



מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree

Master of Science

Submitted to the Scientific Council of the

Weizmann Institute of Science

Rehovot, Israel

עבודת גמר (תזה) לתואר

מוסמך למדעים

מוגשת למועצה המדעית של

מכון ויצמן למדע

רחובות, ישראל

By

Alisa Greenberg

מאת

אליסה גרינברג

איתור סיסטמטי של גורמים תאיים המשפיעים על דיוק התרגום
תחת הפרעות גנטיות ופיזיולוגיות

Systematic detection of cellular factors affecting
translation accuracy upon physiological and genetic
perturbations

Advisor:

Prof. Yitzhak
Pilpel

מנחה:

פרופסור יצחק
פלפל

January 2019

שבט תשע"ט

Abstract

Protein homeostasis (proteostasis) plays a key role in health and fitness of all living organisms. To maintain proteostasis, cells invest much resources and energy in regulation of each step in proteins' life including translation, folding, modification, trafficking and degradation. Regulation of the proteome is extremely important considering that the loss of proteostasis is related to aging and protein conformational diseases as neurodegenerative diseases, metabolic disorders and cancer. The ribosome plays a major part in maintenance of proteostasis, since its speed and accuracy levels determine the copy number and quality of all newly synthesized proteins.

While DNA mutations and transcription errors are well studied and could be detected by nucleotide sequencing, measuring translation errors, although orders of magnitude more abundant, can be challenging. In this work we used and developed a new translation “error meter” systems to measure the translation errors within different cells and conditions, and in different genetic background. These systems are based on a reporter protein in which a codon of functionally essential amino acid is substituted by codon for a different amino acid, resulting in a non-functional protein whose activity can be rescued by specific amino acid misincorporation.

Luciferase-based error meter enabled us to measure different types of mistranslation across multiple cell lines, drug treatments and stress conditions. These assays had revealed a difference in mistranslation trends between the first and second codon positions, relative to mistranslation of third codon position and stop codon readthrough. Mistranslation rate of the first and second codon positions was more conserved across different cell lines, and found to be relatively resistant to changes imposed by environmental stresses as starvation and antibiotic treatment. A comparison of WI-38 fibroblasts and embryonic kidney cells (HEK) had revealed higher mistranslation rate among the former, as measured by all error meter variants. To our surprise, we have observed higher accuracy of the cancer-analogous WI-38 fast cells, relative to WI-38 slow cells representing primary cell characteristics. Another surprising finding was a reduction in error rate under media starvation conditions, possibly suggesting an adaptation mechanism which increases the accuracy of translation upon nutrient scarcity.

We employed the luciferase system to scan a siRNA library in order to detect genes whose activity affects the translation error rate. Silencing repressors of mTOR/S6K had increased the

error rate, since activation of these pathways leads to ribosome acceleration and reduction of accuracy rising from the innate trade-off between translation speed and fidelity.

Since luciferase signal detection requires cell lysis, we have designed and created a fluorescence-based error meter system which can be employed in living single cells. This new system had revealed a variance of mistranslation levels within a population of genetically identical cells, and could be employed for high-throughput methods in the future.

Table of Contents

Abstract	2
Introduction	5
Goals of the study	7
Materials and methods	7
Cell lines	7
Medium	8
Plasmids	8
Cloning	9
Cell transfection	13
Dual luciferase assay	14
FACS analysis	14
Statistical analysis	14
Results	15
Translational accuracy of different cell lines	15
Measurement of G418-induced translation errors	20
Effect of media starvation on translational accuracy	24
Scanning siRNA library for translation accuracy modifiers	27
In-silico simulation of eGFP-based error meter	33
Design of enzyme-based fluorescent error meter	36
Application of zipGFP-TEV error meter for detection of starvation-induced stress	39
Discussion	41
Literature	45

Introduction

Individual cells and whole organisms are constantly exposed to intrinsic and environmental stress, which might damage the proteome and cause loss of protein homeostasis¹. In healthy cells maintenance of proteostasis is achieved by highly regulated and coordinated activity of several cellular processes such as protein synthesis, post-translation modification, folding, trafficking and degradation². The vast amount of resources and energy cells invest in these systems indicate their importance for the cellular health and fitness, as well as the potential harm of erroneous proteins³. From translation to degradation, Each step in the lifetime of a protein molecule could be impaired by acute stress and chronic accumulation of damage, which might lead to protein conformational diseases as neurodegenerative diseases, metabolic disorders, and even cancer^{2,4}.

The ribosome and the protein synthesis machinery have major influence on the cellular proteome, and modulating different components of these systems was shown to protect the proteome from possible damage⁵. Specific mutations or depletion of some ribosomal proteins and translation factors can affect the lifespan of evolutionarily disparate organisms such as yeast, worms, flies and mice¹. In addition, a comparison of rodent species with wide range of lifespans had found strong negative correlation between the translation error rate and the maximal life span of the species⁶, i.e. long-lived species featured more accurate translation machinery.

Accuracy is a critical attribute of the ribosome since translation errors might lead to production of dysfunctional and misfolded proteins, leading to reduced cellular efficiency and putting the cell in danger of forming toxic aggregates. While genetic mutations and transcription errors can be detected by DNA and RNA sequencing, measuring translation errors, although orders of magnitude more abundant⁷, can be more challenging.

In this work, we will use a translation “error meter” system to measure the averaged translation error rate within a cell. The system is based on a reporter protein in which a codon of functionally essential amino acid is substituted by codon for a different amino acid resulting in a non-functional protein. Translation error (and also order-of-magnitude-less- frequent transcription errors) can rescue the protein’s activity by misincorporating the functional amino acid at the mutated position, leading to the appearance of a detectable signal. The rate of translation errors thus can be deduced from the mutant reporter’s signal.

The luciferase error meter system can be applied for detection and measurement of translation error in multiple cellular conditions and treatments, revealing their effect on the cellular translation accuracy and proteostasis. However, for technical reasons the system cannot be used in living cells, a limitation that complicates its application for scanning a library of individual living cells that harbor a deletion of a certain gene each.

The issue of measuring translation errors in living cells in high-throughput manner can be addressed by designing a system based on expression and measurement of fluorescent signals. In this project we have designed and built GFP based error-meter that allows highly sensitive, single cell measurement of translation error rate. Our initial plan was to base our error meter on the GFP gene that will be inactivated by mutation of functionally essential codon. However, after simulating this system in-silico we realized that it won't be a suitable translation "error meter", since the predicted fluorescent signal falls below the detection limit of the measurement method (FACS) we planned to use. Considering the limited number of GFP molecules per cell, the relatively weak signal of single GFP molecule and the error rate known from the literature, we decided to construct a signal-amplifying "error meter".

zipGFP is a system of split GFP whose signal depends on cleavage by the TEV protease⁸. The activity of TEV protease can be reduced significantly as a result of mutation in one of the amino acids of this enzyme's catalytic triad^{9,10}, preventing it from cleaving and thus activating the fluorescent signal of zipGFP. As TEV protease can catalyze and activate multiple zipGFP molecules, a single translation error that rescue the activity of TEV results in activation of many GFP molecules. here we show that co-transfection of zipGFP and mutated TEV protease can be used as a translation "error meter", and the fluorescent signal can be detected and analysed by flow cytometer.

Implementation of both "error meter" systems in multiple cell lines, cellular conditions and genetic backgrounds will provide better understanding of the cellular factors and mechanisms affecting translation accuracy. We have observed variation in translation error rates between environmental conditions, different cell line and even individual cells within a population. However, some general trends consistently appear across all of these conditions, implicating the existence of mechanisms regulating the rate of different types of mistranslation. We believe this project will expose some of the underlying mechanisms and conditions balancing the cellular need

for constant protein synthesis while maintaining sufficient accuracy for maintenance of a healthy proteome.

Goals of the study

The goals of this study were to construct and apply translation “error meter” systems in order to reveal the relation between multiple genes, cell lines and cellular conditions on the accuracy of protein synthesis. To address this goal, we planned to:

1. Measure the translation accuracy of cells under different genetic, physiological and environmental conditions. We planned to achieve this goal by employing the luciferase translation error meter system in cell lines with different genetic background, as well as well as treatment with drugs affecting the ribosome and creating environmental stress by medium starvation.
2. Identify genes that contribute to the translation fidelity of the cell. We planned to do so by scanning a siRNA library with the luciferase translation error meter system.
3. Construct new and employ new translation “error meter” system on the basis of a fluorescent protein, suitable for measuring errors in individual living cells. This system will contain separate constantly expressed fluorescent reporters for signal normalization, and antibiotic resistance genes for future selection of stable clones.

Materials and methods

Cell lines

- WI-38 immortalized cells* - This cell line was created in Varda Roter’s lab from primary human embryonic lung fibroblasts (WI-38) and were described previously¹¹¹². Briefly, The primary cells were infected with a recombinant retrovirus encoding the telomerase hTERT, in order to allow the bypass of replicative senescence. WI-38/ hTERT^{slow} (refer to as WI38 slow), is an early progeny of the immortalized cells, collected 150 population doublings (PDLs) after the telomerase infection. These cells are characterized with sensitivity to contact inhibition and low proliferation rate, comparable to the primary cells. In contrast, WI-38/ hTERT^{fast} (refer to as WI38 fast) were collected ~350 PDLs after the infection, at a time

when the culture had accumulated genetic alterations. and significantly accelerated its proliferation rate. Unlike WI38 slow and primary cells, WI38 fast is resistant to contact inhibition and differs in gene expression patterns and Karyotype^{11,12}.

- ii. *HEK 293* – Human embryonic kidney epithelial cells. This cell line was kindly given us by Prof. Yosef Shaul's lab.
- iii. *HEK 293T*- A highly transfectable derivative of human embryonic kidney 293 cells. This cell line was kindly given us by Prof. Noam Stern-Ginossar's lab.
- iv. *HeLa* – Human epithelial cell line originates from cervical cancer tissue. This cell line was kindly given to us by Prof. Menachem Rubinstein.

Medium

- i. MEM-EAGLE with Earle's salts base (BI, Beit-Haemek, Israel), supplemented with 10% FCS, 1% penicillin/ streptomycin (P/S) and 1% L-Glutamine. This medium was used as a base medium for WI38 slow and fast cells.
- ii. High glucose Dulbecco's Modified Eagle Medium (BI, Beit-Haemek, Israel, Supplemented with 10% FCS, 1% penicillin/ streptomycin (P/S) and 1% L-Glutamine. This medium was used as base medium for HeLa, HEK 293 and HEK 293T cell lines. For dualZIP plasmids selection, 500 µg/ml Geneticin (G418) was added to the medium and refreshed every two days.
- iii. *For serum starvation assay*, Low glucose (1 g/l D-Glucose) Dulbecco's Modified Eagle Medium, supplemented with 1% FCS, 1% penicillin streptomycin (P/S) and 1% L-Glutamine was used

Plasmids

- i. *Luciferase* - The luciferase plasmids collection was described previously⁶, and was a kind gift from Prof. Andrei Seluanov's and Vera Gorbunova's lab^{6,13}. The collection consists of wild-type Firefly and Renilla luciferase plasmids, and four plasmids encoding a mutated versions of the Firefly reporter. The catalytically important amino acid Lysine 529¹⁴ (AAA) was mutated in the first, second and thirds positions to Glutamic acid (GAA), Isoleucine (ATA) and Asparagine (AAT), respectively. Premature stop codon was inserted by mutating Cysteine 81

(CGA) into stop codon (TGA). These mutants are catalytically impaired or truncated proteins, while translation mistake of amino acid misincorporation or stop codon readthrough can rescue the activity of the reporter.

- ii. *zipGFP and TEV plasmids* – ZipGFP1-10_TEV, ZipGFP11_TEV and pcDNA3.1 TEV was a gift from Xiaokun Shu (Addgene plasmids # 81242, #81243 and #64276). Self-assembling split GFP consists of two fragments, one containing the first 10 β -sheets (GFP1-10) and the second part is the last β -sheet (GFP11)⁸. In order to prevent spontaneous assembly, each GFP part was flanked on both termini with the heterodimerizing “zipper-like” E5 and K5 coiled coils¹⁵. The larger GFP1-10 fragment is linked to mCherry fluorophore by viral self-cleaving peptide T2A¹⁶, thus allowing normalization of the GFP signal by the constant co-expression of mCherry.

Cloning

- i. *DualZIP* – restriction free cloning by Gibson assembly was used in order to create a single plasmid expressing both zipGFP fragments. The vector plasmid zipGFP1-10 and the zipGFP11 sequence insert were amplified separately by PCR. Primers for zipGFP11 insert include overlapping regions to the zipGFP1-10 vector. The FWR primer has an extension of the viral self-cleaving peptide P2A (marked in green)¹⁶. Primers for zipGFP1-10 vector include overlapping regions of the zipGFP11 insert and P2A peptide. Primers used for PCR amplification:

primer	Forward primer	Reverse primer
vector zipGFP1-10	CCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC CAG	GCTTCAGCAGGCTGAAGTTA GTAGCTCCGCTTCCTTACTTG TACAGCTCGTCCATGC
insert P2A- zipGFP11	AAGTAAGGAAGCGGAGCTACTA ACTTCAGCCTGC TGAAGCAGGCTGGCGACGTGGAGGAGAACCCTGG ACCTAAGCTTGCCACCATGGGC	GGCAACTAGAAAGGCACAGT CGAGGCTGATCAGCGGG

The PCR amplification reactions were conducted using iProof master mix (X2), according to standard protocol, with $T_m = 58^\circ\text{C}$. PCR products were treated with DpnI enzyme (NEB).

After PCR clean-up (Promega), samples were run in 1% agarose gel to ensure that the PCR product is composed of a single amplicon in the appropriate size. Gibson assembly reaction was conducted using NEBuilder HiFi DNA Assembly Master Mix (NEB) according to standard protocol¹⁷. To find recombinant plasmid, colonies that grow under ampicillin selection were sequenced verified. Primers used for sequencing are Forward CMV promoter (CGCAAATGGGCGGTAGGCGTG) and Forward primer annealing in the middle of mCherry (GAGGACTACACCATCGTGG). The second cloning step was the removal of the original stop codon of the mCherry gene, in order to enable the expression of downstream P2A- zipGFP11. “Back to back” restriction free cloning was used to remove the stop codon, where the primers anneal at either side of the targeted deletion sequence. Whole plasmid amplification proceeds outwards from this area, thus excluding this region from the PCR product. Primes used for amplification are:

Forward mCherry stop codon remove – GGAAGCGGAGCTACTAACTTCAG, Reverse mCherry stop codon remove – CTTGTACAGCTCGTCCATGCCG. The primers were phosphorylated using T4 PNK (NEB) protocol before the PCR reaction. The PCR amplification was conducted using iProof master mix (X2) according to standard protocol, with $T_m = 58.5^\circ\text{C}$. Original plasmids were degraded using Dpn1, followed by PCR clean-up and room-temperature overnight ligation of the plasmids’ ends using T4 DNA ligase (NEB). The ligation product was transformed into DH5 α competent bacteria, and colonies that grow under ampicillin selection were tested by sequencing of the purified plasmid using plasmid miniprep kit. Forward primer annealing in the middle of mCherry (GAGGACTACACCATCGTGG) was used for sequencing.

- ii. *TEV-BFP* – restriction free cloning by Gibson assembly was used in order to link TEV protease to tagBFP fluorophore in order to enable us to control for transfection efficiency and expression levels of the enzyme. A plasmid containing Hygromycin resistance gene linked by T2A to the blue fluorophore tagBFP was kindly given to us by Prof. Igor Ulitsky’s lab. This plasmid is based on addgene plasmid #62348 where the antibiotic resistance gene was changed from Puromycin to Hygromycin. The vector plasmid pcDNA3.1 TEV (addgene) and the T2A-BFP sequence insert were amplified separately by PCR. Primers for each one of the fragment were designed to have flanking regions with overlap to the other fragment.

Primer	Forward primer	Reverse primer
TEV vector	CTGTGCCTTCTAGTTGCCAGC	TCTAGAGCGGCCTTCTTTAAGCTG AGTGGCTTCCTTAAC
T2A-BFP insert	TTAAGGAAGCCACTCAGCTTAAAGAAGGCC GCTCTAGAGGAGAGGGCAGAGGAAGTCTCC	AGGGGCAAACAACAGATGGCTGG C

The PCR amplifications were conducted using iProof master mix (X2), according to standard protocol, with $T_m = 68^\circ\text{C}$ and $T_m = 61^\circ\text{C}$ for the vector and insert (respectively). Original plasmids were degraded using Dpn1, followed by PCR clean-up and running the cleaned products in 1% agarose gel to ensure that the PCR product is composed of a single amplicon in the appropriate size. Gibson assembly reaction was conducted using NEBuilder HiFi DNA Assembly Master Mix (NEB) according to standard protocol¹⁷. Primers used for sequencing are T2A validation – CTGTTCAGGAGGAATAACGGCAC (anneals at the middle of TEV), and BFP validation – GCTAGTAGCCAGGATGTTCG (anneals at the middle of tagBFP). After several failed attempts to remove the stop codon at the end of TEV using Gibson assembly and “back to back” cloning, we have consulted the cloning unit about other methods we can use. With the kind help of Dr. Yoav Peleg we applied Transfer-PCR (TPCR)¹⁸, a restriction-free cloning method in which the fragment amplification and integration happen in a single PCR reaction. A single pair of primers was designed to anneal on each side of the sequence targeted for deletion: F_tPCR_stop remove- AAAGAAGGCCGCTCTGGAGAGGGCAGAGGAAGTC, R_tPCR_stop remove- GTATGTGGTGACTCTCTCCCATG. The PCR amplifications were conducted using 10ng of vector and insert plasmids (pcDNA3.1 TEV and T2A-tagBFP), 20nM of forward and reverse primers, dNTPs (10mM each), 10ul of 5x Phusion buffer and 0.8ul of 2U/ μl Phusion polymerase (Finnzymes) in final volume of 50ul. The PCR program included 30 cycles, with $T_m = 60^\circ\text{C}$, with 5 minute elongation step. To eliminate the original plasmid, PCR products were incubated with 1 μl DpnI enzyme for 1 hour at 37°C , and transformed into DH5 α competent bacteria with no heat inactivation of DpnI reaction or PCR product clean up. Colonies that grow under ampicillin selection were tested by sequencing of the purified plasmid using plasmid miniprep kit. Primers used for sequencing are T2A validation – CTGTTCAGGAGGAATAACGGCAC (anneals at the middle of TEV), and BFP validation – GCTAGTAGCCAGGATGTTCG (anneals at the middle of tagBFP).

- iii. *TEV antibiotic resistance replacement* –Neomycin is the mammalian selection marker found in pcDNA3.1 TEV and the zipGFP plasmids zipGFP1-10, zipGFP11 and the newly cloned dualZIP plasmids. In order to enable selection for cells expressing both TEV and dualZIP, the Neomycin resistance of pcDNA3.1 TEV was replaced by Hygromycin resistance gene. Restriction free Gibson assembly was performed to insert HygR resistance gene from the Hyg-T2A-BFP kindly given to us by Prof. Igor Ulitsky’s lab. Megaprimers for each one of the fragment were designed to have flanking regions with overlap to the other fragment.

Primer	Forward primer	Reverse primer
Hygromyc in insert	GACAGGATGAGGATCGTTTCGATGTC TGTCGAGAAGTTTCTGATCG	GTCGCTTGGTCGGTCATTTCTTCCTCT GCCCTCTCCTCC
TEV vector	GGAGGAGAGGGCAGAGGAAGAAATG ACCGACCAAGCGAC	CGATCAGAACTTCTCGACAGACATC GAAACGATCCTCATCCTGTC

The PCR amplifications were conducted using iProof master mix (X2), according to standard protocol, with $T_m = 57^\circ\text{C}$. Original plasmids were degraded using Dpn1, followed by PCR clean-up and running the cleaned products in 1% agarose gel to ensure that the PCR product is composed of a single amplicon in the appropriate size. . Gibson assembly reaction was conducted using NEBuilder HiFi DNA Assembly Master Mix (NEB) according to standard protocol¹⁷. Primers used for sequencing are HygR validation – CAATGTCCTGACCGACAATG (anneals in the middle of HygR).

- iv. *TEV protease mutagenesis* – Creation of translation “Error meter” system based on TEV protease requires a mutant version of TEV whose activity can be rescued by misincorporation of the WT amino acid during translating of the mutated codon. Candidates for mutagenesis were codons for the catalytic triad residues His46, Asp81, and Cys151¹⁰, whose mutation in the structurally similar enzyme lead to significant reduction in activity⁹. H46R was chosen since it was shown to reduce the activity of TEV and can be represented both in near-cognate and non-cognate codons to Histidine. The mutagenesis was done by “back to back” restriction free cloning was used to remove the stop codon, where the primers anneal at either side of the targeted deletion sequence. Whole plasmid amplification proceeds

outwards from this area, thus excluding this region from the PCR product. Primes used for amplification are:

Primer	Forward primer	Reverse primer
H46R near-cognate	CTG TTCAGGAGGAATAACGGCAC	GCGCTTGTTGGTGATGATGAAGGGGC
H46R non-cognate	CTG TTCAGGAGGAATAACGGCAC	CCTCTTGTTGGTGATGATGAAGGGGC

The primers were phosphorylated using T4 PNK (NEB) protocol before the PCR reaction. The PCR amplification reactions were conducted using iProof master mix (X2), according to standard protocol, with $T_m = 60^\circ\text{C}$. Original plasmids were degraded using Dpn1, followed by PCR clean-up and room-temperature overnight ligation of the plasmids' ends using T4 DNA ligase (NEB). The ligation product was transformed into DH5 α competent bacteria, and colonies that grow under ampicillin selection were tested by sequencing of the purified plasmid using plasmid miniprep kit. Forward T7 sequencing primer TAATACGACTCACTATAGGG was used for sequencing.

Cell transfection

- i. *WI38 fast and slow* – cells were transfected 24 hours post seeding, at approximate 70-80% confluency. Media was replaced 60 minutes before transfection. Appropriate amounts of DNA and Polyjet transfection reagent (Polyplus-transfection SA) were separately diluted into equivalent volumes of serum free MEM. The diluted Polyjet was added into the diluted DNA, vortexed and incubated for 15 minutes in room temperature. After the incubation the Polyjet-DNA mixture was added drop-wise into the freshly changed media of each well.
- ii. *HEK-293, HEK-293T and HeLa cell lines* - cells were transfected 24 hours post seeding, at approximate 70-80% confluency. Appropriate amounts of DNA and jetPEI transfection reagent (Polyplus-transfection SA) were separately diluted into equivalent volumes of 150 mM NaCl provided with the reagent kit. The diluted jetPEI was added into the diluted DNA, vortexed and incubated for 25 minutes in room temperature. After the incubation the jetPEI-DNA mixture was added drop-wise into the freshly changed media of each well.

Dual luciferase assay

Dual-Luciferase assay kit (promega) was used to perform analysis on cells co-transfected with Renilla and Firefly luciferase plasmids (respective ratio of 1:19 in plasmid amount). Cells were harvested and lysed using the supplied passive lysis buffer, and slowly shaken at room temperature for 15 minutes. The lysate was transferred into black 96-well plate, and the luciferase signal was read in Veritas microplate luminometer. Each lysate sample was mixed with 100ul of Luciferase assay reagent followed by measurement of Firefly luciferase luminescence. 100ul of Stop&Glo reagent was added to each well, stopping the activity of Firefly luciferase and activating renilla luciferase. The ratio between Firefly and Renilla luciferase was used as a normalized measure of translational accuracy. In experiments were several cell lines or treatments were applied the normalized signal of Firefly luciferase mutants is presented as percentage of the wild type variant.

FACS analysis

Preliminary FACS experiments were conducted using BD Accuri C6 personal flow cytometer (BD Biosciences). 293T cells were transfected with different combinations of the TEV-zipGFP system plasmids and collected to FACS analysis 24 hours post-transfection. The sample cells were trypsinized, washed and suspended in PBS. GFP and mCherry signals were measured by FL1 and FL3 lasers, respectively. Later experiments including WT and mutant versions of TEV-BFP and dualZIP were performed in the flow cytometry unit using BD LSR II flow cytometer. GFP, mCherry and tagBFP were measured using the following respective lasers: Blue 488nm (detector C), Yellow-Green 561nm (detector D), and Violet 405nm (detector B, based on Alexa Fluor 405 settings). Samples were prepared in similar manner to the preliminary experiments, except of resuspension in FACS buffer instead of PBS. Results analysis was performed using the FlowJo software, where gating was based on cell size, morphology and fluorescence (mCherry and tagBFP) above background signal of unstained cells.

Statistical analysis

Mann-Whitney U test (Wilcoxon rank sum test) was used in comparisons of different groups within an experiment, and normalized results of combined biological repeats. This statistical test was chosen since it does not require the assumption of normal distribution of the population¹⁹, and most of examined populations were non normally distributed. Spearman's ranked

correlation was applied for used for estimation of correlation between order of signal appearances while comparing siRNA library scanning repeats²⁰. P values lower than 0.05 were considered as statistically significant. The calculations were performed using Python and the statistical package Scipy²¹.

Results

We wish to investigate how different genetic and environmental factors affect the translational accuracy of different cell lines. recent work by Ke et al. had demonstrated a strong correlation between translational accuracy and maximal longevity of different rodent species, measured by a luciferase-based translation error meter system⁶. This intriguing result had inspired us to further investigate the effect of genetic and environmental background on translational accuracy, both by the existing luciferase system as well as a new fluorescent system constructed by us.

Translational accuracy of different cell lines

The first aim of the project was to examine the effect of genetic background on the translation error rate. WI-38 fast and WI-38 slow are immortalized cell lines derived from primary human embryonic lung fibroblasts (see Materials and Methods). Even though these cell lines share a common ancestor, they differ in their proliferation rate, sensitivity to contact inhibition, gene expression patterns and karyotype^{11,12}. While WI-38 slow shares many characteristics with the primary cells, WI-38 fast exhibit cancer-like properties as aggressive proliferation, up-regulation of genes associated with malignancy and reduced expression of tumor suppressor genes. The translational accuracy of these cell lines was measured by the luciferase error meter system, based on four firefly luciferase mutants, each inactivated by insertion of either a premature stop codon, or by point mutation in different position at a catalytically important codon. Transfection efficiency and expression levels are normalized by co-transfection of the firefly luciferase variants with WT renilla luciferase. The signal of each luciferase was measured by a luminometer, where cell lysate was mixed with reagents activating the luciferases and measuring their signal (Materials and Methods).

The error patterns of WI-38 fast and slow cell lines were similar to each other, as the variant mutated the third position had significantly higher signal then variants mutated in the other

codon positions and the premature stop codon variant (Figure). This translation error pattern of WI-38 lung fibroblasts resembles the pattern of rodent skin fibroblasts whose error rate was measured with the same system⁶. The rodent fibroblasts had similar error rate for first position, second position and premature stop codon mutants, while the error rate of the third position was significantly higher.

Biological synthesis processes as DNA replication and protein translation have basic trade-off between speed and accuracy, as a result of the kinetic parameters of their components²². We hypothesize that the rapid proliferation rate of WI-38 cells will be accompanied by increased error rate, due to the increased need for newly synthesized proteins in the dividing cells. In contrast to our hypothesis, our results showed exactly the opposite: WI-38fast cells exhibited significantly lower error rate, as measured by most mutated versions of Firefly luciferase (Figure 1).

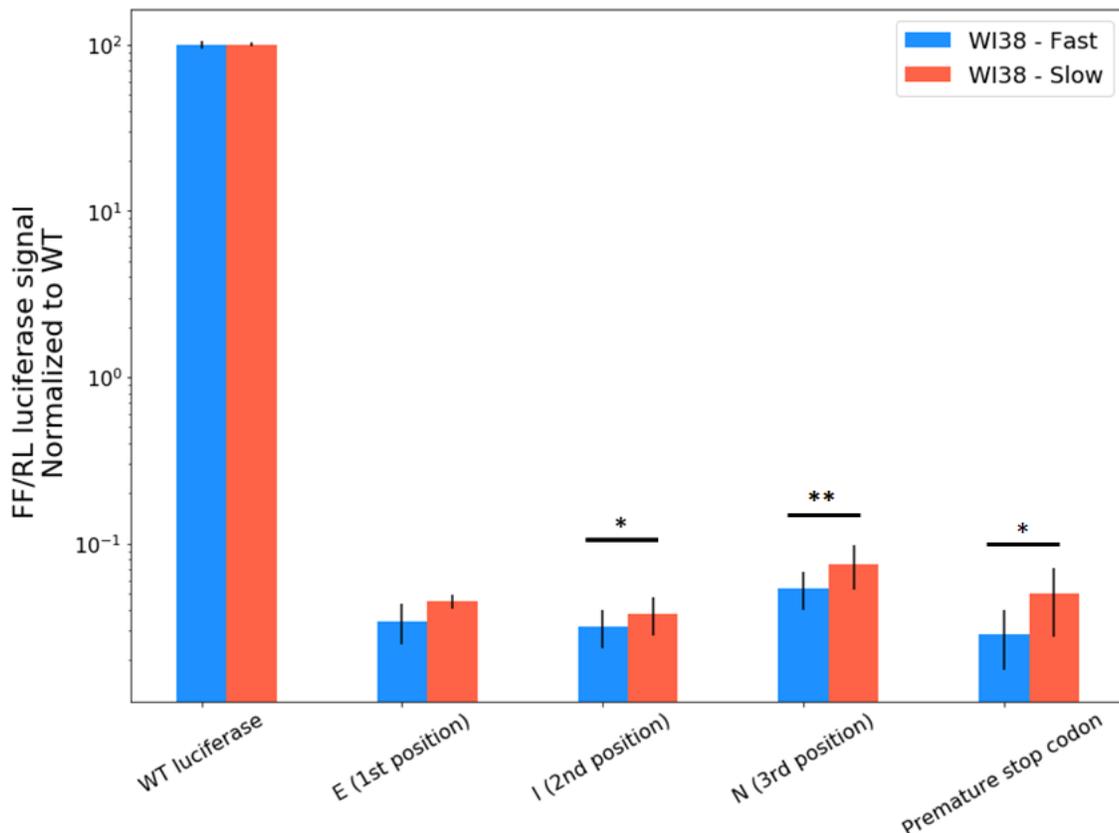


Figure 1 - Comparison of translation mistake rates in WI-38 slow and WI-38 fast cells. The error rates are estimated based on the firefly luciferase signal of the various mutants. The signal is normalized to constant expression of co-transfect renilla luciferase vector. Presented here are the ratio between the

firefly and the renilla luciferases of each mutated version of firefly luciferase normalized to WT firefly signal at the same cell line. The WI-38 fast cells exhibited significantly lower error rate in most of the mutant version of the luciferase error meter. Translation error rates measured by 3rd position mutant were significantly higher in both cell lines, $P < 0.05$. (* P value < 0.05 , ** P value < 0.01 , Mann–Whitney U test).

By applying the luciferase error meter system to human embryonic kidney (HEK) 293 and 293T cells we have discovered an error pattern that differs from the human lung and rodent skin fibroblasts. While WI-38 derived cell lines were prone to make translation errors mainly on the thirds position, both HEK cell-lines had expressed significantly higher error rate in the first and second positions than the thirds and stop codons mutants (Figure 2).

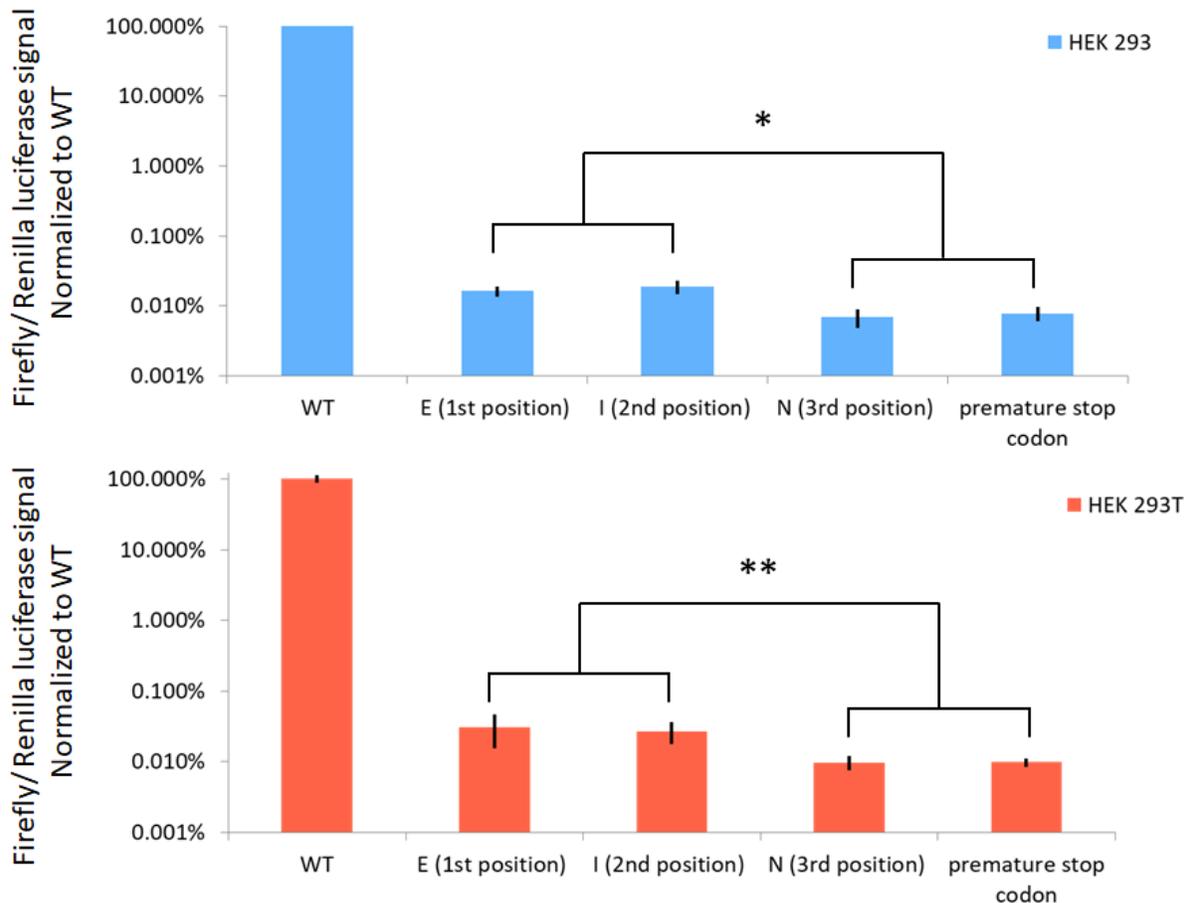


Figure 2 - A comparison of translation mistake rates in HEK 293 and HEK 293T cells. The error rates are estimated based on the firefly luciferase signal of the various mutants, normalized by constant

expression of WT renilla luciferase. Presented here are the ratio between the firefly and the renilla luciferases of each mutated version of firefly luciferase normalized to WT firefly signal at the same cell line. In both cells line the translation errors rate measured by 1st and 2nd position variants were significantly higher than 3rd and premature stop codon variants (* P value <0.05, ** P value <0.01 Mann–Whitney U test).

Besides the variation in mistranslation patterns, WI-38 and HEK derived cell lines differed at the basal error rate measured by most error meter variants. Generally, WI-38 fast and slow cell lines had higher translation error rated compared to HEK 293 and 293T cell lines, as measured by most error meters (Figure 3). All four cell lines maintained across all variants the same relative order of error rates, with WI-38 slow being the most error prone, with accuracy gradually increasing in WI-38 fast, HEK 293T and HEK 293 being the most accurate cell lines. The largest difference between these cell lines was in mistranslation rates at the codon's third position, with 12 fold increase between HEK 293 to WI-38 slow. WI-38 and HEK derived cell lines displayed distinctive error rate in stop codon readthrough, were the signal of WI-38 slow measured by the premature stop codon construct was 8 times higher than the signal of HEK 293. Although the trend of error rates was maintained across all luciferase variants, the differences between the cell lines were lessened in the first and second position error meters (3.6 and 2.2 fold change between the highest and lowest error rates, respectively). By comparing the maximal error rate measured by each error meter and the standard deviation of signals within each error meter group, the second position mutant was found to have the lowest maximal error rate, and most unvarying signal between all error meters. At the other side of the scale, third position error meter had measured the highest error rate, and whose signal was the most volatile signal.

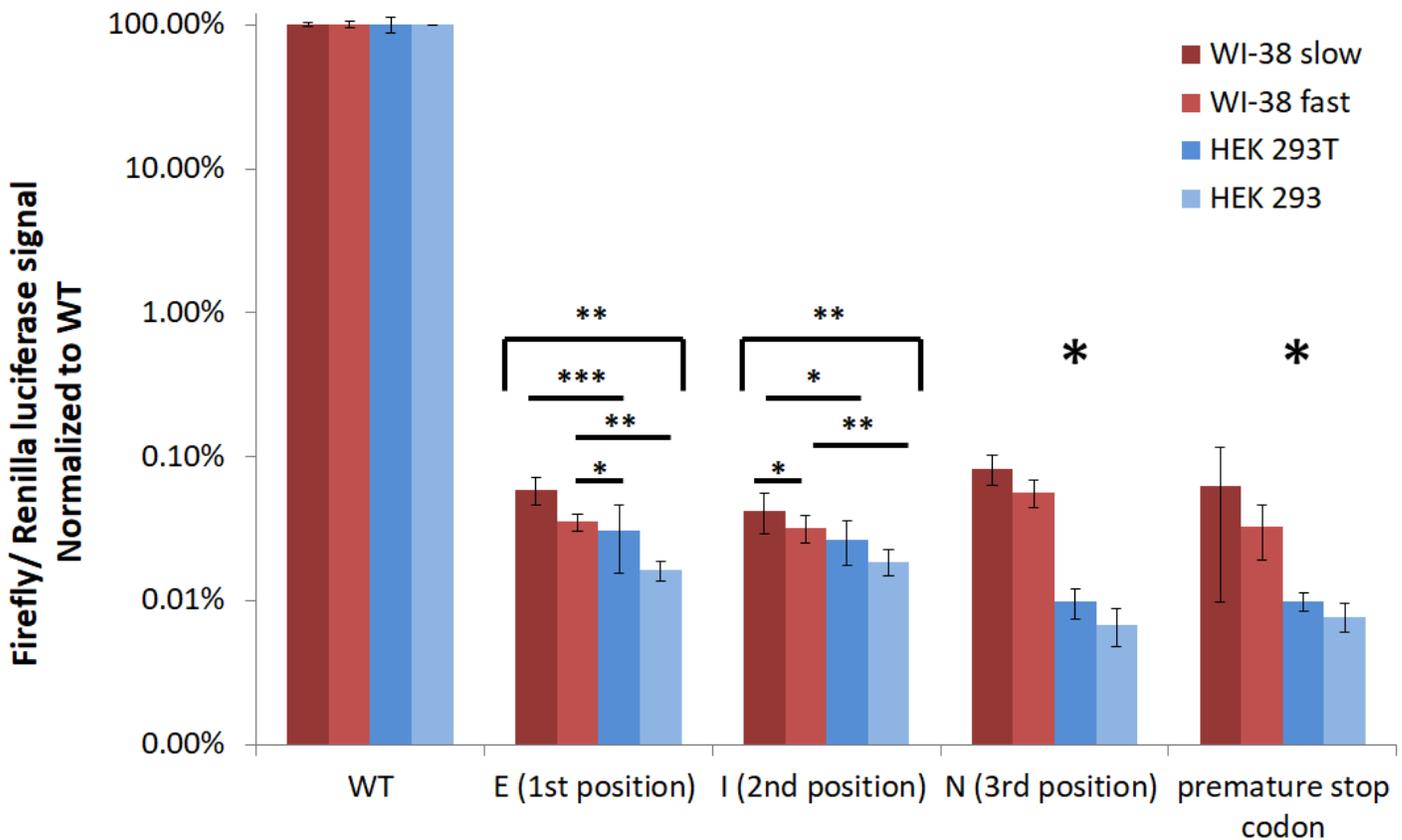


Figure 3 - A comparison of translation mistake rates in WI-38 and HEK derived cell lines. The error rates are estimated based on the firefly luciferase signal of the various mutants, normalized by constant expression of WT renilla luciferase. Presented here are the ratio between the firefly and the renilla luciferases of each mutated version of firefly luciferase normalized to WT firefly signal at the same cell line. Generally WI-38 fast and slow cell lines had higher translation error rates compared to HEK 293 and 293T cell lines, as measured by most error meters (* P value <0.05, ** P value <0.01 Mann–Whitney U test). The large asterisk sign above the 3rd position and premature stop codon variants indicates that all compared pairs of cell lines were significantly different.

A possible explanation for difference in mistranslation trends between the first and second codon positions to the third position and the premature stop codon is the potential damage to the cell resulting from this kind of translation error rates. Translation error in the first position will cause incorporation of different amino acid for all codons except some case of 6-box codons of arginine and leucine. The potential damage of mistranslating the second position is even greater, since it is guaranteed to change the amino acid and the misincorporated amino acid is likely to have opposite chemical characteristics than the encoded amino acid. Hydrophilic amino acids are

encoded by codons with A in the second position, while codons for hydrophobic amino acids have U in the second position. As a result translation error of the second position might change the chemical properties of the amino acid in a way that might disrupt the structural and functional properties of the secondary structure or even the whole protein²³.

On the other hand, translation mistakes of the third position are not as likely to change the amino acid do to redundancy is number of codons relative to amino acids, and the wobble of tRNA molecules on the third position of the codon. Similarly, stop codon readthrough might impose smaller potential damage than translation errors of the first and second positions. Stop codon readthrough differs from other mistranslation events, since the number of stop codons in each protein is significantly smaller than codon for amino acids. In addition, incorporation of any amino acid can dramatically alter the amino acid sequence downstream of the mistranslated stop codon, thus might impose an additional evolutionary cost for the organism. . As a result, the potential damage of stop codon readthrough is reduced by the low probability of this error to occur due to the scarcity of potential targets.

The potential damage of each translation error type might impose different evolutionary pressure on accurate translation of different codon positions and stop codons. Following this hypothesis positions with high penalty for mistranslation will have lower maximal error rate and narrower distribution of error rates between cell line, since the evolutionary pressure would push the translation machinery to be as accurate as possible under the demand to maintain high speed translation²². Types of mistranslation which impose lower danger to the cells would be expected to have variable error rates across different cell lines and organisms, since the lower evolutionary pressure can tolerate wider spectrum of error rate and solutions for the speed-accuracy tradeoff. This hypothesis can explain the gradual increase in signal variability of each error meter, starting from the relatively stable second position variant, through first position, premature stop codon up to the third position error meter.

Measurement of G418-induced translation errors

While the above results show a significant difference in translation error rate and patterns between cell lines, we wanted to convince ourselves that the error meter's signal is a genuine result of mistranslation. In order to do so, we looked for a condition known to increase mistranslation in human cells which will serve us as a positive control for the ability of the

system to measure translational accuracy. Aminoglycosides are a class of antibiotic drugs altering protein synthesis by targeting different sites within functional centers of both ribosomal sub units. While aminoglycosides can affect both prokaryotic and eukaryotic cells, their effect on translation differs between these groups²⁴. While many aminoglycosides efficiently inhibit bacterial protein synthesis by affecting all stages of translation (initiation, elongation, termination and recycling), the main effect on eukaryotic cells is increased rate of stop codon readthrough. As a result some aminoglycosides are commonly used broad-spectrum antibacterial treatments²⁵, while others are being tested as potential treatments for PSC (Premature Stop Codon) associated human diseases²⁴.

We chose the mammalian aminoglycoside Geneticin (G418) for our assay, since it was found to be an effective inducer of stop codon readthrough by PSC luciferase system²⁶. G418 increases mistranslation by binding the decoding center of the small subunit (helix 44) and changing its conformation to one that allows tRNA accommodation, even if it is not the cognate amino acid for the translated codon. Some structural characteristics of the eukaryotic ribosome prevent completely stable binding of G418, thus limiting its capacity to increase the general error rate as it does in bacteria^{24,25}. Nonetheless, the interaction between G418 and the human ribosome is sufficient to allow tRNAs to compete with release factors and accommodate an amino acid when translating a stop codon.

HEK 293 and HeLa cell lines transfected with PSC luciferase mutant and treated with G418 had shown significant increase in error rate in a time depended manner (Figure 4). While both cell lines are genetically sensitive to G418, the treatment has significantly higher effect on HEK 293 cells, whose signal increased linearly and reached its maximal value after 32 hours when the experiment ended (36 fold increase relative to control). The signal of HeLa cells creates a sigmoid curve that reaches a plateau of ~9.5 fold increase after 24 hours.

The strong response of HEK 293 to the treatment compelled us to use it as the cell line in which the effect of G418 will be tested on all the luciferase error meter variants. These cells had a significant and remarkable increase in error rate when treated with different concentrations of G418, displaying an ~80 fold increase when treated with high concentration of 1000 (ug/ml) for 24 hours (Figure 4). These results show the luciferase error meter is a sensitive system for measurement of wide range of error rate.

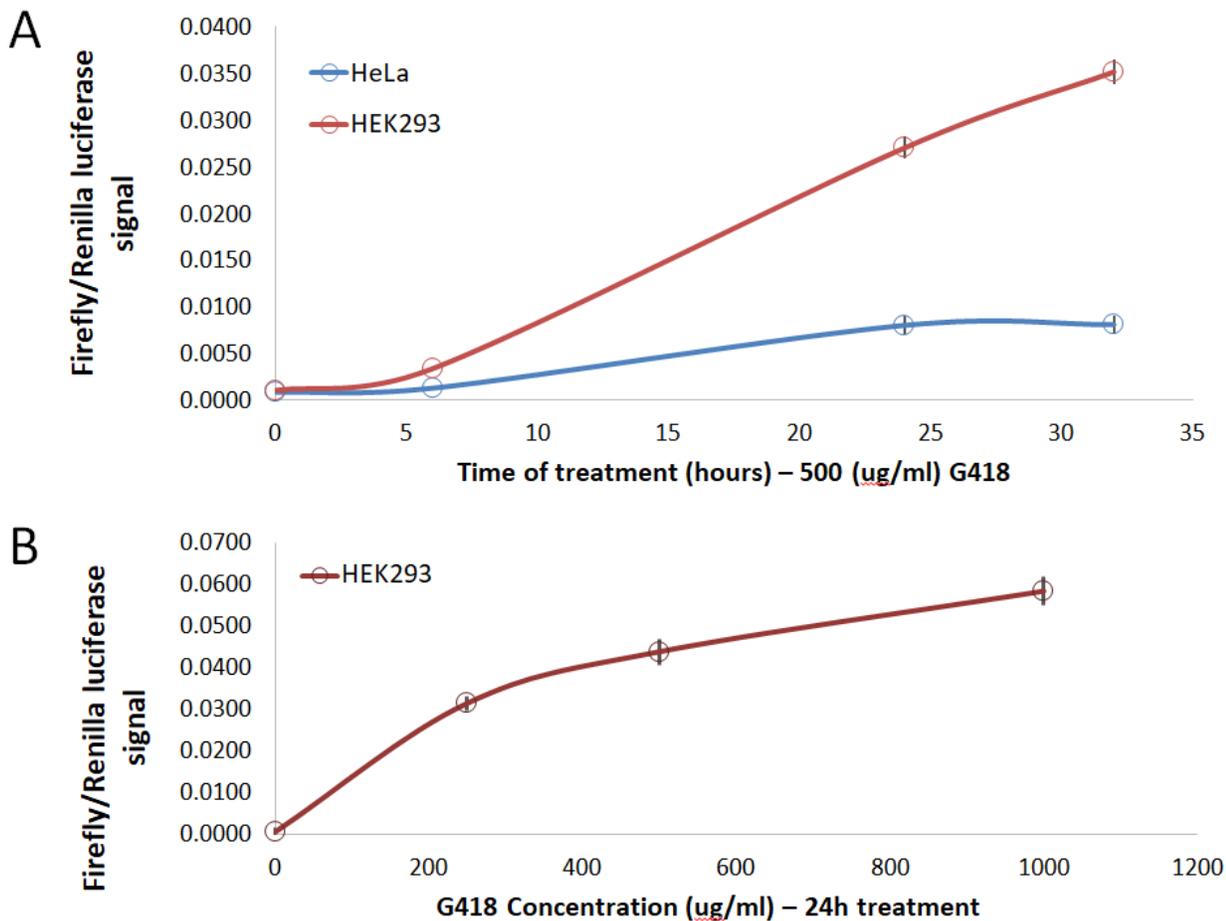


Figure 4 - Effect of the antibiotic Geneticin (G418) on stop codon read through. The error rates are estimated based on the PSC firefly luciferase variant, normalized by constant expression of WT renilla luciferase. Multiple cells lines, treatment durations and drug concentrations were used in order to estimate the optimal conditions for further experiment. (A) HEK 293 and HeLa cell lines were treated with 500(ug/ml) G418 for for 6, 24 and 36 hours before signal reading. The effect of G418 on HEK 293 cells was significantly higher than on HeLa cells for each of the treatment time points. (B) HEK 293 cells were treated with media containing 250, 500 and 1000 ug/ml of G418 for 24 hours. The treatment lead to 44, 61 and 81 fold increase in signal, relative to untreated control cells.

Following these experiments we chose the 24 hour time point and 500 (ug/ml) G418 concentration as our treatment regime for cells expressing the full luciferase error meter system. This treatment induced a 60 fold increase in the signal of PSC luciferase, resembling the results of the concentration-dependent preliminary experiments. In contrast, the amino acid misincorporation rates were mostly unchanged by the G418 treatment. The first and second codon positions were not affected at all by the treatment, and thought the treatment had

significantly increased the error rate in the third position the effect size was relatively small (1.8 fold increase relative to control).

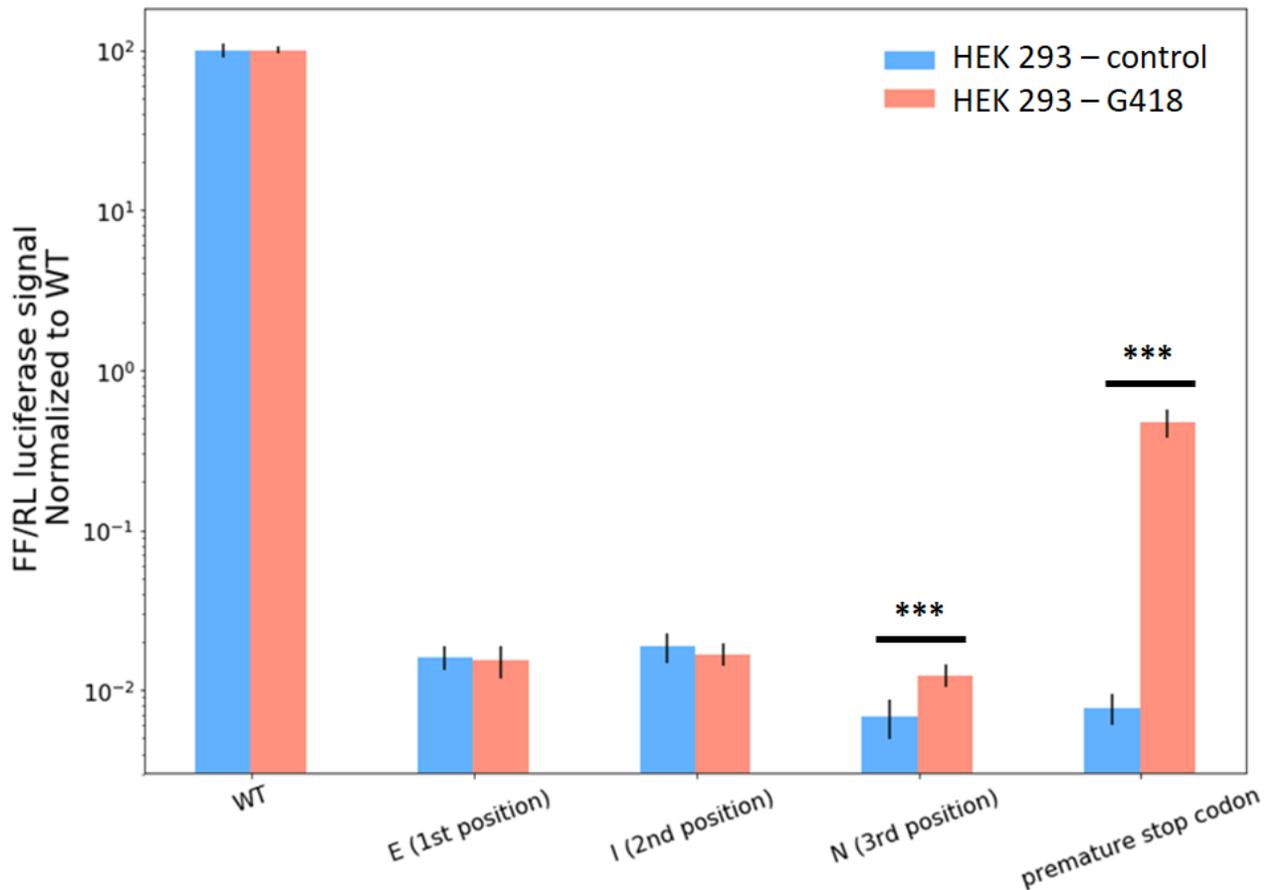


Figure 5 - Effect of the antibiotic Geneticin (G418) on HEK 293 cell line. The error rates are estimated based on the firefly luciferase signal of the various mutants, normalized by constant expression of WT renilla luciferase. Presented here are the ratio between the firefly and the renilla luciferases of each mutated version of firefly luciferase normalized to WT firefly signal at the same treatment condition. The cells were treated with media containing G418 (500 ug/ml) 24 hours before reading of the signal. The treatment had significantly increased the translation error rates on FF variants with mutation at the third position and premature stop codon (approx. 1.8 and 60 fold, respectively). (***) P value <0.001, Mann-Whitney U test).

As expected, G418 treatment had caused a sharp rise in stop codon readthrough rates, but it was hard for us to predict the effect on amino acid misincorporation rate, since the effect of aminoglycosides on eukaryotic and mammalian cells is as studied as in bacteria. Besides the

effect of G418, stop codon readthrough mistakes have are more likely to be detected by the luciferase system, since incorporation of any amino acid would rescue the signal. The error meter variants of the different codon position will detect a signal only when mistranslated with a specific destination amino acid, thus reducing the reversal probability by 1/19.

The structure of the eukaryotic ribosome prevents a stable binding of G418, thus limiting its ability to cause general mistranslation as it does in bacteria²⁴. From all three variants detecting mistranslation at different codon positions, only the third position was affected by G418. This difference between the positions might suggest that the ribosome is more tolerable towards mistranslation of the third position, since it is less likely to change the amino acid sequence of the protein. While this hypothesis is plausible, we need to take in account that HEK 293 had the lowest error rate of all cell lines measured by the luciferase error meter, and that in HEK derived cell lines the third position and PSC variants had the lowest error rate. These two factors make the third position error rate of HEK 293 cells to be the lowest error rate of all position and cell line combination tested by us in the previous section. As a result, conducting the G418 assay on cell lines with different basal error rates and patterns (as WI-38 fast and slow) would be beneficial to our understanding of the ribosomal tolerance for translation errors at different codon positions.

Effect of media starvation on translational accuracy

Following the results of the G418 error induction assay, we wanted to examine the effect of environmental stress on translational accuracy. Our first intuition was to use heat shock stress, but unfortunately the luciferase reporter is extremely sensitive to small temperature changes²⁷. Stress conditions as nutrient starvation and oxidative stress increase mistranslation, both by elevating misloading and mispairing rates. However, some levels of mistranslation might be beneficial to the cell, when deacetylated tRNAs and mistranslated proteins serve as signals for general and oxidative stress response²⁸⁻³⁰. We decided to induce stress trough nutrient starvation, which was found to increase rates of mistranslation in human cells³¹, although the cells in this assay were co-transfected with an error meter and a tRNA gene prone to mischarging by the reversing amino acid.

HEK293T cells expressing the luciferase error meter system were starved for 24 and 48 hours with “starvation media”, containing low amounts of glucose and serum. Media for cells in both starvation and control conditions was changed daily and replaced by fresh media of the same type.

Media starvation of 48 hours had significantly reduced the translation error rate across all luciferase mutant variants. While reduction in mistranslation had appeared in the first and second positions only after 48 hour of starvation, 24 hour treatment was sufficient to reduce the translation error rate measured by the third position and PSC variants (Figure 6). The latter variants differed in their error pattern, as the third position variant’s signal was reduced in a time dependent manner, and the PSC variant had reached its minimal error rate after 24 hours and were not affected from longer starvation. Beside the difference in reaction time to starvation, the error meters differed in the effect size of error rate reduction. Even though the first two positions had reduced their error rate only 48 hours, they expressed the largest effect with 46% and 35% reduction in error rate, respectively. On the contrary, the third position and PSC constructs had reacted to starvation after 24 hours, but gained smaller reduction of translation error rates (21% and 14% decrease from control to 48 hour starvation conditions, respectively).

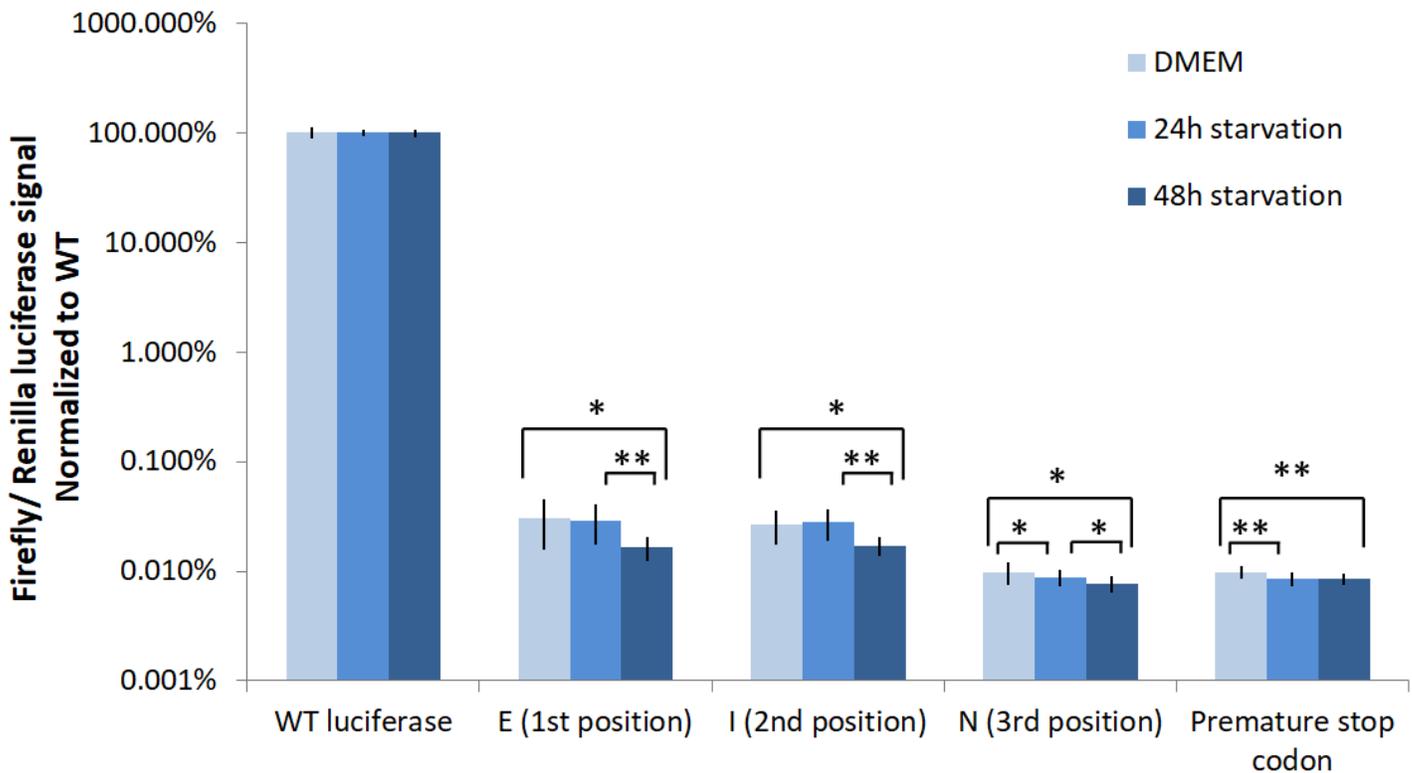


Figure 6 - Effect of media starvation on HEK293T cell. The error rates are estimated based on the firefly luciferase signal of the various mutants, normalized by constant expression of WT renilla luciferase. Presented here are the ratio between the firefly and the renilla luciferases of each mutated version of firefly luciferase normalized to WT firefly signal at the same treatment condition. The cells were treated with low glucose DMEM with 1% FCS for 24 and 48 hours before reading of the signal. The treatment had significantly reduced the error rate when compared with control cells growing in regular DMEM (high glucose, 10% FCS). Third position and PSC variants were effected even by short 24h starvation regime, while variants mutated in the first and second positions were effected only after 48h starvation. (* P value <0.05, ** P value <0.01, Mann–Whitney U test).

These results counter our expectation of observing an increase in mistranslation rates as a result of media starvation. Most of the literature regarding the effect of starvation on translation accuracy was conducted by starving bacteria and yeast to specific amino acids, and the results mostly showed misincorporation of other amino acids in place of the scarce one^{28,29}. In our assay the amino acid composition was normal, and instead the cells were exposed to carbon source and serum depletion. This starvation program might have included stress response reaction, whose effect on general translational accuracy might reduce the translation error rate. In addition, reduction of available nutrients reduces the translation speed through mTOR/S6K signaling pathways, thus leading to increased accuracy^{5,32}.

While media starvation had increased the accuracy of translation across all error meters, the speed and extent of the effect varied between mutants. Error meter for the first and second codon positions had slow and relatively large increase in accuracy, when the third position and PSC variants had quick but small accuracy gain. The behavior difference between the first and second pairs of error meters was observed in the previous experiments is maintained here, and could be influenced both by general trends of mistranslation and specific characteristics of the HEK 293 cell line.

As discussed in previous sections, mistranslation of the first and second codon positions impose greater potential damage than third position mistranslation or stop codon readthrough. The first and second codon positions were more resistant to mistranslation changes caused by starvation, but once it occurred the error rate was significantly reduced. HEK 293 have relatively low general error rate, with the first two codon positions displaying the highest error rates. Taken

together, the HEK specific error pattern and the effects of starvation might show us that in this cell line the accuracy of translating first and second positions is not optimized, and some levels of mistranslation are tolerated. On the contrary, translation error rates of third position and PSC are reduced quickly during starvation, but their basal translation accuracy is relatively high and can't be improved as much as in the other error meters.

Scanning siRNA library for translation accuracy modifiers

The second goal of this project was to identify genes that modify the translational accuracy of the cell. In order to do so we have purchased several siRNA libraries from the collection of the genomic repository unit, with the kind help of Dr. Ghil Jona. The first step of our screen included 54 siRNAs, targeting genes whose function is translation related (determined by their GO annotation). Additional 38 genes were selected semi-randomly, taking in account their expression level and making sure that the translation related and randomly chosen siRNAs will target genes with similar expression distribution. (Table 1)

Translation related genes				randomly selected genes		
AIRE	EIF2AK3	FBLN5	RPS6KA1	A4GALT	ERCC4	SDHA
DAPK1	EIF2AK4	MTRF1L	RPS6KA3	COPB2	ERH	SDHD
DAPK3	EIF2B1	RPL10	RPS6KA4	CRKL	EVC2	SPHK2
DENR	EIF2B4	RPL11	RPS6KB1	CSNK1G3	EZH1	SRC
DNAJC3	EIF2B5	RPL13A	RPS6KB2	CSNK2B	F11R	SRPK2
DOC1	EIF3S2	RPL32	RPS9	CTTN	FANCA	UBE2S
DSTN	EIF3S3	RPL37	RSU1	DGKE	FANCD2	USF2
EEF1A1	EIF3S4	RPL37A	SART1	DOK1	PTMA	ZNF148
EEF1A2	EIF3S5	RPL5	SOCS5	DUSP8	RAPGEF3	ZNF398
EEF1B2	EIF3S7	RPS16	STK35	DYRK1A	RBKS	
EEF2	EIF3S8	RPS19	ZBTB16	E2F5	RHEB	
EEF2K	EIF4B	RPS27	ZFP36L1	EPHA1	ROCK1	
EGFR	EIF4E	RPS29		EPHA4	RP2	
EIF1AY	EPRS	RPS4X		EPHA7	S100A10	

Table 1 – list of translation related and randomly selected siRNAs composing our library.

The siRNAs were co-transfected with the luciferase error meter second position variant and the normalizing renilla luciferase. In all the repeats of the experiment the firefly and renilla luciferase signals had strong and significant correlation. Based on this correlation of the whole

population of siRNAs we have calculated a trend line, predicting the expected firefly luciferase signal based on a given renilla signal. (figure 7)

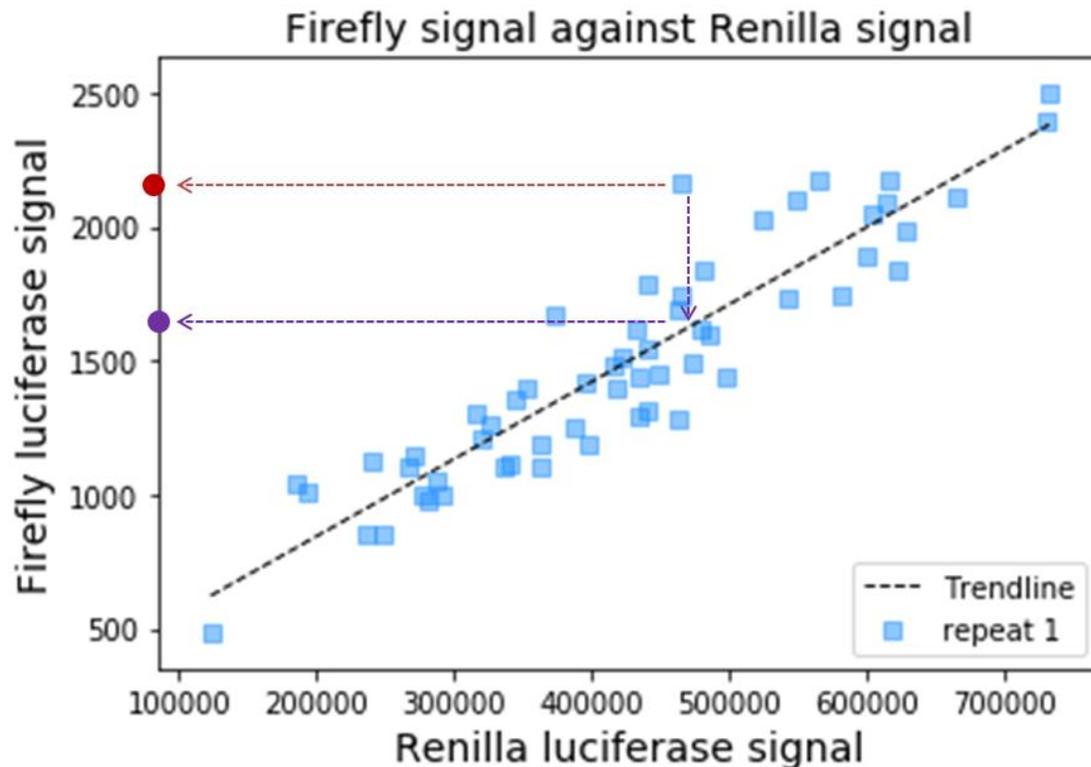


Figure 7 – firefly luciferase signal prediction. .The expected firefly signal is calculated from the regression line between measured renilla and firefly signals of cell (black line). The ratio between the measured and expected firefly signals (red and purple marks, respectively) is used as a measure of the contribution of the siRNA target gene to the accuracy of translation.

The ratio between the measured and expected firefly luciferase shows the influence of the silenced gene on translation – ratio greater than one indicates that the gene is a translation proof reader, genes who increase the error rate will have ratio smaller than one, and genes whose deletion doesn't affect the accuracy of translation will have ratio of one (figure 8).

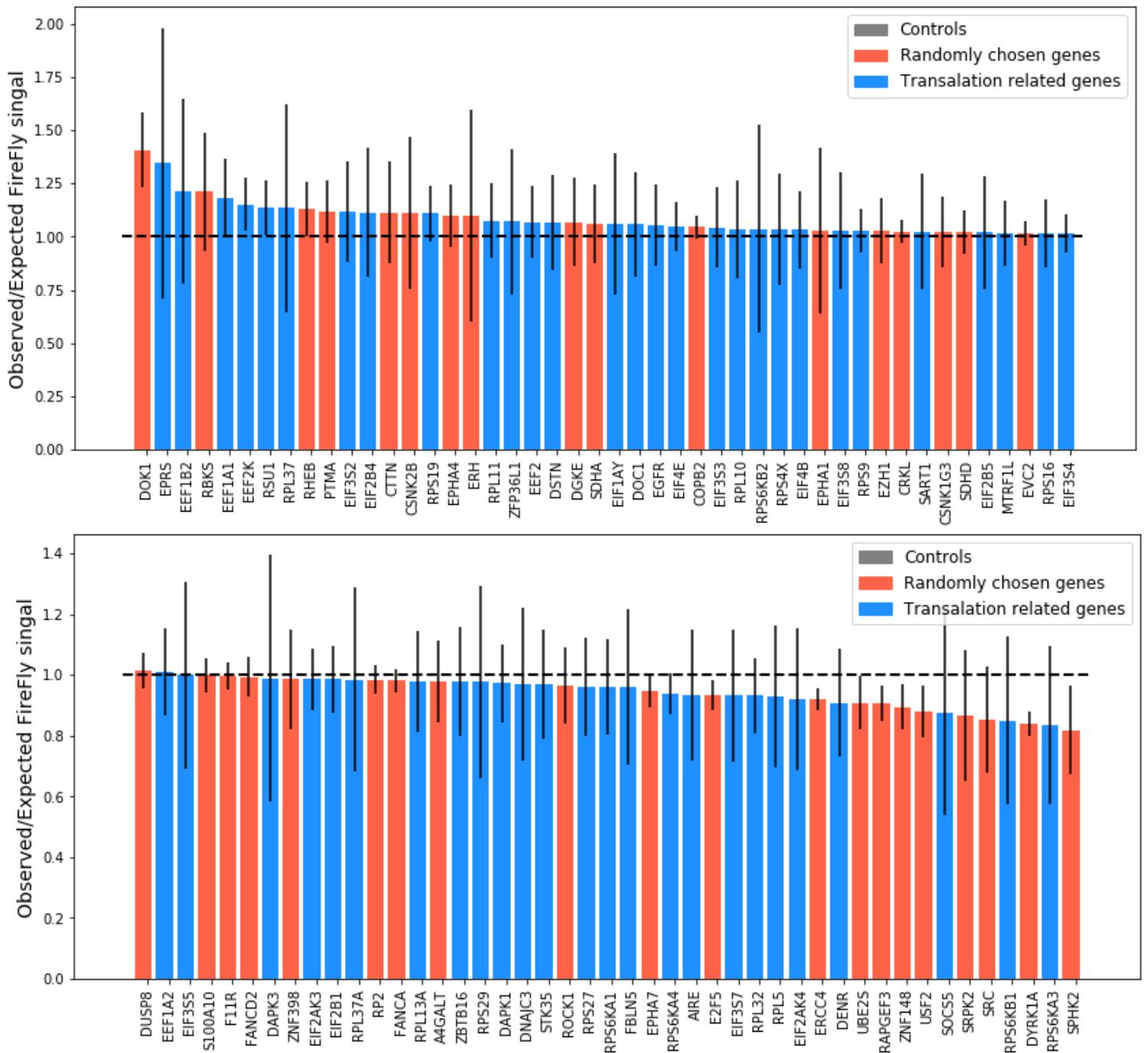


Figure 8 – Effect of gene silencing on translation accuracy. HEK 293 cells were treated with siRNA library containing translation-related genes (blue), randomly chosen genes (red). Effect of each siRNA is calculated as ratio between measured firefly signal (second position error meter) and the expected values calculated by the renilla signal and the regression line of the whole population. siRNAs whose signal is higher than one (indicated by black dashed line) increase the error rate, while siRNAs whose signal is under one reduce the error rate. Average signal and standard deviation are calculated by results of 3-6 biological repeats.

luciferase signal across repeats, divided to groups transfected with different siRNAs. Treatment with siRNA against translation related genes (determined by GO annotation, marked in blue) produced higher translation error rate than treatment with siRNA against genes randomly selected from the library (marked in red). (* P value <0.05, Mann–Whitney U test).

We have examined the annotations and functions of genes whose signal significantly altered the error rate. Largest increase in mistranslation was observed upon silencing DOK1, a cytoplasmic scaffolding protein that provides a docking platform for the assembly of signaling complexes³³. DOK1 recruits the Ras repressor RasGAP and serves as upstream negative regulator of the Ras-Erk pathway³⁴ (Figure 10). This pathway converges downstream with the mTOR-S6K pathway, activating together translation initiation and elongation factors and increasing the protein synthesis rate at the expense of accuracy levels^{5,32,35}.

The speed of translation elongation is dependent on the phosphorylation status of Threonine 56 in the elongation factor eEF2. When this position is phosphorylated by the kinase eEF2K the translation speed is relatively low, but the activity S6K and Erk pathways suppresses eEF2K and causes an acceleration of translation elongation. Similarly, silencing eEF2K and the Ras repressor RSU1 had led to an increase in translation error rate due to loss of eEF2 phosphorylation^{32,36}.

The role of Rheb (Ras homolog enriched in brain) in increasing the translation error rate is more complicated, due to its dual role in protein synthesis³⁷. The classic role of Rheb is activation of mTOR pathway by inducing a conformational change of mTORC1 that realigns the catalytic residues to their functional positions³⁸. Recently discovered function of Rheb moderates its effect on translation by inactivation of the initiation factor eIF2 α through phosphorylation by PERK³⁷, and thus reducing the translation rate in a manner that might increase the accuracy.

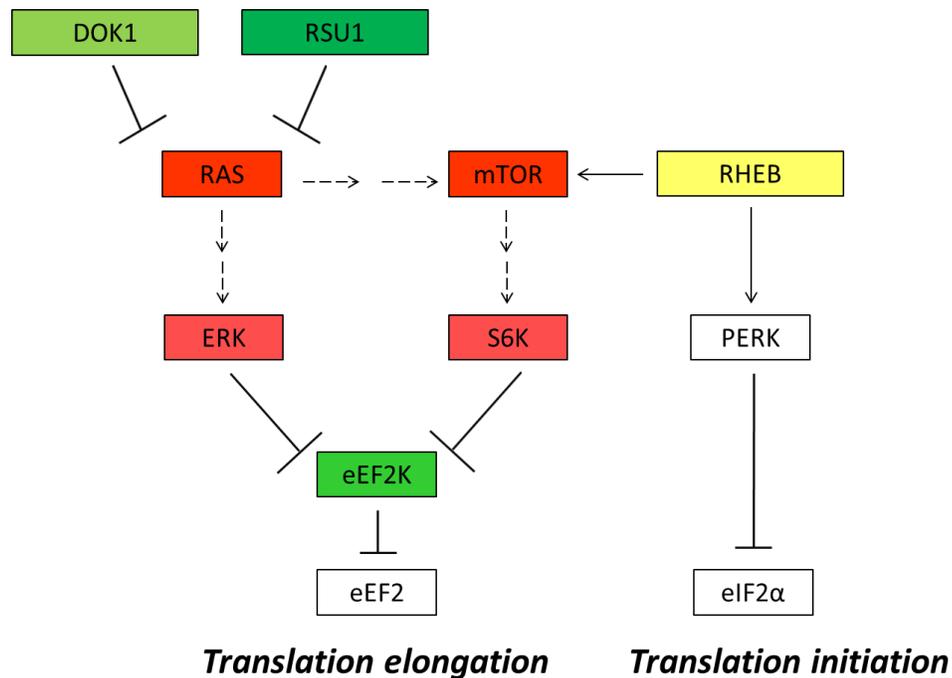


Figure 10 – regulatory connections of high-signal siRNA candidates to translation. Silencing of DOK1, RSU1 and eEF2K (green) increases the mistranslation rate by accelerating translation elongation of eEF2, through RAS-ERK and mTOR- S6K pathways (red). RHEB (yellow) plays a dual role in translation, since it increases the elongation rate through mTOR but reduces the initiation rate through PERK.

On the other end of the scale, silencing of several genes had increased the translational accuracy measured by our system. Total number of 11 siRNAs were found to reduce the error rate, and this set is consisted of 10 randomly chosen genes and the ‘non targeting’ control. This control siRNA targets firefly luciferase with different codon usage relative to our constructs, but some levels of off-target silencing might explain the reduction in measured firefly signal. No change in error rate was measured with the other negative control, “RISC free” siRNA which is chemically modified to prevent uptake and processing by RISC.

The largest effect was attributed to silencing of SPHK2, a sphingosine kinase that was found to be overexpressed and essential for survival and migration of multiple cancer types. Inhibition of SPHK2 by was found to be effective in suppression of cancer proliferation as well as inhibition of AKT/mTOR/S6K pathways,³⁹ suggesting a positive regulation role of SPHK2 upstream of these pathways. Additional examples for genes whose silencing had reduced the translation error rate appear in table 2.

Table 2

Gene name	Basic function ⁴⁰	Relation to translation error rate
DYRK1A (dual-specificity tyrosine phosphorylation-regulated kinase A1)	Phosphorylates serine/threonine and tyrosine residues. Associated with brain development disorders, down syndrome and cell cycle.	Silencing could reduce the translation error rate since members of the DYRK family activate the translation initiation factor EIF5ε ⁴¹ .
RAPGEF3 (Rap Guanine Nucleotide Exchange Factor 3)	Guanine nucleotide exchange factor for the cAMP-dependent GTPases RAP1A and RAP2A. Upon activation RAPGEF3 assembles a signaling complex which activates PI3K gamma complex.	Silencing could reduce the translation error rate since RAPGEF3 increases protein synthesis through RAS pathway and PI3K activation of mTOR ⁴² .
UBE2S (Ubiquitin Conjugating Enzyme E2 S)	Promotes progression through mitosis by inducing ubiquitination and degradation of specific anaphase promoting complex substrates.	Silencing of UBE2S was shown to reduce cell proliferation ⁴³ , which might reduce the activation of proliferation-promoting pathways as RAS-ERK and AKT-mTOR.

In-silico simulation of eGFP-based error meter

The luciferase error meter had shown to be sensitive and easy to apply system for detection of translation errors in multiple conditions and cell lines. Nevertheless, the technical requirement of cell lysis limits the application of this system for high throughput methods, such as scanning large deletion libraries in single cell manner. As a result, we have decided to design a fluorescence based translation error meter, in which the signal indicating the error rate could be measured in living cells. Our initial plan for a fluorescent error meter was point mutating eGFP in a catalytically important codon, conceptually inspired by the design of the luciferase error meter system. We have decided to create in-silico simulation of this system before we build it, in order to determine if this system will be sensitive enough to detect rare events as translation

errors. The main factor tested in this system is the ability to produce a signal within the detection limits of FACS machines (about 500 fluorescent molecules per cell).

General rate of 10^{-4} translation errors per amino acid²⁸ was determined as the basal (X1) mistranslation rate, and the system was tested on wide ranges of values relative to this rate. Since misincorporation of only specific amino acid will cause reversal of the signal, the basal probability of signal appearance by mistranslation is $(1/19)*10^{-4}$.

While translation errors are order of magnitude more abundant than transcription error, each transcription error can result in production of dozens to hundreds of protein molecules influenced by the error. We determined the basal rate of transcription error to be 10^{-5} error per base²⁸, making the reversal probability to a specific nucleotide due to transcription error to be $(1/3)*10^{-5}$. DNA mutations were not included in the simulation for the low mutation rate $\sim 10^{-8}$ mutation per base per generation, and the small number of generations per experiment (roughly one cell division per day in the fastest human cells)⁴⁴. Protein molecules are 5,000-10,000 times more abundant than mRNA molecules in yeast and mammalian cells^{45,46}, and based on these numbers we determined that a single mRNA molecule will lead to approximately 100 protein molecules. A certain level of noise in expression was inserted to the system by randomly choosing a number from a normal distribution ($\mu = 100$, $\sigma^2 = 30$).

We have used the reversal probabilities from translation and transcription to calculate the number of active GFP molecules per cell, across populations with different translation error rates. In order to do so, we had to estimate the number of GFP protein and mRNA molecules within the cells. Human cells contain $1-2*10^9$ protein molecules per cell⁴⁷, and we tested our system on different expression levels of GFP (1-10% of the proteome) at different error rates (Figure 11). We have modeled the probability to produce different levels of active GFP molecules by Poisson distribution, where κ is range of active molecules between 0 to 5000, and λ (average number of events) is calculated by multiplying the transcription error probability times the estimated mRNA copy number, plus the product of the translation error probability and the estimated protein copy number.

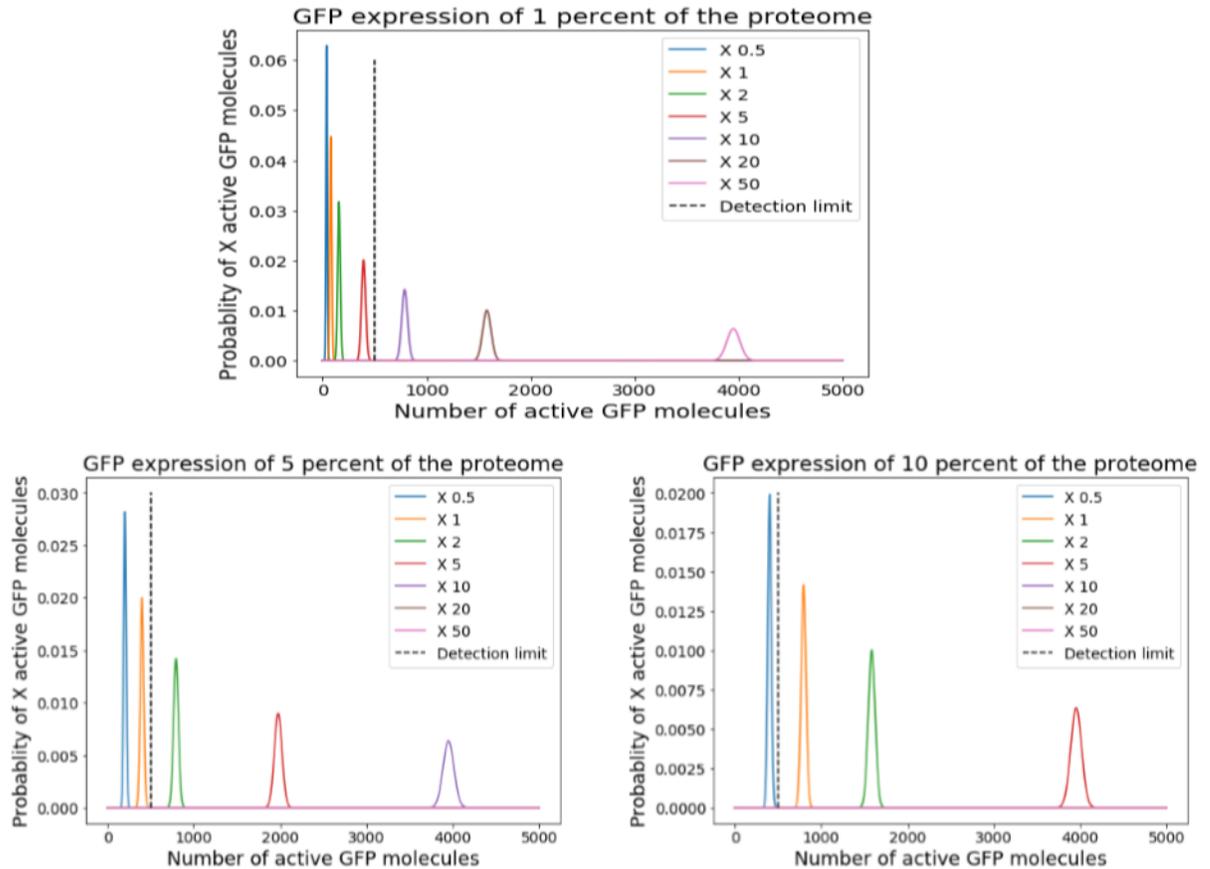


Figure 11 – In silico simulation of GFP based error meter. Number of active GFP molecules per cell was estimated based on transcription error rate and various levels expression levels and mistranslation rates. These parameters were used for calculating the average number of events λ of Poisson distribution. The minimal detection limit of FACS (black dashed line) determines whether a population with a certain error rate can be detected at a specific expression level.

The simulation's results have shown us that for low and medium expression level examined (1 and 5 percent of the proteome) the system would not be able to detect the signal of cells with basal error rate. In these expression levels only cells with respective 10 and 2 fold increases in error rate will be detected. Only when the expression level had reached 10 percent of the proteome the number of molecules was sufficient for detection of cells with basal translation error rate, and even in this case the system could identify increases in error rate, but not conditions that improve accuracy. Achieving this expression level in human cell would be technically complex and burdensome for the cells, while reliance on this expression level will reduce the stability and reliability of the system.

Design of enzyme-based fluorescent error meter

The simulation's results had shown us we would not be able to detect signals of a system in which single translation error results in single active protein molecule. Signal amplification can be achieved by designing an enzyme-based system, conceptually similar to the luciferase error meter. In these systems an enzyme is mutated in catalytically important position, and each translation error results in one active molecule which can produce signal from multiple substrate molecules. We chose to base our error meter on zipGFP⁸, a system of split GFP whose self-assembly is prevented by two peptides flanking each GFP fragment and forming a coiled coil that prevents their binding (Materials and Methods). One of the peptides in each fragment contains the cleavage site of the viral enzyme TEV protease, thus the activity of this enzyme can restore the fluorescence of many zipGFP molecules (Figure 12).

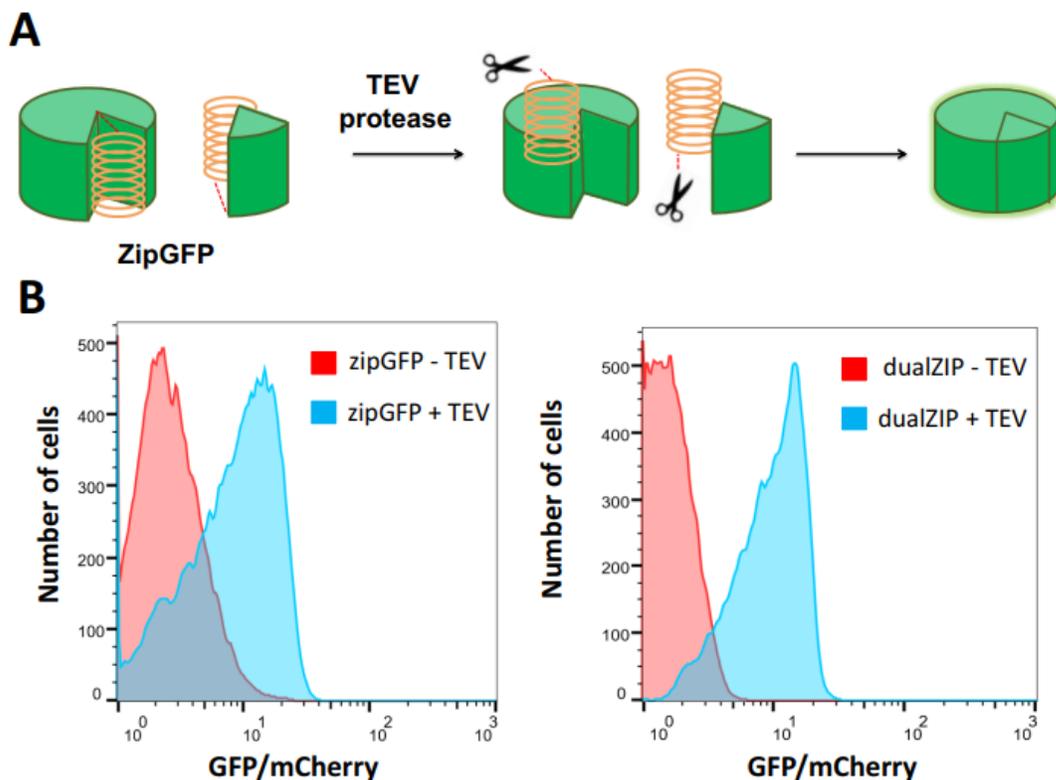


Figure 12 – enzyme activated fluorescence system based on zipGFP and TEV protease. (A) Diagram describing the activation of zipGFP by TEV protease. Two split GFP fragments (green) are “zipped” by flanking peptides forming a coiled coil (orange) which prevents the self-assembly of GFP. One of the peptides contains the target site of TEV protease (red dashed line), thus cleavage by

TEV releases the fragments from the blocking coiled coil and allows the assembly of active GFP. (B) Co-transfection of zipGFP (left figure) or the newly constructed dualZIP (right figure) with TEV protease leading to ~5 fold increase in the normalized GFP/mCherry signal relative to control cells transfected only with zipGFP (-TEV).

HEK293 cells were transfected with three plasmids expressing both zipGFP parts (zipGFP1-10 and zipGFP11) and active TEV protease. The transfection and expression efficiencies were normalized by mCherry signal, co-expressed with zipGFP1-10. Transfection with all three plasmids lead to 4-5 increase in GFP/mCherry ratio, analogous to previously published results⁸ (Figure 12).

In order to make the transfection more efficient and reduce the number of co-transfected plasmid we have cloned zipGFP11 gene into zipGFP1-10 plasmid, in a manner allowing their co-expression within a single reading frame. In addition, TEV protease was fused to tagBFP fluorophore, allowing us to control for its transfection and expression efficiency as well (Material and Methods). The newly constructed dualZIP and TEV-BFP plasmids had similar catalytic activity and signals as the original constructs, and the measured GFP signal of each cell was normalized by its mCherry signal the BFP signal of the sample.

Creation of translation error meter system based on TEV protease requires a mutant version of TEV whose activity can be rescued by misincorporation of the WT amino acid during translating of the mutated codon. Candidates for mutagenesis were codons for the catalytic triad residues His46, Asp81, and Cys151¹⁰, whose mutation in the structurally similar enzyme lead to significant reduction in activity⁹. H46R was chosen since it was shown to reduce the activity of TEV and can be represented both in near-cognate and non-cognate codons to Histidine. In addition, previous work of our lab had found increased mismatch rate of codons with Guanine in the second position to be recognized as Adenosine and decoded by tRNAs with U in their second position²⁹. This finding suggests that four-box arginine (CGN) will have higher probability to be mistranslated as histidine codons CAT/C. We chose for CGC as our near cognate codon because of its relative prevalence in the human genome⁴⁸, representing about third of four-box arginine codon appearances and two times more abundant than CGT. In addition, decoding of arginine CGC is done solely by wobble, since

the human genome does not have a tRNA gene with matching anticodon, and might experience an increased probability for mistranslation^{49,50}. Six-box arginine codon AGG was chosen for our non-cognate H46R TEV variant (Figure 13).

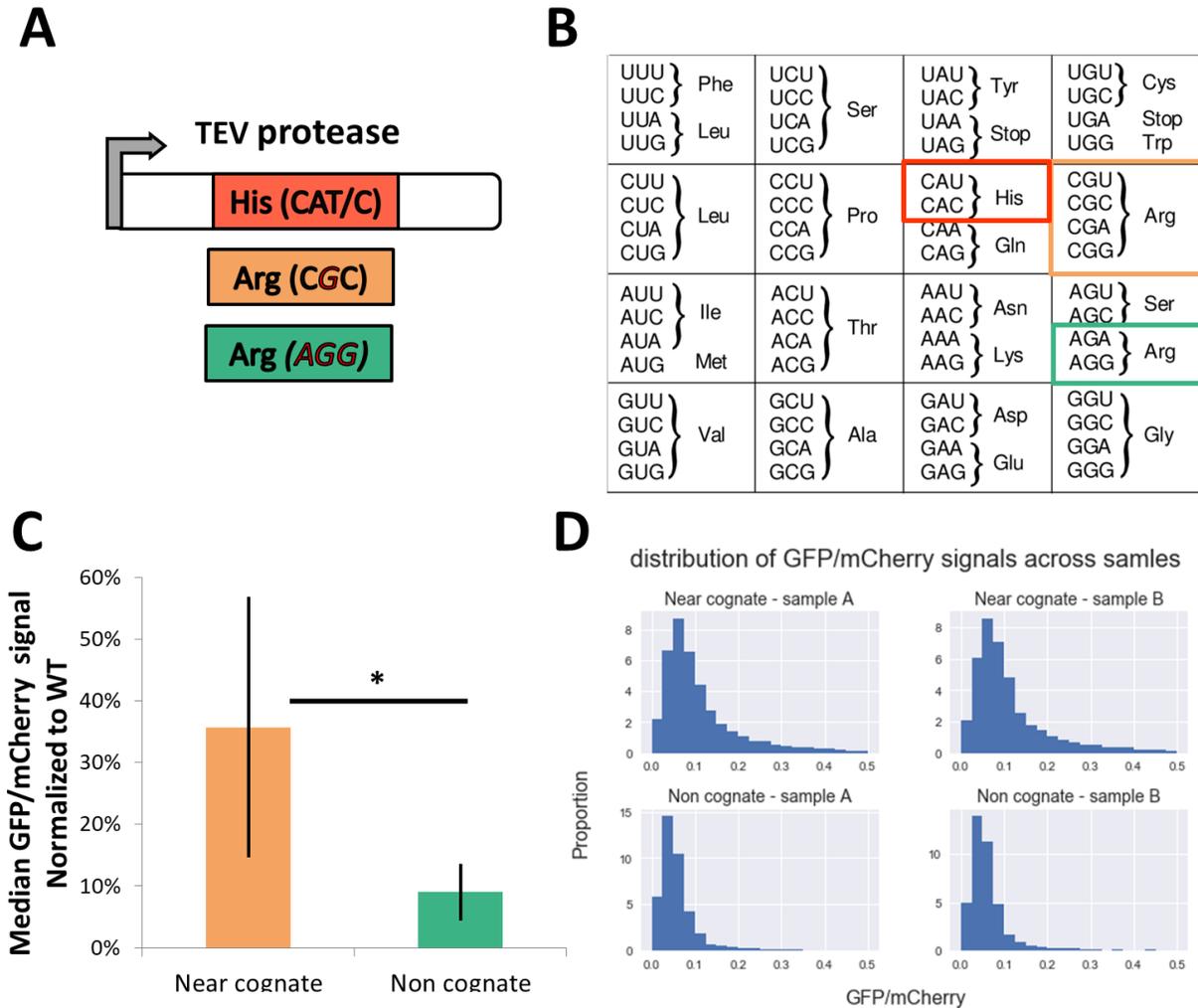


Figure 13 – measurement of translation errors by dualZIP-TEV error meter. (A) Schematic design of the TEV protease variants. A second position TEV error meter was built by point mutating the catalytically important codon histidine 46 (red) into the near cognate codon four-box arginine CGC (orange). A non-cognate variant of the H46R mutant was constructed by mutating the whole codon to the six-box arginine AGG (green). (B) A codon table with marks indicating the destination WT amino acid codons (red), the four-box near cognate arginine codons (orange) and six-box non-cognate arginine codons (green). (C) Distribution of the normalized GFP/mCherry signal of the TEV H46R error meter variants, normalized by the median BFP signal of the sample. The near-cognate construct is characterized by wider distribution of signals and higher median value of the population. (D) Average median

GFP/mCherry signal of both TEV H46R variants, as measured by several biological repeats ($n = 4$). The mistranslation error of the near-cognate codon was significantly higher than the error rate of the non-cognate construct (* P value <0.05 , Mann–Whitney U test).

HEK293T cells co-transfected with dualZIP and different TEV-BFP variants were analyzed by flow cytometer, while removing events representing with abnormal shape and size, and cells whose mCherry and tagBFP fluorescence was below the background. Analysis of the normalized GFP/mCherry signal of the remaining cells had shown that the H46R mutation had significantly reduced the activity of TEV protease, while the extent of reduction varied between mutants (Figure 14). The non-cognate mutant is characterized by relatively narrow distribution of signals, reaching an average of just $9 \pm 4.6\%$ the activity of wild type TEV protease. In contrast to these results, the signal of near-cognate mutant had significantly higher error rate and wider signal distribution, with $35.7 \pm 21\%$ of the average WT activity. The basal activity of both mutant TEV variants are supposed to be similar, since they have the same amino acid sequence. Hence we can deduce from the results that the main factor contributing to difference in activity levels is the sequential distance from any histidine codon (1 for the near cognate and 3 for the non-cognate variants).

Application of zipGFP-TEV error meter for detection of starvation-induced stress

After establishing our system and validating that the measured error rate is not amino acid dependent, we wanted to apply our fluorescent error meter for detection of translation errors under stress conditions. Since the dualZIP plasmid contains neomycin resistance gene we couldn't use many of the antibiotic drugs from the aminoglycoside group, including G418. Therefore we have decided to apply media starvation for 48 hours, a treatment regime which had reduced mistranslation rate as measured by our luciferase error meter. HEK293T cells were co-transfected with dualZIP and TEV-BFP variants, and were introduced to starvation media 24 hours post transfection. The samples were analyzed by flow cytometer in a similar manner to our previous experiments with this system, but in this case we have conducted the analysis on cell

populations expressing all transfection markers (mCherry and BFP) or just mCherry when comparing the samples to cells not transfected with any TEV variant.

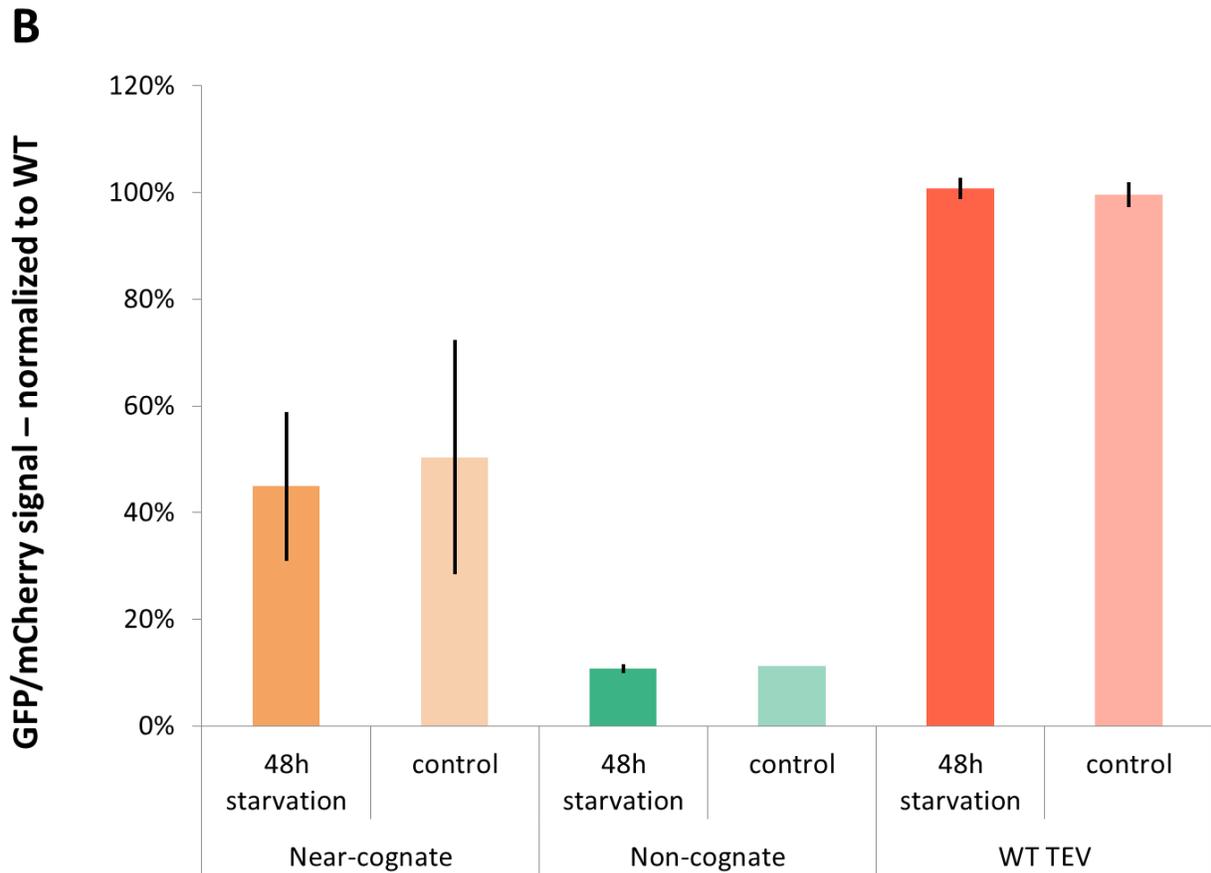
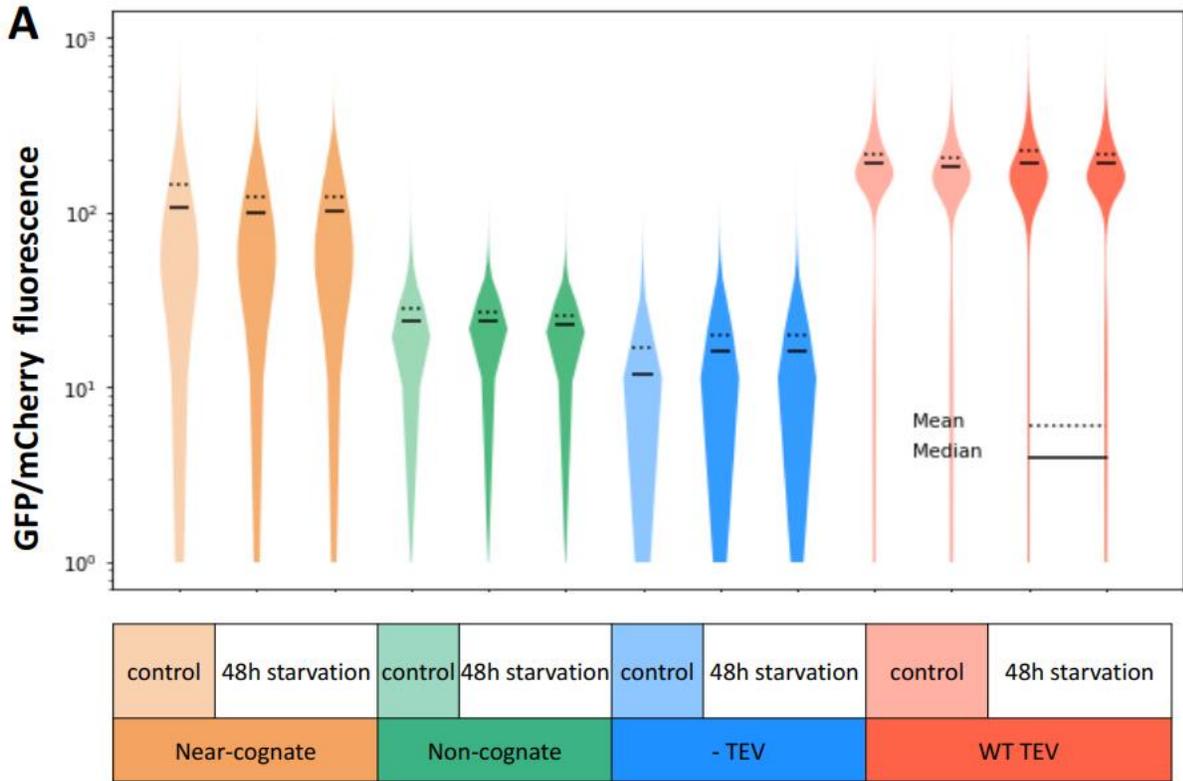


Figure 14 – Measurement of mistranslation under starvation condition by dualZIP – TEV error meter. (A) comparison of all TEV variants under starvation and control (regular media) conditions, FACS result gating of cells expressing mCherry above background. The near-cognate mutant TEV had significantly higher error rate than the non-cognate mutant, and exhibited much wider distribution of signal levels. Treatment with starvation media haven't lead to significant change in the normalized GFP fluorescence across repeats including in the background signal of dualZIP and the signal of WT TEV protease (B) Average median GFP/mCherry signal of all TEV variants normalized by the median BFP signal of the sample, as measured by several biological repeats (n = 3). Starvation treatment had reduced the variance and median of the near-cognate TEV mutant in non-significant manner (Mann–Whitney U test), and had no effect on WT and non-cognate TEV variants.

The results of this assay maintained the same trend as our previous experiments, where the TEV mutants had displayed significantly lower signals than the WT, but still higher signal than the background (-TEV cells transfected only with zipGFP, Figure). In both conditions the near-cognate TEV mutant had higher error rate with wider span of signal levels than the non-cognate mutant. Even though the effect of starvation reduced the signal and narrowed the distribution of the near-cognate construct, the effect wasn't significant ($50\pm 22\%$ and $45\pm 14\%$ for full media and starvation conditions, respectively). The error rate measured by the non-cognate mutant wasn't influenced by the starvation treatment, and stayed at 11% in both conditions. These results might suggest that the zipGFP-TEV error meter system is not sensitive enough to measure the change after 48 hours, and longer starvation regime might lead to detect a significant signal in the near cognate mutant.

Discussion

In this work we aimed to reveal external and internal factors and mechanisms which influence the accuracy of translation. We addressed this issue by employing two sets of translation “error meters” systems to measure the average accuracy of protein synthesis. These systems are based on a reporter protein whose activity is dramatically reduced by mutating functionally essential amino acid to a different amino acid. Translation error can rescue the protein's activity by misincorporating the functional amino acid at the mutated position, leading to the appearance of

a detectable signal. This signal, which represents the error rate, allows us to deduce the average translation error rate within the cell or the population.

We have used a luciferase-based error meter for measuring different types of mistranslation across multiple cell lines, drug treatments and stress conditions. These assays had revealed a difference in mistranslation trends between the first and second codon positions, relative to mistranslation of third codon position and stop codon readthrough. Mistranslation rate of the first and second codon positions was relatively similar across cell lines varying in cell type, genetic background, gene expression patterns, proliferation rate and more. Moreover, the first and second codon positions were found to be more resistant to changes imposed by external stresses as antibiotic treatment and media starvation. This difference in mistranslation pattern can be derived from diverse evolutionary pressure and potential damage caused by each type of translation error. Mistranslation of the first two codon positions has extremely high probability to alter the amino acid sequence, and in some cases can dramatically change the chemical properties and secondary structure of the protein. On the contrary, in many cases mistranslation of the third position won't cause amino acid substitution as a result of the wobble effect, especially within four-box codons. While stop codon readthrough can change the amino acid sequence, it is less likely to happen as a result of stop codon rarity in the coding sequence. In addition, the danger of sequence alteration by stop codon readthrough can be significantly reduced by insertion of additional stop codon with short distances between them⁵¹.

First we have used the luciferase system to measure different types of mistranslation across several cell lines deriving from lung fibroblasts (WI-38 fast and slow) and embryonic kidney cells (HEK 293 and 293T). This comparison had revealed elevated translation accuracy of the cancer-analogous WI-38 fast cells, relative to WI-38 slow cells representing primary cell characteristics. Generally, WI-38 derived fibroblasts have expressed higher error rate when compared with HEK derived cells. The later cell lines have been used for testing the effect of external stresses as antibiotic drug treatment and media starvation.

The ribosome-targeting aminoglycoside G418 is known to promote stop codon readthrough^{24,26}, but in our assay was additionally found to significantly increase the translation error rate at the third position of the codon. On the other hand, glucose and media starvation have reduced mistranslation across all error meter constructs in a time-dependent manner. The reduction in

error rate can be explained by adaptation of the cell to the stress conditions, and repression of mTOR pathway as a result of nutrient depletion. Downstream signals of this pathway induce activation of translation initiation and elongation and increase the synthesis speed on the behalf of accuracy^{5,32,35}. As a result, upstream inhibition of mTOR by media starvation can slow down translation and increase the accuracy of each produced protein.

Similar pattern was observed when we have scanned a siRNA library for genes whose silencing will alter the translation accuracy. Silencing negative regulators of RAS-ERK and mTOR-S6K pathways had increased their signaling, leading to downstream acceleration of translation followed by an increase of the error rate. This effect was observed as well when silencing the eEF2K, the enzyme who directly phosphorylates and thus slows down the translation elongation factor eEF2.

As we have seen, the luciferase error meter system can be applied for detection of translation errors across many cell lines and conditions, but technical characteristics of the system prevent it from being widely used in high throughput methods. Luciferase provides a readout that is based on an enzymatic reaction with its substrate Luciferin, which takes place in an extracellular medium and requires lysis of the cells expressing the reporter. Thus, translation error meter system based on luciferase would be difficult to apply for library of individual living cells that harbor a deletion of a certain gene each.

We have designed and built a fluorescence-based error meter that can measure translation accuracy in living cells. This system produces a detectible signal even at low error rate due to enzymatic amplification, where the mutated error meter is a TEV protease whose activity unlocks a fluorescent signal of inactivated split GFP system (zipGFP)⁸. We chose to mutate one of the codons belonging to the catalytic triad, whose mutation in the structurally similar enzyme lead to significant reduction in activity^{9,10}. The chosen Histidine 46 to Arginine mutation was shown to reduce the activity of TEV, and Arginine codons can be represented both in near-cognate (single codon position change) and non-cognate (three position difference) codons to Histidine.

Since both mutants encode identical amino acid sequence, they presumed to have similar background activity, with differences resulting mainly from translation errors (even though the order-of-magnitude less frequent transcription errors can rescue the signal as well). While the

non-cognate mutant had expressed constantly low signal, the signal of near-cognate mutants of significantly higher, with wide range of translation error rates even within each sample. The small variance within the non-cognate population suggests that mistranslation events are extremely rare between codons with 3 position difference, making background activity of the mutant to be the main component of the signal. On the other hand, the significantly higher and wider distribution of near-cognate signals can be attributed to combination of background activity (similar to the signal of the non-cognate variant) and mistranslation-induced activity rescue. Moreover, we have observed large variance of near-cognate signals within each sample, suggesting a certain level of stochasticity in mistranslation, even within genetically and environmentally homogeneous populations.

In conclusion, accuracy of protein synthesis varies between environmental conditions, different cell line and even individual cells within a population. However, some general trends consistently appear across all of these conditions, implicating the existence of mechanisms regulating the rate of different types of mistranslation. Our siRNA library results had shown some of these mechanisms, especially in relation to proliferation-related signaling pathways as RAS and mTOR. Application of our TEV-zipGFP error meter on much larger CRISPR deletion libraries can expand our knowledge of translation accuracy even farther, when scanning thousands of genes with a potential influence on this mechanism. Since protein synthesis is a necessary part in proliferation and survival of each living organism, we hope that our research will contribute to the accumulating knowledge regarding the accuracy and dynamics of the process.

Literature

1. Steffen, K. K. & Dillin, A. A Ribosomal Perspective on Proteostasis and Aging. *Cell Metab.* **23**, 1004–1012 (2016).
2. Morimoto, R. I. & Cuervo, A. M. Proteostasis and the Aging Proteome in Health and Disease. *Journals Gerontol. Biol. Sci. Cite J. as J Gerontol A Biol Sci Med Sci* **69**, 33–38 (2014).
3. Allan Drummond, D. & Wilke, C. O. The evolutionary consequences of erroneous protein synthesis. *Nat. Rev. Genet.* **10**, 715–724 (2009).
4. Kaushik, S. & Cuervo, A. M. Proteostasis and aging. *Nat. Med.* **21**, 1406–1415 (2015).
5. Gonskikh, Y. & Polacek, N. Alterations of the translation apparatus during aging and stress response. *Mech. Ageing Dev.* 0–1 (2016). doi:10.1016/j.mad.2017.04.003
6. Ke, Z. *et al.* Translation fidelity coevolves with longevity. *Aging Cell* 1–6 (2017). doi:10.1111/acel.12628
7. Connections between translation, transcription and replication error-rates. *Biochimie* **73**, 1517–1523 (1991).
8. To, T. *et al.* Rational Design of a GFP-Based Fluorogenic Caspase Reporter for Imaging Apoptosis In Vivo. *Cell Chem. Biol.* **23**, 875–882 (2016).
9. Hwang, D. C. *et al.* Characterization of active-site residues of the NIa protease from tobacco vein mottling virus. *Mol Cells* **10**, 505–511 (2000).
10. Phan, J. *et al.* Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**, 50564–72 (2002).

11. Milyavsky, M. *et al.* Transcriptional Programs following Genetic Alterations in *p53*, *INK4A*, and *H-Ras* Genes along Defined Stages of Malignant Transformation. *Cancer Res.* **65**, 4530–4543 (2005).
12. Lindvall, C. *et al.* Molecular characterization of human telomerase reverse transcriptase-immortalized human fibroblasts by gene expression profiling: activation of the epiregulin gene. *Cancer Res.* **63**, 1743–7 (2003).
13. Azpurua, J. *et al.* Naked mole-rat has increased translational fidelity compared with the mouse, as well as a unique 28S ribosomal RNA cleavage. *Proc. Natl. Acad. Sci.* **110**, 17350–17355 (2013).
14. Branchini, B. R., Murtiashaw, M. H., Magyar, R. A. & Anderson, S. M. The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry* **39**, 5433–5440 (2000).
15. De Crescenzo, G., Litowski, J. R., Hodges, R. S. & O'Connor-McCourt, M. D. Real-time monitoring of the interactions of two-stranded de novo designed coiled-coils: Effect of chain length on the kinetic and thermodynamic constants of binding. *Biochemistry* **42**, 1754–1763 (2003).
16. Szymczak, A. L. *et al.* Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. *Nat. Biotechnol.* **22**, 589–594 (2004).
17. Sambrook, J. & Russell, D. W. *Molecular cloning: A laboratory Manual. 3rd edn.* (Cold Spring Harbor Laboratory Press, 2001).
18. Erijman, A., Dantes, A., Bernheim, R., Shifman, J. M. & Peleg, Y. Transfer-PCR (TPCR): A highway for DNA cloning and protein engineering. *J. Struct. Biol.* **175**, 171–177 (2011).

19. McKnight, P. E. & Najab, J. Mann-Whitney U Test. in *The Corsini Encyclopedia of Psychology* 1–1 (John Wiley & Sons, Inc., 2010). doi:10.1002/9780470479216.corpsy0524
20. Myers, L. & Sirois, M. J. Spearman Correlation Coefficients, Differences between. in *Encyclopedia of Statistical Sciences* (John Wiley & Sons, Inc., 2006). doi:10.1002/0471667196.ess5050.pub2
21. Oliphant, T. E. SciPy: Open source scientific tools for Python. *Comput. Sci. Eng. Computing*, 10–20 (2007).
22. Banerjee, K., Kolomeisky, A. B. & Igoshin, O. A. Elucidating interplay of speed and accuracy in biological error correction. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5183–5188 (2017).
23. Chiusano, M. L. *et al.* Second codon positions of genes and the secondary structures of proteins. Relationships and implications for the origin of the genetic code. *Gene* **261**, 63–9 (2000).
24. Prokhorova, I. *et al.* Aminoglycoside interactions and impacts on the eukaryotic ribosome. *Proc. Natl. Acad. Sci.* 201715501 (2017). doi:10.1073/pnas.1715501114
25. Shakil, S., Khan, R., Zarrilli, R. & Khan, A. U. Aminoglycosides versus bacteria – a description of the action, resistance mechanism, and nosocomial battleground. *J. Biomed. Sci.* **15**, 5–14 (2007).
26. Heier, C. R. & DiDonato, C. J. Translational readthrough by the aminoglycoside geneticin (G418) modulates SMN stability in vitro and improves motor function in SMA mice in vivo. *Hum. Mol. Genet.* **18**, 1310–1322 (2009).
27. Harrison, E. M., Garden, O. J., Ross, J. A. & Wigmore, S. J. Firefly luciferase terminally degraded by mild heat exposure: Implications for reporter assays. *J. Immunol. Methods* **310**, 182–185 (2006).

28. Mohler, K. & Ibba, M. Translational fidelity and mistranslation in the cellular response to stress. *Nat. Microbiol.* **2**, 17117 (2017).
29. Mordret, E. *et al.* Systematic detection of amino acid substitutions in proteome reveals a mechanistic basis of ribosome errors. *bioRxiv* 255943 (2018). doi:10.1101/255943
30. Fan, Y. *et al.* Protein mistranslation protects bacteria against oxidative stress. *Nucleic Acids Res.* **43**, 1740–8 (2015).
31. Lant, J. T. *et al.* Visualizing tRNA-dependent mistranslation in human cells. *RNA Biol.* **0**, (2017).
32. Sherman, M. Y. & Qian, S.-B. Less is more: improving proteostasis by translation slow down. *Trends Biochem. Sci.* **38**, 585–591 (2013).
33. Oxley, C. L. *et al.* An integrin phosphorylation switch: the effect of beta3 integrin tail phosphorylation on Dok1 and talin binding. *J. Biol. Chem.* **283**, 5420–6 (2008).
34. Mashima, R., Hishida, Y., Tezuka, T. & Yamanashi, Y. The roles of Dok family adapters in immunoreceptor signaling. *Immunol. Rev.* **232**, 273–285 (2009).
35. Mendoza, M. C., Er, E. E. & Blenis, J. The Ras-ERK and PI3K-mTOR pathways: cross-talk and compensation. *Trends Biochem. Sci.* **36**, 320–328 (2011).
36. Dougherty, G. W., Jose, C. & Cutler, M. Lou. The Rsu-1-PINCH1-ILK complex is regulated by Ras activation in tumor cells. *Eur. J. Cell Biol.* **87**, 721–734 (2008).
37. Tyagi, R. *et al.* Rheb Inhibits Protein Synthesis by Activating the PERK-eIF2 α Signaling Cascade. *Cell Rep.* **10**, 684–693 (2015).

38. Yang, H. *et al.* Mechanisms of mTORC1 activation by RHEB and inhibition by PRAS40. *Nature* **552**, 368–373 (2017).
39. Xun, C. *et al.* Targeting sphingosine kinase 2 (SphK2) by ABC294640 inhibits colorectal cancer cell growth in vitro and in vivo. *J. Exp. Clin. Cancer Res.* **34**, 94 (2015).
40. Safran, M. *et al.* GeneCards Version 3: the human gene integrator. *Database* **2010**, baq020-baq020 (2010).
41. Woods, Y. L. *et al.* The kinase DYRK phosphorylates protein-synthesis initiation factor eIF2Bepsilon at Ser539 and the microtubule-associated protein tau at Thr212: potential role for DYRK as a glycogen synthase kinase 3-priming kinase. *Biochem. J.* **355**, 609–15 (2001).
42. Banerjee, U. & Cheng, X. Exchange protein directly activated by cAMP encoded by the mammalian *rapgef3* gene: Structure, function and therapeutics. *Gene* **570**, 157–67 (2015).
43. Garnett, M. J. *et al.* UBE2S elongates ubiquitin chains on APC/C substrates to promote mitotic exit. *Nat. Cell Biol.* **11**, 1363–1369 (2009).
44. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
45. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.* **25**, 117–124 (2007).
46. Li, J. J., Bickel, P. J. & Biggin, M. D. System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ* **2**, e270 (2014).
47. Milo, R. & Phillips, R. *Cell biology by the numbers*. (Garland Science,

2015).

48. Castro-Chavez, F. Most Used Codons per Amino Acid and per Genome in the Code of Man Compared to Other Organisms According to the Rotating Circular Genetic Code. *Neuroquantology* **9**, (2011).
49. Chan, P.P. & Lowe, T. . GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, 184–189 (2016).
50. Zhang, Z., Shah, B. & Bondarenko, P. V. G/U and Certain Wobble Position Mismatches as Possible Main Causes of Amino Acid Misincorporations. *Biochemistry* **52**, 8165–8176 (2013).
51. Johnson, L. J. *et al.* Stops making sense: translational trade-offs and stop codon reassignment. *BMC Evol. Biol.* **11**, 227 (2011).