgcctctcaggta
            tatcgcctagtagctatagactctcaggtcaagtatcgagtatcaagtggtacactcaggtgtgcgtatagtgc
            cagctagagtactctgatctctcaggaatcgtgaaagttacgatctcatc</INSDSeq_sequence>
        </INSDSeq>
    </SequenceData>
    <SequenceData sequenceIDNumber="3">
        <INSDSeq>
            <INSDSeq_length>180</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
```

```
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
                <INSDFeature>
                    <INSDFeature_key>source</INSDFeature_key>
                    <INSDFeature_location>1..180</INSDFeature_location>
                    <INSDFeature_quals>
                        <INSDQualifier>
                            <INSDQualifier_name>mol_type</INSDQualifier_name>
                            <INSDQualifier_value>other RNA</INSDQualifier_value>
                        </INSDQualifier>
                        <INSDQualifier id="q8">
                            <INSDQualifier_name>organism</INSDQualifier_name>
                            <INSDQualifier_value>synthetic
                            construct</INSDQualifier_value>
                        </INSDQualifier>
                    </INSDFeature_quals>
                </INSDFeature>
            </INSDSeq_feature-table>

            <INSDSeq_sequence>ctattccggctttctttatatccgatcactttcttaactgtacgcctctttcttat
            atcgcctagtagctatagactctttcttcaagtatcgagtatcaagtggtacactttcttgtgcgtatagtgcc
            agctagagtactctgatctctttcttaatcgtgaaagttacgatctcatc</INSDSeq_sequence>
        </INSDSeq>
    </SequenceData>
    <SequenceData sequenceIDNumber="4">
        <INSDSeq>
            <INSDSeq_length>180</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
                <INSDFeature>
                    <INSDFeature_key>source</INSDFeature_key>
                    <INSDFeature_location>1..180</INSDFeature_location>
                    <INSDFeature_quals>
                        <INSDQualifier>
                            <INSDQualifier_name>mol_type</INSDQualifier_name>
                            <INSDQualifier_value>other RNA</INSDQualifier_value>
                        </INSDQualifier>
                        <INSDQualifier id="q9">
                            <INSDQualifier_name>organism</INSDQualifier_name>
                            <INSDQualifier_value>synthetic
                            construct</INSDQualifier_value>
                        </INSDQualifier>
                    </INSDFeature_quals>
                </INSDFeature>
            </INSDSeq_feature-table>

            <INSDSeq_sequence>ctattccggtagaagatatatccgatcatagaagaaactgtacgcctagaagacat
            atcgcctagtagctatagactgagaagacaagtatcgagtatcaagtggtacagagaagagtgcgtatagtgcc
            agctagagtactctgatctagaagacaatcgtgaaagttacgatctcatc</INSDSeq_sequence>
        </INSDSeq>
    </SequenceData>
    <SequenceData sequenceIDNumber="5">
        <INSDSeq>
            <INSDSeq_length>180</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
                <INSDFeature>
                    <INSDFeature_key>source</INSDFeature_key>
                    <INSDFeature_location>1..180</INSDFeature_location>
                    <INSDFeature_quals>
                        <INSDQualifier>
                            <INSDQualifier_name>mol_type</INSDQualifier_name>
                            <INSDQualifier_value>other RNA</INSDQualifier_value>
                        </INSDQualifier>
                        <INSDQualifier id="q10">
                            <INSDQualifier_name>organism</INSDQualifier_name>
                            <INSDQualifier_value>synthetic
                            construct</INSDQualifier_value>
                        </INSDQualifier>
                    </INSDFeature_quals>
```

```
                </INSDFeature>
            </INSDSeq_feature-table>

            <INSDSeq_sequence>ctattccggttatttatatatccgatcattatttaaactgtacgccttatttacat
            atcgcctagtagctatagactgtatttacaagtatcgagtatcaagtggtacagtatttagtgcgtatagtgcc
            agctagagtactctgatcttatttacaatcgtgaaagttacgatctcatc</INSDSeq_sequence>
        </INSDSeq>
    </SequenceData>
    <SequenceData sequenceIDNumber="6">
        <INSDSeq>
            <INSDSeq_length>187</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
                <INSDFeature>
                    <INSDFeature_key>source</INSDFeature_key>
                    <INSDFeature_location>1..187</INSDFeature_location>
                    <INSDFeature_quals>
                        <INSDQualifier>
                            <INSDQualifier_name>mol_type</INSDQualifier_name>
                            <INSDQualifier_value>other RNA</INSDQualifier_value>
                        </INSDQualifier>
                        <INSDQualifier id="q12">
                            <INSDQualifier_name>organism</INSDQualifier_name>
                            <INSDQualifier_value>synthetic
                            construct</INSDQualifier_value>
                        </INSDQualifier>
                    </INSDFeature_quals>
                </INSDFeature>
            </INSDSeq_feature-table>

            <INSDSeq_sequence>gatcactattccggctttctttatatccgatcatatactaaactgtacgcctctgc
            aacatatcgcctagtagctatagactgagcgatcaagtatcgagtatcaagtggtacagtcgcgagtgcgtata
            gtgccagctagagtactctgatctgaccgccaatcgtgaaagttacgatctcatcgc</INSDSeq_sequenc
            e>
        </INSDSeq>
    </SequenceData>
    <SequenceData sequenceIDNumber="7">
        <INSDSeq>
            <INSDSeq_length>187</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
                <INSDFeature>
                    <INSDFeature_key>source</INSDFeature_key>
                    <INSDFeature_location>1..187</INSDFeature_location>
                    <INSDFeature_quals>
                        <INSDQualifier>
                            <INSDQualifier_name>mol_type</INSDQualifier_name>
                            <INSDQualifier_value>other RNA</INSDQualifier_value>
                        </INSDQualifier>
                        <INSDQualifier id="q14">
                            <INSDQualifier_name>organism</INSDQualifier_name>
                            <INSDQualifier_value>synthetic
                            construct</INSDQualifier_value>
                        </INSDQualifier>
                    </INSDFeature_quals>
                </INSDFeature>
            </INSDSeq_feature-table>

            <INSDSeq_sequence>gatcactattccggctttctttatatccgatcactttcttaactgtacgcctctgc
            aacatatcgcctagtagctatagactgagcgatcaagtatcgagtatcaagtggtacagtcgcgagtgcgtata
            gtgccagctagagtactctgatctgaccgccaatcgtgaaagttacgatctcatcgc</INSDSeq_sequenc
            e>
        </INSDSeq>
    </SequenceData>
    <SequenceData sequenceIDNumber="8">
        <INSDSeq>
            <INSDSeq_length>187</INSDSeq_length>
            <INSDSeq_moltype>RNA</INSDSeq_moltype>
            <INSDSeq_division>PAT</INSDSeq_division>
            <INSDSeq_feature-table>
```

```
                    <INSDFeature>
                        <INSDFeature_key>source</INSDFeature_key>
                        <INSDFeature_location>1..187</INSDFeature_location>
                        <INSDFeature_quals>
                            <INSDQualifier>
                                <INSDQualifier_name>mol_type</INSDQualifier_name>
                                <INSDQualifier_value>other RNA</INSDQualifier_value>
                            </INSDQualifier>
                            <INSDQualifier id="q16">
                                <INSDQualifier_name>organism</INSDQualifier_name>
                                <INSDQualifier_value>synthetic
                                construct</INSDQualifier_value>
                            </INSDQualifier>
                        </INSDFeature_quals>
                    </INSDFeature>
                </INSDSeq_feature-table>

                <INSDSeq_sequence>gatcactattccggctttctttatatccgatcactttcttaactgtacgcctctttcttatatcgcctagtagctatagactctttcttcaagtatcgagtatcaagtggtacagtcgcgagtgcgtatagtgccagctagagtactctgatctgaccgccaatcgtgaaagttacgatctcatcgc</INSDSeq_sequence>
            </INSDSeq>
        </SequenceData>
    </ST26SequenceListing>
```

CONCLUSIES

1. Een RNA-molecuul met een gemodificeerde nucleotidesequentie, waarbij de gemodificeerde nucleotidesequentie één of meer sequentie-elementen omvat gekozen uit Tabel 1 en/of Tabel 2 in de UTR van het RNA-molecuul, en waarbij het RNA-molecuul codeert voor een eiwit.

2. RNA-molecuul met een gemodificeerde nucleotidesequentie volgens conclusie 1, waarbij de één of meer sequentie-elementen gekozen uit Tabel 1 in de UTR van het RNA-molecuul resulteren in een verhoogde of verlaagde expressie van het eiwit vertaald van het RNA-molecuul met een gemodificeerd nucleotide sequentie wanneer het in een cel wordt geïntroduceerd, in vergelijking met hetzelfde eiwit vertaald uit een RNA-molecuul dat geen gemodificeerde nucleotidesequentie heeft.

3. RNA-molecuul met een gemodificeerde nucleotidesequentie volgens conclusie 1 of 2, waarbij de UTR 2, 3, 4 5, 6, 7, 8, 9 of 10 sequentie-elementen gekozen uit Tabel 1 omvat, en/of waarbij de UTR omvat 2, 3, 4 5, 6, 7, 8, 9 of 10 exemplaren van hetzelfde sequentie-element gekozen uit tabel 1.

4. RNA-molecuul met een gemodificeerde nucleotidesequentie volgens een van de voorgaande conclusies, waarbij de UTR 2 of meer sequentie-elementen gekozen uit tabel 1 omvat, waarbij elk sequentie-element gescheiden is door een spacer-sequentie.

5. RNA-molecuul met een gemodificeerde nucleotidesequentie volgens een van de voorgaande conclusies, waarbij het door het RNA-molecuul gecodeerde eiwit een recombinant eiwit is.

6. RNA-molecuul met een gemodificeerde nucleotidesequentie volgens een van de voorgaande conclusies, waarbij de gemodificeerde nucleotidesequentie een of meer sequentie-elementen gekozen uit Tabel 1 en/of Tabel 2 in de 3' UTR en/of de 5' UTR van het RNA-molecuul omvat.

7. DNA-molecuul dat codeert voor het RNA-molecuul met een gemodificeerde nucleotidesequentie zoals gedefinieerd in een van de conclusies 1 tot 6, bij voorkeur waarbij het DNA-molecuul een enkelstrengs DNA, een dubbelstrengs DNA of een circulair DNA is, bij voorkeur waarbij het DNA molecuul een vector is, met meer voorkeur een expressievector of een virale vector.

8. Virusdeeltje dat codeert voor het RNA-molecuul met een gemodificeerde nucleotidesequentie zoals gedefinieerd in een van de conclusies 1 tot 6.

9. Cel omvattende het RNA-molecuul met een gemodificeerde nucleotidesequentie volgens een van de conclusies 1 tot 6 of het DNA-molecuul volgens conclusie 7 of het virusdeeltje volgens conclusie 8.

10. Een *in vitro* of *ex vivo* methode voor het moduleren van de expressie van een eiwit in een cel,
waarbij de methode bestaat uit een of meer van:
a) het introduceren van het RNA-molecuul met een gemodificeerde nucleotidesequentie zoals gedefinieerd in één van de conclusies 1 tot 6 in de cel;
b) het inbrengen van het DNA-molecuul volgens conclusie 7 in een cel en het laten overschrijven van het DNA-molecuul;
c) het infecteren van de cel met het virusdeeltje volgens conclusie 8;
d) het zodanig modificeren van een DNA-molecuul in een cel dat het gemodificeerde DNA-molecuul transcribeert naar een RNA-molecuul met een gemodificeerde nucleotidesequentie zoals gedefinieerd in één van de conclusies 1 tot 6; en/of
e) het modificeren van een RNA-molecuul in de cel om een RNA-molecuul te verkrijgen met f) een gemodificeerde nucleotidesequentie zoals gedefinieerd in één van de conclusies 1 tot 6.

11. Een *in vitro* of *ex vivo* methode voor het tot expressie brengen van een eiwit in een cel, waarbij de methode omvat:
het introduceren van een RNA-molecuul met een gemodificeerde nucleotidesequentie zoals gedefinieerd in een van de conclusies 1 tot 6, een DNA-molecuul zoals

gedefinieerd in conclusie 7, het virusdeeltje zoals gedefinieerd in conclusie 8 in een cel, of het verkrijgen van een cel zoals gedefinieerd in conclusie 9, en
het mogelijk maken van de translatie van het eiwit van het RNA-molecuul, en eventueel het isoleren of verkrijgen van het eiwit.

12. RNA-molecuul met een gemodificeerde nucleotidesequentie zoals gedefinieerd in een van de conclusies 1 tot 6, een DNA-molecuul zoals gedefinieerd in conclusie 7, het virusdeeltje zoals gedefinieerd in conclusie 8, de cel zoals gedefinieerd in conclusie 9, of de cel verkregen of verkrijgbaar door de werkwijze volgens conclusies 10 of 11, voor gebruik als geneesmiddel.

13. RNA-molecuul met een gemodificeerde nucleotidesequentie zoals gedefinieerd in een van de conclusies 1 tot 6, een DNA-molecuul zoals gedefinieerd in conclusie 7, het virusdeeltje zoals gedefinieerd in conclusie 8, de cel zoals gedefinieerd in conclusie 9, of de cel verkregen of verkrijgbaar door de werkwijze volgens conclusies 10 of 11 voor gebruik bij het behandelen, voorkomen of verbeteren van een ziekte bij een patiënt, waarbij het gebruik omvat het toedienen van het RNA-molecuul met een gemodificeerde nucleotidesequentie of het DNA-molecuul aan het individu.

14. RNA-molecuul met een gemodificeerde nucleotidesequentie, DNA-molecuul, virusdeeltje of cel voor gebruik volgens conclusie 13, waarbij de ziekte wordt gekozen uit:
kanker, een immuun gerelateerde aandoening, een bloedingsstoornis, een aandoening die verband houdt met over-expressie van een eiwit, een aandoening die verband houdt met onder-expressie van een eiwit.

15. RNA-molecuul met een gemodificeerde nucleotidesequentie, DNA-molecuul, virusdeeltje of cel voor gebruik volgens conclusie 13, waarbij het RNA-molecuul met een gemodificeerde nucleotidesequentie, DNA-molecuul, virusdeeltje of cel een vaccin is.

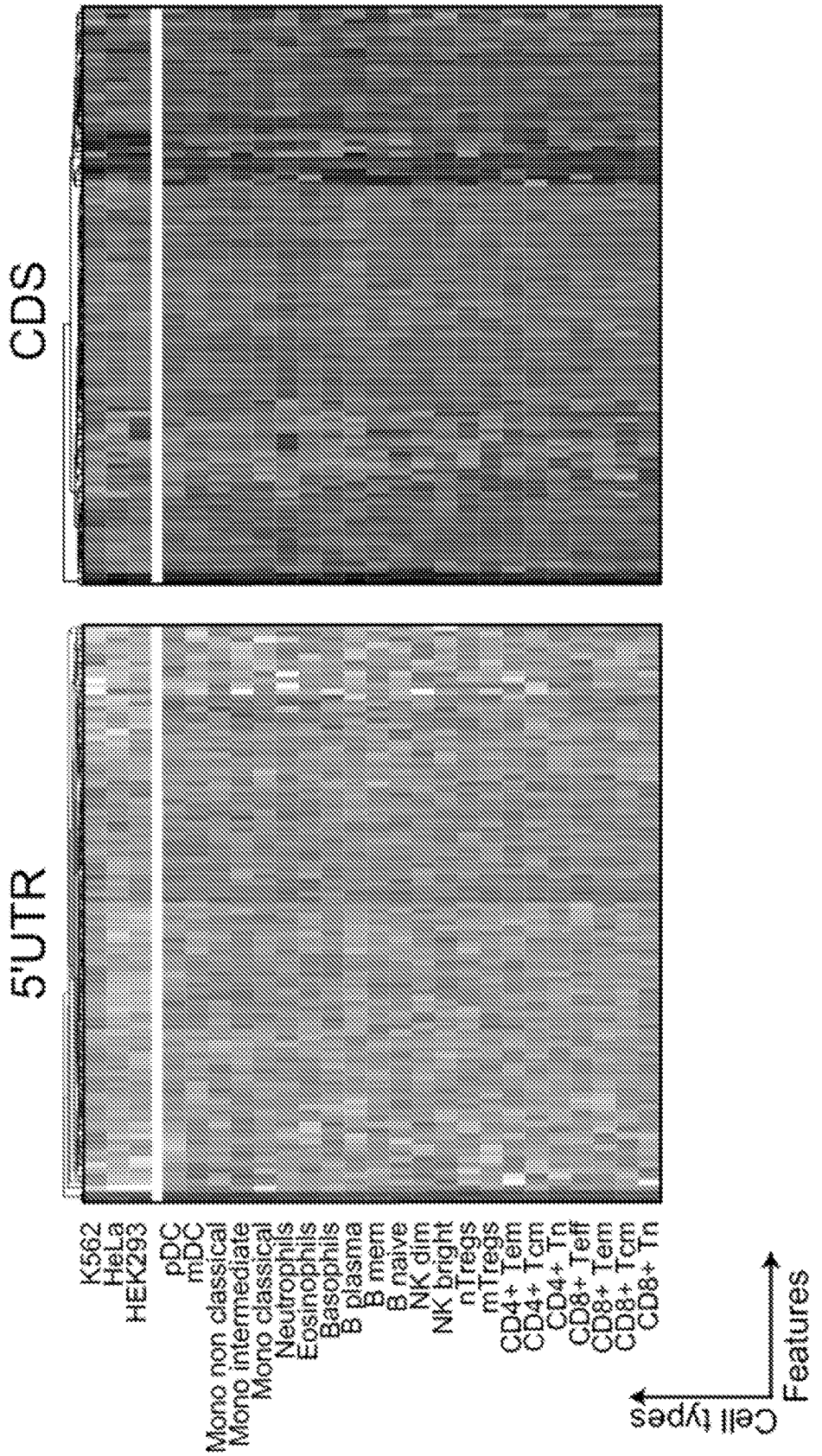
5'UTR
CDS
K562
HeLa
HEK293
pDC
mDC
Mono non classical
Mono intermediate
Mono classical
Neutrophils
Eosinophils
Basophils
B plasma
B mem
B naive
NK dim
NK bright
nTregs
mTregs
CD4+ Tem
CD4+ Tcm
CD4+ Tn
CD8+ Teff
CD8+ Tem
CD8+ Tcm
CD8+ Tn
Cell types
Features

Fig. 1

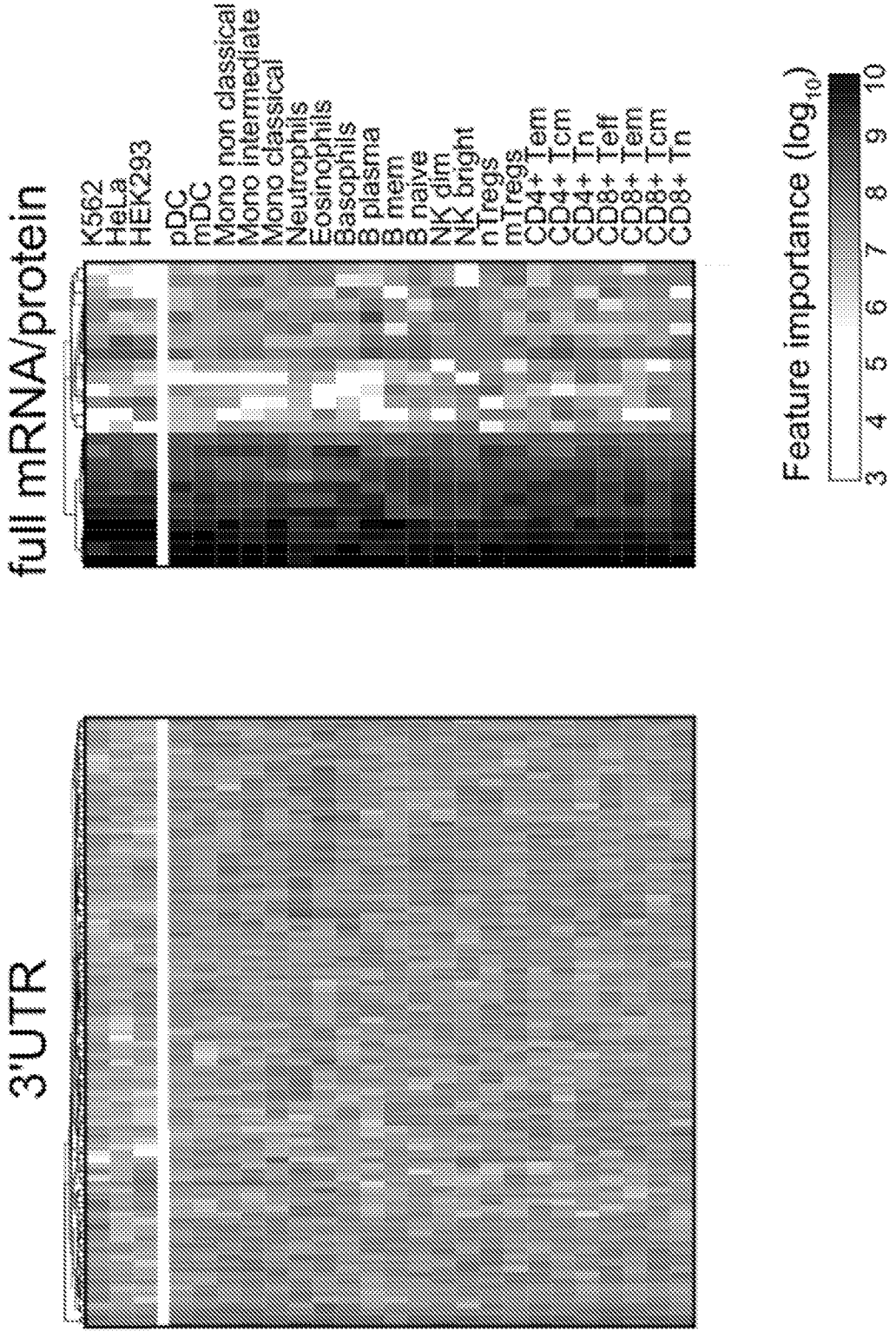

Fig. 1 (continued)

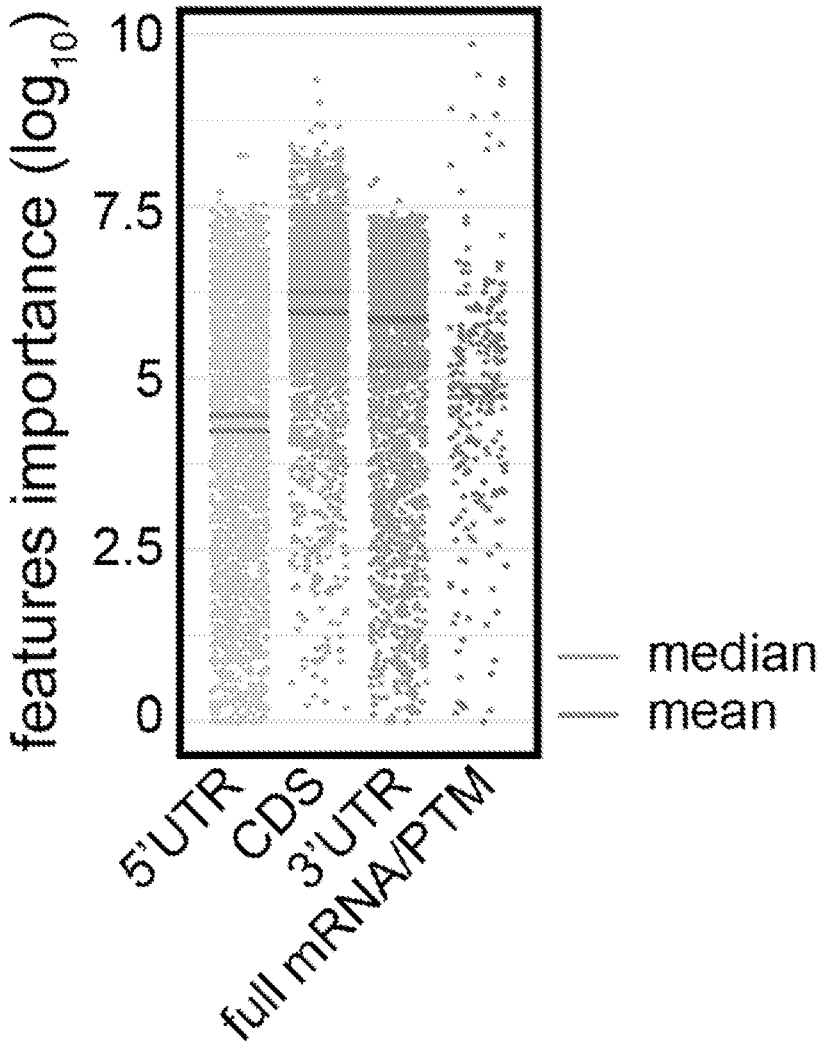
features importance (log10)
10
7.5
5
2.5
0
median
mean
5'UTR
CDS
3'UTR
full mRNA/PTM

Fig. 2

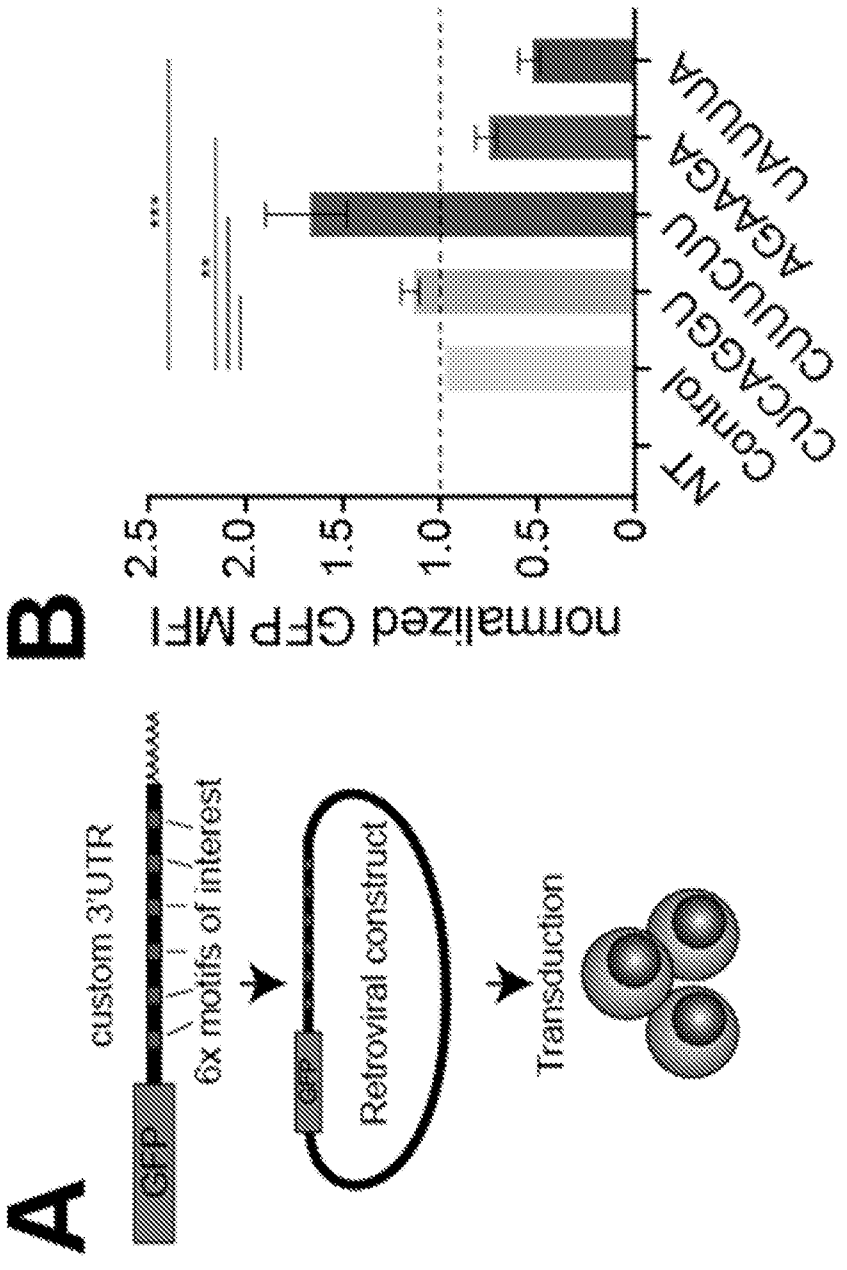
A
GFP
custom 3'UTR
6x motifs of interest
Retroviral construct
LTR
Transduction
B
normalized GFP MFI
2.5
2.0
1.5
1.0
0.5
0
NT
Control
CUCAGGU
CUUUCUU
AGAAAGA
UAUUUUA
***
**

Fig. 3

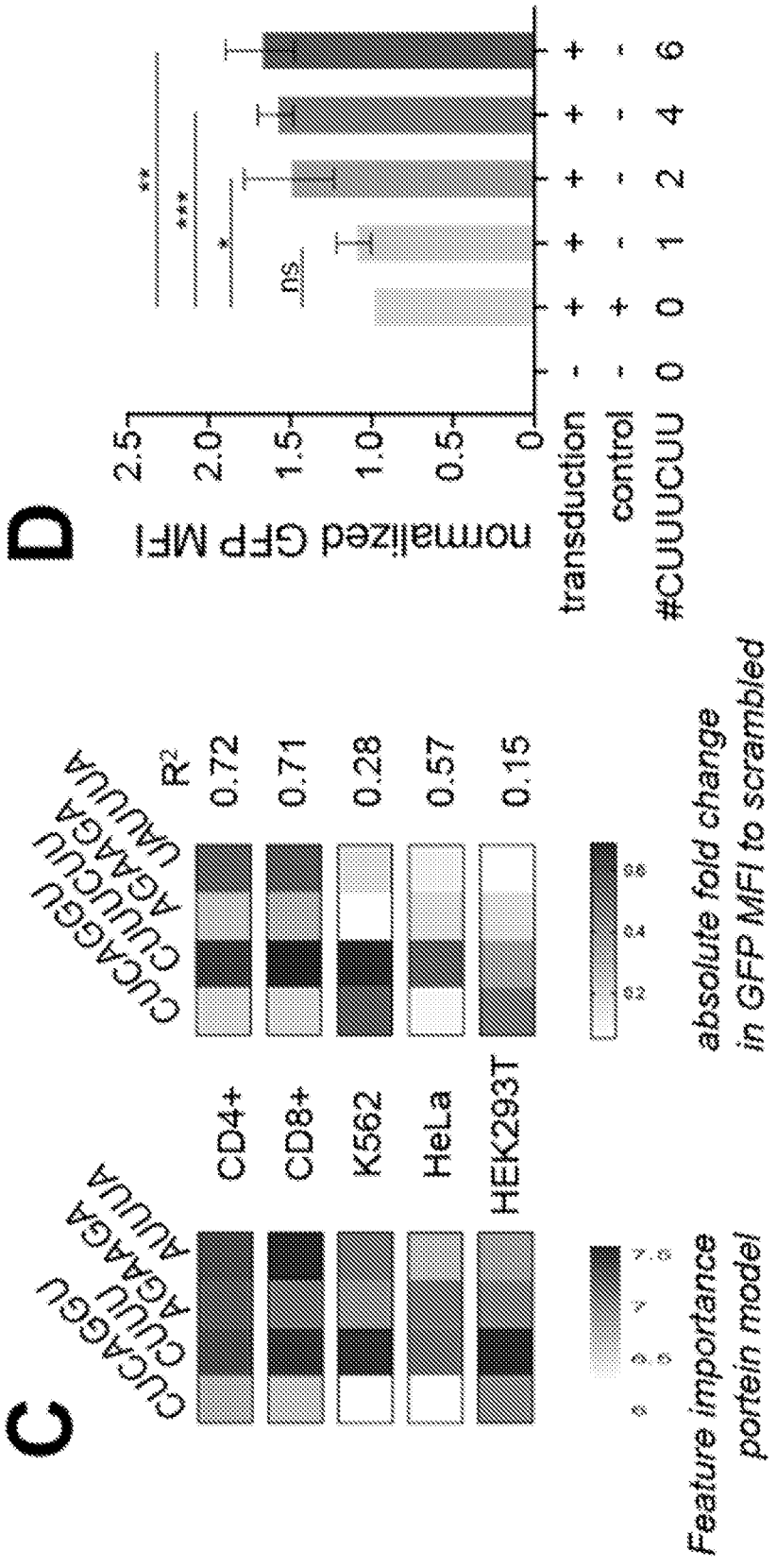

Fig. 3 (continued)

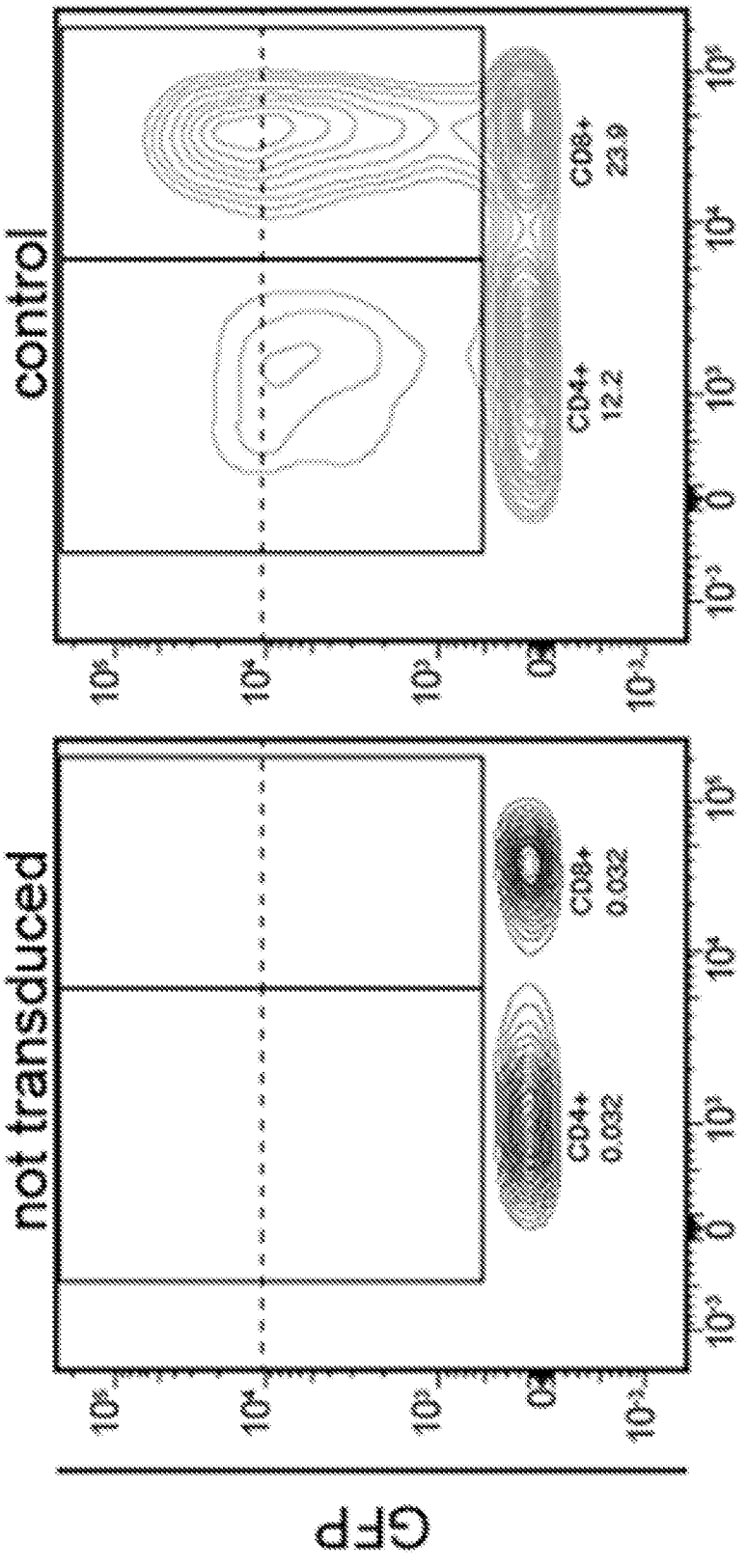

Fig. 4

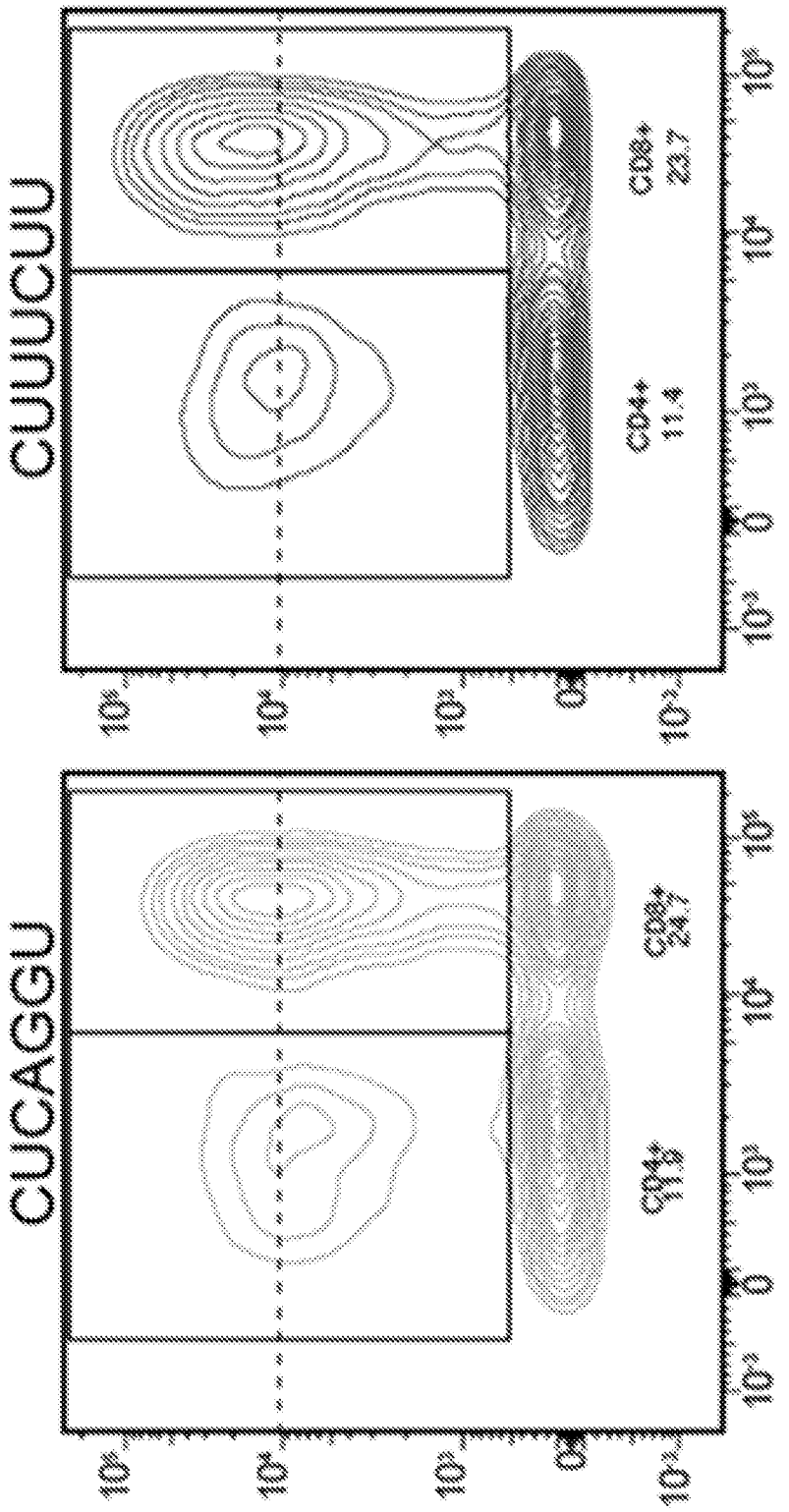

Fig. 4 (continued)

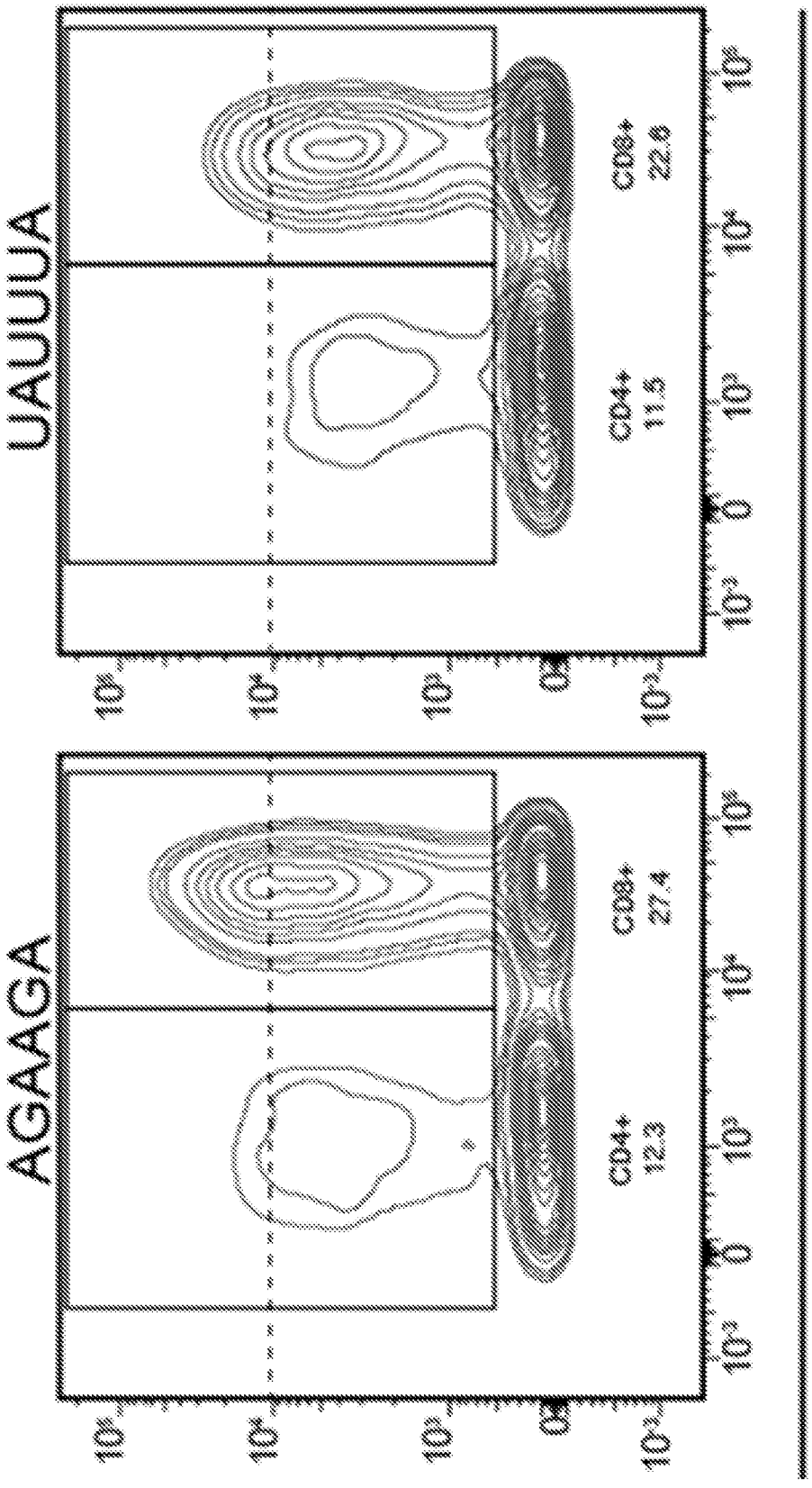

Fig. 4 (continued)

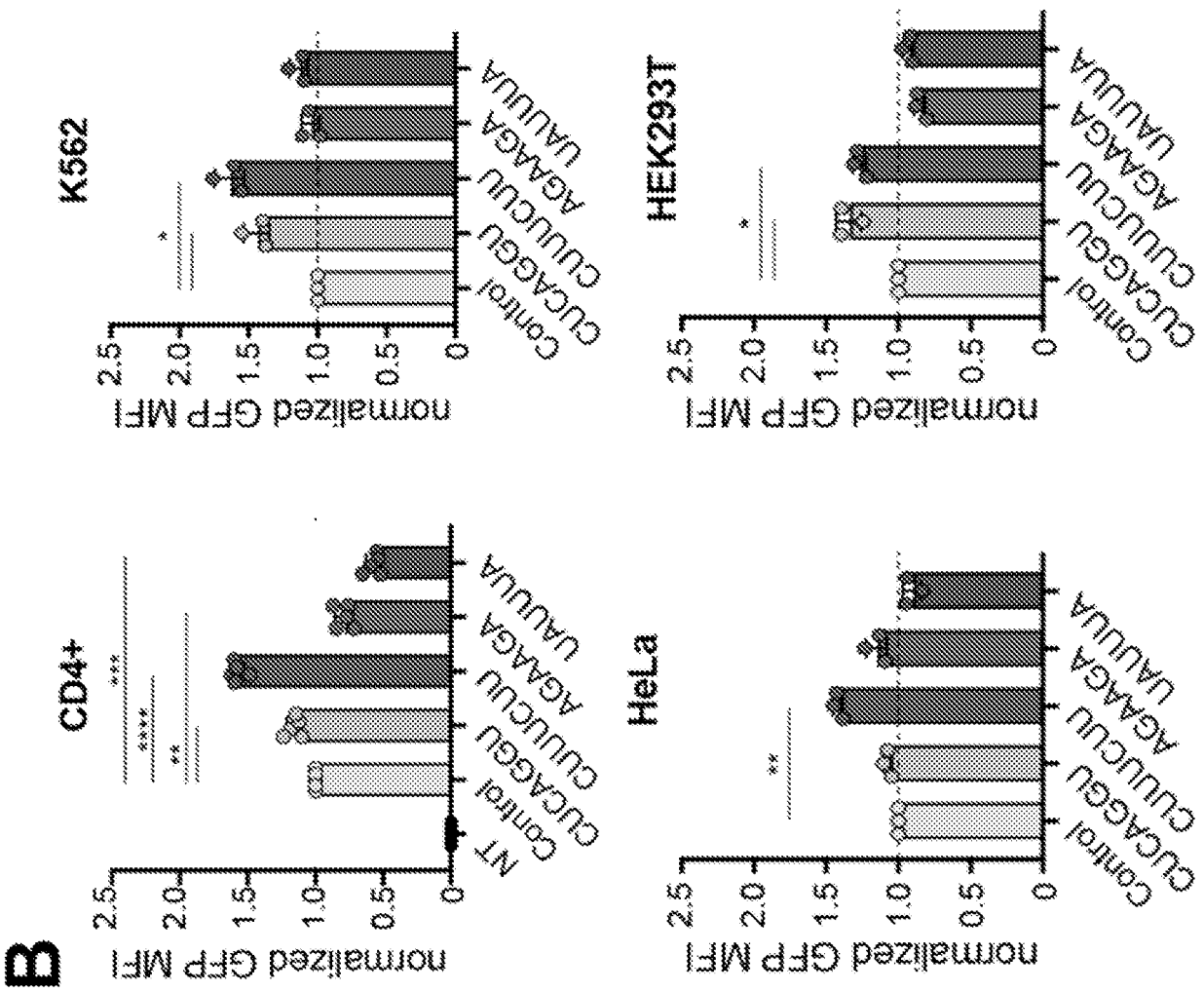
B
K562
HEK293T
CD4+
HeLa
normalized GFP MFI
2.5
2.0
1.5
1.0
0.5
0
NT
Control
CUCAGGU
CUUUCUU
AGAAGA
UAUUUA
*
**
***
****

Fig. 5

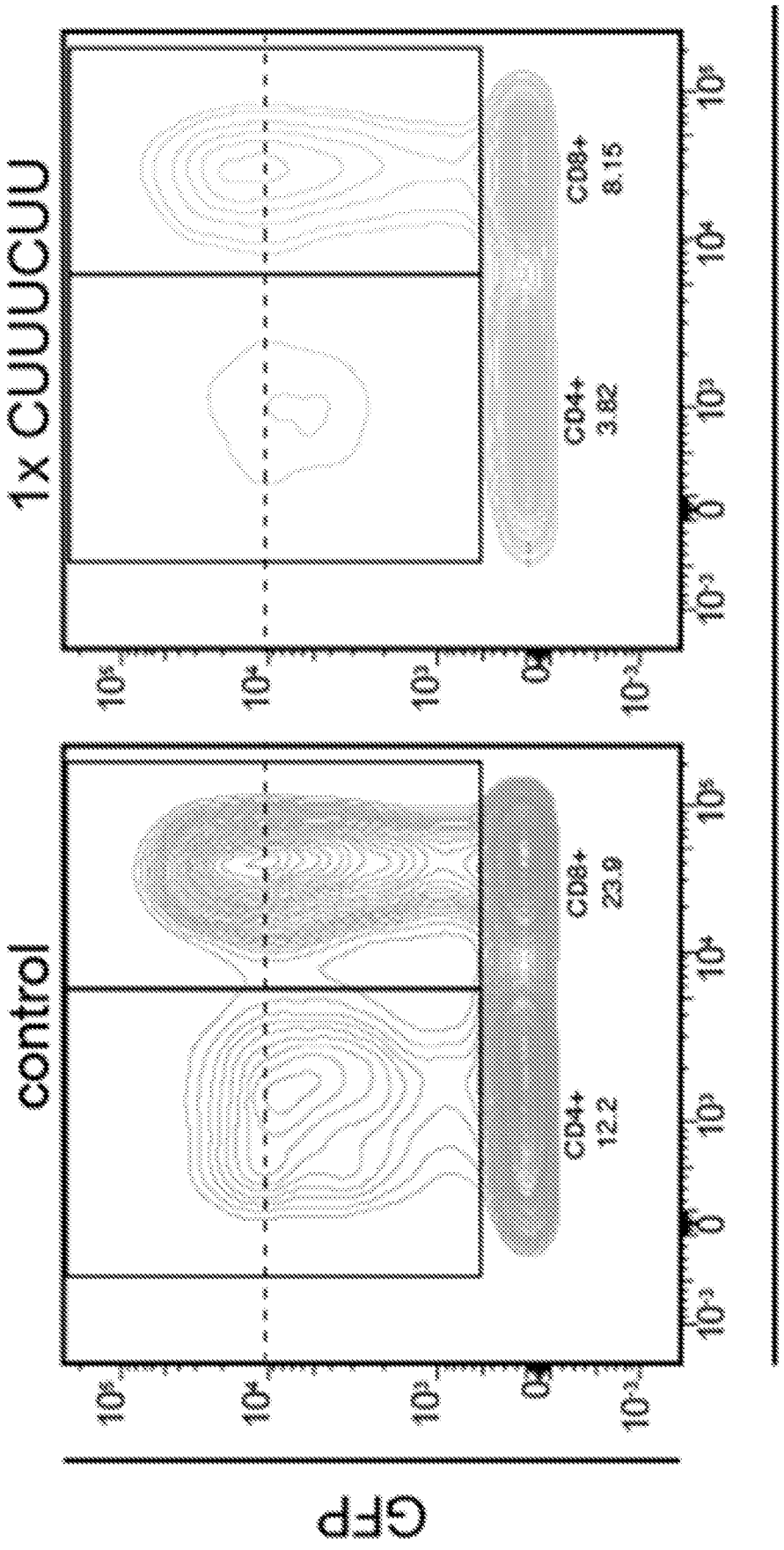

Fig. 6

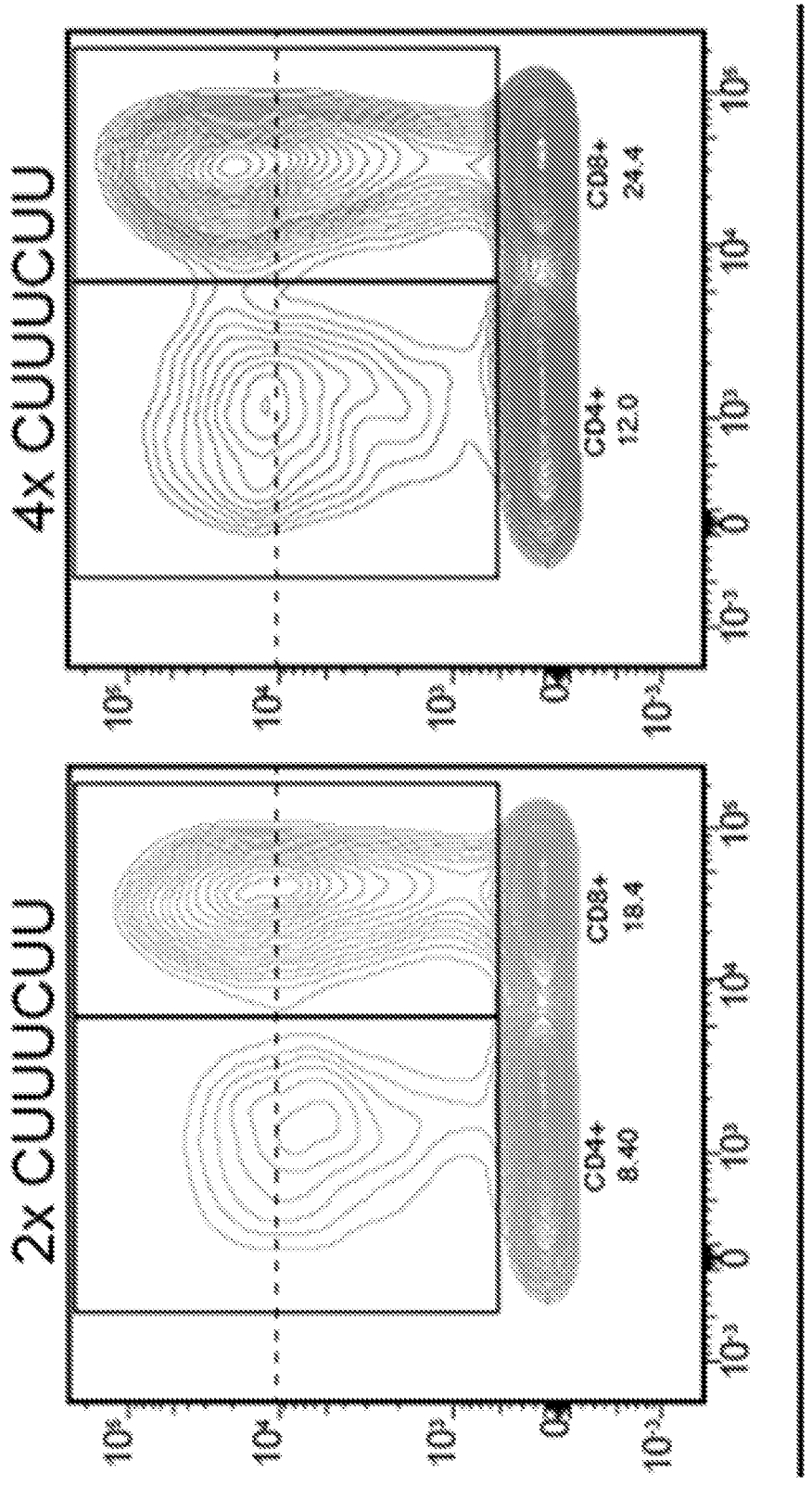

Fig. 6 (continued)

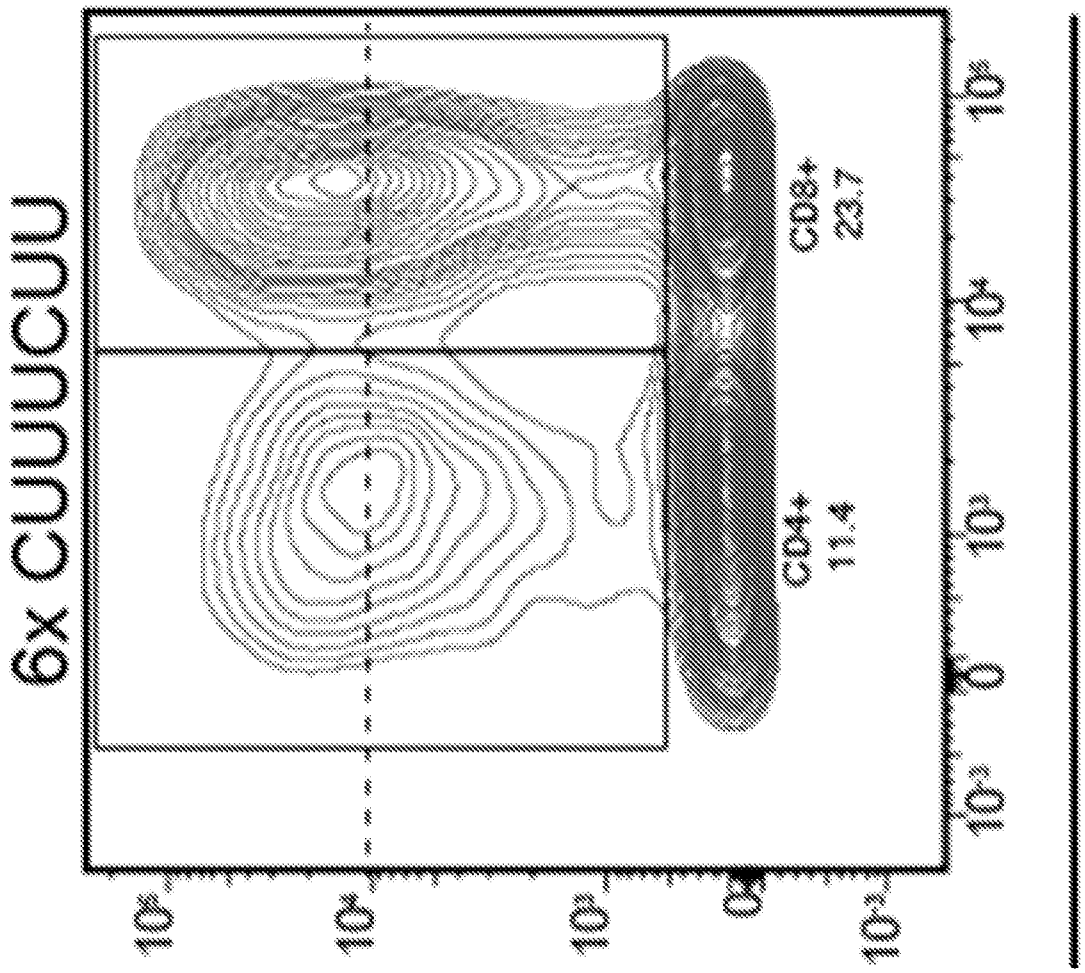

Fig. 6 (continued)

SAMENWERKINGSVERDRAG (PCT)

RAPPORT BETREFFENDE NIEUWHEIDSONDERZOEK VAN INTERNATIONAAL TYPE

IDENTIFICATIE VAN DE NATIONALE AANVRAGE	KENMERK VAN DE AANVRAGER OF VAN DE GEMACHTIGDE
Nederlands aanvraag nr. **2033707**	Indieningsdatum **09-12-2022**
	Ingeroepen voorrangsdatum
Aanvrager (Naam) **Stichting Sanquin Bloedvoorziening**	
Datum van het verzoek voor een onderzoek van internationaal type **21-01-2023**	Door de Instantie voor Internationaal Onderzoek aan het verzoek voor een onderzoek van internationaal type toegekend nr. **SN83149**

I. CLASSIFICATIE VAN HET ONDERWERP (bij toepassing van verschillende classificaties, alle classificatiesymbolen opgeven)

Volgens de internationale classificatie (IPC)

Zie onderzoeksrapport

II. ONDERZOCHTE GEBIEDEN VAN DE TECHNIEK

Onderzochte minimumdocumentatie

Classificatiesysteem	Classificatiesymbolen
IPC	**Zie onderzoeksrapport**

Onderzochte andere documentatie dan de minimum documentatie, voor zover dergelijke documenten in de onderzochte gebieden zijn opgenomen

III.		**GEEN ONDERZOEK MOGELIJK VOOR BEPAALDE CONCLUSIES**	(opmerkingen op aanvullingsblad)
IV.	**X**	**GEBREK AAN EENHEID VAN UITVINDING**	(opmerkingen op aanvullingsblad)

ONDERZOEKSRAPPORT BETREFFENDE HET RESULTAAT VAN HET ONDERZOEK NAAR DE STAND VAN DE TECHNIEK VAN HET INTERNATIONALE TYPE

Nummer van het verzoek om een onderzoek naar de stand van de techniek
NL 2033707

A. CLASSIFICATIE VAN HET ONDERWERP

INV. C12N15/63 C12N15/67 C12N15/113 C12N15/86
ADD.

Volgens de Internationale Classificatie van octrooien (IPC) of zowel volgens de nationale classificatie als volgens de IPC.

B. ONDERZOCHTE GEBIEDEN VAN DE TECHNIEK

Onderzochte miminum documentatie (classificatie gevolgd door classificatiesymbolen)

C12N C40B C07K

Onderzochte andere documentatie dan de mimimum documentatie, voor dergelijke documenten, voor zover dergelijke documenten in de onderzochte gebieden zijn opgenomen

Tijdens het onderzoek geraadpleegde elektronische gegevensbestanden (naam van de gegevensbestanden en, waar uitvoerbaar, gebruikte trefwoorden)

EPO-Internal, WPI Data, BIOSIS, CHEM ABS Data, Sequence Search, EMBASE

C. VAN BELANG GEACHTE DOCUMENTEN

Categorie °	Geciteerde documenten, eventueel met aanduiding van speciaal van belang zijnde passages	Van belang voor conclusie nr.
	EENHEID VAN UITVINDING ONTBREEKT **zie aanvullingsblad B** **-----**	
A	**TONG XIN ET AL: "Enhancement of p53 expression in keratinocytes by the bioflavonoid apigenin is associated with RNA-binding protein HuR", MOLECULAR CARCINOGENESIS, deel 48, nr. 2, 1 februari 2009 (2009-02-01), bladzijden 118-129, XP93050742, US ISSN: 0899-1987, DOI: 10.1002/mc.20460 * bladzijde 122, rechter kolom, alinea 2; figuur 4 ***	**1-15**

	-/--	

[X] Verdere documenten worden vermeld in het vervolg van vak C.

[X] Leden van dezelfde octrooifamilie zijn vermeld in een bijlage

° Speciale categorieën van aangehaalde documenten

"A" niet tot de categorie X of Y behorende literatuur die de stand van de techniek beschrijft

"D" in de octrooiaanvrage vermeld

"E" eerdere octrooi(aanvrage), gepubliceerd op of na de indieningsdatum, waarin dezelfde uitvinding wordt beschreven

"L" om andere redenen vermelde literatuur

"O" niet-schriftelijke stand van de techniek

"P" tussen de voorrangsdatum en de indieningsdatum gepubliceerde literatuur

"T" na de indieningsdatum of de voorrangsdatum gepubliceerde literatuur die niet bezwarend is voor de octrooiaanvrage, maar wordt vermeld ter verheldering van de theorie of het principe dat ten grondslag ligt aan de uitvinding

"X" de conclusie wordt als niet nieuw of niet inventief beschouwd ten opzichte van deze literatuur

"Y" de conclusie wordt als niet inventief beschouwd ten opzichte van de combinatie van deze literatuur met andere geciteerde literatuur van dezelfde categorie, waarbij de combinatie voor de vakman voor de hand liggend wordt geacht

"&" lid van dezelfde octrooifamilie of overeenkomstige octrooipublicatie

Datum waarop het onderzoek naar de stand van de techniek van internationaal type werd voltooid	Verzenddatum van het rapport van het onderzoek naar de stand van de techniek van internationaal type
13 juli 2023	

Naam en adres van de instantie	De bevoegde ambtenaar
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	**Petri, Bernhard**

2

ONDERZOEKSRAPPORT BETREFFENDE HET RESULTAAT VAN HET ONDERZOEK NAAR DE STAND VAN DE TECHNIEK VAN HET INTERNATIONALE TYPE

Nummer van het verzoek om een onderzoek naar de stand van de techniek
NL 2033707

C.(Vervolg). VAN BELANG GEACHTE DOCUMENTEN

Categorie °	Geciteerde documenten, eventueel met aanduiding van speciaal van belang zijnde passages	Van belang voor conclusie nr.
A	CHUNXI ZENG ET AL: "Leveraging mRNA Sequences and Nanoparticles to Deliver SARS-CoV-2 Antigens In Vivo", ADVANCED MATERIALS, VCH PUBLISHERS, DE, deel 32, nr. 40, 2 september 2020 (2020-09-02), bladzijde n/a, XP071875293, ISSN: 0935-9648, DOI: 10.1002/ADMA.202004452 * bladzijde 3, rechter kolom, regels 47-52; figuur 4a; tabel S4 *	1-15
A	-& ZENG CHUNXI ET AL: "Leveraging mRNA Sequences and Nanoparticles to Deliver SARS-CoV-2 Antigens In Vivo - Supporting Information", ADVANCED MATERIALS, deel 32, nr. 40, 2 september 2020 (2020-09-02), bladzijde 2004452, XP93052391, DE ISSN: 0935-9648, DOI: 10.1002/adma.202004452 Gevonden op het Internet: URL:https://onlinelibrary.wiley.com/doi/full-xml/10.1002/adma.202004452> * tabellen S4, S2 *	1-15

X	WO 2016/107877 A1 (CUREVAC AG [DE]) 7 juli 2016 (2016-07-07) * bladzijde 24, regel 27 - bladzijde 27, regel 9; sequenties 306, 311, 377, 378 * * bladzijde 53, regel 16 - bladzijde 55, regel 24 * * bladzijde 65, regel 29 - bladzijde 121, regel 51 *	1-15

X	LAGNADO C A ET AL: "AUUUA IS NOT SUFFICIENT TO PROMOTE POLY(A) SHORTENING AND DEGRADATION OF AN MRNA: THE FUNCTIONAL SEQUENCE WITHIN AU-RICH ELEMENTS MAY BE UUAUUUA(U/A)(U/A)", MOLECULAR AND CELLULAR BIOLOGY, AMERICAN SOCIETY FOR PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS, US, deel 14, nr. 12, 1 december 1994 (1994-12-01), bladzijden 7984-7995, XP009023841, ISSN: 0270-7306 * figuur 1 *	1-15

	-/--	

ONDERZOEKSRAPPORT BETREFFENDE HET RESULTAAT VAN HET ONDERZOEK NAAR DE STAND VAN DE TECHNIEK VAN HET INTERNATIONALE TYPE

Nummer van het verzoek om een onderzoek naar de stand van de techniek
NL 2033707

C.(Vervolg). VAN BELANG GEACHTE DOCUMENTEN

Categorie °	Geciteerde documenten, eventueel met aanduiding van speciaal van belang zijnde passages	Van belang voor conclusie nr.
X	E. JACINTO-LOEZA ET AL: "Minigene-like inhibition of protein synthesis mediated by hungry codons near the start codon", NUCLEIC ACIDS RESEARCH, deel 36, nr. 13, 1 augustus 2008 (2008-08-01), bladzijden 4233-4241, XP055059444, ISSN: 0305-1048, DOI: 10.1093/nar/gkn395 * samenvatting *	1-15
A	YAN QINGQING ET AL: "Depletion of Arabidopsis SC35 and SC35-like serine/arginine-rich proteins affects the transcription and splicing of a subset of genes", PLOS GENETICS, deel 13, nr. 3, 8 maart 2017 (2017-03-08), bladzijde e1006663, XP093055625, USA ISSN: 1553-7390, DOI: 10.1371/journal.pgen.1006663 * samenvatting *	1-15
X	US 2006/014146 A1 (SOUCAILLE PHILIPPE [FR] ET AL) 19 januari 2006 (2006-01-19) * alineas [0009], [0165]; sequentie 32 *	1-15

GEBREK AAN EENHEID VAN UITVINDING

Octrooiaanvrage Nr.:
SN 83149
NL 2033707

AANVULLINGSBLAD B

De Instantie belast met het uitvoeren van het onderzoek naar de stand van de techniek heeft vastgesteld dat deze aanvrage meerdere uitvindingen bevat, te weten:

1. conclusies: 1-15(gedeeltelijk)

RNA molecule comprising a sequence element in the UTR of the RNA and wherein the sequence element is represented as No 1 in Table 1

2-1795. conclusies: 1-15(gedeeltelijk)

as invention 1, wherein the sequence element is represented as No 2 , 3, 4 ... , 1794, 1795, respectively

---

1796-3670. conclusies: 1-15(gedeeltelijk)

as invention 1, wherein the sequence element is represented as No 1796, 1797, ... , 3669, 3670 of Table 2, respectively

---

Het vooronderzoek werd tot het eerste onderwerp beperkt.

ONDERZOEKSRAPPORT BETREFFENDE HET RESULTAAT VAN HET ONDERZOEK NAAR DE STAND VAN DE TECHNIEK VAN HET INTERNATIONALE TYPE

Informatie over leden van dezelfde octrooifamilie

Nummer van het verzoek om een onderzoek naar de stand van de techniek

NL 2033707

In het rapport genoemd octrooigeschrift	Datum van publicatie	Overeenkomend(e) geschrift(en)	Datum van publicatie
WO 2016107877 A1	07-07-2016	AU 2015373404 A1	25-05-2017
		BR 112017009835 A2	26-12-2017
		CA 2966092 A1	07-07-2016
		CN 107124889 A	01-09-2017
		EP 3240558 A1	08-11-2017
		EP 3494982 A1	12-06-2019
		JP 6907116 B2	21-07-2021
		JP 2018501802 A	25-01-2018
		JP 2021168664 A	28-10-2021
		KR 20170100660 A	04-09-2017
		RU 2017127203 A	01-02-2019
		SG 10201906673W A	27-09-2019
		SG 11201704681Q A	28-07-2017
		US 2018148727 A1	31-05-2018
		US 2019345504 A1	14-11-2019
		WO 2016107877 A1	07-07-2016
US 2006014146 A1	19-01-2006	AT 401422 T	15-08-2008
		AU 2003230982 A1	03-11-2003
		CA 2483338 A1	30-10-2003
		DK 1501947 T3	17-11-2008
		EP 1501947 A2	02-02-2005
		JP 2005523015 A	04-08-2005
		US 2006014146 A1	19-01-2006
		WO 03089605 A2	30-10-2003

WRITTEN OPINION

File No.	Filing date *(day/month/year)*	Priority date *(day/month/year)*	Application No.
SN83149	09.12.2022		NL2033707

International Patent Classification (IPC)
INV. C12N15/63 C12N15/67 C12N15/113 C12N15/86

Applicant
Stichting Sanquin Bloedvoorziening

This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☒ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☒ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the application
- ☒ Box No. VIII Certain observations on the application

Examiner
Petri, Bernhard

Application number

WRITTEN OPINION

NL2033707

Box No. I Basis of this opinion

1. This opinion has been established on the basis of the latest set of claims filed before the start of the search.

2. With regard to any **nucleotide and/or amino acid sequence** disclosed in the application, this opinion has been established on the basis of a sequence listing:

   a. ☒ forming part of the application as filed.

   b. ☐ furnished subsequent to the filing date for the purposes of search,

      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the application as filed.

3. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the application, this opinion has been established to the extent that a meaningful opinion could be formed without a WIPO Standard ST.26 compliant sequence listing.

4. Additional comments:

Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

The questions whether the claimed invention appears to be novel, to involve an inventive step, or to be industrially applicable have not been examined in respect of

☐ the entire application

☒ claims Nos. 1-15(gedeeltelijk)

because:

☐ the said application, or the said claims Nos. relate to the following subject matter which does not require a search *(specify)*:

☐ the description, claims or drawings *(indicate particular elements below)* or said claims Nos. are so unclear that no meaningful opinion could be formed *(specify)*:

☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed *(specify)*:

☒ no search report has been established for the whole application or for said claims Nos. 1-15(gedeeltelijk)

☐ a meaningful opinion could not be formed without the sequence listing; the applicant did not furnish a sequence listing complying with WIPO Standard ST.26.

☐ See Supplemental Box for further details.

WRITTEN OPINION

Box No. IV Lack of unity of invention

1. The requirement of unity of invention is not complied with for the following reasons:

   see separate sheet

2. This report has been established in respect of the following parts of the application:

   ☐ all parts.

   ☒ the parts relating to claims Nos. (see Search Report)

Box No. V Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty	Yes:	Claims	
	No:	Claims	1-15(gedeeltelijk)
Inventive step	Yes:	Claims	
	No:	Claims	1-15(gedeeltelijk)
Industrial applicability	Yes:	Claims	1-15(gedeeltelijk)
	No:	Claims	

2. Citations and explanations

   see separate sheet

Box No. VIII Certain observations on the application

see separate sheet

1 The application refers to the bioinformatic identification of 5' and 3' UTR elements predicted to correlate with increased or decreased expression rates.

Five elements, i.e.

CUUUCUU - SEQ ID NO: 3,

CUCAGGU - SEQ ID NO: 2,

UAUUUA - SEQ ID NO: 5, and

AGAAGA - SEQ ID NO: 4

were tested in a randomized scrambled 3' UTR of 180 nt in a GFP reporter assay.

Models for predicting mRNA and protein abundance based on a library of sequence features extracted from the human transcriptome, including GC content, transcript length, RNA-binding protein motifs, predicted miR seed score and codon usage are developed and validated on published expression data from HEK293T, HeLa, K562 and PB-derived human T-, B-cells, DCs, monocytes and GC subsets.

The model itself is not disclosed.

Cited Prior Art

2 Reference is made to the following documents:

D1 TONG XIN ET AL: "Enhancement of p53 expression in keratinocytes by the bioflavonoid apigenin is associated with RNA-binding protein HuR",
MOLECULAR CARCINOGENESIS,
deel 48, nr. 2, 1 februari 2009 (2009-02-01), bladzijden 118-129,
XP93050742,
US
ISSN: 0899-1987, DOI: 10.1002/mc.20460

D2 CHUNXI ZENG ET AL: "Leveraging mRNA Sequences and Nanoparticles to Deliver SARS-CoV-2 Antigens In Vivo",
ADVANCED MATERIALS, VCH PUBLISHERS, DE,
deel 32, nr. 40, 2 september 2020 (2020-09-02), bladzijde n/a,

XP071875293,
ISSN: 0935-9648, DOI: 10.1002/ADMA.202004452 ; -& ZENG CHUNXI ET AL: "Leveraging mRNA Sequences and Nanoparticles to Deliver SARS-CoV-2 Antigens In Vivo - Supporting Information", ADVANCED MATERIALS,
deel 32, nr. 40, 2 september 2020 (2020-09-02), bladzijde 2004452, XP93052391,
DE
ISSN: 0935-9648, DOI: 10.1002/adma.202004452
Gevonden op het Internet:
URL:https://onlinelibrary.wiley.com/doi/full-xml/10.1002/adma.202004452

D3 WO 2016/107877 A1 (CUREVAC AG [DE]) 7 juli 2016 (2016-07-07)

D4 LAGNADO C A ET AL: "AUUUA IS NOT SUFFICIENT TO PROMOTE POLY(A) SHORTENING AND DEGRADATION OF AN MRNA: THE FUNCTIONAL SEQUENCE WITHIN AU-RICH ELEMENTS MAY BE UUAUUUA(U/A)(U/A)",
MOLECULAR AND CELLULAR BIOLOGY, AMERICAN SOCIETY FOR PHARMACOLOGY AND EXPERIMENTAL THERAPEUTICS, US,
deel 14, nr. 12, 1 december 1994 (1994-12-01), bladzijden 7984-7995, XP009023841,
ISSN: 0270-7306

D5 E. JACINTO-LOEZA ET AL: "Minigene-like inhibition of protein synthesis mediated by hungry codons near the start codon",
NUCLEIC ACIDS RESEARCH,
deel 36, nr. 13, 1 augustus 2008 (2008-08-01), bladzijden 4233-4241, XP055059444,
ISSN: 0305-1048, DOI: 10.1093/nar/gkn395

D6 YAN QINGQING ET AL: "Depletion of Arabidopsis SC35 and SC35-like serine/arginine-rich proteins affects the transcription and splicing of a subset of genes",
PLOS GENETICS,
deel 13, nr. 3, 8 maart 2017 (2017-03-08), bladzijde e1006663,

XP093055625,
USA
ISSN: 1553-7390, DOI: 10.1371/journal.pgen.1006663

D7 US 2006/014146 A1 (SOUCAILLE PHILIPPE [FR] ET AL) 19 januari 2006 (2006-01-19)

Re Item IV

Lack of unity of invention

3 The application relates to the development of a model for the prediction of mRNA/protein expression levels.

Tables 1 and 2 describe the sequence elements that were identified using a ML algorithm to correlated mRNA expression levels and sequence data to protein levels. The inventors found that up to 61 % of observed mRNA levels and up to 63% of protein expression levels could be explained based on the presence of sequence elements as defined herein.

Table 1 describes sequence elements that are specifically identified in the 5' UTR and thus anticipated to correlate with a modulated protein expression when introduced in the 5' UTR of the RNA molecule.

Table 2 describes sequence elements that are specifically identified in the 3' UTR and thus anticipated to correlate with a modulated protein expression when introduced in the 3' UTR of the RNA molecule.

Four particular sequence elements i.e. CUUUCUU - SEQ ID NO: 3, CUCAGGU - SEQ ID NO: 2, UAUUUA - SEQ ID NO: 5, and AGAAGA - SEQ ID NO: 4 were tested in a scambled 3' UTR test sequence for effects on expression in a GFP reporter assay.

3'UTRs containing AGAAGA motifs (SEQ ID NO: 4) or UAUUUA motifs (SEQ ID NO: 5) were found to suppressed the GFP protein expression.

The CUCAGGU and CUUUCUU motifs (SEQ ID NO: 1 and SEQ ID NO: 2) in the 3'UTR augmented GFP protein levels compared to the scrambled control (SEQ ID NO: 1) 3'UTR (Fig. 3A).

4 Thus the technical problem which can be extracted from the application as filed resides in the identification of predicted motifs which can be used to modulate mRNA/protein expression (cf [0004] and [0005] of the application as filed).

5 The alleged solution is represented in the 3670 alternative sequence motives depicted in Tables 1 and 2.

6 These 3670 alternative solutions however do not relate to a group of inventions so linked as to form a single inventive concept.

The reason being as follows:

Where a group of inventions is claimed, the requirement of unity of invention shall be fulfilled only when there is a technical relationship among those inventions involving one or more of the same or corresponding special technical features. The expression "special technical features" shall mean those features which define a contribution which each of the claimed inventions considered as a whole makes over the prior art.

For chemical alternatives as in the instant case the requirement for same or corresponding special technical features is considered met if the alternatives are of a similar nature, i.e. all alternatives have a common property and activity **and** share a significant structural element. In the instant case the 3670 sequence motifs do not share a particular sequence motif which is responsible for the common activity, and which leads to a technical contribution over the art. Thus no significant structural element is shared by the claimed alternative 3670 alternative UTR motives.

7 The fact that all sequence motifs are motifs expected to modulate mRNA/protein expression when integrated into 3' or 5' UTRs of target mRNAs cannot constitute a special technical feature, as such sequence motifs are known in the art.

D1 discloses page 122, right column, 2nd paragraph and Fig 4 fusion of a fragment from the 3'UTR of p53 comprising AUUUA to a reporter gene and monitoring the effects on translational upregulation by apigenin.

D2 discloses at page 3, right column, lines 47-52 and Fig 4 and Tables S2 and S4 3' and 5' UTR motifs and their integration into UTR for comprehensive UTR engineering of a mRNA for expression of SARS-CoV2 antigens. The motifs are miRNA binding sites integrated into de novo designed NCA-7 and NCA-8 (cf Table S2) and QRE1/2, R3U, and ARE integrated into S27a.

8 As a consequence and due to the fact that no other technical features can be distinguished which in the light of the prior art could be regarded as special technical features, there is lack of unity.

9 Thus the technical content of the present application is rearranged into 3670 individual objective problems with independent solutions as follows:

10

Invention 1.	Problem: Provision of an mRNA with modulated expression Solution: Claims: 1-15(all partially) RNA molecule comprising a sequence element in the UTR of the RNA and wherein the sequence element is represented as No 1 in Table 1
Inventions 2-1795.	Problem: Provision of an mRNA with modulated expression Solution: Claims: 1-15(all partially) as invention 1, wherein the sequence element is represented as No 2 , 3, 4 ... , 1794, 1795, respectively
Inventions 1796-3670.	Problem: Provision of an mRNA with modulated expression Solution: Claims: 1-15(all partially) as invention 1, wherein the sequence element is represented as No 1796, 1797, ... , 3669, 3670 of Table 2, respectively

11 Each of these separate inventions/group of inventions has its own relevant prior art to be determined, requiring independent searches to be conducted.

As a consequence, the search of all of these separate inventions/group of inventions would have required a considerable additional effort. Only the first invention has been searched.

Please also consider the possibility that separate search strategies may lead to overlapping sets of documents. This can only be judged based on hindsight after the search effort has been spent and does not imply that such an additional search did not require considerable additional effort.

12 The search has been conducted for invention 1, i.e. sequence motif 'CUCUUU'. Albeit formally belonging to other inventions, to cover the experimental part of the application, also sequence motifs CUUUCUU - SEQ ID NO: 3, CUCAGGU - SEQ ID NO: 2, UAUUUA - SEQ ID NO: 5, and AGAAGA - SEQ ID NO: 4 have been included into the search.

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

Invention 1: RNA comprising 'CUCUUU'. or 'CUUUCUU', 'CUCAGGU', 'UAUUUA' or 'AGAAGA'.

Novelty

13 Claims 1-15 lack prima facie novelty. RNAs comprising the claimed motifs in their UTRs are obviously known form the prior art. In fact said motifs have been determined in publicly available mRNAs (cf. page 26 line 27 of the application as filed).

14 The initial search for each of the short motifs reveals an extremely large number of potential hits. More than 4.000.000 nucleic acids exhibiting each of the claimed motifs were detected in databases such as Chemical Abstracts REGISTRY. It is not possible in a comprehensive way to limit the search to UTRs of mRNAs only. Thus the representative citation of D3 does not reflect the entire state of the art of potential novelty destroying nucleic acids.

15 In addition to the observed novelty overflow which in fact rendered a meaningful search over the entire scope impossible it is observed that the claimed motifs are rather abundant even in UTRs know to affect expression of mRNAs and proteins thereof.

Exemplary cited reference is D3.

D3 discloses at page 24 line 27 - page 27, line 9, page 53 line 16 - page 55 line 24 and page 65 line 29 - page 121 line 51 the identification 3' and 5' UTRs for

The following table provides detected motifs in exemplary UTRs disclosed in D3:

SEQ ID NO	CUCUUU	CUUUCUU	CUCAGGU	UAUUUA	AGAAGA
378	35-40				

306		25-31, 282-288			
370			13-19		
311				381-386, 416-421	
377					177-182

16 D4 discloses ARE motif UAUUUA in reporter assay constructs. Thus D4 explicitly anticipates subject-matter in relation to SEQ ID NO 5.

17 Similarly D5 discloses AGAAGA motif in lacZ constructs to inhibit expression. Thus D5 explicitly anticipates subject-matter in relation to SEQ ID NO 4.

18 D6 discloses motif AGAAGA as binding site of splicing factors SC35/SC35-like proteins.

19 D7 also discloses motif AGAAGA, as ribosome binding site, introduced into the upstream region of lacZ.

20 This analysis also applies to method claim 11 as said claim merely defines the introduction of modified nucleic acids (cf below, section VIII).

Inventive step

21 Due to the high abundance of short sequence motives such as 'CUCUUU'. or 'CUUUCUU', 'CUCAGGU', 'UAUUUA' or 'AGAAGA' even explicitly in UTRs of mRNAs, no distinguishing features are present which could characterize product claims. Such no sensible analysis of inventive step is possible.

22 The identification of particular motifs which can be used to modulated expression could give rise to use claims directed at such particular motifs (cf however above D4-D7 for motifs UAUUUA and AGAAGA).

23 Furthermore e.g. D4 sheds doubts that such motives may be universally applicable outside the particular experimental context, as in D4 the motif UAUUUA was found ineffective as destabilizing element.

Re Item VIII

Certain observations on the application

24 The application lacks conciseness. The claims are directed at mRNAs defined in relation to sequence elements depicted in Tables 1 and 2. Tables 1 and 2 are depicted at pages 33 - 44 and 45 - 56 and depict 5' UTR and 3' UTR motifs, respectively. In total 3670 UTR motifs are thus the individualized subject of claims 1-15.

It is not possible to execute 3670 independent structure searches for mRNAs comprising the motifs of Tables 1 or 2.

25 The above notwithstanding it is observed that the claimed motifs seem to be very abundant. An initial search for SEQ ID NO 3, revealed a total of 4.780.389 hits in sequence databases such as REGISTRY of Chemical abstracts.

Given this enormous large number of documents relevant to the issue of novelty it is impossible to determine which parts of claims 1-15 may be said to define subject-matter for which protection might legitimately be sought. For these reasons, a meaningful search of the whole claimed subject-matter of claims 1-15 can not be carried out.

26 Claim 1 in essential parts of the definition lack clarity. Claim 1 seeks to define mRNA molecules comprising motifs which obviously are comprised in mRNA sequences of the prior art. Process features such as "modified" cannot have a limitation on the definition of such product claims, as for a given sequence it is not recognizable/testable whether such a claimed modification has taken place. This is in particular true for modifications which are as such not recognizable, such as e.g gene editing modifications.

Unclear expressions cannot be used to distinguish subject-matter for which protection is sought, and if having regard to the invention unclear terms are essential, they cannot be maintained in the claims.