Synthesis of Designer Proteins by Manipulating Translation

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Rajanya Bhattacharjee

geboren am 19 Juli, 1995 in Kolkata

Mainz, 2024

Dekan: [redacted]

1 Berichterstatter: Prof. Dr. Edward A. Lemke

2 Berichterstatter: [redacted]

Tag der mündlichen Prüfung: 24.01.2025

VERSICHERUNG

Name: Rajanya Bhattacharjee

Hiermit versichere ich gemäß § 12, (2) der Promotionsordnung vom 01.04.2018:
(zutreffendes ist angekreuzt.)

☐ Ich habe die heute als Dissertation vorgelegte Arbeit selbst angefertigt und ausschließlich die angegebenen Quellen und Hilfsmitte verwendet.

☐ Ich habe oder hatte die jetzt als Dissertation vorgelegte Arbeit noch an keiner anderen deutschen oder ausländischen Hochschule oder vergleichbaren Einrichtung zur Erlangung eines akademischen Grades eingereicht.

☐ Ich habe noch kein Promotions- PhD,- oder ein vergleichbares Graduierungsverfahren im Promotionsfach erfolglos beendet.

☐ Ich habe noch kein Promotions- PhD,- oder ein vergleichbares Graduierungsverfahren im Promotionsfach erfolgreich beendet.

☐ Für die Anfertigung der vorgelegten Arbeit wurde keine entgeltliche Hilfe Dritter, insbesondere eine Promotionsberatung oder – vermittlung in Anspruch genommen.

Mainz, den 04.02.25 Rajanya Bhattacharjee
(Unterschrift)

Table of Contents

Abstract

Proteins, one of the primary workhorses of a living cell, are composed of 22 different amino acids. The impressive diversity in protein structure and function that has been achieved by only 22 building blocks draws attention to the untapped potential of expanding the repertoire of canonical amino acids. Genetic code expansion (GCE) is an *in vivo* protein modification technology that allows the incorporation of non-canonical amino acids (ncAAs) in a protein of interest (POI). Most commonly GCE utilizes a special pair of aminoacyl tRNA synthetase (aaRS) and its cognate tRNA to repurpose a stop codon to incorporate an ncAA instead of signaling the termination of translation. These GCE-specific aaRS/tRNA pairs need to be orthogonal to the endogenous aaRS/tRNA pairs to minimize unspecific modification of the host proteome.

Decades of research towards perfecting the GCE technology has facilitated the incorporation of over 500 ncAAs of diverse chemical properties into POIs. These ncAAs can potentially have most desired chemical handle and can render new functionalities to the POIs, for example, post-translational modifications (PTMs), ability to bio-orthogonally react with probes including synthetic dyes, to name a few. Despite being an extremely powerful technology, GCE is not devoid of limitations. Some of the pressing challenges include lack of mRNA specificity leading to unspecific ncAA incorporation in the host proteome, limited availability of codons for reassignment and generation of truncated POI due to failure of GCE. To mitigate the first problem in mammalian cells Reinkemeier *et al*. developed membraneless orthogonally translating organelles (OTOs) to confine the GCE process. mRNA of the POI was selectively recruited to the OTOs thereby minimizing unspecific ncAA incorporation in the host proteome.

In order to truly harness the potential of GCE it is necessary to expand beyond the three available stop codons. This thesis explores the possibility of reassigning sense codons to perform mRNA specific GCE in a POI in mammalian cells. Sense codons have been reassigned both in site-specific and residue-specific manner. In site-specific sense codon reassignment, the codon to be reassigned occurs only at pre-determined locations in the gene of the POI. On the other hand, for residue-specific reassignment, a particular sense codon that naturally occurs in the gene of the POI is chosen and all occurrences of the selected codon are attempted to be reassigned. Although proteome wide residue-specific sense codon reassignment has been performed before, this technology has not been applied in a mRNA-specific manner.

A set of 8 sense codons could be successfully reassigned site-specifically in EGFP-based reporters in HEK293T cells. The highest fold change selectivity of 14fold for site-specific sense codon reassignment was achieved for the CTA codon with an OTO as compared to the non-selective, cytoplasmic GCE machinery. For residue-specific sense codon reassignment, 61 sense codons were individually reassigned in two different reporters to select the suitable codon for reassignment. mRNA selective residue-specific sense codon reassignment by OTOs was successfully demonstrated for multiple sense codons in vimentin::mcerulean in HEK293T cells. Both mRNA selective site- and residue-specific

sense codon reassignment technologies could be used for labelling POIs for confocal microscopy applications.

The success of selective residue-specific sense codon reassignment is the key step in the development of a novel protein labelling technology that would allow *in vivo* visualization of protein shapes when combined with super resolution microscopy, thereby elucidating yet invisible sub-cellular structures. Besides fluorescent microscopy applications, both site- and residue-specific sense codon reassignment hold the potential of facilitating *in vivo* synthesis of artificial biopolymers of potentially any functionality and the work of this thesis brings us one step closer to achieving greater feats than billions of years of evolution.

Zusammenfassung

Proteine, eines der wichtigsten Arbeitspferde einer lebenden Zelle, bestehen aus 22 verschiedenen Aminosäuren. Die beeindruckende Vielfalt der Proteinstruktur und Funktion, die mit nur 22 Bausteinen erreicht wurde, lenkt die Aufmerksamkeit auf das ungenutzte Potenzial der Erweiterung des Repertoires kanonischer Aminosäuren. Die „Genetic Code Expansion“ (GCE) ist eine Technologie zur In-vivo-Proteinmodifikation, die den Einbau nicht-kanonischer Aminosäuren (ncAAs) in ein Protein von Interesse (POI) ermöglicht. In der Regel wird bei GCE ein spezielles Paar aus Aminoacyl-tRNA-Synthetase (aaRS) und der zugehörigen tRNA verwendet, um ein Stoppcodon umzuwandeln und eine ncAA einzubauen, anstatt den Abbruch der Translation zu signalisieren. Diese GCE-spezifischen aaRS/tRNA-Paare müssen orthogonal zu den endogenen aaRS/tRNA-Paaren sein, um eine unspezifische Veränderung des Wirtsproteoms zu minimieren.

Die jahrzehntelange Forschung zur Perfektionierung der GCE-Technologie hat den Einbau von über 500 ncAAs mit unterschiedlichen chemischen Eigenschaften in POIs ermöglicht. Diese ncAAs können potenziell fast jede gewünschte chemische Eigenschaft haben und den POI neue Funktionen verleihen, z. B. posttranslationale Modifikationen (PTMs), die Fähigkeit, bio-orthogonal mit Sonden einschließlich synthetischer Farbstoffe zu reagieren, um nur einige zu nennen. Obwohl es sich um eine äußerst leistungsfähige Technologie handelt, ist die GCE nicht frei von Einschränkungen. Zu den dringenden Herausforderungen gehören die mangelnde mRNA-Spezifität, die zu unspezifischem Einbau von ncAA in das Wirtsproteom führt, die begrenzte Verfügbarkeit von Codons für die Neuzuordnung und die Erzeugung von verkürzten POI aufgrund des Scheiterns von GCE. Um das erste Problem in Säugetierzellen zu entschärfen, entwickelten Reinkemeier *et al*. membranlose orthogonal translatierende Organellen (OTOs), welche den GCE mRNA spezifisch machen. Die mRNA der POI wurde selektiv zu den OTOs rekrutiert, wodurch der unspezifische Einbau von ncAA in das Wirtsproteom minimiert wurde.

Um das Potenzial von GCE wirklich nutzen zu können, ist es notwendig, über die drei verfügbaren Stoppcodons hinauszugehen. In dieser Arbeit wird die Möglichkeit untersucht, bereits codierende Codons neu zuzuordnen, um mRNA-spezifische GCE in einem POI in Säugetierzellen durchzuführen. Codierende Codons wurden sowohl Positionsspezifisch als auch Aminosäurenspezifisch neu zugewiesen. Bei der Positionsspezifischen Neuzuweisung von bereits codierenden Codons kommt das neu zuzuweisende Codon nur an vorher festgelegten Stellen im Gen des POI vor. Bei der Aminosäurenspezifischen Neuzuweisung hingegen wird ein bestimmtes Codon, das natürlicherweise im Gen des POI vorkommt, ausgewählt, und es wird versucht, alle Codons derselben Art neu zuzuweisen. Obwohl eine Proteom-weite, Aminosäurenspezifische Neuzuweisung bereits durchgeführt wurde, ist diese Technologie nicht mRNA spezifisch.

Eine Reihe von 8 bereits codierenden Codons konnte erfolgreich Positionsspezifisch in EGFP-basierten Reportern in HEK293T-Zellen neu zugewiesen werden. Die höchste Selektivität von 14 fach für die Positionsspezifische Neuzuweisung des CTA-Codons mit einer OTO im Vergleich zu der nicht-selektiven, zytoplasmatischen GCE-Maschinerie

erreicht. Für die Aminosäurenspezifische Neuzuweisung wurden alle 61 codierenden Codons in zwei verschiedenen Reportern einzeln neu zugewiesen, um das geeignete Codon für die Neuzuweisung auszuwählen. Die selektive Aminosäurenspezifische Neuzuweisung durch OTOs wurde erfolgreich für mehrere Codons in dem Reporter Vimentin::mcerulean in HEK293T-Zellen nachgewiesen. Sowohl mRNA-selektive Positionsspezifische als auch Codonspezifische Technologien zur Neuzuweisung von Codons konnten für die Markierung von POIs für konfokale Mikroskopieanwendungen verwendet werden.

Der Erfolg der selektiven, Aminosäurenspezifisch Neuzuordnung von Codons ist der entscheidende Schritt bei der Entwicklung einer neuartigen Technologie zur Proteinmarkierung, die in Kombination mit hochauflösender Mikroskopie die In-vivo-Visualisierung von Proteinformen und damit die Aufklärung bisher unsichtbarer subzellulärer Strukturen ermöglichen würde. Neben Anwendungen in der Fluoreszenzmikroskopie haben sowohl die Positions als auch Aminosäurenspezifische Neuzweisung von Codons das Potenzial, die In-vivo-Synthese künstlicher Biopolymere mit potenziell beliebiger Funktionalität zu erleichtern, und die in dieser Arbeit geleistete Arbeit bringt uns einen Schritt näher an das Erreichen größerer Ziele, als die Errungenschaft von Milliarden Jahren der Evolution.

Acknowledgements

List of Abbreviations

3IF	3-iodo-phenylalanine
5MTP	5-methoxy-L-tryptophan
AAV2	adeno-associated virus
AA	PylRS from *Methanosarcina mazei* harbouring the mutations N346A and C348A
AcK	acetylated lysine
ACPK	N^{ε}-((((1*R*,2*R*)-2-azidocyclopentyl)oxy)carbonyl)-l-lysine
ADC	antibody-drug conjugates
ADAR	adenosine deaminase acting on RNA
AF	PylRS from *Methanosarcina mazei* harbouring the mutations Y306A and Y384F
AHA	azidohomoalanine
APC	antibody-polymer conjugates
BCNK	bicyclo[6.1.0]non-4-ynyl lysine carbamate
BOCK	t-butyloxycarbonyl-l-lysine
BONCAT	bio-orthogonal non-canonical amino acid tagging
CAGE	conjugative assembly genome engineering
Cas proteins	CRISPR associated proteins
CbzK	(2S)-2-amino-6-{[(benzyloxy)carbonyl]amino}hexanoic acid
C. elegans	*Caenorhabditis elegans*
CuAAC	copper catalyzed alkyl-azide cycloaadition
CRISPR	clustered Regularly Interspaced Short Palindromic Repeats
Cy5-H-Tetrazine	3-(p-Benzylamino)-1,2,4,5-tetrazine - Cy5
DBCO	dibenzocyclooctyne
Dh	*Desulfitobacterium hafniense*
DHFR	dihydrofolate reductase
DiZPK	3-(3-methyl-3H-diazirine-3- yl)-propaminocarbonyl-Nε -L-lysine
DiZSeK	{Ne-3-[(3-methyl-3H-diazirine-3-yl)]-propaminocarbonyl-g-seleno-l-lysine}
eBK	ε-tBoc-lysine
EGFP	enhanced GFP
Ec / *E.coli*	*Escherichia coli*
EPR	electron paramagnetic resonance
Evolink	evolution and link
F	full length FUS
ΔF	truncated FUS or FUS without the RNA binding domain [FUS(1-267)]
FACS	fluorescence activated cell sorting
FADS	fetal akinesia deformation sequence
FKBP	FK506 binding protein
FP	fluorescent protein
FUNCAT	fluorescent non-canonical amino-acid tagging
G1	*Methanogenic archaeon ISO4-G1*
GCE	Genetic Code Expansion
GFP	green fluorescent protein
GlnRS	glutaminyl synthetase

H5	*Methanogenic archaeon ISO4-H5*
HOCouA	7-hydroxycoumarine alanine
HPG	homopropargylglycine
IEDDA	inverse-electron demand Diels-Alder reaction
iPASS	as Identification of Permissive Amber Sites for Suppression
IR	infrared
LCK	LCK tyrosine kinase
LeuRS	leucyl-RS/ leucyl synthetase
Ma	*Methanomethylophilus alvus*
MAGE	Multiplex automated genome engineering
Mb	*Methanosarcina barkeri*
MCP	MS2 coat protein
Mm / *M. mazei*	*Methanosarcina mazei*
Mj */ M. jannaschii*	*Methanocaldococcus jannaschhii*
MINFLUX	microscopy with minimum photon fluxes
Ml	*Methanomassiliicoccus luminyensis*
Mt	*Methanonatronarchaeum termitum*
ncAA	non-canonical amino acid
NES	nuclear export signal
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NPC	nuclear pore complex
NUP	nucleoporin
ON	overnight
OMeY	O-methyl-L-tyrosine
OTO	orthogonally translating organelle
PACE	phage assisted continuous evolution
PAGE	polyacrylamide gel electrophoresis
pAcF	p-acetyl-L-phenylalanine
pAzF	p-azido-L-phenylalanine
PBD	pyrrolobenzodiazepine
PBS	Phosphate-buffered saline
pBpa	p-benzoyl-L-phenylalanine
pCNPhe	p-cyanophenylalanine
PCR	polymerase chain reaction
PEI	polyethylenimine
pFF	p-fluorophenylalanine
PheRS	phenylalanyl synthetase
pIF	p-iodo-L-phenylalanine
pNCSF	para-isothiocyanate phenylalanine
POI	protein of interest
pThr	phosphothreonine
PTM	post-translational modification
pTyr	phosphotyrosine
Pumby	pumilio-based assembly
PumHD	pumilio homology domains
PYP	photoactive yellow protein
PylRS	pyrrolysyl synthetase

Pyl-tRNA	pyrrolysyl tRNA
RBD	RNA binding domain
REPAIR	RNA editing for programmable A to I replacement
RESCUE	RNA editing for specific C to U exchange
RT	room temperature
S.cerevisiae	*Sachharomyces cerevisiae*
SCOK	cyclooctyne-lysine
SORT	stochastic orthogonal recoding of translation
SPAAC	strain-promoted alkyl-azide cycloaddition
sTyr	sulfotyrosine
SRM	super-resolution microscopy
SRNS	steroid-resistant nephrotic syndrome
aaRS	synthetase/ aminoacyl tRNA synthetase
TCO*AK	trans-cyclooct-2-en-L lysine
TetBu	3-(6-butyl-1,2,4,5-tetrazin-3-yl)-phenylalanine
TyrRS	tyrosyl synthetase
Tfm-Phe	tri-fluoromethylphenylalanine
TMP	trimethoprime
TMSiPhe	trimethylsilyltyrosine
VADER	virus-assisted directed evolution of tRNA

List of Figures and Tables

Figures

Supplementary Figures

Appendix Figures

Tables

1 Introduction

Expanding the amino acid repertoire beyond the 22 canonical amino acids holds the key to improving upon billions of years of evolution and opening a plethora of scientific possibilities. Genetic code expansion (GCE) is a technology that facilitates synthesis of designer proteins with non-canonical amino acids (ncAAs) of desired functionality.[1] Some of the early instances of GCE include *in vitro* incorporation of ncAAs in proteins of interest (POIs) by artificially charged tRNAs.[2] The Schultz lab in 1989 developed GCE to achieve *in vitro* synthesis of larger proteins like β-lactamase.[3] Following decades of research GCE has made great strides towards advancement resulting in today's sophisticated cellular machinery capable of hijacking the host translation system to enable *in vivo* synthesis of proteins modified with ncAAs.[4]

At present more than 500 distinct ncAAs have been incorporated in different POIs rendering them with myriads of new functionalities.[1] Depending on the chemical handles, these ncAAs can enable POIs to bio-orthogonally react with compatible probes including fluorescent dyes, introduce post-translational modifications (PTMs), incorporate heavy atoms to allow protein structure elucidation by X-ray crystallography and NMR spectroscopy, to name a few.[5] Particularly noteworthy is the contribution of GCE in significantly reducing the size of fluorescent protein labels to a single synthetic dye molecule.[6] Combined with advanced microscopy techniques GCE holds the promise of elucidating yet invisible subcellular structures.

The central theme of this thesis is protein engineering by addressing one of the pressing challenges of GCE (figure 1). A key intended future application for the work in this thesis is to develop a novel protein labelling strategy to visualize the shape of proteins *in vivo* with super-resolution microscopy. Hence in the following subsections first a brief overview of current fluorescent labelling strategies will be discussed to highlight the advantage of GCE enabled incorporation of fluorescent probes over other available fluorescent tags (section 1.1). The GCE technology itself will be detailed in substantial depth in sections 1.3 and 1.4. Section 1.3.2 will cover applications of GCE in various other fields of life sciences to underscore the immense potential of this technology. Despite the progress, GCE still suffers from certain limitations. The limitations of GCE will be discussed in section 1.4 along with the advancements made so far to mitigate them. Finally, the key motivations and aims of this thesis will be summarized in section 1.5.

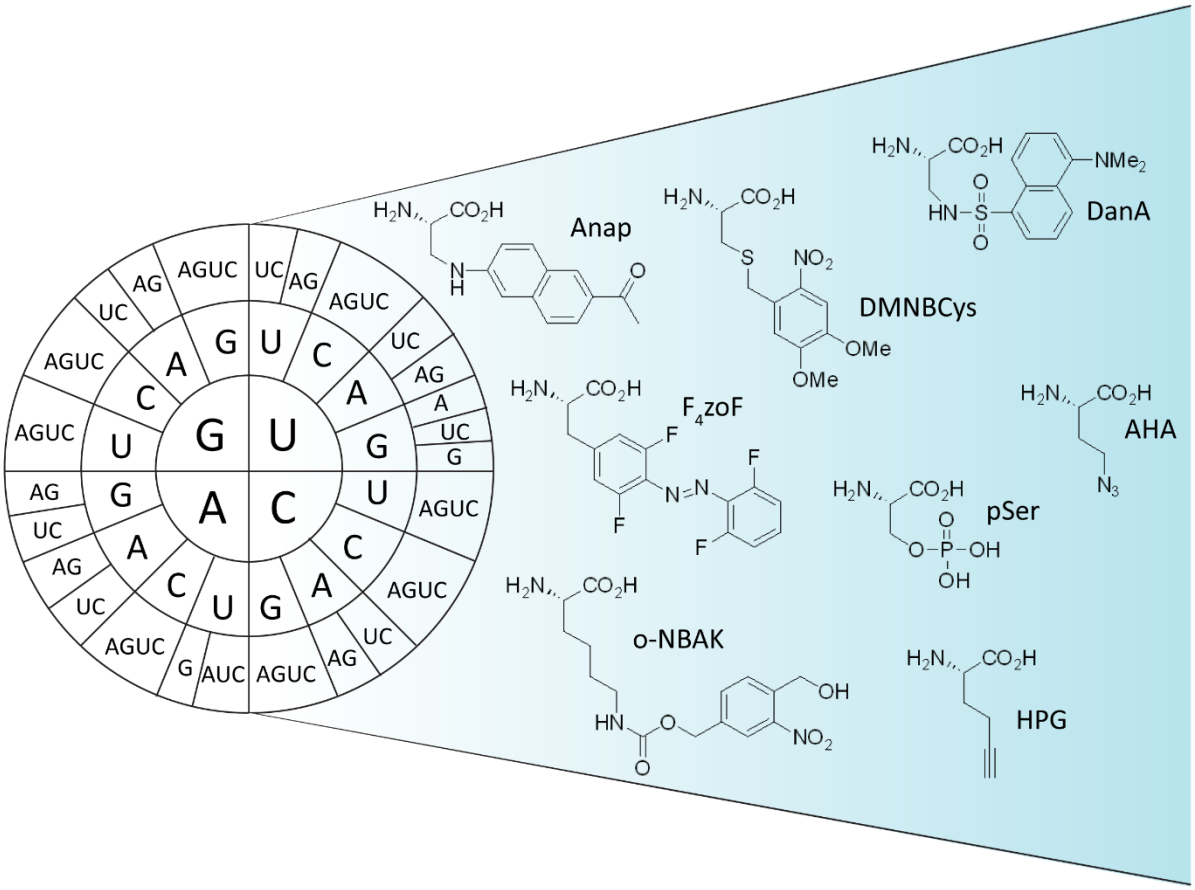


Figure 1| Potentially any sense codon can be reassigned to incorporate any ncAAs in a POI by GCE. The central theme of this thesis is to manipulate translation to efficiently repurpose sense codons to incorporate desired ncAAs in a POI.

1.1 Fluorescent protein labelling techniques currently in use

Fluorescence microscopy is an indispensable tool for investigating cellular phenomena allowing us a glimpse into subcellular localization of individual proteins of interest, providing insights regarding individual protein functionality and molecular mechanisms underlying diseases. Fluorescence refers to the phenomenon by virtue of which certain molecules can absorb light of a specific wavelength and emit light of a longer wavelength. Binding of such fluorescent probes to biomolecules enables visualization and monitoring of the labelled biomolecules by diverse microscopy approaches on subsequent excitation by light of relevant wavelength. For decades the aim has always been to resolve the minute details of cellular components and today, with the groundbreaking innovation of super-resolution microscopy, it is possible to achieve 1-3nm resolution in three dimensions with MINFLUX technology.[7]

However, besides technical advancements, the choice of fluorophore also plays a pivotal role in resolving subcellular components. The scientific advancements brought about by fluorescent microscopy and the pertinent need to achieve the highest possible resolution has inspired years of research into improving fluorescent probes and labelling techniques. The ideal probe needs to be as bright as possible at the same time being

minimally invasive to preserve the native properties of the POI. In the following subsections various protein labelling strategies currently in use have been discussed.

1.1.1 Immuno-fluorescence

To date one of the most common and popular techniques for labelling proteins is by antibody conjugated synthetic dyes.[8] In case of direct immunostaining, the fluorophore is conjugated to the primary antibody itself whereas for indirect labelling the dye is conjugated to the secondary antibody. This method is not suitable for live cell imaging. For immunostaining cells need to be fixed and subsequently permeabilized to allow the entry of the sufficiently large antibodies. Nevertheless, this method has played a pivotal role in answering fundamental biological questions and allowed visualization of subcellular structures.

1.1.2 Fluorescent proteins

Discovery of the green fluorescent protein (GFP) revolutionized fluorescence microscopy in live cells and was awarded a Nobel prize in 2008.[9,10] This fluorescent protein (FP) can simply be expressed *in vivo* as a fusion construct with any POI to investigate protein activity, subcellular localization, gene expression, perform colocalization studies, among others.[8] The success of GFP has inspired robust research into the development of a range of FPs with improved spectroscopic properties and diverse emission spectra. Some examples include mPlum (emission- 649nm), mCherry (emission – 610nm), tdTomato (emission – 581nm) , mCitrine (emission – 529nm), EYFP (emission – 527nm), mEmerald (emission – 509nm), mCerulean (emission – 475/503nm) etc.[11]

Currently there exists a diverse palette of FPs with properties suitable for varying needs. In selecting a particular fluorescent protein, there are a few parameters that need to be considered. Brightness is one of the most important characteristics of a FP and is defined as the product of the molar absorption coeffiecient at the excitation wavelength and quantum yield.[12] To ensure adequate expression of functionally active FPs in mammalian cell-based applications, it is imperative that the protein can efficiently fold under respective cell growth conditions. Under prolonged light exposure, FPs are likely to get photobleached. Hence photostability becomes a relevant parameter depending on the nature of the experiment. Recently Hirano *et al*. engineered StayGold, a green light emitter, that boasts of more than 10fold higher photostability as compared to contemporary FPs.[13] In the case of multi-color imaging experiments, it is prudent to not select FPs with overlapping spectra to avoid cross-talk between the different probes.

Photoactivable or photoswitchable FPs mark a noteworthy advancement in FP engineering efforts. Fluorescent properties of such FPs can either be activated or altered on exposure to light of certain wavelengths.[14,15] A single mutation, T203H, in GFP produces PA-GFP which can be activated by UV-light.[16] Examples of photoswitchable FPs that undergo light-induced spectral alterations include PS-CFP, pAmRFP1, Kaede.[17–19] Dronpa, on the other hand, is an FP that undergoes a conversion from fluorescent green-form to a non-fluorescent form on exposure to blue light.[20] These FPs find frequent applications in pulse-chase experiments, protein-protein interaction studies as well as in super-resolution microscopy.

FPs, despite their indispensability, are not ideal probes. These are proteins of about 27kDa and their fusion to a POI may alter the native characteristics of the POI.

1.1.3 Probes derived from protein-binding ligands and small molecules

FKBP' is an engineered variant of the FK506 binding protein (FKBP) that selectively binds to a synthetic ligand with a much higher affinity as compared to the wild type FKBP.[21] This allows fusion constructs of FKBP' with any POI to be selectively labelled by fluorescent derivatives of the synthetic ligand.[22,23] Other examples include fluorescent derivatives of trimethoprime (TMP). TMP is an inhibitor of the *E. coli* dihydrofolate reductase (DHFR). Since it has a much higher specificity towards *E. coli* DHFR than the mammalian counterparts, the TMP- *E. coli* DHFR system can be used for *in vivo* imaging applications similar to the FKBP' system.[24] FlAsH-EDT_2 (4,5-bis(1,3,2-dithiarsolan-2-yl)fluorescein) on the other hand is a nonfluorescent biarsenical dye that can bind to POIs containing a CCPGCC motif. On undergoing reaction with the tetracysteine peptide, FlAsH-EDT_2 gets converted to its fluorescent form thereby labelling the POI [25]

1.1.4 Self-labelling protein domains and peptide tags

Some of the widely used examples in this category include SNAP, CLIP and Halo tags, amongst others. SNAP tag is a 20kDa enzyme, O^6-alkylguanine DNA alkyltransferase (AGT), that can utilize O^6-alkylguanine, O^6- benzylguanine and their derivatives as substrates. It can be fused to a POI and subsequently labelled with a fluorescent, cell permeable derivative of O^6-benzylguanine via a cysteine residue, thereby facilitating *in vivo* labelling of the fusion construct.[26] CLIP tag is a variant of the AGT protein and works on a similar principle. However, it uses O^6-benylcytosine as substrate.[27] Halo tag (33kDa), on the other hand, is derived from a bacterial hydrolase, haloalkane dehalogenase and irreversibly binds to chloroalkane based synthetic ligands.[28] Fluorescent dyes designed with such a chloroalkane linker can effectively label POIs fused with a Halo tag. Photoactive yellow protein (PYP) is another example of a self-labelling protein domain of bacterial origin. It is only 14kDa in size and can form a thioester bond with 7-hydroxycoumarin-3-carboxylic acid, 4-hydroxycinnaminic acid as well as their derivatives.[29] Besides entire protein domains, short peptide tags can also offer effective tools for protein labelling. Sunbul *et al.* engineered the reactive self-labelling tag (ReacTR) from a TexasRed fluorophore-binding peptide (TR512) precursor.[30] The disadvantage of the ReacTR is that it is limited to only the TexasRed dye. The dC10α tag is another such peptide tag containing two cysteine residues that can react with corresponding maleimide moieties of a fluorescent dye.[31] In order to prevent unspecific binding of the maleimide containing fluorophore to thiol groups other than those on the dC10α tag, methoxy substituents could be added to the maleimides. These substituents were shown to not affect the interaction of the maleimides to the reactive cysteine residues of the dC10α tag, thereby reducing background labelling.[31]

1.1.5 Labelling via metal chelating ligands

Peptide tags like histidine and tetra-aspartate tags are known to interact with metal chelating ligands namely Nickel bound nitrilotriacetic acid (NTA) and zinc based DpaTyr

respectively. POIs fused with histidine or tetra-aspartate tags can respectively be labelled with NTA[32] or DpaTyr [33] bound probes.

1.1.6 Enzyme mediated labelling

Certain enzymes can be used to catalyze the binding of labelled substrates to a POI. These enzymes recognize a peptide sequence or a small protein, which when fused to the POI can facilitate the subsequent labelling step. Sortase A is an example of such an enzyme. It can catalyze peptide bond formation between the threonine of a LPXTG motif and an oligoglycine sequence.[34,35] Sortase A has been used for cell surface labelling as well as for facilitating *in vivo* fusion of GFP with POIs in mammalian cells.[35,36] Other examples include phosphopantetheinyl-transferases (PPTases), biotin ligase and engineered lipoic acid ligase (LplA) variants.[19,37] PPTases catalyze binding of CoA bearing probes to POIs fused with a carrier protein that PPTases can recognize.[38] The bacterial PPTases do not cross-react with the mammalian carrier proteins thereby minimizing background labelling. In order to avoid interfering with cellular processes involving CoA, this method is only used for labelling surface proteins.[19] Similarly, biotin ligases site-specifically add biotin to certain peptide motifs that can subsequently be labelled with streptavidin fused probes.[39,40] The peptide motifs recognized by bacterial biotin ligases often differ from recognition motifs of mammalian biotin ligases.[19] Hence biotin ligases of bacterial origin can be used for *in vivo* labelling in mammalian cells with minimal background. Engineered LplAs on the other hand add short chain fatty acid derivatives to specific peptide sequences fused with the POI.[37,41]

1.1.7 Non-canonical amino acid based labelling

Most of the methods listed above involve fusion of proteins (at least about 14 kDa) or peptide tags to a POI. With the remarkable strides made in microscopy techniques, especially in super-resolution microscopy, it is possible to resolve objects 1-3nm apart.[7] However, fusion of POIs with fluorescent proteins or peptide tags can adversely affect the resolution. The precise visualization of cellular structures by advanced microscopy needs fluorescent probes to be as small as possible. ncAA – based labelling is the ideal solution in this case since the size of the probe can be drastically reduced to a single amino acid or a single dye molecule.[42,43] Additionally, the probe can be introduced at any desired site on the POI. These ncAAs can be incorporated into target proteins *in vivo* by GCE, which forms the central theme of this thesis and will be discussed in substantial detail in sections 1.3 and 1.4. ncAAs can either themselves be fluorescent or specifically bind to fluorescent dyes via bio-orthogonal reactions. Both cases are discussed in section 1.3.2.3.

1.3 Genetic Code Expansion (GCE)

GCE is an *in vivo* protein modification technology that facilitates the incorporation of ncAAs of potentially any desired functionality into a POI.[44] Most commonly this is achieved by an orthogonal pair of aminoacyl-tRNA synthetase (aaRS)/tRNA capable of decoding stop codons and repurposing them to introduce ncAAs in a POI (figure 2). To minimize cross-talk with the host translation system it is imperative for the GCE-specific aaRS/tRNA to not recognize endogenous aaRS/tRNA pairs. In this respect aaRS/tRNA

pairs isolated from the archaea, *Methanosarcina mazei*, are considered excellent candidates for GCE since they are orthogonal in both prokaryotic and eukaryotic systems. Derivatives of the pyrrolysyl synthetase (PylRS) from *M. mazei* (Mm) comprise some of the widely used GCE-specific aaRS/tRNA pairs and boast an impressive range of specificity towards several ncAAs[45,46]. Other such pairs include engineered variants of PylRS/Pyl-tRNA from *Methanosarcina barkeri* (Mb) and *Methanomethylophilus alvus* (Ma),[46] tyrosyl synthetase (TyrRS)/Tyr-tRNA from *Methanocaldococcus jannaschhii* (Mj), leucyl synthetase (leuRS)/tRNA, glutaminyl synthetase (GlnRS)/tRNA, TyrRS/tRNA from *Escherichia coli* (Ec) and phenylalanyl synthetase (PheRS)/tRNA, TyrRS/tRNA from yeast, to name a few.[1,47] Extensive evolution and rational design of various aaRS/tRNA pairs have made it possible to genetically encode over 500 distinct ncAAs in POIs.[1] Some of the efforts dedicated towards engineering GCE specific aaRS and tRNAs will be discussed in the following section 1.3.1.

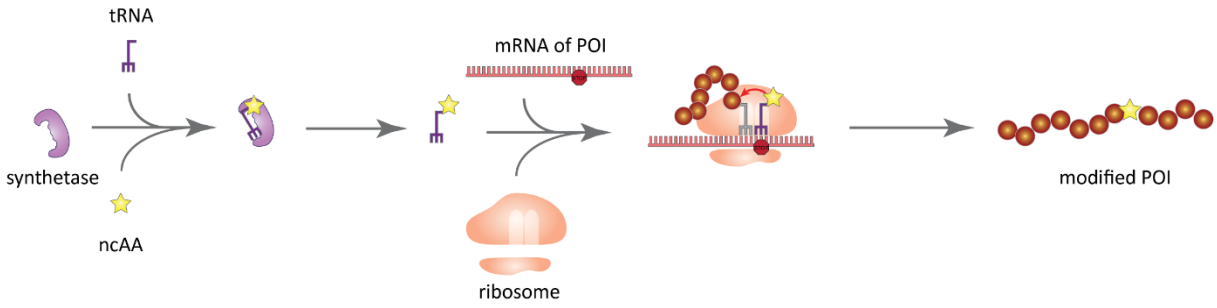


Figure 2| Genetic Code Expansion. GCE utilizes synthetase/tRNA pairs that are capable of decoding stop codons and repurpose these stop codons to incorporate ncAAs into POIs.

1.3.1 Engineering of GCE machinery

aaRS and tRNA engineering constitute a significant part of the GCE optimization efforts. One of the primary goals of aaRS engineering is to expand the substrate specificity to allow incorporation of the ever-growing list of new ncAAs. Besides the substrate binding pocket, editing domains of aaRSs as well as domains involved in interaction with the cognate tRNA are also relevant targets for engineering approaches. On the other hand, structural optimization of tRNAs is mainly driven by the motivation of achieving increased expression levels in the host system. Both aaRS and tRNAs are also evolved to generate more orthogonal pairs for simultaneous incorporation of distinct ncAAs.

1.3.1.1 Engineering of aaRS

Directed evolution is an iterative selection process that is frequently used for developing modified aaRS.[48] In this method pre-selected single or multiple residues of an aaRS are mutated to generate a library of variants containing every possible combination of 20 canonical amino acids in these positions. Members of this plasmid library are expressed in a suitable host system and subjected to repeated cycles of positive and negative selection pipelines to select for certain properties.[49,50] *E. coli* is one of the widely used platforms for directed evolution of GCE components. For expanding the substrate specificity of an aaRS, the positive selection tests are typically carried out in presence of the desired ncAA and the negative ones in the absence of the ncAA. Positive selections often include the expression of a selective marker containing the amber codon, for

example antibiotic resistance genes.[51] The host organisms expressing aaRS variants that can successfully decode the stop codon can survive in growth media containing the respective antibiotic. Negative selections are needed to eliminate promiscuous aaRS variants that can accept canonical amino acids. The selective markers for such tests usually consist of a stop codon containing gene coding for a product lethal to the host organism, for example barnase in case of *E. coli*.[44] In absence of the desired ncAA, any aaRS variant that can incorporate a canonical amino acid at the stop codon will synthesize the toxic product leading to cell death. Many of the aaRS variants used in eukaryotes are evolved in bacterial systems. However optimal performances of such aaRS in eukaryotes can be hindered due to the higher level of cellular complexity of eukaryotes as compared to the prokaryotes. Hence effort has been directed toward developing directed evolution platforms based on eukaryotic hosts. *Saccharomyces cerevisiae* was used to generate a polyspecific variant of EcTyrRS capable of accepting p-acetyl-L-phenylalanine (pAcF), p-benzoyl-L-phenylalanine (pBpa), p-azido-L-phenylalanine (pAzF), O-methyl-L-tyrosine (OMeY), and p-iodo-L-phenylalanine (pIF).[52] A high-throughput evolution platform was developed by Stieglitz *et al.* in *S. cerevisiae* to generate a range of aaRS variants having expanded ncAA specificity as well as highly selective variants with the ability to distinguish between structurally similar ncAAs.[53] In this case positive selection was based on the expression of a stop codon harboring gene encoding a fluorescent protein. Cells expressing the fluorescent protein in presence of the desired ncAAs were subsequently collected by fluorescence activated cell sorting (FACS).

Phage display is another elegant method for selection of evolved aaRS variants.[50,54] In this case a gene coding for the coat protein of bacteriophage is fused at its N terminal to an antibody binding epitope tag gene containing a stop codon. Upon transformation of the phagemid containing the aaRS variants and the cognate tRNA in *E. coli*, synthesis of the coat protein and subsequent propagation of the bacteriophage is only possible on successful stop codon suppression by the aaRS variants. These epitope presenting bacteriophages can be subsequently purified by antibody interaction along with the packaged aaRS variant gene. Another phage-based evolution strategy is phage assisted continuous evolution (PACE).[55] In this case non-infectious bacteriophages lacking the gene for protein III are utilized. Protein III is necessary for bacteriophage infection of *E. coli* and subsequent propagation.[56] Hence *E. coli* is supplied with the protein III gene containing a stop codon along with aaRS variants and their cognate tRNAs. Effective aaRS mutants can produce propagating bacteriophages by successful suppression of the stop codon in the protein III gene and the infectious bacteriophages are subsequently collected.

Besides directed evolution, rational designing strategies are also effective for developing improved aaRS variants. One such aaRS is EcLeuRS with a mutation at site 265.[57] The basis of this mutation is to improve the interaction of the EcLeuRS with its cognate tRNA. To enable stop codon decoding, a G to C mutation is necessary at position 34 in the anticodon loop of the EcLeu-tRNA. However, G34 is recognized by Asp265 of the EcLeuRS, thereby leading to the hypothesis that the G34C mutation creates a physical

gap in the interaction of EcLeuRS/tRNA and adversely affects the recognition. To compensate for this gap, site 265 of EcLeuRS was mutated with amino acids having longer side chains than aspartic acid. Of the five chosen alternatives Asp265Arg resulted in the highest yield of GCE modified fluorescent reporter with 186% increase in fluorescence intensity as compared to the non-mutated EcLeuRS.

Other instances of aaRS engineering include the addition of a nuclear export signal to the MmPylRS by Nikic *et al*.[58] Upon *in vivo* click-labelling of GCE modified POI with fluorescent dyes, they observed bright unspecific nuclear signal. A search for the cause of such background labelling led to further investigation of the PylRS structure which revealed the unexpected presence of a nuclear localization signal (NLS). The cause of the unspecific signal was due to click-reaction of fluorescent dyes to ncAA charged tRNAs bound to the nuclear localized PylRS molecules. Addition of a stronger nuclear export signal (NES) allowed enrichment of the PylRS into the cytoplasm and improved accessibility to the translation machinery.

A recent development in aaRS engineering involves tRNA display technology. The elegance of this method lies in the fact that it facilitates selection of aaRSs capable of accepting monomers that are poor ribosomal substrates.[59] The drawback of evolution strategies based on translation readout is that the new aaRS substrates need to be accepted by the ribosome. Hence extending aaRS substrate specificity beyond α-L-amino acids is challenging and ribosomes could not be evolved to accept non-canonical monomers (substrates other than α-L-amino acids or structurally similar monomers). With the tRNA display method the Chin lab was able to incorporate β-amino acids and α,α-disubstituted amino acids into a POI thereby expanding the class of monomers that can be utilized for GCE.[59]

1.3.1.2 Engineering of tRNA

A significant part of tRNA engineering efforts is directed towards improving tRNA expression levels in the host system, especially in the case of eukaryotic hosts. Quite often archaeal and bacterial tRNAs and their derivatives are utilized for GCE in eukaryotes. However, most tRNAs of archaeal and prokaryotic origin lack the consensus sequences A- and B box elements necessary for tRNA expression in eukaryotic hosts by RNA polymerase III.[60–62] Some of the common strategies include addition of relevant promoter elements to the tRNA genes or expression of the tRNAs under appropriate promoters for example U6, H1 and 7SK in mammalian cells[63] and RPR1 and SNR52 in yeast.[64] It is important to note that the simple addition of promoter elements, that is the A- and B box sequences is not always sufficient to ensure adequate tRNA expression. When such an approach failed for MbPyl-tRNA in *S. cerevisiae*, Hancock *et al*. made use of an unconventional, dicistronic gene of *S. cerevisiae*.[65] The arginine-$tRNA_{ucu}$ and asparagine-$tRNA_{GUC}$ in *S. cerevisiae* is transcribed as a single precursor tRNA and subsequently processed to yield two different tRNAs. Exchanging the asparagine-$tRNA_{GUC}$ gene with the MbPyl-tRNA gene allowed sufficient expression of the latter.

Similar to aaRS engineering, directed evolution and rational design of the tRNA structure has resulted in a number of variants with enhanced stop codon suppression capability.

The Pyl-tRNA isolated from *M. mazei* is one of the widely used GCE enabling tRNAs. Hence structural optimization of the Pyl-tRNA plays an important role in GCE enhancement. The archeal Pyl-tRNA has a secondary structure that differs distinctly from the mammalian counterparts, except for some similarities to the mitochondrial serine tRNA. The design principles implemented by Serfling *et al.* are based on the hypothesis that this structural divergence adversely affects interaction of the Pyl-tRNA with mammalian translation machinery.[66] The best performing tRNA variants from Serfling *et al.* were the result of addition of PylRS recognition motifs to *Bos taurus* mitochondrial serine $tRNA_{CUA}$ with subsequent inclusion of the A-box motif in the D arm and various mutations, especially between the acceptor arm and the T-arm. With the improved tRNA variants M15 and C15, a 2– 5fold increase in intracellular expression level of the tRNAs was observed as compared to the native Pyl-tRNA.

Another noteworthy advancement in tRNA engineering is the development of the virus-assisted directed evolution of tRNA (VADER) platform for directed evolution of tRNAs in a mammalian system.[67] An adeno-associated virus (AAV2) delivery system is utilized to ensure each cell is supplied with a single tRNA variant. The cells are subsequently transfected with MbPylRS gene as well as genes necessary for viral propagation. Viral reproduction was coupled to successful GCE, hence viral propagation was observed only in presence of the desired ncAA. In order to eliminate non-orthogonal tRNA variants that can interact with endogenous synthetases, ncAA with azide group was incorporated in the viral capsid. Virus particles with azide containing capsid proteins could be selectively isolated by the strain-promoted alkyl-azide cycloaddition (SPAAC) reaction between the azido group and dibenzocyclooctyne (DBCO).

1.3.2 Applications of GCE

The immense potential of the GCE technology can truly be realized from the indelible mark it has left in diverse fields of life sciences, biotechnology and medicine. The continuously growing palette of ncAAs boasts of a wide range of designer chemical handles that add new functionalities to a POI. With tailor-made ncAAs it is possible to add critical PTMs to recombinant proteins expressed in bacterial host, optically control cellular processes, regulate gene expression, minimally label POIs with synthetic dyes, to name a few. In this section a closer look will be taken at some of the defining applications of GCE.

1.3.2.1 Addition of post-translational modifications in POIs by GCE

Post-translational addition of chemical groups to proteins play an important role in fine-tuning cellular processes in eukaryotes[68,69] and perturbation of the PTM landscape has been implicated in a number of diseases, thus highlighting the importance of understanding the precise role of such modifications.[70] GCE offers an elegant method to study diverse PTMs by facilitating site-specific incorporation of ncAAs with desired PTMs or mimetic groups in any POI. Considering phosphorylation is one of the most abundant PTMs, much effort has been directed towards expression of site-specifically phosphorylated proteins by GCE. Certain methanogenic archaea possess an aaRS known as SepRS that charge $tRNA_{CYS}$ with O-phosphoserine (Sep). The Sep is subsequently converted to cysteine by the enzyme SepCysS. The Söll group utilized the SepRS isolated

from *Methanococcus maripaludis* along with an engineered variant of the $MjtRNA_{CYS}$ to facilitate Sep incorporation in *E. coli* in response to an amber codon.[71] Owing to the reversibility of phosphorylation, often it is necessary to knock out endogenous phosphoserine phosphatases to ensure retention of the phosphate group. Additionally engineering of the prokaryotic elongation factor EF Tu also plays an important role in mediating optimal interaction of the Sep charged tRNA and the ribosome. Similar strategies have resulted in the incorporation of phosphotyrosine (pTyr) by an evolved MjTyrRS in *E. coli*.[72] However, in this case manipulation of the host transport machinery was necessary to allow adequate bio-availability of pTyr. Besides Sep and pTyr, phosphothreonine (pThr) incorporation by GCE has also been reported.[73]

Along with phosphorylation, other relevant PTMs introduced by GCE include acetylation, sulfonation, nitration and methylation. Acetylation is known to influence DNA replication and repair pathways, chromatin structure, subcellular protein localization, among others.[74–76] Hence *in vivo* encoding of acetylated lysine has facilitated several important studies in both prokaryotes and eukaryotes. Histone acetylation is a significant phenomenon influencing chromatin structure and transcription. Understanding the effect of histone acetylation on cellular processes can not only answer fundamental questions but also unearth molecular mechanisms underlying diseases like cancer.[77] Elsässer *et al*. encoded acetylated lysine in six pre-determined sites in histone H3 in mammalian cells to study the effect on gene expression.[78]

Tyrosine sulfonation on protein surfaces enhances protein-protein interactions. This PTM is commonly observed in secreted and membrane proteins and plays an important role in signaling pathways, viral entry of cells like HIV infection, to name a few.[79,80] E51, an antibody against the HIV coat protein gp120, has five sulfonated tyrosine residues in its V_HCDR3 loop which is crucial for interaction with gp120.[81–83] To obtain a better understanding of the role of individual sulfonated tyrosine residue in E51-gp120 binding, Li *et al*. generated E51 antibodies incorporated with sulfotyrosine (sTyr) by GCE.[84] GCE was chosen as the preferred method since it yielded a homogenous population of E51 sulfoforms thereby simplifying, to some extent, investigation of the sulfonation landscape.

Post-translational nitration occurs due to oxidative stress[85,86] and its consequence can be studied by site-specific incorporation of nitrated tyrosine derivatives into POIs.[87] Methylation on the other hand is challenging to introduce by GCE. Methylated lysine, for example, is structurally very similar to lysine and hence GCE-specific synthetases cannot distinguish between the two.[1] An alternative approach would be to use lysine analogs with caged methionine groups like Boc-protected N-methyl lysine.[88]

1.3.2.2 Optical control of cellular processes by GCE

Photo-sensitive ncAAs are often used to control *in vivo* protein functionality. Photocaged ncAAs form a subset of this category and contain a protective chemical group that shields the active moiety. On exposure to light of specific wavelength these ncAAs can be activated by degrading the shielding group.[89] The use of photo-caged ncAAs offer a non-invasive method for controlling subcellular protein localization, signaling pathways and

gene expression. Lemke *et al*. demonstrated that nuclear trafficking of a transcription factor Pho4 in *S.cerevisiae* could be optically controlled by incorporating a photocaged serine derivative 4,5-dimethoxy-2-nitrobenzylserine (DMNB-Ser) at specific locations in Pho4. DMNB-Ser prevented phosphorylation of serine which was necessary for nuclear export of Pho4 by Msn5. On exposure to blue light the caging group of DMNB-Ser could be degraded thereby facilitating phosphorylation of serine and subsequent nuclear export. [90] Gautier *et al*. showed that introduction of a caged lysine in the NLS prevented the import of NLS fused EGFP into the nucleus. On exposure to UV light a 4fold increase in the nuclear-to-cytoplasmic ratio of EGFP signal was observed.[91] Similarly, it is possible to control the activity of various polymerases, recombinases or nucleases by incorporation of photo-caged ncAAs. Examples include incorporation of caged tyrosines in Cre recombinase active site for inducible DNA recombination,[92] in a zinc-finger nuclease system (ZFN) to control DNA cleavage[93] and introduction of photo-caged lysine and tyrosine in critical domains of T7RNA polymerase for regulating gene expression,[94] among others.

Caging groups are indeed useful for controlling cellular pathways, however the irreversible decaging process poses a limitation for the usage of photo-caged ncAAs as on-off switches. In this case ncAAs designed with photoisomerizable side-chains can be an alternative since they undergo reversible chemical changes on exposure to light. ncAAs with an azobenzene moiety and a thiol reactive group, when incorporated into a POI, can react with a neighbouring cysteine group, thereby forming a photoisomerizable azobenzene bridge. These photo-switchable ncAAs find widespread application in *in vivo* control of protein conformational changes and thereby study of protein function. The thiol reactive groups have been optimized to eliminate the need of harmful UV light exposure to induce reaction with the cysteine residues in the POI. Benzyl-chloride is an example of such a thiol-reactive moiety that can react with a neighboring cysteine by proximity-enabled nucleophilic substitution. Efforts have also been made towards developing photo-switchable ncAAs that can be controlled by visible light instead of toxic UV light. Hoppmann *et al*. developed a derivative of the azobenzene containing ncAAs with a fluorinated phenyl ring that could undergo reversible changes on exposure to green light. Photo-switchable ncAAs have been incorporated into proteins in both *E. coli* and mammalian cells and have been shown to induce structural changes in proteins like calmodulin.[95]

Yet another class of light-sensitive ncAAs is the photo-crosslinking ncAAs. Once incorporated into a POI, such ncAAs, on exposure to light of specific wavelength, can undergo reaction with proximal biomolecules. Hence photo-crosslinking ncAAs are an asset for studying *in vivo* protein-protein interactions. pBpa, pAzF, 3-(3-methyl-3H-diazirine-3- yl)-propaminocarbonyl-Nε -L-lysine (DiZPK), N^{ε}-((((1*R*,2*R*)-2-azidocyclopentyl)oxy)carbonyl)-l-lysine (ACPK) are some examples of photo-crosslinking ncAAs.[96–98] Lin *et al*. reported the use of DiZPK and ACPK for unearthing pathogenic and defense mechanisms of enteropathogenic bacteria like *Shigella*.[98] pAzF has been used in elucidating the activation mechanism of G-protein-coupled corticotropin

receptor (CRF1R) in mammalian cells as well as for elucidating the binding site of peptide exendin-4 on human glucagon-like peptide-1 receptor (GLP-1R).[99]

1.3.2.3 Study of protein functionality by GCE

The diverse chemistry of the >500 available ncAAs allow site-specific incorporation of a myriad of probes for investigating protein structure and function by nuclear magnetic resonance (NMR) spectroscopy, infrared (IR) spectroscopy, electron paramagnetic resonance (EPR), X-ray crystallography and fluorescence microscopy.[1] Jackson *et al*. reported the utility of Tri-fluoromethylphenylalanine (tfm-Phe) to study confirmational changes of nitroreductase and histidinol dehydrogenase.[100] Similarly, trimethylsilyltyrosine (TMSiPhe) has been used for probing structural information of phospho-β2 adrenergic receptor/β-arrestin-1(β-arr1) by NMR spectroscopy.[101] Schultz *et al*. used ncAA p-cyanophenylalanine (pCNPhe) as an IR spectrocopy probe to gain insights into structural changes induced in myoglobin by ligand binding.[102] For EPR measurements ncAAs with spin labels are of great value as demonstrated by Schmidt *et al*. with a nitroxide containing lysine derivative.[103] Introduction of heavy atoms into POI by GCE finds application in study of protein structures by X-ray crystallography. The heavy atoms play an important role for phase determination in X-ray crystallography. The incorporation of p-iodophenylalanine into T4 lysozyme and its subsequent crystallization was reported by Xie *et al*.[104] GCE has contributed significantly towards minimizing the size of fluorescent probes and brought it down to a single amino acid. Fluorescent ncAAs can be used for FRET measurements to study protein conformational changes, to track labelled POI as well as for monitoring subcellular localization of POIs. Many fluorescent ncAAs have been designed with emission wavelengths ranging from 450nm to about 550nm. Some examples include 7-hydroxycoumarine alanine (HOCouA) and its derivatives, L-3-(6- acetylnaphthalen-2-ylamino)-2-aminopropionic acid (Anap) and dansylalanine (DanA).[105–107]

An alternative strategy for introducing diverse probes into POI includes the application of bio-orthogonal reactions between ncAAs and the corresponding probes with compatible chemical handles.[1,108] The beauty of these bio-orthogonal reactions is that even in the crowded cellular environment, the reactions occur specifically between the desired components. A wide range of bio-orthogonal reactions exist including copper catalyzed alkyl-azide cycloaadition (CuAAC),[109] SPAAC,[110] inverse-electron demand Diels-Alder reaction (IEDDA),[111] Staudinger ligation,[112] ketone/hydroxylamine condensation[1], nitrile-aminothiol condensation,[113] nucleophilic substitution reaction[1], Palladium catalyzed reaction [114] and photo-click cycloaddition.[115] Of the above listed reactions, IEDDA is one of the most frequently implemented and is also the most relevant for the work done in this thesis. IEDDA occurs between strained alkenes/ alkynes and tetrazine moieties. The advantages of using IEDDA are it is generally very fast, does not produce any toxic by-products and unlike CuAAC, no additional components are necessary. The ease of introduction of a diverse range of probes into POIs by IEDDA has inspired synthesis of numerous ncAAs with compatible chemical handles including trans-cyclooctene, bicyclononyne, cyclopropene and norbornene derivatives. Some of the commonly used ones include trans-cyclooctene lysine (2'-

TCOK), bicyclo[6.1.0]non-4-ynyl lysine carbamate (BCNK), cyclooctyne-lysine (SCOK).[111,116,117]

1.3.2.4 Therapeutic applications of GCE

GCE technology has contributed significantly to the development of new therapeutic strategies against a number of diseases including cancer. Antibody-drug conjugates, bi-specific antibodies, antibody-antisense conjugates, generation of GCE-dependent virus for development of vaccines are some of the notable examples.[118] Antibody-drug conjugates (ADCs) constitute a powerful therapy against cancer. The antibody is responsible for targeting the small molecule or drug to the cancerous tissues thereby reducing toxic side effects of the drug to healthy tissues. Incorporation of ncAAs with reactive moieties into desired antibodies allow binding of drug molecules with compatible chemical handles by bio-orthogonal reactions.[119] Unlike electrophilic modification of cysteine or lysine residues, ADCs generated by click chemistry between ncAA and desired drug molecule yield a homogenous product.[120] One such example is the modification of anti-Her2/neu antibody with the ncAA *N*6-((2-azidoethoxy)carbonyl)-l-lysine. Click reaction mediated by the azide group enabled conjugation of auristatin F or a pyrrolobenzodiazepine (PBD) with the anti-Her2/neu antibody, generating ADCs that showed promising results in mouse model.[121] siRNA based therapeutic approaches have also benefited from GCE. One of the major challenges in this case is the selective delivery of the siRNA to the target cell. Antibody-polymer conjugates (APCs) formed by reaction between a site-specifically incorporated ncAA into the antibody and a cationic polymer can achieve tissue-specific targeting of siRNA. The cationic polymer binds to the siRNA and the antibody ensures its delivery to the target cell. Example of APCs generated by GCE include *p*AcF incorporated anti-Her2 antibodies combined with an aminooxy-tethered cationic polymer.[122] Selective gene silencing by siRNA has substantial potential as countermeasures against diseases like cancer, neurodegenerative as well as infectious diseases. Bispecific antibodies comprise yet another instance of GCE enabled therapeutics. These antibodies, as the name suggests, can recognize two targets. Anti Her2/ anti CD3 is an example of such an antibody. pAcF was incorporated site-specifically into both anti Her2 and anti CD3, subsequently these two antibodies were coupled by click reaction with a linker containing two reactive groups.[123]

Protein based therapeutics are often conjugated to polyethylene glycol (PEG) to have better pharmacokinetics and biocompatibility.[124] The usual mechanism of PEG conjugation yields heterogenous products, hence ncAA mediated bio-orthogonal reaction to functionalized PEG is preferred. Besides PEGylation, certain ncAAs, for example p-fluorophenylalanine (pFF), para-isothiocyanate phenylalanine (pNCSF) among others, can themselves enhance protein stability.[125,126]

GCE is a useful method for generating live attenuated virus for vaccine development. Typically, a stop codon is inserted into the genes of an essential viral protein which can only be decoded by GCE machinery in the presence of certain ncAAs, thereby coupling replication of the virus to GCE. Such attenuated virus was developed for influenza A and HIV with commendable results.[127–129]

A recent development in anti-cancer therapy includes engineering of chimeric antigen receptor T cells (CAR-T). CARs are modified to allow T-cells to specifically recognize antigens on the surface of cancer cells. GCE can be used for the development of switchable CAR-T cells. T-cells with engineered CAR capable of recognizing fluorescein isothiocyanate (FITC) is one such example.[130] ncAA incorporated antibodies could bio-orthogonally react with FITC molecules which were subsequently recognized by the CAR-T cells. This method could be used to conjugate any antibody to FITC thereby altering specificity of the CAR-T cells with minimal effort. Also, since the conjugation process was dependent on ncAA availability, this could be used as a switch for controlling CAR-T cell activity. Another instance of GCE enabled cell-based therapy is ncAA triggered therapeutic switch (NATS) against diabetes.[131] NATS consists of a mammalian cell containing GCE-specific aaRS/tRNA and insulin gene modified with a stop codon. These cells could be implanted into mice models and on feeding the mice with ncAA containing cookies the insulin production could be induced.

1.4 Limitations of Genetic Code Expansion & the State-of-the-Art

Undoubtedly, the technology of GCE has opened a plethora of scientific possibilities, however it is still quite far from being perfect. Some of the pertinent issues with classical GCE include its lack of mRNA specificity, limited availability of codons for reassignment and generation of truncated POIs due to premature termination of translation at the stop codons meant to be reassigned.[132] Laborious research has resulted in several elegant strategies to mitigate these challenges, however more success has been achieved in prokaryotes than in the eukaryotes. Owing to the higher complexity of cellular organization, a direct transfer of optimized technologies, in many cases, is not possible from prokaryotic to eukaryotic systems. This section is dedicated to discussing the advances made so far in mitigating drawbacks of GCE and the challenges that still remain. For every limitation of GCE, mitigating strategies applicable for prokaryotic systems are discussed first, followed by strategies for eukaryotic systems. For example, section 1.4.1 covers the strategies for rendering mRNA specificity to GCE. Here first orthogonal ribosomes (section 1.4.1.1) and genome reprogramming (section 1.4.1.2) are discussed, both of which have been developed in prokaryotic hosts, followed by synthetic membraneless organelles (section 1.4.1.3) which is applicable for eukaryotic systems. A similar structure has also been followed for section 1.4.2.

1.4.1 Strategies to render GCE mRNA specific

Orthogonality of the GCE machinery is key to ensuring minimum unspecific modification of the host proteome. Utilizing rarest occurring stop codons like amber codon for reassignment and GCE specific aaRS/tRNA pairs that do not interfere with host translation are effective measures but not sufficient. Additional strategies are necessary to restrict the desired modification to only the POI.

1.4.1.1 Orthogonal ribosomes for selective mRNA translation

mRNAs in bacteria and archaea are recruited to the ribosome by the interaction of the Shine-Dalgarno sequence, located about 8 base pairs upstream of the start codon, with the corresponding binding site on the ribosome.[133–135] Utilizing this interaction, in

1987 Hui and de Boer engineered ribosomes to specifically recruit and translate mRNAs with matching altered Shine-Dalgarno sequences.[136] Decades later, the same concept would form the basis for the development of orthogonal ribosome (O-ribosome)/mRNA pairs (figure 3) for selective GCE by the Chin lab.[137] In combination with a suitable GCE specific aaRS/tRNA pair, the O-ribosome system was shown to incorporate ncAA in response to the quadruplet codon AGGA in *E. coli* in a mRNA-specific manner.[137] The success of the O-ribosome concept inspired further optimization with the goal of GCE enhancement and resulted in several improved versions including ribo-X, ribo-Q and ribo-T. Ribo-X harbors two mutations namely U531G and U534A in the 16s rRNA and was engineered with the goal of hindering interaction with the release factor, thereby favoring amber suppression.[138] In combination with the relevant aaRS/tRNA , ribo-X could achieve 3- and 20- fold enhancement in single and dual amber codon suppression, respectively. Ribo-Q is another O-ribosome that has comparable amber suppression efficiency as ribo-X and additionally can also decode quadruplet codons.[139] In both ribo-Q and ribo-X, the smaller subunit of the ribosome has been engineered, while the larger subunit is exchanged freely between the wild-type and the engineered small units. In order to develop completely orthogonal ribosomes Jewett and Mankin labs tried an alternative approach.[140] They engineered ribosomes known as ribo-T, where the two subunits were tethered or connected by short RNA linkers. The length of the linkers was a critical parameter in ensuring the native interactions between the two subunits as well as with other translation components were not impaired. Improved variant of ribo-T/mRNAs could be successfully used for *in vivo* incorporation of pAzF in GFP in *E. coli*.[141] However, these tethered ribosomes have certain limitations. Ribo-Ts assemble slower than their wild-type counterparts thereby adversely affecting *E. coli* growth.[1] Translation initiation and termination by Ribo-T ribosomes have also been shown to be slightly impaired.[142] To develop upgraded variants of tethered ribosomes, the Jewett lab established a platform named Evolink (evolution and link) to enable high-throughput evolution of ribo-Ts.[143] Additionally attempts have also been made to combine the winning traits of O-ribosomes and tethered ribosomes and have resulted in the development of O-stapled-ribosomes.[142]

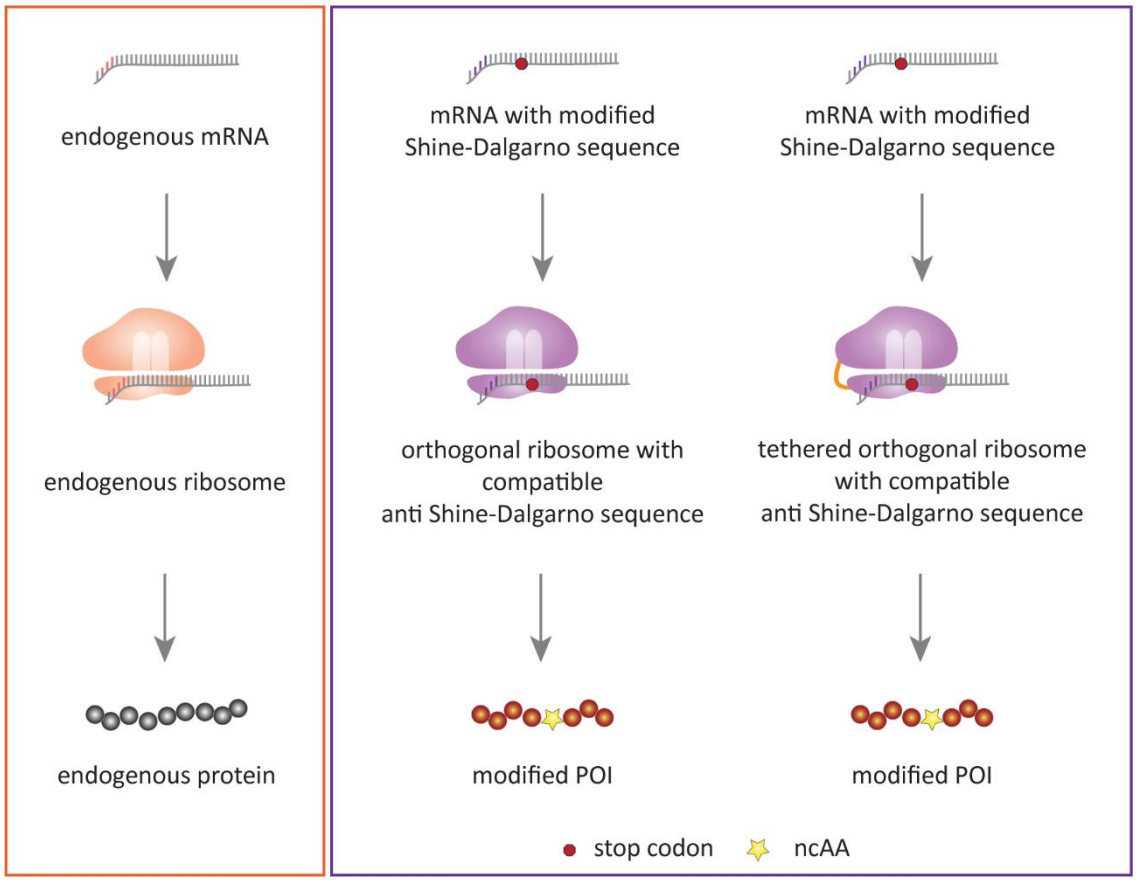


Figure 3| Orthogonal ribosomes selectively process mRNAs with compatible anti Shine-Dalgarno sequence.

1.4.1.2 Genome reprogramming and synthesis

61 sense codons code for 20 amino acids, which means there are redundant codons that can be potentially used for GCE. Reprogramming the entire genome of an organism to eliminate selected sense or stop codons can free up these codons for ncAA incorporation, at the same time solving the issue of unspecific host proteome modification. Multiplex automated genome engineering (MAGE) is a high-throughput genome editing technology where genetic modifications are introduced by repetitive cycles in a bacterial genome.[144] It is followed by conjugative assembly genome engineering (CAGE) to assemble the modified genomic fragments into the complete reprogrammed bacterial genome.[145,146] MAGE and CAGE facilitated the development of an engineered *E. coli* strain C321.ΔA where the TAG stop codon was replaced with the TAA codon and the gene for release factor RF1, specific for TAG codon, was also removed for added effect.[145] This strain could be used for GCE resulting in pAzF or 2-naphthylalanine (NapA) incorporation in a GFP reporter.[146] Despite success of the C321.ΔA strain, it is relevant to note that it had 355 off-target mutations and slower growth rate. Hence further engineering of the strain was performed, guided by the outcome of sequence comparisons between C321.ΔA and its ancestors' genomes.[147] Yokoyama and Sakamoto groups engineered two more *E. coli* strains, B-95.ΔA and B-95.ΔAΔfabR.[147] In this case instead of substituting all amber codons, only 95 of the 273 TAG codons were replaced. Similar to the C321.ΔA strain, the RF1 gene was also eliminated in these two strains. Alternative to MAGE and CAGE based genome

reprogramming, it is also possible to synthesize entire genomes. Replicon excision method for enhanced genome engineering through programmed recombination (REXER) is one of the methods enabling genome synthesis.[148] It was used to develop the syn61 and syn61Δ13 strains of *E. coli*. In these strains the TAG stop codon and two serine codons, TCG and TCA have been substituted with synonymous codons TAA, AGC and AGT.[149,150] Additionally, in the syn61Δ13 strain the genes for RF1 and the tRNAs decoding the substituted serine codons were deleted. Further optimization efforts yielded the Syn61Δ3(ev5) strain which was subsequently used for GCE applications.[150]

Besides ribosome and genome engineering, the use of artificial base pairs can also mitigate the issue of mRNA selectivity since such base pairs do not naturally occur in the host system. The artificial base pairs are discussed in detail in section 1.4.2.1.

1.4.1.3 Synthetic membraneless organelles to confine GCE

Selective GCE in eukaryotes is quite a challenge considering the elegant concepts of O-ribosome, artificial base pairs or complete genome synthesis are either not possible owing to the different cellular organization as compared to prokaryotes or pose a herculean effort. Eukaryotes are known to have diverse subcellular compartments, known as organelles, to separate distinct biochemical processes. Reinkemeier *et al*. took advantage of this strategy and attempted to build synthetic organelles to compartmentalize the entire process of GCE, thereby minimizing crosstalk with the host translational machinery (figure 4).[4,151,152] These organelles, referred to as orthogonally translating organelles (OTOs) are generated by the combined effect of phase separation and spatial targeting. The key components of an OTO are: (1) The anchoring domain: This domain is responsible for targeting the OTO to different subcellular locations like the plasma membrane, the endoplasmic reticulum membrane, the Golgi membrane etc. (2) The assembler domain: The purpose of the assembler domain is to create a condensate enriched in the target mRNA and the GCE machinery. Multiple assemblers were tested by Reinkemeier *et al*. The best performing OTOs utilize a phase separating protein as the assembler. This protein is fused to the GCE-specific synthetase and generate the scaffold of the OTO. (3) The RNA binding domain (RBD): This is comprised of different RNA binding proteins, for example MS2 bacteriophage coat protein (MCP), λN22 etc which can specifically bind to their cognate RNA motifs or loops like MS2 and BoxB respectively. The mRNA of the POI is tagged with such RNA loops at the 3'-UTR and is selectively recruited to the OTO by the specific interaction between the RNA loops and the respective RBDs. Since the OTOs are membraneless, all other components necessary for translation can self-recruit into the OTOs. In their pioneering work, the Lemke lab demonstrated a selectivity of about 8fold for amber suppression with OTOs localized at the microtubule plus ends as compared to cytoplasmic GCE that is GCE without an organelle.[4] Furthermore, they developed thin film-like OTOs on different membrane surfaces whose small dimension allowed generation of multiple such OTOs in the same cell.[151] Two such film-like OTOs facilitated simultaneous suppression of the same amber codon, TAG, in two different POIs to incorporate two distinct ncAAs respectively. Generation of multiple orthogonal translation systems relies on a set of prerequisites. (1) Independent assembly: The OTOs must not intermix with each other. (2)

Selective RNA recruitment: Each OTO must recruit distinct mRNAs of interest. (3) Distinct ncAA specificity: The aaRSs enriched in the respective OTOs should not have the same ncAA substrate. With these film-like OTOs more than an order of magnitude fold change in selectivity was observed as compared to the classical cytoplasmic GCE. The OTOs have been extensively used for fluorescence microscopy applications in mammalian cells.[151] Recently the mitochondrial membrane anchored OTO was used for dual amber suppression in the disordered region of nucleoporin 98 (NUP98).[153] The distance between the two ncAAs was varied and on subsequent labelling with donor and acceptor dyes fluorescence lifetime imaging of the FRET pair was performed. This elucidated information regarding the conformation of NUP98 in the central core of the nuclear pore complex.

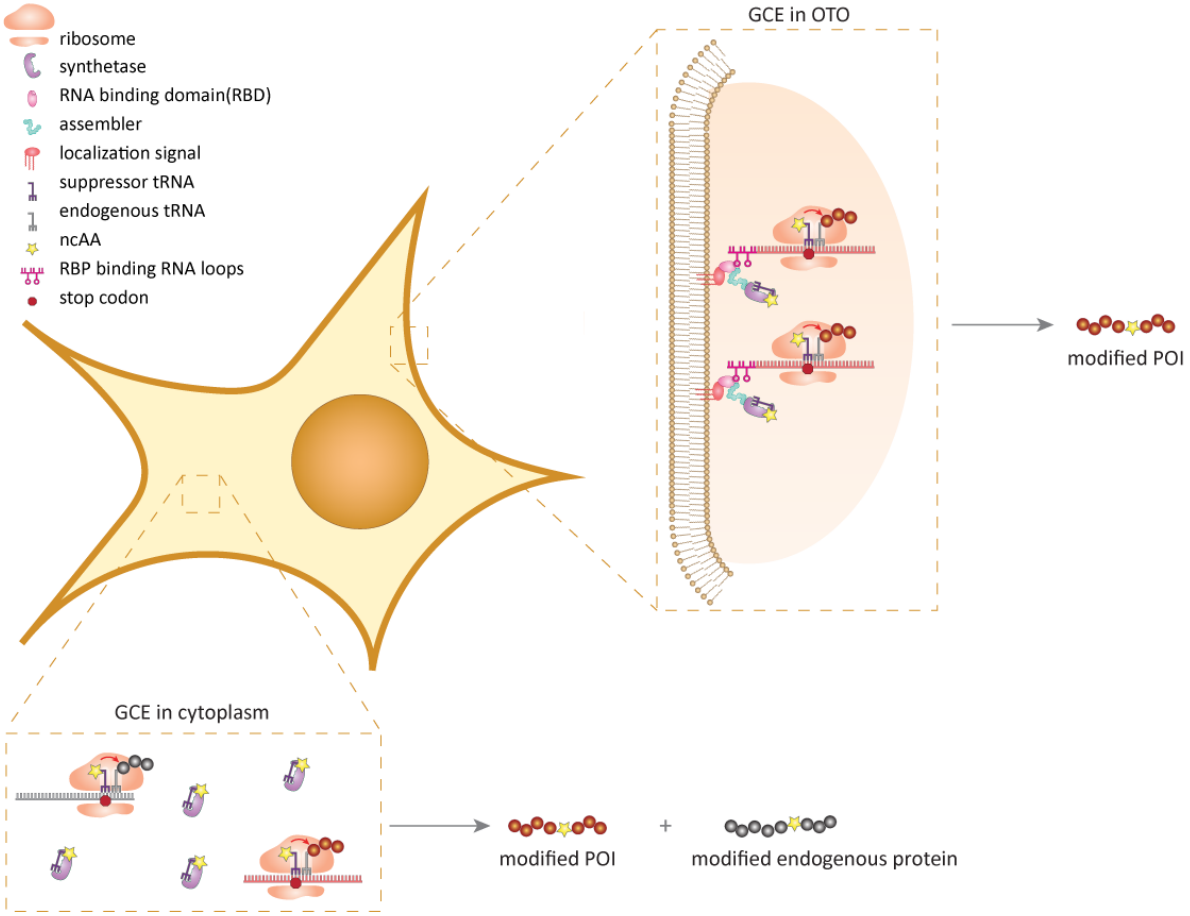


Figure 4| OTOs render mRNA specificity to GCE. OTOs, formed by the combined effect of spatial targeting and phase separation, confine GCE to an artificial membraneless compartment in the cell. The mRNA of interest is tagged with RNA loops that can interact specifically with certain RBDs. This RNA loop-RBD interaction is utilized to selectively recruit the target mRNA to the OTO. The GCE specific synthetase is enriched in the OTO and other components necessary for translation can freely gain access into the OTO. As a result, upon utilizing the OTO the issue of unspecific modification of stop codons in the host proteome can be mitigated. In comparison, in the absence of the OTO, when GCE machinery is present in the cytoplasm, both the POI as well as the host proteome get modified with ncAAs.

1.4.2 Strategies against limited number of codons for reassignment

The far-reaching potential of GCE has been established without a doubt by the myriad of enabling applications and justifies the limitless pursuit of refining this technology to perfection. The ability to establish orthogonal genetic codes in host organisms opens up

the possibility of synthesizing designer biopolymers of potentially any desired functionality. However, being constrained to only the three available stop codons for reassignment poses a hindrance toward achievement of such a goal. Besides the potential of artificial biopolymers with ncAAs, development of which might still be years away, just the incorporation of multiple probes in a POI can simplify the study of their respective effects on the POI or lead to the development of novel microscopic techniques as explained in section 3.9.3. Currently the alternative to stop codons for GCE are artificial base pairs, quadruplet codons and sense codons and this section will discuss each along with their advantages and limitations. Similar to section 1.4.1, in this section as well first strategies exclusively in prokaryotes is discussed in sections 1.4.2.1 followed by approaches applicable to both prokaryotes and eukaryotes in sections 1.4.2.2 and 1.4.2.3.

1.4.2.1 Artificial base pairs

Artificial base pairs refer to synthetic nucleic acids that can substitute the naturally occurring base pairs in DNA and RNA. Combinations of these artificial base pairs can form new codons to be subsequently reassigned to ncAAs. Initial research in this direction dates back to 1989 with the development of iso-C and iso-G, constitutional isomers of C and G, by the Benner group.[154] The respective deoxyribonucleotide as well as the ribonucleotides could form base pairs *in vitro* giving rise to synthetic DNA and RNA molecules. These bases could also be recognized as substrates by the DNA polymerase Klenow enzyme and the bacterial T7 RNA polymerase leading to the generation of a six-lettered genetic code. This code was translated *in vitro* to a peptide chain containing the ncAA 3-iodotyrosine.[155] However, the iso-dC was not stable and was prone to hydrolyzation to Uracil.[154] As the next step, the same group developed dκ and dπ.[156] The pyrimidine dκ was also compatible for pairing with the natural base deoxyxanthosine, however the latter's tendency to undergo depurination resulted in the selection of the artificial purine dπ for the base pair dκ-dπ. Another such pair dS-dY was shown effective for the generation of additional codon for the incorporation of the ncAA 3-chlorotyrosine into the protein Ras in an *E. coli* derived cell-free system.[157] An improved derivative of the iso-C::iso-D was also developed, known as dZ-dP , which exhibited high stability and fidelity rates.[158,159] Yet another noteworthy class of artificial base pairs are the Hachimoji nucleic acids.[160]

The strategies explored so far for the development of artificial base pairs include isomers of the canonical nucleic acids as explained above, bases with alternative binding interactions other than classical hydrogen bonds, for example hydrophobic interactions or interactions based on copper or silver ions and nucleic acids with completely altered backbone chemistry like the xeno nucleic acids (XNAs). The dNaM-d5SICS and dNaM-dTPT3 (figure 5), developed by the Romsberg group, are notable examples of hydrophobic interaction based artificial base pairs with a remarkable amplification fidelity of 99.9%.[161,162]

The approach of expanding the genetic code at the nucleic acid level promises remarkable advancements. These artificial base pairs can generate completely orthogonal information transfer systems in host organisms. However, besides

optimization of the bases themselves there are more roadblocks that need to be overcome to utilize these for *in vivo* GCE. Although some classes of artificial bases are compatible with existing polymerases, bases like the XNAs require development of new replication and transcription machineries. Biosynthesis of these chemically derived bases also poses quite a challenge. Research toward solving this issue is focused on external supply of certain precursors to the cells which can be converted to the desired artificial base *in vivo* by certain enzymes. The success of this strategy relies on efficient uptake of these natural/synthetic precursors by the host organism. The result of an extensive screen of NTR transporters in *E. coli* led to the discovery of the transporter NTT2 from the diatom algae *Phaeodactylum tricornutum* that could import the dNTPs d5SICSTP and dNaMTP.[163] Replication and transcription of templates containing the base pair d5SICS–dNaM was observed in *E. coli* demonstrating the effectiveness of the NTT2 transporter and subsequent *in vivo* incorporation of ncAA Nε -[(2-propynyloxy)carbonyl]-L-lysine (PRK) was achieved in sfGFP by reassigning codons composed of this base pair.[164–166] It is important to note here that the flanking sequence of the artificial codon plays a significant role in determining its final retention rate as it was possible to vary this rate from a 100% to complete loss by mutating the flanking sequence.[165] An updated version of the NTT2 containing *E. coli* was generated by engineering the NTT2, modifying the base-pairs and optimizing the codon context for better artificial base pair retention.[166]

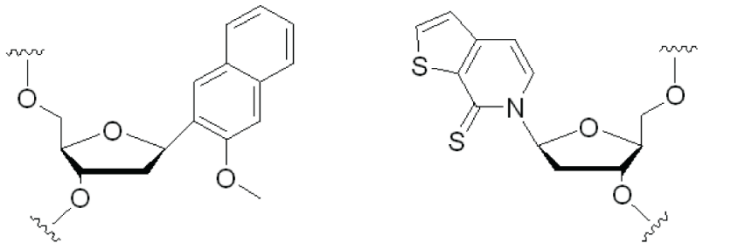


Figure 5| Examples of artificial base pairs

1.4.2.2 Quadruplet codons

Quadruplet codons offer an alternate strategy to overcome the limited availability of codons for reassignment to ncAAs. These are composed of combinations of four bases

and expand the genetic alphabet by 256 codons. In *E. coli* it has already been possible to incorporate 4 distinct ncAAs in a POI by reassigning 4 different quadruplet codons namely CUAG, UAGA, AGGA and AGUA.[167] The Schultz lab was able to develop this technology for the eukaryotes and use AGGA and UAGA codons to incorporate ncAA in POI in 293T cells.[168,169] The use of quadruplet codons also proved to be a promising approach toward generating attenuated virus for vaccine development. Engineered HIV viruses dependent on UAGA decoding machinery for successful replication and propagation, were generated, thereby minimizing the risk of host infection.[169] Greiss and coworkers were the first to implement the use of quadruplet codons in multicellular organisms for ncAA incorporation and synthesized Cre recombinase containing photocaged lysine and caspase-3 containing photocaged cysteine in *C. elegans*.[170] Usage of these photocaged ncAAs allowed optical control of processes like gene expression and cell ablation respectively.

A prerequisite for efficient quadruplet codon decoding is the availability of the corresponding appropriate aaRS/tRNA pairs. In cases where the anticodon loop of the tRNA does not play a role in its recognition by the synthetase, an extension of the anticodon loop to read the quadruplet codon could be a solution. The PylRS/Pyl-tRNA is such a system and was utilized for decoding the AGGA codon by the Schultz lab. However, there was still the possibility that the mutation of the anticodon loop could adversely affect the tRNA ribosome interaction. Hence the Pyl-tRNA needed to be evolved first. The four bases of the anticodon loop were randomized and the final Pyl-tRNA candidate was chosen after rounds of positive and negative selection tests performed in *E. coli*.[169] Mills *et al.* performed a comparative study with 11 variants of quadruplet codon decoding Pyl-tRNAs including 5 pre-validated ones.[171] The Pyl-tRNAs with anticodon NCUA (N = A/U/G/C) were found to be non-functional in HEK293T cells whereas $tRNA_{UCCU(Ev2)}$ was evaluated as the best working candidate.[168] Further instances of tRNA engineering include the mutations designed by the Greiss lab for their $tRNA_{UAGA}$ to mimic the endogenous tRNA of *C. elegans*.[170]

1.4.2.3 Reassignment of sense codons

Artificial base pairs and quadruplet codons are excellent strategies for expanding the availability of codons for GCE, however both have their limitations and await further development to enable versatile applicability. An alternative in this case would be to turn to the 61 available sense codons. Sense codon reassignment can be done either residue specifically or as recently shown, in a site-specific manner.

In case of residue specific sense codon reassignment, ncAAs are incorporated proteome wide in response to naturally occurring sense codons in the host organism.(figure 6) Endogenous aaRS/tRNA pairs, for example, methionine RS (MetRS)/tRNA are promiscuous enough to accept ncAAs having close structural similarity to the cognate canonical amino acid, in this case methionine. Taking advantage of this phenomena, methionine analogues like azidohomoalanine (AHA) and homopropargylglycine (HPG) could be incorporated in the host proteome by repurposing ATG codons.[172] Residue specific sense codon reassignment is commonly used in proteomics studies where newly synthesized proteins in response to specific stimuli or in selected cell types and

developmental stages are investigated. AHA or HPG, by virtue of the azide or alkyne group respectively can bio-orthogonally react with compatible probes for affinity purification. The proteins purified by this azide / alkyne mediated click chemistry can subsequently be identified by mass spectrometry analysis. Based on this application, the Tirrell and Schumann groups developed the technique bio-orthogonal non-canonical amino acid tagging (BONCAT). The applicability of BONCAT has been proven in various host systems including mammalian cells, zebrafish, mouse and plant models. In order to study nascent proteins in a specific cell type Alvarez-Castelao *et al.* expressed an engineered MetRS on Cre-recombinase induction in a mouse model.[173] Using this system it was possible to selectively label the newly synthesized proteins in excitatory principal neurons and Purkinje neurons thereby enabling identification of more than 200 proteins that are affected by certain external stimuli. Taupathy induced effect on protein expression could be studied in K369I tau transgenic K3 mouse model by the use of BONCAT technology.

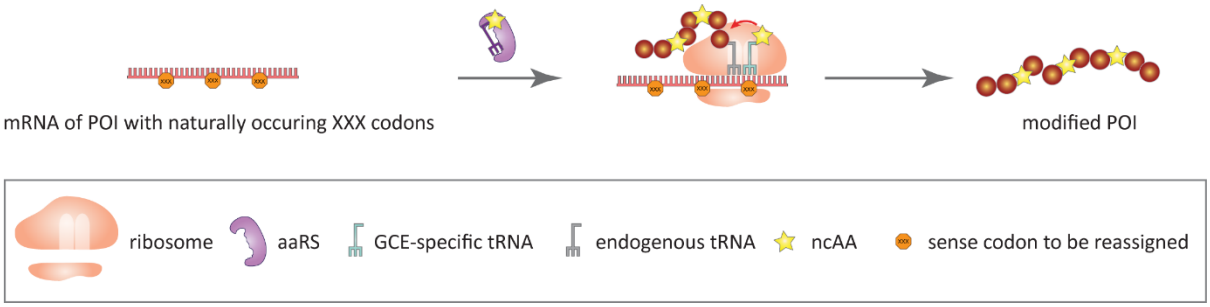


Figure 6| Residue-specific sense codon reassignment. In residue-specific sense codon reassignment a naturally occurring sense codon in the host transcriptome (XXX) is reassigned to incorporate ncAAs.

Besides tags for affinity purification, fluorescent dyes with compatible groups can also click-react with the methionine homologues used in BONCAT to allow fluorescent imaging of the host proteome and the technique is termed as fluorescent non-canonical amino acid tagging (FUNCAT).[174] FUNCAT, when combined with proximity ligation assay (FUNCAT-PLA) can enable the detection of specific proteins. In FUNCAT-PLA two primary antibodies are necessary, one against the POI and the other against the reactive groups of the incorporated ncAA. The secondary antibodies are supplied with oligonucleotides that can only ligate when the two primary antibodies are on the POI. Subsequent to ligation, the oligonucleotides are amplified by rolling cycle amplification and labelled with compatible probes. Another variant of BONCAT is stochastic orthogonal recoding of translation (SORT). Unlike BONCAT, SORT utilizes anticodon mutated GCE-specific tRNAs and evolved aaRSs to reassign any sense codon to any desired ncAA occurring in the host transcriptome. Depending on the subsequent detection methodology SORT can be classified as either SORT-M or SORT-E. SORT-M refers to stochastic orthogonal recoding of translation with chemoselective modification and has been utilized to fluorescently label newly synthesized proteins in specific tissues at specific developmental stages in *Drosophila melanogaster*.[175] SORT-E on the other hand is useful when the goal is to specifically enrich the ncAA incorporated proteins and to identify these proteins by mass spectrometry analysis.[176] The plethora of applications in proteomics studies prove residue specific sense codon reassignment to

be a powerful technology that can harness the potential of repurposing the 61 sense codons. However, this technology can neither be applied for site-specific ncAA incorporation nor has it been, so far shown, to selectively modify a specific POI.

Recent work by Ding *et al*. demonstrates the use of rare occurring sense codons for site-specific incorporation of ncAA in a POI.[177] Unlike stop codon suppression, sense codon reassignment faces the tremendous challenge of competing with the abundant endogenous tRNAs. Rare codons have a frequency of usage less than 1% and a lower abundance of decoding tRNAs, thereby making them preferred choice for reassignment. Their rarity of occurrence minimizes chances of unspecific modification of the host proteome and there are less endogenous tRNAs to compete with. In this study the rare TCG codon was selected for site-specific ncAA incorporation in mammalian cells. Multiple optimization steps including removal of the canonical amino acid from the growth media ,tRNA engineering, knocking down of endogenous tRNA, stable expression of GCE components, designing flanking sequence of the TCG codon based on the concept of codon context effect on reassignment were employed to achieve incorporation of upto 6 ncAAs in a POI in response to 6 TCG codons. The ncAA incorporation was detected by gel-based assays and mass spectrometry. Although it marks a significant advancement in expanding GCE to sense codons, it is not certain if the technology will be suitable for applications like labelling POI for fluorescent microscopy. An in-depth discussion of this publication and comparison with the work done in this thesis is covered in section 3.7.

1.5 Objectives – mRNA selective site- and residue-specific sense codon reassignment in mammalian cells

GCE empowers us to ultimately synthesize artificial designer biopolymers. One of the steps to reach this challenging goal would be to expand the set of blank codons available for ncAA incorporation. Harnessing the potential of sense codon reassignment is one of the strategies to take classical GCE beyond the three stop codons. Hence the two major aims of this thesis are to accomplish site- (1.5.1) and residue-specific (1.5.2) sense codon reassignment selectively in a POI in mammalian cells.

1.5.1 mRNA selective site-specific sense codon reassignment in mammalian cells

Aim one of this thesis is to demonstrate site-specific sense codon reassignment in a mRNA selective manner in mammalian cells. In case of site-specific sense codon reassignment, similar to stop codon suppression, specific positions are selected in the gene of the POI and these positions are mutated to introduce the sense codon meant for reassignment. The gene of the POI needs to be codon optimized so that the sense codon for reassignment does not occur at any other site besides the ones pre-selected. The major aim of performing selective site-specific sense codon reassignment is divided into the following, 1.5.1.1 and 1.5.1.2, steps.

1.5.1.1 Site-specific reassignment of sense codons selectively in a POI by film-like OTOs

First step of achieving site-specific reassignment of sense codons selectively in a POI in mammalian cells, is to prove the ability of the film-like OTOs to render mRNA selectivity to site-specific sense codon reassignment for a selected set of sense codons.

1.5.1.2 Application of mRNA selective site-specific sense codon reassignment to fluorescently label POIs for subsequent confocal microscopy experiments

After establishing the applicability of the OTOs to render mRNA selectivity to site-specific sense codon reassignment, my second step is to apply this technology to site-specifically label a POI for visualization by confocal microscopy.

1.5.2 mRNA selective residue-specific sense codon reassignment in mammalian cells

Aim two of this thesis is to achieve mRNA selective residue-specific sense codon reassignment in mammalian cells. As discussed in section 1.4.2.3 residue-specific sense codon reassignment has found widespread use in proteomics studies. However, the immense potential of this technique to incorporate probes in a mRNA selective manner has not been explored. My aim of performing selective residue-specific sense codon reassignment is divided into the following,1.5.2.1, 1.5.2.2 and 1.5.2.3, three steps.

1.5.2.1 Establishing a screening pipeline to select suitable sense codons for residue-specific reassignment in a POI

As a first step towards performing mRNA selective residue-specific sense codon reassignment, a microscopy-based screening pipeline will be established to select suitable sense codons for reassignment for each POI.

1.5.2.2 Demonstrating the applicability of film-like OTOs to render mRNA selectivity to residue-specific sense codon reassignment

In the second step of aim two, the sense codons selected from the screening experiment (section 1.5.2.1), will be reassigned individually in a POI to demonstrate the ability of the OTOs to render mRNA selectivity to residue-specific sense codon reassignment in mammalian cells.

1.5.2.3 Application of mRNA selective residue-specific sense codon reassignment to selectively label a POI for subsequent imaging by confocal microscopy

Finally, in step 3, mRNA selective residue-specific sense codon reassignment will be used to fluorescently label a POI to facilitate subsequent visualization of the POI by confocal microscopy in distinct mammalian cell lines.

Together, these objectives form the fundament for future applications of GCE for the development of a novel protein labelling technique to visualize shape of proteins *in vivo* by SRM and *in vivo* synthesis of designer biopolymers.

2 Materials & Methods

2.1 Materials

Table 1: List of buffers

Buffer	Specifications
1X Phosphate-buffered saline (PBS)	Gibco™, ThermoFisher Scientific (10010-015)
Hepes	Carl Roth (9105.4)
10x Transport Buffer (TB)	200 mM HEPES 1.1 M Potassium acetate 50 mM Sodium acetate 20 mM Magnesium acetate 10 mM EGTA
1x TB + PEG6000 + DTT	1X TB 5 mg PEG6000/mL 1X TB 2 mM DTT
MOPS SDS Running Buffer (20X)	NuPAGE™, ThermoFisher Scientific (NP0001)
50X Tris-acetate-EDTA (TAE) buffer	IMB, Mainz core facility
10X T4 DNA ligase buffer	Lucigen (F88912-1)
5X FastDigest Green-buffer	ThermoFisher Scientific, Invitrogen
5X Phusion HF buffer	Thermo Scientific (F-518)
5X SuperFi II buffer	Invitrogen/ThermoFisher Scientific
Annealing buffer	200 µL 1M TrisHCl pH 7.5 25 µL 5M NaCl 100 µL 1M $MgCl_2$ 575 µL water
Radioimmunoprecipitation assay buffer (RIPA)	Supplied with GFP-Trap® Magnetic Particles M-270 Kit from chromoTek
Lysis buffer for western blot	1X Complete™, EDTA-free protease inhibitor cocktail 1.2 cU/µL SmNuclease 1 mM $MgCl_2$ in RIPA buffer
2X SDS buffer for elution	120 mM Tric/Cl pH 6.8 20% glycerol 4% SDS 0.04% bromophenol blue (can be eliminated if concentration of the eluted protein needs to be measured using the Pierce reagent) 10% β-mercaptoethanol

Table 2: List of enzymes and mastermix

Enzyme	Specifications
NotI	Thermo Fisher Scientific (FD0595)
MluI	Thermo Fisher Scientific (FD0564)
BshTI	Thermo Fisher Scientific (FD1464)
ApaI	Thermo Fisher Scientific (FD1414)
NheI	Thermo Fisher Scientific (FD0974)
BglII	Thermo Fisher Scientific (FD0083)
DpnI	Thermo Fisher Scientific (FD1704)
BpiI	Thermo Fisher Scientific (ER1012)
Eco31I	Thermo Fisher Scientific (FD0294)
Phusion™ Hot Start II High-Fidelity DNA-Polymerase	Thermo Fisher Scientific (F549L)
Platinum™ SuperFi II DNA-Polymerase	Thermo Fisher Scientific (12361010)
EconoTaq polymerase	Lucigen (30031-1-LU)
T4 DNA ligase	Lucigen (30241-1-LU)
2X Gibson mastermix	IMB core facility
Alkaline phosphatase (FastAP)	Thermo Fisher Scientific (EF0651)
Sm nuclease	IMB core facility

Table 3: List of antibodies

Antibody	Specifications
Anti-flag (mouse)	Sigma-Aldrich (F3165)
Anti-myc (mouse)	Cell Signaling (2276)
Anti-mouse IRDye 800CW	LI-COR Biosciences (926-32212)
Anti-mouse Alexa 488	Thermo Fisher Scientific (A-11001)

Table 4: List of ncAAs

ncAA	Specifications
TCO*AK	Sichem (SC-8008)
BOCK	Iris Biotech (HAA1096.0025)
SCOK	Sichem (SC-8000)
CbzK	Sigma-Aldrich (96840)
3IF	Chem-Impex International Inc., CAS 20846-39-3
100mM stock solution of ncAAs were prepared in 0.2M NaOH in 15% DMSO	

Table 5: List of dyes

Dyes	Specifications
LD655-H-Tetrazine	Lumidyne
Cy5-H-Tetrazine	Conju Probes/ Jena Bioscience (CP-4006)

Table 6: List of kits

Kit	Specifications
PureLink™ Quick Plasmid Miniprep Kit	Invitrogen, ThermoFisher Scientific (K210011)
PureLink™ HiPure Plasmid Maxiprep Kit	Invitrogen, ThermoFisher Scientific (K210017)
PureLink™ PCR Purification Kit	Invitrogen, ThermoFisher Scientific (K310002)
Monarch DNA Gel Extraction Kit	New England Biolabs (T1020L)
Pierce™ 660nm Protein Assay Reagent Ionic Detergent Compatibility Reagent (IDCR)	Thermo Scientific (22660) Thermo Scientific (22663)
GFP-Trap® Magnetic Particles M-270 Kit	chromotek (AB_2827592)

Table 7: List of other chemicals

Chemicals	Specifications
Dulbecco's Modified Eagle Medium (DMEM), high glucose	Gibco™, ThermoFisher Scientific (41965062)
DMEM++++	DMEM high glucose 9% FBS 1% L-glutamine 1% Sodium pyruvate 1% Penicillin-Streptomycin
DMEM, low glucose, pyruvate	Gibco™, ThermoFisher Scientific (31885023)
DMEM+++	DMEM, low glucose, pyruvate 9% FBS 1% L-glutamine 1% Penicillin-Streptomycin
DMEM, low glucose, pyruvate, no glutamine, no phenol red	Gibco™, ThermoFisher Scientific (11880028)
Digitonin	Panvreac (A1905)
Fetal Bovine Serum (FBS)	Sigma-Aldrich (F7524)
Sodium pyruvate (100mM)	Thermo Fisher Scientific (11360039)
Penicillin-Streptomycin (10.000 U/ml)	Thermo Fisher Scientific (15140122)
L-Glutamine (200 mM)	Thermo Fisher Scientific (25030081)
Trypsin-EDTA (0.05%), phenolred	Thermo Fisher Scientific (25300054)
Polyethylenimin (PEI)	Sigma-Aldrich (408727)
JetPRIME	Polyplus (11407)
JetPRIME buffer	Polyplus
Lysogeny Broth (LB) media mix	Carl Roth (X964.4)
LB agar plates	IMB media kitchen
SOC medium	Bacto-Tryptone 2% Yeast-Extract 0.5% NaCl 5M

	$MgCl_2$ 1M KCl 1M $MgSO_4$ 1M Glucose 20%
Milk powder	Carl Roth (T145.3)
Bovine Serum Albumin (BSA)	Sigma-Aldrich (A7906-50G)
Pre-diluted protein assay standards (BSA standards)	Thermo Scientific (23208)
Ampicillin-sodium salt BioChemica	AppliChem (A0839,0010)
Kanamycin sulfate	AppliChem (A1493,0050)
PageRuler™ Prestained Protein Ladder, 10 to 180 kDa	Thermo Fisher Scientific (26616)
SDS loading dye (5X)	312.5 mM Tris-Cl pH 6.8 25% Glycerol 10% SDS 5mM β-Mercaptoethanol 0.05% Bromophenol blue
GeneRuler 1 kb Plus DNA Ladder	ThermoFisher Scientific (SM1331)
DNA loading dye (6X)	40% (w/v) sucrose in water 0.25% (w/v) Bromophenol blue
Agarose	Sigma-Aldrich (A9539-500G)
Ultrapure Agarose	Invitrogen (16500-100)
Ethidium Bromide	Fluka (46066)
Paraformaldehyde (PFA)	Sigma-Aldrich (158127)
Deoxynucleotide (dNTP) Solution Mix	New England Biolabs (N0447S)
$MgCl_2$ (50 mM)	Thermo Scientific, F-510Mg
Complete™, EDTA-free protease inhibitor cocktail	Sigma-Aldrich (118753580001)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich (472301)
phenylmethylsulfonyl fluoride (PMSF) BioChemica	AppliChem (A0999.0025)
PEG(polyethylene glycol)6000	Sigma (81253)
Potassium acetate	Merck KGaA (1.04820.1000)
Sodium acetate	Merck KGaA (1.06268.1000)
Magnesium acetate	Merck KGaA (1.05819.0250)
ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA)	PanReac AppliChem (A0878,0025)
Tween 20	Sigma (P1379)
Triton X-100	Sigma (T8787)

Table 8: List of plasmids generated for this thesis

(For sequences please refer to Appendix II)

Vectors used for cloning : pUC57-Kan**,** pBI, pcDNA3.1/Zeo(+)

Plasmid	**Cloning strategy**
Pyl-tRNA^{XXX} (pUC57-Kan_U6Full_aa XXX, where U6 – promoter, aa- amino acid, XXX- codon to be reassigned)	
pUC57-Kan_U6Full_Ser TCT	Quick change mutagenesis
pUC57-Kan_U6Full_Ser TCC	Quick change mutagenesis
pUC57-Kan_U6Full_Ser TCA	Quick change mutagenesis
pUC57-Kan_U6Full_Ser AGT	Quick change mutagenesis
pUC57-Kan_U6Full_Thr_ACT	Quick change mutagenesis
pUC57-Kan_U6Full_Thr ACA	Quick change mutagenesis
pUC57-Kan_U6Full_Cys_TGT	Quick change mutagenesis
pUC57-Kan_U6Full_Tyr _TAT	Quick change mutagenesis
pUC57-Kan_U6Full_Tyr TAC	Quick change mutagenesis
pUC57-Kan_U6Full_Gln CAG	Quick change mutagenesis
pUC57-Kan_U6Full_Gly GGT	Quick change mutagenesis
pUC57-Kan_U6Full_Gly GGC	Quick change mutagenesis
pUC57-Kan_U6Full_Asp AAT	Quick change mutagenesis
pUC57-Kan_U6Full_Asp AAC	Quick change mutagenesis
pUC57-Kan_U6Full_Gln CAA	Quick change mutagenesis
pUC57-Kan_U6Full_Gly GGA	Quick change mutagenesis
pUC57-Kan_U6Full_Gly GGG	Quick change mutagenesis
pUC57-Kan_U6Full_Ala GCT	Quick change mutagenesis

pUC57-Kan_U6Full_Ala GCC	Quick change mutagenesis
pUC57-Kan_U6Full_Ala GCA	Quick change mutagenesis
pUC57-Kan_U6Full_Ala GCG	Quick change mutagenesis
pUC57-Kan_U6Full_Val_GTT	Quick change mutagenesis
pUC57-Kan_U6Full_Val_GTC	Quick change mutagenesis
pUC57-Kan_U6Full_Val_GTA	Quick change mutagenesis
pUC57-Kan_U6Full_Val_GTG	Quick change mutagenesis
pUC57-Kan_U6Full_Leu_CTT	Quick change mutagenesis
pUC57-Kan_U6Full_Leu_CTC	Quick change mutagenesis
pUC57-Kan_U6Full_Leu_CTG	Quick change mutagenesis
pUC57-Kan_U6Full_Leu_TTA	Quick change mutagenesis
pUC57-Kan_U6Full_Leu_TTG	Quick change mutagenesis
pUC57-Kan_U6Full_Ile_ATT	Quick change mutagenesis
pUC57-Kan_U6Full_Ile_ATC	Quick change mutagenesis
pUC57-Kan_U6Full_Ile_ATA	Quick change mutagenesis
pUC57-Kan_U6Full_Met_ATG	Quick change mutagenesis
pUC57-Kan_U6Full_Trp_TGG	Quick change mutagenesis
pUC57-Kan_U6Full_Phe_TTT	Quick change mutagenesis
pUC57-Kan_U6Full_Phe_TTC	Quick change mutagenesis
pUC57-Kan_U6Full_Pro_CCT	Quick change mutagenesis
pUC57-Kan_U6Full_Pro_CCA	Quick change mutagenesis
pUC57-Kan_U6Full_Asp_GAT	Quick change mutagenesis
pUC57-Kan_U6Full_Asp_GAC	Quick change mutagenesis

pUC57-Kan_U6Full_Glu_GAA	Quick change mutagenesis
pUC57-Kan_U6Full_Glu_GAG	Quick change mutagenesis
pUC57-Kan_U6Full_Lys_AAA	Quick change mutagenesis
pUC57-Kan_U6Full_Lys_AAG	Quick change mutagenesis
pUC57-Kan_U6Full_Arg_AGA	Quick change mutagenesis
pUC57-Kan_U6Full_Arg_AGG	Quick change mutagenesis
pUC57-Kan_U6Full_Arg_CGC	Quick change mutagenesis
pUC57-Kan_U6Full_Arg_CGA	Quick change mutagenesis
pUC57-Kan_U6Full_Arg_CGG	Quick change mutagenesis
pUC57-Kan_U6Full_His_CAT	Quick change mutagenesis
pUC57-Kan_U6Full_His_CAC	Quick change mutagenesis
pUC57-Kan_U6Full_2xtRNA_Thr_ACC	Cloning strategy by Rohilla *et al*.[178]
pUC57-Kan_U6Full_3xtRNA_Thr_ACC	Cloning strategy by Rohilla *et al*.
pUC57-Kan_U6Full_5xtRNA_Thr_ACC	Cloning strategy by Rohilla *et al*.
pUC57-Kan_U6Full_7xtRNA_Thr_ACC	Cloning strategy by Rohilla *et al*.
Reporters XXX - codon to be reassigned, CCC codon was used to introduce proline at site 66 of EGFP to generate the inactive/dark $EGFP^{66P}$. Hence the dark EGFP is denoted as $EGFP^{66CCC}$ in the following plasmid names.	
pBI_myc_vimentin_mcerulean_2xMS2	Gibson assembly
pBI_myc_$vimentin^{116TAG}$_mcerulean_2xMS2	Quick change mutagenesis
pEGFP_NUP153_2xMS2	Annealed oligonucleotide to introduce MS2 loops

pBI_CMV1_Flag_EGFP^{39CTA}_4xBoxB_CMV2_Flag_EGFP^{39CTA}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39ATA}_4xBoxB_CMV2_Flag_EGFP^{39ATA}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39TTA}_4xBoxB_CMV2_Flag_EGFP^{39TTA}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39CGT}_4xBoxB_CMV2_Flag_EGFP^{39CGT}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39TCG}_4xBoxB_CMV2_Flag_EGFP^{39TCG}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39TGC}_4xBoxB_CMV2_Flag_EGFP^{39TGC}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39CCG}_4xBoxB_CMV2_Flag_EGFP^{39CCG}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39ACG}_4xBoxB_CMV2_Flag_EGFP^{39ACG}_EGFP^{66CCC}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39CTA}_EGFP^{66CCC}_4xBoxB_Flag_CMV2_EGFP^{39CTA}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39ATA}_EGFP^{66CCC}_4xBoxB_Flag_CMV2_EGFP^{39ATA}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39CGT}_EGFP^{66CCC}_4xBoxB_Flag_CMV2_EGFP^{39CGT}_ 2xMS2	Gibson assembly & restriction-ligation
pBI_CMV1_Flag_EGFP^{39TGC}_EGFP^{66CCC}_4xBoxB_Flag_CMV2_EGFP^{39XXX}_ 2xMS2	Gibson assembly &

	restriction-ligation
pBI_CMV1_Flag_EGFP^{39CCG}_EGFP^{66CCC}_4xBoxB_Flag_CMV2_EGFP^{39XXX}_2xMS2	Gibson assembly & restriction-ligation

Table 9 : List of plasmids obtained from Lemke lab repository and used for experiments in this thesis

pcDNA3.1_LCK::MCP::AF::F::AF [151]
pcDNA3.1_LCK::MCP::AA::F::AA
pcDNA3.1_LCK::MCP::F::AF [151]
pcDNA3.1_LCK::λN22::AF::F::AF
pcDNA.3.1_LCK::F::AF
pcDNA3.1_LCK::MCP::ΔFUS::AF
pcDNA3.1_LCK::MCP::AF
pcDNA3.1_NES::AF [58]
pUC57-Kan_U6Full_tRNA_Pro_CCC
pUC57-Kan_U6Full_tRNA_Thr_ACC
pUC57-Kan_U6Full_tRNA_Cys_TGC
pUC57-Kan-_U6Full_tRNA_Arg_CGT
pUC57-Kan_U6Full_tRNA_Ser_TCG
pUC57-Kan_U6Full_tRNA_Leu_CTA
pUC57-Kan_U6Full_tRNA_Pro_CCG
pUC57-Kan_U6Full_tRNA_Thr_ACG
pUC57-Kan_U6Full_tRNA_Ser_AGC
pBI_CMV1_Flag_EGFP^{39CGT}_4xBoxB

Table 10: List of equipments

Equipment	**Manufacturer**
Centrifuges (5910R, 5910Ri, 5427R)	Eppendorf
GenePulser Xcell	Bio-Rad
Thermomixer C	Eppendorf
DeNovix DS-11 Spectrophotometer	DeNovix Inc.
Trans-Blot Turbo Transfer System	Bio-Rad
myFuge mini centrifuge	Benchmark Scientific
PowerPac Universal	Bio-Rad
XCell Surelock Novex Mini-Cell	Thermo Fisher Scientific
Mini-Sub Cell GT Cell	Bio-Rad

Pipetus (pipette-controller)	Hirschmann
Vortex Mixer	Greiner bio-one
Precision balance (0.01 – 3500 g)	KERN & SOHN GmbH
Analytical balance (0.1 mg – 320 g)	KERN & SOHN GmbH
Absorbance 96 plate reader	Enzo Life Sciences
ChemiDoc MP imaging system	Bio-Rad
T100 Thermal Cycler	Bio-Rad
Cell culture laminar flow workbench (Herasafe 2030i)	Thermo Scientific
Milli-Q Water purification system	Merck KGaA
Incubator	Heraeus Intruments
CO2-Incubator (HeraCell 150i)	Thermo Fisher Scientific
CellDrop Brightfield Cell Counter	DeNovix Inc.
Confocal microscope (TCS SP5)	Leica Microsystems
Custom built microscope	Picoquant
Odyssey imager	LI-COR Biosciences

Table 11: List of consumables

Item	**Manufacturer**
Latex gloves	KimTech
Nitrile gloves	Ansell, VWR
10 µL pipet tips	Gilson, Thermo Fisher Scientific
20 µL pipet tips	Thermo Fisher Scientific
200 µL pipet tips	Gilson, Thermo Fisher Scientific
1000 µL pipet tips	Gilson, Thermo Fisher Scientific
2 mL serological pipet	Corning Inc.
5 mL serological pipet	Corning Inc.
10 mL serological pipet	Corning Inc.
25 mL serological pipet	Corning Inc.
0.2 mL PCR tubes	Starlab Group
1.7 mL reaction tubes	Eppendorf

2.0 mL reaction tubes	Eppendorf
15 mL conical tubes	Greiner
50 mL conical tubes	Greiner, Eppendorf
14 mL culture tubes (round bottom)	Corning Inc.
µ-Slide 8 Well Glass Bottom	Ibidi
6 well dishes	Thermo Fisher Scientific
24 well dishes	Thermo Fisher Scientific
P10 dishes	Greiner
GenePulser® cuvettes	Bio-Rad
Inoculation loops	Sigma Aldrich
Glass beads	Sigma Aldrich
NuPAGE™ 4-12% Bis-Tris Gel	Thermo Fisher Scientific
Trans-Blot Turbo Midi 0.2 µm Nitrocellulose Transfer Packs	Bio-Rad

Table 12: List of software used

Adobe Illustrator 2024
ChemDraw
FIJI
Image Lab V5
Image Studio Lite Ver 5.2
Microsoft Excel
Microsoft Word
SnapGene

2.2 Methods

2.2.1 Molecular cloning

2.2.1.1 Polymerase Chain Reaction (PCR) and purification of the amplified product

To perform PCR, primers were used at a final concentration of 0.25 µM. 1µL of template DNA (stock concentration of 1 – 10 ng/µL) was added per 50 µL of reaction volume. Polymerase-specific buffer (5X) was added to a final concentration of 1X. dNTP (stock concentration 10 mM) was added to a final concentration of 0.2 µM. 0.5 µL of Phusion Hot Start II High-Fidelity DNA-polymerase or 1 µL of Platinum SuperFi II DNA-polymerase was added per 50ul of reaction volume. Water was used to make up the volume. The PCR was performed using a Bio-Rad T100 Thermal Cyler. The PCR cycle consisted of the following steps:

1 Initial denaturation for 45s at 98°C

2 Denaturation for 10s at 98°C

3 Annealing for 30s (temperature varies depending on the primers in use)

4 Elongation at 72°C (duration depends upon the length of the target DNA, speed of the polymerases used is 30s/kb)

5 Steps 2-4 are repeated 35 times

6 Final elongation for 7 mins at 72°C

7 Hold at 10°C

Following PCR, an aliquot of the reaction mixture was run in a 1% agarose gel at 150V for 30mins. The gel was cast with 1% agarose in TAE buffer containing 0.1 µg/mL ethidium bromide. The reaction mixture with a final concentration of 1X DNA loading dye was loaded into the agarose gel. Prior to purification of the PCR products, the PCR reaction mixture was incubated with fast digest DpnI (1 µL / 50 µL reaction volume) restriction enzyme to digest the template DNA.

PCR product was purified either by Quick PCR purification kit from Invitrogen following manufacturer's instructions or by gel extraction. For gel extraction, the PCR product was run in a 0.75% agarose gel at 100V for 1hr or till the desired band was sufficiently separated. The required bands were cut out under the exposure of UV light. The Monarch DNA gel extraction kit from New England Biolabs was used according to the manufacturer's instructions to extract the PCR product from the gel pieces.

DNA concentration was measured using a nanodrop spectrophotometer from DeNovix.

2.2.1.2 Colony PCR

To perform colony PCR necessary primers were used at a final concentration of 1 µM, dNTPs were used at a final concentration of 200 µM, EconoTaq DNA polymerase was used at a final concentration of 0.05 U/µL, and 1X EconoTaq buffer was used. The volume was made upto 50 µL with water for each reaction. Bacteria from a single colony was added per reaction as source of template DNA. The PCR reaction was performed using a Bio-Rad T100 Thermal Cyler. The PCR cycle consisted of the following steps:

1 Initial denaturation for 2 mins at 94°C

2 Denaturation for 30s at 94°C

3 Annealing for 30s at 57°C

4 Elongation at 72°C (duration depends upon the length of the target DNA, speed of EconoTaq polymerase is 1 min/kb)

5 Steps 2-4 are repeated 35 times

6 Final elongation for 15 mins at 72°C

7 Hold at 10°C

Agarose gel electrophoresis was performed with the PCR products. Plasmid extraction from incubated bacterial colony was performed for the colonies that contained the desired plasmid DNA.

2.2.1.3 Cloning strategies

2.2.1.3.1 Gibson assembly

Gibson assembly is an efficient cloning strategy to insert multiple gene fragments in a cloning vector. The inserts need to have an overlap of 20-40 base pairs. Three enzymes are needed for Gibson assembly, namely, an exonuclease that cleaves from the 5' end, a DNA polymerase and a ligase. The exonuclease removes bases from the 5' end of the overlap regions of the inserts allowing the single stranded complementary regions of the inserts to anneal. This is followed by the DNA polymerase that fills resulting gaps and finally the ligase seals any nicks in the double stranded plasmid DNA. In all the clonings performed in this thesis by Gibson assembly an overlap of 40 base pairs was chosen. These overlaps were added to the inserts by PCR. A 2X Gibson assembly mastermix containing all the necessary enzymes was used and made up half of the total reaction volume. A reaction volume of 20 µL was typically used, consisting of 10 µL of Gibson mastermix and the rest of the volume was made up by the vector and inserts. Water was used to make up volume whenever necessary. 100 ng of digested vector was used. A stoichiometric ratio of 1:5 was maintained between the vector and each of the inserts. The reaction mixture was incubated for 1 hr at 50°C followed by transformation into electro-competent TOP10 *E. coli* cells. The transformed cells were incubated for 1 hr at 37°C and 700 rpm. After incubation the cells were plated in appropriate antibiotic containing LB agar plates.

2.2.1.3.2 Mutagenesis

Quick change mutagenesis was used to insert mutations at selected locations in different genes. Partially overlapping primers were designed containing the desired mutation.

2.2.1.3.3 Cloning plasmids with multiple Pyl-tRNA gene copies

The cloning strategy from Rohilla *et al*. was used for cloning multiple Pyl-tRNA copies in a plasmid[178]

2.2.1.3.4 Restriction-ligation cloning

In the case of clonings performed by restriction-ligation, the inserts were amplified by PCR. Recognition sites for selected restriction enzymes were added to the inserts by PCR. Both the inserts and the vector were digested by same restriction enzymes to generate compatible sticky ends which were ligated by T4 ligase. A total reaction volume of 10 µL and 100 ng of digested vector was typically used for every reaction. A stoichiometric ratio of 1:5 of vector and each insert was used. T4 ligase buffer was added to a final concentration of 1X along with 0.5µL T4 ligase. The reaction mixture was incubated for 30 mins at room temperature (RT) followed by transformation in TOP10 *E. coli* cells and plating in appropriate antibiotic containing LB agar plates.

2.2.1.3.5 Annealing oligonucleotides

Cloning with annealing primers was performed to add RNA loops, for example MS2 and BoxB loops to plasmid DNA. Annealing DNA primers were designed to have compatible sticky ends with the corresponding digested vector. The primer annealing reaction was performed with each primer at a final concentration of 5 µM, 1X annealing buffer (table 1) and volume was made up to 100 µL with water. The annealing reaction mixture was incubated at 95°C for 5 mins and then allowed to cool down. The annealed primers were cloned into the respective digested vector following the same ligation procedure as described in section 5.2.1.3.4 (0.2 µL of the annealed reaction mixture was used as insert for ligation).

2.2.1.4 Transformation

Plasmid DNAs were transformed in electro-competent TOP10 *E. coli* cells using the electroporator GenePulser Xcell™ from Bio-Rad. Each transformation consisted of 1 µL plasmid DNA and 50 µL of TOP10 *E. coli* cells. Immediately after electroporation 50 µL of SOC media was added and the cells were incubated for at least 30 mins at 37°C and 700 rpm. Subsequently the cells were plated in LB agar plates with appropriate antibiotics and incubated overnight (ON) at 37°C.

2.2.1.5 Plasmid Isolation

To perform mini- or maxipreps of the plasmid DNAs, a single colony of transformed TOP10 cells was picked from the LB agar plate and incubated ON in appropriate amount of LB medium with suitable antibiotic (0.05 mg/mL) at 37°C and 180 rpm. Plasmid isolation was performed with either PureLink™ Quick plasmid Miniprep kit or PureLink™ HiPure Plasmid Maxiprep kit from Thermo Fisher Scientific following the manufacturer's instructions.

2.2.2 Cell culture

Human embryonic kidney 293T (HEK293T) cells were grown in P10 dishes at 37°C and 5% CO_2 in DMEM++++ (table 7). Cells were split at 80% confluency. 1 million cells were seeded per P10 plates which were split after a growth period of 48 hrs. For splitting (P10 plates), the cells were first washed with 10 mL sterile 1X PBS followed by addition of 1 mL 0.05% trypsin. The plates were incubated at 37°C and 5% CO_2 for 5 mins after addition of trypsin to detach the cells from the surface of the plate. Following trypsinization, 9 ml of DMEM++++ was added to the cells and the concentration of the cell suspension was measured using a DeNovix cell counter. For HeLaK the same procedure was followed except for the media and the number of cells seeded. HeLaK cells were grown in DMEM+++ (table 7) media and 700000 cells were seeded in P10 dishes for splitting after 48 hrs.

2.2.3 Transfection

For transfection with plasmid DNA, cells were seeded in glass-bottom chambered coverslips for imaging experiments or plastic dishes for western blot experiments. 55000 HEK293T cells were seeded per well of an 8 well chambered coverslip. For other culturing plates the number of HEK293T cells seeded was scaled according to the area of each well. In the case of HeLaK cells 35000 cells were seeded per well of an 8 well chambered coverslip. The cells were transfected after 24 hrs of growth. 600 ng of plasmid DNA was

transfected per well of an 8 well chambered coverslip, 6 µg per well of a 6 well cell culture dish and 24 µg per P10 dish. The transfection mix consisted of the plasmid DNA, transfection reagent and the buffer specific to the transfection reagent used. For each well, the volume of the buffer used was 10% of the volume of media added to the well. For example, 250 µL of media is added per well of an 8 well chambered coverslip, hence transfection mix for each well of such a dish would have 25 µL of the transfection reagent specific buffer. Polyethylenimine (PEI) (3µg per µg DNA) was used as the transfection reagent for HEK293T cells and JetPRIME (2 µL per µg DNA) was used for HeLaK cells. For transfecting HEK293T cells, the plasmids and PEI were mixed in DMEM (low glucose, pyruvate, no glutamine, no phenol red), vortexed for 10secs and incubated at room RT for 15 mins before adding to the cells. For transfecting HeLaK cells, the plasmids and JetPrime were mixed in JetPRIME buffer, vortexed for 10s and incubated at room temperature for 10 mins before adding to the cells.

2.2.4 Immunofluorescence

Cells were washed with 1X PBS followed by fixation with 2% paraformaldehyde (PFA). Cells were incubated at RT for 10mins with 2%PFA. After incubation with PFA cells were washed with 1X PBS followed by permeabilization. For permeabilization cells were incubated for 15 mins at RT with 0.5% Triton X-100 in 1X PBS. After permeabilization cells were again washed with 1X PBS twice and blocked for 1hr at RT in 3% BSA in 1X PBS (blocking solution). Following the blocking step cells were incubated ON with primary antibody (diluted in the blocking solution) at 4°C. After incubation with primary antibody, cells were washed twice with 1X PBS and incubated with secondary antibody (diluted in blocking solution) for 1hr at RT. Following incubation with secondary antibody, cells were again washed twice with 1X PBS and subsequently imaged in 1X PBS.

Primary antibody	**Dilution**
anti-myc (mouse)	1:4000
Secondary antibody	**Dilution**
anti-mouse Alexa 488	1:2000

2.2.5 Western Blot

Following 24 hrs of incubation with ncAAs, transfected cells were first washed with 1X PBS. Cells were detached from the respective cell culture dishes by incubating with trypsin at 37°C for 5 mins. The amount of trypsin added was 1/10th of the recommended volume of media per well. For example, for a 6-well cell culture dish the amount of media added per well is 2 mL and amount of trypsin added for detaching the cells is 200 µL. After incubation with trypsin, media was added to reach the recommended volume per well (2 mL for each well of a 6 well cell culture dish) and the detached cells were collected. The cell suspension was centrifuged at 4°C for 4 mins at 400 rcf. The media was discarded, and the cells were resuspended in 1X PBS and centrifuged at 4°C for 4 mins at 400 rcf. The 1X PBS was discarded, and the cell pellets were either lysed immediately or flash frozen with liquid nitrogen and stored at -80°C until further steps.

The cell pellets were thawed on ice. After thawing the pellets were lysed on ice by resuspending them in the lysis buffer for western blot (the details are provided in table 1). Cell pellet obtained from one well of a 6 well cell culture dish was lysed in 60 µl of lysis buffer for 20 mins on ice. Cell pellet obtained from one P10 cell culture dish (10^6 – 10^7 cells) was lysed in 200 µl of lysis buffer for 30 mins on ice. The cell suspensions were mixed by pipetting every 10 mins. After lysis the protein concentration of the lysate was measured using Pierce™ 660 nm protein assay reagent following the manufacturer's instructions. The absorbance was measured with the plate reader (table 10). Subsequently the cell lysates were labelled with 3-(p-Benzylamino)-1,2,4,5-tetrazine - Cy5 (Cy5-H-Tetrazine).

SDS loading buffer (section 5.1) at a final concentration of 1X was added to the labelled cell lysate and incubated for 5 mins at 95°C. SDS polyacrylamide gel electrophoresis (SDS-PAGE) was performed with the cell lysates. An equal amount of total protein (25 µg or 35 µg) was loaded onto the gel for each sample for a particular experiment. Gels were run in MOPS running buffer for 45 mins at 200 V.

Polyacrylamide Gels were blotted on nitrocellulose membranes using the Trans-Blot Turbo Transfer System from Bio-Rad.

The blots were blocked for 1 hr at RT in 5% milk powder in 1X PBS followed by ON incubation with primary antibody (diluted in 5% milk in 1X PBS) at 4°C. After incubation with the primary antibody blots were washed 3 times for 10 mins at RT with 0.2% Tween20 in 1X PBS. Following the washing step blots were incubated in secondary antibody (diluted in 5% milk in 1X PBS) for 1 hr at RT followed by washing 3 times for 10 mins at RT with 0.2% Tween 20 in 1X PBS. The blots were subsequently imaged with the LiCor Odyssey imager at 800 nm to visualize the antibody labelling and imaged with ChemiDoc MP imaging system from Bio-Rad at 629/630nm to visualize the Cy5-H-Tetrazine labelling.

Primary antibody	**Dilution**
anti-flag (mouse)	1:10000
anti-myc (mouse)	1:1000
Secondary antibody	**Dilution**
anti-mouse IRDye 800CW	1:10000

2.2.6 GFP bead purification

After 24 hrs of incubation with ncAAs, transfected cells were detached from the cell culture dishes by incubation with trypsin and cell pellets were subsequently collected as explained in section 2.2.5. The cell pellets were either lysed immediately or flash frozen in liquid nitrogen and stored at -80°C.

Protein purification from HEK293T cells using GFP beads was performed using the GFP-Trap Magnetic Particles M-270 from ChromoTek following the manufacturer's instructions.

The frozen cell pellets were thawed on ice followed by lysis. Pellet from 10^6-10^7 cells were incubated with 200 µL of the lysis buffer (included in the ChromoTek kit) with 1 mM PMSF

for 30mins on ice. The cell suspension was mixed by pipetting every 10 mins. Subsequently 300 µL of dilution buffer (included in the ChromoTek kit) with 1 mM PMSF was added to the cell lysate.

25 µL of GFP bead slurry was used to purify protein from 10^6-10^7 cells. The beads were equilibrated with ice cold wash buffer (included in the ChromoTek kit). 25 µL of GFP bead slurry was resuspended in 500 µL of wash buffer by inverting gently in a 1.5 mL tube. The beads were then separated using magnets and the wash buffer was discarded.

Cell lysates diluted with dilution buffer with 1 mM PMSF were added to the equilibrated beads and rotated end-over-end for 1 hr at 4°C to allow the proteins to bind to the beads. Following the binding step, beads were separated from the suspension with magnets and the supernatant was discarded. The beads were washed twice by resuspending in 500 µL of wash buffer, followed by magnetic separation of the beads and removal of the supernatant.

After washing the proteins bound to the GFP beads were labelled with Cy5-H-Tetrazine by resuspending the beads in 40 µL of dilution buffer with 50 µM Cy5-H-Tetrazine. The beads were incubated for 1 hr at 37°C with the dye solution. Following labelling, the beads were separated by magnets and the supernatant along with the excess dye was discarded.

The labelled proteins were eluted from the beads by suspending the beads in 1X SDS buffer for elution(table 1) and incubating for 5 mins at 95°C. The beads were then separated, and the eluted purified protein was collected. SDS-PAGE and western blot were performed with the purified samples as described in section 2.2.5. For all experiments with GFP bead purified protein, an equal volume of eluted protein was added for each sample in a particular experiment. For quantifications, Cy5 signal for each sample was always normalized to the immunolabel (western blot) signal to account for differences in protein expression between different samples.

2.2.7 Labelling cells with cell impermeable clickable dye for confocal microscopy

Cells were washed twice with 1X TB containing PEG6000 and DTT (table 1). After washing, cells were permeabilized by incubating with digitonin (0.04 mg/mL) for 10 mins at RT. Following permeabilization, cells were washed twice with 1X TB containing PEG6000 and DTT. Permeabilized cells were labelled with LD655-H-Tetrazine (100 nM in 1X TB containing PEG6000 and DTT) by incubating 10 mins at RT. After labelling cells were first washed twice at RT with 1X TB containing PEG6000 and DTT. Subsequently the cells were incubated for 30 mins at 37°C with 1X TB containing PEG6000 and DTT, while exchanging the buffer at 10 mins interval. The 30 min washing step was performed to remove the excess dye.

2.2.8 Labelling cell lysate with clickable dye

Cell lysate was incubated with a final concentration of 10 µM Cy5-H-tetrazine for 1 hr at 37°C.

2.2.9 Sample preparation for submission to IMB proteomics core facility for mass spectrometry analysis

Necessary plasmids were transfected in HEK293T cells as described in section 2.2.3 and grown for 24 hrs in the presence of 250 μM 3IF. The POI was purified by GFP bead purification (section 2.2.6) and SDS-PAGE was performed with the eluted POI. The band corresponding to the POI was cut out and submitted to the IMB proteomics core facility. The POI was digested using trypsin prior to mass spectrometry analysis.

2.2.10 Data analysis

2.2.10.1 Image analysis

Image analysis for confocal microscopy data was performed using the software FIJI (FIJI is just image J). Colocalization analysis was performed using the FIJI plugin JaCop.[179]

For selectivity analysis, first for every imaging frame, the GCE specific click-label (LD655) signal that colocalized with the signal from standard labelling (immunofluorescence or fused fluorescent protein) was extracted. To do so, the image obtained from the standard label channel was duplicated, gaussian filter was applied, followed by an appropriate threshold , and a binary mask was created from this image. In the binary mask all pixels corresponding to signal were given a value of 1 and all other pixels had a value of 0. The binary mask was multiplied to the image from the LD655 channel. The resulting image had only the LD655 signal that colocalized with the mask and from this image the total intensity of the colocalized LD655 was calculated using FIJI. The total LD655 intensity was also calculated for every imaging frame with FIJI. Selectivity was calculated as:

$$\frac{Total\ intensity_{LD655\ signal\ colocalized\ with\ standard\ labelling\ signal}}{Total\ intensity_{total\ LD655\ signal}}$$

2.2.10.2 Analysis for in-gel assay

The immunolabelled protein bands were visualized by scanning the blot at 800 nm using the LiCor Odyssey imager. The band intensities were quantified using Image Studio Lite software.

The Cy5 labelled protein bands were visualized by illuminating the blot at 629/630nm using the ChemiDoc MP imaging system from Bio-Rad at 629/630nm. The band intensities were quantified using the Bio-Rad analysis software Image Lab V5.

2.2.11 Figure generation

Illustrations and data figures were made using Adobe Illustrator and ChemDraw.

2.2.12 Plasmid design

Plasmids were designed with SnapGene.

3 Results

3.1 Site-specific sense codon reassignment in a POI by film-like OTOs

The first aim of this thesis was to enable mRNA selective site-specific sense codon reassignment in mammalian cells. This section is dedicated to discussing the experiments performed to validate successful site-specific sense codon reassignment to incorporate ncAAs selectively in a POI in mammalian cells by using OTOs.

3.1.1 Validation of mRNA-specific sense codon reassignment in mammalian cells by in-gel assay

Similar to classical stop codon suppression, in site-specific sense codon reassignment, the ncAA is incorporated at a predetermined site in the POI. In this case a suitable site is chosen in the POI and the respective gene is engineered to introduce the sense codon, to be reassigned, only at the selected site.

To estimate selectivity of sense codon reassignment by film-like OTOs an in-gel assay was designed with EGFP reporters. An extended EGFP(EGFP^{39X}::EGFP^{66P} , X refers to the ncAA incorporated) and a normal EGFP (EGFP^{39X} , X refers to the ncAA incorporated) were expressed simultaneously in mammalian cells from the same plasmid. Both the extended and the normal EGFP contained the sense codon for reassignment at the permissive site 39. Additionally, the extended EGFP was fused C terminally with an inactive/dark EGFP harbouring the mutation 66P at its chromophore. The logic behind fusion of the dark EGFP was to separate the extended and the normal EGFPs in SDS polyacrylamide gel electrophoresis (SDS-PAGE). The mRNAs of both the reporters were encoded with distinct RNA motifs to allow them to be targeted to OTOs enriched with the corresponding RNA binding domains (RBDs). In this case two pairs of RNA motifs and their corresponding binding partners have been utilized, namely MS2 loops – MS2 bacteriophage coat protein (MCP) and BoxB loops – λN22. On using OTOs enriched with MCP, only mRNA with MS2 loops would be recruited to the OTO and subsequently only the corresponding POI should be modified. Similarly, mRNAs with BoxB loops should be recruited to OTOs enriched with λN22 and desired ncAA should be incorporated only in the corresponding POI (figure 7).

To selectively detect the POI, the cell lysate was labelled with a fluorescent dye having a clickable chemical handle. Such dyes are capable of specifically reacting with ncAAs harbouring a compatible chemical group (section 1:3:2:3). Throughout this work IEDDA (section 1.3.2.3) reaction is used for ncAA-specific labelling of modified POIs. SDS-PAGE was performed with the labelled lysate followed by western blotting. Scanning the blot at a wavelength specific for the click-labelled fluorescent dye should show only the ncAA incorporated modified proteins (figure 8). On the other hand, both the extended and normal EGFPs should be visible on immunolabelling the blot against the epitope tag flag, common to both the POIs, with an antibody conjugated fluorescent dye.

The original EGFP reporter and in-gel assay design was conceived by former members of the lab, Dr. Christopher D. Reinkemeier and [redacted]. Sense codon reassignment with the original EGFP reporters have been shown in the Master thesis of [redacted]. The reporters used in this thesis are updated versions of the original ones. In the original

reporters, RNA motifs were added only to the gene of the extended EGFP reporter. In-gel assay was performed with only OTOs enriched with MCP where the mRNA of only the extended EGFP should be recruited to the OTO and the mRNA of the normal EGFP should remain in the cytoplasm. Subsequently the extended EGFP should be modified with the desired ncAA while the normal EGFP should not be modified. The cell lysate was click-labelled with Cy5-H-Tetrazine, and gel electrophoresis was performed. Cy5 scan of the gel showed only the modified reporter labelled with Cy5 whereas EGFP fluorescence showed both the extended and normal EGFPs. The assay was performed for two sense codons CTA and TCG.

In contrast to the original reporters designed by Dr. Reinkemeier and [redacted] in the reporters used in this thesis the mRNAs of both the extended and normal EGFPs were tagged with distinct RNA motifs that can bind to their cognate RBDs; MS2 loops for extended EGFP and BoxB loops for normal EGFP. In-gel assay was performed using two different OTOs, LCK::MCP::AF::F::AF and LCK::λN22::AF::F::AF (individual components of a fusion construct are separated by ::). LCK refers to the N-terminal domain of rodent LCK tyrosine kinase which targets the OTO to the plasma-membrane, AF refers to the PylRS from *M. mazei* harbouring the mutations Y306A and Y384F and F refers to full length fused in sarcoma (FUS) which plays the role of the assembler. Depending on the OTO in use either the extended or the normal EGFP would be modified with an ncAA. A second set of reporters were also designed for this thesis, where the loops were interchanged, that is BoxB loops were encoded with the extended EGFP gene and MS2 loops were encoded with the normal EGFP gene. Using the two different OTOs and the two sets of reporters the effect of different mRNA recruiting systems and subsequent selectivity of sense codon reassignment were tested and quantified. Also, a set of 8 sense codons were tested for the in-gel assay A depicted in figure 9 and 5 sense codons were tested for the in-gel assay B depicted in figure 10. Unlike the experiments performed in [redacted] Master thesis, in this dissertation instead of EGFP fluorescence western blot analysis with immunolabelling against epitope tag flag fused to both the reporters was performed to detect the total expressed EGFP reporters (both modified and unmodified by GCE).

I supervised a bachelor student [redacted], who worked on site-specific sense codon reassignment. She has used the reporters used in this thesis for her Bachelor thesis.

All the data in this thesis have been acquired by me. Some plasmids have been used from the Lemke lab resources. In section 2.1 table 8 all plasmids generated for this thesis and in table 9 those taken from the lab resources have been listed.

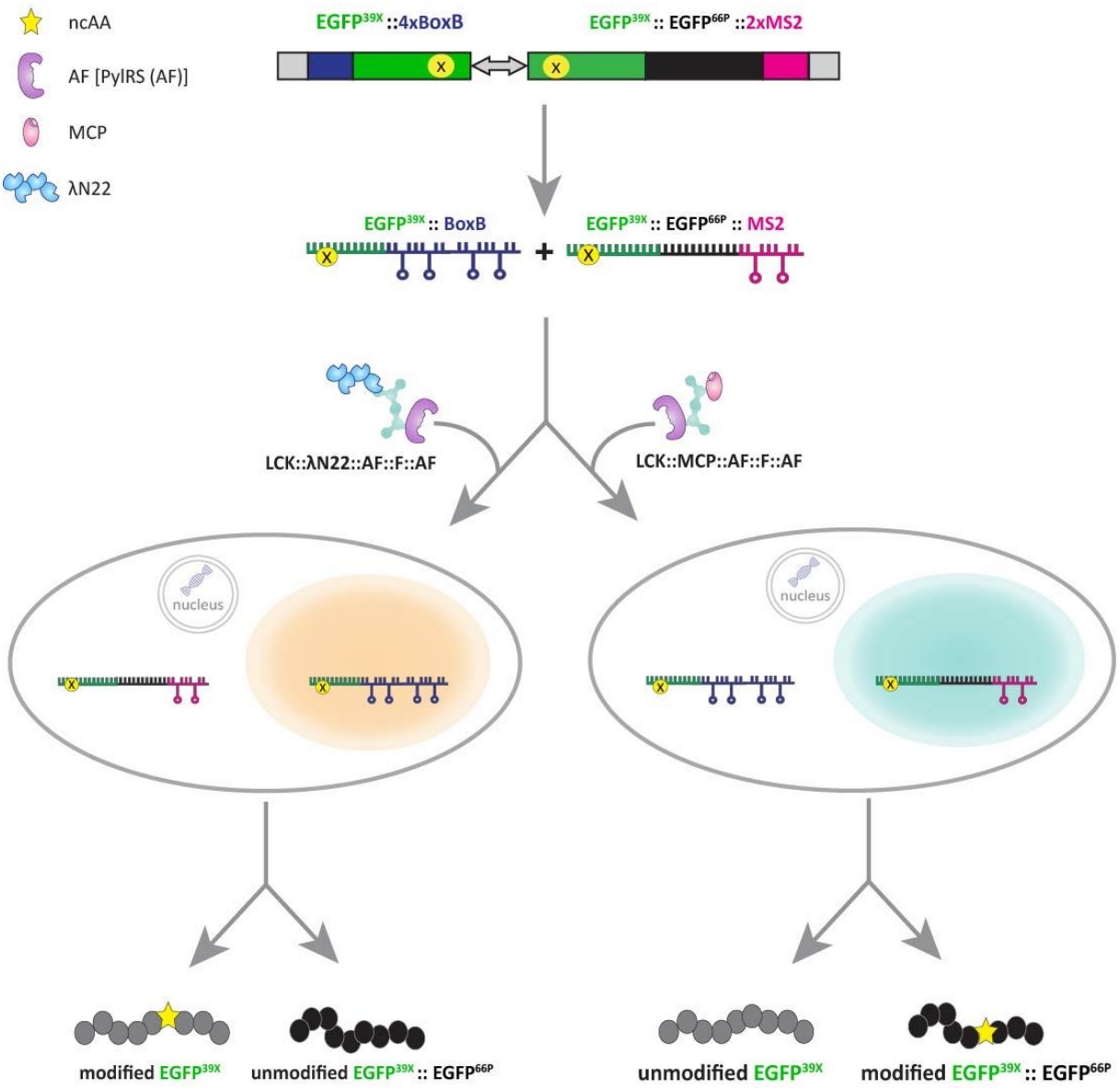


Figure 7| Illustration depicting EGFP reporter design for in-gel validation of mRNA selective site-specific sense codon reassignment. The two reporters, normal EGFP ($EGFP^{39X}$) and extended EGFP ($EGFP^{39X}$::$EGFP^{66P}$) are expressed simultaneously from a single plasmid. Both reporter genes are engineered to introduce a selected sense codon (denoted by a X in a yellow circle) at a position corresponding to site 39 in EGFP. The mRNA of the normal EGFP is tagged with 4xBoxB loops that can interact with the RBD λN22. When an OTO fused with λN22 is used, the mRNA of the normal EGFP is selectively recruited to the OTO and modified to incorporate a desired ncAA. The extended EGFP should not, ideally, be modified in this case. Similarly, the mRNA of the extended EGFP is tagged with 2xMS2 loops that can interact with MCP. On using the OTO with MCP, the mRNA of extended EGFP gets recruited to the OTO selectively and subsequently only extended EGFP should be modified with ncAA.

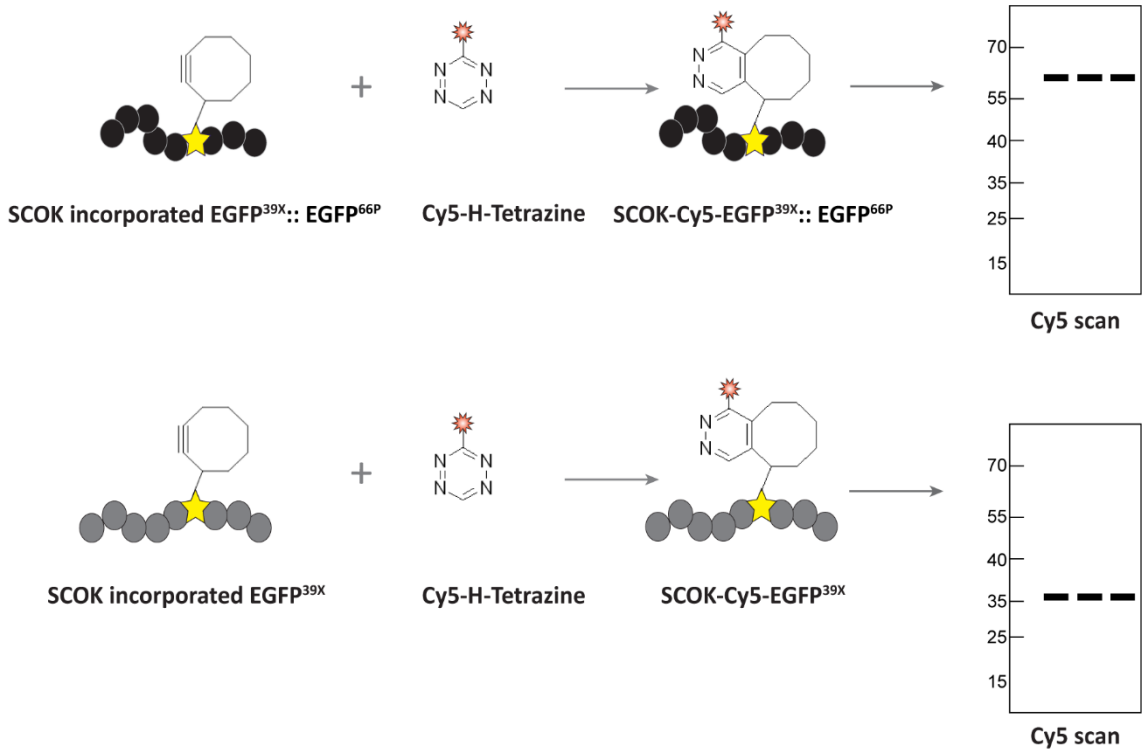


Figure 8| Click-labelling of SCOK incorporated EGFP reporters with Cy5-H-Tetrazine. Lysates of mammalian cells expressing GCE components are labelled with Cy5-H-Tetrazine. Cy5-H-Tetrazine can bio-orthogonally react with the ncAA SCOK by IEDDA reaction. SDS-PAGE can be performed by the labelled cell lysate, followed by western blot. Only the POIs incorporated with SCOK should be labelled with Cy5-H-Tetrazine and be visible in the Cy5 scan of the blot.

3.1.2 Quantifying selectivity of reassignment for multiple sense codons using EGFP reporters by in-gel assay

To determine the selectivity of reassignment, two in-gel assays were performed namely in-gel assay A and in-gel assay B. In-gel assays A and B differ only in the choice of reporters and sense codons selected for reassignment. In in-gel assay A the mRNA of extended EGFP was tagged with MS2 loops and the mRNA of the normal EGFP was tagged with BoxB loops. A set of 8 sense codons, comprising rare (occurrence less than <10/ thousand codons) and abundant codons, was selected for reassignment in in-gel assay A. The selected codons are listed as follows: leucine codons CTA (7.2 occurrence/thousand codons), ATA (7.5 occurrence/thousand codons), TTA (7.7 occurrence/thousand codons); threonine codon ACG (6.1 occurrence/thousand codons); arginine codon CGT (4.5 occurrence/thousand codons); proline codon CCG (6.9 occurrence/thousand codons); serine codon TCG (4.4 occurrence/thousand codons) and cysteine codon TGC (12.6 occurrence/thousand codons).[180] In in-gel assay B the EGFP reporters and RNA binding loops combination was interchanged. Here the mRNA of the extended EGFP was tagged with BoxB loops and mRNA of normal EGFP was tagged with MS2 loops. In contrast to in-gel assay A, in in-gel assay B, the extended EGFP should be modified with an ncAA when an OTO enriched with λN22 is used. On the other hand, the normal EGFP should be modified with an ncAA when an OTO enriched with MCP is used. A set of 5 sense codons namely, CTA, ATA, TGC, CGT and CCG were reassigned in in-gel assay B. Owing to time constraints the reporters corresponding to the codons TCG, ACG and TTA for in-gel assay

B could not be cloned. The overall design of the in-gel assays is explained in detail in section 3.1.1 and a brief outline is also provided in figures 7 and 8.

HEK293T cells were transfected with three plasmids in a 1:1:1 ratio. The three plasmids were: 1) the reporter plasmid expressing both the extended and normal EGFPs; 2) the plasmid coding for Pyl-tRNA with the corresponding anticodon for the sense codon to be reassigned and 3) the construct for cytoplasmic PylRS (NES::AF, NES::AF refers to the PylRS(AF) fused with an NES) or the construct for an OTO. Two different OTOs were used here namely OTO with MCP designated as LCK::MCP::AF::F::AF and OTO with λN22 designated as LCK::λN22::AF::F::AF. The selectivity achieved by the OTO was compared against the non-selective cytoplasmic NES::AF system. An OTO, designated as LCK::AF::F::AF, lacking the mRNA recruiting RBD was also used as one of the controls to demonstrate the necessity of the mRNA recruitment system for successful mRNA specific sense codon reassignment. Either the clickable ncAA cyclooctyne lysine (SCOK) or the non-clickable ncAA (2S)-2-amino-6-{[(benzyloxy)carbonyl]amino}hexanoic acid (CbzK) was incorporated in the POI. The cell lysates were subsequently labelled with Cy5-H-Tetrazine which can only click react with SCOK by IEDDA reaction. Samples where CbzK was incorporated in the POI served as a negative control.

It is important to note that for all sense codons in both in-gel assays A and B (figures 9 and 10) Cy5 labelling of the cell lysate shows significantly higher background incorporation of SCOK in samples with the cytoplasmic NES::AF as compared to those with the OTOs. This serves as a clear demonstration of the success of the OTO system in reducing unspecific modification of the host proteome by GCE.

In in-gel assay A (figure 9) CTA, ATA and CGT codons were amongst the best working codons. For CTA, ATA and CGT codon reassignment a clear Cy5-H-Tetetrazine labelled band was visible corresponding to the POI intended for modification (extended EGFP is marked by yellow arrows and normal EGFP by red arrows in figure 9). As expected, predominantly, the extended EGFP reporter was labelled with Cy5 when the OTO LCK::MCP::AF::F::AF was used and the normal EGFP was labelled on application of the OTO LCK::λN22::AF::F::AF. For the 8 sense codons a Cy5 labelled band corresponding to the normal EGFP was observed in the relevant lanes. However for codons CCG (figure 9C), TTA (figure 9D), TCG (figure 9D), TGC (figure 9E) and ACG (figure 9E) labelled extended EGFP band was barely visible in the lane corresponding to LCK::MCP::AF::F::AF. This is an indication that the BoxB-λN22 RNA recruitment system is more efficient than the MS2-MCP based system. No bands were detected in the lanes corresponding to the non-clickable CbzK, in the Cy5 scan of the blots, as expected. The signal corresponding to immunolabelling against an epitope tag, in this case flag tag (both extended and normal EGFP reporters were N-terminally fused with a flag tag), served as an estimation of the total expression of the two EGFP reporters and was used to normalize the Cy5 signal for the subsequent analysis pipeline.

In in-gel assay B (figure 10), CTA and CGT codons proved to be the best codons for reassignment. In-line with the outcome of in-gel assay A, in in-gel assay B also target mRNA recruitment to the OTO by the interaction of BoxB -λN22 proved to be better than

recruitment by MS2-MCP interaction. In in-gel assay B, for all samples where GCE was performed with the OTO LCK::λN22::AF::F::AF, bands corresponding to the modified POI i.e extended EGFP (for in-gel assay B the mRNA of the extended EGFP was tagged with BoxB loops and hence should be selectively recruited to OTOs with λN22) was observed in the Cy5 scan of the blots. For CCG and TGC codons band corresponding to the normal EGFP for samples where GCE was performed with the OTO LCK::MCP::AF::F::AF was either not detected or very faint, respectively. Bands corresponding to reporter POIs were also either very faint or not detected for samples with the OTO LCK::AF::F::AF which lacked the RBD necessary for recruiting the target mRNA into the OTO. Thus, demonstrating that the mRNA recruitment system is necessary for selective sense codon reassignment by the OTOs. No Cy5 labelled bands were observed for samples where the POI was modified with the non-clickable ncAA CbzK. Similar to in-gel assay A, in in-gel assay B also both extended and normal EGFPs were immunolabelled against the N-terminally fused flag tag. Hence both the POIs were visible in the immunolabel scan of the blots (labelled WB in figure 10).

To quantify selectivity, first Cy5 signals from the reporter bands were divided by the corresponding signal from immunolabelling to account for the difference in expression levels between different samples. Let this ratio be denoted as 'A'. The second step was to calculate, for each case, the ratio of 'A' for the POI intended for modification and 'A' for the POI that was not meant to be modified. For example in the samples where LCK::MCP::AF::F::AF was used, the extended EGFP was ideally supposed to be modified in in-gel assay A. Hence the ratio to be calculated was ($A_{\text{EXTENDED EGFP}}$ / $A_{\text{NORMAL EGFP}}$). Similarly for the samples with LCK::λN22::AF::F::AF, it was the normal EGFP that should be modified in in-gel assay A. Hence in this case the ratio to be calculated was ($A_{\text{NORMAL EGFP}}$ / $A_{\text{EXTENDED EGFP}}$). Let this ratio be generally referred to as ($A_{\text{POI INTENDED FOR NCAA INCORPORATION}}$ / $A_{\text{POI NOT INTENDED FOR NCAA INCORPORATION}}$) and denoted as 'B'. 'B' is an estimation of how selective the sense codon reassignment is since it takes into account the extent of unspecific modifications. Since the cytoplasmic system is not selective, the fold change in the value 'B' for site-specific sense codon reassignment with different OTOs as compared to site-specific sense codon reassignment with the cytoplasmic NES::AF is a reliable measure of the selectivity achieved for sense codon reassignment using the OTOs. In figures 11 (for in-gel assay A) and 12 (for in-gel assay B) the (B_{OTO} / $B_{\text{NES::AF}}$) defined as fold change selectivity is plotted for different codons.

For most cases for in-gel assay A (figure 11) a higher fold change selectivity was observed for site-specific sense codon reassignment by the OTO LCK::λN22::AF::F::AF, reaching a maximum of about 14fold for the CTA codon, followed by the ATA codon with about 9fold. In case of the OTO LCK::MCP::AF::F::AF the highest fold change selectivity of about 10fold was observed for both CTA and CGT codons. The error bars in figure 11 represent standard deviation of data obtained from two biological replicates of in-gel assay A. The blots shown in figure 11 are representative of two biological replicates of in-gel assay A. For in-gel assay B (figure 12) fold change selectivity for codons ATA, CGT, CCG and TGC were higher when LCK::λN22::AF::F::AF was used for sense codon reassignment as compared to when LCK::MCP::AF::F::AF was. For samples with the OTO LCK::λN22::AF::F::AF,

highest fold change selectivity of about 8fold was recorded for the site-specific reassignment of the CGT codon, followed by about 6fold for the site-specific reassignment of the CTA codon. In the case of LCK::MCP::AF::F::AF , highest fold change selectivity of about 8fold was recorded for the CTA codon, followed by about 6fold for CGT codon. The result of in-gel assay B was obtained from a single biological replicate.

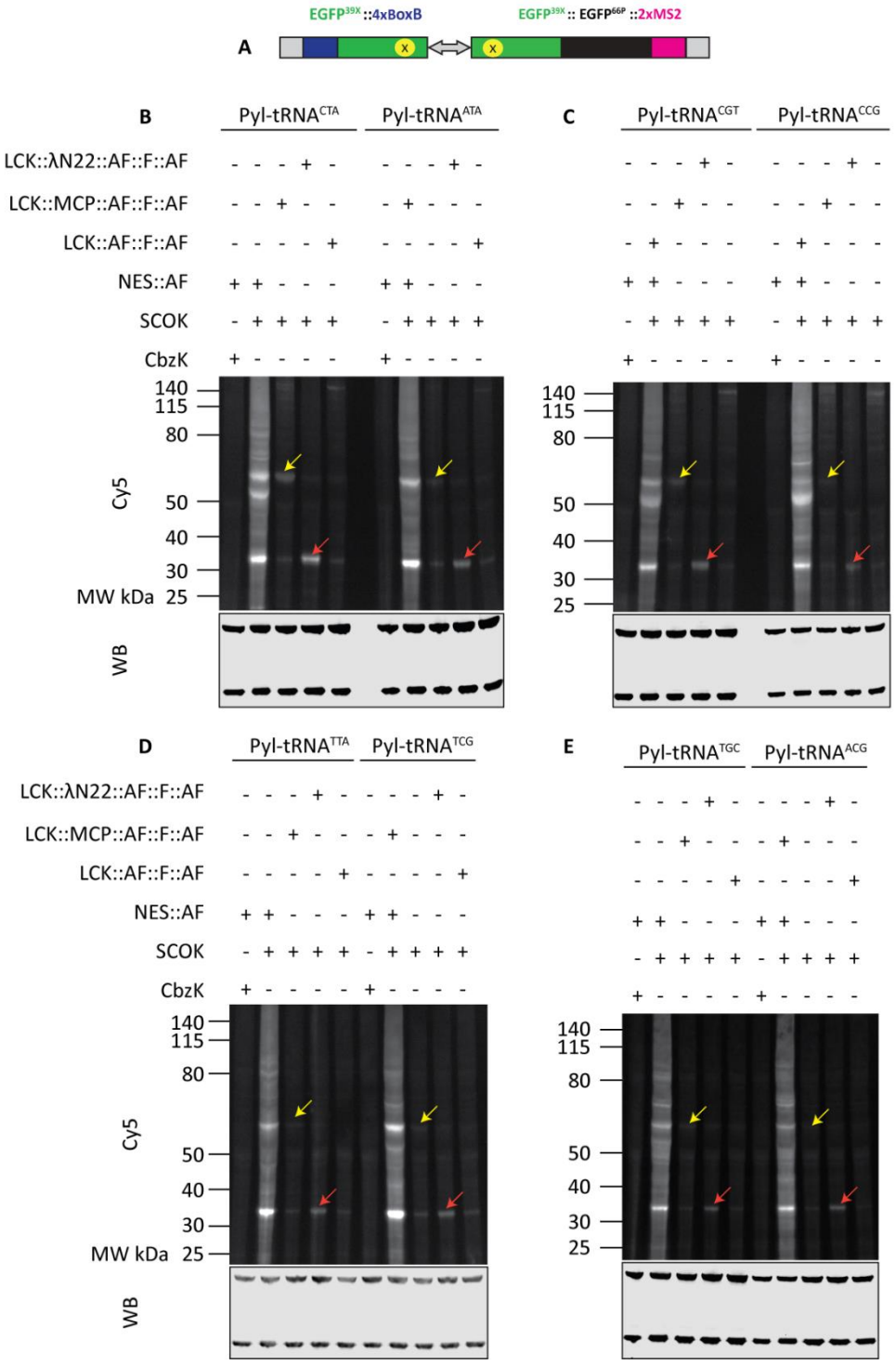


Figure 9| In-gel assay A for validating sense codon reassignment in HEK293T cells. Bands corresponding to the extended EGFP are marked with yellow arrows and those representing the normal EGFP are marked with red arrows. (XXX) in Pyl-tRNA^{XXX} denotes the codon that has been reassigned. Cy5 scan of the blots are marked as "Cy5" and western blots or immunolabel scans of the blots are marked as "WB. (A) EGFP reporters used for in-gel assay A. (B),(C),(D) and (E) respectively show the Cy5 and immunolabel scan of the blots for reassigning CTA and ATA codons; CGT and CCG codons; TTA and TCG codons; TGC and ACG codons respectively . In in-gel A assay the extended EGFP mRNA was tagged with MS2 loops hence in samples where GCE was performed with LCK::MCP::AF::F::AF, the extended EGFP should be selectively modified and visible in the Cy5 scan of the blot. On the other hand the mRNA of the normal EGFP was tagged with BoxB loops and hence on using the OTO LCK::λN22::AF::F::AF, the normal EGFP should be selectively modified and visible in the Cy5 scan of the blots. Samples where GCE was performed with the non-selective, cytoplasmic NES::AF was included to compare the selectivity of sense codon reassignment achieved by the OTOs. Samples with the OTO LCK::AF::F::AF (lacks the RBD necessary to selectively recruit target mRNA into the OTO) were also included to demonstrate the necessity of the RBD-RNA loops interaction for performing selective GCE by the OTOs. Only reporters incorporated with SCOK can be click labelled with the compatible dye Cy5-H-Tetrazine. Non-clickable CbzK was used as a negative control and no bands should be visible in the Cy5 scan for the samples with CbzK.

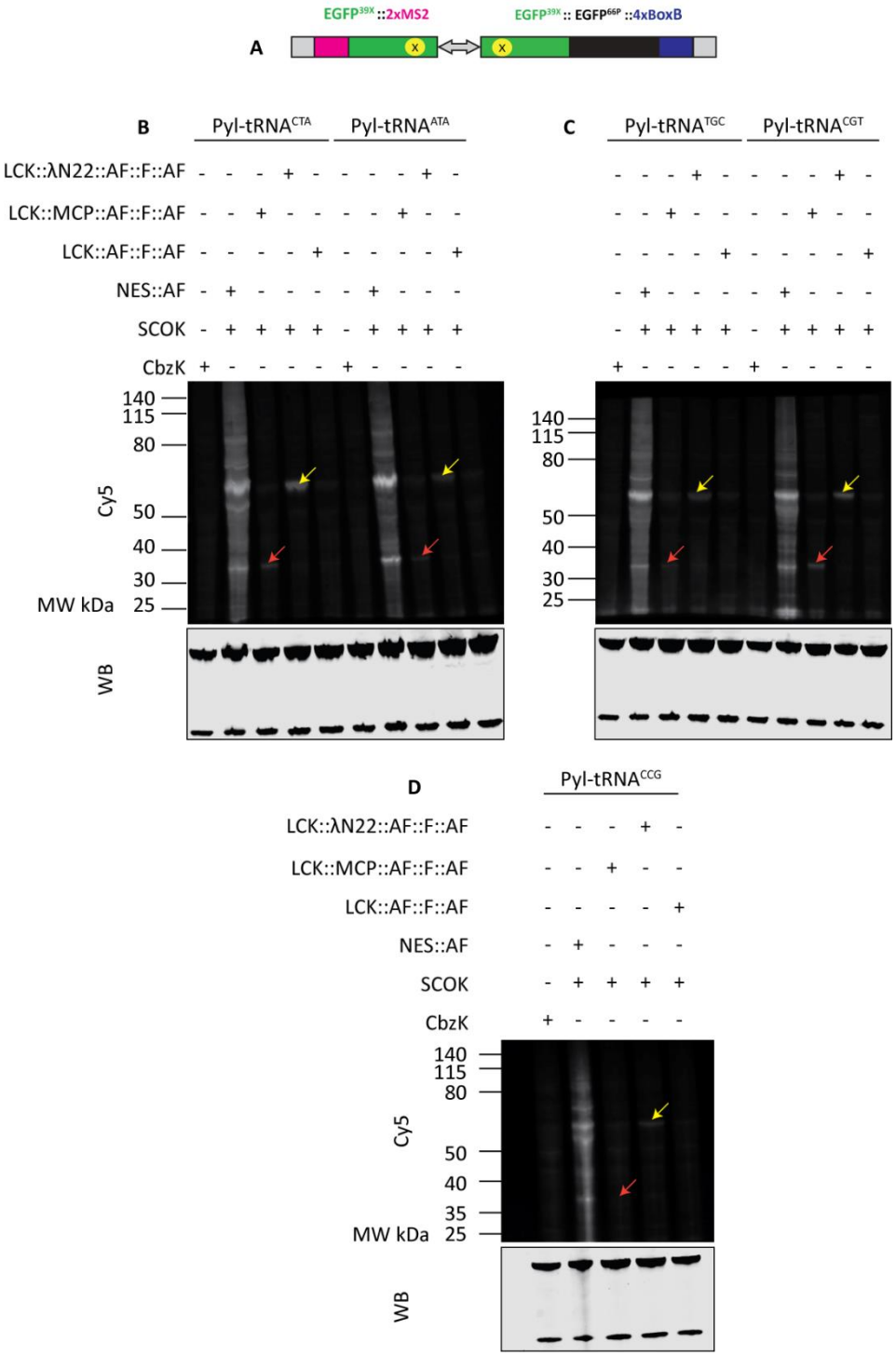


Figure 10| In-gel assay B for validating sense codon reassignment in HEK293T cells. Bands corresponding to the extended EGFP are marked with yellow arrows and those representing the normal EGFP are marked with red arrows. (XXX) in Pyl-tRNA^{XXX} denotes the codon that has been reassigned. Cy5 scan of the blots are marked as "Cy5" and western blots or immunolabel scans of the blots are marked as "WB".(A) EGFP reporters used for in-gel assay B.(B), (C) and (D) respectively show the Cy5 and immunolabel scans of the blots for reassigning CTA and ATA codons; TGC and CGT codons; and CCG codon respectively.In in-gel assay B the normal EGFP mRNA was tagged with MS2 loops hence in samples where GCE was performed with LCK::MCP::AF::F::AF, the normal EGFP should be selectively modified and visible in the Cy5 scan of the blot. On the other hand the mRNA of the extended EGFP was tagged with BoxB loops and hence on using the OTO LCK::λN22::AF::F::AF, the extended EGFP should be selectively modified and visible in the Cy5 scan of the blots. Samples where GCE was performed with the non-selective, cytoplasmic NES::AF was included to compare the selectivity of sense codon reassignment achieved by the OTOs. Samples with the OTO LCK::AF::F::AF (lacks the RBD necessary to selectively recruit target mRNA into the OTO) were also included to demonstrate the necessity of the RBD-RNA loops interaction for performing selective GCE by the OTOs. Only reporters incorporated with SCOK can be click labelled with the compatible dye Cy5-H-Tetrazine. Non-clickable CbzK was used as a negative control and no bands should be visible in the Cy5 scan for the samples where CbzK.was incorporated in the POI.

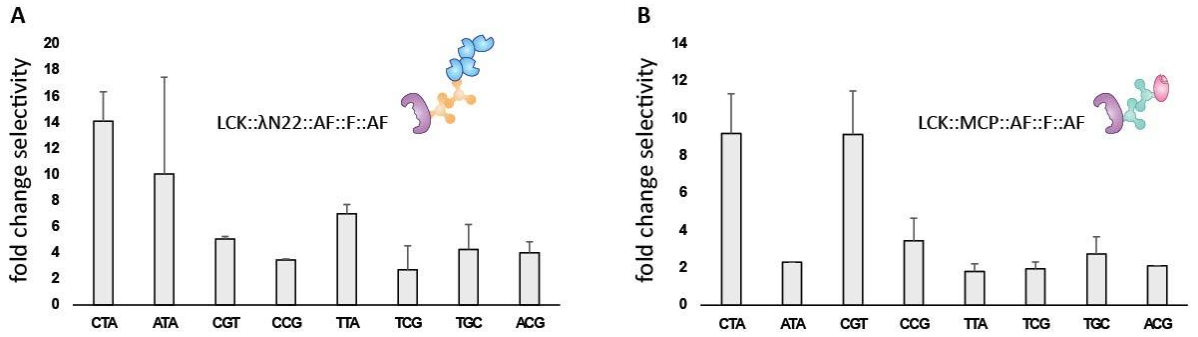


Figure 11| Fold change selectivity (B_{OTO} / $B_{NES::AF}$) of sense codon reassignment determined from in-gel assay A for 8 selected codons. (A) shows the results for sense codon reassignment with the OTO LCK::λN22::AF::F::AF and (B) shows the results for sense codon reassignment with OTO LCK::MCP::AF::F::AF. Detailed explanation of the analysis is provided in section 3.1.2.

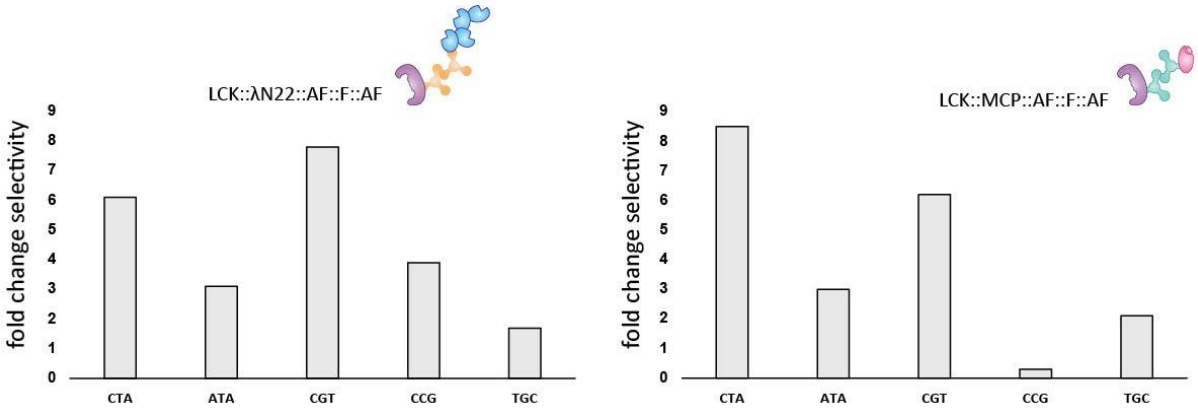


Figure 12| Fold change selectivity (B_{OTO} / $B_{NES::AF}$) of sense codon reassignment determined from in-gel assay B for 5 selected codons. (A) shows the results for sense codon reassignment with OTO LCK::λN22::AF::F::AF and (B) shows the results for sense codon reassignment with OTO LCK::MCP::AF::F::AF. Detailed explanation of the analysis is provided in section 3.1.2.

3.1.3 Site-specific sense codon reassignment enabled fluorescent labelling of vimentin::mcerulean for confocal microscopy

A prime motivation for reassigning sense codons to ncAAs is to enable incorporation of distinct probes site-specifically in a POI, including fluorescent dyes for imaging applications. To test the applicability of site-specific sense codon reassignment to introduce fluorescent dyes in a POI for visualization by confocal microscopy, vimentin fused C-terminally with a fluorescent protein mcerulean (vimentin::mcerulean) was used as a reporter. The mRNA of vimentin::mcerulean was tagged with MS2 loops to allow selective recruitment into MCP-based OTOs. A CTA codon at site 164 in vimentin was reassigned to incorporate clickable ncAA *trans*-cyclooct-2-en-L lysine (TCO*AK) / non-clickable ncAA t-butyloxycarbonyl-l-lysine (BOCK) using the OTO LCK::MCP::AF::F::AF. The CTA codon did not occur at any other site of the reporter construct. The experiment was performed in HEK293T cells, and the cells were labelled with clickable synthetic dye LD655-H-Tetrazine. LD655-H-Tetrazine can specifically react with TCO*AK by IEDDA reaction. NES::AF was used to demonstrate the non-selectivity of the cytoplasmic system and LCK::AF::F::AF (OTO lacking the mRNA recruiting RBD) was used to highlight the

significance of the selective mRNA recruitment machinery for mRNA specific sense codon reassignment. The signal from the mcerulean channel served as a standard labelling method for reference. Colocalization of the GCE dependent click-label LD655 channel with the mcerulean channel is a positive indication that the POI was site-specifically labelled by sense codon reassignment.

In figure 13 it can be seen that the sample with LCK::MCP::AF::F::AF shows colocalization between the mcerulean and the LD655 channels (M2-0.46, M2 is Mander's coeffiecient 2. M2 signifies the fraction of the LD655 image that colocalizes with the mcerulean image. M2 was calculated using the Jacop plugin in FIJI. The average M2 value obtained from 4 different imaging frames has been reported here. An M2 value of 1 signifies complete colocalization.) A significant amount of non-colocalized unspecific labelling can be observed in the LD655 channel corresponding to the NES::AF sample, as predicted. No LD655 signal was expected in the non-clickable BOCK sample and the same was observed in figure 13. A primarily non-colocalized, low intensity signal was detected for the samples with the OTO LCK::AF::F::AF demonstrating the necessity of the mRNA recruitment machinery for mRNA specific sense codon reassignment.

To quantify the selectivity of CTA reassignment, a ratio of the LD655 signal (total intensity) colocalized with mcerulean signal and total LD655 signal (total intensity) was calculated for all samples. Hence selectivity here is defined as:

$$\left[\frac{Total\ intensity_{LD655\ signal\ colocalized\ with\ mcerulean\ signal}}{Total\ intensity_{Total\ LD655\ signal}}\right]_{OTO\ or\ NES::AF}$$

Fold change selectivity for sense codon reassignment with LCK::MCP::AF::F::AF or LCK::AF::F::AF compared to sense codon reassignment with the non-selective NES::AF sample is depicted as a bar plot in figure 13. Each bar represents the value:

$$\frac{Selectivity_{OTO}}{Selectivity_{NES::AF}}$$

The fold change selectivity value represented by each bar in figure 13 is the mean fold change selectivity obtained from two biological replicates. The error bars represent the standard deviation. For each sample of each biological replicate, mean selectivity value was calculated from at least 3 imaging frames. A fold change selectivity value higher than 1 would signify selective sense codon reassignment by the OTO. The plasma-membrane localized OTO LCK::MCP::AF::F::AF demonstrated a 1.4fold change selectivity as compared to NES::AF, whereas that recorded for the OTO lacking RBD for selective mRNA recruitment, LCK::AF::F::AF was 0.5fold.

The incorporation of ncAA by mRNA selective site-specific sense codon reassignment was also validated by mass spectrometry. The CGT codon was reassigned in $EGFP^{39CGT}$ to incorporate 3-iodo-phenylalanine (3IF) by LCK::MCP::AA::F::AA (AA refers to PylRS(AA) with the mutations C346A, N348A) in HEK293T cells. $EGFP^{39CGT}$ was purified by GFP beads and submitted for mass spectrometry analysis. It was possible to detect $EGFP^{39CGT}$ fragment (mass spectrometry measurements were performed after digesting $EGFP^{39CGT}$ with trypsin) with 3IF incorporated at position 39 (supplementary figure 4).

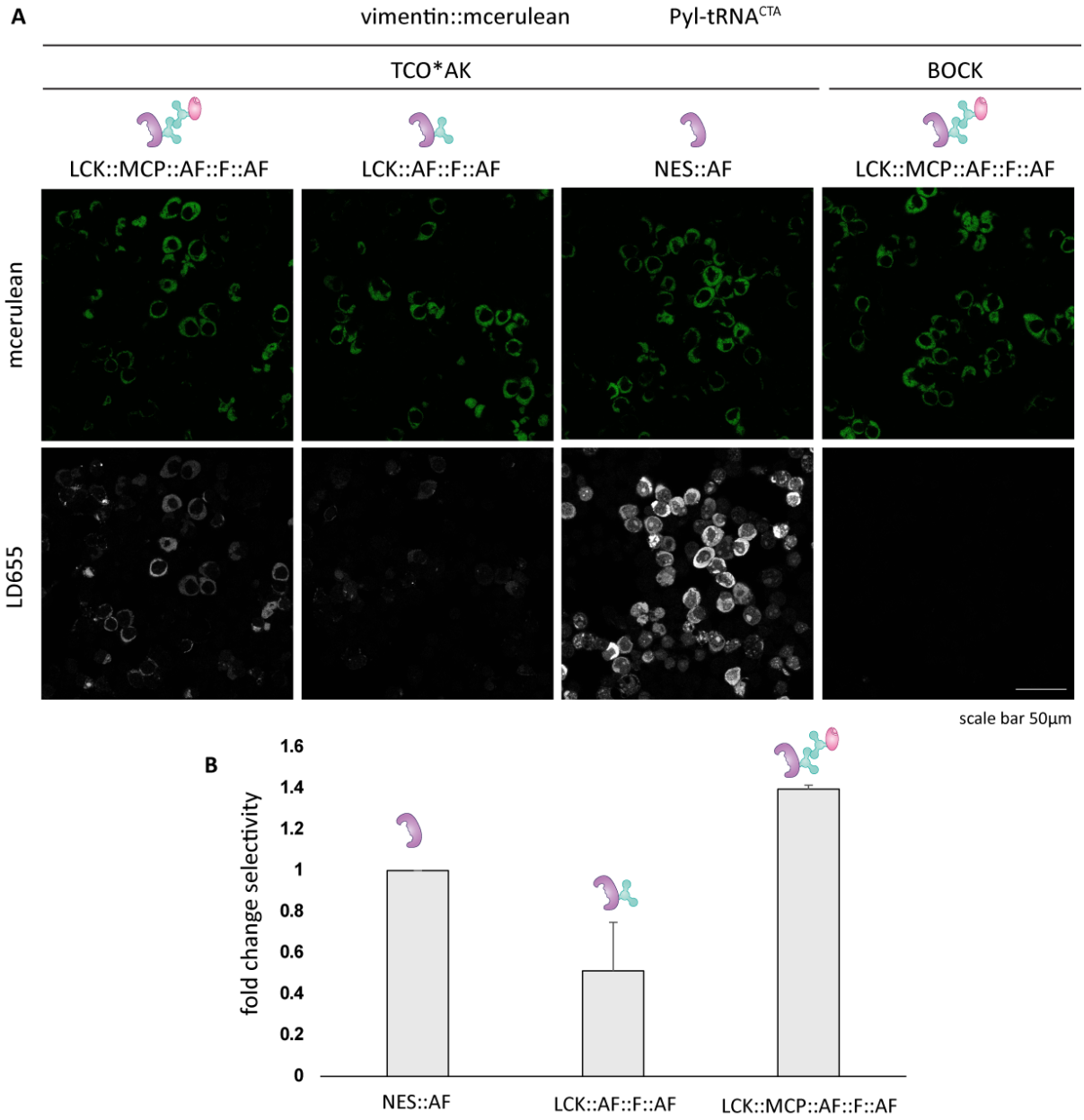


Figure 13| Confocal images of vimentin::mcerulean click labelled with LD655-H-Tetrazine.(A) the upper row shows the mcerulean signal and the lower row shows the LD655 signal. The codon CTA in vimentin::mcerulean was reassigned to incorporate clickable ncAA TCO*AK (first 3 columns) and non-clickable BOCK (last column). (B) Bar plot depicting quantification of selectivity of CTA reassignment. Detailed explanation of the analysis is provided in section 3.1.3. The error bars represent the standard deviation of the fold change value obtained from different imaging frames of a single replicate.

3.2 Residue-specific sense codon reassignment

The second objective of this thesis was to perform mRNA selective residue-specific sense codon reassignment in mammalian cells to incorporate ncAAs in a POI. In the following sections first the concept of mRNA selective residue-specific sense codon reassignment will be explained along with the assay designed for selecting suitable codons for reassignment in reporter POIs and the corresponding result of codon selection (section 3.2.1). In the subsequent sections results will be described validating OTO enabled mRNA selective residue-specific sense codon reassignment in mammalian cells for selected sense codons (sections 3.2.2-3.2.7).

3.2.1 Selection of suitable sense codons for reassignment in reporters by confocal microscopy-based screening

In case of residue-specific sense codon reassignment a specific sense codon (XXX) is chosen which occurs naturally in the gene of the POI and is repurposed to incorporate an ncAA. Unlike site-specific sense codon reassignment, for residue-specific reassignment no mutation is necessary in the gene of the POI. Ideally all XXX codons present in the gene of interest are aimed to be repurposed. However, depending on the efficiency of residue-specific sense codon reassignment, GCE is likely to result in a mixed population consisting of POIs with varying number of XXX codons reassigned (figure 14). This strategy has found widespread use in proteomics studies as summarized in section 1.4.2.3, however its potential as a method for incorporating multiple probes exclusively into a POI has remained untapped as mentioned.

Adding mRNA specificity with the help of OTOs to residue-specific sense codon reassignment (figure 14) can facilitate a myriad of scientific possibilities (sections 4.9.2 and 4.9.3). To validate the effectiveness of OTOs for selective residue-specific sense codon reassignment, appropriate reporters are necessary along with suitable codons for reassignment. To identify sense codons suitable for reassignment, for each reporter all 61 sense codons were individually reassigned to incorporate a clickable ncAA in mammalian cells. The cells were subsequently labelled with a compatible fluorescent dye capable of bio-orthogonally reacting with the incorporated ncAA. Immunofluorescence against an epitope tag fused to the POI or a fluorescent protein (EGFP) fused to the POI served as the standard labelling technique. Colocalization between the confocal microscopy images obtained by exciting the click-labelled dye with the images obtained from the standard labelling technique would mean that the POI could be successfully labelled by GCE.

HEK293T cells were transiently transfected with three plasmids (in 1:1:1 ratio) expressing the reporter, an OTO / NES::AF and the Pyl-tRNA^{XXX} (XXX – sense codon reassigned) respectively. The reporters chosen for this experiment to select suitable sense codons for reassignment were the fusion constructs vimentin::mcerulean and nucleoporin (NUP)153::EGFP. The plasma membrane localized OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate clickable ncAA TCO*AK/ non-clickable ncAA BOCK in both the reporters. The mRNAs of both the reporters were tagged with MS2 loops. Cells were labelled with LD655-H-Tetrazine since it can bio-orthogonally react with TCO*AK by IEDDA reaction. Cells where BOCK was incorporated in the reporters should not show any

signal from LD655 since BOCK does not react with LD655-H-Tetrazine. Non-clickable BOCK was used as a negative control for both the reporters to determine the extent of unspecific interaction of the dye with cellular components. In case of vimentin::mcerulean immunolabelling against N-terminally fused epitope tag myc with Alexa 488 conjugated antibody served as the standard label whereas for NUP153::EGFP, the EGFP signal was recorded as the standard label. Two additional controls were included for NUP153::EGFP namely samples where GCE was performed by the OTO LCK::AF::F::AF (OTO lacking mRNA-recruiting RBD MCP) and samples where GCE was performed by the non-selective cytoplasmic NES::AF. Cells were subsequently imaged using a confocal microscope.

Figures 15-30 depict the results for the confocal microscopy experiments for codon selection in vimentin::mcerulean. Images corresponding to some sense codons showed colocalization to varied extents between immunofluorescence (anti-myc Alexa 488) and LD655 channels whereas for the rest of the sense codons non-colocalized, diffused signal was observed in LD655 channel. Some of the sense codons showing promising results include GAT (M2– 0.53), ACC (M2- 0.59), ATC (M2-0.62). M2 refers to Mander's coefficient 2 which quantifies the fraction of the LD655 image that overlaps with the corresponding immunofluorescence image. A M2 value of 1 signifies complete colocalization. M2 has been calculated using the JACoP plugin in FIJI.[179] The value mentioned for each codon is an average calculated from 3 imaging frames. The LD655 channels corresponding to the samples where BOCK was incorporated in the POI showed no signal in all cases, as expected. A selectivity analysis was also performed for the 61 codons (Figure 31). Selectivity is defined as:

$$\left[\frac{Total\ intensity_{LD655\ signal\ colocalized\ with\ signal\ from\ immunofluorescence}}{Total\ intensity_{total\ LD655\ signal}}\right]_{XXX}$$

XXX refers to the sense codon reassigned. The value represented by each bar is the average of at least 3 imaging frames. A selectivity value of 1 would be the most desired outcome. Selectivity value 1 would mean the images obtained from the immunofluorescence and LD655 channels completely colocalize and there is no unspecific background LD655 labelling, thus signifying that vimentin::mcerulean could be selectively labelled by GCE.

Figure 32 and appendix I figures 1-24 depict a subset of the confocal microscopy experiment results for NUP153::EGFP. The codon selection with NUP153::EGFP was not as successful as with vimentin::mcerulean. Only in very few cases few cells with colocalized nuclear rims between EGFP (standard label) and LD655 (GCE-specific label) channels for samples where GCE was performed by the OTO LCK::MCP::AF::F::AF could be observed (appendix I figure 1 (TCT), appendix I figure 3 (AGC), appendix I figure 12 (GCC)). No signal was recorded in the LD655 channel of the BOCK samples as expected.

The differing outcomes for the two reporters highlight the need for screening all sense codons for every POI, which is quite a laborious process. Vimentin::mcerulean was chosen for further experiments due to the promising outcome of the codon selection process.

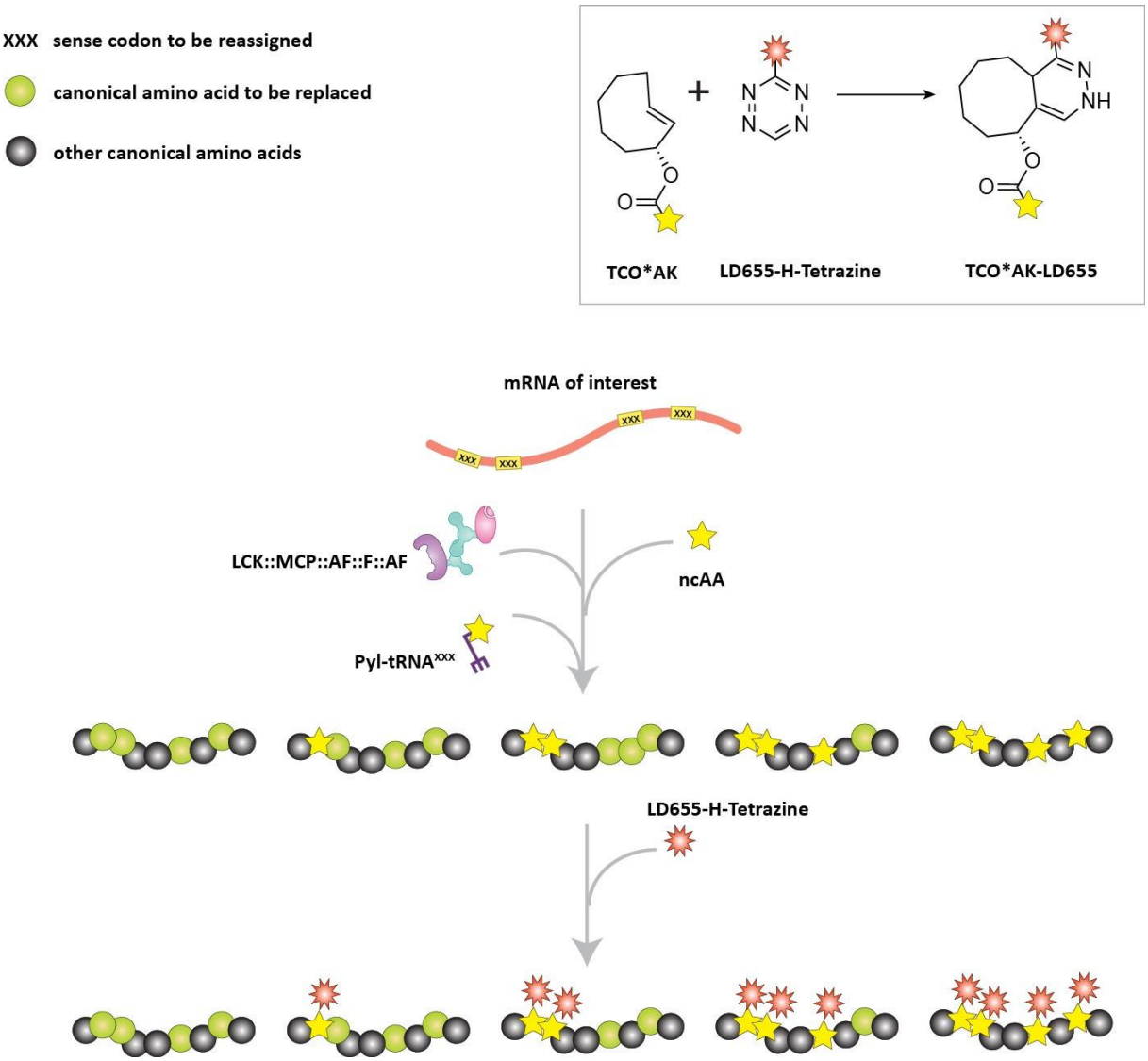


Figure 14| Illustration depicting mRNA selective residue-specific sense codon reassignment and the subsequent click-labelling of the fluorescent dye to the ncAA. In residue-specific sense codon reassignment, a particular codon XXX which occurs naturally in the mRNA of interest is reassigned by the GCE machinery to incorporate desired ncAAs. OTOs are used to render mRNA selectivity to residue-specific sense codon reassignment. The mRNA of the POI is tagged with RBD binding RNA loops to facilitate recruitment into the OTOs. Depending on the efficiency of the process, either all XXX codons occurring in the mRNA will be reassigned to incorporate equal number of ncAAs or none of the codons will be reassigning resulting in the expression of the wild type POI or a mixed product will be generated containing POIs with differing number of incorporated ncAAs. On incorporation of clickable ncAAs like TCO*AK in the POI, equal number of compatible synthetic dyes can be introduced which can react bio-orthogonally with TCO*AK by IEDDA reaction. The click reaction between TCO*AK and the dye LD655-H-Tetrazine with the compatible chemical handle tetrazine is shown in the inset.

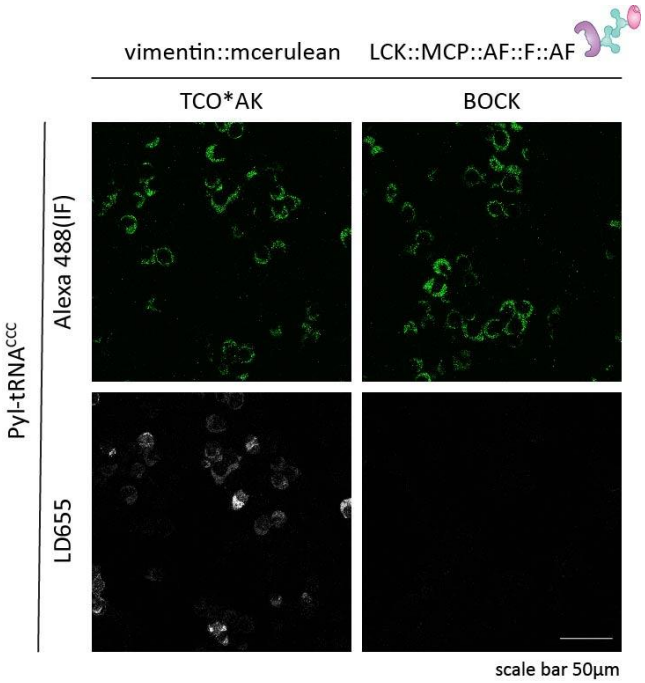


Figure 15| Confocal imaging results for codon selection for vimentin::mcerulean (codon CCC). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. Either the clickable ncAA TCO*AK or the non-clickable ncAA BOCK was incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning CCC codon in vimentin::mcerulean.

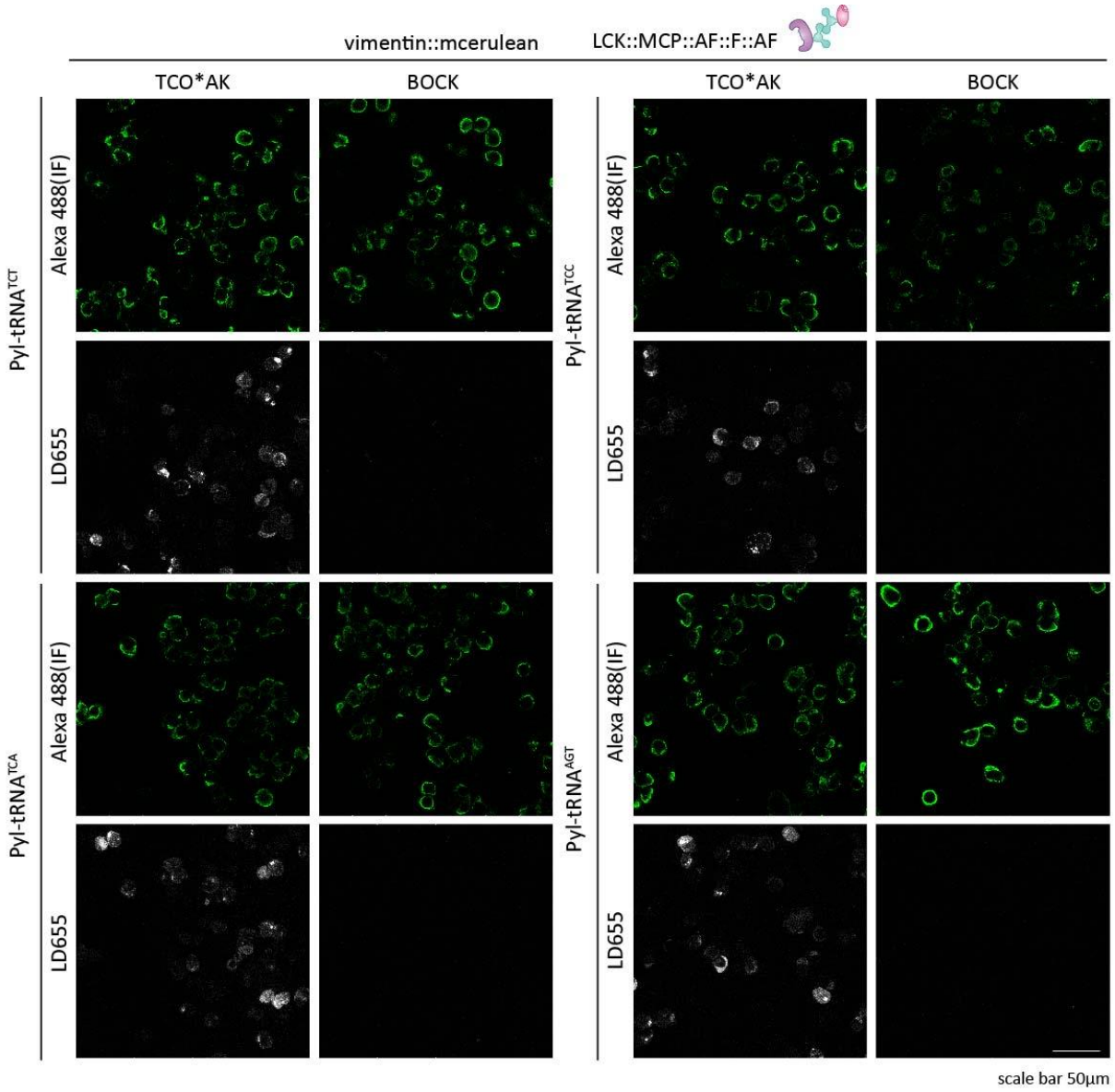


Figure 16| Confocal imaging results for codon selection for vimentin::mcerulean (codons TCT, TCC, TCA, AGT). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons TCT, TCC, TCA and AGT in vimentin::mcerulean.

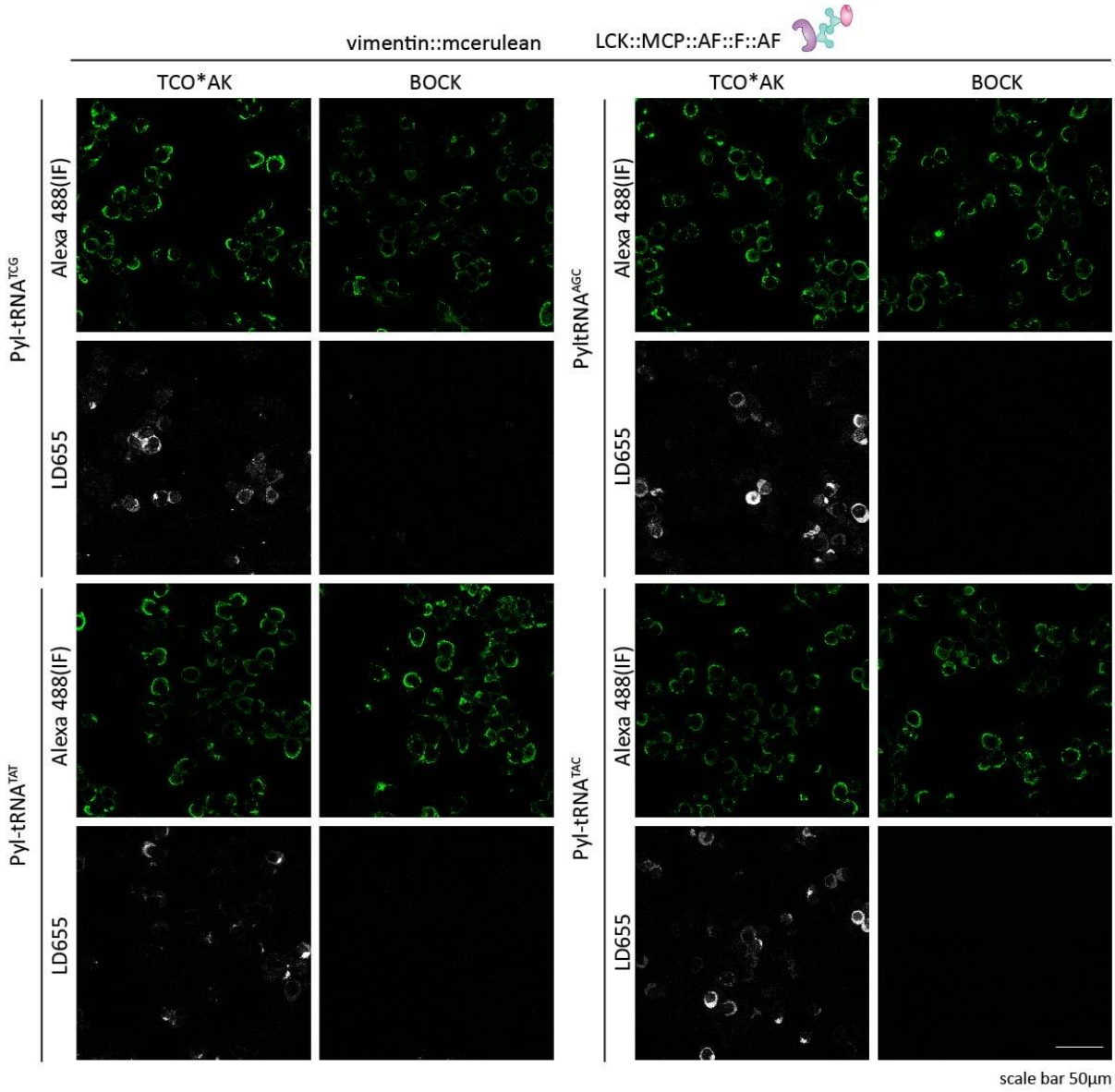


Figure 17| Confocal imaging results for codon selection for vimentin::mcerulean(codons TCG, ACG, TAT, TAC). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons TCG, ACG, TAT and TAC in vimentin::mcerulean.

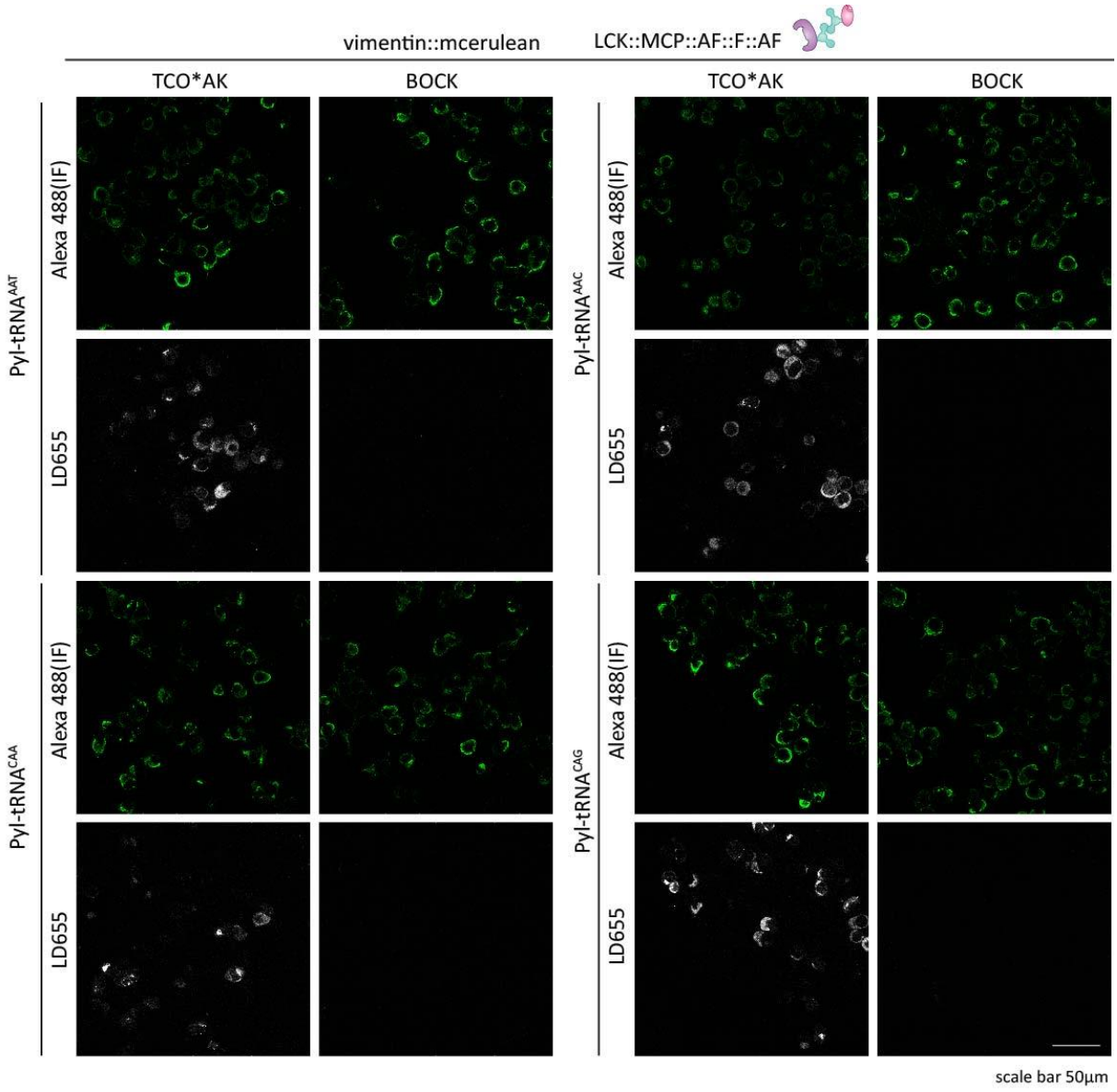


Figure 18| Confocal imaging results for codon selection for vimentin::mcerulean (codons AAT, AAC, CAA, CAG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons AAT, AAC, CAA and CAG in vimentin::mcerulean.

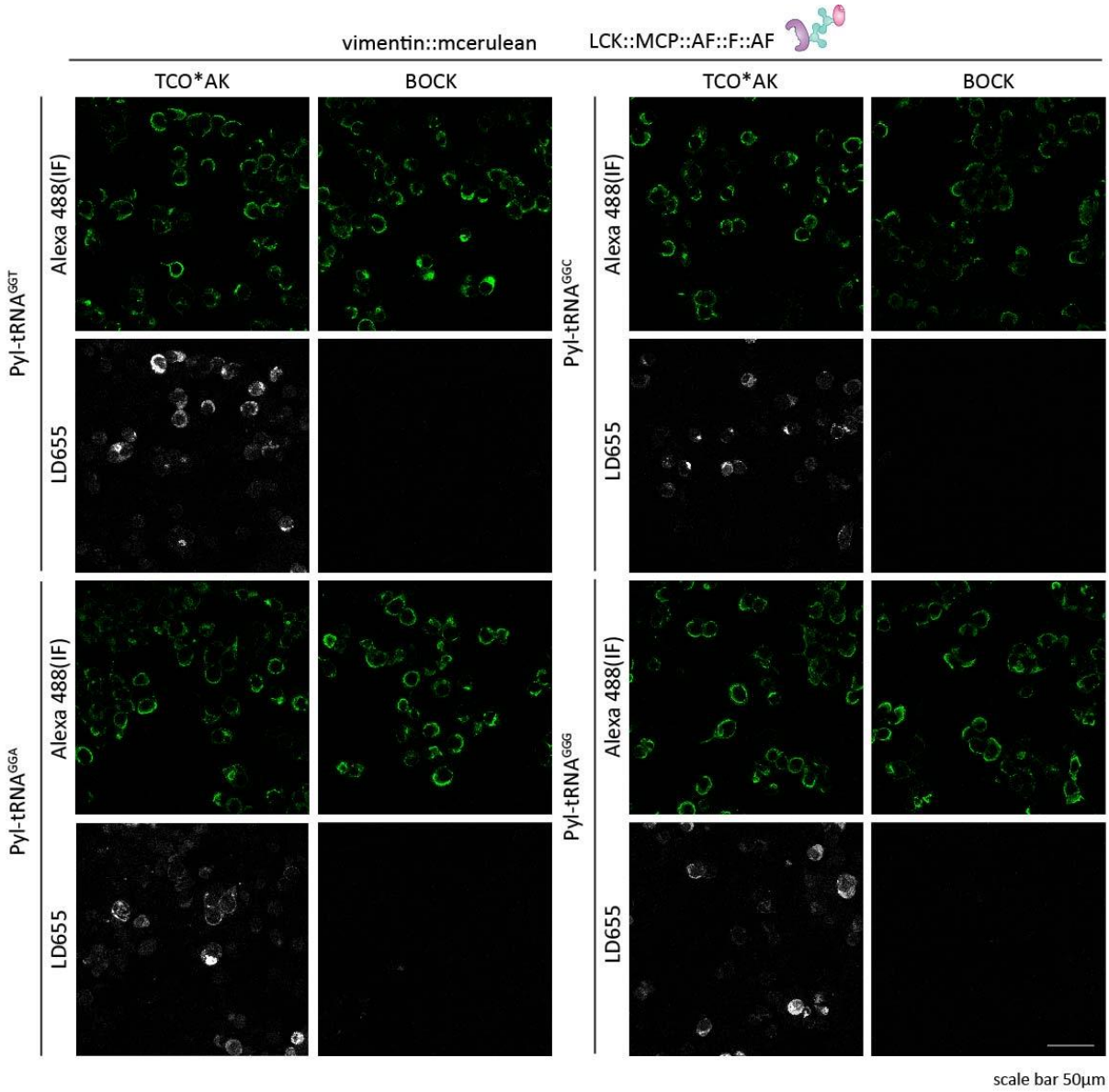


Figure 19| Confocal imaging results for codon selection for vimentin::mcerulean (GGT, GGC, GGA, GGG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons GGT, GGC, GGA and GGG in vimentin::mcerulean.

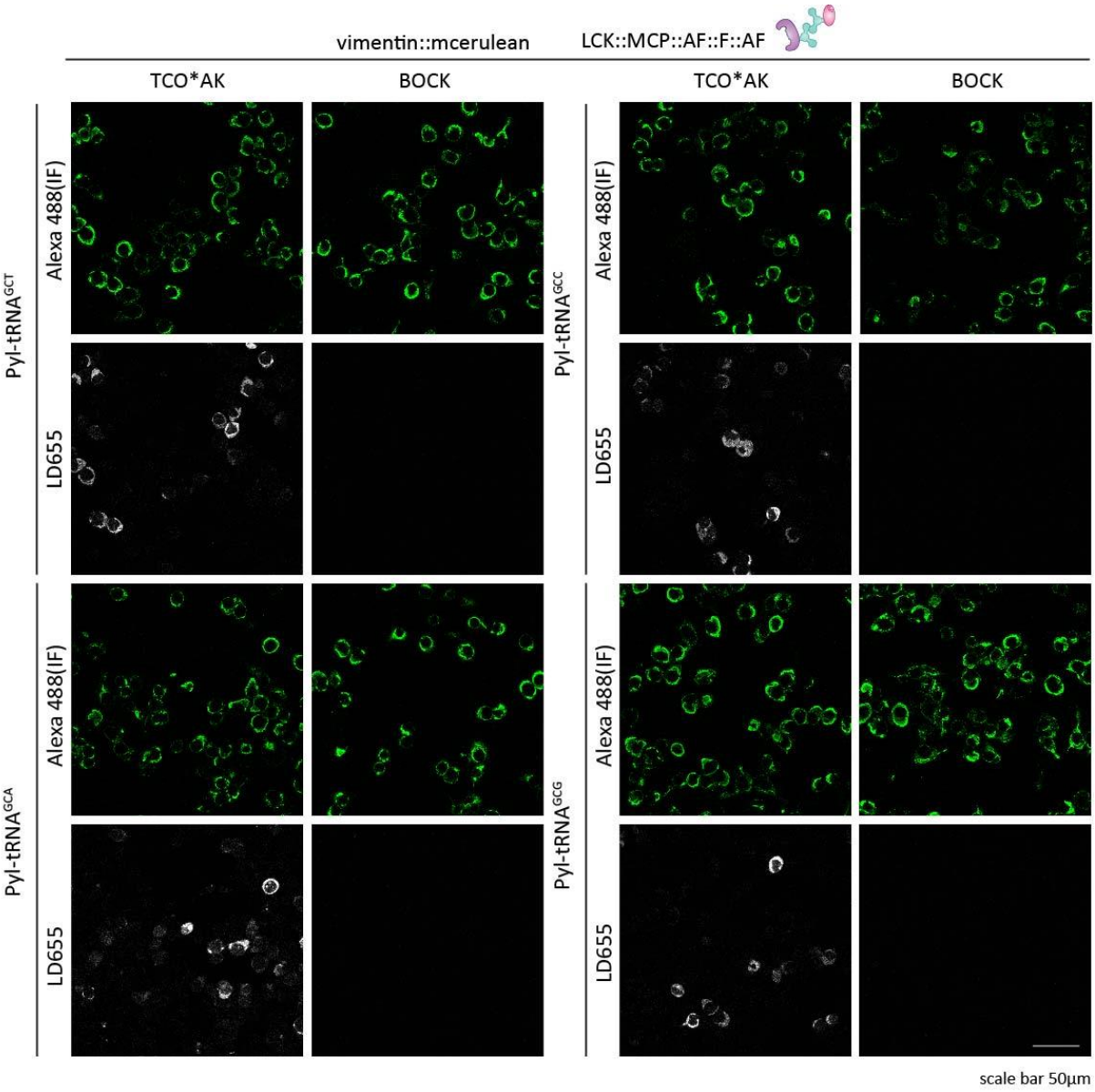


Figure 20| Confocal imaging results for codon selection for vimentin::mcerulean (codons GCT, GCC, GCA, GCG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons GCT, GCC, GCA and GCG in vimentin::mcerulean.

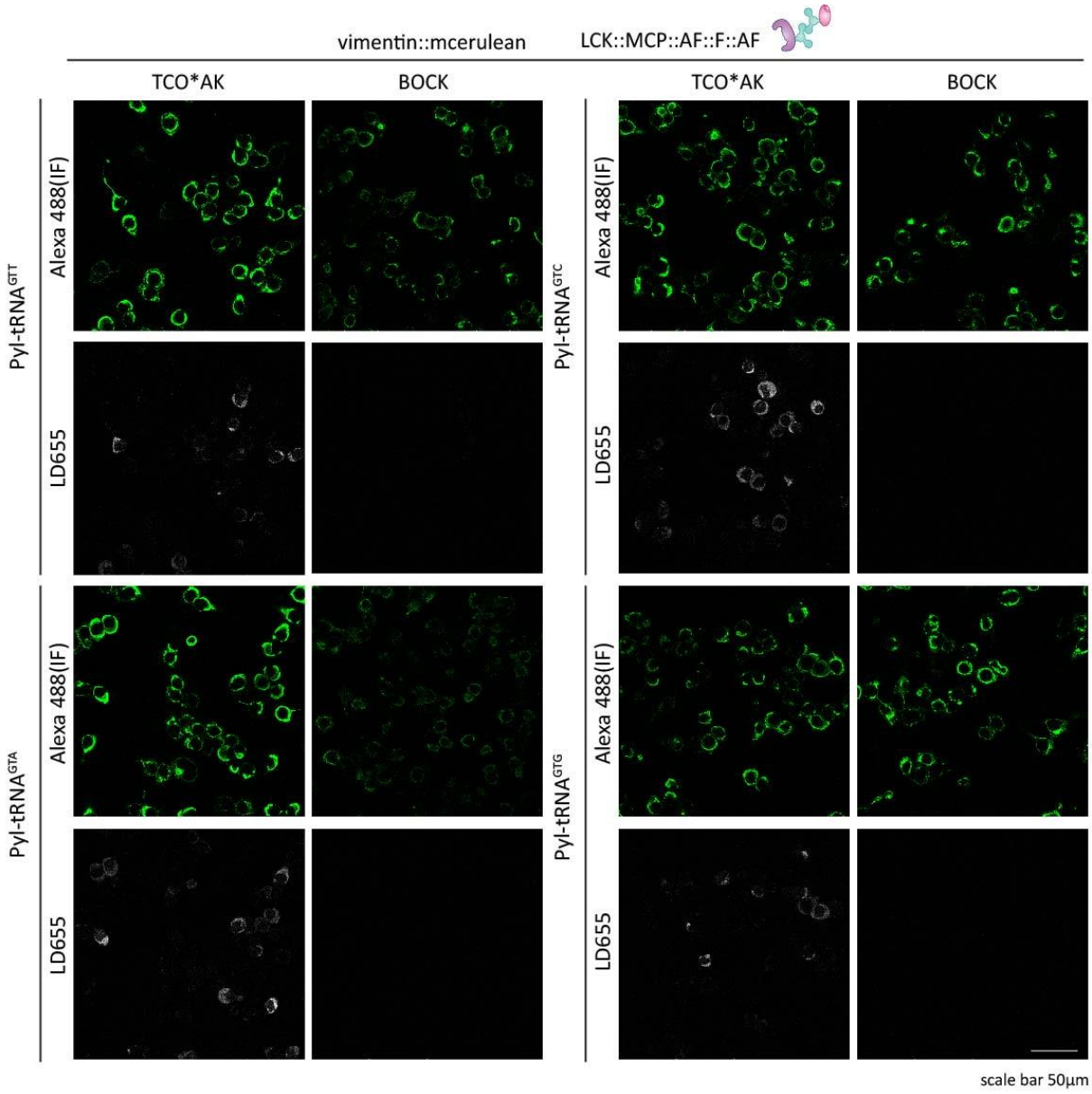


Figure 21| Confocal imaging results for codon selection for vimentin::mcerulean (GTT, GTC, GTA, GTG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons GTT, GTC,GTA and GTG in vimentin::mcerulean.

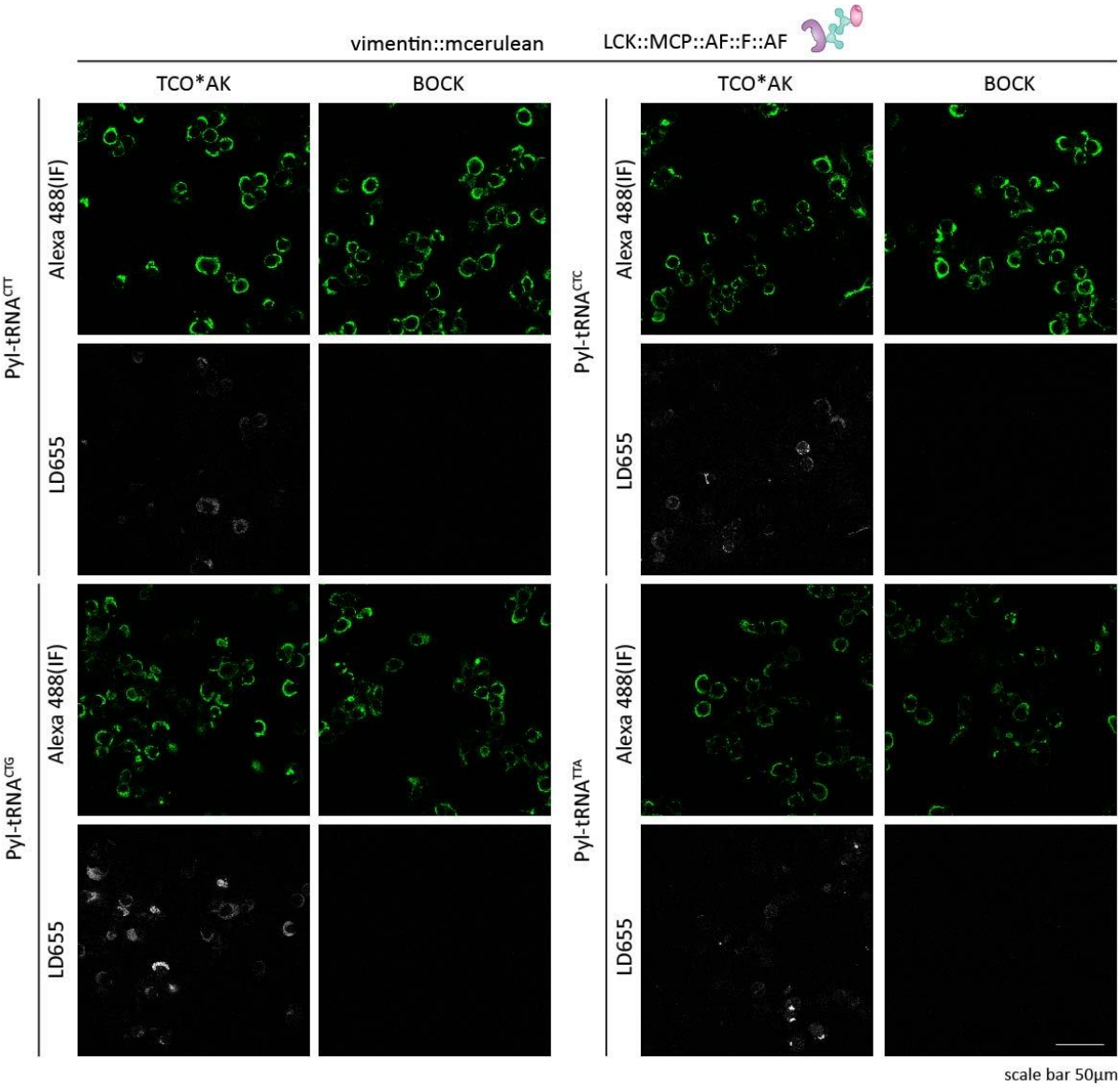


Figure 22| Confocal imaging results for codon selection for vimentin::mcerulean (CTT, CTC, CTG, TTA). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons CTT, CTC, CTG and TTA in vimentin::mcerulean.

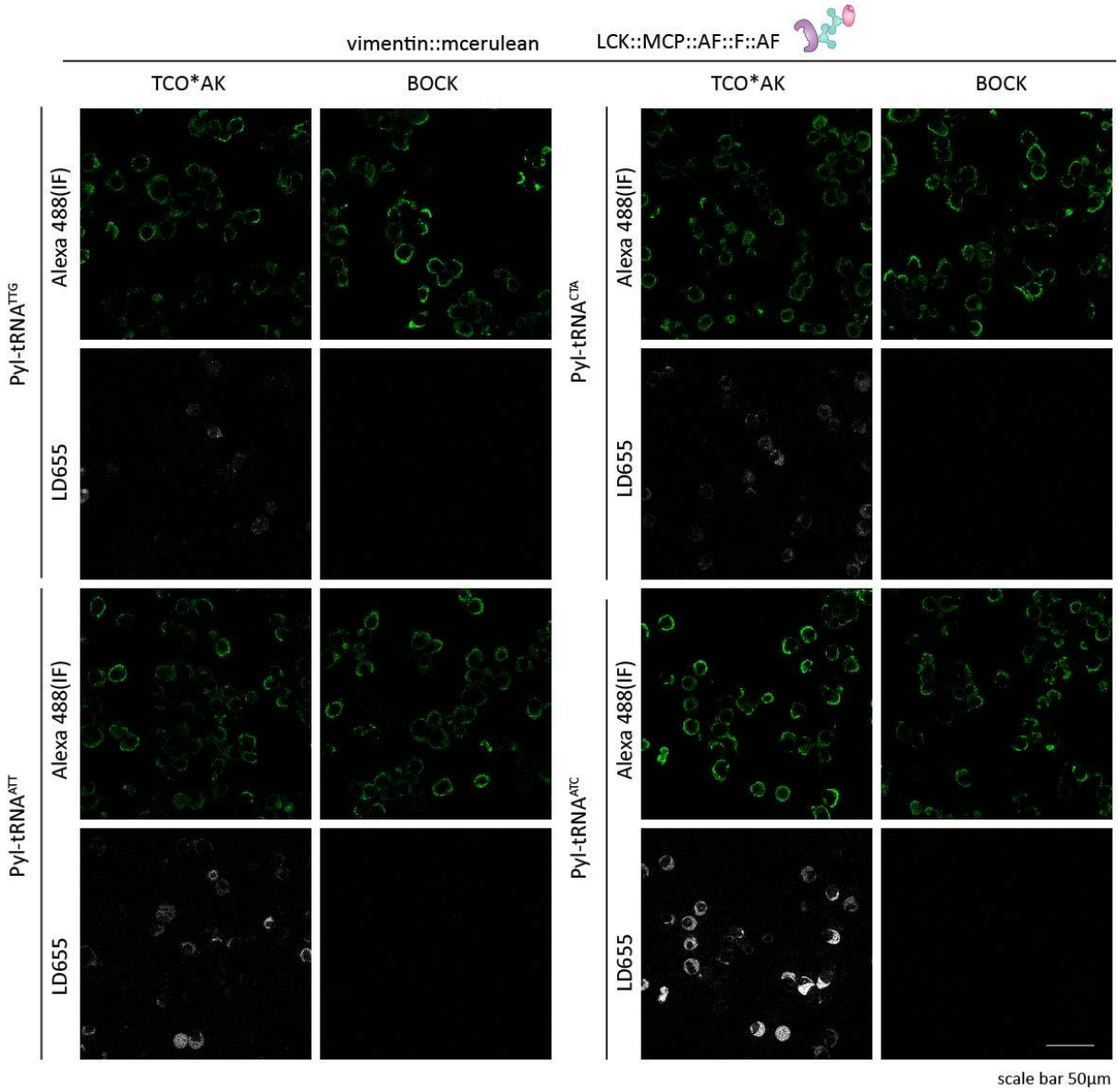


Figure 23| Confocal imaging results for codon selection for vimentin::mcerulean (TTG, CTA, ATT, ATC). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons TTG, CTA, ATT and ATC in vimentin::mcerulean.

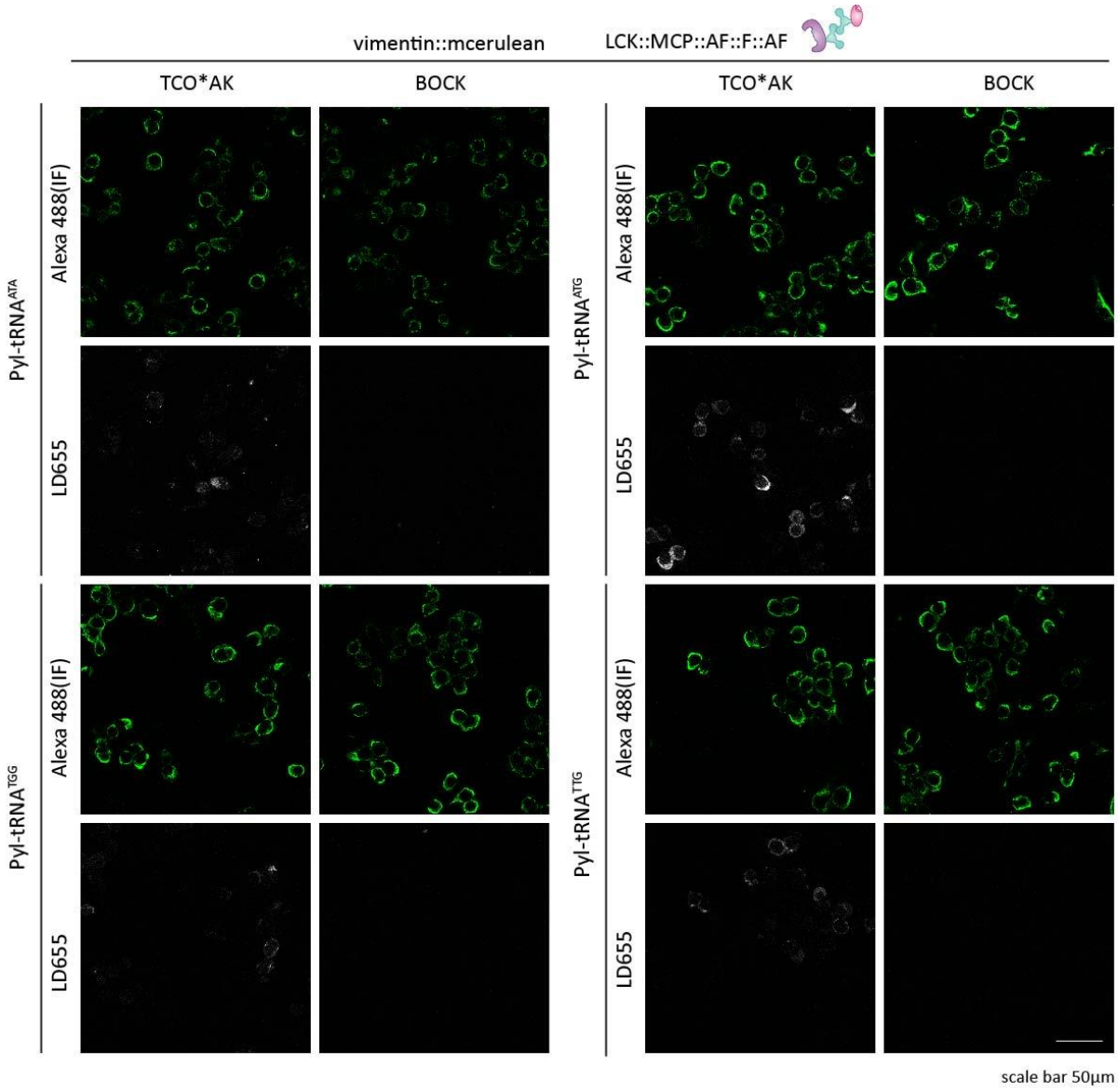


Figure 24| Confocal imaging results for codon selection for vimentin::mcerulean (ATA, ATG, TGG, TTG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons ATA, ATG, TGG and TTG in vimentin::mcerulean.

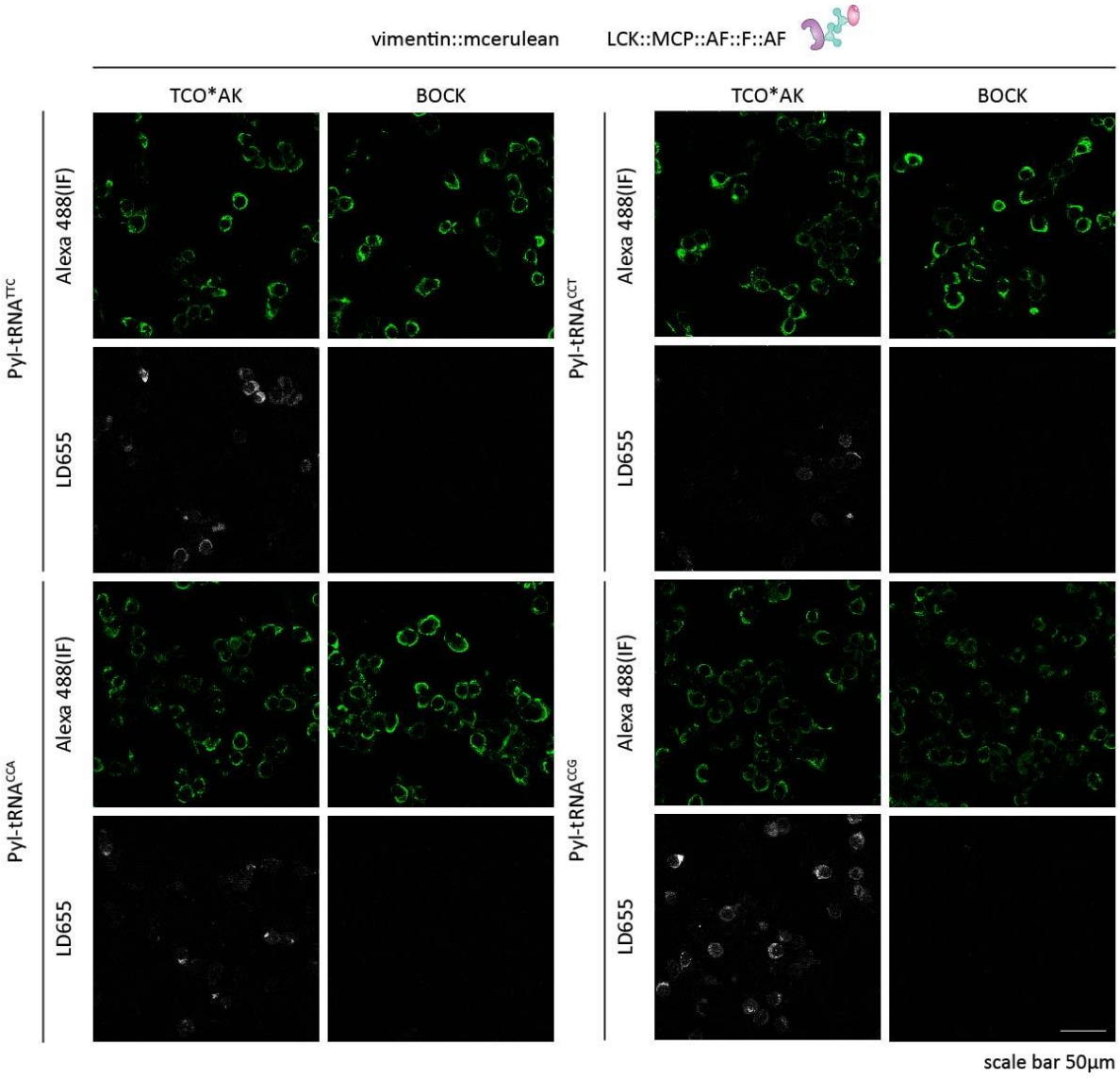


Figure 25| Confocal imaging results for codon selection for vimentin::mcerulean (TTC, CCT, CCA, CCG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons TTC, CCT, CCA and CCG in vimentin::mcerulean.

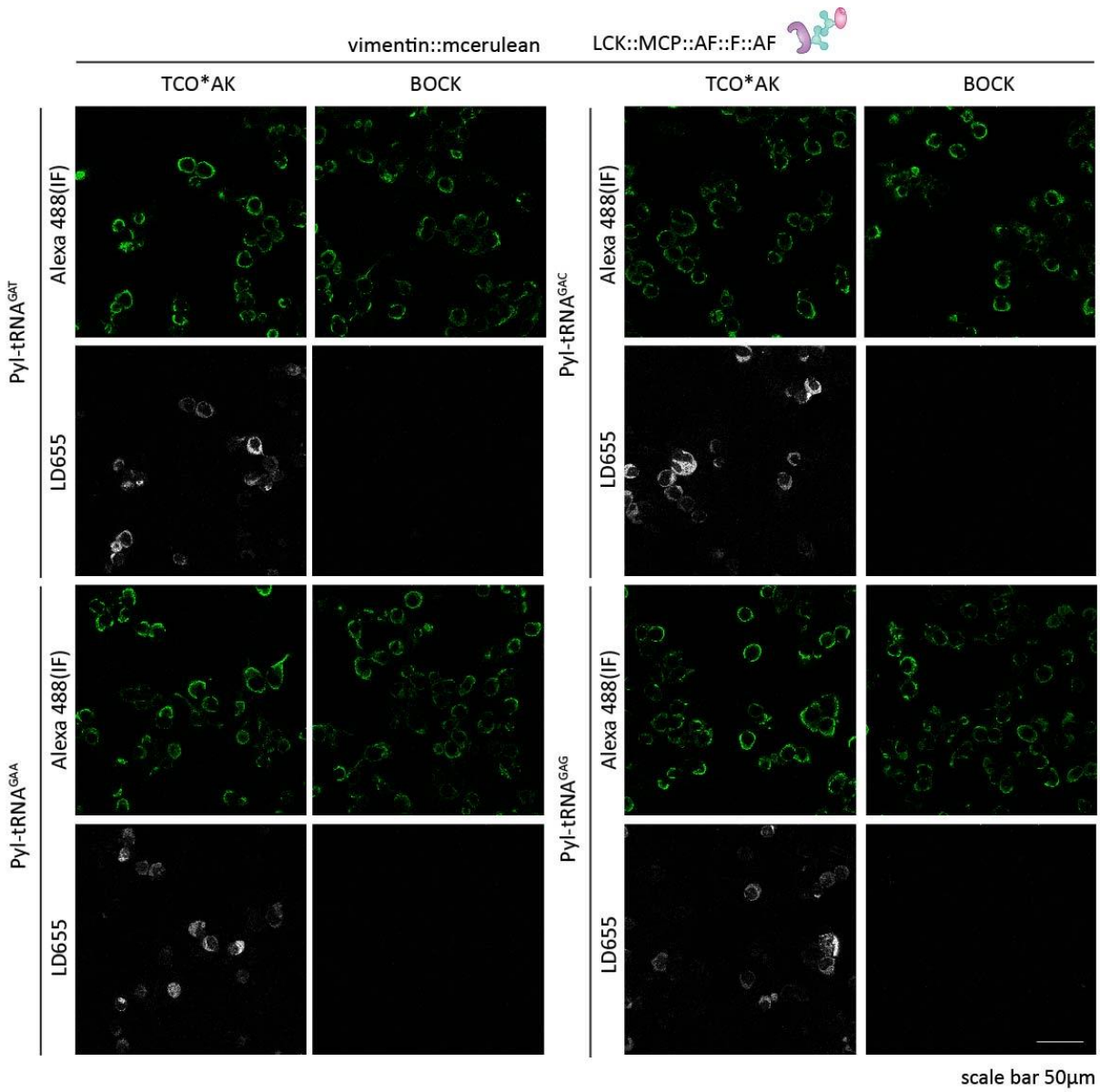


Figure 26| Confocal imaging results for codon selection for vimentin::mcerulean (GAT, GAC, GAA, GAG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons GAT, GAC, GAA and GAG in vimentin::mcerulean.

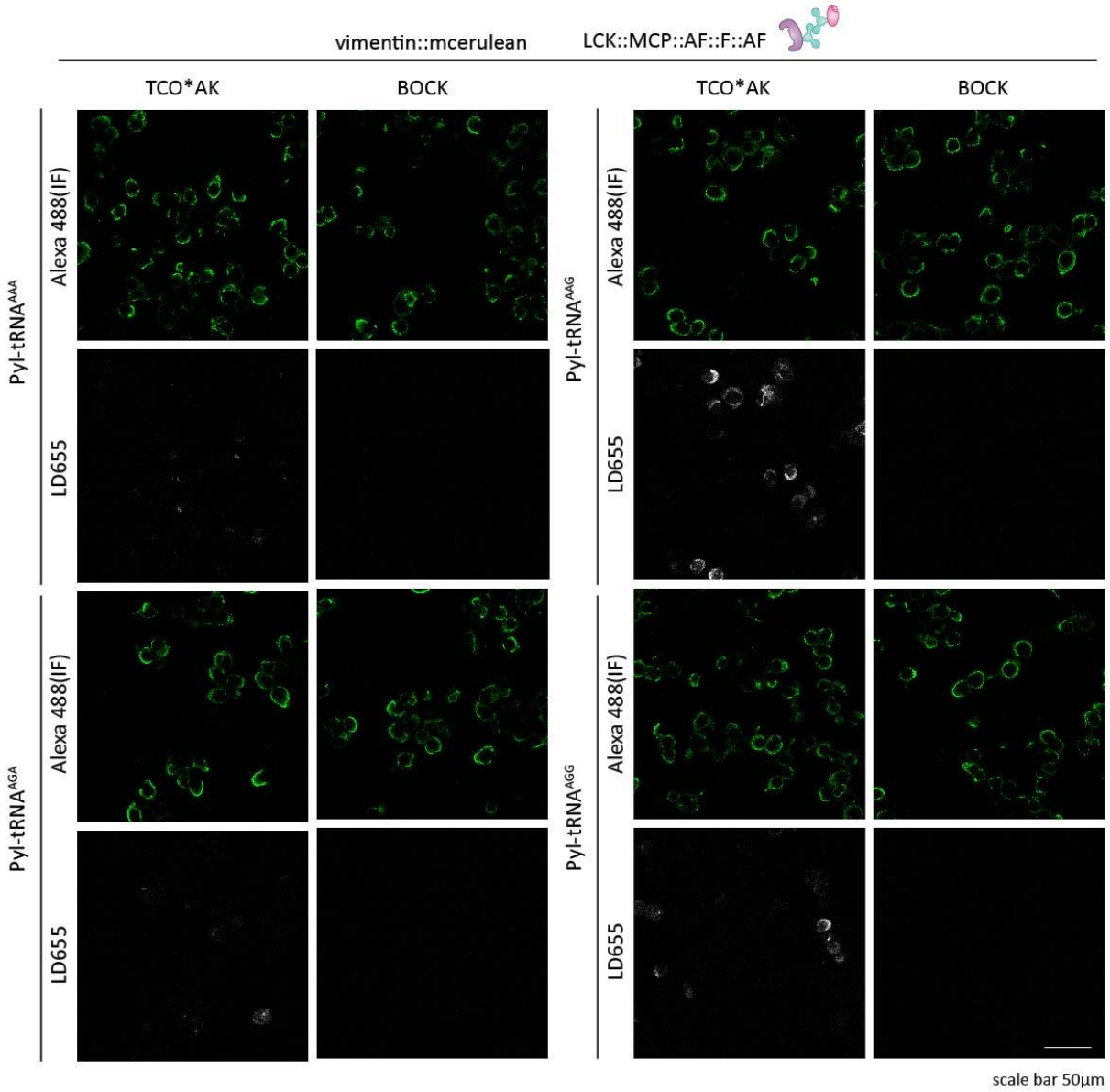


Figure 27| Confocal imaging results for codon selection for vimentin::mcerulean (codons AAA, AAG, AGA, AGG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons AAA, AAG, AGA and AGG in vimentin::mcerulean.

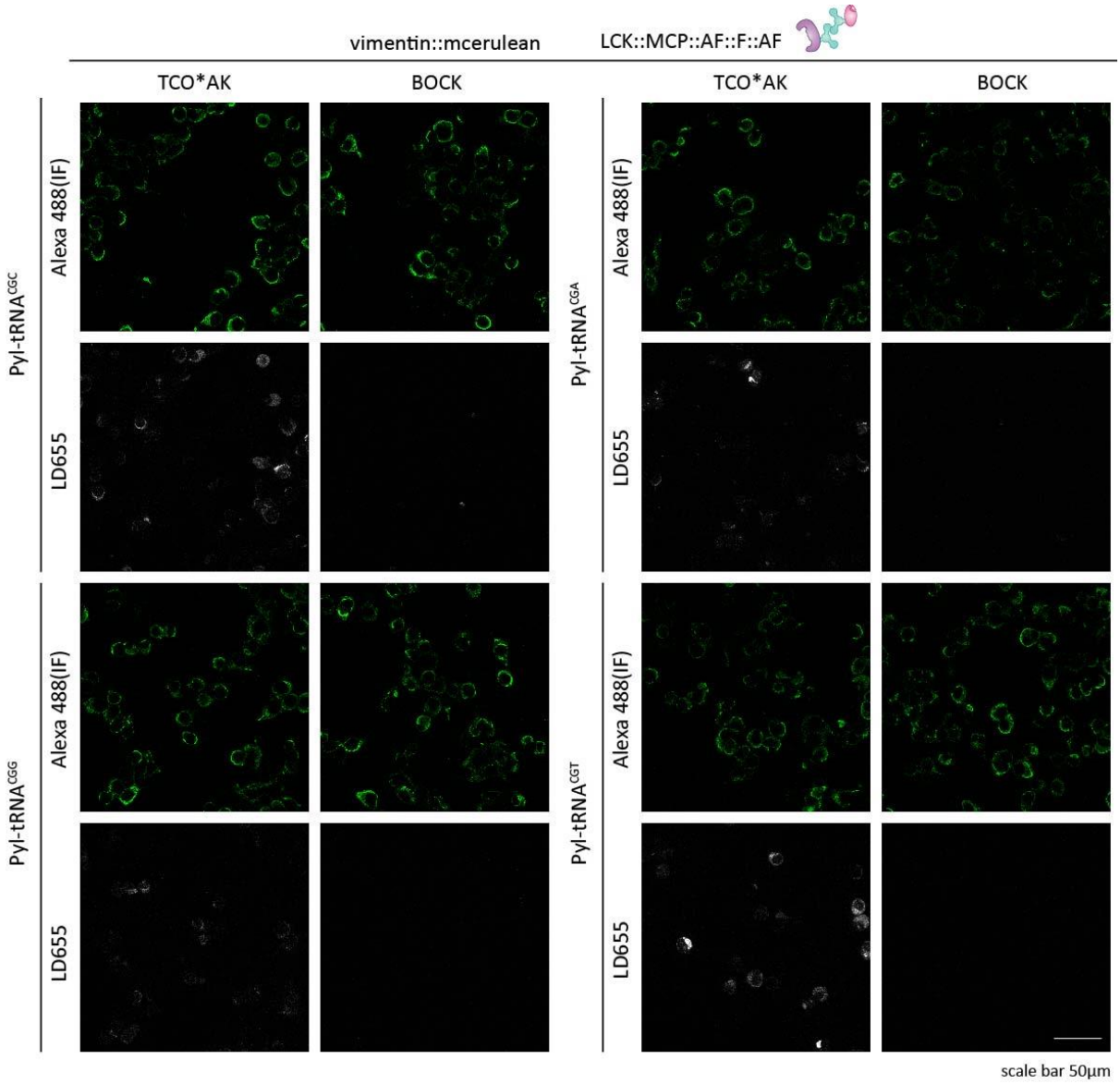


Figure 28| Confocal imaging results for codon selection for vimentin::mcerulean (codons CGC, CGA, CGG, CGT). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons CGC, CGA, CGG and CGT in vimentin::mcerulean.

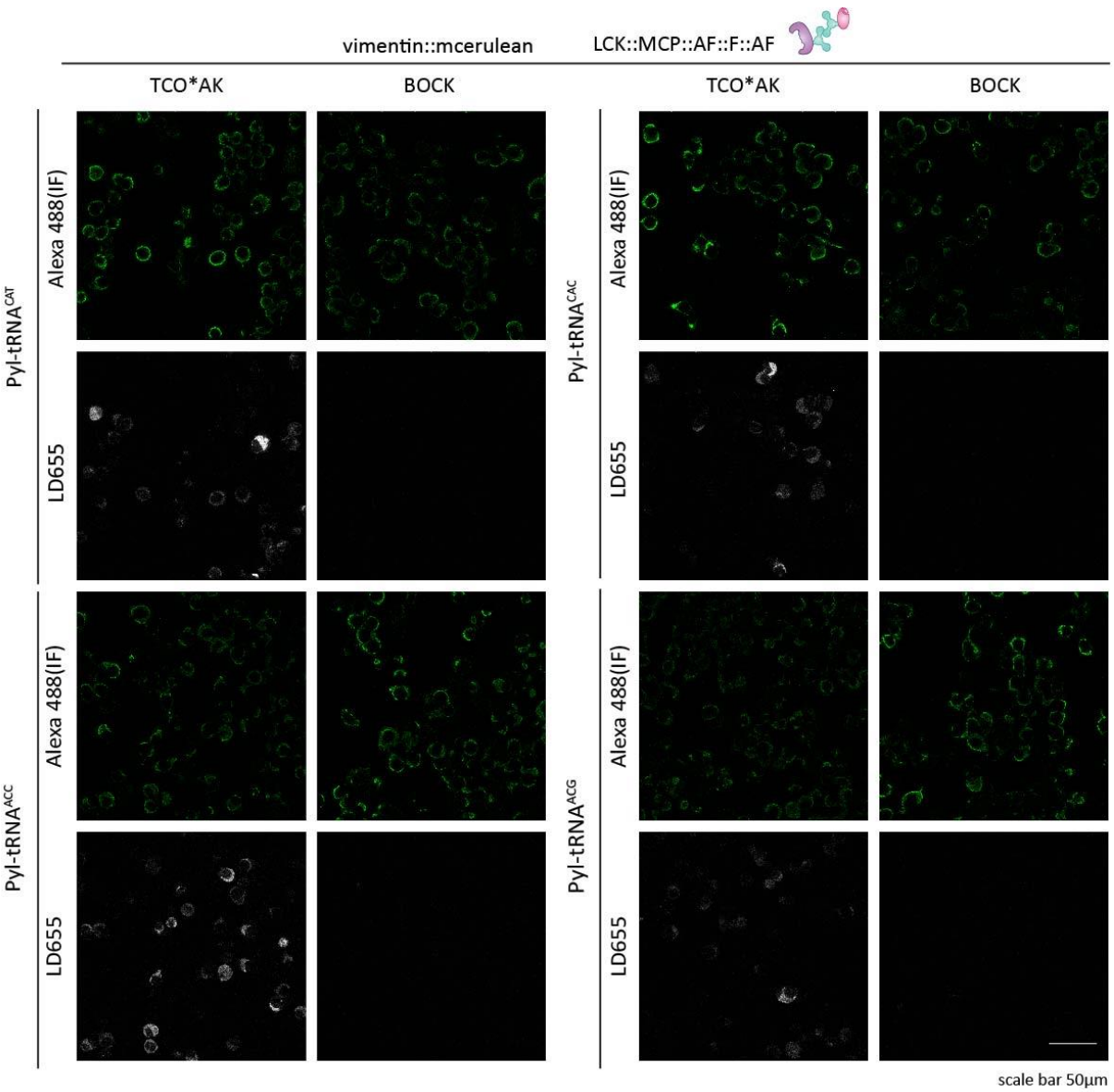


Figure 29| Confocal imaging results for codon selection for vimentin::mcerulean (codons CAT, CAC, ACC, ACG). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons CAT, CAC, ACC and ACG in vimentin::mcerulean.

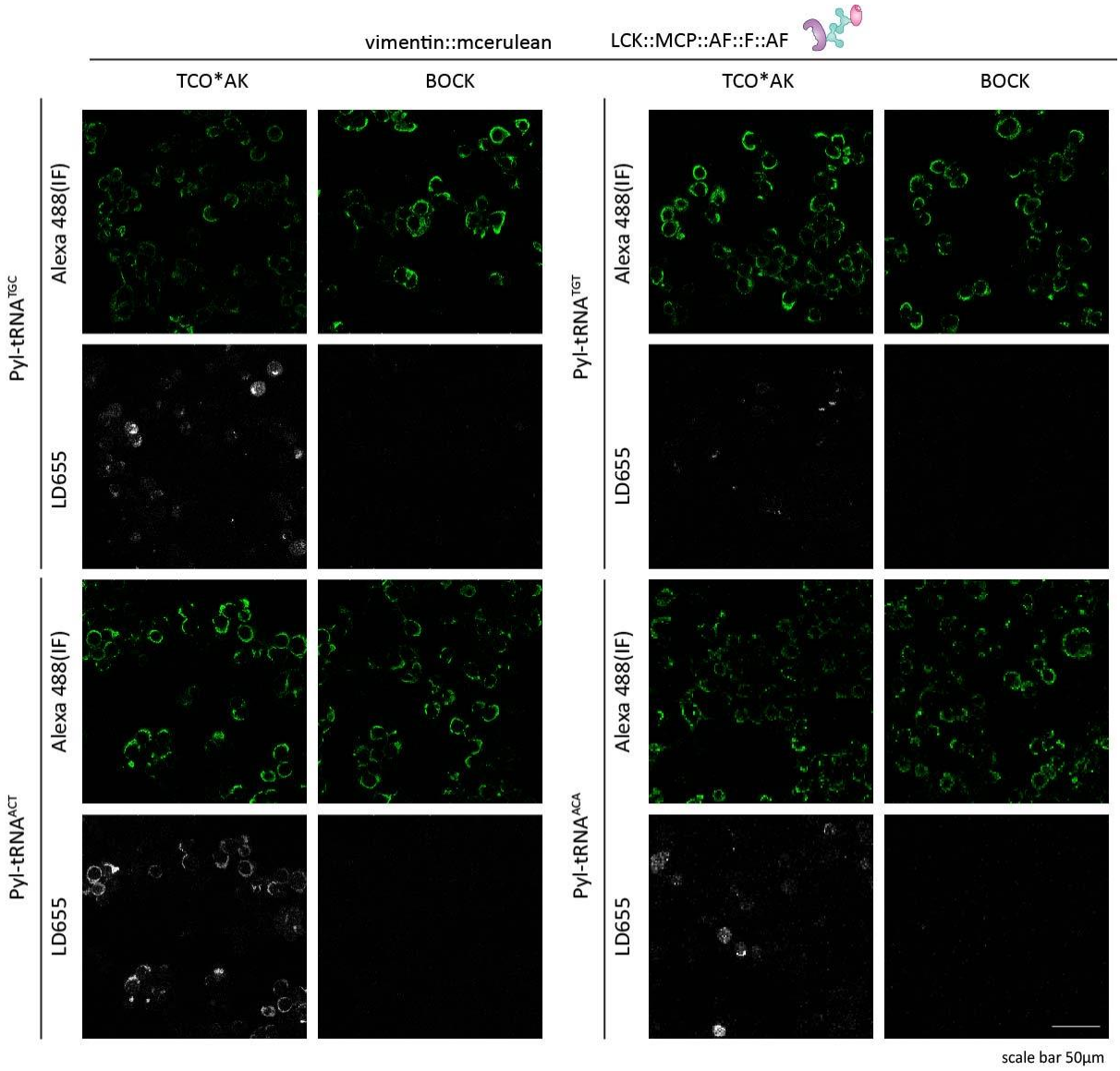
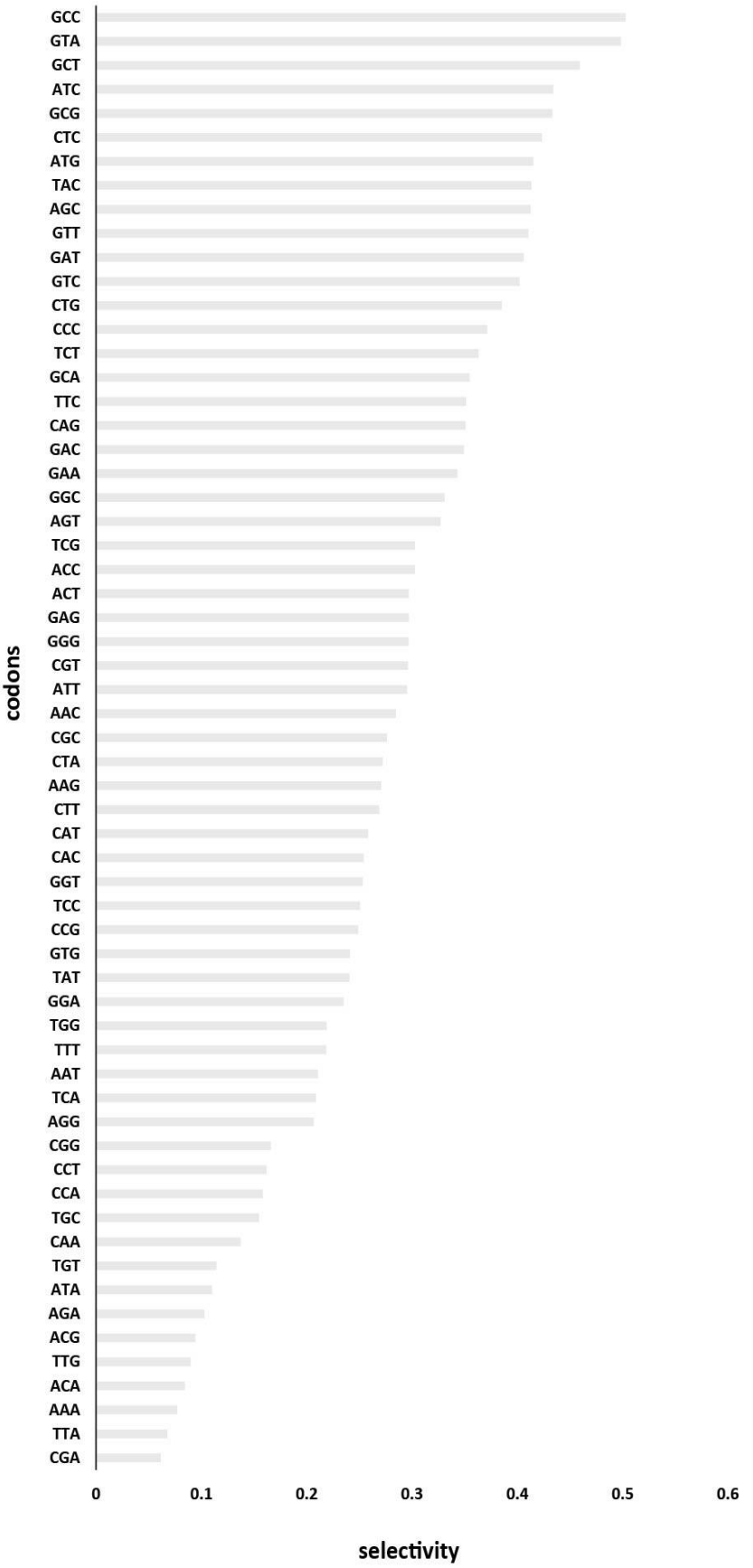


Figure 30| Confocal imaging results for codon selection for vimentin::mcerulean (GAT, GAC, ACT, ACA). The OTO LCK::MCP::AF::F::AF was used for sense codon reassignment in HEK293T cells. All necessary plasmids were transiently transfected in the cells. The ncAAs TCO*AK and BOCK were individually incorporated in vimentin::mcerulean. Cells were labelled with LD655-H-Tetrazine which can bio-orthogonally react with TCO*AK by IEDDA reaction. BOCK was used as a negative control since it cannot click react with LD655-H-Tetrazine. Additionally immunofluorescence (IF) was performed against a myc tag fused N-terminally to vimentin::mcerulean and the secondary antibody was conjugated with the dye Alexa 488. XXX in the Pyl-tRNA^{XXX} refers to the sense codon that was reassigned. The imaging data shown in this figure depicts the outcome of reassigning codons GAT, GAC, ACT and ACA in vimentin::mcerulean.

Figure 31| Selectivity analysis for reassignment of sense codons in vimentin::mcerulean. Selectivity is calculated as the ratio between total LD655 signal that colocalized with the immunofluorescence signal and the total LD655 signal (section 3.2.1). Extraction of the LD655 signal that colocalized with the vimentin signal was performed using FIJI. Calculation of total intensity values for the images in the LD655 channel was also done with FIJI. Detailed data analysis steps are described in section 2.2.9.1. A selectivity value of 1 would mean that the mcerulean and LD655 signals completely colocalized and hence only vimentin::mcerulean was selectively labelled by GCE.



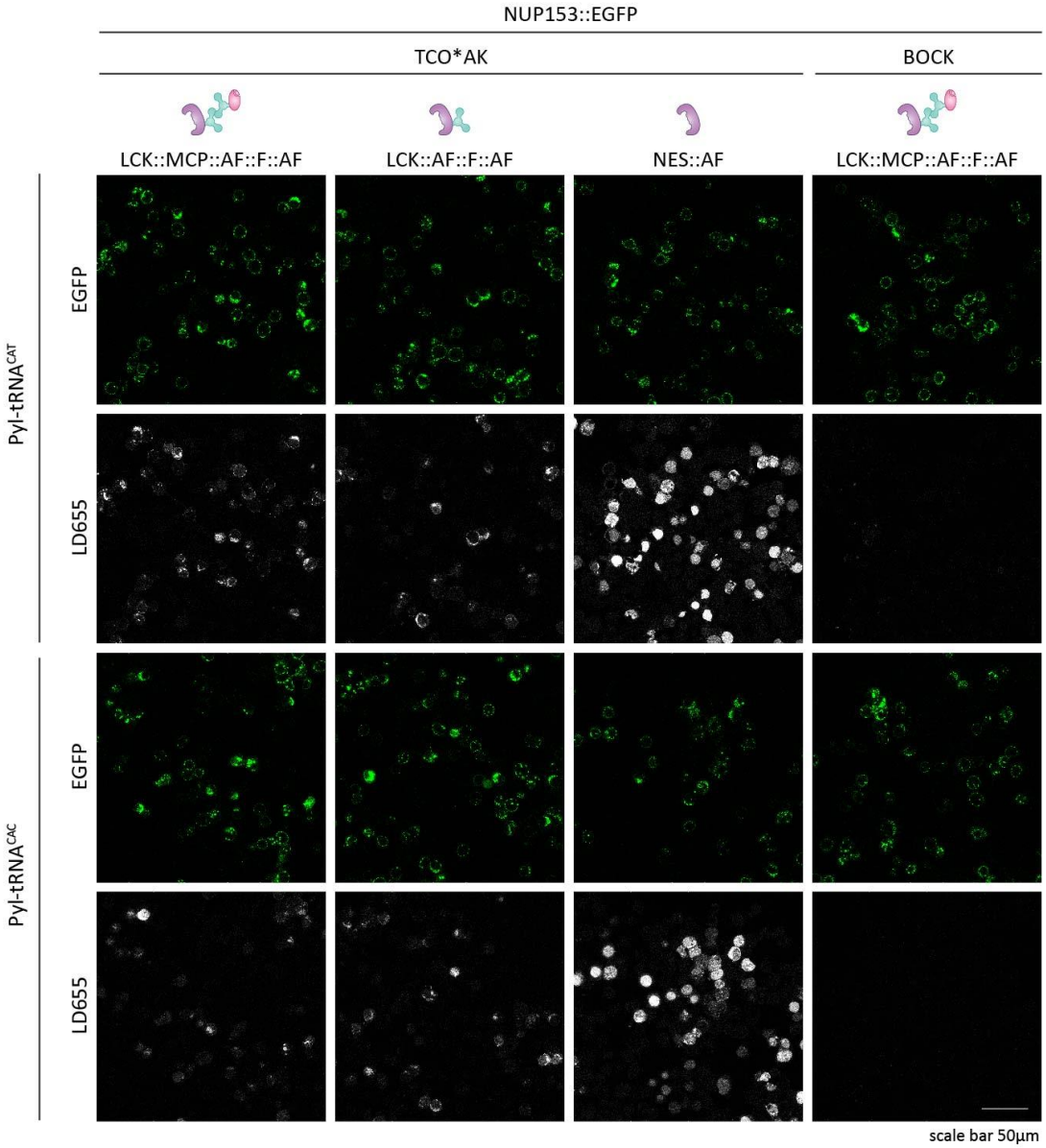


Figure 32| Confocal imaging results for codon selection for NUP153::EGFP (codons CAT, CAC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate TCO*AK/BOCK in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons CAT and CAC individually in NUP153::EGFP.

3.2.2 In-gel assay for validation of residue-specific sense codon reassignment for multiple codons in vimentin::mcerulean

A set of 10 sense codons; AGC, ACC, CTA, ACT, TAC, AAC, GCT, ATC, TTC and GAT, that showed promising results in the confocal microscopy-based codon selection experiments (section 3.2.1, figures 15-31) for vimentin::mcerulean was selected for an in-gel assay. Here these 10 sense codons were reassigned individually in vimentin::mcerulean to incorporate the clickable ncAA SCOK by the OTO LCK::MCP::AF::F::AF in HEK293T cells. Cells were transfected with three plasmids expressing the reporter vimentin::mcerulean, the OTO LCK::MCP::AF::F::AF and Pyl-tRNA^{XXX} (XXX- sense codon reassigned) respectively. mRNA of vimentin::mcerulean was tagged with MS2 loops to allow selective recruitment into the OTO LCK::MCP::AF::F::AF by MCP-MS2 interaction. The cell lysate was labelled with SCOK compatible fluorescent dye Cy5-H-Tetrazine followed by western blotting. Since Cy5-H-Tetrazine specifically click reacts with SCOK, only proteins with incorporated SCOK should be visible on a Cy5 scan of the blot. Immunolabelling was performed against myc tag fused N-terminally with vimentin::mcerulean to provide an estimation of total vimentin::mcerulean (modified and unmodified by GCE) expression.

Figure 33A shows a labelled band corresponding to vimentin::mcerulean in the Cy5 scan for all the 10 codons, thus demonstrating successful mRNA selective residue-specific sense codon reassignment. Additionally, ratio of the Cy5 band intensities for each codon and the corresponding immunolabel signal was plotted in figure 33B as an estimation of yield of POI modified with ncAA. The yield of SCOK incorporated vimentin::mcerulean recorded for the ACC and TTC codon reassignment were relatively higher than the yield recorded for the rest of the 8 codons. For most of the further experiments the ACC codon was used for residue-specific reassignment in vimentin::mcerulean.

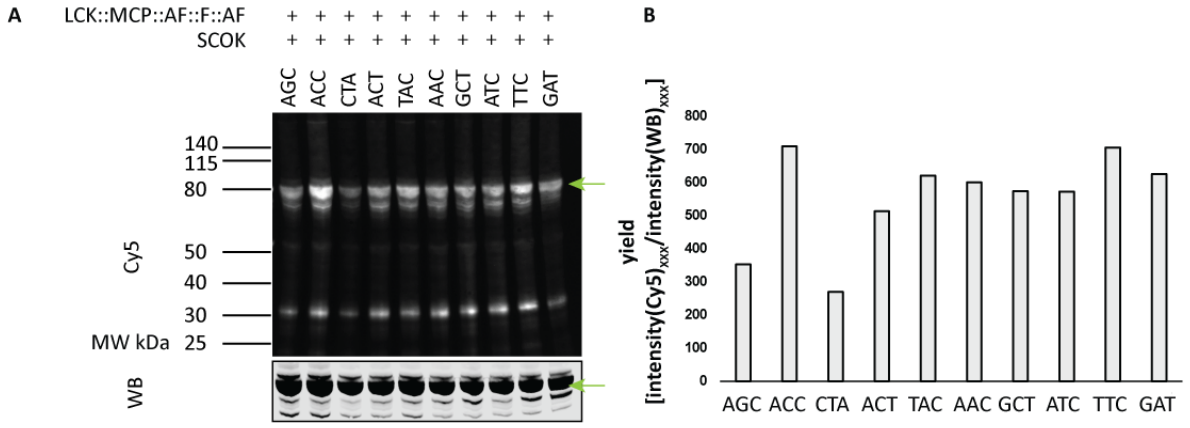


Figure 33| In-gel assay depicting mRNA selective residue-specific sense codon reassignment in vimentin::mcerulean for multiple sense codons in HEK293T cells. Depicts the in-gel assay outcome for reassigning a set of 10 sense codons individually in vimentin::mcerulean in HEK293T cells by the OTO LCK::MCP::AF::F::AF. The height of vimentin::mcerulean bands is marked by the green arrow. (B) Bar plot showing the yield of modified vimentin::mcerulean obtained by individually reassigning each of the selected 10 sense codons (yield is defined as the ratio of the Cy5 signal and the corresponding immunolabel signal (denoted as intensity$_{WB}$) for vimentin::mcerulean bands) .

3.2.3 OTO renders mRNA-selectivity to residue-specific sense codon reassignment

To demonstrate the applicability of the OTO system to render mRNA-selectivity to residue-specific sense codon reassignment, in-gel assay as well as confocal microscopy-based experiments were performed. A set of 6 codons based on the results of section 3.2.1and 3.2.2 was selected for the in-gel assay. Codons TAC, ACC, AGC, ATC and ACT were some of the suitable sense codons for reassignment, exhibiting relatively higher selectivity values for sense codon reassignment (figure 31) as well as promising result in the in-gel assay in section 3.2.2, figure 33. The ACG codon represented the sense codons not suitable for efficient residue-specific reassignment in vimentin::mcerulean. Each of the 6 codons were individually reassigned to incorporate SCOK in vimentin::mcerulean using the OTO LCK::MCP::AF::F::AF in HEK293T cells. MCP in the OTO facilitated selective recruitment of MS2 tagged mRNA of vimentin::mcerulean to the OTO. To demonstrate the necessity of the MCP-MS2 interaction for performing mRNA selective sense codon reassignment, in-gel assay was performed with an OTO lacking MCP, namely LCK::AF::F::AF. The cytoplasmic NES::AF represented non-selective sense codon reassignment. Additionally, a sample with only LCK::MCP::AF::F::AF and the respective Pyl-tRNA^{XXX}, without the reporter vimentin::mcerulean was also included. The OTO system is membraneless, hence, although not actively recruiting, it is still possible for unspecific mRNA to gain access to the OTO, leading to background ncAA incorporation. The additional control experiment with only LCK::MCP::AF::F::AF and respective Pyl-tRNA^{XXX}, without the reporter vimentin::mcerulean, would allow observation of the extent of unspecific protein modification by the OTO in absence of the overexpressed reporter. Samples where non-clickable ncAA CbzK was incorporated in vimentin::mcerulean was included as a negative control. The cell lysate was labelled with Cy5-H-Tetrazine that can click react with SCOK but not with CbzK. Subsequently western blot was performed with the Cy5 labelled cell lysates. The blots were also immunolabelled against an N-terminally fused myc tag in vimentin::mcerulean. The signal from the immunolabel was used as an estimation for total vimentin::mcerulean expressed (modified and unmodified by GCE).

In figure 34 it can be seen that for all the 6 cases there is a marked decrease in non-specific SCOK incorporation in samples where GCE was performed with the OTO LCK::MCP::AF::F::AF as compared to the samples where GCE was performed with the non-selective cytoplasmic NES::AF. This shows the effectiveness of the OTO for performing selective sense codon reassignment. Cy5 labelled band corresponding to vimentin::mcerulean was observed for all 6 codons,for ACG codon the band was quite faint as expected. Cy5 labelled vimentin bands were either absent or barely visible for all samples with LCK::AF::F::AF. The bar plot in figure 34 depicts the fold change yield of modified vimentin::mcerulean for samples with LCK::MCP::AF::F::AF as compared to those with LCK::AF::F::AF. Fold change yield is defined as:

$$\left[\frac{\left(\frac{Intensity_{Cy5}}{Intensity_{WB}}\right)_{LCK::MCP::AF::F::AF}}{\left(\frac{Intensity_{Cy5}}{Intensity_{WB}}\right)_{LCK::AF::F::AF}}\right]_{XXX}$$

WB refers to the western blot or immunolabelling signal and XXX refers to the codon reassigned. A fold change yield higher than 1 would mean higher yield of ncAA modified vimentin::mcerulean when GCE was performed with the OTO LCK::MCP::AF::F::AF as compared to when GCE was performed with the OTO LCK::AF::F::AF (lacks MCP necessary for selectively recruiting MS2 tagged vimentin::mcerulean mRNA). The fold change yield for all 6 codons were higher than 1, thereby highlighting the necessity of the mRNA recruiting machinery for efficiently performing mRNA selective residue-specific sense codon reassignment. It can also be noted that samples with only LCK::MCP::AF::F::AF and Pyl-tRNA^{XXX} did not show visible nonspecific ncAA incorporation in the Cy5 scan. No signal was observed for the samples where non-clickable CbzK was incorporated in vimentin::mcerulean. The in-gel assay for ACC codon reassignment was performed twice. Data corresponding to the second biological replicate is available in supplementary figure 3. The results shown in supplementary figure 3 are in-line with the conclusion drawn from figure 35.

For the confocal microscopy experiment the ACC codon was reassigned to incorporate TCO*AK in vimentin::mcerulean in HEK293T cells (figure 35). All necessary plasmids expressing the reporter and GCE machinery were transiently transfected in the cells. The cells were labelled with LD655-H-Tetrazine which can click react with TCO*AK. The signal from the fused fluorescent protein mcerulean was considered as standard labelling. The confocal microscopy images obtained from exciting mcerulean and LD655 individually were compared. Colocalized images were observed for samples where sense codon was reassigned with the OTO LCK::MCP::AF::F::AF (M2-0.59, M2 is the Mander's coefficient signifying the fraction of the image from LD655 channel that colocalizes with the image from the mcerulean channel. M2 was calculated with the plugin Jacop in FIJI and shows the average M2 value obtained from 6 imaging frames from 2 biological replicates) whereas mostly non-colocalized low intensity signal was observed in the LD655 channel corresponding to LCK::AF::F::AF. Unspecific LD655 labelling was observed for the samples where GCE was performed with the non-selective cytoplasmic NES::AF and no LD655 signal was recorded for the sample with non-clickable ncAA BOCK. Selectivity of sense codon reassignment was calculated as:

$$\left[\frac{Total\ intensity_{LD655\ signal\ colocalized\ with\ mcerulean\ signal}}{Total\ intensity_{total\ LD655\ signal}}\right]_{OTO\ or\ NES::AF}$$

Fold change selectivity ($selectivity_{OTO}/selectivity_{NES::AF}$) in figure 35 (bar plot) shows about 1.8fold change for mRNA selective residue-specific sense codon reassignment by LCK::MCP::AF::F::AF as compared to GCE performed with non-selective cytoplasmic NES::AF system and 1.2fold change for LCK::AF::F::AF. The bar plot depicts mean fold change values calculated from at least six imaging frames from two biological replicates. The error bar represents the standard deviation.

Thus, both the in-gel assay as well as the confocal microscopy experiments successfully demonstrate the effectiveness of film-like OTO for selective residue-specific sense codon reassignment.

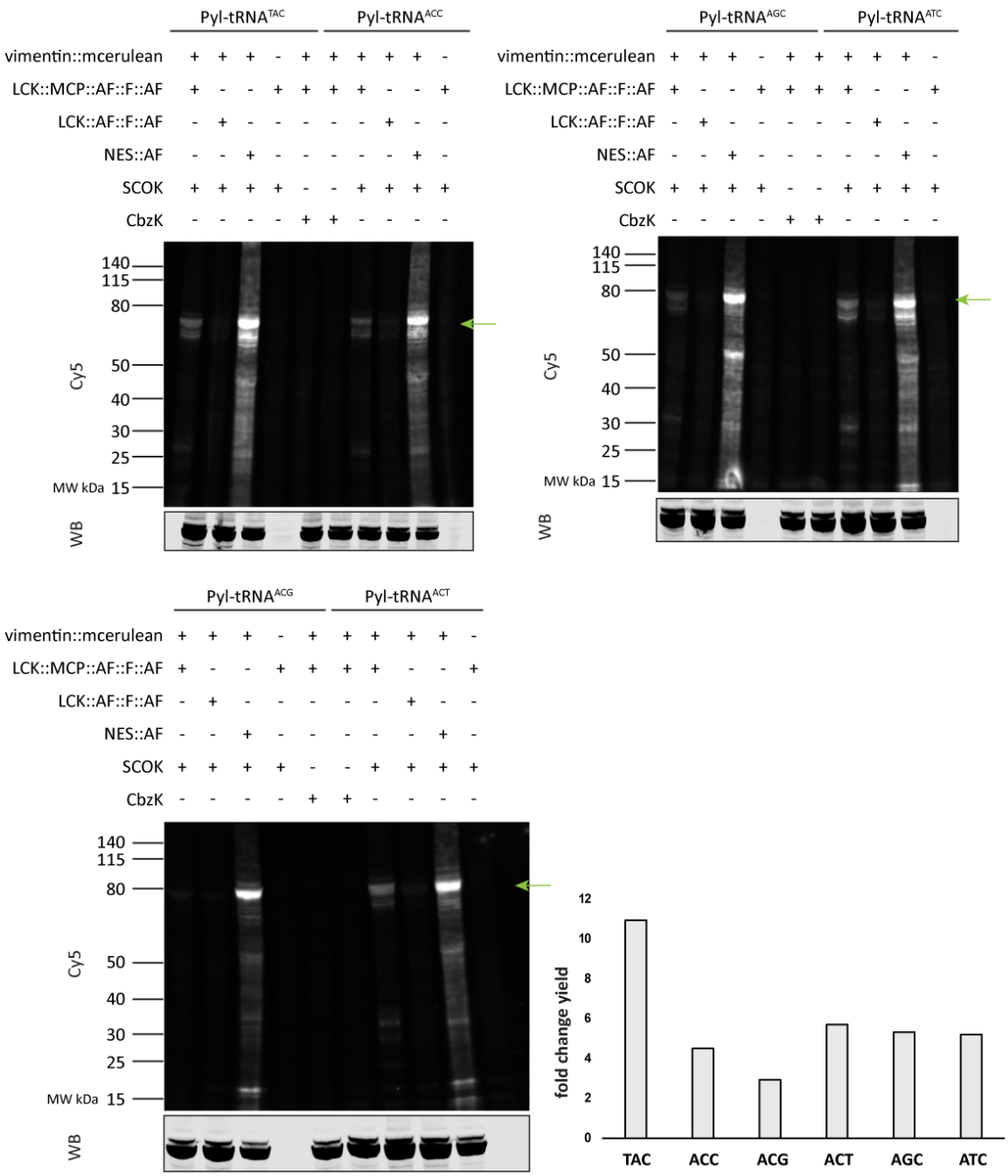


Figure 34| Results of the in-gel assay demonstrating effect of the OTO in reducing unspecific sense codon reassignment. The codons TAC, ACC,, AGC, ATC, ACG were individually reassigned in vimentin::mcerulean in HEK293T cells. The cell lysates were labelled with Cy5-H-Tetrazine. SDS-PAGE, followed by western blot was performed with the labelled cell lysates. XXX in Pyl-tRNA^{XXX} refers to the sense codon reassigned. A clear Cy5 labelled band corresponding to vimentin::mcerulean (marked by green arrow) can be seen in the lanes where GCE was performed by the OTO LCK::MCP::AF::F::AF for codons TAC, ACC, AGC, ATC and ACT while a very faint band can be seen for the ACG codon. All lanes corresponding to the samples where GCE was performed by cytoplasmic NES::AF shows significantly higher unspecific Cy5 labelling as compared to the samples where GCE was performed with the OTO LCK::MCP::AF::F::AF. The immunolabel scan of the blot is marked as WB. The bar plot depicts the fold change yield of modified vimentin::mcerulean obtained for samples where sense codons were reassigned by the OTO LCK::MCP::AF::F::AF as compared to samples where sense codons were reassigned with LCK::AF::F::AF (OTO lacking the RBD for selective recruitment of target mRNA). The fold change yield has been defined in section 3.2.3.

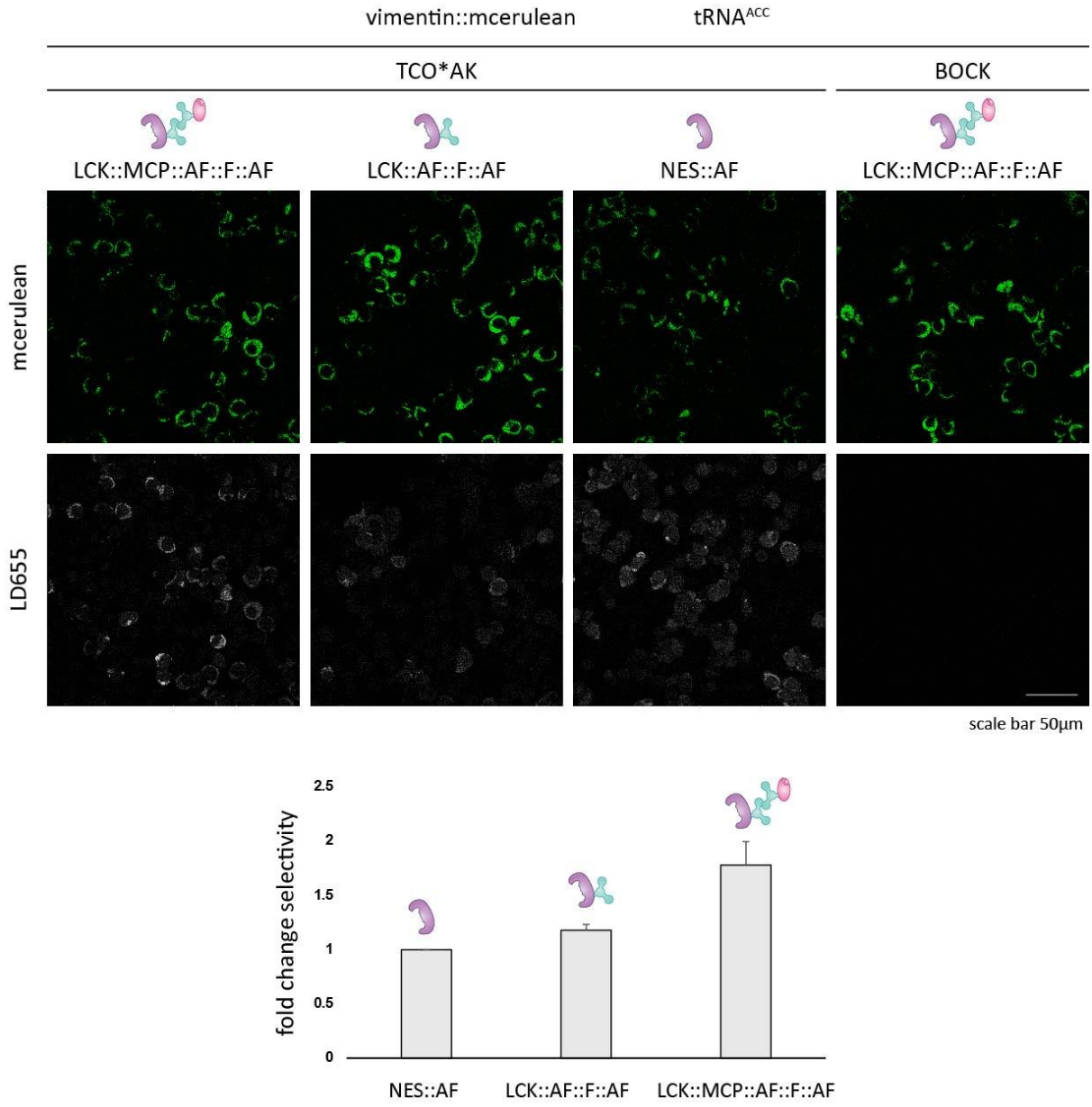


Figure 35| Confocal images showing reassignment of ACC in vimentin::mcerulean in HEK293T cells. The images in the first and last columns (from left) depict the outcome of residue-specific sense codon reassignment by the OTO LCK::MCP::AF::F::AF in vimentin::mcerulean. In the first column, the ncAA incorporated was TCO*AK which can click react with LD655-H-Tetrazine. In the last column from left, the non-clickable ncAA BOCK was incorporated and hence no signal was observed in the LD655 channel. The second and third columns(from left) depicts HEK293T cells where GCE was performed using the cytoplasmic NES::AF and the OTO LCK::AF::F::AF (lacking mRNA recruiting RBD MCP) respectively. The bar plot depicts fold change selectivity of samples with LCK::AF::F::AF and LCK::MCP::AF::F::AF as compared to NES::AF. Fold change selectivity has been defined as $selectivity_{OTO}/selectivity_{NES::AF}$. Selectivity was calculated as defined in section 3.2.3.

3.2.4 Estimation of average number of ncAAs incorporated in vimentin::mcerulean by reassigning the ACC codon

In sections 3.2.2, 3.2.3, I showed that it is possible to selectively modify POIs by residue-specific sense codon reassignment. Since this strategy has the potential to incorporate multiple ncAAs it is important to determine how many ncAAs get introduced on an average for a certain POI for a certain sense codon. For this purpose the reporters vimentin::mcerulean and vimentin^{116TAG}::mcerulean were used. The mRNA of both the reporters was tagged with MS2 loops to facilitate selective recruitment to MCP-based OTOs. In vimentin::mcerulean the ACC codon, which occurs 27 times in its genetic sequence, was reassigned to SCOK /CbzK in HEK293T cells. SCOK/CbzK was also incorporated in vimentin^{116TAG}::mcerulean by amber suppression in HEK293T cells. Cell lysates were labelled with Cy5-H-Tetrazine followed by western blotting. Proteins modified with SCOK would be visible on the Cy5 scan of the blot since Cy5-H-Tetrazine click reacts with SCOK, on the other hand proteins modified with CbzK would not be visible on the Cy5 scan of the blot since CbzK does not react with Cy5-H-Tetrazine. CbzK was used as a negative control to determine the extent of unspecific Cy5-H-Tetrazine interactions with cellular components. The blots were also immunolabelled against a myc tag fused N-terminally to vimentin::mcerulean / vimentin^{116TAG}::mcerulean. The signal from the anti-myc labelling served as an estimation of the total amount of POI expressed (modified and unmodified by GCE).

Now, for vimentin^{116TAG}::mecerulean there is only one codon for reassignment (TAG) and hence only one SCOK could be incorporated and labelled. Considering ($Intensity_{CY5}/Intensity_{WB}$, WB refers to the signal from anti-myc labelling) for vimentin^{116TAG}::mcerulean bands to be 1, the fold change of the same ratio for the bands corresponding to vimentin::mcerulean will be a measure of the average number of Cy5-H-Tetrazine introduced in vimentin::mcerulean, which in turn is a measure of the number of ncAAs incorporated. A fold change value higher than 1 would be an indication of multiple ncAA incorporation in vimentin::mcerulean. This experiment to determine the number of ncAAs incorporated in vimentin::mcerulean by ACC codon reassignment was performed with a range of SCOK concentrations. The results for the in-gel assay to estimate the number of ncAAs incorporated in vimentin::mcerulean by ACC codon reassignment, with the broadest range of ncAA concentrations(0 µM – 2 mM) is depicted in figure 36. Data for the same in-gel assay with shorter ncAA concentration range is provided in supplementary figures 1 and 2.

Figure 36A shows the Cy5 and immunolabel scans of the blot for vimentin::mcerulean (left) and vimentin^{116TAG}::mcerulean(right). The vimentin::mcerulean band height is marked by green arrows on the Cy5 scans of the blots. Fig 36B depicts the fold change in band intensities. Each bar represents the value :

$$\left[\frac{(Intensity_{Cy5}/Intensity_{WB})_{ACC}}{(Intensity_{Cy5}/Intensity_{WB})_{TAG}}\right]_{ncAA\ concentration}$$

WB refers to the signal from anti-myc labelling and is an estimation of total POI expressed.The fold change intensity values show a dependence on ncAA concentration,

reaching its highest value of 0.6fold at 250 µM SCOK. However a decrease is observed at 500µM SCOK followed by a slight rise for the higher concentrations. For none of the SCOK concentrations a fold change intensity value higher than 1 was observed. Hence incorporation of multiple ncAAs could not be proved. The same conclusion was also drawn from the in-gel assay results depicted in supplementary figures 1 and 2. The lanes on the Cy5 scan of the blot corresponding to samples where CbzK was incorporated by GCE showed no Cy5 labelled bands as expected.

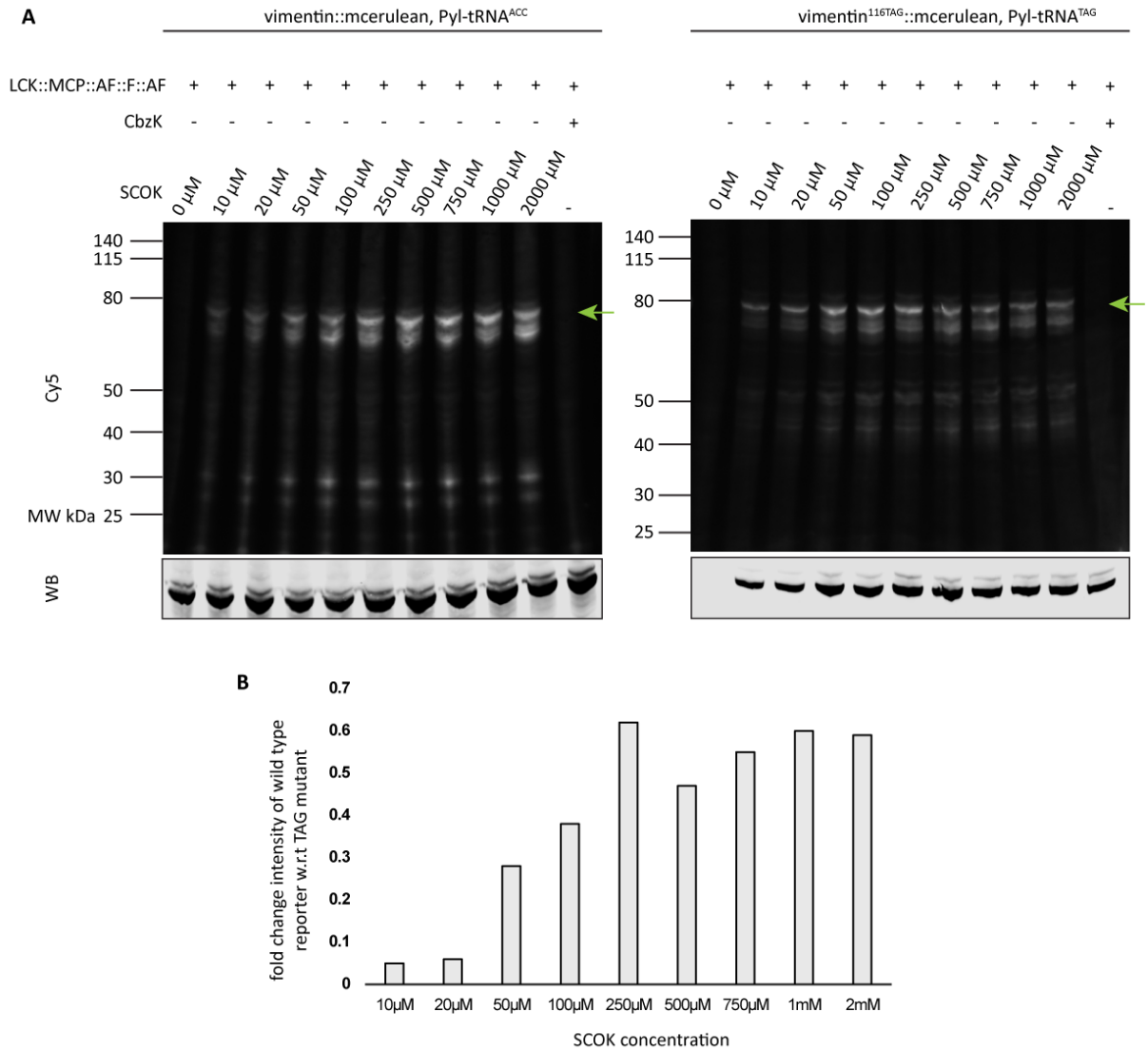


Figure 36| Estimation of number of ncAAs incorporated in vimentin::mcerulean by reassigning the ACC codon. (A) Shows the Cy5 and immunolabel (denoted as WB) scans of the blots for samples with vimentin::mcerulean (left) and vimentin^{116TAG}::mcerulean (right). The in-gel assay is described in detail in section 3.2.4. (B) Depicts the bar plot showing fold change intensity (defined in section 3.2.4) of vimentin::mcerulean samples compared to vimentin^{116TAG}::mcerulean. Interpretation of fold change intensity is explained in detail in section 3.2.4. A fold change intensity lower than 1 signifies that multiple ncAA incorporation in vimentin::mcerulean could not be proved.

3.2.5 Increased expression of Pyl-tRNA^{ACC} to enable incorporation of multiple ncAAs in vimentin::mcerulean

One of the prime challenges of sense codon reassignment is competition from the abundant endogenous tRNAs. Two ways to mitigate this issue would be to either lower the amount of endogenous tRNAs or increase the expression of GCE specific tRNA. Here the Pyl-tRNA^{ACC} expression was increased with an aim to improve efficiency of mRNA selective residue-specific sense codon reassignment and subsequently enable multiple ncAA incorporation.

For all previous experiments Pyl-tRNA^{ACC} was overexpressed in HEK293T cells by transiently transfecting a plasmid encoding a single copy of the Pyl-tRNA^{ACC} gene. To increase the expression of Pyl-tRNA , four plasmids were cloned encoding two, three, five and seven copies of the Pyl-tRNA^{ACC} gene. To determine the effect of increased Pyl-tRNA^{ACC} expression on residue-specific sense codon reassignment by OTO, in-gel assay was performed. HEK293T cells were transfected with three plasmids, respectively expressing the reporter vimentin::mcerulean, the OTO LCK::MCP::AF::F::AF and either one, two, three, five or seven copies of the Pyl-tRNA^{ACC}. The mRNA of vimentin::mcerulean was tagged with MS2 loops. Along with sense codon reassignment, amber suppression was also performed to compare and evaluate the performance of residue-specific sense codon reassignment as well as to determine whether increasing Pyl-tRNA^{ACC} expression enabled multiple ncAA incorporation. To perform amber suppression, HEK293T cells were transfected with plasmids expressing the reporter vimentin^{116TAG}::mcerulean , the OTO LCK::MCP::AF::F::AF and a single copy of Pyl-tRNA^{TAG}. The mRNA of vimentin^{116TAG}::mcerulean was tagged with MS2 loops. ACC codon was reassigned in vimentin::mcerulean and amber (TAG) codon was suppressed in vimentin^{116TAG}::mcerulean to incorporate SCOK/CbzK. Cell lysates were labelled with Cy5-H-Tetrazine, which reacts bio-orthogonally with clickable ncAA SCOK. CbzK was used as a negative control since it does not click react with Cy5-H-Tetrazine. SDS-PAGE and western blot was performed with the labelled lysates. The proteins modified with SCOK and labelled by Cy5-H-Tetrazine were visualized in the Cy5 scan of the blot. Total amount of reporter protein (modified and unmodified by GCE) was visualized by immunolabelling the blot against myc tag fused N-terminally to both vimentin::mcerulean and vimentin^{116TAG}::mcerulean (figure 37). For all samples the ncAA concentration used was 250 µM.

Figure 37 (left) depicts the Cy5 and the immunolabel scans of the blot. Fold change intensity was calculated, as detailed in section 3.2.4, for each sample with Cy5-H-Tetrazine labelled POI. Each bar corresponding to Pyl-tRNA^{ACC} in the bar plot depicted in figure 37 represents the value:

$$\left[\frac{(Intensity_{Cy5}/Intensity_{WB})_{ACC}}{(Intensity_{Cy5}/Intensity_{WB})_{TAG}}\right]_{Yx\ Pyl-tRNA^{ACC}}$$

where Y refers to the number of Pyl-tRNA^{ACC} gene copies encoded in the tRNA expression plasmid transfected to the HEK293T cells and Intensity$_{WB}$ refers to the signal obtained from anti-myc immunolabelling of the POIs.

As explained in section 3.2.4 a fold change intensity value higher than 1 signifies incorporation of multiple ncAAs in the POI. Figure 37 shows that for none of the vimentin::mcerulean samples fold change intensity value of higher than 1 was reached. Hence incorporation of multiple ncAAs in vimentin::mcerulean by residue-specific sense codon reassignment could not be proved. Moreover, the highest value was recorded for 1x Pyl-tRNA^{ACC} which was unexpected. The increase in Pyl-tRNA^{ACC} expression did not show a positive impact on sense codon reassignment. However, due to time constraints this experiment could only be performed once. Further studies are necessary to conclusively validate the role of increased GCE specific tRNA expression for residue-specific sense codon reassignment.

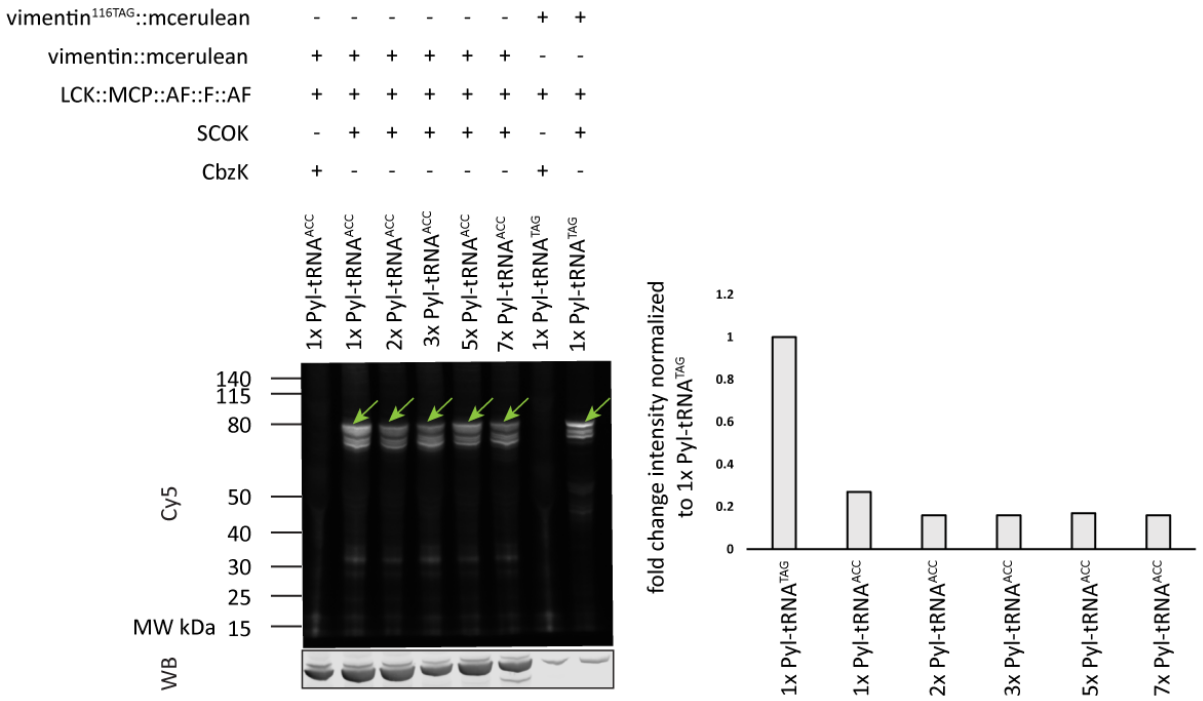


Figure 37| The effect of increasing Pyl-tRNA^{ACC} expression on ACC codon reassignment in vimentin::mcerulean in HEK293T cells. HEK293T cells were transfected with plasmids expressing vimentin::mcerulean / vimentin^{116TAG}::mcerulean, YxPyl-tRNA^{ACC} (Y – number of copies of the Pyl-tRNA^{ACC} gene being expressed from the transfected plasmid) or Pyl-tRNA^{TAG} and the OTO LCK::MCP:;AF::F::AF. SCOK / CbzK was incorporated in the POIs. The green arrows on the Cy5 scan mark vimentin::mcerulean (both wild type and TAG mutant). The anti-myc (myc tag was fused N-terminally to both vimentin::mcerulean and vimentin^{116TAG}::mcerulean)immunolabel scan of the blot is denoted as WB. The bar plot shows the fold change intensities as defined in section 3.2.5. Interpretation of the fold change intensity is explained in sections 3.2.4 and 3.2.5. A fold change intensity lower than 1 signifies incorporation of multiple ncAA in vimentin::mcerulean could not be proven.

3.2.6 Using OTO without the RNA binding domain of FUS for reassigning ACC codon in vimentin::mcerulean

The initial OTOs, designed by Reinkemeier *et al.*, were developed using multiple strategies including only spatial targeting, only phase separation and combination of both. Combination of spatial targeting to a subcellular location and phase separation yielded the best performing OTOs.[4] OTOs were designed with either SPD5 or full length FUS

fused to the synthetase PylRS(AF) (denoted as AF throughout the thesis) as the phase separating domain. The OTO with full length FUS achieved highest selectivity for stop codon suppression. Further improvements on the OTOs by Reinkemeier *et al*. resulted in film-like OTOs on membrane surfaces. These OTOs were also built with full length FUS.[151] For the study in this thesis high performing, already established OTOs, containing full length FUS as the phase separating domain, were selected from those designed by Reinkemeier *et al*.

However, FUS is an RNA binding protein [181] and it might be possible for its RNA binding domain to recruit unspecific RNAs into the OTO or have a bias towards the mRNA of the reporter in use that is vimentin::mcerulean. In section 3.2.3 figure 34, control samples without the reporter, only with the OTO LCK::MCP::AF::F::AF and the respective Pyl-tRNA^{XXX} (XXX refers to the sense codon to be reassigned) were included for 6 sense codons to determine unspecific sense codon reassignment. No bands were detected in the respective lanes in Cy5 scan of the blots, thus showing full length FUS did not recruit unspecific mRNAs into the OTO to a detectable extent.

In this section an in-gel assay is described where an OTO with full length FUS(F), LCK::MCP::F::AF as well as an OTO with truncated FUS (ΔF) (FUS$^{2\text{-}267}$ that is FUS without the RNA binding domain), LCK::MCP::ΔF::AF, were used individually to reassign ACC codon in vimentin::mcerulean (the mRNA of vimentin::mcerulean was tagged with MS2 loops) to incorporate SCOK/CbzK in HEK293T cells. Three additional samples were also included where GCE was performed by the OTO LCK::F::AF (OTO without the mRNA recruiting RBD), LCK::MCP::AF(OTO without the phase separating domain) and the cytoplasmic NES::AF respectively. Cell lysates were labelled with Cy5-H-Tetrazine followed by western blot. Proteins modified with SCOK would be visible in the Cy5 scan of the blot since Cy5-H-Tetrazine click reacts with SCOK. The sample where CbzK was incorporated in vimentin::mcerulean was included as the negative control since CbzK does not click react with Cy5-H-Tetrazine. The blots were also immunolabelled against the myc tag fused to vimentin::mcerulean.

Figure 38 shows the Cy5 and the immunolabel scan of the blot. Each bar in the bar plot in figure 38 represents the value:

$$\left[\frac{Cy5\ intensity_{vimentin::mcerulean}\ (Intensity_{Cy5})}{immunolabel\ intensity_{vimentin::mcerulean}\ (Intensity_{WB})}\right]_{OTO}$$

which is a measure of the efficiency of mRNA selective residue-specific sense codon reassignment by the OTOs. The Cy5 intensity of the vimentin::mcerulean band is divided by the corresponding immunolabel intensity in order to account for the differences in POI expression level between different samples. In the Cy5 scan a clear band for the modified POI, vimentin::mcerulean is observed for both the OTOs LCK::MCP::F::AF and LCK::MCP::ΔF::AF. The ratio of the Cy5 and immunolabel signals for vimentin::mcerulean for the sample with LCK::MCP::ΔF::AF is higher than that recorded for the sample with LCK::MCP::F::AF. This serves as an indication that the RNA binding domain of full length FUS does not have a selective affinity towards the mRNA of the reporter vimentin::mcerulean. If full length FUS did have an affinity for the mRNA of

vimentin::mcerulean, then the OTO LCK::MCP::F::AF would have had stronger recruiting ability for vimentin::mcerulean mRNA than the OTO LCK::MCP::ΔF::AF which contains truncated FUS without the RNA binding domain (ΔF), resulting in higher efficiency of sense codon reassignment for LCK::MCP::F::AF than for LCK::MCP::ΔF::AF.

Similar to all in-gel assays throughout the thesis a significant amount of unspecific labelling was observed for the NES::AF system as compared to the OTOs as shown in the Cy5 scan of the blot in figure 38. Faint bands corresponding to vimentin::mcerulean were also observed for the samples with the OTOs LCK::MCP::AF (OTO lacking the phase separating domain) and LCK::F::AF (OTO lacking the mRNA recruiting protein MCP) in the Cy5 scan of the blot in figure 38. For both LCK::MCP::AF and LCK::F::AF, the ratio of Cy5 intensity and immunolabel intensity for vimentin::mcerulean was lower than that recorded for the fully functional OTOs LCK::MCP::F::AF and LCK::MCP::ΔF::AF. Thus, eliminating either the mRNA recruiting domain or the phase separating domain has adverse effects on mRNA selective residue-specific sense codon reassignment by OTOs.

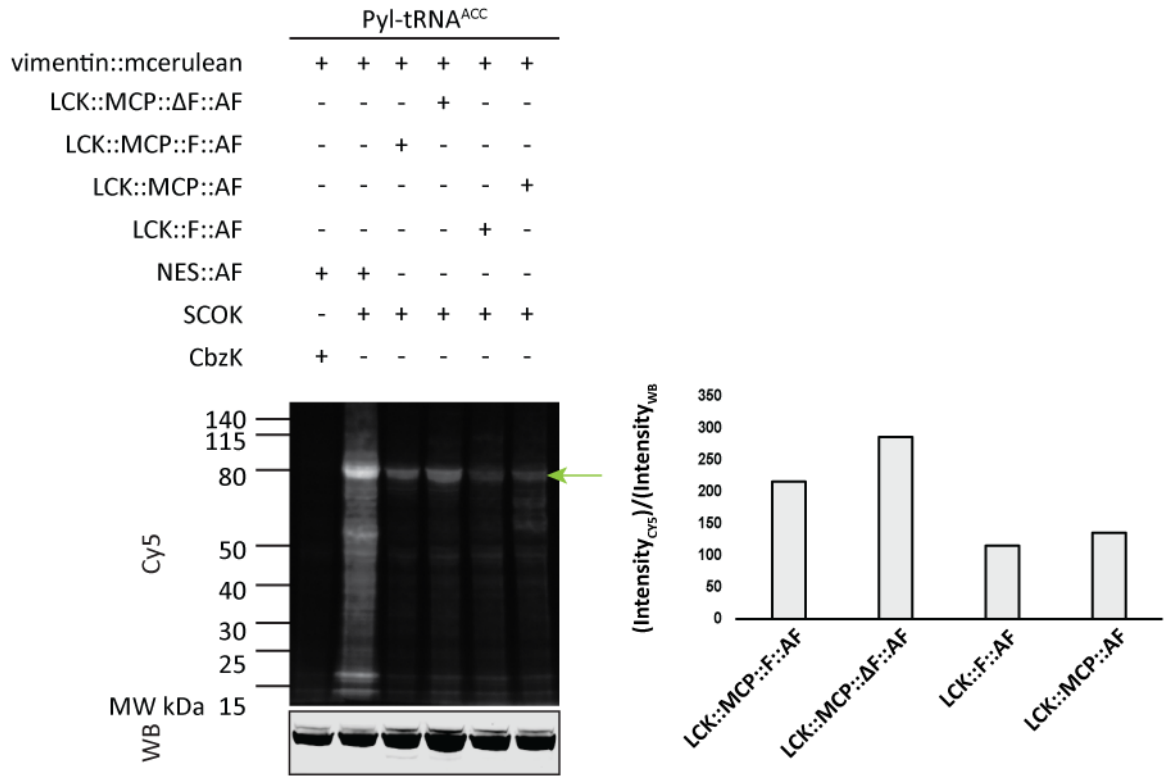


Figure 38| Result of in-gel assay comparing performance of OTOs with full length FUS (F) or truncated FUS (ΔF) (FUS lacking the RNA binding domain). HEK293T cells were transfected with plasmids expressing vimentin::mcerulean, Pyl-tRNA^{ACC} and an OTO / NES::AF respectively. The ACC codon was reassigned in vimentin::mcerulean to incorporate SCOK/CbzK. The cells lysates were labelled with Cy5-H-Tetrazine. Western blot was performed with the cell lysates and the blots were immunolabelled with anti-myc primary antibody and secondary antibody conjugated with a dye. The immunolabel served as an estimation of total amount of POI expressed. Proteins modified with SCOK were visible in the Cy5 scan of the blot. The green arrow marks the height of vimentin::mcerulean bands on the Cy5 scan of the blot. The immunolabel scan of the blot is denoted as WB. Ratio of Cy5 signal (Intensity$_{Cy5}$) and immunolabel signal (Intensity$_{WB}$) for vimentin::mcerulean bands for the samples with different OTOs have been depicted in the bar plot.

3.2.7 Labelling of vimentin::mcerulean by residue-specific sense codon reassignment in HelaK cells

The results shown in section 3.2.3, confirm successful mRNA selective residue-specific sense codon reassignment by OTOs in HEK293T cells. To prove the applicability of the method to incorporate ncAAs in POIs in cell lines other than HEK293T, ACC codons in vimentin::mcerulean (the mRNA of vimentin::mcerulean was tagged with MS2 loops) were reassigned in HeLaK cells using LCK::MCP::AF::F::AF. HeLaK cells were also chosen due to their larger size compared to HEK293T cells and consequent ease of visualizing vimentin fibers. The ACC codon was reassigned to incorporate TCO*AK in vimentin::mcerulean and the cells were labelled with LD655-H-Tetrazine which bio-orthogonally reacts with TCO*AK by IEDDA reaction. The mcerulean signal was considered as the standard label. Colocalization between mcerulean and LD655 channels would serve as a positive indication of successful incorporation of TCO*AK in vimentin::mcerulean by GCE. Figure 39 shows images of mcerulean and LD655 channels depicting clear fibrous structure of vimentin in HeLaK cells. Thus, mRNA selective residue-specific sense codon reassignment by OTOs could be extended to HeLaK cells.

Aside from in-gel and imaging experiments, validation of ncAA incorporation by mRNA selective residue-specific sense codon reassignment was also performed by mass-spectrometry. ACC codons were reassigned in vimentin::mcerulean to incorporate the ncAA 3IF by the OTO LCK::MCP::AA::F::AA in HEK293T cells. Subsequent to digestion by trypsin, fragments of vimentin::mcerulean with incorporated 3IF in predicted location could be detected by mass-spectrometry thereby confirming successful reassignment of ACC codon by sense codon reassignment (data is shown in supplementary figure 5).

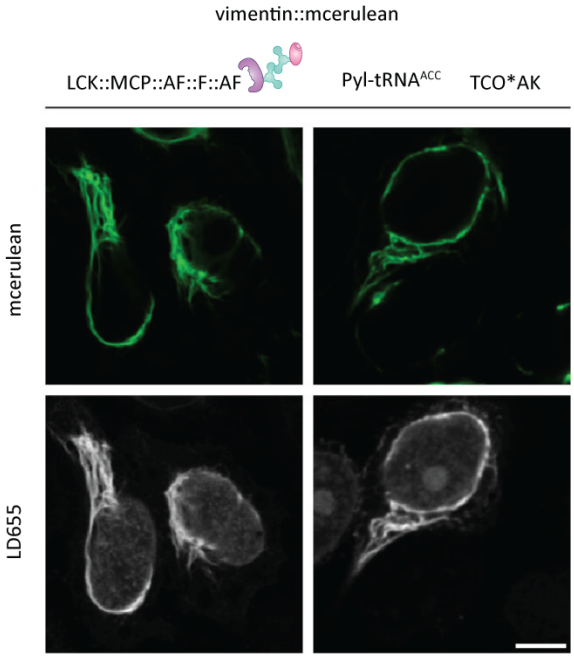


Figure 39| Labelling of vimentin::mcerulean by ACC codon reassignment in HeLaK cells using LCK::MCP::AF::F::AF. The upper row (images in green) shows the images from the mcerulean channel and the lower row (images in gray) shows images from the GCE specific, clickable dye LD655 channel.

4 Discussion

In this thesis incorporation of ncAAs in a POI by mRNA selective site-specific (sections 1.5 and 3.1) as well as residue-specific sense codon reassignment (sections 1.5 and 3.2) by OTOs (section 1.4.1.3) in mammalian cells has been accomplished. In the following sections (4.1 – 4.6), key findings of this dissertation will be discussed. A recent study by Ding *et al.*,[177] which was published during the course of my PhD, where site-specific sense codon reassignment was performed in mammalian cells will be discussed in detail in section 4.7. Finally future perspectives of the work in this thesis will be summarized in section 4.9.

4.1 OTOs render sense codon reassignment mRNA specific

4.1.1 Site-specific sense codon reassignment

In the case of site-specific sense codon reassignment, the codon to be reassigned occurs only at the predetermined location in the gene of the POI. Selective sense codon reassignment was demonstrated using in-gel assay and confocal microscopy experiments. The in-gel assay was conducted with two EGFP reporters expressed simultaneously from the same plasmid. One of the reporters, designated as extended EGFP ($EGFP^{39X}$::$EGFP^{66P}$, where X refers to the ncAA incorporated), was a fusion protein comprising of a normal EGFP and an inactive EGFP containing a mutation at site 66. The second reporter was only a normal EGFP ($EGFP^{39X}$, where X refers to the ncAA incorporated). In $EGFP^{39X}$::$EGFP^{66P}$, fusion of the inactive EGFP to the normal EGFP created a difference in molecular weight between $EGFP^{39X}$::$EGFP^{66P}$ (extended EGFP) and $EGFP^{39X}$ (normal EGFP). Thus, it was possible to separately detect the extended EGFP and the normal EGFP in SDS-PAGE. The mRNAs of both the reporters contained distinct RNA motifs that could interact specifically with their cognate RBDs. Depending on which RBD was enriched in the OTO, only one of the mRNAs should get recruited into the OTO and ideally only the corresponding reporter protein should be modified with an ncAA (figure 7). The ncAA could be detected by labelling with tetrazine functionalized Cy5 dye capable of bio-orthogonally reacting with the ncAA. Site 39 of the normal EGFPs in both the reporters was chosen for incorporation of ncAA. The extended and normal EGFP reporters could be used to estimate the selectivity of site-specific sense codon reassignment in HEK293T cells.

Two sets of in-gel assays were performed with the two EGFP reporters where the combination of the reporters and the RBD binding RNA loops were interchanged. In in-gel assay A (section 3.1.2, figure 9), the mRNA of the extended EGFP reporter was fused with MS2 loops and was recruited to OTOs enriched with MCP whereas the mRNA of the normal EGFP reporter was fused with BoxB loops and was recruited to OTOs enriched with λN22. Hence for in-gel assay A ideally the extended EGFP should be modified with the desired ncAA when OTO with MCP is used and the normal EGFP should be modified with an ncAA when OTO with λN22 is used. mRNA selective site-specific sense codon reassignment was validated for a set of 8 codons, CTA, TTA, ATA, TCG, TGC, ACG, CGT and CCG in in-gel assay A. Initially rare codons (CTA, TTA, ATA, TCG, ACG. CGT and CCG) were chosen for site-specific sense codon reassignment. The abundance of isoacceptor

tRNAs decoding rare codons is lower as compared to that for the abundant codons.[182] Hence performing sense codon reassignment with rare codons would have lower competition from respective endogenous tRNAs. However, I wanted to develop mRNA selective site-specific sense codon reassignment as a method applicable to both rare and abundant codons. Hence the abundant codon TGC was added to show that it is possible to perform mRNA selective site-specific sense codon reassignment with abundant codons as well. To detect the presence of the incorporated ncAA SCOK in the POI, cell lysates were labelled with Cy5-H-Tetrazine which can click-react with SCOK, followed by western blot. Cy5 scan of the blots (figure 9) show, for the 8 codons, a significant reduction in unspecific background labelling in the samples with OTOs as compared to the samples where GCE had been performed with non-specific cytoplasmic NES::AF. Unspecific background Cy5 labelling signifies unspecific modification of the host proteome by GCE. Selectivity of sense codon reassignment with OTOs as compared to cytoplasmic NES::AF was calculated as described in section 3.1.2. The highest fold change selectivity of sense codon reassignment by the OTO LCK::λN22::AF::F::AF was recorded as 14fold for the sense codon CTA and the highest fold change selectivity for the OTO LCK::MCP::AF::F::AF was recorded as 9fold for the sense codon CTA.

In in-gel assay B (section 3.1.2, figure 10), the mRNA of the extended EGFP was tagged with BoxB loops and hence recruited to OTOs enriched with λN22. On the other hand, the mRNA of the normal EGFP was tagged with MS2 loops and was recruited to OTOs enriched with MCP. In contrast to in-gel assay A, in in-gel assay B the extended EGFP would be modified with an ncAA when OTOs with λN22 are used and the normal EGFP would be modified when OTOs with MCP are used. A set of 5 codons, CTA, CGT, CCG, TGC and ATA, were reassigned in in-gel assay B (due to time constrains EGFP reporters for 5 sense codons were cloned for in-gel assay B). The Cy5-H-Tetrazine labelling of cell lysate and subsequent selectivity estimation was performed similar to in-gel assay A. All the samples in in-gel assay B where sense codon reassignment was performed with the cytoplasmic, non-selective NES::AF, much higher unspecific Cy5 labelling was observed compared to samples where the OTOs were utilized(figure 10). The highest fold change selectivity of sense codon reassignment with λN22-based OTO (LCK::λN22::AF::F::AF) as compared to NES::AF was recorded as 8fold for the codon CGT while the highest fold change for MCP-based OTO (LCK::MCP::AF::F::AF) was recorded as 8fold for the codon CTA.

Effective application of the OTOs to render site-specific sense codon reassignment mRNA selective was also validated by confocal microscopy experiments (section 3.1.3). The microscopy experiments were performed with the fusion construct vimentin::mcerulean (the mRNA of vimentin::mcerulean was tagged with MS2 loops), where the CTA codon was reassigned to incorporate TCO*AK at position 164 in vimentin using the OTO, LCK::MCP::AF::F::AF in HEK293T cells. The cells were subsequently labelled with LD655-H-Tetrazine which can react specifically with TCO*AK by IEDDA (section 1.3.2.3) reaction. Colocalized images were observed on exciting mcerulean and LD655 for samples with OTO, which served as a positive indication that the reporter vimentin::mcerulean could be modified with the ncAA TCO*AK (figure 13). In contrast to

samples with OTO, those with the cytoplasmic NES::AF showed higher unspecific background LD655 labelling. Fold change selectivity of site-specific sense codon reassignment with OTO as compared to reassignment with the non-selective, cytoplasmic NES::AF was calculated as discussed in section 3.1.3. A fold change selectivity higher than 1 would indicate that the OTO can perform site-specific sense codon reassignment in an mRNA selective manner. The fold change selectivity for site-specific sense codon reassignment with OTO LCK::MCP::AF::F::AF in comparison to reassignment with NES::AF was recorded to be 1.4fold. The fold change selectivity of 1.4fold represents the mean obtained from two biological replicates. For each biological replicate average selectivity values (defined in section 3.1.3) from at least 3 imaging frames were used to calculate the fold change selectivity. The error bar in figure 13 represents the standard deviation of the fold change selectivity values obtained from the two biological replicates of the confocal microscopy-based experiment to demonstrate the applicability of the OTO to render mRNA selectivity to site-specific sense codon reassignment. The standard deviation of the fold change selectivity values for site-specific sense codon reassignment with the OTO LCK::MCP::AF::F::AF as compared to NES::AF was 0.02.

Hence both the in-gel assays A and B and the confocal microscopy experiments showed that mRNA selective site-specific sense codon reassignment could be achieved by confining GCE to the OTOs.

4.1.2 Residue-specific sense codon reassignment

In residue-specific sense codon reassignment, a sense codon, XXX, is selected for reassignment and all XXX codons occurring naturally in the gene of the POI are attempted to be repurposed to incorporate ncAAs. As discussed in section 1.4.2.3, this technique has been widely used for proteome wide incorporation of ncAA in a host,[172–176] however, to the best of my knowledge, so far mRNA selective residue-specific sense codon reassignment has not been performed.

In this thesis in-gel assays were performed to validate mRNA selective residue-specific sense codon reassignment using OTOs in mammalian cells. 6 codons, namely TAC, ACC, AGC, ATC, ACG and ACT, were reassigned individually to incorporate SCOK in the fusion construct vimentin::mcerulean (section 3.2.3). These 6 codons were chosen based on the results observed in sections 3.2.1 and 3.2.2. TAC, ACC, AGC, ATC and ACT were some of the codons that were found suitable for reassignment since for these codons it was possible to observe colocalized images between the mcerulean and LD655 channels (figures 17, 23, 29 and 30). Colocalization between mcerulean and LD655 images is a positive indication that it was possible to label vimentin::mcerulean by GCE. The selectivity of residue-specific sense codon reassignment for codons TAC, ACC, AGC, ATC and ACT were also relatively high amongst the 61 codons that were individually reassigned in vimentin::mcerulean (figure 31). Additionally in the in-gel assay performed in section 3.2.2, a clear band corresponding to vimentin::mcerulean could be observed in the Cy5 scan of the blot when the codons TAC, ACC, AGC, ATC and ACT were individually reassigned in vimentin::mcerulean. The observation of a band for vimentin::mcerulean in the Cy5 scan of the blot signifies successful incorporation of a

clickable ncAA in vimentin::mcerulean. On the other hand the ACG codon was chosen to represent sense codons that were deemed unsuitable for efficient sense codon reassignment in vimentin::mcerulean.

The cell lysate was subsequently labelled with Cy5-H-Tetrazine which can bio-orthogonally react with SCOK by IEDDA reaction. On performing western blot, the proteins modified with SCOK could be visualized by Cy5 scan of the blot. Figure 34 shows that for all the 6 codons there was significant decrease in unspecific background Cy5 labelling in samples with OTO LCK::MCP::AF::F::AF as compared to the samples with the cytoplasmic NES::AF, thus validating the success of OTOs in rendering mRNA selectivity to residue-specific sense codon reassignment.

4.2 BoxB-λN22 vs MS2-MCP for mRNA recruitment into the OTO

mRNA selective site-specific sense codon reassignment has been demonstrated with two OTOs (section 1.4.1.3) differing in their mRNA recruitment machinery (section 3.1.2). The OTO LCK::MCP::AF::F::AF utilized the MCP-MS2 interaction to recruit target mRNA into the OTO while the OTO LCK::λN22::AF::F::AF recruited mRNA by the λN22-BoxB interaction. Figure 9 depicts the results of in-gel assay A (described in sections 3.1.2 and 4.1.1) where ideally only the extended EGFP should be modified with an ncAA and visible in the Cy5 scan of the blot when LCK::MCP::AF::F::AF is used for sense codon reassignment, whereas only normal EGFP should be modified and visible in the Cy5 scan of the blot when GCE is performed with LCK::λN22::AF::F::AF. For 5 out of 8 codons, the fold change selectivity (calculated as described in section 3.1.2) of site-specific sense codon reassignment recorded for samples with LCK::λN22::AF::F::AF was higher than the fold change selectivity recorded for samples with LCK::MCP::AF::F::AF (figure 11). A similar trend was observed for in-gel assay B. In in-gel assay B the combination of reporter and RNA loops was interchanged as compared to in-gel assay A. In in-gel assay B higher fold change selectivity of site-specific sense codon reassignment was recorded for 4 out of 5 codons when GCE was performed by LCK::λN22::AF::F::AF (figure 12).

4.3 Use of full length FUS as the assembler domain of the OTOs

The OTOs used throughout this thesis were formed by the combined effect of spatial targeting and phase separation. During the initial development of the OTOs, Reinkemeier *et al.* had tested multiple strategies for OTO assembly including only spatial targeting to a predetermined subcellular localization, only phase separation and combination of both approaches.[4] The highest selectivity and efficiency for amber suppression was achieved by an organelle formed by utilizing full length FUS as the phase separating domain and truncated kinesin for localization to microtubule plus ends. In a follow up study by Reinkemeier *et al.*, improved film-like OTOs were developed by phase-separation on different membrane surfaces which also used full length FUS as the phase separating domain.[151] Two such highly mRNA selective and efficient OTOs, LCK::MCP::AF::F::AF and LCK::λN22::AF::F::AF were chosen for performing sense codon reassignment in this thesis. Full length FUS has proven to contribute positively towards the selectivity and efficiency of the OTOs. However, since FUS is a RNA binding protein,[181] it might be possible that it has a natural affinity for mRNA of any of the reporters used or might recruit

unspecific mRNAs into the OTO. In-gel assays were performed to rule out such scenarios for both site-specific as well as residue-specific sense codon reassignment and the outcome will be discussed in sections 4.3.1 and 4.3.2.

4.3.1 Site-specific sense codon reassignment

In-gel assays A and B were performed to quantify the selectivity of site-specific sense codon reassignment in HEK293T cells (section 3.1.2, figures 9,10,11 & 12 and section 4.1.1). Two sets of EGFP based reporters were used for the in-gel assays A and B. The combination of reporters and RNA loops necessary for mRNA recruitment to the OTOs was interchanged for the reporters used in in-gel assays A and B as follows:

In-gel assay A: extended EGFP mRNA was tagged with MS2 loops and normal EGFP mRNA was tagged with BoxB loops. Hence upon using the OTO LCK::MCP::AF::F::AF for GCE, ideally, only the extended EGFP should be modified with the desired ncAA and on using LCK::λN22::AF::F::AF only the normal EGFP should be modified with the ncAA.

In-gel assay B: extended EGFP mRNA was tagged with BoxB loops and normal EGFP mRNA was tagged with MS2 loops. Hence upon using the OTO LCK::MCP::AF::F::AF, ideally, only the normal EGFP should be modified with the desired ncAA and on using LCK::λN22::AF::F::AF only the extended EGFP should be modified with the ncAA.

For both in-gel assays A and B, on labelling cell lysates with Cy5-H-Tetrazine followed by western blot, only proteins with incorporated clickable ncAA SCOK should be visible in the Cy5 scan of the blot since Cy5-H-Tetrazine bio-orthogonally reacts with SCOK.

In figures 9 & 10, the effect of interchanging the reporter-RNA loop combination is clearly reflected. In figure 9, for the 8 sense codons (CTA, ATA, CGT, CCG, TTA, TCG, TGC and ACG), the Cy5 scan of the blots primarily show modified extended EGFP for samples with LCK::MCP::AF::F::AF and modified normal EGFP for samples with LCK::λN22::AF::F::AF. Only in case of CCG and ACG codons, the efficiency of sense codon reassignment was low and clear bands corresponding to extended EGFP was not visible for samples with LCK::MCP::AF::F::AF. In line with the expectation, figure 10 shows, for the 5 sense codons (CTA, ATA, TGC, CGT and CCG) primarily the modified normal EGFP is visible in the Cy5 scan of the blots for samples with the OTO LCK::MCP::AF::F::AF and modified extended EGFP is visible for samples with LCK::λN22::AF::F::AF. The outcome of in-gel assays A and B clearly show that the mRNA recruitment to the OTOs is dependent solely on the interaction of the RNA binding domain of the OTOs (MCP or λN22) and the RNA loops MS2 or BoxB. Bias towards recruitment of either one of the EGFP reporter mRNAs due to the presence of full length FUS was not detected. Additionally, site-specific sense codon reassignment, for both in-gel assay A and B, was also performed with an OTO with impaired mRNA recruitment ability. LCK::AF::F::AF lacks the RBD MCP or λN22 and as expected, in all samples with LCK::AF::F::AF faint or no bands corresponding to modified reporters were visible in the Cy5 blots (figures 9 & 10). In some cases faint bands were visible for samples with LCK::AF::F::AF. A possible reason for the faint bands in the Cy5 blot, despite lack of mRNA recruitment ability of LCK::AF::F::AF could be that since the OTOs are not membrane-bound, a subset of the overexpressed reporter mRNAs can still gain access into the OTOs, in turn generating GCE modified reporters. Also, unspecific

GCE modified protein bands were not detected for samples with LCK::AF::F::AF. Hence it can be concluded that full length FUS did not have any detectable contribution in recruiting unspecific mRNAs into the OTO or bias towards recruitment of either of the EGFP reporters.

4.3.2 Residue-specific sense codon reassignment

Most of the assays (except some samples in the in-gel assay described in section 3.2.6, figure 38) for mRNA selective residue-specific sense codon reassignment were performed with the OTO LCK::MCP::AF::F::AF. Similar to site-specific sense codon reassignment, experiments were also performed for mRNA selective residue-specific sense codon reassignment to show that the RNA binding domain of full length FUS in the OTO did not have a bias towards the reporter mRNA (vimentin::mcerulean) or recruit unspecific mRNAs into the OTO. Section 3.2.3, figure 34 (also discussed in section 4.1.2), shows the result of an in-gel assay performed where selected sense codons were individually reassigned to incorporate SCOK in vimentin::mcerulean using OTOs in HEK293T cells. The cell lysates were labelled with Cy5-H-Tetrazine followed by western blot. For all the 6 sense codons (TAC, ACC, AGC, ATC, ACG and ACT), samples were included where sense codon reassignment was performed with the OTO LCK::AF::F::AF (LCK::AF::F::AF lacks MCP and cannot selectively recruit MS2 tagged vimentin::mcerulean mRNA). Bands corresponding to GCE modified vimentin::mcerulean was either not observed in the Cy5 scan of the blots (Cy5-H-tetrazine click reacts with compatible ncAAs like SCOK, hence only proteins modified with clickable ncAAs can be visualized in Cy5 scan of the blot) or faint bands were observed. Fold change yield of GCE modified reporter (defined in section 3.2.3, in all cases differences in protein expression levels were taken into account) for samples with fully functional OTO LCK::MCP::AF::F::AF as compared to samples with LCK::AF::F::AF was higher than 1 for all 6 sense codons. The highest fold change yield of about 11fold was recorded for TAC codon and the lowest of about 3fold was recorded for ACG codon. Fold change yield higher than 1 signifies higher expression of GCE modified vimentin::mcerulean for samples with LCK::MCP::AF::F::AF as compared to samples with LCK::AF::F::AF. Hence, selective mRNA recruitment to OTO is predominantly dependent on MCP-MS2 interaction and affinity of full length FUS towards vimentin::mcerulean mRNA was not detected.

In section 3.2.6, figure 38 in-gel assay was also performed where ACC codon was reassigned in vimentin::mcerulean with an OTO designed with only the disordered domain of FUS (designated as ΔF), LCK::MCP::ΔF::AF. The efficiency of sense codon reassignment by LCK::MCP::ΔF::AF was compared with the performance of the OTO with full length FUS, LCK::MCP::F::AF (figure 38). LCK::MCP::ΔF:AF recorded slightly higher yield of GCE modified reporter (defined in section 3.2.6 as $Intensity_{Cy5}/Intensity_{WB}$, where WB stands for western blot and signifies intensity of the immunolabel signal. The signal from the immunolabel is an estimation of total amount of vimentin::mcerulean expressed, both modified and unmodified) as compared to LCK::MCP::F::AF. Hence affinity of full length FUS towards vimentin::mcerulean mRNA and the possibility of full length FUS enhancing vimentin::mcerulean mRNA recruitment for the OTO LCK::MCP::F::AF could be ruled out. Furthermore, lower yield of GCE modified

vimentin::mcerulean was recorded for samples with the OTO LCK::F::AF (lacking mRNA recruiting RBD, MCP) compared to the yield obtained on performing GCE with either LCK::MCP::ΔF::AF or LCK::MCP::F::AF. Hence showing that the interaction of MCP-MS2 is driving the selective recruitment of MS2 tagged vimentin::mcerulean mRNA into the OTO.

In figure 34, samples were included for all the selected sense codons where the reporter vimentin::mcerulean was excluded, only the OTO LCK::MCP::AF::F::AF and the Pyl-tRNA corresponding to the sense codon to be reassigned was included. No unspecific bands were observed in the Cy5 scan of the blot for any of the codons which shows that full length FUS did not recruit unspecific mRNAs to the OTO to a detectable extent.

In this thesis sense codon reassignment (both site and residue-specific) was performed with different reporters and in all cases it could be shown that the selectivity of the OTOs relied on the interaction of the RBD of the OTO and compatible RNA loops. Similar outcome can also be drawn for stop codon suppression from the works of Reinkemeier *et al*.[151] The applicability of the OTOs to perform mRNA selective GCE is not reliant on the usage of reporters, the mRNAs of which might have an affinity to full length FUS. Detectable unspecific GCE due to recruitment of non-target mRNAs was also not observed in this thesis. That being said, improvements are always possible and the design of the OTOs can be optimized further as discussed in section 4.9.1.

4.4 Multiple ncAA incorporation by residue-specific sense codon reassignment

mRNA selective residue-specific sense codon reassignment attempts to repurpose all the naturally occurring XXX (sense codon selected for reassignment) codons in the mRNA of a POI to incorporate desired ncAAs. Hence in principle this method should allow multiple ncAA incorporation in a POI. To estimate the number of ncAAs that could, in practice, be incorporated by residue-specific sense codon reassignment, an in-gel assay was performed by repurposing the ACC codon in vimentin::mcerulean by the OTO LCK::MCP::AF::F::AF in HEK293T cells. The ACC codon reassignment in vimentin::mcerulean was performed for 10 different ncAA concentrations; the ncAA concentration ranged from 0 µM to 2 mM. Efficacy of ACC codon reassignment was evaluated against amber (TAG) suppression of vimentin^{116TAG}::mcerulean by LCK::MCP::AF::F::AF in HEK293T cells.(section 3.2.4, figure 36) The basic idea was that vimentin^{116TAG}::mcerulean has only one site for ncAA incorporation. Hence on subsequent labelling of cell lysate with clickable dye Cy5-H-Tetrazine, which can bio-orthogonally react with the incorporated ncAA SCOK, only one dye molecule would be introduced per molecule of vimentin^{116TAG}::mcerulean. However since the ACC codon occurs multiple times in vimentin::mcerulean gene, there is the possibility of incorporating multiple ncAAs and subsequently introducing multiple dyes per molecule of vimentin::mcerulean on labelling the cell lysate with Cy5-H-Tetrazine. On performing western blot with the labelled cell lysates, Cy5 scan of the blot would show only the proteins modified with the ncAA SCOK. On the other hand immunolabelling the blot against myc tag fused N-terminally to both vimentin^{116TAG}::mcerulean as well as vimentin::mcerulean would show total amount of both proteins, thereby serving as a measure for the expression levels. Now

if we consider the ratio of the Cy5 intensity and the immunolabel intensity (the Cy5 intensity is divided by the immunolabel intensity to account for the variation in expression levels between different samples) of the vimentin^{116TAG}::mcerulean band as one (since every labelled vimentin^{116TAG}::mcerulean molecule can have only one dye incorporated); the fold change for the same value(Cy5 intensity/immunolabel intensity) for the vimentin::mcerulean bands should give an estimation of the average number of dyes introduced in vimentin::mcerulean by ACC codon reassignment. A fold change intensity value higher than 1 would be an indication of incorporation of multiple dye molecules, hence multiple ncAA incorporation in vimentin::mcerulean. However as shown in figure 36, the fold change intensity values for all vimentin::mcerulean samples were lower than 1, the highest being 0.6fold. Hence there was no conclusive proof of multiple ncAA incorporation by mRNA selective residue-specific sense codon reassignment.

A possible reason for inability to incorporate multiple ncAAs in a POI could be overall low efficiency of residue-specific sense codon reassignment by the OTOs. A major hindrance to sense codon reassignment is competition from endogenous tRNAs leading to incorporation of canonical amino acids instead of ncAAs. Hence lowering the amount of respective endogenous tRNAs by knocking down or increasing expression of the GCE specific tRNAs would be beneficial for improving efficiency of residue-specific sense codon reassignment. Efforts directed towards increasing the expression of GCE specific tRNAs and the outcome have been covered in section 3.2.5, figure 37 as well as in section 4.5.

4.5 Effect of increasing tRNA expression on residue-specific sense codon reassignment

An important aspect of mRNA selective residue-specific sense codon reassignment was to enable the incorporation of multiple ncAAs in a POI. In-gel assay performed to determine the number of ncAAs incorporated in vimentin::mcerulean by reassigning the ACC codon by the OTO LCK::MCP::AF::F::AF (section 3.2.4 figure 36 and also discussed in section 4.4) did not show indication of more than one ncAA being incorporated. The first approach, that was explored in this thesis, to facilitate multiple ncAA incorporation was to devise strategies to improve the efficiency of mRNA selective residue-specific sense codon reassignment.

A common measure for GCE enhancement is increasing the amount of GCE specific tRNAs, as had been shown by multiple studies. Parrish *et al*. compared amber suppression efficiency in *C. elegans* by generating two different *C. elegans* strains stably expressing either a single copy of EcLeu-tRNA or three copies of EcLeu-tRNA.[183] A Luciferase containing an amber codon was used as the reporter and the strain of *C.elegans* expressing three copies of EcLeu-tRNA showed 19fold higher luciferase activity as compared to the strain expressing a single copy of EcLeu-tRNA. A similar increase in stop codon suppression in mammalian cells by increasing suppressor tRNA expression was demonstrated by Coin *et al*. [184] and Schmied *et al*. [185] among others. Multiple copies of GCE specific tRNA were also utilized to improve efficiency of sense codon reassignment in the recent work by Ding *et al*.[177] Hence, I sought to perform OTO

enabled mRNA selective residue-specific sense codon reassignment in HEK293T cells by transfecting plasmids encoding varying number of GCE specific tRNA genes.

Four different plasmids were cloned carrying two, three, five and seven copies respectively of the Pyl-tRNA gene with the mutated anticodon TGG. In-gel assay was performed (section 3.2.5, figure 37) where plasmids encoding one, two, three, five and seven copies of the Pyl-tRNA^{ACC} were used individually to reassign the ACC codons in vimentin::mcerulean by the OTO LCK::MCP::AF::F::AF in HEK293T cells. The mRNA of vimentin::mcerulean was tagged with MS2 loops to facilitate selective recruitment in the OTO LCK::MCP::AF::F::AF. The efficiency of sense codon reassignment obtained for each case was also compared against the efficiency of amber suppression in vimentin^{116TAG}::mcerulean (mRNA of vimentin^{116TAG}::mcerulean was tagged with MS2 loops) by LCK::MCP::AF::F::AF and plasmid encoding a single copy of Pyl-tRNA^{TAG}. Fold change intensity (defined in section 3.2.4) values for all the samples where ACC codon was reassigned in vimentin::mcerulean as compared to amber suppression was less than 1. A fold change intensity value lower than 1 does not indicate multiple ncAA incorporation (interpretation of fold change intensity is explained in detail in section 3.2.4 and 4.4). Unexpectedly the highest fold change intensity value amongst the vimentin::mcerulean samples was recorded for the sample where Pyl-tRNA was expressed from a plasmid encoding only a single copy of the Pyl-tRNA gene. It seemed that increasing the expression of tRNA did not have a positive effect on mRNA selective residue-specific sense codon reassignment. However, owing to time constraints the in-gel assay was only performed once. Further studies are necessary to ascertain the exact effect of increasing expression of GCE specific tRNA on selective residue-specific sense codon reassignment.

Another aspect to consider for future studies to enable multiple ncAA incorporation in a POI by residue-specific sense codon reassignment, would be the use of engineered variants of the Pyl-tRNA that have been designed to improve interaction with the mammalian translation machinery, for example the M15 and C15 Pyl-tRNA variants from Serfling *et al.*[66] (section 1.3.1.2). Pyl-tRNAs derived from the M15 and C15 variants have proven to be beneficial for sense codon reassignment as shown by Ding *et al.*[177] Alternatively Pyl-tRNA A2.1 developed by the Chatterjee lab using the VADER [67] (section 1.3.1.2) platform would also be a promising candidate since it has shown higher ncAA incorporation activity compared to the wild type Pyl-tRNA variant. The VADER platform itself could also be used to evolve new tRNA variants for enhanced sense codon reassignment.

4.6 Consideration of codon usage bias and codon context effects for selection of sense codons for reassignment

Codon usage bias and codon context effects are two phenomena that can play an important role in determining the suitability of a codon for reassignment to incorporate ncAAs. Hence codon usage bias will be discussed in detail in section 4.6.1, followed by codon context effect in section 4.6.2.

4.6.1 Role of codon usage bias in reassignment efficiency of sense codons

Most of the amino acids are coded by multiple sense codons which are referred to as synonymous codons. However, the synonymous codons are not all equally utilized. The preferential usage of certain codons over others to code for the same amino acid is known as codon usage bias.[186] This phenomenon has been hypothesized to stem from a number of factors. The GC content of the genome dictates the preferential usage of codons with G/C in the third position. Organisms with a GC rich genome, for example, *Mycobacterium tuberculosis*, will select for codons ending with G/C.[187] Similarly, the opposite effect can be observed for organisms with a low GC content genome, for example *Plasmodium falciparum*.[188] Another dominant hypothesis is that codon usage bias is related to translation efficiency. Studies have shown that codons having higher abundance of corresponding isoacceptor tRNAs tend to be preferentially selected. A comparative analysis of the codon usage bias in *E.coli* and *Saccharomyces cerevisiae* by Ikemura states that apart from the organism-specific pattern of codon usage, within the same organism, depending on the gene category, certain codons are preferred over others.[182] A clear distinction can be observed in the codon usage bias between highly expressed and low expressed genes. Upon quantification of the abundance of isoacceptor tRNAs in different organisms like *E. coli*, *Salmonella typhimurium* and *S. cerevisiae*, a positive correlation could be observed between the codon usage and the abundance of the respective tRNAs.[182]

Codon usage bias is also related to the position of a codon in the gene.[189] Close to the 5' end codons are selected with the primary goal of facilitating efficient ribosomal initiation. mRNA secondary structure within the first 40 nucleotides from the 5' end is important for ribosomal initiation and hence codons with a tendency to form strong secondary structure which can hinder the same are preferentially excluded. In such cases in *E. coli* A/T rich rare codons are often selected due to their ability to form weak mRNA secondary structure. Additionally preferential codon selection can be observed in the vicinity of intron-exon boundaries. Among other factors, 5' GT and 3' AG dinucleotides of introns are known to play an important role in facilitating correct splicing. Recognition of splicing partners may be hindered if there is AG dinucleotide at the 5' end of a non-first intron or GT at the 3' end of non-last introns. Hence at intron-exon boundaries codon usage bias exists to avoid such a situation.[190]

Thus, depending on the host organism as well as the location of the codon in the respective mRNA, certain codons will be more suitable for efficient GCE as compared to others.

4.6.2 Codon context effects

The local environment or the flanking sequences of a codon are one of the determining factors of the reassignment efficiency of the codon. This phenomenon is known as codon context effect and much work has been done to explore the role of flanking sequences in stop codon suppression.[191] In the case of stop codons, the base immediately downstream of the codon is particularly significant. In eukaryotes, purines at this position favor translation termination whereas cytosine promotes suppression of the stop

codon.[192] It is also important to note that the pattern of codon context effect differs between organisms. The effect observed in the case of humans is not the same as that demonstrated *for E. coli*.

While most studies focused mainly on the effect of the base immediately 3' of the stop codon on its suppression efficiency, Bartoschek *et al.* investigated the role of the entire local environment that is both upstream and downstream sequences of the TAG codon on amber suppression.[193] They generated mouse embryonic stem cells (mESC) as well as HEK293T cells stably expressing PylRS/Pyl-tRNA from *M. mazei*. In these engineered cell lines, with the help of the stop codon decoding PylRS/Pyl-tRNA, they used a variant of SORT (section 1.4.2.3) to suppress all the naturally occurring TAG codons. The ncAA bicyclo[6.1.0]non-4-ynyl lysine carbamate (BCNK) incorporated proteome was click-labelled with functionalized biotin. The modified proteome was enriched by biotin-streptavidin interaction. Mass spectrometry-based analysis was used to determine the amber suppression efficiency at various naturally occurring amber codons and this data was used to build an informatics-based tool for predicting amber codon context effects in mammalian cells.[193] This predictor was designated as Identification of Permissive Amber Sites for Suppression (iPASS). In line with previous studies,[191,192] Bartoschek *et al.* also found an enrichment of C at +4 position (TAG refers to +1, +2 and +3 position from 5' to 3') for efficiently suppressed TAG codons. According to this study by Bartoschek *et al.*, 6 base pairs both upstream and downstream of the amber codon have an influence over final amber suppression efficiency. The accuracy of the iPASS tool was validated by a fluorescence-based assay in multiple POIs including H2A, H3 and Dnmt3b in cell lines stably expressing necessary GCE machinery. The suppression efficiency obtained experimentally was compared to the values predicted by iPASS and a linear correlation was obtained proving the accuracy of the tool. iPASS is useful for not only predicting suppression efficiency of specific amber codons in a particular sequence context, but also for optimizing flanking sequences to improve amber suppression efficiency.

Similar to stop codons, flanking sequences also influence reassignment of sense codons as was observed by Ding *et al.* in the context of TCG codon reassignment (section 4.7.3).[177]

Throughout this thesis it has been observed that different sense codons exhibit different efficiencies of reassignment for both site- and residue-specific sense codon reassignment (sections 3.1.2, 3.2.2 and 3.2.3). Codon usage bias and codon context effects can be two of the factors influencing the suitability of different sense codons for efficient reassignment.

4.7 A review and comparison of sense codon reassignment in mammalian cells between the 2024 publication by the Lin lab (Ding *et al.*, Science, 2024) and the work done in this dissertation

The recent work by Ding *et al.*[177] is a step forward towards harnessing the potential of sense codon reassignment in eukaryotes and share similar objectives with this dissertation.[177] Hence it is relevant to comparatively discuss the strategies used in both the approaches and to identify areas of improvement. In the following subsections

different aspects of sense codon reassignment in mammalian cells have been individually reviewed. In each subsection, the relevant strategies used by Ding *et al*. have been first reviewed and then comparisons have been made with the methodology used in this thesis.

4.7.1 Selection of codons: rare vs abundant

As mentioned earlier in the thesis, one of the major challenges of sense codon reassignment is the competition from endogenous tRNAs. In rare codons, the corresponding tRNAs are also less abundant. That is certainly an incentive for using only rare sense codons for GCE, however that would mean limiting oneself to only a subset of the 61 available codons.

Ding *et al*. plotted codon frequency of occurence (occurrence/1000 codons) of the human genome against abundance of the respective tRNAs. Guided by this plot they selected seven of the rarest codons namely TCG, ACG, CCG, GTA, CTA, GCG and CGT. These codons were further evaluated for unspecific ncAA incorporation in the host proteome of HEK293T cells by dot blot. Pyl-tRNA^{XXX} (XXX- sense codon decoded by the Pyl-tRNA) from *M. mazei* and an engineered variant of the cognate PylRS (TetRS) were expressed in HEK293T cells to facilitate proteome wide incorporation of the ncAA 3-(6-butyl-1,2,4,5-tetrazin-3-yl)-phenylalanine (TetBu) by repurposing each of the seven selected rare codons individually. TetBu, by virtue of a tetrazine moiety, can click-react with TCO functionalized biotin, thereby allowing the detection of the modified proteome by biotin-streptavidin interaction on a dot blot. On quantification of the relative intensities, reassignment of CTA and TCG codons resulted in the least proteome wide ncAA incorporation. Furthermore, site-specific sense codon reassignment was performed in EGFP to estimate recoding rate of the seven rare codons by mass-spectrometry and western blot assays. In both cases recoding rate was calculated as follows:

$$\frac{Intensity_{ncAA\ incorporated\ POI}}{(Intensity_{ncAA\ incorporated\ POI} + Intensity_{wild\ type\ POI})} \times 100\%$$

TCG codon had the highest recoding rate. Based on the estimation of background incorporation and recoding rate, the TCG codon was chosen for further experiments.

In this thesis, on the other hand, sense codons were not selected based on their rarity of occurrence in the host genome. Both for site-specific as well as residue-specific sense codon reassignment, especially for the later, rare and abundant codons were used (section 3). For residue-specific sense codon reassignment, the codons were selected based on the POI and with a focus on subsequent fluorescent labelling-based application (section 3.2.1). To utilize the incorporated ncAAs for introducing probes by bio-orthogonal reactions, it is important for the ncAA to be accessible to the probes. It is possible that after folding of the POI, the incorporated ncAA is no longer sufficiently accessible for subsequent labelling steps. In this case even with high recoding rate of a certain codon, the final application will not be successful. For multiple ncAA incorporation using residue-specific sense codon reassignment, it might also be relevant to take into account how many times a certain codon occurs in the DNA sequence of the POI. Hence,

depending on the goal, it is relevant to perform codon selection for every POI. It was observed in section 3.2.1 and in the Appendix I that reassignment of same codons in vimentin::mcerulean and NUP153::EGFP did not result in similar extent of success.

4.7.2 Mitigating unspecific background incorporation of ncAA in the host proteome

Reassigning sense codons without mRNA specificity will lead to unspecific ncAA incorporation in the host proteome. In this respect using sense codons with low abundance does help in mitigating the issue to a certain extent. Ding *et al.* evaluated the background incorporation for reassigning seven rare codons as discussed in section 4.7.1 and the outcome acted as one of the selection parameters for the final codon of choice. They also observed by ribo-seq that overexpressing the reporter resulted in engaging majority of the ribosome (50%) for translation of the reporter. This phenomenon combined with enhanced GCE machinery (discussed in more detail in section 4.7.3) and either knocking down of one of the TCG decoding endogenous genes or depletion of serine (canonical amino acid coded by TCG) from the growth media, resulted in boosting selective TCG codon reassignment in the POI. Additionally they have also designed a FUS::λN22 based organelle system to confine GCE in a membraneless organelle in line with the OTO concept developed by Reinkemeier *et al.*[4,151] and used throughout the work of this dissertation. This FUS::λN22 organelle was indeed shown to decrease the background incorporation further. They did not use the organelle for further experiments due to the complexity of the system. However, since the organelle performed quite well in reducing unspecific ncAA incorporation, as observed by Ding *et al.,* the utilization of the organelles would have a strong potential to enable the optimized site-specific sense codon reassignment process developed by Ding *et al.* to be used for applications such as selective fluorescent labelling of POIs.

In this thesis, the OTO system has been utilized to render mRNA specificity to sense codon reassignment. It has been consistently shown in this thesis (sections 3.1.2 figures 9-12, section 3.1.3, figure 13) that despite using rare codons, there is considerable background labelling and the OTO could significantly lower the non-specific host proteome modification leading to a highest fold change selectivity of 14fold for site-specific sense codon reassignment by OTO as compared to the cytoplasmic non-selective NES::AF (section 3.1.2, figure 11). The applicability of the OTOs for mitigating unspecific GCE has also been discussed in section 4.1.

4.7.3 Strategies to enhance efficiency of sense codon reassignment

Out of the seven rare codons selected by Ding *et al.*, the highest recoding rate of 8.5% was reached by the TCG codon. Effort directed by Ding *et al.* towards boosting the recording rate was focused on engineering the GCE machinery, deprivation of respective canonical amino acid and optimization of the flanking sequence of the TCG codon based on codon context effects. Each of the strategies will be discussed individually in this section.

Enhancement of GCE-specific tRNA expression as well as engineering of the tRNA to improve its interaction with the translation machinery constitute one of the common approaches for GCE optimization.(section 1.3.1.2) Here they used engineered variants of the Pyl-tRNA, C15 and M15 from Serfling *et al.* (section 1.3.1.2) as the starting point and

designed several variants based on C15, M15 as well as fusion of the C15 and M15 Pyl-tRNAs. The best performing Pyl-tRNA was a variant of the C15 and M15 fusion, designated as Pyl-tRNA-CM15. Application of the Pyl-tRNA-CM15 along with the NES::TetRS led to 3.5fold increase in the recoding rate(defined in 4.7.1). Subsequently a further boost to 67% recoding rate was achieved by expressing instead of one, copies of Pyl-tRNA-CM15.

Besides designing more efficient GCE machinery, lowering the incorporation of canonical amino acids into the POI is another approach for enhancing the recoding rate. This can be achieved by either decreasing the respective endogenous tRNA or using growth media deprived of the respective canonical amino acid, in this case, serine. The Ser-tRNA^{TGC} is coded by four alleles in humans. Ding *et al*. generated stable cell lines by knocking out the alleles individually. The cell line with knocked out allele TRS-CG-A4-1, designated as 4C23, was selected as the best performing cell line. A recoding rate of 88% was achieved by the combined effect of using optimized GCE machinery in the 4C23 cell line. Alternatively, the use of growth media devoid of serine led to a recoding rate of 83.5%.

The recoding rate also depends, to a considerable extent, on the flanking sequence of the codon to be reassigned as discussed in section 3.6.2. Similar to amber suppression, the relevance of the flanking regions was also observed for TCG codon reassignment by Ding *et al*. Depending on the position of the TCG codon in the EGFP reporter the recoding rate varied from 11% to 83.5%. To develop a computational tool to optimize flanking sequences for TCG codon reassignment they designed a pipeline comprising of proteomics studies and a linear regression model. First the ncAA {Ne-3-[(3-methyl-3H-diazirine-3-yl)]-propaminocarbonyl-g-seleno-l-lysine} DiZSeK was incorporated proteome wide at TCG codons by suitable GCE machinery. The modified proteome was analyzed by mass spectrometry and the data was used to study the probability of occurrence of the base pairs surrounding the reassigned TCG codons. Additionally, a set of random flanking sequences were also generated for the TCG codon in the EGFP reporter and subsequently the recoding rate of the TCG was determined. Both these datasets were used to develop a linear regression algorithm-based predictor to design optimum flanking sequences (6 base pairs both upstream and downstream of the TCG) for enhanced TCG codon reassignment.

The strategy of canonical amino acid starvation is simple and effective; however, it is not applicable for essential canonical amino acids. For example, the ACC codon has been used for reassignment in multiple experiments in section 3.2 of this thesis. ACC codon codes for threonine and since it is an essential amino acid, starvation of the same would be detrimental to the cells. The effect of optimizing flanking codon sequence was not explored in this thesis. Site-specific sense codon reassignment can certainly benefit from engineering POI genes based on codon context effect considerations. However, one of the motivations for developing residue-specific sense codon reassignment was to enable labelling of POIs with minimal or no engineering at the gene level. The idea is to finally extend the method to conveniently label endogenous proteins without the need to mutate the corresponding gene. Hence for further development and optimization of mRNA

selective residue-specific sense codon reassignment focus should be directed towards engineering of the GCE machinery.

4.7.4 Incorporation of multiple ncAAs by codon reassignment

As a result of the cumulated effect of several optimization steps, Ding *et al*. were successful in incorporating a single ncAA multiple times by reassigning TCG codons as well as multiple distinct ncAAs by repurposing a combination of TCG and stop codons in a POI. Upto six TCG codons were reassigned to incorporate TetBu in an EGFP reporter and the performance was compared to amber suppression. Higher yield of ncAA incorporated POI was observed for TCG codon reassignment as compared to amber suppression for encoding one, three and six ncAAs. The highest recoding rate of 91% was achieved for POI containing three ncAAs. Furthermore, simultaneous repurposing of stop as well as sense codons was performed to incorporate multiple distinct ncAAs, namely 5-methoxy-L-tryptophan (5MTP), OMeY, BOCK and TetBu in an EGFP reporter. In this case the yield of POI modified with four distinct ncAAs was not compared with the yield of wild type POI.

Multiple ncAA incorporation could not be observed in this thesis. Competition from endogenous tRNAs could be one of the reasons. Increasing expression of GCE-specific Pyl-tRNA did not lead to an improvement of the efficiency of mRNA selective residue-specific sense codon reassignment (section 3.2.5, figure 37). Increasing Pyl-tRNA expression further combined with knocking down of corresponding endogenous tRNAs would be the recommended next step towards optimizing sense codon reassignment both in a site-specific and residue-specific manner.

4.8 Sense codon vs stop codon reassignment

Currently stop codon suppression, especially amber (TAG) codon, is the widely adopted means of performing GCE in eukaryotes. The use of sense codons for synthesizing proteins with ncAAs was, until very recently (section 4.7), restricted to proteome wide residue-specific sense codon reassignment for methods like BONCAT, FUNCAT and SORT (section 1.4.2.3). Stop codon suppression, despite being a powerful protein modification technology, suffers from a number of limitations as covered in section 1.4. Sense codon reassignment holds the potential to mitigate some of the drawbacks of stop codon suppression. As has been emphasized multiple times in this thesis, one of the main advantages of sense codon reassignment over stop codon suppression is the higher availability of codons which can be repurposed to incorporate ncAAs in POIs. Going beyond the three stop codons is imperative for unleashing the full potential of GCE and has been discussed in section 4.9.2.3. However there are also other aspects of GCE where sense codon reassignment shines.

Stop codon suppression generates truncated POI on failure of GCE. Competition of the GCE machinery with the release factor often results in premature termination of translation at the stop codon intended for suppression, thereby generating C-terminally truncated POI.[185] Additionally, owing to the existence of alternate translation initiation sites, for example, internal ribosome recognition elements, translation can also begin after the stop codon intended for suppression, resulting in N-terminally truncated POIs.[194]These truncated POIs can have an unknown extent of toxicity towards the host.

On the other hand, in the case of sense codon reassignment truncated POIs are not generated. Instead, the wild type POI is expressed when GCE fails. Also, stop codon suppression, unlike sense codon reassignment, is hindered by cellular pathways like the Nonsense Mediated Decay (NMD) pathway which degrades transcripts with premature stop codons.[195] Sense codon reassignment, on the other hand, suffers from competition from the endogenous tRNAs. Improvements to mitigate this limitation are necessary to harness the complete potential of sense codon reassignment.

4.9 Future perspectives

4.9.1 Modification of endogenous POIs by mRNA selective residue-specific sense codon reassignment by OTOs without any engineering of the POI gene

Currently selective recruitment of the mRNA of interest into the OTOs is facilitated by interaction of certain RNA motifs and their corresponding RNA binding proteins, for example, MS2 loops and MCP, BoxB loops and λN22 etc. The limitation of this approach is that the gene of interest always needs to be engineered to code for these RNA motifs. In this section possible alternative mRNA recruitment mechanisms will be discussed.

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNAs and corresponding CRISPR associated (Cas) proteins find widespread usage in gene editing applications.[196] Cas13 is a class 2 type VI Cas protein with the ability to degrade single stranded target RNA molecules.[197] CrRNA guided Cas13 proteins with defective RNase activity can serve as a tool for selective mRNA recruitment. The applicability of dCas13 (engineered Cas13 with impaired RNase activity) based targeting system has already been shown to carry out modifications in target RNA. Adenosine deaminase acting on RNA (ADAR) facilitates the conversion of adenosine (A) to ionosine (I). Cox *et al*. developed the RNA editing platform known as RNA Editing for Programmable A to I Replacement (REPAIR) where a dCas13 variant was fused to ADAR2 to carry out RNA-specific A to I modification.[198] Another RNA editing platform known as RNA Editing for specific C to U Exchange (RESCUE) utilized dCas13 based recruitment of an engineered variant of ADAR (ADAR2dd) to enable C to U modification in target RNA.[199] With appropriate CrRNAs dCas13 could be used to target mRNAs of interest to the OTOs. Another option would be to use the RNA binding proteins of the Puf family. These proteins contain the Pumilio Homology domains (PUM-HD) that facilitate RNA binding.[200] The PUM-HD consists of 8 repeats known as PUM repeats and each of them is responsible for recognizing an RNA base. The sequence specificity of these repeats can be altered to target these proteins to specific RNA sequences.[201] Based on the PUM-HDs, the Boyden lab developed a set of four canonical protein modules, each of which can recognize an RNA base.[202] Chains of varying lengths consisting of different composition of these protein modules known as Pumby (Pumilio-based assembly) would allow recognition of any target sequence. Combined with residue-specific sense codon reassignment, both alternate recruiting mechanisms would allow GCE in POIs without the need of any engineering at the gene level. This would allow convenient usage of mRNA selective residue-specific sense codon reassignment for modifying endogenous POIs of a host organism.

4.9.2 *In vivo* synthesis of artificial designer biopolymers

Despite significant advancements in GCE achieved by years of research, the potential of the technology to synthesize artificial biopolymers *in vivo* is still untapped. Some of the pre-requisites for synthesizing completely artificial biopolymers by GCE include development of GCE-specific mutually orthogonal aaRS/tRNA pairs, expanding the range of substrates that can be accepted by the ribosome and increasing the availability of codons for reassignment. The work done in this thesis contributes towards mitigating the issue of limited availability of codons for GCE. Together with the advancements in the development of mutually orthogonal aaRS/tRNA pairs and increasing the repertoire of aaRS substrates beyond α-L-amino acids, sense codon reassignment holds the key for unleashing the complete potential of GCE. It is important to understand each of these concepts to comprehend the path towards achieving the future goal of *in vivo* generation of artificial biopolymers of desired functionality. Hence in the following sections state-of-the-art in mutually orthogonal GCE-specific aaRS/tRNA development (section 4.9.2.1) and in expanding the substrate specificity of aaRS beyond α-L-amino acids (section 4.9.2.2) will be covered along with the role of these advancements towards the future perspective of the work done in this thesis. In section 4.9.2.3 the significance of this dissertation towards achieving *in vivo* designer biopolymers will be summarized.

4.9.2.1 Development of orthogonal aaRS/tRNA pairs

Simultaneous incorporation of multiple distinct ncAAs in a POI requires, among other things, GCE specific aaRS/tRNA pairs that are orthogonal to each other as well as to the host endogenous aaRS/tRNA pairs. Currently only a handful of GCE-specific aaRS/tRNA pairs fulfil this criterion. This section will cover some of the studies where mutually orthogonal aaRS/tRNA pairs have been identified or developed for GCE. Most of these mutually orthogonal aaRS/tRNA pairs have been used for multiple stop codon suppression. These pairs could be further engineered to enable simultaneous reassignment of different sense codons to incorporate distinct ncAAs.

In *E. coli* derivatives of the tyrosine RS from *M. jannaschii* (Mj) and the cognate tRNA can be used with derivatives of PylRS/tRNA pairs from *M. mazei* (Mm) or *M. barkeri* (Mb) for simultaneous incorporation of distinct ncAAs in a POI. It was possible to simultaneously suppress TAG and TAA codons in an EGFP reporter in *E. coli* by MbPylRS/tRNA and MjTyrRS/tRNA respectively to incorporate the following pairs of ncAAs: pAcF and AzK, pAcF and ε-tBoc-lysine (eBK), pAzF and eBK, and OMeY and eBK.[203] Triple stop codon suppression in an engineered *E.coli* strain was demonstrated by the Chatterjee lab.[204] They developed an *E. coli* tryptophanyl RS (EcTrpRS)/tRNA pair to use in combination with MjTyrRS/tRNA and MmPylRS/tRNA to facilitate simultaneous incorporation of three distinct ncAAs in a POI. In the engineered *E.coli* strain the endogenous EcTrpRS/tRNA was replaced by its counterpart from yeast thereby allowing the use of a variant of the EcTrpRS/tRNA for GCE. Since all three stop codons were repurposed for ncAA incorporation, an alternate strategy had to be devised for correct termination of POI translation. A poly-histidine tag followed by a tobacco etch-virus (TEV) protease cleavage site and 3 TAA stop codons were added at the C-terminal of the POI. Cleavage by TEV protease allowed the generation of the correct full-length POI.[204]

In mammalian cells dual suppression of TAA and TAG stop codons was demonstrated by Xiao *et al*. Mutually orthogonal MmPylRS/MbPyl-tRNA and engineered EcTyrRS/tRNA pairs were used in HEK293T cells as TAA and TAG suppressors respectively.[205] In another study by the Chatterjee group MmPylRS/tRNA and EcLeuRS/tRNA were used as TGA and TAG suppressors in mammalian cells.[206] MmPylRS does not recognize the anticodon of the cognate tRNA, hence the MmPylRS/tRNA pair can be used to suppress different stop codons by simply altering the anticodon sequence of the cognate tRNA. Simultaneous TAA and TAG suppression was also achieved by Meineke *et al*.[207] In this case MmPylRS/tRNA was used as the TAA suppressor and an engineered derivative of EcTyrRS/tRNA served as the TAG suppressor.

A vast range of mutually orthogonal aaRS/tRNA pairs with distinct amino acid specificity would be the key to reassigning multiple sense codons to incorporate multiple different ncAAs. Hence it is necessary to go beyond the limited number of currently available orthogonal GCE-specific aaRS/tRNA pairs and direct efforts towards identification and subsequent evolution/engineering of new aaRS/tRNA pairs. The Chin lab developed several orthogonal aaRS/tRNA pairs for GCE including triply orthogonal aaRS/tRNA pairs with distinct ncAA and codon specificity. Continued effort dedicated to evolving orthogonal aaRS/tRNA pairs, engineering them for sense codon reassignment combined with mRNA specific sense codon reassignment would accomplish the final goal of *in vivo* designer biopolymer synthesis. Hence a brief overview of the strategy adopted, and the advancements made so far by the Chin lab for developing new aaRS/tRNA will be covered in the following paragraph.

The widely used PylRS from *M. mazei* and *M. barkeri* are known to consist of two domains.[208–210] The C-terminal domain is the catalytical domain necessary for accepting the amino acid and charging the cognate tRNA. The N-terminal domain on the other hand enhances binding affinity with the tRNA by interacting with its T-arm and variable loop.[211] Both N- and C-terminal domains are necessary to achieve detectable *in vivo* GCE.[208,212] In the case of *Desulfitobacterium hafniense (Dh)*, unlike *M.mazei* and *M.barkeri*, the N- and C- terminal domains of the PylRS are expressed as two distinct polypeptides which are subsequently assembled.[213] While searching for PylRS genes with similarity to the catalytic domain of MmPylRS and the N-terminal domain of DhPylRS, Willis *et al.* identified new PylRS/tRNA candidates from different organisms to test their amber suppression efficiency in *E.coli*.[213] Among the newly identified candidates was also a class of PylRS (ΔNPylRS) which had sequence similarity to the C-terminal domain of MmPylRS but lacked the N-terminal domain. ΔNPylRS/tRNA pairs from *Methanomethylophilus alvus* (Ma), *Methanogenic archaeon ISO4-G1* (G1), *Methanogenic archaeon ISO4-H5* (*H5*), *Methanonatronarchaeum termitum* (*Mt*), and *Methanomassiliicoccus luminyensis* (*Ml*) were further evaluated for their performance in amber suppression in *E. coli*. This study resulted in the identification of two ΔNPylRS/tRNA pairs, namely G1PylRS/tRNA and MaPylRS/tRNA that are orthogonal to *E.coli* endogenous aaRS/tRNA pairs as well as to MmPylRS/tRNA.[213] Additionally, the MaPylRS/tRNA was engineered to develop variants with enhanced amber suppression

capabilities. The discovery of the ΔNPylRS served as the starting point for development of more mutually orthogonal aaRS/tRNA pairs for GCE.

Based on sequence similarities ΔNPylRS/tRNA pairs can be divided into classes A and B. Class A ΔNPylRSs preferentially interact with class A tRNAs and the same is observed for class B ΔNPylRS/tRNA pairs.[214] Dunkelmann *et al.* investigated the orthogonality and amber suppression efficiency of 11 selected ΔNPylRS/tRNA pairs in *E. coli*.[214] All 11 tRNAs were found to be orthogonal to the *E.coli* translational machinery. Out of the 11 ΔNPylRS/tRNA pairs, 8 demonstrated detectable amber suppression in *E.coli*. Further studies were performed with 88 combinations of ΔNPylRS/tRNA pairs arising from the 11 tRNAs and 8 functional ΔNPylRSs. Out of the 88 combinations it was possible to identify 18 mutually orthogonal pairs. Following the promising results, Dunkelmann *et al.* proceeded to develop triple mutually orthogonal PylRS/tRNA pairs. Since MmPyl-tRNA could recognize several class A and B ΔNPylRSs, there was a need to find Pyl-tRNAs that would be compatible with MmPylRS and orthogonal to class A and B ΔNPylRSs. This search led to the identification of Spe Pyl-tRNA which performed almost similar to the cognate MmPyl-tRNA. After extensive engineering efforts 12 triply orthogonal PylRS/tRNA pairs were developed. Each pair comprised of MmPylRS/Spe Pyl-tRNA, specific class B ΔNPylRS/ evolved AlvΔNPyl-tRNA and specific class B ΔNPylRS/ evolved IntΔNPyl-tRNA. These pairs were further engineered to alter their ncAA and codon specificity to facilitate incorporation of distinct ncAAs in a POI. Finally, a set of three mutually orthogonal PylRS/tRNA pair was developed that could incorporate ncAAs in POI in response to the amber stop codon and two quadruplet codons, AGGA and AGTA. In 2023 the same group developed quadruply and quintuply mutually orthogonal PylRS/tRNA pairs. The ncAA specificity of these pairs are yet to be engineered to facilitate incorporation of multiple distinct ncAAs in a POI.[215]

Although significant progress has been made, a lot more is still left to be done to develop more GCE-specific aaRS/tRNA pairs mutually orthogonal at all levels including ncAA specificity to allow simultaneous incorporation of multiple distinct ncAAs .

4.9.2.2 Going beyond α-L-amino acids as building blocks for protein synthesis

In vivo synthesis of designer biopolymers requires expansion of the repertoire of monomers that can be incorporated by the translation machinery. At present more than 500 ncAAs with a variety of chemical handles have been incorporated into POIs *in vivo*. However true diversity in artificial biopolymer synthesis by GCE can only be achieved by going beyond canonical α-L-amino acids. Recently Dunkelmann *et al.* devised a strategy for discovering aaRS capable of charging their cognate tRNAs with monomers that are poor ribosomal substrates. They could successfully incorporate, α-α-disubstituted and β-linked monomers into a POI in *E.coli*.[59]. Manipulation of the translation machinery to enable incorporation of monomers other than α-L-amino acids represents a step forward towards unleashing the full potential of GCE. However, incorporation of multiple such monomers requires availability of codons for reassignment which can be fulfilled by the technology of sense codon reassignment developed in this thesis.

4.9.2.3 Availability of blank codons for reassignment

Ensuring availability of blank codons for ncAA incorporation is another challenge that needs to be overcome for *in vivo* synthesis of biopolymers by GCE. Years of research dedicated to solving this problem has resulted in a number of strategies including use of codons made from artificial base pairs, quadruplet codons as well as reassignment of sense codons. State-of-the-art in all these approaches have been discussed in section 1.4.2. So far *in vivo* site-specific sense codon reassignment in a POI has been shown for only a few sense codons. This thesis shows the reassignment of multiple sense codons, both rare and abundant in a mRNA-specific manner in mammalian cells. It is a step forward towards efficient reassignment of all sense codons for incorporation of desired ncAAs. Combined with advances in identification and evolution of orthogonal GCE-specific aaRS/tRNA pairs and expansion of the repertoire of monomers that can be processed by the ribosome, sense codon reassignment has the potential to realize the goal of *in vivo* synthesis of artificial designer biopolymers.

4.9.3 Shape microscopy

A prime motivation for developing mRNA selective residue-specific sense codon reassignment was to develop a novel fluorescent labelling technique to allow visualization of protein shapes *in vivo*. The realization of this goal requires mRNA selective residue-specific sense codon reassignment and subsequent incorporation of multiple fluorescent labels in a POI. Combined with SRM this labelling technique will allow the *in vivo* visualization of protein shapes. The success of this approach will provide the scientific community with the means to investigate still invisible subcellular structures. One such example will be discussed in this section.

The hallmark of eukaryotic cellular organization is the segregation of different biochemical processes into distinct organelles and the nucleus is such an organelle which serves as the seat of crucial cellular processes like DNA replication and transcription. [216,217] Molecular traffic between the nucleus and the cytoplasm is tightly regulated by the nuclear pore complex (NPC).[218] It is a large macromolecular complex of 120MDa (human) composed of multiple copies of 30 different proteins known as nucleoporins (NUPs). Dysfunctional NPC and impaired nucleocytoplasmic transport have been implicated in a number of diseases. Some examples include steroid-resistant nephrotic syndrome (SRNS), a kidney disease caused by mutations in NUP93, NUP205;[219] fetal akinesia deformation sequence (FADS) caused by mutations in NUP88;[220] a cardiac disease atrial fibrillation caused by mutation in NUP155 [221]etc. The significance of the NPC for maintaining cell physiology and the need to unearth molecular mechanisms of diseases caused by NPC dysfunction demand a thorough understanding of the structure of the NPC. The NPC consists of the following substructures: the nuclear and cytoplasmic rings, the inner pore ring, the cytoplasmic filaments, the nuclear basket and the central core.[218,222] Decades of laborious research using techniques like electron microscopy, SRM, X-ray crystallography, cryo-electron tomography has facilitated construction of a near-atomic model of the NPC.[223–226] However, the picture is incomplete. Detailed structure of the nuclear basket remains elusive. The NUPs comprising the nuclear basket are primarily disordered

and cryo-electron tomography, the most advanced technique instrumental in revealing substructures of the NPC is not suitable for visualizing disordered proteins due to their dynamic nature. This creates a need for the development of alternative strategies to shed light onto such subcellular structures. Efficient mRNA selective residue-specific sense codon reassignment would allow incorporation of multiple ncAAs at different sites of a NUP in the nuclear basket, for example NUP153. These ncAAs can be click-labelled with compatible fluorescent dyes and subsequently imaged by advanced SRM techniques like MINFLUX. A significant reduction in probe size combined with high resolution of MINFLUX would be pivotal in revealing the elusive structure of the nuclear basket.

The work of this thesis shows that it is possible to use OTOs to significantly minimize unspecific GCE and achieve mRNA selective residue-specific sense codon reassignment. It also shows the utility of this method for imaging applications. However, the method still needs further development to enable incorporation of multiple ncAAs that can subsequently click-react with fluorescent dyes.

5 Conclusion

This thesis demonstrates the applicability of mRNA-specific sense codon reassignment as a technique for fluorescent labelling of POI. Furthermore, the success of residue-specific sense codon reassignment exclusively in a POI and its subsequent use for imaging applications, pave the way for development of a novel technique to incorporate multiple fluorescent dyes in a POI. In combination with SRM, it will allow visualization of the shape of proteins *in vivo* as well as shed light on the still invisible subcellular structures, for example, the nuclear basket of the NPC. Furthermore, with the development of OTOs with alternate mRNA recruitment mechanisms, as discussed in section 4.9.1, it will be possible to conveniently use this labelling technique for endogenous proteins without the need of any engineering of the POI at the gene level.

In general, sense codon reassignment provides a solution to one of the pressing challenges of GCE, that is, limited availability of blank codons for ncAA incorporation. Combined with development of mutually orthogonal GCE-specific aaRS/tRNA pairs and expansion of the repertoire of building blocks that can be processed by the translation machinery, harnessing the potential of sense codon reassignment will facilitate *in vivo* synthesis of designer biopolymers and achieve feats greater than that made possible by billions of years of evolution.

Bibliography

[1] C. Jann, S. Giofré, R. Bhattacharjee, E.A. Lemke, Cracking the Code: Reprogramming the Genetic Script in Prokaryotes and Eukaryotes to Harness the Power of Noncanonical Amino Acids, Chem. Rev. 124 (2024) 10281–10362. https://doi.org/10.1021/acs.chemrev.3c00878.

[2] T.G. Heckler, Y. Zama, T. Naka, S.M. Hecht, Dipeptide formation with misacylated tRNAPhes, J Biol Chem 258 (1983) 4492–4495.

[3] C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins, Science 244 (1989) 182–188. https://doi.org/10.1126/science.2649980.

[4] C.D. Reinkemeier, G.E. Girona, E.A. Lemke, Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes, Science 363 (2019) eaaw2644. https://doi.org/10.1126/science.aaw2644.

[5] J.W. Chin, Expanding and reprogramming the genetic code, Nature 550 (2017) 53–60. https://doi.org/10.1038/nature24031.

[6] I. Nikić, T. Plass, O. Schraidt, J. Szymański, J.A.G. Briggs, C. Schultz, E.A. Lemke, Minimal Tags for Rapid Dual-Color Live-Cell Labeling and Super-Resolution Microscopy, Angewandte Chemie International Edition 53 (2014) 2245–2249. https://doi.org/10.1002/anie.201309847.

[7] R. Schmidt, T. Weihs, C.A. Wurm, I. Jansen, J. Rehman, S.J. Sahl, S.W. Hell, MINFLUX nanometer-scale 3D imaging and microsecond-range tracking on a common fluorescence microscope, Nat Commun 12 (2021) 1478. https://doi.org/10.1038/s41467-021-21652-z.

[8] B.N.G. Giepmans, S.R. Adams, M.H. Ellisman, R.Y. Tsien, The Fluorescent Toolbox for Assessing Protein Location and Function, Science 312 (2006) 217–224. https://doi.org/10.1126/science.1124618.

[9] R.Y. Tsien, The green fluorescent protein, Annu Rev Biochem 67 (1998) 509–544. https://doi.org/10.1146/annurev.biochem.67.1.509.

[10] H. Morise, O. Shimomura, F.H. Johnson, J. Winant, Intermolecular energy transfer in the bioluminescent system of Aequorea, Biochemistry 13 (1974) 2656–2662. https://doi.org/10.1021/bi00709a028.

[11] N.C. Shaner, G.H. Patterson, M.W. Davidson, Advances in fluorescent protein technology, Journal of Cell Science 120 (2007) 4247–4260. https://doi.org/10.1242/jcs.005801.

[12] K.-L. Wong, J.-C.G. Bünzli, P.A. Tanner, Quantum yield and brightness, Journal of Luminescence 224 (2020) 117256. https://doi.org/10.1016/j.jlumin.2020.117256.

[13] M. Hirano, R. Ando, S. Shimozono, M. Sugiyama, N. Takeda, H. Kurokawa, R. Deguchi, K. Endo, K. Haga, R. Takai-Todaka, S. Inaura, Y. Matsumura, H. Hama, Y. Okada, T. Fujiwara, T. Morimoto, K. Katayama, A. Miyawaki, A highly photostable and bright green fluorescent protein, Nat Biotechnol 40 (2022) 1132–1142. https://doi.org/10.1038/s41587-022-01278-2.

[14] M. Fernández-Suárez, A.Y. Ting, Fluorescent probes for super-resolution imaging in living cells, Nat Rev Mol Cell Biol 9 (2008) 929–943. https://doi.org/10.1038/nrm2531.

[15] J. Lippincott-Schwartz, G.H. Patterson, Photoactivatable fluorescent proteins for diffraction-limited and super-resolution imaging, Trends in Cell Biology 19 (2009) 555–565. https://doi.org/10.1016/j.tcb.2009.09.003.

[16] J.N. Henderson, R. Gepshtein, J.R. Heenan, K. Kallio, D. Huppert, S.J. Remington, Structure and mechanism of the photoactivatable green fluorescent protein, J Am Chem Soc 131 (2009) 4176–4177. https://doi.org/10.1021/ja808851n.

[17] V.V. Verkhusha, A. Sorkin, Conversion of the monomeric red fluorescent protein into a photoactivatable probe, Chem Biol 12 (2005) 279–285. https://doi.org/10.1016/j.chembiol.2005.01.005.

[18] E.A. Souslova, D.M. Chudakov, Photoswitchable cyan fluorescent protein as a FRET donor, Microsc Res Tech 69 (2006) 207–209. https://doi.org/10.1002/jemt.20278.
[19] K. Lang, J.W. Chin, Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins, Chem. Rev. 114 (2014) 4764–4806. https://doi.org/10.1021/cr400355w.
[20] M. Andresen, A.C. Stiel, S. Trowitzsch, G. Weber, C. Eggeling, M.C. Wahl, S.W. Hell, S. Jakobs, Structural basis for reversible photoswitching in Dronpa, Proceedings of the National Academy of Sciences 104 (2007) 13005–13009. https://doi.org/10.1073/pnas.0700629104.
[21] T. Clackson, W. Yang, L.W. Rozamus, M. Hatada, J.F. Amara, C.T. Rollins, L.F. Stevenson, S.R. Magari, S.A. Wood, N.L. Courage, X. Lu, F. Cerasoli, M. Gilman, D.A. Holt, Redesigning an FKBP–ligand interface to generate chemical dimerizers with novel specificity, Proceedings of the National Academy of Sciences 95 (1998) 10437–10442. https://doi.org/10.1073/pnas.95.18.10437.
[22] K.M. Marks, P.D. Braun, G.P. Nolan, A general approach for chemical labeling and rapid, spatially controlled protein inactivation, Proceedings of the National Academy of Sciences 101 (2004) 9982–9987. https://doi.org/10.1073/pnas.0401609101.
[23] A.F.L. Schneider, C.P.R. Hackenberger, Fluorescent labelling in living cells, Current Opinion in Biotechnology 48 (2017) 61–68. https://doi.org/10.1016/j.copbio.2017.03.012.
[24] L.W. Miller, Y. Cai, M.P. Sheetz, V.W. Cornish, In vivo protein labeling with trimethoprim conjugates: a flexible chemical tag, Nat Methods 2 (2005) 255–257. https://doi.org/10.1038/nmeth749.
[25] B.A. Griffin, S.R. Adams, R.Y. Tsien, Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells, Science 281 (1998) 269–272. https://doi.org/10.1126/science.281.5374.269.
[26] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, A general method for the covalent labeling of fusion proteins with small molecules in vivo, Nat Biotechnol 21 (2003) 86–89. https://doi.org/10.1038/nbt765.
[27] A. Gautier, A. Juillerat, C. Heinis, I.R. Corrêa, M. Kindermann, F. Beaufils, K. Johnsson, An engineered protein tag for multiprotein labeling in living cells, Chem Biol 15 (2008) 128–136. https://doi.org/10.1016/j.chembiol.2008.01.007.
[28] G.V. Los, L.P. Encell, M.G. McDougall, D.D. Hartzell, N. Karassina, C. Zimprich, M.G. Wood, R. Learish, R.F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D.H. Klaubert, R.F. Bulleit, K.V. Wood, HaloTag: a novel protein labeling technology for cell imaging and protein analysis, ACS Chem Biol 3 (2008) 373–382. https://doi.org/10.1021/cb800025k.
[29] Y. Hori, H. Ueno, S. Mizukami, K. Kikuchi, Photoactive Yellow Protein-Based Protein Labeling System with Turn-On Fluorescence Intensity, J. Am. Chem. Soc. 131 (2009) 16610–16611. https://doi.org/10.1021/ja904800k.
[30] M. Sunbul, L. Nacheva, A. Jäschke, Proximity-Induced Covalent Labeling of Proteins with a Reactive Fluorophore-Binding Peptide Tag, Bioconjugate Chem. 26 (2015) 1466–1469. https://doi.org/10.1021/acs.bioconjchem.5b00304.
[31] Y. Chen, C.M. Clouthier, K. Tsao, M. Strmiskova, H. Lachance, J.W. Keillor, Coumarin-based fluorogenic probes for no-wash protein labeling, Angew Chem Int Ed Engl 53 (2014) 13785–13788. https://doi.org/10.1002/anie.201408015.
[32] E.G. Guignet, R. Hovius, H. Vogel, Reversible site-selective labeling of membrane proteins in live cells, Nat Biotechnol 22 (2004) 440–444. https://doi.org/10.1038/nbt954.
[33] A. Ojida, K. Honda, D. Shinmi, S. Kiyonaka, Y. Mori, I. Hamachi, Oligo-Asp Tag/Zn(II) Complex Probe as a New Pair for Labeling and Fluorescence Imaging of Proteins, J. Am. Chem. Soc. 128 (2006) 10452–10459. https://doi.org/10.1021/ja0618604.

[34] S.K. Mazmanian, G. Liu, H. Ton-That, O. Schneewind, Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall, Science 285 (1999) 760–763. https://doi.org/10.1126/science.285.5428.760.
[35] T. Tanaka, T. Yamamoto, S. Tsukiji, T. Nagamune, Site-Specific Protein Modification on Living Cells Catalyzed by Sortase, ChemBioChem 9 (2008) 802–807. https://doi.org/10.1002/cbic.200700614.
[36] K. Strijbis, E. Spooner, H.L. Ploegh, Protein Ligation in Living Cells Using Sortase, Traffic 13 (2012) 780–789. https://doi.org/10.1111/j.1600-0854.2012.01345.x.
[37] M. Fernández-Suárez, H. Baruah, L. Martínez-Hernández, K.T. Xie, J.M. Baskin, C.R. Bertozzi, A.Y. Ting, Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes, Nat Biotechnol 25 (2007) 1483–1487. https://doi.org/10.1038/nbt1355.
[38] Z. Zhou, P. Cironi, A.J. Lin, Y. Xu, S. Hrvatin, D.E. Golan, P.A. Silver, C.T. Walsh, J. Yin, Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases, ACS Chem Biol 2 (2007) 337–346. https://doi.org/10.1021/cb700054k.
[39] T.M.S. Tang, D. Cardella, A.J. Lander, X. Li, J.S. Escudero, Y.-H. Tsai, L.Y.P. Luk, Use of an asparaginyl endopeptidase for chemo-enzymatic peptide and protein labeling †Electronic supplementary information (ESI) available. See DOI: 10.1039/d0sc02023k, Chem Sci 11 (2020) 5881–5888. https://doi.org/10.1039/d0sc02023k.
[40] I. Chen, M. Howarth, W. Lin, A.Y. Ting, Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase, Nat Methods 2 (2005) 99–104. https://doi.org/10.1038/nmeth735.
[41] D.S. Liu, L.G. Nivón, F. Richter, P.J. Goldman, T.J. Deerinck, J.Z. Yao, D. Richardson, W.S. Phipps, A.Z. Ye, M.H. Ellisman, C.L. Drennan, D. Baker, A.Y. Ting, Computational design of a red fluorophore ligase for site-specific protein labeling in living cells, Proceedings of the National Academy of Sciences 111 (2014) E4551–E4559. https://doi.org/10.1073/pnas.1404736111.
[42] A. Nygaard, L.G. Zachariassen, K.S. Larsen, A.S. Kristensen, C.J. Loland, Fluorescent non-canonical amino acid provides insight into the human serotonin transporter, Nat Commun 15 (2024) 9267. https://doi.org/10.1038/s41467-024-53584-9.
[43] I. Nikić, E.A. Lemke, Genetic code expansion enabled site-specific dual-color protein labeling: superresolution microscopy and beyond, Curr Opin Chem Biol 28 (2015) 164–173. https://doi.org/10.1016/j.cbpa.2015.07.021.
[44] L. Wang, A. Brock, B. Herberich, P.G. Schultz, Expanding the Genetic Code of Escherichia coli, Science 292 (2001) 498–500. https://doi.org/10.1126/science.1060077.
[45] T. Yanagisawa, M. Kuratani, E. Seki, N. Hino, K. Sakamoto, S. Yokoyama, Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-tRNA Synthetase, Cell Chem Biol 26 (2019) 936-949.e13. https://doi.org/10.1016/j.chembiol.2019.03.008.
[46] N.G. Koch, N. Budisa, Evolution of Pyrrolysyl-tRNA Synthetase: From Methanogenesis to Genetic Code Expansion, Chem. Rev. 124 (2024) 9580–9608. https://doi.org/10.1021/acs.chemrev.4c00031.
[47] M.A. Shandell, Z. Tan, V.W. Cornish, Genetic Code Expansion: A Brief History and Perspective, Biochemistry 60 (2021) 3455–3469. https://doi.org/10.1021/acs.biochem.1c00286.
[48] F.H. Arnold, Directed Evolution: Bringing New Chemistry to Life, Angewandte Chemie International Edition 57 (2018) 4143–4148. https://doi.org/10.1002/anie.201708408.
[49] D.R. Liu, P.G. Schultz, Progress toward the evolution of an organism with an expanded genetic code, Proceedings of the National Academy of Sciences 96 (1999) 4780–4785. https://doi.org/10.1073/pnas.96.9.4780.

[50] M. Pastrnak, P.G. Schultz, Phage selection for site-specific incorporation of unnatural amino acids into proteins in vivo, Bioorganic & Medicinal Chemistry 9 (2001) 2373–2379. https://doi.org/10.1016/S0968-0896(01)00157-2.
[51] M. Amiram, A.D. Haimovich, C. Fan, Y.-S. Wang, H.-R. Aerni, I. Ntai, D.W. Moonan, N.J. Ma, A.J. Rovner, S.H. Hong, N.L. Kelleher, A.L. Goodman, M.C. Jewett, D. Söll, J. Rinehart, F.J. Isaacs, Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids, Nat Biotechnol 33 (2015) 1272–1279. https://doi.org/10.1038/nbt.3372.
[52] A. Deiters, T.A. Cropp, M. Mukherji, J.W. Chin, J.C. Anderson, P.G. Schultz, Adding Amino Acids with Novel Reactivity to the Genetic Code of Saccharomyces Cerevisiae, J. Am. Chem. Soc. 125 (2003) 11782–11783. https://doi.org/10.1021/ja0370037.
[53] J.T. Stieglitz, J.A. Van Deventer, High-Throughput Aminoacyl-tRNA Synthetase Engineering for Genetic Code Expansion in Yeast, ACS Synth. Biol. 11 (2022) 2284–2299. https://doi.org/10.1021/acssynbio.1c00626.
[54] F. Tian, M.-L. Tsao, P.G. Schultz, A Phage Display System with Unnatural Amino Acids, J. Am. Chem. Soc. 126 (2004) 15962–15963. https://doi.org/10.1021/ja045673m.
[55] K.M. Esvelt, J.C. Carlson, D.R. Liu, A system for the continuous directed evolution of biomolecules, Nature 472 (2011) 499–503. https://doi.org/10.1038/nature09929.
[56] D.I. Bryson, C. Fan, L.-T. Guo, C. Miller, D. Söll, D.R. Liu, Continuous directed evolution of aminoacyl-tRNA synthetases, Nat Chem Biol 13 (2017) 1253–1260. https://doi.org/10.1038/nchembio.2474.
[57] J.K. Takimoto, K.L. Adams, Z. Xiang, L. Wang, Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells, Mol. BioSyst. 5 (2009) 931–934. https://doi.org/10.1039/B904228H.
[58] I. Nikić, G. Estrada Girona, J.H. Kang, G. Paci, S. Mikhaleva, C. Koehler, N.V. Shymanska, C. Ventura Santos, D. Spitz, E.A. Lemke, Debugging Eukaryotic Genetic Code Expansion for Site-Specific Click-PAINT Super-Resolution Microscopy, Angew Chem Int Ed Engl 55 (2016) 16172–16176. https://doi.org/10.1002/anie.201608284.
[59] D.L. Dunkelmann, C. Piedrafita, A. Dickson, K.C. Liu, T.S. Elliott, M. Fiedler, D. Bellini, A. Zhou, D. Cervettini, J.W. Chin, Adding α,α-disubstituted and β-linked monomers to the genetic code of an organism, Nature 625 (2024) 603–610. https://doi.org/10.1038/s41586-023-06897-6.
[60] G. Galli, H. Hofstetter, M.L. Birnstiel, Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements, Nature 294 (1981) 626–631. https://doi.org/10.1038/294626a0.
[61] S. Sharp, D. DeFranco, T. Dingermann, P. Farrell, D. Söll, Internal control regions for transcription of eukaryotic tRNA genes., Proceedings of the National Academy of Sciences 78 (1981) 6657–6661. https://doi.org/10.1073/pnas.78.11.6657.
[62] H.J. Drabkin, H.-J. Park, U.L. Rajbhandary, Amber Suppression in Mammalian Cells Dependent upon Expression of an Escherichia coli Aminoacyl-tRNA Synthetase Gene, Molecular and Cellular Biology 16 (1996) 907–913. https://doi.org/10.1128/MCB.16.3.907.
[63] H.-C. Yin, X.-Y. Chen, W. Wang, Q.-W. Meng, Identification and comparison of the porcine H1, U6, and 7SK RNA polymerase III promoters for short hairpin RNA expression, Mamm Genome 31 (2020) 110–116. https://doi.org/10.1007/s00335-020-09838-0.
[64] Q. Wang, L. Wang, New Methods Enabling Efficient Incorporation of Unnatural Amino Acids in Yeast, J. Am. Chem. Soc. 130 (2008) 6066–6067. https://doi.org/10.1021/ja800894n.
[65] S.M. Hancock, R. Uprety, A. Deiters, J.W. Chin, Expanding the Genetic Code of Yeast for Incorporation of Diverse Unnatural Amino Acids via a Pyrrolysyl-tRNA Synthetase/tRNA Pair, J. Am. Chem. Soc. 132 (2010) 14819–14824. https://doi.org/10.1021/ja104609m.
[66] R. Serfling, C. Lorenz, M. Etzel, G. Schicht, T. Böttke, M. Mörl, I. Coin, Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in

mammalian cells, Nucleic Acids Research 46 (2018) 1–10. https://doi.org/10.1093/nar/gkx1156.
[67] D. Jewel, R.E. Kelemen, R.L. Huang, Z. Zhu, B. Sundaresh, X. Cao, K. Malley, Z. Huang, M. Pasha, J. Anthony, T. van Opijnen, A. Chatterjee, Virus-assisted directed evolution of enhanced suppressor tRNAs in mammalian cells, Nat Methods 20 (2023) 95–103. https://doi.org/10.1038/s41592-022-01706-w.
[68] S. Ramazi, J. Zahiri, Post-translational modifications in proteins: resources, tools and prediction methods, Database 2021 (2021) baab012. https://doi.org/10.1093/database/baab012.
[69] G.A. Khoury, R.C. Baliban, C.A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database, Sci Rep 1 (2011) 90. https://doi.org/10.1038/srep00090.
[70] S. Shafi, A. Singh, P. Gupta, P.A. Chawla, F. Fayaz, A. Sharma, F.H. Pottoo, Deciphering the Role of Aberrant Protein Post-Translational Modification in the Pathology of Neurodegeneration, CNS Neurol Disord Drug Targets 20 (2021) 54–67. https://doi.org/10.2174/1871527319666200903162200.
[71] H.-S. Park, M.J. Hohn, T. Umehara, L.-T. Guo, E.M. Osborne, J. Benner, C.J. Noren, J. Rinehart, D. Söll, Expanding the Genetic Code of Escherichia coli with Phosphoserine, Science 333 (2011) 1151–1154. https://doi.org/10.1126/science.1207203.
[72] C. Fan, K. Ip, D. Söll, Expanding the genetic code of Escherichia coli with phosphotyrosine, FEBS Letters 590 (2016) 3040–3047. https://doi.org/10.1002/1873-3468.12333.
[73] M.S. Zhang, S.F. Brunner, N. Huguenin-Dezot, A.D. Liang, W.H. Schmied, D.T. Rogerson, J.W. Chin, Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing, Nature Methods 14 (2017) 729. https://doi.org/10.1038/nmeth.4302.
[74] Z.A. Wang, P.A. Cole, The Chemical Biology of Reversible Lysine Post-translational Modifications, Cell Chemical Biology 27 (2020) 953–969. https://doi.org/10.1016/j.chembiol.2020.07.002.
[75] W. Dang, K.K. Steffen, R. Perry, J.A. Dorsey, F.B. Johnson, A. Shilatifard, M. Kaeberlein, B.K. Kennedy, S.L. Berger, Histone H4 lysine 16 acetylation regulates cellular lifespan, Nature 459 (2009) 802–807. https://doi.org/10.1038/nature08085.
[76] S. Zhao, W. Xu, W. Jiang, W. Yu, Y. Lin, T. Zhang, J. Yao, L. Zhou, Y. Zeng, H. Li, Y. Li, J. Shi, W. An, S.M. Hancock, F. He, L. Qin, J. Chin, P. Yang, X. Chen, Q. Lei, Y. Xiong, K.-L. Guan, Regulation of Cellular Metabolism by Protein Lysine Acetylation, Science 327 (2010) 1000–1004. https://doi.org/10.1126/science.1179689.
[77] M. Hu, F. He, E.W. Thompson, K. (Ken) Ostrikov, X. Dai, Lysine Acetylation, Cancer Hallmarks and Emerging Onco-Therapeutic Opportunities, Cancers (Basel) 14 (2022) 346. https://doi.org/10.3390/cancers14020346.
[78] S.J. Elsässer, R.J. Ernst, O.S. Walker, J.W. Chin, Genetic code expansion in stable cell lines enables encoded chromatin modification, Nature Methods 13 (2016) 158. https://doi.org/10.1038/nmeth.3701.
[79] Y.-S. Yang, C.-C. Wang, B.-H. Chen, Y.-H. Hou, K.-S. Hung, Y.-C. Mao, Tyrosine Sulfation as a Protein Post-Translational Modification, Molecules 20 (2015) 2138–2164. https://doi.org/10.3390/molecules20022138.
[80] V. Stewart, P.C. Ronald, Sulfotyrosine residues: Interaction specificity determinants for extracellular protein–protein interactions, Journal of Biological Chemistry 298 (2022) 102232. https://doi.org/10.1016/j.jbc.2022.102232.
[81] C. Huang, M. Venturi, S. Majeed, M.J. Moore, S. Phogat, M.-Y. Zhang, D.S. Dimitrov, W.A. Hendrickson, J. Robinson, J. Sodroski, R. Wyatt, H. Choe, M. Farzan, P.D. Kwong, Structural basis of tyrosine sulfation and VH-gene usage in antibodies that recognize the HIV type 1 coreceptor-binding site on gp120, Proceedings of the National Academy of Sciences 101 (2004) 2706–2711. https://doi.org/10.1073/pnas.0308527100.

[82] T. Dorfman, M.J. Moore, A.C. Guth, H. Choe, M. Farzan, A Tyrosine-sulfated Peptide Derived from the Heavy-chain CDR3 Region of an HIV-1-neutralizing Antibody Binds gp120 and Inhibits HIV-1 Infection*, Journal of Biological Chemistry 281 (2006) 28529–28535. https://doi.org/10.1074/jbc.M602732200.
[83] J.J. Chiang, M.R. Gardner, B.D. Quinlan, T. Dorfman, H. Choe, M. Farzan, Enhanced Recognition and Neutralization of HIV-1 by Antibody-Derived CCR5-Mimetic Peptide Variants, Journal of Virology 86 (2012) 12417. https://doi.org/10.1128/JVI.00967-12.
[84] X. Li, J. Hitomi, C.C. Liu, Characterization of a Sulfated Anti-HIV Antibody Using an Expanded Genetic Code, Biochemistry 57 (2018) 2903–2907. https://doi.org/10.1021/acs.biochem.8b00374.
[85] J.S. Beckman, W.H. Koppenol, Nitric oxide, superoxide, and peroxynitrite: the good, the bad, and ugly, American Journal of Physiology-Cell Physiology 271 (1996) C1424–C1437. https://doi.org/10.1152/ajpcell.1996.271.5.C1424.
[86] H. Ischiropoulos, Biological Tyrosine Nitration: A Pathophysiological Function of Nitric Oxide and Reactive Oxygen Species, Archives of Biochemistry and Biophysics 356 (1998) 1–11. https://doi.org/10.1006/abbi.1998.0755.
[87] H. Neumann, J.L. Hazen, J. Weinstein, R.A. Mehl, J.W. Chin, Genetically Encoding Protein Oxidative Damage, J. Am. Chem. Soc. 130 (2008) 4028–4033. https://doi.org/10.1021/ja710100d.
[88] D.P. Nguyen, M.M. Garcia Alai, P.B. Kapadnis, H. Neumann, J.W. Chin, Genetically Encoding Nϵ-Methyl-l-lysine in Recombinant Histones, J. Am. Chem. Soc. 131 (2009) 14194–14195. https://doi.org/10.1021/ja906603s.
[89] N. Wu, A. Deiters, T.A. Cropp, D. King, P.G. Schultz, A Genetically Encoded Photocaged Amino Acid, J. Am. Chem. Soc. 126 (2004) 14306–14307. https://doi.org/10.1021/ja040175z.
[90] E.A. Lemke, D. Summerer, B.H. Geierstanger, S.M. Brittain, P.G. Schultz, Control of protein phosphorylation with a genetically encoded photocaged amino acid, Nat Chem Biol 3 (2007) 769–772. https://doi.org/10.1038/nchembio.2007.44.
[91] A. Gautier, D.P. Nguyen, H. Lusic, W. An, A. Deiters, J.W. Chin, Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells, J. Am. Chem. Soc. 132 (2010) 4086–4088. https://doi.org/10.1021/ja910688s.
[92] J. Luo, E. Arbely, J. Zhang, C. Chou, R. Uprety, J.W. Chin, A. Deiters, Genetically encoded optical activation of DNA recombination in human cells †Electronic supplementary information (ESI) available: Experimental protocols. See DOI: 10.1039/c6cc03934k Click here for additional data file., Chem Commun (Camb) 52 (2016) 8529–8532. https://doi.org/10.1039/c6cc03934k.
[93] C. Chou, A. Deiters, Light-Activated Gene Editing with a Photocaged Zinc-Finger Nuclease, Angew Chem Int Ed Engl 50 (2011) 6839–6842. https://doi.org/10.1002/anie.201101157.
[94] J. Hemphill, C. Chou, J.W. Chin, A. Deiters, Genetically Encoded Light-Activated Transcription for Spatiotemporal Control of Gene Expression and Gene Silencing in Mammalian Cells, J. Am. Chem. Soc. 135 (2013) 13433–13439. https://doi.org/10.1021/ja4051026.
[95] C. Hoppmann, V.K. Lacey, G.V. Louie, J. Wei, J.P. Noel, L. Wang, Genetically Encoding Photoswitchable Click Amino Acids in Escherichia coli and Mammalian Cells, Angewandte Chemie International Edition 53 (2014) 3932–3936. https://doi.org/10.1002/anie.201400001.
[96] J.W. Chin, A.B. Martin, D.S. King, L. Wang, P.G. Schultz, Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli, Proceedings of the National Academy of Sciences 99 (2002) 11020–11024. https://doi.org/10.1073/pnas.172226299.
[97] J.W. Chin, S.W. Santoro, A.B. Martin, D.S. King, L. Wang, P.G. Schultz, Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia coli, J. Am. Chem. Soc. 124 (2002) 9026–9027. https://doi.org/10.1021/ja027007w.

[98] S. Lin, Z. Zhang, H. Xu, L. Li, S. Chen, J. Li, Z. Hao, P.R. Chen, Site-Specific Incorporation of Photo-Cross-Linker and Bioorthogonal Amino Acids into Enteric Bacterial Pathogens, J. Am. Chem. Soc. 133 (2011) 20581–20587. https://doi.org/10.1021/ja209008w.
[99] I. Coin, V. Katritch, T. Sun, Z. Xiang, F.Y. Siu, M. Beyermann, R.C. Stevens, L. Wang, Genetically Encoded Chemical Probes in Cells Reveal the Binding Path of Urocortin-I to CRF Class B GPCR, Cell 155 (2013) 1258–1269. https://doi.org/10.1016/j.cell.2013.11.008.
[100] J.C. Jackson, J.T. Hammill, R.A. Mehl, Site-Specific Incorporation of a 19F-Amino Acid into Proteins as an NMR Probe for Characterizing Protein Structure and Reactivity, J. Am. Chem. Soc. 129 (2007) 1160–1166. https://doi.org/10.1021/ja064661t.
[101] Q. Liu, Q. He, X. Lyu, F. Yang, Z. Zhu, P. Xiao, Z. Yang, F. Zhang, Z. Yang, X. Wang, P. Sun, Q. Wang, C. Qu, Z. Gong, J. Lin, Z. Xu, S. Song, S. Huang, S. Guo, M. Han, K. Zhu, X. Chen, A.W. Kahsai, K.-H. Xiao, W. Kong, F. Li, K. Ruan, Z. Li, X. Yu, X. Niu, C. Jin, J. Wang, J. Sun, DeSiphering receptor core-induced and ligand-dependent conformational changes in arrestin via genetic encoded trimethylsilyl 1H-NMR probe, Nat Commun 11 (2020) 4857. https://doi.org/10.1038/s41467-020-18433-5.
[102] K.C. Schultz, L. Supekova, Y. Ryu, J. Xie, R. Perera, P.G. Schultz, A genetically encoded infrared probe, J Am Chem Soc 128 (2006) 13984–13985. https://doi.org/10.1021/ja0636690.
[103] M.J. Schmidt, A. Fedoseev, D. Bücker, J. Borbas, C. Peter, M. Drescher, D. Summerer, EPR Distance Measurements in Native Proteins with Genetically Encoded Spin Labels, ACS Chem. Biol. 10 (2015) 2764–2771. https://doi.org/10.1021/acschembio.5b00512.
[104] J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon, P.G. Schultz, The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination, Nat Biotechnol 22 (2004) 1297–1301. https://doi.org/10.1038/nbt1013.
[105] D. Summerer, S. Chen, N. Wu, A. Deiters, J.W. Chin, P.G. Schultz, A genetically encoded fluorescent amino acid, Proceedings of the National Academy of Sciences 103 (2006) 9785–9789. https://doi.org/10.1073/pnas.0603965103.
[106] A. Chatterjee, J. Guo, H.S. Lee, P.G. Schultz, A Genetically Encoded Fluorescent Probe in Mammalian Cells, J. Am. Chem. Soc. 135 (2013) 12540–12543. https://doi.org/10.1021/ja4059553.
[107] J. Wang, J. Xie, P.G. Schultz, A Genetically Encoded Fluorescent Amino Acid, J. Am. Chem. Soc. 128 (2006) 8738–8739. https://doi.org/10.1021/ja062666k.
[108] H.C. Kolb, M.G. Finn, K.B. Sharpless, Click Chemistry: Diverse Chemical Function from a Few Good Reactions, Angewandte Chemie International Edition 40 (2001) 2004–2021. https://doi.org/10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.
[109] A.J. Link, D.A. Tirrell, Cell surface labeling of Escherichia coli via copper(I)-catalyzed [3+2] cycloaddition, J Am Chem Soc 125 (2003) 11164–11165. https://doi.org/10.1021/ja036765z.
[110] J.C. Jewett, E.M. Sletten, C.R. Bertozzi, Rapid Cu-Free Click Chemistry with Readily Synthesized Biarylazacyclooctynones, J. Am. Chem. Soc. 132 (2010) 3688–3690. https://doi.org/10.1021/ja100014q.
[111] B.L. Oliveira, Z. Guo, G.J.L. Bernardes, Inverse electron demand Diels–Alder reactions in chemical biology, Chem. Soc. Rev. 46 (2017) 4895–4950. https://doi.org/10.1039/C7CS00184C.
[112] A.J. Link, M.K.S. Vink, D.A. Tirrell, Presentation and Detection of Azide Functionality in Bacterial Cell Surface Proteins, J. Am. Chem. Soc. 126 (2004) 10598–10602. https://doi.org/10.1021/ja047629c.
[113] G. Liang, H. Ren, J. Rao, A biocompatible condensation reaction for controlled assembly of nanostructures in living cells, Nature Chem 2 (2010) 54–60. https://doi.org/10.1038/nchem.480.

[114] N. Li, R.K.V. Lim, S. Edwardraja, Q. Lin, Copper-Free Sonogashira Cross-Coupling for Functionalization of Alkyne-Encoded Proteins in Aqueous Medium and in Bacterial Cells, J. Am. Chem. Soc. 133 (2011) 15316–15319. https://doi.org/10.1021/ja2066913.
[115] W. Song, Y. Wang, Z. Yu, C.I.R. Vera, J. Qu, Q. Lin, A Metabolic Alkene Reporter for Spatiotemporally Controlled Imaging of Newly Synthesized Proteins in Mammalian Cells, ACS Chem. Biol. 5 (2010) 875–885. https://doi.org/10.1021/cb100193h.
[116] E. Kozma, O. Demeter, P. Kele, Bio-orthogonal Fluorescent Labelling of Biopolymers through Inverse-Electron-Demand Diels-Alder Reactions, Chembiochem 18 (2017) 486–501. https://doi.org/10.1002/cbic.201600607.
[117] M.L. Blackman, M. Royzen, J.M. Fox, Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels−Alder Reactivity, J. Am. Chem. Soc. 130 (2008) 13518–13519. https://doi.org/10.1021/ja8053805.
[118] Y. Huang, T. Liu, Therapeutic applications of genetic code expansion, Synth Syst Biotechnol 3 (2018) 150–158. https://doi.org/10.1016/j.synbio.2018.09.003.
[119] C.H. Kim, J.Y. Axup, P.G. Schultz, Protein conjugation with genetically encoded unnatural amino acids, Current Opinion in Chemical Biology 17 (2013) 412–419. https://doi.org/10.1016/j.cbpa.2013.04.017.
[120] J.Y. Axup, K.M. Bajjuri, M. Ritland, B.M. Hutchins, C.H. Kim, S.A. Kazane, R. Halder, J.S. Forsyth, A.F. Santidrian, K. Stafin, Y. Lu, H. Tran, A.J. Seller, S.L. Biroc, A. Szydlik, J.K. Pinkstaff, F. Tian, S.C. Sinha, B. Felding-Habermann, V.V. Smider, P.G. Schultz, Synthesis of site-specific antibody-drug conjugates using unnatural amino acids, Proceedings of the National Academy of Sciences 109 (2012) 16101–16106. https://doi.org/10.1073/pnas.1211023109.
[121] M.P. VanBrunt, K. Shanebeck, Z. Caldwell, J. Johnson, P. Thompson, T. Martin, H. Dong, G. Li, H. Xu, F. D'Hooge, L. Masterson, P. Bariola, A. Tiberghien, E. Ezeadi, D.G. Williams, J.A. Hartley, P.W. Howard, K.H. Grabstein, M.A. Bowen, M. Marelli, Genetically Encoded Azide Containing Amino Acid in Mammalian Cells Enables Site-Specific Antibody-Drug Conjugates Using Click Cycloaddition Chemistry, Bioconjug Chem 26 (2015) 2249–2260. https://doi.org/10.1021/acs.bioconjchem.5b00359.
[122] H. Lu, D. Wang, S. Kazane, T. Javahishvili, F. Tian, F. Song, A. Sellers, B. Barnett, P.G. Schultz, Site-specific Antibody-polymer Conjugates for siRNA Delivery, J Am Chem Soc 135 (2013) 13885–13891. https://doi.org/10.1021/ja4059525.
[123] Y. Cao, J.Y. Axup, J.S.Y. Ma, R.E. Wang, S. Choi, V. Tardif, R.K.V. Lim, H.M. Pugh, B.R. Lawson, G. Welzel, S.A. Kazane, Y. Sun, F. Tian, S. Srinagesh, T. Javahishvili, P.G. Schultz, C.H. Kim, Multiformat T-Cell-Engaging Bispecific Antibodies Targeting Human Breast Cancers, Angewandte Chemie International Edition 54 (2015) 7022–7027. https://doi.org/10.1002/anie.201500799.
[124] M. Swierczewska, K.C. Lee, S. Lee, What is the future of PEGylated therapies?, Expert Opinion on Emerging Drugs 20 (2015) 531–536. https://doi.org/10.1517/14728214.2015.1113254.
[125] C.-Y. Yang, P.D. Renfrew, A.J. Olsen, M. Zhang, C. Yuvienco, R. Bonneau, J.K. Montclare, Improved Stability and Half-Life of Fluorinated Phosphotriesterase Using Rosetta, ChemBioChem 15 (2014) 1761–1764. https://doi.org/10.1002/cbic.201402062.
[126] W. Xuan, J. Li, X. Luo, P.G. Schultz, Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins, Angewandte Chemie International Edition 55 (2016) 10065–10068. https://doi.org/10.1002/anie.201604891.
[127] N. Wang, Y. Li, W. Niu, M. Sun, R. Cerny, Q. Li, J. Guo, Construction of a Live-Attenuated HIV-1 Vaccine through Genetic Code Expansion, Angewandte Chemie International Edition 53 (2014) 4867–4871. https://doi.org/10.1002/anie.201402092.
[128] Z. Yuan, N. Wang, G. Kang, W. Niu, Q. Li, J. Guo, Controlling Multicycle Replication of Live-Attenuated HIV-1 Using an Unnatural Genetic Switch, ACS Synth. Biol. 6 (2017) 721–731. https://doi.org/10.1021/acssynbio.6b00373.

[129] L. Si, H. Xu, X. Zhou, Z. Zhang, Z. Tian, Y. Wang, Y. Wu, B. Zhang, Z. Niu, C. Zhang, G. Fu, S. Xiao, Q. Xia, L. Zhang, D. Zhou, Generation of influenza A viruses as live but replication-incompetent virus vaccines, Science 354 (2016) 1170–1173. https://doi.org/10.1126/science.aah5869.
[130] K. Tamada, D. Geng, Y. Sakoda, N. Bansal, R. Srivastava, Z. Li, E. Davila, Redirecting gene-modified T cells toward various cancer types using tagged antibodies, Clin Cancer Res 18 (2012) 6436–6445. https://doi.org/10.1158/1078-0432.CCR-12-1449.
[131] C. Chen, G. Yu, Y. Huang, W. Cheng, Y. Li, Y. Sun, H. Ye, T. Liu, Genetic-code-expanded cell-based therapy for treating diabetes in mice, Nat Chem Biol 18 (2022) 47–55. https://doi.org/10.1038/s41589-021-00899-z.
[132] R. Bhattacharjee, E.A. Lemke, Potential vs Challenges of Expanding the Protein Universe With Genetic Code Expansion in Eukaryotic Cells, Journal of Molecular Biology 436 (2024) 168807. https://doi.org/10.1016/j.jmb.2024.168807.
[133] J. Shine, L. Dalgarno, The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites, Proceedings of the National Academy of Sciences 71 (1974) 1342–1346. https://doi.org/10.1073/pnas.71.4.1342.
[134] J.-D. Wen, S.-T. Kuo, H.-H.D. Chou, The diversity of Shine-Dalgarno sequences sheds light on the evolution of translation initiation, RNA Biol 18 (2021) 1489–1500. https://doi.org/10.1080/15476286.2020.1861406.
[135] N. Malys, Shine-Dalgarno sequence of bacteriophage T4: GAGG prevails in early genes, Mol Biol Rep 39 (2012) 33–39. https://doi.org/10.1007/s11033-011-0707-4.
[136] A. Hui, H.A. de Boer, Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in Escherichia coli., Proceedings of the National Academy of Sciences 84 (1987) 4762–4766. https://doi.org/10.1073/pnas.84.14.4762.
[137] O. Rackham, J.W. Chin, A network of orthogonal ribosome·mRNA pairs, Nat Chem Biol 1 (2005) 159–166. https://doi.org/10.1038/nchembio719.
[138] K. Wang, H. Neumann, S.Y. Peak-Chew, J.W. Chin, Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion, Nat Biotechnol 25 (2007) 770–777. https://doi.org/10.1038/nbt1314.
[139] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J.W. Chin, Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome, Nature 464 (2010) 441–444. https://doi.org/10.1038/nature08817.
[140] C. Orelle, E.D. Carlson, T. Szal, T. Florin, M.C. Jewett, A.S. Mankin, Protein synthesis by ribosomes with tethered subunits, Nature 524 (2015) 119–124. https://doi.org/10.1038/nature14862.
[141] E.D. Carlson, A.E. d'Aquino, D.S. Kim, E.M. Fulk, K. Hoang, T. Szal, A.S. Mankin, M.C. Jewett, Engineered ribosomes with tethered subunits for expanding biological function, Nat Commun 10 (2019) 3920. https://doi.org/10.1038/s41467-019-11427-y.
[142] W.H. Schmied, Z. Tnimov, C. Uttamapinant, C.D. Rae, S.D. Fried, J.W. Chin, Controlling orthogonal ribosome subunit interactions enables evolution of new function, Nature 564 (2018) 444–448. https://doi.org/10.1038/s41586-018-0773-z.
[143] D.S. Kim, A. Watkins, E. Bidstrup, J. Lee, V. Topkar, C. Kofman, K.J. Schwarz, Y. Liu, G. Pintilie, E. Roney, R. Das, M.C. Jewett, Three-dimensional structure-guided evolution of a ribosome with tethered subunits, Nat Chem Biol 18 (2022) 990–998. https://doi.org/10.1038/s41589-022-01064-w.
[144] H.H. Wang, F.J. Isaacs, P.A. Carr, Z.Z. Sun, G. Xu, C.R. Forest, G.M. Church, Programming cells by multiplex genome engineering and accelerated evolution, Nature 460 (2009) 894–898. https://doi.org/10.1038/nature08187.
[145] M.J. Lajoie, A.J. Rovner, D.B. Goodman, H.-R. Aerni, A.D. Haimovich, G. Kuznetsov, J.A. Mercer, H.H. Wang, P.A. Carr, J.A. Mosberg, N. Rohland, P.G. Schultz, J.M. Jacobson, J.

Rinehart, G.M. Church, F.J. Isaacs, Genomically Recoded Organisms Expand Biological Functions, Science 342 (2013) 357–360. https://doi.org/10.1126/science.1241459.
[146] F.J. Isaacs, P.A. Carr, H.H. Wang, M.J. Lajoie, B. Sterling, L. Kraal, A.C. Tolonen, T.A. Gianoulis, D.B. Goodman, N.B. Reppas, C.J. Emig, D. Bang, S.J. Hwang, M.C. Jewett, J.M. Jacobson, G.M. Church, Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement, Science 333 (2011) 348–353. https://doi.org/10.1126/science.1205822.
[147] T. Mukai, H. Hoshi, K. Ohtake, M. Takahashi, A. Yamaguchi, A. Hayashi, S. Yokoyama, K. Sakamoto, Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon, Sci Rep 5 (2015) 9699. https://doi.org/10.1038/srep09699.
[148] K. Wang, J. Fredens, S.F. Brunner, S.H. Kim, T. Chia, J.W. Chin, Defining synonymous codon compression schemes by genome recoding, Nature 539 (2016) 59–64. https://doi.org/10.1038/nature20124.
[149] J. Fredens, K. Wang, D. de la Torre, L.F.H. Funke, W.E. Robertson, Y. Christova, T. Chia, W.H. Schmied, D.L. Dunkelmann, V. Beránek, C. Uttamapinant, A.G. Llamazares, T.S. Elliott, J.W. Chin, Total synthesis of Escherichia coli with a recoded genome, Nature 569 (2019) 514–518. https://doi.org/10.1038/s41586-019-1192-5.
[150] W.E. Robertson, L.F.H. Funke, D. de la Torre, J. Fredens, T.S. Elliott, M. Spinck, Y. Christova, D. Cervettini, F.L. Böge, K.C. Liu, S. Buse, S. Maslen, G.P.C. Salmond, J.W. Chin, Sense codon reassignment enables viral resistance and encoded polymer synthesis, Science 372 (2021) 1057–1062. https://doi.org/10.1126/science.abg3029.
[151] C.D. Reinkemeier, E.A. Lemke, Dual film-like organelles enable spatial separation of orthogonal eukaryotic translation, Cell 184 (2021) 4886-4903.e21. https://doi.org/10.1016/j.cell.2021.08.001.
[152] C.D. Reinkemeier, E.A. Lemke, Condensed, Microtubule-coating Thin Organelles for Orthogonal Translation in Mammalian Cells, J Mol Biol 434 (2022) 167454. https://doi.org/10.1016/j.jmb.2022.167454.
[153] M. Yu, M. Heidari, S. Mikhaleva, P.S. Tan, S. Mingu, H. Ruan, C.D. Reinkemeier, A. Obarska-Kosinska, M. Siggel, M. Beck, G. Hummer, E.A. Lemke, Visualizing the disordered nuclear transport machinery in situ, Nature 617 (2023) 162–169. https://doi.org/10.1038/s41586-023-05990-0.
[154] C. Switzer, S.E. Moroney, S.A. Benner, Enzymatic incorporation of a new base pair into DNA and RNA, J. Am. Chem. Soc. 111 (1989) 8322–8323. https://doi.org/10.1021/ja00203a067.
[155] J.D. Bain, C. Switzer, R. Chamberlin, S.A. Benner, Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code, Nature 356 (1992) 537–539. https://doi.org/10.1038/356537a0.
[156] J.A. Piccirilli, S.A. Benner, T. Krauch, S.E. Moroney, S.A. Benner, Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet, Nature 343 (1990) 33–37. https://doi.org/10.1038/343033a0.
[157] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa, S. Yokoyama, An unnatural base pair for incorporating amino acid analogs into proteins, Nat Biotechnol 20 (2002) 177–182. https://doi.org/10.1038/nbt0202-177.
[158] Z. Yang, F. Chen, J.B. Alvarado, S.A. Benner, Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System, J. Am. Chem. Soc. 133 (2011) 15105–15112. https://doi.org/10.1021/ja204910n.
[159] Z. Yang, D. Hutter, P. Sheng, A.M. Sismour, S.A. Benner, Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern, Nucleic Acids Research 34 (2006) 6095–6101. https://doi.org/10.1093/nar/gkl633.
[160] S. Hoshika, N.A. Leal, M.-J. Kim, M.-S. Kim, N.B. Karalkar, H.-J. Kim, A.M. Bates, N.E. Watkins, H.A. SantaLucia, A.J. Meyer, S. DasGupta, J.A. Piccirilli, A.D. Ellington, J. SantaLucia, M.M. Georgiadis, S.A. Benner, Hachimoji DNA and RNA: A genetic system with

eight building blocks, Science 363 (2019) 884–887. https://doi.org/10.1126/science.aat0971.
[161] Y.J. Seo, G.T. Hwang, P. Ordoukhanian, F.E. Romesberg, Optimization of an Unnatural Base Pair toward Natural-Like Replication, J. Am. Chem. Soc. 131 (2009) 3246–3252. https://doi.org/10.1021/ja807853m.
[162] D.A. Malyshev, K. Dhami, H.T. Quach, T. Lavergne, P. Ordoukhanian, A. Torkamani, F.E. Romesberg, Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet, Proceedings of the National Academy of Sciences 109 (2012) 12005–12010. https://doi.org/10.1073/pnas.1205176109.
[163] D.A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J.M. Foster, I.R. Corrêa, F.E. Romesberg, A semi-synthetic organism with an expanded genetic alphabet, Nature 509 (2014) 385–388. https://doi.org/10.1038/nature13314.
[164] F.E. Romesberg, Creation, Optimization, and Use of Semi-Synthetic Organisms that Store and Retrieve Increased Genetic Information, Journal of Molecular Biology 434 (2022) 167331. https://doi.org/10.1016/j.jmb.2021.167331.
[165] Y. Zhang, J.L. Ptacin, E.C. Fischer, H.R. Aerni, C.E. Caffaro, K. San Jose, A.W. Feldman, C.R. Turner, F.E. Romesberg, A semi-synthetic organism that stores and retrieves increased genetic information, Nature 551 (2017) 644–647. https://doi.org/10.1038/nature24659.
[166] Y. Zhang, B.M. Lamb, A.W. Feldman, A.X. Zhou, T. Lavergne, L. Li, F.E. Romesberg, A semisynthetic organism engineered for the stable expansion of the genetic alphabet, Proceedings of the National Academy of Sciences 114 (2017) 1317–1322. https://doi.org/10.1073/pnas.1616443114.
[167] D.L. Dunkelmann, S.B. Oehm, A.T. Beattie, J.W. Chin, A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design, Nat. Chem. 13 (2021) 1110–1117. https://doi.org/10.1038/s41557-021-00764-5.
[168] W. Niu, P.G. Schultz, J. Guo, An Expanded Genetic Code in Mammalian Cells with a Functional Quadruplet Codon, ACS Chem. Biol. 8 (2013) 1640–1645. https://doi.org/10.1021/cb4001662.
[169] Y. Chen, Y. Wan, N. Wang, Z. Yuan, W. Niu, Q. Li, J. Guo, Controlling the Replication of a Genomically Recoded HIV-1 with a Functional Quadruplet Codon in Mammalian Cells, ACS Synth. Biol. 7 (2018) 1612–1617. https://doi.org/10.1021/acssynbio.8b00096.
[170] Z. Xi, L. Davis, K. Baxter, A. Tynan, A. Goutou, S. Greiss, Using a quadruplet codon to expand the genetic code of an animal, Nucleic Acids Research 50 (2022) 4801–4812. https://doi.org/10.1093/nar/gkab1168.
[171] E.M. Mills, V.L. Barlow, A.T. Jones, Y.-H. Tsai, Development of mammalian cell logic gates controlled by unnatural amino acids, Cell Reports Methods 1 (2021) 100073. https://doi.org/10.1016/j.crmeth.2021.100073.
[172] D.C. Dieterich, A.J. Link, J. Graumann, D.A. Tirrell, E.M. Schuman, Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT), Proceedings of the National Academy of Sciences 103 (2006) 9482–9487. https://doi.org/10.1073/pnas.0601637103.
[173] B. Alvarez-Castelao, C.T. Schanzenbächer, C. Hanus, C. Glock, S. Tom Dieck, A.R. Dörrbaum, I. Bartnik, B. Nassim-Assir, E. Ciirdaeva, A. Mueller, D.C. Dieterich, D.A. Tirrell, J.D. Langer, E.M. Schuman, Cell-type-specific metabolic labeling of nascent proteomes in vivo, Nat Biotechnol 35 (2017) 1196–1201. https://doi.org/10.1038/nbt.4016.
[174] S. tom Dieck, L. Kochen, C. Hanus, M. Heumüller, I. Bartnik, B. Nassim-Assir, K. Merk, T. Mosler, S. Garg, S. Bunse, D.A. Tirrell, E.M. Schuman, Direct visualization of newly synthesized target proteins in situ, Nat Methods 12 (2015) 411–414. https://doi.org/10.1038/nmeth.3319.
[175] T.S. Elliott, F.M. Townsley, A. Bianco, R.J. Ernst, A. Sachdeva, S.J. Elsässer, L. Davis, K. Lang, R. Pisa, S. Greiss, K.S. Lilley, J.W. Chin, Proteome labeling and protein identification in

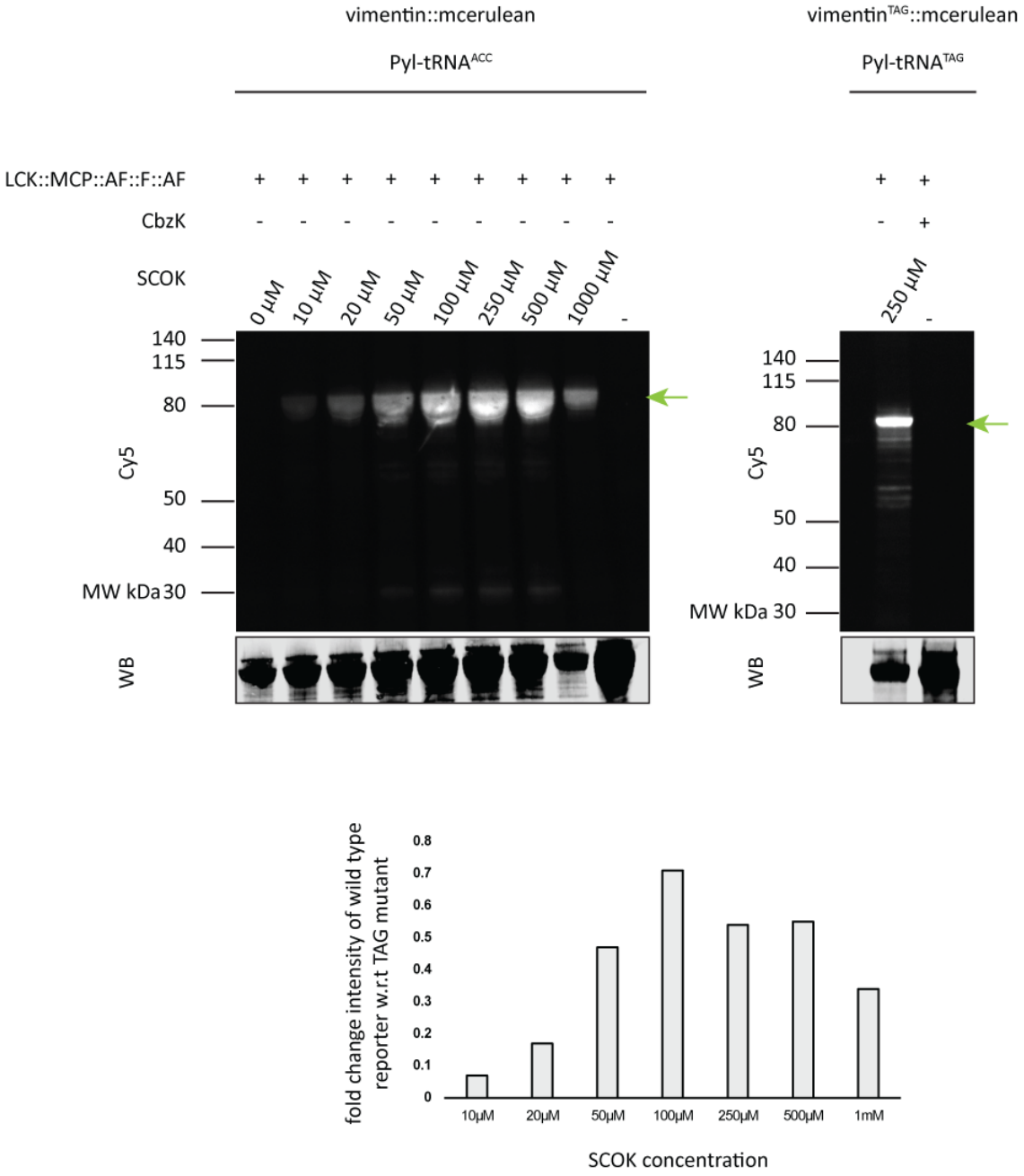
specific tissues and at specific developmental stages in an animal, Nat Biotechnol 32 (2014) 465–472. https://doi.org/10.1038/nbt.2860.
[176] T.S. Elliott, A. Bianco, F.M. Townsley, S.D. Fried, J.W. Chin, Tagging and Enriching Proteins Enables Cell-Specific Proteomics, Cell Chemical Biology 23 (2016) 805–815. https://doi.org/10.1016/j.chembiol.2016.05.018.
[177] W. Ding, W. Yu, Y. Chen, L. Lao, Y. Fang, C. Fang, H. Zhao, B. Yang, S. Lin, Rare codon recoding for efficient noncanonical amino acid incorporation in mammalian cells, Science 384 (2024) 1134–1142. https://doi.org/10.1126/science.adm8143.
[178] K.J. Rohilla, K.N. Ovington, A.A. Pater, M. Barton, A.J. Henke, K.T. Gagnon, Systematic microsatellite repeat expansion cloning and validation, Hum Genet 139 (2020) 1233–1246. https://doi.org/10.1007/s00439-020-02165-z.
[179] S. Bolte, F.P. Cordelières, A guided tour into subcellular colocalization analysis in light microscopy, Journal of Microscopy 224 (2006) 213–232. https://doi.org/10.1111/j.1365-2818.2006.01706.x.
[180] Y. Nakamura, T. Gojobori, T. Ikemura, Codon usage tabulated from international DNA sequence databases: status for the year 2000, Nucleic Acids Res 28 (2000) 292. https://doi.org/10.1093/nar/28.1.292.
[181] D. Dormann, C. Haass, Fused in sarcoma (FUS): an oncogene goes awry in neurodegeneration, Mol Cell Neurosci 56 (2013) 475–486. https://doi.org/10.1016/j.mcn.2013.03.006.
[182] T. Ikemura, Codon usage and tRNA content in unicellular and multicellular organisms., Molecular Biology and Evolution 2 (1985) 13–34. https://doi.org/10.1093/oxfordjournals.molbev.a040335.
[183] A.R. Parrish, X. She, Z. Xiang, I. Coin, Z. Shen, S.P. Briggs, A. Dillin, L. Wang, Expanding the Genetic Code of Caenorhabditis elegans Using Bacterial Aminoacyl-tRNA Synthetase/tRNA Pairs, ACS Chem. Biol. 7 (2012) 1292–1302. https://doi.org/10.1021/cb200542j.
[184] I. Coin, M.H. Perrin, W.W. Vale, L. Wang, Photo-Cross-Linkers Incorporated into G-Protein-Coupled Receptors in Mammalian Cells: A Ligand Comparison, Angewandte Chemie International Edition 50 (2011) 8077–8081. https://doi.org/10.1002/anie.201102646.
[185] W.H. Schmied, S.J. Elsässer, C. Uttamapinant, J.W. Chin, Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1, J. Am. Chem. Soc. 136 (2014) 15577–15583. https://doi.org/10.1021/ja5069728.
[186] A. Iriarte, G. Lamolle, H. Musto, Codon Usage Bias: An Endless Tale, J Mol Evol 89 (2021) 589–593. https://doi.org/10.1007/s00239-021-10027-z.
[187] S.G.E. Andersson, P.M. Sharp, Codon usage in the Mycobacterium tuberculosis complex, Microbiology 142 (1996) 915–925. https://doi.org/10.1099/00221287-142-4-915.
[188] H. Musto, H. Rodriguez-Maseda, G. Bernardi, Compositional properties of nuclear genes from *Plasmodium falciparum*, Gene 152 (1995) 127–132. https://doi.org/10.1016/0378-1119(94)00708-Z.
[189] J.B. Plotkin, G. Kudla, Synonymous but not the same: the causes and consequences of codon bias, Nat Rev Genet 12 (2011) 32–42. https://doi.org/10.1038/nrg2899.
[190] S.T. Parvathy, V. Udayasuriyan, V. Bhadana, Codon usage bias, Mol Biol Rep 49 (2022) 539–565. https://doi.org/10.1007/s11033-021-06749-4.
[191] M.K. Phillips-Jones, F.J. Watson, R. Martin, The 3' Codon Context Effect on UAG Suppressor tRNA is Different in *Escherichia coli* and Human Cells, Journal of Molecular Biology 233 (1993) 1–6. https://doi.org/10.1006/jmbi.1993.1479.
[192] K.K. McCaughan, C.M. Brown, M.E. Dalphin, M.J. Berry, W.P. Tate, Translational termination efficiency in mammals is influenced by the base following the stop codon., Proceedings of the National Academy of Sciences 92 (1995) 5431–5435. https://doi.org/10.1073/pnas.92.12.5431.

[193] M.D. Bartoschek, E. Ugur, T. Nguyen, G. Rodschinka, M. Wierer, K. Lang, S. Bultmann, Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells, Nucleic Acids Research 49 (2021) e62. https://doi.org/10.1093/nar/gkab132.
[194] E.F. Joest, C. Winter, J.S. Wesalo, A. Deiters, R. Tampé, Efficient Amber Suppression via Ribosomal Skipping for In Situ Synthesis of Photoconditional Nanobodies, ACS Synth. Biol. 11 (2022) 1466–1476. https://doi.org/10.1021/acssynbio.1c00471.
[195] N. Amrani, M.S. Sachs, A. Jacobson, Early nonsense: mRNA decay solves a translational problem, Nat Rev Mol Cell Biol 7 (2006) 415–425. https://doi.org/10.1038/nrm1942.
[196] P.D. Hsu, E.S. Lander, F. Zhang, Development and Applications of CRISPR-Cas9 for Genome Engineering, Cell 157 (2014) 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010.
[197] G. Zhu, X. Zhou, M. Wen, J. Qiao, G. Li, Y. Yao, CRISPR–Cas13: Pioneering RNA Editing for Nucleic Acid Therapeutics, BioDesign Research 6 (2024) 0041. https://doi.org/10.34133/bdr.0041.
[198] D.B.T. Cox, J.S. Gootenberg, O.O. Abudayyeh, B. Franklin, M.J. Kellner, J. Joung, F. Zhang, RNA editing with CRISPR-Cas13, Science 358 (2017) 1019–1027. https://doi.org/10.1126/science.aaq0180.
[199] O.O. Abudayyeh, J.S. Gootenberg, B. Franklin, J. Koob, M.J. Kellner, A. Ladha, J. Joung, P. Kirchgatterer, D.B.T. Cox, F. Zhang, A cytosine deaminase for programmable single-base RNA editing, Science 365 (2019) 382–386. https://doi.org/10.1126/science.aax7063.
[200] Y.-Y. Zhao, M.-W. Mao, W.-J. Zhang, J. Wang, H.-T. Li, Y. Yang, Z. Wang, J.-W. Wu, Expanding RNA binding specificity and affinity of engineered PUF domains, Nucleic Acids Res 46 (2018) 4771–4782. https://doi.org/10.1093/nar/gky134.
[201] C.-G. Cheong, T.M.T. Hall, Engineering RNA sequence specificity of Pumilio repeats, Proceedings of the National Academy of Sciences 103 (2006) 13635–13639. https://doi.org/10.1073/pnas.0606294103.
[202] K.P. Adamala, D.A. Martin-Alarcon, E.S. Boyden, Programmable RNA-binding protein composed of repeats of a single modular unit, Proceedings of the National Academy of Sciences 113 (2016) E2579–E2588. https://doi.org/10.1073/pnas.1519368113.
[203] A. Chatterjee, S.B. Sun, J.L. Furman, H. Xiao, P.G. Schultz, A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli, Biochemistry 52 (2013) 10.1021/bi4000244. https://doi.org/10.1021/bi4000244.
[204] J.S. Italia, P.S. Addy, S.B. Erickson, J.C. Peeler, E. Weerapana, A. Chatterjee, Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites, J. Am. Chem. Soc. 141 (2019) 6204–6212. https://doi.org/10.1021/jacs.8b12954.
[205] H. Xiao, A. Chatterjee, S. Choi, K.M. Bajjuri, S.C. Sinha, P.G. Schultz, Genetic Incorporation of Multiple Unnatural Amino Acids into Proteins in Mammalian Cells, Angewandte Chemie International Edition 52 (2013) 14080–14083. https://doi.org/10.1002/anie.201308137.
[206] Y. Zheng, P.S. Addy, R. Mukherjee, A. Chatterjee, Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells, Chem. Sci. 8 (2017) 7211–7217. https://doi.org/10.1039/C7SC02560B.
[207] B. Meineke, J. Heimgärtner, R. Caridha, M.F. Block, K.J. Kimler, M.F. Pires, M. Landreh, S.J. Elsässer, Dual stop codon suppression in mammalian cells with genomically integrated genetic code expansion machinery, Cell Reports Methods 3 (2023) 100626. https://doi.org/10.1016/j.crmeth.2023.100626.
[208] R. Jiang, J.A. Krzycki, PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine, J Biol Chem 287 (2012) 32738–32746. https://doi.org/10.1074/jbc.M112.396754.

[209] T. Suzuki, C. Miller, L.-T. Guo, J.M.L. Ho, D.I. Bryson, Y.-S. Wang, D.R. Liu, D. Söll, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat Chem Biol 13 (2017) 1261–1266. https://doi.org/10.1038/nchembio.2497.
[210] J.M. Kavran, S. Gundllapalli, P. O'Donoghue, M. Englert, D. Söll, T.A. Steitz, Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation, Proc Natl Acad Sci U S A 104 (2007) 11268–11273. https://doi.org/10.1073/pnas.0704769104.
[211] K. Nozawa, P. O'Donoghue, S. Gundllapalli, Y. Araiso, R. Ishitani, T. Umehara, D. Söll, O. Nureki, Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality, Nature 457 (2009) 1163–1167. https://doi.org/10.1038/nature07611.
[212] S. Herring, A. Ambrogelly, S. Gundllapalli, P. O'Donoghue, C.R. Polycarpo, D. Söll, The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity, FEBS Lett 581 (2007) 3197–3203. https://doi.org/10.1016/j.febslet.2007.06.004.
[213] J.C.W. Willis, J.W. Chin, Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs, Nature Chem 10 (2018) 831–837. https://doi.org/10.1038/s41557-018-0052-5.
[214] D.L. Dunkelmann, J.C.W. Willis, A.T. Beattie, J.W. Chin, Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids, Nat. Chem. 12 (2020) 535–544. https://doi.org/10.1038/s41557-020-0472-x.
[215] A.T. Beattie, D.L. Dunkelmann, J.W. Chin, Quintuply orthogonal pyrrolysyl-tRNA synthetase/tRNAPyl pairs, Nat Chem 15 (2023) 948–959. https://doi.org/10.1038/s41557-023-01232-y.
[216] S.F. Banani, H.O. Lee, A.A. Hyman, M.K. Rosen, Biomolecular condensates: organizers of cellular biochemistry, Nat Rev Mol Cell Biol 18 (2017) 285–298. https://doi.org/10.1038/nrm.2017.7.
[217] A. Zhukov, V. Popov, Eukaryotic Cell Membranes: Structure, Composition, Research Methods and Computational Modelling, Int J Mol Sci 24 (2023) 11226. https://doi.org/10.3390/ijms241311226.
[218] M. Beck, E. Hurt, The nuclear pore complex: understanding its function through structural insight, Nat Rev Mol Cell Biol 18 (2017) 73–89. https://doi.org/10.1038/nrm.2016.147.
[219] D.A. Braun, C.E. Sadowski, S. Kohl, S. Lovric, S.A. Astrinidis, W.L. Pabst, H.Y. Gee, S. Ashraf, J.A. Lawson, S. Shril, M. Airik, W. Tan, D. Schapiro, J. Rao, W.-I. Choi, T. Hermle, M.J. Kemper, M. Pohl, F. Ozaltin, M. Konrad, R. Bogdanovic, R. Büscher, U. Helmchen, E. Serdaroglu, R.P. Lifton, W. Antonin, F. Hildebrandt, Mutations in nuclear pore genes NUP93, NUP205 and XPO5 cause steroid-resistant nephrotic syndrome, Nat Genet 48 (2016) 457–465. https://doi.org/10.1038/ng.3512.
[220] E. Bonnin, P. Cabochette, A. Filosa, R. Jühlen, S. Komatsuzaki, M. Hezwani, A. Dickmanns, V. Martinelli, M. Vermeersch, L. Supply, N. Martins, L. Pirenne, G. Ravenscroft, M. Lombard, S. Port, C. Spillner, S. Janssens, E. Roets, J. Van Dorpe, M. Lammens, R.H. Kehlenbach, R. Ficner, N.G. Laing, K. Hoffmann, B. Vanhollebeke, B. Fahrenkrog, Biallelic mutations in nucleoporin NUP88 cause lethal fetal akinesia deformation sequence, PLoS Genet 14 (2018) e1007845. https://doi.org/10.1371/journal.pgen.1007845.
[221] X. Zhang, S. Chen, S. Yoo, S. Chakrabarti, T. Zhang, T. Ke, C. Oberti, S.L. Yong, F. Fang, L. Li, R. de la Fuente, L. Wang, Q. Chen, Q.K. Wang, Mutation in nuclear pore component NUP155 leads to atrial fibrillation and early sudden cardiac death, Cell 135 (2008) 1017–1027. https://doi.org/10.1016/j.cell.2008.10.022.
[222] C. Strambio-De-Castillia, M. Niepel, M.P. Rout, The nuclear pore complex: bridging nuclear transport and gene regulation, Nat Rev Mol Cell Biol 11 (2010) 490–501. https://doi.org/10.1038/nrm2928.
[223] P. Fontana, Y. Dong, X. Pi, A.B. Tong, C.W. Hecksel, L. Wang, T.-M. Fu, C. Bustamante, H. Wu, Structure of cytoplasmic ring of nuclear pore complex by integrative cryo-EM and AlphaFold, Science 376 (2022) eabm9326. https://doi.org/10.1126/science.abm9326.

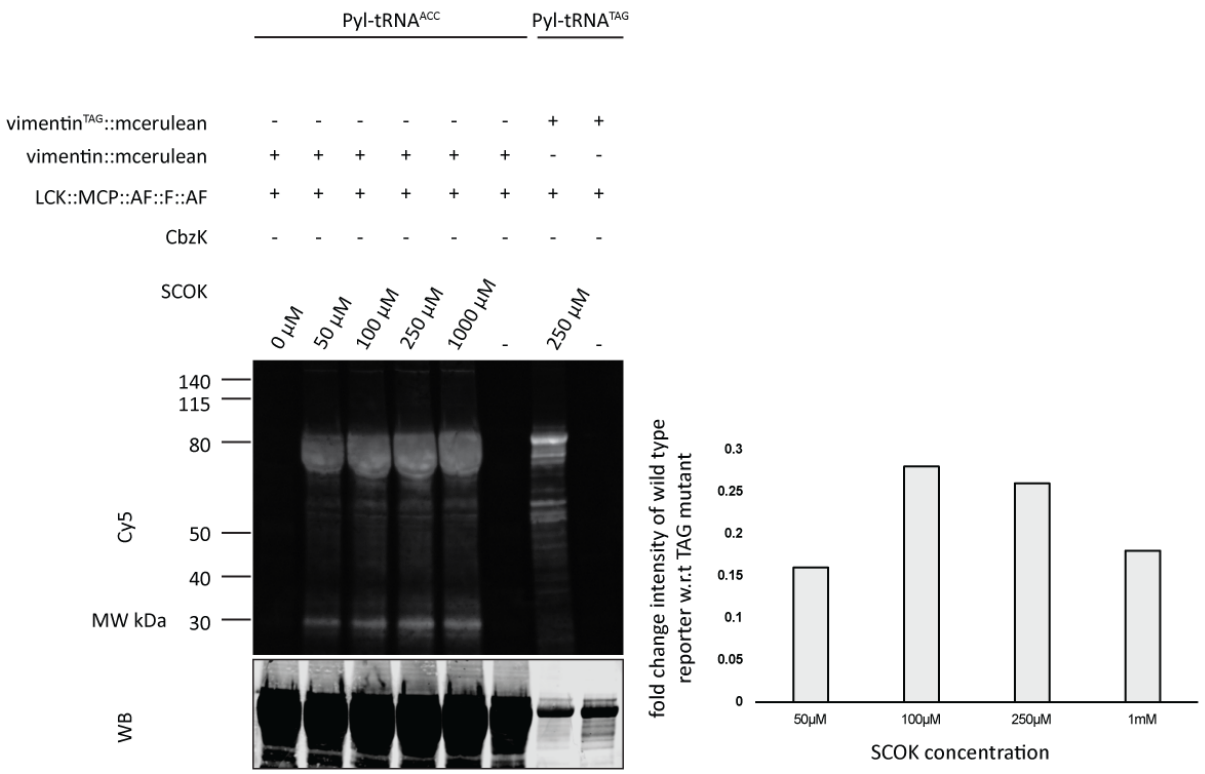
[224] C.J. Bley, S. Nie, G.W. Mobbs, S. Petrovic, A.T. Gres, X. Liu, S. Mukherjee, S. Harvey, F.M. Huber, D.H. Lin, B. Brown, A.W. Tang, E.J. Rundlet, A.R. Correia, S. Chen, S.G. Regmi, T.A. Stevens, C.A. Jette, M. Dasso, A. Patke, A.F. Palazzo, A.A. Kossiakoff, A. Hoelz, Architecture of the cytoplasmic face of the nuclear pore, Science 376 (2022) eabm9129. https://doi.org/10.1126/science.abm9129.
[225] X. Zhu, G. Huang, C. Zeng, X. Zhan, K. Liang, Q. Xu, Y. Zhao, P. Wang, Q. Wang, Q. Zhou, Q. Tao, M. Liu, J. Lei, C. Yan, Y. Shi, Structure of the cytoplasmic ring of the Xenopus laevis nuclear pore complex, Science 376 (2022) eabl8280. https://doi.org/10.1126/science.abl8280.
[226] S. Petrovic, D. Samanta, T. Perriches, C.J. Bley, K. Thierbach, B. Brown, S. Nie, G.W. Mobbs, T.A. Stevens, X. Liu, G.P. Tomaleri, L. Schaus, A. Hoelz, Architecture of the linker-scaffold in the nuclear pore, Science 376 (2022) eabm9798. https://doi.org/10.1126/science.abm9798.

Supplementary data

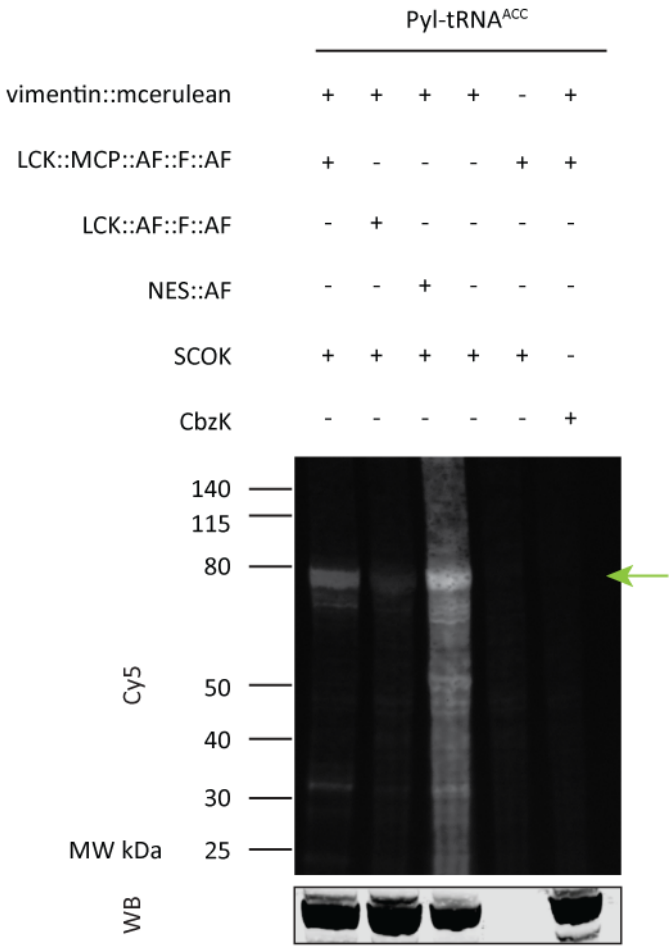


Supplementary Figure 1| In-gel assay to estimate the number of ncAAs incorporated in vimentin::mcerulean. (A)HEK293T cells were transfected with plasmids expressing the OTO LCK::MCP::AF::F::AF, the reporter vimentin::mcerulean and Pyl-tRNA$^{ACC.}$ The ACC codon was reassigned in vimentin::mcerulean by LCK::MCP::AF::F::AF to incorporate SCOK or CbzK. ACC codon was reassigned in vimentin::mcerulean under varying SCOK concentrations, resulting in 8 different samples where GCE was performed with 0 µM, 10 µM, 20 µM, 50 µM, 100 µM, 250 µM, 500 µM and 1 mM SCOK respectively. Vimentin::mcerulean was purified with GFP beads from HEK293T cells and labelled with Cy5-H-Tetrazine. SDS-PAGE was performed with equal volume of labelled protein from each sample, followed by western blot. The proteins modified with SCOK should be visible in the Cy5 scan of the blot since Cy5-H-Tetrazine click-reacts with SCOK by IEDDA reaction. The sample where CbzK (cells were incubated with 1mM CbzK) was incorporated should not be visible in the Cy5 scan of the blot since CbzK does not react with Cy5-H-Tetrazine. Height of vimentin::mcerulean is marked by a green arrow on the Cy5 scan. The blot was also

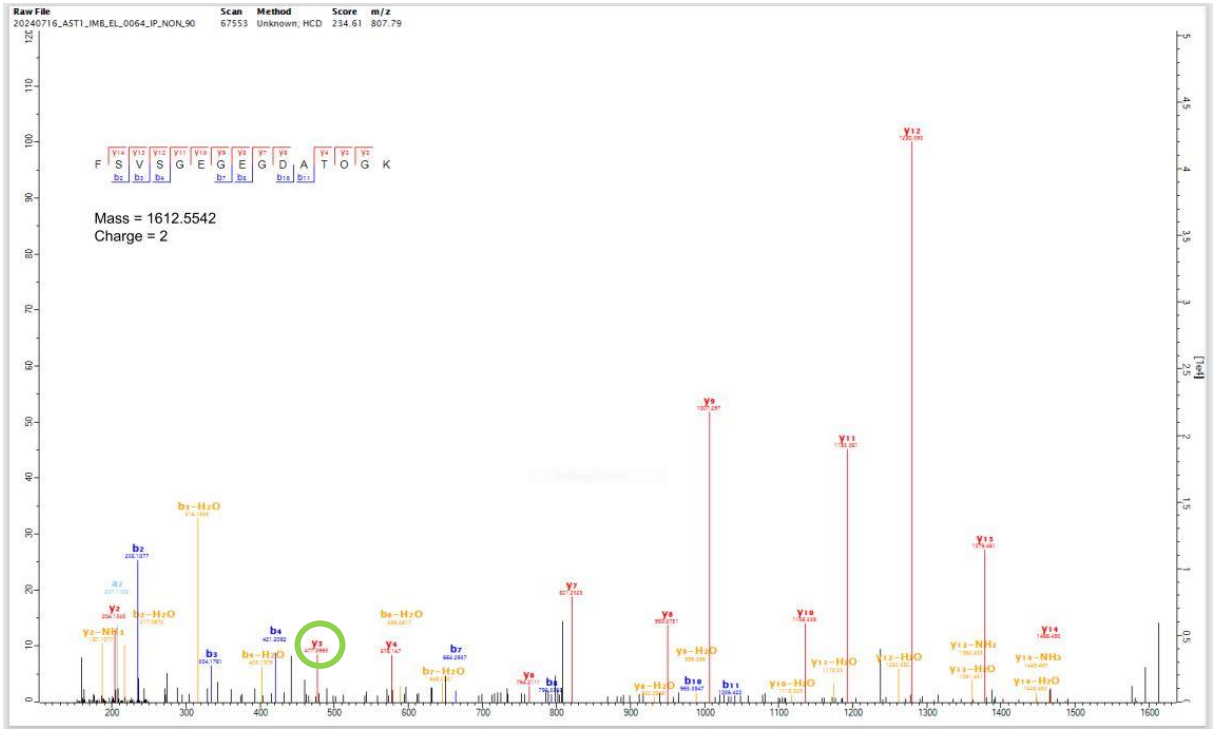
immunolabelled against the myc antibody fused N-terminally to vimentin::mcerulean. The immunolabel scan of the blot is marked as WB (B) HEK293T cells were transfected with plasmids expressing the OTO LCK::MCP::AF::F::AF, the reporter vimentin^{116TAG}::mcerulean and Pyl-tRNA$^{TAG.}$ The TAG codon was suppressed in vimentin::mcerulean by LCK::MCP::AF::F::AF to incorporate SCOK or CbzK (ncAA concentration used was 250μM, which was the established optimized ncAA concentration for TAG suppression). Vimentin^{116TAG}::mcerulean was purified with GFP beads and labelled with Cy5-H-Tetrazine. Vimentin^{116TAG}::mcerulean is marked by a green arrow. SDS-PAGE and western blot was performed with equal volume of each sample. Similar to (A) proteins incorporated with SCOK was visible in the Cy5 scan of the gel and those with CbzK were not visible. The blot was also immunolabelled against the myc antibody fused N-terminally to vimentin^{116TAG}::mcerulean. The immunolabel scan of the blot is marked as WB. (C) Fold change intensity was calculated as described in section 3.2.4. For all calculations Cy5 intensity of each band was divided by immunolabel intensity of the same band to account for different amount of total POI loaded for different samples. The interpretation of fold change intensity is also described in section 3.2.4. A fold change intensity lower than 1fold signifies that incorporation of multiple ncAAs could not be proven.



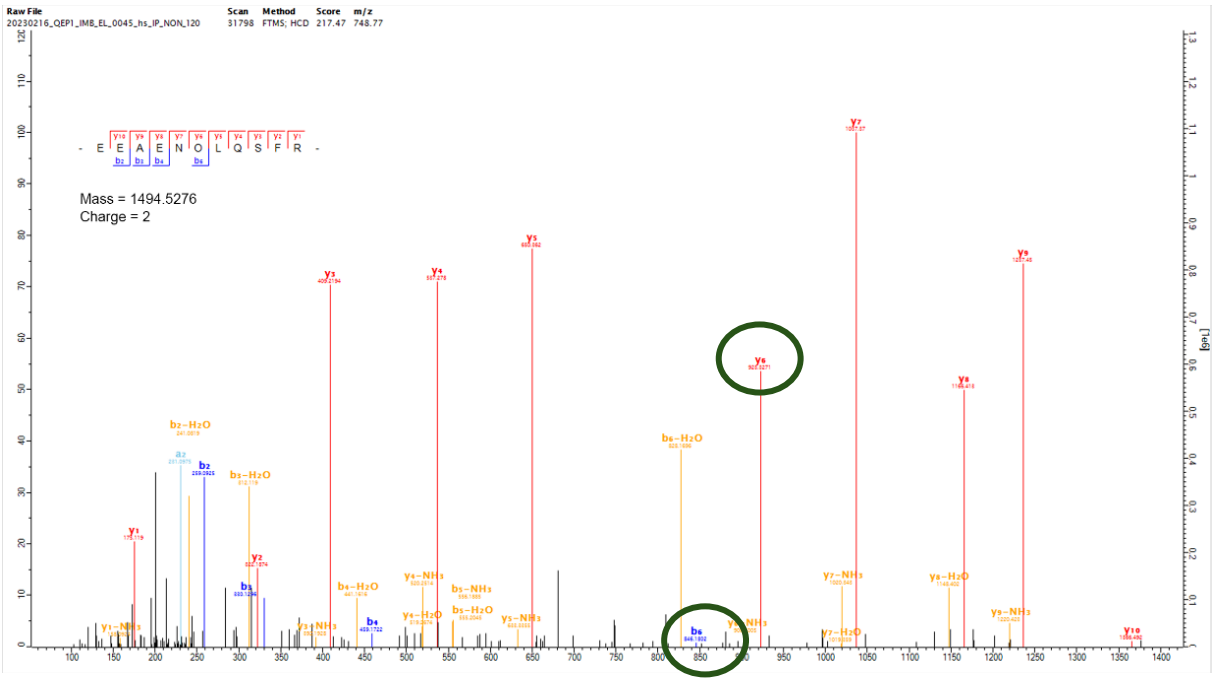
Supplementary Figure 2| In-gel assay to estimate the number of ncAAs incorporated in vimentin::mcerulean. (left) HEK293T cells were transfected with plasmids expressing the OTO LCK::MCP::AF::F::AF, the reporter vimentin::mcerulean or vimentin^{116TAG}::mcerulean and Pyl-tRNA^{ACC} or Pyl-tRNA$^{TAG.}$ The ACC codon was reassigned in vimentin::mcerulean by LCK::MCP::AF::F::AF to incorporate SCOK or CbzK (cells were incubated with 1mM CbzK). ACC codon was reassigned in vimentin::mcerulean under varying SCOK concentrations, resulting in 5 different samples where GCE was performed with 0 µM, 5 0µM, 100 µM,250 µM and 1 mM SCOK respectively. The TAG codon was suppressed in vimentin::mcerulean by LCK::MCP::AF::F::AF to incorporate SCOK or CbzK (ncAA concentration used was 250µM, which was the established optimized ncAA concentration for TAG suppression). Vimentin::mcerulean was purified with GFP beads from HEK293T cells and labelled with Cy5-H-Tetrazine. SDS-PAGE was performed with equal volume of labelled protein from each sample , followed by western blot . The proteins modified with SCOK should be visible in the Cy5 scan of the blot since Cy5-H-Tetrazine click-reacts with SCOK by IEDDA reaction. The sample where CbzK was incorporated should not be visible in the Cy5 scan of the blot since CbzK does not react with Cy5-H-Tetrazine. Height of vimentin::mcerulean/vimentin^{116TAG}::mcerulean is marked by a green arrow on the Cy5 scan of the blot. The blot was also immunolabelled against the myc antibody fused N-terminally to vimentin::mcerulean/ vimentin^{116TAG}::mcerulean. The immunolabel scan of the blot is marked as WB (right) Fold change intensity was calculated as described in section 3.2.4. For all calculations Cy5 intensity of each band was divided by immunolabel intensity of the same band to account for different amount of total POI loaded for different samples. The interpretation of fold change intensity is also described in section 3.2.4. A fold change intensity lower than 1fold signifies that incorporation of multiple ncAAs could not be proven.



Supplementary Figure 3| In-gel assay to demonstrate the effect of OTO in rendering residue-specific sense codon reassignment mRNA selective. The ACC codon was reassigned in vimentin::mcerulean to incorporate SCOK / CbzK in HEK293T cells. Cell lysates were labelled with Cy5-H-Tetrazine followed by western blot. Proteins incorporated with SCOK were visible in the Cy5 scan of the blot since Cy5-H-Tetrazine click reacts with SCOK. Non-clickable CbzK was used as a negative control and proteins incorporated with CbzK were not visible in the Cy5 scan of the blot. Immunolabelling was also performed against an N-terminally fused myc tag to vimentin::mcerulean. The immunolabel scan of the blot is marked as WB (WB – western blot). The sample where GCE was performed with the OTO LCK::MCP::AF::AF shows much lower unspecific Cy5 labelling compared to the sample where the non-selective cytoplasmic NES::AF was used. The sample where the ACC codon was reassigned with the OTO LCK::AF::F::AF (lacks the RBD necessary for recruiting target mRNA to the OTO) only a faint band could be observed for Cy5 labelled vimentin::mcerulean demonstrating the need for the RBD domain i.e MCP for efficient recruitment of the MS2 tagged vimentin::mcerulean mRNA and subsequent sense codon reassignment. The sample with LCK::MCP::AF::F::AF and Pyl-TRNA^{ACC}, without vimentin::mcerulean, does not show any unspecific Cy5 labelled bands. Hence unspecific mRNAs were not recruited to the OTO to a detectable extent in the absence of the target reporter mRNA.

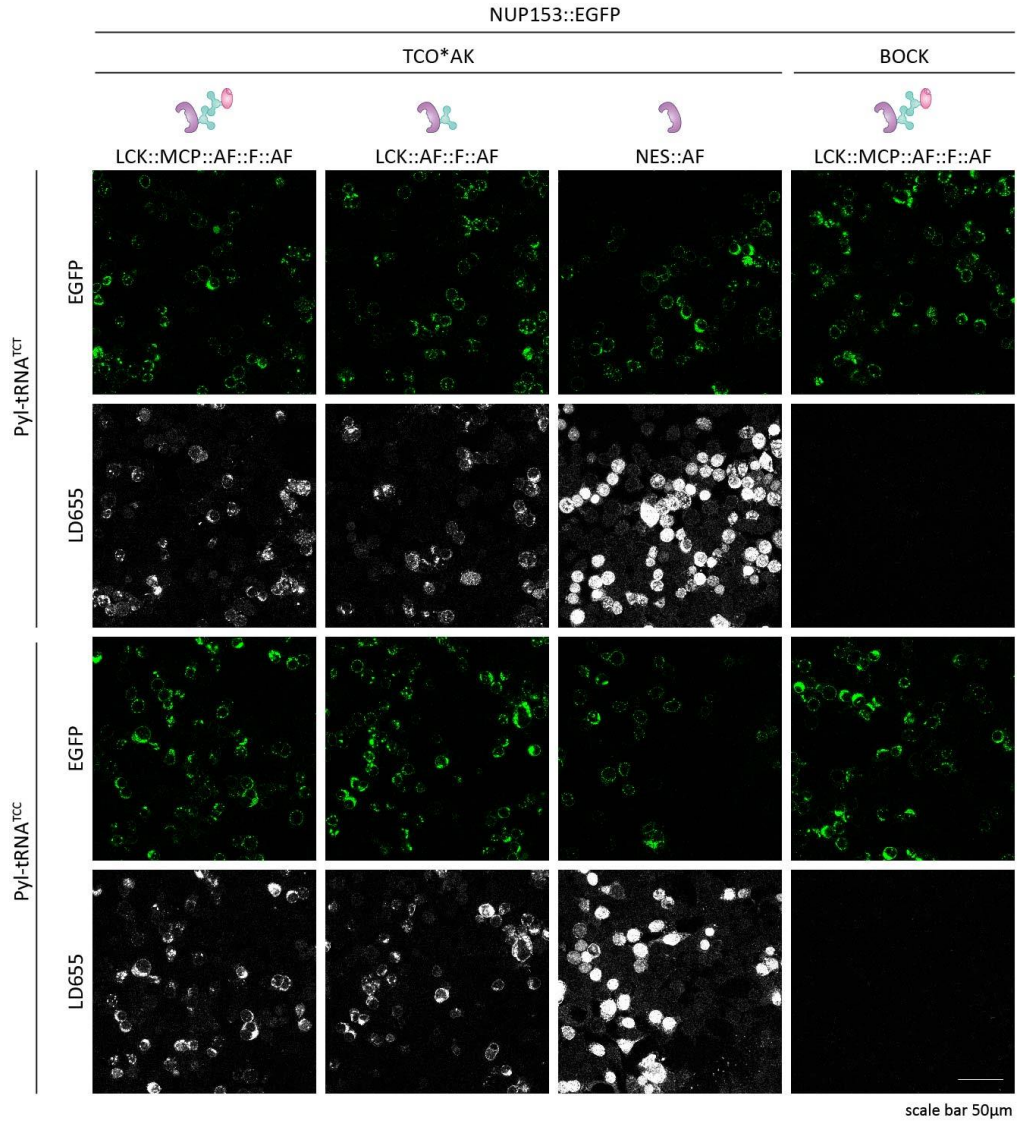


Supplementary Figure 4| Validation of incorporation of ncAA in EGFP^{39CGT} by mRNA selective residue-specific sense codon reassignment by mass spectrometry. The ncAA 3IF was incorporated in EGFP^{39CGT} at site 39 by reassigning the CGT codon by the OTO LCK::MCP::AF::F::AF in HEK293T cells. SDS-PAGE was performed with the cell lysate and the band corresponding to EGFP^{39CGT} was cut out and submitted for mass spectrometry analysis. EGFP^{39CGT} was digested using trypsin prior to the measurement. The peak marked by green circle signify detection of EGFP^{39CGT} fragment with incorporated 3IF.

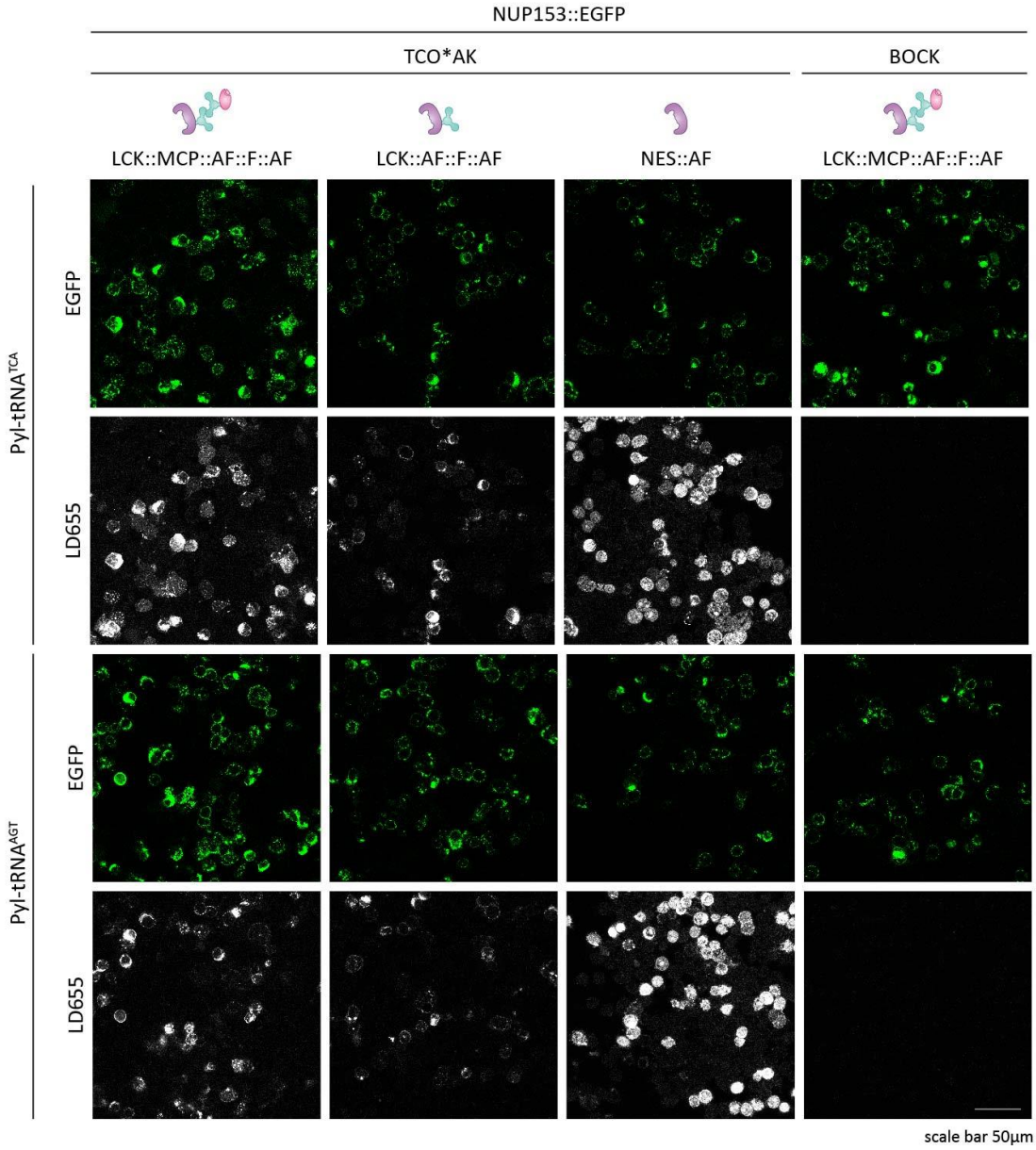


Supplementary Figure 5| Validation of incorporation of ncAA in vimentin::mcerulean by mRNA selective residue-specific sense codon reassignment by mass spectrometry. The ncAA 3IF was incorporated in vimentin::mcerulean by reassigning the ACC codon by the OTO LCK::MCP::AF::F::AF in HEK293T cells. SDS-PAGE was performed with the cell lysate and the band corresponding to vimentin::mcerulean was cut out and submitted for mass spectrometry analysis. Vimentin::mcerulean was digested using trypsin prior to the measurement. The peaks marked by green circle signify detection of vimentin::mcerulean fragments with incorporated 3IF.

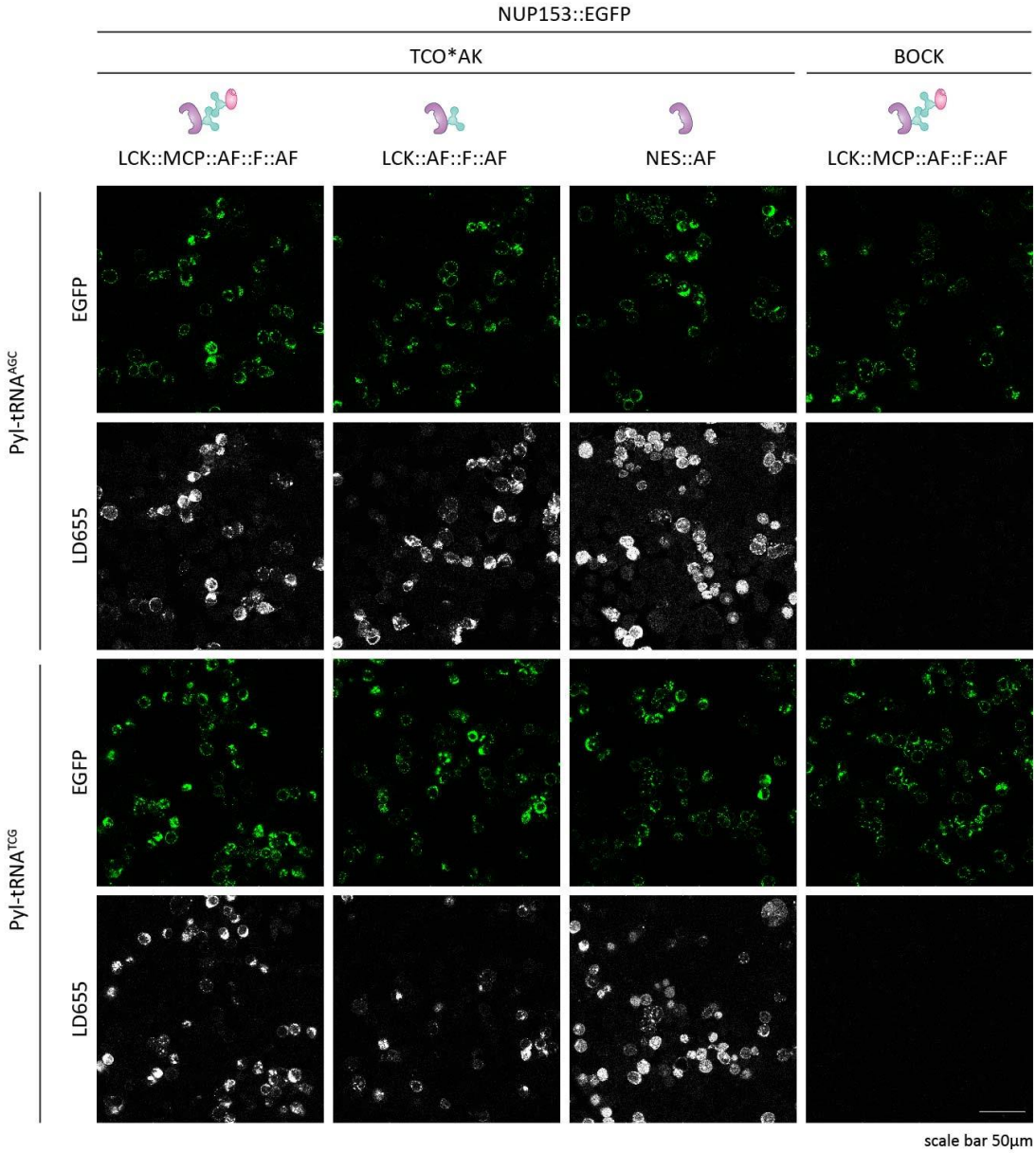
Appendix I



Appendix I Figure 1| Confocal imaging results for codon selection for NUP153::EGFP (codons TCT, TCC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons TCT and TCC individually in NUP153::EGFP.

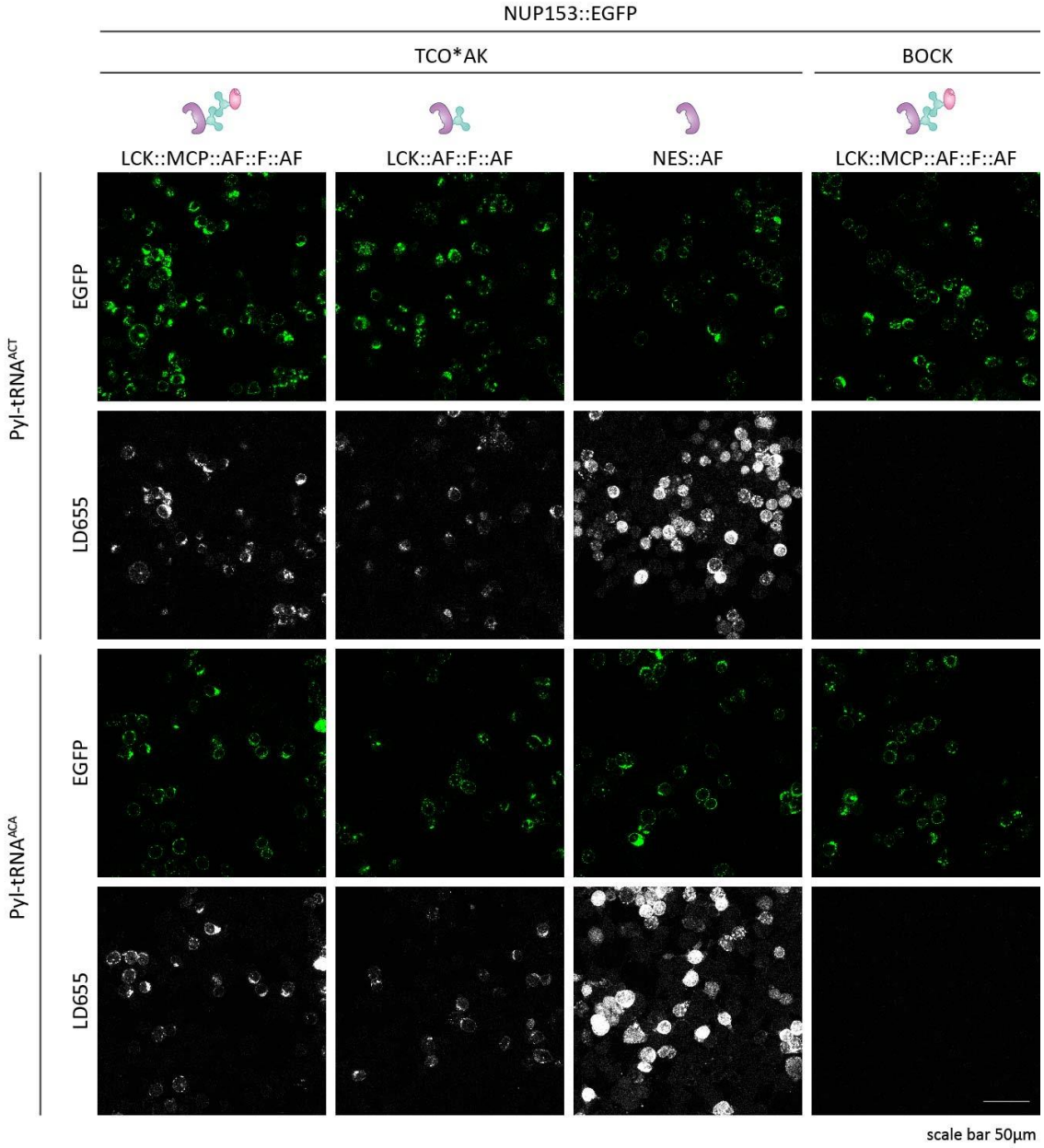


Appendix I Figure 2| Confocal imaging results for codon selection for NUP153::EGFP (codons TCA, AGT). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons TCA and AGT individually in NUP153::EGFP.

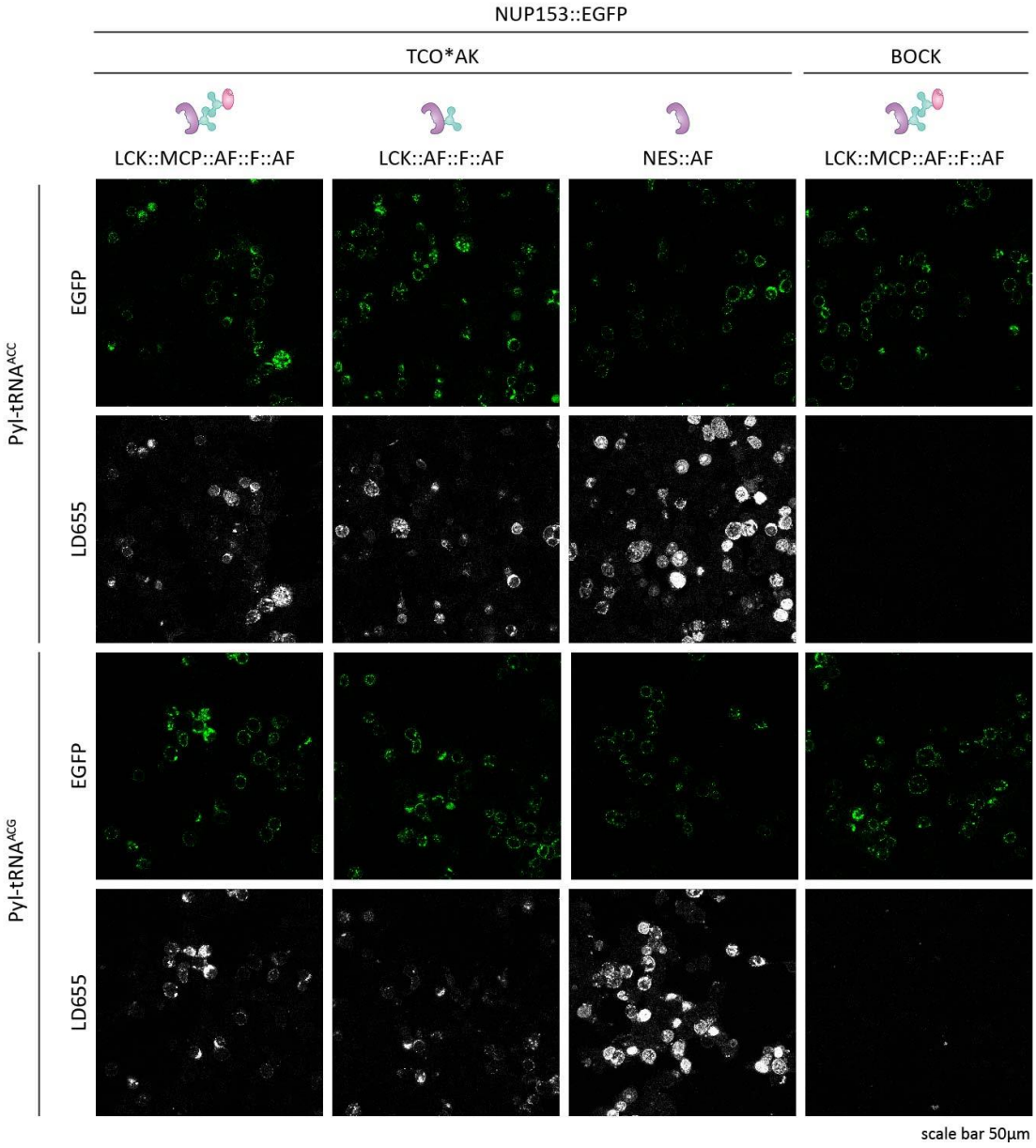


Appendix I Figure 3| Confocal imaging results for codon selection for NUP153::EGFP (codons AGC, TCG). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons AGC and TCG

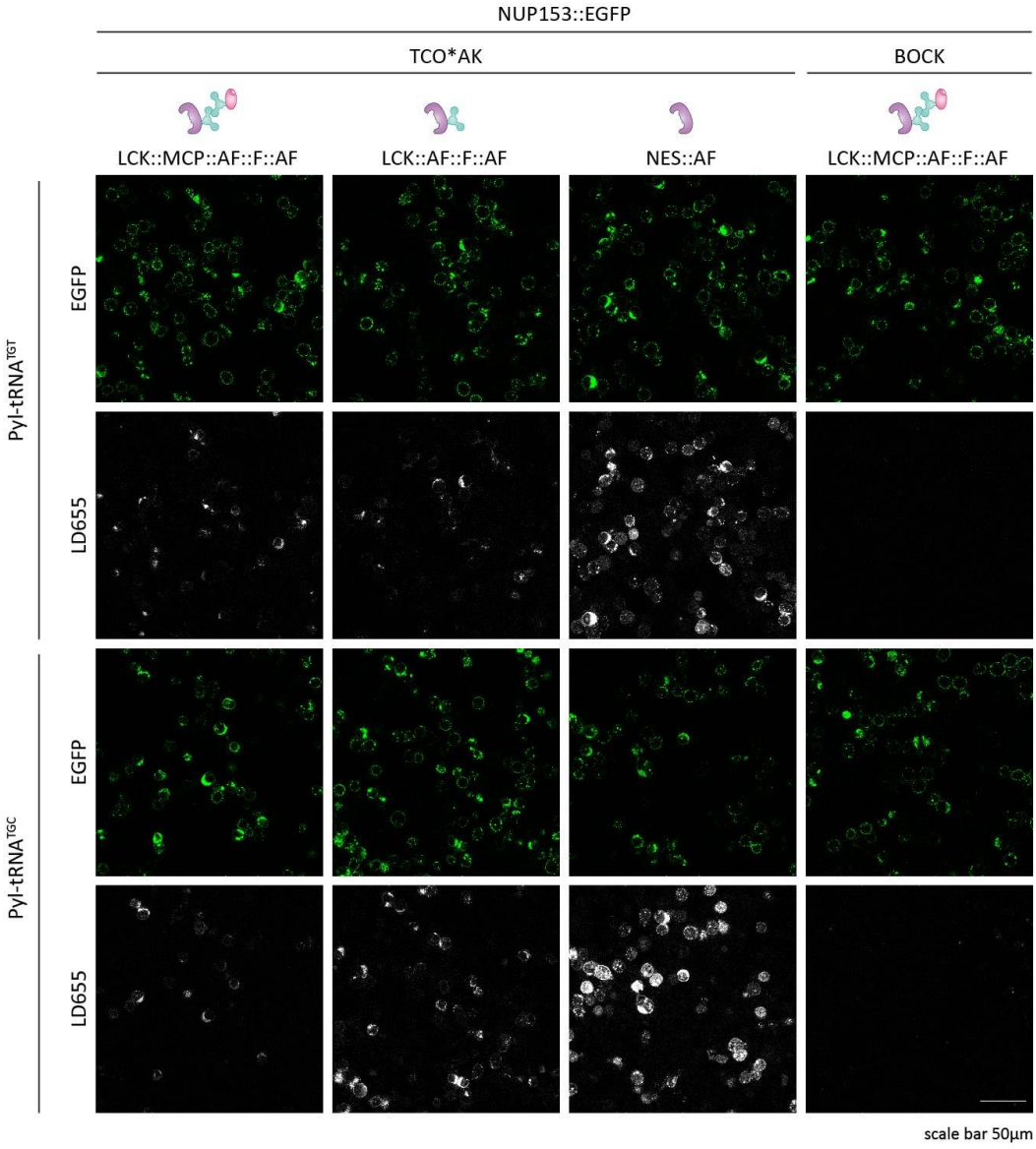
individually in NUP153::EGFP.



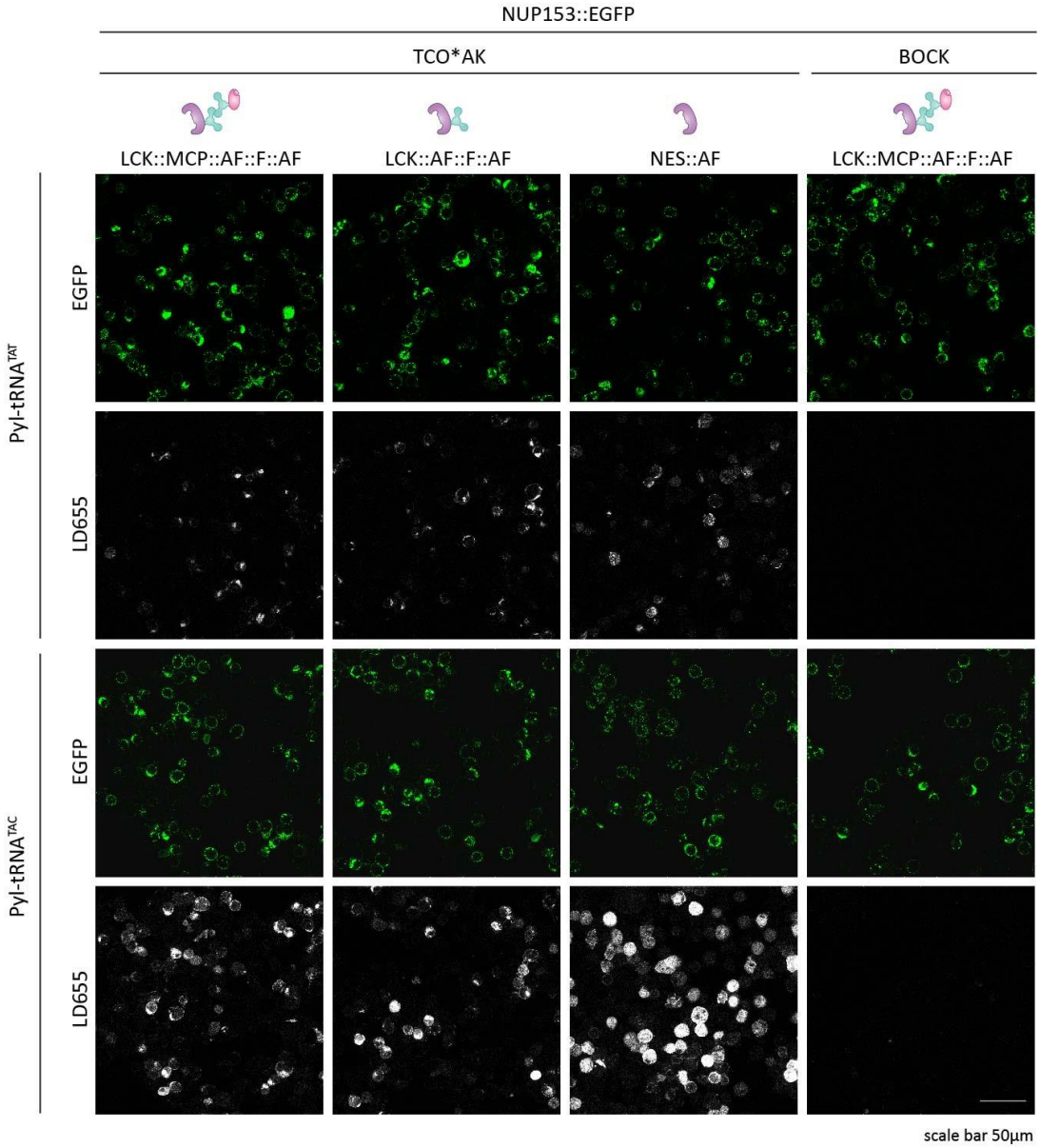
Appendix I Figure 4| Confocal imaging results for codon selection for NUP153::EGFP (codons ACT, ACA). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons ACT and ACA individually in NUP153::EGFP.



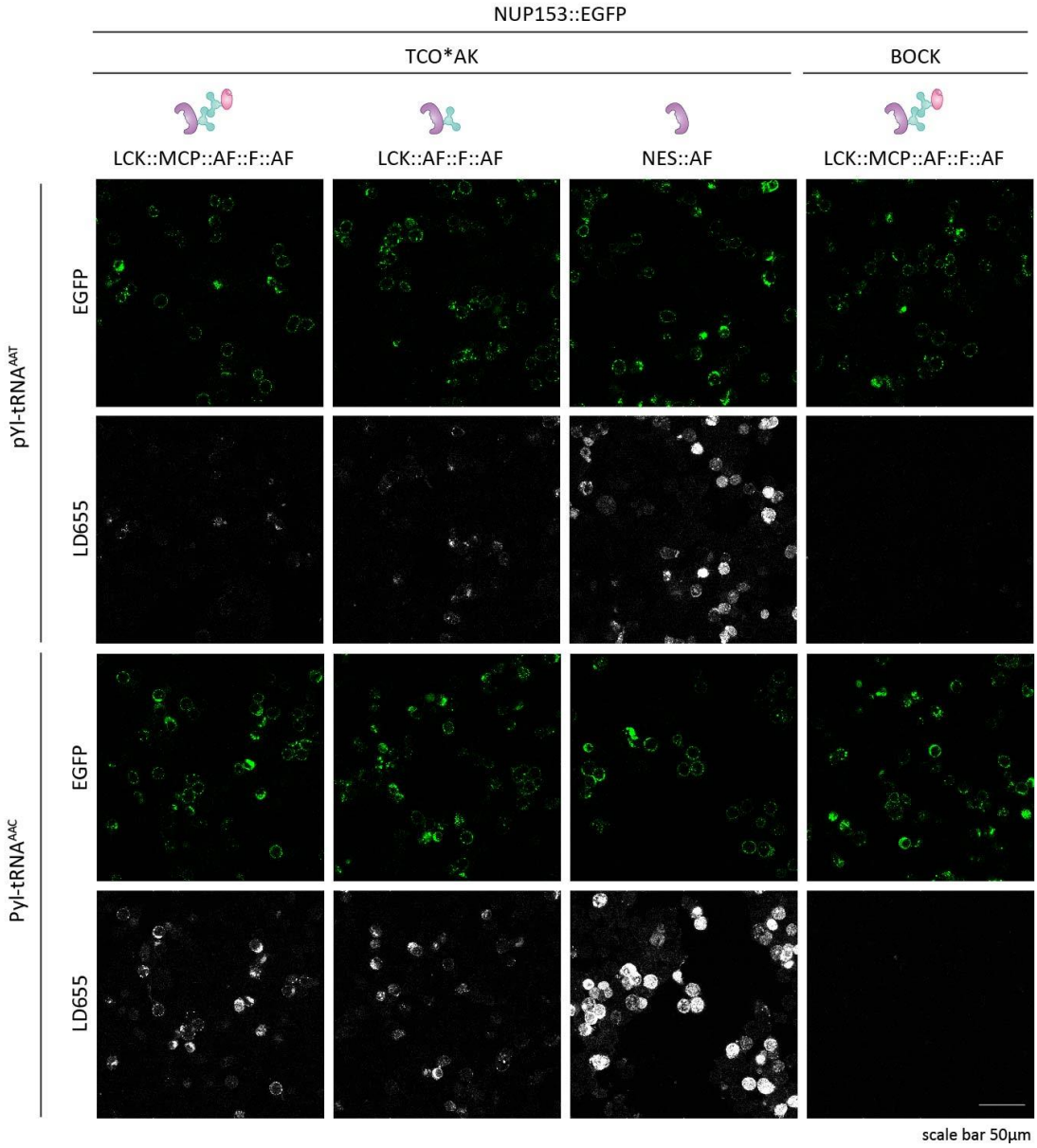
Appendix I Figure 5| Confocal imaging results for codon selection for NUP153::EGFP (codons ACC, ACG). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons ACC and ACG individually in NUP153::EGFP.



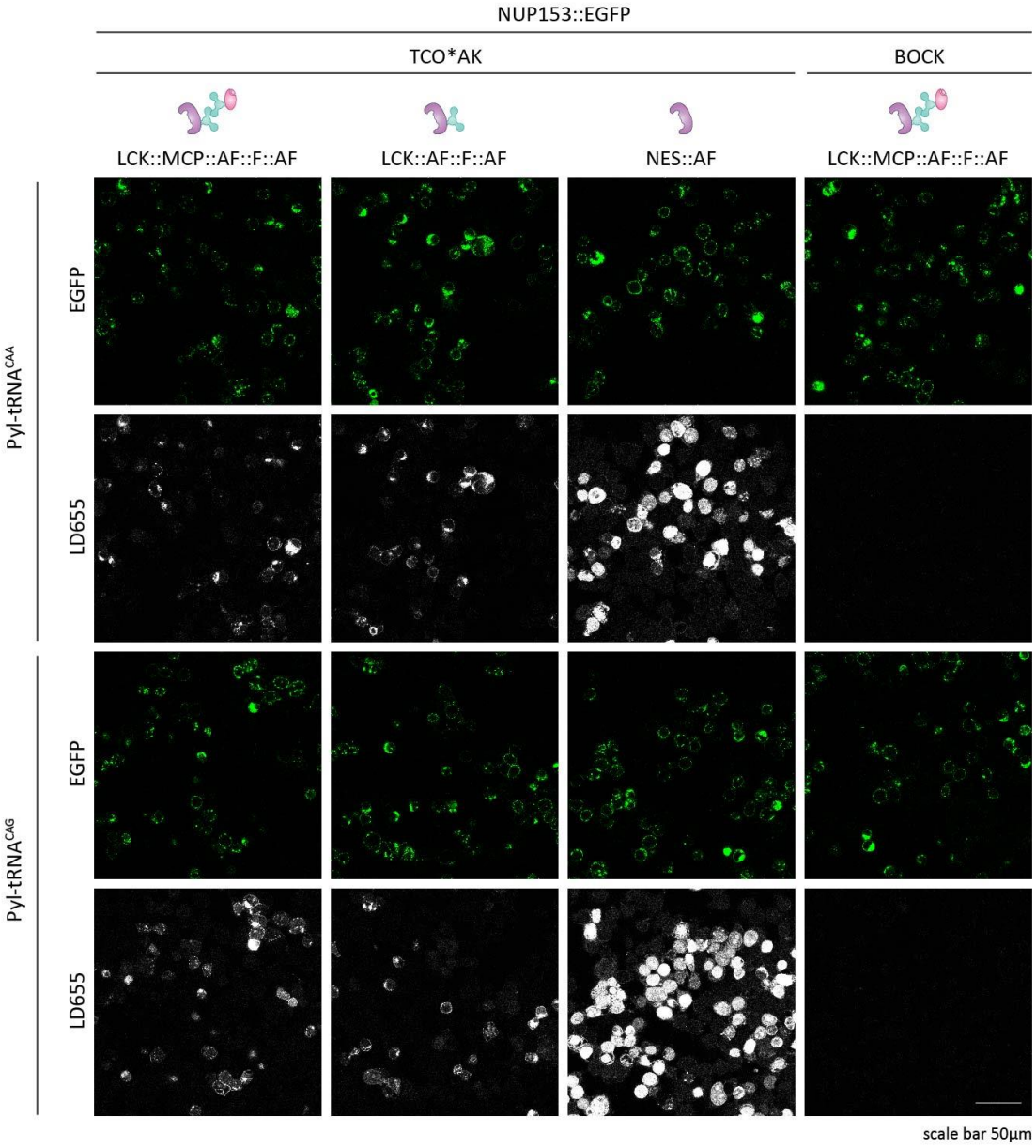
Appendix I Figure 6| Confocal imaging results for codon selection for NUP153::EGFP (codons TGT, TGC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons TGT and TGC individually in NUP153::EGFP.



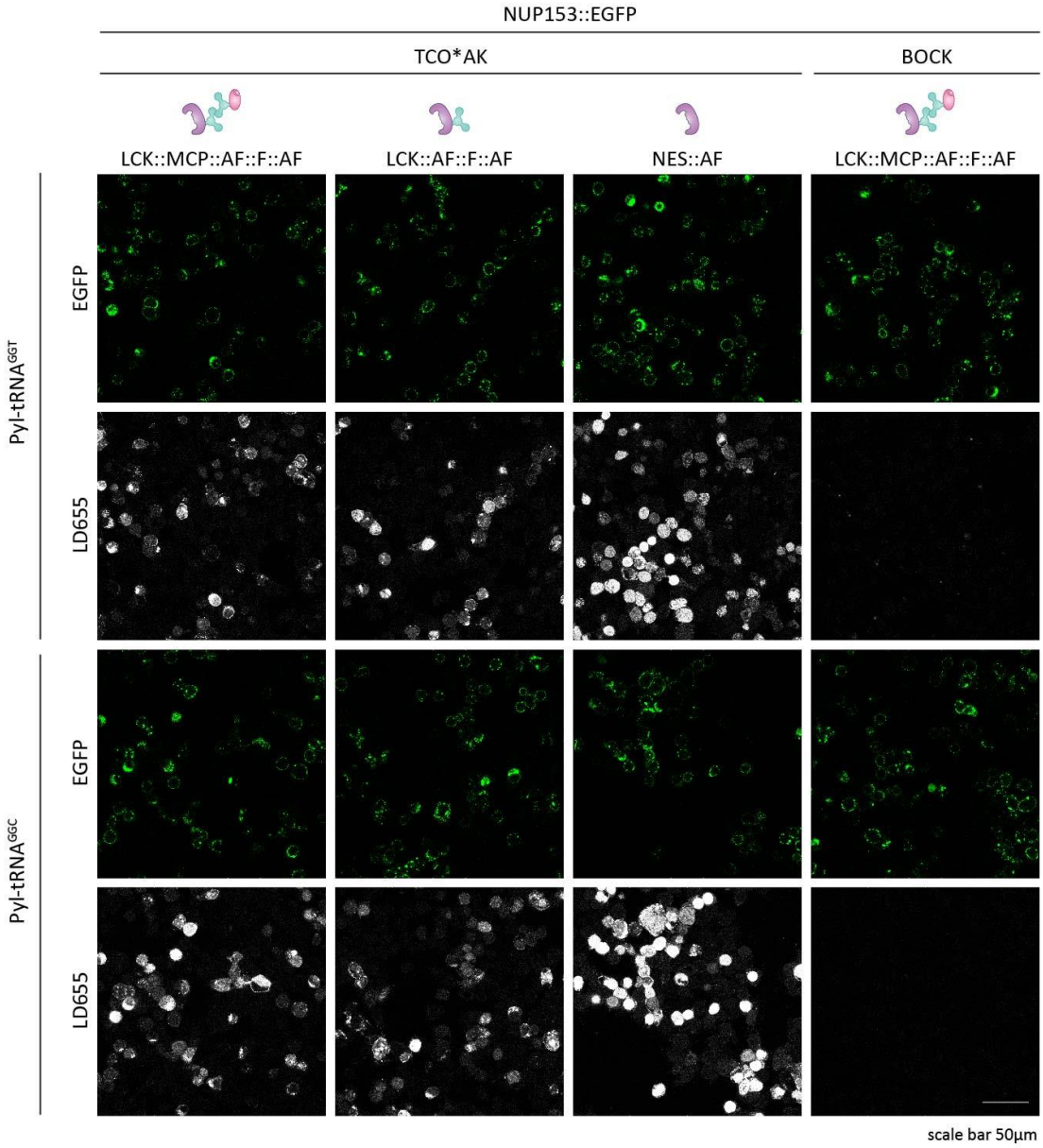
Appendix I Figure 7| Confocal imaging results for codon selection for NUP153::EGFP (codons TAT, TAC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons TAT and TAC individually in NUP153::EGFP.



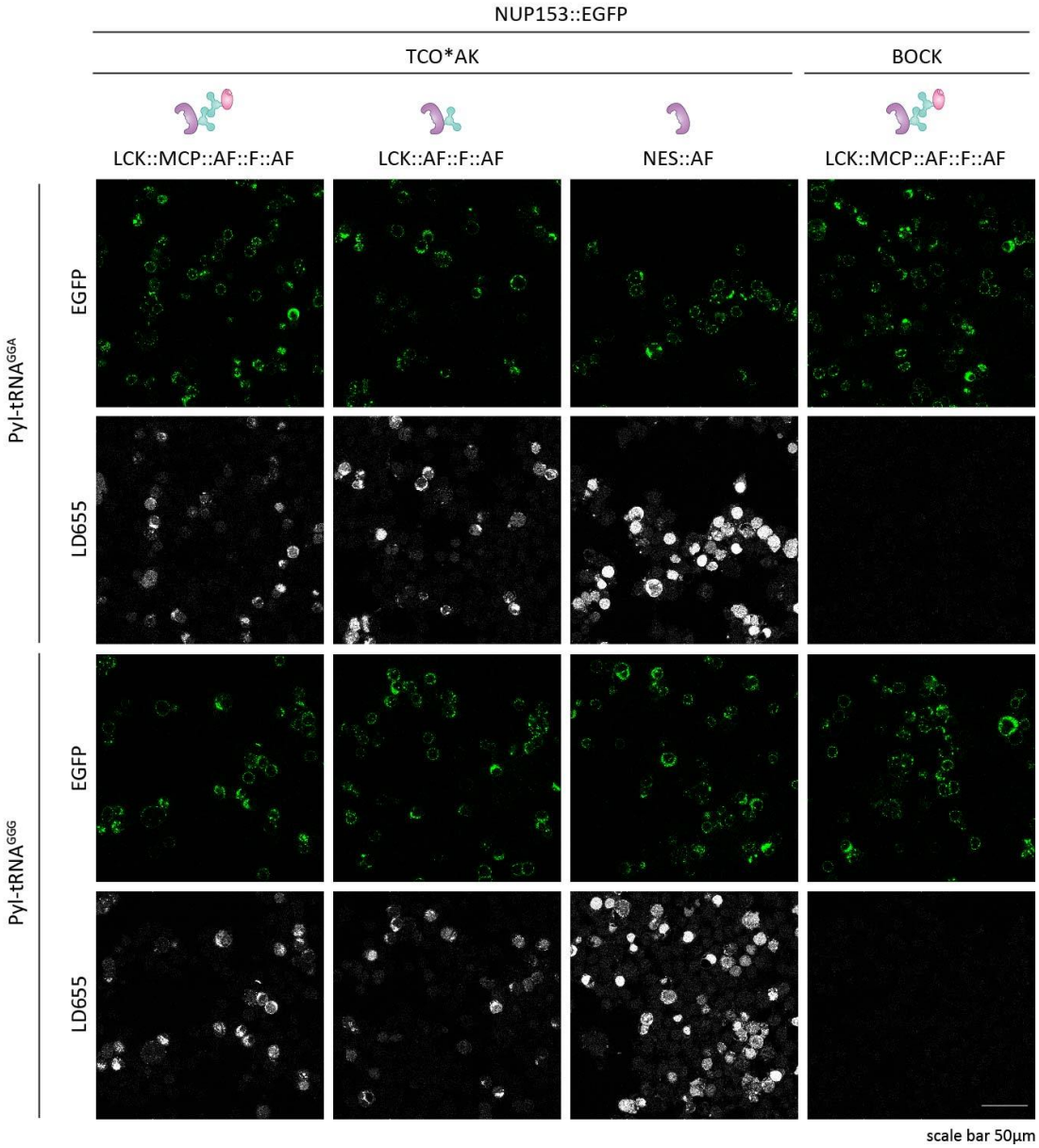
Appendix I Figure 8| Confocal imaging results for codon selection for NUP153::EGFP (codons AAT, AAC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons AAT and AAC individually in NUP153::EGFP.



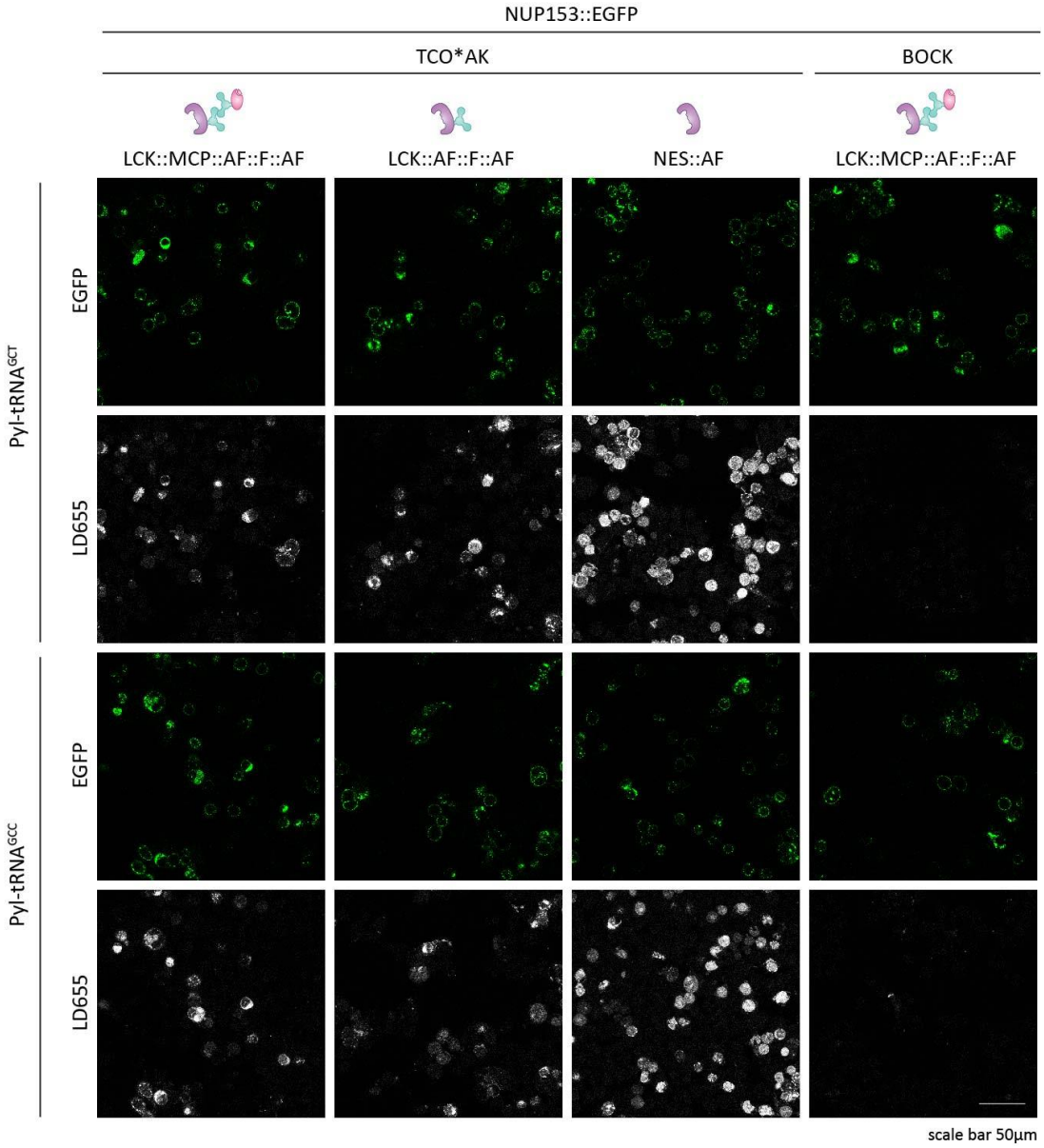
Appendix I Figure 9| Confocal imaging results for codon selection for NUP153::EGFP (codons CAA, CAG). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons CAA and CAG individually in NUP153::EGFP.



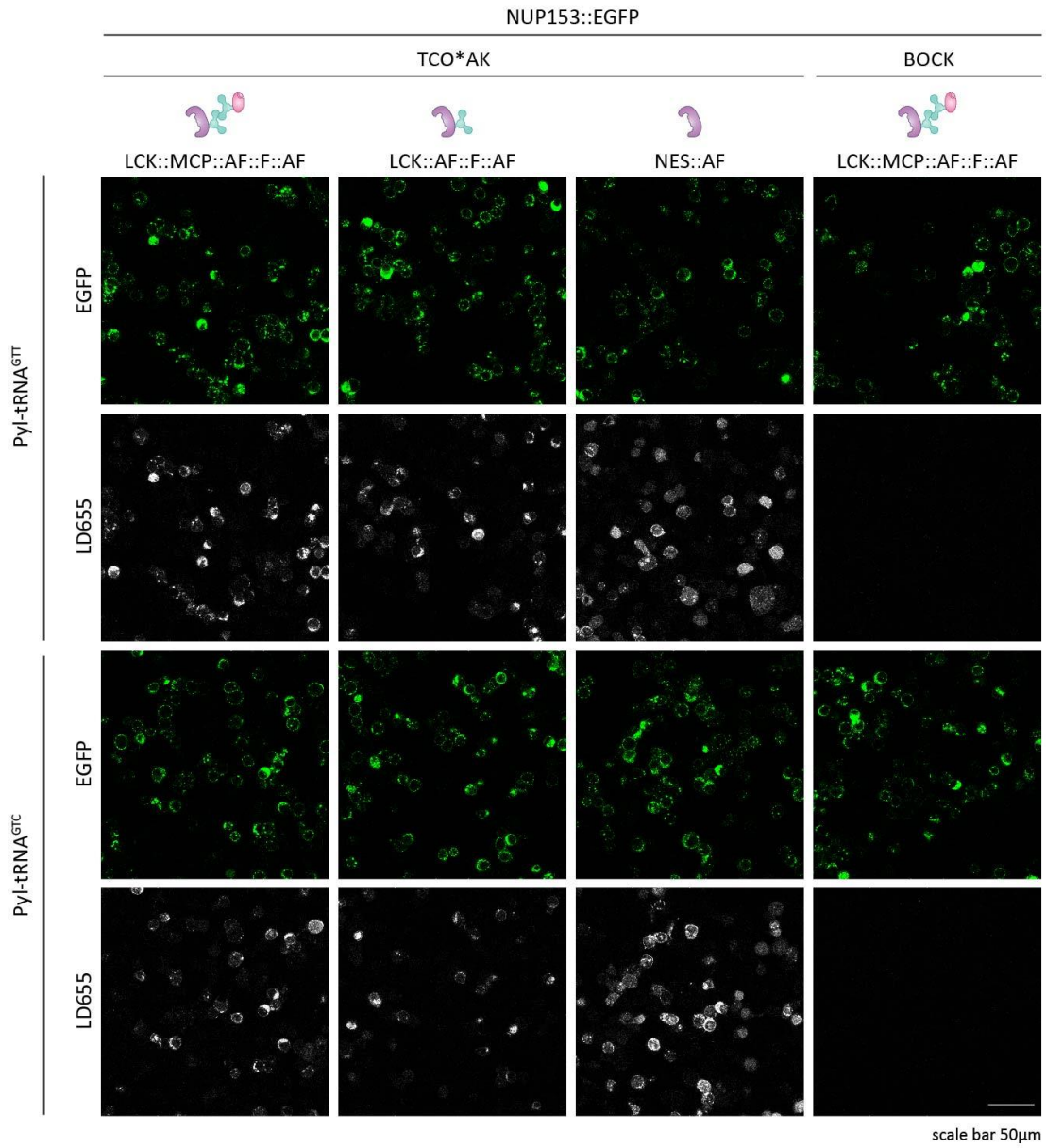
Appendix I Figure 10| Confocal imaging results for codon selection for NUP153::EGFP (codons GGT, GGC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons GGT and GGC individually in NUP153::EGFP.



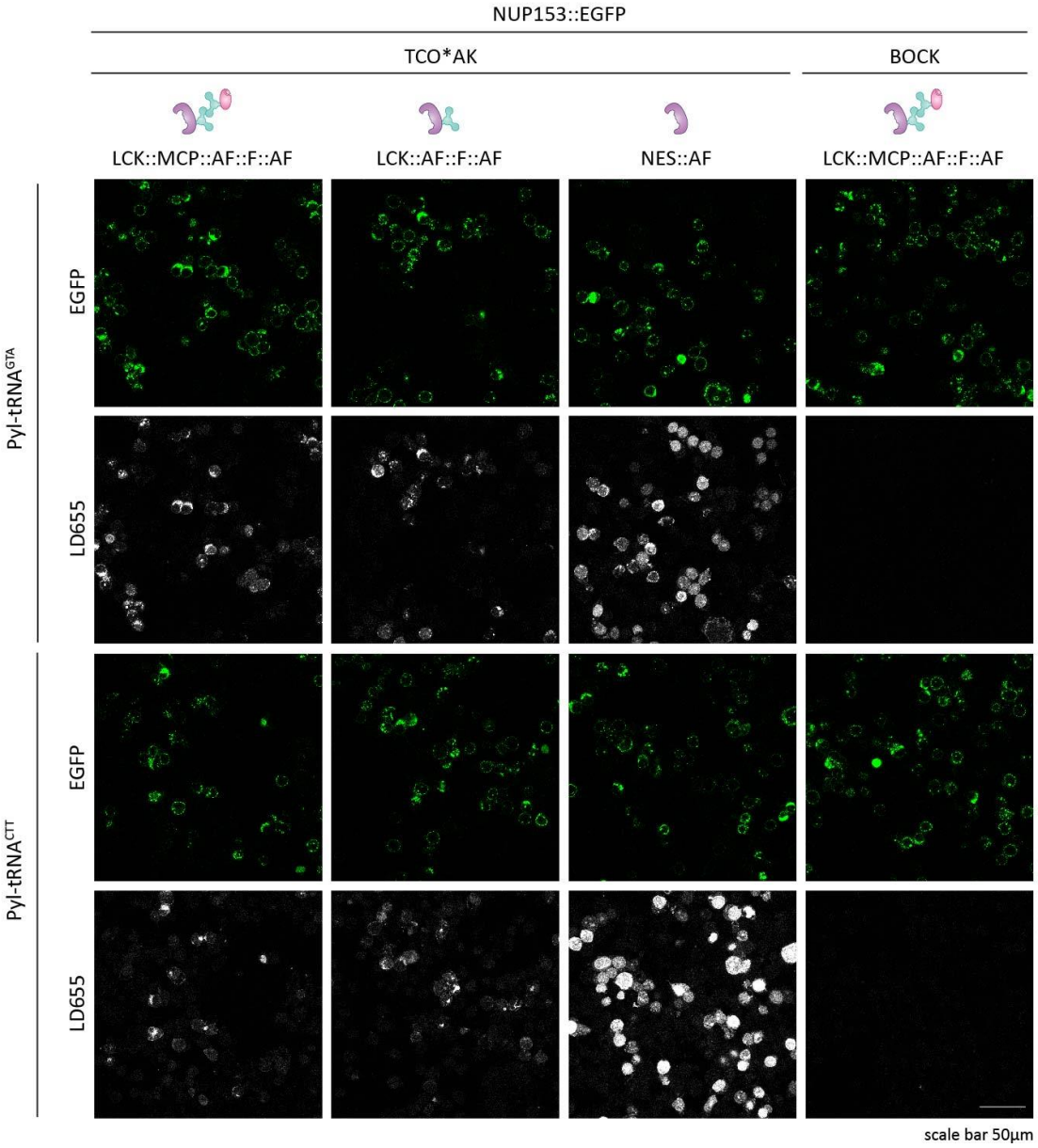
Appendix I Figure 11| Confocal imaging results for codon selection for NUP153::EGFP (codons GGA, GGG). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons GGA and GGC individually in NUP153::EGFP.



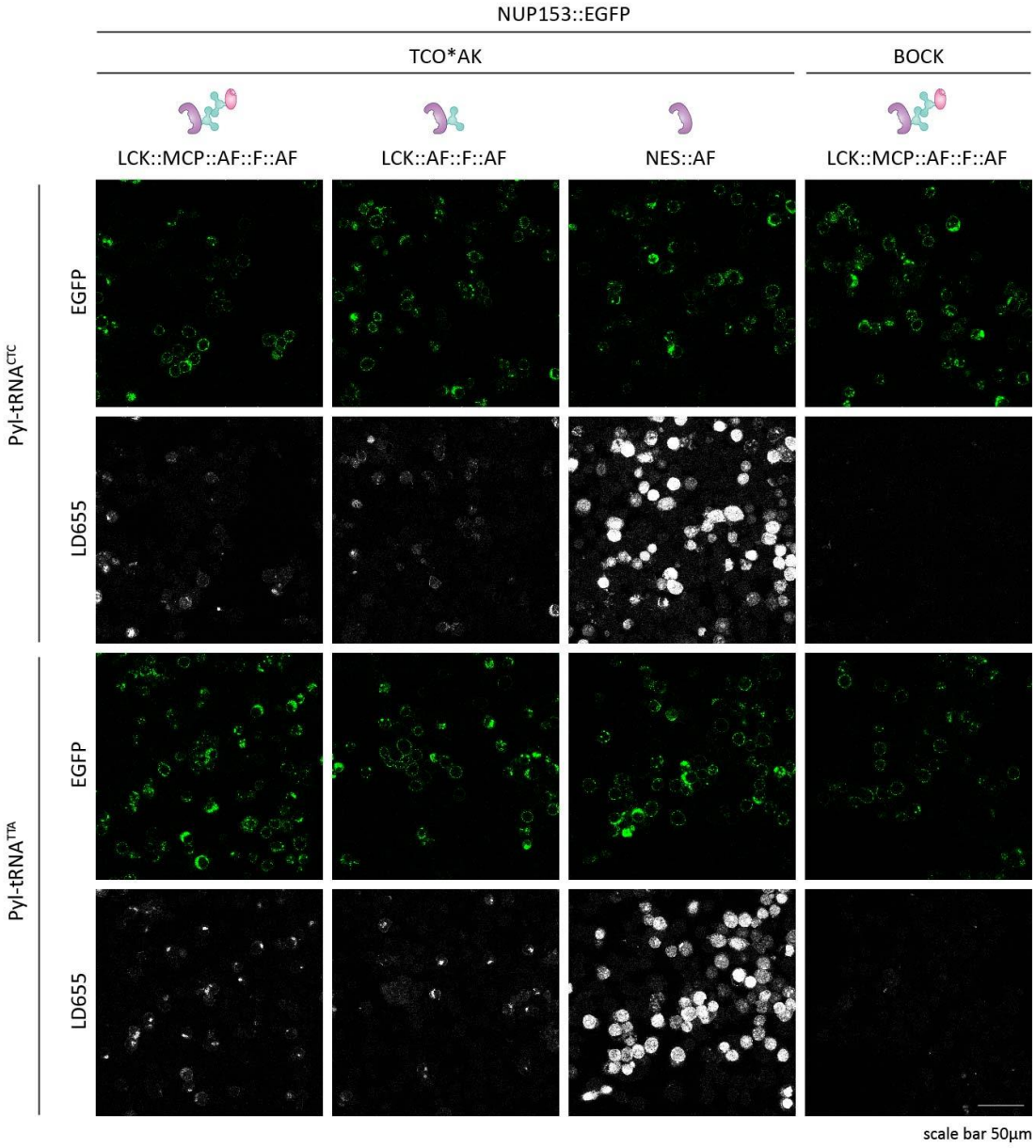
Appendix I Figure 12| Confocal imaging results for codon selection for NUP153::EGFP (codons GCT, GCC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons GCT and GCC individually in NUP153::EGFP.



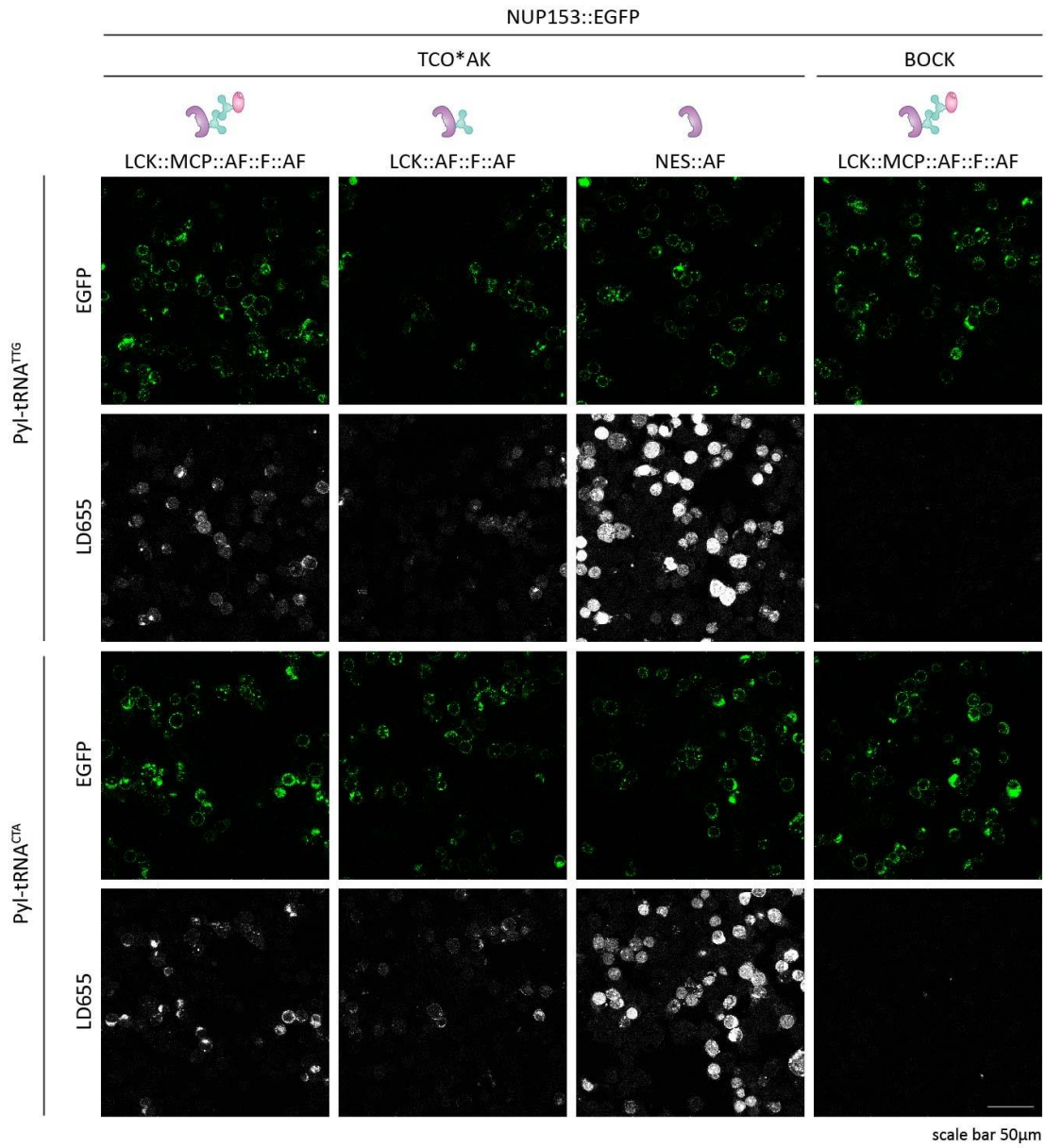
Appendix I Figure 13| Confocal imaging results for codon selection for NUP153::EGFP (codons GTT, GTC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons GTT and GTC individually in NUP153::EGFP.



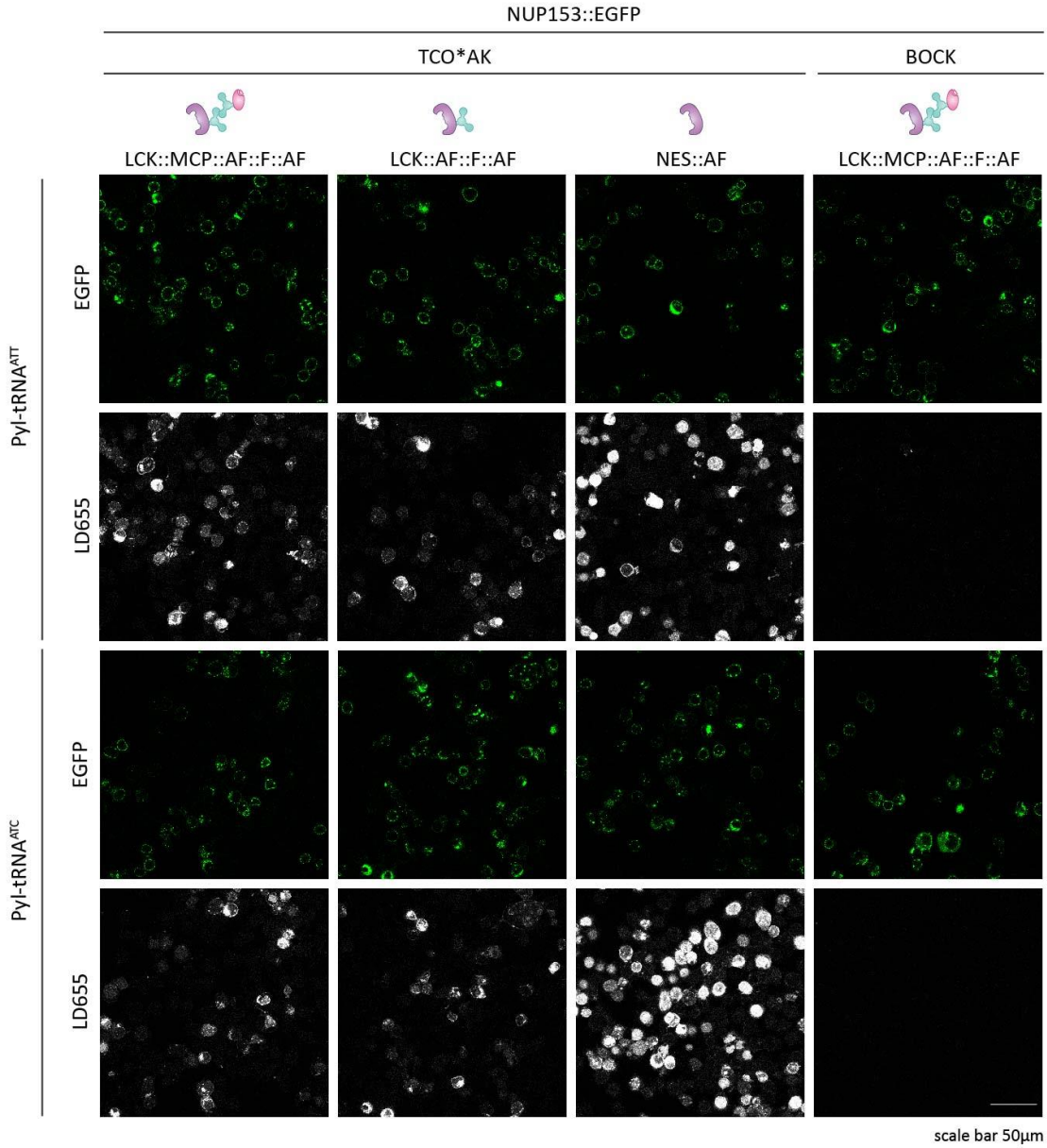
Appendix I Figure 14| Confocal imaging results for codon selection for NUP153::EGFP (codons GTA, CTT). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons GTA and CTT individually in NUP153::EGFP.



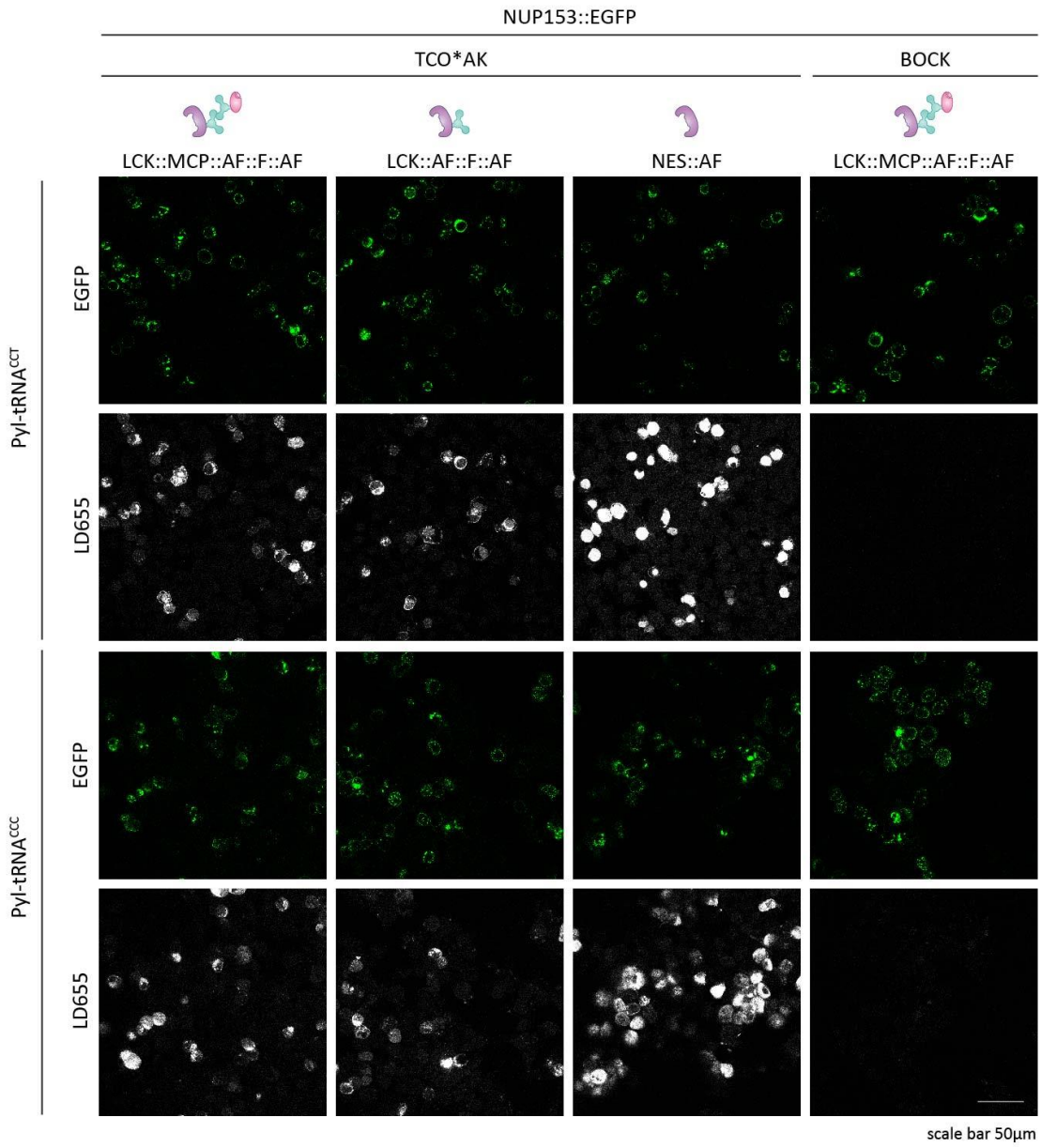
Appendix I Figure 15| Confocal imaging results for codon selection for NUP153::EGFP (codons CTC, TTA). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons CTC and TTA individually in NUP153::EGFP.



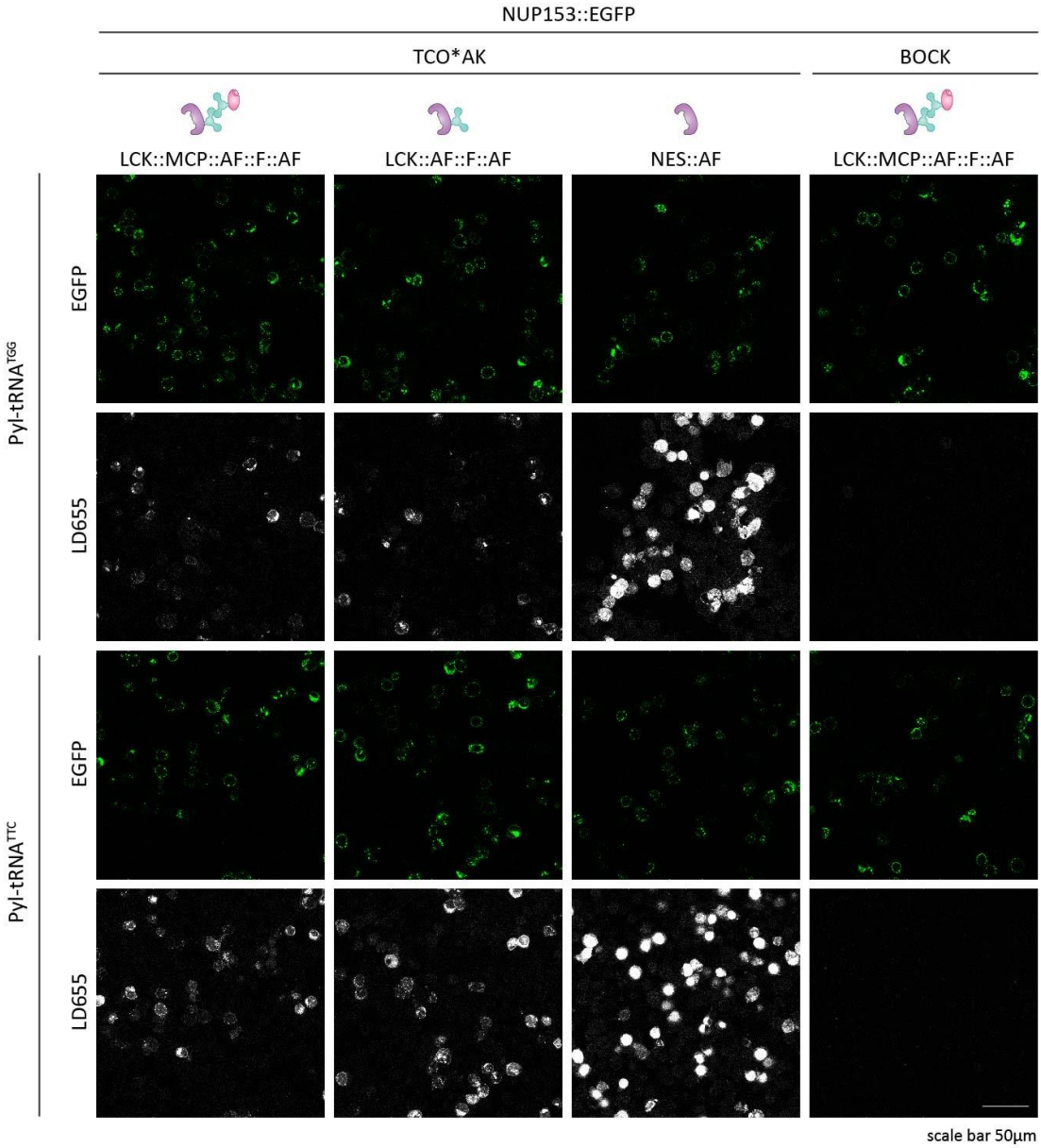
Appendix I Figure 16| Confocal imaging results for codon selection for NUP153::EGFP (codons TTG, CTA). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons TTG and CTA individually in NUP153::EGFP.



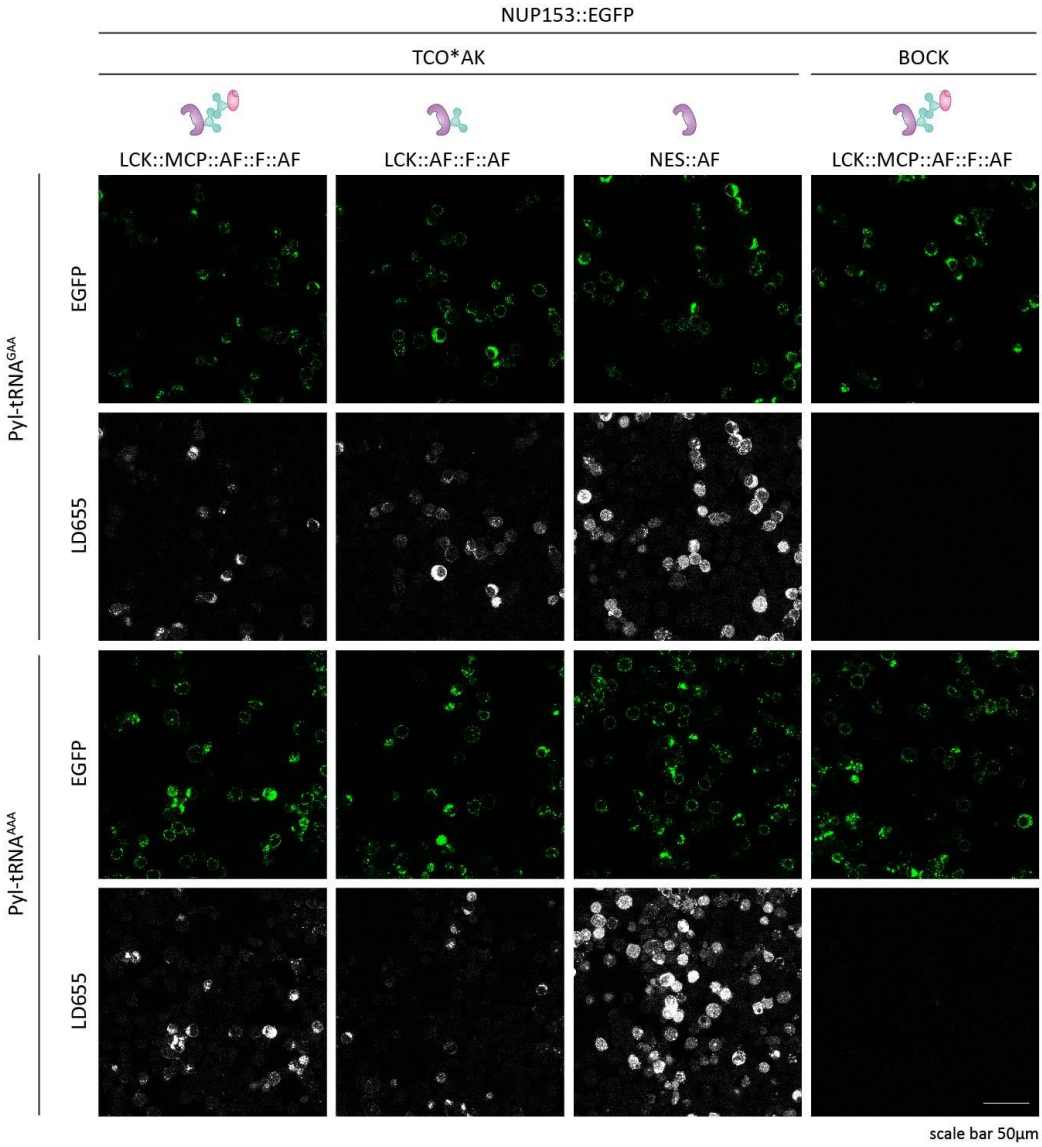
Appendix I Figure 17| Confocal imaging results for codon selection for NUP153::EGFP (codons ATT, ATC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons ATT and ATC individually in NUP153::EGFP.



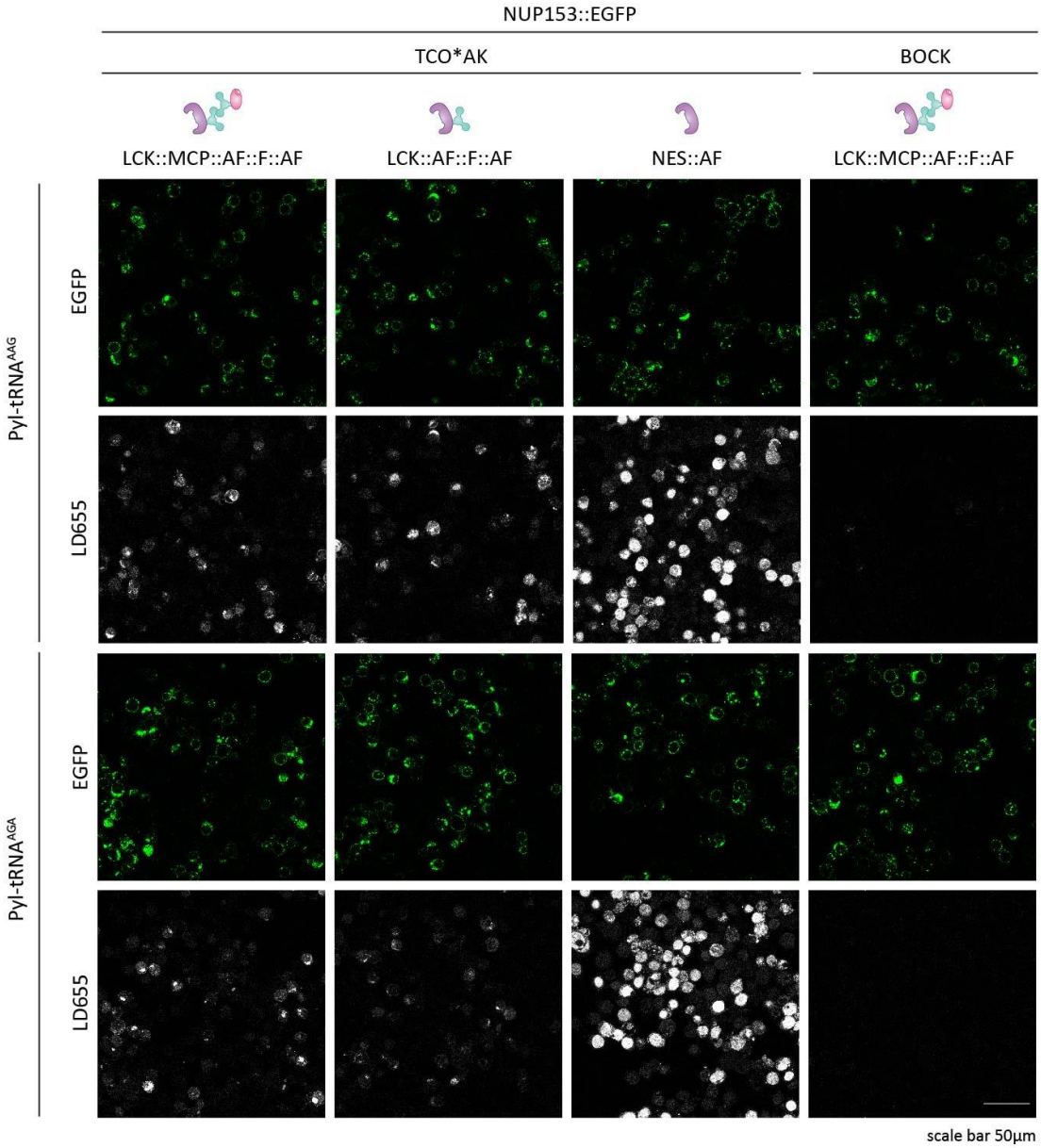
Appendix I Figure 18| Confocal imaging results for codon selection for NUP153::EGFP (codons CCT, CCC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons CCT and CCC individually in NUP153::EGFP.



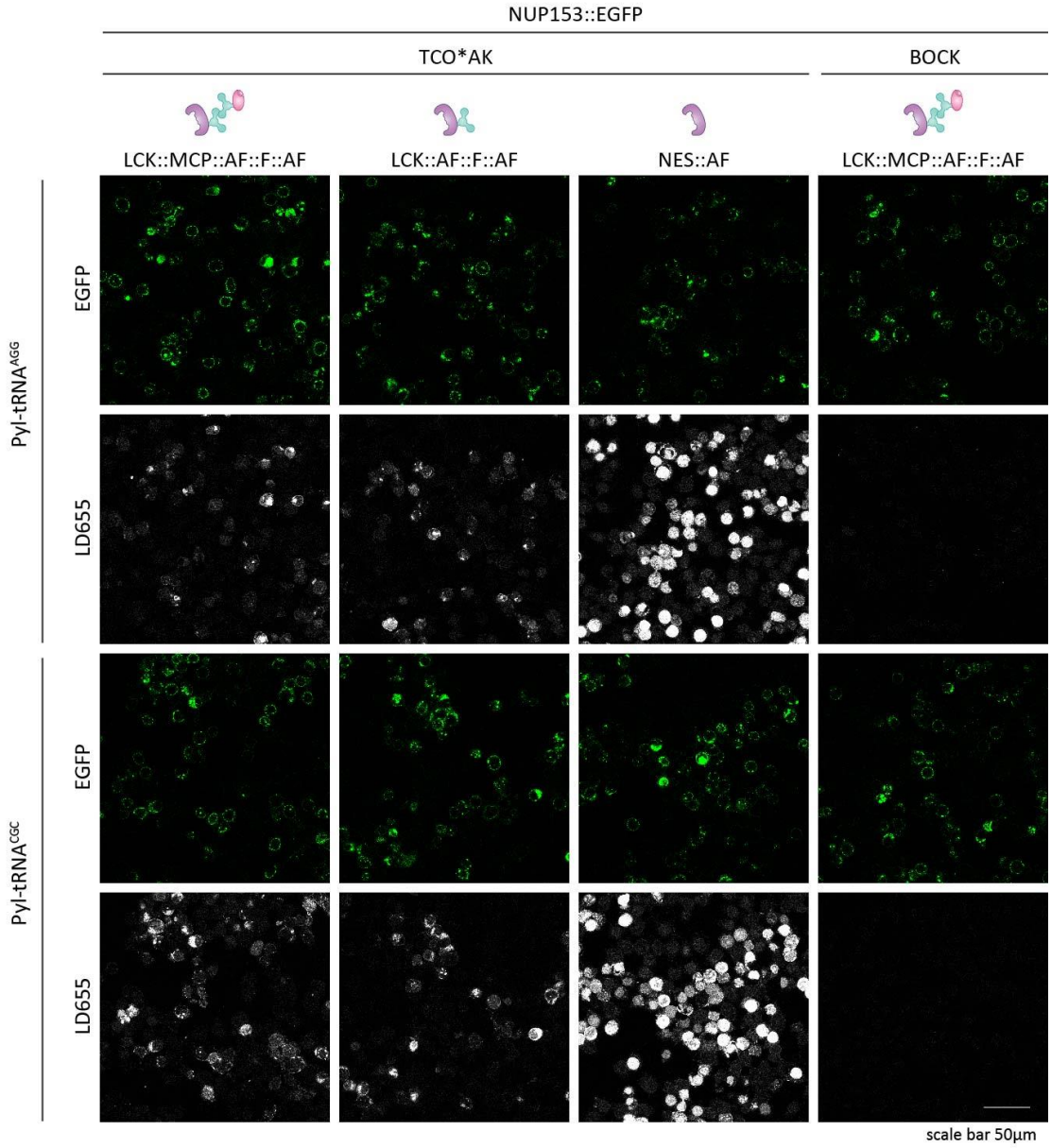
Appendix I Figure 19| Confocal imaging results for codon selection for NUP153::EGFP (codons TGG, TTC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons TGG and TTC individually in NUP153::EGFP.



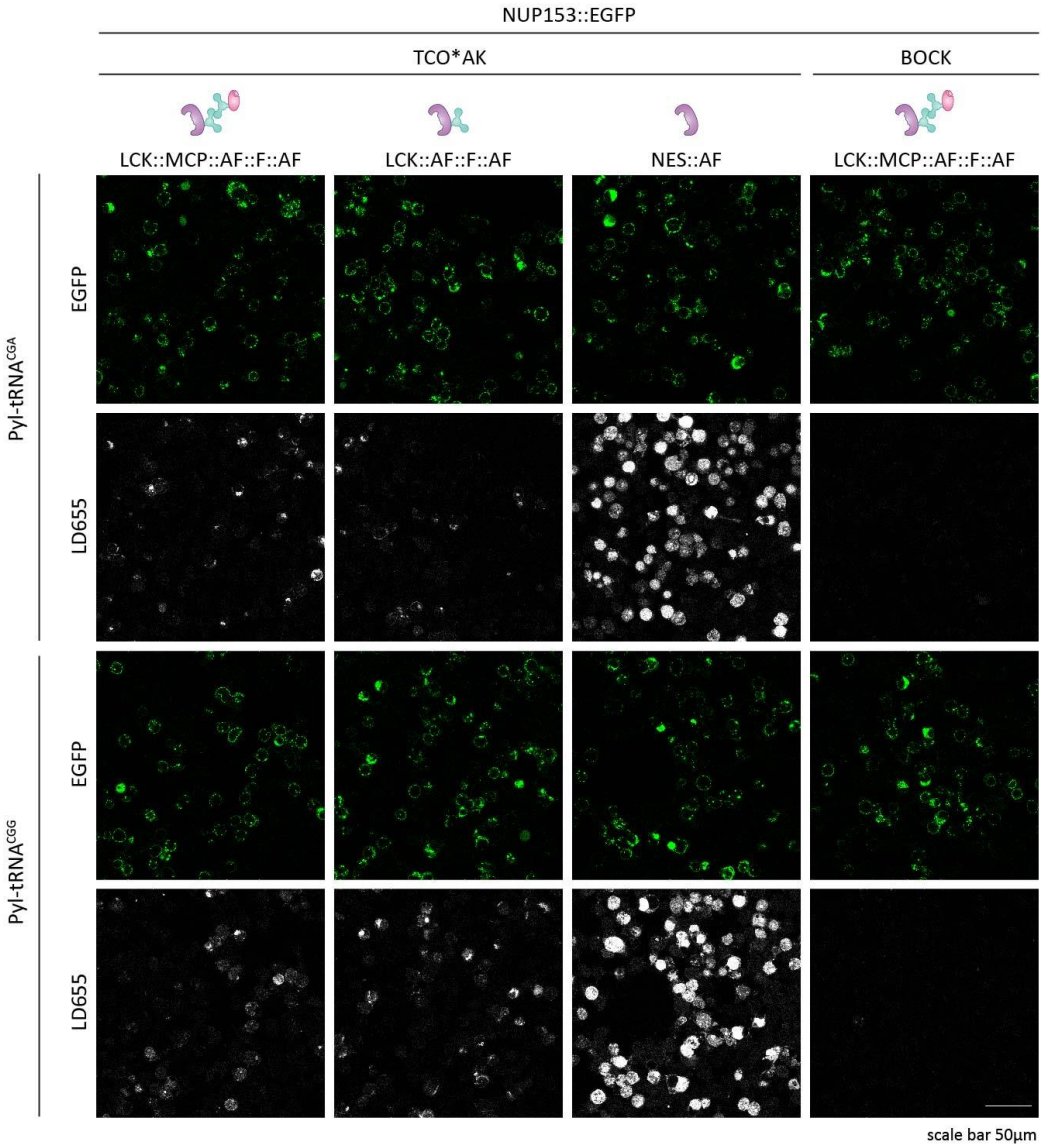
Appendix I Figure 20| Confocal imaging results for codon selection for NUP153::EGFP (codons GAA, AAA). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons GAA and AAA individually in NUP153::EGFP.



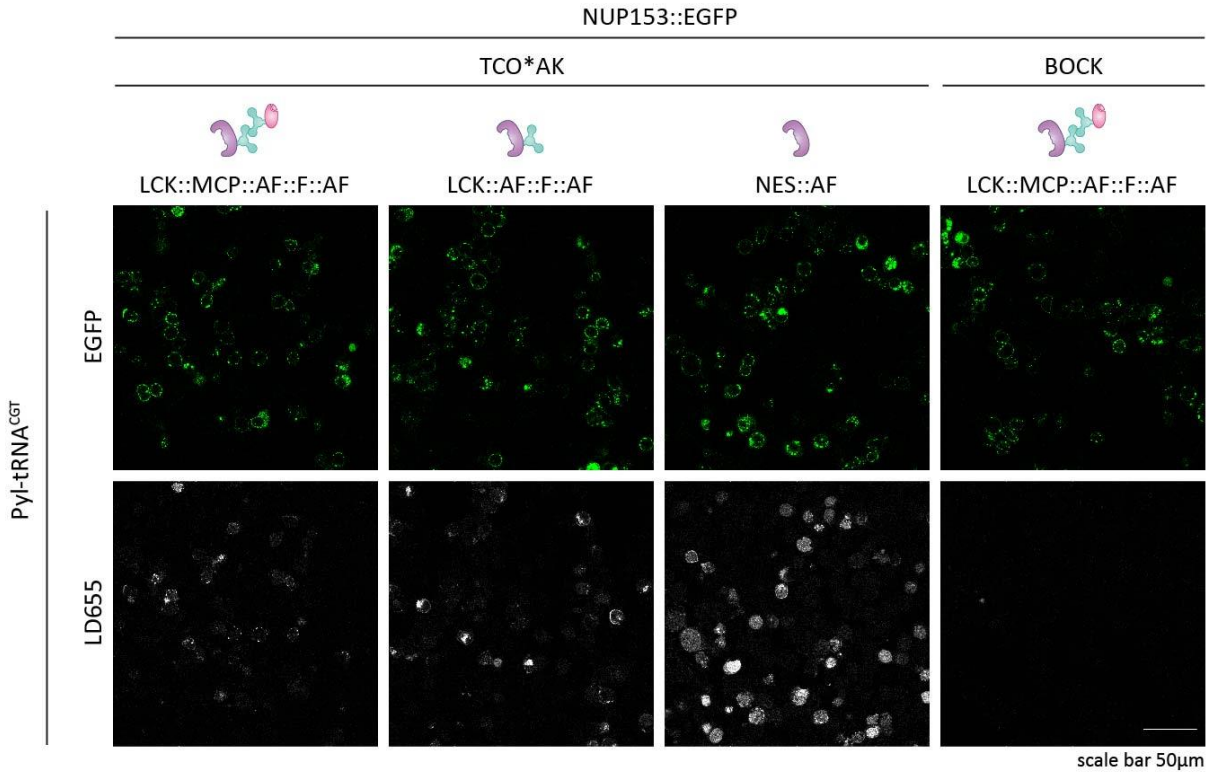
Appendix I Figure 21| Confocal imaging results for codon selection for NUP153::EGFP (codons AAG, AGA). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons AAG and AGA individually in NUP153::EGFP.



Appendix I Figure 22| Confocal imaging results for codon selection for NUP153::EGFP (codons AGG, CGC). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons AGG and CGC individually in NUP153::EGFP.



Appendix I Figure 23| Confocal imaging results for codon selection for NUP153::EGFP (codons CGA, CGG). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codons CGA and CGG individually in NUP153::EGFP.



Appendix I Figure 24| Confocal imaging results for codon selection for NUP153::EGFP (codon CGT). The confocal imaging experiment was performed as described in section 3.2.1.The OTO LCK::MCP::AF::F::AF was used to reassign sense codons to incorporate ncAAs in NUP153::EGFP in HEK293T cells (first and last columns from the left). Sense codon reassignment was also performed with an OTO lacking the mRNA recruiting domain MCP, namely LCK::AF::F::AF (second column from the left). The purpose of using LCK::AF::F::AF was to demonstrate the need of MCP-MS2 interaction for selective sense codon reassignment by facilitating preferential recruitment of MS2 tagged mRNA of interest into the OTO. Sense codon was reassigned in NUP153::EGFP with non-selective cytoplasmic NES::AF (third column from left) to show unspecific GCE as compared to samples with the fully functional OTO LCK::MCP::AF::F::AF. TCO*AK was incorporated in NUP153::EGFP (columns 1 to 3 from left) and subsequently detected by labelling the cells with LD655-H-Tetrazine that reacts with TCO*AK by IEDDA reaction. BOCK was incorporated in NUP153::EGFP as a negative control (last column from left) since it does not react with LD655-H-Tetrazine. EGFP signal from the fusion construct NUP153::EGFP was used as a standard label. This figure depicts the outcome for reassigning the codon CGT individually in NUP153::EGFP.

Appendix II

Sequences of plasmids generated for this thesis (5' - 3')

The sequence of BoxB loops with linker has been highlighted in blue in the normal EGFP reporter sequence. The sequence of MS2 loops with linker has been highlighted in magenta in the extended EGFP sequences. The codon to be reassigned is highlighted in green. To generate the inactive/dark EGFP, the CCC codon was engineered to the position corresponding to site 66 in order to incorporate proline at site 66.

Reporter: pBI_CMV1_Flag::EGFP^{39CTA}_4xBoxB_CMV2_Flag::EGFP^{39CTA}::EGFP^{66CCC}_2xMS2 (codon optimized to have CTA only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39CTA}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**CTA**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39CTA}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**CTA**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA

GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGGATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39CGT}_4xBoxB_CMV2_Flag::EGFP^{39CGT}::EGFP^{66CCC}_2xMS2 (codon optimized to have CGT only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39CGT}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**CGT**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39CGT}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG

CGAGGGCGAGGGCGACGCCACC**CGT**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39TTA}_4xBoxB_CMV2_Flag::EGFP^{39TTA}::EGFP^{66CCC}_2xMS2 (codon optimized to have TTA only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39TTA}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**TTA**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA

GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39TTA}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**TTA**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39TTA}_4xBoxB_CMV2_Flag::EGFP^{39TTA}::EGFP^{66CCC}_2xMS2 (codon optimized to have TTA only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39ATA}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG

CGAGGGCGAGGGCGACGCCACC**ATA**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39ATA}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**ATA**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39CCG}_4xBoxB_CMV2_Flag::EGFP^{39CCG}::EGFP^{66CCC}_2xMS2 (codon optimized to have CCG only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39CCG}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**CCG**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39CCG}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**CCG**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG

ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39TCG}_4xBoxB_CMV2_Flag::EGFP^{39TCG}::EGFP^{66CCC}_2xMS2
(codon optimized to have TCG only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39TCG}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**TCG**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39TCG}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**TCG**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA

GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39ACG}_4xBoxB_CMV2_Flag::EGFP^{39ACG}::EGFP^{66CCC}_2xMS2 (codon optimized to have TCG only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39ACG}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**ACG**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39ACG}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG

CGAGGGCGAGGGCGACGCCACC**ACG**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39TGC}_4xBoxB_CMV2_Flag::EGFP^{39TGC}::EGFP^{66CCC}_2xMS2 (codon optimized to have TGC only at the predetermined site)

Vector: pBI

Sequence: Flag::EGFP^{39TGC}_4xBoxB

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**TGC**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA

GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGCGGCCGCA
TCGATTATAAGCTTTGTACAGCCCTGAAAAAGGGCTGGAGCCCTGAAAAAGGGCATTTGCCC
TGAAAAAGGGCGTCCACGCCCTGAAAAAGGGC

Sequence: Flag::EGFP^{39TGC}::EGFP^{66CCC}_2xMS2

ATGGGCCGCCTGGAGAGCACCCCCCCCAAGAAGAAGCGCAAGGTGGAGGACAGCGCCA
GCGACTACAAGGACGACGACGACAAGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTG
GTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG
CGAGGGCGAGGGCGACGCCACC**TGC**GGCAAGCTGACCCTGAAGTTCATCTGTACCACCG
GCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCTACGGCGTGCAGTGT
TTCAGCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAG
GGCTACGTGCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC
CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT
CAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGT
GTACATCATGGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAA
CATCGAGGACGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGC
GACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAA
GGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCA
TCACCCTGGGCATGGACGAGCTGTACAAGTCCGGAGGAGCACCAGGAAGTGCTGGTTCTG
CTGCTGGTAGTGGAGGTACCGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCA
TCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGGCGAGGGC
GAGGGCGACGCCACCTATGGCAAGCTGACCCTGAAGTTCATCTGTACCACCGGCAAGCTG
CCCGTGCCCTGGCCCACCCTGGTGACCACCCTGACCCCCGGCGTGCAGTGTTTCAGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA
AGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGG
ACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCAT
GGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCC
GTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGAGCGCCCTGAGCAAGGACCCCAA
CGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGG
GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTAAGGCCCAGATCTCCTGAGGC
TGCAGCCTACTAGTCCTAGAAAACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAA
CATGAGGATCACCCATGT

Reporter: pBI_CMV1_Flag::EGFP^{39XXX}::EGFP^{66CCC}_4xBoxB_CMV2_Flag::EGFP^{39XXX}_2MS2
(XXX refers to the codon being reassigned)

Vector: pBI

The sequences of EGFP^{39XXX}::EGFP^{66CCC} and EGFP^{39XXX} are same as that mentioned for the pBI_CMV1_Flag::EGFP^{39TGC}_4xBoxB_CMV2_Flag::EGFP^{39TGC}::EGFP^{66CCC}_2xMS2 reporters.

Reporter: pBI_CMV1_myc::vimentin::mcerulean_2xMS2

Vector: pBI (source- Clontech)

Sequence:
GCCACCATGGAGCAGAAGCTGATCTCAGAGGAGGACCTGATGTCCACCAGGTCCGTGTCCT
CGTCCTCCTACCGCAGGATGTTCGGCGGCCCGGGCACCGCGAGCCGGCCGAGCTCCAG
CCGGAGCTACGTGACTACGTCCACCCGCACCTACAGCCTGGGCAGCGCGCTGCGCCCC
AGCACCAGCCGCAGCCTCTACGCCTCGTCCCCGGGCGGCGTGTATGCCACGCGCTCCTC
TGCCGTGCGCCTGCGGAGCAGCGTGCCCGGGGTGCGGCTCCTGCAGGACTCGGTGGAC
TTCTCGCTGGCCGACGCCATCAACACCGAGTTCAAGAACACCCGCACCAACGAGAAGGTG
GAGCTGCAGGAGCTGAATGACCGCTTCGCCAACTACATCGACAAGGTGCGCTTCCTGGAG
CAGCAGAATAAGATCCTGCTGGCCGAGCTCGAGCAGCTCAAGGGCCAAGGCAAGTCGCG
CCTGGGGGACCTCTACGAGGAGGAGATGCGGGAGCTGCGCCGGCAGGTGGACCAGCTAA
CCAACGACAAAGCCCGCGTCGAGGTGGAGCGCGACAACCTGGCCGAGGACATCATGCG
CCTCCGGGAGAAATTGCAGGAGGAGATGCTTCAGAGAGAGGAAGCCGAAAACACCCTGCA
ATCTTTCAGACAGGATGTTGACAATGCGTCTCTGGCACGTCTTGACCTTGAACGCAAAGTGGA
ATCTTTGCAAGAAGAGATTGCCTTTTTGAAGAAACTCCACGAAGAGGAAATCCAGGAGCTGC
AGGCTCAGATTCAGGAACAGCATGTCCAAATCGATGTGGATGTTTCCAAGCCTGACCTCACG
GCTGCCCTGCGTGACGTACGTCAGCAATATGAAAGTGTGGCTGCCAAGAACCTGCAGGAGG
CAGAAGAATGGTACAAATCCAAGTTTGCTGACCTCTCTGAGGCTGCCAACCGGAACAATGAC
GCCCTGCGCCAGGCAAAGCAGGAGTCCACTGAGTACCGGAGACAGGTGCAGTCCCTCAC
CTGTGAAGTGGATGCCCTTAAAGGAACCAATGAGTCCCTGGAACGCCAGATGCGTGAAATG
GAAGAGAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGCAGGATGA
GATTCAGAATATGAAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGT
TAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGCTGCTGGAAGGCGAGGAGAGC
AGGATTTCTCTGCCTCTTCCAAACTTTTCCTCCCTGAACCTGAGGGAAACTAATCTGGATTCAC
TCCCTCTGGTTGATACCCACTCAAAAAGGACACTTCTGATTAAGACGGTTGAAACTAGAGATG
GACAGGTTATCAACGAAACTTCTCAGCATCACGATGACCTTGAAGGGGATCCACCGGTCGC
CACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCT
GGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGG
CCCACCCTCGTGACCACCCTGAGCTGGGGCGTGCAGTGCTTCGCCCGCTACCCCGACCA
CATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCAC
CATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGA
CACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCT
GGGGCACAAGCTGGAGTACAACGCCATCCACGGCAACGTCTATATCACCGCCGACAAGCA
GAAGAACGGCATCAAGGCCAACTTCGGCCTCAACTGCAACATCGAGGACGGCAGCGTGC
AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCG
ATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGC
TGTACAAGTAAGCGGCCGCATCGATTATAAGCTTTGTACACCTAGAAAACATGAGGATCACCC
ATGTCTGCACCTCGACACTAGAAAACATGAGGATCACCCATGT

Reporter: pBI_CMV1_myc::vimentin^{116TAG}::mcerulean_2xMS2

Vector: pBI

GCCACCATGGAGCAGAAGCTGATCTCAGAGGAGGACCTGATGTCCACCAGGTCCGTGTCCT
CGTCCTCCTACCGCAGGATGTTCGGCGGCCCGGGCACCGCGAGCCGGCCGAGCTCCAG

CCGGAGCTACGTGACTACGTCCACCCGCACCTACAGCCTGGGCAGCGCGCTGCGCCCC
AGCACCAGCCGCAGCCTCTACGCCTCGTCCCCGGGCGGCGTGTATGCCACGCGCTCCTC
TGCCGTGCGCCTGCGGAGCAGCGTGCCCGGGGTGCGGCTCCTGCAGGACTCGGTGGAC
TTCTCGCTGGCCGACGCCATCAACACCGAGTTCAAGAACACCCGCACCAACGAGAAGGTG
GAGCTGCAGGAGCTGAATGACCGCTTCGCCTAGTACATCGACAAGGTGCGCTTCCTGGAGC
AGCAGAATAAGATCCTGCTGGCCGAGCTCGAGCAGCTCAAGGGCCAAGGCAAGTCGCGC
CTGGGGGACCTCTACGAGGAGGAGATGCGGGAGCTGCGCCGGCAGGTGGACCAGCTAAC
CAACGACAAAGCCCGCGTCGAGGTGGAGCGCGACAACCTGGCCGAGGACATCATGCGC
CTCCGGGAGAAATTGCAGGAGGAGATGCTTCAGAGAGAGGAAGCCGAAAACACCCTGCAA
TCTTTCAGACAGGATGTTGACAATGCGTCTCTGGCACGTCTTGACCTTGAACGCAAAGTGGAA
TCTTTGCAAGAAGAGATTGCCTTTTTGAAGAAACTCCACGAAGAGGAAATCCAGGAGCTGCA
GGCTCAGATTCAGGAACAGCATGTCCAAATCGATGTGGATGTTTCCAAGCCTGACCTCACGG
CTGCCCTGCGTGACGTACGTCAGCAATATGAAAGTGTGGCTGCCAAGAACCTGCAGGAGGC
AGAAGAATGGTACAAATCCAAGTTTGCTGACCTCTCTGAGGCTGCCAACCGGAACAATGACG
CCCTGCGCCAGGCAAAGCAGGAGTCCACTGAGTACCGGAGACAGGTGCAGTCCCTCACC
TGTGAAGTGGATGCCCTTAAAGGAACCAATGAGTCCCTGGAACGCCAGATGCGTGAAATGGA
AGAGAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGCAGGATGAGA
TTCAGAATATGAAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCTGCTCAATGTTA
AGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGCTGCTGGAAGGCGAGGAGAGCA
GGATTTCTCTGCCTCTTCCAAACTTTTCCTCCCTGAACCTGAGGGAAACTAATCTGGATTCACT
CCCTCTGGTTGATACCCACTCAAAAAGGACACTTCTGATTAAGACGGTTGAAACTAGAGATGG
ACAGGTTATCAACGAAACTTCTCAGCATCACGATGACCTTGAAGGGGATCCACCGGTCGCC
ACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTG
GACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCA
CCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGC
CCACCCTCGTGACCACCCTGAGCTGGGGCGTGCAGTGCTTCGCCCGCTACCCCGACCAC
ATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGAC
ACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTG
GGGCACAAGCTGGAGTACAACGCCATCCACGGCAACGTCTATATCACCGCCGACAAGCAG
AAGAACGGCATCAAGGCCAACTTCGGCCTCAACTGCAACATCGAGGACGGCAGCGTGCAG
CTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
ACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTA
CAAGTAAGCGGCCGCATCGATTATAAGCTTTGTACACCTAGAAAACATGAGGATCACCCATGT
CTGCACCTCGACACTAGAAAACATGAGGATCACCCATGT

Reporter: pEGFP_NUP153_2xMS2 (MS2 loops were added by me to the existing pEGFP_NUP153 plasmid from the Lemke lab repository)

Vector: pEGFP

Sequence:

ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC
GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTA

CGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCA
CCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGA
AGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCT
TCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACC
CTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGG
CACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGA
ACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCG
CCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAA
CCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACA
TGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACA
AGTCCGGCCGGACTCAGATCTCGAGAGCGTCTGGTGCTGGCGGTGTTGGTGGAGGAGGTG
GGGGTAAAATTCGTACTCGTCGCTGTCATCAAGGTCCGATTAAACCGTATCAGCAGGGACGT
CAGCAACATCAGGGTATTCTGAGCCGTGTGACCGAAAGCGTGAAAAACATTGTGCCGGGTT
GGCTGCAACGTTATTTCAACAAAAATGAGGATGTGTGTTCGTGTTCTACCGATACCAGTGAAGT
TCCTCGTTGGCCGGAAAACAAAGAAGATCACCTGGTGTATGCCGATGAAGAATCGAGCAATA
TCACCGATGGCCGTATTACTCCTGAACCGGCGGTTAGTAACACTGAAGAACCGTCAACCACA
AGCACAGCATCGAACTATCCAGATGTCCTGACTCGCCCTTCTCTGCACCGTTCTCACCTGAA
CTTTAGCATGCTGGAATCACCAGCTCTGCATTGTCAGCCGTCTACCAGTAGTGCCTTCCCGAT
TGGCTCTAGTGGCTTTTCGCTGGTCAAAGAGATCAAAGACTCGACCTCTCAACATGACGATGA
TAACATTAGCACGACCTCGGGTTTTAGTAGCCGTGCCTCCGATAAAGACATTACCGTGAGCAA
AAACACCTCTCTGCCGCCTCTGTGGAGTCCTGAAGCCGAACGCTCTCATAGTCTGTCTCAGC
ACACAGCCACCAGTTCCAAAAAACCAGCCTTCAACCTGAGCGCCTTTGGTACACTGTCACC
GAGCCTGGGAAATTCCTCTATCCTGAAAACATCACAGCTGGGCGATAGTCCGTTTTATCCGG
GCAAAACGACGTATGGTGGTGCCGCTGCTGCTGTTCGCCAGTCTAAACTGCGTAACACTCC
GTATCAAGCTCCAGTCCGTCGCCAAATGAAAGCAAAACAACTGTCGGCCCAGTCTTATGGTG
TGACAAGCTCTACAGCTCGTCGTATCCTGCAAAGTCTGGAGAAAATGTCATCTCCGCTGGCA
GATGCCAAACGTATTCCGTCCATTGTGAGCAGTCCGCTGAATAGCCCGCTGGACCGTAGTG
GGATCGATATCACCGACTTCCAAGCCAAACGTGAGAAAGTGGATAGCCAGTATCCGCCTGTA
CAACGTCTGATGACCCCGAAACCGGTTTCAATTGCCACGAATCGTAGCGTGTATTTCAAACC
GTCACTGACCCCTAGTGGTGAGTTTCGTAAAACAAATCAGCGTATCGACAACAAATGCTCTAC
CGGGTATGAAAAAAACATGACGCCGGGACAGAATCGTGAACAACGTGAATCTGGCTTCTCTT
ATCCGAACTTTAGTCTGCCGGCAGCAAATGGTCTGAGTAGCGGTGTAGGAGGTGGTGGGGG
CAAAATGCGCCGTGAACGTCACGCCTTTGTGGCCTCTAAACCTCTGGAAGAAGAAGAGATG
GAGGTTCCTGTACTGCCGAAAATCAGTCTGCCTATCACCTCTTCAAGTCTGCCGACCTTCAAC
TTTTCTAGTCCGGAAATCACAACCTCTAGCCCGTCACCGATTAATAGCAGTCAAGCACTGACG
AATAAAGTCCAAATGACCTCACCGAGTTCTACGGGTTCTCCGATGTTCAAATTCTCTAGTCCTA
TCGTGAAATCAACCGAAGCGAACGTCCTGCCTCCTTCTAGTATTGGGTTCACCTTTAGCGTCC
CAGTGGCCAAAACAGCTGAACTGAGCGGTAGCAGTAGTACTCTGGAACCGATTATCAGCTCA
AGCGCCCATCATGTCACTACCGTGAATAGCACAAACTGTAAAAAAACGCCGCCTGAGGACT
GTGAAGGACCGTTTCGTCCTGCCGAAATCCTGAAAGAAGGTTCCGTCCTGGACATTCTGAAA
TCTCCGGGATTTGCCTCTCCTAAAATCGACTCTGTTGCCGCTCAACCAACTGCCACATCACC
GGTGGTTTATACTCGTCCGGCGATTAGCAGTTTTAGCAGTAGTGGCATCGGTTTTGGTGAATCC
CTGAAAGCTGGCTCATCTTGGCAGTGTGACACCTGCCTGCTGCAAAACAAAGTGACCGATAA
CAAATGTATTGCCTGTCAGGCCGCCAAACTGTCTCCTCGTGATACAGCCAAACAGACCGGC

ATCGAAACCCCTAATAAAAGCGGGAAAACGACCCTGTCAGCAAGTGGTACGGGATTTGGGG
ACAAATTCAAACCTGTGATCGGCACATGGGACTGTGACACTTGTCTGGTACAGAACAAACCA
GAAGCGATCAAATGTGTGGCCTGTGAAACGCCTAAACCTGGAACATGTGTGAAACGTGCCCT
GACTCTGACTGTTGTGTCAGAAAGCGCCGAAACCATGACGGCAAGCAGCTCATCCTGTACTG
TGACTACCGGGACTCTGGGATTTGGTGACAAATTCAAACGCCCGATTGGTTCCTGGGAATGC
TCCGTGTGTTGTGTGAGCAATAATGCCGAGGACAACAAATGTGTGTCCTGTATGAGCGAGAAA
CCTGGCAGCTCTGTTCCTGCTAGCAGCTCTAGCACAGTTCCTGTTAGTCTGCCTAGTGGTGGT
TCTCTGGGTCTGGAAAAATTCAAAAAACCTGAAGGAAGCTGGGATTGTGAGCTGTGCCTGGTA
CAGAATAAAGCGGATAGCACGAAATGTCTGGCCTGTGAGTCAGCCAAACCAGGGACTAAAA
GCGGCTTTAAAGGCTTCGACACGTCGAGCAGTTCTAGTAACAGCGCCGCCTCATCATCTTTC
AAATTTGGGGTGAGCAGCTCCTCTAGTGGTCCTAGTCAAACACTGACCTCTACCGGAAACTTC
AAATTCGGCGATCAGGGTGGCTTCAAAATTGGTGTCTCCTCTGATTCGGGTAGCATTAACCCG
ATGAGTGAGGGGTTCAAATTCAGCAAACCAATTGGCGATTTCAAATTCGGTGTGTCGTCTGAAT
CCAAACCTGAAGAAGTCAAAAAAGACAGCAAAAACGACAATTTCAAATTCGGCCTGTCTAGT
GGTCTGTCTAATCCGGTTAGCCTGACCCCGTTTCAGTTCGGGGTGTCTAATCTGGGTCAGGAA
GAGAAAAAAGAGGAGCTGCCTAAAAGTTCATCTGCCGGGTTCAGTTTTGGTACAGGCGTGAT
CAATAGCACTCCAGCACCAGCCAATACAATCGTGACGAGCGAGAACAAATCGAGCTTCAAC
CTGGGGACAATCGAAACGAAAAGCGCCAGTGTAGCGCCATTCACGTGTAAAACCTCCGAG
GCAAAAAAAGAAGAGATGCCGGCCACAAAAGGTGGATTCTCATTCGGCAACGTGGAACCG
GCTAGCCTGCCATCAGCAAGCGTGTTTGTACTGGGCCGTACCGAGGAGAAACAGCAGGAA
CCTGTTACTAGCACCAGTCTGGTCTTTGGTAAAAAAGCCGACAACGAAGAACCGAAATGTCA
GCCAGTGTTCAGCTTCGGCAATAGCGAACAGACGAAAGACGAAAACAGCAGCAAATCGAC
GTTCAGCTTCAGTATGACGAAACCGAGCGAAAAAGAAAGTGAGCAGCCAGCAAAAGCAACG
TTCGCCTTTGGAGCACAGACATCAACCACAGCCGATCAAGGAGCAGCGAAACCAGTTTTCA
GTTTTCTGAATAACAGCTCAAGCAGCAGTTCTACACCAGCAACCTCAGCAGGTGGTGGGATC
TTTGGATCAAGCACCTCATCCAGCAATCCGCCAGTGGCAACATTCGTGTTTGGCCAGAGCA
GTAATCCGGTGTCATCTTCAGCATTTGGGAATACCGCCGAGAGTAGCACATCACAGTCTCTGC
TGTTCTCACAGGACTCTAAACTGGCAACCACCTCTTCTACTGGTACAGCGGTTACCCCGTTTG
TGTTCGGTCCGGGAGCATCATCCAATAATACCACGACGTCGGGCTTTGGGTTTGGTGCCACG
ACAACAAGCAGTAGCGCTGGTAGCAGCTTTGTCTTTGGCACAGGTCCTTCAGCACCTTCTGC
TTCACCAGCTTTCGGAGCCAATCAGACTCCGACATTCGGACAGTCACAGGGTGCCTCTCAA
CCAAATCCTCCGGGTTTTGGCAGTATTAGCAGTAGTACCGCCCTGTTCCCGACCGGTAGTCA
ACCGGCACCGCCAACATTTGGAACGGTTAGCAGTAGTAGTCAGCCTCCGGTGTTTGGACAA
CAACCGAGCCAGAGCGCCTTCGGATCAGGAACGACCCCTAATAGTAGCAGTGCCTTCCAG
TTCGGTAGCAGTACCACCAACTTCAACTTCACGAACAATAGCCCGTCAGGTGTGTTCACGTTT
GGCGCCAATTCTTCTACCCCAGCGGCAAGTGCTCAACCTTCAGGCTCAGGTGGATTTCCTTT
CAACCAGTCACCAGCAGCGTTTACTGTTGGTTCTAACGGGAAAAACGTTTTCAGTAGCAGCG
GCACCTCGTTTTCTGGTCGTAAAATCAAAACGGCCGTTCGTCGCCGTAAATAATCTAGAGGC
GGCCCGGGATCCACATGAGGATCACCCATGTCTGCACCTCGACACTAGAAAACATGAGGAT
CACCCATGT

Pyl-tRNA^{XXX}: pUC57-Kan_U6Full_ aa_XXX (aa – amino acid, XXX- codon to be reassigned)

Vector: pUC57_Kanamycin

Sequence:

CAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAA
GGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACAAAATACGTG
ACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAATTATGTTTTAAAATGGACTATCATA
TGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACA
CCGGAAACCTGATCATGTAGATCGAATGGACT**YYY**AATCCGTTCAGCCGGGTTAGATTCCCG
GGGTTTCCG

YYY – anticodon (52 Pyl-tRNA^{XXX} were generated for 52 codons by varying the anticodon (**YYY**) sequence, please refer to table 8 for the list of the plasmids)

Curriculum Vitae