

DISSECTING PROTEIN QUALITY CONTROL MECHANISMS OF MISLOCALIZED PROTEINS

Dissertation

zur Erlangung des Grades

Doktor der Naturwissenschaften

am Fachbereich Biologie

Der Johannes Gutenberg-Universität Mainz

Jia Jun Fung

geboren am 19.02.1993

in Klang, Malaysia

Mainz, January 2023

Dekan: Prof. Dr. Eckhard Thines

1. Gutachter: [REDACTED]

2. Gutachter: [REDACTED]

Tag der mündlichen Prüfung: 10.01.2023

PREFACE

The research presented in this dissertation was carried out in the research group of [REDACTED] at the [REDACTED]. The contents in this thesis were written solely by myself and all resources and works of others are clearly cited and referenced.

Through the duration of the PhD study, I was supported and guided by my supervisor [REDACTED] in terms of discussion and experimental designs. All the experiments and analyses were performed by me with the exception of the following:

- Part of the initial extraction of the MoBY1 library plasmids in a high-throughput format using a 96-well direct miniprep kit was assisted by [REDACTED].
- Flow cytometry measurement setup using BD LSRFortessa SORP flow cytometer was done with the help of the IMB Flow Cytometry core facility (main contact: [REDACTED]).
- Strain identity verifications of some tFT strains by colony PCR were done by [REDACTED]. Additionally, [REDACTED] also helped with the mini-prepping of one of the MoBY libraries for Next Generation Sequencing of the libraries.
- Library prep for Next Generation Sequencing was carried out by the IMB Genomics core facility and the raw sequencing data was processed by the IMB Bioinformatics core facility (main contact: [REDACTED]).

TABLE OF CONTENTS

TABLE OF CONTENTS	I
ABSTRACT	III
ZUSAMMENFASUNG	V
LIST OF TABLES	VII
LIST OF FIGURES	VIII
LIST OF ABBREVIATIONS	X
INTRODUCTION	1
1.1 <i>Protein homeostasis</i>	1
1.1.1: Protein QC by the Ubiquitin Proteasome System	3
1.1.2: Protein sequestration as a QC mechanism	4
1.2 <i>Mislocalized proteins</i>	6
1.2.1: Mistargeting or loss of targeting signal as a source of MLPs	8
1.2.2: Compromised protein import leads to protein mislocalization	9
1.2.3: Excess proteins/orphaned proteins as source of MLPs	10
1.2.4: Methods of recognition of MLPs for removal	11
1.3 <i>Project aims and goals</i>	13
METHODOLOGY	14
1.1 <i>High-throughput studies using Synthetic Genetic Array (SGA)</i>	14
1.2 <i>Following protein fate using a tandem fluorescent protein timer (tFT)</i>	15
1.3 <i>Mislocalizing protein through overexpression</i>	16
1.4 <i>Expectations on attenuation behaviour upon overexpression</i>	18
RESULTS	20
CHAPTER 1: SETUP AND PREPARATION FOR STUDYING PROTEIN MISLOCALIZATION	20
1.1 <i>Constructing and characterizing the tFT library</i>	20
1.2 <i>Preparing the overexpression libraries</i>	22
1.3 <i>Assembly of libraries into a screening layout</i>	23
CHAPTER 2: MISLOCALIZING PROTEINS THROUGH OVEREXPRESSION	25
2.1 <i>MLPs that are attenuated upon low level overexpression (tFT-MoBY1)</i>	25
2.2 <i>MLPs that are attenuated upon stronger overexpression (tFT-MoBY2)</i>	28
2.3 <i>Very strong overexpression leads to growth defects and attenuated MLPs (tFT-mORF)</i>	32
CHAPTER 3: SEQUENCING OF MOBY PLASMID LIBRARIES	35
3.1 <i>Data processing and clean up</i>	36
3.2 <i>Mutations in MoBY2 plasmids unlikely to have impacted results of overexpression screen</i>	39
CHAPTER 4: PROPERTIES OF ATTENUATED PROTEINS	40
4.1 <i>Properties of all proteins with attenuated abundance</i>	40
4.2 <i>Breakdown of tFT-MoBY2 screen results into different attenuation categories</i>	46
4.2.1: Proteins that are destabilized but not attenuated	47
4.2.2: Proteins that are attenuated but not destabilized	49

4.2.3: Proteins that are attenuated and destabilized.....	53
CHAPTER 5: UNDERSTANDING THE DIFFERENCE IN ATTENUATION BEHAVIOUR OF SUBUNITS OF COMPLEXES	55
5.1 <i>Properties of attenuated subunits</i>	55
5.2 <i>Overexpression and knockout screen across subunits within complexes</i>	58
5.2.1: Endoplasmic Reticulum Membrane Complex (EMC)	58
5.2.2: Vacuolar Adenosine-Tri-Phosphatase (V-ATPase) Complex	62
5.2.3: CCR4-NOT Complex	66
5.2.4: Glucose Induced Degradation (GID) Complex	69
5.2.5: Summary of the overexpression and knockout screen across subunits within complexes	70
CHAPTER 6: UNDERSTANDING THE DIFFERENCE IN ATTENUATION BEHAVIOUR OF MITOCHONDRIAL PROTEINS.....	71
6.1 <i>Overexpression levels of cytosolic and mitochondrial ribosomes are similar</i>	73
6.2 <i>tFT-tagged mitochondrial proteins localize correctly to the mitochondria with exceptions</i>	74
6.3 <i>Growth on alternate carbon sources and mild heat stress does not lead to increased attenuation of mitochondrial proteins</i>	79
CHAPTER 7: SEARCHING THE PROTEOME FOR DOSAGE COMPENSATION GENES	81
DISCUSSION.....	83
1.1 <i>Studying protein mislocalization through protein overexpression</i>	83
1.2 <i>Properties of MLPs in determining attenuation behaviour</i>	84
1.3 <i>Multiple modes of regulation and attenuation of MLPs upon overexpression</i>	86
1.4 <i>Mitochondrial proteins are resistant to mislocalization by protein overexpression</i>	87
1.5 <i>Concluding remarks and outlooks</i>	88
METHODS & MATERIALS	89
YEAST METHODS AND PLASMIDS.....	89
COMPARISON AND CHARACTERIZATION OF THE MCH-MNG TFT.....	89
PREPARATION OF OVEREXPRESSION LIBRARIES.....	91
TFT-OVEREXPRESSION SCREENS	93
TFT-KNOCKOUT SCREEN	95
SUBUNITS TFT-OVEREXPRESSION AND TFT-KNOCKOUT SCREENS	95
PLASMID COPY NUMBER AND GENE EXPRESSION LEVEL QUANTIFICATION	96
NEXT GENERATION SEQUENCING OF MOBY PLASMID LIBRARIES	97
CONFIRMATION OF PROTEIN LOCALIZATION VIA MICROSCOPY.....	99
APPENDIX.....	101
SUPPLEMENTARY TABLES	101
INSTRUMENTS AND SOFTWARE.....	112
REFERENCE LIST	113
ACKNOWLEDGEMENTS	127
CURRICULUM VITAE	128

ABSTRACT

Protein homeostasis plays an important role in maintaining a healthy proteome within the cell with its quality control systems. The breakdown of protein homeostasis leads to the accumulation of mutated, misfolded, damaged, and mislocalized proteins, many of which are associated with various diseases. Mislocalized proteins are proteins that fail to reach their native compartment or assemble into their native complexes. Mislocalization can occur due to intrinsic inefficiencies of targeting, folding, trafficking, or complex assembly. Various diseases have been associated with or arise due to mislocalization of proteins, and furthermore, the extent of mislocalization increases with aging and in aneuploidy. It is therefore important to better understand the cellular mechanisms that handle mislocalized proteins.

To this end, I sought to systematically identify proteins that are degraded upon mislocalization and then dissect and characterize the properties of these proteins. In order to accomplish this, I established a proteome-wide platform to study the fate of mislocalized proteins upon overexpression of individual proteins using a tandem fluorescent protein timer tag. The tag provides information on protein abundance and turnover. Using this approach, I screened 4211 proteins for their behaviour upon overexpression at different levels using low and high copy plasmids with endogenous and GAL1 promoters. My results show that ~15% of proteins are attenuated upon overexpression with a 2 μ plasmid with endogenous promoters. These proteins are distinct from proteins that were attenuated in aneuploidy conditions. In addition, the main characteristics of the attenuated proteins are that they are enriched for subunits of complexes and essential genes, are highly synthesized, and have long half-lives. Aside from the 15% of attenuated proteins, I also found that another ~8% of proteins were destabilized but not attenuated upon overexpression. I showed evidence from the data and the protein properties that support my hypothesis for various methods that the cell may be using to manage and deal with protein mislocalization upon overexpression, depending on the protein's change in protein levels and stability.

I am able to show that the data also recapitulates the known degradation of cytosolic ribosomal proteins upon overexpression, however, the mitochondrial ribosomal proteins lack any attenuation. In fact, there was a depletion of mitochondrial proteins from the attenuated proteins. Only with an even stronger overexpression, I see attenuation of more mitochondrial proteins but still very minimal for the mitochondrial ribosomal proteins. I also showed that in general, the tagged mitochondrial proteins still localize correctly upon overexpression of an untagged version of the protein. The lack of attenuation of mitochondrial proteins did not change with increased growth temperature or on non-fermentable media.

Focusing on the subunits of complexes, I found that not all subunits within a complex were similarly attenuated. This difference in attenuation behaviour is likely not due to synthesis rates or half-lives, but rather, presence of chaperones or direct interaction partners, assembly factors, and assembly sequence of a complex. Through overexpressing and knocking out every partner subunit within the ER membrane complex, I was able to dissect and propose an assembly model for this complex with the information on its structure. I tested several other complexes but not all were clear and easily interpretable.

Based on my research, I present a large dataset of proteins that are attenuated upon overexpression which can be used to further dissect the protein quality control mechanisms that deal with mislocalized proteins. My research paves the way for further research into the proteins that are attenuated and/or destabilized upon overexpression and the likely mechanisms that are in play for each. I also present a method to dissect the assembly sequence of a complex and subunit pairs that are interdependent for stability.

ZUSAMMENFASSUNG

Die Proteinhomöostase spielt eine wichtige Rolle bei der Aufrechterhaltung eines gesunden Proteoms innerhalb der Zelle mit ihren Qualitätskontrollsystemen. Eine Störung der Proteinhomöostase führt zur Anhäufung mutierter, fehlgefalteter, beschädigter und fehllokalisierter Proteine, von denen viele mit verschiedenen Krankheiten in Verbindung stehen. Fehllokalisierte Proteine sind Proteine, die ihr natives Kompartiment nicht erreichen oder sich nicht zu ihren nativen Komplexen zusammenfügen. Die Fehllokalisierung kann auf intrinsische Ineffizienzen bei der Zielbestimmung, der Faltung, dem Transport oder der Komplexbildung zurückzuführen sein. Verschiedene Krankheiten sowie deren Entstehung werden mit der Fehllokalisierung von Proteinen in Verbindung gebracht. Des Weiteren nimmt das Ausmaß der Fehllokalisierung mit dem Alter sowie bei Aneuploidie zu. Daher ist es wichtig, die zellulären Mechanismen, die mit fehllokalisierten Proteinen umgehen, besser zu verstehen.

Zu diesem Zweck habe ich versucht, systematisch Proteine zu identifizieren, die bei Fehllokalisierung abgebaut werden, und dann die Eigenschaften dieser Proteine zu analysieren und zu charakterisieren. Dazu habe ich eine proteomweite Plattform etabliert, um das Schicksal fehllokalisierter Proteine während der Überexpression einzelner Proteine mit Hilfe eines Tandem-Fluoreszenzprotein-Timer-Tags zu untersuchen. Der Tag liefert Informationen über die Abundanz und den Umsatz von Proteinen. Mit diesem Ansatz untersuchte ich 4211 Proteine auf ihr Verhalten bei verschiedenen Graden der Überexpression unter Verwendung von Plasmiden mit niedriger und hoher Kopienzahl und endogenen und GAL1-Promotoren. Meine Ergebnisse zeigen, dass bei der Überexpression mit einem 2 μ -Plasmid und Verwendung von endogenen Promotoren ungefähr 15 % der Proteine vermindert nachweisbar sind. Diese Proteine unterscheiden sich von den Proteinen, die unter Aneuploidiebedingungen vermindert nachweisbar waren. Darüber hinaus sind die Hauptmerkmale dieser Proteine eine Anreicherung von Untereinheiten von Komplexen und essentiellen Genen, dass sie in hohem Maße synthetisiert werden und eine lange Halbwertszeit haben. Neben den 15 % der vermindert nachweisbaren Proteine habe ich auch festgestellt, dass weitere 8 % der Proteine durch Überexpression zwar destabilisiert, aber nicht vermindert nachweisbar wurden. Diese Daten, sowie die Eigenschaften der Proteine, unterstützen meine Hypothese über verschiedene mögliche Methoden, mit denen die Zelle die Fehllokalisierung von Proteinen abhängig von deren Menge und Stabilität bewältigen kann.

Ich kann zeigen, dass die Daten auch den schon bekannten Abbau der zytosolischen, ribosomalen Proteine bei Überexpression aufzeigen, während bei den mitochondrialen, ribosomalen Proteinen keine Verminderung zu verzeichnen ist. Vielmehr waren unter den vermindert

nachweisbaren Proteinen nur wenige mitochondriale Proteine. Erst bei einer noch stärkeren Überexpression sehe ich eine Verminderung von mehr mitochondrialen Proteinen, aber immer noch eine sehr minimale für die mitochondrialen, ribosomalen Proteine. Des Weiteren habe ich gezeigt, dass die getaggten mitochondrialen Proteine im Allgemeinen auch bei Überexpression einer nicht markierten Version des Proteins korrekt lokalisiert werden. Die fehlende Reduktion der mitochondrialen Proteine änderte sich nicht bei erhöhter Wachstumstemperatur oder auf nicht fermentierbaren Medien.

Bei der Betrachtung der Untereinheiten von Komplexen stellte ich fest, dass nicht alle Untereinheiten innerhalb eines Komplexes in gleicher Weise reduziert waren. Dieser Unterschied im Verhalten ist wahrscheinlich nicht auf die Syntheseraten oder Halbwertszeiten zurückzuführen, sondern vielmehr auf das Vorhandensein von Chaperonen oder direkten Interaktionspartnern, Assembly-Faktoren und der Assembly-Sequenz eines Komplexes. Durch Überexpression und Knock-Out aller Partner-Untereinheiten des ER-Membrankomplexes konnte ich ein Modell für den Zusammenbau des Komplexes mit den Informationen über seine Struktur erstellen. Ich habe mehrere andere Komplexe getestet, aber nicht alle waren klar und leicht zu interpretieren.

Auf der Grundlage meiner Forschung präsentiere ich einen großen Datensatz von Proteinen, die bei Überexpression vermindert nachweisbar werden. Dies kann zur weiteren Analyse der Qualitätskontrollmechanismen, die mit fehllokalisierten Proteinen zu tun haben, genutzt werden. Meine Forschung ebnet den Weg für die weitere Erforschung der Proteine, die bei Überexpression reduziert und/oder destabilisiert werden, sowie der wahrscheinlichen Mechanismen, die dabei im Spiel sind. Außerdem stelle ich eine Methode vor, mit der sich die Aufbaureihenfolge eines Komplexes und der Untereinheitenpaare, die für ihre Stabilität voneinander abhängig sind, aufschlüsseln lassen.

LIST OF TABLES

Table 1: Disease-associated mutations of proteostasis network components.....	2
Table 2: Diseases associated with mislocalized proteins.....	7
Supplementary Table 1: Yeast strains used in this study.....	101
Supplementary Table 2: Plasmids used in this study.....	104
Supplementary Table 3: Yeast and <i>E. coli</i> libraries used in this study.....	105
Supplementary Table 4: Primers for qPCR and RT-qPCR.....	106
Supplementary Table 5: General primers	107
Supplementary Table 6: Sources for datasets used in data analysis	109
Supplementary Table 7: Curated list of ORFs with known dosage compensation behaviour upon overexpression	110
Supplementary Table 8: Instruments used in this study	112
Supplementary Table 9: Softwares used in this study	112
Supplementary Table 10: R packages used for data analysis	112

LIST OF FIGURES

Figure 1: Components of the UPS and the various roles of ubiquitin in cellular functions.	4
Figure 2: Location of various compartments that abnormal proteins can be sequestered into.	5
Figure 3: Mislocalized proteins and their possible fates.	6
Figure 4: The steps taken to efficiently construct a haploid carrying modifications of interest from two different haploids in an array format using the SGA method.	15
Figure 5: Construction of the tFT strain from a C-SWAT starting strain and donor plasmid.	16
Figure 6: Schematic of the overexpression plasmids used in this study.	17
Figure 7: Expected behaviour of proteins mislocalized through overexpression.	19
Figure 8: Characterizing the mCherry-mNeonGreen tFT and comparing it to mCherry-sfGFP.	21
Figure 9: tFT-overexpression screen layout and process.	24
Figure 10: Low level overexpression with CEN plasmids lead to attenuation of a small proportion of proteins many of which are subunits of complexes.	27
Figure 11: Stronger overexpression with 2 μ plasmids lead to attenuation of a larger proportion of proteins many of which are subunits of complexes and can recapitulate literature findings.	29
Figure 12: Very strong overexpression with 2 μ plasmids and GAL1 promoter lead to widespread growth defects but filtered data is still usable.	33
Figure 13: NGS sequencing outcomes of MoBY1 and MoBY2 plasmid libraries.	38
Figure 14: Variant annotations of SNPs in MoBY2 plasmids used in tFT-MoBY2 screen.	39
Figure 15: Attenuated subunits from the tFT-MoBY2 overexpression screen are from various complexes of different sizes but not all subunits within a complex are attenuated.	41
Figure 16: Properties of attenuated hits from the tFT-MoBY2 overexpression screen.	42
Figure 17: Properties of ORFs that are destabilized but not attenuated.	48
Figure 18: Properties of ORFs that are attenuated but not destabilized.	51
Figure 19: Properties of ORFs that are have attenuated mNG levels and ratio.	54
Figure 20: Properties of attenuated subunits of complexes within tFT-MoBY2 screen.	56
Figure 21: OST complex assembly and synthesis rate correlation.	57
Figure 22: Results of overexpression and knockout crosses for ER membrane complex (EMC).	59
Figure 23: Proposed model assembly pathway for EMC.	61
Figure 24: Results of overexpression and knockout crosses for V-ATPase.	65
Figure 25: Results of overexpression and knockout crosses for CCR4-NOT complex.	67
Figure 26: Results of overexpression and knockout crosses for GID complex.	69

Figure 27: Mitochondrial proteins are resistant to attenuation upon overexpression and mitochondrial ribosomal subunits are not as readily attenuated as their cytosolic counterparts.....	72
Figure 28: Plasmid copy number comparison between cytosolic and mitochondrial ribosomal subunits.	73
Figure 29: Most tFT-tagged mitochondrial ribosomal proteins localize correctly to the mitochondria and some mislocalized upon overexpression.	74
Figure 30: Mitochondrial ribosomal protein localization with mNG-tagged strains.....	75
Figure 31: <i>Img2</i> and <i>Mrp122</i> show difference in import behaviour when untagged or smaller tagged options exist but overall there is no preferential attenuation of the exogenous protein.....	78
Figure 32: Attenuation behaviour of mitochondrial proteins remain unchanged when grown on alternate carbon sources and at higher temperatures.	80
Figure 33: tFT library strains may still harbour the initial donor plasmid and so interferes with SGA selection when crossed with the YKO library.	82

LIST OF ABBREVIATIONS

AF	allele frequency
AS	anchor sequence
BAM	binary alignment map
DUBs	deubiquitylating enzymes
EMC	ER membrane complex
ER	endoplasmic reticulum
ERAD	endoplasmic reticulum associated degradation
ESR	environmental stress response
GET	Guided Entry of Tail-Anchored
GID	Glucose Induced Degradation
GPI	glycosylphosphatidylinositol
GQC	Golgi quality control
GRAVY	Grand Average of Hydrophathy
HECT	homologous to E6-AP c-terminus
HSPs	heat shock proteins
IGV	Integrative Genomics Viewer
IMB	Institute of Molecular Biology
INQ	intranuclear quality control site
IPOD	insoluble protein deposit
JUNQ	juxtannuclear quality control site
MAD	mitochondria associated degradation
mCH	mCherry
MLPs	mislocalized proteins
mNG	mNeonGreen
MTS	mitochondrial targeting signal
NLS	nuclear localization signal
OE	overexpression
ORF	open reading frame

PCN	plasmid copy number
PTM	post-translational modification
QC	quality control
qPCR	quantitative PCR
RBR	RING-between-RiNG
RING	Really Interesting New Gene
RQC	ribosome-associated quality control
RT-qPCR	Reverse transcription qPCR
sfGFP	superfolderGFP
SGA	synthetic genetic array\
SNP	single nucleotide polymorphism
SWAT	SWAp-Tag
TA	tail-anchored
tFT	tandem fluorescent protein timer
TMT	tandem mass tag
TRC	transmembrane recognition complex
UPR	unfolded protein response
UPS	Ubiquitin Proteasome System

INTRODUCTION

“A cell is like a factory that produces many different products (proteins) that need to be folded, assembled into larger products, and be sent to the correct location within the factory for it to be used. Along the assembly line, the quality control mechanisms check that these products are correct before allowing further processing. However, if the product is faulty or at the wrong location, it needs to be removed for destruction or if still usable, remodelled and relocated. If the quality control mechanisms breakdown, then there will be chaos with wrong and faulty products being produced and sent to the wrong places.”

While an oversimplification, the analogy above contains the key points that describe protein homeostasis and the topic that this PhD thesis will address and explore.

1.1 Protein homeostasis

Maintaining a balanced and functional proteome is key to the survival of the cell. Protein homeostasis is the process of maintaining a balanced and functional proteome through protein synthesis, folding, assembly, localization, export, and removal (Hartl, Bracher and Hayer-Hartl, 2011; Oromendia, Dodgson and Amon, 2012; Sala, Bott and Morimoto, 2017). Beginning from translation till forming a fully functional protein, a nascent protein goes through various steps of folding, assembling, and targeting. If at any point along the process problems occur and a protein ends up being abnormal, the cell's protein quality control (QC) systems need to identify the abnormal protein, try and correct it if possible, or else remove it (Chen *et al.*, 2011; Kong *et al.*, 2021). These protein QC systems are highly important to maintain protein homeostasis and so its failure is often associated with aging, neurodegenerative diseases, and cancer (Labbadia and Morimoto, 2015; Schneider and Bertolotti, 2015; Klaipts, Jayaraj and Hartl, 2018). Some of the known QC mechanisms include, but are not limited to, endoplasmic reticulum associated degradation (ERAD), Golgi quality control (GQC), endo-lysosome quality control, ribosome-associated quality control (RQC), mitophagy, and mitochondria-associated degradation (MAD) (Youle and Narendra, 2010; Amm, Sommer and Wolf, 2014; Joazeiro, 2019; Sun and Brodsky, 2019; Song, Herrmann and Becker, 2020; Sardana and Emr, 2021). Failures in these protein QC systems can lead to various diseases such as amyotrophic lateral sclerosis, Paget's disease, Parkinson's disease, and many more (Labbadia and Morimoto, 2015; Rinaldi, Mäger and Wood, 2016; George *et al.*, 2018) (Table 1). Therefore, a clear understanding of

protein QC systems is not only important to improve our basic knowledge on their functions but it could lead to identifying causes of diseases at a molecular level that can then be targeted for a treatment.

Table 1: Disease-associated mutations of proteostasis network components

Disease	Gene	Protein	Function
Amyotrophic Lateral Sclerosis	UBQLN2	Ubiquilin-2	Ubiquitin-like protein
	OPTN	Optineurin	Autophagy
	VAPB	VAPB	Vesicle trafficking
Parkinson's Disease	PARK1	α -synuclein	E3 ubiquitin ligase
	HSP9A	Mortalin	Mitochondrial HSP70
	DNAJC13	RME8	Receptor-mediated endocytosis
Charcot–Marie–Tooth disease	HSPB1	HSP27	Small heat shock protein
Charcot–Marie–Tooth disease type 2L	HSPB8	HSP22	Small heat shock protein
Angelman syndrome	UBE3A	UBE3A	E3 ubiquitin ligase
Dilated Cardiomyopathy with Ataxia syndrome	DNAJC19	TIMM14	Mitochondrial import
Cardiovascular disease	HSPA8	HSC70	Protein folding
Dominantly inherited myopathy	DNAJB6	DNAJB6	Protein folding
Paget's disease, Amyotrophic Lateral Sclerosis	SQSTM1	p62	Autophagy
Spinocerebellar Ataxia (SCA3) / Machado–Joseph disease	ATXN3	Ataxin-3	Deubiquitylase
Spastic paraplegia (SPG15)	ZFYVE26	Spastizin	Autophagosome formation

Source: Labbadia and Morimoto, 2015; Rinaldi, Mäger and Wood, 2016

1.1.1: Protein QC by the Ubiquitin Proteasome System

In order to carry out the QC of abnormal proteins, the cell could degrade these abnormal proteins and recycle the amino acids to be re-used (Chen *et al.*, 2011). The cell utilizes a host of tools which includes the Ubiquitin Proteasome System (UPS). The UPS consists of several components such as ubiquitin, E1-ubiquitin activating enzymes, E2-ubiquitin conjugating enzymes, E3-ubiquitin ligases, deubiquitylating enzymes (DUBs), and the proteasome (Finley *et al.*, 2012). The E1, E2, and E3 enzymes work together to ubiquitinate proteins to target them for degradation by the proteasome, while the DUBs work to remove ubiquitin from proteins (Kulathu and Komander, 2012; Suresh, Pascoe and Andrews, 2020; Dikic and Schulman, 2022). The E1, E2, and E3 enzymes work in a pyramid-like structure, generally, with only one (yeast) or two (human) E1 enzymes, tens of E2 enzymes and hundreds of different E3 enzymes present in the cell (Figure 1A) (Weissman, 2001; Groettrup *et al.*, 2008). The E3 ligases provide the substrate specificity in the process of ubiquitination and while many E3 ligases are known, the exact targets for every E3 are not well known (Iconomou and Saunders, 2016). The UPS also has a significant level of redundancy and multiplicity as a single E3 ligase may target multiple proteins or a single protein may be targeted by multiple E3 enzymes. The E3 ligases can be differentiated into several families depending on the domains present which are the RING (Really Interesting New Gene) domain, HECT (homologous to E6-AP c-terminus) domain, U-Box domain, and RBR (RING-between-RING) domain (Hatakeyama and Nakayama, 2003; Deshaies and Joazeiro, 2009; Wang *et al.*, 2020; Yang *et al.*, 2021). Additionally, the RING E3 ligase family utilizes a set of different substrate receptors for substrate recognition such as the F-box, BC-box, BTB, and DCAF (Harper and Schulman, 2021).

Besides the enzymes, ubiquitin is also an important part of the UPS as it is the molecule that is added onto proteins as a post-translational modification (PTM) to target them to the proteasome for degradation. For a ubiquitin molecule to be added onto a target protein, it is first activated by an E1 enzyme in an ATP-dependent manner. Then, the ubiquitin is transferred onto the active cysteine site of an E2 enzyme via a thioester bond, and lastly, the E3 ligase interacts with the target protein and the ubiquitin carrying E2 and finishes the process by catalysing the transfer of ubiquitin from the E2 to the target protein (Finley *et al.*, 2012). While targeting for degradation is one of its main roles, ubiquitin has been shown to have a role in various other cellular processes such as autophagy, cell signalling, protein trafficking, kinase modifications, and cell cycle regulation (Figure 1B) (Song and Luo, 2019; Dang, Nie and Wei, 2020; Yang *et al.*, 2021). This myriad of functions are attributed to ubiquitin's versatility in being added as single molecules onto proteins (mono-ubiquitination or multi-mono-ubiquitination) or formation of a chain of ubiquitin linked by its lysine residues (poly-

ubiquitination). Ubiquitin is generally conjugated onto a protein via lysine residues and itself has 7 different lysine residues that can also be ubiquitinated to build various poly-ubiquitin chains. The type of function a ubiquitin molecule is associated with depends on the type of linkage within the poly-ubiquitin chains (Figure 1B) (Finley *et al.*, 2012).

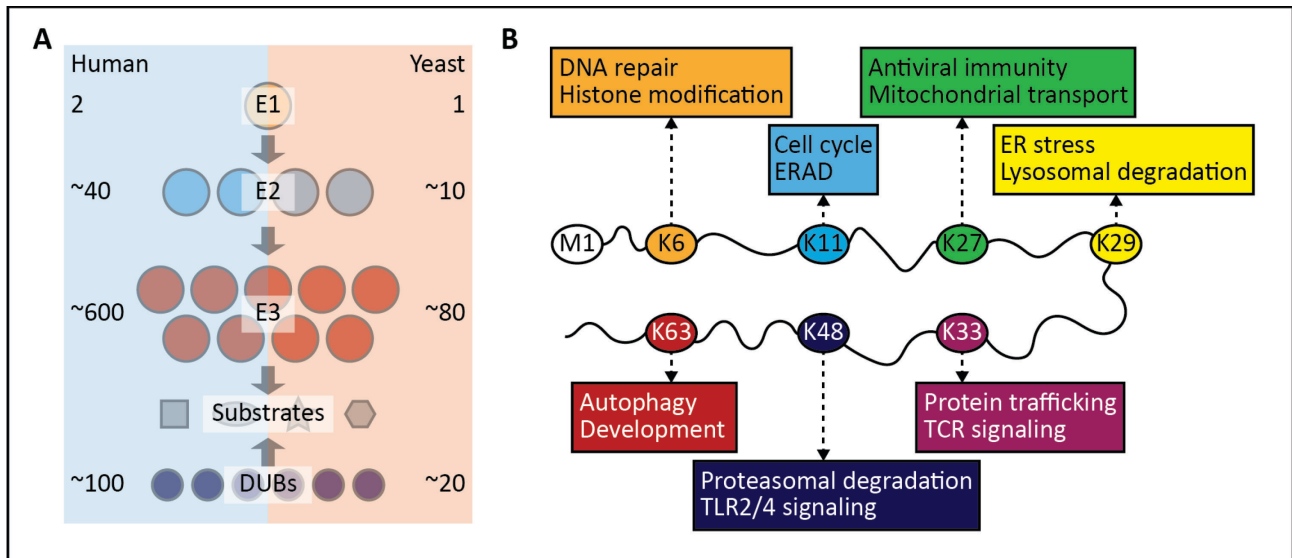


Figure 1: Components of the UPS and the various roles of ubiquitin in cellular functions.

(A) The UPS consists of E1, E2, E3, and DUB enzymes. The number of enzymes in each category are shown in the shaded regions for humans (blue) and the yeast *S. cerevisiae* (orange). The enzymes work in a cascade whereby one or two E1 enzyme activates a ubiquitin molecule, transfers it to tens of different E2s, and then with the help of hundreds of different E3s for substrate recognition, attaches the ubiquitin to the target protein. DUBs function to remove the ubiquitin added onto proteins. E3s and DUBs both have a degree of redundancy and multiplicity on substrate recognition. Within the UPS, eventually, the ubiquitinated substrate is targeted to the proteasome for degradation. (B) A ubiquitin molecule with the initiator methionine indicated as M1 and the seven different lysines (K6, K11, K27, K29, K33, K48, and K63) that would be possible for another ubiquitin to be attached. The major roles of the different ubiquitin chains built on the different lysines are shown.

1.1.2: Protein sequestration as a QC mechanism

In complementation to degradation, studies have shown that cells can also utilize sequestration and re-folding to maintain protein homeostasis (Chen *et al.*, 2011; Hill, Hanzén and Nyström, 2017; Kumar, Mathew and Stirling, 2022). Various forms of stress to the cell can lead to transcriptional upregulation of molecular chaperones such as the heat shock proteins (HSPs) in cases such as the induction of the environmental stress response (ESR) or unfolded protein response (UPR). Here, the cell utilizes a network of chaperones to bind to abnormal proteins that are misfolded and attempt

re-folding said proteins as this is likely faster in response to the stressors compared to degradation (Chen *et al.*, 2011). HSPs are also utilized to sequester abnormal proteins into compartments or inclusions in times of stress which can reduce the immediate burden on the QC mechanisms, preventing harmful interactions. While such responses using molecular chaperones can be beneficial, the inherent risk of this alternative method in QC of abnormal proteins is that it depends on the capacity of the chaperones and that the inclusions may be difficult to be resolved.

Examples of such sequestration compartments in the cytosol are the IPOD (insoluble protein deposit) and the JUNQ (juxtannuclear quality control site) (Kaganovich, Kopito and Frydman, 2008; Miller, Mogk and Bukau, 2015; Sontag, Samant and Frydman, 2017) (Figure 2, brown and blue regions). The two compartments have been shown to hold different types of proteins in different states for different end outcomes. The proteins sequestered into JUNQ are mainly misfolded proteins that can be extracted by chaperones such as Hsp104 to be refolded or sent for degradation. On the other hand, proteins in IPOD are generally insoluble aggregated proteins which are localized close to the vacuole near pre-autophagosomal structures. Within the nucleus, another such compartment has been identified and named as INQ (Intranuclear quality control site) (Miller *et al.*, 2015; Kumar, Mathew and Stirling, 2022). The INQ is similar to JUNQ whereby it is associated with various chaperones for disaggregation and re-folding, as well as, proteasomes for degradation (Figure 2, green region).

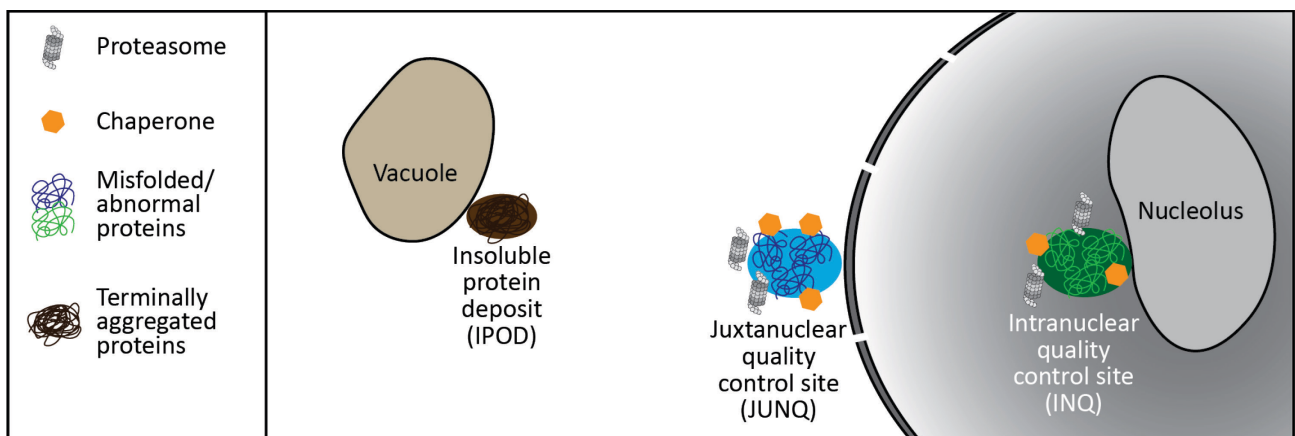


Figure 2: Location of various compartments that abnormal proteins can be sequestered into. The insoluble protein deposit (IPOD) is regularly found associated to the vacuole. Juxtannuclear quality control site (JUNQ) is localized in the cytosol but near the nucleus. Intranuclear quality control site (INQ) is located within the nucleus associated to the nucleolus. Both JUNQ and INQ are generally composed of misfolded/abnormal proteins that are brought there by sequestration chaperones, disaggregating chaperones, folding chaperones, and proteasomes. IPOD is usually composed of terminally aggregated proteins that can no longer be re-folded.

1.2 Mislocalized proteins

The protein QC systems identify and act upon a diverse range of abnormal proteins such as misfolded, mutated, damaged, or mislocalized proteins. In my PhD thesis, I will focus on mislocalized proteins (MLPs) and their QC systems. As the word mislocalize denotes, MLPs are proteins that fail to reach their native compartment to perform their functions (Figure 3A) (Hessa *et al.*, 2011; Hegde, 2014). Orphan proteins, proteins that are not assembled into complexes due to being in excess or lacking a binding partner, can be considered as a subset of MLPs since they fail to assemble into their native complexes (Figure 3B) (Juszkiewicz and Hegde, 2018).

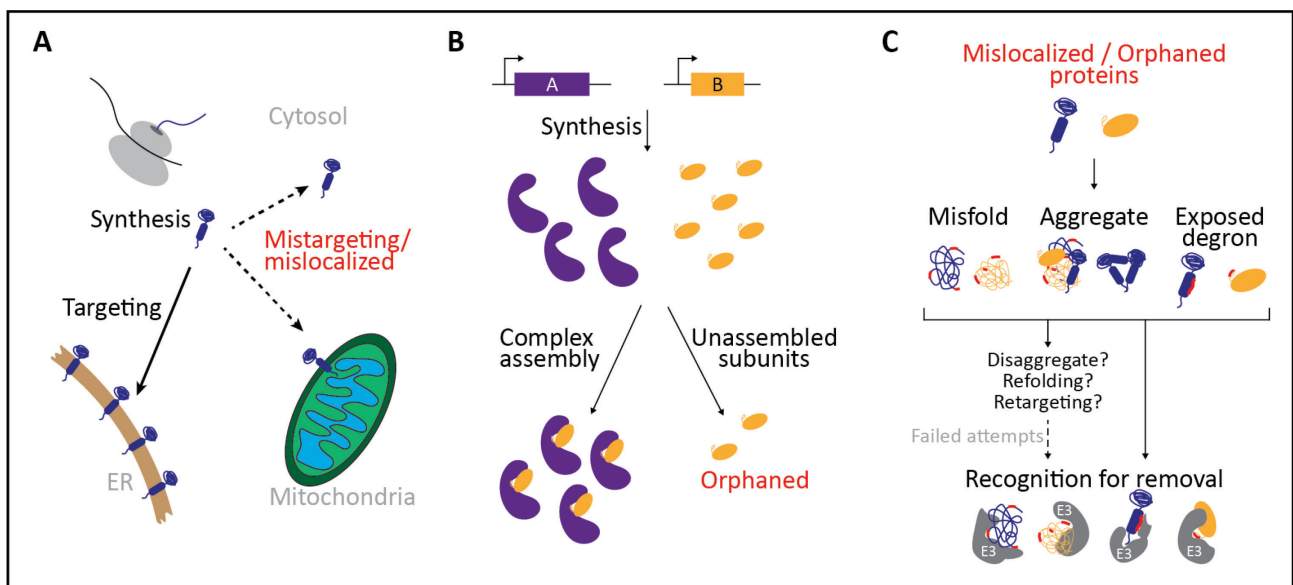


Figure 3: Mislocalized proteins and their possible fates.

(A) A protein that is synthesized and meant for the ER, but is mistargeted to the mitochondria or into the cytosol is a mislocalized protein. (B) In a hypothetical scenario of a two subunit complex of subunits A and B, one subunit is produced in slight excess than its partner. Post-complex assembly, the excess subunits are orphaned and are mislocalized proteins. (C) The mislocalized and orphaned proteins could possibly then misfold or expose previously buried interfaces which may then lead to aggregation. These exposed interfaces (red segments on the protein), which are usually buried upon proper folding, localization, or complex assembly, may be degrons that can then be recognized by an E3 and marked for removal. These abnormal proteins may also be given a second chance by attempting to disaggregate, refold, and/or re-engage again with targeting machinery to re-localize. However, when refolding and re-targeting again fails, the proteins need to be removed.

These MLPs have a tendency to misfold in the cytosol, have exposed surfaces that can be recognized as degrons, as well as form inappropriate interactions which may lead to aggregation, all of which are then subject to QC mechanisms for removal (Figure 3C) (Hegde and Zavodszky, 2019). In some cases, however, these MLPs may be given a second chance at re-folding and/or re-targeting

(Figure 3C) (Dederer and Lemberg, 2021). In terms of relevance to diseases, MLPs have been shown to be associated with various diseases such as Hutchinson-Gilford progeria syndrome, cancer, Thalassemia, etc (Table 2). Therefore, a clearer understanding of MLPs and their QC systems is important for dissecting the causes of diseases at a molecular level and may allow for the development of novel therapeutic strategies.

Table 2: Diseases associated with mislocalized proteins.

Disease	Gene/Protein	Normal localization	Abnormal localization	Reference
Amyotrophic Lateral Sclerosis	TDP-43	Nucleus	Cytoplasm	(Suk and Rousseaux, 2020)
Triple A syndrome	ALADIN	Nuclear Pore Complex	Cytoplasm	(Cronshaw and Matunis, 2003)
Cancer	Retinoblastoma protein	Nucleus	Cytoplasm	(Wang and Li, 2014)
	Mucin-13	Transmembrane	Cytoplasm, nucleus	
	EBP50	Cytoplasm	Nucleus	
	KL-6 mucin	Cytoplasm, cell surface	Nucleus	
Cystic Fibrosis	CFTR	Plasma membrane	ER	(Bridges and Bradbury, 2018)
GLUT1 deficiency syndrome	GLUT1	Plasma membrane	Endosomes, Golgi	(Meyer <i>et al.</i> , 2018)
Hutchinson-Gilford progeria syndrome	Ubc9	Nucleus	Cytoplasm	(Kelley <i>et al.</i> , 2011)
Fukuyama-type congenital muscular dystrophy	Fukutin	Golgi	ER	(Tachikawa <i>et al.</i> , 2012)
Thalassemia	α -globin, β -globin	In complex as hemoglobin $\alpha_2\beta_2$	No complex formation OR formation of β_4 tetramers	(Higgs, 2013)

1.2.1: Mistargeting or loss of targeting signal as a source of MLPs

Most proteins rely on localization signal sequences that are present within the protein sequence to be targeted to various compartments (Hung and Link, 2011; Juskiewicz and Hegde, 2018). These localization signals are generally not specific motifs, but rather, rely on certain properties such as hydrophobicity, length, and charge. Regardless if it is properties or a specific sequence, these localization signals can be easily disrupted if there are mutations within the amino acid sequence or a mistake in recognising similar properties, which then leads to proteins being mistargeted to the wrong compartments (Juskiewicz and Hegde, 2018). Furthermore, studies in mammalian cells have shown that efficiency in signals for the secretory pathway ranges between 95% and 60%, which indicates even with the best signal there is still a 5% chance of mislocalizing (Levine *et al.*, 2005).

Firstly, an example of a MLP that occurs due to mistargeting is Pex15. Pex15 is a peroxisomal tail-anchored (TA) protein that relies on the GET (Guided Entry of Tail-Anchored) pathway for correct localization to the peroxisomal membrane (Schuldiner *et al.*, 2008). In a later study, Pex15 was shown to mislocalize more to the mitochondria instead of localizing to the peroxisome when the GET pathway or Msp1 (ATAD1 in mammals) is compromised (Okreglak and Walter, 2014). Indeed, another study also found Pex15 (PEX26 in human) and Gos1 (GOS28 in human), a Golgi TA protein, to both be mislocalized to the mitochondria upon compromising the GET pathway and Msp1 (Chen *et al.*, 2014). The GET pathway is the targeting pathway for these proteins and Msp1 is an AAA-ATPase that plays an important role in extracting these wrongly targeted proteins in the mitochondria. These wrongly targeted proteins that are extracted can either be sent for degradation or for re-targeting to the correct localization (Okreglak and Walter, 2014; Dederer and Lemberg, 2021). The Msp1 clients such as Pex15, Fmp32, and Gem1, that are extracted are then targeted for degradation through the E3 ubiquitin ligase Doa10 (Dederer *et al.*, 2019). The existence of proteins such as Msp1 that plays a role in extracting wrongly targeted proteins reasons that mislocalization does take place within the cell and there must be substrates for it to carry out its function.

Next, in the transmembrane recognition complex (TRC) pathway (mammalian equivalent of the GET pathway), a ribosome associating factor BAG6 can promote the association of TA proteins with their targeting factors (e.g. TRC40) for insertion into the endoplasmic reticulum (ER) (Mariappan *et al.*, 2010). Interestingly, BAG6 can also carry out the QC of mislocalized TA proteins by working together with the E3 ligase RNF126 to ubiquitinate the captured mislocalized TA proteins to target it for degradation (Rodrigo-Brenni, Gutierrez and Hegde, 2014). Another study showed that a source of mislocalized proteins that are BAG6 substrates can come from the AAA-ATPase p97 (Cdc48 in yeast) that extracts misfolded proteins from the ER membrane into the cytosol (Hu *et al.*,

2020). Additionally, the BAG6 complex is also an important part of MLP QC machinery in removing mammalian prion proteins in the cytosol to prevent it from aggregating, interacting inappropriately, and leading to cell death (Hessa *et al.*, 2011).

Besides that, the loss or disruption of targeting signals, such as the ER signal sequences, can also lead to mislocalization of various proteins. The yeast proteins carboxypeptidase Y and vacuolar proteinase A mislocalizes into the cytosol and misfolds if the ER signal sequence is removed (Prasad, Kawaguchi and Ng, 2010). These proteins are then targeted for degradation by the E3 ligases San1 and Ubr1. In human cells, the membrane protein IL-2 receptor α subunit (IL-2 α) was shown to also mislocalize to the cytosol upon truncation of its ER signal sequence (SS) (Suzuki and Kawahara, 2016). The degradation of IL-2 α Δ SS MLP is dependent on UBQLN4 for recognition and BAG6 for ubiquitylation. The examples here remove the ER signal sequences of the proteins which would be similar to a truncation mutant that could occur due to a mutation. It is also possible that a mutation in the signal sequence itself may disrupt the targeting process and lead to mislocalization.

1.2.2: Compromised protein import leads to protein mislocalization

Subsequently, a different scenario where MLPs can arise is when protein import across membranes is compromised in situations such as faulty channel proteins, mutations on the imported protein, or presence of excess proteins leading to clogged channels. A recent study by Shakya *et al.* (2021), showed that blocking mitochondrial import leads to mitochondrial proteins mislocalizing to various different sub-cellular compartments where some are then degraded. One source of inhibition of mitochondrial import is due to the increase in oxidative stress (Wright *et al.*, 2001), which increases with aging and disease (Liguori *et al.*, 2018; Chen, Kroemer and Kepp, 2020). In conditions where there is an import block at the mitochondrial outer membrane, a study showed that Cis1 recruits Msp1 to the TOM complex to extract the proteins that are stuck in the channel and blocking import (Weidberg and Amon, 2018). Besides that, sometimes the components of a transport channel itself can become mislocalized and this causes a downstream effect of more protein mislocalization due to missing import channels. One such example is in Alzheimer's disease, where the tau protein interacts with proteins of the nuclear pore complex, NUP98 and NUP62, and causes mislocalization of these nuclear pore proteins into the cytosol which then contributes to fibrilization and aggregation (Eftekharzadeh *et al.*, 2018). The mislocalization of these nuclear pore proteins and the aggregation further disrupts the complex structure and function which then causes a cascade of problems as nuclear protein import is impaired, which then leads to accumulation of other mislocalized proteins.

Next, the E3 ligase Doa1, which is involved in targeting some Msp1 clients for degradation, also functions within the pre-insertional ERAD pathway which monitors cytosolic proteins for secretory proteins that fail to be translocated into the ER and therefore are mislocalized (Ast *et al.*, 2014). A protein with a glycosylphosphatidylinositol (GPI) anchor sequence (AS) is translated by the ribosome on the cytosolic side of the ER and it would need to translocate into the ER. Doa10 works together with Ubc7 to ubiquitinate these GPI AS-bearing proteins on the cytosolic side of the ER and Ubp1, a deubiquitylase, works to remove the ubiquitin molecule to allow these proteins to try and translocate. However, if these proteins remain mislocalized in the cytosol the chance of degradation by the proteasome would increase.

1.2.3: Excess proteins/orphaned proteins as source of MLPs

Besides faults in protein targeting and import, presence of excess proteins are also sources of MLPs. These excess proteins can be from partner subunits that are not perfectly stoichiometrically produced or by dysregulation of levels of proteins, especially proteins that require dedicated chaperones for holding or folding. Aneuploid cancer is one such example of a condition in which there is generally a deregulation in protein levels which leads to presence of excess proteins (Geiger, Cox and Mann, 2010; Stingle *et al.*, 2012; Bonney, Moriya and Amon, 2016). In the study carried out by Dephoure *et al.*, (2014) in which they measured protein levels in aneuploid yeast strains carrying a single extra chromosome, they found that about 15% of proteins on the additional chromosome were attenuated. The proteins that were attenuated due to being in excess were enriched for subunits of complexes, which is not as surprising as these subunits likely cannot form their native complexes due to mismatched levels (i.e. mislocalized orphan proteins). It is also quite likely that the excess proteins could mislocalize to the wrong compartment(s) due to the overwhelmed protein QC system. In another study using the same set of aneuploid yeast, Brennan *et al.*, (2019) found that about 25% of proteins on the additional chromosome are aggregated. These aggregated proteins do not overlap well with the proteins found to be attenuated by degradation in Dephoure *et al.*, (2014) and so the two groups of proteins are considered to be mutually exclusive.

Additionally, in terms of orphan proteins of complexes, the cytosolic ribosomal proteins are good examples that excess orphan proteins are very rapidly degraded by the UPS to maintain a good balance of subunits present within the cell (Sung, Reitsma, *et al.*, 2016). The degradation of these excess orphan proteins is achieved through the action of E2 enzymes Ubc4/Ubc5 and the E3 ubiquitin ligase Tom1 (HUWE1 in humans) (Sung, Porras-Yakushi, *et al.*, 2016; An and Harper, 2020).

Besides HUWE1, UBE2O has also been implicated in the degradation of RPL24 and the ubiquitination of RPL3, RPL8 which are human ribosomal subunits, and ubiquitination of α -globin (HBA1), a subunit of haemoglobin, in experiments using reticulocyte lysate (Yanagitani, Juszkiwicz and Hegde, 2017). Similar to excess cytosolic ribosomal subunits, OST complex subunits in yeast have also been shown to be degraded when in excess by ERAD components Doa10 and Hrd1 (Mueller *et al.*, 2015). Interestingly, Mueller *et al.* (2015) found that the subunits that were already fully assembled into a complex were not perturbed, but only the excess unassembled subunits were degraded. Subsequently, a proportion of subunits of complexes were shown to be degraded non-exponentially, that is, the proteins are less stable at early time points and stabilize over time (Mcshane *et al.*, 2016). These findings indicate that the QC systems manage complex levels by preferentially removing excess orphan subunits rather than destroying an already assembled complex.

Furthermore, aging tends to go hand-in-hand with reduced or impaired gene regulations as well as impaired QC systems, both of which may also cause presence of excess proteins, and so, a source of MLPs (Higuchi-Sanabria *et al.*, 2018; Koyuncu *et al.*, 2021; Kumar and Lapierre, 2021). For example, studies in *C. elegans* show that with aging, there is a loss in proteostasis capacity leading to increased protein mislocalization and aggregation (Ben-Zvi, Miller and Morimoto, 2009). Also in *C. elegans*, there is a general loss of ubiquitination with aging through upregulation of DUBs, and this could lead to reduced degradation and protein mislocalization (Koyuncu *et al.*, 2021).

1.2.4: Methods of recognition of MLPs for removal

The utilization of many of the UPS components in degrading MLPs brings up the question on how are these MLPs being recognized. Recognition of substrates for degradation can be achieved through specific motifs known as degrons (Timms and Koren, 2020). These degrons can occur at the terminals, also known as N-degron or C-degron, or even internally. Approximately 60% and 80% of proteins in yeast and humans respectively are N-terminally acetylated and this acetylation creates a specific signal for the Acetyl/N-degron (Ac/N-degron) pathway (Hwang, Shemorry and Varshavsky, 2010; Varshavsky, 2019). Thus it is likely that the Ac/N-degron may play an important role in recognition of MLPs and targeting for degradation. The acetylated N-terminal may be shielded from recognition upon proper folding or formation of complexes. This shielding phenomenon has been shown for several subunits of complexes such as Hcn1 in the APC/C ubiquitin ligase of *S. pombe* and Cog1 in the Golgi-associated COG complex in budding yeast (Shemorry, Hwang and Varshavsky, 2013). Similarly for the C-degrons, it has been shown that C-degron pathways are involved in

recognition and removal of MLPs (Yeh *et al.*, 2021). This recognition can be for membrane proteins that are inserted in the wrong direction, truncated or cleaved proteins when mislocalized, and also shielded C-terminals such as in a ubiquitin molecule (Lin *et al.*, 2018; Yeh *et al.*, 2021).

While the QC systems governing the MLPs are not well known, they are definitely important for cell survival and are highly relevant in aging and disease. Some of these QC systems for MLPs have been found and described over the years, but there are potentially many more QC systems and pathways for MLPs that are yet to be identified. Thus, the questions remains on what are the other QC systems and pathways for MLPs and how are MLPs being recognized?

1.3 Project aims and goals

The larger goal of the project is to systematically dissect quality control systems for MLPs through understanding the properties that define and differentiate MLPs and identifying the mechanisms that lead to the removal of these abnormal proteins.

For my PhD thesis specifically, I first aimed to track the fate of proteins when they are mislocalized through overexpression and determine which proteins are attenuated when mislocalized. By challenging the cell's QC systems with excess proteins, I expect that the necessary chaperone and targeting systems become overwhelmed and subunits would exceed their available binding partners within a complex. Thus, the cell would need to remove these MLPs which leads to a decrease in protein levels (i.e. the protein is attenuated).

After identifying MLPs that are attenuated upon overexpression, I will look into various protein properties that would characterise these attenuated MLPs. I will analyze the properties such as synthesis rates, unstable N-termini, protein disorderness, hydrophobicity, and sub-cellular localization on their role in determining the attenuation behaviour of MLPs.

Using the information gathered on MLPs that are attenuated upon overexpression, I will then also test whether previous findings, whereby majority of degraded proteins upon overexpression are subunits of multi-molecular complexes, hold true. If it holds true, I intend to follow up on the outcome by looking at the attenuation behaviour of each subunit upon overexpression of another partner subunit to understand how a subunit affects its partner subunits' likelihood of being attenuated.

Additionally, I will also look into if there are other methods that the cell uses to cope with the overexpression of these proteins, subunits and non-subunits, besides degradation.

The outcomes from all of this exploration will hopefully provide a clearer picture on which MLPs are attenuated upon overexpression and the characteristics of these proteins that increase the likelihood of being attenuated. Additionally, the large dataset will provide a starting point to understand MLPs and further try to identify the protein QC systems involved in surveillance and removal of MLPs.

I intend to follow the fate of mislocalized proteins through overexpression of individual proteins across the proteome. In order to carry out my research, I needed a model organism to work with that would allow for a high-throughput screen of the entire proteome, a method to follow the fate of a protein upon mislocalization, and a method to induce mislocalization. I would follow the fate of a tagged endogenous protein upon mislocalization and I expect that the level of the tagged protein would be attenuated (less than expected) if it is subjected to QC systems for removal.

1.1 High-throughput studies using Synthetic Genetic Array (SGA)

For my PhD project, I chose to work with the model organism *Saccharomyces cerevisiae*. An advantage in using *S. cerevisiae* (from hereafter referred to as yeast) is that its proteome is well described and annotated, genetic manipulations can be relatively easily introduced/performed, and there are many options of various knock-out, overexpression, and tagged-ORF (open reading frame) libraries of the entire yeast proteome available (Winzeler *et al.*, 1999; Ghaemmaghami *et al.*, 2003; Duina, Miller and Keeney, 2014). Additionally, by using yeast, I can perform high-throughput screens in arrayed formats in order to test the entire proteome within a reasonable time frame.

To achieve the high-throughput screening of the proteome, I utilized the Synthetic Genetic Array (SGA) method to select for the haploid yeast strains containing the modification and perturbation that I need (Tong *et al.*, 2001; Tong and Boone, 2006). I mated haploid yeast strains carrying a tagged protein of interest with another haploid yeast strain of an opposite mating type carrying the perturbation (Figure 4). The mated strains are then put through sporulation and several selection steps to finally obtain haploid strains carrying both the tagged protein and perturbation which are then used for measurements.

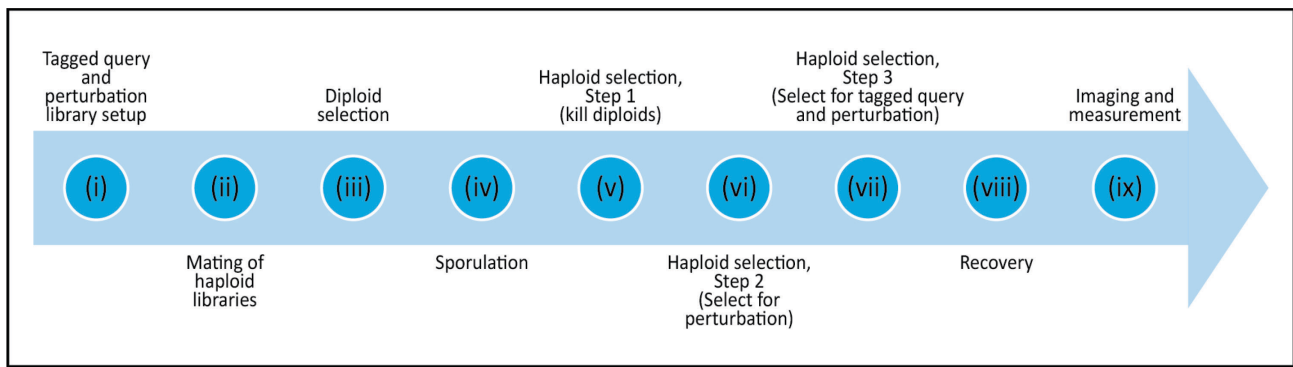


Figure 4: The steps taken to efficiently construct a haploid carrying modifications of interest from two different haploids in an array format using the SGA method.

(i) The yeast strains carrying the proteins to be studied (tagged query) and yeast strains carrying the perturbation to be applied are arrayed onto agar as colonies. (ii) The two haploid libraries of opposite mating types are mated with each other. (iii) Diploid selection is performed by using antibiotics and/or amino-acid dropouts plates to select for diploids that carry both traits, tagged query and perturbation. (iv) Diploids are then placed on sporulation agar. (v-vii) From sporulation, haploids are selected using a series of antibiotics and/or amino-acid dropouts plates to select for haploid strains that carry both desired traits, tagged query and perturbation. (viii) A recovery step with similar antibiotics and/or amino-acid dropouts plates as in vii is carried out to even out growth differences from all the selections. (ix) The final haploids strains are then ready for imaging and measurement. Colonies are grown for 24 hours before imaging and colony fluorescence measurements.

1.2 Following protein fate using a tandem fluorescent protein timer (tFT)

In order to study mislocalized proteins, I needed a method to track and follow the fate of a protein upon mislocalization. To this end, I decided to use a tandem fluorescent protein timer (tFT) tag to follow the fate of a protein. The tFT tag consists of a fast maturing green fluorescent protein and a slower maturing red fluorescent protein (Khmelniskii *et al.*, 2012). The fluorescent readout from the green signal provides an indicator of protein abundance while the ratio of red to green signal can provide an indicator of protein stability. By comparing the fluorescent readouts of a perturbed strain against wild type, I then determined the changes in abundance and protein turnover upon perturbation.

In order to perform the experiments in a high-throughput manner, I needed a way to tag the entire yeast proteome very quickly with my chosen tFT tag. For this, I used the SWAp-Tag (SWAT) method and the C-SWAT library to tag each protein at the C-terminal with the tFT tag through SGA (Yofe *et al.*, 2016; Meurer *et al.*, 2018; Weill *et al.*, 2018) (Figure 5). I decided to use the Type-II SWAp-Tag procedure to tag the proteins so that I have a hygromycin selection marker to use in the SGA selection process and for future crosses. In the end, I would have a tFT library which consists of ORFs tagged with the tFT tag at the C-terminal with an ADH1 terminator.

For this project, I constructed tFT libraries using two different tFTs, mCherry-sfGFP (mCH-sfGFP) and mCherry-mNeonGreen (mCH-mNG). I first compared the two tFT libraries to each other to characterise them. After comparing the two tFT libraries, I decided to use the mCH-mNG tFT for further experiments. The details are discussed further in the Chapter 1, Section 1.1.

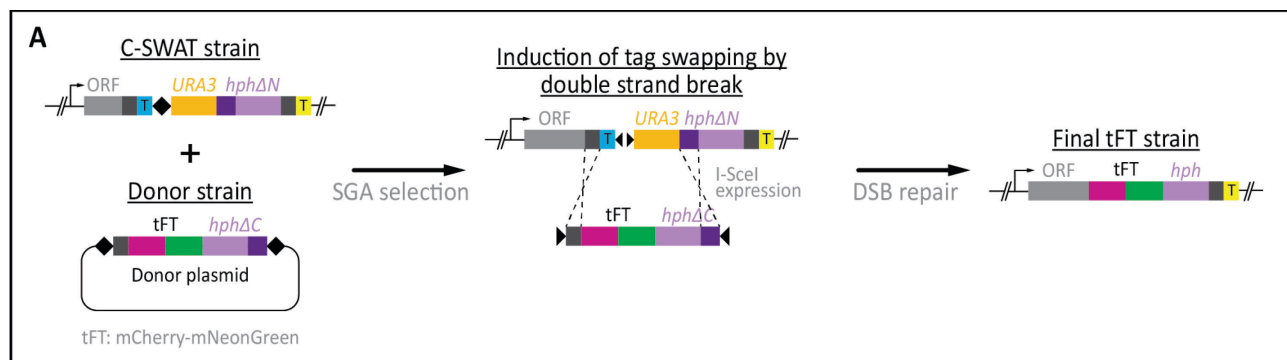


Figure 5: Construction of the tFT strain from a C-SWAT starting strain and donor plasmid. The C-SWAT strain is crossed with a donor strain carrying the tFT donor plasmid. The tFT donor plasmid contains the mCH-mNG tFT and part of a hygromycin resistance cassette. The final haploid strains carrying both the C-SWAT cassette and the tFT donor plasmid are selected using the SGA method. Induction of tag swapping is done by inducing expression of I-SceI, a homing enzyme, to cause double strand breaks at the designated cut site on the C-SWAT cassette and the tFT donor plasmid. The tFT tag is then inserted at the desired locus through double strand break repair which also reforms the full hygromycin resistance cassette for selection of the final strain.

1.3 Mislocalizing protein through overexpression

The next piece needed for my project is a method to induce protein mislocalization. Although protein mislocalization occurs within the cell, the level of mislocalization is likely too low to be detected without any perturbation. Therefore, I decided to use protein overexpression as a method to induce mislocalization. The presence of excess proteins through overexpression should lead to a stronger attenuation of MLPs which can then be better detected and quantified. Protein overexpression also mimics the condition in aneuploid cancer cells when protein production is dysregulated due to extra chromosomes being present. However, by studying MLPs through individual protein overexpression, I expect that the outcomes arising from minimal perturbations to the cell will be unbiased and not due to off-target or global effects.

To achieve different levels of overexpression, I used two types of overexpression plasmids, a low-copy centromeric plasmid with a CEN/ARS origin sequence or a high-copy plasmid with a 2 μ origin sequence. The CEN plasmid usually exists as 1-2 copies within the cell and are relatively stable

(Mumberg, Müller and Funk, 1995). The 2micron (2μ) plasmid is a high-copy plasmid with copy numbers of around 20 copies (Christianson *et al.*, 1992). The copy number of these 2μ plasmids are further affected by the presence of different antibiotic and auxotrophic selection markers on the plasmid that can lead to higher or lower copy numbers, likely due to differing burden on the cell (Karim, Curran and Alper, 2013). (Note: The 2μ plasmids that are commonly used nowadays should not be mistaken with the 2μ circle plasmid which does exist in much higher copies of around 40-100 copies due to presence of a site-specific recombinase FLP and the Flp recombination target site (Zakian, Brewer and Fangman, 1979; Chan *et al.*, 2013; Yen-Ting-Liu *et al.*, 2014)).

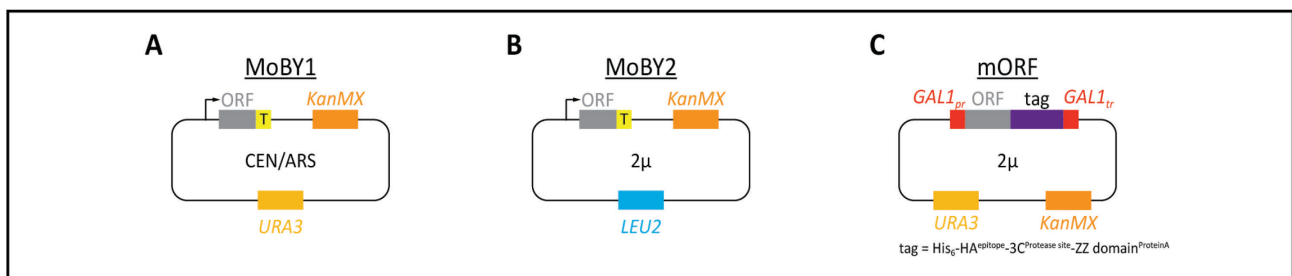


Figure 6: Schematic of the overexpression plasmids used in this study.

Important features on the overexpression plasmids from the MoBY1 (A), MoBY2 (B), and mORF (C) libraries. The MoBY1 (A) and MoBY2 (B) library plasmids have the ORF under its own endogenous promoter and terminator and both carry a KanMX selection marker. The MoBY1 (A) plasmid is a CEN plasmid and has a *URA3* selection marker while MoBY2 (B) is a 2μ plasmid and has a *LEU2* selection marker. The mORF (C) library plasmids are 2μ plasmids and have the ORF under a GAL1 promoter and terminator, and tagged with an affinity purification tag. The mORF (C) plasmids also carry a KanMX and a *URA3* selection marker.

There are three previously constructed overexpression libraries that I decided to use which would cover a range of different overexpression levels. Firstly, the MoBY1 library which consists of CEN plasmids carrying the ORF with its endogenous promoter and terminator present (Figure 6A) (Ho *et al.*, 2009). This is a low level overexpression with the expectation that there would be 1-2 copies of the plasmid per cell and with the endogenous promoter, the expression level would also be increased by a similar fold. Secondly, the MoBY2 library which consists of 2μ plasmids carrying the ORF with its endogenous promoter and terminator (Figure 6B) (Ho, 2011; Magtanong *et al.*, 2011). This library would lead to a stronger level of overexpression since the expectation is that there will be about 20 copies of the plasmid per cell which again should scale the expression levels similarly since it is also using the endogenous promoter and terminator. Thirdly, the mORF library which consists of 2μ plasmids carrying the ORF under a GAL promoter for very strong overexpression

(Figure 6C) (Gelperin *et al.*, 2005). While the plasmid copy number (PCN) should be similar to that of the MoBY2 library, the GAL promoter would lead to increased transcript levels, depending on the native promoter. Since it is on a different promoter, the levels of overexpression would technically be much stronger for a very low expressed protein as compared to a protein that is normally expressed at levels close to that of the GAL1 gene. Also, the ORF in the mORF library plasmids are tagged with a triple affinity tag His₆-HA^{epitope}-3C^{Protease site}-ZZ domain^{ProteinA}.

In summary, I would expect that the fold overexpression to be same for every ORF within the MoBY1 and MoBY2 libraries respectively, however, for the mORF library, the fold overexpression depends on the relative strength of the GAL1 promoter versus the native promoter and the absolute overexpression levels is expected to be similar for different ORFs. Additionally, for all three overexpression libraries, I needed a control plasmid as a comparison to simulate the native condition but still have all the needed selection markers for SGA. Therefore, I chose to use the plasmid carrying the ORF YGR045C, an uncharacterized and putative protein of unknown function (Fisk *et al.*, 2006), as the control plasmid for each library.

1.4 Expectations on attenuation behaviour upon overexpression

Now that I have outlined all the components needed, I next outline the expectation for the different outcomes from the overexpression screens that will be conducted. In the overexpression screens, measuring the red and green fluorescence levels of the tFT-tagged strain with the control plasmid constitutes the native phenotype and so provides a readout of the protein abundance (green fluorescent protein signal) and stability (red-to-green signal ratio) at native conditions (Figure 7A). Measuring the red and green fluorescence levels of the tFT-tagged strain with the corresponding overexpression plasmid will provide a readout of protein abundance and stability in the overexpression condition (Figure 7B). Then, I would compare the green fluorescence levels and red-to-green ratio in the overexpression condition against the native condition.

There are two likely outcomes that I expect from the overexpression screen. Upon overexpression, one outcome would be that the protein is not a substrate of QC systems for MLPs upon mislocalization through overexpression and so the fluorescence and ratio levels of the overexpression condition should be similar to the native condition (Figure 7B(i)). The second outcome is that the protein is a substrate of QC systems for MLPs and so the fluorescence and ratio levels would be lower in the overexpression condition than the native condition which indicates that

the protein is attenuated (Figure 7B(ii)). The expectations stated above are with the assumption that the cell does not differentiate between the tagged and untagged version of proteins.

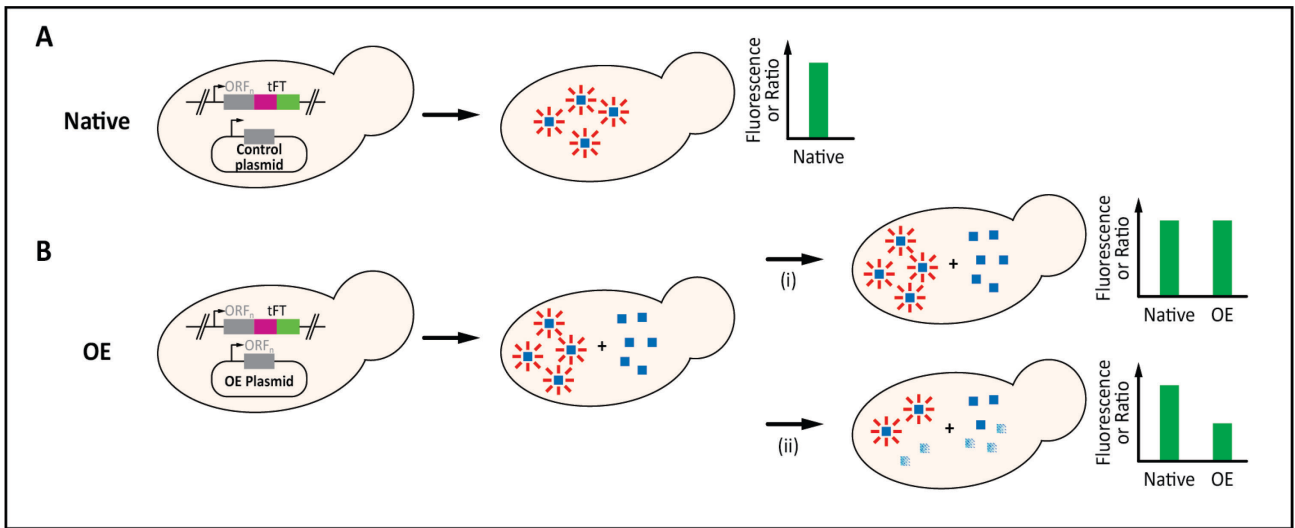


Figure 7: Expected behaviour of proteins mislocalized through overexpression

(A) A tFT strain with the endogenous protein tagged with tFT carrying a control plasmid gives the fluorescence/ratio level of the native condition. (B) A tFT strain with the same tFT-tagged endogenous protein carrying an overexpression plasmid with the corresponding protein that is untagged gives the fluorescence/ratio level of the overexpression condition (OE). With overexpression, (i) if the protein is not attenuated by the cell, then the fluorescence/ratio level in the OE condition remains similar to the native condition; (ii) if the protein is attenuated by the cell, then the fluorescence/ratio level is expected to decrease in the OE condition compared to the native condition.

CHAPTER 1: Setup and preparation for studying protein mislocalization

This chapter will detail the preparation and characterization of the tFT library, preparation and checks on the overexpression libraries, and layout of the tFT-overexpression screens.

1.1 Constructing and characterizing the tFT library

As noted before, in order to follow the fate of proteins upon overexpression to induce mislocalization, I decided to use the tFT tag to determine changes in protein abundance and stability.

Here, I needed to decide on which fluorescent proteins that I would need to use in the tFT. The red fluorescent protein mCherry was chosen since it has been well characterized in use as a tFT with the sfGFP green fluorescent protein (Khmelniskii *et al.*, 2012). This mCH-sfGFP tFT would allow for studying degradation of proteins with half-lives of between ~10 minutes to ~8 hours. However, the sfGFP fluorescent protein was then later described to have incomplete proteasomal degradation due to its tight folds which leads to accumulation of tFT fragments which may affect the tFT's ability as a degradation reporter (Khmelniskii *et al.*, 2016).

With this in mind, I needed a different green fluorescent protein to be used in the tFT tag with mCherry since I would want to mainly first determine changes in protein abundance and so issues with sfGFP makes it less suitable. The mNeonGreen (mNG) fluorescent protein was characterized in Shaner *et al.*, (2013) and has been shown to be a much brighter green FP than sfGFP. mNG also does not bear the incomplete degradation problem as sfGFP (Khmelniskii *et al.*, 2016). This two reasons together make it a better choice for my experiments.

Thus, I set out to first characterize the mCH-mNG tFT tag by constructing the mCH-mNG tFT library through SGA tagging using the C-SWAT library and the SWAp-Tag method. From the constructed mCH-mNG tFT library, I chose 20 strains from the tFT library to assess the tagging efficiency via flow cytometry. The flow cytometry results show that the tagging efficiency of the C-SWAT system is remarkable as I can achieve a near 100% (>95%) tagging efficiency to tag every ORF with the mCH-mNG tFT (Figure 8A).

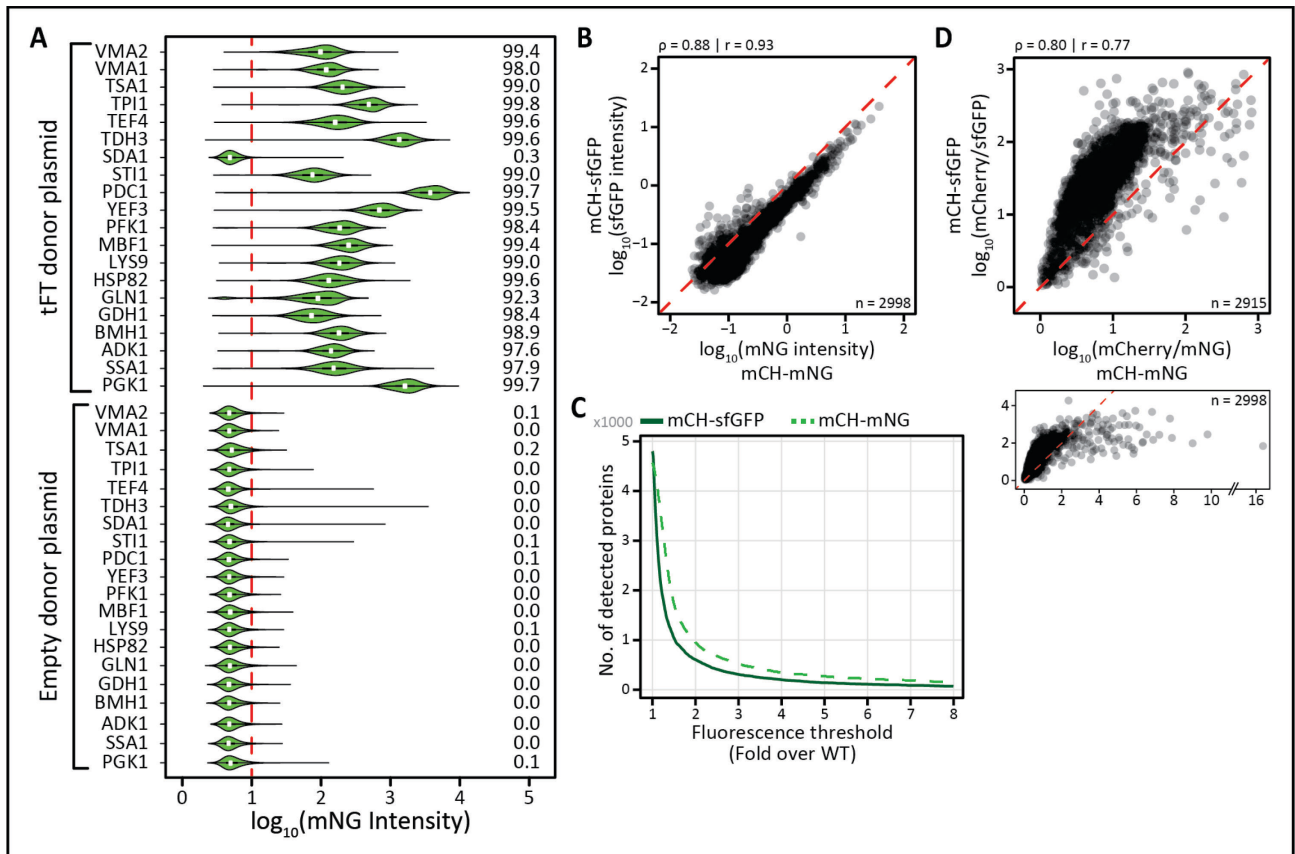


Figure 8: Characterizing the mCherry-mNeonGreen tFT and comparing it to mCherry-sfGFP. (A) Representative flow cytometry results of tFT tagged strains constructed using the tFT donor plasmid and non-tFT tagged strains constructed using an empty donor plasmid. Violin plots show the distribution of the mNG intensity in the mCherry-mNG tFT strains. Red dotted line indicates threshold for calling non-fluorescent cells. Numbers on the right indicate the percentage of cells that are above the threshold. (B) Log₁₀ green fluorescence intensity of the tFT strains for mCH-sfGFP against mCH-mNG as measured through colony fluorescence measurements. (C) Number of proteins detected at different fold over wild-type thresholds for mCH-sfGFP (solid line) and mCH-mNG (dotted line) tFTs. (D) Log₁₀ red-to-green ratio of the tFT strains for mCH-sfGFP against mCH-mNG as measured through colony fluorescence measurements with limited axis (top) and unlimited axis (bottom).

Using the same process, I also constructed the mCH-sfGFP tFT library for comparison against the mCH-mNG tFT. As shown in Figure 8B, the fluorescence level of mNG correlates very well with sfGFP but mNG has a higher intensity. With this higher intensity, I also have a better signal-to-noise outcome from the mCH-mNG tFT as I can detect more proteins than the mCH-sfGFP tFT with the same fluorescence threshold on fold over wild type (Figure 8C). However, in terms of ratio, I can see that mCH-mNG tFT has less resolution at the lower ratios where the bulk of data is located (Figure 8D). This outcome is likely due to the accumulation of incompletely degraded tFT fragments that facilitate better detection of very low abundant proteins that are extremely unstable (Khmelnikii *et*

al., 2016). From this point on, “tFT library” will refer to the mCH-mNG tFT library unless otherwise specified.

Additionally, I also constructed another tFT library still with mCH-mNG, however, using a different donor plasmid constructed with a different selection marker (NatMX instead of KanMX). This library was created in lieu of the results that when performing a yeast knockout tFT screen, I could not perform proper haploid selection as the yeast knockout collection was constructed using KanMX marker as the knockout cassette (See Chapter 7). In brief, the final tFT strain possibly still carried some of the original donor plasmid even without maintained selection pressure throughout the tFT library construction after sporulation. The presence of this “leftover” donor plasmid then interferes with further SGA selections as the marker can no longer be used and so reduces the number of available selection markers.

1.2 Preparing the overexpression libraries

While the MoBY2 and mORF libraries are both available and present in yeast, the MoBY1 library exists as an *E. coli* library. Thus, I first have to mini-prep the plasmids out of the *E. coli* strains and then transform the plasmids into the yeast strain BY4741. BY4741 was chosen as it is the same background as the MoBY2 library and also a compatible mating type to the tFT library.

To start off, the MoBY1 library plasmids were extracted using a 96-well mini-prep kit which provided the plasmids in a 96-well plate format. The plasmid extraction was done together with the lab technician [REDACTED]. I then performed some digestion checks and sequenced some of the plasmids to confirm the plasmid identities and that they were in the correct well before proceeding with transformation into BY4741. Transformation was also done in a high-throughput format in a 96-deep well plates and plated out using the Singer RoToR for single colony selection, storage, and further use.

In the process of sequencing some of the plasmids from the MoBY1 library, I noted that a handful of plasmids have some point mutations within them. After the overexpression screens were completed, during validation, more plasmids with mutations were found. The decision was then made within the project group to fully sequence all plasmids in the MoBY1 and MoBY2 library. These two libraries were fully sequenced using next generation sequencing and the outcomes are discussed in Chapter 3.

1.3 Assembly of libraries into a screening layout

For the overexpression screens, the tFT, MoBY1, MoBY2, and mORF libraries were re-arrayed using the ROTOR pinning robot (Singer Instruments) into a 1536 colony array format. Each 4x4 grid would have the matching ORF from the tFT and overexpression libraries in triplicates of the tFT strain and a non-fluorescent strain crossed with the overexpression strain carrying the corresponding ORF and the native condition control strain carrying the plasmid with *YGR045C* ORF (Figure 9A & 9B). A set of reference strains were placed on each plate at the bottom right position in a 2x2 grid. The set of reference strains will be used for normalizations to correct for plate effects across different plates and spatial effects within the same plate. On the plates are also border strains consisting of dummy strains to reduce any border effects due to more availability of nutrients to the outermost colonies in the array. A more detailed explanation on the layouts and reasons for the various different strains when conducting high-throughput screens with tFTs can be found in Fung, Blöcher-Juárez and Khmelinskii (2022).

I then conducted the overexpression screens by crossing the tFT library strains with the overexpression libraries laid out as described. Using the SGA selection method, haploid strains carrying both the tFT tagged ORF and overexpression plasmid are selected (Figure 9C). I then measured the colony fluorescence of the tFT-MoBY1 and tFT-MoBY2 screen set using a plate reader. For the tFT-mORF screen, I first transferred the strains onto galactose media to induce overexpression and then carried out colony fluorescence measurements using a plate reader.

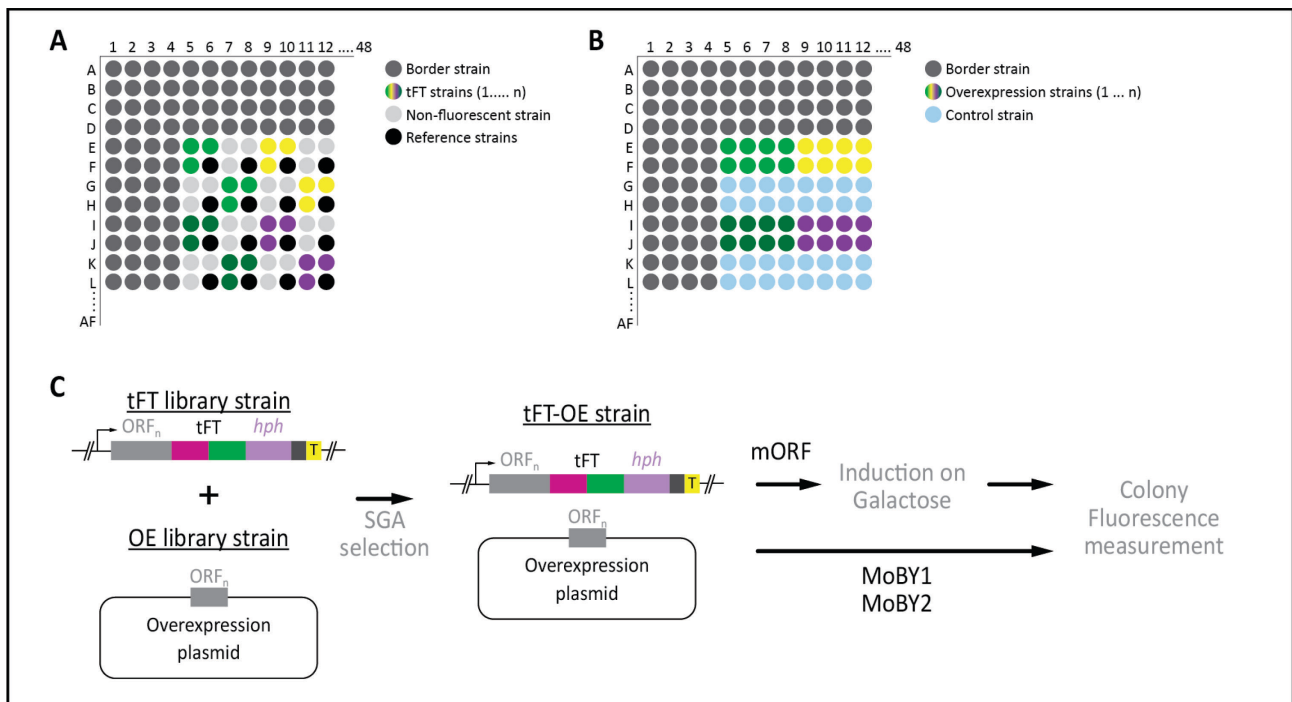


Figure 9: tFT-overexpression screen layout and process.

(A-B) Layout of the colonies arrayed in a 32x48 (row x column) format for the tFT library (A) and the overexpression libraries (B). (A) The tFT strains are laid out in triplicates in every 2x2 section and the same ORF is present six times in a 4x4 grid (coloured). Non-fluorescent strains are also placed in triplicates and present 6 times per 4x4 grid (light gray). Within each 4x4 grid, there are 4 positions that are filled with reference strains from the reference set (black). The set of reference strains is same across all plates. (B) Overexpression strains, strains carrying overexpression plasmid, are positioned in the top two rows for every 4x4 grid (coloured). The ORFs match in position to the tFT library layout. On the bottom two rows for every 4x4 grid are the control strains carrying the plasmid with the ORF YGR045C on the plasmid (light blue). (A & B) Border strains (dark gray) are dummy strains placed in the outermost 4 columns and rows in a 1536-array format. (C) The outline of how the overexpression screen is performed. The tFT library strain is crossed with the overexpression library strain carrying the corresponding ORF on the plasmid. The strains undergo SGA selection for the final haploid strain that contains both components. For the mORF overexpression screen, there is an extra step of induction on galactose media. Then, the colonies on the final step are measured using a plate reader to determine colony fluorescence.

CHAPTER 2: Mislocalizing proteins through overexpression

This chapter will lay out the results of the three different overexpression screens conducted (tFT-MoBY1, tFT-MoBY2, and tFT-mORF) and provide an insight on the proteins that are found to be attenuated upon overexpression. Detailed discussion on the properties of the attenuated proteins will be in Chapter 4.

2.1 MLPs that are attenuated upon low level overexpression (tFT-MoBY1)

This sub-section describes the results from the tFT-MoBY1 overexpression screen. In brief, the MoBY1 overexpression library was crossed with the tFT library carrying the mCH-mNG tFT tag on each ORF. Haploid strains containing both the tFT and OE plasmid were selected using the SGA method. After which, the haploids were then measured on a plate reader to obtain colony fluorescence measurements and imaged for segmentation to measure colony size..

The screen showed that only about 4.5% of proteins are significantly attenuated upon overexpression ($\geq 20\%$ reduction, $p < 0.1$) (Figure 10A). The PCN for tFT-MoBY1 strains were determined by qPCR analysis to range between 1-3 copies with an average of 1.6 copies (Figure 10B). Through this set of results, I can already make a note on proteins that are extremely sensitive to overexpression and are probably prone to mislocalization at even such a low level of OE. Furthermore, even with the low level of overexpression, I could show that cytosolic ribosomal subunits, which are known to be attenuated upon overexpression (Sung, Reitsma, *et al.*, 2016), were found to be mildly attenuated within this screen (Figure 10C).

Additionally, I performed a GO term enrichment analysis on cellular component and noted that the attenuated proteins were enriched with subunits of protein complexes (Figure 10D). This enrichment remains even after excluding the ribosomal subunits from the dataset to avoid the large number of attenuated ribosomal proteins biasing the outcome. Further breakdown of the screen results show that many of the attenuated proteins in terms of mNG level did not have a significant reduction in ratio (Figure 10E, bottom row). As pointed out in Figure 8D, the ratio measurements from the mCH-mNG tFT is less sensitive for a proportion of proteins, and so, it is likely that the tFT is unable to determine the small changes in ratio at this level of overexpression.

Next, Dephoure and colleagues showed that about 15% of proteins on the duplicated chromosome in yeast aneuploid cells are attenuated when measuring protein abundance using tandem mass tag (TMT)-based mass spectrometry (hereafter referred to as TMT dataset) (Dephoure *et al.*,

2014). The overexpression screen done with MoBY1 plasmids should be similar in the level of overexpression to that of the duplicated chromosome condition (i.e. one extra copy), however, the level of attenuation I measured was much lower. Comparing the attenuated proteins from the TMT dataset on aneuploid yeast to my tFT-MoBY1 overexpression screen, I can see that there are some proteins that overlap but there are still many other proteins that are attenuated only in my individual overexpression method (Figure 10F). The difference in level of attenuation across the proteome and number of overlapping hits may be due to the likelihood that in an aneuploid condition, overexpression of the many different proteins on a single duplicated chromosome can lead to off-target effects or a more global response such as the environmental stress response (ESR) (Torres *et al.*, 2007; Zhu *et al.*, 2018). Thus, my results show a strong argument that my method of individual protein overexpression, provides a clearer picture of the response to individual proteins being in excess and mislocalizing.

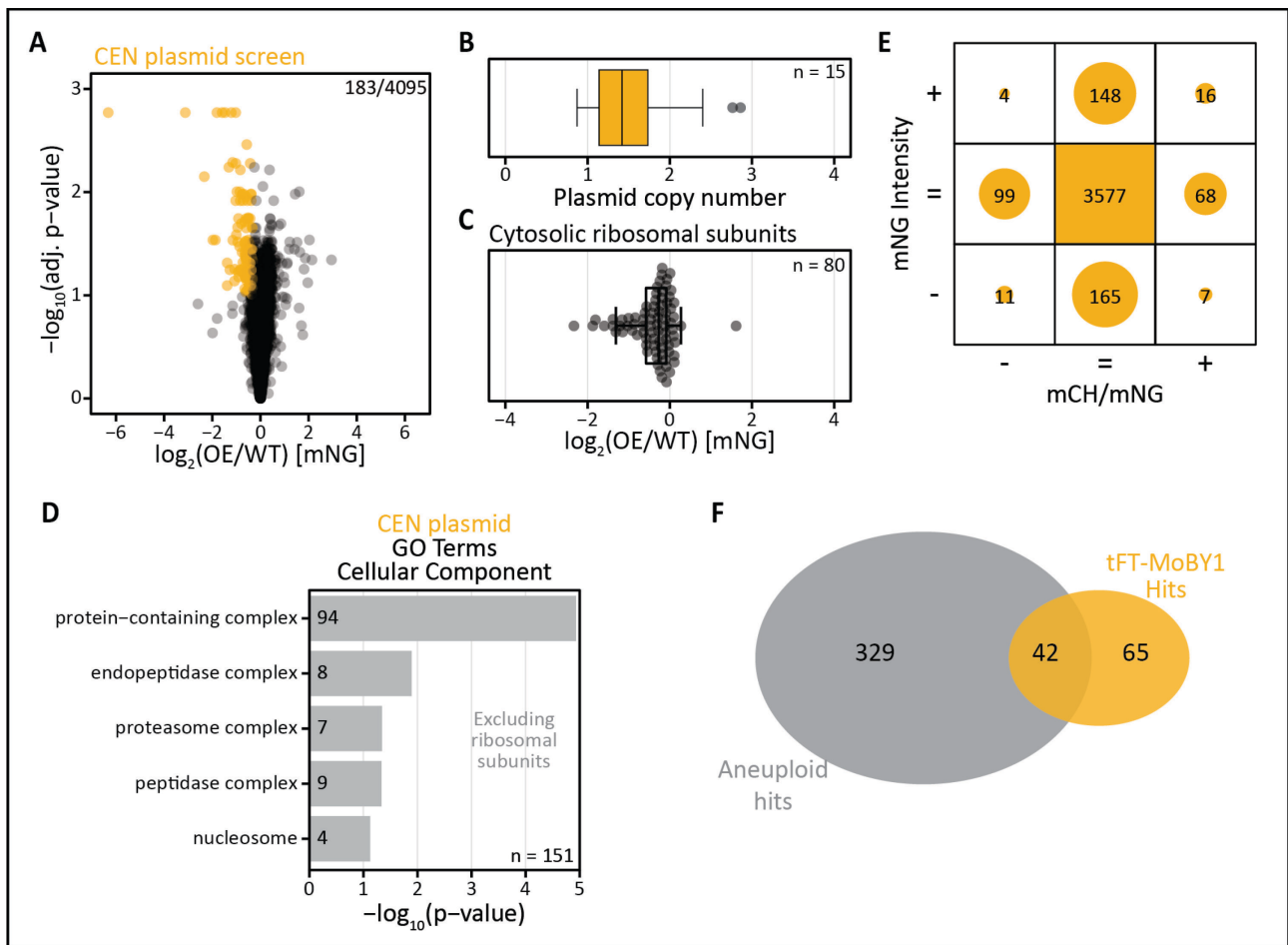


Figure 10: Low level overexpression with CEN plasmids lead to attenuation of a small proportion of proteins many of which are subunits of complexes.

(A) Volcano plot showing the \log_2 fold change of mNG levels upon overexpression (OE) compared against non-overexpression condition (WT). Significantly attenuated ORFs are highlighted in orange ($\geq 20\%$ decrease; $p < 0.1$). (B) Plasmid copy number of 15 strains from the screen as determined by qPCR. (C) \log_2 fold change of cytosolic ribosomal subunits within the screen. (D) GO term enrichment analysis results for cellular component after exclusion of ribosomal subunits. (E) Balloon plot describing the results of the overexpression screen with rows showing the significant increase (+), no change (=), and significant decrease (-) of mNG levels and columns showing respective changes in red-to-green ratio. Sizes of circles are proportional to number of ORFs in that category. (F) Venn diagram on the overlap of tFT-MoBY1 attenuated hits against the attenuated hits found in the tandem mass tag (TMT) dataset of aneuploid strains from Dephore *et al.* (2014).

2.2 MLPs that are attenuated upon stronger overexpression (tFT-MoBY2)

This sub-section describes the results from the tFT-MoBY2 overexpression screen. In brief, the MoBY2 overexpression library was crossed with the tFT library carrying the mCH-mNG tFT tag on each ORF. Haploid strains containing both the tFT and OE plasmid were selected using the SGA method. After which, the haploids were then measured on a plate reader to obtain colony fluorescence measurements and imaged for segmentation to measure colony size..

Here, with a higher level of overexpression from the MoBY2's 2 μ plasmid, I found that ~14.6% of proteins (566/3889) tested were significantly attenuated ($\geq 20\%$ reduction, $p < 0.1$) (Figure 11A). The PCN for tFT-MoBY2 strains were determined by qPCR analysis to range between 1-7 copies with an average of 4.8 copies (Figure 11B). It is likely that the KanMX marker present on the overexpression plasmid led to a lower level of PCN than the expected ~20 copies, a behaviour which has been noted in literature (Karim, Curran and Alper, 2013). I further quantified the expression level of the ORFs by performing RT-qPCR on RNA obtained from the same strains used for PCN quantification. The fold increase in gene expression levels correlate nicely with the gene copy number, which is the quantified PCN + 1 (Figure 11C). The RT-qPCR results also show that the overexpression using the 2 μ plasmid with endogenous promoter and terminator leads to a relative overexpression based on PCN rather than an absolute overexpression, i.e. with a heterologous GAL promoter and terminator.

It is also important to highlight that with the stronger level of overexpression, there also comes a mild cost on cell fitness (Figure 11D). The scatter plot in Figure 11D shows that about 4.8% of proteins screened had a significant reduction in cell size upon overexpression with the stronger 2 μ plasmid ($\geq 20\%$ reduction, $p < 0.1$). From this 186 proteins, 97 are proteins that were also attenuated out of the 566 proteins. It is unlikely that the cell size effect will affect too much the analysis of the results from this screen. Nonetheless, it is still important to keep this observation in mind.

As a first check, I showed that the cytosolic ribosomal subunits were also attenuated within this screen as expected from their behaviour described in literature (Figure 11E). Additionally, the level of attenuation of the attenuated cytosolic ribosomal subunits is stronger with a higher level of overexpression than in the tFT-MoBY1 screen, with the median log₂ fold change for tFT-MoBY2 at -1.14 against -0.72 for tFT-MoBY1. The data from this screen was also able to recapitulate findings in literature of proteins that are attenuated upon overexpression (Figure 11F, red) (See Supplementary Table 7 for list of curated ORFs with known dosage compensation behaviour upon overexpression). The proteins that have been tested in literature to not be attenuated upon overexpression is also not

attenuated within my screen results (Figure 11F, white). Taken together, the similar attenuation behaviour of proteins from literature and within this screen results together with the attenuation behaviour of the cytosolic ribosomal proteins, I am confident on the reliability of the outcomes of the screen.

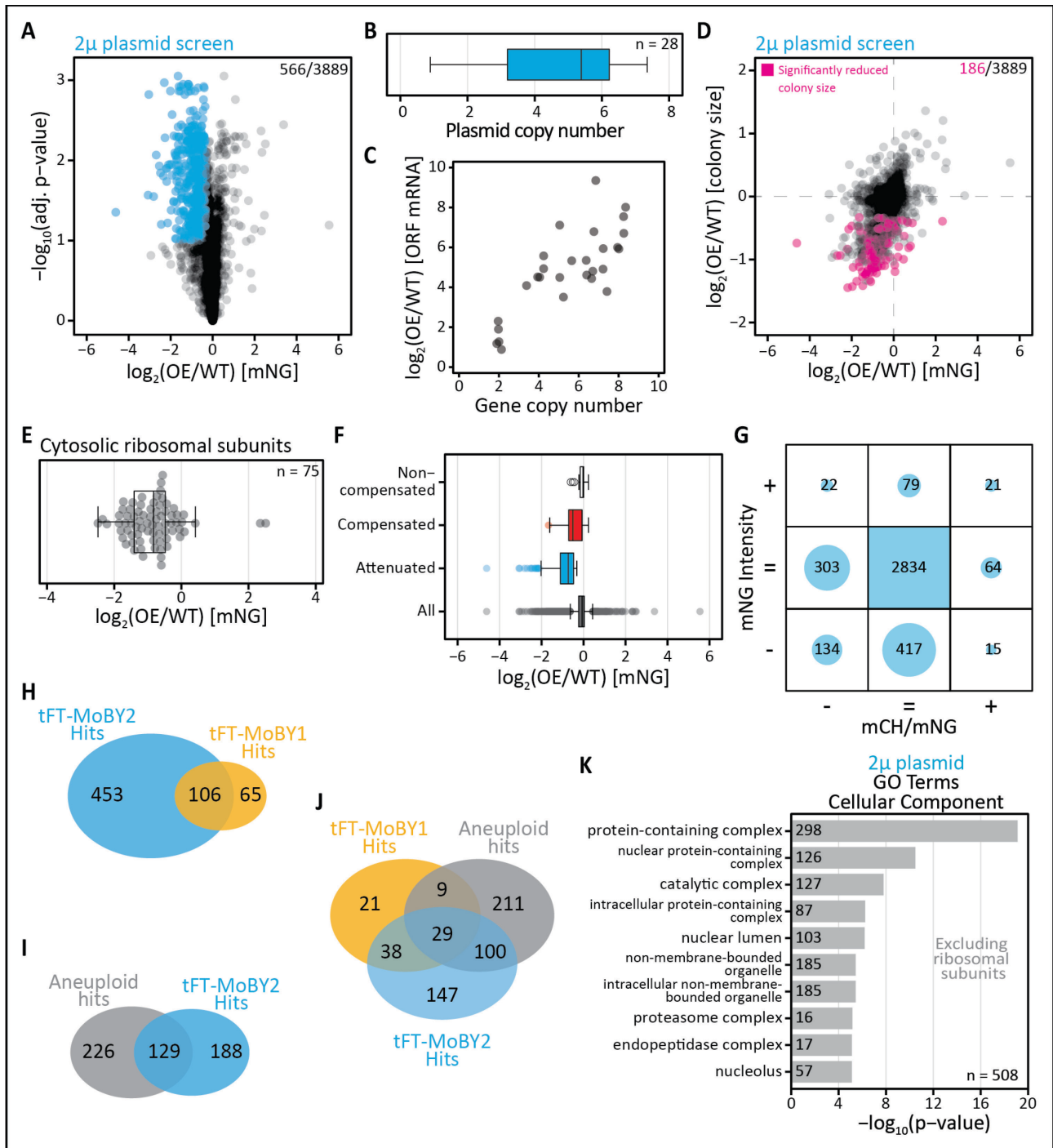


Figure 11: Stronger overexpression with 2 μ plasmids lead to attenuation of a larger proportion of proteins many of which are subunits of complexes and can recapitulate literature findings.

(Legend continues on next page)

(A) Volcano plot showing the log₂ fold change of mNG levels upon overexpression (OE) compared against non-overexpression condition (WT). Significantly attenuated ORFs are highlighted in blue ($\geq 20\%$ decrease; $p < 0.1$). (B) Plasmid copy number of 28 strains from the screen as determined by qPCR. (C) Log₂ fold change of ORF mRNA levels quantified via RT-qPCR against gene copy number quantified via qPCR (plasmid copy number + 1). (D) Comparison of log₂ fold change in colony size against the log₂ fold change in mNG levels upon overexpression. Strains with a significant decrease in colony size is highlighted in pink ($\geq 20\%$ decrease; $p < 0.1$). (E) Log₂ fold change of cytosolic ribosomal subunits within the screen. (F) Boxplots showing ORFs from literature that have been shown to be dosage compensated (red) or otherwise (white) and their behaviour within the overexpression screen, and the distribution of the attenuated hits (blue) and all ORFs tested within the screen (gray). (G) Balloon plot describing the results of the overexpression screen with rows showing the significant increase (+), no change (=), and significant decrease (-) of mNG levels and columns showing respective changes in red-to-green ratio. Sizes of circles are proportional to number of ORFs in that category. (H) Venn diagram on the overlap of tFT-MoBY2 and tFT-MoBY1 attenuated hits. (I) Venn diagram on the overlap of tFT-MoBY2 attenuated hits against the attenuated hits found in the TMT dataset of aneuploid strains from Dephoure *et al.* (2014). (J) Venn diagram on the overlap of attenuated hits of tFT-MoBY2 and tFT-MoBY1 screens, and the attenuated hits from the TMT dataset of aneuploid strains from Dephoure *et al.* (2014). (K) GO term enrichment analysis results for cellular component after exclusion of ribosomal subunits.

The breakdown of the screen results show that many of the attenuated proteins in terms of mNG level still did not have a significant reduction in ratio (Figure 11G). However, the proportion of proteins that are attenuated in mNG level and ratio has increased by about 4.3-fold (5.46% in tFT-MoBY1 vs 23.67% in tFT-MoBY2). The results in changes in ratio may not reflect the full picture on all changes in protein turnover, but it has already shown that with stronger overexpression, there are more proteins that are attenuated and also destabilized strong enough for the tFT to show a significant difference in ratio. On the other hand, there are also a large proportion of proteins that are significantly attenuated in terms of ratio but are not attenuated in terms of mNG levels. Could there be a separate mode of attenuation by the cell when it comes to handling MLPs through protein overexpression? I further analyzed the attenuated proteins in terms of either ratio (stability) or mNG (abundance), or both and discuss this further in Chapter 4 Section 4.2.

The attenuated proteins from the tFT-MOBY2 screen overlaps with that of the tFT-MoBY1 screen with many more hits found in the latter (Figure 11H). However, there are a handful of hits that were identified in the tFT-MoBY1 screen but not here. I performed a GO term enrichment analysis on these 65 proteins and found no enrichment, and then I further checked that the lack of attenuation in tFT-MoBY2 screen was not due to colony size differences. I can only speculate that maybe with the lower level of overexpression, the cell opted for attenuating these proteins to keep it in check, but with the higher level of overexpression, the cell attempts to upregulate other factors to handle the

proteins but not remove it. I also noted that these 65 proteins were relatively low in abundance and so any changes would be amplified and seem significant.

Besides that, while the percentage of attenuated proteins from the tFT-MoBY2 screen is similar to that from Dephoure and colleagues' results (Dephoure *et al.*, 2014), the attenuated proteins from this screen still do not overlap well with their aneuploid data (Figure 11I). Making the comparison between the tFT-MoBY1, tFT-MoBY2, and aneuploid TMT data, I can clearly see that the attenuated hits from the MoBY overexpression does not overlap well with the aneuploid data (Figure 11J). Here again, I would emphasize that my method provides a cleaner readout from individual protein's effect rather than the effect from a whole chromosome worth of overexpression.

While it is as expected that there is higher number of proteins that are attenuated with the stronger overexpression, however, it is even more interesting that the cell does not "care" about its other 75% protein's levels being more than normal. What differentiates the attenuated proteins from the non-attenuated proteins? To further understand the properties of these attenuated proteins, I performed a GO term enrichment analysis on cellular component and noted, once again, that the attenuated hits are enriched with subunits of protein complexes (Figure 11K). The enrichment for subunits of protein complexes is much stronger than in the tFT-MoBY1 screen. This enrichment brings up the question on what makes having excess protein subunits makes the subunit itself more susceptible to attenuation upon overexpression.

I will discuss in more detail on the properties of the attenuated proteins in Chapter 4 Section 4.1, and also expand on several aspects of it in Chapter 5 and Chapter 6. I also conducted further analysis and experiments on this trend of enriched attenuated subunits and discuss it in Chapter 5.

2.3 Very strong overexpression leads to growth defects and attenuated MLPs (tFT-mORF)

This last sub-section describes the results from the tFT-mORF overexpression screen. In brief, the mORF overexpression library was crossed with the tFT library carrying the mCH-mNG tFT tag on each ORF. Haploid strains containing both the tFT and OE plasmid were selected using the SGA method. After which, the haploids were transferred onto a galactose media to induce overexpression and then measured on a plate reader to obtain colony fluorescence measurements and imaged for segmentation to measure colony size.

The overexpression screen performed with the mORF library led to strong growth defects in the strains upon induction of overexpression. I found 1864/4000 (46.6%) ORFs were significantly attenuated upon overexpression ($\geq 20\%$ reduction, $p < 0.1$) (Figure 12A). However, a large number of these proteins (1726/4000) have a strong significant decrease in colony size ($\geq 20\%$ reduction, $p < 0.1$) upon overexpression (Figure 12B) with 1113 significantly attenuated ORFs affected.

Due to this issue, I decided to apply a filter on the change in colony size to remove ORFs that had a strong change in colony size and proceed to analyze the results to see what I can determine. This threshold was set at 20% with p -value < 0.1 , whereby any ORFs with more than 20% decrease in size and p -value < 0.1 are filtered off. With the filtering in place, the data shows that 751/2277 (33%) ORFs were attenuated upon overexpression and these ORFs were used for further analysis (Figure 12C).

Performing the quality control checks, I still found that ribosomal subunits were attenuated as expected and also proteins known to be attenuated upon overexpression from literature also behaves similarly within this screen (Figure 12D & 12E). Most importantly, the proteins that are found to not be attenuated upon overexpression remains not attenuated in this dataset as well (Figure 12E). The results of these checks indicate that the filtered data is still usable and reliable.

Viewing the distribution of proteins across changes in mNG level and changes in ratio, I see that a large number of proteins have a decrease in mNG levels but no change in the red-to-green ratio (Figure 12F). The proportion of proteins that are attenuated in mNG levels and ratio compared to all proteins that are attenuated (9.85%) is lower than in the tFT-MoBY2 screen but higher than in tFT-MoBY1 screen. Since a large proportion of proteins are only now beginning to be attenuated (much less negative \log_2 fold change in mNG for tFT-mORF only hits in Figure 12I), it could be that the change in ratio is again not large enough for the tFT to detect.

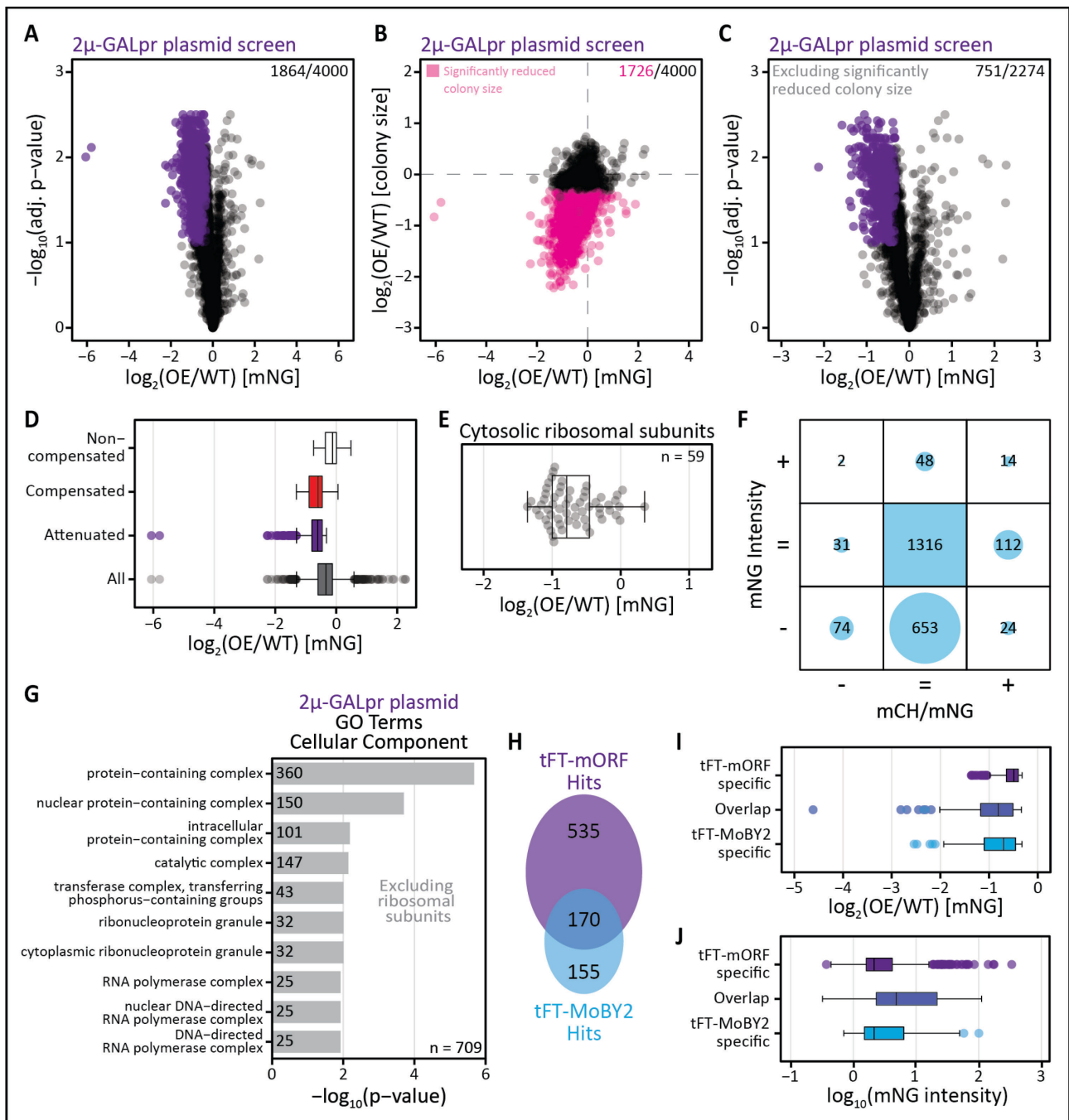


Figure 12: Very strong overexpression with 2 μ plasmids and GAL1 promoter lead to widespread growth defects but filtered data is still usable.

(A) Volcano plot showing the log₂ fold change of mNG levels upon overexpression (OE) compared against non-overexpression condition (WT). Significantly attenuated ORFs are highlighted in purple ($\geq 20\%$ decrease; $p < 0.1$). (B) Comparison of log₂ fold change in colony size against the log₂ fold change in mNG levels upon overexpression. ORFs with a significant reduction in colony size is highlighted in pink ($\geq 20\%$ decrease; $p < 0.1$). (C) Same as in (A) but after removal of ORFs that have a significantly reduced colony size. (D) Boxplots showing ORFs from literature that have been shown to be dosage compensated (red) or otherwise (white) and their behaviour within the overexpression screen, and the distribution of the attenuated hits (blue) and all ORFs tested within the screen (gray). (E) Log₂ fold change of cytosolic ribosomal subunits within the screen.

(Legend continues on next page)

(F) Balloon plot describing the results of the overexpression screen with rows showing the significant increase (+), no change (=), and significant decrease (-) of mNG levels and columns showing respective changes in red-to-green ratio. Size of circles are proportional to number of ORFs in that category. (G) GO term enrichment analysis results for cellular component after exclusion of ribosomal subunits. (H) Venn diagram on the overlap of tFT-mORF and tFT-MoBY2 attenuated hits. (I) Comparison of \log_2 fold change of mNG levels between attenuated hits that occur in both screens or the screen specific hits as per (H). The \log_2 fold change data for tFT-mORF specific category is from tFT-mORF screen and data for the tFT-MoBY2 specific and overlap categories are from the tFT-MoBY2 screen. (J) Comparison of \log_{10} mNG levels between attenuated hits that occur in both screens or the screen specific hits as per (H). The \log_{10} mNG data is from tFT-mORF screen.

Next, performing a GO term enrichment analysis on the tFT-mORF hits showed enrichment for subunits of protein complexes in agreement with the other two screens (Figure 12G). I also find GO terms on ribonucleoprotein granule and RNA polymerase complex appearing.

When looking at the overlap of attenuated hits between tFT-mORF and tFT-MoBY2 screens, I can see that the tFT-mORF screen with a much stronger overexpression leads to more hits being identified (Figure 12H). However, the level of mNG attenuation is not as strong for the tFT-mORF screen specific hits (Figure 12I). For the set of tFT-mORF screen specific hits, it could be likely that these proteins are much more tolerant to overexpression and are only now crossing the limit of the QC systems that try to maintain them and so turns to the QC systems that will remove them. In support of this argument, I can see that the tFT-mORF specific hits are mostly low abundant and relatively much lower than the hits coming from the overlap of both screens, so the overexpression from the MoBY2 plasmids might not be enough (Figure 12J). However, it is unclear as to why the tFT-MoBY2 specific hits are not attenuated in this screen since they are also similar in being low abundant as the tFT-mORF specific hits.

Overall, while the data from this screen is still usable, a large portion of the proteome is already filtered off. Thus, I would not be able to paint a full picture on the properties of attenuated proteins by excluding so many proteins. Moving forward, further analysis and discussions will be based mostly on the data from tFT-MoBY1 and tFT-MoBY2 screens.

CHAPTER 3: Sequencing of MoBY plasmid libraries

In the process of validating the screen results and performing checks on the overexpression plasmids. I found that a handful of the MoBY1 and MoBY2 plasmids had point mutations within them. Another PhD student in the lab noted the same for a handful of MoBY2 plasmids she was using for her experiments.

Based on the publication on the construction of the MoBY1 plasmid library, the plasmids were sequenced to confirm the plasmid identities (Ho *et al.*, 2009). The sequencing was done only from the promoter of the KanMX module into the 3' ORF junction, and from the section after the KanMX module into KanMX, which means that the entire ORF was not sequenced for all plasmids. The full ORFs were also not sequenced for the MoBY2 library plasmids (Ho, 2011). Since the plasmid libraries were never sequenced fully for identifying errors, I decided to sequence the MoBY overexpression libraries to determine the extent of mutations within the plasmids and use the information to further clean up the results, if necessary.

The MoBY1 library was already present as an *E. coli* library and could be used directly to mini-prepare the plasmids and send for sequencing. The MoBY2 library at hand exists as a yeast library and would have been slightly more difficult to process. The plasmids would have to be prepped out of yeast (which has a low yield), and then transformed into *E. coli*, and be mini-prepped again for enough material for sequencing. So here, I am really thankful to the Boone Lab for providing an *E. coli* version of the MoBY2 library which I can then directly use for library sequencing.

In brief, the two libraries were grown, pooled, mini-prepped, and submitted to the IMB Genomics core facility to perform next generation sequencing. I performed the miniprep for the MoBY2 plasmid library while [REDACTED] performed the miniprep for the MoBY1 plasmid library. The raw data obtained from sequencing was then processed first by the IMB Bioinformatics core facility and the results from variant calling were provided. Using the sequencing and variant calling results provided, I then performed my own further processing, checks, and detailed analysis.

The results are described here in this chapter.

3.1 Data processing and clean up

The resulting single nucleotide polymorphisms (SNPs) from sequencing the MoBY1 and MoBY2 plasmid libraries show a large clear overlap (Figure 13A). Both plasmid libraries also had additional SNPs that were not found in the other. It is possible that mutations were gained or lost in the process of generating the MoBY2 plasmid from MoBY1. Besides that, it is also possible that the SNPs may have not passed the various quality control checks and filter threshold criteria set when the raw sequencing data was processed by the Bioinformatics CF.

Before looking deeper, I first highlight an important detail on how the MoBY plasmids were constructed and the effect this has on interpreting the results. The ORF in the MoBY plasmids were amplified using primer sequences that anneal to ~900bp upstream of the start codon and ~250bp downstream from the stop codon (Figure 13B) (Ho *et al.*, 2009). Therefore, when assembling the sequencing results, there are sequences that will inevitably come from two or more plasmids, termed overlap region (Figure 13C(i)). The examples given in Figure 13C(i) shows that the regions highlighted in blue have sequencing results originating from the plasmid carrying CCR4 and FUN26 to the left (left box), or from three different plasmids carrying genes CCR4, ATS1, or FUN30 (right box). The average read depth of these overlap regions do correlate nicely with the sum of average read depth of the regions flanking it, provided the neighbouring region is not an overlap region as well (Figure 13D). However, an issue that arises from the presence of overlapping regions is that SNPs within the region can belong to two or more plasmids. This is especially important if the allele frequency (AF) of the SNP is not near 100%, and so the SNP may only belong to one plasmid and not the other.

Starting with the more straightforward results, I first look at the SNPs that are in the non-overlap regions that belongs to only one plasmid. The majority of these SNPs are at an AF of 0.9 and higher, which I set as the threshold to call a clean SNP (i.e. clear nucleotide change with no multiple alternatives) (Figure 13E). However, in a non-overlap region there are also SNPs that have an AF less than 0.9 (termed as mixed SNPs because not all reads have the same base), which most likely means that there are mixed plasmid populations (Figure 13C(iia)). The mixed plasmid could be coming from a mixture of plasmids co-existing within a cell and all cells have the two versions, or that there are two separate groups of cells in the same well from the library and each group carries a different version. Based on tests done by streaking the culture from a well, isolating individual colonies, and sequencing the plasmids, a clean colony with only one version of the plasmid can be obtained. The possibility to isolate strains with a single version of the plasmid leans more on that there are two separate groups of cells in the same well from the library and each group carries a different version.

For mixed SNPs in the non-overlapping region, it is clear to call that plasmid as mixed since there are two or more versions of the plasmid with the same ORF but different SNPs at the same library position.

Next, I look into the results from the overlap regions and from the AF plot for the overlap regions, it is clear that mixed SNPs are the majority in this region (Figure 13F). Using the same threshold as before, if a SNP is present at an AF above 0.9, it is defined as being present in all plasmids that are within the overlap. However, for the mixed SNPs (AF<0.9) present in the overlap regions, it is not as direct since the SNP may also only be present in the plasmid for one ORF but not for the other. To try and clear up the mixed SNP in overlap regions, I used the ratio of average read depth of the next or previous region to the sum of the neighbouring regions and compared it to the AF of the mixed SNP. If the AF matches either ratio, then it is very likely that the SNP is from the neighbouring region (either previous or next). From this analysis, I only re-assigned the SNPs to neighbouring regions if the AF is not 0.5 \pm 0.05 and matches very closely to the calculated ratios (AF – ratio = 0 \pm 0.25SD). I then manually checked on some of these re-assigned SNPs and found that the method was sound. Unfortunately, this method of calculation and comparison is unable to allow re-assignment when the neighbouring region is also an overlapping region since the average read depth of a region is not close to the sum of neighbouring regions, and so the ratio is inaccurate (Figure 13D, red data points). For the remainder mixed SNPs that were not checked, I assumed that the SNP is present in all plasmids involved in the overlap region. Fortunately, the mixed SNP would only be inside the ORF for one plasmid and upstream or downstream of the ORF of the other plasmids involved, in most cases.

Besides looking into the SNPs in general, I also looked specifically into insertions and deletions (INDELs) within the promoter and terminator regions that are maybe large enough to cause an issue in the protein's transcription. These INDELs were manually checked, and removed if found to be wrongly annotated. Additionally, from the variant annotation data on the type of mutation that a SNP causes, I made a subset of the data annotated with critical mutations (i.e. stop gained, stop lost, start lost, and frameshift) and manually checked to confirm the annotation. These critical mutations are termed so because the mutations would render the protein non-functional or incomplete.

Overall, I found that about 50% (2229/4480) of the plasmids in the MoBY2 library had some kind of mutation, but the highly disruptive critical mutations were not very common (Figure 13G). The most common type of mutation found was single codon changes (missense variants) that could not be clearly determined if it would have affected the protein as a whole.

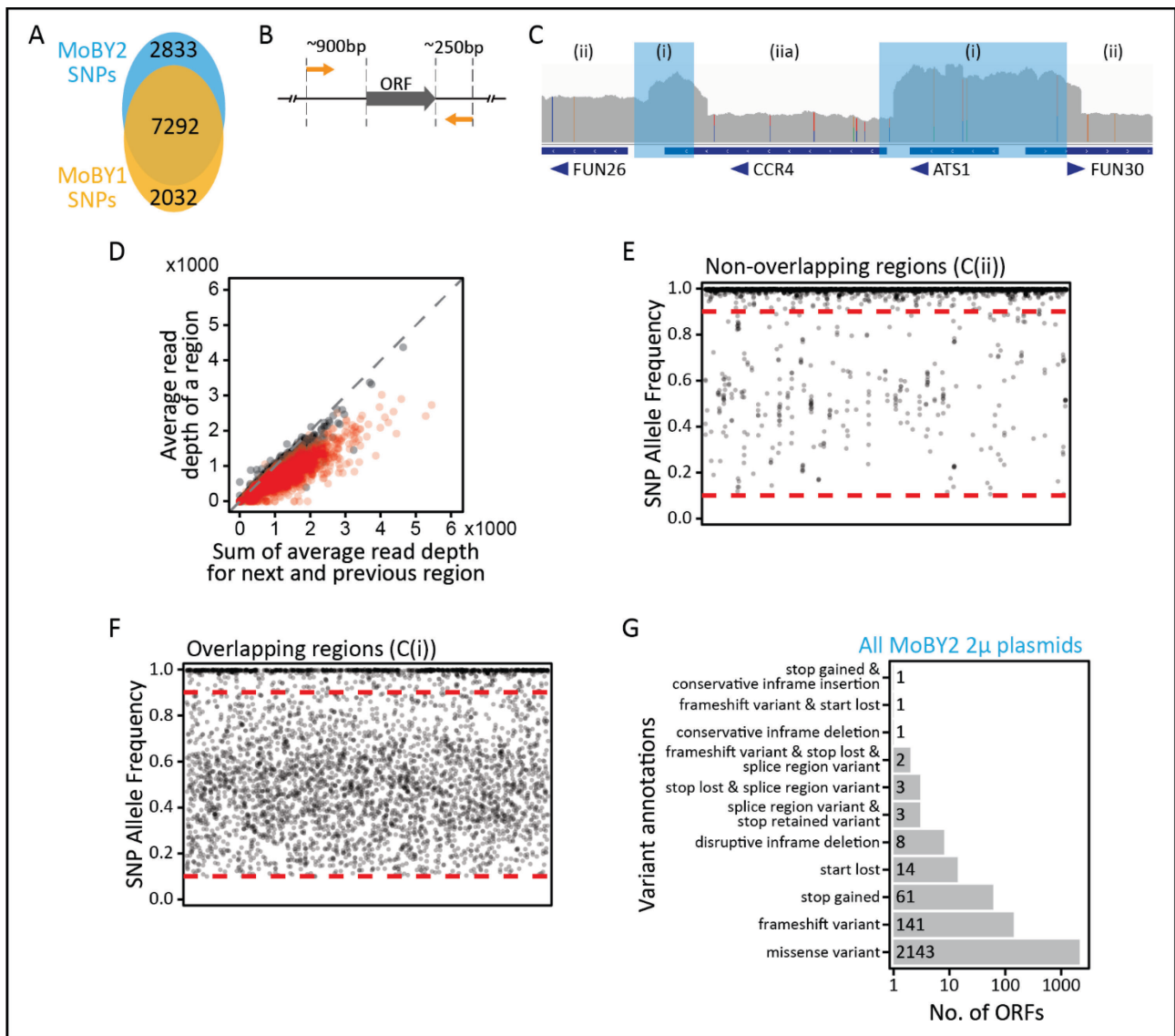


Figure 13: NGS sequencing outcomes of MoBY1 and MoBY2 plasmid libraries.

(A) Venn diagram of overlap between SNPs called in the MoBY1 and MoBY2 plasmid libraries. (B) Diagram on how the ORF construct was amplified to be used and inserted into the overexpression plasmids. (C) Screenshot from the genome browser of a section of the sequencing results. Sections highlighted in blue are coming from overlapping sequences of plasmids that have adjacent ORFs (i) and non-overlapping regions are not highlighted (ii and iia). SNPs that are called by the algorithm can show >90% allele frequency (clear) (ii) or <90% allele frequency (mixed) (iia). (D) Correlation between the average read depth of a region that is an overlap (C(i)) and the sum of average read depth of the next and previous regions (adjacent regions). Adjacent regions that are also overlap regions are highlighted in red. (E) Allele frequency distribution of SNPs within non-overlapping regions (C(ii)) arranged on the x-axis by chromosome. (F) Allele frequency distribution of SNPs within overlapping regions (C(i)) arranged on the x-axis by chromosome. (G) Number of ORFs with the annotated variant in the whole MoBY2 plasmid library. One ORF may have multiple SNPs that fall into different variant annotations.

3.2 Mutations in MoBY2 plasmids unlikely to have impacted results of overexpression screen

Next, sub-setting the finalized SNP data to only ORFs that were within the tFT-MoBY2 screen, I noted that 1951 ORFs had some kind of mutation out of the 3889 ORFs and the majority are missense variants (Figure 14). I then performed a Fisher's test to check if there is an enrichment of ORFs with mutations within the attenuated hits and found there to be none ($p = 0.8558$; odds = 1.02). Additionally, I also checked if critical mutations were enriched and found none ($p = 0.05245$; odds = 0.62).

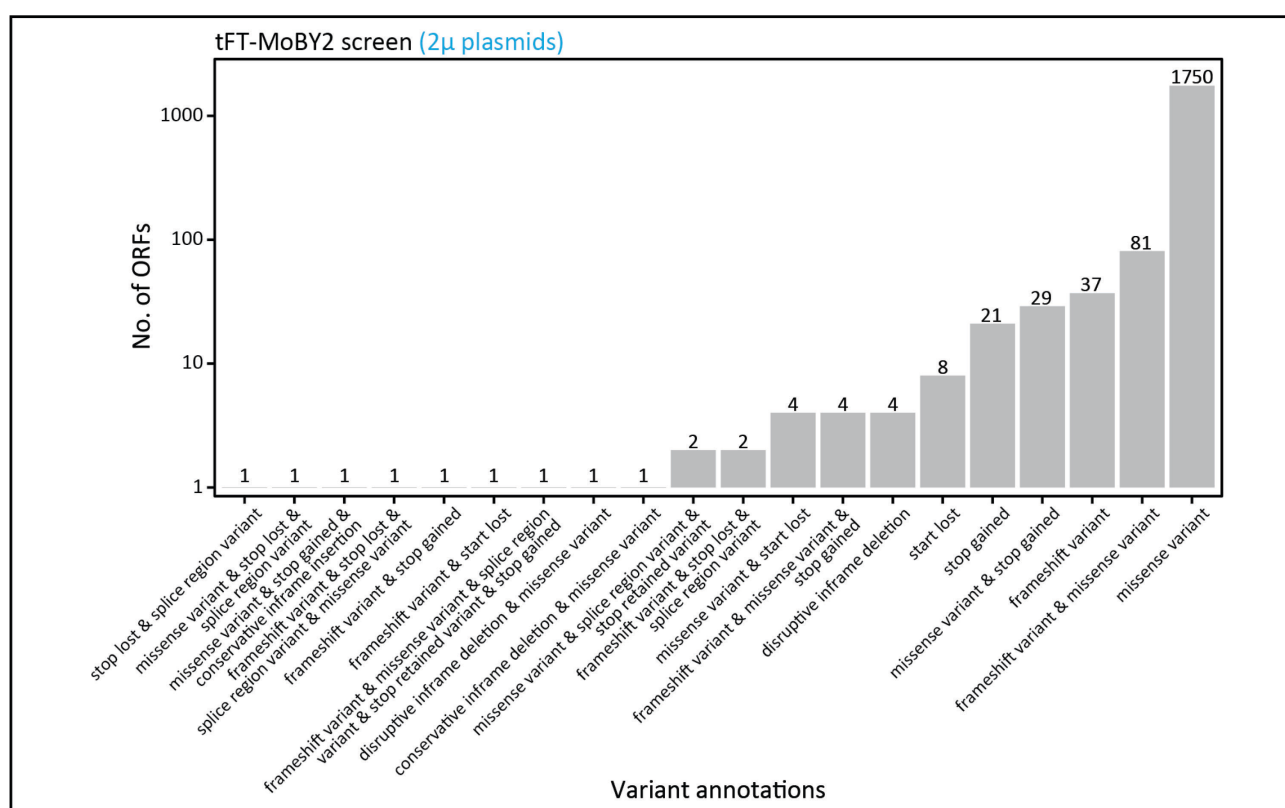


Figure 14: Variant annotations of SNPs in MoBY2 plasmids used in tFT-MoBY2 screen

Bar chart of number of ORFs carrying the SNP with the annotated variant categories for all MoBY2 plasmids used in the tFT-MoBY2 screen. Each ORF belongs to only one annotation category (only one bar) so there are no duplicate ORFs in the count. An ORF may have several SNPs that have different variant annotations, and so, these are summarized into one category.

The outcome of the sequencing results provide a useful database on mutations found in the MoBY libraries for future reference when using these libraries. The results also conclude that mutations on the overexpression plasmids likely do not affect the outcome of the tFT-MoBY screens performed.

CHAPTER 4: Properties of attenuated proteins

In this chapter, I take a deeper look into the attenuated proteins identified from the overexpression screens with a special focus on the tFT-MoBY2 screen hits. I provide an overview of the properties that define the attenuated proteins and highlight two avenues of interest that I looked into further in this thesis. I also breakdown the attenuated proteins into several categories based on their mNG and ratio attenuation behaviour to describe the different ways the cell may deal with protein overexpression.

4.1 Properties of all proteins with attenuated abundance

Across all three overexpression screens, I have found that the attenuated proteins are highly enriched with subunits of protein complexes. This enrichment for complex subunits among attenuated proteins have been seen in aneuploid conditions with an extra chromosome (Dephoure *et al.*, 2014). In the case of my screens, I first wanted to understand if the enrichment may be from just several large complexes with many subunits. Based on Figure 15A, I can clearly see that attenuated subunits come from a variety of complexes of different sizes. Logically, larger complexes that have more number of subunits, will be more likely to have at least one attenuated subunit (Figure 15B). Besides that, I also find that not all subunits within the same complex are similarly attenuated (Figure 15C). What differentiates these protein subunits within the same complex from being attenuated upon mislocalization? This is the first avenue of interest that I will study further and discuss the outcomes in Chapter 5.

Next, I found that the attenuated proteins are enriched with essential proteins (Fisher's test: $p = 4.136e-16$, odds = 2.39). Could this mean that the cell prefers to maintain these essential proteins at a specific level to avoid upsetting the pathways and processes it deems important?

Additionally, the attenuated proteins also have a significantly longer half-life (Figure 16A). Unsurprisingly, by increasing the stringency on calling attenuated hits (i.e. proteins that are more strongly attenuated), the half-life of the more stringent hits are even longer (Figure 16A). Besides half-life, I also found that the attenuated proteins generally have a significantly higher synthesis rate and the trend remains even with removal of cytosolic ribosomal subunits (Figure 16B). Removing the cytosolic ribosomal proteins from the synthesis rate analysis was important as many of the subunits fall within the attenuated category and majority have a very high synthesis rate. The observations on synthesis rates and protein half-life adds to the argument that this group of overexpressed proteins are attenuated because the cell needs to maintain these proteins at a fixed level

since it is produced rapidly and remains longer within the cell; any perturbations to the protein level from overexpression would lead then force the cell to correct it.

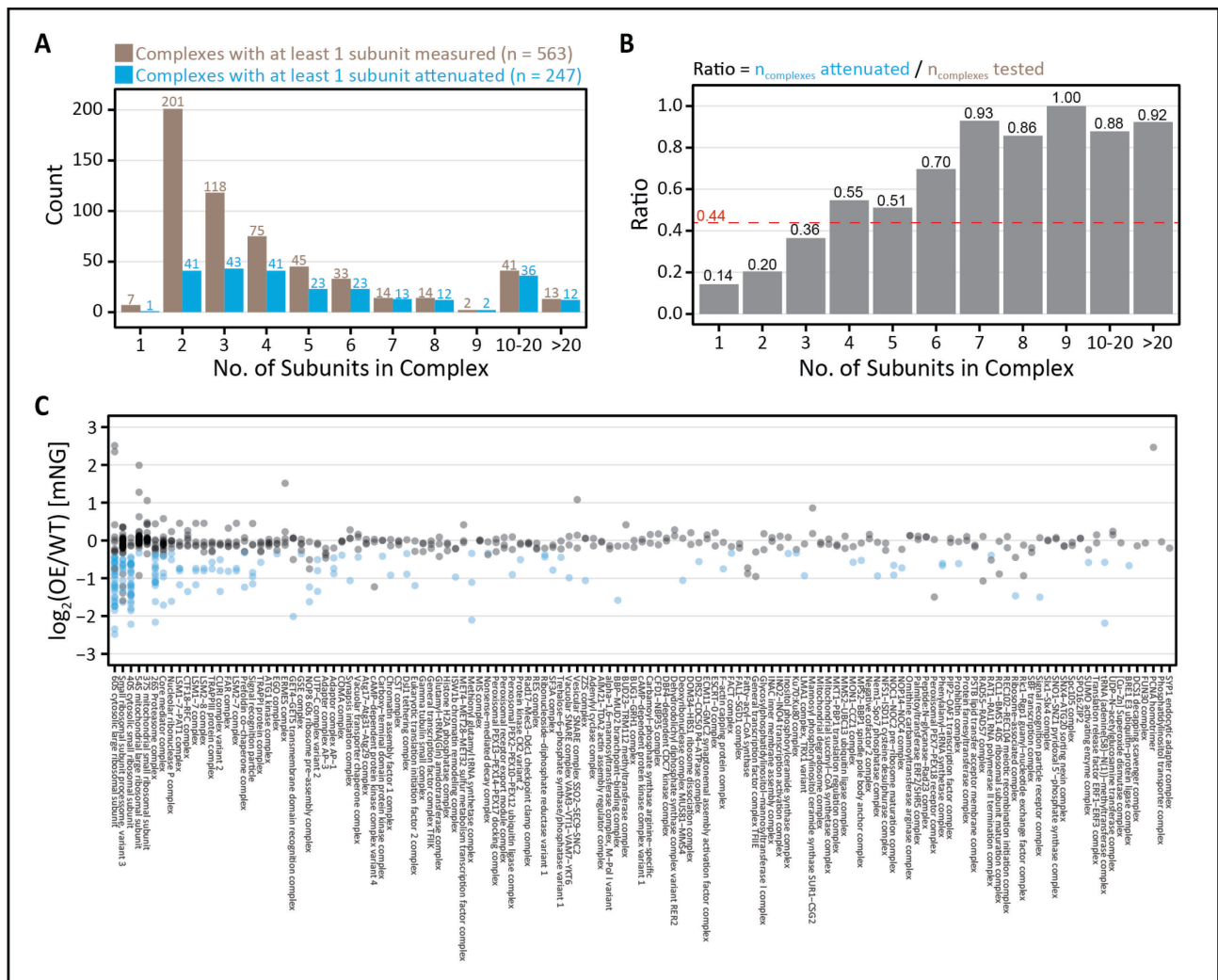


Figure 15: Attenuated subunits from the tFT-MoBY2 overexpression screen are from various complexes of different sizes but not all subunits within a complex are attenuated.

(A) Number of complexes with at least 1 subunit tested (gray) and at least 1 subunit attenuated (blue) categorized by size of complex based on number of individual subunits (Stoichiometry is not accounted). (B) Ratio of complexes with attenuated subunits to complexes tested categorized by size of complex based on number of individual subunits (Stoichiometry is not accounted). Red dotted line represents the ratio of the total number of complexes with attenuated subunits to total number of complexes tested. (C) Behaviour of different subunits within each complex. Attenuated hits are highlighted in blue. Complexes shown are limited to only complexes where all subunits were detected or those larger than 15 subunits. Subunit annotation data is obtained from the EBI Complex portal and Meldal *et al.* (2021).

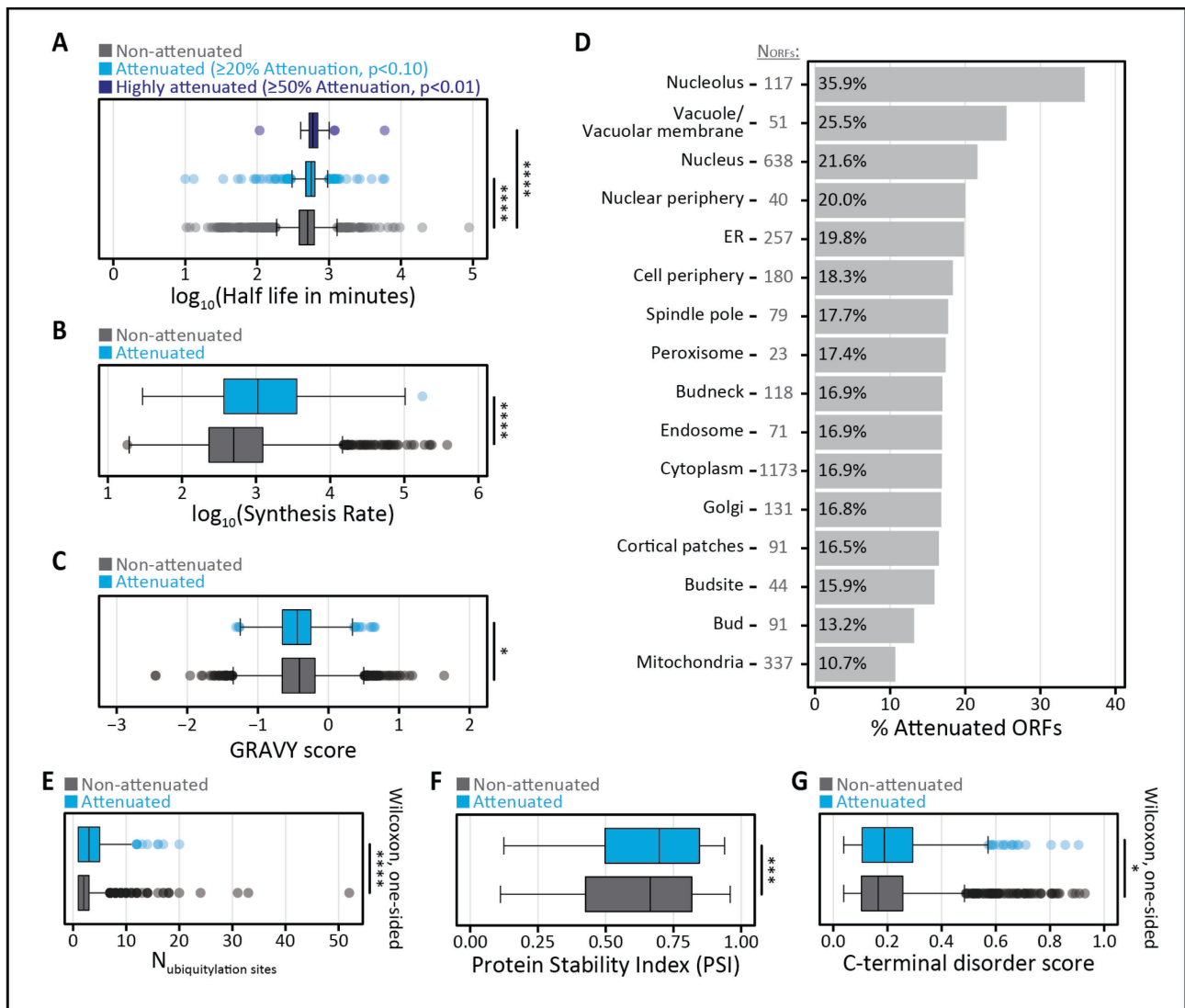


Figure 16: Properties of attenuated hits from the tFT-MoBY2 overexpression screen.

(A) Comparison of \log_{10} protein half-life in minutes of attenuated hits at different stringency thresholds (blue: $\geq 20\%$ decrease, $p < 0.1$; purple: $\geq 50\%$ decrease, $p < 0.01$) and non-attenuated hits (gray). Protein half-life data obtained from Christiano *et al.* (2014). (B) Comparison of \log_{10} synthesis rate of attenuated hits against non-attenuated hits within the screen. Synthesis rate data obtained from Taggart and Li (2018). (C) Comparison of GRAVY (Grand Average of Hydrophathy) scores of attenuated hits against non-attenuated hits. (D) Percentage of attenuated ORFs at different cellular localizations. Number of ORFs per localization is noted in gray on the y-axis. Localization data is obtained from Chong *et al.* (2015). (E) Comparison of number of ubiquitylation sites on a protein between attenuated hits and non-attenuated ORFs. Ubiquitylation site data is obtained from Swaney *et al.* (2013). (F) Comparison of protein stability index (PSI) of wild-type yeast N-termini between attenuated hits and non-attenuated hits. PSI data is obtained from Kats *et al.* (2018). (G) Comparison of C-terminal disorder scores between attenuated and non-attenuated hits. Disorder scores data obtained from van der Lee *et al.* (2014). (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$)

I then turned to look at physicochemical properties that may give a hint on what differentiates the attenuated and non-attenuated proteins. Approximately 1/3 of the eukaryotic proteins are integral membrane proteins with hydrophobic transmembrane domains that will be inserted into a lipid bilayer (McDowell *et al.*, 2020). These hydrophobic domains when exposed in the cytosol can be targets for factors that steer the protein towards targeting or degradation (Rodrigo-Brenni, Gutierrez and Hegde, 2014). Therefore, I first looked at the GRAVY score of the attenuated proteins compared to the non-attenuated proteins to see if hydrophobicity would play a role in determining attenuation. I found that the attenuated proteins are significantly less hydrophobic ($p = 0.01569$, $t = -2.4208$) (Figure 16C). Increasing the stringency of calling attenuated hits led to a stronger trend where the GRAVY score of attenuated proteins is even less hydrophobic.

Seeing as the attenuated hits are less hydrophobic, I then wondered if proteins with transmembrane domains (i.e. membrane embedded proteins) are depleted from the attenuated hits since these transmembrane domains tend to usually be hydrophobic in nature. As expected, I found that the attenuated hits are depleted of proteins with predicted transmembrane domains (Fisher's Test: $p = 0.03569$, odds = 0.78). Interestingly, this depletion is stronger in the tFT-MoBY1 screen (Fisher's Test: $p = 0.01607$, odds = 0.600), likely because the level of overexpression is not as strong as in the tFT-MoBY2 screen and so does not exceed a threshold where the cell finds it to be a problem. I would argue here that since membranes play an important role to maintain different properties of the different sub-compartments within the cell, the cell may want to protect the membranes much more and so, there are more efficient quality control systems to handle excess proteins than other cellular compartments. These systems do not necessarily need to remove the proteins but just ensure that they are properly targeted, inserted, and functional, and thereby reducing the need to expend further energy for destroying and then re-synthesizing. However, there will be a limit for when the protein levels are present in such an excess that it is too overwhelming for the cell, then the cell attenuates the protein levels.

Consequently, I wondered if proteins at certain cellular localizations may have a higher tendency to be attenuated. If so, I would predict that proteins in membranous compartments would be depleted from attenuated proteins based on the trend on depletion of proteins with predicted transmembrane domains. Looking at protein localizations, I see that at a lower overexpression level (tFT-MoBY1), the attenuated hits from this screen are significantly depleted of proteins from mitochondria (Fisher's Test: $p = 0.0221$, odds = 0.501) and Golgi (Fisher's Test: $p = 0.024$, odds = 0.229), and while not significant, there is also a strong depletion of proteins from the endosomes (Fisher's Test: $p = 0.127$, odds = 0.227). This observation agrees with the trend of depleted proteins

with transmembrane domains. There is also a mild enrichment of proteins localized in the cytoplasm (Fisher's Test: $p = 0.0418$, odds = 1.35). Then, at a higher overexpression level (tFT-MoBY2), only the significant depletion of mitochondrial proteins remain (Fisher's Test: $p = 9.38e-5$, odds = 0.501) (Figure 16D). However, proteins localizing to the nucleolus (Fisher's Test: $p = 3.02e-6$, odds = 2.69) and the nucleus (Fisher's Test: $p = 0.00847$, odds = 1.36) are significantly enriched (Figure 16D). As far as the levels of overexpression caused by the MoBY plasmids, the mitochondria remained consistently depleted of attenuated proteins. The depletion trend is only no longer seen when looking at the tFT-mORF screen data. These results show that with increasing overexpression levels, the QC systems focused on targeting and insertion to membranes are gradually overwhelmed and the QC systems opt to degrade the excess proteins. Why is there such a strong and consistent depletion of mitochondrial proteins from the attenuated hits? This is the second avenue of interest that I will study further and discuss the outcomes in Chapter 6.

Another goal of the project is to determine the protein quality control system that is responsible for removal of the mislocalized protein. To this end, I wanted to understand how the attenuated proteins detected from my screens are removed. A major player in protein degradation is the Ubiquitin-Proteasome System (UPS) which has been noted to be important for the degradation of mislocalized and orphan proteins (see review by Kong *et al.* (2021)). Ubiquitination is an important step to target a protein for degradation by the proteasome. So, I decided to look into if the attenuated proteins are more likely to have ubiquitylation sites and I found that the attenuated hits have a significantly higher number of ubiquitylation sites (Figure 16E). The presence of more ubiquitylation sites may indicate that the attenuation is achieved through degradation by the ubiquitin proteasome system. Then, the follow up question is how are the proteins being recognized?

One possible method for proteins to be recognized for degradation is through their N-termini or N-degron. To check for the role of N-degrons on the degradation of the attenuated proteins, I compared the hits against a dataset of protein stability index (PSI) values from a study where they determined how protein turnover was affected by different yeast N-termini (referred to as N-terminome) (Kats *et al.*, 2018). The PSI values ranges from 0-1 with lower values indicating that the N-terminal is unstable and may contain an N-degron, and higher values indicating that the N-terminal is stable and less likely to contain an N-degron. I found that the attenuated hits have a significantly higher PSI (Wilcoxon rank sum test: $p = 8.19e-4$, $W = 676079$) and are actually depleted of from the pool of proteins with $PSI < 0.5$ (Fisher's Test: $p = 0.002114$, odds = 0.70) (Figure 16F). The results show that N-degrons do not play a major role in the degradation of the attenuated hits.

The next step is then to look at the C-terminal of the protein as a possible recognition site for protein degradation. There is no published study directly addressing the stability of C-termini similar to the N-termini study by Kats *et al.* (2018), but I can use the next available measure, disorderness. While disordered regions are not equivalent to degrons, studies have shown that exposed disordered regions can form weak but problematic interactions with other proteins and be used for recruitment into aggregates (Grousl *et al.*, 2018; Macossay-Castillo *et al.*, 2019). Additionally, an unstructured disorder region is needed for an efficient degradation by the proteasome (Prakash *et al.*, 2004). A study on intrinsically disordered segments in a protein showed that proteins with a long terminal or internal disordered regions have significantly shorter half-lives, indicating that disorderness plays a role in affecting protein turnover (van der Lee *et al.*, 2014). Here, I would hypothesize that if disorderness plays a role in degradation of attenuated proteins, the attenuated proteins would have a higher disorder score either internal or at the C-terminal. Using this dataset on protein disorder and the disorder scores assigned to each protein, I found that attenuated proteins had a significantly higher C-terminal disorder score (Figure 16G). This result indicates that C-terminal disordered regions may play a role in targeting MLPs for degradation. However, I do need to highlight that the tFT-tagged proteins are tagged on the C-terminal which makes the interpretation here not as straightforward.

As a summary for this part, I have shown that attenuated proteins are enriched with subunits of complexes and essential proteins. The attenuated proteins generally have a high synthesis rate and long protein half-life. In terms of localization, the attenuated proteins are enriched in cytosolic and nuclear proteins but are depleted of mitochondrial proteins. The attenuated proteins are also less hydrophobic and are depleted of proteins with predicted transmembrane domains. Lastly, in terms of degradation and its recognition, the attenuated proteins have significantly more ubiquitylation sites which may indicate that the UPS is involved with the degradation, and the disorderness of the proteins' C-terminal may play a role for recognition.

4.2 Breakdown of tFT-MoBY2 screen results into different attenuation categories

So far, I have focused the analysis on just proteins that are attenuated in terms of decrease in mNG levels. With the use of a tFT, I can also look at changes in red-to-green ratio as a readout of stability. While the cell attenuates directly the level of ~14.6% of proteins, there is a non-trivial number of proteins (303/3889) that fall into the category of reduced ratio (destabilized) but are not attenuated on mNG levels. What does the cell do to these ~7.8% of proteins and how are they different? My theory here is that the cell must have more ways than just degradation to handle protein overexpression and mislocalization. There are studies that have shown that cells in heat shock can temporarily sequester proteins into aggregates that then later dissociate to continue folding and function (Wallace *et al.*, 2015). Besides that, even in aneuploid cells, studies have shown that while a proportion of proteins are attenuated, another separate and distinct population of proteins that are on the extra chromosome tend to aggregate (Brennan *et al.*, 2019).

Therefore, in addition to the properties analyzed above, I categorized the data from the tFT-MoBY2 screen into a few interesting groups based on changes in mNG fluorescence which represent changes in abundance and changes in red-to-green ratio which represents changes in protein stability. The three categories that I analyzed further are:

- (1) proteins that are destabilized but not attenuated,
- (2) proteins that are attenuated but not destabilized, and
- (3) proteins that are attenuated and destabilized.

I propose several hypothesis as to why the proteins in each group behave as they do, provide some hints from the data already obtained, and suggest possible future experiments to look further into strengthening the hypothesis proposed.

4.2.1: Proteins that are destabilized but not attenuated

Within this group, there are 303 proteins that are destabilized but not attenuated (Figure 17A). While the mCH-mNG tFT is not as sensitive in detecting changes in protein turnover as mCH-sfGFP, I would argue that because of that, if a protein has a strong enough change in ratio to be detected then it must be a true change in protein turnover. Accordingly, the points made below on the trends I see remain when I increase the stringency for calling proteins with an attenuated ratio from $\geq 20\%$ decrease and $p < 0.1$ to $\geq 50\%$ decrease and $p < 0.05$.

For a protein to have a decrease in ratio, it would mean that there is a decrease in its stability which likely comes from increased protein turnover, but this would mean that the protein levels should also be attenuated, which is not the case. Thus, in order for a protein to have no change in mNG levels but a decrease in ratio, the cause is then likely not due to protein turnover but a folding/maturation slow down.

I propose that proteins in this group are not degraded but rather the nascent proteins are sequestered into aggregates that may be resolved later for folding once the stress, in this case protein overexpression, is relieved or more of its folding factors or binding partners are available. This type of sequestration for later resolution is not unheard of as it has been described in literature that upon a mild heat stress, cells aggregate proteins to try and slow down harmful effects (Wallace *et al.*, 2015). These aggregates are then later resolved and the proteins are removed from the aggregates to continue re-folding and maturing instead of being sent for degradation.

Several properties of proteins within this group hint that a maturation kinetics slowdown is the case. First of all, the proteins within this group have a significantly lower synthesis rate and are less abundant (Figure 17C). Additionally, the proteins also have a significantly shorter half-life, indicating that in normal conditions, the protein is already degraded relatively faster (Figure 17D). A low synthesis rate and high turnover indicates that the cell would prefer not to have too much of it. However, since the protein levels are already low, if the cell goes for increasing degradation when there is an excess, there may be a risk of removing too much and end up having not enough.

The alternative to immediately going for degradation would be to sequester the excess to slow down maturation and control the resolution from the aggregate. Comparing with data about known aggregates that occur in aneuploid cells with an extra chromosome (Brennan *et al.*, 2019), the proteins within this group are enriched for proteins that tend to aggregate (Fisher's Test: $p = 0.04273$, odds = 1.90) and have a slightly higher, albeit non-significant, propensity for aggregation. Another hint as to why protein degradation does not play a big role here is that proteins from this group have

significantly less number of ubiquitylation sites and it is the only category among the three to have this property (Figure 17F).

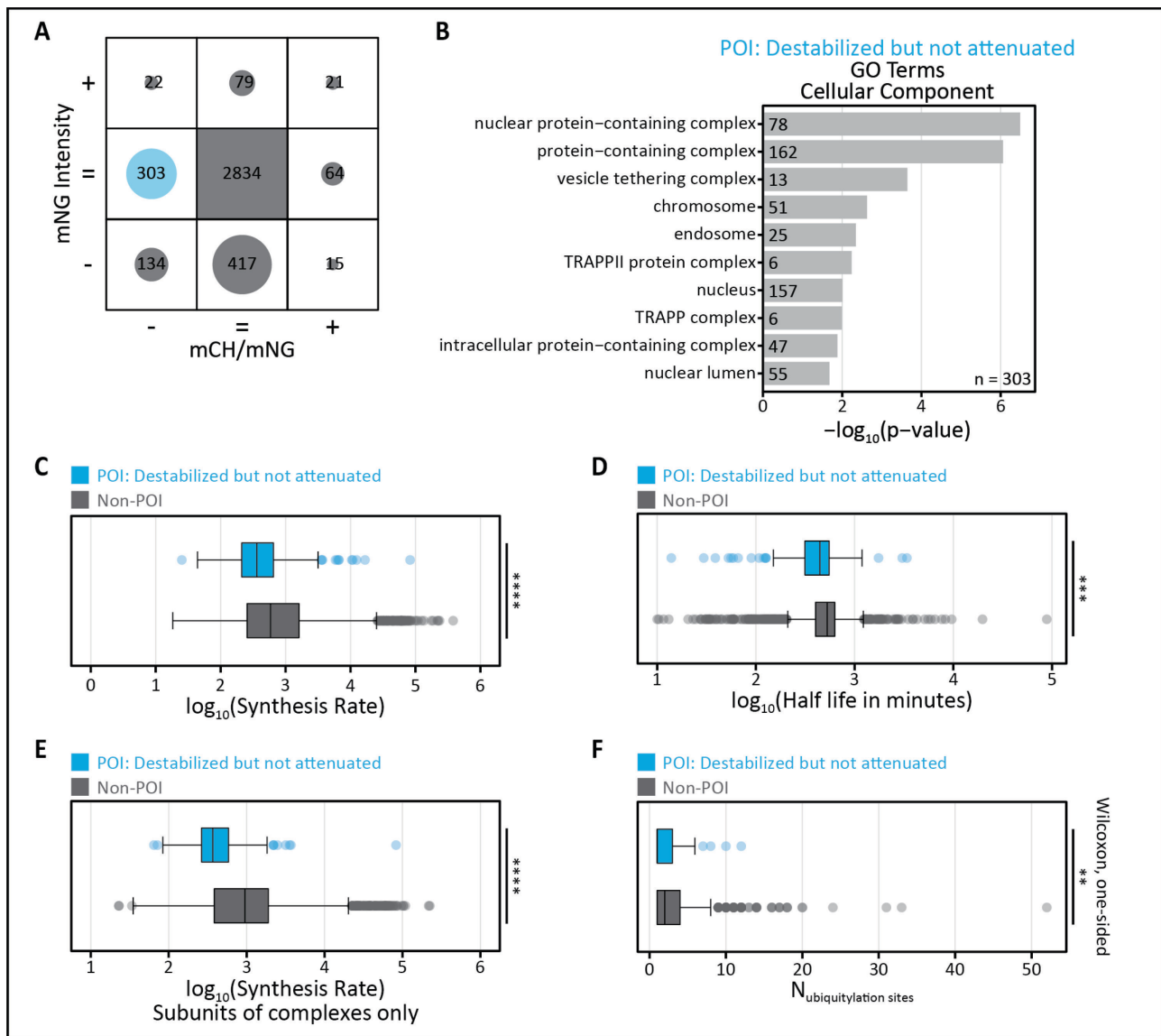


Figure 17: Properties of ORFs that are destabilized but not attenuated

(A) Balloon plot of the 2 μ overexpression screen results as in Figure 11G with the category being discussed highlighted in blue. The highlighted proteins are proteins of interest (POI) that were destabilized but not attenuated (B) GO term enrichment analysis results for cellular component of the POI. (C-F) Comparison of POI against non-POI in terms of (C) log₁₀ synthesis rate, (D) log₁₀ protein half-life, (E) log₁₀ synthesis rate of subunits, and (F) number of ubiquitylation sites. Data sources: Protein half-life data (Christiano *et al.*, 2014); Synthesis rate data (Taggart and Li, 2018); Ubiquitylation site data (Swaney *et al.*, 2013); Subunits annotation (Meldal *et al.*, 2021). (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001)

Additionally, the GO term enrichment analysis indicates that subunits of protein complexes are also highly enriched in this group (Figure 17B). I noted that the subunits of complexes that are within this category have a significantly lower synthesis rate as well (Figure 17E). It is possible then that the cell does actually attempt to manage the protein subunits that are non-attenuated in terms of mNG upon overexpression through an alternate pathway to, in the end, have proper level of complexes forming.

Naturally, the follow up experiment that is missing here would be to image the proteins from this group and see if there are aggregates forming upon overexpression. Another alternative would be to use differential centrifugation to extract aggregates and determine if the proteins from this group is highly enriched in the aggregates.

4.2.2: Proteins that are attenuated but not destabilized

Within this group there are 417 proteins that are attenuated but not destabilized (Figure 18A). I find this group of proteins to be highly interesting because there could be two possibilities that are not mutually exclusive from one another that may explain the behaviour.

The first hypothesis I would propose is that proteins in this group are attenuated, either partially or fully, through autoregulation or post-transcriptional regulation. Through pre-translational regulation, the levels of proteins can decrease but the maturation and degradation kinetics should remain the same and so the ratio remains unchanged.

The second hypothesis that I propose is that there is a bias in degradation of newly synthesized proteins compared to mature proteins that are maybe already assembled into a complex or localized to a compartment and performing its function. Here, mature and already in-complex proteins are less likely to be degraded so the ratio remains steady but the degradation of newer proteins leads to the decrease in green fluorescence. This biased degradation of newer subunits than those that are already assembly have been described in literature (Mcshane *et al.*, 2016).

Let us look at the data that would argue for both hypothesis. First of all, the proteins in this group have a significantly higher synthesis rate and longer half-life (Figure 18C and 18D) indicating that the cell is producing these proteins very rapidly and keeping it around for longer. Thus, instead of relying solely on degradation to regulate its levels, which may not be fast enough or become overwhelmed, the cell reduces the synthesis in the pre-translational stages via mechanisms such as promoter autoregulation, mRNA splicing, mRNA degradation, or translational initiation slowdown.

A GO term enrichment analysis on proteins within this group leads to enrichment for terms such as stress granule and ribonucleoprotein granule (Figure 18B). These terms refers to non-membrane bound organelles consisting of mRNA and proteins that form when translation initiation is limiting, for example, during overexpression.

To further strengthen the argument for both pre-translational regulation and biased degradation of newly synthesized proteins, we can look at the ribosomal subunits. Ribosomal subunits have been shown to have some level of autoregulation (Roy *et al.*, 2020) and has also been shown to rapidly degrade excess produced unassembled ribosomal subunits (Sung, Reitsma, *et al.*, 2016). A large proportion of ribosomal subunits that were attenuated actually fall into this category, with some of the ribosomal subunits shown to be autoregulated to an extent. Interestingly, the ribosomal subunits that are within this group and found to be autoregulated have a much stronger attenuation than ribosomal proteins that are not autoregulated (Figure 18E). It is likely that the stronger attenuation comes from a combination of both the autoregulation and degradation of excess subunits.

Additionally, I quantified the mNG gene expression levels in a handful of strains, not restricted to only this group, through RT-qPCR to initially show that all the attenuation comes from post-translational mechanisms. From the scatter plot in Figure 18F, I can see that in majority of the cases, attenuation of the mNG level as measured on colony is not due to transcriptional attenuation and there is no bias on whether or not the protein is a subunit of a complex. Unintentionally, I also found HMO1, a protein known to be autoregulated, to be strongly attenuated in both its mNG expression level and mNG fluorescence level (Xiao *et al.*, 2011). Along with HMO1, I also see other proteins that have not been described to be autoregulated but have a reduced mNG expression upon overexpression (e.g. Zuo1, Emc5, Hsp78, Nar1). The change in mNG gene expression results indicates that some of these proteins may potentially be transcriptionally attenuated. However, not all these proteins are within this category, so maybe there are also other pathways that work together with autoregulation to control the protein levels.

Besides that, I also find that the proteins within this group have a significantly higher C-terminal disorder score (Figure 18G). The higher disorder score at the C-terminal could possibly indicate that the proteins are recognized by the C-terminal for degradation. Additionally, I would suggest that the recognition at the C-terminal would also allow time for a protein to fold or engage with chaperones/partners as it is synthesized to allow it to mature as much as possible. However, once produced and if it is not needed, it can be recognized by its C-terminal for removal.

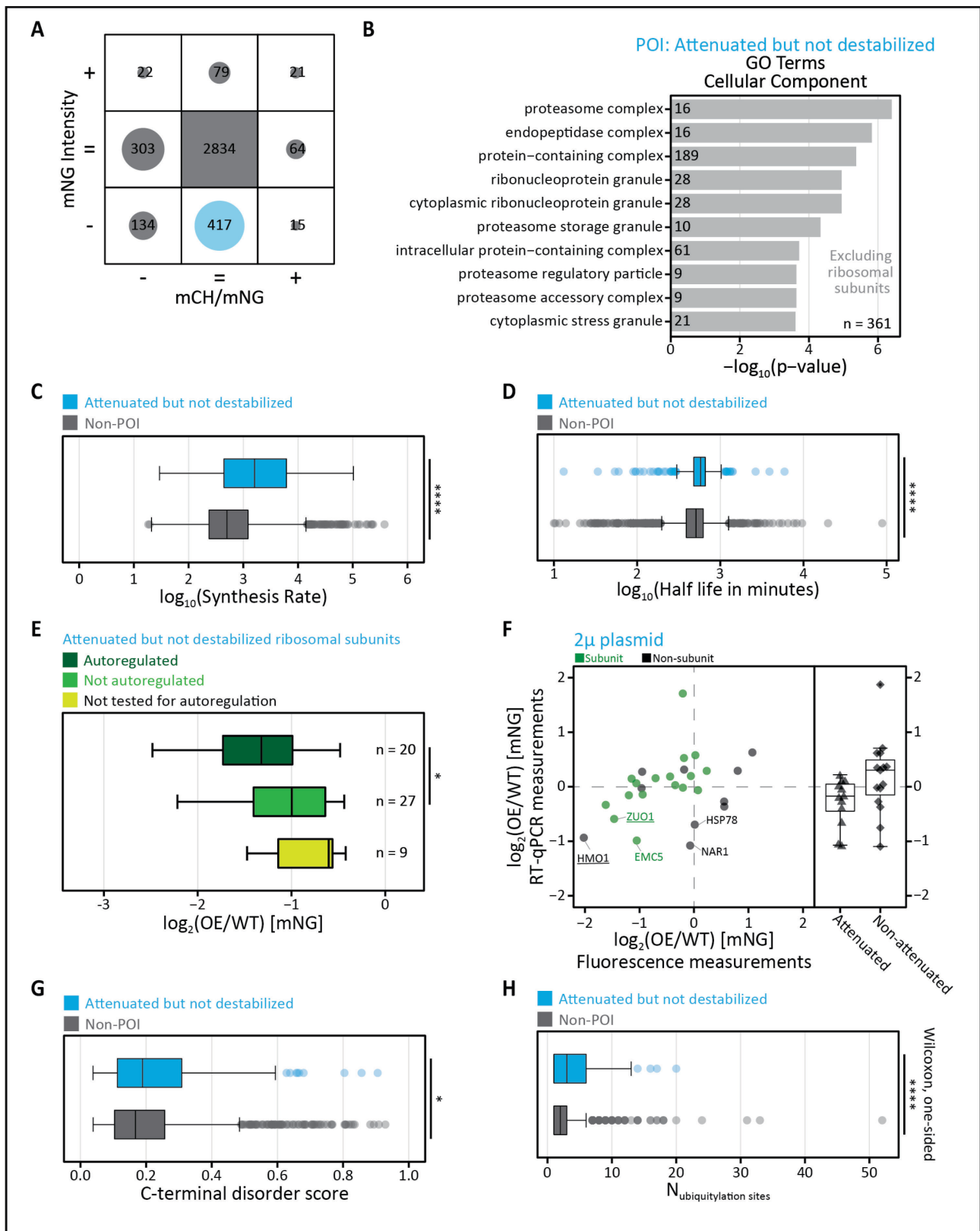


Figure 18: Properties of ORFs that are attenuated but not destabilized

(A) Balloon plot of the 2 μ overexpression screen results as in Figure 11G with the category being discussed highlighted in blue. The highlighted proteins will be referred as proteins of interest (POI)
 (B) GO term enrichment analysis results for cellular component of the POI. (C-D) Comparison of POI against non-POI in terms of (C) \log_{10} synthesis rate and (D) \log_{10} protein half-life.

(Legend continues on next page)

(E) Comparison of log₂ fold change in mNG levels between ribosomal subunits (within POI) that are known to be autoregulated, not autoregulated, or not tested in the dataset from Roy *et al.* (2020). (F) Log₂ fold change of mNG expression level upon overexpression as measured by RT-qPCR correlated with log₂ fold change of mNG levels from colony fluorescence measurements (left) and categorized on mNG level attenuation behaviour from colony fluorescence measurements (right). Genes with ≥25% decrease in mNG expression levels are labelled. Underlined genes are POI. (G-H) Comparison of POI against non-POI in terms of (G) C-terminal disorder score and (H) number of ubiquitylation sites. Data sources: Protein half-life data (Christiano *et al.*, 2014); Synthesis rate data (Taggart and Li, 2018); Ubiquitylation site data (Swaney *et al.*, 2013); Subunits annotation (Meldal *et al.*, 2021). (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001)

Another possibility would be that, instead of being the target for recognition and removal, the more disordered C-terminal causes inappropriate interactions which leads to its detection as aberrant proteins destined for removal. In line with this, the proteins within this group also have significantly more ubiquitylation sites which could have a signalling function and if not deubiquitinated, it is then marked for degradation (Figure 18H).

This group of proteins can be further studied through more RT-qPCR measurements to determine if there are more proteins with attenuated mNG expression levels that could hint at a sort of autoregulation or pre-translational regulation. As a further test, one could swap the promoter of a protein that is autoregulated with a non-autoregulated promoter, and by doing so expect that the downregulation effect would be abolished. To try and understand further on biased degradation of newly synthesized proteins and maybe identify new substrates under this umbrella, one could reproduce the global pulse-chase experiment conducted by Mcshane *et al.* (2016) but in yeast cells to detect non-exponentially degraded proteins.

4.2.3: Proteins that are attenuated and destabilized

Within this group there are 134 proteins that are both attenuated and destabilized, which indicate that they are likely to be degraded (Figure 19A). This group of protein is highly useful to study and look for factors that play a role in degrading mislocalized proteins since this group is the most likely to be attenuated due to increased protein turnover.

A GO term enrichment analysis shows that the group is also enriched with subunits of protein complexes and that 105 out of the 133 are annotated within this GO term (Figure 19B). These subunits of complexes have a significantly lower synthesis rate than the remainder subunits (Figure 19C). The behaviour is in contrast to the proteins in the previous section that have a higher synthesis rate. Interestingly, the attenuated subunits of complexes still have a significantly longer half-life (Figure 19D). I would suggest that maybe the cell can afford to rely on degradation of these proteins since the synthesis is slower even though the proteins remain longer in the cell.

Additionally, the proteins are also found to have a lower average overall disorder score (Figure 19F). The lower disorder score seen covers both subunits and non-subunits, however, for the non-subunits the C-terminal and internal disorder scores are also significantly lower. Could a lower disorder indicate that more specific and targeted degradation machinery is needed?

The proteins within this group are already being used for further studies by another colleague in the lab. She has shown in her work that proteins from this group are stabilized upon proteasomal inhibition which directly implies that the proteins are being degraded upon overexpression by the UPS. Further work on this would be to identify the different E3 ligases that are responsible for the degradation of these proteins and DUBs that can reverse the phenotype.

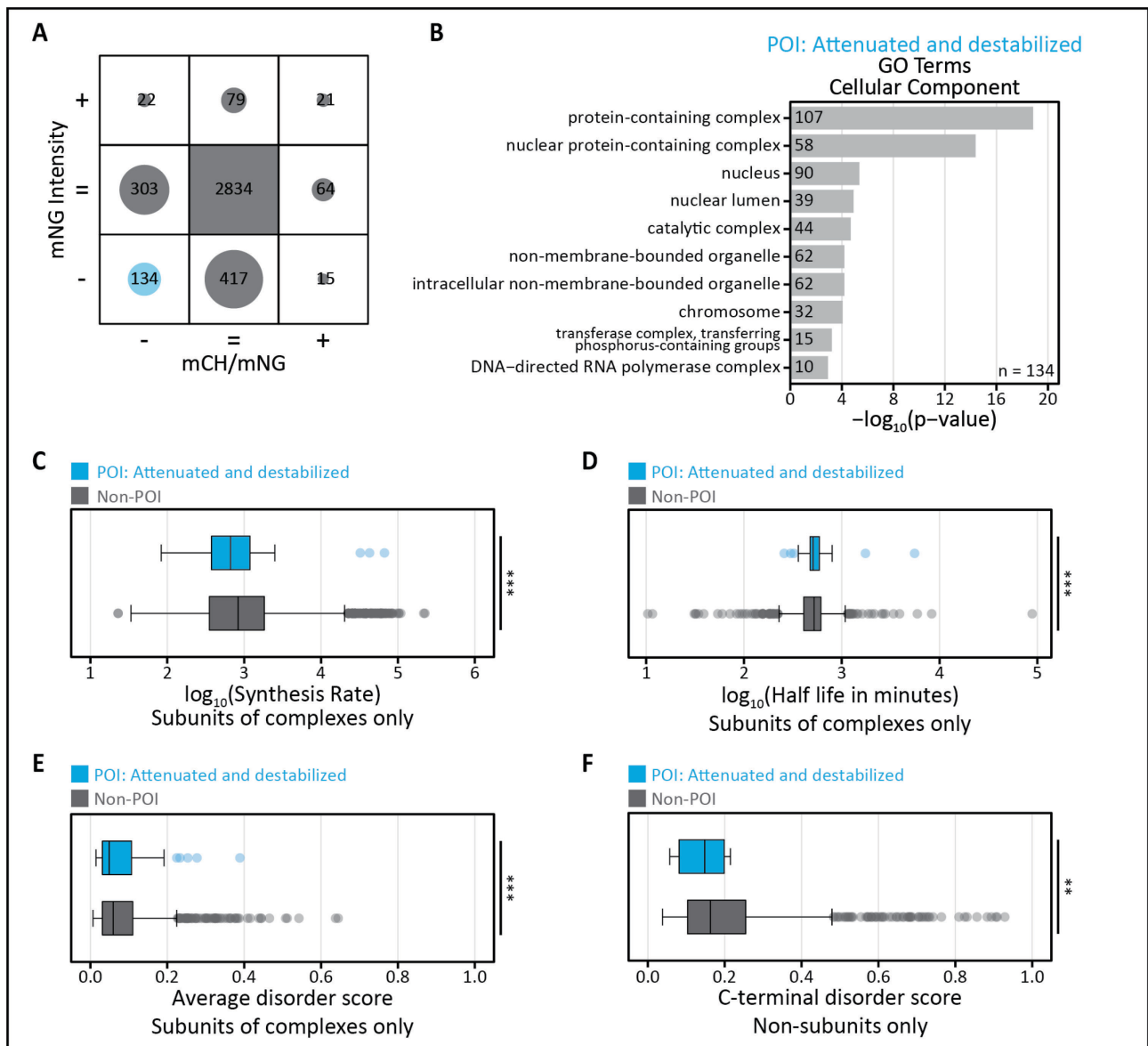


Figure 19: Properties of ORFs that have attenuated mNG levels and ratio.

(A) Balloon plot of the 2μ overexpression screen results as in Figure 11G with the category being discussed highlighted in blue. The highlighted proteins will be referred as proteins of interest (POI) (B) GO term enrichment analysis results for cellular component of the POI. (C-F) Comparison of POI against non-POI in terms of (C) \log_{10} synthesis rate of subunits, (D) \log_{10} protein half-life of subunits, (E) average disorder score of subunits, and (F) C-terminal disorder score of non-subunits. Data sources: Protein half-life data (Christiano *et al.*, 2014); Synthesis rate data (Taggart and Li, 2018); Ubiquitylation site data (Swaney *et al.*, 2013); Subunits annotation (Meldal *et al.*, 2021). (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$)

CHAPTER 5: Understanding the difference in attenuation behaviour of subunits of complexes

As seen from the results shown for the three overexpression screens, one of the most common attribute of the attenuated proteins is that they are highly enriched with subunits of protein complexes. This enrichment comes from attenuated subunits across various different complexes regardless of size. While the probability of a complex having a subunit that is attenuated increases as the complex size increases, not all subunits within a complex behaves the same way, that is, not all subunits are attenuated together.

In this chapter, I look at some properties that may explain the difference in attenuation behaviour between the attenuated and non-attenuated subunits. In a bid to understand further on the attenuation of subunits in relation to levels of its partner subunits, I also carry out a screen on select complexes to overexpress or knockout every partner subunit within the complex. I would like to understand how the synthesis rate and half-life play a role in determining attenuation of subunits.

Additionally, large multi-subunit complexes assemble from various different subunits and sometimes intermediate sub-complexes form first before the final complex formation takes place (Kane, Tarsio and Jianzhong, 1999; De Wulf, McAinsh and Sorger, 2003; Mueller *et al.*, 2015). Thus, I would like to know if the assembly sequence of a complex may play a major role in determining protein attenuation.

5.1 Properties of attenuated subunits

As discussed in an earlier chapter, the N-terminal of proteins may contain an N-degron that leads to recognition and degradation. Could it be that the N-terminus for the attenuated subunits is hidden upon complex formation, and so when present in excess without a binding partner, this region is exposed for recognition? I compared the attenuated subunits against the non-attenuated subunits using the N-terminome dataset of PSI values and found that there is no clear trend to support this train of argument (Figure 20A).

In a similar fashion, after checking the N-terminal I looked into C-terminal or internal disordered regions and if they could play a role in promoting degradation? Checking this data, there is some significant difference between the attenuated and non-attenuated subunits where the former has a higher C-terminal disorder score (Figure 20B). As noted earlier, the tFT-tag is on the C-terminal of the protein, and so the results here is not easily interpretable. Besides that, the internal disorder

score was not significantly different between the attenuated and non-attenuated subunits, but I still cannot rule out that the exposed internal interaction surfaces are used for recognition and just may not be disordered.

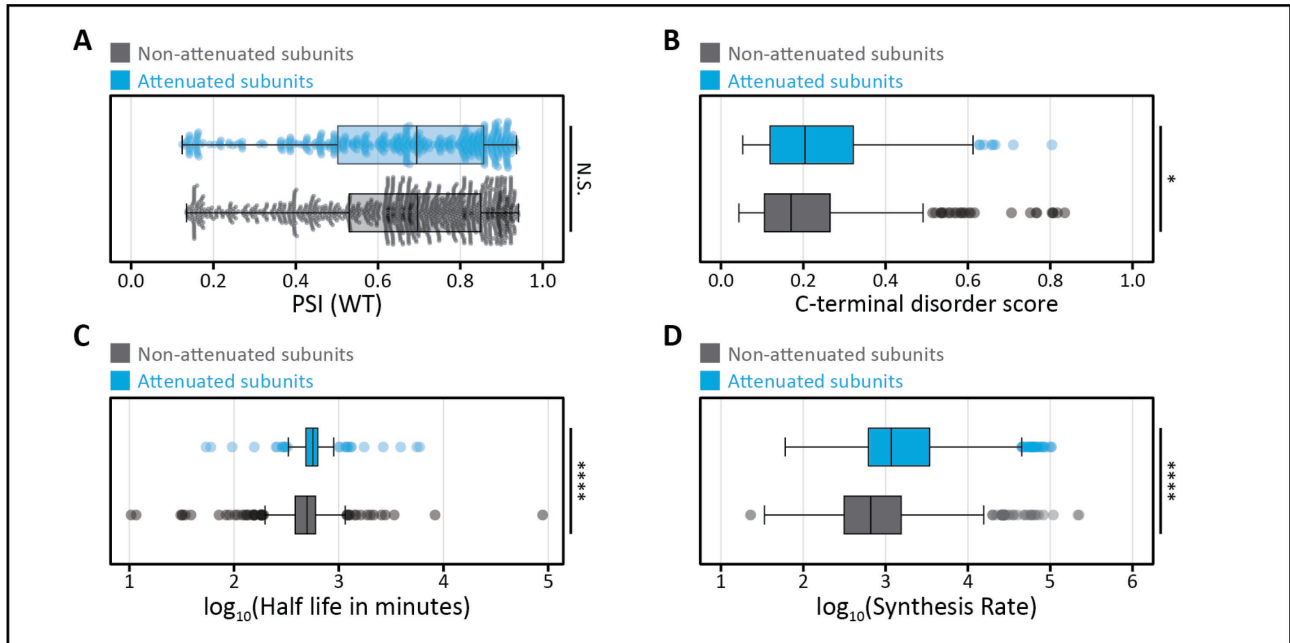


Figure 20: Properties of attenuated subunits of complexes within tFT-MoBY2 screen.

(A-D) Comparison of attenuated subunits against non-attenuated subunits in terms of (A) protein stability index (PSI) of wild-type yeast N-termini, (B) C-terminal disorder score, (C) \log_{10} protein half-life, and (D) \log_{10} synthesis rate. Data sources: Protein half-life data (Christiano *et al.*, 2014); Synthesis rate data (Taggart and Li, 2018); PSI data (Kats *et al.*, 2018); Subunits annotation (Meldal *et al.*, 2021). (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$)

Besides that, attenuated subunits also have significantly higher synthesis rates and half-life (Figure 20C and 20D). However, once the values are median centered per complex, this trend disappears. The two observations taken together indicate that while synthesis rate and half-life may have a role in general in determining attenuation behaviour, but within a given complex, this role is very much diminished. On the other hand, taking a per-complex approach in median centering the synthesis rates and half-life may not be the best method since subunits of a complex are not all synthesized at the same time, assemble together instantaneously, and may have stable intermediate sub-complexes.

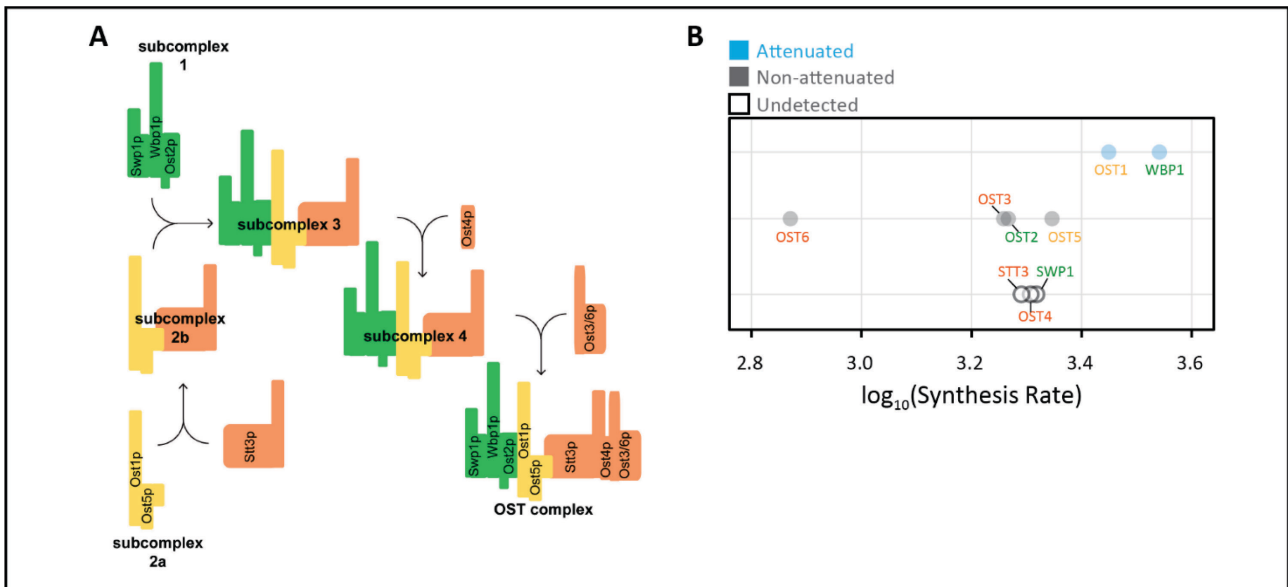


Figure 21: OST complex assembly and synthesis rate correlation.

(A) Cartoon in Figure 7 of Mueller *et al.* (2015) showing their proposed assembly sequence of OST complex. (B) Synthesis rate of OST complex subunits categorized by attenuated (blue points), not attenuated (gray points), and not detected in tFT2-MoBY screen (white points). The gene names are coloured to match the colour scheme used in (A).

Some well-known complexes have been shown to have a step-wise assembly and may have intermediate sub-complexes that are relatively stable (Kane, Tarsio and Jianzhong, 1999; Mueller *et al.*, 2015). For example, in Figure 21A, the OST complex forms in a step-wise manner with different subunits forming smaller sub-complexes before assembling further. The knock-out of one subunit usually affects other subunits within its own sub-complex much more, and also affects subunits that are downstream to its assembly but not upstream. Interestingly, using the OST complex as an example, the attenuated subunits obtained from the screen has the highest synthesis rates among all subunits and also within its own sub-complex (Figure 21B). Coincidentally, the attenuated subunits are also among the first in the sequence of assembly. Since median-centering subunits synthesis rates per-complex does not show a difference between attenuated and non-attenuated subunits, from this example, I would like to know then if the assembly sequence plays a heavier role in determining if a subunit should be attenuated upon overexpression.

5.2 Overexpression and knockout screen across subunits within complexes

In order to understand further, the many possibilities that determines a subunit's attenuation behaviour, I conducted a screen that overexpresses or knocks out each subunit within the complex and follow the fate of its partners. For this set of experiments, I chose to focus on the ER membrane complex (EMC), Vacuolar ATPase (V-ATPase) complex, CCR4-NOT complex, and the Glucose Induced Degradation (GID) complex, all of which have a clearly defined list of subunits and their structures mostly resolved. All four complexes have a mix of subunits that were attenuated upon overexpression of itself and subunits that were not attenuated.

5.2.1: Endoplasmic Reticulum Membrane Complex (EMC)

The EMC is an 8 subunit complex (9 in humans) localized at the ER membrane that plays a major role in the biogenesis of many integral membrane proteins through promoting insertion and stabilizing transmembrane domains for polytopic membrane protein's biogenesis (O'donnell *et al.*, 2020; Volkmar and Christianson, 2020). The structure of the protein has been recently fully resolved and it unveiled two different binding pockets that are used for its role in biogenesis of integral membrane proteins (Miller-Vedam *et al.*, 2020). So far, there are no known assembly factors needed for the assembly of the EMC in yeast, but WNK1 has been shown to be an assembly factor in humans that binds to EMC2 pre-assembly (Pleiner *et al.*, 2021). Using the structural information, I can make a better informed decision when interpreting the overexpression and knockout screen results as I can see which subunits are physically interacting with each other.

Based on the overexpression and knockout screens done for EMC subunits, I can see that synthesis rate does not play a role in determining attenuation of a subunit upon overexpression of itself (See EMC1, EMC2, EMC5, and EMC7 in Figure 22B and 22D). I would also point out that for this particular complex, the range of distribution of synthesis rate is also very small, much smaller as compared to the OST complex subunits in Figure 21B. Protein half-life also does not seem to play a major role in determining attenuation behaviour of the EMC subunits (Figure 22E).

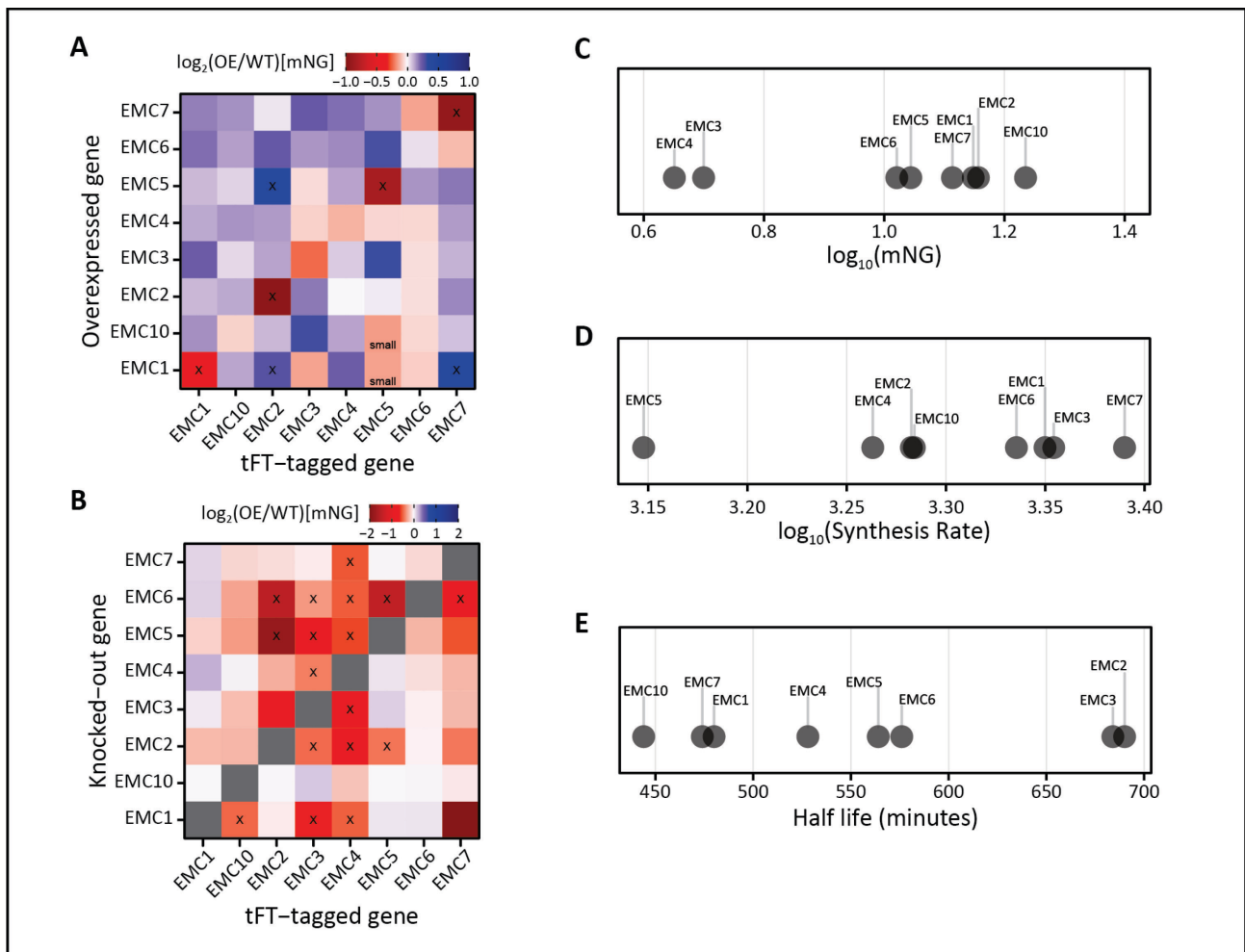


Figure 22: Results of overexpression and knockout crosses for ER membrane complex (EMC). (A-B) Heatmap of \log_2 fold change in mNG levels upon overexpression (A) or knockout (B) of another subunit within the complex. Pairs with significant changes in mNG levels ($\geq 20\%$ decrease or increase; $p < 0.1$) are marked with an 'x'. Pairs with significant changes in colony size are marked with either 'small' or 'large'. Gray squares indicate no measurements available. (C) Expression levels of EMC subunits of the tFT-tagged strain crossed with a knockout control strain (*ygr045cΔ*). (D) \log_{10} synthesis rate of EMC subunits obtained from Taggart and Li (2018). (E) Protein half-life of EMC subunits obtained from Christiano *et al.* (2014).

Using the data, I can also determine subunits that are dependent on another subunit to be able to accumulate. This dependency behaviour may indicate that the protein pair interacts strongly to form a sub-complex, or one's protein stability is directly affected by another. The clear example here would be that *Emc2* depends on *Emc5* for stability since the overexpression of *Emc5* leads to an accumulation and a knockout of *Emc5* leads to a decrease of *Emc2* (Figure 22A and 22B). Similar observations can be made for *Emc1* and *Emc7* (Note: the attenuation behaviour seen for *Emc7*-tFT strain crossed with *emc1Δ* strain is not significant only because of one of the three technical replicates behaving far different.).

The results here also show that, for a relatively small complex, I can determine which subunits are most likely to be the most stable and the starting subunit in an assembly sequence. The knockout of all partner subunits did not affect the levels of Emc1 and Emc6 whereas the knockout of Emc1 and Emc6 leads to decrease in mNG levels for almost all of the other subunits. Taken together, the data argues that these two are probably the initial subunit. One interesting observation is that, if Emc6 is the starting subunit because the knockout of it leads to decreased levels of almost all other subunits, the overexpression of Emc6 should lead to accumulation of the other subunits but this is not the case. This observation implies that the cell also maintains a separate level of control on the level of the full complex by not just forming more of the complex when the starting subunit is present. Besides the starting subunit, I can also determine which subunits are most likely to be at the end of the complex assembly sequence. The Emc3 and Emc4 levels are affected by the knockout of almost all subunits within the complex which indicate that they may be at the end of the complex assembly since the lack of any subunits leads to lack of complex formation and degradation of the subunit.

Using the overall data from the EMC screens, I would like to propose a model on the assembly pathway of the EMC. As established, Emc6 is likely the starting subunit (Figure 23A(i)). Emc2 depends on Emc5 for stability and thus probably forms a temporary sub-complex which goes on to bind with Emc6 (Figure 23A(ii & iii)). Then, binding of Emc1 comes into play since the knockout of it did not affect the levels of Emc2, Emc5, and Emc6 (Figure 23A(iv)). Lastly, Emc3 and Emc4, which also probably forms a sub-complex of itself that is highly unstable, binds to the rest of the complex to complete it (Figure 23A(v)). Emc3 and Emc4 is likely the last to assemble as the knockout of everything else leads to its attenuation. From the data, I am unable to place at which step exactly where the peripheral subunits, Emc7 and Emc10, joins the complex (Figure 23A(vi)). However, it is likely that Emc7 binds to Emc1 to cause some kind of conformational change that allows for Emc4's stability in the rest of the complex. The knockout and overexpression of Emc10 does not affect any other subunit and so depend solely on presence of Emc1.

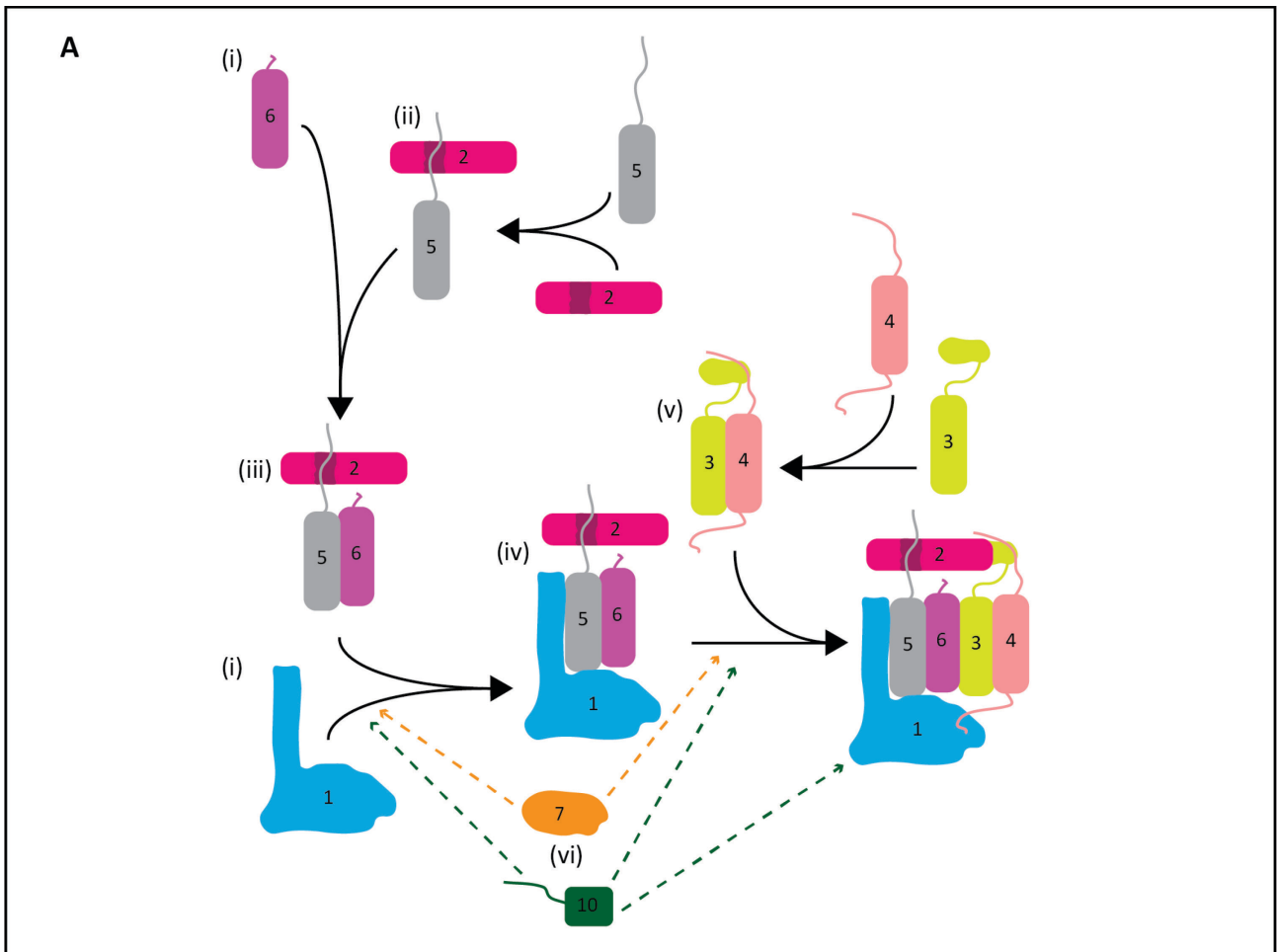


Figure 23: Proposed model assembly pathway for EMC.

(i) Emc1 and Emc6 are most likely at the start of the assembly chain. (ii) Emc2 binds with Emc5 to form a temporary sub-complex where Emc2 is stabilized by binding with Emc5. (iii) This sub-complex (ii) then binds with Emc6 to form the next sub-complex. (iv) Emc1 then binds to the combined product to have the penultimate sub-complex. (v) Emc3 and Emc4 interact with each other and then binds to the sub-complex of Emc1, Emc2, Emc5, and Emc6 (iv). (vi) Emc7 and Emc10 can bind at several possible points within the chain of assembly as they both depend on interaction with Emc1 only. However, it is likely that Emc7 binds before Emc4 joins the complex.

5.2.2: Vacuolar Adenosine-Tri-Phosphatase (V-ATPase) Complex

The V-ATPase is a large complex consisting of 15 subunits that form two major sub-complexes termed as V₀ and V₁ sub-complexes. The V₀ sub-complex is largely membrane embedded and consists of 6 subunits and one of two subunits, Vph1 or Stv1, that determine if the complex will be localized in the vacuole or Golgi respectively (Vasanthakumar *et al.*, 2019). The central proton ring in the V₀ sub-complex consists of Vma11, Vma16, and 8 copies of Vma3, while the other subunits are all present as single copies (Mazhab-Jafari *et al.*, 2016). The V₁ sub-complex is wholly cytosolic and consists of 8 subunits for which 4 of these subunits are present in three copies per complex (Vma1, Vma2, Vma4, & Vma10) (Oot *et al.*, 2016).

The assembly of the complex is relatively complicated as it involves multiple chaperones (for the V₀ subcomplex), an ATP-dependent step to start the assembly (for the V₁ sub-complex), and the RAVE complex in association and dissociation of the V₁ sub-complex from the V₀ (Tomashek, Garrison and Klionsky, 1997; Smardon, Tarsio and Kane, 2002; Ryan, Graham and Stevens, 2008). Besides that, early studies have shown that the complex can form in a concerted manner where V₁ subunits can begin its formation on Vph1/Stv1 or in an independent manner where V₁ and V₀ sub-complexes form individually and are then assembled together at the target destination (Kane, Tarsio and Jianzhong, 1999). The V₁ sub-complex can also reversibly dissociate from the full complex in conditions of low glucose and the free V₁ sub-complex then depends on the RAVE complex to maintain its stability and prevent degradation (Smardon, Tarsio and Kane, 2002; Smardon *et al.*, 2014).

Although the complex assembly is relatively complex, studies have also shown that there are some intermediate steps of smaller sub-complexes forming that are relatively stable before forming the V₀ and V₁ sub-complexes. Since I would also like to know if the attenuation behaviour of subunits would be clearer when making comparisons within these smaller intermediate sub-complexes rather than the full complex, I would hope that studying the V-ATPase may shed some light in this direction. Additionally, since some of these subunits are also known to rely on presence of chaperones for stability, it could be that these subunits are more tolerant to overexpression.

Based on the screen using V-ATPase subunits, it is also clear that neither synthesis rate nor protein half-life alone is the main player in determining attenuation of a subunit upon overexpression of itself (See VMA4, VMA8, STV1, and VPH1 in Figure 24A, 24D, & 24E). Since the assembly process of the V-ATPase into its complex is relatively well described, I wanted to see if I could build an assembly sequence from my data that would match to the proposed assembly process (similar to

how I did for the EMC subunits in Chapter 5 Section 5.2.1). Unfortunately, the knockout of most V-ATPase genes seem to have a synthetic lethal effect with the tFT-tagged genes (Figure 24B). The lack of data from the knock-out screen makes it near impossible to build an assembly sequence of the whole complex based on overexpression data alone. It also hinders the attempt to understand the attenuation behaviour of different subunits within the whole complex upon overexpression. Aside from the missing data, having such a large complex also makes interpreting the data more difficult since there are more interaction pairs to look into.

However, using the available data, there are several interesting points to highlight. First, a different factor may explain the attenuation for Stv1 and Vph1 upon overexpression of itself, the presence of chaperones. Stv1 and Vph1 are subunits of the V0 sub-complex which determine if the complex remains in the Golgi and endosomal system, or localizes to the vacuole. These two subunits are also known to be bound to two chaperones, Vma22 and Vma12 before assembly into the V0 sub-complex (Malkus *et al.*, 2004). Studies have shown that the half-life of the Stv1 and Vph1 subunit decreases dramatically if the chaperones are knocked-out. Therefore, in a similar fashion, overexpression of these subunits leads to insufficient chaperones which also leads to an attenuation in their levels. While not statistically significant, the overexpression of Vph1 leads to a decrease in Stv1 levels and vice versa, also likely due to competition for the same set of chaperones. The V-ATPase V0 sub-complex is also an interesting case for the role of a chaperone since it is known that Pkr1 is needed for the assembly efficiency of the sub-complex and knockout of it reduces the number of full complex formed to ~5-10% (Davis-Kaplan *et al.*, 2006). The major point I want to highlight here is that having a chaperone system for folding stability and assembly efficiency of subunits within a complex, adds a layer of difficulty to understand why certain subunits were attenuated while others not.

If a protein subunit requires a chaperone system to maintain its stability, overexpression of the subunit above the levels of its chaperones should lead to destabilization and degradation. The example here would be for Vph1 and Stv1, where the steady state levels of its chaperones are not really high. It could also be that the lack of attenuation of some subunits of complexes was due to presence of excess chaperone or a robust chaperone system that can handle the excess. Therefore, I could further use overexpression to spot subunits of complexes that may need a chaperone for its stability before assembly since overexpression of a protein well above the level of its chaperones should lead to degradation.

Using the screens result together with the information on the structure of the V-ATPase, I can still attempt to dissect and understand the behaviour of some subunits upon overexpression. For

example, the overexpression of Vma8 leads to increased levels of Vma2 and Vma10 of the V1 subcomplex, and Vma6, Vma9, and Vma16 of the V0 sub-complex. Vma8 is the central stalk of the V1 sub-complex that interacts with c-ring of the V0 sub-complex and so maybe the overexpression of this subunit allows for better stabilization between the two sub-complexes.

Next, looking at the results in small pairwise fashion, I can again highlight some protein pairs that are dependent on one or each other. Vma4 probably depends on Vma2 for its stability since the overexpression of Vma2 leads to increased Vma4 levels but not vice versa (Figure 24A). Based on the V-ATPase structure, Vma4 is a stalk peripheral subunit that binds onto Vma2 in the full complex and so the results of the screen are in agreement with the structure that Vma4 is very likely to depend on Vma2 for its stability. Another such pair, Vma9 and Vma11, may also be interdependent for maintaining its levels since the overexpression of one leads to the increase in the other. In this case, the structure of the complex does not show a direct physical interaction between those two subunits, but Vma9 and the ring subunits (Vma3, 11, & 16) associate via Vma21 (Compton, Graham and Stevens, 2006), which may be bridging the interaction between Vma9 and Vma11 in the presence of excess of either subunits.

Overall, the results from this V-ATPase complex highlight the challenges in using this method to dissect complex assembly and subunit behaviours in large complexes, even more so when there are chaperones involved. However, the data can still be used to make sense of some protein interactions and identify partner pairs that are either interdependent or one dependent on the other. Additionally, overexpression can also be used to pinpoint possible subunits that depend on chaperones for stability before assembly.

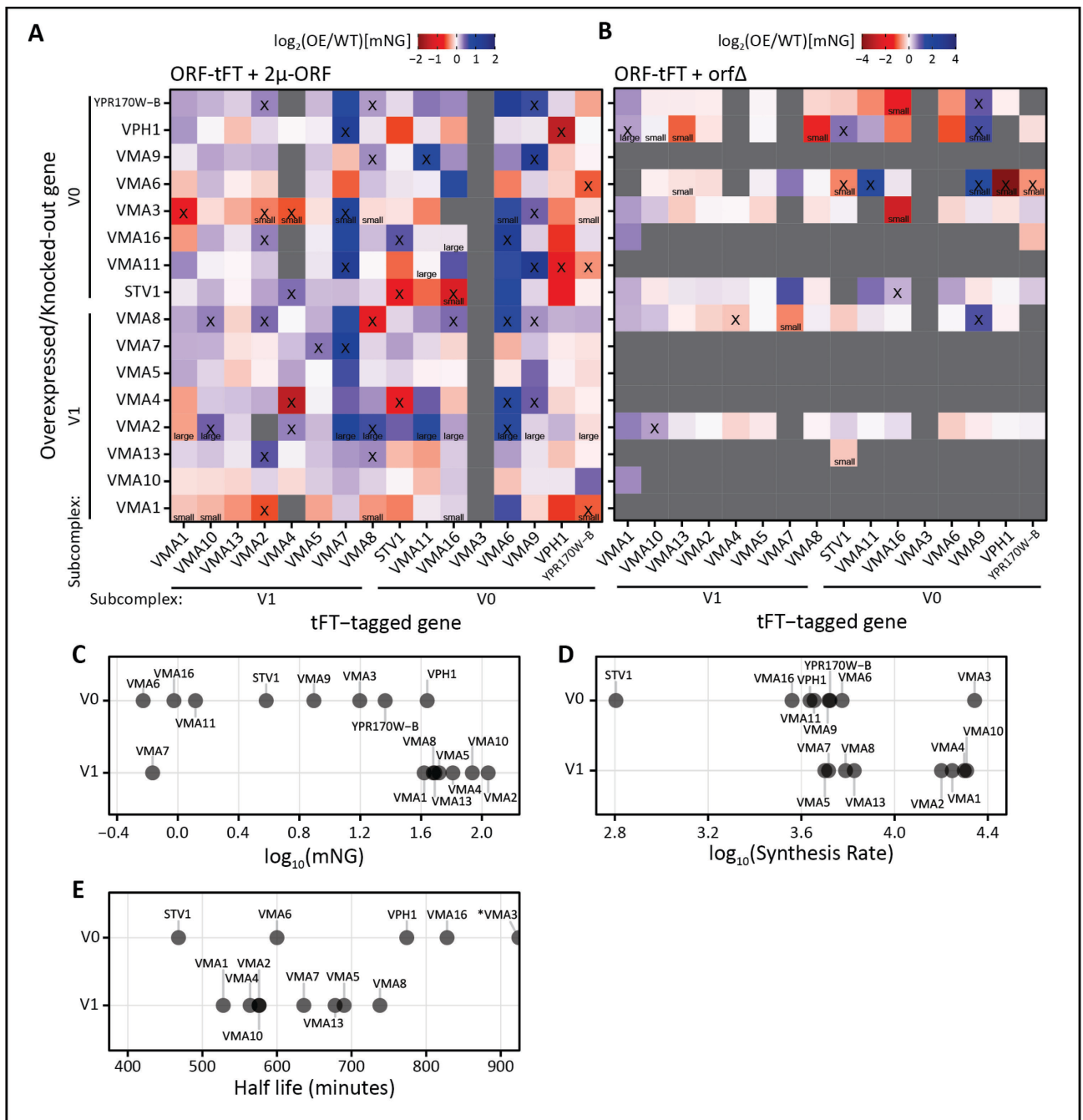


Figure 24: Results of overexpression and knockout crosses for V-ATPase.

(A-B) Heatmap of log₂ fold change in mNG levels upon overexpression (A) and knockout (B) of another subunit within the complex. Pairs with significant changes in mNG levels ($\geq 20\%$ decrease or increase; $p < 0.1$) are marked with an 'x'. Pairs with significant changes in colony size are marked with either 'small' or 'large'. Gray squares indicate no measurements available. (C) Expression levels of V-ATPase subunits of the tFT-tagged strain crossed with a knockout control strain (*ygr045c* Δ). (D) Log₁₀ synthesis rate of V-ATPase subunits obtained from Taggart and Li (2018). (E) Protein half-life of V-ATPase subunits obtained from Christiano *et al.* (2014). *Vma3 (partial data point) has a half-life longer than the measurement range by Christiano *et al.* (2014) and so is annotated as >100 hours (>6000 minutes).

5.2.3: CCR4-NOT Complex

The CCR4-NOT complex is present as a complex of 1.0 MDa and 1.9 MDa in yeast (Chen *et al.*, 2001). Studies trying to identify the full complex almost always manages to pull down a core complex that matches the 1.0 MDa size, and this complex consists of 9 subunits: Ccr4, Pop2 (Caf2), Caf40, Caf130, Cdc39 (Not1), Cdc36 (Not2), Not3, Mot2 (Not4), and Not5 (Nasertorabi *et al.*, 2011). The 1.9 MDa complex probably consists of this core complex with other proteins that associate transiently with the complex such as Btt1, Dbf2, Cal4 (Chen *et al.*, 2001; Cui *et al.*, 2008; Pillet *et al.*, 2022). The complex is organized on Cdc39 as the main scaffold for all other component, and can be defined into modules such as the NOT module, POP2-CCR4 module, and CAF130-BTT1 module (Cui *et al.*, 2008).

While being a relatively simple complex of just 9 subunits, the complex is involved in many cellular processes due to its different modules. This complex plays an essential role in transcription regulation through interaction of Not2 and Not5 with the TFIID transcription factor complex, mRNA decay via the POP2-CCR4 deadenylase activity, and protein ubiquitylation through Not4 (a ubiquitin E3 ligase) and then targeting for proteasomal degradation together with POP2-CCR4 (Collart and Timmers, 2004; Kandasamy, Pradhan and Palanimurugan, 2021). Interestingly, the function of each module is largely unaffected by disruptions of subunits from another module and the stability of each module in the complex depends mainly on the scaffold Cdc39, which is essential in yeast (Bai *et al.*, 1999). Since the modules and functions are so distinct and separate, and there are other subunits that transiently associate with the complex, I would also predict that the assembly of each module would be independent of the others (i.e. there is no stepwise assembly for the full complex).

Based on the screen using CCR4-NOT complex, attenuation due to overexpression of itself only occurs for Cdc39 and Not5 (Figure 25A). Caf40 and Not3 also show strong attenuation but it is not statistically significant (Figure 25A). This complex presents another interesting point and challenge for using this screening method to decipher subunit behaviour because the complex functions in sort of a modular fashion with the modules involved in different functions, the Ccr4 subunit being able to function on its own outside the complex (Maillet *et al.*, 2000; Chen, Chiang and Denis, 2002), as well as having essential subunits that cannot be knocked-out.

Firstly, even though Ccr4 has been shown to be functional and active on its own without being bound to the CCR4-NOT complex, I still see that the knockout of Pop2 leads to attenuation of Ccr4 which may argue that a majority of Ccr4 still binds to Pop2 in the complex than being by itself (Figure 25B and 25C). The overexpression of Pop2, however, does not lead to accumulation of Ccr4 (Figure

25A). In Figure 25D, the steady state levels of Pop2 is quite low compared to Ccr4, so maybe the overexpression is not enough to accommodate accumulating more Ccr4 interactions. I could also speculate that Pop2 can form promiscuous interactions with Caf130 that pulls it away from Ccr4, since it seems that overexpression of Caf130 can lead to accumulation in Pop2 but also leads to an attenuation in Ccr4 levels albeit non-significant.

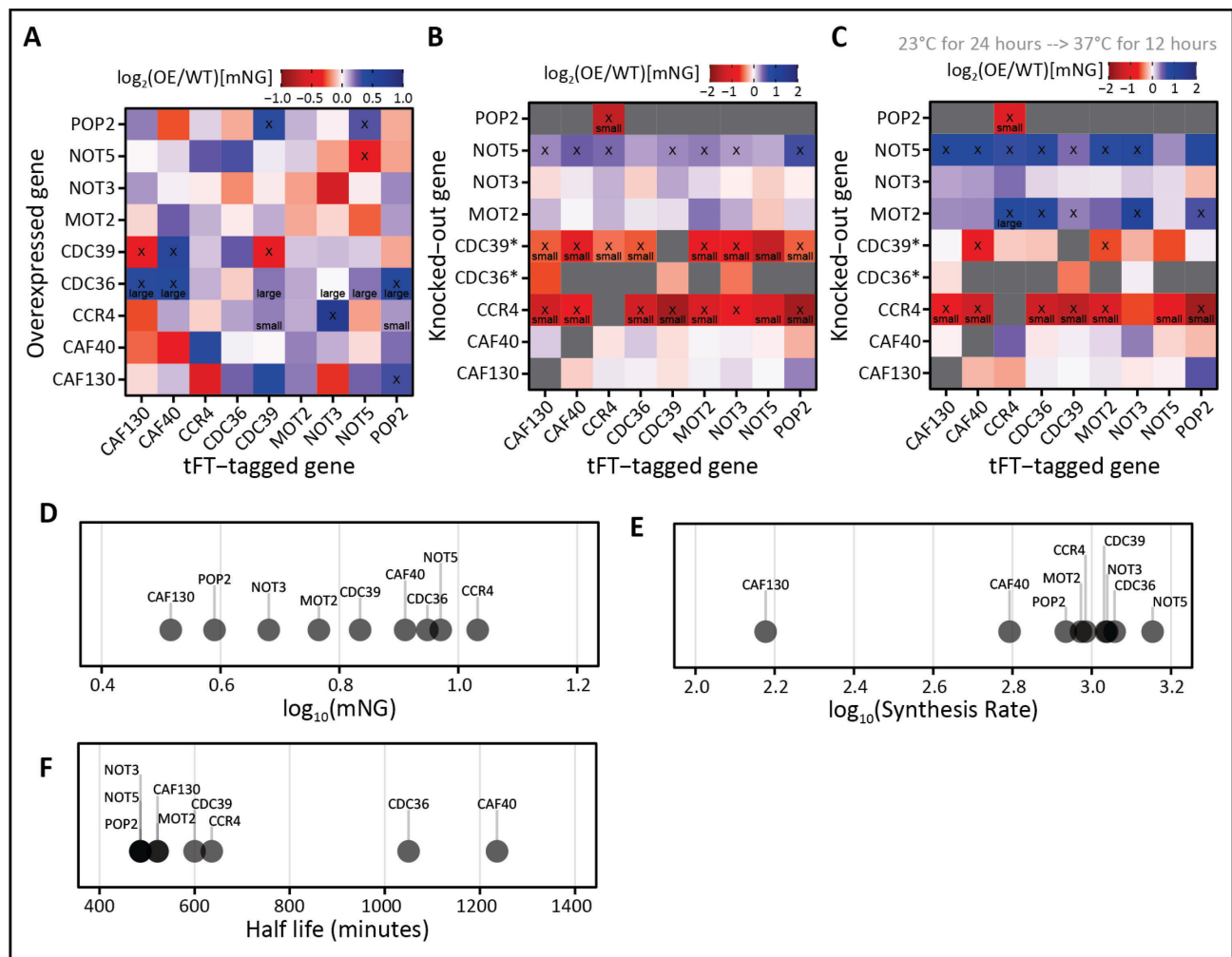


Figure 25: Results of overexpression and knockout crosses for CCR4-NOT complex.

(A-C) Heatmap of \log_2 fold change in mNG levels upon overexpression (A) and knockout (B & C) of another subunit within the complex. Strains in (C) were incubated at 23°C for 24 hours then transferred to 37°C for 12 hours before measurement. Strains carrying temperature sensitive alleles instead of a full knockout are marked with * next to their gene names in (B) and (C). Pairs with significant changes in mNG levels ($\geq 20\%$ decrease or increase; $p < 0.1$) are marked with an 'x'. Pairs with significant changes in colony size are marked with either 'small' or 'large'. Gray squares indicate no measurements available. (D) Expression levels of CCR4-NOT complex subunits of the tFT-tagged strain crossed with a knockout control strain (*ygr045c*). (E) \log_{10} synthesis rate of CCR4-NOT subunits obtained from Taggart and Li (2018). (F) Protein half-life of CCR4-NOT subunits obtained from Christiano *et al.* (2014).

Additionally, any protein that plays an important role in maintaining cellular fitness, even if the strain is viable, will make deciphering colony fluorescence results difficult since colony size and fluorescence levels are correlated. For example, Ccr4 has been noted to play a crucial role in mRNA deadenylation to regulate mRNA decay, affect cell cycle progression when knocked-out, and is even involved in promoting degradation of ubiquitin-dependent substrates (Liu *et al.*, 1998; Kandasamy, Pradhan and Palanimurugan, 2021). Due to its many important roles, knockout of Ccr4 gave viable colonies after crossing with the tFT strains and nearly all partner subunits were attenuated, but most strains had cellular growth defects (Figure 25B and 25C). The outcome for this type of protein is unclear then as to whether it was direct (i.e. due to the partners physically interacting) or indirect (i.e. due to a global change or remodelling). In the case of Ccr4, it is likely a global effect due the growth defect for all except Not3. However, the knockout of Ccr4 leads to decreased Not3 levels while overexpression of Ccr4 leads to increased Not3 levels without any growth effect, which could indicate a direct interaction. Similar to previous examples of this behaviour, this means that Not3 is dependent on Ccr4 for its level in the complex. This result is interesting because Not3 has not been shown to interact directly with Ccr4.

Another challenge in this experiment is deciphering the outcomes for temperature sensitive alleles, since in most cases the protein is still present for binding, just either malformed or inactive when shifted to restrictive growth temperatures. For the *cdc39-ts* allele, the shift from permissive to non-permissive growth temperature only lead to decreased Caf40 and Mot2 levels, and possibly also Not5 but not the other subunits (Figure 25C). The *cdc39-ts* allele carries a single point mutation from Trp to Arg at position 1753 which is in the midst of the binding region for Mot2 and Not5 (residues 1430-1862) (Collart and Timmers, 2004). The binding site of Caf40 is not clearly established as it seems to interact with Caf40, which is in the C-terminal end, and also the N-terminal 1100 residues of Cdc39 (Bai *et al.*, 1999; Chen *et al.*, 2001). Newer studies show that Mot2 has a conserved sequence motif that binds CAF40 (Keskeny *et al.*, 2019), which could then indicate that maybe Caf40 is interacting with Cdc39 at the N-terminal directly and at the C-terminal via Mot2. So, the results would seem to argue that Trp at position 1753 is important for the binding of Mot2 to Cdc39, and by disrupting that binding, it also affects the stability of Caf40. Interestingly, overexpression of Cdc39 lead to increased levels of Caf40 but not others. Taken together with the temperature sensitive allele results, the stability of Caf40 most likely depend directly on Cdc39.

As mentioned before, the exact binding of Caf40 is unclear as it seems to interact with both the C and N termini of Cdc39. Caf130, another component of the CCR4-NOT complex, also exhibit such interactions with both terminal regions of Cdc39 (Collart and Timmers, 2004). Interestingly, the

overexpression of Cdc39 leads to opposing effects on Caf130 and Caf40 levels, whereby Caf130 decreases upon overexpression of Cdc39 and Caf40 increases (Figure 25A). Studies carried out on gene expression changes with knockouts of Caf40 and Caf130 indicate that the two affect very different sets of genes (Cui *et al.*, 2008). It is unclear to me at this point in time as to why there could be such differences.

5.2.4: Glucose Induced Degradation (GID) Complex

The results of the screen with the GID complex subunits are shown in Figure 25. Based on previous experiments by other colleagues working with the GID subunits, it is known that the GID subunits 1, 2, 4, 5, 8, and 9 that are C-terminally tagged with the tFT are non-functional. Regardless, the screen was still performed and the results obtained did not show any striking outcomes. Since I do not have the information as to whether these proteins still physically interact and form the complex, I am unable to interpret and analyze the data further.

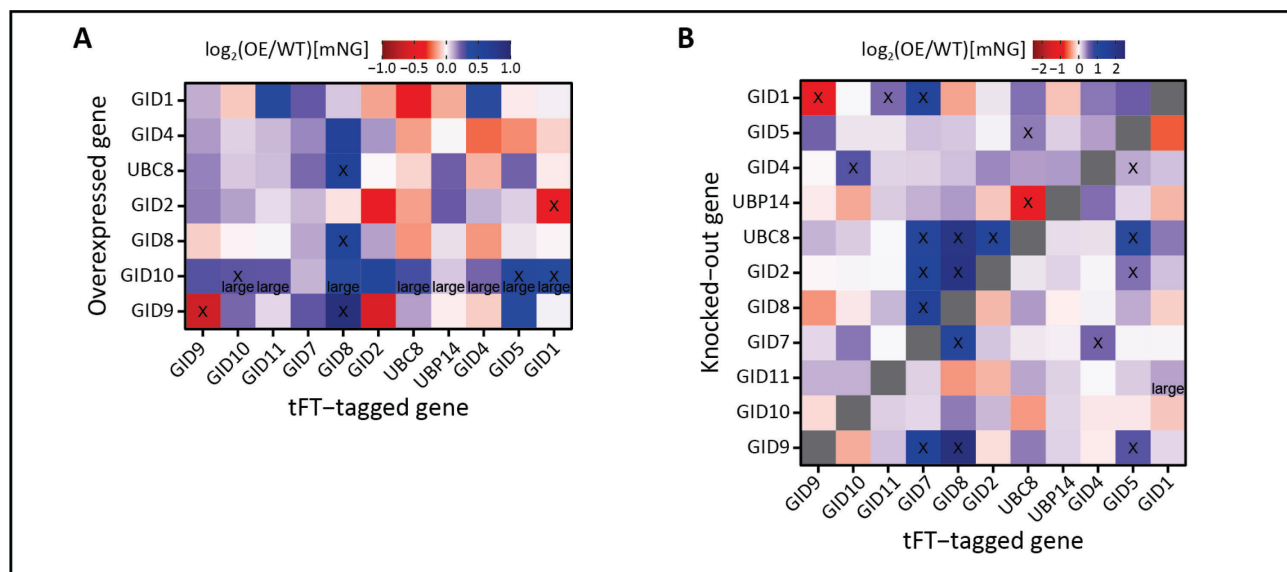


Figure 26: Results of overexpression and knockout crosses for GID complex.

(A-B) Heatmap of log₂ fold change in mNG levels upon overexpression (A) or knockout (B) of another subunit within the complex. Pairs with significant changes in mNG levels ($\geq 20\%$ decrease or increase; $p < 0.1$) are marked with an 'x'. Pairs with significant changes in colony size are marked with either 'small' or 'large'. Gray squares indicate no measurements available.

5.2.5: Summary of the overexpression and knockout screen across subunits within complexes

From the overexpression screens conducted with the tFT-MoBY2 library, I found that synthesis rate does play a large role in differentiating between the attenuated subunits and non-attenuated subunits in general but not at a per-complex level. Through the overexpression and knockout screens for the select complexes, synthesis rate alone really does not play a role in determining the attenuation behaviour of a subunit. However, I have shown that the complex assembly sequence is important and that chaperones also play a role in determining if a protein would be attenuated upon overexpression of itself.

Additionally, I have shown that this screening method of crossing each partner with a knockout or overexpression of another partner subunit can be used to deduce a complex's assembly sequence, if there is such a step-wise assembly. The technique also can be used to determine pairs of proteins that are dependent on one another for their stability and may pinpoint novel interactions that may not have been reported. However, as with all methods, the challenges or limitations for this method is that:

- (1) As the complex size increases, interpretation of a full complex becomes complicated;
- (2) If essential proteins are involved, interaction outcomes may not be clear as the proteins are still present and may bind to other subunits, but just not functional;
- (3) If a protein's function, when impaired, can lead to growth defects, then the colony fluorescence measurement is also affected.

While the usage of this method is limited, it may still provide new and unexpected insights on complex subunits relationship.

CHAPTER 6: Understanding the difference in attenuation behaviour of mitochondrial proteins

In Chapter 2 and 4 on the overexpression screen results, I have noted that there is a significant depletion of mitochondrial proteins from the pool of attenuated proteins in the tFT-MoBY1 and tFT-MoBY2 screens. However, in the tFT-mORF screen, there is no longer a depletion of the mitochondrial proteins from the attenuated hits. The overlap of mitochondrial proteins that are attenuated upon overexpression within the three screens is shown in Figure 27A. I can see that as the overexpression level increases, the number of attenuated proteins increases as well (Figure 27B). Amusingly, in the tFT-MoBY2 screen, the mitochondrial ribosomal counterparts were found to not be attenuated in contrast to its cytosolic counterpart (Figure 27C). Even with the much stronger overexpression in the tFT-mORF screen, only a handful of mitochondrial ribosomal subunits were attenuated (Figure 27D). There is so far nothing known yet about the behaviour of mitochondrial ribosomal proteins upon overexpression.

I was then interested to try and understand why is there a difference in attenuation behaviour between cytosolic and mitochondrial proteins upon overexpression. In this section, I set out to answer several questions using mainly the ribosomal proteins because there are cytosolic and mitochondrial counterparts for most ribosomal proteins. The questions that I would try and answer are:

- (1) Could a difference in overexpression level between the mitochondrial and cytosolic ribosomal proteins lead to difference in attenuation behaviour?
- (2) Are the mitochondrial proteins being imported correctly into the mitochondria?
- (3) Could placing strains at more challenging conditions lead to attenuation of mitochondrial proteins?

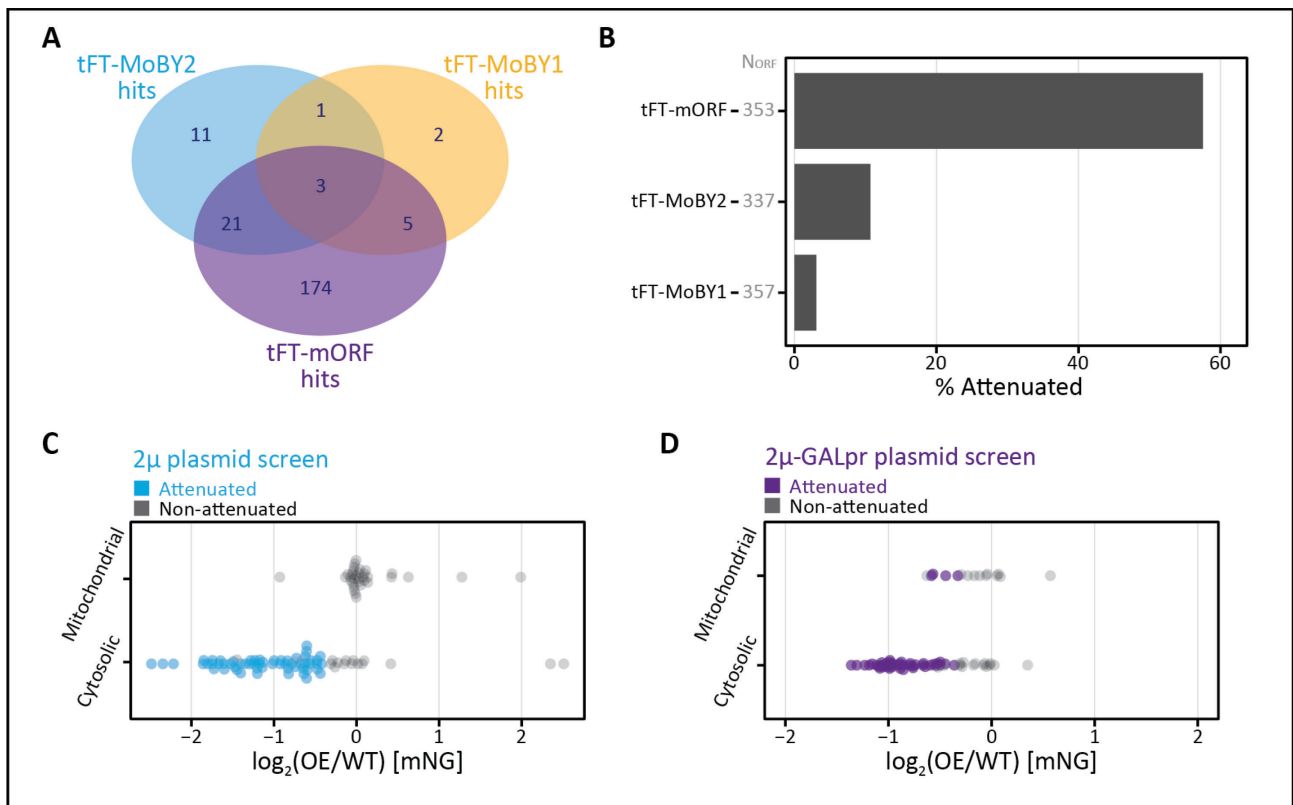


Figure 27: Mitochondrial proteins are resistant to attenuation upon overexpression and mitochondrial ribosomal subunits are not as readily attenuated as their cytosolic counterparts. (A) Venn diagram of overlap of attenuated mitochondrial proteins between tFT-MoBY1, tFT-MoBY2, and tFT-mORF screens. (B) Percentage of attenuated ORFs in the three screens described in A. The total number of mitochondrial proteins tested are shown in gray on the y-axis. (C & D) Distribution of \log_2 fold change of mNG levels upon overexpression of mitochondrial and cytosolic ribosomal subunits in the tFT-MoBY2 screen (C) or tFT-mORF screen (D).

6.1 Overexpression levels of cytosolic and mitochondrial ribosomes are similar

In Chapter 2 Section 2.2, I have shown that the plasmid copy number for the tFT-MOBY2 screen ranges from 2-7 copies and as seen in Figure 27A, increasing the level of overexpression leads to more mitochondrial proteins being attenuated. The hypothesis would be that mitochondrial ribosomal proteins are not being overexpressed as highly as their cytosolic counterparts and so this difference in overexpression levels is the reason for the lack of attenuated mitochondrial proteins.

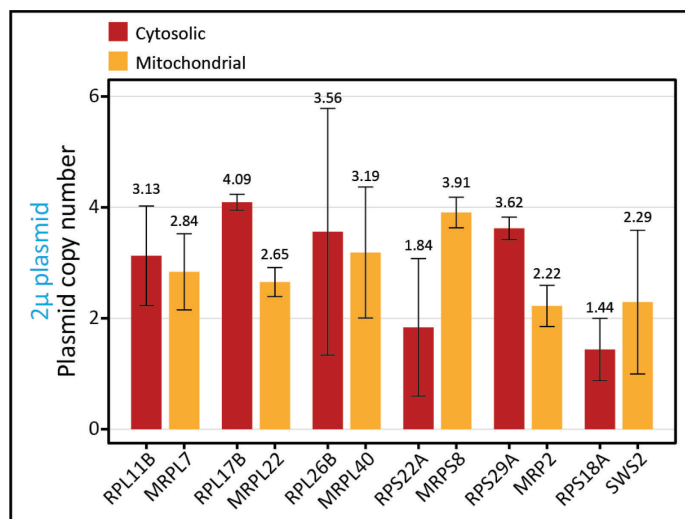


Figure 28: Plasmid copy number comparison between cytosolic and mitochondrial ribosomal subunits.

Plasmid copy numbers of several ribosomal subunits with the cytosolic counterpart in red and mitochondrial counterpart in orange. Two biological replicates were done for each sample. Error bars are the standard deviation and values above the bar is the mean.

So, I quantified the plasmid copy number for different strains carrying the overexpression plasmid of ribosomal proteins through qPCR. I can see that there is no difference in plasmid copy number between the two types of ribosomal proteins, both, when comparing all cytosolic against all mitochondrial or when comparing each mitochondrial ribosomal subunit with its direct cytosolic counterpart (Figure 28). Since I have also already shown that plasmid copy number scale fairly linearly with the fold increase in gene expression level (Figure 11C), I have no reason to believe that the level of overexpression between the cytosolic and mitochondrial ribosomal subunits are vastly different. Based on this set of results, I would argue that the lack of attenuation between the cytosolic and mitochondrial ribosomal subunits in the tFT-MOBY2 screen is not due to different overexpression levels. The results also show that with the same level of overexpression, mitochondrial proteins are more resistant to being attenuated upon overexpression, and that a much stronger level of overexpression is needed to push mitochondrial proteins to be attenuated.

6.2 tFT-tagged mitochondrial proteins localize correctly to the mitochondria with exceptions

It has been shown in literature that mitochondrial proteins are degraded upon import block (Shakya *et al.*, 2021), but through my screen, I am comparing two tFT-tagged strains to each other, one with an overexpression plasmid carrying the untagged ORF and one with a control plasmid. If the tFT-tagged proteins were having issues with import into the mitochondria, I probably would not detect a difference in levels since in both conditions the tFT-tagged proteins are already being degraded. So, I would first like to check and validate that the tFT-tagged mitochondrial proteins are localizing correctly at the mitochondria through fluorescence microscopy. I imaged the strains carrying an endogenously tFT-tagged ORF with either a control plasmid or a MoBY2 overexpression plasmid with the corresponding untagged ORF (Figure 29A).

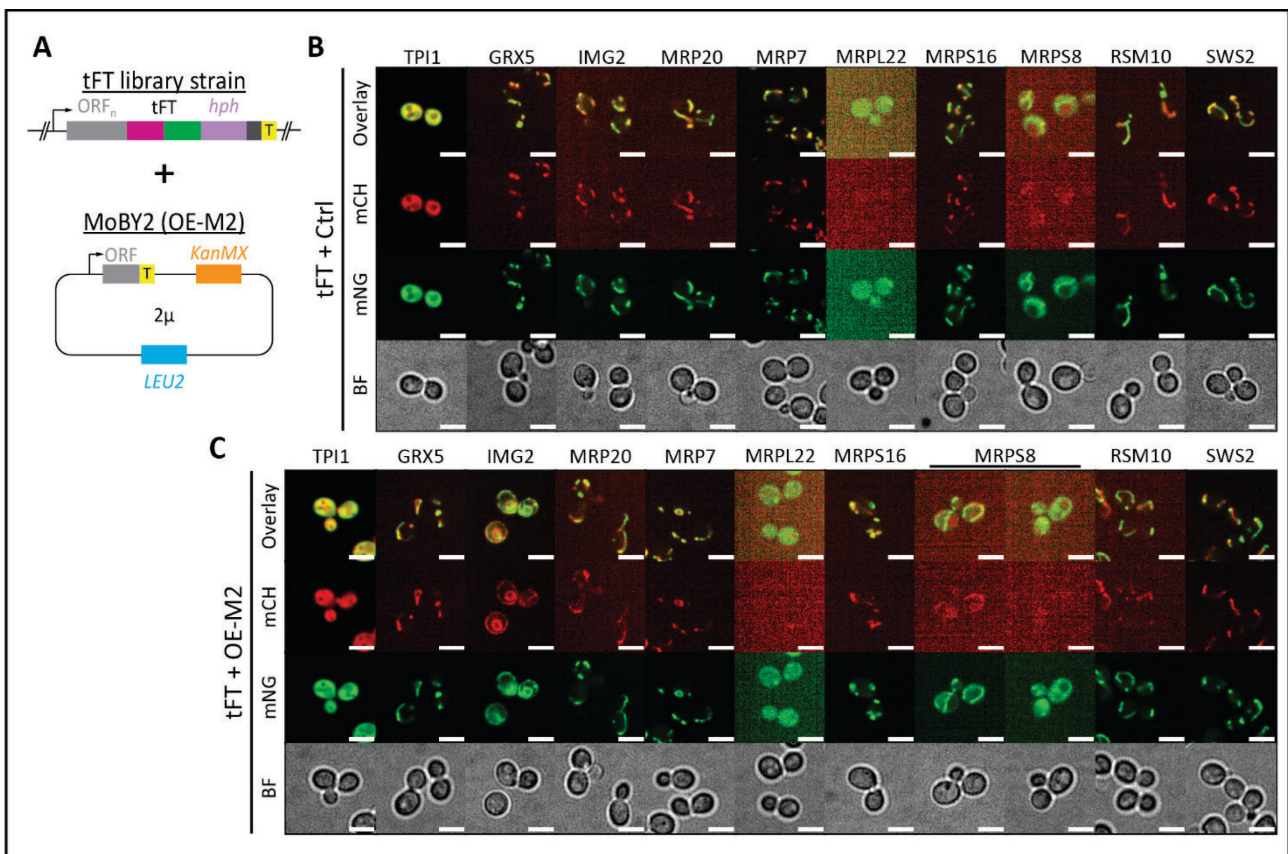


Figure 29: Most tFT-tagged mitochondrial ribosomal proteins localize correctly to the mitochondria and some mislocalized upon overexpression.

(A) Setup of the strain imaged where the endogenous ORF is tagged with the mCH-mNG tFT tag and the overexpression plasmid (OE-M2) is MoBY2 like. As a control plasmid for the non-overexpression condition, the MoBY2 plasmid with YGR045C ORF is used (Ctrl). (B & C) Images of strains with endogenous tFT-tagged ORF and expressing either the (B) control plasmid (Ctrl) or (C) overexpression plasmid (OE-M2). Images are individually adjusted in ImageJ on the intensity to show the best/clearest image. The images should not be compared in terms of intensity.

Firstly, the results show that the majority of tFT-tagged proteins localize correctly to the mitochondria with Mrpl22 being the exception (Figure 29B). The Mrpl22 signal seems to be mainly cytosolic. I also noted that while the fluorescence signals were mainly mitochondrial, there were also some diffuse signal within the cytosol as well. This diffuse cytosolic signal could indicate that there is already some mislocalization due to import inefficiency. Upon overexpression with the MoBY2 plasmids, there is no clear change in localization for all proteins except *Img2* and *Mrps8* (Figure 29C). Intriguingly, *Img2*-tFT mislocalizes to the ER upon overexpression with the untagged copy of *Img2* and *Mrps8*-tFT has some cells that are mislocalizing fully to the cytosol (Figure 29C). The observation of *Img2*-tFT and *Mrps8*-tFT mislocalising upon overexpression with an untagged copy raises the question on whether the cell is differentiating between the tFT-tagged endogenous copy and the untagged exogenous copy and is just selectively degrading one or the other. This observation on the change in localization, however, is not a common occurrence.

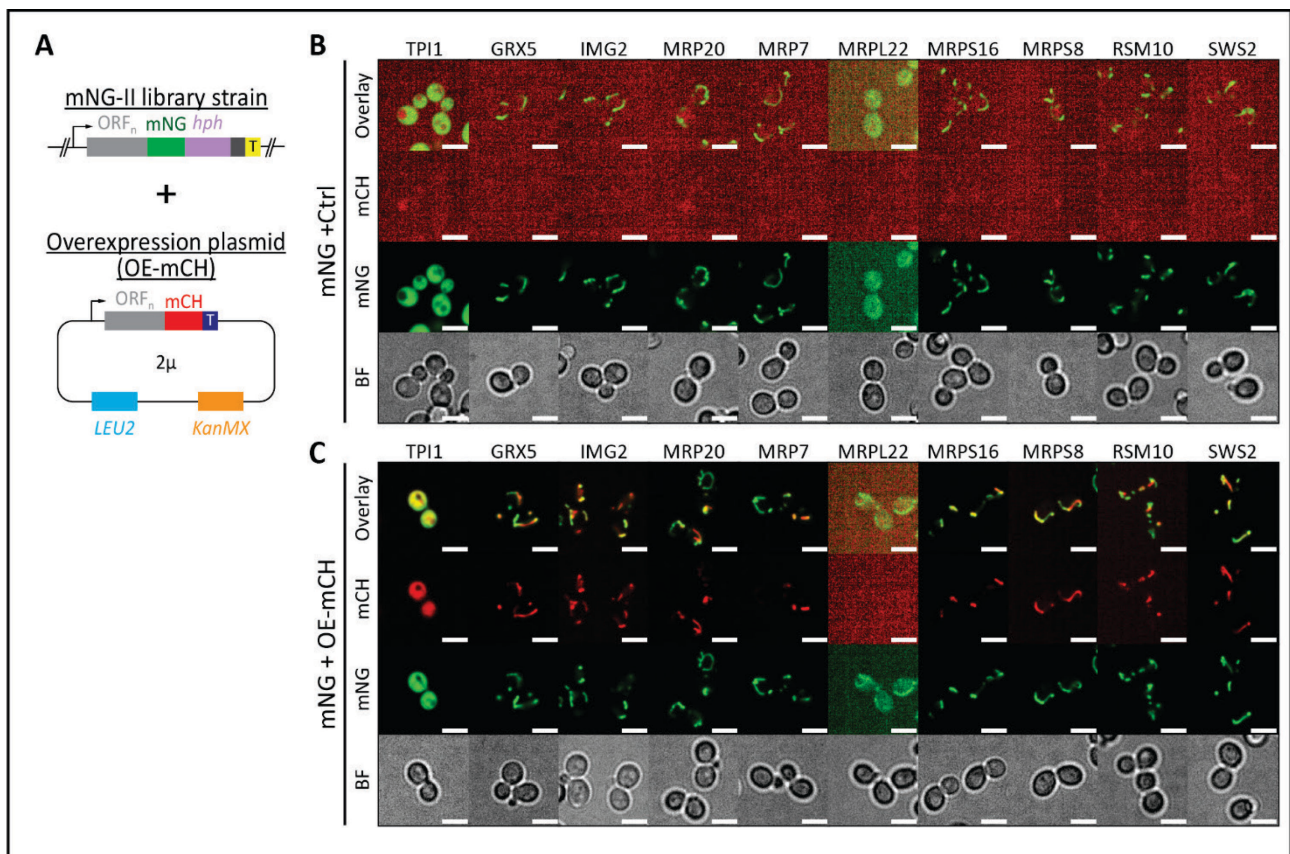


Figure 30: Mitochondrial ribosomal protein localization with mNG-tagged strains.

(A) Setup of the strain imaged where the endogenous ORF is tagged with mNG and the overexpression plasmid (OE-mCH) is MoBY2 like but the ORF is tagged with mCherry. As a control plasmid for the non-overexpression condition, the MoBY2 plasmid with YGR045C ORF is used (Ctrl). (B & C) Images of strains with endogenous mNG-tagged ORF and expressing either the (B) control plasmid (Ctrl) or (C) overexpression plasmid with ORF-mCH (OE-mCH). Images are individually adjusted in ImageJ on the intensity to show the best/clearest image. The images should not be compared in terms of intensity.

Next, I wanted to confirm also the localization of the exogenous mitochondrial proteins, the proteins that are expressed from the plasmid. For this experiment, I constructed the overexpression plasmid by tagging the ORF with mCherry. I then imaged the strains carrying an endogenously mNG-tagged ORF with either a control plasmid or an overexpression plasmid with the corresponding ORF tagged with mCherry (Figure 30A).

Using a smaller tag (mNG only vs mCH-mNG tFT), I noted that there are no obvious differences seen in localization for most proteins (Figure 30B). However, the localization of Mrps8-mNG and Mrpl22-mNG does appear more mitochondrial. Could it be possible that the tag affects some proteins more than others? Regardless, I see no change in localization nor attenuation upon overexpression in the mNG endogenous tag (Figure 30C). Bizarrely, *Img2*-mNG did not mislocalize to the ER upon overexpression with *Img2*-mCH as did *Img2*-tFT with overexpression of untagged *Img2* (Figure 30C).

I then decided to take a further look into the *Img2* phenotype. I next introduced the *Img2*-mCH overexpression plasmid into the *Img2*-tFT strain to determine where the overexpression plasmid is localising in this strain. Surprisingly, introducing the mCH-tagged ORF on a plasmid led to the localization of the endogenous protein into the mitochondria instead of the ER and the exogenous protein also localized to the mitochondria (Figure 31A). Based on these results, I hypothesize that, for *Img2*, the cell makes do with the available protein when there is no choice but prefers the untagged version of the protein. I think that *Img2*-tFT is poorly imported into the mitochondria, and upon overexpression, the mitochondria imports the untagged version of *Img2* instead of *Img2*-tFT. This mislocalization is not seen in the *Img2*-mNG strain with 2 μ -*IMG2*-mCH plasmid because both the endogenous and exogenous versions have a tag and so the cell does not differentiate between it and imports both versions. I would expect that if I introduced the overexpression of the untagged version of *Img2* into the *Img2*-mNG strain, it would also cause mislocalization of the tagged version into the ER.

In the experiment to look further into the *Img2* phenotype, I also introduced the ORF-mCH plasmid into the other ORF-tFT strains to confirm the localization of the exogenous copy in the tFT background. I found no difference in localization of the tFT-tagged ORF between the overexpression with untagged or mCherry tagged ORF for the same set of ORFs as above with exceptions for *Img2*-tFT (described above) and Mrpl22-tFT. The exogenous mCH-tagged copy of the protein localized similarly as the endogenous tFT-tagged copy. For Mrpl22, introducing the Mrpl22-mCH overexpression plasmid into the Mrpl22-tFT strain lead to the exogenous copy localizing to the correct localization but the endogenous tagged strain remains mislocalized into the cytosol (Figure

31B). The observation here is odd because the same Mrpl22-mCH plasmid was introduced into the endogenous Mrpl22-mNG strain but none of the exogenous or endogenous tagged proteins localized to the mitochondria (Figure 31B). Firstly, it is clear that tagged Mrpl22 is very poorly imported into the mitochondria in comparison to the other ORFs tested and regardless of type of tag. In the Mrpl22-mNG strain with 2 μ -MRPL22-mCH plasmid, both copies are tagged, and so both are equally poorly imported. However, in the Mrpl22-tFT strain with 2 μ -MRPL22-mCH plasmid, one tag is larger than the other, so possibly the protein with the smaller tag is preferred for import.

Besides visually annotating the localization of the proteins, I quantified the fluorescence levels in the microscopy images. I could show that in general there was no attenuation of the mNG levels upon overexpression regardless of whether it was endogenously tagged with the tFT or mNG only (Figure 31C & 31D). In Figure 31C, the mislocalization of Img2-tFT upon overexpression lead to a significant increase in mNG signal, and although there is a mild but significant accumulation for Mrpl22-tFT, the results may not be accurate since majority of the protein is mislocalizing into the cytosol.

Then, looking at the comparison of ORF-tFT with 2u-ORF-mCH against ORF-tFT with 2u-YGR045C or ORF-mNG with 2u-ORF-mCH against ORF-mCHmNG with 2u-YGR045C, I can see that the level of protein overexpression (mCH signal) is around 4-fold (Figure 31E and 31F). Since the previous quantification of plasmid copy number of these strains were around 2.85-fold average (Figure 28), the 4-fold protein overexpression shows that there does not seem to be selective attenuation of the exogenous protein.

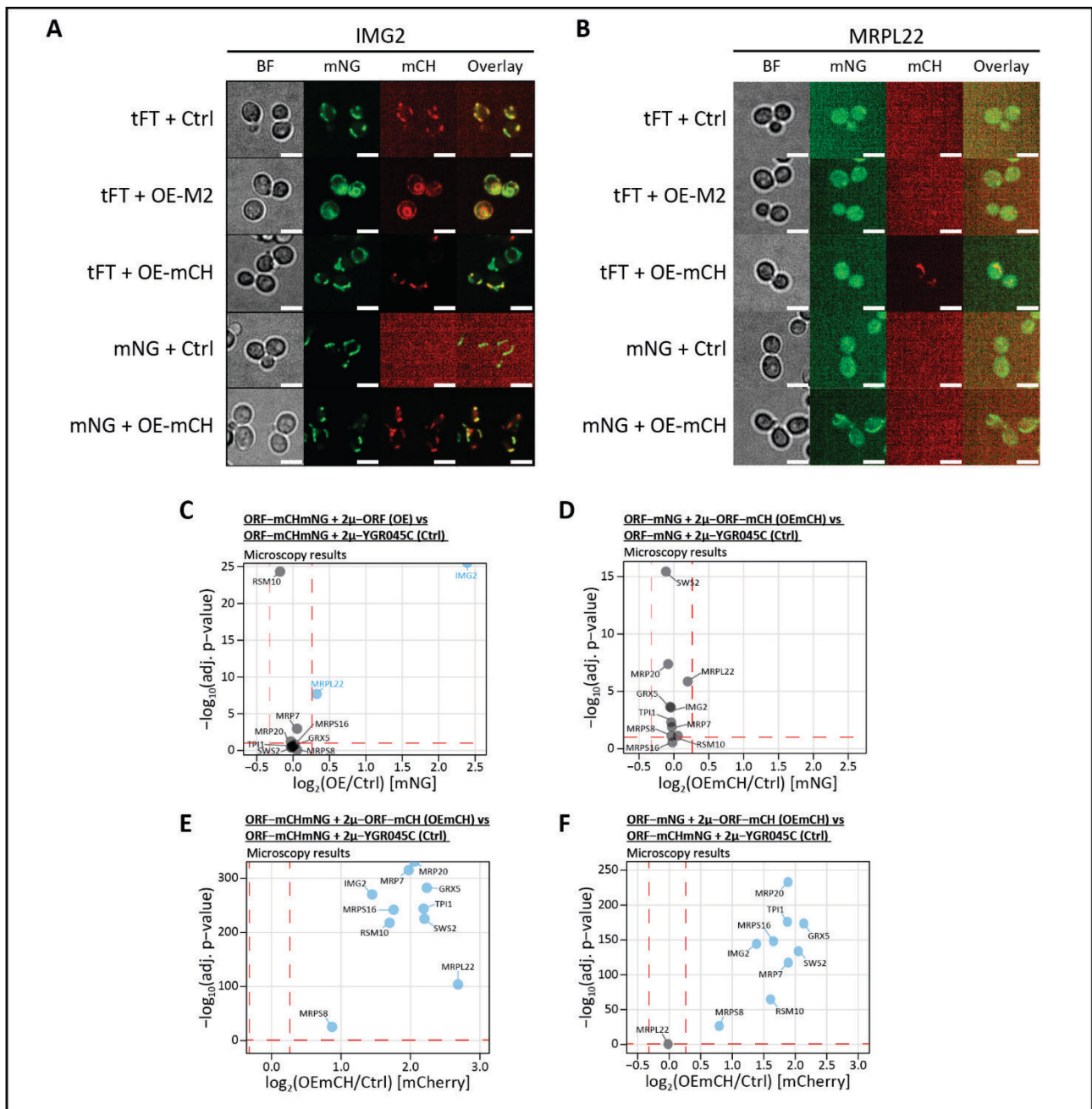


Figure 31: Img2 and Mrpl22 show difference in import behaviour when untagged or smaller tagged options exist but overall there is no preferential attenuation of the exogenous protein. (A & B) Localization of Img2 (A) and Mrpl22 (B) tagged endogenously with tFT or mNG, and carrying overexpression plasmids with an untagged ORF (OE-M2), mCH-tagged ORF (OE-mCH), or the control plasmid with YGR045C ORF (Ctrl). Images are individually adjusted in ImageJ on the intensity to show the best/clearest image. The images should not be compared in terms of intensity. (C-D) Volcano plots on the \log_2 fold change of mNG levels as quantified from microscopy images. Comparisons are done between ORF-tFT + 2 μ -ORF against ORF-tFT + 2 μ -YGR045C (C) and ORF-mNG + 2 μ -ORF-mCH against ORF-mNG + 2 μ -YGR045C (D). (E-F) Volcano plots on the \log_2 fold change of mCherry levels as quantified from microscopy images. Comparisons are done between ORF-tFT + 2 μ -ORF-mCH against ORF-tFT + 2 μ -YGR045C (E) and ORF-mNG + 2 μ -ORF-mCH against ORF-tFT + 2 μ -YGR045C (F).

6.3 Growth on alternate carbon sources and mild heat stress does not lead to increased attenuation of mitochondrial proteins

Increasing the overexpression levels from the 2 μ with endogenous promoter to 2 μ with GAL promoter led to an increase in number of mitochondrial proteins that were attenuated (Figure 27A and 27B). The observation suggests that mitochondrial proteins are more tolerant to overexpression but still has a limit to which it tolerates overexpression. I wondered then if by placing the strains on alternate carbon sources, where the function of the mitochondria becomes more important, would I see more attenuation upon overexpression with the MoBY2 plasmids? In addition to the alternate carbon source, I also decided to challenge the cells by growing it at a higher temperature to see if the mild heat stress would lead to increased attenuation. Essentially, would the mitochondria reduce its toleration of excess proteins and increase its stringency on quality control of mitochondrial proteins in stress conditions?

The attenuation behaviour of mitochondrial proteins upon overexpression did not change when grown on other carbon sources such as ethanol, glycerol, and acetate (Figure 32). Even when the cells were grown at 37°C in combination with the alternate carbon source, the attenuation behaviour of the mitochondrial proteins did not change. I also noted that all strains did not grow well on glycerol media at 30°C and so no colony fluorescence measurements were obtained.

The results here solidifies further that the mitochondria is very tolerant to overexpression of mitochondrial proteins, even when the cell experiences mild heat stress or needs to use non-fermentative carbon sources. Only through very strong overexpression can there be mitochondrial protein attenuation, which argues that the mechanisms in place that ensure mitochondrial proteins are properly folded, inserted, interacting with the right partners, localized correctly within the mitochondria, and maintains its functioning status, are much more robust than the rest of the cell.

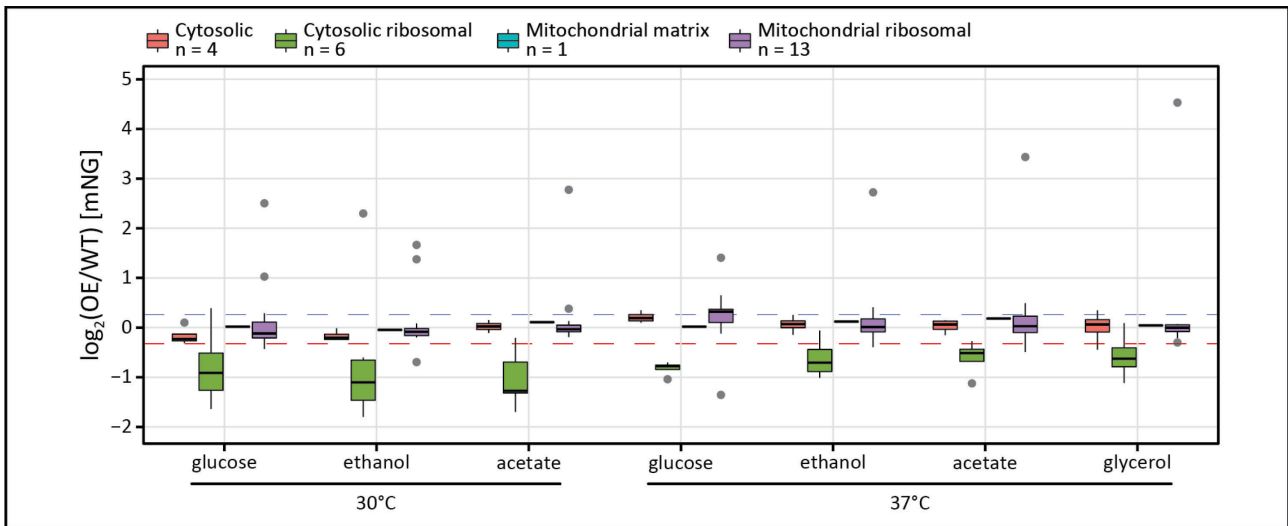


Figure 32: Attenuation behaviour of mitochondrial proteins remain unchanged when grown on alternate carbon sources and at higher temperatures.

Distribution of the log₂ fold change of mNG levels for the different categories of ORFs tested grown on agar with different carbon sources and under different incubation temperatures. Threshold on the log₂ fold change of mNG levels for calling attenuated proteins are shown on the plot as red dotted line $\geq 20\%$ decrease and gray dotted line for $\geq 20\%$ increase.

CHAPTER 7: Searching the proteome for dosage compensation genes

In this chapter, I will describe the experiment carried out to identify dosage compensation genes within the yeast proteome using the tFT library and the Yeast Knockout Collection (hereafter referred to as YKO library), and the problem that occurred which prevented the final analysis of the data obtained. Dosage compensation in yeast has been shown to occur at the post-translational level and is a mechanism to buffer against genetic perturbations that may occur (Ishikawa *et al.*, 2017). Some studies have been conducted to identify various dosage compensated genes but so far no study has been done on the whole proteome, only smaller subsets (Springer, Weissman and Kirschner, 2010; Ishikawa *et al.*, 2017). I therefore set out to use the tFT library I constructed with the YKO library to search the proteome for dosage compensation genes.

The tFT library was crossed with the YKO library strains and diploids strains were selected, which I then measured the colony fluorescence at the diploid stage. I would compare the fluorescence level of the diploid strain that is ORF-tFT/orf Δ against the diploid strain that is ORF-tFT/ORF. If the ORF tested is a dosage compensated ORF, then I would expect the mNG fluorescence level in the ORF-tFT/orf Δ strain to be double that of the ORF-tFT/ORF strain. Additionally, I took the diploid strains and continued with the SGA method to try and isolate haploids with both ORF-tFT and orf Δ , which would not be possible and the strain should not grow. In this way, I can also confirm that the ORF-tagged gene and the knockout gene corresponded or in other words, the knockout strain has the correct gene knocked-out.

After sporulation, I moved the strains onto haploid selection media which should end up with no growth for the ORF-tFT strains that were crossed with its corresponding orf Δ . Here, I found that for the majority of the screen, most of the haploid strains continued to grow (Figure 33A). This observation was confounding at first because it is unlikely that there are so many positions where the YKO library strain is wrong. After some checks were done, I found that some of the tFT strain still somehow maintained the SWAT donor plasmid used for the tFT library construction and this plasmid has a KanMX marker present. Thus, some of the tFT library strains could still grow on media containing G418 (Figure 33B & 33C). The YKO library was also constructed by using KanMX to knockout the ORFs. Due to this issue, I cannot be sure that the diploid strains that were selected and measured were actually diploids that are ORF-tFT/orf Δ or just the tFT strain with the leftover donor plasmid. I also cannot just stamp the tFT library onto G418 media to determine which well positions have an issue as there is a mix of cells with and without the leftover donor plasmid (Figure 33C). As shown in Figure 33C, the replica plate of a tFT strain from YPD+Hygromycin had some colonies that could grow on G418 but not many.

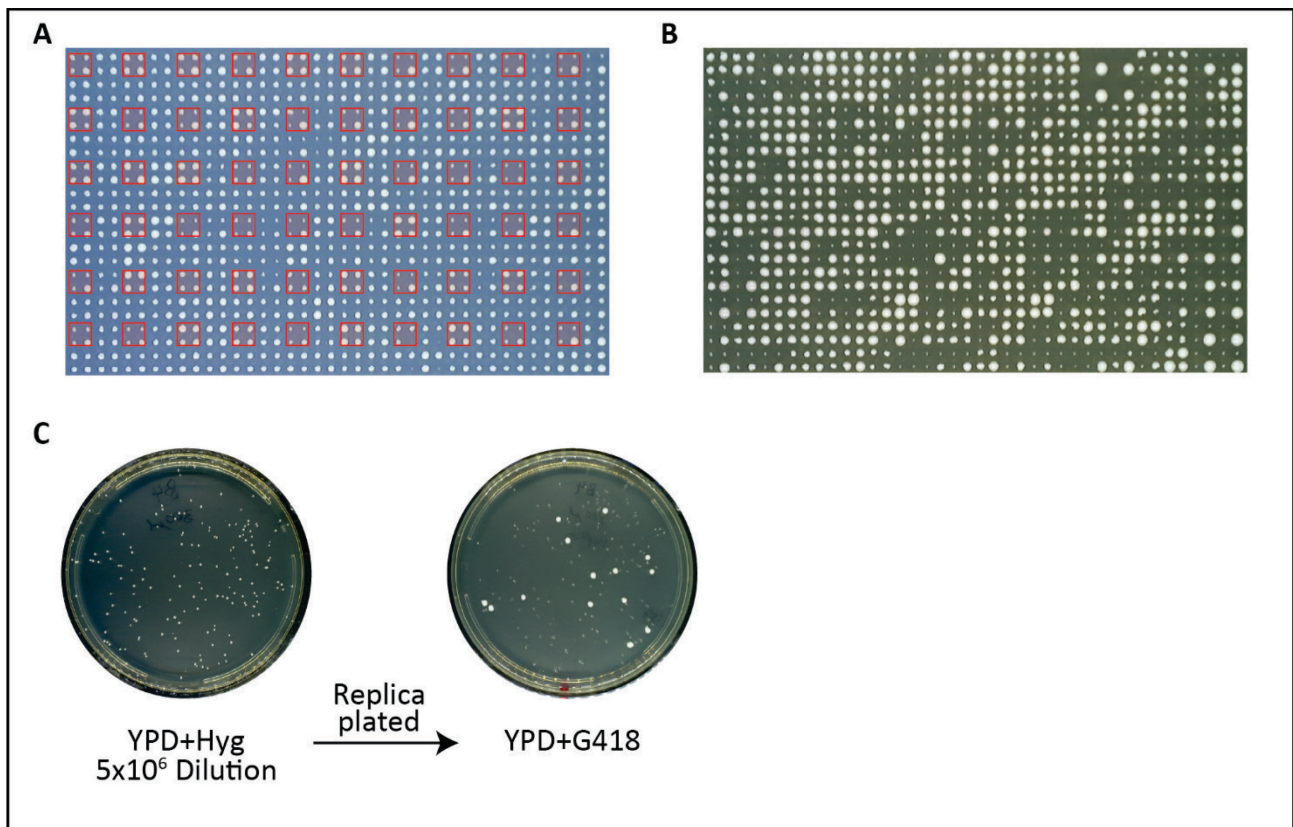


Figure 33: tFT library strains may still harbour the initial donor plasmid and so interferes with SGA selection when crossed with the YKO library.

(A) Photograph of an example plate from the tFT-YKO screen at the last haploid selection step. The red boxes mark the positions where the colonies should not be growing (except for the bottom right as that is a reference strain). (B) Photograph of an example plate of the tFT library stamped on YPD+G418 agar plate. The strains should not grow on G418 media. (C) A random colony of a tFT strain was picked and inoculated into 200ul of sterile PBS, diluted by 5×10^6 times and plated out onto YPD+Hyg. Once there was growth on the plate, the colonies were replica plated onto YPD+G418. The strain should not grow on G418 media.

Due to this issue, before I proceeded with more experiments on crossing the tFT library strains and YKO library strains for the screens in Chapter 5 Section 5.2, I also constructed another tFT library with the mCH-mNG tFT but the donor plasmid had NatMX on it instead of KanMX. Due to time constraints, the experiment was not repeated with the new tFT library. However, with this newer tFT library constructed with a NatMX marker on the donor plasmid, one should be able to perform the screen on the entire proteome to identify dosage compensated genes.

DISCUSSION

1.1 Studying protein mislocalization through protein overexpression

The results from my tFT overexpression screens of differing overexpression levels showed that the cell is quite tolerant to mislocalization through protein overexpression. I've shown that there are about 4.5% of the proteome (183 proteins attenuated in tFT-MoBY1) studied which are very intolerant to being in excess, as even a low level of overexpression lead to proteins attenuation. For some proteins (535 proteins that are tFT-mORF specific), only overexpression at a very high level (2 μ with GAL1 promoter) would lead to attenuation of its level. However, as I have also shown with the tFT-MoBY2 2 μ overexpression screen, there is also a set of 303 proteins that are destabilized but not attenuated. These proteins are separate and distinct from the 566 proteins that were attenuated upon overexpression but are likely still being regulated by the cell through other QC mechanisms. Taken together, there are 869 proteins (22.3%) that have a decrease in either abundance and/or stability upon overexpression.

The data from the overexpression screens were also able to recapitulate the findings in literature of proteins that are known to be attenuated upon overexpression, however, the attenuated proteins are distinctly separate from those attenuated in an aneuploid condition. I reason that this is due to the likelihood that in aneuploid cells, there is a much more global response due to the presence of many different excess proteins which may obscure the actual reason for the attenuation. The reasons may include, but are not limited to, cell cycle delays, metabolic disruptions, genomic instability, alterations to the transcriptome, and proteotoxic stress (Torres *et al.*, 2007; Williams *et al.*, 2008; Li *et al.*, 2010; Sheltzer *et al.*, 2011; Stinglele *et al.*, 2012; Zhu *et al.*, 2012, 2018; Oromendia and Amon, 2014).

Alternatively, rather than using overexpression to study protein mislocalization, one can modify signal sequences of proteins to mistarget proteins to different compartments and directly cause mislocalization. Sequences such as the mitochondrial targeting signal (MTS), nuclear localization signal (NLS), ER retention signal are well studied and described in the literature (Pelham, 1990; Omura, 1998; Kunze and Berger, 2015; Lu *et al.*, 2021). Hypothetically, one could tag all proteins with the NLS or ER retention signal and determine its fate upon mistargeting to those compartments. However, in this case, studies will have to be done per sub-cellular compartment and also proteins are being unnaturally mislocalized to those compartments which may not happen in a physiologically relevant context. Besides that, it could also happen that nothing will affect these MLPs in the target

compartment as there may be a lack of the correct QC machinery that recognizes or removes these MLPs.

All in all, studying MLPs through individual protein overexpression provides a more unbiased manner to screen the proteome as I am not purposefully mistargeting the proteins but rather allow them to mislocalize as they would in a regular cell when there is excess. Through this method, I would also highlight individual proteins that are more likely to be a problem than others when its level is dysregulated in diseases and aging. Therefore, my approach on individual protein overexpression provides a much cleaner readout on what happens to a single protein when it is overexpressed and mislocalizes.

Then, the follow up question in using the set of attenuated proteins as potential substrates for QC systems of MLPs is to ask what are the QC systems involved in removal of MLPs? What is the contribution of the UPS and autophagy pathways in removing these proteins? Which components of the pathways are involved?

1.2 Properties of MLPs in determining attenuation behaviour

Based on the analysis and comparisons of the proteins that were found to be attenuated, a few clear characteristics and properties stood out. These attenuated proteins are enriched with essential proteins, have a high synthesis rate, are relatively very stable with long half-lives, and are enriched with subunits of protein complexes. The observation here on the enrichment of subunits of complexes tracks with what has been reported in literature that the cell actively regulates and controls the stoichiometry of the subunits of protein complexes (Dephoure *et al.*, 2014; Ishikawa *et al.*, 2017; Brennan *et al.*, 2019; Taggart *et al.*, 2020). Furthermore, the enrichment of essential proteins and subunits of complexes together also makes sense as many of the cell's essential functions are carried out by multi-subunit complexes (Dezso, Oltvai and Barabási, 2003; Hart, Lee and Marcotte, 2007; Hart *et al.*, 2014; Meldal *et al.*, 2021). Since the attenuated proteins are produced rapidly and remains longer in the cell, I argue that my results show that these attenuated proteins, in general, are much more tightly regulated due to its importance in cell survival as well as its probability to cause greater harm when left unchecked.

In a per-complex setting, these two properties (synthesis rates and half-lives) become less important while other factors take a larger role. These factors include presence of chaperones or assembly factors that are needed to maintain the stability of subunits until they assemble, assembly factors that determine the efficiency of assembly, and also the assembly sequence of a complex (as

seen for the results of the EMC and V-ATPase subunits in Chapter 5 Section 5.2). The tFT-MoBY2 overexpression screen results also show an interesting observation, in that not all subunits within the same complex are equally attenuated. Taken together, the need for chaperones to maintain the stability and also promote complex formation could be a strong reason why some subunits in a complex were attenuated while others are not. It is interesting to postulate that one could detect novel assembly factors by screening for proteins that when overexpressed can reverse the attenuation behaviour of an attenuated overexpressed subunit. The benefit of overexpression in comparison to knockouts is that there would be less of a chance of causing growth defects, especially if the assembly factor is also needed for other proteins or functions.

Additionally, in some cases, the likelihood for a mislocalized subunit to be attenuated depends more on its direct interaction partners in intermediate sub-complexes rather than the presence of the whole complex. For example, the dependency of Emc2 on Emc5 for its stability but not vice versa. Interestingly, the human EMC2 has been shown to bind to WNK1 for its stability when in the cytosol and the human EMC8 cap for stability in the complex, but there are no homologs of these two proteins in yeast (Pleiner *et al.*, 2021). Thus, maybe yeast Emc5 is covering the role of WNK1 and/or EMC8 in yeast. The results from the screens conducted with the subunits of complexes and its partners hint that these attenuated proteins were attenuated because they may have exposed surfaces when not bound to its interaction partner(s) or chaperones since this is a way for MLPs to be recognized (Shemorry, Hwang and Varshavsky, 2013; Yanagitani, Juskiewicz and Hegde, 2017; Lin *et al.*, 2018; Inglis *et al.*, 2020). To this end, I did look into whether the N-termini of the attenuated subunits could play a large role in recognition and destabilization but the results showed that the attenuated subunits were not enriched with unstable N-termini. Results from Ishikawa, Ishihara and Moriya (2020) also agree that the Ac/N-degron pathway only plays a partial role in control of protein levels for subunits of complexes although their study was focused only on the RNase P/MRP subunits. At the C-terminal end, disorderness was significantly higher for attenuated proteins although the difference is not convincing enough to state that it plays a major role, but maybe this role was diminished as the proteins were tagged at the C-terminal with the tFT tag.

One avenue that was not explored, and is outside the scope of this thesis, was on the possibility of co-translational and post-translational assembly of the complexes also playing a role in determining the attenuation behaviour of the subunits. Co-translational assembly is described as the engagement of a mature subunit to the nascent chain of the partner subunit that is being synthesized or two nascent chains that engage each other during synthesis, while post-translational assembly is the formation of the complex after the subunits are already fully formed and folded (Gloge *et al.*, 2014; Wells,

Bergendahl and Marsh, 2015; Williams and Dichtl, 2018). Several different subunits have been described to be co-translationally assembled into their complexes in the recent years and more will likely be found with advances in ribosomal profiling (Kassem, Villanyi and Collart, 2017; Shiber *et al.*, 2018; Panasenko *et al.*, 2019; Seidel *et al.*, 2022). Through co-translational assembly, the protein that is being engaged while synthesized would already have a partner subunit to interact with immediately upon release from the ribosome which brings about the benefit of reducing promiscuous interactions, aggregation, and regulation of complex localization and stoichiometry (Schwarz and Beck, 2019). In terms of my overexpression screens, are the attenuated subunits upon overexpression mainly assembled co-translationally? I would expect that these co-translationally assembled subunits, when produced in excess, may not have enough binding partners and so the protein is then released and then maybe targeted for degradation. Alternatively, the cell could slow down translation through ribosomal pausing (e.g. Rpt1 and Rpt2 in Panasenko *et al.* (2019)) which may lead to ribosomal collisions and then triggering the RQC pathway. To answer this question, further studies would need to be done on more subunits of complexes to determine those that are co-translationally assembled.

1.3 Multiple modes of regulation and attenuation of MLPs upon overexpression

When taking a deeper look at the proteins that were attenuated and/or destabilized, I categorized the proteins into three categories with different attenuation behaviours and proposed different methods for regulation when being in excess. First, the proteins that were destabilized but not attenuated are likely to be regulated at the maturation and folding stage, as these are proteins that the cell would not want to directly degrade due to its already low levels, slow synthesis rates, and short half-life. Since this group of proteins is also enriched in subunits of complexes, perhaps having a slowed down maturation and folding allows for more of its partner subunits to be produced and ready to engage. It would also be tempting to propose that the proteins within this group are the subunits that are co-translationally assembled and so upon overexpression do not have enough binding partners, but instead of degradation, they are sequestered into temporary compartments while awaiting partners to assemble and fully mature.

The next category are the proteins that were attenuated but not destabilized and so are likely to be either pre-translationally regulated or the degradation is biased towards newly synthesized proteins rather than mature, functional, and assembled proteins. Interestingly, these two methods of regulation may not be mutually exclusive as studies on ribosomal subunits have shown that there is a biased degradation of excess newly synthesized subunits and also a level of autoregulation of some

of the subunits when in excess (Sung, Reitsma, *et al.*, 2016; Roy *et al.*, 2020). I have also shown, through my results, that ribosomal subunits that are both autoregulated and have biased degradation tend to be more strongly attenuated upon overexpression, indicating that the two pathways may work together for a stronger effect. Further studies to see if it is mainly biased degradation that plays a major role can be performed as it has already been shown in mammalian cells (Mcshane *et al.*, 2016).

Lastly, the proteins that are attenuated and destabilized are likely to be regulated via protein degradation. Work by another PhD student in the lab using a subset of these proteins has shown that inhibition of proteasomal degradation leads to accumulation of these attenuated proteins. Additionally, this group of proteins were also found to have lower overall disorder scores. Taken together, recognition of these MLPs may require specific substrate receptors, instead of depending on disordered regions, and that the proteins here are being degraded by the UPS.

Just as the cell has many pathways and mechanisms to handle various cellular stressors, it is just as logical to expect that the cell would have various different methods in handling MLPs that occur through overexpression. It is also quite likely that the cell uses many different methods to fine tune the protein levels within the cell to maintain protein homeostasis (Taggart *et al.*, 2020). While the question on how does the cell handle MLPs through overexpression is partially answered here through the different mechanisms proposed, the unanswered question is why do the remaining ~65-75% of proteins not get attenuated? Do these non-attenuated proteins have an even better mechanism that prevents the proteins from mislocalizing and causing issues within the cell? Or how do they escape the QC systems that recognize and capture MLPs?

1.4 Mitochondrial proteins are resistant to mislocalization by protein overexpression

Lastly, through my overexpression screens, I have found that mitochondrial proteins are relatively resistant to being attenuated upon overexpression. The initial assumption of the overexpression screens is that the cell should not differentiate between the tagged endogenous copy and the untagged exogenous copy. This assumption remains true as I have shown that for majority of the mitochondrial proteins I tested, the tagged copy of the protein does localize correctly to the mitochondria, even upon overexpression with an untagged copy, and that there is likely no biased degradation of either copy. Otherwise, the mislocalization of these mitochondrial proteins happen regardless of presence of the untagged version, as is the case for Mrpl22 and Mrps8. The results of mislocalization of Img2-tFT upon overexpression with the untagged version is an exception and not the rule. In this case I can only speculate that the cell is preferring to utilize the untagged exogenous

copy instead of the tFT-tagged copy because tagging the exogenous copy with a smaller mCherry tag already leads to the tFT-tagged copy localizing in the mitochondria again. Besides that, the lack of attenuation behaviour for mitochondrial proteins remains even when grown on non-fermentative media and at mild heat stress conditions. What protection mechanisms are present in the cell for the mitochondrial proteins to make it so resilient to excess proteins?

1.5 Concluding remarks and outlooks

In conclusion, the experiments carried out in this study provide a rich resource of proteins that can be used for further studies on the different QC systems and mechanisms that handle MLPs. From my results, I find that the cellular protein QC systems are very robust in handling excess proteins and preventing mislocalization, particularly for the mitochondria, but are also able to remove MLPs when it overwhelms the QC system. However, the mechanisms used to allow for such resilience to excess proteins for the mitochondria remains to be studied further.

Additionally, through this study, I have also proposed several different avenues that the cell uses to handle MLPs upon overexpression. Further studies can be conducted to better identify the exact QC mechanisms that are utilized by the cell for each MLP such as sequestration, autoregulation, biased degradation, and recognition and degradation by the UPS and/or autophagy. Through understanding the exact QC mechanisms, the causes of diseases at a molecular level can also be dissected and allow for work on targeted treatments.

Lastly, I have highlighted the reasons as to why some subunits are more likely to be attenuated in a complex than others (i.e. presence of chaperones, assembly factors, and assembly sequence), but there is still one interesting category that remains unexplored in this thesis, that is the co-translational assembly of complex subunits. In parallel, I have also proposed and tested a method to study protein complex assembly sequence and to dissect subunit pairs that are interdependent for stability. The information acquired can be used together with structural data to further understand subunit-subunit interactions and relationship within a complex. All in all, the work shown also contributes to the better understanding of subunits of complexes and its formation, which in turn is important in understanding the impact of mislocalized orphan proteins on protein homeostasis and diseases.

Yeast methods and plasmids

Yeast strains used in this study are listed in Supplementary Table 1. Yeast genome manipulations (gene deletions and tagging) were performed using standard procedures based on PCR targeting (Janke *et al.*, 2004). The gene deletions and tagging were confirmed by PCR. Plasmids used in this study are listed in Supplementary Table 2. Construction of new plasmids were done using NEBuilder HiFi DNA Assembly (New England Biolabs) in which the components are PCR amplified from existing plasmids, restriction digested from existing plasmids, or PCR amplified from genomic DNA, unless otherwise stated. PCR amplification of inserts were done using Velocity DNA polymerase (Bioline) and colony PCR for strain verifications were done with in-house HF-DNA polymerase from IMB protein production core facility.

Yeast strains were grown at 30°C in synthetic complete medium with 2% w/v glucose unless otherwise stated. For induction of genes under the GAL1 promoter, strains were grown with 2% w/v raffinose and 2% w/v galactose in replacement of glucose. Yeast array manipulations were done using the ROTOR pinning robot (Singer Instruments) in 96-, 384-, and 1536-colony format, depending on the step and requirements. The pinning robot was also used for yeast manipulations to move from liquid (well plates) to liquid (well plates), liquid to solid format and vice versa. Imaging of yeast agar plates were done using the PhenoBooth (Singer Instruments) and colony fluorescence measurements were done using a plate reader with a monochromator (TECAN), unless otherwise stated.

Comparison and characterization of the mCH-mNG tFT

The tFT donor plasmids were constructed using NEBuilder HiFi DNA Assembly where the inserts, mCH-sfGFP or mCH-mNG, were PCR amplified from pMaM17 and pMaM332 respectively, and the backbone used is BamHI and SpeI digested pMaM482. The tFT plasmids (pJJF001 and pJJF003) and pMaM482 (non-fluorescent control) were then transformed into the SWAT donor strain, yMaM639 to obtain the tFT donor strains yJJF0013 and yJJF0014, and non-fluorescent donor strain yJJF0015 respectively.

The full C-SWAT library was then crossed with the tFT donor strains in a 1536-array format on agar. The crossing and tag swapping was done by sequentially pinning the strains on the appropriate media according to the SGA method (Baryshnikova *et al.*, 2010) as follows:

- mating of C-SWAT and donor strains on YPD plates (10 g/l yeast extract (BD Biosciences), 20 g/l bacto peptone (BD Biosciences), 20 g/l glucose (Merck), 20 g/l bacto agar (BD Biosciences))
- selection of diploids on SC(MSG)-Ura + G418 plates (1.7 g/l yeast nitrogen base without amino acids and ammonium sulfate (BD Biosciences), 1 g/l monosodium glutamic acid (MSG) (Sigma-Aldrich), 2 g/l amino acid dropout mix SC(MSG)-Ura, G418 (200 mg/l, Invivogen), 20 g/l glucose, 20 g/l bacto agar)
- sporulation for 6 days on SPO plates (20 g/l potassium acetate (Sigma-Aldrich), 20 g/l bacto agar) at 23 °C
- selection of haploids, step 1, on SC(MSG)-His/Arg/Lys/Ura + canavanine/thialysine plates (50 mg/l canavanine (Sigma-Aldrich), 50 mg/l thialysine (Sigma-Aldrich))
- selection of haploids, step 2, on SC(MSG)-His/Arg/Lys/Ura + canavanine/thialysine/G418 plates (50 mg/l canavanine, 50 mg/l thialysine, 200 mg/l G418)
- selection of haploids, step 3, on SC(MSG)-His/Arg/Lys/Ura + canavanine/thialysine/G418/clonNAT plates (50 mg/l canavanine, 50 mg/l thialysine, 200 mg/l G-418, 100 mg/l clonNAT (Werner BioAgents))
- induction of I-SceI enzyme for tag swapping on SC-His Raf/Gal plates (6.7 g/l yeast nitrogen base without amino acids (BD Biosciences), 2 g/l amino acid dropout mix SC-His, 20 g/l galactose (Serva), 20 g/l raffinose (Sigma-Aldrich), 20 g/l agar) - done twice
- counter-selection against the acceptor module on SC-His + 5-FOA plates (20 g/l glucose, 1 g/l 5-FOA (Apollo Scientific, PC4054))
- recovery after counter-selection on SC(MSG)-His/Arg/Lys + canavanine/thialysine plates (50 mg/l canavanine, 50 mg/l thialysine)

The strains at the end of the process were used for flow cytometry and fluorescence measurements.

For checking the tagging efficiency via flow cytometry, 20 strains from each library, the two tFT libraries and the non-fluorescent control library, were taken and grown overnight in well plates at 30°C containing SC-His + Adenine media (250mg/L adenine (Sigma-Aldrich)). The strains are then diluted 1:200 in fresh media in a 96-well plate before being measured on the BD LSRFortessa

SORP flow cytometer with the high-throughput sampler attachment. Measurements were done with the laser BL488nm 530/30-A and YG561nm 610/20-A for the red and green signal respectively with minimum of 20,000 cells measured per sample. Settings on the high-throughput sampler were sample flow rate = 0.5 μ l/sec, sample volume = 15 μ l, mixing volume = 50 μ l, mixing speed = 180 μ l/sec, number of mixes = 3, wash volume = 400 μ l, and no BLR. Data obtained were then exported as .fcs files and imported in R for further analysis. The cells were gated on first on FSC-Area and SSC-Area for 'live cells', then on SSC-Area and SSC-Width for 'single cells', then lastly on the BL488 (green) and YG526 (red) channels for positive cells. Non-fluorescent cells were used as negative controls to determine the threshold for the red and green channel. The \log_{10} (mNG) levels of the strains were reported and percentage of cells above the threshold set based on fluorescence levels of the non-fluorescent cells were calculated for each strain.

For the fluorescence measurements, strains from the final step were pinned onto SC-His + Adenine plates. After 24 hours of growth, the plates were imaged on the PhenoBooth (Singer Instruments) and the colony fluorescence was measured on a plate reader. The colony fluorescence measurements were done using a plate reader (TECAN) with filters for the red channel (excitation wavelength of 580/20, emission wavelength of 625/35, 50% mirror) and green channel (excitation wavelength of 485/20, emission wavelength of 530/25, Dichroic 510 mirror). The optimal gain is first obtained from the first plate and then adjusted to be lower (~15 from optimal gain) to avoid overexposure and all plates are measured with this manual gain. The colony fluorescence measurement data and plate images were then imported into R (R Core Team, 2021) for data analysis. After removing dummy strains in the border of the plates, the fluorescence measurements were first corrected for across plate effects and within plate effects using the reference strains included in the plate. Then, the signal-to-noise (tFT against non-fluorescent control) ratio and background (non-fluorescent control) subtracted fluorescence values were calculated. Lastly, the red-to-green ratios were then calculated from the background subtracted values. The values were \log_{10} transformed when used for plotting.

Preparation of overexpression libraries

MoBY1 plasmids were extracted from the MoBY1 E. coli library (Dharmacon) which exists in a 96-well plate format. The E. coli strains in the 96-well plates were pinned onto LB Kan plates (100 μ g/ml kanamycin) using the ROTOR pinning robot and grown overnight at 37°C. The colonies on agar were then inoculated into 2ml deep well plates with LB Kan media and grown overnight at 37°C. The

plasmids were then extracted using the DirectPrep 96 Miniprep kit (Qiagen) as per the kit's protocol. Briefly, the overnight cultures in the deep well plates were harvested by centrifugation at 1,500g for 5 minutes. The supernatant in each well was removed and the pellets were re-suspended in the well plates with Buffer P1. Then, Buffer P2 followed by Buffer DP3 and isopropanol was added to each sample well; for each step the mixture is mixed thoroughly by inversion. The lysates were then pipetted into the wells of the DirectPrep 96 plate and drawn through by vacuum. Buffer PE was then added to each well and drawn through by vacuum. Lastly, the plasmids were eluted with 35ul of sterile double distilled water (ddH₂O) using vacuum suction into the 96-well plates provided by the kit. Plasmids from several wells were selected for restriction digests to confirm the identity of the plasmid and that the library annotation was correct. Once confirmed, the plasmids were used for transformation into BY4741.

For high-throughput transformation, 20µl of competent cells were placed into the wells of a 96 deep well plate. Plasmids (25ng minimum) were added into each well according to its original position in the library and allowed to incubate at room temperature for 15 minutes. Then, 120µl of LiPEG (40% w/v PEG3350 (Sigma-Aldrich), 1mM EDTA, 10mM Tris-HCl pH8.0, 100mM lithium acetate (Sigma-Aldrich)) were added into each well and incubated at room temperature for 15 minutes. The deep well plate is then sealed with an air permeable foil (Steinbrenner) and then placed into a water bath for heat shock at 42°C for 40 minutes. After the heat shock, the plate was placed on ice for 2 minutes and then centrifuged at 1,500 g for 3 minutes. The supernatant was removed and 1.5ml of YPD is added to each well for the cells to recover overnight at 30°C. After recovery, the plate was centrifuged as before and 1.3ml of supernatant was removed. The cell pellets were re-suspended in the remaining media and transferred into a normal 96-well plate. Using the ROTOR pinning robot, the cultures were pinned out onto SC-URA plates using the 7x7 program where each position of well-plate is pinned in a square grid of 7x7 in order to dilute the cells to get single colonies towards the end of the grid. The pinned plate was then incubated at 30°C for up to two days for colonies to form. Then, using the ROTOR pinning robot again, strains were pinned from the 7x7 grids into the standard 96-array format. The yeast strains carrying the MoBY1 plasmids were stored as the MoBY1 yeast library in glycerol for future use.

The MoBY1, MoBY2, and mORF yeast libraries were re-arrayed into the format needed for the overexpression screens. Using the library annotation files, the ORF positions were matched between the different overexpression libraries and the tFT library to generate the coordinates for a new layout to re-array the strains. For re-arraying, the yeast strains were first pinned onto appropriate agar plates for each library in a 384-array format. The libraries were then re-arrayed using the ROTOR

pinning robot with the Stinger attachment (Singer Instruments) which uses metal pins to pick a colony and transfer it to a new plate based on coordinates given for the source and target in the desired layout. These metal pins were washed and sterilized before re-use. The re-arrayed libraries were then stored in glycerol in 96-well plates for future use.

tFT-overexpression screens

The tFT donor plasmid was constructed using NEBuilder HI-FI DNA assembly where the insert, mCH-mNG, was PCR amplified from pMaM332 and the backbone used is BamHI and SpeI digested pMaM484. The donor strain yJFF0018 was constructed using the strain yMaM1205 by replacing the LEU2 with SpHIS5 (also known as HIS3MX6). The tFT donor strain yJFF0027 was then constructed by transforming the tFT plasmid pJFF004 into the donor strain yJFF0018. In parallel, the plasmid pMaM484 was also transformed into the yJFF0018 to have a non-fluorescent donor strain yJFF0028 when creating the tFT library. The tFT library for the overexpression screen was then constructed in the same procedure using yJFF0027 and yJFF0028 as before (see ‘Comparison and characterization of the mCH-mNG tFT’) except that recovery after counter-selection was done on SC(MSG)-His/Arg/Lys + canavanine/thialysine/ hygromycin (400mg/L hygromycin B Gold (Invivogen)) since it is using the type-II tag. The tagging efficiency for this tFT library construction was also quantified as before via flow cytometry and confirmed to be similar to the previous constructed library. Once confirmed, the tFT library was then stored in glycerol in 96-well plates for future use.

The tFT library was also re-arrayed into the format needed for the overexpression screen (as noted in ‘Preparation of overexpression libraries’). The re-arrayed tFT library and overexpression libraries were then pinned onto agar and expanded into final array format as described in Chapter 1 Section 1.3. The tFT library and overexpression libraries were then crossed in a 1536-array format on agar. The crossing was done by sequentially pinning the strains on the appropriate media according to the SGA method (Baryshnikova *et al.*, 2010) as follows:

- mating of tFT and overexpression library strains on YPD plates
- selection of diploids on SC(MSG)-Ura + hygromycin plates for tFT-MoBY1 and tFT-mORF, SC(MSG)-LEU + hygromycin plates for tFT-MoBY2 – done twice
- sporulation for 6-7 days on SPO plates at 23 °C
- selection of haploids, step 1, on SC(MSG)-His/Arg/Lys/Ura + canavanine/thialysine plates for tFT-MoBY1 and tFT-mORF and SC(MSG)-His/Arg/Lys/Leu + canavanine/thialysine plates for tFT-MoBY2

- selection of haploids, step 2, on SC(MSG)-His/Arg/Lys/Ura + canavanine/thialysine/hygromycin plates for tFT-MoBY1 and tFT-mORF and SC(MSG)-His/Arg/Lys/Leu + canavanine/thialysine/hygromycin plates for tFT-MoBY2 (400mg/L hygromycin B Gold (Invivogen))

The strains at the end of the process were used for fluorescence measurements.

For the fluorescence measurements, strains from the final step were pinned onto SC(MSG)-His/Arg/Lys/Ura + canavanine/thialysine/hygromycin + Adenine plates and SC(MSG)-His/Arg/Lys/Leu + canavanine/thialysine/hygromycin + Adenine plates for tFT-MoBY1 and tFT-MoBY2 respectively. For tFT-mORF, the final strains were pinned onto SC(MSG)-His/Arg/Lys/Ura + canavanine/thialysine/hygromycin + Adenine plates without glucose but with 2% w/v raffinose and 2% w/v galactose to induce overexpression. After 24 hours of growth, the plates were imaged on the PhenoBooth (Singer Instruments) and the colony fluorescence was measured on a plate reader. The colony fluorescence measurements were done using a plate reader (TECAN) with a monochromator for the red channel (excitation wavelength of 586/10, emission wavelength of 612/10, 50% mirror) and the green channel (excitation wavelength of 506/5, emission wavelength of 524/5, Dichroic 510 mirror). Before measuring the full set of plates, the optimal gain is first determined and then two manual gains are set, one much lower (~-15 from optimal gain) and one higher (~+5 from optimal gain) than the optimal gain. The colony fluorescence measurement data and plate images were then imported into R (R Core Team, 2021) for data analysis.

Firstly, any missing data due to overexposure from the high gain is recalculated from the low gain values using the median ratio of high to low gain across all plates measured. The plate images are segmented using an R package *gitter* (Wagih and Parts, 2014) to obtain colony sizes. Then, measurements from dummy strains in the border colonies are removed and all measurements of empty positions, based on the images, are replaced with NA. The fluorescence measurements were first corrected for across plate effects and within plate effects using the reference strains included in the plate. Then, the background (non-fluorescent control) subtracted fluorescence values and red-to-green ratios were calculated locally for each 4x4 grid. The log₂ fold change of the overexpression against native condition for the final fluorescence and ratio levels were then calculated and statistical analyses were performed to determine significance. The calculated p-values were then adjusted using the Benjamini-Hochberg correction.

tFT-knockout screen

The tFT library constructed for the overexpression screen was also used to cross with the Yeast Knockout Collection. The Yeast Knockout Collection was first re-arrayed into the same ORF layout as the tFT library.

The re-arrayed tFT library and knockout library were then pinned onto agar and expanded into final array format as described in Chapter 1 Section 1.3. The strain with YGR045C knockout was used as the native control for the knockout condition. The tFT library and knockout library were then crossed in a 1536-array format on agar. The crossing was done by sequentially pinning the strains on the appropriate media according to the SGA method (Baryshnikova *et al.*, 2010) as follows:

- mating of tFT and knockout library strains on YPD plates
- selection of diploids on SC(MSG) Complete + G418/hygromycin plates – done twice
- sporulation for 6-7 days on SPO plates at 23 °C
- selection of haploids, step 1, on SC(MSG)-His/Arg/Lys + canavanine/thialysine plates
- selection of haploids, step 2, on SC(MSG)-His/Arg/Lys/ + canavanine/thialysine/hygromycin/G418

The strains after the diploid selection process were used for fluorescence measurements. The strains at the end, second haploid selection step, should be largely not growing.

For the fluorescence measurements, diploid strains from the diploid selection step were pinned onto SC(MSG) Complete + G418/hygromycin + Adenine plates. After 24 hours of growth, the plates were imaged on the PhenoBooth (Singer Instruments) and the colony fluorescence was measured on the plate reader. Colony fluorescence measurements and data analysis were performed as before (see ‘tFT-overexpression screens’). The strains at the final haploid selection steps were imaged on the PhenoBooth (Singer Instruments) to confirm that the knocked out protein position is correct.

Subunits tFT-overexpression and tFT-knockout screens

For the subunit tFT-overexpression screens, MoBY2 plasmids for the selected subunits of the selected complexes were taken from the MoBY2 E. coli library. The plasmids were checked by restriction digest and Sanger sequencing to confirm that there are no mutations present in the ORF on the plasmids. For plasmids where the ORF has mutations, the corrected plasmids were constructed either via Q5 site-directed mutagenesis kit (NEB) or NEBuilder HiFi DNA Assembly. The plasmids were then transformed into BY4741. For the subunit tFT-knockout screens, the strains were streaked out

from the Yeast Knockout Collection and verified via colony PCR. Strains that were incorrect were constructed by using PCR targeting with standard S1/S2 primers (Janke *et al.*, 2004) to knockout the ORF of interest with a KanMX cassette in the BY4741 background. The tFT strains were streaked out from the mCH-mNG tFT library and were confirmed to not grow on G418 containing media (i.e. no remaining KanMX bearing plasmid).

The tFT, overexpression, and knockout strains were then arrayed into the format where each position had an ORF corresponding to either itself or a partner subunit from the same complex. The tFT strains were then crossed with both overexpression and knockout strains and using SGA method. The SGA crossings were the same as the sections above for the tFT-overexpression and tFT-knockout screens, except that for this subunit screen, the tFT-knockout cross was imaged and measured at the final haploid stage instead of diploid.

After the final recovery step, the colonies were pinned onto imaging plates containing adenine. After 24 hours of growth, the plates were imaged on the PhenoBooth (Singer Instruments) and the colony fluorescence was measured on the plate reader. Colony fluorescence measurements and data analysis were performed as before (see ‘tFT-overexpression screens’).

Plasmid copy number and gene expression level quantification

Yeast strains were inoculated into the appropriate selection media and grown overnight at 30°C. The overnight cultures were then diluted to OD₆₀₀ = 0.2 and grown again at 30°C until OD₆₀₀ reaches ~0.8. The cultures were then harvested for genomic DNA and RNA extraction.

For genomic DNA (gDNA) extraction, 2ml of culture was harvested by centrifugation at 19,000g for 3 minutes. The supernatant was then removed and the pellet was subjected to gDNA extraction using the Gentra Puregene Yeast/Bact. Kit (Qiagen). Briefly, the pellet was incubated with the lytic enzyme and then lysed. The proteins were then precipitated out and the supernatant transferred to a clean tube. The DNA was then precipitated out by isopropanol, washed with ethanol and air dried, and then hydrated in the DNA hydration solution provided by the kit.

For RNA extraction, 30 ml of culture was harvested by centrifugation at 3,000g for 15 minutes. RNA extraction was done using the hot phenol method. Briefly, the harvested pellet was re-suspended in AE buffer (50mM Na-Acetate pH 5.3 and 10mM EDTA) and 10% SDS was added to the suspension. Then, pre-equilibrated phenol was added to the suspension and the mixture was incubated at 65°C for 5 minutes then chilled on ice. The mixture is centrifuged at 21,000g for 3

minutes at 4°C and the aqueous layer is transferred to a new tube. A phenol/chloroform/isoamyl mix was then added to this new tube and the mixture is incubated at room temperature for 5 minutes. Repeat the centrifugation and move the top layer to a new tube again, and here, 3M Na-Acetate pH 5.3 was added to the solution and mixed well. The mixture was then added with 100% ethanol and incubate at -20°C for about 30 minutes. The mixture was centrifuged and washed with 80% ethanol. Lastly, the ethanol was removed after centrifugation and the RNA pellet was hydrated with RNase free water and treated with DNaseI (Qiagen).

Quantification of plasmid copy number was done by qPCR using the extracted gDNA with primers for the endogenous control gene ACT1, URA3 gene for MoBY1 plasmids, and LEU2 gene for MoBY2 plasmids. The samples are compared against a reference strain that has only one URA3 or one LEU2 gene in the genome, for MoBY1 and MoBY2 respectively. qPCR was performed using the ViiA7 Real-Time PCR system (Applied Biosystems) and EVA Green low ROX qPCR master mix from IMB protein production core facility.

Quantification of gene expression levels were done by RT-qPCR using the extracted RNA. Reverse transcription of RNA was performed using SuperScript III First-Strand Synthesis System for RT-qPCR (ThermoFisher) and oligo(dT)₂₀ primer as per kit protocol. The synthesized cDNA is then used for qPCR with primers for the endogenous control gene ACT1, mNG, and gene-specific oligos for each ORF tested. Comparison was done per ORF between the overexpression and native conditions. qPCR was performed using the ViiA7 Real-Time PCR system (Applied Biosystems) and EVA Green low ROX qPCR master mix from IMB protein production core facility.

Ct values obtained from ViiA7 Real-Time PCR system (Applied Biosystems) were exported using QuantStudio™ Real-Time PCR Software and imported into R (R Core Team, 2021) for data analysis and calculations. Fold difference of the OE condition for plasmid copy number compared to a reference strain and for gene expression compared to native condition was calculated using the $2^{-\Delta\Delta C_T}$ method (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008).

Next Generation Sequencing of MoBY plasmid libraries

In brief, the E. coli versions of the two MoBY libraries were grown as colonies on agar and then the colonies were washed and pooled together for each library. The libraries were then mini-prepped and the final plasmid pools were submitted to the IMB Genomics core facility to perform the library preparation for next generation sequencing with NextSeq 500 Midoutput Flowcell kit.

The pool of plasmid DNA was sheared with a Covaris S2 focused ultrasonicator (Parameters: Intensity=0.5, Duty Cycle=20%, Cycles per burst=50, Temperature=20°C, Water Level=15, Treatment Time=55sec). NGS library prep was then performed using NEBNext Ultra II DNA Library Prep Kit for Illumina Version 6.0, 3/20. Libraries were prepared with a starting amount of 100 ng of fragmented DNA and were amplified in 3 PCR cycles. The libraries were profiled in a High Sensitivity DNA chip on a 2100 Bioanalyzer (Agilent technologies) and quantified using the Qubit dsDNA HS Assay Kit, in a Qubit 2.0 Fluorometer (Life technologies) and by qPCR (sparQ Universal Library Quant Kit; Quanta Bio). Both samples were pooled in equimolar ratio and sequenced on 1 NextSeq 500 Midoutput Flowcell, PE for 2x 154 cycles plus 7 cycles for the index read.

The outcome of the sequencing data was then processed first by the IMB Bioinformatics core facility. Raw data quality was assessed with FastQC v0.11.8. Raw reads were trimmed using Cutadapt v1.18 to remove adapter sequences and were mapped with BWA v0.7.15 against the UCSC sacCer3 (S288C) reference genome. After duplicate removal with picard v2.20 MarkDuplicates we obtained a mean sequencing depth of >300. For variant calling of the plasmid pools we applied the GATK recommendations for somatic short variant discovery with Mutect2 (GATK v4.2.0.0) in 'Tumor-only' mode applying the Linked De Bruijn graph option to recover better haplotypes (--linked-de-bruijn-graph). Variants were filtered with GATK FilterMutectCalls by applying the filter criteria minimum allele fraction of 10% (--min-allele-fraction 0.1), minimum of 10 unique reads supporting the alternate allele (--unique-alt-read-count 10) and minimum of 5 alternate allele reads per strand (--min-reads-per-strand 5) additional to the default parameter. Filtered variants were annotated with snpEff v5.0. Overlap and non-overlap regions were determined using information from the primers obtained from (Ho *et al.*, 2009) and an average read depth per region was calculated.

Using the sequencing results, region-based annotations, and variant annotations provided, further processing and detailed analysis was performed in R (R Core Team, 2021) using the vcfR package for reading in vcf files and the Integrative Genomics Viewer (IGV) was used for manual visualization of the sequencing data from BAM (binary alignment map) files. The threshold for calling clear SNPs was set at 0.9 using information from non-overlap regions. Mixed plasmids were then differentiated using a combination of comparing the allele frequency (AF) of the SNP to the ratio of average read depth of the neighbouring region against sum of neighbouring regions and manual checks by directly looking at the sequencing results BAM files in IGV. The ratio was calculated by taking the average read depth of the next or previous region (neighbouring region) to the SNP and dividing it by the sum of these two regions. The SNP is more likely to belong to the plasmid that has the region where the ratio matches closer to the AF of the SNP. Only SNPs with the AF – ratio = 0

+/- 0.25 standard deviation and that the region it is assigned to is not also an overlap region then it is marked to be re-assigned to the correct plasmid and removed from the unaffected plasmid. Critical mutations were manually checked by directly looking at the sequencing results BAM files in IGV to determine if it was any sequencing artefact and/or if it should be re-assigned (in case if the SNP AF <0.9). All the information is then combined into a final table with the variant annotation and SNPs that have been marked from all the checks that need to be removed are removed.

Confirmation of protein localization via microscopy

For protein localization studies, strains were chosen and taken from the tFT library (constructed for the overexpression screens) and mNG-II library (Meurer *et al.*, 2018) for further modifications. MoBY2 plasmids for the selected ORFs were taken from the MoBY2 E. coli library and were checked by restriction digest and Sanger sequencing to confirm that there are no mutations present in the ORF on the plasmids. Overexpression plasmids with mCherry tagged ORF were constructed using PCR amplified mCherry from pMM151, PCR amplified ORF sequence from genomic DNA, and plasmid backbone p5587 digested with NruI and NdeI. The selected tFT library strains were then transformed individually with corresponding MoBY2 plasmids, mCherry-tagged overexpression plasmids, and the MoBY2 control plasmid carrying the ORF YGR045C. The selected mNG-II library strains were then transformed individually with corresponding mCherry-tagged overexpression plasmids and the MoBY2 control plasmid carrying the ORF YGR045C. These strains were then used for localization studies via fluorescence microscopy together with non-fluorescent strains from the non-fluorescent control library.

For fluorescence microscopy, yeast strains were inoculated into the appropriate selection media and grown overnight at 30°C in a 96-well plate. In the morning, the plate was diluted using the ROTOR pinning robot by pinning with 96-long pins from the overnight culture plate into a new 96-well plate with fresh media. The diluted plate was then grown at 30°C for 6 hours before transferring 60µl of culture 384-well plates (PhenoPlate 384-well microplates, PerkinElmer). Before the transferring the culture into the 384-well microscopy plate, the plate is first treated with 2mg/ml Concanavalin A (Sigma-Aldrich) for 30 minutes and then washed twice with sterile ddH₂O.

Microscopy imaging was then performed using the Opera Phenix (high-throughput spinning disk confocal microscope, Perkin Elmer) in the IMB microscopy core facility. Single plane images were taken from 10 fields of view for each well using the following setting and order: 561 nm

excitation and 570-630 nm emission filter for red channel, brightfield, and 488 nm excitation and 500-550 nm emission filter for green channel.

Images were then analyzed using ImageJ and Cellpose, an algorithm for cellular segmentation (Stringer *et al.*, 2021). First, yeast cells that are in focus with no artefacts were selected using Cellpose and a self-trained model through an ImageJ plugin (<https://github.com/BioImaging-NKI/Cellpose-Fiji>), which in the end generates a mask of the cells for each image. This mask was then used to select the cells and perform quantification of the median intensity values for the red and green channel, and also quantifications on cell size and circularity. The measurement data were then imported into R (R Core Team, 2021) for data analysis. Firstly, cells were filtered based on area, roundness and circularity to avoid overly small or large cells and irregular shaped cells (this can arise from the algorithm for cellular segmentation giving an incorrect mask). Then, the fluorescence values of all sample strains were background corrected using the mean intensity of non-fluorescent strains. The \log_2 fold change values when comparing between the different categories were calculated for red fluorescence, green fluorescence, and red-to-green ratio using background corrected fluorescence intensity values.

APPENDIX

Supplementary tables

Supplementary Table 1: Yeast strains used in this study

Strain	Background	Genotype	Reference
BY4741	S288c	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	(Baker Brachmann <i>et al.</i> , 1998)
Y8205	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2	(Tong and Boone, 2007)
yMaM639	Y8205	leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2	(Meurer <i>et al.</i> , 2018)
Y7092	S288c	MATalpha can1Δ::STE2pr-SpHIS5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0	(Tong and Boone, 2007)
yMaM1205	Y7092	can1Δ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI	(Meurer <i>et al.</i> , 2018)
yMaM312	Y8205	PDC1-sfGFPΔC-CYC1term-URA3-sfGFPΔN	Obtained from Matthias Meurer, Knop Lab
yJF0013	yMaM639	pJF001 (pRS41K mCherry-sfGFP C-SWAT type I donor)	This study
yJF0014	yMaM639	pJF003 (pRS41K mCherry-mNeonGreen C-SWAT type I donor)	This study
yJF0015	yMaM639	pMaM482 (pRS41K CEN ARS KanMx C-SWAT type I donor)	This study
yJF0018	yMaM1205	MATalpha lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::STE3pr-SpHIS5-ADH1term-GAL1pr-NLS-I-SCEI	This study
yJF0027	yJF0018	pJF004 (pRS41K mCherry-mNeonGreen-ADH1term-TEFpr-hphΔC C-SWAT type II donor)	This study
yJF0028	yJF0018	pMaM484 (pRS41K CEN ARS KanMx C-SWAT type II donor)	This study
yE10038	BY4742	MATalpha his3Δ leu2Δ lys2Δ ura3Δ ybr209w::KanMX-GPD-mNG-Myc-(GS)x3-SceI-FCY1-SceI-T8term-URA3prom-5'URA3-Barcode-lox71-GalCre can1::MFa1pr-HIS3-MFalpha1pr-LEU2 fcy1::HphMX SAL1 CAT5(91M) MIP1(661T)	Obtained from █████
yJF0142	BY4741	pMoBY2_VMA3 (2μ KanMX LEU2 VMA3)	This study
yJF0143	BY4741	pMoBY2_VMA5 (2μ KanMX LEU2 VMA5)	This study
yJF0144	BY4741	pMoBY2_MES1 (2μ KanMX LEU2 MES1)	This study
yJF0145	BY4741	pMoBY2 EMC1 (2μ KanMX LEU2 EMC1)	This study
yJF0146	BY4741	pJF030 (2μ KanMX LEU2 VMA9)	This study
yJF0147	BY4741	pJF031 (2μ KanMX LEU2 NOT5)	This study
yJF0148	BY4741	pJF032 (2μ KanMX LEU2 EMC3)	This study
yJF0149	BY4741	pJF033 (2μ KanMX LEU2 VPH1)	This study
yJF0150	BY4741	pJF034 (2μ KanMX LEU2 VMA10)	This study
yJF0151	BY4741	pJF035 (2μ KanMX LEU2 VMA11)	This study
yJF0152	BY4741	pJF037 (2μ KanMX LEU2 VMA13)	This study
yJF0153	BY4741	pJF036 (2μ KanMX LEU2 YPR170W-B)	This study
yJF0154	BY4741	pJF040 (2μ KanMX LEU2 POP2)	This study
yJF0155	BY4741	pJF039 (2μ KanMX LEU2 CCR4)	This study
yJF0156	BY4741	pJF038 (2μ KanMX LEU2 CAF130)	This study
yJF0157	BY4741	pJF044 (2μ KanMX LEU2 CDC39)	This study
yJF0158	BY4741	pJF046 (2μ KanMX LEU2 STV1)	This study
yJF0159	BY4741	pJF041 (2μ KanMX LEU2 MOT2)	This study
yJF0160	BY4741	pJF042 (2μ KanMX LEU2 GUS1)	This study
yJF0161	BY4741	pJF043 (2μ KanMX LEU2 NOT3)	This study
yJF0162	BY4741	pJF047 (2μ KanMX LEU2 CAF40)	This study

Strain	Background	Genotype	Reference
yJF0163	BY4741	pMoBY2_VOA1 (2μ KanMX LEU2 VOA1)	This study
yJF0164	BY4741	pMoBY2_VMA1 (2μ KanMX LEU2 VMA1)	This study
yJF0165	BY4741	pMoBY2_VMA2 (2μ KanMX LEU2 VMA2)	This study
yJF0166	BY4741	pMoBY2_VMA4 (2μ KanMX LEU2 VMA4)	This study
yJF0167	BY4741	pMoBY2_VMA6 (2μ KanMX LEU2 VMA6)	This study
yJF0168	BY4741	pMoBY2_VMA7 (2μ KanMX LEU2 VMA7)	This study
yJF0169	BY4741	pMoBY2_VMA8 (2μ KanMX LEU2 VMA8)	This study
yJF0170	BY4741	pMoBY2_VMA16 (2μ KanMX LEU2 VMA16)	This study
yJF0173	BY4741	pMoBY2_ARC1 (2μ KanMX LEU2 ARC1)	This study
yJF0174	BY4741	pMoBY2_SOP4 (2μ KanMX LEU2 SOP4)	This study
yJF0175	BY4741	pMoBY2 EMC4 (2μ KanMX LEU2 EMC4)	This study
yJF0176	BY4741	pMoBY2 EMC5 (2μ KanMX LEU2 EMC5)	This study
yJF0177	BY4741	pMoBY2 EMC6 (2μ KanMX LEU2 EMC6)	This study
yJF0178	BY4741	pMoBY2 EMC10 (2μ KanMX LEU2 EMC10)	This study
yJF0179	BY4741	pMoBY2_CDC36 (2μ KanMX LEU2 CDC36)	This study
yJF0185	BY4741	pJF045 (2μ KanMX LEU2 EMC2)	This study
yJF0186	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRPS8-mNeonGreen-hphNT1 pJF048 (2μ LEU2 AmpR KanMX MRPS8-mCherry)	This study
yJF0187	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 SWS2-mNeonGreen-hphNT1 pJF049 (2μ LEU2 AmpR KanMX SWS2-mCherry)	This study
yJF0188	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 GRX5-mNeonGreen-hphNT1 pJF050 (2μ LEU2 AmpR KanMX GRX5-mCherry)	This study
yJF0189	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRPS16-mNeonGreen-hphNT1 pJF051 (2μ LEU2 AmpR KanMX MRPS16-mCherry)	This study
yJF0190	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRPL22-mNeonGreen-hphNT1 pJF052 (2μ LEU2 AmpR KanMX MRPL22-mCherry)	This study
yJF0191	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRP20-mNeonGreen-hphNT1 pJF053 (2μ LEU2 AmpR KanMX MRP20-mCherry)	This study
yJF0192	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 TPI1-mNeonGreen-hphNT1 pJF054 (2μ LEU2 AmpR KanMX TPI1-mCherry)	This study
yJF0193	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRPS8-mNeonGreen-hphNT1 pMoBY2- YGR045C (2μ LEU2 AmpR YGR045C KanMX)	This study
yJF0194	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 SWS2-mNeonGreen-hphNT1 pMoBY2- YGR045C (2μ LEU2 AmpR YGR045C KanMX)	This study
yJF0195	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 GRX5-mNeonGreen-hphNT1 pMoBY2- YGR045C (2μ LEU2 AmpR YGR045C KanMX)	This study
yJF0196	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRPS16-mNeonGreen-hphNT1 pMoBY2-YGR045C (2μ LEU2 AmpR YGR045C KanMX)	This study

Strain	Background	Genotype	Reference
yJF0197	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRPL22-mNeonGreen-hphNT1 pMoBY2-YGR045C (2μ LEU2 AmpR YGR045C KanMX)	This study
yJF0198	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 MRP20-mNeonGreen-hphNT1 pMoBY2-YGR045C (2μ LEU2 AmpR YGR045C KanMX)	This study
yJF0199	S288c	MATa leu2Δ0::GAL1pr-NLS-I-SCEI-natNT2 can1Δ::STE2pr-SpHIS5 lyp1Δ::STE3pr-LEU2 his3Δ1 ura3Δ0 TPI1-mNeonGreen-hphNT1 pMoBY2-YGR045C (2μ LEU2 AmpR YGR045C KanMX)	This study
yJF0200	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI MRPS8-mCherry-mNG hph pJF048 (2μ LEU2 AmpR KanMX MRPS8-mCherry)	This study
yJF0201	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI SWS2-mCherry-mNG hph pJF049 (2μ LEU2 AmpR KanMX SWS2-mCherry)	This study
yJF0202	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI GRX5-mCherry-mNG hph pJF050 (2μ LEU2 AmpR KanMX GRX5-mCherry)	This study
yJF0203	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI MRPS16-mCherry-mNG hph pJF051 (2μ LEU2 AmpR KanMX MRPS16-mCherry)	This study
yJF0204	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI MRPL22-mCherry-mNG hph pJF052 (2μ LEU2 AmpR KanMX MRPL22-mCherry)	This study
yJF0205	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI MRP20-mCherry-mNG hph pJF053 (2μ LEU2 AmpR KanMX MRP20-mCherry)	This study
yJF0206	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI TPI1-mCherry-mNG hph pJF054 (2μ LEU2 AmpR KanMX TPI1-mCherry)	This study
yJF0125	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI RSM10-mCherry-mNG hph pJF017 (2μ LEU2 AmpR KanMX RSM10-mCherry)	This study
yJF0127	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI MRP7-mCherry-mNG hph pJF021 (2μ LEU2 AmpR KanMX MRP7-mCherry)	This study
yJF0128	S288c	MATalpha his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lyp1Δ can1Δ::STE3pr-SpHIS5-GAL1pr-I-SCEI IMG2-mCherry-mNG hph pJF022 (2μ LEU2 AmpR KanMX IMG2-mCherry)	This study

Supplementary Table 2: Plasmids used in this study

Name	Description	Reference
pMaM17	pFA6a mCherry-sfGFP-KanMx	(Khmelinskii <i>et al.</i> , 2012)
pMaM332	pFA6a mCherry-mNeonGreen-kanMX	(Khmelinskii <i>et al.</i> , 2016)
pMaM482	pRS41K CEN ARS kanMX (type I donor template)	(Meurer <i>et al.</i> , 2018)
pMaM484	pRS41K CEN ARS kanMX (type II donor template)	(Meurer <i>et al.</i> , 2018)
pJF001	pRS41K mCherry-sfGFP (C-SWAT type I donor)	This study
pJF003	pRS41K mCherry-mNeonGreen (C-SWAT type I donor)	This study
pJF004	pRS41K mCherry-mNeonGreen-ADH1term-TEFpr-hphΔC (C-SWAT type II donor)	This study
pJF023	pJF004 ΔKanMX::NatMX (C-SWAT type II donor)	This study
pJF029	pMaM484 ΔKanMX::NatMX (C-SWAT type II donor)	This study
pHU2797	pYM TAP TEFpr-SpHIS5-TEFterm	Obtained from Helle Ulrich's lab
pMoBY2_YGR045C	2μ LEU2 AmpR YGR045C KanMX MoBY2 library plasmid. Position: 19NP B1	(Ho, 2011; Magtanong <i>et al.</i> , 2011)
P5476	pMAGIC-2u-LEU2#1 starting vector ; the 2u from this vector is amplified from yep352 while the leu2marker is amplified from yep351; this is the starting recipient vector for yeast ORF cloning project	
P5587	2u empty barcoded vector LEU+, ampicillin/carbenicillin resistance G418/KAN cassette flanked with barcodes Uptag TACGGAGTGGTCAGTTACTT DNtag AGGTAGCCTATACTGCTCTT A6 clone	Obtained from Charlie Boone's lab
pMM151	pFA6a-mCherry-ADH1term-KanMX	(Khmelinskii <i>et al.</i> , 2011)
pJF030	P5476 (2μ LEU2 AmpR) VMA9 KanMX	This study
pJF031	P5476 (2μ LEU2 AmpR) NOT5 KanMX	This study
pJF032	P5476 (2μ LEU2 AmpR) EMC3 KanMX	This study
pJF033	P5587 (2μ LEU2 AmpR KanMX) VPH1	This study
pJF034	P5587 (2μ LEU2 AmpR KanMX) VMA10	This study
pJF035	P5587 (2μ LEU2 AmpR KanMX) VMA11	This study
pJF036	P5587 (2μ LEU2 AmpR KanMX) YPR170W-B	This study
pJF037	P5587 (2μ LEU2 AmpR KanMX) VMA13	This study
pJF038	P5587 (2μ LEU2 AmpR KanMX) CAF130	This study
pJF039	P5587 (2μ LEU2 AmpR KanMX) CCR4	This study
pJF040	P5587 (2μ LEU2 AmpR KanMX) POP2	This study
pJF041	P5476 (2μ LEU2 AmpR KanMX) MOT2 MoBY2 library plasmid; mutation corrected by SDM	This study
pJF042	2μ LEU2 AmpR GUS1 KanMX MoBY2 library plasmid; mutation corrected by SDM	This study
pJF043	P5476 (2μ LEU2 AmpR KanMX) NOT3 MoBY2 library plasmid; mutation corrected by SDM	This study
pJF044	P5587 (2μ LEU2 AmpR KanMX) CDC39	This study
pJF045	P5587 (2μ LEU2 AmpR KanMX) EMC2	This study
pJF046	P5587 (2μ LEU2 AmpR KanMX) STV1	This study
pJF047	P5476 (2μ LEU2 AmpR KanMX) CAF40 MoBY2 library plasmid; mutation corrected by SDM	This study
pJF048	P5587 (2μ LEU2 AmpR KanMX) MRPS8-mCH	This study
pJF049	P5587 (2μ LEU2 AmpR KanMX) SWS2-mCH	This study
pJF050	P5587 (2μ LEU2 AmpR KanMX) GRX5-mCH	This study
pJF051	P5587 (2μ LEU2 AmpR KanMX) MRPS16-mCH	This study

Name	Description	Reference
pJF052	P5587 (2μ LEU2 AmpR KanMX) MRPL22-mCH	This study
pJF053	P5587 (2μ LEU2 AmpR KanMX) MRP20-mCH	This study
pJF054	P5587 (2μ LEU2 AmpR KanMX) TPI1-mCH	This study
pJF017	P5476 (2μ LEU2 AmpR KanMX) RSM10-mCH MoBY2 library plasmid was cut and mCH tag inserted after ORF	This study
pJF021	P5476 (2μ LEU2 AmpR KanMX) MRP7-mCH MoBY2 library plasmid was cut and mCH tag inserted after ORF	This study
pJF022	P5476 (2μ LEU2 AmpR KanMX) IMG2-mCH MoBY2 library plasmid was cut and mCH tag inserted after ORF	This study

Supplementary Table 3: Yeast and *E. coli* libraries used in this study

Name	Genotype	Source	Reference
MoBY1 (<i>E. coli</i>)	BUN20 [Δ lac-169 rpoS(Am) robA1 creC510 hsdR514 Δ uidA(MluI):pir-116 endA(BT333) recA1 F'(lac+ pro+ Δ oriT:tet)] Plasmid: p5472 CEN CmR URA3 ORF KanMX	Dharmacon	(Ho <i>et al.</i> , 2009)
MoBY1 (Yeast)	BY4741 MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0 Plasmid: p5472 CEN CmR URA3 ORF KanMX	This thesis	
MoBY2 (<i>E. coli</i>)	P5530 [lacI ^Q rmB3 Δ lacZ4787 hsdR514 Δ (araBAD)567 Δ (rhaBAD)568 galU95 Δ endA9:FRT Δ recA635:FRT umuC:ParaBAD-I-SceI-FRT] Plasmid: p5476 2μ LEU2 AmpR ORF KanMX	Obtained from Charlie Boone's lab	
MoBY2 (Yeast)	BY4741 MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0 Plasmid: p5476 2μ LEU2 AmpR ORF KanMX	Dharmacon	(Ho, 2011; Magtanong <i>et al.</i> , 2011)
mORF	Y258 MATa, pep4-3, his4-580, ura3-52, leu2-3,112 Plasmid: BG1805 URA3 GAL1pr-ORF-His ₆ -HA ^{epitope} -3C ^{Protease site} -ZZ domain ^{ProteinA} -GAL1term	Obtained from Brian Luke's lab	(Gelperin <i>et al.</i> , 2005)
Yeast Knockout Collection	BY4741 orf Δ ::KanMX	Obtained from Brian Luke's lab	(Giaever <i>et al.</i> , 2002)
C-SWAT	BY4741 ORF-C_SWAT-URA3	Obtained from Michael Knop's lab	(Meurer <i>et al.</i> , 2018)
mNG-II	MATa leu2 Δ 0::GAL1pr-NLS-I-SCEI-natNT2 can1 Δ ::STE2pr-SpHIS5 lyp1 Δ ::STE3pr-LEU2 his3 Δ 1 ura3 Δ 0 ORF-mNeonGreen-hphNT1	Obtained from Michael Knop's lab	(Meurer <i>et al.</i> , 2018)
tFT (mCH-mNG) type-II	MATalpha ORF-mCherry-mNeonGreen-hphNT1 lyp1 Δ his3 Δ 1 leu2 Δ 0 ura3 Δ 0 met15 Δ 0 can1 Δ ::STE3pr-SpHIS5-ADH1term-GAL1pr-NLS-I-SCEI Note: Donor plasmid uses KanMX	This thesis	
tFT (mCH-mNG) type-II Version 2	MATalpha ORF-mCherry-mNeonGreen-hphNT1 lyp1 Δ his3 Δ 1 leu2 Δ 0 ura3 Δ 0 met15 Δ 0 can1 Δ ::STE3pr-SpHIS5-ADH1term-GAL1pr-NLS-I-SCEI Note: Donor plasmid uses NatMX	This thesis	

Supplementary Table 4: Primers for qPCR and RT-qPCR

Gene	Forward primer (5'-3')	Reverse primer (5'-3')
ACS2	CACGAAAACGTCTCGGAAGC	GGTGTGATGTGTCTGCATCAC
ACT1	CCCAGGTATTGCCGAAAGAATGC	TTTGTGGAAAGGTAGTCAAAGAAGCC
AFB1	CCAATTCTGTTCCTTTCTCGACAGC	ACCATTCTGTGATTGACTAGTTCTGG
CAC2	CCCGTATTTTACGAAACGTGCCA	GACCATGCTAAATCAGTTATGGGTGAG
DSN1	CGATGAGCTAAGGTACCAGTCG	TCCCTTCAACAGCTGTTGAGG
EMC5	ACGATGGGAAGATAATATCCCAAGG	TGAACGTCAACAAAGCTTGGAG
ERG25	GCTGAACACCACGATTTGCAT	CTTCTCTGGAGGCCTTAGCTTC
ERP1	GCCACTATTGACATTCTACATGCC	TGACAATCCACCACATAGCAG
ERP2	CTTTGACTGACGAGCATGAAGC	AGACATGTTTCTCCATTCTCTGGC
FMP41	ACAGGATATTTTCAGAGTCAAGTGCTC	AGTCAAAATGATATCACCCGGTTC
GCD11	GGTGGTCTGATTGGTGTGG	CCATCTGTTTTGACACCTAATAGACG
GUS1	GCAAATTCGCCGCCATCTTC	GGCAGAATACTTCAAGCCGCC
GUT1	CTGTCGGTTCTGGCAGTGG	CATATTGGCTGCAATGGCTGC
HMO1	GCTTGCCACCTTTATCCTCAACTG	GTGGTAGGGTACCGTTTTTCTTTGC
HSP78	TCCACCTGAGTTCATCAATCG	AGCCAATCCTTCGCTTCATCAG
KanMX	CGATGCGCCAGAGTTGTTTCTG	CGCAGTGGTGAGTAACCATGC
LCB2	ATGCCGGCTCCTGTTTTAGC	CAGCCACACCGTAGACAATAAATCC
LEU2	CTGTGGAGGAAACCATCAAGAACG	TGGGATAACGGAGGCTTCATCG
LSM12	TGCAGAGGGTCAGTTCATTTTCG	GGATTGGTTACTTCCCTCATGTAGC
LSP1	CCTTATTGGAGCTGCTAGATGACTC	GAAAGAGTTGGTTTACTGCTGACG
mNG	CACTTTCGCTAAGCCAATGGCT	CCATACCCATAACGTCAGTGAAAGC
MPM1	ACGATGCTGAATATGCTTCTGC	TAGTAATGTCTACCGTGCCATCG
NAR1	CTACCACTCGGACTACATCGAGG	CCATTGCTAGCTCCTCTCCGT
NSL1	AGGATGAGTATTTGGCACAATTGG	ACAATTCATTGAGTGAGGCAACG
OST3	GGTCAAGGGTATAACAATCTTGCGG	GCAGGACTCTTTATCGTGAACACG
POP1	GCACACTACTCGAACGAGGTC	TCAAATCATGTACCTCTGGAAGTGG
POP5	CGACGTCGATGGACTGATCG	GTCATTATCGCTGAGGTGTGATG
PRE5	CATCTTTTAGAATTCCAACCTTCCGG	CCAGCTTTGATTAGTTCATCCGG
PRI1	GCCCAAGCAAGACTCTCAC	CACACAGACATTCCCCGTAGC
PRT1	GATGCAATGGAAGCTGACACC	TCGATGGTGGTAAAGTCATCGC
RAD1	GCTTCAAAAGCTTTTCATGGAATGAG	CGTTATCATCAGTGGTCTTACCAGG
RPL26B	CGTTTCTCTGACAGAAGAAAGG	GACTTCATCGTCTTCTGATTGG
RPP1	AGGAATAAGAGCCACAAACAGACC	CATGGCCTAATTGTTTCAGCATCC
RPT2	CGGGTGAGAATGCACCAAGTATTG	GTGGCCATAATCACCTTCACATCC
SCD6	GTAATCGCCCTCCTCAATCC	GATTCATCCTGCTTGTGAACAGC
UBC1	GCATGGTCGCCAGTGATAAACT	GGCGTATAACCTCGTCCATAGTGC
URA3	GAGACATGGGTGGAAGAGATGAAGG	CCTGTAGAGACCACATCATCCACG
UTP18	AAATAGGTGGCTTGCGGTAGG	GCACAGTATTTGACCATCAGGGG
ZUO1	AAGCTGAAGCTAAGCCAAGG	CCTTAGCAGAGTTACGGATGGC

Supplementary Table 5: General primers

Name	Sequence (5'-3')	Description/Notes
Constructing tFT donor plasmids and tFT donor strain		
F-mCherry-GB1	GGTTCTGGCGGTGGCGGATCCATGGTGAGCAAGGGCG	Amplify tFT insert for constructing tFT donor plasmids.
R-sfGFP-GB1	CGATATCTCAAAGAAGAACTACTAGTTACTTATAAAGC TCGTCCATTCCG	Amplify tFT insert for constructing tFT donor plasmids. For mCH-sfGFP type I tFT. Forward primer is F-mCherry-GB1
R-sfGFP-GB2	AAATTCGCTTATTTAGAAGTACTAGTTTACTTATAAAG CTCGTCCATTCCG	Amplify tFT insert for constructing tFT donor plasmids. For mCH-sfGFP type II tFT. Forward primer is F-mCherry-GB1
R-mNeon-GB1	CGATATCTCAAAGAAGAACTACTAGTTTACTTGTACAAT TCGTCCATACCC	Amplify tFT insert for constructing tFT donor plasmids. For mCH-mNG type I tFT. Forward primer is F-mCherry-GB1
R-mNeon-GB2	AAATTCGCTTATTTAGAAGTACTAGTTTACTTGTACAA TTCGTCCATACCC	Amplify tFT insert for constructing tFT donor plasmids. For mCH-mNG type II tFT. Forward primer is F-mCherry-GB1
R-SpHIS5-N	AACTTTAGAGGCAATTAATTTGTGTAGGAAAGGCAAA ATACTATCAAAATTTTCATGGGTAGGAGGGCTTTGTAG	Amplify SpHIS5 from pHU2797 for swapping out LEU2 from yMaM1205
F-ADH1t-C	TTTTTATAACTTATTTAATAATAAAAATCATAAATCATA AGAAATTCGCCTCGAGTTACAACACTCCCTTCGTGCT	Amplify SpHIS5 from pHU2797 for swapping out LEU2 from yMaM1205
Constructing and correcting MoBY2 overexpression plasmids		
F_GUS1_SDM_at1987	AATGTTAAGGaTATGAAGATTGGTG	For site-directed mutagenesis of GUS1
R_GUS1_SDM_at1987	CAAGTCAGCAATGGCATC	For site-directed mutagenesis of GUS1
F_NOT3_SDM_at37	GTCGATAGGGtCTTTAAAAAATTAAC	For site-directed mutagenesis of NOT3
R_NOT3_SDM_at37	CTCCTGCTGTAATTTTCTATG	For site-directed mutagenesis of NOT3
F_MOT2_SDM_at1243	GCTGCAGAGGaATACAAAGATC	For site-directed mutagenesis of MOT2
R_MOT2_SDM_at1243	AAGAGATTGCTTCTTCTTG	For site-directed mutagenesis of MOT2
F_STV1_SDM_at1605	TGACAATAfTCAAATCAGGCTGGCAATGGCCTTCCACTTT C	For site-directed mutagenesis of STV1
R_STV1_SDM_at1605	AGCCTGATTTGAaTATTGTCATAGATTTGAAAAGATAT CATTGTACAACAGC	For site-directed mutagenesis of STV1
F_STV1_SDM_at1842	ACATAaTTGGTAACTTTATCCCCGTTTAGTGTATTATGCA ATCAATATTTGG	For site-directed mutagenesis of STV1
R_STV1_SDM_at1842	GGGATAAAGTTACCAaTATGTCCACTTTAGAGTTTTTAG CTCTGTAATTGATATAAG	For site-directed mutagenesis of STV1
F_CAF40_SDM_at874	AAGGTTCTACgCCGTAACAAATG	For site-directed mutagenesis of CAF40
R_CAF40_SDM_at874	TCCAAGGTGGCACAGATG	For site-directed mutagenesis of CAF40
F_VMA2_SDM_at595	CCTACCAAGGaTGTTTCATGATG	For site-directed mutagenesis of VMA2
R_VMA2_SDM_at595	TCTCACAAACCAGCCTG	For site-directed mutagenesis of VMA2
F_MoBY2_p5587_VMA10	atcgataagcttgacgtccatagCGCCATACTTGCAAATTGC	Amplify VMA10 insert from genomic DNA
R_MoBY2_p5587_VMA10	agagacctgtggacatcgcgccgcGTAATCCCCATCGGCGTTTAG	Amplify VMA10 insert from genomic DNA
F_MoBY2_p5587_VMA11	atcgataagcttgacgtccatagGCTTATTCCTCTTCGGTTAG	Amplify VMA11 insert from genomic DNA
R_MoBY2_p5587_VMA11	agagacctgtggacatcgcgccgcGATGCCATGGATACTACG	Amplify VMA11 insert from genomic DNA
F_MoBY2_p5587_YPR170W-B	atcgataagcttgacgtccatagAGTGCAAACACAGAAGAG	Amplify YPR170W-B insert from genomic DNA

Name	Sequence (5'-3')	Description/Notes
R_MoBY2_p5587_YPR1 70W-B	agagacctgtggacatcgcgccgcACGCTCATATTTTGTGCATAC	Amplify YPR1 70W-B insert from genomic DNA
F_MoBY2_p5587_EMC2	atcgataagcttgacgtccatagCCTCAATCAATCTGAAAGAAC	Amplify EMC2 insert from genomic DNA
R_MoBY2_p5587_EMC2	agagacctgtggacatcgcgccgcCTGCTCAAATGCATGTC	Amplify EMC2 insert from genomic DNA
F_MoBY2_p5587_VMA 13	atcgataagcttgacgtccatagCTGCAAAGCGACGCTGTG	Amplify VMA13 insert from genomic DNA
R_MoBY2_p5587_VMA 13	agagacctgtggacatcgcgccgcCCATTAATACCCGACTTCGGC	Amplify VMA13 insert from genomic DNA
F_MoBY2_p5587_CAF1 30	atcgataagcttgacgtccatagCTGATGCAATTCTACCTTAC	Amplify CAF130 insert from genomic DNA
R_MoBY2_p5587_CAF1 30	agagacctgtggacatcgcgccgcGGAACCCATGTTTAACAC	Amplify CAF130 insert from genomic DNA
F_MoBY2_p5587_VPH1	atcgataagcttgacgtccatagGAAACTCAGATAAGTCGATATG	Amplify VPH1 insert from genomic DNA
R_MoBY2_p5587_VPH1	agagacctgtggacatcgcgccgcGATTTTCGATTCTAACGTTACC	Amplify VPH1 insert from genomic DNA
F_MoBY2_p5587_CCR4	gctgggtaccggccccccctcgagCGTGCTCTAGCACTGATAAG	Amplify CCR4 insert from genomic DNA
R_MoBY2_p5587_CCR4	agagacctgtggacatcgcgccgcGGATTTCGTTATAGAATACAAGATTC	Amplify CCR4 insert from genomic DNA
F_MoBY2_p5587_POP2	gctgggtaccggccccccctcgagCTCATTATACCAGTCATTCC	Amplify POP2 insert from genomic DNA
R_MoBY2_p5587_POP2	agagacctgtggacatcgcgccgcGTCATTCTTTCATTAATCGC	Amplify POP2 insert from genomic DNA
F_MoBY2_p5587_CDC3 9	gctgggtaccggccccccctcGTCGACGGCTCTCAAAGTACCAG	Amplify CDC39 insert from genomic DNA
R_MoBY2_p5587_CDC3 9	agagacctgtggacatcgcgccgcCATAAGGTAATTGCAGCTC	Amplify CDC39 insert from genomic DNA
F_MoBY2_p5587_STV1	gctgggtaccggccccccctcGTCGACGCCATTTTGTATTGTAGTCAAC	Amplify STV1 insert from genomic DNA
R_MoBY2_p5587_STV1	agagacctgtggacatcgcgccgcCTACGACAGTTTCCGATTTTG	Amplify STV1 insert from genomic DNA
Constructing mCherry tagged ORF overexpression plasmids		
F-Linker-mCH-KanNterm	ggtggcggttctgcccgtggcggtaccATGGTGAGCAAGGGCGAGGA	Amplify mCherry insert
R-InKanNterm-mCH	ctgattgcccgacattatcgCGAGCCATTATACCCATATAAATCA GC	Amplify mCherry insert
F-480bpUP-MRPS8	atcgataagcttgacgtccatagCTGATAATGTTGTCTTCGGTTTG	Amplify MRPS8 insert from genomic DNA
R-Linker-MRPS8-noStop	ggatccgccaccgccagaaccgccaccTTTACTCTGCATAGTACTTCT C	Amplify MRPS8 insert from genomic DNA
F-479bpUP-SWS2	atcgataagcttgacgtccatagCCTCAACGCGAACGATAGATC	Amplify SWS2 insert from genomic DNA
R-Linker-SWS2-noStop	ggatccgccaccgccagaaccgccaccCTTGCCAAAGATACAAGACC	Amplify SWS2 insert from genomic DNA
F-490bpUP-GRX5	atcgataagcttgacgtccatagGACGACTGTAGAAGTATGCAAATC	Amplify GRX5 insert from genomic DNA
R-Linker-GRX5-noStop	ggatccgccaccgccagaaccgccaccACGATCTTTGGTTTCTTCTTC	Amplify GRX5 insert from genomic DNA
F-496bpUP-MRPS16	atcgataagcttgacgtccatagCCTCAACGGCAGTTAATATTATAG	Amplify MRPS16 insert from genomic DNA
R-Linker-MRPS16- noStop	ggatccgccaccgccagaaccgccaccCTCCAATGTTTCCATTCTTTC	Amplify MRPS16 insert from genomic DNA
F-498bpUP-MRPL22	atcgataagcttgacgtccatagCAGTTCAGAACAATCACAAGGG	Amplify MRPL22 insert from genomic DNA
R-Linker-MRPL22- noStop	ggatccgccaccgccagaaccgccaccCCACTTGTATACGCCGCTG	Amplify MRPL22 insert from genomic DNA
F-479bpUP-MRP20	atcgataagcttgacgtccatagCATCTCAGAAATTGAGTTATTTAT AC	Amplify MRP20 insert from genomic DNA
R-Linker-MRP20-noStop	ggatccgccaccgccagaaccgccaccATGTAGATCTTCTATGTACCT G	Amplify MRP20 insert from genomic DNA
F-493bpUP-TPI1	atcgataagcttgacgtccatagCTGTGAGGACCTTAATACATTC	Amplify TPI1 insert from genomic DNA

Name	Sequence (5'-3')	Description/Notes
R-Linker-TPI1-noStop	ggatccgccaccgccagaaccgccaccGTTTCTAGAGTTGATGATATC AAC	Amplify TPI1 insert from genomic DNA
R-InKanMX-273	CGCAGTGGTGAGTAACCATGC	Primer binds within KanMX
F-YDR041W-L-mCherry	GAACTATTGGATAGCCCCGATTTTAAAAAACATTTGGAA AAGAAAggtggcggttctggcgggcatCATGGTGAGCAAGGG CGAG	Amplify mCherry with arms into RSM10. This was an early attempt to use yeast gap-repair cloning to construct the plasmid. Reverse primer is R-InKanMX-273.
F-YNL005C-L-mCherry	GTTGAAGTCATTGCTAGAAAGTAGGCGGGCTTTTTGAGC AAGCTTggtggcggttctggcgggcatCATGGTGAGCAAGGGC GAG	Amplify mCherry with arms into MRP7. This was an early attempt to use yeast gap-repair cloning to construct the plasmid. Reverse primer is R-InKanMX-273.
F-YCR071C-L-mCherry	GGCAATGCAGTAGAGGCTGTCAAAAGGGTATTAACCAA GAAATTTggtggcggttctggcgggcatCATGGTGAGCAAGGG CGAG	Amplify mCherry with arms into IMG2. This was an early attempt to use yeast gap-repair cloning to construct the plasmid. Reverse primer is R-InKanMX-273.

Supplementary Table 6: Sources for datasets used in data analysis

Dataset Name	Values	Reference
N-terminome	PSI (Protein Stability Index)	(Kats <i>et al.</i> , 2018)
Synthesis Rate	Protein synthesis rates	(Taggart and Li, 2018)
Half-life	Protein half-life in minutes	(Christiano <i>et al.</i> , 2014)
Ubiquitylation	Number of ubiquitylation sites	(Swaney <i>et al.</i> , 2013)
Protein disorder	Average / N-terminal / C-terminal / Internal disorder score	(Van Der Lee <i>et al.</i> , 2014)
Aneuploid	Change in protein levels	(Dephoure <i>et al.</i> , 2014)
Essential ORFs	List of essential ORFs	(Després <i>et al.</i> , 2020) *
Protein subunits	List of protein subunits	(Meldal <i>et al.</i> , 2015)
Autoregulation of ribosomal subunits	Change in protein levels	(Roy <i>et al.</i> , 2020)
Protein localization	Localization score	(Chong <i>et al.</i> , 2015)

*Retrieved from link provided in the article: http://www-sequence.stanford.edu/group/yeast_deletion_project/Essential_ORFs.txt

Supplementary Table 7: Curated list of ORFs with known dosage compensation behaviour upon overexpression

ORF	Gene	Dosage Compensated	Reference	Source
YAL061W	BDH2	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL060W	BDH1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL059W	ECM1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL058W	CNE1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL055W	PEX22	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL049C	AIM2	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL046C	AIM1 (BOL3)	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL044W-A	BOL1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL044C	GCV3	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL041W	CDC24	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL039C	CYC3	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL038W	CDC19	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL023C	PMT2	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL021C	CCR4	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL017W	PSK1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL016W	TPD3	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL015C	NTG1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL012W	CYS3	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL011W	SWC3	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL010C	MDM10	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL008W	FUN14	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAL005C	SSA1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR002W	NUP60	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR002C-A	ERP1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR003W	SWD1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR007C	RFA1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR014C	BUD14	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR015W	ADE1	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR019C	CDC15	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YAR028W	YAR028W	No	Ishikawa <i>et al.</i> (2017)	Fig.S1
YBL018C	POP8	No	Ishikawa <i>et al.</i> (2017)	Fig.S5
YOR091W	TMA46	No	Ishikawa <i>et al.</i> (2017)	Fig.4
YIR010W	DSN1	No	Ishikawa <i>et al.</i> (2017)	Fig.4
YML012W	ERV25	No	Ishikawa <i>et al.</i> (2017)	Fig.4
YGL022W	STT3	Yes	Mueller <i>et al.</i> (2015)	Fig.3
YMR149W	SWP1	Yes	Mueller <i>et al.</i> (2015)	Fig.3
YOR103C	OST2	Yes	Mueller <i>et al.</i> (2015)	Fig.3
YJL002C	OST1	Yes	Mueller <i>et al.</i> (2015)	Fig.3
YAL036C	RBG1	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YAL034W-A	MTW1	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YAL033W	POP5	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YAL027W	SAW1	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YAL007C	ERP2	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YPL233W	Nsl1	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YNL282W	POP3	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YGR030C	POP6	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4

ORF	Gene	Dosage Compensated	Reference	Source
YBR167C	POP7	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YHR062C	RPP1	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YIR015W	RPR2	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YGL200C	EMP24	Yes	Ishikawa <i>et al.</i> (2017)	Fig.4
YMR142C	RPL13B	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YLR344W	RPL26A	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YER056C-A	RPL34A	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YMR194W	RPL36A	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YDR447C	RPS17B	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YDR450W	RPS18A	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YER074W	RPS24A	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YIL069C	RPS24B	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YNL031C	HHT2	Yes	Sung, Reitsma, <i>et al.</i> (2016)	Fig.1
YGL223C	COG1	Yes	Shemorry, Hwang and Varshavsky (2013)	Fig.4b

Instruments and software

Supplementary Table 8: Instruments used in this study

Instrument	Supplier
ROTOR pinning robot	Singer Instruments
Stinger attachment for ROTOR	Singer Instruments
PhenoBooth	Singer Instruments
Spark 20M plate reader with filters	TECAN
Spark 20M plate reader with monochromator	TECAN
Chemidoc imaging system	BioRad
Opera phenix	PerkinElmer
BD LSRFortessa SORP with high throughput sampler	BD
ViiA7 Real-Time PCR system	Applied Biosystems
Liquidator 96	Steinbrenner
BioShake XP, microtiter plate shaker	Analytik Jena

Supplementary Table 9: Softwares used in this study

Software	Source
Microsoft Office 2016	Microsoft
Adobe Illustrator 2020	Adobe
R Studio (2022.02.3 Build 492)	https://posit.co/products/open-source/rstudio/
R (versions 3.5.0 and 4.1.1)	(R Core Team, 2021) https://www.R-project.org/
R Markdown (Ver. 2.18.1)	https://rmarkdown.rstudio.com/
ImageJ (Ver. 1.53t)	https://imagej.nih.gov/ij/
Cellpose (through ImageJ)	(Stringer <i>et al.</i> , 2021)
ImageLab (Ver. 5.2.1)	BioRad
Phobius (through R)	(Käll, Krogh and Sonnhammer, 2004)
BD FACSDiva (Ver. 9.0.1)	BD
QuantStudio™ Real-Time PCR Software (Ver. 1.3)	Applied Biosystems
Integrative Genomics Viewer (Ver. 2.10.0)	(Robinson <i>et al.</i> , 2011) https://software.broadinstitute.org/software/igv/home

Supplementary Table 10: R packages used for data analysis

R package	Source
gitter	(Wagih and Parts, 2014)
tidyverse	(Wickham <i>et al.</i> , 2019)
knitr	(Xie, 2015, 2022)
openxlsx	(Schauberger and Walker, 2021)
VennDiagram	(Chen, 2022)
gplots	(Warnes <i>et al.</i> , 2022)
ragp	(Dragičević <i>et al.</i> , 2020)
vcfR	(Knaus and Grünwald, 2017)

REFERENCE LIST

- Amm, I., Sommer, T. and Wolf, D. H. (2014) 'Protein quality control and elimination of protein waste: The role of the ubiquitin-proteasome system', *Biochimica et Biophysica Acta - Molecular Cell Research*. Elsevier, 1843(1), pp. 182–196. doi: 10.1016/j.bbamcr.2013.06.031.
- An, H. and Harper, J. W. (2020) 'Ribosome Abundance Control Via the Ubiquitin-Proteasome System and Autophagy', *Journal of Molecular Biology*. Academic Press, pp. 170–184. doi: 10.1016/j.jmb.2019.06.001.
- Ast, T. *et al.* (2014) 'A cytosolic degradation pathway, prERAD, monitors pre-inserted secretory pathway proteins', *Journal of Cell Science*. Company of Biologists Ltd, 127(14), pp. 3017–3023. doi: 10.1242/JCS.144386/259835/AM/A-CYTOSOLIC-DEGRADATION-PATHWAY-PRERAD-MONITORS.
- Bai, Y. *et al.* (1999) 'The CCR4 and CAF1 Proteins of the CCR4-NOT Complex Are Physically and Functionally Separated from NOT2, NOT4, and NOT5', *Molecular and Cellular Biology*. American Society for Microbiology, 19(10), pp. 6642–6651. doi: 10.1128/MCB.19.10.6642/ASSET/27239D4B-6B4A-44F1-8713-75683156E956/ASSETS/GRAPHIC/MB1091752007.JPEG.
- Baker Brachmann, C. *et al.* (1998) 'Designer Deletion Strains derived from *Saccharomyces cerevisiae* S288C: a Useful set of Strains and Plasmids for PCR-mediated Gene Disruption and Other Applications', *Yeast*, 14, pp. 115–132. doi: 10.1002/(SICI)1097-0061(19980130)14:2.
- Baryshnikova, A. *et al.* (2010) 'Synthetic genetic array (SGA) analysis in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*', *Methods in enzymology*. Methods Enzymol, 470(C), pp. 145–179. doi: 10.1016/S0076-6879(10)70007-0.
- Ben-Zvi, A., Miller, E. A. and Morimoto, R. I. (2009) 'Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(35), pp. 14914–14919. doi: 10.1073/PNAS.0902882106/SUPPL_FILE/0902882106SI.PDF.
- Bonney, M. E., Moriya, H. and Amon, A. (2016) 'Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes', pp. 898–903. doi: 10.1101/gad.261743.115.898.
- Brennan, C. M. *et al.* (2019) 'Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells', *Genes & Development*. Cold Spring Harbor Laboratory Press. doi: 10.1101/gad.327494.119.
- Bridges, R. J. and Bradbury, N. A. (2018) 'Cystic fibrosis, cystic fibrosis transmembrane conductance regulator and drugs: Insights from cellular trafficking', *Handbook of Experimental Pharmacology*. Springer New York LLC, 245, pp. 385–425. doi: 10.1007/164_2018_103/COVER.
- Chan, K. M. *et al.* (2013) 'The 2 micron plasmid of *Saccharomyces cerevisiae*: A miniaturized selfish genome with optimized functional competence', *Plasmid*. Academic Press, 70(1), pp. 2–17. doi: 10.1016/J.PLASMID.2013.03.001.
- Chen, B. *et al.* (2011) 'Cellular strategies of protein quality control', *Cold Spring Harbor Perspectives in Biology*, 3(8), pp. 1–14. doi: 10.1101/cshperspect.a004374.
- Chen, G., Kroemer, G. and Kepp, O. (2020) 'Mitophagy: An Emerging Role in Aging and Age-Associated Diseases', *Frontiers in Cell and Developmental Biology*. Frontiers Media S.A., 8, p. 200.

doi: 10.3389/FCELL.2020.00200/BIBTEX.

Chen, H. (2022) ‘VennDiagram: Generate High-Resolution Venn and Euler Plots’. Available at: <https://cran.r-project.org/package=VennDiagram>.

Chen, J. *et al.* (2001) ‘Purification and characterization of the 1.0 MDa CCR4-NOT complex identifies two novel components of the complex’, *Journal of Molecular Biology*. Academic Press, 314(4), pp. 683–694. doi: 10.1006/JMBI.2001.5162.

Chen, J., Chiang, Y. C. and Denis, C. L. (2002) ‘CCR4, a 3′–5′ poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylase’, *The EMBO Journal*. John Wiley & Sons, Ltd, 21(6), pp. 1414–1426. doi: 10.1093/EMBOJ/21.6.1414.

Chen, Y.-C. *et al.* (2014) ‘Msp1/ATAD1 maintains mitochondrial function by facilitating the degradation of mislocalized tail-anchored proteins’, *The EMBO Journal*. John Wiley & Sons, Ltd, 33(14), pp. 1548–1564. doi: 10.15252/EMBJ.201487943.

Chong, Y. T. *et al.* (2015) ‘Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis.’, *Cell*. Elsevier, 161(6), pp. 1413–24. doi: 10.1016/j.cell.2015.04.051.

Christiano, R. *et al.* (2014) ‘Global Proteome Turnover Analyses of the Yeasts *S.cerevisiae* and *S.pombe*’, *Cell Reports*. Elsevier B.V., 9(5), pp. 1959–1965. doi: 10.1016/j.celrep.2014.10.065.

Christianson, T. W. *et al.* (1992) ‘Multifunctional yeast high-copy-number shuttle vectors’, *Gene*. Elsevier, 110(1), pp. 119–122. doi: 10.1016/0378-1119(92)90454-W.

Collart, M. A. and Timmers, H. T. M. (2004) ‘The Eukaryotic Ccr4-Not Complex: A Regulatory Platform Integrating mRNA Metabolism with Cellular Signaling Pathways?’, *Progress in Nucleic Acid Research and Molecular Biology*. Academic Press, 77, pp. 289–322. doi: 10.1016/S0079-6603(04)77008-7.

Compton, M. A., Graham, L. A. and Stevens, T. H. (2006) ‘Vma9p (Subunit e) Is an Integral Membrane V0 Subunit of the Yeast V-ATPase’, *Journal of Biological Chemistry*. Elsevier, 281(22), pp. 15312–15319. doi: 10.1074/JBC.M600890200.

Cronshaw, J. M. and Matunis, M. J. (2003) ‘The nuclear pore complex protein ALADIN is mislocalized in triple A syndrome’, *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 100(10), pp. 5823–5827. doi: 10.1073/PNAS.1031047100.

Cui, Y. *et al.* (2008) ‘Genome wide expression analysis of the CCR4-NOT complex indicates that it consists of three modules with the NOT module controlling SAGA-responsive genes’, *Molecular Genetics and Genomics*. Springer, 279(4), pp. 323–337. doi: 10.1007/S00438-007-0314-1/TABLES/3.

Dang, F., Nie, L. and Wei, W. (2020) ‘Ubiquitin signaling in cell cycle control and tumorigenesis’, *Cell Death & Differentiation* 2020 28:2. Nature Publishing Group, 28(2), pp. 427–438. doi: 10.1038/s41418-020-00648-0.

Davis-Kaplan, S. R. *et al.* (2006) ‘PKR1 Encodes an Assembly Factor for the Yeast V-Type ATPase’, *Journal of Biological Chemistry*. Elsevier BV, 281(42), pp. 32025–32035. doi: 10.1016/s0021-9258(19)84116-9.

Dederer, V. *et al.* (2019) ‘Cooperation of mitochondrial and er factors in quality control of tail-anchored proteins’, *eLife*. eLife Sciences Publications Ltd, 8. doi: 10.7554/ELIFE.45506.

Dederer, V. and Lemberg, M. K. (2021) ‘Transmembrane dislocases: a second chance for protein targeting’, *Trends in Cell Biology*. Elsevier, 31(11), pp. 898–911. doi: 10.1016/J.TCB.2021.05.007.

- Dephoure, N. *et al.* (2014) ‘Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast’, *eLife*. eLife Sciences Publications Limited, 3(2014), pp. 1–27. doi: 10.7554/eLife.03023.
- Deshaies, R. J. and Joazeiro, C. A. P. (2009) ‘RING Domain E3 Ubiquitin Ligases’, *Annual Review of Biochemistry*. Annual Reviews, 78, pp. 399–434. doi: 10.1146/ANNUREV.BIOCHEM.78.101807.093809.
- Després, P. C. *et al.* (2020) ‘Perturbing proteomes at single residue resolution using base editing’, *Nature Communications* 2020 11:1. Nature Publishing Group, 11(1), pp. 1–13. doi: 10.1038/s41467-020-15796-7.
- Dezso, Z., Oltvai, Z. N. and Barabási, A. L. (2003) ‘Bioinformatics Analysis of Experimentally Determined Protein Complexes in the Yeast *Saccharomyces cerevisiae*’, *Genome Research*. Cold Spring Harbor Laboratory Press, 13(11), pp. 2450–2454. doi: 10.1101/GR.1073603.
- Dikic, I. and Schulman, B. A. (2022) ‘An expanded lexicon for the ubiquitin code’, *Nature Reviews Molecular Cell Biology* 2022. Nature Publishing Group, pp. 1–15. doi: 10.1038/s41580-022-00543-1.
- Dragičević, M. B. *et al.* (2020) ‘ragp: Pipeline for mining of plant hydroxyproline-rich glycoproteins with implementation in R’, *Glycobiology*. Oxford Academic, 30(1), pp. 19–35. doi: 10.1093/GLYCOB/CWZ072.
- Duina, A. A., Miller, M. E. and Keeney, J. B. (2014) ‘Budding Yeast for Budding Geneticists: A Primer on the *Saccharomyces cerevisiae* Model System’, *Genetics*. Oxford University Press, 197(1), p. 33. doi: 10.1534/GENETICS.114.163188.
- Eftekharzadeh, B. *et al.* (2018) ‘Tau Protein Disrupts Nucleocytoplasmic Transport in Alzheimer’s Disease’, *Neuron*. Cell Press, 99(5), pp. 925–940.e7. doi: 10.1016/J.NEURON.2018.07.039.
- Finley, D. *et al.* (2012) ‘The ubiquitin-proteasome system of *Saccharomyces cerevisiae*’, *Genetics*. Genetics Society of America, 192(2), pp. 319–360. doi: 10.1534/genetics.112.140467.
- Fisk, D. G. *et al.* (2006) ‘*Saccharomyces cerevisiae* S288C genome annotation: a working hypothesis’, *Yeast*. John Wiley & Sons, Ltd, 23(12), pp. 857–865. doi: 10.1002/YEA.1400.
- Fung, J. J., Blöcher-Juárez, K. and Khmelinskii, A. (2022) ‘High-Throughput Analysis of Protein Turnover with Tandem Fluorescent Protein Timers’, *Methods in Molecular Biology*. Humana Press Inc., 2378, pp. 85–100. doi: 10.1007/978-1-0716-1732-8_6/FIGURES/4.
- Geiger, T., Cox, J. and Mann, M. (2010) ‘Proteomic changes resulting from gene copy number variations in cancer cells’, *PLoS Genetics*, 6(9). doi: 10.1371/journal.pgen.1001090.
- Gelperin, D. M. *et al.* (2005) ‘Biochemical and genetic analysis of the yeast proteome with a movable ORF collection’, *Genes and Development*. Cold Spring Harbor Laboratory Press, 19(23), pp. 2816–2826. doi: 10.1101/gad.1362105.
- George, A. J. *et al.* (2018) ‘A comprehensive atlas of E3 ubiquitin ligase mutations in neurological disorders’, *Frontiers in Genetics*. Frontiers Media S.A., 9(FEB), p. 29. doi: 10.3389/FGENE.2018.00029/FULL.
- Ghaemmaghami, S. *et al.* (2003) ‘Global analysis of protein expression in yeast’, *Nature* 2003 425:6959. Nature Publishing Group, 425(6959), pp. 737–741. doi: 10.1038/nature02046.
- Giaever, G. *et al.* (2002) ‘Functional profiling of the *Saccharomyces cerevisiae* genome’, *Nature* 2002 418:6896. Nature Publishing Group, 418(6896), pp. 387–391. doi: 10.1038/nature00935.

- Gloge, F. *et al.* (2014) ‘Co-translational mechanisms of protein maturation’, *Current Opinion in Structural Biology*. Elsevier Current Trends, pp. 24–33. doi: 10.1016/j.sbi.2013.11.004.
- Groettrup, M. *et al.* (2008) ‘Activating the ubiquitin family: UBA6 challenges the field’, *Trends in Biochemical Sciences*. Elsevier Current Trends, 33(5), pp. 230–237. doi: 10.1016/J.TIBS.2008.01.005.
- Grousl, T. *et al.* (2018) ‘A prion-like domain in Hsp42 drives chaperone-facilitated aggregation of misfolded proteins.’, *The Journal of cell biology*. Rockefeller University Press, 217(4), pp. 1269–1285. doi: 10.1083/jcb.201708116.
- Harper, J. W. and Schulman, B. A. (2021) ‘Cullin-RING Ubiquitin Ligase Regulatory Circuits: A Quarter Century Beyond the F-Box Hypothesis’, *Annual Review of Biochemistry*. Annual Reviews, 90, pp. 403–429. doi: 10.1146/ANNUREV-BIOCHEM-090120-013613.
- Hart, G. T., Lee, I. and Marcotte, E. R. (2007) ‘A high-accuracy consensus map of yeast protein complexes reveals modular nature of gene essentiality’, *BMC Bioinformatics*. BioMed Central, 8, p. 236. doi: 10.1186/1471-2105-8-236.
- Hart, T. *et al.* (2014) ‘Measuring error rates in genomic perturbation screens: gold standards for human functional genomics’, *Molecular Systems Biology*. John Wiley & Sons, Ltd, 10(7), p. 733. doi: 10.15252/MSB.20145216.
- Hartl, F. U., Bracher, A. and Hayer-Hartl, M. (2011) ‘Molecular chaperones in protein folding and proteostasis’, *Nature*. Nature Publishing Group, 475(7356), pp. 324–332. doi: 10.1038/nature10317.
- Hatakeyama, S. and Nakayama, K. I. I. (2003) ‘U-box proteins as a new family of ubiquitin ligases’, *Biochemical and Biophysical Research Communications*. Academic Press, 302(4), pp. 635–645. doi: 10.1016/S0006-291X(03)00245-6.
- Hegde, R. S. (2014) ‘Msp1: patrolling mitochondria for lost proteins’, *The EMBO Journal*. Wiley-Blackwell, 33(14), pp. 1509–1510. doi: 10.15252/embj.201488930.
- Hegde, R. S. and Zavodszky, E. (2019) ‘Recognition and Degradation of Mislocalized Proteins in Health and Disease’, *Cold Spring Harbor Perspectives in Biology*. Cold Spring Harbor Laboratory Press, 11(11), p. a033902. doi: 10.1101/CSHPERSPECT.A033902.
- Hessa, T. *et al.* (2011) ‘Protein targeting and degradation are coupled for elimination of mislocalized proteins’, *Nature*. NIH Public Access, 475(7356), pp. 394–399. doi: 10.1038/nature10181.
- Higgs, D. R. (2013) ‘The Molecular Basis of α -Thalassemia’, *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 3(1). doi: 10.1101/CSHPERSPECT.A011718.
- Higuchi-Sanabria, R. *et al.* (2018) ‘A Futile Battle? Protein Quality Control and the Stress of Aging’, *Developmental Cell*. Cell Press, 44(2), pp. 139–163. doi: 10.1016/J.DEVCEL.2017.12.020.
- Hill, S. M., Hanzén, S. and Nyström, T. (2017) ‘Restricted access: spatial sequestration of damaged proteins during stress and aging’, *EMBO Reports*. EMBO Press, 18(3), pp. 377–391. doi: 10.15252/embr.201643458.
- Ho, C. H. *et al.* (2009) ‘A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds’, *Nature Biotechnology*. NIH Public Access, 27(4), pp. 369–377. doi: 10.1038/nbt.1534.
- Ho, C. H. (2011) *Molecular barcoded plasmid yeast ORF library: Linking bioactive compounds to their cellular targets and mapping dosage suppressor networks*. Available at: <https://tspace.library.utoronto.ca/handle/1807/29745> (Accessed: 2 June 2018).

- Hu, X. *et al.* (2020) ‘RNF126-Mediated Reubiquitination Is Required for Proteasomal Degradation of p97-Extracted Membrane Proteins’, *Molecular Cell*. Cell Press, 79(2), pp. 320-331.e9. doi: 10.1016/j.molcel.2020.06.023.
- Hung, M.-C. and Link, W. (2011) ‘Protein localization in disease and therapy.’, *Journal of cell science*. The Company of Biologists Ltd, 124(Pt 20), pp. 3381–92. doi: 10.1242/jcs.089110.
- Hwang, C. S., Shemorry, A. and Varshavsky, A. (2010) ‘N-terminal acetylation of cellular proteins creates specific degradation signals’, *Science*. American Association for the Advancement of Science, 327(5968), pp. 973–977. doi: 10.1126/SCIENCE.1183147/SUPPL_FILE/HWANG.SOM.PDF.
- Iconomou, M. and Saunders, D. N. (2016) ‘Systematic approaches to identify E3 ligase substrates’, *Biochemical Journal*. Portland Press Ltd, 473(22), pp. 4083–4101. doi: 10.1042/BCJ20160719.
- Inglis, A. J. *et al.* (2020) ‘Differential Modes of Orphan Subunit Recognition for the WRB/CAML Complex’, *Cell Reports*. Elsevier B.V., 30(11), pp. 3691-3698.e5. doi: 10.1016/j.celrep.2020.02.084.
- Ishikawa, K. *et al.* (2017) ‘Post-Translational Dosage Compensation Buffers Genetic Perturbations to Stoichiometry of Protein Complexes’, *PLoS Genetics*, 13(1), p. e1006554. doi: 10.1371/journal.pgen.1006554.
- Ishikawa, K., Ishihara, A. and Moriya, H. (2020) ‘Exploring the Complexity of Protein-Level Dosage Compensation that Fine-Tunes Stoichiometry of Multiprotein Complexes’, *PLoS Genetics*. Edited by M. Jovanovic. Public Library of Science, 16(10), p. e1009091. doi: 10.1371/journal.pgen.1009091.
- Janke, C. *et al.* (2004) ‘A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes’, *Yeast*. John Wiley & Sons, Ltd, 21(11), pp. 947–962. doi: 10.1002/YEA.1142.
- Joazeiro, C. A. P. (2019) ‘Mechanisms and functions of ribosome-associated protein quality control’, *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, p. 1. doi: 10.1038/s41580-019-0118-2.
- Juszkiewicz, S. and Hegde, R. S. (2018) ‘Quality Control of Orphaned Proteins’, *Molecular Cell*. Cell Press, 71(3), pp. 443–457. doi: 10.1016/J.MOLCEL.2018.07.001.
- Kaganovich, D., Kopito, R. and Frydman, J. (2008) ‘Misfolded proteins partition between two distinct quality control compartments’, *Nature*. Nature Publishing Group, 454(7208), pp. 1088–1095. doi: 10.1038/nature07195.
- Käll, L., Krogh, A. and Sonnhammer, E. L. L. (2004) ‘A combined transmembrane topology and signal peptide prediction method’, *Journal of Molecular Biology*. Academic Press, 338(5), pp. 1027–1036. doi: 10.1016/j.jmb.2004.03.016.
- Kandasamy, G., Pradhan, A. K. and Palanimurugan, R. (2021) ‘Ccr4-Not complex subunits Ccr4, Caf1, and Not4 are novel proteolysis factors promoting the degradation of ubiquitin-dependent substrates by the 26S proteasome’, *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. Elsevier BV, 1868(6), p. 119010. doi: 10.1016/j.bbamcr.2021.119010.
- Kane, P. M., Tarsio, M. and Jianzhong, L. (1999) ‘Early Steps in Assembly of the Yeast Vacuolar H⁺-ATPase’, *Journal of Biological Chemistry*. Elsevier, 274(24), pp. 17275–17283. doi: 10.1074/JBC.274.24.17275.
- Karim, A. S., Curran, K. A. and Alper, H. S. (2013) ‘Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications.’, *FEMS yeast research*. NIH Public Access, 13(1), pp. 107–16. doi: 10.1111/1567-1364.12016.

- Kassem, S., Villanyi, Z. and Collart, M. A. (2017) 'Not5-dependent co-translational assembly of Ada2 and Spt20 is essential for functional integrity of SAGA', *Nucleic Acids Research*, 45(3), pp. 1186–1199. doi: 10.1093/nar/gkw1059.
- Kats, I. *et al.* (2018) 'Mapping Degradation Signals and Pathways in a Eukaryotic N-terminome', *Molecular Cell*. Cell Press, 70(3), pp. 488–501.e5. doi: 10.1016/J.MOLCEL.2018.03.033.
- Kelley, J. B. *et al.* (2011) 'The Defective Nuclear Lamina in Hutchinson-Gilford Progeria Syndrome Disrupts the Nucleocytoplasmic Ran Gradient and Inhibits Nuclear Localization of Ubc9', *Molecular and Cellular Biology*. American Society for Microbiology, 31(16), pp. 3378–3395. doi: 10.1128/MCB.05087-11/ASSET/F930BC57-6506-4432-BB54-2517C2959A8F/ASSETS/GRAPHIC/ZMB9991091680014.JPEG.
- Keskeny, C. *et al.* (2019) 'A conserved CAF40-binding motif in metazoan NOT4 mediates association with the CCR4–NOT complex', *Genes & Development*. Cold Spring Harbor Laboratory Press, 33(3–4), pp. 236–252. doi: 10.1101/GAD.320952.118.
- Khmelinskii, A. *et al.* (2011) 'Seamless Gene Tagging by Endonuclease-Driven Homologous Recombination', *PLOS ONE*. Public Library of Science, 6(8), p. e23794. doi: 10.1371/JOURNAL.PONE.0023794.
- Khmelinskii, A. *et al.* (2012) 'Tandem fluorescent protein timers for in vivo analysis of protein dynamics', *Nature Biotechnology*. Nature Publishing Group, 30(7), pp. 708–714. doi: 10.1038/nbt.2281.
- Khmelinskii, A. *et al.* (2016) 'Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers.', *Molecular biology of the cell*. American Society for Cell Biology, 27(2), pp. 360–70. doi: 10.1091/mbc.E15-07-0525.
- Klaips, C. L., Jayaraj, G. G. and Hartl, F. U. (2018) 'Pathways of cellular proteostasis in aging and disease', *Journal of Cell Biology*, 217(1), pp. 51–63. doi: 10.1083/jcb.201709072.
- Knaus, B. J. and Grünwald, N. J. (2017) 'vcfr: a package to manipulate and visualize variant call format data in R', *Molecular Ecology Resources*. John Wiley & Sons, Ltd, 17(1), pp. 44–53. doi: 10.1111/1755-0998.12549.
- Kong, K. Y. E. *et al.* (2021) 'Quality control of mislocalized and orphan proteins', *Experimental Cell Research*. Academic Press, 403(2), p. 112617. doi: 10.1016/j.yexcr.2021.112617.
- Koyuncu, S. *et al.* (2021) 'Rewiring of the ubiquitinated proteome determines ageing in *C. elegans*', *Nature* 2021 596:7871. Nature Publishing Group, 596(7871), pp. 285–290. doi: 10.1038/s41586-021-03781-z.
- Kulathu, Y. and Komander, D. (2012) 'Atypical ubiquitylation — the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages', *Nature Reviews Molecular Cell Biology* 2012 13:8. Nature Publishing Group, 13(8), pp. 508–523. doi: 10.1038/nrm3394.
- Kumar, A., Mathew, V. and Stirling, P. C. (2022) 'Nuclear protein quality control in yeast: The latest INquiries', *Journal of Biological Chemistry*. Elsevier, 298(8), p. 102199. doi: 10.1016/J.JBC.2022.102199.
- Kumar, A. V. and Lapierre, L. R. (2021) 'Location, location, location: subcellular protein partitioning in proteostasis and aging', *Biophysical Reviews* 2021 13:6. Springer, 13(6), pp. 931–941. doi: 10.1007/S12551-021-00890-X.
- Kunze, M. and Berger, J. (2015) 'The similarity between N-terminal targeting signals for protein import into different organelles and its evolutionary relevance', *Frontiers in Physiology*. Frontiers

Media S.A., 6(SEP), p. 259. doi: 10.3389/FPHYS.2015.00259/BIBTEX.

Labbadia, J. and Morimoto, R. I. (2015) 'The Biology of Proteostasis in Aging and Disease', *Annual Review of Biochemistry*. Annual Reviews, 84(1), pp. 435–464. doi: 10.1146/annurev-biochem-060614-033955.

Van Der Lee, R. *et al.* (2014) 'Classification of intrinsically disordered regions and proteins', *Chemical Reviews*. American Chemical Society, 114(13), pp. 6589–6631. doi: 10.1021/CR400525M/ASSET/IMAGES/CR400525M.SOCIAL.JPEG_V03.

Levine, C. G. *et al.* (2005) 'The efficiency of protein compartmentalization into the secretory pathway', *Molecular Biology of the Cell*. American Society for Cell Biology, 16(1), pp. 279–291. doi: 10.1091/MBC.E04-06-0508/ASSET/IMAGES/LARGE/ZMK0010529570008.JPEG.

Li, M. *et al.* (2010) 'The ATM-p53 pathway suppresses aneuploidy-induced tumorigenesis', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 107(32), pp. 14188–14193. doi: 10.1073/PNAS.1005960107/SUPPL_FILE/PNAS.201005960SI.PDF.

Liguori, I. *et al.* (2018) 'Oxidative stress, aging, and diseases', *Clinical Interventions in Aging*. Dove Press, 13, p. 757. doi: 10.2147/CIA.S158513.

Lin, H.-C. *et al.* (2018) 'C-Terminal End-Directed Protein Elimination by CRL2 Ubiquitin Ligases.', *Molecular cell*. Elsevier, 70(4), pp. 602–613.e3. doi: 10.1016/j.molcel.2018.04.006.

Liu, H. Y. *et al.* (1998) 'The NOT proteins are part of the CCR4 transcriptional complex and affect gene expression both positively and negatively', *The EMBO Journal*. John Wiley & Sons, Ltd, 17(4), pp. 1096–1106. doi: 10.1093/EMBOJ/17.4.1096.

Livak, K. J. and Schmittgen, T. D. (2001) 'Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method', *Methods*. Academic Press, 25(4), pp. 402–408. doi: 10.1006/METH.2001.1262.

Lu, J. *et al.* (2021) 'Types of nuclear localization signals and mechanisms of protein import into the nucleus', *Cell Communication and Signaling*. BioMed Central Ltd, 19(1), pp. 1–10. doi: 10.1186/S12964-021-00741-Y/FIGURES/1.

Macossay-Castillo, M. *et al.* (2019) 'The Balancing Act of Intrinsically Disordered Proteins: Enabling Functional Diversity while Minimizing Promiscuity', *Journal of Molecular Biology*. Academic Press, 431(8), pp. 1650–1670. doi: 10.1016/J.JMB.2019.03.008.

Magtanong, L. *et al.* (2011) 'Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell', *Nature Biotechnology*. Nature Publishing Group, 29(6), pp. 505–511. doi: 10.1038/nbt.1855.

Maillet, L. *et al.* (2000) 'The essential function of not1 lies within the Ccr4-not complex', *Journal of Molecular Biology*. Academic Press, 303(2), pp. 131–143. doi: 10.1006/JMBI.2000.4131.

Malkus, P. *et al.* (2004) 'Role of Vma21p in assembly and transport of the yeast vacuolar ATPase', *Molecular Biology of the Cell*. American Society for Cell Biology, 15(11), pp. 5075–5091. doi: 10.1091/MBC.E04-06-0514/ASSET/IMAGES/LARGE/ZMK0110428780010.JPEG.

Mariappan, M. *et al.* (2010) 'A ribosome-associating factor chaperones tail-anchored membrane proteins', *Nature* 2010 466:7310. Nature Publishing Group, 466(7310), pp. 1120–1124. doi: 10.1038/nature09296.

Mazhab-Jafari, M. T. *et al.* (2016) 'Atomic model for the membrane-embedded VO motor of a

- eukaryotic V-ATPase', *Nature*. Europe PMC Funders, 539(7627), p. 118. doi: 10.1038/NATURE19828.
- McDowell, M. A. *et al.* (2020) 'Structural Basis of Tail-Anchored Membrane Protein Biogenesis by the GET Insertase Complex', *Molecular cell*. NLM (Medline), 80(1), pp. 72–86. doi: 10.1016/j.molcel.2020.08.012.
- Mcshane, E. *et al.* (2016) 'Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation Kinetic Analysis of Protein Stability Reveals Age-Dependent Degradation', *Cell*. Elsevier Inc, 167, pp. 803–815. doi: 10.1016/j.cell.2016.09.015.
- Meldal, B. H. M. *et al.* (2015) 'The complex portal--an encyclopaedia of macromolecular complexes.', *Nucleic acids research*. Oxford University Press, 43(Database issue), pp. D479-84. doi: 10.1093/nar/gku975.
- Meldal, B. H. M. *et al.* (2021) 'Analysing the yeast complexome—the Complex Portal rising to the challenge', *Nucleic Acids Research*. Oxford Academic, 49(6), pp. 3156–3167. doi: 10.1093/NAR/GKAB077.
- Meurer, M. *et al.* (2018) 'Genome-wide C-SWAT library for high-throughput yeast genome tagging', *Nature Methods*. Nature Publishing Group, 15(8), p. 1. doi: 10.1038/s41592-018-0045-8.
- Meyer, K. *et al.* (2018) 'Mutations in Disordered Regions Can Cause Disease by Creating Dileucine Motifs', *Cell*. Cell Press, 175(1), pp. 239-253.e17. doi: 10.1016/J.CELL.2018.08.019.
- Miller-Vedam, L. E. *et al.* (2020) 'Structural and mechanistic basis of the EMC-dependent biogenesis of distinct transmembrane clients', *eLife*. eLife Sciences Publications Ltd, 9, pp. 1–124. doi: 10.7554/ELIFE.62611.
- Miller, S. B. M. *et al.* (2015) 'Compartment-specific aggregates direct distinct nuclear and cytoplasmic aggregate deposition.', *The EMBO journal*, 34(6), pp. 778–97. doi: 10.15252/embj.201489524.
- Miller, S. B. M., Mogk, A. and Bukau, B. (2015) 'Spatially Organized Aggregation of Misfolded Proteins as Cellular Stress Defense Strategy', *Journal of Molecular Biology*. Academic Press, 427(7), pp. 1564–1574. doi: 10.1016/J.JMB.2015.02.006.
- Mueller, S. *et al.* (2015) 'Protein degradation corrects for imbalanced subunit stoichiometry in OST complex assembly.', *Molecular Biology of the Cell*. American Society for Cell Biology, 26(14), pp. 2596–608. doi: 10.1091/mbc.E15-03-0168.
- Mumberg, D., Müller, R. and Funk, M. (1995) 'Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds', *Gene*. Elsevier, 156(1), pp. 119–122. doi: 10.1016/0378-1119(95)00037-7.
- Nasertorabi, F. *et al.* (2011) 'Insights into the structure of the CCR4-NOT complex by electron microscopy', *FEBS Letters*. John Wiley & Sons, Ltd, 585(14), pp. 2182–2186. doi: 10.1016/J.FEBSLET.2011.05.071.
- O'donnell, J. P. *et al.* (2020) 'The architecture of EMC reveals a path for membrane protein insertion', *eLife*. eLife Sciences Publications Ltd, 9, pp. 1–56. doi: 10.7554/eLife.57887.
- Okreglak, V. and Walter, P. (2014) 'The conserved AAA-ATPase Msp1 confers organelle specificity to tail-anchored proteins.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 111(22), pp. 8019–24. doi: 10.1073/pnas.1405755111.
- Omura, T. (1998) 'Mitochondria-targeting sequence, a multi-role sorting sequence recognized at all

- steps of protein import into mitochondria', *Journal of biochemistry*. J Biochem, 123(6), pp. 1010–1016. doi: 10.1093/OXFORDJOURNALS.JBCHEM.A022036.
- Oot, R. A. *et al.* (2016) 'Crystal structure of yeast V1-ATPase in the autoinhibited state', *The EMBO Journal*. John Wiley & Sons, Ltd, 35(15), pp. 1694–1706. doi: 10.15252/EMBJ.201593447.
- Oromendia, A. B. and Amon, A. (2014) 'Aneuploidy: implications for protein homeostasis and disease', *Disease Models & Mechanisms*. David H. Koch Institute for Integrative Cancer Research, Department of Biology and Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA., 7(1), pp. 15–20. doi: 10.1242/dmm.013391.
- Oromendia, A. B., Dodgson, S. E. and Amon, A. (2012) 'Aneuploidy causes proteotoxic stress in yeast', *Genes Dev*. Koch Institute for Integrative Cancer Research, Department of Biology, Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02138, USA., 26, pp. 2696–2708. doi: 10.1101/gad.207407.112.
- Panasenko, O. O. *et al.* (2019) 'Co-translational assembly of proteasome subunits in NOT1-containing assemblyosomes', *Nature Structural and Molecular Biology*. Nature Publishing Group, 26(2), pp. 110–120. doi: 10.1038/s41594-018-0179-5.
- Pelham, H. R. B. (1990) 'The retention signal for soluble proteins of the endoplasmic reticulum', *Trends in biochemical sciences*. Trends Biochem Sci, 15(12), pp. 483–486. doi: 10.1016/0968-0004(90)90303-S.
- Pillet, B. *et al.* (2022) 'Dedicated chaperones coordinate cotranslational regulation of ribosomal protein production with ribosome assembly to preserve proteostasis', *eLife*. eLife Sciences Publications Ltd, 11. doi: 10.7554/ELIFE.74255.
- Pleiner, T. *et al.* (2021) 'WNK1 is an assembly factor for the human ER membrane protein complex', *Molecular Cell*, 81(13), pp. 2693–2704.e12. doi: 10.1016/J.MOLCEL.2021.04.013.
- Prakash, S. *et al.* (2004) 'An unstructured initiation site is required for efficient proteasome-mediated degradation', *Nature Structural & Molecular Biology 2004 11:9*. Nature Publishing Group, 11(9), pp. 830–837. doi: 10.1038/nsmb814.
- Prasad, R., Kawaguchi, S. and Ng, D. T. W. (2010) 'A nucleus-based quality control mechanism for cytosolic proteins', *Molecular Biology of the Cell*. The American Society for Cell Biology, 21(13), pp. 2117–2127. doi: 10.1091/MBC.E10-02-0111/ASSET/IMAGES/LARGE/ZMK0131095040008.JPEG.
- R Core Team (2021) 'R: A Language and Environment for Statistical Computing'. Vienna, Austria. Available at: <https://www.r-project.org/>.
- Rinaldi, C., Mäger, I. and Wood, M. J. (2016) 'Proteostasis and Diseases of the Motor Unit', *Frontiers in Molecular Neuroscience*. Frontiers, 0, p. 164. doi: 10.3389/FNMOL.2016.00164.
- Robinson, J. T. *et al.* (2011) 'Integrative genomics viewer', *Nature Biotechnology 2011 29:1*. Nature Publishing Group, 29(1), pp. 24–26. doi: 10.1038/nbt.1754.
- Rodrigo-Brenni, M. C., Gutierrez, E. and Hegde, R. S. (2014) 'Cytosolic Quality Control of Mislocalized Proteins Requires RNF126 Recruitment to Bag6', *Molecular Cell*. Cell Press, 55(2), pp. 227–237. doi: 10.1016/J.MOLCEL.2014.05.025.
- Roy, B. *et al.* (2020) 'Autoregulation of yeast ribosomal proteins discovered by efficient search for feedback regulation', *Communications Biology*. Nature Publishing Group, 3(1), pp. 1–9. doi: 10.1038/s42003-020-01494-z.

- Ryan, M., Graham, L. A. and Stevens, T. H. (2008) 'Voa1p functions in V-ATPase assembly in the yeast endoplasmic reticulum', *Molecular Biology of the Cell*. The American Society for Cell Biology, 19(12), pp. 5131–5142. doi: 10.1091/MBC.E08-06-0629/ASSET/IMAGES/LARGE/ZMK0120887900010.JPEG.
- Sala, A. J., Bott, L. C. and Morimoto, R. I. (2017) 'Shaping proteostasis at the cellular, tissue, and organismal level.', *The Journal of cell biology*. Rockefeller University Press, 216(5), pp. 1231–1241. doi: 10.1083/jcb.201612111.
- Sardana, R. and Emr, S. D. (2021) 'Membrane Protein Quality Control Mechanisms in the Endo-Lysosome System', *Trends in Cell Biology*. Elsevier Ltd, pp. 269–283. doi: 10.1016/j.tcb.2020.11.011.
- Schauberger, P. and Walker, A. (2021) 'openxlsx: Read, Write and Edit xlsx Files'. Available at: <https://cran.r-project.org/package=openxlsx>.
- Schmittgen, T. D. and Livak, K. J. (2008) 'Analyzing real-time PCR data by the comparative CT method', *Nature Protocols*. Nature Publishing Group, 3(6), pp. 1101–1108. doi: 10.1038/nprot.2008.73.
- Schneider, K. and Bertolotti, A. (2015) 'Surviving protein quality control catastrophes--from cells to organisms.', *Journal of cell science*. The Company of Biologists Ltd, 128(21), pp. 3861–9. doi: 10.1242/jcs.173047.
- Schuldiner, M. *et al.* (2008) 'The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane', *Cell*. Cell Press, 134(4), pp. 634–645. doi: 10.1016/J.CELL.2008.06.025.
- Schwarz, A. and Beck, M. (2019) 'The Benefits of Cotranslational Assembly: A Structural Perspective', *Trends in Cell Biology*. Elsevier Ltd, pp. 791–803. doi: 10.1016/j.tcb.2019.07.006.
- Seidel, M. *et al.* (2022) 'Co-translational assembly orchestrates competing biogenesis pathways', *Nature Communications 2022 13:1*. Nature Publishing Group, 13(1), pp. 1–15. doi: 10.1038/s41467-022-28878-5.
- Shakya, V. P. *et al.* (2021) 'A nuclear-based quality control pathway for non-imported mitochondrial proteins', *eLife*. doi: 10.7554/eLife.61230.
- Shaner, N. C. *et al.* (2013) 'A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*', *Nature Methods*. Nature Publishing Group, 10(5), pp. 407–409. doi: 10.1038/nmeth.2413.
- Sheltzer, J. M. *et al.* (2011) 'Aneuploidy drives genomic instability in yeast', *Science*. David H. Koch Institute for Integrative Cancer Research and Howard Hughes Medical Institute (HHMI), Massachusetts Institute of Technology, Cambridge, MA 02139, USA., 333, pp. 1026–1030. doi: 10.1126/science.1206412.
- Shemorry, A., Hwang, C.-S. and Varshavsky, A. (2013) 'Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway', *Molecular Cell*. Cell Press, 50(4), pp. 540–551. doi: 10.1016/J.MOLCEL.2013.03.018.
- Shiber, A. *et al.* (2018) 'Cotranslational assembly of protein complexes in eukaryotes revealed by ribosome profiling', *Nature*. Nature Publishing Group, 561(7722), pp. 268–272. doi: 10.1038/s41586-018-0462-y.
- Smardon, A. M. *et al.* (2014) 'The RAVE complex is an isoform-specific V-ATPase assembly factor in yeast', *Molecular Biology of the Cell*. The American Society for Cell Biology, 25(3), pp. 356–367. doi: 10.1091/MBC.E13-05-0231/ASSET/IMAGES/LARGE/356FIG8.JPEG.

- Smardon, A. M., Tarsio, M. and Kane, P. M. (2002) ‘The RAVE Complex Is Essential for Stable Assembly of the Yeast V-ATPase’, *Journal of Biological Chemistry*. Elsevier, 277(16), pp. 13831–13839. doi: 10.1074/JBC.M200682200.
- Song, J., Herrmann, J. M. and Becker, T. (2020) ‘Quality control of the mitochondrial proteome’, *Nature Reviews Molecular Cell Biology* 2020 22:1. Nature Publishing Group, 22(1), pp. 54–70. doi: 10.1038/s41580-020-00300-2.
- Song, L. and Luo, Z. Q. (2019) ‘Post-translational regulation of ubiquitin signaling’, *Journal of Cell Biology*. The Rockefeller University Press, 218(6), pp. 1776–1786. doi: 10.1083/JCB.201902074.
- Sontag, E. M., Samant, R. S. and Frydman, J. (2017) ‘Mechanisms and Functions of Spatial Protein Quality Control’, *Annual Review of Biochemistry*. Annual Reviews , 86, pp. 97–122. doi: 10.1146/ANNUREV-BIOCHEM-060815-014616.
- Springer, M., Weissman, J. S. and Kirschner, M. W. (2010) ‘A general lack of compensation for gene dosage in yeast’, *Molecular Systems Biology*. EMBO Press, 6(1), p. 368. doi: 10.1038/msb.2010.19.
- Stingele, S. *et al.* (2012) ‘Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells’, *Molecular Systems Biology*. European Molecular Biology Organization, 8(1110), p. 608. doi: 10.1038/msb.2012.40.
- Stringer, C. *et al.* (2021) ‘Cellpose: a generalist algorithm for cellular segmentation’, *Nature Methods*. Nature Research, 18(1), pp. 100–106. doi: 10.1038/s41592-020-01018-x.
- Suk, T. R. and Rousseaux, M. W. C. (2020) ‘The role of TDP-43 mislocalization in amyotrophic lateral sclerosis’, *Molecular Neurodegeneration*. BioMed Central Ltd, 15(1), pp. 1–16. doi: 10.1186/S13024-020-00397-1/FIGURES/2.
- Sun, Z. and Brodsky, J. L. (2019) ‘Protein quality control in the secretory pathway’, *Journal of Cell Biology*. Rockefeller University Press, pp. 3171–3187. doi: 10.1083/jcb.201906047.
- Sung, M. K., Porras-Yakushi, T. R., *et al.* (2016) ‘A conserved quality-control pathway that mediates degradation of unassembled ribosomal proteins’, *Elife*. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, United States. Proteome Exploration Laboratory, Division of Biology and Biological Engineering, Beckman Institute, California Institute of Technology, Pasadena, U, 5. doi: 10.7554/eLife.19105.
- Sung, M. K., Reitsma, J. M., *et al.* (2016) ‘Ribosomal proteins produced in excess are degraded by the ubiquitin-proteasome system’, *Molecular Biology of the Cell*. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125. Proteome Exploration Laboratory, Division of Biology and Biological Engineering, Beckman Institute, California Institute of Technology, Pasadena, CA 91, 27(17), pp. 2642–2652. doi: 10.1091/mbc.E16-05-0290.
- Suresh, H. G., Pascoe, N. and Andrews, B. (2020) ‘The structure and function of deubiquitinases: lessons from budding yeast’, *Open Biology*. The Royal Society, 10(10). doi: 10.1098/RSOB.200279.
- Suzuki, R. and Kawahara, H. (2016) ‘UBQLN4 recognizes mislocalized transmembrane domain proteins and targets these to proteasomal degradation’, *EMBO reports*. John Wiley & Sons, Ltd, 17(6), pp. 842–857. doi: 10.15252/EMBR.201541402.
- Swaney, D. L. *et al.* (2013) ‘Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation’, *Nature Methods* 2013 10:7. Nature Publishing Group, 10(7), pp. 676–682. doi: 10.1038/nmeth.2519.
- Tachikawa, M. *et al.* (2012) ‘Mislocalization of fukutin protein by disease-causing missense mutations can be rescued with treatments directed at folding amelioration’, *Journal of Biological*

Chemistry. Elsevier, 287(11), pp. 8398–8406. doi: 10.1074/jbc.M111.300905.

Taggart, J. C. *et al.* (2020) ‘Keeping the Proportions of Protein Complex Components in Check’, *Cell Systems*. Cell Press, pp. 125–132. doi: 10.1016/j.cels.2020.01.004.

Taggart, J. C. and Li, G.-W. (2018) ‘Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes’, *Cell Systems*. Cell Press, 7(6), pp. 580–589.e4. doi: 10.1016/J.CELS.2018.11.003.

Timms, R. T. and Koren, I. (2020) ‘Tying up loose ends: The N-degron and C-degron pathways of protein degradation’, *Biochemical Society Transactions*. Portland Press Ltd, pp. 1557–1567. doi: 10.1042/BST20191094.

Tomashek, J. J., Garrison, B. S. and Klionsky, D. J. (1997) ‘Reconstitution in Vitro of the V1 Complex from the Yeast Vacuolar Proton-translocating ATPase: ASSEMBLY RECAPITULATES MECHANISM’, *Journal of Biological Chemistry*. Elsevier, 272(26), pp. 16618–16623. doi: 10.1074/JBC.272.26.16618.

Tong, A. H. Y. *et al.* (2001) ‘Systematic genetic analysis with ordered arrays of yeast deletion mutants’, *Science*. American Association for the Advancement of Science, 294(5550), pp. 2364–2368. doi: 10.1126/science.1065810.

Tong, A. H. Y. and Boone, C. (2006) ‘Synthetic genetic array analysis in *Saccharomyces cerevisiae*.’, *Methods in molecular biology (Clifton, N.J.)*. Humana Press, Totowa, NJ, 313, pp. 171–192. doi: 10.1385/1-59259-958-3:171/FIGURES/7.

Tong, A. H. Y. and Boone, C. (2007) ‘High-Throughput Strain Construction and Systematic Synthetic Lethal Screening in *Saccharomyces cerevisiae*’, in *Methods in Microbiology*. Academic Press, pp. 369–707. doi: 10.1016/S0580-9517(06)36016-3.

Torres, E. M. *et al.* (2007) ‘Effects of aneuploidy on cellular physiology and cell division in haploid yeast’, *Science*. Center for Cancer Research, Howard Hughes Medical Institute, Massachusetts Institute of Technology, E17-233, 40 Ames Street, Cambridge, MA 02139, USA., 317, pp. 916–924. doi: 10.1126/science.1142210.

van der Lee, R. *et al.* (2014) ‘Intrinsically Disordered Segments Affect Protein Half-Life in the Cell and during Evolution’, *Cell Reports*. Cell Press, 8(6), pp. 1832–1844. doi: 10.1016/J.CELREP.2014.07.055.

Varshavsky, A. (2019) ‘N-degron and C-degron pathways of protein degradation’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(2), pp. 358–366. doi: 10.1073/PNAS.1816596116/SUPPL_FILE/PNAS.1816596116.SAPP.PDF.

Vasanthakumar, T. *et al.* (2019) ‘Structural comparison of the vacuolar and Golgi V-ATPases from *saccharomyces cerevisiae*’, *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 116(15), pp. 7272–7277. doi: 10.1073/PNAS.1814818116/-/DCSUPPLEMENTAL.

Volkmar, N. and Christianson, J. C. (2020) *Squaring the EMC - How promoting membrane protein biogenesis impacts cellular functions and organismal homeostasis*, *Journal of Cell Science*. Company of Biologists Ltd. Available at: <https://journals.biologists.com/jcs/article/133/8/jcs243519/224934/Squaring-the-EMC-how-promoting-membrane-protein> (Accessed: 22 April 2021).

Wagih, O. and Parts, L. (2014) ‘gitter: a robust and accurate method for quantification of colony sizes

- from plate images.’, *G3 (Bethesda, Md.)*. Genetics Society of America, 4(3), pp. 547–52. doi: 10.1534/g3.113.009431.
- Wallace, E. W. J. *et al.* (2015) ‘Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress’, *Cell*, 162(6), pp. 1286–1298. doi: 10.1016/j.cell.2015.08.041.
- Wang, X. and Li, S. (2014) ‘Protein mislocalization: Mechanisms, functions and clinical applications in cancer’, *Biochimica et Biophysica Acta - Reviews on Cancer*. Elsevier, pp. 13–25. doi: 10.1016/j.bbcan.2014.03.006.
- Wang, Y. *et al.* (2020) ‘HECT E3 ubiquitin ligases - Emerging insights into their biological roles and disease relevance’, *Journal of Cell Science*. Company of Biologists Ltd, 133(7). doi: 10.1242/JCS.228072/224840.
- Warnes, G. R. *et al.* (2022) ‘gplots: Various R Programming Tools for Plotting Data’. Available at: <https://cran.r-project.org/package=gplots>.
- Weidberg, H. and Amon, A. (2018) ‘MitoCPR—A surveillance pathway that protects mitochondria in response to protein import stress’, *Science*, 360(6385). doi: 10.1126/science.aan4146.
- Weill, U. *et al.* (2018) ‘Genome-wide SWAp-Tag yeast libraries for proteome exploration’, *Nature Methods*. Nature Publishing Group, 15(8), pp. 617–622. doi: 10.1038/s41592-018-0044-9.
- Weissman, A. M. (2001) ‘Themes and variations on ubiquitylation’, *Nature Reviews Molecular Cell Biology* 2001 2:3. Nature Publishing Group, 2(3), pp. 169–178. doi: 10.1038/35056563.
- Wells, J. N., Bergendahl, L. T. and Marsh, J. A. (2015) ‘Co-translational assembly of protein complexes’, *Biochemical Society Transactions*. Portland Press Ltd, 43, pp. 1221–1226. doi: 10.1042/BST20150159.
- Wickham, H. *et al.* (2019) ‘Welcome to the Tidyverse’, *Journal of Open Source Software*. The Open Journal, 4(43), p. 1686. doi: 10.21105/JOSS.01686.
- Williams, B. R. *et al.* (2008) ‘Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells’, *Science*. American Association for the Advancement of Science , 322(5902), pp. 703–709. doi: 10.1126/SCIENCE.1160058/SUPPL_FILE/WILLIAMS-SOM.PDF.
- Williams, N. K. and Dichtl, B. (2018) ‘Co-translational control of protein complex formation: A fundamental pathway of cellular organization?’, *Biochemical Society Transactions*. Portland Press Ltd, pp. 197–206. doi: 10.1042/BST20170451.
- Winzeler, E. A. *et al.* (1999) ‘Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis’, *Science*. American Association for the Advancement of Science, 285(5429), pp. 901–906. doi: 10.1126/SCIENCE.285.5429.901/SUPPL_FILE/1040380S1_THUMB.GIF.
- Wright, G. *et al.* (2001) ‘Oxidative Stress Inhibits the Mitochondrial Import of Preproteins and Leads to Their Degradation’, *Experimental Cell Research*. Academic Press, 263(1), pp. 107–117. doi: 10.1006/EXCR.2000.5096.
- De Wulf, P., McAinsh, A. D. and Sorger, P. K. (2003) ‘Hierarchical assembly of the budding yeast kinetochore from multiple subcomplexes’, *Genes and Development*, 17(23), pp. 2902–2921. doi: 10.1101/gad.1144403.
- Xiao, L. J. *et al.* (2011) ‘Expression of yeast high mobility group protein HMO1 is regulated by TOR signaling’, *Gene*. Elsevier, 489(1), pp. 55–62. doi: 10.1016/J.GENE.2011.08.017.
- Xie, Y. (2015) *Dynamic Documents with {R} and knitr*. 2nd edn. Boca Raton, Florida: Chapman and Hall/CRC. Available at: <https://yihui.org/knitr/>.

- Xie, Y. (2022) ‘knitr: A General-Purpose Package for Dynamic Report Generation in R’. Available at: <https://yihui.org/knitr/>.
- Yanagitani, K., Juszkievicz, S. and Hegde, R. S. (2017) ‘UBE2O is a quality control factor for orphans of multiprotein complexes.’, *Science (New York, N.Y.)*. American Association for the Advancement of Science, 357(6350), pp. 472–475. doi: 10.1126/science.aan0178.
- Yang, Q. *et al.* (2021) ‘E3 ubiquitin ligases: styles, structures and functions’, *Molecular Biomedicine* 2021 2:1. Springer, 2(1), pp. 1–17. doi: 10.1186/S43556-021-00043-2.
- Yeh, C. *et al.* (2021) ‘The C-degron pathway eliminates mislocalized proteins and products of deubiquitinating enzymes’, *The EMBO Journal*. EMBO, p. e105846. doi: 10.15252/embj.2020105846.
- Yen-Ting-Liu *et al.* (2014) ‘The partitioning and copy number control systems of the selfish yeast plasmid: an optimized molecular design for stable persistence in host cells’, *Microbiology spectrum*. NIH Public Access, 2(5). doi: 10.1128/MICROBIOLSPEC.PLAS-0003-2013.
- Yofe, I. *et al.* (2016) ‘One library to make them all: Streamlining the creation of yeast libraries via a SWAp-Tag strategy’, *Nature Methods*. Europe PMC Funders, 13(4), pp. 371–378. doi: 10.1038/nmeth.3795.
- Youle, R. J. and Narendra, D. P. (2010) ‘Mechanisms of mitophagy’, *Nature Reviews Molecular Cell Biology* 2011 12:1. Nature Publishing Group, 12(1), pp. 9–14. doi: 10.1038/nrm3028.
- Zakian, V. A., Brewer, B. J. and Fangman, W. L. (1979) ‘Replication of each copy of the yeast 2 micron DNA plasmid occurs during the S phase’, *Cell*. Elsevier, 17(4), pp. 923–934. doi: 10.1016/0092-8674(79)90332-5.
- Zhu, J. *et al.* (2012) ‘Karyotypic Determinants of Chromosome Instability in Aneuploid Budding Yeast’, *PLOS Genetics*. Public Library of Science, 8(5), p. e1002719. doi: 10.1371/JOURNAL.PGEN.1002719.
- Zhu, J. *et al.* (2018) ‘Cellular Stress Associated with Aneuploidy’, *Developmental Cell*. Elsevier Inc., 44(4), pp. 420–431. doi: 10.1016/j.devcel.2018.02.002.

ACKNOWLEDGEMENTS

CURRICULUM VITAE