(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 March 2026 (26.03.2026)

WIPO | PCT

(10) International Publication Number
WO 2026/062371 A2

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12N 15/67* (2006.01)

(21) International Application Number:
PCT/GB2025/052026

(22) International Filing Date:
15 September 2025 (15.09.2025)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
2413707.7 17 September 2024 (17.09.2024) GB

(71) Applicants: **CAMBRIDGE ENTERPRISE LIMITED** [GB/GB]; The Old Schools, Cambridge Cambridgeshire CB2 1TN (GB). **UNIVERSITY OF KENT** [GB/GB]; The Registry, Canterbury Kent CT2 7NZ (GB).

(72) Inventors: **DE LIMA HEDAYIOGLU, Fabio**; University of Kent, The Registry, Canterbury Kent CT2 7NZ (GB). **KURUSAMY, Sathishkumar**; University of Kent, The Registry, Canterbury Kent CT2 7NZ (GB). **MULRONEY, Thomas**; University of Cambridge, c/o Cambridge Enterprise Ltd, Hauser Forum, 3 Charles Babbage Road, Cambridge Cambridgeshire CB3 0GT (GB). **SMALES, Christopher, Mark**; University of Kent, The Registry, Canterbury Kent CT2 7NZ (GB). **THAVENTHIRAN, James**; University of Cambridge, c/o Cambridge Enterprise Ltd, Hauser Forum, 3 Charles Babbage Road, Cambridge Cambridgeshire CB3 0GT (GB). **VON DER HAAR, Tobias**; University of Kent, The Registry, Canterbury Kent CT2 7NZ (GB). **WILLIS, Anne**; University of Cambridge, c/o Cambridge Enterprise Ltd, Hauser Forum, 3 Charles Babbage Road, Cambridge Cambridgeshire CB3 0GT (GB).

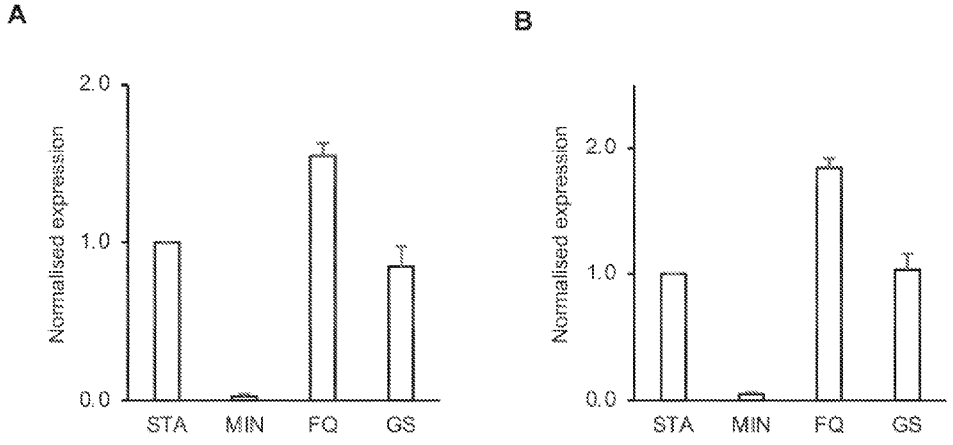
(74) Agent: **HGF LIMITED**; 4th Floor, 1 City Square, Leeds LS1 2ES (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UY, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MU, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: OPTIMISED RNAS

Figure 1



(57) Abstract: The present invention provides methods of producing RNAs optimised to have higher translation efficiency. Also provided are RNAs produced by such methods and their use as medicaments. Also provided are computer implemented methods for designing such RNAs.

WO 2026/062371 A2

[Continued on next page]

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

Optimised RNAs

Field of Invention

The present invention provides methods of producing RNAs optimised to have higher translation efficiency. Also provided are RNAs produced by such methods and their use as medicaments. Also provided are computer implemented methods for designing such RNAs.

Background

In vitro transcribed messenger RNA (IVT mRNA) has emerged as a promising avenue for developing novel therapeutic interventions. While IVT mRNA holds great potential for future therapies, efforts to enhance the production of full-length, active proteins are imperative for realising its clinical utility[1]. Researchers typically employ codon optimization strategies to enhance translation efficiency, recognizing the pivotal role of the open reading frame (ORF) in mRNA vaccine efficacy[2]. Additionally, optimizing mRNA secondary structure and stability, often achieved through increasing the GC mRNA content of the coding sequence, can bolster mRNA vaccine safety, efficacy, and stability[3]. However, achieving optimal translation efficiency and *in vivo* protein expression requires careful consideration of factors such as codon usage frequency and host tRNA availability, which vary across organisms and tissues[4]. Further optimization efforts must also address host cell-dependent factors, including tRNA complement and dynamics of tRNA processing by the ribosome, to optimize translation rates and mitigate transcript degradation pathways[5]. Despite advances in optimization algorithms, concerns regarding their effectiveness have been raised. Moreover, the presence of numerous "slow" codons in vaccine sequences, decoded by less abundant tRNAs, may contribute to elongation stalling[6]. Overall, the wider utility of mRNAs as tools and therapeutics will benefit from technology to increase or tailor mRNA translation efficiency and stability.

There is a need for improved RNAs, such as therapeutic mRNAs.

There is a need for improved methods of producing RNAs, such as therapeutic mRNAs.

There is a need for improved methods of designing RNAs, such as therapeutic mRNAs.

There is a need for improved methods and means for increasing translation efficiency and/or fidelity of mRNAs, such as therapeutic mRNAs.

Brief summary of the disclosure

Draft Claims

In a first aspect there is provided, a nucleic acid molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected based one or more properties of the codon, wherein the one or more properties comprise one or more of:

a level of tRNA isoacceptor expression in a target cell;

codon decoding speeds; and/or

binding affinity between the codon and cognate tRNA anticodon.

In certain embodiments, at least one codon has been selected based on the level of tRNA isoacceptor expression in a target cell.

In certain embodiments, all codons for one amino acid have been selected based on the one or more properties of the codon.

In certain embodiments, all codons for all amino acids have been selected based on the one or more properties of each codon.

In certain embodiments, the nucleic acid sequence is configured to increase translation efficiency and/or translation fidelity in the target cell.

In certain embodiments, the target cell comprises an in vivo or in vitro cell.

In certain embodiments, the target cell comprises a skeletal muscle cell and wherein the codon for:

alanine comprises the sequence 5’ GCU 3’;

arginine comprises the sequence 5’ AGG 3’;

aspartic acid comprises the sequence 5’ GAC 3’;

glutamic acid comprises the sequence 5’ GAG 3’;

glycine comprises the sequence 5’ GGC 3’;

leucine comprises the sequence 5’ CUU 3’;

lysine comprises the sequence 5’ AAG 3’;

proline comprises the sequence 5’ CCU 3’;

serine comprises the sequence 5’ AGC 3’;

asparagine comprises the sequence 5’ AAC 3’;

glutamine comprises the sequence 5' CAG 3';

histidine comprises the sequence 5' CAC 3';

valine comprises the sequence 5' GUG 3'; and/or

methionine comprises the sequence 5' AUG 3'; or

wherein the target cell comprises an HEK293T cell and wherein the codon for:

alanine comprises the sequence 5' GCU 3';

arginine comprises the sequence 5' CGU 3';

aspartic acid comprises the sequence 5' GAC 3';

glutamic acid comprises the sequence 5' CAA 3';

glycine comprises the sequence 5' GGC 3';

leucine comprises the sequence 5' CUG 3';

lysine comprises the sequence 5' AAG 3';

proline comprises the sequence 5' CCU 3';

serine comprises the sequence 5' AGC 3';

asparagine comprises the sequence 5' AAC 3';

glutamine comprises the sequence 5' CAG 3';

histidine comprises the sequence 5' CAC 3';

valine comprises the sequence 5' GUU 3';

isoleucine comprises the sequence 5' AUA 3';

threonine comprises the sequence 5' ACU 3';

tryptophan comprises the sequence 5' UGG 3';

tyrosine comprises the sequence 5' UAC 3';

phenylalanine comprises the sequence 5' UUC 3'; and/or

methionine comprises the sequence 5' AUG 3'

In a second aspect there is provided, a nucleic acid molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids and encoding a signal peptide selected from:

alpha-1-antichymotrypsin (AACT) signal peptide; or

Human Serum albumin (HSA) signal peptide.

In certain embodiments, the nucleic acid molecule is for translation in target cell, wherein the target cell is a muscle cell, liver cell, kidney cell, cardiomyocyte and/ or antigen presenting cell.

In certain embodiments, the target cell is a liver cell and the signal peptide comprises alpha-1-antichymotrypsin (AACT) signal peptide.

In certain embodiments, the target cell is a muscle cell and the signal peptide comprises human Serum albumin (HSA) signal peptide.

In certain embodiments, the alpha-1-antichymotrypsin (AACT) signal peptide comprises an amino acid sequence according to SEQ ID NO: 10 and/or wherein human Serum albumin (HSA) signal peptide comprises an amino acid sequence according to SEQ ID NO: 22.

In a third aspect, there is provided, a nucleic acid molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids in a target cell and encoding a 5' and 3' untranslated region (UTR):

wherein the 5' UTR is derived from a MYL1 mRNA and the 3' UTR is derived from a MYL1 mRNA (MUTR2);

wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR4);

wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a TNNC1 mRNA (MUTR5);

wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR6);

wherein the 5' UTR is derived from a MYH1 mRNA and the 3' UTR is derived from a MYH1 mRNA (MUTR7);

wherein the 5' UTR is derived from a MYL2 mRNA and the 3' UTR is derived from a MYL2 mRNA (MUTR8);

wherein the 5' UTR is derived from a CKM mRNA and the 3' UTR is derived from a CKM mRNA (MUTR10);

wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a COX6A2 mRNA (MUTR11);

wherein the 5' UTR is derived from a ORM1 mRNA and the 3' UTR is derived from a ORM1 mRNA (LUTR1);

wherein the 5' UTR is derived from a SAA2 mRNA and the 3' UTR is derived from a SAA2 mRNA (LUTR2);

wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR3);

wherein the 5' UTR is derived from a CFHR2 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR5);

wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a SAA1 mRNA (LUTR6);

wherein the 5' UTR is derived from a GC mRNA and the 3' UTR is derived from a GC mRNA (LUTR7);

wherein the 5' UTR is derived from a AHSG mRNA and the 3' UTR is derived from a AHSG mRNA (LUTR8);

wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a APOA1 mRNA (LUTR9); or

wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a GC mRNA (LUTR10).

In certain embodiments:

the 5' UTR comprises or consists of SEQ ID NO: 37 and the 3' UTR comprises or consists of SEQ ID NO: 38 (MUTR2);

the 5' UTR comprises or consists of SEQ ID NO: 41 and the 3' UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

the 5' UTR comprises or consists of SEQ ID NO: 43 and the 3' UTR comprises or consists of SEQ ID NO: 44 (MUTR5);

the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 46 (MUTR6);

the 5' UTR comprises or consists of SEQ ID NO: 47 and the 3' UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 49 and the 3' UTR comprises or consists of SEQ ID NO: 50 (MUTR8);

the 5' UTR comprises or consists of SEQ ID NO: 53 and the 3' UTR comprises or consists of SEQ ID NO: 54 (MUTR10);

the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

the 5' UTR comprises or consists of SEQ ID NO: 56 and the 3' UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

the 5' UTR comprises or consists of SEQ ID NO: 58 and the 3' UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

the 5' UTR comprises or consists of SEQ ID NO: 64 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 66 (LUTR6);

the 5' UTR comprises or consists of SEQ ID NO: 67 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 69 and the 3' UTR comprises or consists of SEQ ID NO: 70 (LUTR8);

the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 63 (LUTR9) or

the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR10).

In certain embodiments, the target cell is a muscle cell and:

the 5' UTR comprises or consists of SEQ ID NO: 37 and the 3' UTR comprises or consists of SEQ ID NO: 38 (MUTR2);

the 5' UTR comprises or consists of SEQ ID NO: 41 and the 3' UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 46 (MUTR6);

the 5' UTR comprises or consists of SEQ ID NO: 47 and the 3' UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 49 and the 3' UTR comprises or consists of SEQ ID NO: 50 (MUTR8);

the 5' UTR comprises or consists of SEQ ID NO: 53 and the 3' UTR comprises or consists of SEQ ID NO: 54 (MUTR10);

the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

the 5' UTR comprises or consists of SEQ ID NO: 56 and the 3' UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

the 5' UTR comprises or consists of SEQ ID NO: 58 and the 3' UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

the 5' UTR comprises or consists of SEQ ID NO: 64 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 66 (LUTR6);

the 5' UTR comprises or consists of SEQ ID NO: 67 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 69 and the 3' UTR comprises or consists of SEQ ID NO: 70 (LUTR8); or

the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR10).

In certain embodiments, the target cell is a liver cell and:

the 5' UTR comprises or consists of SEQ ID NO: 37 and the 3' UTR comprises or consists of SEQ ID NO: 38 (MUTR2);

the 5' UTR comprises or consists of SEQ ID NO: 41 and the 3' UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

the 5' UTR comprises or consists of SEQ ID NO: 43 and the 3' UTR comprises or consists of SEQ ID NO: 44 (MUTR5);

the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 46 (MUTR6);

the 5' UTR comprises or consists of SEQ ID NO: 47 and the 3' UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 49 and the 3' UTR comprises or consists of SEQ ID NO: 50 (MUTR8);

the 5' UTR comprises or consists of SEQ ID NO: 53 and the 3' UTR comprises or consists of SEQ ID NO: 54 (MUTR10);

the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

the 5' UTR comprises or consists of SEQ ID NO: 56 and the 3' UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

the 5' UTR comprises or consists of SEQ ID NO: 58 and the 3' UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

the 5' UTR comprises or consists of SEQ ID NO: 64 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

the 5' UTR comprises or consists of SEQ ID NO: 67 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 69 and the 3' UTR comprises or consists of SEQ ID NO: 70 (LUTR8); or

the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR10).

In certain embodiments, the target cell is an HEK293T cell and

the 5' UTR comprises or consists of SEQ ID NO: 41 and the 3' UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

the 5' UTR comprises or consists of SEQ ID NO: 47 and the 3' UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

the 5' UTR comprises or consists of SEQ ID NO: 56 and the 3' UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

the 5' UTR comprises or consists of SEQ ID NO: 58 and the 3' UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

the 5' UTR comprises or consists of SEQ ID NO: 64 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 66 (LUTR6);

the 5' UTR comprises or consists of SEQ ID NO: 67 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

the 5' UTR comprises or consists of SEQ ID NO: 69 and the 3' UTR comprises or consists of SEQ ID NO: 70 (LUTR8);

the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 63 (LUTR9); or

the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR10).

In a fourth aspect there is provided, a molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids in a target cell and comprising at least one alternative reading frame sequence which encodes an alternative translation product that differs from the translation product of in-frame translation of the nucleic acid, wherein the nucleic acid comprises a nucleic acid sequence encoding at least two consecutive stop codons (tandem stop codon) for reducing translation of the alternative translation products.

In certain embodiments, the at least two consecutive stop codons each comprise a nucleic acid sequence consisting of UAA or UAG.

In certain embodiments, the nucleic acid comprises two consecutive stop codons and wherein the two consecutive stop codons together comprise a nucleic acid sequence comprising or consisting of:

UAAUAA;

UAAUAG;

UAGUAA; or

UAGUAG.

In certain embodiments, the at least one tandem stop codon is encoded in a +1 reading frame from the translation product of in-frame translation.

In certain embodiments, the nucleic acid comprises an RNA and wherein at least one tandem stop codon comprise at least one N1-methyl pseudouridine.

In certain embodiments of any aspect, the nucleic acid molecule comprises an mRNA or a DNA molecule encoding an mRNA.

In a fifth aspect there is provided, a nucleic acid molecule comprising a nucleic acid sequence encoding any one or more of the first to fourth aspects in any combination.

In a sixth aspect there is provided, an in vitro transcribed mRNA comprising a nucleic acid sequence encoding any one or more of the first to fourth aspects in any combination.

In certain embodiments, the nucleic acid molecule is an mRNA or encodes an mRNA, or the in vitro transcribed mRNA, and wherein the mRNA comprises a therapeutic mRNA.

In certain embodiments, the nucleic acid molecule or in vitro transcribed mRNA comprises a nucleic acid sequence encoding a poly(A) tail wherein the poly(A) tail comprises 97 to 135 adenine residues; optionally comprises 97 adenine residues.

In certain embodiments, the nucleic acid molecule is an mRNA, or the in vitro transcribed mRNA and the mRNA comprises a m7G-5'ppp5'GpG 5' cap.

In certain embodiments, the nucleic acid molecule is an mRNA, or the in vitro transcribed mRNA, and the mRNA comprises one or more modified ribonucleotide.

In a seventh aspect there is provided, a method of producing an optimised nucleic acid molecule for translation, the method comprising;

determining preferred codons in a target cell for at least one amino acid;

producing a nucleic acid molecule that comprises a nucleic acid sequence encoding one or more codons for translation to one or more amino acids, wherein at least one codon for the first amino acid comprises the preferred codon;

wherein determining preferred codons comprises:

determining a level of tRNA isoacceptor expression in a target cell for at least one first amino acid;

determining codon decoding speed for at least one first amino acid; and/or

determining binding affinity between a codon and cognate tRNA anticodon for at least one first amino acid.

In certain embodiments, determining a level of tRNA isoacceptor expression comprises:

determining the most abundantly expressed tRNA anti-codon for the at least one first amino acid and wherein the preferred codon comprises the complement of the most abundantly expressed tRNA anti-codon; or

determining the ratio of the level of expression of cognate to near-cognate tRNAs for the at least one first amino acid and wherein the preferred codon comprises the complement of the anti-codon with the lowest ratio of near-cognate to cognate tRNA.

In certain embodiments, the method further comprises determining the preferred codon for one or more further amino acids.

In certain embodiments, the nucleic acid molecule comprises a nucleic acid sequence encoding at least one codon for the each of the one or more further amino acids, wherein the at least one codon for each further amino acid comprises the preferred codon for each further amino acid.

In certain embodiments, all codons for each of the at least one first and/or further amino acids comprise the preferred codon.

In certain embodiments, the nucleic acid molecule comprises an mRNA or a DNA molecule encoding an mRNA.

In certain embodiments, the producing further comprises in vitro transcribing the DNA molecule to produce the mRNA.

In certain embodiments, the mRNA comprises a therapeutic mRNA.

In certain embodiments, the producing further comprises adding a nucleic acid sequence encoding a poly(A) tail; optionally wherein the poly(A) tail comprises 97 to 135 adenine residues; further optionally wherein the poly(A) tail comprises 97 adenine residues.

In certain embodiments, the method comprises enzymatically adding a or the poly(A) tail optionally, as described herein.

In certain embodiments, the producing further comprises adding a m7G-5'ppp5'GpG 5' cap.

In certain embodiments, the mRNA comprises one or more modified ribonucleotide.

In an eighth aspect there is provided, use of a nucleic acid molecule encoding a 5' and 3' UTR:

wherein the 5' UTR is derived from a MYLPF mRNA and the 3' UTR is derived from a MYLPF mRNA (MUTR2);

wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR4);

wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR6);

wherein the 5' UTR is derived from a MYH1 mRNA and the 3' UTR is derived from a MYH1 mRNA (MUTR7);

wherein the 5' UTR is derived from a MYL2 mRNA and the 3' UTR is derived from a MYL2 mRNA (MUTR8);

wherein the 5' UTR is derived from a CKM mRNA and the 3' UTR is derived from a CKM mRNA (MUTR10);

wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a COX6A2 mRNA (MUTR11);

wherein the 5' UTR is derived from a ORM1 mRNA and the 3' UTR is derived from a ORM1 mRNA (LUTR1);

wherein the 5' UTR is derived from a SAA2 mRNA and the 3' UTR is derived from a SAA2 mRNA (LUTR2);

wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR3);

wherein the 5' UTR is derived from a CFHR2 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR5);

wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a SAA1 mRNA (LUTR6);

wherein the 5' UTR is derived from a GC mRNA and the 3' UTR is derived from a GC mRNA (LUTR7);

wherein the 5' UTR is derived from a AHSG mRNA and the 3' UTR is derived from a AHSG mRNA (LUTR8); or

wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a GC mRNA (LUTR10);

for translation of an mRNA in a muscle cell.

In a ninth aspect there is provided use of a nucleic acid molecule encoding a UTR:

wherein the 5' UTR is derived from a MYLPF mRNA and the 3' UTR is derived from a MYLPF mRNA (MUTR2);

wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR4);

wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a TNNC1 mRNA (MUTR5);

wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR6);

wherein the 5' UTR is derived from a MYH1 mRNA and the 3' UTR is derived from a MYH1 mRNA (MUTR7);

wherein the 5' UTR is derived from a MYL2 mRNA and the 3' UTR is derived from a MYL2 mRNA (MUTR8);

wherein the 5' UTR is derived from a CKM mRNA and the 3' UTR is derived from a CKM mRNA (MUTR10);

wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a COX6A2 mRNA (MUTR11);

wherein the 5' UTR is derived from a ORM1 mRNA and the 3' UTR is derived from a ORM1 mRNA (LUTR1);

wherein the 5' UTR is derived from a SAA2 mRNA and the 3' UTR is derived from a SAA2 mRNA (LUTR2);

wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR3);

wherein the 5' UTR is derived from a CFHR2 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR5);

wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a SAA1 mRNA (LUTR6);

wherein the 5' UTR is derived from a GC mRNA and the 3' UTR is derived from a GC mRNA (LUTR7);

wherein the 5' UTR is derived from a AHSG mRNA and the 3' UTR is derived from a AHSG mRNA (LUTR8); or

wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a GC mRNA (LUTR10;

for translation of an mRNA in a liver cell.

In a tenth aspect there is provided, use of an alpha-1-antichymotrypsin (AACT) signal peptide for translation of an mRNA in a liver cell.

In an eleventh aspect there is provided, use of a human Serum albumin (HSA) signal peptide for translation of an mRNA in a muscle cell.

In certain embodiments, the nucleic acid sequence encoding:

the human Serum albumin (HSA) signal peptide comprises or consists of a nucleic acid sequence according to SEQ ID NO: 23 or 24; or

the alpha-1-antichymotrypsin (AACT) signal peptide comprises or consists of a nucleic acid sequence according to SEQ ID NO: 11 or 12.

In a twelfth aspect there is provided, a nucleic acid molecule or in vitro transcribed mRNA of as described herein or a pharmaceutical composition thereof for use as a medicament.

In a thirteenth aspect there is provided, a nucleic acid molecule or in vitro transcribed mRNA of as described herein or a pharmaceutical composition thereof for use as a vaccine.

In a fourteenth aspect there is provided, a nucleic acid molecule or in vitro transcribed mRNA of as described herein or a pharmaceutical composition thereof for use in inducing an immune response in a subject.

In a fifteenth aspect there is provided, a method of treating or preventing a disease in a subject in need thereof, comprising administering a nucleic acid molecule or in vitro transcribed mRNA as described herein or a pharmaceutical composition thereof to the subject.

In a sixteenth aspect there is provided, a method of vaccinating a subject in need thereof, comprising administering a nucleic acid molecule or in vitro transcribed mRNA as described herein or a pharmaceutical composition thereof to the subject.

In a seventeenth aspect there is provided, a method of inducing an immune response in a subject in need thereof, comprising administering a nucleic acid molecule or in vitro transcribed mRNA as described herein or a pharmaceutical composition thereof to the subject.

In an eighteenth aspect there is provided, a computer-implemented method for designing an optimised RNA sequence for translation, the method comprising:

(a) receiving a target translated protein sequence, one or more user-defined objectives for the optimised RNA sequence, and one or more parameters associated with each user-defined objective and corresponding to the optimised RNA sequence, the one or more parameters including:

codon composition efficiency for translation of the RNA sequence to the target translated protein sequence;

presence or absence of specified nucleotide repeats;

GC content;

desirability of specified nucleotide sub-sequences; and

location of one or more stop codons;

(b) receiving deviation information for each parameter, wherein the deviation information is for processing a nucleotide sequence and is indicative of a degree of deviation of a nucleotide sequence from a target value of the associated parameter;

(c) determining a score function for each parameter based on the deviation information, wherein each score function is usable for processing a nucleotide sequence to generate a score indicative of the degree of deviation between the nucleotide sequence and the target value of the parameter associated with that score function;

(d) selecting, from amongst a plurality of nucleotide sequences, a set of random nucleotide sequences, wherein each random nucleotide sequence comprises a nucleotide sequence that encodes the target translated protein sequence;

(e) generating one or more parameter scores for each random nucleotide sequence based on the score functions determined for the corresponding one or more parameters;

(f) generating a total score for each random nucleotide sequence based on the one or more parameter scores generated for that random nucleotide sequence;

(g) assigning a rank to each random nucleotide sequence in the set of random nucleotide sequences based on the generated total scores of the random nucleotide sequences from a lowest rank to a highest rank; and

removing, from the plurality of nucleotide sequences, one or more of the lowest ranked random nucleotide sequences based on the assigned ranking of each random nucleotide sequence, to provide a reduced plurality of nucleotide sequences;

(h) performing a plurality of iterative steps including repeating steps (d) to (g),wherein the plurality of nucleotide sequences in step (d) in an iterative step corresponds to the reduced plurality of nucleotide sequences of step (g) in the previous iterative step;

(i) wherein step (h) is performed until a stop condition is met to provide a set of optimised RNA sequences, wherein each optimised RNA sequence has a total score above a score threshold;

wherein each optimised RNA sequence comprises at least one different sequence property.

In certain embodiments, the random nucleotide sequence assigned with the lowest rank has a total score indicative of the one or more parameter scores having the greatest degree of deviation from the target values associated with the respective one or more parameters and the random nucleotide sequence assigned with the highest rank has a total score indicative of the one or more parameter scores having the smallest degree of deviation from the target values associated with the respective one or more parameters.

In certain embodiments, the method further comprises identifying the highest ranked random nucleotide sequence for the set of random nucleotide sequences in a first iterative step and for the set of random nucleotide sequences in a second iterative step, the first and second iterative steps among the plurality of iterative steps; and

determining a difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step;

comparing the difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step with a difference threshold,

wherein the stop condition comprises one or more of:

a predetermined number of iterations;

based on the comparison, the difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step being below the difference threshold; and

at least one selected random nucleotide sequence has a total score above an optimisation threshold.

In certain embodiments, the user defined objectives comprise one or more of target application, biological activity, expression profile, manufacturability, secretion profile and/or storability.

In certain embodiments, the method further comprises:

selecting one or more sequence elements from a database based on the one or more user defined objectives and the one or more parameters; optionally

wherein the sequence elements comprise one or more of a 5'-UTR, a 3'-UTR, a poly(A) tail length, signal peptide sequence, an aptamer sequence, protein binding sequences, tandem stop codon and/or nucleic acid binding sequences.

combining the selected sequence elements with each optimised RNA sequence to provide a set of full-length mRNA sequences; and

outputting a signal indicative of a set of optimised full-length mRNA sequences,

wherein the set of optimised full-length mRNA sequences comprises full-length mRNA sequences which comprise at least one optimised RNA sequence which meets the score threshold of step (i) above.

In certain embodiments, the set of optimised full-length mRNA sequences comprises full-length mRNA sequences which comprise an optimised RNA sequence which has a total score below the score threshold of step (i), the method further comprises

repeating steps (d) to (i) above for the optimised full-length mRNA sequences; and

outputting a signal indicative of a set of optimised full-length mRNA sequences, comprising the optimised full-length mRNA sequence identified as having the highest total score.

In certain embodiments, the:

target application comprises one or more of: administration route, end user purpose, target tissue type, target cell type, cellular localisation, and/or cellular processing ;

codon composition efficiency is determined using one or more of: codon decoding times, codon preference, codon usage frequency, codon usage patterns, abundance of tRNA isoacceptors and/or ratio of cognate to near-cognate tRNAs;

specified nucleotide sub-sequences comprise: frameshifting sequences, ribosomal slippery sequences, transcription terminator sequences, RNA polymerase stalling sequences, aptamer sequences, secondary structure forming sequences, restriction enzyme sites, RNA binding protein binding sites and/or frameshifted premature termination codons;

biological activity comprises one or more of: immunogenicity, cellular interactions, adjuvant activity, enzymatic activity and/or cellular effects;

expression profile comprises expression duration, expression location and/or expression level;

manufacturability comprises yield of encoded protein;

secretion profile comprises localisation of the RNA and/or encoded protein in a target cell; and/or

storability comprises stability of the RNA at predetermined environmental conditions.

In certain embodiments,

step (g) further comprises combining n nucleotides of the 5' end of at least one first removed nucleotide sequence with L – n nucleotides of the 3' end of a second removed nucleotide sequence to provide a mixed sequence; and/or

randomly mutating at least one of the removed nucleotide sequences to generate a randomly mutated sequence, carrying out steps (e) to (g) for each randomly mutated sequence and selecting at least one optimised randomly mutated sequence based on the score threshold.

In certain embodiments,

in the plurality of iterative steps of step (h), the set of random nucleotide sequences in a given iterative step is selected to include the non-removed nucleotide sequences present from the set of random nucleotide sequences in the previous iterative step and one or more nucleotide sequences that were not included in the set of random nucleotide sequences in the previous iterative step,

wherein the set of random nucleotide sequences in the given iterative step further comprises one or more mixed sequences and/or optimised randomly mutated sequences.

In some embodiments, the set of random nucleotide sequences comprises one or more seeding nucleotide sequences, wherein each seeding nucleotide sequence comprises a nucleotide sequence that has a predetermined parameter score for one of the one or more parameters and encodes the target translated protein sequence.

In a nineteenth aspect there is provided, computer readable instructions which, when executed by one or more processors, cause the one or more processors to perform the method of the eighteenth aspect.

In a twentieth aspect there is provided, an apparatus for designing an optimised RNA sequence for translation, the apparatus comprising:

a memory arranged to store machine-readable instructions;

an input unit arranged to receive an input; and

processing circuitry arranged to operably execute the stored machine-readable instructions to:

(a) receive, via the input unit, a target translated protein sequence, one or more user-defined objectives for the optimised RNA sequence, and one or more parameters associated with each user-defined objective and corresponding to the target translated protein sequence, the one or more parameters including:

codon composition efficiency for translation of the RNA sequence to the target translated protein sequence;

presence or absence of specified nucleotide repeats;

GC content; and

desirability of specified nucleotide sub-sequences;

(b) receive, via the input unit, deviation information for each parameter, the deviation information for processing a nucleotide sequence and indicative of a degree of deviation of a nucleotide sequence from a target value of the associated parameter;

(c) determine a score function for each parameter based on the deviation information, wherein each score function is usable for processing a nucleotide sequence to generate a score indicative of a degree of deviation between the nucleotide sequence and the target value of the parameter associated with that score function;

(d) select, from amongst a plurality of nucleotide sequences, a set of random nucleotide sequences, wherein each random nucleotide sequence comprises a nucleotide sequence that encodes the target translated protein sequence;

(e) generate one or more parameter scores for each random nucleotide sequence based on the score functions determined for the corresponding one or more parameters;

(f) generate a total score for each random nucleotide sequence based on the one or more parameter scores generated for that random nucleotide sequence;

(g) assign a rank to each random nucleotide sequence in the set of random nucleotide sequences based on the generated total scores of the random nucleotide sequences from a lowest rank to a highest rank; and

remove, from the plurality of nucleotide sequences, one or more of the lowest ranked random nucleotide sequences based on the assigned ranking of each random nucleotide sequence, to provide a reduced plurality of nucleotide sequences;

(h) perform a plurality of iterative steps including repeating steps (d) to (g), wherein the plurality of nucleotide sequences in step (d) in an iterative step corresponds to the reduced plurality of nucleotide sequences of step (g) in the previous iterative step,

(i) wherein step (h) is repeated until a stop condition is met to provide a set of optimised RNA sequences, wherein each optimised RNA sequence has a total score above a score threshold.

wherein each optimised RNA sequence comprises at least one different sequence property.

In a twenty-first aspect there is provided, a system for designing an optimised RNA sequence for translation, the system comprising:

the apparatus of the twentieth aspect; and

a second apparatus comprising a database including one or more sequence elements, wherein the apparatus of the twentieth aspect is configured to retrieve the one or more sequence elements from the second apparatus for providing a set of full-length mRNA sequences.

In some embodiments, the set of random nucleotide sequences comprises one or more seeding nucleotide sequences wherein each seeding nucleotide sequence comprises a nucleotide sequence that has a predetermined parameter score for one of the one or more parameters and encodes the target translated protein sequence.

Throughout the description and claims of this specification, the words “comprise” and “contain” and variations of them mean “including but not limited to”, and they are not intended to (and do not) exclude other moieties, additives, components, integers or steps.

Throughout the description and claims of this specification, the singular encompasses the plural unless the context otherwise requires. In particular, where the indefinite article is used, the specification is to be understood as contemplating plurality as well as singularity, unless the context requires otherwise.

Features, integers, characteristics, compounds, chemical moieties or groups described in conjunction with a particular aspect, embodiment or example of the invention are to be understood to be applicable to any other aspect, embodiment or example described herein unless incompatible therewith.

Various aspects of the invention are described in further detail below.

Brief description of the Figures

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

Figure 1 shows FQ (FQ RNA corresponds to an RNA where codons have been selected based on near-cognate:cognate ratios using methods described herein) coding sequence variant increases nluc mRNA translation efficiency. STA, FQ, MIN, or GS nLuc mRNAs were transfected into HEK293T cells and translation efficiency assessed by nLuc protein synthesis (n=3 replicated experiments). **A** shows data for unmodified mRNAs and **B**, (N)1-methylpseudouridylated mRNAs. Data are mean +SE secreted luciferase activity in cell culture medium normalised to STA nLuc levels.

Figure 2 shows mRNA coding sequence optimisation increases mRNA vaccination efficacy in mice. **A**, Schematic showing mRNA construct design containing the receptor binding domain of SARS-CoV-2, the influenza HA transmembrane domain and an MHC class I epitope (SIINFEKL – SEQ ID NO: 72). **B**, Neutralising antibody titres at 50% inhibition (NT50) against wildtype SARS-CoV-2 1 week after immunisation with a 20 µg dose of ST-RBD mRNA or FQ-RBD mRNA. **C**, Representative plots (left) and summary data (right) showing production of IFNγ produced by CD8+ splenocytes from mice vaccinated with 20 µg mRNA in response to SIINFEKL (SEQ ID NO: 72) stimulation. **D**, Neutralising antibody titres at 50% inhibition (NT50) against wildtype SARS-CoV-2 1 week after final immunisation from mice immunised with 2 doses of 4 µg ST-RBD mRNA or FQ-RBD mRNA 3 weeks apart. **E**, Representative plots (left) and summary data (right) showing production of IFNγ produced by CD8+ splenocytes from mice vaccinated with 2 doses of 4 µg mRNA 3 weeks apart in response to SIINFEKL (SEQ ID NO: 72) stimulation. Statistically significant differences between groups were determined by one-way ANOVA.

Figure 3 shows isoacceptor tRNA expression in target cells. **A**, HEK293T or **B**, primary human skeletal muscle cells. Data are transcripts per million for aggregated isodecoder tRNA read counts.

Figure 4 shows isoacceptor tRNA-based sequence optimisation increases nluc mRNA translation efficiency**. A**, Coding sequence-optimised nLuc mRNA (ISOpt) or GS nLuc mRNA were transfected into HEK293T cells and translation efficiency assessed by nLuc protein synthesis (n=3 replicated experiments). **B**, An nLuc mRNA optimised for expression in SKMC cells (SKMC-Opt, n=4), or control nLuc mRNA (GS, n=3), were transfected into primary SKMCs and translation efficiency assessed by nLuc protein synthesis. Data are mean secreted luciferase activity in cell culture medium normalised to GS nLuc levels (**A**), or raw RLU (**B**). '*', $P<0.05$, Welch's unpaired one-tailed T-test.

Figure 5 shows effects of ER signal peptide on nLuc mRNA expression in Huh7, HEK293T, and SKMCs. nLuc mRNAs encoding different signal peptide sequences were transfected into **A**, Huh7, **B**, HEK293T, or **C**, SKMCs and expression assessed by nluc protein synthesis (n=6 replicated experiments). Data are mean+SE secreted luciferase activity in cell culture medium normalised to NT nLuc levels.

Figure 6 shows impact of specific cell-derived UTRs on nLuc expression from *IVT*mRNAs in different cell types. nLuc mRNAs containing different skeletal muscle-origin untranslated region (UTR) sequences were transfected into **A**, SKMCs, **B**, HEK293T, or **C**, HuH7s. nLuc mRNAs containing different liver-origin UTR sequences were transfected into **D**, Huh7, **E**, HEK293T, or **F**, SKMCs. **A**-**F**, mRNA translation efficiency was assessed by nluc protein synthesis (n=3 replicated experiments). Data are mean+SE secreted luciferase activity in cell culture medium normalised to control nLuc levels (CON).

Figure 7 shows Poly(A) tail length influences nLuc protein expression from IVTmRNAs. **A**, Estimated polyA tail length for E.coli poly(A) polymerase units (U) under standard reaction conditions. **B**, nLuc mRNAs were polyadenylated using the indicated units of E.coli poly(A) polymerase, purified, were transfected into HEK293T cells, and expression assessed by nluc protein synthesis. Data are mean+SE secreted luciferase activity in cell culture medium normalised to 0 U poly(A) polymerase nLuc levels. **C**, nLuc mRNAs were synthesised containing either m7G-5'ppp5'GpG or m7G5'ppp5'2'OMeApG, purified, transfected into SKMCs, and expression assessed by nluc protein synthesis (n=4 replicated experiments). Data are mean secreted luciferase activity in cell culture medium after 24 hours. '*', $P<0.05$, Welch's unpaired one-tailed T-test.

Figure 8 shows overview of sequence optimisation method. **A**, Examples of possible functions relating sequence parameters to scores. In each case the user-defined objective is to meet the parameter value (for example, GC content) indicated by the red line, at which point the score is 1. The different functions define how quickly movement away from the target parameter set by the objective is penalised. **B**, Flow chart illustrating key processes in assembling high-performing sequences for mRNA sequence optimisation.

Figure 9 shows a method 900 of designing an optimised RNA sequence for translation according to embodiments of the disclosure.

Figure 10 shows an apparatus 1002 for designing an optimised RNA sequence for translation in accordance with embodiments of the disclosure.

Figure 11 shows affinity measurements can identify preferred codons. **A**) RNA oligonucleotides containing specific anticodon (Top line) and codon (bottom line) sequences coupled to anchor nucleotides (grey) were annealed (**UCAGUCGA**UAI (SEQ ID NO: 83) –

bold underline = anchor), and the affinity between codons and anticodons determined using temperature-dependent CD spectroscopy. **B**) affinity measurements for modified and unmodified AUU (Ile) codons and cognate (UAI) and near-cognate (GAA) tRNA anticodons. The uridine modification, Ψ, reduces the difference in binding affinity between the cognate and near-cognate anticodon with an AUU codon, indicating that this codon is less preferred in modified sequences compared to unmodified ones. C, expression levels of a model sequence (nanoLuciferase) derived from modified or unmodified, *in vitro* transcribed RNAs transfected into HEK293 cells. The RNA sequences are identical except for the use of isoleucine encoding codons, which are either AUU or AUC. The results demonstrate that codon:anticodon affinity measurements can be used to identify preferred codons for use in sequence design strategies.

Figure 12 shows UGA stop codon is sufficient to prevent full protein expression but not to prevent T cell antigen presentation. (a) Distribution of stop codons encoded in the +1 frame of SARS-CoV-2 viral spike sequence (top) and Pfizer BioNTech mRNA vaccine spike sequence (bottom). (b) Schematic showing FLuc and FLuc U*GA mRNA constructs and how they are translated in frame or in the +1 frame. (c) Western blot (anti-Flag epitope) of lysate of untransfected cells, cells transfected with FLuc mRNA, or cells transfected with FLuc U*GA mRNA. Asterisk identifies a band at the 65.5 KDa the molecular weight of full length FLuc protein. Blot was also stained for beta actin as a loading control. (d) Luciferase activity from cell lysates of untransfected cells, cells transfected with FLuc or FLuc U*GA constructs. FLuc versus FLuc U*GA, P<0.0001 (one-way ANOVA) (e) Proportion of CD8+ OT-I T cells activated (CD69+) when incubated with untransfected cells or cells transfected with FLuc or FLuc U*GA modified mRNA. (f) Proportion of SIINFEKL (SEQ ID NO: 72) Multimer+ CD8+ splenocytes of mice that were immunised with FLuc mRNA or FLuc U*GA mRNA. FLuc versus FLuc U*GA, P=0.0004 (one-way ANOVA) (g) Proportion of SIINFEKL (SEQ ID NO: 72) Multimer+ T cells identified in spleen of mice immunised with FLuc mRNA or FLuc-mRNA, that are terminally differentiated $KLRG1^{+}CD127^{-}$. Representative plots are shown on the left and summary data is shown on the right. (h) IFN-γ production of splenocytes when incubated with or without SIINFEKL (SEQ ID NO: 72). Representative are shown on the left and summary statistics of median fluorescence intensity of IFN-γ+ cells on the right. (i) Peripheral blood mononuclear cells (PBMC) IFN-γ ELISpot responses from human donors vaccinated with AstraZeneca ChAdOx1 nCoV-19 or Pfizer BNT162b2 stimulated with peptides covering the Pfizer 3' UTR sequence.

Figure 13 shows **an vitro system for antigen evaluation from modified mRNA therapeutics.** (a) HEK293T cells and sorted HEK293T cells transduced with an $H\text{-}2K^{b}$ expressing plasmid stained with an anti-$H\text{-}2K^{b}$ antibody. Representative histograms are

shown on the left and summary data is shown on the right. HEK293T versus HEK293T H-2K^b, *P*<0.0001 (unpaired t test) (b) H-2K^b transduced HEK293T cells incubated with or without SIINFEKL (SEQ ID NO: 72) stained with the 25-D1.15 antibody clone specific for the SIINFEKL (SEQ ID NO: 72) presented on H-2K^b. Representative histograms are shown on the left and summary data is shown on the right. Untreated versus +SIINFEKL (SEQ ID NO: 72), *P*<0.0001 (unpaired t test) (c) H-2Kb transduced HEK293T cells transfected with the listed modified mRNA constructs stained with the 25-D1.16 antibody. Representative histograms are shown on the left and summary data is shown on the right. No SIINFEKL (SEQ ID NO: 72) versus In frame SIINFEKL (SEQ ID NO: 72), *P*<0.0001; No SIINFEKL (SEQ ID NO: 72) versus 3x In frame SIINFEKL (SEQ ID NO: 72), *P*<0.0001, In frame SIINFEKL (SEQ ID NO: 72) versus 3x In frame SIINFEKL (SEQ ID NO: 72), *P*<0.0001 (one-way ANOVA). (d) Schematic of the co-culture system in which OT-I cells are layered 24 hours after transfection of H-2K^b transduced HEK293T cells with modified mRNA constructs. Intended (in frame) and mistranslated (+1 frame and 3'UTR) T cell antigens can be detected by measuring CD69 upregulation on OT-I CD8+ T cells (e) Proportion of CD8+ OT1 T cells activated (CD69+) when incubated with untransfected cells or cells transfected with listed modified mRNA constructs. Representative plots are shown on the left and summary data is shown on the right. No SIINFEKL (SEQ ID NO: 72) versus In frame SIINFEKL (SEQ ID NO: 72), *P*<0.0001; No SIINFEKL (SEQ ID NO: 72) versus 3x In frame SIINFEKL (SEQ ID NO: 72), *P*<0.0001, In frame SIINFEKL (SEQ ID NO: 72) versus 3x In frame SIINFEKL (SEQ ID NO: 72), *P*<0.0001 (one-way ANOVA).

Figure 14 shows CD8+ T cells specific for +1 encoded antigen can limit repetitive doses. Proportion of SIINFEKL (SEQ ID NO: 72) Multimer+ CD8+ T cells in the draining lymph node (a) or muscle (b), taken from mice 24 hours or 48 hours after vaccination with modified mRNA COVID-19 vaccines with either SIINFEKL (SEQ ID NO: 72) encoded in the +1 frame (+1 SIINFEKL (SEQ ID NO: 72)) or not encoded (No SIINFEKL (SEQ ID NO: 72)). dLN No SIINFEKL (SEQ ID NO: 72) (48h) versus +1 SIINFEKL (SEQ ID NO: 72) (48h), *P*=0.0256; muscle +1 SIINFEKL (SEQ ID NO: 72) (24h) vs +1 SIINFEKL (SEQ ID NO: 72) (48h), *P*=0.0398; No SIINFEKL (SEQ ID NO: 72) (48h) versus +1 SIINFEKL (SEQ ID NO: 72) (48h), *P*=0.0003 (one-way ANOVA). (c) Comparative phenotype of SIINFEKL (SEQ ID NO: 72) Multimer+ cells in the draining lymph node and muscle from mice 24 hours or 48 hours after vaccination with +1 SIINFEKL (SEQ ID NO: 72) or No SIINFEKL (SEQ ID NO: 72) constructs. Draining lymph node is shown in orange and muscle is shown in purple. Representative plots are shown on the left and data from individual mice is displayed on the right. CD127+KLRG1- 24h dLN versus muscle, *P*=0.02; CD127+KLRG1- 48h dLN versus muscle, *P*=0.0027; CD127-KLRG1+ 24h dLN versus muscle, *P*=0.0201; CD127-KLRG1+

48h dLN versus muscle, *P*=0.0029 (paired t test). (d) Immunofluorescent staining of FFPE sections of muscle taken 48 hours after immunisation, with either No SIINFEKL (SEQ ID NO: 72) (top) or +1 SIINFEKL (SEQ ID NO: 72) (bottom). (e) RBD titre 21 days after transfer of OT-I or WT effector T cells and immunisation with either No SIINFEKL (SEQ ID NO: 72) or +1 SIINFEKL (SEQ ID NO: 72) constructs. No SIINFEKL (SEQ ID NO: 72) (+OT-I transfer) versus +1 SIINFEKL (SEQ ID NO: 72) (+OT-I transfer), *P*=0.0307; +1 SIINFEKL (SEQ ID NO: 72) (+OT-I transfer) versus +1 SIINFEKL (SEQ ID NO: 72) (+WT transfer), *P*=0.0254; +1 SIINFEKL (SEQ ID NO: 72) (+OT-I transfer) versus No SIINFEKL (SEQ ID NO: 72) (WT transfer), *P*=0.0033 (one-way ANOVA) (f) SARS-CoV-2 live virus neutralisation titre from mice in receipt of 3 weekly doses of either No SIINFEKL (SEQ ID NO: 72) or +1 SIINFEKL (SEQ ID NO: 72) COVID-19 vaccines. Day 42 No SIINFEKL (SEQ ID NO: 72) versus +1 SIINFEKL (SEQ ID NO: 72), *P*= 0.0266; day 63 No SIINFEKL (SEQ ID NO: 72) versus +1 SIINFEKL (SEQ ID NO: 72), *P*=0.00966 (Mann-Whitney tests).

Figure 15 shows strategies to prevent the expression of antigens encoded in the +1 frame. (a) Proportion of CD8+ OT-I T cells activated when incubated with cells transfected with modified mRNA COVID-19 vaccines with either SIINFEKL not encoded (No SIINFEKL), encoded in the +1 frame (+1 SIINFEKL), or with an equivalent to the U*208C synonymous mutation previously reported to remove the slippery site[1] (U*925C). (b) Proportion of CD8+ OT1 T cells activated when incubated with cells transfected with modified mRNA COVID-19 vaccines with SIINFEKL encoded in frame immediately after in frame U*GA, U*AG or U*AA stop codons. U*GA versus U*AA, *P*<0.0001; U*GA versus U*AG, P<0.0001 (one-way ANOVA) (c) Proportion of CD8+ OT1 T cells activated when incubated with cells transfected with modified mRNA COVID-19 vaccines with +1 frame encoded SIINFEKL immediately after +1 frame encoded single or tandem U*AG or U*AA stop codons. U*AA versus U*AAU*AA, P=0.0289; U*AA versus U*AG, P=0.0469; U*AA versus U*AGU*AG, P=0.0026 (one-way ANOVA). (d) Proportion of CD8+ T cells IFN-γ+ after SIINFEKL restimulation of splenocytes from mice immunised with modified mRNA COVID-19 vaccines with +1 frame encoded SIINFEKL immediately after +1 frame encoded single or tandem U*AA stop codons compared to the No SIINFEKL vaccine. Representative plots are shown on the left and summary data is shown on the right. +1 SIINFEKL versus U*AA, *P*=0.0174; +1 SIINFEKL versus U*AAU*AA, *P*=0.0004 (one-way ANOVA).

Figure 16 shows a modified mRNA COVID-19 vaccine without evidence of cryptic antigen expression encoded in either the +1 frame or 3' UTR. (a) Schematic showing modified mRNA COVID-19 vaccines with SIINFEKL encoded in the +1 frame (+1 SIINFEKL), optimised with the insertion of multiple synonymous changes leading to +1 encoding of single or tandem U*AA, or U*AG stops with SIINFEKL encoded in the +1 frame (Optimised

+1 SIINFEKL) or optimised with the insertion of multiple synonymous changes leading to +1 encoding of single or tandem U*AA, or U*AG stops without SIINFEKL encoded (Optimised). (b) Proportion of CD8+ OT-I T cells activated when incubated with cells transfected with PfizerRBD+SIINFEKL, Optimised+SIINFEKL or Optimised modified mRNA COVID-19 vaccines. PfizerRBD+SIINFEKL versus Optimised+SIINFEKL, $P<0.0001$; PfizerRBD+SIINFEKL versus Optimised, $P<0.0001$ (one-way ANOVA). Proportion of CD8+ T cells (c) SIINFEKL Multimer+ in the muscle or spleen or (d) IFN-γ+ after restimulation of splenocytes with ± SIINFEKL from mice immunised with PfizerRBD+SIINFEKL, Optimised+SIINFEKL or Optimised modified mRNA COVID-19 vaccines. Muscle PfizerRBD+SIINFEKL versus Optimised+SIINFEKL, $P=0.0235$; PfizerRBD+SIINFEKL versus Optimised, $P=0.0341$; Spleen PfizerRBD+SIINFEKL versus Optimised+SIINFEKL, $P=0<0.0001$; PfizerRBD+SIINFEKL versus Optimised, $P<0.0001$ (one-way ANOVA) (e) Proportion of CD8+ T cells expressing activation markers in an activation induced marker (AIM) assessment after stimulation with an overlapping peptide pool covering the RBD in splenocytes from mice immunised with PfizerRBD+SIINFEKL, Optimised+SIINFEKL or Optimised modified mRNA COVID-19 vaccines. %CD25+ T cells PfizerRBD+SIINFEKL versus Optimised+SIINFEKL, $P=0.0007$; PfizerRBD+SIINFEKL versus Optimised, $P=0.0084$; % CD69+ T cells PfizerRBD+SIINFEKL versus Optimised+SIINFEKL, $P=0.0017$ (one-way ANOVA) (f) SARS-CoV-2 live virus neutralisation titre from mice immunised with PfizerRBD+SIINFEKL or Optimised+SIINFEKL modified mRNA COVID-19 vaccines. IFN-γ ELISpot responses from splenocytes immunised with BNT162b2 or Optimised+SIINFEKL modified mRNA COVID-19 vaccines stimulated with overlapping peptide pools covering (g) the +1 frame of the RBD sequence or (h) 3'UTR of the Pfizer sequence. (g) BNT162b2 versus Optimised+SL8, P=0.0238 (unpaired t test); (h) BNT162b2 versus Optimised+SIINFEKL, P=0.0246 (unpaired t test).

Figure 17 shows schematics of the mRNA non-coding and coding sequences of SARS-CoV-2 RDB domain containing different UTRs derived from muscle or liver and containing an MHC class I epitope (SIINFEKL – SEQ ID NO: 72).

Figure 18 shows B cell antigen response determined by flow cytometry using the different UTRs and RBD-SIINFEKL constructs described in Figure x and the SIINFEKL antigen assay.

Figure 19 shows the percentage of cells showing a positive T cell antigen response using the different UTRs and RBD-SIINFEKL constructs described in Figure 17.

The patent, scientific and technical literature referred to herein establish knowledge that was available to those skilled in the art at the time of filing. The entire disclosures of the issued

patents, published and pending patent applications, and other publications that are cited herein are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference. In the case of any inconsistencies, the present disclosure will prevail.

Various aspects of the invention are described in further detail below.

Detailed Description

Provided herein are nucleic acid molecules that encode optimized mRNAs. The nucleic acid molecules may be DNA (e.g. a transcription template) or RNA. The term mRNA refers to an RNA molecule that encodes a protein. mRNA may refer to a ribonucleic acid molecule (RNA) that has been transcribed from a DNA sequence by an RNA polymerase enzyme, and interacts with a ribosome to synthesize protein encoded by DNA. Generally, mRNA is classified into two sub-classes: pre-mRNA and mature mRNA. Precursor mRNA (pre-mRNA) is mRNA that has been transcribed by RNA polymerase but has not undergone any post-transcriptional processing (e.g., 5′capping, splicing, editing, and polyadenylation) and may therefore include 5’ untranslated region (UTR), introns and/or one or more 3’ UTRs. Mature mRNA has been modified via post-transcriptional processing (e.g., spliced to remove introns and polyadenylated region) and is capable of interacting with ribosomes to perform protein synthesis. The particular nucleic acid molecule sequence composition and length of an mRNA will depend on the protein encoded by the mRNA. Traditionally, the basic components of an mRNA molecule include at least a coding region, a 5′UTR, a 3′UTR, a 5′ cap and a poly-A tail. In vitro transcribed (IVT) mRNA may function as mRNA but is distinguished from wild-type mRNA in their functional and/or structural design features, which serve to overcome existing problems of effective polypeptide production using nucleic-acid based therapeutics. For example, IVT mRNA may be chemically modified. For example, the polynucleotide “ATCG” may be chemically modified to “AT-5meC-G”.

The optimized mRNAs described herein may be referred to as “modified”. In the context of the invention, a modified optimized mRNA refers to an optimized mRNA that has been modified to introduce chemically modified nucleotides and/or genetic (nucleic acid molecule sequence) modifications.

The optimized mRNAs described herein may comprise naturally occurring ribonucleotides and/or non-naturally occurring ribonucleotides (e.g. canonical nucleotides) such as chemically modified nucleotides. In some examples, the optimized mRNAs provided herein may include at least one chemically modified ribonucleotide. In some examples, the chemically modified ribonucleotide may be selected from the group consisting of

pseudouridine, N1-methylpseudouridine (1-methylΨ), 2-thiouridine, 4′-thiouridine, 5-methylcytosine, 2-thio-l-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-pseudouridine, 2-thio-5-aza-uridine , 2-thio-dihydropseudouridine, 2-thio-dihydrouridine, 2-thio-pseudouridine, 4-methoxy-2-thio-pseudouridine, 4-methoxy-pseudouridine, 4-thio-1-methyl-pseudouridine, 4-thio-pseudouridine, 5-aza-uridine, dihydropseudouridine, 5-methoxyuridine, and 2′-O-methyl uridine. Other exemplary chemical modifications useful in optimized mRNAs described herein include those listed in US Published patent application 2015/0064235 which is incorporated herein.

In some examples, the optimized mRNAs provided herein include at least one N1-methylpseudouridine.

The optimized mRNAs provided herein may be pre-mRNAs or mature mRNAs. In some examples, the optimized mRNAs provided herein may include one or more features of a pre-mRNA but not all features of a pre-mRNA. For example, the optimized mRNAs provided herein may include a poly-adenylation, a 5'UTR and a 3' UTR sequence but not include introns. In some examples, the optimized mRNAs provided herein may include one or more features of a mature mRNA, but not all features of a mature mRNA

In some examples, the optimized mRNAs provided herein (or nucleic acid molecules encoding said optimized mRNAs) are therapeutic mRNAs (i.e. therapeutic optimized mRNAs). "Therapeutic mRNA" refers to an mRNA molecule (e.g., an in vitro transcribed (IVT) mRNA) that encodes a therapeutic protein. Therapeutic proteins mediate a variety of effects in a host cell or a subject in order to treat a disease or ameliorate the signs and symptoms of a disease. For example, a therapeutic protein can replace a protein that is deficient or abnormal, augment the function of an endogenous protein, provide a novel function to a cell (e.g., inhibit or activate an endogenous cellular activity, or act as a delivery agent for another therapeutic compound (e.g., an antibody-drug conjugate). Therapeutic mRNA may be useful for the treatment or prevention of the following diseases and conditions: infectious diseases (such as bacterial infections, viral infections, parasitic infections), cell proliferation disorders (such as cancer), genetic disorders, inflammatory disease, cardiovascular disorders, metabolic diseases, allergic disease, neurodegenerative diseases, protein or enzyme deficiency disorder and/or autoimmune diseases.

Examples of therapeutic mRNAs are Pfizer and BioNtech's BNT162b2 (Covid-19), Moderna's mRNA-1273 (Covid-19), mRNA-2416 (solid tumour or lymphoma), MRT5005 (cystic fibrosis), mRNA-2752 (solid tumour or lymphoma), AZD-8601 (heart failure), NY-ESO-1 (multiple myeloma, synovial sarcoma, melanoma), CTX001 (β-thalassemia), SB-728mR-HSPC (HIV), SB-728mR-T (HIV), BNT163 (HSV2), BNT164 (tuberculosis), BNT165

(malaria), BNT167 (shingles), BNT161 (influenza), BNT153 (undisclosed cancers), BNT152 (undisclosed cancers), BNT142 (undisclosed cancers), BNT141 (undisclosed cancers), BNT131 (undisclosed cancers), BNT122 (melanoma), colorectal cancer, BNT116 (non-small cell lung carcinoma), BNT115 (ovarian cancer), BNT113 (head and neck cancer), BNT112 (prostate cancer), BNT111 (melanoma), mRNA-1345 (Respiratory syncytial virus), mRNA-1010 (influenza), mRNA-1647 (cytomegalovirus), mRNA-4157/V940 (melanoma), mRNA-3927 (Propionic acidemia), mRNA-0184 (heart failure), VX-522 (cystic fibrosis). The uses of each mRNA therapy are shown in brackets.

Therapeutic mRNA molecules are generally synthesized in a laboratory (e.g., by in vitro transcription). mRNA can be isolated from tissues or cells by a variety of methods. For example, a total RNA extraction can be performed on cells or a cell lysate, and the resulting extracted total RNA can be purified (e.g., on a column comprising oligo-dT beads) to obtain extracted mRNA. Alternatively, mRNA can be synthesized in a cell-free environment, for example, by in vitro transcription (IVT). IVT is a process that permits template-directed (e.g. via an IVT DNA template) synthesis of a ribonucleic acid molecule (RNA) (e.g., messenger RNA (mRNA)). It is based, generally, on the engineering of a DNA template that includes a bacteriophage promoter sequence upstream of the sequence of interest, followed by transcription using a corresponding RNA polymerase. In vitro mRNA transcripts, for example, may be used as therapeutics in vivo to direct ribosomes to express protein therapeutics within targeted tissues.

An “in vitro transcription template (IVT),” as used herein, refers to deoxyribonucleic acid molecule (DNA) suitable for use in an IVT reaction for the production of messenger RNA (mRNA). In some examples, an IVT template encodes a 5′ untranslated region, contains an open reading frame, and encodes a 3′ untranslated region and a poly(A) tail. The particular nucleotide sequence composition and length of an IVT template will depend on the mRNA of interest encoded by the template. IVT mature mRNA preparation includes several steps, linear DNA template obtainment, IVT, 5′ capping, and poly(A) tail adding.

In some examples, the nucleic acid molecules provided herein are in vitro transcription templates (i.e. DNA suitable for IVT) or transcription templates (i.e. DNA suitable for transcription in a subject).

A “5′ untranslated region (UTR)” refers to a region of an mRNA that is directly upstream (i.e., 5′) from the start codon (i.e., the first codon of an mRNA transcript translated by a ribosome) that does not encode a protein or peptide. IVT is performed by linearizing plasmid DNA templates or PCR templates requiring at least a promoter and the corresponding mRNA construct sequence. IVT may be carried out by adding polymerases (T7, T3, or SP6) but

requires additional capping. Uncapped mRNA is rapidly degraded by RNase and contains a 5′-ppp group, which causes greater immune stimulation and can be treated with phosphatase to reduce undesirable efficacy. Two methods may be implemented for the capping of IVT mRNA: co-transcriptional capping and posttranscriptional capping. Cap dinucleotide mixtures containing four nucleoside triphosphates (NTPs) are incorporated at the 5′ end of the RNA with RNA polymerase during co-transcriptional capping. Co-transcriptional capping processing has permitted coordinated transcription with mRNA capping. Poly(A) tails of IVT mRNAs are normally encoded in the DNA template or attached to IVT mRNA by enzymatic polyadenylation. The former may have more precise control of the length of the poly(A) tail. IVT mRNAs may be mixed with RNA polymerase and DNA templates after synthesis; thus, purification of IVT mRNA may be needed, including removing immunostimulatory contaminants, free ribonucleotides, short mRNA and DNA templates. Generally, Dnase is used to degrade excess DNA templates. Commercial purification kits may be used to purify and separate the synthesized mRNA, followed by precipitation using ethanol or isopropanol, which can remove most contaminants and obtain high-purity mRNA, and then the mRNA may be precipitated with high concentrations of LiCl or alcohol-based precipitation, chromatographic methods (molecular exclusion chromatography, ion-exchange chromatography, or affinity chromatography with immobilized oligo-dT), or elution from a silica membrane column, which removes proteins, free nucleotides or other components but not dsRNA impurities. Reversed-phase HPLC may be used to remove dsRNA contaminants from the transcription reaction solution.

A “3′ untranslated region (UTR)” refers to a region of an mRNA that is directly downstream (i.e., 3′) from the stop codon (i.e., the codon of an mRNA transcript that signals a termination of translation) that does not encode a protein or peptide.

A “polyA tail” is a region of mRNA that is downstream, e.g., directly downstream (i.e., 3′), from the 3′ UTR that contains multiple, consecutive adenosine monophosphates. A polyA tail may contain 10 to 300 adenosine monophosphates. For example, a polyA tail may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290 or 300 adenosine monophosphates. In some examples, a polyA tail contains 50 to 250 adenosine monophosphates. In a relevant biological setting (e.g., in cells, in vivo, etc.), the poly(A) tail functions to protect mRNA from enzymatic degradation, e.g., in the cytoplasm, and aids in transcription termination, export of the mRNA from the nucleus, and translation. However, in some examples, mRNA molecules provided herein do not comprise a polyA tail (such molecules are referred to as “tailless”).

An “open reading frame” is a continuous stretch of DNA or RNA beginning with a start codon e.g., methionine (ATG) and ending with a stop codon (e.g., TAA, TAG or TGA) and encodes a protein or peptide.

In some examples, the optimised mRNAs provided herein may provide an improved translation efficiency in comparison to a control mRNA. The control mRNA may be any mRNA that does not include at least one of the preferred codons, signal peptides, UTR pairs, optimised poly(A) tail and/or 5’ cap as described herein.

In some examples, the optimised mRNAs provided including any one or more of the preferred codons, signal peptides, UTR pairs, optimised poly(A) tail and/or 5’ cap as described herein may have an increased translation efficiency in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150 or more amino acid codons which are preferred codons as described herein may have an increased translation efficiency in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising a 5’ and 3’ UTR pair as described herein may have an increased translation efficiency in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising an optimised poly(a) tail as described herein may have an increased translation efficiency in a target cell as described herein of at least 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%, 55%,

60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising a nucleic acid molecule sequence encoding a signal peptide as described herein may have an increased translation efficiency in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising an optimised 5' cap as described herein may have an increased translation efficiency in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

In some examples, the optimised mRNAs provided herein may provide an improved translation fidelity in comparison to a control mRNA. The control mRNA may be any mRNA that does not include at least one of the preferred codons, signal peptides, UTR pairs, optimised poly(A) tail and/or 5' cap as described herein.

In some examples, the optimised mRNAs provided including any one or more of the preferred codons, signal peptides, UTR pairs, optimised poly(A) tail and/or 5' cap as described herein may have an increased translation fidelity in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150 or more amino acid codons which are preferred codons as described herein may have an increased translation fidelity in target cell as described herein of at least 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%,

55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising a 5' and 3' UTR pair as described herein may have an increased translation fidelity in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising an optimised poly(a) tail as described herein may have an increased translation fidelity in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising a nucleic acid molecule sequence encoding a signal peptide as described herein may have an increased translation fidelity in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

For example, an mRNA comprising an optimised 5' cap as described herein may have an increased translation fidelity in a target cell as described herein of at least 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%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 120%, 130%, 140%, 150% or 200% in comparison to a control mRNA.

The nucleic acid molecules described herein may be isolated nucleic acid molecules. An "isolated" nucleic acid is substantially separated away from other nucleic acid sequences with which the nucleic acid is normally associated, such as from the chromosomal or extrachromosomal DNA of a cell in which the polynucleotide naturally occurs. The term also embraces nucleic acids that are biochemically purified so as to substantially remove

contaminating nucleic acids and other cellular components. Isolated nucleic acids are substantially free of sequences (such as protein encoding sequences) that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the polynucleotide) in the genomic DNA of the organism from which the nucleic acid is derived. For example, the isolated nucleic acid can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotide sequences that naturally flank the nucleic acid in genomic DNA of the cell from which the nucleic acid is derived.

In some examples, the nucleic acid molecules are purified nucleic acid molecules. The term "purified," relates to the isolation of a molecule or compound in a form that is substantially free of contaminants normally associated with the molecule or compound in a native or natural environment, or substantially enriched in concentration relative to other compounds present when the compound is first formed, and means having been increased in purity as a result of being separated from other components of the original composition. The term "purified nucleic acid" is used herein to describe a nucleic acid sequence which has been separated, produced apart from, or purified away from other biological compounds including, but not limited to polypeptides, lipids and carbohydrates, while affecting a chemical or functional change in the component (e.g. a nucleic acid may be purified from a chromosome by removing protein contaminants and breaking chemical bonds connecting the nucleic acid to the remaining DNA in the chromosome).

Target Cells

In some examples, the target cell is a human cell. In some examples, the target cell is an in vitro cell, an ex vivo cell or an in vivo cell. For example, the levels of tRNA isoacceptors or anticodons thereof may be determined from a cell that has been obtained from a subject and is maintained in vitro or ex vivo. In some examples, the levels of tRNA isoacceptors or anticodons thereof may be determined from a sample obtained from a subject's in vivo cell.

In some examples, the target cell may be selected from muscle cells, liver cells, kidney cells, immune cells, nervous system cells, fat cells, bone cells, blood cells and stem cells.

In some examples, the target cell may be a muscle cell such as a skeletal muscle cell, smooth muscle cell or cardiac cell. In some examples, the target cell is a skeletal muscle cell.

In some examples, the target cell may be a liver cell such as hepatocytes, parenchymal cells, non-parenchymal cells, sinusoidal endothelial cells, Kupffer cells, stellate cells, and lymphocytes. In some examples, the target cell is a hepatocyte.

In some examples, the target cell may be a kidney cell, such as an embryonic kidney cell (such as HEK293 cell), specialized filtration cells (e.g., mesangial cells and podocytes), epithelial cells (e.g., tubule brush border cells, thin segment cells, thick ascending limb cells, distal tubule cells), specialized resorption cells (e.g., collecting duct principal cells, collecting duct intercalated cells), and renal interstitial cells (e.g., fibroblasts, immune cells, macrophages, dendritic cells, and perivascular cells). In some examples, the target cell is a human embryonic kidney cell.

In some examples, the target cell may be an immune cell. "immune cell" includes any cell within the immune system. Examples include white blood cells, such as lymphocytes (e.g. T lymphocytes or T cells, B cells or natural killer cells), dendritic cells, follicular dendritic cells and granulocytes.

In some examples, the target cell is an antigen-presenting cell. "Antigen-presenting cell" refers to cells capable of acquiring, processing, presenting, or displaying at least one antigen or antigenic fragment on (or at) its cell surface. In general, the term "antigen-presenting cell" can be any cell that aids the enhancement of an immune response or immune tolerance (e.g., from the T-cell or B-cell arms of the immune system) to an antigen. For example, a cell that displays or presents an antigen normally or preferentially with a class II major histocompatibility molecule or complex to an immune cell may be denoted as an "antigen-presenting cell". Antigen-presenting cells include but are not limited to astrocytes, oligodendrocytes, microglia, macrophages, B cells, dendritic cells and precursors thereof.

In some examples, the target cell is a nervous system cell. Nervous system cells include nerve cells, neural stem cells and neural progenitor cells. Examples of nerve cells include inhibitory nerve cells, excitatory nerve cells, dopamin-producing nerve cells, cranial nerves, intervening nerves, and optic nerves. Alternatively, nervous system cells may be motor nerve cells, oligodendrocyte progenitor cells, astrocytes, oligodendrocytes or the like.

In some examples, the target cell is a bone cell such as osteoblasts, osteoclasts, and osteocytes.

In some examples, the target cell is a blood cell, such as red blood cells, platelets, monocytes, lymphocytes, neutrophils, eosinophils, basophils, and macrophages.

In some examples, the target cell is a stem cell. "Stem cell" refers to a multipotent cell having the capacity to self renew and to differentiate into multiple cell lineages. Stem cells include pluripotent stem cells, somatic stem cells, embryonic stem cells, induced pluripotent stem, hematopoietic stem cells, mesenchymal stem cells, neural stem cells, epithelial stem cells. In some examples, the target cell is not a totipotent cell.

In some examples, the target cell is a cardiac cell, such as a cardiomyocyte. "Cardiomyocyte" refers to heart muscle cells. Cardiomyocytes may be ventricular-, atrial-, and/or nodal-type cardiomyocytes, or a mixed population of cardiomyocytes. Cardiomyocytes may exhibit one or more features including, capacity to beat spontaneously, are electrically mature, metabolically mature, contractility mature, exhibit appropriate expression of one or more gene markers (e.g., TNNI3, TNNT1, MYH6, MYH7, KCNJ2, RyR, and REST), exhibit appropriate expression of one or more quiescence markers (e.g., p16 and p130), and/or exhibit appropriate morphological characteristics (e.g., rod shaped cells and organized sarcomeres).

Optimised Codons

The nucleic acid molecules provided herein encode one or more codons that have been optimised using the methods described herein in order to provide improved translation efficiency and/or fidelity when translated in a cell.

Also provided are methods of producing nucleic acid molecules including preferred codons.

"Fidelity of translation" refers to the accuracy of translation of an mRNA. For example, the translation of the desired protein (comprising the amino acids encoded by the in-frame codons) which is produced by translation occurring in the intended reading frame (i.e. correct or in-frame translation). "Efficiency of translation" refers to the speed at which a protein is translated from an RNA molecule and/or the amount of protein translated from an RNA molecule.

The codons encoded by the nucleic acid molecules described herein may be selected using various methods described below. Codons that may have been selected for improved properties may be referred to herein as "preferred codons". Preferred codon refers to a codon that encodes for the transcription of a desired amino acid that has been selected from other codons that encode for the transcription of the same amino acid.

The preferred codons may be selected based on a level of tRNA isoacceptor expression in a target cell. Isoacceptor tRNAs refers to tRNAs that have different anticodons but still carry the same amino acid. A particular tRNA may be denoted herein according to its aminoacylating amino acid, which is indicated such as in $tRNA^{Leu}$ or tRNA-Leu. A particular tRNA may additionally be denoted according to its anticodon, indicated, such as $tRNA^{Leu}$ (CAG) or tRNA-Leu-CAG.

For example, the preferred codon encoding an amino acid may be selected based on the amount or abundance of each isoacceptor tRNA for the amino acid in the target cell. The

level of tRNA isoacceptor for at least one or each amino acid may be determined using any suitable method. For example, using mRNA sequencing methods such as RNA sequencing (e.g. small RNA sequencing).

RNA sequencing may include the use of methods such as amplification (PCR) based methods (reverse transcriptase PCR (RT-PCR) and quantitative reverse transcriptase PCR (qRT-PCR), or RNAseq (next generation sequencing, also referred to as second generation sequencing or massively parallel sequencing). Methods of DNA sequencing are known and include methods similar to those used for determining mRNA sequences.

Next Generation Sequencing (NGS) (second generation sequencing or massively parallel sequencing; Mardis, E. R. (2008). As there are many NGS technologies available, there are small differences in the methodology for RNA sequencing. The following is a description of how RNA sequencing using NGS works in general. Total RNA is extracted from the sample of interest using a common RNA extraction method. Post-extraction processes can be used to enrich the RNA sample. Complimentary DNA (cDNA) is then synthesised using extracted RNA. cDNA is then used as the template for RNA sequencing. NGS uses variations of sequencing by synthesis (SBS) chemistry (*Fuller, C. W., et al. (2009). The challenges of sequencing by synthesis. Nature biotechnology, 27(11), 1013-1023*). With cDNA as a template, new nucleotide fragments, known as reads, are synthesised base by base, with each incorporated base recorded during sequencing (Fuller, 2009). The data output from RNA sequencing is a list of all the reads generated, and their sequence (Fuller, 2009 and Metzker, 2010). This data undergoes quality assessment (Patel, R. K., & Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PloS one, 7(2), e30619).

Small RNA sequencing is a technique to isolate and sequence small RNA species, such as tRNAs.

In some examples, the preferred codons are determined by determining the levels of expression of tRNA anticodons for at least one or each amino acid in a target cell. Determining the level of expression of tRNA anticodons may be done using any suitable methods known in the field. For example, determining levels of expression may utilise RNA sequencing methods described above. In addition, or alternatively, tRNA anticodon expression levels may be determined using RT-PCR, real time RT-PCR, next generation sequencing, western blot, dot blot, enzyme linked immunosorbent assay (ELISA). Differential expression of genes may be determined by performing RNA expression analysis. RNA may be extracted from a target cell and the level of RNA may be quantified by hybridisation of probes to provide a gene count.

The level of expression or gene count, of each gene may then be normalised based on the expression levels of a number of housekeeping genes.

Determination of the amount of tRNA isoacceptors or expression levels of tRNA anticodons for tRNA isoacceptors may be used to select specific codons to be encoded in the nucleic acid. For example, the nucleic acid molecule may encode the codon for an amino acid that corresponds to the most abundant isoacceptor tRNA and anticodon thereof for the same amino acid detected in a target cell. For example, the nucleic acid molecule may encode the codon that corresponds to the most abundantly expressed tRNA anticodon for the same amino acid detected in a target cell.

In some examples, preferred codons may be determined by analysis of codon decoding speed or rate in a target cell. Codon decoding rate may be determined using methods such as ribosome profiling. Ribosome profiling may be used to determine the time, extent, rate and/or fidelity of ribosome decoding of a particular codon of an RNA, which in turn is determined by the amount of time a ribosome spends at a particular codon (also referred to as dwell time). The latter may be referred to as a "codon elongation rate" or a "codon decoding rate". Relative dwell time of ribosomes between two locations in an RNA, instead of the actual or absolute dwell time at a single location, can also be determined by comparing the number of sequencing reads of protected mRNA fragments at each location (e.g., a codon) (see O'Connor et al., (2016) Nature Commun 7: 12915).

In some examples, preferred codon selection may be determined using the ratio of near-cognate to cognate tRNA. Methods of determining the ratio of near-cognate to cognate tRNA can be found in “Chu, Dominique, David J. Barnes, and Tobias Von Der Haar. "The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae." Nucleic acids research 39.15 (2011): 6705-6714”. In some examples, the preferred codon may be the codon with the lowest ratio of near-cognate to cognate tRNAs. The ratio might vary from zero to ~50, and the preferred codon may be the codon which has the lowest ratio of all codons encoding the same amino acid.

In some examples, preferred codons may be determined by analysis of the binding affinity between a codon and its cognate tRNA anticodon. For example, RNA oligonucleotides may be chemically synthesized using standard phosphoramidite chemistry. mRNA codon and tRNA anticodon sequence may be synthesised fused to standard anchor sequences (see Figure 11A). A set of oligonucleotides comprising only the anchor sequences, without added codon/anticodon sequences, may be used as controls. Oligos may then be resuspended in buffer, and pairs of oligos containing codon and anticodon sequences are mixed. The mixed oligos are heated in defined temperature increments, and circular dichroism (CD) spectra are

recorded after each temperature increase. Melting points are determined from the temperature-dependent CD spectra, by plotting the spectral shift at 265 nm against the temperature. For more details see “Wang, Xiaoyu, et al. "Biophysics of artificially expanded genetic information systems. Thermodynamics of DNA duplexes containing matches and mismatches involving 2-amino-3-nitropyridin-6-one (Z) and imidazo [1, 2-a]-1, 3, 5-triazin-4 (8H) one (P)." ACS Synthetic Biology 6.5 (2017): 782-792”.

In some examples, preferred codons may be determined by analysis of frequency with which codons are used in a genome. For example, the relative usage of all codons in the human genome is known, and more frequently used codons are usually found in more highly expressed mRNAs. Preferred codons may thus be determined by analysing which codons are most frequently used in natural genes.

It will be understood that the methods of determining preferred codons for a target cell may be carried out for at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more amino acids encoded by a nucleic acid molecule as described herein. For example, all the amino acids encoded by a nucleic acid molecule may be encoded by a preferred codon determined by one or more of the methods described above. In some examples, only a portion of the codons are preferred codons. In some examples, a target cell may not have a preferred codon for one or more amino acids and as such, any codon for that amino acid may be used. In some examples, an amino acid may have more than one preferred codon and the preferred codons encoded for each occurrence of the codon may include a single preferred codon or a mixture of preferred codons. In some examples, where there is no preferred codon or multiple preferred codons determined, the codon for the amino acid encoded by the nucleic acid molecule may selected based on other factors or properties. For example, selected based on one or more of the parameters such as target application or specified nucleotide sub-sequences as described herein.

After determination of preferred codons, the nucleic acid molecule may be produced incorporating the preferred codons using methods known in the field. For example, for the production of DNA molecules encoding optimised mRNAs as described herein, such as solid-phase DNA synthesis and ligation methods, DNA assembly methods (e.g. using BioBricks, and Golden Gate cloning methods), long-overlap based assembly methods (such as Gibson assembly, MODAL, or BASIC).

Production of optimised mRNAs as described herein may done using the methods described above (i.e. de novo synthesis) or may be produced by in vitro transcription as described herein. For example, a DNA molecule encoding the optimised mRNA may be produced and used as a template for in vitro transcription.

In some examples, the methods include adding a poly(A) tail, as described herein. For example, an optimised poly(A) tail as described herein (e.g. comprising about 97 to 100 adenine residues). In particular, the methods may include enzymatically adding a poly(A) tail, such as an optimised poly(A) tail as described herein. For example, the poly(A) tail may be added by contacting the mRNA with E.Coli poly(A) polymerase.

In some examples, the methods include capping the optimised mRNA molecule with a 5' capping agent. For example, capping the optimised mRNA molecule using a m7G-5'ppp5'GpG 5' cap or m7G5'ppp5'2'OMeApG. In particular, the optimised mRNA may be capped using a m7G-5'ppp5'GpG 5' cap.

The methods may further include purification of the nucleic acid. For example, purification of the optimised mRNA. Methods of purifying nucleic acid molecules are well known and purification can be carried out using suitable commercially available kits. In some examples, purification is carried out using methods based on phenol-chloroform extraction, spin column purification, and magnetic bead-based method.

Muscle Cell Preferred Codons

As shown in the examples provided herein, preferred codons may be determined for specific target cell types. For example, the target cell may be a muscle cell, in such examples, the set of preferred codons may be referred to herein as "muscle cell preferred codons". Muscle cell preferred codons are not to be construed as limited to muscle cells and may be preferred codons in other cell types.

In some examples, the target cell is a muscle cell (such as a human skeletal muscle cell) and the preferred codon for alanine is 5' GCU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one alanine residue is encoded by the codon 5' GCU 3'. In some examples, all alanine residues are encoded by the codon 5' GCU 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for arginine is 5' AGG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one arginine residue is encoded by the codon 5' AGG 3'. In some examples, all arginine residues are encoded by the codon 5' AGG 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for aspartic acid is 5' GAC 3'. As such, in one example, there is

provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one aspartic acid residue is encoded by the codon 5' GAC 3'. In some examples, all aspartic acid residues are encoded by the codon 5' GAC 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a glutamic acid is 5' GAG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one glutamic acid residue is encoded by the codon 5' GAG 3'. In some examples, all glutamic acid residues are encoded by the codon 5' GAG 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a glycine is 5' GGC mRNA 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one glycine residue is encoded by the codon 5' GGC mRNA 3'. In some examples, all glycine residues are encoded by the codon 5' GGC mRNA 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a leucine is 5' CUU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one leucine residue is encoded by the codon 5' CUU 3'. In some examples, all leucine residues are encoded by the codon 5' CUU 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a lysine is 5' AAG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one lysine residue is encoded by the codon 5' AAG 3'. In some examples, all lysine residues are encoded by the codon 5' AAG 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a proline is 5' CCU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one proline residue is encoded by the codon 5' CCU 3'. In some examples, all proline residues are encoded by the codon 5' CCU 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a serine is 5' AGC mRNA 3'. As such, in one example, there

is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one serine residue is encoded by the codon 5' AGC mRNA 3'. In some examples, all serine residues are encoded by the codon 5' AGC mRNA 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a asparagine is 5' AAC 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one asparagine residue is encoded by the codon 5' AAC 3'. In some examples, all asparagine residues are encoded by the codon 5' AAC 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a glutamine is 5' CAG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one glutamine residue is encoded by the codon 5' CAG 3'. In some examples, all glutamine residues are encoded by the codon 5' CAG 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a histidine is 5' CAC 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one histidine residue is encoded by the codon 5' CAC 3'. In some examples, all histidine residues are encoded by the codon 5' CAC 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a valine is 5' GUG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one valine residue is encoded by the codon 5' GUG 3'. In some examples, all valine residues are encoded by the codon 5' GUG 3'.

In some examples, the target cell is a skeletal muscle cell (such as a human skeletal muscle cell) and the preferred codon for a methionine is 5' AUG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a skeletal muscle cell wherein at least one methionine residue is encoded by the codon 5' AUG 3'. In some examples, all methionine residues are encoded by the codon 5' AUG 3'.

Kidney Cell Preferred Codons

As shown in the examples provided herein, preferred codons may be determined for specific target cell types. For example, the target cell may be a kidney cell, in such examples, the set of preferred codons may be referred to herein as "kidney cell preferred codons". It is well known that human embryonic kidney cells are used in the field as a model for a number of different cells. As such, kidney cell preferred codons are not to be construed as limited to kidney cells and may be preferred codons in other cell types.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for alanine is 5' GCU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one alanine residue is encoded by the codon 5' GCU 3'. In some examples, all alanine residues are encoded by the codon 5' GCU 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for arginine is 5' CGU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one arginine residue is encoded by the codon 5' CGU 3'. In some examples, all arginine residues are encoded by the codon 5' CGU 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for aspartic acid is 5' GAC 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one aspartic acid residue is encoded by the codon 5' GAC 3'. In some examples, all aspartic acid residues are encoded by the codon 5' GAC 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for glutamic acid is 5' CAA 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one glutamic acid residue is encoded by the codon 5' CAA 3'. In some examples, all glutamic acid residues are encoded by the codon 5' CAA 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for glycine is 5' GGC mRNA 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one glycine residue is encoded by the

codon 5' GGC mRNA 3'. In some examples, all glycine residues are encoded by the codon 5' GGC mRNA 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for leucine is 5' CUG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one leucine residue is encoded by the codon 5' CUG 3'. In some examples, all leucine residues are encoded by the codon 5' CUG 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for lysine is 5' AAG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one lysine residue is encoded by the codon 5' AAG 3'. In some examples, all lysine residues are encoded by the codon 5' AAG 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for proline is 5' CCU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one proline residue is encoded by the codon 5' CCU 3'. In some examples, all proline residues are encoded by the codon 5' CCU 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for serine is 5' AGC mRNA 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one serine residue is encoded by the codon 5' AGC mRNA 3'. In some examples, all serine residues are encoded by the codon 5' AGC mRNA 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for asparagine is 5' AAC 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one asparagine residue is encoded by the codon 5' AAC 3'. In some examples, all asparagine residues are encoded by the codon 5' AAC 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for glutamine is 5' CAG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one glutamine residue is encoded by the codon 5' CAG 3'. In some examples, all glutamine residues are encoded by the codon 5' CAG 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for histidine is 5' CAC 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one histidine residue is encoded by the codon 5' CAC 3'. In some examples, all histidine residues are encoded by the codon 5' CAC 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for valine is 5' GUU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one valine residue is encoded by the codon 5' GUU 3'. In some examples, all valine residues are encoded by the codon 5' GUU 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for isoleucine is 5' AUA 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one isoleucine residue is encoded by the codon 5' AUA 3'. In some examples, all isoleucine residues are encoded by the codon 5' AUA 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for threonine is 5' ACU 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one threonine residue is encoded by the codon 5' ACU 3'. In some examples, all threonine residues are encoded by the codon 5' ACU 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for tryptophan is 5' UGG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one tryptophan residue is encoded by the codon 5' UGG 3'. In some examples, all tryptophan residues are encoded by the codon 5' UGG 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for tyrosine is 5' UAC 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one tyrosine residue is encoded by the codon 5' UAC 3'. In some examples, all tyrosine residues are encoded by the codon 5' UAC 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for phenylalanine is 5' UUC 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one phenylalanine residue is encoded by the codon 5' UUC 3'. In some examples, all phenylalanine residues are encoded by the codon 5' UUC 3'.

In some examples, the target cell is a kidney cell (such as a human embryonic kidney cell, e.g., HEK293T cell) and the preferred codon for methionine is 5' AUG 3'. As such, in one example, there is provided a nucleic acid molecule such as an mRNA (or DNA encoding the mRNA) for translation in a kidney cell wherein at least one methionine residue is encoded by the codon 5' AUG 3'. In some examples, all methionine residues are encoded by the codon 5' AUG 3'.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic optimised mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

Signal Peptides

Provided herein are nucleic acid molecules that encode a signal peptide derived from alpha-1-antichymotrypsin (AACT) signal peptide or Human Serum albumin (HSA) signal peptide.

In particular, the nucleic acid molecule may be an mRNA (such as a therapeutic optimised mRNA as described herein) or a DNA molecule encoding an optimised mRNA.

In some examples, the encoded signal peptide is an endoplasmic reticulum (ER) signal peptide. The nucleic acid molecule sequence encoding the ER signal peptide may be

preferred to as an ER secretion signal sequence. As such, in some examples, the nucleic acid molecules described herein may include an ER secretion signal sequence. The ER signal peptide may be encoded so that the translated peptide includes the ER signal peptide at the N-terminus of the peptide. Therefore, the nucleic acid molecules described herein may include a nucleic acid molecule sequence located 5' of a protein encoded by the nucleic acid. In some examples, the ER signal peptide is suitable for binding by a signal recognition particle (SRP) in a target cell and for translocation of the peptide during translation to the endoplasmic reticulum.

The signal peptide may have effects on the efficiency of translation as well as provide secretion of translated proteins. For example, the nucleic acid molecule may be a therapeutic optimised mRNA encoding an antigen (such as a viral antigen) that needs to be secreted from the target cell to provide a therapeutic effect. In addition, factors such as protein yield, cellular localisation, secretion efficiency, and post-translational modifications may vary depending on the signal peptide used.

AACT is a member of the serpin family of proteins, a group of proteins that inhibit serine proteases. The gene encoding AACT (*SERPINA3*) is overexpressed in the liver. This gene is one in a cluster of serpin genes located on the q arm of chromosome 14. Although its physiological function is unclear, it can inhibit neutrophil cathepsin G and mast cell chymase, both of which can convert angiotensin-1 to the active angiotensin-2. Human AACT is represented by UniProtKB ID P01011.

The signal sequence of AACT is considered to comprise amino acid residues 1 to 23 of the sequence provided by UniProtKB number P01011. In some examples, the AACT signal peptide provided herein comprises at least residues 1 to 23 of human AACT:

MERMLPLLALGLLAAGFCPAVLC (SEQ ID NO: 71)

In some examples, the AACT signal peptide provided herein comprises at least residues 1 to 25 of human AACT:

MERMLPLLALGLLAAGFCPAVLCHP (SEQ ID NO: 10)

In some examples, the nucleic acid molecule encodes a signal peptide having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 10.

In some examples, the nucleic acid molecule encodes a signal peptide comprising SEQ ID NO: 10. In some examples, the nucleic acid molecule encodes a signal peptide comprising SEQ ID NO: 71.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide comprising SEQ ID NO: 10.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide consisting of SEQ ID NO: 10.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide comprising of SEQ ID NO: 10 and a nucleic acid molecule sequence encoding a protein (such as a therapeutic protein) wherein the nucleic acid molecule sequence encoding the signal peptide is located upstream (5') of the nucleic acid molecule sequence encoding a protein.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide consisting of SEQ ID NO: 10 and a nucleic acid molecule sequence encoding a protein (such as a therapeutic protein) wherein the nucleic acid molecule sequence encoding the signal peptide is located upstream (5') of the nucleic acid molecule sequence encoding a protein.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 10 comprises a nucleic acid molecule sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 11.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 10 comprises a nucleic acid molecule sequence comprising SEQ ID NO: 11.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 10 comprises a nucleic acid molecule sequence consisting of SEQ ID NO: 11.

In some examples, the nucleic acid molecule sequence encoding the AACT signal peptide may comprise one or more preferred codons as described herein (i.e. one or more codons that have been optimised using the methods described herein).

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 10 comprises a nucleic acid molecule sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 12.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 10 comprises a nucleic acid molecule sequence comprising SEQ ID NO: 12.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 10 comprises a nucleic acid molecule sequence consisting of SEQ ID NO: 12.

HSA is the serum albumin found in human blood. It is the most abundant protein in human blood plasma. The gene for HSA is located on chromosome 4 in locus 4q13.3. HSA

functions in the regulation of blood plasma colloid osmotic pressure and acts as a carrier protein for a wide range of endogenous molecules including hormones, fatty acids, and metabolites, as well as exogenous drugs. Additionally, HSA exhibits an esterase-like activity with broad substrate specificity. The encoded preproprotein is proteolytically processed to generate the mature protein. Human AACT is represented by UniProtKB ID P02768.

The signal sequence of HSA is considered to comprise amino acid residues 1 to 18 of the sequence provided by UniProtKB ID P02768. In some examples, the HSA signal peptide provided herein comprises at least residues 1 to 18 of human HSA:

MKWVUFISLLFLFSSAYS (SED ID NO: 22)

In some examples, the nucleic acid molecule encodes a signal peptide having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 22.

In some examples, the nucleic acid molecule encodes a signal peptide comprising SEQ ID NO: 22.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide comprising SEQ ID NO: 22.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide consisting SEQ ID NO: 22.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide comprising of SEQ ID NO: 22 and a nucleic acid molecule sequence encoding a protein (such as a therapeutic protein) wherein the nucleic acid molecule sequence encoding the signal peptide is located upstream (5') of the nucleic acid molecule sequence encoding a protein.

In some examples, the nucleic acid molecule includes a nucleic acid molecule sequence encoding a signal peptide consisting of SEQ ID NO: 22 and a nucleic acid molecule sequence encoding a protein (such as a therapeutic protein) wherein the nucleic acid molecule sequence encoding the signal peptide is located upstream (5') of the nucleic acid molecule sequence encoding a protein.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 22 comprises a nucleic acid molecule sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 23.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 22 comprises a nucleic acid molecule sequence comprising SEQ ID NO: 23.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 22 comprises a nucleic acid molecule sequence consisting of SEQ ID NO: 23.

In some examples, the nucleic acid molecule sequence encoding the HSA signal peptide may comprise one or more preferred codons as described herein (i.e. one or more codons that have been optimised using the methods described herein).

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 22 comprises a nucleic acid molecule sequence having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 24.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 22 comprises a nucleic acid molecule sequence comprising SEQ ID NO: 24.

In some examples, the nucleic acid molecule sequence encoding the signal peptide according to SEQ ID NO: 22 comprises a nucleic acid molecule sequence consisting of SEQ ID NO: 24.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding an AACT signal peptide as described above. In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding an HSA signal peptide as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding an AACT signal peptide as described above. In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding an HSA signal peptide as described above. In some examples, one or more codons of the encoded signal peptide comprise a preferred codon.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding an AACT signal peptide as described above. In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding an HSA signal peptide as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding an AACT signal peptide as described above. In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding an HSA signal peptide as described above. In some examples, one or more codons of the encoded signal peptide comprise a preferred codon.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide, as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, an AACT signal peptide as described herein may provide improved translation efficiency in liver cells in particular. As such, there is provided use of an AACT signal peptide as described herein for translation of optimised mRNA in a liver cell. For example, use of an AACT signal peptide comprising or consisting of an amino acid sequence according to SEQ ID NO: 10. For example, use of a nucleic acid molecule comprising a nucleic acid molecule sequence encoding an AACT signal peptide comprising or consisting of an amino acid sequence according to SEQ ID NO: 10. For example, use of a nucleic acid molecule comprising a nucleic acid molecule sequence according to SEQ ID NO:

Optimised UTR Pairs

Provided herein are untranslated regions (UTRs) that may improve the efficiency of translation of an optimised mRNA, as described herein. The UTRs provided herein may be provided as pairs including one 5' UTR and one 3' UTR which may both be included or encoded by a nucleic acid molecule as described herein at respective ends of the nucleic acid molecules described herein. The pairs of UTRs may be referred to using a designation based on the cell type from which the UTRs are derived. For example, MUTR1 refers to a pair of UTRs derived from genes expressed in muscle (M) cells. As such, use of the designations refers to a specific pair of UTRs as detailed in Tables 1 and 2 below:

Table 1 - Muscle-specific UTR combinations

	Gene name	5′ UTR	5′ UTR SEQ ID NO	3′ UTR	3′ UTR SEQ ID NO
MUTR 1	myosin light chain, phosphorylatable, fast skeletal muscle (MYLPF)	MYLPF	35	MYLPF	36
MUTR 2	myosin light chain 1 (MYL1)	MYL1	37	MYL1	38
MUTR 3	actin, alpha 1, skeletal muscle (ACTA1)	ACTA1	39	ACTA1	40
MUTR 4	troponin C1, slow skeletal and cardiac type (TNNC1) & ENO3	TNNC1	41	ENO3	42
MUTR 5	troponin T1, slow skeletal type (TNNT1)	TNNT1	43	TNNT1	44
MUTR 6	enolase 3 (ENO3)	ENO3	45	ENO3	46
MUTR 7	myosin heavy chain 1 (MYH1)	MYH1	47	MYH1	48
MUTR 8	myosin light chain 2 (MYL2)	MYL2	49	MYL2	50
MUTR 9	troponin I2, fast skeletal type (TNNI2)	TNNI2	51	TNNI2	52
MUTR 10	creatine kinase, M-type (CKM)	CKM	53	CKM	54
MUTR 11	enolase 3 (ENO3) & cytochrome c oxidase subunit 6A2 (COX6A2)	ENO3	45	COX6A2	55

Table 2 - Liver-specific UTR combinations

	Gene name	5′ UTR	5′ UTR SEQ ID NO	3′ UTR	3′ UTR SEQ ID NO
LUTR 1	orosomucoid 1 (ORM1)	ORM1	56	ORM2	57
LUTR 2	serum amyloid A2 (SAA2)	SAA2	58	SAA2	59
LUTR 3	serum amyloid A4, constitutive (SAA4)	SAA4	60	SAA4	61
LUTR 4	apolipoprotein A1 (APOA1)	APOA1	62	APOA1	63
LUTR 5	complement factor H related 2 (CFHR2) & SAA4	CFHR2	64	SAA4	61
LUTR 6	serum amyloid A1 (SAA1)	SAA1	65	SAA1	66
LUTR 7	GC, vitamin D binding protein (GC)	GC	67	GC	68
LUTR 8	alpha 2-HS glycoprotein (AHSG)	AHSG	69	AHSG	70
LUTR 9	serum amyloid A4, constitutive (SAA4) & APOA1	SAA4	60	APOA1	63
LUTR 10	serum amyloid A1 (SAA1) & GC	SAA1	65	GC	68

"Untranslated region" refers to a region positioned at both sides of a coding region and is not to be translated into a protein. mRNAs may include a 5′-untranslated region (5′-UTR) at the 5′ terminal end and a 3′-untranslated region (3′-UTR) at the 3′ terminal end of a mRNA.

"Coding region" refers to a region to be translated into a protein in mRNA.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from Myosin Light Chain, Phosphorylatable, Fast Skeletal Muscle (MYLPF) mRNA and a nucleic acid molecule encoding a 3' UTR derived from Myosin Light Chain, Phosphorylatable, Fast Skeletal Muscle (MYLPF) mRNA (MUTR1).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from MYLPF mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from MYLPF mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from MYLPF mRNA and a 3' UTR derived from MYLPF mRNA.

In some examples, the 5' UTR derived from MYLPF mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 35. In some examples, the 5' UTR derived from MYLPF mRNA comprises a nucleic acid molecule

sequence according to SEQ ID NO: 35. In some examples, the 5' UTR derived from MYLPF mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 35.

In some examples, the 3' UTR derived from MYLPF mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 36. In some examples, the 3' UTR derived from MYLPF mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 36. In some examples, the 3' UTR derived from MYLPF mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 36.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYLPF mRNA and a 3' UTR derived from MYLPF mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYLPF mRNA and a 3' UTR derived from MYLPF mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYLPF mRNA and a 3' UTR derived from MYLPF mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYLPF mRNA and a 3' UTR derived from MYLPF mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a

preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from Myosin Light Chain (MYL1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from Myosin Light Chain (MYL1) mRNA (MUTR2).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from MYL1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from MYL1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from MYL1 mRNA and a 3' UTR derived from MYL1 mRNA.

In some examples, the 5' UTR derived from MYL1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 37. In some examples, the 5' UTR derived from MYL1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 37. In some examples, the 5' UTR derived from MYL1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 37.

In some examples, the 3' UTR derived from MYL1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 38. In some examples, the 3' UTR derived from MYL1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 38. In some examples, the 3' UTR derived from MYL1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 38.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL1 mRNA and a 3' UTR derived from MYL1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or

more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL1 mRNA and a 3' UTR derived from MYL1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL1 mRNA and a 3' UTR derived from MYL1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL1 mRNA and a 3' UTR derived from MYL1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR2 (i.e. SEQ ID NOs: 37 and 38) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein in muscle cells.

In some examples, the UTR pair of MUTR2 (i.e. SEQ ID NOs: 37 and 38) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYL1 mRNA (SEQ ID NO: 37) and a 3' UTR derived from MYL1 mRNA (SEQ ID NO: 38) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from actin, alpha1, skeletal muscle (ACTA1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from actin, alpha1, skeletal muscle (ACTA1) mRNA (MUTR3).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ACTA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from ACTA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ACTA1 mRNA and a 3' UTR derived from ACTA1 mRNA.

In some examples, the 5' UTR derived from ACTA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 39. In some examples, the 5' UTR derived from ACTA1 mRNA comprises a nucleic acid molecule

sequence according to SEQ ID NO: 39. In some examples, the 5' UTR derived from ACTA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 39.

In some examples, the 3' UTR derived from ACTA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 40. In some examples, the 3' UTR derived from ACTA1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 40 In some examples, the 3' UTR derived from ACTA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 40.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ACTA1 mRNA and a 3' UTR derived from ACTA1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ACTA1 mRNA and a 3' UTR derived from ACTA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ACTA1 mRNA and a 3' UTR derived from ACTA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ACTA1 mRNA and a 3' UTR derived from ACTA1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a

preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic optimised mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from Troponin C1, Slow Skeletal And Cardiac Type (TNNC1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from enolase 3 (ENO3) mRNA (MUTR4).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from TNNC1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from TNNC1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from TNNC1 mRNA and a 3' UTR derived from TNNC1 mRNA.

In some examples, the 5' UTR derived from TNNC1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 41. In some examples, the 5' UTR derived from TNNC1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 41. In some examples, the 5' UTR derived from TNNC1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 41.

In some examples, the 3' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 42. In some examples, the 3' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 42. In some examples, the 3' UTR derived from ENO3 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 42.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNC1 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon

comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNC1 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNC1 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNC1 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR4 (i.e. SEQ ID NOs: 41 and 42) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein in muscle cells.

In some examples, the UTR pair of MUTR4 (i.e. SEQ ID NOs: 41 and 42) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 41) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 42) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from Troponin T1, Slow Skeletal Type (TNNT1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from Troponin T1, Slow Skeletal Type (TNNT1) mRNA (MUTR5).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from TNNT1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from TNNT1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from TNNT1 mRNA and a 3' UTR derived from TNNT1 mRNA.

In some examples, the 5' UTR derived from TNNT1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 43. In some examples, the 5' UTR derived from TNNT1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 43. In some examples, the 5' UTR derived from TNNT1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 43.

In some examples, the 3' UTR derived from TNNT1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 44. In some examples, the 3' UTR derived from TNNT1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 44. In some examples, the 3' UTR derived from TNNT1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 44.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNT1 mRNA and a 3' UTR derived from TNNT1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNT1 mRNA and a 3' UTR derived from TNNT1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNT1 mRNA and a 3' UTR derived from TNNT1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNT1 mRNA and a 3' UTR derived from TNNT1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR5 (i.e. SEQ ID NOs: 43 and 44) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 43) and a 3' UTR derived from TNNC1 mRNA (SEQ ID NO: 44) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 43) and a 3' UTR derived from TNNC1 mRNA (SEQ ID NO: 44) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 43) and a 3' UTR derived from TNNC1 mRNA (SEQ ID NO: 44) as described herein may be for use in liver cells. As such, there is provided use of an mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from TNNC1 mRNA (SEQ ID NO: 43) and a 3' UTR derived from TNNC1 mRNA (SEQ ID NO: 44) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from enolase 3 (ENO3) mRNA and a nucleic acid molecule encoding a 3' UTR derived from enolase 3 (ENO3) mRNA (MUTR6).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ENO3 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from ENO3 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from ENO3 mRNA (MUTR6).

In some examples, the 5' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 45. In some examples, the 5' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 45. In some examples, the 5' UTR derived from ENO3 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 45.

In some examples, the 3' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%,

85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 46. In some examples, the 3' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 46. In some examples, the 3' UTR derived from ENO3 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 46.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from ENO3 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or

more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR6 (i.e. SEQ ID NOs: 45 and 46) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein in muscle cells.

In some examples, the UTR pair of MUTR6 (i.e. SEQ ID NOs: 45 and 46) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from ENO3 mRNA (SEQ ID NO: 46) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from myosin heavy chain 1 (MYH1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from myosin heavy chain 1 (MYH1) mRNA (MUTR7).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from MYH1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from MYH1 mRNA. In some examples, there is provided a nucleic acid

molecule comprising a 5' UTR derived from MYH1 mRNA and a 3' UTR derived from MYH1 mRNA (MUTR7).

In some examples, the 5' UTR derived from MYH1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 47. In some examples, the 5' UTR derived from MYH1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 47. In some examples, the 5' UTR derived from MYH1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 47.

In some examples, the 3' UTR derived from MYH1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 48. In some examples, the 3' UTR derived from MYH1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 48. In some examples, the 3' UTR derived from MYH1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 48.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYH1 mRNA and a 3' UTR derived from MYH1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYH1 mRNA and a 3' UTR derived from MYH1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYH1 mRNA and a 3' UTR derived from MYH1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYH1 mRNA and a 3' UTR derived from MYH1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or

more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR7 (i.e. SEQ ID NOs: 47 and 48) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as described herein in muscle cells.

In some examples, the UTR pair of MUTR7 (i.e. SEQ ID NOs: 47 and 48) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as

described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYH1 mRNA (SEQ ID NO: 47) and a 3' UTR derived from MYH1 mRNA (SEQ ID NO: 48) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from myosin light chain 2 (MYL2) mRNA and a nucleic acid molecule encoding a 3' UTR derived from myosin light chain 2 (MYL2) mRNA (MUTR8).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from MYL2 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from MYL2 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from MYL2 mRNA and a 3' UTR derived from MYL2 mRNA (MUTR8).

In some examples, the 5' UTR derived from MYL2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 49. In some examples, the 5' UTR derived from MYL2 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 49. In some examples, the 5' UTR derived from MYL2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 49.

In some examples, the 3' UTR derived from MYL2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 50. In some examples, the 3' UTR derived from MYL2 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 50. In some examples, the 3' UTR derived from MYL2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 50.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL2 mRNA and a 3' UTR derived from MYL2 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL2 mRNA and a 3' UTR derived from MYL2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL2 mRNA and a 3' UTR derived from MYL2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from MYL2 mRNA and a 3' UTR derived from MYL2 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR8 (i.e. SEQ ID NOs: 49 and 50) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid

molecule sequence comprising a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein in muscle cells.

In some examples, the UTR pair of MUTR8 (i.e. SEQ ID NOs: 49 and 50) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from MYL2 mRNA (SEQ ID NO: 49) and a 3' UTR derived from MYL2 mRNA (SEQ ID NO: 50) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from troponin I2, fast skeletal type (TNNI2) mRNA and a nucleic acid molecule encoding a 3' UTR derived from troponin I2, fast skeletal type (TNNI2) mRNA (MUTR9).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from TNNI2 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from TNNI2 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from TNNI2and a 3' UTR derived from TNNI2 mRNA (MUTR9).

In some examples, the 5' UTR derived from TNNI2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 51. In some examples, the 5' UTR derived from TNNI2 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 51. In some examples, the 5' UTR derived from TNNI2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 51.

In some examples, the 3' UTR derived from TNNI2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 52. In some examples, the 3' UTR derived from TNNI2 mRNA comprises a nucleic acid molecule

sequence according to SEQ ID NO: 52. In some examples, the 3' UTR derived from TNNI2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 52.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNI2 mRNA and a 3' UTR derived from TNNI2 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNI2 mRNA and a 3' UTR derived from TNNI2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNI2 mRNA and a 3' UTR derived from TNNI2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from TNNI2 mRNA and a 3' UTR derived from TNNI2 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified

ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from creatine kinase, M-type (CKM) mRNA and a nucleic acid molecule encoding a 3' UTR derived from creatine kinase, M-type (CKM) mRNA (MUTR10).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from CKM mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from CKM mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from CKM mRNA and a 3' UTR derived from CKM mRNA (MUTR10).

In some examples, the 5' UTR derived from CKM mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 53. In some examples, the 5' UTR derived from CKM mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 53. In some examples, the 5' UTR derived from CKM mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 53.

In some examples, the 3' UTR derived from CKM mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 54. In some examples, the 3' UTR derived from CKM mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 54. In some examples, the 3' UTR derived from CKM mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 54.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CKM mRNA and a 3' UTR derived from CKM mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CKM mRNA and a 3' UTR derived from CKM mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CKM mRNA and a 3' UTR derived from CKM mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CKM mRNA and a 3' UTR derived from CKM mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR10 (i.e. SEQ ID NOs: 53 and 54) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein in muscle cells.

In some examples, the UTR pair of MUTR10 (i.e. SEQ ID NOs: 53 and 54) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CKM mRNA (SEQ ID NO: 53) and a 3' UTR derived from CKM mRNA (SEQ ID NO: 54) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from enolase 3 (ENO3) mRNA and a nucleic acid molecule encoding a 3' UTR derived from cytochrome c oxidase subunit 6A2 (COX6A2) mRNA (MUTR11).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ENO3 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from COX6A2 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from COX6A2 mRNA (MUTR11).

In some examples, the 5' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 45. In some examples, the 5' UTR derived from ENO3 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 45. In some examples, the 5' UTR derived from ENO3 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 45.

In some examples, the 3' UTR derived from COX6A2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 55. In some examples, the 3' UTR derived from COX6A2 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 55. In some examples, the 3' UTR derived from COX6A2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 55.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from COX6A2 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from COX6A2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from COX6A2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ENO3 mRNA and a 3' UTR derived from COX6A2 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of MUTR11 (i.e. SEQ ID NOs: 45 and 55) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein in muscle cells.

In some examples, the UTR pair of MUTR11 (i.e. SEQ ID NOs: 45 and 50) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ENO3 mRNA (SEQ ID NO: 45) and a 3' UTR derived from COX6A2 mRNA (SEQ ID NO: 55) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from orosomucoid 1 (ORM1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from orosomucoid 1 (ORM1) mRNA (LUTR1).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ORM1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from ORM1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from ORM1and a 3' UTR derived from ORM1 mRNA (LUTR1).

In some examples, the 5' UTR derived from ORM1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 56. In some examples, the 5' UTR derived from ORM1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 56. In some examples, the 5' UTR derived from ORM1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 56.

In some examples, the 3' UTR derived from ORM1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 57. In some examples, the 3' UTR derived from ORM1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 57. In some examples, the 3' UTR derived from ORM1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 57.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ORM1 mRNA and a 3' UTR derived from ORM1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ORM1 mRNA and a 3' UTR derived from ORM1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ORM1 mRNA and a 3' UTR derived from ORM1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from ORM1 mRNA and a 3' UTR derived from ORM1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR1 (i.e. SEQ ID NOs: 56 and 57) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein in muscle cells.

In some examples, the UTR pair of LUTR1 (i.e. SEQ ID NOs: 56 and 57) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule

sequence comprising a 5' UTR derived from ORM1 mRNA (SEQ ID NO: 56) and a 3' UTR derived from ORM1 mRNA (SEQ ID NO: 57) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from serum amyloid A2 (SAA2) mRNA and a nucleic acid molecule encoding a 3' UTR derived from serum amyloid A2 (SAA2) mRNA (LUTR2).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA2 mRNA.In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from SAA2 mRNA.In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA2 mRNA and a 3' UTR derived from SAA2 mRNA (LUTR2).

In some examples, the 5' UTR derived from SAA2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 58. In some examples, the 5' UTR derived from SAA2 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 58. In some examples, the 5' UTR derived from SAA2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 58.

In some examples, the 3' UTR derived from SAA2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 59. In some examples, the 3' UTR derived from SAA2 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 59. In some examples, the 3' UTR derived from SAA2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 59.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA2 mRNA and a 3' UTR derived from SAA2 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA2 mRNA and a 3' UTR derived from SAA2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA2 mRNA and a 3' UTR derived from SAA2 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA2 mRNA and a 3' UTR derived from SAA2 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR2 (i.e. SEQ ID NOs: 58 and 59) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein in muscle cells.

In some examples, the UTR pair of LUTR2 (i.e. SEQ ID NOs: 58 and 59) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA2 mRNA (SEQ ID NO: 58) and a 3' UTR derived from SAA2 mRNA (SEQ ID NO: 59) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from serum amyloid A4, constitutive (SAA4) mRNA and a nucleic acid molecule encoding a 3' UTR derived from serum amyloid A4, constitutive (SAA4) mRNA (LUTR3).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA4 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from SAA4 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from SAA4 mRNA (LUTR3).

In some examples, the 5' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 60. In some examples, the 5' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 60. In some examples, the 5' UTR derived from SAA4 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 60.

In some examples, the 3' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 61. In some examples, the 3' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 61. In some examples, the 3' UTR derived from SAA4 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 61.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR3 (i.e. SEQ ID NOs: 60 and 61) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in muscle cells.

In some examples, the UTR pair of LUTR3 (i.e. SEQ ID NOs: 60 and 61) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA4 mRNA (SEQ ID NO: 60) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from apolipoprotein A1 (APOA1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from apolipoprotein A1 (APOA1) mRNA (LUTR4).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from APOA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from APOA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from APOA1 mRNA and a 3' UTR derived from APOA1 mRNA (LUTR4).

In some examples, the 5' UTR derived from APOA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 62. In some examples, the 5' UTR derived from APOA1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 62. In some examples, the 5' UTR derived from APOA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 62.

In some examples, the 3' UTR derived from APOA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 63. In some examples, the 3' UTR derived from APOA1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 63. In some examples, the 3' UTR derived from APOA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 63.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from APOA1 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from APOA1 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from APOA1 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from APOA1 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide

as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from complement factor H related 2 (CFHR2) mRNA and a nucleic acid molecule encoding a 3' UTR derived from serum amyloid A4, constitutive (SAA4) mRNA (LUTR5).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from CFHR2 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from SAA4 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from CFHR2 mRNA and a 3' UTR derived from SAA4 mRNA (LUTR5).

In some examples, the 5' UTR derived from CFHR2 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 64. In some examples, the 5' UTR derived from CFHR2 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 64. In some examples, the 5' UTR derived from CFHR2 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 64.

In some examples, the 3' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 61. In some examples, the 3' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 61. In some examples, the 3' UTR derived from SAA4 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 61.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CFHR2 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CFHR2 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CFHR2 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from CFHR2 mRNA and a 3' UTR derived from SAA4 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR5 (i.e. SEQ ID NOs: 64 and 61) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in muscle cells.

In some examples, the UTR pair of LUTR5 (i.e. SEQ ID NOs: 64 and 61) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from CFHR2 mRNA (SEQ ID NO: 64) and a 3' UTR derived from SAA4 mRNA (SEQ ID NO: 61) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from serum amyloid A1 (SAA1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from serum amyloid A1 (SAA1) mRNA (LUTR6).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from SAA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA1and a 3' UTR derived from SAA1 mRNA (LUTR6).

In some examples, the 5' UTR derived from SAA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 65. In some examples, the 5' UTR derived from SAA1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 65. In some examples, the 5' UTR derived from SAA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 65.

In some examples, the 3' UTR derived from SAA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 66. In some examples, the 3' UTR derived from SAA1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 66. In some examples, the 3' UTR derived from SAA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 66.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from SAA1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from SAA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from SAA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from SAA1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR6 (i.e. SEQ ID NOs: 65 and 66) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from SAA1 mRNA (SEQ ID NO: 66) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from SAA1 mRNA (SEQ ID NO: 66) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from SAA1 mRNA (SEQ ID NO: 66) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from SAA1 mRNA (SEQ ID NO: 66) as described herein in muscle cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from GC, vitamin D binding protein (GC) mRNA and a nucleic acid molecule encoding a 3' UTR derived from GC, vitamin D binding protein (GC) mRNA (LUTR7).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from GC mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from GC mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from GC mRNA and a 3' UTR derived from GC mRNA (LUTR7).

In some examples, the 5' UTR derived from GC mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 67. In some examples, the

5' UTR derived from GC mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 67. In some examples, the 5' UTR derived from GC mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 67.

In some examples, the 3' UTR derived from GC mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 68. In some examples, the 3' UTR derived from GC mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 68. In some examples, the 3' UTR derived from GC mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 68.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from GC mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from GC mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from GC mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from GC mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a

preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR7 (i.e. SEQ ID NOs: 67 and 68) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in muscle cells.

In some examples, the UTR pair of LUTR7 (i.e. SEQ ID NOs: 67 and 68) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from GC mRNA (SEQ ID NO: 67) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from alpha 2-HS glycoprotein (AHSG) mRNA and a nucleic acid molecule encoding a 3' UTR derived from alpha 2-HS glycoprotein (AHSG) mRNA (LUTR8).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from AHSG mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from AHSG mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from AHSG mRNA and a 3' UTR derived from AHSG mRNA (LUTR8).

In some examples, the 5' UTR derived from AHSG mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 69. In some examples, the 5' UTR derived from AHSG mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 69. In some examples, the 5' UTR derived from AHSG mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 69.

In some examples, the 3' UTR derived from AHSG mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 70. In some examples, the 3' UTR derived from AHSG mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 70. In some examples, the 3' UTR derived from AHSG mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 70.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from AHSG mRNA and a 3' UTR derived from AHSG mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from AHSG mRNA and a 3' UTR derived from AHSG mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from AHSG mRNA and a 3' UTR derived from AHSG mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a

preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from AHSG mRNA and a 3' UTR derived from AHSG mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR8 (i.e. SEQ ID NOs: 69 and 70) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ ID NO: 70) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ ID NO: 70) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ ID NO: 70) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ ID NO: 70) as described herein in muscle cells.

In some examples, the UTR pair of LUTR8 (i.e. SEQ ID NOs: 69 and 70) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ

ID NO: 70) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ ID NO: 70) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ ID NO: 70) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from AHSG mRNA (SEQ ID NO: 69) and a 3' UTR derived from AHSG mRNA (SEQ ID NO: 70) as described herein in liver cells.

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from serum amyloid A4, constitutive (SAA4) mRNA and a nucleic acid molecule encoding a 3' UTR derived from apolipoprotein A1 (APOA1) mRNA (LUTR9).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA4 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from APOA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from APOA1 mRNA (LUTR9).

In some examples, the 5' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 60. In some examples, the 5' UTR derived from SAA4 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 60. In some examples, the 5' UTR derived from SAA4 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 60.

In some examples, the 3' UTR derived from APOA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 63. In some examples, the 3' UTR derived from APOA1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 63. In some examples, the 3' UTR derived from APOA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 63.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA4 mRNA and a 3' UTR derived from APOA1 mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided a UTR pair comprising a nucleic acid molecule encoding a 5' UTR derived from serum amyloid 1 (SAA1) mRNA and a nucleic acid molecule encoding a 3' UTR derived from GC, vitamin D binding protein (GC) mRNA (LUTR10).

In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA1 mRNA. In some examples, there is provided a nucleic acid molecule comprising a 3' UTR derived from GC mRNA. In some examples, there is provided a nucleic acid molecule comprising a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from GC mRNA (LUTR10).

In some examples, the 5' UTR derived from SAA1 mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 65. In some examples, the 5' UTR derived from SAA1 mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 65. In some examples, the 5' UTR derived from SAA1 mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 65.

In some examples, the 3' UTR derived from GC mRNA comprises a nucleic acid molecule sequence comprising a nucleic acid molecule sequence having at least 70%, 75%, 80%, 85%, 90%, 95%, 99% or 100% sequence identity to SEQ ID NO: 68. In some examples, the 3' UTR derived from GC mRNA comprises a nucleic acid molecule sequence according to SEQ ID NO: 68. In some examples, the 3' UTR derived from GC mRNA consists of a nucleic acid molecule sequence according to SEQ ID NO: 68.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, there is provided a DNA molecule encoding an optimised mRNA comprises one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above and comprising a nucleic acid molecule sequence encoding a 5' UTR derived from SAA1 mRNA and a 3' UTR derived from GC mRNA as described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, there is provided an optimised mRNA molecule comprising one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected using one or more of the methods described above.

In some examples, the optimised mRNA is a therapeutic mRNA. In some examples, the optimised mRNA has been transcribed by in vitro transcription. In some examples, the optimised mRNA comprises one or more modified ribonucleotides. That is to say that one or more of the ribonucleotides of the optimised mRNA are replaced with a modified ribonucleotide as described herein. For example, one or more uracils in the optimised mRNA may be (N)1-methylpseudouridine (1-methylΨ).

In some examples, the UTR pair of LUTR10 (i.e. SEQ ID NOs: 65 and 68) as described above may increase efficiency of translation in muscle cells in particular. As such, a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in muscle cells. As such, there is provided use of a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in muscle cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in muscle cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in muscle cells.

In some examples, the UTR pair of LUTR10 (i.e. SEQ ID NOs: 65 and 68) as described above may increase efficiency of translation in liver cells in particular. As such, a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in liver cells. As such, there is provided use of a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in liver cells.

In some examples, an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein may be for use in liver cells. As such, there is provided use of an optimised mRNA or DNA encoding such an optimised mRNA comprising a nucleic acid molecule sequence comprising a 5' UTR derived from SAA1 mRNA (SEQ ID NO: 65) and a 3' UTR derived from GC mRNA (SEQ ID NO: 68) as described herein in liver cells.

Optimised Poly(A) Tail and 5' cap

The optimised mRNAs or DNAs encoding such mRNAs provided herein may include an optimised poly (A) tail. For example, the optimised poly(A) may comprise at least about 90 adenine residues. In some examples, the optimised poly(A) tail comprises at least about 95 adenine residues. In some examples, the optimised poly(A) tail comprises at least about 97 adenine residues. In some examples, the optimised poly(A) tail comprises from about 95 to about 135 adenine residues. In some examples, the optimised poly(A) tail comprises from about 97 to about 135 adenine residues. In some examples, the optimised poly(A) tail comprises from about 95 to about 100 adenine residues. In some examples, the optimised poly(A) tail comprises from about 97 to about 100 adenine residues. In some examples, the optimised poly(A) tail comprises about 97 adenine residues.

In some examples, the optimised poly(A) tail may be added using E.coli poly(A) polymerase. In such cases, the poly(A) tail length may be defined by the number of E.coli poly(A) polymerase units. For example, the optimised poly(A) tail comprises from 5 to 10 E.coli poly(A) polymerase units. In some examples, the optimised poly(A) tail comprises 5 E.coli poly(A) polymerase units.

In some examples, the methods described herein comprise enzymatically adding a poly(A) tail to the optimised mRNAs as described herein. As such, the optimised mRNAs as described herein may include an enzymatically added poly(A) tail.

In some examples, the optimised mRNAs as described herein may include a 5' cap. In particular, the 5' cap may be m7G-5'ppp5'GpG. In particular, a m7G-5'ppp5'GpG 5' cap may provide improved translation efficiency and/or fidelity in muscle cells. As such, an mRNA including the 5' cap m7G-5'ppp5'GpG may be for use in muscle cells. For example, in muscle cells of a subject.

Tandem (Consecutive) Stop Codons

The optimised mRNAs or DNAs encoding said mRNAs provided herein may include at least two consecutively arranged stop codons. That is to say the optimised mRNA or DNA encoding said mRNA may include two stop codons adjacent to each other. Two adjacent stop codons may be referred to herein as a tandem stop codon. Two consecutive stop codons may be referred to herein as a tandem stop codon.

The tandem stop codon may be encoded in an alternative reading frame to a protein encoded by the optimised mRNA or DNA encoding said mRNA. Proteins are translated by reading tri-nucleotides (codons) from the 5' to the 3' end, starting with the amino acid methionine as the start (initiation) codon. Each codon is translated into a single amino acid. The code itself is degenerate, meaning that a particular amino acid can be specified by more than one codon. A shift of any number of nucleotides that is not divisible by 3 in the reading frame will cause subsequent codons to be read differently from that of the intended or correct reading frame. This effectively changes the ribosomal reading frame leading to the production of alternative polypeptides encoded by mRNA. The translation of a frameshifted codon may be referred to as out-of-frame translation, and the products (i.e. alternative proteins or polypeptides) may be referred to as out-of-frame products or proteins or as alternative products or proteins (i.e. alternative to the product encoded by the non-frameshifted (correct or in frame translation product) open reading frame). Out-of-frame and/or alternative translation products may be more immunogenic in comparison to the in-frame translation product. The out-of-frame and/or alternative translation products may also have a reduced efficacy in comparison to the in-frame translation product.

The tandem stop codon may be encoded in a frameshifted reading frame in comparison to the reading frame for production of the desired protein encoded by the optimised mRNA. For example, the tandem stop codon may be encoded in a -1, -2, +1, or +2 frameshifted reading frame. In some examples, the tandem stop codon may be encoded in a +1 frameshifted reading frame.

The tandem stop codon may include any two of the three stop codons: UAA; UAG; or UGA. In some examples, the tandem stop codon comprise a nucleic acid sequence selected from one of UAAUAA, UAAUAG, UAGUA, or UAGUAG.

In some examples, the optimised mRNAs or DNAs encoding said mRNAs provided herein comprise more than one tandem stop codon. For example, 1, 2, 3, 4, or 5 tandem stop codons. Tandem stop codons may be encoded out-of-frame at any point in the nucleic acid sequence where an alternative product may be produced due to frameshifted translation.

In some examples, the tandem stop codon comprises one or more N1-methyl pseudouridine. In some examples, only one of the consecutive stop codons comprises N1-methyl

pseudouridine. In some examples, both consecutive stop codons comprise at least one N1-methyl pseudouridine.

The tandem stop codons may reduce the production of unwanted polypeptides due to ribosomal slippage and/or frameshifting. As such, the tandem stop codons may help to reduce the immunogenicity of the nucleic acid, such as an mRNA. In some examples, the tandem stop codons may reduce the production of polypeptides that elicit a T-cell response. In some examples, the tandem stop codons may reduce the production of functional T cell antigens.

Combinations

It will be understood that optimised mRNAs or nucleic acid molecules encoding such mRNAs may include one or more of the preferred codons, signal peptides, UTR pairs, optimised poly(A) tail, tandem stop codons and/or 5' cap as described herein.

For example, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein and one or more of the signal peptides, UTR pairs, optimised poly(A) tail and/or 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein and a nucleic acid sequence encoding signal peptide as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein and an UTR pair as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein and an optimised poly(A) tail as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described

herein and at least one tandem stop codon as described herein. In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein and an UTR pair as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein and an optimised poly(A) tail as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein and at least one tandem stop codon as described herein. In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an UTR pair as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an UTR pair as described herein and an optimised poly(A) tail as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an UTR pair as described herein and at least one tandem stop codon as described herein. In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an optimised poly(A) tail as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include at least one tandem stop codon and a 5' cap as described herein as described herein. In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein and an UTR pair as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein and optimised poly(A) tail as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein and at least one tandem stop codon as described herein. In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule)) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, an UTR pair as described herein and optimised poly(A) tail as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, an UTR pair as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, an UTR pair as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, optimised poly(A) tail as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide

as described herein, an UTR pair as described herein and optimised poly(A) tail as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, optimised poly(A) tail as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an UTR pair as described herein, optimised poly(A) tail as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an UTR pair as described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an UTR pair as described herein, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include optimised poly(A) tail as described herein, at least one tandem stop codon described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein and optimised poly(A) tail as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, optimised poly(A) tail as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, an UTR pair as described herein, optimised poly(A) tail as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, an UTR pair as described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described

herein, an UTR pair as described herein, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, at least one tandem stop codon described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein, optimised poly(A) tail and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein, optimised poly(A) tail and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein, at least one tandem stop codon and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include a nucleic acid sequence encoding signal peptide as described herein, optimised poly(A) tail as described herein, at least one tandem stop codon and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include an UTR pair as described herein, optimised poly(A) tail as described herein, at least one tandem stop codon and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein an UTR pair as described herein, optimised poly(A) tail as described herein and at least one tandem stop codon as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein an UTR pair as

described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, optimised poly(A) tail, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, an UTR pair as described herein, optimised poly(A) tail, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that includes a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein, optimised poly(A) tail, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein, optimised poly(A) tail, at least one tandem stop codon as described herein and a 5' cap as described herein.

In some examples, there is provided nucleic acid molecules encoding an optimised mRNA (e.g. a DNA or RNA molecule) that include one or more preferred codons as described herein, a nucleic acid sequence encoding signal peptide as described herein, an UTR pair as described herein, optimised poly(A) tail as described herein and a 5' cap as described herein.

Medical Uses

The nucleic acid molecules described herein (e.g. DNA transcription templates or optimized mRNAs) may be used for a number of in vitro and in vivo uses. Reference to medicals uses,

nucleic acid molecules described herein for use in methods of treatment and methods of manufacturing a medicament using the nucleic acid molecules described herein are all to be understood to relate to methods of treating a subject using the nucleic acid molecules described herein. Thus in one example, there is provided a nucleic acid molecule as described herein for manufacture of a medicament for treating any of the disorders, conditions or diseases described herein. In one example, there is provided a nucleic acid molecule as described herein for use in treating any of the disorders, conditions or diseases described herein. In one example, there is provided a nucleic acid molecule as described herein for treating any of the disorders, conditions or diseases described herein. In one example, there is provided a method of treating a subject in need thereof comprising administering a nucleic acid molecule as described herein. For example, the method is a method for treating any of the disorders, conditions or diseases described herein.

The nucleic acid molecules described herein may be for use as medicaments. For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of preventing or treating a disease or condition in a subject. The disease or condition treated may depend on the protein encoded by the mRNA or DNA template described herein. In general, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in treating any diseases or conditions that may benefit from the administration of an mRNA or protein translated therefrom. The optimized mRNAs or DNA templates encoding said mRNAs may be for use or used in methods of treating infectious diseases (such as bacterial infections, viral infections, parasitic infections), cell proliferation disorders (such as cancer), genetic disorders, inflammatory disease, cardiovascular disorders, metabolic diseases, allergic disease, neurodegenerative diseases, protein or enzyme deficiency disorder and/or autoimmune diseases.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating a genetic disorder. “Genetic disorder” refers to a congenital or acquired disease caused by chromosomal or mitochondrial DNA abnormalities, and examples thereof include Down syndrome, Wilson disease, Edwards syndrome, Patau syndrome, Turner syndrome, Klinefelter syndrome, Apert syndrome, Crouzon syndrome, 22q11.2 deletion syndrome, Williams syndrome, Laurence-Moon-Biedl syndrome, Prader-Willi syndrome, Angelman syndrome, Kallmann syndrome, Aicardi-Goutieres syndrome, Miller-Dieker syndrome, Rubinstein-Taybi syndrome, Cornelia de Lange syndrome, cri-du-chat syndrome, super female, super male, and mitochondrial disease.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating a protein or enzyme deficiency disorder.

"Protein or enzyme deficiency disorders" refers to any disease or disorder that is associated with a subject lacking one or more proteins or enzymes or lacking sufficient activity of a protein or enzyme leading to symptoms and adverse effects in the subject. Examples of protein or enzyme deficiency disorders include Pompe disease, mucopolysaccharidosis types I, II, and VI, hemophilias A and B hypercystinia, Danon's disease, myoclonic renal failure syndrome, sialic acid storage disorders such as ISSD, Salla disease and moderately severe Salla disease, Niemann-Pick disease C1 and C2, mucolipid accumulation disease type IV; neuronal ceroid lipofuscinosis includes but not limited to ceroid lipofuscinosis 1 Types (Haltia-Santavuori disease and INCL), neuronal ceroid lipofuscinosis type 2 (Jansky-Bielschowsky disease), cereoid lipofuscinosis type 3 (Batten-Spielmeyer-Sjogren disease), waxy Lipofuscinosis type 4 (Parry's disease and Kufs A and B), cereoid lipofuscinosis type 5 (late infant Finnish type), cereofuscinosis type 6 (Lake-Cavanagh) Or Indian type), cereoid lipofuscinosis type 7 (Turkish type), cereoid lipofuscinosis type 8 (Northern epilepsy, epilepsy mental disorder), cereoid lipofuscinosis 9, cereoid lipofuscinosis 10, cereoid lipofuscinosis 11, cereoid lipofuscinosis 12, cereoid Lipofuscinosis 13, Cereofuscinosis 14; Lysosomal-related organelle disorders including but not limited to Hermansky-Pudlak disease type 1, Hermansky-Pudlak disease type 2, Hermansky-Pudlak disease type 3, Hermansky-Pudlak disease type 4, Hermansky-Pudlak disease type 5, Hermansky-Pudlak disease type 6, Hermansky-Pudlak disease type 7, Hermansky-Pudlak disease type 8, Hermansky-Pudlak disease type 9, Griscelli syndrome 1 (Elejalde syndrome)), Griscelli syndrome 2, Chédiak–Higashi disease, lysosomal storage disorders, e.g. Hurler syndrome, Niemann-Pick disease, Tay-Sachs disease, Gaucher disease, Fabry disease or Krabbe disease; Phenylketonuria; mitochondrial disorders; Friedreich ataxia; peroxisomal disorders, e.g. Zellweger syndrome or Adrenoleukodystrophy; metal metabolism disorders, e.g. Wilson disease or hemochromatosis; organic acidemias, e.g. methylmalonic acidemia or propionic acidemia; urea cycle disorders, e.g. ornithine transcarbamylase deficiency or citrullinemia and/or β-thalassemia. Other examples of enzyme deficiency disorders include Type I diabetes mellitus, which results from the patient's failure to produce insulin.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating cardiovascular disorders. "Cardiovascular disease" or "cardiovascular disorder" refers to diseases affecting the heart or blood vessels or both. For example, cardiovascular disease includes arrhythmia (atrial or ventricular or both); atherosclerosis and its sequelae; angina; cardiac rhythm disturbances; myocardial ischemia; myocardial infarction; cardiac or vascular aneurysm; vasculitis, stroke; peripheral obstructive arteriopathy of a limb, an organ, or a tissue; reperfusion injury following ischemia of the brain, heart, kidney or other organ or tissue; endotoxic, surgical, or traumatic shock;

hypertension, valvular heart disease, heart failure, abnormal blood pressure; shock; vasoconstriction (including that associated with migraines); vascular abnormality, insufficiency limited to a single organ or tissue.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating an autoimmune disease. “Autoimmune disease” refers to a disease or condition in which a subject's immune system has an aberrant immune response against a substance that does not normally elicit an immune response in a healthy subject. Examples of autoimmune diseases that may be treated include Acute Disseminated Encephalomyelitis (ADEM), Acute necrotizing hemorrhagic leukoencephalitis, Addison's disease, Agammaglobulinemia, Alopecia areata, Amyloidosis, Ankylosing spondylitis, Anti-GBM/Anti-TBM nephritis, Antiphospholipid syndrome (APS), Autoimmune angioedema, Autoimmune aplastic anemia, Autoimmune dysautonomia, Autoimmune hepatitis, Autoimmune hyperlipidemia, Autoimmune immunodeficiency, Autoimmune inner ear disease (AIED), Autoimmune myocarditis, Autoimmune oophoritis, Autoimmune pancreatitis, Autoimmune retinopathy, Autoimmune thrombocytopenic purpura (ATP), Autoimmune thyroid disease, Autoimmune urticaria, Axonal or neuronal neuropathies, Balo disease, Behcet's disease, Bullous pemphigoid, Cardiomyopathy, Castleman disease, Celiac disease, Chagas disease, Chronic fatigue syndrome, Chronic inflammatory demyelinating polyneuropathy (CIDP), Chronic recurrent multifocal ostomyelitis (CRMO), Churg-Strauss syndrome, Cicatricial pemphigoid/benign mucosal pemphigoid, Crohn's disease, Cogans syndrome, Cold agglutinin disease, Congenital heart block, Coxsackie myocarditis, CREST disease, Essential mixed cryoglobulinemia, Demyelinating neuropathies, Dermatitis herpetiformis, Dermatomyositis, Devic's disease (neuromyelitis optica), Discoid lupus, Dressler's syndrome, Endometriosis, Eosinophilic esophagitis, Eosinophilic fasciitis, Erythema nodosum, Experimental allergic encephalomyelitis, Evans syndrome, Fibromyalgia, Fibrosing alveolitis, Giant cell arteritis (temporal arteritis), Giant cell myocarditis, Glomerulonephritis, Goodpasture's syndrome, Granulomatosis with Polyangiitis (GPA) (formerly called Wegener's Granulomatosis), Graves' disease, Guillain-Barre syndrome, Hashimoto's encephalitis, Hashimoto's thyroiditis, Hemolytic anemia, Henoch-Schonlein purpura, Herpes gestationis, Hypogammaglobulinemia, Idiopathic thrombocytopenic purpura (ITP), IgA nephropathy, IgG4-related sclerosing disease, Immunoregulatory lipoproteins, Inclusion body myositis, Interstitial cystitis, Juvenile arthritis, Juvenile diabetes (Type 1 diabetes), Juvenile myositis, Kawasaki syndrome, Lambert-Eaton syndrome, Leukocytoclastic vasculitis, Lichen planus, Lichen sclerosus, Ligneous conjunctivitis, Linear IgA disease (LAD), Lupus (SLE), Lyme disease, chronic, Meniere's disease, Microscopic polyangiitis, Mixed connective tissue disease (MCTD), Mooren's ulcer,

Mucha-Habermann disease, Multiple sclerosis, Myasthenia gravis, Myositis, Narcolepsy, Neuromyelitis optica (Devic's), Neutropenia, Ocular cicatricial pemphigoid, Optic neuritis, Palindromic rheumatism, PANDAS (Pediatric Autoimmune Neuropsychiatric Disorders Associated with Streptococcus), Paraneoplastic cerebellar degeneration, Paroxysmal nocturnal hemoglobinuria (PNH), Parry Romberg syndrome, Parsonnage-Turner syndrome, Pars planitis (peripheral uveitis), Pemphigus, Peripheral neuropathy, Perivenous encephalomyelitis, Pernicious anemia, POEMS syndrome, Polyarteritis nodosa, Type I, II, & III autoimmune polyglandular syndromes, Polymyalgia rheumatica, Polymyositis, Postmyocardial infarction syndrome, Postpericardiotomy syndrome, Progesterone dermatitis, Primary biliary cirrhosis, Primary sclerosing cholangitis, Psoriasis, Psoriatic arthritis, Idiopathic pulmonary fibrosis, Pyoderma gangrenosum, Pure red cell aplasia, Raynauds phenomenon, Reactive Arthritis, Reflex sympathetic dystrophy, Reiter's syndrome, Relapsing polychondritis, Restless legs syndrome, Retroperitoneal fibrosis, Rheumatic fever, Rheumatoid arthritis, Sarcoidosis, Schmidt syndrome, Scleritis, Scleroderma, Sjogren's syndrome, Sperm & testicular autoimmunity, Stiff person syndrome, Subacute bacterial endocarditis (SBE), Susac's syndrome, Sympathetic ophthalmia, Takayasu's arteritis, Temporal arteritis/Giant cell arteritis, Thrombocytopenic purpura (TTP), Tolosa-Hunt syndrome, Transverse myelitis, Type 1 diabetes, Ulcerative colitis, Undifferentiated connective tissue disease (UCTD), Uveitis, Vasculitis, Vesiculobullous dermatosis, Vitiligo, or Wegener's granulomatosis (i.e., Granulomatosis with Polyangiitis (GPA).

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating a neurodegenerative diseases. “Neurodegenerative disease” refers to a disease or condition in which the function of a subject's nervous system becomes impaired. Examples of neurodegenerative diseases that may be treated include Alexander's disease, Alper's disease, Alzheimer's disease, Amyotrophic lateral sclerosis, Ataxia telangiectasia, Batten disease (also known as Spielmeyer-Vogt-Sjogren-Batten disease), Bovine spongiform encephalopathy (BSE), Canavan disease, Cockayne syndrome, Corticobasal degeneration, Creutzfeldt-Jakob disease, frontotemporal dementia, Gerstmann-Straussler-Scheinker syndrome, Huntington's disease, HIV-associated dementia, Kennedy's disease, Krabbe's disease, kuru, Lewy body dementia, Machado-Joseph disease (Spinocerebellar ataxia type 3), Multiple sclerosis, Multiple System Atrophy, Narcolepsy, Neuroborreliosis, Parkinson's disease, Pelizaeus-Merzbacher Disease, Pick's disease, Primary lateral sclerosis, Prion diseases, Refsum's disease, Sandhoffs disease, Schilder's disease, Subacute combined degeneration of spinal cord secondary to Pernicious Anaemia, Schizophrenia, Spinocerebellar ataxia (multiple

types with varying characteristics), Spinal muscular atrophy, Steele-Richardson-Olszewski disease, or Tabes dorsalis.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating a metabolic disease. “Metabolic disease” refers to a disease or condition in which a subject's metabolism or metabolic system (e.g., function of storing or utilizing energy) becomes impaired. Examples of metabolic diseases that may be treated include diabetes (e.g., type I or type II), obesity, metabolic syndrome, or a mitochondrial disease (e.g., dysfunction of mitochondria or aberrant mitochondrial function).

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating an inflammatory disease. “Inflammatory disease” refers to a disease or condition characterized by aberrant inflammation (e.g. an increased level of inflammation compared to a control, such as a healthy person not suffering from a disease). Examples of inflammatory diseases include traumatic brain injury, arthritis, rheumatoid arthritis, psoriatic arthritis, juvenile idiopathic arthritis, multiple sclerosis, systemic lupus erythematosus (SLE), myasthenia gravis, juvenile-onset diabetes, diabetes mellitus type 1, Guillain-Barre syndrome, Hashimoto's encephalitis, Hashimoto's thyroiditis, ankylosing spondylitis, psoriasis, Sjogren's syndrome, vasculitis, glomerulonephritis, auto-immune thyroiditis, Behcet's disease, Crohn's disease, ulcerative colitis, bullous pemphigoid, sarcoidosis, ichthyosis, Graves ophthalmopathy, inflammatory bowel disease, Addison's disease, Vitiligo, asthma, allergic asthma, acne vulgaris, celiac disease, chronic prostatitis, inflammatory bowel disease, pelvic inflammatory disease, reperfusion injury, sarcoidosis, transplant rejection, interstitial cystitis, atherosclerosis, and atopic dermatitis.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating cancer. The optimized mRNAs or DNA templates encoding said mRNAs provided herein may help to induce effective tumour-reactive T-cell responses to a tumour. When for use in treating cancer, the optimized mRNAs or DNA templates encoding said mRNAs encode a tumour-associated epitope. The optimized mRNAs or DNA templates encoding said mRNAs described herein may effectively help generate a population of immune cells, in particular of CD8+ effector T cells (also known as cytotoxic T lymphocytes (CTLs)). The immune cells induced by administration of the optimized mRNAs or DNA templates encoding said mRNAs described herein may be reactive to the epitope, or epitopes, translated from the optimized mRNAs or DNA templates encoding said mRNAs described herein. These immune cells are then primed for the killing of cancer cells that present the same or similar epitopes. Such optimized mRNAs or DNA templates encoding said mRNAs for use in treating cancer may be referred to as “cancer vaccines” or “cancer immunotherapy vaccines”.

The medical uses and methods of treating cancer may include administering to a subject in need thereof a therapeutically effective amount of an optimized mRNAs or DNA templates encoding said mRNAs as described herein.

The medical uses and methods of treatment described herein may be used in the treatment of a wide range of cancers. Tumours may be of mesenchymal or epithelial origin. Cancers include cancers of the colon, rectum, cervix, breast, lung, stomach, uterus, skin, mouth, tung, lips, larynx, kidney, bladder, prostate, brain, and blood cells. The medical uses and methods of treatment described herein may be used in the treatment of solid tumours.

Suitably, a cancer to be treated by a medical use or method of treatment described herein may be a solid tumour selected from but not limited to the group consisting of: pancreatic ductal adenocarcinoma, pancreatic cancer; breast cancer; melanoma; non-small cell lung cancer; small cell lung cancer; nasopharyngeal cancer; hepatocellular cancer; colorectal cancer; oesophageal cancer; gastric cancer; anal cancer; small intestine cancer; mesothelioma; kidney cancer; renal cell carcinoma; bladder cancer; prostate cancer; ovarian cancer; vulval cancer; cervical cancer; penile cancer; uveal melanoma; retinoblastoma; sarcoma; osteosarcoma; glioblastoma; adrenocortical carcinoma; neuroblastoma; Wilms tumour; endometrial cancer; and thyroid cancer.

In reference to cancer, the terms " treatment" and " treating" should be taken as encompassing therapy undertaken in order to prevent, slow down, or reduce an undesired physiological change or disorder, such as the growth, development or spread of cancer. Beneficial or desired results include but are not limited to the alleviation of symptoms, diminishment of the extent of disease, stabilized state of disease (which is to say, a disease that is not worsening), delay or slowing of disease progression, de-staging the tumour (e.g., changing from borderline resectable to amendable for surgical resection), amelioration or palliation of the disease state, and remission (either partial or total).

Treatment may bring about prolonged survival as compared to expected survival if not receiving treatment. Alternatively, or additionally, treatment may provide a patient with an improved standard of life as compared to that which would be expected if not receiving treatment.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating allergic disease. An “allergic disease” refers to a condition caused by hypersensitivity of the immune system to typically harmless substances in the environment. Allergic diseases include but are not limited to, asthma, hypersensitivity lung diseases, rhinitis, rhino-conjunctivitis, rhinosinusitis, atopic eczema, contact dermatitis, allergic conjunctivitis (intermittent and persistent), vernal conjunctivitis

(hay fever), atopic keratoconjunctivitis, giant papillary conjunctivitis, urticaria (hives), angioedema, hypersensitivity pneumonitis, eosinophilic bronchitis, vasculitis, hypersensitivity vasculitis, antineutrophil cytoplasmic antibody (ANCA) associated vasculitis, Wegner's granulomatosis, Churg Strauss vasculitis, microscopic polyangiitis, temporal arteritis, celiac disease, mastocytosis, and anaphylaxis.

In particular, the optimized mRNAs or DNA templates encoding said mRNAs described herein may encode a protein that includes an allergenic epitope. It will be apparent that in the majority of cases that the allergenic epitope is an epitope from or derived from an allergen that causes allergy symptoms or allergic reaction in a subject.

In some examples, the optimized mRNAs or DNA templates encoding said mRNAs described herein that encode a protein that includes an allergenic epitope may be for use in methods of allergy immunotherapy (AIT). In some examples, subcutaneous allergy immunotherapy (SCIT). Optimized mRNAs or DNA templates encoding said mRNAs for use in AIT and/or SCIT may be referred to as allergy vaccines.

In general, AIT comprises administering an allergen to the patient in order to treat an allergy to that allergen of the patient, i.e., reducing current or future immune response, such as an allergen-specific IgE response and/or histamine release by mastocytes and/or granulocytes induced by the allergen, and/or manifestation of clinical symptoms of allergy. Immunotherapy is conventionally carried out by repeatedly administering a mono-dose or incremental doses of an allergen to a patient in need thereof, thereby resulting in an adaptive immune response of the patient who becomes desensitised to the allergen.

During AIT or SCIT, increasing doses of the allergen or allergenic epitope are administered, followed by a maintenance dose for several years, with the goal of inducing immunological changes leading to symptom amelioration while on therapy, as well as sustained desensitization off AIT or SCIT (immune tolerance).

Typically, at the start of AIT or SCIT, subjects receive increasing doses of the allergen or allergenic epitope at weekly intervals over several weeks to months, under tightly monitored medical supervision. The gradual dose escalation enables tolerability to therapy and mitigates the risk of severe hypersensitivity reactions related to allergen administration.

For example, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use in methods of treating infectious diseases. “Infectious disease” refers to a disease which results from an infection. Infection is a condition caused by the invasion of an organism by a foreign agent (i.e., an infectious agent). Infectious agents include but are not limited to, bacteria, fungi, viruses, viroids, nematodes (e.g., parasites such as roundworms and pinworms), arthropods (e.g., mites, fleas, lice, ticks), and macroparasites

(e.g., tapeworms). Common infectious diseases include bacterial and viral infections. The optimized mRNAs or DNA templates encoding said mRNAs described herein, when for use in methods of treating infectious disease, may include an epitope or antigen from or derived from the pathogen causing the infectious disease.

The optimized mRNAs or DNA templates encoding said mRNAs described herein may be particularly useful in prevention or treatment of infectious diseases caused by intracellular pathogens. For example, viruses (e.g., CMV, HIV, SARS viruses, such as COVID-19, coronaviruses), bacteria (e.g., Listeria, Mycobacteria, Salmonella (e.g., S. typhi) enteropathogenic Escherichia coli (EPEC), enterohaemorrhagic Escherichia coli (EHEC), Yersinia, Shigella, Chlamydia, Chlamydophila, Staphylococcus, Legionella), protozoa (e.g., Taxoplasma), fungi, and intracellular parasites (e.g., Plasmodium (e.g., P. vivax, P. falciparum, P. ovale, and P. malariae). The compositions and formulations described herein may reduce humoral response against an immunogenic immunomodulator, including epitopes derived from such intracellular pathogens and increase cellular mediated response.

"Treatment" in relation to infectious diseases refers to any administration of a therapeutic optimized mRNA as described herein that partially or completely alleviates, ameliorates, relieves, inhibits, delays the onset of, reduces the severity of and/or reduces the incidence of one or more symptoms or features of an infectious disease or the predisposition toward the disease. Such treatment may be of a subject who does not exhibit signs of the relevant disease, and/or of a subject who exhibits only early signs of the disease. Alternatively or additionally, such treatment may be of a subject who exhibits one or more established signs of the relevant disease. As such the term "treating" in reference to infectious diseases refers to the vaccination of a subject. “Prevention" refers to a delay of onset of an infectious disease. Prevention may be considered complete when onset of an infectious disease or disorder has been delayed for a predefined period of time.

In some examples, the optimized mRNAs or DNA templates encoding said mRNAs described herein may be for use as a vaccine. As such, in some examples provided herein are immunogenic compositions comprising a therapeutic optimized mRNA as described. In some examples, the compositions are vaccine compositions. The terms "immunogenic composition" and "immunological composition" and "immunogenic or immunological composition" refer to compositions that elicit an immune response against an antigen or immunogen after administration into a subject. The terms "vaccine" and "vaccine composition" refers to compositions that induce a protective immune response against the antigen of interest or which efficaciously protects against the antigen; for instance, after administration to the subject, elicits a protective immune response against the targeted antigen or immunogen.

In some examples, the methods of preventing and/or treating infectious disease is a method of vaccination. "Vaccination" refers to the administration of a therapeutic optimized mRNA as described herein intended to generate an immune response, for example to a disease-causing pathogen. Vaccination can be administered before, during, and/or after exposure to a disease-causing pathogen, and in some examples, before, during, and/or shortly after exposure to the agent. In some examples, vaccination includes multiple administrations, appropriately spaced in time, of a therapeutic optimized mRNA as described herein.

In some examples, the optimized mRNAs or DNA templates encoding said mRNAs may be for use in vaccination against a virus. For example, for use in vaccination of a subject against viruses of the retroviridae, orthmyxoviridae, paramyxoviridae, arenaviridae, bunyaviridae, flaviviridae, filoviridae, togaviridae, picornaviridae, caliciviridae and coronaviridae families. Examples of such viruses include, but are not limited to, adenovirus, rhinovirus, hepatitis, immunodeficiency virus, polio, measles, Ebola, Coxsackie, Rhino, West Nile, small pox, encephalitis, yellow fever, Dengue fever, influenza (including human, avian, and swine), lassa, lymphocytic choriomeningitis, junin, machuppo, guanarito, hantavirus, Rift Valley Fever, La Crosse, California encephalitis, Crimean-Congo, Marburg, Japanese Encephalitis, Kyasanur Forest, Venezuelan equine encephalitis, Eastern equine encephalitis, Western equine encephalitis, severe acute respiratory syndrome (SARS), parainfluenza, respiratory syncytial, Punta Toro, Tacaribe and pachindae.

In some examples, the virus is an influenza (including human, avian, and swine) or a severe acute respiratory syndrome (SARS) virus. In some examples, the virus is a coronaviridae virus. In some examples, the virus is Covid-19.

As used herein, the terms “treat”, “treating” and "treatment" generally are taken to include an intervention performed with the intention of preventing the development or altering the pathology of a condition, disorder or symptom (e.g., an allergic disease, infectious disease, etc.). Accordingly, "treatment" refers to both therapeutic treatment and prophylactic or preventative measures (such as vaccination), wherein the object is to prevent or slow down (lessen) the targeted condition, disorder or symptom. “Treatment” therefore encompasses a reduction, slowing or inhibition of disease symptoms, for example, at least 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or 100% when compared to before treatment.

As used herein the term “subject” generally refers to an individual, e.g., a human, having or at risk of having a specified condition, disorder or symptom. The subject may be a patient, i.e., a subject in need of treatment in accordance with the invention. The subject may have received treatment for the condition, disorder or symptom. Alternatively, the subject has not been treated prior to treatment in accordance with the present invention.

The optimized mRNAs or DNA templates encoding said mRNAs described herein generally can be administered to the subject by any conventional route, including injection or by gradual infusion over time. The administration may, for example, by intramuscular, intravascular, intracavity, intracerebral, intralesional, rectal, subcutaneous, intradermal, epidural, intrathecal, and percutaneous administration.

The methods of treatment and medical uses described herein may provide optimized mRNAs or DNA templates encoding said mRNAs as described herein to a recipient via any suitable route of administration.

The optimized mRNAs or DNA templates encoding said mRNAs can be administered via any desired route of administration. The therapeutic optimized mRNAs, or medical uses, may make use of a route of administration selected from the group consisting of: intravenous (iv) administration; subcutaneous (sc) administration; intramuscular (im) administration; intradermal (id) administration; sublingual (sl) administration; and intranasal administration.

The skilled person will be able to determine suitable forms of the optimized mRNAs or DNA templates encoding said mRNAs of the invention for use with the desired route of administration.

In some examples, the optimized mRNAs or DNA templates encoding said mRNAs as described may be administered via a route selected from intratumoral, inhalation, or intracardiac injection.

The optimized mRNAs or DNA templates encoding said mRNAs described herein are for administration in an effective amount. An “effective amount” is an amount that alone, or together with further doses, produces the desired (therapeutic or non-therapeutic) response. The effective amount to be used will depend, for example, upon the therapeutic (or non-therapeutic) objectives, the route of administration, and the condition of the patient/subject. For example, the suitable dosage of a therapeutic optimized mRNA of the invention for a given patient/subject will be determined by the attending physician (or person administering the composition), taking into consideration various factors known to modify the action of the optimized mRNAs or DNA templates encoding said mRNAs of the invention for example severity and type of disease, body weight, sex, diet, time and route of administration, other medications and other relevant clinical factors. The dosages and schedules may be varied according to the particular condition, disorder or symptom of the overall condition of the patient/subject. Effective dosages may be determined by either in vitro or in vivo methods.

As detailed above, certain preferred codons, signal peptides and/or UTR pairs may be more effective in certain cell types. As such, it will be understood that administration of the nucleic acid molecules described herein may be to particular target cells depending on which cell

the preferred codons, signal peptides and/or UTR pairs provide the greatest increase in translation efficiency and/or fidelity.

As such, there is provided herein a nucleic acid molecule comprising a nucleic acid molecule sequence encoding one or more codons for translation to one or more amino acids, wherein at least one codon comprises a muscle cell preferred codon as described herein for use as described herein wherein the nucleic acid molecule is administered to a subject's muscle cells.

As such, there is provided herein a nucleic acid molecule comprising a nucleic acid molecule sequence encoding one or more codons for translation to one or more amino acids, wherein at least one codon comprises a kidney cell preferred codon as described herein for use as described herein wherein the nucleic acid molecule is administered to a subject's kidney cells.

As such, there is provided herein a nucleic acid molecule comprising a nucleic acid molecule sequence encoding an AACT signal peptide as described herein wherein the nucleic acid molecule is administered to a subject's liver cells.

As such, there is provided herein a nucleic acid molecule comprising a nucleic acid molecule sequence encoding an HSA signal peptide as described herein wherein the nucleic acid molecule is administered to a subject's muscle cells.

As such, there is provided herein a nucleic acid molecule comprising a nucleic acid molecule sequence encoding a:

a. 5' UTR is derived from a MYLPF mRNA and a 3' UTR is derived from a MYLPF mRNA (MUTR2);
b. 5' UTR is derived from a TNNC1 mRNA and a 3' UTR is derived from a ENO3 mRNA (MUTR4);
c. 5' UTR is derived from a ENO3 mRNA and a 3' UTR is derived from a ENO3 mRNA (MUTR6);
d. 5' UTR is derived from a MYH1 mRNA and a 3' UTR is derived from a MYH1 mRNA (MUTR7);
e. 5' UTR is derived from a MYL2 mRNA and a 3' UTR is derived from a MYL2 mRNA (MUTR8);
f. 5' UTR is derived from a CKM mRNA and a 3' UTR is derived from a CKM mRNA (MUTR10);

g. 5' UTR is derived from a ENO3 mRNA and a 3' UTR is derived from a COX6A2 mRNA (MUTR11);

h. 5' UTR is derived from a ORM1 mRNA and a 3' UTR is derived from a ORM1 mRNA (LUTR1);

i. 5' UTR is derived from a SAA2 mRNA and a 3' UTR is derived from a SAA2 mRNA (LUTR2);

j. 5' UTR is derived from a SAA4 mRNA and a 3' UTR is derived from a SAA4 mRNA (LUTR3);

k. 5' UTR is derived from a CFHR2 mRNA and a 3' UTR is derived from a SAA4 mRNA (LUTR5);

l. 5' UTR is derived from a SAA1 mRNA and a 3' UTR is derived from a SAA1 mRNA (LUTR6);

m. 5' UTR is derived from a GC mRNA and a 3' UTR is derived from a GC mRNA (LUTR7);

n. 5' UTR is derived from a AHSG mRNA and a 3' UTR is derived from a AHSG mRNA (LUTR8); or

o. 5' UTR is derived from a SAA1 mRNA and a 3' UTR is derived from a GC mRNA (LUTR10);

wherein the nucleic acid molecule is administered to a subject's muscle cells.

As such, there is provided herein a nucleic acid molecule comprising a nucleic acid molecule sequence encoding a:

a. 5' UTR is derived from a MYLPF mRNA and a 3' UTR is derived from a MYLPF mRNA (MUTR2);

b. 5' UTR is derived from a TNNC1 mRNA and a 3' UTR is derived from a ENO3 mRNA (MUTR4);

c. 5' UTR is derived from a TNNC1 mRNA and a 3' UTR is derived from a TNNC1 mRNA (MUTR5);

d. 5' UTR is derived from a ENO3 mRNA and a 3' UTR is derived from a ENO3 mRNA (MUTR6);

e. 5' UTR is derived from a MYH1 mRNA and a 3' UTR is derived from a MYH1 mRNA (MUTR7);

f. 5' UTR is derived from a MYL2 mRNA and a 3' UTR is derived from a MYL2 mRNA (MUTR8);

g. 5' UTR is derived from a CKM mRNA and a 3' UTR is derived from a CKM mRNA (MUTR10);

h. 5' UTR is derived from a ENO3 mRNA and a 3' UTR is derived from a COX6A2 mRNA (MUTR11);

i. 5' UTR is derived from a ORM1 mRNA and a 3' UTR is derived from a ORM1 mRNA (LUTR1);

j. 5' UTR is derived from a SAA2 mRNA and a 3' UTR is derived from a SAA2 mRNA (LUTR2);

k. 5' UTR is derived from a SAA4 mRNA and a 3' UTR is derived from a SAA4 mRNA (LUTR3);

l. 5' UTR is derived from a CFHR2 mRNA and a 3' UTR is derived from a SAA4 mRNA (LUTR5);

m. 5' UTR is derived from a SAA1 mRNA and a 3' UTR is derived from a SAA1 mRNA (LUTR6);

n. 5' UTR is derived from a GC mRNA and a 3' UTR is derived from a GC mRNA (LUTR7);

o. 5' UTR is derived from a AHSG mRNA and a 3' UTR is derived from a AHSG mRNA (LUTR8); or

p. 5' UTR is derived from a SAA1 mRNA and a 3' UTR is derived from a GC mRNA (LUTR10

wherein the nucleic acid molecule is administered to a subject's liver cells.

It will be understood that optimised mRNAs or nucleic acid molecules encoding such mRNAs may include one or more of the preferred codons, signal peptides, UTR pairs, optimised poly(A) tail and/or 5' cap as described herein.

Compositions

In some examples, the optimized mRNAs or DNA templates encoding said mRNAs as described herein, are formulated as compositions. In some examples, the optimized mRNAs or DNA templates encoding said mRNAs as described herein, may be provided as part of a

pharmaceutical formulation or composition. Advantageously, such formulations may be administered to a human subject in need thereof (as described elsewhere herein).

The compositions, including the optimized mRNAs or DNA templates encoding said mRNAs as described, may include delivery agents specific to the administration and delivery of mRNA to a subject. For example, the compositions may include optimized mRNAs or DNA templates encoding said mRNAs encapsulated in lipids, polymers, or dendrimers. In some examples, the compositions may include cell-penetrating peptides that may or may not be linked or may link in use (i.e. covalently or non-covalently) to the optimized mRNA or the translation product thereof. For example, optimized mRNAs or DNA templates encoding said mRNAs may be delivered to a target tissue or cell by the use of lipid nanoparticles, autologous T cells, CAR-T cells, plasmid DNA, modified CD34+ hHSPCs, cytotoxic T lymphocytes, or T cells. For example, see Kowalski PS, Rudra A, Miao L, Anderson DG. Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. Mol Ther. 2019;27(4):710-728. doi:10.1016/j.ymthe.2019.02.012 and Qin S, Tang X, Chen Y, et al. mRNA-based therapeutics: powerful and versatile tools to combat diseases. Signal Transduct Target Ther. 2022;7(1):166. Published 2022 May 21. doi:10.1038/s41392-022-01007-w.

A pharmaceutical formulation and the compositions described herein may comprise optimized mRNAs or DNA templates encoding said mRNAs as described herein, along with a pharmaceutically acceptable excipient, adjuvant, diluent and/or carrier.

Compositions and formulations may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives, compatible carriers, supplementary immune potentiating agents such as adjuvants and cytokines and optionally other therapeutic agents or compounds.

As used herein, "pharmaceutically acceptable" refers to a material that is not biologically or otherwise undesirable, i.e., the material may be administered to an individual along with the selected optimized mRNAs or DNA templates encoding said mRNAs without causing any undesirable biological effects or interacting in a deleterious manner with any of the other components of the pharmaceutical formulation in which it is contained.

Excipients are natural or synthetic substances formulated alongside an active ingredient (e.g. a optimized mRNAs or DNA templates encoding said mRNAs as provided herein), included for the purpose of bulking-up the formulation or to confer a therapeutic enhancement on the active ingredient in the final dosage form, such as facilitating drug absorption or solubility. Excipients can also be useful in the manufacturing process to aid in the handling of the active substance concerned, such as by facilitating powder flowability or

non-stick properties, in addition to aiding in vitro stability, such as prevention of denaturation over the expected shelf life. Pharmaceutically acceptable excipients are well known in the art. A suitable excipient is, therefore, easily identifiable by one of ordinary skill in the art. By way of example, suitable pharmaceutically acceptable excipients include water, saline, aqueous dextrose, glycerol, ethanol, and the like.

Adjuvants are pharmacological and/or immunological agents that modify the effect of other agents in a formulation. Pharmaceutically acceptable adjuvants are well-known in the art. A suitable adjuvant is, therefore, easily identifiable by one of ordinary skill in the art. Merely by way of example, a pharmaceutical formulation may comprise an adjuvant selected from the group consisting of: AS03; AddaS03; AS04; MF59; AddaVax; Poly I:C; R848; Cpg; virus-like particles; virosomes; MPL; and flagellin protein.

Diluents are diluting agents. Pharmaceutically acceptable diluents are well-known in the art. A suitable diluent is, therefore, easily identifiable by one of ordinary skill in the art.

Carriers are non-toxic to recipients at the dosages and concentrations employed and are compatible with other ingredients of the formulation. The term “carrier” denotes an organic or inorganic ingredient, natural or synthetic, with which the active ingredient is combined to facilitate the application. Pharmaceutically acceptable carriers are well known in the art. A suitable carrier is, therefore, easily identifiable by one of ordinary skill in the art.

Methods of designing Optimised mRNAs

Provided herein are methods of designing an optimised nucleic acid sequences (such as mRNAs) for translation. The methods may be implemented by a suitable computing device. In some examples, the methods may further include producing the design mRNAs using methods described above. For example, using de novo DNA synthesis and subsequent in vitro transcription.

Fig. 9 shows a method 900 of designing an optimised RNA sequence for translation according to methods described herein.

In step 902, method 900 comprises receiving a target translated protein sequence, one or more user-defined objectives for the optimised RNA sequence, and one or more parameters associated with each user-defined objective, whereby the one or more parameters correspond to the target translated protein sequence.

The objectives may be defined by a user and may be referred to as user defined objectives. The user defined objectives may include any desired use and/or outcome of the use of the

RNA. For example, user defined objectives may include one or more of target application, biological activity, expression profile, manufacturability, secretion profile and/or storability.

Target application may include one or more of administration route, end user purpose, target tissue type, target cell type, cellular localisation, and/or cellular processing.

In some examples, target application may include expression of the encoded protein in a specific target cell (such as a liver cell). In such cases, 5' and 3' UTRs may be those described herein that have been shown to have higher efficiency in liver cells (such as MUTR2, MUTR4, MUTR6, MUTR7, MUTR8, MUTR10, MUTR11, LUTR1, LUTR2, LUTR3, LUTR5, LUTR6, LUTR7, LUTR8, or LUTR10). In such cases, an aptamer specific to a target in liver cells may be selected.

In some examples, the target application may be a route of administration, such as intravenous, intramuscular, subcutaneous, rectal, vaginal, intrathecal, intraoccular, intranasal or inhalation administration. Administration route may be effected by any one of the parameters disclosed herein as well as by sequence elements such as UTRs (e.g. UTRs that are suited for certain cell types), aptamers (e.g. that help with uptake into specific cells such as cardiomyocytes – for example see Philippou, Styliana et al. "Selective Delivery to Cardiac Muscle Cells Using Cell-Specific Aptamers." Pharmaceuticals (Basel, Switzerland) vol. 16,9 1264. 6 Sep. 2023, doi:10.3390/ph16091264) and codon selection as described herein.

In some examples, the target application may be end user purpose. End user purpose refers to, the eventual use of the produced RNA molecule designed by the methods described herein. For example, end user purpose may include use of the RNA for vaccination, production of therapeutic cell lines, or production of a therapeutic protein. It will be understood that end user purposes may overlap. For example, vaccination may include the production of a therapeutic protein in a subject's target cells. However, production therapeutic proteins is not limited to vaccination and may include the production of therapeutics proteins to help treat or prevent diseases associated with the therapeutic protein encoded. For example, therapeutic proteins include antibodies, growth factors, cytokines, immune system proteins, enzymes, interferons, interleukins, and thrombolytics and the like.

In some examples, the target application may be cellular localisation. Cellular localisation refers to the localisation in a target cell of the RNA and/or encoded protein after production (i.e. translation) of the protein by the cell. For example, cellular localisation includes localisation to the cytoplasm, organelles or cellular compartments such as the nucleus, mitochondria, exosomes, plasma membrane, endoplasmic reticulum, peroxisomes,

lysosomes and vacuoles. Cellular localisation may be controlled by the presence of signal sequences within the RNA, for example, in the coding region and/or untranslated regions (e.g. 5' and/or 3' UTRs) of an RNA, and/or signal sequences in the encoded protein.

Cellular processing refers to the interaction of an RNA and/or the encoded protein with any of the processing machinery of a target cell.

Biological activity refers to the activity that the encoded protein or part thereof may have in a target cell. In some examples, biological activity includes one or more of immunogenicity, cellular interactions, adjuvant activity, enzymatic activity and/or cellular effects. For example, in the cell of a subject to whom the RNA or encoded protein is administered. For example, biological activity may refer to the enzymatic activity of an encoded protein, effects on cells, effects on cellular pathways, cellular interactions, immunogenicity and/or adjuvant activity. Biological activity may be affected by one or more of the parameters described below such as codon selection, sequence elements, untranslated regions and other parameters described herein. Biological activity may also include unintended biological activity as well as intended biological activity. For example the generation of unintended translation products due to ribosomal frameshifting, the activation of biological pathways which can reduce protein synthesis in target cells, or misfolding of an encoded protein.

Expression profile may include one or more of expression duration, expression location and/or expression level. Expression duration refers to how long the RNA persists and is translated in target cells, which may be related to parameters such as mRNA secondary structure, codon usage, and GC content.

Expression levels refers to the levels of the encoded protein which can be achieved from a given dose of the designed RNA, which may be related to how efficiently mRNA can be processed by the cellular protein synthesis machinery.

Expression location where the RNA is expressed in a cell or subject. Expression location may be affected by one or more of the parameters described herein.

Secretion profile includes secretion pathway and/or secretion efficiency. Secretion pathway refers to the processing of the RNA and/or the encoded protein by a target cell. For example, processing of the RNA and/or encoded protein to or by cells via secretory vesicles such as exosomes and lysosomes.

Secretion efficiency refers to how efficiently the encoded protein is secreted (if required), which is related to the efficiency of the signal sequence and to RNA motifs in the open reading frame and untranslated regions.

Secretion may be affected by any one or more of the parameters described herein.

Manufacturability refers to the RNA yield from in vitro transcription of a given amount of template DNA, which is related (for example) to the presence of motifs that induce termination of the RNA polymerase, or that lead to polymerase drop-off or to polymerase stalling.

Storability refers to how stable the formulated mRNA is estimated to be at a given temperature and other conditions, and what other storage conditions need to be applied, which is related to the secondary structure content of the mRNA and its intended formulation. The one or more parameters include:

i. codon composition efficiency for translation of the RNA sequence to the target translated protein sequence;

ii. presence or absence of specified nucleotide repeats;

iii. GC content;

iv. desirability of specified nucleotide sub-sequences; and/or

v. location of one or more stop codons.

In some examples, the user defined objectives relate to desired properties of the final product designed using the methods described herein. The user defined objectives do not directly relate to sequence of nucleotides selected. In contrast, parameters relate to specific sequences of nucleotides which may have properties that help to achieve the user defined objectives. For example, if the user wishes to provide an optimised RNA or mRNA with improved stability and biological activity, the GC content may be a relatively low value. In addition, user objectives may be achieved by selection of sequence elements as described herein. For example, the user defined objective of expression in a target cell, sequence elements that are known or expected to provide expression and/or improved expression in the target cell may be selected. In addition, parameters that provide for expression and/or improved expression in the target cell may also be selected.

Codon composition efficiency for translation refers the selection of codons that are most efficiently transcribed, for example, based on the target application. Condon composition efficiency may be determined using codon decoding times, codon preference, codon usage frequency, codon usage patterns and/or abundance of tRNA isoacceptors and/or ratio of cognate to near-cognate tRNAs. For example, codon composition efficiency may be determined using the methods described above. In some examples, codon composition efficiency is linked to target application, such as target cell. For example, codons may be selected based on the preferred codons as described herein determined using methods as described herein.

Nucleotide repeat refers to a nucleic acid or a region of a nucleic acid comprising any short sequence of 1-20 nucleobase residues in length (e.g., a dinucleotide, trinucleotide, tetranucleotide, pentanucleotide, hexanucleotide sequence, etc.) wherein the short sequence is repeated 2 or more times (e.g., 2, 3, 4, 5, 10, 15, 20, 50, 100, 200, 500, or more repeats). For example, desirable repeats may include 5'UTRs and may help improve stability of the RNA and/or efficiency of translation and so may be desirable nucleotide repeats. For example, undesirable repeats may include repeats of modified uridines , may decrease stability of the RNA and/or efficiency of translation and so may be undesirable nucleotide repeats.

"GC content" is the fraction or percentage of total nucleobase residues in a nucleic acid sequence that are guanine residues, cytosine residues, or analogues thereof. For example, a 100 nt sequence that contains exactly 30 cytosines, exactly 30 guanines, exactly one cytosine analogues, and exactly one guanine analogues has a GC richness of 62%. Although high GC-content may cause problems for mRNA secondary structure, the higher GC sequence translates 100-fold higher than a low GC sequence. GC-rich sequences may also help reduce immunogenicity.

Nucleotide sub-sequences may include any one or more of frameshifting sequences, ribosomal slippery sequences, transcription terminator sequences, RNA polymerase stalling sequences, aptamer sequences, secondary structure forming sequences, restriction enzyme sites, RNA binding protein binding sites and/or frameshifted premature termination codons.

The term "frameshifting sequence" is used to refer to any nucleic acid sequences, such as an RNA sequence or DNA sequence encoding an optimised mRNA described herein, that may cause or increase the likelihood of frameshifting of a ribosome (i.e. translation or ribosomal frameshifting).

Frameshifting nucleic acid sequences may be any sequences that are known or predicted to lead to ribosomal frameshifting. Methods to identify or predict frameshifting nucleic acid sequences are known. For example, US20080103745A1 describes a model for predicting frameshifts. Other methods for predicting ribosomal frameshifts include those described in Moon, S. et al., LNCS, 2004, 3036: 334-34; Hammell, A. B. et al., Genomic Res., 1999, 9: 417-427); Bekaert, M. et al., Bioinformatics, 2003, 19: 327-335) and Shah, A. A. et al., Bioinformatics, 2002, 18: 1046-1053.

Known frameshifting nucleic acid sequences that may cause a +1 frameshift include: sequences of UUUUGA (SEQ ID NO: 73), UCCUGA (SEQ ID NO: 74) or CCCUGA (SEQ ID NO: 75); spacer components having a spacer with 4 to 11 nucleotides; and/or secondary structures capable of designating stem-loops or pseudoknots.

Known frameshifting nucleic acid sequences that may cause a −1 or +1 frameshift include sites that comprise sequentially a sequence of XXXYYYZ (SEQ ID NO: 76), wherein X is any nucleotide, wherein Y is A or U, and wherein Z is A, U, or C; space components with 4 to 11 nucleotides; and/or secondary structures component capable of designating stem-loops or pseudoknots.

In some examples, the frameshifting nucleic acid sequence may comprise a sequence of PPPX (SEQ ID NO: 77), wherein X is any nucleotide, and PPP is a trinucleotide repeat of any nucleotide. In some examples, the frameshifting nucleic acid sequence may comprise a sequence of m1Ψm1Ψm1ΨX (SEQ ID NO: 78), wherein X is any nucleotide and m1Ψ is (N)1-methylpseudouridine. In some examples, the frameshifting nucleic acid sequences may comprise a sequence of CUUAGG(SEQ ID NO: 79), CUUGAC (SEQ ID NO: 80), CAGCAG (SEQ ID NO: 81), or UCUGCGG (SEQ ID NO: 82).

The frameshifting nucleic acid sequences may, in some cases, be described as ribosomal slippery sequences. As such, in some examples, the modified therapeutic mRNAs of the invention may comprise one or more ribosomal slippery sequences.

Pseudoknots are secondary RNA substructures that contain two or more stem-loop motifs with intercalated stems. The pseudoknot or stem-loop structure in the mRNA is thought to result in pausing of the ribosome, resulting in eventual frameshifting.

In some examples, frameshifting nucleic acid sequence may also cause ribosomal stalling. In some examples, frameshifting nucleic acid sequences are any sequence that may cause ribosomal stalling. Sequences that cause ribosomal stalling are known and may include sequences that encode mRNA secondary structure, runs of rare or difficult-to-decode codons, and codons encoding certain amino acids such as proline, glycine, positively charged amino acids, and negatively charged amino acids.

Other frameshifting nucleic acid sequences will be known. For example, the frameshifting nucleic acid sequence may be sequence as identified in a database such as the FSDB (see Moon S, Byun Y, Han K. FSDB: a frameshift signal database. Comput Biol Chem. 2007;31(4):298-302. Doi:10.1016/j.compbiolchem.2007.05.004) or the PRFdb (see Belew, Ashton T., et al. “PRFdb: a database of computationally predicted eukaryotic programmed-1 ribosomal frameshift signals.” BMC genomics 9.1 (2008): 1-7.).

“Transcription terminator sequence” refers to the nucleic acid sequence immediately downstream of an encoded protein and is responsible for the termination of the transcription of the gene of interest. Transcription terminator sequences are widely distributed sequences in prokaryotic and eukaryotic genomes Some examples of transcription terminator sequences include, poly (A)-dependent pathway terminators, r-dependent sites, Rho-

dependent termination (RDT) sequences, and intrinsic terminators. A specific example of a T7 terminator includes the sequence ATCTGTT (SEQ ID NO: 85) in the non-template strand (see for example, He, Biao, et al. "Characterization of an unusual, sequence-specific termination signal for T7 RNA polymerase." Journal of Biological Chemistry 273.30 (1998): 18802-18811).

In some examples, the one or more parameters is location of one or more stop codons. The stop codons may be in frame stop codons or stop codons that are out-of-frame. For example, one or more stop codons may be may be located in the open reading frame of an encoded protein so as to allow for correct production of the encoded protein.

In some examples, one or more stop codons may be located out-of-frame. Proteins are translated by reading tri-nucleotides (codons) from the 5' to the 3' end, starting with the amino acid methionine as the start (initiation) codon. Each codon is translated into a single amino acid. A shift of any number of nucleotides that is not divisible by 3 in the reading frame will cause subsequent codons to be read differently from that of the intended or correct reading frame. This effectively changes the ribosomal reading frame leading to the production of alternative polypeptides. The translation of a frameshifted codon may be referred to as out-of-frame translation, and the products (i.e. alternative proteins or polypeptides) may be referred to as out-of-frame products or proteins or as alternative products or proteins (i.e. alternative to the product encoded by the non-frameshifted (correct or in frame translation product) open reading frame). For example, one or more stop codons may be located in a -2, -1, +1, and/or +2 frameshifted position. For example, a tandem stop codon as described herein may be located in a frameshifted reading frame in comparison to the reading frame for production of the desired protein encoded by the RNA sequence. For example, a tandem stop codon may be encoded in a -1, -2, +1, or +2 frameshifted reading frame. In some examples, a tandem stop codon may be encoded in a +1 frameshifted reading frame.

In some examples, the sequence elements comprise aptamer sequence as described herein. "Aptamer" or "aptamer sequence" refers to a nucleic acid sequence having a specific binding affinity for a target, e.g., a target molecule specific to a target cell. Aptamers may enhance uptake into target cells or translation inside the target cells.

Secondary structure forming sequences refers to any nucleic acid sequence that may lead to the formation of secondary structures such as hairpins, stem loops and pseudoknots. Formation of RNA secondary structure may be predicted using methods and computing devices well known in the field.

Restriction enzyme sites refers to a target nucleic acid sequence that is recognized and cleaved by a restriction enzyme. Restriction enzymes are well known in the art. Such sites may help with manipulation and production of the designed nucleic acids.

RNA protein binding sites refers to sequences that may be specifically or preferably bound by RNA binding proteins. RNA binding protein refers to a protein that contains an RNA binding domain and thus has a specific or general affinity for RNA. Exemplary RNA binding proteins are proteins involved in alternative splicing, RNA editing, polyadenylation, nuclear export, mRNA localization, and control of translation.

Frameshifted premature termination codons refers. “Premature termination codon” or “premature stop codon” refers to a stop codon in an mRNA (prior to the endogenous or desired termination codon) as the result of a mutation (i.e. nucleic acid modification). Premature termination codon (PTC) may include one of three stop codons: UAA; UAG; or UGA. A frameshifted premature stop codon is a stop codon which is encoded in a frameshifted reading frame. For example, when translated in-frame, the translation product is unaffected by the PTC. However, if frameshifting occurs (i.e. a -1, -2, +1, or +2 frameshift), the frameshifted reading frame comprises at least one codon that encodes a PTC that results in termination of translation and production of a truncated protein that may be degraded when produced in a subject.

In step 904, method 900 comprises receiving deviation information for each parameter. The deviation information is for processing a nucleotide sequence and is indicative of a degree of acceptable deviation of a nucleotide sequence from a target value of the associated parameter.

In step 906, method 900 comprises determining a score function for each parameter based on the deviation information. Each score function is usable for processing a nucleotide sequence to generate a score indicative of a degree of deviation between the nucleotide sequence and the target value of the parameter associated with that score function. Examples and a further discussion of the score function and deviation information are described in relation to Fig. 8 below.

In step 908, method 900 comprises selecting, from amongst a plurality of nucleotide sequences, a set of random nucleotide sequences. Each random nucleotide sequence comprises a nucleotide sequence that encodes the target translated protein sequence.

In step 910, method 900 comprises generating one or more parameter scores for each random nucleotide sequence based on the score functions determined for the corresponding one or more parameters.

In some examples, the set of random nucleotide sequences includes one or more seeding nucleotide sequences. Seeding nucleotide sequences include nucleotide sequences that encode the target translated protein sequence and have a predetermined parameter score for one of the parameters. For example, the set random nucleotide sequences may comprise one or more a seeding sequence that has a predetermined parameter score for one of:

i. codon composition efficiency for translation of the RNA sequence to the target translated protein sequence;

ii. presence or absence of specified nucleotide repeats;

iii. GC content;

iv. desirability of specified nucleotide sub-sequences; and/or

v. location of one or more stop codons.

A “predetermined parameter score”, as used herein, is a parameter score that meets a predetermined threshold parameter score, ideally a predetermined threshold corresponding to a desired or optimised score for the parameter. The predetermined parameter score may be a relatively high parameter score. For example, the predetermined parameter score may be a score with a low degree of deviation from the target value of the parameter associated with the corresponding score function. For example, for the parameter of GC content, the desired GC content may be 50% and the seeding nucleotide sequence may comprise a GC score of 50%. That is to say that the seeding nucleotide sequence may have parameter score equal to the optimised parameter score for one of the parameters, or a parameter core that is $\pm$10%, $\pm$15%, or $\pm$20% of the optimised score. Without being bound by theory, including seeding sequences in the set of random sequences may improve the efficiency of the method (for example by reducing the number of required iterations) and/or the provision of a greater number of optimised nucleotide sequences that have a total score greater than the score threshold.

Determination of a parameter score for seeding nucleotide sequences may be carried using any suitable methods or calculations for individual sequences. For example, by calculating GC content or providing a seeding nucleotide sequence that includes properties that match the corresponding parameter. For example, includes stop codons at specific desired locations, has the desired presence or absence of specified nucleotide repeats, has a specified codon composition and/or includes specific desired sub-sequences.

In some examples, the random set of sequences may include multiple seeding sequences with a predetermined parameter score for a single parameter, i.e. such that multiple seeding sequences have a predetermined score for the same parameter. In some examples, the

random set of sequences may include multiple seeding sequences, each of which has a predetermined parameter score for a different parameter, i.e. such that each seeding sequences is associated with a unique parameter, optionally where each parameter has a unique seeding sequence. In some examples, the random set of sequences may include at least one seeding sequence having a predetermined parameter score for each parameter.

In step 912, method 900 comprises generating a total score for each random nucleotide sequence based on the one or more parameter scores generated for that random nucleotide sequence. In particular, once the parameter scores have been generated for each parameter across the set of random nucleotide sequences in step 910, the total score for each random nucleotide sequence is generated in step 912. The total score may be generated based on the one or more parameter scores by any suitable means, such as using a statistical or rules-based algorithm, trained model, etc. For example, the total score for a given random nucleotide sequence may be the sum, product or average of the one or more parameter scores of that random nucleotide sequence.

In step 914, method 900 comprises assigning a rank to each random nucleotide sequence in the set of random nucleotide sequences based on the generated total scores of the random nucleotide sequences from a lowest rank to a highest rank.

In step 916, method 900 comprises removing, from the plurality of nucleotide sequences, one or more of the lowest ranked random nucleotide sequences based on the assigned ranking of each random nucleotide sequence, to provide a reduced plurality of nucleotide sequences.

In step 918, method 900 comprises performing a plurality of iterative steps including iteratively repeating steps 908, 910, 912 and 914. In each successive iterative step, the plurality of nucleotide sequences in step 908 (i.e. from which the set of random nucleotide sequences is selected) corresponds to the reduced plurality of nucleotide sequences provided in step 916 in the previous iterative step. Taking for example the plurality of nucleotide sequences as including an n number of iterative steps, when iterative step (n) provides a reduced plurality of nucleotide sequences at step 916, a subsequent iterative step (n+1) is performed in step 918, such that step 908 of the subsequent iterative step (n+1) includes selecting a set of random nucleotide sequences from the reduced plurality of nucleotide sequences provided at step 916 in the iterative step (n).

In some embodiments, the set of random nucleotide sequences in the subsequent iterative step (n+1) may be selected to include the non-removed nucleotide sequences present from the set of random nucleotide sequences in the previous iterative step (n) (i.e. those same nucleotide sequences that have been maintained within the plurality of nucleotide

sequences from the previous iterative step (n)) and to also include one or more new nucleotide sequences from among the plurality of nucleotide sequences but which were not included in the set of random nucleotide sequences in one or more of the previous iterative steps. In such embodiments, the number of such new nucleotide sequences may correspond to the number of nucleotide sequences removed at step 916 in the preceding iterative step (n-1). In doing so, each successive iterative step effectively acts to replenish a set of random nucleotide sequences that has been reduced at step 916 in the preceding iterative step by virtue of the removal. In such examples, the set of random nucleotide sequences in the given iterative step (n) further comprises one or more mixed sequences and/or optimised randomly mutated sequences.

As such, step 918 includes iteratively repeating steps 908 to 916 for each iterative step until a stop condition is met to provide a set of optimised RNA sequences at step 920, wherein each optimised RNA sequence has a total score above a score threshold. Here, the score threshold may be predetermined e.g. via a user input and corresponds to a minimum acceptable total score of the optimised RNA. It will be appreciated that the method generally provides optimised RNA sequences with total scores above the score threshold, by virtue of the iterative steps being performed at step 918 to reduce the plurality of nucleotide sequences by removing those with the lowest ranked scores. However, in some embodiments, to ensure that the optimised RNA sequences have a total score above the score threshold, the method 900 may optionally further comprise comparing, following step 912 in each iterative step, the total score of each random nucleotide sequence with the score threshold and removing from the plurality of nucleotide sequences any random nucleotide sequences having a total score above the score threshold.

Each optimised RNA sequence comprises at least one different sequence property. For example, each optimised RNA sequence may comprise a different open reading frame (ORF), GC content, presence or absence of specified nucleotide repeats, specified nucleotide sub-sequences and/or sequence elements (when included for an optimised mRNA as described below). It will be understood that when only a single optimised RNA sequence is returned to the user that no sequence differences will exist as there is no comparator for the difference.

Referring back to step 906, examples of the deviation information and score function will now be described, with reference to FIG. 8A. In the present example, the deviation information is predetermined for each parameter and may be set according to a user input, which may for example be received via a suitable computing device (an example of which is described in more detail with reference to FIG. 10). Here, the deviation information may be understood to mean a degree of deviation allowed of the nucleotide sequence from the

target value for a given parameter. As such, the deviation information may be considered to be a threshold indicative of a level of precision, for example a variance or uncertainty within which the nucleotide sequence may be considered to match the target value. The deviation information may for example comprise one or more numerical values, where each deviation value may correspond to a respective parameter. For example, if the deviation value is relatively low, this effectively signifies that it is important for the nucleotide sequence to match its target value with relatively high precision and thus allow for a relatively low variance with respect to the target value. On the other hand, if the deviation value is relatively high, this effectively signifies that the nucleotide sequence may match its target value with relatively less precision and thus allow for a relatively high amount of variance with respect to the target value. However, the disclosure is not limited to the deviation information being received as one or more numerical values, and in some embodiments, the deviation information may be derived from a ranking of the parameters. In such embodiments, the deviation information may include ranking information for ranking each parameter relative to one another. For example, more highly ranked parameters may be determined to allow for a relatively lower tolerance in the variance from the target parameter value (i.e. indicating a lower degree of deviation), whereas lower ranked parameters may be determined to allow for a relatively greater degree of deviation from the target parameter value.

Based on the deviation value, step 906 includes determining the score function accordingly. The score function may be determined using any suitable means, such as a statistical or rules-based algorithm, or a trained model. Examples of trained models may include large language models for processing natural language queries (e.g. a generative pre-trained transformer) arranged to process the inputs of steps 902 and 904 to generate and output the score function. The score function is understood to be a mathematical function of the deviation information and may have any suitable relationship with the deviation information. For example, the score function may be a polynomial function, where the deviation information is related to the degree of the polynomial function, as represented by:

$$x \propto y^n$$

where x is the score, y is the parameter value and n is the degree of the polynomial related to the deviation information. As such, the score function may be determined based on the deviation information. Fig. 8A shows some examples of score functions that may be determined. Each of the score functions plot the score against the value of a particular parameter of a nucleotide sequence, where the target value of the parameter is provided with a score of 1. As shown in FIG. 8A(a) (corresponding to the left graph), the deviation information may be set relatively low to provide for example a triangular score function,

which permits a nucleotide sequence to only deviate from the target parameter value with a relatively small degree of deviation. FIG. 8A(b) (corresponding to the centre graph) shows an example where the deviation information has been set relatively higher than in Fig. 8A(a), providing for example a parabolic score function to allow for a greater degree of deviation from the parameter target value. FIG. 8A(c) (corresponding to the right graph) shows an example where the deviation information has been set relatively higher than in Fig. 8A(b), such that a higher degree of polynomial (e.g. cubic) function is provided to allow for a yet greater degree of deviation from the target parameter value. It will be appreciated that the disclosure is not limited to this, as the above described polynomial relationship between the degree of deviation and the score function is merely an illustrative example. In particular, any suitable mathematical function may be generated based on the deviation information.

In some embodiments, the score threshold may be derived from the deviation information. For example, if a random nucleotide sequence is determined to have one or more parameter scores that deviate from the target value greater than permitted by the degree of deviation indicated by the deviation information, the method 900 further comprises removing that random nucleotide sequence from the plurality of nucleotide sequences, thereby helping to provide optimised RNA sequences at step 920. In step 914 of the present example, the random nucleotide sequence assigned with the lowest rank may have a total score indicative of the one or more parameter scores having the greatest degree of deviation from the target values associated with the respective one or more parameters. The random nucleotide sequence assigned with the highest rank may have a total score indicative of the one or more parameter scores having the smallest degree of deviation from the target values associated with the respective one or more parameters.

In some examples of the disclosure, step 920 provides the set of optimised RNA sequences by identifying the highest ranked random nucleotide sequence for the set of random nucleotide sequences in a first iterative step and the highest ranked random nucleotide sequence for the set of random nucleotide sequences in a second iterative step. Here, the first iterative step may be the preceding iteration to the second iterative step. In such embodiments, step 920 further includes determining a difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step, and comparing the difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step with a difference threshold. In such embodiments, the stop condition may comprise one or more of: a predetermined number of iterations; the difference in the total score of the highest ranked random nucleotide sequence being below the difference threshold; and at least one selected random nucleotide sequence has a total score above an optimisation threshold. Here, the

optimisation threshold differs from the score threshold. In particular, the optimisation threshold is understood to correspond to a total score greater than the score threshold, such that whilst every optimised RNA sequence has a total score above the score threshold, not every optimised RNA sequence has a total score above the optimisation threshold, since the optimisation threshold is used as a stop condition.

Optionally, the user defined objectives comprise one or more of target application, biological activity, expression profile, manufacturability, secretion profile and/or storability as described above.

Optionally the method may include selecting one or more sequence elements which may then be combined with the optimised RNA sequence to provide a full length mRNA sequence (e.g. including 5' and 3' UTR sequences, poly(A) tail, signal peptide sequence, protein binding sequences, tandem stop codon, nucleic acid binding sequences and any other nucleic acid sequences required for translation of the optimised RNA sequence into a protein when in use, such as when administered to a target cell or subject). It will be understood that addition of sequence elements may not be required and the optimised RNA sequence may be combined with any suitable sequence elements after output of the optimised RNA sequence. In some examples, each full length mRNA sequence provided in a set may have different sequence elements to each other (i.e. different sequence properties).

The sequence elements may be provided in a database and selected accordingly. In some examples, the sequence elements may comprise one or more of the desirable or undesirable sub-sequences described above.

Sequence elements may include any one or more of a 5'-UTR, a 3'-UTR, a poly(A) tail length, signal peptide sequence, an aptamer sequence, protein binding sequences and/or nucleic acid binding sequences . For example, the sequence elements may include at least the 5' UTR and 3' UTR pairs described above. For example, the sequence elements may include an optimised poly(A) tail as described above. For example, the sequence elements may include at least the signal peptide sequences as described above.

In some examples, the sequence elements comprise aptamer sequence as described herein. "Aptamer" or "aptamer sequence" refers to a nucleic acid sequence having a specific binding affinity for a target, e.g., a target molecule specific to a target cell. Aptamers may enhance uptake into target cells or translation inside the target cells. In some examples, the sequence elements comprise one or more tandem stop codons as described herein.

Protein binding sequences may be any sequence known to have specific affinity for one or more proteins. For example, protein binding sequences may be sequences that have affinity

for any known RNA-binding proteins. As used herein, RNA-binding proteins refer to proteins that bind to RNA molecules, are generally found in the cytoplasm and nucleus, and are important in formation of ribonucleoproteins (RNP). RNA-binding proteins contain various structural motifs, such as RNA recognition motif (RRM), dsRNA binding domain, zinc finger and others.

Nucleic acid binding sequences may be any sequence known to have affinity for an RNA binding nucleic acid. For example, the RNAs may include sequences that bind to small interfering RNAs (siRNAs). siRNA binding sites include the coding region and the 3′-untranslated region (3-UTR) of mRNA. Examples of siRNA binding sequences are provided in Ui-Tei, Kumiko et al. “Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference.” Nucleic acids research vol. 32,3 936-48. 9 Feb. 2004, doi:10.1093/nar/gkh247 and Jopling, Catherine. “Liver-specific microRNA-122: Biogenesis and function.” RNA biology vol. 9,2 (2012): 137-42. doi:10.4161/rna.18827.

A set of sequence elements may be selected from the database based on user defined objectives. For example, user defined objectives may be any one or more of the objectives described above. In some examples, the user defined objectives are as described herein.

For example, the target application may be expression of the encoded protein in a specific target cell (such as a liver cell). In such cases, 5’ and 3’ UTRs mat be those described herein that have been shown to have higher efficiency in liver cells (such as MUTR2, MUTR4, MUTR6, MUTR7, MUTR8, MUTR10, MUTR11, LUTR1, LUTR2, LUTR3, LUTR5, LUTR6, LUTR7, LUTR8, or LUTR10). In such cases, an aptamer specific to a target in liver cells may be selected.

In some embodiments, following step 920, the method 900 comprises outputting a signal indicative of a set of optimised full-length mRNA sequences, whereby the set of optimised full-length mRNA sequences comprises full-length mRNA sequences which comprise at least one optimised RNA sequence which meets the predetermined score threshold of step 920. In this case where the set of optimised full-length mRNA sequences comprises full-length mRNA sequences which comprise an optimised RNA sequence which has a total score below the predetermined score threshold, method 900 further comprises iteratively repeating steps 908 to 920 for the optimised full-length mRNA sequences and outputting a signal indicative of a set of optimised full-length mRNA sequences, comprising the optimised full-length mRNA sequence identified as having the highest total score.

In the present example, step 914 further comprises combining n nucleotides of the 5' end of at least one first removed nucleotide sequence with L – n nucleotides of the 3’ end of a second removed nucleotide sequence to provide a mixed sequence. However, the

disclosure is not limited to this, and in some examples, step 914 includes randomly mutating at least one removed sequence to generate a randomly mutated sequence, carrying out steps 910 to 914 for each randomly mutated sequence and selecting at least one optimised randomly mutated sequence based on a predetermined threshold score. In further examples, step 914 may include performing both i.e. combining n nucleotides of the 5' end of at least one first removed nucleotide sequence with L – n nucleotides of the 3' end of a second removed nucleotide sequence to provide a mixed sequence, and randomly mutating at least one removed sequence to generate a randomly mutated sequence, carrying out steps 910 to 914 for each randomly mutated sequence and selecting at least one optimised randomly mutated sequence based on the predetermined score threshold.

The method 900 may be implemented by any suitable means, such as a suitable computing device, an example of which is described with reference to FIG. 10.

FIG. 10 shows an apparatus 1002 for designing an optimised RNA sequence for translation as described herein. The apparatus 1002 may be used to implement the method 900. The apparatus 1002 includes processing circuitry 1004, a memory 1008 and an input/output I/O interface 1006.

The processing circuitry 1004 may include any suitable processer and/or combination of processors. For example, the processing unit 1001 may include one or more of a Central Processing Unit (CPU) and a Graphical Processing Unit (GPU). The processing circuitry may comprises for example comprise an input means and an output means. The input means may comprise an electrical input of the processing circuitry. The output means may comprise an electrical output of the processing circuitry. In some examples, the input means and output means may be unified such as in the form of a network interface which inputs and outputs data, for example to a communication bus of communication circuitry. The processing circuitry may therefore receive data from the communication bus and output data onto the communication bus.

The memory 1008 may include volatile memory and/or non-volatile/persistent memory. The memory 1008 may, for example, be used to store data such as an operating system, instructions to be executed by the processing unit (e.g. in the form of software to be executed by the processing unit), configuration information related to the apparatus 1002, session information and/or configuration or registration information associated with any other device in communication with the apparatus 1002. For example, the memory 1008 may be used to store instructions for executing any of the methods disclosed herein.

At least the processing circuitry 1004 is connected to an input/output I/O interface 1006. The I/O interface 100 facilitates communication with one or more other devices in communication

with it, and includes communication circuitry to facilitate this communication. For example, the I/O interface 1006 may be operable to transmit and/or receive communications to/from other devices in a network. In some examples, the I/O interface 1003 may be operable to transmit and/or receive communications over an air interface. For example, the I/O interface 1003 may include a transmitter and/or a receiver for transmitting and/or receiving wireless communication (e.g. radio frequency signals). In some examples, the I/O interface 1003 may include a transceiver configured to receive and transmit wireless communication (e.g. radio frequency signals). In some examples, the I/O interface 1003 may be operable to additionally or alternatively communicate over one or more wired connections. The communication circuitry may be provided in one or more communication devices, and may be of any suitable type for communicating with the sensor to receive data signals indicative of the plurality of measurements. For example, the communication circuitry may include a wireless module to receive the signals from the sensor wirelessly, such as a Bluetooth®, Wi-Fi, WLAN and/or network data module. The communication circuitry may comprise a receiver in addition to a transmitter or may comprise a transceiver adapted to both transmit information to and receive information from the processing circuitry.

Optionally, the apparatus 1002 may further include a display, which may for example be for displaying information to a user, such as information corresponding to the set of optimised full-length mRNA sequences. The display may comprise any suitable electronic display such as a touch sensitive display. The display may be connected to at least the processing circuitry 1004. The processing circuitry 1004 may generate display signals which are sent to the display in order to cause the display to display information. Optionally, the apparatus 1002 further includes an input unit arranged to receive a user input. The display together with the input unit may incorporate for example a graphical user interface (GUI).

Furthermore, in some examples the apparatus may be in communication with another apparatus having a display, such that information relating to the random nucleotide sequences and/or optimised full-length mRNA sequences are output to the display of the another apparatus so as to be remotely displayed.

In some examples, the information relating to the random nucleotide sequences and/or any of the information derived therefrom, such as the optimised full-length mRNA sequences may be stored on a database in a cloud networking environment. In such examples, the apparatus is in communication with the database which may be arranged remotely on a server in the cloud networking environment. The server includes communication circuitry and processing circuitry. However, the server is not limited to this and may be any suitable computing device. The apparatus is arranged to store information on the database and also

retrieve information from the database. However, it will be understood that the disclosure is not limited to this, and the database may in other examples be stored on the apparatus itself.

Also provided herein are machine readable instructions, which when executed by processing circuitry, cause the processing circuitry to carry out a method of designing a tag sequence for specifically binding to a tag binding sequence as described herein.

Also provided herein is a machine readable data storage medium having tangibly stored thereon the machine readable instructions of designing a tag sequence for specifically binding to a tag binding sequence as described herein.

With reference to FIG. 9, the apparatus 1002 may thus perform each step of the method 900. In particular, the apparatus 1002 may have stored on the memory 1008 instructions, which is executed by the processing circuitry 1004 to perform the steps of the method 900. The instructions may for example be any suitable machine readable instructions, such as those described herein (e.g. statistical or rules-based algorithm, trained model, etc.).

The apparatus 1002 for example may initially receive via its input unit one or more user inputs specifying a target translated protein sequence, one or more user-defined objectives, one or more parameters, and a deviation information described in relation to steps 902 and 904. The machine readable instructions (e.g. rules based algorithm, trained model, etc.) may then process the user inputs to determine a score function for each parameter based on the deviation information in step 906.

Where functional units have been described as circuitry, the circuitry may be general purpose processor circuitry configured by program code to perform specified processing functions. The circuitry may also be configured by modification to the processing hardware. Configuration of the circuitry to perform a specified function may be entirely in hardware, entirely in software or using a combination of hardware modification and software execution. Program instructions may be used to configure logic gates of general purpose or special-purpose processor circuitry to perform a processing function.

Circuitry may be implemented, for example, as a hardware circuit comprising custom Very Large Scale Integrated, VLSI, circuits or gate arrays, off-the-shelf semiconductors such as logic chips, transistors, or other discrete components. Circuitry may also be implemented in programmable hardware devices such as field programmable gate arrays, FPGA, programmable array logic, programmable logic devices, A System on Chip, SoC, graphics processing units, GPU, or the like.

Machine readable program instructions may be provided on a transitory medium such as a transmission medium or on a non-transitory medium such as a storage medium. Such

machine readable instructions (computer program code) may be implemented in a high level procedural or object oriented programming language. However, the program(s) may be implemented in assembly or machine language, if desired. In any case, the language may be a compiled or interpreted language, and combined with hardware implementations. Program instructions may be executed on a single processor or on two or more processors in a distributed manner.

While the making and using of various embodiments of the present invention are discussed in detail below, it should be appreciated that the present invention provides many applicable inventive concepts that can be embodied in a wide variety of specific contexts. The specific embodiments discussed herein are merely illustrative of specific ways to make and use the invention and do not delimit the scope of the invention.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Current Protocols in Molecular Biology (Ausubel, 2000, Wiley and son Inc, Library of Congress, USA); Molecular Cloning: A Laboratory Manual, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (Harries and Higgins eds. 1984); Transcription and Translation (Hames and Higgins eds. 1984); Culture of Animal Cells (Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells and Enzymes (IRL Press, 1986); Perbal, A Practical Guide to Molecular Cloning (1984); the series, Methods in Enzymology (Abelson and Simon, eds. -in-chief, Academic Press, Inc., New York), specifically, Vols.154 and 155 (Wu et al. eds.) and Vol. 185, "Gene Expression Technology" (Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (Miller and Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods in Cell and Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook of Experimental Immunology, Vols. I-IV (Weir and Blackwell, eds., 1986); and Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Terms such as "a", "an" and "the" are not intended to refer to only a singular entity, but include the general class of which a specific example may be used for illustration. The terminology herein is used to describe specific embodiments of the invention, but their usage does not delimit the invention, except as outlined in the claims.

Unless otherwise indicated, nucleic acid molecules are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation,

respectively. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art.

Aspects of the invention are demonstrated by the following non-limiting examples.

EXAMPLES

Example 1 – Translation Optimisation

Materials and Methods

IVT mRNA synthesis

The preparation of IVT mRNAs was based on *in vitro* transcription from DNA templates. For mRNAs described in Fig 1, *in vitro* transcription was performed with the IVTpro™mRNA Synthesis kit (Takara, Cat. #6141) according to the manufacturer's instructions. A 20 µL transcription reaction contained a maximum of 2 µg linear DNA template, 100 mM of each NTP, CleanCap AG (Trilink Biotenchnologies, N7113),2 µL enzyme mix and 2 uL Transcription Buffer. After a total incubation for 2 h at 37°C, the DNA was digested by addition of 1 µL/2 U Turbo DNase for 15 min at 37°C. For modified IVT mRNAs, N^1-methylpseudouridine (Trilink Biotechnologies, N1019) was substituted for uridine triphosphate at an equivalent concentration. The mRNA was purified using a commercially available RNeasy Plus Kit (ID: 74134, Qiagen) according to the manufacturer's instruction and kept frozen at −80 °C until required for further use. For mRNAs in Fig 2, Fig 5-6, and Fig 7B, DNA fragments were obtained from Telesis Bio and PCR amplified to produce a DNA template. mRNA was synthesised by in vitro transcription using the HiScribe T7 mRNA Kit with CleanCap (New England Biolabs, E2040S) with UTP substituted for N1-methylpseudoUTP (TriLink Biotechnologies, San Diego, USA, N-1081). Polyadenylated tails were added enzymatically using the E. coli Poly(A) Polymerase Kit (NEB, M0276L) and then purified using the RNeasy Mini Kit (Qiagen, 74004). Purified mRNA was quantified using the Qubit 3 Fluorometer and analysed for integrity by agarose gel electrophoresis before being stored at -80C. For mRNAs in Fig 4 and Fig 7C, mRNAs were synthesised according to a previously described protocol[1]. CleanCap AG (TriLink N-7113) was used according to the manufacturer's recommended concentration and molar ratio to obtain m7G5'ppp5'2'OMeApG nLuc mRNA (Fig 7C).

Cell culture and transfection

HEK293T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Gibco) containing 2 mM L-glutamine, supplemented with 10% foetal bovine serum (v/v), 100 U/ml penicillin, and 0.1 mg/ml streptomycin at 37°C in 5% CO_2 incubators. Human Skeletal Muscle Cells (SKMCs) were obtained from Promocell (C-12530) and cultured in Skeletal Muscle Cell Growth Medium (Promocell C-23060). For IVT mRNA transfections in Fig 2, Fig 5-6, and Fig 7B, HEK293 or SKMC cells were seeded in 24 well plates, and transfected after 24 hours using 1 ug mRNA:Lipofectamine-2000 prepared in a serum free media-Opti-MEM (Gibco 31985062). Transfection was performed in a complete DMEM supplemented with 10% FBS. For other IVT mRNA transfections, 250 ng mRNA:Lipofectamine-2000 was used.

Luciferase assay

Luciferase activity was measured using a Nano-Glo luciferase assay kit from Promega according to the manufacturer's instructions (Nano-Glo Luciferase Assay System, N1110). Luciferase activities were normalised to levels from controls, where described, to obtain relative expression levels, or reported as raw values.

LNP formulation

Lipids were dissolved in 100% ethanol and mixed in the molar ratio indicated below to a final concentration of 8mg/ml. mRNA was diluted in 10mM citrate buffer at pH 4 and mixed with lipids at a 3:1 mRNA:lipid ratio using a NanoAssemblr Ignite (Precision Nanosystems). LNPs were then diluted in sterile PBS and concentrated using a centrifugal filter with a 100 kDa molecular weight cut off. Concentrated LNPs were passed through a sterile filter prior to mouse immunisations.

Table 3

Lipid	**Molar ratio**
DSPC	9.4
ALC-0159	1.7
ALC-0315	46.3
Cholesterol	42.6

Mouse immunisation

C57BL/6J mice (8-12 weeks) were purchased from Charles River laboratories and immunised intramuscularly with mRNA-LNP. When given multiple doses mice were immunised 3 weeks apart into the same site of primary immunisation. Mice were culled 1 week after final vaccination and spleens and serum were collected. Spleens were mashed and filtered through 70um cell strainers. Red blood cells were lysed using RBC lysis buffer (155mM NH_4Cl, 12 mM $NaHCO_3$, 0.1 mM EDTA) and then filtered through 40um cell strainers.

IFNγ assay

Splenocytes (1 x 10^6 cells) were incubated in RPMI 1640 media with 1ul BD GolgiPlug and IL2 (20ng/ul), with or without SIINFEKL peptide (SEQ ID NO: 72) (1uM). Splenocytes were incubated for 6h at 37C before extracellularly staining with CD19 (PE-Cy7), F4/80 (PE-Cy7), NK1.1 (PE-Cy7), Ly6G (PE-Cy7), CD8a (APC), CD3 (APC-Fire750) and CD44 (BV650) and intracellularly staining for IFNγ (PE). Flow cytometry was performed using a BD Fortessa Analyser and data were analysed in FlowJo (BD).

Table 4

Antibody	Dye	Supplier	Dilution
Anti-mouse CD8a	APC	Biolegend	1:200
Anti-mouse CD3	APC-Fire 750	Biolegend	1:100
Anti-mouse CD44	BV650	Biolegend	1:200
Anti-mouse INF-y	PE	Biolegend	1:100
Anti-mouse CD19	PE-Cy7	Biolegend	1:200
Anti-mouse F4/80	PE-Cy7	Biolegend	1:200
Anti-mouse NK1.1	PE-Cy7	Biolegend	1:200
Anti-mouse Ly6G	PE-Cy7	Biolegend	1:200

Live virus Neutralisation

Serum was collected by cardiac puncture 1 week after immunisation and was assessed as previously described[2].

tRNA sequencing

Small RNA was purified from HEK293T cells using the mirVana miRNA isolation kit (Invitrogen AM1560). Total RNA was purified from SKMCs using Trizol (Invitrogen) extraction followed by isopropanol precipitation. RNA was demethylated using purified *E.coli* AlkB, which was a gift from the Bushell Lab, and deacylated in 100 mM Tris-Cl pH 9.0 for 45 minutes. Treated RNA was polyadenylated using *E.coli* polyA polymerase (Ambion 1350), and reverse-transcribed using SuperScript III (Invitrogen) and Illumina TruSeq Small RNA 5'-adapter-oligodT. cDNA was polydeoxyguanylated using Terminal Transferase (NEB M0315), and PCR-amplified with Illumina TruSeq Universal Primer and 5' Illumina Index Sequence primers containing a custom spacer sequence and oligodC-3'. PCR cDNA products were purified by PAGE, pooled, and sequenced on an Illumina MiSeq system (150 cycles, single end, v3). Reads were analysed as previously detailed[3]. Figure 2 displays average data for n=3 replicated HEK293T or SKMC cultures.

Results & Discussion

Coding sequence optimisation

To investigate the impact of codon usage on nLuc mRNA translation, we designed four mRNA sequences which encode the same nLuc polypeptide but differ in their synonymous codon content (SEQ ID NOs: 1 – 4). STA (SEQ ID NO: 4) contained a coding sequence which was used as a reference standard[7], FQ (SEQ ID NO: 1) contained a coding sequence mimicking codon usage predicted from codon usage in the human genome, predicted to exhibit high expression, and MIN (SEQ ID NO: 3) contained a coding sequence which was predicted to be poorly translated as it was a poor mimic of human cell codon usage. GS mRNA (SEQ ID NO: 3) contained an nLuc coding sequence designed using a codon optimisation algorithm proprietary to GenScript and served as a state-of-the-art comparison. The codon usage of GS mRNA is also similar to codon usage in BNT162b2, an mRNA-based SARS-CoV-2 vaccine[8]. We synthesised these mRNAs using either canonical ribonucleotides (unmodified), or by replacing uridine-5'-triphosphate with (N)1-methylpseudouridine-5'-triphosphate (modified). We then transfected each mRNA into HEK293T cells for 24 hours and measured secreted nLuc levels in cell culture medium to assess translation efficiency. There were varying levels of nLuc expression with different variants in both unmodified and modified versions in HEK293T cells (Fig 1). The minimal variant, as predicted, showed very low expression (Fig 1). FQ mRNA variant exhibited higher expression than other mRNAs, whether (N)1-methylpseudouridylated or unmodified (Fig 1). These data showed that codon selection strategy in FQ mRNA produced a more efficiently translated mRNA sequence than other approaches.

Improving mRNA translation efficiency has been shown to improve mRNA therapeutic efficacy[9]. Therefore, we designed and synthesised reporter mRNAs for evaluation in a mouse model of mRNA vaccination. These mRNAs encoded a fusion polypeptide consisting of SARS-CoV-2 Spike Glycoprotein Receptor-binding Domain (RBD), the Influenza A haemagglutinin transmembrane domain, and the hen egg ovalbumin MHC Class I SIINFEKL (SEQ ID NO: 72) epitope (Fig 2A). FQ-RBD utilised RBD coding sequence using the same codon strategy as FQ nLuc mRNA, while ST-RBD used the RBD coding sequence from BNT162b2, a clinically approved COVID19 mRNA vaccine, which was used as a state-of-art comparison, similarly to GS nLuc mRNA. We vaccinated mice with (N)1-methylpseudouridylated FQ-RBD mRNA or (N)1-methylpseudouridylated ST-RBD and measured (i) Spike RBD antibody responses and (ii) CD8+ T cell immunogenicity to SIINFEKL (SEQ ID NO: 72), after either a single dose vaccination schedule, or following two doses. FQ-RBD mRNA vaccination produced higher anti-SARS-CoV-2 antibody titres after either one or two doses than ST-RBD (Fig 2b and 2d). After one dose, CD8+ T cell responses for FQ-RBD- and ST-RBD-vaccinated mice were similar (Fig 2c). However, after two doses, CD8+ T cell responses were greater for FQ-RBD-vaccinated mice than ST-RBD-vaccinated mice (Fig 2e). These data suggest that codon selection strategy in FQ-RBD increased mRNA vaccine efficacy and support prior studies demonstrating that mRNA sequence optimisation can improve development of new mRNA therapeutics.

mRNA translation efficiency is related to the efficiency of each ribosome-catalysed elongation cycle during protein synthesis. An elongation cycle starts with the binding of an aminoacyl-tRNA-eEF2-GTP complex to the elongating ribosome A-site. Productive A-site binding of the cognate aminoacyl-tRNA-eEF1A-GTP complex leads to GTP hydrolysis and eEF1A-GDP dissociation. Then, peptidyl transfer and peptide bond formation occurs between the P-site peptidyl-tRNA and the incoming aminoacyl tRNA, which is catalysed by the ribosome and leads to an extended peptidyl-tRNA. Then, the ribosome translocates to the next codon in the mRNA ORF, the deacyl-tRNA is shifted into the ribosome E-site, and the extended peptidyl-tRNA is newly accommodated in the P-site, which is facilitated by binding and hydrolysis of eEF2/GTP. These processes result in extension of the nascent polypeptide-tRNA by one residue and regenerates the A-site-vacant ribosome ready for a subsequent aminoacyl-tRNA, which completes the elongation cycle[10]. The availability of substrate aminoacyl-tRNAs is an important factor for translation elongation efficiency, and decreased availability is known to decrease protein synthesis rates via multiple mechanisms. We and others have shown that mRNA synonymous codon usage can affect mRNA translation efficiency, and we sought to improve strategies for designing mRNA coding

sequences by tuning IVT mRNA codon usage (tRNA demand) with isoacceptor aminoacyl-tRNA substrate levels (tRNA supply).

We measured isoacceptor tRNA abundance in HEK293T cells and primary skeletal muscle cells using high-throughput RNA-sequencing and used these data to design mRNAs with altered codon usage (IVT mRNA tRNA demand) complementary to target cell tRNA supply. For example, during protein synthesis leucine residues are encoded by the codons CUC, CUU, UUG, CUG, UUA, or CUA. These codons are decoded by either tRNA-Leu-AAG, tRNA-Leu-CAA, tRNA-Leu-CAG, tRNA-Leu-UAA, or tRNA-Leu-UAG. In HEK293T cells, tRNA-Leu-CAG levels are approximately 2-fold higher than tRNA-Leu-AAG (Fig 3). Therefore, in a simple example where other factors are equal, an IVT mRNA containing mostly GUG codons would be expected to be decoded more efficiently in HEK293T cells than another, synonymous IVT mRNA containing mostly CUU codons. We synthesised an nLuc mRNA (ISOpt – SEQ ID NO: 5), which was designed using HEK293T tRNA abundance and additional factors (e.g. predicted ribosome density), transfected this mRNA into HEK293T cells, and compared nLuc protein synthesis levels to those from cells transfected with GS mRNA. After 48 hours ISOpt mRNA produced significantly greater levels of nLuc than GS mRNA (Fig 4A). These data suggest that codon optimisation based on measured isoacceptor tRNA abundance can increase IVT mRNA translation efficiency compared to existing approaches. In addition, nLuc levels from ISOpt mRNA translation continued to increase from 24 to 48 hours, whereas nLuc levels from GS mRNA translation did not increase (Fig 4A). This observation may be due to initiation of GS mRNA decay sooner than ISOpt mRNA decay due to cellular sensing of suboptimal mRNA codon usage, which is supported by recent studies[11].

Current mRNA-based infectious disease vaccines make use of intramuscular injection routes, which is known to transfect multiple cell types, including skeletal muscle cells. We determined tRNA abundance in primary human skeletal muscle cells (SKMCs), used these data to design and synthesise another bespoke nLuc mRNA (SKMC-Opt – SEQ ID NO: 6), transfected this mRNA into SKMCs, and measured nLuc synthesis to evaluate whether mRNA translation efficiency in a therapeutic target cell population could be improved using isoacceptor tRNA concentrations to inform mRNA codon usage. SKMC-Opt mRNA translation after 24 hours was approximately 8-fold greater than GS mRNA (Fig 4B). Our findings show that changes in mRNA codon usage can improve translation efficiency, and support the notion that codon optimality is not fixed and common to all human cells, but varies between different cell types with different phenotypes[12]. Furthermore, these data suggest that IVT mRNA translation can be selectively modulated in different cells based on

differential isoacceptor tRNA expression, which would be an important discovery for mRNA development and may permit more targeted and efficacious therapies.

ER signal peptide assessment

An ER secretion signal sequence, also known as an ER signal peptide, is a short amino acid sequence at the N-terminus of a polypeptide that directs the protein to the secretory pathway of a cell. This pathway involves the polypeptide being synthesized by ribosomes and then either co-translationally or post-translationally transported into the ER. The co-translational import process is that by which most ER bound proteins enter the ER. The signal sequence is recognised by the Signal Recognition Particle (SRP) as the polypeptide emerges from the ribosome tunnel and attenuates translation of the mRNA until the SRP binds the SRP receptor whereby the polypeptide is transferred into the SEC61 translocon, the SRP dissociates, and translation resumes and the growing polypeptide is translocated co-translationally into the ER. Secretory proteins are subsequently processed and trafficked through the Golgi apparatus, and eventually secreted from the cell[13]. By including a secretion signal sequence in the IVT mRNA coding sequence, the polypeptide product will include this sequence and will be directed to the secretory pathway as a result. Encoding ER signal peptides in IVT mRNAs is necessary for expression of recombinant proteins which are secreted for their proper biological function, or for expressing recombinant proteins which require secretion to exert their therapeutic effects, for example IVT mRNAs encoding virus antigens[13-15].

Studying how different signal peptides affect protein expression levels is useful for optimising the production of recombinant proteins. By comparing the performance of various signal peptides, researchers can identify which signal peptide is most effective in driving high levels of protein expression. Factors such as protein yield, cellular localisation, secretion efficiency, and post-translational modifications may vary depending on the signal peptide used. We designed nLuc mRNAs which encode different ER signal peptides, transfected these mRNAs into different cells, and measured nLuc protein synthesis to evaluate their effects (SEQ ID NOs: 7 – 27). Our results revealed that among the tested signal peptides, alpha-1-antichymotrypsin (AACT), Human Serum albumin (HSA), and Native Wuhan strain of the COVID-19 spike protein (NT) signal peptides consistently gave high levels of nLuc expression in Huh7 liver cells (Fig 5A). Notably, the liver-secreted signal peptide yielded a comparable pattern of nLuc expression in both HEK293T cells and SKMCs, indicating its potential for consistent protein production across diverse cellular environments (Fig 5B, C). Interestingly, the signal peptides AACT and HSA demonstrated similar levels of nLuc

expression compared to the native signal peptide (NT), utilized in mRNA-1273 and BNT162b2 COVID19 mRNA vaccines (Fig 5).

In summary, AACT, HSA, and NT demonstrated promising potential for enhancing protein production across various cell types when utilised as signal peptides in IVT mRNAs.

UTR screening

Untranslated regions (UTRs) of eukaryotic mRNAs are integral components affecting various aspects of gene expression, including translation efficiency, mRNA localization, and stability[17-20]. Both the 5′ and 3′ UTRs harbour sequence elements capable of modulating translation in a manner that can either enhance or inhibit protein synthesis. Some 5'UTRs, characterized by highly structured segments, possess the ability to impede mRNA translation, whereas the presence of specific structural motifs, such as internal ribosome entry sites (IRESs), within this region can promote translation under specific cellular conditions. The 5'UTR is also critical for the recruitment and assembly of the translation initiation complex. The 3′ UTR plays a pivotal role in conferring mRNA stability, facilitating mRNA localization to specific cellular compartments, and orchestrating cap-dependent translation through intricate interactions between the 3′-poly(A) tail and the 5′ cap structure. Additionally, the 3′ UTR serves as the primary site for microRNA-mediated regulation of translation, adding another layer of complexity to gene expression control.

To explore the selection of cell and tissue-specific UTRs, we analysed publicly available GTEx and RNA-seq data from human skeletal muscle tissue, liver Huh7, and liver tissue[20,21]. We selected 21 UTRs derived from the top 50 highly expressed transcripts from these data, and designed nLuc mRNAs containing these UTR sequences to evaluate using our reporter system (Tables 1 and 2 and SEQ ID NOs: 35 to 70). For comparison, we also designed an nLuc mRNA (CON) which contained UTR sequences from mRNA-1273, a clinically approved COVID19 mRNA vaccine, which served as a state-of-the-art comparison. Seven UTRs from skeletal muscle cell mRNAs (MYL1, TNNC1, ENO3, MYH1, MYL2, CKM, COX6A2, Table 1) significantly increased protein expression from IVT mRNAs in SKMCs compared to CON UTR (Figure 6A). We also assessed nLuc synthesis in transfected HEK293T cells, which showed that some muscle UTRs also exerted a substantial impact on mRNA translation in these cells (Fig 6B). We also assessed nLuc synthesis in IVT mRNA-transfected Huh7 cells. Interestingly, while the UTRs altered nLuc synthesis in HuH7s, the majority did not correlate with nLuc levels observed in SKMCs (Fig 6C). Only three UTR combinations increased nLuc expression in HuH7s (Fig 6C). Similar to our findings in

HEK293T cells, the majority of mRNA expression did not align with that observed in SKMCs – only three UTR combinations increased nLuc expression, suggesting a conserved influence of these UTRs across different cell types (Fig 6).

We investigated the impact of liver cell derived UTR combinations (Table 2) on nLuc mRNA expression by conducting initial transfections of IVT mRNA into HuH7s and evaluating nLuc expression levels. Seven liver-derived UTR combinations increased nLuc expression compared to CON UTR combination (Fig 6D). We subsequently tested these UTRs in HEK293T cells and SKMCs. Most of the UTRs exerted a significant impact in both HEK293T cells and SKMCs, and similarly to their effect in HuH7s (Fig 6E, F). Although these UTRs were derived from mRNAs highly translated in liver cells[22], their effect on mRNA translation efficiency appears to be common to other cells.

In conclusion, our findings provide valuable insights into the role of UTRs in regulating IVT mRNA translation. We have identified UTRs that support high level expression from IVT mRNAs across cell types and others that support high expression in some cell types but not others. These sequences, and the choice of these sequences, affect expression of target proteins from mRNAs and exhibit cell-specific effects on IVT mRNA translation, which could be used to achieve more targeted expression of IVT mRNA-encoded therapeutic proteins. We expect that further investigations into the mechanisms underlying UTR-mediated regulation of mRNA expression will contribute to the development of more effective mRNA-based therapeutics.

Assessment of mRNA polyA tail length

The poly(A) tail is a crucial element in maintaining mRNA stability and optimizing translation efficiency. It functions by inhibiting exonuclease-mediated mRNA degradation and facilitates binding to poly(A)-binding proteins (PABPs), working in synergy with 5′ m7G cap sequences to regulate translational efficiency[23]. There are two main methods of polyadenylation for engineered mRNA in IVTmRNAs: traditional enzymatic polyadenylation, which adds the poly(A) tail to the 3′ end of mRNA without regulating tail length, and designing a fixed-length poly(A) sequence on a DNA template, allowing for control over tail length during transcription[24].

In mammalian cells, actively translated mRNAs typically feature 100–250 adenosine residues in their poly(A) tails[25]. An optimal poly(A) tail length can significantly enhance translation efficiency and mRNA stability and prior research indicates that as the poly(A) tail size increases to around 120 nucleotides, the corresponding protein expression level similarly increases[26]. To further investigate this phenomenon, we synthesized nLuc IVT

mRNAs with varying lengths of poly(A) tails and assessed their translation efficiency in HEK293T cells. Our findings revealed a distinct trend between increasing poly(A) tail length and increasing recombinant protein synthesis. Notably, enzymatically added poly(A) tails led to a substantial increase in nLuc expression up to a specific threshold. However, beyond a poly(A) tail length exceeding approximately 100 base pairs, we observed no further enhancement in protein expression levels (Fig 7A and B). These data provide valuable insight into the relationship between poly(A) tail length and protein expression dynamics from IVT mRNA.

Assessment of mRNA 5'-cap structures on mRNA translation efficiency

Eukaryotic mRNAs typically contain modified 5'-nucleotides, termed the mRNA 5'-cap, which regulate mRNA translation initiation and stability[27]. However, it is currently unclear how mRNA 5'-cap synthesis for IVT mRNAs affect translation efficiency. To gain a better understanding, we synthesised two nLuc mRNAs containing either m7G-5'ppp5'GpG or m7G5'ppp5'2'OMeApG, two mRNA 5'-cap structures which can be produced in IVT mRNAs. These otherwise identical nLuc mRNAs were transfected into SKMCs to assess the impact of each of the mRNA 5'-cap structures on IVT mRNA translation efficiency. Translation efficiency of m7G-5'ppp5'GpG nLuc mRNA was significantly greater than m7G5'ppp5'2'OMeApG nLuc mRNA in SKMCs (Fig 7C). These data suggest that different mRNA 5'-cap structures can significantly alter IVT mRNA translation efficiency, and that IVT mRNAs containing m7G-5'ppp5'GpG exhibit increased mRNA translation efficiency, which are important factors for designing and manufacturing efficacious therapeutic mRNAs.

Computer code and software for design of improved IVT mRNAs

Our studies have produced methods and insights for design of mRNA sequences with favourable properties, for example improved translation efficiency and cell specificity (Fig 1, Fig 2-7). We have produced computer code and software which integrates these developments and can be used as a tool to design highly optimised mRNA sequences encoding a user-defined polypeptide sequence. This tool enables users to design mRNAs with favourable properties, such as those demonstrated in the above mRNAs, and includes attributes/parameters which are specific or desirable to user-specific applications.

Example 2 – Design Software

The following attributes/parameters are included:

(DESIRED PRODUCT PROPERTIES)

Expression levels - the levels of the encoded protein which can be achieved from a given dose of the designed mRNA, which is related to how efficiently mRNA can be processed by the cellular protein synthesis machinery.

Expression duration - for how long the mRNA persists and is translated in target cells, which is related to parameters such as mRNA secondary structure, codon usage, and GC content.

Secretion efficiency - how efficiently the encoded protein is secreted (if required), which is related to the efficiency of the signal sequence and to RNA motifs in the open reading frame and untranslated regions.

Intended biological activity of the mRNA, such as acting as a vaccine adjuvant, which is related to the affinity of the RNA to innate immune receptors.

Unintended biological activity of the mRNA or the encoded protein, such as the generation of unintended translation products such as ribosomal frameshifting, the activation of biological pathways which can reduce protein synthesis in target cells, or misfolding of the protein.

Together, the above attributes estimate the efficacy of the mRNA. Other relevant attributes that can be manipulated by sequence design include:

Manufacturability - the RNA yield from *in vitro* transcription of a given amount of template DNA, which is related (for example) to the presence of motifs that induce termination of the RNA polymerase, or that lead to polymerase drop-off or to polymerase stalling.

Storability - how stable the formulated mRNA is estimated to be at a given temperature and other conditions, and what other storage conditions need to be applied, which is related to the secondary structure content of the mRNA and its intended formulation.

The above attributes are critical to mRNA manufacturing, therapeutic efficacy, and also drug safety, but are also useful for understanding or estimating manufacturing cost and feasibility.

Method

1) Users can specify the following:
    a) a target protein sequence
    b) a target application, such as “intravenous injection targeting liver uptake”.
    c) specific objectives for optimisation of a nucleic acid sequence that encodes the target protein. Examples of such objectives include:
        i) The codon composition, with reference to a look-up table that gives an efficiency measure for each codon. Examples for such efficiency measures include.

- Experimentally measured or computationally predicted codon decoding times.
- Codon preferences derived from experimentally determined tRNA levels.
- Observed usage frequencies in particular genomes.
- Codon usage patterns known to function particularly well for the application stated by the user.

ii) The GC content of the sequence.

iii) The presence or absence of specific nucleotide repeats.

iv) Specific desirable or undesirable sub-sequences, such as:

- "Slippery sequences" which can lead to ribosome frameshifting.
- Motifs that interfere with manufacturability e.g. transcription terminators or sequences that cause polymerase stalling or drop-off.
- Aptamer sequences that add functionality to the transcript.
- Restriction enzyme sites that interfere with or are required for DNA manipulation.
- Binding sites for RNA binding proteins.

d) A measure of the relative importance of each objective.

2) A scoring system is assigned for each Optimisation Objective that can be applied to any possible sequence, where the highest score represents a sequence that meets the objective exactly or as exactly as possible, and the lowest score represents the sequence that least meets the scoring objective. The relationship between a parameter corresponding to an objective, such as GC content, is connected to the score by one of a series of possible mathematical functions (Fig 8A).

3) A pool of randomly chosen sequences is manipulated so that the sequences approach the "optimal" state, defined as the state with the highest possible sum of scores for all objectives defined by the user (ie where no individual score can be improved without reducing one or more other scores. Technically this is sometimes termed "Pareto Optimisation"). In detail this comprises the following steps:

- Selection of a pool of random sequences encoding the specified target protein.
- Scoring of all sequences in the pool against each optimisation objective using the scoring function.

- Summation of the scores for all objectives for each sequence to give a single overall score.
- Ranking of all sequences from best to worst overall score, and removal of a fixed number of the lowest scoring sequences from the pool.
- Mixing of some of the remaining sequences in the pool by “cross-over”, where cross-over means that:
  - Two sequences are removed from the pool.
  - The first n nucleotides of one sequence are combined with the last l-n nucleotides of another sequence (where l is the length of the sequence) and vice versa.
  - The new sequences are returned to the pool.
  - This procedure is repeated a set number of times.
- Changing of individual sequence by random mutation, where:
  - A sequence is removed from the pool.
  - A random set of codons is selected for mutagenesis.
  - New sequences are generated which contain all possible combinations of codons encoding the original amino acid sequence in the selected positions.
  - The resulting mutated sequences are scored against each objective.
  - Scores are summed and the single top-scoring sequence is returned to the pool.
  - This procedure is repeated a set number of times.
- The pool is replenished to its original size with a new set of randomly chosen sequences.
- This cycle is repeated until a stop condition is met, where stop conditions include:
  - A specified number of iterations.
  - The difference in score of the top scoring sequences from one cycle to the next becomes smaller than a specified threshold .
  - If a sequence in the pool meets a specified top score.

4) The top scoring sequences from the optimisation cycle constitute a selection of open reading frames which are candidate sequences for assembly of the full IVT mRNA sequence.

5) Sequence elements other than the open reading frame are selected from a knowledge base that matches these elements to the application stated by the user. For the application example stated above, “intravenous injection targeting liver uptake” this might include selection of 5’- and 3’-UTRs know to function well in liver but not in other tissues, and/or aptamers that enhance uptake into cells or translation inside the cells.

6) A pool of candidate full-length IVTmRNA sequences are then assembled using the best matching 5’-UTRs, 3’-UTRs, and ORFs.

7) If all objectives specified by the user could be processed in the optimisation cycle, the top scoring sequence or sequences (depending on what the user specifies) with the best UTRs for the application are returned to the user.

8) If objectives were specified that cannot be processed in the optimisation cycle for reasons of computational efficiency or because they relate to parameters that are altered by the presence of the 5’-and 3’-UTRs (examples include computationally predicted secondary structure content or objectives that require application of computationally expensive deep learning models), these objectives are scored for a limited number of fully assembled sequences and the top scoring sequences following this final scoring are returned to the user.

A graphical representation of the method is shown in Figure 8B. A particular aspect of the method that enables efficient computational execution of the optimisation cycle is the computational representation of sequences as a single 8-bit integer per codon, and where individual nucleotides of the codon are represented in a bit-wise fashion with bits one and two representing the first nucleotide of the codon; bits 3 and 4 represent the second nucleotide, and bits 5 and 6 represent the third nucleotide (Tables X-Y). Representation in this number format accelerates a number of computational operations, including searching for particular sub-sequences and determining complementary sequences, both of which can be achieved by applying a bitwise XOR operation.

Example 1 & 2 Conclusions

Our study has evaluated different aspects of mRNA design to produce optimal mRNAs. These results have demonstrated how new approaches to mRNA coding sequence design, UTR choice, ER signal peptide choice, polyA tail length, and 5’-cap structure can increase

mRNA translation efficiency and mRNA therapeutic efficacy. Our data show how mRNAs can be designed to be selectively expressed in different tissues, which is an unexplored area of mRNA technology development but would be highly desirable for new therapeutic mRNAs. Using these insights, we have produced software tools and computer code which allow other researchers to design mRNAs which contain these features. These developments increase access to mRNA technology for research and drug development and will improve mRNA therapeutic efficacy in future applications.

EXAMPLE 3

Summary

Modified mRNA therapeutics are used as vaccines to prevent infections and treat cancer, and there is hope that ongoing research will enable this therapeutic modality to treat most human diseases. This revolutionary platform works because the mRNA sequence acts as a code, which ribosomes translate into a therapeutically active protein. However, the sequence of the modified mRNA code can be misread, leading to off-target immune responses[1]. This misreading occurs due to slippage of the ribosome reading frame, resulting in the generation of peptides encoded in the +1 reading frame of the mRNA sequence. How modified mRNA sequences are best optimised to prevent the generation of aberrant T cell stimulating peptide antigens is unknown. Here, using a novel direct frameshift-antigen detection system, it is show that encoding tandem stop codons can prevent antigen generation by developing a working modified mRNA COVID-19 vaccine without evidence of misdirected immunity. This example shows that whilst the UGA stop codon, found extensively in the +1 frame of existing modified mRNA COVID-19 vaccines, effectively prevents the production of frameshift proteins, it fails to prevent the creation of functional T cell antigens. This is evidenced by preclinical experiments and the detection of T cell immunity directed to peptides encoded downstream of these stop codons in individuals vaccinated with the Pfizer BNT162b2 modified mRNA vaccine. Further, aberrant antigen expression is linked to inflammation, cell death, and a reduced vaccination response upon repeated dosing. By incorporating tandem repeats of more stringent UAA or UAG stop codons in the +1 frame of the mRNA sequence, the detection of frameshift-derived antigens was effectively eliminated. This strategy was employed to develop and preclinically test a modified mRNA COVID-19 vaccine that demonstrated live-virus neutralising capacity without signs of misdirected T-cell immunity. Beyond vaccines, this system and optimisation strategy paves the way for developing safer novel modified mRNA therapeutics for repetitive dose indications whilst reducing the risk of harmful immune responses.

Main

The current success of modified mRNA therapeutics relies on the Nobel prize-winning discovery of Kariko and Weissman. They found that the substitution of the nucleoside uridine with modified nucleosides within the mRNA code enabled modified mRNA translation by ribosomes into therapeutically active proteins[2]. However, our recent research demonstrated that N1-methyl pseudouridine modified mRNA is not decoded by ribosomes normally. 'Slippery' sequences within the mRNA code comprising runs of N1-methyl pseudouridine (1-methylΨ or U*) could cause ribosomes to slip their reading frame and start decoding the mRNA code in the +1 frame, generating an off-target product, peptides encoded by the wrong reading frame of the mRNA sequence[1]. Although these unintended peptides only comprised ~10% of the translated product, this would be significant if these fragments stimulated T-cell immunity. We found in patients immunised with the 1-methylΨ modified mRNA Pfizer BNT162b2 vaccine, but not in patients immunised with the AstraZeneca ChAdOx-1 nCoV-19 DNA vaccine, that we could detect T cell responses to peptides encoded by the +1 frame of the Pfizer vaccine. This was a highly surprising and unanticipated observation considering the 59 +1 encoded U*GA stop codons distributed across the +1 frame of the modified mRNA sequence of the Pfizer vaccine (Fig. 12a); these would have been expected to limit mRNA translation in the +1-reading frame either by nonsense mediated decay or translational termination. However, in addition to the degradation of both long and short lived stable functional proteins, peptide T cell antigens can be produced as defective ribosomal products (DRiPs)[3]. It has been suggested that the mRNA translational requirements for DRiPs differs from that of stable proteins with both specialised immunoribosomes[4] and non-canonical translation initiation factors[5] involved in generating these antigens. We investigated whether U*GA stop codons, although able to prevent the production of stable proteins, fails to prevent the translation required for the short peptide fragments that are T cell antigens.

To enable direct assessment of antigen production from modified mRNA, we used an *in vitro* antigen expression system as traditional mRNA translational expression systems may not have the required sensitivity for the detection of peptide antigens[6]. We first transduced HEK 293T cells with the H-2K^b murine MHC Class I molecule and confirmed its expression (Fig.13a). The transduced cells were able to present the immunogenic SIINFEKL (SEQ ID NO: 72) peptide of the OVA protein as these cells only stained positive with the H-2K^b/SIINFEKL (SEQ ID NO: 72) specific 25-D1.16 antibody[7] when incubated with SIINFEKL (SEQ ID NO: 72) peptide (Fig.13b). This monoculture system could readily detect SIINFEKL (SEQ ID NO: 72) expression following transfection of a modified mRNA vector encoding three in frame copies of the SIINFEKL (SEQ ID NO: 72) antigen (Fig.13c). For direct

quantification of T cell activation, we cocultured with transgenic OT-I CD8+ T cells specific for the SIINFEKL (SEQ ID NO: 72) peptide, layered onto the mRNA-transfected H-2K^{b} expressing HEK 293T cells. Greater than 65% of CD8+ T cells were activated when modified mRNA constructs expressing SIINFEKL (SEQ ID NO: 72) were used for transfection, with less than <2% expressing the activation marker when mRNA constructs not expressing SIINFEKL (SEQ ID NO: 72) were used (Fig.13d-e). Using this system and by placing SIINFEKL (SEQ ID NO: 72) in alternate reading frames or after stop codons or synonymous mutations, in modified mRNA constructs, we could evaluate both the sequence requirements for SIINFEKL (SEQ ID NO: 72) expression and the sequence optimisation strategies to prevent its expression.

To investigate the fidelity of the U*GA stop codon, we modified the previously described +1 frameshift reporter containing the amino-terminal of firefly luciferase (NFLuc) and the carboxy-terminal of firefly luciferase in the +1 frame (+1 CFLuc) to now include SIINFEKL (SEQ ID NO: 72) encoded in the +1 frame in the c-terminus of the construct[1]. If translated normally, a catalytically inactive NFLuc is translated; however, if a +1 frameshift occurs, a full-length, catalytically active SIINFEKL (SEQ ID NO: 72) tagged protein would be generated. An alternative modified construct was made, FLuc U*GA, in which a +1 encoded premature U*GA termination codon was inserted immediately prior to the +1 CFLuc, potentially preventing the expression of full-length luciferase and SIINFEKL (SEQ ID NO: 72). In this construct, production of full-length, functional FLuc and SIINFEKL (SEQ ID NO: 72) expression would depend on the readthrough of the +1 encoded U*GA stop codon (Fig. 12b). Transfection of 1-methylΨ modified FLuc construct resulted in expression of the catalytically active, full-length protein (Fig. 12c-d). In contrast, transfecting with the FLuc U*GA construct resulted in no detectable full-length protein or catalytic activity, indicating that the stop codons were sufficient to prevent stable full-length protein expression (Fig. 12c-d). However, when the same transfected cells were cocultured with OT-I CD8+ T cells, despite the difference in full-length protein expression, both constructs were similarly able to activate the T cells (Fig. 12e). These results are consistent with +1 encoded antigens being DRiPs which due to their extremely short half-lives are not detectable as part of the stable proteome[8-10] but due to the prolonged half-life of peptide-MHC Class I complexes[11,12] can stimulate T cells. The OT-I TCR to SIINFEKL (SEQ ID NO: 72) antigen is particularly high avidity, so to assess if these constructs drove sufficient SIINFEKL (SEQ ID NO: 72) expression *in vivo* to activate a polyclonal repertoire CD8+ T cells of natural avidity, we immunised mice and assessed clonal expansion and differentiation using SIINFEKL (SEQ ID NO: 72) MHC Class I multimers one week later. Although expansion was reduced in mice

immunised with the FLuc U*GA construct, both constructs generated populations of SIINFEKL (SEQ ID NO: 72) reactive cells indicating substantial *in vivo* antigen expression (Fig. 12f). Alongside their ability to drive clonal expansion these constructs were equally able to drive differentiation with equivalent proportions of terminally differentiated cytotoxic KLRG1+CD127- cells making up the SIINFEKL (SEQ ID NO: 72) reactive population (Fig. 12g). There was no evidence of marked avidity differences in the SIINFEKL (SEQ ID NO: 72) reactive cells from these mice, as restimulation with a fixed concentration of SIINFEKL (SEQ ID NO: 72) elicited a comparable average IFN-γ production in splenocytes from these mice (Fig. 12h). These results show how, despite extensive U*GA stop codons in the +1 frame, T cell responses to +1 encoded peptides could be generated by the 1-methylΨ modified COVID-19 vaccines. Furthermore, this demonstrates that the evaluation of *in vitro* translated proteins from 1-methylΨ modified mRNA therapeutics, included in the regulatory submissions of both approved 1-methylΨ modified mRNA vaccines[13,14], provides little insight into off-target T cell antigen expression. Given that translation termination of the Spike protein from the BNT162b2 vaccine is mediated by a U*GA stop codon, we explored whether T cell immunity could be directed towards peptides encoded in-frame from the 3' untranslated region (UTR) of this vaccine. IFN-γ ELISpot responses were detected after stimulation with 3' UTR peptides in samples from people immunised with the Pfizer BioNTech vaccine but not in those immunised with the ChAdOx nCoV-19 DNA vaccine (Fig. 12i). This indicates that in-frame stop codon readthrough in 1-methylΨ modified therapeutics is a potential additional source of misdirected immunity.

Our finding that +1 encoded antigen drives expansion and differentiation of CD8+ T cells into cytotoxic effectors, potentially capable of killing mRNA expressing cells, led us to hypothesise that this could have a negative effect on target protein expression upon repetitive dosing of 1-methylΨ modified mRNA therapeutics. To evaluate this further, we made candidate COVID-19 vaccine constructs in which we fused the receptor binding domain (RBD) of the SARS-CoV-2 Spike protein to the transmembrane domain of the HA protein of influenza (for membrane tethering) either with or without +1 encoded SIINFEKL (SEQ ID NO: 72). By comparing the immune response of these vaccines, we could assess how an immunogenic +1 encoded peptide impacted the target RBD-vaccine response. Following 3-week interval dosing of these vaccines, we could only detect SIINFEKL (SEQ ID NO: 72) specific CD8+ T cells in mice that had received the +1 encoded SIINFEKL (SEQ ID NO: 72) COVID-19 vaccine (Fig. 14a-b). The percentage of SIINFEKL (SEQ ID NO: 72) specific T cells increased in the muscle and draining lymph node (dLN) between 24 and 48hrs after the final immunisation. Analysis showed that, in the muscle, there was a significantly higher percentage of terminally differentiated SIINFEKL (SEQ ID NO: 72)-

specific CD8+ T cells (CD127- KLRG1+) compared to the draining lymph node (dLN). Conversely, the percentage of memory SIINFEKL (SEQ ID NO: 72)-specific CD8+ T cells (CD127+ KLRG1-) was higher in the dLN compared to the muscle (Fig. 14c). Histological assessment of the muscle tissue demonstrated co-localisation of CD8+ T cells with the cytotoxic granule protein, perforin, and other cells expressing the apoptotic marker cleaved caspase 3 but this was only detectable in the muscle tissue of mice that had received the +1 encoded SIINFEKL (SEQ ID NO: 72) vaccine (Fig. 14d). These results prompted the hypothesis that terminally differentiated CD8+ T cells reactive to the off-target +1 encoded antigen could kill cells that are translating the mRNA in subsequent doses and thereby limit the intended target response. To evaluate this, we transferred either SIINFEKL (SEQ ID NO: 72) specific OT-I or wildtype control cytotoxic effector T cells into mice at the same time as a single dose vaccination with these vaccines. After 3 weeks, we found that both the RBD antibody titre and RBD-specific CD8+ T cells were significantly lower in mice that had received the OT-I CD8+ T cell transfer and were immunised with +1 SIINFEKL (SEQ ID NO: 72) encoding vaccine (Fig. 14e), demonstrating that +1 encoded antigen reactive T cells can limit the target protein of 1-methylΨ modified mRNA therapeutics. In keeping with this conclusion, mice immunised with multiple doses of mRNA constructs with +1 encoded SIINFEKL (SEQ ID NO: 72) had a significantly reduced live SARS-CoV-2 virus neutralisation titre following multiple doses compared with the same construct without SIINFEKL (SEQ ID NO: 72) (Fig. 14f).

We evaluated whether our coculture system could identify strategies to reduce +1 encoded antigen expression. First, we introduced the equivalent synonymous mutation as our previous report to remove the slippery site. This mutation does not alter the in frame amino acid sequence but alters the +1 amino acid sequence, which should reduce slippage, as the in frame and +1 sequences can no longer be decoded as the same amino acid. However, we found that this mutation did not affect T cell activation (Fig. 15a), possibly because ribosome slippage had already occurred earlier in the mRNA sequence. Next, we tested whether inserting U*AA or U*AG stop codons immediately before the SIINFEKL (SEQ ID NO: 72) sequence could better terminate antigen translation. Although the insertion of these more stringent stop codons reduced T cell activation, residual T cell activation could still be detected (Fig 15b). However, the use of tandem stop codons (either U*AAU*AA or U*AGU*AG) reduced the CD8+ T cell activation (Fig 15c). We validated these results *in vivo* by immunising mice. Mice immunised with the modified mRNA COVID-19 vaccine in which tandem UAA stop codons were inserted upstream to the +1 encoded SIINFEKL (SEQ ID NO: 72) had no excess SIINFEKL (SEQ ID NO: 72) specific CD8+ T cells (Fig. 15d), suggesting that incorporating tandem UAA or UAG stops could be effective in limiting

antigen expression encoded by the +1 frame or in the 3'UTR of 1-methylΨ modified mRNA therapeutics.

Building on these results we sought to develop a working COVID-19 vaccine in which +1 frameshift and in frame 3'UTR antigen expression was prevented. We therefore tested a COVID-19 vaccine in which (1) the slippery site was removed, (2) a tandem UAA stop terminated the in frame coding sequence and (3) single or tandem U*AA or U*AG stop codons were generated in the +1 frame with synonymous in frame nucleotide substitutions. We compared this sequence optimised +1 encoded SIINFEKL (SEQ ID NO: 72) expressing vaccine with a non-optimised (PfizerRBD) +1 encoded SIINFEKL (SEQ ID NO: 72) vaccine and the optimised construct in which the +1 encoded SIINFEKL (SEQ ID NO: 72) was removed (Fig. 16a). *In vitro* assessment of these constructs ability to express functional antigen demonstrated no significant difference between the sequence optimised +1 encoded SIINFEKL (SEQ ID NO: 72) expressing construct and the construct without SIINFEKL (SEQ ID NO: 72), with both exhibiting significantly reduced ability to activate OT-I CD8+ T cells than the PfizerRBD construct (Fig. 16b). These vaccines were then assessed *in vivo* by immunising animals with a three-dose schedule. Mice in receipt of the PfizerRBD construct had significantly elevated SIINFEKL (SEQ ID NO: 72) specific CD8+ T cells in the spleen and muscle, with no significant difference noted between the optimised constructs, indicating that our optimisation strategy had prevented significant +1 encoded antigen expression (Fig. 16c-d). We then assessed if the expression of the +1 encoded SIINFEKL (SEQ ID NO: 72) impacted the target Spike vaccine response. Mice in receipt of the sequence optimised COVID-19 vaccine had elevated levels of Spike reactive T cells (Fig. 16e) and higher live virus neutralisation capacity (Fig. 16f) consistent with our earlier findings that expression of the +1 encoded SIINFEKL (SEQ ID NO: 72) limited the in frame target vaccine response after repetitive doses. We compared our optimised vaccine in the +1 frame with Pfizer BNT162b2 vaccine and found reduced ELISpot responses when stimulated with overlapping peptide pools covering the +1 frame of the RBD sequence and the 3'UTR (Fig. 16g-h). Taken together these results indicate that the optimised vaccine without SIINFEKL (SEQ ID NO: 72) encoded is both effective and has no detectable evidence of misdirected T cell immunity to either +1 frame encoded or in frame 3'UTR antigens.

Conclusions

The +1 frame encoded U*GA stop codons found throughout Pfizer BNT162b2 and Moderna mRNA-1273 vaccine (73 +1 encoded U*GA stops) may have resulted from sequence optimisation, based on codon usage, aimed at maximising protein expression. In people, the most frequent codons for leucine and valine are CUG and GUG, which when followed by nucleotides encoding isoleucine, threonine, arginine, and asparagine, as well as certain

nucleotide triplets encoding serine, lead to the formation of a +1 encoded UGA stop codon. The readthrough of the UGA termination codon in 1-methylΨ modified mRNA is consistent with previous reports for pseudouridine modified mRNA[15,16]. Although optimisation strategies that result in this codon in the +1 frame can be advantageous in increasing target protein expression and preventing expression of stable protein encoded from the +1 frame, we have shown that it is not sufficient to prevent expression of +1 encoded T cell antigens. Demonstrating the effectiveness of an *in vitro* system for assessing antigen expression from the +1 frame, we find that tandem use of the UAA and UAG stop codons can reduce +1 encoded antigen expression which in addition to ribosome frameshifting could also arise from alternative translational initiation in the +1 frame of the mRNA sequence. Using this information we have made an effective COVID-19 vaccine with no detectable off-target immune responses -- an approach that could be adopted to make an effective vaccine against current circulating variants of SARS-CoV-2. Whilst currently there is no evidence that mistranslated T cell antigens cause harm, we have shown that unintended T cell antigens, can limit the response to the target RBD antigen in the case of a COVID-19 vaccine. Although important for vaccines, this is perhaps most relevant for mRNA therapies aimed at making a therapeutically active protein, in which a cytotoxic T cell response targeting modified mRNA expressing cells will reduce production of the active protein. Consequently, this report shows how sequence optimisation can be used to both minimise mistranslation related immune responses and maintain therapeutic efficacy from modified mRNA therapeutics.

Methods

Plasmids and mRNA

DNA fragments were obtained from GeneWiz and cloned into a plasmid vector, using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen, 450245). Fragments were PCR amplified from plasmids to produce a DNA template for mRNA synthesis. mRNA was synthesised by in vitro transcription using the HiScribe T7 mRNA Kit with CleanCap (New England Biolabs, E2040S) with UTP substituted for N1-methylpseudoUTP (TriLink Biotechnologies, N-1081). Polyadenylated tails were added enzymatically using the E. coli Poly(A) Polymerase Kit (New England Biolabs, M0276L) and then purified using the RNeasy Mini Kit (Qiagen, 74004). Purified mRNA was quantified using the Qubit 3 Fluorometer (Invitrogen) using the Qubit RNA Broad Range Assay Kit (Invitrogen, Q10211) and analysed for integrity by agarose gel electrophoresis before being stored at -80°C.

Cell culture, transfection and coculture assay

HEK293T and HEK293T H-2K^b cells were cultured in High Glucose DMEM (Gibco, 41965039) supplemented with 10% (v/v) FBS and 50 U/ml penicillin-streptomycin. Cells were grown at 37°C, 95% humidity and 5% CO_2. When seeded for transfection cells were plated in antibiotic free DMEM supplemented with 10% (v/v) FBS.

Lentiviral transduction

HEK293T cells were co-transfected with envelope plasmid pMD2.G (100ng), packaging plasmid psPAX2 (900ng), and lentiviral transfer H-2K^b plasmid (1µg) using TransIT-LT1 (Mirus Bio, MIR2304). Cells were incubated for 18 hours at 37 °C before growth media was replaced. Cells were cultured for a further 24 hours before media was harvested and filtered (0.45µm). Filtered virus was then spun onto HEK293T cells for 30 minutes with polybrene (8µg/ml) and cells cultured overnight. Transduced cells were treated with puromycin to select successfully transduced cells.

Coculture Assay

HEK293T H-2K^b cells were plated at 5x10^4/well in 24 well plates approximately 16 hours before transfection. Cells were transfected with 500ng of mRNA using Lipofectamine 3000 (Invitrogen, L3000015) following manufacturer recommendations. After 24 hours, OT-I naïve CD8+ T cells were isolated using the mouse Naïve CD8+ T cell isolation kit (Miltenyi Biotec, 130-096-543) and were layered over transfected cells (1x10^4 OT-I T cells/well). After 24 hours cells were assessed by flow cytometry.

Western blot

HEK293T H-2K^b cells were plated at 5x10^5 cells/well in 6 well plates and transfected with 2.5µg mRNA and cultured for 24 hours. Cells were lysed with RIPA buffer (Thermo Scientific, 89900) containing protease/phosphatase inhibitor (Thermo Scientific, 78442). Cell lysates were mixed with 4x LDS PAGE buffer (Invitrogen, NP0007) and denatured at 70°C for 10 minutes. Cooled samples were run on NuPAGE 4-12% Bis-Tris Mini Protein Gels (Invitrogen) and transferred to a nitrocellulose membrane. Membranes were probed using an anti-Flag antibody (1/2000) (Invitrogen, MA1-91878) and anti-mouse-HRP antibody (1/10000) (Invitrogen, G-21040) and detected with Pierce ECL Western Blotting Substrate (Thermo Scientific, 32106). Membranes were stripped with Restore Western Blot Stripping Buffer (Thermo Scientific, 21059) and re-probed with anti-β-actin (1/4000) (Abcam, AB8227)

and anti-rabbit-HRP antibody (1/10000) (Invitrogen, G21234). Membranes were imaged using a ChemiDoc MP Imaging System (BioRad).

Luciferase assay

HEK293T H-2K^b cells were plated at 5x10^4/well in 24 well plates and transfected with 500ng mRNA and cultured for 24 hours. Luciferase activity was assessed using the Luciferase Assay System (Promega E4550), following manufacturer protocols and were measured using a VICTOR Nivo Multimode Plate Reader (PerkinElmer).

Flow cytometry

Cells were stained with specific antibody cocktails (Table 5). Samples were acquired on the BD LSRFortessa using FACSDIVA software (BD Bioscience). FCS files were exported and analysed using FlowJo v10.7.2 (BD Bioscience) software.

Table 5 Flow cytometry antibodies details

Coculture Assay	Source	Identifier	Clone	Concentration/Dilution
Mouse monoclonal anti-mouse H-2Kb	Leinco Technologies	Y200	Y-3	1/500
Anti-mouse IgG2b APC	Biolegend	406712	RMG2b-1	1/500
Anti-mouse H-2Kb bound to SIINFEKL APC/Fire750	Biolegend	141614	25-D1.16	1/100
Rat monoclonal anti-mouse CD8b APC	Biolegend	140410	53-5.8	1/200
Armenian Hamster monoclonal anti-mouse CD69 BV711	Biolegend	104537	H1.2F3	1/50
Rat monoclonal anti-mouse CD45 FITC	Biolegend	103108	30-F11	1/100

SIINFEKL specific T cells	Source	Identifier	Clone	Concentration/Dilution
Mouse anti-mouse CD45.2 BUV395	BD Biosciences	564616	104	1/200
Armenian hamster monoclonal anti-mouse TCR β APC/Fire750	Biolegend	109246	H57-597	1/200
Rat monoclonal anti-mouse CD4 BV605	Biolegend	116027	RM4-4	1/200
Rat monoclonal anti-mouse CD8a BV421	Biolegend	100737	53-6.7	1/200
SIINFEKL Dextramer PE	Immudex	JD02163		10ul/test
Rat monoclonal anti-mouse CD62L BV711	Biolegend	104445	MEL-14	1/200
Rat monoclonal anti-mouse/human CD44 BV650	Biolegend	103049	IM7	1/200
Rat monoclonal anti-mouse CD127 AF488	Biolegend	135018	A7R34	1/200
Syrian hamster monoclonal anti-mouse/human KLRG1 APC	Biolegend	138412	2F1	1/150
Mouse monoclonal anti-mouse CX3CR1 BV785	Biolegend	149029	SA011F11	1/200

IFN-γ assay	Source	Identifier	Clone	Concentration/Dilution
Armenian hamster monoclonal anti-mouse TCR β APC/Fire750	Biolegend	109246	H57-597	1/200
Rat monoclonal anti-mouse CD8a APC	Biolegend	100712	53-6.7	1/200
Rat monoclonal anti-mouse/human CD44 BV650	Biolegend	103049	IM7	1/200
Rat monoclonal anti-mouse IFN-γ PE	Biolegend	505808	XMG1.2	1/100
AIM assay	**Source**	**Identifier**	**Clone**	**Concentration/Dilution**
Armenian hamster monoclonal anti-mouse TCR β APC/Fire750	Biolegend	109246	H57-597	1/200
Rat monoclonal anti-mouse CD8a PerCP/Cyanine5.5	Biolegend	100734	53-6.7	1/200
Rat monoclonal anti-mouse PD-1 BV421	Biolegend	135221	29F.1A12	1/200
Armenian hamster anti-mouse CD69 BV650	Biolegend	104541	H1.2F3	1/200
Rat monoclonal anti-mouse CD107a BV711	Biolegend	121631	1D4B	1/200
Rat monoclonal anti-mouse CD25 PE	Biolegend	113704	A18246A	1/200

Lipid nanoparticle formulation

Lipids were dissolved in 100% ethanol and mixed DSPC: Cholesterol: ALC-0159: ALC-0315 in a molar ratio of 9.4:42.6:1.7:46.3 to a final concentration of 8mg/ml. mRNA was diluted in 10mM citrate buffer at pH 4 and mixed with lipids at a 3:1 mRNA:lipid ratio using a NanoAssemblr Ignite (Precision Nanosystems). LNPs were then diluted in sterile PBS and concentrated using a centrifugal filter (100 kDa MWCO). Concentrated LNPs were passed through a 0.2µm sterile filter and quantified using the Quant-iT RiboGreen RNA Kit (Invitrogen, R11490).

Ethics statement

Animal experiments were licensed by the UK Home Office according to the Animals Scientific Procedures Act 1986 (License PP6047951) and approved by local ethics committees from the University of Cambridge. Human sample collection and analysis was conducted in accordance with the principles of good clinical practice and following approved protocols of the NIHR National BioResource. Samples were collected with the written informed consent of all study participants under the NIHR National BioResource-Research Tissue Bank ethics (research ethics committee (REC): 17/EE/0025) and from the PITCH study. PITCH is a substudy of the SIREN study, which was approved by the Berkshire REC, Health Research 250 Authority (IRAS ID 284460, REC reference 20/SC/0230), with PITCH recognized as a substudy on 2 December 2020. SIREN is registered with ISRCTN (Trial ID: 252 ISRCTN11041050). Some participants were recruited under aligned study protocols. In Liverpool, some participants were recruited under the study Human immune responses to acute virus infections (16/NW/0170), approved by North West - Liverpool Central REC on 8 March 2016, and amended on 14 September 2020 and 4 May 2021. In Oxford, participants were recruited under the GI Biobank Study 16/YH/0247, approved by the REC at Yorkshire & The Humber - Sheffield REC on 29 July 2016, which was amended for this purpose on 8 June 2020. The study was conducted in compliance with all relevant ethical regulations for work with human participants, and according to the principles of the Declaration of Helsinki (2008) and the International Conference on Harmonization Good Clinical Practice guidelines. Written informed consent to publish clinical and genetic data, as well as for study participation, was obtained for all participants enrolled in the study.

Mouse immunisations

C57BL/6J mice (8-12 weeks) were immunised intramuscularly with mRNA-LNP. When given multiple doses mice were immunised 3 weeks apart into the same site of primary immunisation. Blood samples were collected via saphenous vein bleeds at the indicated time points. Mice were culled at the indicated time point after final vaccination and spleens, draining lymph node, muscle tissue and serum were collected. Spleens and lymph node were mashed and filtered through 70µm cell strainers. Red blood cells were lysed using RBC lysis buffer (155mM NH_4Cl, 12 mM $NaHCO_3$, 0.1 mM EDTA) and then filtered through 40µm cell strainers. Cells were taken directly for immunofluorescent staining of SIINFEKL specific T cells or were cryopreserved. The biceps femoris muscle was dissected and processed using the Skeletal Muscle Dissociation kit (Miltenyi Biotec, 130-098-305), as described in the manufacturers protocol. Red blood cells were lysed and cells were stained immediately for flow cytometry analysis.

IFN-γ assay

Splenocytes ($1x10^6$) were treated with Brefeldin A (1µl/$1x10^6$ cells) (BD Bioscience, 55029), IL-2 (20ng/ml) (Biolegend, 575406) and stimulated with the SIINFEKL peptide (0µM or 1µM) (Invivogen, vac-sin). Cells were incubated at 37°C for 6 hours, and then were stained for extracellular surface markers. Cells were then fixed and permeabilised using the BD Cytofix/Cytoperm™Fixation/Permeabilization Kit (BD Biosciences, 554714) and stained for intracellular IFN-γ.

Effector cell differentiation and transfer

96-well plates were coated with 10µg/ml anti-CD3 (Invitrogen, 17A2, 16-0032085) in PBS overnight at 4°C. CD8+ T cells were isolated from WT or OT-I spleens using the CD8a+ T Cell Isolation Kit, mouse (Miltenyi Biotec 130-104-075). T cells were seeded at a density of $5x10^4$ cells/well in complete RPMI (Gibco, 21875-034) supplemented with FBS 10% (v/v), 1x GlutaMAX™ (Gibco, 35050061), 55mM 2-mercaptoethanol, 50U/ml penicillin-streptomycin, 2ng/ml IL-7 (Biolegend, 577802), 20ng/ml IL-2 (Biolegend, 575406) and 5µg/ml anti-CD28 (Invitrogen, 16-0281-85) final concentration and incubated at 37°C. At 48 and 96 hours, all cells were detached, counted and re-seeded in 96-well plates at $5x10^4$ cells/well in complete media with the exclusion of anti-CD28. At 5 days, all cells were detached, washed, counted and phenotypically assessed by flow cytometry prior to intravenous transfer of $5x10^6$ T cells per mouse.

ELISA

Plates were coated with trimeric spike protein (1µg/ml) (Abcam, AB288548) at 4°C overnight. Plates were then blocked with 1% BSA, 0.2% Tween at room temperature for 30 minutes. Sample serum was incubated on the plate at room temperature for 2 hours, and then incubated with anti-mouse IgG HRP (1/1000) (Invitrogen, 62-6520) for 1 hour. Plates were then incubated with TMB substrate (Thermo Scientific, N301) for 20 minutes before the reaction was stopped by addition of 1M HCl. Absorbance at 450nm was measured using a VICTOR Nivo Multimode Plate Reader (PerkinElmer).

NT50

Luminescent HEK293T-ACE2-30F-PLP2 reporter cells (clone B7) expressing ACE2 and SARS-CoV-2 Papain-like protease-activatable circularly permuted firefly luciferase (FFluc) are available from the National Institute for Biological Standards and Control (NIBSC, www.nibsc.org, catalogue number 101062). Sera were heat-inactivated at 56 °C for 30 min before use, and neutralising antibody titres at 50% inhibition (NT_{50s}) measured as previously described. For purposes of visualisation and ranking, samples for which visual inspection of the titration curve indicated inhibition at low dilutions, but which did not meet criteria for quantification, were assigned an arbitrary NT_{50} of 4.

Activation induced marker (AIM) assay

Splenocytes were thawed in RPMI 1640 medium supplemented with 50U/ml penicillin–streptomycin, containing 0.01% (v/v) Benzonase nuclease. Cells were washed and then incubated for 2 hours at 37 °C, 5% CO_2 in RPMI1640 medium, 10% (v/v) FBS and 1% (v/v) penicillin–streptomycin. Cells ($7.5x10^5$ cells/sample) were stimulated with peptides (4µg/ml) in RPMI1640 medium, IL-2 (20ng/ml), IL-7 (2ng/ml) and 1/200 anti-CD107a antibody for 18 hours. Unstimulated, DMSO treated, and Concanavalin A treated cells were included as controls. Cell surface markers were stained for 30 minutes at 4°C.

T cell cytokine production by ELISPOT

IFN-γ ELISpot assays were carried out as previously described using the human IFN-γ ELISpot PLUS kit (ALP). Overlapping peptide pools corresponding to: in-frame SARS-CoV-2 spike RBD and TMD; peptides predicted to occur by +1 translation of these domains; and peptides corresponding to the 3'UTR of the BNT162b2 vaccine. Peptides were 15-mers with

11 amino acid overlaps and obtained from Mimotopes. Cryopreserved splenocytes and PBMCs were thawed in RPMI1640 medium supplemented with 50 U/ml penicillin–streptomycin, containing 0.01% (v/v) Benzonase nuclease. Cells were washed and then incubated for 2-3 hours at 37°C, 5% CO_2 in RPMI1640 medium, 10% (v/v) FBS and 50U/ml penicillin–streptomycin. Pre-coated IFN-γ ELISpot 96-well plates (MabTech) were washed three times with PBS and then blocked with RPMI1640 medium, 10% (v/v) FBS and 50 U/ml penicillin–streptomycin for 45 min. Overlapping peptide pools were plated at 4µg/ml with dimethylsulfoxide (DMSO) as the negative control. Cells ($2x10^5$/well) were resuspended in RPMI1640 medium, 10% (v/v) FBS and 50 U/ml penicillin-streptomycin. IL-2 (20ng/ml) and IL-7 (2ng/ml) was added to assess 3'UTR responses. Cells were discarded, and plates were washed with PBS 0.05% (v/v) Tween and incubated with IFN-γ detector antibody (clone 7-B6-1, 1µg/ml) for 3 hours at room temperature. Washed plates were then incubated with streptavidin alkaline phosphatase antibody (1µg/ml) for 1 hour 30 minutes. Plates were washed and then colour development was carried out using 1-step NBT/BCIP substrate solution. Filtered NBT/BCIP was added to each well for 5 minutes at room temperature after which development was stopped with cold water. Plates were dried at room temperature before quantification using an AID iSpot Spectrum EliSpot Reader (AID EliSpot Software version 7.0, Autoimmun Diagnostika). Spot counts were expressed as DMSO corrected spot forming units (SFU) per million cells.

EXAMPLE 4

In order to assess *in vitro* the antigenicity potential of SARS-CoV-2 receptor binding domain (RBD) constructs containing the UTRs identified, the constructs shown in Figure 17 were generated which contained either Pfizer of Moderna spike vaccine UTRs or UTRs identified here along with the influenza HA transmembrane domain and MHC class I epitope SIINFEKL. Specifically constructs were generated containing the UTRs from muscle UTR 4 (MUTR4 – including a 5' UTR according to SEQ ID NO 41 and a 3' UTR according to SEQ ID NO: 42) and liver UTR1, 2 and 3 (LUTR1– including a 5' UTR according to SEQ ID NO 56 and a 3' UTR according to SEQ ID NO: 57, LUTR2– including a 5' UTR according to SEQ ID NO 58 and a 3' UTR according to SEQ ID NO: 59, LUTR3– including a 5' UTR according to SEQ ID NO 60 and a 3' UTR according to SEQ ID NO: 61). IVT mRNAs were then prepared by *in vitro* transcription from DNA templates as described under IVT mRNA synthesis above and used in an *in vitro* antigen expression system based on HEK293T cells transfected to express the H-$2K^b$ murine MHC Class I molecule. Direct quantification of B and T cell activation was determined by coculture with B and T cells specific for the SINFEKL peptide (SEQ ID NO: 72) with IVT mRNA transfected H-$2K^b$ expressing HEK 293T cells.

Sequences

Name	Sequence	Notes	TYPE	SEQ ID NO
FQ	GAAAUAAGAGAGAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC**AUGAAGUGGGUGACCUUCAUCA** **GCCUGCUGUUCCUGUUCAGCAGCGCCUACAGC**GUGUUCACCCUGGAGGACUUCGUGGGCGACUGG AGACAGACCGCCGGCUACAACCUGGACCAGGUGCUGGAGCAGGGCGGCGUGAGCAGCCUGUUCCAG AACCUGGGCGUGAGCGUGACCCCCAUCCAGAGAAUCGUGCUGAGCGGCGAGAACGGCCUGAAGAUC GACAUCCACGUGAUCAUCCCCUACGAGGGCCUGAGCGGCGACCAGAUGGGCCAGAUCGAGAAGAUC UUCAAGGUGGUGUACCCCGUGGACGACCACCACUUCAAGGUGAUCCUGCACUACGGCACCCUGGUG AUCGACGGCGUGACCCCCAACAUGAUCGACUACUUCGGCAGACCCUACGAGGGCAUCGCCGUGUUC GACGGCAAGAAGAUCACCGUGACCGGCACCCUGUGGAACGGCAACAAGAUCAUCGACGAGAGACUGA UCAACCCCGACGGCAGCCUGCUGUUCAGAGUGACCAUCAACGGCGCGUGACCGGCUGGAGACUGUGCG AGAGAAUCCUGGCCUGAUGAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCC CUAAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAA AAACAUUUAUUUUCAUUGC	Underline = UTRs Bold = signal peptide poly(a) not depicted	RNA	1
GS	GAAAUAAGAGAGAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC**AUGAAAUGGGUCACGUUUCAUCU** **CUUUGCUUUUCCUGUUUUCAUCCGCGUACUCCG**CGUGUUCACCCUGGAGGACUUCGUCGGGGACUGG CGCCAGACGGGCUGGCUACAACCUGGACCAGGUCCUGGAACAGGGCGGAGUGUCGUCUCUCCUUCCAG AACCUGGGCGUUAGUGUGACUCCCAUCCAGCGCAUCGUGCUGAGCGGGGAGAAUGGGCUGAAGAUU GACAUCCACGUCAUCAUACCUUACGAGGGCCUCAGCGGAGACCAGAUGGGCCAAAUUGAGAAGAUCU UCAAGGUGGUGUACCCAGUGGAUGACCACCACUUUAAAGUAAUCCUGCAUUACGGCACACUAGUGAU UGAUGGUGUGACUCCGAAUAUGAUCGAUUACUUCGGUAGGCCCUAUGAGGGCAUCGCCGUGUUCGA CGGCAAGAAGAUCACCGUAACCGGCACCCUUUGGAACGGCAACAAGAUCAUCGACGAGCGUCUCAUU AACCCUGACGGCUCCCUGCUGUUUCGCGUGACCAUCAACGGUUGGCGCCUGUGCGAG CGGAUUUUGGCCUGAUAAGCUCGCUUUCUUGCUGUCCCAAUUUCUAUUAAAGGUUCCUUUGUUCCCU AAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAA ACAUUUAUUUUCAUUGC	Underline = UTRs Bold = signal peptide poly(a) not depcited	RNA	2

MIN	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC**AUGAAGUGGGUCACCUUUAUCU** **CCUUAUUAUUUUUUAUUUUCCUCCGCCUAUUCC**GUCUUUUACCUUAGAGGAUUUUUGUCGGUGAUUGGC GCCAGACCGCCGGUUAUAAUUUAGAUCAGGUCUUAGAGCAGGGUGGUGUCUCCUCCUUAUUUCAGA AUUUAGGUGUCUCCGUCACCCCCAUCCAGCGCAUCGUCUUAUCCGGUGAGAAUGGUUUAAAGAUCG AUAUCCAUGUCAUCAUCCCCCUAUGAGGGUUUAUCCGGUGAUCAGAUGGGUCAGAUCGAGAAGAUCUU UAAGGUCGUCUAUCCCCGUCGAUGAUCAUCAUUUUUAAGGUCAUCUUACAUUAUGGUACCUUAGUCAUC GAUGGUGUCACCCCCAAUAUGAUCGAUUAUUUUGGUCGCCCCUAUGAGGGUAUCGCCGUCUUUGAU GGUAAGAAGAUCACCGUCACCGGUACCUUAUGGAAUGGUAAUAAGAUCAUCGAUGAGCCGCUUAAUCA AUCCCGAUGGUUCCUUAUUAUUUUCGCGUCACCAUCAAUGGUGUCACCGGUUGGCGCUUAUGUGAGC GCAUCUUAGCCUGAUAAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCCUAA GUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAAAC AUUUAUUUUUCAUUGC	Underline = UTRs Bold = signal peptide poly(a) not depcited	R N A	3
STA	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC**AUGAAGUGGGUAACCUUUAUUU** **CCCUUCUUUUUCUCUUUAGCUCGGCUUAUUCC**GUCUUCACACUCGAAGAUUUCGUUGGGGACUGGC GACAGACAGCCGGCUACAACCUGGACCAAGUCCUUGAACAGGGAGGUGUGUCCAGUUUGUUUCAGA AUCUCGGGGUGUCCGUAACUCCGAUCCAAAGGAUUGUCCUGAGCGCGGUGAAAAUGGGCUGAAGAUCG ACAUCCAUGUCAUCAUCCCGUAUGAAGGUCUGAGCGGCGACCAAAUGGGCCAGAUCGAAAAAAUUUUU UAAGGUGGUGUACCCUGUGGAUGAUCACUCACUUUAAGGUGAUCCUGCACUAUGGCACACUGGUAAU CGACGGGGUUACGCCGAACAUGAUCGACUAUUUUCGGACGGCCGUAUGAAGGCAUCGCCGUGUUUCGA CGGCAAAAAGAUCACUGUAACAGGGACCCUGUGGAACGGCAACAAAAUUAUCGACGAGCGCCUGAUC AACCCCGACGGCUCCCCUGCUGUUCCGAGUAACCAUCAACGGAGUGACCGGCUGGCGGCUGUGCGAA CGCAUUCUGGCGUGAUAAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCCU AAGUCCAACUACUAAACUGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAA ACAUUUAUUUUUCAUUGC	Underline = UTRs Bold = signal peptide poly(a) not depcited	R N A	4
IS-OPT	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC**AUGAAAUGGGUUACUUUCAUUA** **GCCUGCUGUUCCUGUUCAGCAGCGCUUACAGC**GUUUUUCACUCUGGAAGACUUCGUUGGAGACUGGA GACAAACUGCUGGAUACAACCUGGACCAAGUUUCUGGAACAAGGAGGAGUUAGCAGCCUGUUCCAAAA CCUGGGAGUUAGCGUUACUCCUAUUCAAAGAAUUGUUCUGAGCGGAGAAAACGGACUGAAAAUUGAC AUUCACGUUAUAUUAUUCCUUACGAAGGACUGAGCGGAGACCAAAUGGGACAAAUUGAAAAAAUUUUCA AAGUUGUUUACCCUGUUGACGACCACCACUUCAAAGUUAUUCUGCACUACGGAACUCUGGUUAUUGA CGGAGUUACUCCUAACAUGAUUGACUACUUCGGAAGACCUUACGAAGGAAUUGCUGUUUUCGACGGA AAAAAAUUACUGUUACUGGAACUCUGUGGGAAACAAAAUUAUUGACGAAAGACUGAUUAACCC UGACGGAAGCCUGCUGUUCAGAGUUACUAUUAACGGAGUUACUGGAUGGAGACUGUGCGAAAGAAU UCUGGCUUAGUGAAAGCUUGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCC	Underline = UTRs Bold = signal peptide poly(a) not depcited	R N A	5

	UAAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAA AACAUUUAUUUUCAUUGCGGAUCC			
SKMC-OPT	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGCCACC**AUGAAGUGGGUGACGUUCAUUA GCCUUCUCUUCCUCCUUCAGCAGCGCUUACAGC**GUGUUCACUCUUGAGGACUUCGUGGGCGACUGGA GGCAGACUGCUGGCUACAACCUUGACCAGGUGCUUGAGCAGGGCGGCGUGAGCAGCCUCUUCCAGA ACCUUGGCGUGAGCGUGACUCCUAUUCAGAGGAUUGUGCUUAGCGGCGAGAACGGCCUUAAGAUUG ACAUUCACGUGAUAUAUUCCUUACGAGGGCCUUAGCGGCGACCAGAUGGGCCAGAUUGAGAAGAUCU UCAAGGUGGUGUACCCUGUGGACGACCACCACUUCAAGGUGAUUCUUCACUACGGCACUCUUGUGA UUGACGGCGUGACUCCUAACAUGAUUGACUACUUCGGCAGGCCUUACGAGGGCAUUGCUGUGUUCG ACGGCAAGAAGAUUACUGUGACUGGCACUCUUUGGAACGGCAACAAGAUUAUUGACGAGAGGCUUAU UAACCCUGACGGCAGCCUUCUCUUCAGGGUGACUAUUAACGGCGUGACUGGCUGGAGGCUUUGCGA GAGGAUUCUUGCUUGAUAAAAGCUUGCUCGCUUUCUUGCUGUCCAAUUUCUACUUUAAAGGUUCCUUU GUUCCCUAAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUA AUAAAAAACAUUUAUUUUCAUUGC			
A1AG	MALSWVLUVLSLLPLLEA		A A	7
A1AG NT	AUGGCGCUGUCCUGGGUUCUUACAGUCCUGAGCCUCCUACCUCUGCUGGAAGCC		R N A	8
A1AG Optimised NT	AUGGCCCUGAGCUGGGUGCUGACCGUGCUGAGCCUGCUGCCCCUGCUGGAGGCC		R N A	9
AACT	MERMLPLLALGLLAAGFCPAVLCHP		A A	10
AACT NT	AUGGAGAGAAUGUUACCUCUCCUGGCUCUUGGGGCUCUUGGCGGCUGGGUUCUGCCCUGCUGUCCU CUGCCACCCU		R N A	11

AACT optimised NT	AUGGAGAGAAUGCUGCCCCUGCUGGCCCUGGGCCUGCUGGCCGCCGGCUUCUGCCCCGCCGUGCU GUGCCACCCC		RNA 12
A1AT	MPSSVSWGILLLAGLCCLVPVSLA		RNA 13
A1AT NT	AUGCCGUCUUCUGUCUCGUGGGGCAUCCUCCUGCUGGCAGGCCUGUGCUGCCUGGUCCCUGUCUC CCUGGCU		RNA 14
A1AT optimised NT	AUGCCCAGCAGCGUGAGCUGGGGCAUCCUGCUGCUGGCCGGCCUGUGCUGCCUGGUGCCCGUGAG CCUGGCC		RNA 15
PEBP1	MGWUMRLVUAALLLGLMMVVUG		RNA 16
PEBP2 NT	AUGGGUUGGACAAUGAGGCUGGUCACAGCAGCACUGUUACUGGGUCUCAUGAUGGUGGUCACUGGA		RNA 17
PEBP3 optimised NT	AUGGGCUGGACCAUGAGACUGGUGACCGCCGCCCUGCUGCUGGGCCUGAUGAUGGUGGUGACCGG C		RNA 18
HPT	MSALGAVIALLLWGQLFA		RNA 19
HPT NT	AUGAGUGCCCUGGGAGCUGUCAUUGCCCUCCUGCUCUGGGGACAGCUUUUUGCA		RNA 20
HPT optimised NT	AUGAGCGCCCUGGGCGCCGUGAUCGCCCUGCUGUGGGGCCAGCUGUUCGCC		RNA 21
HSA	MKWVUFISLLFLFSSAYS		RNA 22
HSA NT	AUGAAGUGGGUAACCUUUAUUUCCCUUCUUUUUCUCUUUAGCUCGGCUUAUUCC		RNA 23
HSA optimised NT	AUGGUGUUCACCCUGGAGGACUUCGUGGGCGACUGGAGACAGACCGCCGGCUAC		RNA 24

NT	MFVFLVLLPLVSS		AAA 25
NT NT	AUGUUCGUGUUCCUGGUGCUGCUGCCCCCUGGUGAGCAGC		RNA 26
NT optimised NT	AUGUUCGUGUUCCUGGUGCUGCUGCCCCUGGUGAGCAGC		RNA 27
FQ nLuc + AIAG1	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAACAGCCACC*AUGGCCCUGAGCUGGGUGCUGA* *CCGUGCUGAGCCUGCUGCCCCCUGCUGGAGGCCG*GUGUUCACCCUGGAGGACUUCGUGGGCGACUGG AGACAGACCGCCGGCUACAACCUGGACCAGGUGCUGGAGCAGGGCGGCGUGAGCAGCCUGUUCCAG AACCUGGGCGUGAGCGUGACCCCCAUCCAGAGAAUCGUGCUGAGCGGCGAGAACGGCCUGAAGAUC GACAUCCACGUGAUCAUCCCCUACGAGGGCCUGAGCGGCGACCAGAUGGGCCAGAUCGAGAAGAUC UUCAAGGUGGUGUACCCCGUGGACGACCACCACUUCAAGGUGAUCCUGCACUACGGCACCCUGGUG AUCGACGGCGUGACCCCCAACAUGAUCGACUACUUCGGCAGACCCUACGAGGGCAUCGCCGUGUUC GACGGCAAGAAGAUCACCGUGACCGGCACCCUGUGGAACGGCAACAAGAUCAUCGACGAGAGACUGA UCAACCCCGACGGCAGCCUGCUGUUCAGAGUGACCAUCAACGGCGUGACCGGCUGGAGACUGUGCCG AGAGAAUCCUGGCCUGAUGAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCCC CUAAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAA AAACAUUUAUUUUUCAUUGC	Underline = signal peptide	RNA 28
FQ nLuc + AACT	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAACAGCCACC*AUGGAGAAUGCUGCCCCUGC* *UGGCCCUGGGCCUGCUGGCCGGCCUUCUGCCCCGCCGUGCUGUGCCACCC*GUGUUCACCCUG GAGGACUUCGUGGGCGACUGGAGACAGACCGCCGGCUACAACCUGGACCAGGUGCUGGAGCAGGGC GGCGUGAGCAGCCUGUUCCAGAACCUGGGCGUGAGCGUGACCCCCAUCCAGAGAAUCGUGCUGAGC GGCGAGAACGGCCUGAAGAUCGACAUCCACGUGAUCAUCCCCUACGAGGGCCUGAGCGGCGACCAG AUGGGCCAGAUCGAGAAGAUCUUCAAGGUGGUGUACCCCGUGGACGACCACCACUUCAAGGUGAUC CUGCACUACGGCACCCUGGUGAUCGACGGCGUGACCCCCAACAUGAUCGACUACUUCGGCAGACCC UACGAGGGCAUCGCCGUGUUCGACGGCAAGAAGAUCACCGUGACCGGCACCCUGUGGAACGGCAAC AAGAUCAUCGACGAGAGACUGAUCAACCCCGACGGCAGCCUGCUGUUCAGAGUGACCAUCAACGGCG UGACCGGCUGGAGACUGUGCGAGAGAAUCCUGGCCUGAUGAGCUCGCUUUCUUGCUGUCCAAUUUC UAUUAAAGGUUCCUUUGUUCCCUAAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGC AUCUGGAUUCUGCCUAAUAAAAAACAUUUAUUUUCAUUGC	Underline = signal peptide	RNA 29

FQ nLuc + A1AT	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC*AUGCCCAGCAGCGUGAGCUGGG* *GCAUCCUGCUGCUGGCCGGCCUGUGCUGCCUGGUGCCCGUGAGCCUGGCC*GUGUUCACCCUGGAG GACUUCGUGGGGCGACUGGAGACAGACCGCCGCCGGCUACAACCUGGACCAGGUGCUGGAGCAGGGGCGG CGUGAGCAGCCUGUUCCAGAACCUGGGCGUGAGCGUGACCCCCAUCCAGAGAAUCGUGCUGAGCGGCGG CGAGAACGGCCUGAAGAUCGACAUCCACGUGAUCAUCCCCUACGAGGGCCUGAGCGGCGACCAGAU GGGCCAGAUCGAGAAGAUCUUCAAGGUGGUGUACCCCGUGGACGACCACCACUUCAAGGUGAUCCU GCACUACGGCACCCUGGUGAUCGACGGCGUGACCCCCAACAUGAUCGACUACUUCGGCAGACCCUA CGAGGGCAUCGCCGUGUUCGACGGCAAGAAGAUCACCGUGACCGGCACCCUGUGGAACGGCAACAA GAUCAUCGACGAGAGACUGAUCAACCCCGACGGCAGCCUGCUGUUCAGAGUGACCAUCAACGGCGU GACCGGCUGGAGACUGUGCGAGAGAAUCCUGGCCUGAUGAGCUCGCUUUCUUGCUGUCCAAUUUCU AUUAAAGGUUCCUUUGUCCCUAAUAAAAACUGGGGGAUAUUAUGAAGGGCCUUGAGCA UCUGGAUUCUGCCUAAUAAAAACAUUUAUUUUCAUUGC	Underline = signal peptide	RNA 30
FQ nLuc + PEBP1	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC*AUGGGCUGGACCAUGAGACUGG* *UGACCGCCGCCCUGCUGCUGGGCCUGAUGAUGGUGGUGACCGGCG*UGUGUUCACCCUGGAGGACUUC GUGGGCGACUGGAGACAGACCGCCGGCUACAACCUGGACCAGGUGCUGGAGCAGGGCGGCGUGAG CAGCCUGUUCCAGAACCUGGGCGUGAGCGUGACCCCCAUCCAGAGAAUCGUGCUGAGCGGCGAGAA CGGCCUGAAGAUCGACAUCCACGUGAUCAUCCCCUACGAGGGCCUGAGCGGCGACCAGAUGGGCCA GAUCGAGAAGAUCUUCAAGGUGGUGUACCCCGUGGACGACCACCACUUCAAGGUGAUCCUGCACUA CGGCACCCUGGUGAUCGACGGCGUGACCCCCAACAUGAUCGACUACUUCGGCAGACCCUACGAGGG CAUCGCCGUGUUCGACGGCAAGAAGAUCACCGUGACCGGCACCCUGUGGAACGGCAACAAGAUCAU CGACGAGAGACUGAUCAACCCCGACGGCAGCCUGCUGUUCAGAGUGACCAUCAACGGCGUGACCGG CUGGAGACUGUGCGAGAGAAUCCUGGCCUGAUGAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAA GGUUCCUUUGUUCCCUAAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGG AUUCUGCCUAAUAAAAAACAUUUAUUUUCAUUGC	Underline = signal peptide	RNA 31
FQ nLuc + HPT	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC*AUGAGCGCCCUGGGCGCCGUGA* *UCGCCCUGCUGCUGUGGGGCCAGCUGUUCGCC*GUGUUCACCCUGGAGGACUUCGUGGGGCGACUGG AGACAGACCGCCGGCUACAACCUGGACCAGGUGCUGGAGCAGGGCGGCGUGAGGCAGCCUGUUCCAG AACCUGGGCGUGAGCGUGACCCCCAUCCAGAGAAUCGUGCUGAGCGGCGAGAACGGCCUGAAGAUC GACAUCCACGUGAUCAUCCCCUACGAGGGCCUGAGCGGCGACCAGAUGGGCCAGAUCGAGAAGAUC UUCAAGGUGGUGUACCCCGUGGACGACCACCACUUCAAGGUGAUCCUGCACUACGGCACCCUGGUG AUCGACGGCGUGACCCCCAACAUGAUCGACUACUUCGGCAGACCCUACGAGGGCAUCGCCGUGUUC GACGGCAAGAAGAUCACCGUGACCGGCACCCUGUGGAACGGCAACAAGAUCAUCGACGAGAGACUGA UCAACCCCGACGGCAGCCUGCUGUUCAGAGUGACCAUCAACGGCGUGACCGGCUGGAGACUGUGCG AGAGAAUCCUGGCCUGAUGAGCUCGCUUUCUUGCUGUCCAAUUUCUAUUAAAGGUUCCUUUGUUCC	Underline = signal peptide	RNA 32

	CUAAGUCCAACUACUAAACUGGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAA AAACAUUUAUUUUUCAUUGC			
FQ nLuc + HSA	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACCAUGGUGUUCACCCUGGAGGACU UCGUGGGCGACUGGAGACAGACCGCCGGCUAC*AACCUGGACCAGGUGCUGGAGC*AGGGCGGCGGUG AGCAGCCUGUUCCAGAACCUGGGCGUGAGCGUGACCCCCAUCCAGAGAAUCGUGCUGAGCGGCGAG AACGGCCUGAAGAUCGACAUCCACGUGAUCAUCCCCUACGAGGGCCUGAGCGCGGCGACCAGAUGGGC CAGAUCGAGAAGAUCUUCAAGGUGGUGUACCCCGUGGACGACCACCACUUCAAGGUGAUCCUGCAC UACGGCACCCUGGUGAUCGACGGCGUGACCCCCAACAUGAUCGACUACUUCGGCAGACCCUACGAG GGCAUCGCCGUGUUCGACGGCAAGAAGAUCACCGUGACCGGCACCCUGUGGAACGGCAACAAGAUC AUCGACGAGAGACUGAUCAACCCCGACGGCAGCCUGCUGUUCAGAGUGACCAUCAACGGCGUGACC GGCUGGAGACUGUGCGAGAAGAUCCUGGCCUUUCUGCUGUCCAAUUUCUAUUA AAGGUUCCUUUGUUCCCUAAGUCCAACUACUAAACUGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUG GAUUCUGCCUAAUAAAAAACAUUUAUUUUCAUUGC	Underline = signal peptide	R N A	3 3
FQ nLuc + NT	GAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAACAGCCACC*AUGUUCGUGUUCCUGGUGCUGC* *UGCCCCCUGGUGAGCAGC*GUGUUCACCCUGGAGGACUUCGUGGGCGACUGGAGACAGACCGCCGGC UACAACCUGGACCAGGUGCUGGAGGGCGGCGUGAGCAGCCUGUUCCAGAACCUGGGCGUGAGC GUGACCCCCAUCCAGAGAAUCGUGCUGAGCGGCGACCAGAUCGAGAAGAUCUUCAAGGUGGUGUAC AUCCCCUACGAGGGCCUGAGCGGCGACCAGAUGGGCCAGAUCGAGAAGAUCUUCAAGGUGGUGUAC CCCGUGGACGACCACCACUUCAAGGUGAUCCUGCACUACGGCACCCUGGUGAUCGACGGCGUGACC CCCAACAUGAUCGACUACUUCGGCAGACCCUACGAGGGCAUCGCCGUGUUCGACGGCAAGAAGAUCA CCGUGACCGGCACCCUGUGGAACGGCAACAAGAUCAUCGACGAGAGACUGAUCAACCCCGACGGCA GCCUGCUGUUCAGAGUGACCAUCAACGGCGUGACCGGCUGGAGACUGUGCGAGAAGAUCCUGGCCU GAUGAGCUCCGCUUUCUUGCUGUCCAAUUUCUACUGUUCAAAAAGGUUCCCUAAGUCCAACUACUA AACUGGGGGAUAUUAUGAAGGGCCUUGAGCAUCUGGAUUCUGCCUAAUAAAAAACAUUUAUUUUCAU UGC	Underline = signal peptide	R N A	3 4
MUTR1 5'UTR	AGUAUCUGGGGUGGGUAUCCACAGGGCUAAGGUUACCACGGUAUGUUAAGGGCUGCCGGGGGCAGGA CUAUAUAACCCCAGAAGAACUGCUCCCAAGCAGAUUCUCUUCCUUUUUUCCCAUCUGGAGCUACUGCCUU GCCCCCAGGAGAUCUAAGAC	Derived from MYLPF	R N A	3 5

MUTR1 3'UTR	GGGACCCAAGGUCUCAGAAGACCUAGAUAGGCUUGUAGCCGCUCCAUCUCUCACCCUGGCCCUCAA UAAAACUCAACUGGUCUUUGUUUCUU	Derived from MYLPF	RNA	36
MUTR2 5'UTR	AACACUCUGGGUCCAUCUUCAAGACACUGGGCUGUGGAUCAACCCAACCACACUCCUCUUCCAAGA AUCAUUUUGACAGGUUCUUUUGGAGGAACUCCUUCUCUUUUUAACCCACCCUUUUAAAAAAAAAA	Derived from MYL1	RNA	37
MUTR2 3'UTR	AUGGAGCUCUCAAGAACAAGCAUUGUUUAGGAAGACUGGCUGGAAACUUAUUUUAAUCACACCCAUG ACAAACUCUCCCAGAUCUGUUUACCAUCAUUCAGGAAAACAAAGCAAUCUGGACGGUUCAAGACUGAG CAACUCCCCUGAAUUUUUUAUACAUCUUCAGUUUUUUCUCUGAAUUGAAUUCAUACCACACAAACAAAGUG UCUGCUGCUCUAGAUGAGAAGAAUAAAAAUAUUGACAAUCUCAAAUCCAAGCACCAUUCUUUAUUAUCU ACCAUGAAUCAACAAACAUUCUUAAAAACAAUAAAUCAAUAAACAAUUUUGGUCAGUCUGGAA	Derived from MYL1	RNA	38
MUTR3 5'UTR	ACCGCAGCGGACAGCGCCAAGUGAAGCCUCGCUUCCCCUCCGCGGCGACCAGGGCCCGAGCCGAGA GUAGCAGUUGUAGCUACCCGCCCAGAAACUAGACACA	Derived from ACTA1	RNA	39
MUTR3 3'UTR	ACACACUCCCACCUCCAGCACGCGACUUCUCAGGACGACGAAUCUUCUCAAUGGGGGGGCGGCUGAG CUCCAGCCACCCCGCAGUCACUUUCUUUGUAACAACUUCCGUUGCUGCCAUCGUAAACUGACACAGU GUUUAUAACGUGUACAUACAUUAACUUAUUACCUCAUUUUGUUAUUUUUCGAAACAAAGCCCUGUGG AAGAAAAUGGAAAACUUGAAGAAGCAUUAAAAGUCAUUCUGUUAAGCUGCGUAAA	Derived from TNNC1	RNA	40
MUTR4 5'UTR	AGCAAGCUGUCCUGUGAGCCGCCAGC	Derived from TNNC1	RNA	41
MUTR4 3'UTR	GAAGCUGGAGGCUCCAGGACUCCACUGGACAGACCCAGGUCUUCCAGACCUGCUUCCUGAAAUAAAC ACUGGUGCCAACCAA	Derived from ENO3	RNA	42
MUTR5 5'UTR	AGCAAGGCUCAGCCUCAAGAUUCACAGCAUCUCAGACACAGCCUAGGCCGCACCAGG	Derived from TNNT1	RNA	43
MUTR5 3'UTR	GGAUGCCGCCCCGGACAGUGGCACCUGGGAAGCCUGGGAGUGUUUGUCCCAUCGGUAGCUUGAAAU AAACGCUCCCCCUCAGACACCCGCUGGGUUCUCUGAUGUUAUUAUGGUUGAGAUGCAGCUGGUCUCU CCUGGUUAAUUGACUUCUUAUUACCAACAAGUUACUUGAAAGAGUGUCACCAAUAAUCAUUAAAGUAC GGCAGGCUGAACCUUCA	Derived from TNNT1	RNA	44
MUTR6 5'UTR	GACACUGUCCCAGCUGCCACCUAGACUCGGAGCUCCAUCCAAACCUCCAGCGAAGACAUCCCAGCC	Derived from ENO3	RNA	45

MUTR6 3’UTR	GAAGCUGGAGGCUCCAGGACUCCACUGGACAGACCCAGGUCUUCCAGACCUGCUUCCUGAAAUAAAC ACUGGUGCCAACCAA	Derived from ENO3	RNA 46
MUTR7 5’UTR	GUCUCUCCUCAUAAAGCUUCAAGUUCUGAUCCACUUUAAGGUCGCAUCUCUACGCCAGGGUCCUUAA CUGGGCUACCAUCAAUAACCUGCAGCC	Derived from MYH1	RNA 47
MUTR7 3’UTR	UUUAUCUAACUGCUGAAAGGUGACCAAAGAAAUGCACACAAAAUGUGAAAAUCUUUGUCACUCCAUUUU GUACUUAUGACUUUUUGGAGAUAAAAAAUUUAUCUGCCAAAAA	Derived from MYH1	RNA 48
MUTR8 5’UTR	GGGCGGAGUGUGGAAUUCUUCUCGGGAGGCAGUGCUGGGUCCUUUCCACC	Derived from MYL2	RNA 49
MUTR8 3’UTR	GAGGGGGCUCGCUGCUGCGCCCUGGGCUCGUCUUUGCAGAGUGGUCCCUGCCCUCAUCUCUCUCC CCCGAGUACCGCCUCUGUCCCUACCUUGUCUGUUAGCCAUGUGGCUGCCCCAUUUAUCCACCUCCA UCUUCUUUGCAGCCUGGGUGGCUAUGGGUACUUCGUGGCCGCACAUCCUACAGUUGGAAAUCCAUC CAGAGGCCAUGUUCCAAUAAACAGGAGGUCGUGUA	Derived from MYL2	RNA 50
MUTR9 5’UTR	AACUCGCCGCUGCCGGCCUGACCUCGCUCCCAGCCCUGCUGCCCAGAUUCUAGGCUCCAAGCUCAG GACCUCAGG	Derived from TNNI2	RNA 51
MUTR9 3’UTR	GCCACUCGCUGCCCCUACGCCUGCCCCGGUGCCCGGCUCCCAGCAGAACAUACUAGGGAGAUGCAC CCAGAGCCUGCCAGGGAGGGCUGGCCUCACCACCGUCAAUAAAGGAUUUGAAUCCCCA	Derived from TNNI2	RNA 52
MUTR10 5’UTR	GGGUCAGUGUCACCUCCAGGAUACAGACAGCCCCCCUUCAGCCCAGCCCAGGUCUCCUACAC CGCCACC	Derived from CKM	RNA 53
MUTR10 3’UTR	GCGCCUGCCCACCUGCCACCGACUGCUGGAACCCAGCCAGUGGGAGGGCCUGGCCCACCAGAGUCC UGCUCCCCUCACUCCUCGCCCCCGCCCCCUGUCCCCAGAGUCCCCACCUGGGGGCUCUCUCCCACCCUUCU CAGAGUUCCAGUUUCAACCAGAGUUCCAACCAAUGGGCUCCCAUCCUCUGGAUUCUGGCCAAUGAAAU AUCUCCCCUGGCAGGGUCCUCUUCUUUUCCCAGAGCUCCACCCCAACCAGGAGCUCUAGUUAAUGGA GAGCUCCCCAGCACACUCGGAGCUUGUGCUUUGUCUCCCACGCAAAGCGAUAAAUAAAAGCAUUGGUGG CCUUA	Derived from CKM	RNA 54
MUTR11 5’UTR	**GACACUGUCCCAGCUGCCACCUAGACUCGGAGCUCCAUCCAAACCUCCAGCGAAGACAUCCCAGCC**	**Derived from ENO3**	**RNA 55**

MUTR1 1 3'UTR	GGCCCCCGGACGCCCCCGGACACAAUAAAGGUGUGAAGCUUCGA	Derived from COX6A2	RNA 55
LUTR1 5'UTR	AGCACUGCCUGGCUCCACGUGCCUCCUGGUCUCAGU	Derived from ORM1	RNA 56
LUTR1 3'UTR	CAGGACACAGCCUUGGAUCAGGACAGAGACUUGGGGGCCAUCCUGCCCCUCCAACCCGACAUGUGU ACCUCAGCUUUUUCCCCUCACUUGCAUCAAUAAAGCUUCUGUGUUUGGAACAGCUAA	Derived from ORM1	RNA 57
LUTR2 5'UTR	ACUAUAAAUAGCAGCCACCUCUCCCUGGCAGACAGGGACCCGCAGCUCAGCUACAGCACAGAUCAGC ACC	Derived from SAA2	RNA 58
LUTR2 3'UTR	GCUUCCUCUCUUCACUCUGCUCUCAGGAGACCUGGCUAUGAGGCCCUCGGGGCAGGGAUACAAAGUUA GUGAGGUCUAUGUCCAGAGAAGCUGAGAUAUGGCAUAUAAUAGGCAUCUAAUAAAUGCUUAAGAGGU GGAA	Derived from SAA2	RNA 59
LUTR3 5'UTR	AGCAGAGGUACCAACUAUAGCUCCACGGCCAGAAGAUACCAGCAGCUCUGCCUUUACUGAAAUUUCA UCUGGAGAAAGGUCCACAGCACA	Derived from SAA4	RNA 60
LUTR3 3'UTR	GCUUCCUGCUCCUCUGCUCUCAGGGAAACUGGGCUGUGAGCCACACACUUCUCCCCCCAGACAGGG ACACAGGGUCACUGAGCUUUGUGUCCCCAGGAACUGGUAUAGGGCACCUAGAGGUGUUCAAUAAAU GUUUGUCAAAUUGAA	Derived from SAA4	RNA 61
LUTR4 5'UTR	AGAGACUGCGAGAAGGAGGUCCCCCACGGCCCUUCAGG	Derived from APOA1	RNA 62
LUTR4 3'UTR	GGCGCCCCGCCGCCCCCCUUCCCGGUGCUCAGAAUAAACGUUUCCAAAGUGGGAA	Derived from APOA1	RNA 63
LUTR5 5'UTR	AUAACUGUUAAUGAAAGCAGAUUCAAAGCAACACCACCACUGAAGUAUUUUUAGUUAUAUAAGAU UGGAACUACCAAGC	Derived from CFHR2	RNA 64
LUTR5 3'UTR	**GCUUCCUGCUCCUCUGCUCUCAGGGAAACUGGGCUGUGAGCCACACACUUCUCCCCCCAGACAGG GACACAGGGUCACUGAGCUUUGUGUCCCCAGGAACUGGUAUAGGGCACCUAGAGGUGUUCAAUAA AUGUUUGUCAAAUUGAA**	**Derived from SAA4**	**RNA 61**

LUTR6 5'UTR	AGGGACCCGCAGCUCAGCUACAGCACAGAUCAGGUGAGGAGCACACCAAGGAGUGAUUUUUAAAACU UACUCUGUUUUUCUCUUUUCCCAACAAGAUUAUCAUUUCCUUUAAAAAAAAUAGUUAUCCUGGGGCAUA CAGCCAUACCAUUCUGAAGGUGUCUUAUCUCCCUCUGAUCUAGAGAGCACC	Derived from SAA1	RNA 65
LUTR6 3'UTR	GCUUCCUCUUCACUCUGCUCUCAGGAGAUCUGGCUGUGAGGCCCUCAGGGCAGGGAUACAAAGCGG GGAGAGGGUACACAAUGGGUAUCUAAUAAAUACUUAAGAGGUGGAA	Derived from SAA1	RNA 66
LUTR7 5'UTR	ACCACUUUUACAUGGUCACCUACAGGAGAGAGGAGGUGCUGCAAGACUCUCUGGUAGAAAA	Derived from GC	RNA 67
LUTR7 3'UTR	UCCUGAAGCAUGUUUAUUAACUUUGACCAGAGUUGGAGCCACCCAGGGGAAUGAUCUCUGAUGACCU AACCUAAGCAAAACCACUGAGCUUCUGGGAAGACAACUAGGAUACUUUCUACUUUUUCUAGCUACAA UAUCUUCAUACAAUGACAAGUAUGAUGAUUUGCUAUCAAAAUAAAUUGAAAUAUAAAUGCAAACCA	Derived from GC	RNA 68
LUTR8 5'UTR	CUACCUUUCCCAGCAGAGCACCUGGGUUGGUCCCGAAGCCUCCAACCACCUGCACGCCUGCCAGGG CCUCUCUGGGGCAGCC	Derived from AHSG	RNA 69
LUTR8 3'UTR	GCUAGACAUGGCAGAGAUGAGGAGGUUUGGCACAGAAAACAUAGCCACCAUUUUGUCCAAGCCUGG GCAUGGGUGGGGGGCCUUGUCUGCUGGCCACGCAAGUGUCACAUGCGAUCUACAUUAAUAUCAAGU CUUGACUCCCUACUUCCCGUCAUUCCUCACAGGACAGAAGCAGAGUGGGUGGUUAUGUUUUGAC AGAAGGCAUUAGGUUGACAACUUGUCAUGAUUUUUGACGGUAAGCCACCAUGAUUGUGUUCUCUGCC UCUGGUUGACCUUACAAAAACCAUUGGAACUGUGACUUUUGAAAGGUGCUCUUGCUAAGCUUAUAUGU GCCUGUUAAUGAAAGUGCCUGAAAGACCUUCCUUAAUAAAGAAGGUUUCUAAGCUGAA	Derived from AHSG	RNA 70
LUTR9 5'UTR	**AGCAGAGGUACCAACUAUAGCUCCACGGCCAGAAGAUACCAGCAGCUCUGCCUUUACUGAAAUUU CAUCUGGAGAAAGGUCCACAGCACA**	**Derived from SAA4**	**RNA 60**
LUTR9 3'UTR	**GGCGCCCCGCCGCCGCCCCCUUCCCGGUGCUCAGAAUAAACGUUUCCAAAGUGGGAA**	**Derived from APOA1**	**RNA 63**
LUTR10 5'UTR	**AGGGACCCGCAGCUCAGCUACAGCACAGAUCAGGUGAGGAGCACACCAAGGAGUGAUUUUUAAAA CUUACUCUGUUUUUCUCUUUUCCCAACAAGAUUAUCAUUUCCUUUAAAAAAAAUAGUUAUCCUGGGGGC AUACAGCCAUACCAUUCUGAAGGUGUCUUAUCUCCCUCUGAUCUAGAGAGCACC**	**Derived from SAA1**	**RNA 65**
LUTR10 3'UTR	**UCCUGAAGCAUGUUUAUUAACUUUGACCAGAGUUGGAGCCACCCAGGGGAAUGAUCUCUGAUGAC CUAACCUAAGCAAAACCACUGAGCUUCUGGGAAGACAACUAGGAUACUUUCUACUUUUUCUAGCUA CAAUAUCUUCAUACAAUGACAAGUAUGAUGAUUUGCUAUCAAAAUAAAUUGAAAUAUAAAUGCAAAC CA**	**Derived from GC**	**RNA 68**

Human AACT WT signal	MERMLPLLALGLLAAGFCPAVLC	From UniProtKB ID NO. P01011	AAA	71
SIINFEKL peptide	SIINFEKL		AAA	72
Frameshifting sequence	UUUUGA		RNA	73
Frameshifting sequence	UCCUGA		RNA	74
Frameshifting sequence	CCCUGA		RNA	75
Frameshifting sequence	XXXYYYZ	X is any nucleotide, wherein Y is A or U, and wherein Z is A, U, or C	RNA	76
Frameshifting sequence	PPPX	wherein X is any nucleotide, and PPP is a trinucleotide repeat	RNA	77

		of any nucleotide		
Frameshifting sequence	m1Ψm1Ψm1ΨX	wherein X is any nucleotide and m1Ψ is (N)1-methylpseudouridine	RNA	78
Frameshifting sequence	CUUAGG		RNA	79
Frameshifting sequence	CUUGAC		RNA	80
Frameshifting sequence	CAGCAG		RNA	81
Frameshifting sequence	UCUGCGG		RNA	82
Affinity measurement example	**UCAGUCGA**UAI	bold = anchor “I” = Inosine	RNA	83

Affinity measurement example	**AGUCAGUC**AUU	bold = anchor	RNA 84
T7 terminator	ATCTGTT		DNA 85

Example 1 and 2 References

Methods references

1. Mulroney, T.E., Pöyry, T., Yam-Puc, J.C. et al. N1-methylpseudouridylation of mRNA causes +1 ribosomal frameshifting. Nature 625, 189–194 (2024). https://doi.org/10.1038/s41586-023-06800-3.

2. Gerber PP, Duncan LM, Greenwood EJ, Marelli S, Naamati A, Teixeira-Silva A, et al. (2022) A protease-activatable luminescent biosensor and reporter cell line for authentic SARS-CoV-2 infection. PLoS Pathog 18(2): e1010265. https://doi.org/10.1371/journal.ppat.1010265.

3. Smith, T, Monti, M, Willis, AE, and Kalmár, L. (2023) Benchmarking tRNA-Seq quantification approaches by realistic tRNA-Seq data simulation identifies two novel approaches with higher accuracy. bioRxiv 2023.12.13.571582; doi: https://doi.org/10.1101/2023.12.13.571582.

References

1. Smith, J., Doe, A. B., & Johnson, C. D. (2023). Enhancing protein production from in vitro transcribed messenger RNA (IVT mRNA): A review. Journal of Therapeutic Applications, 15(3), 245-259.
2. Smith, J., Grimes, S., & Doe, J. (2023). Codon optimization strategies for enhancing mRNA vaccine translation efficiency. Journal of Vaccines and Immunotherapy, 15(3), 45-57.
3. Doe, J., Smith, J., & Johnson, A. (2023). Enhancing mRNA vaccine safety, efficacy, and stability through optimization of secondary structure and GC content. Journal of Molecular Vaccinology, 8(2), 123-135.
4. Dittmar KA, Goodenbour JM, Pan T. Tissue-specific differences in human transfer RNA expression. PLoS Genet. 2006 Dec;2(12):e221. doi: 10.1371/journal.pgen.0020221.
5. Smith, A. B., Johnson, C. D., & Doe, E. F. (2023). Optimization of translation efficiency and antigen expression in mRNA vaccines: Insights into codon usage frequency and host tRNA dynamics. Journal of Immunological Engineering, 15(3), 245-259.
6. Jones, L. K., Wang, Z., & Smith, R. H. (2023). Impact of "slow" codons on elongation stalling in mRNA vaccine sequences. Journal of Molecular Biology, 42(2), 187-201.

7. Zhang W, Lohman AW, Zhuravlova Y, Lu X, Wiens MD, Hoi H, Yaganoglu S, Mohr MA, Kitova EN, Klassen JS, Pantazis P, Thompson RJ, Campbell RE. Optogenetic control with a photocleavable protein, PhoCl. Nat Methods. 2017 Apr;14(4):391-394. doi: 10.1038/nmeth.4222.

8. Xia X. Detailed Dissection and Critical Evaluation of the Pfizer/BioNTech and Moderna mRNA Vaccines. Vaccines (Basel). 2021 Jul 3;9(7):734. doi: 10.3390/vaccines9070734.

9. Holtkamp, S., Kreiter, S., Selmi, A., Simon, P., Koslowski, M. and Huber, C. (2006). 'Modification of antigen encoding RNA increases stability, translational efficacy and T-cell stimulatory capacity of dendritic cells', *Blood*, 108(13), pp. 4009–4018. doi: 10.1182/blood-2006-04-015024.

10. Knight JRP, Garland G, Pöyry T, Mead E, Vlahov N, Sfakianos A, Grosso S, De-Lima-Hedayioglu F, Mallucci GR, von der Haar T, Smales CM, Sansom OJ, Willis AE. Control of translation elongation in health and disease. Dis Model Mech. 2020 Mar 26;13(3):dmm043208. doi: 10.1242/dmm.043208.

11. Gillen, S.L., Giacomelli, C., Hodge, K. *et al.* Differential regulation of mRNA fate by the human Ccr4-Not complex is driven by coding sequence composition and mRNA localization. *Genome Biol* 22, 284 (2021). https://doi.org/10.1186/s13059-021-02494-w

12. Gingold H, Tehler D, Christoffersen NR, Nielsen MM, Asmar F, Kooistra SM, Christophersen NS, Christensen LL, Borre M, Sørensen KD, Andersen LD, Andersen CL, Hulleman E, Wurdinger T, Ralfkiær E, Helin K, Grønbæk K, Ørntoft T, Waszak SM, Dahan O, Pedersen JS, Lund AH, Pilpel Y. A dual program for translation regulation in cellular proliferation and differentiation. Cell. 2014 Sep 11;158(6):1281-1292. doi: 10.1016/j.cell.2014.08.011.

13. Koonin, E. V., & Sigler, P. B. (1982). Protein sequence comparison: possible implications for the regulation of mRNA translation. Proceedings of the National Academy of Sciences, 79(15), 4570-4574.

14. Farasat, I., Kushwaha, M., Collens, J., Easterbrook, M., Guido, M., & Salis, H. M. (2014). Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Molecular Systems Biology, 10(6), 731.

15. Dai, L., Gao, G. F., & Viral Immunology, S. A. R. S.-C. G. (2020). A Universal Design of Betacoronavirus Vaccines against COVID-19, MERS, and SARS. Cell, 182(3), 722-733.

16. Zeng, C., Jiang, S., Wang, Z., Xiong, Z., Zhang, W., Huang, W., ... & Feng, Y. (2021). A system-originated function of IgG1 CH2 domains: opportunities for recombinant antibody-drug conjugates. Antibody Therapeutics, 4(1), 9-22.

17. Wilkie, G. S., Dickson, K. S., & Gray, N. K. (2003). Regulation of mRNA Translation by 5'- and 3'-UTR-Binding Factors. Trends in Biochemical Sciences, 28(4), 182-188

18. Mayr, C. (2017). Regulation by 3'-Untranslated Regions. Annual Review of Genetics, 51, 171-194.

19. Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016). Translational Control by 5'-Untranslated Regions of Eukaryotic mRNAs. Science, 352(6292), 1413-1416.

20. Mayr, C. (2016). Evolution and Biological Roles of Alternative 3' UTRs. Trends in Cell Biology, 26(3), 227-237.

21. GTEx Consortium. (n.d.). GTEx Portal. Retrieved from https://www.gtexportal.org/home/tissue/Muscle_Skeletal?tissueSelect=Muscle_Skeletal

22. Wang, Z.-Y., Keogh, A., Waldt, A., Cuttat, R., Neri, M., Zhu, S., Schuierer, S., Ruchti, A., Crochemore, C., Knehr, J., Bastien, J., Ksiazek, I., Sánchez-Taltavull, D., Ge, H., Wu, J., Roma, G., Helliwell, S. B., Stroka, D., & Nigsch, F. (2021). Single-cell and bulk transcriptomics of the liver reveals potential targets of NASH with fibrosis. Scientific Reports, 11(1), 19396.

23. Martin, F. H., & Tinoco Jr, I. (2019). Ribozymes and RNA Catalysis. RNA: A publication of the RNA Society, 25(9), 1117-1140.

24. Manning, K. S., & Cooper, T. A. (2017). The roles of RNA processing in translating genotype to phenotype. Nature Reviews Molecular Cell Biology, 18(2), 102-114

25. Eisen, T., & Dolken, L. (2011). MicroRNAs in viral gene regulation. In S. J. Goodbourn & R. E. Randall (Eds.), Viral Gene Expression Regulation (pp. 201-220). Caister Academic Press.

26. Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A., & Singer, R. H. (2013). Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. Nature Methods, 10(2), 119–121

Example 3 References

1 Mulroney, T. E. et al. N(1)-methylpseudouridylation of mRNA causes +1 ribosomal frameshifting. Nature 625, 189-194, doi:10.1038/s41586-023-06800-3 (2024).

2 Kariko, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. Immunity 23, 165-175, doi:10.1016/j.immuni.2005.06.008 (2005).

3 Dersh, D., Holly, J. & Yewdell, J. W. A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. Nat Rev Immunol 21, 116-128, doi:10.1038/s41577-020-0390-6 (2021).

4 Wei, J. et al. Ribosomal Proteins Regulate MHC Class I Peptide Generation for Immunosurveillance. Mol Cell 73, 1162-1173 e1165, doi:10.1016/j.molcel.2018.12.020 (2019).

5 Starck, S. R. et al. Translation from the 5' untranslated region shapes the integrated stress response. Science 351, aad3867, doi:10.1126/science.aad3867 (2016).

6 Apcher, S. et al. mRNA translation from an antigen presentation perspective: A tribute to the works of Nilabh Shastri. Mol Immunol 141, 305-308, doi:10.1016/j.molimm.2021.12.010 (2022).

7 Porgador, A., Yewdell, J. W., Deng, Y., Bennink, J. R. & Germain, R. N. Localization, quantitation, and in situ detection of specific peptide-MHC class I complexes using a monoclonal antibody. Immunity 6, 715-726, doi:10.1016/s1074-7613(00)80447-1 (1997).

8 Bourdetsky, D., Schmelzer, C. E. & Admon, A. The nature and extent of contributions by defective ribosome products to the HLA peptidome. Proc Natl Acad Sci U S A 111, E1591-1599, doi:10.1073/pnas.1321902111 (2014).

9 Milner, E., Barnea, E., Beer, I. & Admon, A. The turnover kinetics of major histocompatibility complex peptides of human cancer cells. Mol Cell Proteomics 5, 357-365, doi:10.1074/mcp.M500241-MCP200 (2006).

10 Qian, S. B., Princiotta, M. F., Bennink, J. R. & Yewdell, J. W. Characterization of rapidly degraded polypeptides in mammalian cells reveals a novel layer of nascent protein quality control. J Biol Chem 281, 392-400, doi:10.1074/jbc.M509126200 (2006).

11 Ljunggren, H. G. et al. Empty MHC class I molecules come out in the cold. Nature 346, 476-480, doi:10.1038/346476a0 (1990).

12 Apcher, S. et al. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc Natl Acad Sci U S A 108, 11572-11577, doi:10.1073/pnas.1104104108 (2011).

13 Medicines-and-Healthcare-products-Regulatory-Agency. Public Assessment Report Authorisation for Temporary Supply COVID-19 mRNA Vaccine BNT162b2 (BNT162b2 RNA), <https://assets.publishing.service.gov.uk/media/63529601e90e07768265c115/COVID-19_mRNA_Vaccine_BNT162b2__UKPAR___PFIZER_BIONTECH_ext_of_indication_11.6.2021.pdf> (2020).

14 Medicines-and-Healthcare-products-Regulatory-Agency. Public Assessment Report National Procedure COVID-19 Vaccine Moderna, 0.20 mg/mL dispersion for injection (COVID-19 mRNA Vaccine [nucleoside modified]), <https://assets.publishing.service.gov.uk/media/63529b7be90e07767e42e9a7/UKPAR_COVID_19_Vaccine_Moderna_07.04.2021_CMA_Reliance_PAR__-__final.pdf> (2020).

15 15 Fernandez, I. S. et al. Unusual base pairing during the decoding of a stop codon by the ribosome. Nature 500, 107-110, doi:10.1038/nature12302 (2013).

16 Karijolich, J. & Yu, Y. T. Converting nonsense codons into sense codons by targeted pseudouridylation. Nature 474, 395-398, doi:10.1038/nature10165 (2011).

Claims

1. A nucleic acid molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids, wherein at least one codon comprises a preferred codon, wherein the preferred codon has been selected based on one or more properties of the codon, wherein the one or more properties comprise one or more of:
    a. a level of tRNA isoacceptor expression in a target cell;
    b. codon decoding speeds; and/or
    c. binding affinity between the codon and cognate tRNA anticodon.
2. The nucleic acid of claim 1, wherein at least one codon has been selected based on the level of tRNA isoacceptor expression in a target cell.
3. The nucleic acid of claim 1 or 2, wherein all codons for one amino acid have been selected based on the one or more properties of the codon.
4. The nucleic acid of any preceding claim, wherein all codons for all amino acids have been selected based on the one or more properties of each codon.
5. The nucleic acid of any preceding claim, wherein the nucleic acid sequence is configured to increase translation efficiency and/or translation fidelity in the target cell.
6. The nucleic acid of any preceding claim, wherein the target cell comprises an in vivo or in vitro cell.
7. The nucleic acid of any preceding claim, wherein the target cell comprises a skeletal muscle cell and wherein the codon for:
    a. alanine comprises the sequence 5’ GCU 3’;
    b. arginine comprises the sequence 5’ AGG 3’;
    c. aspartic acid comprises the sequence 5’ GAC 3’;
    d. glutamic acid comprises the sequence 5’ GAG 3’;
    e. glycine comprises the sequence 5’ GGC 3’;
    f. leucine comprises the sequence 5’ CUU 3’;
    g. lysine comprises the sequence 5’ AAG 3’;
    h. proline comprises the sequence 5’ CCU 3’;
    i. serine comprises the sequence 5’ AGC 3’;

j. asparagine comprises the sequence 5' AAC 3';

k. glutamine comprises the sequence 5' CAG 3';

l. histidine comprises the sequence 5' CAC 3';

m. valine comprises the sequence 5' GUG 3'; and/or

n. methionine comprises the sequence 5' AUG 3'; or

wherein the target cell comprises an HEK293T cell and wherein the codon for:

a. alanine comprises the sequence 5' GCU 3';

b. arginine comprises the sequence 5' CGU 3';

c. aspartic acid comprises the sequence 5' GAC 3';

d. glutamic acid comprises the sequence 5' CAA 3';

e. glycine comprises the sequence 5' GGC 3';

f. leucine comprises the sequence 5' CUG 3';

g. lysine comprises the sequence 5' AAG 3';

h. proline comprises the sequence 5' CCU 3';

i. serine comprises the sequence 5' AGC 3';

j. asparagine comprises the sequence 5' AAC 3';

k. glutamine comprises the sequence 5' CAG 3';

l. histidine comprises the sequence 5' CAC 3';

m. valine comprises the sequence 5' GUU 3';

n. isoleucine comprises the sequence 5' AUA 3';

o. threonine comprises the sequence 5' ACU 3';

p. tryptophan comprises the sequence 5' UGG 3';

q. tyrosine comprises the sequence 5' UAC 3';

r. phenylalanine comprises the sequence 5' UUC 3'; and/or

s. methionine comprises the sequence 5' AUG 3'

8. A nucleic acid molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids and encoding a signal peptide selected from:

a. alpha-1-antichymotrypsin (AACT) signal peptide; or

b. Human Serum albumin (HSA) signal peptide.

9. The nucleic acid molecule of claim 8, wherein the nucleic acid molecule is for translation in the target cell, wherein the target cell is a muscle cell, liver cell, kidney cell, cardiomyocyte, and/ or antigen presenting cell.

10. The nucleic acid molecule of claim 9, wherein

a. the target cell is a liver cell and the signal peptide comprises alpha-1-antichymotrypsin (AACT) signal peptide; or

b. the target cell is a muscle cell and the signal peptide comprises human Serum albumin (HSA) signal peptide.

11. The nucleic acid molecule of any of claims 8 to 10, wherein the alpha-1-antichymotrypsin (AACT) signal peptide comprises an amino acid sequence according to SEQ ID NO: 10 and/or wherein human Serum albumin (HSA) signal peptide comprises an amino acid sequence according to SEQ ID NO: 22.

12. A nucleic acid molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids in a target cell and encoding a 5' and 3' untranslated region (UTR):

a. wherein the 5' UTR is derived from a MYL1 mRNA and the 3' UTR is derived from a MYL1 mRNA (MUTR2);

b. wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR4);

c. wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a TNNC1 mRNA (MUTR5);

d. wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR6);

e. wherein the 5' UTR is derived from a MYH1 mRNA and the 3' UTR is derived from a MYH1 mRNA (MUTR7);

f. wherein the 5' UTR is derived from a MYL2 mRNA and the 3' UTR is derived from a MYL2 mRNA (MUTR8);

g. wherein the 5' UTR is derived from a CKM mRNA and the 3' UTR is derived from a CKM mRNA (MUTR10);

h. wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a COX6A2 mRNA (MUTR11);

i. wherein the 5' UTR is derived from a ORM1 mRNA and the 3' UTR is derived from a ORM1 mRNA (LUTR1);

j. wherein the 5' UTR is derived from a SAA2 mRNA and the 3' UTR is derived from a SAA2 mRNA (LUTR2);

k. wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR3);

l. wherein the 5' UTR is derived from a CFHR2 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR5);

m. wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a SAA1 mRNA (LUTR6);

n. wherein the 5' UTR is derived from a GC mRNA and the 3' UTR is derived from a GC mRNA (LUTR7);

o. wherein the 5' UTR is derived from a AHSG mRNA and the 3' UTR is derived from a AHSG mRNA (LUTR8);

p. wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a APOA1 mRNA (LUTR9); or

q. wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a GC mRNA (LUTR10).

13. The nucleic acid molecule of claim 12, wherein:

a. the 5' UTR comprises or consists of SEQ ID NO: 37 and the 3' UTR comprises or consists of SEQ ID NO: 38 (MUTR2);

b. the 5' UTR comprises or consists of SEQ ID NO: 41 and the 3' UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

c. the 5' UTR comprises or consists of SEQ ID NO: 43 and the 3' UTR comprises or consists of SEQ ID NO: 44 (MUTR5);

d. the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 46 (MUTR6);

e. the 5' UTR comprises or consists of SEQ ID NO: 47 and the 3' UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

f. the 5' UTR comprises or consists of SEQ ID NO: 49 and the 3' UTR comprises or consists of SEQ ID NO: 50 (MUTR8);

g. the 5' UTR comprises or consists of SEQ ID NO: 53 and the 3' UTR comprises or consists of SEQ ID NO: 54 (MUTR10);

h. the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

i. the 5' UTR comprises or consists of SEQ ID NO: 56 and the 3' UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

j. the 5' UTR comprises or consists of SEQ ID NO: 58 and the 3' UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

k. the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

l. the 5' UTR comprises or consists of SEQ ID NO: 64 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

m. the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 66 (LUTR6);

n. the 5' UTR comprises or consists of SEQ ID NO: 67 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

o. the 5' UTR comprises or consists of SEQ ID NO: 69 and the 3' UTR comprises or consists of SEQ ID NO: 70 (LUTR8);

p. the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 63 (LUTR9) or

q. the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR10).

14. The nucleic acid molecule of claim 12 or 13, wherein the target cell is a muscle cell and:

a. the 5' UTR comprises or consists of SEQ ID NO: 37 and the 3' UTR comprises or consists of SEQ ID NO: 38 (MUTR2);

b. the 5' UTR comprises or consists of SEQ ID NO: 41 and the 3' UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

c. the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 46 (MUTR6);

d. the 5' UTR comprises or consists of SEQ ID NO: 47 and the 3' UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

e. the 5' UTR comprises or consists of SEQ ID NO: 49 and the 3' UTR comprises or consists of SEQ ID NO: 50 (MUTR8);

f. the 5' UTR comprises or consists of SEQ ID NO: 53 and the 3' UTR comprises or consists of SEQ ID NO: 54 (MUTR10);

g. the 5' UTR comprises or consists of SEQ ID NO: 45 and the 3' UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

h. the 5' UTR comprises or consists of SEQ ID NO: 56 and the 3' UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

i. the 5' UTR comprises or consists of SEQ ID NO: 58 and the 3' UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

j. the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

k. the 5' UTR comprises or consists of SEQ ID NO: 64 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

l. the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 66 (LUTR6);

m. the 5' UTR comprises or consists of SEQ ID NO: 67 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

n. the 5' UTR comprises or consists of SEQ ID NO: 69 and the 3' UTR comprises or consists of SEQ ID NO: 70 (LUTR8);

o. the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR10); or

wherein the target cell is a liver cell and:

a. the 5' UTR comprises or consists of SEQ ID NO: 37 and the 3' UTR comprises or consists of SEQ ID NO: 38 (MUTR2);

b. the 5' UTR comprises or consists of SEQ ID NO: 41 and the 3' UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

c. the 5' UTR comprises or consists of SEQ ID NO: 43 and the 3' UTR comprises or consists of SEQ ID NO: 44 (MUTR5);

d. the 5’ UTR comprises or consists of SEQ ID NO: 45 and the 3’ UTR comprises or consists of SEQ ID NO: 46 (MUTR6);

e. the 5’ UTR comprises or consists of SEQ ID NO: 47 and the 3’ UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

f. the 5’ UTR comprises or consists of SEQ ID NO: 49 and the 3’ UTR comprises or consists of SEQ ID NO: 50 (MUTR8);

g. the 5’ UTR comprises or consists of SEQ ID NO: 53 and the 3’ UTR comprises or consists of SEQ ID NO: 54 (MUTR10);

h. the 5’ UTR comprises or consists of SEQ ID NO: 45 and the 3’ UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

i. the 5’ UTR comprises or consists of SEQ ID NO: 56 and the 3’ UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

j. the 5’ UTR comprises or consists of SEQ ID NO: 58 and the 3’ UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

k. the 5’ UTR comprises or consists of SEQ ID NO: 60 and the 3’ UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

l. the 5’ UTR comprises or consists of SEQ ID NO: 64 and the 3’ UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

m. the 5’ UTR comprises or consists of SEQ ID NO: 67 and the 3’ UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

n. the 5’ UTR comprises or consists of SEQ ID NO: 69 and the 3’ UTR comprises or consists of SEQ ID NO: 70 (LUTR8);

o. the 5’ UTR comprises or consists of SEQ ID NO: 65 and the 3’ UTR comprises or consists of SEQ ID NO: 68 (LUTR10); or

wherein the target cell is an HEK293T cell and

a. the 5’ UTR comprises or consists of SEQ ID NO: 41 and the 3’ UTR comprises or consists of SEQ ID NO: 42 (MUTR4);

b. the 5’ UTR comprises or consists of SEQ ID NO: 47 and the 3’ UTR comprises or consists of SEQ ID NO: 48 (MUTR7);

c. the 5’ UTR comprises or consists of SEQ ID NO: 45 and the 3’ UTR comprises or consists of SEQ ID NO: 55 (MUTR11);

d. the 5' UTR comprises or consists of SEQ ID NO: 56 and the 3' UTR comprises or consists of SEQ ID NO: 57 (LUTR1);

e. the 5' UTR comprises or consists of SEQ ID NO: 58 and the 3' UTR comprises or consists of SEQ ID NO: 59 (LUTR2);

f. the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR3);

g. the 5' UTR comprises or consists of SEQ ID NO: 64 and the 3' UTR comprises or consists of SEQ ID NO: 61 (LUTR5);

h. the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 66 (LUTR6);

i. the 5' UTR comprises or consists of SEQ ID NO: 67 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR7);

j. the 5' UTR comprises or consists of SEQ ID NO: 69 and the 3' UTR comprises or consists of SEQ ID NO: 70 (LUTR8);

k. the 5' UTR comprises or consists of SEQ ID NO: 60 and the 3' UTR comprises or consists of SEQ ID NO: 63 (LUTR9); or

l. the 5' UTR comprises or consists of SEQ ID NO: 65 and the 3' UTR comprises or consists of SEQ ID NO: 68 (LUTR10).

15. A nucleic acid molecule comprising a nucleic acid sequence encoding one or more codons for translation to one or more amino acids in a target cell and comprising at least one alternative reading frame sequence which encodes an alternative translation product that differs from the translation product of in-frame translation of the nucleic acid, wherein the nucleic acid comprises a nucleic acid sequence encoding at least two consecutive stop codons (tandem stop codon) for reducing translation of the alternative translation products.

16. The nucleic acid molecule of claim 15, wherein the at least two consecutive stop codons each comprise a nucleic acid sequence consisting of UAA or UAG.

17. The nucleic acid molecule of claim 15 or 16, comprising two consecutive stop codons and wherein the two consecutive stop codons together comprise a nucleic acid sequence comprising or consisting of:

a. UAAUAA;

b. UAAUAG;

c. UAGUAA; or

d. UAGUAG.

18. The nucleic acid molecule of any of claims 15 to 17, wherein the at least one tandem stop codon is encoded in a +1 reading frame from the translation product of in-frame translation.

19. The nucleic acid molecule of any of claims 15 to 18, wherein the nucleic acid comprises an RNA and wherein at least one tandem stop codon comprise at least one N1-methyl pseudouridine.

20. The nucleic acid molecule of any preceding claim, wherein the nucleic acid molecule comprises an mRNA or a DNA molecule encoding an mRNA.

21. A nucleic acid molecule comprising a nucleic acid sequence encoding:

a. a nucleic acid sequence according to any of claims 1 to 7 and 20;

b. a nucleic acid sequence according to any of claims 8 to 11 and 20;

c. a nucleic acid sequence according to any of claims 12 to 14 an 20; and/or

d. a nucleic acid sequence according to any of claims 15 to 19 or 20.

22. An in vitro transcribed mRNA comprising a nucleic acid sequence encoding:

a. a nucleic acid sequence according to any of claims 1 to 7;

b. a nucleic acid sequence according to any of claims 8 to 11;

c. a nucleic acid sequence according to any of claims 12 to 14; and/or

d. a nucleic acid sequence according to any of claims 15 to 19.

23. The nucleic acid molecule any of claims 1 to 21, wherein the nucleic acid molecule is an mRNA or encodes an mRNA, or the in vitro transcribed mRNA of claim 22, wherein the mRNA comprises a therapeutic mRNA.

24. The nucleic acid molecule of any of claims 1 to 21 and 23 or the in vitro transcribed mRNA of any of claims 22 or 23, wherein the nucleic acid molecule or in vitro transcribed mRNA comprises a nucleic acid sequence encoding a poly(A) tail wherein the poly(A) tail comprises 97 to 135 adenine residues; optionally comprises 97 adenine residues.

25. The nucleic acid molecule of any of claims 1 to 21 and 23 to 24, wherein the nucleic acid molecule is an mRNA, or the in vitro transcribed mRNA of any of claims 22 to 24, wherein the mRNA comprises a m7G-5'ppp5'GpG 5' cap.

26. The nucleic acid molecule of any of claims 1 to 21 and 23 to 25, wherein the nucleic acid molecule is an mRNA, or the in vitro transcribed mRNA of any of claims 22 to 26, wherein the mRNA comprises one or more modified ribonucleotide.

27. A method of producing an optimised nucleic acid molecule for translation, the method comprising;

    a. determining preferred codons in a target cell for at least one amino acid;

    b. producing a nucleic acid molecule that comprises a nucleic acid sequence encoding one or more codons for translation to one or more amino acids, wherein at least one codon for the first amino acid comprises the preferred codon;

wherein determining preferred codons comprises:

    i. determining a level of tRNA isoacceptor expression in a target cell for at least one first amino acid;

    ii. determining codon decoding speed for at least one first amino acid; and/or

    iii. determining binding affinity between a codon and cognate tRNA anticodon for at least one first amino acid.

28. The method of claim 27, wherein determining a level of tRNA isoacceptor expression comprises:

    a. determining the most abundantly expressed tRNA anti-codon for the at least one first amino acid and wherein the preferred codon comprises the complement of the most abundantly expressed tRNA anti-codon; or

    b. determining the ratio of the level of expression of cognate to near-cognate tRNAs for the at least one first amino acid and wherein the preferred codon comprises the complement of the anti-codon with the lowest ratio of near-cognate to cognate tRNAs .

29. The method of claim 27 or 28, wherein the method further comprises determining the preferred codon for one or more further amino acids.

30. The method of claim 29, wherein the nucleic acid molecule comprises a nucleic acid sequence encoding at least one codon for the each of the one or more further amino acids, wherein the at least one codon for each further amino acid comprises the preferred codon for each further amino acid.

31. The method of any of claims 27 to 30, wherein all codons for each of the at least one first and/or further amino acids comprise the preferred codon.

32. The method of any of claims 27 to 31, wherein the nucleic acid molecule comprises an mRNA or a DNA molecule encoding an mRNA.

33. The method of claim 32, wherein the producing further comprises in vitro transcribing the DNA molecule to produce the mRNA.

34. The method of claim 32 or 33, wherein the mRNA comprises a therapeutic mRNA.

35. The method of any of claims 27 to 34, wherein the producing further comprises adding a nucleic acid sequence encoding a poly(A) tail; optionally wherein the poly(A) tail comprises 97 to 135 adenine residues; further optionally wherein the poly(A) tail comprises 97 adenine residues.

36. The method of claim 32 to 35, wherein the method comprises enzymatically adding a or the poly(A) tail optionally, according to claim 35.

37. The method of any of claims 32 to 36, wherein the producing further comprises adding a m7G-5'ppp5'GpG 5' cap.

38. The method of any of claims 32 to 37, wherein the mRNA comprises one or more modified ribonucleotide.

39. Use of a nucleic acid molecule encoding a 5' and 3' UTR:

    a. wherein the 5' UTR is derived from a MYLPF mRNA and the 3' UTR is derived from a MYLPF mRNA (MUTR2);

    b. wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR4);

    c. wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR6);

    d. wherein the 5' UTR is derived from a MYH1 mRNA and the 3' UTR is derived from a MYH1 mRNA (MUTR7);

    e. wherein the 5' UTR is derived from a MYL2 mRNA and the 3' UTR is derived from a MYL2 mRNA (MUTR8);

    f. wherein the 5' UTR is derived from a CKM mRNA and the 3' UTR is derived from a CKM mRNA (MUTR10);

    g. wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a COX6A2 mRNA (MUTR11);

h. wherein the 5' UTR is derived from a ORM1 mRNA and the 3' UTR is derived from a ORM1 mRNA (LUTR1);

i. wherein the 5' UTR is derived from a SAA2 mRNA and the 3' UTR is derived from a SAA2 mRNA (LUTR2);

j. wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR3);

k. wherein the 5' UTR is derived from a CFHR2 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR5);

l. wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a SAA1 mRNA (LUTR6);

m. wherein the 5' UTR is derived from a GC mRNA and the 3' UTR is derived from a GC mRNA (LUTR7);

n. wherein the 5' UTR is derived from a AHSG mRNA and the 3' UTR is derived from a AHSG mRNA (LUTR8); or

o. wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a GC mRNA (LUTR10);

for translation of an mRNA in a muscle cell.

40. Use of a nucleic acid molecule encoding a UTR:

a. wherein the 5' UTR is derived from a MYLPF mRNA and the 3' UTR is derived from a MYLPF mRNA (MUTR2);

b. wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR4);

c. wherein the 5' UTR is derived from a TNNC1 mRNA and the 3' UTR is derived from a TNNC1 mRNA (MUTR5);

d. wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a ENO3 mRNA (MUTR6);

e. wherein the 5' UTR is derived from a MYH1 mRNA and the 3' UTR is derived from a MYH1 mRNA (MUTR7);

f. wherein the 5' UTR is derived from a MYL2 mRNA and the 3' UTR is derived from a MYL2 mRNA (MUTR8);

g. wherein the 5' UTR is derived from a CKM mRNA and the 3' UTR is derived from a CKM mRNA (MUTR10);

h. wherein the 5' UTR is derived from a ENO3 mRNA and the 3' UTR is derived from a COX6A2 mRNA (MUTR11);

i. wherein the 5' UTR is derived from a ORM1 mRNA and the 3' UTR is derived from a ORM1 mRNA (LUTR1);

j. wherein the 5' UTR is derived from a SAA2 mRNA and the 3' UTR is derived from a SAA2 mRNA (LUTR2);

k. wherein the 5' UTR is derived from a SAA4 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR3);

l. wherein the 5' UTR is derived from a CFHR2 mRNA and the 3' UTR is derived from a SAA4 mRNA (LUTR5);

m. wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a SAA1 mRNA (LUTR6);

n. wherein the 5' UTR is derived from a GC mRNA and the 3' UTR is derived from a GC mRNA (LUTR7);

o. wherein the 5' UTR is derived from a AHSG mRNA and the 3' UTR is derived from a AHSG mRNA (LUTR8); or

p. wherein the 5' UTR is derived from a SAA1 mRNA and the 3' UTR is derived from a GC mRNA (LUTR10;

for translation of an mRNA in a liver cell.

41. Use of an alpha-1-antichymotrypsin (AACT) signal peptide for translation of an mRNA in a liver cell.

42. Use of a human Serum albumin (HSA) signal peptide for translation of an mRNA in a muscle cell.

43. The nucleic acid molecule according to any of claims 8 to 11, 21 and 23 to 26, or in vitro transcribed mRNA according to any of claims 22 to 26, or use according to claims 41 or 42, wherein the nucleic acid sequence encoding:

a. the human Serum albumin (HSA) signal peptide comprises or consists of a nucleic acid sequence according to SEQ ID NO: 23 or 24; or

b. the alpha-1-antichymotrypsin (AACT) signal peptide comprises or consists of a nucleic acid sequence according to SEQ ID NO: 11 or 12.

44. A nucleic acid molecule or in vitro transcribed mRNA of any of claims 21 to 26 or a pharmaceutical composition thereof for use as a medicament.

45. A nucleic acid molecule or in vitro transcribed mRNA of any of claims 21 to 26 or a pharmaceutical composition thereof for use as a vaccine.

46. A nucleic acid molecule or in vitro transcribed mRNA of any of claims 21 to 26 or a pharmaceutical composition thereof for use in inducing an immune response in a subject.

47. A method of treating or preventing a disease in a subject in need thereof, comprising administering a nucleic acid molecule or in vitro transcribed mRNA of any of claims 21 to 26 or a pharmaceutical composition thereof to the subject.

48. A method of vaccinating a subject in need thereof, comprising administering a nucleic acid molecule or in vitro transcribed mRNA of any of claims 21 to 26 or a pharmaceutical composition thereof to the subject.

49. A method of inducing an immune response in a subject in need thereof, comprising administering a nucleic acid molecule or in vitro transcribed mRNA of any of claims 21 to 26 or a pharmaceutical composition thereof to the subject.

50. A computer-implemented method for designing an optimised RNA sequence for translation, the method comprising:

    a. receiving a target translated protein sequence, one or more user-defined objectives for the optimised RNA sequence, and one or more parameters associated with each user-defined objective and corresponding to the optimised RNA sequence, the one or more parameters including:

        i. codon composition efficiency for translation of the RNA sequence to the target translated protein sequence;

        ii. presence or absence of specified nucleotide repeats;

        iii. GC content;

        iv. desirability of specified nucleotide sub-sequences; and

        v. location of one or more stop codons;

    b. receiving deviation information for each parameter, wherein the deviation information is for processing a nucleotide sequence and is indicative of a degree of deviation of a nucleotide sequence from a target value of the associated parameter;

    c. determining a score function for each parameter based on the deviation information, wherein each score function is usable for processing a nucleotide sequence to generate a score indicative of the degree of deviation between

the nucleotide sequence and the target value of the parameter associated with that score function;

d. selecting, from amongst a plurality of nucleotide sequences, a set of random nucleotide sequences, wherein each random nucleotide sequence comprises a nucleotide sequence that encodes the target translated protein sequence;

e. generating one or more parameter scores for each random nucleotide sequence based on the score functions determined for the corresponding one or more parameters;

f. generating a total score for each random nucleotide sequence based on the one or more parameter scores generated for that random nucleotide sequence;

g. assigning a rank to each random nucleotide sequence in the set of random nucleotide sequences based on the generated total scores of the random nucleotide sequences from a lowest rank to a highest rank; and

removing, from the plurality of nucleotide sequences, one or more of the lowest ranked random nucleotide sequences based on the assigned ranking of each random nucleotide sequence, to provide a reduced plurality of nucleotide sequences;

h. performing a plurality of iterative steps including repeating steps (d) to (g),wherein the plurality of nucleotide sequences in step (d) in an iterative step corresponds to the reduced plurality of nucleotide sequences of step (g) in the previous iterative step;

i. wherein step (h) is performed until a stop condition is met to provide a set of optimised RNA sequences, wherein each optimised RNA sequence has a total score above a score threshold.

wherein each optimised RNA sequence comprises at least one different sequence property.

51. The method of claim 50, wherein the random nucleotide sequence assigned with the lowest rank has a total score indicative of the one or more parameter scores having the greatest degree of deviation from the target values associated with the respective one or more parameters and the random nucleotide sequence assigned with the highest rank has a total score indicative of the one or more parameter scores having the smallest degree of deviation from the target values associated with the respective one or more parameters.

52. The method of claim 50 or 51, further comprising identifying the highest ranked random nucleotide sequence for the set of random nucleotide sequences in a first iterative step and for the set of random nucleotide sequences in a second iterative step, the first and second iterative steps among the plurality of iterative steps; and

determining a difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step;

comparing the difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step with a difference threshold,

wherein the stop condition comprises one or more of:

a. a predetermined number of iterations;

b. based on the comparison, the difference in the total score of the highest ranked random nucleotide sequence in the first iterative step and the second iterative step being below the difference threshold; and

c. at least one selected random nucleotide sequence has a total score above an optimisation threshold.

53. The method of any one of claims 50 to 52, wherein the user defined objectives comprise one or more of target application, biological activity, expression profile, manufacturability, secretion profile and/or storability.

54. The method of any one of claims 50 to 53, further comprising:

a. selecting one or more sequence elements from a database based on the one or more user defined objectives and the one or more parameters; optionally

wherein the sequence elements comprise one or more of a 5'-UTR, a 3'-UTR, a poly(A) tail length, signal peptide sequence, an aptamer sequence, protein binding sequences, tandem stop codon and/or nucleic acid binding sequences.

b. combining the selected sequence elements with each optimised RNA sequence to provide a set of full-length mRNA sequences; and

c. outputting a signal indicative of a set of optimised full-length mRNA sequences,

wherein the set of optimised full-length mRNA sequences comprises full-length mRNA sequences which comprise at least one optimised RNA sequence which meets the score threshold of claim 50 (i).

55. The method of claim 54, wherein the set of optimised full-length mRNA sequences comprises full-length mRNA sequences which comprise an optimised RNA sequence which has a total score below the score threshold of claim 50 (i), the method further comprises

    a. repeating steps (d) to (i) of claim 50 for the optimised full-length mRNA sequences; and

    b. outputting a signal indicative of a set of optimised full-length mRNA sequences, comprising the optimised full-length mRNA sequence identified as having the highest total score.

56. The method of any of claims 50 to 55, wherein:

    a. target application comprises one or more of: administration route, end user purpose, target tissue type, target cell type, cellular localisation, and/or cellular processing ;

    b. codon composition efficiency is determined using one or more of: codon decoding times, codon preference, codon usage frequency, codon usage patterns, abundance of tRNA isoacceptors and/or ratio of cognate to near-cognate tRNAs;

    c. specified nucleotide sub-sequences comprise: frameshifting sequences, ribosomal slippery sequences, transcription terminator sequences, RNA polymerase stalling sequences, aptamer sequences, secondary structure forming sequences, restriction enzyme sites, RNA binding protein binding sites and/or frameshifted premature termination codons;

    d. biological activity comprises one or more of: immunogenicity, cellular interactions, adjuvant activity, enzymatic activity and/or cellular effects;

    e. expression profile comprises expression duration, expression location and/or expression level;

    f. manufacturability comprises yield of encoded protein;

    g. secretion profile comprises localisation of the RNA and/or encoded protein in a target cell; and/or

    h. storability comprises stability of the RNA at predetermined environmental conditions.

57. The method of any of claims 50 to 56,

wherein step (g) of claim 50 further comprises combining n nucleotides of the 5' end of at least one first removed nucleotide sequence with L – n nucleotides of the 3' end of a second removed nucleotide sequence to provide a mixed sequence; and/or

randomly mutating at least one of the removed nucleotide sequences to generate a randomly mutated sequence, carrying out steps (e) to (g) for each randomly mutated sequence and selecting at least one optimised randomly mutated sequence based on the score threshold.

58. The method of any of claims 50 to 57,

wherein, in the plurality of iterative steps of step (h) of claim 50, the set of random nucleotide sequences in a given iterative step is selected to include the non-removed nucleotide sequences present from the set of random nucleotide sequences in the previous iterative step and one or more nucleotide sequences that were not included in the set of random nucleotide sequences in the previous iterative step,

wherein the set of random nucleotide sequences in the given iterative step further comprises one or more mixed sequences and/or optimised randomly mutated sequences.

59. The method of any of claims 50 to 58, wherein the set of random nucleotide sequences comprises one or more seeding nucleotide sequences wherein each seeding nucleotide sequence comprises a nucleotide sequence that has a predetermined parameter score for one of the one or more parameters and encodes the target translated protein sequence.

60. Computer readable instructions which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 50 to 59.

61. An apparatus for designing an optimised RNA sequence for translation, the apparatus comprising:

a memory arranged to store machine-readable instructions;

an input unit arranged to receive an input; and

processing circuitry arranged to operably execute the stored machine-readable instructions to:

a. receive, via the input unit, a target translated protein sequence, one or more user-defined objectives for the optimised RNA sequence, and one or more parameters associated with each user-defined objective and corresponding to the target translated protein sequence, the one or more parameters including:

i. codon composition efficiency for translation of the RNA sequence to the target translated protein sequence;

ii. presence or absence of specified nucleotide repeats;

iii. GC content; and

iv. desirability of specified nucleotide sub-sequences;

b. receive, via the input unit, deviation information for each parameter, the deviation information for processing a nucleotide sequence and indicative of a degree of deviation of a nucleotide sequence from a target value of the associated parameter;

c. determine a score function for each parameter based on the deviation information, wherein each score function is usable for processing a nucleotide sequence to generate a score indicative of a degree of deviation between the nucleotide sequence and the target value of the parameter associated with that score function;

d. select, from amongst a plurality of nucleotide sequences, a set of random nucleotide sequences, wherein each random nucleotide sequence comprises a nucleotide sequence that encodes the target translated protein sequence;

e. generate one or more parameter scores for each random nucleotide sequence based on the score functions determined for the corresponding one or more parameters;

f. generate a total score for each random nucleotide sequence based on the one or more parameter scores generated for that random nucleotide sequence;

g. assign a rank to each random nucleotide sequence in the set of random nucleotide sequences based on the generated total scores of the random nucleotide sequences from a lowest rank to a highest rank; and

remove, from the plurality of nucleotide sequences, one or more of the lowest ranked random nucleotide sequences based on the assigned ranking of each random nucleotide sequence, to provide a reduced plurality of nucleotide sequences;

h. perform a plurality of iterative steps including repeating steps (d) to (g), wherein the plurality of nucleotide sequences in step (d) in an

iterative step corresponds to the reduced plurality of nucleotide sequences of step (g) in the previous iterative step,

i. wherein step (h) is repeated until a stop condition is met to provide a set of optimised RNA sequences, wherein each optimised RNA sequence has a total score above a score threshold.

wherein each optimised RNA sequence comprises at least one different sequence property.

62. The method of claim 61, wherein the set of random nucleotide sequences comprises one or more seeding nucleotide sequences wherein each seeding nucleotide sequence comprises a nucleotide sequence that has a predetermined parameter score for one of the one or more parameters and encodes the target translated protein sequence.

63. A system for designing an optimised RNA sequence for translation, the system comprising:

the apparatus of claim 61 or 62; and

a second apparatus comprising a database including one or more sequence elements, wherein the apparatus of claim 61 or 62 is configured to retrieve the one or more sequence elements from the second apparatus for providing a set of full-length mRNA sequences.

Figure 1

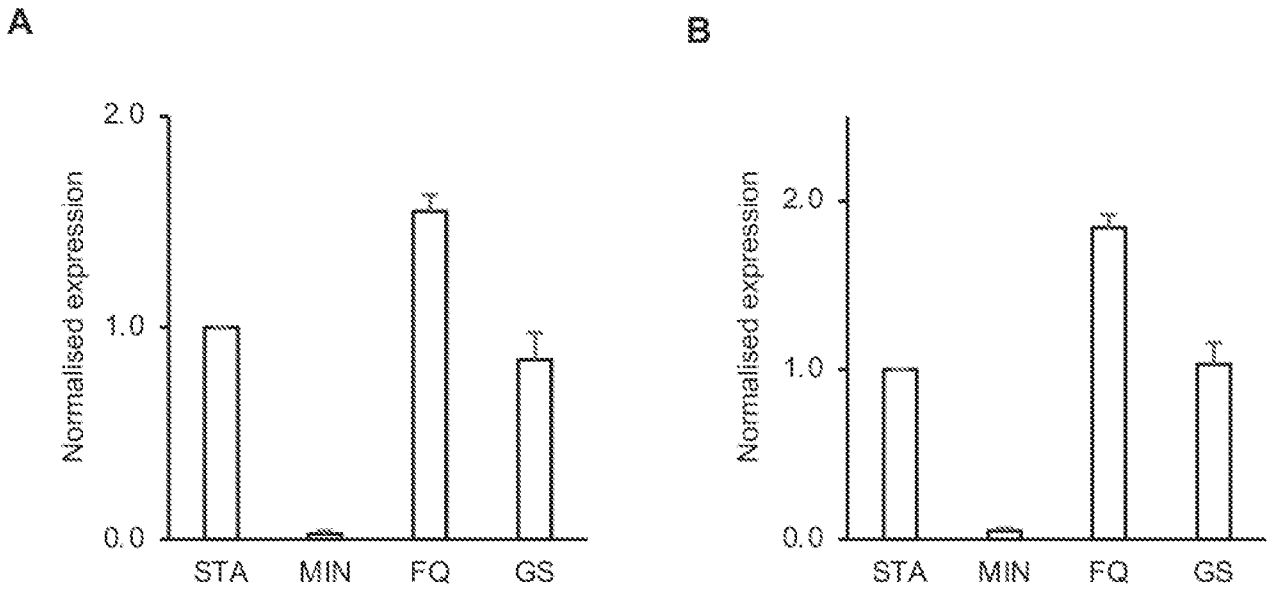


Figure 2

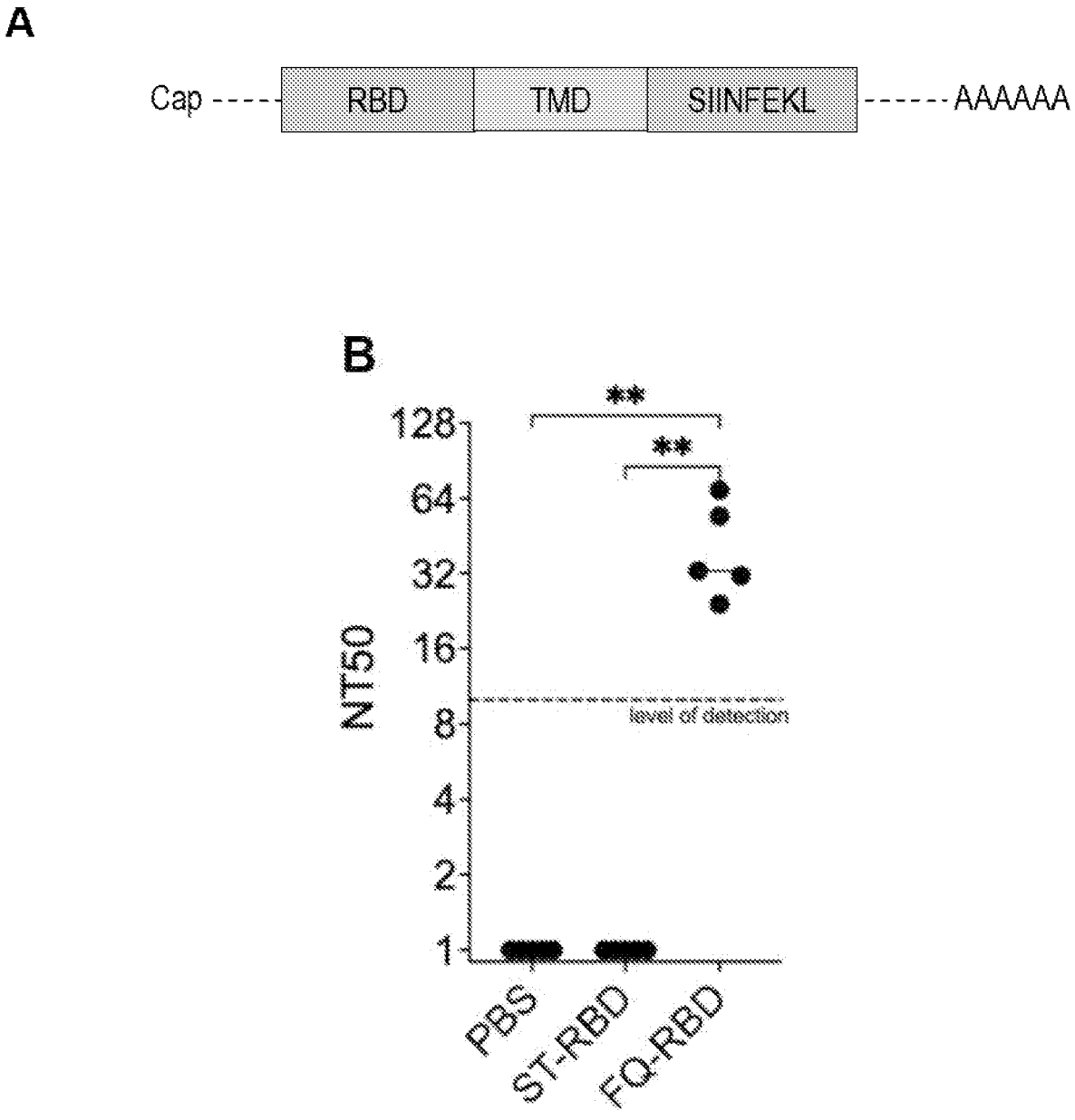
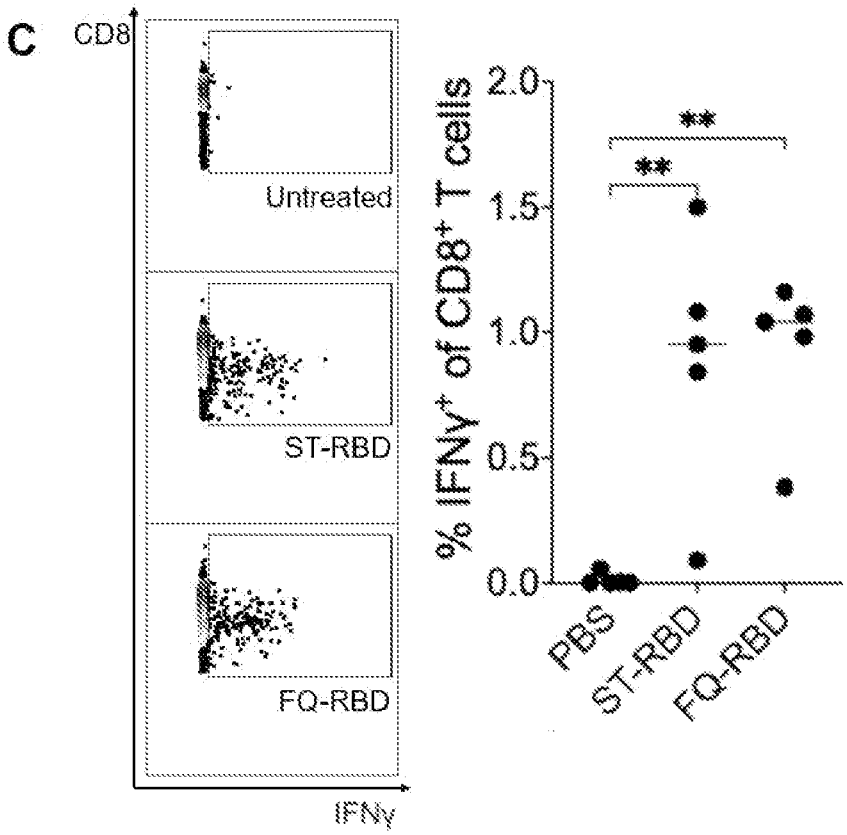

A
Cap
RBD
TMD
SIINFEKL
AAAAAA
B
NT50
128
64
32
16
8
4
2
1
**
**
level of detection
PBS
ST-RBD
FQ-RBD

Figure 2 continued



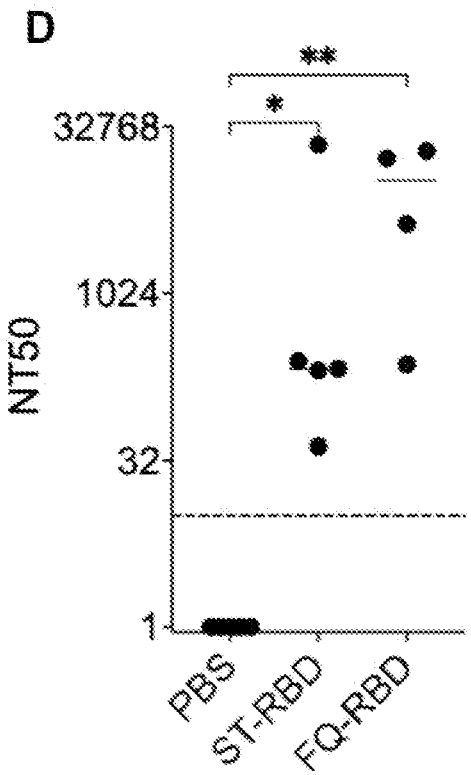


Figure 2 continued

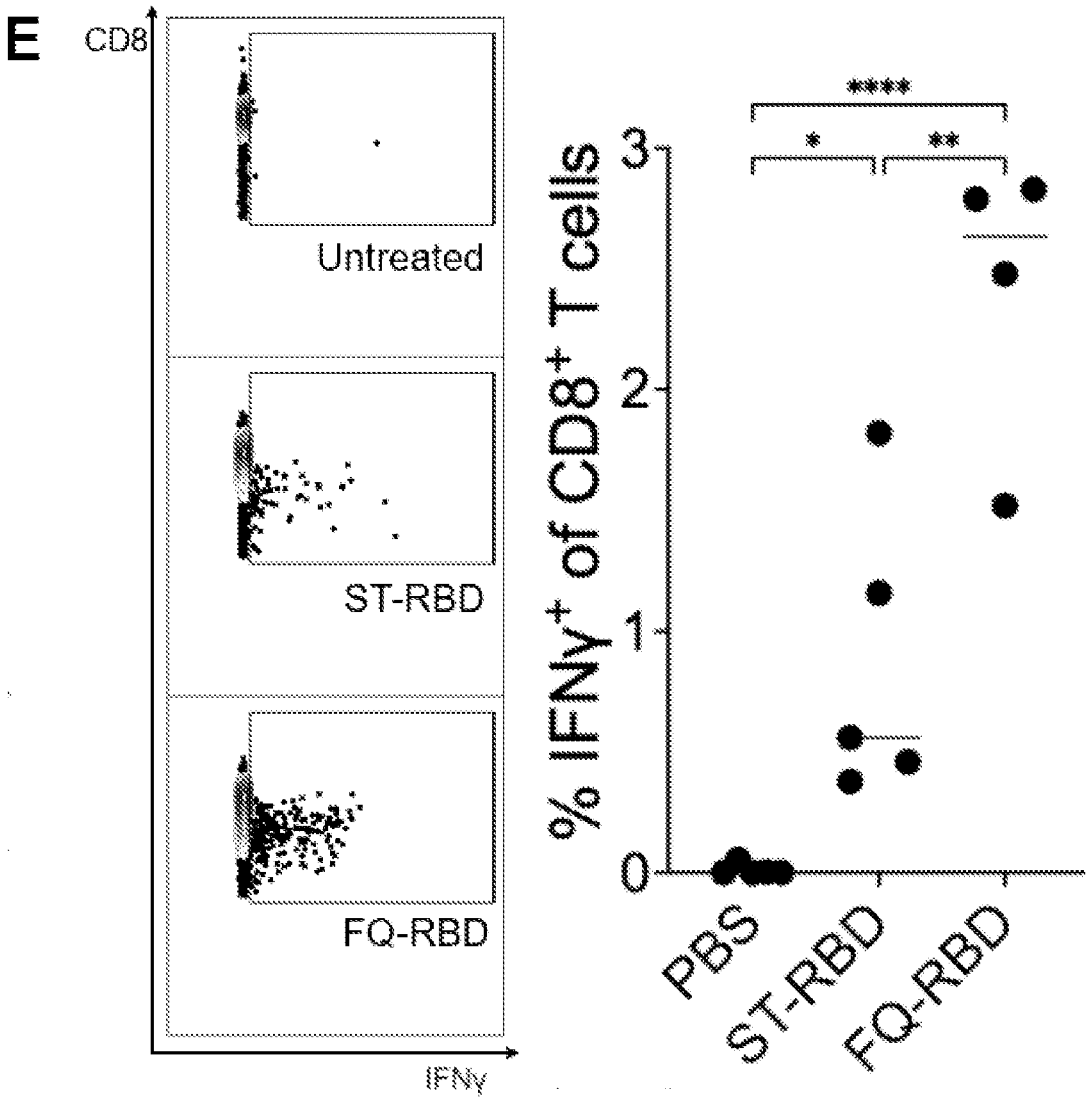


Figure 3

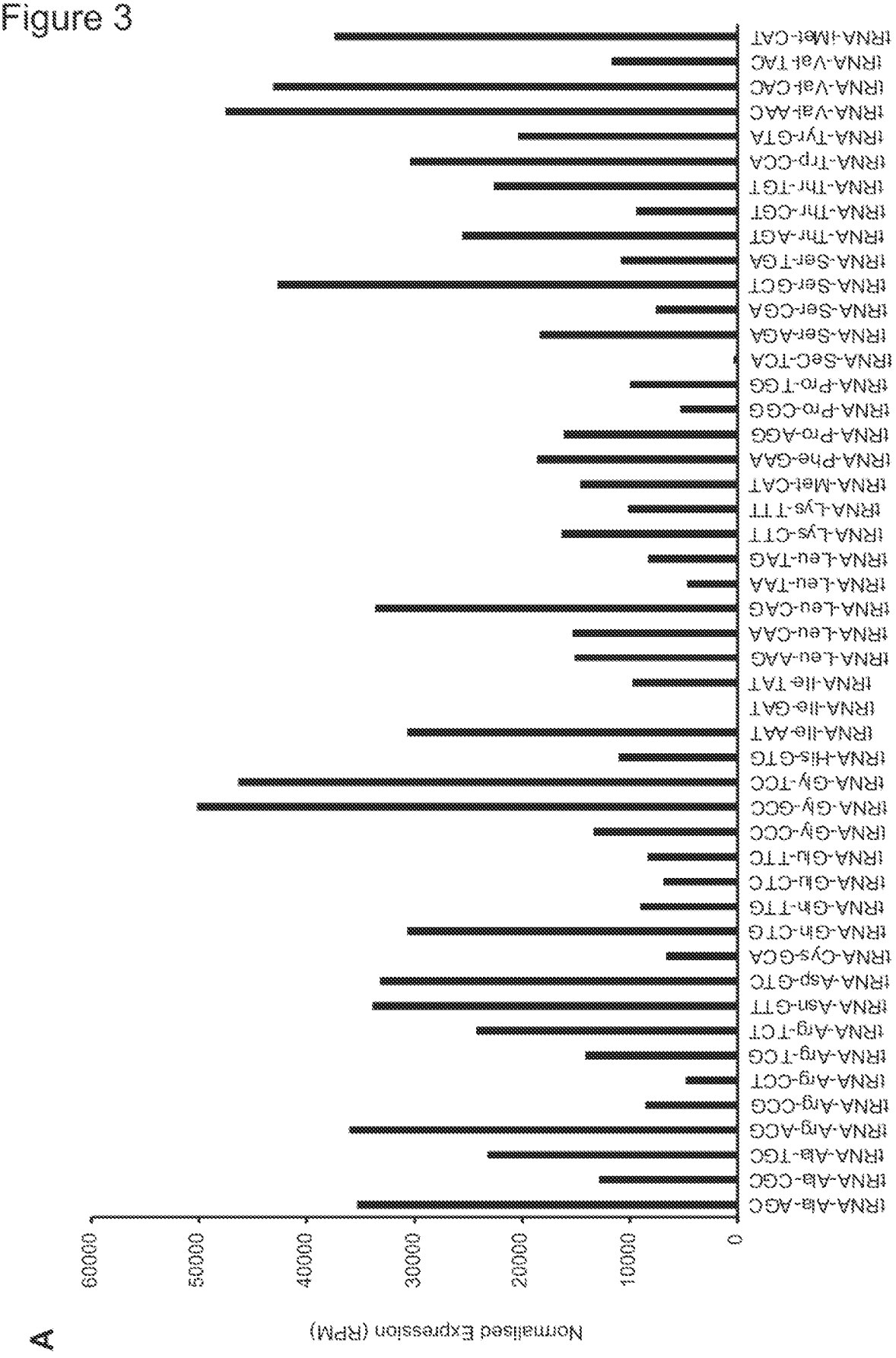

A
Normalised Expression (RPM)
60000
50000
40000
30000
20000
10000
0
tRNA-Ala-AGC
tRNA-Ala-CGC
tRNA-Ala-TGC
tRNA-Arg-ACG
tRNA-Arg-CCG
tRNA-Arg-CCT
tRNA-Arg-TCG
tRNA-Arg-TCT
tRNA-Asn-GTT
tRNA-Asp-GTC
tRNA-Cys-GCA
tRNA-Gln-CTG
tRNA-Gln-TTG
tRNA-Glu-CTC
tRNA-Glu-TTC
tRNA-Gly-CCC
tRNA-Gly-GCC
tRNA-Gly-TCC
tRNA-His-GTG
tRNA-Ile-AAT
tRNA-Ile-GAT
tRNA-Ile-TAT
tRNA-Leu-AAG
tRNA-Leu-CAA
tRNA-Leu-CAG
tRNA-Leu-TAA
tRNA-Leu-TAG
tRNA-Lys-CTT
tRNA-Lys-TTT
tRNA-Met-CAT
tRNA-Phe-GAA
tRNA-Pro-AGG
tRNA-Pro-CGG
tRNA-Pro-TGG
tRNA-Sec-TCA
tRNA-Ser-AGA
tRNA-Ser-CGA
tRNA-Ser-GCT
tRNA-Ser-TGA
tRNA-Thr-ACT
tRNA-Thr-CGT
tRNA-Thr-TGT
tRNA-Trp-CCA
tRNA-Tyr-GTA
tRNA-Val-AAC
tRNA-Val-CAC
tRNA-Val-TAC
tRNA-iMet-CAT

Figure 3 continued

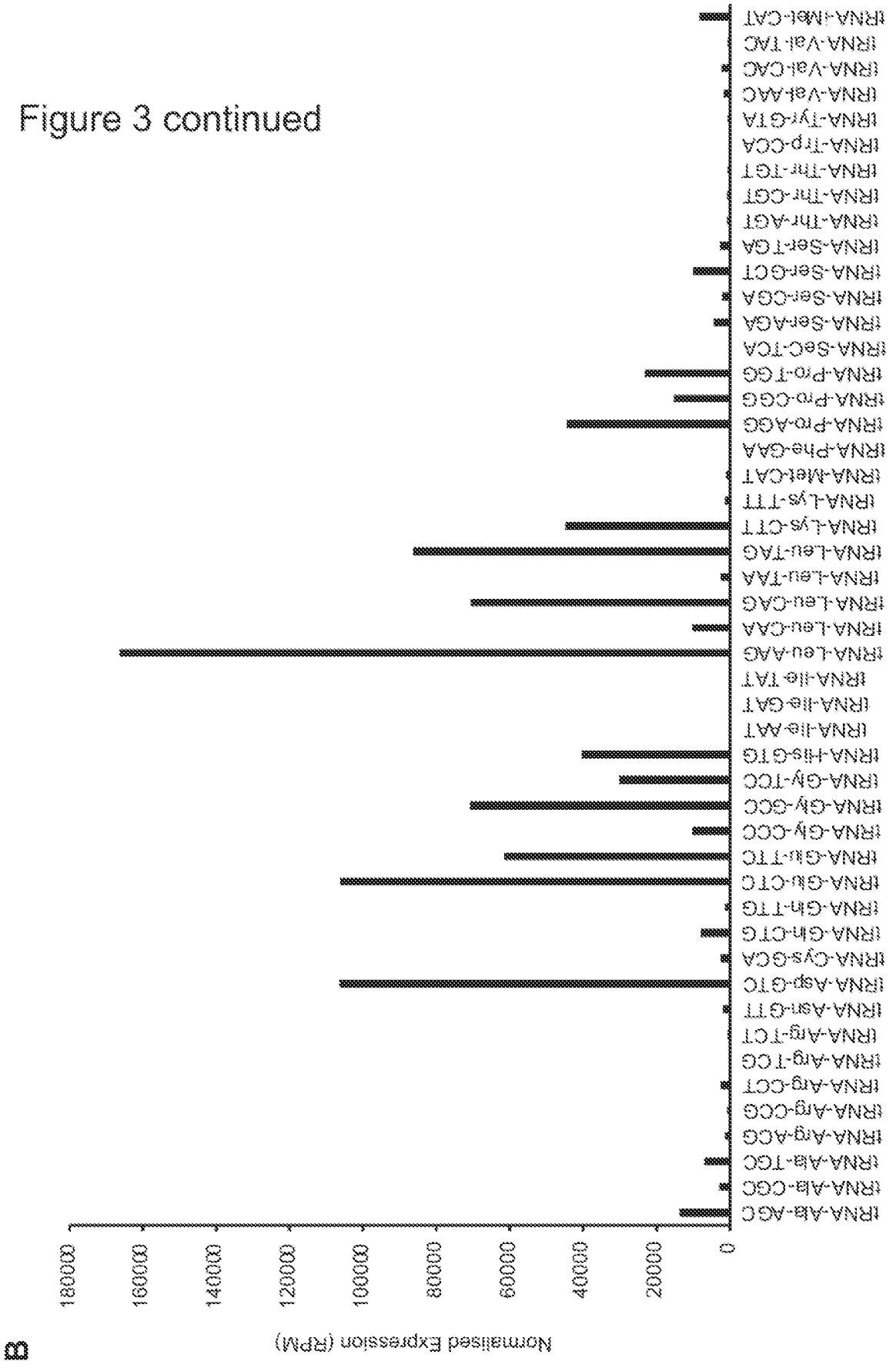
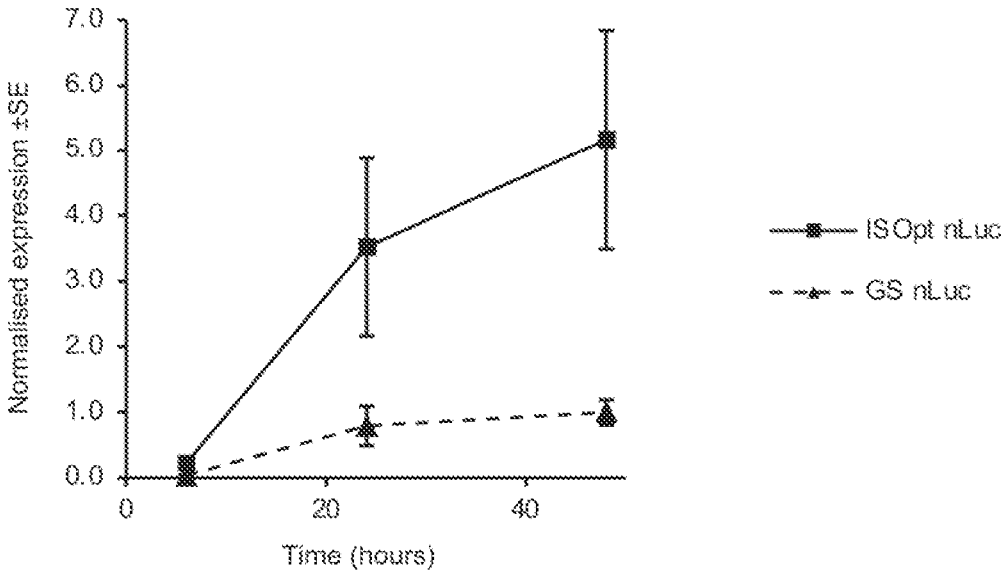


Figure 4

A


7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Normalised expression ±SE
0
20
40
Time (hours)
ISOpt nLuc
GS nLuc

B

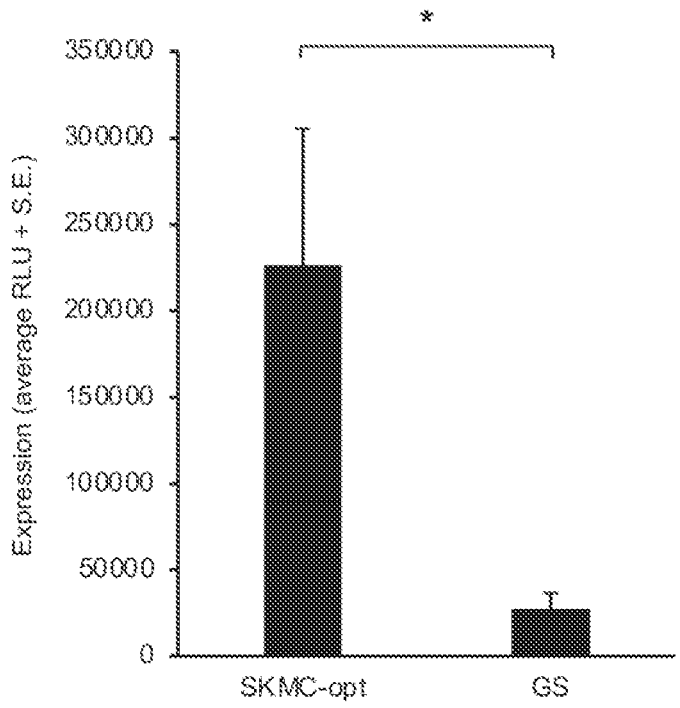

*
350000
300000
250000
200000
150000
100000
50000
0
Expression (average RLU + S.E.)
SKMC-opt
GS

Figure 5

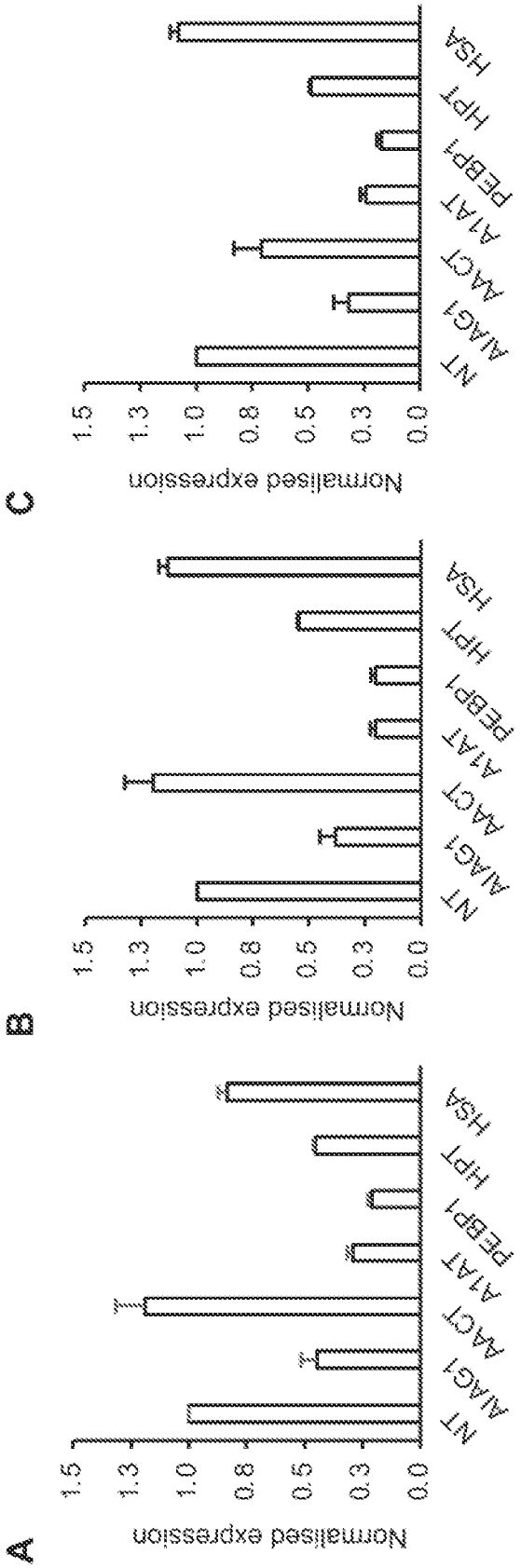
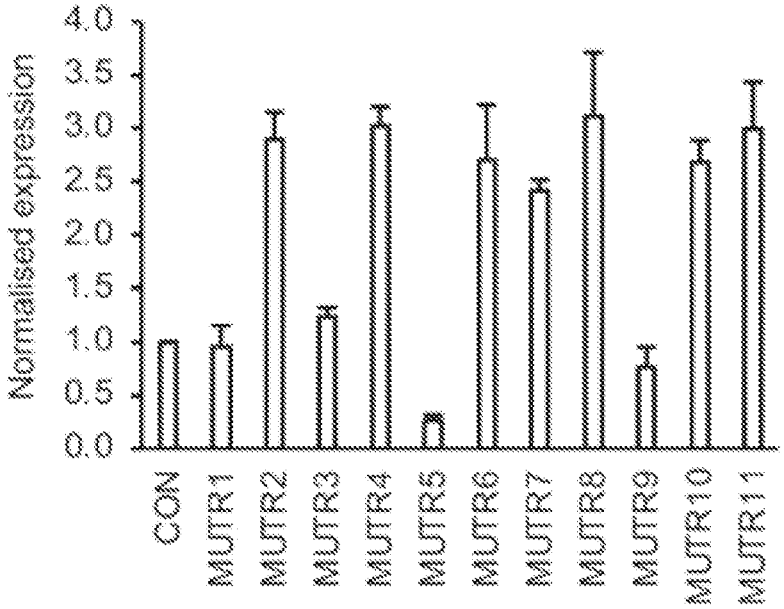

A
B
C
Normalised expression
1.5
1.3
1.0
0.8
0.5
0.3
0.0
NT
AIAG1
AACT
AIAT
PEBP1
HPT
HSA

Figure 6

A


Normalised expression
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
CON
MUTR1
MUTR2
MUTR3
MUTR4
MUTR5
MUTR6
MUTR7
MUTR8
MUTR9
MUTR10
MUTR11

B

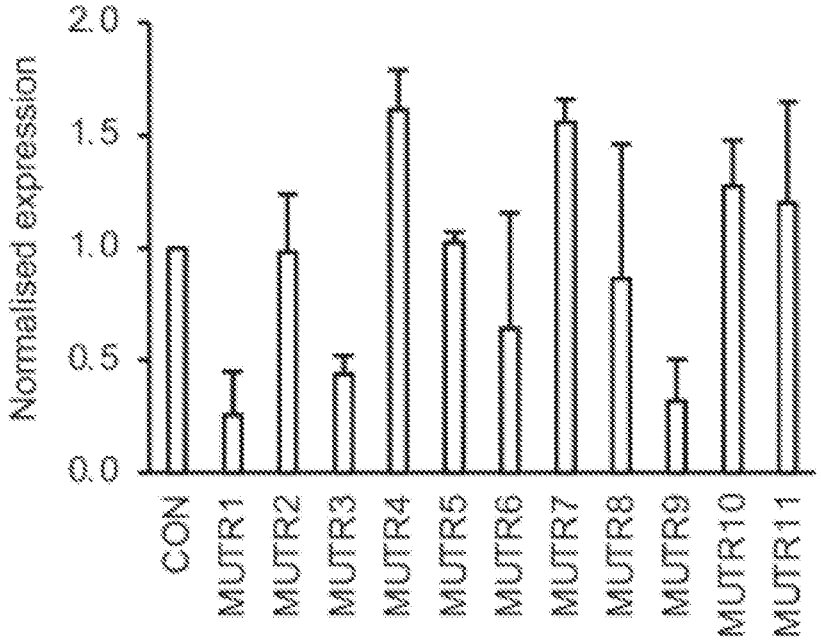
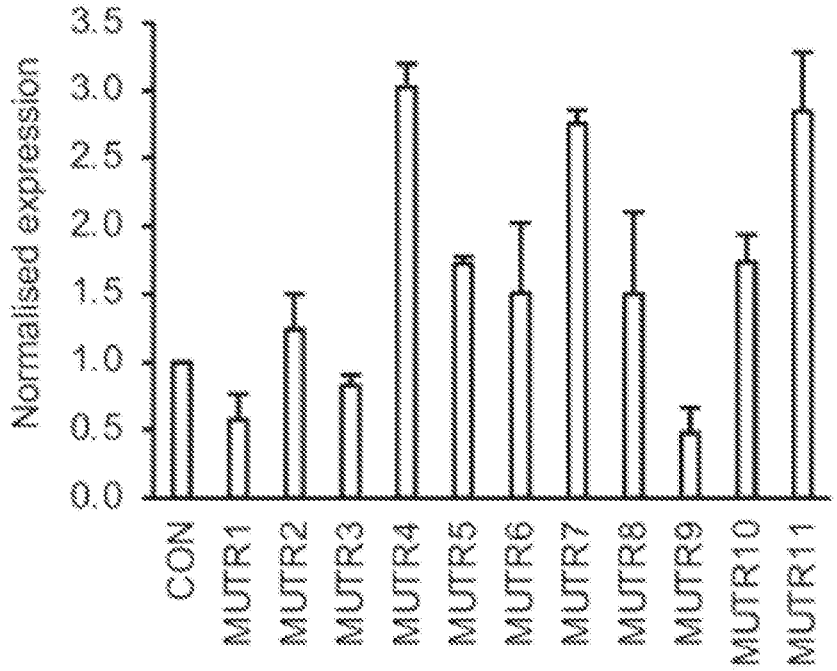

Normalised expression
2.0
1.5
1.0
0.5
0.0
CON
MUTR1
MUTR2
MUTR3
MUTR4
MUTR5
MUTR6
MUTR7
MUTR8
MUTR9
MUTR10
MUTR11

Figure 6 continued

C



D

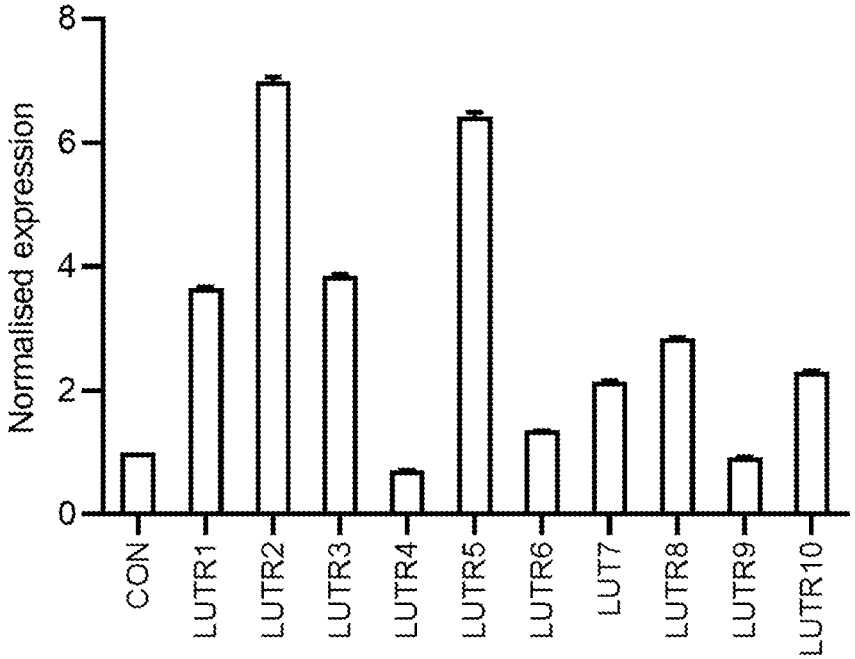
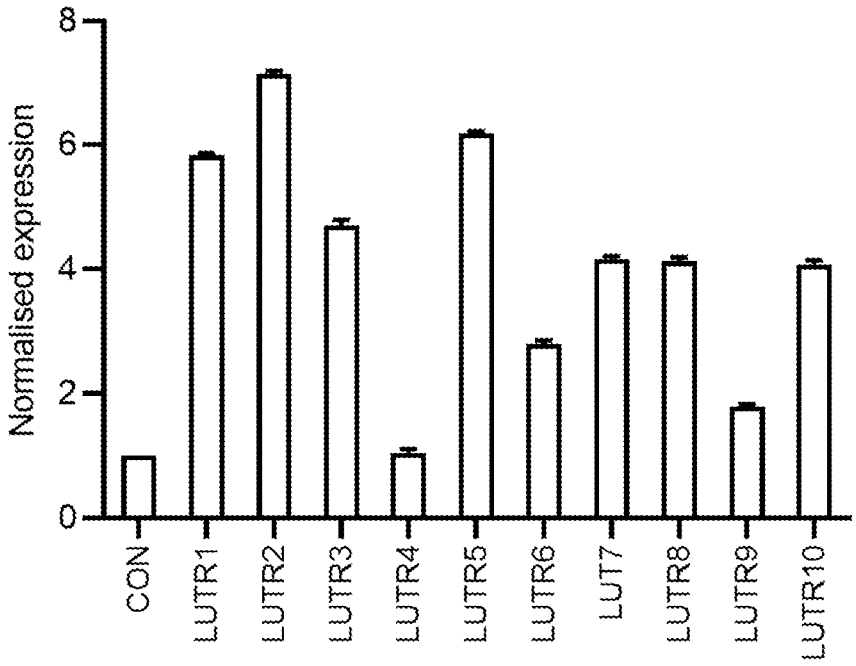


Figure 6 continued

E



F

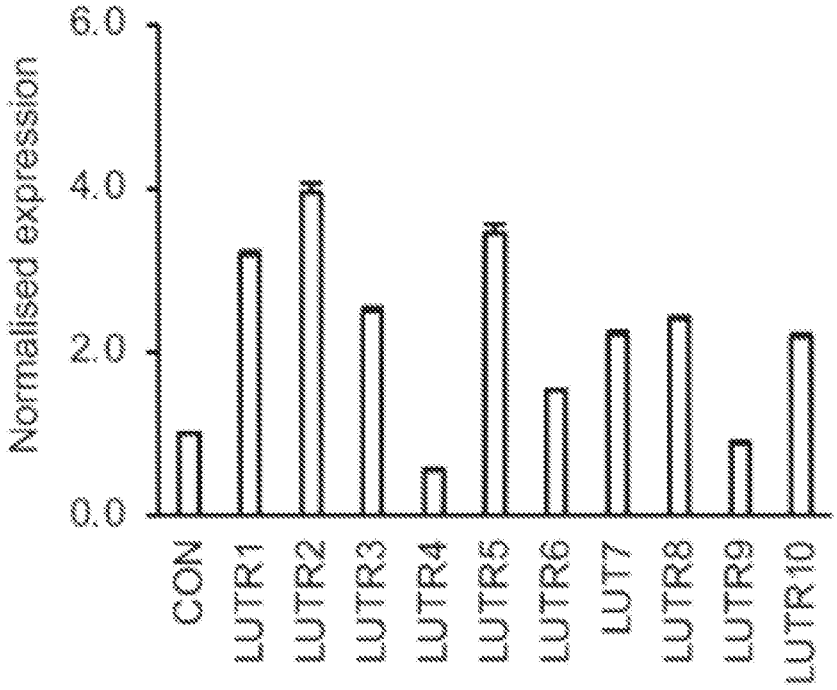


Figure 7


A
0 U NO POLY (A)
2.5 U AAAAAAAAAAAAAAAA +47
5 U AAAAAAAAAAAAAAAAAAAAAAAA +97
10 U AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA +135
15 U AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA +167

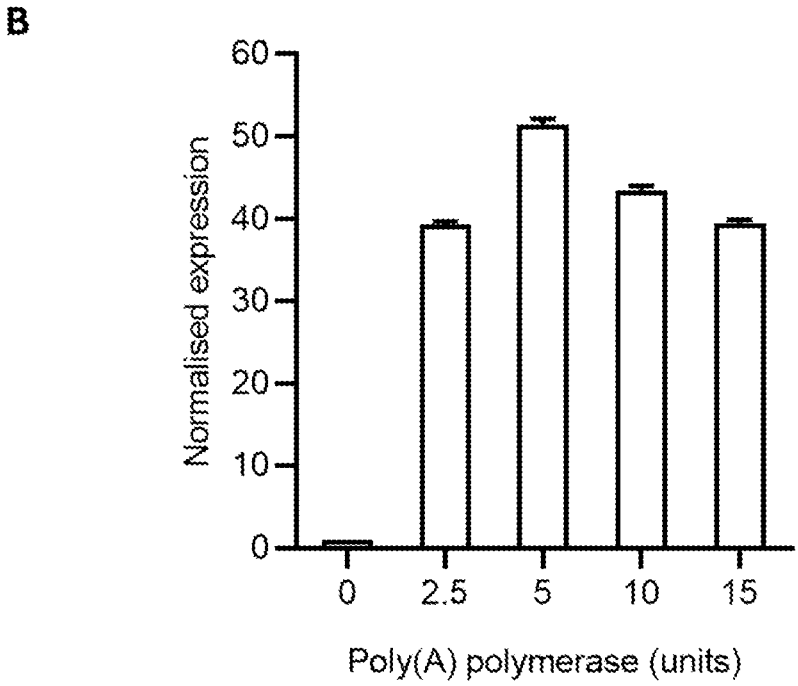

B
Normalised expression
60
50
40
30
20
10
0
0
2.5
5
10
15
Poly(A) polymerase (units)

Figure 7 continued

C

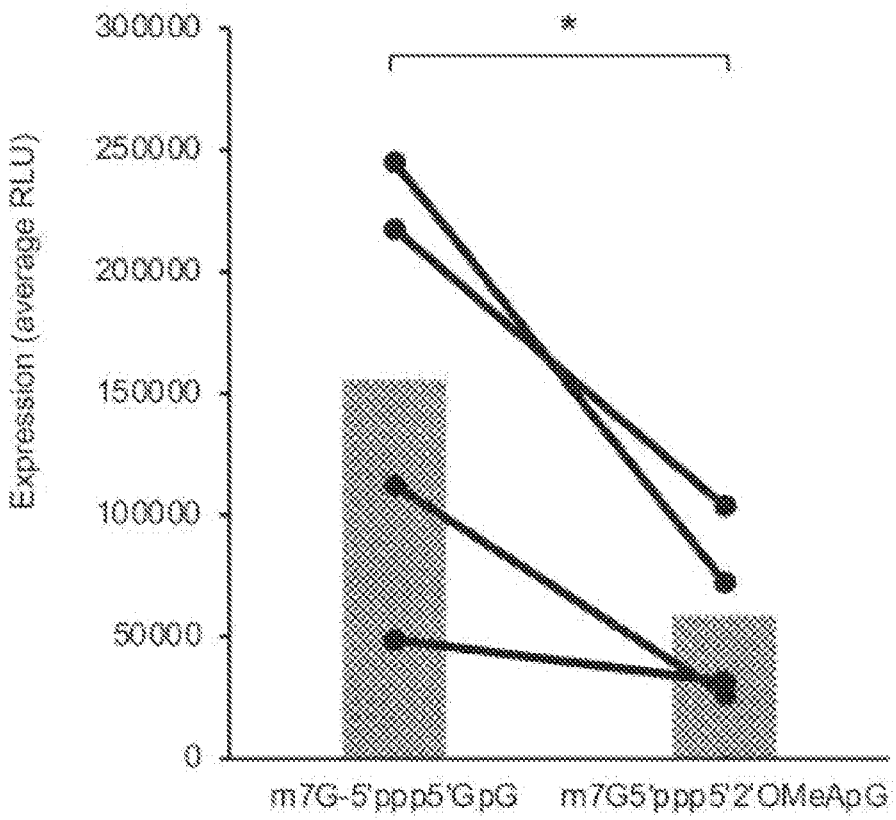
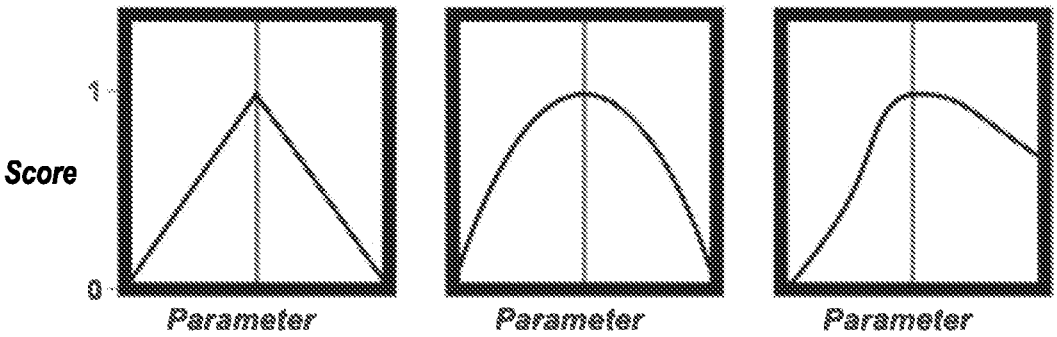


Figure 8

A


Score
1
0
Parameter
Parameter
Parameter

B

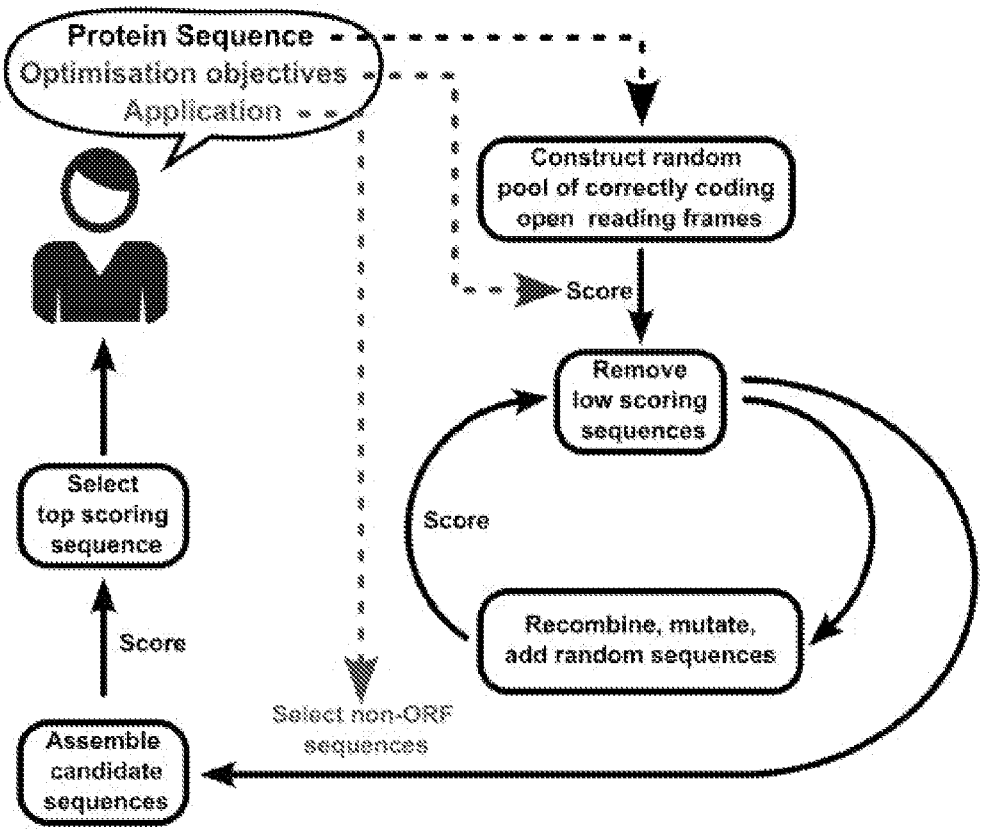

Protein Sequence
Optimisation objectives
Application
Construct random pool of correctly coding open reading frames
Score
Remove low scoring sequences
Score
Recombine, mutate, add random sequences
Select top scoring sequence
Score
Select non-ORF sequences
Assemble candidate sequences

Figure 9

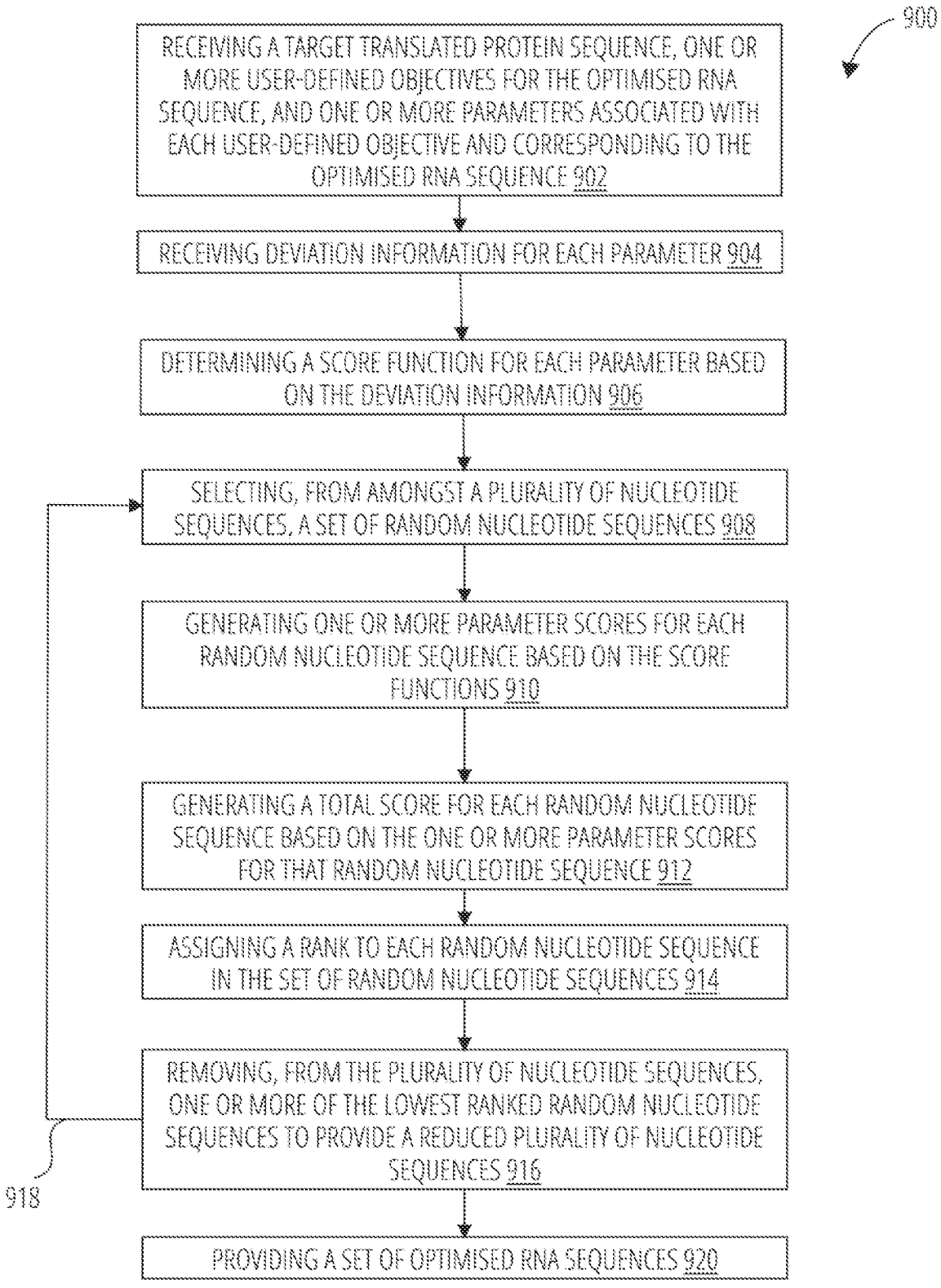

900
RECEIVING A TARGET TRANSLATED PROTEIN SEQUENCE, ONE OR MORE USER-DEFINED OBJECTIVES FOR THE OPTIMISED RNA SEQUENCE, AND ONE OR MORE PARAMETERS ASSOCIATED WITH EACH USER-DEFINED OBJECTIVE AND CORRESPONDING TO THE OPTIMISED RNA SEQUENCE 902
RECEIVING DEVIATION INFORMATION FOR EACH PARAMETER 904
DETERMINING A SCORE FUNCTION FOR EACH PARAMETER BASED ON THE DEVIATION INFORMATION 906
SELECTING, FROM AMONGST A PLURALITY OF NUCLEOTIDE SEQUENCES, A SET OF RANDOM NUCLEOTIDE SEQUENCES 908
GENERATING ONE OR MORE PARAMETER SCORES FOR EACH RANDOM NUCLEOTIDE SEQUENCE BASED ON THE SCORE FUNCTIONS 910
GENERATING A TOTAL SCORE FOR EACH RANDOM NUCLEOTIDE SEQUENCE BASED ON THE ONE OR MORE PARAMETER SCORES FOR THAT RANDOM NUCLEOTIDE SEQUENCE 912
ASSIGNING A RANK TO EACH RANDOM NUCLEOTIDE SEQUENCE IN THE SET OF RANDOM NUCLEOTIDE SEQUENCES 914
REMOVING, FROM THE PLURALITY OF NUCLEOTIDE SEQUENCES, ONE OR MORE OF THE LOWEST RANKED RANDOM NUCLEOTIDE SEQUENCES TO PROVIDE A REDUCED PLURALITY OF NUCLEOTIDE SEQUENCES 916
918
PROVIDING A SET OF OPTIMISED RNA SEQUENCES 920

Figure 10

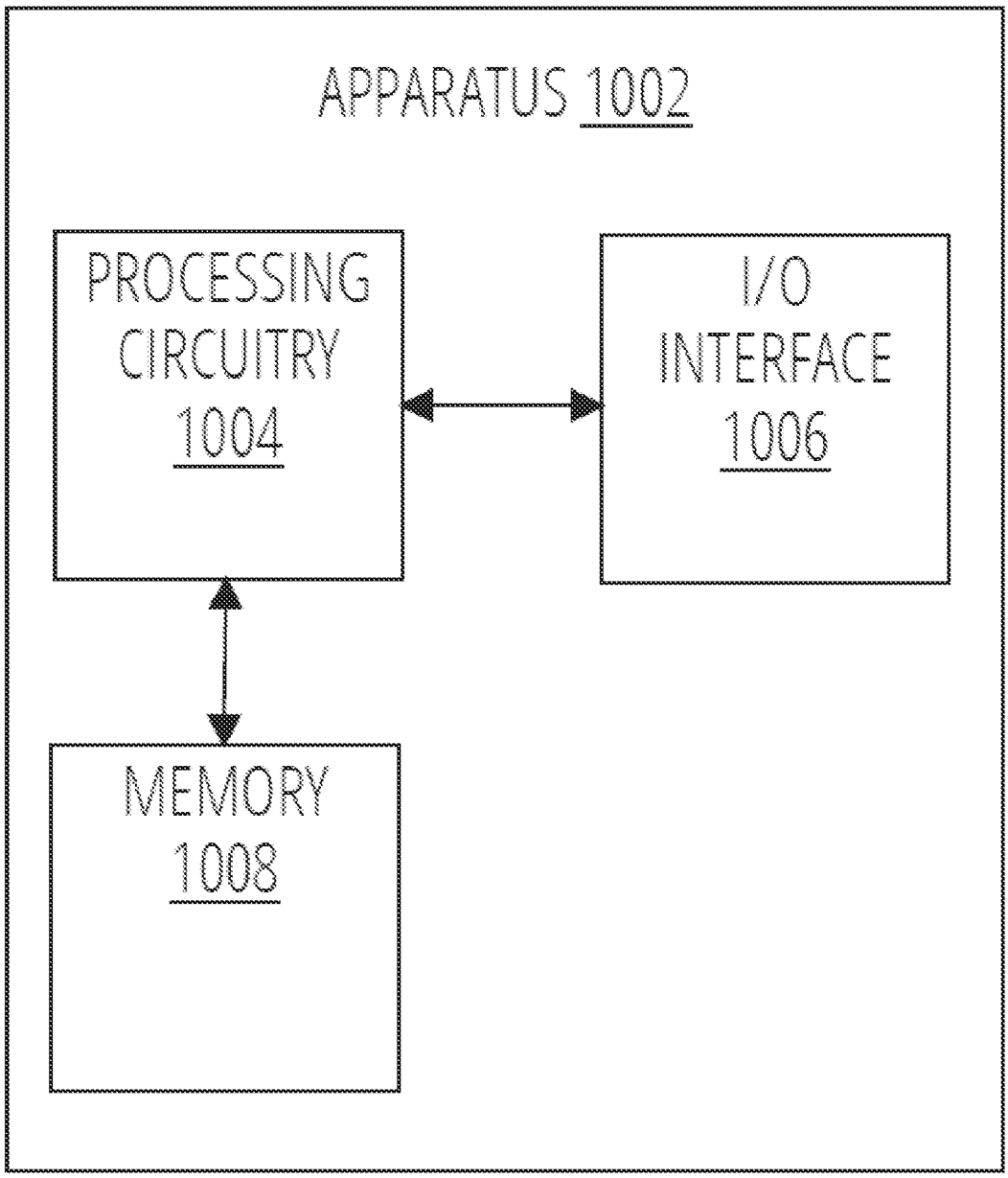

APPARATUS 1002
PROCESSING CIRCUITRY 1004
I/O INTERFACE 1006
MEMORY 1008

Figure 11

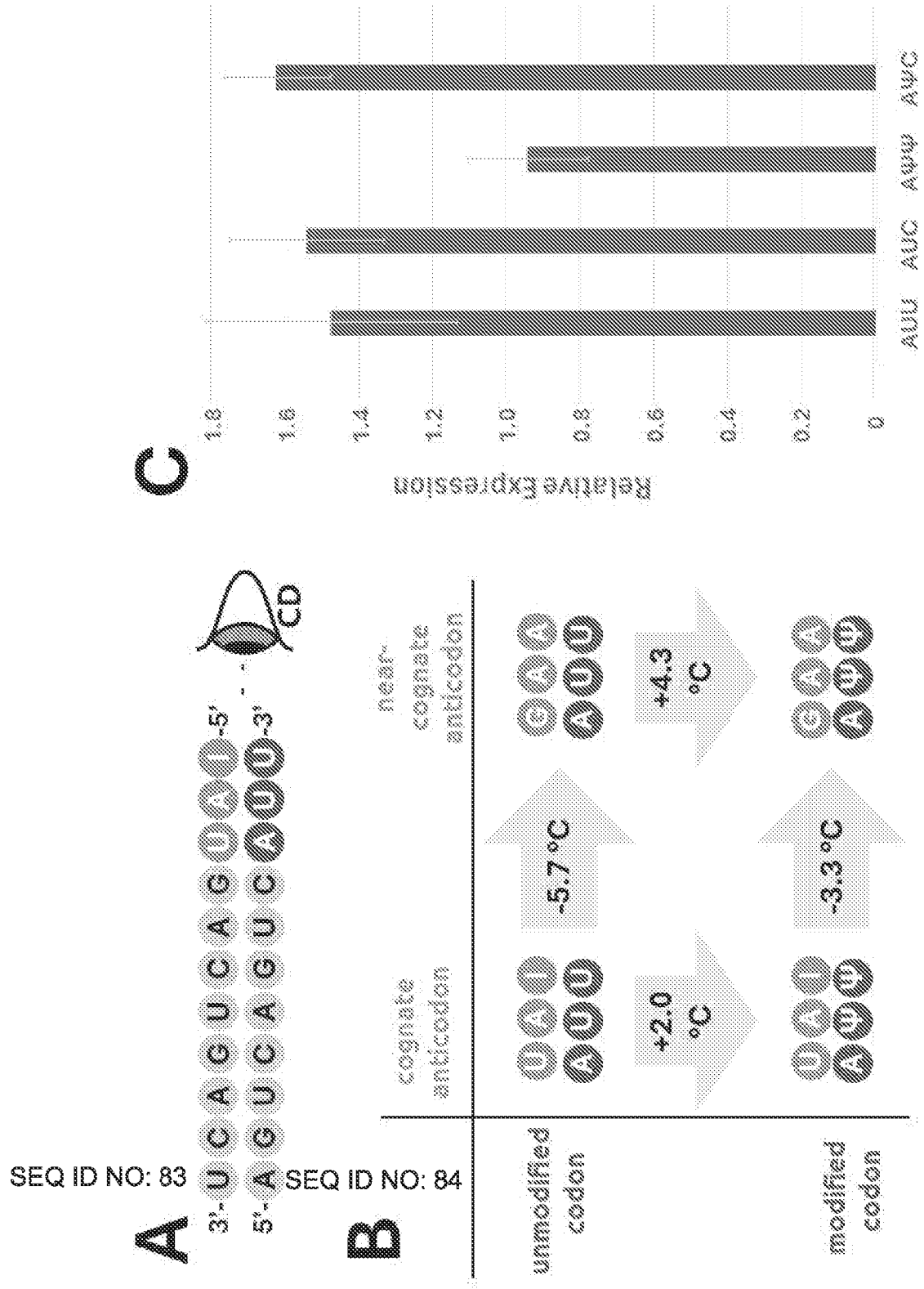
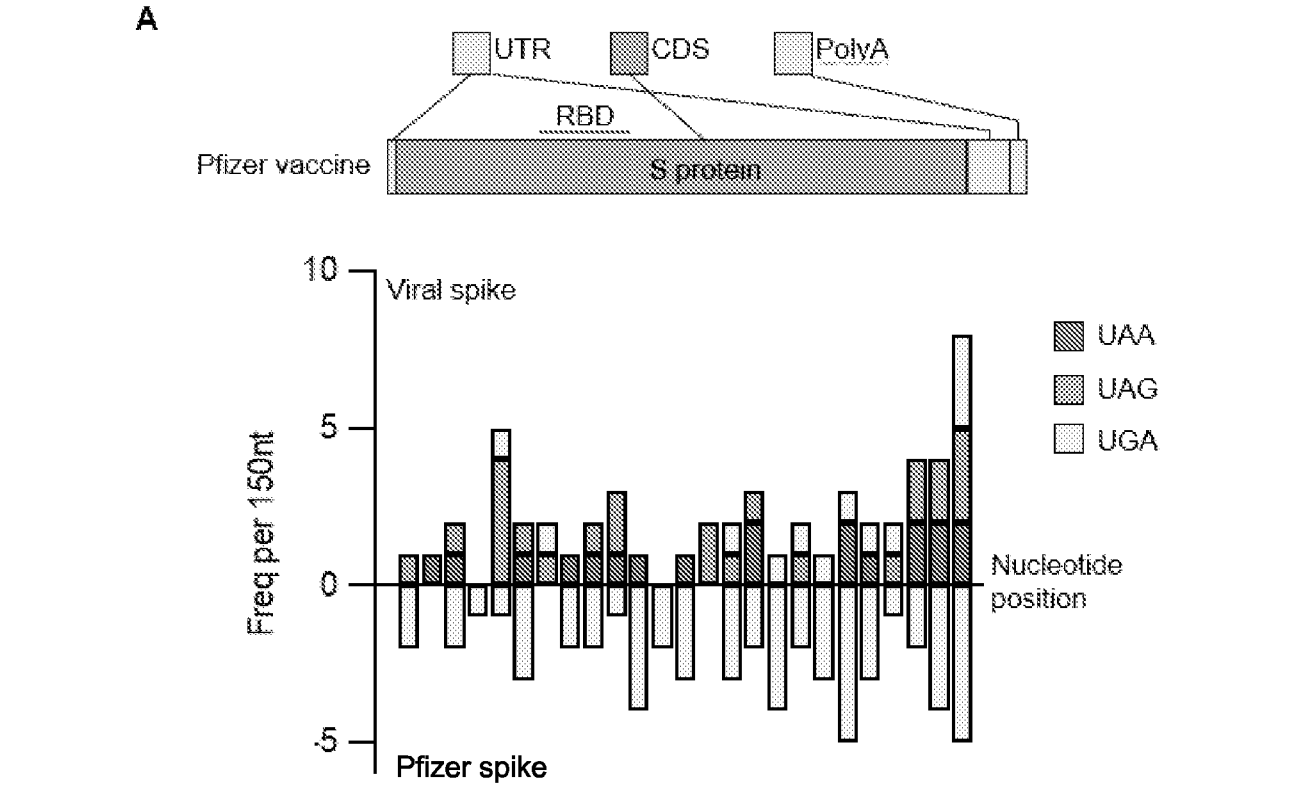

C
Relative Expression
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0
AUU
AUC
AψU
AψC
A
SEQ ID NO: 83
3'- U C A G U C A G U C A G U C A -5'
5'- A G U C A G U C A G U C A U -3'
SEQ ID NO: 84
CD
B
cognate anticodon
near-cognate anticodon
unmodified codon
modified codon
-5.7 °C
+2.0 °C
+4.3 °C
-3.3 °C

Figure 12


A
UTR
CDS
PolyA
RBD
Pfizer vaccine
S protein
10
5
0
-5
Viral spike
Pfizer spike
Freq per 150nt
UAA
UAG
UGA
Nucleotide position

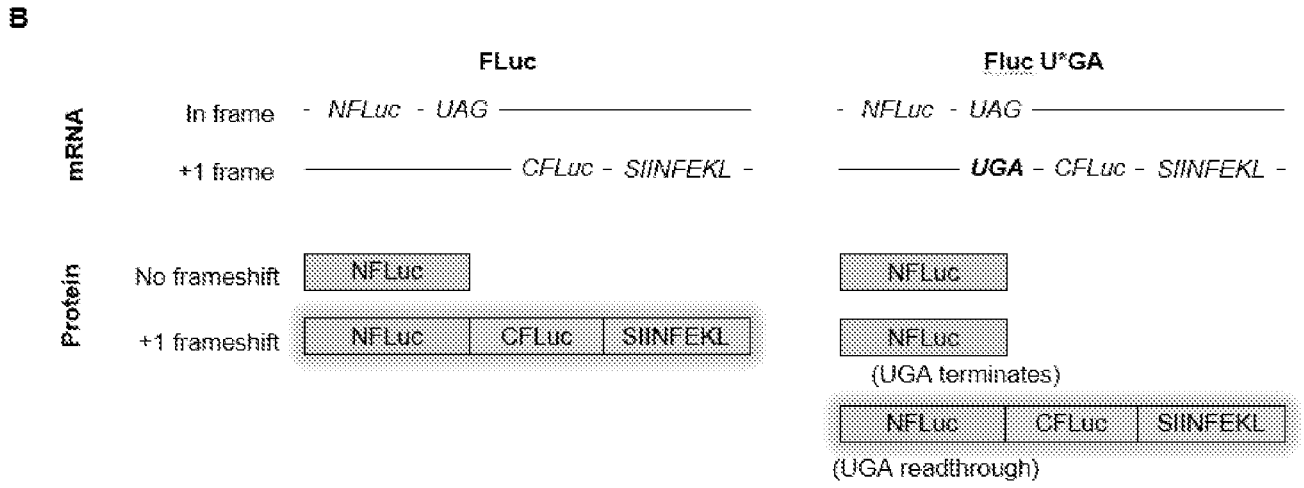

B
FLuc
Fluc U*GA
mRNA
In frame
+1 frame
- NFLuc - UAG
CFLuc - SIINFEKL -
UGA - CFLuc - SIINFEKL -
Protein
No frameshift
+1 frameshift
NFLuc
CFLuc
SIINFEKL
(UGA terminates)
(UGA readthrough)

Figure 12 continued

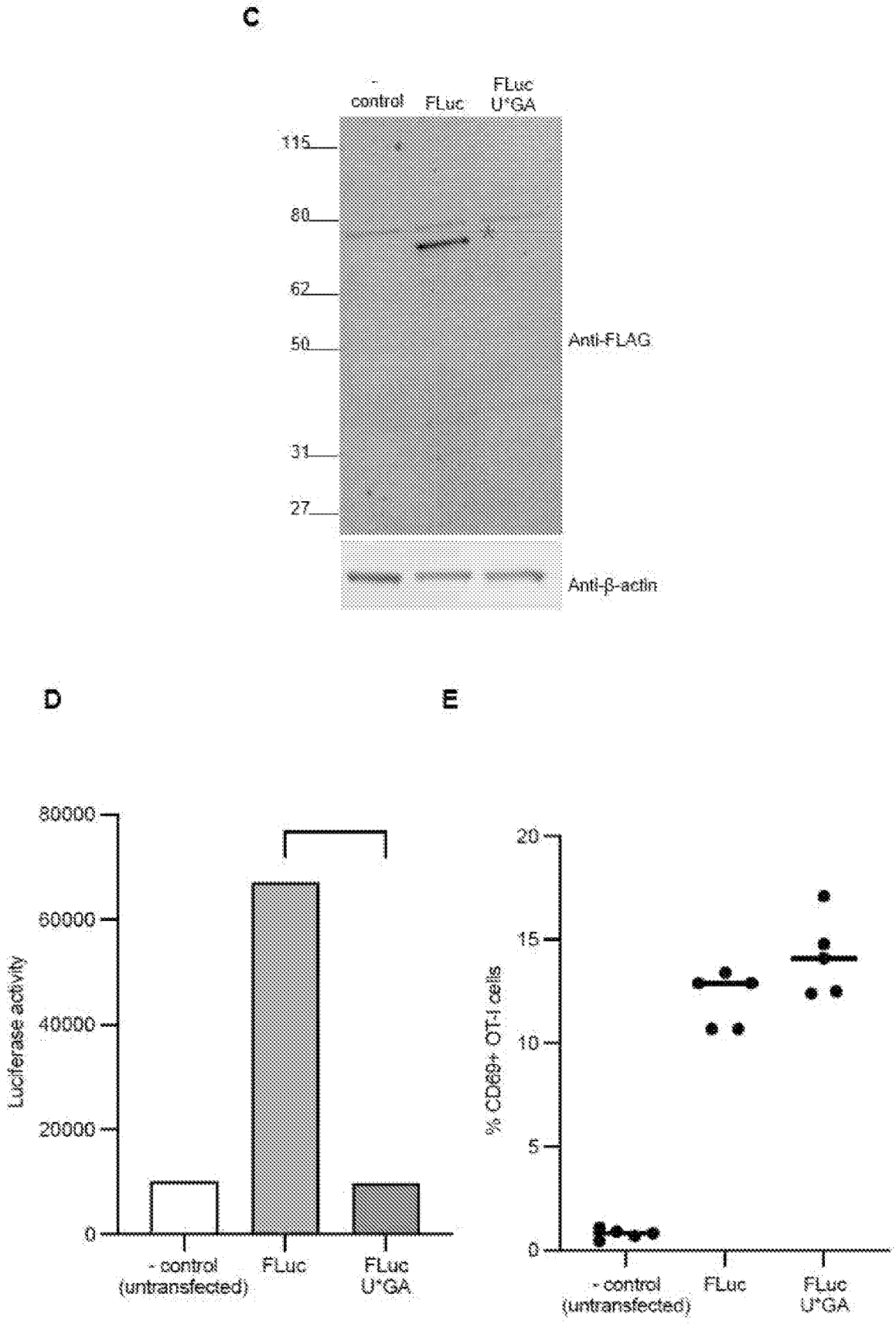
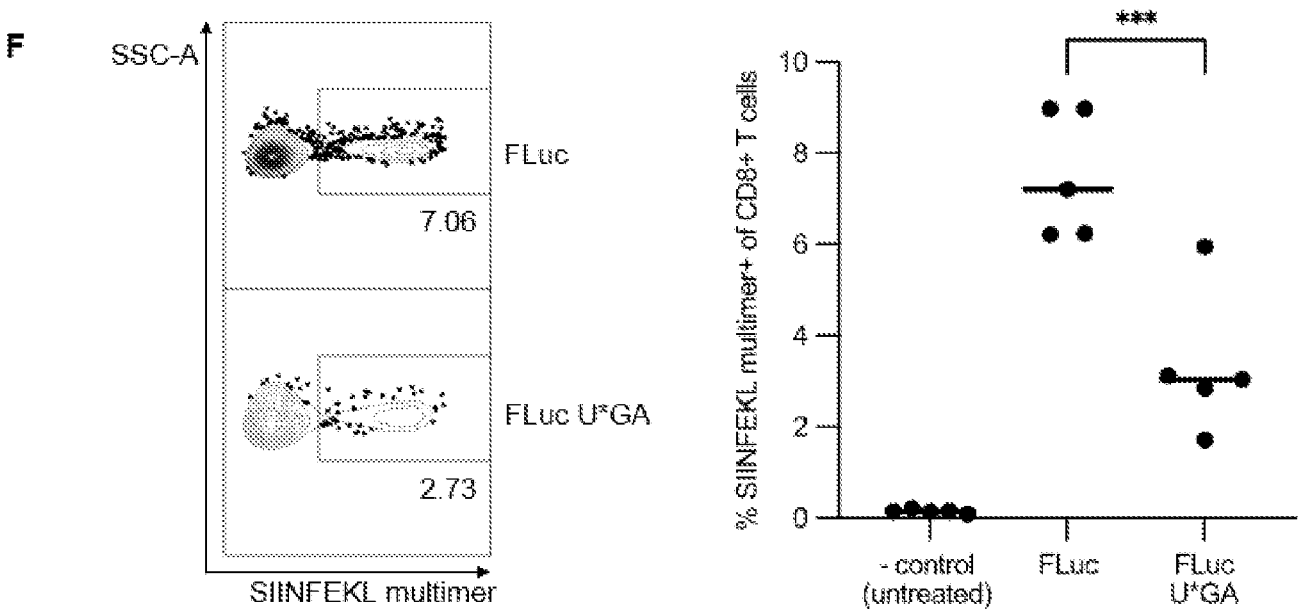


Figure 12 continued



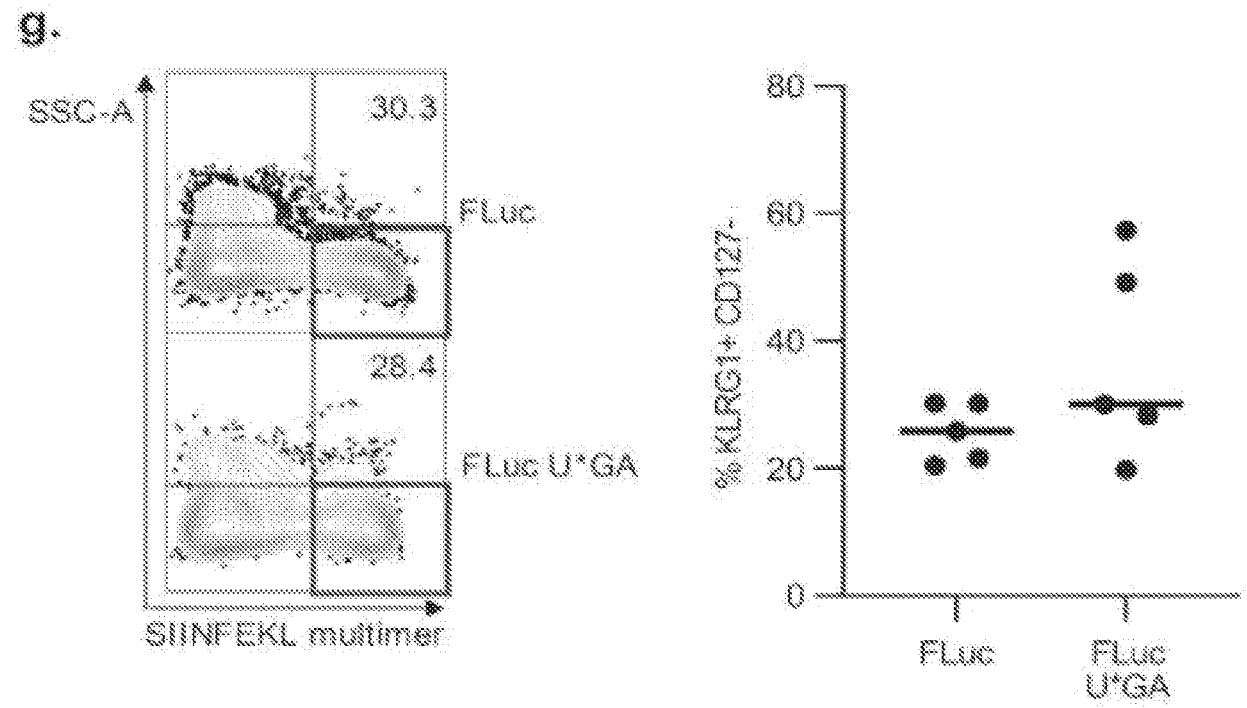
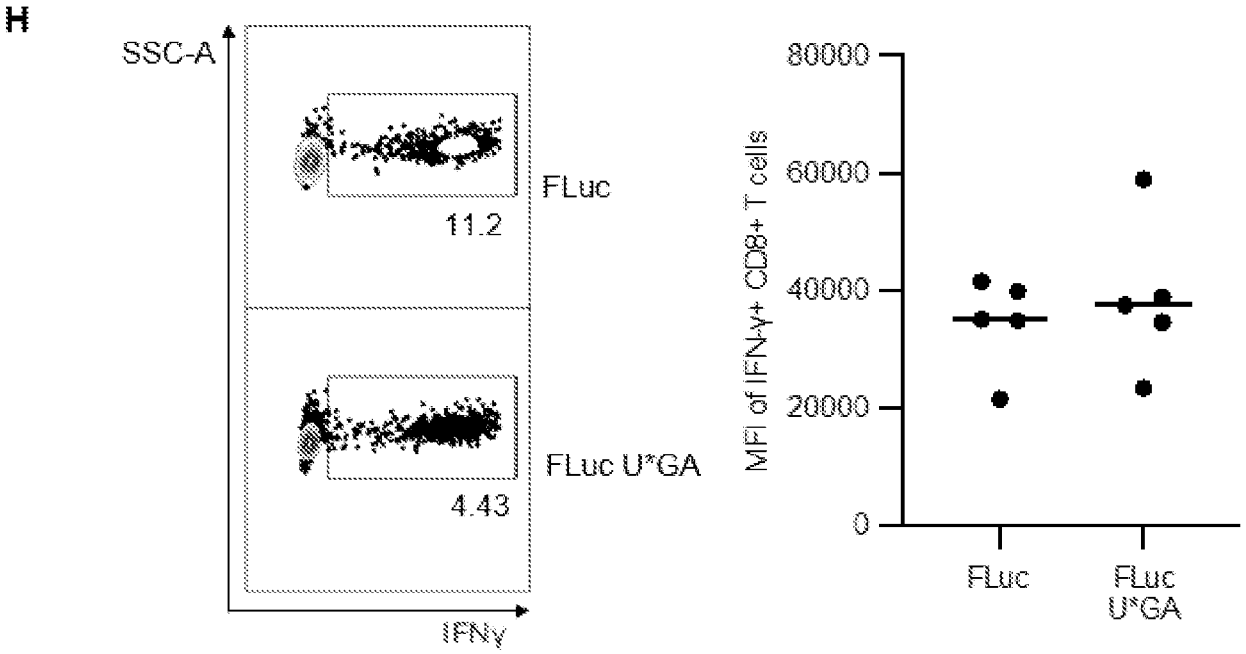


Figure 12 continued



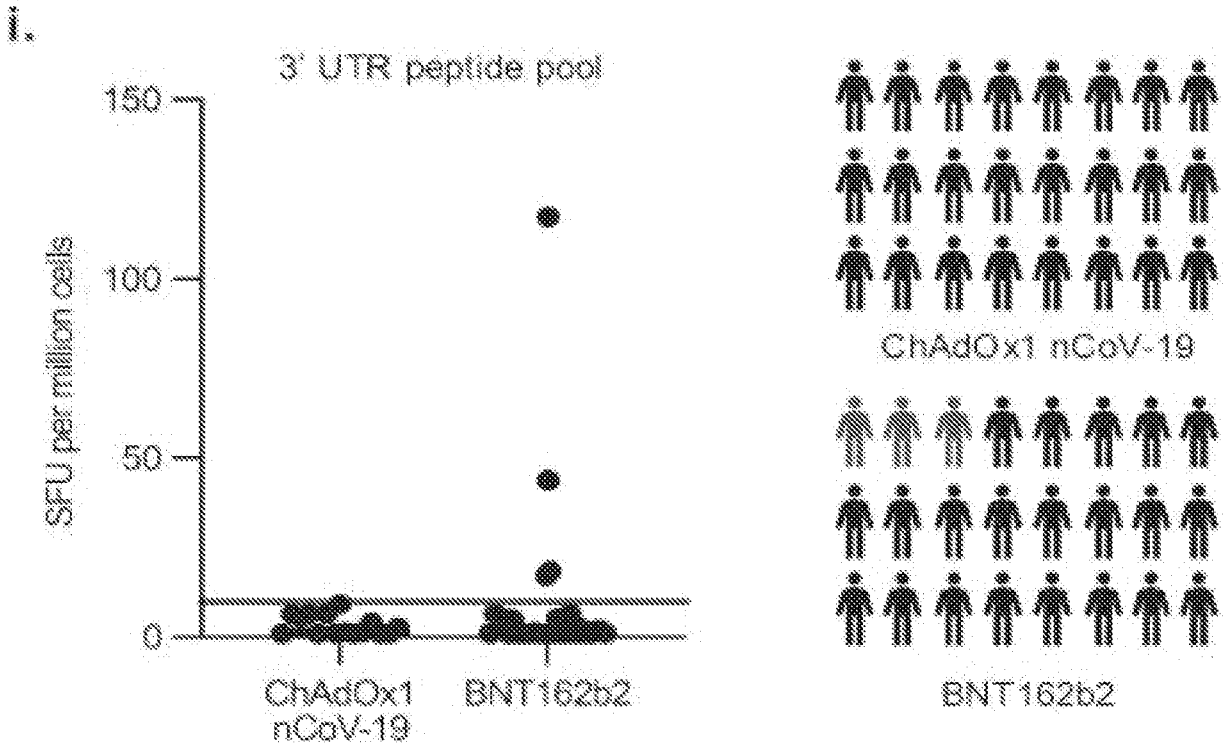


Figure 13

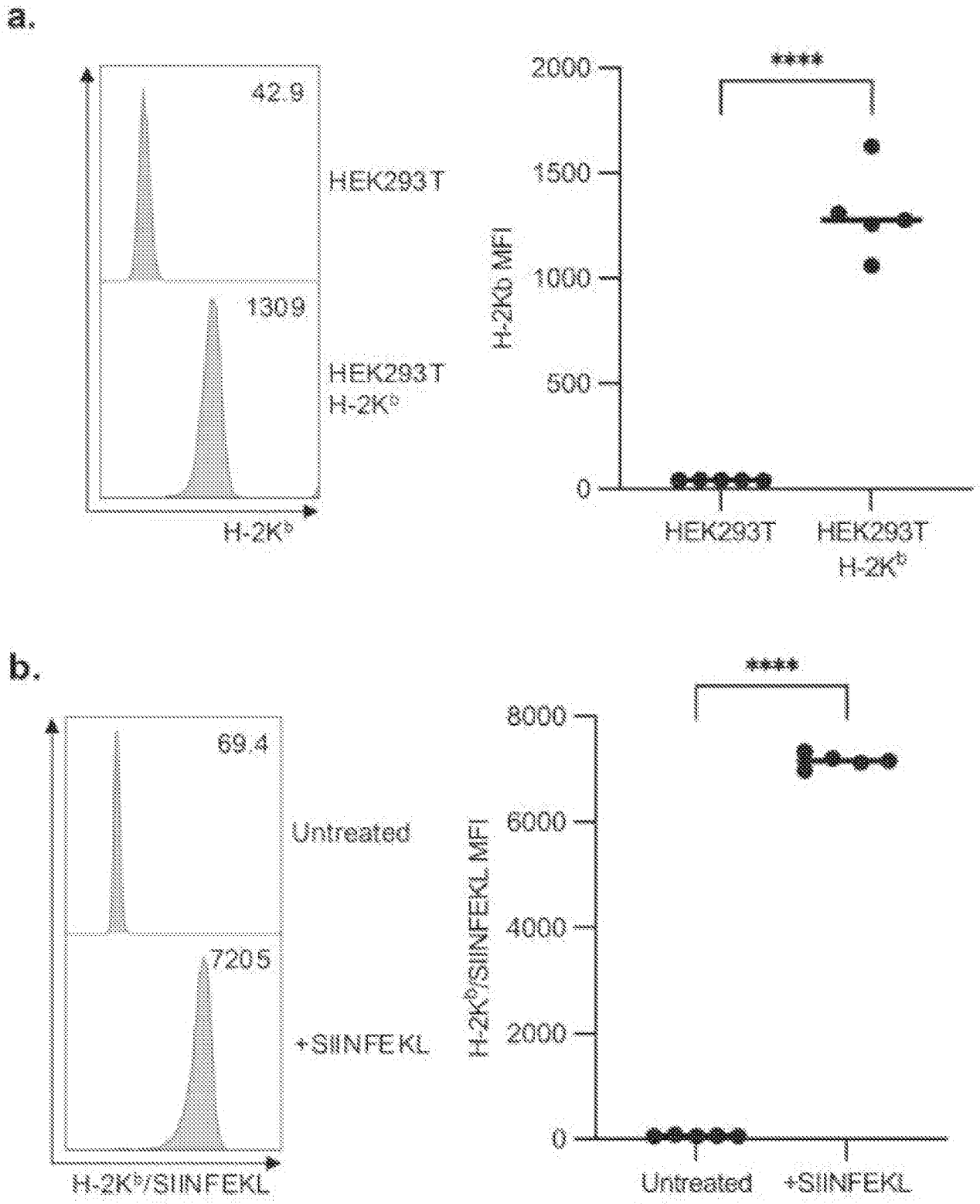
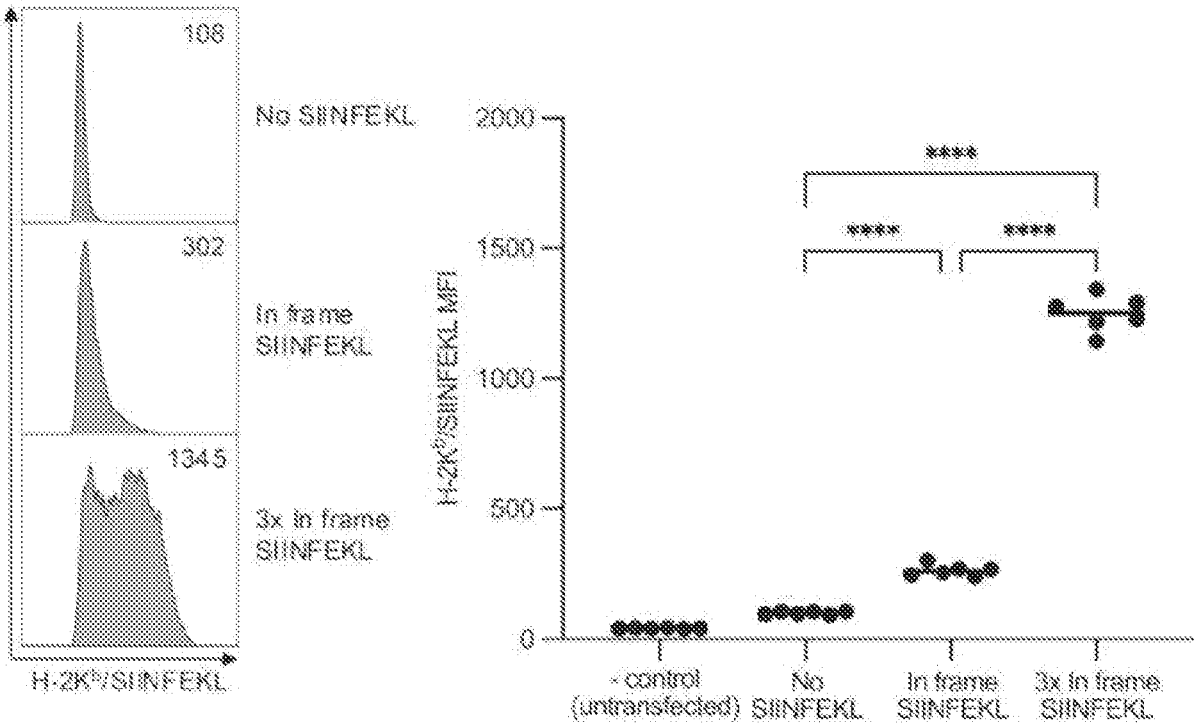

a.
42.9
HEK293T
1309
HEK293T
H-2K^b
H-2K^b
H-2K^b MFI
2000
1500
1000
500
0
****
HEK293T
HEK293T
H-2K^b
b.
69.4
Untreated
7205
+SIINFEKL
H-2K^b/SIINFEKL
H-2K^b/SIINFEKL MFI
8000
6000
4000
2000
0
****
Untreated
+SIINFEKL

Figure 13 continued

C.



D

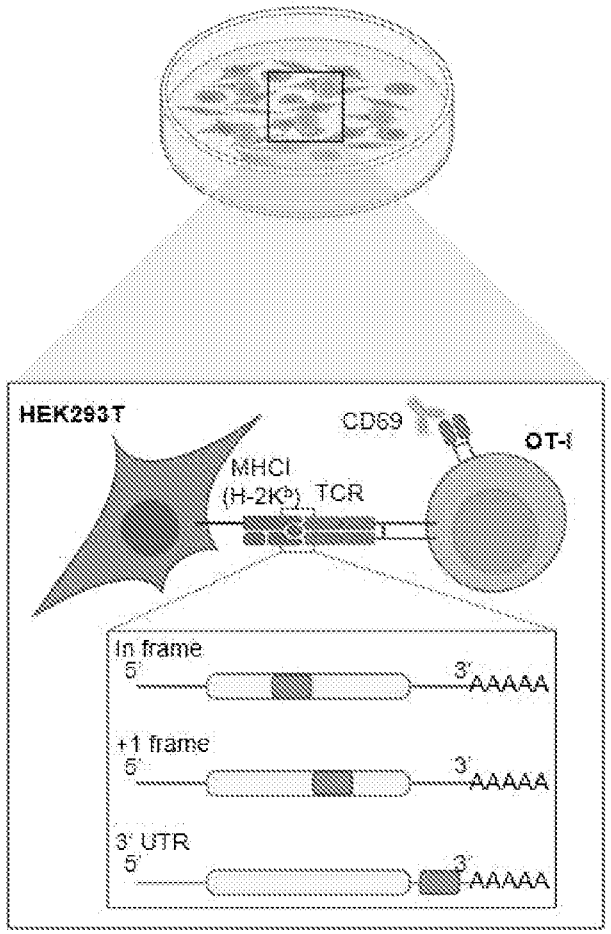


Figure 13 continued

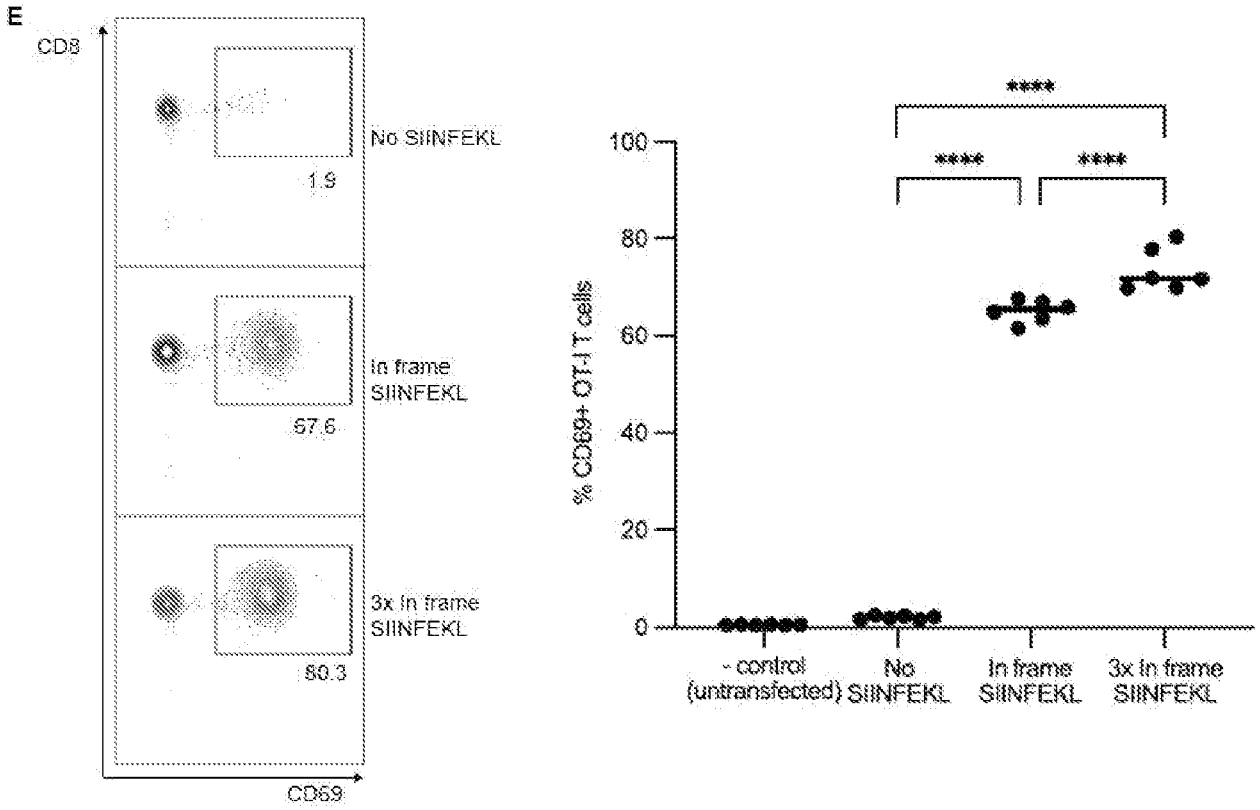


Figure 14

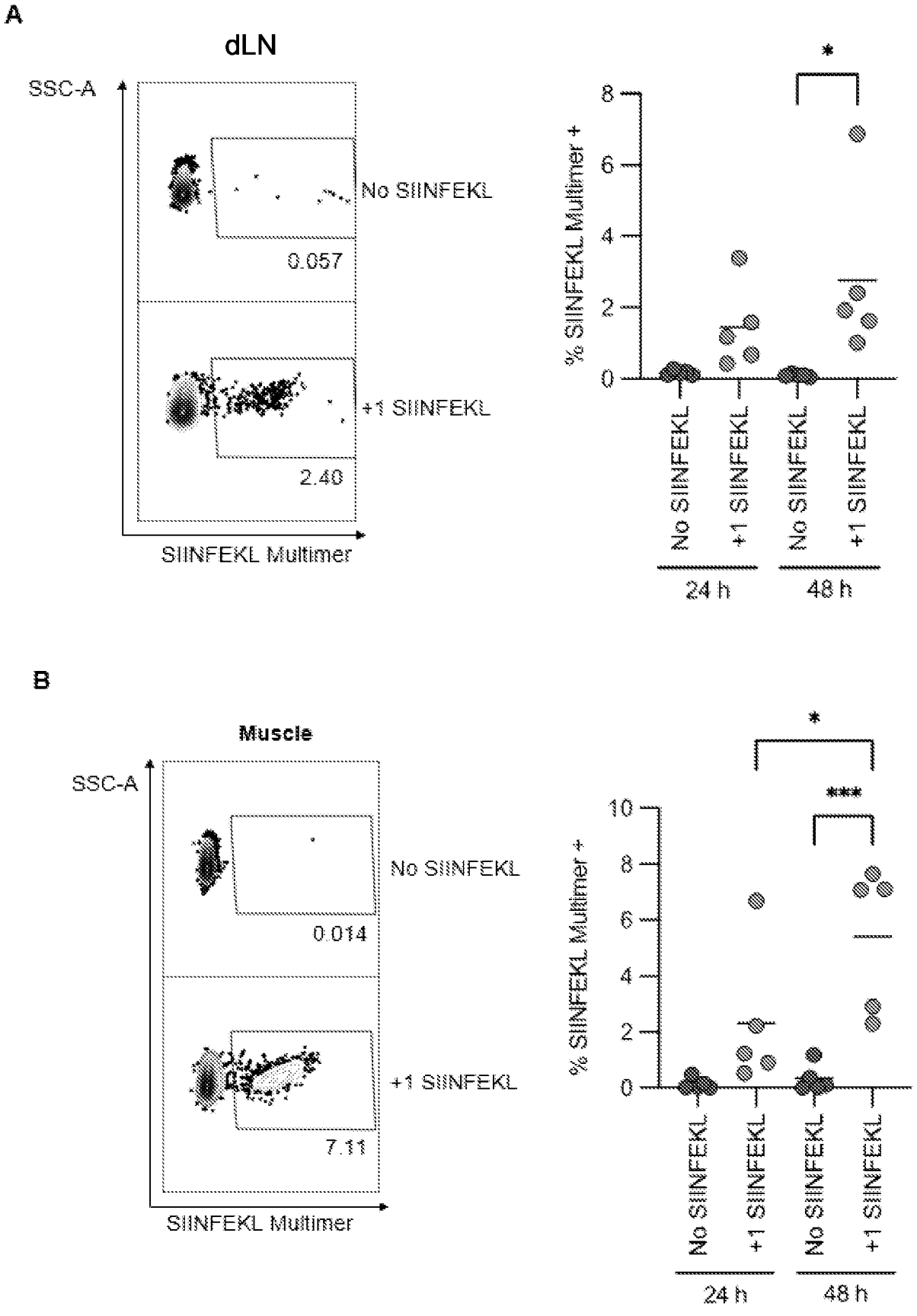

A
dLN
SSC-A
No SIINFEKL
0.057
+1 SIINFEKL
2.40
SIINFEKL Multimer
% SIINFEKL Multimer +
No SIINFEKL
+1 SIINFEKL
No SIINFEKL
+1 SIINFEKL
24 h
48 h
B
Muscle
SSC-A
No SIINFEKL
0.014
+1 SIINFEKL
7.11
SIINFEKL Multimer
% SIINFEKL Multimer +
No SIINFEKL
+1 SIINFEKL
No SIINFEKL
+1 SIINFEKL
24 h
48 h

Figure 14 continued

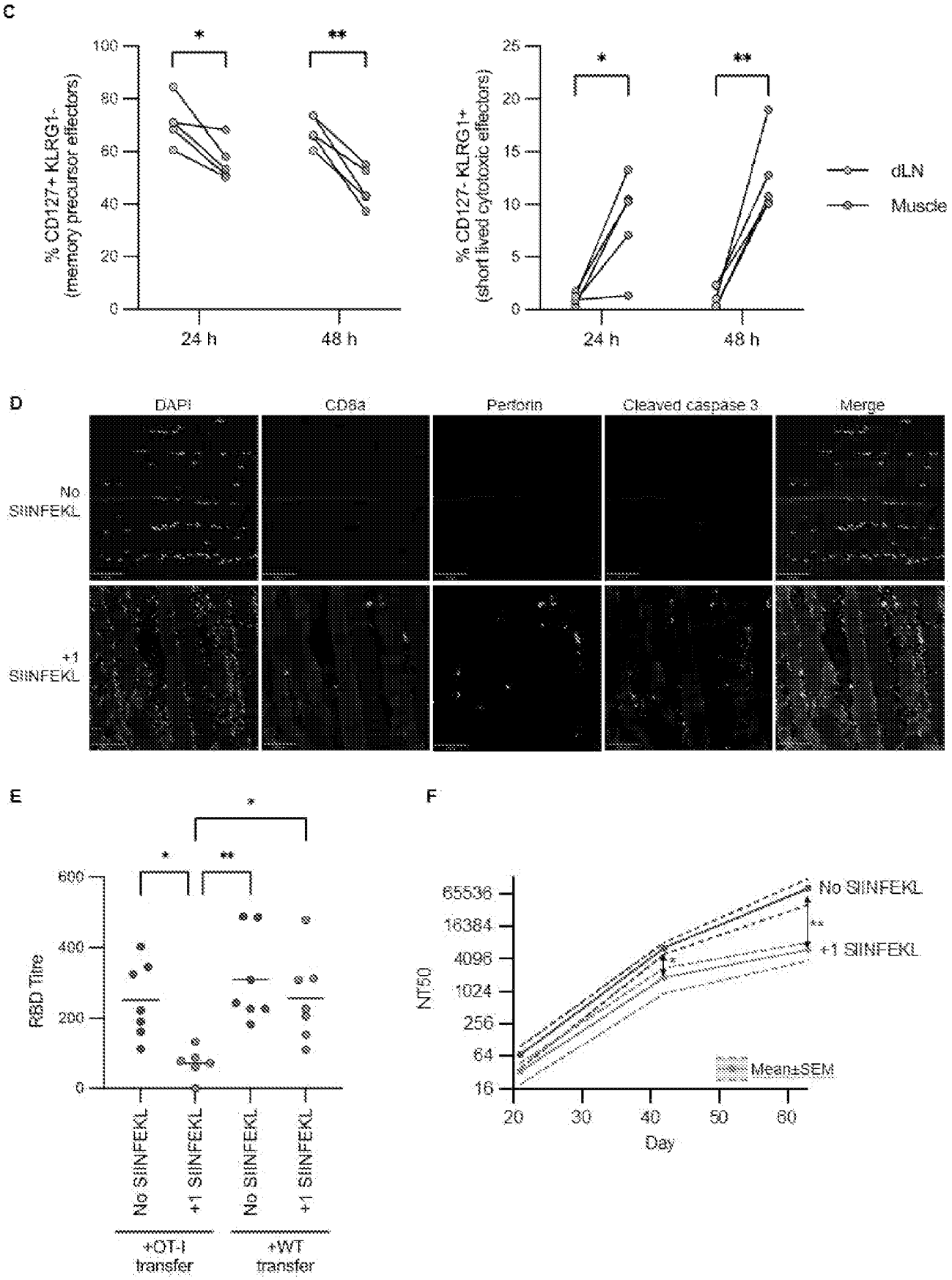


Figure 15

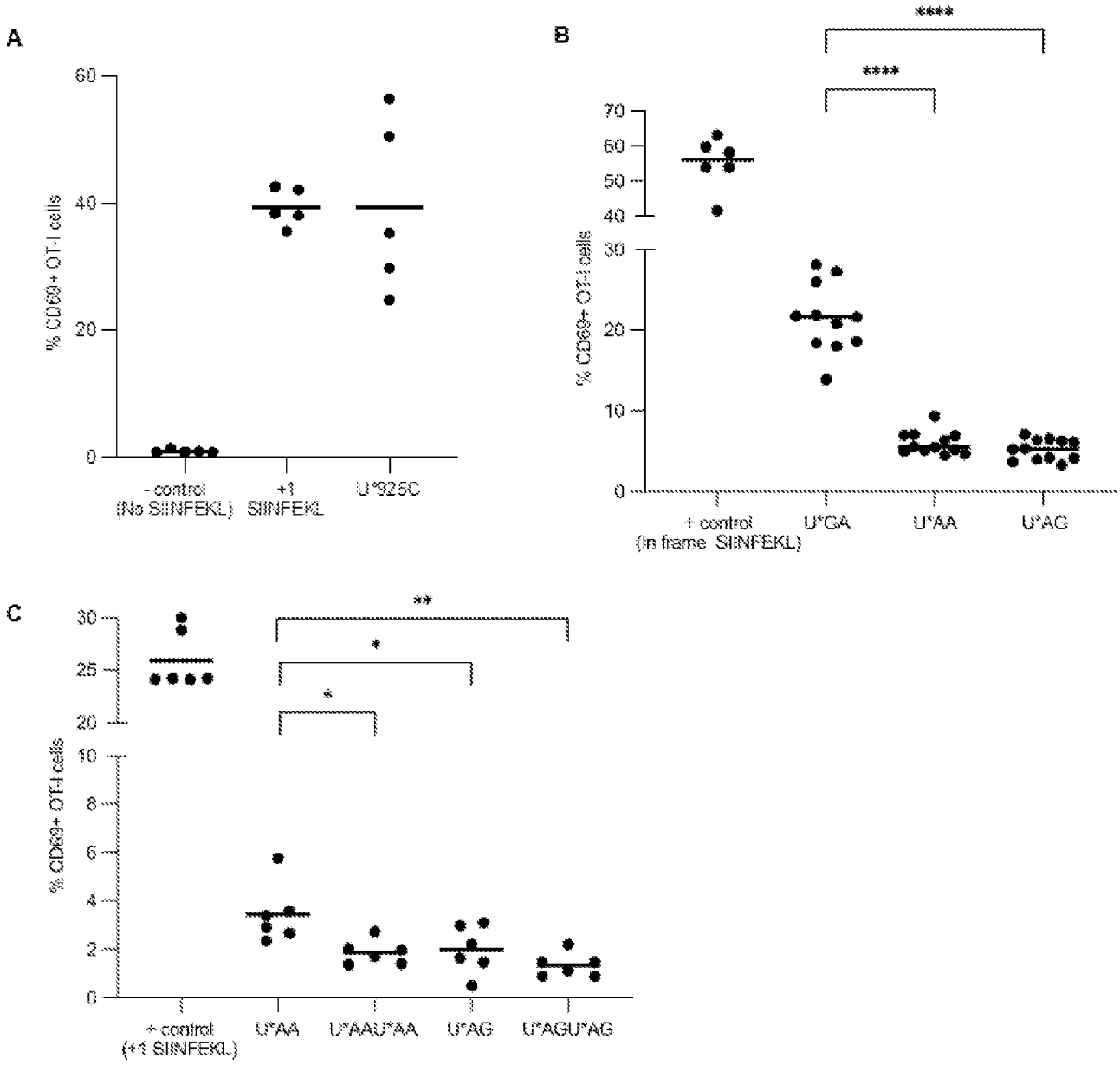

A
% CD69+ OT-I cells
60
40
20
0
- control
(No SIINFEKL)
+1
SIINFEKL
U*925C
B
****
****
% CD69+ OT-I cells
70
60
50
40
30
20
10
0
+ control
(in frame SIINFEKL)
U*GA
U*AA
U*AG
C
**
*
*
% CD69+ OT-I cells
30
25
20
10
8
6
4
2
0
+ control
(+1 SIINFEKL)
U*AA
U*AAU*AA
U*AG
U*AGU*AG

Figure 15 continued

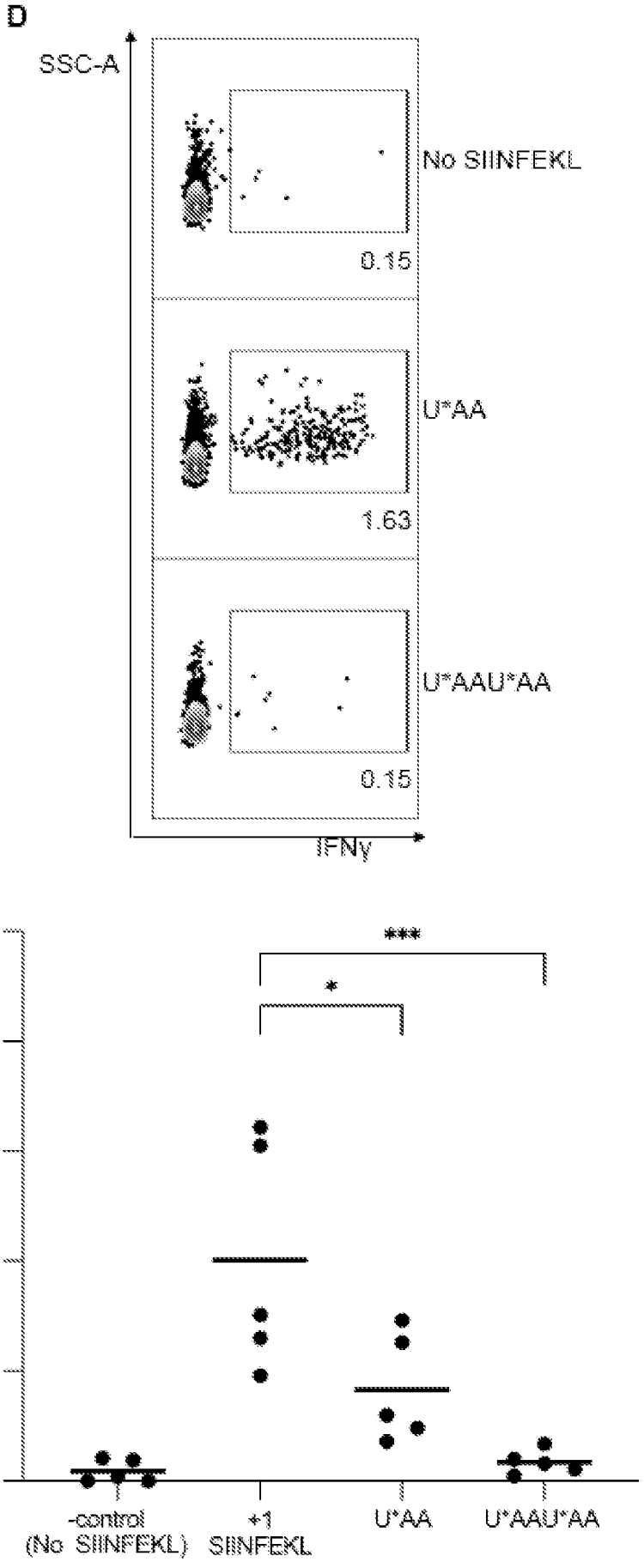
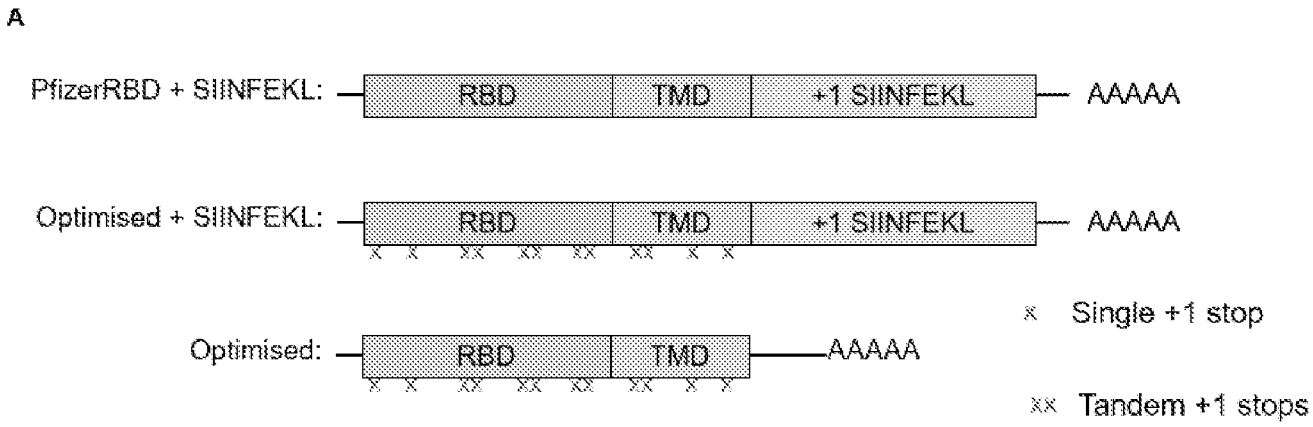


Figure 16


A
PfizerRBD + SIINFEKL:
RBD
TMD
+1 SIINFEKL
AAAAA
Optimised + SIINFEKL:
RBD
TMD
+1 SIINFEKL
AAAAA
Optimised:
RBD
TMD
AAAAA
x Single +1 stop
xx Tandem +1 stops

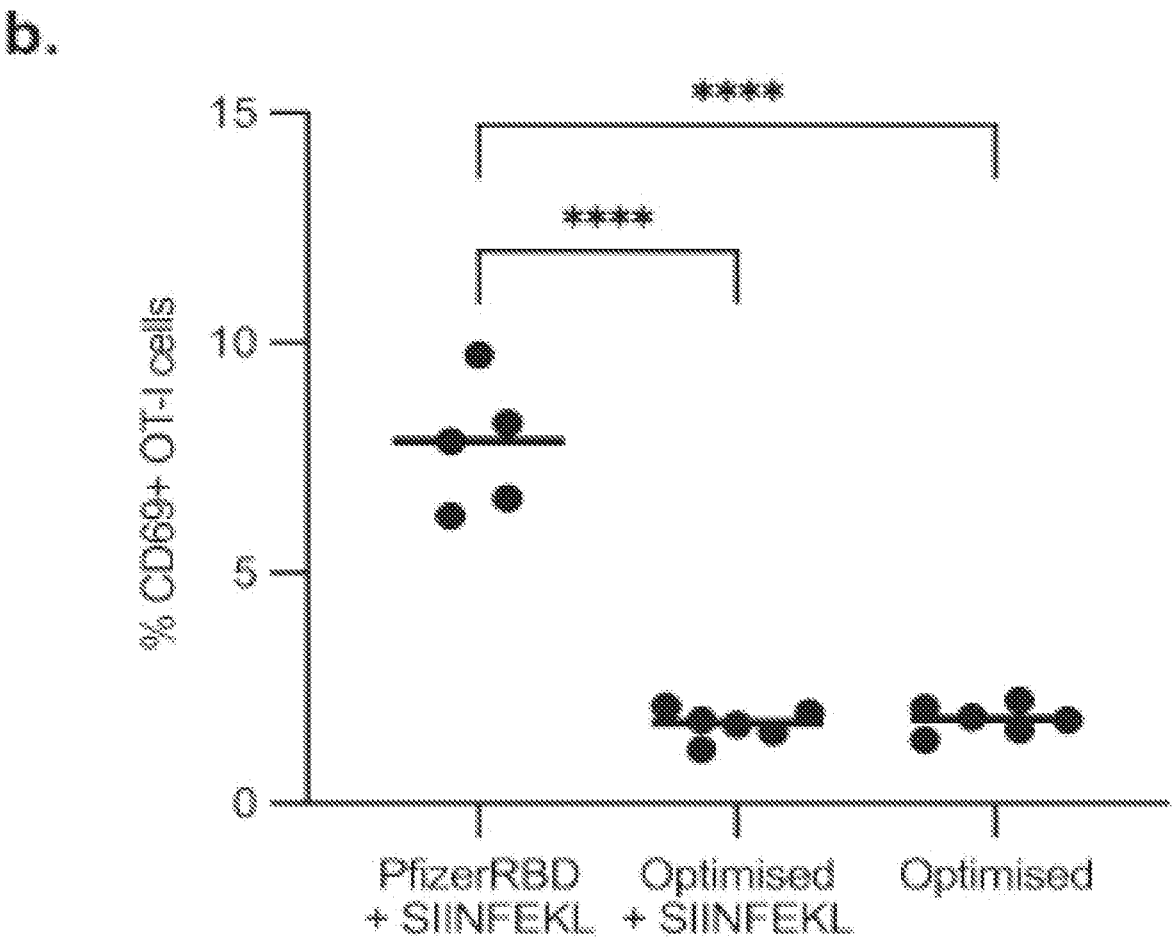
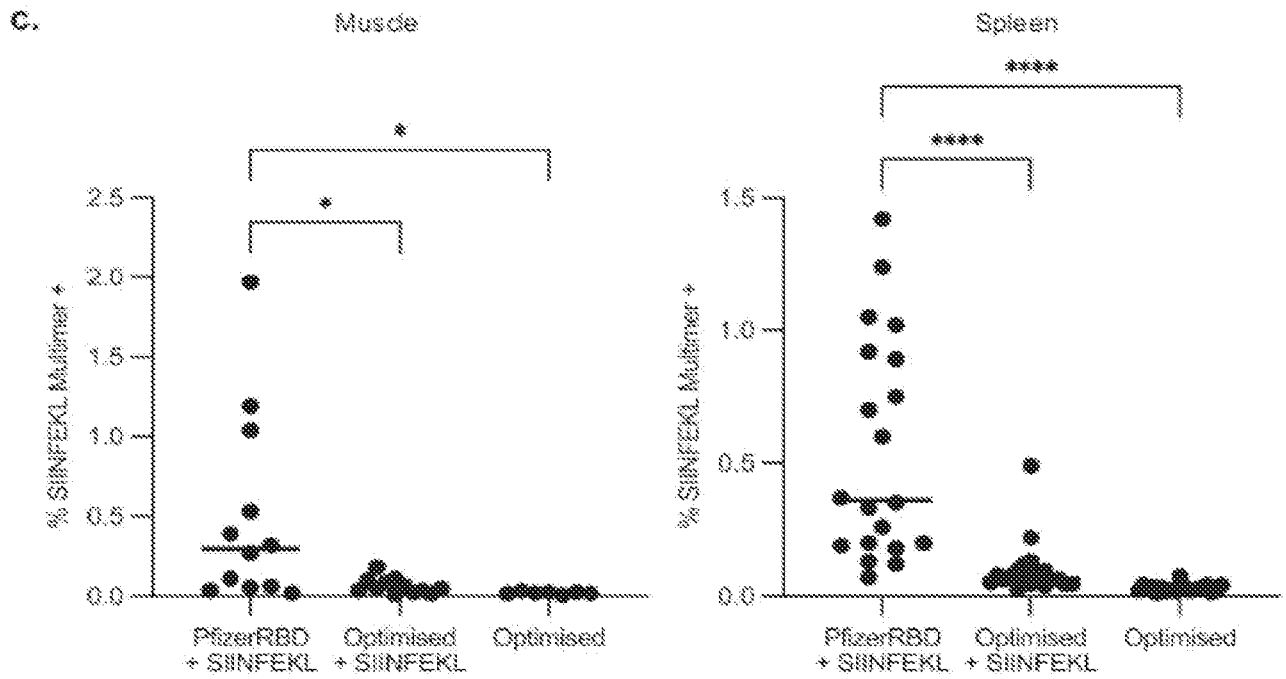

b.
****
****
15
10
5
0
% CD69+ OT-I cells
PfizerRBD + SIINFEKL
Optimised + SIINFEKL
Optimised

Figure 16 continued



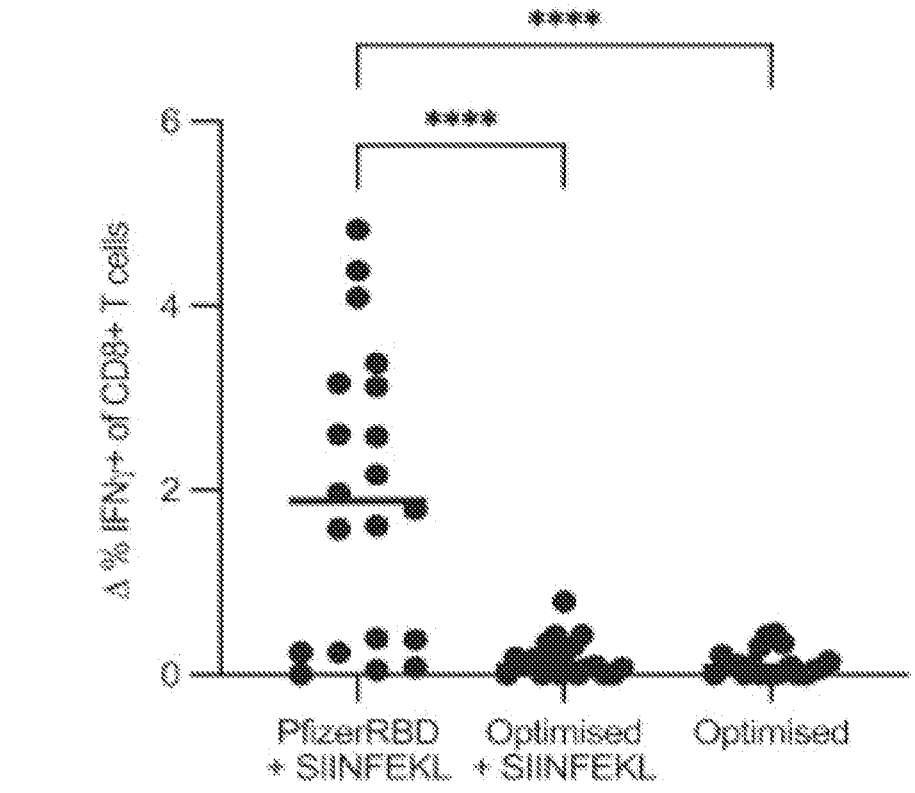


Figure 16 continued

e.

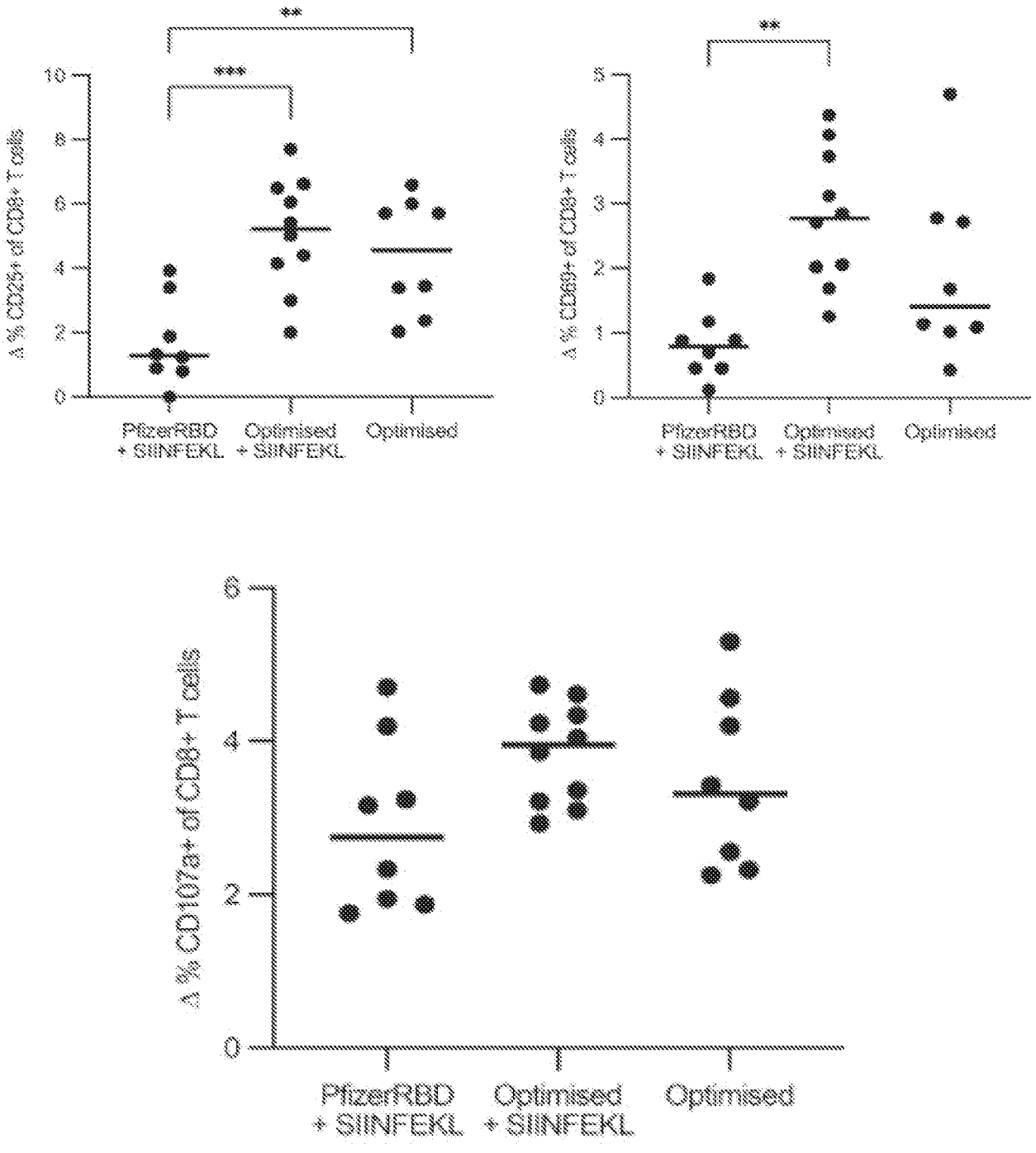


Figure 16 continued

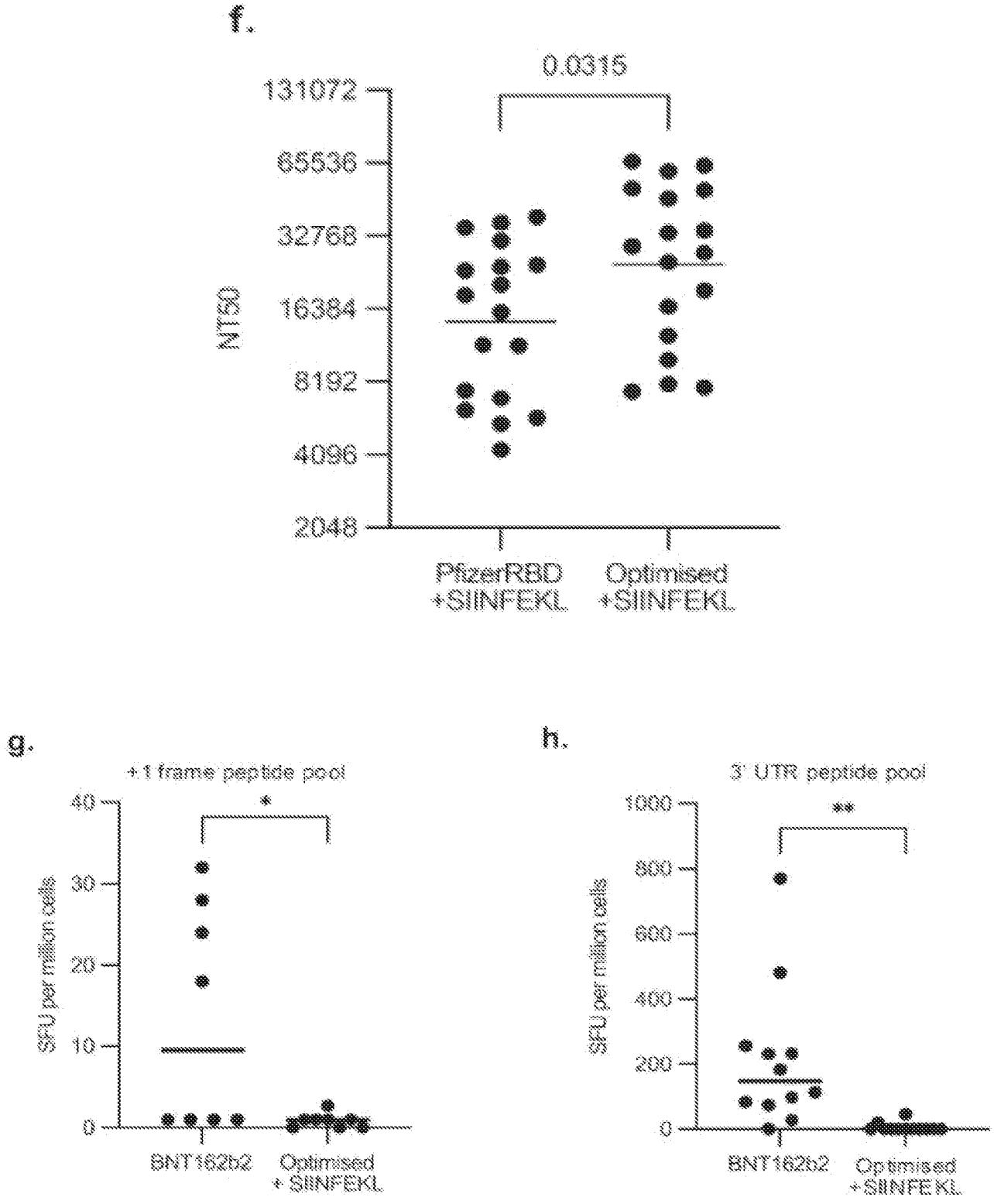


Figure 17

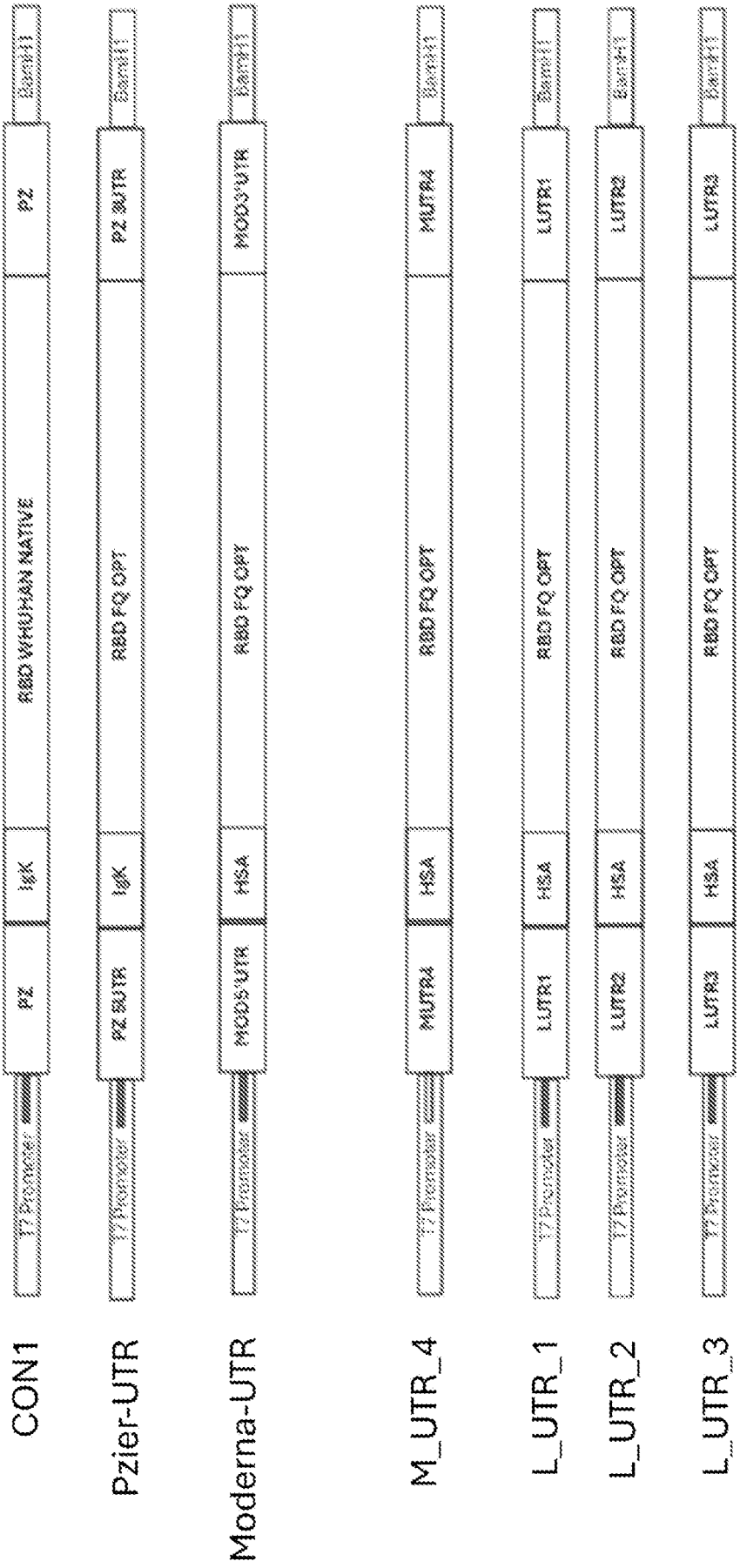

CON1
T7 Promoter
PZ
IgK
RBD WHUHAN NATIVE
PZ
BamH1
Pzier-UTR
T7 Promoter
PZ 5UTR
IgK
RBD FQ OPT
PZ 3UTR
BamH1
Moderna-UTR
T7 Promoter
MOD5'UTR
HSA
RBD FQ OPT
MOD3'UTR
BamH1
M_UTR_4
T7 Promoter
MUTR4
HSA
RBD FQ OPT
MUTR4
BamH1
L_UTR_1
T7 Promoter
LUTR1
HSA
RBD FQ OPT
LUTR1
BamH1
L_UTR_2
T7 Promoter
LUTR2
HSA
RBD FQ OPT
LUTR2
BamH1
L_UTR_3
T7 Promoter
LUTR3
HSA
RBD FQ OPT
LUTR3
BamH1

Figure 18

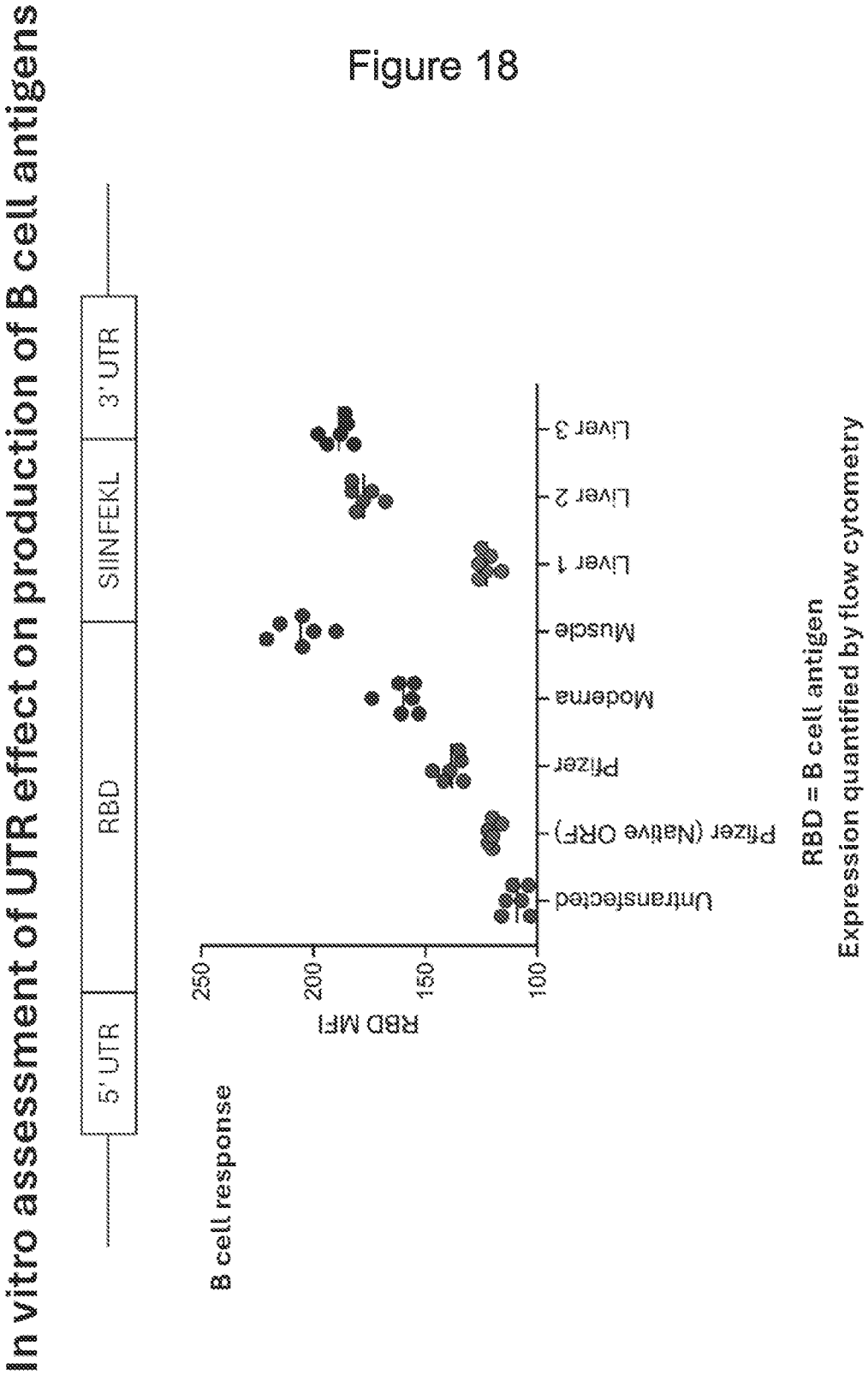
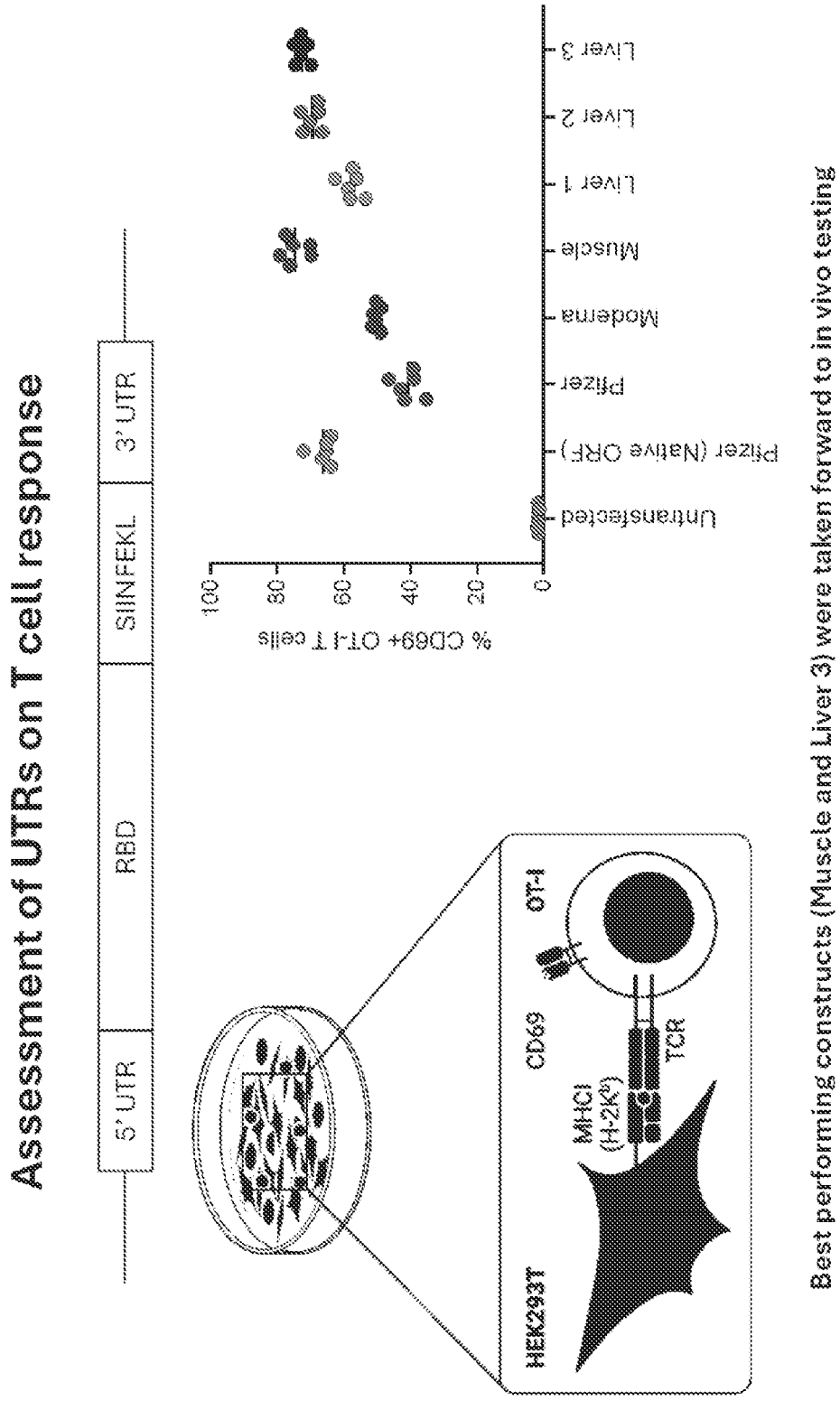

In vitro assessment of UTR effect on production of B cell antigens
5' UTR
RBD
SIINFEKL
3' UTR
B cell response
RBD MFI
250
200
150
100
Untransfected
Pfizer (Native ORF)
Pfizer
Moderna
Muscle
Liver 1
Liver 2
Liver 3
RBD = B cell antigen
Expression quantified by flow cytometry

Figure 19


Assessment of UTRs on T cell response
5' UTR
RBD
SIINFEKL
3' UTR
% CD69+ OT-I T cells
100
80
60
40
20
0
Untransfected
Pfizer (Native ORF)
Pfizer
Moderna
Muscle
Liver 1
Liver 2
Liver 3
HEK293T
MHCI (H-2K^b)
CD69
TCR
OT-I
Best performing constructs (Muscle and Liver 3) were taken forward to in vivo testing