**Charles University**
**Faculty of Science**

Study programme: Biology
Branch of study: Genetics, molecular biology and virology


**Bc. Michaela Kučerová**

# Cysteine tRNA regulates protein synthesis in human cell lines

# Cysteinová tRNA reguluje proteosyntézu v lidských buněčných liniích

Diploma thesis

Supervisor: RNDr. Petra Beznosková, Ph.D.

Prague, 2021

**Prohlášení:**

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 26.04.2021

…….

Bc. Michaela Kučerová

**Poděkování**

Na tomto místě bych chtěla poděkovat své školitelce, RNDr. Petře Beznoskové, Ph.D. za čas mně věnovaný, užitečné rady a vedení mé práce. Též děkuji Dr. rer. nat. Leoši Shivaya Valáškovi, DSc. za možnost pracovat v Laboratoři regulace genové exprese. Celému týmu laboratoře pak děkuji za příjemně strávené chvíle a podporu.

## Abstract

A significant number of known human genetic diseases is associated with nonsense mutations leading to the introduction of a premature termination codon into the coding sequence. A termination codon can be read through by its near-cognate tRNA (tRNA with two anticodon nucleotides base-pairing with a stop codon); potentially generating C-terminally extended protein variants. In yeast, UGA stop codon was described to be read through by tRNA-Trp and tRNA-Cys. Similar was observed for tRNA-Trp in human HEK293T cell line. The aim of this thesis was to investigate if human tRNA-Cys can act as a near-cognate tRNA in human HEK293T cell line. There are two isoacceptors which constitute the tRNA-Cys family, with ACA and GCA anticodon. There are 1 and 23 isodecoders to the ACA and GCA anticodons, respectively. Here, altogether as many as nine tRNA-Cys isodecoders (distinct in their sequence and with varying levels of expression) were tested for their ability to increase UGA readthrough in HEK293T using p2luci and pSGDluc dual-luciferase reporter vectors. In both p2luci and pSGDluc, we observed that at least one tRNA-Cys isodecoder, tRNA-Cys-GCA-4-1, is capable of significantly elevating the UGA readthrough levels when overexpressed in HEK293T. This indicates that similarly to yeast, tRNA-Cys is capable of incorporating at the UGA stop codon *via* readthrough to regulate protein synthesis in human cell lines.

**Keywords:** cysteine tRNA, near-cognate tRNA, proteosynthesis, stop codon, readthrough, Firefly, *Renilla*

## Abstrakt

Velká část lidských genetických chorob je asociovaná s *nonsense* mutacemi vedoucími ke vzniku předčasného terminačního kodonu v kódující sekvenci. Terminační kodon může být pročten pomocí *near-cognate* tRNA (tRNA se dvěma bazemi antikodonu párujícími se stop kodonem), což nazýváme tzv. *readthrough;* toto může vést k tvorbě C-terminálně extendovaných variant proteinu. Schopnost indukovat *readthrough* UGA stop kodonu byla v kvasinkách popsána u tRNA-Trp a tRNA-Cys. V lidské buněčné linii HEK293T byla popsána u lidské tRNA-Trp. Cílem této práce bylo zjistit, zda také lidská tRNA-Cys může fungovat jako *near-cognate* tRNA v lidské buněčné linii HEK293T. Rodina tRNA-Cys je tvořena dvěma isoakceptory, jeden nese antikodon ACA, druhý GCA. Existuje 1 *isodecoder* ACA a 23 *isodecoderů* GCA isoakceptoru. V této práci bylo testováno celkem 9 *isodecoderů* lidské tRNA-Cys (lišících se sekvencí a mírou exprese) pro schopnost pročíst stop kodon UGA v HEK293T za použití p2luci a pSGDluc *dual-luciferase* vektorů. Pozorovali jsme, že v p2luci i pSGDluc je minimálně jeden *isodecoder,* tRNA-Cys-GCA-4-1, schopen zvýšit *readthrough* stop kodonu UGA při zvýšení exprese tohoto *isodecoderu* v HEK293T. Tyto výsledky naznačují, že stejně jako v kvasinkách, je i v lidských buněčných liniích tRNA-Cys schopna inkorporovat se na stop kodon UGA pomocí *readthrough* a regulovat tak proteosyntézu v lidských buněčných liniích.



**Klíčová slova:** cysteinová tRNA, *near-cognate* tRNA, proteosyntéza, stop kodon, *readthrough*, Firefly, *Renilla*

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

**A-site** – aminoacyl tRNA binding site

**AAGs –** aminoglycoside antibiotics

**aa-tRNA –** aminoacyl-tRNA

**aaRS / ARS –** aminoacyl-tRNA synthetase

**Ago** – Argonaute protein

**AQP4** – Aquaporin 4

**ARM-Seq –** AlkB-facilitated RNA methylation sequencing

**ATP10D** – ATPase phospholipid transporting 10D (putative)

**B** – not A (G or C or T)

**BSA** – Bovine serum albumin

**CDH23** – Cadherin Related 23

**CTFV –** Colorado tick fever virus

**CysRS / CARS –** cysteinyl-tRNA synthetase

***D. melanogaster*** – *Drosophila melanogaster*

**DDX58** – DExD/H-Box Helicase 58

**dH$_2$O** – distilled $H_2O$

**DM-tRNA-seq –** demethylase tRNA sequencing

***E. coli*** – *Escherichia coli*

**eIF4G –** eukaryotic initiation factor 4G

**eRF1** – eukaryotic release factor 1

**eRF3 –** eukaryotic release factor 3

**E-site** – exit tRNA binding site

**GlnRS** – glutaminyl-tRNA synthetase

**GtRNAdb** – Genomic tRNA database

**HEPES** – N-(2-Hydroxyethyl) piperazine-N-2-ethan sulphonic acid

**iso-tRNA-CP** - isodecoder-specific tRNA gene contribution profile

**kDa** – kilodalton

**LB** – Luria-Bertani medium

**LDHB –** Lactate dehydrogenase B

**LDHBx –** extended form of LDHB

**m$^7$G –** 7-Methylguanosine

**MAPK10** – Mitogen-Activated Protein Kinase 10

**MCS –** multiple cloning sites

**MDH1 –** Malate dehydrogenase 1

**Mo-MuLV –** Moloney murine leukemia virus

**mRNA** – messenger RNA

**N** – any nucleotide

**nc-tRNA –** near-cognate tRNA

**NMD** – nonsense-mediated mRNA decay

**OPRK1** – Opioid Receptor Kappa 1

**OPRL1** – Opioid related nociceptin receptor 1

**P-site** – peptidyl-tRNA binding site

**PAGE** – poly-acryl amid gel electrophoresis

**PABP** – poly(A) binding protein

**PBS –** phosphate buffer saline

**PMSF** – phenylmethylsulphonyl fluoride

**pre-tRNA –** precursor-tRNA

**PTC –** premature termination codon

**PTS1 –** peroxisomal targeting signal type 1

**R** – Purine (A or G)

**RISC –** RNA induced silencing complex

**RNAse P –** ribonuclease P

**RNase Z –** ribonuclease Z

**RPM –** reads per milion

***S. cerevisiae*** – *Saccharomyces cerevisiae*

**SDS** – sodium dodecyl sulphate

**SIRPB1** – Signal Regulatory Protein Beta 1

**SMA –** spinal muscular atrophy

**TBE** – Tris/Boric Acid/EDTA

**TBS –** Tris Buffered Saline

**TBS-T** – TBS/Tween

**TG –** Tris-Glycine

**tiRNAs –** tRNA-derived stressed-induced fragments

**TMEM86B** – Transmembrane Protein 86B

**TMV –** tobacco mosaic virus

**TOG –** terminal oligoguanine motif

**tRF** – tRNA fragment

**tRFs –** tRNA-derived fragments

**tRNA –** transfer RNA

**TrpRS –** tryptophanyl-tRNA synthetase

**UTR –** untranslated region

**VDR –** Vitamin-D receptor

**VDRx –** extended form of VDR

**VEGF-A –** Vascular endothelial growth factor A

**VEGF-Ax** – extended form of VEGF-A

**WB** – Western blot

**Y** – pyrimidine (C or T)

**YAMAT-Seq –** Y-shaped Adapter-ligated MAture tRNA sequencing

**YB-1 –** Y-box binding protein 1

# 1 INTRODUCTION

It is estimated that a significant number (~ 11 %) of all described gene lesions causing human genetic diseases is generated by a nonsense mutation leading to the introduction of a premature termination codon (PTC) into the coding sequence, resulting in generation of a shorter polypeptide (Mort et al. 2008). This could be potentially solved by promoting translational stop codon readthrough and be thus of the highest importance in clinics, however, no actually effective readthrough-inducing drug exists today.

Naturally, stop codons can be read through by a near-cognate tRNA (with two of the three tRNA anticodon nucleotides complementary to a stop codon) with very low frequency. Its programmed form, which leads to generation of C-terminally extended proteins with potentially novel functions, can lead to much higher efficiencies. Stop codon readthrough (basal readthrough as well as the programmed form) is affected by many factors, including the stop codon context, *cis*-acting elements on mRNA or trans-acting factors. The precise molecular mechanism of how these factors promote readthrough remains unknown. The programmed stop codon readthrough was first described in viruses which use it as an advantageous mechanism to increase the coding capacity of their tiny genomes. It was further described in mammals, including humans. Quite recently, a number of human genes were described to undergo readthrough leading to generation of C-terminally extended protein variants. The function of some of these protein variants was suggested, however, further research is necessary. All of this indicates that the problematics of stop codon readthrough is a rapidly evolving and exciting scientific field of study where a lot remains to be unravelled.

It was shown in our laboratory and by others that in yeast, tRNA-Tyr, tRNA-Trp and tRNA-Cys can incorporate at their near-cognate stop codons. Very recently, Beznoskova et al. 2021 demonstrated that tRNA-Tyr and tRNA-Trp, when overexpressed using a dual-luciferase plasmid construct, promote UAG/UAA and UGA readthrough, respectively, also in human HEK293T cell line and suggested that these two tRNAs effectively read through their near-cognate stop codons when overexpressed. In this thesis, we try to investigate if also human tRNA-Cys can read through its near-cognate UGA stop codon in human HEK293T cell line when overexpressed.

# 2 LITERATURE REVIEW

## 2.1 Introducing (cysteine) tRNA

### 2.1.1 The importance of tRNA molecules (with emphasis on tRNA-Cys)

#### 2.1.1.1 Multiple roles of tRNA

tRNA molecules are non-coding RNA molecules, usually about 73 - 93 nucleotides in length, charged with specific amino acid on their 3'OH end (Rich and Rajbhandary 1976). tRNAs are the crucial elements in the elongation phase of translation: a ternary complex consisting of a cognate elongating aminoacyl-tRNA (aa-tRNA), eIF1A and GTP is delivered to the ribosome A-site. After codon recognition by aa-tRNA, GTP is hydrolysed by eIF1A and eEF1A-GDP is released which allows the aa-tRNA to be accommodated in the A-site. The eEF1A-GDP is recycled to eEF1A-GTP by eEF1B exchange factor. A peptidyl transferase centre of the ribosome then positions the aa-tRNA and peptidyl-tRNA for catalysis. The peptide bond formation is believed to be promoted by eIF5A favourably positioning the two substrates. The growing peptide is transferred from the peptidyl-tRNA in the P-site to the new aa-tRNA in the A-site and a new, one amino acid longer peptidyl-tRNA is made. During peptide bond catalysis, these tRNAs are transferred into hybrid P/E and A/P states, where the acceptor ends are in the E- and P-sites and the anticodon loops in the P- and A-sites. Translocation of the new peptidyl-tRNA and the deacetylated tRNA to the P-and E-sites, respectively, is promoted by binding of the eEF2-GTP complex and followed by eEF2-GDP release. The deacetylated tRNA is released and a new cycle of elongation begins when next eEF1A-GTP-aa-tRNA complex binds the A-site. The mechanism is depicted in Fig. 1 and reviewed in detail in (Dever and Green 2012, Dever, Dinman and Green 2018).

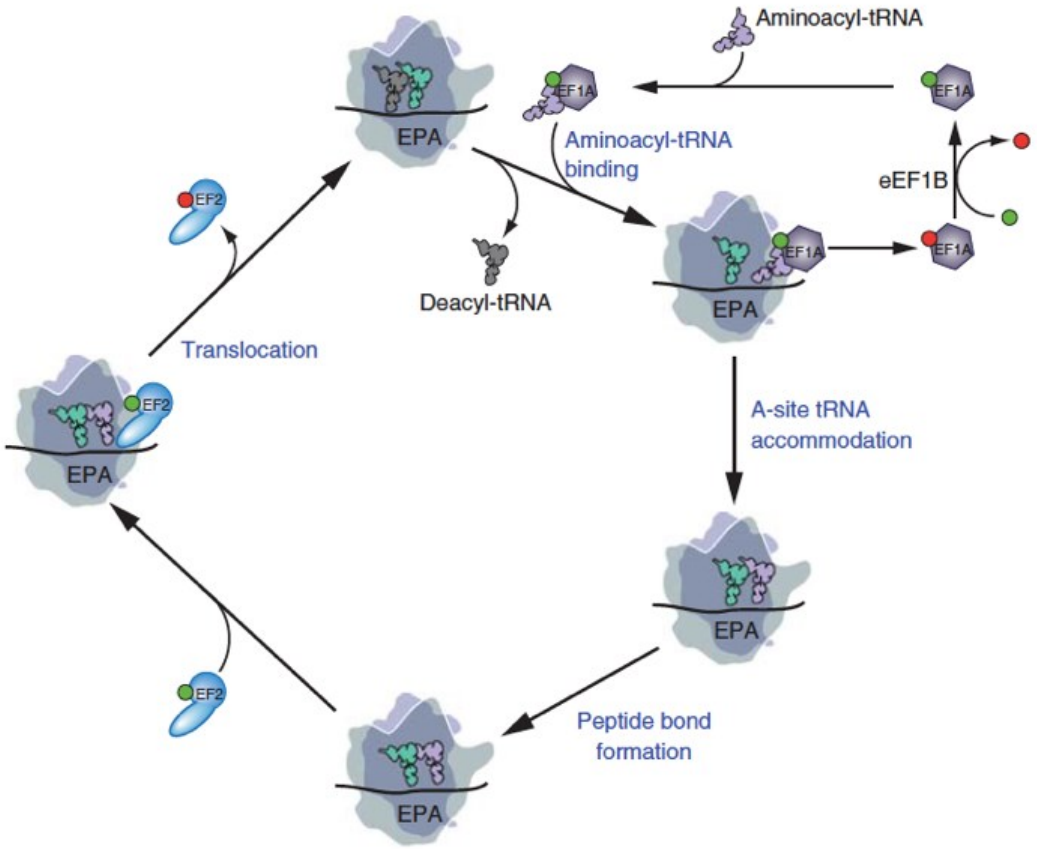



**Figure 1. Eukaryotic translation elongation.** Aminoacyl-tRNA (aa-tRNA) is delivered to the ribosome A-site in a ternary complex consisting of aa-tRNA/eEF1A-GTP. After anticodon-codon binding, GTP is hydrolysed by eIF1A, eEF1A-GDP complex is released and aa-tRNA is accommodated in the A-site. The eEF1A-GDP is recycled to eEF1A-GTP by eEF1B exchange factor. Peptide bond is formed and translocation of the new peptidyl-tRNA and the deacetylated tRNA is mediated by eIF2-GTP. eEF2-GDP is released, and new cycle begins when new ternary complex enters the A-site (Dever and Green 2012).

Apart from their canonical role in proteosynthesis, tRNAs have a wide range of noncanonical functions such as global gene expression modulation, apoptosis modulation, amino acid transfer during posttranslational protein modification, priming of retrotransposon and retroviruses reverse transcription, are involved in many diseases' onset and severity, have a role in a cell stress response, *etc*. These regulatory functions are reviewed in detail in (Avcilar-Kucukgoze and Kashina 2020).

Among those, the main focus has been drawn to tRNA fragments generated by cleavage of tRNA molecules. These fragments can be divided into two groups. The first group are tRNA halves (also called tRNA-derived stressed-induced fragments, tiRNAs), 5' tiRNAs

and 3’ tiRNAs, which are about 31-40 nucleotides long, begin at the 5’or 3’ end of the tRNA molecule and are cleaved in anticodon (Kumar, Kuscu and Dutta 2016, Avcilar-Kucukgoze and Kashina 2020). Cleavage of tRNA is performed by stress-activated ribonuclease angiogenin (Yamasaki et al. 2009). These can have a role in translational inhibition. It was suggested that transfection of 5’ tiRNAs (and not 3’tiRNAs) extracted from angiogenin-treated U2OS cells into U2OS cell line caused global proteosynthesis levels reduction by about 20 % (Yamasaki et al. 2009). Furthermore, it was shown with synthetic 5’tiRNAs derived from tRNA-Cys and tRNA-Ala that these especially can mediate significant reduction in translation levels. Both tRNA-Ala and tRNA-Cys have a terminal oligoguanine (TOG) motif at their 5’ends (with four to five guanine residues) (Ivanov et al. 2011). These guanine residues were found to form G-quadruplex (G4) structures required for translation inhibition (Ivanov et al. 2014). tRNA-Ala and tRNA-Cys are the only two tRNAs which possess the 5’TOG and its presence is conserved among many living organisms; moreover, these motifs are very rare in other tRNAs. Mutational analysis shown that at least four guanines at the 5’end of tiRNAs are necessary for inhibition of translation. The translational initiation inhibition was explained by interaction with the translational repressor Y-box binding protein 1 (YB-1) which leads to displacement of the cap-binding complex eIF4F from the $m^7G$ cap structures on mRNAs and stress granules formation (Ivanov et al. 2011).

The second group are tRNA-derived fragments (tRFs), which are about 14-30 nucleotides long, similar in size to microRNA, tRF-3 and tRF-5, generated from 3’ and 5’ ends of mature tRNA (cleaved at the T-loop or D-loop, respectively) and tRF-1 generated from 3’ends of the primary transcript tRNA. tRF-3s and tRF-5s can be divided into subgroups according to their length. Another type are tRF-2s, which contain only the anticodon stem and loop. tRNA fragments are present in all domains of life. Although it has been suggested that the tRFs could be generated by Dicer endonuclease, further data opposed, showing that Dicer depletion does not prevent from tRFs formation and thus, it may be possible that tRFs can be generated by some unknown enzyme or multiple enzymes specifically, however, further research is necessary. The detailed information about tRNA fragments is reviewed in (Kumar et al. 2016, Avcilar-Kucukgoze and Kashina 2020).

It was suggested that tRFs may function similarly to miRNA. It was shown that some tRFs associate with human Argonaute (Ago) proteins 1, 3, 4 and not 2 (although Ago 2 is the

main human “silencing-related” one) and use their 5’ seed sequences to associate with target mRNA (Kumar et al. 2014). Moreover, it was shown using HEK293T cells that tRF-3s generated from tRNA-Cys and tRNA-Leu downregulate the expression of luciferase reporters with a complementary sequence in the 3’UTR (3’ untranslated region) by about 60 %, and this repression is dependent on Ago protein and independent on Dicer or Drosha. It was also shown that complex of tRF-3 and target mRNA associates with GW182 proteins in Ago-containing RISC (RNA-induced silencing complex), which are necessary for miRNA-induced translational repression. tRFs could thus be able to enter RISC and mediate translational repression (Kuscu et al. 2018).

#### 2.1.1.2 The sequence and the structure of tRNAs

The tRNA genes are transcribed by RNA Polymerase III. The tRNA molecule itself comprises internal promoter type II, with sequence boxes named A Box and B Box (Paule and White 2000, Geiduschek and Tocchinivalentini 1988). The sequence for human A Box and B Box can be written as **TRGYNNARNNG** and R**GTTC**R**A**NN**C**Y, respectively, where N means any nucleotide, R means purine (G/A) and Y means pyrimidine (C/T) (conserved nucleotides are highlighted in bold) (Hamada et al. 2001, Canella et al. 2010). Some nucleotides within the A Box and the B Box are highly conserved because of the tRNA tertiary folding structure, some are variable. This variability within the A Box and B Box between human isodecoders may be the cause of different tRNA isodecoders‘ expression rates in tissues (Goodenbour and Pan 2006).

A transcribed eukaryotic precursor-tRNA (pre-tRNA) molecule undergoes maturation, where the 5’end is processed by ribonuclease P (RNAse P) removing the 5’ leader sequence; the 3’end is processed by ribonuclease Z (RNase Z) removing the 3’ trailer sequence immediately following the discriminator base ($N_{73}$) and a short 5’CCA sequence required for aminoacylation is added by tRNA nucleotidyl transferase. tRNAs containing introns are then spliced. Subsequently, tRNAs are posttranscriptionally modified and specifically aminoacylated by their proper aminoacyl-tRNA synthetases. For detailed review on this topic please see (Hopper and Phizicky 2003, Phizicky and Hopper 2010).

The structure of mature tRNA is very specific, secondary structure (Fig. 2A) termed as a “cloverleaf”, tertiary structure (Fig. 2B) termed as an “L-shape”. The secondary

"cloverleaf " structure of tRNA molecule was first suggested in 1965 on yeast alanine tRNA, showing the sequence folds into four base-paired stem regions connected with loop regions (Holley et al. 1965). These regions were further named as follows: acceptor stem, TΨC stem and TΨC loop (Ψ being pseudouridine, T-loop), variable loop (variable stem only if long variable arm is present), anticodon stem and anticodon loop and dihydrouridine stem and dihydrouridine loop (D-loop). A stem and a loop are usually referred to as an "arm" (Rich and Rajbhandary 1976). The tertiary "L-shape" was first described in 1973 on yeast phenylalanine tRNA, which revealed that the acceptor arm and the T arm form one "side" of the "L-shape", whilst the D- arm and anticodon arm form the other one (Kim et al. 1973).

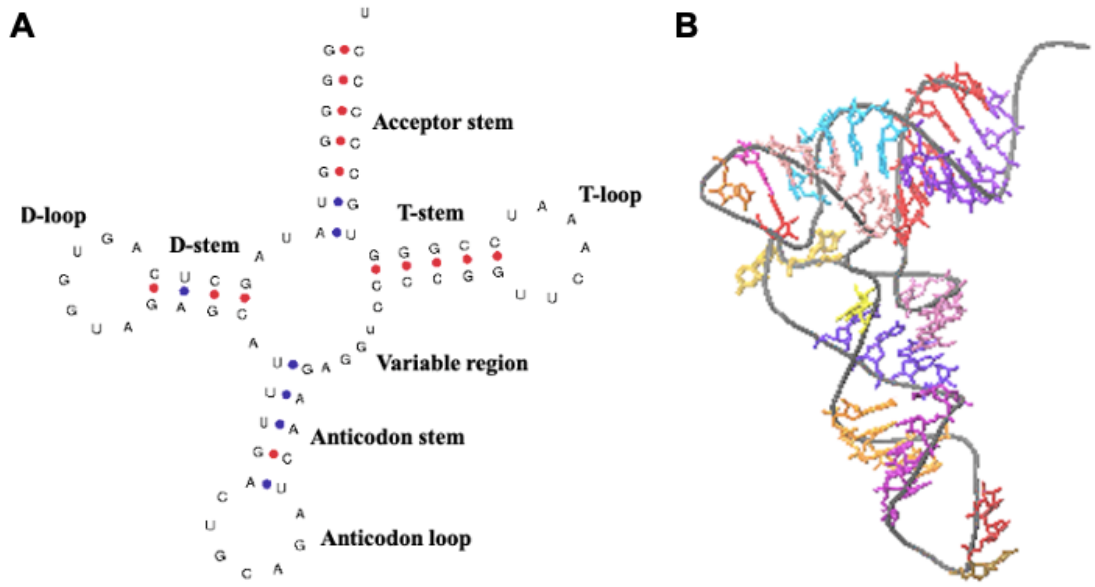



**Figure 2. The sequence, the secondary structure prediction (A) and the tertiary structure prediction (B) of human tRNA-Cys-GCA-2-1.** The sequence and the secondary structure prediction were downloaded from http://gtrnadb.ucsc.edu. The model of the tertiary structure prediction was downloaded from http://lowelab.ucsc.edu/tRNAscan-SE/.

tRNA molecules are usually highly modified, each tRNA bearing approximately 13 modifications. These modifications are required for maintaining the molecule structural stability and function *in vivo*. Some modifications are directly connected with decoding, those are modifications in the anticodon loop, mainly modified position 34 (anticodon position 1) and position 37 (immediately after anticodon). Other modifications in the tRNA

body can have many other functions, mainly related to the molecule structural stability (Pan 2018).

The described tRNA modifications are gathered in the MODOMICS database, however, no human cysteine tRNA modifications are present in this database so far. No records are also present for human cytosolic tRNA-Asp, tRNA-Ile, tRNA-Lys, tRNA-Pro and tRNA-Trp (Machnicka et al. 2013, Boccaletto et al. 2018). However, some modifications for tRNA-Cys were suggested by novel high-throughput sequencing methods, namely $m^1A58$ and $m^1G37$ modifications for both tRNA-Cys-GCA-2-2 and tRNA-Cys-GCA-4-1 isodecoders, using demethylase tRNA sequencing (DM-tRNA-seq) in HEK293T cells (Clark et al. 2016). $m^1A58$ modification was also described before for human tRNA-Cys-GCA fragments using ARM-Seq (AlkB-facilitated RNA methylation sequencing) in B cell-derived human cell lines, GM05372 and GM12878 (Cozen et al. 2015).

#### 2.1.1.3 Aminoacylation of (cysteine) tRNA

Aminoacylation of tRNA is one of the most important steps of proteosynthesis. The esterification of a correct amino acid to the tRNA is catalyzed by its cognate aminoacyl-tRNA synthetase (aaRS or ARS). Aminoacyl-tRNA synthetases are named according to the tRNA they charge, e.g., cysteinyl-tRNA synthetase (CysRS or CARS) charges tRNA-Cys with cysteine. Altogether 23 aaRSs have been described so far, which can be divided into two classes and several subclasses. CysRS belongs to the subclass Ia, charging with thiolated amino acids (Gomez and Ibba 2020). The recognition of amino acid is initiated by electrostatic interactions between phosphate backbone of the tRNA and positively charged protein residues of aaRS; the tRNA then shows a conformational change in the catalytic site of the aaRS. Kinetic determinantion based on $K_{cat}$ is then used to eliminate non-cognate tRNA and finally, the correct tRNA isoacceptor is chosen based on identity determinants, which can be sequence or structural (Gomez and Ibba 2020). For example, in case of AlaRS, the only identity determinant is a single G3:U70 base pair of tRNA-Ala (Hou and Schimmel 1988).

In case of tRNA-Cys, the GCA anticodon and position U73 are considered the general identity determinants for aminoacylation (Pallanck, Li and Schulman 1992). These are also universally conserved in evolution as shown for *E. coli*, yeast and human tRNA-Cys.

The evolution of tRNA identity determinants and coevolution of particular synthetase domains are demonstrated by the evolution of G15:C48 pair in yeast (differing from G15:G48 in *E. coli*), and presence of G15:C48 pair and G37 in anticodon loop in humans (differing from A37 in *E. coli* and yeast). tRNAs usually have R15:Y48 (purine:pyrimidine) tertiary base pair, called the Levitt pair (Levitt 1969). In *E. coli*, the non-conserved G15:G48 Levitt pair is required for correct aminoacylation and is not present in either yeast or humans (Lipman and Hou 1998, Hou, Westhof and Giege 1993).

It is largely important to prevent mischarging and even though the incorporation of tRNA is highly specific, some aaRSs (10 of the 23 described) have evolved editing mechanisms. CysRS does not possess such an activity (Gomez and Ibba 2020). It does not actually require an editing activity, because it prevents mischarging by non-cognate serine tRNA sufficiently by presence of a zinc ion in the active site of CysRS. The cysteine molecule forms there a specific zinc-thiolate interaction with nearly ideal trigonal bipyramidal geometry of which the serine molecule is not capable (Zhang et al. 2003, Newberry, Hou and Perona 2002).

#### 2.1.1.4 tRNA isodecoders and their variability

The genetic code is degenerate, which means that multiple codons exist which code for the same amino acid. There are 61 possible sense codons (and 3 stop codons) according to the genetic code, however, in human, only 49 tRNA isoacceptors are present (Goodenbour and Pan 2006); the remaining codons are decoded by wobble base pairing on codon position 3 (described first in (Crick 1966). The isoacceptors are tRNA molecules having different anticodon but charged with the same amino acid. Altogether 21 isoacceptor families exist, one for each amino acid and one for selenocysteine. For each single isoacceptor, multiple tRNAs differing in their sequence can exist, which are called the isodecoders. Altogether 274 different tRNA isodecoder species are produced from 446 human tRNA genes (Goodenbour and Pan 2006).

As of this writing, according to GtRNAdb 2.0 (in the hg38 reference human genome), there are 29 genes leading to production of cysteine tRNAs with mature score > 50. Two isoacceptors represent the tRNA-Cys family – with ACA and GCA anticodon, of which, there exist 1 and 23 isodecoders, respectively. Two isodecoder species of these are

encoded by four genes, namely tRNA-Cys-GCA-2 isodecoder family (tRNA-Cys-GCA-2-1, tRNA-Cys-GCA-2-2, tRNA-Cys-GCA-2-3 and tRNA-Cys-GCA-2-4) and tRNA-Cys-GCA-9 isodecoder family (similarly 1-4). The ACA isoacceptor, and in this case also the only isodecoder, has an intron inside the sequence and is classified as pseudogene in GtRNAdb 2.0 nowadays.

There are only minor nucleotide changes among the multiple isodecoders of the same isoacceptor type but these could account for different expression rates in human tissues or developmental stages (Goodenbour and Pan 2006). It was shown using ChIP-Seq Pol III that the Pol III occupancy on tRNA genes varies among different cell types in strength and location (mouse, rat, human, macaque, dog and opossum livers) which showed that the isodecoder expression may depend on a cell type (Kutter et al. 2011).

For instance, in mouse there are five isodecoders of tRNA-Arg-UCU family, with a nervous system specific isodecoder gene Tr20, which differs in sequence and has no introns (whilst the other isodecoders do have intron). It was shown that expression of this isodecoder is central nervous system specific (60 % of all expression of this isodecoder family is in brain), whereas all four of the other isodecoders are expressed in all tissues (Pan 2018, Ishimura et al. 2014).

It was also demonstrated that different tRNA isodecoders show different ability to suppress UAG stop codon. Altogether 31 isodecoders from 12 isoacceptor families of serine, leucine and alanine tRNAs (4, 5, 3 isoacceptors; 21, 7 and 10 isodecoders, respectively) were changed to all bear the CUA anticodon. These tRNAs were selected since the determinant for recognition of aminoacyl-tRNA synthetase and aminoacylation is located at different positions of the tRNA than anticodon (long-variable arm in case of leucine and serine tRNA, G3:U70 base-pair in the acceptor stem for alanine tRNA). *In vitro* transcribed tRNA isoforms were co-transfected into HeLa cells with plasmid DNA encoding a GFP mutant containing single UAG stop codon. The tRNA isoforms were tested for their suppression activity. A large difference in suppression activity was observed; some molecules shown near background levels of suppression, some shown 20-fold increase above background (although all isodecoders were equally stable and shown similar aminoacylation levels *in vivo*), which showed that the suppression ability of different isodecoders may differ significantly. In this study, the long variable arm was described to be important for

suppression – in case of tRNA-Ser, when expanded by 2-bp or 3-bp insertion mutation, the suppression increased significantly (Geslain and Pan 2010).

To summarize, all of this indicates that the expression of isodecoders is variable and some isodecoders may potentially have other roles apart from the canonical codon-recognition and amino-acid-incorporation role in translation.

## 2.2 Introducing the phenomenon of translational readthrough

### 2.2.1 Translational termination

Translation, a process of proteosynthesis, comprises four stages: a) *initiation*, where the ribosome is assembled at the mRNA AUG initiation codon with the initiator tRNA bound to the P-site of the ribosome b) *elongation*, where (cognate) aminoacyl-tRNAs enter the ribosome's A-site, a peptide bond is formed after correct decoding and a polypeptide is continuously made, c) *termination*, where a stop codon is recognized by specific release factors and release of the nascent polypeptide from the ribosome is triggered, d) *recycling*, where the used ribosomal subunits are dissociated and the mRNA and deacetylated tRNA are released to be used again in another round of translation. The mechanism of translation is reviewed in detail in (Kapp and Lorsch 2004).

In eukaryotes, two classes of release factors exist. Single class 1 factor, eRF1, which decodes all three stop codons UGA, UAG, UAA and triggers the release of the nascent peptide. Single class 2 release factor exists, a GTPase protein eRF3 (Kapp and Lorsch 2004). It was shown on the crystal structure of eRF1 that the incorporation of eRF1 to ribosome's A-site happens by molecular mimicry; the eRF1 molecule mimics the shape of the tRNA molecule, and thus can be incorporated to the A-site of the ribosome efficiently (Song et al. 2000).

The mechanism of eukaryotic termination of translation is depicted in Fig. 3. First, a ternary complex consisting of eRF1/eRF3-GTP enters the A-site of the ribosome and a stop codon is recognized by eRF1 (Fig. 3B). GTP bound by eRF3 is then hydrolysed (Fig. 3C). After GTP hydrolysis, eRF1 extends its conformation so that its M domain's catalytic GGQ motif

accommodates in the peptidyl-transferase centre close to the CCA end of the peptidyl-tRNA in the ribosome P-site and the peptide release is triggered (Fig. 3D), probably analogously to bacterial machinery, by nucleophilic attack of water molecule to the peptidyl-tRNA ester bond. Both eRF1 and eRF3 support each other – eRF1 enhances GTP binding to eRF3 by being a GTP dissociation inhibitor; stop codon recognition by eRF1 accelerates GTP hydrolysis by eRF3; and eRF3 enhances the peptide release activity of eRF1. The detailed mechanism of eukaryotic translation termination is reviewed in (Hellen 2018).

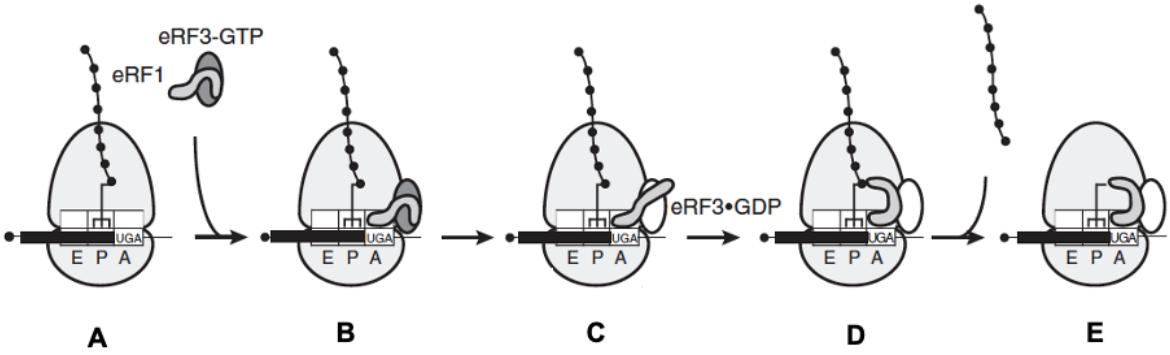



**Figure 3. Eukaryotic translation termination. A)** In the pre-termination complex there is peptidyl-tRNA in the P-site and stop codon in the A-site of the ribosome. **B)** A ternary complex eRF1/eRF3-GTP enters the A-site and the stop codon is recognized by eRF1. **C)** GTP is hydrolysed by eRF3 and **D)** eRF1 accommodates in the peptidyl-transferase centre and induces peptide release. **E)** After proper termination**,** eRF1 and (possibly) eRF3-GDP stay associated with the post-termination complex (Hellen 2018).

### 2.2.2 The importance of translational stop codon readthrough

Translational stop codon readthrough is a naturally occurring situation on ribosome that can arise when recognition of a stop codon by a release factor (and thus proper termination) competes with incorporation of a near-cognate tRNA (tRNA with two of the three anticodon nucleotides complementary to a stop codon sequence) in the A-site of the ribosome. When readthrough occurs, the stop codon is "redefined as sense", an amino acid is incorporated and translation on mRNA continues until the next stop codon in the same open reading frame is reached, which leads to production of C-terminally extended proteins. The mechanism is depicted in Fig. 4 (Dabrowski, Bukowy-Bieryllo and

Zietkiewicz 2015). The natural frequency of translational readthrough (also termed basal readthrough) is quite low, < 0.1 % (Keeling et al. 2012).

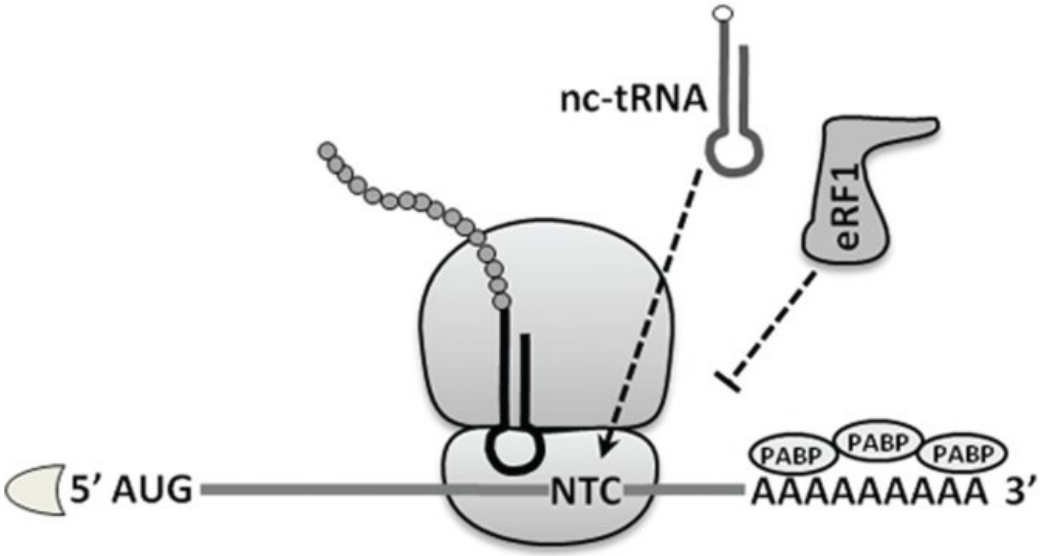



**Figure 4. Mechanism of translational stop codon readthrough.** During translational stop codon readthrough, a near-cognate tRNA (nc-tRNA) is incorporated in the ribosome A-site (instead of a release factor eRF1) and proteosynthesis is thus not terminated properly at a normal termination codon (NTC) (Dabrowski et al. 2015).

The efficiency of basal readthrough (i.e. the termination efficiency) varies depending on the identity of the stop codon, nucleotide context in the immediate vicinity of the stop codon (discussed in 3.2.5), *cis*-acting mRNA elements (Firth et al. 2011) or possibly trans-acting factors (Beznoskova et al. 2015, Beissel et al. 2019). Naturally occurring readthrough can also be programmed, which leads to generation of C-terminally extended proteins with potentially new functions or biological roles in a cell (Loughran et al. 2018, Eswarappa et al. 2014, Schueren et al. 2014, Loughran et al. 2014). It is probable that programmed readthrough is affected by similar factors as basal readthrough but can also involve mRNA-specific factors. Nonetheless, all of the regulatory factors promoting programmed readthrough have not been documented yet (Martins-Dias et al. 2021).

Originally, translational readthrough was reported in viruses, such as the tobacco mosaic virus (TMV) (Pelham 1978), the Sindbis virus (Li and Rice 1993) or the Moloney murine leukemia virus (Mo-MuLV) (Yoshinaka et al. 1985). In 1980, the very first non-viral example of natural translational readthrough was unravelled; where tRNA-Trp isolated from the rabbit reticulocyte lysate shown to be suppressing the UGA stop codon on a rabbit

beta haemoglobin mRNA (Geller and Rich 1980). Later on, translational readthrough was described in multiple organisms including *Drosophila* (Dunn et al. 2013), yeast (Namy et al. 2003) and mammals. Several examples of natural programmed readthrough in mammals follow to demonstrate the diverse pool of the readthrough-undergoing genes.

It was shown that programmed translational readthrough generates an extended variant of the vascular endothelial growth factor A (VEGF-A), VEGF-Ax, in mammalian endothelial cells. In this case, a 63-nt RNA segment following the stop codon on the VEGFA mRNA encodes the extension of the protein and acts as a *cis*-acting element; a *trans*-acting factor, the heterogeneous nuclear ribonucleoprotein hnRNP A2/B binds this element and promotes programmed stop codon readthrough. This study suggested that the readthrough-generated VEGF-Ax variant shows an anti-angiogenic activity, in contrast to the pro-angiogenic activity of the original protein (Eswarappa et al. 2014). This function was, however, questioned in a different study by demonstrating the pro-angiogenic activity of VEGF-Ax (Xin et al. 2016).

Another example could be the lactate dehydrogenase B (LDHB), where the readthrough generated extended LDHBx form contains a hidden peroxisomal targeting signal type 1 (PTS1) in the extension. It was shown that the extended form LDHBx localizes to peroxisomes and is capable of transporting the lactate dehydrogenase A (LDHA) into the peroxisomes. The LDHBx contains a relevant stop codon context, UGA CUAG, described in Loughran, 2014 and its readthrough is performed without any *cis*-acting elements (Schueren et al. 2014). A C-extended protein variant containing PTS1 and localizing to peroxisomes was reported also for malate dehydrogenase 1 (MDH1). Just like in the previous case, this gene has the UGA CUAG context (Stiebler et al. 2014).

The levels of programmed readthrough can be quite high, e.g., the OPRL1 gene, encoding a G protein-coupled opiate receptor, has both relevant stop codon context and 5' *cis*-elements and shows the readthrough levels of ~ 31 % in HEK293T cells in pDluc (Loughran et al. 2014).

Recently, a readthrough-generated C-extended variant of vitamin-D receptor (VDR), VDRx, has been described. It was shown by co-immunoprecipitation analysis that these two protein variants may form homodimers and heterodimers with each other. VDRx

shows reduced transcriptional response to its natural ligand calcitriol. No specific subcellular targeting motif was identified within the extension of VDRx and the subcellular localization is similar to VDR. The function of this extended variant is yet to be unravelled (Loughran et al. 2018).

Interestingly, it was suggested that miRNAs could potentially also affect stop codon readthrough efficiency. Let7a miRNA was shown to bind a downstream *cis*-acting element on AGO1 mRNA to promote readthrough, leading to generation of an extended protein variant named Ago1x. Ago1x was shown to load miRNA onto its target mRNA without inducing post-transcriptional gene silencing (since it cannot interact with GW182 protein); it was suggested that this extended form acts as a competitive inhibitor of miRNA pathway (Singh et al. 2019).

Up until today the precise molecular mechanism of readthrough induction by the sequence context, the *cis*-acting elements and trans-acting factors has not been revealed. Undoubtedly, new genes undergoing programmed translational readthrough will be described in future with possibly new or yet unknown biological functions and will hopefully help to discover more specific and general information regarding programmed readthrough.

### 2.2.3 Near-cognate tRNAs known to incorporate at stop codons

A near-cognate tRNA (nc-tRNA) is a tRNA with two of its three anticodon nucleotides complementary to particular stop codon nucleotides; several naturally occurring tRNAs are nc-tRNAs. The interaction between the stop codon and the nc-tRNA is performed mostly by a mismatch at either position 1 or 3 of the stop codon (Roy et al. 2015). The list of possible near-cognate codons to UAA, UAG and UGA stop codons with mismatch at position 1 or 3 is found in Table 1.

**Table 1. The list of codons with mismatch at position 1 or 3 of the UAA/UAG/UGA stop codons (eukaryotic genome).** The inserted amino acids are indicated.

| UAA | | UAG | | UGA | |
|---|---|---|---|---|---|
| GAA | Glu | GAG | Glu | GGA | Gly |
| AAA | Lys | AAG | Lys | AGA | Arg |
| CAA | Gln | CAG | Gln | CGA | Arg |
| UAC | Tyr | UAC | Tyr | UGG | Trp |
| UAU | Tyr | UAU | Tyr | UGC | Cys |
| | | | | UGU | Cys |

In yeast it was shown that Gln, Lys and Tyr can be inserted at UAA and UAG stop codon and Trp, Arg and Cys at UGA stop codon. No Glu was observed to be inserted at UAA or UAG (Blanchet et al. 2014). The same incorporation was observed in a different study, which showed that for UAA both Tyr and Gln were inserted at much higher frequency than Lys (45.5 ± 5 % and 54 ± 7 % and 0.5 ± 0.1 %). These amino acids were also inserted at UAG, where the most frequent amino acid was Gln (80 ± 6 %) and both Tyr and Lys were inserted at low frequencies (11 ± 4 % and 9 ± 2 %). Consistently, Glu was never observed to be incorporated at UAA or UAG (Roy et al. 2015).

For the UGA stop codon, there exist two types of nc-tRNAs with mismatch at stop codon position 3 – tRNA-Trp (CCA) and tRNA-Cys (GCA, ACA; depicted in Fig. 5) and two nc-tRNA with mismatch at stop codon position 1 – tRNA-Arg (UCU) and tRNA-Gly (UCC) (see Table 1). It was shown that the UGA stop codon is decoded by nc-tRNAs which mispair at position 3 of the stop codon in majority of the cases (Roy et al. 2015, Beznoskova et al. 2015). In the previously mentioned study, for this particular stop codon, Trp, Arg and Cys amino acids were incorporated, with Trp being inserted at much higher frequency (86 ± 4 %) than Arg (7 ± 3 %) and Cys (7 ± 2 %). Insertion of Gly at the UGA stop codon was not detected (Roy et al. 2015). The insertion of Trp, Arg and Cys at UGA was also described long before in case of gag-pol junction in Mo-MuLV (Feng et al. 1990).

Similar results were observed in yeast in our laboratory, the UGA stop codon being decoded by both tRNA-Trp (CCA) and tRNA-Cys (GCA). Incorporation of tRNA-Arg (UCU) or tRNA-Gly (UCC) with mismatch at position 1 was not detected. This study also suggested that eIF3, eukaryotic initiation factor 1, promotes incorporation of nc-tRNAs with a mismatch at position 3 of a programmed stop codon (for all of the three stop codons).

The eIF3 bound to programmed pre-termination complexes (on stop codons in unfavourable termination contexts) impairs the eRF1 decoding of the stop codon position 3 and allows incorporation of particular nc-tRNAs (Beznoskova et al. 2015).


**Figure 5. Decoding of UGA stop codon by a near-cognate tRNA-Cys with GCA anticodon.** The UGA stop codon can be read through by a near-cognate tRNA with mismatch at position 3 of the stop codon. This specific tRNA and stop codon were used as a general representative of the mechanism since the thesis focuses particularly on UGA stop codon and tRNA-Cys-GCA.

### 2.2.4 The role of the stop codon and its context

It was shown in an analysis of human genes that stop codons naturally occur with different frequencies – the most frequent one is the UGA stop codon (Opal, ~ 49 %), then UAA (Ochre, ~ 28 %) and the least frequent is UAG (Amber, ~ 23 %), proving that not only sense codons exhibit codon usage bias (Sun et al. 2005). Not only show the stop codons the usage bias, also, they show different probability of being read through – with UGA being the most probably read through and UAA leading the most probably to termination (Napthine et al. 2012, Loughran et al. 2014, Skuzeski et al. 1991).

Instead of the traditional three-letter stop codon, a four-letter stop codon theory was suggested for the translational termination process - in *S. cerevisiae* and *D. melanogaster,* highly expressed genes are usually terminated at UAA(A/G), in eukaryotes generally at UAA(A/G) and UGA(A/G), and in human mostly at UGA(A/G), with the fourth letter depending on the GC content of the organism (Brown et al. 1990). The base immediately following the stop codon (sometimes referred to as "+4") may even contribute to whether termination or incorporation of selenocysteine happens – in mammals, UGA followed by C or U was preferentially suppressed (1:3 ratio termination : selenocysteine incorporation),

UGA followed by A or G was three times more likely to be terminated than suppressed (McCaughan et al. 1995).

It was shown using yeast in our laboratory that the nature of the +4 nucleotide determines the character of the near-cognate tRNA incorporated. Specifically, UGA-G and UGA-C are preferentially read through by cysteine tRNA, UGA-A incorporates tryptophan (Beznoskova, Gunisova and Valasek 2016). Also, in yeast, the readthrough levels increase the most if the base immediately following any of the stop codons is cytosine, and the efficiency of readthrough increases in order: UGA-U < UGA-G < UGA-A < UGA-C (Beznoskova et al. 2015, Dabrowski et al. 2015). An explanation was proposed, that for any of the stop codons, if the +4 nucleotide is cytosine, then the ability of eRF1 to recognize a stop codon properly is compromised and thus the readthrough levels increase (Beznoskova et al. 2016).

Unsurprisingly, stop codon readthrough is dependent not only on the +4 nucleotide, but on the nucleotide context around the stop codon. It was described for the Colorado tick fever virus (CTFV) that the sequence following the stop codon can influence the efficiency of stop codon readthrough and shown that UGA is the leakiest stop codon (the most susceptible to readthrough) and UAA is the most strictly used as a termination codon (Napthine et al. 2012).

A largely known context to stimulate readthrough is the tobacco mosaic virus (TMV) nucleotide context. In case of TMV, the genomic RNA of this virus translates into two proteins, a smaller 126 kDa and a larger 183 kDa one, the latter being produced as a product of UAG stop codon readthrough (Pelham 1978, Goelet et al. 1982), probably by a natural cytosolic near-cognate tRNA-Tyr of the tobacco plant (Beier et al. 1984). This readthrough is again dependent on the stop codon 3'context sequence, here being UAG-CAR-YYA (Skuzeski et al. 1991). It was shown in plants that all of the three stop codons can be read through when placed in the TMV context, with UAG being the leakiest (Skuzeski et al. 1991). A similar sequence, UAG-CAR-NBA, is known to stimulate readthrough in yeast (Namy, Hatin and Rousset 2001).

In human cells, a different experiment tested seven human readthrough-candidate genes with highly conserved UGA stop codons and experimentally shown that four of those,

OPRK1, OPRL1, MAPK10 and AQP4 have the UGA stop codon immediately followed by a CUAG sequence and show significantly increased readthrough levels. The CUAG context was shown to be the most efficient for the UGA stop codon (Loughran et al. 2014). The OPRL1 gene, when placed in the TMV context (CAAUUA), showed lower level of readthrough (about 15 %) than with its native CUAG context (about 31 %). When mutating the CAAUUA to CUAUUA and thus partially restoring the CUAG motif, the original (CUAG) level of readthrough was restored (Loughran et al. 2014). Further study identified altogether 23 human candidate genes with UGA immediately followed by CUAG (six of those, OPRK1, OPRL1, MAPK10, AQP4 (Loughran et al. 2014), LDHB (Schueren et al. 2014), MDH1 (Stiebler et al. 2014), were described to undergo readthrough before), the remaining 17 were tested, which showed for all the 17 tested genes the readthrough levels ranging from 1.3 % to 6.7 %. Among those, five genes, ATP10D, CDH23, DDX58, SIRPB1 and TMEM86B showed readthrough levels about 5 %. The highest readthrough was observed for the vitamin D receptor mRNA, 6.7 % (Loughran et al. 2018).

Not only is the 3'context important, but the 5'context upstream of the stop codon has also been discussed, however, not so many publications focus on the role of the 5' context.
In yeast, it was suggested that the P-site tRNA (in -1 codon position relative to the stop codon) and the chemical character of the amino acid in -2 position relative to the stop codon affect the readthrough efficiency (Mottagui-Tabar, Tuite and Isaksson 1998). Further study in yeast suggested that the P-site tRNA and the identity of the penultimate amino acid are not the main 5'determinants of UAG readthrough efficiency, and suggested presence of two adenines at the two positions preceding the UAG stop codon instead; it was hypothesized that these adenines modify the mRNA structure at the P-site of the ribosome (Tork et al. 2004).

In mammals, the role of the -1 amino acid on the UAG readthrough efficiency was investigated using NIH3T3 mice cells and revealed that it is not the main 5'determinant. A large variability in UAG readthrough levels was observed for the same amino acid depending on the codon used (for example glycine is encoded by four codons and there is a 5-fold difference in the readthrough levels when GGA or GGU is used). The role of the -1 nucleotide was then investigated. It was demonstrated that the identity of the nucleotide at the -1 position affects readthrough and the readthrough efficiency gradually decreases in nucleotide order A > G > Y (pyrimidine), with A being associated

with the highest readthrough. Concerning the frequency, a strong bias was revealed on this position, with C being the most frequent nucleotide. However, no correlation of frequency and readthrough efficiency was observed – both A and U show almost comparable low frequency at -1 position, however, A provides the highest readthrough levels and U the lowest (Cassan and Rousset 2001).

To summarize, this data indicate that the stop codon and its context play an important role in the readthrough efficiency, however, there is still a lot to investigate about this interesting topic, probably mainly about the upstream 3' context importance, which has not been studied as much as the downstream 5' context.

### 2.2.5 The therapeutic potential of programmed readthrough

It is not uncommon that a premature termination codon (PTC) arises on mRNA. When such mRNA bearing a PTC is translated, truncated protein variants leading to gain-of-function or dominant-negative effects could result. Thus, the cell usually directly degrades these stop codon-containing mRNAs by a mechanism of nonsense-mediated mRNA decay (NMD, reviewed in (Frischmeyer and Dietz 1999).

However, this mechanism may not be 100 % sufficient. It is estimated that about 11 % of all described gene lesions causing human genetic diseases and about 20 % of all single base pair substitutions affecting gene coding regions associated with human diseases is generated by nonsense mutations leading to the introduction of a premature termination codon (PTC) into the coding sequence (Mort et al. 2008). PTCs can account for such diseases as cystic fibrosis, spinal muscular atrophy (SMA), Duchenne muscular dystrophy or haemophilia, thus, searching for an effective drug that would induce translational readthrough of PTC and even partially restore the activity of a wild-type protein is desired (Lee and Dougherty 2012).

In case of naturally occurring stop codons, readthrough occurs with 0.001– 0.1 % frequency, whilst in case of PTCs, the frequency is higher, 0.01 – 1 % (Keeling et al. 2012). The fact that the "real" stop codons, which usually reside "at the end" of a gene are usually read through with lower frequency could imply that when readthrough-inducing drugs are used, the "normal" genes would not be affected so much (Keeling et al. 2012). This was supported by primer extension inhibition (toeprinting) assay, which showed that the

translational machinery pauses at PTCs, but not at normal stop codons – such a pause may increase the propensity of readthrough occurrence (Keeling et al. 2012, Amrani et al. 2004). Another explanation was proposed, based on an information that during termination, interaction of poly(A) binding protein (PABP) with eIF3 is stimulatory for peptide release. Since normal stop codons reside "at the end" of a gene and PTCs usually reside randomly, far upstream from normal poly(A) tails, this interaction with PABP may be lacking. However, the mRNA molecule circularizes during the process of translation, by interaction of PABP and eIF4G, thus the distance between PABP and eRF3 may not be the only explanation for this (Keeling et al. 2012).

Readthrough can be induced pharmacologically. The mechanism of a PTC readthrough is depicted in Fig. 6. Among the most studied molecules which could mediate PTC readthrough belong aminoglycoside antibiotics (AAGs, which lead to incorporation of nc-tRNA), but many other molecules have been challenged, including suppressor tRNAs (tRNAs with anticodon perfectly complementary to a stop codon). However, usually the main problem with PTC readthrough-inducing drugs resides in their toxicity. At this time, no actually effective, generally used PTC readthrough-inducing drug exists, that would be both efficient and safe (Dabrowski, Bukowy-Bieryllo and Zietkiewicz 2018, Lee and Dougherty 2012).

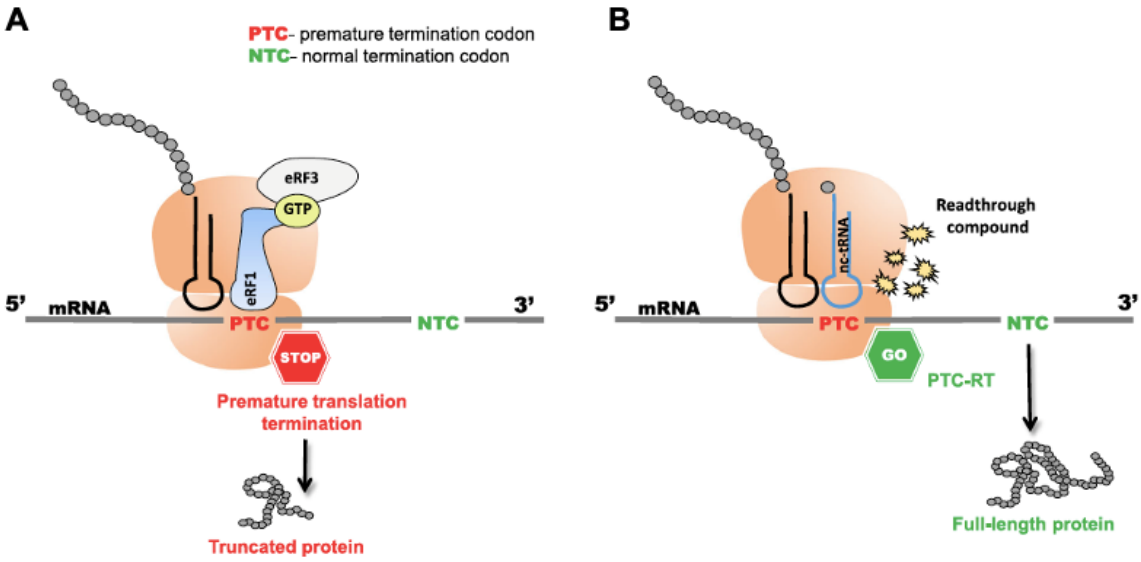



**Figure 6. Schematic representation of readthrough on mRNA containing a PTC**. **A)** When no readthrough occurs, translation is terminated at PTC and truncated protein is generated. **B)** When readthrough is induced (here indicted by a "Readthrough compound"), a PTC is recoded by a near-cognate tRNA and translation continues until the next termination codon is met (Dabrowski et al. 2018).

# 3 AIMS OF THE THESIS

The aim of this thesis was to investigate if tRNA-Cys could function as a near-cognate tRNA in human cell lines by decoding the UGA stop codon *via* stop codon readthrough. Several specific aims were set:

**Aim 1:**

**Description**: The aim was to identify the tRNA-Cys isodecoders which are expressed in human cell lines.

**Strategy**: tRNA isoforms are today only computationally predicted. Thus, published data from tRNA-Seq and ChIP-Seq were analyzed to select the highest scoring molecules according to their expression.

**Aim 2:**

**Description**: The aim was to investigate if tRNA-Cys could act as a near-cognate/ readthrough-inducing tRNA in human cell lines, since it was described for yeast tRNA-Cys before (Beznoskova et al. 2015, Beznoskova et al. 2016).

**Strategy**: HEK293T human cells were transfected with p2luci or pSGDluc where the mature tRNA sequences of the selected tRNA-Cys isodecoders were placed under the U6 promoter. The isodecoder's ability to increase UGA readthrough was assessed using dual-luciferase reporter assay.

**Aim 3:**

**Description**: The aim was to identify if the anticodon of tRNA-Cys (GCA) is capable of decoding the Opal stop codon (UGA).

**Strategy**: tRNA-Gln-GCT previously reported as a functional suppressor tRNA was used and its anticodon was mutated from GCT to GCA. This sequence, placed under the U6 promoter, was tested similarly to the tRNA-Cys isodecoders.

# 4 MATERIAL AND METHODS

## 4.1 Laboratory equipment

### 4.1.1 Centrifuges

Beckman Coulter Allegra® X-15R Centrifuge, rotor SX 4750A

Beckman Coulter optima™ L-90K Ultracentrifuge, rotor SW 41

Biosan Centrifuge/vortex Multi Spin MSC-300

Eppendorf Centrifuge 5414D

### 4.1.2 Electrophoresis

Bio-Rad Criterion Cell

Bio-Rad Mini-Sub Cell GT Cell

Bio-Rad Wide-Sub Cell GT Cell

Bio-Rad PowerPac Basic power supply

Bio-Rad PowerPac HC power supply

Bio-Rad PowerPac Universal power supply

### 4.1.3 Other equipment

Beckman Coulter DU® 530 Life Science UV/VIS Spectrophotometer

Biosan Bio RS-24 Rotator

Biosan Mini Rocker MR-1

Biosan Multi RS-60 Rotator

Eppendorf Thermomixer Comfort

Syngene G:BOX iChemi – gel documentation and analysis system

Multitron INFORS HT shaker

Nunc™ MicroWell™ 96-Well, Nunclon Delta-Treated, Flat-Bottom Microplate (Thermo Scientific™)

NanoDrop™ One Microvolume UV-Vis Spectrophotometer (Thermo Scientific™)

Neubauer-improved CE-marked counting chamber, 0.1mm depth, Marienfeld Germany

Scientific Industries Vortex genie 2

Sanyo MCO-18AIC $CO_2$ incubator

Infinite® 200 PRO (TECAN)

## 4.2 Chemicals

10x TBE (Bio-Rad)

10x TG (Bio-Rad)

10x PBS (Bio-Rad)

Agar (Serva Electrophoresis™)

Bradford Reagent (Sigma-Aldrich®)

Bromophenol Blue (Sigma-Aldrich®)

Albumin, Bovine, Fraction V (MP Biomedicals)

Bacto Peptone (Becton, Dickinson and Company)

Bacto Tryptone (Becton, Dickinson and Company)

β-Mercaptoethanol (Sigma-Aldrich®)

Bright-Glo™ Luciferase Assay Reagent (Promega)

Criterion™ TGX™ Precast Gels, 4-20 %, 1.0 mm, 12+2 Well Comb, 45 µl (Bio-Rad)

CutSmart™ Buffer

EDTA-free Complete Protease Inhibitor Mix tablets (Roche)

Electroporation cuvettes (Eppendorf)

Ethidium bromide (Serva Electrophoresis™)

Glo Lysis Buffer (GLB), 1X (Promega)

NTP (NEB®)

NEBuffer™ 3.1

SDS solution 20 % (Sigma-Aldrich®)

SeaKem® LE Agarose (Lonza)

Subcloning Efficiency™ DH5α™ Competent Cells (Invitrogen™)

SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific™)

Tris (Serva Electrophoresis™)

Triton X-100 (Sigma-Aldrich®)

TurboFect™ Transfection Reagent

TWEEN® 20 (Sigma-Aldrich®)

UltraPure™ DNase/RNase-Free Distilled Water (Invitrogen™)

*Renilla*-Glo™ Luciferase Assay Reagent (Promega)

## 4.3 Solutions

**Blocking solution**: 1x TBS-T, 5 % (w/v) low-fat milk or BSA

**Blotting buffer:** 20 % (v/v) methanol; 1x TG

**1x GA HEPES:** 10 mmol.l$^{-1}$ HEPES, pH 7.5, 62.5 mmol.l$^{-1}$ KCl, 2.5 mmol.l$^{-1}$ MgCl2

**Sample buffer for agarose gel electrophoresis 6x:** 30 % (v/v) glycerol, 0.25 % (w/v) bromophenol blue

**Sample loading buffer for WB 2x:** 250 mmol.l$^{-1}$ Tris-HCl (pH 6.8), 20 % (v/v) glycerol, 4 % (w/v) SDS, 2 % (v/v) beta-mercaptoethanol, 0.4 % (w/v) bromophenol blue

**Sample loading buffer for SDS-PAGE 4x:** 1 mol.l$^{-1}$ Tris-HCl (pH 6.8), 40 % (v/v) glycerol; 8 % (w/v) SDS, 0.06% (w/v) bromophenol blue, 1.47 % (v/v) beta-mercaptoethanol

**SDS-PAGE running buffer:** 1x TG, 0.1 % (w/v) SDS

**TBS-T buffer:** 1x TBS, 0.1 % (w/v) Tween® 20

## 4.4 Antibodies

### 4.4.1 Primary antibodies

Anti-CARS antibody (ab235536) (Abcam) – rabbit polyclonal antibody to human CARS, reacts with human/mouse, suitable for WB (recommended dilution 1:2.000 – 1:5.000)

Recombinant RabMAb® Anti-CARS antibody [EPR7121] (ab126714) (Abcam) – rabbit monoclonal antibody to CARS, reacts with human, suitable for WB (recommended dilution 1:10.000 – 1:50.000)

Anti-CARS antibody produced in rabbit, Prestige Antibodies® Powered by Atlas Antibodies (HPA002383) (Sigma-Aldrich®) – rabbit polyclonal antibody to CARS, reacts with rat/mouse/human, suitable for WB (dilution not specified, generally for Prestige Antibodies Sigma-Aldrich® recommends 1:250-1:500 dilution)

Anti-RPS14 antibody (ab174661) (Abcam) – rabbit polyclonal to RPS14, reacts with mouse/human, suitable for WB (recommended dilution 1:2.000 – 1:10.000)

### 4.4.2 Secondary antibodies

ECL™ Anti-Rabbit IgG, Horseradish Peroxidase linked whole antibody from donkey, Lot.116599202 (Sigma-Aldrich®), diluted 1:10.000

## 4.5 Enzymes and inhibitors

T4 Ligase (Roche)
HEPES (Serva Electrophoresis™)
RNase inhibitor (Roche)
Leupeptin (Sigma-Aldrich®)
Aprotinin (Sigma-Aldrich®)
PMSF (Serva Electrophoresis™)
DTT (Sigma-Aldrich®)
Proteinase K(Sigma-Aldrich®)
Trypsin (Sigma-Aldrich®)
Triton X-100 (Sigma-Aldrich®)
Pepstatin (Roche)
cOmplete™ Protease Inhibitor Mix tablets (Roche)
BamHI (NEB®)
SalI (NEB®)
MscI (NEB®)
PstI (NEB®)
NcoI (NEB®)

## 4.6 Markers

TrackIt ™ 1 Kb Plus DNA Ladder (Invitrogen™)

1 kb Plus DNA (Invitrogen™)

Precision Plus Protein™ Dual Color Standards (Bio-Rad)

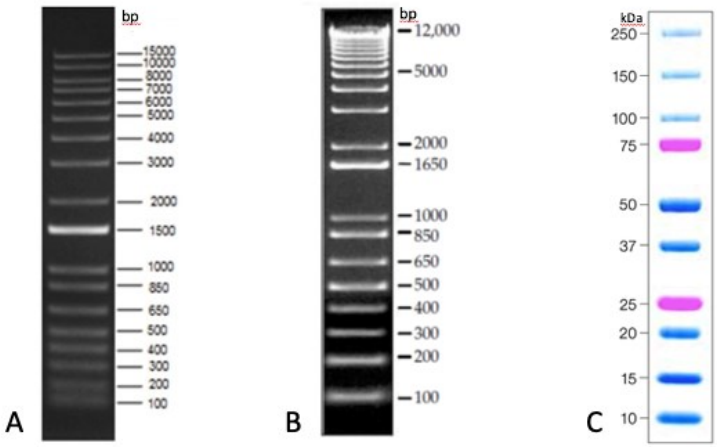



**Figure 7. Representation lot of markers from distributors' official website.** Markers displayed as from the left: A: The TrackIt™ 1 Kb Plus DNA Ladder (Invitrogen™) run on 1 % agarose gel with ethidium bromide. B: 1 kb Plus DNA (Invitrogen™) on 0.9 % agaro se gel with ethidium bromide. C: Precision Plus Protein™ Dual Color Standards (Bio-Rad), on a 4-20 % Criterion™ TGX™ gel.

## 4.7 Plasmids

**Table 2. The list of plasmids used in this study** which were not constructed directly during the course of this work.

| Plasmid name | Plasmid description | Source of reference |
|---|---|---|
| PBB171 | p2luci-UGA-TMV-U6-no tRNA | Beznoskova et al. 2021 |
| PBB172 | p2luci-CAG-TMV-U6-no tRNA | Beznoskova et al. 2021 |
| B1145 | pSGDluc-UGA-CUAG-no tRNA | Loughran et al. 2017 |
| B1146 | pSGDluc-UGG-CUAG-no tRNA | Loughran et al. 2017 |
| PBB210 | p2luci-UGA-TMV-U6-tC01 | Petra Beznoskova |
| PBB200 | p2luci-CAG-TMV-U6-tC01 | Petra Beznoskova |
| PBB212 | p2luci-UGA-TMV-U6-tC02 | Petra Beznoskova |
| PBB201 | p2luci-CAG-TMV-U6-tC02 | Petra Beznoskova |
| PBB215 | p2luci-UGA-TMV-U6-tC03 | Petra Beznoskova |
| PBB202 | p2luci-CAG-TMV-U6-tC03 | Petra Beznoskova |

| | | |
|---|---|---|
| PBB267 | p2luci-UGA-TMV-U6-tC04 | Petra Beznoskova |
| PBB268 | p2luci-CAG-TMV-U6-tC04 | Petra Beznoskova |

## 4.8 Oligonucleotides

### 4.8.1 GeneArt™ Strings™ DNA Fragments (Invitrogen™)

**Table 3. The list of GeneArt™ Strings™ DNA Fragments used in this study.**

| Name | Sequence (5´to 3´) |
|---|---|
| **tC06 (for tRNA-Cys-GCA-2-2)** | CTGATAAATGCTTCAATAATAGTTTAAACGGATCCTGGC CAAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAG ATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATC ATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC TTTATATATCTTGTGGAAAGGACGAAACACCGGGGCCCG GGGGTATAGCTCAGTGGTAGAGCATTTGACTGCAGATCA AGAGGTCCCCGGTTCAAATCCGGGTGCCCCCTGGTACCTA TTGAAAAATTTTTTGGAACCCGGGGTCGACGTTTAAACTAT TGAAAAAGGAAGAGTATG |
| **tC07 (for tRNA-Cys-GCA-14-1)** | CTGATAAATGCTTCAATAATAGTTTAAACGGATCCTGGC CAAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAG ATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATC ATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC TTTATATATCTTGTGGAAAGGACGAAACACCGGGGCCCG GGGGTATAGCTCAGGGGTAGAGCATTTGACTGCAGATCA AGAAGTCCCCGGTTCAAATCCGGGTGCCCCCTGGTACCTA TTGAAAAATTTTTTGGAACCCGGGGTCGACGTTTAAACTAT TGAAAAAGGAAGAGTATG |

| | |
|---|---|
| **tC08**<br>**(for tRNA-Cys-GCA-17-1)** | CTGATAAATGCTTCAATAATAGTTTAAACGGATCCTGGC<br>CAAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT<br>CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAG<br>ATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA<br>GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG<br>TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATC<br>ATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC<br>TTTATATATCTTGTGGAAAGGACGAAACACCGGGGCCCG<br>GGGATATAGCTCAGGGGTAGAGCATTTGACTGCAGATCA<br>AGAGGTCCCCGGTTCAAATCCGGGTGCCCCCCGGTACCTA<br>TTGAAAAATTTTTTGGAACCCGGGGTCGACGTTTAAACTA<br>TTGAAAAAGGAAGAGTATG |
| **tC09**<br>**(for tRNA-Cys-GCA-9)** | CTGATAAATGCTTCAATAATAGTTTAAACGGATCCTGGC<br>CAAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT<br>CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAG<br>ATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA<br>GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG<br>TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATC<br>ATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC<br>TTTATATATCTTGTGGAAAGGACGAAACACCGGGGCCCG<br>GGGGTATAGCTCAGGGGTAGAGCATTTGACTGCAGATCA<br>AGAGGTCCCTGGTTCAAATCCAGGTGCCCCCTGGTACCTA<br>TTGAAAAATTTTTTGGAACCCGGGGTCGACGTTTAAACTA<br>TTGAAAAAGGAAGAGTATG |
| **tC10**<br>**(for tRNA-Cys-GCA-4-1)** | CTGATAAATGCTTCAATAATAGTTTAAACGGATCCTGGC<br>CAAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT<br>CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAG<br>ATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA<br>GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG<br>TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATC<br>ATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC<br>TTTATATATCTTGTGGAAAGGACGAAACACCGGGGCCCG<br>GGGGTATAGCTCAGTGGTAGAGCATTTGACTGCAGATCA<br>AGAGGTCCCTGGTTCAAATCCGGGTGCCCCCTGGTACCTA<br>TTGAAAAATTTTTTGGAACCCGGGGTCGACGTTTAAACTA<br>TTGAAAAAGGAAGAGTATG |

| | |
|---|---|
| **tQ-GCA (for tRNA-Gln-GCA)** | CTGATAAATGCTTCAATAATAGTTTAAACGGATCCTGGC CAAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATT CCTTCATATTTGCATATACGATACAAGGCTGTTAGAGAG ATAATTGGAATTAATTTGACTGTAAACACAAAGATATTA GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGG TAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATC ATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC TTTATATATCTTGTGGAAAGGACGAAACACCGGGGCCCG GTTCCATGGTGTAATGGTTAGCACTCTGGACTGCAAATCC AGCGATCCGAGTTCAAATCTCGGTGGAACCTGGTACCTATT GAAAAATTTTTTGGAACCCGGGGTCGACGTTTAAACTATTG AAAAAGGAAGAGTATG |

## 4.9 Bacterial cultivation

### 4.9.1 Escherichia coli strain

| Strain | Genotype | Source |
|---|---|---|
| DH5α | fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 | Invitrogen™ |

### 4.9.2 Permanent strain collections and strain storage

Stock bacterial strains were stored in 50 % glycerol in a -80 °C freezer. When necessary, the bacteria were stored for a short period of time on agar plates or in liquid media at 4 °C.

### 4.9.3 Bacterial cultivation media and plates

**Luria – Bertani (LB) medium:** 1 % (w/v) bacto tryptone; 0.5 % (w/v) bacto yeast extract; 0.5 % (w/v) NaCl

**LB-Amp medium**: LB; 0.015 % (w/v) ampicillin

**LB-Amp plates:** 1 % (w/v) bacto tryptone; 0.5 % (w/v) bacto yeast extract; 0.5 % (w/v) NaCl; 2.5 % (w/v) agar; 0.015 % (w/v) ampicillin

**SOC medium:** 2 % (w/v) bacto tryptone; 0.5 % (w/v) bacto yeast extract; 0.06 % (w/v) NaCl; 0.02 % (w/v) KCl; 0.35 % (w/v) glucose

## 4.10 Mammalian cell culture cultivation

### 4.10.1 HEK293T cell line

The HEK293T cell line (ATCC® CRL-3216™) was purchased from original vendor's e-shop website and was handled according to the vendor's instructions.

### 4.10.2 Cell culture cultivation media

**DMEM:** Dulbecco's Modified Eagle's Medium (Sigma-Aldrich)
**DMEM with 10 % FBS:** Dulbecco's Modified Eagle's Medium (Sigma-Aldrich), 10 % (v/v) Fetal Bovine Serum (Thermofisher Scientific)

### 4.10.3 Cell culture cultivation plastics

Tissue culture dish 100, Techno Plastic Products®, product no. 93100
Tissue culture flask, 75 cm2, Techno Plastic Products®, product no. 90076
Tissue culture test plate, 6 wells, Techno Plastic Products®, product no. 92406
Tissue culture test plate, 24 wells, Techno Plastic Products®, product no. 92424

## 4.11 Working with bacterial strain

### 4.11.1 Bacterial strain cultivation

Bacteria were grown in a liquid media at 37 °C while shaking at 210 rpm or on solid agar plates at 37 °C in an incubator.

## 4.12 Working with cell cultures

### 4.12.1 Cell culture cultivation

The cells were incubated at 37 °C in a humidified incubator with 5 % $CO_2$. The cells were cultivated in cell cultivation plastic dishes (see 4.10.3), with number of cells and amount of the medium adjusted according to Table 4. It is important to use proper amount of medium, approximately 0.2 ml to 0.3 ml of medium for 1 $cm^2$ of culture vessel growth surface area is recommended; using more medium may result in too slow diffusion of oxygen to cells, whilst using too little medium can mean too little nutrients for the cells to live on. Generally, minimum of 1 x $10^5$ cells/$cm^2$ can be expected to grow in an attached monolayer in culture, however, this number may vary depending on the cell line type and culture conditions (online source, [3]).

**Table 4. Summarization of the necessary information for different cell culture plastics.** Information taken from online sources [3] and [4.]

| | **Surface area ($cm^2$)** | **Seeding density** | **Cells at confluency*** | **Total volume of dish/well (ml)** |
|---|---|---|---|---|
| 60mm dish | 21.5 | 0.8 x $10^6$ | 3.2 x $10^6$ | 4.3 – 6.5 |
| 100mm dish | 56.7 | 2.2 x $10^6$ | 8.8 x $10^6$ | 11.3 – 17.0 |
| 6-well plate | 9.6 | 0.3 x $10^6$ | 1.2 x $10^6$ | 1.9 – 2.9 |
| 24-well plate | 1.9 | 0.05 x $10^6$ | 0.24 x $10^6$ | 0.4 – 0.6 |
| T-75 flask | 75 | 2.1 x $10^6$ | 8.4 x $10^6$ | 15 – 22.5 |

*can vary with cell type, estimated here for HeLa cells

### 4.12.2 Passaging cells

The cells were checked every day under the microscope to monitor their confluency (percentage of surface covered by cell monolayer). The culture was passaged when the cells were about 80 % confluent. The cells should not become overconfluent because it could negatively affect their viability and gene expression. All equipment that was to be placed to the flow box was sprayed with 70 % ethanol to prevent from possible contamination of the working area and cells. The medium, 1x PBS and trypsin were preheated to 37 °C.

The medium was aspirated from the side of the flask or plate (not to touch the cells) by using aspiration pump. For 75 $cm^2$ flask, 5 ml of 1x PBS was gently added to wash the cells from residual medium and then aspirated. 2 ml of 0.025 % trypsin was added to the flask and incubated shortly at 37 °C until the cells have detached. A new flask was prepared in the meantime. After detaching of the cells, the 2 ml of trypsin-cell suspension was transferred to a tube with fresh medium and 10 µl of this suspension was used to count the number of cells present. This was performed under the microscope using a glass haemocytometer, the Neubauer counting chamber.

When passaging, most cells should not be split more than 1:10 (based on the confluent flask surface area) as the seeding density will be too low for the cells to survive (online source, [1]). Thus, an appropriate counted volume of cell suspension was inoculated into a new flask. The confluency was anticipated since we presume that the cell number doubles every day. The flasks were incubated at 37 °C in a humidified incubator with 5 % $CO_2$. HEK293T cells are typically passaged when the flask is about 80 % confluent and they are split to a ratio of 1:10 - 1:20 (online source, [2]).

### 4.12.3 Counting cells in Neubauer-improved counting chamber

The cells were counted under the microscope using a counting chamber. The glass haemocytometer (Neubauer-improved CE-marked counting chamber, 0.1 mm depth, Marienfeld Germany) was covered with a cover glass and then the cell suspension was applied to the edge of the haemocytometer to allow the cell suspension to be drawn in by capillary action. Only cells present within a small square or present on a left line or bottom line of the small square were counted Fig. 8.

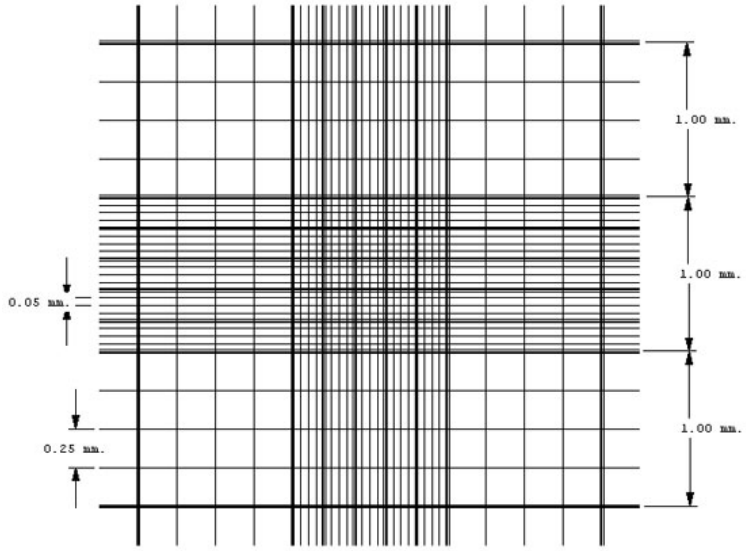



**Figure 8. The Neubauer-improved counting chamber.** The whole area of the chamber measures 9 $mm^2$. The chamber consists of 9 large squares (1 $mm^2$ each). Each of the 9 squares is divided by single lines into 16 small squares. The central large square is divided into 25 medium size squares (0.04 $mm^2$ each) with double or triple lines. Each of these 25 squares is again divided into 16 small squares (0.0025 $mm^2$ each) with single lines. The chamber depth is 0.1 mm. The picture was downloaded from https://laboratoryinfo.com/manual-cell-counting-neubauer-chamber/.

The cells were counted in each of the 4 large squares. At the end of the procedure the average of the 4 large squares and cell concentration was calculated. To calculate the number of cells in 1 ml we multiplied the cell average by $10^4$.

## 4.13 DNA manipulation

### 4.13.1 Plasmid DNA isolation – QIAprep® Spin Miniprep Kit (Qiagen®)

Commercially distributed QIAprep® Spin Miniprep kit was used for the isolation of recombinant plasmids. The kit contained: Buffer P1, Buffer P2 (RNase A added), Buffer PE, QIAprep® spin column, Collection tube (2 ml).

*E. coli* was cultivated 16 hours prior to being harvested for isolation of the plasmid DNA. Isolation was then performed following seller's instructions. First of all 4 ml of the culture was pelleted by centrifugation at ~ 16.000 g for 1 minute at room temperature. The pellet was resuspended in 250 µl of Buffer P1, which contained RNase A. For lysis of the cells 250 µl of Buffer P2 was added next and the solution was mixed thoroughly by inverting the tube 4-6 times until the solution became clear. After mixing, 350 µl of neutralization buffer Buffer N3 was added and the solution was mixed immediately and throughly by inverting the tube 4-6 times. The mixture was then centrifuged at ~ 16.000 g for 10 minutes. QIAprep® spin column was inserted into a provided collection tube. The supernatant containing the plasmid DNA was applied to the QIAprep® spin column to bind DNA and the tube was centrifuged at ~ 16.000 g for 60 seconds and then the flow-through liquid was discarded. Next the QIAprep® spin column was washed by adding 500 µl of Buffer PB and centrifuged at ~ 16.000 g for 60 seconds. The flow-through was discarded again. 750 µl of Buffer PE was added next and the tube was centrifuged at ~ 16.000 g for 60 seconds. The flow-through was discarded again and the QIAprep® spin column was centrifuged for another 1 minute to remove residual wash buffer. The QIAprep® spin column was placed into a clean 1.5 ml microcentrifuge tube. To elute DNA, 30-50 µl of water was added to the QIAprep® spin column, let stand for 1 minute and then centrifuged again at ~ 16.000 g for 60 seconds. After this isolation the concentration of DNA was measured using Nanodrop™.

### 4.13.2 DNA concentration measurement

The DNA concentration was measured using the Nanodrop™ according to manufacturer's instructions. For analysis, 1.5 µl of sample was used. After measurement, the $A_{260}/A_{280}$ and $A_{260}/A_{230}$ ratio was checked.

### 4.13.3 Restriction digestion reaction

For restriction digestion reaction the proper enzyme was chosen that would cut within a desired position of DNA molecule. Sizes of desired fragments were predicted using Clone Manager software. For the enzyme chosen, the number of U/ml was checked, as well as its ability to digest in different temperatures and buffers. Reaction conditions were set according to manufacturer´s recommendations (NEB®, Roche). When digesting with more enzymes, the compatibility of buffers and temperatures was checked. 5 – 10 U of enzyme was used for 1 µg of DNA. The volume of the reaction and the amount of every reagent was set according to the table below (Table 5). The digestion was confirmed using agarose gel electrophoresis according to the protocol for electrophoresis (see 4.13.5) and Sanger sequencing (see 4.13.8).

**Table 5. The recommended reagent volumes for different restriction reaction volumes.**

| Reaction volume (µl) | Enzyme (Units) | DNA (µg) | 10x Buffer (µl) |
|---|---|---|---|
| 10 | 1 | 0.1 - 0.2 | 1 |
| 25 | 5 | 0.5 | 2.5 |
| 50 | 10 | 1 | 5 |

### 4.13.4 Ligation

Ligation reaction was performed in total reaction volume of 15 µl with 1:3 vector:insert molar ratio. First of all, the proper amount of vector DNA and insert DNA was mixed and incubated for 1 minute at 60 °C. After this incubation, the mixture was kept on ice for 5 minutes. Next 1 µl of the 10 x ligation buffer and 0.75 µl of T4 DNA ligase (1 U/µl) was added and the mixture was incubated for 20 minutes on ice and then kept in 16 °C overnight (minimal incubation period should be 16 hours). A control ligation reaction with $H_2O$ instead of insert DNA was always prepared to check for vector self-ligation. The ligated product was then transformed to the competent cells of *E. coli* strain DH5α by heat shock method according to the protocol for heat shock transformation (see 4.14.1).

### 4.13.5 Agarose gel electrophoresis

For the preparation of agarose gel the proper amount of agarose was dissolved in a desired volume of 1x TBE by heating in a microwave oven (e.g., 1 g of agarose in 100 ml of 1x TBE for a 1 % gel). After cooling the agarose down to approximately 60 °C, ethidium bromide was added to final concentration 0.5 μg/ml. The mixture was stirred to disperse the ethidium bromide and then poured to a gel rack with a comb already inserted. After the gel became solid, the comb was removed and the gel in the rack was placed into a tank with 1x TBE with ethidium bromide added to final concentration 0.5 μg/ml. The 1 Kb Plus DNA Ladder (Invitrogen™) was used as a marker and loaded to the gel first. The samples were mixed with 6x Sample buffer (1 volume of Sample Buffer to 5 volumes of our sample) and loaded onto the gel. After loading the samples, the electrophoresis was run at 5 V/cm of gel for approximately 1 hour. Each 15 minutes the gel was being taken picture of under a UV transilluminator and monitored if the fragments could be already distinguished properly by size.

### 4.13.6 Isolation of DNA from the gel - QIAquick® Gel extraction Kit (Qiagen®)

QIAquick® Gel extraction Kit distributed by the Qiagen® company was used for the isolation of desired fragments of DNA from agarose gel. The kit contained: Buffer QG, Buffer PE (ethanol added), QIAquick® column, collection tube (2 ml). The DNA band of our interest was excised from the agarose gel with a clean and sharp scalpel. The isolation was performed according to the vendor's protocol. The gel slice was weighed in a microcentrifuge tube. Next 3 volumes of Buffer QG to 1 volume of the gel (100 mg gel equals to 100 μl) was added to the gel slice in the tube. This mixture was incubated at 50 °C for 10 minutes and vortexed every 2 minutes to help dissolve the gel. After the incubation 1 gel volume of isopropanol was added to the sample and the sample was mixed. The QIAquick® spin column was placed in a provided 2 ml collection tube. The sample was then applied to the QIAquick® column and centrifuged at ~ 16.000 g for 60 seconds. The flow-through was discarded and the QIAquick® column was placed into the same tube. To wash, 750 ml of Buffer PE was added to the QIAquick® column and centrifuged at ~ 16.000 g for 60 seconds. The flow-through was then discarded and the QIAquick® column was placed back into the same tube. To remove residual wash buffer, the QIAquick® column in the provided collection tube was centrifuged at ~ 16.000 g for another 60 seconds. After centrifugation, the QIAquick® column was placed into

a clean 1.5 ml microcentrifuge tube. To elute DNA, 30-50 µl of water was added, let stand for 60 seconds and then centrifuged at ~ 16.000 g for 60 seconds. Finally, the concentration of DNA was measured using Nanodrop™.

### 4.13.7 DNA purification - QIAquick® PCR Purification Kit (Qiagen®)

This kit was used for DNA purification. The kit contained: Buffer PE, Buffer PB, QIAquick® column, collection tube (2 ml). The purification was performed according to the seller's instructions. One volume of DNA was added to 4 volumes of Buffer PB and the sample was mixed. The QIAquick® column was placed into a provided 2 ml collection tube. To bind DNA, the sample was applied to the QIAquick® column and centrifuged at ~ 16.000 g for 60 seconds. The flow-through was discarded and the QIAquick® column was placed back into the same tube. To wash, 750 µl of Buffer PE was added to the QIAquick® column and centrifuged at ~ 16.000 g for 60 seconds. The flow-through was discarded and the QIAquick® column was placed back to the same tube. The QIAquick® column was centrifuged once more in the provided 2 ml collection tube for 60 seconds to remove residual wash buffer. The QIAquick® column was then placed into a clean 1.5 ml microcentrifuge tube. To elute DNA, 30-50 µl of water was added, let stand for 1 minute and then centrifuged at ~ 16.000 g for 60 seconds. After this isolation the concentration of DNA was measured using Nanodrop™.

### 4.13.8 Sequencing

After restriction digestion confirmation, the positive clones were sent for Sanger sequencing to SEQme s.r.o. Each sequencing reaction was prepared by mixing of 500 ng of template DNA, 2.5 µl of 10 mmol primer and appropriate amount of water, so that the final volume of reaction was 10 µl. The sequencing results were than analysed using the BioEdit software and the Clone Manager software.

### 4.13.9 Sequencing analysis

The sequences obtained from DNA sequencing were aligned with the sequences desired using the BioEdit sequence alignment editor and the confirmed sequences were selected.

## 4.14 Introduction of nucleic acids into target cells

### 4.14.1 Transformation of *E. coli* cells DH5α by heat shock method

The competent cells were taken out of the -80 °C freezer and let thaw on ice. A reasonable amount of plasmid DNA was added into a new tube (1-10 ng should be enough, however, the concentration was varied upon the experiment requirements). 200 µl of cells were added to the plasmid DNA. The mixture was incubated on ice for 30 minutes. The cells were then heat shocked at 42 °C for 40 seconds. Straightaway after heat shock, 400 µl of SOC medium was added to each reaction. This was incubated on shaker for 45 minutes at 37 °C and 1.000 rpm. After the incubation, the samples were centrifuged for 30 seconds and the medium was poured away so that there was about 200 µl left. The cells were than resuspended in the left medium and dispersed on an LB plate with proper antibiotic. All the plasmids used in this thesis contain an ampicillin resistance gene.

### 4.14.2 Transfection with plasmid DNA using TurboFect™ Transfection Reagent

The cells were plated so that at time of transfection they were ~ 50 % confluent (equals plating ~ 40.000 cells per well in a 24-well plate on day 1, transfection on day 2). No-FBS DMEM was preheated to 37 °C. For single reaction 100 µl of no-FBS DMEM was mixed with appropriate amount of DNA (for UGA and sense control reporters 200 ng of DNA was used) and vortexed for 10 seconds. Next the appropriate amount of TurboFect™ Transfection Reagent was added into each reaction and vortexed thoroughly for 10-15 seconds to make sure the transfection reagent in properly dispersed in tube (1 µl of TurboFect™ added to 500 ng of DNA). The mixture was incubated for 20 minutes at room temperature. After incubation, 100 µl of transfection mixture was added per well to a 24-well plate, drop by drop to cover the full surface of the well. The suspension was mixed gently by moving the plate carefully. The cells were then incubated for 24 hours at 37 °C in a humidified incubator with 5 % $CO_2$. After 24-hour-incubation, the cells were handled according the protocol for readthrough measurement.

## 4.15 Protein manipulation

### 4.15.1 Whole cell extract preparation for analysis by Western blot

When harvesting cells, three analogous whole cell extract preparation protocols were used during this study. Direct lysis on plate using 1x Glo Lysis Buffer (Promega), direct lysis on plate with 1x Loading Buffer (for WB only) and whole cell extract preparation using centrifugation. The way of whole cell extract preparation did not affect the specificity or sensitivity of antibody binding during Western blot.

#### 4.15.1.1 Direct lysis on plate using Glo Lysis Buffer (GLB), 1X (Promega)

The cells were plated so that at the time of the whole cell extract preparation they were ~ 80 % confluent (equals plating ~ 40.000 cells/well in a 24-well plate on day 1, transfection on day 2, harvesting day 3). 1x Glo Lysis Buffer (Promega) was equilibrated to room temperature prior to experiment. The centrifuge was cooled down to 4 °C. The medium was aspirated away from the cells. Next 200 µl of the 1x Glo Lysis Buffer (Promega) was added. The cells were being lysed on plate for 5 minutes at 25 °C while shaking at 550 rpm. The suspension was transferred to a new 1.5 ml Eppendorf tube already kept on ice. From this moment on, the lysate was kept on ice. The sample was then centrifuged for 5 minutes at ~ 16.000 g and 4 °C. The supernatant (about 180 µl) was transferred to a new 1.5 ml Eppendorf tube and was handled according to the SDS-PAGE sample preparation protocol (4.15.2) prior to Western blot.

#### 4.15.1.2 Direct lysis on plate using 1x Loading Buffer (for Western blot only)

1x Loading Buffer was prepared by mixing 200 µl of 2x Loading Buffer with 200 µl of 1x PBS. Next, 200 µl of 1x Loading Buffer was added into each well of a plate and mixed thoroughly. The samples were then transferred to a new 1.5 ml Eppendorf tube and incubated for 5 minutes at 95 °C prior to SDS-PAGE. Alternatively, the samples can be stored at -20 °C.

#### 4.15.1.3 Lysate preparation using precleaning centrifugation step

The lysis buffer was prepared by mixing all the components at room temperature according to the table below (Table 6). First the medium was aspirated away from the cells. The cells were carefully washed with 200 µl of room tempered 1x PBS. The 1x PBS was also aspirated away. Next 200 µl of the lysis buffer was added, the cells were scraped off with the cell scraper and the lysate was transferred to a new 1.5 ml Eppendorf tube. The sample was incubated on ice for 5 minutes and vortexed 3 times in between. Next the sample was centrifuged for 5 minutes at 4 °C and ~ 16.000 g. The supernatant was transferred to a new 1.5 ml Eppendorf tube kept on ice. The protein concentration was then measured using the Bradford reagent (see 4.15.5) and sample was handled according to protocol for SDS-PAGE sample preparation (see 4.15.2).

**Table 6. The composition of the lysis buffer.**

| Reagent | Amount |
|---|---|
| 1x GA HEPES | 5 ml |
| 1M DTT | 5 µl |
| Aprotinin (1 mg/ml) | 5 µl |
| Leupeptin (1 mg/ml) | 5 µl |
| Pepstatin (1 mg/ml in methanol) | 5 µl |
| PMSF (100 mM in isopropanol) | 50 µl |
| 20 % Triton X-100 | 250 µl |
| cOmplete™, EDTA-free Protease Inhibitor Mix tablets (Roche) | 1 tablet |

### 4.15.2 SDS-PAGE sample preparation

The whole cell extract samples for SDS-PAGE were mixed with 4x Sample loading buffer and incubated for 5 minutes at 95 °C.

### 4.15.3 SDS-PAGE

Criterion™ TGX™ Precast Gel (4-20 %, 1.0 mm, 12+2 Well Comb, 45 µl (Bio-Rad)) was used for the SDS-PAGE. The gel in a plastic storage tray was handled according to the manual of the vendor. The wells of the gel were washed with distilled water to remove all possible non-polymerized polyacrylamide first. The gel was then put to a tank for SDS-PAGE and the tank was filled with SDS-PAGE running buffer. Next the visible marker (Plus Protein™ Standards Dual Color, Bio-Rad) was loaded. The marker volume was loaded in ratio 2:1, at the start and at the end of the gel, respectively (to be able to tell the orientation of the gel/membrane afterwards). Preheated samples were loaded next with maximum volume 30 µl. The SDS-PAGE was run at voltage 200 V for about 60 minutes until the 10 kDa marker reached the gel bottom.

### 4.15.4 Western blot

#### 4.15.4.1 Transfer to nitrocellulose membrane

The ice-cold Blotting buffer was poured to a tray and the blotting sponges soaked wet. The membrane was placed in a different tray with blotting buffer as well. After the SDS-PAGE run finished, the gel was put onto white pieces of chromatography paper (Whatman), the membrane was put to a gel and another layer of Whatman paper was placed to the membrane. This was then placed in between the soaked sponges and the whole sandwich with the sponges was then into the blotting device (cathode, plastic net, sponge, Whatman paper, gel, nitrocellulose membrane, Whatman paper, 2 sponges, plastic net, anode). It is important to make sure the whole sandwich is firmly tightened to prevent moving of the gel/membrane and get rid of all air bubbles, which could impede the successful transfer. The blotting tank was filled with cooled blotting buffer. The blotting device was put to the cold-room and the blotting was run at 25 V with maximum of 2 A for 1.5 hour. After the blotting was finished, the blotting device was reassembled, and the membrane was carefully washed in protein-free TBS-T for 10 minutes to remove possible residual methanol. We usually do this washing step in our laboratory, because we empirically found that some antibodies provide worse outcomes without first washing the methanol off. After the washing step, the membrane was incubated in a blocking medium (5 % low fat milk or BSA in TBS-T) for 1 hour. Blocking of the membrane is an important step, because the proteins from blocking medium bind to the whole surface of

the membrane and prevent from non-specific antibody binding. After blocking, the membrane was again washed in TBS-T for 10 minutes. When probing with different antibodies, the membrane was cut specifically. After this step, the membrane strips were sealed in a plastic foil and incubated with specifically diluted primary antibodies overnight on a shaker in a cold-room at 4 °C. To remove all non-specifically bound antibodies the membrane strips were washed in high volume of TBS-T twice, each time for 10 minutes. Next the membrane strips were sealed to a plastic foil again, and incubated for one hour, with secondary antibodies this time (diluted 1:10.000 in TBS-T). After the one-hour-incubation, the membrane strips were washed 3 times for 5 minutes in TBS-T to remove all non-specifically bound secondary antibodies. When handling smaller proteins, it is better not to wash too excessively, because it could also wash our proteins away. Thus, we wash for a shorter period of time in larger volume of TBS-T.

#### 4.15.4.2 Chemiluminiscent detection

The membrane strips were incubated with the SuperSignal West Femto Maximum Sensitivity Luminol and SuperSignal West Femto Maximum Sensitivity Stable Peroxidase Buffer (1:1 solution) for 1 minute. Pictures of gradually increasing exposures were taken using the G:BOX iChemi device and interpreted using PC. When selecting pictures to be presented, the oversaturated pixels in consecutive exposures were detected first using SnapGene® programme make sure the picture was as much informative as possible.

### 4.15.5 Protein concentration measurement using Bradford reagent

For measurement of protein concentration, 1 µl of the whole cell extract was added to 1 ml of the Bradford reagent and the concentration of protein was measured at 550 nm using Beckman Coulter DU® 530 Life Science UV/VIS Spectrophotometer.

## 4.16 Dual Luciferase Assay System

### 4.16.1 Readthrough measurement using Bright-Glo™ and *Renilla*-Glo™ Luciferase Assay Systems (Promega)

The cells were grown in a 24-well plate so that at time of readthrough measurement they were ~ 80 % confluent (equals plating ~ 40 000 cells per well on day 1, transfection on day 2, readthrough measurement on day 3). The readthrough measurement was performed 24 hours post-transfection (for protocol of transfection see 4.14.2). The Bright-Glo™ Luciferase Assay Reagent (Promega) was taken out of the -80 °C freezer first and let thaw to gain room temperature. Both *Renilla*-Glo™ Luciferase Assay Buffer (Promega) and 1x Glo Lysis Buffer (Promega) were taken out of the fridge and let stand to gain room temperature. Next, the medium was aspirated away from each well of the plate and 200 µl of 1x Glo Lysis Buffer (Promega) was added with a repeater pipette into each well of a plate. The plate was then incubated at 25 °C for 5 min while shaking at 550 rpm. In between, two white Nunc™ MicroWell™ 96-well plates were prepared (Thermo Scientific™), one for Firefly, one for *Renilla* luminiscence measurement. The cell lysate suspension was then transferred to desired wells of a 96-well-plate. The volume was varied according to the plasmid context – 40 µl of suspension for the pSGDluc vectors, where the readthrough values are high, 80 µl for the p2luci vectors, where the readthrough values are lower. *Renilla*-Glo™ Luciferase Assay Buffer (Promega) was mixed with *Renilla*-Glo™ Luciferase Assay Substrate (kept on -20 °C) in ratio 100:1. The *Renilla* mix was then added into each well of the *Renilla* plate (in ratio 1:1 Renilla mix to cell suspension in well; i.e. for single well 80 µl of *Renilla* mix added to 80 µl of cell lysate suspension). The *Renilla* plate was incubated for 15 minutes at 25 °C and 350 rpm. Meantime Brigh-Glo™ Luciferase Assay Reagent (Promega) was added into each well of the Firefly plate in ratio 1:1 Bright-Glo™ to cell suspension in well just like with *Renilla*. The Firefly plate was incubated for 5 minutes at 25 °C and 350 rpm. The luminescence was measured straightaway using Infinite® 200 PRO (TECAN) and i-control™ software.

### 4.16.2 Readthrough analysis

The reporter used here consists of the *Renilla* gene, a test tRNA sequence (or control tRNA sequence) and the Firefly gene. The Firefly gene is only expressed when stop codon readthrough occurs. The readthrough efficiency is then estimated by using following formula:

$$Readthrough\ (\%) = \frac{\frac{F\ stop}{R\ stop}}{\frac{F\ sense}{R\ sense}} \ x\ 100$$

The Firefly activities are divided by the *Renilla* activities, for both the sense and the stop codon. Normalization to *Renilla* helps overcome possible discrepancies caused by different cell viability, transfection efficiency or errors in pipetting. The normalized value for stop codon is then divided by the normalized value for sense codon. To get the final readthrough efficiency in %, the outcoming value is multiplied by 100.

# 5 RESULTS

## 5.1 Selection of tRNA-Cys isodecoders from bioinformatically predicted human tRNA-Cys genes

The initial aim of this thesis was to choose a human cysteine tRNA molecule that is highly expressed and translationally active *in vivo.* Two isoacceptors constitute the tRNA-Cys family – with ACA and GCA anticodon, of which there exist 1 and 23 isodecoders, respectively. The isodecoders' mature tRNA sequences differ only slightly, however, these nucleotide changes, although minor, may have a role in overall tRNA expression rate or even stop codon recognition. Thus, we decided to select several isodecoders as a substitute for an ideal cysteine tRNA proxy.

The sequences of cysteine tRNA genes were taken from Genomic tRNA Database ; GtRNAdb 2.0 (Chan and Lowe 2016). The GtRNAdb 2.0 uses the tRNAscan-SE 2.0 search tool to detect and functionally predict tRNA genes and annotate them in genomes (Chan and Lowe 2016, Lowe and Chan 2016, Chan et al. 2019). Since all of these tRNAs are only bioinformatically predicted, it is possible that some of the isoforms may not be highly expressed or actually existing *in vivo.* On the other hand, to support the trustworthiness of this database, it was experimentally shown using ChIP-Seq (Pol III) that all of the experimentally detected human tRNAs have been computationally predicted in GtRNAdb before (224 ChIP-Seq detected / 625 total predicted) (Kutter et al. 2011).

First, mature tRNAscan-SE score was taken into account – which, generally, when higher than 50, may imply the tRNA is an existing one and forms a secondary structure; when too low, may indicate a pseudogene (Pan 2018). Indeed, a recent tRNA-Seq study showed that in HEK293T cells the tRNAs with tRNAscan-SE scores above 55 contributed the most to the total tRNA pool (Torres et al. 2019). All of the 23 tRNA-Cys-GCA isodecoders have the mature tRNA score higher than 50. The ACA isoacceptor, and in this case also the only isodecoder, with its mature tRNA score 46.2, has an intron inside the sequence and is classified as pseudogene today.

Another feafure that could be taken into account is the “rank of isodecoder” section in GtRNAdb which shows how high particular tRNA scores relative to all other isodecoders in the isodecoder family (Chan and Lowe 2016). The rank of isodecoder can be directly told from its gene name, for cysteine tRNA isodecoders, tRNA-Cys-GCA-1-1 being the highest ranking one.

To assess the expression of tRNA-Cys isodecoders, three publicly available expression datasets were analysed. The analysis of Solexa Illumina HiSeq 2000 small RNA-Seq gene counts in HEK293T cells, accession no. GSE114904 (Torres et al. 2019) is presented in Supplementary Table 1. The analysis of the data from Polymerase III (Pol III) chromatin immunoprecipitation followed by sequencing (ChIP-Seq) from liver cancer cell lines HepG2, Huh7 and healthy human liver, accession no. E-MTAB-958 (Rudolph et al. 2016; the tRNA expression values are publicly available online in Table S4: https://doi.org/10.1371/journal.pgen.1006024.s016) is presented in Supplementary Table 2. The analysis of YAMAT-seq (Y-shaped Adapter-ligated MAture tRNA sequencing) in breast cancer cell lines MCF-7, SK-BR-3 and BT-20, accession no. SRP096584 (Shigematsu et al. 2017) is presented in Supplementary Table 3. The tRNA gene counts expressed as RPM (reads per milion) were divided into five groups according to their expression relative to the whole data set, 100 % being the isodecoder with the highest expression values. These were then classified into five groups (marked with a plus sign, “+”): “+” meaning 0-20 %; “++” meaning 20-40 %; “+++” meaning 40-60 %; “++++” meaning 60-80 %; “+++++” meaning 80-100 %. The results were then compared for all the three datasets and the highest scoring isodecoders were selected to be further tested experimentally. The Table 7 summarizes the isodecoders selected acording to the analysis described above.

The expression of isodecoder tRNA genes relative to their whole isodecoder set (based on percentage of sequencing reads) was bioinformatically evaluated by applying a numerical analysis, isodecoder-specific tRNA gene contribution profile (iso-tRNA-CP), to several tRNA-Seq datasets from HEK293T cells. Overall, this showed (consistently with our analysis), that tRNA-Cys-GCA-2 (tC06) and tRNA-Cys-GCA-4-1 (tC10) contribute the most to the cysteine tRNA isodecoder dataset. Iso-tRNA-CP also showed that tRNA expression is cell-type and tissue specific, comparing tRNA-Seq datasets from the HEK293T cells and human brain (Torres et al. 2019).

**Table 7. The tRNA-Cys isodecoders selected by the expression analysis which were further used experimentally**. The table shows the proportion of individual tRNA-Cys isodecoders expression relative to the whole isodecoder set of tRNA-Cys-GCA. ND – Not detected. The range of expression is indicated by a plus sign ("+"), "+++++" meaning the highest expression. The tC06 sequence can be found in the human genome for four identical gene copies (tRNA-Cys-GCA-2-1, tRNA-Cys-GCA-2-2, tRNA-Cys-GCA-2-3, tRNA-Cys-GCA-2-4), exactly the same as for the tC09 sequence (tRNA-Cys-GCA-9-1, tRNA-Cys-GCA-9-2, tRNA-Cys-GCA-9-3, tRNA-Cys-GCA-9-4).

| Thesis name | Gene name | Tores, 2019 | Kutter, 2016 | | | Shigematsu, 2017 | | |
|---|---|---|---|---|---|---|---|---|
| | | HEK293T | HepG2 | Huh7 | Liver-Adult | MCF-7 | SKBR-3 | BT-20 |
| tC06 | tRNA-Cys-GCA-2-1 | +++ | ND | ND | ND | +++++ | +++++ | ++++ |
| | tRNA-Cys-GCA-2-2 | +++++ | ++ | +++ | ++ | | | |
| | tRNA-Cys-GCA-2-3 | + | +++++ | +++++ | +++++ | | | |
| | tRNA-Cys-GCA-2-4 | ++ | ND | ND | ND | | | |
| tC07 | tRNA-Cys-GCA-14-1 | + | +++ | ++++ | +++ | + | + | + |
| tC08 | tRNA-Cys-GCA-17-1 | +++++ | +++ | +++++ | ++ | ND | ND | ND |
| tC09 | tRNA-Cys-GCA-9-1 | + | ND | ND | ND | + | + | + |
| | tRNA-Cys-GCA-9-2 | ++ | + | + | + | | | |
| | tRNA-Cys-GCA-9-3 | ++ | ND | ND | ND | | | |
| | tRNA-Cys-GCA-9-4 | ++ | + | + | + | | | |
| tC10 | tRNA-Cys-GCA-4-1 | + | ++++ | +++++ | ++++ | +++ | + | ++ |

Four more isodecoders were tested experimentally, summarized in Table 8, which were used in our lab prior to this study and were selected by a different approach, an analysis of the sequence and the secondary structure. It is obvious that not all of these isodecoders score high in the expression analysis. tC01 was selected as a mammalian representative of 72-nt tRNA-Cys, tC02 as a mammalian representative of 73-nt tRNA-Cys. tC03 was selected for its distinct sequence, structure and presence of an intron. tC04 was selected for its distinct sequence and structure. A full list of all of the analysed isodecoders and their expression scores can be found in Supplemetary Table 5. The sequences of all selected tRNA-Cys isodecoders are aligned in Supplementary Figure 1.

**Table 8. The tRNA-Cys isodecoders selected by the sequence analysis and further used experimentally**. The table shows the proportion of individual tRNA-Cys-GCA (ACA) isodecoders expression relative to the whole isodecoder set of tRNA-Cys-GCA (ACA). ND – Not detected. The range of expression indicated by a plus sign ("+"), "+++++" meaning the highest expression.

| Thesis name | Gene name | Tores, 2019 | Kutter, 2016 | | | Shigematsu, 2017 | | |
|---|---|---|---|---|---|---|---|---|
| | | HEK293T | HepG2 | Huh7 | Liver-Adult | MCF-7 | SKBR-3 | BT-20 |
| tC01 | tRNA-Cys-GCA-8-1 | + | + | + | + | + | + | + |
| tC02 | tRNA-Cys-GCA-5-1 | ++ | + | + | + | ++++ | + | ++ |
| tC03 | tRNA-Cys-ACA-1-1 | + | ND | ND | ND | ND | ND | ND |
| tC04 | tRNA-Cys-GCA-22-1 | + | ND | ND | ND | ND | ND | ND |

## 5.2 A and B Box sequence comparison of human tRNA-Cys isodecoder genes

As mentioned before, multiple sequence differences among human isodecoder genes are located in the A Box and B Box sequence which could account for isodecoders' different expression rates in tissues or developmental stages (Goodenbour and Pan 2006). To investigate if there is a correlation between differences in the A Box and B Box sequences of tRNA-Cys isodecoders and their expression (in other words: if the highly/lowly expressed genes share similarities or differences in the A and B Box sequences), the sequences of the five most highly and five least expressed isodecoders as well as the A and B Box sequences of all tRNA-Cys-GCA isodecoder genes were compared using online WebLogo generator (Crooks et al. 2004). The sequence alignment of the highly expressed isodecoder genes can be found in Supplementary Figure 2 and of the least expressed isodecoder genes in Supplementary Figure 3. The logos are depicted in Fig. 9 and the detailed description follows:

The prescribed consensus sequence for the A Box is: TRGYNNARNNG, where N means any nucleotide, R means purine (G/A) and Y means pyrimidine (C/T). In the A Box logo prepared from all tRNA-Cys-GCA isodecoder genes, we can see there is a variation especially in the A Box sequence position 8 and 9 (Fig. 9c). For the highly expressed genes, in the A Box position 9, there is T in the tRNA-Cys-GCA-2 (tC06) and tRNA-Cys-GCA-4-1 (tC10) gene (Fig. 9a), which is never present in this position in any of the least expressed tRNA-Cys-GCA genes (Fig. 9b). Generally, the A Box position 8 allows incorporation of purine (G/A), however, in the logo prepared from the least

expressed isodecoders (Fig. 9b), there is C in this position in tRNA-Cys-GCA-23-1; we can say that the A Box logo of the least expressed tRNA-Cys-GCA isodecoders slightly differs from the prescribed consensus sequence. In contrast, the highly expressed genes always precisely follow the prescribed A Box sequence and have "T" in the A Box sequence position 9 (the importance of which will be further discussed in the next chapters).

The prescribed consensus sequence for the B Box is: RGTTCRANNCY. When the B Box sequences of all tRNA-Cys-GCA isodecoder genes are compared, we see there is variability especially in the B Box sequence positions 1, 2 and 11 (Fig. 9f). The B Box sequence of all the highly expressed genes is identical and follows canonical prescribed sequence (Fig. 9d). The B Box logo is more variable when the lowly expressed genes are aligned (Fig. 9e); this logo is very similar to the one prepared from all the cysteine isodecoder genes. Notably, position 2 of the B Box sequence should always be G, however, lowly expressed tRNA-Cys-GCA-19-1 shows A in this position. To conclude, we can say that the highly expressed genes precisely follow the prescribed sequence for the B Box (the sequence is identical) and that any variability in this sequence may correspond with the isodecoders' different expression rates.

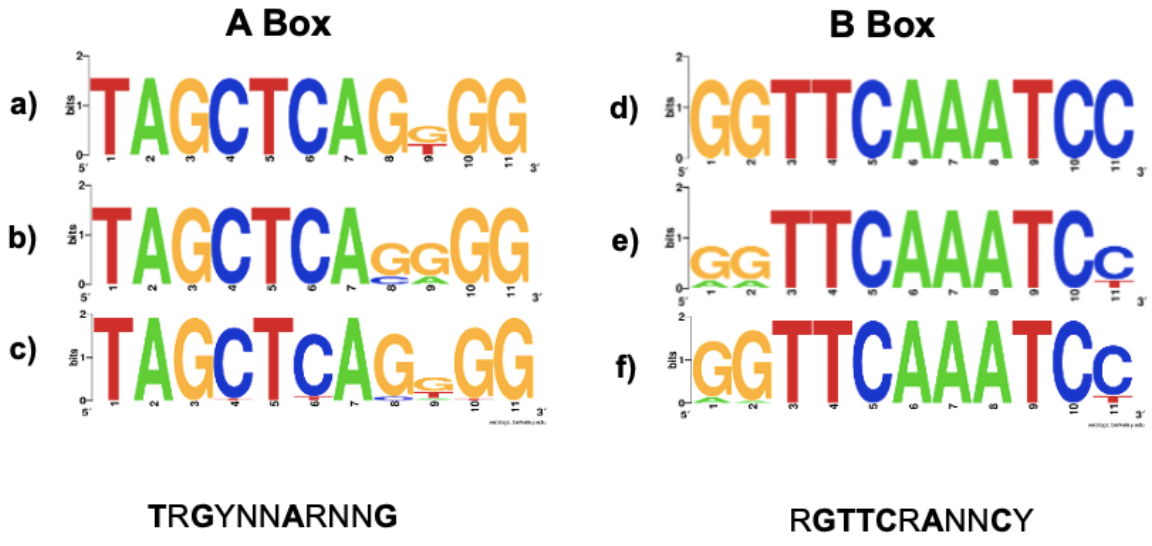



**Figure 9. The A Box and the B box sequence logo comparison for tRNA-Cys-GCA isodecoder genes. a)** The A Box logo prepared from the five most highly expressed cysteine tRNA genes: tRNA-Cys-GCA-2, tRNA-Cys-GCA-4-1, tRNA-Cys-GCA-9, tRNA-Cys-GA-14-1 and tRNA-Cys-GCA-17-1 genes. **b)** The A Box logo prepared from the least expressed cysteine tRNA genes: tRNA-Cys-GCA-3-1, tRNA-Cys-GCA-10-1, tRNA-Cys-GCA-19-1, tRNA-Cys-GCA-22-1

and tRNA-Cys-GCA-23-1. **c)** The A Box logo prepared from all human tRNA-Cys genes. **d)** The B Box logo prepared from the five most highly expressed tRNA-Cys genes. **e)** The B Box logo prepared from the least expressed tRNA-Cys genes. **f)** The B Box logo prepared from all human tRNA-Cys genes. The logos were generated using https://weblogo.berkeley.edu/.

Altogether, this data are in line with previous research and show that the sequence differences among different isodecoders located in the A and B Box (and deviations from the prescribed consensus sequence) may correspond with different tRNA isodecoders' expression levels in a cell- and tissue-specific way.

## 5.3 Working with reporter tRNA expression plasmids

### 5.3.1 Description of the p2luci and pSGDluc reporter systems

Two dual-luciferase reporter systems were used in this work, p2luci and pSGDluc. These are constructs bearing two luciferases, *Renilla* and Firefly, separated by a test sequence containing UGA stop codon, or, in control purposes, an in-frame CAG sense codon replacing the stop codon in case of p2luci or an in-frame UGG sense codon in case of pSGDluc. The synthesis of Firefly luciferase is dependent on bypassing the stop codon *via* readthrough – only then is the Firefly luciferase expressed. The reporters used in this thesis are slightly modified; they contain a U6 promoter and the specific tested tRNA gene is subcloned under this promoter (Fig. 10, Beznoskova et al. 2021). The U6 promoter was used here because it was experimentally shown in HEK293T to enhance the expression of some tRNAs (Koukuntla et al. 2013). In the reporter systems used here, the stop codon (and sense codon in control purposes) is followed by a nucleotide context known to promote readthrough, in p2luci the Tobacco Mosaic Virus (TMV) (Pelham 1978) context and in pSGDluc the CUAG context (Loughran et al. 2017).


**Figure 10. A schematic description of the reporter used in this study.** The reporter codes for both the *Renilla* and Firefly luciferases and the tested tRNA on a single construct. A cassette containing a DNA sequence of a mature tRNA under the U6 promoter (or without any tRNA in control purposes) can be specifically cut within the A and B multiple cloning sites (MCS) and inserted into a vector of interest with corresponding restriction sites (Beznoskova et al. 2021).

The p2luci is the first plasmid construct described that allows measuring activities of both *Renilla* and Firefly luciferases in a single-tube assay. It was generated from modified pRL-SV40 plasmid. In p2luci, Firefly luciferase gene is placed in-frame after *Renilla* luciferase gene and an alternative polylinker is inserted between the *Renilla* and Firefly genes. The p2luci plasmid has the major expression features of pRL-SV40: a SV40 early enhancer/promoter to initiate transcription in eukaryotic cells, a T7 promoter in front of the *Renilla* luciferase gene, a blunt end restriction site behind the Firefly luciferase gene, a chimeric intron to increase transport of mRNA into cytoplasm, a SV40 late polyadenylation signal (Grentzmann et al. 1998). When the stop/sense codon between the two luciferases is decoded, a fused protein is created.

Since it is possible that the fused product may cause distortions of the separate reporter activities, a pSGDluc vector was created, that allows continuous translation of unfused proteins (Loughran et al. 2017). The pSGDluc was constructed from a pDluc plasmid (a derivative of p2luci) (Fixsen and Howard 2010). The pSGDluc uses a viral StopGo, 33 amino acid sequence ending with P-G-P motif (proline-glycine-proline). When StopGo is encoded within a single long ORF, it induces translation of two separate polypeptides. The synthesis of a peptide bond within the Gly-Pro sequence is disengaged with a mechanism proposed, in which the interaction of the nascent polypeptide with the exit tunnel of the ribosome inhibits peptide bond formation and the nascent polypeptide is released – translation then continues at the proline codon ("pseudo-reinitiation") (Brown and Ryan 2010). The pSGDluc has two StopGo sequences - one at the 3'end of the *Renilla* luciferase gene and second at the 5' end of the Firefly luciferase gene. The main features of pSGDluc remain the same as of the p2luci vector: a SV40 early enhancer/promoter, a T7 promoter, a SV40 late polyadenylation signal (Loughran et al. 2017).

### 5.3.2 Construction of tRNA-Cys reporter constructs using the pSGDluc vector

The inserts bearing the tRNA under the U6 promoter were prepared using the GeneArt™ Strings™ DNA Fragments (see 4.8.1). Each of the DNA fragment was first digested by SalI-BamHI and the 387-bp product was then purified using the QIAquick® PCR Purification Kit (Qiagen®) (see 4.13.7). The inserts were ligated into B1145/B1146 vectors in 1:3 vector:insert molar ratio according to the protocol for ligation (see 4.13.4). The ligation reaction was then transformed into *E. coli* strain DH5α according to the protocol for heat shock transformation (see 4.14.1) and the cells were spread onto an LB-Amp plate and grown for 24 hours. Selected clones were then inoculated into a liquid LB-Amp medium and bacteria was grown for 16 hours prior to plasmid DNA isolation. The isolation was then performed using the QIAprep® Spin Miniprep Kit (Qiagen®) (see 4.13.1). After the DNA isolation, the subcloning was confirmed by a specific restriction reaction (see 4.13.3). For inserts subcloned into B1145 and B1146 plasmids, the products were control-digested by PstI enzyme with expected fragment sizes 5074 bp and 1164 bp (Figs. 11 and 12). The plasmids confirmed by restriction digest were then verified by sequencing (see 4.13.8). The full list of pSGDluc plasmids created for the purpose of this study is shown in table below (Table 9).

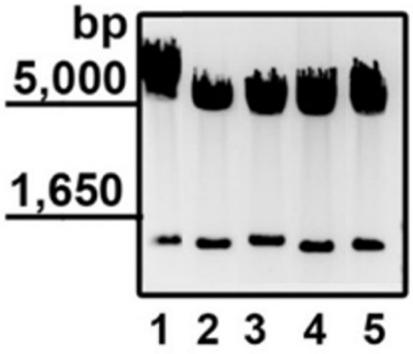



**Figure 11. Confirmation of subcloning of desired tRNA-Cys inserts into pSGDluc-UGA (B1145) by restriction digest.** The reaction was performed using PstI enzyme with expected fragment sizes 5074 bp and 1164 bp. Electrophoresis was performed with ~ 400 ng of DNA run on a 0.8 % agarose gel at 5 V/cm. Lanes: 1 – MK1, 2 – MK3, 3 – MK4, 4 – MK5, 5 – MK7.

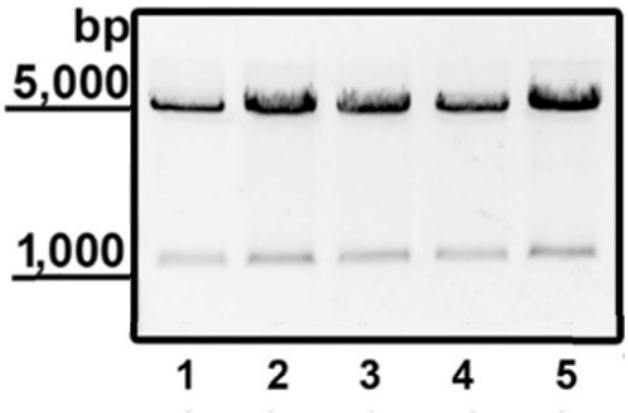



**Figure 12. Confirmation of subcloning of desired tRNA-Cys inserts into pSGDluc-UGG (B1146) by restriction digest.** The reaction was performed using PstI enzyme with expected fragment sizes 5074 bp and 1164 bp. Electrophoresis was performed with ~ 400 ng of DNA run on a 0.8 % agarose gel at 5 V/cm. Lanes: 1 – MK9, 2 – MK11, 3 – MK13, 4 – MK15, 5 –MK17.

**Table 9: The full list of plasmids created for the purpose of this study with tRNA-Cys inserts subcloned into pSGDluc-UGA (B1145) and pSGDluc-UGG (B1146).** pMK1 and pMK9 were prepared by inserting the 387-bp SalI-BamHI fragment from GeneArt™ Strings™ DNA Fragments tC06 sequence (for tRNA-Cys-GCA-2) into SalI-BamHI digested B1145 and B1146, respectively. The other vectors were prepared accordingly – pMK3 and pMK11 with GeneArt™ Strings™ DNA Fragments tC07 sequence (for tRNA-Cys-GCA-14-1); pMK4 and pMK13 with GeneArt™ Strings™ DNA Fragments tC08 sequence (for tRNA-Cys-GCA-17-1); pMK5 and pMK15 with GeneArt™ Strings™ DNA Fragments tC09 sequence (for tRNA-Cys-GCA-9); pMK7 and pMK17 with GeneArt™ Strings™ DNA Fragments tC10 sequence (for tRNA-Cys-GCA-4-1).

| **Plasmid name** | **Plasmid description** |
|---|---|
| pMK1 | pSGDluc-UGA-CUAG-U6-tC06 |
| pMK9 | pSGDluc-UGG-CUAG-U6-tC06 |
| pMK3 | pSGDluc-UGA-CUAG-U6-tC07 |
| pMK11 | pSGDluc-UGG-CUAG-U6-tC07 |
| pMK4 | pSGDluc-UGA-CUAG-U6-tC08 |
| pMK13 | pSGDluc-UGG-CUAG-U6-tC08 |
| pMK5 | pSGDluc-UGA-CUAG-U6-tC09 |
| pMK15 | pSGDluc-UGG-CUAG-U6-tC09 |
| pMK7 | pSGDluc-UGA-CUAG-U6-tC10 |
| pMK17 | pSGDluc-UGG-CUAG-U6-tC10 |

### 5.3.3 Construction of tRNA-Cys reporter constructs using the p2luci vector

The inserts bearing the tRNA under the U6 promoter were prepared by digesting a plasmid (with our desired insert sequence) that was created before for the B1145/B1146 vectors (Table 8). The selected plasmids were first digested by MscI-KpnI and the restriction reaction was then confirmed by electrophoresis. Subsequently, the expected 351-bp fragment was excised from the gel and the DNA was isolated according to the protocol for isolation of DNA from gel (see 4.13.6) using the QIAquick® Gel extraction Kit (Qiagen®). The inserts were then ligated into PBB171/PBB172 vectors in 1:3 vector:insert molar ratio according to the protocol for ligation (see 4.13.4). The ligation reaction was transformed into *E. coli* strain DH5α according to the protocol for heat shock transformation (see 4.14.1) and the cells were spread onto an LB-Amp plate and grown for 24 hours. Selected clones were then inoculated into a liquid LB-Amp medium and bacteria was grown for 16 hours prior to plasmid DNA isolation. The isolation was then performed using the QIAprep® Spin Miniprep Kit (Qiagen®) (see 4.13.1). After isolation of the DNA, the subcloning was confirmed by specific restriction reaction (see 4.13.3). For inserts subcloned into PBB171 and PBB172, the products were control-digested by PstI with the expected fragment sizes 3576 bp and 2210 bp (Fig. 13). The plasmids confirmed by restriction digest were then verified by sequencing (see 4.13.8). The full list of p2luci plasmids further used in this study is shown in table below (Table 10).

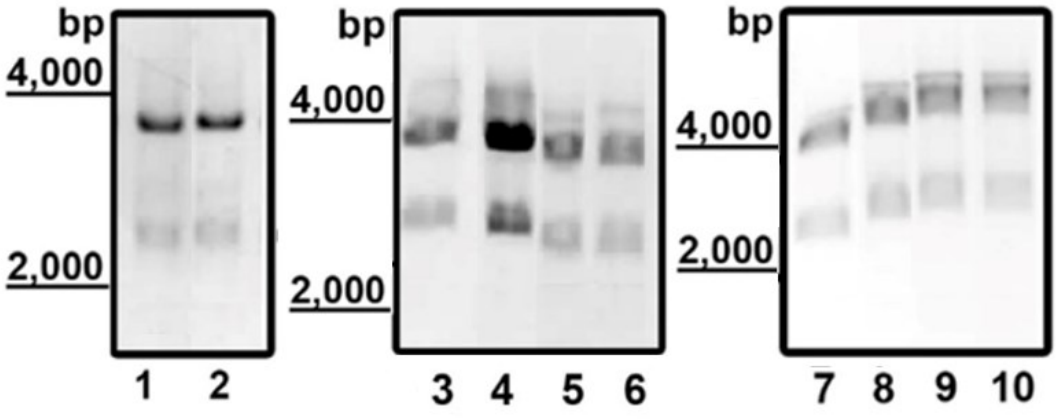



**Figure 13. Confirmation of subcloning of desired tRNA-Cys inserts into p2luci-UGA (PBB171) and p2luci-CAG (PBB172) by restriction digest**. The reaction was performed using PstI enzyme with expected fragment sizes 3576 bp and 2210 bp. Electrophoresis was performed with ~ 400 ng of DNA run on a 0.8 % agarose gel at 5 V/cm. Lanes: 1 – MK20, 2 – MK19, 3 – MK22, 4 – MK24, 5 – MK26, 6 – MK28, 7 – MK32, 8 – MK36, 9 – MK34, 10 – MK30.

**Table 10. The full list of plasmids created for the purpose of this study with tRNA-Cys inserts subcloned into p2luci-UGA (PBB171) and p2luci-CAG (PBB172).** pMK20 and pMK19 were prepared by inserting the 351-bp MscI-KpnI fragment from pMK1 into MscI-KpnI digested PBB171 and PBB172, respectively. The other plasmids were prepared accordingly – pMK24 and pMK22 using a fragment from pMK3; pMK28 and pMK26 using pMK13; pMK32 and pMK30 using pMK5; pMK36 and pMK34 using pMK7.

| Plasmid name | Plasmid description |
|---|---|
| pMK20 | p2luci-UGA-TMV-U6-tC06 |
| pMK19 | p2luci-CAG-TMV-U6-tC06 |
| pMK24 | p2luci-UGA-TMV-U6-tC07 |
| pMK22 | p2luci-CAG-TMV-U6-tC07 |
| pMK28 | p2luci-UGA-TMV-U6-tC08 |
| pMK26 | p2luci-CAG-TMV-U6-tC08 |
| pMK32 | p2luci-UGA-TMV-U6-tC09 |
| pMK30 | p2luci-CAG-TMV-U6-tC09 |
| pMK36 | p2luci-UGA-TMV-U6-tC10 |
| pMK34 | p2luci-CAG-TMV-U6-tC10 |

### 5.3.4 Readthrough measurement of overexpressed tRNA-Cys isodecoders in p2luci

The readthrough results for p2luci constructs will be described before results for pSGDluc constructs because tC01, tC02, tC03 and tC04 isodecoders were subcloned and initially measured in our laboratory before I started the project. To be able to compare the readthrough values they must be measured at the same time. Therefore, tC01, tC02, tC03 and tC04 were remeasured in the same run with further selected isodecoders tC06, tC07, tC08, tC09 and tC10. The measurements were performed as described in protocol for readthrough measurement (see 4.16.1).

However, since each tRNA isoform was subcloned into two constructs (bearing either sense or stop codon) we first decided to investigate if overexpression of any tRNA-Cys isodecoders in the sense control plasmid affects the Firefly to *Renilla* ratio in HEK293T cells. An initial experiment was performed in p2luci, which shown comparable ratio values for all tested tRNAs as well as for the reporter with no tRNA (Fig. 14), proving that tRNA-Cys expressed under U6 promoter does not affect the expression of Firefly-*Renilla*

fusion protein in our reporter. Similar results were observed for both tRNA-Tyr and tRNA-Trp (Beznoskova et al. 2021). Based on this uniform result we decided to only use p2luci-CAG-no_tRNA as our sense codon control for further readthrough estimation. Since each readthrough experiment was designed so that there were biological duplicates for each plasmid and each experiment was repeated at least three times, by using only one sense codon control we saved a lot of material and were able to perform a single repetition with all isodecoders on the same plate.

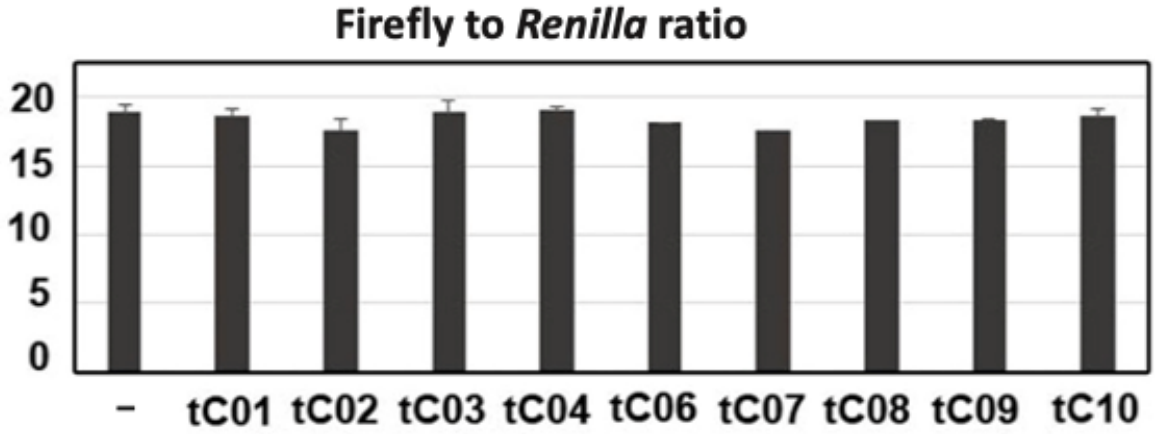



**Figure 14. Overexpression of human tRNA-Cys isodecoders in p2luci reporter bearing CAG sense codon does not affect Firefly to *Renilla* ratio in HEK293T cells**. The F/R ratio values measured for plasmids bearing sense codon are comparable, thus, for further experiments, the exact numerical value for each sense-codon plasmid would be neglected and only the reporter with no tRNA (-) value be used for readthrough calculation.

After we decided to only use p2luci-CAG-no_tRNA as 100 % for calculation of readthrough efficiency of all isodecoders in p2luci, we estimated UGA readthrough according to the protocol described in chapter 4.16.1 and realised that based on the UGA readthrough levels measured in p2luci, the results can be divided into three groups: i) overexpression of tC10 and tC06 increases UGA readthrough; ii) overexpression of tC09, tC01 and tC02 slightly increases UGA readthrough; iii) the mature tRNA sequences of tC07, tC03, tC04 and tC08 placed under the U6 promoter did not change the levels of UGA readthrough. The readthrough results and other interesting features of tRNAs in these groups will be discussed in detail in the upcoming text.

### 5.3.4.1 Overexpression of tC10 and tC06 increases UGA readthrough in HEK293T

tRNA-Cys-GCA-4-1 (tC10) and tRNA-Cys-GCA-2 (tC06) showed significantly increased UGA readthrough levels when overexpressed in HEK293T cells in p2luci (Fig. 15). Overexpressed tC10 showed ~ 6.0 % of UGA readthrough ($P < 0.01$) and overexpressed tC06 showed ~ 4.9 % of UGA readthrough ($P < 0.05$), compared to ~ 3.1 % of UGA readthrough estimated for the reporter with no tRNA (Fig. 15, Supplementary Table 6). The unique features that might be contributing to or directly causing this readthrough effect, together with the additional information we gathered, are described for each isoform separately in the text below.

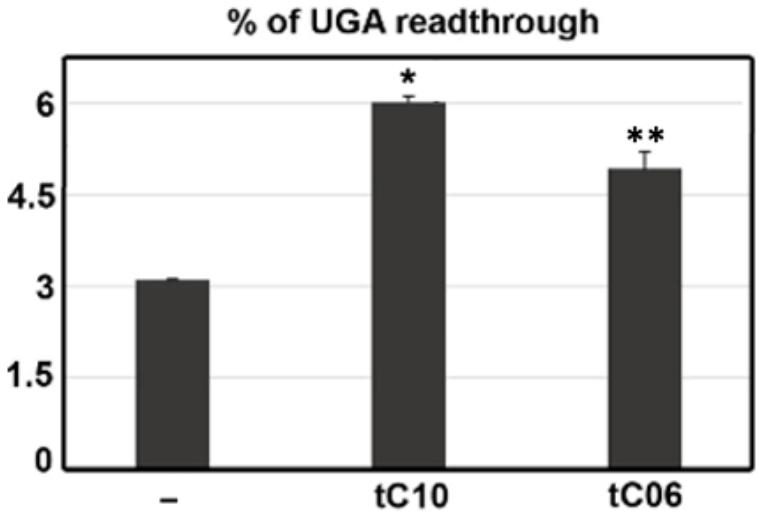



**Figure 15. Overexpression of tC10 and tC06 increases UGA readthrough in HEK293T in p2luci.** The changes in the readthrough levels relative to the reporter with no tRNA (–) were analysed by Student's t-test (mean+SD; n=2); statistically significant increase with $P < 0.01$ is marked with asterisk "*" and with $P < 0.05$ with "**". The readthrough levels for all tested tRNA molecules with their exact numerical values can be found in Supplementary Table 6.

**tRNA-Cys-GCA-4-1 (tC10)**

In p2luci, overexpressed tC10 repeatedly showed the highest UGA readthrough level. tC10 was the second highest ranking isodecoder in ChIP-Seq data analysis (Supplementary Table 2) and the third highest ranking isodecoder in YAMAT-Seq (Supplementary Table 3). Interestingly, in tRNA-Seq data from Torres, 2019, this tRNA showed only a low level of expression (Supplementary Table 1). The sequence and the secondary structure of this isodecoder seem canonical (Fig. 16), showing no unusual residues or introns. The mature tRNA score for this isodecoder is 81.8, the second highest of all the tRNA-Cys isodecoders (Supplementary Table 6). Since overexpression of this isodecoder shown the highest UGA readthrough level, the structures of all the other tRNA-Cys isodecoders tested were compared to this one, demonstrating nucleotide changes that may cause the loss of the readthrough inducing properties.

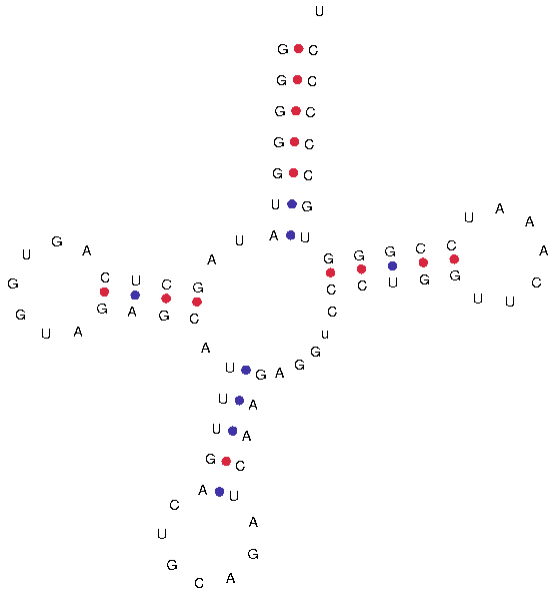


**Figure 16. The sequence and the secondary structure of tRNA-Cys-GCA-4-1 (tC10)**. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

**tRNA-Cys-GCA-2 (tC06)**

In p2luci, overexpressed tC06 showed repeatedly the second highest UGA readthrough level. tC06 was the highest ranking isodecoder in all the three expression analyses (Table 7, Supplementary Table 5). In the human genome, tC06 sequence can be found for four identical gene copies, tRNA-Cys-GCA-2-1, tRNA-Cys-GCA-2-2,

tRNA-Cys-GCA-2-3 and tRNA-Cys-GCA-2-4. The sequence and the secondary structure of this isodecoder (Fig. 17) seem canonical, showing no unusual residues or introns. This isodecoder has the highest mature tRNA score of all tRNA-Cys isodecoders, 81.9 (Supplementary Table 6). Compared to tC10, in the sequence of tC06 there is cytosine instead of uracil in the same position of T-arm (with stronger G:C base pairing in tC06 and G:U base pairing in tC10, compare Fig. 16 and Fig. 17). This change, that keeps the T-arm intact, has only mild effect on readthrough properties, if any.

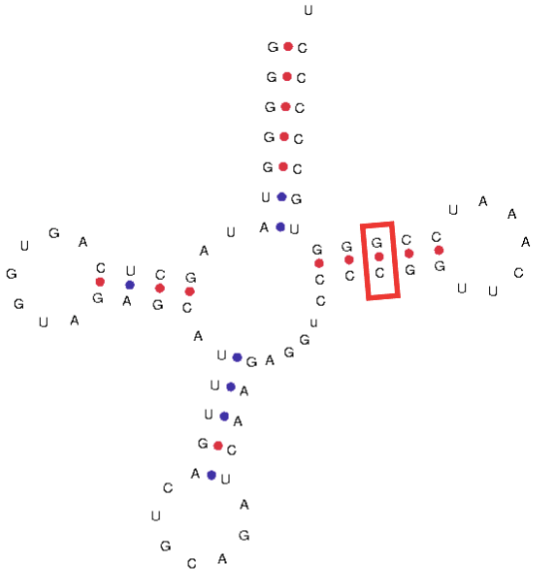


**Figure 17. The sequence and the secondary structure of tRNA-Cys-GCA-2 (tC06).** The difference in base pairing (from tC10) is indicated by a red rectangle. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

Of all tRNA-Cys isodecoders tested, both tC06 and tC10 showed the highest, significantly increased UGA readthrough levels when compared to the reporter with no tRNA. These were also the isodecoders which scored high in the expression analyses and had the highest mature tRNA score as predicted by tRNAscan-SE. In other words: these two tRNAs highly likely are expressed endogenously in human cells and if so, they have the capacity to incorporate at UGA stop codon during translational readthrough.

### 5.3.4.2 Overexpression of tC09, tC01 and tC02 slightly increases UGA readthrough in HEK293T

tRNA-Cys-GCA-9 (tC09), tRNA-Cys-GCA-8-1 (tC01) and tRNA-Cys-GCA-5-1 (tC02) showed slightly increased UGA readthrough levels when overexpressed in HEK293T cells in p2luci (Fig. 18). Overexpressed tC09 showed ~ 4.5 % of UGA readthrough ($P < 0.1$) and overexpressed tC02 showed ~ 4.2 % of UGA readthrough, compared to ~ 3.3 % of UGA readthrough estimated for the reporter with no tRNA (Fig. 18, Supplementary Table 6). Overexpressed tC01 showed ~ 4.0 % of UGA readthrough ($P < 0.1$), compared to ~ 3.1 % of UGA readthrough estimated for the reporter with no tRNA (Fig. 18, Supplementary Table 6).

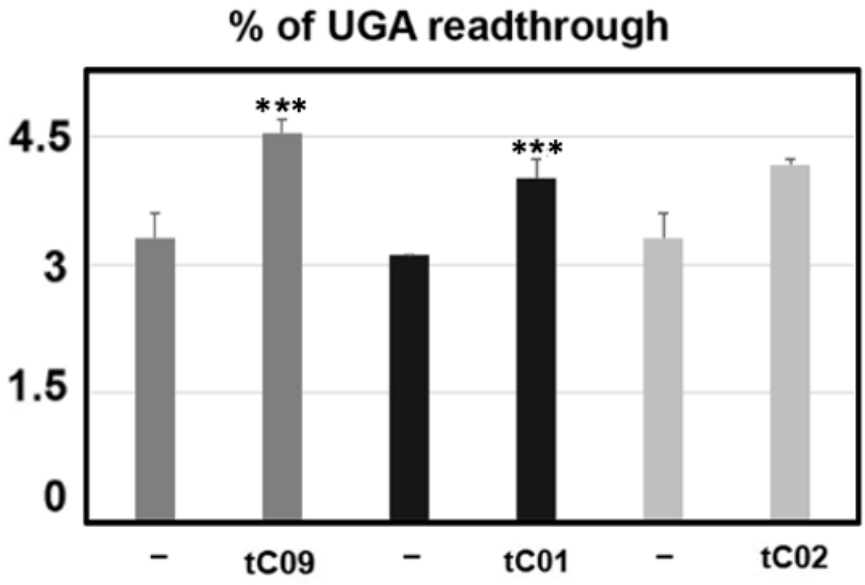



**Figure 18. Overexpression of tC09, tC01 and tC02 slightly increases UGA readthrough in HEK293T in p2luci.** The changes in the readthrough levels relative to the reporter with no tRNA (–) were analysed by Student's t-test (mean+SD; n=2); statistically significant increases with $P < 0.1$ are marked with asterisks "***". The readthrough levels for all tested tRNA molecules with their exact numerical values can be found in Supplementary Table 6.

**tRNA-Cys-GCA-9 (tC09)**

In p2luci, overexpressed tC09 showed slightly increased UGA readthrough level with $P < 0.1$ (Fig. 18, Supplementary Table 6). tC09 was detected in the expression analyses with generally lower levels than the molecules with the high capability of readthrough increase (Table 7, Supplementary Table 5). In the human genome, tC09 sequence can be found for four identical gene copies, tRNA-Cys-GCA-9-1, tRNA-Cys-GCA-9-2, tRNA-Cys-GCA-9-3 and tRNA-Cys-GCA-9-4. The sequence and the secondary structure of this isodecoder (Fig. 19) seem canonical, showing no unusual residues or introns. The mature tRNA score is lower than for tRNAs in the previous group, 77.2. Compared to tC10, in the sequence of tC09 there is guanine instead of uracil in the same position of D-loop and adenine instead of guanine in the same position of T-arm (with A:U base pairing in tC09 and G:U base pairing in tC10; compare Fig. 16 and Fig. 19).

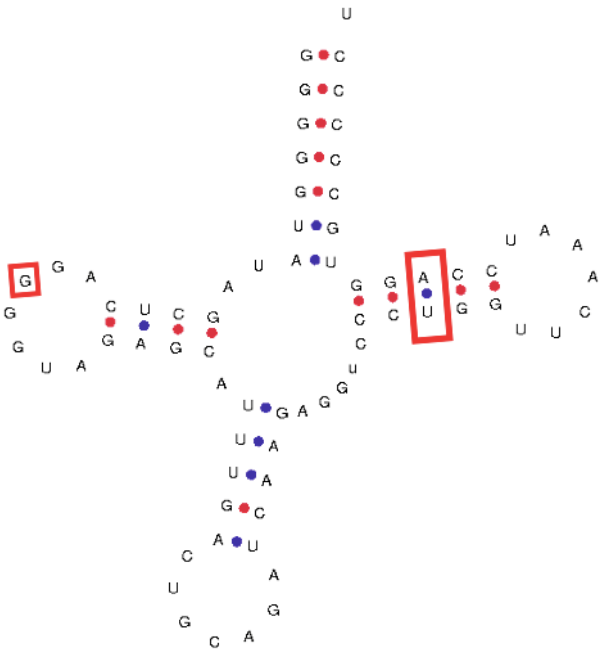


**Figure 19**. **The sequence and the secondary structure of tRNA-Cys-GCA-9 (tC09).** The differences from tC10 are indicated by red rectangles. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

**tRNA-Cys-GCA-8-1 (tC01)**

In p2luci, overexpressed tC01 showed slightly increased UGA readthrough level with $P < 0.1$ (Fig. 18, Supplementary Table 6). tC01 was detected with quite low numbers in all the three expression analyses (Table 8, Supplementary Table 5). The sequence and the secondary structure of this isodecoder (Fig. 20) seem canonical, showing no unusual residues or introns. The mature tRNA score is lower than for any of the tRNAs in the previous group, 77.3. Compared to tC10, in the sequence of tC01 there is guanine instead of uracil in the same position of D-loop and cytosine instead of uracil in the same position of T-arm (with stronger G:C base pairing in tC01 and G:U base pairing in tC10; compare Fig. 16 and Fig. 20).

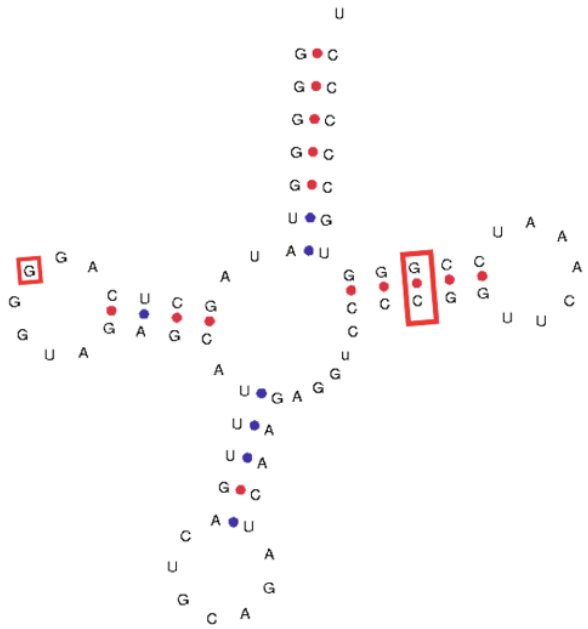


**Figure 20. The sequence and the secondary structure of tRNA-Cys-GCA-8-1 (tC01).** The differences from tC10 are indicated by red rectangles. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

**tRNA-Cys-GCA-5-1 (tC02)**

In p2luci, overexpressed tC02 showed slightly increased UGA readthrough level (Fig. 18, Supplementary Table 6). tC02 was the second highest ranking isodecoder in YAMAT-Seq (Supplementary Table 3), however, this isodecoder was detected with quite low expression levels in both tRNA-Seq from Torres, 2019 (Supplementary Table 1) and ChIP-Seq (Supplementary Table 2). Its mature tRNA score is the highest in this group, but lower than

for tRNAs in the previous group, 78.7. The sequence and the secondary structure (Fig. 21) of this isodecoder seem quite canonical. The major structural difference resides in the D-loop, which is 8 nucleotides long in case of tC02 (with an additional guanine leading to extension of the D-loop), whilst the D-loop of tC10 is 7 nucleotides long; other pairing within the molecule is not affected by this extension. Compared to tC10, in the sequence of tC02 there is cytosine instead of uracil in the same position of T-arm (with stronger G:C base pairing in tC02 and G:U base pairing in tC10; compare Fig. 16 and Fig. 21).

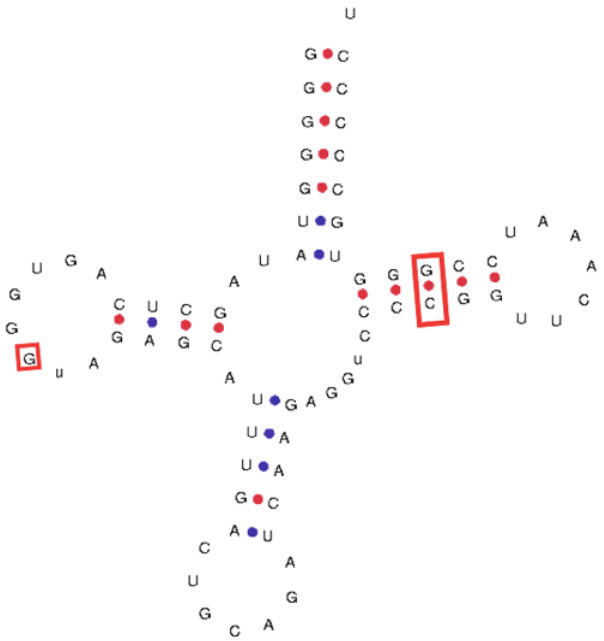


**Figure 21. The sequence and the secondary structure of tRNA-Cys-GCA-5-1 (tC02).** The differences from tC10 are indicated by red rectangles. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

Overexpression of abovementioned isodecoders led to a mild increase in UGA readthrough levels. The readthrough increase was significant with $P < 0.1$ only in case of overexpressed tC09 and tC01. In the expression analyses, these isodecoders in general had lower expression levels than the isodecoders categorized here as showing high capability of readthrough increase. For all of them, the mature tRNA score predicted by tRNAscan-SE is also lower than for tC10 and tC06. Compared to tC10, which showed the highest readthrough levels in p2luci, these isodecoders have nucleotide differences in their sequence that might be contributing to the lower capability to incorporate at UGA stop codon during translational readthrough.

### 5.3.4.3 The mature tRNA sequences of tC07, tC03, tC04 and tC08 placed under the U6 promoter in p2luci did not change UGA readthrough in HEK293T

The mature tRNA sequences of tRNA-Cys-GCA-14-1 (tC07), tRNA-Cys-ACA-1-1 (tC03), tRNA-Cys-GCA-22-1 (tC04), tRNA-Cys-GCA-17-1 (tC08), placed under the U6 promoter, did not change the levels of UGA readthrough (compared to the level of a control plasmid) in HEK293T cells in p2luci (Fig. 22). The UGA readthrough levels with their exact numerical values can be found in Supplementary Table 6.

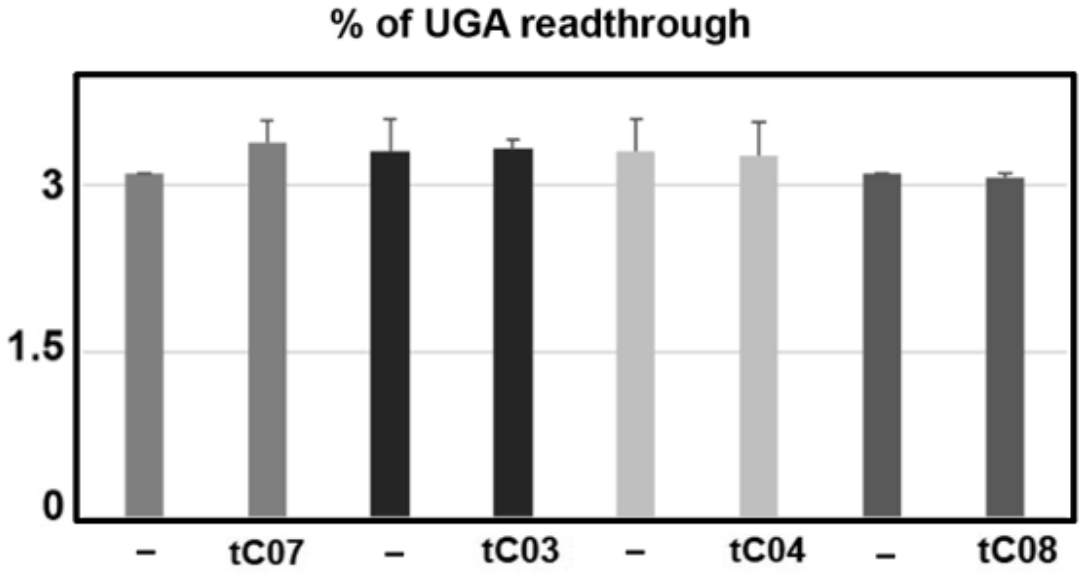



**Figure 22. The mature tRNA sequences of tC07, tC03, tC04 and tC08 placed under the U6 promoter did not change UGA readthrough in p2luci in HEK293T.** The "–" indicates a control reporter with no tRNA. The readthrough levels for all tested tRNA molecules with their exact numerical values can be found in Supplementary Table 6.

**tRNA-Cys-GCA-14-1 (tC07)**

In the tRNA-Seq from Torres, 2019 and YAMAT-Seq, tC07 was detected with low numbers, however, in ChIP-Seq analysis, significant expression levels (about +++) were observed (Supplementary Table 5). Moreover, ChIP-Seq levels were much higher than for e.g. tC01 or tC02 (Supplementary Table 2). tC07 has quite low mature tRNA score, 71.6 (compared to the isodecoders in the previous group, see in Supplementary Table 6). The sequence and the secondary structure of this isodecoder (Fig. 23) seem canonical, showing no unusual residues or introns. Compared to tC10, in the sequence of tC07 there

is guanine instead of uracil in the same position of D-loop, adenine instead of guanine in the same position of variable loop and cytosine instead of uracil in the same position of T-arm (with stronger G:C base pairing in tC07 and G:U base pairing in tC10; compare Fig. 16 and Fig. 23).

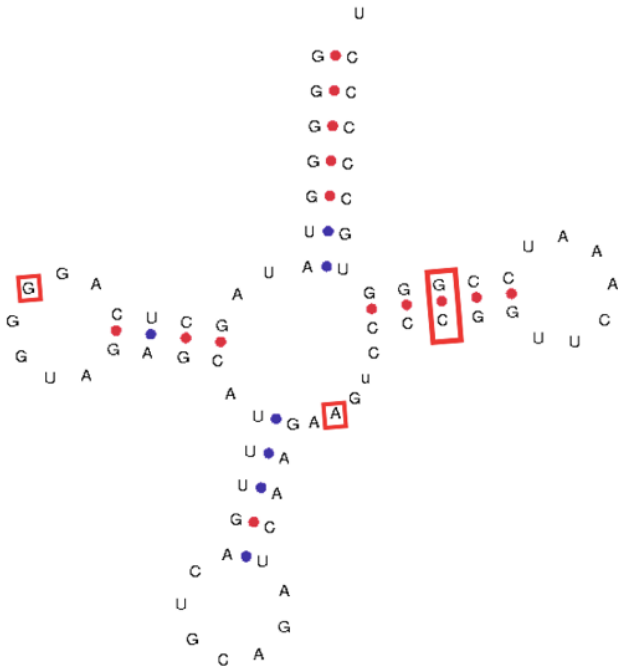


**Figure 23. The sequence and the secondary structure of tRNA-Cys-GCA-14-1 (tC07).** The differences from tC10 are indicated by red rectangles. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

**tRNA-Cys-ACA-1-1 (tC03)**

tRNA-Cys-ACA-1-1 is the only isodecoder molecule of the tRNA-Cys-ACA isoacceptor. In the expression analyses, tC03 was detected only in Torres 2019 tRNA-Seq, with very low expression levels (Supplementary Table 5, Supplementary Table 1). Its very low mature tRNA score is 46.2, which may imply that this molecule will not fold into a functional structure. Both the sequence and the secondary structure of tC03 are very different when compared to the canonical tRNA structure of tC10; compare Fig. 16 and Fig. 24. tC03 contains intron between sequence positions 37 and 133 and is categorized as a primary pseudogene in GtRNAdb these days. According to GtRNAdb, namely U4:U69, U6:U67, U29:U41, U51:U63 atypical features are present in tC03 (Fig. 24).

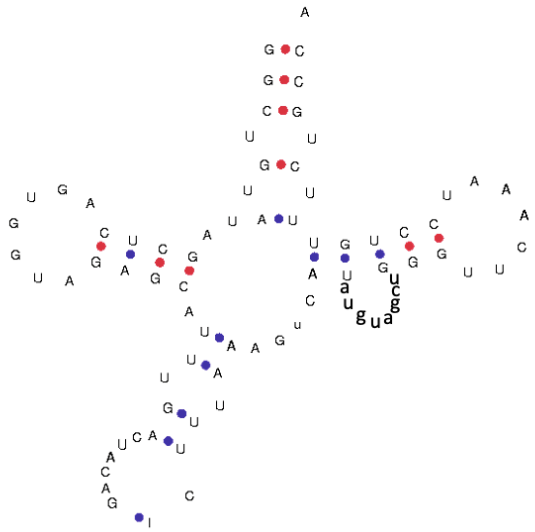


**Figure 24. The sequence and the secondary structure of tRNA-Cys-ACA-1-1 (tC03).** This tRNA contains intron and is categorized as a primary pseudogene in GtRNAdb. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtRNAdb.ucsc.edu.

### tRNA-Cys-GCA-22-1 (tC04)

In the expression analysis, tRNA-Cys-GCA-22-1 was detected only in tRNA-Seq data from Torres, 2019, with low expression levels (Supplementary Table 5, Supplementary Table 1). The mature tRNA score of this isodecoder is low, 64.9. Compared to tC10, the sequence and the secondary structure of tC04 seem a bit different (compare Fig. 16 and Fig. 25), the acceptor stem of tC04 is shorter and there is a bulge in it (Fig. 26). The very 3' nucleotide of the acceptor stem of tC04 is adenine, compared to uracil in the same position of tC10. In tC04, there is guanine instead of uracil in the same position of D-loop, cytosine instead of uracil in the same position of T-arm (with stronger G:C base pairing in tC04 and G:U base pairing in tC10), U:A base pairing in T-arm (and C:G base pairing in the same positions of tC10; compare Fig. 16 and Fig. 25).

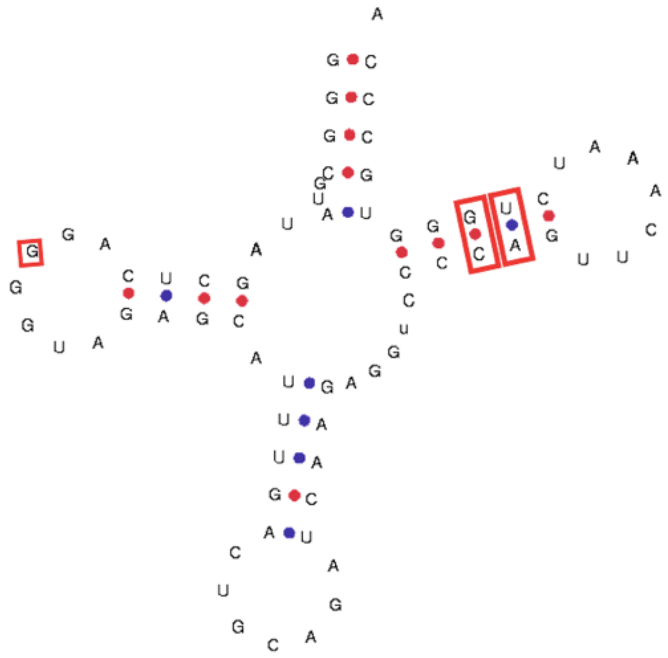


**Figure 25. The sequence and the secondary structure of tRNA-Cys-GCA-22-1 (tC04).**
The differences from tC10 are indicated by red rectangles. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

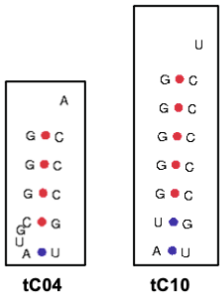



**Figure 26. Comparison of the acceptor stem of tC04 and tC10.**

### tRNA-Cys-GCA-17-1 (tC08)

tC08 was the second highest ranking isodecoder in tRNA-Seq data from Torres, 2019 (Supplementary Table 1). Similarly to tRNA-Cys-GCA-14-1, this isodecoder also ranked significantly high in ChIP-Seq analyses (Supplementary Table 2). On the other hand, this molecule was not detected by YAMAT-Seq (Supplementary Table 3, Supplementary Table 4). The mature tRNA score of this isodecoder is quite low, 70.7. Compared to tC10, in the sequence of tC08 there is guanine instead of uracil in the same position of D-loop,

cytosine instead of uracil in the same position of T-arm (with stronger G:C base pairing in tC08 and G:U base pairing in tC10), adenine instead of guanine in the acceptor stem (with unusual A:C base pairing in tC08 and G:C base pairing in tC10). The very 3' nucleotide of the acceptor stem of tC08 is cytosine (instead of uracil in tC10; compare Fig. 16 and Fig. 27). tC08 was included to this study because of its significant expression levels predicted by ChIP-Seq and tRNA-Seq. Since it has quite low mature tRNA score and is structurally different, we believe that although some tRNA molecules can be expressed significantly in different cell lines, it is possible that they are not translationally active.

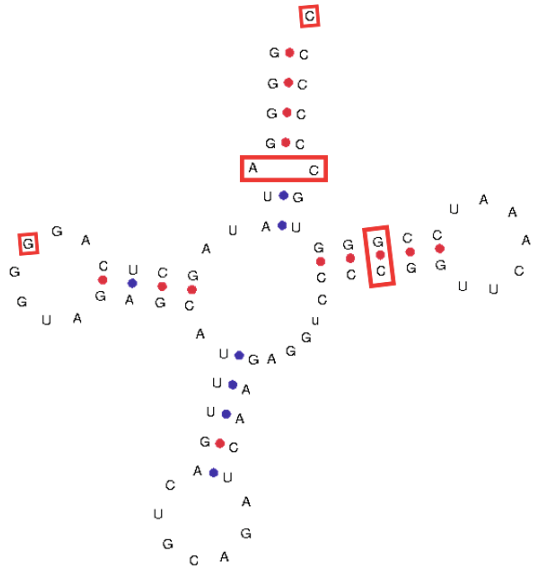


**Figure 27. The sequence and the secondary structure of tRNA-Cys-GCA-17-1 (tC08).** The differences from tC10 are indicated by red rectangles. Base pairing with three hydrogen bonds is indicated by red dots and with two hydrogen bonds by blue dots. The secondary structure prediction was downloaded from http://gtrnadb.ucsc.edu.

Three of the four isodecoders in this group (tC03, tC04, tC08) are structurally different. The mature tRNA scores predicted by tRNAscan-SE are lower than for the isodecoders described here to increase UGA readthrough when overexpressed. Low scores may indicate that these molecules do not fold into functional tRNAs. We did not observe any changes in UGA readthrough levels when the abovementioned tRNA sequences were placed under the U6 promoter in p2luci. However, since we are not able to quantify the rate of overexpression, we cannot say if the tRNAs were expressed from our reporter/folded properly/or aminoacylated properly. Thus, we cannot conclude if these isodecoder molecules do or do not have the ability to increase UGA readthrough.

Regarding all tRNA-Cys isodecoders tested, we saw that the mature tRNA scores as predicted by tRNAscan-SE correlated with the readthrough-induction capability of each isodecoder. The isodecoders with the highest UGA readthrough levels had the highest scores, the isodecoders with only a mild readthrough increase had lower scores. Although we cannot directly interpret the results concerning the isodecoder sequences with no increase of UGA readthrough, these isodecoders had the lowest scores. The same correlation was observed for tRNA-Tyr and tRNA-Trp in Beznoskova et al. 2021.

### 5.3.5 Readthrough measurement of overexpressed tRNA-Cys isodecoders in pSGDluc

The UGA readthrough levels measured in pSGDluc slightly vary from those measured in p2luci. In pSGDluc, the only significant increase of UGA readthrough with $P < 0.01$ is observed when tC10 is overexpressed, consistent with the results measured for p2luci. Overexpressed tC06 showed the second highest UGA readthrough with $P < 0.05$ in p2luci, however, in pSGDluc, no significant increase was measured (Fig. 28). tC01, tC02, tC03 and tC04 were not tested in pSGDluc.

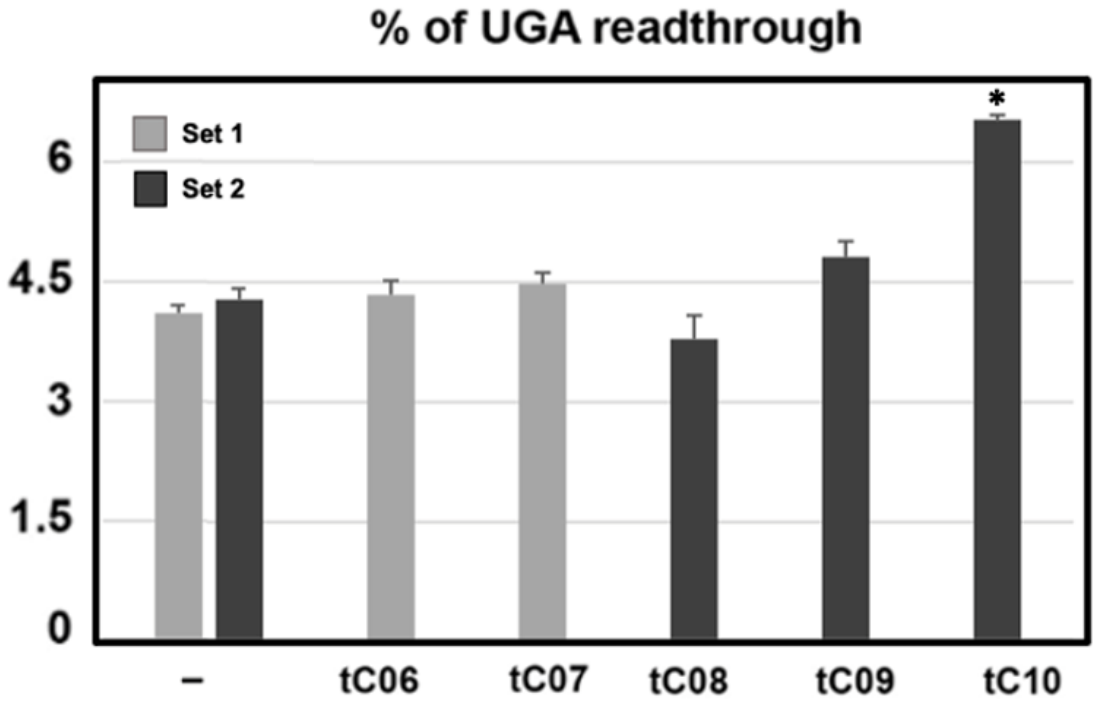



**Figure 28. The UGA readthrough levels measured using pSGDluc.** These isodecoders were measured separately as set 1 (gray) set 2 (black). The changes in the readthrough levels relative to the reporter with no tRNA (–) were analysed by Student's t-test (mean+SD; n=2); statistically significant increase with $P < 0.01$ is marked with asterisk "*". The UGA readthrough levels for all tested tRNA molecules with their exact numerical values for pSGDluc can be found in Supplementary Table 7.

When we align the sequences of all the tested tRNA-Cys isodecoders (Fig. 29), we can see that both tC10 and tC06 (which increase UGA readthrough in p2luci the most when overexpressed) share a "T" in sequence position 16, whereas all of the other selected isodecoders have a "G" in this position. tC10 (with the very highest level of UGA readthrough) differs from tC06 only by a single nucleotide, a "T" in the sequence

position 50 (Fig. 29). This “T” in the sequence position 50 is present also in the sequence of tC09 (Fig.29), which showed the third highest UGA readthrough levels in p2luci. This might potentially indicate that presence of “T” in both the sequence positions 16 and 50 may be the cause of tC10 being the isodecoder with the highest UGA readthrough-inducing ability.


**Figure 29. The sequence alignment of all tested tRNA-Cys isodecoders.** tC03 not included. The highest scoring tC10 and tC06 both have a “T” instead of otherwise canonical “G” in sequence position 16 (indicated by red rectangles; this is also a change in the A Box sequence when compared to the other isodecoders, see chapter 6.2). In tC10, there is another unique “T” in sequence position 50 (indicated by a red rectangle). The same position 50 “T” is present in the sequence of tC09, which showed the third highest UGA readthrough levels. In the sequence of tC09, there is an additional unique “A” in position 62 (gray rectangle).

## 5.4 Working with artificially prepared tRNA-Gln-GCA

### 5.4.1 Selection of human tRNA for anticodon alteration

Another approach, complementary to analysis of cysteine tRNA bearing the GCA anticodon, is to use another tRNA, mutate its anticodon to GCA and test it for UGA stop codon recognition. Noteworthy, for most of the tRNAs, the anticodon nucleotides are important for recognition by particular aminoacyl-tRNA synthetase and correct amino acid charging.

The construction of a functional suppressor tRNA obtained by a mutation in the anticodon was described before (Koukuntla et al. 2013). This amber UAG suppressor was created from a naturally occurring tRNA-Gln-CTG-1 (Roy, Cooke and Buckland 1982), by changing the CTG anticodon to CTA; particularly from tRNA-Gln since glutamine is one of the most commonly mutated amino acids in nonsense mutations in human genes (Koukuntla et al. 2013). The human U6 SnRNA promoter (which is much stronger that the

Pol III promoter within the tRNA gene) was subcloned upstream of the glutamine tRNA sequence, based on previous results showing that this led to an increase of expression in case of human tRNA-Tyr (Lai et al. 2004). Indeed, the Amber suppressor tRNA-Gln-CTA showed significantly higher suppression efficacy (Koukuntla et al. 2013). The same was observed when recreated in our reporter system (Beznoskova et al. 2021).

Since it was shown for tRNA-Gln-CTG-1 that substitution in the anticodon provided a functional tRNA mutant with high suppression activity at UAG stop codon (Koukuntla et al. 2013), we created a mutant version of the same tRNA, where the CTG anticodon was mutated to GCA (Fig. 30c) to test this version for its ability to recognize the UGA stop codon. The selected gene was tRNA-Gln-CTG-1; this sequence can be found in the human genome for five identical gene copies, tRNA-Gln-CTG-1-1, tRNA-Gln-CTG-1-2, tRNA-Gln-CTG-1-3, tRNA-Gln-CTG-1-4 and tRNA-Gln-CTG-1-5.

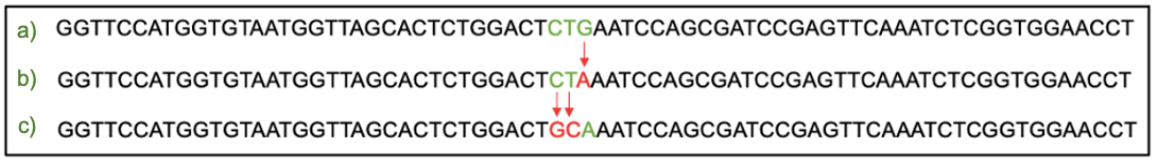



**Figure 30. The sequences of mature tRNAs**. **a)** Naturally occurring human tRNA-Gln-CTG-1 with anticodon CTG. **b)** Mutated version of Amber UAG suppressor tRNA-Gln-CTA prepared in Koukuntla, 2013. **c)** Mutated version of the same tRNA prepared in this study, tRNA-Gln-GCA.

### 5.4.2 Construction of artificially prepared tRNA-Gln-GCA reporter constructs using the pSGDluc vector

The insert bearing the mutated sequence of mature tRNA-Gln-GCA depicted in Fig. 30c under the U6 promoter was prepared using the GeneArt™ Strings™ DNA Fragments (see 4.8.1) exactly as described for preparation of aforementioned tRNA-Cys inserts. The GeneArt™ Strings™ DNA Fragments sequence tRNA-Gln-GCA for the mutant version of glutamine tRNA was first digested by SalI-BamHI and the 387-bp product was then purified using the QIAquick® PCR Purification Kit (Qiagen®) (see 4.13.7). The insert was ligated into B1145/B1146 vectors in 1:3 vector:insert molar ratio according to the protocol for ligation (see 4.13.4). The ligation reaction was then transformed into *E. coli* strain DH5α according to the protocol for heat shock transformation (see 4.14.1)

and the cells were spread onto an LB-Amp plate and grown for 24 hours. Selected clones were then inoculated into liquid LB-Amp medium and bacteria was grown for 16 hours prior to plasmid DNA isolation. The isolation was then performed using the QIAprep® Spin Miniprep Kit (Qiagen®) (see 4.13.1). After the DNA isolation, the subcloning was confirmed by specific restriction reaction (see 4.13.3). For inserts subcloned into B1145 and B1146, the products were control-digested by NcoI enzyme with expected fragment sizes 3399 bp and 2376 bp (Fig. 31). The plasmids confirmed by restriction digest were then verified by sequencing (see 4.13.8) and the full list of pSGDluc-tRNA-Gln-GCA plasmids is shown in Table 11.

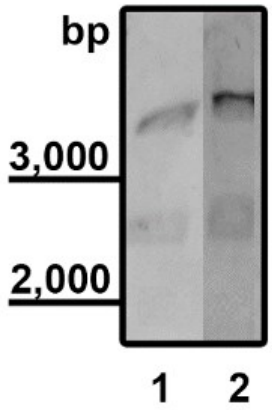



**Figure 31. Confirmation of subcloning of desired tRNA-Gln-GCA insert into pSGDluc-UGA (B1145) and pSGDluc-UGG (B1146) by restriction digest.** The reaction was performed using NcoI enzyme with expected fragment sizes 3399 bp and 2376 bp. Electrophoresis was performed with ~ 400 ng of DNA run on a 0.8 % agarose gel at 5 V/cm. Lane 1 – pMK38, lane 2 – pMK40.

**Table 11. The full list of plasmids created for the purpose of this study with tRNA-Gln-GCA insert subcloned into pSGDluc-UGA (B1145) and pSGDluc-UGG (B1146).** pMK38 and pMK40 were prepared by inserting the 387-bp SalI-BamHI fragment from GeneArt™ Strings™ DNA Fragments sequence (for tRNA-Gln-GCA) into SalI-BamHI digested B1145 and B1146, respectively.

| Plasmid name | Plasmid description |
|---|---|
| pMK38 | pSGDluc-UGA-U6-tQ-GCA |
| pMK40 | pSGDluc-UGG-U6-tQ-GCA |

### 5.4.3 Construction of artificially prepared tRNA-Gln-GCA reporter constructs using the p2luci vector

The insert bearing the mutated sequence of mature tRNA-Gln-GCA depicted in Fig. 30c under the U6 promoter was prepared by digesting the pMK38 plasmid created before for the pSGDluc vector (Table 11) by MscI-KpnI. The restriction reaction was subsequently confirmed by electrophoresis and the desired 351-bp fragment was excised from the gel and the DNA was isolated according to the protocol for isolation of DNA from the gel (see 4.13.6) using the QIAquick® Gel extraction Kit (Qiagen®). The inserts were then ligated into PBB171/PBB172 vectors in 1:3 vector:insert molar ratio according to the protocol for ligation (see 4.13.4). The ligation reaction was transformed into *E. coli* strain DH5α according to the protocol for heat shock transformation (see 4.14.1) and the cells were spread onto an LB-Amp plate and grown for 24 hours. Selected clones were then inoculated into liquid LB-Amp medium and bacteria was grown for 16 hours prior to plasmid DNA isolation. The isolation was then performed using the QIAprep® Spin Miniprep Kit (Qiagen®) (see 4.13.1). After the DNA isolation, the subcloning was confirmed by specific restriction reaction (see 4.13.3). For inserts subcloned into PBB171 and PBB172, the products were control-digested by NcoI enzyme with the expected fragment sizes 3697 bp and 1797 bp (Fig. 32). The plasmids confirmed by restriction digest were then verified by sequencing (see 4.13.8) and the full list of p2luci-tRNA-Gln-GCA plasmids is shown in a table below (Table 12).

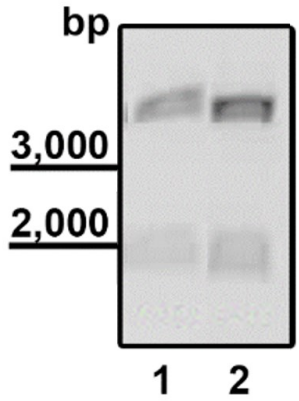



**Figure 32. Confirmation of subcloning of desired tRNA-Gln-GCA insert into p2luci-UGA (PBB171) and p2luci-CAG (PBB172) by restriction digest**. The reaction was performed using NcoI enzyme with expected fragment sizes 3697 bp and 1797 bp. Electrophoresis was performed with ~ 400 ng of DNA run on a 0.8 % agarose gel at 5 V/cm. Lane 1 – pMK44, lane 2 – pMK42.

**Table 12. The full list of plasmids created for the purpose of this study with tRNA-Gln-GCA insert subcloned into p2luci-UGA (PBB171) and p2luci-CAG (PBB172).** pMK44 and pMK24 were prepared by inserting the 351-bp MscI-KpnI fragment from pMK38 into MscI-KpnI digested PBB171 and PBB172, respectively.

| Plasmid name | Plasmid description |
|---|---|
| pMK44 | p2luci-UGA-U6-tQ-GCA |
| pMK42 | p2luci-CAG-U6-tQ-GCA |

### 5.4.4 Readthrough measurement of tRNA-Gln-GCA in pSGDluc vector

The readthrough measurement and analysis were performed according to the particular protocols (see 4.16.1). The estimated percentage of UGA readthrough for tRNA-Gln-GCA is depicted below. No increase of UGA readthrough is observed, compared to the reporter with no tRNA (Fig. 33).

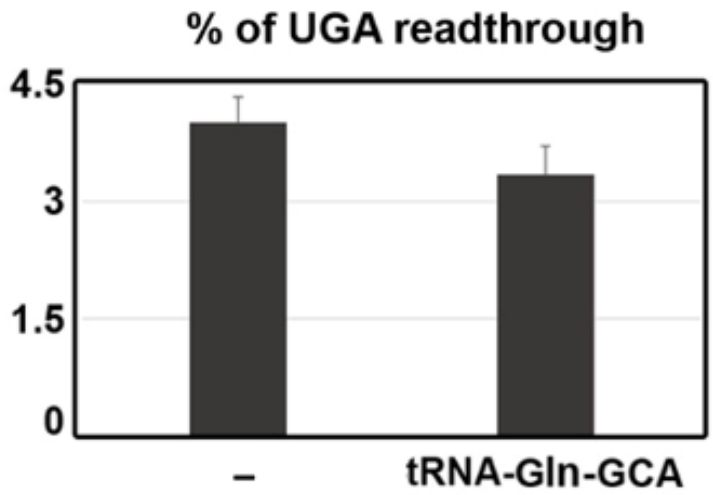



**Figure 33. The sequence of an artificially prepared tRNA-Gln-GCA under the U6 promoter in pSGDluc did not change the level of UGA readthrough in HEK293T.** The UGA readthrough values of the tested construct seem lower than of the reporter with no tRNA (–), however, the decrease is not statistically significant as calculated by the Student's t-test statistical formula used**.** The readthrough values of stop-codon-containing plasmids (B1145 and MK38) were normalized to their sense-codon-containing counterparts (B1146 and MK40). The exact numerical values of UGA readthrough of tRNA-Gln-GCA tested in pSGDluc are listed in Supplementary Table 8.

### 5.4.5 Readthrough measurement of tRNA-Gln-GCA in p2luci vector

The readthrough measurement and analysis was performed according to the particular protocol (see 4.16.1). The percentage of UGA readthrough efficiency is depicted below (Fig. 34), showing no change in the UGA readthrough levels for the traditional p2luci reporter-system (PBB171 and PBB172) just the same as for the new StopGo sequence-containing pSGDluc (B1145 and B1146).

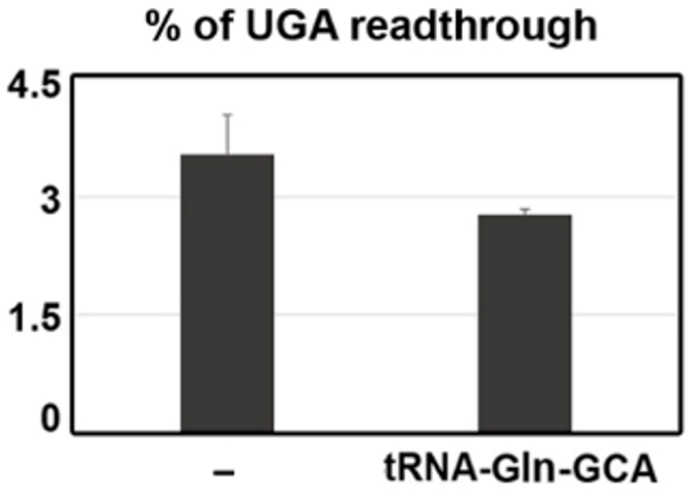



**Figure 34. The sequence of an artificially prepared tRNA-Gln-GCA under the U6 promoter in p2luci did not change the level of UGA readthrough in HEK293T.** Just like for the pSGDluc, the UGA readthrough values of the tested construct seem lower than of the reporter with no tRNA (–), however, the decrease is not statistically significant as calculated by the Student's t-test statistical formula used**.** The stop-codon-containing plasmids (PBB171 and MK44) were normalized to their particular sense-codon-containing counterparts (PBB172 and MK42). The exact numerical values of UGA readthrough of tRNA-Gln-GCA tested in p2luci are listed in Supplementary Table 8.

## 5.5 CARS protein from HEK293T cells visualization

Another approach that could be used to confirm that some isodecoders can modify UGA readthrough levels would be to downregulate the levels of cysteine aminoacyl-tRNA synthetase (CARS), thus lowering the levels of cysteine tRNAs in HEK293T cells and measure UGA readthrough levels in these cells after the CARS downregulation. The readthrough levels should be correspondingly decreased when CARS is downregulated. In this introductory chapter we demonstrate that we are capable of detecting CARS protein by Western blot. Current experiments focus on downregulation and its optimalization and will try to confirm again that cysteine tRNA can modify UGA readthrough efficiency in HEK293T cells.

### 5.5.1 Optimalization of Western blot for CARS visualization

To detect CARS protein by Western blot, a commercially distributed anti-CARS antibodies were purchased. The first anti-CARS antibody tested, monoclonal ab126714 (Abcam), turned out to be inefficient even in high concentrations (Fig. 35C and Fig. 36G). Another anti-CARS antibody was ordered, polyclonal ab235536 (Abcam), which showed almost the same results (Fig. 35A, 35B and Fig. 36E, 36F). A new antibody was ordered from different vendor, polyclonal HPA002383 (Sigma-Aldrich®), and for troubleshooting purposes, an experiment was designed that would test different whole cell extract preparations, different antibodies, antibody dilutions and blocking agents. Two identical membranes were always used, one for being blocked in 5 % BSA (Fig. 35), one for being blocked in 5 % low-fat milk (Fig. 36). Three different types of whole cell extract preparation were performed – a) direct lysis on plate using 1x Glo Lysis Buffer (Promega) (see 4.15.1.1), b) Direct lysis on plate with 1x Loading Buffer (see 4.15.1.2), c) Lysate preparation using precleaning centrifugation step (see 4.15.1.3). Anti-RPS14, ab174661 (Abcam), 1:1000 dilution in 5 % low-fat milk was used as a loading control (Fig. 35 and Fig. 36).

Each membrane was then used for probing with different antibodies:

A: anti-CARS ab235536 (Abcam), 1:2000 dilution in 5 % BSA

B: anti-CARS ab235536 (Abcam), 1:500 dilution in 5 % BSA

C: anti-CARS, ab126714 (Abcam), 1:1000 dilution in 5 % BSA

D: anti-CARS HPA002383 (Sigma-Aldrich®), 1:1000 dilution in 5 % BSA

E: anti-CARS ab235536 (Abcam), 1:2000 dilution in 5 % low-fat milk

F: anti-CARS ab235536 (Abcam), 1:500 dilution in 5 % low-fat milk

G: anti-CARS, ab126714 (Abcam), 1:1000 dilution in 5 % low-fat milk

H: anti-CARS HPA002383 (Sigma-Aldrich®), 1:1000 dilution in 5 % low-fat milk

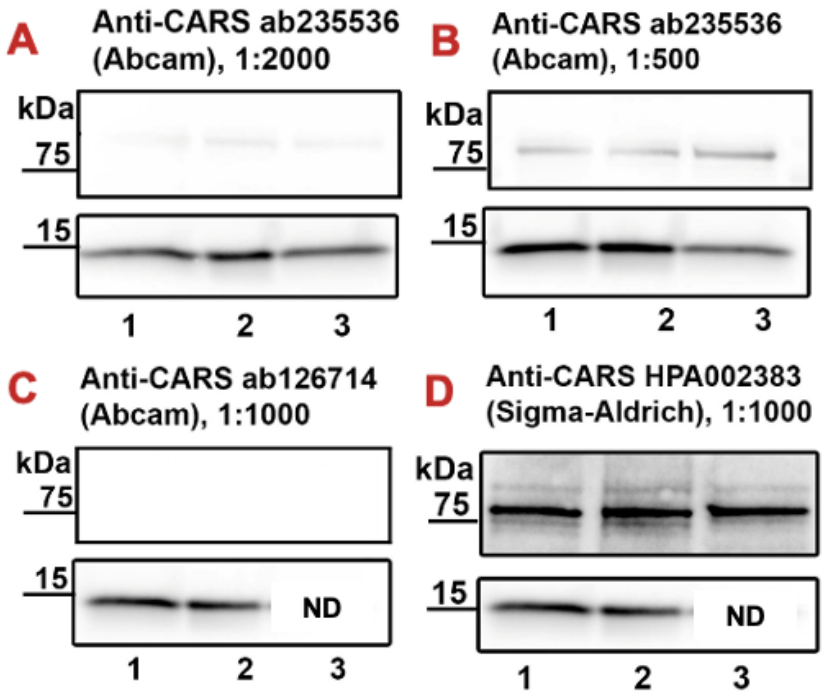



**Figure 35. Membranes blocked in 5 % BSA.** Anti-RPS14, ab174661 (Abcam), 1:1000 dilution in 5 % low-fat milk was used as a loading control. ND – not detected. Membrane C shows very faint band at expected position when blocked in 5 % BSA, however, recommended dilution for this antibody is 1:2000-1:5000.

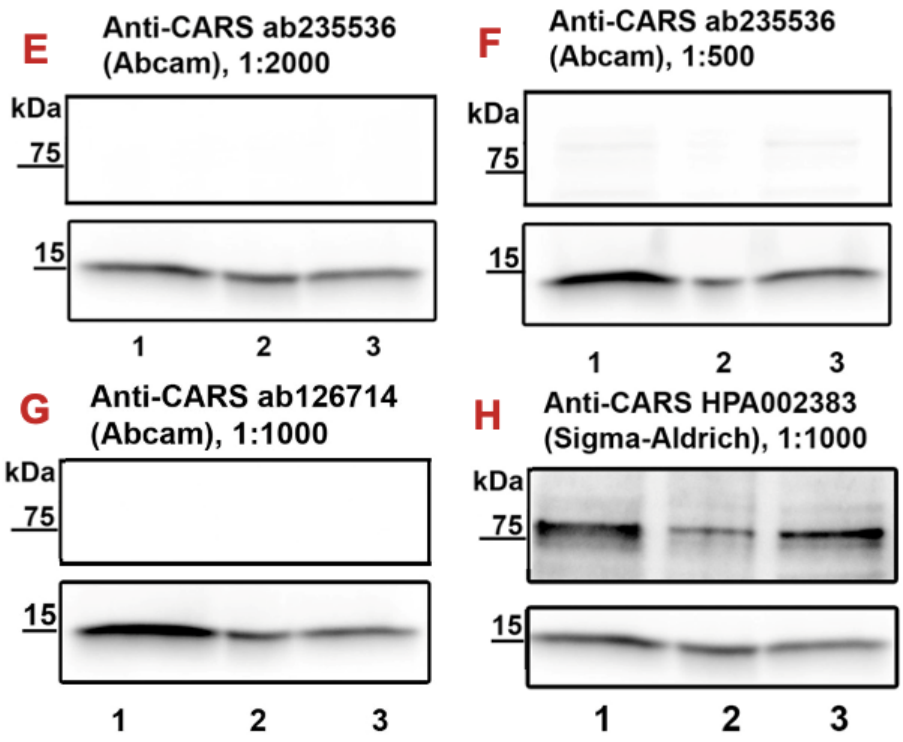



**Figure 36. Membranes blocked in 5 % low-fat milk**. Anti-RPS14, ab174661 (Abcam), 1:1000 dilution in 5 % low-fat milk was used as a loading control. Only HPA002383 shown band at expected position.

To summarize, only HPA002383 (Sigma-Aldrich®) showed a clear and visible band at expected position in recommended 1:1000 dilution in 5 % BSA as well as 5 % low-fat milk (Fig. 35D and 36H). By using this antibody, we were able to detect CARS protein in HEK293T using Western blot. This antibody is intended to be used in further experiments.

# 6 DISCUSSION

## 6.1 Selection of tRNA-Cys isodecoders from bioinformatically predicted human tRNA-Cys genes

The initial aim of this study was to identify a functional tRNA-Cys molecule. Since tRNAs in GtRNAdb are only computationally predicted and we do not know which genes encode *in vivo* functional tRNAs, we analysed three expression datasets acquired by different methods from different cell lines: tRNA-Seq data from Torres, 2019, ChIP-Seq data from Rudolph, 2016 and YAMAT-Seq data from Shigematsu, 2017 and selected five highly expressed tRNA-Cys genes.

We are aware of the fact that the expression of tRNA isodecoders varies a lot and not all tRNA isodecoder genes are expressed in all cells and tissues the same rate (Torres et al. 2019, Kutter et al. 2011). Indeed, we saw that the expression of tRNA-Cys isodecoder genes varies a lot as predicted by the expression analyses. For instance, tRNA-Cys-GCA-2-3 was the highest ranking isodecoder in all the three cell lines in ChIP-Seq (+++++), however, in HEK293T tRNA-Seq data from Torres, 2019, it had only low expression (+). Conversely, tRNA-Cys-GCA-2-2 was the highest ranking isodecoder in Torres, 2019 data (+++++), however, showed only a mild expression in ChIP-Seq data (++). Interestingly, we also saw significant variation within a single dataset in case of tRNA-Cys-GCA-2-1. In Torres, 2019 data, the expression in measurement CTRL3.1 was "+++++", in CTRL4.1 it was only "+". Moreover, this isodecoder was not detected by ChIP-Seq at all. All of this could suggest that these analyses provided only an insight into the actual levels of expression.

Undoubtedly, there are both strenghts and weaknesses to each of the attitudes, tRNA-Seq and ChIP-Seq. Discerning and quantifying the mature tRNAs using tRNA-Seq may be challenging since the secondary structure of tRNA and tRNA posttranscriptional modifications may impair sequencing and it can be diffucult to distinguish mature tRNA reads from pre-RNAs or tRNA fragments. Also, similarity in tRNA sequences may impair aligning to the specific tRNA gene (Torres et al. 2019). To challenge some of these

problems, for example, in highthrouput YAMAT-seq method, modified Y-shaped adaptor molecules are ligated by T4 RNA ligase 2 specifically to mature tRNAs and resulting cDNAs are subjected to Illumina HiSeq 2500 sequencing system (Shigematsu et al. 2017).

In ChIP-Seq, only the actively expressed tRNAs are detected. During ChIP-Seq, a sequence bound by the polymerase is subsequently sequenced (Schmidt et al. 2009), and short sequences up and downstream of the tRNA are detected, which may help to align the tRNA gene precisely to the right position in genome. Still, it can be found in literature that a poor correlation of tRNA levels quantified by tRNA-Seq and ChIP-Seq was observed, and stalling of Pol III was suggested to explain for this fact. On the other hand, further data were arguing against Pol III arrest – not having enough data supporting either of these hypotheses may imply that rather than for precise quantification of particular tRNA levels, expression levels predicted by ChIP-Seq might better be considered a "marker of ongoing transcription" only (Orioli et al. 2016).

Taken together, we selected the five highly expressed tRNA-Cys isodecoder genes by combining the information gathered from both the ChIP-Seq and tRNA-Seq attitudes, which was the most useful for our purposes – identifying the functional tRNA-Cys molecules. Four more tRNA-Cys isodecoders were used, which were used in our laboratory in past, but those were not scoring especially high in the expression analysis performed here. As was described before, the expression of isodecoders is tissue and cell type specific and also different isodecoders can have different suppression ability (summarized in chapter 2.1.1.3). Thus, we cannot rule out the possibility that the lowly expressed tRNA-Cys isodecoders (as analysed here) will be functional in a cell specific way.

## 6.2 A and B Box sequence comparison of human tRNA-Cys isodecoder genes

It is still quite not resolved, why there exist multiple isodecoders to a single tRNA isoacceptor species. They may be possibly expressed specifically in different tissues or during different cell states. Indeed it was demonstrated that tRNA isodecoders show different expression rates in different tissues (Kutter et al. 2011, Torres et al. 2019) and suggested that the sequence variability in the A and B Boxes of the tRNA molecules could account for different tRNA isodecoder expression rates in tissues (Goodenbour and Pan

2006). In this thesis, the logos prepared from the sequences of the A and B Box of the five most expressed and the five least expressed tRNA-Cys isodecoders were compared to search for possible correlation of the A and B Box sequences and isodecoders' expression levels. On the one hand, we saw that all the highly expressed genes precisely followed the canonical consensus sequence for both the A and the B Box. On the other hand, the genes with the lowest expression levels had variability in their sequence (either the sequence was different than the prescribed one, or the sequence followed the consensus, but differed from the sequence of the highly expressed genes). Altogether, this analysis is in line with previous research and supported the idea that sequence differences in the A Box and the B Box among different isodecoders may correlate with different tRNA isodecoders' expression rates in cells and tissues.

## 6.3 Identifying tRNA-Cys isodecoders which increase UGA readthrough when overexpressed

During the course of this thesis as many as 9 tRNA-Cys isodecoders with both ACA (1) and GCA (8) anticodons were tested for their ability to act as readthrough inducing near-cognate tRNA molecules. These were tested using two reporter plasmid systems, p2luci, where a fused *Renilla*-Firefly product is translated, and pSGDluc, where two separate proteins are made.

The UGA readthrough levels measured in p2luci slightly vary from the levels measured in pSGDluc, however, we are aware of the fact that we use two different contexts, the TMV context in p2luci and the -CUAG context in pSGDluc and we thus cannot compare the readthrough levels between the two vectors. Interestingly, a variation in the readthrough levels was described before for the same context, in an experiment comparing pSGDluc to the traditional dual-luciferase vector system pDluc (Fixsen and Howard 2010), a plasmid system similar to p2luci (Grentzmann et al. 1998). These results showed that the readthrough levels measured when transfecting HEK293T with pDluc may differ from those observed when transfecting with pSGDluc. For instance, about 17 % of readthrough was observed for human OPRL1 mRNA in pSGDluc, whereas about 30 % of readthrough was observed for the same mRNA in pDluc. The Firefly protein levels were checked by immunoblotting and densitometry (to tell if the reporter activities and abundances are concordant), which showed that the OPRL1 readthrough level measured in pSGDluc is

more accurate (Loughran et al. 2017). In my case, it could be interesting to test the isodecoder sequences in the TMV context in pSGDluc, so that we would be able to compare the levels measured in both contexts.

In p2luci, two molecules, tRNA-Cys-GCA-4-1 (tC10) and tRNA-Cys-GCA-2 (tC06) showed almost ~ 100 % and ~ 50 % increase in readthrough levels, respectively, when overexpressed in HEK293T. Another three molecules showed mild increase of UGA readthrough, namely tRNA-Cys-GCA-9 (tC09), tRNA-Cys-GCA-8-1 (tC01) and tRNA-Cys-GCA-5-1 (tC02). The remaining mature tRNA sequences placed under the U6 promoter did not change UGA readthrough levels as compared to the reporter with no tRNA – those were tRNA-Cys-GCA-22-1 (tC04), tRNA-Cys-GCA-14-1 (tC07), tRNA-Cys-GCA-17-1 (tC08) and tRNA-Cys-ACA-1-1 (tC03). In pSGDluc, only tRNA-Cys-GCA-4-1 (tC10) showed significantly increased level of UGA readthrough (~ 50 % increase) with $P < 0.01$ as compared to the UGA readthrough level of the reporter with no tRNA. Surprisingly, we did not observe any changes in UGA readthrough when the mature tRNA sequence of tC06 was placed under the U6 promoter in pSGDluc.

Since tC10 showed significantly increased UGA readthrough in both p2luci and pSGDluc, the sequences of all the tested tRNA-Cys isodecoders were then compared to tC10 to search for the unique nucleotide features of tC10 that could potentially account for the elevated levels of UGA readthrough. The nucleotide differences are presented in detail for better understanding directly together with the readthrough results in chapter 5.3.4. Compared to all the other tested isodecoders, in the mature tRNA sequence of tC10 there are unique thymines in the sequence positions 16 and 50. This “T” in sequence position 16 is also present in tC06, which, when overexpressed, showed the second highest UGA readthrough levels in p2luci. The third highest UGA readthrough in p2luci was observed for overexpressed tC09, which has a “T” in position 50. Presence of “T” in both sequence positions 16 and 50 of tC10 was thus suggested to have a potential role in UGA readthrough increase, by a mechanism which remains to be unravelled.

Although tC06 has a “T” in position 16 and showed significantly increased levels in p2luci, it did not do so in pSGDluc. We believe that this could be due to the different contexts used and that it would be interesting to test this sequence with TMV context in pSGDluc.

Some of the sequences, when placed under the U6 promoter, did not change the UGA readthrough levels in both p2luci and pSGDluc. Honestly, we can't simply say they do not have the ability to increase UGA readthrough when overexpressed in our reporter system. In this system we are not able to precisely quantify the exact levels of expressed tRNAs yet and have thus no exact information regarding the rate of overexpression. When we do not see any increase of UGA readthrough, we can't directly say whether the particular tRNA is expressed from our reporter, folded properly or if it is aminoacylated. Thus, we can only interpret the results where we get a statistically significant increase of UGA readthrough – in which case we can say that the particular tRNA gets overexpressed and this overexpression leads to an increase in the levels of UGA readthrough using our reporter constructs in HEK293T cells. Concerning my data, the results indicate that overexpressed human tRNA-Cys-GCA-4-1 (tC10) elevates UGA readthrough in human cells lines just like was suggested for human tRNA-Trp (for UGA) and tRNA-Tyr (for UAG and UAA) in our laboratory recently (Beznoskova et al. 2021).

The four tRNA-Cys-GCA isodecoders mentioned to be used in our laboratory before (tC01, tC02, tC03, tC04) all showed low levels or mild increase of UGA readthrough when overexpressed in p2luci in HEK293T cell. Beznoskova et al. 2021 say that the overexpression of these in the same p2luci system resulted in a translational shut down and suggests possibly explaining this by generation of 5' tRNA-derived fragments (as described in chapter 2.1.1.1). In accord with this, in my experiments, we similarly observed decreased levels of *Renilla* in all p2luci constructs tested here (compared to the levels of a control construct) indicating the possible translational shut down. To confirm if anything like translational shut down indeed happens, more experiments than just comparing the *Renilla* levels as measured here would be required. For now, we can only say that the lower-than-control *Renilla* levels were observed for all tRNA-Cys isodecoders tested in p2luci. Despite this fact, in multiple independent experiments we repeatedly observed UGA readthrough increase for the tRNA-Cys isodecoders mentioned above.

Just like in Beznoskova et al. 2021 for tRNA-Tyr and tRNA-Trp, correlation between the mature tRNA scores predicted by tRNAscan-SE and the tRNA-Cys isodecoders' ability to elevate UGA readthrough in p2luci was observed. We suggested that this indicates that tRNAscan-SE is a reliable tool for predicting the tRNA functionality and activity *in vivo.*

Concerning the stop codon context, as mentioned before, it was shown in our laboratory in yeast, that UGA-G and UGA-C is preferentially read through by tRNA-Cys and UGA-A by tRNA-Trp (Beznoskova et al. 2016). It could be interesting to investigate if this finding will also be applicable to humans. So far, we have no data regarding this, and more research is necessary.

## 6.4 Working with artificially prepared tRNA-Gln-GCA

One of the aims of this study was to test if the GCA anticodon is capable of UGA stop codon recognition. The abovementioned results identified the tRNA-Cys isodecoders capable of increasing UGA readthrough when overexpressed in our reporter system and thus confirmed that the GCA anticodon is indeed capable of UGA stop codon recognition.

In parallel with those experiments, we decided to use a complementary approach. We designed a mutated version of glutamine tRNA, tRNA-Gln-GCA, by changing the anticodon of a naturally occurring tRNA-Gln-CTG from CTG to GCA. We selected this molecule since its CTG anticodon was mutated to CTA in human cell line and showed significant suppression activity at the UAG stop codon (Koukuntla et al. 2013). The mutated tRNA-Gln-GCA prepared during the course of this thesis was tested using the p2luci and pSGDluc constructs bearing the UGA stop codon (described before in 5.3.1.).

This artificially prepared sequence, when placed under the U6 promoter, did not change the readthrough level using both p2luci and pSGDluc. The fact that we do not see an effect on UGA readthrough does not necessarily mean that this tRNA can't recognize the GCA anticodon. It is possible that we simply are not able to create a translationally active tRNA with cysteine GCA anticodon from the tRNA-Gln sequence *in vivo* by this approach. Theoretically, the mutation in the anticodon of the tRNA-Gln that we made may result in the tRNA not being expressed from our reporter at all. Again, we only can interpret the effect of a particular overexpressed tRNA when we see that overexpression of that tRNA leads to a statistically significant increase in UGA readthrough levels.

For instance, tRNA-Gln was also shown to be incorporating efficiently at UAG in yeast in our laboratory recently (Beznoskova et al. 2019), however, human tRNA-Gln-CTG did not

increase UAG readthrough in human cells in our reporter system (Beznoskova et al. 2021). In their latest paper, Beznoskova et al. 2021 emphasise that the suppressor tRNA-Gln-CTA shows a very strong suppression activity yet differs from the natural tRNA-Gln-CTG only by a single nucleotide. They hypothesise that possibly the wild-type tRNA-Gln-CTG is properly processed as well as the suppressor tRNA and suggest that possibly this tRNA does not increase the UAG readthrough levels in human contrary to yeast.

## 6.5 Future plan: CARS downregulation

Lastly, we thought of to using an opposite, yet complementary approach to confirm our findings – to downregulate the cysteinyl-tRNA synthase (CARS) levels in HEK293T cells using siRNA and perform the UGA readthrough measurement with these cells just like was done in Beznoskova et al. 2021 for tryptophanyl-tRNA synthetase (WARS) for tRNA-Trp and tyrosyl-tRNA synthetase (YARS) for tRNA-Tyr. We would expect to see the readthrough levels decrease. We show here an initial experiment where we visualize the natural level of CARS using Western blot.

The further experiments will thus focus on siRNA-downregulation, optimalization of the downregulation and its visualisation by qPCR and Western blot and will hopefully confirm our hypothesis – that cysteine tRNA regulates proteosynthesis in human cell lines *via* stop codon readthrough.

# 7 CONCLUSIONS

- In the initial part of this thesis, the five most highly expressed tRNA-Cys-GCA isodecoders were selected, based on analysis of three publicly available quantitation datasets.

- Second, the human tRNA-Cys's ability to function as a near-cognate/readthrough-inducing tRNA in human cell line HEK293T was assessed. Taken together, at least one isodecoder of all tested, tRNA-Cys-GCA-4-1 (named here as tC10), shown statistically significant increase in UGA readthrough levels ($P < 0.01$) in both p2luci and pSGDluc vectors.

- Notably, we observed a correlation in the mature tRNA score as predicted by tRNAscan-SE and the overexpressed molecule's ability to increase UGA readthrough.

- Although it is apparent by the previously mentioned results that the GCA anticodon of tRNA-Cys is capable of decoding the UGA stop codon, we did not see UGA readthrough increase in either of the vector systems when we used artificially prepared sequence of tRNA-Gln-GCA. We can't conclude anything regarding this tRNA's ability of UGA readthrough increase since we are not able to tell if we created a translationally active tRNA, if it was expressed from our reporter, folded or aminoacylated properly.

# 8 LIST OF REFERENCES

* Review article


*Amrani, N., R. Ganesan, S. Kervestin, D. A. Mangus, S. Ghosh & A. Jacobson (2004) A faux 3 '-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature,* 432**,** 112-118.

*Avcilar-Kucukgoze, I. & A. Kashina (2020) Hijacking tRNAs From Translation: Regulatory Functions of tRNAs in Mammalian Cell Physiology. *Frontiers in Molecular Biosciences,* 7.

Beier, H., M. Barciszewska, G. Krupp, R. Mitnacht & H. J. Gross (1984) UAG readthrough during TMV RNA translation: isolation and sequence of two tRNAs with suppressor activity from tobacco plants. *Embo Journal,* 3**,** 351-356.

Beissel, C., B. Neumann, S. Uhse, I. Hampe, P. Karki & H. Krebber (2019) Translation termination depends on the sequential ribosomal entry of eRF1 and eRF3. *Nucleic Acids Research,* 47**,** 4798-4813.

Beznoskova, P., L. Bidou, O. Namy & L. S. Valasek (2021): Increased expression of tryptophan and tyrosine tRNAs elevates stop codon readthrough in human cell lines. *Nucleic Acid Research* (accepted)

Beznoskova, P., S. Gunisova & L. S. Valasek (2016) Rules of UGA-N decoding by near-cognate tRNAs and analysis of readthrough on short uORFs in yeast. *Rna,* 22**,** 456-466.

Beznoskova, P., Z. Pavlikova, J. Zeman, C. E. Aitken & L. S. Valasek (2019) Yeast applied readthrough inducing system (YARIS): an invivo assay for the comprehensive study of translational readthrough. *Nucleic Acids Research,* 47**,** 6339-6350.

Beznoskova, P., S. Wagner, M. E. Jansen, T. von der Haar & L. S. Valasek (2015) Translation initiation factor elF3 promotes programmed stop codon readthrough. *Nucleic Acids Research,* 43**,** 5099-5111.

Blanchet, S., D. Cornu, M. Argentini & O. Namy (2014) New insights into the incorporation of natural suppressor tRNAs at stop codons in Saccharomyces cerevisiae. *Nucleic Acids Research,* 42**,** 10061-10072.

Boccaletto, P., M. A. Machnicka, E. Purta, P. Piatkowski, B. Baginski, T. K. Wirecki, V. de Crecy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm & J. M. Bujnicki (2018) MODOMICS: a database of RNA modification pathways. 2017 update. *Nucleic Acids Research,* 46**,** D303-D307.

Brown, C. M., P. A. Stockwell, C. N. A. Trotman & W. P. Tate (1990) Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Research,* 18**,** 6339-6345.

Canella, D., V. Praz, J. H. Reina, P. Cousin & N. Hernandez (2010) Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. *Genome Research,* 20**,** 710-721.

Cassan M, JP. Rousset (2001) UAG readthrough in mammalian cells (2001) Effect of upstream and downstream stop codon contexts reveal different signals. *BMC Mol Biol*; 2:3.

Chan, P. P. & T. M. Lowe (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Research,* 44**,** D184-D189.

Chan, P. P., B. Y. Lin, A. J. Mak, T. M. Lowe (2019) tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. Preprint posted on www.biorxiv.org. doi: https://doi.org/10.1101/614032

Clark, W. C., M. E. Evans, D. Dominissini, G. Q. Zheng & T. Pan (2016) tRNA base methylation identification and quantification via high-throughput sequencing. *Rna,* 22**,** 1771-1784.

Cozen, A. E., E. Quartley, A. D. Holmes, E. Hrabeta-Robinson, E. M. Phizicky & T. M. Lowe (2015) ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nature Methods,* 12**,** 879-884.

Crick, F. H. C. (1966) Codon--anticodon pairing: the wobble hypothesis. *Journal of Molecular Biology,* 19**,** 548-555.

Crooks, G. E., G. Hon, J. M. Chandonia & S. E. Brenner (2004) WebLogo: A sequence logo generator. *Genome Research,* 14**,** 1188-1190.

*Dabrowski, M., Z. Bukowy-Bieryllo & E. Zietkiewicz (2015) Translational readthrough potential of natural termination codons in eucaryotes - The impact of RNA sequence. *Rna Biology,* 12**,** 950-958.

*--- (2018) Advances in therapeutic use of a drug-stimulated translational readthrough of premature termination codons. *Molecular Medicine,* 24.

*Dever, T. E., J. D. Dinman & R. Green (2018) Translation Elongation and Recoding in Eukaryotes. *Cold Spring Harbor Perspectives in Biology,* 10.

*Dever, T. E. & R. Green (2012) The Elongation, Termination, and Recycling Phases of Translation in Eukaryotes. *Cold Spring Harbor Perspectives in Biology,* 4**,** 16.

Dunn, J. G., C. K. Foo, N. G. Belletier, E. R. Gavis & J. S. Weissman (2013) Ribosome profiling reveals pervasive and regulated stop codon readthrough in Drosophila melanogaster. *Elife,* 2.

Eswarappa, S. M., A. A. Potdar, W. J. Koch, Y. Fan, K. Vasu, D. Lindner, B. Wiiiard, L. M. Graham, P. E. DiCorieto & P. L. Fox (2014) Programmed Translational Readthrough Generates Antiangiogenic VEGF-Ax. *Cell,* 157**,** 1605-1618.

Feng, Y. X., T. D. Copeland, S. Oroszlan, A. Rein & J. G. Levin (1990) Identification of amino acids inserted during suppression of UAA and UGA termination codons at the gag-pol junction of Moloney murine leukemia virus. *Proceedings of the National Academy of Sciences of the United States of America,* 87**,** 8860-8863.

Firth, A. E., N. M. Wills, R. F. Gesteland & J. F. Atkins (2011) Stimulation of stop codon readthrough: frequent presence of an extended 3 ' RNA structural element. *Nucleic Acids Research,* 39**,** 6679-6691.

Fixsen, S. M. & M. T. Howard (2010) Processive Selenocysteine Incorporation during Synthesis of Eukaryotic Selenoproteins. *Journal of Molecular Biology,* 399**,** 385-396.

*Frischmeyer, P. A. & H. C. Dietz (1999) Nonsense-mediated mRNA decay in health and disease. *Human Molecular Genetics,* 8**,** 1893-1900.

*Geiduschek, E. P. & G. P. Tocchinivalentini (1988) Transcription by RNA polymerase III. *Annual Review of Biochemistry,* 57**,** 873-914.

Geller, A. I. & A. Rich (1980) A UGA termination suppression tRNATrp active in rabbit reticulocytes. *Nature,* 283**,** 41-46.

Geslain, R. & T. Pan (2010) Functional Analysis of Human tRNA Isodecoders. *Journal of Molecular Biology,* 396**,** 821-831.

Goelet, P., G. P. Lomonossoff, P. J. G. Butler, M. E. Akam, M. J. Gait & J. Karn (1982) Nucleotide sequence of tobacco mosaic virus RNA. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences,* 79**,** 5818-5822.

*Gomez, M. A. R. & M. Ibba (2020) Aminoacyl-tRNA synthetases. *Rna,* 26**,** 910-936.
Goodenbour, J. M. & T. Pan (2006) Diversity of tRNA genes in eukaryotes. *Nucleic Acids Research,* 34**,** 6137-6146.
Grentzmann, G., J. A. Ingram, P. J. Kelly, R. F. Gesteland & J. F. Atkins (1998) A dual-luciferase reporter system for studying recoding signals. *Rna,* 4**,** 479-486.
Hamada, M., Y. Huang, T. M. Lowe & R. J. Maraia (2001) Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. *Molecular and Cellular Biology,* 21**,** 6870-6881.
*Hellen, C. U. T. (2018) Translation Termination and Ribosome Recycling in Eukaryotes. *Cold Spring Harbor Perspectives in Biology,* 10.
Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick & A. Zamir (1965) Structure of a ribonucleic acid. *Science,* 147**,** 1462-1465.
*Hopper, A. K. & E. M. Phizicky (2003) tRNA transfers to the limelight. *Genes & Development,* 17**,** 162-180.
Hou, Y. M. & P. Schimmel (1988) A simple structural feature is a major determinant of the identity of a transfer RNA. *Nature,* 333**,** 140-145.
Hou, Y. M., E. Westhof & R. Giege (1993) An unusual RNA tertiary interaction has a role for the specific aminoacylation of a transfer RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 90**,** 6776-6780.
Ishimura, R., G. Nagy, I. Dotu, H. H. Zhou, X. L. Yang, P. Schimmel, S. Senju, Y. Nishimura, J. H. Chuang & S. L. Ackerman (2014) Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. *Science,* 345**,** 455-459.
Ivanov, P., M. M. Emara, J. Villen, S. P. Gygi & P. Anderson (2011) Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. *Molecular Cell,* 43**,** 613-623.
Ivanov, P., E. O'Day, M. M. Emara, G. Wagner, J. Lieberman & P. Anderson (2014) G-quadruplex structures contribute to the neuroprotective effects of angiogenin-induced tRNA fragments. *Proceedings of the National Academy of Sciences of the United States of America,* 111**,** 18201-18206.
*Kapp, L. D. & J. R. Lorsch (2004) The molecular mechanics of eukaryotic translation. *Annual Review of Biochemistry,* 73**,** 657-704.
*Keeling, K. M., D. Wang, S. E. Conard & D. M. Bedwell (2012) Suppression of premature termination codons as a therapeutic approach. *Critical Reviews in Biochemistry and Molecular Biology,* 47**,** 444-463.
Kim, S. H., G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl & A. Rich (1973) Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science,* 179**,** 285-288.
Koukuntla, R., W. J. Ramsey, W. B. Young & C. J. Link (2013) U6 promoter-enhanced GlnUAG suppressor tRNA has higher suppression efficacy and can be stably expressed in 293 cells. *Journal of Gene Medicine,* 15**,** 93-101.
Kumar, P., J. Anaya, S. B. Mudunuri & A. Dutta (2014) Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. *Bmc Biology,* 12.
*Kumar, P., C. Kuscu & A. Dutta (2016) Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). *Trends in Biochemical Sciences,* 41**,** 679-689.
Kuscu, C., P. Kumar, M. Kiran, Z. L. Su, A. Malik & A. Dutta (2018) tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *Rna,* 24**,** 1093-1105.

Kutter, C., G. D. Brown, A. Goncalves, M. D. Wilson, S. Watt, A. Brazma, R. J. White & D. T. Odom (2011) Pol III binding in six mammals shows conservation among amino acid isotypes despite divergence among tRNA genes. *Nature Genetics,* 43**,** 948-955.

Lai, D. Z., S. J. Weng, C. Wang, L. Q. Qi, C. M. Yu, L. Fu & W. Chen (2004) Small antisense RNA to cyclin D1 generated by pre-tRNA splicing inhibits growth of human hepatoma cells. *Febs Letters,* 576**,** 481-486.

*Lee, H. L. R. & J. P. Dougherty (2012) Pharmaceutical therapies to recode nonsense mutations in inherited diseases. *Pharmacology & Therapeutics,* 136**,** 227-266.

Levitt, M. (1969) Detailed molecular model for transfer ribonucleic acid. *Nature,* 224**,** 759-763.

Li, G. P. & C. M. Rice (1993) The signal for translational readthrough of a UGA codon in Sindbis virus RNA involves a single cytidine residue immediately downstream of the termination codon. *Journal of Virology,* 67**,** 5062-5067.

Lipman, R. S. A. & Y. M. Hou (1998) Aminoacylation of tRNA in the evolution of an aminoacyl-RNA synthetase. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 13495-13500.

Loughran, G., M. Y. Chou, I. P. Ivanov, I. Jungreis, M. Kellis, A. M. Kiran, P. V. Baranov & J. F. Atkins (2014) Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Research,* 42**,** 8928-8938.

Loughran, G., M. T. Howard, A. E. Firth & J. F. Atkins (2017) Avoidance of reporter assay distortions from fused dual reporters. *Rna,* 23**,** 1285-1289.

Loughran, G., I. Jungreis, I. Tzani, M. Power, R. I. Dmitriev, I. P. Ivanov, M. Kellis & J. F. Atkins (2018) Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response. *Journal of Biological Chemistry,* 293**,** 4434-4444.

Lowe, T. M. & P. P. Chan (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research,* 44**,** W54-W57.

Machnicka, M. A., K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki & H. Grosjean (2013) MODOMICS: a database of RNA modification pathways-2013 update. *Nucleic Acids Research,* 41**,** D262-D267.

Martins-Dias P, L. Romão (2021) Nonsense suppression therapies in human genetic diseases. *Cell Mol Life Sci*. Epub ahead of print. doi: 10.1007/s00018-021-03809-7

McCaughan, K. K., C. M. Brown, M. E. Dalphin, M. J. Berry & W. P. Tate (1995) Translational termination efficiency in mammals is influenced by the base following the stop codon. *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 5431-5435.

Mort, M., D. Ivanov, D. N. Cooper & N. A. Chuzhanova (2008) A meta-analysis of nonsense mutations causing human genetic disease. *Human Mutation,* 29**,** 1037-1047.

Mottagui-Tabar, S., M. F. Tuite & L. A. Isaksson (1998) The influence of 5 ' codon context on translation termination in Saccharomyces cerevisiae. *European Journal of Biochemistry,* 257**,** 249-254.

Namy, O., G. Duchateau-Nguyen, I. Hatin, S. Hermann-Le Denmat, M. Termier & J. P. Rousset (2003) Identification of stop codon readthrough genes in Saccharomyces cerevisiae. *Nucleic Acids Research,* 31**,** 2289-2296.

Namy, O., I. Hatin & J. P. Rousset (2001) Impact of the six nucleotides downstream of the stop codon on translation termination. *Embo Reports,* 2**,** 787-793.

Napthine, S., C. Yek, M. L. Powell, T. D. K. Brown & I. Brierley (2012) Characterization of the stop codon readthrough signal of Colorado tick fever virus segment 9 RNA. *Rna,* 18**,** 241-252.

Newberry, K. J., Y. M. Hou & J. J. Perona (2002) Structural origins of amino acid selection without editing by cysteinyl-tRNA synthetase. *Embo Journal,* 21**,** 2778-2787.

Orioli, A., V. Praz, P. Lhote & N. Hernandez (2016) Human MAF1 targets and represses active RNA polymerase III genes by preventing recruitment rather than inducing long-term transcriptional arrest. *Genome Research,* 26**,** 624-635.

Pallanck, L., S. H. Li & L. H. Schulman (1992) The anticodon and discriminator base are major determinants of cysteine tRNA identity in vivo. *Journal of Biological Chemistry,* 267**,** 7221-7223.

*Pan, T. (2018) Modifications and functional genomics of human transfer RNA. *Cell Research,* 28**,** 395-404.

*Paule, M. R. & R. J. White (2000) Transcription by RNA polymerases I and III. *Nucleic Acids Research,* 28**,** 1283-1298.

Pelham, H. R. B. (1978) Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature,* 272**,** 469-471.

*Phizicky, E. M. & A. K. Hopper (2010) tRNA biology charges to the front. *Genes & Development,* 24**,** 1832-1860.

*Rich, A. & U. L. Rajbhandary (1976) Transfer RNA: molecular structure, sequence, and properties. *Annual Review of Biochemistry,* 45**,** 805-860.

Roy, B., J. D. Leszyk, D. A. Mangus & A. Jacobson (2015) Nonsense suppression by near-cognate tRNAs employs alternative base pairing at codon positions 1 and 3. *Proceedings of the National Academy of Sciences of the United States of America,* 112**,** 3038-3043.

Roy, K. L., H. Cooke & R. Buckland (1982) Nucleotide sequence of a segment of human DNA containing the three tRNA genes. *Nucleic Acids Research,* 10**,** 7313-7322.

Rudolph, K. L. M., B. M. Schmitt, D. Villar, R. J. White, J. C. Marioni, C. Kutter & D. T. Odom (2016) Codon-Driven Translational Efficiency Is Stable across Diverse Mammalian Cell States. *Plos Genetics,* 12**,** 23.

Schmidt, D., M. D. Wilson, C. Spyrou, G. D. Brown, J. Hadfield & D. T. Odom (2009) ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods,* 48**,** 240-248.

Schueren, F., T. Lingner, R. George, J. Hofhuis, C. Dickel, J. Gartner & S. Thoms (2014) Peroxisomal lactate dehydrogenase is generated by translational readthrough in mammals. *Elife,* 3.

Shigematsu, M., S. Honda, P. Loher, A. G. Telonis, I. Rigoutsos & Y. Kirino (2017) YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Research,* 45**,** 11.

Singh, A., L. E. Manjunath, P. Kundu, S. Sahoo, A. Das, H. R. Suma, P. L. Fox & S. M. Eswarappa (2019) Let-7a-regulated translational readthrough of mammalian AGO1 generates a microRNA pathway inhibitor. *Embo Journal,* 38.

Skuzeski, J. M., L. M. Nichols, R. F. Gesteland & J. F. Atkins (1991) The signal for a leaky UAG stop codon in several plant viruses includes the two downstream codons. *Journal of Molecular Biology,* 218**,** 365-373.

Song, H. W., P. Mugnier, A. K. Das, H. M. Webb, D. R. Evans, M. F. Tuite, B. A. Hemmings & D. Barford (2000) The crystal structure of human eukaryotic release factor eRF1 - Mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell,* 100**,** 311-321.

Stiebler, A. C., J. Freitag, K. O. Schink, T. Stehlik, B. A. M. Tillmann, J. Ast & M. Bolker (2014) Ribosomal Readthrough at a Short UGA Stop Codon Context Triggers Dual Localization of Metabolic Enzymes in Fungi and Animals. *Plos Genetics,* 10.

Sun, J. C., M. Chen, J. L. Xu & J. H. Luo (2005) Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. *Journal of Molecular Evolution,* 61**,** 437-444.

Tork, S., I. Hatin, J. P. Rousset & C. Fabret (2004) The major 5 ' determinant in stop codon read-through involves two adjacent adenines. *Nucleic Acids Research,* 32**,** 415-421.

Torres, A. G., O. Reina, C. S. O. Attolini & L. R. de Pouplana (2019) Differential expression of human tRNA genes drives the abundance of tRNA-derived fragments. *Proceedings of the National Academy of Sciences of the United States of America,* 116**,** 8451-8456.

Xin, H., C. L. Zhong, E. Nudleman & N. Ferrara (2016) Evidence for Pro-angiogenic Functions of VEGF-Ax. *Cell,* 167**,** 275-284.

Yamasaki, S., P. Ivanov, G. F. Hu & P. Anderson (2009) Angiogenin cleaves tRNA and promotes stress-induced translational repression. *Journal of Cell Biology,* 185**,** 35-42.

Yoshinaka, Y., I. Katoh, T. D. Copeland & S. Oroszlan (1985) Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *Proceedings of the National Academy of Sciences of the United States of America,* 82**,** 1618-1622.

Zhang, C. M., T. Christian, K. J. Newberry, J. J. Perona & Y. M. Hou (2003) Zinc-mediated amino acid discrimination in cysteinyl-tRNA synthetase. *Journal of Molecular Biology,* 327**,** 911-917.

**Online sources:**

[1] Abcam. Cell culture guidelines. Available online:

https://www.abcam.com/ps/pdf/protocols/cell_culture.pdf

[2] Horizon™. HEK293T cells. Technical manual. Available online:

https://horizondiscovery.com/-/media/Files/Horizon/resources/Technical-manuals/tla-hek293t-cell-line-manual.pdf

[3] CORNING®. Surface Areas and Guide for Recommended Medium Volumes for Corning® Cell Culture Vessels. Available online:

https://www.corning.com/catalog/cls/documents/application-notes/CLS-AN-209.pdf

[4] Thermo Fisher Scientific. Useful information for various sizes of cell culture dishes and flasks. Available online: https://www.thermofisher.com/cz/en/home/references/gibco-cell-culture-basics/cell-culture-protocols/cell-culture-useful-numbers.html

# 9 Supplementary material

**Supplementary Table 1. All tRNA-Cys-GCA isodecoders detected by tRNA-Seq in HEK293T.** The expression values are expressed as RPM (reads per million). Both columns show results for HEK293T cell line, named CTRL (3.1) and CTRL (4.1). The levels of expression are demonstrated using a plus sign "+", + meaning 0-20 %; ++ meaning 20-40 %; +++ meaning 40-60 %; ++++ meaning 60-80 %; +++++ meaning 80-100 % of expression (100 % being the highest scoring value). Altogether 23 tRNA-Cys-GCA isodecoders were described so far for the GCA isoacceptor and single molecule for the ACA isoacceptor. The table shows 33 tRNA-Cys molecules, including tRNA-Cys-ACA-1-1, which is categorized as primary pseudogene by tRNAscan SE 2.0, tRNA-Cys-GCA-25-1, which is categorized as primary pseudogene and tRNA-Cys-GCA-24, which is categorized as a molecule with uncertain function.

| | tRNA-Seq results in HEK293T cell line | | Expression analysis | |
|---|---|---|---|---|
| Gene name | CTRL(3.1) | CTRL(4.1) | CTRL(3.1) | CTRL(4.1) |
| tRNA-Cys-GCA-1-1 | 128.91 | 626.21 | ++ | +++ |
| tRNA-Cys-GCA-2-1 | 432.58 | 343.95 | +++++ | + |
| tRNA-Cys-GCA-2-2 | 512.79 | 1343.08 | +++++ | +++++ |
| tRNA-Cys-GCA-2-3 | 101.70 | 134.59 | + | + |
| tRNA-Cys-GCA-2-4 | 210.56 | 525.27 | +++ | ++ |
| tRNA-Cys-GCA-3-1 | 12.89 | 72.90 | + | + |
| tRNA-Cys-GCA-4-1 | 75.92 | 200.01 | + | + |
| tRNA-Cys-GCA-5-1 | 174.75 | 391.61 | ++ | + |
| tRNA-Cys-GCA-6-1 | 121.75 | 312.17 | ++ | + |
| tRNA-Cys-GCA-7-1 | 48.70 | 65.42 | + | + |
| tRNA-Cys-GCA-8-1 | 35.81 | 262.63 | + | + |
| tRNA-Cys-GCA-9-1 | 35.81 | 105.61 | + | + |
| tRNA-Cys-GCA-9-2 | 98.83 | 568.26 | + | +++ |
| tRNA-Cys-GCA-9-3 | 111.72 | 660.79 | ++ | +++ |
| tRNA-Cys-GCA-9-4 | 95.97 | 565.46 | + | +++ |
| tRNA-Cys-GCA-10-1 | 17.19 | 66.36 | + | + |
| tRNA-Cys-GCA-11-1 | 1.43 | 17.76 | + | + |
| tRNA-Cys-GCA-12-1 | 157.56 | 1032.78 | ++ | ++++ |
| tRNA-Cys-GCA-13-1 | 14.32 | 25.24 | + | + |
| tRNA-Cys-GCA-14-1 | 53.00 | 290.67 | + | + |
| tRNA-Cys-GCA-15-1 | 35.81 | 46.73 | + | + |
| tRNA-Cys-GCA-16-1 | 35.81 | 31.78 | + | + |
| tRNA-Cys-GCA-17-1 | 423.98 | 1283.26 | +++++ | +++++ |
| tRNA-Cys-GCA-18-1 | 61.59 | 106.55 | + | + |
| tRNA-Cys-GCA-19-1 | 10.03 | 79.44 | + | + |
| tRNA-Cys-GCA-20-1 | 17.19 | 43.93 | + | + |
| tRNA-Cys-GCA-21-1 | 144.67 | 521.53 | ++ | ++ |
| tRNA-Cys-GCA-22-1 | 38.67 | 40.19 | + | + |
| tRNA-Cys-GCA-23-1 | 75.92 | 88.79 | + | + |
| tRNA-Cys-GCA-24-1 | 7.16 | 26.17 | + | + |
| tRNA-Cys-GCA-25-1 | 14.32 | 143.93 | + | + |
| tRNA-Cys-ACA-1-1 | 8.59 | 10.28 | + | + |

**Supplementary Table 2. The tRNA-Cys-GCA isodecoders detected by ChIP-Seq in HepG2, Huh7 and healthy liver adult cell lines.** The expression values are expressed as RPM (reads per million). The levels of expression are demonstrated using a plus sign "+", + meaning 0-20 %; ++ meaning 20-40 %; +++ meaning 40-60 %; ++++ meaning 60-80 %; +++++ meaning 80-100 % of expression (100 % being the highest scoring value).

| | ChiP-Seq levels in cell lines | | | Expression analysis | | |
|---|---|---|---|---|---|---|
| **Gene name** | **HepG2** | **Huh7** | **Liver-Adult** | **HepG2** | **Huh7** | **Liver-Adult** |
| **tRNA-Cys-GCA-1-1** | 147.40 | 249.46 | 66.77 | + | + | + |
| **tRNA-Cys-GCA-2-2** | 2239.61 | 2906.76 | 2505.33 | ++ | +++ | ++ |
| **tRNA-Cys-GCA-2-3** | 5732.34 | 4918.50 | 10120.96 | +++++ | +++++ | +++++ |
| **tRNA-Cys-GCA-4-1** | 4202.66 | 3990.54 | 6029.02 | ++++ | +++++ | ++++ |
| **tRNA-Cys-GCA-5-1** | 374.60 | 571.56 | 101.81 | + | + | + |
| **tRNA-Cys-GCA-8-1** | 236.95 | 105.00 | 90.39 | + | + | + |
| **tRNA-Cys-GCA-9-2** | 674.61 | 319.77 | 68.82 | + | + | + |
| **tRNA-Cys-GCA-9-4** | 1107.73 | 893.89 | 29.34 | + | + | + |
| **tRNA-Cys-GCA-11-1** | 429.12 | 40.22 | 36.42 | + | + | + |
| **tRNA-Cys-GCA-14-1** | 3436.97 | 3650.05 | 4853.66 | +++ | ++++ | +++ |
| **tRNA-Cys-GCA-15-1** | 596.14 | 300.58 | 3.71 | + | + | + |
| **tRNA-Cys-GCA-17-1** | 2559.11 | 4052.71 | 2353.60 | +++ | +++++ | ++ |
| **tRNA-Cys-GCA-20-1** | 326.50 | 142.15 | 52.27 | + | + | + |
| **tRNA-Cys-GCA-24-1** | 802.39 | 319.45 | 46.55 | + | + | + |

**Supplementary Table 3. The tRNA-Cys-GCA isodecoders detected by YAMAT-Seq in MCF7, SKBR3 and BT20 cell lines** with their numerical values. The expression values are expressed as RPM (reads per million).

| Gene name | MCF7 cell line | | | SKBR3 cell line | | | BT20 cell line | | |
|---|---|---|---|---|---|---|---|---|---|
| | MCF7_A | MCF7_B | MCF7_C | SKBR3_A | SKBR3_B | SKBR3_C | BT20_A | BT20_B | BT20_C |
| tRNA-Cys-GCA-1-1 | 13.29 | 6.78 | 8.24 | 0.36 | 0.24 | 0.00 | 3.66 | 4.23 | 4.24 |
| tRNA-Cys-GCA-2-1 | 133.88 | 74.82 | 91.33 | 334.06 | 362.83 | 179.70 | 110.28 | 115.32 | 117.25 |
| tRNA-Cys-GCA-3-1 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| tRNA-Cys-GCA-4-1 | 51.65 | 93.17 | 37.42 | 5.86 | 6.27 | 3.39 | 29.42 | 35.11 | 30.80 |
| tRNA-Cys-GCA-5-1 | 107.85 | 3.51 | 78.48 | 2.10 | 2.60 | 1.44 | 28.04 | 28.99 | 28.53 |
| tRNA-Cys-GCA-6-1 | 10.71 | 0.08 | 7.61 | 0.14 | 0.41 | 0.25 | 0.97 | 1.53 | 0.94 |
| tRNA-Cys-GCA-7-1 | 3.59 | 0.32 | 2.77 | 0.00 | 0.00 | 0.00 | 0.48 | 0.45 | 1.18 |
| tRNA-Cys-GCA-8-1 | 1.83 | 1.28 | 1.27 | 0.58 | 0.33 | 0.17 | 0.41 | 0.72 | 0.24 |
| tRNA-Cys-GCA-9-1 | 16.07 | 0.08 | 11.57 | 0.80 | 0.98 | 0.17 | 7.04 | 8.19 | 8.33 |
| tRNA-Cys-GCA-10-1 | 0.00 | 0.00 | 0.32 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| tRNA-Cys-GCA-11-1 | 0.00 | 36.69 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| tRNA-Cys-GCA-12-1 | 1.83 | 0.00 | 1.82 | 0.14 | 0.08 | 0.00 | 0.41 | 0.18 | 0.00 |
| tRNA-Cys-GCA-13-1 | 6.51 | 3.67 | 4.36 | 0.51 | 0.16 | 0.08 | 0.90 | 0.54 | 1.02 |
| tRNA-Cys-GCA-14-1 | 14.85 | 11.25 | 10.54 | 0.14 | 0.24 | 0.25 | 0,00 | 0.09 | 0.16 |
| tRNA-Cys-GCA-15-1 | 0.68 | 11.81 | 0.40 | 0.00 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 |
| tRNA-Cys-GCA-16-1 | 2.51 | 1.68 | 1.59 | 0.00 | 0.08 | 0.00 | 0.90 | 0.54 | 0.31 |
| tRNA-Cys-GCA-17-1 | 0.00 | 0.00 | 0,00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| tRNA-Cys-GCA-18-1 | 0.07 | 1.28 | 0.08 | 0.00 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 |
| tRNA-Cys-GCA-19-1 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.00 |

**Supplementary Table 4. The tRNA-Cys-GCA isodecoders detected by YAMAT-Seq in MCF7, SKRB3 and BT20 cell lines** expressed with a plus sign "+", + meaning 0-20 %; ++ meaning 20-40 %; +++ meaning 40-60 %; ++++ meaning 60-80 %; +++++ meaning 80-100 % of expression (100 % being the highest scoring value).

| | Expression analysis | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gene name | MCF7_A | MCF7_B | MCF7_C | Average | SKBR3_A | SKBR3_B | SKBR3_C | Average | BT20_A | BT20_B | BT20_C | Average |
| tRNA-Cys-GCA-1-1 | + | + | + | + | + | + | 0 | + | + | + | + | + |
| tRNA-Cys-GCA-2 | +++++ | +++++ | +++++ | +++++ | +++++ | +++++ | +++++ | +++++ | +++++ | ++++ | +++++ | ++++ |
| tRNA-Cys-GCA-3-1 | + | 0 | 0 | ND | 0 | 0 | 0 | ND | 0 | 0 | 0 | ND |
| tRNA-Cys-GCA-4-1 | ++ | +++++ | +++ | +++ | + | + | + | + | ++ | ++ | ++ | ++ |
| tRNA-Cys-GCA-5-1 | +++++ | + | +++++ | ++++ | + | + | + | + | ++ | ++ | ++ | ++ |
| tRNA-Cys-GCA-6-1 | + | + | + | + | + | + | + | + | + | + | + | + |
| tRNA-Cys-GCA-7-1 | + | + | + | + | 0 | 0 | + | ND | 0 | + | + | + |
| tRNA-Cys-GCA-8-1 | + | + | + | + | + | + | + | + | + | + | + | + |
| tRNA-Cys-GCA-9 | + | + | + | + | + | + | + | + | + | + | + | + |
| tRNA-Cys-GCA-10-1 | 0 | 0 | + | ND | 0 | 0 | 0 | ND | 0 | 0 | 0 | ND |
| tRNA-Cys-GCA-11-1 | 0 | ++ | + | + | 0 | 0 | 0 | ND | 0 | 0 | 0 | ND |
| tRNA-Cys-GCA-12-1 | + | - | + | + | + | + | 0 | + | + | + | 0 | + |
| tRNA-Cys-GCA-13-1 | + | + | + | + | + | + | + | + | + | + | + | + |
| tRNA-Cys-GCA-14-1 | + | + | + | + | + | + | + | + | 0 | + | + | + |
| tRNA-Cys-GCA-15-1 | + | + | + | + | 0 | 0 | + | ND | 0 | 0 | 0 | ND |
| tRNA-Cys-GCA-16-1 | + | + | + | + | 0 | + | 0 | ND | + | + | + | + |
| tRNA-Cys-GCA-17-1 | 0 | 0 | 0 | ND | 0 | 0 | 0 | ND | 0 | 0 | 0 | ND |
| tRNA-Cys-GCA-18-1 | + | + | + | + | 0 | 0 | 0 | ND | + | 0 | 0 | ND |
| tRNA-Cys-GCA-19-1 | 0 | 0 | 0 | ND | 0 | 0 | 0 | ND | 0 | + | 0 | ND |

**Supplementary Table 5. All tRNA-Cys isodecoders detected by aforementioned methods** summarized with their expression levels expressed with a plus sign (+), + meaning 0-20 %; ++ meaning 20-40 %; +++ meaning 40-60 %; ++++ meaning 60-80 %; +++++ meaning 80-100 % of expression (100 % being the highest scoring value).

| Gene name | tRNA-Seq in HEK293T | | ChiP-Seq in cell lines | | | YAMAT-Seq in cell lines | | |
|---|---|---|---|---|---|---|---|---|
| | CTRL(3.1) | CTRL(4.1) | HepG2 | Huh7 | Liver-Adult | MCF7 | SKBR3 | BT20 |
| tRNA-Cys-GCA-1-1 | ++ | +++ | + | + | + | + | + | + |
| tRNA-Cys-GCA-2-1 | +++++ | + | ND | ND | ND | +++++ | +++++ | ++++ |
| tRNA-Cys-GCA-2-2 | +++++ | +++++ | ++ | +++ | ++ | | | |
| tRNA-Cys-GCA-2-3 | + | + | +++++ | +++++ | +++++ | | | |
| tRNA-Cys-GCA-2-4 | +++ | ++ | ND | ND | ND | | | |
| tRNA-Cys-GCA-3-1 | + | + | ND | ND | ND | ND | ND | ND |
| tRNA-Cys-GCA-4-1 | + | + | ++++ | +++++ | ++++ | +++ | + | ++ |
| tRNA-Cys-GCA-5-1 | ++ | + | + | + | + | ++++ | + | ++ |
| tRNA-Cys-GCA-6-1 | ++ | + | ND | ND | ND | + | + | + |
| tRNA-Cys-GCA-7-1 | + | + | ND | ND | ND | + | ND | + |
| tRNA-Cys-GCA-8-1 | + | + | + | + | + | + | + | + |
| tRNA-Cys-GCA-9-1 | + | + | ND | ND | ND | + | + | + |
| tRNA-Cys-GCA-9-2 | + | +++ | + | + | + | | | |
| tRNA-Cys-GCA-9-3 | ++ | +++ | ND | ND | ND | | | |
| tRNA-Cys-GCA-9-4 | + | +++ | + | + | + | | | |
| tRNA-Cys-GCA-10-1 | + | + | ND | ND | ND | ND | ND | ND |
| tRNA-Cys-GCA-11-1 | + | + | + | + | + | + | ND | ND |
| tRNA-Cys-GCA-12-1 | ++ | ++++ | ND | ND | ND | + | + | + |
| tRNA-Cys-GCA-13-1 | + | + | ND | ND | ND | + | + | + |
| tRNA-Cys-GCA-14-1 | + | + | +++ | ++++ | +++ | + | + | + |
| tRNA-Cys-GCA-15-1 | + | + | + | + | + | + | ND | ND |
| tRNA-Cys-GCA-16-1 | + | + | ND | ND | ND | + | ND | + |
| tRNA-Cys-GCA-17-1 | +++++ | +++++ | +++ | +++++ | ++ | ND | ND | ND |
| tRNA-Cys-GCA-18-1 | + | + | ND | ND | ND | + | ND | ND |
| tRNA-Cys-GCA-19-1 | + | + | ND | ND | ND | ND | ND | ND |
| tRNA-Cys-GCA-20-1 | + | + | + | + | + | ND | ND | ND |
| tRNA-Cys-GCA-21-1 | ++ | ++ | ND | ND | ND | ND | ND | ND |
| tRNA-Cys-GCA-22-1 | + | + | ND | ND | ND | ND | ND | ND |
| tRNA-Cys-GCA-23-1 | + | + | ND | ND | ND | ND | ND | ND |
| tRNA-Cys-GCA-24-1 | + | + | + | + | + | ND | ND | ND |
| tRNA-Cys-GCA-25-1 | + | + | ND | ND | ND | ND | ND | ND |
| tRNA-Cys-ACA-1-1 | + | + | ND | ND | ND | ND | ND | ND |

**Supplementary Table 6**. **All tRNA-Cys-GCA isodecoders tested for their ability to increase UGA readthrough** summarized with their numerical readthrough levels in p2luci construct compared to the reporter with no tRNA (empty vector, eV). The changes in the readthrough levels relative to the reporter with no tRNA (eV) were analysed by Student's t-test (mean+SD; n=2); statistically significant increases are marked with asterisk (*) as follows: P < 0.01* and P < 0.05** and P < 0.1***.

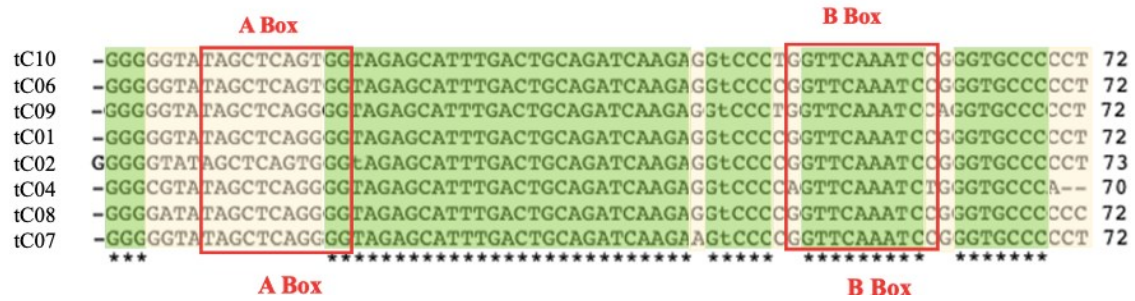
| Name | Mature tRNA score | Empty vector readthrough mean in % (p2luci) | SD (eV) | Tested tRNA readthrough mean in % (p2luci) | SD (tested tRNA) | Student's t-test |
|---|---|---|---|---|---|---|
| **tC10** | 81.8 | 3.11 | 0.005 | 6.02 | 0.11 | 26.52* |
| **tC06** | 81.9 | 3.11 | 0.005 | 4.92 | 0.30 | 6.11* |
| **tC09** | 77.2 | 3.31 | 0.30 | 4.54 | 0.16 | 3.68* |
| **tC01** | 77.3 | 3.11 | 0.005 | 4.04 | 0.24 | 3.79* |
| **tC02** | 78.7 | 3.31 | 0.30 | 4.16 | 0.07 | 2.82 |
| **tC07** | 71.6 | 3.11 | 0.005 | 3.39 | 0.20 | 1.36 |
| **tC03** | 46.2 | 3.31 | 0.30 | 3.33 | 0.08 | 0.09 |
| **tC04** | 64.9 | 3.31 | 0.30 | 3.27 | 0.31 | 0.09 |
| **tC08** | 70.7 | 3.11 | 0.005 | 3.07 | 0.04 | 1.17 |

**Supplementary Table 7. All tRNA-Cys-GCA isodecoders tested for their ability to increase UGA readthrough** summarized with their numerical readthrough levels in pSGDluc construct compared to the reporter with no tRNA (empty vector, eV). The changes in the readthrough levels relative to the reporter with no tRNA (eV) were analysed by Student's t-test (mean+SD; n=2); statistically significant increase with P < 0.01 is marked with asterisk (*).

| Name | Mature tRNA score | Empty vector readthrough mean in % (pSGDluc) | SD (eV) | Tested tRNA readthrough mean in % (pSGDluc) | SD (tested tRNA) | Student's t-test |
|---|---|---|---|---|---|---|
| **tC10** | 81.8 | 4.27 | 0.14 | 6.54 | 0.06 | 14.43* |
| **tC07** | 71.6 | 4.10 | 0.10 | 4.49 | 0.14 | 2.29 |
| **tC09** | 77.2 | 4.27 | 0.14 | 4.82 | 0.19 | 2.28 |
| **tC06** | 81.9 | 4.10 | 0.10 | 4.34 | 0.18 | 1.18 |
| **tC08** | 70.7 | 4.27 | 0.14 | 3.79 | 0.30 | 1.46 |

**Supplementary Table 8. The UGA readthrough levels of artificially prepared tRNA-Gln-GCA construct** measured in both p2luci and pSGDluc vectors.

| Name | Empty vector readthrough mean in % | SD (eV) | Tested tRNA readthrough mean in % | SD (tested tRNA) | Student's t-test |
|---|---|---|---|---|---|
| p2luci_tRNA-Gln-GCA | 3.53 | 0.50 | 2.76 | 0.08 | 1.52 |
| pSGDluc_tRNA-Gln-GCA | 4.01 | 0.32 | 3.34 | 0.35 | 1.39 |




**Supplementary Figure 1. Alignment of selected tRNA-Cys-GCA isodecoders tested for their ability to increase UGA readthrough**. The sequences of human A Box and B Box can be written as **TRGYNNARNNG** and R**GTTCRA**NNCY, respectively. tC02 is "shifted" in this alignment, since it has an insertion in D-loop, however, the A Box sequence remains unaffected.


**Supplementary Figure 2. Sequence alignment of the most highly expressed tRNA-Cys isodecoders** selected by the expression analysis performed during the course of this thesis.


**Supplementary Figure 3. Sequence alignment of the lowly expressed tRNA-Cys isodecoders** selected by the expression analysis performed during the course of this thesis.