

# **Asymmetric protein inheritance in yeast replicative aging**

## **Dissertation**

Zur Erlangung des Grades  
Doktor der Naturwissenschaften

Am Fachbereich Biologie  
der Johannes Gutenberg-Universität Mainz

vorgelegt von

**Karla Azucena Juárez Núñez**

geboren am 30.04.1992

in Michoacán, Mexiko

7. Dezember 2023, Mainz

Dekan: Prof. Dr. Eckhard Thines

1. Gutachter: Dr. Anton Khmelinskii

2. Gutachter: Prof. Dr. René Ketting

Tag der mündlichen Prüfung: 07.12.2023

## Table of Contents

Preface.....	6
List of publications.....	7
First author publications.....	7
Contributor in other research publications.....	7
Summary.....	8
Zusammenfassung.....	10
CHAPTER 1. GENOME-WIDE ASSESSMENT OF TAG POSITION IN YEAST FITNESS.....	12
INTRODUCTION.....	12
Fusion tags as a tool to study proteins of interest.....	12
Commonly used tags.....	13
AIM OF THIS STUDY.....	14
RESULTS.....	15
Liquid and agar based competition assays using FPs as tags.....	15
Using a non-fluorescent tag for competition assays.....	17
Creation of genome-wide Halo tagged libraries and their competition.....	19
Characteristics of N-winners and C-winners.....	26
Validation of tag presence and fitness effect.....	29
DISCUSSION.....	32
Liquid and agar based competition assays using FPs as tags.....	32
Using a non-fluorescent tag for competition assays.....	33
Creation of genome-wide Halo tagged libraries and their competition.....	34
Characteristics of N-winners and C-winners.....	37
Validation of tag presence and fitness effect.....	38
MATERIAL AND METHODS.....	41
Yeast competition assay in liquid media.....	41
Yeast competition assay on agar media.....	42
Measurement of cytosolic fluorescence of donor strains for the proof-of-principle (PoP).....	42
Creation of N and C-tagged libraries using Halo as a tag.....	43
Yeast whole cell extract using TCA.....	45
SUPPLEMENTARY.....	46
CHAPTER 2. ASYMMETRIC PROTEIN INHERITANCE IN YEAST REPLICATIVE AGING.....	49
INTRODUCTION.....	49

Aging and cellular senescence .....	49
The hallmarks of aging from mammals to single-celled organisms.....	50
Replicative aging and asymmetric cell division and in budding yeast.....	51
Tandem fluorescent proteins timers (tFTs) as a red-out of protein age .....	52
Protein-inheritance pattern in budding yeast .....	55
AIM OF THIS STUDY .....	56
Objectives.....	56
RESULTS .....	57
Selection of FP for tFTs and generation of mScarlet and mScarlet-H .....	57
Evaluation of different tFTs as reporters for protein degradation.....	57
Evaluation of different tFTs as reporters for protein protein-age based segregation .....	61
Generation of a donor strain with markers for image segmentation.....	62
Construction of genomewide C-tagged libraries and their validation .....	67
DISCUSSION.....	71
Evaluation of different tFTs as reporters for protein degradation.....	71
Evaluation of different tFTs as reporters for protein protein-age based segregation .....	73
Generation of a donor strain with markers for image segmentation.....	74
Construction of genomewide C-tagged libraries .....	75
MATERIAL AND METHODS.....	78
Molecular techniques used for multiple experiments .....	78
Preparation of budding yeast competent cells .....	78
Budding yeast transformation .....	79
Plasmid extraction from yeast strains .....	80
Yeast whole cell extract using TCA .....	80
Confocal fluorescence microscopy .....	81
Generation of plasmids carrying different tFTs .....	82
Evaluation of different tFTs as reporters for protein degradation.....	87
Construction of genome-wide yeast libraries .....	89
Measuring fluorescence signal using Flow cytometry.....	92
Budding index.....	92
Alpha-factor G1 synchronization.....	93
Spotting assays .....	96
SUPPLEMENTARY .....	97

Appendix I .....	101
List of abbreviations .....	101
REFERENCES .....	103
Appendix II .....	<b>Error! Bookmark not defined.</b>
Acknowledgements.....	<b>Error! Bookmark not defined.</b>
Curriculum vitae.....	<b>Error! Bookmark not defined.</b>

## **Preface**

I hereby declare that I am the sole author and composer of this report and that no other source, other than those listed, have been used. Furthermore, I declare that I have acknowledged the work of others by providing detailed references of their cited work.

Throughout this study, I have received support for experimental design and data analysis and interpretation from my supervisor Dr. Anton Khmelinskii. For the efficient analysis of the imaged genome-wide library of ORF-tFT described in Chapter 2, we created an automated image segmentation pipeline in collaboration with Mohammed Ali, Danver Hans Värvi and Leopold Parts from the University of Tartu, Estonia. The results of this analysis are still pending until receiving complete raw output files from our collaborators, and therefore nothing shown in this thesis was evaluated using such a pipeline. Nevertheless, I would like to acknowledge all of the continuous work from the collaborators to finalize this automated image analysis pipeline.

Place, Date

Signature

## List of publications

### First author publications

1. An adaptive teosinte *mexicana* introgression modulates phosphatidylcholine levels and is associated with maize flowering time. Barnes, C.Allison., Rodríguez-Zapata, Fausto., **Juárez Núñez Karla A.** *et al.* PNAS (2022) <https://www.pnas.org/doi/abs/10.1073/pnas.2100036119>
2. High-Throughput Analysis of Protein Turnover with Tandem Fluorescent Protein Timers. Fung J.J., **Blöcher-Juárez K.**, Khmelinskii A. In: Pérez-Torrado R. (eds) The Unfolded Protein Response. Methods in Molecular Biology, vol 2378. Humana, New York, NY. (2022) [https://doi.org/10.1007/978-1-0716-1732-8\\_6](https://doi.org/10.1007/978-1-0716-1732-8_6)

### Contributor in other research publications

1. A B73xPalomero Toluqueño mapping population reveals local adaptation in Mexican highland maize. Pérez-Limón, Sergio. *et al.* G3: Genes, Genomes, Genetics. (2022) <https://doi.org/10.1093/g3journal/jkab447>
2. Demonstration of local adaptation of maize landraces by reciprocal transplantation. Janzen, M. Garret. *et al.* Evolutionary Applications (2022). <https://onlinelibrary.wiley.com/doi/10.1111/eva.13372>

## Summary

In contemporary research, the utilization of diverse tags like fluorescent proteins or small epitopes, for the investigation of proteins of interest has become ubiquitous. Normalizing the extensive usage of tags has made us overlook the critical question of how each tag influences the functionality, subcellular localization and interactions with protein-protein networks of the respective protein of interest.

In this study I sought to shed light on this topic by systematically analyzing the effect of external tags on different proteins in budding yeast, *Saccharomyces cerevisiae*.

To achieve this goal, I chose to assess the impact of tag position, at C-terminus or N-terminus, on a given protein. In order to conduct a genome-wide comparison of the tag position effects, I selected the non-fluorescent protein Halo tag. The Halo tag is a commonly used tag that has a similar size to commonly used fluorescent proteins, rendering it a suitable model for extrapolating its effects to tags of similar sizes. Additionally, for distinguishing between C-terminally and N-terminally tagged strains, I decided to use different cytosolic markers for each strain type.

I conducted a comparative analysis between ORF-3xMyc-Halo and Halo-3Myc-ORF libraries, which I also constructed, to evaluate the effects of tagging proteins at the C-terminus and at the N-terminus. To ensure a fair comparison between the two, I developed a high-throughput yeast competition assay in liquid medium. This study involved competing 5,340 strain pairs over a four-day period, with the identification of C-tagged and N-tagged cells accomplished through cytosolic fluorescence measurements via flow cytometry.

After the genome-wide competition, I noted two significant observations. First, there is a general trend towards enhanced strain fitness when ORFs are tagged at the C-terminus. I identified 292 ORFs that demonstrate improved fitness when tagged at the N-terminus, 4310 ORFs that show no significant difference in fitness based on tag position and 649 ORFs that exhibit enhanced fitness when tagged at the C-terminus. Secondly, essential genes appear to be particularly susceptible to the impact of tagging, in contrast to non-essential genes. This observation aligns with the notion that any disturbance to essential genes has a more detrimental effect.

In this study, I generated a comprehensive genome-wide catalog that identifies the protein terminus at which introducing a tag is less detrimental to the final strain's fitness.

After this initial finding, I proceeded to leverage this information to investigate replicative aging in budding yeast. Given the organisms' asymmetrical replication, I employed a tandem fluorescent protein timer (tFT) to track proteins with age-dependent subcellular distributions during mitosis. After evaluating various tFT options, I opted to create a library in which 5,561 ORFs were tagged at the C-terminus with the selected tFT, mScarlet-I-mNeonGreen.

Following the establishment of a high-throughput imaging protocol, I conducted imaging of the aforementioned library to obtain protein age data of each protein pool in both the mother and the daughter cells during mitosis. After manually annotating proteins that displayed asymmetrical segregation based on their age, I identified 775



proteins of interest. It is important to emphasize that these hits will require further validation through our automated image analysis pipeline. However, this initial exploration suggests a promising approach for identifying candidate proteins whose age may correlate with cellular damage and, ultimately, the aging of the cell receiving the older protein pool. Completing this analysis will also yield a new catalog of proteins in budding yeast that could be potentially acting as aging factors.

## Zusammenfassung

In der heutigen Forschung ist der Einsatz verschiedenster Proteinmarkierungen, mittels beispielsweise fluoreszierenden Proteinen oder kleinen Epitopen, zur Untersuchung von spezifischen Proteinen allgegenwärtig geworden. Der zur Normalität gewordene Einsatz von solch umfangreichen Markierungen hat uns jedoch die kritische Frage übersehen lassen, inwieweit eben jene Markierungen die Funktionalität, die subzelluläre Lokalisation und die Interaktionen in Protein-Protein-Netzwerken eines jeweiligen Proteins beeinflusst.

In dieser Arbeit habe ich versucht dieses Thema näher zu beleuchten, indem ich die Auswirkungen von Markierungen auf verschiedene Proteine in der Bierhefe *Saccharomyces cerevisiae* systematisch analysiert habe.

Um dieses Ziel zu erreichen, entschied ich mich die Auswirkungen der Markierungsposition am C-Terminus oder N-Terminus eines gegebenen Proteins zu bewerten. Um einen genomweiten Vergleich der Auswirkungen der Markierungsposition durchzuführen, wählte ich die nicht-fluoreszierende Markierung namens Halo-Tag aus. Der Halo-Tag ist eine häufig verwendete Markierung, die eine ähnliche Größe wie andere beliebte, fluoreszierende Proteine besitzt. Diese Eigenschaft macht es zu einem geeigneten Modell um seine Auswirkungen auf Markierungen ähnlicher Größe zu extrapolieren. Zusätzlich entschied ich mich für die Verwendung unterschiedlicher zytoplasmatischer Markierungen für C-terminal und N-terminal markierte Stämme, um diese voneinander zu unterscheiden.

Ich führte einen Vergleich zwischen ORF-3xMyc-Halo und Halo-3xMyc-ORF-Bibliotheken durch, die ich ebenfalls erstellte, um die Auswirkungen von C-terminaler und N-terminaler Markierung zu bewerten. Um einen fairen Vergleich zwischen beiden herzustellen, entwickelte ich eine Hochdurchsatz-Konkurrenzuntersuchung in flüssigem Medium. Diese Studie umfasste die Konkurrenz von 5340 Hefestamm-Paaren über einen Zeitraum von vier Tagen, wobei die Identifikation von C-markierten und N-markierten Zellen durch zytoplasmatische Fluoreszenzmessungen mittels Durchflusszytometrie erfolgte.

Nach der genomweiten Konkurrenzuntersuchung stellte ich zwei wichtige Beobachtungen fest. Erstens, es besteht eine allgemeine Neigung zu einer verbesserten Fitness der Hefestämme, wenn ORFs am C-Terminus markiert sind. Ich identifizierte 292 ORFs die bei Markierung des N-Terminus eine verbesserte Fitness zeigten. Bei 4.310 ORFs konnten keine signifikanten Unterschiede in der Fitness in Abhängigkeit der Markierungsposition festgestellt werden, und 649 ORFs zeigten eine erhöhte Fitness bei Markierung des C-Terminus. Zweitens, essentielle Gene scheinen besonders anfällig für die Auswirkungen der Markierung zu sein. Diese Beobachtung steht im Einklang mit der Auffassung, dass Störungen in essentiellen Genen schwerwiegendere Auswirkungen haben.

Mit dieser Studie leiste ich einen Beitrag zur wissenschaftlichen Gemeinschaft, indem ich einen umfassenden, genomweiten Katalog vorstelle, der die bevorzugte Markierungsposition eines Proteins identifiziert und dadurch die Fitness-Nachteile des endgültigen Stammes minimiert.

Auf Grundlage der erhobenen Daten, untersuchte ich zudem das Altern durch Replikation in *Saccharomyces cerevisiae*. Aufgrund der asymmetrischen Replikation der Bierhefe verwendete ich einen Tandem-Fluoreszenzprotein-Timer (tFT), um Proteine mit altersabhängigen, subzellulären Lokalisationen während der Mitose zu verfolgen. Nachdem ich verschiedene tFT-Optionen geprüft hatte, entschied ich mich eine Bibliothek zu erstellen, in der 5.561 ORFs am C-Terminus mit dem ausgewählten tFT namens mScarlet-I-mNeonGreen markiert wurden.

Um Daten bezüglich des Alters für jede Proteingruppe in sowohl Mutter- als auch Tochterzellen während der Mitose zu erhalten, wurde die erstellte Bibliothek unter Verwendung eines Hochdurchsatz-Bildgebungs-Protokolls untersucht. Nach der manuellen Annotation von Proteinen, die eine asymmetrische Segregation basierend auf ihrem Alter zeigten, identifizierte ich 775 Proteine für zukünftige Studien. Es ist wichtig zu betonen, dass eine weitere Validierung durch unseren automatisierten Bildanalysen-Algorithmus erforderlich ist. Dennoch deuten diese vorläufigen Ergebnisse auf einen vielversprechenden Ansatz hin, mit welchem Proteine identifiziert werden können, deren Alter womöglich mit zellulären Schäden und letztendlich mit dem Altern der Zelle korrelieren. Eine abschließende Analyse wird zu einem neuen Proteinkatalog führen, welcher potentielle Alternsfaktoren identifiziert.

## **CHAPTER 1. GENOME-WIDE ASSESSMENT OF TAG POSITION IN YEAST FITNESS**

Nowadays the use of tags to study proteins of interest is so extensive, that we have stopped asking ourselves how each tag affects each protein counterpart's function, localization and protein-protein interaction networks.

### **INTRODUCTION**

While working on creating a library for protein segregation identification, I had to decide at which termini to tag the ORFs. Affinity tags for purification and fluorescent tags for visualization have been used ubiquitously in biology. Nevertheless the introduction of a tag is not trivial; fusion proteins are different from their native form and depending on the nature of each, their functionality, localization, expression levels, stability, etc, might change (Terpe 2003, Yofe et al. 2016, Woestenenk et al. 2004).

Therefore, in parallel to the proteome segregation study in yeast I started to investigate the effect of tagging each ORF at the C-terminus and at the N-terminus.

### **Fusion tags as a tool to study proteins of interest**

A commonly employed read-out for assessing protein function is its abundance and localization (Ho, Baryshnikova, and Brown 2018), therefore, it is imperative to recognize that the introduction of a tagging system into the experimental framework should not be underestimated.

Affinity tags for purification and fluorescent tags for visualization have been used ubiquitously in biology. Over the years, numerous studies have assessed protein abundance in yeast, each with a distinct research objective.

Previously, a pivotal study compiled a genome-wide catalog of yeast proteins localization (Huh et al. 2003). In this research, 6029 strains were constructed, with each protein being tagged endogenously at the carboxy-terminal (C-terminus) with the green fluorescent protein (GFP). The ORF-GFP library was imaged by fluorescence microscopy, finding out that only 4156 strains showed fluorescence signals above background. The fact that only 75 % of the yeast proteome was detectable in microscopy might be due to protein abundance, nevertheless, protein detection could be improved by using a brighter FP. However, it is important to consider that some proteins that were not detected might have been affected by the presence of the GFP tag at the C-terminus.

The genome-wide effect of protein tags have been indirectly studied while assessing the protein expression or localization using different FPs and epitopes, like GFP, sfGFP and TAP-tag (Huh et al. 2003; Yofe et al. 2016; Ghaemmaghami et al. 2003). By comparing endogenous protein localization between studies, we could better speculate that differences in such might be due to the difference in the tag (size, 3D structure, etc). Nonetheless, the comparison is not always possible since the protein tagging has not been consistent. Some studies have used not only a different FP or epitope for their purpose, but also a different protein-termini to introduce such tag. Additionally, different tagging approaches have been widely used, tagging proteins

endogenously or by using different heterologous terminators or different heterologous promoters.

### **Commonly used tags**

A diversity of polypeptides and proteins have been used as tags for studying proteins of interest. Fusion tags exist in a wide-range of sizes, from large fluorescent proteins like sfGFP (26.8 kDa), mNeonGreen (26.6 kDa) and mCherry (26.7 kDa) to smaller peptide tags such as HA (1.1 kDa), 6xHis (0.8 kDa) and myc (1.2 kDa) (Lambert, 2019; Sainz-Baggetto et al. 2017).

Besides the tag itself, another fact to consider when tagging a protein of interest is the location of the tag. Tagging proteins at the C-terminus allows for conservation of endogenous expression levels, although for some proteins carrying a localization signal at the C-terminus would not be an option (Gauss et al, 2005; Costello et al, 2017). On the other hand, N-terminus tagging requires the use of a non-endogenous promoter, which normally promotes stronger expression than the endogenous one (Gauss et al, 2005).

For protein purification one common tag is the His tag. Previously, a study has shown how the protein solubility of recombinant proteins produced in *E. coli*, from ORFs coming from different microorganisms, is equally affected by the 6xHis tag presence at the C-terminus and at the N-terminus (Busso et al. 2004). It was also shown in human cells, how the 6xHis tag, when placed at C-terminus and at the N-terminus, leads to negative effects on the protein solubility, which is also protein-specific (Woestenenk et al. 2004).

The effect of different tags has also been studied in *Saccharomyces cerevisiae*. Proteins like Ddc1, Cln2, Pkc1 and Rad53 have shown an increase in protein instability when tagged at the C-terminus with 3xHis (Sainz-Baggetto et al. 2017). In the same study, those proteins exhibited different expression levels when tagged with the GFP tag instead, some of those changes were explained by a parallel decrease in protein stability. Tagging proteins with small tags can influence tertiary structure, native function and stability depending on tag location and its amino acid composition. In the case of relatively large tags such as GFP, there may be an impact on protein synthesis (Sainz-Baggetto et al. 2017).

A recent study from Weill and collaborators showed that liquid competition of *Saccharomyces cerevisiae* strains where the same protein is tagged at opposite termini provides a good read-out of fitness, thus offering insights into the ideal position for tagging each protein (Weill et al. 2019).

Recognizing that both the nature and size of a tag can impact any protein of interest in *Saccharomyces cerevisiae*, I decided to systematically assess the fitness effect of positioning the non-fluorescent Halo tag at the C-terminus and N-terminus of each yeast ORF.

### **AIM OF THIS STUDY**

The prevalence of protein tags in research has led to overlook the consequences of their presence. To address this gap, I decided to perform a systematic evaluation of the impact of protein tags on fitness.

To systematically compare the effects of external tags at both the N-terminus and C-terminus, the Halo tag was a strong candidate due to its widespread use and similar size to other commonly used fluorescent proteins (33 kDa). To distinguish between N-tagged and C-tagged strains, a distinct cytosolic marker will be utilized.

My proposal involved co-cultivating N-tagged and C-tagged populations for each protein of interest. By monitoring changes in population percentages over time, I used this as a fitness indicator.

This study aims to illuminate the individual effects of tagging each protein in the budding yeast genome at the C-terminus or N-terminus. The insights gained will assist other researchers in making informed decisions about where to tag their protein of interest and the potential effects of tagging at either terminus.

## **RESULTS**

### **Liquid and agar based competition assays using FPs as tags**

In order to better understand the effect of tag position on yeast fitness, I decided to reproduce a previously published yeast competition assay (Weill et al. 2019). The selected competition study used 57 ORFs that were tagged at the N-terminus and at the C-terminus tagged with GFP. Strains where the proteins were tagged at the C-terminus were distinguished from N-terminus tagged strains by the incorporation of a mCherry as cytosolic marker. Each pair of strains was co-cultivated for four days in liquid and the population percentage of each strain was quantified everyday by measuring each population's percentage with flow cytometry.

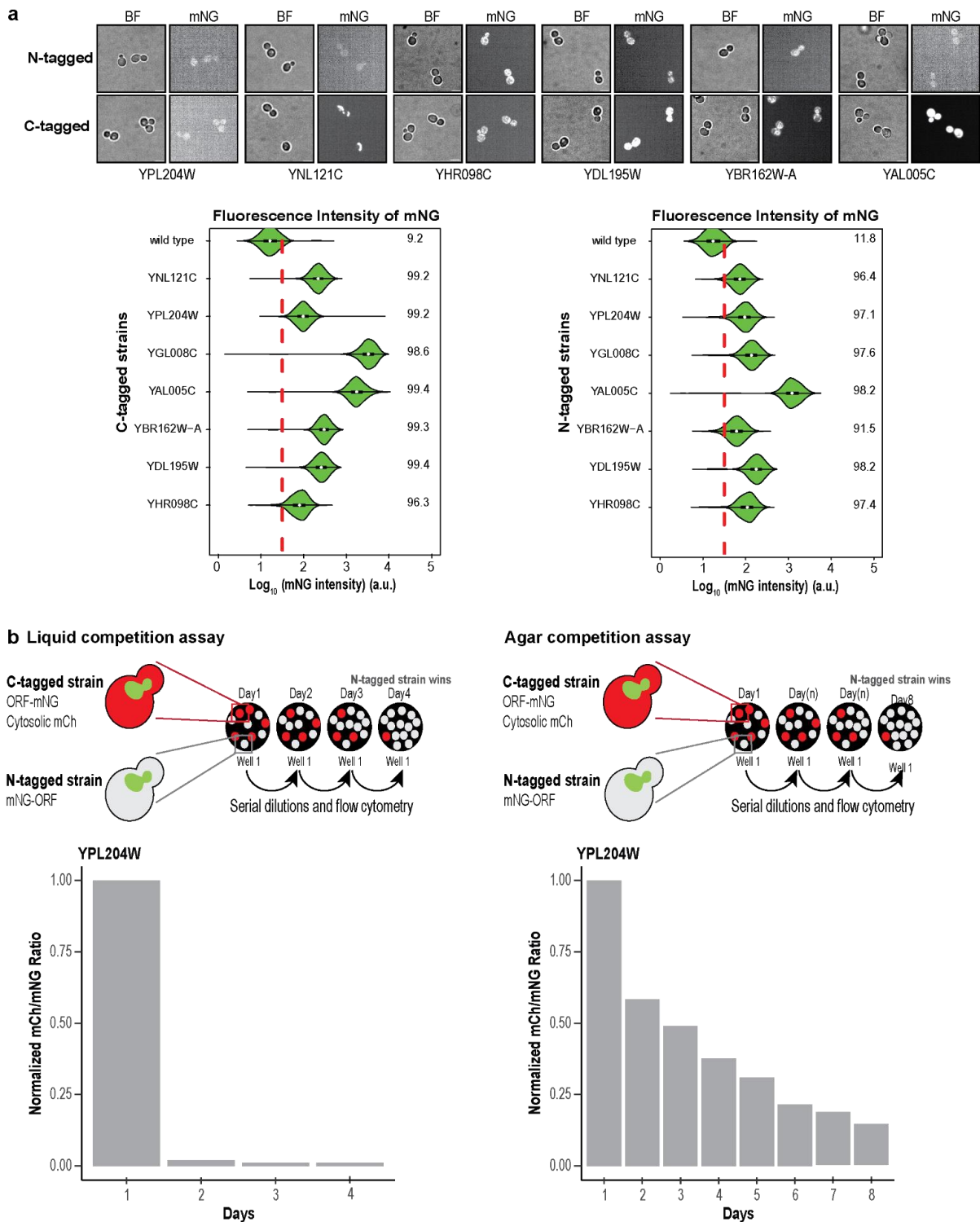
Since mNeonGreen is brighter in yeast than GFP (Meurer et al. 2018), I started by creating two donor strains that could be distinguish by their fluorescence, one carrying mCherry as cytosolic marker and one with no cytosolic marker. The first donor strain was used for tagging ORFs at the C-terminus with mNeonGreen, while the non-fluorescent strain was used to tag ORFs at the N-terminus with mNeonGreen. Using the SGA method (Baryshnikova et al. 2010), I tagged 7 ORFs that were also used in the study of Weill et al. 2017.

Tagging efficiency of all strains was quantified by imaging the different strains by confocal fluorescence microscopy and manually counting the cells with mNG signal (data not shown). As well, I performed flow cytometry of all 14 strains to have a confirmation of the tagging efficiency (Figure 1a), previously scored by microscopy. In general, the majority of the strains exhibited a tagging efficiency above 90 %, for both N-terminus and C-terminus tagging.

In the fluorescence microscopy of the strain-pairs, I observed that for some proteins tagging at the N-terminus or at the C-terminus resulted in a different localization (Figure 1a), or in some cases the localization seemed to be the same with an apparent difference in the expression level.

After confirming that per ORF, both N-terminus and C-terminus tagged strains exhibited almost a 100 % tagging efficiency (Figure 1a), I proceeded to reproduce the previously established yeast liquid competition assay (Weill et al. 2019). Since the technical assesment of a genome-wide competition assay would be easier on agar, I decided to adapt the liquid competition assay protocol to also perform the assay on agar. As yeast cultures grow slower on agar, I decided to compete the strains on agar for eight days instead of four (Figure 1b).

Here, I only show the results for one of the tested ORFs, YPL204W (Figure 1b), where both the N-terminus and the C-terminus tagged strains were competed in liquid form and on agar. By the end of both competition assays, the normalized mCh/ mNG ratio decreased. This reflecting on the negative fitness effect of the C-tagged population and the N-tagged population improved fitness through time. Hereafter, I will refer to such ORFs as N-winners, where the N-tagged version of the ORF seems to lead to a fitness benefit of the strain. For this ORF, YPL204W, I obtained the same fitness read-out as the one previously reported in literature (Weill et al. 2019).



**Figure 1.** Competition assay of strains carrying N-terminus and C-terminus tags. **a.** (up) Localization of several proteins according to the position of the tag. Brightfield (BF) and mNeonGreen (mNG) channels were acquired with 500ms of exposure and 70% laser power. Images were auto-contrasted. Scale bar, 5  $\mu$ m. (down) ORFs tagged at the C-terminus and at the N-terminus. Tagging efficiencies were determined by measuring 10000 cells per strain in flow cytometry. **b.** (up) Scheme of the liquid competition assay proposed by Weill et al. 2019 and my adapted protocol for agar competition. A protein that localizes to the nucleus is tagged with mNG at N-terminus



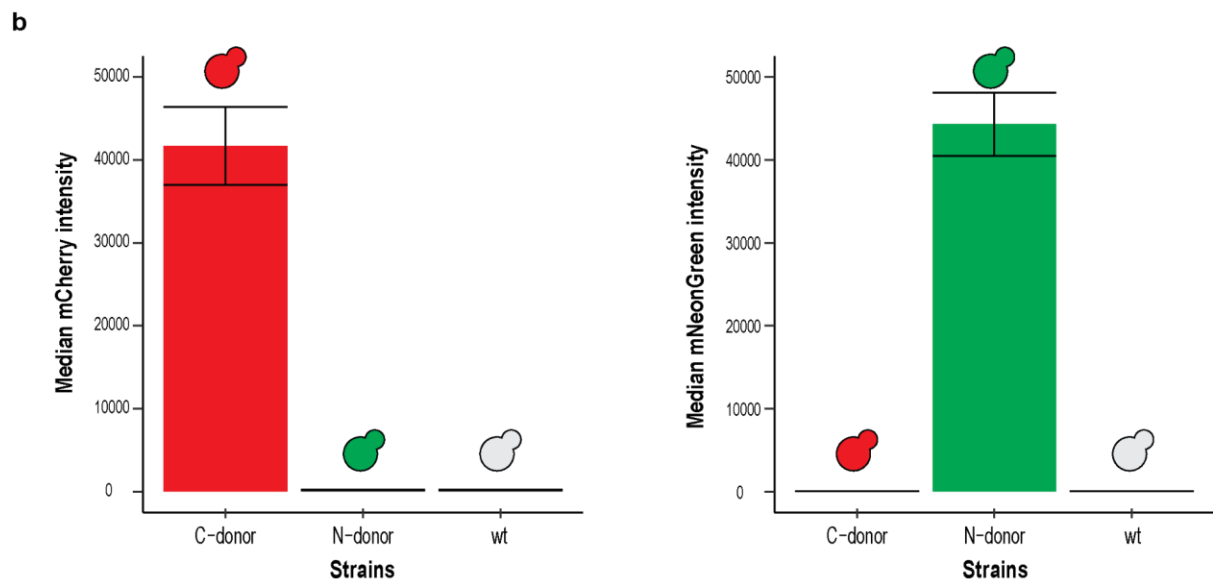
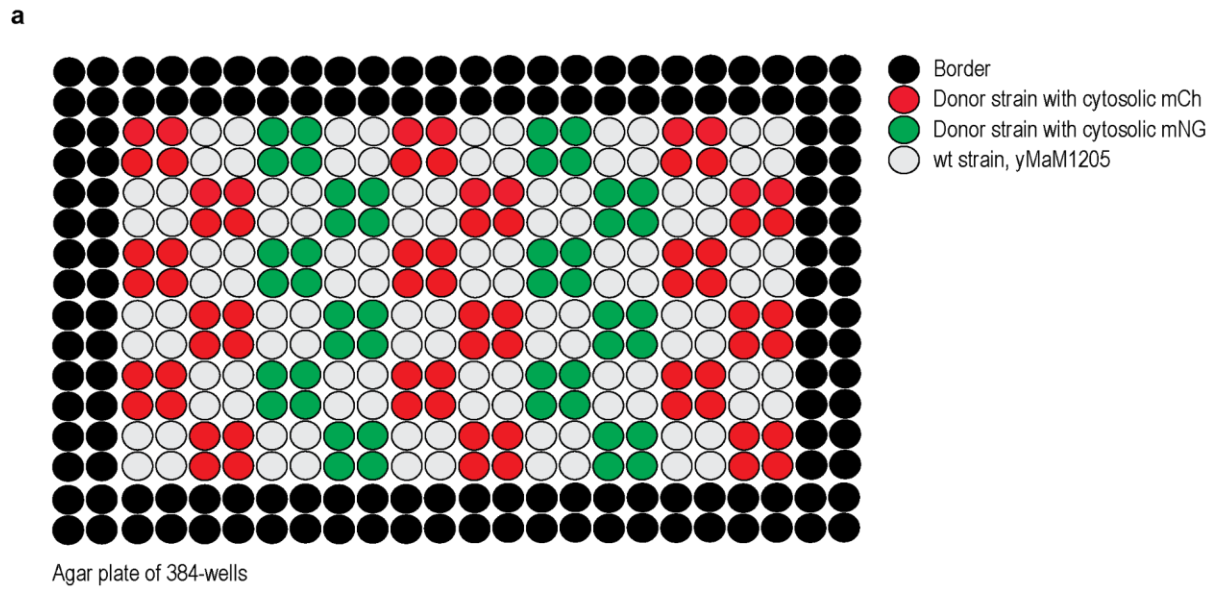
and C-terminus. (down) The ORF YPL204W was tagged at the N-terminus and at the C-terminus with mNG, C-tagged strain expressed the mCh as cytosolic marker while the N-tagged strain was non-fluorescent in the cytosol. Both strains were co-cultivated and grown for 4 days (liquid competition assay) or 8 days (agar competition assay). The mCh/mNG ratios were normalized to the day 1 ratio, in each competition assay.

### **Using a non-fluorescent tag for competition assays**

In order to systematically compare the effect of the tag's position on yeast fitness, I decided to use the same tag for N-terminus and C-terminus. However, to distinguish between the N-tagged and C-tagged population of cells, I required an additional marker. For that reason I selected the non-fluorescent Halo tag for N-terminus and C-terminus tagging, and two different FP as cytosolic markers to distinguish between the two competing populations, mCherry as cytosolic marker for C-tagged strains and mNeonGreen as cytosolic marker for N-tagged strains.

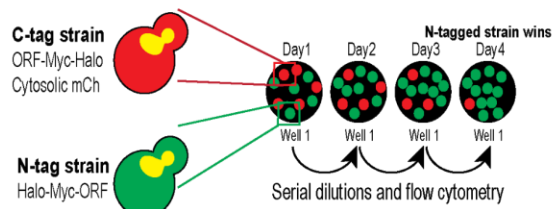
To make sure we could distinguish both cytosolic markers intensities from background auto-fluorescence, I measured the fluorescence of both donor strains together with the non-fluorescent wt strain, arrayed in the same agar plate (Figure 2). Once I confirmed that mCherry and mNeonGreen brightness as cytosolic markers was above background, I transformed existing donor Halo tag plasmids into their respective yeast donor strain. Afterwards, I used such plasmids to endogenously tag the above mentioned ORFs at the N-terminus (Halo-myc-ORF) and C-terminus (ORF-myc-Halo).

I then repeated the liquid and agar competition assays to assess the fitness effect of tag position, when using Halo as a tag (Figure 3). For the liquid competition assay I determined population percentages as a fitness read-out, using flow cytometry. For determining the fitness effect of the agar-based competition assay, I measured fluorescence intensity with a plate reader.

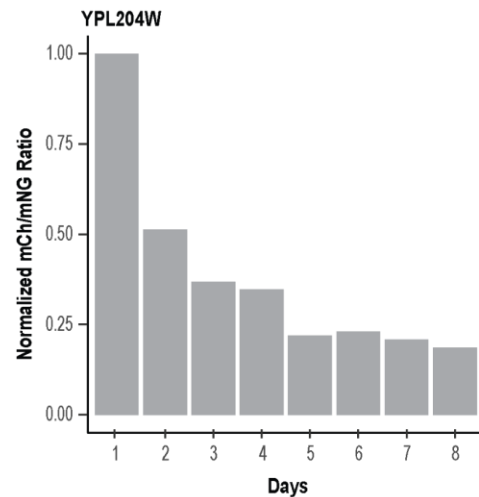
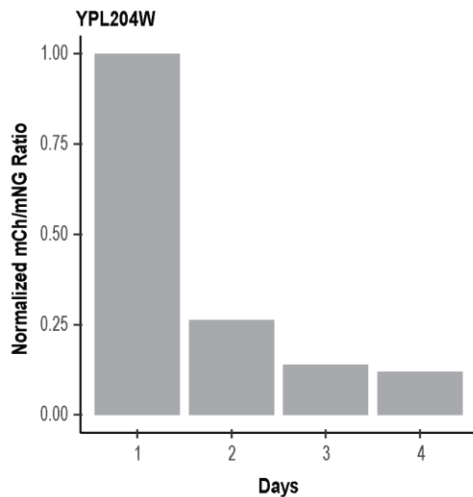
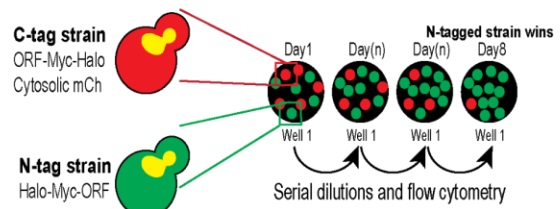


**Figure 2.** C-donor and N-donor strains carrying different fluorescent cytosolic markers. **a.** Agar plate with arrayed donor strains and non-fluorescent wt strain grown in a 384-format. **b.** The fluorescence intensities were measured using a plate reader (SparkControl 20M plate reader with monochromator, TECAN) and the settings indicated in Table 1. Error shows standard deviation calculated with 72 technical replicates of the C-donor strains and with 48 technical replicates of the N-donor strain.

### a Liquid competition assay



### Agar competition assay



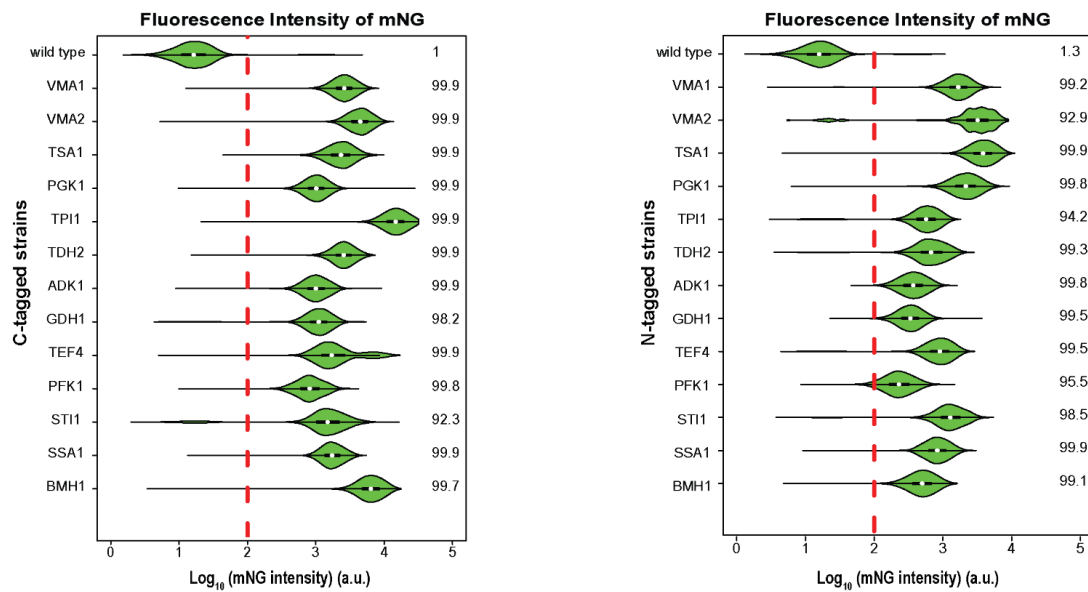
**Figure 3.** Proof-of-principle for competition assays using the non-fluorescent Halo tag. **a.** Experimental design for liquid and agar-based competition assays. N-tagged strains expressed the mNG cytosolic marker while C-tagged strains expressed the mCh cytosolic marker. A protein that localizes to the nucleus is tagged with non-fluorescent Halo tag at N-terminus and C-terminus. **b.** Liquid and agar competition assay of the ORF YPL204W tagged at the N-terminus and at the C-terminus with Halo-myc and myc-Halo, respectively. The mCh/mNG ratio was normalized to the ratio at day 1, in each competition assay.

Again, I observed the same fitness trends in both liquid and agar-based competition assays for the given ORFs. Therefore, I decided to construct two genome-wide libraries using Halo as tag. Following the construction of such libraries, I established a high-throughput genome-wide competition protocol and assessed the genome-wide effect of Halo tag position on fitness.

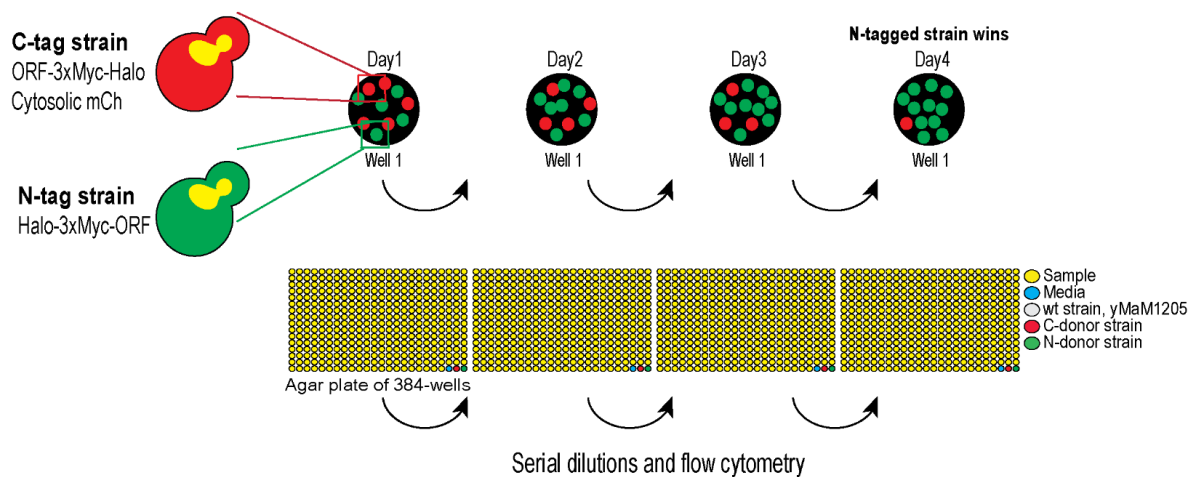
### Creation of genome-wide Halo tagged libraries and their competition

In order to determine genome-wide the effect of the tag position on fitness, I used the SWAp-Tag approach (Yofe et al. 2016) and constructed two libraries in which all ORFs in the C-SWAT (5760 ORFs) and in the N-SWAT (6144 ORFs) libraries were tagged endogenously with the 3xMyc-Halo and Halo-3xMyc tags respectively. As previously established, the donor strain for the C-tagging was carrying the mCherry as cytosolic marker while the N-donor strain was carrying the mNeonGreen as cytosolic marker (Figure 4b).

**a Reference ORFs tagged with mNG at the N or at the C-terminal**



**b Liquid competition assay**



**Figure 4.** Tagging efficiency of libraries and genome-wide liquid competition assay. **a.** Comparison of the tagging efficiency obtained for control ORFs when tagged at the C-terminus and N-terminus with Halo-3xMyc and 3xMyc-Halo. The percentage of cells with green fluorescence above background per protein is shown; around 10000 cells were measured per strain using flow cytometry. **b.** Experimental design for the genome-wide competition assay in liquid where N-tagged strains expressed the mNG cytosolic marker while C-tagged strains expressed the mCh cytosolic marker. Genome-wide Halo was used as tag; generated strain pairs were co-cultivated in a well (indicated as Sample well) for competition during 4 days.

During the construction of the libraries, I also included a reference plate containing 13 strains, present in both C-SWAT and N-SWAT libraries, expressing high abundance proteins and which had been used as a reference of the tagging efficiency when constructing new libraries using the SGA method (Meurer et al. 2018). These strains were crossed against donor strains carrying mNeonGreen as a tag. After the construction of the libraries, the tagging efficiency of the reference plate was assessed

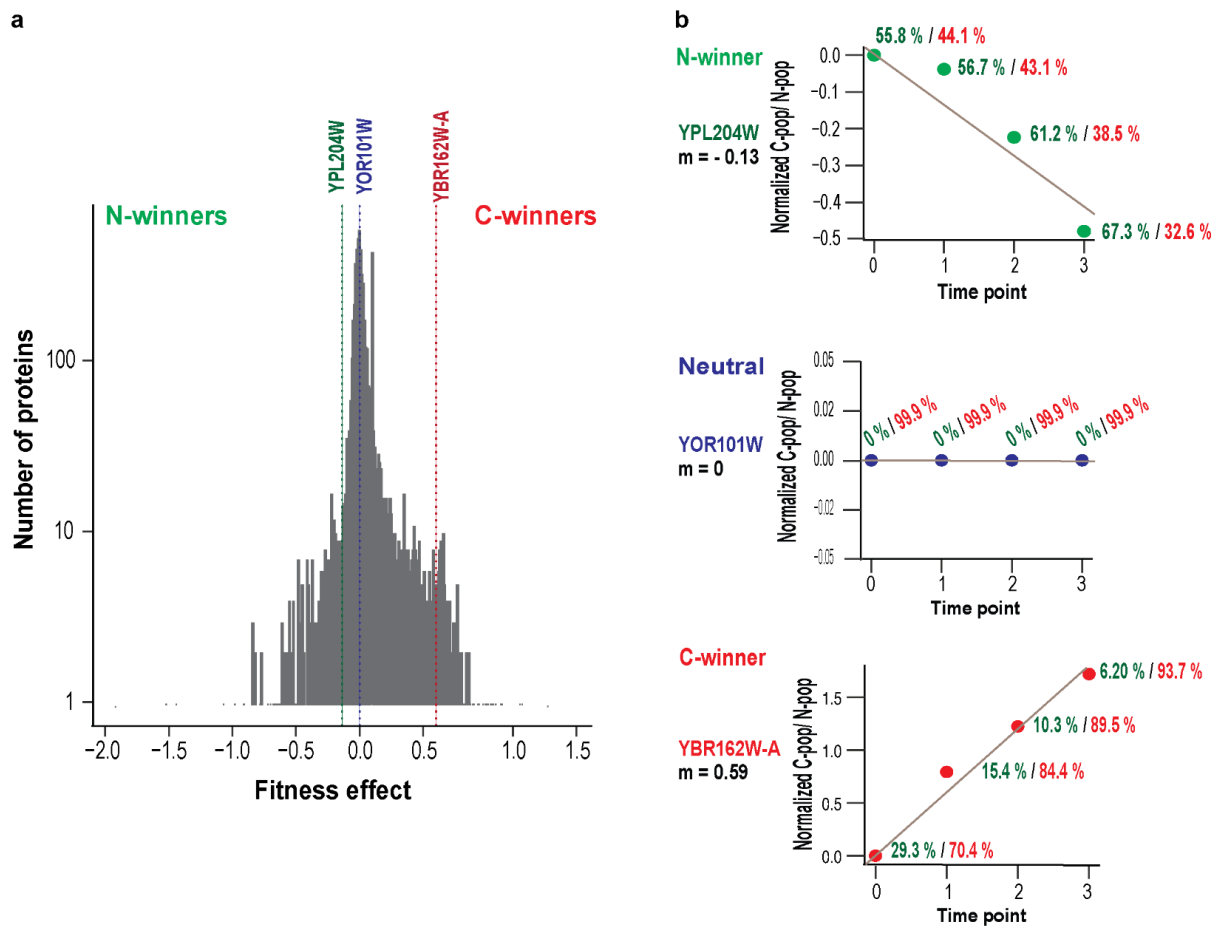
by measuring the cell percentage carrying the mNeonGreen tag in flow cytometry. The tagging efficiency for the C-tagged strains varied between 92.3% and 99.9% and for their N-tagged counterparts the efficiency ranged from 93.9% and 99.9% (Figure 4a). Based on this, I assumed the high-throughput (HTS) tagging procedure with Halo was also successful for the majority of the strains of the libraries.

After the library construction, I re-arrayed 5340 ORFs which were shared between the C-3xmyc-Halo and N-Halo-3xMyc libraries. After the re-array of both libraries on agar plates, I inoculated pairs of C-tagged and the corresponding N-tagged counterpart strains in a 96-well plate with liquid media. These cultures were grown at normal conditions for 24 h, diluted after saturation and grown again until an optimal OD for flow cytometry. The competition assay lasted four days as established previously (Figure 4b). The measuring time per 96-well plate was optimized to 32 min.

In total, I competed and measured 58 different 96-well plates for 4 days each. Every plate had 92 sample wells and 4 control wells. The control wells corresponded to: one media well, one well inoculated with a non-fluorescent strain corresponding to the background of the donor strains (yMaM1205) and one well per donor strain (Figure 4b). Having these controls along the competition assay allowed me to calculate the cross-contamination per plate by checking the population percentage from the C-donor well into the N-donor well. The cross-contamination in the screen was on average 0.11% across the 58 plates, along the 4 days.

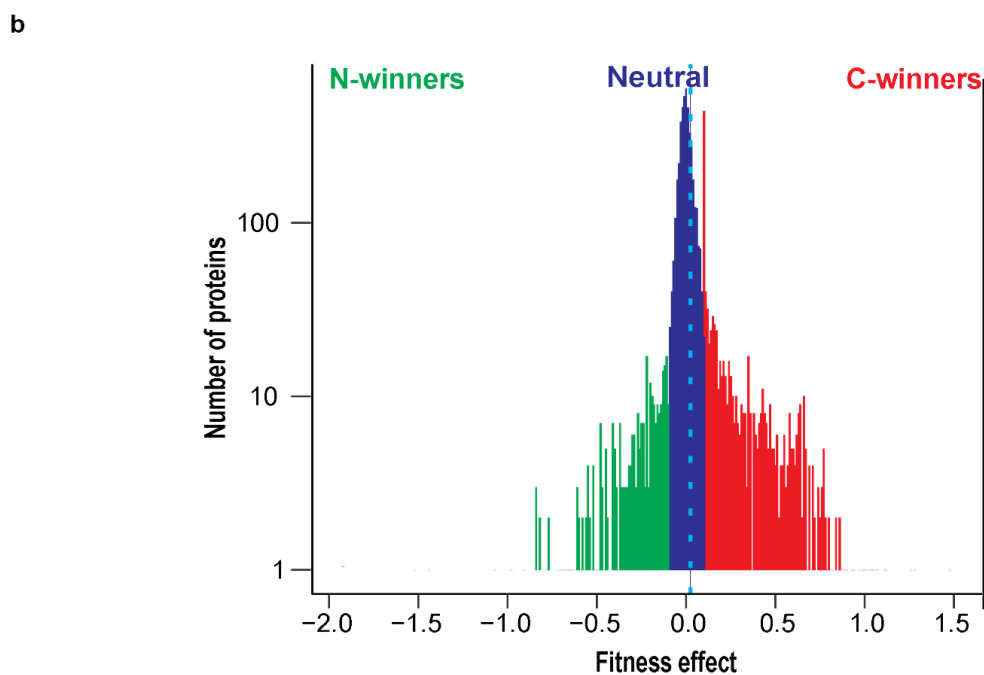
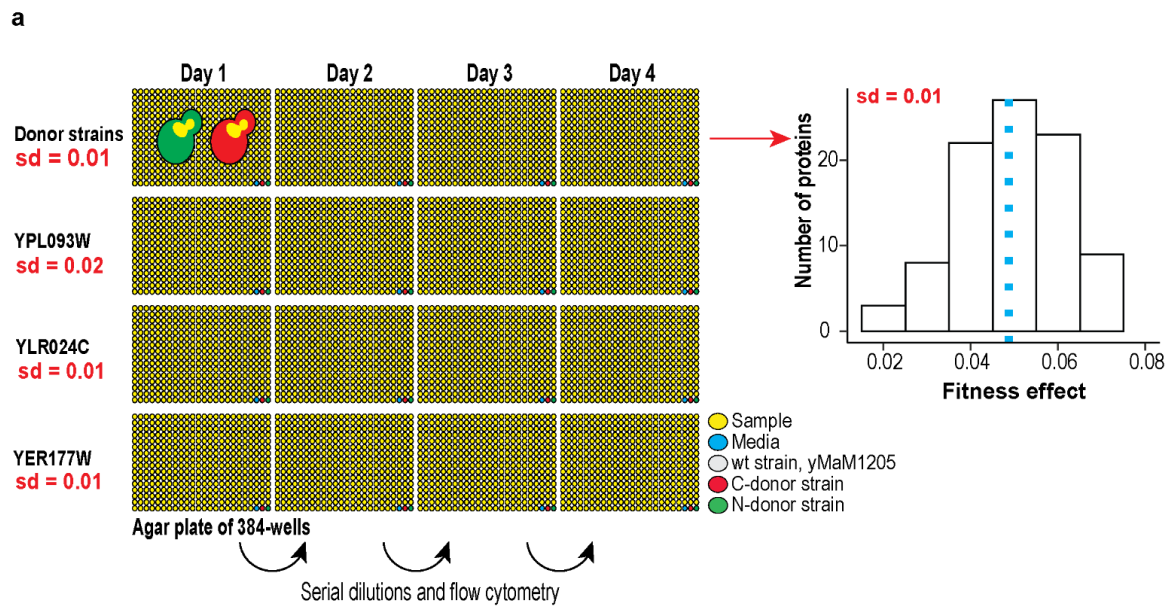
In order to estimate the relative fitness difference between strain pairs, I calculated the ratio of percentage of C-tagged cells/ N-tagged cells per well, per day. I then normalized the ratio per ORF, to the ratio of day 1 of the competition assay.

Then I fitted a linear regression model to the time points (x-axis) against the log of the normalized C-population/ N-population ratio (y-axis) (Figure 5b). The slope of the fitted model was positive if the C-tagged strain exhibited a better fitness than the N-tagged strain; and the slope had a negative value if the N-tagged strain showed a better fitness instead (Figure 5b).



**Figure 5.** Genome-wide distribution of the effect of the tag position on fitness. **a.** Genome-wide distribution of fitness effect, indicated with dashed lines a potential N-winner ORF (dark green), a potential neutral ORF (dark blue) and a potential C-winner (dark red). **b.** The C-tagged/ N-tagged population percentages, shown next to each data point, per day were normalized to the value at day 1 (time point 0). A linear regression model was fitted to the log<sub>10</sub> of the normalized C-population/ N-population ratio (y-axis) against the time points (x-axis). The slope ( $m$ ) of the fitted is shown on the left side of each competition plot.

YOR101W is an example of how an ORF whose slope is neutral, might look across four days of competition. For this ORF, there was no colony for the initial N-SWAT library and therefore, no final N-tagged Halo-3xMyc strain.

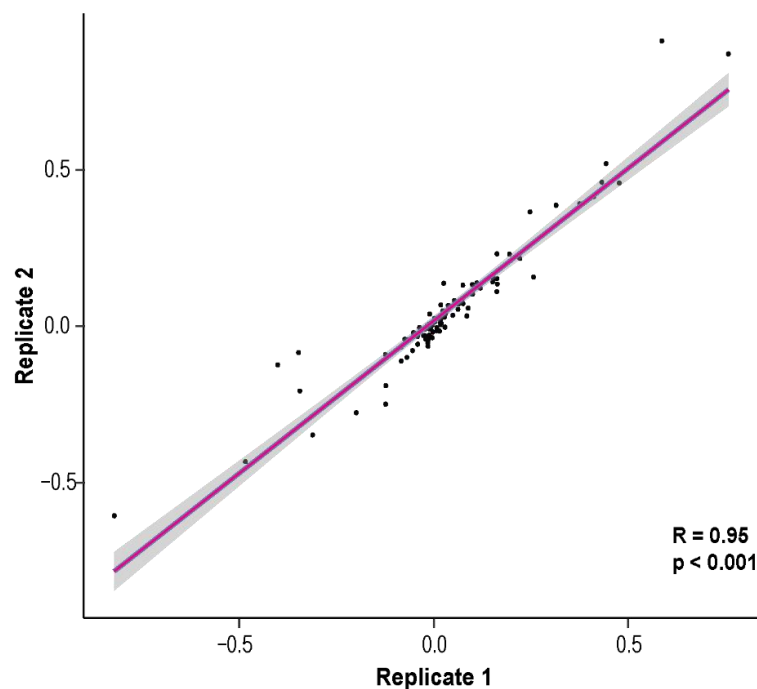


**Figure 6.** Genome-wide threshold determination for fitness effect for 5340 ORFs. **a.** Measurement of the technical noise (sd) in the flow cytometer while competing pairs of strains. The distribution of the fitness effect detected in 92-wells of N-donor vs C-donor competition. Y-axis shows the protein count, the light blue dashed line indicates the mean of the distribution. **b.** Genome-wide distribution of effect on fitness of the tag position. Y-axis indicates the log<sub>10</sub> of the protein count and x-axis the slope of the fitted model that is considered as the read-out for the fitness effect of the tag position. Mean slope (0.02) of the genome-wide competition assay is indicated as a light blue dashed line.

In order to determine the technical noise during the competition assay, I competed 4 different pairs of N-tagged and C-tagged strains for the ORFs YPL093W, YLR024C, YER177W and also the initial donor strains carrying only a cytosolic marker. I competed the strain pairs along 92-wells, for 4 days, and calculated the standard

deviation (sd) of the fitness effect along the plate per strain (Table S1). I determined that the sd per strain pair, where the maximum sd was 0.02 and the minimum was 0.01, among those strains. Therefore, I decided to use the maximum sd value for setting the threshold for significant fitness effect of the tag position; I set the threshold using a five sigma statistical significance value ( $5 \times 0.02 = 0.1$ ). Using  $\pm 0.1$  fitness effect value as threshold, I found 292 ORFs in the screen are N-winners (there is a fitness benefit when tagged at the N-terminus over the C-terminus), 4310 ORFs are neutral towards the position of the tag and 649 ORFs are C-winners (Figure 6b).

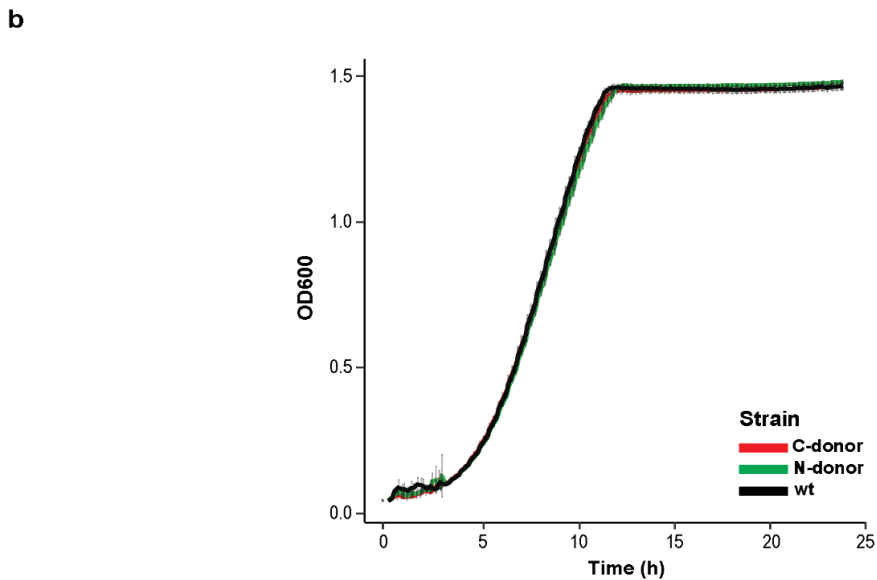
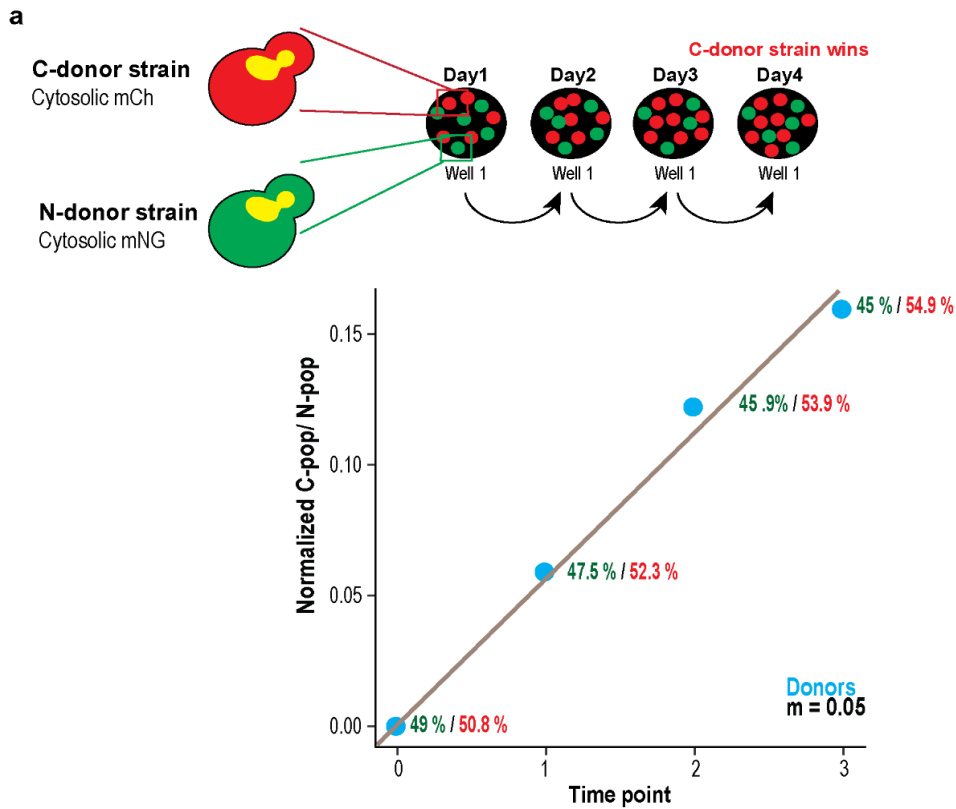
Among the 58 different 96-plates competed and measured in flow cytometry, I included a biological replicate for plate 1, which was measured at the end of the competition assay (plate 59). After the screen, I check how reproducible were the results for fitness effect across the 92-wells in those two replicates. As seen in Figure 7, there was a very good correlation between the fitness effect found for the 92 strains in those two replicates, having a statistical significance indicated with p-value  $< 0.001$  and a linear correlation coefficient of  $R = 0.95$ .



**Figure 7.** Reproducibility of high-throughput genome-wide liquid competition assay. Correlation between the fitness effect of two biological replicates for 92 strain pairs. Replicate1 (plate 1) on the x-axis and Replicate2 (plate 59) on the y-axis. Linear correlation fit 'y ~ x', with a correlation accuracy  $p < 0.001$  and correlation coefficient  $R = 0.95$ .

When I competed the C-donor and N-donor strains for determining the noise of the measurement in the flow cytometry (Figure 6a) I realized the average slope (fitness effect) across 92-competition wells was 0.048 (indicated with a light blue dashed line), which meant the C-donor strains was showing a fitness benefit compared to the N-donor strain (Figure 6a).



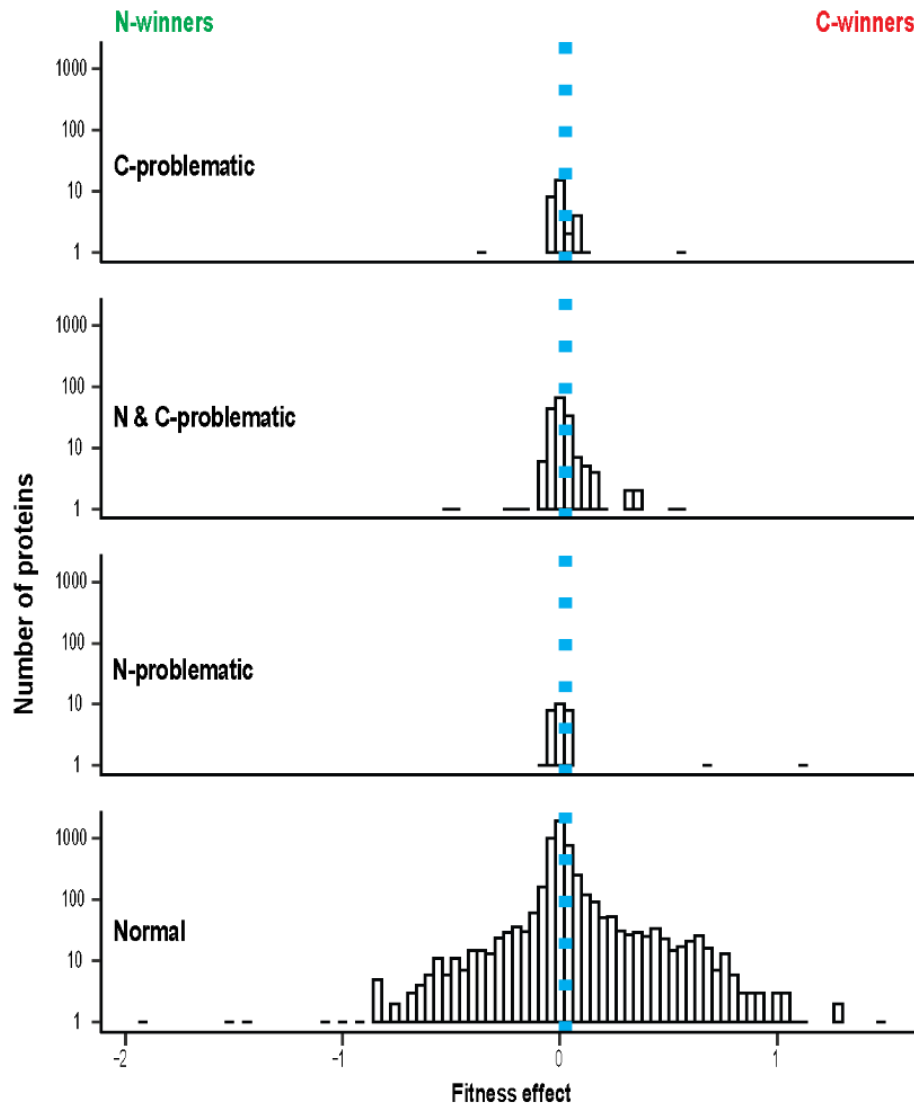


**Figure 8.** Characterization of the donor strains used for the construction of libraries. **a.** Population percentages in well E06 during the competition assay performed for the previous determination of noise during the measurement in Figure 6a. The C-tagged/ N-tagged population percentages, shown next to each data point, per day were normalized to the value at day 1 (time point 0). A linear regression model was fitted to the log10 of the normalized C-population/ N-population ratio (y-axis) against the time points (x-axis). The slope ( $m$ ) of the fitted is shown on the bottom right. **b.** 24 h growth curve of N-donor, C-donor and wt strain (yMaM1205) grown in SC-Leu+Ade media.

The C-donor strain carrying mCherry as cytosolic markers seems to have a fitness advantage over the N-donor strains carrying the mNeonGreen as cytosolic marker

when competed on liquid for four days (Figure 8a). This advantage of the C-donor over the N-donor was not noticeable in a growth curve (Figure 8b).

### Characteristics of N-winners and C-winners



**Figure 9.** Relative fitness effect of the tag position of ORFs difficult to tag. Y-axis indicates the log<sub>10</sub> of the protein count and x-axis the slope of the fitted model. Mean slope is indicated as a light blue dashed line.

Previously, it was shown that seamless tagging achieved by DNA double-strand break repair by homologous recombination is inefficient for a series of ORFs. For ORFs where the start codon flanked by similar stretches longer than 28 bp and for ORFs with the stop codon flanked by similar stretches longer than 28 bp, there is a potential for ectopic module excision during the seamless N-terminus and C-terminus tagging, respectively (Khmelinskii et al. 2011).

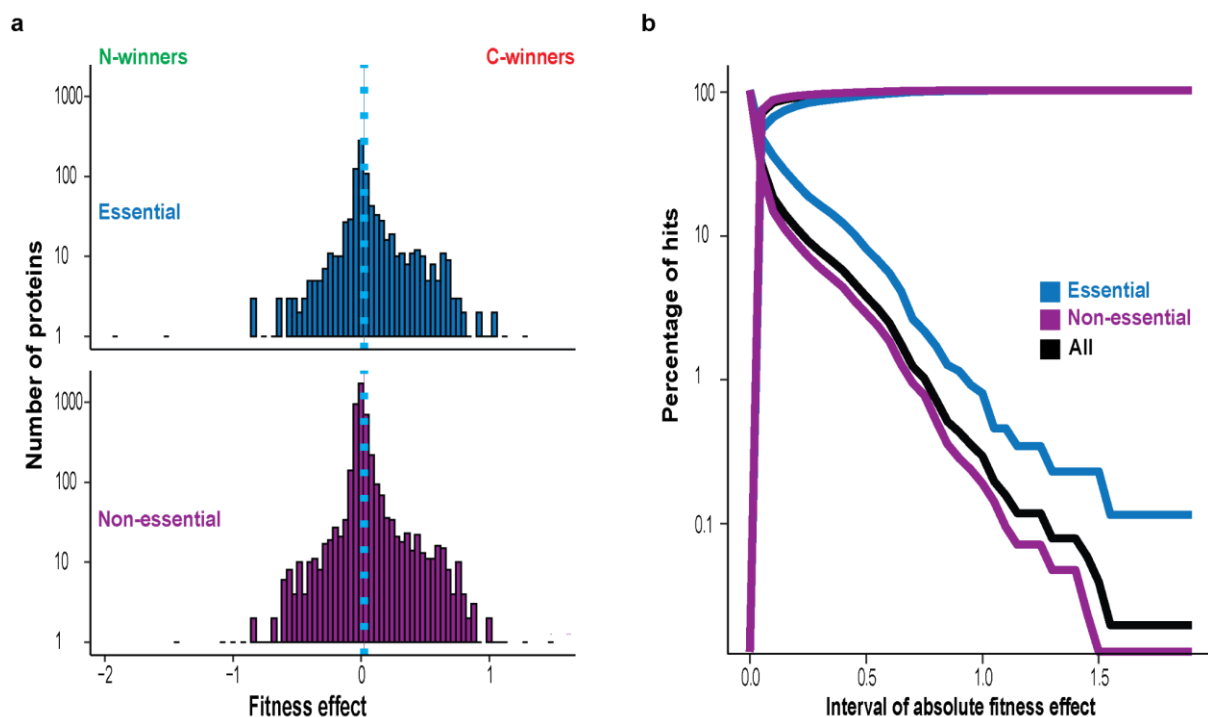
Knowing this, I checked for those ORFs in the competition assay (Figure 9) and classified them in three categories: ORFs that could not be to tag at the C-terminus (C-problematic), ORFs that could have problems when tagged at the N-terminus (N-

problematic) and ORFs that would have problems when tagged at any termini (N-C-problematic).

Looking at where within the distribution (Figure 9) were those ORFs falling, I observed that out of 32 C-problematic ORFs, 29 were within the Neutral range, 2 were C-winners and 1 was N-winner. From the 29 N-problematic ORFs, 27 were Neutral and 2 were C-winners. The majority of the C-problematic and N-problematic ORFs were identified as Neutral ones within my competition assay.

From the 176 ORFs reported to be both N-problematic and C-problematic, 155 were among the Neutral range in my screen, 5 were N-winners and 16 C-winners.

I then checked for where essential and non-essential ORFs were located within the genome-wide fitness effect distribution. For this, I used a previously generated list of 898 genes reported as essential (Després et al. 2020). I annotated as non-essential 4354 ORFs which I tested but were not present in the previous published study.



**Figure 10.** Essential and non-essential ORFs in the genome-wide fitness effect distribution.

**a.** Fitness effect of essential and non-essential ORFs. Y-axis indicates the log<sub>10</sub> of the protein count and x-axis the slope of the fitted model. Mean slope is indicated as a light blue dashed line. **b.** Count of hits (N-winners and C-winners) give different thresholds for the absolute fitness effect in the genome-wide competition screen. Y-axis indicated the log<sub>10</sub> of the number of hits and the x-axis the absolute value of the fitness effect.

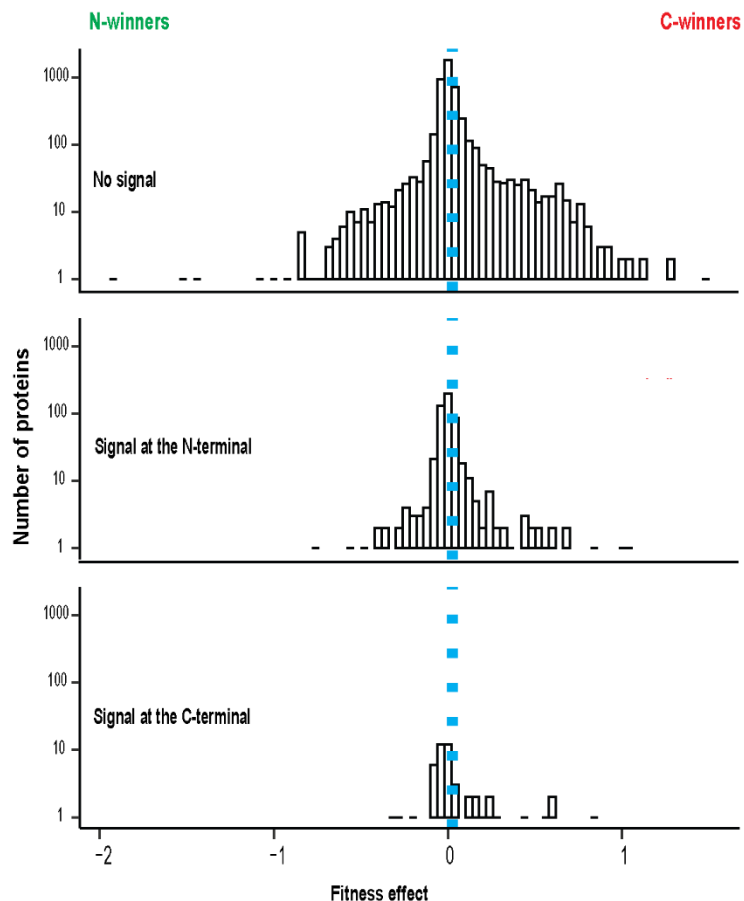
Among the essential ORFs, 11% were identified as N-winners, 65% as Neutral and 23% as C-winners. In contrast, among the 4354 non-essential ORFs, 4.3% were categorized as N-winners, 10% as C-winners, with the remainder falling into the Neutral category (Figure 10a).

I also tested how many hits we would get for different given fitness effect thresholds. For this purpose, I defined a hit as an ORF located either at the left or at the right side of a given threshold. In Figure 10 panel b, I generated a vector of different thresholds for the absolute fitness effect of all ORFs of the genome-wide competition assay. This, to have more information on the existence of a trend to have more hits (C-winners and N-winners) for a specific ORF-type (essential or non-essential) independently of the threshold selected for the slope.

Based on each threshold I calculated how many ORFs were located to the left side of the given threshold and how many were on the other side, and we converted this number to percentage. I observed that there is a difference between essential and non-essential ORFs, having more hits (for both N-winners and C-winners) for essential ORFs (Figure 10a, blue line). Together with John Fing Jia, we also performed a t-test for this dataset (data not shown), and we concluded that there is a significant enrichment (p-value =  $6.194e^{-5}$ ) of essential genes within the N-winners and C-winners.

I then looked at the behavior in the competition screen of 561 ORFs reported/predicted to have a signal peptide (SP) at the N-terminus (Yofe et al. 2016). As expected, we see that ORFs with a SP at the N-terminus tend to be C-winners.

From different literature, I also created a list of 51 ORFs reported to be TA-proteins, containing a TMD at the C-terminus. I observed that ORFs carrying a SP at the N-terminus are more likely to be C-winners (Figure 11). On the other hand, ORFs carrying a TMD at the C-terminus tend to be N-winners (Figure 11).

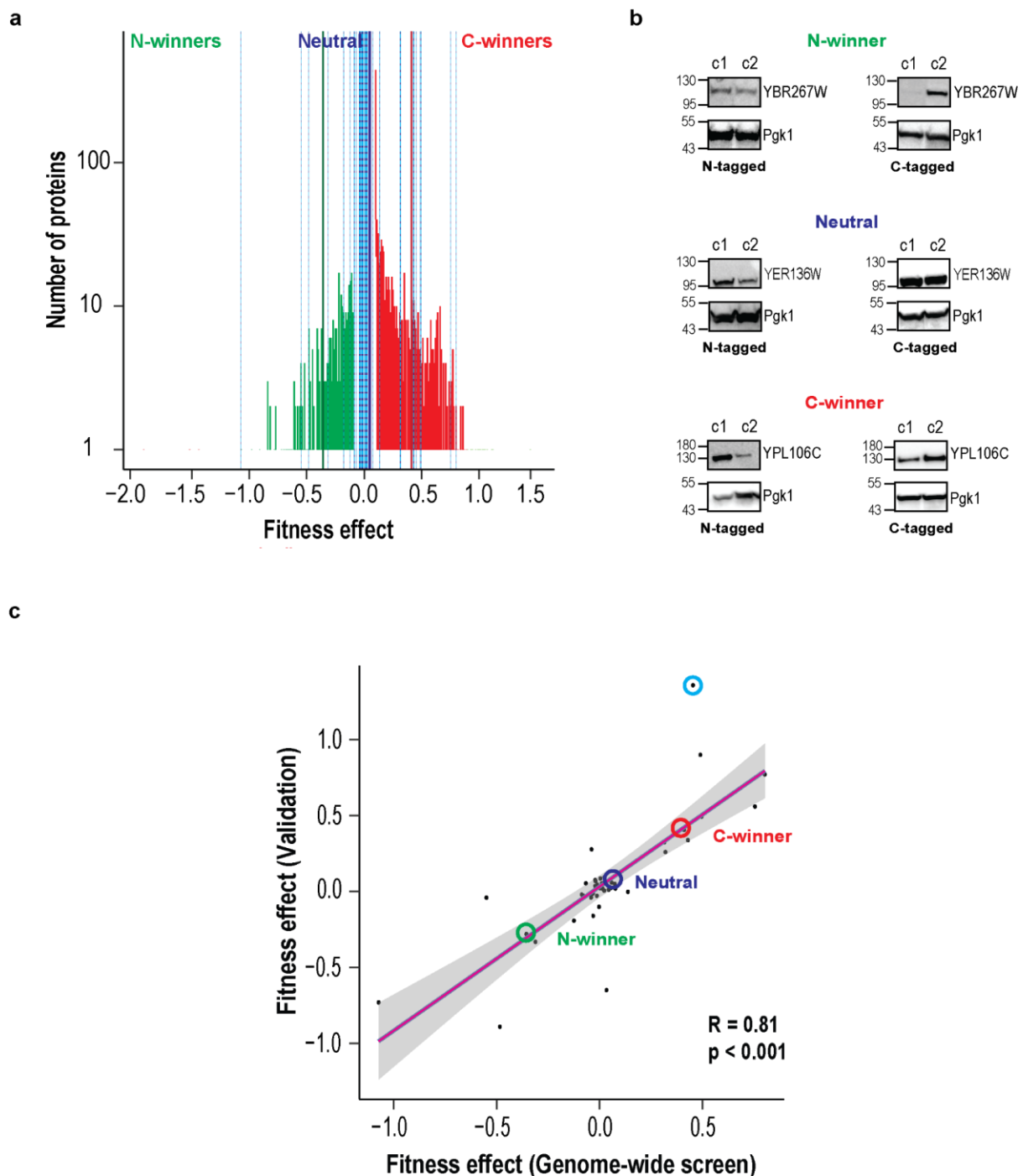


**Figure 11.** Genome-wide presence of signals at the C-terminus and N-terminus. Y-axis indicates the log<sub>10</sub> of the protein count and x-axis the slope of the fitted model. Mean slope is indicated as a light blue dashed line.

### Validation of tag presence and fitness effect

During the creation of both genome-wide Halo libraries I did not have a direct approach to measure their tagging efficiency. Since Halo is a non-fluorescent tag, I decided to perform an additional experiment to validate the presence of such tag. For this, I worked together with two masters students.

We randomly selected 45 strains of each final C-tagged (ORF-3xMyc-Halo) and N-tagged (Halo-3Myc-ORF) libraries, that were covering all the fitness effect range of the genome-wide competition assay (Figure 12a). These 45 strain-pairs were then streaked for single colonies, from which three clones were patched per strain. We extracted proteins from two clones per strain, and performed western-blot (wb) using anti-Myc as primary antibody to validate the presence of the 3xMyc tag (Figure 12b). The strains for which their western-blot are shown, are highlighted with a solid line in the genome-wide distribution of fitness effect (Figure 12a).



**Figure 12.** Validation of the presence of the tag in 45 N-tagged and C-tagged strains. **a.** Indicated by light blue dashed lines is the position of the 45 selected strains for validation. Strains randomly picked for validation of the presence of the tag. **b.** Protein extracts were transferred to 10 % acrylamide membranes, which were incubated with anti-Myc, mouse monoclonal antibody (IMB protein core facility). **c.** Correlation between the fitness effect value obtained during the genome-wide competition screen (x-axis) and the value obtained from the Validation subset (y-axis). Linear correlation fit 'y ~ x', with a correlation accuracy  $p < 0.001$  and correlation coefficient  $R = 0.81$ .

Through western-blotting, we were able to identify positive clones carrying the 3xMyc tag. Nevertheless, not both clones tested were consistently positive in all strains, in

some cases like clone 1 for N-tagged YBR267W strain has a very dim signal intensity compared to clone 2 (Figure 12b). In other cases (data not shown), both clones for both strains had the same signal intensity, or only one clone was positive, or both clones were negative. Therefore, we selected only one positive clone per strain pair, and we competed the strains in liquid during 4 days and measured the population percentages every 24 h, as previously described.

After this competition assay, we again fitted a linear model to the time points against the log<sub>10</sub> of the normalized C-population/ N-population ratio. We then compared the fitness effect, slope of the fitted model, in each competition pair of the validation experiment against the fitness effect value obtained in the genome-wide screen for the same strain pairs (Figure 12c). We calculated the correlation coefficient and its accuracy, which were  $R = 0.81$  and  $p < 0.001$ , for the overall 45 strain pairs used for validation. The same strains for which the  $w_s$  are shown, are also indicated with a circle in the fitness effect correlation (Figure 12c).

In general, the presence of the tag and the strong correlation between the fitness effects of the subset strains used for validation and their corresponding effects in the genome-wide screen confirmed that the observed fitness effects were indeed attributable to the tag's presence at a specific protein terminus.

## **DISCUSSION**

Since fusion proteins are different from their native form, some tags at specific protein termini can affect their functionality, localization, expression levels, stability (Terpe 2003, Yofe et al. 2016, Woestenenk et al. 2004) and probably also the downstream biological processes in which they are involved.

### **Liquid and agar based competition assays using FPs as tags**

In order to systematically compare the effect of the tag position on the fitness of the resulting strain, I decided to compete pairs of strains where the same protein of interest is tagged either at the N-terminus or at the C-terminus. This, assuming that if the protein's function, expression or localization is not affected by the position of the tag, we would get the same population percentage for N-terminus and C-terminus tagged strains after the competition. If the tagging of a given protein at the N-terminus represents an advantage or is less deleterious than tagging at the opposite protein terminal, the N-tagged strain population percentage should be larger at the end of the competition, making this strain a "N-winner". If by the end of the competition assay, the most abundant population corresponds to the C-tagged strain, this would translate into such POI being a "C-winner", and tagging it at the C-terminus is either advantageous or less deleterious than the opposite tagging position.

Additionally, one needs to keep in mind the possibility that for some proteins, the tag at a given position might be more advantageous than no tag at all.

To test if this competition approach would be informative in yeast, I decided to reproduce a previous liquid competition assay for budding yeast in which 57 proteins were tagged at either C-terminus or at the N-terminus with GFP (Weill et al. 2019). While the C-tagged strain expressed a mCherry as cytosolic marker, the N-tagged strain was non-fluorescent, making it possible to quantify each population percentage at any given time point by using flow cytometry. I randomly chose 12 of the ORFs used in that study, and created two strains per POI, where I tagged them with mNG at the N-terminus expressing no cytosolic marker or tagging them at the C-terminus and expressing a mCh in the cytosol.

To confirm the tagging efficiency of those strains, I performed flow cytometry (Figure 1a) to detect the mNG signal (N-tagged strains) or both mNG and mCh signals (C-tagged strains).

As a second approach for determining the tagging efficiency, I also performed confocal fluorescence microscopy to manually count and calculate how many cells per fov per strain were carrying the mNG signal (Figure 1a). Through both approaches the tagging efficiency of all 24 strains was above 97.5 %. As seen in bottom part of Figure 1 panel a, the tagging efficiency with mNG was slightly higher for C-tagged strains than for N-tagged strains for 11 ORFs, only YNL121C had equal tagging efficiency at both protein termini.

Some of the proteins tested showed the same protein localization when tagged with mNG at the N-terminus and at the C-terminus, like the ORF YBR162W-A exhibited an ER localization in both strains (Figure 1a) as previously reported when tagged with GFP (Weill et al. 2019). In other cases, strains showed a different localization when



tagged at the N or at the C-terminus. The ORF YNL121C tagged at the N-terminus seemed to have a cytosolic localization while at the C-terminus it localized clearly to the mitochondria as expected, since it is one of the components of the translocase of outer membrane (TOM) complex. The mislocalization of YNL121C when tagged at the N-terminus can be explained since it was reported before that this gene carries a signal peptide (SP) at the N-terminus required for its correct targeting (Yofe et al. 2016).

ORFs such as YAL005C and YBR162W-A, had the same localization when tagged at C and N-terminus with mNG and also with GFP (Weill et al. 2019) (Figure 1a).

After competing the 12 strain pairs in liquid and agar media, we obtained the same trends of C-tagged and N-tagged population increase and/ or decrease as the study we were reproducing (Weill et al. 2019). I showed an example for the competition read-out of the ORF YPL204W (Figure 1b), where we can observe that in both competition approaches this ORF behaves as a N-winner, where the mCh population (C-tagged) decreases along time compared to the mNG-only population (N-tagged).

### **Using a non-fluorescent tag for competition assays**

Since my aim was to comparatively study the effect that the tag position has on fitness, I needed to choose a tag. For this, I considered the possibility that the effect of the tag on a given protein, might change not only based on which protein-termini it is incorporated but also based on the size of the tag. For this reason, I decided to use the non-fluorescent Halo tag, whose size (33 kDa) is very similar to commonly used FP and therefore the outcome of this study could be extrapolated to the effect that any other FPs or epitopes of a similar size would have on each protein. Moreover, several Halo tag ligands are already available, which make possible the use of the same genetic construct to perform protein isolation, protein interaction assays, imaging etc (England, Luo, and Cai 2015). For this reason, I considered constructing two genome-wide libraries carrying the Halo tag, as a valuable resource that could be used for different studies.

Before constructing new libraries, I carried out a proof-of-principle experiment (Figure 3), where I repeated the liquid and agar-based competition assay with the same 12 ORFs but now using Halo-Myc or Myc-Halo as N or C-tags respectively. From liquid and agar competition assays, I obtained for all the ORFs the same fitness effect than before with mNG as a tag. Having this confirmation, I decided to create the two genome-wide libraries where all ORFs available in the N and C-SWAT libraries (Yofe et al. 2016; Meurer et al. 2018) are now tagged by the Halo tag instead of the SWAT module. I used the SGA (A. H. Y. Tong and Boone 2006) approach to perform the high-throughput tagging of the ORFs; hence I created a N-tagged library (Halo-3xMyc-ORF) of 6144 strains and a C-tagged library (ORF-Halo-3xMyc) of 5760 strains.

The proof-of-principle experiment was performed using Myc-Halo and Halo-Myc tags, due to their immediate availability in the laboratory. Nevertheless, for the creation of the genome-wide libraries I decided to use three copies of the Myc tag instead, to increase the detection sensitivity of any strain in western blot or immunoprecipitation experiments.

## Creation of genome-wide Halo tagged libraries and their competition

Since the Halo tag is a non-fluorescent protein, during the SGA procedure for high-throughput tagging, I included an additional plate carrying high abundance proteins that were tagged at the N-terminus or at the C-terminus with mNG instead. I measured the tagging efficiency in those strains with flow cytometry (Figure 4a). All strains had a good tagging efficiency, where almost for all ORFs the tagging efficiency for N-tagged and C-tagged pairs was the same. Nevertheless, for ORFs like SSA1/YAL005C, PGK1/ YCR012W and ADK1/YDR226W, there was a bigger tagging efficiency difference between strain pairs.

Interestingly, when previously I used mNG as a tag there was a consistent tagging efficiency difference between N-tagged and C-tagged strains and for all ORFs tested (except YNL121C). In that assessment, the N-tagged strains had a lower tagging efficiency compared to their corresponding C-tagged counterparts, in around 1 % to 2 % difference (Figure 1a).

Nevertheless, based on the measured tagging efficiency of the reference ORFs tagged with mNG (Figure 4a) I assumed the high-throughput tagging procedure with Halo was also successful for the majority of the strains of the libraries.

Once libraries were constructed and their tagging efficiency estimated based on a reference subset, I proceeded to re-arrayed the final N-tagged and C-tagged strains which had a counterpart in the opposite library. Even though both liquid and agar competition assays lead to the same readouts, performing a liquid competition assay and seeing a fitness difference took half of the time. Therefore, after rearraying, I performed a genome-wide competition assay in liquid of both subsets of the Halo libraries (Figure 4b). In total, 5340 strain pairs were competed for four days, where the population percentage of each day was measured using flow cytometry.

Afterwards, I decided to estimate the effect of the tag position on fitness by calculating the ratio of percentage of C-tagged cells/ N-tagged cells per well, per day. Normalizing the ratio per ORF, to the ratio of day 1 and fitting a linear regression model to the time points (x-axis) against the log of the normalized C-population/ N-population ratio (y-axis) (Figure 5b).

When the slope of the fitted model was positive, the C-tagged strain exhibited a better fitness than the N-tagged strain (C-winner), ending up with a higher percentage of the C-tagged population by the end of the competition. If the slope had a negative value, the N-tagged strain was the most abundant population at the end of the competition (N-winner). And in the cases where both population percentages remained more or less constant along the competition, the slope value was close to 0 (Figure 5b).

When plotting the genome-wide distribution of the fitness effect (Figure 5a), I realized the overall distribution looked almost normal, with a mean fitness effect value of 0.02 the distribution seemed slightly skewed over the right of the x-axis. This could mean that all the C-tagged strains could have a slight benefit over the N-tagged strains, which might be due to the different cytosolic markers of each original donor strain.

When a threshold for distinguishing between N-winners and C-winners was not decided, I assumed all strains located at the left side of the mean of the distribution

were most likely N-winners. Contrary, the strains at the right hand side of the distribution's mean were in majority potential C-winners. The distribution of the fitness effect had a range of -1.9 to 1.2.

When looking through individual ORF competition results across the distribution (Figure 5b), I confirmed that YPL204W is a N-winner and also that YBR162W-A is a C-winner in my competition screen, as it was previously reported (Weill et al. 2019). Which meant, that without setting up any precise threshold, the left and right sides of the distribution's mean were a good starting point. When checking the N-tagged and C-tagged population percentages for each day for the only ORF in my screen for which the fitness effect was 0, YOR101W, I realized there was no N-tagged strain present at any time point. So I went and looked at the photographs I took from the initial N-SWAT and C-SWAT libraries and confirmed that this ORF had a spot in the N-SWAT library but this contains no colony, reason why I do not have a colony in the final Halo-3xMyc-ORF library.

To understand how each slope value translates in terms of population percentages, I looked at the N-tagged (in green) and C-tagged (in red) strain populations at each time point (Figure 5b). Having a slope of 0.59 translated into a final population ratio of 23.9 % N-tagged and 70.4% C-tagged. When fitness effect size is smaller, the slope value is also close to the center of the distribution, like for YPL204W  $m = -0.13$  with a N-tagged population of 67.3 % and C-tagged population of 32.6 % at the end of the competition.

In order to estimate the technical and biological noise in the final distribution, I decided to select ORFs located with different fitness effect value, therefore I chose YPL093W with  $m = -0.65$ , YLR024C with  $m = -0.3$  and YER177W with  $m = 0.02$ . Together with those, I decided to also compete the initial C-donor against the N-donor strains, which carry no tag but differ in the cytosolic marker that each one is expressing. I competed each strain pair across 92 wells (Figure 6a) in order to calculate the standard deviation (sd) across a plate.

Across all strain pairs tested, the sd was very consistent, for YLR024C, YER177W and for the donor strains the  $sd = 0.01$  while for YPL093W the  $sd = 0.02$ . Taking the maximum sd value, I set the significance threshold at five sigma ( $5\sigma = \pm 0.1$ ). With this threshold, out of 5340 ORFs competed (Figure 6, b), there were 292 N-winners (12%), 4310 ORFs were neutral (80%) and there were 649 C-winners (5%). Having this information available, anyone could look up for their POI and decide at which protein termini tagging to tag their POI in order to minimize the effect on fitness.

This dataset can also aid the identification of proteins that might contain an unknown localization signal either at the N-terminus or at the C-terminus.

Previously, when estimating the tagging efficiency of the genome-wide libraries using a reference plate of ORFs tagged with mNG instead Halo tag (Figure 4a) I observed that almost all reference strains had a high tagging efficiency, where the N-tagged strain had a very similar tagging efficiency as the corresponding C-tagged counterpart. However, for three reference ORFs there was a bigger difference between strain pairs. After we set the threshold for defining N-winner, neutral ORFs and C-winner, I revisited this observation thinking that the difference in tagging efficiency with mNG as a tag,

could have been an early indication of the effect of the position of any tag for the fitness of a strain when tagging those specific ORFs.

Nevertheless, in this case this does not seem to be an indication of a fitness effect caused by the position of the tag, since for SSA1, for example, the tagging efficiencies were 98.4 % for the N-tagged strain and 92.3 % for the C-tagged strain and in the screen SSA1 had a fitness effect  $f = -0.05$ , falling within the neutral ORFs.

For PGK1 the tagging efficiency for N-tagged strain was 93.9 % and 99.9 % for the C-tagged strain, and during the competition screen its fitness effect was calculated in  $.026$ , being a neutral ORF as well. Finally, for ADK1 the tagging efficiency for the N-tagged strain was 94.2 % and for the C-tagged strain was 99.8 %; and falling within the neutral ORFs classification having a fitness effect of  $0.06$  in the screen.

Additionally, I also checked two reference ORFs where the tagging efficiency when tagged with mNG was the same for N and C-tagged strains: TDH2/YJR009C and VMA1/ YDL185W. For both ORFs, the tagging efficiency was very high, 99.8 % and 99.9 % respectively. Nevertheless TDH2 lies within the neutral range of the competition assay with a fitness effect value of  $-0.01$  while VMA1 is a C-winner with a fitness effect value of  $0.75$ .

To confirm the reproducibility of our findings, during the genome-wide screen I included a technical replicate of the Plate 1 measured in flow cytometry; this replicate was measured at the end of the competition assay. I used those two replicates to calculate the correlation between the fitness effect detected in each one (Figure 7). For the 92 ORFs tested, I got an overall all good correlation. Concluding my genome-wide competition assay is reproducible.

After confirming the reproducibility, I revisited the results for N-donor and C-donor competition (Figure 7), in which the average fitness effect was  $0.048$ . This fitness effect value indicates a fitness advantage of the C-donor over the N-donor strain, which was unexpected. Genetically, both donor strains come from the same background (yMaM1205) and the only difference is that the N-donor strain has mNG as a cytosolic marker while the C-donor strain has mCherry as such. Since both FPs have a very similar molecular weight, mNG being  $26.6$  kDa and mCh  $26.7$  kDa, the reason for mNG to compromise the strain fitness should be other than its size. One possibility would be that the degradation of mNG might be incomplete, as reported before for sfGFP (Khmelinskii et al. 2016), which could be constrain for the overall homeostasis of the cell.

As seen in the Figure 8 panel a, when looking at the competition outcome of one well of N-donor against C-donor, the fitness effect of  $m = 0.05$  translates into a 5% more abundance of the C-tagged population over the N-tagged one. Even when the reason behind this advantage difference is yet unknown, we could use this information to correct the overall fitness effect distribution. When performing a growth curve of the donor strain and the wt strain (yMaM1205), this fitness difference is not observable in this assay (Figure 8b). It remains an open question, why mCherry provides a fitness advantage over the mNeonGreen as a cytosolic marker.

## Characteristics of N-winners and C-winners

Having set a threshold for N-winners and C-winners across the yeast genome, I continued gathering information of common characteristics of N-winner, C-winners and Neutral ORFs.

It is known, that a set of yeast ORFs harboring a stretch or repeats at the start codon or at the stop codon, tend to have problems when endogenously tagged at N-terminus or C-terminus respectively (Khmelniskii et al. 2011). Within the final Halo libraries, only 32 ORFs were reported to be difficult to tag at the C-terminus (C-problematic), 29 ORFs were described as problematic to be tagged at the N-terminus (N-problematic), 176 ORFs were reported to have problems to be tagged at both C-terminus and N-terminus and the rest 5015 ORFs have not been reported to have any difficulty when tagged endogenously by DNA double-strand break repair by homologous recombination.

Interestingly, I observed that out of 32 C-problematic ORFs, 91% were falling within the Neutral range (Figure 9). Among the 29 N-problematic ORFs, 93% were Neutral as well.

Finally, from the 176 ORFs reported to be both N-problematic and C-problematic, 155 were among the Neutral range in my screen, 5 were N-winners and 16 C-winners. This is quite interesting, because it implies that within my competition screen, when ORFs have problems to be effectively tagged at both termini, the C-tagged strain would have a fitness advantage. This could happen if the C-terminus was less likely to end up tagged, and competing a “wt” strain against a N-tagged strain would have a benefit. Another explanation would be that the N-terminus tagging for such ORFs might be less likely, therefore they will have an advantage over the strain where tagging occurred at the opposite end.

Overall, I would need to gather a larger list of ORFs reported to be problematic to be tagged. The current dataset does not help to elucidate if there are trends due to this problematic tagging condition.

Continuing the characterization of N-winners, Neutral ORFs and C-winners, I proceeded to evaluate if there was a trend for gene essentiality and their susceptibility to the position of the tag. Within the 5340 tested ORFs (Figure 10a), 898 are reported to be essential (Després et al. 2020) and the remaining 4354 non-essential ORFs.

I observed that for both essential and non-essential ORFs, a better fitness is achieved when they are tagged at the C-terminus (Figure 10a). Within the essential class 33% of the ORFs were hits, 11% being N-winners and 23% C-winners. And within the 4354 non-essential ORFs 14.3% were hits, 4.3% N-winners and 10% were C-winners.

By calculating the number of hits I would get for different fitness effect thresholds and by defining a hit as an ORF located either at the left or at the right side of a given threshold, again I observed that there are more hits for essential genes than for non-essential genes (Figure 10b). This is congruent with the idea that any perturbation to an essential gene will lead to worse consequences than when perturbing a non-essential gene.

I considered that a reason for a larger number of C-winners over N-winners, regardless of being essential or non-essential ORFs, could be the presence of different signals (like signal peptides) located at the N-terminus of such ORFs. After manually curating a list of ORFs carrying a signal at the N-terminus (such as SP), or at the C-terminus (like TMDs), I looked at where in the distribution (Figure 11) ORFs carrying those signal types would be. In order to draw solid conclusions from this analysis, I need to further expand the list of signals in the curated list. Nevertheless, it seems that the presence of signals at the N-terminus is one of the reasons for such ORFs to have a fitness advantage when tagged at the C-terminus instead (Figure 11). As well, the presence of signals at the C-terminus could be the explanation for ORFs behaving as N-winners.

### **Validation of tag presence and fitness effect**

The previous genome-wide competition assay indicates that the relative fitness of some strains is affected by the tag position with respect to the ORF of interest. In my genome-wide competition assay, I tagged 5340 ORFs of *Saccharomyces cerevisiae*, at the C and at the N terminal with the non-fluorescent Halo tag followed by a 3xMyc tag. Nevertheless, since Halo tag is a non-fluorescent tag, direct validation of the presence of a tag at the N-terminus or C-terminus of the POIs was not possible.

Therefore, I decided to validate my previous findings by randomly selecting 45 strains across the distribution range of the fitness effect (Figure 12a), streak them for single colonies and performing western blots for detecting the 3xMyc tag located always in between the Halo tag and the POI.

Among the 45 pairs of strains and two clones used per strain for the detection of the 3xMyc tag by western blotting (Table S2), I observed some degree of heterogeneity in the frequency of finding positive clones (Figure 12b).

For eight ORFs (18% of the tested ORFs), when testing two clones per strain type, I found only one positive clone for the N-tagged strain and two for the C-tagged strain. On the other hand, for four ORFs (9% of the tested ORFs) I found two positive clones for the N-tagged strain but only one for the C-tagged counterpart. For the rest 73% of the tested ORFs, both of the evaluated clones for N-tagged and C-tagged strains were positive.

This heterogeneity of positive clones could indicate that the tagging efficiency of the reference ORFs (Figure 4a) was not representative of the genome-wide tagging efficiency when using Halo tag and 3xMyc as tags instead of mNG.

Since for some of the ORFs that did not have a good correlation between the validation and the screen experiments, this could be an indication that either N or C strain population in the final library might have had a lower tagging efficiency and the competition was not fair. This is a possibility supported by the fact that for such outliers (Figure 12c) both clones for both strains were found positive and the band size was around the expected size. One example of such ORFs is YFL045 (Figure 12c, circled in light blue), for which all clones were positive when checked by wb but which relative fitness effect value was very contrasting in between the screen and the validation, being  $m = 0.45$  and  $m = 1.35$  respectively. This would mean that it is possible that the C-tagged strain might have had a lower tagging efficiency in the final colony of the

screen compared to the clone used for the validation. I also observed the same for other ORFs that had a low correlation between validation and screen.

Another explanation that some spontaneous mutation arose for some proteins which have a higher mutation rate, such mutations could lead to the loss of protein function and therefore lack of protein signal in wb for some ORFs (Gou, Bloom, and Kruglyak 2019; Willensdorfer, Bürger, and Nowak 2007). On the same line of thinking, even when for the majority of the strain pairs I detected positive clones only, for some pairs the clones, either of the N-tagged clones or the C-tagged ones, exhibited a different signal intensity. This fact could also indicate the presence of mutation in the protein which leads to different abundances of the same (Figure 12b).

I can also not discard the possibility of recombination problems for some ORFs, due to the presence of the triple copy of Myc tag, even when I used different codon usage for each copy. Therefore, I will create two new donor plasmids for N and C-terminus tagging with mNG-3xMyc, mNG-Halo-3xMyc and 3xMyc-mNG and 3xMyc-Halo-mNG respectively. I will use these donor plasmids to transform non-fluorescent donor strains in order to tag the reference ORFs (Figure 4a), whose mNG signal will be measured by flow cytometry. In this way, I will make sure that no heterogeneity in tagging efficiency is occurring either due to the 3xMyc nor the Halo tag presence. Again, this experiment will confirm that these tags are not causing any problem to the reference ORFs only. However, the hypothesis that the tagging efficiency in my genome-wide libraries is homogenous and any heterogeneity in wbs might be due to loss of function by spontaneous mutations arising more frequently for some proteins than others could still be valid; as well, I still could consider the idea that tagging efficiency for the reference ORFs is not representative of the genome-wide range.

Besides the heterogeneity within the frequency of positive clones per ORF, the western blot results seemed correct and reliable since across all of them the bands detected for each pair of strains (N-tagged and C-tagged per ORF) were around the expected size, or a little bit larger. Moreover, the bands of all positive clones per strain pair were always the same size.

While observing a signal in a western blot reflected the functionality of the protein and the presence of the 3xMyc tag, it did not confirm that the tag was present at the correct locus. This is why, I randomly chose three of the validation strain pairs (Table S2, indicated with \*) and confirmed that the tag was at the correct locus by PCR-amplifying the ORF-3xMyc-Halo and Halo-3xMyc-ORF junctions and sequenced them by the Sanger method.

Finally, after analyzing the western blot results and finding positive clones for all of the tested strains (Figure 12b), I selected one clone for each N-tagged strain and C-tagged strain and repeated the liquid competition assay. The relative fitness effect of the tag position per strain was calculated as described before. Comparing the fitness effect value for these 45 strains in the Validation against their values obtained during the genome-wide screen, the correlation was  $R = 0.81$  with an accuracy of  $p < 0.001$  (Figure 12c).

This high correlation in between relative fitness effects from both independent experiments, confirms not only that our method is reproducible, but also that the fitness effects observed in the genome-wide competition assay are a direct read-out of the

consequences of the position of the tag for each individual protein in the yeast genome.

All these experiments directly confirmed the presence of the 3xMyc tag and indirectly verified the presence of the Halo tag. To directly validate the presence of the Halo tag across the entire library, I propose selecting a subset of strains and optimizing the use of a Halo tag antibody for western blot detection. However, this approach would confirm the tag presence but only for a limited subset of strains.



## **MATERIAL AND METHODS**

### **Yeast competition assay in liquid media**

I used a flow cytometer (BD LSRFortessa), and its acquisition software BD FACSDiva v9.0.1, for the quantification of the percentage of fluorescent signals in yeast cells.

The flow cytometer was used together with its auto-sampler for 96-well plate format in the high-throughput (HTS) mode. The settings for the loader were as follows: sample flow rate 0.5  $\mu\text{L}/\text{s}$ , sample volume 25  $\mu\text{L}$ , mixing volume 100  $\mu\text{L}$ , mixing speed 100  $\mu\text{L}/\text{s}$ , number of mixes 3 and wash volume 800  $\mu\text{L}$ . The gating of the fluorescent populations was defined manually. Firstly, the whole cell population was filtered based on cell size by using the forward and side scatter (FSC and SSC) to distinguish from double cells and single cells. In the case of the genome-wide C-tagged tFT library, the fluorescent populations were defined manually in four gates: non-fluorescent gate, cells with green fluorescence, cells with red fluorescence and double positive cells (DP which had as we expected both green and red fluorescence). Per strain, 10000 cells were measured in the flow cytometer.

The protocol to prepare cells for fluorescent measurement in the mentioned flow cytometer is as follows:

1. Under sterile conditions fill up the 96-well plate (PhenoPlate 96-well microplates, PerkinElmer) wells needed with 150  $\mu\text{L}$  of the required medium
2. Seal the plate with an air-o-seal. Close it with its plastic lid, place it in the incubator at 30 °C and 220 rpm. Add a box's lid on top of it to prevent evaporation. Grow to saturation (overnight)
3. Next day, dilute the saturated cultures and grow them for 6 hours to  $6\text{-}8 \times 10^6$  cells/mL
4. Perform the quality check of the lasers performance in the flow cytometer, prime the HTS plate-sampler and proceed to measure the Single-cell fluorescence intensities by the 488 and 561 nm lasers, for the red and green fluorescence respectively

**Note 1:** For the genome-wide high-throughput flow cytometry, I used SC low fluorescence medium (6.7 g/ L bacto yeast nitrogen base without amino acids and without folic acid and riboflavin (LoFlo, Formedium), 2 g/ L amino acid dropout mix, 2% (w/v) glucose) supplemented with 300 mg/L adenine (A8626, Sigma-Aldrich).

Note 2: For flow cytometry with plates and HTS standard acquiring mode, the next settings were Events to record 10 000, Sample flow rate ( $\mu\text{L}/\text{s}$ ) 0.5, Sample volume ( $\mu\text{L}$ ) 100, Mixing volume ( $\mu\text{L}$ ) 100, Mixing speed ( $\mu\text{L}/\text{s}$ ) 180, #Mixes 3 and Wash volume ( $\mu\text{L}$ ) 800.

## Yeast competition assay on agar media

The agar-based competition assay, which was measured using a plate reader (SparkControl 20M plate reader with monochromator, TECAN).

## Measurement of cytosolic fluorescence of donor strains for the proof-of-principle (PoP)

In order to confirm that mCherry and mNeonGreen were bright enough as cytosolic markers, I transformed them into the strain yMaM1205 and measured their fluorescence on agar (Figure 2) using a plate reader (SparkControl 20M plate reader with monochromator, TECAN).

The donor strains measured were the following: yKBJ0147 as C-donor strain (MAT $\alpha$  lyp1 $\Delta$  his3 $\Delta$ 1 leu2 $\Delta$ 0 ura3 $\Delta$ 0 met15 $\Delta$ 0 can1 $\Delta$ ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI leu2 $\Delta$ ::natNT2-TEF1pr-mCh-CYC1ter) and the N-donor strain yKBJ0148 (yMaM1205 (MAT $\alpha$  lyp1 $\Delta$  his3 $\Delta$ 1 leu2 $\Delta$ 0 ura3 $\Delta$ 0 met15 $\Delta$ 0 can1 $\Delta$ ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI) leu2 $\Delta$ ::natNT2-TEF1pr-mNG-CYC1ter).

I followed the following steps for measuring fluorescence on agar plates with the plate reader:

1. I inoculated a 96-well plate with the respective strains (Figure 2, a), with 150  $\mu$ L per well of SC-Leu+Ade low fluorescence filter sterilized media. I grew it overnight at 30 °C and 220 rpm.
2. Next day I stamped onto an agar plate of SC-Leu+Ade in 384-format. I let it grow for 24h at 30 °C.
3. Next day I re-stamped from a 384-agar plate onto a new agar plate in the same format. I let it grow for 24 h at 30 °C.
4. I took a photo of it and measured the fluorescence of mNG and mCh in the plate reader.
5. For the analysis of the plate measurement:
  - Calculate the median and mean for the negative control (wt yMaM1205)
  - Calculate the median and average values of the mCh fluorescence of the yKBJ0147-4 (C-donor) and the yKBJ0148-1 (N-donor) strains

**Table 1.** Settings for the fluorescence intensity measurements, of donor strains carrying FPs as cytosolic markers. Measurements performed in the plate reader.

FT	Excitation wavelength	Emission wavelength	Excitation bandwidth	Emission bandwidth	Gain
mCherry	586 nm	612 nm	10 nm	10 nm	102
mNeonGreen	506 nm	524 nm	5 nm	5 nm	86

For mCh the mirror setting was 50% Mirror and for mNG Dichroic 510. For both FPs the number of flashes was set to 10.

After making sure donor strains had a bright enough cytosolic fluorescence, I proceeded to transform the following donor plasmids into their respective donor strain: C donor plasmid pKEK017 (pRS41K-myc-HaloTag) and N donor plasmid pKEK016 (pSD-N1-HaloTag-myc). The final donor strains were created and named as yKBJ0147-4 (C-donor) and yKBJ0148-1 (N-donor).

### **Creation of N and C-tagged libraries using Halo as a tag**

The donor plasmid type I (endogenous) for N-tagging pKBJ050-8 (pSD-N1-HaloTag-3xMyc) was constructed by linearizing the plasmid pKEK016 with BamHI-HF and SpeI-HF, a fragment of 536 0bp was gel-purified. The insert was generated by PCR-amplifying the Halo tag from pKEK016 with the oligos 20/100 and 21/1; a gel-purified fragment of 1042 bp was ligated to the vector by Gibson assembly. The newly assembled plasmid was transformed into DH5 $\alpha$  (C2987 NEB) cells and plated on LB+Amp, and it was validated by sequencing with the oligos 21/2 and 21/3.

The donor plasmid type I (Endogenous) for C-tagging pKBJ051-1 (pRS41K-3Myc-HaloTag) was constructed by digesting the plasmid pKEK017 with BamHI-HF and SpeI-HF; the vector of 5172 bp was gel-purified. For the insert, Halo tag was PCR-amplified from pKEK017 with the oligos 21/4 and 21/5. The insert of 1045 bp was gel-purified and assembled together with the vector by Gibson assembly. This plasmid was then transformed into DH5 $\alpha$  (C2987 NEB) cells and plated on LB+Amp. The new plasmid was sequenced with the oligos 21/6 and 21/7.

When a Halo tag is genetically encoded, there are several dyes and ligands that can be used to detect it (England, Luo, and Cai 2015). Nevertheless, cell labeling in live yeast has a low efficiency of the dye retention or ligand introduction, making it necessary to aid the staining by cell electroporation which might affect yeast physiology in unknown ways (Ball et al. 2016; Stagge et al. 2013).

It has been shown that disrupting the gene PDR5, which encodes a transporter for xenobiotic compounds resistance, increases the dye incorporation for C-terminus fusions with Halo tag (Ball et al. 2016). Therefore, I decided to disrupt this gene in the donor strains for genome-wide tagging with Halo tag, in case our laboratory or someone else would envision using dyes or Halo ligands with this library in the future.

For the donor strains construction I used the genetic background of yMaM1205 (MAT $\alpha$  lyp1 $\Delta$  his3 $\Delta$ 1 leu2 $\Delta$ 0 ura3 $\Delta$ 0 met15 $\Delta$ 0 can1 $\Delta$ ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI). To construct the N-donor strain, I linearized the plasmid pKBJ048 with ZraI and DraIII enzymes, then the fragment of 5340 bp, which corresponded to the leu2dC-natNT2-TEF1pr-mNG-CYC1ter was gel-purified and transformed into competent cells of the strain yMaM1205. The new strain yKBJ0148-1 was plated on SC(MSG)-LEU/ARG/LYS+NAT and its validation was performed by detection of the mNG signal under confocal fluorescence microscopy.

I prepared competent cells of this new strain and transformed them with the Pdr5-HphNT1 deletion cassette amplified from the plasmid pFA6a-hphNT1. This new strain

yKBJ0201-2 was validated by its growth on YPD+Hyg and by colony PCR with the oligos 3/80 and 2/57 to detect the presence of HphNT1-promoter region (TEFpr).

Finally, this donor strain yKBJ0201-2 was transformed with the donor plasmid type I for N-terminus tagging pKBJ050-8; validated by its growth on YPD+G418.

This final N-donor strain has the next genotype: yMaM1205 (MAT $\alpha$  lyp1 $\Delta$  his3 $\Delta$ 1 leu2 $\Delta$ 0 ura3 $\Delta$ 0 met15 $\Delta$ 0 can1 $\Delta$ ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI) leu2 $\Delta$ ::natNT2-TEF1pr-mNG-CYC1ter pdr5 $\Delta$ ::hphNT1 pSD-N1-KanMx-HaloTag-3Myc

yKBJ0201-2 (2087) competent cells transformed with donor plasmid pKBJ050-8 (1226), plated 20% on YPD+G418. Validation: Grows on YPD+G418.

In order to create the C-donor strain, I also started by linearizing a plasmid carrying the cytosolic marker for this strain. The plasmid pKBJ047-1 was linearized with ZraI and HpaI and the fragment of 4816 bp was gel-purified and transformed into yeast yMaM1205 strain competent cells. The new strain yKBJ0147 was plated on SC(MSG)-LEU/ARG/LYS+NAT and the mCh signal was validated by confocal fluorescence microscopy.

This new strain was transformed with Pdr5-HphNT1 deletion cassette amplified from pFA6a-hphNT1 plasmid. The transformation was plated on YPD+Hyg and the deletion of PDR5 was validated by Colony PCR with the oligos 3/80 and 2/57 for checking the presence of HphNT1-promoter region (TEFpr). This new strain was named yKBJ0200.

Finally, the strain yKBJ0200 was transformed with the donor plasmid pKBJ051 and plated on YPD+G418 media. This final C-donor strain has the next genotype: yMaM1205 (MAT $\alpha$  lyp1 $\Delta$  his3 $\Delta$ 1 leu2 $\Delta$ 0 ura3 $\Delta$ 0 met15 $\Delta$ 0 can1 $\Delta$ ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI) leu2 $\Delta$ ::natNT2-TEF1pr-mCh-CYC1ter pdr5 $\Delta$ ::hphNT1 pRS41K-KanMx-3Myc-HaloTag.

In order to perform high-throughput swapping of the SWAT module in the N-SWAT and C-SWAT libraries (Yofe et al. 2016; Meurer et al. 2018), I used the synthetic genetic array (SGA) method (A. H. Tong et al. 2001; Baryshnikova et al. 2010). For the SGA I crossed the donor strains, carrying the Halo-3xMyc and 3xMyc-Halo tags respectively. Through a series of replica-pinning on different selection media, I generated two new N and C-tagged libraries with our desired tag. The genetic crossing was semi-automated, using the pinning robot (Singer Instruments). The SGA steps I followed for the construction of the N-Endo-Halo-3Myc and C-Endo-3Myc-Halo libraries were the following:

1. Grow the C-SWAT and N-SWAT libraries on SC-Ura and grow the donor strains on YPD+G418 on 384-colony array format
2. Mate the N and C-SWAT libraries and the respective N and C-donor strains on YPD
3. Select for diploids on SC(MSG)-Ura+G418 plates. Repeat the selection twice
4. Pin the diploids on Sporulation media and let grow for 7 days at 23 °C
5. Select for haploids:

Haploid selection 1: SC(MSG)-Leu/Arg/Lys/Ura + Can/Thi  
Haploid selection 2: SC(MSG)-Leu/Arg/Lys/Ura + Can/Thi/G418  
Haploid selection 3: SC(MSG)-Leu/Arg/Lys/Ura + Can/Thi/G418/Nat  
Haploid selection 4: SC(MSG)-Leu/Arg/Lys/Ura + Can/Thi/G418/Nat+Hyg

6. Induction of I-SceI enzyme for tag swapping on SC-Leu + Raf/Gal. Repeat the induction twice
7. Counter selection on SC-Leu + 5 FOA against strains carrying the acceptor module. Repeats this selection twice
8. Pin the new two library strains on SC-Leu+Ade plates and grow for at least 1 day to have enough material to measure tagging efficiency and to prepare glycerol stocks
9. Measure tagging efficiency of your reference plates by following the protocol "Measuring fluorescence signal using Flow cytometry"
10. Inoculate 384-well plates (Greiner Bio-One) with 50  $\mu$ L of SC-Leu+Gly(15%) media. Seal with air-o-seals and grow strains at 30 °C in a standing incubator for 2 days. For these two libraries, two copies of each were stored in 384-format at - 80 °C

In order to validate the presence of the tag in the final library, I performed fluorescent measurements of 21 strains expressing proteins known for their high abundance in yeast and which have previously been used for tagging efficiency validation in other studies (Meurer et al. 2018). These strains were crossed to a N-donor strain carrying mNG as tag and to a C-donor strains carrying as well mNG as tag. They followed almost the same SGA procedure as the other two libraries. I measured green fluorescence of these strains using flow cytometry and their tagging efficiency was scored as the percentage of cells with fluorescence in each fluorescent population. The settings for the measurements in flow cytometry are described in the respective protocol above for "Measuring fluorescence signal using Flow cytometry".

**Note:** Take photographs of all of the plates in several steps along the SGA. For sure take an initial photograph of N and C-SWAT libraries and a final photograph of the libraries on SC-Leu+Ade plates.

### Yeast whole cell extract using TCA

First: Have cold ddH<sub>2</sub>O on ice (1 ml per sample) ready before starting the protocol. Work on ice as indicated below.

1. Collect 2-3 OD of cells in microcentrifuge tubes by centrifugation at 3.200 rpm for 2 min at room temperature
2. Resuspend cell pellet in 1 ml of cold water by vortexing. Keep tubes on ice  
Alternatively, resuspend some cells from a fresh patch in 1 ml of cold water  
Alternatively, just use 1 ml of a culture (only for synthetic media, rich media have too much material that will contaminate whole cell extracts)
3. Add 150  $\mu$ l of 1.85 M NaOH. Mix by vortexing. Incubate 10 min on ice

4. Add 150  $\mu$ l of 55% (w/v) TCA (w/v). Mix by vortexing. Incubate for 10 min on ice
5. Centrifuge at 14000 rpm for 15 min at 4°C. Remove as much supernatant as possible by aspiration, centrifuge again for 1 min and remove remaining supernatant
6. Add HU buffer to the pellet (precipitated proteins) at 50  $\mu$ l HU buffer per OD. Store samples at -20°C or proceed
7. Resuspend pellet by vortexing. If supernatant turns yellow, add 1-2  $\mu$ l of 2 M Tris-HCl pH 8 to neutralize
8. Heat samples for 15 min at 65°C with mixing to fully dissolve the pellet. Centrifuge at 14000 rpm for 10 min before loading on the gel. Load 10-15  $\mu$ l per well

**Note:** If necessary, dilute MW marker with HU buffer so as to load the same volume per well. Load empty wells with the same volume of HU buffer. HU buffer (high-urea buffer): 8 M urea, 5% SDS, 200 mM Tris-HCl pH 6.8, 0.1 mM EDTA, bromophenol blue (according to taste). Make 10 ml aliquots in 15 ml falcon tubes and freeze at -20°C. Before use, add DTT to 15 mg/ml final concentration. SDS-PAGE: ~30 mA per mini-gel, 150-200 V (max).

### SUPPLEMENTARY

**Table S1.** Table for noise measurement results.

Strain pairs	SD of slope	Avg slope (Validation)	Avg slope (Screen)
Donors	0.012	0.049	NA
YPL093W	0.023	-0.601	-0.653
YLR024C	0.014	-0.315	-0.303
YER177W	0.017	0.133	0.025

**Table S2.** List of strains that were tested and results of detection of 3xMyc by western blot. \* indicates the strains for which the ORF-3xMyc-Halo and Halo-3xMyc-ORF junctions were confirmed as correct by Sanger sequencing.

ORF	N-tagged strain positive clones	C-tagged strain positive clones
YJR010C-A	2	2
YKR095W-A	2	2
YPL106C	1	2
YER136W	2	2

YBR049C	2	2
YER157W	2	2
YFR011C	2	2
YLR074C	2	1
YOR279C	2	2
YKR083C	2	2
YKL216W	2	2
YGL121C	2	1
YGL234W	2	2
YOL109W	2	2
YBL023C	2	2
YBR025C	2	2
YBR267W	2	2
YHR173C	2	2
YHR037W	2	2
YBR181C	1	2
YPR108W	1	2
YFR033C	2	2
YML004C	2	1
YMR140W	2	1
YLR028C	2	2
YBR101C	2	2
YBR137W	2	2
YFL045C	2	2
YBL002W	2	2
YPL160W	2	2
YGL009C	1	2
YMR020W	1	2

YGR234W	2	2
YNL121C	2	2
YDL167C	2	2
YOR253W	1	2
YJR009C	2	2
YLR077W	2	2
YML028W	2	2
YIL138C *	2	2
YDR074W *	2	2
YPL094C	2	2
YMR110C *	2	2
YIR001C	1	2
YDL051W	1	2

YIL138C\* m = - 0.5 N-winner

YDR074W\* m = - 0.48 N-winner

YMR110C\* m = 0.75 C-winner



## **CHAPTER 2. ASYMMETRIC PROTEIN INHERITANCE IN YEAST REPLICATIVE AGING**

### **INTRODUCTION**

Is aging the result of a genetic program or is simply an inevitable, random consequence of life? The question of whether aging is a deliberate process or merely an incidental byproduct or epiphenomenon of life remains a topic of ongoing investigation, and there is no definitive answer to this question (Helfand and Rogina 2003).

### **Aging and cellular senescence**

“Aging can be defined as the time-related deterioration of the physiological functions necessary for survival and fertility” (Gilbert 2000). The aging phenomena can be analyzed from two different perspectives, from the organismal level and from the cellular level.

At the organismal level aging is marked by the loss of fitness, which encompasses a decline in various physiological and functional aspects (López-Otín et al. 2013; Brian K. Kennedy and Pennypacker 2014). This includes reduced organ function, impaired immune responses, diminished cognitive abilities and an increased susceptibility to diseases (Brian K. Kennedy and Pennypacker 2014). One of the key challenges in aging research is to understand the mechanisms underlying these progressive declines in fitness and identify interventions that can ameliorate the aging process and/ or which can increase the healthspan (Brian K. Kennedy and Pennypacker 2014; López-Otín et al. 2016).

At the cellular level, the existence of cellular senescence was described almost six decades ago, by scientists who realized that serially *in vitro* cultured diploid human cell strains have a limited lifetime (Hayflick 1965). This further incapacity to proliferate was for the first time suspected to be a consequence of aging or senescence at the cellular level (Hayflick 1965). In different characterizations of several human cell strains from embryonic origin, the same research group concluded that the number of generations expected from a human diploid cell strain did not depend on the number of subcultivations but rather on the number of cell doublings. Concluding that each strain has the same doubling potential (Hayflick 1965).

These discoveries pointed at the fact that the same lineages of cells seemed to exhibit the same lifespan; which raised the question whether or not cellular aging is a genetic program.

The fact that cells don't proliferate forever in culture, contrary to many cancer cells, made cellular senescence to be considered as a potential cancer and tumor-suppressor mechanism (Steinkraus, Kaeberlein, and Kennedy 2008). On a different end, cellular senescence has also been considered an aging agent. This, due to the observations that tissue regeneration and repair deteriorate with time, and tissue renewal and functionality is essential for complex organisms (Steinkraus, Kaeberlein, and Kennedy 2008).

Understanding causes and consequences of aging at the cellular level, sheds light into how this process is connected to aging at the whole organismal level. Nevertheless, whether cellular senescence properly recapitulates organismal aging is still debatable.

### **The hallmarks of aging from mammals to single-celled organisms**

Along the years, diverse approaches have been used to try to find correlation between aging and different plausible causes of it. Recently, the aging field has established that to a certain extent, aging seems to be influenced by different conserved genetic pathways (López-Otín et al. 2013; Bitto et al. 2015).

In order to consider a given process as a hallmark of aging, three criterias have been proposed: it should be present in normal aging, its exacerbation should accelerate aging while its amelioration would retard aging (López-Otín et al. 2013) The proposed aging hallmarks are: genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient-sensing, mitochondrial dysfunction, cellular senescence, stem cell exhaustion, and altered intercellular communication (López-Otín et al. 2013).

While these hallmarks of aging are often associated with mammalian aging, many of them are evolutionarily conserved in simpler organisms. For instance, processes like genomic instability, telomere shortening, epigenetic alterations and mitochondrial dysfunction, which are key components of aging in mammals, have also been observed in various model organisms (López-Otín et al. 2013). The study of aging in these organisms provides valuable insights in the conserved mechanisms of aging and the potential for interventions to extend lifespan and promote healthy aging.

The long life span of common multicellular model organisms has always been a limitation to study aging, nevertheless, studies performed in models such as rodents have provided us with key knowledge about the genetic and environmental factors involved in this process. Studies using mammal cells in cell culture rose as a solution to the long life span constraint (Steinkraus, Kaeberlein, and Kennedy 2008). This alternative has been informative to understand different cellular aspects involved in cellular senescence, but remains questionable how translatable this knowledge is to understand organismal aging (Campisi and d'Adda di Fagagna 2007).

Studies involving various eukaryotic and invertebrate model organisms, like flies, worms and yeast, have significantly advanced our understanding of fundamental biological conserved processes, including aging (Brian K. Kennedy and Pennypacker 2014)(Steinkraus, Kaeberlein, and Kennedy 2008) & (Brian K. Kennedy and Pennypacker 2014). Due to their simpler genetic manipulation and short lifespan, single-celled organisms have played a significant role in better understanding the cellular, molecular and genetic basis of aging in mammals (Kenyon 2010; Fontana, Partridge, and Longo 2010).

Notably, research in the yeast *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* has shed light on essential aspects of aging, such as replicative aging and telomere attrition. Additionally, model organisms like *Caenorhabditis elegans* have contributed to our understanding of aging pathways, particularly those related to lifespan extension through caloric restriction and other interventions (Houtkooper, Williams, and Auwerx 2010).

Over the years, the single-celled eukaryote budding yeast, *Saccharomyces cerevisiae*, has been extensively used as a model for replicative aging. This, as the aging process in *S. cerevisiae* exhibits similarities to the hallmarks previously described for mammals (Janssens and Veenhoff 2016) .

### **Replicative aging and asymmetric cell division in budding yeast**

Among the studied species, most of the knowledge of aging is based primarily on measurements on age at the moment of death and survivorship curves (Helfand and Rogina 2003). Based on these read-outs, we have been able to shed light on which genetic or/ and environmental alterations change these read-outs in a given population (Helfand and Rogina 2003).

In *Saccharomyces cerevisiae*, there are two key approaches to interpret aging: Replicative life span (RLS) and chronological life span (CLS) (Steinkraus, Kaeberlein, and Kennedy 2008). The RLS assay determines the number of times that a cell divides. While in the CLS assay the read-out is the chronological time that the cells are unable to divide before losing viability (Janssens and Veenhoff 2016).

In all organisms, ranging from bacteria to mammals, asymmetric cell division is a common process whereby asymmetric inheritance of cellular components gives rise to two daughter cells with different characteristics and fates (Shahriyari and Komarova 2013). The asymmetry in cell fate is an outcome of the asymmetry of cell fate determining factors (Neumüller and Knoblich 2009). During cell division, fate determinants may be polarized within a cell and then segregated asymmetrically (Venkei and Yamashita 2018).

In the budding yeast, *Saccharomyces cerevisiae*, the division of cells is asymmetrical, producing two cells that have different characteristics (Higuchi-Sanabria et al. 2014): mother and daughter. During the text, the terms daughter or bud will be used interchangeably.

While the mother can go through a limited number of divisions before senescence, daughter cells are born rejuvenated irrespective of the age of their mother cell (Higuchi-Sanabria et al. 2014; Steinkraus, Kaeberlein, and Kennedy 2008) and their cell cycle is longer than their mothers previous cell cycle (Hartwell and Unger 1977). Additionally, mother cells have a larger size than daughter cells (Hartwell and Unger 1977).

The asymmetry in yeast replicative aging has been thought to occur due to the asymmetric inheritance of senescence factors like extrachromosomal ribosomal DNA circles (ERCs) (Ganley et al. 2009; Sinclair and Guarente 1997), damaged proteins (Erjavec et al. 2007) and less-fit mitochondria (Vevea et al. 2014), which are retained within the mother cell.

ERCs are formed as a result of genome instability (Ganley et al. 2009; Saka et al. 2013), their accumulation increases with age and leads to a shorter life span (Sinclair and Guarente 1997). It has been proposed that high abundance of ERCs limits lifespan by overloading the replication or transcription machinery of the cell thus leading to an inability to replicate or transcribe DNA (Sinclair and Guarente 1997).

An asymmetrically segregated senescence factor are damaged protein aggregates (Janssens and Veenhoff 2016) like carbonylated proteins. These molecules are asymmetrically distributed between mother and daughter during cytokinesis (Aguilaniu et al. 2003). One cause of this uneven segregation is the presence of the cytoskeleton, which is required for the formation of a diffusion barrier at the bud neck (Erjavec et al. 2007). The flow through the bud neck is also restricted by the presence of septin proteins, which are responsible for the asymmetric distribution of some ER-PM contact-site proteins like Ist2 (Sugiyama and Tanaka 2019) where the old molecules remain in the mother while the daughter gets newly synthesized Ist2. Nevertheless, physical impediments are not the only explanation for asymmetrical segregation.

It has been shown pre-existing plasma membrane-associated transporters of the multidrug resistance protein family (MDR) remain associated to the cell cortex of the mother and new proteins are synthesized and deposited in the daughter cell (Eldakak et al. 2010). This being also true for septin mutant *cdc12*, concluding that in the case of MDR proteins the tight association to the cortex prevents the segregation of pre-existing molecules into the daughter (Eldakak et al. 2010).

A known protein known to exhibit asymmetric segregation in budding yeast is the plasma proton pump (Pma1), which molecules are segregated based on protein age; the old protein pool segregates to the mother cell while the newly synthesized protein pool remains in the daughter cell (Khmelinskii et al. 2016).

In these cases of functionality-dependent segregation, the age of the protein correlates with damage and with replicative aging. During my project, I will focus on this type of asymmetric segregation, which correlates with senescence.

### **Tandem fluorescent proteins timers (tFTs) as a red-out of protein age**

Fluorescent proteins (FP) have allowed us to characterize proteins of interest by indicating to us their cellular localization, their accumulation or decrease upon different conditions (Chudakov et al. 2010). Nevertheless, coupled techniques that allowed the simultaneous study of protein turnover and subcellular trafficking presented different constraints that did not allow for their use in living cells (Khmelinskii et al. 2012).

Those barriers were overcome with the use of tandem fluorescent protein timers (tFTs), which are a fusion of two single-color fluorescent proteins that have very different maturation kinetics and very separable emission spectra (Khmelinskii et al. 2012; Khmelinskii and Knop 2014). After translation, each fluorescent protein within a tFT matures with different kinetics, thus the ratio of the two different fluorescent intensities represents a readout for protein age (Khmelinskii et al. 2012).

In order to design a tFT useful for each research purpose, one could exploit the great diversity of single-color fluorescent proteins. In order to be able to study wider molecular-times, one would try to fuse two fluorescent proteins with very different kinetics, one carrying a fast-maturing fluorophore and one with a slow-measuring fluorophore. In order to increase the detection sensitivity of the tFT, the fast-maturing FP should be as fast-maturing as possible while the slow-maturing FP should be as slow as the process that wants to be studied (Khmelinskii and Knop 2014).

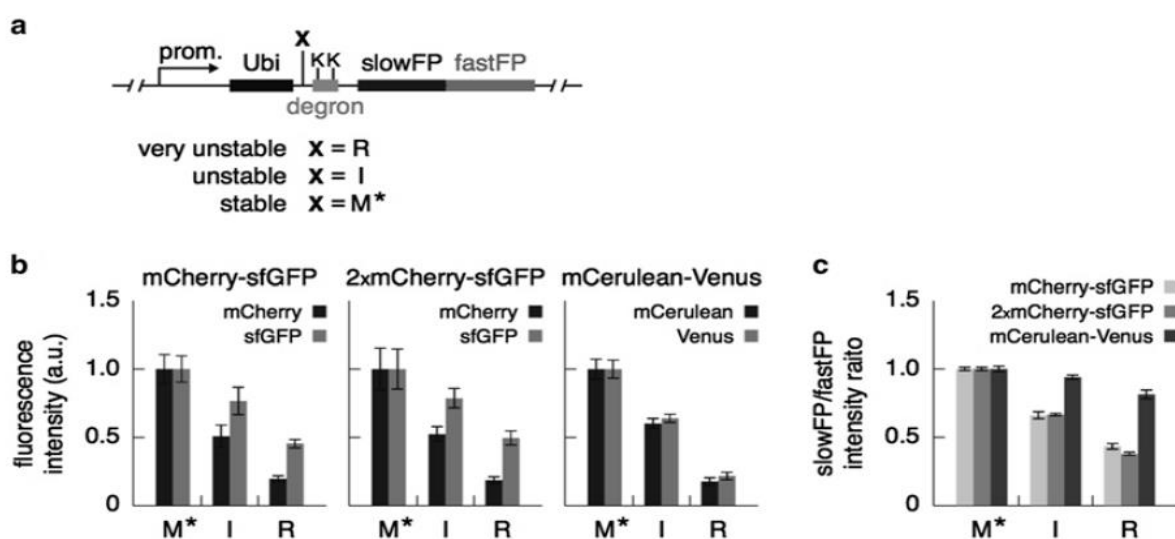
Another important characteristic of the FPs used within a tFT, is their monomeric state of the FP. This state minimizes different artifacts that occur when fusion proteins are expressed at different levels than the endogenous ones (Shemiakina et al. 2012).

Through random mutagenesis on a sequence based on mCherry and multiple RFPs, mScarlet, mScarlet-H and mScarlet-I were created. mScarlet is a monomeric red FP that outperforms other common red FPs as fusion tag, exhibiting a remarkable brightness in mammalian cells (Bindels et al. 2017), same is true for its two variants with a single amino acid substitution mScarlet-H and mScarlet-I.

In budding yeast mNeonGreen and mScarlet-I were reported to be three times brighter (Meurer et al. 2018) than other FPs used in previous genome-wide libraries like: GFP endogenously used as a tag at the C-terminus (Huh et al. 2003), mCherry-sfGFP endogenously used as a tag at the N-terminus (Khmelniskii et al. 2014) and endogenously tagged mCherry and GFP at the N-terminus (Yofe et al. 2016).

Different sets of slowFP-fastFP fusions have been tested for their usefulness as tFTs. One approach to test this, is to evaluate the ability of a potential tFT to report on protein degradation by using synthetic proteins with different half-lives (Ubi-X-tFT) whose degradation rate depends on a single amino acid (X) (N-degrons) (Khmelniskii and Knop 2014; Varshavsky 2011; Fung, Blöcher-Juárez, and Khmelniskii 2022). After translation, cleavage of the ubiquitin moiety by deubiquitinating enzymes, exposes a new N-terminus starting with the X residue present, which determines the stability of the given X-tFT fusion (Bachmair, Finley, and Varshavsky 1986; Bachmair and Varshavsky 1989; Varshavsky 2019). Accordingly, if the red/green fluorescence ratio of yeast strains expressing different Ubi-X-tFT constructs increases with the increasing stability of the X-tFT fusions, the given tFT behavior is proven, see Figure A (Fung, Blöcher-Juárez, and Khmelniskii 2022; Khmelniskii and Knop 2014).

When designing a tFT, the relative brightness of each of the FPs within the tFT is also important, specially when the tFT will be used for imaging purposes (Fung, Blöcher-Juárez, and Khmelniskii 2022; Khmelniskii and Knop 2014).



**Figure A.** Experimental validation of tFT behavior of different slowFP-fastFP fusions. **a.** N-degron constructs encoding a ubiquitin moiety, which after co-translational cleavage exposes the X-residue at the N-terminus. Constructs with unstable X-

residues at the N-terminus are targeted for proteasomal degradation. **b.** Fluorescence intensity of different slowFP and fastFP **c.** slowFP/fastFP fluorescence intensity ratios of yeast colonies expressing N-degron constructs followed by the indicated tFTs (Khmelninskii and Knop 2014).

In different model organisms, different combinations of FPs have been successfully used as tFT, such is the case of mCherry-sfGFP, 2xmCherry-sfGFP and in yeast (Khmelninskii and Knop 2014; Khmelninskii et al. 2012), tagged at the carboxy-terminus RFP-sfGFP in zebrafish (Donà et al. 2013) tagging protein of interests (POI) at the N-terminus with mCherry-sfGFP, tdTomato-sfGFP, mKate2-sfGFP, RFP-sfGFP, fm-mCherry-sfGP and td-fm-mCherry-sfGFP in *Drosophila* embryos (Durrieu et al. 2018), tagging POIs at the C-terminus with mCherry-sfGFP in plants (*Arabidopsis thaliana* and *Nicotina benthamiana*) (Zhang et al. 2019)

Monomeric variants of mCherry (mCh) and of the superfolder green fluorescent protein (sfGFP) have demonstrated to be a good combination within a tFT due to their different maturation kinetics. mCherry displays a maturation half time of 40 min (Merzlyak et al. 2007) while sfGFP fluoresces in minutes after synthesis (Pédélec et al. 2006). When a protein of interest is tagged with mCherry-sfGFP, shortly after synthesis a pool of this fusion should emit mainly green fluorescence, and gradually the red fluorescence emission will increase as mCherry's fluorophore matures. This gradual time-dependent change in the ratio of red to green fluorescence, provides a valuable estimate of the protein age of a pool of molecules (Khmelninskii et al. 2012, 2016).

Fusing a tFT, like mCh-sfGFP, to a protein of interest can be a useful approach to study protein inheritance in model organisms with asymmetric division like *Saccharomyces cerevisiae*. During mitosis, proteins with age-dependent subcellular location can be tracked using a tFT to assess if protein trafficking between mother and daughter compartment is protein-age dependent. In the case of budding yeast, during divisions, the preexisting spindle pole body (SPB) segregates to the daughter cell while newly synthesized SPB stays in the mother cell (Pereira et al. 2001). In a previous study, cells expressing the SPB component Spc42 tagged with a tFT (Spc42-mCh-sfGFP) exhibited a higher mCh/sfGFP ratio in the bud (Rb) than in the mother (Rm) (Khmelninskii et al. 2012). Confirming that the Rb/Rm ratio is a good indicator for protein-age difference in between bud and daughter compartments during mitosis. Therefore, the average age of a protein pool is determined by their production rate, the maturation kinetics of the FPs within the tFT and by the degradation kinetics of the proteins within the fusion (Khmelninskii and Knop 2014).

In budding yeast some commonly used greenFPs, especially sfGFP, can partially resist proteasomal degradation, which results in cytosolic accumulation of cytosolic greenFP upon protein degradation of the PO-tFT fusions (Khmelninskii et al. 2016). This is important to consider when designing a tFT and when interpreting the results of the studies carried out with it. This since the increase in greenFP signal in the cytosol could be misleading for studying protein localization, protein dynamics upon different conditions, and would also provide an overestimation of the amount of newly synthesized protein fusions (Khmelninskii et al. 2016).

For a protein to be targeted by the proteasome, a degradation signal (degron) and an initiation site are required (Schrader, Harstad, and Matouschek 2009). Consequently, the degradation of a tFT fusion will start from the site where the POI is placed,

therefore the product after proteasomal degradation will depend on the order of the FPs within a tFT and its position relative to the POI (C or N-terminus tag) (Khmelniskii et al. 2016).

### **Protein-inheritance pattern in budding yeast**

On a smaller level, there have been studies that tried to characterize the segregation of specific organelles in between mother and daughter cells in budding yeast. Thanks to these, it is known that symmetric divisions occur more often as the mother cell ages (B. K. Kennedy, Austriaco, and Guarente 1994) and one of the characteristics of this symmetry is the loss of mitochondria membrane potential and the improper segregation of active mitochondria to the daughter cells (Lai et al. 2002; McFaline-Figueroa et al. 2011).

During asymmetric divisions, the retention of lower-functioning mitochondria in the mother is associated with age-associated fitness decline (McFaline-Figueroa et al. 2011). Nevertheless, an advantage of an asymmetric segregation of certain components, like damaged proteins and less-fit mitochondria, would be the guarantee of generating damage-free offspring able to preserve the cell line.

In a recent study, genome-wide quantitative proteomics was used to characterize the protein segregation of old and newly synthesized protein between mother and bud (Sugiyama and Tanaka 2019). By using the alpha-factor pheromone cell cultures were synchronized in G1 phase and at the same time labeled with Cy5 dye. After the release of the arrest, Cy5-labeled cells were identified as mothers and unlabeled ones as daughters. To distinguish from pre-existing and newly synthesized proteins, they grew the cells using a SILAC approach to mark the newly produced proteins. Protein synthesis and degradation was stopped by fixating the cells with methanol. Using flow cytometry, cells were sorted into two populations: daughter-enriched and mother-enriched. These fractions were then sent to LC-MS/MS to quantify the pre-existing (old) and newly synthesized proteins. In this study, a total of 2039 proteins were detected with their mass spectrometry-based approach.

From those, 56 were classified as segregating asymmetrically, where old protein segregated to the mother. From these proteins with asymmetric index, 20 were previously reported to have that segregation pattern, therefore they only discovered 36 new asymmetrically segregated proteins. They also identified 1829 symmetrically segregated proteins and 152 could not be classified in a category due to inconsistency of results between their two replicates.

Comparing this protein segregation dataset (Sugiyama and Tanaka 2019) to a dataset where the abundance of 5748 proteins among 21 quantitative studies was unified under the same unit (Ho, Baryshnikova, and Brown 2018), I realized the proteins identified by the mass spectrometry approach were mainly proteins with high abundance (Figure S1).

## AIM OF THIS STUDY

Despite the significant discoveries made in yeast as a model for cellular aging, the quality control mechanisms maintaining asymmetric inheritance of cellular components are still unknown. In this project, I want to know which proteins from *Saccharomyces cerevisiae* are segregated based on protein age and how those segregation patterns are maintained. As well, I would like to check which of the asymmetrically segregated proteins are acting as senescence factors.

For this aim, I decided to create a library where all ORFs are tagged at the C-terminus with the bright mScarlet-I-mNeonGreen tFT. Performing fluorescence microscopy of this library could allow the detection of more proteins which normally exhibit a low abundance and that might be invisible under different approaches.

## Objectives

- 1) To determine the genome-wide protein segregation pattern in budding yeast
  - To identify, using fluorescence microscopy, which proteins are asymmetrically segregated based on protein-age between mother and daughter cell
  - Classify the type of asymmetry of each protein based on protein-age
  - To investigate which are the known mechanisms underlying that asymmetric segregation for each protein
  
- 2) To investigate what is the relation between protein-age based asymmetric inheritance and replicative aging
  - To identify which of the asymmetric phenotypes identified are playing a role in the replicative aging of budding yeast



## **RESULTS**

### **Selection of FP for tFTs and generation of mScarlet and mScarlet-H**

Based on previous reports in mammalian cells of mScarlet (mS), mScarlet-H (mSH) and mScarlet-I (mSI) red FPs being brighter and slower-maturing than mCherry (mCh) (Bindels et al. 2017) and mScarlet-I being brighter than mCherry in budding yeast (Meurer et al. 2018) we decided to test these three red FP as part of tFTs, as well as the well known mCherry FP.

For fast-maturing green FP for tFTs, I decided to use the well-characterized sfGFP (Khmelinskii et al. 2012, 2016). Additionally, I decided to test mNeonGreen as a fast-maturing FP since it has been reported to be brighter than common green FPs in yeast (Meurer et al. 2018).

In total, I selected two fast-maturing green FPs and four slow-maturing red FPs to be combined and evaluated for their behavior as tFTs (Table S3).

Since the sfGFP-mCherry fusion has been proven before to behave as a tFT (Khmelinskii et al. 2012, 2016), I decided to include this in my evaluation of other fluorescent fusions as tFTs and use it as a control.

I decided to use the mScarlet-I sequence, already available in the plasmid pYD13 (Meurer et al. 2018) in our laboratory, as the initial sequence to generate the additional two mScarlet variations reported in mammalian cells. I used Quick change PCR site-directed mutagenesis (Table 6) to create the single amino acid substitution I74T (from isoleucine to threonine in the 74th amino acid) generating the mScarlet variant.

For generating the mScarlet-H variant I used a gap repair-cloning approach. I amplified the previously created plasmid carrying the mScarlet variant, and I transformed this into budding yeast together with annealed oligos which were carrying the three amino acid substitution at the position 164 (M164H), to obtain a histidine instead of a methionine. After having a yeast strain carrying a plasmid with the new mScarlet-H variant, I extracted this plasmid and transformed *E. coli* DH10beta electrocompetent cells (Top10 C404052) by electroporation.

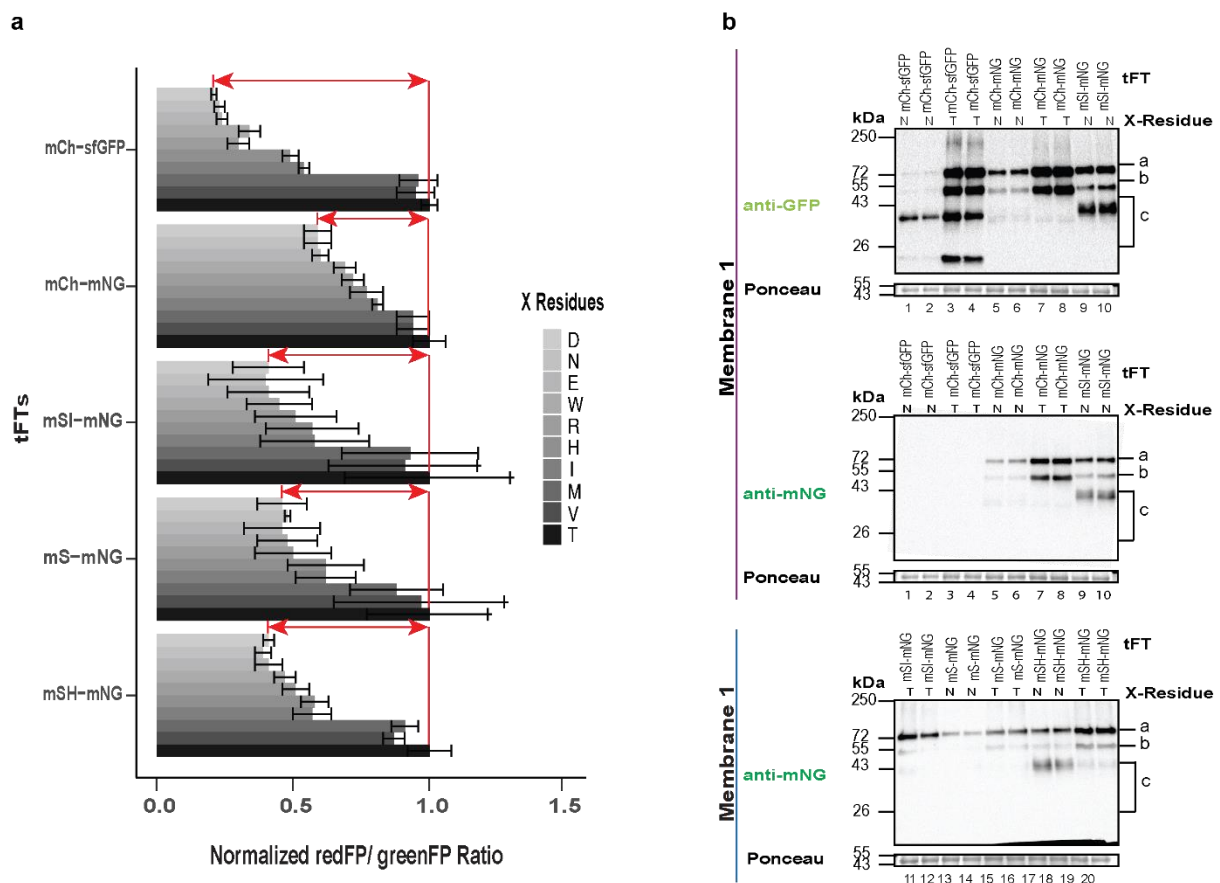
### **Evaluation of different tFTs as reporters for protein degradation**

Since there is evidence that the relative position of some greenFPs within a tFT affects, in a different magnitude, the cytosolic accumulation of fragments produced after proteasomal degradation of the POI-tFT or tFT-POI fusion (Khmelinskii et al. 2016) it is important to characterize different slowFP-fastFP as reporters of protein degradation.

To be on the safe side, I decided to create different tFTs candidates in which the green FP would be placed as distal to the degradation initiation site as possible. Meaning, for C-terminus tagging of any POI our construct would have the order "POI-redFP-greenFP" while for N-terminus tagging the order would be "greenFP-redFP-POI".

In order to evaluate protein degradation using different tFTs, we cloned different slow-maturing red FPs in combination with different fast-maturing green FPs (Table S4) as tFTs in the backbone plasmid pAnB19 (pRS413-GPDpr-Ubi-EcoRV-STOP-eK-mCherry-sfGFP). The backbone plasmid carried a ubiquitin moiety under the GPD promoter at the N-terminus, followed by a stop codon and X-tFT. These elements make it possible to test if our tFTs are good reporters of protein degradation by using the approach described in above (Figure A).

We know that based on the nature of the amino acid residues placed at the N-terminus of a protein, their propensity to be degraded by the proteasome changes (Khmelinskii and Knop 2014; Varshavsky 2011; Fung, Blöcher-Juárez, and Khmelinskii 2022). Therefore, I decided to test ten different amino acid residues as X-residues within our ubiquitin fusion plasmids. The X-residues selected were: asparagine (N), aspartic acid (D), glutamic acid (E), tryptophan (W), arginine (R), histidine (H), isoleucine (I), methionine (M), valine (V) and threonine (T).



**Figure 13.** Characterization of different tFTs as reporters of protein degradation. **a.** The ratio of each Ubi-X-tFT was normalized to the value of threonine-tFT (T-tFT), which was the most stable construct across all tFTs used; this normalization was performed per X-tFT construct. We used a plate reader (SparkControl 20M plate reader with monochromator, TECAN) to measure fluorescence intensity of X-tFT constructs arrayed in a 1536 colony format on agar. The graph shows values for one biological replicate per X-tFT construct. For mCherry-sfGFP constructs 24 technical replicates were measured per biological replicate of each X-tFT, for the rest of the tFTs 16 technical replicates were measured per biological replicate of each X-tFT. Error bar indicated for each X-tFT. **b.** Membrane one was incubated first with anti-GFP (top

image), followed by incubation with anti-mNG (middle image). Membrane two was only incubated with anti-GFP. Two technical replicates per X-tFT were used. Degradation fragments: (a) full length X-tFT, (b) a mCherry<sup>ΔN</sup>-greenFP product of mCherry hydrolysis during cell extract preparation (Gross et al. 2000; Shemiakina et al. 2012) and (c) processed-tFT fragments.

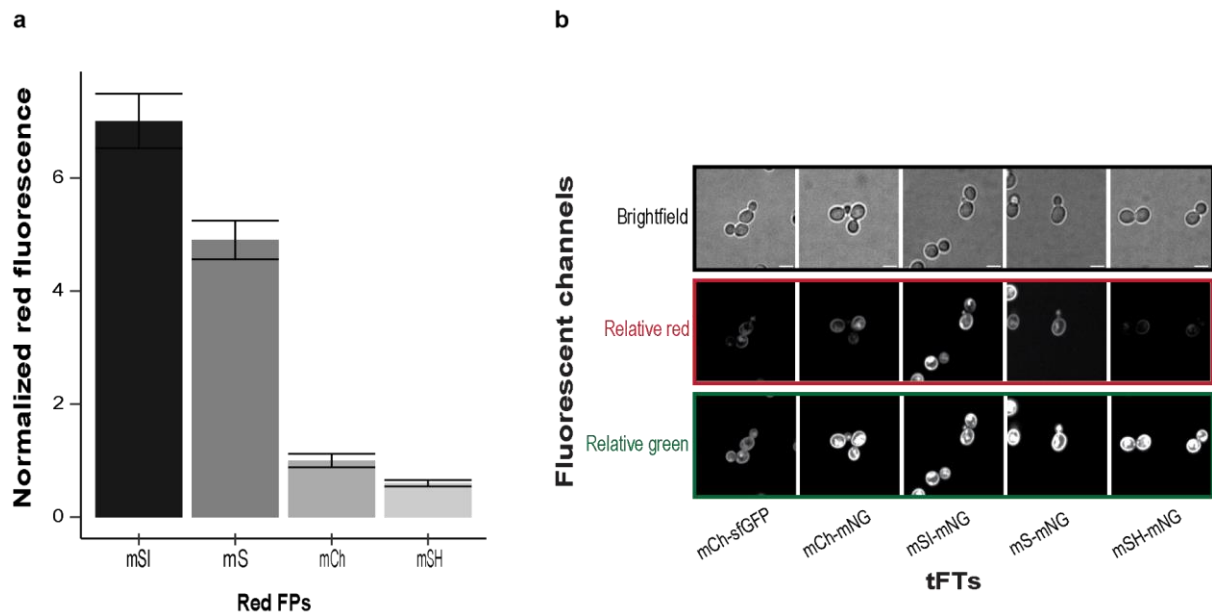
The different Ubi-X-tFT constructs will exhibit a different degradation rate depending on the identity of the single amino acid (X) placed at the N-terminus (Khmelinskii and Knop 2014; Varshavsky 2011; Fung, Blöcher-Juárez, and Khmelinskii 2022). Co-translational cleavage of the ubiquitin moiety will expose the X-residue, determining the stability of each protein fusion according to the N-end rule pathway (Bachmair, Finley, and Varshavsky 1986; Bachmair and Varshavsky 1989; Varshavsky 2019). Thus, the red/green fluorescence ratio of yeast strains expressing different Ubi-X-tFT constructs will directly correlate with their stability (Figure A).

As seen in Figure 13 panel a, the five different tFTs tested can be used as reporters for protein degradation. With all tFTs we were able to distinguish between very unstable constructs N-tFT and very stable constructs T-tFT.

I proceeded to examine the performance of the tFTs as reporters for protein degradation. With the most stable and unstable constructs previously evaluated (Figure 13a) N-tFTs and T-tFTs I performed a western blot experiment (Figure 13b). Using all the tFTs as a C-terminus tag, having the red FPs more proximal to the degradation initiation site. As seen in Figure 13b, all N-tFTs exhibit a dimmer signal compared to the stable constructs T-tFTs, confirming all the tFTs tested are good reporters of protein degradation.

Accumulation of processed tFT fragments (<43 kDa) are observed only for constructs with the mCh-sfGFP tFT (Figure 13b). This could be, as reported before, due to the incomplete proteasomal degradation of sfGFP (Khmelinskii et al. 2016).

In the case of mSI-mNG carrying the most unstable residue N, we observed a larger signal of processed tFT fragments compared to T-mSI-mNG (Figure 13b), same pattern was detected for mSH-mNG. For both N-mS-mNG and T-mS-mNG, there was no detectable signal for processed tFT fragments, which might indicate a different degradation in between mScarlet variants.



**Figure 14.** Relative brightness of the different red FPs tested in tFTs. **a.** Whole colony measurements of 28 technical replicates per T-tFT, grown in the same agar plate, were measured in a plate reader (TECAN). The excitation and emission used were 569 nm 593 nm, respectively. **b.** Representative fluorescence microscopy images of dividing cells expressing the plasma membrane proton pump Pma1 tagged with indicated tFTs at the C-terminus. The fluorescent channels acquired were brightfield (BF), green channel and red channel. Scale bar, 5  $\mu$ m. The panel of images for each fluorescent channel was contrasted based on the settings of mScarlet-I or mNeonGreen, respectively. All fluorescent images were acquired using 800 ms of exposure and 70 % of laser power.

As we wanted to determine the segregation pattern based on fluorescent microscopy, the *in vivo* brightness in yeast of the red FPs was a main parameter for the selection of a red FP. Therefore, I performed whole colony red fluorescent measurements of the T-tFT strains (Figure 14a). For these measurements, in the borders and across the agar plate, I placed a non-fluorescent strain as negative control (yKBJ0052 strain, p413-GDP uncut transformed into BY4741) and as reference I used a strain carrying the construct T-mCherry-mNeonGreen (yKBJ0021). After background correction and normalization to the reference strain, the red fluorescence of each tFT was normalized for protein expression levels using the fluorescence of each green FP. The final red fluorescence of each FP was finally normalized to the red fluorescence of mCherry for comparison.

As we can see in Figure 14a, mScarlet-I was the brightest red FP tested. In order to validate this findings, I performed a confocal fluorescent microscopy of strains where the Pma1 proteins was tagged with each tFT at the C-terminus (Figure 14b). This microscopy shows that mNeonGreen is brighter than sfGFP, as previously reported. Additionally, I confirmed that the brightest of the red FP testes was mScarlet-I.

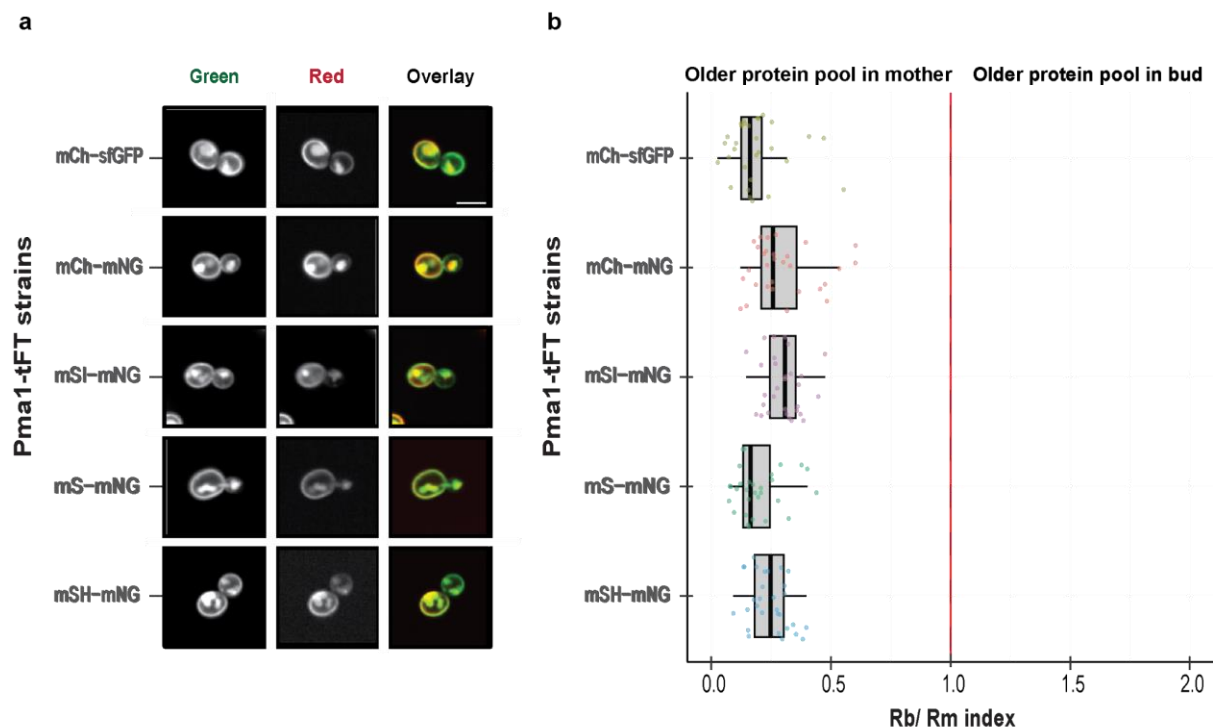
Considering the brightness results, mScarlet-I-mNeonGreen appears to be the most suitable tFT for detecting a broader range of proteins throughout the entire genome.

## Evaluation of different tFTs as reporters for protein protein-age based segregation

In order to determine the genome-wide protein-age based segregation in budding yeast, we needed to construct a genome-wide library where all available ORFs within the budding yeast genome were tagged with a given tFT. Therefore, it was important to select very bright FPs to allow the visualization of as many low abundant proteins as possible. Thus, we decided to evaluate the previously mentioned tFTs also as reporters for protein-age based segregation validated using confocal fluorescence microscopy.

In order to evaluate the tFTs as reporters of protein age, I needed to use them as a tag for a protein known to segregate between mother and daughter cells based on protein age. Therefore I decided to use the plasma membrane protein Pma1 (Khmelinskii et al. 2016), for which it is known that the preexisting (old) protein pool localizes at the mother compartment while the newly synthesized protein goes to the bud side.

I generated five different yeast strains where Pma1 was tagged with the five different tFTs at the C-terminus. The yeast strains were grown to exponential phase to guarantee that the majority of the cells were dividing during the microscopy. I performed confocal fluorescence microscopy of unsynchronized yeast cell cultures and confirmed that mScarlet-I and mNeonGreen are the brightest red and green FPs tested (Figure 14).



**Figure 15.** Characterization of tFTs as reporters for protein age. **a.** Representative fluorescence microscopy images of dividing cells expressing the plasma membrane proton pump Pma1 tagged with the different tFTs tested. Mother (m) and bud (b) compartments are indicated. All images are auto contrasted. **b.** Quantification of Pma1 localized to the plasma membrane in mother (Rm) and bud

(Rb) compartments of dividing cells ( $n = 30$  cells for each tFT). Correction for auto-fluorescence was performed using the strain yEI0061. Centerlines mark the medians, box limits indicate the 25th and 75th percentiles, and whiskers extend to fifth and 95<sup>th</sup> percentiles. All fluorescent images were acquired using 800 ms of exposure and 70 % of laser power. Scale bar, 5  $\mu\text{m}$ .

From the same microscopy images, I manually segmented 30 dividing pairs per Pma1-tFT and quantified their red and green fluorescence at the mother (m) and at the bud (b) compartments (Figure 14a). Pma1 is already known to segregate asymmetrically based on protein age, where the older protein pool localizes to the mother and the newly synthesized protein mainly localized to the bud (Khmelinskii et al. 2016).

Therefore, to characterize the tFTs as reporters for protein age, I calculated the Rb/Rm ratio for each Pma1-tFT strain. In the case of Pma1, the red/ green ratio at the bud (Rb) is expected to be less than 1, since there is a larger amount of green fluorescence (newly synthesized protein) than red fluorescence in the bud. While for the red/ green ratio at the mother compartment (Rm) the opposite is expected. Pma1, or any other protein that segregates asymmetrically with the older protein pool at the mother, under normal conditions has to have a  $Rb/ Rm < 1$ . Symmetrically segregated proteins will have a Rb/ Rm around 1 and proteins in which the older protein pool localizes to the bud, the Rb/ Rm ratio should be larger than 1.

As seen in Figure 14a, we are able to detect the asymmetric segregation of Pma1 using all tFTs tested. For all tFTs the cells exhibit a lower red/ green intensity ratio (R) at the plasma membrane in the bud (Rb) than in the mother cell (Rm).

As the Rb/Rm ratio was not that different in between tFTs, the maturation of sfGFP and mNeonGreen might be similarly fast for both. As well, the slow maturation of the different red FPs does not vary greatly in between them.

After the various evaluation of different tFTs as reporters for protein degradation and of protein age, I chose mScarlet-I-mNeonGreen as the tFT to study of genome-wide segregation based on protein age in yeast.

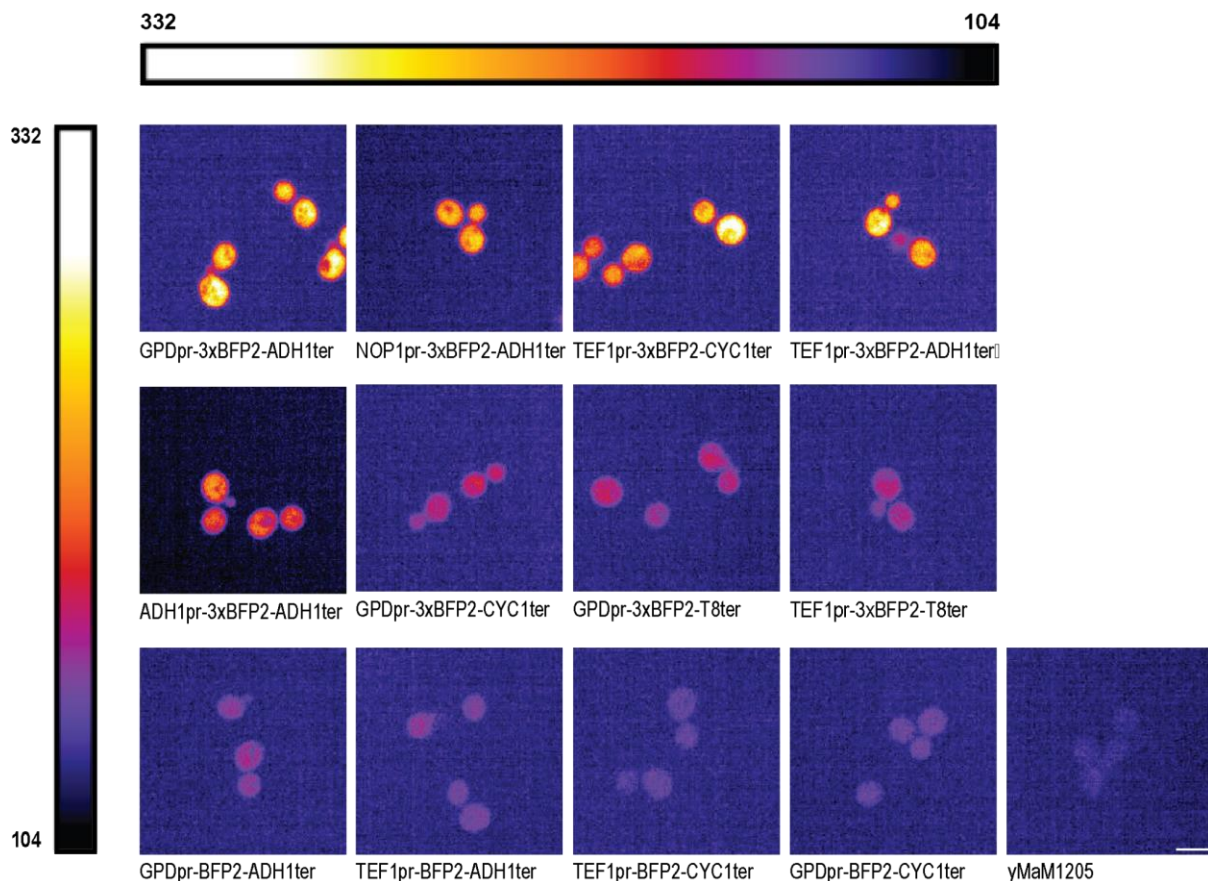
### **Generation of a donor strain with markers for image segmentation**

After selecting mScarlet-I-mNeonGreen as a tFT to study protein segregation in the yeast genome, I needed to construct a suitable donor plasmid carrying the tFT of interest. For this, I also decided to construct a donor plasmid of the type II, with which we can select for strains carrying the tag of interest via reconstitution of the hygromycin resistance marker (hph) (Meurer et al. 2018). The donor plasmid type II I constructed was named pKBJ011-1.

For the high-throughput tagging of the yeast ORFs, the donor plasmid needed to be transformed into a yeast donor strain of MAT $\alpha$  mating type, in order to cross it to the C-SWAT strains which are MAT $\alpha$ .

For the aim of determining the genome-wide protein segregation in yeast, I wanted to use fluorescent microscopy for the phenotyping of the final ORF-tFT-Hph library. Therefore, it was needed to include some markers within the donor strain that would help us to segment the cells (mother and bud compartments) in the final genome-wide

microscopy. For better background correction, I also wanted to grow negative control cells (no tFT present) together with sample cells (expressing tFT). Hence, I needed to distinguish both cell populations during the segmentation.



**Figure 16.** Relative blue fluorescence of yeast strains with different cytosolic markers. The control strain, yMaM1205, corresponds to the background strain used to generate the other strains. The blue fluorescence intensity of all images was normalized to the brightest strain (GPDpr-3xBFP2-ADH1ter). Scale bar, 5  $\mu$ m.

I decided to use the same blue fluorescent protein BFP2 as a cytosolic marker for cell segmentation of control and sample cells. In order to differentiate the two populations per image, I needed the cytosolic markers to have different intensity. Reason why I decided to use a single copy or a triple copy of the BFP2. I also generated cytosolic markers with a range of blue intensities by combining several promoters (Pr) and terminators (Ter). Finally, after constructing different plasmids carrying the different combinations of Pr-1x/3xBFP2-Ter, I transformed them into yeast yMaM1205 competent cells (Figure 16).

I performed confocal fluorescent microscopy of the different cytosolic strains and concluded that the brightest cytosolic marker was the GPDpr-3xBFP2-ADH1ter, while the strain with the lowest cytosolic blue intensity was the one with the cytosolic marker GPDpr-BFP2-CYC1ter (Figure 16). Since I needed an additional brighter bud neck marker to identify dividing-pairs, I decided to use the dimmest cytosolic marker for sample cells (GPDpr-BFP2-CYC1ter) and the brightest one (GPDpr-3xBFP2-ADH1ter) for negative control cells.

For distinguishing dividing cells from close-by sample cells, I selected the non-essential septin Shs1 as a good candidate to tag with a FP. The Shs1 is a component of the septin ring, localizing to the mother-bud neck and required during the cytokinesis. Upon Shs1 deletion, cytokinesis is defective and yeast chains are formed across cell cycles (Iwase et al. 2007).

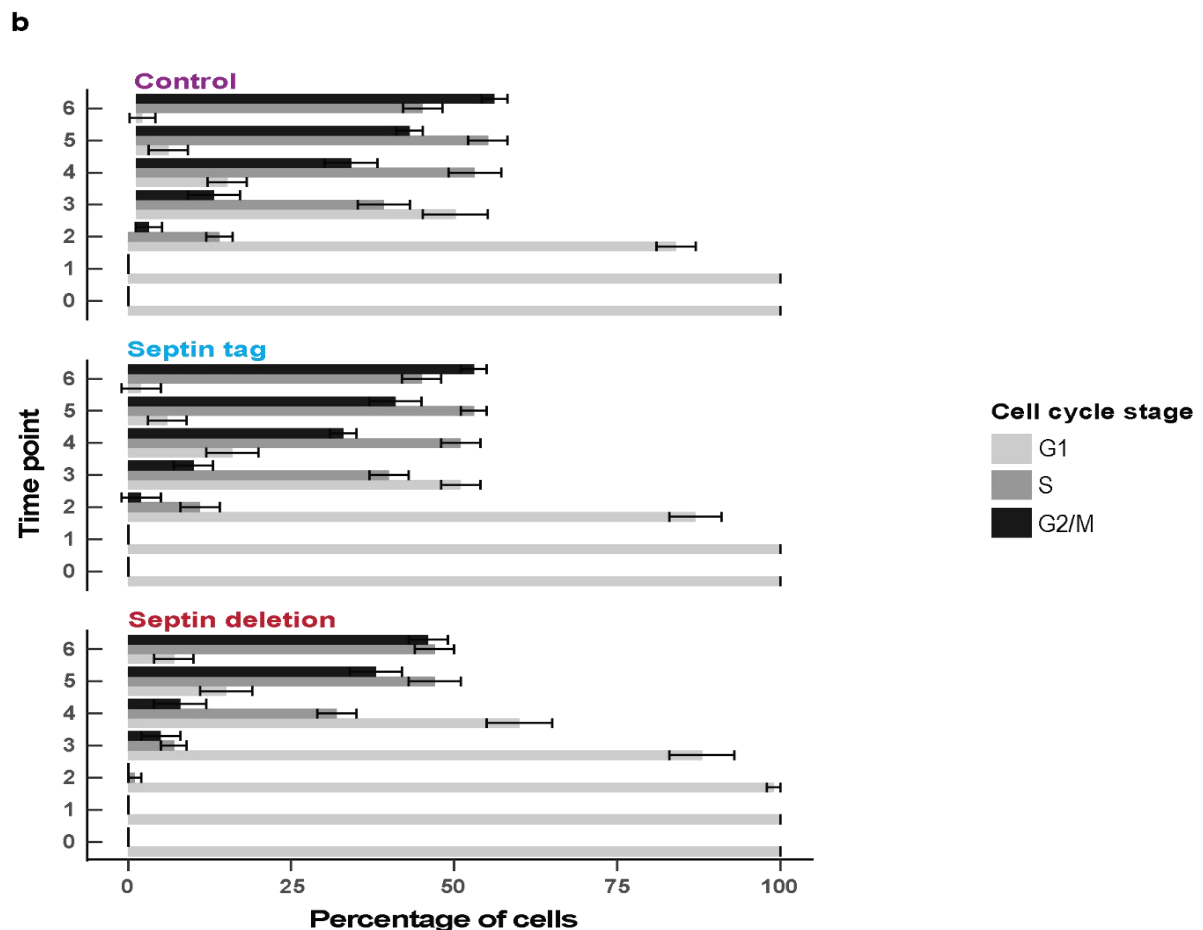
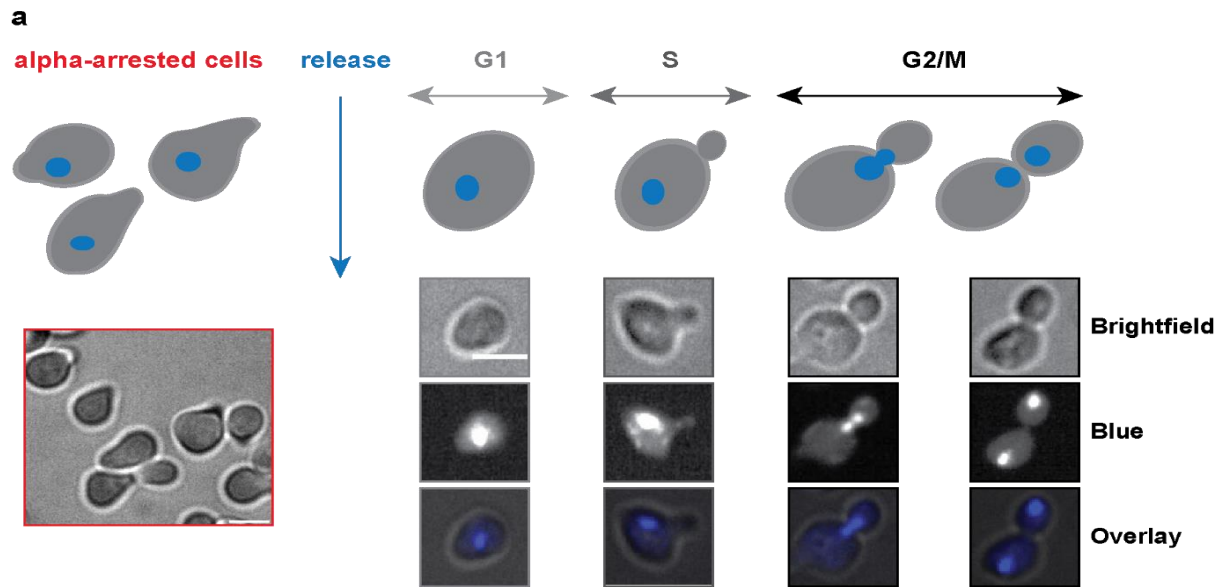
Using the selected cytosolic markers for sample and negative control cells, we tagged the Shs1 septin with the 3xBFP2-ADH1ter cassette. Together with our image-analysis collaborators, we evaluated the separability of sample and negative control cells based on their blue intensity only (not shown); we confirmed that we accurately distinguished between sample and negative control cells, negative control cells exhibiting a cytosolic blue intensity around three times higher than the sample cells. For this separability test we used sample cells where the Pma1 was tagged at the C-terminus with mScarlet-I-mNeonGreen, therefore we confirmed that the cell population identified as sample cells based on the dim blue intensity was correct by checking the Pma1 signal under the red and green channels.

To confirm that tagging this septin does not affect cell division in yeast, I decided to use the genetic background BY4741 (MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 0) and created three strains: one expressing the previously selected cytosolic marker for sample cells (GPDpr-BFP2-CYC1ter), one strain expressing the same cytosolic marker in addition to Shs1 tagged with the 3xBFP2-ADH1ter cassette and one strain where I replaced the Shs1 by the hphNT1 cassette.

These three strains were intentionally mating type MatA to use the alpha-factor hormone for synchronizing the cultures in G1 phase. I decided to synchronize the three different strains to better characterize any cell division difference in between them. After the release of synchronization, I collected cells every 15 minutes and imaged each time point using confocal fluorescence microscopy. Using a mechanical counter I quantified the number of cells per strain in G1, S or G2/M phase (Figure 17a) My cytological classification was based on the previously published (Rosebrock 2017).

Using the Shs1-3xBFP2-ADH1ter as a bud neck marker seems to not affect the cell-cycle progression of synchronized control and septin-tagged cells (Figure 17b). As expected, the Septin deletion strain (Shs1 $\Delta$ ) has a slower progression through the cell cycle since the deletion of this septin affects cytokinesis (Iwase et al. 2007).

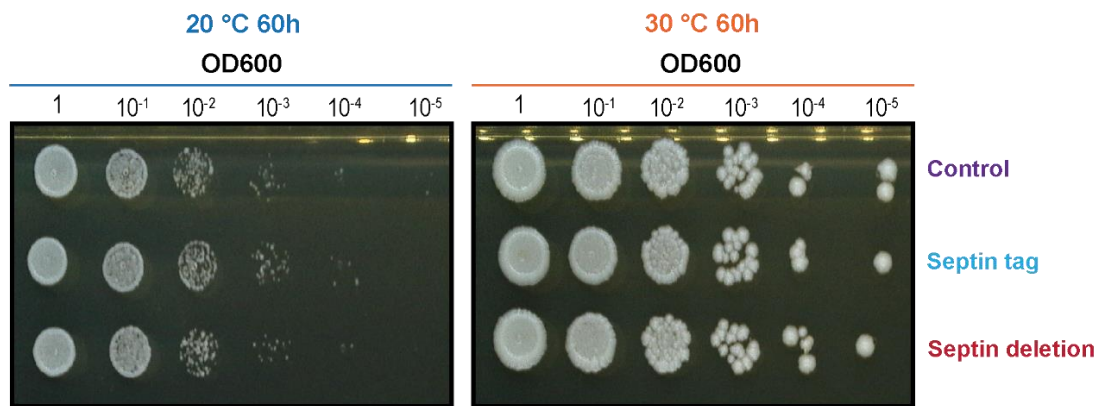




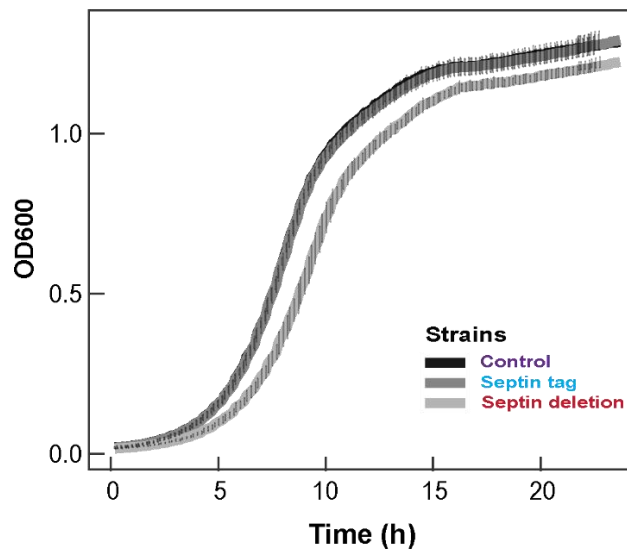
**Figure 17.** Cell cycle progression of strains with a cytosolic marker and septin tags. **a.** Phenotype of  $\alpha$ -factor arrested yeast cells and cytological classification after the release. Confocal fluorescence microscopy of Control cells (cytosolic GPDpr-BFP2-CYC1ter marker) shown as an example. Brightfield (BF) and Blue channels were acquired with 500ms of exposure and 70% laser power. Scale bar, 5  $\mu$ m. **b.** Percentage of cells at different cell cycle stages, across different time points after  $\alpha$ -factor release. Control strain (cytosolic GPDpr-BFP2-CYC1ter marker), Septin tag

strain (cytosolic GPDpr-BFP2-CYC1ter marker and Shs1-3xBFP2-ADH1ter) and Septin deletion strain (Shs1 $\Delta$ ::hphNT1) were grown at the same time using the same conditions. Time points were collected every 15 minutes after the release, t0 being the only time point taken before the arrest.

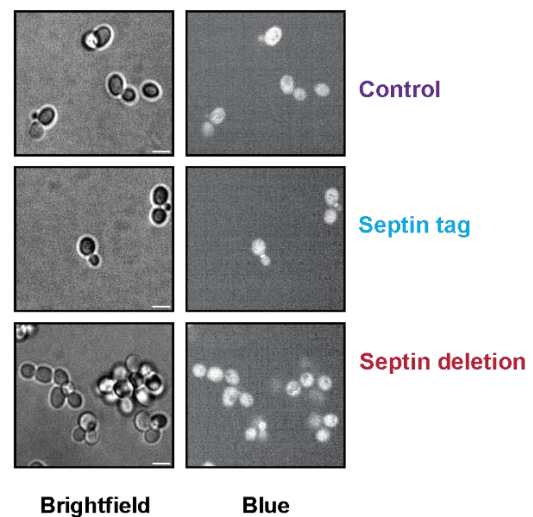
**a**



**b**



**c**



**Figure 18.** Spotting assay of strains with S288C background. Representative spotting assay, from four technical plate replicates. **a.** Serial dilutions of different strains were spotted on YPD+Nat agar plates and grown for 60h. Three technical replicates were grown per temperature treatment, only one representative plate of each treatment is shown. **b.** Growth curve of different strains grown for 24h in YPD liquid media at 30 °C and 220 rpm. **c.** Confocal fluorescence microscopy of different strains. Brightfield (BF) and blue channels were acquired with 500ms of exposure and 70% laser power. Scale bar, 5  $\mu$ m.

It has been reported that the deletion of Shs1 causes cold-sensitivity in the W303 genetic background (Iwase et al. 2007). Therefore, I decided to test if the previous strains (Figure 17) presented cold-sensitivity. To test the growth upon cold temperature, I performed a spotting assay where I tested five different dilutions of each strain (Figure 18a). The spotting assay revealed no clear difference between the strains, neither at 20 °C nor at 30 °C.

I further assessed the growth in liquid media of those three strains (Figure 18b), in this case it was more obvious that the Control strain (cytosolic GPDpr-BFP2-CYC1ter marker) and the Septin tag strain (cytosolic GPDpr-BFP2-CYC1ter marker and Shs1-3xBFP2-ADH1ter) have a more similar growth. I also observed how the Septin deletion strain (Shs1 $\Delta$ ::hphNT1) seems to have a slower growth compared to the other strains.

Finally, when performing confocal fluorescence microscopy of unsynchronized cultures in exponential phase for the above mentioned strains (Figure 18c), I observed formation of chains only for the Septin deletion strain. While the Control strain and the Septin strain have a similar phenotype.

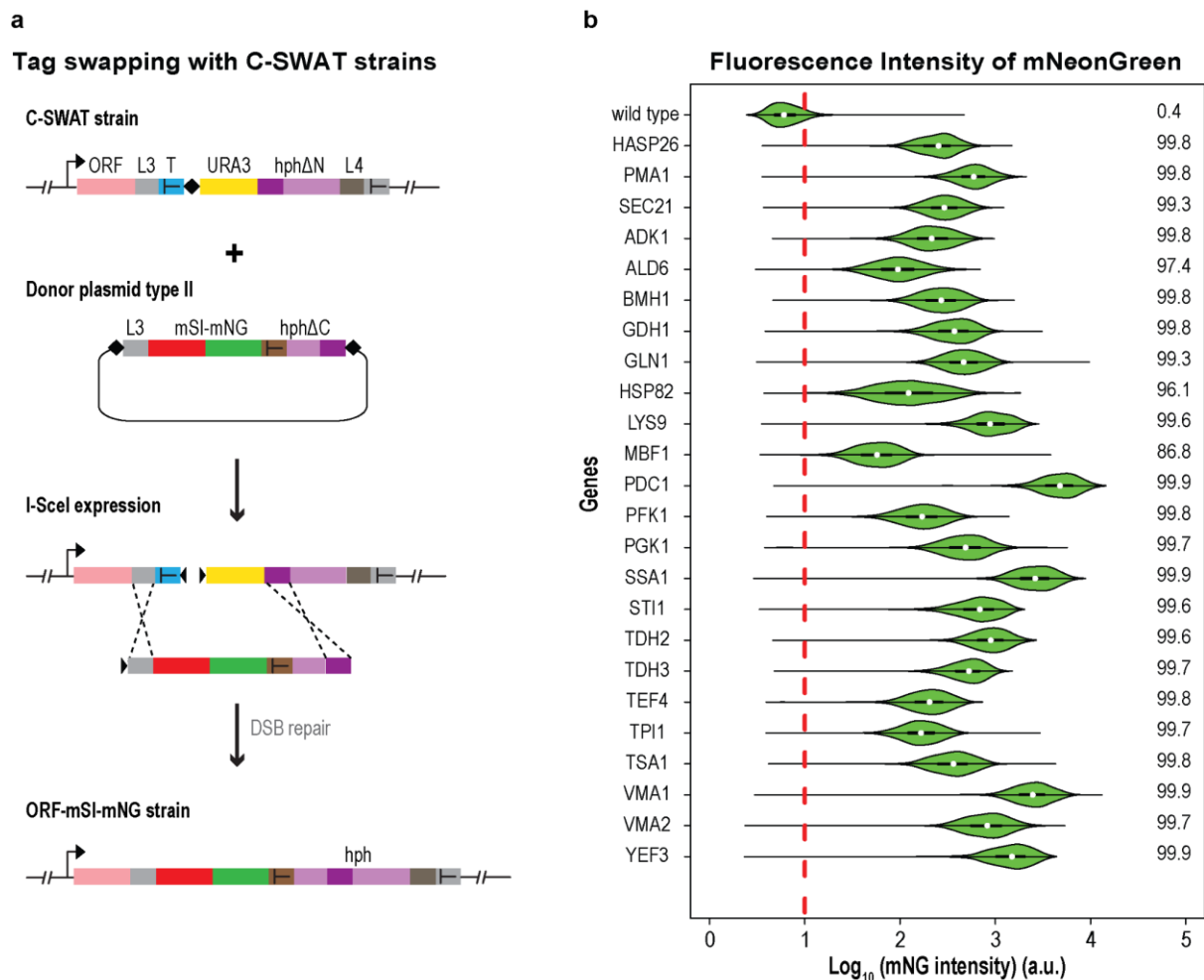
After concluding that using the Shs1-3xBFP2-ADH1ter as a marker for bud and mother identification does not affect cell division, we reevaluated the separability of negative control cells expressing a bright blue cytosolic marker (GPDpr-3xBFP2-ADH1ter) and samples cells expressing a dim blue cytosolic marker (GPDpr-BFP2-CYC1ter) and now also expressing a bright blue septin tag (Shs1-3xBFP2-ADH1ter).

The separability test confirmed that even with presence of the bright septin tag, the sample cells are still separable from the negative control cells, based on their blue intensity only.

### **Construction of genomewide C-tagged libraries and their validation**

After choosing cytosolic and septin markers for the cell segmentation, I proceeded to create the final donor strain by transforming competent cells of the genetic background yMaM1205 with the dim blue cytosolic marker and the bright septin tag. After this strain was created, the donor plasmid carrying the tFT (mScarlet-I-mNeonGreen) was also transformed into this strain.

In order to use tFTs as a tool to determine genome-wide protein segregation, the SWAT libraries were used to tag all available ORFs from the yeast genome with the desired tFT. The C and N-SWAT libraries (Yofe et al. 2016; Meurer et al. 2018) are composed of strains in which each ORF is tagged with an acceptor module. The acceptor module at the C or N-terminus the high-throughput swapping of this module with the tag of interest. Thanks to the design of the SWAT libraries, donor plasmids with different selection strategies can be designed. In this project, I used mScarlet-I-mNeonGreen to tag all ORFs at the C-terminus using a selection reconstitution tagging approach (Figure 19a).



**Figure 19.** Construction of a non-endogenous ORF-mScarlet-I-mNeonGreen library. **a.** By crossing the already existing C-SWAT library to a donor plasmid carrying the tag of interest (tFT) and inducing the expression of I-SceI, double-strand breaks (DSBs) are generated at the positions indicated in the acceptor module of the C-SWAT strains and at the donor plasmid. By homologous recombination, DSBs are repaired leading to substitution of the acceptor module with a tag. New strains carrying the tFT of interest can be selected by the lack of the URA3 present in the starting SWAT acceptor module.

**b.** Comparison of the tagging efficiency obtained with the donor plasmid type II (pKBJ011) carrying the mScarlet-I-mNeonGreen tFT. The high-throughput protein tagging was performed using the SGA for the whole C-SWAT library. Protein tagging efficiency was tested for the 24 indicated genes only. The percentage of cells with red and green fluorescence above background per protein is shown; around 10000 cells were measured per strain using flow cytometry.

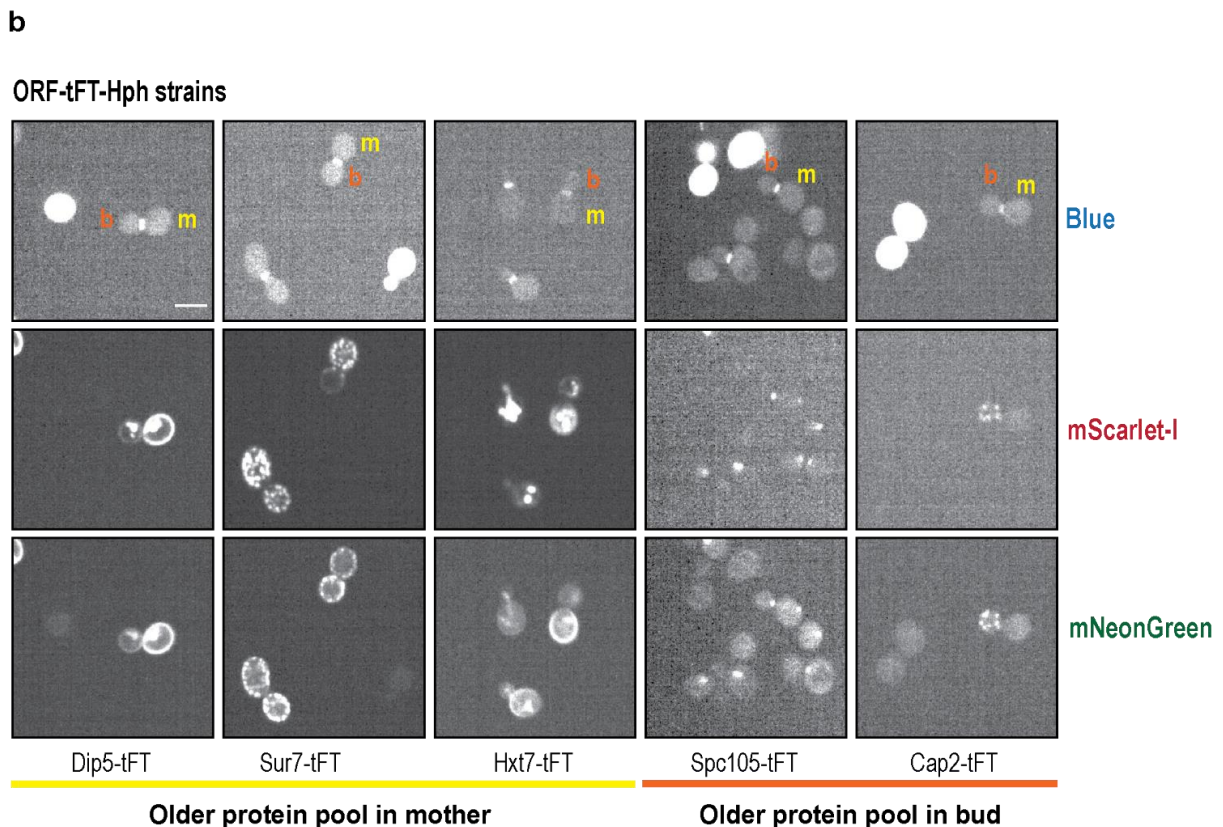
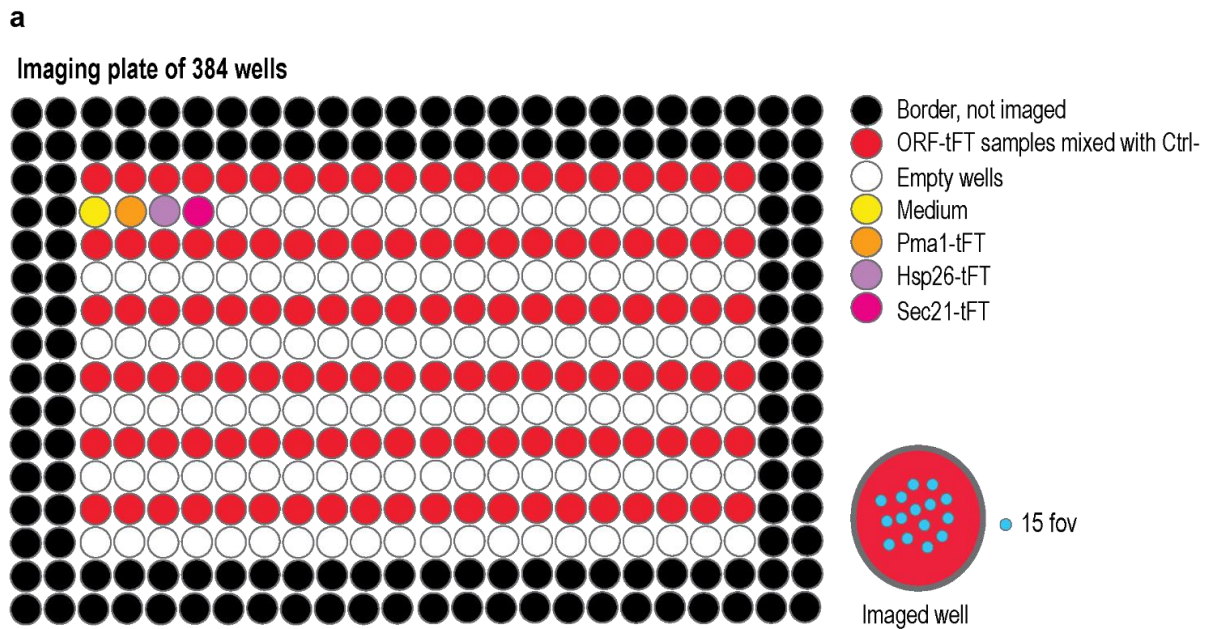
The C-SWAT library contains 5,661 ORFs (Meurer et al. 2018), therefore, the new library I constructed contains 5,661 ORFs (around 93 % of the yeast ORFs) tagged at the C-terminus by mScarlet-I-mNeonGreen-Hyg.

For the new library, I picked 24 strains (ORF-mScarlet-I-mNeonGreen) that are known to have a high protein abundance and which have been used in previous studies as reference to test for tag efficiency (Meurer et al. 2018) using a flow cytometer, and

confirmed that the tagging efficiency across those strains was close to 100 % per strain (Figure 19b). This tagging efficiency will be extrapolated to the rest of the library.

After confirming the successful tagging of the new library, I proceeded to perform high-throughput microscopy of each strain, using the Opera Phenix microscope. To image the 5,561 strains, I used 48 imaging plates per technical replicate, and I imaged in total two technical replicates.

In each plate, I imaged 120 ORF-tFT strains mixed with the negative control strain. I also included in each plate the next controls: a well of medium, a well with the Pma1-tFT strain, a well with the Hsp26-tFT strain and a well with the Sec21-tFT strain. Those control strains were streaked from a single clone and will serve as a control for possible batch effects along the microscopy experiment. Each plate, 193 wells, was imaged within 1 hour.



**Figure 20.** Representation of one 384-imaging plate from genome-wide microscopy. **a.** Each plate was imaged using the Opera Phenix microscope, under the 63x water objective. Channels acquired were BF (500 ms of exposure and 100% laser power), red channel (500 ms and 70%), green channel (500 ms and 100%) and blue channel (500 ms and 100%). Per well 15 fields of view were acquired. **b.** Representative images of asymmetrically segregated proteins. All images correspond to one field of view of the given strain-tFT from the ORF-mS-I-mNG-hphNT1 library. All images were auto contrasted. Scale bar, 5  $\mu$ m.

While the creation and imaging of the library was finalized, the pipeline for the image analysis is still under development. For that reason, I decided to annotate myself the protein segregation phenotype that I observe for each strain of the ORF-mS-I-mNG-hphNT1 library.

In order to visually annotate the library, I decided to only annotate whether a protein seemed to be strongly asymmetrically segregated, and additionally, I noted down if the asymmetry consisted of the older protein pool being localized in the mother or if the older protein pool was segregated to the bud.

In total, visually I was able to identify 775 proteins having an asymmetric segregation between mother and bud compartments. Some examples of the final strains of the ORF-mS-I-mNG-hphNT1 library are shown in Figure 20b.

## **DISCUSSION**

### **Evaluation of different tFTs as reporters for protein degradation**

After performing whole colony fluorescent measurements and determining the redFP/greenFP ratio of different X-tFT constructs (Figure 13a), it is clear that with all tFTs we are able to distinguish between very unstable constructs like D-tFT and very stable ones like T-tFT. Therefore, we can conclude that the five tFTs tested are good reporters of protein degradation.

Nevertheless, we observed that the mCherry-mNeonGreen tFT showed the narrowest dynamic range (indicated with red arrows in Figure 13a) of the redFP/greenFP ratio between X-tFT constructs. Contrary, using the rest of the tFTs, we were able to better differentiate between very stable (N-tFT) and very unstable (T-tFT) constructs. The mCherry-sfGFP showed the largest dynamic range, followed by all mScarlets-mNeonGreen tFTs. For studying protein stability a tFT with a wider dynamic range, like mCherry-sfGFP and mScarlets-mNeonGreen, is recommended, since they would allow detecting finer variation in the stability of a protein.

It has been described how the order of the FPs within a tFT can affect its time range due to the incomplete proteasomal degradation of some greenFPs like sfGFP (Khmelinskii et al. 2016). This incomplete degradation of sfGFP increases the difference between mCherry/sfGFP fluorescence ratio of stable and unstable fusions, as we see when using mCherry-sfGFP with very unstable residues, like D and N, we observed how the red/green fluorescence ratio was considerably lower than when using the rest of the tFTs.

As part of the characterization of the tFTs, I also performed a western blot using unstable and stable constructs with each tFT (Figure D). As seen in Figure D panel a, this membrane contained tFTs with sfGFP but also with mNG, nevertheless the image on the left represents this membrane incubated with anti-GFP antibody. As observed, anti-GFP antibody is useful to detect sfGFP (lanes 1 - 4) and mNeonGreen (lanes 5 - 10).

Taking into account that sfGFP and GFP have a nucleotide similarity of 95% (based on NCBI blast), it is understandable that a GFP antibody would recognize sfGFP as well. On the other hand, the nucleotide similarity between GFP and mNG is 34%,

which confirms that the anti-GFP antibody is not specific enough. When the same membrane was incubated with anti-mNG, only constructs carrying mNeonGreen were detected, concluding that the anti-mNG antibody is very specific.

In Figure 13a we confirm that all the tFTs are good reporters for protein stability, since we can differentiate the unstable constructs (N-tFT) from the stable ones (T-tFTs). For all tFTs, the signal for the very stable constructs is considerably higher than for the unstable constructs.

In all membranes in Figure 13b, we observe different degradation products (a, b and c fragments) in each lane. In the western blot, we observed a band corresponding to the molecular weight (MW) of a full length X-tFT (>55 kDa). We also detected a band in between 43 and 55 kDa which would correspond to a shorter mCherry<sup>ΔN</sup>-greenFP product of mCherry hydrolysis during cell extract preparation (Gross et al. 2000; Shemiakina et al. 2012). As we also detected this band for the other redFPs, this would indicate the mScarlets potentially undergo a sort of hydrolysis during protein extraction. Aligning the sequences of all red FPs tested in Figure 13, regardless of the 87 % of similarity between mCherry and mScarlet variants, I observed how the N-terminus of all mScarlet variants has a 24 fragment completely different to mCherry sequence which might make a difference in the degradation in between these red FPs.

As described in previous studies (Gross et al. 2000), upon harsh treatments like boiling, DsRed undergoes hydrolysis after the Phe-65. The nucleotide similarity between DsRed and mCherry is 87%, and 85% of similarity for DsRed and mScarlet-I and mScarlet, and 86% to mScarlet-H. Taking into account this nucleotide similarity, and the presence of the amino acid Phenylalanine at position 65th, we could hypothesize that the rest of the red FPs could also undergo a similar hydrolysis during the protein extraction protocol, that would lead to the presence at different amounts of mCherry<sup>ΔN</sup>-greenFP products (Figure D, b) upon protein degradation of the tFTs.

Not significant similarity found for FusionRed and mCherry nor FusionRed and mScarlet-I (based on NCBI blast). Nevertheless, it has been shown that FusionRed exhibits a partial cleavage around the 65<sup>th</sup> amino acid position of the chromophore (Shemiakina et al. 2012). This study also showed how mCherry and FusionRed both undergo partial cleavage upon incubation at 99 °C, leading to fragments of a bit larger than 30 kDa (full length of each FP), a fragment at 20 kDa and a fragment at around 10k Da for FusionRed and a fragment a little bit larger than 10kDa for mCherry on SDS-PAGE. This would reinforce the possibility of other red FPs, like mScarlet variants, to undergo partial cleavage under different denaturing conditions.

As reported before, incomplete degradation of the sfGFP fold will result in tFT fragments that remain in the cytosol (Khmelinskii et al. 2016), with a molecular weight in between 25 and 37 kDa and observed in a western blot as two bands in with very similar molecular weight. I also observed these processed tFT bands with the five tFTs tested (Figure D fraction c), detecting one band in between 34 and 43 kDa for almost all tFTs. This fraction c has almost an undetectable signal for constructs with mS-mNG (Figure D panel b, lanes 13 - 16).

For mCherry-sfGFP only, I also observed a smaller band in between 17 and 26 kDa. These processed-tFT fragments bands have been previously reported for other mCherry-greenFP tFTs, especially prominently for mCherry-sfGFP. These fragments



are supposed to be the result of incomplete proteasomal degradation of the greenFPs and are reported to exhibit a molecular weight in between 25 and 37 kDa.

This different observation in between the processed-tFT fragments in my western blot and in the cited study (Khmelinskii et al. 2016), might be due to the different materials and conditions in each laboratory. I used a 12 % acrylamide SDS-PAGE while they used a gradient SDS-PAGE (NuPAGE Novex 4-12% Bis-Tris protein gels; Thermo Fisher Scientific, Waltham, MA). Additionally, I used the mouse monoclonal anti-GFP (11814460001 Roche) and the secondary antibody goat anti-Mouse IgG (H+L) Cross-Adsorbed (G-21040ThermoFisher), while in the mentioned study they used a mixture of the same anti-GFP antibody and mouse anti-HA antibody, as for secondary antibody they used goat anti-mouse IgG-HRP antibody (sc-2005; Santa Cruz Biotechnology, Dallas, TX). These could be some reasons why we observed a different pattern for the processed tFT fragments.

Another interesting observation was the higher signal of extra processed tFTs mSH-mNG, compared to the rest of the tFTs which did not carry sfGFP, for both N and T-mSH-mNG constructs. mScarlet-H has a three nucleotide mutation, compared to mScarlet-I and mScarlet, that changes the nature of one 164<sup>th</sup> amino acid. This single amino acid change has been reported to make mScarlet-H the slowest maturing of these three mScarlets (Bindels et al. 2017) mammalian cells. mScarlet-H has reported to exhibit a maturation time of 264 min, while mScarlet-I and mScarlet exhibit maturation times of 174 and 23 min, respectively. Since this single amino acid substitution not only affected the maturation time of mScarlet-H, but also made it the most photostable of these mentioned mScarlets, we could speculate that this mutation could be also affecting its degradation.

I confirmed that in yeast, mScarlet-I was brighter than mCherry Figure 14 as reported before (Meurer et al. 2018). mScarlet-I was also the brightest of the mScarlet variants in yeast, as reported in mammalian cells (Bindels et al. 2017).

### **Evaluation of different tFTs as reporters for protein protein-age based segregation**

As we observed in Figure 13a, tFTs like mCherry-sfGFP and mScarlets-mNeonGreen are tFTs with a wide dynamic range, which allows us to discriminate in between a larger range of protein ages (Fung, Blöcher-Juárez, and Khmelinskii 2022). Using confocal fluorescence microscopy, I confirmed that mScarlet-I was the brightest of the red FPs tested, *in vivo* (Figure 14b). On the other hand, mNeonGreen is brighter than sfGFP as it was previously reported before in budding yeast (Meurer et al. 2018).

Besides the brightness, I was also interested in selecting a tFT with a large dynamic range to be able to detect a wider variety of protein ages. Hence, I was interested to select the fastest maturing green FP and the slowest maturing red FP.

In mammalian cells, mScarlet-I was reported to be the brightest of the mScarlet variants, followed by mScarlet and mScarlet-H. In the same cell type, the slowest maturing FP was mScarlet-H, followed by mScarlet and mScarlet-I (Bindels et al. 2017). My results show that, as in mammalian cells, in budding yeast the brightest mScarlet is mScarlet-I, followed by mScarlet and mScarlet-H (Figure 14). Differently

than in mammalian cells, in yeast mScarlet seems to be the one with the slowest maturation, succeeded by mScarlet-H and mScarlet-I (Figure 15b).

As observed in Figure 15a, we were able to detect the asymmetric segregation of Pma1 using all tFTs tested. For which, we conclude that all the candidate tFTs can, indeed, be used as such and they are good reporters of protein age in yeast.

Published data shows that the median Rb/ Rm for Pma1-mCherry-sfGFP was around 0.3, similarly to Pma1-mCherry-mNeonGreen-HA was around 0.3 (Khmelinskii et al. 2016). In my assessment, the median Rb/ Rm for Pma1-mCherry-sfGFP was around 0.2 and 0.3 for Pma1-mCherry-mNeonGreen; slight differences in the final Pma1-tFT Rb/ Rm might be present due to the use of a different microscope and settings used for its determination.

Altogether, I decided to select mScarlet-I as the red FP to be used in a tFT. mScarlet-I being the brightest red FP, but the fastest maturing mScarlet variant tested in yeast, will allow us to detect a larger amount of proteins while assessing genome-wide protein segregation. For follow up experiments on proteins with known segregation patterns, we could envision making use of a slower-maturing red FP like mScarlet.

As for the green FP, I decided to use mNeonGreen due to its brightness in yeast.

### **Generation of a donor strain with markers for image segmentation**

It is important to keep in mind that depending on which locus the different blue cytosolic markers (Figure 16) are integrated into, their final fluorescence or localization might be different than the one determined by me when introduced in the *leu2Δ* locus. Even when expression and localization levels are not dependent on gene locus, there could be special cases where this does not apply.

It has been reported that the deletion of *Shs1* causes cold-sensitivity in the W303 genetic background (Iwase et al. 2007), reflected in slow growth at 25 °C for *Shs1Δ* strains, and not growth at 20 °C.

Therefore, I decided to test whether the septin tagged strain presents cold-sensitivity due to perturbation of the protein by the tag, as seen for the *Shs1* deletion. After performing spotting assays of all the strains, I did not observe any difference in growth in between strains at 20 °C. The strains also grew in a similar fashion at 30 °C (Figure 18c). The lack of cold-sensitivity might be due to the different genetic background of my strains, which is BY4741. The BY4741 strain is a well studied and commonly used genotype in the yeast research, and it was part of a set of deletion strains generated from the S288C genetic background (Brachmann et al. 1998). On the other hand, the W303 genetic background has derived in several variants, one widely used being the W303-K6011 and sharing 85% of its genome with S288C (Ralser et al. 2012). Very likely, due to some genetic differences between the W303 and the S288C background, the *Shs1Δ* is cold-sensitive in the first one but not in the latter. The 2 populations could be competing. This might not have an effect in the way protein segregates, but could affect the lesser or greater percentage of sample cells.

In order to further evaluate the growth differences between strains, I decided to use a more sensitive method to detect any phenotypic differences between strains, therefore

I generated growth curves. When assessing the growth in liquid, I confirmed that the Septin tag strain (cytosolic GPDpr-BFP2-CYC1ter marker and Shs1-3xBFP2-ADH1ter) has a growth tendency very similar to the one of the Control strain (cytosolic GPDpr-BFP2-CYC1ter marker). On the other hand, the Septin deletion strain (Shs1 $\Delta$ ::hphNT1) exhibits a slower growth compared to the other strains. This confirms that the tag used for the Shs1 septin is not affecting the cell growth of the strain.

Lastly, to further characterize those strains, I analyzed their phenotype by confocal fluorescence microscopy of unsynchronized cultures in exponential phase (Figure 18c). By this the observation was that the Septin tag strain cells have a similar phenotype as the Control cell, where dividing-pairs can be distinguished along the field of view. In contrast, the Septin deletion strain exhibited chain formation among cells, which made it impossible to distinguish between dividing-pairs. As it has been reported before, the formation of chains is a distinctive indicator for cytokinesis defect (Iwase et al. 2007). All this evidence indicates that the Shs1-3xBFP2-ADH1ter tag, nor the presence of a cytosolic marker are causing cytokinesis defects in the Septin strain.

### **Construction of genomewide C-tagged libraries**

After deciding for mScarlet-I-mNeonGreen as the tFT of choice to study the protein segregation in yeast, I proceeded to construct and test donor plasmids with this tFT. I created donor plasmids (Table S5) to tag proteins at the C-terminus and N-terminus, as well as to tag proteins endogenously (tagging Type I) and under a marker, hygromycin, reconstitution approach (tagging Type II).

After the generation of those donor plasmids, I decided to tag all of the yeast genome at the C-terminus, since the majority of the available protein abundance and localization datasets that I could use to compare my results to, are studies where the yeast genes were tagged C-terminusly with a FP or an epitope (Ho, Baryshnikova, and Brown 2018).

Regarding the tagging strategy, an endogenous tagging approach allows to preserve the gene regulatory sequences downstream of the tag, with the constraint that there is no marker for positive selection present in the final strain. While using the selection reconstitution tagging, a truncated hygromycin marker (hph $\Delta$ C) is placed within the donor plasmid (mScarlet-I-mNeonGreen-ADH1term-TEFpr-hph $\Delta$ C), so the hygromycin marker is functional after a successful tag-swapping procedure. While the selection reconstitution tagging provides a reliable positive selection of swapped events, the seamless tagging provides the opportunity to explain possible differences in protein levels after tagging or to notice a different segregation pattern, which could be associated with regulatory events at 3'UTR (Yofe et al. 2016; Meurer et al. 2018).

Since my aim was to assess protein segregation by fluorescence microscopy, I opted for using a non-endogenous tagging approach to take advantage of the presence of reconstituted hygromycin as marker for positive selection. In follow up libraries, an endogenous tagging at both N and C-termini could be performed and phenotype through fluorescence microscopy as well.

When I was optimizing the microscopy protocol, I observed that there was a significant difference in the cellular density at the beginning and at the end of the imaging per plate, this was expected since I performed live imaging and the procedure lasted 1 h (Figure S2). To contrast the possible effect of this technical detail in the phenotyping outcome, I imaged two replicates of my library. In each replicate I inverted the order in which the strains were imaged, therefore I ended up having one set of low and one set of high cell density images per strain.

After constructing the ORF-mS-I-mNG-hphNT1 library I assessed the tagging efficiency of 20 strains, which had been previously used as reference due to their high abundance (Meurer et al. 2018). Additionally, I also measured the tagging efficiency for other strains one expressing a high abundance protein (Pdc1) that I had used in previous experiments and other three strains expressing the proteins used as controls for batch effect during the high-throughput microscopy (Hsp26, Pma1 and Sec21). For all of those 24 strains, the tagging efficiency was above 98% (Figure 19b). This percentage of cells successfully tagged per strain, led me to extrapolate that the rest of the library strains were also tagged successfully since they all were created simultaneously.

After imaging 5,561 strains, I manually annotated their segregation phenotype while the imaging pipeline is still under development. After this visual annotation, I was able to identify 775 proteins which had a strong asymmetric segregation phenotype between mother and daughter cells.

Revising the previously published yeast protein segregation catalog (Sugiyama and Tanaka 2019), referred to as “Segregatome” henceforth, I realized they managed to detect mainly proteins with high abundance (Figure S1). Proteins like Sur7 and Spc105 were not detected in that study, nevertheless, in my library they exhibited a bright enough signal to annotate their segregation pattern by eye (Figure 20b). In the published yeast protein abundance dataset, hereafter Unification dataset, (Ho, Baryshnikova, and Brown 2018) Sur7 exhibits a relatively high abundance, with a median molecules per cell of 7334 which is above the median of the whole abundance dataset. On the other hand, Spc105 was listed with a median molecules per cell of 1969, which in turn represents a low abundance.

These two examples of proteins, with high and low abundance, that were previously not detected by mass spectrometry but were detected under our own approach, are a good confirmation that our method to detect protein segregation is more sensitive.

Nevertheless, among the proteins I detected as asymmetric, I also found Dip5 and Hxt7, which were reported to segregate asymmetrically in the mentioned study (Sugiyama and Tanaka 2019). Examples like these served as controls, and as a confirmation that our approach is reliable.

Out of 56 proteins found asymmetric in the Segregatome study (Figure S1), I identified only 20 of those as showing an asymmetric segregation in my library; these overlapping proteins were the ones which had the highest asymmetric index in the Segregatome study. From the rest of the proteins with asymmetric index in the Segregatome study, 3 ORFs were not present in the initial C-SWAT library that I used to construct my library and 31 ORFs were instead symmetric in my annotation and 2 proteins had unclear segregation when analyzed by eye.

In the Segregatome study, out of a total of 2039 detected proteins, 152 were annotated as “Not consistent”, since their segregation phenotype was contrasting in between their two replicates. Some of those “Not consistent” ORFs, had a clear asymmetric phenotype in both replicates of my library, like YMR297W. In the Unification dataset, this ORF was reported to have a median abundance of 5557 molecules per cell, which is higher than the median abundance of the whole dataset that was set at 2069 molecules per cell and the maximum abundance was 659598.

Therefore, we could hypothesize that ORFs would be difficult to classify under the Segregatome approach when their abundance is medium or low, as their classification efficiency might decrease in proportion with protein abundance as seen in Figure B1.

Another example of “Not consistent” ORF in the Segregatome study was the ORF YCL025C, which in my library clearly segregates symmetrically (data not shown here) and its abundance was reported to be very close to the median abundance of the Unification dataset.

The protein Cap2 (Figure 20b), visually, had an asymmetric segregation pattern in my library, where the newly synthesized and also the old protein pool seemed to mainly localize to the mother compartment. Contrastingly, Cap2 had been previously classified as symmetrically segregated in the Segregatome study. It is important to mention that my manual annotation of the protein segregation was done as a temporary resource. For the strain Cap2-tFT, for example, it is unclear whether the segregation is really asymmetrical or if both old and newly synthesized protein pools are exclusively localized in the bud, at least during mitosis. These finer details might be cleared up after the automated quantification is finalized.

While manually annotating the segregation phenotypes, I realized for punctae-looking proteins, an additional imaging will be needed where multiple planes are acquired to be able to cover a larger volume of the cell and accurately determine the segregation phenotype of these proteins. An example of such punctae-like proteins are the different components of the spindle pole body (SPB). One of the components of the SPB, the Spc42 protein is known to segregate asymmetrically (Figure B), inheriting the old protein pool to the bud and the newly synthesized protein to the mother cell (Khmelniskii et al. 2012). In my library for Spc42, some cells have this asymmetric phenotype and others a symmetric one. These different phenotypes were dependent on how well focused the small buds were, and whether this small protein was completely visible at 2  $\mu\text{m}$  of our acquired plane.

In order of increasing abundance (Ho, Baryshnikova, and Brown 2018) B components: Spc72, Spc97, Spc98, Spc34, Spc29, Spc25, Spc19, Spc42, Spc105, Spb4, Spc110, Spc24, Spc1, Spb1, Spc3, Spc2. From these SPB components, only Spc34, Spb4, Spc3 and Spc2 were detected in the Segregatome study.

Besides the increased sensitivity to detect low abundance proteins, my microscopy approach also has the advantage of being live cell imaging. Through the culturing protocol I optimized that even when cultures are unsynchronized, the majority of cells are in exponential phase.

As a pending calculation, when the automated imaging pipeline is ready we need to assess the tagging efficiency per ORF. By calculating how many of the sample cells

per fov had green and red signal, we could have a list of tagging efficiency per ORF. This would be specially informative to have a hint, that ORFs with low tagging efficiency in my library, might carry some reported or unknown signal peptides at the C-terminus that make their tagging in this termini detrimental. Maybe for such low tagging efficiency ORFs, tagging at their N-terminus would be a suggestion.

Additionally, it would be informative to try to calculate how strong are the segregation phenotypes across the library. This could be estimated by getting a percentage of how dividing-pairs per fov show the asymmetric/symmetric phenotype. Until now, it is still an open question how strong each of these phenotypes might be, or if they are changing across the cell-cycle.

As a final remark, it is important to keep in mind that the segregation catalog that we provide in this study reflects the segregation of proteins only during a single cell cycle. Therefore, it will be required to design an experiment, like time-lapse microscopy, to follow the protein segregation patterns through multiple cell cycles and confirm which of these asymmetrically segregated proteins identified maintains that phenotype over time.

## **MATERIAL AND METHODS**

### **Molecular techniques used for multiple experiments**

The same gel-purification kit, PCR purification kit and Miniprep kit were used repeatedly throughout the study, otherwise indicated.

For the ligation of vector and inserts for cloning, the Gibson assembly method (Gibson et al. 2009) was used, along with the Gibson assembly master mix from NEB (E211) and the NEB 5-alpha Competent *E. coli* (C2987), otherwise indicated.

The nouns primers and oligos were used interchangeably along the text. When referring to a primer, the primer will be indicated with the number of the box where it is stored, over the number of the primer position within that box, example primer 4/65 refers to the primer 65th in the box 4th.

### **Preparation of budding yeast competent cells**

Heat-denature carrier DNA (Invitrogen UltraPure Salmon Sperm DNA solution) at 95 °C for 10 min. Then leave it on ice.

1. Prepare 50 ml of a liquid culture ( $OD_{600} \sim 0.8$ ):

- inoculate the culture from a single colony/ patch the night before and incubate at 30°C and shaking at 220 rpm (use selective medium if necessary, e.g. to maintain CEN/2 $\mu$  plasmids)

- measure the OD in the morning. If the OD is too high (> 1), dilute the culture back to  $OD \sim 0.15$  in 50 mL fresh medium. Place the dilution in the shaker until the culture reaches  $OD 0.6-0.8$  (3-4 h at 30 °C considering a doubling time of  $\sim 2$  h)

2. Spin down the culture at 3200 rpm for 2 min at room temperature
  3. Discard the supernatant and wash cells in 25 ml of sterile H<sub>2</sub>O. Spin down
  4. Discard the supernatant and wash cells in 12.5 ml of LiSorb. Spin down
  5. Discard the supernatant. Spin down for 1 min and remove remaining LiSorb with a pipette
  6. Add 300 µl of LiSorb
- Prepare aliquots of 43 µL of the competent cells in and freeze at -80 °C

If the competent cells are used directly:

7. Add 7 µL of denatured carrier DNA and resuspend cells by pipetting
8. Use cells for transformation directly

**Note 1:** competent cells can be frozen after step 6, without adding carrier DNA.

**Note 2:** the volumes of LiSorb and carrier DNA in steps 6 and 7 are for a 50 mL culture at OD 0.8. Adjust according to the volume and density of the culture.

### Budding yeast transformation

Use freshly prepared competent cells (which have a higher transformation efficiency) or frozen competent cells (with carrier DNA). Do not keep cells on ice. Work at room temperature. Do not refreeze competent cells.

Before starting the transformation protocol, set a water bath to 42 °C.

1. Mix 43 µL of competent cells and 7 µL of heat-denatured carrier DNA (or 50 µl of competent cells if the carrier DNA was previously added) with sample DNA (up to 5 µL):

- for cassette integration, use 5 µL of PCR product (~1000 ng)
- for plasmid integration, use ~200 ng of digested plasmid
- for plasmid transformation, use ~25 ng of DNA (in 1-5 µL volume)

Incubate at room temperature for 15 min.

2. Add 300 µL of LiPEG, mix by vortexing and incubate for 15 min at room temperature
3. Add 30 µL of DMSO, mix by vortexing and incubate for 10 min at 42°C
4. Spin down cells at 1900 g (or 4.500 rpm) for 3 min at room temperature and remove the supernatant by aspiration
- 5a. For auxotrophic selection markers, resuspend the cells in 100 µL of sterile PBS

and plate out on appropriate selection plates

**5b.** For antibiotic resistance markers, allow cells to recover before plating out:

- resuspend cells in 1 mL of YPD medium
- incubate for 6 h or overnight at 30 °C (or 23 °C for *ts* strains)
- spin down the cells at 3200 rpm for 3 min at room temperature, remove 900 µL of medium
- resuspend the cells in the remaining 100 µL and plate out on appropriate selection plates
- Place the plates in the 30 °C incubator and let them grow for two days

**Note 1:** An alternative step 3 is to heat shock for 40 min at 42°C without adding DMSO

### Plasmid extraction from yeast strains

Yeast plasmid extraction was performed using Zymolyase and Miniprep methods. After growing saturated cultures, around  $1.0-1.5 \times 10^7$  cells were centrifuged at 600 g for 2 min. Then, the supernatant was then removed and the pellet was resuspended in the buffer containing Zymolyase for the digestion of the yeast cell wall. The mixture was incubated at 37 °C for 60 min. Subsequent steps were the same as for *E. coli* plasmid extraction using the Miniprep kit from Qiagen. Next, 5 -10 µl of isolated DNA was used to transform into *E. coli* cells for propagation. Plasmids were then extracted using the Miniprep kit from Qiagen. The correctness of all plasmids was always confirmed by restriction enzyme digestion followed by Sanger sequencing to confirm different DNA sequences within each plasmid.

### Yeast whole cell extract using TCA

First: Have cold ddH<sub>2</sub>O on ice (1 ml per sample) ready before starting the protocol. Work on ice as indicated below.

1. Collect 2-3 OD of cells in microcentrifuge tubes by centrifugation at 3200 rpm for 2 min at room temperature
2. Resuspend cell pellet in 1 ml of cold water by vortexing. Keep tubes on ice  
Alternatively, resuspend some cells from a fresh patch in 1 ml of cold water  
Alternatively, just use 1 ml of a culture (only for synthetic media, rich media have too much material that will contaminate whole cell extracts)
3. Add 150 µl of 1.85 M NaOH. Mix by vortexing. Incubate 10 min on ice
4. Add 150 µl of 55% (w/v) TCA (w/v). Mix by vortexing. Incubate for 10 min on ice
5. Centrifuge at 14000 rpm for 15 min at 4 °C. Remove as much supernatant as possible by aspiration, centrifuge again for 1 min and remove remaining supernatant



6. Add HU buffer to the pellet (precipitated proteins) at 50  $\mu$ L HU buffer per OD. Store samples at -20 °C or proceed

7. Resuspend pellet by vortexing. If supernatant turns yellow, add 1-2  $\mu$ L of 2 M Tris-HCl pH 8 to neutralize

8. Heat samples for 15 min at 65 °C with mixing to fully dissolve the pellet. Centrifuge at 14000 rpm for 10 min before loading on the gel. Load 10-15  $\mu$ l per well

**Note:** If necessary, dilute MW marker with HU buffer so as to load the same volume per well. Load empty wells with the same volume of HU buffer. HU buffer (high-urea buffer): 8 M urea, 5% SDS, 200 mM Tris-HCl pH 6.8, 0.1 mM EDTA, bromophenol blue (according to taste). Make 10 mL aliquots in 15 ml falcon tubes and freeze at -20°C. Before use, add DTT to 15 mg/mL final concentration. SDS-PAGE: ~30 mA per mini-gel, 150-200 V (max).

### Confocal fluorescence microscopy

In order to validate the clones, POI-tFT, POI-FP or FP-POI, or deletion of a given POI, I used microscopy to verify this. I used the high-throughput spinning disk confocal microscope Opera Phenix.

Work under sterile conditions is required along the whole procedure. Turn on your burner.

1. Fill up 96-well plate wells, as many wells as the number of clones to validate, with 150  $\mu$ L of the required medium (supplemented with adenine to decrease autofluorescence of cells)

2. Seal the plate with an air-o-seal. Close it with its plastic lid, place it in the incubator at 30 °C and 220 rpm. Add a box's lid on top of it to prevent evaporation. Grow to saturation (overnight)

3. Next day using the pinning robot (Singer Instruments), inoculate a new 96-well plate pre-filled with 150  $\mu$ L of the required media per well with the saturated culture

4. Grow for 6 h to  $0.5 \times 10^7$  cells/ mL

5. 60  $\mu$ L of 6 h culture per well were transferred to 384-well plates (PhenoPlate 384-well microplates, PerkinElmer) previously coated with 2 mg/ mL concanavalin A (C2010, Sigma-Aldrich) solution in water (Khmelinskii and Knop 2014)

6. While waiting, thaw aliquots of 2 mg/ mL concanavalin A (C2010, Sigma-Aldrich). Spin at 16000 x g for 5 min at 4 °C before use. Use 70  $\mu$ L of the supernatant to coat each well. Let it rest on the bench for at least 30 min. Just before going to the microscope, remove the concanavalin A with a pipette, wash the wells with 100  $\mu$ L ddH<sub>2</sub>O water and proceed to transfer the 150  $\mu$ L of your 6 h culture to each coated well

7. Go to the microscope. Imaging was performed at room temperature on a high-throughput spinning disk confocal microscope (Opera Phenix, PerkinElmer), equipped

with four laser lines (405nm, 488nm, 568nm, 640nm), a 63X water-immersion objective (NA 1.15) and two 16-bit CMOS cameras (2160 by 2160 pixels). For each well, single plane images of at least 15 fields of view were acquired with 2x2 pixel binning (corresponding to 300 nm pixel size) in the following order: red channel (599 nm emission filter, 561 nm excitation), brightfield (0 nm emission filter, 740 nm excitation), blue channel (456 nm emission filter, 405 nm excitation), mNeonGreen (522 nm emission filter, 488 nm excitation). After optimizing this protocol, I determined a focused plane for my haploid yeast strains at 2  $\mu$ m, which was always the plane depth acquired.

**Note 1:** In all the microscopy experiments, image one well containing media only, to correct for autofluorescence of the media.

**Note 2:** When staining the cells with Calcofluor, add it during step 3.

**Note 3:** For the genome-wide high-throughput microscopy, I used SC low fluorescence medium (6.7 g/ L bacto yeast nitrogen base without amino acids and without folic acid and riboflavin (LoFlo, Formedium), 2 g/ L amino acid dropout mix, 2% (w/v) glucose) supplemented with 300 mg/L adenine (A8626, Sigma-Aldrich). For the genome-wide high-throughput microscopy I used 500 ms of exposure for all fluorescent channels and 70 % of laser power, while for brightfield I used 100 % of laser power.

### Generation of plasmids carrying different tFTs

In order to evaluate different tFTs as reporters of protein degradation, we used the plasmid pAnB19 (Kats et al. 2018) as a backbone. The plasmid pAnB19 carried the necessary elements (GPDpr-Ubiquitin-X-e<sup>k</sup>) to evaluate the half lives of different synthetic proteins in which a single amino acid (X) determines their stability upon cleavage of the ubiquitin moiety and exposure of X residue as the new N-terminus (see Figure A).

I cloned each tFT after the e<sup>k</sup> sequence of the pAnB19 plasmid, and created five new plasmids (Table 2). Since the backbone plasmid pAnB19 already carried the mCherry-sfGFP tFT, I just additionally cloned the CYC terminator (CYCter) after the tFT. Therefore to construct the pKBJ001 plasmid (GPDpr-Ubi-EcoRV-STOP-eK-mCherry-sfGFP-CYC1ter) I digested the pAnB199 plasmids with the restriction enzymes XhoI and ApaI at 37 °C for 30 minutes (Table 3), followed by gel-purification (QIAquick Gel Extraction Kit Cat. No. 28706X4) of the largest fragment of 7480 bp which was used as a vector. The insert, CYCter, was PCR-amplified from the plasmid pEI076 (Table 3 and Table 4) and gel-purified. Vector and insert were assembled by the Gibson assembly method (Gibson et al. 2009) using the Gibson assembly master mix from NEB (E211) and the NEB 5-alpha Competent *E. coli* is a derivative of the popular DH5 $\alpha$  (C2987). The final construct was validated by restriction digestion with the same enzymes used initially, XhoI and ApaI, and by Sanger sequencing of the CYCter insertion (oligos 4/60, 4/61).

**Table 2.** List of Promoter-Ubi-EcoRV-STOP-eK-tFT-terminator plasmids constructed.

Plasmid ID	Backbone and tFT
pKBJ001	GDPpr-Ubi-EcoRV-STOP-eK-mCherry-sfGFP-CYC1ter
pKBJ002.1	GDPpr-Ubi-EcoRV-STOP-eK-mCherry-mNeonGreen-CYC1ter
pKBJ003.1	GDPpr-Ubi-EcoRV-STOP-eK-mScarletI-mNeonGreen-CYC1ter
pKBJ004.1	GDPpr-Ubi-EcoRV-STOP-eK-mScarlet-mNeonGreen-CYC1ter
pKBJ005.1	GDPpr-Ubi-EcoRV-STOP-eK-mScarletH-mNeonGreen-CYC1ter

The backbone for all the plasmids constructed in Table 2 was the plasmid pRS413.

**Table 3.** Restriction enzyme (RE) digestion.

Component	10 $\mu$ L reaction
Plasmid DNA (1000 ng)	1 $\mu$ L
RE 1	0.5 $\mu$ L
RE 2	0.5 $\mu$ L
10x rCutSmart buffer	1 $\mu$ L (1x)
ddH <sub>2</sub> O	To 7 $\mu$ L

Most of the restriction enzyme digestions in this study were carried out at 37 °C for 30 min, since the majority of the REs used were available in their High-fidelity version from NEB. Otherwise different conditions are indicated. \* For AarI, the buffer 10x Tango+ oligo buffer from Termpfisher was used instead of rCutSmart buffer.

**Table 4.** PCR mix for the amplification of inserts for cloning.

Component	25 $\mu$ L reaction
ddH <sub>2</sub> O	15.75 $\mu$ L
HF-buffer	2.5 $\mu$ L
DMSO (100 %)	1 $\mu$ L (4 %)

Betaine (5 M)	2.5 $\mu$ L
MgCl <sub>2</sub> (50 mM)	0.25 $\mu$ L
DNA template (1000 ng)	1 $\mu$ L
dNTPs (10 mM)	0.5 $\mu$ L
forward (fw) primer 10 $\mu$ M	0.5 $\mu$ L
reverse (rv) primer 10 $\mu$ M	0.5 $\mu$ L
HF-Polymerase	0.5 $\mu$ L

High-fidelity (HF) buffer and Polymerase were acquired from the Protein production core facility in IMB.

**Table 5.** PCR cycling conditions for the amplification of inserts for cloning.

Step	Temperature (C) - Time (s)
Initial denaturation	98 - 30
Denaturation	98 - 10
Annealing	T <sub>m</sub> (primers)* - 30
Elongation	72 - 30 per kbp (go to denaturation 30 cycles)
Final elongation	72 - 300
Pause	4

The plasmid pKBJ002.1 carrying **mCherry-mNeonGreen** (pRS413-GPDpr-Ubi-EcoRV-STOP-eK-mCherry-mNeonGreen-CYC1term) was constructed using the plasmid pKBJ001 as a backbone. The backbone was digested using AarI and XhoI following the conditions in Table 3\*, and the fragment of the vector was gel-purified. The insert, mNeonGreen sequence, was PCR-amplified (Table 4 and Table 5) from the plasmid pYD10 (Meurer et al. 2018). Vector and insert were assembled by the Gibson assembly method (Gibson et al. 2009). The final plasmid was validated by restriction digestion with the same enzymes used initially, and by Sanger sequencing of the CYCter insertion (oligos 4/64, 5/65).

Plasmid pKBJ003.1 containing the **mScarlet-I-mNeonGreen** tFT (pRS413-GPDpr-Ubi-EcoRV-STOP-eK-mScarletI-mNeonGreen-CYC1term) was constructed using the plasmid pKBJ002.1 as backbone. The plasmid pKBJ002.1 was digested with Sall-HF and AarI. Then, the vector was gel-purified. The insert, mScarlet-I, was PCR-amplified from the plasmid pYD13 (Meurer et al. 2018) and gel-purified. Vector and insert were assembled by the Gibson and transformed into *E. coli* competent cells. The final plasmid was validated by restriction digestion with the enzymes Sall-HF and EcoRV-HF and by sequencing the insertion (oligos 4/64, 5/65).

For constructing the plasmid pKBJ004.1 **mScarlet-mNeonGreen** (pRS413-GPDpr-Ubi-EcoRV-STOP-eK-mScarlet-mNeonGreen-CYC1term), the backbone pKBJ003.1 within the mScarlet-I sequence was mutagenized by PCR site-directed mutagenesis (Bachman 2013) following the condition indicated in Table 6 and Table 7.

For this PCR, fw and rv overlapping primers contained the single amino acid substitution I74T (from isoleucine to threonine in the 74th amino acid) to generate an amplified sequence corresponding to the mScarlet variant. The PCR-amplified and gel-purified plasmid was transformed into *E. coli* competent cells. To confirm the absence of new mutations along our reporter region, we validated the new plasmid by Sanger sequencing of the region GPDpr-Ubi-eK-mScarlet-mNeonGreen-CYC1term (oligos 1/86, 4/61, 5/55, 4/62, 4/63, 6/16, 4/66 and 4/68).

**Table 6.** PCR mix for the amplification of vectors with overlapping primers for site-directed mutagenesis.

Component	25 $\mu$ L reaction
ddH <sub>2</sub> O	8.125 $\mu$ L
Pfu-buffer	1.25 $\mu$ L
DMSO (100 %)	0.625 $\mu$ L (5%)
DNA template (2 ng)	1 $\mu$ L
dNTPs (10 mM)	0.5 $\mu$ L
forward (fw) primer 2.5 $\mu$ M	0.5 $\mu$ L
reverse (rv) primer 2.5 $\mu$ M	0.5 $\mu$ L
Pfu-Polymerase	0.25 $\mu$ L

**Table 7.** PCR cycling conditions for PCR site-directed mutagenesis.

Step	Temperature (C) - Time (s)
Initial denaturation	95 - 300
Denaturation	95 - 50
Annealing	T <sub>m</sub> (primers)* - 50
Elongation	68 - 1020 (2 min per kbp; go to denaturation 18 cycles)
Final elongation	68 - 600

Pause	4
-------	---

For the construction of the plasmid pKBJ005.1 carrying the mScarlet-H-mNeonGreen tFT, I used the gap repair-cloning approach. In order to generate the amino acid change to generate the mScarlet-H from mScarlet, I used two overlapping primers which carried a histidine (CAC) instead of a methionine (ATG) with annealing regions around the position 164<sup>th</sup>. These primers were annealed following the conditions described in Table 8 and Table 9, this new dsDNA was used as an insert. To generate the vector for cloning, I PCR-amplified the plasmid pKBJ004.1 using fw and rv primers (6/14 and 6/15) which were excluding the amino acid position to be mutated (164<sup>th</sup>). The PCR-amplification of the insert was performed using Table 6 and Table 7, then it was gel-purified. Both insert and vector were transformed into BY4741 yeast competent cells.

After having a yeast strain carrying a plasmid with the new mScarlet-H variant, I extracted this plasmid and transformed *E.coli* DH10beta electrocompetent cells (Top10 C404052) by electroporation.

The plasmid pKBJ005.1 was validated by destination with the restriction enzymes EcoRV-HF and Sall-HF and it was Sanger sequenced with oligos 4/68 and 4/69.

**Table 8.** Solution mix used for generating dsDNA inserts from oligonucleotides from single-stranded oligonucleotides with complementary sequences. #NEB catalog # B0201S

Component	30 µL reaction
ddH <sub>2</sub> O	18 µL
fw primer (100 µM)	3 µL
rv primer (100 µM)	3 µL
PNK buffer (10x) <sup>#</sup>	3 µL
NaCl (0.5 M)	3 µL

**Table 9.** Conditions for the generation of dsDNA inserts from oligonucleotides from single-stranded oligonucleotides with complementary sequences.

Steps
1) Heat up mix to 100 °C for 10 min
2) Ramp down to 25 °C over 75 min
3) Use 0.5 µL of the annealed oligos/ insert for cloning with 70 ng of vector

## Evaluation of different tFTs as reporters for protein degradation

Using the plasmids carrying the different tFTs (Table 2), I generated yeast strains carrying X-tFT constructs. The amino acid residues (X) selected were: asparagine (N), aspartic acid (D), glutamic acid (E), tryptophan (W), arginine (R), histidine (H), isoleucine (I), methionine (M), valine (V) and threonine (T). In order to generate such yeast strains, I designed complementary Ubi-X fw and rv oligonucleotides, which annealed the Ubiquitin N-terminus region and to the C-terminus region of the e<sup>k</sup> sequence before the tFT. Each pair of oligos (fw and rv) contained each of the X residues (based on yeast codon usage) to be tested. Each fw and rv oligonucleotide pair was annealed using the procedure described above (Table 9).

In order to incorporate the different X residues into each tFT-construct, I digested overnight the plasmids carrying each Ubi-e<sup>k</sup>-tFT sequence (Table 2). Linearization of the plasmids was performed using the restriction enzyme EcoRV-HF, since there was a cutting site right at the C-terminus of the ubiquitin moiety. The generated vectors were gel-purified. The vectors and annealed oligonucleotides were transformed into yeast BY4741 (MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0) competent cells. As transformation control, we also used the backbone plasmid p413 cut and uncut into the same type of competent cells.

In order to select positive clones for each X-tFT yeast strain, we pinned the yeast strains on agar and used the plate reader to measure green and red fluorescence of eight different clones per X-tFT. I measured two technical replicates per X-tFT clone. As negative control I included the non-fluorescent strain that carried the p413 uncut plasmid; this control was arrayed interspersed with the rest of the clones along the plate. After measuring the green and red fluorescence, the average of the auto-fluorescence detected in the negative control colonies was subtracted to the read-out of the rest of the colonies measured. Based on the corrected green and red fluorescence value per colony, I selected two positive clones per X-tFT when ratios were consistent between replicates and with literature (Kats et al. 2018).

During the Ubi-X-tFT construction, we tried to generate two biological replicates per X-tFT. Nevertheless, for the next X-tFT cases we were able to construct only one biological replicate: mCh-mNG (N, W, R, H, I), mSI-mNG (N, D, W, R, T), mS-mNG (H, M, V), mSH-mNG (W, R, I). Per each X-tFT biological replicate we measured the fluorescence across a given number of technical replicates (n): constructs carrying X-mCh-sfGFP (n = 24), X-mCh-mNG (n = 16), X-mSI-mNG (n = 16), X-mS-mNG (n = 16) and X-mSH-mNG (n = 16). The settings used in the plate reader to measure fluorescence intensity are listed in Table 10.

For determining the fluorescence ratio (redFP/greenFP) for all Ubi-X-tFT constructs we used a plated reader (SparkControl 20M plate reader with monochromator, TECAN) that allowed us to measure fluorescence intensity in colonies arrayed in a 1536 format.

The excitation bandwidth used for all reFP was 10 nm, while for sfGFP and for mNG the excitation bandwidth used was 20 nm and 5 nm respectively. The emission bandwidth used for all reFP was 10 nm, while for sfGFP and for mNG the excitation bandwidth used was 25 nm and 5 nm respectively. For all red and greenFPs excitation and emission mode used was the Monochromator mode.

**Table 10.** Settings for the fluorescence intensity measurements of Ubi-X-tFT performed in the plate reader.

FT	Excitation wavelength (redFP)	Emission wavelength (redFP)	Excitation wavelength (greenFP)	Emission wavelength (greenFP)	Gain (redFP_greenFP)
mCherry-sfGFP	586 nm	612 nm	485	530	95 _ 30
mCherry-mNeonGreen	586 nm	612 nm	506 nm	524 nm	90 _ 90
mScarlet-mNeonGreen	569 nm	594 nm	506 nm	524 nm	85 _ 95
mScarlet-H-mNeonGreen	551 nm	592 nm	506 nm	524 nm	85 _ 95
mSCarlet-I-mNeonGreen	569 nm	593 nm	506 nm	524 nm	85 _ 95

For sfGFP the mirror setting used was Automatic (Dichroic 510), for the rest of red and greenFPs the mirror used was Automatic (50% Mirror). For all FPs the number of flashes was set to 10.

To further validate the selection of the two positive clones per X-tFT, I also performed colony PCR (Table 11) and sent the gel-purified fragment for Sanger sequencing.

**Table 11.** Colony PCR mix for the amplification of DNA fragments.

Component	40 $\mu$ L reaction
ddH <sub>2</sub> O	23.6 $\mu$ L
HF-buffer	4 $\mu$ L
Betaine (5 M)	4 $\mu$ L
MgCl <sub>2</sub> (100 mM)	1 $\mu$ L
DNA template (culture from patch)	1 $\mu$ L
dNTPs (10 mM)	1.6 $\mu$ L
forward (fw) primer 5 $\mu$ M	4 $\mu$ L
reverse (rv) primer 100 $\mu$ M	0.2 $\mu$ L
HF-Polymerase	0.6 $\mu$ L

High-fidelity (HF) buffer and Polymerase were acquired from the Protein production core facility in IMB.



**Table 12.** Colony PCR cycling conditions for the amplification of DNA fragments.

Step	Temperature (C) - Time (s)
Initial denaturation	98 - 300
Denaturation	98 - 30
Annealing	T <sub>m</sub> (primers)* - 30
Elongation	72 - 30 per kbp (go to denaturation 38 cycles)
Final elongation	72 - 300
Pause	4

For further characterization of the tFTs I also performed a western-blot experiment. One validated clone per X-tFT was grown in SC-His media at 30 °C overnight 220 rpm. Next day the OD600 was measured and samples were diluted back to an OD600 of 0.15 and were regrown at 30 °C and 220 rpm, after this dilution, both technical replicates per X-tFT construct were grown until their OD600 was above 1. I proceeded to do whole cell protein extraction and ran the extracted proteins in a 12% acrylamide SDS-PAGE (TGX™ FastCast™ Acrylamide Kit, 12%, BioRad 1610175). Primary antibodies used were mouse monoclonal anti-mNeonGreen (Chromotek) and mouse monoclonal anti-GFP (11814460001 Roche). I used the secondary antibody goat anti-Mouse IgG (H+L) Cross-Adsorbed (G-21040ThermoFisher). As loading control I used the Pgk1 protein, and for its identification I used the primary mouse monoclonal antibody anti-Pgk1 (459250 Invitrogen). For the detection of Pgk1, the same secondary mentioned above was used.

### Construction of genome-wide yeast libraries

For the donor plasmid type II construction (pKBJ011-1, pRS41K-mScarlet-I-mNeonGreen-ADH1term-TEFpr-hphΔC) I linearized the plasmid pMaM484 with BamHI-HF & SpeI-HF and gel-purified fragment of 6711bp. Then I PCR-amplified the mScarlet-I-mNeonGreen tag from the plasmid pKBJ003.1 with oligos 8/100 and 9/2. After gel-purification of the insert, I assembled the vector and insert using the Gibson method and transformed the construct into DH5α (C2987 NEB) cells. Validation of the new donor plasmid was achieved by restriction enzyme digestion with BamHI-HF and SpeI-HF and by sequencing the tFT presence and joints of the cloning with the oligos 6/73 and 9/6.

For the donor strain I used the genetic background of yMaM1205 (MAT<sub>α</sub> lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 can1Δ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI). In order to get two strains, one as negative control and one as donor strain to cross to the C-SWAT library, with the same cytosolic FP as marker but different intensity I decided to use the BFP2 FP. I generated a series of integrative plasmids where the BFP2 was expressed under different promoter and terminator combinations. Each integrative plasmid was linearized with ZraI and HpaI enzymes, the transformation cassette was gel-purified and transformed into competent yeast cells of the strain yMaM1205. All

cassettes were integrated into the *leu2Δ* locus. The strains expressing the cytosolic BFP2 with different intensities are listed in Table 13.

**Table 13.** Strains expressing different cytosolic markers

Strain ID	Cytosolic marker
yKBJ0113	ADH1pr-3xBFP2-ADH1ter
yKBJ0114	GPDpr-BFP2-CYC1ter
yKBJ0115	GPDpr-BFP2-ADH1ter
yKBJ0116	TEF1pr-BFP2-CYC1ter
yKBJ0117	TEF1pr-BFP2-ADH1ter
yKBJ0120	GPDpr-3xBFP2-T8ter
yKBJ0121	TEF1pr-3xBFP2-T8ter
yKBJ0087	GPDpr-3xBFP2-CYC1ter
yKBJ0088	GPDpr-3xBFP2-ADH1ter
yKBJ0089	TEF1pr-3xBFP2-CYC1ter
yKBJ0090	TEF1pr-3xBFP2-ADH1ter
yKBJ0092	NOP1pr-3xBFP2-ADH1ter

All strains in Table 13, were validated by growth on YPD + Nat and by confocal fluorescence microscopy.

The final donor strain was constructed by transforming competent cells of the strain yKBJ0150, which was expressing the dim blue cytosolic marker and the bright septin tag, with the donor plasmid containing the mScarlet-I-mNeonGreen tFT (pKBJ011).

This donor strain was named yKBJ0191-1 (id 1620) and had the next genotype: yMaM1205 (MAT $\alpha$  lyp1 $\Delta$  his3 $\Delta$ 1 leu2 $\Delta$ 0 ura3 $\Delta$ 0 met15 $\Delta$ 0 can1 $\Delta$ ::STE3pr-LEU2-GAL1pr-NLS-I-SCEI) Shs1-3xyomtagBFP2-His3 leu2 $\Delta$ ::natNT2-GPD-yomtagBFP2-CYC1ter pRS41K-mScarlet-I-mNeonGreen-ADH1term-TEFpr-hph $\Delta$ C (C-SWAT typell donor plasmid).

The final yeast strain selected as negative control for microscopy experiments was named yKBJ0149 (id 1421), and it was expressing the brightest blue cytosolic marker tested (GPDpr-3xBFP2-ADH1ter).

In order to perform high-throughput swapping of the SWAT module in the C-SWAT library (Meurer et al. 2018), I used the synthetic genetic array (SGA) method (A. H. Tong et al. 2001; Baryshnikova et al. 2010). For the SGA we crossed a donor strain carrying the tFT tag as well as segmentation markers and selection markers with the C-SWAT library. Through a series of replica-pinning on different selection media, we

generated a new C-tagged library with our desired tag. The genetic crossing was semi-automated, using the pinning robot (Singer Instruments). The SGA steps I followed for the construction of the ORF-mScarlet-I-mNeonGreen-Hyg library (C-mSI-mNG-hphNT1) were the following:

1. Grow the C-SWAT library on SC-Ura and grow the donor strain on YPD+G418 on 384-colony array format
2. Mate the C-SWAT library and the donor strain on YPD
3. Select for diploids on SC(MSG)-Ura+G418 plates. Repeat the selection twice
4. Pin the diploids on Sporulation media and let grow for 7 days at 23 °C
5. Select for haploids:
  - Haploid selection 1: SC(MSG)-Leu/Arg/Lys/Ura + Can/Thi
  - Haploid selection 2: SC(MSG)-Leu/Arg/Lys/Ura + Can/Thi/G418
  - Haploid selection 3: SC(MSG)-Leu/Arg/Lys/Ura + Can/Thi/G418/Nat
  - Haploid selection 4: SC(MSG)-Leu-His/Arg/Lys/Ura + Can/Thi/G418/Nat (repeat twice)
6. Induction of I-SceI enzyme for tag swapping on SC-LEU + Raf/GAL. Repeat the induction twice
7. Counter selection on SC-Leu + 5 FOA against strains carrying the acceptor module
8. Selection for positive strains carrying the new mScarlet-I-mNeonGreen-Hyg on SC(MSG) -Leu/Arg/Lys +Can/Thi/Hyg
9. Inoculate 96-well plates and 384-well plates (Greiner Bio-One) with 200 µL and 50 µL of SC(MSG)-Leu/Arg/Lys+Hyg+Gly(15%) media respectively. Seal with air-o-seals and grow strains at 30 °C in a standing incubator for 2 days. For these two libraries, two copies of each were stored one in 96-format and the second one in 384-format at - 80 °C

In order to validate the presence of the tag in the final library, I performed fluorescent measurements of 21 strains expressing proteins known for their high abundance in yeast and which have previously been used for tagging efficiency validation in other studies (Meurer et al. 2018). I measured the red and green fluorescence of these strains using flow cytometry and their tagging efficiency was scored as the percentage of cells with fluorescence in each fluorescent population. The settings for the measurements in flow cytometry are described in the respective protocol below.

**Note:** Take photographs of all of the plates in several steps along the SGA. For sure take an initial photograph of N and C-SWAT libraries and a final photograph of the libraries on SC(MSG) -Leu/Arg/Lys +Can/Thi/Hyg plates.

## Measuring fluorescence signal using Flow cytometry

I used a flow cytometer (BD LSRFortessa), and its acquisition software BD FACSDiva v9.0.1, for the quantification of the percentage of fluorescent signals in yeast cells.

The flow cytometer was used together with its auto-sampler for 96-well plate format in the high-throughput (HTS) mode. The settings for the loader were as follows: sample flow rate 0.5  $\mu\text{L}/\text{s}$ , sample volume 25  $\mu\text{L}$ , mixing volume 100  $\mu\text{L}$ , mixing speed 100  $\mu\text{L}/\text{s}$ , number of mixes 3 and wash volume 800  $\mu\text{L}$ . The gating of the fluorescent populations was defined manually. Firstly, the whole cell population was filtered based on cell size by using the forward and side scatter (FSC and SSC) to distinguish from double cells and single cells. In the case of the genome-wide C-tagged tFT library, the fluorescent populations were defined manually in four gates: non-fluorescent gate, cells with green fluorescence, cells with red fluorescence and double positive cells (DP which had as we expected both green and red fluorescence). Per strain, 10000 cells were measured in the flow cytometer.

The protocol to prepare cells for fluorescent measurement in the mentioned flow cytometer is as follows:

1. Under sterile conditions fill up the 96-well plate (PhenoPlate 96-well microplates, PerkinElmer) wells needed with 150  $\mu\text{L}$  of the required medium
2. Seal the plate with an air-o-seal. Close it with its plastic lid, place it in the incubator at 30 °C and 220 rpm. Add a box's lid on top of it to prevent evaporation. Grow to saturation (overnight)
3. Next day, dilute the saturated cultures and grow them for 6 hours to  $6\text{-}8 \times 10^6$  cells/mL
4. Perform the quality check of the lasers performance in the flow cytometer, prime the HTS plate-sampler and proceed to measure the Single-cell fluorescence intensities by the 488 and 561 nm lasers, for the red and green fluorescence respectively

**Note 1:** For the genome-wide high-throughput flow cytometry, I used SC low fluorescence medium (6.7 g/ L bacto yeast nitrogen base without amino acids and without folic acid and riboflavin (LoFlo, Formedium), 2 g/ L amino acid dropout mix, 2% (w/v) glucose) supplemented with 300 mg/L adenine (A8626, Sigma-Aldrich).

## Budding index

Through budding index (BI) determination one can evaluate how a given strain is progressing through the cell cycle (Muñoz-Barrera and Monje-Casas 2017). The BI corresponds to the fraction of budded cells in a given population. Therefore, using a mechanical counter one can visually and manually determine the BI as follows:

1. Grow yeast cultures and prepare them for confocal fluorescence microscopy
2. Acquire brightfield and any fluorescence channel required for yours strains

3. Count, aided by the mechanical counter, the number of budded cells. Count the total number of cells in a given field of view (fov) as well
4. Calculate the BI per fov by dividing the budded cells by the total cell number

**Note 1:** Depending on the aim of the experiment, one can make finer classifications and calculate the number of small-budded, large-budded or re-budded cells in the population, for example.

### Alpha-factor G1 synchronization

To more precisely determine the cell-cycle stage of a given cell population, one can chemically synchronize the population using the mating pheromone alpha-factor ( $\alpha$ -factor), which arrests the cells in G1 and M/G1 transition (Rosebrock 2017; Neymann, Wegerhoff, and Engell 2011).

Stock solutions:

Alpha-factor stock solution should be 1 mg/ml in DMSO. Dissolve the alpha-factor powder in DMSO by pipetting. Store in aliquots of 5 mL at -20 °C. For longer storage, keep aliquots at -80 °C. In this study I used the  $\alpha$ -factor by GenScript, RP01002.

DAPI stock solution should have a concentration of at 1mg/ mL in ddH<sub>2</sub>O. Dissolve 1 mg of DAPI powder into 1mL of ddH<sub>2</sub>O, by vortexing and if necessary place the tube of the solution at 60 °C in a water bath for 5 min. Store at -20 °C protected from the light, use an amber flask or aluminum foil.

PBS:DAPI stock solution should be prepared in a ratio of 1:10 000. To prepare 10 mL of PBS:DAPI, use 10 mL of 1xPBS and 1uL of DAPI stock solution. Mix by vortexing and store in a 15mL Falcon tube covered with aluminum foil at 4 °C.

Alpha-factor working concentration:

For *bar1* $\Delta$  strains, use a final concentration of 1  $\mu$ g/ml (1:1000 dilution). For *BAR1* strains, use a final concentration of 10  $\mu$ g/ml (1:100 dilution).

Incubation time:

Alpha-factor is a mating pheromone, so MatA cells will arrest in G1-phase before START. Allow 1.5 doubling times for proper arrest. Check for mating project (shmoos) under microscope about 15 min before the estimated end point. The arrest is considered good when more than 90 % of cells have shmoos. If necessary, incubate further and check in 15 min intervals.

Normally, for cells growing at 30 °C in YPAD, start checking if cells are arrested at 2 h 15 min, maximum: 2 h 30 min. For cells growing at 23 °C in YPAD, look for arrested cells after 2 h 30 min, maximum 3h. In my case, my cells were *BAR1*, grown at 30 °C in SC+Adenine, and were arrested after 2 h15 min.

## Procedure:

**1.** Two days before experiment: set up a pre-culture in about 10 ml medium. Use filter-sterilized medium if doing microscopy. Autoclaved medium drastically enhances the background fluorescence for GFP, RFP, CFP, YFP. DAPI staining is fine in autoclaved medium.

**2.** One day before experiment: If using mutants or a totally unusual medium, determine the doubling time of the culture by diluting a bit of the pre-culture in fresh medium in the morning and following the increase in OD600 over time.

In the late afternoon/ evening, dilute pre-culture in the volume necessary for the experiment to have in the next morning an OD600 of about 0.2 - 0.5.

The cell density of the culture at the starting point, just before adding alpha factor, is an important factor. If the cell density is too high, cells might be in the stationary phase already, the percentage of arrested cells will decrease.

**Note 1:** if the experiment involves only microscopy, an OD600 of about 0.2 works best.

**Note 2:** If doing biochemistry (samples for TCA), start with an OD600 of about 0.5 - 0.6 as in this case more cells will be necessary for making protein extracts.

**3.** Day of experiment: Measure the OD600 of the cultures.

If the OD is higher than expected but lower than 1: dilute cultures back to OD 0.2 - 0.5 and add alpha-factor straightaway.

If the OD is higher than 1: dilute cultures back to OD 0.15 and allow cells to grow for 1 doubling time. Measure OD again and if OK add alpha-factor.

Incubate for 1 ½ doubling time. Check shmoo formation under a microscope.

If more than 90 % of cells have shmoo, transfer cell suspensions to 50 mL Falcon tubes. Centrifuge at 3200 rpm for 2 min at room temperature.

Discard supernatant, gently or use a pipette to get rid of the last part of the supernatant, and wash cells twice with each 50 mL fresh medium (without alpha-factor and pre-warmed at 30 °C if wt strains) by re-suspending cell pellets in medium and centrifuging as before.

Resuspend cells in the required pre-warmed medium and start the time course (t = 0). Grow the culture in flasks in a shaking incubator and sample every 15 min.

**Note:** you should work as quickly as possible when washing alpha-factor away. Cells will reassume growth once alpha-factor concentration gets down. Place a note on the centrifuge to avoid other users while you are releasing your cells. Additionally, prepare all flasks and label all the tubes before you start washing.

**4.** Post-processing of the collected samples: For DAPI staining, use 300 µL of culture and add 700 µL of 100 % ethanol. Vortex briefly to separate the cell conglomerates

and keep on ice. This becomes 70 % EtOH, samples in 70 % Ethanol can be kept for weeks at 4 °C.

Just before inspection, centrifuge cell suspension at 6000 rpm, 2 min and resuspend the cell pellet in PBS-DAPI (PBS + DAPI stock diluted 1:10.000, can be kept at 4 °C, wrapped up in aluminum foil).

The volume of PBS depends on the pellet size, normally 10 - 30 µL PBS-DAPI are fine. In my case I used 30 uL and I mixed it with the pellet by softly pipetting up and down.

**5. Microscopy using Opera Phenix confocal mode:** in advance, coat the necessary wells of a 96-well Perkin Elmar imaging plate (Cell carrier, 96-well format 6055302 or 384-well format 6057300) with 70 uL of Concanavalin A (ConA, Sigma-Aldrich) at 0.1 mg/ mL, let rest for 30 min at RT. The Concanavalin is dissolved in ddH<sub>2</sub>O, and stored at 4 °C in aliquots of 1.5 mL. Before using it, spin at 16 000 x g for 5 min at 4 °C to remove precipitates.

After coating with Concanavalin A for 30 min, pipette it out of each well and store it again in new Eppendorf tubes at 4 °C. ConA can be reused up to the needs of each person. Add 100 uL of ddH<sub>2</sub>O to each well and remove it by aspiration.

Add immediately the 30 uL of cell pellet dissolved in PBS-DAPI and add 40 uL - 80 uL of SC+Ade media (or the required media type) to dilute the sample and have a nice density of cells in the microscopy images. This works if the OD600 was approximately around 0.2 when the alpha-factor was added.

Acquire Brightfield to identify cells and DAPI channel to detect nuclei.

**Note:** For this study, I classified cells in different cell stages as follows: “ 1 DAPI signal and unbudded” corresponded to G1 phase, “1 DAPI signal and budded cell” corresponded to S stage and “2 DAPI signals and budded cell” was considered as G2/M transition stage.

### Growth curve

In this study we used a plate reader (SparkControl 20M plate reader with monochromator, TECAN) to generate all growth curves. Steps are listed next:

**1.** Inoculate your strains of interest in a 24-well plate, use 1 mL of media per well. Grow at 30 °C and 220 rpm overnight or until saturation

**2.** Measure the OD600 of the saturated culture and dilute to an OD600 of 0.02 per well

**Note 1:** When doing a growth curve with Calcofluor in the media, add the dye in this diluting step. For example, add the volume of saturated culture calculated, then add 333.3 µL of the corresponding Calcofluor stock solution and add the remaining volume of the corresponding media.

**Note 2:** For microscopy: 50 µL Calcofluor solution/ 150 µL final volume = 0.333 dilution factor.

3. Grow at 30 °C shaking at 510 rpm in the plate reader, with lid. Measure OD600 every 5 min for up to 24 h

4. Analyze the growth curve by plotting time vs OD600. Subtract the OD600 value of the media to the OD600 value of each well with sample

**Note:** In order to find the optimal Calcofluor concentration to use for staining the cells before microscopy I calculated what the Calcofluor final concentration would be in a microscopy well. In a 96-well plate well the final volume that we will use for microscopy is 150  $\mu$ L, from which 100  $\mu$ L are SC+Ade media and the rest 50  $\mu$ L will be of a Calcofluor solution. Which meant, for microscopy 50  $\mu$ L Calcofluor solution/ 150  $\mu$ L final volume per 96-well plate well equals a dilution factor of 0.333.

**Table 14.** Calcofluor concentrations tested.

<b>Calcofluor stock solutions</b>	<b>Calcofluor final concentration for microscopy</b>
1000 $\mu$ M (1 mg/ mL)	333.33 $\mu$ M
100 $\mu$ M (0.1 mg/ mL)	33.33 $\mu$ M
25 $\mu$ M (0.03 mg/ mL)	8.3 $\mu$ M
5 $\mu$ M (0.005 mg/ mL)	1.6 $\mu$ M
0.05 $\mu$ M (0.005 mg/ mL)	0.016 $\mu$ M (0.0016 mg/ mL)

### **Spotting assays**

1. Grow yeast strains in liquid media until saturation (overnight), 30 °C and 220 rpm

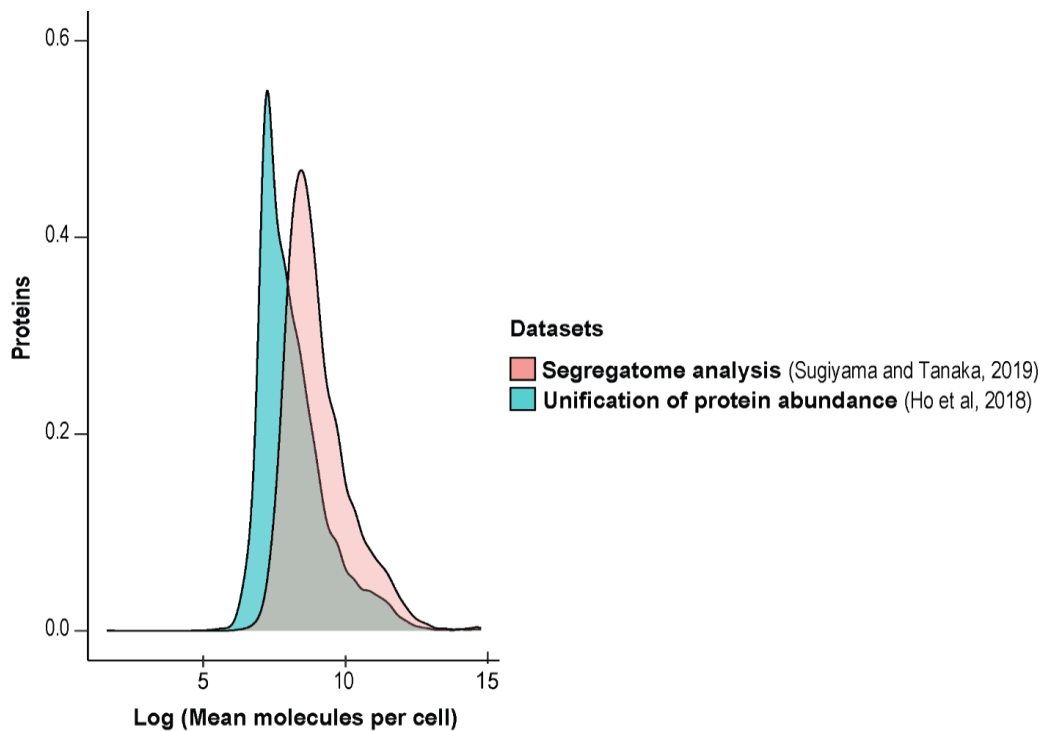
2. Prepare a series of 10 fold dilutions of the strains  
When using a 96-well plate: start with a well with 200  $\mu$ L of saturated culture in 0  $\mu$ L of media, then take 15  $\mu$ L of the first well and use this to inoculate a new well containing 135  $\mu$ L of media, repeat these steps per dilution

3. From each well, use 2.5  $\mu$ L of diluted culture to spot on the agar plate, using a pipette

4. Let grow for 24 h and take a photograph, you can repeat this step as many time as consider necessary to see a growth difference or to conclude there is no difference upon temperature treatment

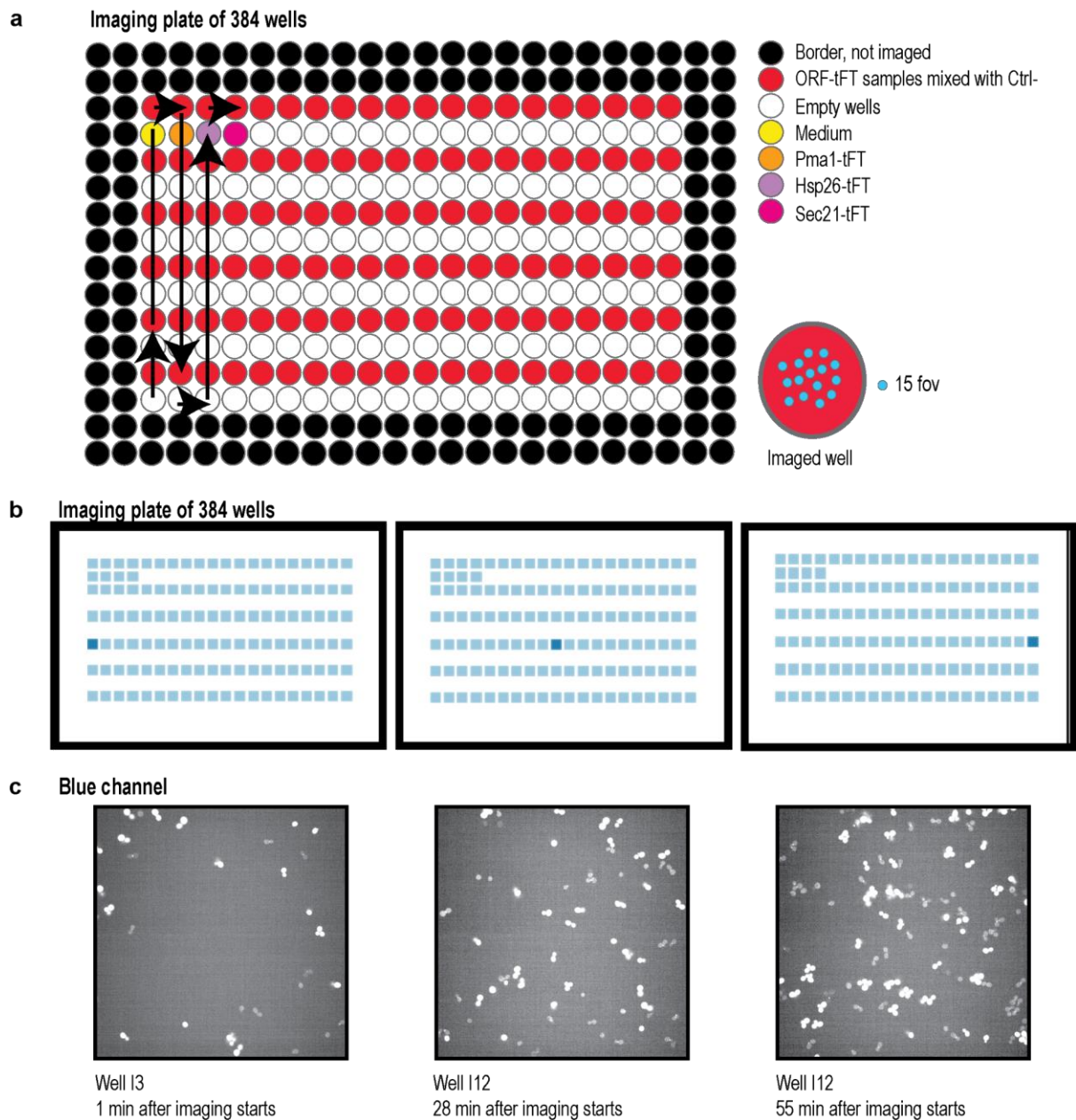


## SUPPLEMENTARY



**Figure S1.** Protein abundance compared in two different genome-wide studies in budding yeast.

In Figure S1, we see the distribution of proteins found by the “Segregatome analysis” based on their mean molecules per cell (Sugiyama and Tanaka 2019). In this “Segregatome study” they classified proteins based the protein segregation of old and newly synthesized protein between mother and bud (Sugiyama and Tanaka 2019). On the other hand, the study termed “Unification of protein abundance” (Ho, Baryshnikova, and Brown 2018) unified 21 different quantitative studies under the same unit, ending up with a total of 5748 proteins for which their protein abundance is now reported as mean molecules per cell.



**Figure S2.** Different cell density across wells of the same microscopy plate.

**a.** The established high-throughput imaging procedure for live cells allows to image a 384-well plate in 1 h. This while imaging three different fluorescent channels, a brightfield mode, imaged 15 fov per well in single plane mode. The imaging order is indicated with black arrows. **b and c.** In live cell microscopy, the yeast cells kept on reproducing along the 1 h of imaging per plate, therefore we end up having lower cell density at the beginning of the plate compared to later wells in the imaging order.

**Table S3.** List and characteristics of the selected fast-maturing green fluorescent proteins and the slow-maturing red fluorescent proteins.

Fluorescent protein	Maturation time (min)	Brightness	Excitation wavelength (nm)	Emission wavelength (nm)
sfGFP	13.6	54.15	485	510
mNeonGreen	10	92.8	506	517
mCherry	40#	15.84	587	610
mScarlet	174	70	569	593
mScarlet-H	264	14.8	551	592
mScarlet-I	36	56.16	569	593

Source: sfGFP (Pédélec et al. 2006), mCherry (Merzlyak et al. 2007) mNeonGreen (Shaner et al. 2013; Cranfill et al. 2016) and mScarlet, mScarlet-H and mScarlet-I (Bindels et al. 2017).

**Table S4.** List of plasmids carrying different tFTs for C-terminus tagging of POIs.

Plasmid-ID	tFT	slow-maturing FP	fast-maturing FP
pKBJ001	mCherry-sfGFP	mCherry	sfGFP
pKBJ002.1	mCherry-mNeonGreen	mCherry	mNeonGreen
pKBJ003.1	mSCarlet-I-mNeonGreen	mSCarlet-I	mNeonGreen
pKBJ004.1	mScarlet-mNeonGreen	mScarlet	mNeonGreen
pKBJ005.1	mScarlet-H-mNeonGreen	mScarlet-H	mNeonGreen

**Table S5.** Donor plasmids for C-terminus and N-terminus tagging under endogenous (Type I) and marker reconstitution (Type II) approaches, carrying the mScarlet-I-mNeonGreen tFT.

<b>Plasmid-ID</b>	<b>tFT</b>	<b>Tagging type</b>
pKBJ010	pRS41K-mScarlet-I-mNeonGreen	C-terminus Seamless (type I)
pKBJ011	pRS41K-mScarlet-I-mNeonGreen-ADH1term-TEFpr-hphΔC	C-terminus Hyg rec (type II)
pKBJ012	pSD-N1-L1-mNeonGreen-mScarlet-I-L2	N-terminus Seamless (type I)
pKBJ013	pSD-N4-hpxdC-GPDpr-mNeonGreen-mScarlet-I-L2	N-terminus Hyg rec (type II)
pKBJ014	pSD-N4-hpxdC-TEF1pr-mNeonGreen-mScarlet-I-L2	N-terminus Hyg rec (type II)
pKBJ015	pSD-N4-hpxdC-NOP1pr-mNeonGreen-mScarlet-I-L2	N-terminus Hyg rec (type II)
pKBJ016	pSD-N7-hpxdC-TEF2pr-SP-mNeonGreen-mScarlet-I-L2	N-terminus Hyg rec (type II)

Hygromycin reconstitution (Hyg rec) as marker reconstitution type II high-throughput tagging approach.

## **Appendix I**

### **List of abbreviations**

BF: brightfield

bp: base pair

ConA: Concanavalin A

dsDNA: double stranded DNA

EtOH: ethanol

Fov: field of view

FP: fluorescent protein

HTS: high-throughput

kbp: kilobase pair

kDa: kilodalton

m: slope

min: minutes

mCh: mCherry

mL: milliliters

mS: mScarlet

mSH: mScarlet-H

mSI: mScarlet-I

ng: nanograms

ng/  $\mu$ L: nanogram per microliters

ORF: open reading frame

PCR: polymerase chain reaction

POI: protein of interest

pr: promoter

RE(s): restriction enzyme(s)

RT: room temperature

ter: terminator

tFT: tandem fluorescent protein timer

TMD: trans-membrane domain

s: second

sd: standard deviation

SP: Signal peptide

SPB: spindle pole body

$\mu\text{L}$ : microliters

## **REFERENCES**

Aguilaniu H, Gustafsson L, Rigoulet M, & Nyström T. (2003). Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science*, 299(5613), 1751-1753. <https://doi.org/10.1126/science.1080418>.

Bachmair A, Finley D, & Varshavsky A. (1986). In vivo half-life of a protein is a function of its amino-terminal residue. *Science*, 234(4773), 179-186. <https://doi.org/10.1126/science.3018930>.

Bachmair, A., & Varshavsky, A. (1989). The degradation signal in a short-lived protein. *Cell*, 56(6), 1019–1032. [https://doi.org/10.1016/0092-8674\(89\)90635-1](https://doi.org/10.1016/0092-8674(89)90635-1).

Bachman, J. (2013). Site-directed mutagenesis. *Methods in Enzymology*, 529, 241–248. <https://doi.org/10.1016/B978-0-12-418687-3.00019-7>.

Ball, D. A., Mehta, G. D., Salomon-Kent, R., Mazza, D., Morisaki, T., Mueller, F., McNally, J. G., & Karpova, T. S. (2016). Single molecule tracking of Ace1p in *Saccharomyces cerevisiae* defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Research*, 44(21), e160. doi: 10.1093/nar/gkw744. PMCID: PMC5137432. PMID: 27566148

Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F. J., Myers, C. L., Andrews, B., & Boone, C. (2010). Synthetic genetic array (SGA) analysis in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Methods in Enzymology*, 470, 145–179. [https://doi.org/10.1016/S0076-6879\(10\)70007-0](https://doi.org/10.1016/S0076-6879(10)70007-0).

Bindels DS, Haarbosch L, van Weeren L, Postma M, Wiese KE, Mastop M, Aumonier S, Gotthard G, Royant A, Hink MA, Gadella TW Jr. (2017). mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nature Methods*, 14(1), 53-56. <https://doi.org/10.1038/nmeth.4074>. PMID: 27869816.

Bitto, A., Wang, A. M., Bennett, C. F., & Kaeberlein, M. (2015). Biochemical genetic pathways that modulate aging in multiple species. *Cold Spring Harbor Perspectives in Medicine*, 5(11). <https://doi.org/10.1101/cshperspect.a025114>.

Brachmann, C. B., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P., & Boeke, J. D. (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(2), 115–132. PMID: 9483801.

Busso, D., Kim, R., & Kim, S. H. (2004). Using an *Escherichia coli* cell-free extract to screen for soluble expression of recombinant proteins. *Journal of Structural and*

Campisi, J., & d'Adda di Fagagna, F. (2007). Cellular senescence: When bad things happen to good cells. *Nature Reviews Molecular Cell Biology*, 8(9), 729–740.

Chudakov, D. M., Matz, M. V., Lukyanov, S., & Lukyanov, K. A. (2010). Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews*, 90(3), 1103-1163. doi: 10.1152/physrev.00038.2009. PMID: 20664080.

Costello, J. L., Castro, I. G., Camões, F., Schrader, T. A., McNeall, D., Yang, J., Giannopoulou, E. A., Gomes, S., Pogenberg, V., Bonekamp, N. A., Ribeiro, D., Wilmanns, M., Jedd, G., Islinger, M., & Schrader, M. (2017). Predicting the targeting of tail-anchored proteins to subcellular compartments in mammalian cells. *Journal of Cell Science*, 130(9), 1675-1687. <https://doi.org/10.1242/jcs.200204>.

Cranfill, P. J., Sell, B. R., Baird, M. A., Allen, J. R., Lavagnino, Z., de Gruiter, H. M., Kremers, G. J., Davidson, M. W., Ustione, A., & Piston, D. W. (2016). Quantitative assessment of fluorescent proteins. *Nature Methods*, 13(7), 557-562. doi: 10.1038/nmeth.3891. PMCID: PMC4927352. PMID: 27240257.

Després, P. C., Dubé, A. K., Seki, M., Yachie, N., & Landry, C. R. (2020). Perturbing proteomes at single residue resolution using base editing. *Nature Communications*, 11(1), 1–13.

Donà, E., Barry, J., Valentin, G., Quirin, C., Khmelinskii, A., Kunze, A., Durupt, F. C., et al. (2019). Directional tissue migration through a self-generated chemokine gradient. *Nature*, 503(7475), 285–289.

Durrieu, L., Kirrmaier, D., Schneidt, T., Kats, I., Raghavan, S., Hufnagel, L., Saunders, T. E., & Knop, M. (2018). Bicoid gradient formation mechanism and dynamics revealed by protein lifetime analysis. *Molecular Systems Biology*, 14(9), e8355. <https://doi.org/10.15252/msb.20188355>.

Eldakak, A., Rancati, G., Rubinstein, B., et al. (2010). Asymmetrically inherited multidrug resistance transporters are recessive determinants in cellular replicative ageing. *Nature Cell Biology*, 12(8), 799-805. <https://doi.org/10.1038/ncb2085>.

England, C. G., Luo, H., & Cai, W. (2015). HaloTag technology: A versatile platform for biomedical applications. *Bioconjugate Chemistry*, 26(6), 975-986. <https://doi.org/10.1021/acs.bioconjchem.5b00191>.

Erjavec, N., Larsson, L., Grantham, J., & Nyström, T. (2007). Accelerated aging and failure to segregate damaged proteins in Sir2 mutants can be suppressed by



overproducing the protein aggregation-remodeling factor Hsp104p. *Genes & Development*, 21(19), 2410-2421. <https://doi.org/10.1101/gad.439307>.

Fontana, L., et al. (2010). Extending Healthy Life Span—From Yeast to Humans. *Science*, 328, 321-326. <https://doi.org/10.1126/science.1172539>.

Fung, J. J., Blöcher-Juárez, K., & Khmelinskii, A. (2022). High-Throughput Analysis of Protein Turnover with Tandem Fluorescent Protein Timers. *Methods in Molecular Biology*, 2378, 85-100. [https://doi.org/10.1007/978-1-0716-1732-8\\_6](https://doi.org/10.1007/978-1-0716-1732-8_6).

Ganley, A. R., Ide, S., Saka, K., & Kobayashi, T. (2009). The effect of replication initiation on gene amplification in the rDNA and its relationship to aging. *Molecular Cell*, 35(5), 683-693. <https://doi.org/10.1016/j.molcel.2009.07.012>.

Gauss, R., Trautwein, M., Sommer, T., & Spang, A. (2005). New modules for the repeated internal and N-terminus epitope tagging of genes in *Saccharomyces cerevisiae*. *Yeast*, 22(1), 1-12. <https://doi.org/10.1002/yea.1187>

Ghaemmaghami, S., Huh, W. K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., & Weissman, J. S. (2003). Global analysis of protein expression in yeast. *Nature*, 425(6959), 737-741. <https://doi.org/10.1038/nature02046>

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A. 3rd, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), 343-345. <https://doi.org/10.1038/nmeth.1318>

Gilbert, S. F. (2000). *Aging: The Biology of Senescence*. In *Developmental Biology* (6th ed.). Sunderland, MA: Sinauer Associates. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK10041/>

Gou, L., Bloom, J. S., & Kruglyak, L. (2019). The Genetic Basis of Mutation Rate Variation in Yeast. *Genetics*, 211(2), 731-740. <https://doi.org/10.1534/genetics.118.301609>.

Gross, L. A., Baird, G. S., Hoffman, R. C., & Tsien, R. Y. (2000). The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences*, 97(22), 11990-11995. <https://doi.org/10.1073/pnas.97.22.11990>.

Hartwell, L. H., & Unger, M. W. (1977). Unequal division in *Saccharomyces cerevisiae* and its implications for the control of cell division. *Journal of Cell Biology*, 75(2 Pt 1), 422-435. <https://doi.org/10.1083/jcb.75.2.422>.

Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*, 37, 614-636. doi: 10.1016/0014-4827(65)90211-9. PMID: 14315085.

Helfand, S. L., & Rogina, B. (2003). Genetics of aging in the fruit fly, *Drosophila melanogaster*. *Annual Review of Genetics*, 37, 329-348. <https://doi.org/10.1146/annurev.genet.37.040103.095211>. PMID: 14616064.

Higuchi-Sanabria, R., Pernice, W. M., Vevea, J. D., Alessi Wolken, D. M., Boldogh, I. R., & Pon, L. A. (2014). Role of asymmetric cell division in lifespan control in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 14(8), 1133-1146. <https://doi.org/10.1111/1567-1364.12216>. PMID: 25263578; PMCID: PMC4270926.

Ho, B., Baryshnikova, A., & Brown, G. W. (2018). Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. *Cell Systems*, 6(2), 192-205.e3. doi: 10.1016/j.cels.2017.12.004. PMID: 29361465.

Houtkooper, R. H., Williams, R. W., & Auwerx, J. (2010). Metabolic networks of longevity. *Cell*, 142(1), 9-14. doi: 10.1016/j.cell.2010.06.029. PMID: 20603007; PMCID: PMC3852811.

Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., & O'Shea, E. K. (2003). Global analysis of protein localization in budding yeast. *Nature*, 425(6959), 686-691. doi: 10.1038/nature02026. PMID: 14562095.

Iwase, M., Luo, J., Bi, E., & Toh-e, A. (2007). Shs1 plays separable roles in septin organization and cytokinesis in *Saccharomyces cerevisiae*. *Genetics*, 177(1), 215–229.

Janssens, G. E., & Veenhoff, L. M. (2016). Evidence for the hallmarks of human aging in replicatively aging yeast. *Microbial Cell Factories*, 15(1), 263–274.

Kats, I., Khmelinskii, A., Kschonsak, M., Huber, F., Knieß, R. A., Bartosik, A., & Knop, M. (2018). Mapping degradation signals and pathways in a eukaryotic N-terminome. *Molecular Cell*, 70(3), 488–501.e5.

Kennedy, B. K., Austriaco Jr, N. R., & Guarente, L. (1994). Daughter cells of *Saccharomyces cerevisiae* from old mothers display a reduced life span. *The Journal of Cell Biology*, 127(6 Pt 2), 1985–1993.

Kennedy, B. K., & Pennypacker, J. K. (2014). Drugs that modulate aging: The promising yet difficult path ahead. *Translational Research: The Journal of Laboratory and Clinical Medicine*, 163(5), 456–465.

Kenyon, C. (2010). The genetics of ageing. *Nature*, *464*, 504–512. <https://doi.org/10.1038/nature08980>.

Khmelinskii, A., Blaszczyk, E., Pantazopoulou, M., Fischer, B., Omnus, D. J., Le Dez, G., ... & Schiebel, E. (2014). Protein quality control at the inner nuclear membrane. *Nature*, *516*(7531), 410–413.

Khmelinskii, A., Keller, P. J., Bartosik, A., Meurer, M., Barry, J. D., Mardin, B. R., ... & Knop, M. (2012). Tandem fluorescent protein timers for in vivo analysis of protein dynamics. *Nature Biotechnology*, *30*(7), 708–714.

Khmelinskii, A., & Knop, M. (2014). Analysis of protein dynamics with tandem fluorescent protein timers. *Methods in Molecular Biology*, *1174*, 195–210.

Khmelinskii, A., Meurer, M., Duishoev, N., Delhomme, N., & Knop, M. (2011). Seamless gene tagging by endonuclease-driven homologous recombination. *PLoS One*, *6*(8), e23794.

Khmelinskii, A., Meurer, M., Ho, C.-T., Besenbeck, B., Füller, J., Lemberg, M. K., ... & Mogk, A. (2016). Incomplete proteasomal degradation of green fluorescent proteins in the context of tandem fluorescent protein timers. *Molecular Biology of the Cell*, *27*(2), 360–370.

Lai, C.-Y., Jaruga, E., Borghouts, C., & Jazwinski, S. M. (2002). A mutation in the ATP2 gene abrogates the age asymmetry between mother and daughter cells of the yeast *Saccharomyces cerevisiae*. *Genetics*, *162*(1), 73–87.

Lambert, T.J. (2019). FPbase: a community-editable fluorescent protein database. *Nature Methods*, *16*(3), 277–278. <https://doi.org/10.1038/s41592-019-0352-8>.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, *153*(6), 1194–1217.

López-Otín, C., Galluzzi, L., Freije, J. M. P., Madeo, F., & Kroemer, G. (2016). Metabolic control of longevity. *Cell*, *166*(4), 802–821.

McFaline-Figueroa, J. R., Vevea, J., Swayne, T. C., Zhou, C., Liu, C., Leung, G., ... & Pon, L. A. (2011). Mitochondrial quality control during inheritance is associated with lifespan and mother-daughter age asymmetry in budding yeast. *Aging Cell*, *10*(5), 885–895.

Merzlyak, E. M., Goedhart, J., Shcherbo, D., Bulina, M. E., Shcheglov, A. S., Fradkov, A. F., ... & Chudakov, D. M. (2007). Bright monomeric red fluorescent protein with an

extended fluorescence lifetime. *Nature Methods*, 4(7), 555-557. <https://doi.org/10.1038/nmeth1062>. PMID: 17572680.

Meurer, M., Duan, Y., Sass, E., Kats, I., Herbst, K., Buchmuller, B. C., ... & Knop, M. (2018). Genome-wide C-SWAT library for high-throughput yeast genome tagging. *Nature Methods*, 15(8), 598–600.

Muñoz-Barrera, M., & Monje-Casas, F. (2017). Analysis of cell cycle progression in *Saccharomyces cerevisiae* using the budding index and tubulin staining. *Methods in Molecular Biology*, 1505, 35–44.

Neumüller, R. A., & Knoblich, J. A. (2009). Dividing cellular asymmetry: Asymmetric cell division and its implications for stem cells and cancer. *Genes & Development*, 23(23), 2675–2699.

Neymann, T. C., Wegerhoff, S., & Engell, S. (2011). Synchronization of budding yeast cultures. *IFAC Proceedings Volumes*, 44(1), 8384–8390.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 24(1), 79–88.

Pereira, G., Tanaka, T. U., Nasmyth, K., & Schiebel, E. (2001). Modes of spindle pole body inheritance and segregation of the Bfa1p-Bub2p checkpoint protein complex. *The EMBO Journal*, 20(22), 6359–6370.

Ralser, M., Kuhl, H., Ralser, M., Werber, M., Lehrach, H., Breitenbach, M., & Timmermann, B. (2012). The *Saccharomyces cerevisiae* W303-K6001 cross-platform genome sequence: Insights into ancestry and physiology of a laboratory mutt. *Open Biology*, 2(8), 120093.

Rosebrock, A. P. (2017). Synchronization and arrest of the budding yeast cell cycle using chemical and genetic methods. *Cold Spring Harbor Protocols*, 2017(1). <https://doi.org/10.1101/pdb.prot088724>.

Saka, K., Ide, S., Ganley, A. R. D., & Kobayashi, T. (2013). Cellular senescence in yeast is regulated by rDNA noncoding transcription. *Current Biology*, 23(18), 1794–1798.

Schrader, E. K., Harstad, K. G., & Matouschek, A. (2009). Targeting proteins for degradation. *Nature Chemical Biology*, 5(11), 815–822.

Shahriyari, L., & Komarova, N. L. (2013). Symmetric vs. asymmetric stem cell divisions: An adaptation against cancer? *PLoS One*, *8*(10), e76195.

Shaner, N. C., Lambert, G. G., Chammas, A., Ni, Y., Cranfill, P. J., Baird, M. A., ... & Davidson, M. W. (2013). A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nature Methods*, *10*(5), 407–409.

Shemiakina, I., Ermakova, G., Cranfill, P., et al. (2012). A monomeric red fluorescent protein with low cytotoxicity. *Nature Communications*, *3*, 1204. <https://doi.org/10.1038/ncomms2208>.

Sinclair, D. A., & Guarente, L. (1997). Extrachromosomal rDNA Circles--a Cause of Aging in Yeast. *Cell*, *91*(7), 1033-1042.

Stagge, F., Mitronova, G. Y., Belov, V. N., Wurm, C. A., & Jakobs, S. (2013). SNAP-, CLIP- and Halo-Tag Labelling of Budding Yeast Cells. *PLoS One*, *8*(10), e78745.

Steinkraus, K. A., Kaeberlein, M., & Kennedy, B. K. (2008). Replicative Aging in Yeast: The Means to the End. *Annual Review of Cell and Developmental Biology*, *24*, 29-54.

Sugiyama, S., & Tanaka, M. (2019). Distinct Segregation Patterns of Yeast Cell-Peripheral Proteins Uncovered by a Method for Protein Segregatome Analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(18), 8909-8918.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, *60*, 523-533.

Terpe, K. (2005). Protein Tags. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/3-540-29623-9\\_3650](https://doi.org/10.1007/3-540-29623-9_3650).

Tong, A. H., Evangelista, M., Parsons, A. B., Xu, H., Bader, G. D., Pagé, N., ... Robinson, M. (2001). Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants. *Science*, *294*(5550), 2364-2368.

Tong, A. H. Y., & Boone, C. (2006). Synthetic Genetic Array Analysis in *Saccharomyces Cerevisiae*. *Methods in Molecular Biology*, *313*, 171-192.

Varshavsky, A. (2011). The N-End Rule Pathway and Regulation by Proteolysis. *Protein Science: A Publication of the Protein Society*, *20*(8), 1298-1345.

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*, 116(2), 358-366.

Venkei, Z. G., & Yamashita, Y. M. (2018). Emerging Mechanisms of Asymmetric Stem Cell Division. *The Journal of Cell Biology*, 217(11), 3785-3795.

Vevea, J. D., Swayne, T. C., Boldogh, I. R., & Pon, L. A. (2014). Inheritance of the Fittest Mitochondria in Yeast. *Trends in Cell Biology*, 24(1), 53-60.

Weill, U., Krieger, G., Avihou, Z., Milo, R., Schuldiner, M., & Davidi, D. (2019). Assessment of GFP Tag Position on Protein Localization and Growth Fitness in Yeast. *Journal of Molecular Biology*, 431(3), 636-641.

Willensdorfer, M., Bürger, R., & Nowak, M. A. (2007). Phenotypic Mutation Rates and the Abundance of Abnormal Proteins in Yeast. *PLoS Computational Biology*, 3(11), e203.

Woestenenk, E. A., Hammarström, M., van den Berg, S., Härd, T., & Berglund, H. (2004). His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. *Journal of Structural and Functional Genomics*, 5(3), 217-229. <https://doi.org/10.1023/b:jsfg.0000031965.37625.0e>

Yofe, I., Weill, U., Meurer, M., Chuartzman, S., Zalckvar, E., Goldman, O., ... Ben-Dor, S. (2016). One Library to Make Them All: Streamlining the Creation of Yeast Libraries via a SWAp-Tag Strategy. *Nature Methods*, 13(4), 371-378.

Zhang, H., Linster, E., Gannon, L., Leemhuis, W., Rundle, C. A., Theodoulou, F. L., & Wirtz, M. (2019). Tandem Fluorescent Protein Timers for Noninvasive Relative Protein Lifetime Measurement in Plants. *Plant Physiology*, 180(2), 718-731.







