Dissecting Specificity of the Human Arg/N-degron Pathway

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

am Fachbereich Biologie

der Johannes Gutenberg-Universität Mainz

vorgelegt von

Andrea Coti

geboren am 03.03.1994

in Kikinda, Serbien

Mainz, Juli 2024

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

JG|U

Dekan: Prof. Dr. Eckhard Thines

1. Gutachter: Dr. Anton Khmelinskii

2. Gutachter: Prof. Dr. Helle Ulrich

Tag der mündlichen Prüfung: 12.09.2024

Preface

I hereby declare that I am the sole author and composer of this report and that no other source or learning aids, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of said work. I hereby also declare that this report has not been prepared for another examination, either wholly or excerpts thereof.

Throughout this study, I have received support in experimental design and data analysis from my supervisor [redacted] contributed significantly to the data analysis presented in this thesis. The experiments involving *Saccharomyces cerevisiae* were conducted in collaboration with [redacted], while those involving ProP-PD technology were carried out in partnership with the [redacted] An experiment with Mass Spectrometry was done in collaboration with [redacted]

Place, Date _____ Signatur _____

Table of Contents

Summary

The Arg/N-degron pathway was the first discovered N-degron pathway in 1986 in Varshavsky's lab. This is an important pathway for targeting proteins for degradation based on their N-terminal residues. In this pathway, specific residues are recognized by UBR family E3 ubiquitin ligases and marked for proteasomal degradation. This pathway plays a significant role in regulating protein stability and function, impacting various cellular processes and diseases, including cancer, neurodegeneration, and cardiovascular development.

It is important to focus on the N-terminal part of proteins to expand our understanding of this pathway. For that, I tested the stability of the library containing all variants of 20 amino acids at the first two positions (Ubi-XZ library) by Global Protein Stability (GPS) profiling. The eK-sfGFP-P2A-mCherry reporter, which showed the highest precision, was chosen for this purpose. I showed that the Ubi-XZ library is highly reproducible, as evidenced by high consistency across replicates. Besides that, the stability of fourteen constructs containing specific dipeptides compared with the same dipeptides in the Ubi-XZ library revealed a strong correlation. The key result of the Ubi-XZ library is the unexpected stability of N-terminal E and Q residues, which are typically targeted for degradation by the Arg/N-degron pathway. To explore further the reasoning behind this observation I explored various hypotheses.

Differential expression of key enzymes in the Arg/N-degron pathway might impair its functionality. However, transcription levels of ATE1, NTAN1, NTAQ1, UBR1, UBR2, UBR4, and UBR5, which are part of this pathway, compared with other genes in the HEK293 cell line indicated that these genes are expressed, suggesting that the deviation is not due to a lack of these proteins. Another hypothesis involved substrate specificity of enzymes, particularly NTAQ1, which targets N-terminal Q for deamidation. Experiments with human NTAN1 and NTAQ1 expressed in yeast cells lacking the Nta1 enzyme confirmed that NTAN1 and NTAQ1 are specific to their substrates.

Two studies showed that acetylation of N-terminal residues prevents degradation by the Arg/N-degron pathway. To investigate this in the context of my results, 12 constructs with dipeptides beginning with N, Q, E, or D and a second residue of L, R, or D were expressed in the ATE1 knockout cell line. These dipeptides underwent acetylation, matching the stability pattern in the Ubi-XZ library and preventing the Arg/N-degron pathway from targeting N-terminal Q and N residues. Future research should identify which NAT enzymes acetylate N-terminal E and D, with NatH as a potential candidate.

Proteomic Peptide Phage Display (ProP-PD) identified sequences on the N-terminus of human proteins binding to NTAN1, NTAQ1, or ATE1. An N-terminal ProP-PD library with all human protein isoforms was used, finding 14 binders for NTAQ1 and 7 for ATE1 isoform 1. Interestingly, binders for NTAQ1 did not start with Q, and those for ATE1 isoform 1 did not predominantly start with E or D. Further testing the stability of these peptides in cells lacking NTAQ1 or ATE1 could reveal new degrons.

This study highlights acetylation as a mechanism preventing degradation by the Arg/N-degron pathway.

Zusammenfassung

Der Arg/N-Degron-Weg war der zuerst entdeckte N-Degron-Weg, im Varshavskys Labor im Jahre 1986. Dies ist ein wichtiger Weg, um Proteine basierend, auf ihren N-terminalen Aminosäurenresten, gezielt für den Proteinabbau auszuwählen. In diesem Weg werden bestimmte Reste von E3-Ubiquitinligasen der UBR-Familien erkannt und für den proteasomalen Abbau markiert. Dieser Weg spielt eine wichtige Rolle bei der Regulierung der Proteinstabilität und -funktion und beeinflusst verschiedene zelluläre Prozesse und Krankheiten, darunter Krebs, Neurodegeneration und kardiovaskuläre Entwicklung.

Es ist wichtig, sich auf den N-terminalen Teil von Proteinen zu konzentrieren, um unser Wissen dieses Weges zu erweitern. Dazu habe ich die Stabilität der Bibliothek mit allen Varianten der kanonischen 20 Aminosäuren an den ersten beiden Positionen (Ubi-XZ-Bibliothek) durch globales Proteinstabilitätsprofiling (GPS) getestet. Zu diesem Zweck wurde der eK-sfGFP-P2A-mCherry-Reporter ausgewählt, der die höchste Präzision zeigte. Ich habe gezeigt, dass die Ubi-XZ-Bibliothek hochgradig reproduzierbar ist, was durch eine hohe Konsistenz über Replikate hinweg belegt wird. Außerdem ergab die Stabilität von vierzehn Konstrukten mit spezifischen Dipeptiden im Vergleich mit den gleichen Dipeptiden in der Ubi-XZ-Bibliothek eine starke Korrelation. Das wichtigste Ergebnis der Ubi-XZ-Bibliothek ist die unerwartete Stabilität der N-terminalen E- und Q-Reste, die typischerweise über den Arg/N-Degron-Signalweg abgebaut werden. Um die Gründe für diese Beobachtung weiter zu erforschen, habe ich verschiedene Hypothesen untersucht.

Die unterschiedliche Expression wichtiger Enzyme im Arg/N-Degron-Signalweg könnte dessen Funktionalität beeinträchtigen. Die Transkriptionsniveaus von ATE1, NTAN1, NTAQ1, UBR1, UBR2, UBR4 und UBR5, die Teil dieses Signalwegs sind, im Vergleich mit anderen Genen in der HEK293-Zelllinie wiesen jedoch darauf hin, dass diese Gene exprimiert werden, was darauf hindeutet, dass die Abweichung nicht auf einen Mangel dieser Proteine zurückzuführen ist. Eine andere Hypothese betraf die Substratspezifität von Enzymen, insbesondere NTAQ1, das zur Desamidierung auf das N-terminale Q abzielt. Experimente mit humanem NTAN1 und NTAQ1, exprimiert in Hefezellen in Abwesenheit des Nta1-Enzymes, bestätigten, dass NTAN1 und NTAQ1 spezifisch für ihre Substrate sind.

Zwei Studien zeigten, dass die Acetylierung von N-terminalen Resten den Abbau durch den Arg/N-Degron-Weg verhindern. Um dies im Kontext meiner Ergebnisse zu untersuchen, wurden 12 Konstrukte mit Dipeptiden, die mit N, Q, E oder D beginnen, und mit einem zweiten Rest, der L, R oder D war, in der ATE1-Knockout-Zelllinie exprimiert. Diese Dipeptide wurden acetyliert, was dem Stabilitätsmuster in der Ubi-XZ-Bibliothek entspricht und verhindert, dass der Arg/N-Degron-Weg auf die N-terminale Q- und N-Reste abzielt. Zukünftige Forschung sollte identifizieren, welche NAT-Enzyme N-terminale E- und D-Reste acetylieren, wobei NatH ein potenzieller Kandidat ist.

Proteomic Peptide Phage Display (ProP-PD) identifizierte Sequenzen am N-Terminus menschlicher Proteine, die an NTAN1, NTAQ1 oder ATE1 binden. Es wurde eine N-terminale ProP-PD-Bibliothek mit allen menschlichen Proteinisoformen verwendet, wobei 14 Binder für NTAQ1 und 7 für ATE1-Isoform 1 gefunden wurden. Interessanterweise begannen Binder für NTAQ1 nicht mit Q und diejenigen für ATE1-Isoform 1 nicht überwiegend mit E oder D. Weitere Tests der Stabilität dieser Peptide in Zellen ohne NTAQ1 oder ATE1 könnten neue Degrons zutage fördern.

Diese Studie hebt Acetylierung als Mechanismus hervor, der den Proteinabbau über den Arg/N-Degron-Weg verhindert.

Introduction

Cell homeostasis

Protein homeostasis, or proteostasis, is a key requirement for the well-being of any cell. Keeping cells in proteostasis is not an easy job and can encounter many challenges, including managing unfolded, misfolded, or damaged proteins. Improper mechanisms dealing with these challenges are associated with aging and can lead to different diseases like cancers, immunological and metabolic diseases, as well as neurodegenerative disorders such as Huntington's, Alzheimer's, Parkinson's, and amyotrophic lateral sclerosis (ALS) (Dikic 2017a; Labbadia and Morimoto 2015; Yerbury et al. 2016).

To cope with these challenges, cells have a proteostasis network (PN), which consists of ~2000 cellular components. This network is composed of a few main processes: protein synthesis, folding/assembly, trafficking, and degradation by the ubiquitin-proteasome system (UPS) or autophagy-lysosome pathway. Molecular chaperones together with their regulatory cofactors are connecting all of these processes (Klaips, Jayaraj, and Hartl 2018; Labbadia and Morimoto 2015). Since cells are constantly facing internal or external changes as well as stress, they always need to repair or eliminate damaged proteins. If the proteins are not able to be refolded they get degraded by one of two main degradation mechanisms UPS or autophagy-lysosome pathway (Dikic 2017b). The importance of those two degradation mechanisms is also recognized by two Nobel Prizes. One in 2004 in Chemistry for Aaron Ciechanover, Avram Hershko, and Irwin Rose for discovering ubiquitin-mediated protein degradation. In 2016, Yoshinori Ohsumi got the Nobel Prize in Medicine and Physiology for his research on autophagy.

The autophagy-lysosome pathway degrades mostly larger cytoplasmic materials like protein aggregates, organelles, or lipid droplets. For now, there are three known types of autophagy pathways: macroautophagy, chaperone-mediated autophagy (CMA), and microautophagy. The purpose of all of those types is to bring substrates to the lysosome and get it degraded (Ichimiya et al. 2020; Raffeiner et al. 2023). On the other hand, UPS has a goal to primarily degrade short-lived, misfolded, or damaged proteins.

Ubiquitin proteasome system (UPS)

The UPS system involves the conjugation of ubiquitin (Ub), typically on lysine residues of target proteins, marking them for degradation. Ubiquitin is a 76 amino acid protein, which is highly conserved. This protein modification is called ubiquitination, in which three ubiquitin enzymes

are involved in this process, the E1 Ub-activating enzyme, the E2 Ub-conjugating enzyme, and the E3 Ub-protein ligase (Figure 1) (Berndsen and Wolberger 2014; Finley et al. 2012; Hochstrasser 1996; Varshavsky 2011a; Zheng and Shabek 2017). Ubiquitin is activated by the Ub-activating enzyme in the ATP-dependent step and is transferred to the Ub-conjugating enzyme E2 by making a thioester bond. Further E3 Ub-protein ligase transfers it to the substrate. An isopeptide bond is formed between the C-terminal glycine of ubiquitin and the lysine residue on the substrate, but in rare occasions, it can also be serine, threonine, or cysteine (X. Wang, Herr, and Hansen 2012). There are more than 600 E3 ligases in human cells and they are divided into three classes: HECT (Homologous to E6-associated protein C Terminus), RING (Really Interesting New Gene), and RBR (RING-between-RING) (Berndsen and Wolberger 2014; Wei Li et al. 2008). The HECT E3s use an active site cysteine in their domain to form a thioester bond with ubiquitin from an E2 enzyme, before transferring it to the substrate. On the other hand, with RING E3s, ubiquitin is directly transferred from an E2 to a substrate without forming a thioester intermediate with the E3. The third class is a subclass of RING E3s which has two RING domains, one binds to an E2 and another accepts the ubiquitin (Zheng and Shabek 2017).

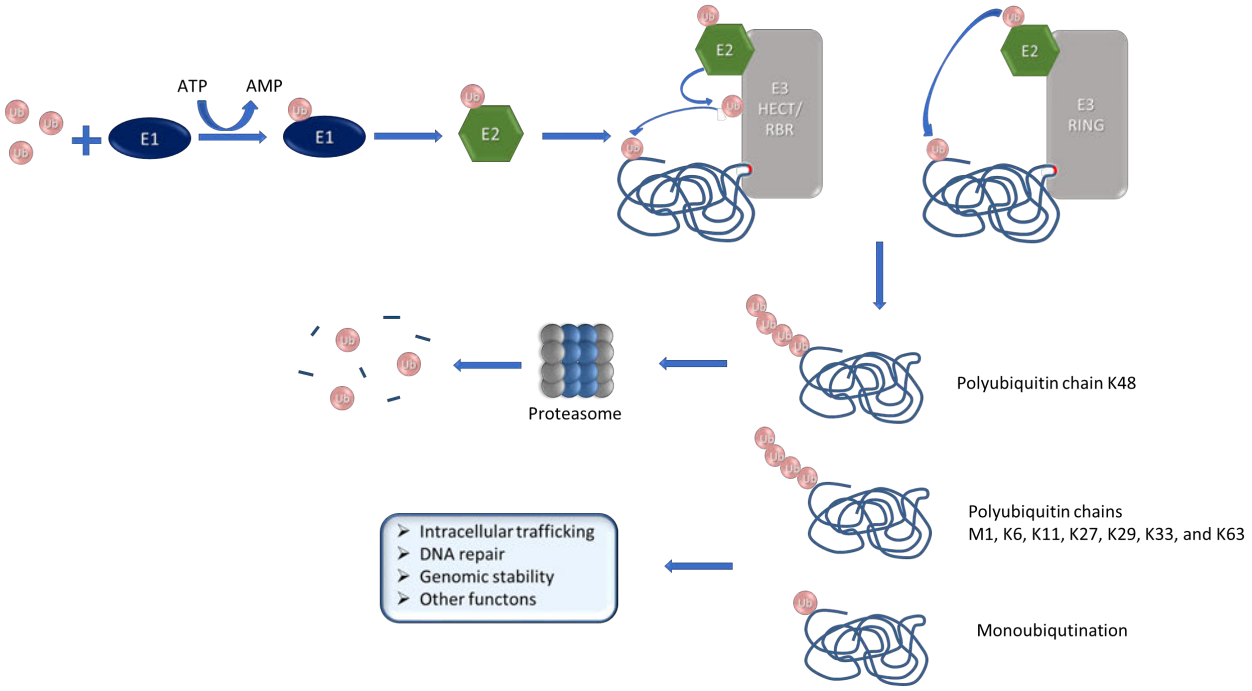


Figure 1. Ubiquitin proteasome system in which ubiquitin is transferred to the substrate for degradation through a cascading process involving E1, E2, and E3 enzymes.

Monoubiquitination is the simplest form of ubiquitination that can also occur in multiple lysine residues and make multi-monoubiquitination. Even though it was thought for a long time that monoubiquitination is a phosphorylation-like modification which can regulate the interaction and activity of substrates, later it was shown that this modification can function as a degradation signal.

Besides monoubiquitination, there are also multiubiquitination and polyubiquitination. There are eight types of homotypic polyubiquitin chains assembled through one of seven internal lysines (K6, K11, K27, K29, K33, K48, and K63) or the α-amino group of the N-terminal methionine (M1). Besides this, the polyubiquitin chains can also be heterotypic (mixed) or branched. The diversity of ubiquitin chains reflects the diversity of signals they produce to get different functions in the cell. This includes proteasomal degradation, selective autophagy, cell signaling pathways like NF-κB activation, protein transport, DNA repair, genomic stability, regulation of the cell cycle, and programmed cell death (Damgaard 2021; W. Li and Ye 2008). The K48 linkage is the predominant lineage in the cells (52% of all linkages in human cells and 29.1 ± 1.9% in yeast cells) which can be recognized by a 26S proteasome and send substrates to degradation (Dammer et al. 2011; Xu et al. 2009).

Ubiquitination processes are reversible and countered by deubiquitinases (DUBs), which cleave ubiquitin from target proteins. The human genome encodes around 100 DUBs, classified into two main categories: cysteine proteases which use cysteine as a catalytic residue, and metalloproteases which use a catalytic serine and zinc as a cofactor (Leznicki and Kulathu 2017; Nijman et al. 2005). Besides recycling ubiquitin, DUBs also regulate the rate of substrate degradation by the proteasome (Farshi et al. 2015).

Degrons

To specifically target particular proteins for degradation by the 26S proteasome, proteins first need to be recognized by particular E3 Ub-protein ligases. For that, the E3 Ub-protein ligase target degradation signals known as degrons (Mészáros et al. 2017; Ravid and Hochstrasser 2008; Zheng and Shabek 2017). Degrons are short linear motifs (SLiMs), containing mostly 2 to 10 amino acids that can be located on the N-terminus, C-terminus, or internal part of a protein (Ella, Reiss, and Ravid 2019; R. Timms and Koren 2020). An important feature of degrons is that they are transferable, meaning that if the sequence is transferred to any protein it would promote the instability of that protein (Ravid and Hochstrasser 2008). Degrons are mostly localized in disordered regions of proteins and there are two major types of them, constitutive and conditional (Lucas and Ciulli 2017; R. Timms and Koren 2020). Constitutive degrons are sequences of amino acids that are mostly hydrophobic and usually buried in protein core or at interaction surfaces between protein subunits. They can become exposed due to improper folding or when the complex fails to assemble properly. External or internal changes and mutations can also influence exposure. On the other hand, conditional degrons are obtained by post-translational modifications (PTM), or they become exposed after protease cleavage. Some of the PTMs are phosphorylation, ligation of Small Ubiquitin-like Modifier (SUMOylation), and hydroxylation (Ella, Reiss, and Ravid 2019).

Protein termini, both N- and C-termini, are favorable sites for degron motifs due to several reasons:

1) N-terminus (42%) and C-terminus (30%) are more likely to be disordered compared to internal regions of proteins (28%) (Lobanov et al. 2010).

2) they are more accessible (Jacob and Unger 2007).

3) protein termini undergo various post-translational processing events, providing opportunities to regulate protein degradation (R. Timms and Koren 2020).

Based on the protein terminus there are two main types of degron pathways, N-degron pathways and C-degron pathways. The N-degron pathways target degrons located at the N-terminus usually obtained by proteolytic cleavage or enzymatic N-terminal-modifications of proteins (Hwang, Shemorry, and Varshavsky 2010b; Piatkov et al. 2014). For now, there are four N-degron pathways known: Arg/N-degron pathway, Ac/N-degron pathway, Pro/N-degron pathway, and Gly/N-degron pathway. The first C-degron was found in 1996 and contained the sequence ANDENYALAA which was targeted for degradation by proteasome-like bacterial protease ClpXP (Keiler, Waller, and Sauer 1996). It took a long time before Elledge and Yin's laboratories discovered a bigger set of C-degrons in human cells in 2018 (Chatr-aryamontri, van der Sloot, and Tyers 2018; Koren et al. 2018; Lin et al. 2018). The C-degrons are predominantly targeted by the Cullin-RING E3 ligases (CRL) which is the largest sub-family of E3 ligases (Lu et al. 2021). The CRLs are multi-subunit complexes containing substrate receptors, adaptors, Cullin scaffolds, and RING-box proteins (Bulatov and Ciulli 2015).

Although both N-terminal and C-terminal degrons are equally important for understanding protein degradation, my research specifically concentrates on N-terminal degrons. Therefore, the following sections will focus more on the N-degron pathway.

N-degron pathways

Arg/N-degron pathway

The first degron pathway discovered by the laboratory of Alexander Varshavsky 37 years ago was the Arg/N-degron pathway (Bachmair, Finley, and Varshavsky 1986). In the Arg/N-degron pathway, E3 ubiquitin ligases target specific unacetylated N-terminal residues for degradation. In yeast, tertiary destabilizing N-terminal residues N and Q may undergo deamidation catalyzed by

Nta1 (Protein N-terminal amidase) and become secondary destabilizing residues (Figure 2A). After that, secondary destabilizing residues, D and E can also be substrates for degradation via arginylation mediated by Arginyltransferase 1 (ATE1) and become primary destabilizing residues. Those residues then can be targeted for degradation by E3 ubiquitin ligase (Baker and Varshavsky 1995; Balzi et al. 1990). There are two types of primary residues, positively charged residues H, K, and R (type I) and bulky hydrophobic residues W, I, F, L, and Y (type II). In addition, a study from 2014 showed that M at the N-terminus followed by W, I, F, L, and Y is an additional primary residue type I (Kim et al. 2014). Primary destabilizing residues are recognized for proteasomal degradation by a physical complex of the RING-type Ubr1 E3 and the HECT-type Ufd4 E3, along with their corresponding E2 enzymes Rad6 and Ubc4/Ubc5, respectively. Bounding Ubr1 to Ufd4 E3 benefits in better processing of ubiquitination (Hwang et al. 2010).

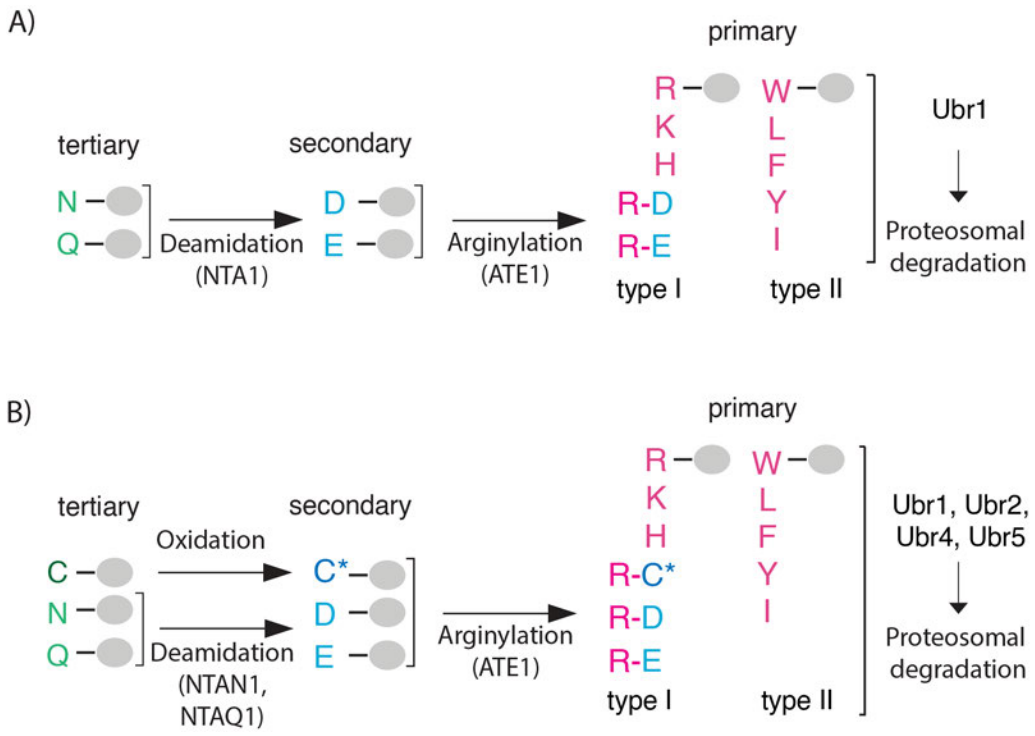


Figure 2. A) Arg/ N-degron pathway in yeast in which primary N and Q at the protein N-terminus are deamidated by NTA1 producing D and E. After arginilation by ATE1 they become primary destabilizing residues, and together with other primary residues can be detected for degradation by Ubr1. B) Arg/ N-degron pathway in mammalian cells, which differs from the yeast pathway in having two deamidation enzymes instead of one, the possibility for oxidated C to become primary destabilizing residues through argiinylation. In mammalian cells N-terminal residues are targeted for degradation by four UBR-family E3 ligases.

In mammalian cells, tertiary destabilizing N-terminal residues N and Q are not targeted for deamidation with only one protein like in yeast, but with two specific proteins (Figure 2B). The N-terminal starting with N can be deamidated by N-terminal asparagine amidohydrolase (NTAN1), while the N-terminal starting with Q can be deamidated by N-terminal glutamine an N-terminal glutamine amidohydrolase (NTAQ1). Interestingly even though NTA1, NTAN1, and NTAQ1 have similar catalytic functions, their sequences are not similar (J. S. Park et al. 2020; M. S. Park et al.

2014). NTAN1 and NTAQ1 are largely located in the cytosol and nucleus (H. Wang et al. 2009). In addition, cleavage of cellular proteins by caspases or calpains can produce fragments with C residue at the N-terminus. The oxidation of the N-terminal C can occur with the involvement of nitric oxide, which is naturally present in mammalian cells but becomes elevated during inflammation, thereby promoting oxidation (Coleman 2001; Hu et al. 2005; Kwon et al. 2002).

Next, Arginyltransferase 1 (ATE1) can target secondary destabilizing resides D, E, and oxidized C for arginylation making them primary residues. ATE1 has two isoforms identical except for a 129-bp region encoding similar yet distinct sequences (Kwon, Kashina, and Varshavsky 1999b). Further primary destabilizing residues can be targeted for degradation by one of the UBR-family E3 ligases (Varshavsky 2011b). The mammalian genome encodes at least seven UBR box proteins, from which Ubr1, Ubr2, Ubr4, and Ubr5 (known as EDD/hHYD) are part of the Arg/N-degron pathway. Ubr1 and Ubr2 are similar in sequence with yeast Ubr1, while the similarity of them with Ubr4 or Ubr5 is mostly fixed just on the UBR domain (Tasaki et al. 2012). While UBR1, UBR2, and UBR4 can bind both type I and type II residues, UBR5 prefers type I N-degrons (Tasaki et al. 2009).

The UBR box, a zinc finger comprising approximately 70 residues, is known for its binding affinity towards type I residues (Choi et al. 2010). The recognition of type II residues by UBR1 and UBR2 is facilitated by their N-domains, which interestingly can be traced back to the evolutionary origins of the bacterial N-recognin ClpS (Lupas and Koretke 2003).

A recent study found that ATE1, NTAN1, NTAQ1, and UBR1-UBE2A/B or UBR2-UBE2A/B can form a complex (Oh et al. 2020). They also proposed so-called substrate channeling, where the substrate can be transferred between active sites of enzymes without the substrate's dissociation into the bulk solution.

Besides targeting proteins for UPS degradation, the Arg/N-degron pathway can target a protein for autophagy degradation. For example, C at the protein terminus can be switched between two pathways. Normally, oxidized C is arginylated by ATE1, leading to K48 polyubiquitination by UBR family E3 ligases. Transient hypoxia causes C to be oxidized, but in the case of prolonged hypoxia or oxidative stress, C is chemically oxidized by reactive oxygen species (ROS), followed by N-terminal arginylation. This degron is then recognized by KCMF-1, in a complex with UBR4, which collaboratively assembles K63/K27-linked ubiquitin chains. These chains then undergo lysosomal degradation via p62-dependent macroautophagy (Ji and Kwon 2017; Yoo et al. 2018).

Numerous proteins have been identified to be regulated by the Arg/N-degron pathway. Many of them are part of pathways that have been linked to the development of cancer, neurodegeneration, and cardiovascular development (Kwon et al. 2002). Some substrates of the Arg/N-degron pathway are proteins TDP43, α-synuclein, PINK1 and Aβ42 related to neurodegenerative disorders like amyotrophic lateral sclerosis (ALS), frontotemporal lobar degeneration (FTLD or FTLD-TDP), primary lateral sclerosis or Parkinson (Brower, Piatkov, and Varshavsky 2013; Eldeeb et al. 2022). Caspases can generate proapoptotic fragments in which

accumulation can increase apoptosis. Arg/N-degron pathway targets those fragments (Cys-RIPK1, Cys-TRAF1, Asp-BRCA1, Leu-LIMK1, Tyr-NEDD9, Arg-BID, Asp-BCLXL, Arg-BIMEL, Asp-EPHA4, and Tyr-MET) for degradation preventing small the proapoptotic signal to lead to apoptosis of the cell (Piatkov, Brower, and Varshavsky 2012). Besides this, the pathway targets proinflammatory fragments which include caspases (Asn-hCASP-1, Gln-hCASP-4, Gln-hCASP-5, Cys-RAB39a, Ile-GRZA) (Leboeuf, Pyatkov, et al. 2020a).

The Arg/N-degron pathway was found to be related to Johanson-Blizzard syndrome since a lack of Ubr1 can cause it (Zenker et al. 2005). Furthermore, many studies done in mice showed different physiological functions to which this pathway contributes (Heo et al. 2023).

In some cases, targeting the Arg/N-degron pathway may provide a therapeutic strategy for these conditions (Eldeeb et al. 2022; Leboeuf, Pyatkov, et al. 2020b; R. Timms and Koren 2020). One of the examples is using siRNA to knock down Ubr1, Ubr2, Ubr4, and Ubr5 which can reduce cancer cell migration and proliferation alongside increasing spontaneous apoptosis (Leboeuf, Abakumova, et al. 2020).

Deeper insight into the specificity of the ubiquitin proteasomal system, especially the Arg/N-degron pathway, can enhance our understanding of protein degradation mechanisms and potentially lead us to novel therapeutic strategies.

Ac/N-degron pathway

N-terminal acetylation is the most common modification in eukaryotes, 80% of human proteins and 60% of yeast proteins are N-terminal acetylated (Aksnes et al. 2016; Khoury, Baliban, and Floudas 2011). In the Ac/N-degron pathway, the removal of the initiator methionine (iMet) from protein N-termini is catalyzed by methionine aminopeptidases (MetAPs) when the second residue belongs to a specific set of amino acids A, S, T, C, V, G, and P (Figure 3). Next, the exposed N-terminus undergoes acetylation mediated by NatA acetyltransferases, which acetylates 40% of human proteins. Besides that, if the first residue is M and the second residue is D, N, E, or Q, NatB will acetylate the N-terminus with M. Furthermore, if the first residue is M and the second one of the following amino acids L, I, F, W, M, Y, H, K or R, then M will be acetylated by NatC, NatE or NatF. In total, there are eight N-terminal acetyltransferases (NAT). Besides the already mentioned five, there are NatD, NatG, and newly found NatH (Aksnes, Ree, and Arnesen 2019; Fluge et al. 2002; Polevoda et al. 2008). NatD is a single-subunit enzyme with a specialized catalytic core that specifically acetylates H2A and H4 histone proteins post their initial N-terminal methionine removal (Mullen et al. 1989). NatG is specific just for plants. The most recently found NatH is located in the cytosol and acetylates post-translationally. It acetylates acidic N termini (DDDI and EEEI) of actin (Drazic et al. 2018). NATs are enzyme complexes containing one of the catalytic subunits NAA10/20/30/40/50/60/70/80. The auxiliary subunits of NatA/B/C are named NAA15,

NAA25, and NAA35/NAA38, respectively (Aksnes, Ree, and Arnesen 2019; Deng and Marmorstein 2021).

The E3 ligases target the acetylated N-terminus of a protein for degradation. In yeast that is the job of two E3 Ub-protein ligases, Doa10 is embedded in the endoplasmic reticulum (ER) and Not4 which is a part of the large Ccr4-Not complex (Hwang, Shemorry, and Varshavsky 2010a; Shemorry, Hwang, and Varshavsky 2013). In mammalian cells, the E3 ligase that is a part of the Ac/N-degron pathway is MARCH6 (also known as TEB4 or MARCHF6) which is a human ortholog of Doa10 (Hassink et al. 2005).

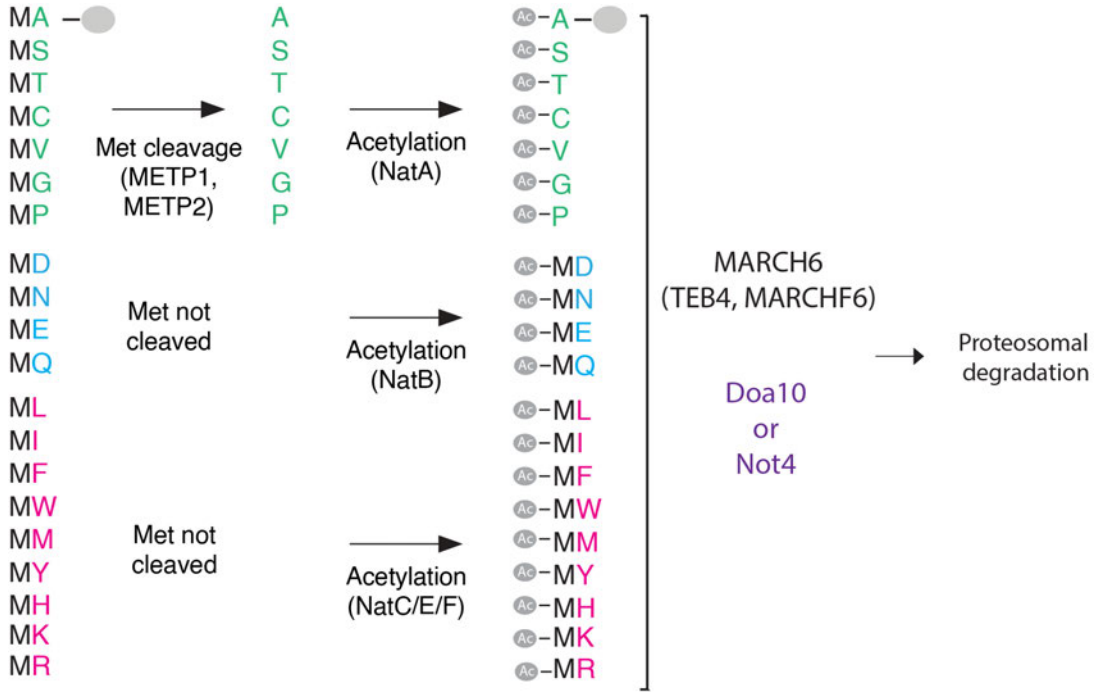


Figure 3. Ac/ N-degron pathway in yeast and mammalian cells in which acetylated N-terminus previously cleaved or not cleaved by methionine aminopeptidase are targeted for proteasomal degradation or my MARCH6 in mammalian cells or with Doa10 or Not4 in yeast (colored in purple on the cartoon).

Even though internal lysines are acetylated by lysine acetyltransferases (KATs), and the reaction is reversible by lysine deacetylases (KDACs), for now, it is not known deacetylases for the N-terminus. This suggests that N-terminal acetylation is very stable and irreversible (Drazic et al. 2016). On the other hand, a study in yeast from 2018 showed that Doa10 targets mostly hydrophobic N-termini arguing that acetylation is rarely recognized as a degron (Kats et al. 2018).

Interestingly, some studies showed that N-terminal acetylation does not always promote degradation but it can also prevent it. In human cells, it was found that NatC acetylates the N-terminus of a set of substrates also recognized by the UBR4-KCMF1 complex (Varland et al. 2023). Furthermore, another study found NatA does N-terminal acetylation of proteins with inhibitor of apoptosis binding motifs (IBMs) at the N-terminus. Like this, it prevents the binding of inhibitors of apoptosis proteins (IAPs) that can lead to cell apoptosis (Mueller et al. 2021). A study in *Arabidopsis thaliana* showed that N-terminal acetylation by NatA masks N-degrons of a subset of

cytosolic proteins and like that prevents their degradation (Linster et al. 2022). Furthermore, the study in yeast found two pathways, MN and EZ pathways, in which degradation was prevented by acetylation (Kats et al. 2018). The MN pathway, Map1 can remove M on the N terminus which is then selected for degradation by the Arg/N-degron pathway, but this can be prevented by N-terminal acetylation done by NatB. Even though Map1 has a low activity in vitro, in cells without NatB, and in certain contexts, Map1 cleaves initial M almost completely. The EZ pathway (where Z represents any of 20 amino acids) involves direct acetylation of N-terminal E residues by NatA, preventing degradation by the Arg/N-end rule pathway.

It is very unlikely that Naa10 influences the turnover of endogenous proteins since NatA only acetylates co-translationally in yeast and there is no known protease that can cleave before E. On the other hand in mammalian cells, a fraction of Naa10 can acetylate the N-termini of EZ post-translationally, so it would be interesting if similar pathways could be found in mammalian cells (Van Damme et al. 2011; Gautschi et al. 2003; Polevoda et al. 2008).

Pro/N-degron pathway

An E3 ligase of this pathway was first described 26 years ago in *Saccharomyces cerevisiae* (Hämmerle et al. 1998). The E3 ligase complex was named GID (glucose-induced degradation) complex and it was shown to target gluconeogenic enzymes, Fbp1 (fructose-1,6-bisphosphatase), Mdh2 (malate dehydrogenase), Pck1 (phosphoenolpyruvate carboxykinase), and Icl1 (isocitrate lyase) for degradation upon switching from growing on ethanol to glucose (Chen et al. 2017; Scherl et al. 2002). It was found that a particular subunit of the GID complex, Gid4 recognizes prolines at the protein N-terminus after methionine gets cleaved by one of two methionine aminopeptidase, MetAP1 or MetAP2 (Figure 4) (Santt et al. 2008). Besides that, it was found that proteins Pck1 starting with SP, Aro10 starting with AP, and Aro10 starting with APPP, after cleavage of S or A by Aminopeptidase Fra1 or Icp55 leaves an exposed P at the N-terminus. This P can be then targeted for degradation by Gid4. On the other hand, it was found that Gid10 also can target N-terminal proline but in particular conditions like starvation or osmotic stresses (Melnykov, Chen, and Varshavsky 2019; Qiao et al. 2020). Those two subunits are interchangeable depending on the conditions. Interestingly enough, a recent study showed the existence of Gid11 that can target exposed threonine at the protein N-terminus after cleavage of the first methionine

by MetAP (Kong et al. 2021). This discovery led to the renaming of the Pro/N-degron pathway into the Pro/Thr/N-degron pathway (Chen et al. 2021).

In mammalian cells, C-terminal to LisH (CTLH) complex is the homolog of the yeast GID complex. Even a very conserved CTLH complex does not have the same function in mammalian cells as the GID complex in yeast (Francis, Han, and Adams 2013). In mammals, this complex is involved in erythropoiesis, organ development, embryogenesis, and cell division, but not in gluconeogenesis like in yeast (Lampert et al. 2018; Soni et al. 2006). Gluconeogenesis in mammals takes place mostly in hepatocytes and also in the kidney cortex and intestine but to a much lower extent (Mutel et al. 2011). Even peptide starting with proline (PGLWKS) was shown by cryo-microscopy to bind to human Gid4 (hGid4), for now, no endogenous substrates with P at the N-terminus were found (Chrustowicz et al. 2022a). For now, the known CTLH substrates are HBP1, ZMYND19, HTRA2 and ARHGAP11A (Bagci et al. 2023; Boldt et al. 2016; Lampert et al. 2018).

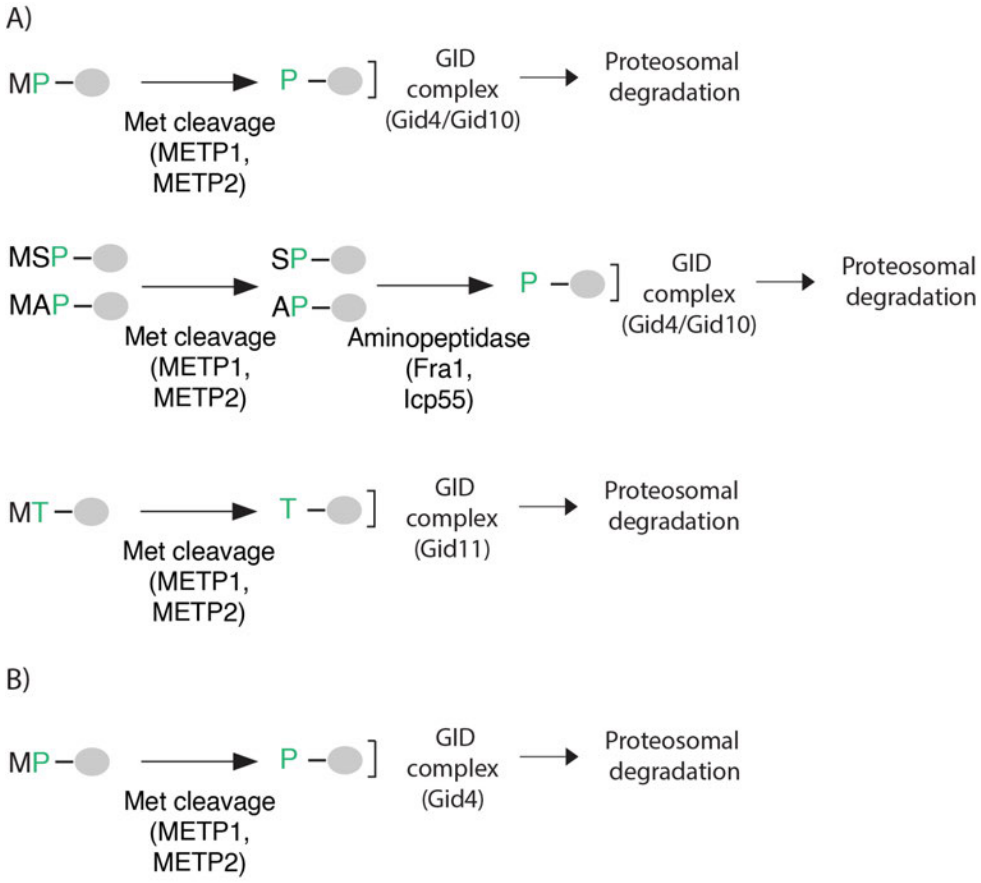


Figure 4. A) Pro/Thr/N-degron pathway in yeast Saccharomyces cerevisiae in which GID complex can target N-terminal P or T for proteasomal degradation. B) Pro/N-degron pathway in mammals in which GID complex can target N-terminal P for proteasomal degradation.

Gly/N-degron pathway

In the recently discovered Gly/N-degron pathway in human cells, two Cullin-RING ligases (CRLs) complexes with substrate adaptors, ZYG11B and ZER1 target N-terminal glycine degrons (Figure 5) (R. T. Timms et al. 2019a). Cullin-RING ligases are a sub-family of RING E3 ligases and it is responsible for up to 20% of ubiquitin-dependent protein turnover in cells (Lydeard, Schulman, and Harper 2013; Soucy et al. 2009). They contain multiple subunits from which Cul1, Cul2, Cul3, Cul4A, Cul4B, and Cul5 in human cells are central scaffolds of a complex. At the C-terminus, the RING protein (either Rbx1 or Rbx2) helps in the recruitment of the E2 conjugating enzyme, while the N-terminus helps with substrate recruitment by particular substrate adaptors. In the case of the Gly/N-degron pathway adaptors, they are ZYG11B and ZER1, which have very similar sequences (29% amino acid similarity) (R. T. Timms et al. 2019a). It is suggested that ZYG11B and ZER1 have overlapping substrate specificity. Additionally, while ZYG11B requires an N-terminal glycine for recognition, ZER1 may recognize substrates with alternative initial residues. One such substrate, derived from KCNT2 begins with MPYL.

Besides this, it was proposed that ZYG11B and ZER1 recognize N-terminal glycine degrons failed to be N-myristoylated. N-myristoylation, a post-translational modification that attaches a 14-carbon fatty acid (myristate) to proteins, occurs exclusively on N-terminal glycine residues. This

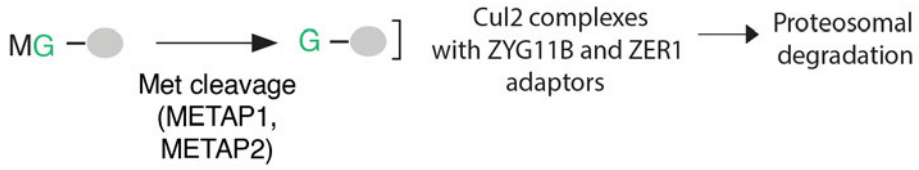


Figure 5. Gly/N-degron pathway in which N-terminal glycine degrons exposed after cleavage by one of methionine aminopeptidase are targeted for degradation by one of two Cullin-RING ligases (CRLs) complexes with substrate adaptors ZYG11B and ZER1

modification is crucial for regulating membrane localization and other properties of several hundred human proteins. It looks like that N-myristoylation shields a protein from degradation by Gly/N-degron pathway (R. T. Timms et al. 2019b).

Materials and Methods

Cell lines and cell culturing

In this thesis, Flp-In T-REx 293 (Thermo Fisher Scientific, R78007), HEK-293FT (Thermo Fisher Scientific, R70007), and HEK-293T (ATCC, CRL-3216) cell lines were used. Cells were grown in a Forma Steri-Cult CO2 Incubator at 37°C with 5% CO2 and 85% of humidity. Cells were grown in

complete DMEM medium containing Dulbecco's Modified Eagle's Medium (DMEM) (Life Technologies, 21969035) supplemented with Penicillin-Streptomycin (10,000 U/mL) (Life Technologies, 15140122), L-Glutamine (200 mM) (Life Technologies, 25030024) and 10% of inactivated Fetal Bovine Serum (FBS) (Life Technologies, 10270106).

Cell passaging

Cells were passaged at around 80% of confluency, which was checked by Invitrogen EVOS FL Digital Inverted Fluorescence Microscope. They were first washed with pre-warmed at 37°C DPBS (Life Technologies, 14190094), and then trypsinized with 0.25% Trypsin-EDTA phenol red (Life Technologies, 25200056) for 5min at 37° C. Trypsin was inactivated by addition of complete DMEM media. After centrifugation at 1000g for 5min, the supernatant was discarded and the pellet was resuspended in complete DMEM. Cells were then split in the desired ratio for further growth.

Cell counting

Scepter 2.0 Cell Counter (Millipore Sigma, 32011202) was used for counting the number of cells after trypsinization. TC20 Automated Cell Counter (BioRad, 1450102) was used for the cells in the IMB S2 lab.

Cell freezing and thawing

The cell lines ready for the passaging were first washed with pre-warmed DPBS and trypsinized with 0.25% Trypsin-EDTA phenol red. Trypsin was inactivated by the addition of complete DMEM and cells were pelleted by centrifugation at 1000g for 5min. The supernatant was discarded and the pellet was resuspended with complete media. Resuspended cells were mixed with an equal amount of freezing medium (20% inactivated FBS, 20% DMSO, 60% complete DMEM), divided into cryovials (Thermo Scientific, 5011-0020), and left to freeze at -80°C overnight in Mr. Frosty Freezing Container (Merck, C1562-1EA). The next day cells were moved to a -150°C freezer for permanent storage.

To thaw the cells, frozen cryovials were placed in a pre-warmed water bath at 37°C. After thawing cells were moved to a new tube and centrifuged at 1000g for 3min. The supernatant was immediately discarded and the pellet was resuspended in complete DMEM and added to a new plate to grow.

Plasmids, cell lines, and oligos used in this thesis

Table 1. All Plasmids used in this thesis

ID	Plasmid name	Strain	Construct
304	pRNA063*	DH5α	pLJMI/Ubi-met-eK-sfGFP_PP2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert met-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_PP2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
305	pRNA064*	DH5α	pLJMI/Ubi-thr-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert thr-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
306	pRNA065-1*	DH5α	pLJMI/Ubi-gly-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert gly-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
307	pRNA066*	DH5α	pLJMI/Ubi-val-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert val-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
308	pRNA067*	DH5α	pLJMI/Ubi-his-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert his-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
309	pRNA068*	DH5α	pLJMI/Ubi-tyr-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert tyr-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
310	pRNA069-1	DH5α	pLJMI/Ubi-arg-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert arg-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)

384	pRNA070*	DH5α	pLJMI/Ubi-asn-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert asn-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
385	pRNA071*	DH5α	pLJMI/Ubi-asp-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert asp-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
386	pRNA072	DH5α	pLJMI/Ubi-lys-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert lys-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
400	pRNA077	DH5α	pLJMI/Ubi-gln-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert gln-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
401	pRNA078	DH5α	pLJMI/Ubi-glu-eK-sfGFP_P2A_mCherry Backbone was made by restriction digestion of pRNA062 (id 303) with BspQI restriction enzyme. Insert glu-lenti_ubi_eK was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
1512	pAC0039-1	DH5α	pLJM1-Ubi-NL-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-NL_1 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1514	pAC0040-1	DH5α	pLJM1-Ubi-QL-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-NL_1 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 2
1516	pAC0041-1	DH5α	pLJM1-Ubi-DL-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-DL_3 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1

1518	pAC0042-1	DH5α	pLJM1-Ubi-EL-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-EL_4 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1520	pAC0043-1	DH5α	pLJM1-Ubi-NR-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-NR_5 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1522	pAC0044-1	DH5α	pLJM1-Ubi-QR-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-QR_6 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1524	pAC0045-1	DH5α	pLJM1-Ubi-DR-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-DR_7 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1526	pAC0046-1	DH5α	pLJM1-Ubi-ER-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-ER_8 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1528	pAC0047-1	DH5α	pLJM1-Ubi-ND-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-ND_9 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1530	pAC0048-1	DH5α	pLJM1-Ubi-QD-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-QD_10 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1532	pAC0049-1	DH5α	pLJM1-Ubi-DD-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-DD_11 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1

1534	pAC0050-1	DH5α	pLJM1-Ubi-ED-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-ED_12 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1536	pAC0051-1	DH5α	pLJM1-Ubi-PK-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-PK_13 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1538	pAC0052-1	DH5α	pLJM1-Ubi-PL-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Insert G1403_AC_E6-Ubi-PL_14 was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer) COLONY 1
1088	pAC009-1	DH5α	pET-3C-GST-His-NTAQ1 Backbone was amplified from pPPCF118-1 by Y.001_pPPCF118_240820_F/ Y.001_pPPCF118_240820_R primers. cDNA of NTAQ1 amplified from Media lab clone with primers pAC009_Gib_NTAQ1_F_2 and pAC009_Gib_NTAQ1_R_2. Verified by colony PCR with pAC009_Gib_NTAQ1_F_2 and pAC009_Gib_NTAQ1_R_2. Sequenced by pAC0011_Che1_GST_F COLONY 1
1089	pAC0010.1-1	DH5α	pET-3C-GST-His-ATE1 Backbone was amplified from pPPCF118-1 by Y.001_pPPCF118_240820_F/ Y.001_pPPCF118_240820_R primers. cDNA of ATE1 isoform 1 ordered from IDT and amplified with primers pAC0010_Gib_ATE1_iso1_F and pAC0010_Gib_ATE1_iso1_R. Verified by colony PCR with pAC0010_Gib_ATE1_iso1_F and pAC0010_Gib_ATE1_iso1_R. Sequenced by T7 COLONY 1
1091	pAC0010.2-6	DH5α	pET-3C-GST-His-ATE1 Backbone was amplified from pPPCF118-1 by Y.001_pPPCF118_240820_F/ Y.001_pPPCF118_240820_R primers. cDNA of ATE1 isoform 2 ordered from IDT and amplified with primers pAC0010_Gib_ATE1_iso1_F and pAC0010_Gib_ATE1_iso1_R. Verified by colony PCR with pAC0010_Gib_ATE1_iso1_F and pAC0010_Gib_ATE1_iso1_R. Sequenced by T7 COLONY 6
1129	pAC0028-2	DH5α	pET-3C-GST-His-NTAN1 Backbone was amplified from pPPCF118-1 by Y.001_pPPCF118_240820_F/ Y.001_pPPCF118_240820_R primers. cDNA of NTAN1 amplified from Media lab clone with primers pAC0028_Gib_NTAN1_F and pAC0028_Gib_NTAN1_R. Verified by restriction digestion by BspQ1. Sequenced by T7 and OPP373 (CTGGAAGTTCTGTTCCAGGGGCCC) COLONY 2
1087	pAC008-1	DH5α	pET-NTAN1-3C-GST-His For Backbone was made by restriction digestion of pAC001 with XhoI. cDNA of NTAN1 amplified from Media lab clone with primers pAC008_Gib_NTAN1_F and pAC008_Gib_NTAN1_R. Verified by colony

			PCR withpAC008_Gib_NTAN1_F and pAC008_Gib_NTAN1_R. Sequenced by T7 COLONY 1
1097	pAC0013-1	DH5α	pET-His-GST-3C-GID4(1-358) Backbone was amplified from pPPCF118-1 byY.001_Gid4_Gib_240820_F/pAC0013_Gib_Gid4_358aa_R. GID4(1-358) with stop codon was amplified from yeast genomic DNA by primers pAC0011_Gib_Gid4_90aa_F/Y.001_Gid4_Gib_240820_R. Verified by colony PCR with Y.001_Gid4_Gib_240820_F and Y.001_Gid4_Che_240820_01_R. Sequenced by T7, pAC0011_Che1_GST_F, Y.001_Gid4_Che_240820_01_F COLONY 1
797	pAC003-2	DH5α	pET-His-GST-3C-GID10 Backbone was amplified from pPPCF118-1 by Y.001_pPPCF118_240820_F/ Y.001_pPPCF118_240820_R primers. Full yeast Gid10 with stop codon was amplified from yeast genomic DNA with primers Y.001_Gid10_Gib_240820_F/Y.001_Gid10_Gib_240820_R. Verified by primers Y.001_Gid10_Gib_240820_F/Y.001_Gid10_Gib_240820_R. Sequenced by T7, Y.001_Gid10_Che_240820_01_F and Y.001_Gid10_Che_04_F COLONY 2
1151	pAC0031-1	DH5α	pET-His-GST-3C-GID4(99-36P2Aa) Backbone was amplified from pPPCF118-1 by Y.001_pPPCF118_240820_F/ Y.001_pPPCF118_240820_R primers. Insert GID4(99-36P2Aa with stop codon was amplified from yeast genomic DNA by primers pAC0031_Gib_Gid4_99aa_F/Y.001_Gid4_Gib_240820_R. Verified by restriction digestion with EcoRV. Sequenced by T7, pAC0011_Che1_GST_F, Y.001_Gid4_Che_240820_01_F and T7term COLONY 1
1153	pAC0032-1	DH5α	pET-His-GST-3C-Gid10(115-293aa) Backbone was amplified from pPPCF118-1 by Y.001_pPPCF118_240820_F/ Y.001_pPPCF118_240820_R primers. Insert Gid10(115-293) with stop codon was amplified from yeast genomic DNA by primers pAC0032_Gib_Gid10_115aa_F /Y.001_Gid10_Gib_240820_R. Verified by restriction digestion with ApaI. Sequenced by T7, pAC0011_Che1_GST_F, Y.001_Gid10_Che_240820_01_F and T7term COLONY 1
529	pPPCF118-1	DH5α	pET-His-GST-3C-attR1 Plasmid from IMB's Protein Production Core Facility; Verified by sequencing with T7 Colony 1
1310	Ubi-XZ library*	DH5α	Ubi_XZ-linker-eGFP-P2A-mCherry Backbone was made by restriction digestion of RNA107 (id 760) with BspQI restriction enzyme. Oligo pool of all variants of amino acids on first two positions (library_pool_linker). In total there are 400 sequences. Oligo pool was amplified with primers PR_pool_ek_eGFP (6/98) and PCR_pool_P2Aa_F (6/96) and cloned into backbone by Gibson. Verified by sequencing with CMVfor (eurofins standard primer)
335	pMDLg/pRRE	DH5α	Addgene plasmid #12251. 3rd generation lentiviral packaging plasmid. Contains Gag and Pol; also requires pRSV-Rev (addgene#12253) and envelope expressing plasmid (addgene#12259)

336	pRSV-Rev	DH5α	Addgene plasmid #12253. 3rd generation lentiviral packaging plasmid. Contains Rev and also requires pMDLg/pRRE (addgene#12251) and envelope expressing plasmid (addgene#12259)
337	pMD2.G	DH5α	Addgene plasmid #12259. VSV-G envelope expressing plasmid
504	psPAX2	DH5α	Addgene plasmid #12260 2nd generation lentiviral packaging plasmid. Can be used with 2nd or 3rd generation lentiviral vectors and envelope expressing plasmid (e.g.)
X	zer1	DH5α	from Ivarsson's lab
X	zyg11b	DH5α	from Ivarsson's lab
760	RNA107*	DH5α	pLJM1-Ubi-stop-linker-eGFP-P2A-mCherry, checked by sequencing
303	pRNA062*	DH5α	pLJM1/Ubi-eK-sfGFP-mCherry
277	pAnB19	DH5α	pRS413-GPDpr-Ubi-EcoRV-STOP-eK-mCherry-sfGFP (Kats et al. 2018)
1929	pAC0057	DH5α	Backbone was made by restriction digestion of pAnB19 with EcoRV. The following inserts were cloned in this backbone separately, G1403_AC_NZ_1, G1403_AC_QZ_2, G1403_AC_DZ_3, and G1403_AC_EZ_4. In total 96 colonies were chosen and sent for sequencing. In total 60 clones were correct.

* Plasmids made by [redacted]

Table 2. All cell lines are made in this thesis. Cell lines containing HEK#3 in their name originated from the Flp-In T-REx 293 cell line. Ordered cell line was sorted by flow cytometry, and clone number 3 was selected for further use.

ID	Cell line name	Construct
79	Hek#3_G62	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-met-eK-sfGFP_2A_mCherry (plasmid pRNA063, id304)
80	Hek#3_G63	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-thr-eK-sfGFP_2A_mCherry (plasmid pRNA064, id305)
81	Hek#3_G64	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-gly-eK-sfGFP_2A_mCherry (plasmid pRNA065, id306)
82	Hek#3_G65	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-val-eK-sfGFP_2A_mCherry (plasmid pRNA066, id307)
83	Hek#3_G66	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-his-eK-sfGFP_2A_mCherry (plasmid pRNA067, id308)
84	Hek#3_G67	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-tyr-eK-sfGFP_2A_mCherry (plasmid pRNA068, id309)
85	Hek#3_G68	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-arg-eK-sfGFP_2A_mCherry (plasmid pRNA069, id310)
86	Hek#3_G69	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-met-eK-sfGFP_2A_mCherry (plasmid pRNA063, id304)

87	Hek#3_G70	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing pLJMI/Ubi-met-eK-sfGFP_2A_mCherry (plasmid pRNA063, id304)
149	**hAC007**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing NL-linker-eGFP-2A-mCherry (pAC0039). Lentiviruses were produced by cell line IvAC007 (id148) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
151	**hAC008**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing QL-linker-eGFP-2A-mCherry (pAC0040). Lentiviruses were produced by cell line IvAC008 (id150) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
153	**hAC009**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing DL-linker-eGFP-2A-mCherry (pAC0041). Lentiviruses were produced by cell line IvAC009 (id152) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
155	**hAC010**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing EL-linker-eGFP-2A-mCherry (pAC0042). Lentiviruses were produced by cell line IvAC010 (id154) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
157	**hAC011**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing NR-linker-eGFP-2A-mCherry (pAC0043). Lentiviruses were produced by cell line IvAC011 (id156) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
159	**hAC012**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing QR-linker-eGFP-2A-mCherry (pAC0044). Lentiviruses were produced by cell line IvAC012 (id158) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
161	**hAC013**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing DR-linker-eGFP-2A-mCherry (pAC0045). Lentiviruses were produced by cell line IvAC013 (id160) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
163	**hAC014**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing ER-linker-eGFP-2A-mCherry (pAC0046). Lentiviruses were produced by cell line IvAC014 (id162) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by

		qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
165	**hAC015**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing ND-linker-eGFP-2A-mCherry (pAC0047). Lentiviruses were produced by cell line IvAC015 (id164) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
167	**hAC016**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing QD-linker-eGFP-2A-mCherry (pAC0048). Lentiviruses were produced by cell line IvAC016 (id166) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
169	**hAC017**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing DD-linker-eGFP-2A-mCherry (pAC0049). Lentiviruses were produced by cell line IvAC017 (id168) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
171	**hAC018**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing ED-linker-eGFP-2A-mCherry (pAC0050). Lentiviruses were produced by cell line IvAC018 (id170) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
173	**hAC019**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing PK-linker-eGFP-2A-mCherry (pAC0051). Lentiviruses were produced by cell line IvAC019 (id172) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
175	**hAC020**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing PL-linker-eGFP-2A-mCherry (pAC0052). Lentiviruses were produced by cell line IvAC020 (id174) in the S2 lab. After infection Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
176	**hAC021**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing NR-linker-eGFP-2A-mCherry (pAC0043). Lentiviruses were produced by cell line IvAC0011 (id156) in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
177	**hAC022**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing NQ-linker-eGFP-2A-mCherry (pAC0044). Lentiviruses were produced by cell line IvAC0012 (id158) in the S2 lab. After infection FHEK#3 KO-D ATE1 cells were downgraded and

		checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
179	**hAC024**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing ND-linker-eGFP-2A-mCherry (pAC0047). Lentiviruses were produced by cell line IvAC0015 (id164) in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
180	**hAC025**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing QD-linker-eGFP-2A-mCherry (pAC0048). Lentiviruses were produced by cell line IvAC0016 (id166) in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
181	**hAC026**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing DD-linker-eGFP-2A-mCherry (pAC0049). Lentiviruses were produced by cell line IvAC0017 (id168) in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
182	**hAC027**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing M-linker-eGFP-2A-mCherry (pRNA063). Lentiviruses were produced by cell line id427 in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
208	**hAC045**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing NL-linker-eGFP-2A-mCherry (pAC0039). Lentiviruses were produced by cell line IvAC007 in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
209	**hAC046**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing QL-linker-eGFP-2A-mCherry (pAC0040). Lentiviruses were produced by cell line IvAC008 in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
210	**hAC047**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing QL-linker-eGFP-2A-mCherry (pAC0040). Lentiviruses were produced by cell line IvAC008 in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
211	**hAC048**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing EL-linker-eGFP-2A-mCherry (pAC0042). Lentiviruses were produced by cell line IvAC0010 in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and

		checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
212	**hAC049**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing ER-linker-eGFP-2A-mCherry (pAC0046). Lentiviruses were produced by cell line IvAC0014 in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
213	**hAC050**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing ED-linker-eGFP-2A-mCherry (pAC0050). Lentiviruses were produced by cell line IvAC0018 in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
178	**hAC023.1**	Starting cell line HEK#3 KO-D ATE1 (id 133) was infected by lentiviruses expressing the plasmid containing DR-linker-eGFP-2A-mCherry (pAC0045). Lentiviruses were produced by cell line IvAC0013 (id160) in the S2 lab. After infection HEK#3 KO-D ATE1 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection and checked by microscopy (mCherry positive).
222	**hAC_Ubi_XZ_library**	Starting cell line Flp-In T-REx 293 was infected by lentiviruses expressing the plasmid containing Ubi-XZ library (id 1310). After transduction, Flp-In T-REx 293 cells were downgraded and checked by qPCR. Positive infected cells were selected by puromycin selection, sorted, and sent for NGS.

Table 3. All oligos used in this thesis

Oligo name	Sequence
met-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTATGGGATCCGGAGCTTGGCTGTT
thr-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTACCGGATCCGGAGCTTGGCTGTT
gly-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTGGCGGATCCGGAGCTTGGCTGTT
val-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTGTGGGATCCGGAGCTTGGCTGTT
his-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTCACGGATCCGGAGCTTGGCTGTT
tyr-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTTACGGATCCGGAGCTTGGCTGTT
arg-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTAGAGGATCCGGAGCTTGGCTGTT
asn-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTAACGGATCCGGAGCTTGGCTGTT
asp-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTGACGGATCCGGAGCTTGGCTGTT
lys-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTAAGGGATCCGGAGCTTGGCTGTT
gln-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTCAGGGATCCGGAGCTTGGCTGTT
glu-lenti_ubi_eK	TCCTGCGTCTGAGAGGTGGTGAGGGATCCGGAGCTTGGCTGTT
PR_pool_ek_eGFP	CCGGTGAACAGCTCCTCGCCCTTGCTTCTGGTGCCCAGGGCGGAGGTGGC
PCR_pool_2aa_F	ACATCCAGAAGGAGTCGACCCTGCACCTGGTCCTGCGTCTGAGAGGTGGT
PCR_pool_2aa_R	GTTCCAGGGGCCCCAATCTCCCTATAGTTC
G1403_AC_E6-Ubi-NL_1	TCCTGCGTCTGAGAGGTGGTAACCTGGCCACCTCCGCCCTGGGCAC

G1403_AC_E6-Ubi-QL_2	TCCTGCGTCTGAGAGGTGGTCAGCTGGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-DL_3	TCCTGCGTCTGAGAGGTGGTGATCTGGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-EL_4	TCCTGCGTCTGAGAGGTGGTGAACTGGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-NR_5	TCCTGCGTCTGAGAGGTGGTAACCGTGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-QR_6	TCCTGCGTCTGAGAGGTGGTCAGCGTGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-DR_7	TCCTGCGTCTGAGAGGTGGTGATCGTGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-ER_8	TCCTGCGTCTGAGAGGTGGTGAACGTGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-ND_9	TCCTGCGTCTGAGAGGTGGTAACGATGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-QD_10	TCCTGCGTCTGAGAGGTGGTCAGGATGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-DD_11	TCCTGCGTCTGAGAGGTGGTGATGATGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-ED_12	TCCTGCGTCTGAGAGGTGGTGAAGATGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-PK_13	TCCTGCGTCTGAGAGGTGGTCCGAAAGCCACCTCCGCCCTGGGCAC
G1403_AC_E6-Ubi-PL_14	TCCTGCGTCTGAGAGGTGGTCCGCTGGCCACCTCCGCCCTGGGCAC
G1403_AC_NZ_1	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATNNKGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
G1403_AC_QZ_2	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTCAANNKGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
G1403_AC_DZ_3	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGATNNKGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
G1403_AC_EZ_4	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGAANNKGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
G1403_AC_E5.3-F_5	CAACATTCAGAAGGAGTCCACCTTACATCTTGTGCTAAGGCTAAGAGGTGGT
G1403_AC_E5.3-R_6	GGGTGGTTTTTCTTTTCACCAGTGAGACGGGCAACAGCCAAGCTCCGGATCC
G14_AC_E9_BlpI+P2A	GCTGGCTAGACACCATGAGCTGAGCCACAAACTTCTCTCTGCTAAAGCAAGCAGGTGATGTTGAAGAAAACCCCGGGCCT
G14_AC_E9_NTAN1_F	GCTAGACACCATGAGCATGCCGCTGCTCGT
G14_AC_E9_NTAN1_R	CAGAGAGAAGTTTGTGGCACTTCCTGGAGAAGAG
G14_AC_E9_NTAQ1_F	GCTAGACACCATGAGCATGGAAGGTAATGGCCCCG
G14_AC_E9_NTAQ1_R	CAGAGAGAAGTTTGTGGCGCAGTTTTTACTG
G14_AC_E9_ATE1_F	GCTAGACACCATGAGCATGGCTTTCTGGGCGGG
G14_AC_E9_ATE1_R	GAGAGAAGTTTGTGGCGTTTCTGAACAGCAGC

G14_AC_E9_GST_F	CACCATGAGCATGTCCCCTATAC
G14_AC_E9_GST_R	GAGAAGTTTGTGGCTTTTGGAGGATGG
G14_AC_E9_NTAN1_F_2	GACACCATGAGCATGCCGCTGCTCGTCGAGG
G14_AC_E9_NTAN1_R_2	GCAGAGAGAAGTTTGTGGCACTTCCTGGAGAAG
Y.001_pPPCF118_240820_F	GATTCGAGGCTGCTAACAAAGCCCG
Y.001_pPPCF118_240820_R	GGGCCCCTGGAACAGAACTTCCAG
pAC0010_Gib_ATE1_iso1_F	CCAGGGGCCCATGGCTTTCTGGGCGGGG
pAC0010_Gib_ATE1_iso1_R	GCCTCGAATCTCAGTTTCTGAACAGCAGCATCC
pAC009_Gib_NTAQ1_F_2	CTGGAAGTTCTGTTCCAGGGGCCCATGGAAGGTAATGGCCCC
pAC009_Gib_NTAQ1_R_2	GGCTTTGTTAGCAGCCTCGAATCTCAGCAGTTTTTACTGCC
pAC0011_Che1_GST_F	GGACCCAATGTGCCTGGA
T7	TAATACGACTCACTATAGGG
pAC0028_Gib_NTAN1_F	TCTGTTCCAGGGGCCCATGCCGCTGCTCGTCGAG
pAC0028_Gib_NTAN1_R	GTTAGCAGCCTCGAATCTCAACTTCCTGGAGAAGAG
OPP373	CTGGAAGTTCTGTTCCAGGGGCCC
pAC008_Gib_NTAN1_F	TTTAAGAAGGAGATATACCATGCCGCTGCTCGTCGAG
pAC008_Gib_NTAN1_R	CCTGGAACAGAACTTCCAGACTTCCTGGAGAAGAGATCTTTTCCC
pAC0013_Gib_Gid4_358aa_R	GCCTCGAATCTCAAGAACAATCACTGGAC
pAC0011_Gib_Gid4_90aa_F	CCAGGGGCCCACAAGGAAGAAATATATTGTTG
Y.001_Gid4_Gib_240820_R	TTTGTTAGCAGCCTCGAATCTCAAGCAAACTCAAAAGAAC
Y.001_Gid4_Che_240820_01_R	GCTAGAGGACAAGAACCCTAATTCC
Y.001_Gid10_Gib_240820_F	AAGTTCTGTTCCAGGGGCCCATGACATCACTGAATATCATG
Y.001_Gid10_Gib_240820_R	TTTGTTAGCAGCCTCGAATCTTAGGCAATTTCAAATCCAC
pAC0031_Gib_Gid4_99aa_F	GTTCCAGGGGCCCCAATCTCCCTATAGTTC
pAC0032_Gib_Gid10_115aa_F	GTTCCAGGGGCCCTCTTTGGACCCGCATGTC
CMVfor	CGCAAATGGGCGGTAGGCGTG
library_pool_linker	TCCTGCGTCTGAGAGGTGGTNNNNNNGCCACCTCCGCCCTGGGCAC

seq20_PCR1_F	CTACACGACGCTCTTCCGATCTGACCCTGCACCTGGTCCT
seq20_PCR1_F2	CTACACGACGCTCTTCCGATCTnnGACCCTGCACCTGGTCCT
seq20_PCR1_F3	CTACACGACGCTCTTCCGATCTnnnnGACCCTGCACCTGGTCCT
seq20_PCR1_F4	CTACACGACGCTCTTCCGATCTnnnnnnnGACCCTGCACCTGGTCCT
seq20_PCR1_bin1_R	GACGTGTGCTCTTCCGATCTCGTGATTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin1_R2	GACGTGTGCTCTTCCGATCTnnCGTGATTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin1_R3	GACGTGTGCTCTTCCGATCTnnnnCGTGATTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin1_R4	GACGTGTGCTCTTCCGATCTnnnnnnnCGTGATTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin2_R	GACGTGTGCTCTTCCGATCTACATCGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin2_R2	GACGTGTGCTCTTCCGATCTnnACATCGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin2_R3	GACGTGTGCTCTTCCGATCTnnnnACATCGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin2_R4	GACGTGTGCTCTTCCGATCTnnnnnnnACATCGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin3_R	GACGTGTGCTCTTCCGATCTGCCTAATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin3_R2	GACGTGTGCTCTTCCGATCTnnGCCTAATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin3_R3	GACGTGTGCTCTTCCGATCTnnnnGCCTAATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin3_R4	GACGTGTGCTCTTCCGATCTnnnnnnnGCCTAATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin4_R	GACGTGTGCTCTTCCGATCTTGGTCATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin4_R2	GACGTGTGCTCTTCCGATCTnnTGGTCATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin4_R3	GACGTGTGCTCTTCCGATCTnnnnTGGTCATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin4_R4	GACGTGTGCTCTTCCGATCTnnnnnnnTGGTCATCGCCCTTGCTTCTGGTG
seq20_PCR1_bin5_R	GACGTGTGCTCTTCCGATCTCACTGTTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin5_R2	GACGTGTGCTCTTCCGATCTnnCACTGTTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin5_R3	GACGTGTGCTCTTCCGATCTnnnnCACTGTTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin5_R4	GACGTGTGCTCTTCCGATCTnnnnnnnCACTGTTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin6_R	GACGTGTGCTCTTCCGATCTATTGGCTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin6_R2	GACGTGTGCTCTTCCGATCTnnATTGGCTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin6_R3	GACGTGTGCTCTTCCGATCTnnnnATTGGCTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin6_R4	GACGTGTGCTCTTCCGATCTnnnnnnnATTGGCTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin7_R	GACGTGTGCTCTTCCGATCTGATCTGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin7_R2	GACGTGTGCTCTTCCGATCTnnGATCTGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin7_R3	GACGTGTGCTCTTCCGATCTnnnnGATCTGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin7_R4	GACGTGTGCTCTTCCGATCTnnnnnnnGATCTGTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin8_R	GACGTGTGCTCTTCCGATCTTCAAGTTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin8_R2	GACGTGTGCTCTTCCGATCTnnTCAAGTTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin8_R3	GACGTGTGCTCTTCCGATCTnnnnTCAAGTTCGCCCTTGCTTCTGGTG
seq20_PCR1_bin8_R4	GACGTGTGCTCTTCCGATCTnnnnnnnTCAAGTTCGCCCTTGCTTCTGGTG
P5_Scriptseq	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T
P7-ix14_Scriptseq	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T
P7-ix15_Scriptseq	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T
P7-ix18_Scriptseq	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT*C*T

For_NewP8_IntLibAmp	CTATTGCTACAAATGCCTATGCA
IntLibAmpRev	GGTGGAGGATCCGGAGG
pAnB19-EQ*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGAACAAGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-EN*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGAAAATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-ED*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGAAGATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-DT*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGATACTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-DI*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGATATTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-DD*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGATGATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-DC*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTGATTGTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-QY*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTCAATATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-QW*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTCAATGGGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-QS*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTCAATCTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-QN*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTCAAAATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-QD*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTCAAGATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-NY*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATTATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-NT*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATACTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-NS*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATTCTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-NN*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATAATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-NK*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATAAAGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-NI*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATATTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-ND*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATGATGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
pAnB19-NC*	CCTTACATCTTGTGCTAAGGCTAAGAGGTGGTAATTGTGGATCCGGAGCTTGGCTGTTGCCCGTCTCAC
NTA1-S1	AATGCTAAAAAGTGCTAGTTTACAGCGATAACTCTATCGTGACATTCAGTGAATGCGTACGCTGCAGGTCGAC
NTA1-S2	ATATGCGTGTAACCTTATTAATTTATAATGTCTAATTTATTATAATATTCCTTCAATCGATGAATTCGAGCTCG

ATE1-S1	TTGTTACAAAAGACAAGATAGTCTCAAGATACAACTGAGAAGAAGCACAATTATGC GTACGCTGCAGGTCGAC
ATE1-S2	TGCTAAAAAATGTTATACATTAGGATTGTTATTAGAATAGTGTGTTGAAGGCTCAAT CGATGAATTCGAGCTCG

* Oligos used in the construction of the yeast library

Generation of single constructs

Plasmids pRNA069-1, pRNA072, pRNA077, and pRNA078

To generate plasmids for single constructs expressed in human cells, all chosen oligonucleotides (with single residues or dipeptides of interest) were amplified with two primers PCR_pool_2aa_F and PR_pool_ek_eGFP. For plasmids, containing single residues fused with reporter Ubi-eK-sfGFP-

PCR component	Volume for 50ul reaction
10x Taq buffer	10ul
Forward primer (PCR_pool_2aa_F) 10uM	1ul
Reverse primer (PR_pool_ek_eGFP) 10uM	1ul
dNTP 10mM	5ul
DNA template (1ug)	1ul
GoTaq Polymerase	1ul
Ultra-pure H20	31ul

Table 4. PCR mix for the amplification of inserts for cloning for plasmids pRNA069-1, pRNA072, pRNA077 and pRNA078.

Step	Temperature	Time	
Initial denaturation	94°C	30s	
Denaturation	94°C	30s	20 cycles
Annealing	58°C	30s	
Elongation	68°C	60s	
Final elongation	68°C	300s	
Pause	4°C	∞	

Figure 5. PCR cycling conditions for the amplification of inserts for cloning for plasmids pRNA069-1, pRNA072, pRNA077 and pRNA078.

P2A-mCherry, GoTaq Polymerase made by IMB Protein Production Core Facility was used. PCR mix was made as shown in Table 4. PCRs were run on Biometra TRIO Thermocycler (Analytikjena, 846-x-070-723) with conditions shown in Table 5. For visualization, 2% agarose gel was used and electrophoresis was run in Mini-Sub Cell GT Systems (BioRad, 1664000EDU) by PowerPac Basic

power supply (BioRad, 1645050). The results of those PCR are shown in Supplementary Figure 27A.

Component	Volume for 50ul reaction
Plasmid DNA (1ug)	3.5ul
rCutSmart Buffer	5ul
BspQI	1ul
Ultra-pure H20	40.5ul

Table 6. Restriction digestion with BspQI.

Backbone RNA107 was opened by restriction digestion with BspQI (New England Biolabs, 10201944). Digestion was done at 50°C for 1h (Table 6). DNA fragments were separated by 0.8% agarose gel electrophoresis and bands were purified by QIAquick Gel Extraction Kit (Qiagen, 28704) (Supplementary Figure 27B). Concentrations were checked by Nano Photometer NP80 (Fisher Scientific, 15485824).

Inserts were cloned into the backbone using Gibson Assembly Master Mix (New England Biolabs, M5510). For this, 10ul of the Gibson Assembly Master Mix was combined with 120ng of backbone, 20ng of insert, and H_2O up to 20ul. The mix was then incubated at 50°C for 1h.

50ul of chemically competent cells NEB 5-alpha Competent *E. coli* (New England Biolabs, C2987) were transformed with 2ul of a mix. After transfection cells were plated on agar plate with ampicillin and grow at 37°C overnight. Colonies were checked by sequencing with CMVfor.

Right colonies were incubated in 5ml of LB with ampicillin at 37°C overnight and plasmids were purified by QIAprep Spin Miniprep Kit (Qiagen, 27104). Plasmids were further used to produce lentiviruses.

Plasmids from pAC0039-1 to pAC0052-1

For plasmids, containing Ubi-XZ-linker-eGFP-P2A-mCherry where XZ represent particular dipeptides (plasmids from pAC0039-1 to pAC0052-1, see Table 1) DNA Velocity Polymerase

PCR component	Volume for 50ul reaction
5x HiFi buffer	10ul
Forward primer (PCR_pool_2aa_F) 10uM	1ul
Reverse primer (PCR_pool_2aa_R) 10uM	1ul
dNTP 10mM	5ul
DNA template (1ug)	1ul
DNA Velocity Polymerase	1ul
Ultra-pure H20	31ul

Table 7. PCR mix for the amplification of inserts for cloning for plasmids from pAC0039-1 to pAC0052-1.

(Meridian Biosences, BIO-21098) was used together with adequate 5xHiFi buffer. PCR mix was made as shown in Table 7. Thermocycle run with conditions shown in Table 8. Bends of desired sizes were cut out and DNA was purified by QIAquick Gel Extraction Kit (Supplementary Figure 27C and D).

The backbone was digested with BspQI like in the previous section (Table 6). Similarly, conditions for Gibson cloning, transformation in *E. coli* cells, sequencing, and DNA purification were done with the same protocols as described in the section "Plasmids pRNA069-1, pRNA072, pRNA077, and pRNA078".

Step	Temperature	Time	
Initial denaturation	98°C	120s	
Denaturation	98°C	30s	20 cycles
Annealing	58°C	30s	
Elongation	72°C	45s	
Final elongation	72°C	300s	
Pause	4°C	∞	

Table 8. PCR cycling conditions for the amplification of inserts for cloning for plasmids from pAC0039-1 to pAC0052-1.

Lentiviral production and transduction of cell lines Hek#3_G62-Hek#3_G70

To produce cell lines Hek#3_G62-Hek#3_G70 (that include cell lines Hek#3_G62, Hek#3_G63, Hek#3_G64, Hek#3_G65, Hek#3_G66, Hek#3_G67, Hek#3_G68, Hek#3_G69, Hek#3_G70), a total of nine distinct lentiviruses were required for transduction. To find out the best concentration of viruses to be used in that experiment I produced first only viruses for cell line Hek#3_G64 with plasmid pRNA065-1 and tested different concentrations of viruses. For producing lentiviruses, $2x10^6$ HEK-293T cells were plated in a 10cm plate and left to be attached to the plate overnight. After 24h, the medium was switched with a mix of complete DMEM with 25uM Chloroquine and incubated for 6h. Next, the medium was changed with DMEM and left for 1h to incubate. Upon that, cells were transfected with 3rd generation of lentiviruses. Transfection was done in the S2 laboratory at the Institute for Molecular Biology. For the transfection, I used Lipofectamin 3000 Transfection Reagent (Invitrogen L3000001). Two mixes were made in the procedure. Mix A contained 250ul of Opti-MEM Reduced Serum Medium and 5ul of Lipo3000 while mix B contained 250ul of Opti-MEM Reduced Serum Medium, 10ul of P300, 5ug of plasmid pMDLg/pRRE, 5ug of plasmid pRSV-Rev, 2.5ug of plasmid pMD2.G and 10ug of plasmid of interest, pRNA065-1. Mix A and mix B were combined (mixAB) and incubated for 15min. Media in 10cm plates was discarded and changed with 2ml of DMEM without FBS. Alter 15min of incubation mixAB was added to the cells. Cells were incubated at 37°C overnight and tomorrow medium was exchanged with 8ml of complete DMEM. After 48h, 8ml of medium with produced viruses was taken and centrifuged

1250rpm for 5min to pellet the particles other than lentiviruses. Supernatant was then taken and filtered through 0.45um filter and used fresh to infect Flp-In T-REx 293 cells.

Day before infection 0.2 x10^6 Flp-In T-REx 293 cells were plated in 6-well plates. After 24h, the medium in which they were growing was thrashed and three different volumes of freshly produced lentiviruses were added (0.5ml, 1ml, or 2.5ml) with complete DMEM with Polybrene (1:800). After another 24h, media with lentiviruses was discarded and cells were moved to 10cm plates to for further grow. The next day media was replaced with media with puromycin (1:1000) to select for infected cells. After one week amount of survived cells was counted by TC20 Automated Cell Counter. Cells infected with 0.5ml of viruses had 0.1x10^6 live cells, infected with 1ml of viruses had 10x10^6 live cells, with 2.5ml had 22x10^6 live cells, and cells that were treated in the same way but without lentiviral infection had 38x10^6 of live cells. I have chosen for my next experiments to further use 1m of viruses that gave 26% of cell survival.

I repeated the same protocol for lentiviral production to make the following cell lines: Hek#3_G62, Hek#3_G63, Hek#3_G64, Hek#3_G65, Hek#3_G66, Hek#3_G67, Hek#3_G68, Hek#3_G69, Hek#3_G70. For the production I used following plasmids: pRNA063, pRNA064, pRNA065-1, pRNA066, pRNA067, pRNA068, pRNA069, pRNA070, pRNA071, pRNA072, pRNA077, pRNA078. Flp-In T-REx 293 cells were infected by 1ml of fresh lentiviruses.

Lentiviral production and transduction for all other cell lines in this thesis

To make cell lines containing dipeptides I used a slightly different protocol than the one described for cell lines Hek#3_G62-Hek#3_G70. For all other cell lines, reporter Ubi-linker-eGFP-P2A-mCherry was used in combination with different dipeptides located between the Ubi moiety and the linker. To produce lentiviruses, HEK-293FT (Thermo Fisher Scientific, R70007) cells were co-transfected with three plasmids psPAX (envelope), MD2G (packaging), and plasmid with a construct of interest. Even HEK-293T cells were employed for the generation of viral vectors containing single residues fused to a reporter describer in the previous part. Cell line HEK-293FT was used for all further experiments due to better efficiency in viral production (Naldini et al. 1996). Three plasmids were co-transfected by Lipofectamine 3000 Transfection Reagent in the S2 laboratory.

The day before transformation, 0.5x10^6 Flp-In T-REx 293 cells or ATE1 KO cells were plated in 6-well plates and left to attach to the bottom of the plate. The next day, 375ul of Opti-MEM Reduced Serum Medium (Gibco, 11058021) and 10.25ul of Lipo3000 were combined in a mix A. Next, 375ul of Opti-MEM Reduced Serum Medium, 8.75ul of P3000 reagent, 1750ng of psPAX plasmid, 750ng of pMD2, and 2500ng of plasmid of interest were mixed in a mix B. Mix A and B were combined (mix AB) and incubated 15min at the room temperature. Medium from the plates was removed and mixAB together with 1.5ml of the new medium was added to the plates, which were then incubated for 6h. After precisely 6h, the medium was removed and 8ml of LPM media

was added (2.5ml of FBS, 100ul of 0.2% pyruvate, and Opti-MEM Glutamax up to a total volume of 50ml). After 48h medium was taken and lentiviruses were concentrated by Lenti-X Concentrator (Takara Bio, PT4421-2). Medium with lentiviruses was first filtered through a 0.45um filter to remove all bigger particles and then mixed in the ratio 3:1 with Lenti-X Concentrator. The mix was incubated overnight at 4°C and further centrifuged at 1.500xg for 45 minutes at 4°C. The supernatant was removed and the pellet with lentiviruses was resuspended with complete DMEM and aliquoted into tubes.

The day before infection, 0.3 $x10^6$ cells were plated in 6-well plates. Next day 2.5ul of fresh lentiviruses were combined with complete DMEM with Polybrene (1:800) up to 6ml. The next day, media with viruses was removed from the plate and put in 10cm plates to grow. In 24h media was changed to a media with puromycin (1:1000). After a week, cells were confirmed by qPCR that they were clean from viruses and ready to be taken from the S2 to S1. Further, the GFP and RFP ratio was measured by Flow Cytometry.

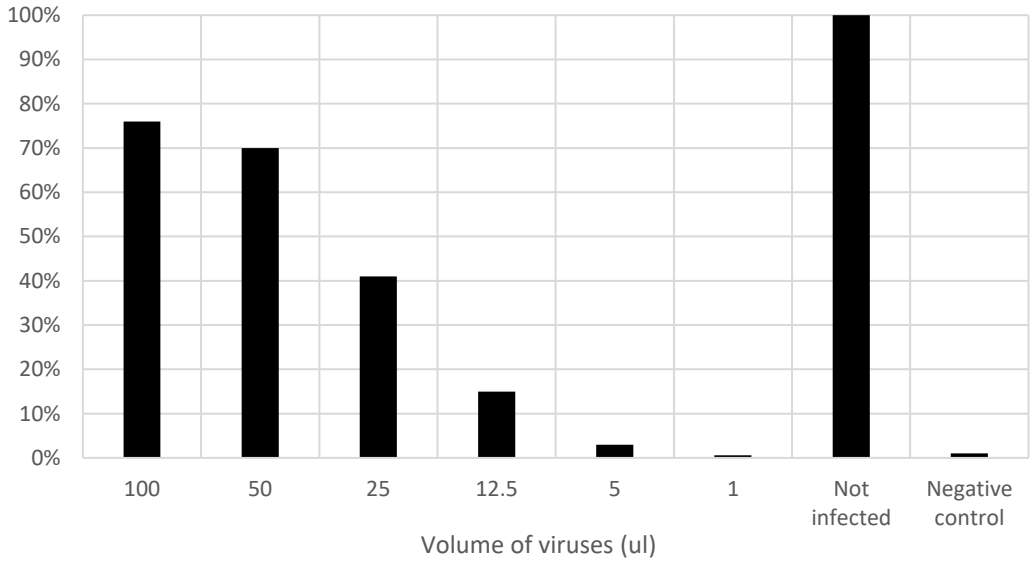


Figure 7. Percentage of live cells 3 days after infection with different volumes of lentiviruses. Y-axis represent different volumes of viruses used for infection of Flp-In T-REx 293 cells. Not infected cells were treated in the same way like infected cells but without infection with viruses. Negative control are cells that are not infected but also not treated with puromycin.

Flow Cytometry

Flow cytometry was done for all cell lines in the same way. Before measurement by Flow Cytometry cells were trypsinized and resuspended in Sorting Buffer (200ul of FBS, 1ml of 2mM EDTA, PBS up to 10ml). Cells from the 1.5ml tubes were transferred in Flow Cytometry tubes and DAPI (1:1000) was added to distinguish dead and live cells. Measurements were performed on a BD LSRFortessa Cell Analyzer (BD Biosciences) and acquisition software BD FACSDiva v9.0.1 (Supplementary Figure 28 and Figure 29). Fluorescence intensities of each single cell intensities were measured using a 488nm laser for green and a 561nm laser for the red signal. Cell size and complexity were determined by forward and side scatter (FSC and SSC). A minimum of 10,000 events were analyzed for each cell line and the resulting data was analyzed by the software Flow Jo and then further with R.

Analysis of cell lines Hek#3_G62-Hek#3_G70

The stability of 12 constructs that differed only in the N-terminal residue fuses with the reporter Ubi-X-eK-sfGFP-P2A-mCherry was measured intwo separate events. First, measured stability of nine constructs R, Y, D, H, V, T, E, Q, M (). To measure the stability of the remaining constructs (G, D, and K), three samples were taken from the first measurement as a control. These samples included V (with the lowest stability), N (with a medium range of stability), and E (with the highest stability). Protein stability normalization was conducted per residue by dividing it with the median of the average of residues V, N, and E. All measurements were conducted in triplicate, except for the controls in the second measurement. Then I combined both experiments together.

Generating of Ubi-XZ library

To produce the Ubi-XZ library expressed in Flp-In T-REx 293 cells, I use a plasmid library (named Ubi-XZ library) made by [redacted] to produce lentiviruses. To produce lentiviruses containing the Ubi-XZ library pool, 2x106 HEK-293FT cells were plated in a 10cm plate and left overnight. The following day, transformation was performed. Mix A was prepared by combining 1.5ml of Opti-MEM Reduced Serum Medium with 41ul of Lipo3000. Simultaneously, mix B was created by combining 1.5ml of Opti-MEM Reduced Serum Medium, 35ul of P3000 reagent, 7ug of psPAX plasmid, 3ug of pMD2, and 10ug of library plasmid. MixAB was then formed by combining mix A and mix B, followed by an incubation period of 15 minutes at room temperature. The medium from the plates was subsequently replaced with mixAB and 6ml of fresh medium, and the plates were incubated for 6 hours. After precisely 6 hours, the medium was removed, and 8ml of LPM media was added. After 48 hours, the lentiviruses were harvested and concentrated using a Lenti-X Concentrator.

To find the best multiplicity of infection (MOI) for viruses I compared the percent of live cells after puromycin selection between cells infected with five different volumes of lentiviruses. I have chosen to infect cells with 1ul, 5ul, 12.5ul, 25ul, 50ul and 100ul of viruses (Figure 7). After counting live cells, I chose a volume of 12.5ul of lentiviruses to use for further infections with this library since MOI should be 0.15.

To produce the Ubi-XZ library in triplicates, $3x10^6$ HEK-293FT cells were plated in three 15cm plates. After 24h, each plate was infected with 12.5ul of fresh lentiviruses with complete DMEM containing Polybrene (1:800) to a final volume of 15ml. The next day, the medium containing viruses was extracted from the plate and splited into two new 15cm plates so that the cell would have space to grow. After 24 hours, the medium was replaced with a puromycin-containing medium (1:1000). Next, cells were ready for profiling using the Global Protein Stability (GPS)–peptidome technology (Figure 14) (Yen et al. 2008a).

Sorting

To prepare cells for sorting, cells grown in 15cm plates were washed twice with prewarmed DPBS. Trypsin was then added to detach them and after 5min of incubation, complete DMEM was added to deactivate trypsin. Cells were pelleted by centrifugation and the supernatant was discarded and replaced with the same amount of Soybean Trypsin Inhibitor (0.05% in DPBS). Next cells were resuspended in 300ul of sorting buffer. Sorting was done by the IMB Flow Cytometry Core Facility on a BD FACSAria III SORP. Cells were sorted into 8 bins depending on their GFP to RFP ratio. In each pin, 100 000 cells were sorted (Nozzle Size: 100 µm (20 psi), ND filter: 2.0, Sort mode: 4-Way Purity). Gates from 2 to 7 are the same sizes, while gates 1 and 8 grou the rest of the cells that have or very low or very high stability ratio (Supplementary Figure 30 and Figure 31). Cells were sorted into eight 1.5ml tubes, directly transferred into 12-well plates and grown to around 8 $x10^6$ cells.

Barcoding for NGS

To be able to distinguish sequencing results between each of the 8 bins and each of the three replicates, two types of NGS barcodes were used. Cells for each of the 8 bins and all three replicates are trypsinized and pelleted. DNA from each of them was extracted and used for the PCRs. Conditions for PCRI are shown in Table 9. Thermocycle conditions are the same as shown in Table 8, except 24 cycles were used instead of 20.

Each PCR reaction was run on 2% agarose gel, and bands with the right size were cut out and purified by QIAquick Gel Extraction Kit. Next DNA of 8 bins were combined together depending on the percentage of cells during cell sorting in each gate. These pools are used to do a second PCR and to add barcodes for each or three replicates. All three replicates had the same forward primer P5_Scriptseq, but different reverse primers. Replicate 1 had reverse primer P7-ix14_Scriptseq, replicate 2 had P7-ix15_Scriptseq, and replicate 3 had P7-ix18_Scriptseq. All other conditions were the same as shown in Table 10. Thermocycler conditions are the same as in Table 8 except only 7 cycles were used.

Bin	Mix NEB Q5 Ultra (ul)	Mix of forward primers 10uM	Mix of reverse primers 10uM	DNA (1000ng)	H2O
Bin 1	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin1_R +seq20_PCR1_bin1_R2+ seq20_PCR1_bin1_R3+s eq20_PCR1_bin1_R4) 1ul	5.9ul	17.1ul
Bin 2	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin2_R +seq20_PCR1_bin2_R2+ seq20_PCR1_bin2_R3+s eq20_PCR1_bin2_R4) 1ul	4.5ul	18.5ul
Bin 3	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin3_R +seq20_PCR1_bin3_R2+ seq20_PCR1_bin3_R3+s eq20_PCR1_bin3_R4) 1ul	4.1ul	18.9ul
Bin 4	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin4_R +seq20_PCR1_bin4_R2+ seq20_PCR1_bin4_R3+s eq20_PCR1_bin4_R4) 1ul	5.2ul	17.8ul
Bin 5	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin5_R +seq20_PCR1_bin5_R2+ seq20_PCR1_bin5_R3+s eq20_PCR1_bin5_R4) 1ul	10ul	13ul
Bin 6	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin6_R +seq20_PCR1_bin6_R2+ seq20_PCR1_bin6_R3+s eq20_PCR1_bin6_R4) 1ul	13.7ul	9.3ul
Bin 7	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin7_R +seq20_PCR1_bin7_R2+ seq20_PCR1_bin7_R3+s eq20_PCR1_bin7_R4) 1ul	6.7ul	16.3ul
Bin 8	25ul	(seq20_PCR1_F+ seq20_PCR1_F2+s eq20_PCR1_F3+se q20_PCR1_F4) 1ul	(seq20_PCR1_bin8_R +seq20_PCR1_bin8_R2+ seq20_PCR1_bin8_R3+s eq20_PCR1_bin8_R4) 1ul	7.3ul	15.7ul

Table 9. PCR mix for barcoding cells for the 8 different bins. This is just an example of one of three replicates. Cells for each bin had a different mix of reverse primers.

PCR component	Volume for 50ul reaction
5x HiFi buffer	10ul
Forward primer (P5_Scriptseq) 10uM	1.3ul
Reverse primer (P7-ix14_Scriptseq or P7-ix15_Scriptseq or P7-ix18_Scriptseq) 10uM	1.3ul
dNTP 10mM	5ul
DNA template (200ng)	16.2ul
DNA Velocity Polymerase	1ul
Ultra-pure H20	15.2ul

Table 10. PCR mix for barcoding cells from three different replicates. All three replicates have the same conditions except different reverse primers.

Reactions of three replicates for the second PCR were run on 2% agarose gel and DNA was purified by QIAquick Gel Extraction Kit. Quality of DNA was checked by Qubit Fluorometric Quantification machine and Agilent 2100 Bioanalyzer system. Samples were sent to IMB Genomic Core Facility for NGS, and analyzed by [redacted] and [redacted].

Generation of ATE1 knockout

To generate ATE1 knockout cell line, CRISPR/Cas9 was used and this experiment was done in collaboration with [redacted]. Two sgRNA were used to make two cuts, the first sgRNA was Ate1_e2_t1_F (caccgTCAGGATCTCATAGACCG) located at the beginning of exon 2, and second Ate1_e4_t2_F (caccGGCTGAAATTGTAAAGGT) located at the beginning of exon 4 (Figure 21A). Guides were cloned into two different backbones containing GPF or mCherry. Both plasmids were transfected into Flp-In T-REx 293 cell line by Lipofectamin 3000 Transfection Reagent (Invitrogen L3000001). Cells that contain both GFP and mCherry signals were sorted into 96-well plates by one.

ATE1 knockout was immunoblotted like describer in the “Immunoblotting” part of this thesis. Besides that to validate knockout particular dipeptides fused to reporter Ubi-linker-eGFP-P2A-mCherry were expressed in this cell line described in the part “Single constructs”.

Immunoblotting

Samples for Mass spectrophotometry we pulled down by magnetic agarose GFP bears and then samples were run on SDS-PAGE. Firstly, protein concentrations were measured by Braford. Next, the samples were incubated at 95°C for 10 min with SDS buffer. Upon that, proteins were separated by SDS-PAGE using 12% gels prepared with TGX FastCast acrylamide kit (Bio-Rad,

1610175). Transfer was done with Trans-Blot Turbo System to the Mini 0.2 µm Nitrocellulose Transfer membrane (Bio-Rad, 1704158). The membrane was subsequently blocked in 5% nonfat dry milk in PBS with 0.1% Tween-20 (PBS-T). Primary antibody incubation took place overnight at 4°C, followed by three washes with PBS-T. HRP-conjugated secondary antibody was added and incubated for 1 hour at room temperature. After another three washes in PBS-T, reactive bands were detected using the ChemiDoc System.

The primary antibody for detecting Mass spectrophotometry samples was anti-GFP (Merck, 11814460001), and secondary Goat anti-Mouse IgG (H+L) Cross-absorbed (Thermo Fisher Scientific, G21040). For the validation, the ATE1 knockout primary body was anti-ATE1 (Sigma, MABS436), and the secondary was Goat anti-rat IgG (H+L) (Thermo Fisher Scientific, 31470).

Mass spectrometry analysis of N-terminal peptides

Cell lines used for mass spectrometry were prepared as described in the section "Lentiviral production and transduction for all other cell lines in this thesis". In total 13 cell lines were prepared and measured. To prepare cells for measurement pull-down with GFP antibody was performed. Cell lines are grouped in two different pools depending on the dipeptide expressed in them. All dipeptides starting with residue N or Q were part of pool 1 and all dipeptides starting with residues E or D were part of pool 2. For pull-down magnetic agarose GFP binder beads from IMB Protein Production Core Facility were used. To estimate the quality of the pull-down samples from different parts of the protocol were run on SDS gel (Supplementary Figure 37). Samples were run on Mass Spectrometry by [redacted].

After the pull-down experiment, proteins were released from the beads by heating at 60°C with orbital shaking in elution buffer (25 mM Tris-HCl pH 8.0, 1%[w/v] SDS) for 15 minutes. Subsequently, proteins underwent DTT reduction, iodoacetamide alkylation, and quenching by DTT, followed by purification using the SP3 method (Hughes et al. 2019). The purified proteins were then eluted twice in 50 µl of 3 M guanidinium chloride, 250 mM MOPS pH 7.9, at 37°C with orbital shaking for 10 minutes. In vitro protein acetylation commenced by adding D6-acetic anhydride (Sigma-Aldrich) to 50 mM, followed by a 30-minute incubation at 37°C with orbital shaking. The process was repeated once, and residual D6-acetic anhydride was neutralized by ammonium bicarbonate buffer. Following a 10-minute incubation at 37°C with orbital shaking, proteins underwent chymotrypsin digestion (1 µg per sample) overnight at 37°C. The resulting peptide solution was acidified with formic acid and purified through solid-phase extraction using C18 StageTips (Empore, 3M Company) (Rappsilber, Ishihama, and Mann 2003)).

Peptide analysis employed an Orbitrap Exploris 480 mass spectrometer (Thermo Fisher Scientific, BRE725533) coupled to an EASY-nLC 1200 UHPLC system (Thermo Fisher Scientific, LC140). Peptides were separated in a 60-cm analytical column (inner diameter: 75 µm, ReproSil-Pur 120 C18-AQ 1.9-µm silica particles, Dr. Maisch GmbH) using online reversed-phase chromatography

with a 90-minute gradient of 2.4-32% acetonitrile and 0.1% formic acid at a nanoflow rate of 250 nl/min. Mass spectrometry operated in data-dependent acquisition mode, with a full scan (resolution: 60,000, scan range: 300-1650 m/z) followed by 15 fragmentation scans via higher energy collision dissociation (HCD; normalized collision energy 30%, resolution: 15,000). Data processing utilized MaxQuant software (version 2.1.3.0) (Cox and Mann 2008) and the Andromeda search engine (Cox et al. 2011) against a target-decoy database comprising UniProt *Homo sapiens* reference proteome, mCherry, reporter protein variants, and common contaminants. The false discovery rate (FDR) was set to 1% at both peptide and protein levels.

To compare protein levels with or without in vivo acetylation at the N-terminus of the reporter protein, N-terminal peptide sequences were extracted from the MaxQuant output file. Peptides exhibiting N-terminal wild-type acetylation were categorized as in vivo acetylation, while those showing N-terminal D3-acetylation or no acetylation were considered without in vivo acetylation. Peptide intensities were then summed separately for each category. The final data was analyzed in R.

Yeast genome manipulation

This part was done in collaboration with [redacted]

Construction of a Ubi-XZ library

The library contains 80 dipeptides where first residue can be D, E, N or Q while second residue can be any of existing amino acids. Dipeptides were fused with tFT described in part "Selecting the optimal approach for measuring protein turnover". In order to make all of those variants, four inserts containing D, E, N or Q with any dipeptide on the second position were ordered with flanking regions. Plasmid pAnB19 (Kats et al. 2018) was linearized by restriction digestion with EcoRV. Conditions were like in Table 6, except incubation was done at 37°C for 1h.

Inserts were cloned into the backbone using Gibson Assembly Master Mix (New England Biolabs, M5510). A mixture containing 10µl of the Master Mix, 120 ng of backbone, 20 ng of insert, and H2O up to 20µl was incubated at 50°C for 1 hour. In total 96 colonies were sent for sequencing from which we got 60 colonies with the right and additional 20 were made manually with oligos with a * in Table 3.

Construction of plasmids with NTA1, NTAN1 and NTAQ1

To linearize the backbone pRS426-GPD was digested using BamHI and HindIII restriction digestion enzymes. The yeast scNTA1 gene was amplified from yMaM330 gDNA using the primer sp426GPD_BamHI_yNTA1_F and p426GPD_HindIII_yNTA_R. Next, NTAN1 was ordered as a

gBlock from IDT and NTAQ1 was amplified from pAC009 with primers 426GPD_BamHI_NTAQ1_F and 426GPD_HindIII_NTAQ1_R. PCR conditions are described in the Table 11 and Table 12.

Yeast transformation and crossing

Transformation in yeast was done like described in Gietz & Woods, 2002. To generate knockout of scNTA1 and human ATE1 like described in Janke et al., 2004. exept PCR conditions were like in with primers NTA1-S1 and NTA1-S2 or ATE1-S1 and ATE1-S2. PCR mix was done as in Table 11 and Table 12, with a difference primers and PCR cycles were 34 instead 40.

PCR component	Volume for 50ul reaction
10x HiFi buffer	5ul
Forward primer 10uM	3ul
Reverse primer 10uM	3ul
dNTP 10mM	1ul
DNA template (100ng)	0.5ul
DNA HF Polymerase	0.5ul
Ultra-pure H20	37ul

Table 11. PCR mix for amplifying NTAQ1 as an inters for cloning

Step	Temperature	Time	
Initial denaturation	95°C	5min	
Denaturation	95°C	30s	40 cycles
Annealing	68°C	30s	
Elongation	72°C	60s	
Final elongation	72°C	7min	
Pause	25°C	5min	

Table 12. PCR cycling conditions for the amplification

Strains	Wild type	Background	Reference	Mutant
BY4741(type a)	nta1Δ::KanMX ate1Δ::hphNT +pAnB19-XZ	+p413-GPD-CYC (yKBJ0052-1)	+pAnB19-TK (yCPO0119)	nta1Δ::KanMX ate1Δ::hphNT +pAnB19-XZ
BY4742(type alpha)	+p426-GPD	+p426-GPD	+p426-GPD	nta1Δ + p426-GPD-NTA variants or ate1Δ + p426-GPD-ATE variants

Table 13. Strains used for crossing. All type a strains were crossed with type alpha

The strains BY4741 (type a) and BY4742 (type alpha) were crossed like in the Table 13. Measuring of sfGFP and mCherry was done by Tecan.

Proteomic peptide-phage display ProP-PD

Phage N-terminal library

N-terminal phage display library was based on N-terminus of all human proteins. Library includes all isoforms of all human proteins with and without initiator methionine together with cleavage sites and microbial infection cleavage sites. Human protein-coding sequences were downloaded from the UniProt database, in total 82 680 unique oligos encoding 39 545 different unique peptides. Depending on presence of initiator methionine sequences were 12 or 11 amino acids long and flanked by primer annealing sites complementary with forward primer For_NewP8_IntLibAmp (CTATTGCTACAAATGCCTATGCA) and reverse primer IntLibAmpRev (GGTGGAGGATCCGGAGG). Custom-designed library was ordered from a commercial provider.

To amplify the oligopool PCR mix was made like in the Table 14. Two step PCR cycling conditions for the amplification of oligopool for N-terminal library. Thermocycle conditions were like in the Table 15.

PCR component	Volume for 50ul reaction
Phusion High-Fidelity PCR Master Mix	25ul
Forward primer (For_NewP8_IntLibAmp) 10uM	2ul
Reverse primer (IntLibAmpRev) 10uM	2ul
DNA template (1ug)	0.4ul
Ultra-pure H20	20.6ul

Table 14. Two step PCR mix for the amplification of oligopool for N-terminal library

Step	Temperature	Time	
Initial denaturation	98°C	30s	
Denaturation	98°C	10s	7 cycles
Annealing	52°C	15s	
Elongation	72°C	10s	
Denaturation	98°C	10s	13 cycles
Annealing	55.5°C	15s	
Elongation	72°C	10s	
Final elongation	72°C	10min	
Pause	4°C	∞	

Table 15. Two step PCR cycling conditions for the amplification of oligopool for N-terminal library

After amplification, electrophoresis for four reactions was done on 2% agarose gel. Bands were purified by QIAquick nucleotide removal kit (Qiagen, 28306) to remove enzymes and salts. Next PicoGreen dsDNA assay kit was used to Quantify the amount of PCR product. The oligonucleotide library is then linked to the primary coat protein P8 of the filamentous M13 phage through a variation of Kunkel mutagenesis (McLaughlin and Sidhu 2013). For this phagemid was introduced into *E. coli* CJ236 cells lacking functional dUTPase and uracil-N-glycosylase. Next, the bacteria containing the phagemid infected with M13K07 helper phage, inducing the packaging of single-stranded phagemid DNA (ssDNA) into phage particles, with uracil substituted for thymine, which are then released into the medium. The resulting dU-ssDNA is isolated from the medium and serves as a template for synthesizing heteroduplex, double-stranded DNA (dsDNA) using the PCR-amplified and phosphorylated oligonucleotide library. This dsDNA library is subsequently introduced into *E. coli* SS320 cells via electroporation, following preinfection with M13 K07. After an overnight incubation, the bacteria generate and amplify the phage library, which can be harvested thereafter.

Recombinant protein production

In order to produce following proteins GST-Gid4(1-358), GST-Gid10(1-292), GST-Gid4(99-362), GST-Gid10(115-292), GST-NTAQ1, GST-ATE1_iso1 (isoform 1), GST-ATE1_iso2 (isoform 2), GST-NTAN1 and NTAN1-GST, firstly plasmids needed to be made. For the proteins having GST of the N-terminus, plasmid pAC001 was digested with XhoI restriction digestion enzyme at 37°C for 1h (Table 16). For all other plasmids pPPCF118-1 was PCRed in two step PCR with forward primer Y.001_pPPCF118_240820_F and reverse Y.001_pPPCF118_240820_R primers except for the plasmid to produce GST-Gid4(1-358) for which reverse primer was pAC0013_Gib_Gid4_358aa_R. PCR was done with HiFi Polymerase produced by IMB Protein Production Core Facility (Table 17).

Component	Volume for 50ul reaction
Plasmid DNA (1ug)	17.8ul
rCutSmart Buffer	5ul
XhoI	1ul
Ultra-pure H20	26.2ul

Table 16. Restriction digestion with XhoI.

PCR component	Volume for 50ul reaction
10x HiFi buffer	5ul
Forward primer (Y.001_pPPCF118_240810_F) 10uM	2.5ul
Reverse primer (Y.001_pPPCF118_240810_R) 10uM	2.5ul
dNTP 10mM	1ul
DNA template (100ng)	0.62ul
DNA HiFi Polymerase	0.5ul
DMSO	2.5ul
Ultra-pure H20	35.38ul

Table 17. PCR mix for amplification of oligo pool for N-terminal library

PCR steps included the one in Table 18.

Step	Temperature	Time	
Initial denaturation	98°C	30s	
Denaturation	98°C	10s	10 cycles
Annealing	57°C	10s	
Elongation	72°C	3min	
Denaturation	98°C	10s	30 cycles
Annealing	65°C	30s	
Elongation	72°C	4min	
Final elongation	72°C	5min	
Pause	4°C	∞	

Table 18. Two step PCR cycling conditions for the amplification of backbone for plasmids used for protein purification.

PCR for inserts for all the cloning was done with HiFi polymerase and had the same conditions as in Table 17 except DNA templates were adjusted depending on the cloning. Oligos that were used as DNA templates can be found in Table 1. Thermocycle steps were described in Table 8 are except 30 cycles were used and annealing temperature was 57°C.

Gibson cloning, transformation and purification were done as described in part "Plasmids pRNA069-1, pRNA072, pRNA077 and pRNA078". Plasmids were submitted to the IMB Protein Production Core Facility for protein purification.

The last two proteins GST-ZER1(519-766) and GST-ZYG11B(480-728) were purified by me with a protocol as described in Ali et al., 2020 and plasmids provided by [redacted].

Phage selection

To select two phage libraries, a new N-terminal library, and an HD2 library, against proteins of interest 96-wells of Maxisorp microtiter plates (NUNC) (Merck, P6366) were coated overnight with proteins (10ug/ul in 100ul of PBS per well). To eliminate unspecific binders parallel plates were coated with GST alone (5ul/ml). Plates were covered with a non-permeable plastic seal and incubated at 4°C overnight with shaking. In parallel Omnimax culture was grown in 2YT with tetracycline overnight at 37°C. For both libraries, four rounds of selection were done and each protein selection was done in triplicate. Tomorrow plates were blocked with 200ul of blocking buffer (PBS, 0.2% BSA) for 1h at 4°C. To prepare the phage library for selection, the appropriate volume of the phage library was diluted 20 times with PBS, and PEG/NaCl (4% PEG-800 (wt/vol) and 0.5 M NaCl) was added to 1/5 of the final volume. For the first round of selection, 100ul of phage library pool was first added to plates coated with GST to eliminate unspecific binding and then after 1h of incubation at 4°C transferred to the plates coated with proteins of interest and incubated for the next 2h. Next plates were washed with cold wash buffer (PBS, 0.5% Tween-20). Phages were then eluted by adding 100ul of *E. coli* SS230 in log-phase and with 2YT to each well. To be able to produce phages, helper phages M13K07 in a final concentration of 10^{10} phage per milliliter were added in each well. After incubation at 37% 45min cultures were transferred to 96-well deep-well blocks with 1ml of 2YT supplemented with kanamycin (25 µg/mL), carbenicillin (100 µg/mL), and isfopropyl-β-D-thiogalactopyranoside (IPTG; 0.4 mM), and shaken overnight at 37 °C. The next day, blocks were centrifuged at 2000xg at 10°C for 10min to remove bacteria. The supernatant was then transferred to a new sterile block, and bacteria were inactivated by heat at 65°C for 20min. From each well 100ul was used for the next round of selection. The next day, plates were again coated with GST or proteins of interest and the whole protocol was repeated. Selections were conducted for four rounds, and the progress was monitored by analyzing aliquots of phage supernatants in a phage ELISA. The goal of ELISA was to detect if the phages were bound to proteins of interest.

After results from ELISA, enriched proteins were selected for NGS. First DNA region on P8 proteins were amplified by unique barcodes previously described by McLaughlin and Sidhu (McLaughlin and Sidhu 2013). Undiluted amplified phage pools (5µL) were PCRed with Phusion High Fidelity DNA polymerase (New England Biolabs, M0530S), with the conditions shown in Table 19 and Table 20. The resulting PCR products were validated through gel electrophoresis (2% agarose gel) of 1µL of PCR products.

PCR component	Volume for 50ul reaction
Phusion High Fidelity DNA polymerase	25ul
Forward primer 0.5uM	5ul
Reverse primer 0.5uM	5ul
DNA template- Phage pool	5ul
Ultra-pure H20	10ul

Table 19. PCR mix for amplification of phage library with particular barcodes for NGS

Step	Temperature	Time	
Initial denaturation	98°C	30s	
Denaturation	98°C	10s	10 cycles
Annealing	57°C	10s	
Elongation	72°C	3min	
Denaturation	98°C	10s	30 cycles
Annealing	65°C	30s	
Elongation	72°C	4min	
Final elongation	72°C	5min	
Pause	4°C	∞	

Table 20. PCR cycling conditions for the amplification of phage library with particular barcodes for NGS

The resulting amplicons were then normalized Mag-Bind magnetic beads and pooled together and concentrated using two columns of a QIAquick PCR purification kit. Pooled amplicons were run at 2% agarose gel and purified with QIA-quick gel extraction kit with extended incubation at the room temperature. Next DNA was eluted with 30ul of TE buffer and concentration was estimated by PicoGreen dye (Invitrogen, 10545213). Samples were then send to Cofactor Genomics (Saint Louis, MO) to NGS. Analysis were done by [redacted] and [redacted].

Results

Selecting the optimal approach for measuring protein turnover

There are a few methods, which can be used to measure protein turnover. While in vitro we can use pulse-chase metabolic labeling followed by immunoprecipitation or mass spectrometry, for experiments in vivo we need different methods (Pratt et al. 2002). In vivo we can determine protein turnover by pulse-labeling fluorescent-protein fusion proteins, using either photo-activation or photobleaching (Hopper et al. 2011). Besides that, tandem fluorescent protein timers (tFTs) are used to measure protein turnover and localization. This timer contains two fluorescent proteins with different maturation kinetics in which the ratio indicates the age of the proteins in a pool. The monomeric mCherry, which matures with a half-life of a ~ 40 min represent slow maturing protein while monomeric sfGFP (superfolder GFP) that matters in few minutes, represent fast maturing protein. Initially, a pool of mCherry-sfGFP molecules exhibits mainly green fluorescence upon synthesis. Over time, they transition to red fluorescence, making the ratio of red to green fluorescence indicative of protein pool age (Khmelinskii et al. 2012a; Khmelinskii and Knop 2014). Although tFT is commonly used in yeast, due to the longer half-life of human proteins this is not the most fitting method for measuring protein degradation in human

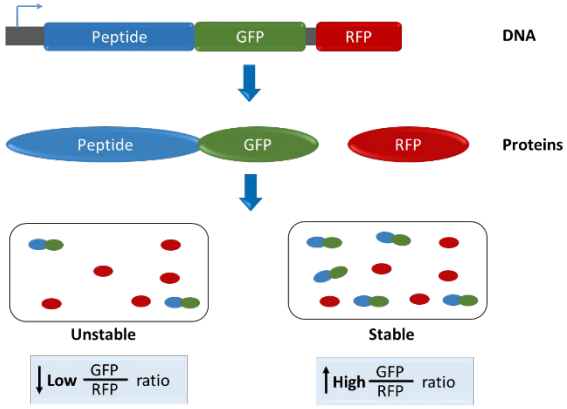


Figure 8. Cartoon of the principles behind ratiometric reporters of protein turnover. After translation, the peptide of interest is fused with GFP, while RFP is an internal control. If the peptide is targeted for degradation, GFP will be degraded as well but levels of RFP will remain constant. Therefore, lower GFP/RFP ratio level indicate that that peptide is more unstable compared to the one with higher ratio.

cells. This problem is overcome by separating two fluorescent proteins in the technique called global protein stability (GPS) profiling (Yen et al. 2008b). The reporter used in GPS contains a green florescent protein (GFP) alongside a separated red fluorescent protein (RFP) that serves as internal reference (Figure 8). The two fluorescent proteins (GFP and RFP) are expressed from the same transcript but make two separated proteins. The ratio between green and red fluorescent protein is directly proportional to the stability and degradation rate of a peptide or a protein fused to this reporter.

In order to find the optimal reporter that will give the most output about protein degradation, in collaboration with [redacted], I examined five different reporters each composed

of the green fluorescent protein in combination with the red fluorescent proteins, separated by an internal ribosome entry site (IRES) or a P2A self-cleaving peptide sequence. In GPS, two fluorescent proteins are produced from a single mRNA strand using an internal ribosome entry site (IRES). The IRES has a function of an additional ribosomal recruitment site allowing translation to occur in the internal part of mRNA. This can cause downstream protein to be expressed at a

Figure 9. Schematic illustration of five different types of reporters to determine the effectiveness of detecting the smallest differences in stability, with the objective of identifying the most sensitive variant. Ubi represents ubiquitin moiety, X represents one of the 12 amino acids, ek and ATSALGT represent two different linkers, sfGFP and eGFP represent two different green fluorescent proteins, P2A and IRES are those different sequences to separate proteins, mCherry and dsRed are two different red fluorescent proteins.

much lower rate than upstream protein, which can potentially cause the problem in measuring protein stabilities. On the other hand, P2A self-cleaving peptide sequence causes the ribosome to skip making peptide bound at the end of the P2A sequence and to produce two proteins from the same transcript (Kats et al. 2018; Kim et al. 2011).

Besides testing IRES and P2A sequence we also wanted to test the green fluorescent proteins, sfGFP, and eGFP, in combination with the red fluorescent proteins, mCherry, and DsRed. The (sfGFP) is a fluorescent protein that is used together with mCherry in the tFT reporter in yeast (Khmelinskii et al. 2012b). sfGFP is made to be much brighter and it matures rapidly compared to wild type GFP (Ilagan et al. 2010; Pédelacq et al. 2005; Zhang, Gurtu, and Kain 1996). However, sfGFP is observed to sometimes not be completely degraded by the proteasome, which prompted us to test eGFP which was previously used in studies with GPS in human cells (Yen et al. 2008b). Between the residues and reporters one of two of linkers were inserted. We wanted to test the ek sequence (extension [e] containing lysine [K]) that is a 36-residue long, unstructured sequence potentially serving as an initiation site for proteasome degradation. It contains two lysines K-15 and K-17 that that can be ubiquitinated (Bachmair and Varshavsky 1989; Prakash et al. 2004; Suzuki and Varshavsky 1999). The second linker we wanted to test was a short linker sequence (ATSALGT) published by Elledge lab (Timms et al. 2019). The focus of their study was to test N-terminal degrons in which they found a new Gly/N-degron pathway. They used the first 23 or 24 amino acids (with and without first methionine) of all human proteins to make a library. In the same study, they used DsRed instead of mCherry, so we wanted to test this RFP as well.

In order to test all the components of a reporter, we separately fused twelve different amino acids (Y, R, D, K, H, G, N, T, E, V, Q, or M) with each of our five chosen reporters. Furthermore, these 12 amino acids are denoted as "X" for the sake of brevity. We chose residues based on their previous reported stability in yeast, and we chose to make a range from lower to higher stability

(Kats et al. 2018). From here on the combinations of different amino acids and reporters are referred to as constructs.

To produce peptides that can start with a residue other than methionine, we used the ubiquitin fusion technique. Each construct was located just after the ubiquitin gene, and since during expression endogenous deubiquitin enzyme cleaves ubiquitin moiety at the N-terminus, the peptide is left exposed (Varshavsky 2005).

Each construct was transduced into the genome of the human embryonic kidney 293 cells (Flp-In T-REx 293 cell line) using lentiviruses. Stability was measured by Flow Cytometry and analyzed using FlowJo and R.

The stability of twelve distinct constructs, characterized solely by variations in the N-terminal residue fused to the reporter Ubi-X-eK-sfGFP-P2A-mCherry, was systematically measured (Figure 10, Supplementary Figure 32, Figure 33, and Figure 34).

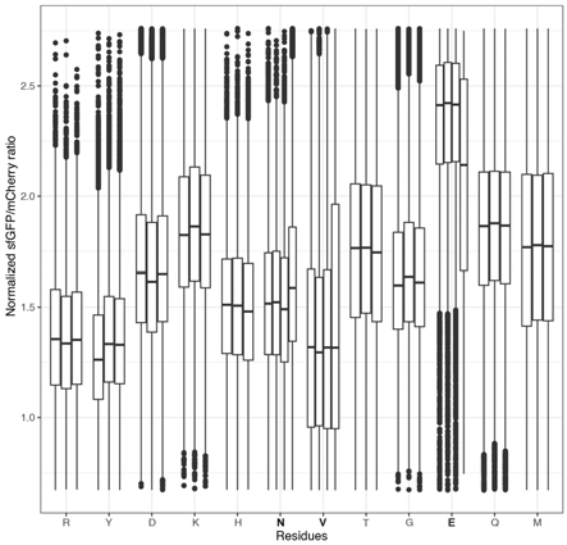


Figure 10. Normalized ratio of sfGFP to mCherry fluorescence for all twelve constructs for all three replicates

Compared to the study by Kats et al. (2018), our analysis revealed notable differences in residue stability (Figure 11). Specifically, our data indicates that the most stable residue is E rather than T, and V exhibits slightly lower stability compared to R, which was identified as the least stable

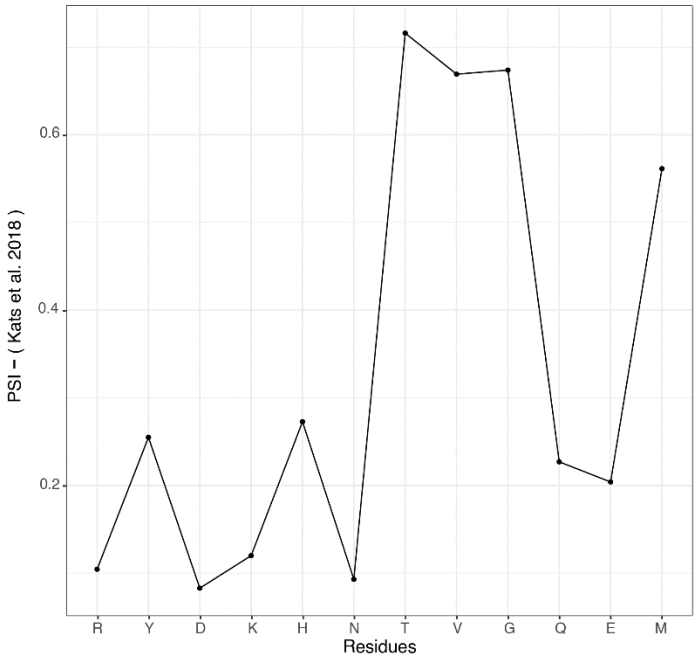


Figure 11. Stability of the first residue fused to the tFT. Data sourced from Kats et al. 2018. Each point represents the mean PSI for peptides starting with a specific residue

residue in the Kats et al. study. Additionally, we observed that Y demonstrates increased stability, while K appears more unstable than reported in the yeast model. These discrepancies may arise from the use of different reporters in the experiments, as well as the differences between the model organisms used in each study.

Data for the Ubi-X-eK-sfGFP-P2A-mCherry were compared with data from four other reporters, Ubi-X-eK-eGFP-P2A-mCherry, Ubi-X-eK-eGFP-IRES-dsRED, Ubi-X-linker-eGFP-P2A-mCherry, Ubi-X-linker-eGFP-IRES-DsRed (Figure 12). The ratio of GFP to RFP for each construct is represented by the mean per replicate since each construct had three biological replicates. Normalization to correct background effects for each construct is achieved by subtracting the mean value of the construct that exhibits the highest ratio from each construct. In this case, that was the mean value of the construct with amino acid E.

Our analysis revealed that linker-eGFP-P2A-mCherry and linker-eGFP-IRES-DsRed reporters exhibit the highest dynamic range compared to other reporters (Figure 12). They were shown to be the most promising reporters for our further studies since the differences in stability are much easier to be detected compared to other reporters. Reporters containing eK sequence did not show a very high dynamic range compared to reporters containing the linker sequence. With this, we showed that even though eK sequence can be successfully used in yeast studies, in studies with human cells it is not efficient enough to be used as a part of a reporter.

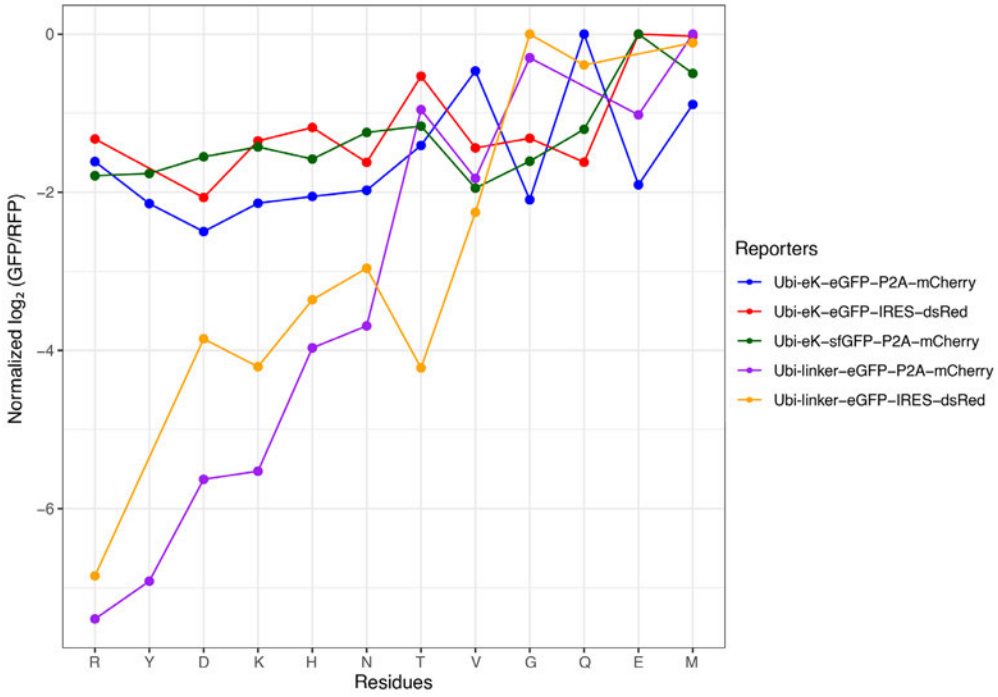


Figure 12. Stability of five different reporters. The mean per replicate represents the ratio of GFP to RFP for each residue. Normalization to correct for background effects involves subtracting the mean ratio of the residue with the highest value. This normalization process is carried out for each reporter.

Comparing the five reporters I tested with the reporter from the study by Timms et al. 2019, I observed that reporters starting with the eK sequence did not match the stability pattern observed in that study (Figure 13). This finding, as previously mentioned, supports the notion that the eK sequence is not optimal for use in human cells. On the other hand, reporters containing

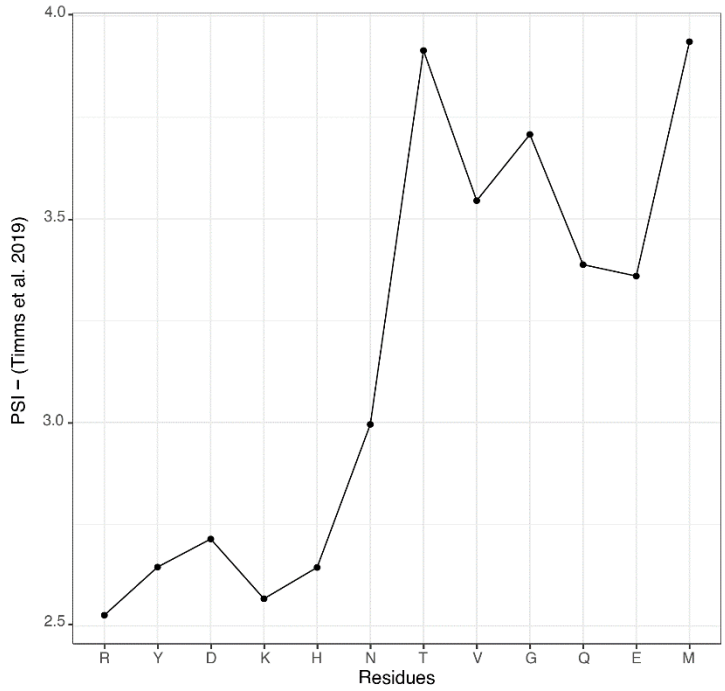


Figure 13. Stability of the first residue fused to the reporter linker-GFP-IRES-dsRed. Data sourced from Timms et al. 2019. Each point represents the mean PSI for peptides starting with a specific residue.

the linker sequence largely aligned with the pattern from the previous study. The reporter linker-eGFP-IRES-dsRed, used in the Timms et al. 2019 study as well, exhibited a similar pattern except for T, which appeared more unstable. This discrepancy suggests that the Ubi-T-linker-eGFP-IRES-dsRed construct may have issues with the IRES sequence in our cell line. In contrast, the reporter linker-eGFP-P2A-mCherry completely matched the observed pattern from the study. With these results, we concluded that linker-eGFP-IRES-dsRed and linker-eGFP-P2A-mCherry reporters are the best reporters for my cell line.

Ubi-XZ library

As revealed in the previous paragraph, the two reporters linker-eGFP-P2A-mCherry and linker-eGFP-IRES-dsRed exhibit the highest dynamic range in measuring peptide stability and matched the results from the previous study by Timms et al. 2019. The main impact on making the decision which of them should be used in our further study made differences between the P2A sequence and the IRES sequence. Since, as explained before, IRES is much longer compared to P2A and downstream protein can be expressed at a much lower rate than upstream protein, we chose to use a reporter containing the P2A sequence.

Upon identifying the most efficient reporter for assessing peptide stability, I wanted to gain a better understanding of the impact of the first two residues at the N-terminus in protein degradation. Two studies with similar goals were done before (Kats et al. 2018; R. T. Timms et al.

2019b). The study conducted in 2018 utilized yeast as a model organism, which, despite its eukaryotic nature akin to human cells, represents a simpler biological system. Consequently, numerous mechanisms governing protein stability in human cells may either be absent or differ significantly in yeast. Besides this, another study done in 2019 utilized the human cell line HEK 293T as a model organism. This study library contained 24 amino acids of primary isoforms of all human proteins (with and without first methionine). This study used a reporter containing IRES sequence which can have problems in the expression levels of proteins. They investigated the stability of the initial two amino acids, although stability assessments were conducted for all 24 amino acids. Making a library that contains only two amino acids at the N-terminus helps to look more deeply at the stability of an N-terminus.

To achieve this goal, I made an Ubi-XZ-linker-eGFP-P2A-mCherry library (named Ubi-XZ hereafter), comprising all possible variants of the 20 amino acids, resulting in a total of 400 dipeptides. These dipeptides were precisely incorporated at the N-terminus and fused to the selected reporter, linker-eGFP-P2A-mCherry. In order to assess the stability profile of the library, I employed the Global Protein Stability (GPS) profiling (Figure 14).

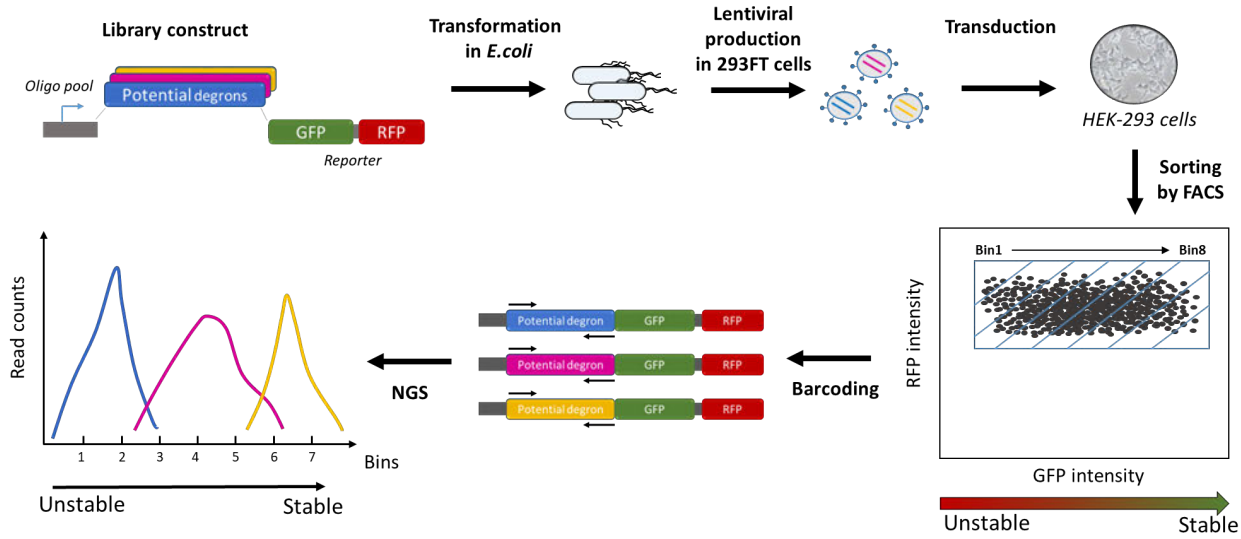


Figure 14. Schematic representation of Global Protein Stability (GPS) profiling of the Ubi-XZ library. The library was made of dipeptides integrated into a vector containing ubiquitin (Ubi) moiety and the reporter. This library was introduced into E. coli through transformation, followed by isolation and purification of the resulting DNA, which was used to produce lentiviruses. After the transduction of a library into the Flp-In T-REx 293 cell line, cells were sorted by Fluorescence-Activated Cell Sorting (FACS). Cells were prepared for Next-Generation Sequencing (NGS) and analyzed in R.

Due to mutations that can happen in the cells, it is possible to get stop codons instead of desired dipeptides in the library. To check the amount of stop codons across all three replicates in the Ubi-XZ library, before producing lentiviruses and transduction, the library was sent for NGS. Data from NGS was analysed then by [redacted]. The analysis revealed that the percentage of stop codons in all three replicates was less than 11% (Figure 15). Even though the probability of

getting a stop codon in one of two positions is 9.2%, we understand that PCR amplification required for barcoding for NGS can increase this percentage. We chose to accept less than 11% of stop codons as indicative of good quality data and proceeded with transduction into Flp-In T-REx 293 cell lines.

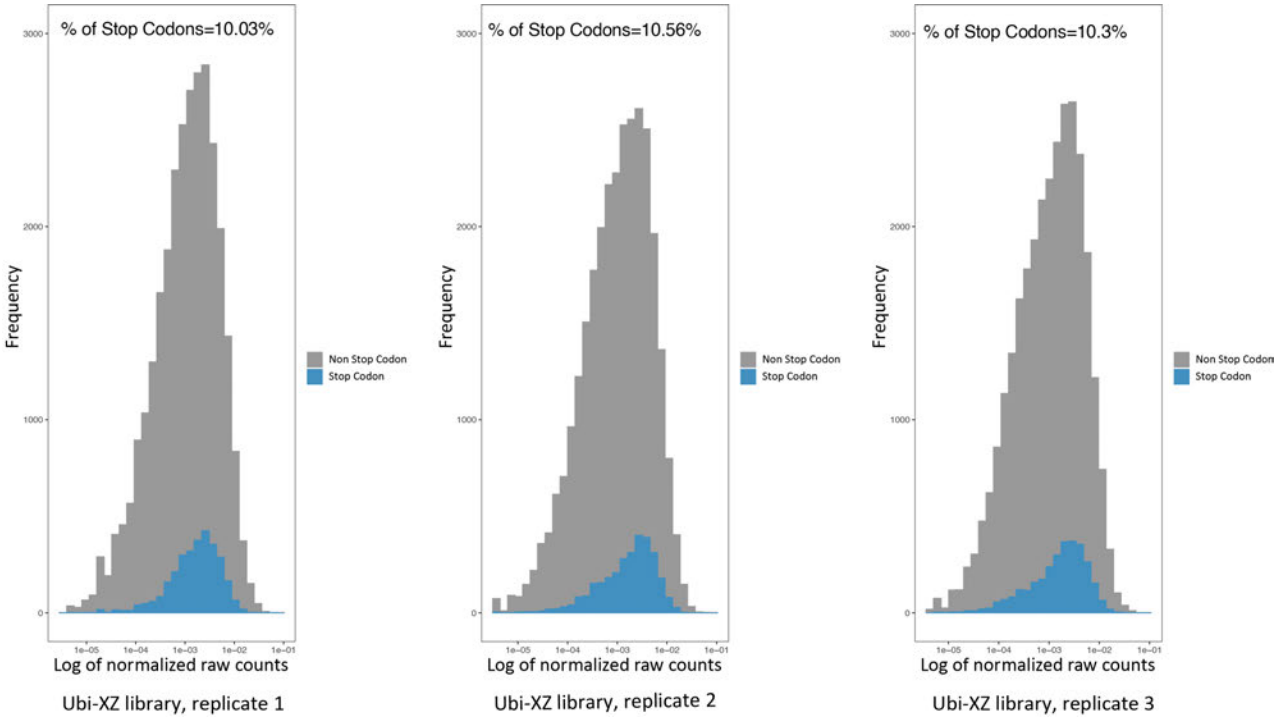


Figure 15. Quality control shows the good quality of all three replicates of Ubi-XZ library. Variation of frequency of normalized raw count in log2 format. The blue color indicates the region of the stop codon and grey non stop codons. Data is showing plots for all three replicates.

The Ubi-XZ library was constructed in biological triplicate. By comparing, our replicates between each other we demonstrates a high degree of correlation indicating that this library is highly reproducible (Figure 16A, Supplementary Figure 35). The results generated from the Ubi-XZ library are visually depicted through a heatmap in Figure 16B, which illustrates all combinations of two residues at the N-terminus. Dipeptides starting with N, D, E, R, K, H, W, L, F, Y, I, and C exhibit instability, consistent with the residues recognized for degradation by the Arg/N-degron pathway. Interestingly, dipeptides starting with Q, considered to be tertiary destabilizing residue within the Arg/N-degron pathway, demonstrate stability, particularly when paired with residues such as L, I, Y, F, W, M, S, E, A, D, and P, with QD emerging as the most stable combination. Furthermore, the secondary destabilizing residue N exhibits instability when paired with any residue at the second position except D. This finding appears to contradict the established understanding of the Arg/N-degron pathway. This suggests either a malfunction within the known pathway or the involvement of an alternative mechanism in preventing degradation.

Dipeptides starting with R and K consistently display instability regardless of the second residue. Next, those featuring D in the second position promote stability irrespective of the first residue, with dipeptide DW identified as the most unstable among them. Apart from dipeptides starting with M, S, and A, other residues recognized by the Ac/N-degron pathway—namely, C, V, G, T, and

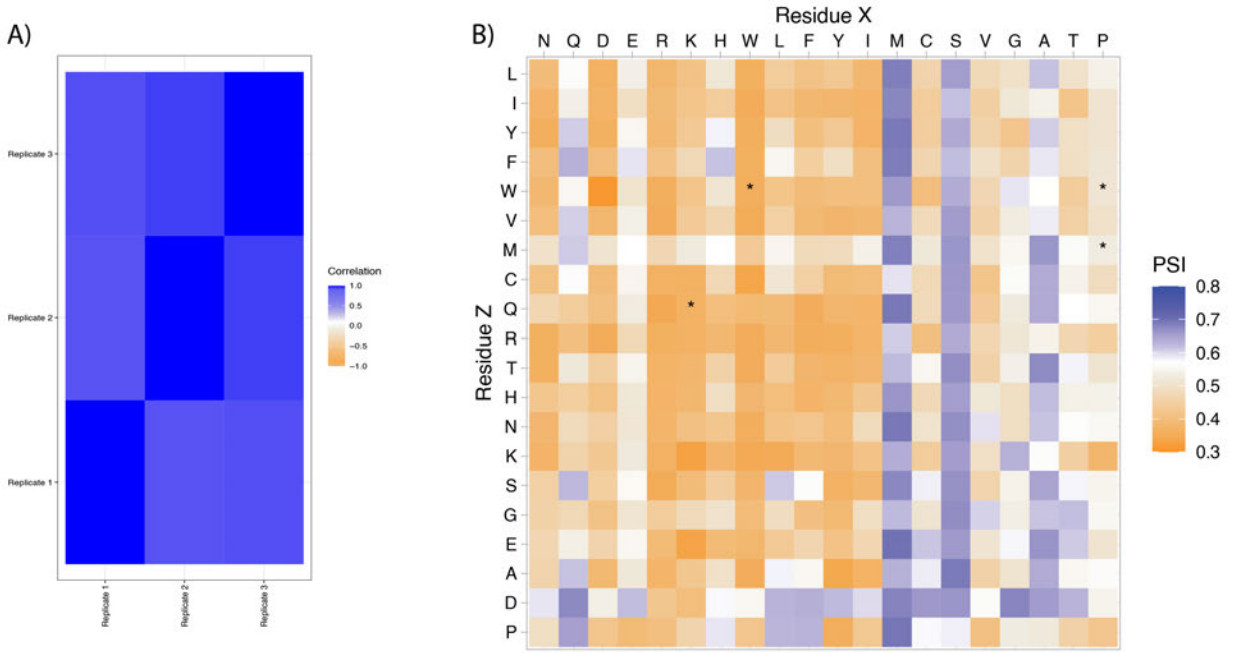


Figure 16. A) Correlation among the PSI of different replicates in the Ubi-XZ library. Correlation is calculated using Pearson correlation; B) *Heatmap displaying the variation of Protein Stability Index (PSI) across the Ubi-XZ library in wild type human cells. The x-axis shows the first residue at the N-terminus and the y-axis represents the second residue. Stable residues have a blue color, while unstable residues have an orange color. The PSI is the regressed PSI from the three replicates. Residues with * represent the most similar residue among all replicates.*

P—are generally found to be unstable. Despite the established status of the Ac/N-degron pathway, it appears that its targeting of all N-terminal degrons is not uniform. Prior investigations in yeast have demonstrated that acetylation does not consistently serve as a degron, a phenomenon which may also be applicable in human cells. Additionally, beyond the Ac/N-degron pathway, N-terminal cysteines can potentially be targeted for degradation by Arg/N-degron pathway, but only if it is previously oxidated (Figure 2)**.**This indicates that acetylation may shield N-terminal cysteine from being targeted by the Arg/N-degron pathway until it is oxidized.

In a previous independent study conducted in the HEK-293T cell line, Timms et al. (2019) employed GPS to investigate the stability of a library encompassing the first 23 residues of primary isoforms of all human proteins. Leveraging this dataset, we conducted a comparative analysis with the Ubi-XZ library. The correlation between the Peptide Stability Index (PSI) of the library from Timms et al. (2019) and the PSI of the Ubi-XZ library indicated a strong correlation coefficient of 0.84 (Figure 17A).

Data from Timms et al. (2019) study was grouped based on the initial two residues at the N-terminus, and the mean values were calculated, generating a heatmap (Figure 17B). Peptides starting with residues expected to be targeted for degradation by the Arg/N-degron pathway in general demonstrated instability. On the other hand, peptides initiating with Q or E, mirroring the findings of the Ubi-XZ library, exhibited higher stability compared to other residues. Additionally,

peptides featuring D in the second position, regardless of the first residue except for R, K, and W, demonstrated notable stability. Consistent with the observations from the Ubi-XZ library, residues at the N-terminus expected to be recognized by the Ac/N-degron pathway displayed overall greater stability, particularly evident in dipeptides beginning with M, S, and A.

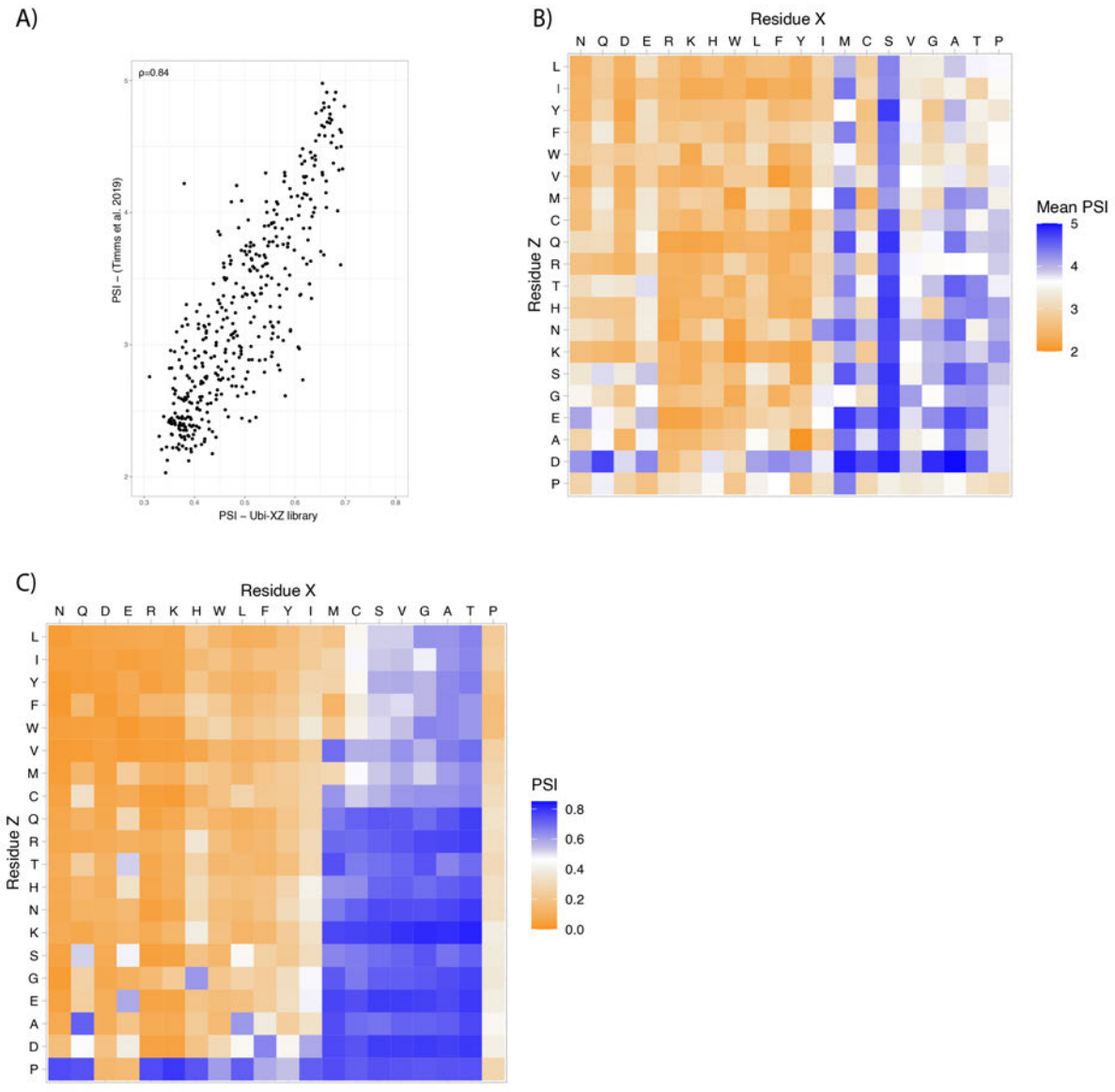


Figure 17. A) Variation of PSI from Timms et al. 2019 study with PSI of Ubi-XZ library (cor: Pearson). B) Heatmap displaying the variation of Protein Stability Index (PSI) across library from Timms et al. 2019 study in wild-type human cell line HEK-293T. The x-axis shows the first residue at the N-terminus and the y-axis represents the second residue. Stable residues have a blue color, while unstable residues have an orange color. C) Heatmap displaying the variation of Protein Stability Index (PSI) across library Kats et al. 2018 study in wild type Saccharomyces Cerevisiae. The x-axis shows the first residue at the N-terminus and the y-axis represents the second residue. Stable residues have a blue color, while unstable residues have an orange color.

Previous research conducted in yeast strain *Saccharomyces cerevisiae* by Kats et al. (2018) confirms the observed trend within the Arg/N-degron pathway seen in the Ubi-XZ library (Figure 17C). Unlike the human Arg/N-degron pathway, which can also target oxidized cysteine residues

at the N-terminus, yeast does not exhibit this capability and this is one of the possible explanations for the increased stability of dipeptides beginning with this residue. Additionally, specific dipeptides starting with Q or E, akin to those identified in the human library, demonstrated enhanced stability. They showed that sometimes particular dipeptides starting with E (EM, EQ, EH, ED, ES, ET, EE) can be acetylated and like that be prevented from being degraded by the Arg/N-degron pathway. On the other hand, the investigation revealed that the Ac/N-degron pathway in yeast rarely recognizes N-terminal degrons but the overall hydrophobicity of the N-terminus is the primary factor influencing the strength of the degrons (Kats et al. 2018).

Comparing the data from the Ubi-XZ library with previous studies has provided us with guidelines for understanding deviations from previously known pathways. The findings from Timms et al. in 2019 strongly support those of the Ubi-XZ library, confirming that the N-termini of most peptides are indeed correctly identified as degrons. However, we also observe deviations in pathways, particularly in the Arg/N-degron pathway, where residues Q and E, expected to be targeted for degradation, exhibit high stability in both our data and that of Timms et al. Furthermore, our data exhibit some differences with studies conducted in yeast, which could be attributed to differences in model organisms. Notably, certain E-starting dipeptides exhibit stability in yeast due to acetylation, suggesting a similar mechanism may exist in human cells—a hypothesis I aim to investigate. Nonetheless, it is crucial to verify the reproducibility of the Ubi-XZ library to rule out experimental noise as a cause for deviations from the N-degron pathway.

Verifying the reproducibility of the Ubi-XZ library

Although the Ubi-XZ library was done in triplicate, showing its high reproducibility, I thought it essential to conduct additional validation to ensure that my results are not merely a product of experimental noise. In order to check the reproducibility of the whole library and to focus on the Arg/N-degron pathway, I choose to make single constructs with the following dipeptides: DD, DL, DR, ED, EL, ER, ND, NL, NR, PK, PL, QD, QL, and QR, fused to the linker-eGFP-P2A-mCherry reporter.

Experiments with single constructs were conducted in a similar meaner to the library. I transduced single constructs into the Flp-In T-REx 293 cell line and measured the stability by Flow Cytometry.

The eGFP/mCherry ratio of fourteen constructs containing particular dipeptides was compared with the PSI of the same dipeptides from the Ubi-XZ library. Results show a correlation of 0.9, indicating that the interesting observations in the Ubi-XZ library heatmap were not a result of noise (Figure 18A, Supplementary Figure 36). Furthermore, the stability between dipeptides with a second residue L, R, or D shows the same trend as stability in the library (Figure 18 B, C, D and Supplementary Figure 36). With this, I confirmed that my data is highly reproducible and it is not a product of experimental noise.

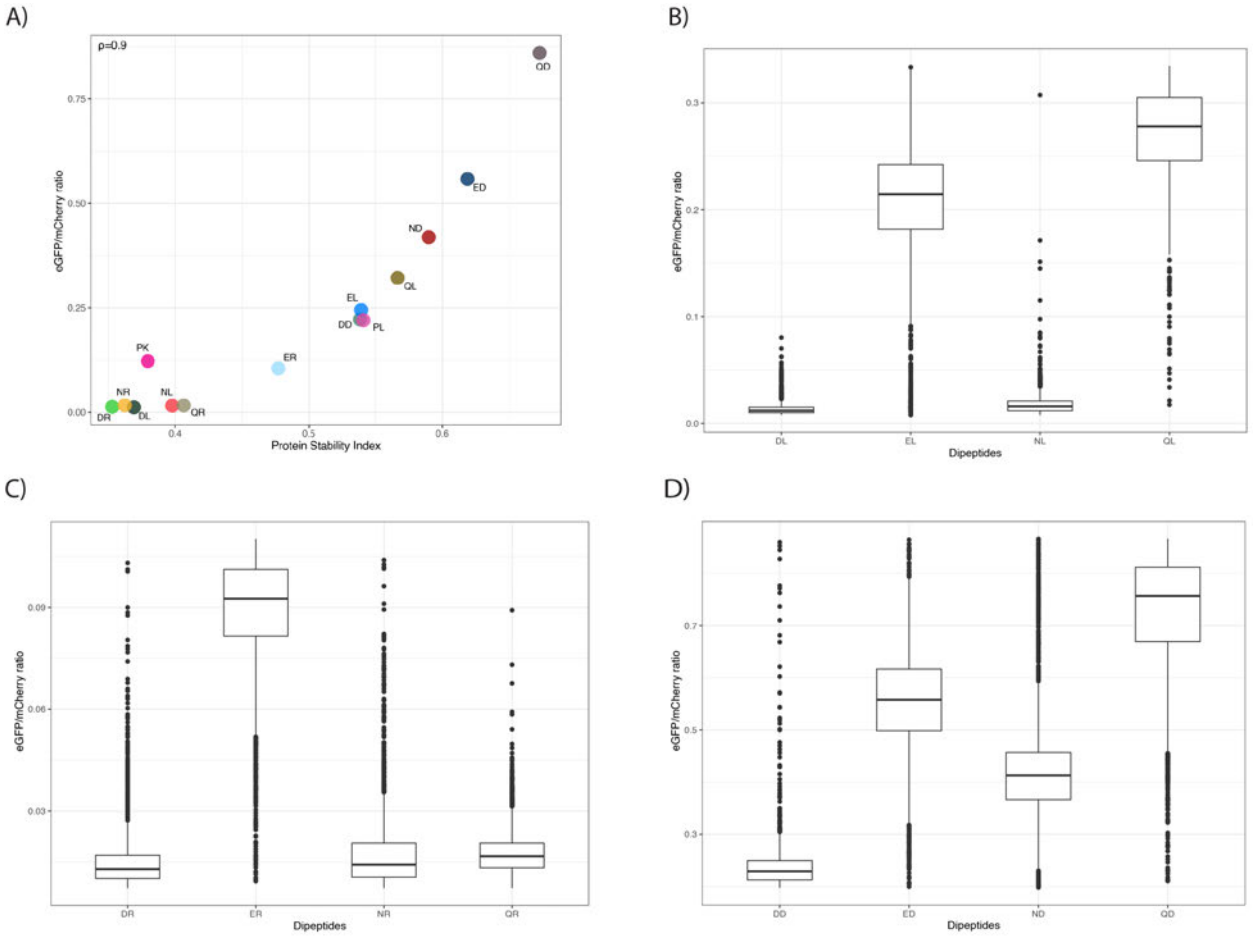


Figure 18. A) Variation of eGFP/mCherry 14 single constructs containing particular dipeptides with PSI of the same dipeptides from the Ubi-XZ library. PSI taken is the regressed PSI from all three replicates and is indicated in the y-axis. The mean of eGFP/mCherry is taken per dipeptide and indicated in the x-axis. Pearson correlation index of 0.9 shows that the stability of translated amino acids from both experiments is correlated. B) Boxplot of eGFP/mCherry ratio of all four dipeptides with a second residue L. C) Boxplot of eGFP/mCherry ratio of all four dipeptides with a second residue R. D) Boxplot of eGFP/mCherry ratio of all four dipeptides with a second residue D.

Given our observation that D and E at the protein N-terminus do not conform to the Arg/N-degron pathway in human cells, I seek to understand the underlying reasons. Several hypotheses arise. Firstly, the deviations from expected degradation patterns could potentially be attributed to differential expression levels of the key enzymes involved in the Arg/N-degron pathway. It is plausible that in the specific cell line utilized for this study, certain components of the machinery may exhibit reduced or negligible expression, leading to an impaired functionality of this degradation pathway. Further investigations into the transcriptional levels of the relevant genes encoding the Arg/N-degron pathway components would be necessary to validate this hypothesis. Another plausible hypothesis to explain the observed deviations involves the substrate specificity of enzymes involved in the Arg/N-degron pathway, especially NTAQ1 that should target N-terminal Q. This differential specificity could lead to preferential targeting or stabilization of certain substrates, resulting in the observed discrepancies from predicted degradation patterns. As mentioned before, the study done in yeast have shown that acetylation of N-terminal residues can prevent their degradation by the Arg/N-degron pathway. It would be interesting to investigate

whether this regulatory mechanism is conserved in human cells. Exploring the interplay between N-terminal acetylation and the recognition mechanisms of the Arg/N-degron machinery could provide insights into the modulation of protein stability and turnover.

Transcript levels of NTAN1, NTANQ1, ATE1, and UBR family E3 ligases

Firstly, I wanted to determine whether the genes associated with the Arg/N-degron pathway were indeed expressed in my cell line. The presence or absence of expression of these genes could potentially elucidate the lack of degradation observed in specific dipeptides targeted by this pathway. To assess this, I examined the transcripts per million (TPM) values for all genes in the HEK293 cell line. The data was taken from the "Human Protein Atlas" database (Nagaoka et al. 2020). Since the HEK293 cell line is a base cell line for Flp-In T-REx 293 that I used in my experiments, it was a great option for comparing expression for my genes of interest (Graham et al. 1977). Upon comparison with the transcript levels of other genes, it was found that six of the selected genes - ATE1, NTAN1, NTAQ1, UBR1, UBR2, UBR4, and UBR5 - are indeed transcribed in the HEK293 cell line. In total, there are 20162 expressed genes with a TPM median of 3.2 while my genes of interest had TPM values of: 11.4 (ATE1), 5.1 (NTAN1), 13.6 (NTAQ1), 7.2 (UBR1), 9.1 (UBR2), 32.1 (UBR4), and 19.4 (UBR5). This indicates that my genes of interest are expressed at a higher level than the average gene in the cell (Figure 19). With this, I confirmed that the unusually

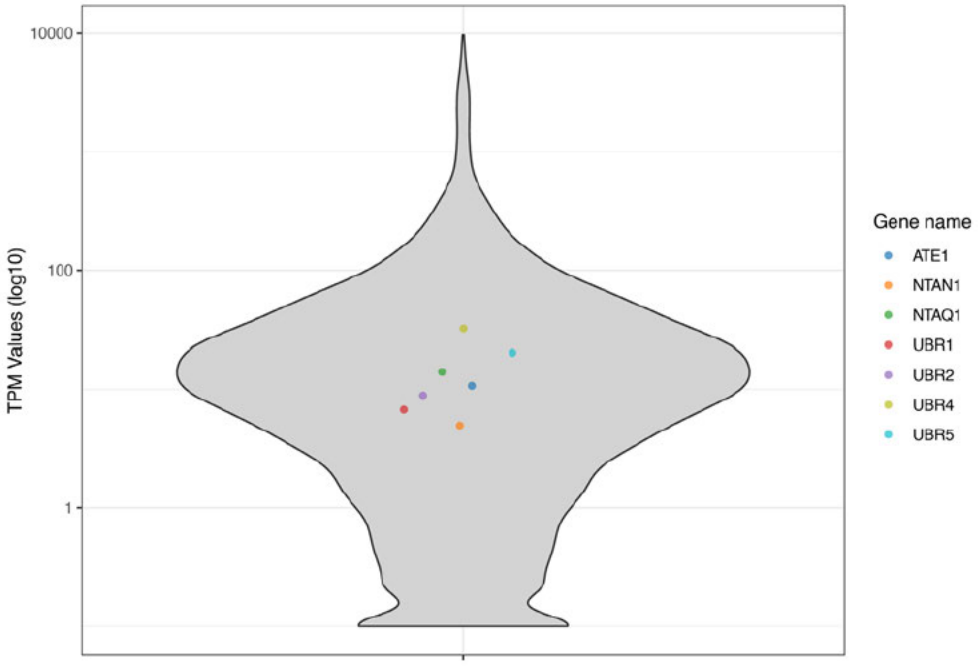


Figure 19. Distributions of TPM values for the seven genes of interest indicated. TPM Values are in log10 scale for better visualization. Log10(median) of TPM is 0.55.

high stability of particular dipeptides in the Ubi-XZ library is not due to the lack of expression of genes related to the Arg/N-degron pathway. This is additionally confirmed by Western Blot for the ATE1 protein in Flp-In T-REx 293 as shown in Figure 21B.

Specificity of human NTAN1 and NTAQ1 in *Saccharomyces cerevisiae*

In contrast to the human Arg/N-degron pathway, the equivalent pathway in *Saccharomyces cerevisiae* exhibits some distinctions. Specifically, in yeast, the deamidation of tertiary destabilizing residues is only done by Nta1 as opposed to the involvement of both NTAN1 and NTAQ1 in humans (Figure 2).

To understand the specificity of human NTAN1 and NTAQ1 my colleague Christian Ochs and I conducted rescue experiments with human NTAN1 and NTAQ1, along with yeast NTA1, in yeast strains expressing a library of 80 dipeptides. Additionally, in these strains, yeast gene *Nta1* was knocked out.

The dipeptide library Ubi-XZ-tFT contains variations where one of the four residues (D, E, N, or Q) served as the first residue at the N-terminus, with all 20 amino acids represented as the second residue. In total 80 dipeptides were fused to the tFT reporter containing mCherry-sfGFP as explained before in the section "Selecting the optimal approach for measuring protein turnover". The stability of the dipeptide library was calculated relative to the stability of the wild-type (WT) strain, which served as the background reference. The outcome of the experiment demonstrated that rescuing the *Nta1* knockout with yeast *Nta1* restored the stability of the library to levels comparable to those observed in the wild-type, as expected (Figure 20).

Furthermore, the introduction of human *NTAN1* into the yeast *NTA1* knockout strain revealed that dipeptides starting with Q showed to be very stable while dipeptides starting with D, E, or N were observed to be more unstable (Figure 20). The introduction of *NTAQ1* into the *Nta1* knockout strain resulted in the stability of dipeptides starting with N while dipeptides starting with D, E, and Q were observed to be more unstable. Interestingly, dipeptides starting with Q and having second residues E or T also showed higher stability. This indicates that those dipeptides are not targeted for deamidation by NTAQ1, which can explain the higher stability of these dipeptides in the Ubi-XZ library. On the other hand, this still does not explain the high stability of other dipeptides starting with Q.

In summary, these findings suggest that similarly to their roles in human cells, both human NTAN1 and NTAQ1 in yeast exhibit a marked specificity for targeting residues beginning with N or Q, respectively, in the degradation process. On the other hand, we didn't find any unusual changes in the specificity of NTAN1 or NTAQ1 that will explain the stability of dipeptides in the Ubi-XZ library, especially once starting with Q.

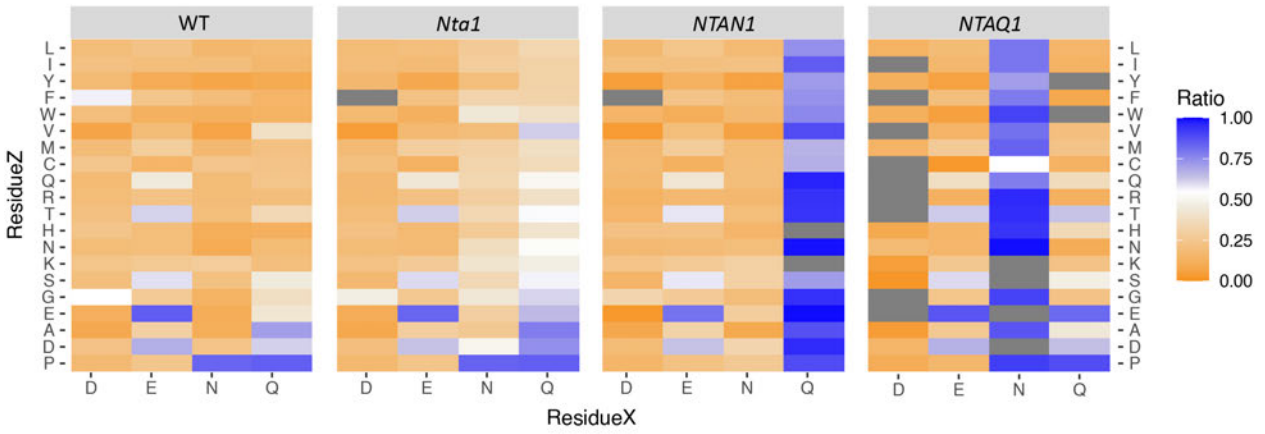


Figure 20. Heatmap displaying the variation of mCherry-sfGFP ratio in the yeast strain lacking yeast NTA1. A higher ratio implies higher stability. Grey squares are NA data. The x-axis shows the first residue at the N-terminus and the y-axis represents the second residue. Stable residues have a blue color, while unstable residues have an orange color

Knockout of ATE1

To investigate the degradation of dipeptides situated at the N-terminus targeted by the Arg/N-degron pathway, CRISPR/Cas9 technology was employed to knockout the human ATE1 gene. This process involved two distinct guide RNAs (gRNAs) to induce separate cleavages within the gene structure. Specifically, the first cleavage targeted exon 2 (E2), followed by a cleavage in exon 4 (E4) (Figure 21A).

To validate the success of ATE1, I conducted the Western Blot using a monoclonal antibody specific to ATE1 (Figure 21B). The absence of a detectable ATE1 protein band in the knockout cell line, compared with the wild-type (WT) cell line, showed the efficacy of the gene editing procedure. Although this outcome provided substantial evidence for the ATE1 knockout, an additional experiment was conducted for further confirmation.

In this experiment, I expressed six Ubi-XZ-eGFP-P2A-Cherry constructs, where XZ represented one of the following dipeptides: ND, NR, QD, QR, DD, and DR (Figure 21C). Those dipeptides were chosen since they are associated with the Arg/N-degron pathway, hypothesizing their stabilization in the absence of ATE1. Additionally, I included a construct Ubi-M-eGFP-P2A-Cherry which was used as a control construct since it is known to be the very stable residue, as shown by previous studies as well as my data from the Ubi-XZ library (Figure 16B and Figure 17B). The results revealed a significant increase in stability for each construct, in the ATE1 knockout cell line compared to the WT. This evidence confirmed the successful establishment of the ATE1 knockout cell line.

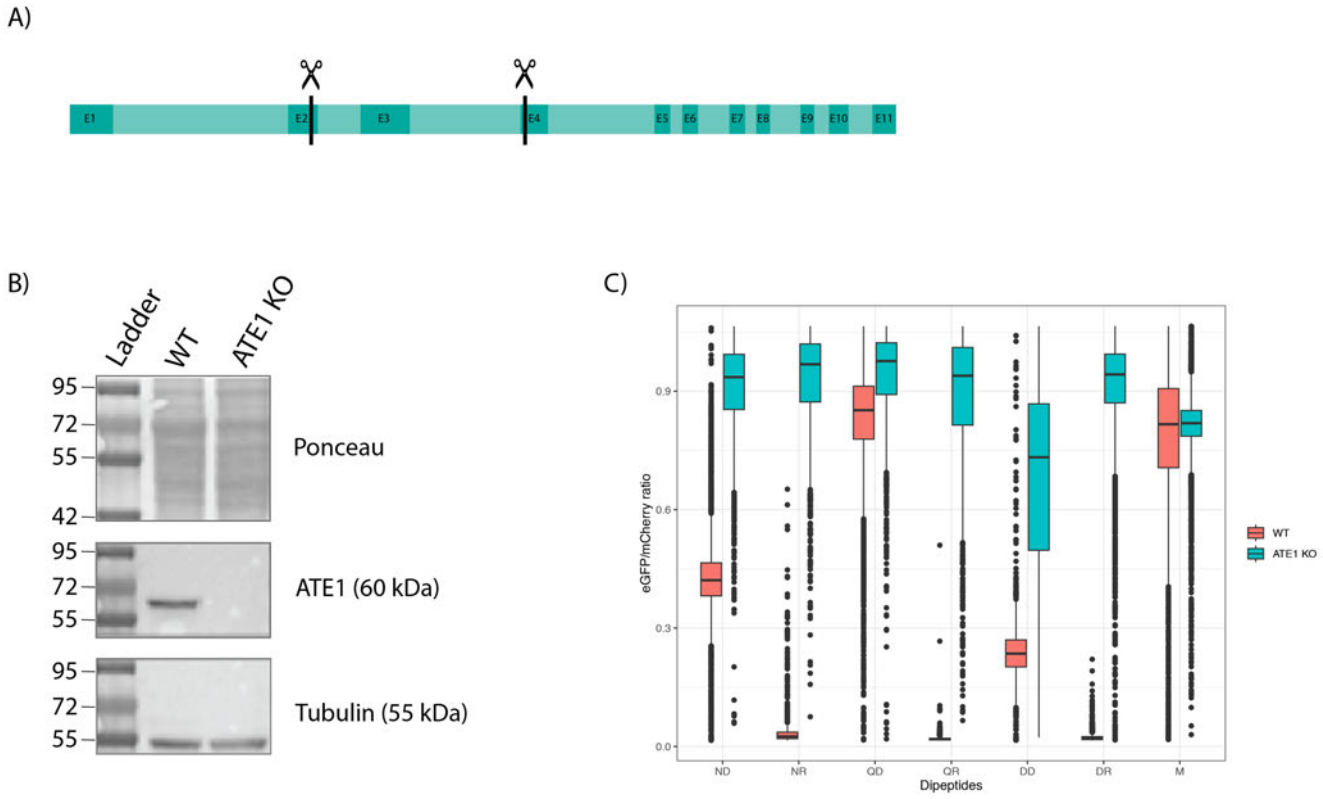


Figure 21. A) Cartoon representing eleven exons of ATE1 gene and positions of cuts done by CRISPR/Cas9 system. B) Western blot results showing absence of ATE1 protein in the ATE1 KO cell line. C) The ratio of eGFP to mCherry fluorescence of six constructs and control expressed in the WT cell line and ATE1 KO cell line.

Acetylation of dipeptides related to Arg/N-degron pathway

I hypothesize that acetylation prevents the targeting of specific dipeptides at the N-terminus by the Arg/N-degron pathway. I conducted the experiment involving the expression of Ubi-XZ-eGFP-P2A-Cherry constructs in the ATE1 knockout cell line, followed by immunoprecipitation with magnetic agarose GFP beads and the analysis of N-terminal acetylation by Mass Spectrometry. To measure in vivo N-terminal acetylation of my constructs, we used D6-acetic anhydride to artificially acetylate all unacetylated sites before chymotrypsin digestion. This is important because we aimed to quantify the difference in acetylation abundance while ensuring that this difference was not due to varying detectability. The protocol was based on the protocol described in a paper by Yeom et al. 2017.

Twelve distinct cell lines were generated, each expressing one of the constructs with NL, NR, ND, QL, QR, QD, DL, DR, DD, EL, ER, and ED, fused with eGFP as a marker. Immunoprecipitation with magnetic agarose GFP beads was performed to selectively isolate the expressed constructs from the cells (Supplementary Figure 37). Acetylation patterns were analyzed between dipeptides with distinct initial residues, providing insights into the potential role of acetylation in modulating the degradation mediated by the Arg/N-degron pathway.

For each individual cell line, we analyzed the presence of acetylated and non-acetylated constructs. Since tertiary destabilizing residues (N and Q) can undergo deamidation, leading to the formation of secondary destabilizing residues (E and D) within the Arg/N-degron pathway, I separated samples into two pools.

Pool 1 contained just samples with constructs starting with N and Q. During the analysis we focused on identifying acetylated and non-acetylated dipeptides carrying N or Q on the end terminus but also the bearing E and D. Pool 2 contained samples with constructs starting with E and D. In our analysis of this pool we focused on dipeptides starting with E and D but we also checked the presence of constructs starting with N and Q and their acetylation.

The results unveiled a distinct pattern in Pool 1 (Figure 22A). The intensity of acetylation was normalized by comparing acetylated and non-acetylated dipeptides measured within a single cell line for each replicate. Each cell line was done in three replicates. The intensity of acetylation in dipeptides starting with N or Q exhibited a notable elevation compared to those with E and D in the same cell line with an exception in the cell line expressing construct with QL in which EL was more acetylated when QL. Observations also revealed acetylation of DD and ED in the cell lines with constructs with ND and QD, respectively.

Furthermore, in Pool 2 intensity of acetylation was normalized like in Pool 1. In this pool we exhibited that cell lines expressing constructs with dipeptides starting with E or D have acetylation of dipeptides starting with E or D except in a cell line expressing construct with DR. Surprisingly, in this pool we detected instances of acetylation in dipeptides originating from N and Q as well, even though this is very unexpected since NTAN1 and NTAQ1 participate in deamidation but the reverse process is not for now known (Figure 22B) (Drazic et al. 2016). The implications of this unexpected observation require further investigation. To better visualize the comparison of acetylation between different cell lines, I created a plot with the percentage of acetylation normalized by the levels of acetylated and non-acetylated forms of each dipeptide for each replicate. This was done for both pools (Figure 22C and D).

In general, the trend of acetylation also matches the stability of the same dipeptides observed in the Ubi-XZ library. For example, if we compare the stability of NR, NL, and ND in the Ubi-XZ library, we will notice that stability increases in that order (Figure 16B). On the other hand, acetylation increases in the order of NR, NL, and ND. This pattern is consistent across sequences beginning with Q, D, or E followed by second residues R, L, and D.

These results suggest that the stability of specific dipeptides in the Ubi-XZ library may be due to acetylation, which prevents them from being targeted for degradation by the Arg/N-degron pathway.

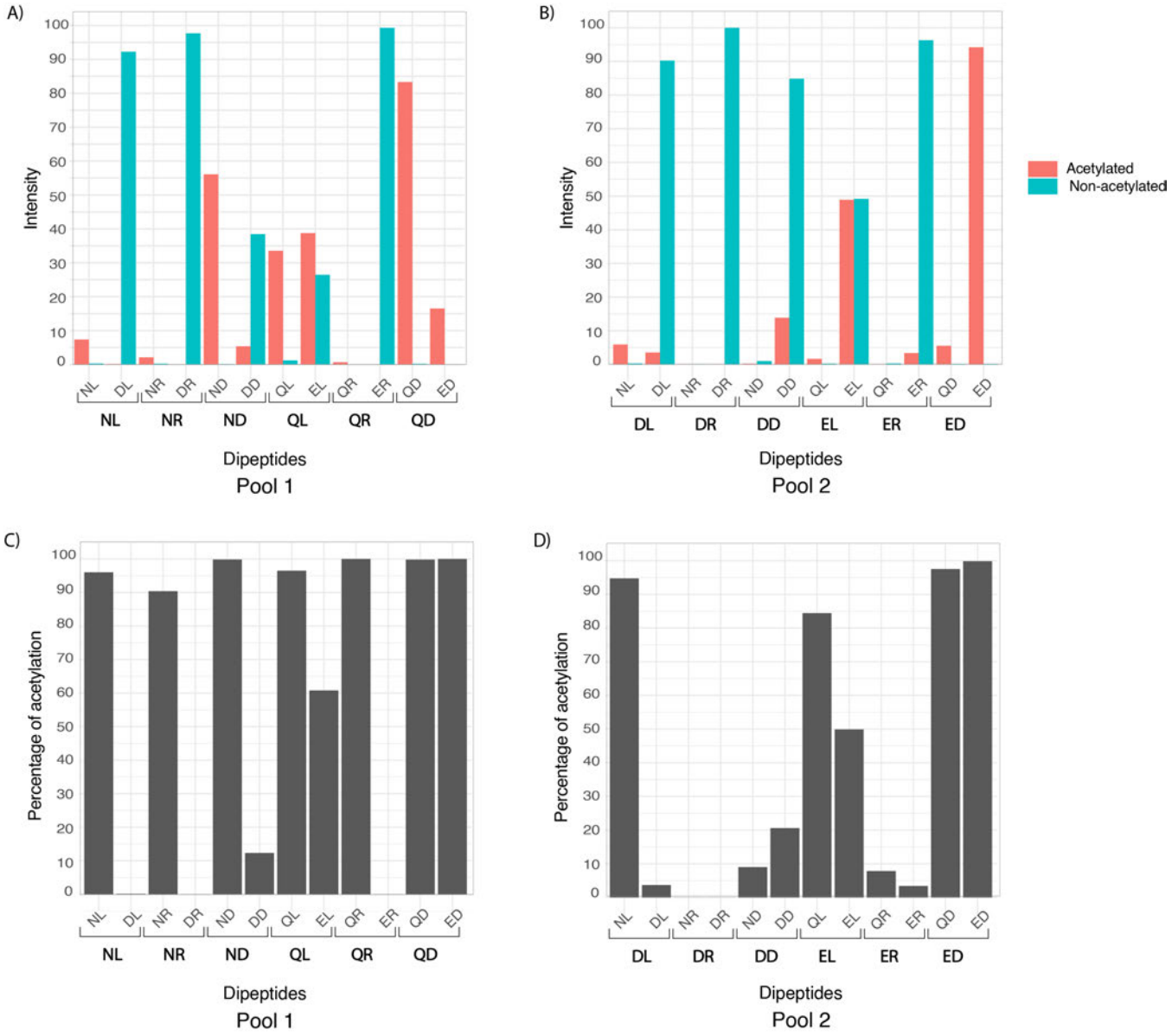


Figure 22. A) Intensity of acetylation normalized by acetylated and non-acetylated dipeptides measured within one cell line for each replicate. The plot shows the average percentage of three replicates for acetylated and non-acetylated dipeptides of Pool 1. B) Same but for Pool 2. C) Percentage of acetylation normalized by acetylated and non-acetylated of each dipeptide for each replicate. The plot shows the average percentage of three replicates for acetylated and non-acetylated dipeptides of Pool 1. D) Same but for Pool 2.

Detecting protein binders of N-terminal library by ProP-PD

In order to understand the bindings of NTAN1, NTAQ1, and ATE1 proteins to the N-terminus of human proteins in vitro, I performed proteomic peptide-phage display (ProP-PD). For this, I made an N-terminal library containing all isoforms of all human proteins with and without initiator methionine and cleavage sites and microbial infection cleavage sites annotated by UniProt. The N-terminal library include in total 82 680 unique oligos encoding 39 545 different unique peptides.

Depending on presence of initiator methionine sequences were 12 or 11 amino acids long. Using a modified Kunkel mutagenesis method, the oligonucleotide library is fused to the major coat protein P8 of the filamentous M13 phage. To be able to expose peptides at the N-terminus, and not at C-terminus or internal part of a phage P8 protein like in previous libraries, I inserted library after the localization signal that undergoes cleavage, thus enabling exposure of the N-terminus for binding purposes (Figure 23) (Davey et al. 2017a; Garrido-Urbani et al. 2016; Ivarsson et al. 2014). Besides new N-terminal library I created, I also used HD2 library already existing in Ivarsson's lab (Benz et al. 2022). This library contains disordered regions fragmented into 16 amino acids in length tiles, with an overlap of 12 amino acids between adjacent peptides. In total library has 938 427 peptides derived from 16 969 proteins.

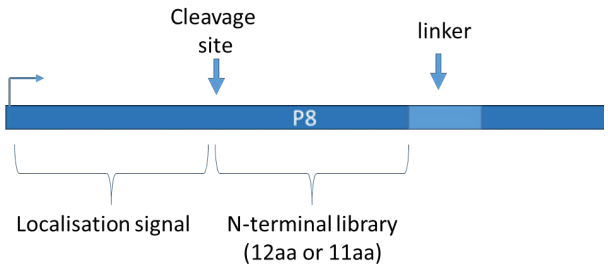


Figure 23. Schematic representation of the N-terminal phage library inserted into the phage P8 protein. N-terminal library containing peptides of 12 or 11 amino acids length is located after cleavage site.

To find binders by ProP-PD, together with the IMB Protein Purification Core Facility we produced the following proteins tagged with GST tag: GST-NTAQ1, GST-ATE1_iso1 (isoform 1), GST-ATE1_iso2 (isoform 2), GST-NTAN1 and NTAN1-GST (Figure 24A). Two isoforms of ATE1 defer only in 129-bp region encoding similar yet distinct sequences (Kwon, Kashina, and Varshavsky 1999a). Full protein GST-ATE1_iso1 was purified twice (GST-ATE1_iso1_I and GST-ATE1_iso1_II), since first elution gave low yield in a second we used elution buffer with increased concentration of imidazole. Besides that, as an additional proteins we also purified a few yeast proteins, truncated in the way to miss part on the N-terminus or C-terminus for better overexpression and purification: GST-Gid4(1-358), GST-Gid10(1-292), GST-Gid4(99-362), and GST-Gid10(115-292) (Figure 24B). Those proteins are two recognition subunits of yeast GID complex that target P at the protein N-terminus for degradation. In this case I used them as a control since ProP-PD was already done with them in a previous study (Chrustowicz et al. 2022b). Additionally, I purified two more human truncated proteins GST-ZER1(519-766) and GST-ZYG11B(480-728) (Figure 24C). Those proteins are chosen as a control since they are proven to target G at the protein N-terminus through the Gly/N-degron pathway in human cells (R. T. Timms et al. 2019a).

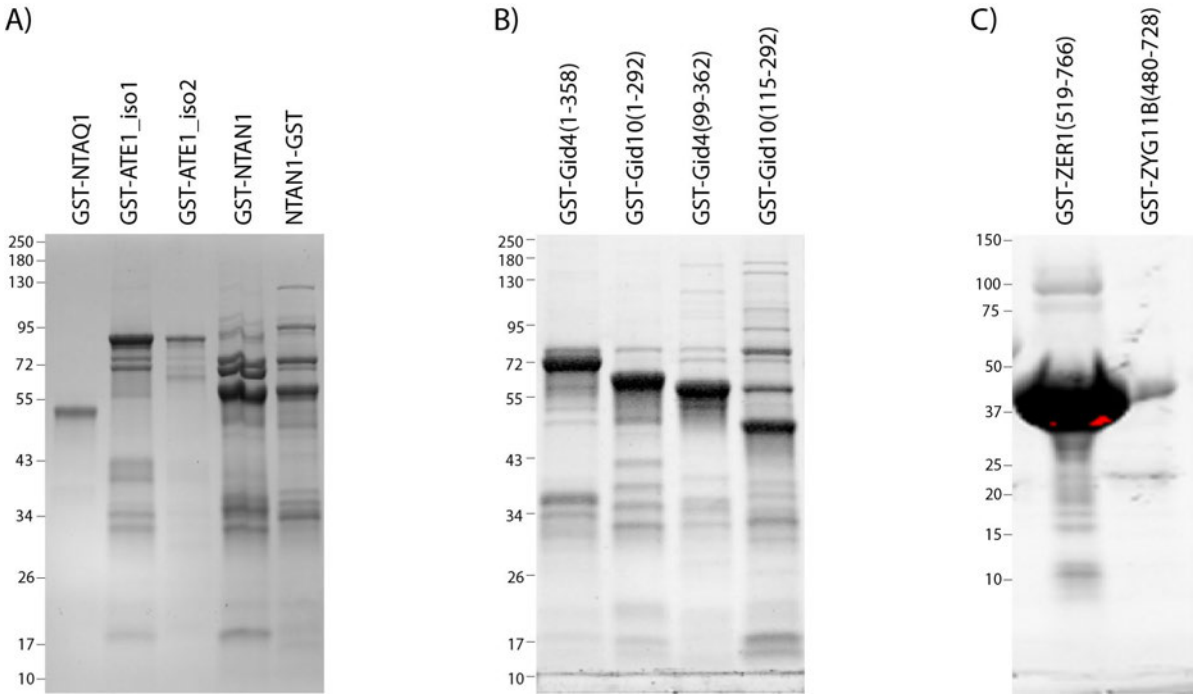


Figure 24. A) Coomassie Blue-stained 12% SDS–PAGE gel showing GST-NTAQ1 (52kDa), GST-ATE1_iso1 (87kDa), GST-ATE1_iso2 (87kDa), GST-NTAN1 (62kDa), NTAN1-GST (62kDa) (3µg of protein in each lane) B) Coomassie Blue-stained 12% SDS–PAGE gel showing GST-Gid4(1-358) (68kDa) , GST-Gid10(1-292) (62kDa), His-GST-Gid4(99-362) (58kDa) and GST-Gid10(115-292) (49kDa) (3µg of protein in each lane). C) Coomassie Blue-stained 4-20% SDS–PAGE gel showing GST-ZER1(519-766) (57kDa) and GST-ZYG11B(480-728) (56kDa)

After undergoing four rounds of phage selection of the N-terminal library and HD2 library against baits of interest, I wanted to detect to which proteins phages bound. For this, I performed pooled phage ELISA to detect the ratio of bait protein ELISA signal to that of control which contained just GST. I used an anti-M13 HRP-conjugate antibody for detection. Proteins that I was using for the new library I did in quadruplicates, while for the already established HD2 library, I used triplicates. My ELISA for the N-terminal library showed phage enrichment for GST-ATE1_iso1_I, GST-NTAQ1, GST-Gid4(99-362), and GST-ZER1(519-766). I also observed phage enrichment in the HD2 library for GST-ATE1_iso1_I and GST-NTAQ1. The threshold was at least double of phages in the sample compared to a negative control containing just a GST. For all phage pools that showed enrichment more than 2 folds higher than GST control, I sent for Next Generation Sequencing (Figure 25).

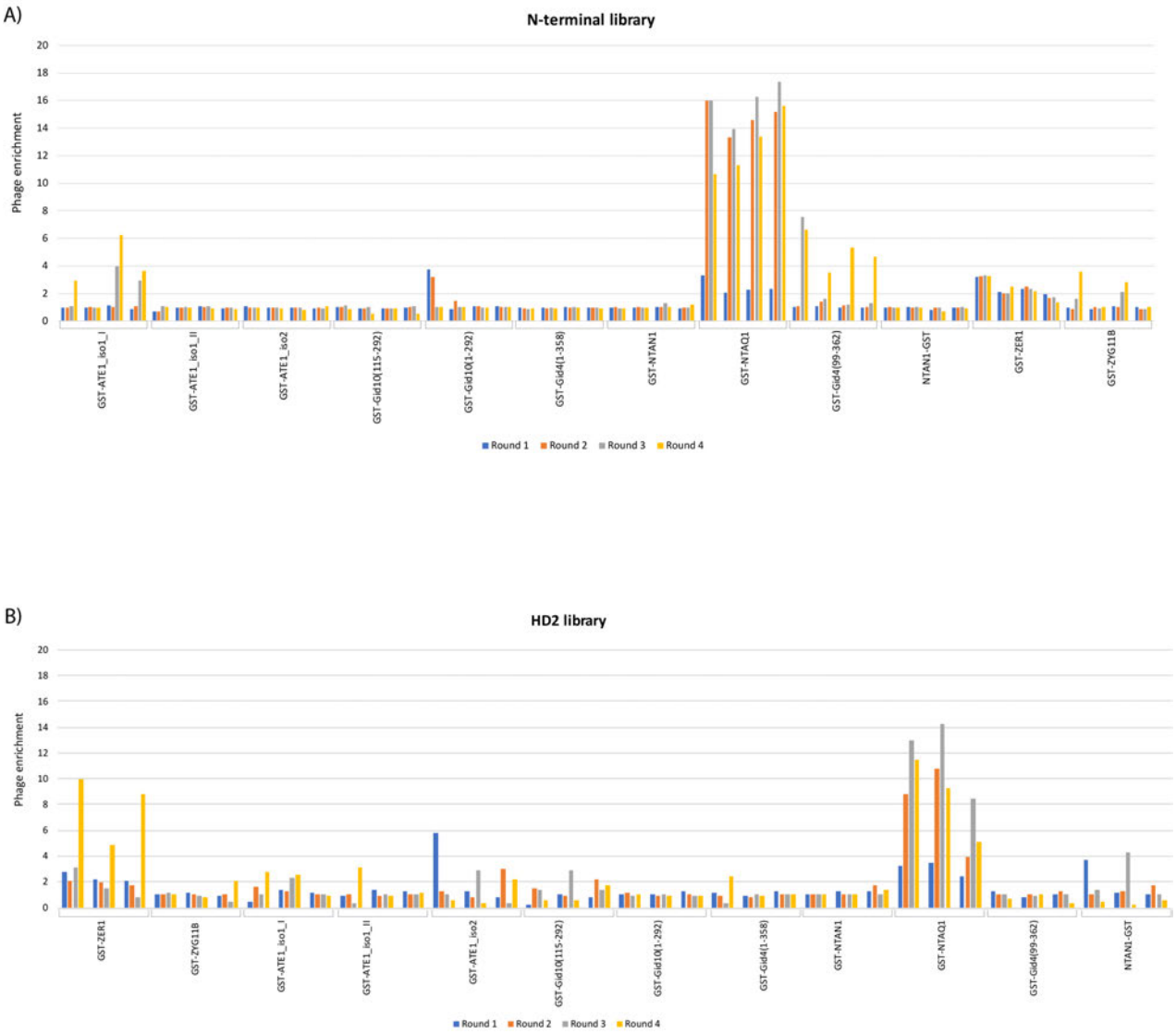


Figure 25. ELISA results of libraries, representing phage enrichment of proteins of interest. The threshold was set to require at least twice the number of phages detected in the sample compared to a negative control containing only GST. A) Phage enrichment of N-terminal library for twelve full or truncated proteins. B) Phage enrichment of HD2 library for twelve full or truncated proteins.

Next Generation Sequencing results for both libraries showed binders from which two gave motifs. After analysis done by [redacted], results show that with N-terminal library we got 14 binders for GST-NTAQ1 and 7 binders for GST-ATE1_iso1_I (Table 21A and B, Supplementary Table 22). With HD2 library for GST-NTAQ1 we got 110 binders and for GST-ATE1_iso1_I 66 binders (Table 21C and D, Supplementary Table 22). Interestingly enough none of the found binders to GST-NTAQ1 from the N-terminal library are starting with Q at the N-terminus. While a few are having Q in the third or fourth position. In the HD2 library we found only four peptides starting with Q at the N-terminus: QGLWLDVHALGELQRQ (Zinc finger protein 436), QWEMETVYSNSEVRNL (Centriolar coiled-coil protein of 110 kDa), QPLQYNLSICPPLLHG (R3H domain-containing protein 2) and QGLWLDVHALGELQRQ (Alanine--tRNA ligase, mitochondrial). Since this library is not made like an N-terminal library with a focus on the N-terminus it is hard to say what the importance of the Q is in this context. On the other hand, binders of GST-ATE1_iso1_I from the N-terminal library did not confirm expectation that peptides with N-

terminal E or D will bind. From 7 peptides 3 of them were staring with A, and none of them with E or D. In the HD2 library, we found two peptides starting with E, EKMLAAKSADGSAPAG (Large neutral amino acids transporter small subunit 1) and ERTHPVPMGTPTPLEP (Zinc finger protein 668). Also, we found two peptides starting with D, DLTEPTQPTRNQCCSN (Ras-related protein Rab-5A), and DTVDTIALSLCGGLAD (Phosphatidate phosphatase LPIN3).

A) GST-NTAQ1 baits from N-terminal library

	Protein	Hit	Specificity	Counts
1	Myosin-9	AQQAADKYLYV	0.995	0.080
2	Small integral membrane protein 42	SSPQLPAFLWD	0.979	0.025
3	Tumor necrosis factor alpha-induced protein 3	MAEQVLPQALYL	0.998	0.018
4	Erythropoietin	MGVHECPAWLWL	0.992	0.015
5	Protein PHTF2	ASKVTDAIVWY	0.997	0.009
6	ADAMTS-like protein 3	ASWTSPWWVLI	0.358	0.008
7	Putative protein PLEKHA9	SELRLCCDLLV	0.996	0.006
8	BPI fold-containing family C protein	CTKTIPVLWGC	0.931	0.003
9	M1-specific T cell receptor beta chain	SNQVLCCVVLC	0.955	0.003
10	Signal peptide peptidase-like 3	AEQTYSWAYSL	1	0.002
11	Tumor necrosis factor alpha-induced protein 3	AEQVLPQALYL	1	0.001
12	Myosin-9	MAQQAADKYLYV	1	0.001
13	GS homeobox 2	SRSFYVDSLII	1	0.001
14	Signal peptide peptidase-like 3	MAEQTYSWAYSL	0.978	0.001

B) GST-ATE1 baits from N-terminal library

	Protein	Hit	Specificity	Counts
1	Transmembrane 9 superfamily member 4	ATAMDWLPWSL	0.97	0.880
2	ADAMTS-like protein 3	ASWTSPWWVLI	0.28	0.004
3	ADAMTS-like protein 3	MASWTSPWWVLI	0.53	0.004
4	V-type proton ATPase subunit C 1	TEFWLISAPGE	0.70	0.002
5	Interleukin-4 receptor subunit alpha	GWLCSGLLFPV	0.93	0.002
6	Putative killer cell immunoglobulin-like receptor	APKLITVLCLG	0.72	0.001
7	Interleukin-4 receptor subunit alpha	MGWLCSGLLFPV	0.37	0.001

C) GST-NTAQ1 baits from HD2 library

	Protein	Hit	Specificity	Counts
1	ARF GTPase-activating protein GIT1	NDAVWLATQNHST	0.701	0.1359
2	Mitogen-activated protein kinase kinase kinase kinase 4	IDPRLLQISPSSGTTV	0.408	0.0684
3	Cytoplasmic tRNA 2-thiolation protein 1	DDMAETVLMNFLR	0.991	0.0487
4	Probable E3 ubiquitin-protein ligase IRF2BPL	NGDLNLQVAPPPPS	0.942	0.0373
5	G patch domain-containing protein 3	DWDVDMSVYYDRI	0.862	0.0282
6	Period circadian protein homolog 3	SGSSDSSIYLTSSVYS	0.950	0.0245
7	Storkhead-box protein 1	MEAESIYINDPTVKP	0.995	0.0193
8	Signal transducing adapter molecule 2	LELYNKLVNEAPVYS	0.929	0.0145
9	Zinc finger CCCH domain-containing protein 15	VSVNDIDLSLYIPRDV	0.931	0.0143
10	E3 ubiquitin-protein ligase NEDD4	WEERQDILGRTYYV	0.380	0.0093

D) GST-ATE1 baits from HD2 library

	Protein	Hit	Specificity	Counts
1	Dynein heavy chain 6, axonemal	YYDSFDTFIRTQFDD	1	0.1538
2	GRIP1-associated protein 1	ALSEEEFQRMQAQL	0.999	0.0593
3	Emerin	YADLSDTELTTLLRRY	0.890	0.0453
4	Coiled-coil domain-containing protein 117	ITEAELCAGPNDWI	0.968	0.0330
5	Spermatogenesis-associated protein 31A3	MENLPFPLKLLSASSI	0.947	0.0137
6	Gap junction delta-3 protein	HRAGKEAGGAEAAA	0.877	0.0062
7	Protein BANP	SEGNLQIHHVGQDG	0.271	0.0053
8	Kinase suppressor of Ras 1	NAFNISAFAHAAPLF	0.909	0.0051
9	FYVE, RhoGEF and PH domain-containing protein 1	NRSNRVCTDCYVAL	0.963	0.0036
10	RING finger protein 10	SDGESDNSDRVPVP	0.820	0.0035

Table 21. A) List of phage display sequences from N-terminal library binding GST-NTAQ1. B) List of phage display sequences from N-terminal library binding GST-ATE1. C) List of first 10 phage display sequences from HD2 library binding GST-NTAQ1. D) List of first 10 phage display sequences from N-terminal binding GST-ATE1. Hit is the identified peptide. Specificity is calculated as the proportion for the current bait of the total NGS read counts of the peptide for all baits screened with this library. Counts is calculated as the fraction of NGS counts for a selection

The motifs from the screen is extracted using SLiMFinder (Edwards, Davey, and Shields 2007). Interestingly, with additional proteins I used to select against N-terminal library I got two motifs. First motif for human protein GST-ZER1(519-766) is [AGS].....[ILV], the logo consists of 48 peptides out of 76 tested in the ProP-PD screen, yielding a coverage of 0.64 and a specificity of 0.007 (Figure 26A). The second motif is found for GST-Gid4(99-362), and it is [GS].....LL. The logo consists of 16 peptides out of 114 tested in the ProP-PD screen, yielding a coverage of 0.13 and a specificity of 0.005 (Figure 26B). Interestingly, the motif for ZER1 does not containng G on the first but on the second position which was unexpected since ZER1 should recognize G at the protein N-terminus (R. T. Timms et al. 2019b). On the other hand, even GID4 is a recognition subunit that should recognize P at the protein N-terminus, or the motif is not containing P. I compared peptides bound to GST-Gid4(99-362) from the N-terminal library with peptides bound to GST-yGid4(115-362) from the phage-displayed library, which consisted of 3.5×10^9 random octapeptides, as reported by Chrustowicz et al. in their 2022 study. This comparison revealed no

similarities between the peptides in the two libraries. Also in my ProP-PD experiment in this study didn't find binders containing P at the N-terminus.

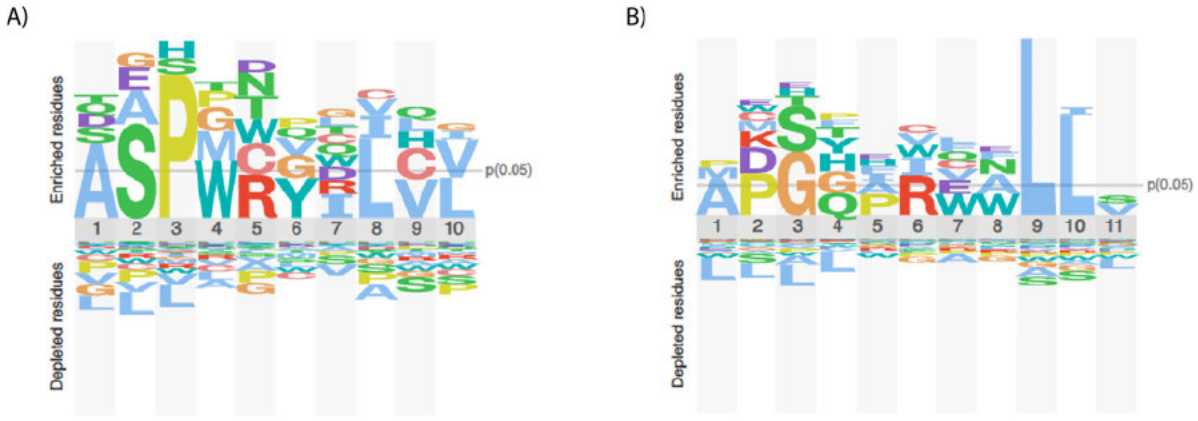


Figure 26. Potential binding motif found from ProP-PD screen using SLiMFinder. A) Motif found in interaction between N-terminal library and GST-ZER1(519-766). B) Motif found in interaction between N-terminal library and GST-Gid4(99-362). X-axis represents the position of amino acid and y-axis represents the relative frequency of amino acid. Positive y-axis showcases enriched amino acid. Numbers are indicating position of an amino acid from the N-terminus where 1 represents the closes tone to the N-terminus.

Discussion

Protein degradation is crucial for maintaining proteostasis and ensuring proper cellular function. To facilitate this process, cells have evolved various mechanisms, including the N-degron pathways, which target proteins for degradation based on the presence of destabilizing N-terminal residues. While much is known about these pathways, recent studies have revealed that their specificity is more complex than previously thought (Kats et al. 2018). The critical role of these pathways in proteostasis and this discovery opened up new doors for further investigation into N-degron pathways.

The Advantage of eK-sfGFP-P2A-mCherry Reporter

Determining the most effective tool for measuring peptide stability is crucial for obtaining the most accurate data. As outlined in the results, five different reporters were tested to identify the one that measures stability with the highest precision. Among these, the linker-eGFP-P2A-mCherry and linker-eGFP-IRES-dsRed reporters exhibited the highest dynamic range in measuring peptide stability.

Reporters containing the eK sequence did not show a high dynamic range compared to those with the linker sequence. This indicates that while the eK sequence can be effectively used in yeast studies, it is not efficient enough for use in human cell studies. Furthermore, sfGFP has been used in yeast studies to measure peptide stability, but it appears to be less suitable for human cells. The reason for this can be that uncomplete degradation of sfGFP, previously shown in yeast study, can potentially have an high effect in human cells (Khmelinskii et al. 2016).

A previous study by Timms et al. 2019 used the linker-eGFP-IRES-dsRed reporter. They expressed library included 24 amino acids of all human proteins fused with this reporter. This reporter displayed a similar pattern to our findings, except for T, that appeared more unstable in our data. On the other hand, the pattern observed with the linker-eGFP-P2A-mCherry reporter was even more consistent with the Timms et al. 2019 study results than linker-eGFP-IRES-dsRed reporter.

To decide which reporter was better for our further experiments we considered the differences of IRES and P2A sequence. Since the IRES is an additional ribosomal sight proteins may be not translated in the same amount which can be a problem for later reporting stability. This was the main reason why I chose linker-eGFP-P2A-mCherry for further studies.

Deviations in Arg/N-Degron Pathway

To better understand N-degron pathways I wanted to focus my research on the first two amino acids at the N-terminus and get better insights into their stability. For that, I chose to use eK-sfGFP-P2A-mCherry reporter which in my previous experiments showed up to be the most advantageous reporter for spotting the smallest differences in protein stability. I produced the Ubi-XZ-linker-eGFP-P2A-mCherry library named Ubi-XZ library, containing all the variants of 20 amino acids at the first two positions. The library showed very high reproducibility and data quality (Figure 15 and Figure 18) as evidenced by the high consistency across three replicates and the relatively low percentage of stop codons. Furthermore, the stability of fourteen constructs containing specific dipeptides was compared with the stability of the same dipeptides in the Ubi-XZ library, revealing a strong correlation.

The Ubi-XZ library revealed that dipeptides starting with C, V, G, T or P are relatively stable which can be explained through the Ac/N-degron pathway. Through this pathway, those amino acids are targeted for acetylation by NatA after cleavage of first methionine and later ubiquitinated by MARCH6 and degraded by the proteasome. On the other hand, M at the N-terminus proves to be very stable, consistent with findings by Timms et al. 2019. Their study demonstrated that the presence of an initial methionine increases the stability of the N-terminus compared to when it is absent. Besides that S and A were also shown to be very stable which is inconsistent with the Ac/N-degron pathway. The one explanation can be that their NatA can recognize them because to their small size and lack of charge, or their initial methionine is never cleaved. However, these hypotheses need further testing.

Besides that, in the study done in yeast showed that residues expected to be targeted for degradation by the Ac/N-degron pathway were significantly more stabilized compared to other dipeptides. This was explained by the observation that the Ac/N-degron pathway is not strictly followed and that hydrophobicity has a greater influence on protein stability (Kats et al. 2018).

The results from the Ubi-XZ library indicated that dipeptides starting with R, K, H, W, L, F, Y, and I are generally very unstable, aligning perfectly with the primary destabilizing residues of the Arg/N-degron pathway. In contrast, the secondary destabilizing residue E and the tertiary

destabilizing residue Q are stable, does not align with the human Arg/N-degron pathway (Varshavsky 2019). This discrepancy was also observed in a 2019 study conducted in human cells, but not in the study conducted in yeast in 2018. This deviation raised several hypotheses regarding its cause. In my further experiments, I aimed to address these hypotheses.

Possible Explanations for deviation in Arg/N-degron pathway

In order to deeper understand the deviation in stability of dipeptides beginning with E or Q, I wanted to test a few different hypotheses. First, these deviations could be due to the differential expression of key enzymes in the Arg/N-degron pathway. Some components might be under-expressed in the specific cell line, impairing the pathway's functionality. To see if this is the case with transcription levels of ATE1, NTAN1, NTAQ1, UBR1, UBR2, UBR4, and UBR5 I compared them with transcription levels of other genes in the HEK293 cell line. I showed that transcripts per million (TPM) values are all higher than the median of other genes, which indicated that expressed in this cell line. This result shows that most probably deviation stability of dipeptides starting with E and Q is not due to the lack of proteins that are participating in detecting those dipeptides for degradation.

Another hypothesis focuses on the substrate specificity of enzymes, particularly NTAQ1, which can target N-terminal Q for deamidation. If this enzyme is not specific enough it can explain our observed deviation in dipeptide stability. To test this, we expressed human NTAN1 and NTAQ1 in yeast cells lacking the Nta1 enzyme. Compared to cells where Nta1 was reintroduced, cells with expressed NTAN1 stabilized dipeptides starting with D, E, and N, but Q. Conversely, cells with expressed NTAQ1 stabilized dipeptides beginning with D, E, and Q, but not N. This indicates that NTAN1 and NTAQ1 are specific to their substrates. This is logical as, despite NTAN1 and NTAQ1 both possessing a conserved cysteine (C) residue crucial for deamination activity, their sequences differ significantly. This also suggests that, from an evolutionary perspective, they were incorporated into the Arg/N-degron pathway independently (M. S. Park et al. 2014).

The yeast study showed that acetylation of N-terminal residues can prevent degradation by the Arg/N-degron pathway (Kats et al. 2018). I wanted to investigate whether this mechanism is present in human cells as well and if that can be the explanation for unexpected stability in our results. To test this I expressed 12 different constructs into cells lacking functional Ate1 gene. Construct contained dipeptides beginning with N, Q, E, or D and as a second residues L, R, or D. My results showed that expressed dipeptides indeed undergo acetylation and even match the stability pattern observed in the Ubi-XZ library. This confirms that acetylation prevents the Arg/N-degron pathway from targeting the N-terminus starting with Q and N at the N-terminus for degradation. Previous studies in plant species *Arabidopsis thaliana* have shown that N-terminal acetylation by NatA is crucial for preventing degradation of a major subset of cytosolic proteins by masking of N-degron motif. Similarly, in the human cell lines, N-terminal acetylation by NatA shields a set of proteins containing N-terminal IAP binding motif-like sequences from binding to

IAP ubiquitin ligases (Linster et al. 2022). Like this, it regulates protein stability and prevents apoptosis (Mueller et al. 2021).Besides this, a recent study showed that NatC can acetylate group of cytosolic proteins and shield them from degradation by masking a motif that can be recognized by UBR4-KCMF1 complex (Varland et al. 2023). Interestingly with all of this, we see that N-terminal acetylation is not promoting degradation but in some cases preventing degradation as well.

The next step would be to understand which of NAT enzymes can acetylate N-terminal E or D. A new NAT enzyme named NatH was shown to detect the N-terminus of actin starting with EEEI or DDDI for post-translationally acetylation (Aksnes, Ree, and Arnesen 2019; Drazic et al. 2018). The next steps would be to do a knockdown of NatH and check the stability of dipeptides starting with N-terminal E or D.

Binding of NTAN1, NTAQ1, and ATE1 to a Peptide Library: A Proteomic Peptide-Phage Display Approach

To find out which sequences on the N-terminus of a human protein can bind to NTAN1, NTAQ1, or ATE1 I employed Proteomic Peptide Phage Display (ProP-PD). Even other approaches can be used to detect protein-protein interactions (PPI) like large-scale yeast-two-hybrid (Y2H) or high-throughput affinity-purification coupled to mass spectrometry (AP-MS), ProP-PD is much better at finding shorter sequences (Davey et al. 2017b; Huttlin et al. 2015; Walhout and Vidal 2001). I made an N-terminal ProP-PD library with all isoforms of all human proteins with and without initiator methionine, with cleavage sites and microbial infection cleavage sites. By utilizing ProP-PD we found 14 binders for NTAQ1 and 7 binders for ATE1 isoform 1. Binders did not show any common motif, and interestingly enough 14 binders of NTAQ1 did not start with Q as expected, but some peptides had Q in the third or fourth position. Similarly, ATE1 isoform 1 did have peptides that start with E or D, but mostly with A or M. It would be interesting to know if those peptides are maybe new degrons or if they may have some different roles. The easiest way to find this out would be to test the stability of those peptides by GPS in cells lacking NTAQ1 or ATE1.

Besides the N-terminal library, I also tested the HD2 library against NTAN1, NTAQ1, and ATE1. This library contains fragmented disordered regions. With the HD2 library, I also got binders for NTAQ1 and ATE1 isoform 1, but not for NTAN1. With this library, I got 110 binders for NTAQ1 and 66 for ATE1 isoform 1. The higher number of binders can be explained also by the fact that the HD2 library is more than 10 times bigger than the N-terminal library.

Additionally, with NTAN1, NTAQ1, and ATE1 I tested the different lengths of two yeast proteins, Gid4 and Gid10, and the full length of two human proteins ZER1 and ZYGB11. Interestingly, by testing them against the N-terminal library I seceded to find two motifs, one for yeast Gid4 and another for ZER1. Yeast Gid4 is a recognition subunit of the GID complex, which can target P at the protein N-terminus. On the other hand, the motif that we found doesn’t contain P and does not overlap with the peptides discovered in a recent similar study (Chrustowicz et al. 2022a). This

showed us that Gid4 can potentially have substrates other than one starting with P at the N-terminus. Besides this, we also found the motif for ZER1 that is [AGS].....[ILV]. Even though it was shown that ZER1 can target G at the protein N-terminus for degradation we found that in our motif G is in a second position after A. It is possible that this motif is not related to degradation and it has some other function. As previously mentioned this can be checked by GPS technique in the future.

Conclusion

Through this thesis, I aimed to explore the specificity of the Arg/N-degron pathway in human cells. By applying the GPS technique, I found an interesting deviation from this pathway and tried to understand which mechanism in human cells is responsible for it. By digging deeper into the subject I found that the unexpected higher stability of Q and E on the protein N-terminus can be explained by a mechanism in which acetylation prevents their recognition by the Arg/N-degron pathway. I excluded the possibility that enzymes that are part of the Arg/N-degron pathway are not expressed by comparing gene expression to other genes in the cells. Besides that, by using yeast as a model system I confirmed that NTAQ1 and NTAN1 evolved to have specifically target Q or N residues N-terminus, respectively. Again excluding this as an option in the explanation the mentioned deviation in the Arg/N-degron pathway. All of this points to the direction that acetylation prevents degradation of the Arg/N-degron pathway. For future study, it would be interesting to find out which of the NAT enzymes is responsible for this acetylation.

Supplementary

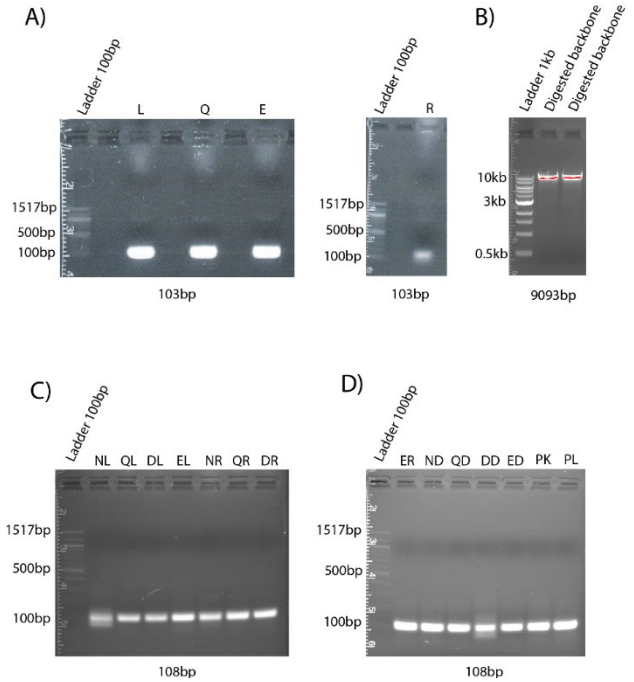


Figure 27. DNA fragments separated by agarose gel electrophoresis. Expected band sizes were written at the bottom of the gels. A) Amplified inserts containing particular single residue. B) Backbone linearized by restriction digestion. C) and D) Amplified inserts containing particular dipeptides.

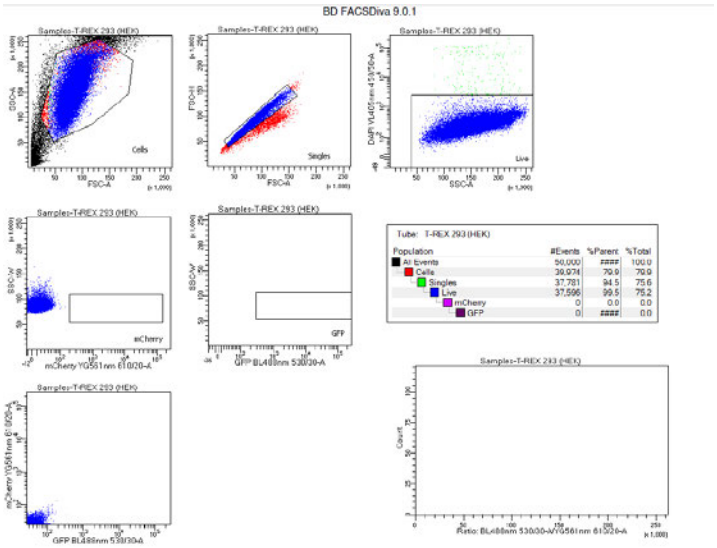


Figure 28. Gating of Flip-In T-Rex cell line by Flow Cytometry with different gates. From left to right population of this cell line, single cells, live cells, cells with mCherry fluorescence and cells with eGFP fluorescence. Next is the table with all gates and with number of events, percentages of the cells in each gate and percentage of total cells. Last two plots are showing from left to the right ratio between mCherry and eGFP and another count of cells and ratio between mCherry and eGFP

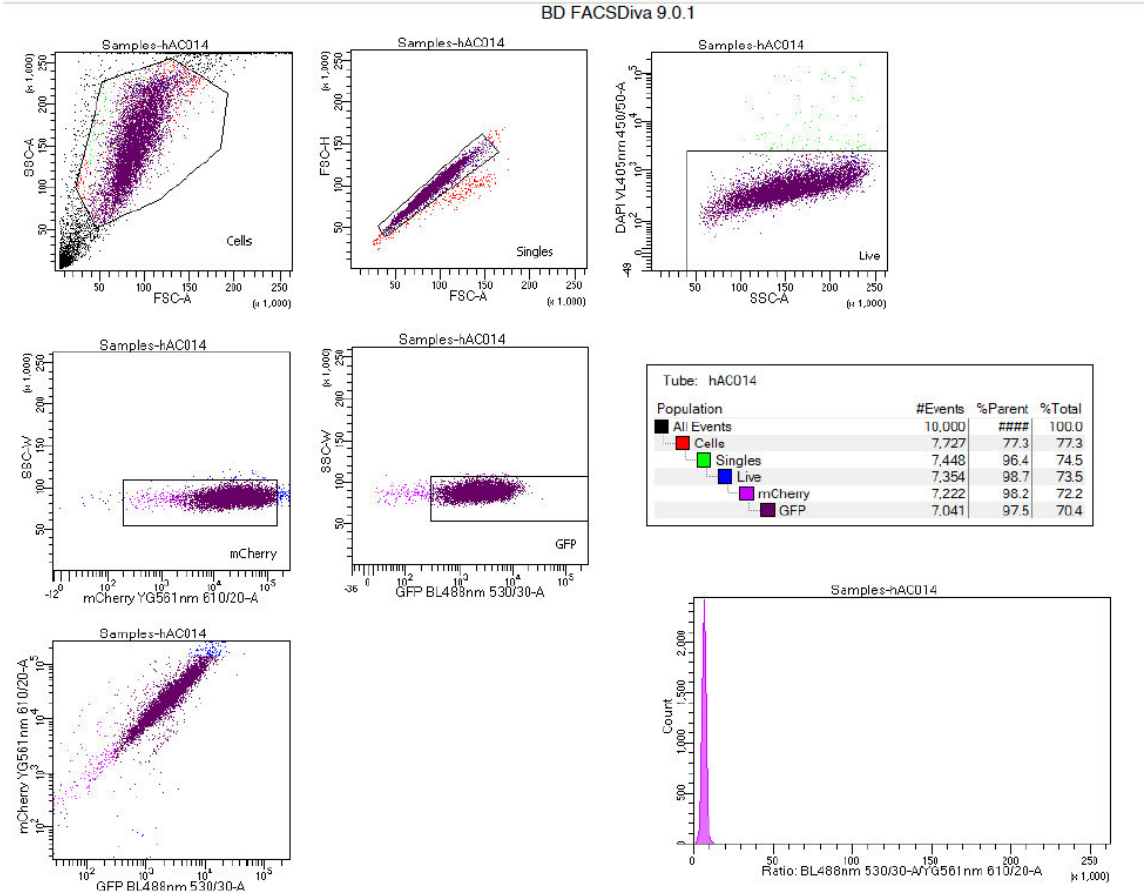
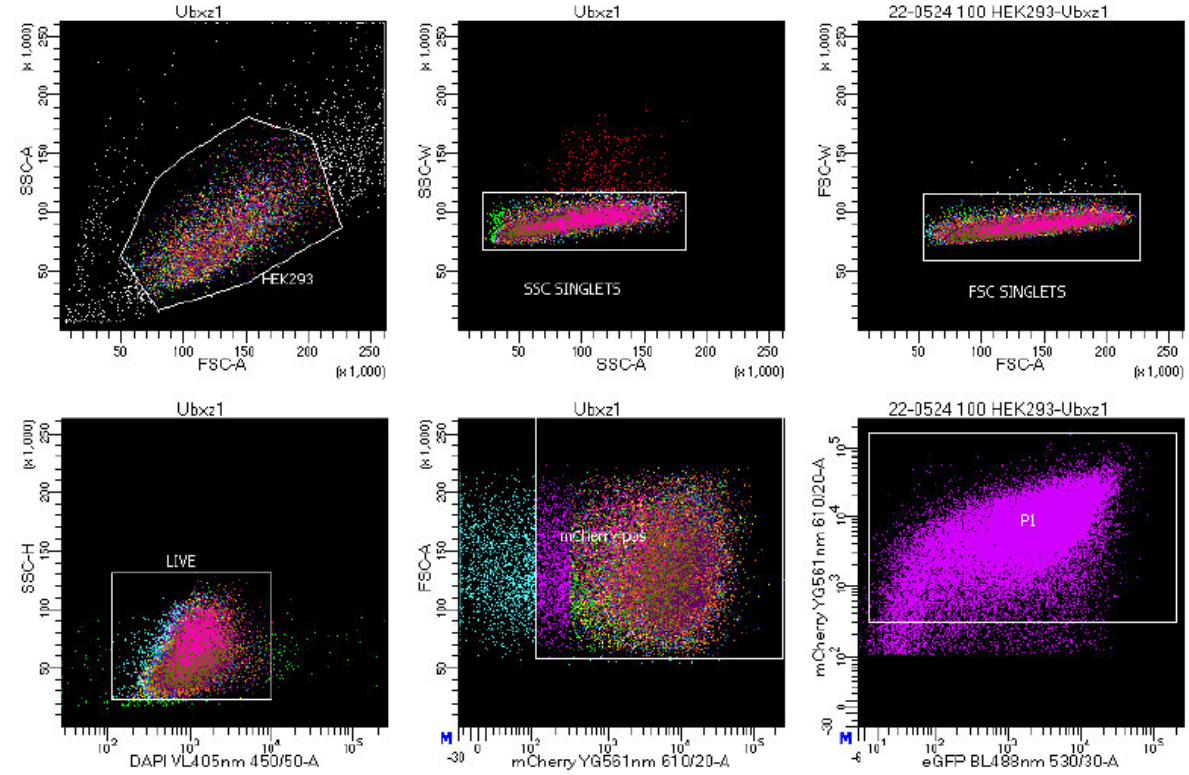


Figure 29. Gating of Flip-In T-Rex cell line with expressed Ubi-ER-ek-eGFP-P2A-mCherry constract by Flow Cytometry with different gates. From left to right population of this cell line, single cells, live cells, cells with mCherry fluorescence and cells with eGFP fluorescence. Next is the table with all gates and with number of events, percentages of the cells in each gate and percentage of total cells. Last two plots are showing from left to the right ratio between mCherry and eGFP and another count of cells and ratio between mCherry and eGFP.



Tube: Ubxz1

Population	#Events	%Parent	%Total
All Events	50,000	####	100.0
HEK293	42,455	84.9	84.9
SSC SINGLETS	42,168	99.3	84.3
FSC SINGLETS	42,126	99.9	84.3
LIVE	41,919	99.5	83.8
mCherry pos	40,128	95.7	80.3
P1	38,620	96.2	77.2
1	2,511	6.5	5.0
2	3,309	8.6	6.6
3	4,682	12.1	9.4
4	6,497	16.8	13.0
5	8,670	22.4	17.3
6	8,580	22.2	17.2
7	3,596	9.3	7.2
8	1,509	3.9	3.0

Figure 30. Gating of Ubi-XZ library by sorter with different gates by BD FACSAria III SORP. From left to right population of this cell line, single cells, live cells, cells with mCherry fluorescence and cells with eGFP fluorescence. Next is the table with all gates and with number of events, percentages of the cells in each gate and percentage of total cells.

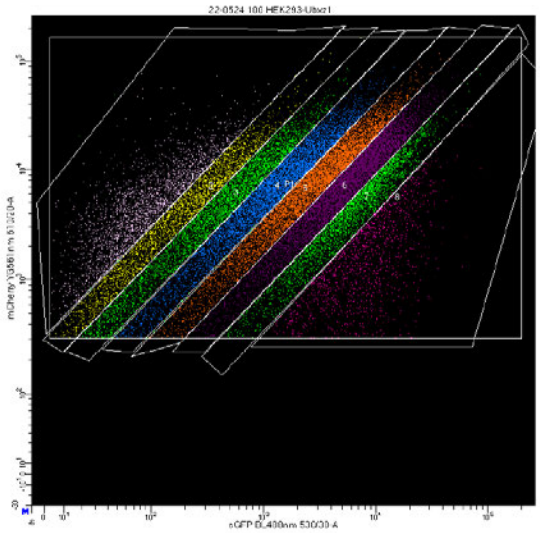


Figure 31. Eight gates made in order to sort cells in eight different bins based on their eGFP to mCherry ratio. Six bins in the middle are the same size and first and last gate group all other cells.

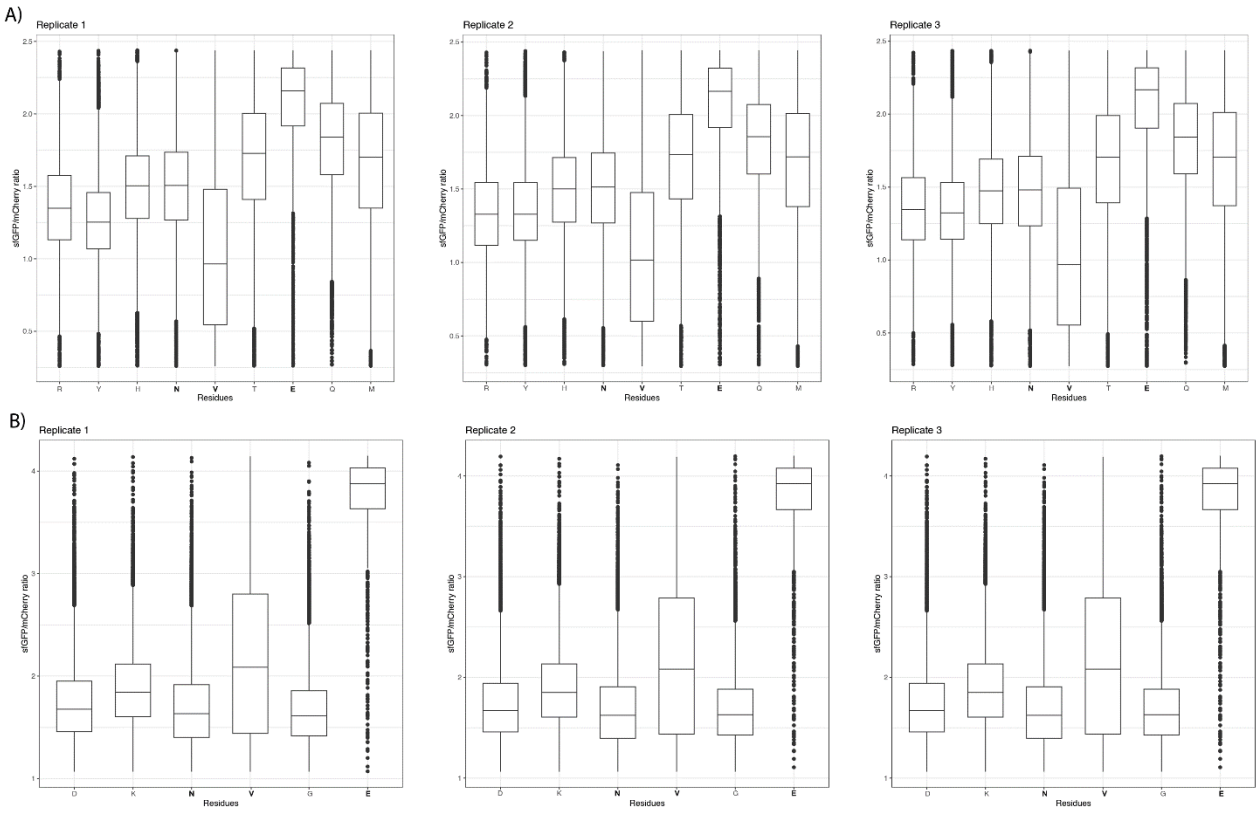


Figure 32. Stability analysis of residues fused with Ubi-ek-sfGFP-P2A-mCherry reporter. To enhance clarity, 10% of the highest and lowest data points have been excluded in all plots. A) The ratio of sfGFP to mCherry fluorescence depicted for nine specific constructs. Each plot represents one of three replicates. B) The ratio of sfGFP to mCherry fluorescence depicted for six specific constructs. Each plot represents one of three replicates

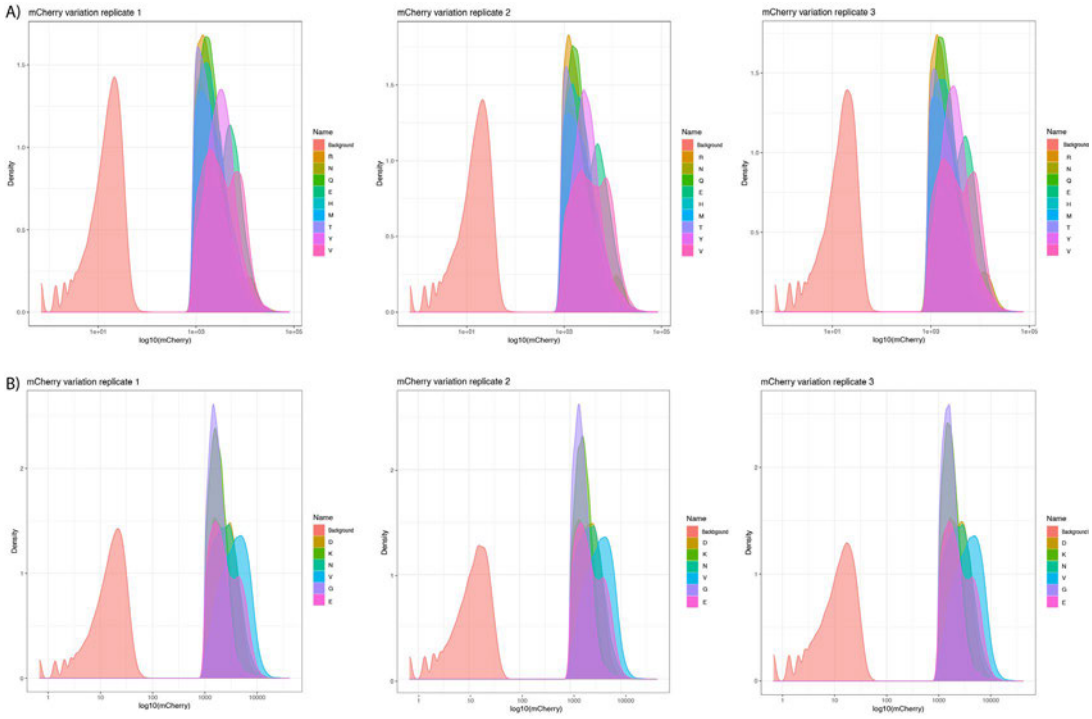


Figure 33. Histogram displaying mCherry density of different constructs in comparison to the background for single residues fused with Ubi-ek-sfGFP-P2A-mCherry reporter. A) mCherry density of nine constructs for three replicates B) mCherry density of six constructs for three replicates

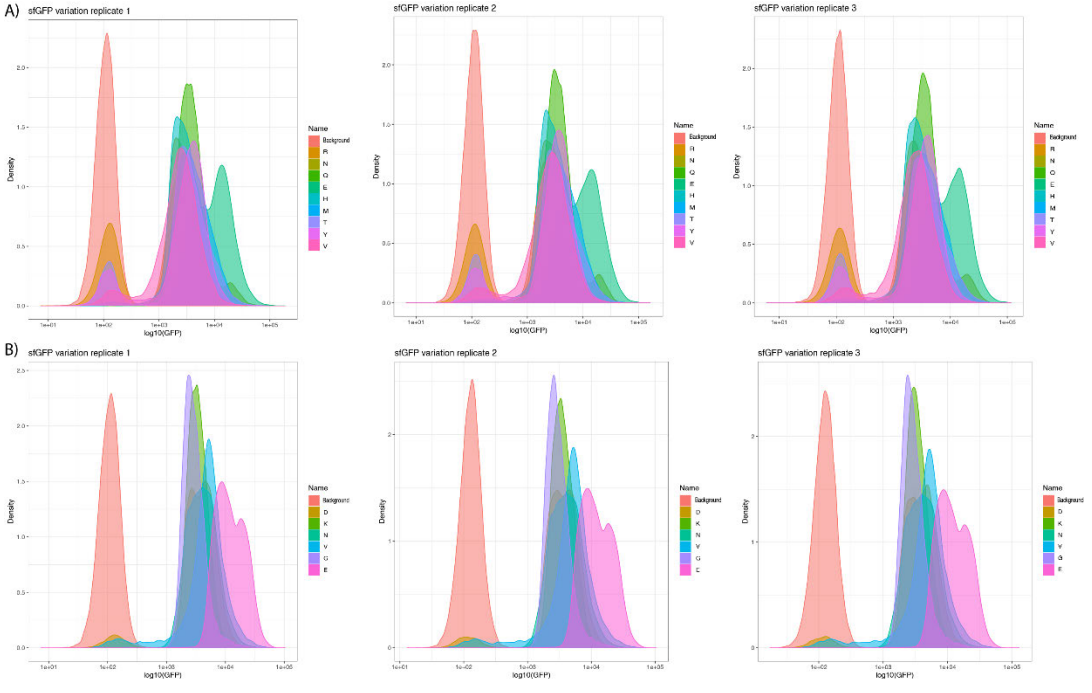


Figure 34. Histogram displaying sfGFP density of different constructs in comparison to the background for single residues fused with Ubi-ek-sfGFP-P2A-mCherry reporter. A) sfGFP density of nine constructs for three replicates B) sfGFP density of six constructs for three replicates

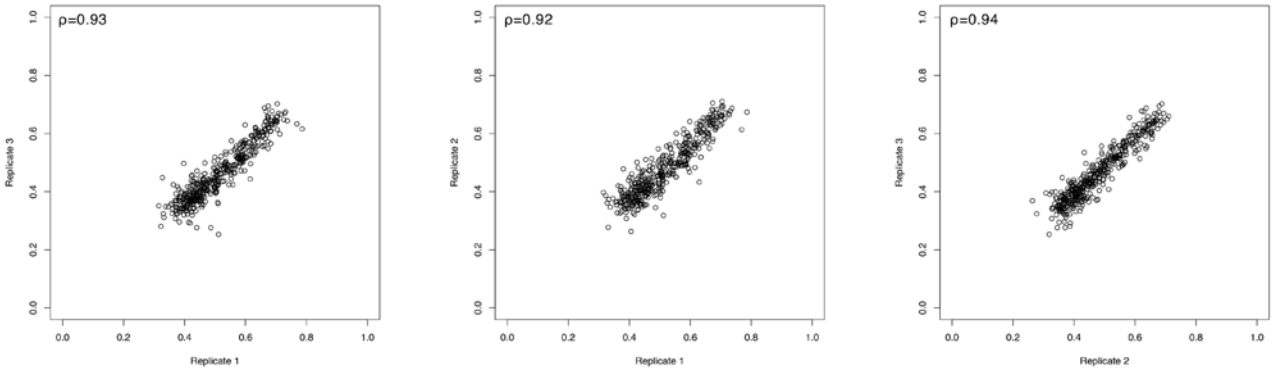


Figure 35. Correlation of pooled PSI between three replicates of Ubi-XZ library. Correlation is calculated using Pearson correlation.

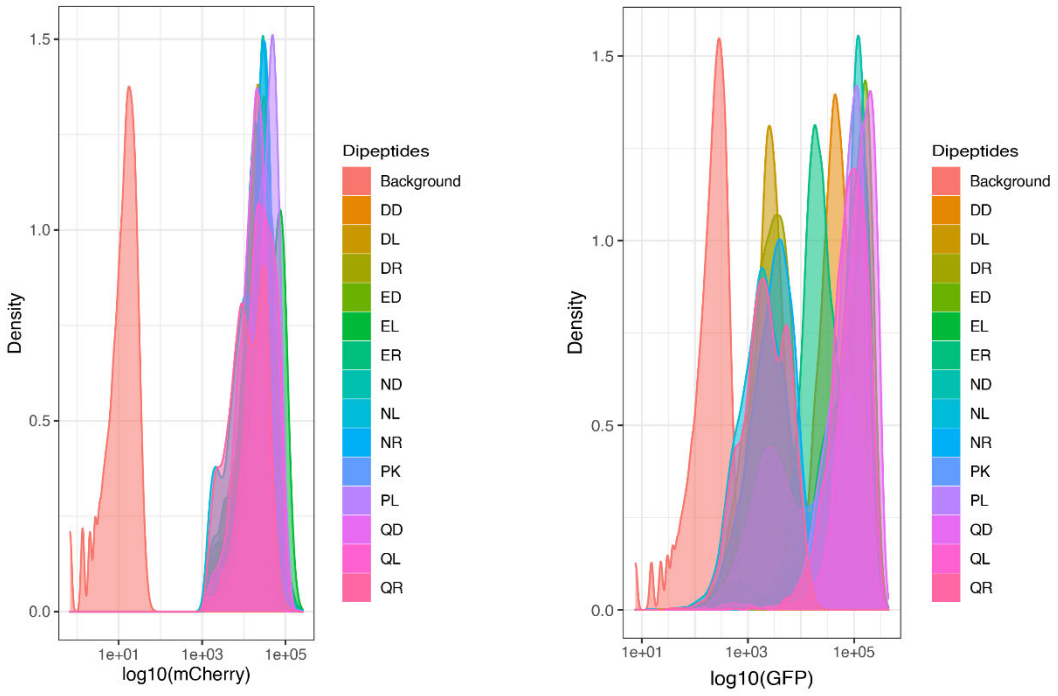


Figure 36. Histograms displaying mCherry or eGFP density of different 14 constructs containing different dipeptides in comparison to the background. X-axis represents density and y-axis log10 of mCherry or eGFP

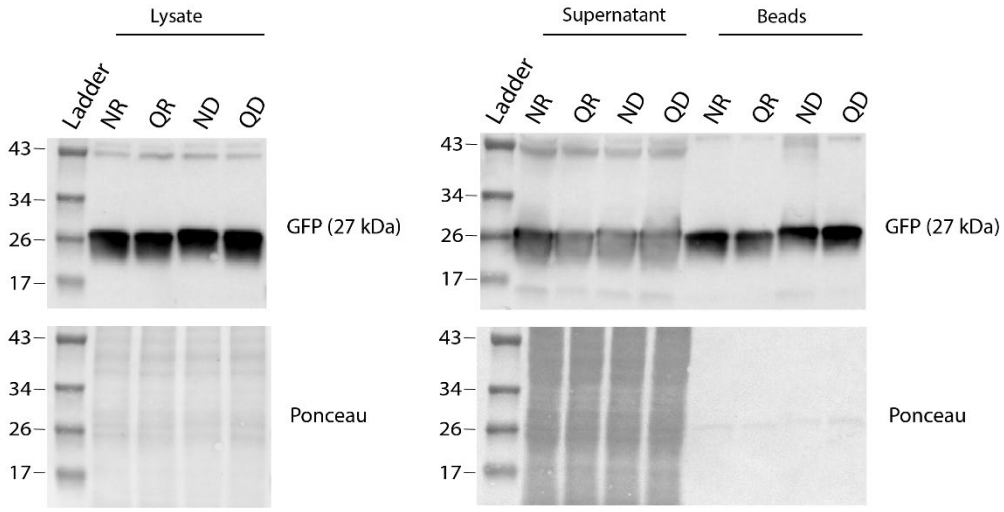


Figure 37. Detection of the different dipeptides conjugated with GFP on a 12% SDS-PAGE. Samples were taken from three different parts of the immunoprecipitation protocol, from the whole cell lysate, supernatant after sample was incubated with magnetic beads, and sample bind to the magnetic beads.

Table 22. List of phage display sequences from N-terminal library or HD2 library binding GST-NTAQ1 or GST-ATE1 proteins. Hit is the identified peptide. Specificity is calculated as the proportion for the current bait of the total NGS read counts of the peptide for all baits screened with this library. Counts is calculated as the fraction of NGS counts for a selection

Protein	Gene name	Hit	Specificity	Counts
GST-NTAQ1 baits from N-terminal library				
Myosin-9	MYH9	AQQAADKYLYV	0.9949	0.0795
Small integral membrane protein 42	SMIM42	SSPQLPAFLWD	0.9789	0.0251
Tumor necrosis factor alpha-induced protein 3	TNFAIP3	MAEQVLPQALYL	0.9977	0.0181
Erythropoietin	EPO	MGVHECPAWLWL	0.9923	0.0146
Protein PHTF2	PHTF2	ASKVTDAIVWY	0.9968	0.0094
ADAMTS-like protein 3	ADAMTSL3	ASWTSPWWVLI	0.3582	0.0082
Putative protein PLEKHA9	PLEKHA9	SELRLCCDLLV	0.9958	0.0065
BPI fold-containing family C protein	BPIFC	CTKTIPVLWGC	0.9313	0.0034
M1-specific T cell receptor beta chain	TRB	SNQVLCCVVLC	0.9551	0.0028
Signal peptide peptidase-like 3	SPPL3	AEQTYSWAYSL	1.0000	0.0019
Tumor necrosis factor alpha-induced protein 3	TNFAIP3	AEQVLPQALYL	1.0000	0.0010
Myosin-9	MYH9	MAQQAADKYLYV	1.0000	0.0009
GS homeobox 2	GSX2	SRSFYVDSLII	1.0000	0.0006
Signal peptide peptidase-like 3	SPPL3	MAEQTYSWAYSL	0.9785	0.0005
GST-ATE1_iso1_I baits from N-terminal library				
Transmembrane 9 superfamily member 4	TM9SF4	ATAMDWLPWSL	0.9656	0.8801
ADAMTS-like protein 3	ADAMTSL3	ASWTSPWWVLI	0.2793	0.0038
ADAMTS-like protein 3	ADAMTSL3	MASWTSPWWVLI	0.5282	0.0038
V-type proton ATPase subunit C 1	ATP6V1C1	TEFWLISAPGE	0.7040	0.0024
Interleukin-4 receptor subunit alpha	IL4R	GWLCSGLLFPV	0.9310	0.0019
Putative killer cell immunoglobulin-like receptor-like protein KIR3DX1	KIR3DX1	APKLITVLCLG	0.7193	0.0012
Interleukin-4 receptor subunit alpha	IL4R	MGWLCSGLLFPV	0.3696	0.0006
GST-NTAQ1 baits from HD2 library				
ARF GTPase-activating protein GIT1	GIT1	NDAVWLATQNHSTLVT	0.7012	0.1359

Mitogen-activated protein kinase kinase kinase kinase 4	MAP4K4	IDPRLLQISPSSGTTV	0.4079	0.0684
Cytoplasmic tRNA 2-thiolation protein 1	CTU1	DDMAETVLMNFLRGDA	0.9895	0.0487
Probable E3 ubiquitin-protein ligase IRF2BPL	IRF2BPL	NGDLNLQVAPPPPSAH	0.9418	0.0373
G patch domain-containing protein 3	GPATCH3	DWDVDMSVYYDRDGGD	0.8624	0.0282
Period circadian protein homolog 3	PER3	SGSSDSSIYLTSSVYS	0.9481	0.0245
Storkhead-box protein 1	STOX1	MEAESIYINDPTVKPI	0.9946	0.0193
Signal transducing adapter molecule 2	STAM2	LELYNKLVNEAPVYSV	0.9209	0.0145
Zinc finger CCCH domain-containing protein 15	ZC3H15	VSVNDIDLSLYIPRDV	0.9312	0.0143
E3 ubiquitin-protein ligase NEDD4	NEDD4	WEERQDILGRTYYVNH	0.3815	0.0093
Sorting nexin-15	SNX15	DGVHVLLQGVPSDPLP	0.6762	0.0076
CASP8-associated protein 2	CASP8AP2	IAVPSSEQEIMHMLRM	0.8260	0.0076
Bromodomain and WD repeat-containing protein 3	BRWD3	TMVAWDRYDTTVITAV	0.6234	0.0071
Transcription factor E2F6	E2F6	NGPIDVYLCEVEQGQT	0.6137	0.0067
Serine/threonine-protein kinase LATS2	LATS2	SLNVDLYELGSTSVQQ	0.4542	0.0067
Insulin receptor substrate 2	IRS2	CGGDSDQYVLMSSPVG	0.2016	0.0056
PH-interacting protein	PHIP	TMVAWDRHDNTVITAV	0.8285	0.0047
Solute carrier family 10 member 6	SLC10A6	VDGDMDLSISMTTCST	0.7340	0.0043
Adenylate kinase 9	AK9	TALREFQRQYEKMEFG	0.9864	0.0043
85/88 kDa calcium-independent phospholipase A2	PLA2G6	PAELHLFRNYDAPETV	0.9325	0.0042
Dysbindin	DTNBP1	ALDVFLNSGGEENTVL	0.8601	0.0042
Zinc finger protein 436	ZNF436	QWGDLTAEEWVSYPLQ	0.2425	0.0041
Integrator complex subunit 2	INTS2	VNPDSLLRNVQSVITT	0.7899	0.0041
DPY30 domain-containing protein 2	DPY30	MEMTEMLKQEEYQIQQ	0.9953	0.0040
Nance-Horan syndrome protein	NHS	SDSEWNYLHHHHDASC	0.6778	0.0036
Katanin-interacting protein	KATNIP	HLEQGFSVYVNGANSE	1.0000	0.0036
Protein SON	SON	TMDAQMLASGTMDAQM	0.8672	0.0034
Nuclear receptor coactivator 3	NCOA3	GNPAVYSMVHMNGSSG	0.9920	0.0033
Serine/threonine-protein kinase haspin	HASPIN	AVPSGLHLPEVSLDRA	0.9785	0.0031
Zinc finger and BTB domain-containing protein 40	ZBTB40	AVTVEDLLDGTVTLIC	0.6493	0.0029

IQCJ-SCHIP1 readthrough transcript protein	IQCJ-SCHIP1	SGMNLQICFVNDSGSD	0.7979	0.0026
MHC class II transactivator	CIITA	TLPHGLWQISEAGTGV	0.3521	0.0026
Protein piccolo	PCLO	TAVDLTSGRVTTGEVM	0.7532	0.0025
Pancreas transcription factor 1 subunit alpha	PTF1A	MDAVLLEHFPGGLDAF	0.4454	0.0025
Nephrocystin-4	NPHP4	MNDWHRIFTQNVLVPP	0.9192	0.0024
Zinc fingers and homeoboxes protein 3	ZHX3	IAQFLSLQQQPPVHAQ	0.5810	0.0022
Zinc finger and BTB domain-containing protein 44	ZBTB44	LENVQYPYQLYIAPST	0.9893	0.0021
Bromodomain-containing protein 9	BRD9	DGSLLYSVVNTAEPDA	0.9324	0.0018
Protein PML	PML	VIGSEVFLPNSNHVAS	0.8170	0.0017
Sarcoma antigen 1	SAGE1	AMSARDLYATVTHNVC	0.2312	0.0016
MAGE-like protein 2	MAGEL2	NLNLSDWEVQSPIQVS	0.9638	0.0016
Ubiquilin-2	UBQLN2	ITENPQLIQNMLSAPY	1.0000	0.0015
Host cell factor 2	HCFC2	HVKWDEVSTVEGYLLQ	0.5491	0.0015
Liprin-alpha-2	PPFIA2	LGDHEWNRTQQIGVLS	0.9643	0.0015
Synphilin-1	SNCAIP	LQQFLEAQKSEGKSLP	0.9000	0.0014
Nucleoporin NUP42	NUP42	MEVWESSGQWMFSVYS	0.6993	0.0014
Intermediate conductance calcium-activated potassium channel protein 4	KCNN4	HAKEVQLFMTDNGLRD	0.7838	0.0014
Spectrin alpha chain, erythrocytic 1	SPTA1	EANVQFQQYLADLHEA	0.7273	0.0014
Chromodomain-helicase-DNA-binding protein 7	CHD7	NAELGQLYWPNTSTLT	0.5776	0.0014
Centriolar coiled-coil protein of 110 kDa	CCC110	QWEMETVYSNSEVRNL	0.3639	0.0013
Synapsin-1	SYN1	SVDMEVLRNGVKVVRS	0.7832	0.0013
R3H domain-containing protein 2	R3HDM2	QPLQYNLSICPPLLHG	0.9717	0.0013
Ornithine decarboxylase antizyme 1	OAZ1	PDACFMAYTFERESSG	0.9800	0.0013
Potassium voltage-gated channel subfamily KQT member 2	KCNQ2	HEELERSFSGFSISQS	0.6538	0.0013
Breast cancer type 1 susceptibility protein	BRCA1	PKDLMLSGERVLQTER	0.7174	0.0012
Intersectin-1	ITSN1	ETDNWDAWAAQPSLTV	0.9149	0.0012
Cyclic AMP-dependent transcription factor ATF-6 alpha	ATF6	TKGQTVLLSQPTVVQL	0.8585	0.0012
Host cell factor 2	HCFC2	TVEGYLLQLSTDLPYQ	0.8200	0.0011

T-cell acute lymphocytic leukemia protein 1	TAL1	PPDDLLQDVLSPNSSC	0.9011	0.0011
Histone deacetylase complex subunit SAP25	SAP25	STAWLSGPELIALTGL	0.2736	0.0011
A-kinase anchor protein 13	AKAP13	SWSSLSYEIPYGDCSV	0.5745	0.0011
Protein angel homolog 1	ANGEL1	TEQWEEDPAVLAWSIA	0.4172	0.0010
Succinate dehydrogenase cytochrome b560 subunit, mitochondrial	SDHC	ESYLELVKSLCLGPAL	1.0000	0.0010
ATPase family AAA domain-containing protein 2B	ATAD2B	REEWNLSTGQARLTSQ	0.8000	0.0010
Rho guanine nucleotide exchange factor 16	ARHGEF16	DPQLYQEIQERGLNTS	0.6417	0.0010
SH3 domain-binding protein 5	SH3BP5	DSAWQEMLNHATQRVM	0.2549	0.0010
Vesicular inhibitory amino acid transporter	SLC32A1	ITAWEAGWNVTNAIQG	0.3203	0.0010
Plakophilin-3	PKP3	MQDGNFLLSALQPEAG	0.3966	0.0009
Cytoplasmic tyrosine-protein kinase BMX	BMX	CTLWEAYANLHTAVNE	0.8571	0.0009
Neuroblastoma breakpoint family member 1	NBPF1	ATNVSMVVSAGPLSSE	0.7875	0.0009
Aladin	AAAS	HDLPLFTETSPTSAPW	0.9036	0.0009
StAR-related lipid transfer protein 9	STARD9	PTSALFVDRASSPILT	1.0000	0.0009
Zinc finger protein 532	ZNF532	ECDCLFMQRDVYISHV	0.6364	0.0009
Transcriptional repressor protein YY1	YY1	MASGDTLYIATDGSEM	0.9531	0.0009
BAG family molecular chaperone regulator 4	BAG4	APPGNLYMTESTSPWP	1.0000	0.0008
Potassium voltage-gated channel subfamily H member 3	KCNH3	PWDPHSLEMVLIGCHG	0.2228	0.0008
Zinc finger homeobox protein 3	ZFHX3	EMVMPSSMFLPAAVPD	0.6701	0.0008
Protein mono-ADP-ribosyltransferase PARP10	PARP10	WQVAERVLQQEHRLQG	0.3609	0.0008
Potassium voltage-gated channel subfamily C member 3	KCNC3	SFLPDLNANAAAWISP	0.3436	0.0008
Protein phosphatase Slingshot homolog 2	SSH2	VSPDIFMQSHSENAIS	0.8028	0.0007
Inositol 1,4,5-triphosphate receptor associated 2	IRAG2	CPESLLQSREYSSLPL	0.4240	0.0007
Ubiquitin carboxyl-terminal hydrolase 20	USP20	TWTYLNSLYGGGPEIA	0.9000	0.0007
Alanine--tRNA ligase, mitochondrial	AARS2	QGLWLDVHALGELQRQ	0.4016	0.0007

Myelin regulatory factor	MYRF	VPTDLHHTQQSQMLHQ	0.4324	0.0007
Mineralocorticoid receptor	NR3C2	TVAESMGLYMDSVRDA	0.8333	0.0007
Serine/threonine-protein kinase 40	STK40	CSQYEFENYMRQQLLL	0.8030	0.0007
Spermidine/spermine N(1)-acetyltransferase-like protein 1	SATL1	LSQQDLNQLVLSQPGL	1.0000	0.0007
Zinc finger protein 74	ZNF74	MLENYQNLLALGPPLH	0.2424	0.0007
CUGBP Elav-like family member 2	CELF2	VTMRNEELLLSNGTAN	0.3556	0.0007
Teneurin-4	TENM4	QENWLLNSNIPLETRN	1.0000	0.0007
Roundabout homolog 3	ROBO3	EAGISLYLAQTARGTA	0.2000	0.0006
Leucine-rich repeat transmembrane neuronal protein 4	LRRTM4	SYDQPVIGYCQAHQPL	0.7833	0.0006
Holliday junction recognition protein	HJURP	TQVDILLQGAEYFECA	0.6522	0.0006
E3 ubiquitin-protein ligase HECTD1	HECTD1	TWDDDYVLKRQFSALV	0.7869	0.0006
Histone-lysine N-methyltransferase, H3 lysine-36 specific	NSD1	KLNQLPSVTLDAVLQG	0.4634	0.0006
Unconventional myosin-XVB	MYO15B	WQDLAGCDFVLDLISQ	0.6494	0.0006
mRNA decay activator protein ZFP36	ZFP36	TAIYESLLSLSPDVPV	0.3537	0.0006
G-protein coupled receptor 3	GPR3	LSLYNALTYYSETTVT	0.9459	0.0005
A-kinase anchor protein 3	AKAP3	NKCVYQSLYMGNEPTP	0.8333	0.0005
PR domain zinc finger protein 4	PRDM4	PVGMEQLTSSSVSNAL	0.2056	0.0005
GST-ATE1_iso1_I baits from HD2 library				
Dynein heavy chain 6, axonemal	DNAAF2	YYDSFDTFIRTQFDDN	1.0000	0.1538
GRIP1-associated protein 1	GRIPAP1	ALSEEEFQRMQAQLLE	0.9987	0.0593
Emerin	EMD	YADLSDTELTTLLRRY	0.8824	0.0453
Coiled-coil domain-containing protein 117	CCDC117	ITEAELCAGPNDWILC	0.9664	0.0330
Spermatogenesis-associated protein 31A5	SPATA31A5	MENLPFPLKLLSASSL	0.9466	0.0137
Gap junction delta-3 protein	GJD3	HRAGKEAGGAEAAAQC	0.8771	0.0062
Protein BANP	BANP	SEGNLQIHHVGQDGQL	0.2301	0.0053
Kinase suppressor of Ras 1	KSR1	NAFNISAFAHAAPLPE	0.9091	0.0051
FYVE, RhoGEF and PH domain-containing protein 1	FGD1	NRSNRVCTDCYVALHG	0.9405	0.0036
RING finger protein 10	RNF10	SDGESDNSDRVPVPSF	0.8202	0.0035
Amphiphysin	AMPH	LPWDLWTTSTDLVQPA	0.2492	0.0032
Bloom syndrome protein	BLM	ILDEDLYINANDQAIA	0.9782	0.0031

Ras-related protein Rab-5A	RAB5A	DLTEPTQPTRNQCCSN	0.9344	0.0029
Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2	AGAP2	SITATPSPRRRSSAAS	0.4589	0.0028
C-type lectin domain family 1 member A	CLEC1A	RTEHRAPSSTWRPVAL	0.2364	0.0026
Ankyrin repeat domain-containing protein 11	ANKRD11	VTVTPSPEGVFSSLQA	0.4658	0.0025
Estrogen receptor beta	ESR2	RCASPVTGPGSKRDAH	0.9492	0.0025
Vacuolar protein sorting-associated protein 13B	VPS13B	HKGQELYSQQDEEQPQ	0.7241	0.0024
Kelch repeat and BTB domain-containing protein 7	KBTBD7	LDSESGSSSSFSDDEV	0.9091	0.0023
Kalirin	KALRN	TYLWEMTSGVEEIPPG	1.0000	0.0022
Prostate tumor-overexpressed gene 1 protein	PTOV1	GALGPIGPSSPGLTLG	0.7581	0.0022
R3H domain-containing protein 2	R3HDM2	PPFQQPMLVPVSQSVQ	0.3333	0.0022
Microprocessor complex subunit DGCR8	DGCR8	VMDVGSGGDGQSELPA	1.0000	0.0021
Fanconi anemia group M protein	FANCM	QLSLSEWRLWQDHPLP	0.9485	0.0019
Pre-mRNA-processing factor 40 homolog A	PRPF40A	LSEGELEKRRRTLLEQ	0.2715	0.0019
Kinase D-interacting substrate of 220 kDa	KIDINS220	SHSGKRGIPHSLSGLQ	1.0000	0.0019
Male-enhanced antigen 1	MEA1	VVADIQDRIQALGLHL	0.7222	0.0017
Chloride intracellular channel protein 6	CLIC6	ASGEAGDSVDAEGPLG	0.7931	0.0017
YY1-associated protein 1	YY1AP1	SPASSMFRKPYVRRRP	0.6337	0.0016
Putative E3 ubiquitin-protein ligase UNKL	UNKL	IPGSLPRAPSLHSPSS	0.4470	0.0016
Neuroblast differentiation-associated protein AHNAK	AHNAK	KLDADMPEVAVEGPNG	0.4095	0.0015
Neuron navigator 2	NAV2	SVVISNPHATMTQQGN	0.9231	0.0015
MARVEL domain-containing protein 2	MARVELD2	PTLHDSERAVSADPLP	1.0000	0.0014
Zinc finger C2HC domain-containing protein 1C	ZC2HC1C	VASNNKIRDPVSEPSV	0.5435	0.0014
Zinc finger protein 784	ZNF784	GGDPSRCHVCGHSCPG	1.0000	0.0014
E3 ubiquitin-protein ligase RBBP6	RBBP6	VKPKPQLSHSSRLSSD	0.8649	0.0013
Fidgetin	FIGN	YSSSTCGSHTVPSLHA	0.2684	0.0012
Bcl-2-like protein 12	BCL2L12	GLREDTLRVLAAFLRR	0.5195	0.0012

Telomere-associated protein RIF1	RIF1	SSPYGAPGTPRMNLSS	0.2683	0.0012
Myomesin-3	MYOM3	GGAGDPRPPQAMEVHR	0.5870	0.0012
IQ motif and SEC7 domain-containing protein 2	IQSEC2	LRSSASESRMSRRIIL	0.4800	0.0012
Zinc finger protein 292	ZNF292	TNLHSNVIPTCEPQSL	0.4074	0.0011
Nuclear body protein SP140	SP140	RKRRKKRGHGWSRMRM	0.3140	0.0011
Proline-rich protein 5-like	PRR5L	RHTVANAHSDIQLLAM	0.4118	0.0010
Tax1-binding protein 1	TAX1BP1	LYKVHLKNTEIENTKL	0.6630	0.0010
Serine/threonine-protein kinase WNK1	WNK1	SLTGVSSSQPIQHPQQ	0.9259	0.0009
IQ motif and SEC7 domain-containing protein 2	IQSEC2	LERGLPYGGSCGGGID	1.0000	0.0008
Cactin	CACTIN	GSSQSDSGEEQSRGQW	0.5066	0.0008
Large neutral amino acids transporter small subunit 1	SLC7A5	EKMLAAKSADGSAPAG	1.0000	0.0007
Microtubule cross-linking factor 1	MTCL1	AARIPAVTLAVTSVAG	0.4412	0.0007
Period circadian protein homolog 3	PER3	GSAASGSSDSSIYLTS	0.8889	0.0007
Vascular endothelial growth factor receptor 3	FLT4	RERLHLGRVLGYGAFG	0.2137	0.0007
DNA repair protein complementing XP-C cells	XPC	PLKSATAKGKKPSKER	0.9118	0.0007
rRNA/tRNA 2'-O-methyltransferase fibrillarin-like protein 1	FBL	GGGGWGSWGGGRGGGG	0.9500	0.0007
Shugoshin 2	SGOL2	SSPELNCNNEINGHTN	1.0000	0.0006
Mitoguardin 2	MIGA2	QSAFSQLRLTPGLRKV	0.6000	0.0006
Olfactory receptor 2G2	OR2G2	RLEPKLHMPMYFFLSH	0.5000	0.0006
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1	SMARCAD1	TSTMDGAIAAALLMFG	0.5833	0.0006
Disks large homolog 5	DLG5	IDTLSSCSQSQTSAST	0.3125	0.0006
Phosphatidate phosphatase LPIN3	LPIN3	DTVDTIALSLCGGLAD	0.2241	0.0005
pre-rRNA 2'-O-ribose RNA methyltransferase FTSJ3	FTSJ3	KGVGRKVRRPAGVRGH	0.7000	0.0005
Zinc finger protein 668	ZNF668	ERTHPVPMGTPTPLEP	0.2105	0.0005
Protein tyrosine phosphatase receptor type C-associated protein	PTPRCAP	ARDSDTEGDLVLGSPG	0.2975	0.0005

Sal-like protein 3	SALL3	PPEKPVTTWLDSKPVL	0.3571	0.0002
WD repeat-containing protein 62	WDR62	VADGLAFHAKRSYQPH	0.2619	0.0001
Proline-rich protein 5-like	PRR5L	ANAHSDIQLLAMATMM	0.7143	0.0000
GST-Gid4(99-362) baits from N-terminal library				
Ras-related protein Rab-30	RAB30	MSMEDYDFLFKI	0.9667	0.4757
Growth/differentiation factor 7	GDF7	MDLSAAAALCLW	0.9958	0.2624
Heat shock 70 kDa protein 4	HSPA4	MSVVGIDLGFQS	0.9970	0.0455
Dehydrogenase/reductase SDR family member 1	DHRS1	MAAPMNGQVCVV	0.8933	0.0056
Cell division cycle protein 23 homolog	CDC23	MAASTSMVPVAV	0.9669	0.0048
Cell division cycle protein 23 homolog	CDC23	AASTSMVPVAV	0.9854	0.0035
Leucine-rich colipase-like protein 1	LRCP1	MAGPGWTLLLLL	0.5864	0.0022
Glucagon-like peptide 1 receptor	GLP1R	MAGAPGPLRLAL	0.9100	0.0021
Plexin-A4	PLXNA4	KAMPWNWTCLL	0.9196	0.0018
Endonuclease domain-containing 1 protein	ENDOD1	MGTARWLALGSL	0.9066	0.0016
DNA polymerase eta	POLH	MATGQDRVVALV	0.4187	0.0015
Putative serpin A13	SERPINA13	MEASRWWLLVTV	0.9282	0.0012
G patch domain and ankyrin repeat-containing protein 1	GPATCH1	MSRPLLITFTPA	0.9295	0.0010
Secretin	SCT	APRPLLLLLL	0.4691	0.0009
Hemopexin	HPX	MARVLGAPVALG	0.8127	0.0008
ADAMTS-like protein 3	ADAMTSL3	MASWTSPWWVLI	0.2740	0.0008
Thrombospondin-4	THBS4	LAPRGAAVLLL	0.4109	0.0007
Humanin-like 12	MTRNR2L12	MAPRGFSCLLLS	0.4818	0.0004
Protein Wnt-2	WNT2	MNAPLGGIWLWL	0.7200	0.0004
Serine protease 41	PRSS41	MGARGALLLALL	0.6806	0.0003
Zinc finger protein 275	ZNF275	MSHPCVSLLGV	0.5979	0.0002
Serine protease inhibitor Kazal-type 14	SPINK14	MAKSFPVFSLLS	0.4774	0.0002
Zinc finger SWIM domain-containing protein 5	ZSWIM5	ADGGEREELLS	0.3000	0.0001
Serpin A11	SERPINA11	MGPAWLWLLGTG	0.5263	0.0001
Cadherin-1	CDH1	GPWSRSLSALL	0.8000	0.0001
Phenylalanine--tRNA ligase alpha subunit	FARSA	ADGQVAELLLR	1.0000	0.0001
Zinc finger SWIM domain-containing protein 5	ZSWIM5	MADGGEREELLS	1.0000	0.0001
Zinc finger protein 275	ZNF275	MMSHPCVSLLGV	0.7059	0.0001

Angiotensin-converting enzyme 2	ACE2	SSSSWLLLSLV	0.3448	0.0001
Fibroblast growth factor 4	FGF4	MSGPGTAAVALL	0.4444	0.0000
Immunoglobulin kappa variable 6D-21	IGKV6-21	SPSQLIGFLLL	1.0000	0.0000
Putative serpin A13	SERPINA13	EASRWWLLVTV	0.6154	0.0000
Pulmonary surfactant-associated protein B	SFTPB	MAESHLLQWLLL	1.0000	0.0000
Glucagon-like peptide 1 receptor	GLP1R	AGAPGPLRLAL	0.2857	0.0000
Humanin-like 8	MTRNR2L8	APRGFSCLLLS	1.0000	0.0000
Pituitary adenylate cyclase-activating polypeptide	ADCYAP1	TMCSGARLALL	1.0000	0.0000
Semaphorin-4C	SEMA4C	APHWAVWLLAA	1.0000	0.0000
Cullin-associated NEDD8-dissociated protein 1	CAND1	MASASYHISNLL	0.6667	0.0000
Serpin A11	SERPINA11	GPAWLWLLGTG	1.0000	0.0000
Alpha-2,8-sialyltransferase 8F	ST8SIA5	RPGGALLALLA	1.0000	0.0000
Cadherin-1	CDH1	MGPWSRSLSALL	1.0000	0.0000
Immunoglobulin kappa variable 6D-21	IGKV6-21	MSPSQLIGFLLL	1.0000	0.0000
Pituitary adenylate cyclase-activating polypeptide	ADCYAP1	MTMCSGARLALL	0.2632	0.0000
Alpha-2,8-sialyltransferase 8F	ST8SIA5	MRPGGALLALLA	0.4000	0.0000
Arylsulfatase A	ARSA	MGAPRSLLLALA	1.0000	0.0000
von Willebrand factor A domain-containing protein 8	VWA8	QSRLLLLGAPG	1.0000	0.0000
Testisin	PRSS21	GARGALLLALL	1.0000	0.0000
Eppin	EPPIN	GSSGLLSLLVL	1.0000	0.0000
Transmembrane 9 superfamily member 4	TM9SF4	MATAMDWLPWSL	1.0000	0.0000
GST-ZER1(519-766) baits from N-terminal library				
ADAMTS-like protein 3	ADAMTSL3	ASWTSPWWVLI	0.3138	0.0972
Fibroblast growth factor 4	FGF4	SGPGTAAVALL	0.9672	0.0318
Interferon-induced 35 kDa protein	IFI35	MSAPLDAALHAL	1.0000	0.0203
DNA polymerase eta	POLH	MATGQDRVVALV	0.4259	0.0087
V-type proton ATPase subunit C 1	ATP6V1C1	TEFWLISAPGE	0.2960	0.0080
FXYD domain-containing ion transport regulator 5	FXYD5	MSPSGRLCLLTI	0.7668	0.0079
Oxytocin-neurophysin 1	OXT	MAGPSLACCLLG	0.8701	0.0048
Matrix metalloproteinase-19	MMP19	MNCQQLWLGFLL	0.8694	0.0046

V-set and immunoglobulin domain-containing protein 10-like 2	VSIG10L2	MVGQRAQHSPVS	1.0000	0.0045
T-cell antigen CD7	CD7	MAGPPRLLLLPL	0.5146	0.0044
BPI fold-containing family B member 1	BPIFB1	MAGPWTFTLLCG	0.7685	0.0043
Putative killer cell immunoglobulin-like receptor-like protein KIR3DX1	KIR3DX1	APKLITVLCLG	0.2807	0.0030
Putative killer cell immunoglobulin-like receptor-like protein KIR3DX1	KIR3DX1	MAPKLITVLCLG	0.2791	0.0027
Histone acetyltransferase type B catalytic subunit	HAT1	MAGFGAMEKFLV	0.7197	0.0025
Guided entry of tail-anchored proteins factor 1	GET1	MSSAAADHWAWL	1.0000	0.0018
Cystatin-11	CST11	MAEPWQALQLL	0.3514	0.0018
UDP-glucuronosyltransferase 2A3	UGT2A3	MRSDKSALVFLL	0.4286	0.0017
Interferon-induced 35 kDa protein	IFI35	SAPLDAALHAL	0.7105	0.0017
Zinc finger protein 275	ZNF275	MSHPCVSLLGV	0.4021	0.0016
Dehydrogenase/reductase SDR family member 1	DHRS1	AAPMNGQVCVV	0.6034	0.0014
Hyaluronidase-3	HYAL3	MTTQLGPALVLG	0.2615	0.0013
Secretory phospholipase A2 receptor	PLA2R1	MLLSPSLLLLL	0.5652	0.0013
Putative oxidoreductase GLYR1	GLYR1	MAAVSLRLGDLV	0.2222	0.0013
Small integral membrane protein 42	SMIM42	MSSPQLPAFLWD	0.7500	0.0010
Dual specificity protein phosphatase 18	DUSP18	MTAPSCAFPVQF	0.5902	0.0009
PILR alpha-associated neural protein	PIANP	MESRMWPALLLS	0.4545	0.0009
Ankyrin repeat and SOCS box protein 8	ASB8	SSSMWYIMQSI	1.0000	0.0009
Serine protease 38	PRSS38	MAAPASVMGPLG	1.0000	0.0008
TBC1 domain family member 3G	TBC1D3G	DVVEVAGSWWA	0.2308	0.0008
S-adenosylhomocysteine hydrolase-like protein 1	AHCYL1	MSMPDAMPLPGV	0.7353	0.0008
Angiotensin-converting enzyme 2	ACE2	SSSSWLLLSLV	0.5172	0.0008

BTB/POZ domain-containing protein KCTD3	KCTD3	MAGGHCGSFPAA	1.0000	0.0007
General transcription factor 3C polypeptide 2	GTF3C2	DTCGVGYVALG	0.5000	0.0007
Ubiquitin carboxyl-terminal hydrolase 17	USP17L2	EDDSLYLGGEW	0.5714	0.0007
Distal membrane-arm assembly complex protein 2	DMAC2	MAAPWASLRLVA	0.5714	0.0007
Glycoprotein Xg	XG	MESWWGLPCLAF	0.6739	0.0007
BTB/POZ domain-containing protein KCTD3	KCTD3	AGGHCGSFPAA	1.0000	0.0006
Nuclear pore complex protein Nup88	NUP88	AAAEGPVGDGE	0.3830	0.0006
Solute carrier family 52, riboflavin transporter, member 1	SLC52A1	MAAPTLGRLVLT	0.7778	0.0005
Cytosolic iron-sulfur assembly component 3	CIAO3	MASPFSGALQLT	0.3226	0.0005
Distal membrane-arm assembly complex protein 2	DMAC2	AAPWASLRLVA	0.8696	0.0005
Otospiralin	OTOA	QACMVPGLALC	0.3077	0.0004
39S ribosomal protein L46, mitochondrial	MRPL46	AAPVRRTLLGV	1.0000	0.0004
Palmitoyl-protein thioesterase 1	PPT1	MASPGCLWLLAV	0.3750	0.0003
Angiotensin-converting enzyme 2	ACE2	MSSSSWLLLSLV	0.5185	0.0003
N-sulphoglucosamine sulphohydrolase	SGSH	SCPVPACCALL	1.0000	0.0002
Complement C1q subcomponent subunit A	C1QA	MEGPRGWLVLCV	1.0000	0.0002
Otospiralin	OTOS	MQACMVPGLALC	1.0000	0.0002
Gamma-glutamyl hydrolase	GGH	ASPGCLLCVLG	1.0000	0.0002
Ankyrin repeat and SOCS box protein 8	ASB8	MSSSMWYIMQSI	1.0000	0.0002
39S ribosomal protein L46, mitochondrial	MRPL46	MAAPVRRTLLGV	0.5000	0.0001
MIEF1 upstream open reading frame protein	MIEF1	MAPWSREAVLSL	0.2632	0.0001
Glycoprotein Xg	XG	ESWWGLPCLAF	0.6667	0.0001
Putative L-aspartate dehydrogenase	ASPDH	ADRGPWRVGVV	1.0000	0.0001
Inter-alpha-trypsin inhibitor heavy chain H6	ITIH6	MSGWRYLICVSF	0.5000	0.0001

N-sulphoglucosamine sulphohydrolase	SGSH	MSCPVPACCALL	0.6000	0.0001
Oxytocin-neurophysin 1	OXT	AGPSLACCLLG	1.0000	0.0001
Palmitoyl-protein thioesterase 1	PPT1	ASPGCLWLLAV	1.0000	0.0001
Retinal dehydrogenase 1	ALDH1A1	SSSGTPDLPVL	1.0000	0.0001
Solute carrier family 52, riboflavin transporter, member 1	SLC52A1	AAPTLGRLVLT	1.0000	0.0001
Gamma-glutamyl hydrolase	GGH	MASPGCLLCVLG	0.2857	0.0000
Proline-rich transmembrane protein 3	PRRT3	ASSPWGCVCGL	1.0000	0.0000

References

Aksnes, Henriette, Adrian Drazic, Michaël Marie, and Thomas Arnesen. 2016. "First Things First: Vital Protein Marks by N-Terminal Acetyltransferases." *Trends in Biochemical Sciences* 41(9): 746–60. http://dx.doi.org/10.1016/j.tibs.2016.07.005.

Aksnes, Henriette, Rasmus Ree, and Thomas Arnesen. 2019. "Co-Translational, Post-Translational, and Non-Catalytic Roles of N-Terminal Acetyltransferases." *Molecular Cell* 73(6): 1097–1114.

Ali, Muhammad, Leandro Simonetti, and Ylva Ivarsson. 2020. "Screening Intrinsically Disordered Regions for Short Linear Binding Motifs." *Methods in molecular biology (Clifton, N.J.)* 2141: 529–52. https://pubmed.ncbi.nlm.nih.gov/32696376/ (October 16, 2022).

Bachmair, Andreas, Daniel Finley, and Alexander Varshavsky. 1986. "In Vivo Half-Life of a Protein Is a Function of Its Amino-Terminal Residue." *Science* 234(4773): 179–86. https://www.science.org/doi/10.1126/science.3018930 (April 15, 2024).

Bachmair, Andreas, and Alexander Varshavsky. 1989. "The Degradation Signal in a Short-Lived Protein." *Cell* 56(6): 1019–32.

Bagci, Halil et al. 2023. "The HGIDGID4 E3 Ubiquitin Ligase Complex Targets ARHGAP11A to Regulate Cell Migration." *bioRxiv*: 2023.07.20.549906. http://biorxiv.org/content/early/2023/07/20/2023.07.20.549906.abstract.

Baker, Rohan T., and Alexander Varshavsky. 1995. "Yeast N-Terminal Amidase: A New Enzyme and Component of the N-End Rule Pathway." *Journal of Biological Chemistry* 270(20): 12065–74. http://dx.doi.org/10.1074/jbc.270.20.12065.

Balzi, E. et al. 1990. "Cloning and Functional Analysis of the Arginyl-TRNA-Protein Transferase Gene ATE1 of Saccharomyces Cerevisiae." *Journal of Biological Chemistry* 265(13): 7464–71.

Benz, Caroline et al. 2022. "Proteome-scale Mapping of Binding Sites in the Unstructured Regions of the Human Proteome." *Molecular Systems Biology* 18(1): 1–27.

Berndsen, Christopher E., and Cynthia Wolberger. 2014. "New Insights into Ubiquitin E3 Ligase Mechanism." *Nature Structural and Molecular Biology* 21(4): 301–7. http://dx.doi.org/10.1038/nsmb.2780.

Boldt, Karsten et al. 2016. "An Organelle-Specific Protein Landscape Identifies Novel Diseases and Molecular Mechanisms." *Nature Communications* 7(May): 1–13.

Brower, Christopher S., Konstantin I. Piatkov, and Alexander Varshavsky. 2013. "Neurodegeneration-Associated Protein Fragments as Short-Lived Substrates of the N-End Rule Pathway." *Molecular Cell* 50(2): 161–71. http://dx.doi.org/10.1016/j.molcel.2013.02.009.

Bulatov, Emil, and Alessio Ciulli. 2015. "Targeting Cullin-RING E3 Ubiquitin Ligases for Drug

Discovery: Structure, Assembly and Small-Molecule Modulation." *Biochemical Journal* 467(3): 365–86.

Chatr-aryamontri, Andrew, Almer van der Sloot, and Mike Tyers. 2018. "At Long Last, a C-Terminal Bookend for the Ubiquitin Code." *Molecular Cell* 70(4): 568–71. https://doi.org/10.1016/j.molcel.2018.05.006.

Chen, Shun Jia et al. 2017. "An N-End Rule Pathway That Recognizes Proline and Destroys Gluconeogenic Enzymes." *Science (New York, N.Y.)* 355(6323). https://pubmed.ncbi.nlm.nih.gov/28126757/ (January 20, 2022).

Chen, Shun Jia, Leehyeon Kim, Hyun Kyu Song, and Alexander Varshavsky. 2021. "Aminopeptidases Trim Xaa-Pro Proteins, Initiating Their Degradation by the Pro/N-Degron Pathway." *Proceedings of the National Academy of Sciences of the United States of America* 118(43): 1–7.

Choi, Woo Suk et al. 2010. "Structural Basis for the Recognition of N-End Rule Substrates by the UBR Box of Ubiquitin Ligases." *Nature Structural and Molecular Biology* 17(10): 1175–81. http://dx.doi.org/10.1038/nsmb.1907.

Chrustowicz, Jakub et al. 2022a. "Multifaceted N-Degron Recognition and Ubiquitylation by GID/CTLH E3 Ligases." *Journal of Molecular Biology* 434(2): 167347. https://doi.org/10.1016/j.jmb.2021.167347.

Coleman, John W. 2001. "Nitric Oxide in Immunity and Inflammation." *International Immunopharmacology* 1(8): 1397–1406.

Cox, Jürgen et al. 2011. "Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment." *Journal of Proteome Research* 10(4): 1794–1805.

Cox, Jürgen, and Matthias Mann. 2008. "MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." *Nature Biotechnology* 26(12): 1367–72.

Damgaard, Rune Busk. 2021. "The Ubiquitin System: From Cell Signalling to Disease Biology and New Therapeutic Opportunities." *Cell Death and Differentiation* 28(2): 423–26. http://dx.doi.org/10.1038/s41418-020-00703-w.

Van Damme, Petra et al. 2011. "Proteome-Derived Peptide Libraries Allow Detailed Analysis of the Substrate Specificities of Nα-Acetyltransferases and Point to HNaa10p as the Post-Translational Actin Nα-Acetyltransferase." *Molecular & Cellular Proteomics : MCP* 10(5). /pmc/articles/PMC3098586/ (April 14, 2022).

Dammer, Eric B. et al. 2011. "Polyubiquitin Linkage Profiles in Three Models of Proteolytic Stress Suggest the Etiology of Alzheimer Disease." *Journal of Biological Chemistry* 286(12): 10457–65. http://dx.doi.org/10.1074/jbc.M110.149633.

Davey, Norman E. et al. 2017a. "Discovery of Short Linear Motif-Mediated Interactions through Phage Display of Intrinsically Disordered Regions of the Human Proteome." *The FEBS Journal* 284(3): 485–98. https://onlinelibrary.wiley.com/doi/full/10.1111/febs.13995

(September 15, 2021).

Deng, Sunbin, and Ronen Marmorstein. 2021. “Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility and Regulation.” *Trends in biochemical sciences* 46(1): 15. /pmc/articles/PMC7749037/ (April 18, 2024).

Dikic, Ivan. 2017a. “Proteasomal and Autophagic Degradation Systems.” *https://doi.org/10.1146/annurev-biochem-061516-044908* 86: 193–224. https://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-061516-044908 (May 15, 2023).

Drazic, Adrian et al. 2018. “NAA80 Is Actin’s N-Terminal Acetyltransferase and Regulates Cytoskeleton Assembly and Cell Motility.” *Proceedings of the National Academy of Sciences of the United States of America* 115(17): 4399–4404. https://www.pnas.org/doi/abs/10.1073/pnas.1718336115 (July 31, 2023).

Drazic, Adrian, Line M. Myklebust, Rasmus Ree, and Thomas Arnesen. 2016. “The World of Protein Acetylation.” *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1864(10): 1372–1401.

Edwards, Richard J., Norman E. Davey, and Denis C. Shields. 2007. “SLiMFinder: A Probabilistic Method for Identifying over-Represented, Convergently Evolved, Short Linear Motifs in Proteins.” *PLoS ONE* 2(10).

Eldeeb, Mohamed A., Mohamed A. Ragheb, Marwa H. Soliman, and Richard P. Fahlman. 2022. “Regulation of Neurodegeneration-Associated Protein Fragments by the N-Degron Pathways.” *Neurotoxicity Research 2021 40:1* 40(1): 298–318. https://link.springer.com/article/10.1007/s12640-021-00396-0 (March 24, 2022).

Ella, Hadar, Yuval Reiss, and Tommer Ravid. 2019. “The Hunt for Degrons of the 26S Proteasome.” *Biomolecules* 9(6).

Farshi, Pershang et al. 2015. 25 Expert Opinion on Therapeutic Patents *Deubiquitinases (DUBs) and DUB Inhibitors: A Patent Review*.

Finley, Daniel, Helle D. Ulrich, Thomas Sommer, and Peter Kaiser. 2012. “The Ubiquitin-Proteasome System of Saccharomyces Cerevisiae.” *Genetics* 192(2): 319–60. /pmc/articles/PMC3454868/?report=abstract (September 8, 2020).

Fluge, Øystein et al. 2002. “NATH, a Novel Gene Overexpressed in Papillary Thyroid Carcinomas.” *Oncogene* 21(33): 5056–68.

Francis, Ore, Fujun Han, and Josephine C. Adams. 2013. “Molecular Phylogeny of a RING E3 Ubiquitin Ligase, Conserved in Eukaryotic Cells and Dominated by Homologous Components, the Muskelin/RanBPM/CTLH Complex.” *PLoS ONE* 8(10).

Garrido-Urbani, Sarah et al. 2016. “Proteomic Peptide Phage Display Uncovers Novel Interactions of the PDZ1-2 Supramodule of Syntenin.” *FEBS Letters* 590(1): 3–12.

Gautschi, Matthias et al. 2003. “ The Yeast N α -Acetyltransferase NatA Is Quantitatively Anchored to the Ribosome and Interacts with Nascent Polypeptides .” *Molecular and*

Cellular Biology 23(20): 7403–14.

Gietz, R. Daniel, and Robin A. Woods. 2002. “Transformation of Yeast by Lithium Acetate/Single-Stranded Carrier DNA/Polyethylene Glycol Method.” *Methods in Enzymology* 350: 87–96.

Graham, F. L., J. Smiley, W. C. Russell, and R. Nairn. 1977. “Characteristics of a Human Cell Line Transformed by DNA from Human Adenovirus Type 5.” *Journal of General Virology* 36(1): 59–72.

Hämmerle, Marcus et al. 1998. “Proteins of Newly Isolated Mutants and the Amino-Terminal Proline Are Essential for Ubiquitin-Proteasome-Catalyzed Catabolite Degradation of Fructose-1,6-Bisphosphatase of Saccharomyces Cerevisiae.” *The Journal of biological chemistry* 273(39): 25000–5. https://pubmed.ncbi.nlm.nih.gov/9737955/ (January 20, 2022).

Hassink, Gerco et al. 2005. “TEB4 Is a C4HC3 RING Finger-Containing Ubiquitin Ligase of the Endoplasmic Reticulum.” *Biochemical Journal* 388(Pt 2): 647. /pmc/articles/PMC1138973/ (April 19, 2024).

Heo, Ah Jung, Su Bin Kim, Yong Tae Kwon, and Chang Hoon Ji. 2023. “The N-Degron Pathway: From Basic Science to Therapeutic Applications.” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1866(2): 194934. https://doi.org/10.1016/j.bbagrm.2023.194934.

Hochstrasser, Mark. 1996. “UBIQUITIN-DEPENDENT PROTEIN DEGRADATION.” (93).

Hopper, A K et al. 2011. “Proteome Half-Life Dynamics.” (February): 764–68.

Hu, Rong Gui et al. 2005. “The N-End Rule Pathway as a Nitric Oxide Sensor Controlling the Levels of Multiple Regulators.” *Nature 2005 437:7061* 437(7061): 981–86. https://www.nature.com/articles/nature04027 (April 20, 2024).

Hughes, Christopher S. et al. 2019. “Single-Pot, Solid-Phase-Enhanced Sample Preparation for Proteomics Experiments.” *Nature Protocols* 14(1): 68–85.

Huttlin, Edward L et al. 2015. “The BioPlex Network of Human Protein Interactions.” *Cell* 162(2): 425–40. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4617211/pdf/nihms-705373.pdf.

Hwang, Cheol Sang, Anna Shemorry, Daniel Auerbach, and Alexander Varshavsky. 2010. “The N-End Rule Pathway Is Mediated by a Complex of the RING-Type Ubr1 and HECT-Type Ufd4 Ubiquitin Ligases.” *Nature Cell Biology 2010 12:12* 12(12): 1177–85. https://www.nature.com/articles/ncb2121 (April 20, 2024).

Hwang, Cheol Sang, Anna Shemorry, and Alexander Varshavsky. 2010a. “N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals.” *Science* 327(5968): 973–77. https://www.science.org/doi/10.1126/science.1183147 (April 19, 2024).

Ichimiya, Tadashi et al. 2020. “Autophagy and Autophagy-related Diseases: A Review.” *International Journal of Molecular Sciences* 21(23): 1–21.

Ilagan, Robielyn P. et al. 2010. “A New Bright Green-Emitting Fluorescent Protein - Engineered Monomeric and Dimeric Forms.” *FEBS Journal* 277(8): 1967–78.

Ivarsson, Ylva et al. 2014. “Large-Scale Interaction Profiling of PDZ Domains through Proteomic Peptide-Phage Display Using Human and Viral Phage Peptidomes.” *Proceedings of the National Academy of Sciences of the United States of America* 111(7): 2542–47.

Jacob, Etai, and Ron Unger. 2007. “A Tale of Two Tails: Why Are Terminal Residues of Proteins Exposed?” *Bioinformatics* 23(2): e225–30. https://dx.doi.org/10.1093/bioinformatics/btl318 (April 15, 2024).

Janke, Carsten et al. 2004. “A Versatile Toolbox for PCR-Based Tagging of Yeast Genes: New Fluorescent Proteins, More Markers and Promoter Substitution Cassettes.” *Yeast* 21(11): 947–62.

Ji, Chang Hoon, and Yong Tae Kwon. 2017. “Crosstalk and Interplay between the Ubiquitin-Proteasome System and Autophagy.” *Molecules and Cells* 40(7): 441–49. https://doi.org/10.14348/molcells.2017.0115.

Kats, Ilia et al. 2018. “Mapping Degradation Signals and Pathways in a Eukaryotic N-Terminome.” *Molecular Cell* 70(3): 488-501.e5. https://doi.org/10.1016/j.molcel.2018.03.033.

Keiler, Kenneth C., Patrick R.H. Waller, and Robert T. Sauer. 1996. “Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA.” *Science* 271(5251): 990–93.

Khmelinskii, Anton et al. 2012a. “Tandem Fluorescent Protein Timers for in Vivo Analysis of Protein Dynamics.” *Nature Biotechnology* 30(7): 708–14.

Khmelinskii, Anton et al. 2016. “Incomplete Proteasomal Degradation of Green Fluorescent Proteins in the Context of Tandem Fluorescent Protein Timers.” *Molecular Biology of the Cell* 27(2): 360–70.

Khmelinskii, Anton, and Michael Knop. “Chapter 13 Protein Timers.” 1174: 195–210.

Khoury, George A., Richard C. Baliban, and Christodoulos A. Floudas. 2011. “Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database.” *Scientific Reports* 1: 1–5.

Kim, Heon Ki et al. 2014. “The N-Terminal Methionine of Cellular Proteins as a Degradation Signal.” *Cell* 156(1–2): 158–69. http://dx.doi.org/10.1016/j.cell.2013.11.031.

Klaips, Courtney L., Gopal Gunanathan Jayaraj, and F. Ulrich Hartl. 2018. “Pathways of Cellular Proteostasis in Aging and Disease.” *Journal of Cell Biology* 217(1): 51–63. http://jcb.rupress.org/content/jcb/early/2017/11/09/jcb.201709072.full.pdf.

Kong, Ka Yiu Edwin et al. 2021. “Timer-Based Proteomic Profiling of the Ubiquitin-Proteasome System Reveals a Substrate Receptor of the GID Ubiquitin Ligase.” *Molecular Cell* 81(11): 2460-2476.e11.

Koren, Itay et al. 2018. “The Eukaryotic Proteome Is Shaped by E3 Ubiquitin Ligases Targeting C-Terminal Degrons.” *Cell* 173(7): 1622-1635.e14. http://dx.doi.org/10.1016/j.cell.2018.04.028.

Kwon, Yong Tae et al. 2002. “An Essential Role of N-Terminal Arginylation in Cardiovascular Development.” *Science* 297(5578): 96–99.

Kwon, Yong Tae, Anna S. Kashina, and Alexander Varshavsky. 1999a. “Alternative Splicing Results in Differential Expression, Activity, and Localization of the Two Forms of Arginyl-TRNA-Protein Transferase, a Component of the N-End Rule Pathway.” *Molecular and cellular biology* 19(1): 182–93. https://pubmed.ncbi.nlm.nih.gov/9858543/ (February 9, 2022).

Labbadia, Johnathan, and Richard I. Morimoto. 2015. “The Biology of Proteostasis in Aging and Disease.” *Annual review of biochemistry* 84: 435. /pmc/articles/PMC4539002/ (March 24, 2022).

Lampert, Fabienne et al. 2018. “The Multi-Subunit GID/CTLH E3 Ubiquitin Ligase Promotes Cell Proliferation and Targets the Transcription Factor Hbp1 for Degradation.” https://doi.org/10.7554/eLife.35528.001 (July 21, 2021).

Leboeuf, Dominique, Tatiana Abakumova, et al. 2020. “Downregulation of the Arg/N-Degron Pathway Sensitizes Cancer Cells to Chemotherapy In Vivo.” *Molecular Therapy* 28(4): 1092–1104. http://www.cell.com/article/S1525001620300538/fulltext (March 15, 2023).

Leboeuf, Dominique, Maxim Pyatkov, Timofei S. Zatsepin, and Konstantin Piatkov. 2020a. “The Arg/N-Degron Pathway—A Potential Running Back in Fine-Tuning the Inflammatory Response?” *Biomolecules* 10(6): 1–16.

Leznicki, Pawel, and Yogesh Kulathu. 2017. “Mechanisms of Regulation and Diversification of Deubiquitylating Enzyme Function.” *Journal of Cell Science* 130(12): 1997–2006. https://dx.doi.org/10.1242/jcs.201855 (April 14, 2024).

Li, W., and Y. Ye. 2008. “Polyubiquitin Chains: Functions, Structures, and Mechanisms.” *Cellular and Molecular Life Sciences* 65(15): 2397–2406.

Li, Wei et al. 2008. “Genome-Wide and Functional Annotation of Human E3 Ubiquitin Ligases Identifies MULAN, a Mitochondrial E3 That Regulates the Organelle’s Dynamics and Signaling.” *PLOS ONE* 3(1): e1487. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0001487 (March 27, 2022).

Lin, Hsiu Chuan et al. 2018. “C-Terminal End-Directed Protein Elimination by CRL2 Ubiquitin Ligases.” *Molecular Cell* 70(4): 602-613.e3.

Linster, Eric et al. 2022. “Cotranslational N-Degron Masking by Acetylation Promotes Proteome Stability in Plants.” *Nature Communications* 13(1): 1–12.

Lobanov, Michail Yu et al. 2010. “Library of Disordered Patterns in 3D Protein Structures.” *PLoS Computational Biology* 6(10).

Lu, Guang et al. 2021. “A Destiny for Degradation: Interplay between Cullin-RING E3 Ligases and Autophagy.” *Trends in Cell Biology* 31(6): 432–44. https://doi.org/10.1016/j.tcb.2021.01.005.

Lucas, Xavier, and Alessio Ciulli. 2017. “Recognition of Substrate Degrons by E3 Ubiquitin Ligases and Modulation by Small-Molecule Mimicry Strategies.” *Current Opinion in Structural Biology* 44: 101–10.

Lupas, Andrei N., and Kristin K. Koretke. 2003. “Bioinformatic Analysis of ClpS, a Protein Module Involved in Prokaryotic and Eukaryotic Protein Degradation.” *Journal of Structural Biology* 141(1): 77–83.

Lydeard, John R., Brenda A. Schulman, and J. Wade Harper. 2013. “Building and Remodelling Cullin-RING E3 Ubiquitin Ligases.” *EMBO Reports* 14(12): 1050–61. https://www.embopress.org/doi/10.1038/embor.2013.173 (April 16, 2024).

McLaughlin, Megan E., and Sachdev S. Sidhu. 2013. “Engineering and Analysis of Peptide-Recognition Domain Specificities by Phage Display and Deep Sequencing.” *Methods in Enzymology* 523: 327–49.

Melnykov, Artem, Shun Jia Chen, and Alexander Varshavsky. 2019. “Gid10 as an Alternative N-Recognin of the Pro/N-Degron Pathway.” *Proceedings of the National Academy of Sciences of the United States of America* 116(32): 15914–23.

Mészáros, Bálint et al. 2017. “Degrons in Cancer.” *Science Signaling* 10(470). http://stke.sciencemag.org/ (September 10, 2020).

Mueller, Franziska et al. 2021. “Overlap of NatA and IAP Substrates Implicates N-Terminal Acetylation in Protein Stabilization.” *Science Advances* 7(3).

Mullen, J. R. et al. 1989. “Identification and Characterization of Genes and Mutants for an N-Terminal Acetyltransferase from Yeast.” *EMBO Journal* 8(7): 2067–75.

Mutel, Elodie et al. 2011. “Control of Blood Glucose in the Absence of Hepatic Glucose Production during Prolonged Fasting in Mice: Induction of Renal and Intestinal Gluconeogenesis by Glucagon.” *Diabetes* 60(12): 3121–31.

Nagaoka, So I. et al. 2020. “ZGLP1 Is a Determinant for the Oogenic Fate in Mice.” *Science* 367(6482).

Naldini, Luigi et al. 1996. “Efficient Transfer , Integration , and Sustained Long-Term Expression of the Transgene in Adult Rat Brains Injected with a Lentiviral Vector.” 93(October): 11382–88.

Nijman, Sebastian M.B. et al. 2005. “A Genomic and Functional Inventory of Deubiquitinating Enzymes.” *Cell* 123(5): 773–86.

Oh, Jang Hyun, Ju Yeon Hyun, Shun Jia Chen, and Alexander Varshavsky. 2020. “Five Enzymes of the Arg/N-Degron Pathway Form a Targeting Complex: The Concept of Superchanneling.” *Proceedings of the National Academy of Sciences of the United States of America* 117(20): 10778–88. https://www.meta.org/papers/five-enzymes-of-the-argn-degron-pathway-form-a/32366662 (February 9, 2022).

Park, Joon Sung et al. 2020. “Structural Analyses on the Deamidation of N-Terminal Asn in the Human N-Degron Pathway.” *Biomolecules* 10(1). /pmc/articles/PMC7022378/ (May 25,

2023).

Park, Mi Seul et al. 2014. “Crystal Structure of Human Protein N-Terminal Glutamine Amidohydrolase, an Initial Component of the N-End Rule Pathway.” *PLOS ONE* 9(10): e111142. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0111142 (March 23, 2023).

Pédelacq, Jean Denis et al. 2005. “Engineering and Characterization of a Superfolder Green Fluorescent Protein.” *Nature Biotechnology 2005 24:1* 24(1): 79–88. https://www.nature.com/articles/nbt1172 (January 8, 2024).

Piatkov, Konstantin I., Christopher S. Brower, and Alexander Varshavsky. 2012. “The N-End Rule Pathway Counteracts Cell Death by Destroying Proapoptotic Protein Fragments.” *Proceedings of the National Academy of Sciences of the United States of America* 109(27).

Piatkov, Konstantin I., Jang Hyun Oh, Yuan Liu, and Alexander Varshavsky. 2014. “Calpain-Generated Natural Protein Fragments as Short-Lived Substrates of the N-End Rule Pathway.” *Proceedings of the National Academy of Sciences of the United States of America* 111(9).

Polevoda, Bogdan et al. 2008. “Yeast Nα-Terminal Acetyltransferases Are Associated with Ribosomes.” *Journal of Cellular Biochemistry* 103(2): 492–508.

Prakash, Sumit et al. 2004. “An Unstructured Initiation Site Is Required for Efficient Proteasome-Mediated Degradation.” *Nature Structural & Molecular Biology 2004 11:9* 11(9): 830–37. https://www.nature.com/articles/nsmb814 (January 12, 2024).

Pratt, Julie M. et al. 2002. “Dynamics of Protein Turnover, a Missing Dimension in Proteomics.” *Molecular & cellular proteomics : MCP* 1(8): 579–91. http://dx.doi.org/10.1074/mcp.M200046-MCP200.

Qiao, Shuai et al. 2020. “Interconversion between Anticipatory and Active GID E3 Ubiquitin Ligase Conformations via Metabolically Driven Substrate Receptor Assembly.” *Molecular Cell* 77(1): 150-163.e9.

Raffeiner, Margot, Shanshuo Zhu, Manuel González-Fuente, and Suayib Üstün. 2023. “Interplay between Autophagy and Proteasome during Protein Turnover.” *Trends in Plant Science* 28(6): 698–714. https://doi.org/10.1016/j.tplants.2023.01.013.

Rappsilber, Juri, Yasushi Ishihama, and Matthias Mann. 2003. “Stop And Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics.” *Analytical Chemistry* 75(3): 663–70.

Ravid, Tommer, and Mark Hochstrasser. 2008. “Diversity of Degradation Signals in the Ubiquitin-Proteasome System.” *Nature Reviews Molecular Cell Biology* 9(9): 679–89. /pmc/articles/PMC2606094/?report=abstract (September 8, 2020).

Santt, Olivier et al. 2008. “The Yeast GID Complex, a Novel Ubiquitin Ligase (E3) Involved in the Regulation of Carbohydrate Metabolism.” *Molecular biology of the cell* 19(8): 3323–33. https://pubmed.ncbi.nlm.nih.gov/18508925/ (January 20, 2022).

Scherl, Alexander et al. 2002. “Functional Proteomic Analysis of Human Nucleolus □.” *Molecular Biology of the Cell* 13(November): 4100–4109.

Shemorry, Anna, Cheol Sang Hwang, and Alexander Varshavsky. 2013. “Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway.” *Molecular Cell* 50(4): 540–51. http://dx.doi.org/10.1016/j.molcel.2013.03.018.

Soni, Shivani et al. 2006. “Absence of Erythroblast Macrophage Protein (Emp) Leads to Failure of Erythroblast Nuclear Extrusion.” *Journal of Biological Chemistry* 281(29): 20181–89. http://dx.doi.org/10.1074/jbc.M603226200.

Soucy, Teresa A. et al. 2009. “An Inhibitor of NEDD8-Activating Enzyme as a New Approach to Treat Cancer.” *Nature* 458(7239): 732–36. https://pubmed.ncbi.nlm.nih.gov/19360080/ (April 16, 2024).

Suzuki, Tetsuro, and Alexander Varshavsky. 1999. “Degradation Signals in the Lysine-Asparagine Sequence Space.” *EMBO Journal* 18(21): 6017–26.

Tasaki, Takafumi et al. 2009. “The Substrate Recognition Domains of the N-End Rule Pathway.” *Journal of Biological Chemistry* 284(3): 1884–95. http://dx.doi.org/10.1074/jbc.M803641200.

Tasaki, Takafumi, Shashikanth M. Sriram, Kyong Soo Park, and Yong Tae Kwon. 2012. “The N-End Rule Pathway.” *Annual Review of Biochemistry* 81(May 2014): 261–89.

Timms, Richard, and Itay Koren. 2020. “Tying up Loose Ends: The N-Degron and C-Degron Pathways of Protein Degradation.” *Biochemical Society Transactions* 48(4): 1557–67. https://doi.org/10.1042/BST20191094 (September 17, 2020).

Timms, Richard T. et al. 2019a. “A Glycine-Specific N-Degron Pathway Mediates the Quality Control of Protein N-Myristoylation.” *Science* 364(6448).

Varland, Sylvia et al. 2023. “N-Terminal Acetylation Shields Proteins from Degradation and Promotes Age-Dependent Motility and Longevity.” *Nature Communications* 14(1).

Varshavsky, Alexander. 2011a. “The N-End Rule Pathway and Regulation by Proteolysis.” *Protein Science* 20(8): 1298–1345.

Varshavsky, Alexander. 2019. “N-Degron and C-Degron Pathways of Protein Degradation.” *Proceedings of the National Academy of Sciences of the United States of America* 116(2): 358–66.

Walhout, Albertha J.M., and Marc Vidal. 2001. “High-Throughput Yeast Two-Hybrid Assays for Large-Scale Protein Interaction Mapping.” *Methods* 24(3): 297–306.

Wang, Haiqing, Konstantin I. Piatkov, Christopher S. Brower, and Alexander Varshavsky. 2009. “Glutamine-Specific N-Terminal Amidase, a Component of the N-End Rule Pathway.” *Molecular cell* 34(6): 686. /pmc/articles/PMC2749074/ (March 23, 2023).

Wang, Xiaoli, Roger A. Herr, and Ted H. Hansen. 2012. “Ubiquitination of Substrates by Esterification.” *Traffic* 13(1): 19–24.

Xu, Ping et al. 2009. “Quantitative Proteomics Reveals the Function of Unconventional Ubiquitin Chains in Proteasomal Degradation.” *Cell* 137(1): 133–45. http://dx.doi.org/10.1016/j.cell.2009.01.041.

Yen, Hsueh Chi Sherry et al. 2008a. “Global Protein Stability Profiling in Mammalian Cells.” *Science* 322(5903): 918–23.

Yeom, Jeonghun et al. 2017. “Comprehensive Analysis of Human Protein N-Termini Enables Assessment of Various Protein Forms.” *Scientific Reports* 7(1): 1–13. http://dx.doi.org/10.1038/s41598-017-06314-9.

Yerbury, Justin J. et al. 2016. “Walking the Tightrope: Proteostasis and Neurodegenerative Disease.” *Journal of Neurochemistry* 137(4): 489–505.

Yoo, Young Dong et al. 2018. “N-Terminal Arginylation Generates a Bimodal Degron That Modulates Autophagic Proteolysis.” *Proceedings of the National Academy of Sciences of the United States of America* 115(12): E2716–24.

Zenker, Martin et al. 2005. “Deficiency of UBR1, a Ubiquitin Ligase of the N-End Rule Pathway, Causes Pancreatic Dysfunction, Malformations and Mental Retardation (Johanson-Blizzard Syndrome).” *Nature Genetics* 37(12): 1345–50.

Zhang, Guohong, Vanessa Gurtu, and Steven R. Kain. 1996. “An Enhanced Green Fluorescent Protein Allows Sensitive Detection of Gene Transfer in Mammalian Cells.” *Biochemical and Biophysical Research Communications* 227(3): 707–11.

Zheng, Ning, and Nitzan Shabek. 2017. “Ubiquitin Ligases: Structure, Function, and Regulation.” *Annual Review of Biochemistry* 86: 129–57.

Abbreviations

AA: Amino acid

ALS: Amyotrophic lateral sclerosis

Aminopeptidase Fra1 or Icp55: Aminopeptidase Fra1 / Intermediate cleavage peptidase 55

ARHGAP11A: Rho GTPase activating protein 11A

Aro10: Aromatic amino acid aminotransferase 10

Arg-BID: Arginine-BH3 interacting-domain death agonist

Arg-BIMEL: Arginine-BCL2-like 11 (BIM extra long)

Asn-hCASP-1: Asparagine-human caspase-1

Asp-BCLXL: Aspartate-B-cell lymphoma-extra large

Asp-BRCA1: Aspartate-breast cancer type 1 susceptibility protein

Asp-EPHA4: Aspartate-EPH receptor A4

ATE1: Arginyl-tRNA--protein transferase 1

ATP: Adenosine triphosphate

Ccr4-Not complex: Carbon catabolite repression 4-Not complex

CMA: Chaperone-mediated autophagy

CRL: Cullin-RING E3 ligases

CTLH complex: C-terminal to LisH complex

Cys-RAB39a: Cysteine-Ras-related protein Rab-39A

Cys-RIPK1: Cysteine-receptor-interacting protein kinase 1

Cys-TRAF1: Cysteine-TNF receptor-associated factor 1

dUTPase: Deoxyuridine triphosphatase

DNA: Deoxyribonucleotide

Doa10: Degradation of alpha2 10

DUBs: Deubiquitinases

DMEM: Dulbecco's Modified Eagle's Medium

dsDNA: Double-stranded DNA

E1: Ub-activating enzyme

E2: Ub-conjugating enzyme

E3: Ub-protein ligase

ELISA: Enzyme-linked immunosorbent assay

ER: Endoplasmic reticulum

FACS: Fluorescence-activated cell sorting

Fbp1: Fructose-1,6-bisphosphatase

FBS: Fetal Bovine Serum

FTLD or FTLD-TDP: Frontotemporal lobar degeneration

gDNA: Genomic DNA

GID complex: Glucose-induced degradation complex

Gln-hCASP-4: Glutamine-human caspase-4

Gln-hCASP-5: Glutamine-human caspase-5

GPS: Global Protein Stability

HECT: Homologous to E6-associated protein C Terminus

HEK-293FT: Human embryonic kidney 293FT cells

HEK-293T: Human embryonic kidney 293T cells

HBP1: HMG-box transcription factor 1

HTRA2: High temperature requirement A serine peptidase 2

IAP: Inhibitor of apoptosis protein

Icl1: Isocitrate lyase 1

Ile-GRZA: Isoleucine-granzyme A

iMet: Initiator Methionine

IRES: Internal ribosome entry site

IPTG: Isopropyl-β-D-thiogalactopyranoside

KCNT2: Potassium sodium-activated channel subfamily T member 2

KCMF-1: Potassium channel modulatory factor 1

Leu-LIMK1: Leucine-LIM domain kinase 1

lp-In T-REx 293: Inducible expression system in T-REx 293 cells

Mdh2: Malate dehydrogenase 2

MetAPs: Methionine aminopeptidases

MOI: Multiplicity of infection

NatA acetyltransferases: N-alpha-acetyltransferase A

NatB acetyltransferases: N-alpha-acetyltransferase B

NatD acetyltransferases: N-alpha-acetyltransferase D

NatF acetyltransferases: N-alpha-acetyltransferase F

NatG acetyltransferases: N-alpha-acetyltransferase G

NatH acetyltransferases: N-alpha-acetyltransferase H

NAA15: N-alpha-acetyltransferase 15

NAA25: N-alpha-acetyltransferase 25

NAA35/NAA38: N-alpha-acetyltransferase 35 / N-alpha-acetyltransferase 38

NGS: Next-generation sequencing

Not4: Negative on TATA-less 4

NTAQ1: N-terminal glutamine amidohydrolase 1

NTAN1: N-terminal asparagine amidohydrolase 1

Nta1: Protein N-terminal amidase

PCR: Polymerase chain reaction

Pck1: Phosphoenolpyruvate carboxykinase 1

PN: Proteostasis network

ProP-PD: Proteomic peptide-phage display

PSI: Protein Stability Index

PTM: Post-translational modifications

Rad6: Radiation sensitive 6

RBR: RING-between-RING

RFP: Red fluorescent protein

RING: Really Interesting New Gene

sfGFP: Superfolder green fluorescent protein

SLiMs: Short Linear Motifs

ssDNA: Single-stranded DNA

SUMOylation: Small Ubiquitin-like Modifier

TEB4 or MARCHF6: Membrane-associated RING-CH-type finger 6

TPM: Transcripts per million

Tyr-MET: Tyrosine-MET proto-oncogene, receptor tyrosine kinase

Tyr-NEDD9: Tyrosine-neural precursor cell expressed, developmentally down-regulated 9

UBR family E3 ubiquitin ligases: Ubiquitin protein ligase E3 component n-recognin family

UBR1: Ubiquitin protein ligase E3 component n-recognin 1

UBR2: Ubiquitin protein ligase E3 component n-recognin 2

UBR4: Ubiquitin protein ligase E3 component n-recognin 4

UBR5: Ubiquitin protein ligase E3 component n-recognin 5

Ub: Ubiquitin

Ubc4/Ubc5: Ubiquitin-conjugating enzyme 4 / Ubiquitin-conjugating enzyme 5

Ufd4: Ubiquitin fusion degradation protein 4

UPS: Ubiquitin Proteasome System

ZER1: Zyg-11 related cell cycle regulator

ZMYND19: Zinc finger MYND domain-containing protein 19

ZYG11B: Zyg-11 family member B

Amino Acids

A: Alanine

C: Cysteine

D: Aspartic acid

E: Glutamic acid

F: Phenylalanine

G: Glycine

H: Histidine

I: Isoleucine

K: Lysine

L: Leucine

M: Methionine

N: Asparagine

P: Proline

Q: Glutamine

R: Arginine

S: Serine

T: Threonine

V: Valine

W: Tryptophan

Y: Tyrosine